VISUALIZING MODEL INFERENCE AND ROBUSTNESS

Christopher Adolph

Department of Political Science and Center for Statistics and the Social Sciences
University of Washington, Seattle
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
Edward Tufte and Scientific Visuals

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

1. **Assume** the reader’s interest and **intelligence**
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**
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
Minard’s Map: The Best Visual Display Ever?

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?
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
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*
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*
A Myth About Scaling

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
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
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…
Surname, Forename M. admitted 3.24.93
Right lower lobe pneumonia, hallucinations, new onset diabetes, history of manic depressive illness

<table>
<thead>
<tr>
<th>-1yr 3.24</th>
<th>4.4.93</th>
<th>4.4.93</th>
<th>-1yr 3.24</th>
<th>4.4.93</th>
<th>4.4.93</th>
<th>-1yr 3.24</th>
<th>4.4.93</th>
<th>4.4.93</th>
</tr>
</thead>
<tbody>
<tr>
<td>WBC</td>
<td>11100 c/µl</td>
<td>11100 c/µl</td>
<td>Psychosis</td>
<td>0</td>
<td>0</td>
<td>Glucose</td>
<td>237 mg/dl</td>
<td>237 mg/dl</td>
</tr>
<tr>
<td>Mood</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

T 98.8° F

<table>
<thead>
<tr>
<th>-1yr 3.24</th>
<th>4.4.93</th>
<th>4.4.93</th>
<th>-1yr 3.24</th>
<th>4.4.93</th>
<th>4.4.93</th>
<th>-1yr 3.24</th>
<th>4.4.93</th>
<th>4.4.93</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>98.8° F</td>
<td>98.8° F</td>
<td>Psychosis</td>
<td>0</td>
<td>0</td>
<td>Glucose</td>
<td>237 mg/dl</td>
<td>237 mg/dl</td>
</tr>
<tr>
<td>Mood</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Discharge. PB MD 1200 4.4.93
No delirium. JT MD 900 4.4.93
Enema given. PAC RN 1100 4.3.93
Will treat for probable constipation. MBM 2245 4.2.93
Vomited three times. RW RN 2230 4.2.93
Left lower lobe infiltrate or atelectasis. AL MD 1500 4.2.93
Alert and oriented. No complaints. PAC RN 1100 4.1.93
Attending to activities of daily living. PAC RN 1100 3.31.93

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 visualization of data

InfoVis: emphasis on fun, exploration, beauty, and “wow”

Scientific Visuals: structured comparison, precision, and inference
Everyone
The essentials — sleeping, eating, and working — take up the better part of the day, often ended with watching television.

---

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

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

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

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

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).
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

### Estimates of relative survival rates, by cancer site

<table>
<thead>
<tr>
<th></th>
<th>% survival rates and standard errors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5 year</td>
</tr>
<tr>
<td>Prostate</td>
<td>98.8 ± 0.4</td>
</tr>
<tr>
<td>Thyroid</td>
<td>96.0 ± 0.8</td>
</tr>
<tr>
<td>Testis</td>
<td>94.7 ± 1.1</td>
</tr>
<tr>
<td>Melanomas</td>
<td>89.0 ± 0.8</td>
</tr>
<tr>
<td>Breast</td>
<td>86.4 ± 0.4</td>
</tr>
<tr>
<td>Hodgkin’s disease</td>
<td>85.1 ± 1.7</td>
</tr>
<tr>
<td>Corpus uteri, uterus</td>
<td>84.3 ± 1.0</td>
</tr>
<tr>
<td>Urinary, bladder</td>
<td>82.1 ± 1.0</td>
</tr>
<tr>
<td>Cervix, uteri</td>
<td>70.5 ± 1.6</td>
</tr>
<tr>
<td>Larynx</td>
<td>68.8 ± 2.1</td>
</tr>
<tr>
<td>Rectum</td>
<td>62.6 ± 1.2</td>
</tr>
<tr>
<td>Kidney, renal pelvis</td>
<td>61.8 ± 1.3</td>
</tr>
<tr>
<td>Colon</td>
<td>61.7 ± 0.8</td>
</tr>
<tr>
<td>Non-Hodgkin’s</td>
<td>57.8 ± 1.0</td>
</tr>
<tr>
<td>Oral cavity, pharynx</td>
<td>56.7 ± 1.3</td>
</tr>
<tr>
<td>Ovary</td>
<td>55.0 ± 1.3</td>
</tr>
<tr>
<td>Leukemia</td>
<td>42.5 ± 1.2</td>
</tr>
<tr>
<td>Brain, nervous system</td>
<td>32.0 ± 1.4</td>
</tr>
<tr>
<td>Multiple myeloma</td>
<td>29.5 ± 1.6</td>
</tr>
<tr>
<td>Stomach</td>
<td>23.8 ± 1.3</td>
</tr>
<tr>
<td>Lung and bronchus</td>
<td>15.0 ± 0.4</td>
</tr>
<tr>
<td>Esophagus</td>
<td>14.2 ± 1.4</td>
</tr>
<tr>
<td>Liver, bile duct</td>
<td>7.5 ± 1.1</td>
</tr>
<tr>
<td>Pancreas</td>
<td>4.0 ± 0.5</td>
</tr>
</tbody>
</table>

*Usually graphics are more effective than tables*
### Estimates of relative survival rates, by cancer site

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

### 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

<table>
<thead>
<tr>
<th>Cancer Site</th>
<th>% survival rates and standard errors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5 year</td>
</tr>
<tr>
<td>Prostate</td>
<td>98.8 0.4</td>
</tr>
<tr>
<td>Thyroid</td>
<td>96.0 0.8</td>
</tr>
<tr>
<td>Testis</td>
<td>94.7 1.1</td>
</tr>
<tr>
<td>Melanomas</td>
<td>89.0 0.8</td>
</tr>
<tr>
<td>Breast</td>
<td>86.4 0.4</td>
</tr>
<tr>
<td>Hodgkin’s disease</td>
<td>85.1 1.7</td>
</tr>
<tr>
<td>Corpus uteri, uterus</td>
<td>84.3 1.0</td>
</tr>
<tr>
<td>Urinary, bladder</td>
<td>82.1 1.0</td>
</tr>
<tr>
<td>Cervix, uteri</td>
<td>70.5 1.6</td>
</tr>
<tr>
<td>Larynx</td>
<td>68.8 2.1</td>
</tr>
<tr>
<td>Rectum</td>
<td>62.6 1.2</td>
</tr>
<tr>
<td>Kidney, renal pelvis</td>
<td>61.8 1.3</td>
</tr>
<tr>
<td>Colon</td>
<td>61.7 0.8</td>
</tr>
<tr>
<td>Non-Hodgkin’s</td>
<td>57.8 1.0</td>
</tr>
<tr>
<td>Oral cavity, pharynx</td>
<td>56.7 1.3</td>
</tr>
<tr>
<td>Ovary</td>
<td>55.0 1.3</td>
</tr>
<tr>
<td>Leukemia</td>
<td>42.5 1.2</td>
</tr>
<tr>
<td>Brain, nervous system</td>
<td>32.0 1.4</td>
</tr>
<tr>
<td>Multiple myeloma</td>
<td>29.5 1.6</td>
</tr>
<tr>
<td>Stomach</td>
<td>23.8 1.3</td>
</tr>
<tr>
<td>Lung and bronchus</td>
<td>15.0 0.4</td>
</tr>
<tr>
<td>Esophagus</td>
<td>14.2 1.4</td>
</tr>
<tr>
<td>Liver, bile duct</td>
<td>7.5 1.1</td>
</tr>
<tr>
<td>Pancreas</td>
<td>4.0 0.5</td>
</tr>
</tbody>
</table>

---

### 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

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

---

### 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

<table>
<thead>
<tr>
<th>Cancer Site</th>
<th>% survival rates and standard errors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5 year</td>
</tr>
<tr>
<td>Prostate</td>
<td>98.8  0.4</td>
</tr>
<tr>
<td>Thyroid</td>
<td>96.0  0.8</td>
</tr>
<tr>
<td>Testis</td>
<td>94.7  1.1</td>
</tr>
<tr>
<td>Melanomas</td>
<td>89.0  0.8</td>
</tr>
<tr>
<td>Breast</td>
<td>86.4  0.4</td>
</tr>
<tr>
<td>Hodgkin’s disease</td>
<td>85.1  1.7</td>
</tr>
<tr>
<td>Corpus uteri, uterus</td>
<td>84.3  1.0</td>
</tr>
<tr>
<td>Urinary, bladder</td>
<td>82.1  1.0</td>
</tr>
<tr>
<td>Cervix, uteri</td>
<td>70.5  1.6</td>
</tr>
<tr>
<td>Larynx</td>
<td>68.8  2.1</td>
</tr>
<tr>
<td>Rectum</td>
<td>62.6  1.2</td>
</tr>
<tr>
<td>Kidney, renal pelvis</td>
<td>61.8  1.3</td>
</tr>
<tr>
<td>Colon</td>
<td>61.7  0.8</td>
</tr>
<tr>
<td>Non-Hodgkin’s</td>
<td>57.8  1.0</td>
</tr>
<tr>
<td>Oral cavity, pharynx</td>
<td>56.7  1.3</td>
</tr>
<tr>
<td>Ovary</td>
<td>55.0  1.3</td>
</tr>
<tr>
<td>Leukemia</td>
<td>42.5  1.2</td>
</tr>
<tr>
<td>Brain, nervous system</td>
<td>32.0  1.4</td>
</tr>
<tr>
<td>Multiple myeloma</td>
<td>29.5  1.6</td>
</tr>
<tr>
<td>Stomach</td>
<td>23.8  1.3</td>
</tr>
<tr>
<td>Lung and bronchus</td>
<td>15.0  0.4</td>
</tr>
<tr>
<td>Esophagus</td>
<td>14.2  1.4</td>
</tr>
<tr>
<td>Liver, bile duct</td>
<td>7.5   1.1</td>
</tr>
<tr>
<td>Pancreas</td>
<td>4.0   0.5</td>
</tr>
</tbody>
</table>

### 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.
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.

<table>
<thead>
<tr>
<th>Cancer Site</th>
<th>% survival rates and standard errors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5 year</td>
</tr>
<tr>
<td>Prostate</td>
<td>98.8</td>
</tr>
<tr>
<td>Thyroid</td>
<td>96.0</td>
</tr>
<tr>
<td>Testis</td>
<td>94.7</td>
</tr>
<tr>
<td>Melanomas</td>
<td>89.0</td>
</tr>
<tr>
<td>Breast</td>
<td>86.4</td>
</tr>
<tr>
<td>Hodgkin's disease</td>
<td>85.1</td>
</tr>
<tr>
<td>Corpus uteri, uterus</td>
<td>84.3</td>
</tr>
<tr>
<td>Urinary, bladder</td>
<td>82.1</td>
</tr>
<tr>
<td>Cervix, uteri</td>
<td>70.5</td>
</tr>
<tr>
<td>Larynx</td>
<td>68.8</td>
</tr>
<tr>
<td>Rectum</td>
<td>62.6</td>
</tr>
<tr>
<td>Kidney, renal pelvis</td>
<td>61.8</td>
</tr>
<tr>
<td>Colon</td>
<td>61.7</td>
</tr>
<tr>
<td>Non-Hodgkin's</td>
<td>57.8</td>
</tr>
<tr>
<td>Oral cavity, pharynx</td>
<td>56.7</td>
</tr>
<tr>
<td>Ovary</td>
<td>55.0</td>
</tr>
<tr>
<td>Leukemia</td>
<td>42.5</td>
</tr>
<tr>
<td>Brain, nervous system</td>
<td>32.0</td>
</tr>
<tr>
<td>Multiple myeloma</td>
<td>29.5</td>
</tr>
<tr>
<td>Stomach</td>
<td>23.8</td>
</tr>
<tr>
<td>Lung and bronchus</td>
<td>15.0</td>
</tr>
<tr>
<td>Esophagus</td>
<td>14.2</td>
</tr>
<tr>
<td>Liver, bile duct</td>
<td>7.5</td>
</tr>
<tr>
<td>Pancreas</td>
<td>4.0</td>
</tr>
</tbody>
</table>
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”.

<table>
<thead>
<tr>
<th></th>
<th>5 year</th>
<th>10 year</th>
<th>15 year</th>
<th>20 year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prostate</td>
<td>99</td>
<td>95</td>
<td>87</td>
<td>81</td>
</tr>
<tr>
<td>Thyroid</td>
<td>96</td>
<td>96</td>
<td>94</td>
<td>95</td>
</tr>
<tr>
<td>Testis</td>
<td>95</td>
<td>94</td>
<td>91</td>
<td>88</td>
</tr>
<tr>
<td>Melanomas</td>
<td>89</td>
<td>87</td>
<td>84</td>
<td>83</td>
</tr>
<tr>
<td>Breast</td>
<td>86</td>
<td>78</td>
<td>71</td>
<td>65</td>
</tr>
<tr>
<td>Hodgkin's disease</td>
<td>85</td>
<td>80</td>
<td>74</td>
<td>65</td>
</tr>
<tr>
<td>Corpus uteri, uterus</td>
<td>84</td>
<td>83</td>
<td>81</td>
<td>67</td>
</tr>
<tr>
<td>Urinary, bladder</td>
<td>82</td>
<td>76</td>
<td>74</td>
<td>79</td>
</tr>
<tr>
<td>Cervix, uteri</td>
<td>71</td>
<td>64</td>
<td>63</td>
<td>60</td>
</tr>
<tr>
<td>Larynx</td>
<td>69</td>
<td>57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rectum</td>
<td>63</td>
<td>46</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>Kidney, renal pelvis</td>
<td>62</td>
<td>55</td>
<td>52</td>
<td>38</td>
</tr>
<tr>
<td>Colon</td>
<td>62</td>
<td>54</td>
<td>50</td>
<td>49</td>
</tr>
<tr>
<td>Non-Hodgkin's</td>
<td>58</td>
<td>55</td>
<td>54</td>
<td>47</td>
</tr>
<tr>
<td>Oral cavity, pharynx</td>
<td>57</td>
<td>46</td>
<td>38</td>
<td>34</td>
</tr>
</tbody>
</table>
What’s missing from the figure that was in the table?

<table>
<thead>
<tr>
<th>Measure of uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard errors</td>
</tr>
</tbody>
</table>

The table had standard errors as measures of uncertainty.
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

<table>
<thead>
<tr>
<th>Country</th>
<th>National</th>
<th>Regional (Lander)</th>
<th>Local (stadt/Gemeind)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pharmaceutical</td>
<td>L1/L2/F/R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary care</td>
<td>L1/F</td>
<td>L2/P</td>
<td></td>
</tr>
<tr>
<td>Sec and Ter care</td>
<td>L1/F</td>
<td>L2/P</td>
<td></td>
</tr>
<tr>
<td>Public Health</td>
<td>F</td>
<td>L1</td>
<td>L2/P</td>
</tr>
<tr>
<td>Belgium</td>
<td></td>
<td>Regional (Regions)</td>
<td>Local (Communes)</td>
</tr>
<tr>
<td>Pharmaceutical</td>
<td>L1/L2/F/R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary care</td>
<td>L1/F</td>
<td>L2/P</td>
<td></td>
</tr>
<tr>
<td>Sec and Ter care</td>
<td>L1/F</td>
<td>L2/P</td>
<td></td>
</tr>
<tr>
<td>Public Health</td>
<td>L1/F</td>
<td>L2/P</td>
<td></td>
</tr>
<tr>
<td>Bulgaria</td>
<td></td>
<td>Regional (Oblasti)</td>
<td>Local (Obshtina)</td>
</tr>
<tr>
<td>Pharmaceutical</td>
<td>L1/L2//R/F</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary care</td>
<td>L1/L2/F</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>Sec and Ter care</td>
<td>L1/L2/F</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>Public Health</td>
<td>L1/L2/F/P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyprus</td>
<td></td>
<td>No elected regional level</td>
<td>Local (eparchies)</td>
</tr>
<tr>
<td>Pharmaceutical</td>
<td>L1/L2/F/R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary care</td>
<td>L1/L2/F/P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sec and Ter care</td>
<td>L1/L2/F/P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public Health</td>
<td>L1/L2/F/P</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

An example from work I did with Scott Greer and Elize Fonseca ("Allocation of Authority in European Health Policy," *Social Science and Medicine*, 2012)
<table>
<thead>
<tr>
<th>Country</th>
<th>Level (National)</th>
<th>Level (Regional)</th>
<th>Level (Local)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>L1/L2/F/R</td>
<td>L1/F L2/P</td>
<td>L1 L2/P</td>
</tr>
<tr>
<td>Belgium</td>
<td>L1/L2/F/R</td>
<td>L1/F L2/P</td>
<td>L1 L2/P</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>L1/L2/F/R</td>
<td>L1/F L2/P</td>
<td>L1 L2/P</td>
</tr>
<tr>
<td>Cyprus</td>
<td>L1/L2/F/R</td>
<td>No elected</td>
<td>L1 L2/F/P</td>
</tr>
</tbody>
</table>

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

**DV: level of government**  
state, regional, or local
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
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

Chris Adolph (University of Washington)
For now, let’s do a little exploratory data analysis (EDA)

Is there even interesting variation in the allocation of authority to explain?
### Problems with this table?

1. Poor distinction 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

<table>
<thead>
<tr>
<th>Country</th>
<th>National</th>
<th>Regional (Land)</th>
<th>Local (stadt/Gemind)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>L1/L2/F/R</td>
<td>L2/P</td>
<td></td>
</tr>
<tr>
<td>Belgium</td>
<td>L1/L2/F/R</td>
<td>L2/P</td>
<td></td>
</tr>
<tr>
<td>Bulgaria</td>
<td>L1/L2/F/R</td>
<td>L2/P</td>
<td></td>
</tr>
<tr>
<td>Cyprus</td>
<td>L1/L2/F/R</td>
<td>L2/P</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Country</th>
<th>National</th>
<th>Regional (Regions)</th>
<th>Local (Communes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgium</td>
<td>L1/L2/F/R</td>
<td>L2/P</td>
<td></td>
</tr>
<tr>
<td>Bulgaria</td>
<td>L1/L2/F/R</td>
<td>L2/P</td>
<td></td>
</tr>
<tr>
<td>Cyprus</td>
<td>L1/L2/F/R</td>
<td>L2/P</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Country</th>
<th>National</th>
<th>Regional (Oblasti)</th>
<th>Local (Obshtina)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bulgaria</td>
<td>L1/L2/F/R</td>
<td>L2/P</td>
<td></td>
</tr>
<tr>
<td>Cyprus</td>
<td>L1/L2/F/R</td>
<td>L2/P</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Country</th>
<th>National</th>
<th>No elected regional level</th>
<th>Local (eparchies)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cyprus</td>
<td>L1/L2/F/R</td>
<td>L2/P</td>
<td></td>
</tr>
</tbody>
</table>

### 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
### Problems with this table?

3. Coding system is opaque (L1? L2? etc.)
<table>
<thead>
<tr>
<th>Country</th>
<th>National</th>
<th>Regional (Länder)</th>
<th>Local (stadt/Gemeinden)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pharmaceutical</td>
<td>L1/L2/F/R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary care</td>
<td>L1/F</td>
<td>L2/P</td>
<td></td>
</tr>
<tr>
<td>Sec and Ter care</td>
<td>L1/F</td>
<td>L2/P</td>
<td></td>
</tr>
<tr>
<td>Public Health</td>
<td>F</td>
<td>L1</td>
<td>L2/P</td>
</tr>
<tr>
<td>Belgium</td>
<td></td>
<td>Regional (Regions)</td>
<td>Local (Communes)</td>
</tr>
<tr>
<td>Pharmaceutical</td>
<td>L1/L2/F/R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary care</td>
<td>L1/F</td>
<td>L2/P</td>
<td></td>
</tr>
<tr>
<td>Sec and Ter care</td>
<td>L1/F</td>
<td>L2/P</td>
<td></td>
</tr>
<tr>
<td>Public Health</td>
<td>L1/F</td>
<td>L2/P</td>
<td></td>
</tr>
<tr>
<td>Bulgaria</td>
<td></td>
<td>Regional (Oblasti)</td>
<td>Local (Obshtini)</td>
</tr>
<tr>
<td>Pharmaceutical</td>
<td>L1/L2//R/F</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary care</td>
<td>L1/L2/F</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>Sec and Ter care</td>
<td>L1/L2/F</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>Public Health</td>
<td>L1/L2/F/P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyprus</td>
<td></td>
<td>No elected regional level</td>
<td>Local (eparchies)</td>
</tr>
<tr>
<td>Pharmaceutical</td>
<td>L1/L2/F/R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary care</td>
<td>L1/L2/F/P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sec and Ter care</td>
<td>L1/L2/F/P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public Health</td>
<td>L1/L2/F/P</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Problems with this table?**

4. Table goes on for four pages! Hard to spot patterns
<table>
<thead>
<tr>
<th>Available Levels of Government</th>
<th>Observed Allocation of Authority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pub Health Provision</td>
<td>Let’s redraw the table as a heatmap</td>
</tr>
<tr>
<td>Sec/Tert Provision</td>
<td></td>
</tr>
<tr>
<td>Primary Provision</td>
<td></td>
</tr>
<tr>
<td>Pub Health Implement</td>
<td></td>
</tr>
<tr>
<td>Sec/Tert Implement</td>
<td></td>
</tr>
<tr>
<td>Primary Implement</td>
<td></td>
</tr>
<tr>
<td>Pub Health Finance</td>
<td></td>
</tr>
<tr>
<td>Primary Finance</td>
<td></td>
</tr>
<tr>
<td>Sec/Tert Finance</td>
<td></td>
</tr>
<tr>
<td>Pharma Finance</td>
<td></td>
</tr>
<tr>
<td>Pub Health Framework</td>
<td></td>
</tr>
<tr>
<td>Primary Framework</td>
<td></td>
</tr>
<tr>
<td>Sec/Tert Framework</td>
<td></td>
</tr>
<tr>
<td>Pharma Regulation</td>
<td></td>
</tr>
<tr>
<td>Pharma Framework</td>
<td></td>
</tr>
<tr>
<td>Pharma Implement</td>
<td></td>
</tr>
</tbody>
</table>
Allocation of authority is highlighted as the shading of each cell.
Rows and columns code covariates, and are sorted by cluster analysis.
<table>
<thead>
<tr>
<th>Available Levels of Government</th>
<th>Observed Allocation of Authority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two Tiers</td>
<td>Local</td>
</tr>
<tr>
<td></td>
<td>Regional</td>
</tr>
<tr>
<td></td>
<td>State</td>
</tr>
</tbody>
</table>

**Rows and columns colors help detect patterns in clustering**
We did not sort by row color, so any emergent pattern is noteworthy.
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?

Unfortunately, cognitive errors are everywhere. But few designers of scientific visuals pay close attention to them.
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 algorithm to accomplish the following:

\[
\text{cognitivelyAdjustedGraphic} \leftarrow \text{correctForErrors(InitialGraphic)}
\]
Unfortunately, cognitive errors are everywhere

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

\[
\text{cognitivelyAdjustedGraphic} \leftarrow \text{correctForErrors(InitialGraphic)}
\]

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
**Graphical elements used to encode data:**

<table>
<thead>
<tr>
<th>More accurate</th>
<th>Less accurate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position on a plane</td>
<td>Color</td>
</tr>
<tr>
<td>Line length</td>
<td></td>
</tr>
<tr>
<td>Angle &amp; slope</td>
<td></td>
</tr>
<tr>
<td>Area</td>
<td></td>
</tr>
<tr>
<td>Volume</td>
<td></td>
</tr>
</tbody>
</table>

Graphical elements are *not* all equal in clarity. People are much better at judging line length than angle or grayscale.

*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

To show and correlate multivariate data, we’d like to use multiple or multifunctional elements

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

Will they all be processed equally... accurately? quickly? with similar intensity? separately? at highest available level of measurement? Unfortunately, NO.

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

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

Income
- Social security, Medicare, and unemployment and other retirement taxes: 30%
- Personal income taxes: 39%
- Excise, customs, estate, gift, and miscellaneous taxes: 6%
- Corporate income taxes: 10%
- Borrowing to cover deficit: 15%

Outlays
- Social security, Medicare, and other retirement: 37%
- National defense, veterans, and foreign affairs: 24%
- Net interest on the debt: 8%
- Physical, human, and community development: 9%
- Social programs: 20%
- Law enforcement and general government: 2%

Budget pies went missing for most of the decade
They returned in 2009 – right after a presidential election
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
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 line plots 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:

85689726984689762689764358922659865986554897689269898
02462996874026557627986789045679232769285460986772098
90834579802790759047098279085790847729087590827908754
98709856749068975786259845690243790472190790709811450
85689726984689762689764458922659865986554897689269898

Same information in both examples.

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

Source: Ware, Information Visualization
Sometimes differences take conscious effort to distinguish. Find the 3’s:

85689726984689762689764358922659865986554897689269898
02462996874026557627986789045679232769285460986772098
90834579802790759047098279085790847729087590827908754
98709856749068975786259845690243790472190790709811450
85689726984689762689764458922659865986554897689269898

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

85689726984689762689764358922659865986554897689269898
02462996874026557627986789045679232769285460986772098
90834579802790759047098279085790847729087590827908754
98709856749068975786259845690243790472190790709811450
85689726984689762689764458922659865986554897689269898

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*
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

*Source:* Ware, *Information Visualization*
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

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**

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*

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.
Using Color (In)effectively

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
Using Color Effectively

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
Available Levels of Government

- Two Tiers
- Three Tiers

Observed Allocation of Authority

- Local
- Regional
- State

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

Chris Adolph (University of Washington)
<table>
<thead>
<tr>
<th>Available Levels of Government</th>
<th>Observed Allocation of Authority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two Tiers</td>
<td>Three Tiers</td>
</tr>
</tbody>
</table>

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
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
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
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 wavelengths by sampling in still more places.

Just like adding bins to a histogram, this gives us a finer view of the distribution.
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!
Primaries and color can be expressed in many equivalent ways.

These are different **colors**
spaces: 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

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

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

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

So why should you use the HSV color space if the OS itself needs colors to be in RGB space to display them? There are two main times RGB is inconvenient. The first is when you want to get a color from a typical user. Since most users don’t understand the nuances of RGB, you need to present them with a way to pick colors which they can understand. Apple’s HSV color picker does this very well. The second time you want to use HSV space is when you have to match colors, or programatically determine if one color is similar to another color. Let’s look at why RGB comparisons are difficult.

The RGB color space is conceptually a cube with one axis representing red, one representing green, and one representing blue, as shown below.

Figure 2: The HSV Cone

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

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

As you can see, where the axes meet at (0,0,0), we have black, and at (255,255,255) (or (1.0,1.0,1.0) if you prefer), we have white. How do we tell if a color in an image is close to a color we picked? We could take the Euclidian distance between the two colors and see if it’s less than a “similarity” parameter. That sounds reasonable, but let’s look at how this works from a perceptual standpoint. Let’s say you allow the user to set a “similarity” threshold for matching all colors that are similar to a chosen color. In RGB space, all points less than or equal to the “similarity” distance from the chosen color form a sphere inside the RGB cube. The user probably thinks of matching a color by choosing “all the bluish tones”, or some similar perceptual way. But the sphere we get in the RGB cube doesn’t

Adventures in HSV Space, page 2

Source: Darrin Cardani, “Adventures in HSV Space”

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
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

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

Source: Colin Ware, Information Visualization
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 Euclidean distances in CIElab space are (approximately) perceptually “equal” to humans
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 steps 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`

```r
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?
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 $\text{Set1}$ 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?
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
Color Cognition Problem 1  Saturation and Size

*Use jewel tones for glyphs*  *Use pastels for 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.
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.
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 substantial 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 substantial 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.
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 substantial 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...
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?
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.
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
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, limiting readers’ and analysts’ understanding of the results.
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!
Key need  Ready-to-use techniques to visually present model results:

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

- for many variables
- for many robustness checks
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

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
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
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 foreseen?
The Challenger launch decision

<table>
<thead>
<tr>
<th>Flt Number</th>
<th>Temp (F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>70</td>
</tr>
<tr>
<td>41b</td>
<td>57</td>
</tr>
<tr>
<td>41c</td>
<td>63</td>
</tr>
<tr>
<td>41d</td>
<td>70</td>
</tr>
<tr>
<td>51c</td>
<td>53</td>
</tr>
<tr>
<td>61a</td>
<td>79</td>
</tr>
<tr>
<td>61c</td>
<td>58</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Flt Number</th>
<th>Temp (F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>70</td>
</tr>
<tr>
<td>41b</td>
<td>57</td>
</tr>
<tr>
<td>41c</td>
<td>63</td>
</tr>
<tr>
<td>41d</td>
<td>70</td>
</tr>
<tr>
<td>51c</td>
<td>53</td>
</tr>
<tr>
<td>61a</td>
<td>79</td>
</tr>
<tr>
<td>61c</td>
<td>58</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Flt Number</th>
<th>Temp (F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>70</td>
</tr>
<tr>
<td>41b</td>
<td>57</td>
</tr>
<tr>
<td>41c</td>
<td>63</td>
</tr>
<tr>
<td>41d</td>
<td>70</td>
</tr>
<tr>
<td>51c</td>
<td>53</td>
</tr>
<tr>
<td>61a</td>
<td>79</td>
</tr>
<tr>
<td>61c</td>
<td>58</td>
</tr>
</tbody>
</table>

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”

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

<table>
<thead>
<tr>
<th>Flt Number</th>
<th>Temp (F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>70</td>
</tr>
<tr>
<td>41b</td>
<td>57</td>
</tr>
<tr>
<td>41c</td>
<td>63</td>
</tr>
<tr>
<td>41d</td>
<td>70</td>
</tr>
<tr>
<td>51c</td>
<td>53</td>
</tr>
<tr>
<td>61a</td>
<td>79</td>
</tr>
<tr>
<td>61c</td>
<td>58</td>
</tr>
</tbody>
</table>

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

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

Flights with O-ring damage

<table>
<thead>
<tr>
<th>Flt Number</th>
<th>Temp (F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>70</td>
</tr>
<tr>
<td>41b</td>
<td>57</td>
</tr>
<tr>
<td>41c</td>
<td>63</td>
</tr>
<tr>
<td>41d</td>
<td>70</td>
</tr>
<tr>
<td>51c</td>
<td>53</td>
</tr>
<tr>
<td>61a</td>
<td>79</td>
</tr>
<tr>
<td>61c</td>
<td>58</td>
</tr>
</tbody>
</table>

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 evidence begins to speak for itself.

What if Morton-Thiokol engineers had made this table before the launch?
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 unprecedented cold
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?
“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.
The Challenger launch decision

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

\[
Pr(Damage) = \frac{1}{1 + e^{-(\hat{\beta}_0 + \hat{\beta}_1 \text{Temp})}}
\]

R gives us this lovely logit output…
Let’s try a simple logit model of damage as a function of temperature:

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

R gives us this lovely logit output...

<table>
<thead>
<tr>
<th>Variable</th>
<th>est.</th>
<th>s.e.</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature (F)</td>
<td>-0.18</td>
<td>0.09</td>
<td>0.047</td>
</tr>
<tr>
<td>Constant</td>
<td>11.9</td>
<td>6.34</td>
<td>0.062</td>
</tr>
<tr>
<td>N</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>log-likelihood</td>
<td>-10.9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Let’s try a simple logit model of damage as a function of temperature:

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

R gives us this lovely logit output…

<table>
<thead>
<tr>
<th>Variable</th>
<th>est.</th>
<th>s.e.</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature (F)</td>
<td>-0.18</td>
<td>0.09</td>
<td>0.047</td>
</tr>
<tr>
<td>Constant</td>
<td>11.9</td>
<td>6.34</td>
<td>0.062</td>
</tr>
<tr>
<td>N</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>log-likelihood</td>
<td>-10.9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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?

How many F's are acceptable?

What's an acceptable Pr(Damage)?

67% CI

95% CI

Launch Temperature (F)

Pr(O−ring Damage)

predicted probability
The Challenger launch decision

In a hearing, Richard Feynmann dramatically showed O-rings lose resilience 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 assessments 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 quantities (for most people, this includes logit coefficients)
- Little effort to transform results to the scale of the quantities of interest → really want the conditional expectation, $\mathbb{E}(y|x)$
- Little effort to make informative statements about estimation uncertainty → 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 ($\text{Vote}_i = 1$), or they don’t ($\text{Vote}_i = 0$).

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?”

<table>
<thead>
<tr>
<th>vote00</th>
<th>age</th>
<th>hsdeg</th>
<th>coldeg</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>49</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>35</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>57</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>63</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>40</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>77</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>43</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>47</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>26</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>48</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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?

---

<table>
<thead>
<tr>
<th></th>
<th>est.</th>
<th>s.e.</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.074</td>
<td>0.017</td>
<td>0.000</td>
</tr>
<tr>
<td>Age$^2$</td>
<td>-0.0004</td>
<td>0.0002</td>
<td>0.009</td>
</tr>
<tr>
<td>High School Grad</td>
<td>1.168</td>
<td>0.178</td>
<td>0.000</td>
</tr>
<tr>
<td>College Grad</td>
<td>1.085</td>
<td>0.131</td>
<td>0.000</td>
</tr>
<tr>
<td>Constant</td>
<td>-3.05</td>
<td>0.418</td>
<td>0.000</td>
</tr>
</tbody>
</table>
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
Run your model as normal. Treat the output as an intermediate step.

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

Calculate or simulate the uncertainty in these final quantities of interest
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
We want to know the behavior of $\mathbb{E}(y|x)$ as we vary $x$.

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:

$x_1 = a, \quad x_2 = b, \quad x_3 = c, \ldots$

- 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 $x$, such as

$$\mathbb{E}(y|x_{\text{scen2}} - y|x_{\text{scen1}}) \quad \text{or} \quad \mathbb{E}(y|x_{\text{scen2}} / y|x_{\text{scen1}})$$
Calculating quantities of interest

Our goal to obtain “quantities of interest,” like

- **Expected Values:** $\mathbb{E}(y|x_c)$
- **Differences:** $\mathbb{E}(y|x_{c2} - y|x_{c1})$
- **Risk Ratios:** $\mathbb{E}(y|x_{c2} / y|x_{c1})$
- or any other function of the above

for some counterfactual $x_c$’s.

For our Voting example, that’s easy – just plug $x_c$ into

$$\mathbb{E}(y|x_c) = \frac{1}{1 + \exp(-x_c \beta)}$$
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?
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)
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×/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
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×/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
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
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.

<table>
<thead>
<tr>
<th>Initial Composition</th>
<th>Hypothetical New Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>FinExp</td>
<td>0.1</td>
</tr>
<tr>
<td>GovExp</td>
<td>0.3</td>
</tr>
<tr>
<td>FMExp</td>
<td>0.1</td>
</tr>
<tr>
<td>CBExp</td>
<td>0.2</td>
</tr>
<tr>
<td>EcoExp</td>
<td>0.3</td>
</tr>
<tr>
<td>Sum</td>
<td>1.0</td>
</tr>
</tbody>
</table>

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 $\beta$ for a component as the “effect” of raising that component.
### American Interest Rate Policy

<table>
<thead>
<tr>
<th>Initial Composition</th>
<th>Hypothetical New Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>FinExp 0.1</td>
<td>$\Delta$FinExp 0.250</td>
</tr>
<tr>
<td>GovExp 0.3</td>
<td>0.15</td>
</tr>
<tr>
<td>FMExp 0.1</td>
<td>0.100</td>
</tr>
<tr>
<td>CBExp 0.2</td>
<td>0.200</td>
</tr>
<tr>
<td>EcoExp 0.3</td>
<td>0.300</td>
</tr>
<tr>
<td>Sum 1.0</td>
<td>1.150</td>
</tr>
</tbody>
</table>

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!
American Interest Rate Policy

<table>
<thead>
<tr>
<th></th>
<th>Initial Composition</th>
<th>Hypothetical New Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>FinExp</td>
<td>0.1</td>
<td>$\Delta$FinExp 0.250</td>
</tr>
<tr>
<td>GovExp</td>
<td>0.3</td>
<td>$=0.15$ 0.300</td>
</tr>
<tr>
<td>FMExp</td>
<td>0.1</td>
<td>$\rightarrow$ 0.100</td>
</tr>
<tr>
<td>CBExp</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>EcoExp</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>Sum</td>
<td>1.0</td>
<td>1.000</td>
</tr>
</tbody>
</table>

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!
### American Interest Rate Policy

<table>
<thead>
<tr>
<th></th>
<th>Initial Composition</th>
<th>Hypothetical New Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>FinExp</td>
<td>0.1</td>
<td>ΔFinExp 0.250</td>
</tr>
<tr>
<td>GovExp</td>
<td>0.3</td>
<td>0.15 0.300</td>
</tr>
<tr>
<td>FMExp</td>
<td>0.1</td>
<td>0.100</td>
</tr>
<tr>
<td>CBExp</td>
<td>0.4</td>
<td>0.400</td>
</tr>
<tr>
<td>EcoExp</td>
<td>0.1</td>
<td>-0.050</td>
</tr>
<tr>
<td>Sum</td>
<td>1.0</td>
<td>1.000</td>
</tr>
</tbody>
</table>

And what if EcoExp (still the reference category) starts out smaller than 0.15? Then our counterfactual would create negative career components!
### American Interest Rate Policy

#### Initial Hypothetical Composition

<table>
<thead>
<tr>
<th>Exp</th>
<th>Initial Composition</th>
<th>Hypothetical New Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>FinExp</td>
<td>0.1</td>
<td>ΔFinExp 0.250</td>
</tr>
<tr>
<td>GovExp</td>
<td>0.3</td>
<td>=0.15</td>
</tr>
<tr>
<td>FMExp</td>
<td>0.1</td>
<td>→</td>
</tr>
<tr>
<td>CBExp</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>EcoExp</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>Sum</td>
<td>1.0</td>
<td>1.000</td>
</tr>
</tbody>
</table>

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.
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 $\beta$s for the composition

<table>
<thead>
<tr>
<th>Composition</th>
<th>Initial</th>
<th>Hypothetical New Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>FinExp</td>
<td>0.1</td>
<td>$\Delta$FinExp</td>
</tr>
<tr>
<td>GovExp</td>
<td>0.3</td>
<td>$\rightarrow$0.15</td>
</tr>
<tr>
<td>FMExp</td>
<td>0.1</td>
<td>$\rightarrow$</td>
</tr>
<tr>
<td>CBExp</td>
<td>0.2</td>
<td>$\rightarrow$</td>
</tr>
<tr>
<td>EcoExp</td>
<td>0.3</td>
<td>$\rightarrow$</td>
</tr>
<tr>
<td>Sum</td>
<td>1.0</td>
<td>1.000</td>
</tr>
</tbody>
</table>
We’ll fit an ordered probit model to the interest rate data:

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

where \( \Phi \) is the Normal CDF and \( \tau \) 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)
Running the model yields the following estimates:

<table>
<thead>
<tr>
<th>EVs</th>
<th>param.</th>
<th>s.e.</th>
<th>EVs</th>
<th>param.</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>FinExp</td>
<td>-0.021</td>
<td>(0.146)</td>
<td>E(Inflation)</td>
<td>0.019</td>
<td>(0.015)</td>
</tr>
<tr>
<td>GovExp</td>
<td>-0.753</td>
<td>(0.188)</td>
<td>E(Unemployment)</td>
<td>-0.035</td>
<td>(0.022)</td>
</tr>
<tr>
<td>FMExp</td>
<td>-1.039</td>
<td>(0.324)</td>
<td>In-Party, election year</td>
<td>-0.182</td>
<td>(0.103)</td>
</tr>
<tr>
<td>CBExp</td>
<td>-0.142</td>
<td>(0.141)</td>
<td>Republican</td>
<td>-0.485</td>
<td>(0.102)</td>
</tr>
<tr>
<td>EcoExp × Repub</td>
<td>0.934</td>
<td>(0.281)</td>
<td>Constant</td>
<td>2.490</td>
<td>(0.148)</td>
</tr>
<tr>
<td>EcoExp × Dem</td>
<td>-0.826</td>
<td>(0.202)</td>
<td>Cutpoint ($\tau$)</td>
<td>3.745</td>
<td>(0.067)</td>
</tr>
</tbody>
</table>

| 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 $XXXExp$ 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 ($\beta$ and $\tau$), but estimates of $\Pr(y_j|x_c, \beta, \tau)$ for hypothetical levels of the EVs $x_c$, which I plot in Figure 1.
### American Interest Rate Policy

#### Table 1: Problematic presentation: FOMC member dissenting votes—Ordered probit parameters.

<table>
<thead>
<tr>
<th>EVs</th>
<th>param.</th>
<th>s.e.</th>
<th>EVs</th>
<th>param.</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>FinExp</td>
<td>−0.021</td>
<td>(0.146)</td>
<td>E(Inflation)</td>
<td>0.019</td>
<td>(0.015)</td>
</tr>
<tr>
<td>GovExp</td>
<td>−0.753</td>
<td>(0.188)</td>
<td>E(Unemployment)</td>
<td>−0.035</td>
<td>(0.022)</td>
</tr>
<tr>
<td>FMExp</td>
<td>−1.039</td>
<td>(0.324)</td>
<td>In-Party, election year</td>
<td>−0.182</td>
<td>(0.103)</td>
</tr>
<tr>
<td>CBExp</td>
<td>−0.142</td>
<td>(0.141)</td>
<td>Republican</td>
<td>−0.485</td>
<td>(0.102)</td>
</tr>
<tr>
<td>EcoExp × Repub</td>
<td>0.934</td>
<td>(0.281)</td>
<td>Constant</td>
<td>2.490</td>
<td>(0.148)</td>
</tr>
<tr>
<td>EcoExp × Dem</td>
<td>−0.826</td>
<td>(0.202)</td>
<td>Cutpoint (τ)</td>
<td>3.745</td>
<td>(0.067)</td>
</tr>
</tbody>
</table>

| N                 | 2957   |        | ln likelihood     | −871.68|        |

How do we interpret these results?
Response variable: FOMC Votes (1 = ease, 2 = accept, 3 = tighten)

<table>
<thead>
<tr>
<th>EVs</th>
<th>param.</th>
<th>s.e.</th>
<th>EVs</th>
<th>param.</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>FinExp</td>
<td>-0.021</td>
<td>(0.146)</td>
<td>E(Inflation)</td>
<td>0.019</td>
<td>(0.015)</td>
</tr>
<tr>
<td>GovExp</td>
<td>-0.753</td>
<td>(0.188)</td>
<td>E(Unemployment)</td>
<td>-0.035</td>
<td>(0.022)</td>
</tr>
<tr>
<td>FMExp</td>
<td>-1.039</td>
<td>(0.324)</td>
<td>In-Party, election year</td>
<td>-0.182</td>
<td>(0.103)</td>
</tr>
<tr>
<td>CBExp</td>
<td>-0.142</td>
<td>(0.141)</td>
<td>Republican</td>
<td>-0.485</td>
<td>(0.102)</td>
</tr>
<tr>
<td>EcoExp × Repub</td>
<td>0.934</td>
<td>(0.281)</td>
<td>Constant</td>
<td>2.490</td>
<td>(0.148)</td>
</tr>
<tr>
<td>EcoExp × Dem</td>
<td>-0.826</td>
<td>(0.202)</td>
<td>Cutpoint (τ)</td>
<td>3.745</td>
<td>(0.067)</td>
</tr>
</tbody>
</table>

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
Looking at this table, two obvious questions 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?
Response variable: FOMC Votes (1 = ease, 2 = accept, 3 = tighten)

<table>
<thead>
<tr>
<th>EVs</th>
<th>param.</th>
<th>s.e.</th>
<th>EVs</th>
<th>param.</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>FinExp</td>
<td>-0.021</td>
<td>(0.146)</td>
<td>E(Inflation)</td>
<td>0.019</td>
<td>(0.015)</td>
</tr>
<tr>
<td>GovExp</td>
<td>-0.753</td>
<td>(0.188)</td>
<td>E(Unemployment)</td>
<td>-0.035</td>
<td>(0.022)</td>
</tr>
<tr>
<td>FMExp</td>
<td>-1.039</td>
<td>(0.324)</td>
<td>In-Party, election year</td>
<td>-0.182</td>
<td>(0.103)</td>
</tr>
<tr>
<td>CBExp</td>
<td>-0.142</td>
<td>(0.141)</td>
<td>Republican</td>
<td>-0.485</td>
<td>(0.102)</td>
</tr>
<tr>
<td>EcoExp × Repub</td>
<td>0.934</td>
<td>(0.281)</td>
<td>Constant</td>
<td>2.490</td>
<td>(0.148)</td>
</tr>
<tr>
<td>EcoExp × Dem</td>
<td>-0.826</td>
<td>(0.202)</td>
<td>Cutpoint (τ)</td>
<td>3.745</td>
<td>(0.067)</td>
</tr>
</tbody>
</table>

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 dissent 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.”

<table>
<thead>
<tr>
<th>Response to an Increase in …</th>
<th>Probability of hawkish dissent</th>
</tr>
</thead>
<tbody>
<tr>
<td>FMExp</td>
<td>0.1% 0.4% 0.8% 2% 4% 8% 20% 40%</td>
</tr>
<tr>
<td>GovExp</td>
<td></td>
</tr>
</tbody>
</table>
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.
Response to an Increase in …

<table>
<thead>
<tr>
<th>Probability of hawkish dissent</th>
<th>Probability of dovish dissent</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1% 0.4% 0.8% 2% 4% 8% 20% 40%</td>
<td>0.1% 0.4% 0.8% 2% 4% 8% 20% 40%</td>
</tr>
</tbody>
</table>

- FMExp
- GovExp
- EcoExp × Dem
- Republican
- In-Party & Election
- E(Unemployment)
- E(Inflation)
- CBExp
- FinExp
- EcoExp × Repub

Change in \( P(\text{hawkish dissent}) \)  
Change in \( P(\text{dovish dissent}) \)

Note that I left out the results for dovish dissent in my presentation. Here they are symmetrical, but with 4 or 5 outcome categories, they may not be
Response to an Increase in …

<table>
<thead>
<tr>
<th>Probability of hawkish dissent</th>
<th>Probability of dovish dissent</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1% 0.4% 0.8% 2% 4% 8% 20% 40%</td>
<td>0.1% 0.4% 0.8% 2% 4% 8% 20% 40%</td>
</tr>
</tbody>
</table>

- **FMExp**
- **GovExp**
- **EcoExp \times Dem**
- **Republican**
- **In-Party & Election**
- **E(Unemployment)**
- **E(Inflation)**
- **CBExp**
- **FinExp**
- **EcoExp \times Repub**

<table>
<thead>
<tr>
<th>Change in P(hawkish dissent)</th>
<th>Change in P(dovish dissent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.03x 0.1x 0.2x 0.5x 1x 2x 5x 10x</td>
<td>0.03x 0.1x 0.2x 0.5x 1x 2x 5x 10x</td>
</tr>
</tbody>
</table>

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!
Response to an Increase in …

<table>
<thead>
<tr>
<th>Probability of hawkish dissent</th>
<th>Probability of dovish dissent</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1% 0.4% 0.8% 2% 4% 8% 20% 40%</td>
<td>0.1% 0.4% 0.8% 2% 4% 8% 20% 40%</td>
</tr>
</tbody>
</table>

**FMExp**

**GovExp**

**EcoExp × Dem**

Republican

In-Party & Election

E(Unemployment)

E(Inflation)

CBExp

FinExp

EcoExp × Repub

Change in P(hawkish dissent)

0.03x 0.1x 0.2x 0.5x 1x 2x 5x 10x

Change in P(dovish dissent)

0.03x 0.1x 0.2x 0.5x 1x 2x 5x 10x

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

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

Neutral career types † (Economics, Business)

Inflation-increasing career types * (Government, Central Bank Staff)

Career Index (CBCC) ‡

Central Bank Independence‡
cbi-c = Cukierman index
cbi-3 = avg of 3 indexes

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

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

Neutral career types† 
(Economics, Business)

Inflation-increasing career types* 
(Government, Central Bank Staff)

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

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

Inflation-increasing
career types*  
(Government, Central Bank Staff)

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.
Comparative Inflation Performance

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

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

+1 sd FinExp

0.0 2.5 5.0
Years after ...

+1 sd FMExp

0.0 2.5 5.0
Years after ...

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.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Expected Sign</th>
<th>DV: ln(Inflation)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sign</td>
<td>1</td>
</tr>
<tr>
<td>FinExp_{j,t-2}</td>
<td>-</td>
<td>-0.14</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td></td>
</tr>
<tr>
<td>FMExp_{j,t-2}</td>
<td>-/+</td>
<td>-0.08</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td></td>
</tr>
<tr>
<td>CBExp_{j,t-2}</td>
<td>+/-</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td></td>
</tr>
<tr>
<td>GovExp_{j,t-2}</td>
<td>+</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td></td>
</tr>
<tr>
<td>CBI_{j,t-2}</td>
<td>-</td>
<td>-0.91</td>
</tr>
<tr>
<td></td>
<td>(0.30)</td>
<td>(0.29)</td>
</tr>
<tr>
<td>CBCC_{j,t-2}^{med}</td>
<td>-</td>
<td>-0.09</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>CBI_{j,t-2} × CBCC_{j,t-2}^{med}</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Imports/GDP)_{j,t-2}</td>
<td>-</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td>(0.26)</td>
<td>(0.25)</td>
</tr>
<tr>
<td>%EcDegree_{j,t-2}</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln (\pi_{j,t-1})</td>
<td></td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>ln (\pi_{j,t-2})</td>
<td></td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.04)</td>
</tr>
</tbody>
</table>

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

Career Index (CBCC)†

| Financial Experience (FinExp) | Financial Ministry Experience (FMExp) |
| GovExp | CB Staff Experience (CBExp) |

Central Bank Independence‡

cbi-c = Cukierman index
cbi-3 = avg of 3 indexes

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 lineplots 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 \textit{R} package that

1. takes as input the \textit{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 \textit{R}’s basic graphics to do this badly
- or get \textit{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
An example tile layout, minus traces
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) → ≈ 2 parties.
  - Proportional representation → 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 of 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?
Next, we can try to figure out what makes the US and Switzerland so different. 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.
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.
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
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
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:** \texttt{linesTile()}, \texttt{pointsile()}, \texttt{polygonTile()}, \texttt{polylinesTile()}, \texttt{and} \texttt{textTile()}
   - **Complex traces:** \texttt{lineplot()}, \texttt{scatter()}, \texttt{ropeladder()}, \texttt{and} \texttt{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.

\[ \text{Pr(Voting)} \text{ given five distinct sets of } x's \]

*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.

\[ \Pr(\text{Voting}) \text{ given five distinct sets of } x's \]

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

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

Hold all but Age constant, then calculate \( \Pr(\text{Voting}) \) 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?
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
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 \( \text{Pr(Vote)} \)
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

<table>
<thead>
<tr>
<th></th>
<th>M1</th>
<th>M2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.075 (0.017)</td>
<td>0.061 (0.017)</td>
</tr>
<tr>
<td>Age²</td>
<td>-0.0004 (0.0002)</td>
<td>-0.0003 (0.0002)</td>
</tr>
<tr>
<td>High School Grad</td>
<td>1.124 (0.180)</td>
<td>1.099 (0.181)</td>
</tr>
<tr>
<td>College Grad</td>
<td>1.080 (0.131)</td>
<td>1.053 (0.132)</td>
</tr>
<tr>
<td>Married</td>
<td></td>
<td>0.373 (0.110)</td>
</tr>
<tr>
<td>Constant</td>
<td>-3.019 (0.418)</td>
<td>-2.866 (0.421)</td>
</tr>
<tr>
<td>log likelihood</td>
<td>-1101.370</td>
<td>-1099.283</td>
</tr>
<tr>
<td>N</td>
<td>1783</td>
<td>1783</td>
</tr>
</tbody>
</table>
Comparing Logistic Regression Models

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

Model 1 (baseline specification)

Model 2 (control for Married)

Logit estimates: 95% confidence interval is shaded

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?

Logit estimates:
95% confidence interval is shaded

Currently Married
Not Married

Probability of Voting

Age of Respondent

Chris Adolph (University of Washington)
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
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 RRls 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 lineplots 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 \{Conservative, Liberal, 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

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:

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

I estimate a zero-inflated
compositional
regression...
Key predictor:
Partisanship of government (PCoG): higher values = more conservative

My claim:

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

I estimate a zero-inflated compositional regression...

---

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

<table>
<thead>
<tr>
<th>Response</th>
<th>Covariates E(sign)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Constant</td>
<td>1.142^{0.124}</td>
<td>1.142^{0.124}</td>
<td>1.308^{a}</td>
<td>−0.139^{0.389}</td>
</tr>
<tr>
<td></td>
<td>PCoG</td>
<td>−1.487^{0.343}</td>
<td>−1.487^{0.342}</td>
<td>−0.691^{0.486}</td>
<td>−1.770^{0.371}</td>
</tr>
<tr>
<td></td>
<td>ConExp_{pre}</td>
<td></td>
<td></td>
<td></td>
<td>1.520^{0.678}</td>
</tr>
<tr>
<td></td>
<td>LibExp_{pre}</td>
<td></td>
<td></td>
<td></td>
<td>1.980^{0.596}</td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>−0.242^{0.101}</td>
<td>−0.242^{0.101}</td>
<td>−0.440^{a}</td>
<td>−1.050^{0.377}</td>
</tr>
<tr>
<td></td>
<td>PCoG</td>
<td>0.171^{0.288}</td>
<td>0.170^{0.280}</td>
<td>0.648^{0.444}</td>
<td>−0.208^{0.322}</td>
</tr>
<tr>
<td></td>
<td>ConExp_{pre}</td>
<td></td>
<td></td>
<td></td>
<td>2.350^{0.645}</td>
</tr>
<tr>
<td></td>
<td>LibExp_{pre}</td>
<td></td>
<td></td>
<td></td>
<td>0.573^{0.539}</td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>0.482^{0.105}</td>
<td>0.482^{0.105}</td>
<td>0.452^{a}</td>
<td>1.710^{0.402}</td>
</tr>
<tr>
<td></td>
<td>PCoG</td>
<td>−0.662^{0.302}</td>
<td>−0.662^{0.302}</td>
<td>−0.163^{0.434}</td>
<td>−0.461^{0.327}</td>
</tr>
<tr>
<td></td>
<td>ConExp_{pre}</td>
<td></td>
<td></td>
<td></td>
<td>−1.960^{0.653}</td>
</tr>
<tr>
<td></td>
<td>LibExp_{pre}</td>
<td></td>
<td></td>
<td></td>
<td>−1.500^{0.564}</td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>0.415^{0.128}</td>
<td>0.381^{0.124}</td>
<td>0.497^{a}</td>
<td>−0.443^{0.440}</td>
</tr>
<tr>
<td></td>
<td>PCoG</td>
<td>−0.390^{0.419}</td>
<td>−0.252^{0.419}</td>
<td>−0.561^{0.482}</td>
<td>−0.147^{0.414}</td>
</tr>
<tr>
<td></td>
<td>ConExp_{pre}</td>
<td></td>
<td></td>
<td></td>
<td>0.314^{0.775}</td>
</tr>
<tr>
<td></td>
<td>LibExp_{pre}</td>
<td></td>
<td></td>
<td></td>
<td>1.470^{0.612}</td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>−0.111^{0.160}</td>
<td>−0.112^{0.147}</td>
<td>0.085^{a}</td>
<td>−0.152^{0.498}</td>
</tr>
<tr>
<td></td>
<td>PCoG</td>
<td>0.557^{0.471}</td>
<td>0.491^{0.446}</td>
<td>0.057^{0.495}</td>
<td>0.546^{0.445}</td>
</tr>
<tr>
<td></td>
<td>ConExp_{pre}</td>
<td></td>
<td></td>
<td></td>
<td>0.071^{0.818}</td>
</tr>
<tr>
<td></td>
<td>LibExp_{pre}</td>
<td></td>
<td></td>
<td></td>
<td>0.007^{0.722}</td>
</tr>
</tbody>
</table>

Composition Model

<table>
<thead>
<tr>
<th>Notes</th>
<th>Normal</th>
<th>Student’s t</th>
<th>Student’s t</th>
<th>Student’s t</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>a, b</td>
<td>b</td>
<td>b</td>
</tr>
<tr>
<td>N</td>
<td>411</td>
<td>411</td>
<td>391</td>
<td></td>
</tr>
<tr>
<td>ln likelihood</td>
<td>−1414.82</td>
<td>−1066.29</td>
<td>−962.84</td>
<td>−985.80</td>
</tr>
<tr>
<td>p-value of LR test against model lacking</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>PCoG</td>
<td>t-dist</td>
<td>f.e.</td>
<td>prev exp</td>
<td></td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>My variable of interest, $X_1$</td>
<td>X.XX</td>
<td>X.XX</td>
<td>X.XX</td>
<td>X.XX</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(X.XX)</td>
<td>(X.XX)</td>
<td>(X.XX)</td>
<td>(X.XX)</td>
<td></td>
</tr>
<tr>
<td>A control</td>
<td>X.XX</td>
<td>X.XX</td>
<td>X.XX</td>
<td>X.XX</td>
<td>X.XX</td>
</tr>
<tr>
<td>I &quot;need&quot;</td>
<td>(X.XX)</td>
<td>(X.XX)</td>
<td>(X.XX)</td>
<td>(X.XX)</td>
<td>(X.XX)</td>
</tr>
<tr>
<td>A control</td>
<td>X.XX</td>
<td>X.XX</td>
<td>X.XX</td>
<td>X.XX</td>
<td>X.XX</td>
</tr>
<tr>
<td>I &quot;need&quot;</td>
<td>(X.XX)</td>
<td>(X.XX)</td>
<td>(X.XX)</td>
<td>(X.XX)</td>
<td>(X.XX)</td>
</tr>
<tr>
<td>A candidate control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alternate measure of $X_1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X.XX</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(X.XX)</td>
</tr>
</tbody>
</table>
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 $E(Y|X)$, 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 $E(Y|X)$, 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 $X$ and a concept $Y$.

2. Choose:
   - a measure of $X$, denoted $X$,
   - a measure of $Y$, denoted $Y$,
   - a set of confounders, $Z$,
   - a functional form, $g(\cdot)$,
   - a probability model of $Y$, $f(\cdot)$.

3. Estimate the probability model $Y \sim f(\mu, \alpha), \mu = g(\text{vec}(X, Z), \beta)$.

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

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

$$CBCC = \text{Conservative Experience} - \text{Liberal Experience}$$

And use my model to predict changes in CBCC and plot them on a ropeladder.
Robustness Ropeladder: Partisan central banker appointment

Estimated increase in Central Bank Conservatism (CBCC) resulting from …

<table>
<thead>
<tr>
<th>Control added</th>
<th>Shifting control from low ($\mu$-1.5 sd) to high ($\mu$+1.5)</th>
<th>Shifting PCoG from From left gov ($\mu$-1.5 sd) to right gov ($\mu$+1.5 sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[None]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Office appointed to</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBI (3 index avg.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBI (Cukierman)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lagged inflation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lagged unemployment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trade openness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Endebtedness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Financial Sector Employment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Financial Sector Score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time trend</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central bank staff size</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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.
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?

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

Neutral career types †
(Economics, Business)

Inflation-increasing career types *
(Government, Central Bank Staff)

Recall the TSCS model of inflation performance

How would we show robustness here?
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.
Robustness for several QoIs at once

Specification

Baseline
Robust Estimation
Add $\pi_{\text{world}}$
Use Cukierman CBI
Omit Imports/GDP
Add Exchange Regime
Add % Left-Appointees
Add Partisan CoG of Aptees
Add % with Econ PhDs

Change in inflation, five years after $+1$ s.d. in ...

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

Specification

Baseline
Robust Estimation
Add $\pi_{\text{world}}$
Use Cukierman CBI
Omit Imports/GDP
Add Exchange Regime
Add % Left-Appointees
Add Partisan CoG of Aptees
Add % with Econ PhDs

Change in inflation, five years after +1 s.d. in ...

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.

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.
Expected percentile

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
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!
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 →
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?
Increase CBCC by +1 sd, given...

Age of Central Banker
- 65 years
- 45 years
- diff-in-diffs

Future Job Matches Policy
- yes
- no
- diff-in-diffs

Monetary Policy Votes
- public
- secret
- diff-in-diffs

Future Jobs and Public Votes
- both
- neither
- diff-in-diffs

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.

Chris Adolph (University of Washington)
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

---

Incr ease CBC C by +1 sd, given...

- Age of Central Banker: 65 years, 45 years, both
- Future Job Matches Policy: yes, no, both
- Monetary Policy Votes: public, secret, neither
- Future Jobs and Public Votes: both, neither, both
Cumulative change in inflation after 5 years

Increase CBCC by +1 sd, given...

Age of Central Banker
- 65 years
- 45 years
- diff-in-diffs

Future Job Matches Policy
- yes
- no
- diff-in-diffs

Monetary Policy Votes
- public
- secret
- diff-in-diffs

Future Jobs and Public Votes
- both
- neither
- diff-in-diffs

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

Cumulative change in inflation after 5 years

Increase CBCC by +1 sd, given...

Age of Central Banker
- 65 years
- 45 years
diff-in-diffs

Future Job Matches Policy
- yes
diff-in-diffs
- no
diff-in-diffs

Monetary Policy Votes
- public
diff-in-diffs
- secret
diff-in-diffs

Future Jobs and Public Votes
- both
diff-in-diffs
- neither
diff-in-diffs

Chris Adolph (University of Washington)
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 × 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 × 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.*

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Hazard ratio</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age &gt; 75</td>
<td>5.78</td>
<td>2.28 – 14.68</td>
</tr>
<tr>
<td>70 &lt; Age ≤ 75</td>
<td>3.48</td>
<td>2.32 – 5.22</td>
</tr>
<tr>
<td>65 &lt; Age ≤ 70</td>
<td>2.01</td>
<td>1.24 – 3.27</td>
</tr>
<tr>
<td>Other Government Experience</td>
<td>1.86</td>
<td>0.82 – 4.23</td>
</tr>
<tr>
<td>Abs diff in PCoG, appt party vs. current</td>
<td>1.67</td>
<td>1.24 – 2.25</td>
</tr>
<tr>
<td>Financial Experience</td>
<td>1.40</td>
<td>0.83 – 2.38</td>
</tr>
<tr>
<td>Finance Ministry Experience</td>
<td>1.34</td>
<td>0.71 – 2.52</td>
</tr>
<tr>
<td>Current PCoG × Inflation</td>
<td>1.05</td>
<td>1.00 – 1.11</td>
</tr>
<tr>
<td>Unemployment</td>
<td>1.04</td>
<td>1.00 – 1.08</td>
</tr>
<tr>
<td>Inflation</td>
<td>1.04</td>
<td>1.01 – 1.07</td>
</tr>
<tr>
<td>Current PCoG × Unemployment</td>
<td>0.95</td>
<td>0.89 – 1.02</td>
</tr>
<tr>
<td>Central Bank Staff Experience</td>
<td>0.90</td>
<td>0.62 – 1.30</td>
</tr>
<tr>
<td>Economics Experience</td>
<td>0.87</td>
<td>0.52 – 1.43</td>
</tr>
<tr>
<td>Current Partisan Center of Gravity (PCoG)</td>
<td>0.86</td>
<td>0.41 – 1.82</td>
</tr>
</tbody>
</table>

| N             | 10,863       |
|log likelihood | −1229.4      |

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.

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

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

**Areas:** Pharmaceuticals, Secondary/Tertiary, Primary Care, Public Health

**Instruments:** Frameworks, Finance, Implementation, Provision

Each combination for each country is a case.
Adolph, Greer & Fonseca consider explanations of whether local, regional, or state-level European governments have power over specific health policy areas and instruments.

**Areas:** Pharmaceuticals, 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 model.
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         Continuous
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 everything 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.
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
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 \( 1 - p \)
Next consider countries with all three levels, holding covariates at means
Probability of allocation of authority by policy type

If we are clever…

- Public Health
- Primary Care
- Secondary/Tertiary
- Pharmaceuticals
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
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”
Residual country effects

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?

I couldn't see it. My wife did: good skiing and historically, mountains strengthened regional autonomy!

Lesson: Share your diagnostic plots!
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
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

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

Chris Adolph (University of Washington)
### Kitchen sink models of 1960 US crime rates, continued

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

Side-by-side robustness ropeladders –
focus is now on comparisons across models, not variables

Chris Adolph (University of Washington)
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

<table>
<thead>
<tr>
<th>Interaction</th>
<th>Counterfactual Strategy</th>
<th>Plot</th>
</tr>
</thead>
<tbody>
<tr>
<td>discrete with discrete</td>
<td>one cf for each combination of values (full factorial)</td>
<td>ropeladder</td>
</tr>
<tr>
<td>continuous with discrete</td>
<td>choose combinations of representative values or</td>
<td>ropeladder</td>
</tr>
<tr>
<td></td>
<td>combine a continuum with each discrete value</td>
<td>lineplot</td>
</tr>
<tr>
<td>continuous with continuous</td>
<td>choose combinations of represented values or</td>
<td>ropeladder</td>
</tr>
<tr>
<td></td>
<td>combine a continuum with each discrete value or</td>
<td>lineplot</td>
</tr>
<tr>
<td></td>
<td>combine a continuum with a continuum</td>
<td>3D functional boxplots</td>
</tr>
</tbody>
</table>
Let’s convert this to a ropeladder ON THE WHITEBOARD
Continuous $\times$ Discrete Interactions: Ropeladders

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

Incr ease CBC C by $+1$ sd, given...

<table>
<thead>
<tr>
<th></th>
<th>Cumulative change in inflation after 5 years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age of Central Banker</td>
<td>65 years</td>
</tr>
<tr>
<td></td>
<td>45 years</td>
</tr>
<tr>
<td>Future Job Matches Policy</td>
<td>yes</td>
</tr>
<tr>
<td>Monetary Policy Votes</td>
<td>public</td>
</tr>
<tr>
<td></td>
<td>secret</td>
</tr>
<tr>
<td>Future Jobs and Public Votes</td>
<td>both</td>
</tr>
<tr>
<td></td>
<td>neither</td>
</tr>
</tbody>
</table>
Continuous × Discrete Interactions: Ropeladders

Cumulative change in inflation after 5 years

Increase CBCC by +1 sd, given...

Age of Central Banker

Future Job Matches Policy

Monetary Policy Votes

Future Jobs and Public Votes

I selected a specific change in my continuous variable and computed counterfactuals under each value of my discrete variable
Let's return to our voting example, where voting was a function of Age, \( \text{Age}^2 \), Education, and Marriage Status.

Suppose we add the interactions \( \text{Age} \times \text{Married} \) and \( \text{Age}^2 \times \text{Married} \).
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 $\times$ Discrete Interactions: Lineplots

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

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

changes to

```r
model2 <-
vote00 ~ age*marriedo + I(age^2)*marriedo + hsdeg + coldeg
```
Continuous $\times$ 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
The results suggest this interaction wasn’t a great idea…
Continuous $\times$ Continuous Interactions: Ropeladders

We’ve already dealt with a continuous $\times$ continuous interaction:

Central banker tenure depended on:

- **Inflation $\times$ Party CoG**
- **Unemployment $\times$ Party CoG**
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 $\times$ Continuous Interactions: Lineplots

CBNA Measure

Long-run Unemployment Under ... Low CBNA High CBNA

5 Year Difference, Low $\rightarrow$ High CBNA

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()`.
Continuous × Continuous Interactions: Lineplots

CBNA Measure

Long-run Unemployment Under ...
Low CBNA
High CBNA
5 Year Difference,
Low → High CBNA

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

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**

<table>
<thead>
<tr>
<th></th>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>high</td>
<td>high</td>
<td>high</td>
<td>high</td>
</tr>
<tr>
<td>high</td>
<td>high</td>
<td>low</td>
<td></td>
</tr>
<tr>
<td>values</td>
<td>high</td>
<td>low</td>
<td>high</td>
</tr>
<tr>
<td>used</td>
<td>high</td>
<td>low</td>
<td>low</td>
</tr>
<tr>
<td>in</td>
<td>low</td>
<td>high</td>
<td>high</td>
</tr>
<tr>
<td>counterfactuals</td>
<td>low</td>
<td>high</td>
<td>low</td>
</tr>
<tr>
<td></td>
<td>low</td>
<td>low</td>
<td>high</td>
</tr>
<tr>
<td></td>
<td>low</td>
<td>low</td>
<td>low</td>
</tr>
</tbody>
</table>

Descriptive names for these combinations essential for presentation in this case.
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)**

<table>
<thead>
<tr>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>values used in counterfactuals</td>
<td>high</td>
<td>high</td>
</tr>
<tr>
<td></td>
<td>continuum low</td>
<td>high</td>
</tr>
<tr>
<td></td>
<td>low</td>
<td>high</td>
</tr>
<tr>
<td></td>
<td>low</td>
<td>low</td>
</tr>
</tbody>
</table>

Use tiling of lineplots to your advantage in this case.
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.
Strategies for Interactions of 3 Continuous Variables

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.

Chris Adolph (University of Washington)
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?**

<table>
<thead>
<tr>
<th></th>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>values used in continuum counterfactuals</td>
<td>high</td>
<td>low</td>
<td></td>
</tr>
</tbody>
</table>

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