Københavns Universitet · 19 – 21 July 2017

VISUALIZING MODEL INFERENCE AND ROBUSTNESS

Christopher Adolph

Department of Political Science *and*

Center for Statistics and the Social Sciences

University of Washington, Seattle

CENTER *for* STATISTICS *and the* SOCIAL SCIENCES

Why Scientific Visualization Should Matter to Social Scientists

Good visuals help social science researchers uncover patterns and relationships we'd otherwise miss

Ever more sophisticated statistical models cry out for clear, easy-to-understand visual representations of model findings

Casual observation suggests good visuals have a big impact on audiences for papers and job talks

Puzzle: Social scientists seldom put as much care into designing visual displays as they devote to crafting effective prose

Goal of the Course

We explore visual techniques for summarizing statistical results and efficiently representing their robustness to alternative modeling assumptions

We implement recommended techniques using `R` packages `tile` & `simcf`

These packages and further course materials are at my website:
`faculty.washington.edu/cadolph`

Plan of the Course

19 July	Session 1:	Effective Visual Display of Data
	2:	Concepts for Visualizing Model Inference
20 July	Session 3:	Tools for Visualizing Model Inference
	4:	Concepts for Visualizing Model Robustness
21 July	Session 5:	Tools for Visualizing Model Robustness
	6:	Workshop on Visualizing Model Inference and Robustness

Want to discuss visualization strategies for your research in Session 6?

Send me a 1-page summary tonight. You may also send data and estimated models for possible inclusion as instructor examples. `cadolph@uw.edu`

Plan of Session 1: Effective Visual Display of Data

1. Principles for effective visual display of data
2. Exemplary visual displays; distinction between scientific displays and InfoVis
3. Cognitive science of visual display of data
4. Effective use of color in visual displays of data

In several beautifully illustrated books Edward Tufte gives the following advice:

1. **Assume** the reader's interest and **intelligence**

Edward Tufte and Scientific Visuals

In several beautifully illustrated books Edward Tufte gives the following advice:

1. **Assume** the reader's interest and **intelligence**
2. Maximize information, minimize ink & space: the **data-ink ratio**

Edward Tufte and Scientific Visuals

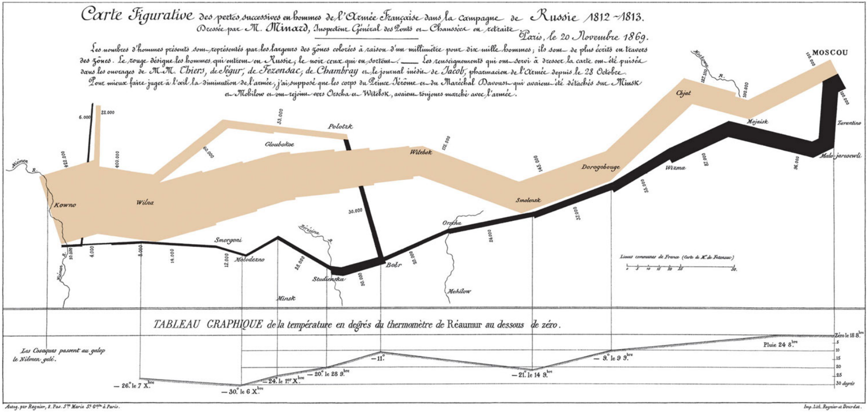
In several beautifully illustrated books Edward Tufte gives the following advice:

1. **Assume** the reader's interest and **intelligence**
2. Maximize information, minimize ink & space: the **data-ink ratio**
3. Show the underlying data and **facilitate comparisons**

Edward Tufte and Scientific Visuals

In several beautifully illustrated books Edward Tufte gives the following advice:

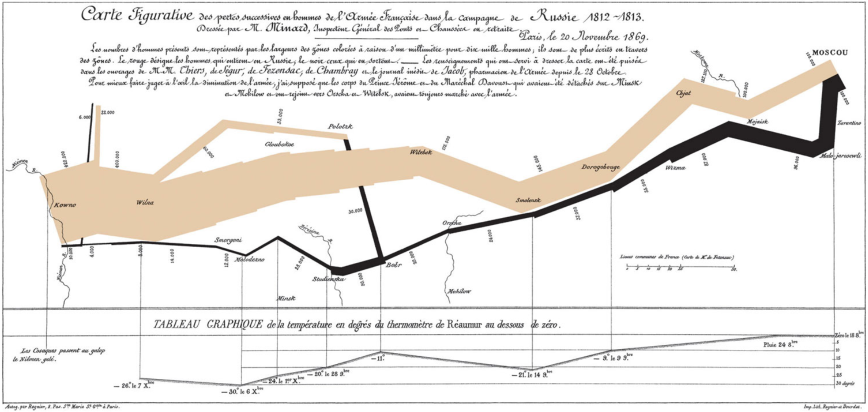
1. **Assume** the reader's interest and **intelligence**
2. Maximize information, minimize ink & space: the **data-ink ratio**
3. Show the underlying data and **facilitate comparisons**
4. Use **small multiples**: repetitions of a basic design



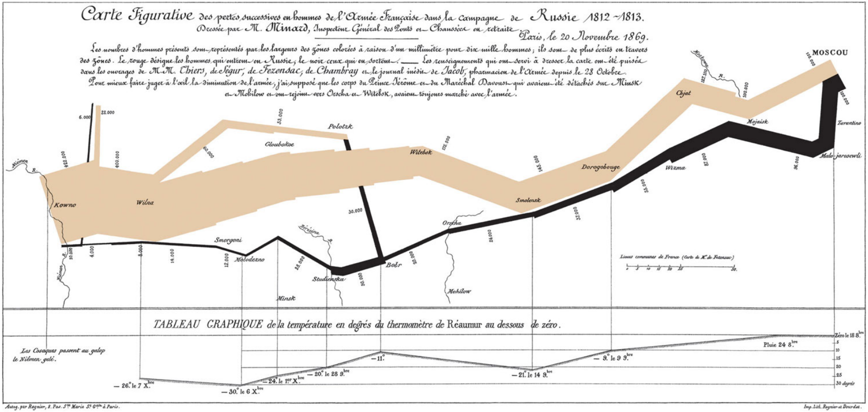
Some plots combine multiple elements into a single "super-plot"

Rich comparisons result if the graphic facilitates comparison across elements

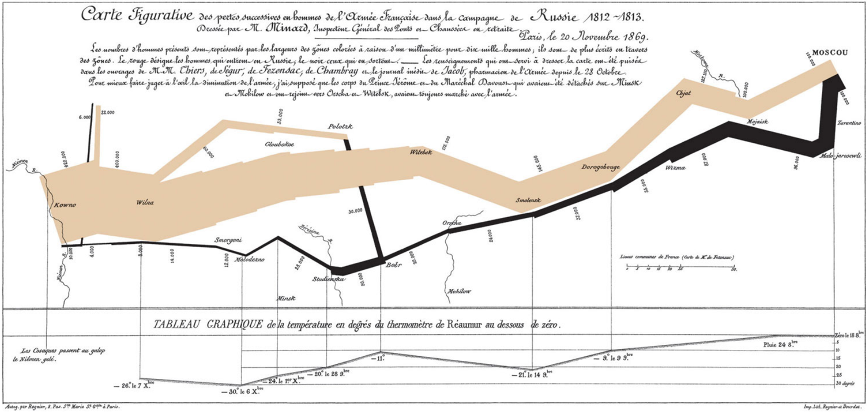
Best example: Minard's display of Napoleon's March on Moscow



How many dimensions?



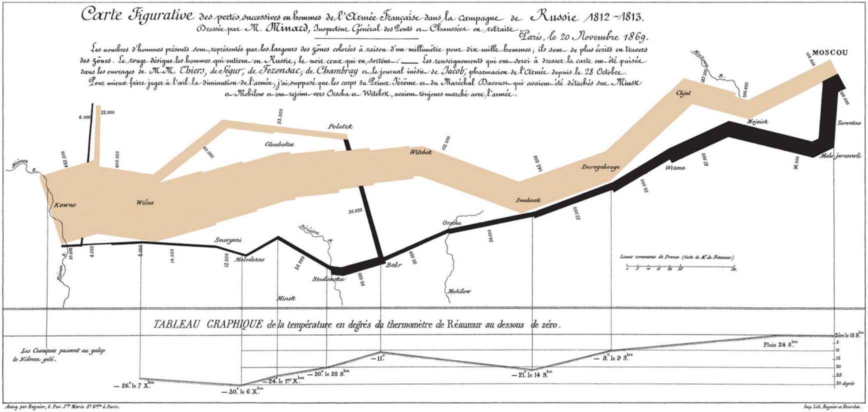
How many dimensions? ***7***



How many dimensions? ***7***

1. Latitude of army & features

   Y-coordinate



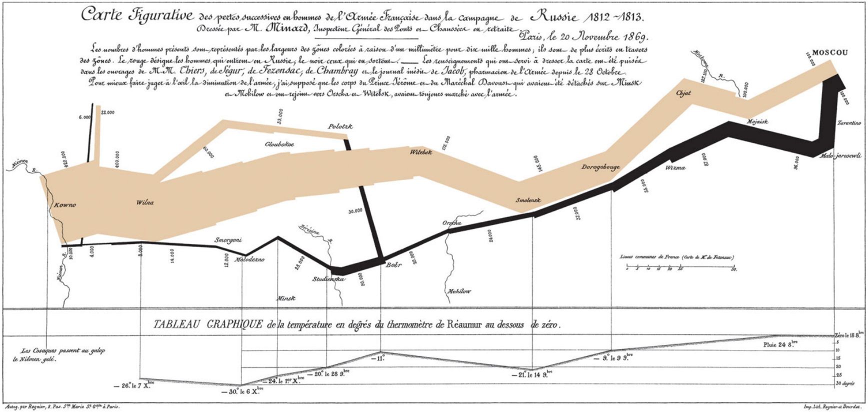
*How many dimensions? **7***

1. Latitude of army & features

   Y-coordinate

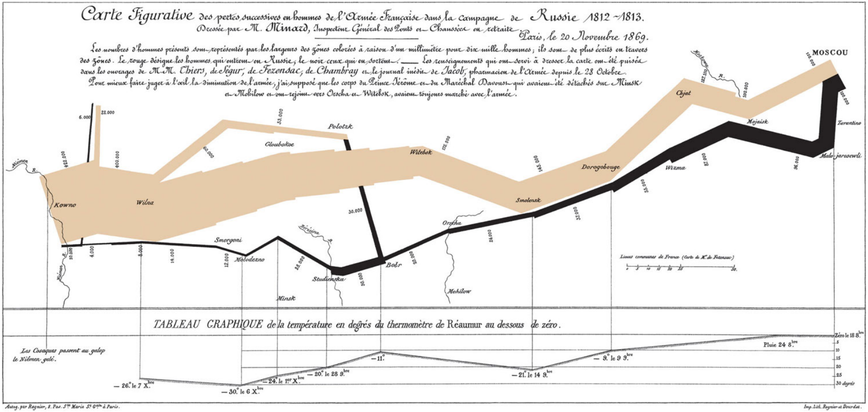
2. Longitude of army & features

   X-coordinate



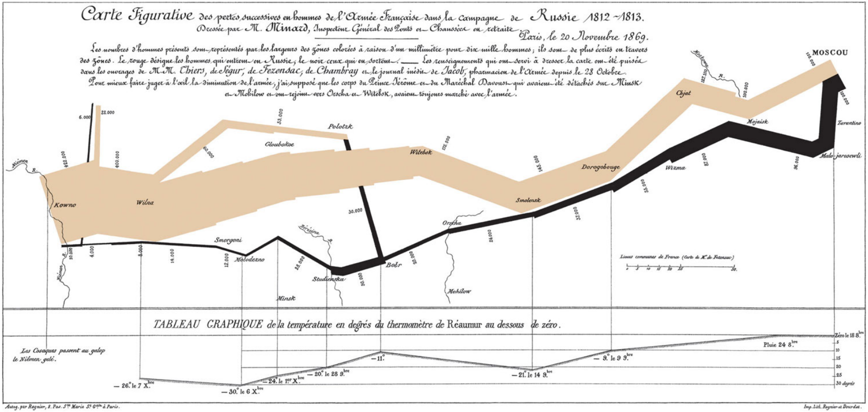
*How many dimensions? **7***

1. Latitude of army & features
   Y-coordinate
2. Longitude of army & features
   X-coordinate
3. Size of army
   width of line, numerals



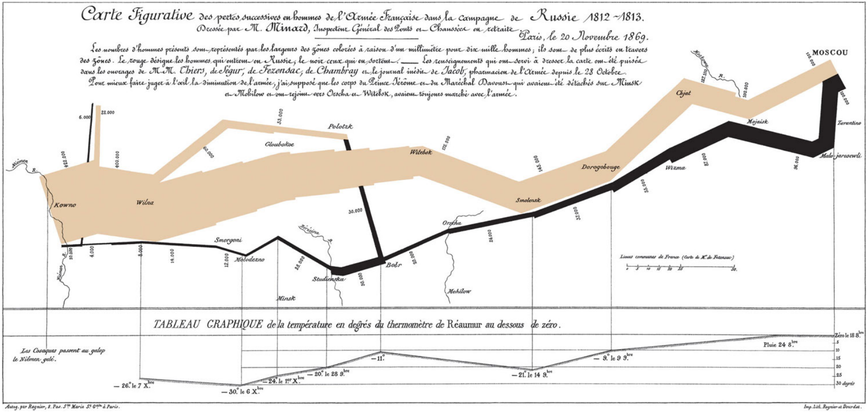
How many dimensions? ***7***

1. Latitude of army & features
   Y-coordinate
2. Longitude of army & features
   X-coordinate
3. Size of army
   width of line, numerals
4. Advance vs. Retreat
   color of line



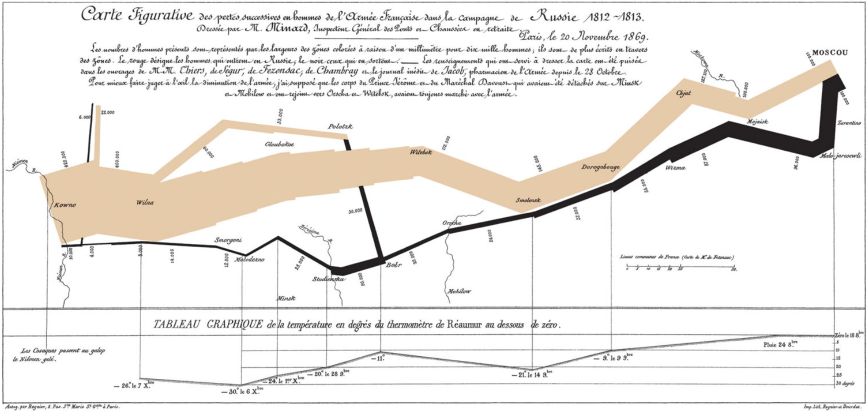
How many dimensions? ***7***

1. Latitude of army & features
   Y-coordinate
2. Longitude of army & features
   X-coordinate
3. Size of army
   width of line, numerals
4. Advance vs. Retreat
   color of line
5. Division of army
   splitting of line



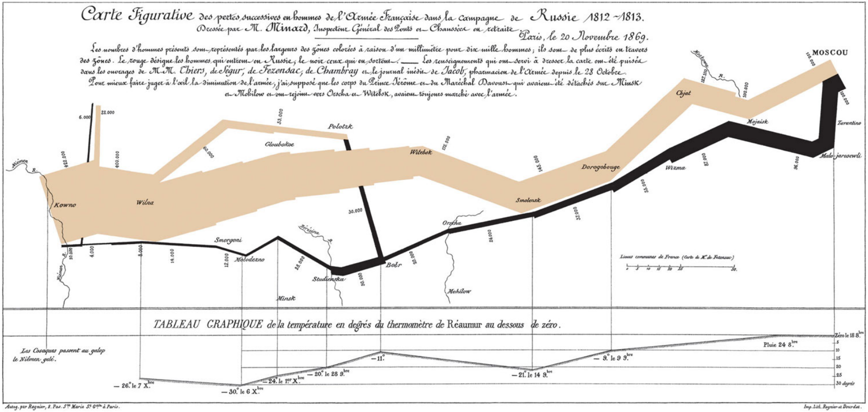
How many dimensions? 7

1. Latitude of army & features
   Y-coordinate
2. Longitude of army & features
   X-coordinate
3. Size of army
   width of line, numerals
4. Advance vs. Retreat
   color of line
5. Division of army
   splitting of line
6. Temperature
   linked lineplot



How many dimensions? ***7***

1. Latitude of army & features
   Y-coordinate
2. Longitude of army & features
   X-coordinate
3. Size of army
   width of line, numerals
4. Advance vs. Retreat
   color of line
5. Division of army
   splitting of line
6. Temperature
   linked lineplot
7. Time
   linked lineplot

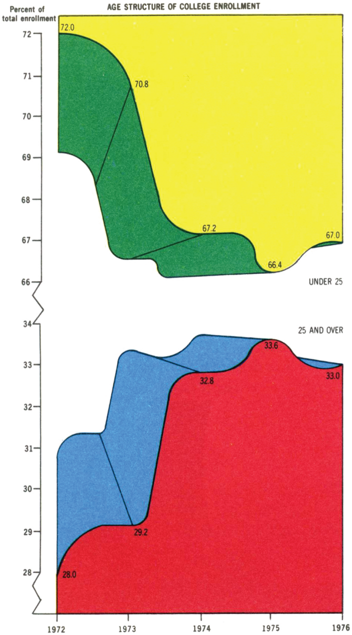


Combines narrative & analysis: a technique mostly lost until this century

May be a spurious relationship here: time and temperature

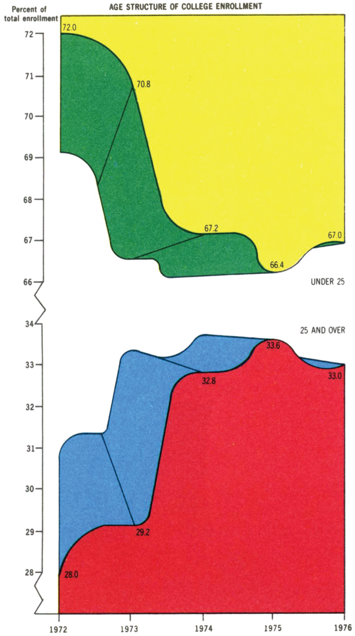
Note the deaths at river crossings – usually, these rivers would be frozen

Did Napoleon choose too warm a winter to invade Russia?



In his first book, Tufte suggested this might be the worst graphic of all time

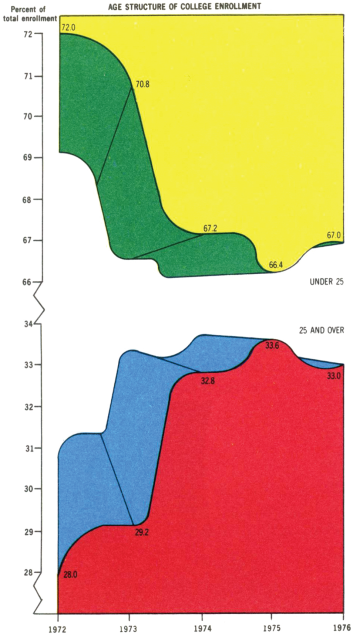
Problems?



In his first book, Tufte suggested this might be the worst graphic of all time

Problems?

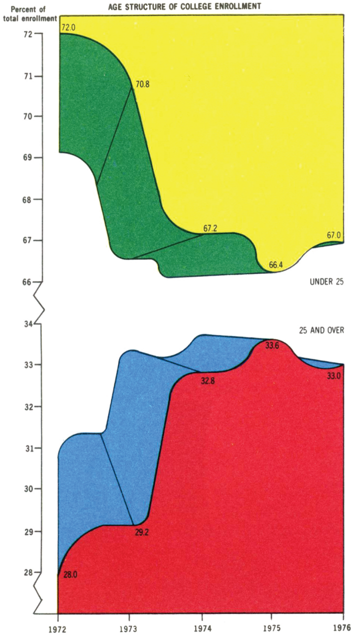
What's the scale?



In his first book, Tufte suggested this might be the worst graphic of all time

Problems?

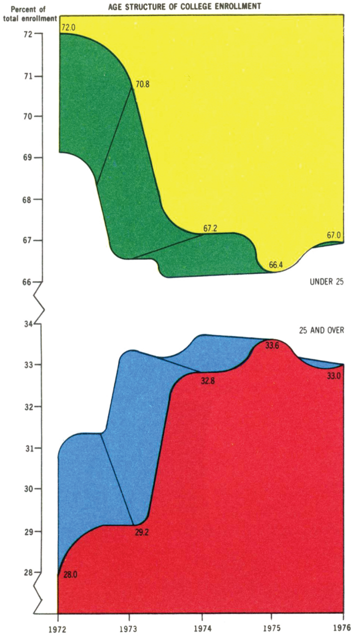
- What's the scale?
- Why the curves?



In his first book, Tufte suggested this might be the worst graphic of all time

Problems?

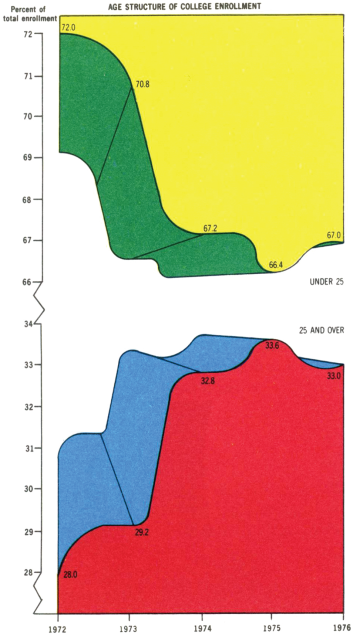
- What's the scale?
- Why the curves?
- Why are there two lines?



In his first book, Tufte suggested this might be the worst graphic of all time

Problems?

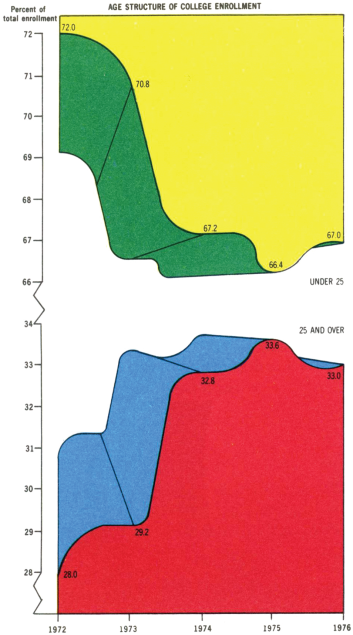
- What's the scale?
- Why the curves?
- Why are there two lines?
- How many data points?



In his first book, Tufte suggested this might be the worst graphic of all time

Problems?

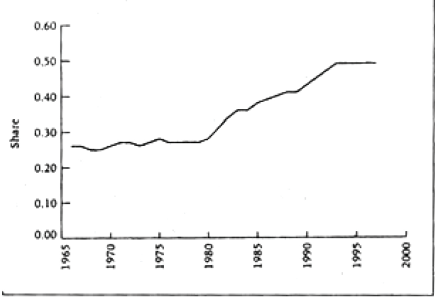
- What's the scale?
- Why the curves?
- Why are there two lines?
- How many data points?
- Why such tiny text?



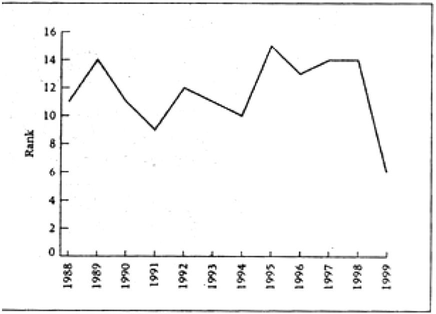
In his first book, Tufte suggested this might be the worst graphic of all time

Problems?

- What's the scale?
- Why the curves?
- Why are there two lines?
- How many data points?
- Why such tiny text?
- Did this need full color?



BY THE NUMBERS: OVER 35 YEARS, CORNELL'S TUITION HAS TAKEN AN INCREASINGLY LARGER SHARE OF ITS MEDIAN STUDENT FAMILY INCOME.

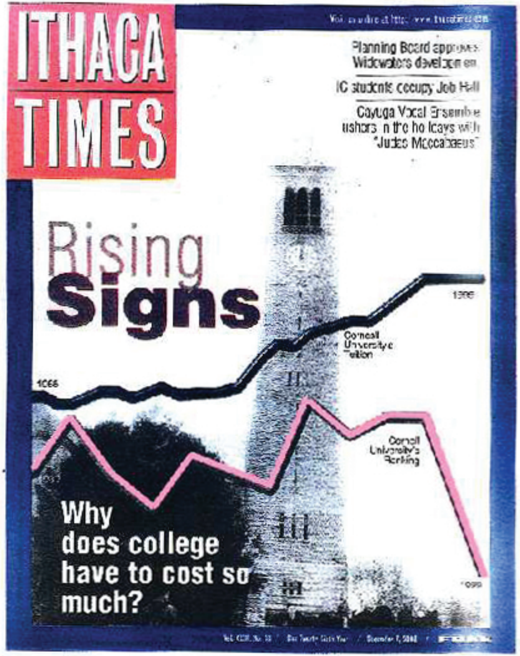


PECKING ORDER: OVER 12 YEARS, CORNELL'S RANKING IN *US NEWS & WORLD REPORT* HAS RISEN AND FALLEN ERRATICALLY.

This graphic may be even worse

What is the claim?

Do you believe it?



An even more misleading combination of these plots on the cover

Scales are what allow us to make comparisons within and *across* graphics

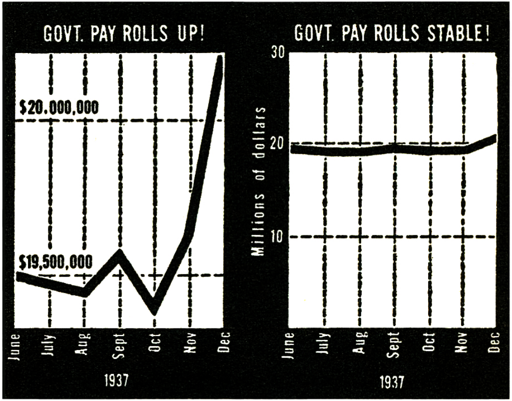
Clearly, careful thought about scaling is essential to making good scientific visuals

A Myth About Scaling

Some people assert as a "rule-of-thumb" that graphical axes must always include zero

Hoff's example at the right is the likely origin of this misleading advice

(To be sure, Hoff never offers it as a rule)



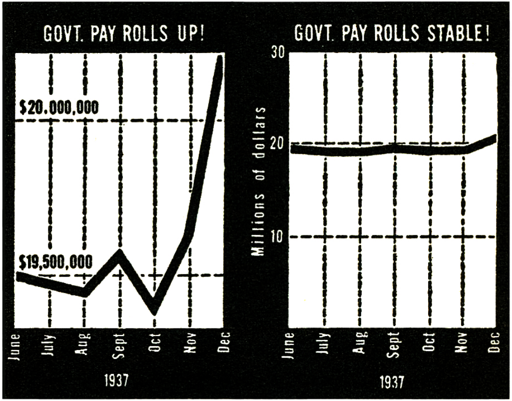
Source: Darrell Hoff, *How to Lie with Statistics*

A Myth About Scaling

Choose scales carefully!

- Start point
- End point
- Units (usual choices:
  linear or log)
- Parallel scales (*optional*)

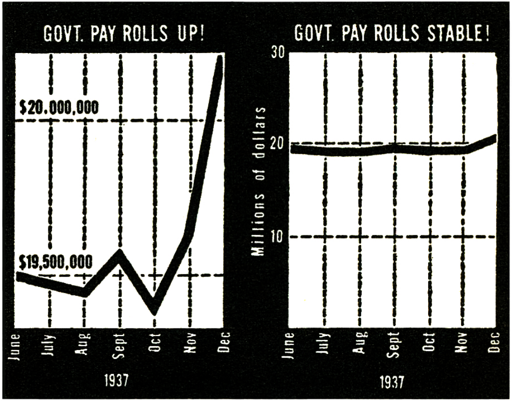
You choices depend on what you want to show and *compare*, not a general rule



Source: Darrell Hoff, *How to Lie with Statistics*

Even in Hoff's example, the left plot is *better* than the right plot

Public budgets are usually very sticky, and 3% changes can be a big deal



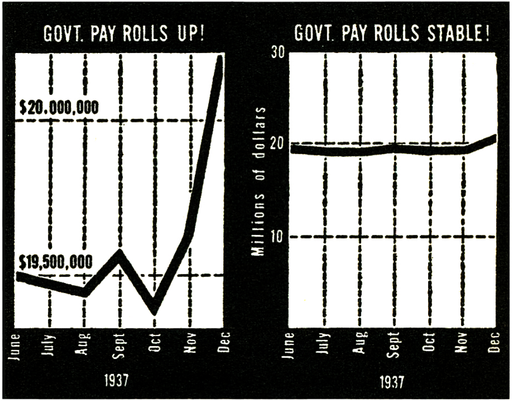
Source: Darrell Hoff, *How to Lie with Statistics*

A Myth About Scaling

Neither plot is ideal – instead, a scale that corresponds to the "usual range" in which budgets might vary would be a better choice

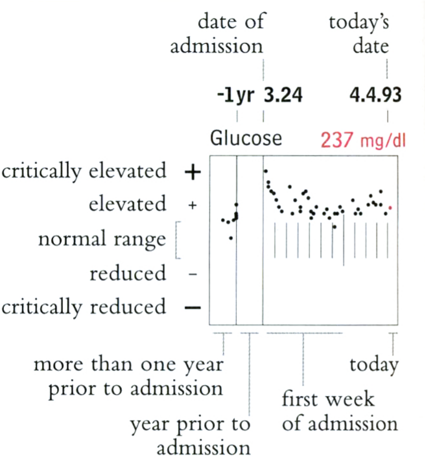
But that suggests the plot is incomplete until compared with *another* set of data

...leading to Tufte's recommendation to plot *small multiples*



Source: Darrell Hoff, *How to Lie with Statistics*

Small Multiples



Tufte proposes medical charts follow the format at left

This chart is annotated for pedagogical purposes

Lots of information; little distracting scaffolding

But the real pay off of this model plot is that it can be repeated once learned...

Small Multiples

A complete layout using small multiples to convey hundreds of pieces of information

Elegant, information-rich, hard to make... Goal: a tool to make this easier

Scientific Visuals and InfoVis

Big data and cheap computing created demand and opportunity for better data visuals in the media

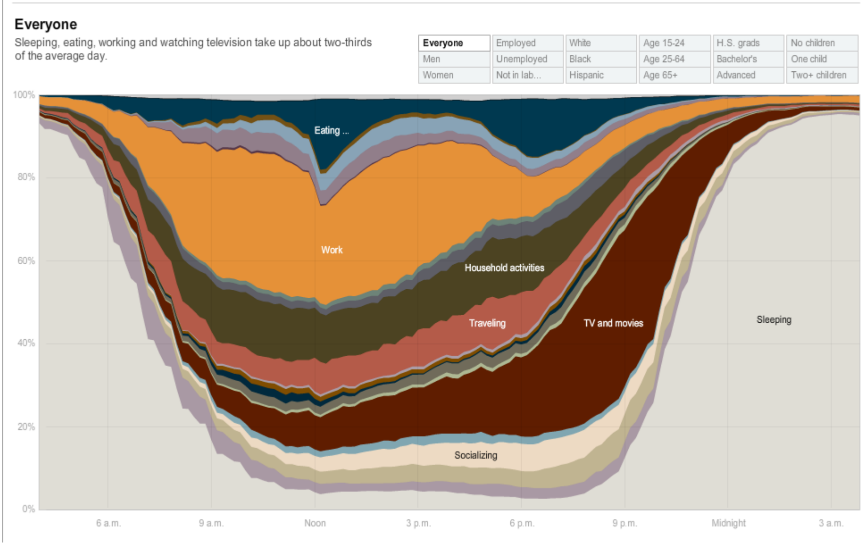
Graphic designers, computer scientists, and journalists have responded:
Information Visualization, or InfoVis

Beautiful, data-rich graphics for exploring public data

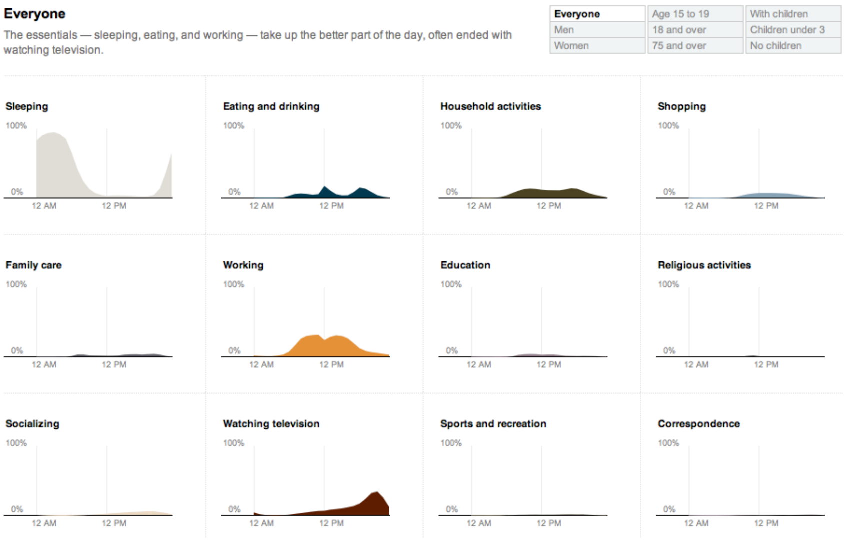
Different goals from scientific visulization of data

InfoVis: emphasis on fun, exploration, beauty, and "wow"

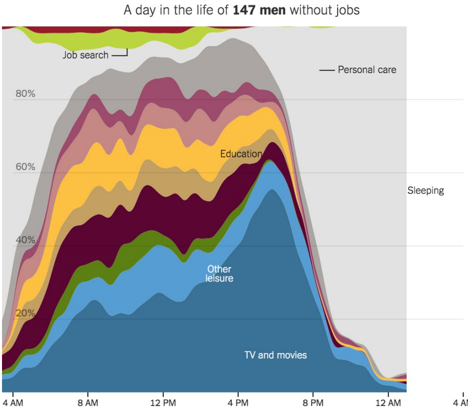
Scientific Visuals: structured comparison, precision, and inference

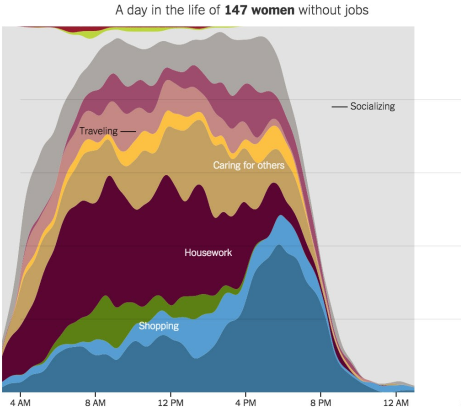


Amanda Cox (CSSS grad) et al, "How different groups spend their day", *New York Times*, `http://www.nytimes.com//interactive/2009/07/31/business/20080801-metrics-graphic.html`



Nathan Yau, "How Americans spend their day," `projects.flowingdata.com/timeuse`



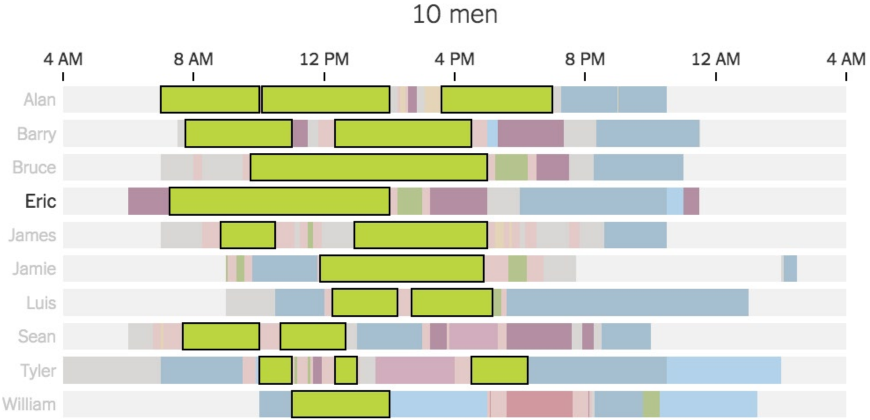


A beautiful synthesis is possible (note the example now focuses on the unemployed)

NYT's Upshot group revisited this problem in January using a blend of data visualization techniques and InfoVis polish

Josh Katz, "How nonemployed Americans spend their weekdays: Men vs. Women," *New York Times*,

`http://www.nytimes.com/interactive/2015/01/06/upshot/how-nonemployed-americans-spend-their-weekdays-men-vs-women.html`



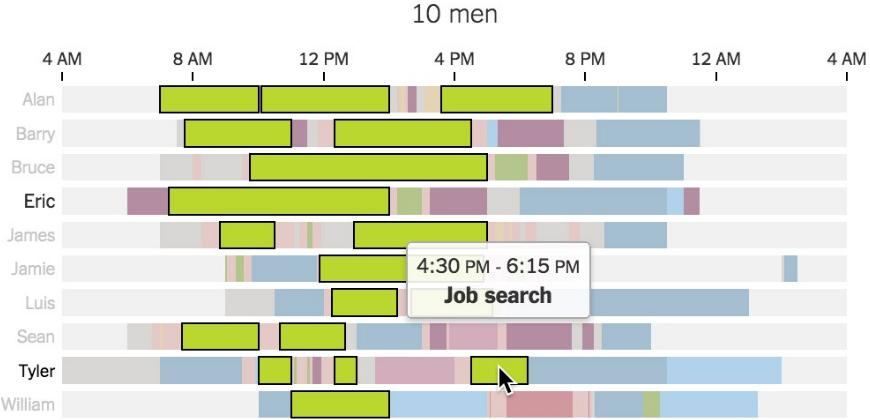
The Upshot's approach follows Tufte's principles:

1. Show as much data as possible

   here, down to lowest level of individual by time by use

Josh Katz, "How nonemployed Americans spend their weekdays: Men vs. Women," *New York Times*,

`http://www.nytimes.com/interactive/2015/01/06/upshot/how-nonemployed-americans-spend-their-weekdays-men-vs-women.html`



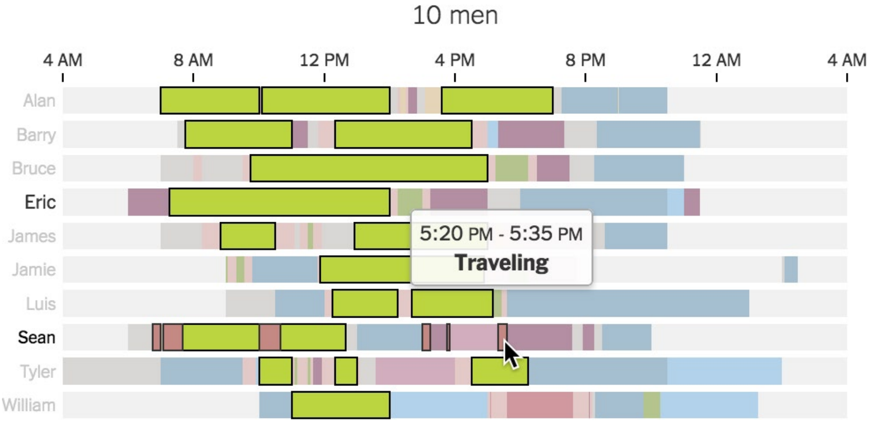
The Upshot's approach follows Tufte's principles:

2. Establish the logic of the graphic with a detailed example

   here, through an interactive display that allows user to select each use

Josh Katz, "How nonemployed Americans spend their weekdays: Men vs. Women," *New York Times*,

`http://www.nytimes.com/interactive/2015/01/06/upshot/how-nonemployed-americans-spend-their-weekdays-men-vs-women.html`



The Upshot's approach follows Tufte's principles:

2. Establish the logic of the graphic with a detailed example
   here, through an interactive display that allows user to select each use

Josh Katz, "How nonemployed Americans spend their weekdays: Men vs. Women," *New York Times*,

`http://www.nytimes.com/interactive/2015/01/06/upshot/how-nonemployed-americans-spend-their-weekdays-men-vs-women.html`

TV and movies

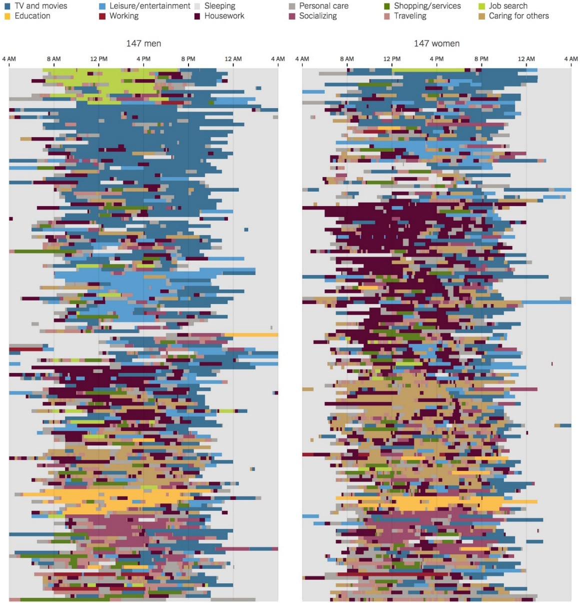
Watching television and movies is a significantly more common activity for the nonemployed than looking for work. For every one person whose main activity was job searching, there were almost six whose main activity was television and movie watching.

The gender breakdown is striking. Of the 65 people who devoted more of their daytime to watching TV and movies than any other activity, 46 are men versus 19 who are women.

With the design established, the Upshot used a series of small multiples to show differences across sexes

Here, they show all respondents whose largest time use was television

Recommendation: sort rows so similar individuals are stacked together (cluster analysis)



Finally, an ambitious graphic shows all data at once

This can only work by *grouping* individuals first by sex (columns of plots),

and then by similarity of time usage (rows of data, sorted via cluster analysis)

Estimates of relative survival rates, by cancer site[9]

	% survival rates and standard errors							
	5 year		10 year		15 year		20 year	
Prostate	98.8	0.4	95.2	0.9	87.1	1.7	81.1	3.0
Thyroid	96.0	0.8	95.8	1.2	94.0	1.6	95.4	2.1
Testis	94.7	1.1	94.0	1.3	91.1	1.8	88.2	2.3
Melanomas	89.0	0.8	86.7	1.1	83.5	1.5	82.8	1.9
Breast	86.4	0.4	78.3	0.6	71.3	0.7	65.0	1.0
Hodgkin's disease	85.1	1.7	79.8	2.0	73.8	2.4	67.1	2.8
Corpus uteri, uterus	84.3	1.0	83.2	1.3	80.8	1.7	79.2	2.0
Urinary, bladder	82.1	1.0	76.2	1.4	70.3	1.9	67.9	2.4
Cervix, uteri	70.5	1.6	64.1	1.8	62.8	2.1	60.0	2.4
Larynx	68.8	2.1	56.7	2.5	45.8	2.8	37.8	3.1
Rectum	62.6	1.2	55.2	1.4	51.8	1.8	49.2	2.3
Kidney, renal pelvis	61.8	1.3	54.4	1.6	49.8	2.0	47.3	2.6
Colon	61.7	0.8	55.4	1.0	53.9	1.2	52.3	1.6
Non-Hodgkin's	57.8	1.0	46.3	1.2	38.3	1.4	34.3	1.7
Oral cavity, pharynx	56.7	1.3	44.2	1.4	37.5	1.6	33.0	1.8
Ovary	55.0	1.3	49.3	1.6	49.9	1.9	49.6	2.4
Leukemia	42.5	1.2	32.4	1.3	29.7	1.5	26.2	1.7
Brain, nervous system	32.0	1.4	29.2	1.5	27.6	1.6	26.1	1.9
Multiple myeloma	29.5	1.6	12.7	1.5	7.0	1.3	4.8	1.5
Stomach	23.8	1.3	19.4	1.4	19.0	1.7	14.9	1.9
Lung and bronchus	15.0	0.4	10.6	0.4	8.1	0.4	6.5	0.4
Esophagus	14.2	1.4	7.9	1.3	7.7	1.6	5.4	2.0
Liver, bile duct	7.5	1.1	5.8	1.2	6.3	1.5	7.6	2.0
Pancreas	4.0	0.5	3.0	1.5	2.7	0.6	2.7	0.8

There are only a few reasons to use a table instead of a graphic:

1. to convey a **handful** of numbers
2. to report precise values for **lookup**
3. to present **many different types** of quantities (i.e., dimensions) for a small number of cases

Usually graphics are more effective than tables

Estimates of relative survival rates, by cancer site[9]

	% survival rates and standard errors							
	5 year		10 year		15 year		20 year	
Prostate	98.8	0.4	95.2	0.9	87.1	1.7	81.1	3.0
Thyroid	96.0	0.8	95.8	1.2	94.0	1.6	95.4	2.1
Testis	94.7	1.1	94.0	1.3	91.1	1.8	88.2	2.3
Melanomas	89.0	0.8	86.7	1.1	83.5	1.5	82.8	1.9
Breast	86.4	0.4	78.3	0.6	71.3	0.7	65.0	1.0
Hodgkin's disease	85.1	1.7	79.8	2.0	73.8	2.4	67.1	2.8
Corpus uteri, uterus	84.3	1.0	83.2	1.3	80.8	1.7	79.2	2.0
Urinary, bladder	82.1	1.0	76.2	1.4	70.3	1.9	67.9	2.4
Cervix, uteri	70.5	1.6	64.1	1.8	62.8	2.1	60.0	2.4
Larynx	68.8	2.1	56.7	2.5	45.8	2.8	37.8	3.1
Rectum	62.6	1.2	55.2	1.4	51.8	1.8	49.2	2.3
Kidney, renal pelvis	61.8	1.3	54.4	1.6	49.8	2.0	47.3	2.6
Colon	61.7	0.8	55.4	1.0	53.9	1.2	52.3	1.6
Non-Hodgkin's	57.8	1.0	46.3	1.2	38.3	1.4	34.3	1.7
Oral cavity, pharynx	56.7	1.3	44.2	1.4	37.5	1.6	33.0	1.8
Ovary	55.0	1.3	49.3	1.6	49.9	1.9	49.6	2.4
Leukemia	42.5	1.2	32.4	1.3	29.7	1.5	26.2	1.7
Brain, nervous system	32.0	1.4	29.2	1.5	27.6	1.6	26.1	1.9
Multiple myeloma	29.5	1.6	12.7	1.5	7.0	1.3	4.8	1.5
Stomach	23.8	1.3	19.4	1.4	19.0	1.7	14.9	1.9
Lung and bronchus	15.0	0.4	10.6	0.4	8.1	0.4	6.5	0.4
Esophagus	14.2	1.4	7.9	1.3	7.7	1.6	5.4	2.0
Liver, bile duct	7.5	1.1	5.8	1.2	6.3	1.5	7.6	2.0
Pancreas	4.0	0.5	3.0	1.5	2.7	0.6	2.7	0.8

Simple ideas for effective tables

1. Minimize the use of guidelines.

Most publishers prohibit vertical lines in tables

Boxes around the whole table are chartjunk

Estimates of relative survival rates, by cancer site[9]

	% survival rates and standard errors							
	5 year		10 year		15 year		20 year	
Prostate	98.8	0.4	95.2	0.9	87.1	1.7	81.1	3.0
Thyroid	96.0	0.8	95.8	1.2	94.0	1.6	95.4	2.1
Testis	94.7	1.1	94.0	1.3	91.1	1.8	88.2	2.3
Melanomas	89.0	0.8	86.7	1.1	83.5	1.5	82.8	1.9
Breast	86.4	0.4	78.3	0.6	71.3	0.7	65.0	1.0
Hodgkin's disease	85.1	1.7	79.8	2.0	73.8	2.4	67.1	2.8
Corpus uteri, uterus	84.3	1.0	83.2	1.3	80.8	1.7	79.2	2.0
Urinary, bladder	82.1	1.0	76.2	1.4	70.3	1.9	67.9	2.4
Cervix, uteri	70.5	1.6	64.1	1.8	62.8	2.1	60.0	2.4
Larynx	68.8	2.1	56.7	2.5	45.8	2.8	37.8	3.1
Rectum	62.6	1.2	55.2	1.4	51.8	1.8	49.2	2.3
Kidney, renal pelvis	61.8	1.3	54.4	1.6	49.8	2.0	47.3	2.6
Colon	61.7	0.8	55.4	1.0	53.9	1.2	52.3	1.6
Non-Hodgkin's	57.8	1.0	46.3	1.2	38.3	1.4	34.3	1.7
Oral cavity, pharynx	56.7	1.3	44.2	1.4	37.5	1.6	33.0	1.8
Ovary	55.0	1.3	49.3	1.6	49.9	1.9	49.6	2.4
Leukemia	42.5	1.2	32.4	1.3	29.7	1.5	26.2	1.7
Brain, nervous system	32.0	1.4	29.2	1.5	27.6	1.6	26.1	1.9
Multiple myeloma	29.5	1.6	12.7	1.5	7.0	1.3	4.8	1.5
Stomach	23.8	1.3	19.4	1.4	19.0	1.7	14.9	1.9
Lung and bronchus	15.0	0.4	10.6	0.4	8.1	0.4	6.5	0.4
Esophagus	14.2	1.4	7.9	1.3	7.7	1.6	5.4	2.0
Liver, bile duct	7.5	1.1	5.8	1.2	6.3	1.5	7.6	2.0
Pancreas	4.0	0.5	3.0	1.5	2.7	0.6	2.7	0.8

Simple ideas for effective tables

2. Report only a few digits.

Don't report non-significant digits

Every extra digit distracts attention from the first, most important one

Estimates of relative survival rates, by cancer site[9]

	% survival rates and standard errors							
	5 year		10 year		15 year		20 year	
Prostate	98.8	0.4	95.2	0.9	87.1	1.7	81.1	3.0
Thyroid	96.0	0.8	95.8	1.2	94.0	1.6	95.4	2.1
Testis	94.7	1.1	94.0	1.3	91.1	1.8	88.2	2.3
Melanomas	89.0	0.8	86.7	1.1	83.5	1.5	82.8	1.9
Breast	86.4	0.4	78.3	0.6	71.3	0.7	65.0	1.0
Hodgkin's disease	85.1	1.7	79.8	2.0	73.8	2.4	67.1	2.8
Corpus uteri, uterus	84.3	1.0	83.2	1.3	80.8	1.7	79.2	2.0
Urinary, bladder	82.1	1.0	76.2	1.4	70.3	1.9	67.9	2.4
Cervix, uteri	70.5	1.6	64.1	1.8	62.8	2.1	60.0	2.4
Larynx	68.8	2.1	56.7	2.5	45.8	2.8	37.8	3.1
Rectum	62.6	1.2	55.2	1.4	51.8	1.8	49.2	2.3
Kidney, renal pelvis	61.8	1.3	54.4	1.6	49.8	2.0	47.3	2.6
Colon	61.7	0.8	55.4	1.0	53.9	1.2	52.3	1.6
Non-Hodgkin's	57.8	1.0	46.3	1.2	38.3	1.4	34.3	1.7
Oral cavity, pharynx	56.7	1.3	44.2	1.4	37.5	1.6	33.0	1.8
Ovary	55.0	1.3	49.3	1.6	49.9	1.9	49.6	2.4
Leukemia	42.5	1.2	32.4	1.3	29.7	1.5	26.2	1.7
Brain, nervous system	32.0	1.4	29.2	1.5	27.6	1.6	26.1	1.9
Multiple myeloma	29.5	1.6	12.7	1.5	7.0	1.3	4.8	1.5
Stomach	23.8	1.3	19.4	1.4	19.0	1.7	14.9	1.9
Lung and bronchus	15.0	0.4	10.6	0.4	8.1	0.4	6.5	0.4
Esophagus	14.2	1.4	7.9	1.3	7.7	1.6	5.4	2.0
Liver, bile duct	7.5	1.1	5.8	1.2	6.3	1.5	7.6	2.0
Pancreas	4.0	0.5	3.0	1.5	2.7	0.6	2.7	0.8

Simple ideas for effective tables

3. Order the table intelligently.

In a 2 dimensional table, order the rows and columns to highlight relationships

You can either

- **diagonalize** – sort based on order, or
- **cluster** – group based on similarity

Estimates of relative survival rates, by cancer site[9]

	% survival rates and standard errors							
	5 year		10 year		15 year		20 year	
Prostate	98.8	0.4	95.2	0.9	87.1	1.7	81.1	3.0
Thyroid	96.0	0.8	95.8	1.2	94.0	1.6	95.4	2.1
Testis	94.7	1.1	94.0	1.3	91.1	1.8	88.2	2.3
Melanomas	89.0	0.8	86.7	1.1	83.5	1.5	82.8	1.9
Breast	86.4	0.4	78.3	0.6	71.3	0.7	65.0	1.0
Hodgkin's disease	85.1	1.7	79.8	2.0	73.8	2.4	67.1	2.8
Corpus uteri, uterus	84.3	1.0	83.2	1.3	80.8	1.7	79.2	2.0
Urinary, bladder	82.1	1.0	76.2	1.4	70.3	1.9	67.9	2.4
Cervix, uteri	70.5	1.6	64.1	1.8	62.8	2.1	60.0	2.4
Larynx	68.8	2.1	56.7	2.5	45.8	2.8	37.8	3.1
Rectum	62.6	1.2	55.2	1.4	51.8	1.8	49.2	2.3
Kidney, renal pelvis	61.8	1.3	54.4	1.6	49.8	2.0	47.3	2.6
Colon	61.7	0.8	55.4	1.0	53.9	1.2	52.3	1.6
Non-Hodgkin's	57.8	1.0	46.3	1.2	38.3	1.4	34.3	1.7
Oral cavity, pharynx	56.7	1.3	44.2	1.4	37.5	1.6	33.0	1.8
Ovary	55.0	1.3	49.3	1.6	49.9	1.9	49.6	2.4
Leukemia	42.5	1.2	32.4	1.3	29.7	1.5	26.2	1.7
Brain, nervous system	32.0	1.4	29.2	1.5	27.6	1.6	26.1	1.9
Multiple myeloma	29.5	1.6	12.7	1.5	7.0	1.3	4.8	1.5
Stomach	23.8	1.3	19.4	1.4	19.0	1.7	14.9	1.9
Lung and bronchus	15.0	0.4	10.6	0.4	8.1	0.4	6.5	0.4
Esophagus	14.2	1.4	7.9	1.3	7.7	1.6	5.4	2.0
Liver, bile duct	7.5	1.1	5.8	1.2	6.3	1.5	7.6	2.0
Pancreas	4.0	0.5	3.0	1.5	2.7	0.6	2.7	0.8

Simple ideas for effective tables

3. Order the table intelligently.

In a 3+ dimensional table, nest the dimensions intelligently.

Note:

Table order applies to
1-dimensional plots,
like dot plots...

and to super tables of plots where rows or columns are categories

Estimates of relative survival rates, by cancer site[9]

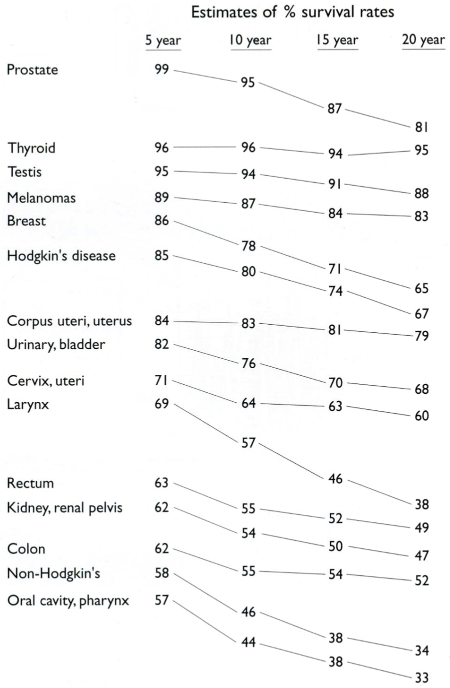
	% survival rates and standard errors							
	5 year		10 year		15 year		20 year	
Prostate	98.8	0.4	95.2	0.9	87.1	1.7	81.1	3.0
Thyroid	96.0	0.8	95.8	1.2	94.0	1.6	95.4	2.1
Testis	94.7	1.1	94.0	1.3	91.1	1.8	88.2	2.3
Melanomas	89.0	0.8	86.7	1.1	83.5	1.5	82.8	1.9
Breast	86.4	0.4	78.3	0.6	71.3	0.7	65.0	1.0
Hodgkin's disease	85.1	1.7	79.8	2.0	73.8	2.4	67.1	2.8
Corpus uteri, uterus	84.3	1.0	83.2	1.3	80.8	1.7	79.2	2.0
Urinary, bladder	82.1	1.0	76.2	1.4	70.3	1.9	67.9	2.4
Cervix, uteri	70.5	1.6	64.1	1.8	62.8	2.1	60.0	2.4
Larynx	68.8	2.1	56.7	2.5	45.8	2.8	37.8	3.1
Rectum	62.6	1.2	55.2	1.4	51.8	1.8	49.2	2.3
Kidney, renal pelvis	61.8	1.3	54.4	1.6	49.8	2.0	47.3	2.6
Colon	61.7	0.8	55.4	1.0	53.9	1.2	52.3	1.6
Non-Hodgkin's	57.8	1.0	46.3	1.2	38.3	1.4	34.3	1.7
Oral cavity, pharynx	56.7	1.3	44.2	1.4	37.5	1.6	33.0	1.8
Ovary	55.0	1.3	49.3	1.6	49.9	1.9	49.6	2.4
Leukemia	42.5	1.2	32.4	1.3	29.7	1.5	26.2	1.7
Brain, nervous system	32.0	1.4	29.2	1.5	27.6	1.6	26.1	1.9
Multiple myeloma	29.5	1.6	12.7	1.5	7.0	1.3	4.8	1.5
Stomach	23.8	1.3	19.4	1.4	19.0	1.7	14.9	1.9
Lung and bronchus	15.0	0.4	10.6	0.4	8.1	0.4	6.5	0.4
Esophagus	14.2	1.4	7.9	1.3	7.7	1.6	5.4	2.0
Liver, bile duct	7.5	1.1	5.8	1.2	6.3	1.5	7.6	2.0
Pancreas	4.0	0.5	3.0	1.5	2.7	0.6	2.7	0.8

The table at left (from Tufte) is effectively designed

It is *diagonalized*, uses *few digits*, and facilitates *lookup*

But tables always limit *comparison*

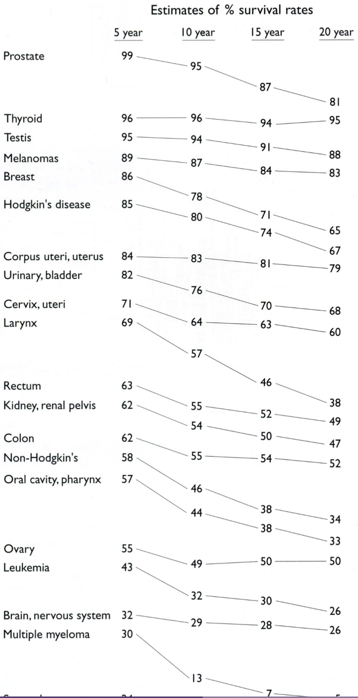
The brain is slower to grasp numerals than graphical representations of numbers



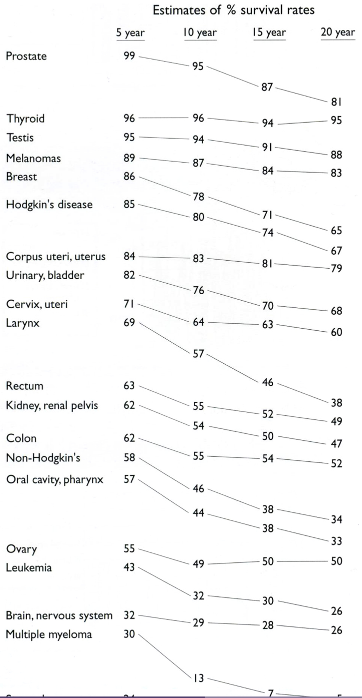
This figure (also from Tufte) is an improvement

It keeps (almost) all the virtues of the table, but also makes comparison easier

Instead of digging information out of the table, it now hits the reader "between the eyes"



What’s missing from the figure that was in the table?



What's missing from the figure that was in the table?

Measures of uncertainty.

The table had standard errors

A major focus of this short course is including uncertainty in plots like this one

Turning Tables into Graphs: European Health Authority

		National	Regional	Local
Austria		**National**	**Regional (Lander)**	**l (stadt/Gemind**
	Pharmaceutical	L1/L2/F/R		
	Primary care	L1/F	L2/P	
	Sec and Ter care	L1/F	L2/P	
	Public Health	F	L1	L2/P
Belgium		**National**	**Regional (Regions)**	**cal (Communes)**
	Pharmaceutical	L1/L2/F/R		
	Primary care	L1/F	L2/P	
	Sec and Ter care	L1/F	L2/P	
	Public Health	L1/F	L2/P	
Bulgaria		**National**	**Regional (Oblasti)**	**ocal (Obshtina)**
	Pharmaceutical	L1/L2//R/F		
	Primary care	L1/L2/F		P
	Sec and Ter care	L1/L2/F		P
	Public Health	L1/L2/F/P		
Cyprus		**National**	**No elected regional level**	**ocal (eparchies)**
	Pharmaceutical	L1/L2/F/R		
	Primary care	L1/L2/F/P		
	Sec and Ter care	L1/L2/F/P		
	Public Health	L1/L2/F/P		

An example from work I did with Scott Greer and Elize Fonseca ("Allocation of Authority in European Health Policy," *Social Science and Medicine*, 2012)

Turning Tables into Graphs: European Health Authority

Austria		**National**	**Regional (Lander)**	**l (stadt/Gemind**
	Pharmaceutical	L1/L2/F/R		
	Primary care	L1/F	L2/P	
	Sec and Ter care	L1/F	L2/P	
	Public Health	F	L1	L2/P
Belgium		**National**	**Regional (Regions)**	**cal (Communes**
	Pharmaceutical	L1/L2/F/R		
	Primary care	L1/F	L2/P	
	Sec and Ter care	L1/F	L2/P	
	Public Health	L1/F	L2/P	
Bulgaria		**National**	**Regional (Oblasti)**	**ocal (Obshtin)**
	Pharmaceutical	L1/L2//R/F		
	Primary care	L1/L2/F		P
	Sec and Ter care	L1/L2/F		P
	Public Health	L1/L2/F/P		
Cyprus		**National**	**No elected regional level**	**ocal (eparchies)**
	Pharmaceutical	L1/L2/F/R		
	Primary care	L1/L2/F/P		
	Sec and Ter care	L1/L2/F/P		
	Public Health	L1/L2/F/P		

Our data code the level of goverment at which European countries lodged authority over specific health policies

DV: level of government state, regional, or local

Country		National	Regional	Local
Austria		**National**	**Regional (Lander)**	**l (stadt/Gemind**
	Pharmaceutical	L1/L2/F/R		
	Primary care	L1/F	L2/P	
	Sec and Ter care	L1/F	L2/P	
	Public Health	F	L1	L2/P
Belgium		**National**	**Regional (Regions)**	**cal (Communes**
	Pharmaceutical	L1/L2/F/R		
	Primary care	L1/F	L2/P	
	Sec and Ter care	L1/F	L2/P	
	Public Health	L1/F	L2/P	
Bulgaria		**National**	**Regional (Oblasti)**	**ocal (Obshtin)**
	Pharmaceutical	L1/L2//R/F		
	Primary care	L1/L2/F		P
	Sec and Ter care	L1/L2/F		P
	Public Health	L1/L2/F/P		
Cyprus		**National**	**No elected regional level**	**ocal (eparchies)**
	Pharmaceutical	L1/L2/F/R		
	Primary care	L1/L2/F/P		
	Sec and Ter care	L1/L2/F/P		
	Public Health	L1/L2/F/P		

The policies are defined along two dimensions, policy area and policy tool:

IV: policy area pharma, primary care, secondary/tertiary care, or public health

IV: policy tool framework (L1), implementation (L2), finance, regulation, provision

Turning Tables into Graphs: European Health Authority

		National	Regional (Lander)	al (stadt/Gemind
Austria	Pharmaceutical	L1/L2/F/R		
	Primary care	L1/F	L2/P	
	Sec and Ter care	L1/F	L2/P	
	Public Health	F	L1	L2/P
Belgium		**National**	**Regional (Regions)**	**cal (Communes)**
	Pharmaceutical	L1/L2/F/R		
	Primary care	L1/F	L2/P	
	Sec and Ter care	L1/F	L2/P	
	Public Health	L1/F	L2/P	
Bulgaria		**National**	**Regional (Oblasti)**	**ocal (Obshtina)**
	Pharmaceutical	L1/L2//R/F		
	Primary care	L1/L2/F		P
	Sec and Ter care	L1/L2/F		P
	Public Health	L1/L2/F/P		
Cyprus		**National**	**No elected regional level**	**ocal (eparchies)**
	Pharmaceutical	L1/L2/F/R		
	Primary care	L1/L2/F/P		
	Sec and Ter care	L1/L2/F/P		
	Public Health	L1/L2/F/P		

We ask to which level (state, regional, or local) a country allocates authority

E.g., a country might use to finance public health

Later, we will model the data to find whether certain areas or tools lodge at certain levels across countries, controlling for other covariates

Austria		**National**	**Regional (Lander)**	**l (stadt/Gemind**
	Pharmaceutical	L1/L2/F/R		
	Primary care	L1/F	L2/P	
	Sec and Ter care	L1/F	L2/P	
	Public Health	F	L1	L2/P
Belgium		**National**	**Regional (Regions)**	**cal (Communes)**
	Pharmaceutical	L1/L2/F/R		
	Primary care	L1/F	L2/P	
	Sec and Ter care	L1/F	L2/P	
	Public Health	L1/F	L2/P	
Bulgaria		**National**	**Regional (Oblasti)**	**ocal (Obshtina)**
	Pharmaceutical	L1/L2//R/F		
	Primary care	L1/L2/F		P
	Sec and Ter care	L1/L2/F		P
	Public Health	L1/L2/F/P		
Cyprus		**National**	**No elected regional level**	**ocal (eparchies)**
	Pharmaceutical	L1/L2/F/R		
	Primary care	L1/L2/F/P		
	Sec and Ter care	L1/L2/F/P		
	Public Health	L1/L2/F/P		

For now, let's do a little *exploratory data analysis* (EDA)

Is there even interesting variation in the allocation of authority to explain?

Turning Tables into Graphs: European Health Authority

		National	Regional (Lander)	l (stadt/Gemind
Austria	Pharmaceutical	L1/L2/F/R		
	Primary care	L1/F	L2/P	
	Sec and Ter care	L1/F	L2/P	
	Public Health	F	L1	L2/P
Belgium		**National**	**Regional (Regions)**	**cal (Communes)**
	Pharmaceutical	L1/L2/F/R		
	Primary care	L1/F	L2/P	
	Sec and Ter care	L1/F	L2/P	
	Public Health	L1/F	L2/P	
Bulgaria		**National**	**Regional (Oblasti)**	**ocal (Obshtina)**
	Pharmaceutical	L1/L2//R/F		
	Primary care	L1/L2/F		P
	Sec and Ter care	L1/L2/F		P
	Public Health	L1/L2/F/P		
Cyprus		**National**	**No elected regional level**	**ocal (eparchies)**
	Pharmaceutical	L1/L2/F/R		
	Primary care	L1/L2/F/P		
	Sec and Ter care	L1/L2/F/P		
	Public Health	L1/L2/F/P		

Problems with this table?

1. Poor distincton between dependent variable & covariates: cell entries code policy tools (IV), rows code policy areas (IV), columns code level (DV)

Turning Tables into Graphs: European Health Authority

		National	Regional (Lander)	al (stadt/Gemind
Austria				
	Pharmaceutical	L1/L2/F/R		
	Primary care	L1/F	L2/P	
	Sec and Ter care	L1/F	L2/P	
	Public Health	F	L1	L2/P
Belgium		**National**	**Regional (Regions)**	**cal (Communes)**
	Pharmaceutical	L1/L2/F/R		
	Primary care	L1/F	L2/P	
	Sec and Ter care	L1/F	L2/P	
	Public Health	L1/F	L2/P	
Bulgaria		**National**	**Regional (Oblasti)**	**ocal (Obshtina)**
	Pharmaceutical	L1/L2//R/F		
	Primary care	L1/L2/F		P
	Sec and Ter care	L1/L2/F		P
	Public Health	L1/L2/F/P		
Cyprus		**National**	**No elected regional level**	**ocal (eparchies)**
	Pharmaceutical	L1/L2/F/R		
	Primary care	L1/L2/F/P		
	Sec and Ter care	L1/L2/F/P		
	Public Health	L1/L2/F/P		

Problems with this table?

2. Alphabetical listing of cases is inefficient for comparison and detection of patterns of variation, such as clusters of similar countries

Turning Tables into Graphs: European Health Authority

		National	Regional (Lander)	l (stadt/Gemind
Austria	Pharmaceutical	L1/L2/F/R		
	Primary care	L1/F	L2/P	
	Sec and Ter care	L1/F	L2/P	
	Public Health	F	L1	L2/P
Belgium		National	Regional (Regions)	cal (Communes)
	Pharmaceutical	L1/L2/F/R		
	Primary care	L1/F	L2/P	
	Sec and Ter care	L1/F	L2/P	
	Public Health	L1/F	L2/P	
Bulgaria		National	Regional (Oblasti)	ocal (Obshtina)
	Pharmaceutical	L1/L2//R/F		
	Primary care	L1/L2/F		P
	Sec and Ter care	L1/L2/F		P
	Public Health	L1/L2/F/P		
Cyprus		National	No elected regional level	ocal (eparchies)
	Pharmaceutical	L1/L2/F/R		
	Primary care	L1/L2/F/P		
	Sec and Ter care	L1/L2/F/P		
	Public Health	L1/L2/F/P		

Problems with this table?

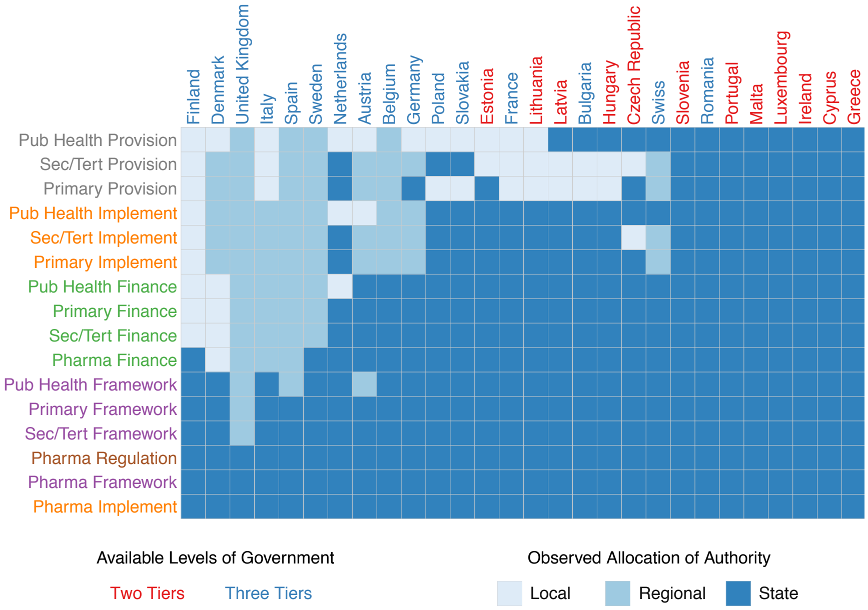
3. Coding system is opaque (L1? L2? etc.)

Turning Tables into Graphs: European Health Authority

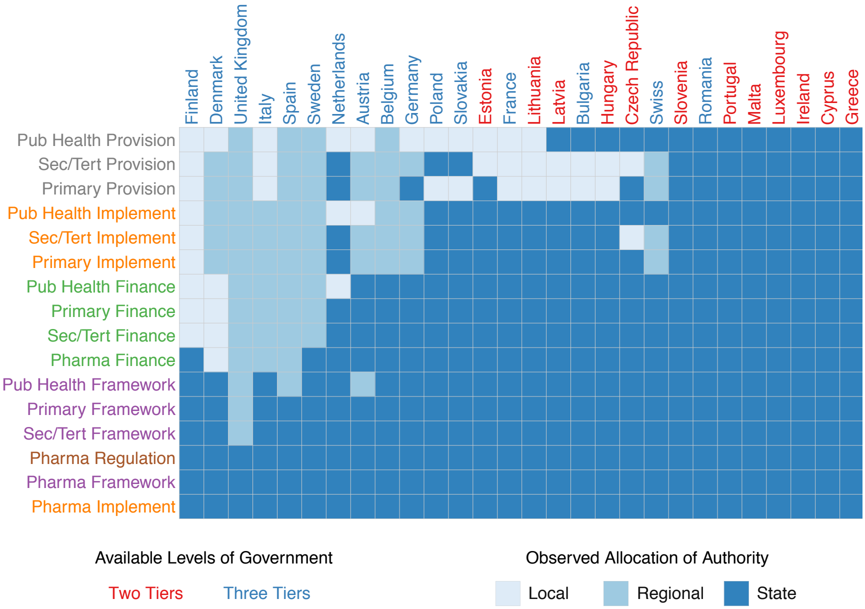
Austria		**National**	**Regional (Lander)**	**l (stadt/Gemind**
	Pharmaceutical	L1/L2/F/R		
	Primary care	L1/F	L2/P	
	Sec and Ter care	L1/F	L2/P	
	Public Health	F	L1	L2/P
Belgium		**National**	**Regional (Regions)**	**cal (Communes**
	Pharmaceutical	L1/L2/F/R		
	Primary care	L1/F	L2/P	
	Sec and Ter care	L1/F	L2/P	
	Public Health	L1/F	L2/P	
Bulgaria		**National**	**Regional (Oblasti)**	**ocal (Obshtina)**
	Pharmaceutical	L1/L2//R/F		
	Primary care	L1/L2/F		P
	Sec and Ter care	L1/L2/F		P
	Public Health	L1/L2/F/P		
Cyprus		**National**	**No elected regional level**	**ocal (eparchies)**
	Pharmaceutical	L1/L2/F/R		
	Primary care	L1/L2/F/P		
	Sec and Ter care	L1/L2/F/P		
	Public Health	L1/L2/F/P		

Problems with this table?

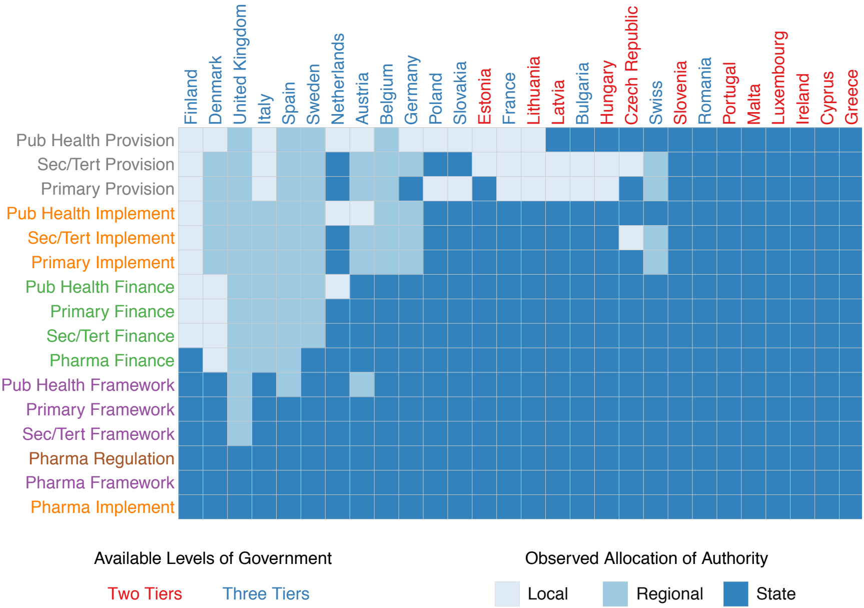
4. Table goes on for four pages! Hard to spot patterns



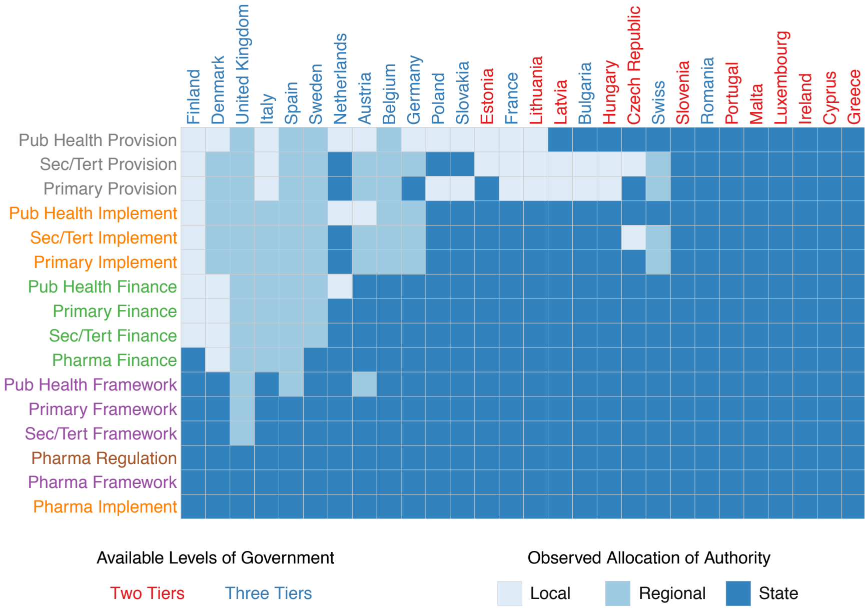
Let's redraw the table as a heatmap



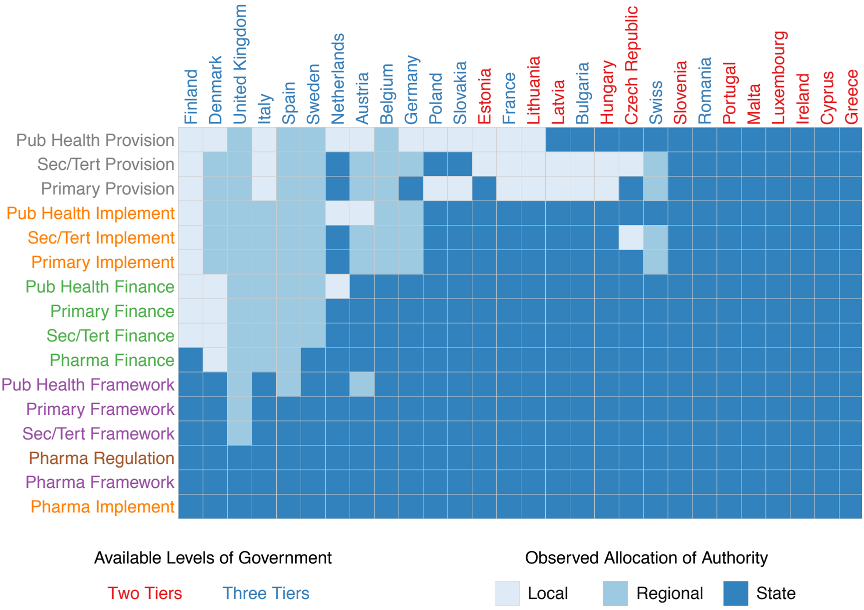
Allocation of authority is highlighted as the shading of each cell



Rows and columns code covariates, and are sorted by cluster analysis



Rows and columns colors help detect patterns in clustering



We did not sort by row color, so any emergent pattern is noteworthy

The Cognitive Science of Visual Displays of Information

Suppose we design the most beautiful, data rich display we can

But we use elements that humans can't perceive: Ultraviolet light.

The limits of human vision render our display useless

Suppose we design the most beautiful, data rich display we can

But we use elements that humans can't perceive: Ultraviolet light.

The limits of human vision render our display useless

Now suppose we design the most beautiful, data rich display we can

We use elements humans *can* perceive, but get systematically wrong

No better? Even worse?

Suppose we design the most beautiful, data rich display we can

But we use elements that humans can't perceive: Ultraviolet light.

The limits of human vision render our display useless

Now suppose we design the most beautiful, data rich display we can

We use elements humans *can* perceive, but get systematically wrong

No better? Even worse?

Unfortunately, cognitive errors are everywhere

But few designers of scientific visuals pay close attention to them

Unfortunately, cognitive errors are everywhere

Ideally, we would have an algorthm to accomplish the following:

```
cognitivelyAdjustedGraphic <- correctForErrors(InitialGraphic)
```

Unfortunately, cognitive errors are everywhere

Ideally, we would have an algorthm to accomplish the following:

```
cognitivelyAdjustedGraphic <- correctForErrors(InitialGraphic)
```

Alas, this does not exist

The cognitive study of graphics is difficult

Hard to systematically understanding how graphical elements combine & interact

Instead, many specific errors known from experiments

These experiments provide warnings about dangerous techniques

To minimize error, we can try to use more reliable graphical elements

The Cognitive Science of Visual Displays of Information

Graphical elements used to encode data:

More accurate	Position on a plane
	Line length
	Angle & slope
↑	Area
	Volume
Less accurate	Color

Source:
Cleveland & McGill, JRSS, 1987

Graphical elements are *not* all equal in clarity

People are much better at judging line length than angle or grayscale

Graphical elements used to encode data:

More accurate	Position on a plane
	Line length
	Angle & slope
↑	Area
	Volume
Less accurate	Color

To show and correlate multivariate data, we'd like to use multiple or multifunctional elememts

Color and size and shape, for example

Source:
Cleveland & McGill, JRSS, 1987

Graphical elements used to encode data:

More accurate	Position on a plane
	Line length
	Angle & slope
↑	Area
	Volume
Less accurate	Color

Source:
Cleveland & McGill, JRSS, 1987

Will they all be processed equally...

accurately?
quickly?
with similar intensity?
separately?
at highest available level of measurement?

Unfortunately, NO.

The Cognitive Science of Visual Displays of Information

Graphical elements used to encode data:

More accurate	Position on a plane
	Line length
	Angle & slope
↑	
	Area
	Volume
Less accurate	Color

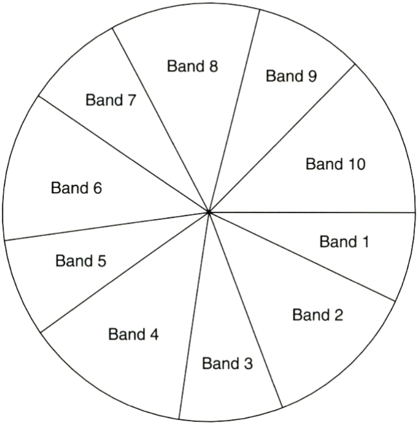
Source:
Cleveland & McGill, JRSS, 1987

Simple advice:

Reserve elements at the top of the list for important variables

Try to avoid using the elements at the bottom of the list to encode *quantitative* data (but redundant usage is fine)

Exception: color *can* effectively encode qualitative data

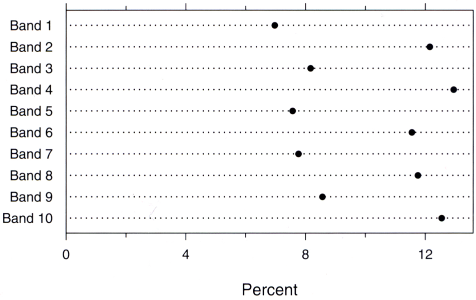


Cognitive failure:
Angular data encoding

Can you describe these data?

The exact sizes of the pies?

Source:
Cleveland, *The Elements of Graphing Data*



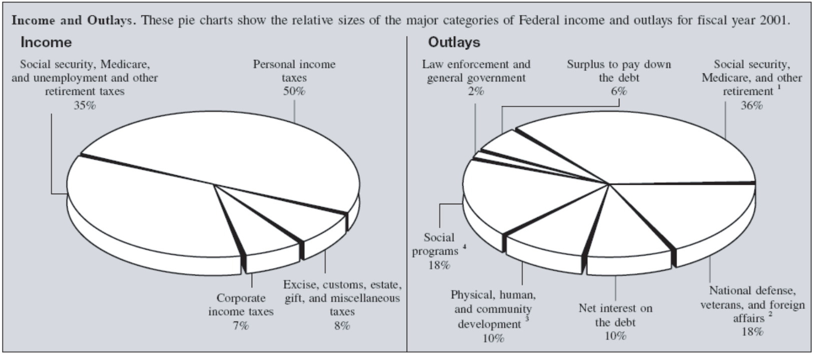
Cognitive solution:
Location data encoding

Did you notice that half of the slices are exactly 50% larger than the others?

Did you guess the exact sizes correctly?

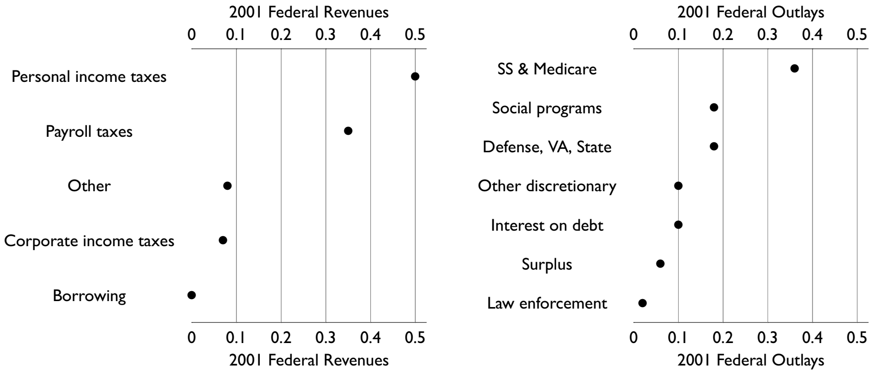
Source:
Cleveland, *The Elements of Graphing Data*

Major Categories of Federal Income and Outlays for Fiscal Year 2001



My favorite pie chart. Budget data printed on the back of the US tax forms.
Would a dot chart be as good? Better?

The Cognitive Science of Visual Displays of Information



I think this is at least as good as the pie.

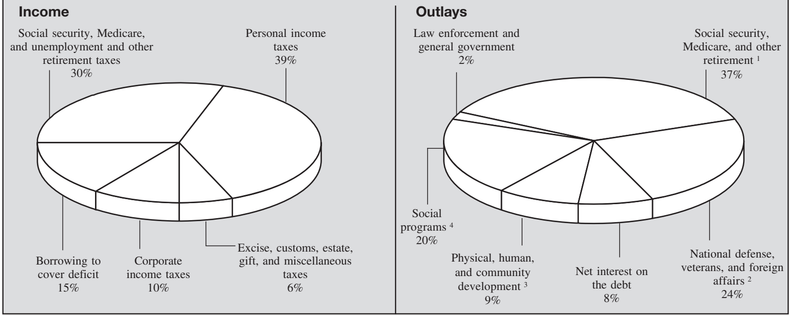
Unlike the pie, could be expanded to, say, 10 categories without much fuss.

Note that I have **diagonalized** the dot plot by sorting.

This is always helpful for reading and comparing data correctly.

Major Categories of Federal Income and Outlays for Fiscal Year 2008

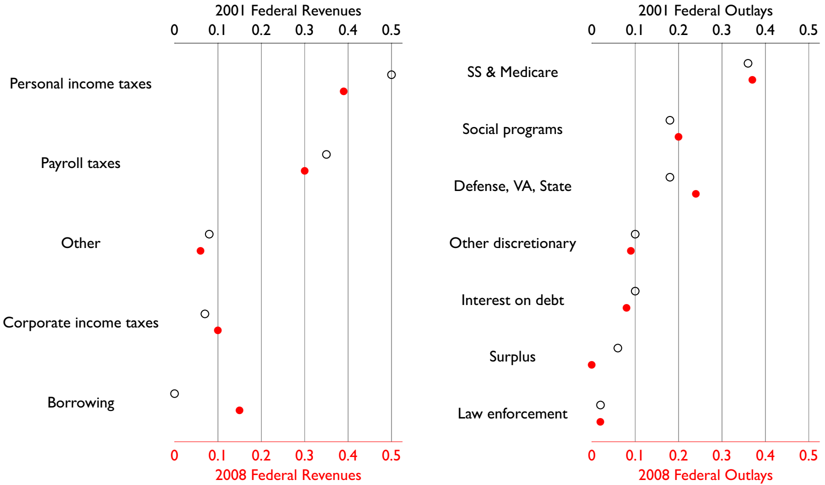
Income and Outlays. These pie charts show the relative sizes of the major categories of federal income and outlays for fiscal year 2008.



Budget pies went missing for most of the decade

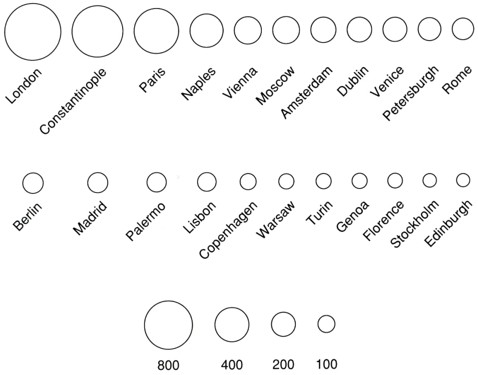
They returned in 2009 – right after a presidential election

The Cognitive Science of Visual Displays of Information



Another advantage of dotplots: easy comparison across plots through integration

This is also why dotplots are more useful than barplots



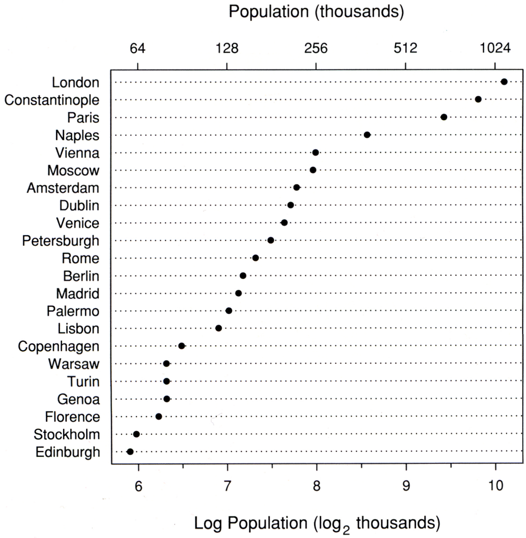
Cognitive failure:
Area data encoding

Is there a smooth increase in city size across these data?

How much bigger was Lisbon than Copenhagen?

Can we even be sure the areas represent population? What if the diameter is what matters?

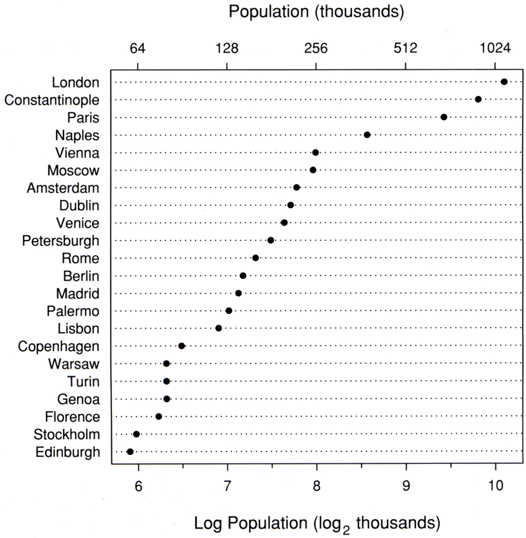
Source:
Cleveland, *The Elements of Graphing Data*



Cognitive solution:
Location data encoding

Lisbon was about fifty percent bigger than Copenhagen

The progression is far from smooth, even on a log scale



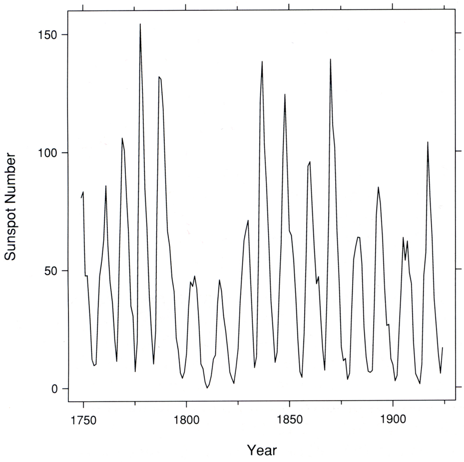
Note on log scaling

Social scientists frequently plot logged data

When you plot a logged variable, you can label the axes using the original, unlogged numbers

This greatly enhances the readability of your plot

And it requires nothing more than using different number labels

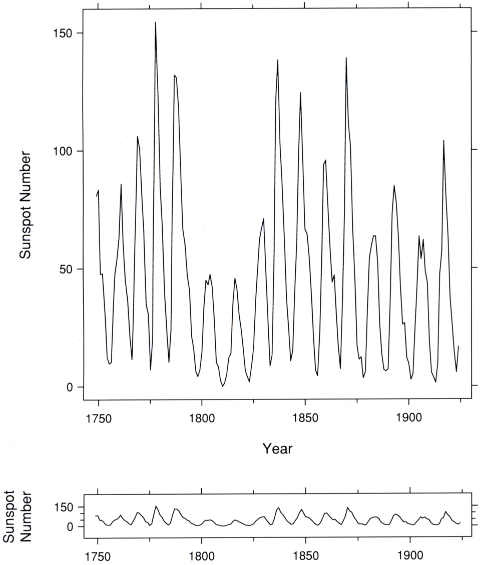


Cognitive failure:
Hard-to-read lineplots

Poor angular perception can make some lineplots hard to read

Are these spikes symmetrical?

Source:
Cleveland, *The Elements of Graphing Data*



Cognitive solution:
better aspect ratios

Choosing a better aspect ratio can make lineplots easier to read

The goal is to make slopes close to 1.0 or −1.0

Cleveland calls this "banking to 45 degrees," because angles near 45 are easier for humans to distinguish

The Cognitive Science of Visual Displays of Information

Sometimes differences take conscious effort to distinguish. Find the 3′s:

856897269846897626897643589226598659865548976892698 98

0246299687402655762798678904567923276928546098677 2098

9083457980279075904709827908579084772908759082790 8754

98709856749068975786259845690243790472190790709811450

85689726984689762689764458922659865986554897689269898

Sometimes differences take conscious effort to distinguish. Find the 3's:

8568972698468976268976435892265986598655489768926989 8
0246299687402655762798678904567923276928546098677209 8
9083457980279075904709827908579084772908759082790875 4
9870985674906897578625984569024379047219079070981145 0
8568972698468976268976445892265986598655489768926989 8

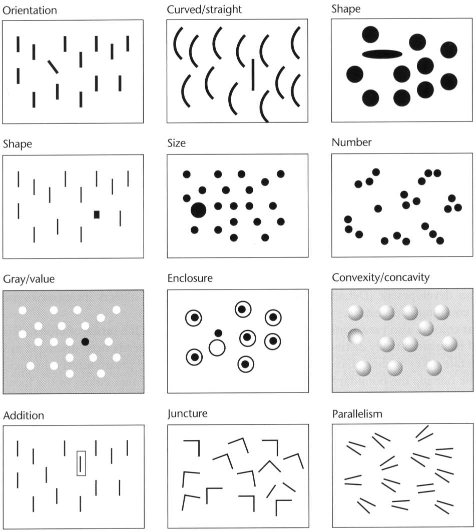
Sometimes encoded data pop right out. Find the 3's:

8568972698468976268976435892265986598655489768926989 8
0246299687402655762798678904567923276928546098677209 8
9083457980279075904709827908579084772908759082790875 4
9870985674906897578625984569024379047219079070981145 0
8568972698468976268976445892265986598655489768926989 8

Same information in both examples.

But our brains process color differences "pre-attentively" – fast & effortlessly

Source: Ware, *Information Visualization*

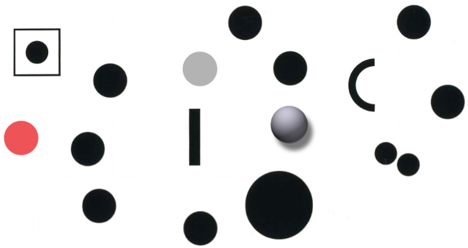


Where possible, **pre-attentive differences** should be exploited

The essence of graphics that "hit you between the eyes"

Tables of numbers seldom if ever achieve this

Source:
Ware, *Information Visualization*

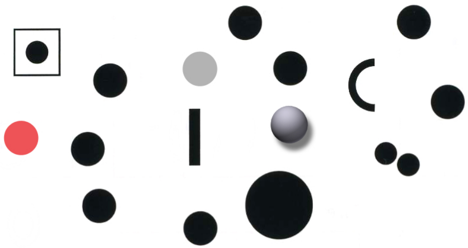


But there's only so much pre-attention to go around

As you add pre-attentive differences, the effect of each diminishes

(though not necessarily equally: some are stronger than others)

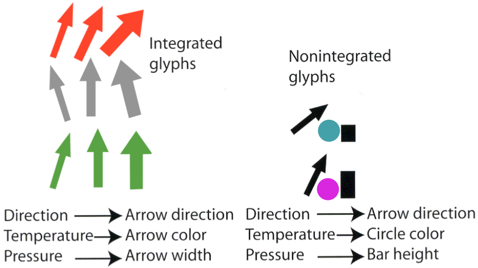
Source:
Ware, *Information Visualization*



The symbols plotted at the left are *glyphs*

Each might represent a single case in a dataset

Source:
Ware, *Information Visualization*



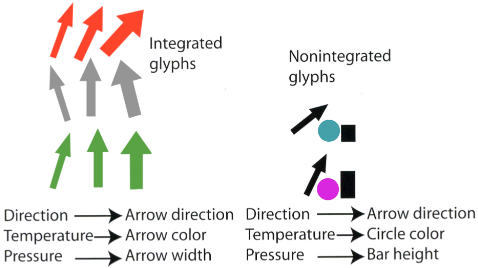
Source:
Ware, *Information Visualization*

The symbols plotted at the left are *glyphs*

Each might represent a single case in a dataset

But each glyph can carry multiple dimensions

The is best achieved using an *integrated* set of glyph characteristics, one per dimension



Number Dimensions of a glyph

$\geq$

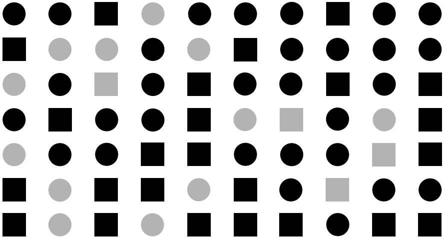
Number of variables encoded

The more variables you encode to dimensions of glyphs, the harder it is to pre-attentively separate the dimensions

Source:

Ware, *Information Visualization*

Quick! Pick out the **gray squares**



Source: Ware, *Information Visualization*

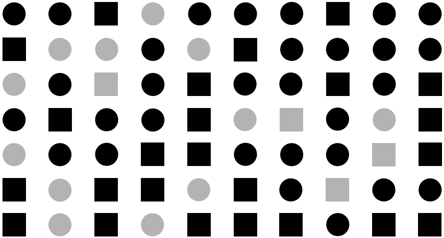
Number Dimensions of a glyph

$\geq$

Number of variables encoded

The more variables you encode to dimensions of glyphs, the harder it is to pre-attentively separate the dimensions

Quick! Pick out the **gray squares**



Source: Ware, *Information Visualization*

Number Dimensions of a glyph

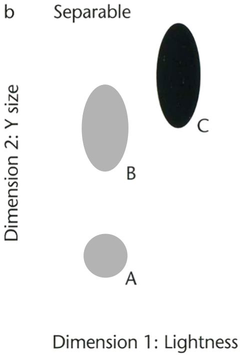
$\geq$

Number of variables encoded

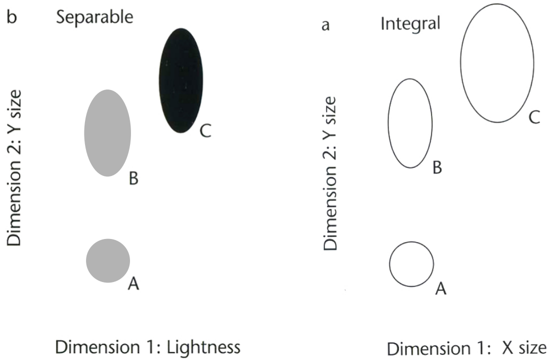
The more variables you encode to dimensions of glyphs, the harder it is to pre-attentively separate the dimensions

This may be an acceptable price for structured comparison across many dimensions

Sometimes, the best graphic – which is the simplest one that makes the desired point – still takes a bit of study to fully comprehend



Take care to choose glyph dimensions that can be cleanly separated



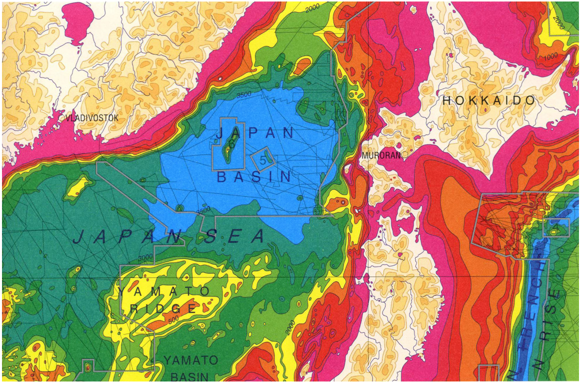
Take care to choose glyph dimensions that can be cleanly separated

Sometimes, dimensions are reinforcing – and tend to blur together.
This makes it harder to extract information from the plot

While striking and chromatic, this map is not clear or useful

The color scale is so ineffectively chosen that we likely wouldn't know this was a map if the place names were missing

Source:
Tufte, *Information Visualization*



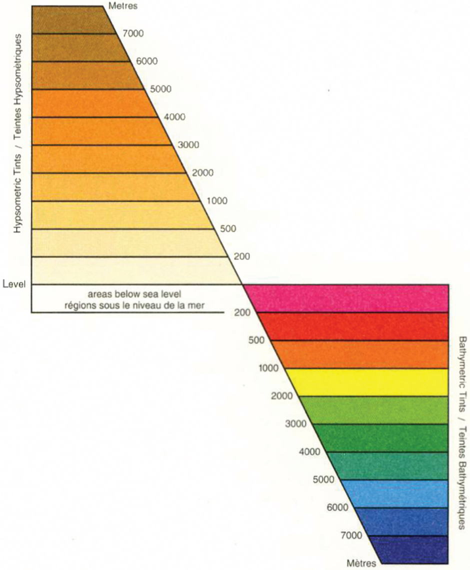
Using Color (In)effectively

The mapmaker used a rainbow scale for underwater depth

Normally, mapmakers maintain a constant hue (or "chroma") for a terrain type, and vary its brightness and saturation

Our eyes can order brightness, but not the rainbow

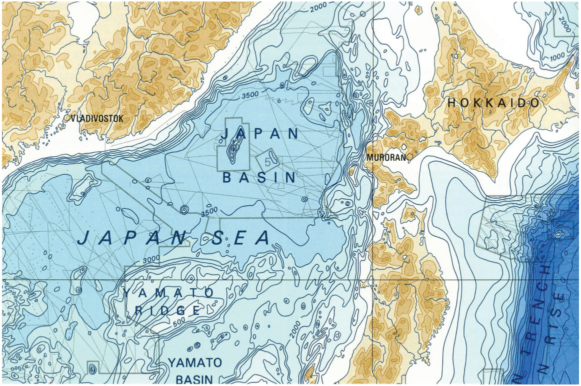
Source:
Tufte, *Information Visualization*



For centuries, cartographers have effectively used color on maps

While not as flashy as a rainbow scale, this map is far more effective for both lookup and comparison

Source:
Tufte, *Information Visualization*

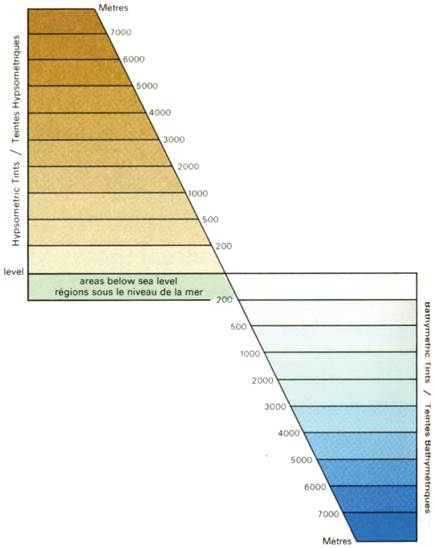


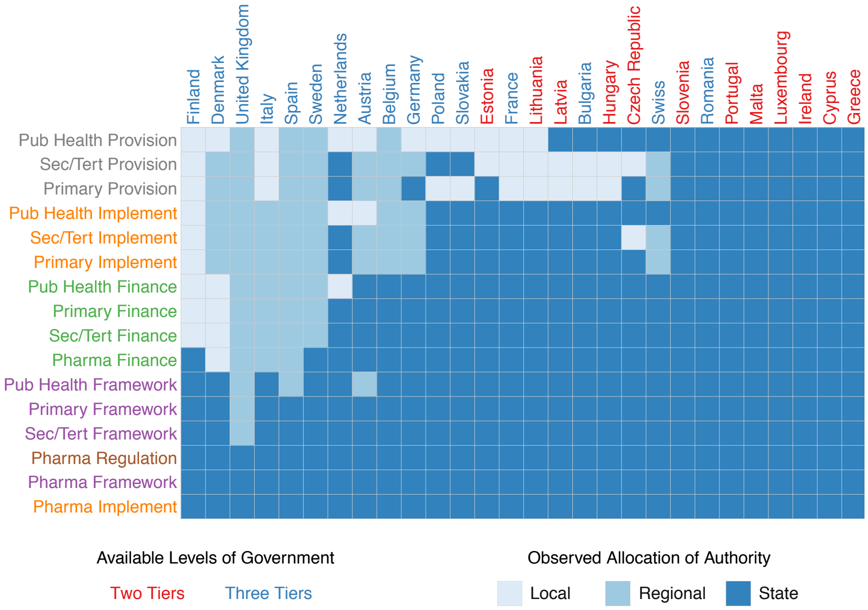
Using Color Effectively

Humans are bad at precise reading of color

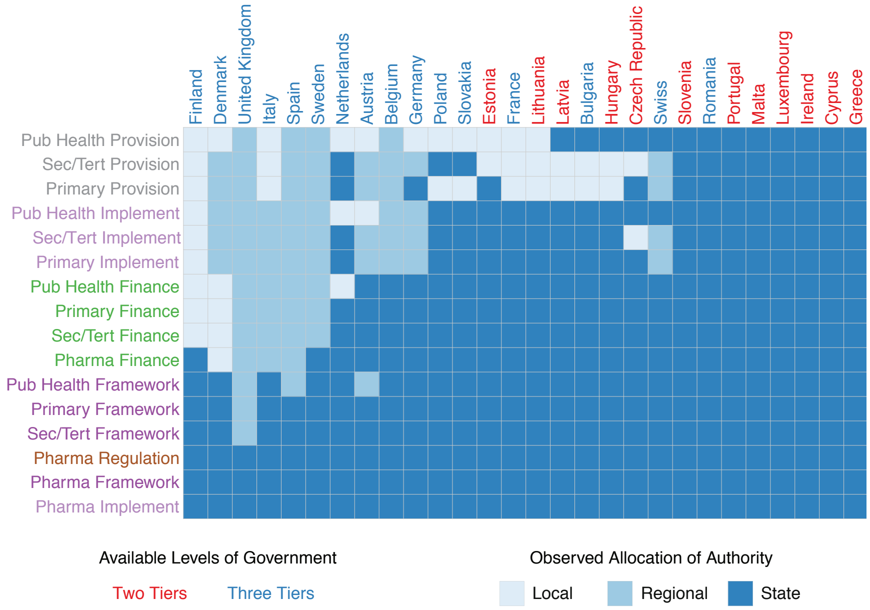
When plotting a quantitative variable on a color scale, care should be taken to find pre-attentively smooth gradients

Source:
Tufte, *Information Visualization*

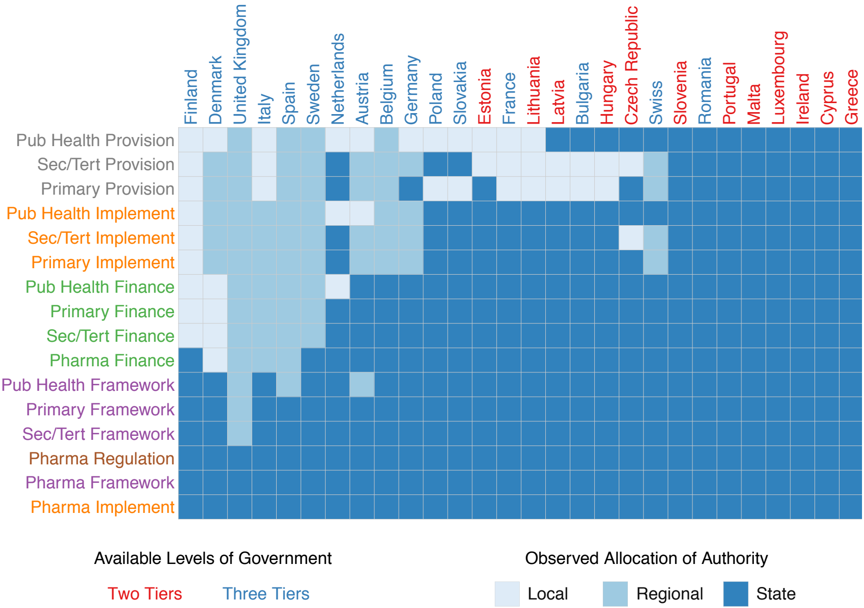




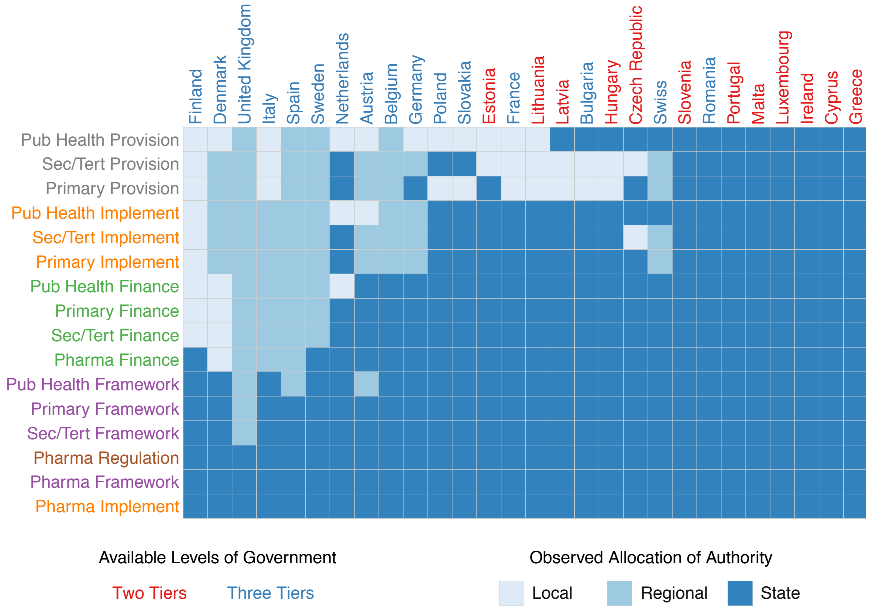
Recall how I used colors to indicate row categories in this heatmap



What if I'd drawn my heatmap with *these* colors instead?



Poor choice of colors for categorical data can distort the data



Choose colors for categories to achieve equal pairwise distinctions

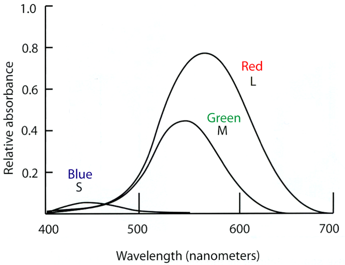
Color Advice

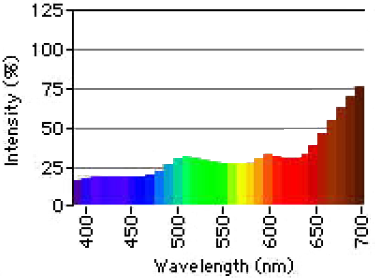
1. Choose colors for quantities using pre-attentively smooth gradients
2. Choose colors for categories to achieve equal pairwise distinctions
3. Avoid overlapping colors with similar brightness (value)
4. Use pastels for large area colors and saturated colors for small points

To understand and implement the above,
we need to know something about the science of color

Color is

- a measure of the wavelength of light
- something we perceive
- a mixture of red, blue, and green
- a mixture of hue, luminosity, and saturation
- broadly categorical; also continuous
- hard to perceive accurately
- hard to reproduce accurately
- an element of many visuals





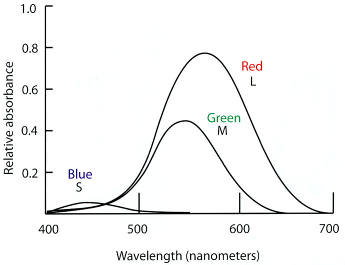
Color Science Vastly Oversimplified by an Outsider

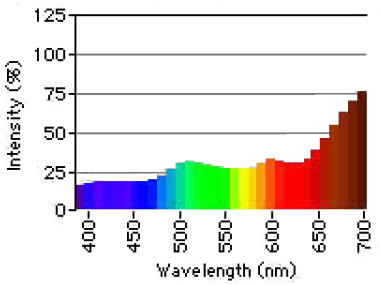
Human eyes contain two kinds of photo-receptive elements:

Rods. Sensitive to brightness
Single photon receptors
Little use in sunlight

Cones. Come in three varieties...

1. Short wavelength (red);
   most sensitive
2. Medium wavelength (green);
   moderately sensitive
3. Long wavelength (blue); weak





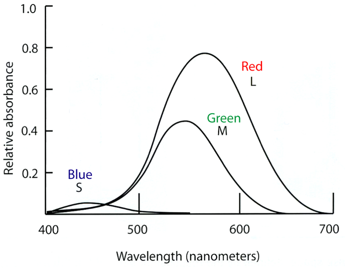
Color Science Vastly Oversimplified by an Outsider

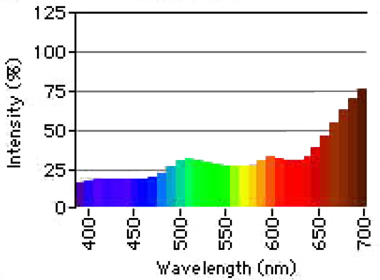
Humans are best at seeing red;
worst at seeing blue

Species vary in color vision ability:

- dogs have only two cones, are red-green colorblind, and see less detail in daylight
- birds have more cones than humans – chickens have 12!

Number of cones = number of primary colors a species perceives.
Mixing the three (human) primaries in different amounts makes any color humans can see.

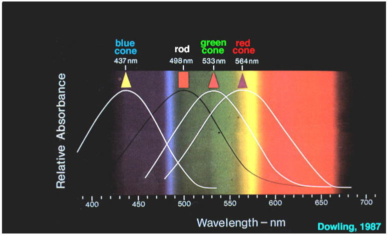


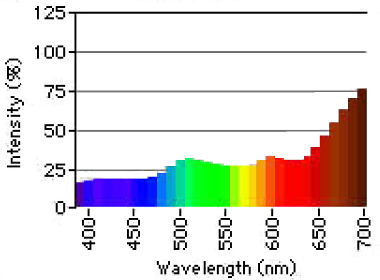


Suppose we used scientific equipment to measure the number of photons received at each of *N* wavelengths throughout the visible spectrum

Clearly, we could choose *N* to be arbitrarily large, and still learn more about the distribution of photon wavelenghts by sampling in still more places

Just like adding bins to a histogram, this gives us a finer view of the distribution

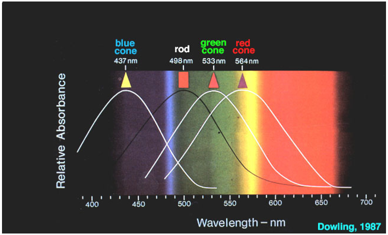


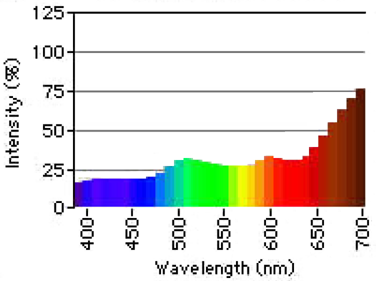


Color Science Vastly Oversimplified by an Outsider

The human eye has cones tuned to just three wavelengths, so our eyes approximate complex histogram as mixture of three densities

Many different colors (that is, distinct histograms on the visual spectrum) that look the "same" to us would look different to a chicken, which samples the distribution in 12 places!





Color Spaces

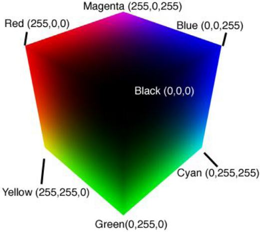


Figure 3a: The black corner (0,0,0) of the RGB Cube

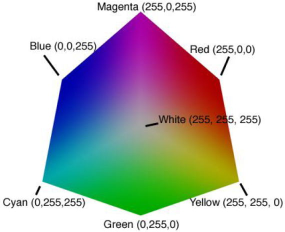


Figure 3b: The white corner (255,255,255) of the RGB Cube

Primaries and color can be expressed in many equivalent ways.

These are different **colorspaces**: mappings from 3 variables to a color

Computer space · **RGB**
Red, Green, Blue

Printer space · **CMYK**
Cyan, Magenta, Yellow, Black

Artist space · **HSV**
Hue, Saturation, Value

Brain space · **CIElab**
Lightness, blue/yellow, red/green

Color Spaces: RGB

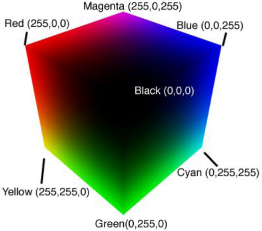


Figure 3a: The black corner (0,0,0) of the RGB Cube

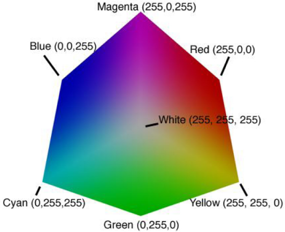


Figure 3b: The white corner (255,255,255) of the RGB Cube

RGB is mainly useful for telling a computer what color of light to display

If you want to tell a printer what color of ink to print, it is easier to use CMYK: cyan, magenta, yellow, and black

But neither color space is useful for choosing colors, either aesthetically or scientifically

Color Spaces: HSV

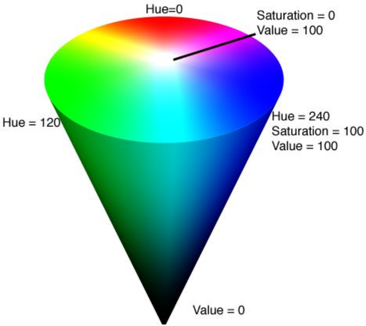


Figure 2: The HSV Cone

Source: Darrin Cardani, “Adventures in HSV Space”

Artists find RGB an inconvenient space to think about color

Instead, they often use HSV:
hue is the “name” of the color

think rainbow

saturation is the richness of the color; desaturated colors have been mixed with gray

think solids vs pastels

value is the brightness of the color; how much white or black is mixed in

think stop sign vs. red wine

Color Spaces: HSV

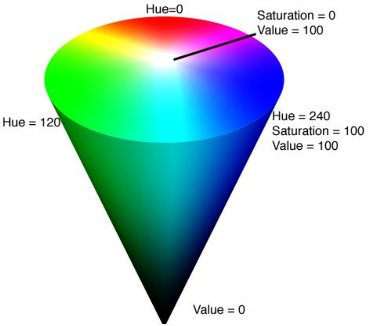


Figure 2: The HSV Cone

HSV is useful for constructing beautiful complementary colors for artistic palettes

Colors on opposite sides of the cone are aesthetically harmonious contrasting colors

Colors 120 degrees apart (color triads) are too

Any HSV color can also be represented in RGB and CMYK, and vice versa

Source: Darrin Cardani, "Adventures in HSV Space"

Color Spaces: Opponent Color Theory

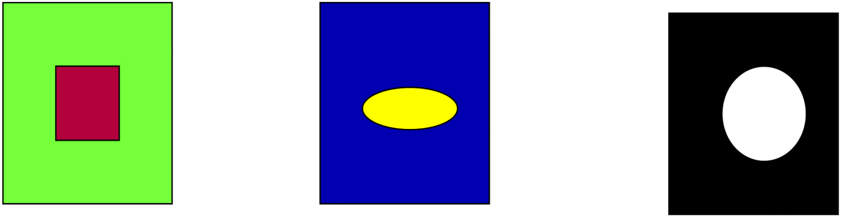
Does your brain read off RGB values from your cones? Or maybe HSV values?

Probably not.

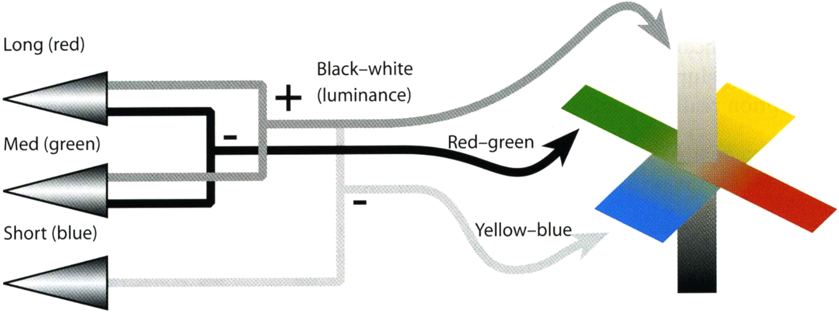
Opponent color theory:

Human optical system converts {S,M,L} cone readouts to three channels

- Redness vs. greenness
- Blueness vs. yellowness
- Brightness



Color Spaces: Opponent Color Theory



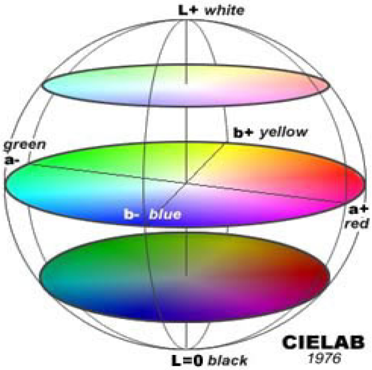
Source: Colin Ware, *Information Visualization*

In other words, red/green and blue/yellow are "opponent colors"

They appear in zero-sum combinations
(no one ever says "the yellowish-blue sweater")

Color blindness: absence or weakness of one set of cones

Color Spaces: CIElab



Opponent color theory suggests a new color space, CIElab

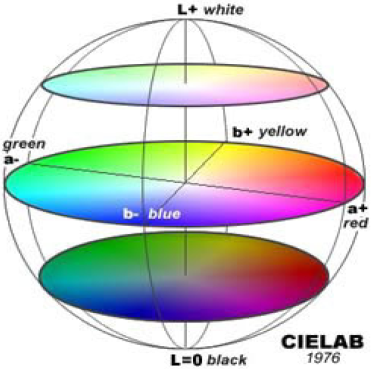
(CIE stands for *Commission internationale de l'éclairage*,

or International Commission on Illumination)

- **l** = luminance (white vs. black)
- **a** = red vs. green
- **b** = blue vs. yellow

Equal Euclidian distances in CIElab space are (approximately) perceptually "equal" to humans

Color Spaces: CIElab



Equal Euclidian distances in CIElab space are (approximately) perceptually "equal" to humans

If CIElab is the brain's color space, it's the best one for choosing colors to convey precise scientific information

If you want to convey distinct categories, choose colors that are well separated in CIElab space

If you want to convey precise numerical steps, choose equal stpes through CIElab space

Does this mean you need to learn a lot of cognitive science before you can make a color graphic?

Not really.

Easy shortcuts available: `RColorBrewer` will choose appropriate colors for you

At left are perceptually equal gradients for different color hues, as suggested by `RColorBrewer`

```
library(RColorBrewer)
display.brewer.all(type="seq")
```

Pick one horizontal strip for your color scale to plot quantitative data

`RColorBrewer` **will also suggest colors for qualitative variables**

Goal here is to make each category equally distinct from the others

```
library(RColorBrewer)
display.brewer.all(type="qual")
```

Why so many choices? Not for aesthetics, but because they solve different color cognition problems

Suppose we use `Pastel1` to encode categories to glyphs

Can you easily tell which color is which?

Suppose we use `Pastel1` to encode categories to glyphs

Can you easily tell which color is which?

Hard to distinguish the hue of small areas of desaturated color

Don't use pastels to color small glyphs

Suppose we use `Pastel1` to encode categories to regions

Can you easily tell which color is which?

Color Cognition Problem 1 *Saturation and Size*

Suppose we use `Pastel1` to encode categories to regions

Can you easily tell which color is which?

Easy to distinguish the hue of large areas of desaturated color

Use pastels to color large regions

Suppose we use `Set1` to encode categories to glyphs

Can you easily tell which color is which?

Suppose we use `Set1` to encode categories to glyphs

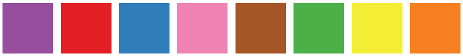
Can you easily tell which color is which?

Easy to distinguish the hue of small areas of saturated color

Use jewel tones to color large regions

Suppose we use `Set1` to encode categories to regions

Would a graph with large bright regions be readable?



Color Cognition Problem 1 *Saturation and Size*

Suppose we use `Set1` to encode categories to regions

Would a graph with large bright regions be readable?

Large areas of saturated color command attention – distract from small details

Avoid jewel tones when coloring large regions

Use jewel tones for glyphs *Use pastels for regions*

Color Cognition Problem 1 — *Saturation and Size*

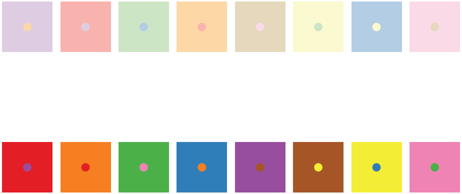
Use jewel tones for glyphs *Use pastels for regions*

Avoid pastel glyphs and saturated regions!

Color Cognition Problem 1 — *Saturation and Size*

Use jewel tones for glyphs *Use pastels for regions*

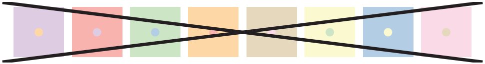
Avoid pastel glyphs and saturated regions!



Color Cognition Problem 1 *Saturation and Size*

Use jewel tones for glyphs *Use pastels for regions*

Avoid pastel glyphs and saturated regions!



Color Cognition Problem 2
Background Contrast

Text is only readable when it differs significantly from the background in *value*

Dark text only works on light backgrounds

Legible text requires value contrast

Legible text requires value contrast

Legible text requires value contrast

Color Cognition Problem 2
Background Contrast

Text is only readable when it differs significantly from the background in *value*

Dark text only works on light backgrounds

Light text only works on dark backgrounds

Legible text requires value contrast

Legible text requires value contrast

Legible text requires value contrast

Color Cognition Problem 2
Background Contrast

Text is only readable when it differs significantly from the background in *value*

Dark text only works on light backgrounds

Light text only works on dark backgrounds

Mid-value backgrounds make muddy images: *avoid*

Legible text requires value contrast

Avoid mid gray backgrounds

Legible text requires value contrast

Color Cognition Problem 2
Background Contrast

Text is only readable when it differs significantly from the background in *value*

Mid-value backgrounds make muddy images: *avoid*

Applies to graphs generally: don't use a gray background to your plot

Warning: gray backgrounds are often the default in `ggplot2` and `Excel`!

Legible text requires value contrast

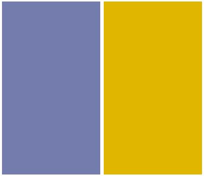
Avoid mid gray backgrounds

Legible text requires value contrast

Common mistaken intuition:

Different hues ("colors") are sufficient to distinguish background and foreground

It is very difficult to read text that is isoluminant with its background color. If clear text material is to be presented it is essential that there be sustantial luminance contrast with the background color. Color contrast is not enough. This particular example is especially difficult because the chromatic difference is in the yellow blue direction. The only exception to the requirement for luminance contrast is when the purpose is artistic effect and not clarity.

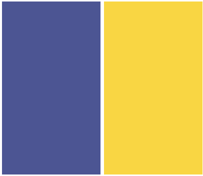


Common mistaken intuition:

Different hues ("colors") are sufficient to distinguish background and foreground

Even two color opposites (blue and yellow) can blend when they have similar values (brightness)

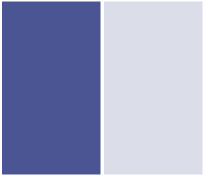
It is very difficult to read text that is isoluminant with its background color. If clear text material is to be presented it is essential that there be sustantial luminance contrast with the background color. Color contrast is not enough. This particular example is especially difficult because the chromatic difference is in the yellow blue direction. The only exception to the requirement for luminance contrast is when the purpose is artistic effect and not clarity.



To avoid unreadable text,
make sure background and foreground have different values

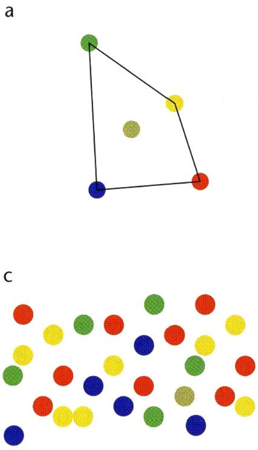
Color Cognition Problem 2 *Background Contrast*

It is very difficult to read text that is isoluminant with its background color. If clear text material is to be presented it is essential that there be sustantial luminance contrast with the background color. Color contrast is not enough. This particular example is especially difficult because the chromatic difference is in the yellow blue direction. The only exception to the requirement for luminance contrast is when the purpose is artistic effect and not clarity.



To avoid unreadable text,
make sure background and foreground have different values

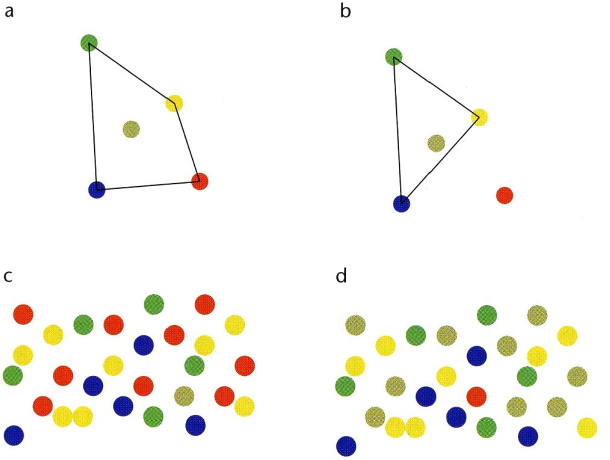
With a large value contrast,
even background and foreground of the same hue can be effective



`RColorBrewer` chooses "equally distinct" colors. How?

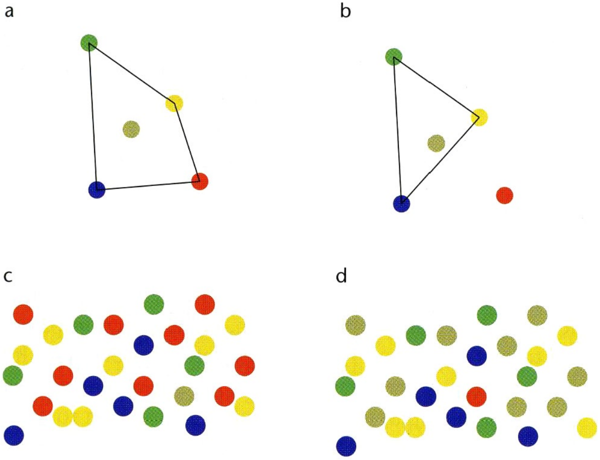
Colors in a are plotted in CIElab space: Interior colors blend in

Choosing Colors as Highlights



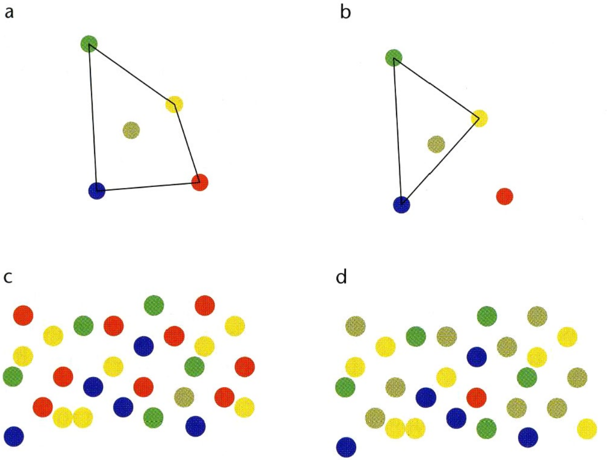
`RColorBrewer` **chooses "equally distinct" colors. How?**

Colors outside the convex *hull* of the other colors *stand out*



`RColorBrewer` chooses "equally distinct" colors. How?

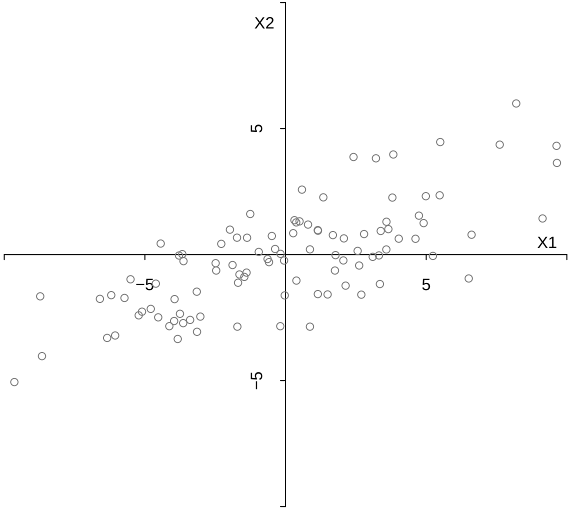
`RColorBrewer` "`qual`" colors are equidistant from each other on the convex hull



If your goal is to *highlight* a point or category,
choose something outside the convex hull of the other colors

Aside: Convex Hulls

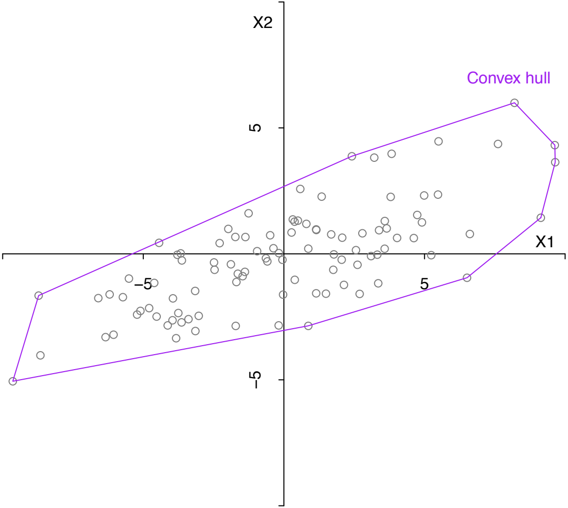
What is a convex hull?



Aside: Convex Hulls

What is a convex hull?

An elastic band wrapped around the cloud of points such that it contains the smallest convex set containing those points

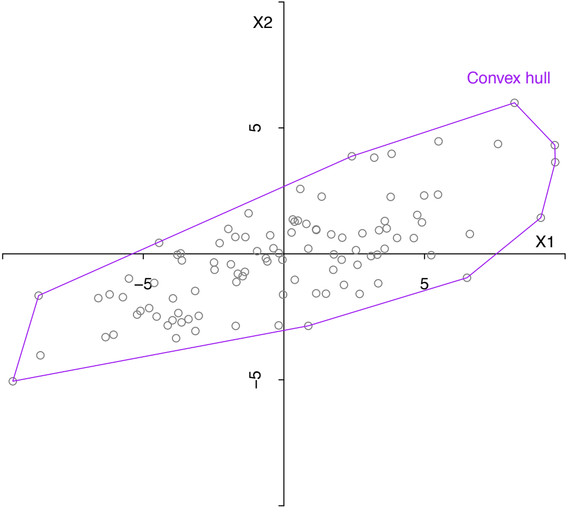


Aside: Convex Hulls

What is a convex hull?

An elastic band wrapped around the cloud of points such that it contains the smallest convex set containing those points

What is a convex set?



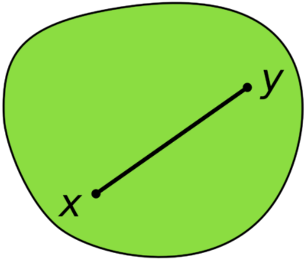
What is a convex hull?

An elastic band wrapped around the cloud of points such that it contains the smallest convex set containing those points

What is a convex set?

If a straight line between any two points in a region remains within that region, that region is a convex set

A Convex Set



Aside: Convex Hulls

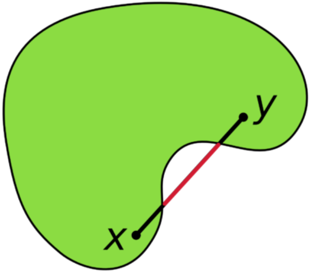
What is a convex hull?

An elastic band wrapped around the cloud of points such that it contains the smallest convex set containing those points

What is a convex set?

If a straight line between any two points in a region remains within that region, that region is a convex set

Not a Convex Set

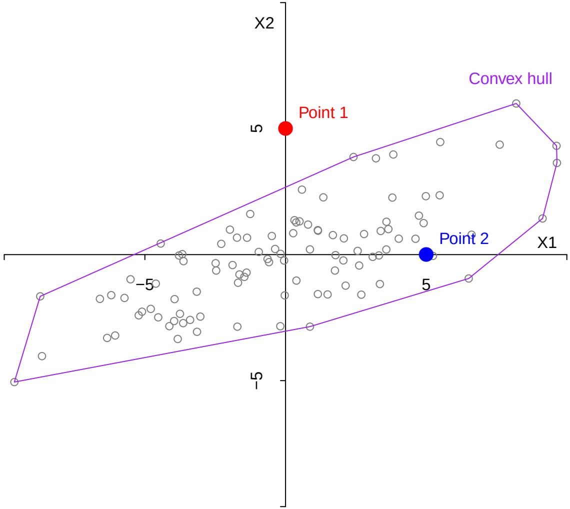


Aside: Convex Hulls

Convex hulls will come up again later when we discuss the difference between extrapolation from a dataset and interpolation from a dataset

Point 2 is interpolated

Point 1 is extrapolated



Summing Up Session 1

So far, we've learned:

Some principles for effective visual display

How to avoid cognitive pitfalls in designing visuals

How to select colors for categorical and quantitative information

Most resources on scientific visualization apply these ideas to *exploratory data analysis* – looking at the data with pictures

But social science is heavy on modeling of data

Next step: Designing effective visuals for understanding *models*

Plan of Session 2: Concepts for Visualizing Model Inference

1. From data exploration to model exploration
2. Graphical approaches to model inference for "simple" models: examples
3. Obtaining Quantities of Interest from models
4. Graphical approaches to model inference for "complex" models

Examples for Session 2

Why did the space shuttle *Challenger* explode?

Source: Tufte

Method: Logistic regression with a single covariate

Who votes in American elections?

Source: King, Tomz, and Wittenberg

Method: Logistic regression with several covariates

How do US central bankers make monetary policy choices?

Source: Adolph, *BBC*, Ch. 4

Method: Ordered probit regression with interactions & compositional variables

What determines cross-national inflation performance?

Source: Adolph, *BBC*, Ch. 3

Method: Time series cross-section regression with compositional variables

Presenting Estimated Models in Social Science

Most empirical work in social science is regression model-driven, with a focus on conditional expectation

Our regression models are

- full of covariates
- often non-linear
- usually involve interactions and transformations

If there is anything we need to visualize well, it is our models

Yet we often just print off tables of parameter estimates

Limits readers' *and analysts'* understanding of the results

Presenting Estimated (Causal) Models in Social Science

What if you work in a causal inference framework?

Still a great need for visualization:

- to show robustness across different techniques
- to show differences across quantities of interest (e.g., ATE vs. ATT)
- to show variation across different kinds of subjects (variation in LATEs/LATTs)

And even if your work with observational data and regression, visualization becomes essential when you want to compare model predictions across many specific cases (examples in Session 6)

What Most Discussions of Statistical Graphics Leave Out

Tufte's books have had a huge impact on information visualization

However, they have two important limits:

Modeling Most examples are either exploratory or very simple models;

Social scientists want cutting edge applications

Tools Need to translate aesthetic guidelines into software

Social scientists are unlikely to do this on their own –
and shouldn't have to!

Goals for Visualizing Model Inference

Key need Ready-to-use techniques to visually present model results:

- for many variables

Goals for Visualizing Model Inference

Key need Ready-to-use techniques to visually present model results:

- for many variables
- for many robustness checks

Goals for Visualizing Model Inference

Key need Ready-to-use techniques to visually present model results:

- for many variables
- for many robustness checks
- showing uncertainty

Goals for Visualizing Model Inference

Key need Ready-to-use techniques to visually present model results:

- for many variables
- for many robustness checks
- showing uncertainty
- without accidental extrapolation

Goals for Visualizing Model Inference

Key need Ready-to-use techniques to visually present model results:

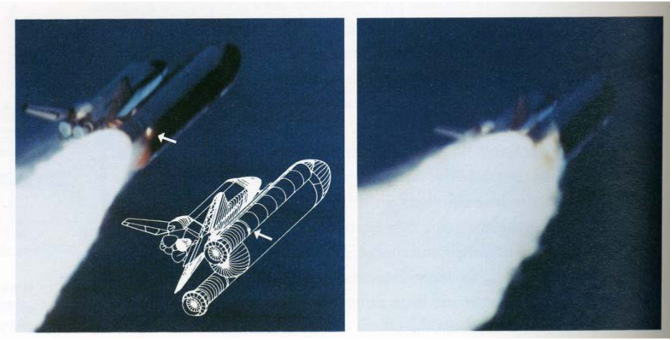
- for many variables
- for many robustness checks
- showing uncertainty
- without accidental extrapolation
- for an audience without deep statistical knowledge

Not covered here Visual displays for data (not model) exploration

For this and other topics, and a reading list, see my course at `faculty.washington.edu/cadolph/vis`

Lots of examples... Too many if we need to discuss methods in detail

The *Challenger* launch decision



In 1986, the Challenger space shuttle exploded moments after liftoff

The decision to launch is one of the most scrutinized in history

Failure of O-rings in the solid-fuel rocket boosters blamed for explosion

Could this failure have been forseen?

The *Challenger* launch decision

Flights with O-ring damage	
Flt Number	Temp (F)
2	70
41b	57
41c	63
41d	70
51c	53
61a	79
61c	58

Data on O-ring failures
at different launch temperatures,
provided to NASA by Morton-Thiokol
hours before launch

Engineers who made this table worried about launching below 53 degrees (Why?)

The *Challenger* launch decision

Flights with O-ring damage	
Flt Number	Temp (F)
2	70
41b	57
41c	63
41d	70
51c	53
61a	79
61c	58

Data on O-ring failures
at different launch temperatures,
provided to NASA by Morton-Thiokol
hours before launch

Engineers who made this table worried about launching below 53 degrees (Why?)

Physical problem: O-ring would erode or "blow-by" in cold temperatures

The *Challenger* launch decision

Flights with O-ring damage	
Flt Number	Temp (F)
2	70
41b	57
41c	63
41d	70
51c	53
61a	79
61c	58

Data on O-ring failures
at different launch temperatures,
provided to NASA by Morton-Thiokol
hours before launch

Engineers who made this table worried about launching below 53 degrees (Why?)

Physical problem: O-ring would erode or "blow-by" in cold temperatures

Failed to convince administrators of danger
Counter-argument:
"damages at low and high temps"

Are there problems with this presentation?
With the use of data?

The *Challenger* launch decision

Flights with O-ring damage	
Flt Number	Temp (F)
2	70
41b	57
41c	63
41d	70
51c	53
61a	79
61c	58

Data on O-ring failures
at different launch temperatures,
provided to NASA by Morton-Thiokol
hours before launch

Are there problems with this presentation?

With the use of data?

The *Challenger* launch decision

Flights with O-ring damage	
Flt Number	Temp (F)
2	70
41b	57
41c	63
41d	70
51c	53
61a	79
61c	58

Data on O-ring failures
at different launch temperatures,
provided to NASA by Morton-Thiokol
hours before launch

Are there problems with this presentation?
With the use of data?

Did not consider successes,
only failures

Selection on the dependent variable

Why sort by launch number?

The *Challenger* launch decision

O-ring damage pre-Challenger, by temperature at launch

Damage?	Temp (F)	Damage?	Temp (F)
Yes	53	**Yes**	70
Yes	57	No	70
Yes	58	No	70
Yes	63	No	72
No	66	No	73
No	67	No	75
No	67	No	76
No	67	No	76
No	68	No	78
No	69	**Yes**	79
Yes	70	No	81

The evidence begins to speak for itself.

What if Morton-Thiokol engineers had made this table before the launch?

The *Challenger* launch decision

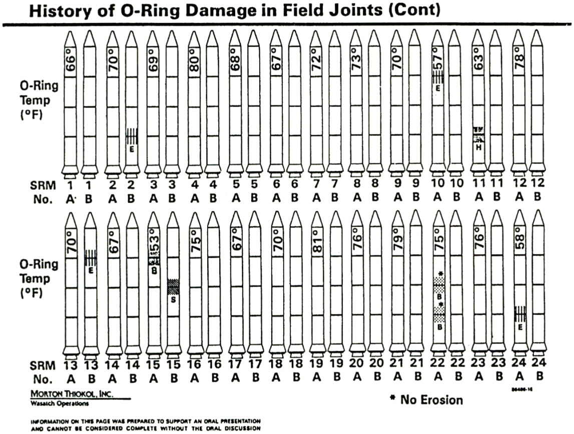
Why didn't NASA make the right decision?

Many answers in the literature:
bureaucratic politics; group think; bounded rationality, etc.

But Edward Tufte thinks it may have been a matter of presentation & modeling:

- Never made the right tables or graphics
- Selected only failure data
- Never considered a simple statistical model

The *Challenger* launch decision

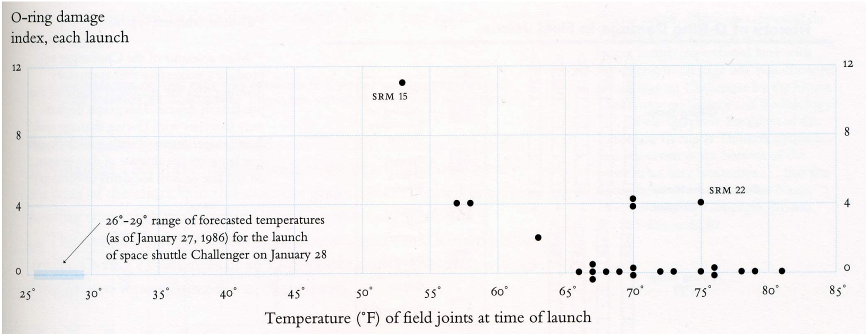


What Morton-Thiokol presented months *after* the disaster

A marvel of poor design – obscures the data, makes analysis *harder*

Can methods commonly used in social science do better?

The *Challenger* launch decision

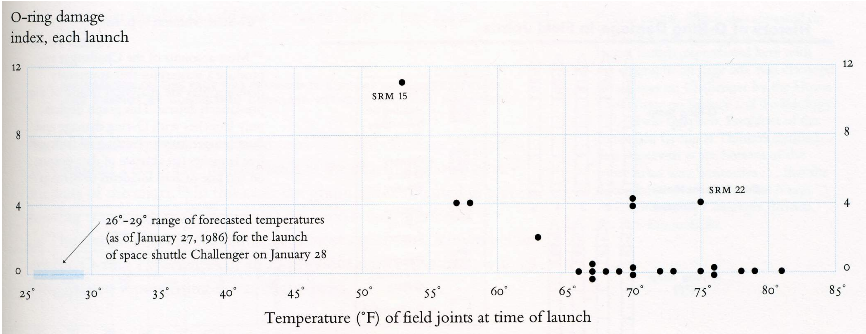


What was the forecast temperature for launch?

26 to 29 degrees Fahrenheit (−2 to −3 degrees C)!

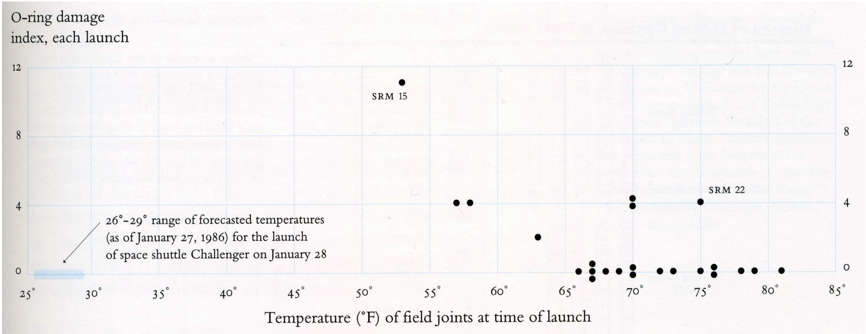
The shuttle was launched in unprecendented cold

The *Challenger* launch decision



Imagine you are the analyst making the launch recommendation.
You've made the scatterplot above. What would you add to it?
Put another way, what do you is the first question you expect to hear?

The *Challenger* launch decision



"What's the chance of failure at 26 degrees?"

The scatterplot suggests the answer is "high," but that's vague.

But what if the next launch is at 58 degrees? Or 67 degrees?

We need a *probability model* and a way to convey that model to the public

Let's try a simple logit model of damage as a function of temperature:

$$\Pr(\text{Damage}) = \text{logit}^{-1}(\hat{\beta}_0 + \hat{\beta}_1 \text{Temp})$$

R gives us this lovely logit output...

Let's try a simple logit model of damage as a function of temperature:

$$\Pr(\text{Damage}) = \text{logit}^{-1}(\hat{\beta}_0 + \hat{\beta}_1 \text{Temp})$$

R gives us this lovely logit output...

Variable	est.	s.e.	p
Temperature (F)	-0.18	0.09	0.047
Constant	11.9	6.34	0.062
N	22		
log-likelihood	-10.9		

The *Challenger* launch decision

Let's try a simple logit model of damage as a function of temperature:

$$\Pr(\text{Damage}) = \text{logit}^{-1}(\hat{\beta}_0 + \hat{\beta}_1 \text{Temp})$$

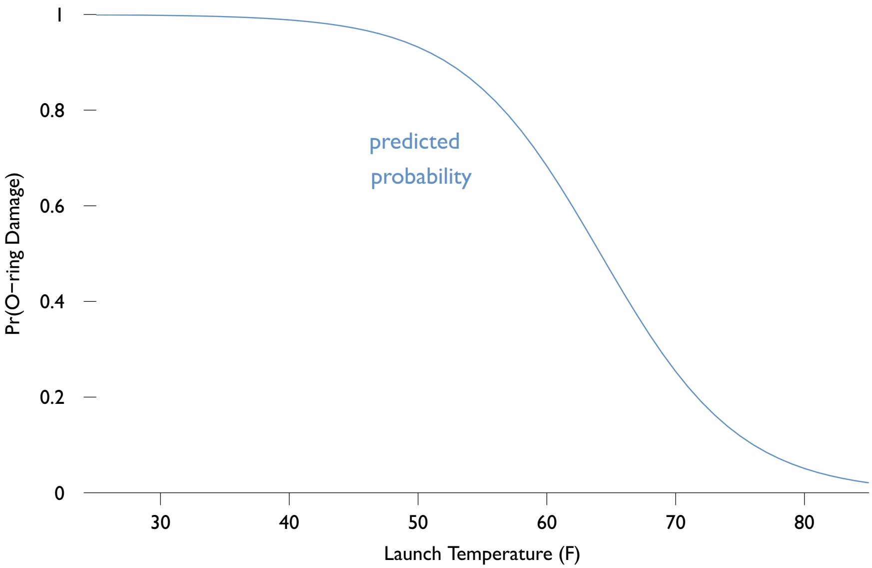
R gives us this lovely logit output...

Variable	est.	s.e.	*p*
Temperature (F)	-0.18	0.09	0.047
Constant	11.9	6.34	0.062
N	22		
log-likelihood	-10.9		

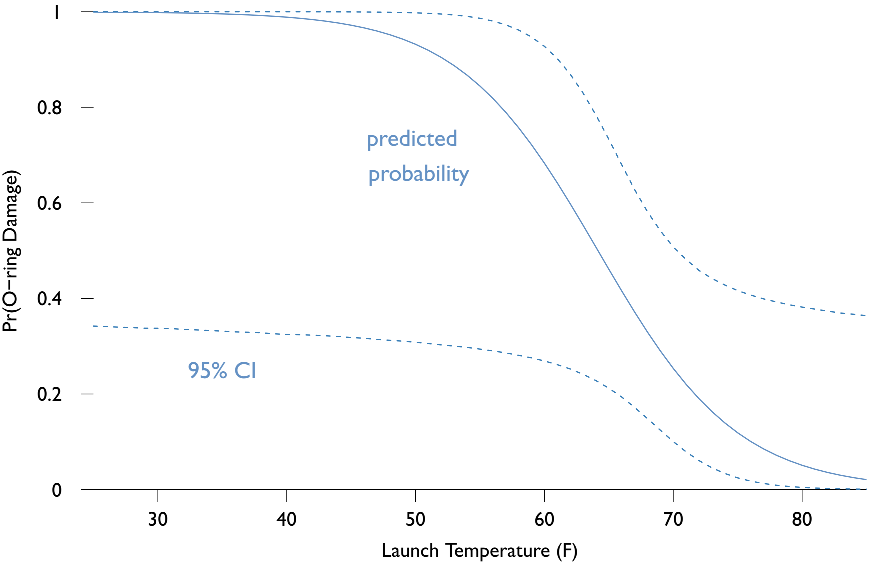
Which most social scientists read as
"a significant negative relationship b/w temperature and probability of damage"

...but that's pretty vague too

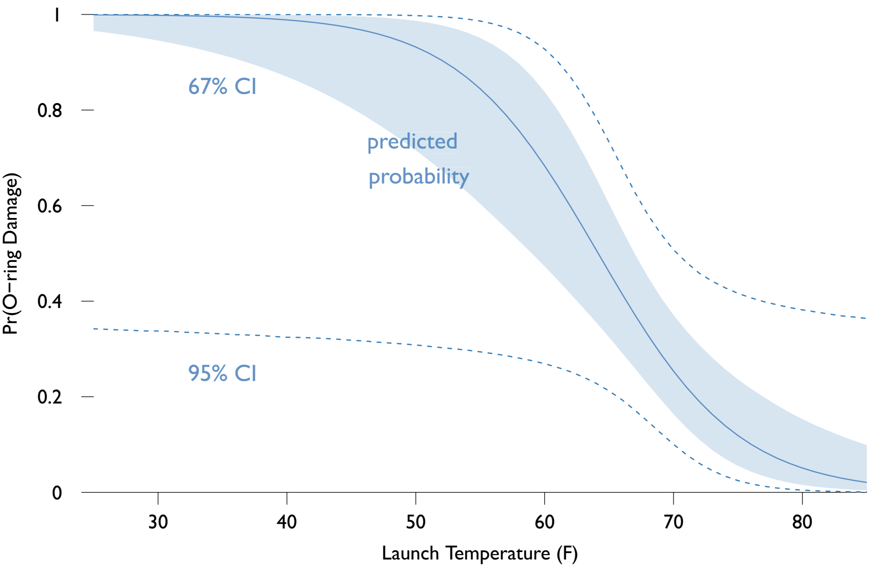
Is there a more persuasive/clear/useful way to present these results?



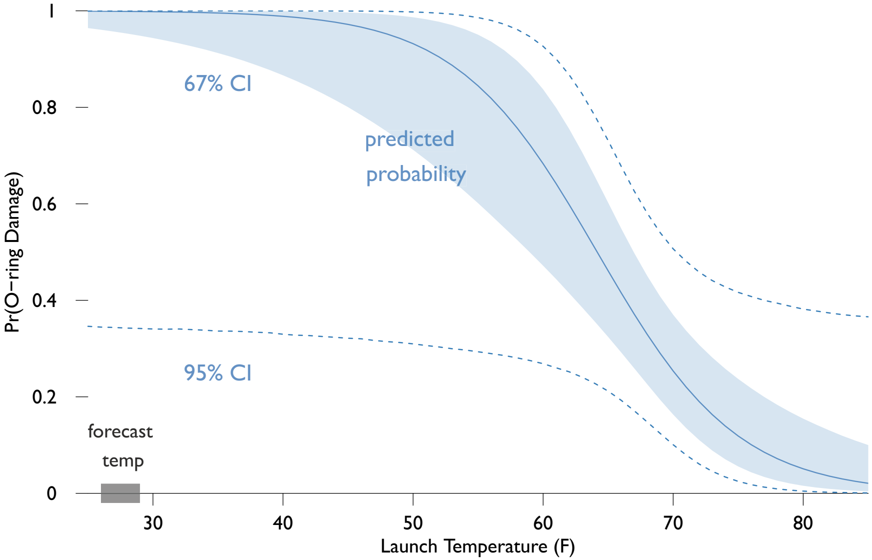
A picture shows model predictions *and* uncertainty



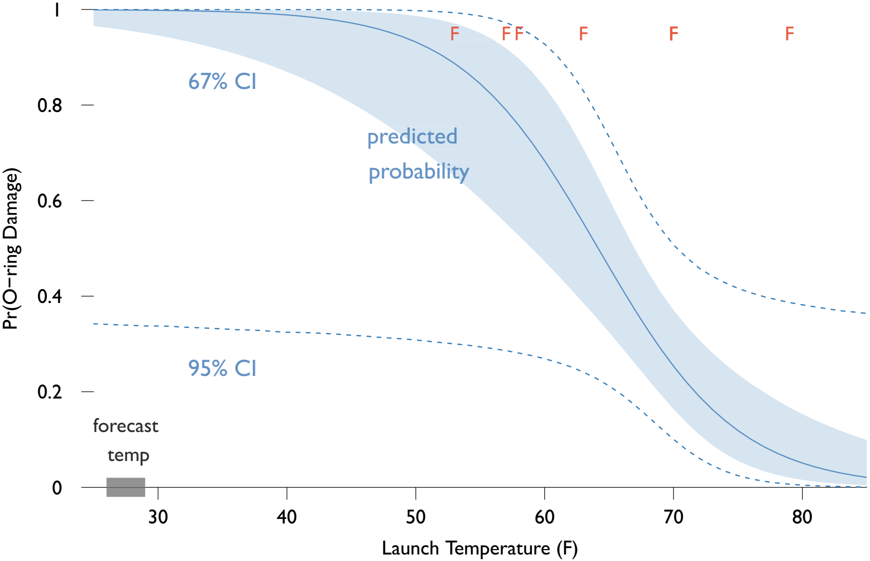
A picture shows model predictions *and* uncertainty



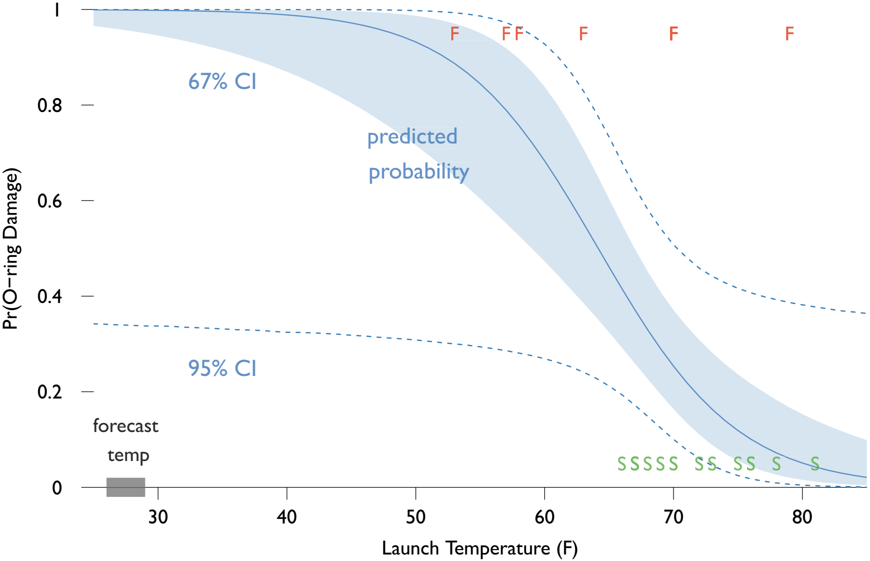
A picture shows model predictions *and* uncertainty



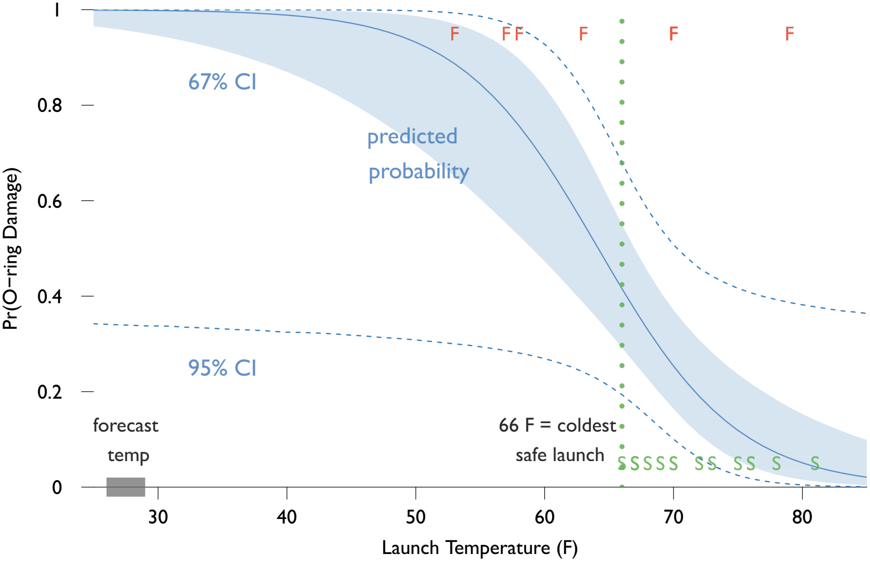
...and gives a more precise sense of how reckless it was to launch at 29 F



When possible, it's good to show the data giving rise to the model

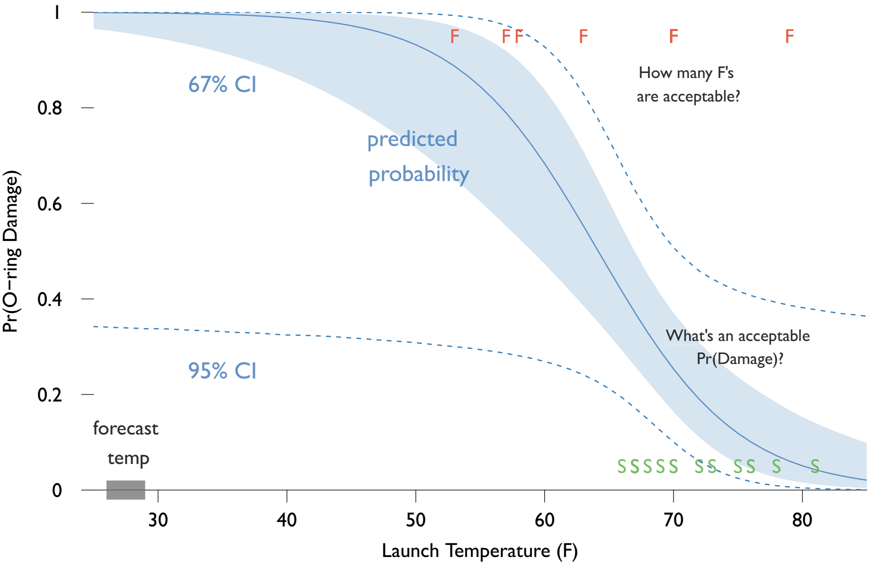


Remembering that the **Failures** are only meaningful compared to **Successes**



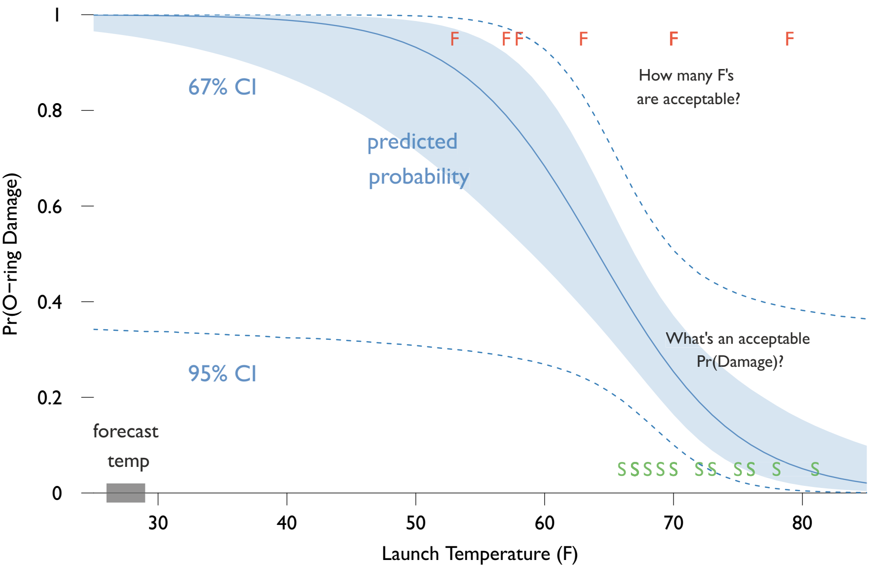
Looking at the data we might think launches <66 F are certain failures

This inference is based on an unstated model



The estimated logit model should give us pause

There is a significant risk of failure across the board



What is an acceptable risk of O-ring failure?

Was the shuttle safe at any temperature?

In a hearing, Richard Feynmann dramatically
showed O-rings lose resilence when cold
by dropping one in his ice water.

Experiment cut through weeks of technical gibberish
concealing flaws in the O-ring

But it shouldn't have taken a Nobel laureate:

any scientist with a year of statistical training could have
used the launch record to reach the same conclusion
And it would take no more than a single graphic to show the result

The *Challenger* launch decision

Lessons for social scientists:

Even relatively simple models and data are easier to understand with visuals

Tables can hide strong correlations

Imagine what might be hiding in datasets with dozens of variables?

Or in models with complex functional forms?

Visuals help make discussion more substantive

See the size of the effect, not just the sign

Make relative judgments of the importance of covariates

Make measured assesments of uncertainty – not just "accept/reject"

Coefficients are not enough

Some limits of typical presentations of statistical results:

- Everything written in terms of arcane intermediate quantites (for most people, this includes logit coefficients)

- Little effort to transform results to the scale of the quantities of interest
  $\rightarrow$ really want the conditional expectation, $\mathbb{E}(y|x)$

- Little effort to make informative statements about estimation uncertainty
  $\rightarrow$ really want to know how uncertain is $\mathbb{E}(y|x)$

- Little visualization at all, or graphs with low data-ink ratios

We will explore a simple dataset using a simple model of voting

People either vote ($\mathrm{Vote}_i = 1$),
or they don't ($\mathrm{Vote}_i = 0$)

Many factors could influence turn-out;
we focus on age and education

Data from National Election Survey in 2000.
"Did you vote in 2000 election?"

vote00	age	hsdeg	coldeg
1	49	1	0
0	35	1	0
1	57	1	0
1	63	1	0
1	40	1	0
1	77	0	0
0	43	1	0
1	47	1	1
1	26	1	1
1	48	1	0
...			

Logit of Decision to Vote, 2000 Presidential NES

	est.	s.e.	*p*-value
Age	0.074	0.017	0.000
Age^2	-0.0004	0.0002	0.009
High School Grad	1.168	0.178	0.000
College Grad	1.085	0.131	0.000
Constant	-3.05	0.418	0.000

Age enters as a quadratic to allow the probability of voting to first rise and eventually fall over the life course

Results look sensible, but what do they mean?

Which has the bigger effect, age or education?

What is the probability a specific person will vote?

An Alternative to Eye-glazing Tables

1. Run your model as normal. Treat the output as an intermediate step.

An Alternative to Eye-glazing Tables

1. Run your model as normal. Treat the output as an intermediate step.

2. Translate your model results back into the scale of the response variable

    - Modeling war? Show the change in probability of war associated with X
    - Modeling counts of crimes committed? Show how those counts vary with X
    - Unemployment rate time series?
      Show how a change in X shifts the unemployment rate over the following t years

An Alternative to Eye-glazing Tables

1. Run your model as normal. Treat the output as an intermediate step.

2. Translate your model results back into the scale of the response variable

   - Modeling war? Show the change in probability of war associated with X
   - Modeling counts of crimes committed? Show how those counts vary with X
   - Unemployment rate time series?
     Show how a change in X shifts the unemployment rate over the following t years

3. Calculate or simulate the uncertainty in these final quantities of interest

An Alternative to Eye-glazing Tables

1. Run your model as normal. Treat the output as an intermediate step.
2. Translate your model results back into the scale of the response variable
    - Modeling war? Show the change in probability of war associated with X
    - Modeling counts of crimes committed? Show how those counts vary with X
    - Unemployment rate time series?
      Show how a change in X shifts the unemployment rate over the following t years
3. Calculate or simulate the uncertainty in these final quantities of interest
4. Present visually as many scenarios calculated from the model as needed

A bit more formally... (KTW, 2000, *AJPS*)

We want to know the behavior of $\mathbb{E}(y|\mathbf{x})$ as we vary $\mathbf{x}$.

In non-linear models with multiple regressors, this gets tricky.

The effect of x_1 depends on all the other x's and $\hat{\beta}$'s

Generally, we will need to make a set of "counterfactual" assumptions:
$x_1 = a, \qquad x_2 = b, \qquad x_3 = c, \quad \ldots$

- Choose $a, b, c, \ldots$ to match a particular counterfactual case of interest *or*
- Hold all but one of the x's at their mean values (or other baseline, such as the factual values by case), then systematically vary the remaining x.

The same trick works if we are after differences in y related to changes in $\mathbf{x}$, such as

$$\mathbb{E}(y|\mathbf{x}_{\text{scen2}} - y|\mathbf{x}_{\text{scen1}}) \qquad \text{or} \qquad \mathbb{E}(y|\mathbf{x}_{\text{scen2}} \;/\; y|\mathbf{x}_{\text{scen1}})$$

Our goal to obtain "quantities of interest," like

- Expected Values: $\mathbb{E}(y|\mathbf{x}_c)$
- Differences: $\mathbb{E}(y|\mathbf{x}_{c2} - y|\mathbf{x}_{c1})$
- Risk Ratios: $\mathbb{E}(y|\mathbf{x}_{c2} \ / \ y|\mathbf{x}_{c1})$
- or any other function of the above

for some counterfactual $\mathbf{x}_c$'s.

For our Voting example, that's easy – just plug $\mathbf{x}_c$ into

$$\mathbb{E}(y|\mathbf{x}_c) = \frac{1}{1 + \exp(-\mathbf{x}_c\beta)}$$

Getting confidence intervals is harder, but there are several options:

- For maximum likelihood models,
  simulate the response conditional on the regressors

  These simulations can easily be summarized as CIs: sort them and take percentiles

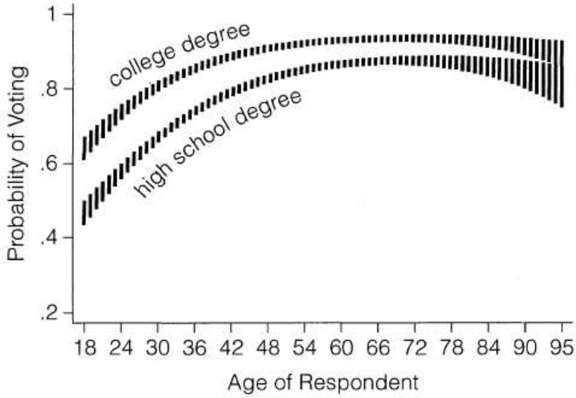
  See King, Tomz, and Wittenberg, 2000, *American Journal of Political Science*, and the `Zelig` or `simcf` packages for `R` or `Clarify` or `margins` for Stata.

- For Bayesian models, usual model output *is* a set of posterior draws

  See Andrew Gelman and Jennifer Hill, 2006, *Data Analysis Using Hierarchical/ Multilevel Models*, Cambridge UP.

Once we have the quantities of interest and confidence intervals,
we're ready to make some graphs...but how?

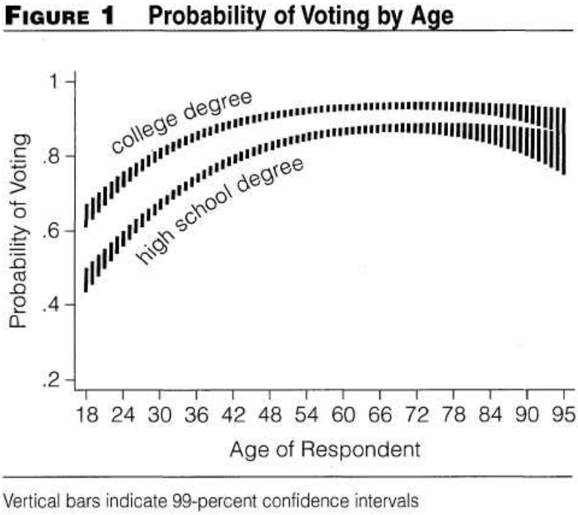
FIGURE 1 **Probability of Voting by Age**



Vertical bars indicate 99-percent confidence intervals

Here is the graph that King, Tomz, and Wittenberg created for this model

How would we make this ourselves?

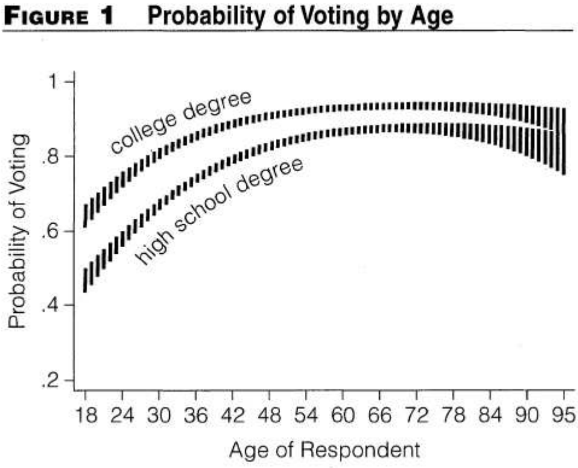


Here is the graph that King, Tomz, and Wittenberg created for this model

How would we make this ourselves?

Could use the default graphics in `Zelig` or `Clarify` (limiting, not as nice as the above)

Could do it by hand (hard)



Vertical bars indicate 99-percent confidence intervals

We'll return to the example in the next session and develop tools for making plots like this one

But note that we don't have to always present model inference in the same format

Students often fixate on plots like this, with a continuous covariate on the x-axis, but there are other options...

American Interest Rate Policy

Example from my own work on central banking (*Bankers, Bureaucrats, and Central Bank Politics*, Cambridge U.P., 2013, Ch. 4)

Federal Reserve Open Market Committee (FOMC) sets interest rates $10\times$/year

Members of the FOMC vote on the Chair's proposed interest rate

Dissenting voters signal whether they would like a higher or lower rate

Dissents are rare but may be symptomatic of how the actual rate gets chosen

Example from my own work on central banking (*Bankers, Bureaucrats, and Central Bank Politics*, Cambridge U.P., 2013, Ch. 4)

Federal Reserve Open Market Committee (FOMC) sets interest rates 10$\times$/year

Members of the FOMC vote on the Chair's proposed interest rate

Dissenting voters signal whether they would like a higher or lower rate

Dissents are rare but may be symptomatic of how the actual rate gets chosen

Many factors could influence interest rate votes:

Individual	Career background
	Appointing party
	Interactions of above
Economy	Expected inflation
	Expected unemployment
Politics	Election cycles

American Interest Rate Policy

My main concern is the individual determinants,
especially career background

I measure career background as a composite variable

Fractions of career spent in each of 5 categories:

Financial Sector	FinExp
Treasury Department	FMExp
Federal Reserve	CBExp
Other Government	GovExp
Academic Economics	EcoExp

These 5 categories plus an (omitted) “Other” must sum to 1.0

American Interest Rate Policy

Because of the composition constraint, to consider the effects of a change in one category, we must adjust the other categories simultaneously

Because of the composition constraint, to consider the effects of a change in one category, we must adjust the other categories simultaneously

	Initial Composition		Hypothetical New Composition
FinExp	0.1	ΔFinExp	0.250
GovExp	0.3	$=$0.15	
FMExp	0.1		
CBExp	0.2	$\rightarrow$	
EcoExp	0.3		
Sum	1.0		1.000

What happens if I increase FinExp by 0.15, but keep all other components the same?

Note – this is close to what I *assume* when I interpret the β for a component as the "effect" of raising that component

	Initial Composition		Hypothetical New Composition
FinExp	0.1	ΔFinExp	0.250
GovExp	0.3	$=$0.15	0.300
FMExp	0.1		0.100
CBExp	0.2	$\rightarrow$	0.200
EcoExp	0.3		0.300
Sum	1.0		1.150

Increasing one component without lowering the combined total of the other components by the same amount leads to a logical fallacy – a career that has 115% total experience!

	Initial Composition		Hypothetical New Composition
FinExp	0.1	ΔFinExp	0.250
GovExp	0.3	$=0.15$	0.300
FMExp	0.1		0.100
CBExp	0.2	$\rightarrow$	0.200
EcoExp	0.3		0.150
Sum	1.0		1.000

Alternatively, if we left out a category (say, EcoExp) as a "reference," we would be implicitly assuming that category alone shrinks to accommodate the increase in FinExp

But that blends the effects of FinExp and EcoExp – so that in our model, the choice of reference category is no longer harmless!

	Initial Composition		Hypothetical New Composition
FinExp	0.1	ΔFinExp	0.250
GovExp	0.3	$=0.15$	0.300
FMExp	0.1		0.100
CBExp	0.4	$\rightarrow$	0.400
EcoExp	0.1		-0.050
Sum	1.0		1.000

And what if EcoExp (still the reference category) starts out smaller than 0.15?

Then our counterfactual would create negative career components!

	Initial Composition		Hypothetical New Composition
FinExp	0.1	ΔFinExp	0.250
GovExp	0.3	$=0.15$	
FMExp	0.1		
CBExp	0.2	$\rightarrow$	
EcoExp	0.3		
Sum	1.0		1.000

When covariates form a composition, we have two problems:

1. to avoid blending effects across components
2. to avoid impossible counterfactuals

I recommend *ratio-preserving counterfactuals*, which uniquely solve both problems

	Initial Composition		Hypothetical New Composition
FinExp	0.1	ΔFinExp	0.250
GovExp	0.3	$=0.15$	0.250
FMExp	0.1		0.083
CBExp	0.2	$\rightarrow$	0.167
EcoExp	0.3		0.250
Sum	1.0		1.000

The transformations above uniquely preserve the ratios among all categories (except FinExp, of course)

Note that now, the effect of a change in one category works through *all* the βs for the composition

We'll fit an ordered probit model to the interest rate data:

$$\begin{aligned}
\Pr(Y_i = \text{ease}|\hat{\beta}, \hat{\tau}) &= \Phi\left(0|X_i\hat{\beta}, 1\right) \\
\Pr(Y_i = \text{assent}|\hat{\beta}, \hat{\tau}) &= \Phi\left(\hat{\tau}|X_i\hat{\beta}, 1\right) - \Phi\left(0|X_i\hat{\beta}, 1\right) \\
\Pr(Y_i = \text{tighten}|\hat{\beta}, \hat{\tau}) &= 1 - \Phi\left(\hat{\tau}|X_i\hat{\beta}, 1\right)
\end{aligned}$$

where Φ is the Normal CDF and τ is a cutpoint

(don't worry if this model is unfamiliar;
suffice it to say we have a nonlinear model and not just linear regression)

American Interest Rate Policy

Running the model yields the following estimates:

Response variable: FOMC Votes (1 = ease, 2 = accept, 3 = tighten)					
EVs	**param.**	**s.e.**	**EVs**	**param.**	**s.e.**
FinExp	−0.021	(0.146)	E(Inflation)	0.019	(0.015)
GovExp	−0.753	(0.188)	E(Unemployment)	−0.035	(0.022)
FMExp	−1.039	(0.324)	In-Party, election year	−0.182	(0.103)
CBExp	−0.142	(0.141)	Republican	−0.485	(0.102)
EcoExp × Repub	0.934	(0.281)	Constant	2.490	(0.148)
EcoExp × Dem	−0.826	(0.202)	Cutpoint (τ)	3.745	(0.067)
N	2957		ln likelihood	−871.68	

Table 1: **Problematic presentation: FOMC member dissenting votes—Ordered probit parameters.** Estimated ordered probit parameters, with standard errors in parentheses, from the regression of a $j = 3$ category variable on a set of explanatory variables (EVs). Although such nonlinear models are often summarized by tables like this one, especially in the social sciences, it is difficult to discern the effects of the EVs listed at right on the probability of each of the j outcomes. Because the career variables *XXX*Exp are logically constrained to a unit sum, even some of the signs are misleading. The usual quantities of interest for an ordered probit model are not the parameters ($\boldsymbol{\beta}$ and τ), but estimates of $\Pr(y_j|\mathbf{x}_c, \boldsymbol{\beta}, \tau)$ for hypothetical levels of the EVs $\mathbf{x}_c$, which I plot in Figure 1.

American Interest Rate Policy

Response variable: FOMC Votes (1 = ease, 2 = accept, 3 = tighten)					
EVs	**param.**	**s.e.**	**EVs**	**param.**	**s.e.**
FinExp	−0.021	(0.146)	E(Inflation)	0.019	(0.015)
GovExp	−0.753	(0.188)	E(Unemployment)	−0.035	(0.022)
FMExp	−1.039	(0.324)	In-Party, election year	−0.182	(0.103)
CBExp	−0.142	(0.141)	Republican	−0.485	(0.102)
EcoExp $\times$ Repub	0.934	(0.281)	Constant	2.490	(0.148)
EcoExp $\times$ Dem	−0.826	(0.202)	Cutpoint (τ)	3.745	(0.067)
N	2957		ln likelihood	−871.68	

How do we interpret these results?

American Interest Rate Policy

Response variable: FOMC Votes (1 = ease, 2 = accept, 3 = tighten)					
EVs	**param.**	**s.e.**	**EVs**	**param.**	**s.e.**
FinExp	−0.021	(0.146)	E(Inflation)	0.019	(0.015)
GovExp	−0.753	(0.188)	E(Unemployment)	−0.035	(0.022)
FMExp	−1.039	(0.324)	In-Party, election year	−0.182	(0.103)
CBExp	−0.142	(0.141)	Republican	−0.485	(0.102)
EcoExp × Repub	0.934	(0.281)	Constant	2.490	(0.148)
EcoExp × Dem	−0.826	(0.202)	Cutpoint (τ)	3.745	(0.067)
N	2957		ln likelihood	−871.68	

Because the model is non-linear,
interpreting coefficients as slopes ($\partial y/\partial\beta$) is grossly misleading

Moreover, the compositional variables are tricky:
If one goes up, the others must go down, to keep the sum =1

Finally, we can't interpret interactive coefficients separately

American Interest Rate Policy

Response variable: FOMC Votes (1 = ease, 2 = accept, 3 = tighten)					
EVs	**param.**	**s.e.**	**EVs**	**param.**	**s.e.**
FinExp	−0.021	(0.146)	E(Inflation)	0.019	(0.015)
GovExp	−0.753	(0.188)	E(Unemployment)	−0.035	(0.022)
FMExp	−1.039	(0.324)	In-Party, election year	−0.182	(0.103)
CBExp	−0.142	(0.141)	Republican	−0.485	(0.102)
EcoExp × Repub	0.934	(0.281)	Constant	2.490	(0.148)
EcoExp × Dem	−0.826	(0.202)	Cutpoint (τ)	3.745	(0.067)
N	2957		ln likelihood	−871.68	

Looking at this table, two obvious question arise:

What is the effect of each covariate on the probability of each kind of vote?

What are the confidence intervals or standard errors for those effects?

American Interest Rate Policy

Response variable: FOMC Votes (1 = ease, 2 = accept, 3 = tighten)					
EVs	**param.**	**s.e.**	**EVs**	**param.**	**s.e.**
FinExp	−0.021	(0.146)	E(Inflation)	0.019	(0.015)
GovExp	−0.753	(0.188)	E(Unemployment)	−0.035	(0.022)
FMExp	−1.039	(0.324)	In-Party, election year	−0.182	(0.103)
CBExp	−0.142	(0.141)	Republican	−0.485	(0.102)
EcoExp × Repub	0.934	(0.281)	Constant	2.490	(0.148)
EcoExp × Dem	−0.826	(0.202)	Cutpoint (τ)	3.745	(0.067)
N	2957		ln likelihood	−871.68	

Cruel to leave this to the reader: it's a lot of work to figure out.

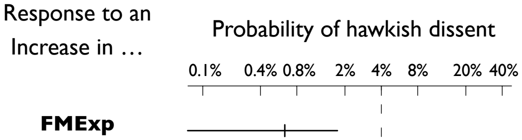
The table above, though conventional, is an intermediate step.

Publishing the table alone is like stopping where Morton-Thiokol did, with pages of technical gibberish – the answers are there, but buried

As the researcher, I should calculate the effects and uncertainty

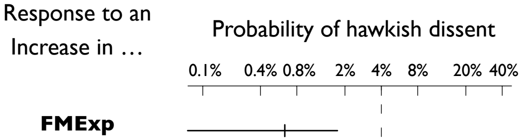
And present them in a readable way

A single graphic achieves both goals

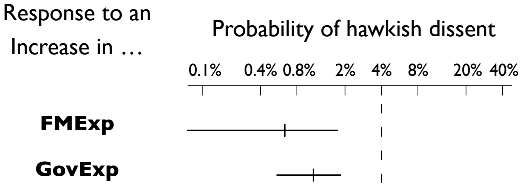


My final graphic will involve small multiples, but explanation should start with a single example

"The average central banker dissents in favor of tighter interest rates 4% of the time. In contrast, former treasury officials in the FOMC dissent 0.6% of the time, with a 95% CI from 0.05% to 2%."



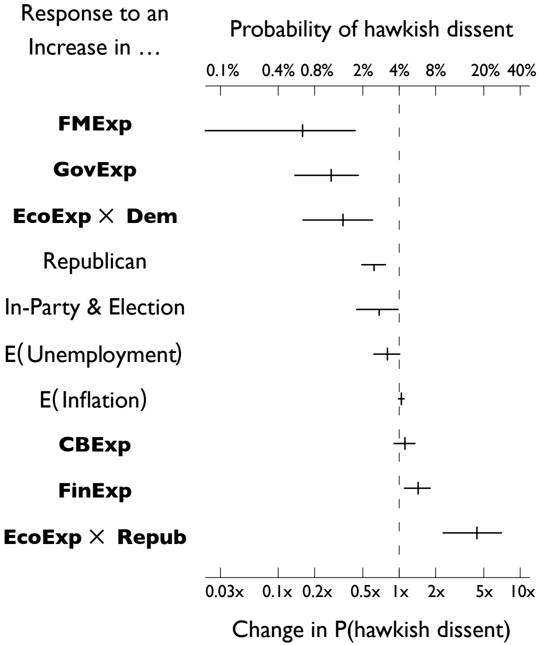
"Other former bureaucrats issue hawkish dissents 1% of the time [95% CI: 0.5 to 2.0], all else equal."



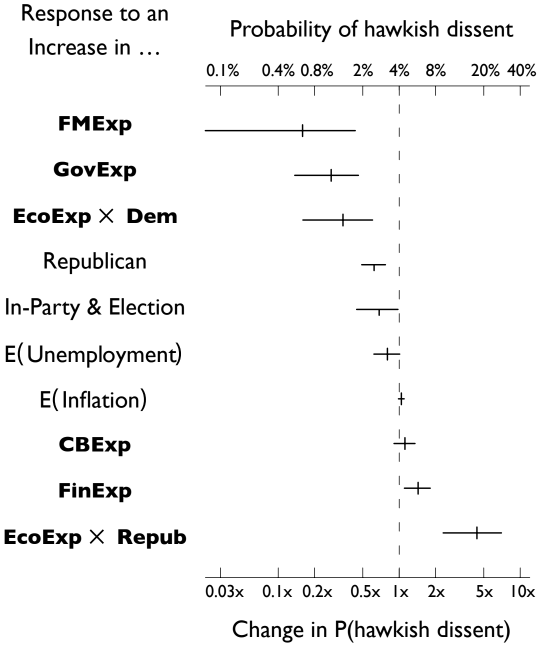
Now that readers understand how to read an individual result, they are ready to explore the graphic on their own.

I can highlight broad trends, then summarize the key findings

But starting by explaining a single instance is critical for effectively using small multiples



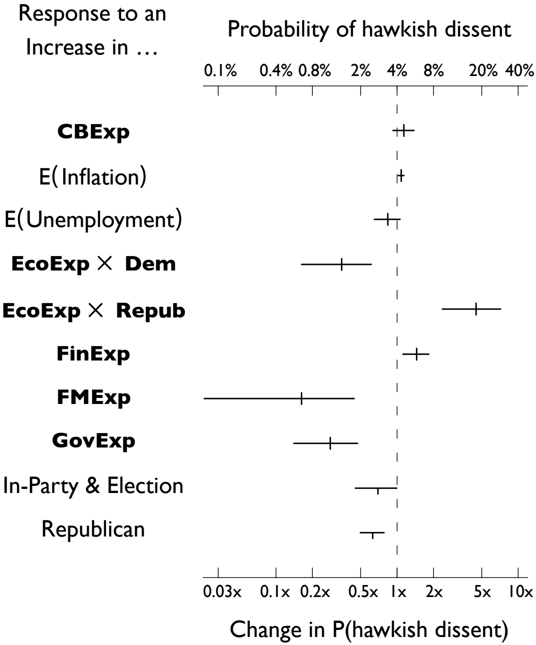
Note that I sorted my scenarios from the smallest to largest effect



Note that I sorted my scenarios from the smallest to largest effect

What if I hadn't?

Most alphabetized dotplots look uninteresting and self-similar

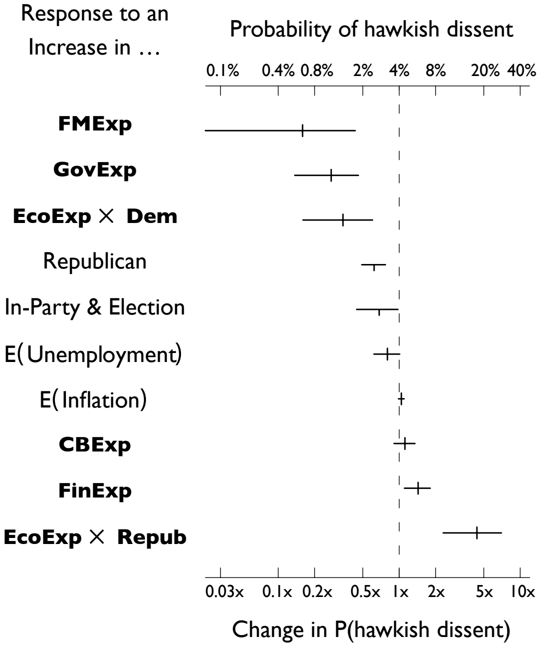


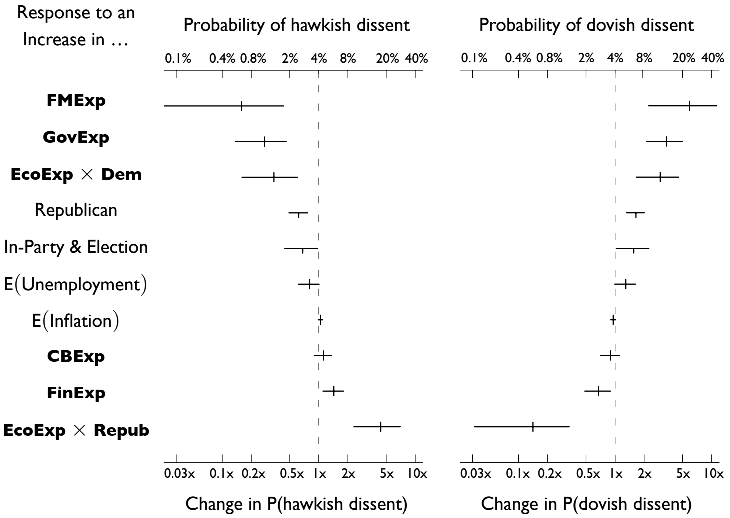
Note that I sorted my scenarios from the smallest to largest effect

What if I hadn't?

Most alphabetized dotplots look uninteresting and self-similar

Diagonalizing makes clear the relative effect sizes *and* whether our variables tend to have large substantive effects

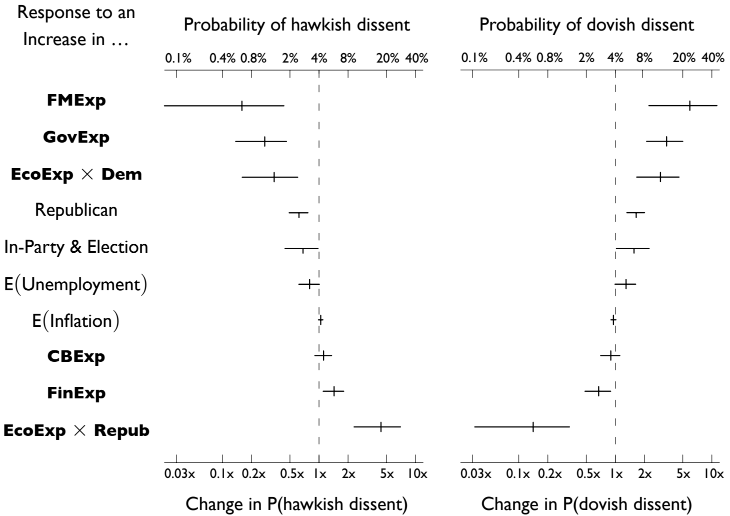




Note that I left out the results for dovish dissents in my presentation

Here they are symmetrical,

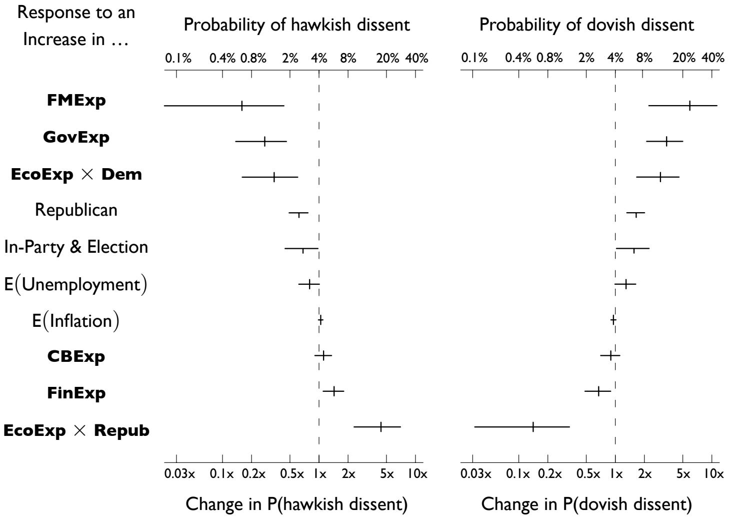
but with 4 or 5 outcome categories, they may not be



How to present ordered probit with many categories?

Temptation: Combine categories before modeling to make a simpler picture

Don't combine categories before estimation – this throws away information!



Instead: Combine categories through simulation *after* estimation

E.g., simulate the probability of "Agree" *or* "Strongly Disagree" on a 5-point scale

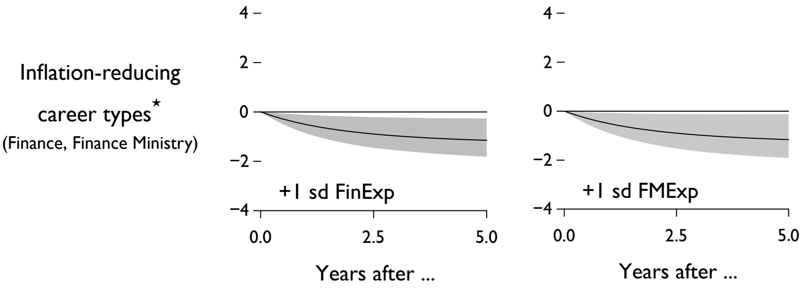
See my MLE lecture on Ordered Choice for examples and `simcf` *code*

Now instead of studying individual central bankers in the United States, we study a panel of 20 central banks across the industrialized world (pre-Euro data)

We ask what effect the average career composition of the central bank policy board has on inflation

Comparative Inflation Performance

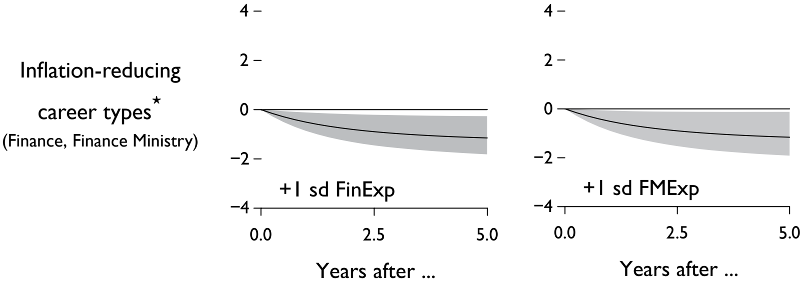
Change in inflation, over time, from changing career composition of the central bank



Now instead of studying individual central bankers in the United States, we study a panel of 20 central banks across the industrialized world (pre-Euro data)

We ask what effect the average career composition of the central bank policy board has on inflation

Change in inflation, over time, from changing career composition of the central bank



We imagine a central bank that initially has central bankers with typical career experience (i.e., the global average in each category)

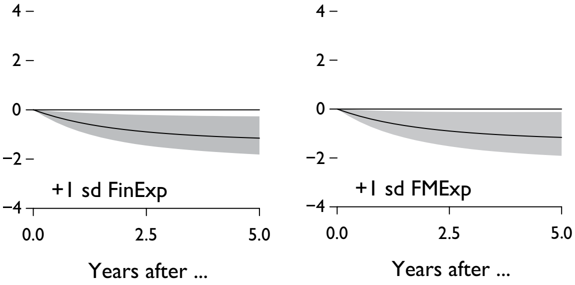
Then, we imagine raising experience in one category (say finance, or FinExp), and use the model to predict how inflation will change over the next 5 years

Change in inflation, over time, from changing career composition of the central bank

Inflation-reducing

career types*

(Finance, Finance Ministry)



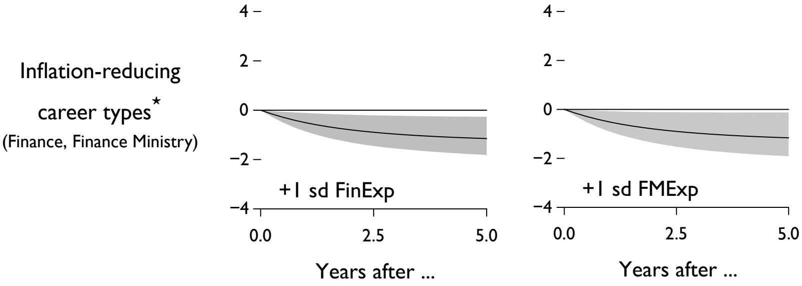
Why not just show a coefficient for each career category?

Two reasons to show the first difference in inflation over time:

1. Raising FinExp means lowering the other categories, so effects are blended across coefficients

Comparative Inflation Performance

Change in inflation, over time, from changing career composition of the central bank



Why not just show a coefficient for each career category?

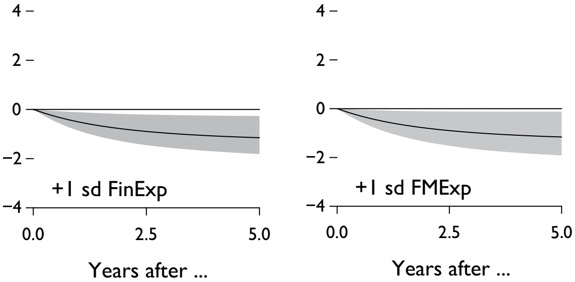
Two reasons to show the first difference in inflation over time:

2. Effects in time series models build over time;
coefficients show (somewhat arbitrary) first period effects

Comparative Inflation Performance

Change in inflation, over time, from changing career composition of the central bank

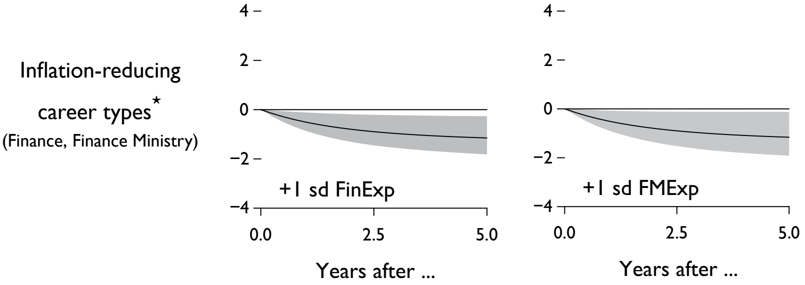
Inflation-reducing career types* (Finance, Finance Ministry)



We simply iterate the KTW simulation algorithm over 5 periods,
computing for each period the difference from inflation under the average board

I used `ldvsimfd()` in the `simcf` package for `R`;
see my course on Panel Data Analysis offered at Essex Summer School

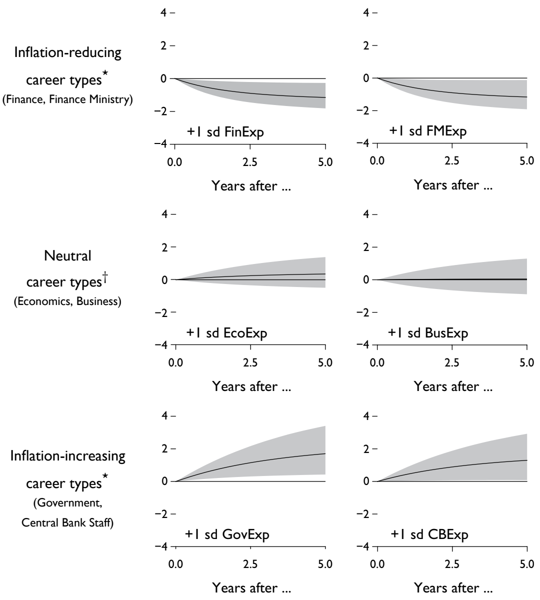
Change in inflation, over time, from changing career composition of the central bank



In the plot above, I show two different scenarios iterated over time: increasing finance experience, or increasing finance ministry experience

Both produce significant reductions in inflation compared to the baseline, and mostly converge to new equilibria after 5 years

Change in inflation, over time, from changing career composition of the central bank



Once we've explained the model, simulation method, and a single plot in our graphic, we can expand to multiple displays

The plot at left replaces an eye-glazing, opaque, and (because of compositional constraints) *misleading* table of regression coefficients

Table 3.7. *Log inflation regressed on central banker characteristics, twenty countries, 1973 to 2000, quarterly.*

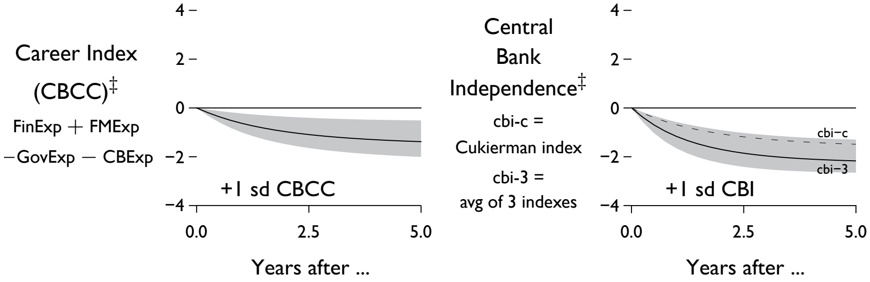
Variable	Expected Sign	DV: ln(Inflation) 1	2	3	4
$\text{FinExp}_{j,t-2}$	$-$	-0.14			-0.09
		(0.08)			(0.07)
$\text{FMExp}_{j,t-2}$	$-/+$	-0.08			-0.13
		(0.06)			(0.06)
$\text{CBExp}_{j,t-2}$	$+/-$	0.12			0.12
		(0.05)			(0.05)
$\text{GovExp}_{j,t-2}$	$+$	0.23			0.19
		(0.08)			(0.08)
$\text{CBI}_{j,t-2}$	$-$	-0.91	-0.92	-0.90	-0.94
		(0.30)	(0.29)	(0.29)	(0.30)
$\text{CBCC}^{\text{med}}_{j,t-2}$	$-$		-0.09	-0.03	
			(0.03)	(0.07)	
$\text{CBI}_{j,t-2} \times \text{CBCC}^{\text{med}}_{j,t-2}$	$-$			-0.12	
				(0.15)	
$(\text{Imports}/\text{GDP})_{j,t-2}$	$-$	-0.02	0.02	0.05	-0.25
		(0.26)	(0.25)	(0.26)	(0.26)
$\%\text{EcDegree}_{j,t-2}$	$-$				0.04
					(0.06)
$\ln \pi_{j,t-1}$		0.97	0.97	0.97	0.96
		(0.04)	(0.04)	(0.04)	(0.04)
$\ln \pi_{j,t-2}$		-0.03	-0.03	-0.03	-0.01
		(0.04)	(0.04)	(0.04)	(0.04)

People often ask, "What if the journal insists on a table instead of the figure?"

In my experience, no one prefers this table to the graph

Give them both, focus your write-up on the graphic, and make sure the graphic explains everything you wanted to get from the table

Change in inflation, over time, from changing career composition of the central bank



No tradeoffs: The small multiple graphs are more accessible to a broad audience and more useful to specialists than a table

You can always include the table as an appendix for those who want to “look under the hood,” but cast your argument in terms of the graphics

Plan of Session 3: Tools for Visualizing Model Inference

1. Introducing the `tile` graphics package

2. Making a scatterplot using `tile`

3. Model Inference with `tile`:
   Voting `lineplot`s of expected values, first differences, and relative risks

Wanted: an easy-to-use R package that

1. takes as input the *output* of estimated statistical models
2. makes a variety of plots for model interpretation
3. plots "triples" (lower, estimate, upper) from estimated models well
4. lays out these plots in a tiled arrangement (small multiples)
5. takes care of axes, titles, and other fussy details

Wanted: an easy-to-use `R` package that

1. takes as input the *output* of estimated statistical models
2. makes a variety of plots for model interpretation
3. plots "triples" (lower, estimate, upper) from estimated models well
4. lays out these plots in a tiled arrangement (small multiples)
5. takes care of axes, titles, and other fussy details

With considerable work, one could

- coerce `R`'s basic graphics to do this badly
- or get `lattice` to do this fairly well for a specific case

But an easy-to-use, general solution is lacking

The `tile` package

My answer is the `tile` package, written using `R`'s `grid` graphics

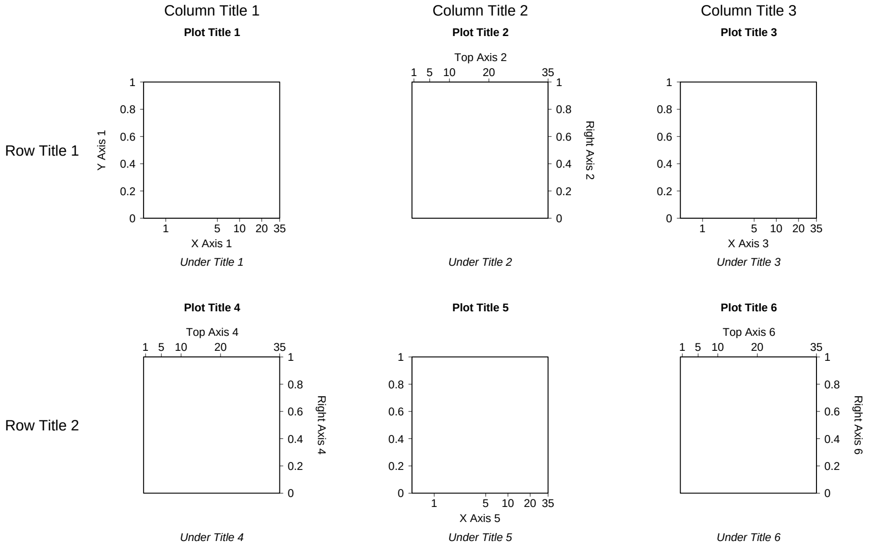
Some basic tile graphic types:

scatter	Scatterplots with fits, CIs, and extrapolation checking
lineplot	Line plots with fits, CIs, and extrapolation checking
ropeladder	Dot plots with CIs and extrapolation checking

Each can take as input draws from the posterior of a regression model

A call to a tile function makes a multiplot layout:

ideal for small multiples of model parameters


Column Title 1
Column Title 2
Column Title 3
Plot Title 1
Plot Title 2
Plot Title 3
Top Axis 2
Row Title 1
Y Axis 1
Right Axis 2
X Axis 1
X Axis 3
Under Title 1
Under Title 2
Under Title 3
Plot Title 4
Plot Title 5
Plot Title 6
Top Axis 4
Top Axis 6
Row Title 2
Right Axis 4
Right Axis 6
X Axis 5
Under Title 4
Under Title 5
Under Title 6

Building a scatterplot: `tile` package warm-up

In my graphics class, I have students build a scatterplot "from scratch"

This helps us see the many choices to make, and implications for:

1. perception of the data
2. exploration of relationships
3. assessment of fit

A good warm up for `tile` before the main event (application to models)

See how `tile` helps follow Tufte's recommendations

Building a scatterplot: Redistribution example

Data on

political party systems

and

redistributive effort from various industrial countries

Source of data & basic plot:

Torben Iversen & David Soskice, 2002,
"Why do some democracies redistribute more than others?" Harvard University.

Building a scatterplot: Redistribution example

Concepts for this example (electoral systems and the welfare state):

Effective number of parties:

- Number of parties varies across countries
- Electoral rules largely determine potential number parties
  - Winner take all (US) $\rightarrow \approx 2$ parties.
  - Proportional representation $\rightarrow$ more parties
- To see this, we need to discount trivial parties and use *effective number of parties*.

Building a scatterplot: Redistribution example

Concepts for this example (electoral systems and the welfare state):

Effective number of parties:

- Number of parties varies across countries
- Electoral rules largely determine potential number parties
  - Winner take all (US) $\rightarrow \approx 2$ parties.
  - Proportional representation $\rightarrow$ more parties
- To see this, we need to discount trivial parties and use *effective number of parties*.

Poverty reduction:

Building a scatterplot: Redistribution example

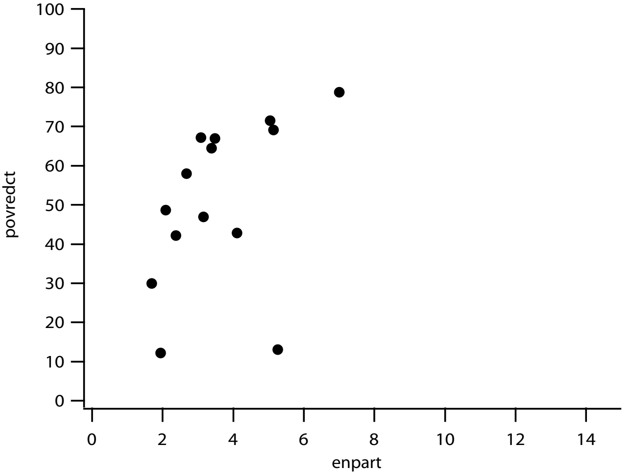
Concepts for this example (electoral systems and the welfare state):

Effective number of parties:

- Number of parties varies across countries
- Electoral rules largely determine potential number parties
  - Winner take all (US) $\rightarrow \approx 2$ parties.
  - Proportional representation $\rightarrow$ more parties
- To see this, we need to discount trivial parties and use *effective number of parties*.

Poverty reduction:

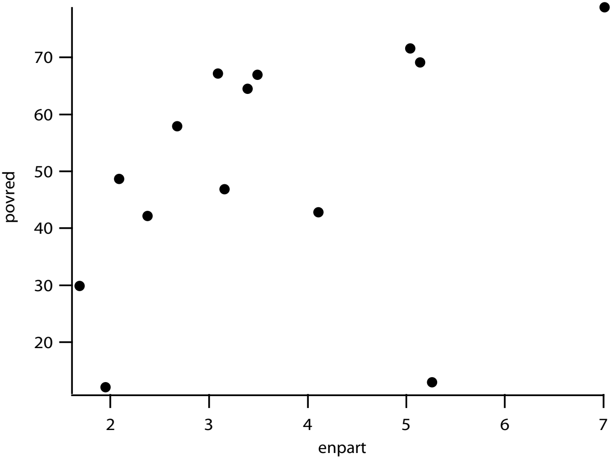
- Percent lifted out of poverty by taxes and transfers.
- Poverty = an income below 50% of mean income.



Initial plotting area is often oddly shaped (I've exaggerated)

This plotting area hides the relationship. Sometimes it can even exclude data!

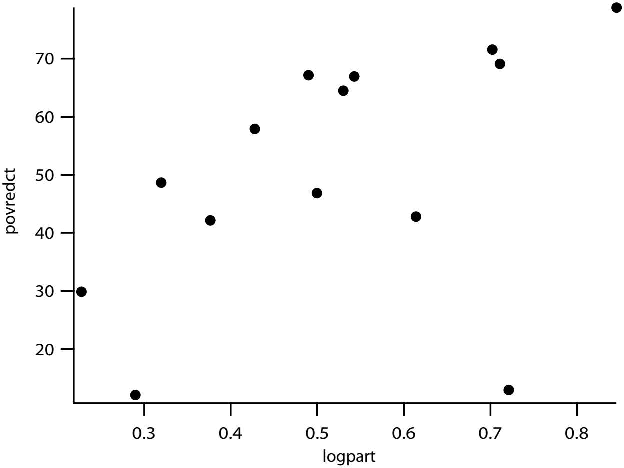
Aside: Filled symbols are good for a little data, but open symbols are better when data overlap



Critical *user* decision: Impose sensible, data based plot limits

Don't just leave this to your package to decide

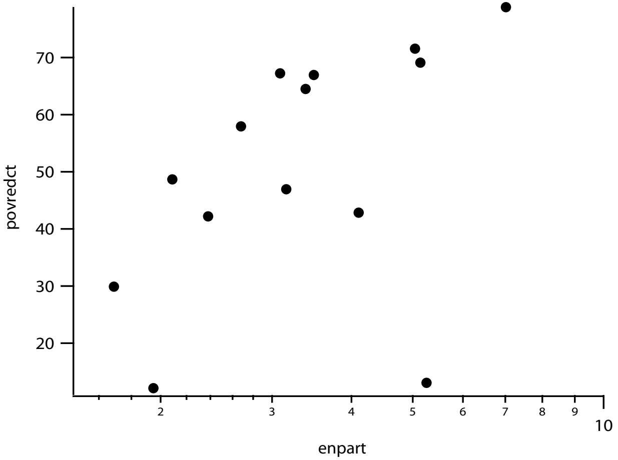
There appears to be a curvilinear relationship. We can bring that out with...



Log scaling.

But why print the exponents?

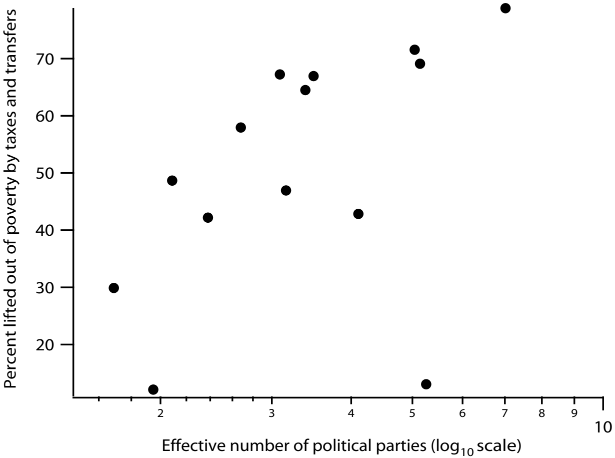
Logs aren't intuitive for many readers, but they don't need to even know we are using them in a graphic...



To make log scales easier for *everyone* to read, use a log scale but supply linear labels...

That is, plot the tick markets at the log values (exponents), but label them with the original linear scale numbers corresponding to those tick marks

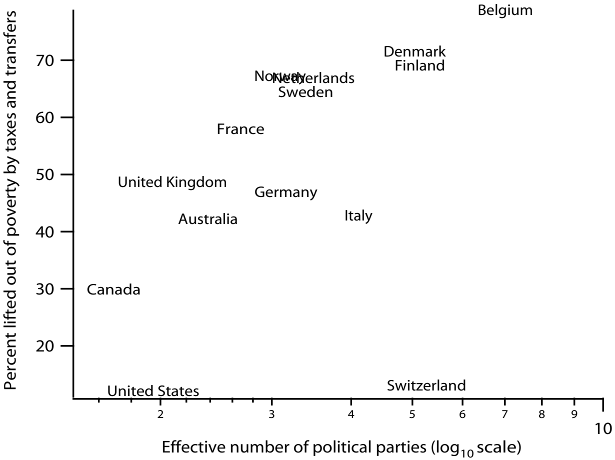
Next problem: Why use abbreviated computer labels for our variables?



Computer labels should stay in the computer

Write out informative axis titles

Next question: What are those outliers?

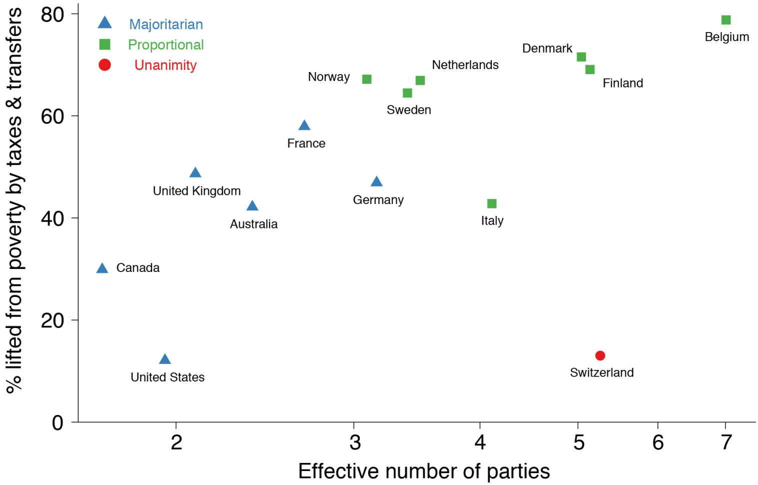


With only a little data & some big outliers, we should show the name of each case as a label

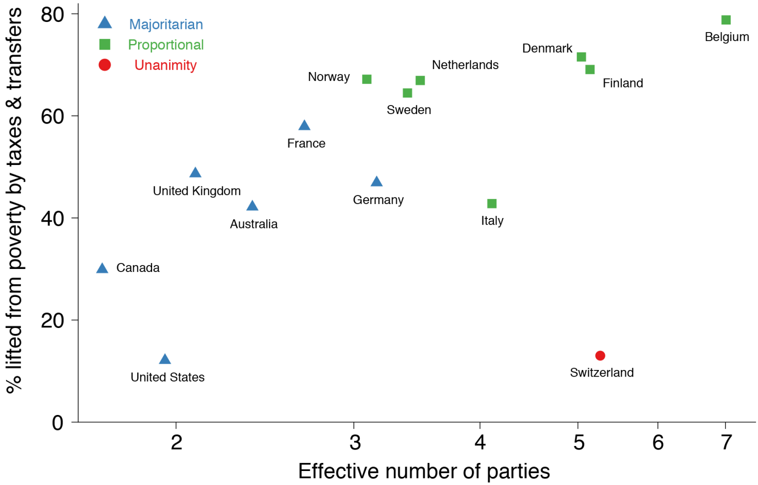
Sometimes we can just replace our plotted points with these labels

Here, let's combine the glyph (symbol) and text label for each point, so that we can use our glyphs to encode a third variable

Next, we can try to figure out what makes the US and Switzerland so different

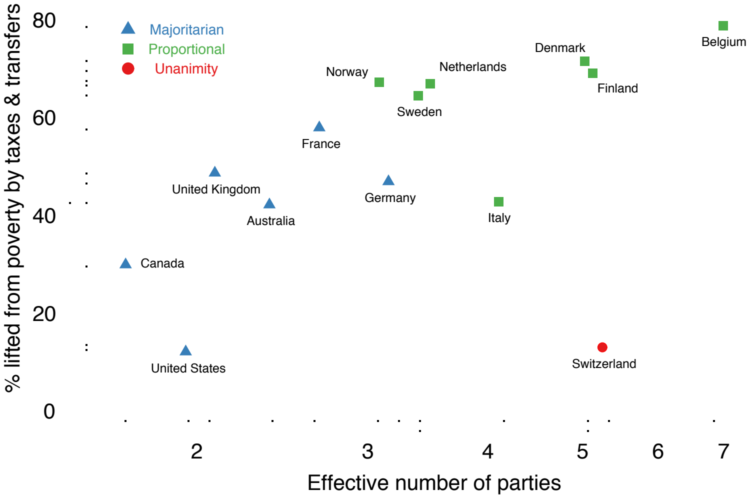


This plot and following plots are made using `scatter` (`tile` package in `R`)



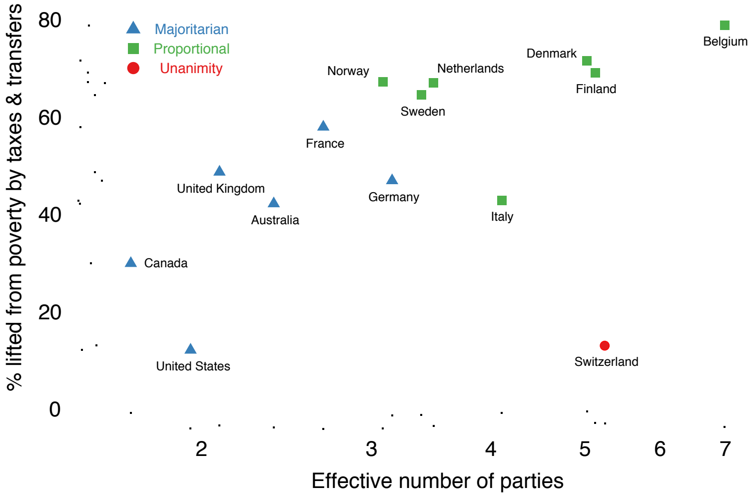
Scatterplots relate two distributions.

Why not make those marginal distributions explicit?

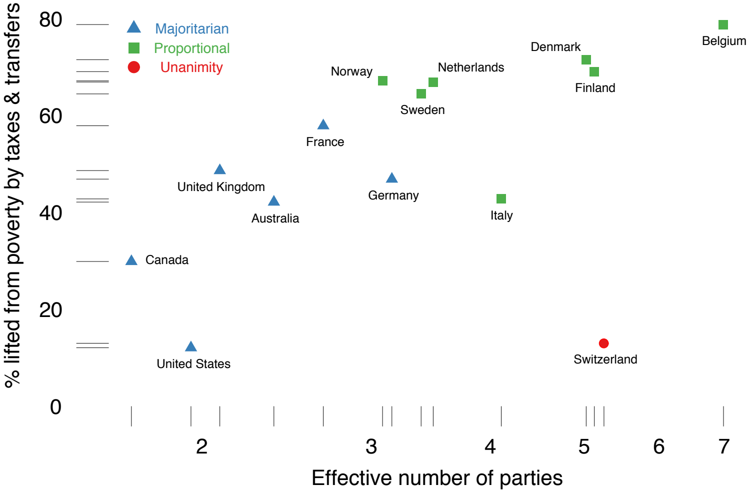


Rugs **accomplish this by replacing the axis lines with the plots**

We could choose any plotting style: from the histogram-like dots...

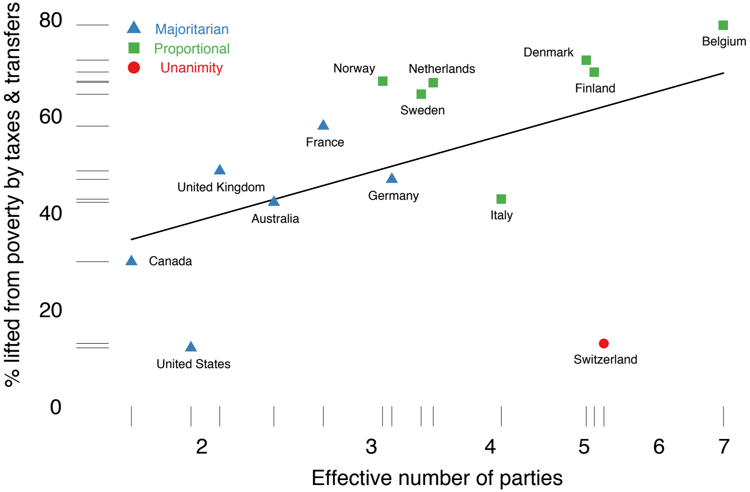


...to a strip of `jittered` data...



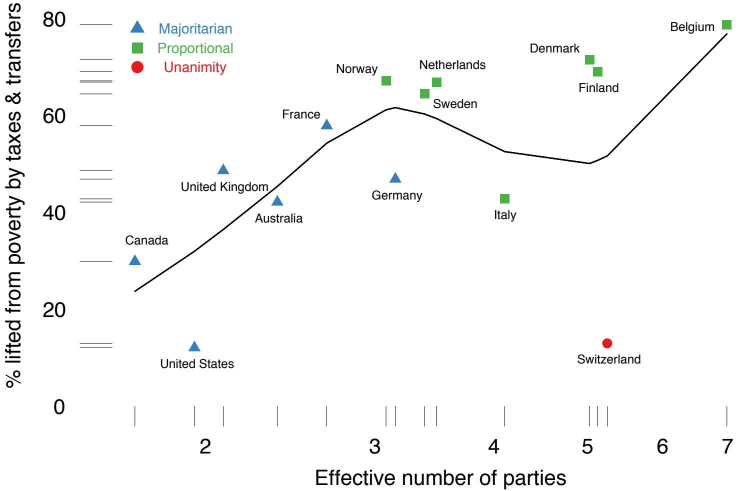
...to a set of very thin `lines` marking each observation

Because we have so few cases, thin lines work best for this example



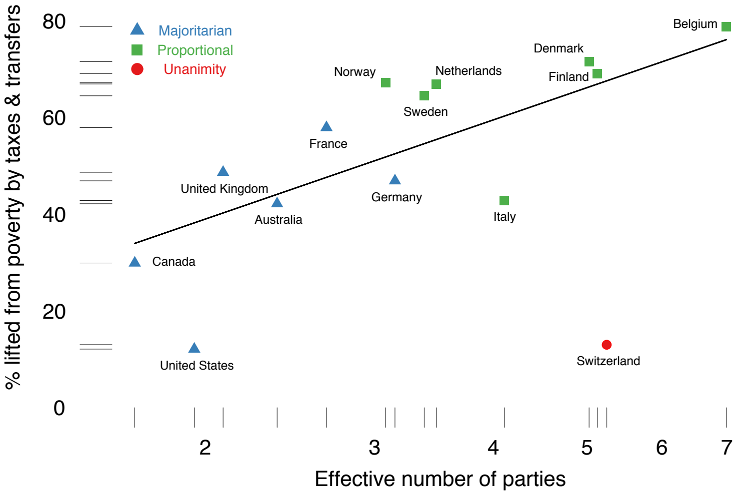
Let's add a parametric model of the data: a least squares fit line

`tile` can do this for us

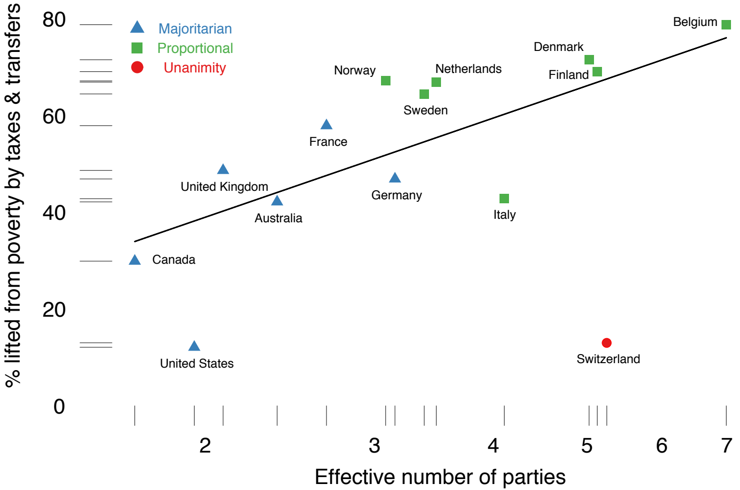


But we don’t have to be parametric

A local smoother, like loess, often helps show non-linear relationships

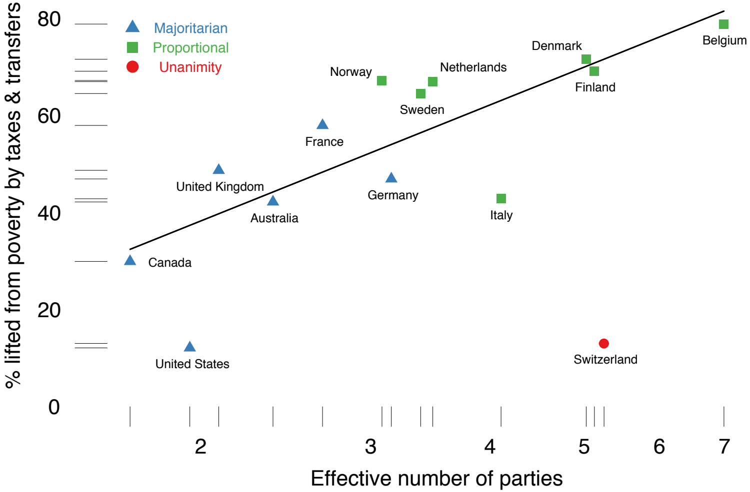


M-estimators weight observations by an influence function to minimize the influence of outliers



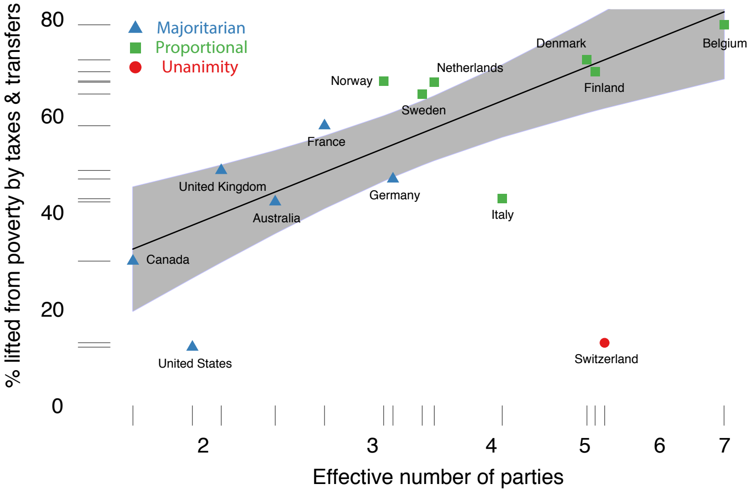
Even with an M-estimator, every outlier has some influence

Thus any one distant outlier can bias the result



A robust and resistant MM-estimator, shown above, largely avoids this problem

Only a (non-outlying) fraction of the data influence this fit. `rlm(method="MM")`



In our final plot, we add 95 percent confidence intervals for the MM-estimator

A measure of uncertainty is essential to reader confidence in the result

Three steps to make tile plots

1. **Create data traces.** Each trace contains the data and graphical parameters needed to plot a single set of graphical elements to one or more plots.

Three steps to make tile plots

1. **Create data traces.** Each trace contains the data and graphical parameters needed to plot a single set of graphical elements to one or more plots.
   - Could be a set of points, *or* text labels, *or* lines, *or* a polygon

1. **Create data traces.** Each trace contains the data and graphical parameters needed to plot a single set of graphical elements to one or more plots.
   - Could be a set of points, *or* text labels, *or* lines, *or* a polygon
   - Could be a set of points *and* symbols, colors, labels, fit line, CIs, and/or extrapolation limits

Three steps to make tile plots

1. **Create data traces.** Each trace contains the data and graphical parameters needed to plot a single set of graphical elements to one or more plots.
   - Could be a set of points, *or* text labels, *or* lines, *or* a polygon
   - Could be a set of points *and* symbols, colors, labels, fit line, CIs, and/or extrapolation limits
   - Could be the data for a dotchart, with labels for each line

Three steps to make tile plots

1. **Create data traces.** Each trace contains the data and graphical parameters needed to plot a single set of graphical elements to one or more plots.
   - Could be a set of points, *or* text labels, *or* lines, *or* a polygon
   - Could be a set of points *and* symbols, colors, labels, fit line, CIs, and/or extrapolation limits
   - Could be the data for a dotchart, with labels for each line
   - Could be the marginal data for a rug
   - **All** annotation must happen in this step

Three steps to make tile plots

1. **Create data traces.** Each trace contains the data and graphical parameters needed to plot a single set of graphical elements to one or more plots.
   - Could be a set of points, *or* text labels, *or* lines, *or* a polygon
   - Could be a set of points *and* symbols, colors, labels, fit line, CIs, and/or extrapolation limits
   - Could be the data for a dotchart, with labels for each line
   - Could be the marginal data for a rug
   - **All** annotation must happen in this step
   - Basic traces: `linesTile()`, `pointsile()`, `polygonTile()`, `polylinesTile()`, and `textTile()`
   - Complex traces: `lineplot()`, `scatter()`, `ropeladder()`, and `rugTile()`

Trace functions in tile

Primitive trace functions:

`linesTile`	Plot a set of connected line segments
`pointsTile`	Plot a set of points
`polygonTile`	Plot a shaded region
`polylinesTile`	Plot a set of unconnected line segments
`textTile`	Plot text labels

Complex traces for model or data exploration:

`lineplot`	Plot lines with confidence intervals, extrapolation warnings
`ropeladder`	Plot dotplots with confidence intervals, extrapolation warnings, and shaded ranges
`rugTile`	Plot marginal data rugs to axes of plots
`scatter`	Plot scatterplots with text and symbol markers, fit lines, and confidence intervals

Three steps to make tile plots

1. **Create data traces.** Each trace contains the data and graphical parameters needed to plot a single set of graphical elements to one or more plots.

2. **Plot the data traces.** Using the `tile()` function, simultaneously plot all traces to all plots.

Three steps to make tile plots

1. **Create data traces.** Each trace contains the data and graphical parameters needed to plot a single set of graphical elements to one or more plots.

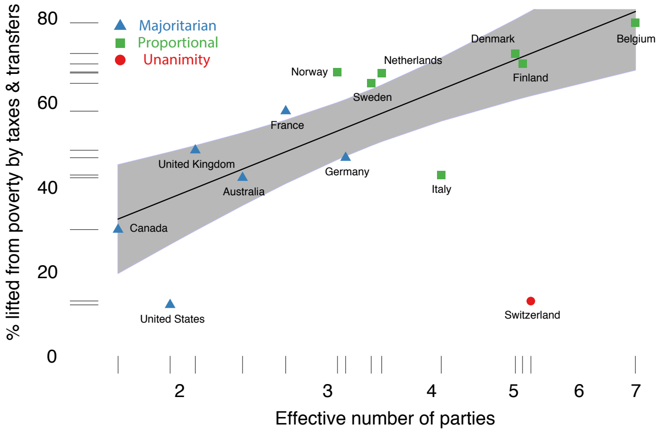
2. **Plot the data traces.** Using the `tile()` function, simultaneously plot all traces to all plots.
   - This is the step where the scaffolding gets made: axes and titles
   - Set up the rows and columns of plots
   - Titles of plots, axes, rows of plots, columns of plots, etc.
   - Set up axis limits, ticks, tick labels, logging of axes

Three steps to make tile plots

1. **Create data traces.** Each trace contains the data and graphical parameters needed to plot a single set of graphical elements to one or more plots.

2. **Plot the data traces.** Using the `tile()` function, simultaneously plot all traces to all plots.

3. **Examine output and revise.** Look at the graph made in step 2, and tweak the input parameters for steps 1 and 2 to make a better graph.



Let's make this plot

CODE EXAMPLE

`inequalityScatter.R`

Plot simulations of QoI

Generally, we want to plot triples: lower, estimate, upper

We could do this for specific **discrete scenarios**, e.g.

Pr(Voting) given five distinct sets of x's

Recommended plot: **Dotplot** with confidence interval lines

Plot simulations of QoI

Generally, we want to plot triples: lower, estimate, upper

We could do this for specific **discrete scenarios**, e.g.

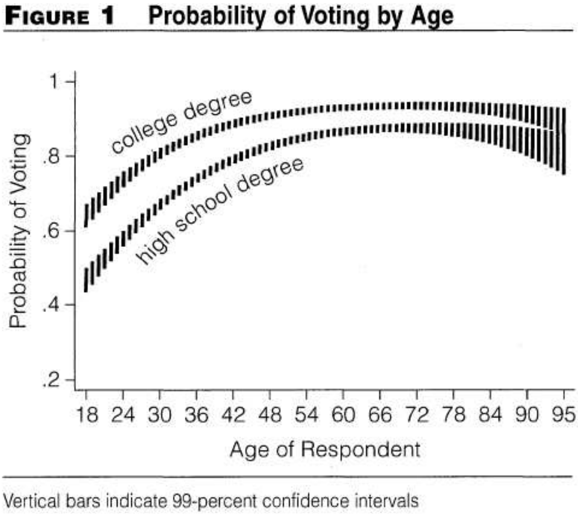
Pr(Voting) given five distinct sets of x's

Recommended plot: **Dotplot** with confidence interval lines

Or for a continuous **stream of scenarios**, e.g.,

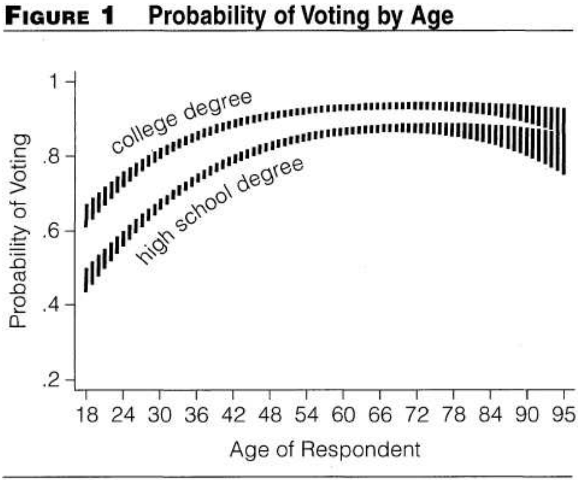
Hold all but Age constant, then calculate Pr(Voting) at every level of Age

Recommended plot: **Lineplot** with shaded confidence intervals



This example is obviously superior to the table of logit coefficients

But is there anything wrong or missing here?



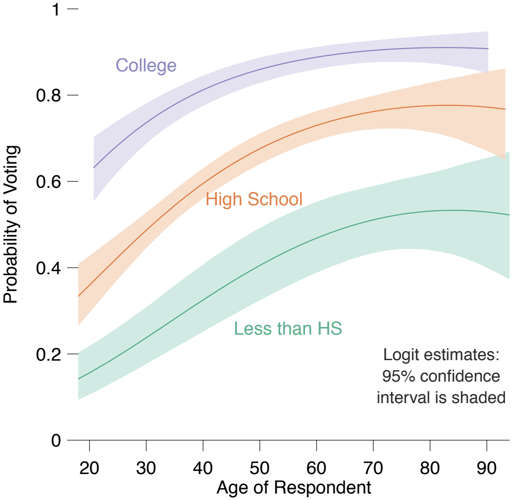
Vertical bars indicate 99-percent confidence intervals

This example is obviously superior to the table of logit coefficients

But is there anything wrong or missing here?

18 year old college grads?!

And what about high school dropouts?



Here is the graphic redrawn in tile

`tile` helps us systematize plotting model results, and helps avoid unwanted extrapolation by limiting results to the convex hull

CODE EXAMPLE

`votingLineplots.R`

Voting Example Redux

Next step: learn to simulate and plot first differences and relative risks

We could do this with our current example.
E.g., hold age fixed and compute the change in Pr(Vote) given an increase in education

But for pedagogical reasons, it will be more useful to add an additional covariate

We now add to our voting model whether the respondent was married

Theory: Marriage should increase voting by increasing concern for a variety of public goods, or by forming ties to a local community, etc.

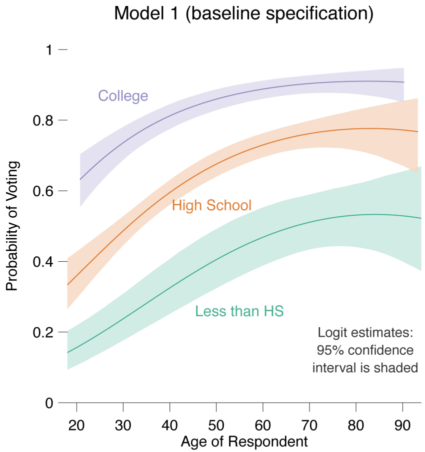
How would this competing model normally be presented?

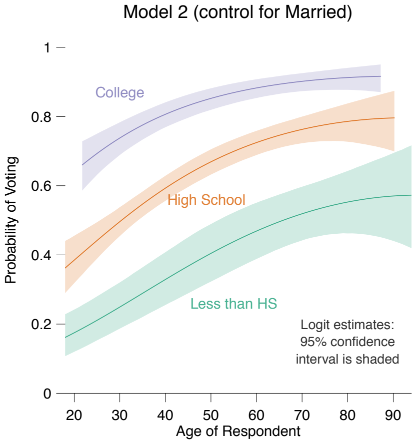
Logit of Decision to Vote, 2000 Presidential NES

	M1	M2
Age	0.075	0.061
	(0.017)	(0.017)
Age^2	-0.0004	-0.0003
	(0.0002)	(0.0002)
High School Grad	1.124	1.099
	(0.180)	(0.181)
College Grad	1.080	1.053
	(0.131)	(0.132)
Married		0.373
		(0.110)
Constant	-3.019	-2.866
	(0.418)	(0.421)
log likelihood	-1101.370	-1099.283
N	1783	1783

Comparing Logistic Regression Models

But we can also compare our results in an intelligible way.

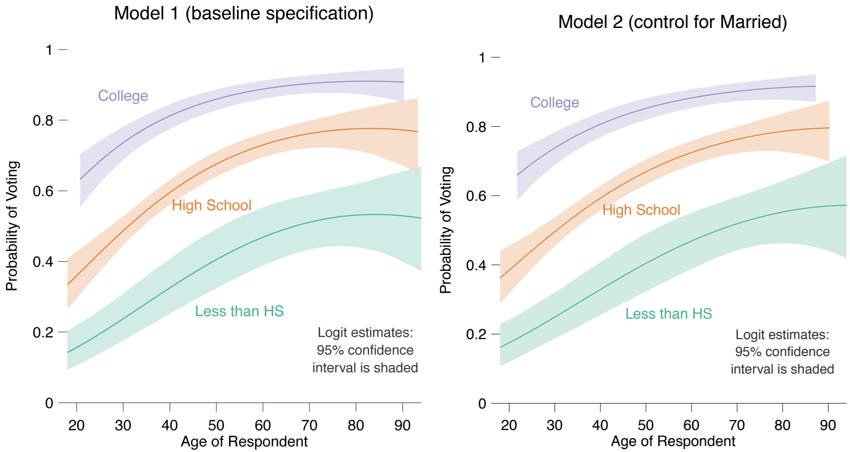




Effects of Age and Education haven't discernably changed

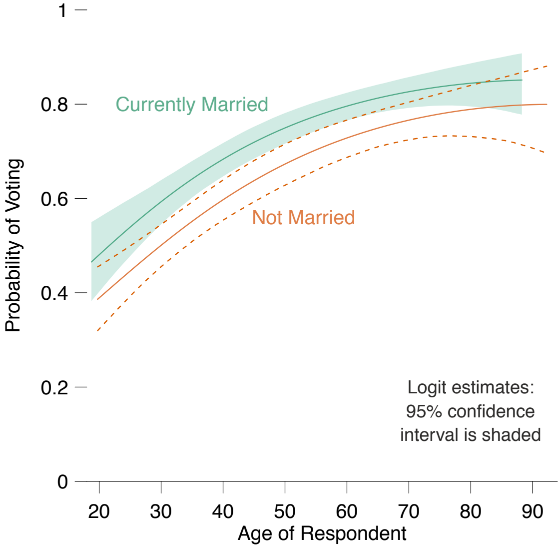
Comparing Logistic Regression Models

But we can also compare our results in an intelligible way.



Our first attempt to show model robustness – we'll find more efficient ways

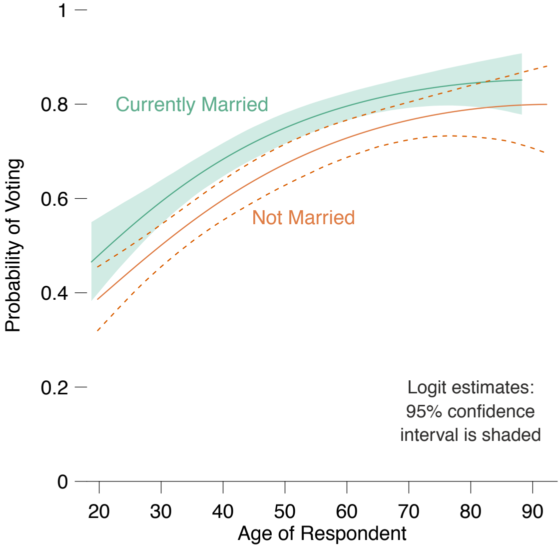
A common misconception about confidence intervals



Ceteris paribus,
marriage has a moderate effect.

Is this effect statistically significant?

A common misconception about confidence intervals



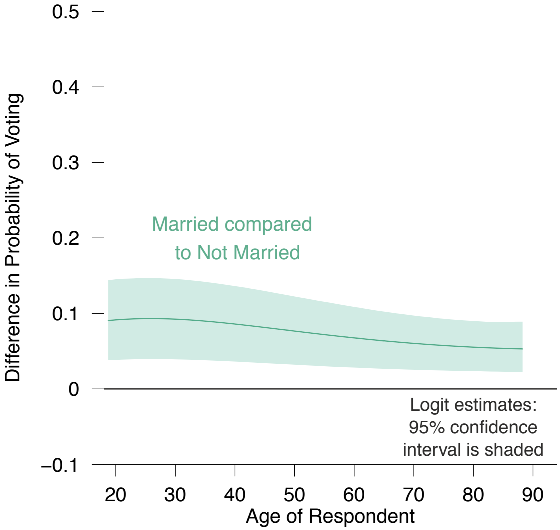
Ceteris paribus, marriage has a moderate effect.

Is this effect statistically significant?

Overlapping CIs of expected values doesn't always imply an insignificant difference

This is not necessarily the case in "close calls"

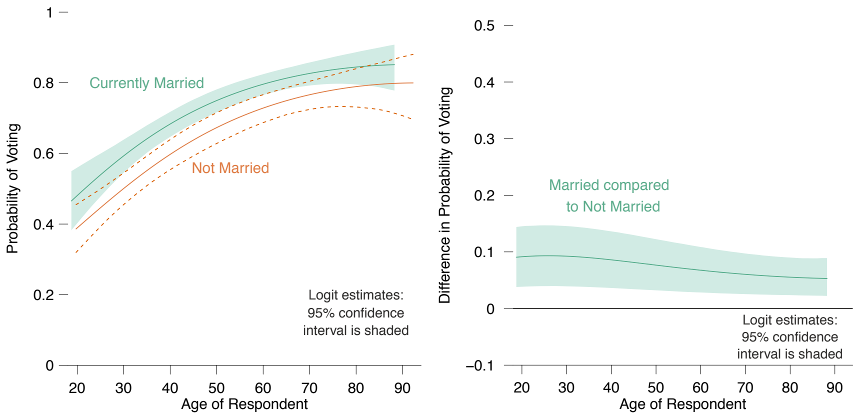
A common misconception about confidence intervals



The right way to assess statistical significance: simulate the CI of the first difference (or relative risk) *directly*

This first difference is always bounded away from zero, hence always significant

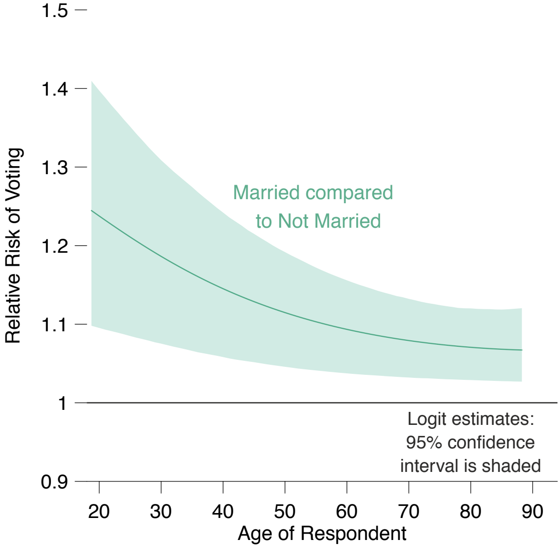
Avoid mistakenly rejecting significant first differences



Expected values estimate both difference & *location*;
demanding a more detailed estimate from the model increases uncertainty

First differences and relative risks estimate the difference only, so they have slightly tighter confidence intervals

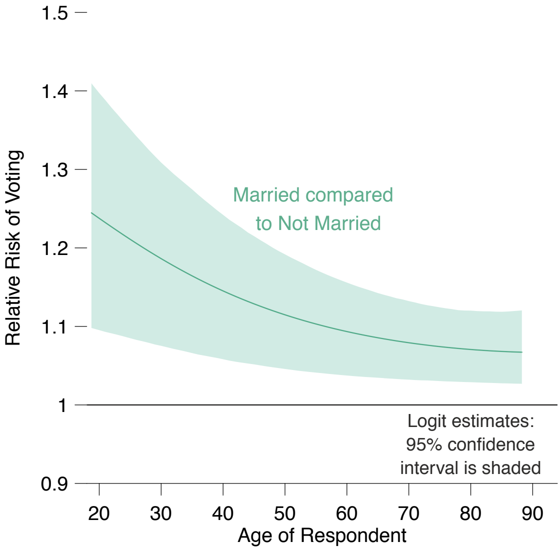
Relative risk plots



Consider showing relative risks instead of (or in support of) first differences

Relative risks show “how many times more likely” a categorical outcome is under the counterfactual

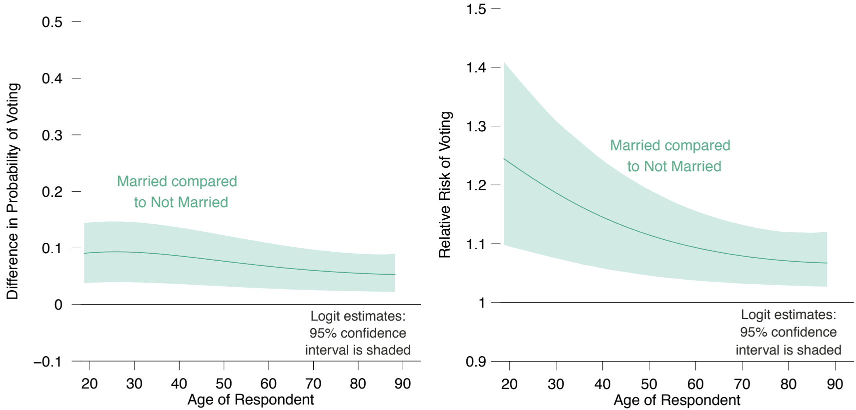
Relative risk plots



For continuous outcomes, RR shows how many times bigger the outcome is under the counterfactual

As with first differences, relative risk should be simulated directly to calculate CIs correctly

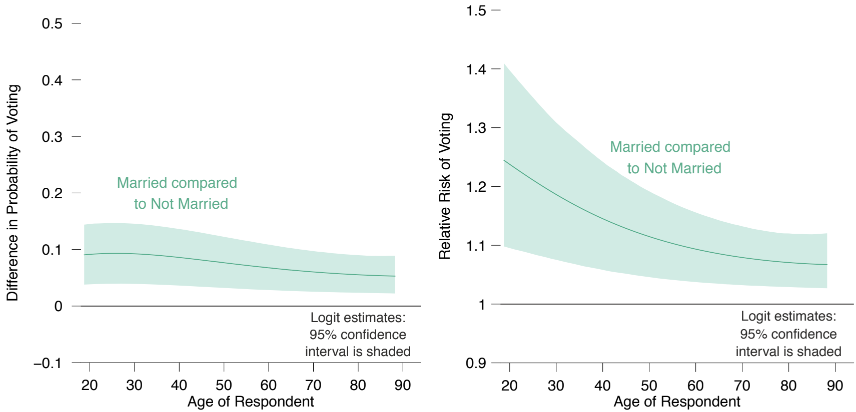
Setting up before-and-after scenarios



Setting up counterfactuals for FDs or RRs is tricky, as we will see in the code

Here I set before and after *age* to the same value (which varied across the plot) but I set Married to different values (0 before, 1 after)

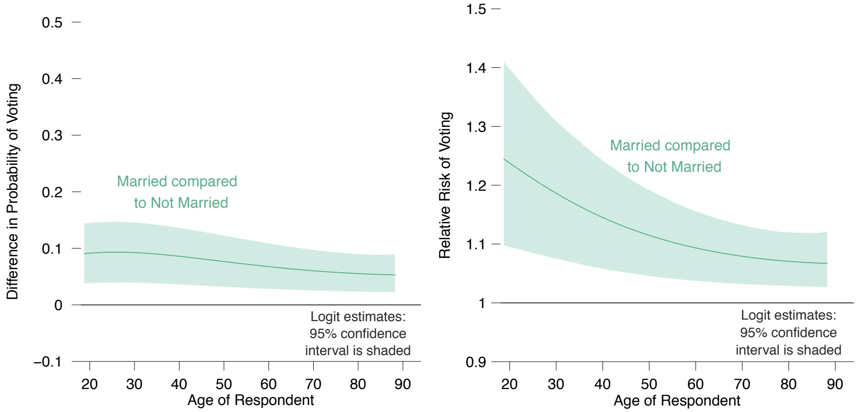
Setting up before-and-after scenarios



Take care in selecting the before and after values of *all* covariates

Most common place to make mistakes, with *huge* substantive consequences

Setting up before-and-after scenarios



RETURN TO CODE EXAMPLE

`votingLineplots.R`

Plan of Session 4: Concepts for Visualizing Model Robustness

1. Compact, systematic presentation of robustness checks
2. Using `ropeladder` plots to show robustness
3. Using `lineplot`s to show robustness
4. Flexible use of `tile` graphics for model inference and robustness

Examples for Session 4

How do Chinese leaders gain power?

Source: Shih, Adolph, and Liu

Method: Bayesian model of partially observed ranks

When do governments choose liberal or conservative central bankers?

Source: Adolph

Method: Zero-inflated compositional data model

How long do central bankers stay in office?

Source: Adolph

Method: Cox proportional hazards model

What explains the tier of European governments controlling health policies?

Source: Adolph, Greer, and Fonseca

Method: Multilevel multinomial logit

When do governments choose liberal or conservative central bankers?

In *Bankers, Bureaucrats, and Central Bank Politics*,
I argue central bankers' career backgrounds explain their monetary policy choices:

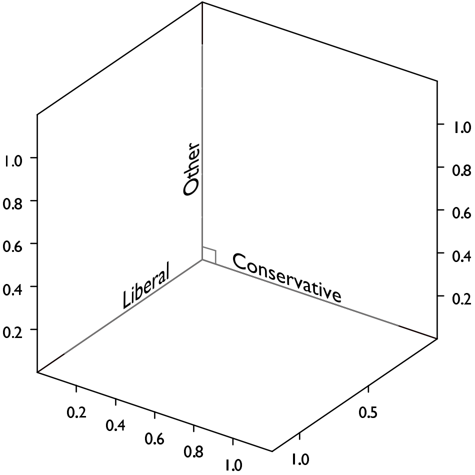
Central bankers with financial sector backgrounds choose more conservative policies, leading to lower inflation but potentially higher unemployment

But how do central banks end up with governors whose careers are conservative?

My claim: more conservative governments should prefer to appoint more conservative central bankers, e.g., those with financial sector backgrounds

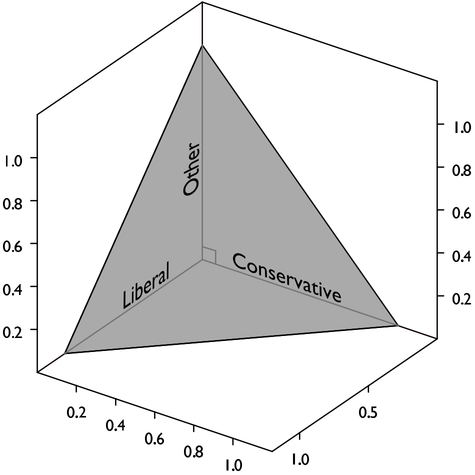
For this model, central bankers career backgrounds are *composed* of shares from liberal, conservative, and "other" career types

The model has a multivariate outcome:
a 3-part composition $\{\text{Conservative}, \text{Liberal}, \text{Other}\}$ that sums to 1



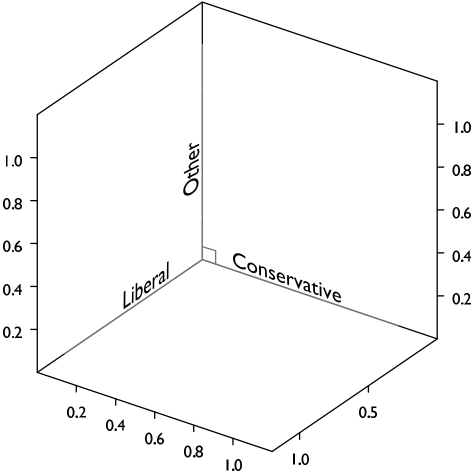
I collect career compositions of central bankers at appointment
from 20 countries over 30 years

How do I visualize a three-part outcome?

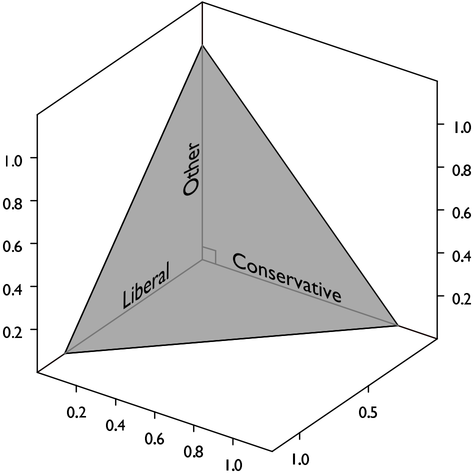


I collect career compositions of central bankers at appointment
from 20 countries over 30 years

How do I visualize a three-part outcome? *Exploit the compositional constraint!*

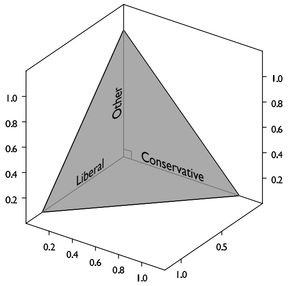


While each of the components {Liberal, Conservative, and Other} can range between 0 and 100%,

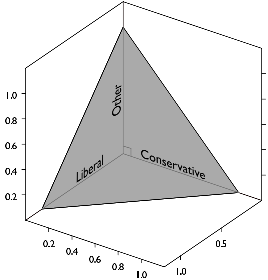


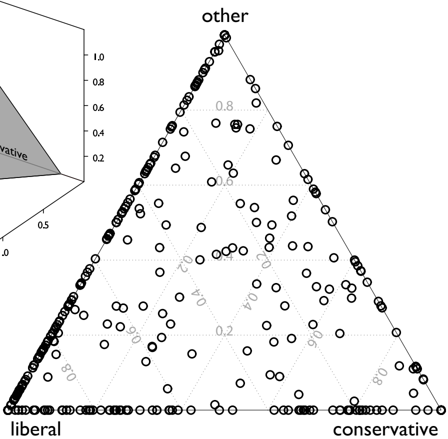
While each of the components {Liberal, Conservative, and Other} can range between 0 and 100%, their sum *must* be 100%

This constrains the possible compositions to the simplex "triangle," which can be represented in 2D even for 3 components



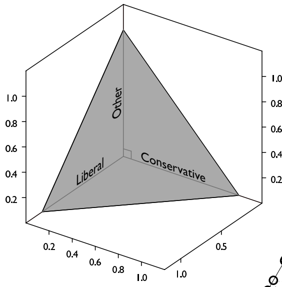
How do I visualize a three-part outcome? *Exploit the compositional constraint!*

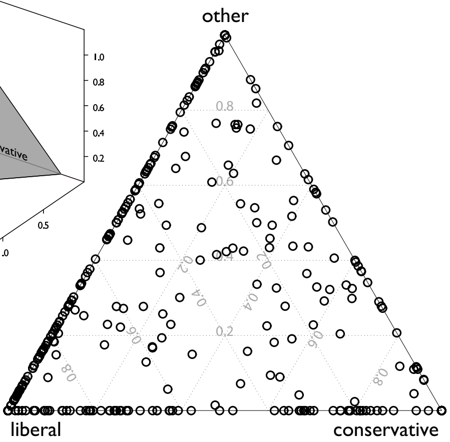




How do I visualize a three-part outcome? *Exploit the compositional constraint!*

We can "pull out" the **simplex**, or the set of points meeting the compositional constraint

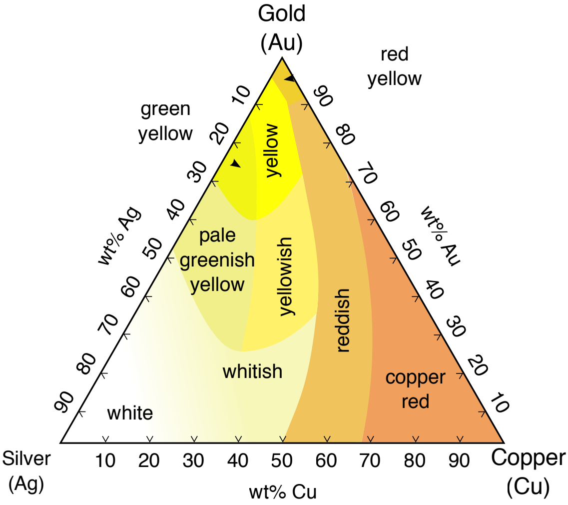




The simplex has 1 less dimension than the composition

The plot of the simplex for a 3-part composition is a *ternary plot*, also known as a triangle or barycentric plot

Aside on ternary plots



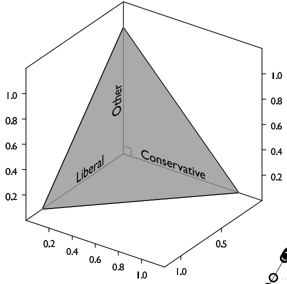
Source: Wikipedia

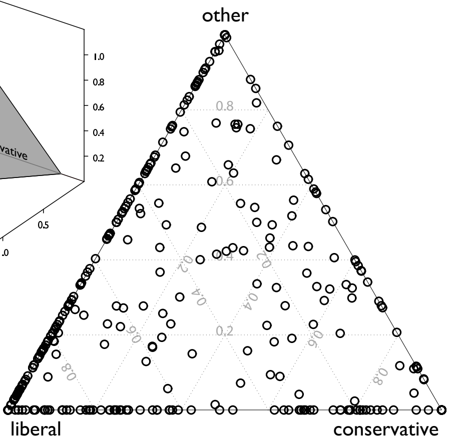
Ternary plots are most used in geology and metallurgy

This plot shows colors of alloys composed on Gold, Silver, and Copper

Key limitation: only works for 3-part compositions

...but you could make one category a "catch-all"





Note that many cases have one or more components at 0

This greatly complicates modeling:
most compositional data models assume all components are non-zero

Key predictor:
Partisanship of government (PCoG): higher values = more conservative

My claim:

$$\{\text{Cons}, \text{Lib}, \text{Oth}\} = f(\text{PCoG}, \text{controls})$$

I estimate a *zero-inflated compositional regression*...

Key predictor: Partisanship of government (PCoG): higher values = more conservative

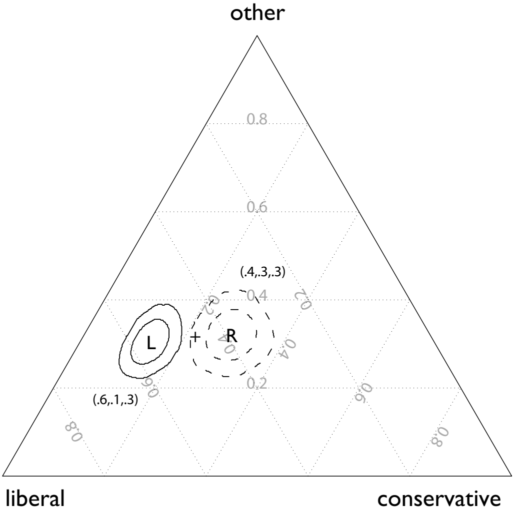
My claim:

$$\{\text{Cons}, \text{Lib}, \text{Oth}\} = f(\text{PCoG}, \text{controls})$$

I estimate a *zero-inflated compositional regression*...

Table 8.3. *Zeros-included compositional data analysis of central banker appointments.*

Response		Covariates	E(sign)	Model 1	Model 2	Model 3	Model 4
Model of non-zeroes	LibExp > 0	Constant		$1.142^{0.124}$	$1.142^{0.124}$	1.308^{a}	$-0.139^{0.389}$
		PCoG	$-$	$-1.487^{0.343}$	$-1.487^{0.342}$	$-0.691^{0.486}$	$-1.770^{0.371}$
		$\text{ConExp}_{\text{pre}}$					$1.520^{0.678}$
		$\text{LibExp}_{\text{pre}}$					$1.980^{0.596}$
	ConExp > 0	Constant		$-0.242^{0.101}$	$-0.242^{0.101}$	-0.440^{a}	$-1.050^{0.377}$
		PCoG	$+$	$0.171^{0.288}$	$0.170^{0.280}$	$0.648^{0.444}$	$-0.208^{0.322}$
		$\text{ConExp}_{\text{pre}}$					$2.350^{0.645}$
		$\text{LibExp}_{\text{pre}}$					$0.573^{0.539}$
	OthExp > 0	Constant		$0.482^{0.105}$	$0.482^{0.105}$	0.452^{a}	$1.710^{0.402}$
		PCoG		$-0.662^{0.302}$	$-0.662^{0.302}$	$-0.163^{0.434}$	$-0.461^{0.327}$
		$\text{ConExp}_{\text{pre}}$					$-1.960^{0.653}$
		$\text{LibExp}_{\text{pre}}$					$-1.500^{0.564}$
Model of composition	$\ln\left(\frac{\text{LibExp}}{\text{OthExp}}\right)$	Constant		$0.415^{0.128}$	$0.381^{0.124}$	0.497^{a}	$-0.443^{0.440}$
		PCoG	$-$	$-0.390^{0.419}$	$-0.252^{0.419}$	$-0.561^{0.482}$	$-0.147^{0.414}$
		$\text{ConExp}_{\text{pre}}$					$0.314^{0.775}$
		$\text{LibExp}_{\text{pre}}$					$1.470^{0.612}$
	$\ln\left(\frac{\text{ConExp}}{\text{OthExp}}\right)$	Constant		$-0.111^{0.160}$	$-0.112^{0.147}$	0.085^{a}	$-0.152^{0.498}$
		PCoG	$+$	$0.557^{0.471}$	$0.491^{0.446}$	$0.057^{0.495}$	$0.546^{0.445}$
		$\text{ConExp}_{\text{pre}}$					$0.071^{0.818}$
		$\text{LibExp}_{\text{pre}}$					$0.007^{0.722}$
		Est. t dfs			$7.779^{2.698}$	$4.730^{1.920}$	$6.900^{3.160}$
Composition Model				Normal	Student's t	Student's t	Student's t
Notes						a,b	b
N				411	411	411	391
ln likelihood				-1414.82	-1066.29	-962.84	-985.80
p-value of LR test				0.000	0.000	0.000	0.000
against model lacking				PCoG	t-dist	f.e.	prev exp



5 nonlinear coefficients aren't the quantity of interest – the expected career composition under partisan government is!

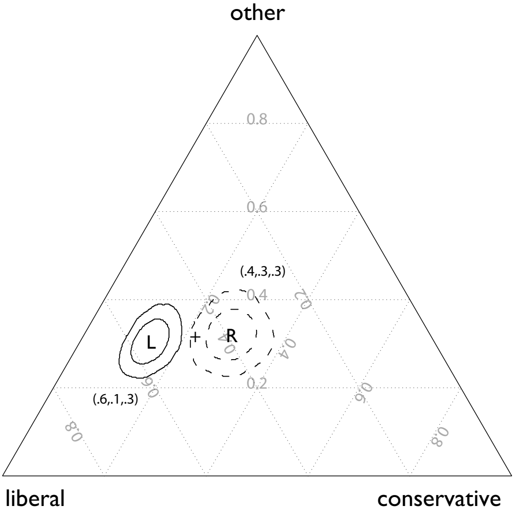
L Left Gov't (-1.5 sd)

R Right Gov't (+1.5 sd)

\+ Average Gov't

Simulation of these components from the model & a ternary plot = Clear results

1- and 2-se confidence regions are computed by kde2d



We find the expected relationship:

Right-wing govts prefer conservative career types

Left-wing govts prefer liberal career types

But do we trust this result? Might it change if we specified our model differently?

Robustness Checks

So far, we've presenting conditional expectations & differences from regressions

But are we confident that these were the "right" estimates?

The language of inference usually assumes we

- correctly specified our model
- correctly measured our variables
- chose the right probability model
- don't have influential outliers, etc.

Robustness Checks

So far, we've presenting conditional expectations & differences from regressions

But are we confident that these were the "right" estimates?

The language of inference usually assumes we

- correctly specified our model
- correctly measured our variables
- chose the right probability model
- don't have influential outliers, etc.

We're never completely sure these assumptions hold.

Most people present one model, and argue it was the best choice

Sometimes, a few alternatives are displayed

The race of the variables

	Model 1	Model 2	Model 3	Model 4	Model 5
My variable of interest, X_1	X.XX (X.XX)	X.XX (X.XX)	X.XX (X.XX)	X.XX (X.XX)	
A control I "need"	X.XX (X.XX)	X.XX (X.XX)	X.XX (X.XX)	X.XX (X.XX)	X.XX (X.XX)
A control I "need"	X.XX (X.XX)	X.XX (X.XX)	X.XX (X.XX)	X.XX (X.XX)	X.XX (X.XX)
A candidate control		X.XX (X.XX)		X.XX (X.XX)	
A candidate control			X.XX (X.XX)	X.XX (X.XX)	
Alternate measure of X_1					X.XX (X.XX)

Robustness Checks

Problems with the approach above?

1. Lots of space to show a few permutations of the model

   Most space wasted or devoted to ancillary info

Robustness Checks

Problems with the approach above?

1. Lots of space to show a few permutations of the model

   Most space wasted or devoted to ancillary info

2. What if we're really interested in $\mathbb{E}(Y|X)$, not $\hat{\beta}$?

   E.g., because of nonlinearities, interactions, scale differences, etc.

Robustness Checks

Problems with the approach above?

1. Lots of space to show a few permutations of the model

   Most space wasted or devoted to ancillary info

2. What if we're really interested in $\mathbb{E}(Y|X)$, not $\hat{\beta}$?

   E.g., because of nonlinearities, interactions, scale differences, etc.

3. The selection of permutations is *ad hoc*.

Robustness Checks

Problems with the approach above?

1. Lots of space to show a few permutations of the model

   Most space wasted or devoted to ancillary info

2. What if we're really interested in $\mathbb{E}(Y|X)$, not $\hat{\beta}$?

   E.g., because of nonlinearities, interactions, scale differences, etc.

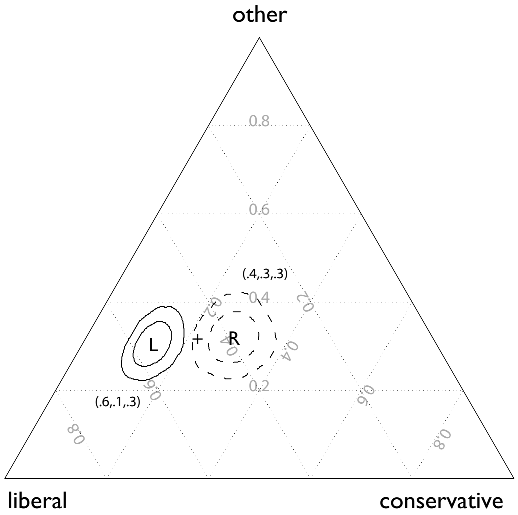
3. The selection of permutations is *ad hoc*.

We'll try to fix 1 & 2.

Objection 3 is harder, but worth thinking about.

Robustness Checks: An algorithm

1. Identify a relation of interest between a concept $\mathcal{X}$ and a concept $\mathcal{Y}$
2. Choose:
   - a measure of $\mathcal{X}$, denoted X,
   - a measure of $\mathcal{Y}$, denoted Y,
   - a set of confounders, Z,
   - a functional form, $g(\cdot)$
   - a probability model of Y, $f(\cdot)$
3. Estimate the probability model $Y \sim f(\mu, \alpha)$, $\mu = g(\mathrm{vec}(X, Z), \beta)$.
4. Simulate the quantity of interest such as $\mathbb{E}(Y|X)$, $\mathbb{E}(Y|X_1 - Y|X_2)$, or $\mathbb{E}(Y|X_1 \ / \ Y|X_2)$ to obtain a point estimate and confidence interval.
5. Repeat 2–4, changing at each iteration one of the choices in step 2.
6. Compile the results in a variant of the dot plot called a *ropeladder*.

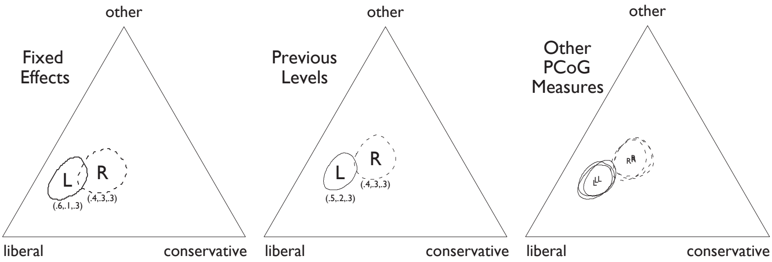


Earlier we reviewed a compositional data model from Ch. 8 of *Bankers, Bureaucrats, and Central Bank Politics*

We used a ternary plot to show the career composition of appointed central bankers depends on the partisanship of the appointing government

How would we show robustness under alternative specifications?

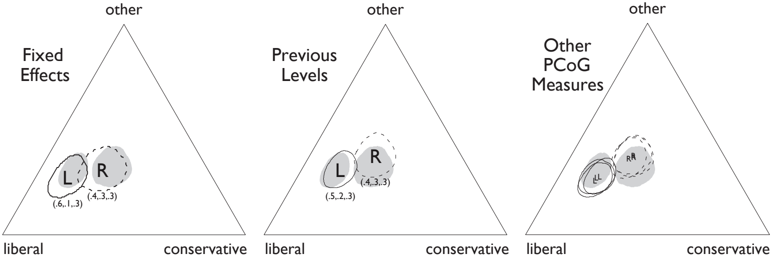
One tool for robustness: Small Multiples & Backup Dancers



Once people understand ternary plots,
they will immediately absorb a small, simplified version

Each of these small multiples shows our result under a different model

The similarity of each plot is immediately obvious here



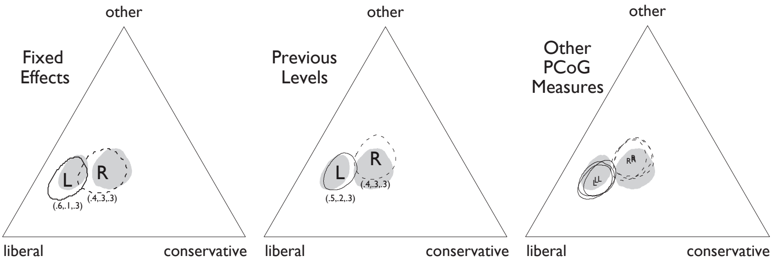
Once people understand ternary plots,
they will immediately absorb a small, simplified version

Each of these small multiples shows our result under a different model

The similarity of each plot is immediately obvious here

If not, putting the original plot in gray in the background helps:
Amanda Cox call these "backup dancers"

One tool for robustness: Small Multiples & Backup Dancers

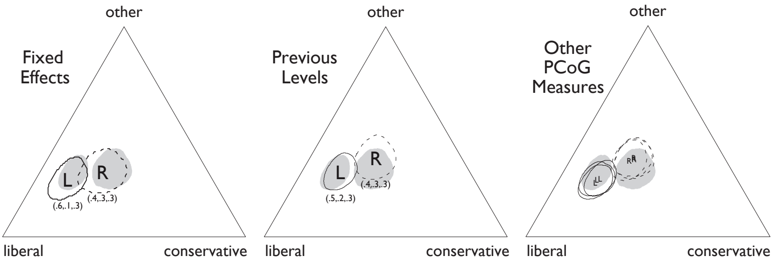


But in this case, I need *lots* of robustness checks

Because of the multiple equations, my statistical model is so demanding it's hard to include many regressors at once

If I try them one at a time, I would fill pages with triangle plots

One tool for robustness: Small Multiples & Backup Dancers



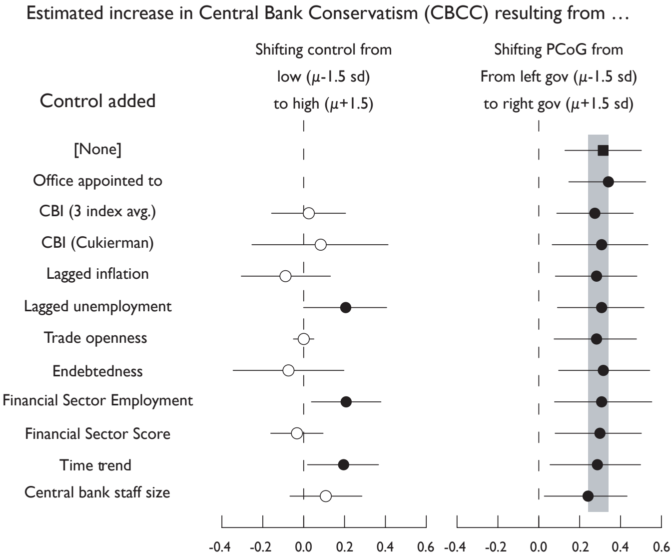
However, the horizontal dimension is the substantively important one: the one that affects affects economic outcomes

So I create a new QoI: Central Banker Career Conservatism (CBCC)

CBCC = Conservative Experience - Liberal Experience

And use my model to predict changes in CBCC and plot them on a *ropeladder*

Robustness Ropeladder: Partisan central banker appointment



Anatomy of a ropeladder plot

I call this a **ropeladder** plot.

The column of dots shows the relationship between y and a specific X under different model assumptions

Each entry corresponds to a different assumption about the specification, or the measures, or the estimation method, etc.

Anatomy of a ropeladder plot

I call this a **ropeladder** plot.

The column of dots shows the relationship between y and a specific X under different model assumptions

Each entry corresponds to a different assumption about the specification, or the measures, or the estimation method, etc.

If all the dots line up, with narrow, similar CIs, we say the finding is robust, and reflects the data under a range of reasonable assumptions

If the ropeladder is "blowing in the wind", we may be skeptical of the finding.
It depends on model assumptions that may be controversial

Anatomy of a ropeladder plot

I call this a **ropeladder** plot.

The column of dots shows the relationship between y and a specific X under different model assumptions

Each entry corresponds to a different assumption about the specification, or the measures, or the estimation method, etc.

If all the dots line up, with narrow, similar CIs, we say the finding is robust, and reflects the data under a range of reasonable assumptions

If the ropeladder is "blowing in the wind", we may be skeptical of the finding.
It depends on model assumptions that may be controversial

The shaded gray box shows the full range of the point estimates for the QoI.

Narrow is better.

Why ropeladders?

1. Anticipate objections on model assumptions, and have concrete answers.

   Avoid: "I ran it that other way, and it came out the 'same'."

   Instead: "I ran it that other way, and look – it made no *substantive* or *statistical* difference worth speaking of."

   Or: "...it makes *this* much difference."

Why ropeladders?

1. Anticipate objections on model assumptions, and have concrete answers.

   Avoid: "I ran it that other way, and it came out the 'same'."

   Instead: "I ran it that other way, and look – it made no *substantive* or *statistical* difference worth speaking of."

   Or: "...it makes *this* much difference."

2. Investigate robustness more thoroughly.

   Traditional tabular presentation would have run to 7 pages, making comparison hard and discouraging a thorough search

Why ropeladders?

1. Anticipate objections on model assumptions, and have concrete answers.

   Avoid: "I ran it that other way, and it came out the 'same'."

   Instead: "I ran it that other way, and look – it made no *substantive* or *statistical* difference worth speaking of."

   Or: "…it makes *this* much difference."

2. Investigate robustness more thoroughly.

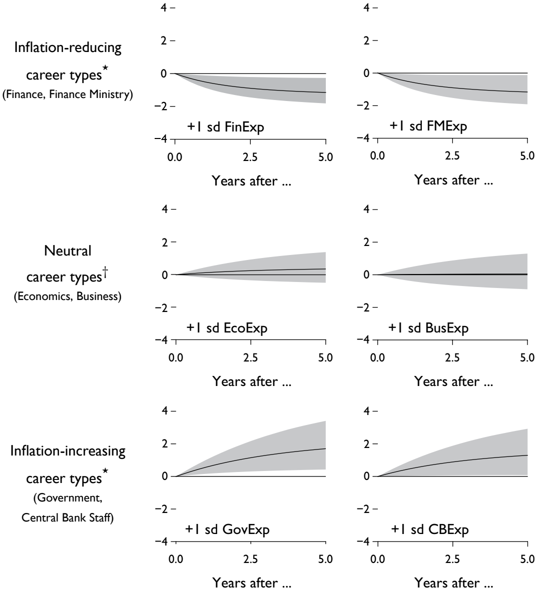
   Traditional tabular presentation would have run to 7 pages, making comparison hard and discouraging a thorough search

3. Find patterns of model sensitivity.

   Two seemingly unrelated changes in specification had the same effect. (Unemployment and Financial Sector Size)

   Turned out to be a missing third covariate. (Time trend)

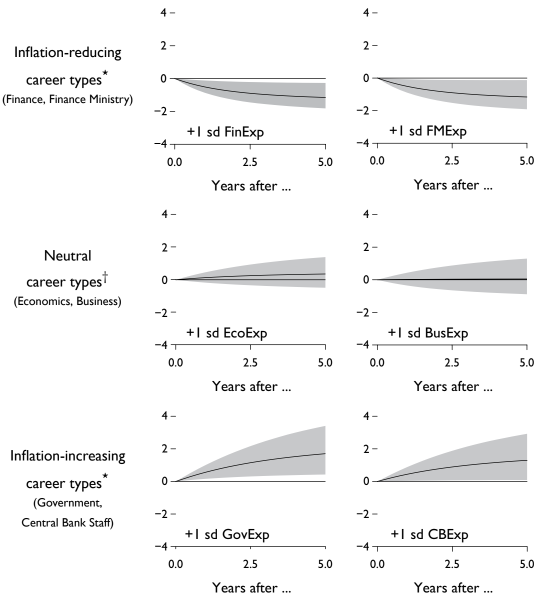
Change in inflation, over time, from changing career composition of the central bank



Recall the TSCS model of inflation performance

How would we show robustness here?

Change in inflation, over time, from changing career composition of the central bank



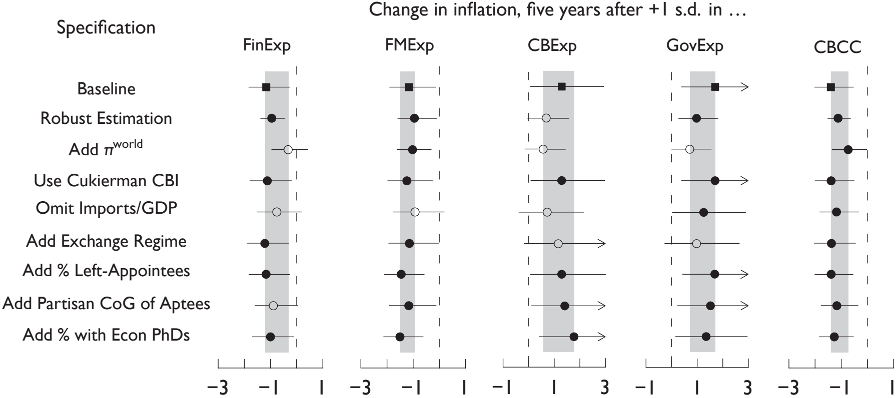
Recall the TSCS model of inflation performance

How would we show robustness here?

Once we understand the dynamics over time, we can simplify our presentation

What if we isolate the 5 year mark, and compare the estimated effects of covariate on inflation at that point under different models?

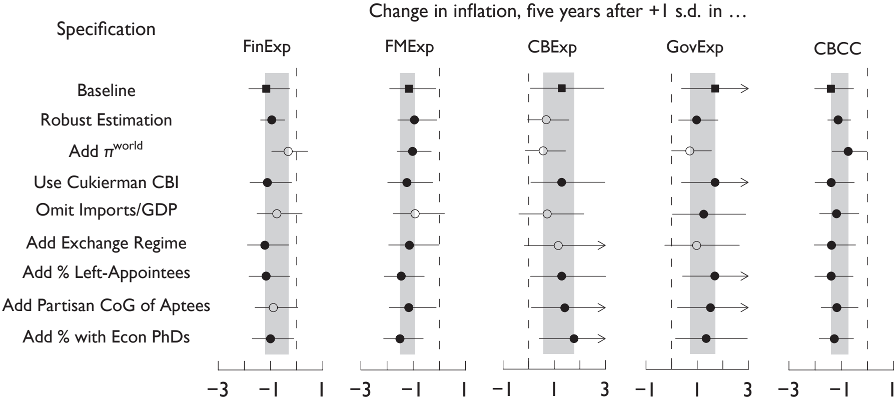
Robustness for several QoIs at once



Each ropeladder, or column, shows the effect of a different variable on the response

That is, reading *across* shows the results from a single model

Reading *down* shows the results for a single question across different models

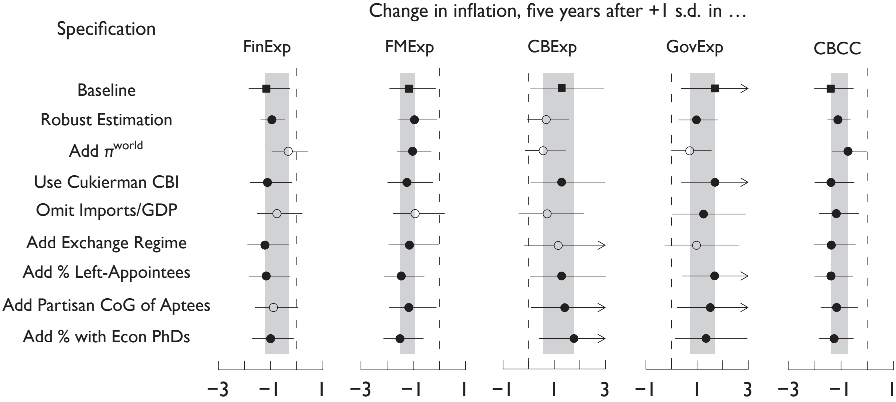


Arrows indicate confidence intervals that extend outside the plot

Choosing our own plotting area using `limits=` is critical for ropeladders

Focus on the area with the point estimates and on any problematic CIs

Robustness for several QoIs at once



To write up robustness, show this graphic and relegate tables to the appendix

You can be specific about the nature of robustness (no “hand-waving”),

yet still write up 8 robustness checks on 5 covariates in 2 pages total

Our tools are flexible

From the first few examples,
you might think lineplots are for model inference
and dotplots (ropeladders) are for model robustness

But these tools are *flexible* and reward creativity

In the following examples,
I use dotplots made with `ropeladder()` to explore models,
then use lineplots to explore robustness

When simulation is the only option: Chinese leadership

Shih, Adolph, and Liu investigate the advancement of elite Chinese leaders in the Reform Period (1982–2002)

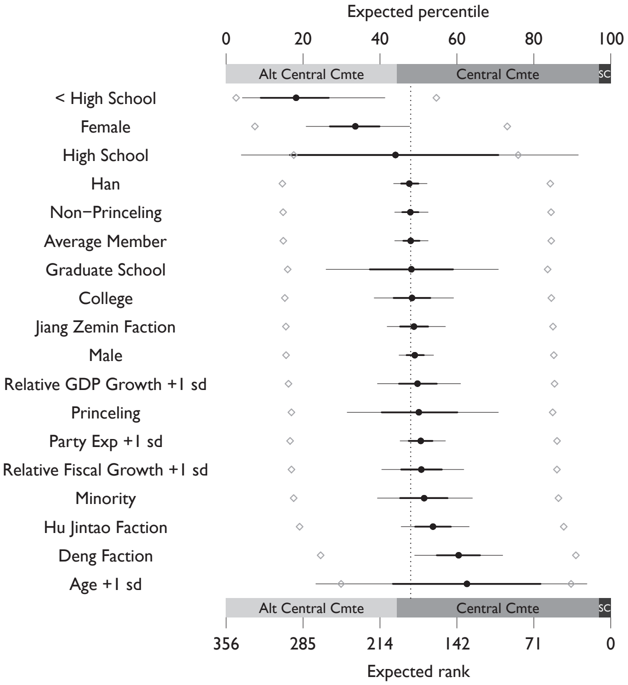
Explain (partially observed) ranks of the top 300 to 500 Chinese Communist Party leaders as a function of:

Demographics	*age, sex, ethnicity*
Education	*level of degree*
Performance	*provincial growth, revenue*
Faction	*birth, school, career, and family ties to top leaders*

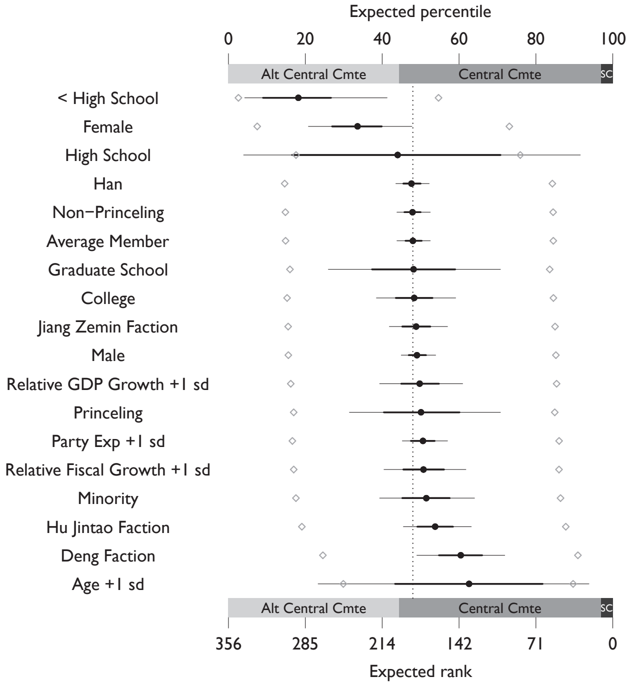
Bayesian model of partially observed ranks of CCP officials

Model parameters difficult to interpret: on a latent scale
and individual effects are conditioned on all other ranked members

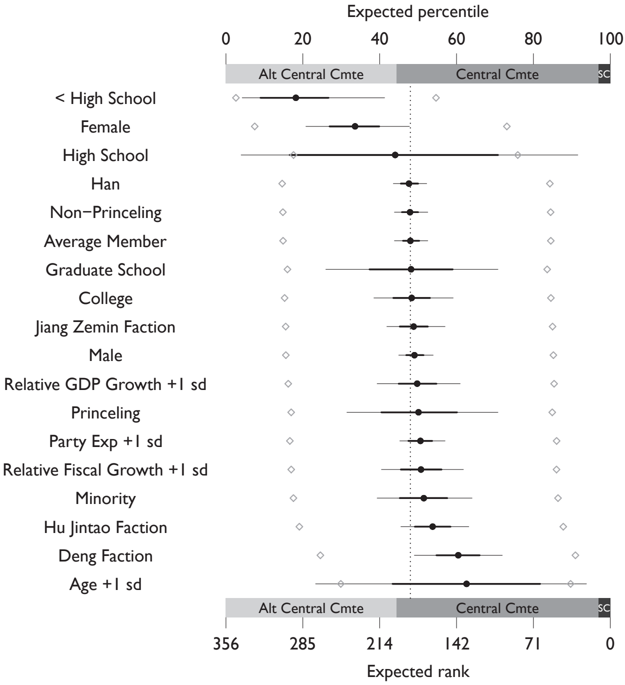
Only solution:
Simulate ranks of hypothetical officials as if placed in the observed hierarchy



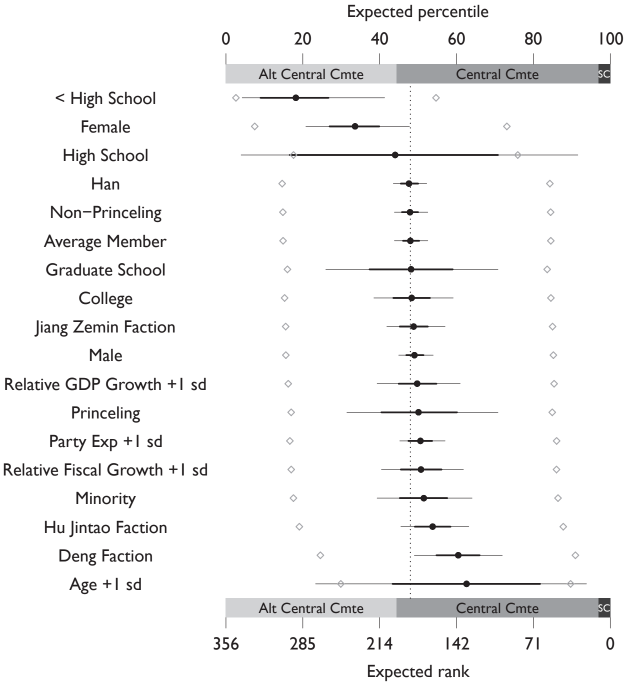
Black circles show expected ranks for otherwise average Chinese officials with the characteristic listed at left



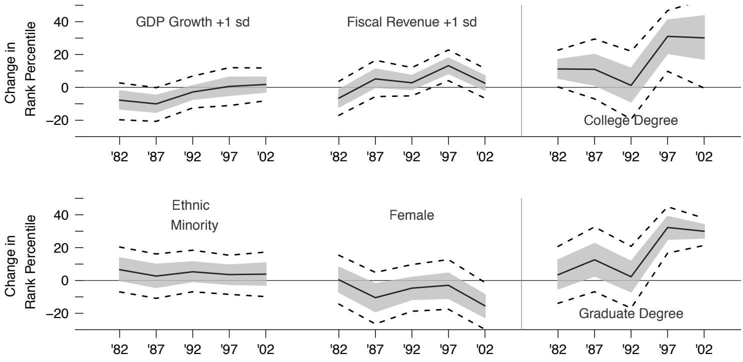
Thick black horizontal lines are 1 std error bars, and thin lines are 95% CIs



Gray triangles are officials with random effects at ± 1 sd; how much unmeasured factors matter

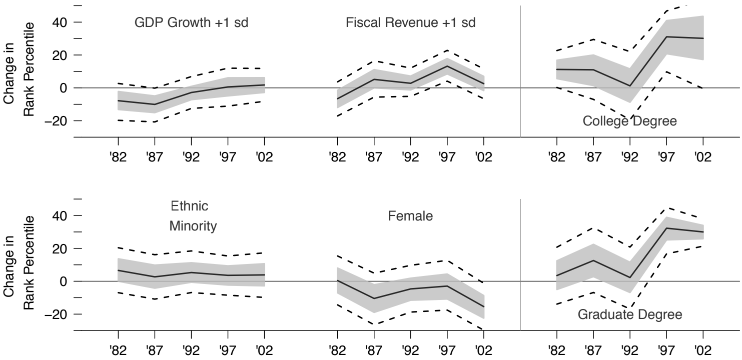


It helps to sort rows of the plot from smallest to largest effect (diagonalization)



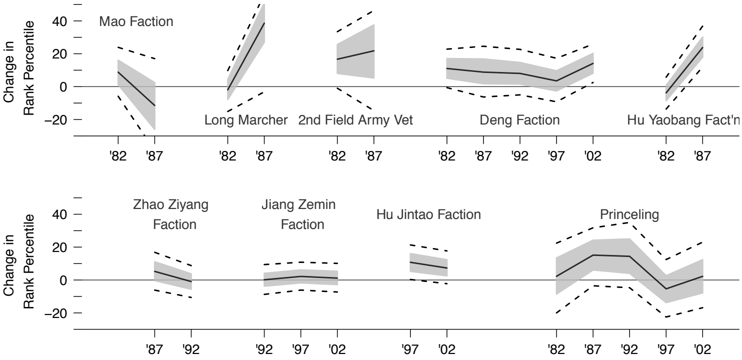
We re-estimate the model *separately* for each year,
leading to a large number of results with varying sets of covariates

A complex lineplot helps organize these results and facilitate comparisons



Note that these results are now *first differences*:

the expected percentile change in rank for an otherwise average official who gains the characteristic noted



Over time, officials' economic performance never matters, but factions often do

Runs counter to the conventional wisdom that meritocratic selection of officials lies behind Chinese economic success

Chinese Officials Robustness

Our findings were controversial: countered the widely accepted belief that Chinese officials are rewarded for economic performance

Critics asked for lots of alternative specifications to probe our results

We used `tile` to show exactly what difference these robustness checks made using overlapping lineplots

We provide detailed one-to-one comparisons of our model with each alternative, for a lengthy appendix...

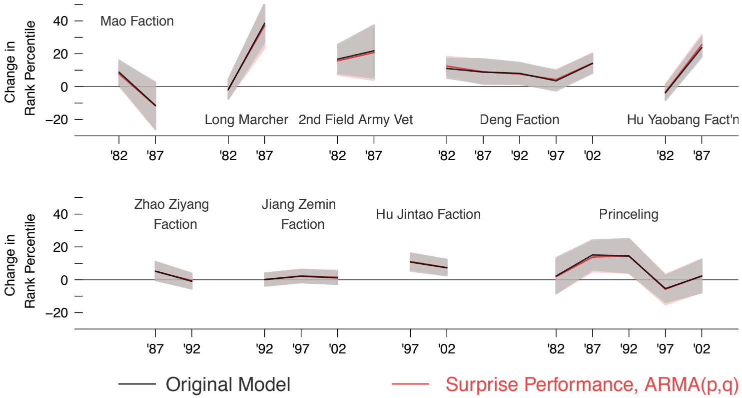
...And a single page summary for the printed article collecting all robustness checks

Some critics worried that our measures of faction were too sensitive, so we considered a more specific alternative

This didn't salvage the conventional wisdom on growth...

But did (unsuprisingly) strengthen our factional results

(Specific measures pick up the strongest ties)

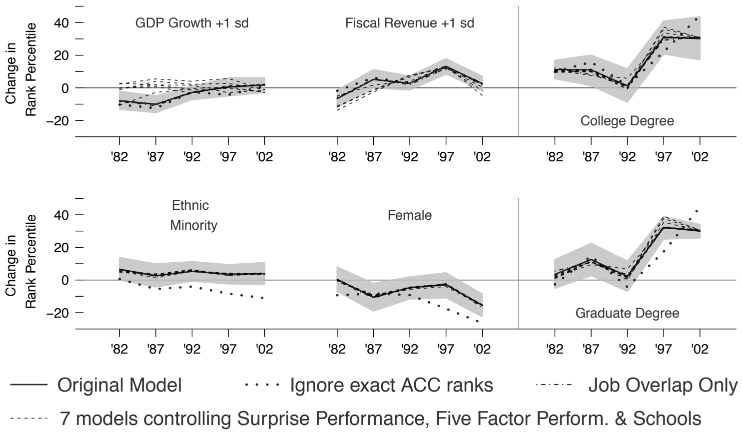


Other critics worried about endogeneity or selection effects flowing from political power to economic performance

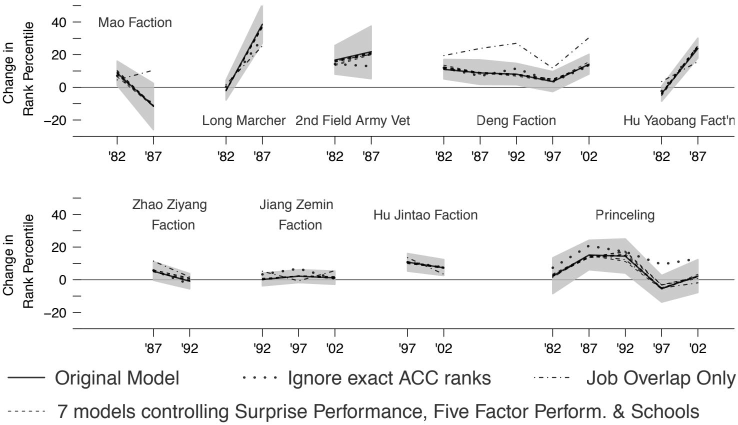
We used measures of unexpected growth to zero in on an official's own performance in office – which still nets zero political benefit

The above summarizes results combined from 2 versions of a model applied over 5 periods, each with 5 multiply imputed datasets (50 models)

But it still takes many pages to show all our robustness checks.
Is there a more efficient way to show that our results stay essentially the same?



In our printed article, we show only this plot,
which overlays the full array of robustness checks



Conveys hundreds of separate findings in a compact, readable form

No knowledge of Bayesian methods or partial rank coefficients required!

More uses of ropeladder traces for model exploration

We will discuss implementation of ropeladders – for robustness and general model inference – in Session 5

But first, let's explore three more uses of ropeladder dotplots that show off the full range of features of these traces:

- Exploring interactive models using differences-in-differences
- Grouping variables and interactions for easier comprehension and explanation
- Grouping categorical responses to multinomial models

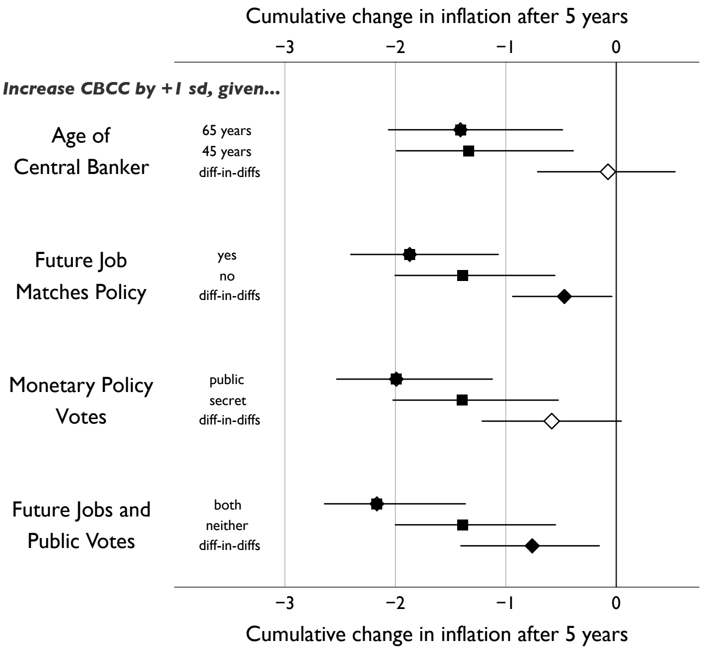
Remember, ropeladders are flexible – surely the *most* flexible way to present models

Be willing to experiment to make your model easier to explain

Recall our comparative inflation example

Central banker careers $\rightarrow$ Inflation performance

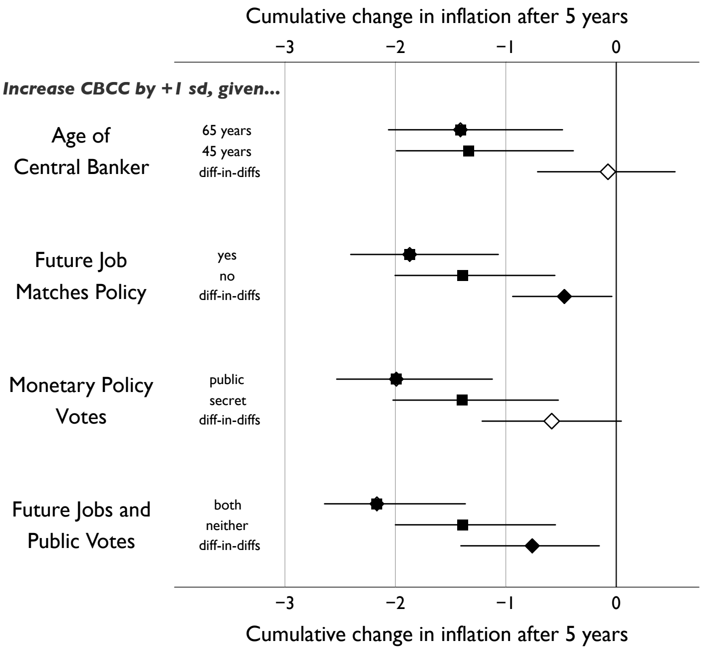
But is this a result of socialization or incentives?



Recall our comparative inflation example

Central banker careers → Inflation performance

But is this a result of socialization or incentives?



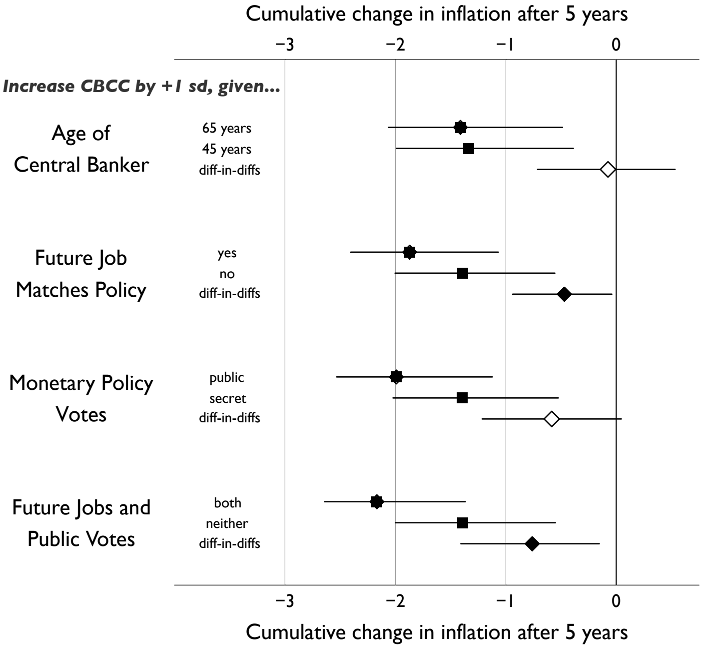
Interactions may reveal the sources of career effects

I interact "conservative" careers with:

Age

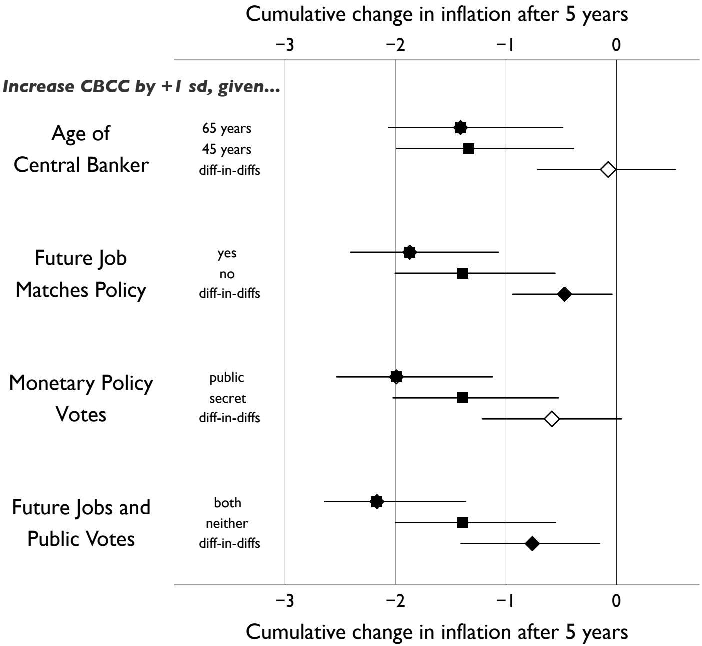
Future Job Awarded

Public Voting



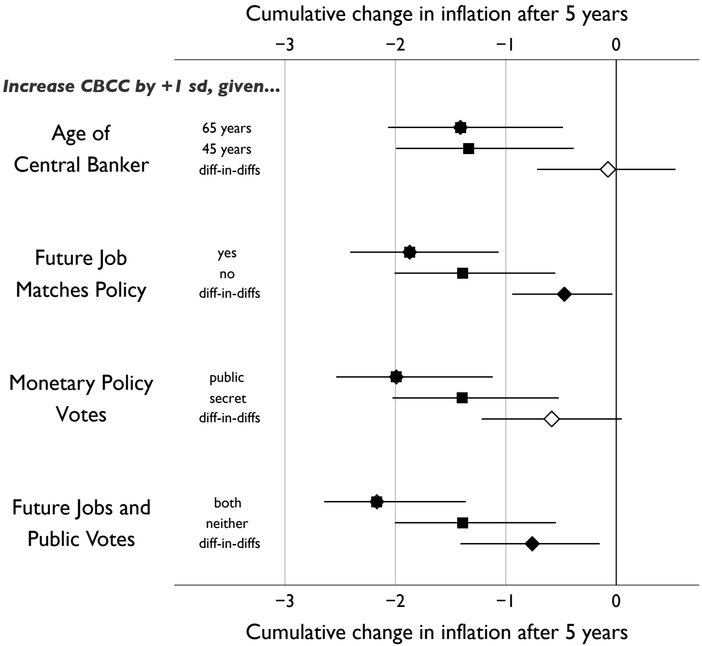
Focus on the "Age of Central Banker" trio of results

This model interacts Career Conservatism (CBCC) with central banker age



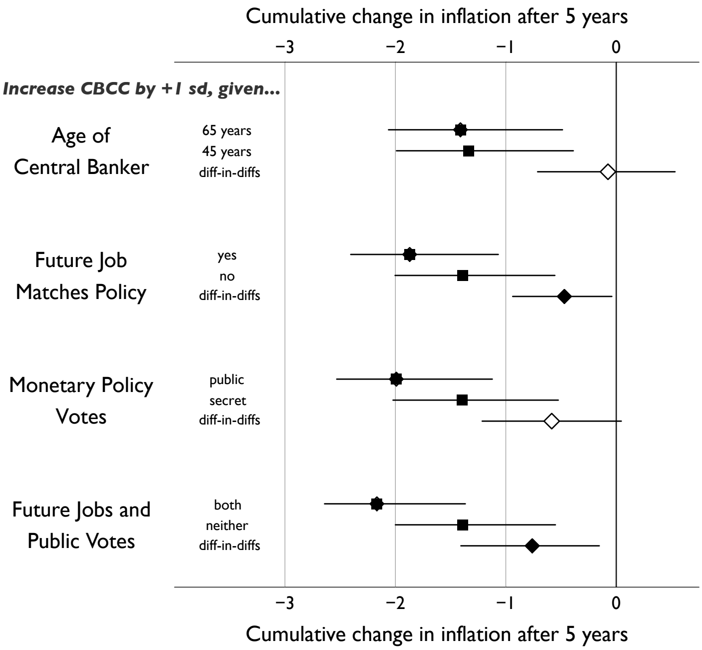
We simulate the effect of +1 sd CBCC given either 65 year old officials or 45 year old officials

We are especially interested in the difference of the first differences across these scenarios

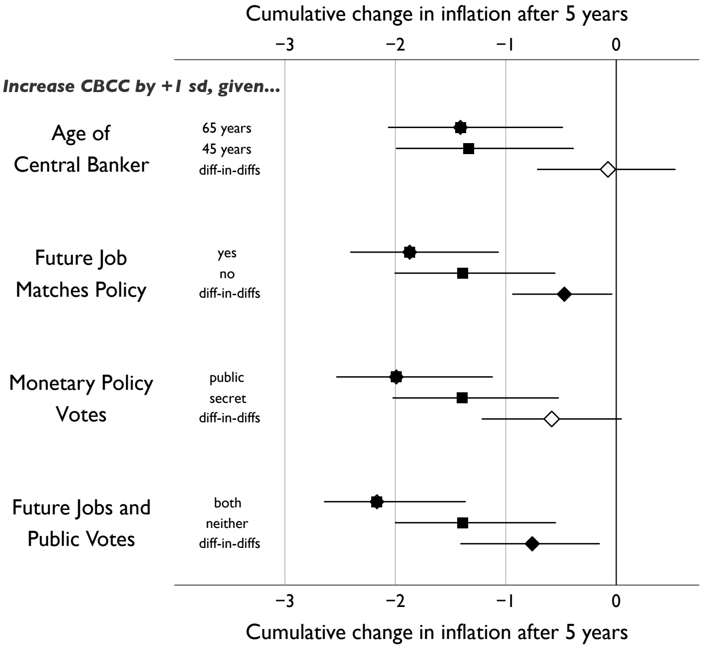


We use the shape of symbols to suggest the "building up" of the full effect for 65 year olds

While open vs. filled indicates significance

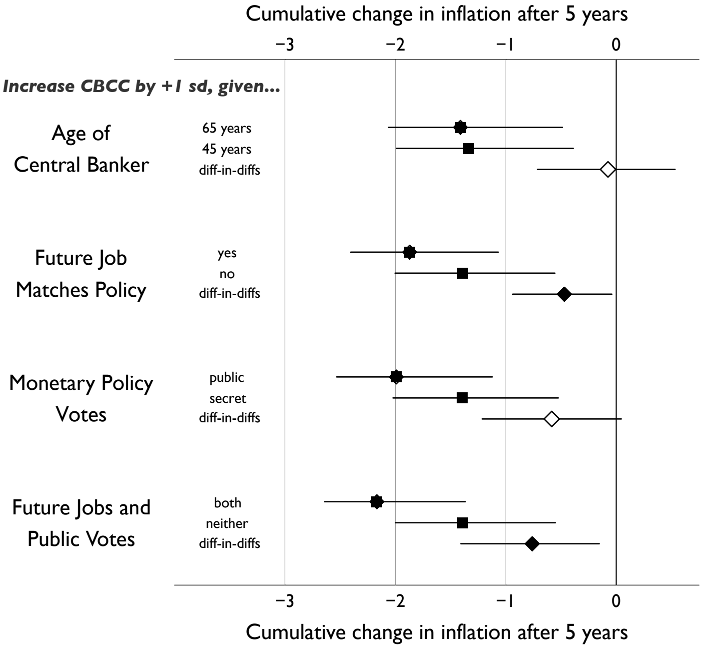


Looking at the whole plot, we find conservatism has bigger inflation-fighting effects when central bankers end up taking jobs in the financial sector

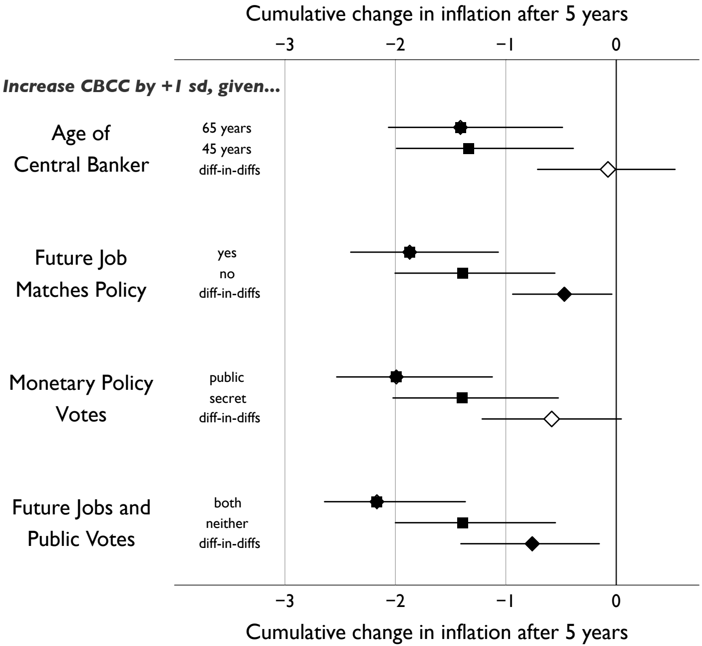


Conservatism may be stronger when votes are public, but this is not quite significant

Age is a wash



Ropeladders can explore interactive effects by working through each combination of values for the interacted covariates



2 keys to success:

Find the minimum number of combinations needed to tell the story

Clear organization – designed *around* the clearest write-up of the result

Chapter 9 of *BBC* explores correlates of central banker tenure in 20 industrialized countries using a Cox proportional hazards model

Covariate

- Age
- Career types
- Economic performance
- Change in government
- Performance $\times$ Party

Last is most interesting: *are central bankers graded on a partisan curve, with the Left penalizing unemployment and the Right inflation?*

Chapter 9 of *BBC* explores correlates of central banker tenure in 20 industrialized countries using a Cox proportional hazards model

Covariate

Age
Career types
Economic performance
Change in government
Performance $\times$ Party

Last is most interesting: *are central bankers graded on a partisan curve, with the Left penalizing unemployment and the Right inflation?*

Table 9.1. *Cox proportional hazards estimates of central banker tenure.*

Covariate	Hazard ratio	95% CI lower	95% CI upper
Age > 75	5.78	2.28	14.68
$70 <$ Age ≤ 75	3.48	2.32	5.22
$65 <$ Age ≤ 70	2.01	1.24	3.27
Other Government Experience	1.86	0.82	4.23
Abs diff in PCoG, appt party vs. current	1.67	1.24	2.25
Financial Experience	1.40	0.83	2.38
Finance Ministry Experience	1.34	0.71	2.52
Current PCoG $\times$ Inflation	1.05	1.00	1.11
Unemployment	1.04	1.00	1.08
Inflation	1.04	1.01	1.07
Current PCoG $\times$ Unemployment	0.95	0.89	1.02
Central Bank Staff Experience	0.90	0.62	1.30
Economics Experience	0.87	0.52	1.43
Current Partisan Center of Gravity (PCoG)	0.86	0.41	1.82
N	10,863	349 individuals	
log likelihood	-1229.4	LR test $p < 10^{-9}$	

Entries are hazard ratios (exponentiated coefficients) and their associated 95 percent confidence intervals. Hazard ratios greater than one indicate factors making retirement/dismissal *more* likely. Confidence intervals are calculated using standard errors clustered by country; significant results are those with lower and upper bounds on the same side of 1.00.

The table is actually fairly interpretable, *except*:

The career covariates are compositional, so their effects are blended

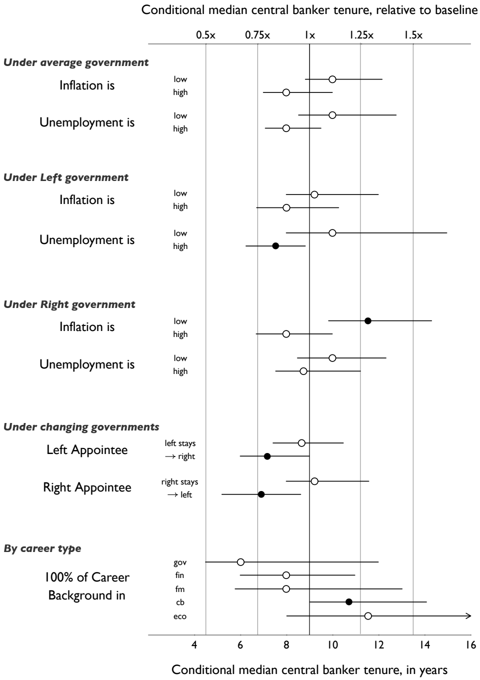
The interaction terms are hard to mentally combine, and it's impossible to get CIs without a computer to help

...so maybe this isn't *that* interpretable

Table 9.1. *Cox proportional hazards estimates of central banker tenure.*

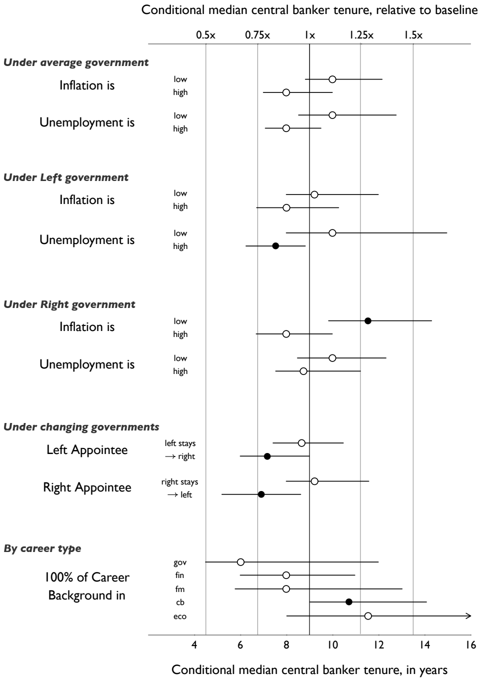
Covariate	Hazard ratio	95% CI lower	95% CI upper
Age > 75	5.78	2.28	14.68
$70 <$ Age ≤ 75	3.48	2.32	5.22
$65 <$ Age ≤ 70	2.01	1.24	3.27
Other Government Experience	1.86	0.82	4.23
Abs diff in PCoG, appt party vs. current	1.67	1.24	2.25
Financial Experience	1.40	0.83	2.38
Finance Ministry Experience	1.34	0.71	2.52
Current PCoG $\times$ Inflation	1.05	1.00	1.11
Unemployment	1.04	1.00	1.08
Inflation	1.04	1.01	1.07
Current PCoG $\times$ Unemployment	0.95	0.89	1.02
Central Bank Staff Experience	0.90	0.62	1.30
Economics Experience	0.87	0.52	1.43
Current Partisan Center of Gravity (PCoG)	0.86	0.41	1.82
N	10,863	349 individuals	
log likelihood	-1229.4	LR test $p < 10^{-9}$	

Entries are hazard ratios (exponentiated coefficients) and their associated 95 percent confidence intervals. Hazard ratios greater than one indicate factors making retirement/dismissal *more* likely. Confidence intervals are calculated using standard errors clustered by country; significant results are those with lower and upper bounds on the same side of 1.00.



We can replace the entire table with a complex dotplot

(Aside: It’s okay to provide handouts of really large plots – they don’t display on LCD projectors well)



Instead of thinking, "What covariates do I plot," ask:

"What is the minimum set of scenarios that will explore the full model space"

The key is picking out counterfactuals that explore effects of both inflation and unemployment under each type of government *and* under each possible change in government

Bonus Example: Allocation of Authority for Health Policy

Adolph, Greer & Fonseca consider explanations of whether local, regional, or state-level European governments have power over specific health policy areas and instruments

Areas: Pharamceuticals, Secondary/Tertiary, Primary Care, Public Health

Instruments: Frameworks, Finance, Implementation, Provision

Each combination for each country is a case

Bonus Example: Allocation of Authority for Health Policy

Adolph, Greer & Fonseca consider explanations of whether local, regional, or state-level European governments have power over specific health policy areas and instruments

Areas: Pharamceuticals, Secondary/Tertiary, Primary Care, Public Health

Instruments: Frameworks, Finance, Implementation, Provision

Each combination for each country is a case

Fiscal federalism suggests lower levels for information-intensive policies and higher levels for policies with spillovers or public goods

Also control for country characteristics and country random effects

With 3 nominal outcomes for each case, need a multilevel multinomial logit

Bonus Example: Allocation of Authority for Health Policy

Covariates:

Policy area	Nominal
Policy instrument	Nominal
Regions old or new	Binary
Country size	Continuous
Number of regions	Continuous
Mountains	Countinuous
Ethnic heterogeneity	Continuous

Tricky part to the model:

some cases have structural zeros for regions (when they don't exist!)

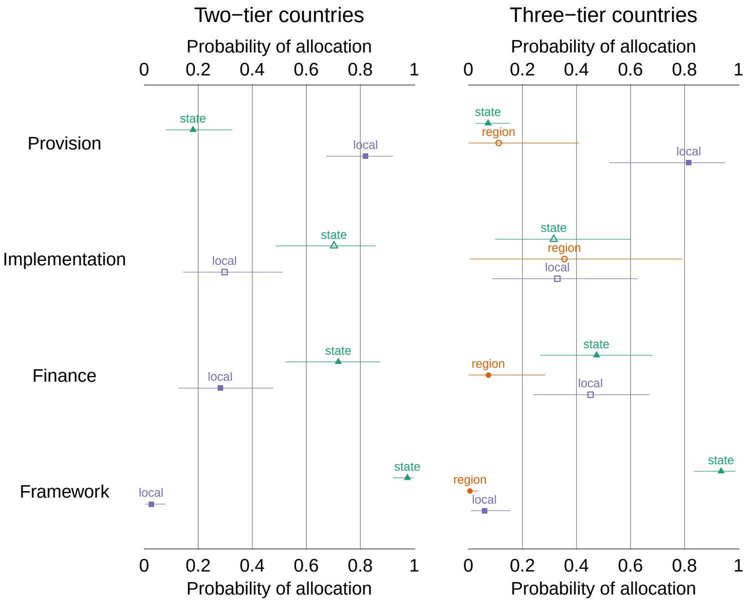
How to set up counterfactuals?

We could set all but one covariate to the mean, then predict the probability of each level of authority given varied levels of the remaining covariate

We should do this separately for countries with and without regional governments

Let's fix everthing but policy instrument to the mean values, then simulate the probability of authority at each level for each instrument

We show the results using a "nested" dot plot, made using `ropeladder()` in the `tile` package



Two−tier countries
Three−tier countries
Probability of allocation
0
0.2
0.4
0.6
0.8
1
Provision
Implementation
Finance
Framework
state
region
local

Special plots for compositional data

Probabilities have a special property: they sum to one

Variables that sum to a constraint are compositional

We can plot a two-part composition on a line,

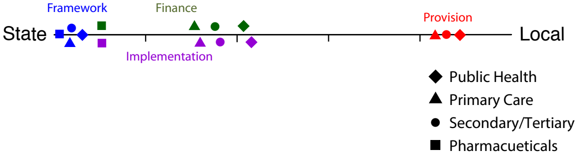
and a three-part composition on a triangular plot

This makes it easier to show more complex counterfactuals, such as every combination of policy area and instrument

But we also need to work harder to explain these plots

Probability of allocation of authority by policy type

Let's start with countries that have no regions, just local and state levels:

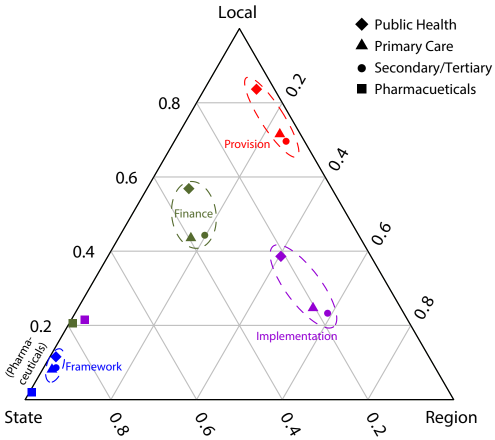


Above holds country characteristics at their means

And predicts the probability of state or local control

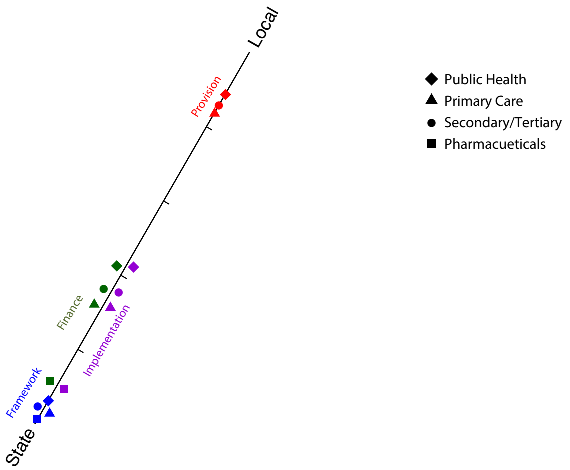
Because of the compositional constraint, these always sum to 1:
if the probability of local is p, the probability of state is $1 - p$

Probability of allocation of authority by policy type



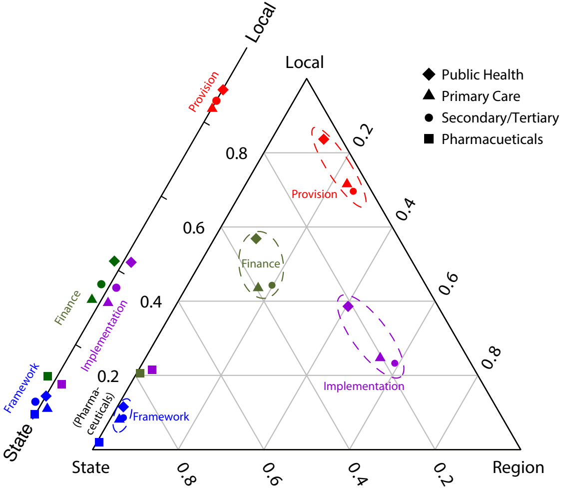
Next consider countries with all three levels, holding covariates at means

Probability of allocation of authority by policy type



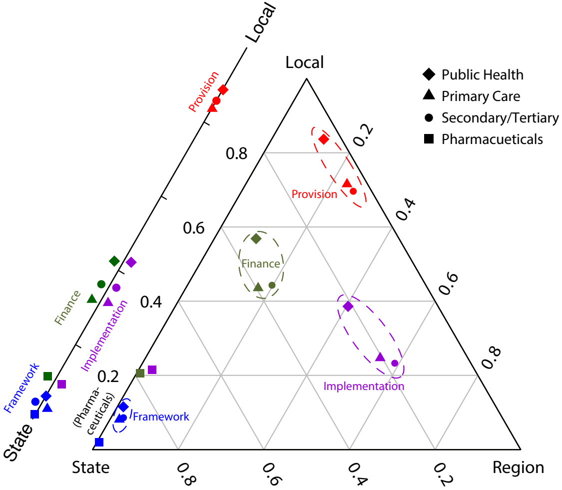
If we are clever...

Probability of allocation of authority by policy type



We can display everything in a single 2D figure

Probability of allocation of authority by policy type

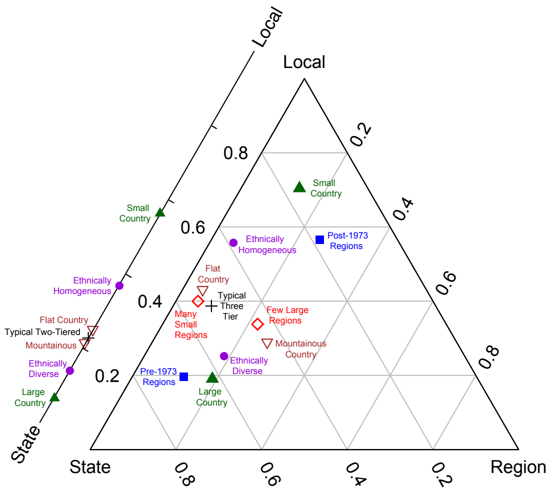


Note the 2-level “line” looks like a projection of the 3-level triangle down to 1D

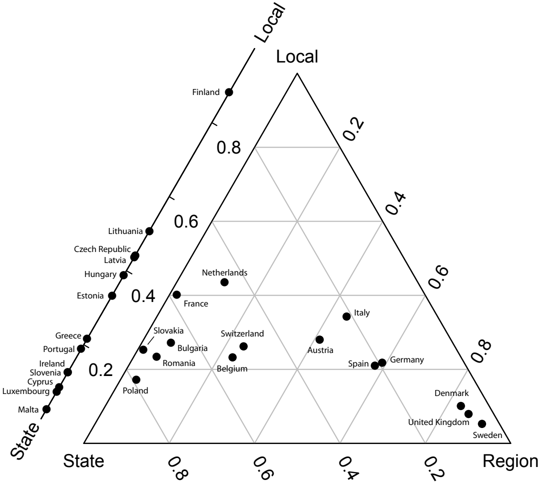
Probability of allocation of authority by country type

We can use the same framework to illustrate other implications of the model

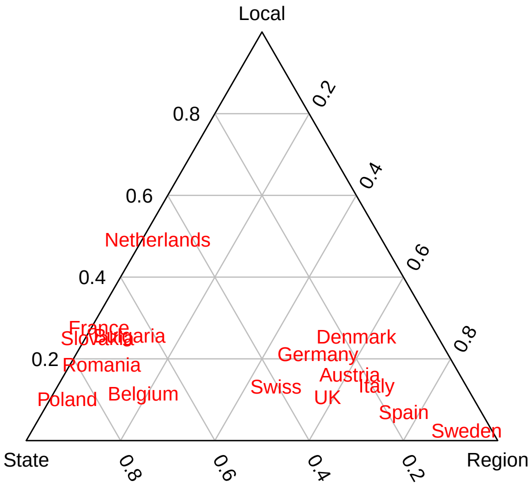
Probability of allocation of authority by country type



Above holds policy area and instrument "at their means"



Looking at the country random effects might suggest omitted variables

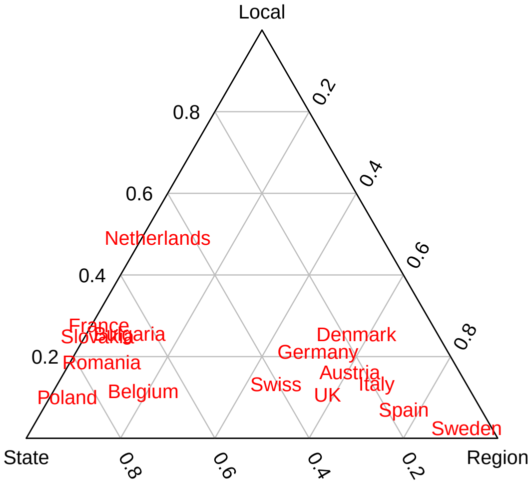


This is the random effects plot from an earlier iteration of the model

(Notice I often don't beautify plots until they are "final")

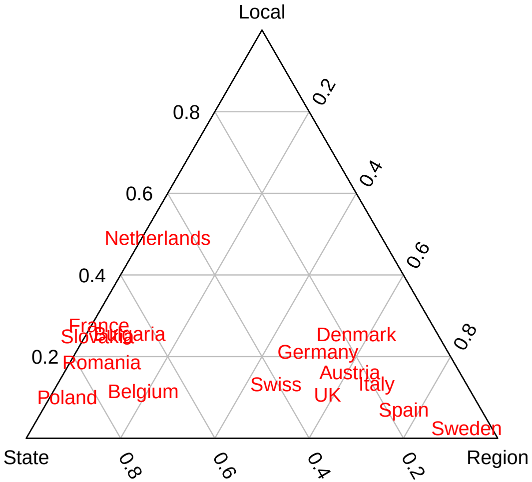
Some residuals appear "clustered" towards regional devolution

What missing variable does this plot suggest?



What missing variable does this plot suggest?

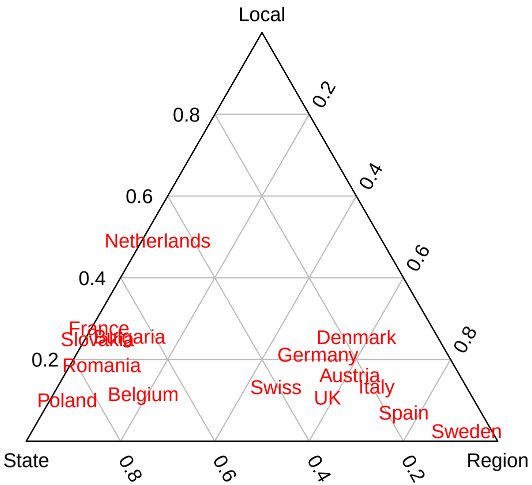
Do the clustered countries have something in common?



What missing variable does this plot suggest?

Do the clustered countries have something in common?

I couldn't see it



What missing variable does this plot suggest?

Do the clustered countries have something in common?

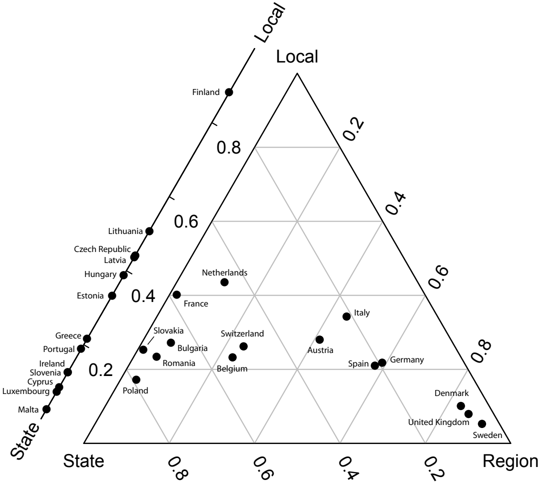
I couldn't see it

My wife did: good skiing

And historically mountains strengthened regional autonomy!

Lesson: Share your diagnostic plots!

Residual country effects



After controlling for mountainousness, no clusters of similar countries remain

Plan of Session 5: Tools for Visualizing Model Robustness

1. Ropeladder robustness examples using U.S. Crime data
2. Visualizing models with interaction terms
3. Workshop on attendee research

Ropeladder Robustness Example: 1960 US crime rates

My apologies – this example isn't particularly substantively interesting or sharp

We have data from each of the 50 US states on crime rates in 1960

And a variety of covariates as seen on the next slide

We will fit a set of models with the same specification but different estimators

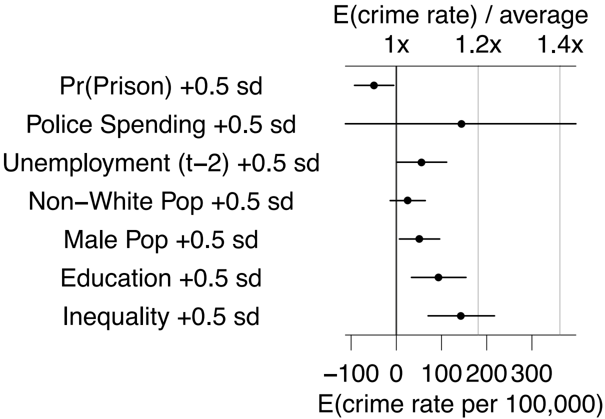
We will then consider several ropeladder-based presentations of robustness

Kitchen sink models of 1960 US crime rates

	Linear	Robust	Poisson	Neg Bin
Constant	-28820.91	-17784.56	-19.08	-15.43
	(10199.82)	(8158.71)	(1.77)	(7.81)
% males aged 14–24	1156.49	2480.55	1.1	1.53
	(522.98)	(418.32)	(0.1)	(0.4)
Southern state	0.97	138.11	0.06	0.06
	(141.49)	(113.18)	(0.02)	(0.11)
Mean education (yrs)	1802.64	1413.62	1.84	2.11
	(590.84)	(472.61)	(0.11)	(0.45)
Police spending 1960	897.54	422.45	0.81	0.76
	(813.8)	(650.95)	(0.15)	(0.62)
Police spending 1959	6.66	651.14	0.01	0.01
	(823.35)	(658.59)	(0.15)	(0.63)
Labor participation	143.91	2235.29	0.63	0.62
	(727.79)	(582.15)	(0.13)	(0.56)
Males per 1000	94.71	-3469.7	-1.46	-2.3
	(1943.8)	(1554.82)	(0.36)	(1.49)
State population	-79.39	-138.58	-0.08	-0.07

	continued			
	Linear	Robust	Poisson	Neg Bin
Nonwhites per 1000	61.25	32.47	0.11	0.11
	(47.85)	(38.28)	(0.01)	(0.04)
Unem, males 14–24	-325.65	-444.95	-0.18	-0.18
	(336.46)	(269.13)	(0.06)	(0.26)
Unem, males 35–39	475.14	895.28	0.39	0.46
	(239.62)	(191.67)	(0.04)	(0.18)
Gross state product, pc	282.31	-196.44	0.69	0.64
	(420.2)	(336.11)	(0.08)	(0.32)
Income inequality	1461.68	943.27	1.68	1.56
	(386.64)	(309.27)	(0.07)	(0.3)
Pr(imprisonment)	-226.39	-443.28	-0.29	-0.31
	(103.39)	(82.7)	(0.02)	(0.08)
E(time in prison)	-69.91	-294.41	-0.16	-0.27
	(184.13)	(147.29)	(0.03)	(0.14)

Ropeladder Robustness Example: US Crime

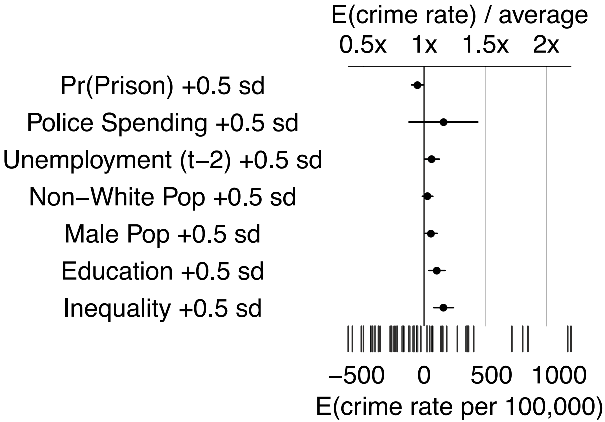


A simple inference dotplot with an extra axis showing relative risk

CODE EXAMPLE

`crimeRopeladders.r`

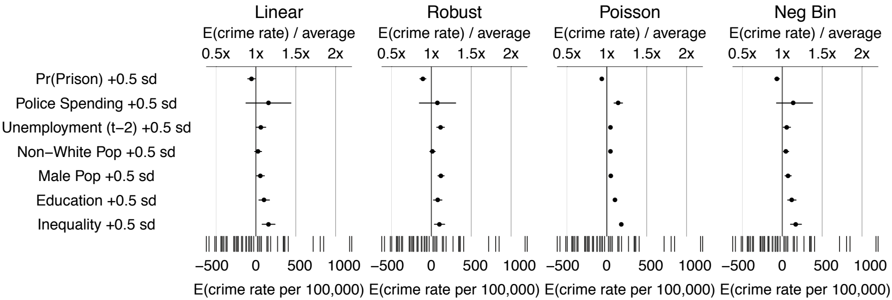
Ropeladder Robustness Example: US Crime



An inference dotplot with a marginal plot of the data

The data vary more widely than the first differences, stretching the plot

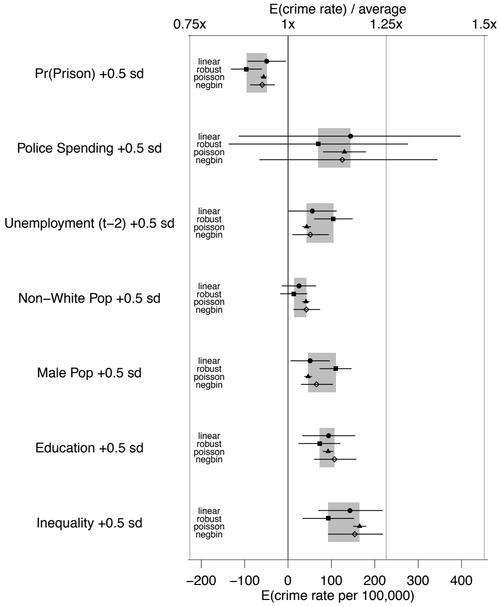
Ropeladder Robustness Example: US Crime



Side-by-side inference dotplots

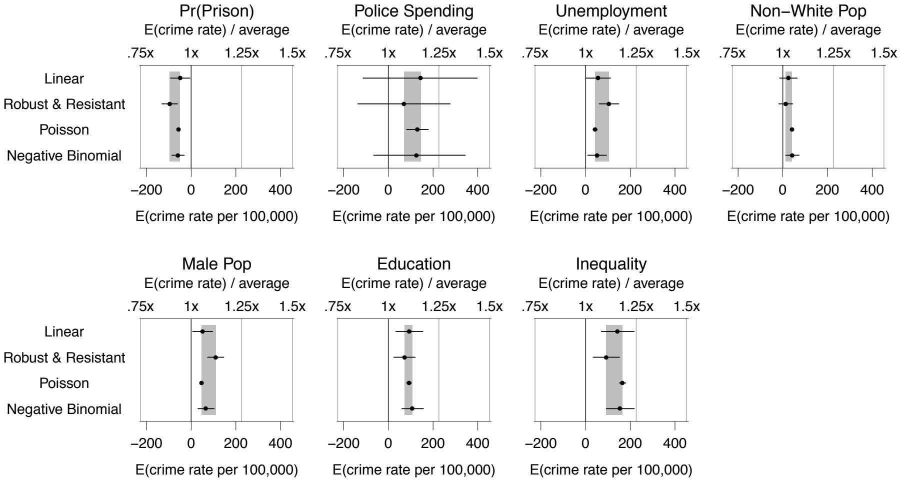
The focus here is comparisons across covariates within models

Ropeladder Robustness Example: US Crime



A superplot of ropeladders

Equal focus on comparisons across covariates and across models



Side-by-side robustness ropeladders –

focus is now on comparisons across models, not variables

How Do I Visualize Interactions of Covariates?

To effectively visualize interactive specifications, you need:

1. A strategy for constructing counterfactuals that survey the model space

2. An algorithm that assembles logically coherent counterfactuals
   and correctly computes QoIs and their CIs

What you don't need: special machinery to calculate "marginal effects"

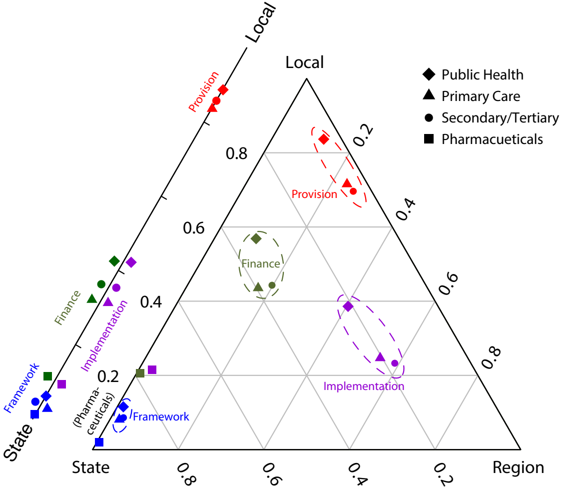
A generic counterfactual and simulation package that can use model formulas will correctly compute EVs, FDs, and RRs of the outcome variable

`simcf` does this – that's basically why it exists

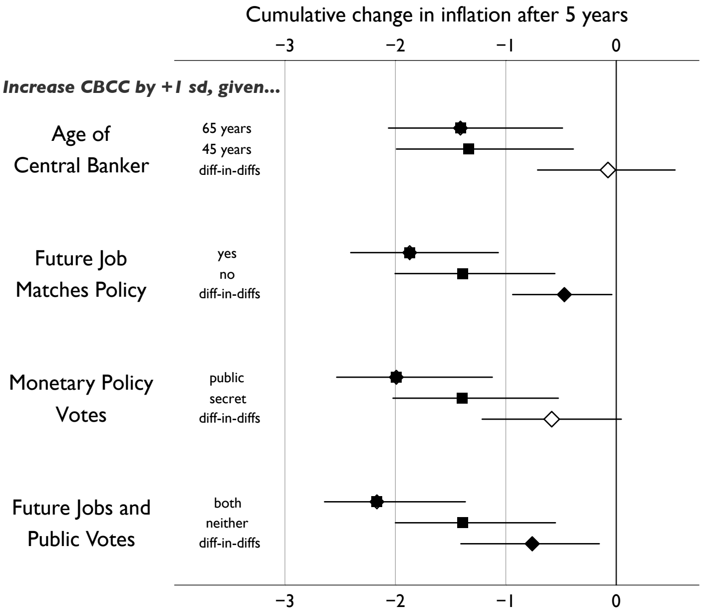
Strategies for Visualizing Interactive Covariates

Interaction	Counterfactual Strategy	Plot
discrete with discrete	one cf for each combination of values (full factorial)	`ropeladder`
continuous with discrete	choose combinations of representative values	`ropeladder`
	or	
	combine a continuum with each discrete value	`lineplot`
continuous with continuous	choose combinations of represented values	`ropeladder`
	or	
	combine a continuum with each discrete value	`lineplot`
	or	
	combine a continuum with a continuum	3D functional boxplots

Discrete × Discrete Interactions: Ropeladders

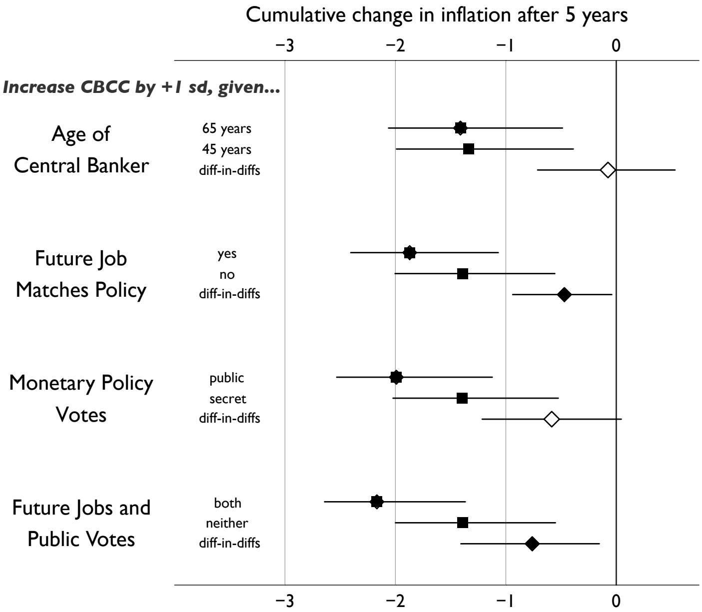


Let's convert this to a ropeladder ON THE WHITEBOARD



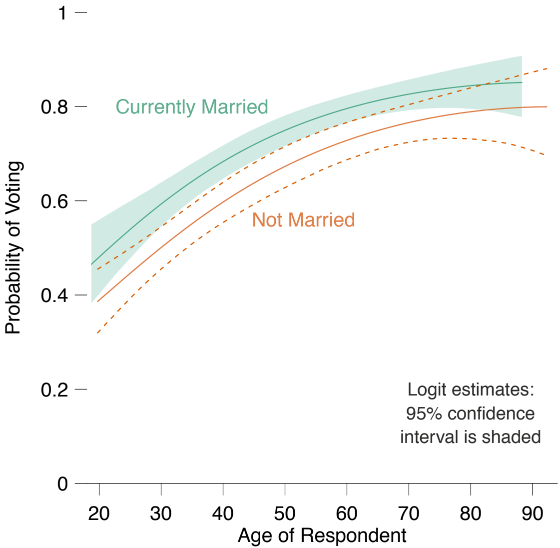
In this model of inflation, a continuous variable – central banker conservatism, CBCC – is interacted with binary variables like whether public votes are taken

Continuous × Discrete Interactions: Ropeladders



I selected a specific change in my continuous variable and computed counterfactuals under each value of my discrete variable

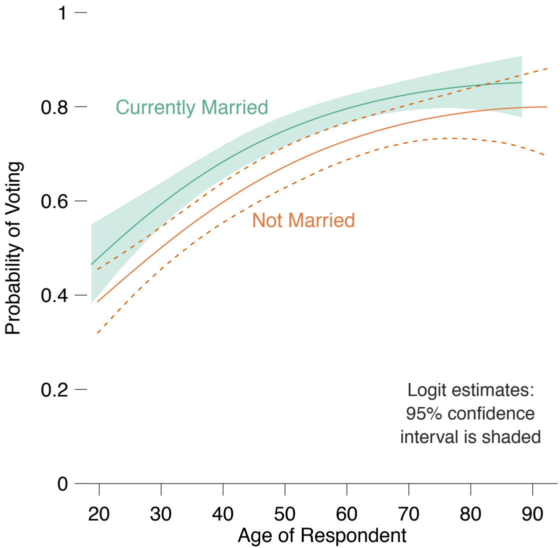
Continuous × Discrete Interactions: Lineplots



Let's return to our voting example, where voting was a function of Age, Age^2, Education, and Marriage Status

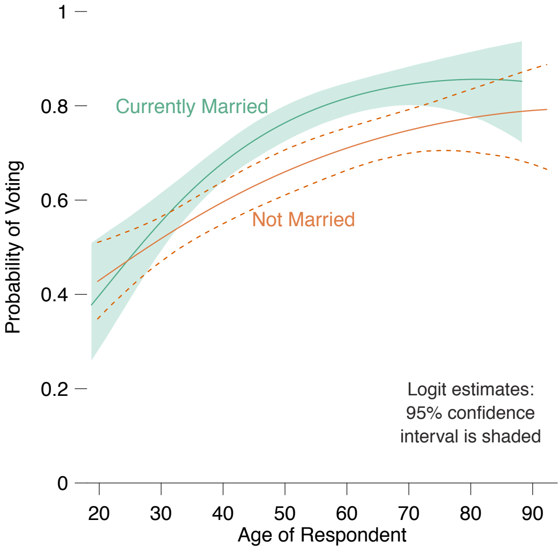
Suppose we add the interactions Age $\times$ Married and Age^2 $\times$ Married

Continuous × Discrete Interactions: Lineplots



Warning: I have no theoretical reason to do so, and the model fit suggests this is an overspecified model. We just want an example of *how* to do this

Continuous × Discrete Interactions: Lineplots



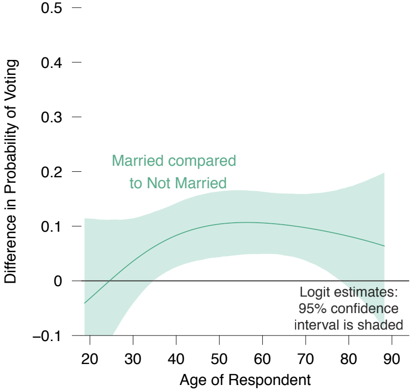
To get the new plots from `votingLineplots.R`, I need to change only one *line*:

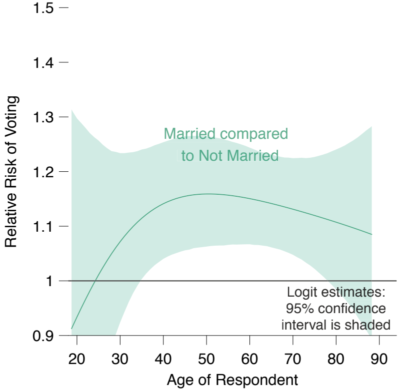
```
model2 <-
vote00 ~ age + I(age^2)
 + hsdeg + coldeg
 + marriedo
```

changes to

```
model2 <-
vote00 ~ age*marriedo
 + I(age^2)*marriedo
 + hsdeg + coldeg
```

Continuous × Discrete Interactions: Lineplots



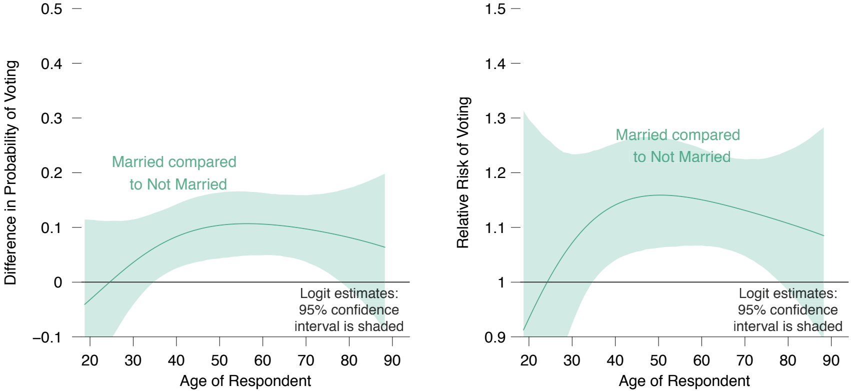


`simcf` takes care of the rest –

it will correctly set up interactions and combines their uncertainty into the QoIs

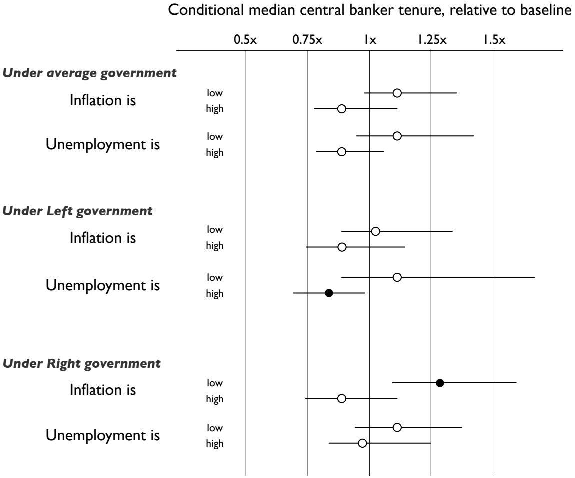
The first differences and relative risks above *are* your marginal effect plots

Continuous × Discrete Interactions: Lineplots



The results suggest this interaction wasn't a great idea...

Continuous × Continuous Interactions: Ropeladders



We've already dealt with a continuous × continuous interaction:

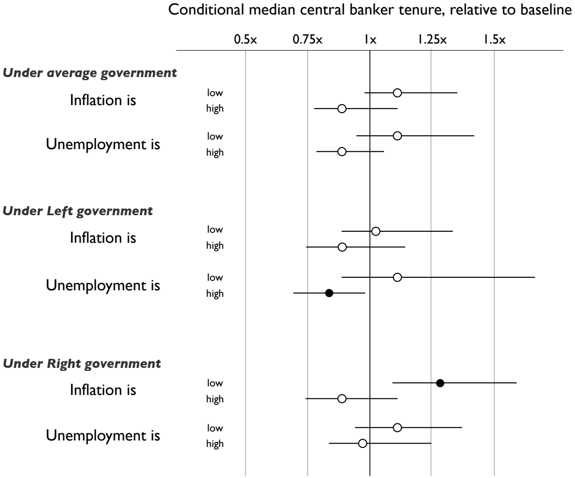
Central banker tenure depended on:

Inflation × Party CoG

and

Unemployment × Party CoG

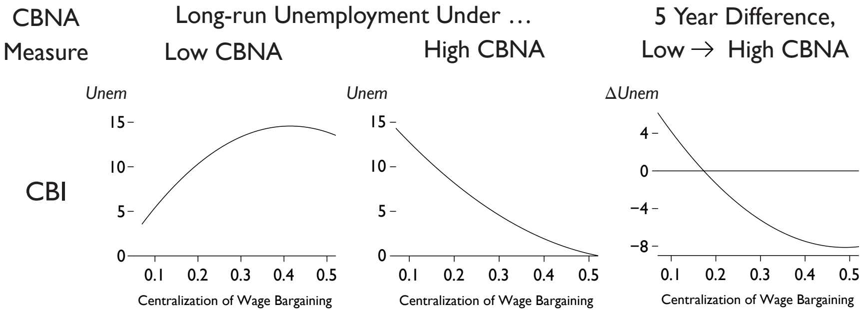
Continuous × Continuous Interactions: Ropeladders



I simply examined every combination of high, low, and average partisanship with high and low inflation or unemployment

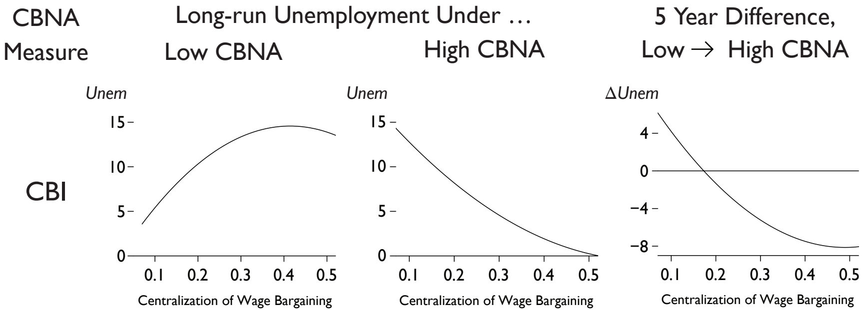
Grouping and labeling the dotplot helps catalog the combinations

Continuous × Continuous Interactions: Lineplots



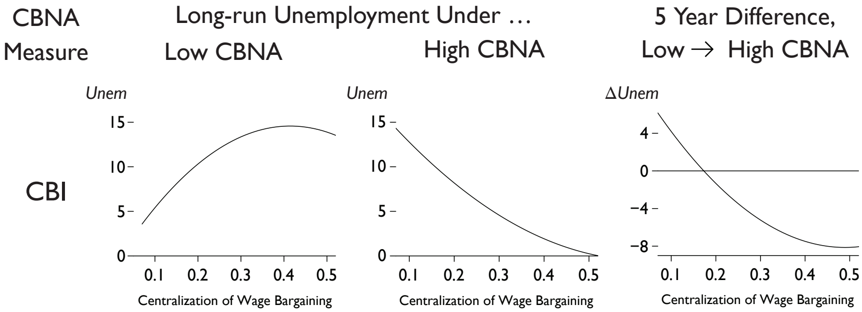
In Chapter 6 of *Bankers, Bureaucrats, and Central Bank Politics*, I consider the interactive effects of central bank "nonaccommodation" (autonomous conservatism) and wage bargaining centralization on unemployment

I build on and test a complex literature positing interactive, nonlinear effects



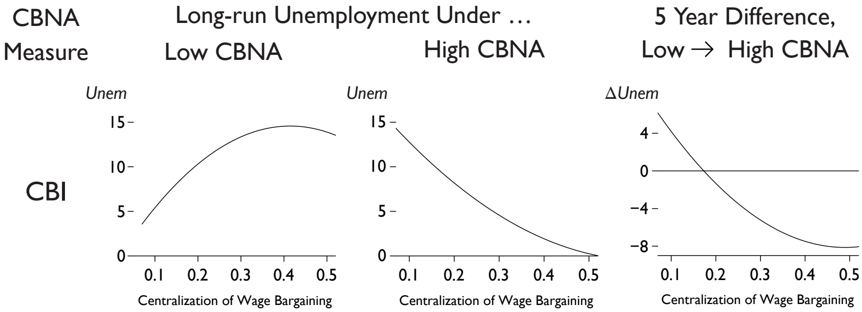
I investigate how different measures of nonaccommodation affect the results

I start with a crude “independence only” measure



The left and middle show expected unemployment across the continuum of CWB for two different levels of CBNA

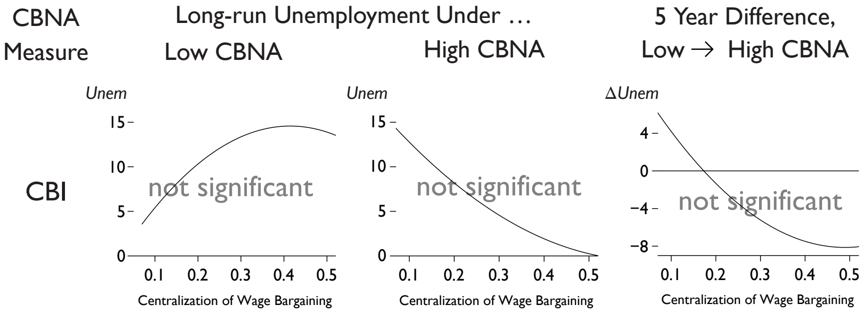
The right plot shows the first difference in unemployment given a change in CBNA at each level of CWB



This is an intuitive measure of the wage-bargaining-conditional effect of nonaccommodation

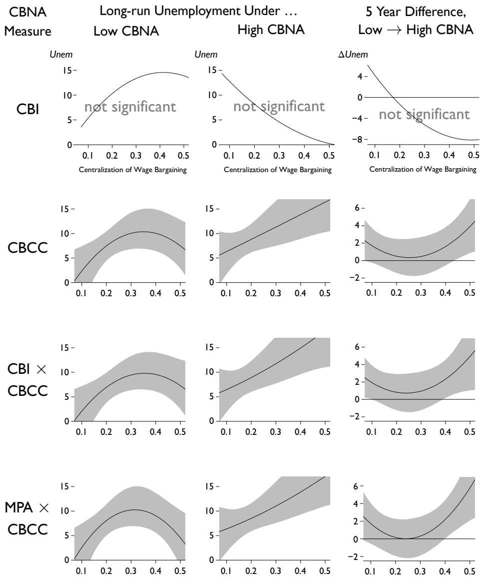
`simcf` can produce this, with the right syntax

Note we've also iterated over time, so you would use `ldvsimfd()`



Why no CIs? Because they would fill the whole plot!

I could make the plot area bigger, but that would make comparison hard

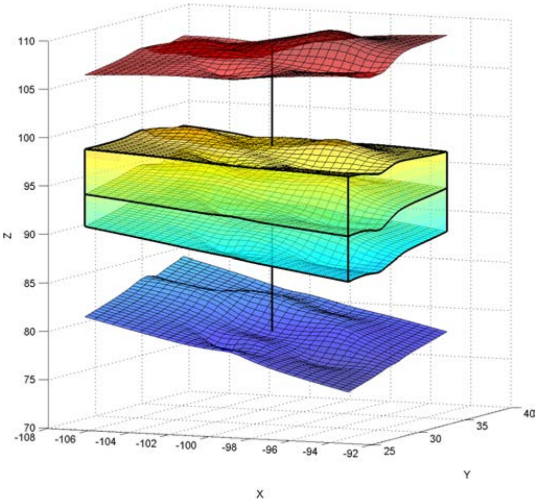


The real goal here is a robustness exercise

Measures of CBNA incorporating career conservatism produce similar and generally more precise results, alone or in combination with different measures of autonomy

Yet another approach to showing robustness – one that emphasizes similarity of fits *and* CIs for conditional relationships

Continuous × Continuous Interactions: 3D Boxplots?



I was long a skeptic of including confidence volumes in 3D plots

This example made me a believer

If it is *really* important to see smooth variation in 2 interacting continuous covariates at the same time, investigate *functional boxplots*

Source: Ying Sun and Marc G. Genton. 2011. "Functional boxplots." *JCGS* 20:2)

Strategies for Interactions of 3 Continuous Variables

Any set of interactions involving 2 or fewer continuous variables can be addressed with the above methods

What if you have 3 continuous variables? Some strategies:

Strategy 1 – Ropeladder

	X_1	X_2	X_3
	high	high	high
	high	high	low
values	high	low	high
used	high	low	low
in	low	high	high
counterfactuals	low	high	low
	low	low	high
	low	low	low

Descriptive names for these combinations essential for presentation in this case

Strategies for Interactions of 3 Continuous Variables

Any set of interactions involving 2 or fewer continuous variables can be addressed with the above methods

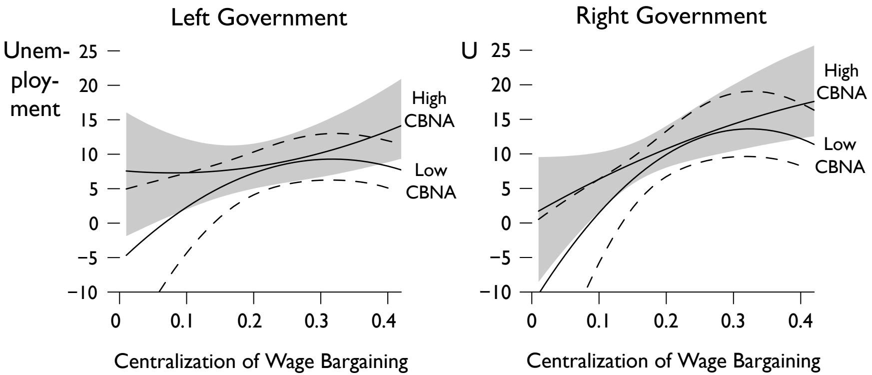
What if you have 3 continuous variables? Some strategies:

Strategy 2 – Lineplots (overlapping and/or side-by-side)

	X_1	X_2	X_3
values used in counterfactuals	contin- uum	high	high
		high	low
		low	high
		low	low

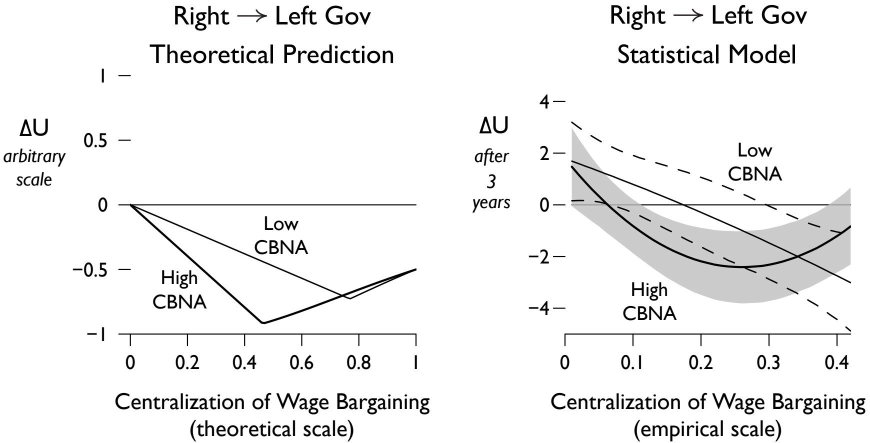
Use tiling of lineplots to your advantage in this case

Strategies for Interactions of 3 Continuous Variables

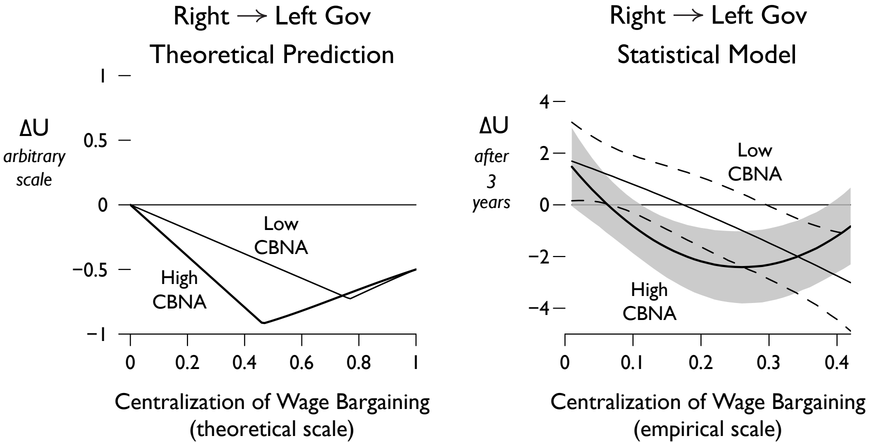


In Chapter 7 of *BBC*, I add a third interactive term to wage bargaining centralization and central bank nonaccommodation: partisanship of government

The theory and model is complex, but graphically, I just plot 4 traces instead of 2



If we plot first differences across partisanship, we're back to a 2 trace plot, but with a separate continuum of first difference for each level of CBNA



For complex models it helps to have a theory and to show it in the same format

– both to justify and to explain the empirical result

Strategies for Interactions of 3 Continuous Variables

Any set of interactions involving 2 or fewer continuous variables can be addressed with the above methods

What if you have 3 continuous variables? Some strategies:

Strategy 3 – Side-by-side 2D boxplots?

	X_1	X_2	X_3
values used in counterfactuals	contin- uum	contin- uum	high low

I've never tackled this problem – but this is the strategy I'd use

Lessons for practice of data analysis

Simulation + Graphics can summarize complex models for a broad audience

You might even find something you missed as an analyst

And even for fancy or complex models, we can and should show uncertainty

Payoff to programming: this is hard the first few times, but gets easier

Code is re-usable, and encourages more ambitious modeling

Teaching with `tile`

`tile` helps clarify data and models in research

Also helps in teaching statistical models

I incorporate this software throughout our graduate statistics sequence at UW

Greatly aids intuitive understanding of models

Find out much more, and download the software, from:

`faculty.washington.edu/cadolph`