# Communicating using graphics

*March 9, 2021*

# Hello – again!

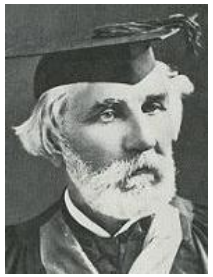
Almost everyone heard lots about graphics in 514, from me. You know how to make beeswarm/violin/scatter plots, and that graphics can aid explanation.

Graphics is a large and active research area – way too big for one class! Today:

- Why communicate visually?
- Principles to apply when making your own
- More on why some graphs work better than others
- Practice at making your own

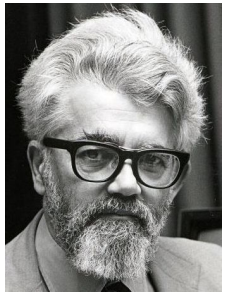
# Why communicate visually? obligatory quotes


*One picture is worth 10,000 words*
Fred Barnard (in a fake Chinese proverb)
*Printer's Ink* 1927



*A picture shows me at a glance what it takes dozens of pages of a book to expound.*
Ivan Turgenev (Russian Novelist)

*Un bon croquis vaut mieux qu'un long discours* – A good sketch is better than a long speech
attr. Napoleon Bonaparte



*1001 words are worth more than a picture*
John McCarthy, computer scientist

# Why communicate visually? efficiency

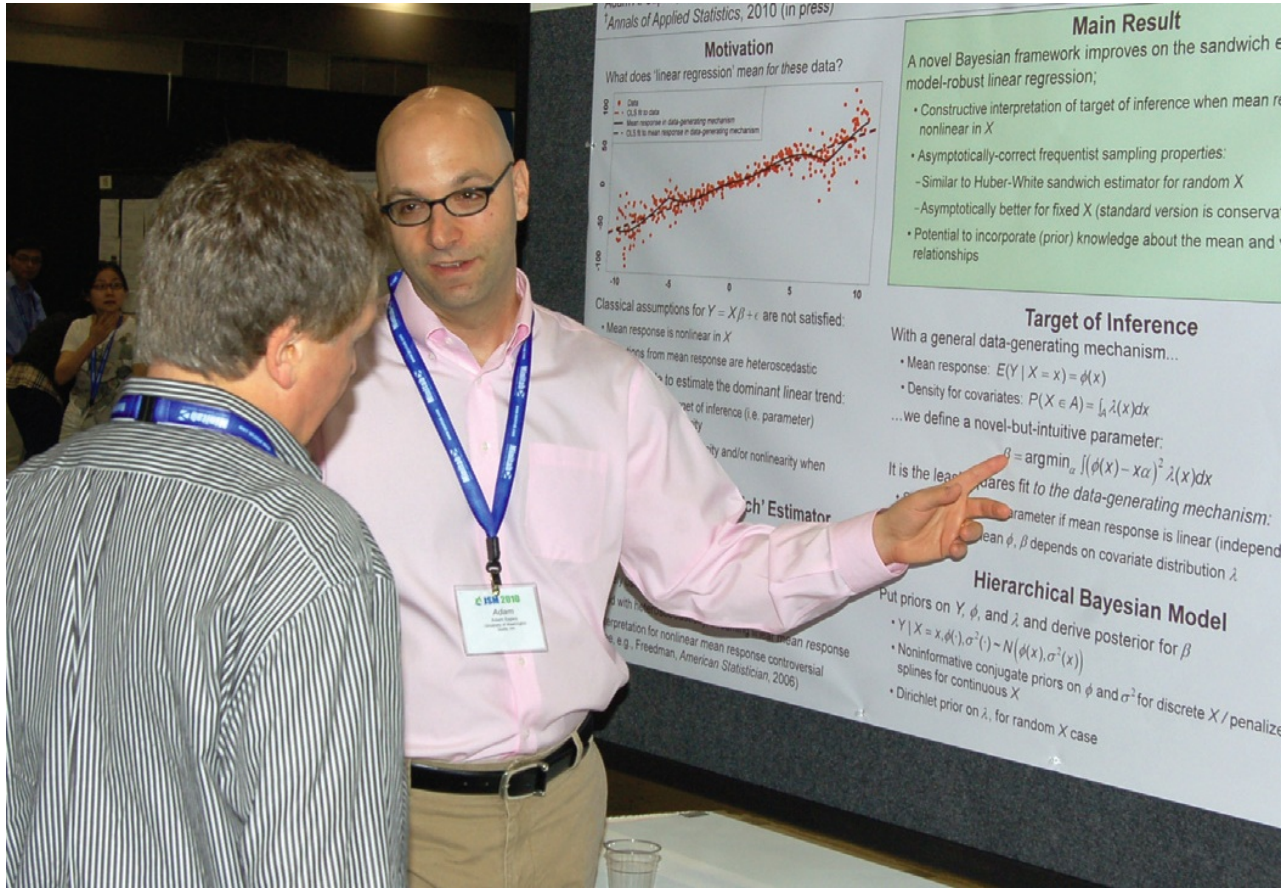
This is a poster session;

**Particularly** at the start of your career, you must present information;

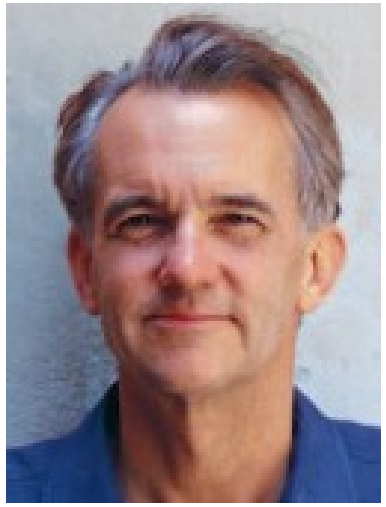
- Comprehensibly – i.e. easily/accurately
- Efficiently – need to keep your audience's attention

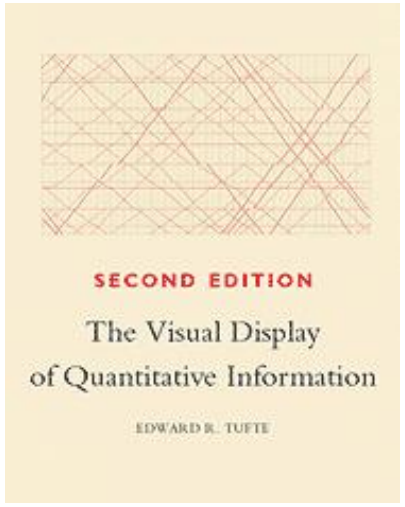
# Why communicate visually? efficiency

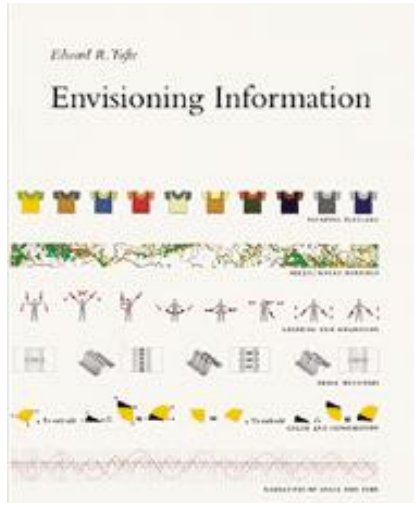
Here's a poster session in Amstat News – where did the audience look first?

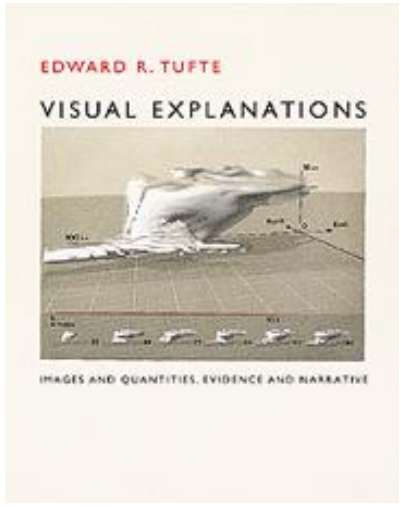


# Principles to apply: from Tufte, and others





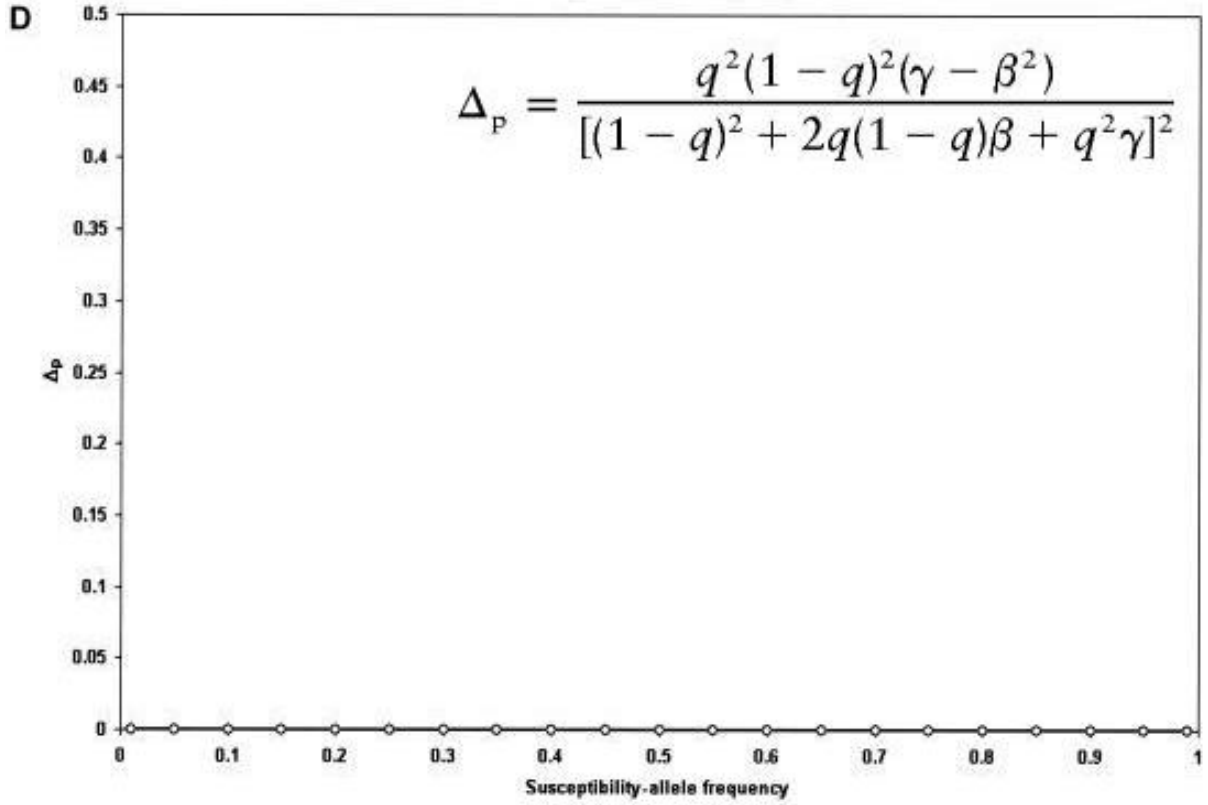




- Serve a reasonably clear purpose
- Show the data
- Avoid distorting what the data have to say
- Encourage the eye to compare different pieces of data

Tufte is great—and easy to read—but rather heuristic. Later work (Stephen Few) adds practical ideas grounded in visual perception research.

# Principles: serve a clear purpose (?)




Wittke-Thompson JK, Pluzhnikov A, Cox NJ (2005) Rational inferences about departures from Hardy-Weinberg equilibrium. *American Journal of Human Genetics* 76:967-986

# Principles: serve a clear purpose

Some **scientific** purposes: (Note close connections to $t$-tests, regression etc)

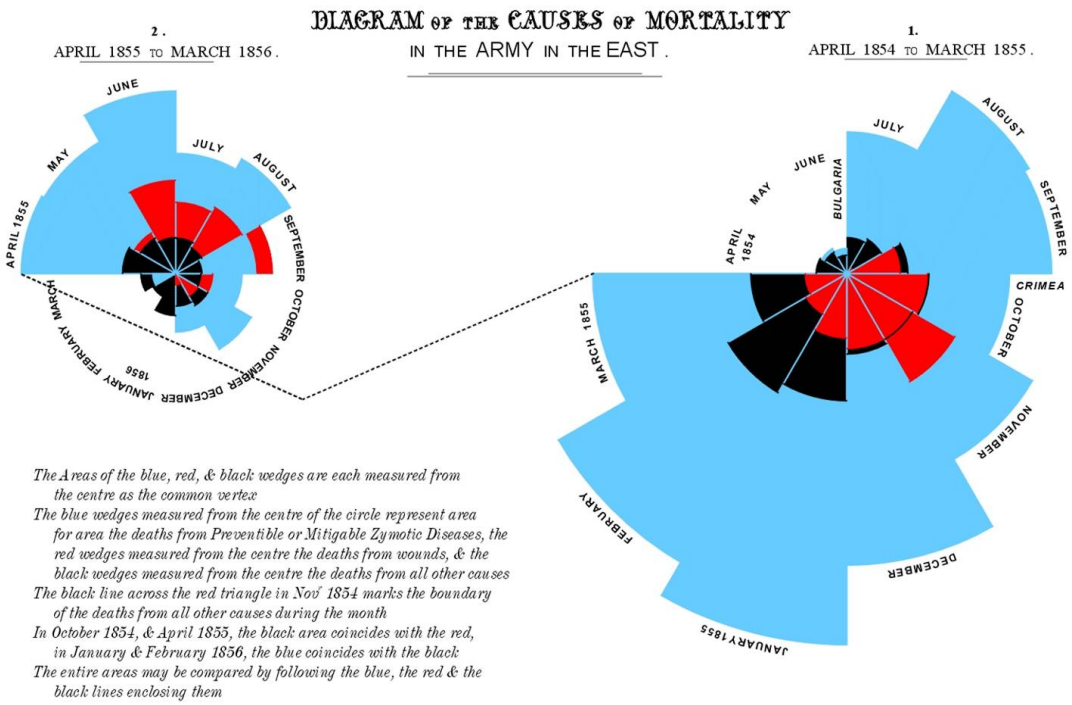
- Histogram/dotchart/beeswarm plot: *summarize* 1D continuous data
- Barchart: *compare* 1D categorical data
- Scatterplots: *show association* of continuous $Y$ and $X$ (or lack of association)
- Mosaic plots: *show association* of categorical $Y$ and $X$ (or lack of association)
- QQ plots: *compare* two continuous distributions; talk about the shift, spread, heavy tails, light tails etc

Recall BIOST 514: these encode value as *position on a common scale* — except mosaic plots, which encode value as area.

In **non-scientific** settings, grabbing the readers' attention may *also* be a goal. This is not 'wrong', but think about whether advice **you** read is for science or sales pitch.

# Principles: serve a clear purpose – example #1

A very early sales pitch, by nurse and statistician Florence Nightingale:

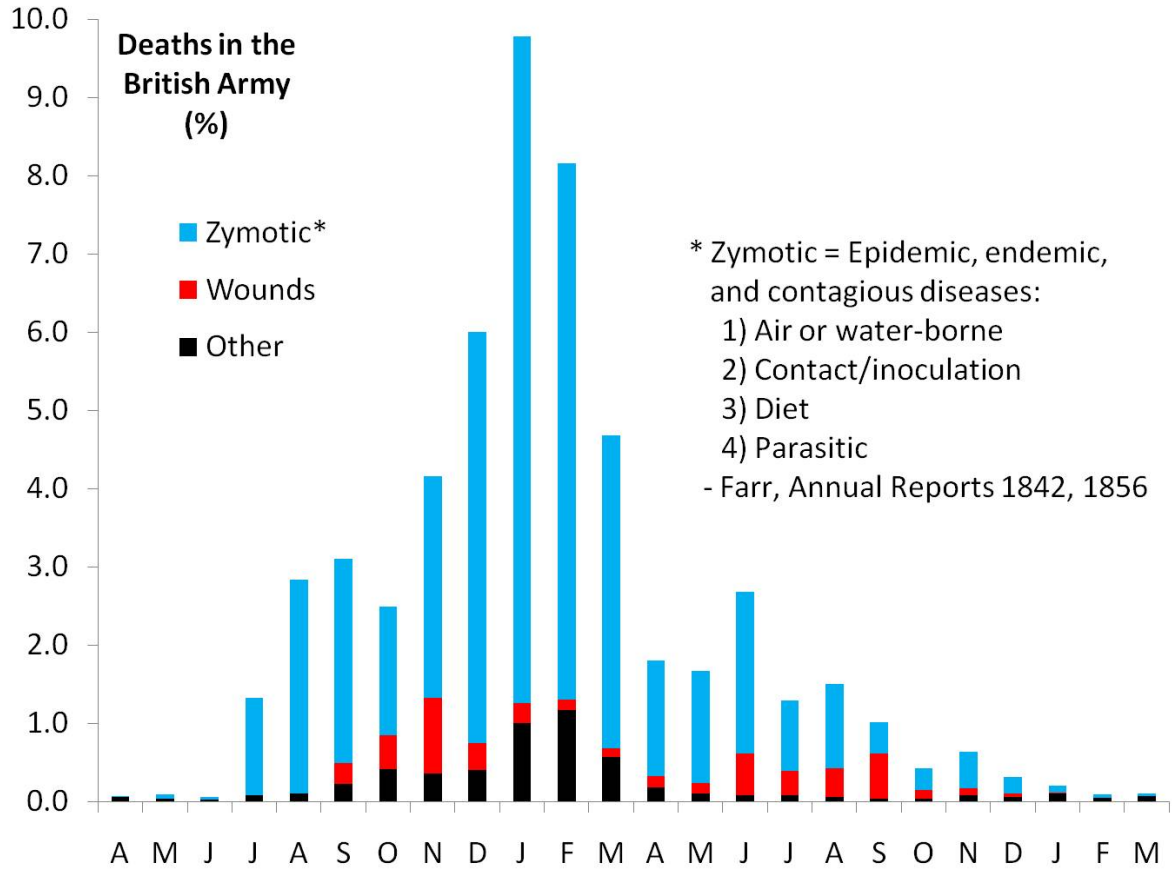



After sanitation changes (early 1856) malnutrition/disease killed far fewer men.

# Principles: serve a clear purpose – example #1

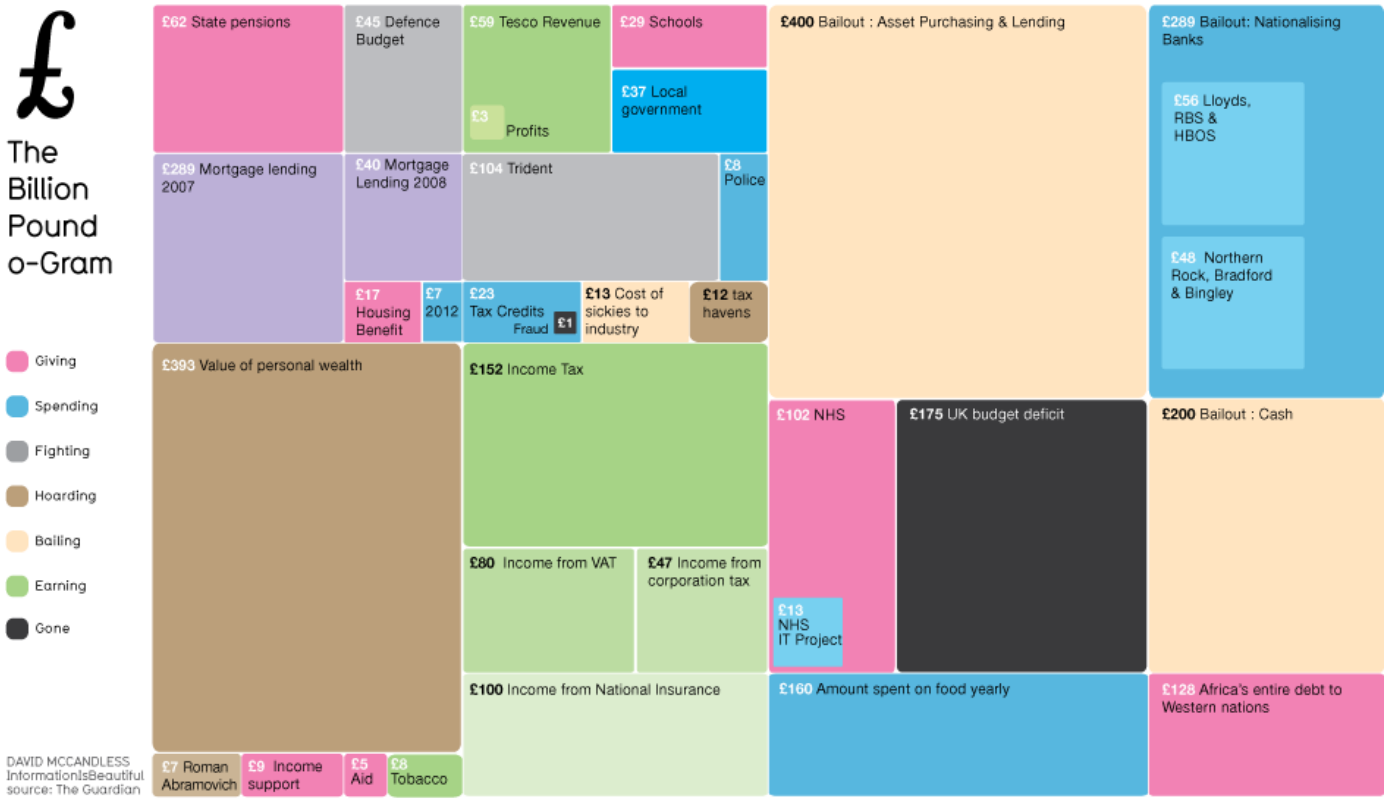
A more default plot:

- Doesn't focus on the 1845/6 difference – Nightingale's main concern, as it was controversial
- Expresses counts as areas, **not** position on a common scale. But the ~square-root transform helps stabilize variance
- Doesn't 'pop', less likely to grab attention



# Principles: serve a clear purpose – example #2

David McCandless* illustrating, for the Guardian, how large numbers compare;

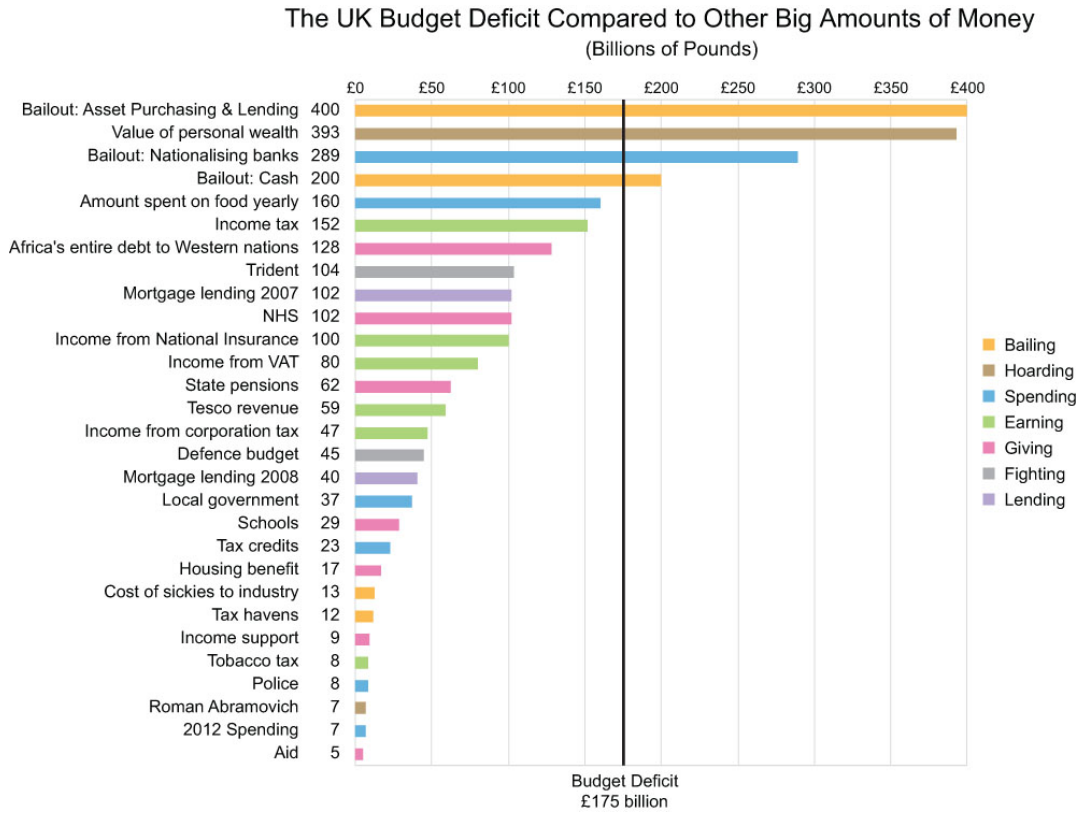



* Writer, designer, 'data journalist', and TED speaker

# Principles: serve a clear purpose – example #2

Large numbers are just numbers—Stephen Few (right) compares them using position on a common scale.

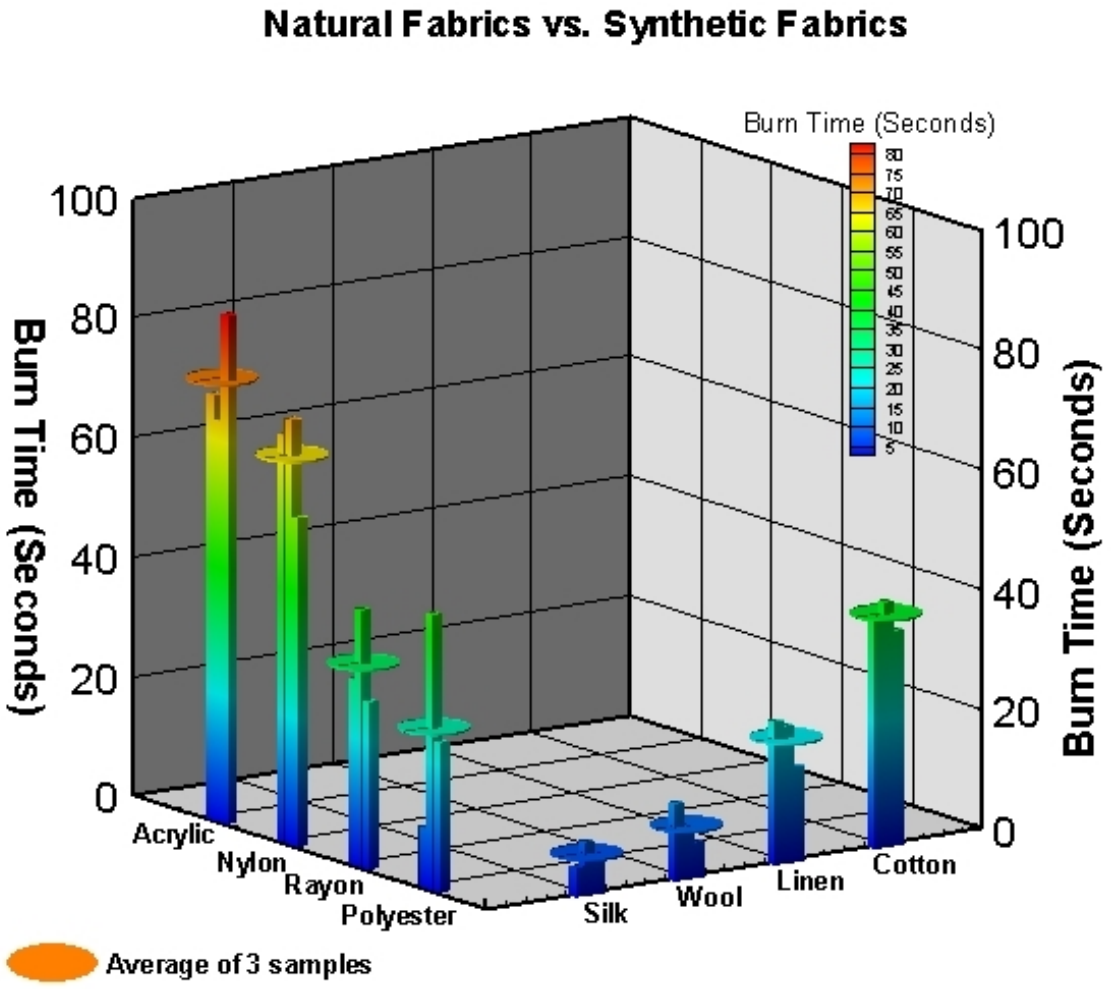
The 'billion pound' plot instead uses area – and the areas compared are far apart.



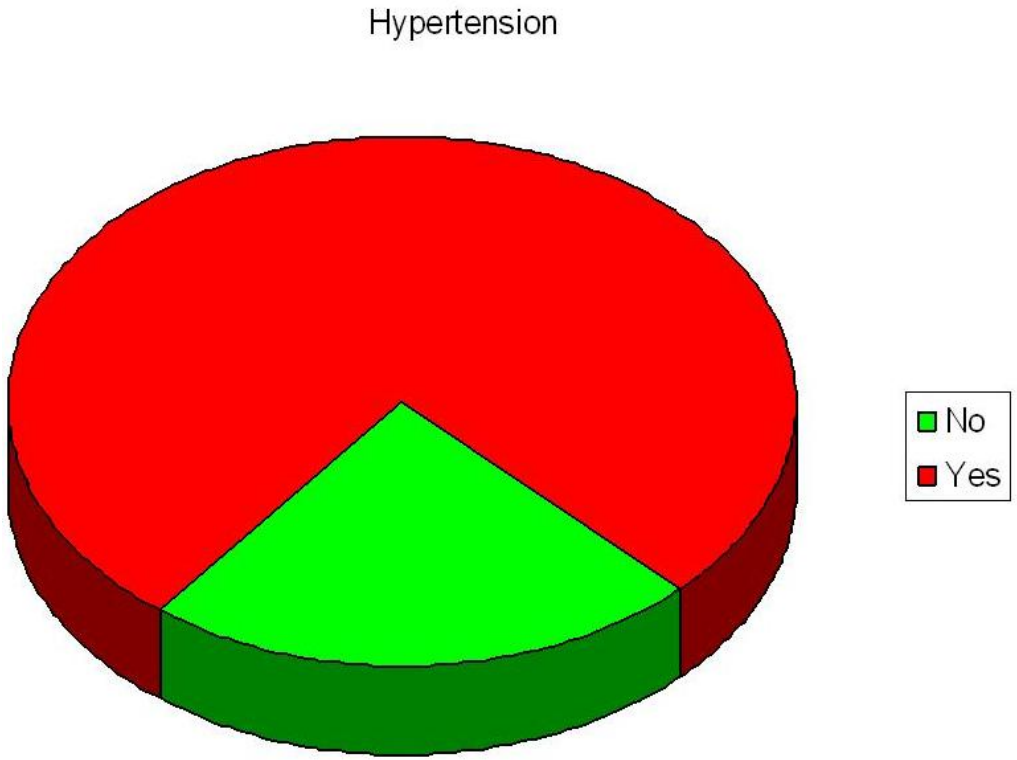

- 2–3 comparisons likely enough? McCandless' 'billion pound' plot is now rare
- Serve a clear purpose $\longrightarrow$ serve a clear **and useful** purpose?

# Principles: show the data

A spiffy 3D plot advertising GAUSS – 8 sets of 3 points each – note how the 'chartjunk' obscures the (simple) data



# Principles: show the data

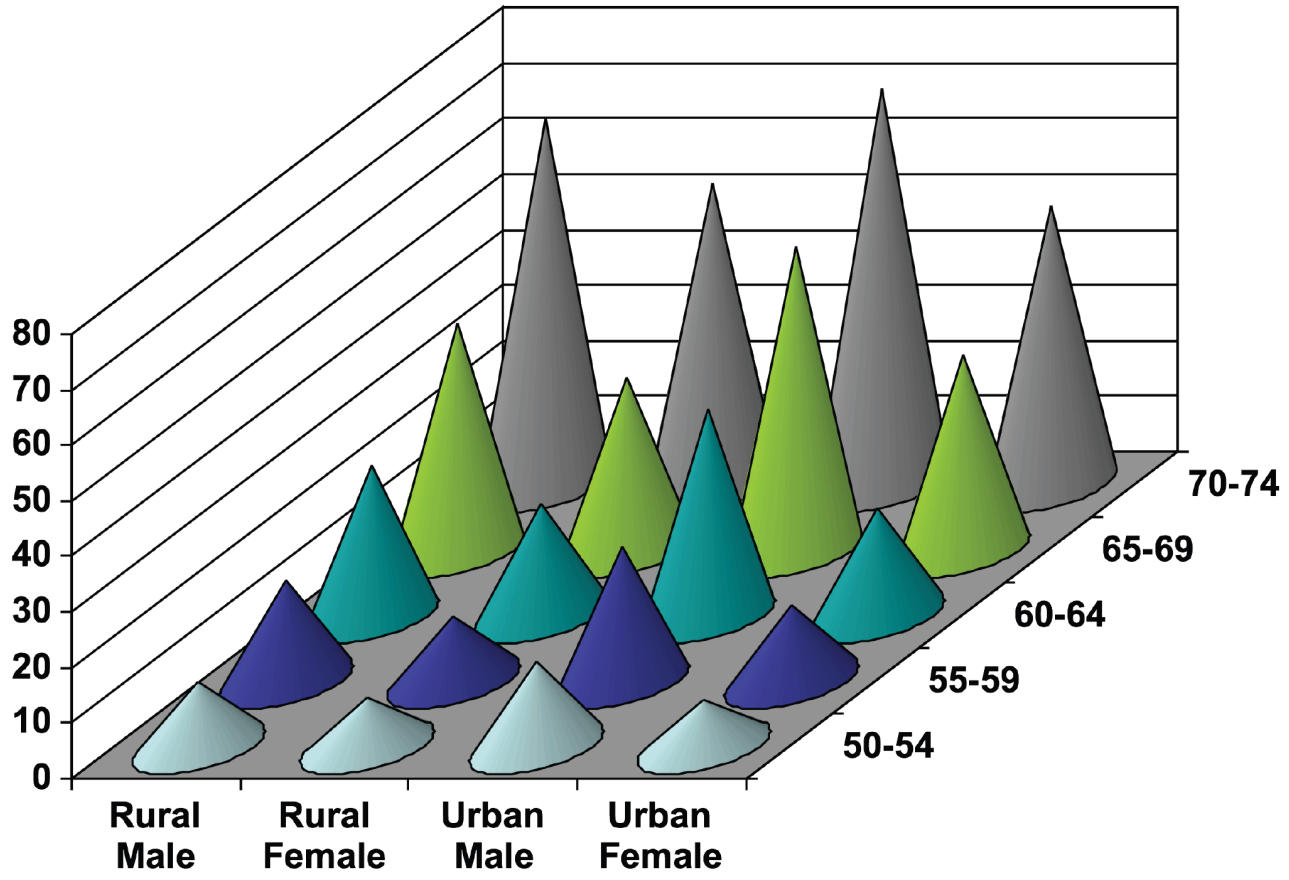



From a real poster; (American Heart Association Epi/Lifestyle conference); three of these (percentages Yes, Female, Yes & Female) were worth a 2×2 table...

# Principles: show the data

Another AHA poster's 'bed of nails' – it's torture!

- **Never** use fake 3D!
- Improved version follows
- Show the data $\longrightarrow$ show the data **clearly**?

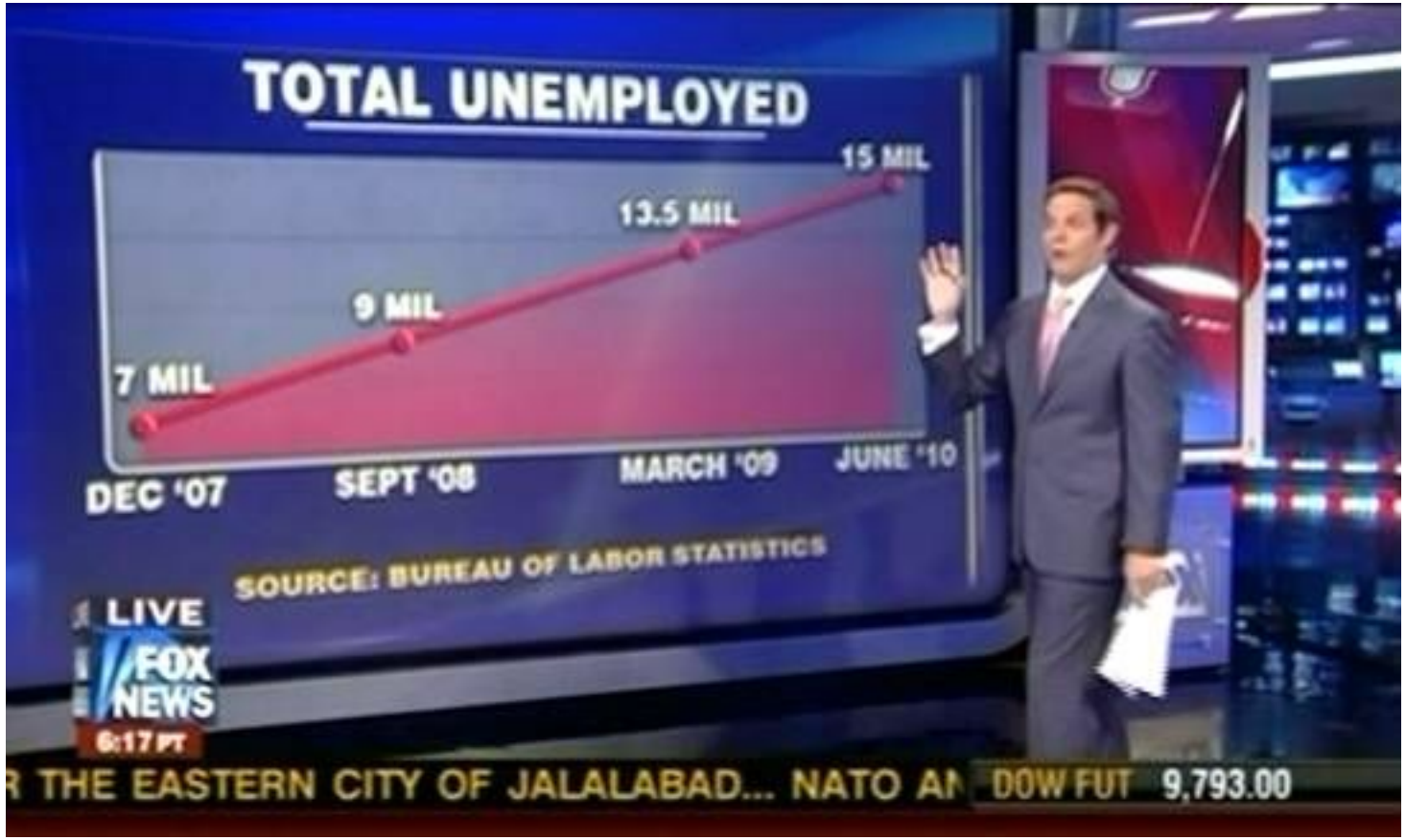


# Principles: Avoid distorting what the data say


"Fair and Balanced" Fox News reporting – "We Report. You Decide" (2010)

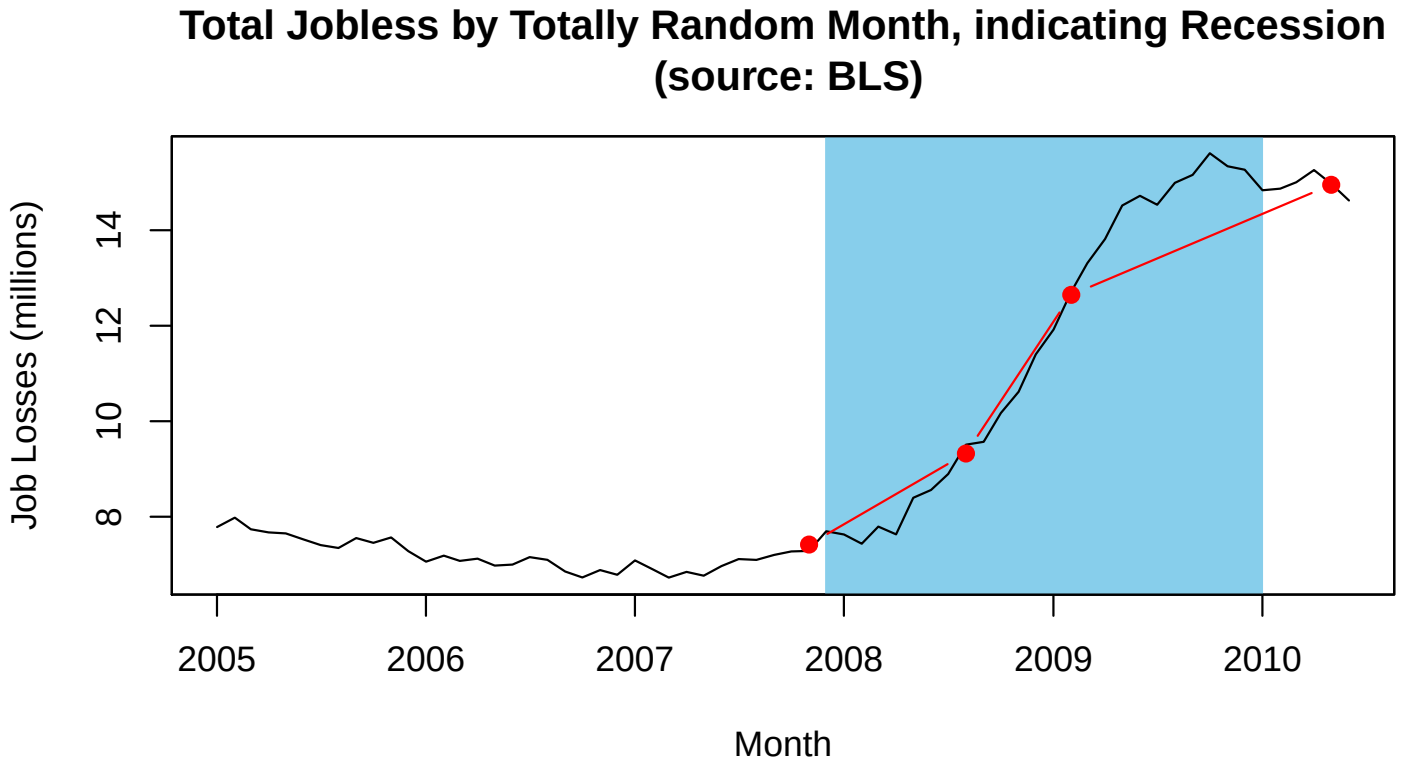
# Principles: Avoid distorting what the data say

They did correct the wildly-wrong title – while **still** distorting BLS's work



# Principles: Avoid distorting what the data say

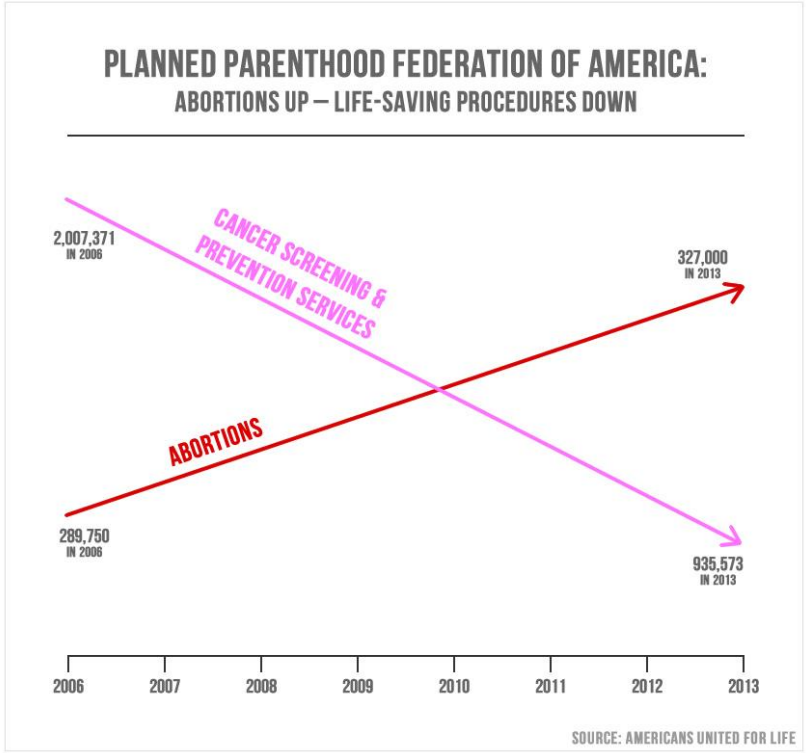
Distorting? Yes – as we see, using **actual** BLS data;



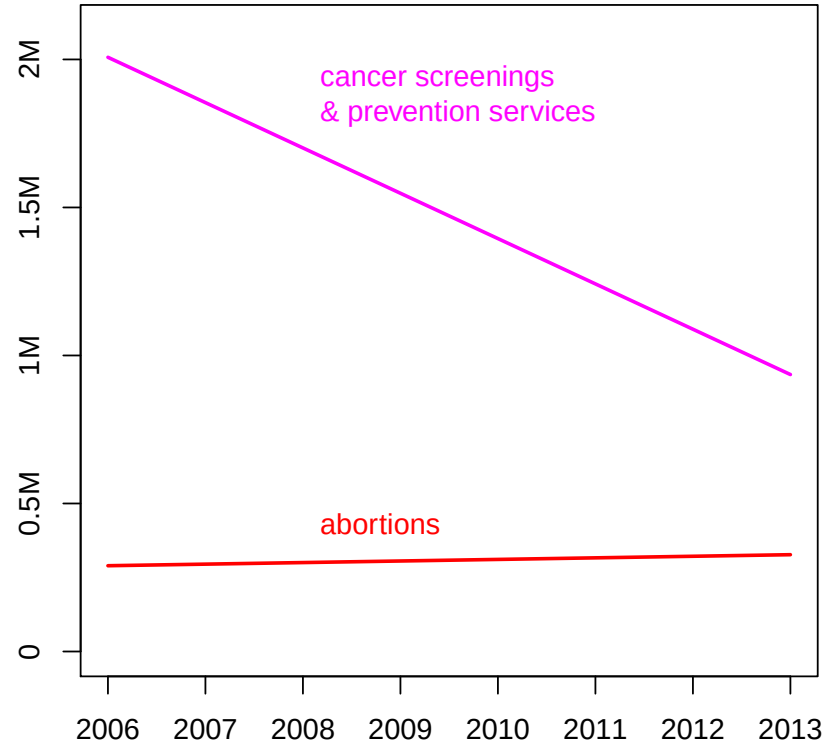
# Principles: Avoid distorting what the data say

*"In pink, that's the reduction in the breast exams, and the red is the increase in the abortions. That's what's going on in your organization."* – Rep Chaffetz

Orig:

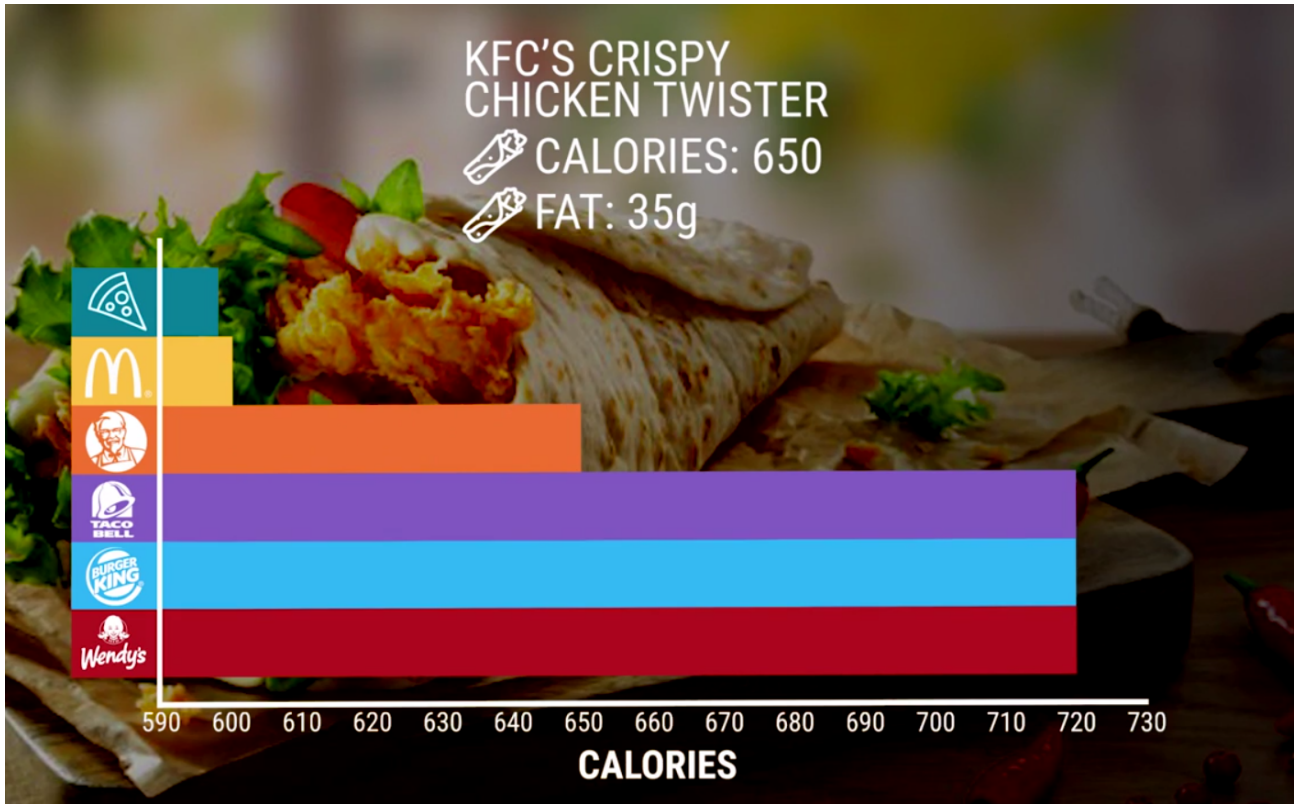



With $y$ axis:



# Principles: Avoid distorting what the data say

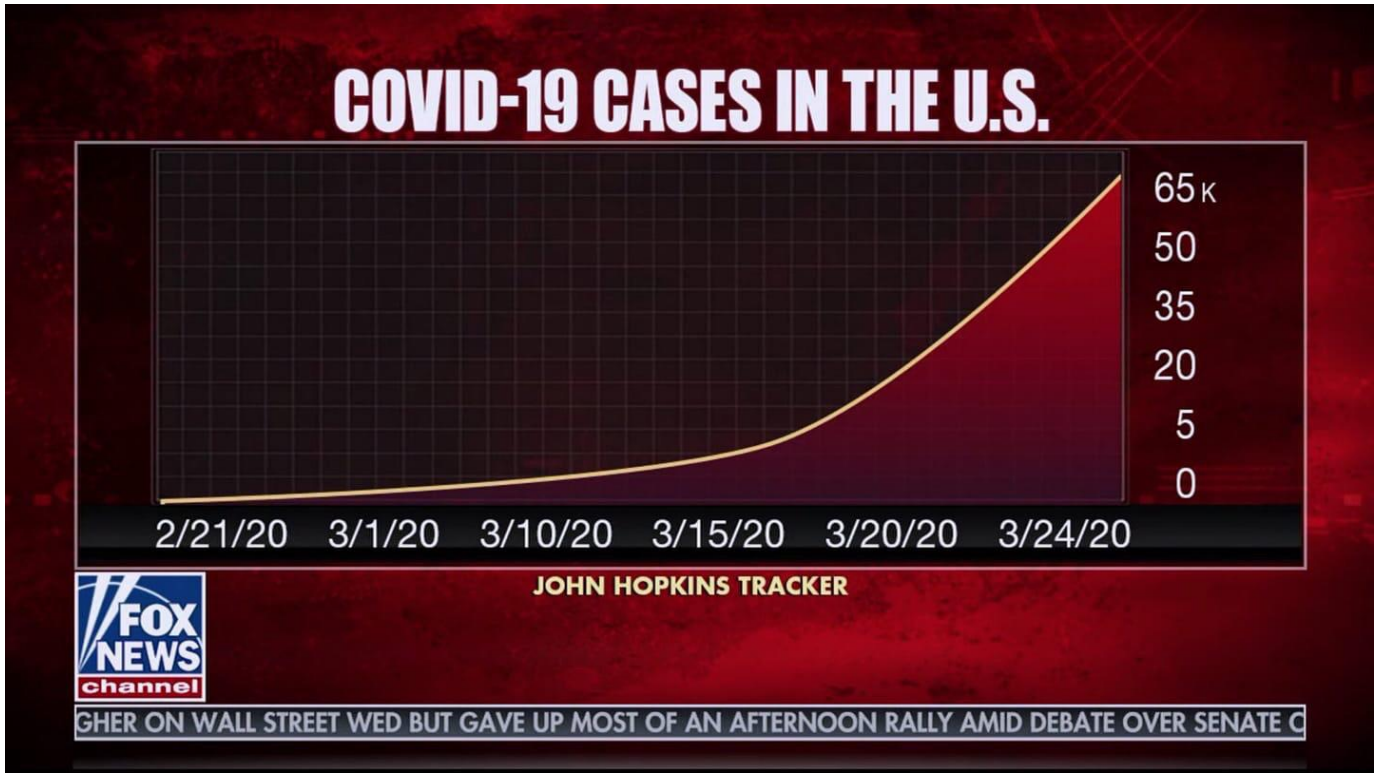
What **really old** trick distorts this bar chart?




Daryl Huff exposed this in *How To Lie With Statistics* (1954).

# Principles: Avoid distorting what the data say

And back to Fox News, who are still at it:




What are the **two** distortions? What is their impact?

# Principles: help the eyes to make comparisons

This is easier said than done. Good graphs, like good statistical analysis, should help your reader *accurately* assess whether;

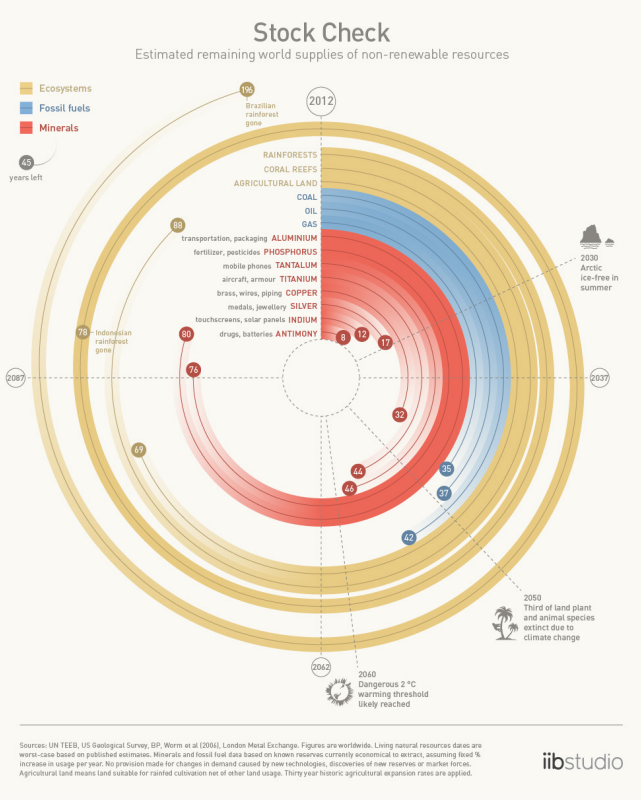
- The effect is there
- The effect is not there
- The data are so uninformative that no-one can tell

To make a graph that does this, a good starting point is 'use the R defaults' – these are based on work at Bell Labs in the 1970s, on early graphics systems, making use of research into how visual perception actually works.

They are a *little* out of date – particularly for plots of 'big data' – but still out-perform defaults from elsewhere.

# Effective Comparisons: pos'n on a common scale

An example from 514, using R's `dotchart()`:

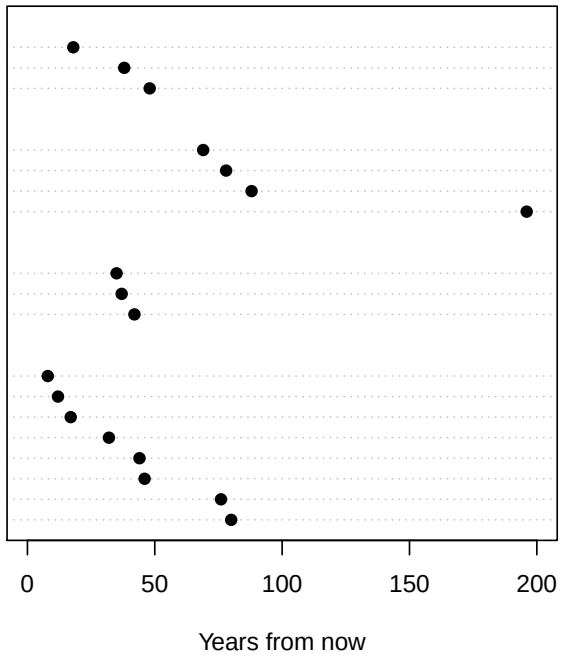



Climate Tipping Points
- Arctic ice−free in summer (worst−case forecast)
- 1/3 of land plant & animal species extinct due to climate change
- 2°C warming threshold likely reached

Ecosystems
- Suitable agricultural land runs out
- Indonesian rainforest completely deforested
- All coral reefs gone
- Amazon completely deforested

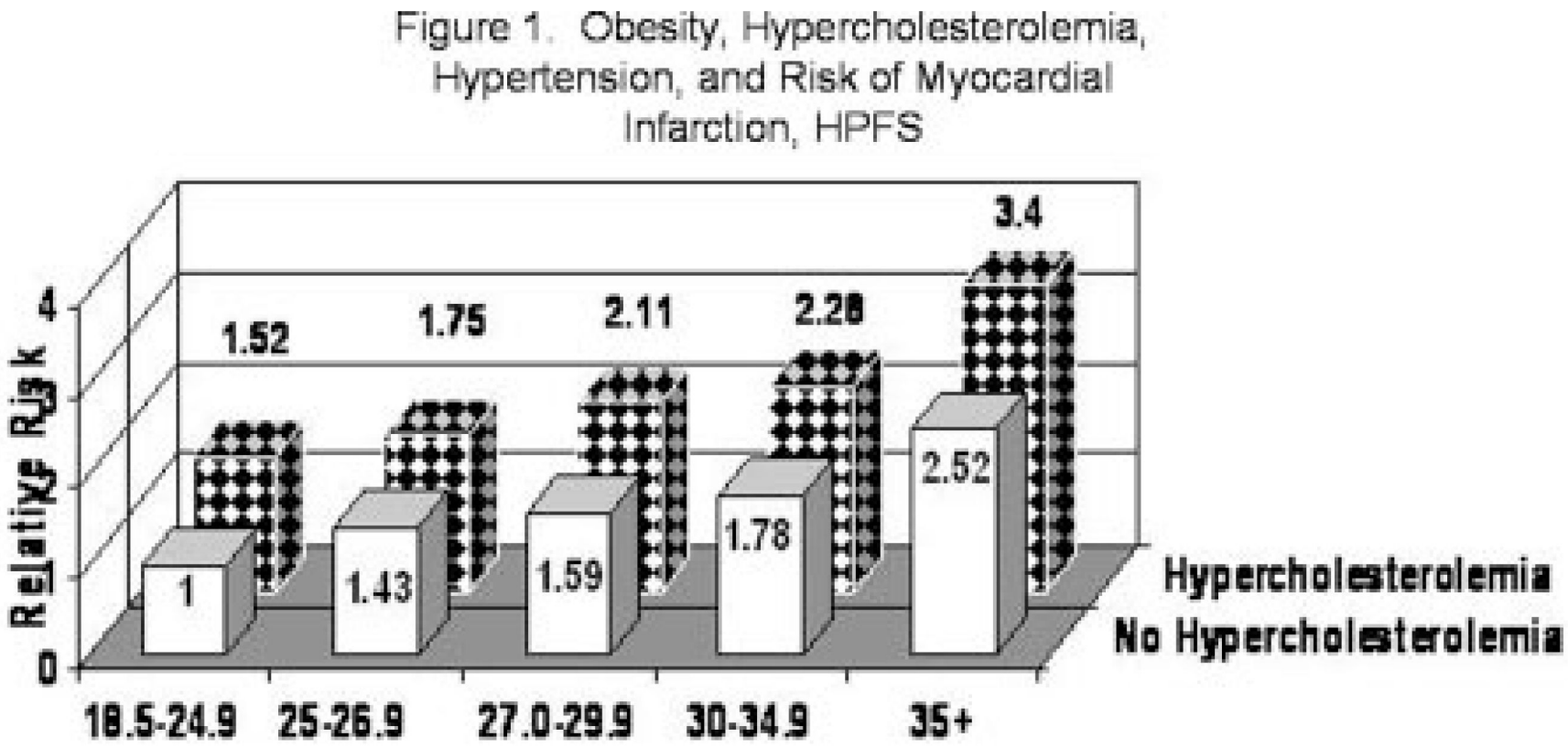
Fossil Fuels
- Gas
- Oil
- Coal

Minerals
- Antimony
- Indium
- Silver
- Copper
- Titanium
- Tantalum
- Phosphorus (phosphate rock)
- Aluminium



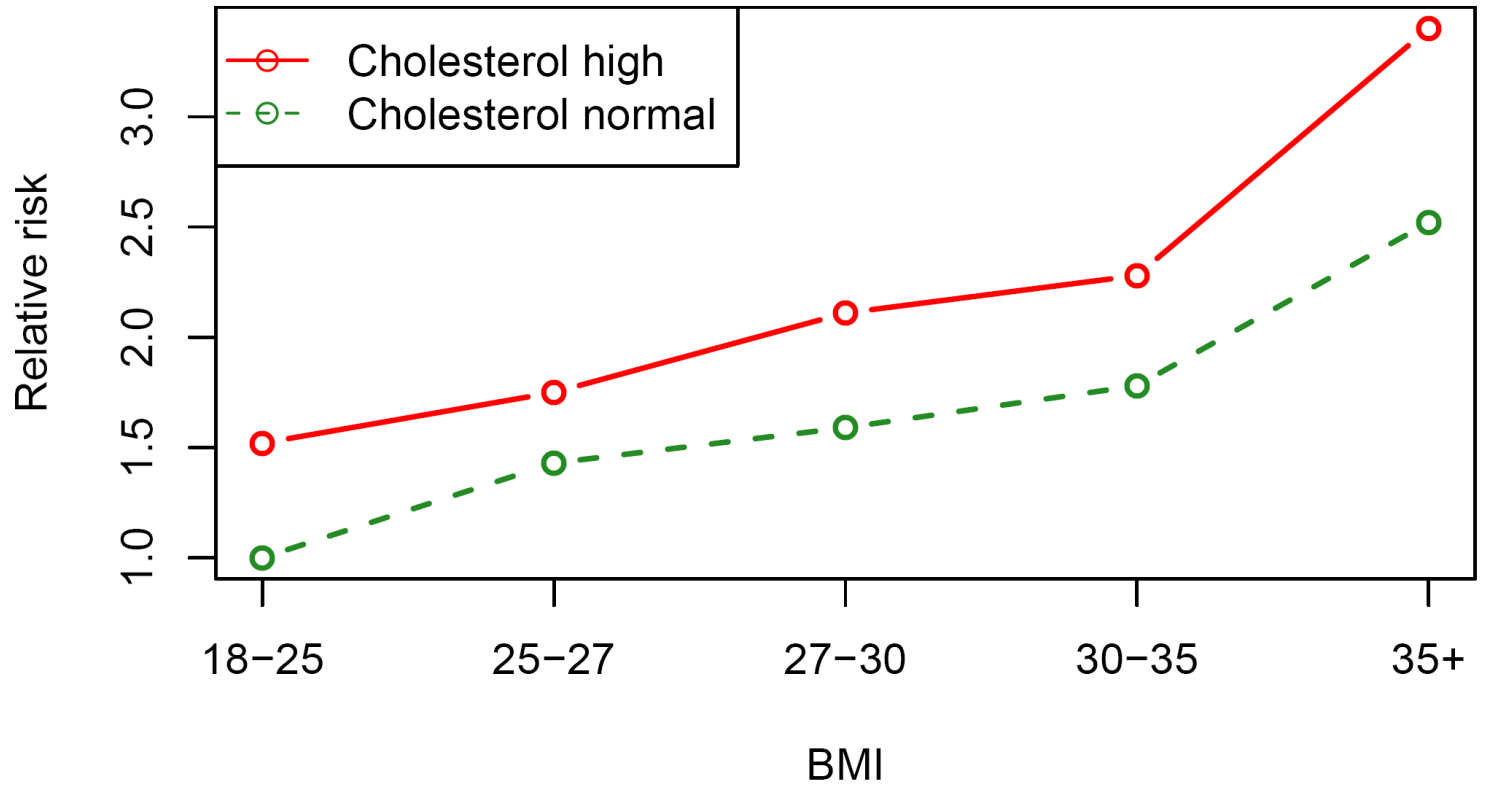
# Effective Comparisons: pos'n on a common scale

And now back to the AHA Epi/Lifestyle conference;



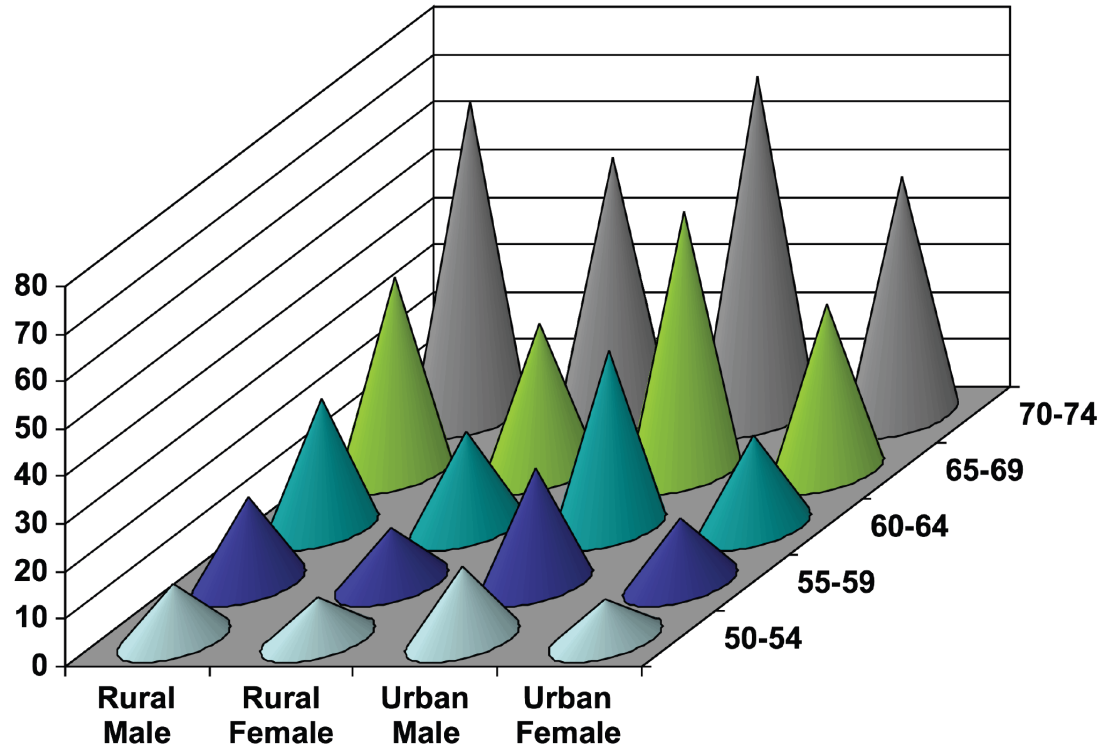
# Effective Comparisons: pos'n on a common scale

How would R do it? (using `type="b"`, CIs would help too)



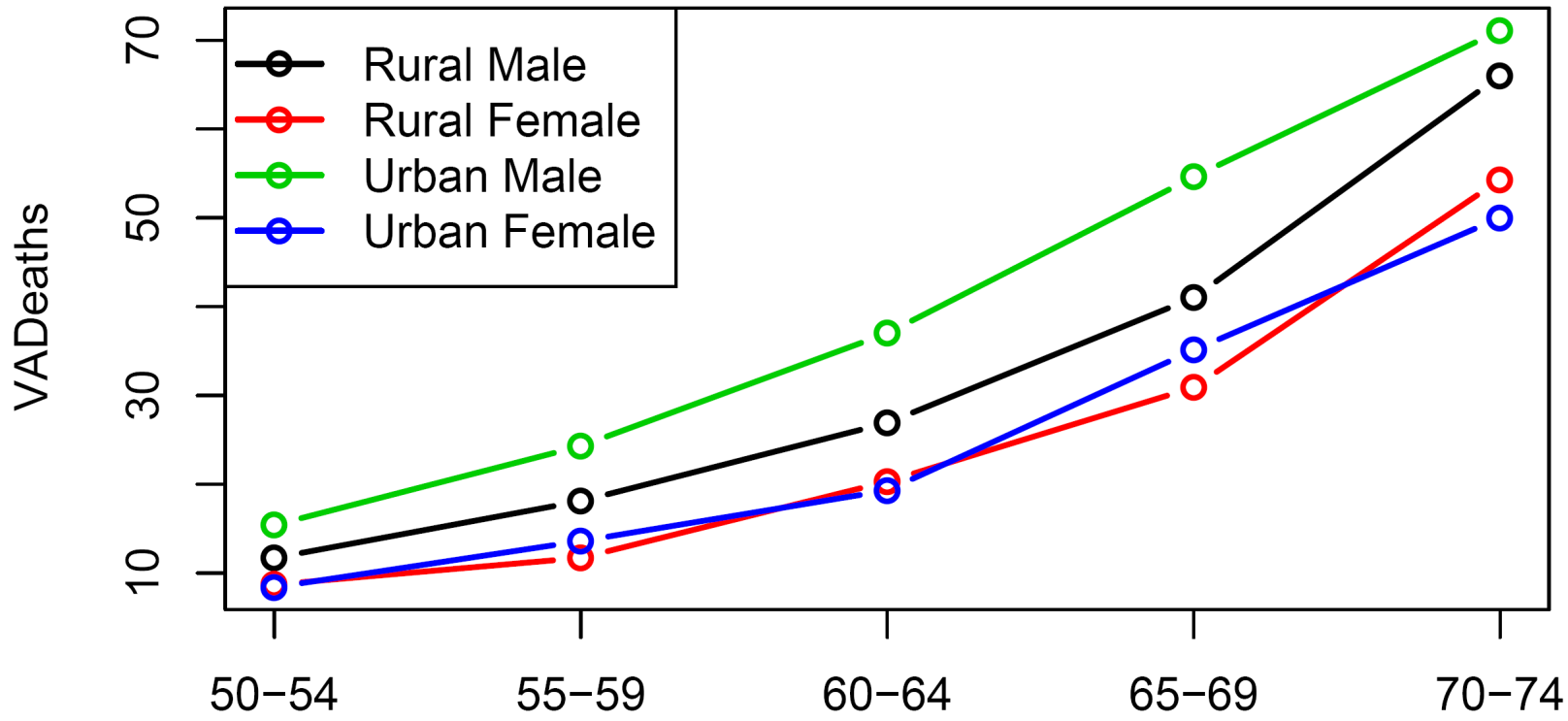
# Effective Comparisons: pos'n on a common scale

The bed of nails returns!



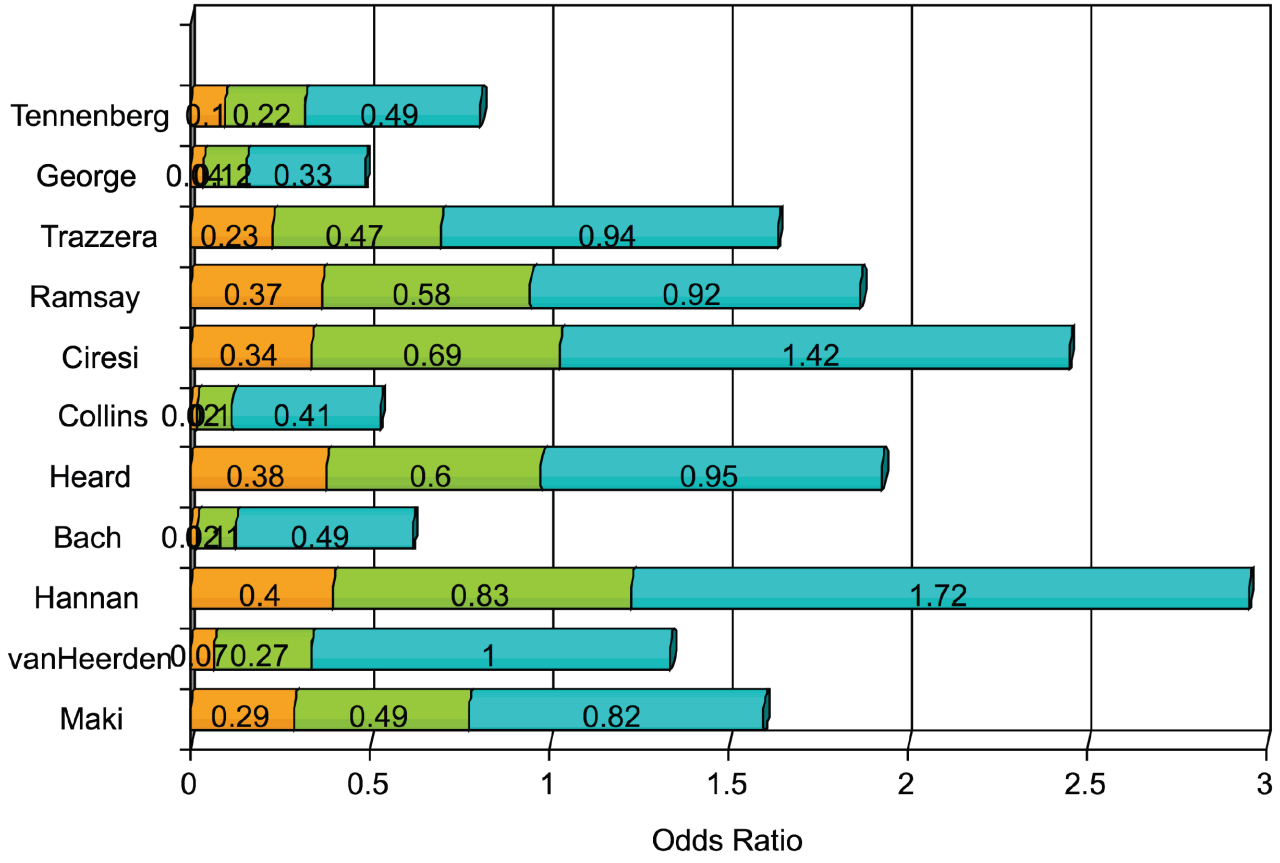
# Effective Comparisons: pos'n on a common scale

Lines indicate *underlying continuum*. Also note overplotting/empty circles;



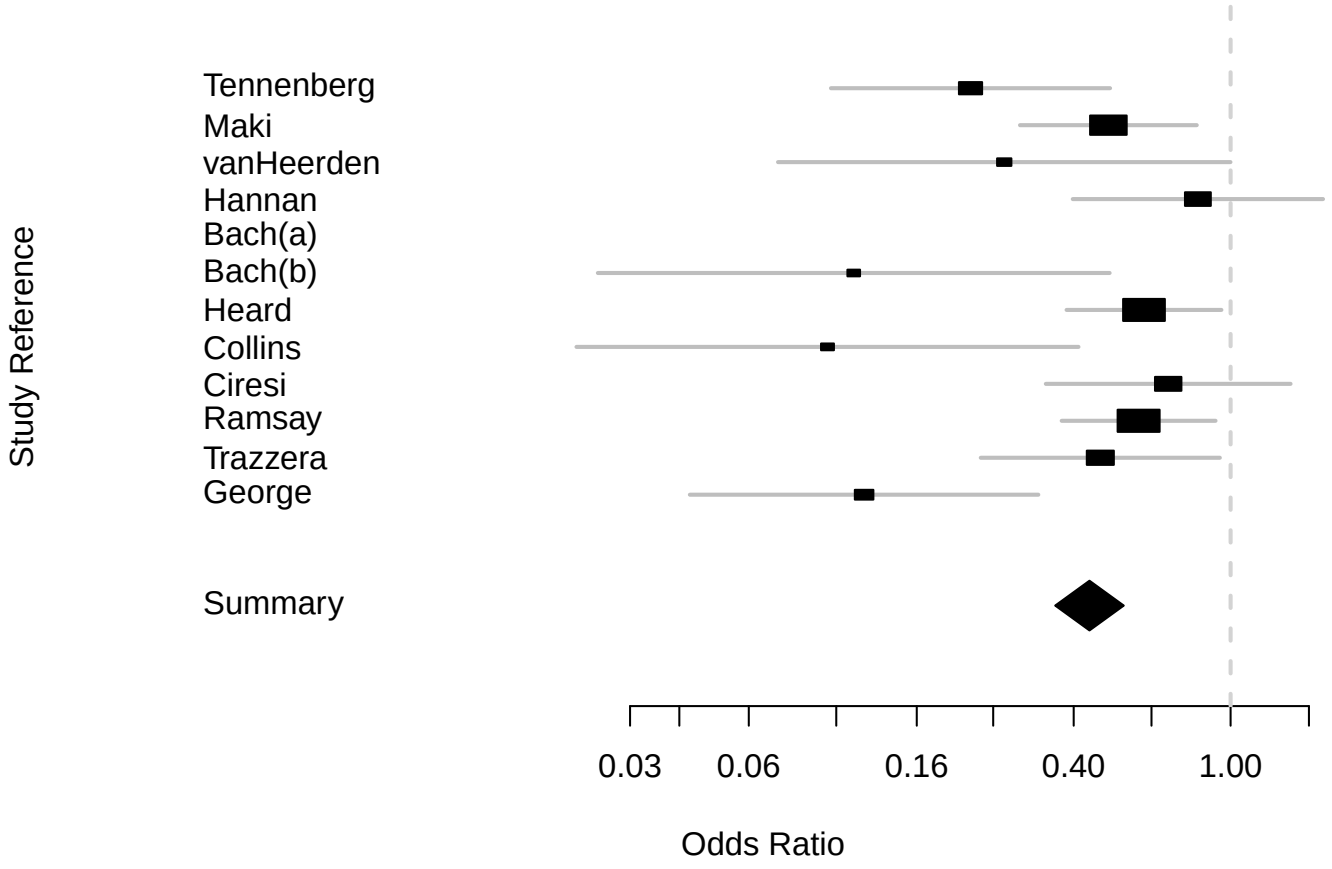
# Effective Comparisons: pos'n on a common scale

How **not** to compare intervals on a common scale:



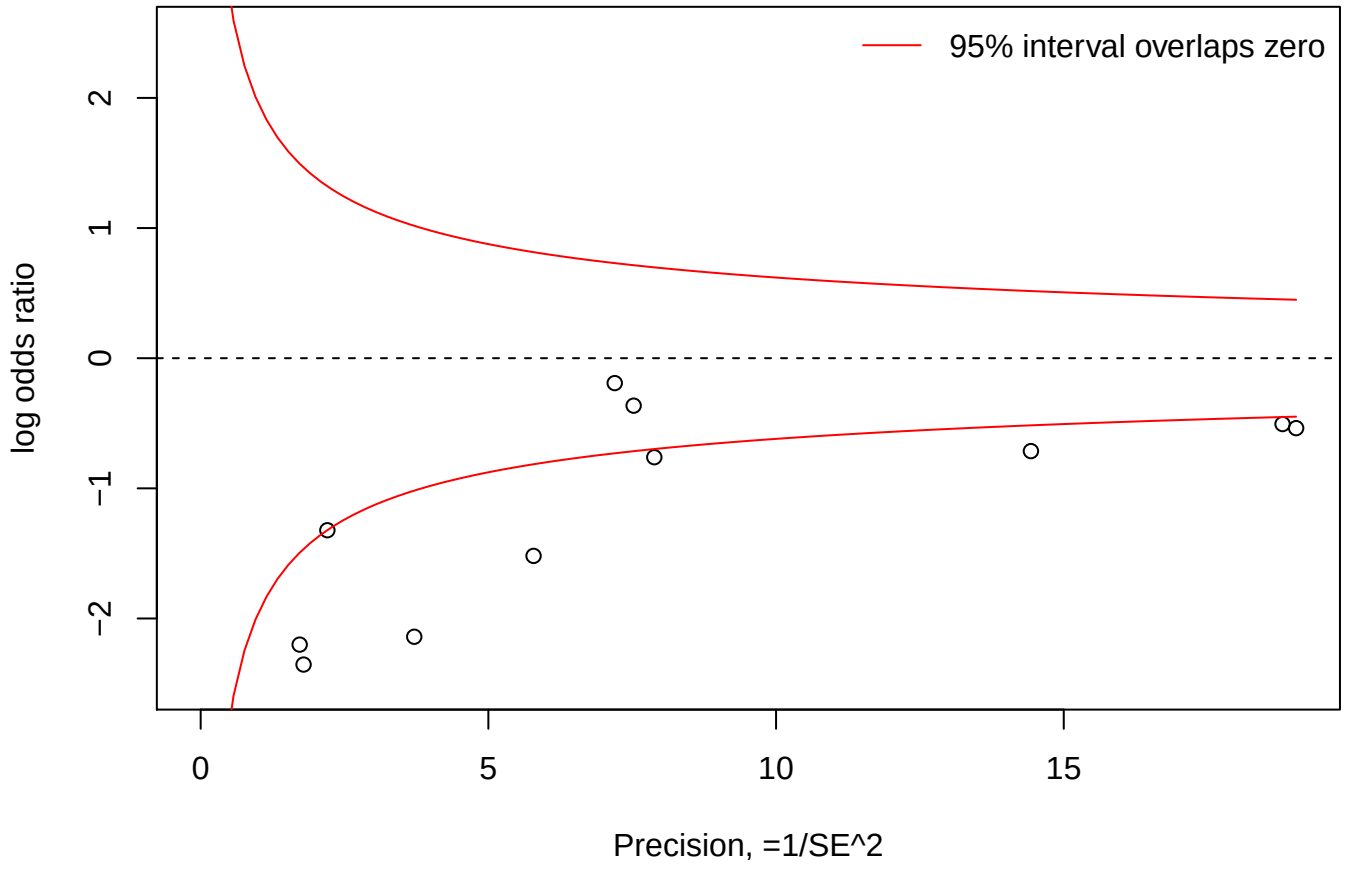
# Effective Comparisons: pos'n on a common scale

Instead using a standard *forest plot*, from the `rmeta` or `metafor` packages:



# Effective Comparisons: pos'n on a common scale

*Funnel plots* show precision ($1/SE^2$) as position, can help show publication bias;



# Effective comparisions: what works well? or not?

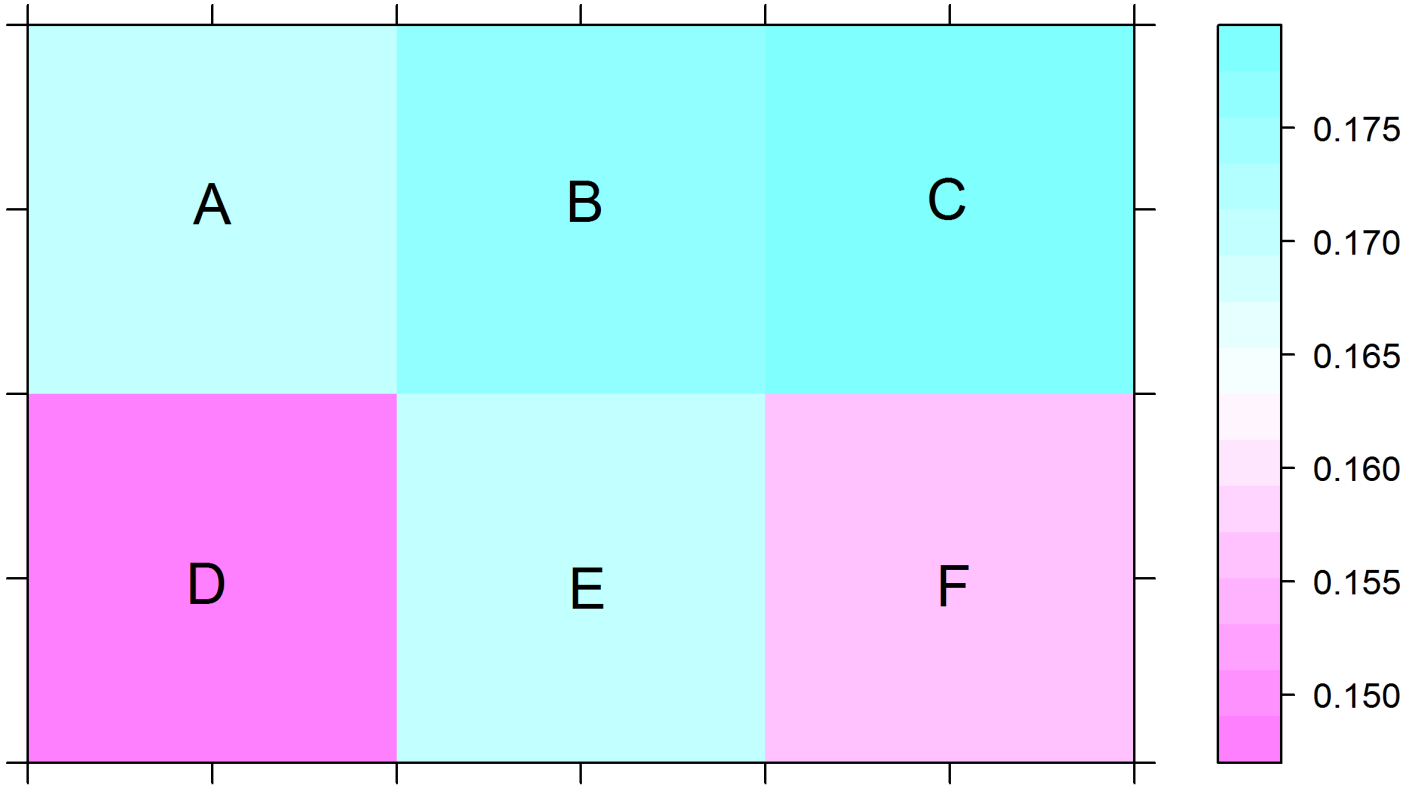
As we've seen, looking cool doesn't mean a comparison **actually works well**. For comparing numeric data, here is an ordering based on perception research;

| Metric | Usage | Accuracy |
|---|---|---|
| Position on common scale | Dot Plot | Best |
| Length | Bar chart | |
| Angle/Slope | Pie chart | |
| Area | Bubble Plot | |
| Volume/Curvature | Fake3D | |
| Color hue, density | e.g. Heat map | Worst |

- See also Cleveland & McGill (JRSSA 1987), and books by Stephen Few and Alberto Cairo
- Let's illustrate this ordering; on the next slide, rank the 6 numbers A,B,...,F – smallest to largest

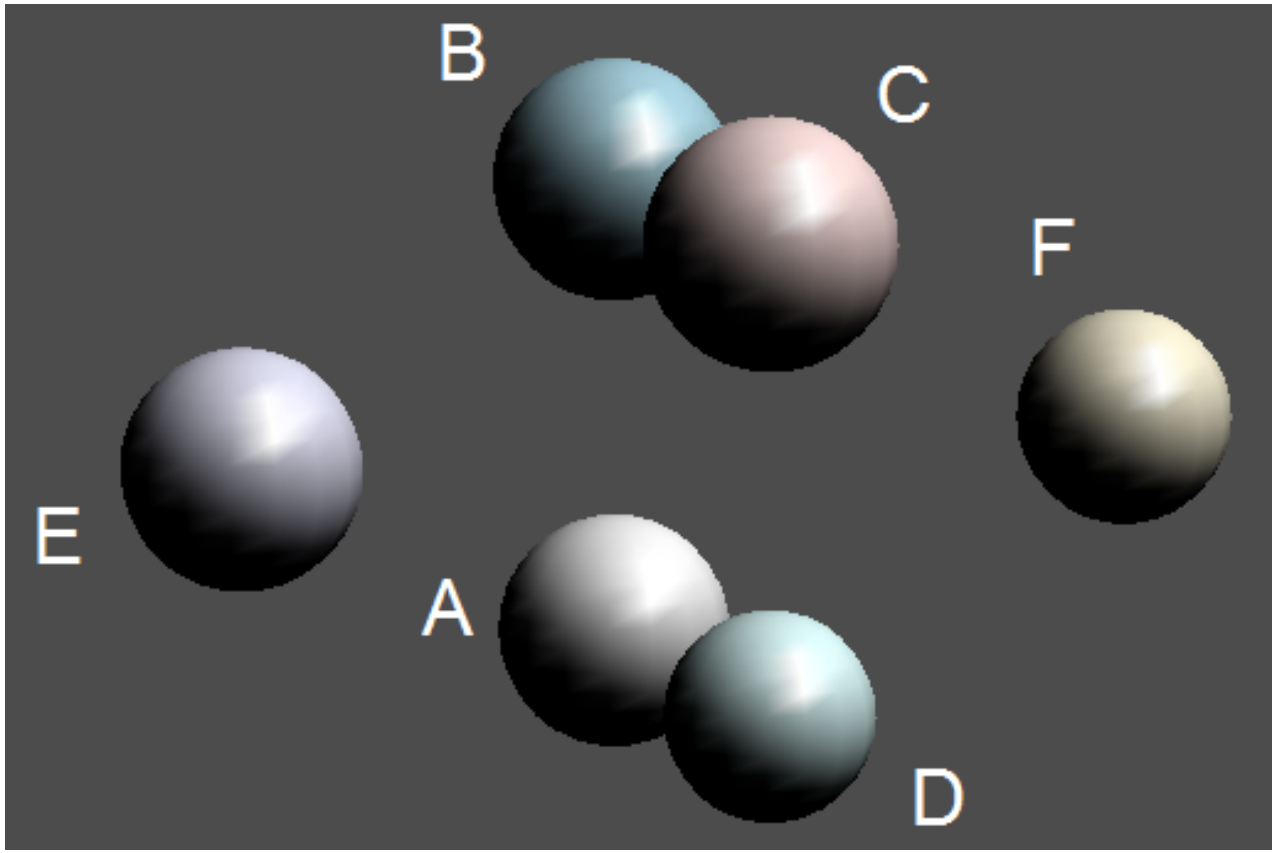
# Effective comparisions: what works well? or not?

Using `image()` or `levelplot()`; (Larger heatmaps can only show trends)



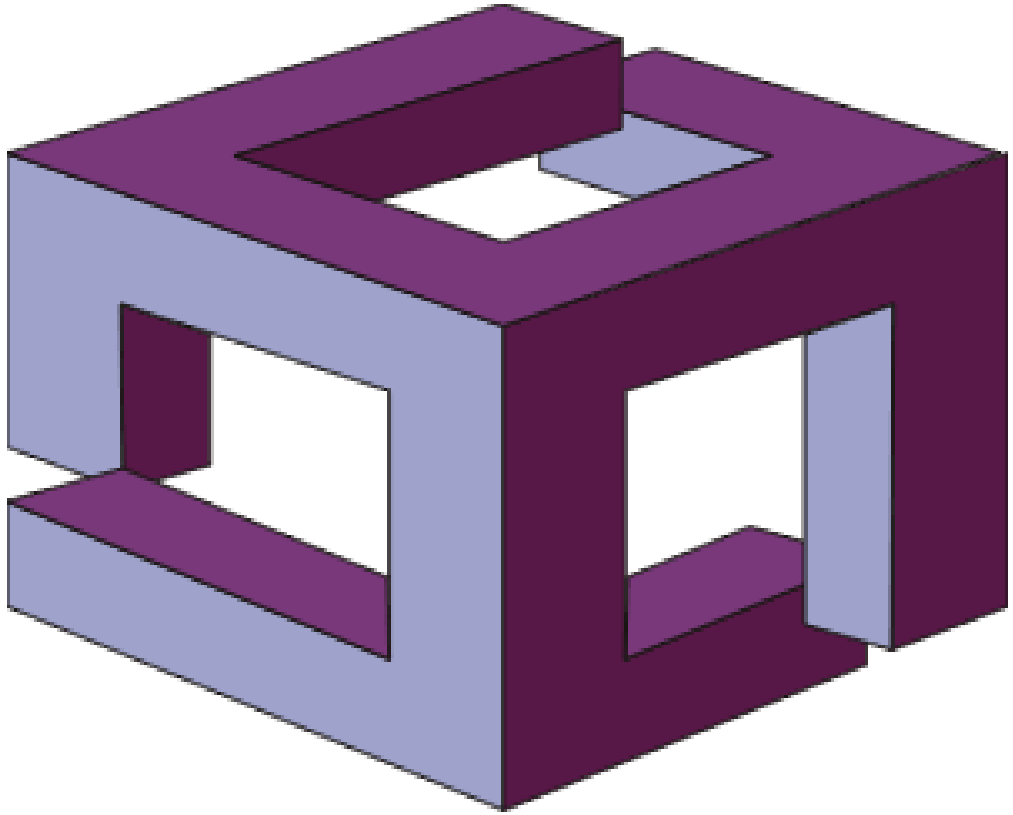
# Effective comparisions: what works well? or not?

R can make fake 3D—with `persp()`, or the `rgl` package—but you know better!



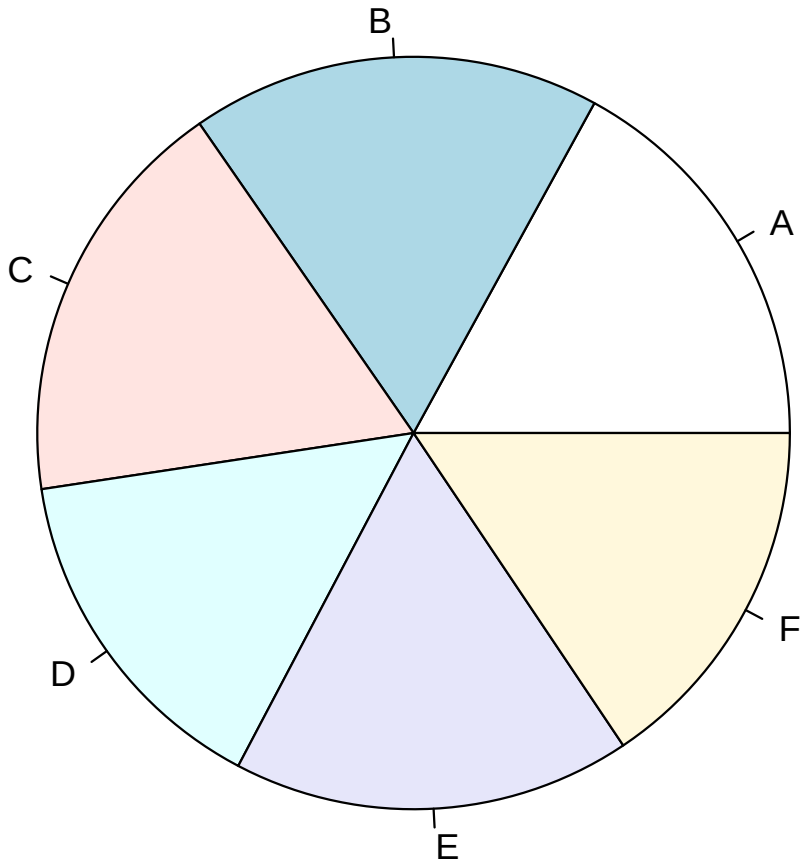
# Effective comparisions: what works well? or not?

Why not fake 3D? Your brain (and everyone else's) is really poor at unpicking 3D information from 'flat' pictures;



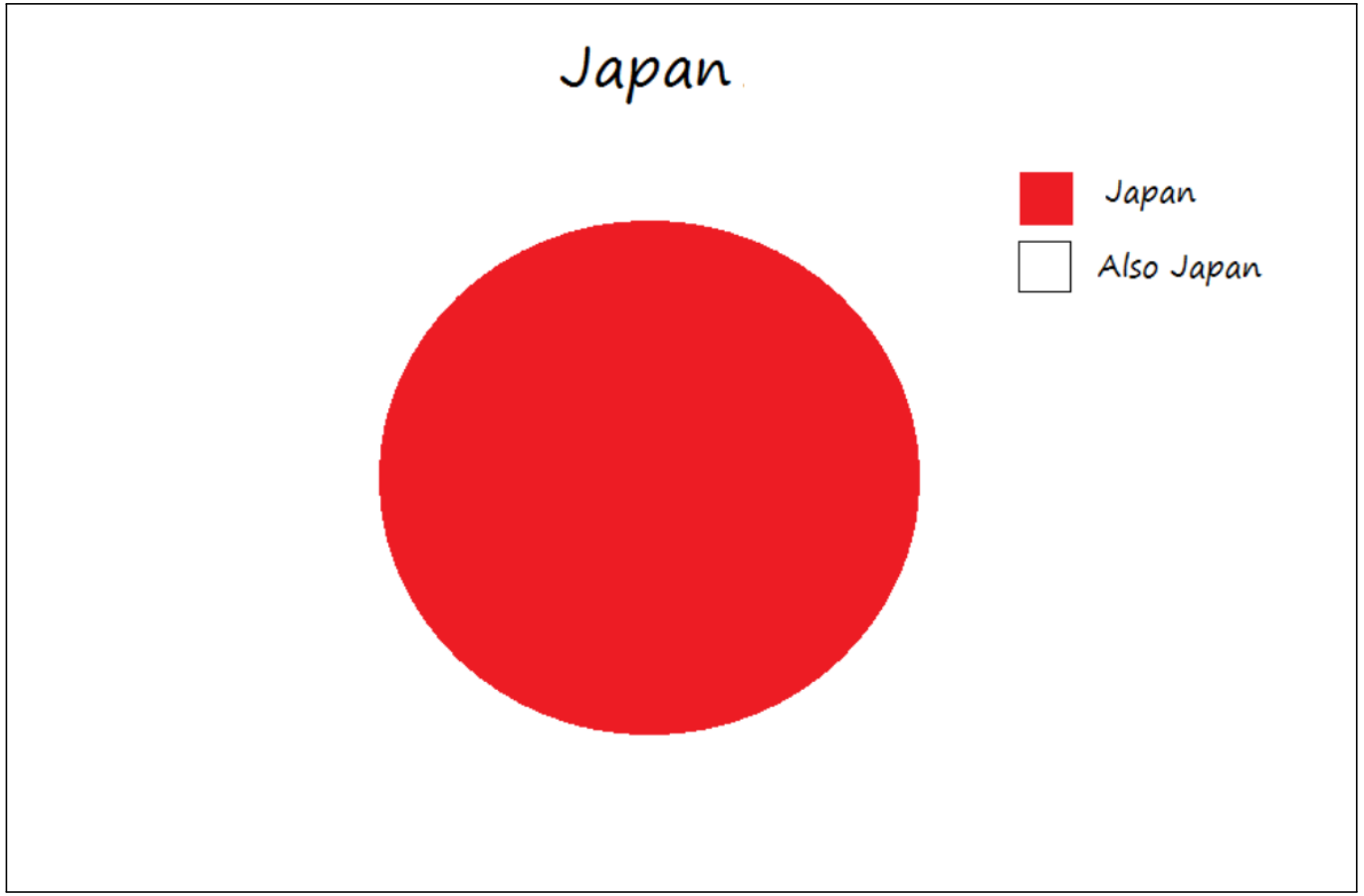
# Effective comparisions: what works well? or not?

Back to our 6 numbers; `pie()` is available, but seldom useful;



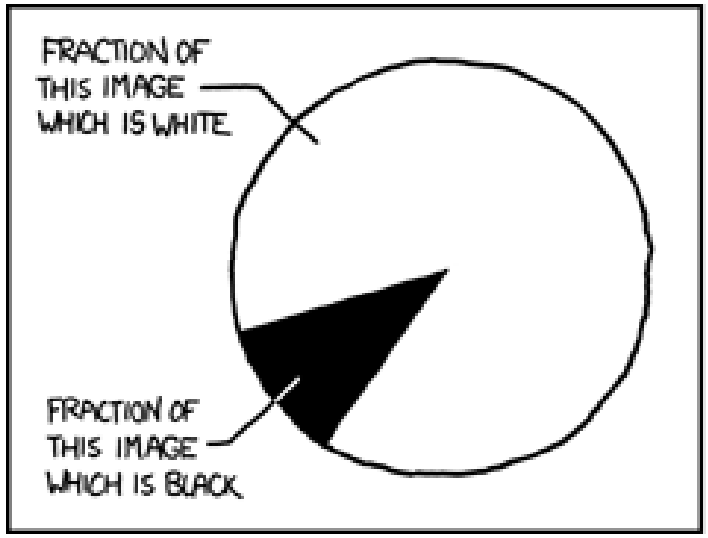
# Effective comparisions: what works well? or not?

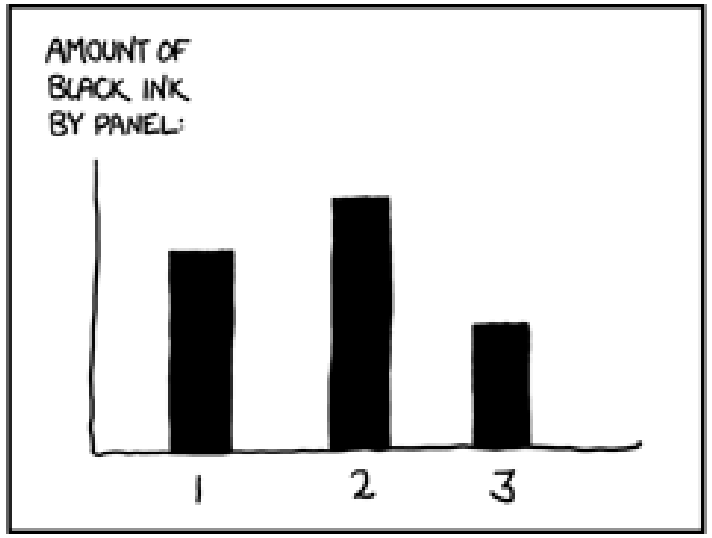
... except for fun (see also 514)

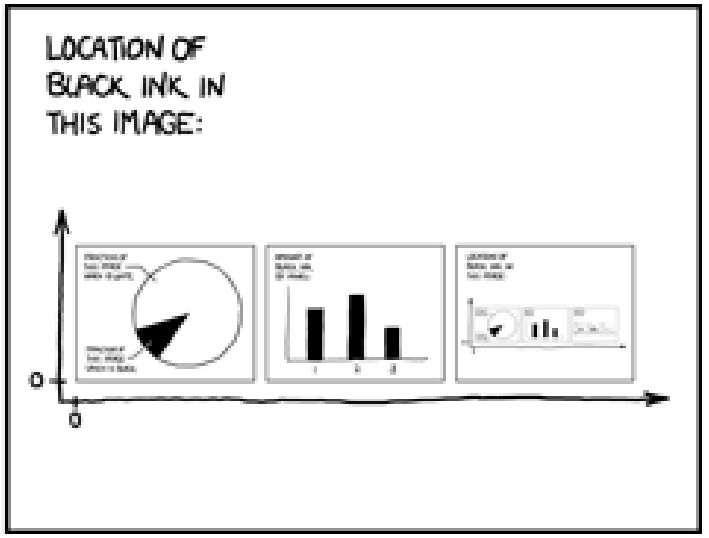


# Effective comparisions: what works well? or not?

... except for fun (see also 514)



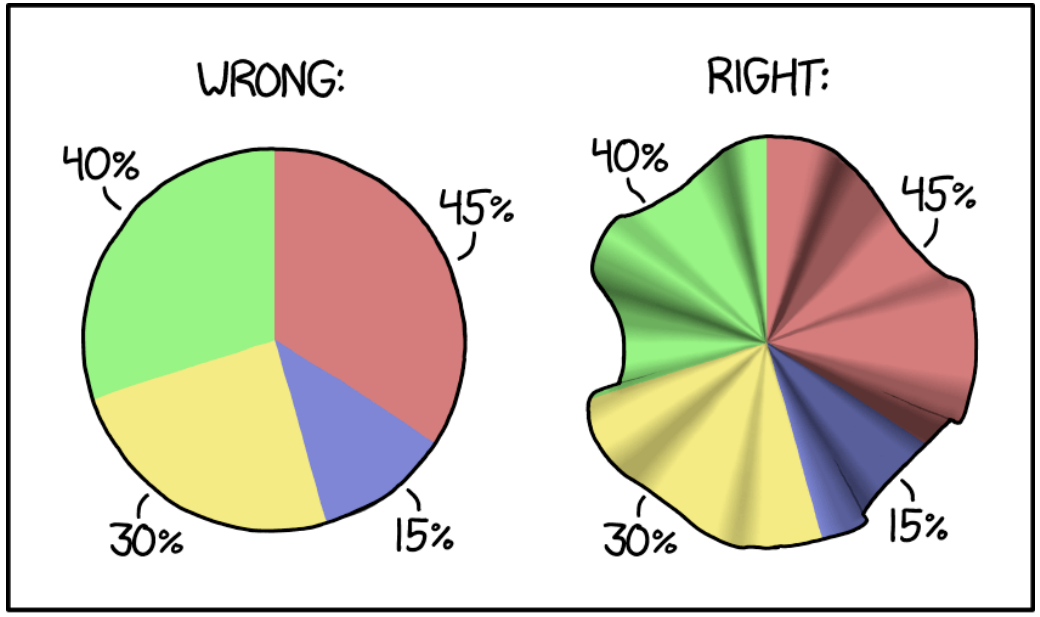




*The contents of any one panel are dependent on the contents of every panel including itself. The graph of panel dependencies is complete and bidirectional, and each node has a loop. The mouseover text has two hundred and forty-two characters*

# Effective comparisions: what works well? or not?

... except for fun (see also 514)

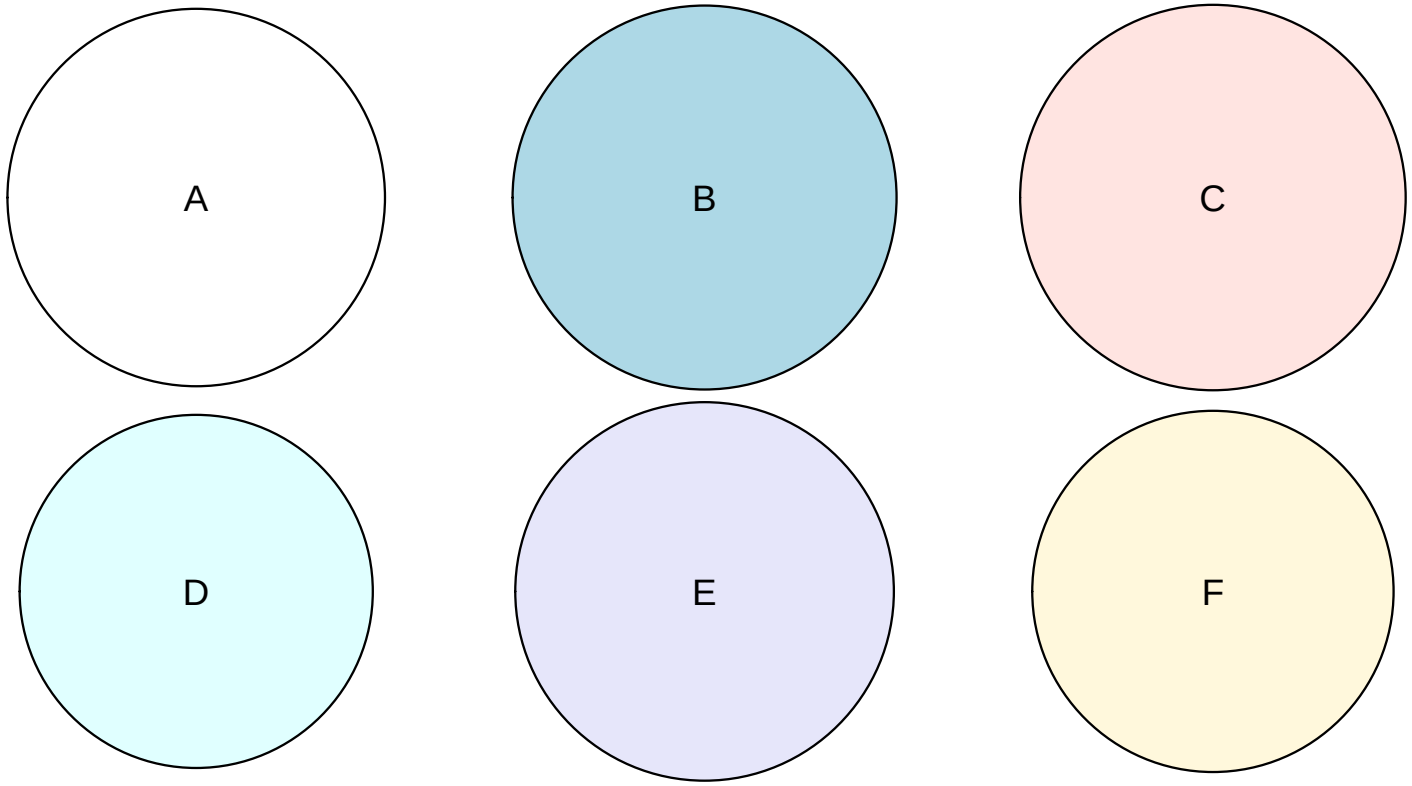



HOW TO MAKE A PIE CHART IF YOUR PERCENTAGES DON'T ADD UP TO 100

*If you can't get your graphing tool to do the shading, just add some clip art of cosmologists discussing the unusual curvature of space in the area.*

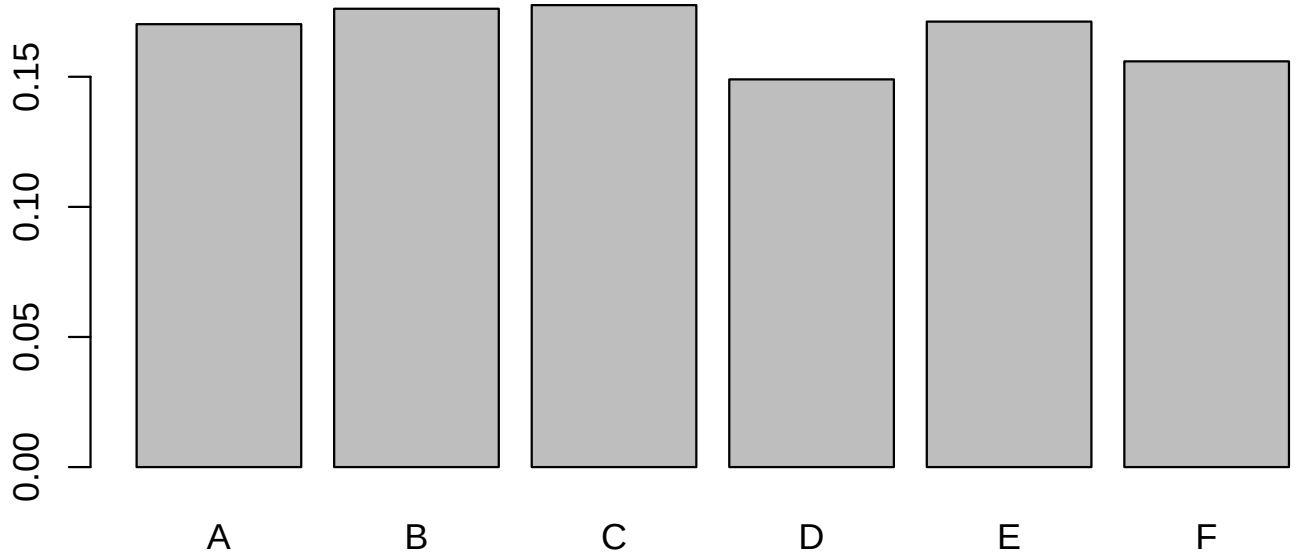
# Effective comparisions: what works well? or not?

Back to our 6 numbers: comparing by area – see `symbols()`



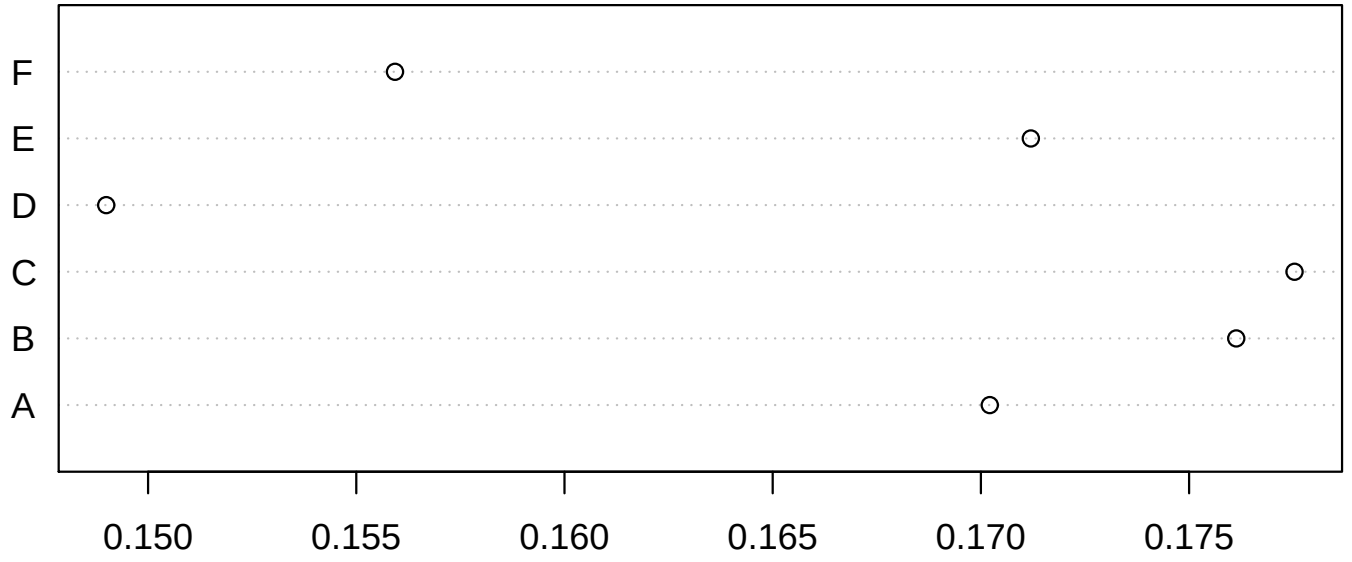
# Effective comparisions: what works well? or not?

Using area more reliably – with a `barplot()` that starts at zero;



# Effective comparisions: what works well? or not?

And finally (and best) — 'position on a common scale'



# Visual perception

Your vision evolved, primarily, to avoid predators and find food – not to read scientific data.

How many 5's in this list?

086010239034521204582510
119454921187766543883695
937945255947375722930620

- This task requires your conscious *attention*
- Your reader has a limited attention span, and memory – here, the bad presentation will distract/annoy them

# Visual perception

Your vision evolved, primarily, to avoid predators and find food – not to read scientific data.

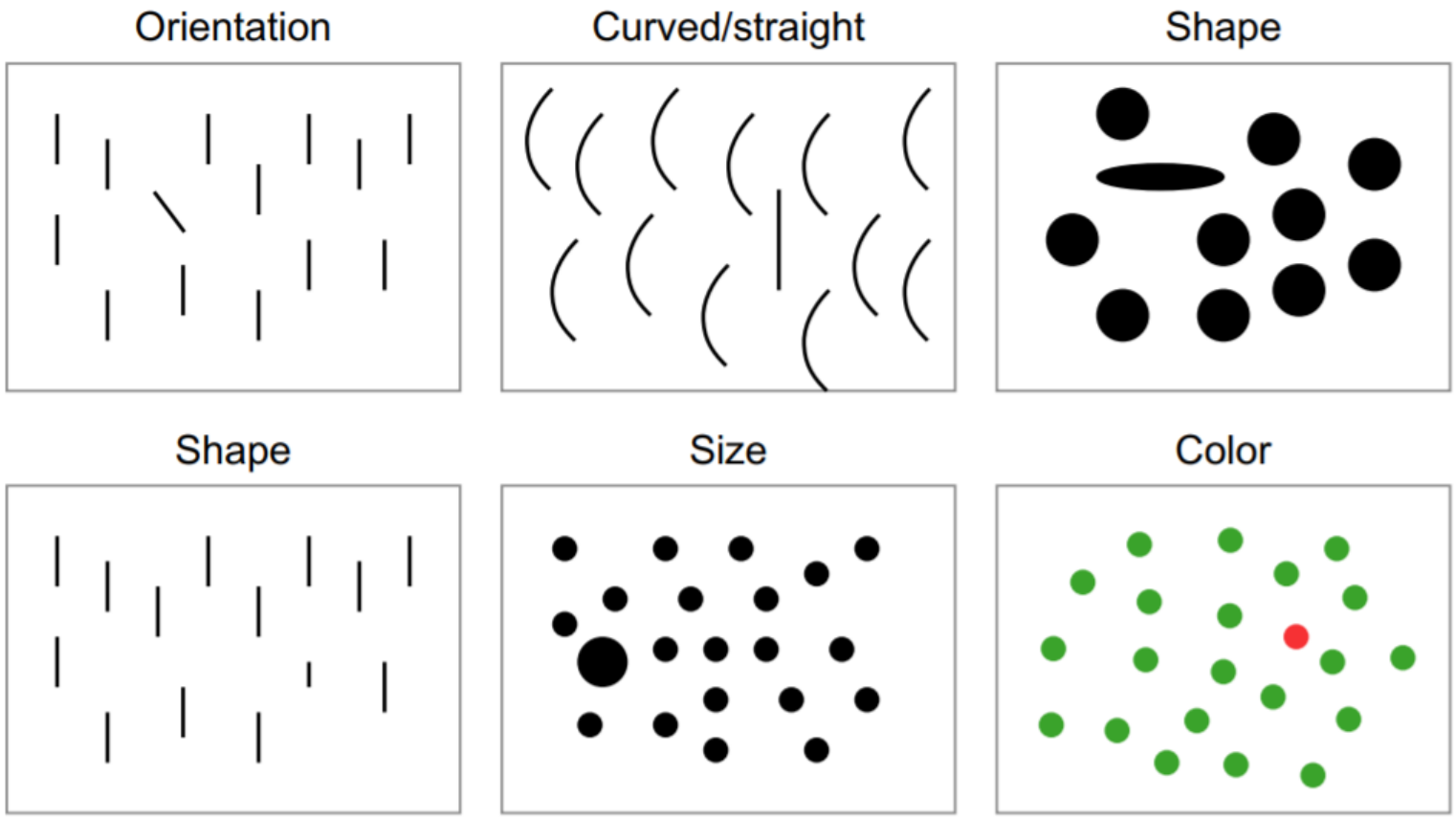
How many 5's in this list?

0860102390345212046825 10
1194549211877665438836 95
9379482559473747229306 20

- Some visual signals (e.g. color) are processed *pre-attentively*
- By using these signals, you make comparisons easy, and avoid distracting/annoying your reader with trivialities

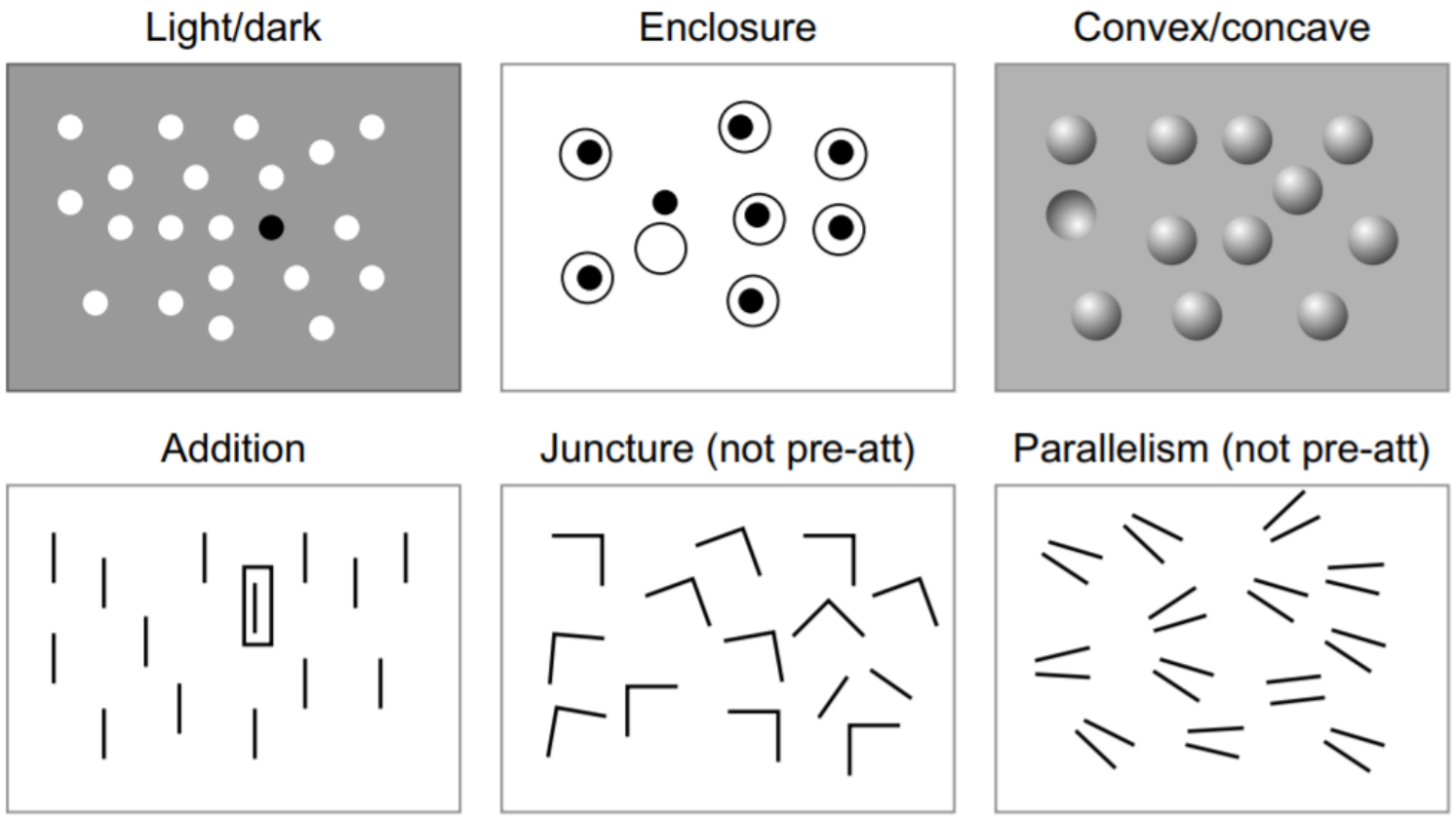
# Visual perception: why graphs work well, or not

From Colin Ware's Visualizing Information; 10 pre-attentive features:



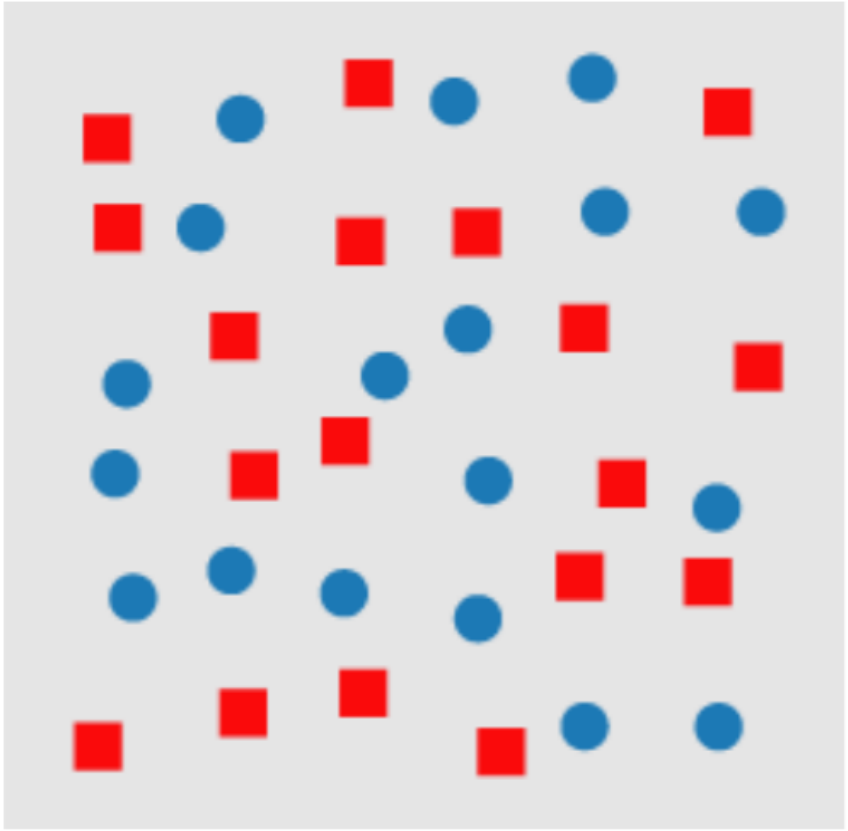
# Visual perception: why graphs work well, or not

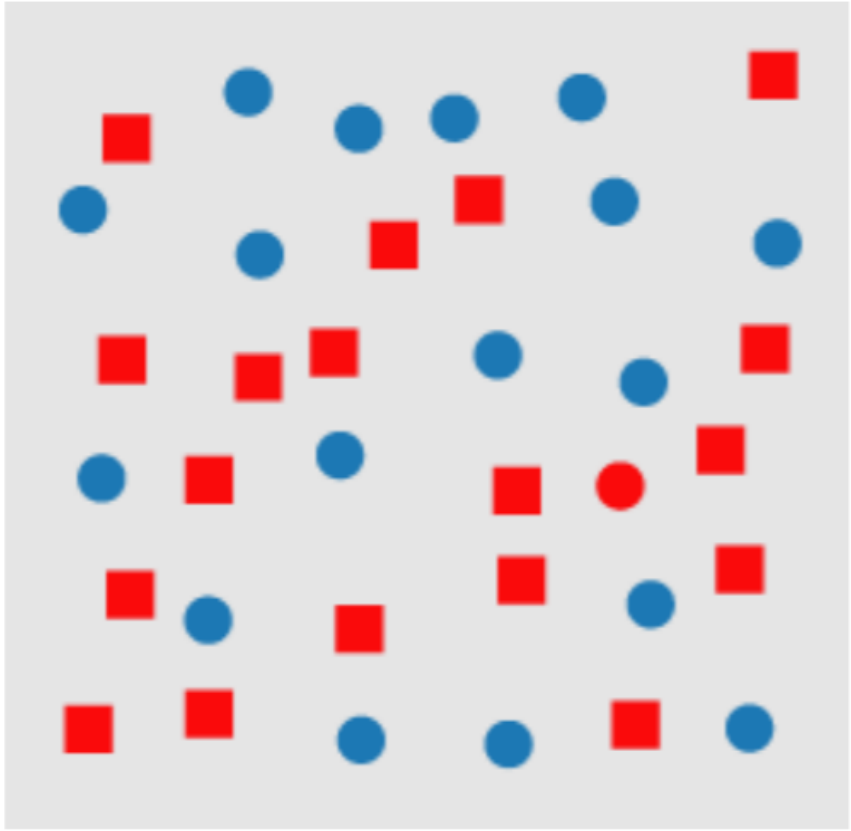
From Colin Ware's Visualizing Information; 10 pre-attentive features:



# Visual perception: why graphs work well, or not

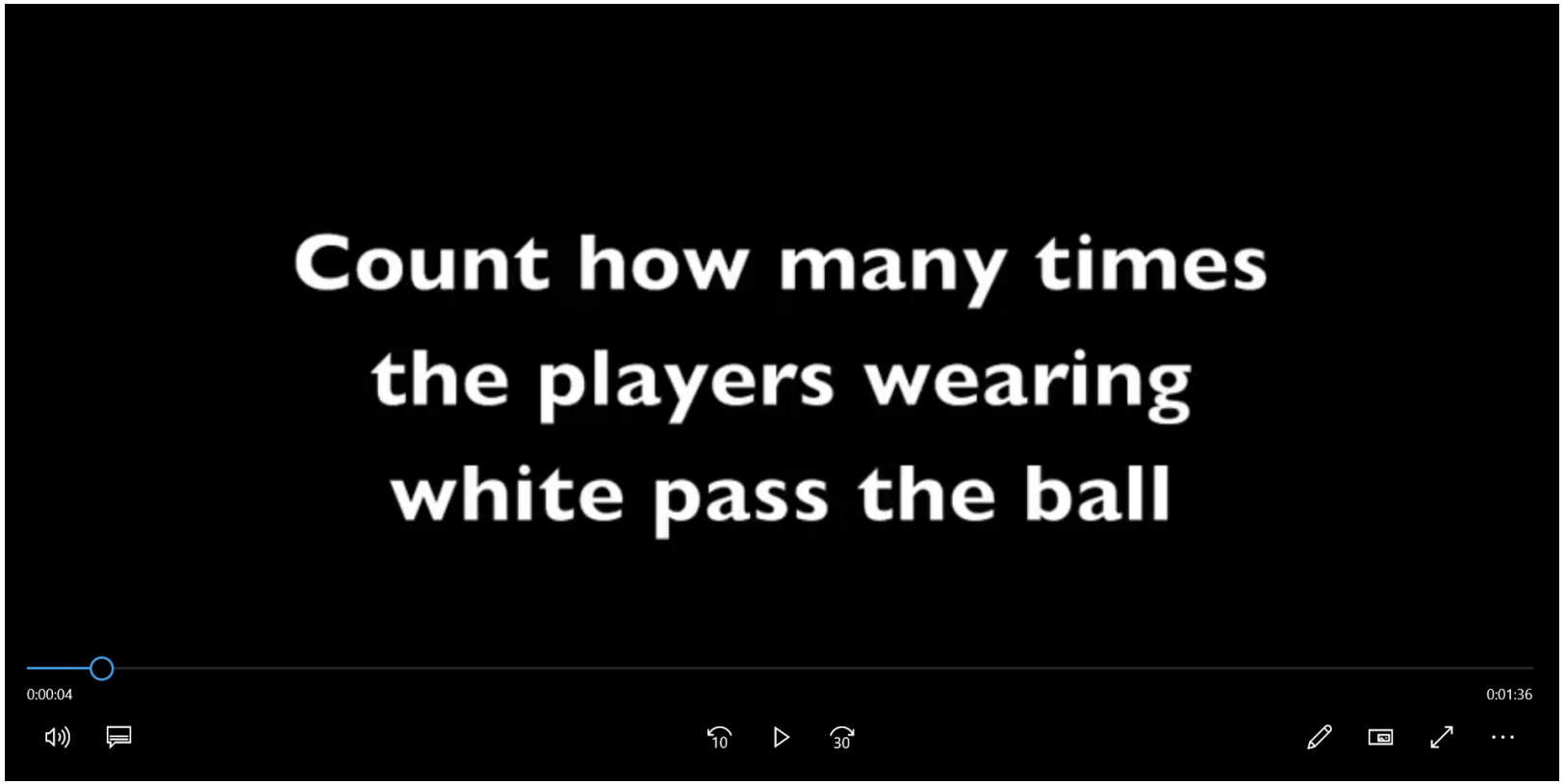
With $\geq 2$ features, we can't pre-attentively process reliably. Find the red circle!



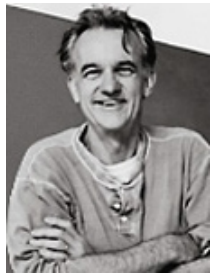


# Visual perception: why graphs work well, or not

What happens when our brains focus attention? Watch this video **carefully**, then answer this anonymous poll.



# Visual perception: using what you know



Some widely-quoted principles from Tufte – for improving graphics;

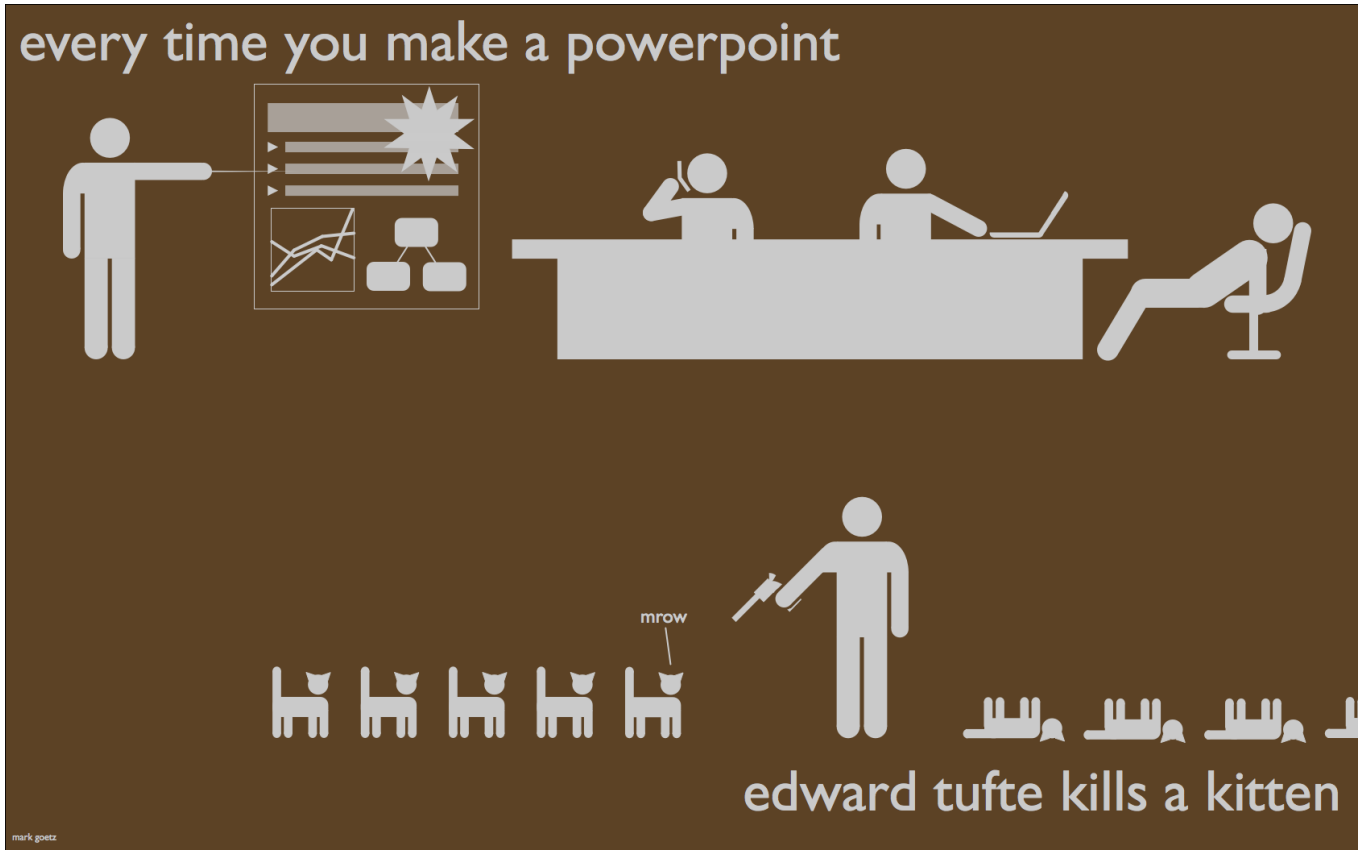
- Above all else, show the data

- Maximize the data-ink ratio (i.e. data ink / total ink)
- Erase non-data-ink (*chartjunk*)
- Erase redundant data-ink
- Revise and edit

These are reasonable guidelines – but don't say anything about what to focus on when editing. Tufte's minimalism also doesn't allow for (pragmatically) using methods familiar to your audience.

# Visual perception: using what you know

Be aware that Tufte (and his strong opinions) are sacrosanct, to some:



# Visual perception: using what you know

If/when choosing a graph gets difficult;

1. Think, fairly hard, about what you want to illustrate
2. Pick a graph you think codes it appropriately
3. *Explain it to someone* – yourself, at first – like you will do with your poster
4. Iterate steps 2 & 3 until convergence at a good solution

- At step 2, borrowing ideas from other people is **just fine** – there are no prizes for originality (unless you're doing datavis)
- If you get stuck, ask for help
- Some high-dimensional patterns are **just too complex** for 2D paper... but we rarely have enough data to say much about them

# Worked Example #1

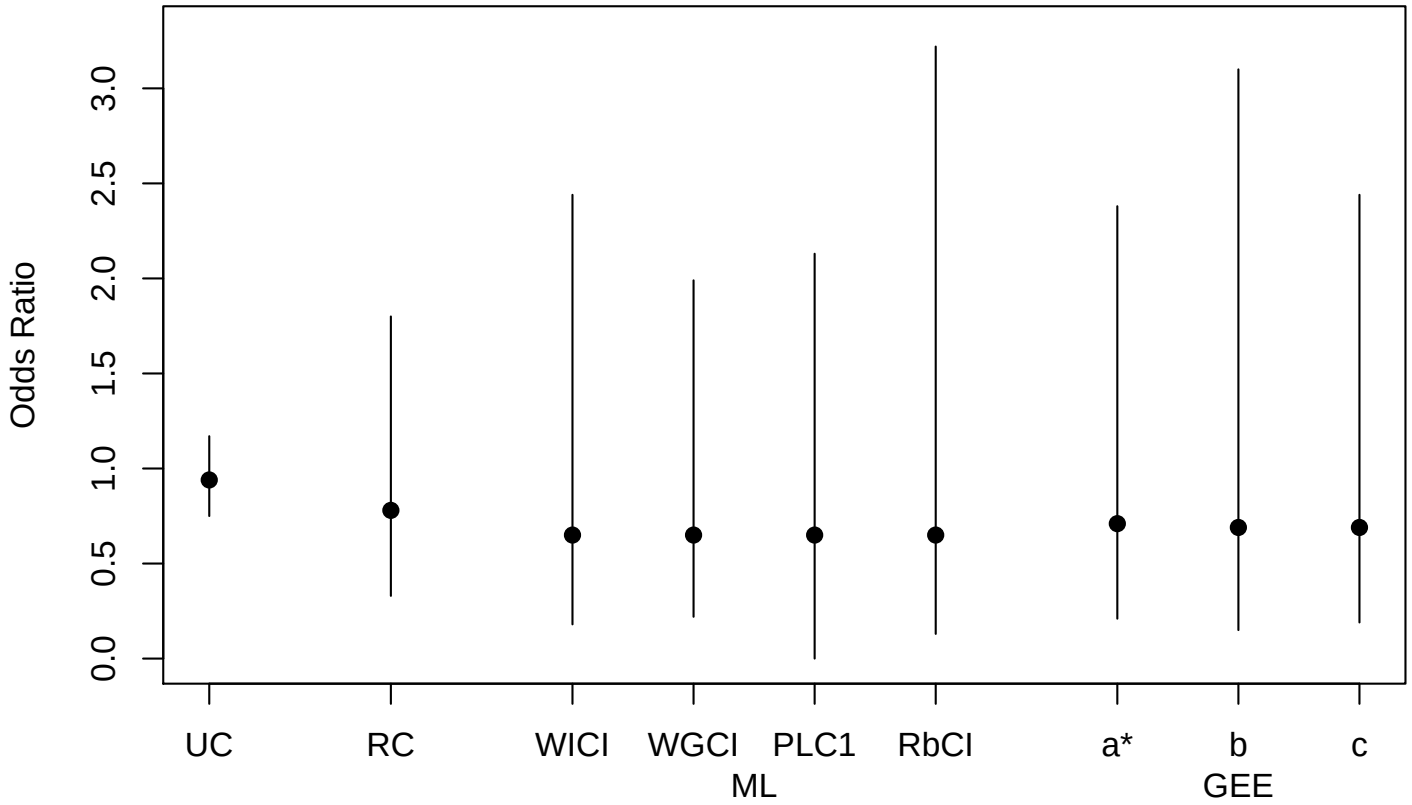
Broadly, statisticians like precision. So why not display precise results in tables?

In *Let's Practice What We Preach: Turning Tables Into Graphs*, Gelman *et al* (2002) compare tables for lookup...

| Method | $\widehat{OR}$ | 95% Interval |
|---|---|---|
| UC | 0.94 | 0.75–1.17 |
| RC | 0.78 | 0.33–1.80 |
| ML-WICI | 0.65 | 0.18–2.44 |
| ML-WGCI | 0.65 | 0.22–1.99 |
| ML-PLCI | 0.65 | 0.00–2.13 |
| ML-RbCI | 0.65 | 0.13–3.22 |
| GEEa*-RBCI | 0.71 | 0.21–2.38 |
| GEEb-RbCI | 0.69 | 0.15–3.10 |
| GEEc-RbCI | 0.69 | 0.19–2.44 |

# Worked Example #1

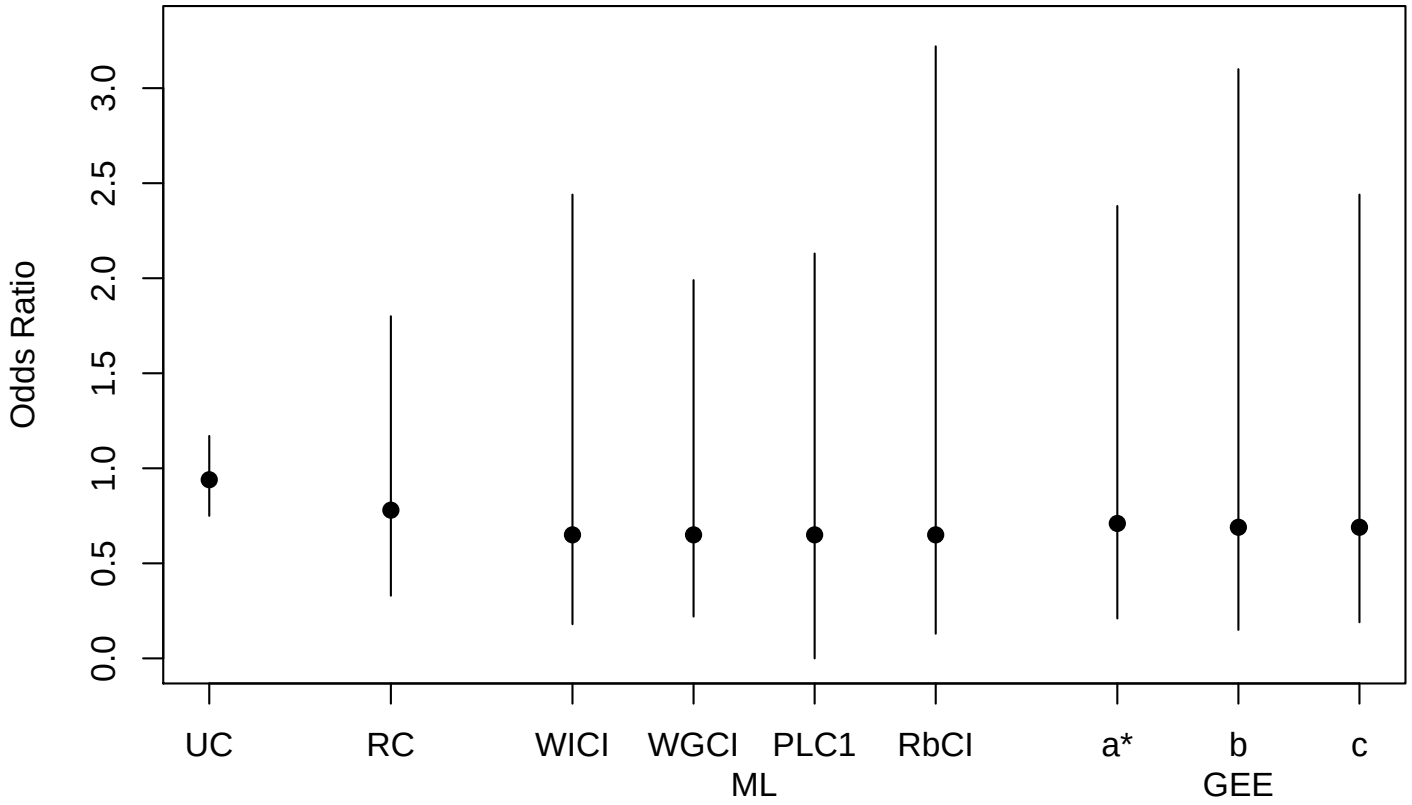
...to graphs, for making comparisons;




- Grouping helps (can also do this in tables)
- Comparisons are far easier, faster than in tables

# Worked Example #1

...to graphs, for making comparisons;




- Log-scale helps compare estimates *and* standard errors, in this case
- ... but zeroes require extra work, outside of any principles

# Worked Example #2

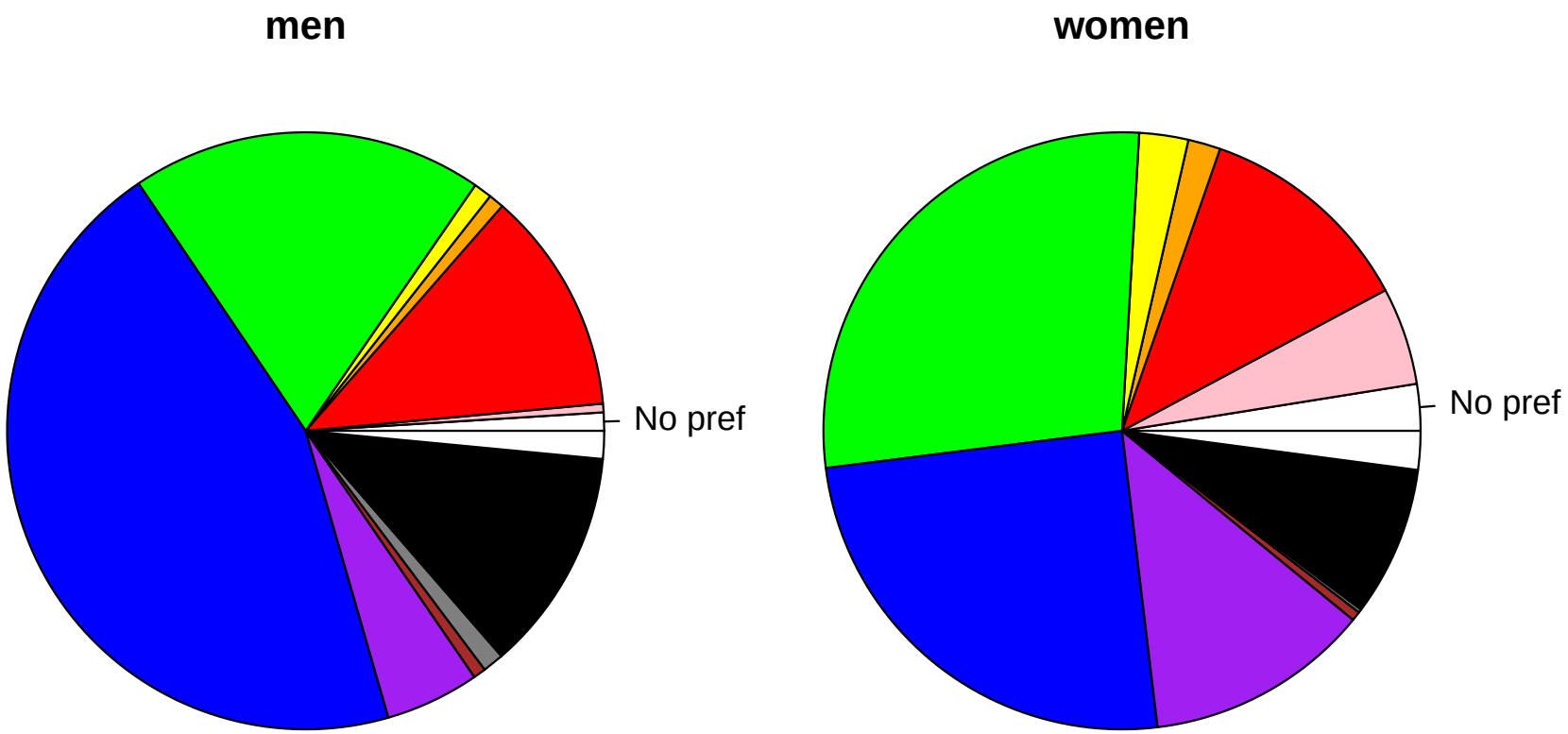
Some data on favorite color, published (!) as Ellis & Ficek 2001

| color | M | F | color | M | F |
|---|---|---|---|---|---|
| No pref | 19 | 95 | blue | 866 | 938 |
| pink | 9 | 199 | purple | 98 | 459 |
| red | 233 | 447 | brown | 13 | 19 |
| orange | 16 | 66 | grey | 22 | 7 |
| yellow | 19 | 100 | black | 233 | 306 |
| green | 367 | 1051 | white | 29 | 79 |

The authors are *"inclined to suspect the involvement of neurohormonal factors"* noting there are *"sex differences in retinal biochemistry and in how the brain processes color information"*.

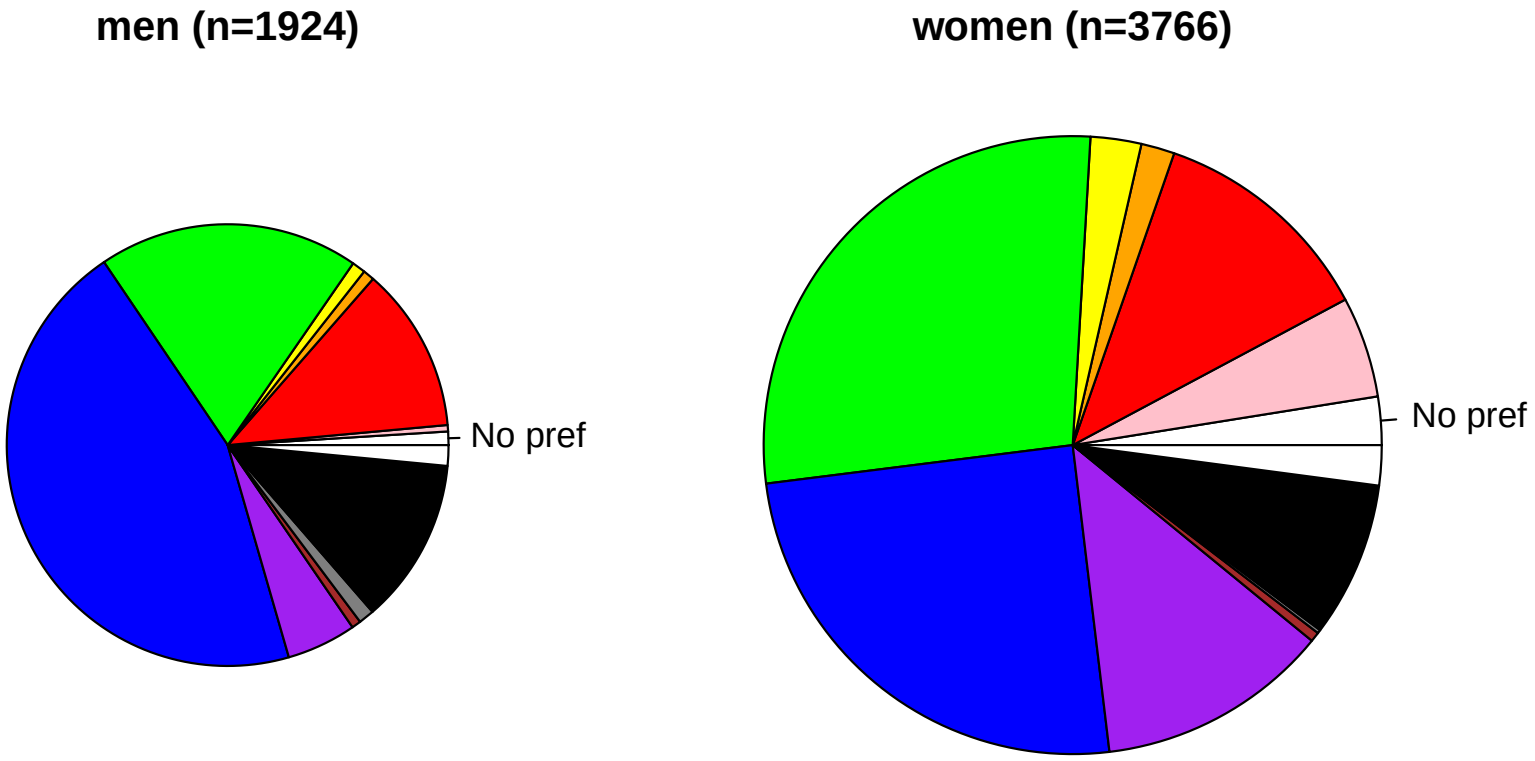
# Worked Example #2

A first attempt; no intervals, comparisons hard



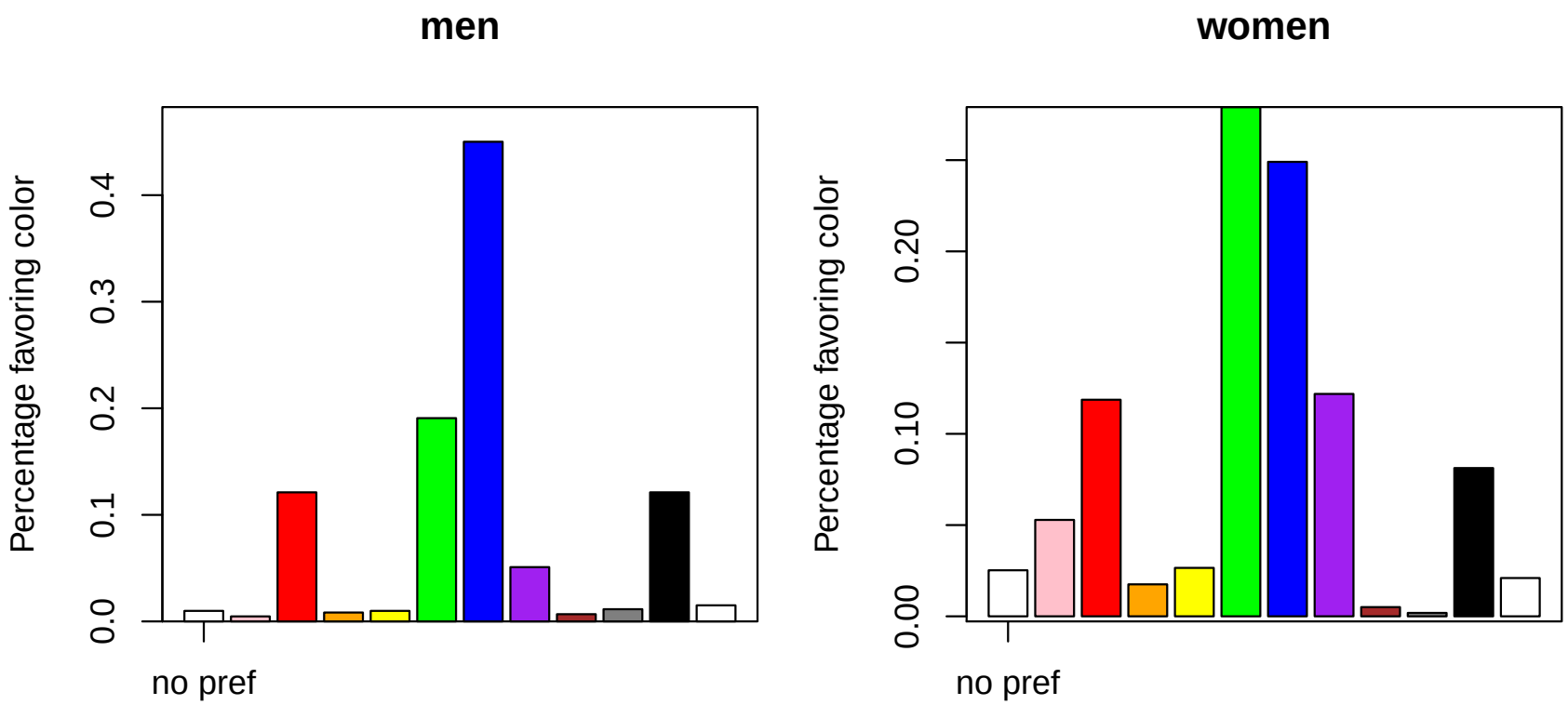
# Worked Example #2

With a *rough* attempt at intervals;



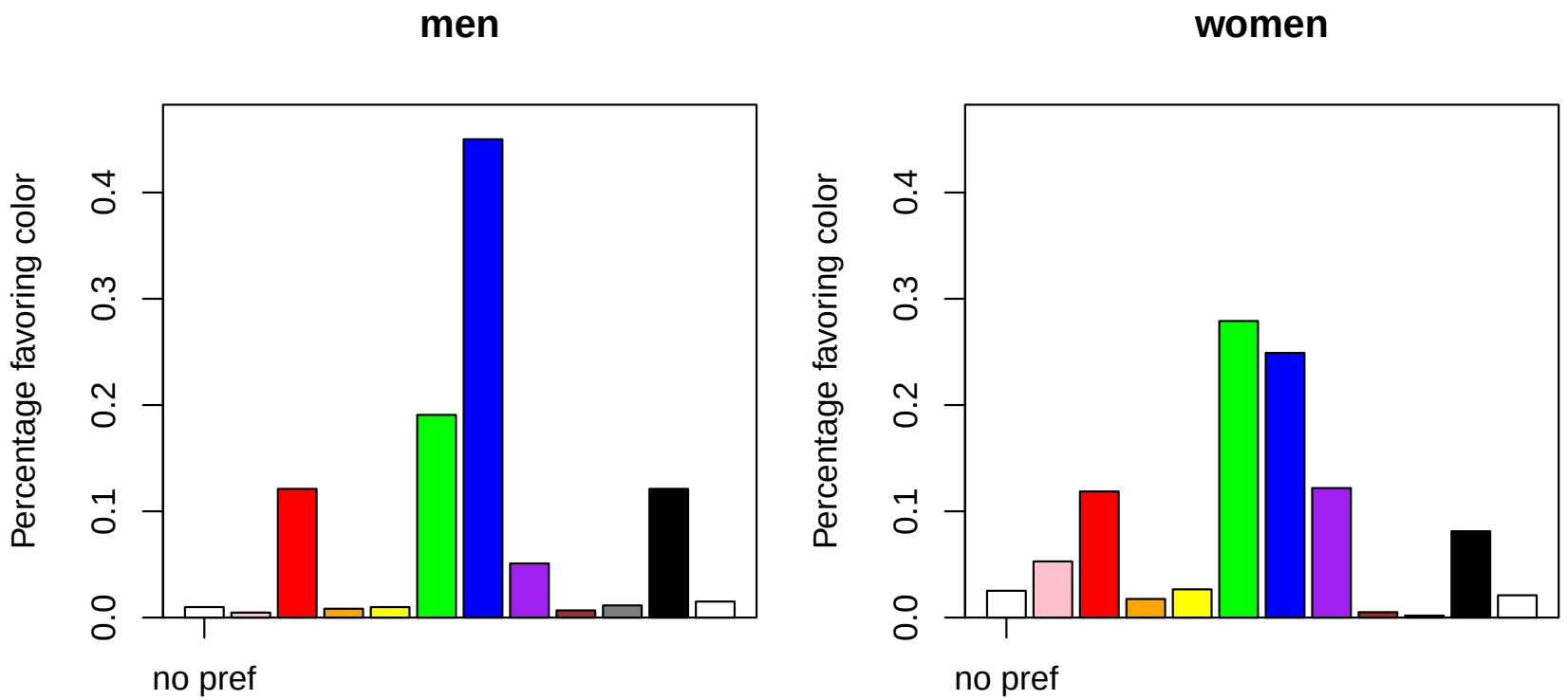
# Worked Example #2

Using position, not area/angle:



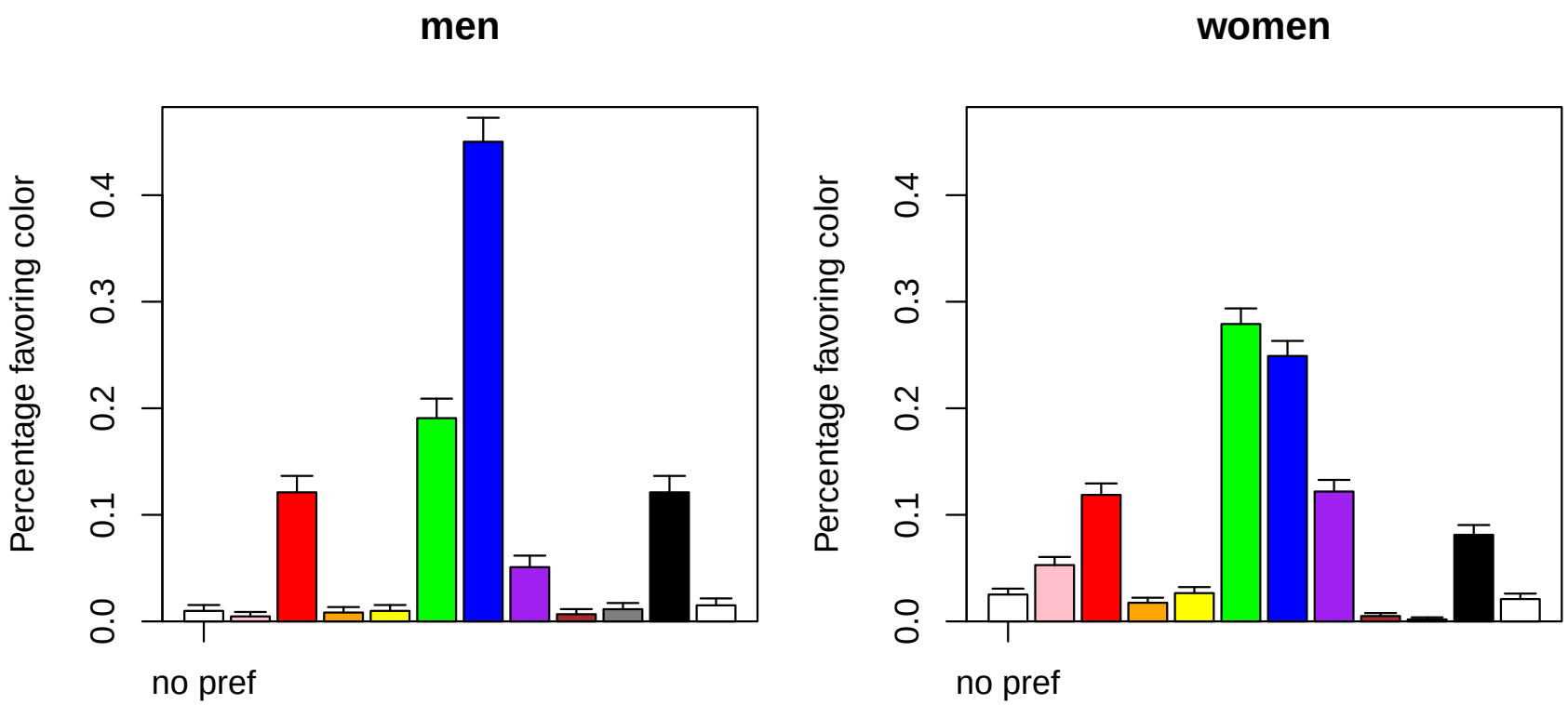
# Worked Example #2

Using position *on a common scale*, not area/angle:



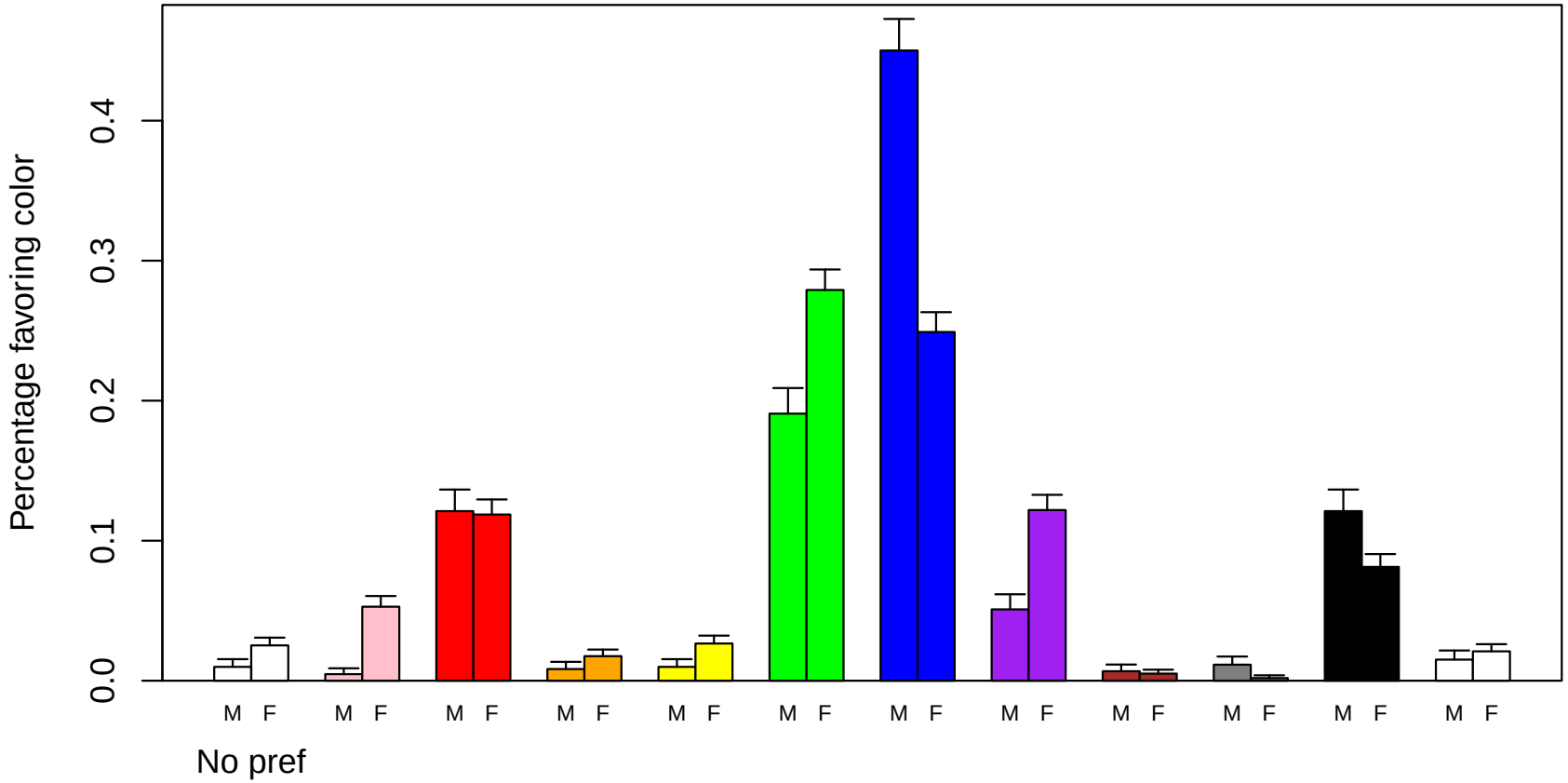
# Worked Example #2

Add uncertainty with antennae – use `segments()`. What does the eye compare?



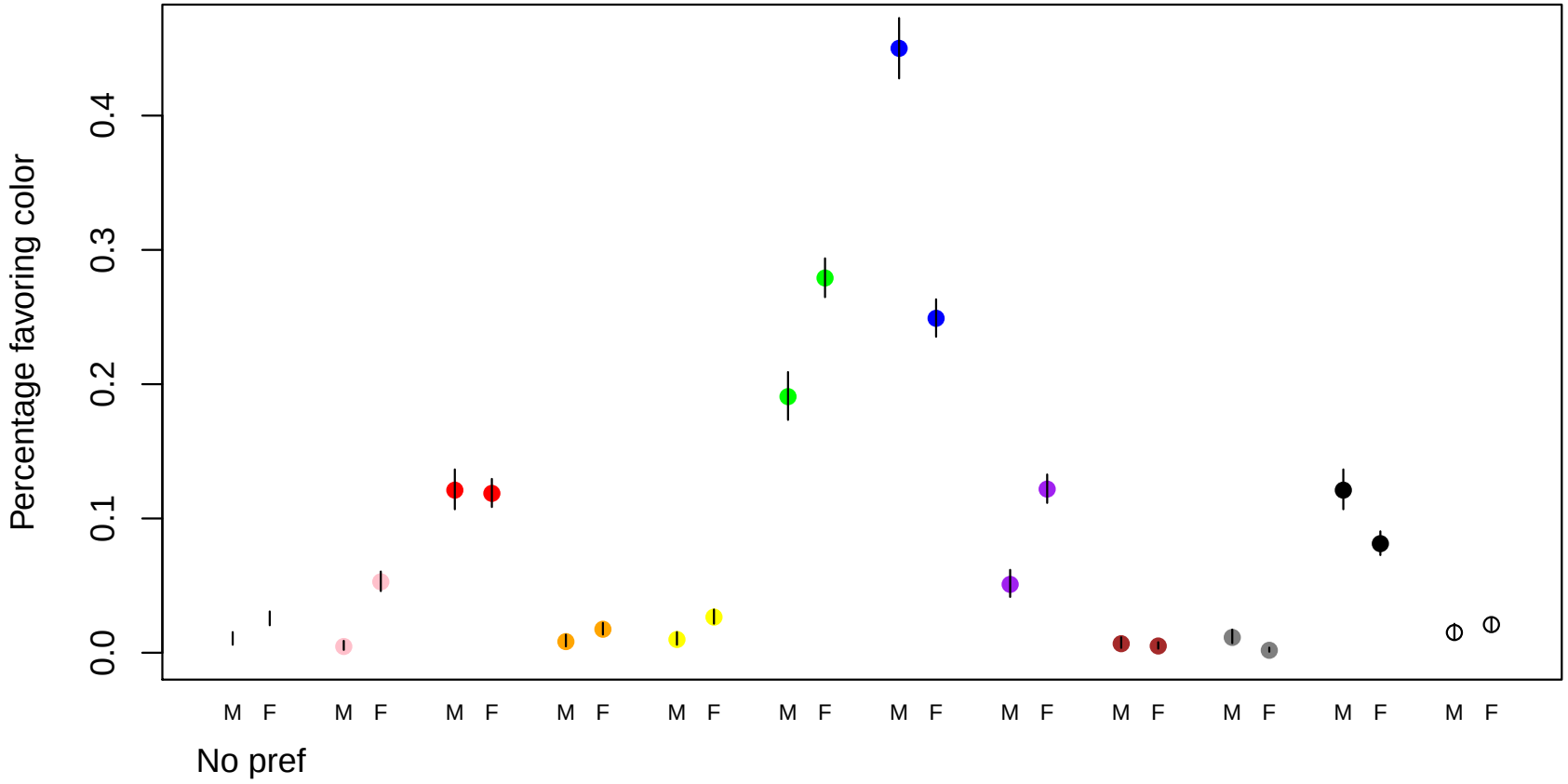
# Worked Example #2

Regroup to stress M/F differences for each color:



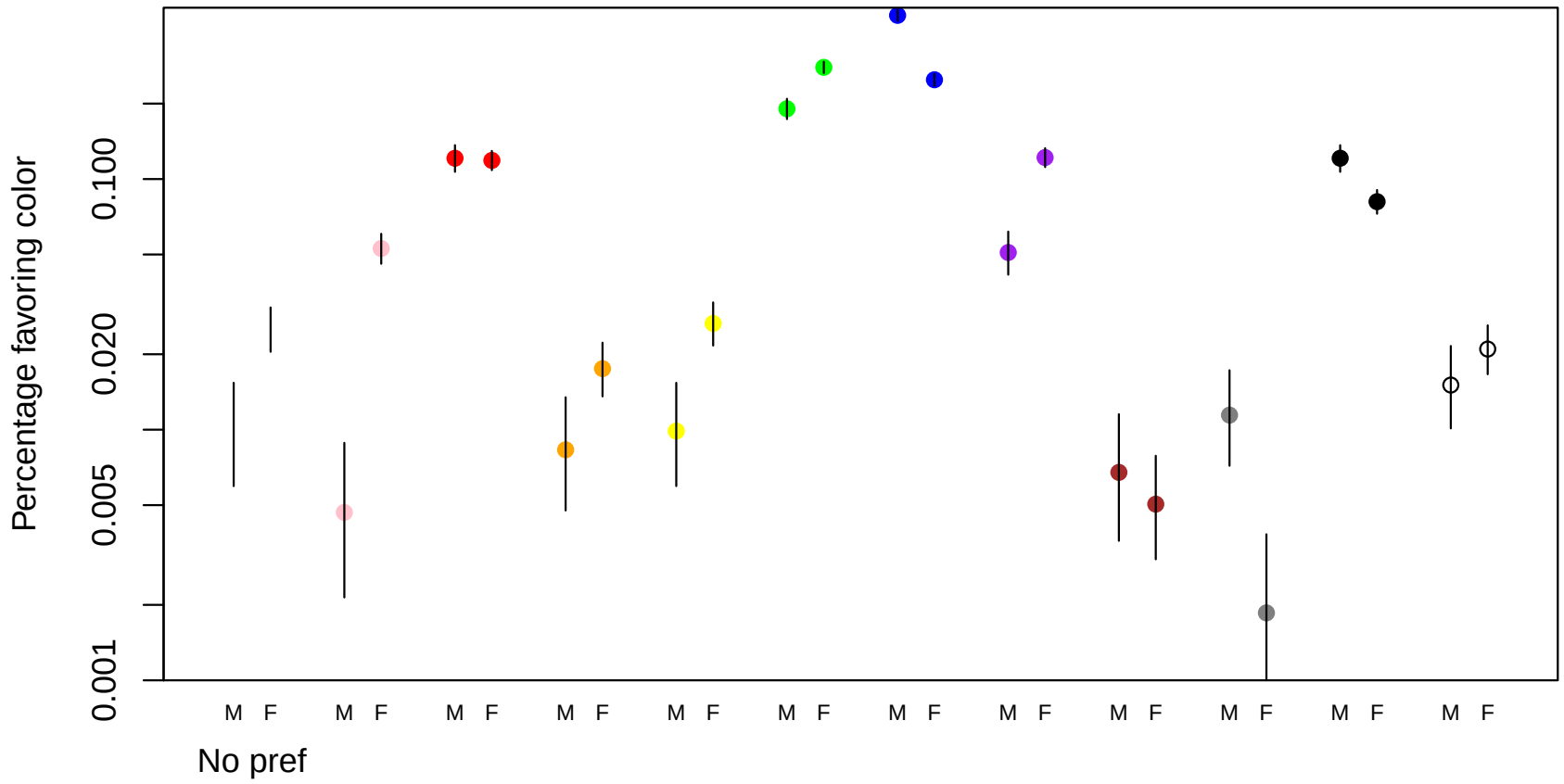
# Worked Example #2

Dump the blocks – use just position, not area. Makes CIs easier to see



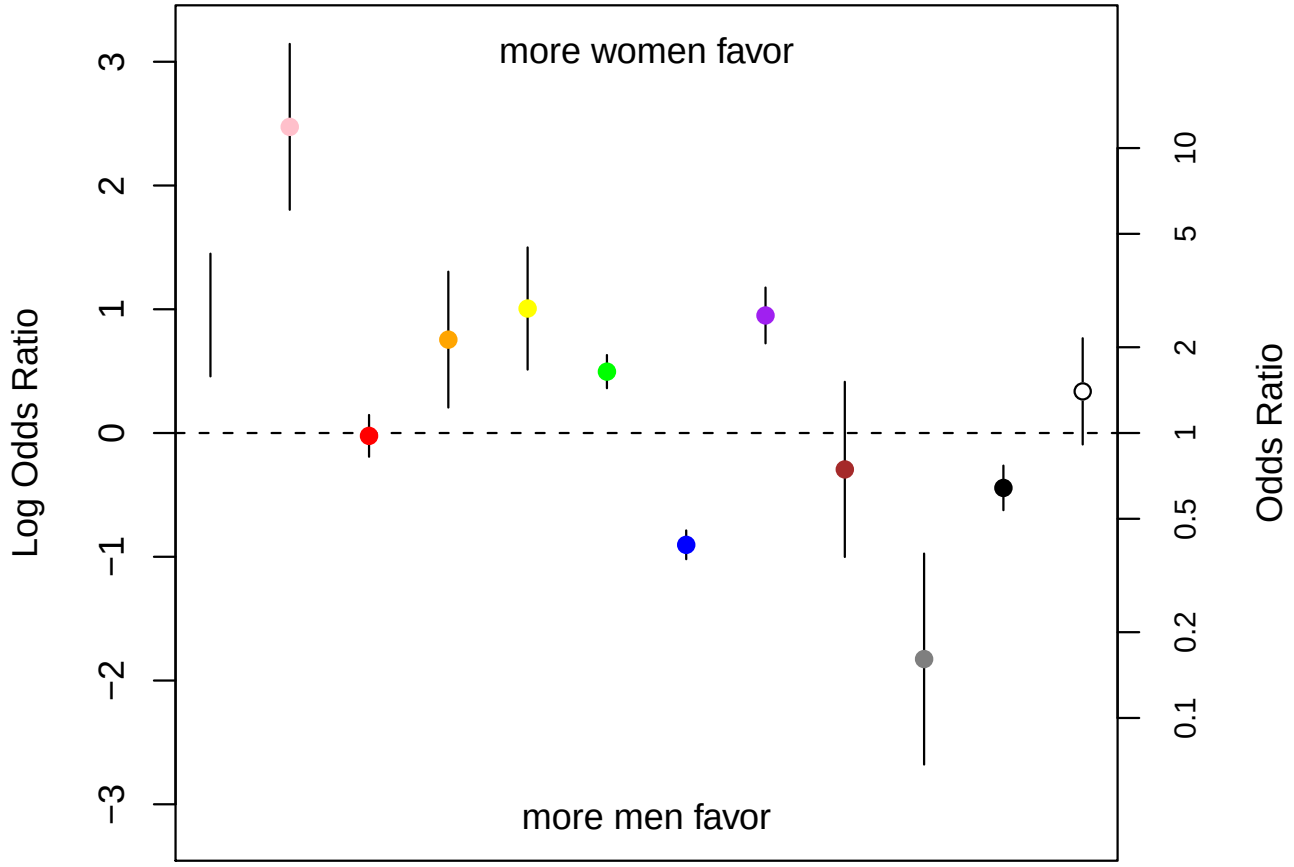
# Worked Example #2

Log-transforming the $y$-axis (`log="y"`) stresses less-popular favorite colors;



# Worked Example #2

To stress only differences, plot only differences; (baseline group irrelevant)



# Worked Example #2

Some lessons from all that;

- To compare items, put them beside each other
- Decide what you want to compare; differences or absolute values?
- Often it will be differences – e.g. regression diagnostics plot residuals, not data
- Minimalist representations (e.g. use of points not areas) are aesthetically 'clean' – and permit e.g. confidence intervals
- Plots will/should evolve, as you decide to stress different results
- Pie charts are rarely useful

# Worked Example #3

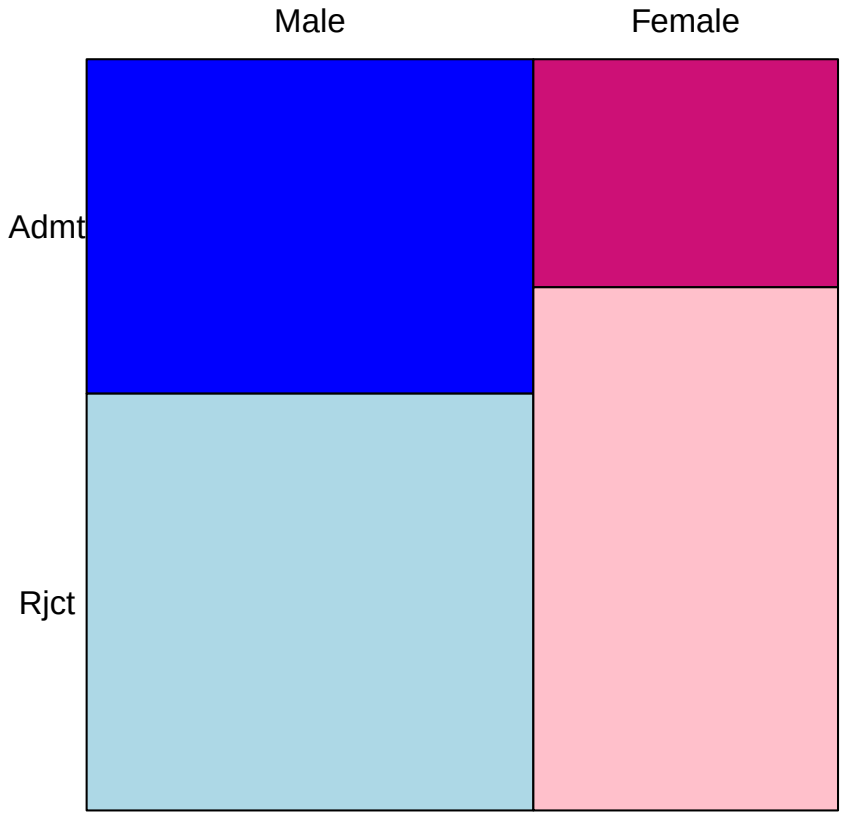
A dataset you saw in 514: in 1973, sex discrimination was suspected in admission to Berkeley;

| Dept | Men | | Women | |
|---|---|---|---|---|
| | $n$ | Admit | $n$ | Admit |
| A | 825 | 0.62 | 108 | 0.82 |
| B | 560 | 0.63 | 25 | 0.68 |
| C | 325 | 0.37 | 593 | 0.34 |
| D | 417 | 0.33 | 375 | 0.35 |
| E | 191 | 0.28 | 393 | 0.24 |
| F | 373 | 0.06 | 341 | 0.07 |
| Total | 2691 | 0.45 | 1835 | 0.30 |

– the 'headlines' compared 45% to 30%. How can we turn this table into a graph?

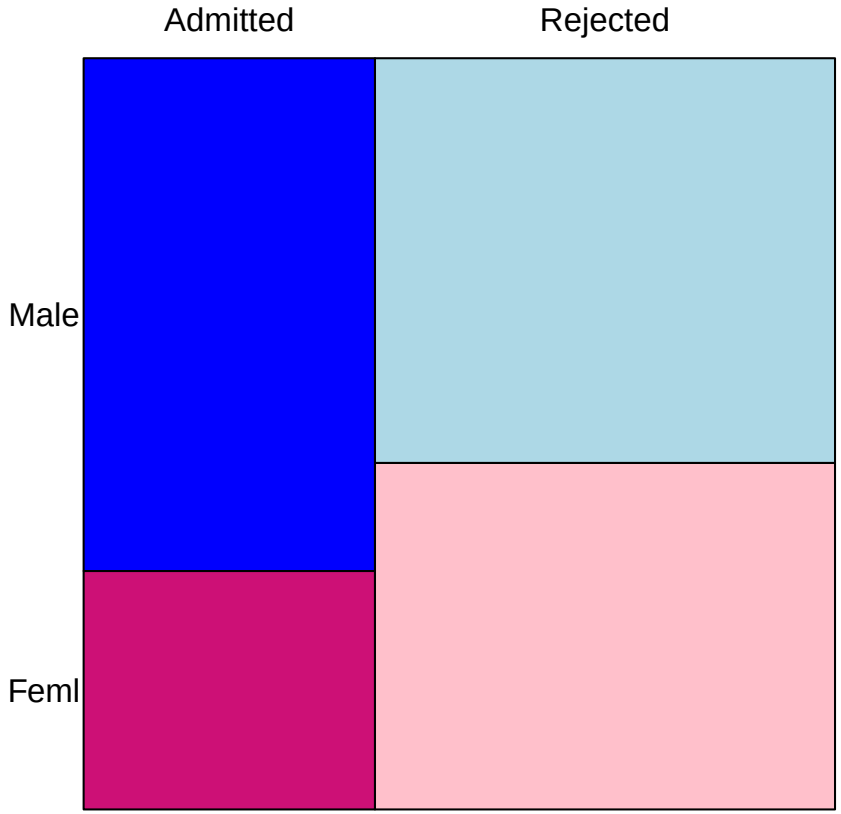
# Worked Example #3

Mosaic plots are a fairly ‘old school’ method...



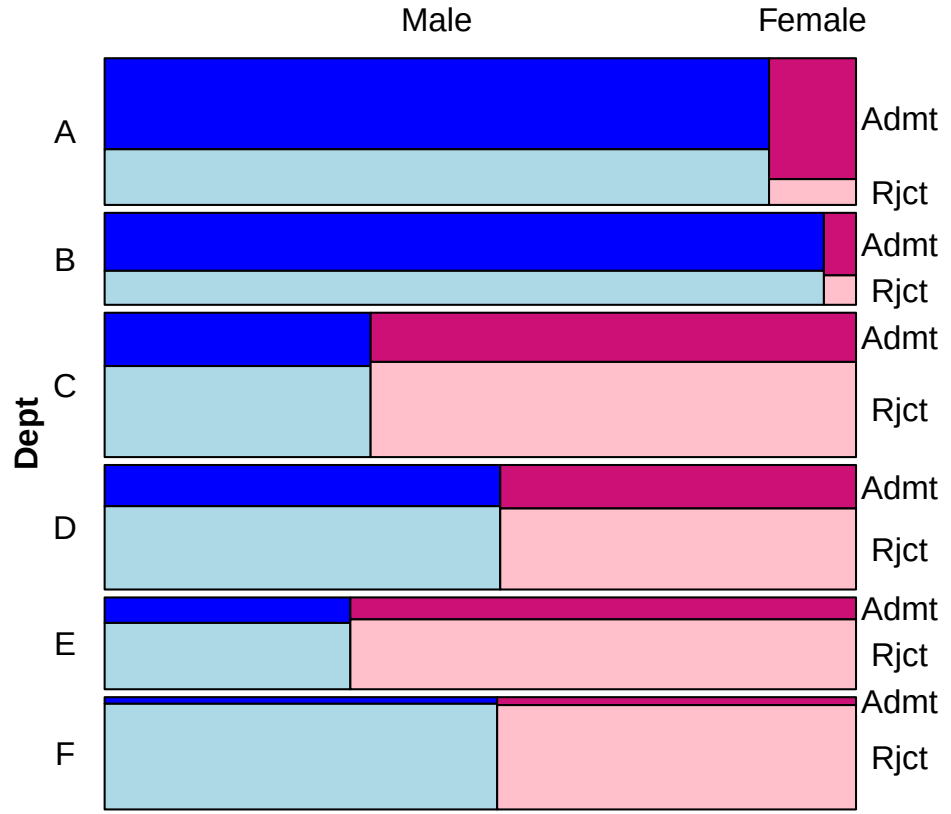
# Worked Example #3

...where conditioning matters;



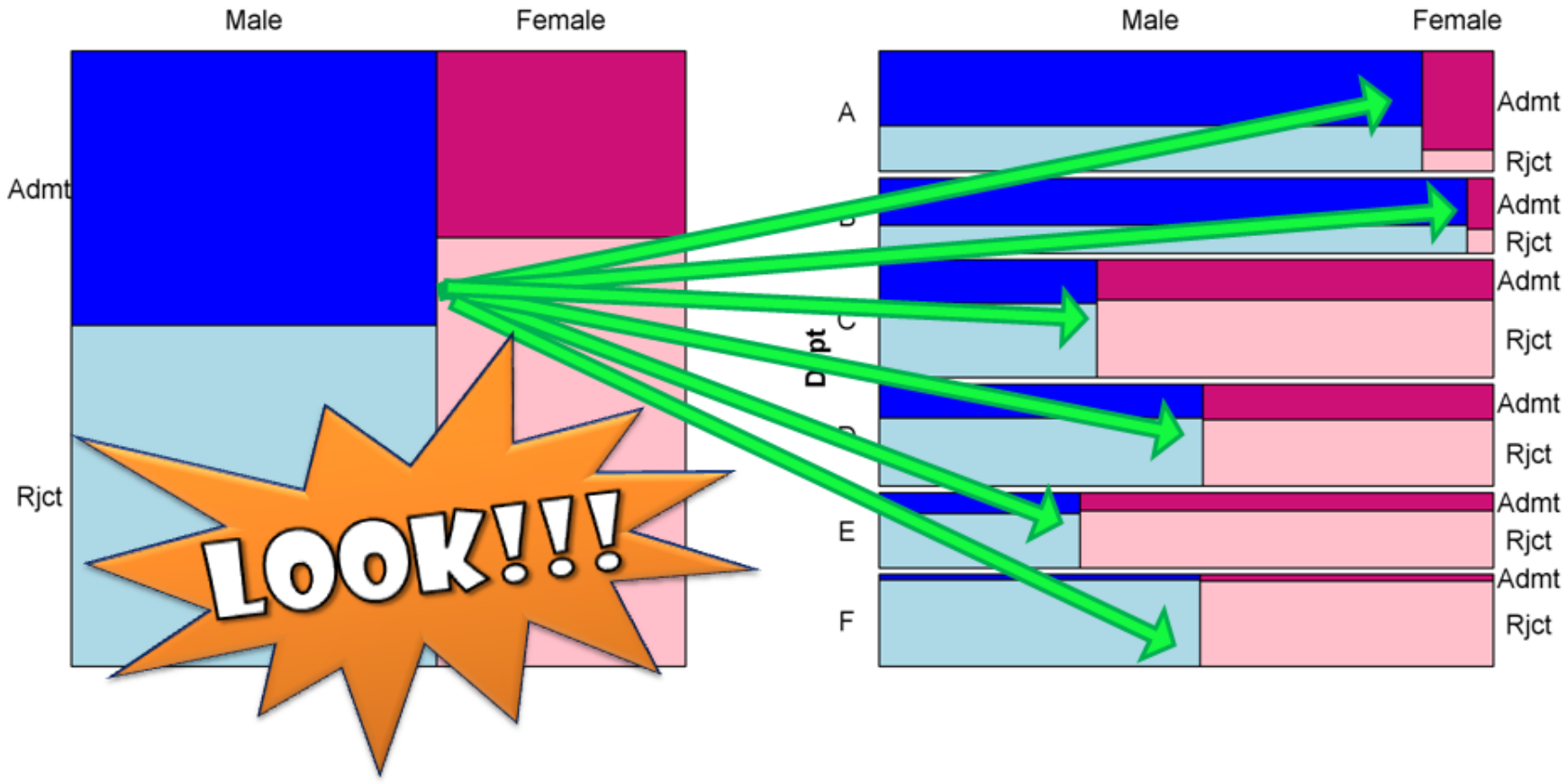
# Worked Example #3

Broken down by department;



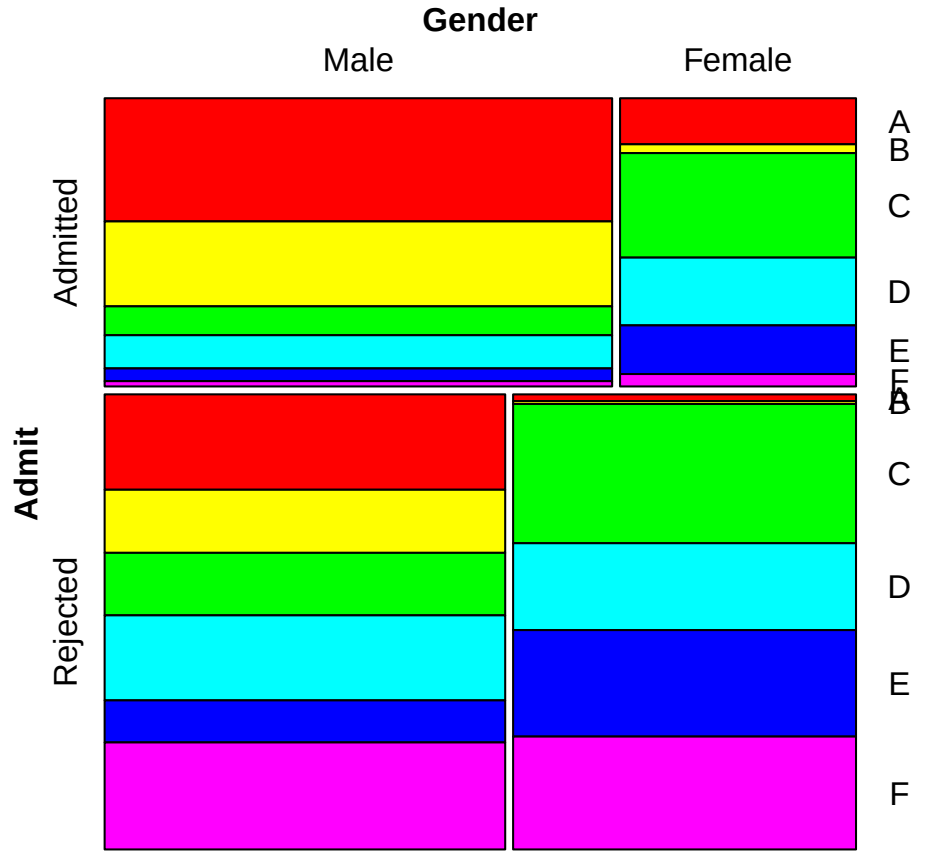
# Worked Example #3

In a talk, one can dramatize the difference;



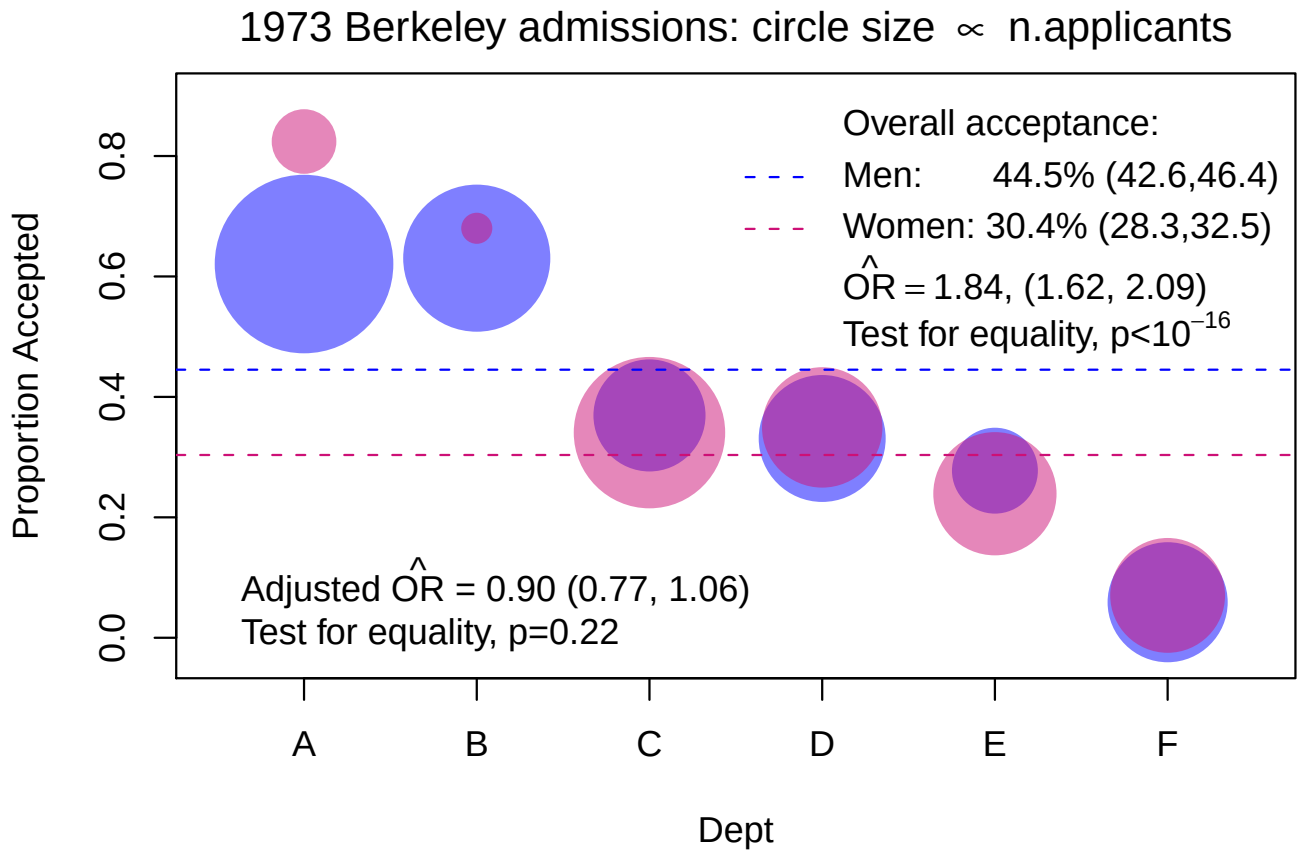
# Worked Example #3

...but this is hard to do on a single plot;



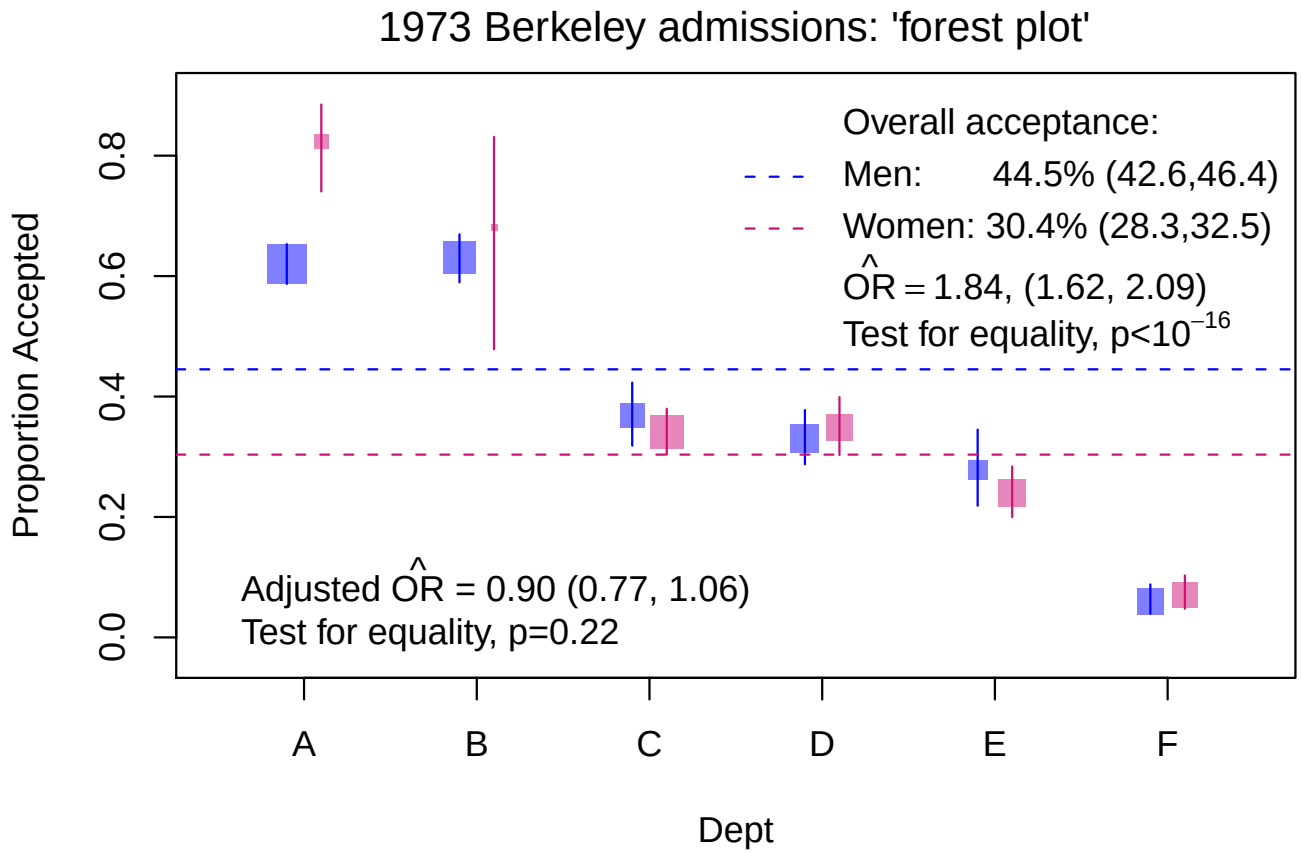
# Worked Example #3

Recall 'position on a common scale'/Tufte;



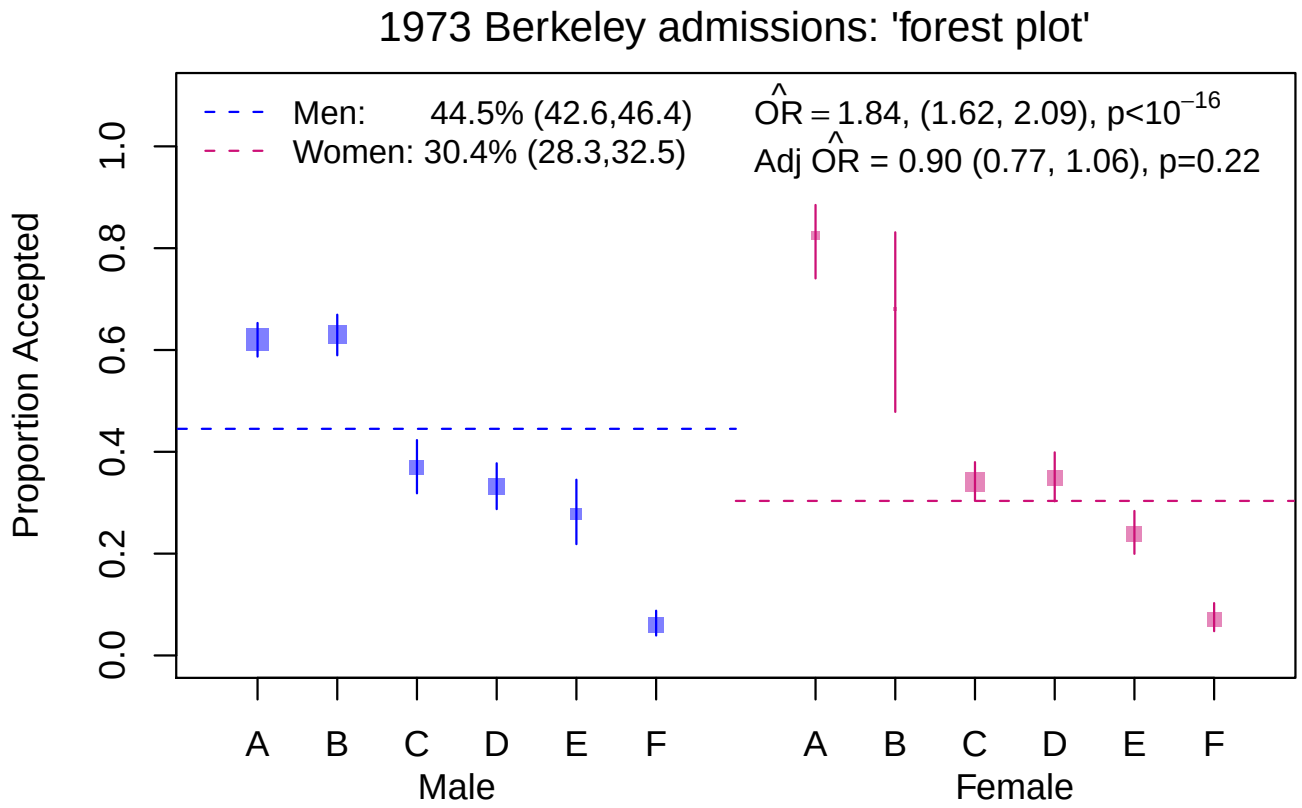
# Worked Example #3

Less ink – but confounding less obvious



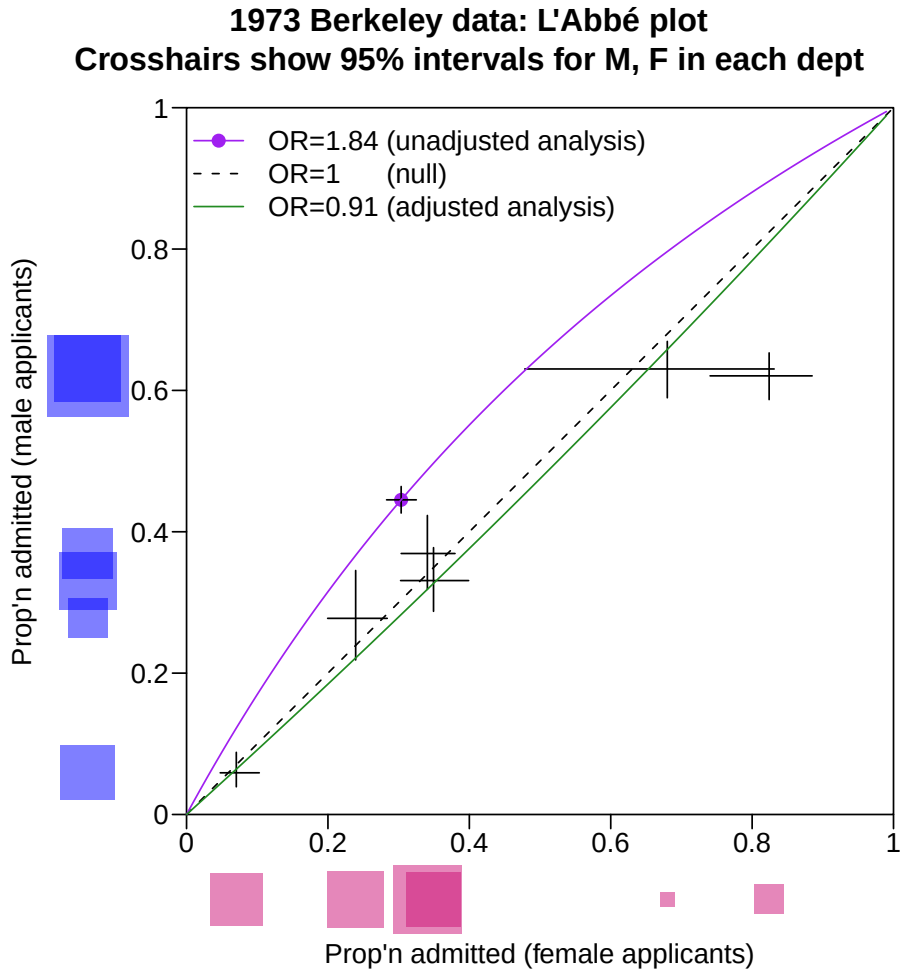
# Worked Example #3

Berkeley-wide comparison of admittance;



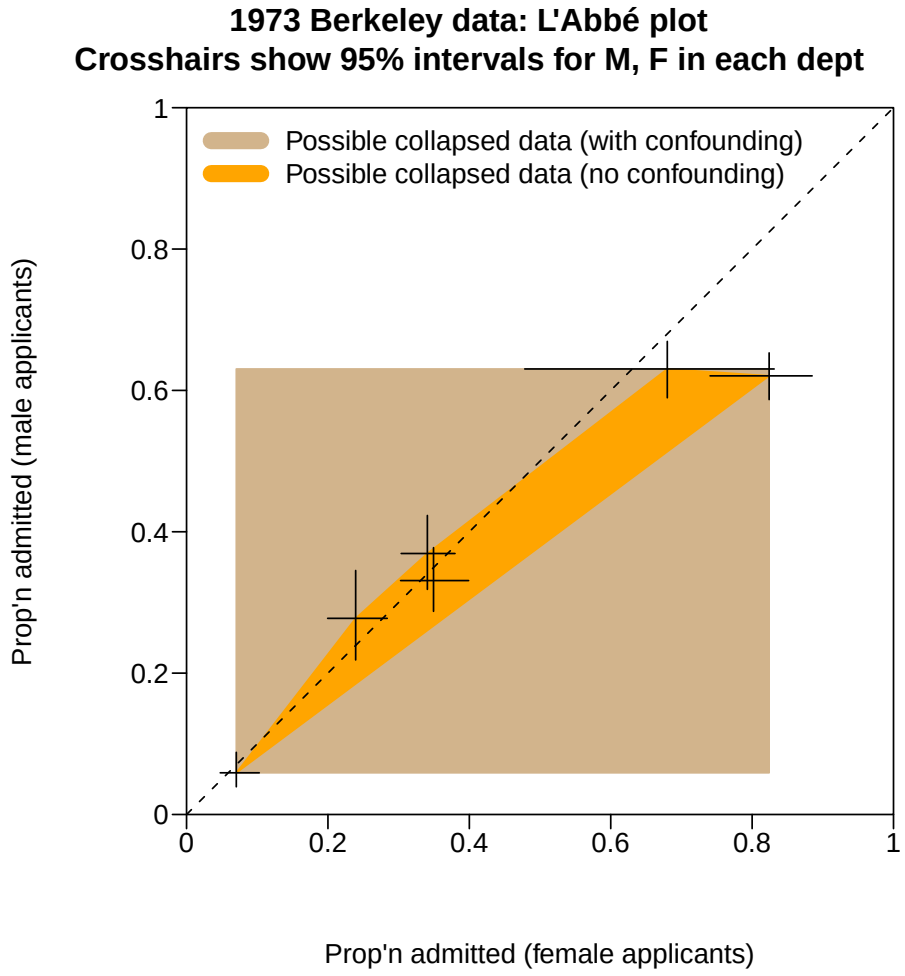
# Worked Example #3

Removing the irrelevant A/B/C ordering;



# Worked Example #3

For discussions of confounding and/or collapsibility;



# Worked Example #3

Some lessons from all that;

- Do you want to compare counts, or proportions? Which is (most) relevant?
- L'Abbé plots (the last version) are a great way to illustrate *just* proportions, in two groups – although they are unfamiliar to some audiences
- Non-collapsibility was for *decades* viewed as weird and non-intuitive – see "Simpson's paradox". With the right graph it's straightforward to see it happening

# Your turn!

- Each breakout room receives one dataset, and a short description of what aspect of it to illustrate
- Take 5 minutes to read the documentation and look at the data
- With your group, discuss what graphics might be effective
- Implement what you think will work
- Be ready to show it to the class, and explain what's good/bad about it

# Special topics: graphs for simulation studies

Simulations studies are very common in methods work;

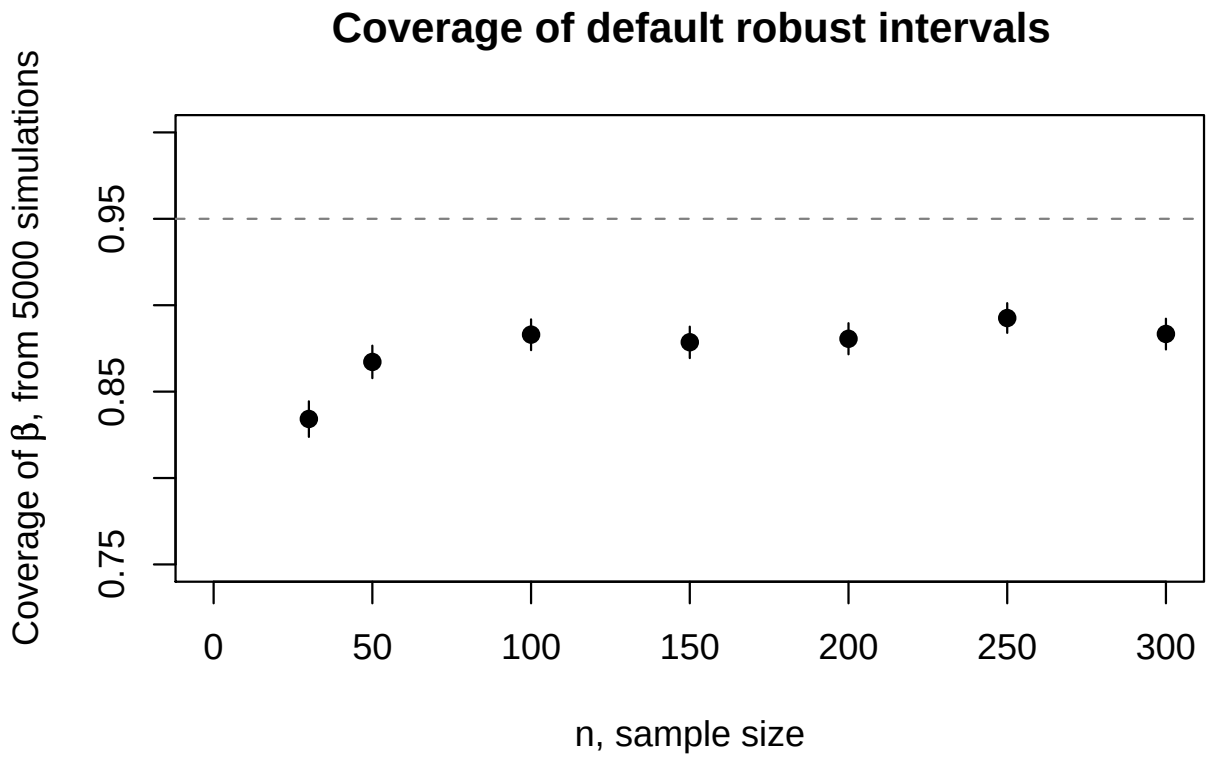
- Tables of estimated coverages (all near 95%) are very common
- Tables of estimated coverages (all near 95%) are **immensely boring**
- Showing the Monte Carlo error can be a challenge

A game for seminars; before the speaker tells you, decide whether they will say;

- "Look how **different** these lines are – **and** mine is best!"
- "Look how **similar** these lines are – **but** mine is best!"

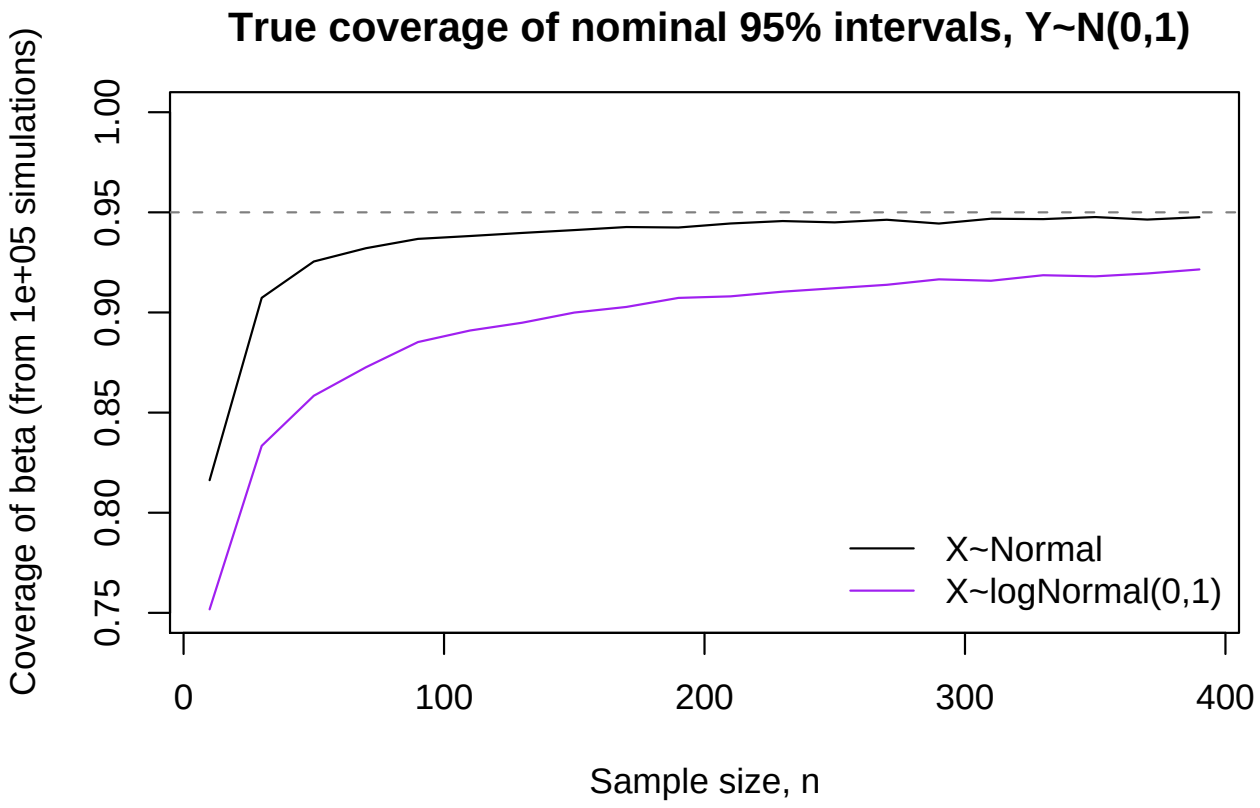
# Special topics: graphs for simulation studies

A typical simulation example; (shows impact of violating regularity conditions)



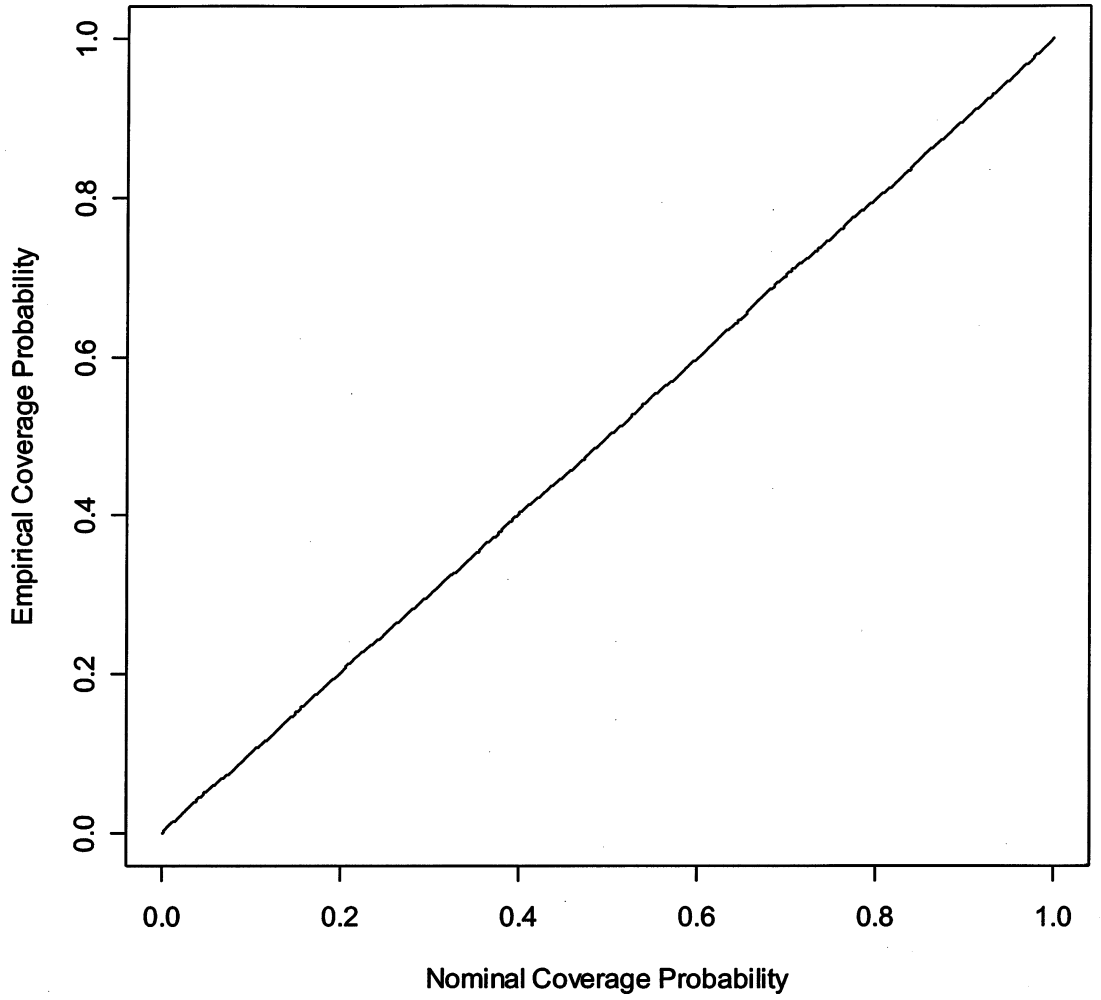
# Special topics: graphs for simulation studies

A typical simulation example – that has negligible Monte Carlo error;



# Special topics: graphs for simulation studies

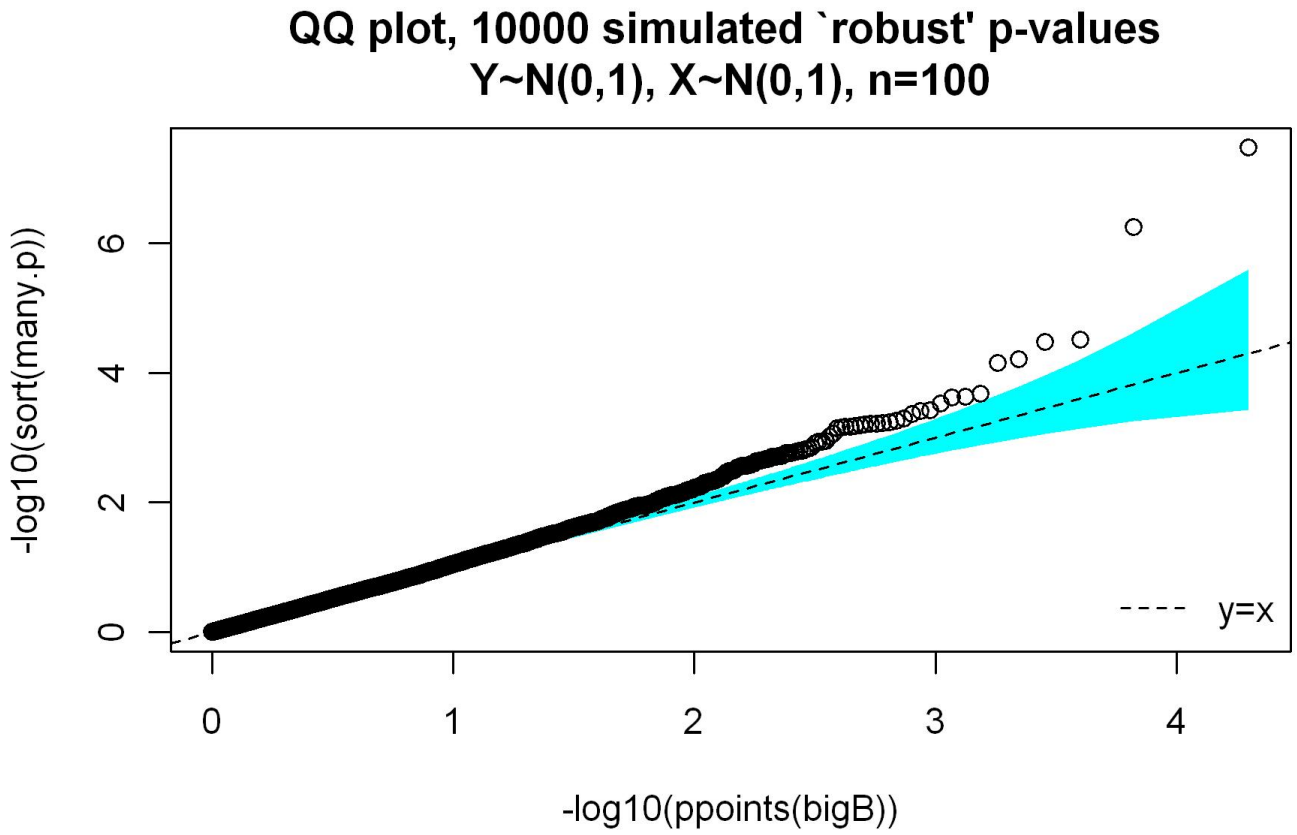
Here's an unhelpful display of a simulation's many $p$-values;




Epstein MP, Satten GA (2003) Inference on haplotype effects in case-control studies using unphased genotype data. *Am Jrnl Hum Genet* 73:1316-1329

# Special topics: graphs for simulation studies

Here's a better one – why is it better?



# Special topics: graphs for simulation studies

Be clear about the point of the simulations:

- It **could be** to show that **with some** $n$ a method controls Type I error rates $\leq 0.05$, or has coverage $= 0.95$. The FDA, for example, would really care about this when approving a trial
  $\Longrightarrow$ *fix the relevant* $\alpha$ *and compare it to dichotomized results*
- It could **instead** be to show how/where the asymptotics break down. For example, with fixed $n$ can we trust results when $\alpha = 0.01$? How about $\alpha = 0.001$ or $10^{-8}$? Accuracy may also depend on the **extent** of assumption violations, e.g. how **much** homoskedasticity is present
  $\Longrightarrow$ *show results varying quantitatively with relevant factors*

Factors **tend** to act linearly on $Z^2$, not $p$, so transforming T1ER/coverage to that scale (approximately) often helps – see $-\log_{10}(p)$ on previous slide.

# Special topics: graphs for simulation studies

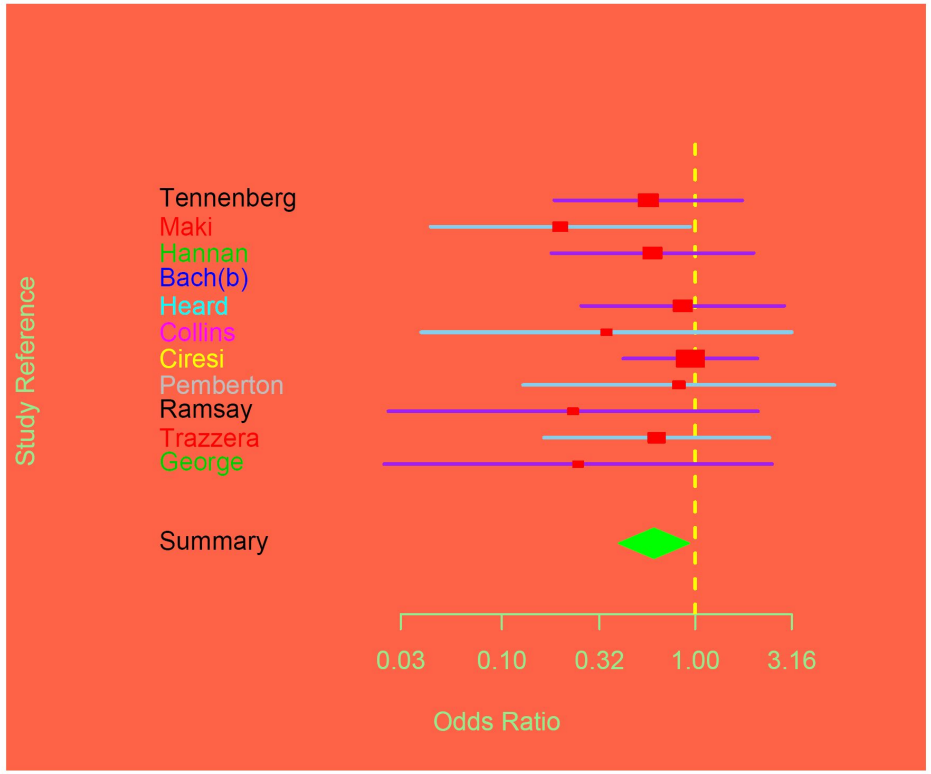
A rule of thumb: ignore any simulation results based on $< 10$ observations, e.g. 10 non-covering simulated CIs, or 10 simulated Type I errors. Why?

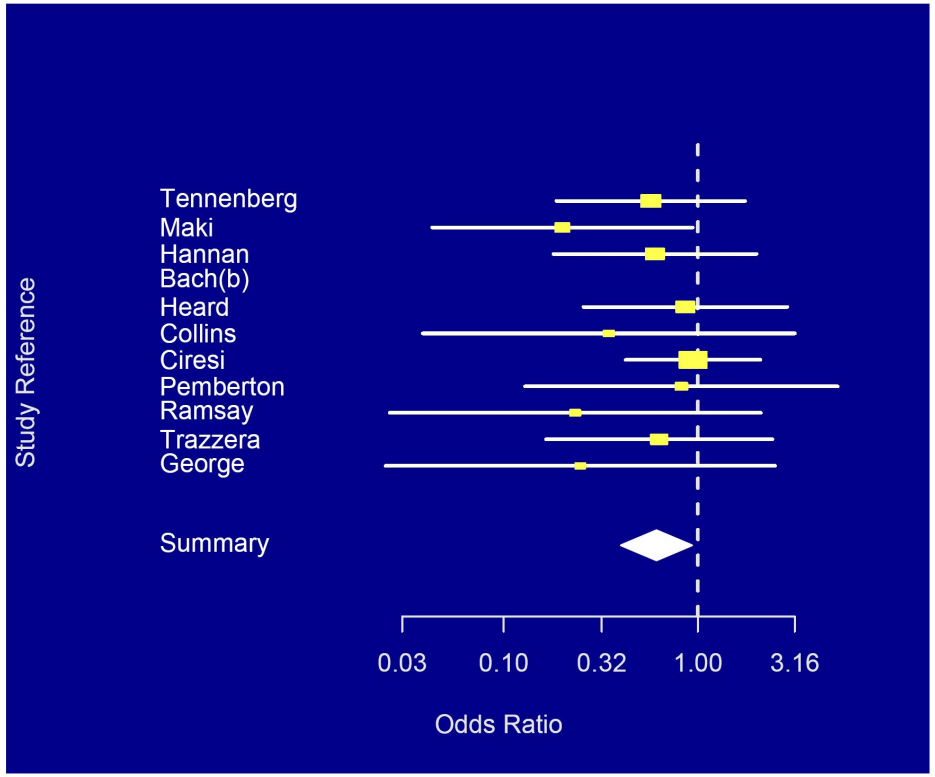
- At this value, it's reasonable to believe the true error rate within half/double the estimated value (see table, right)
- Similar to rule of thumb requiring $n \geq 20$ for estimating a mean, assuming no wildly-heavy tails
- Not getting 10 observations? Run more simulations!
- If you never get 10 observations, beware coding errors and/or unhelpful simulation settings

| $r$ | Exact 95% CI with $10/10^r$ successes |
|---|---|
| 2 | $(0.5, 2)\times 10^{-1}$ |
| 3 | $(0.5, 2)\times 10^{-2}$ |
| 4 | $(0.5, 2)\times 10^{-3}$ |
| 5 | $(0.5, 2)\times 10^{-4}$ |
| 6 | $(0.5, 2)\times 10^{-5}$ |
| 7 | $(0.5, 2)\times 10^{-6}$ |
| 8 | $(0.5, 2)\times 10^{-7}$ |

# Special topics: choosing colors

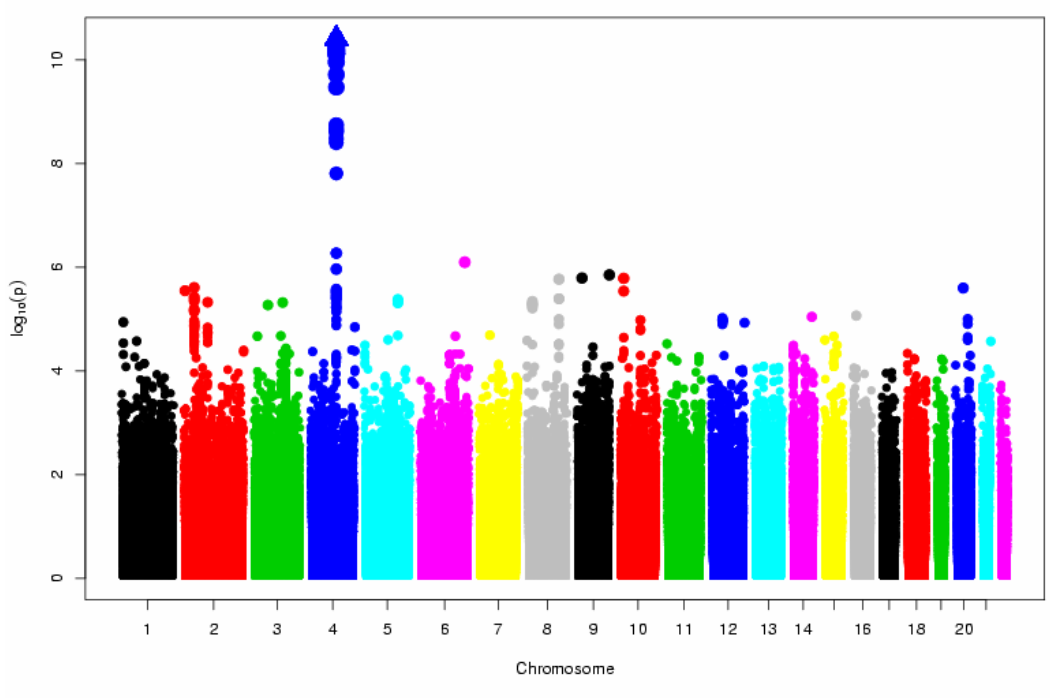
The choice is not just 'does it look cool'?

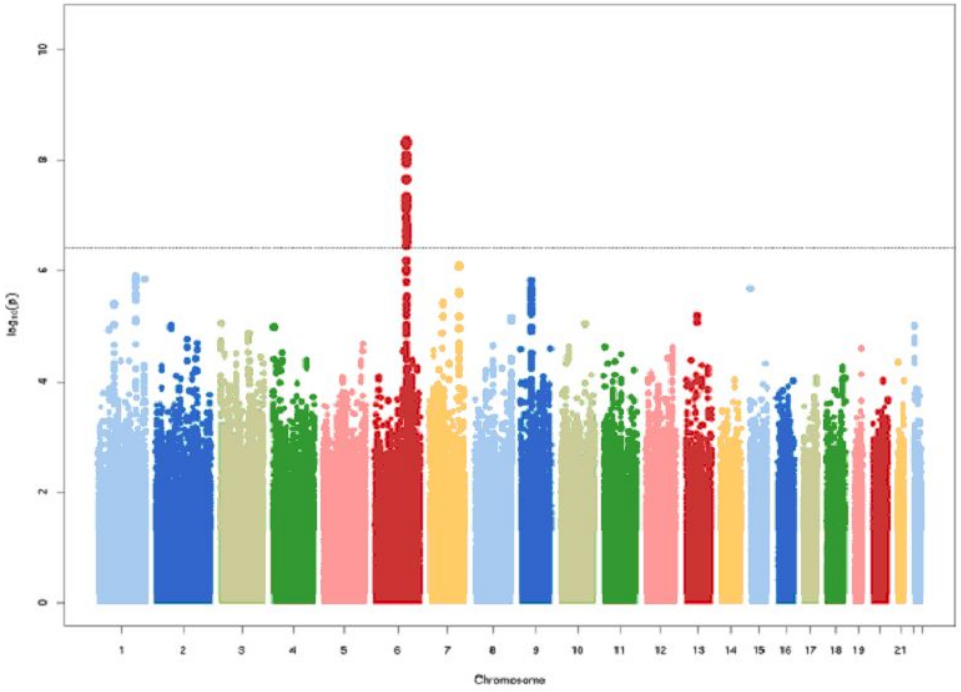


# Special topics: choosing colors

Two 'Manhattan plots', showing $-\log_{10}(p)$ for many multiple tests. Which blobs of color stand out?



# Special topics: choosing colors

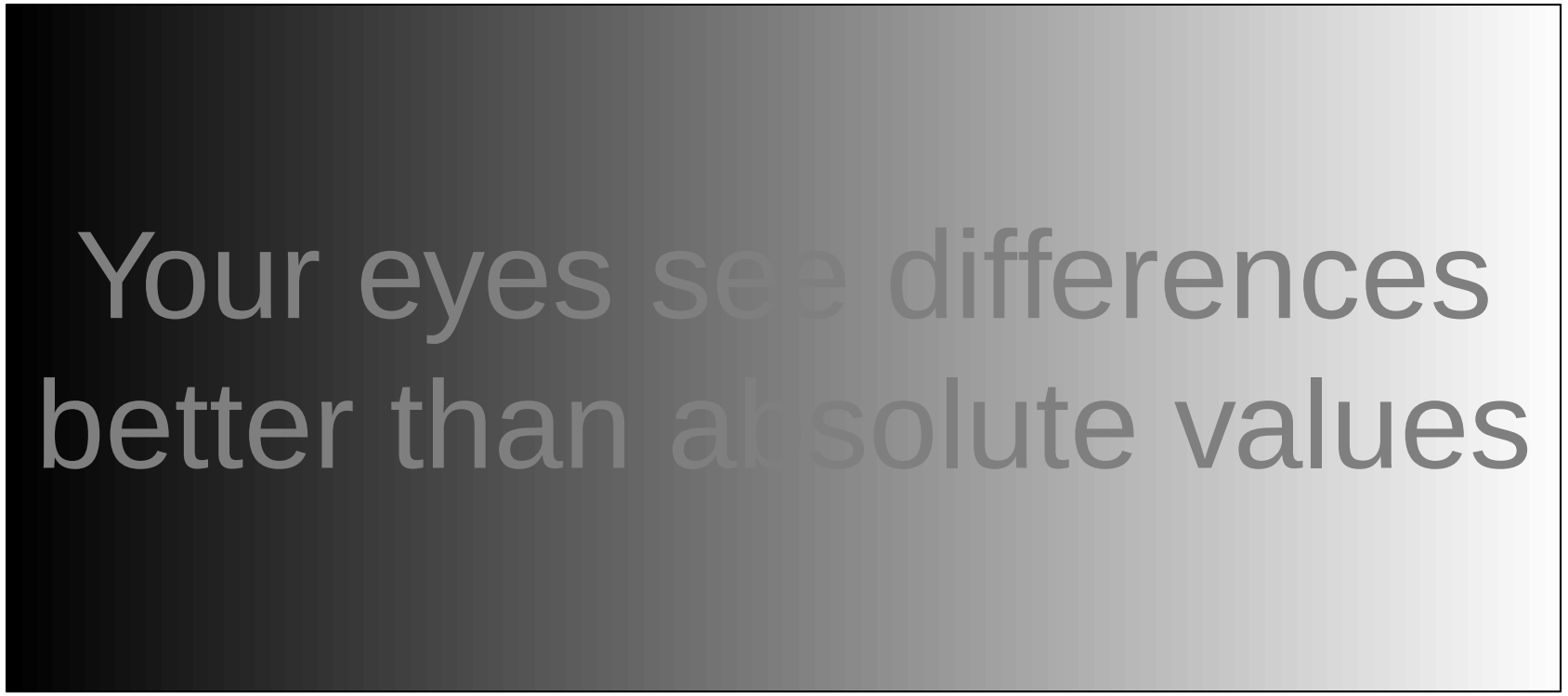
Why? Because...

Your eyes see differences better than absolute values

... and this applies in any color

# Special topics: choosing colors

Why? Because...

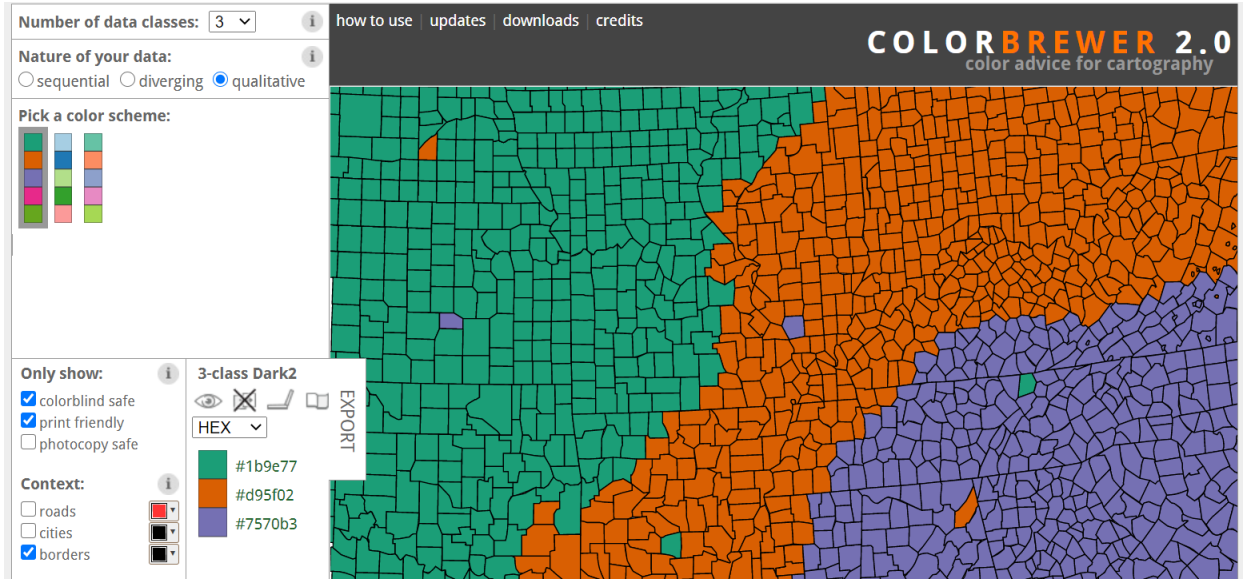



... and light/dark is more obvious than e.g. red/blue

# Special topics: choosing colors

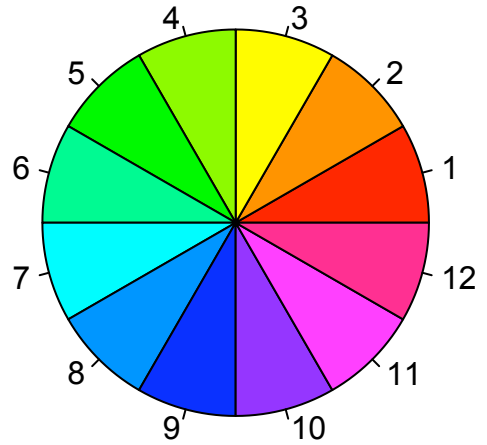
As well as making symbols/text large enough (see BIOST 514, Week 5 Discussion Section) using clearly-contrasting colors will **do most** to help your audience.

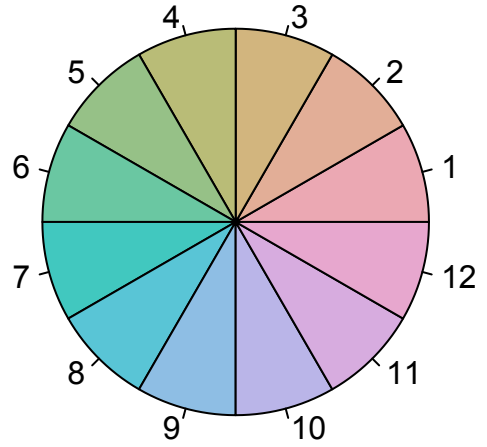
With two colors, check contrast here. When multiple colors are needed, go to ColorBrewer – and R's `RColorBrewer` package – that have color schemes designed for the National Cancer Atlas. Pick hex codes (#RRGGBB) that suit your needs.



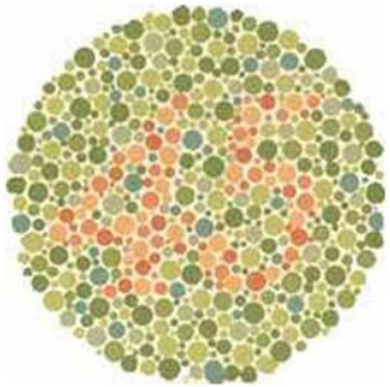
# Special topics: choosing colors

To avoid one color 'popping' out, the `colorspace` package has color schemes based on straight lines in a perceptually-based color space, rather than plain RGB – see right;



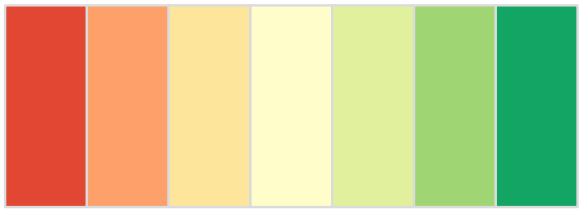



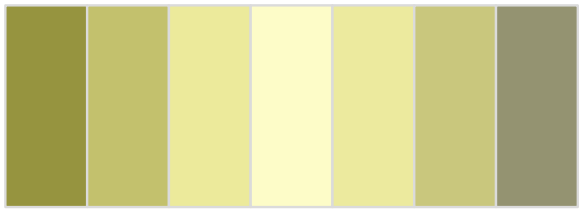
# Special topics: choosing colors

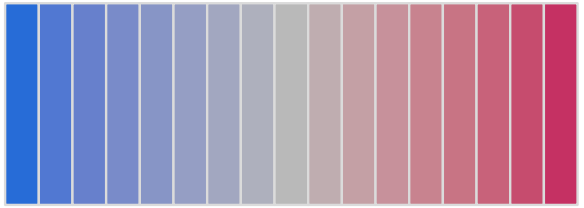


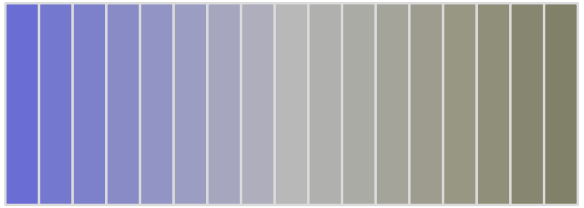
Color blindness affects $\approx 4\%$ of adults – mostly white males.
Red:green color blindness is the most common.

The `dichromat` package attempts to show the impact of red:green color blindness on your R color schemes.









# Special topics: better tables

Eliminate table junk as well as chart junk!

| $r$ | Exact 95% CI with $10/10^r$ successes |
|---|---|
| 2 | $(0.5, 2) \times 10^{-1}$ |
| 3 | $(0.5, 2) \times 10^{-2}$ |
| 4 | $(0.5, 2) \times 10^{-3}$ |

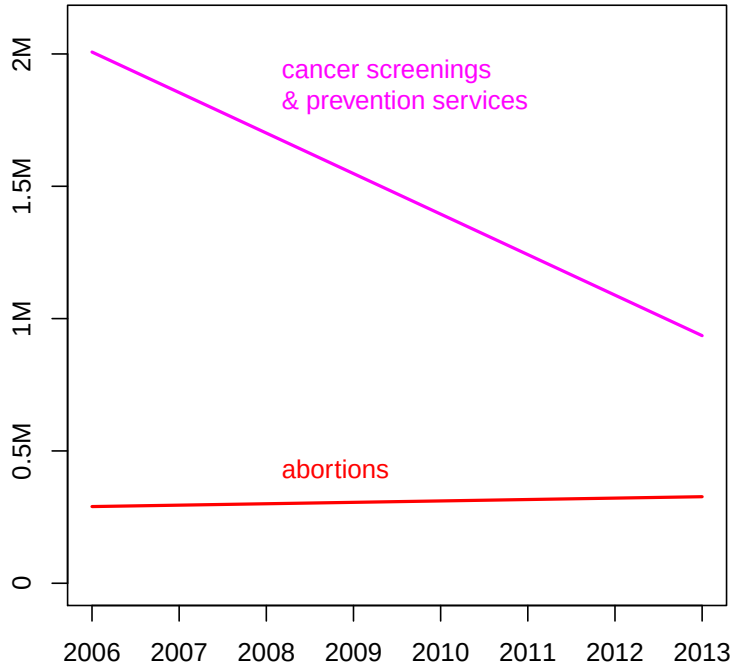
| $r$ | Exact 95% CI with $10/10^r$ successes |
|---|---|
| 2 | $(0.5, 2) \times 10^{-1}$ |
| 3 | $(0.5, 2) \times 10^{-2}$ |
| 4 | $(0.5, 2) \times 10^{-3}$ |

- Drop most of the lines – or tell `xtable()` to drop them.
- Use cell coloring *only* when you really need it
- For more see the APA style guide on tables

# Special topics: label lines directly

Most graphs will need a `legend()`, explaining what the symbols mean.

But as we've seen, no legend can be easier to read:



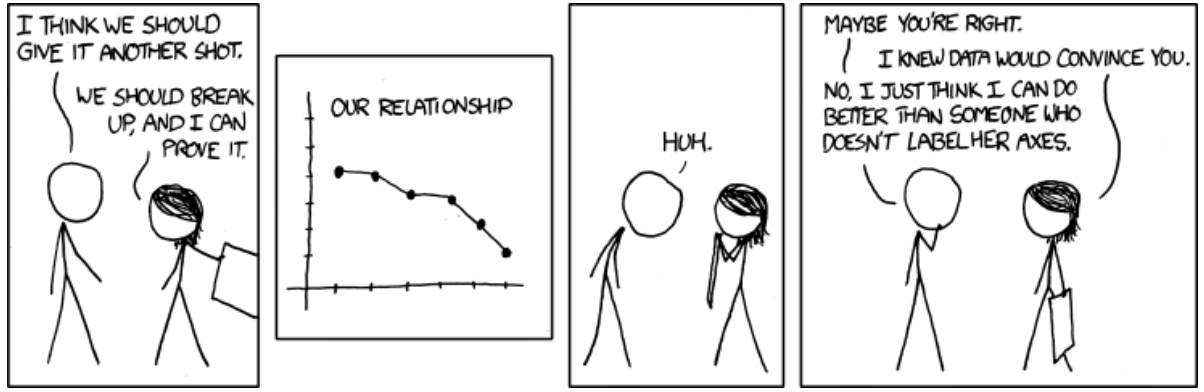

Taking this to extremes gives microtext lines (below) – no R package yet!

# Resources

Thank you for attending! In addition to the hyperlinks in the main slides:

- Excellent graduate-level course materials from Jerzy Wieczorek (was at CMU)
- Excellent undergrad-level course on information visualization – by Ross Ihaka, who started R
- A monograph by Rafe Donahue (Vanderbilt)
- Look around! Use other people's good ideas

Final obligatory XKCD cartoon;




*And if you labeled your axes, I could tell you exactly how MUCH better*