

Militancy Considerations

Non-Violent

Violent

1400 BC

3 BC

610

622

2010

Torah

Bible

Koran

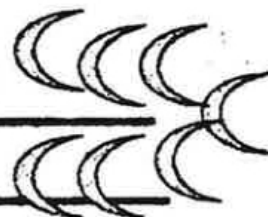
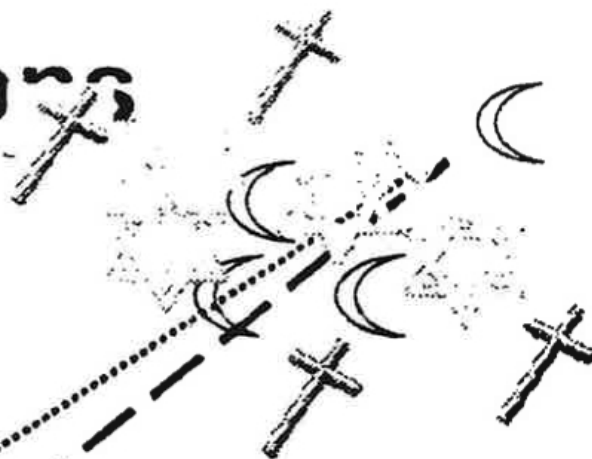
Medina Period

Meccan Period

Adherence by Pious and Devout

Adherence by Pious and Devout

Adherence by Pious and Devout



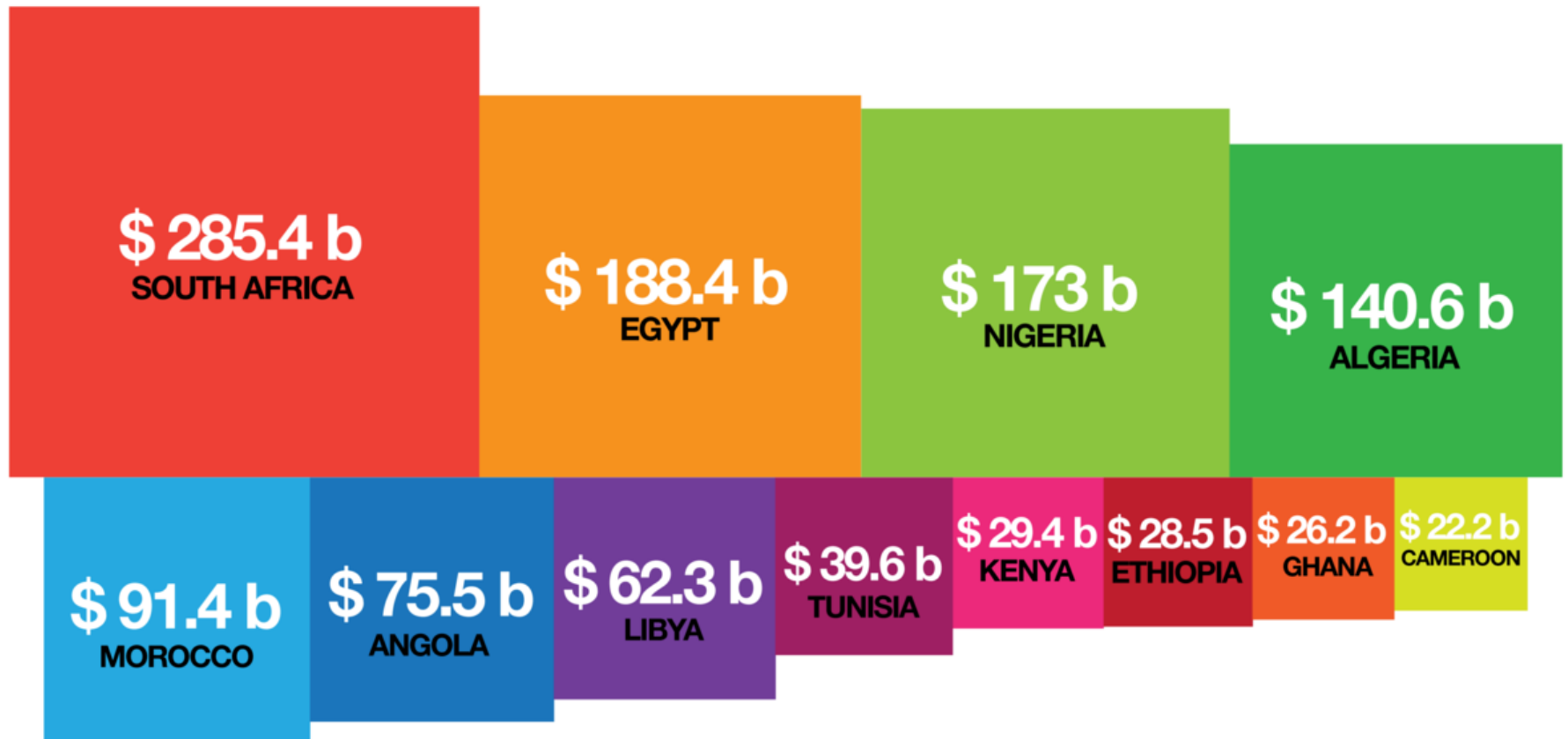
African Countries by GDP

TOP COUNTRIES BY GDP IN U.S. \$ BILLIONS

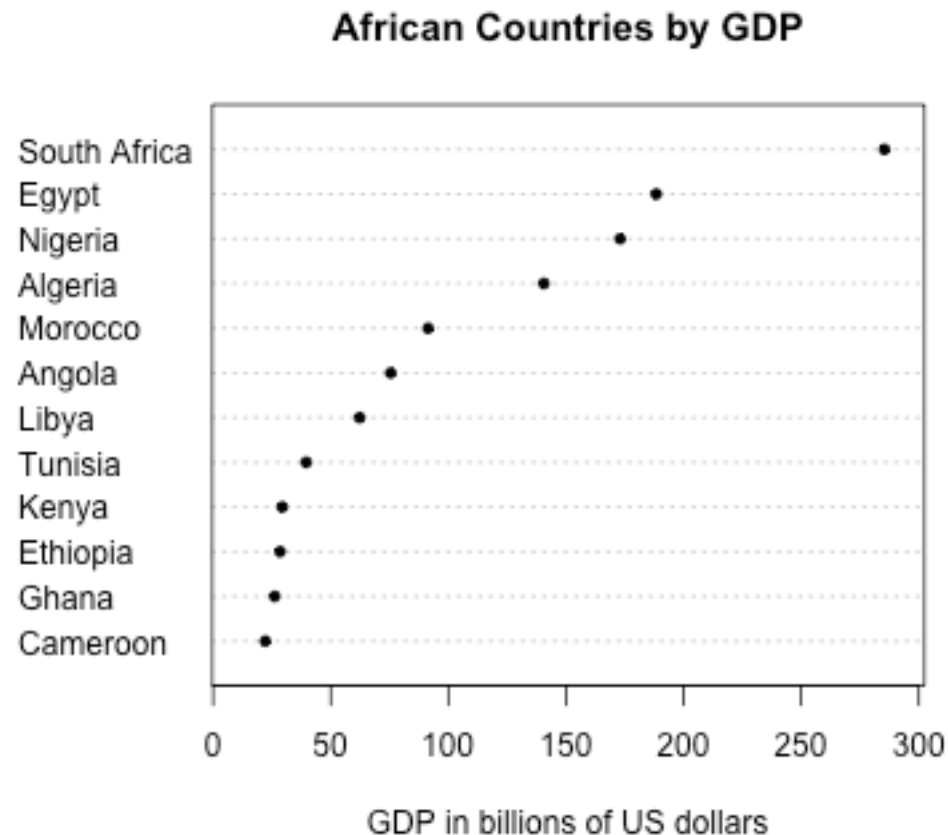
Gross domestic product (GDP) refers to the market value of all final goods and services produced within a country in a given period (2005 - 2009).

GDP CALCULATION

private consumption + gross investment + government spending + (exports - imports)



The informative (but boring) stat graphic



Part 1: The Puzzle

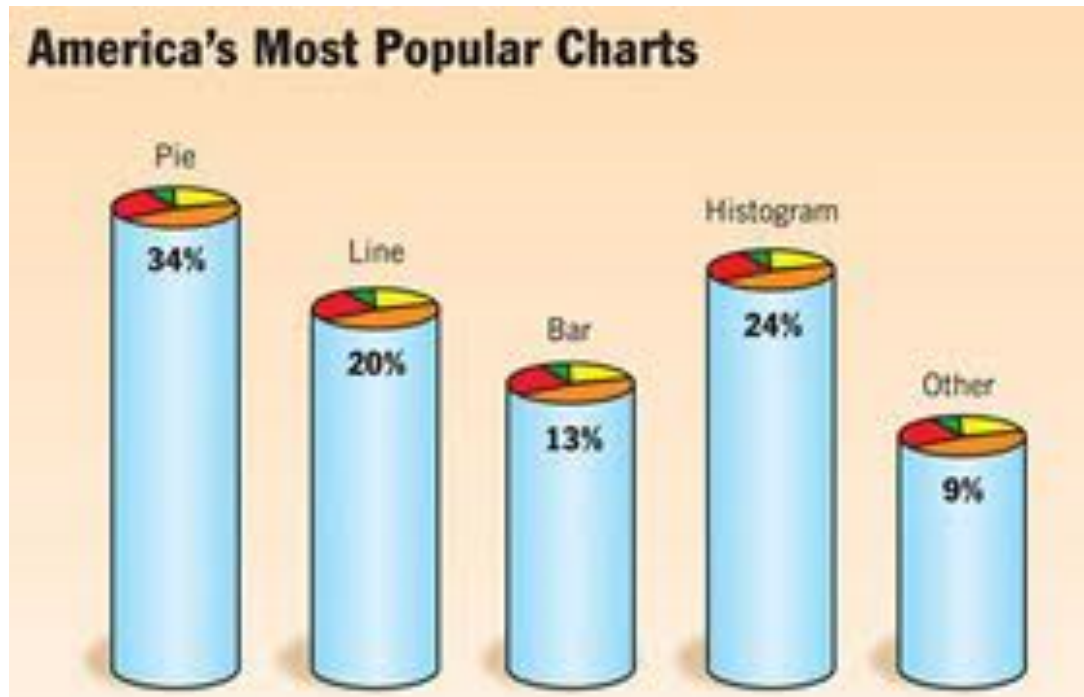
- The “5 best data visualization projects of the year”
- We didn’t like any of them
- And we’re the kind of statisticians who *like* graphics!

Information visualization has become popular . . .



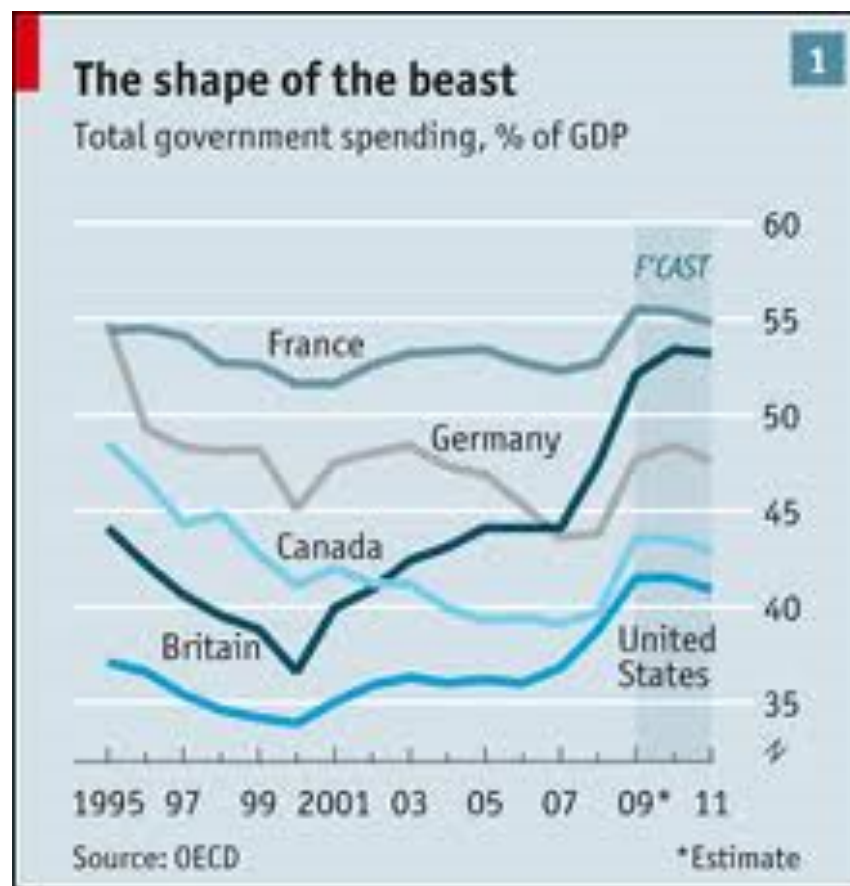
USA Today

... and parodied



The Onion

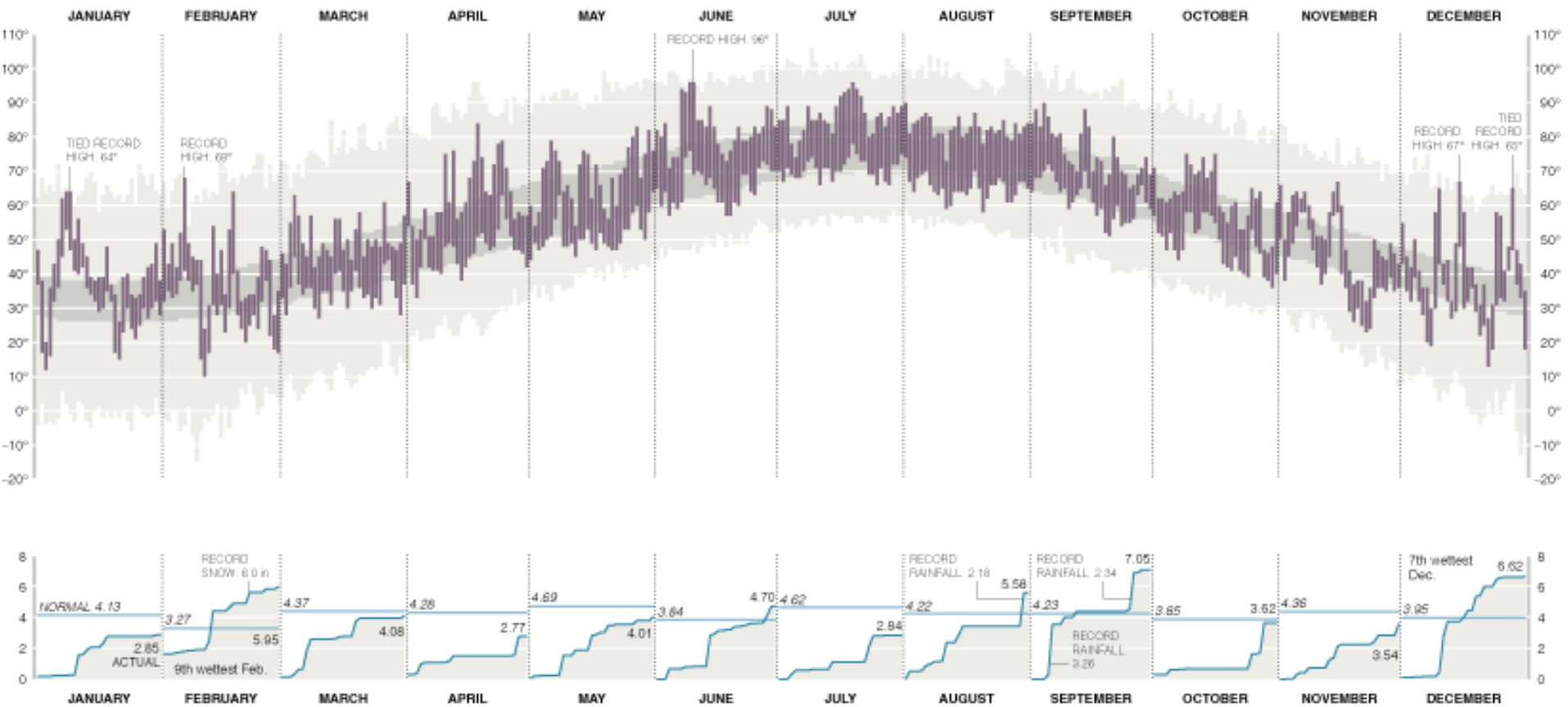
It can be done well . . .



The Economist

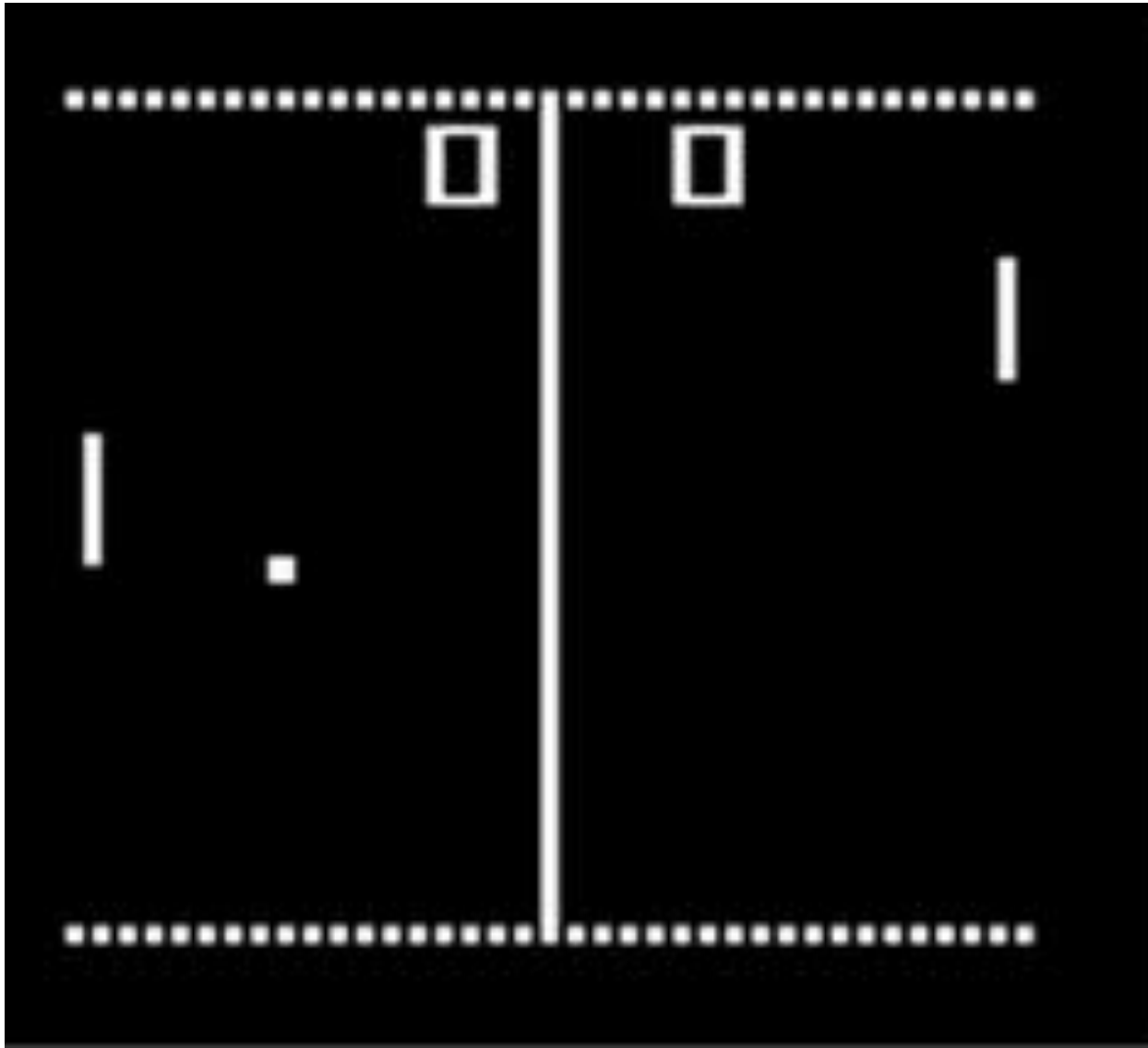
... but to what purpose?

2009 Weather in NYC

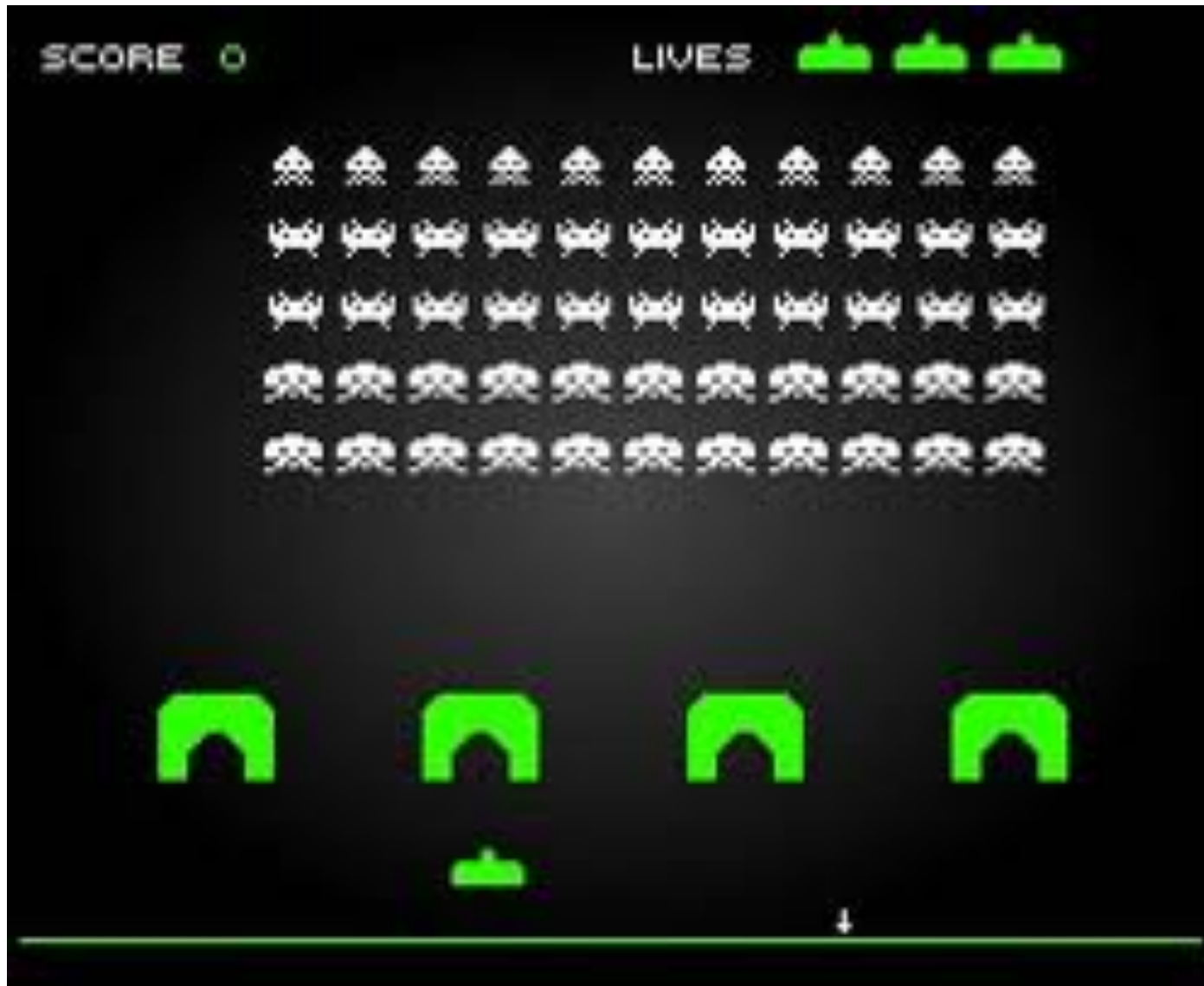


New York Times

Advances in computer graphics . . .



From Pong and Space Invaders . . .



... to Grand Theft Auto



Statistical perspectives on graphics

- Theories of graphics:
 - Bertin (1975), Wilkinson (2005)
 - Graphs as alternative to statistical models
 - Tukey (1977)
 - Graphs for comparisons
 - Cleveland (1985)
- Psychology research on “what works” . . .

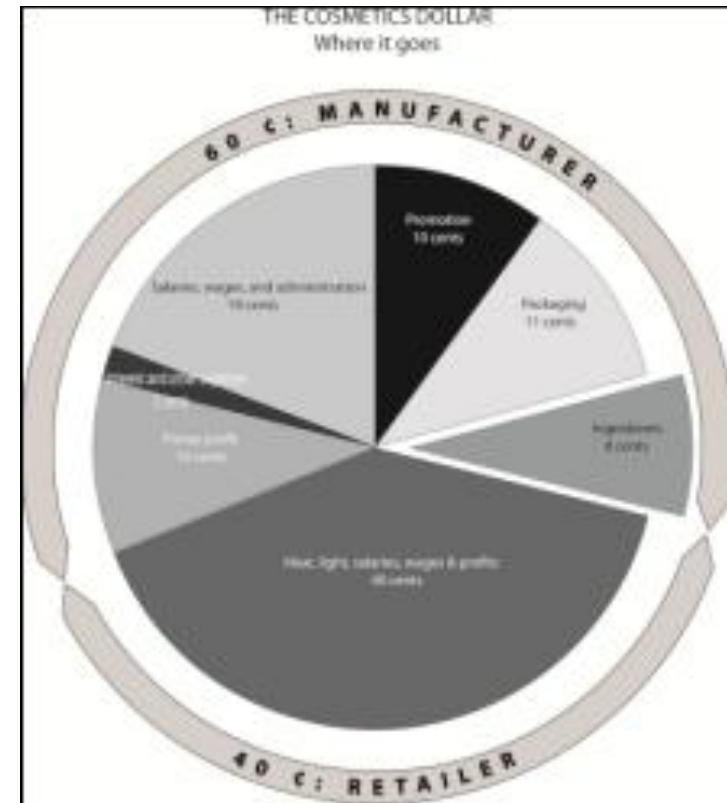
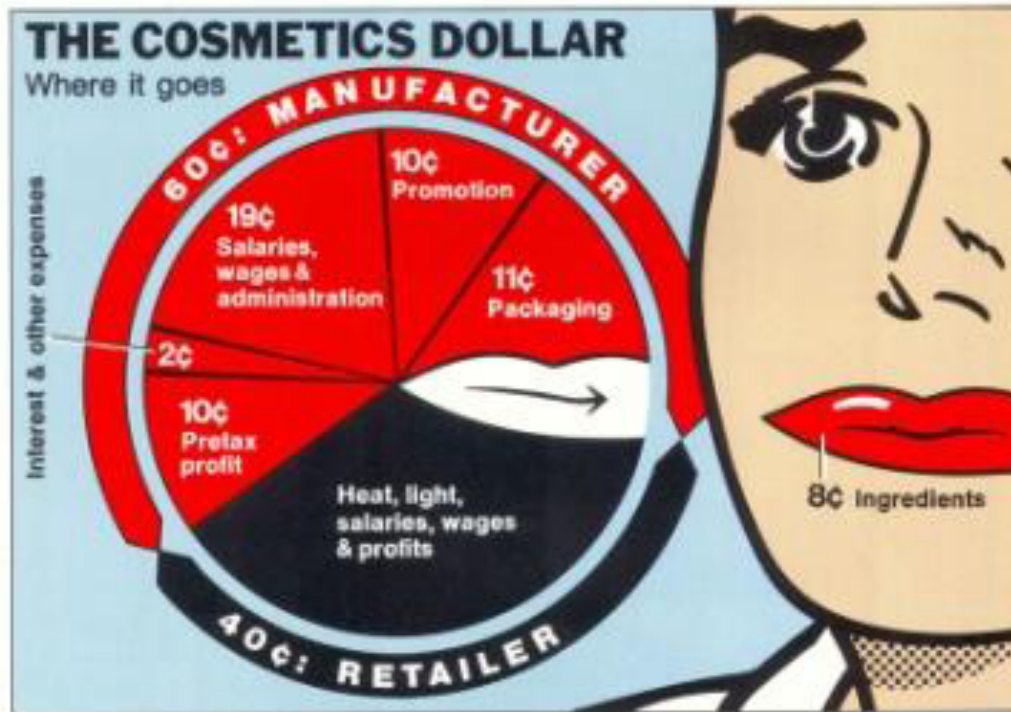
Problems with effectiveness research

Research: Why Chart Junk is More Useful than Plain Graphs

“Yep, **it has been scientifically proven**: the accuracy of people in describing charts with 'chart junk' is no worse than for plain charts, and the recall after a 2-3 week gap was actually significantly better. In addition, people overwhelmingly preferred 'chart junk' diagrams . . .”

- But, before you go and slashdot this . . .

The “chartjunk” study is . . . junk!



- OK. Good chartjunk is better than crap chartjunk

Part 2: Different tools, different goals

- Can we uncover the differences between the values and priorities of infovis and statistical graphics?

5 Best Data Visualization Projects of the Year

December 19, 2008 to Featured, Visualization | [Post on Twitter](#)



5 Best Data Visualization Projects of the Year

Honorable mention: Wordle



Jonathan Feinberg, wordle.com

- Nathan Yau: “It's hard to say what exactly made Wordle so popular, but I [Yau] think it was a mix of randomness, aesthetics, and customization options”
- **Our view:** Visualization as a fun puzzle



#5. The Obama-Clinton divide

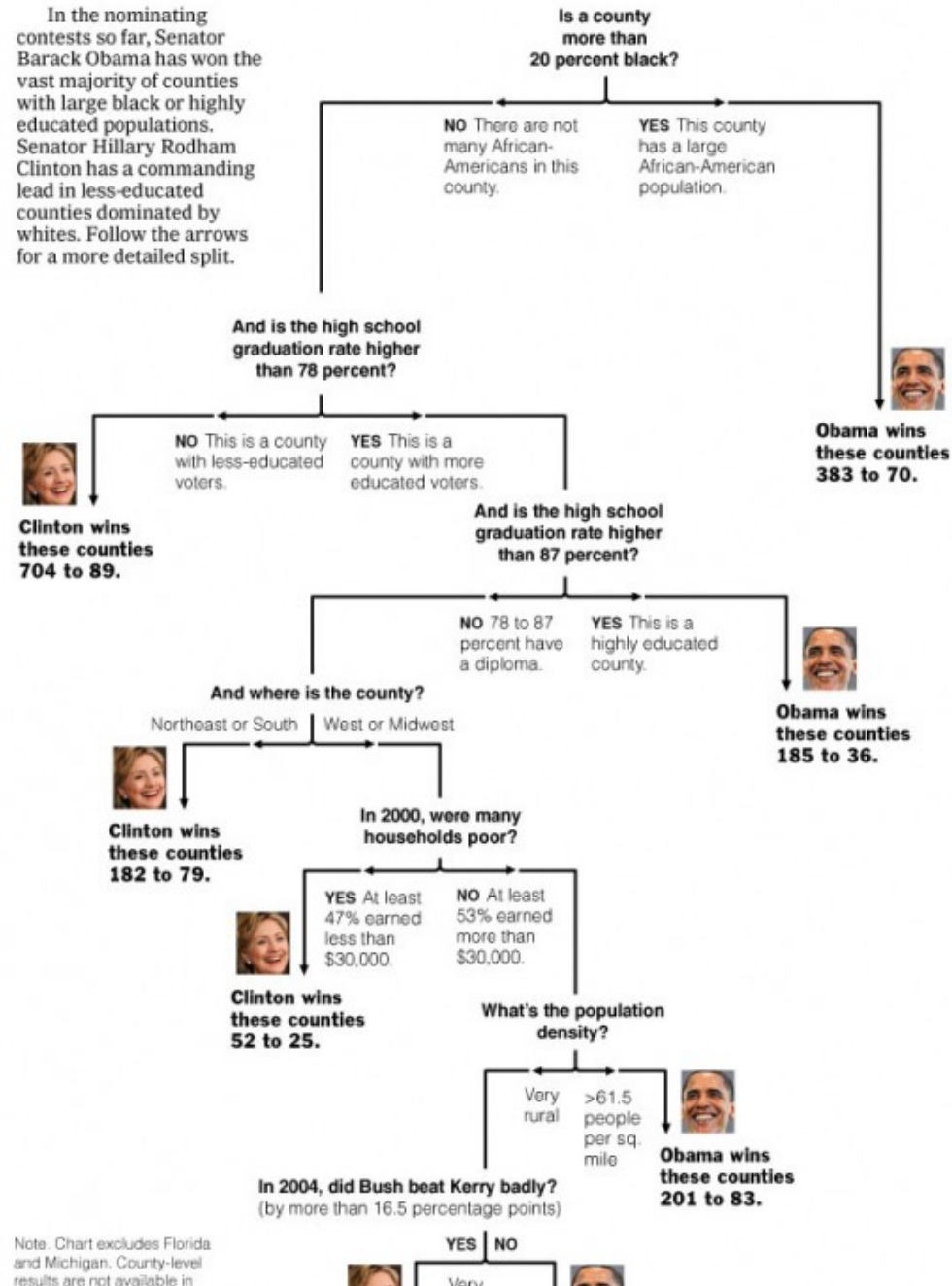
(Amanda Cox,
New York Times)

•Our view:

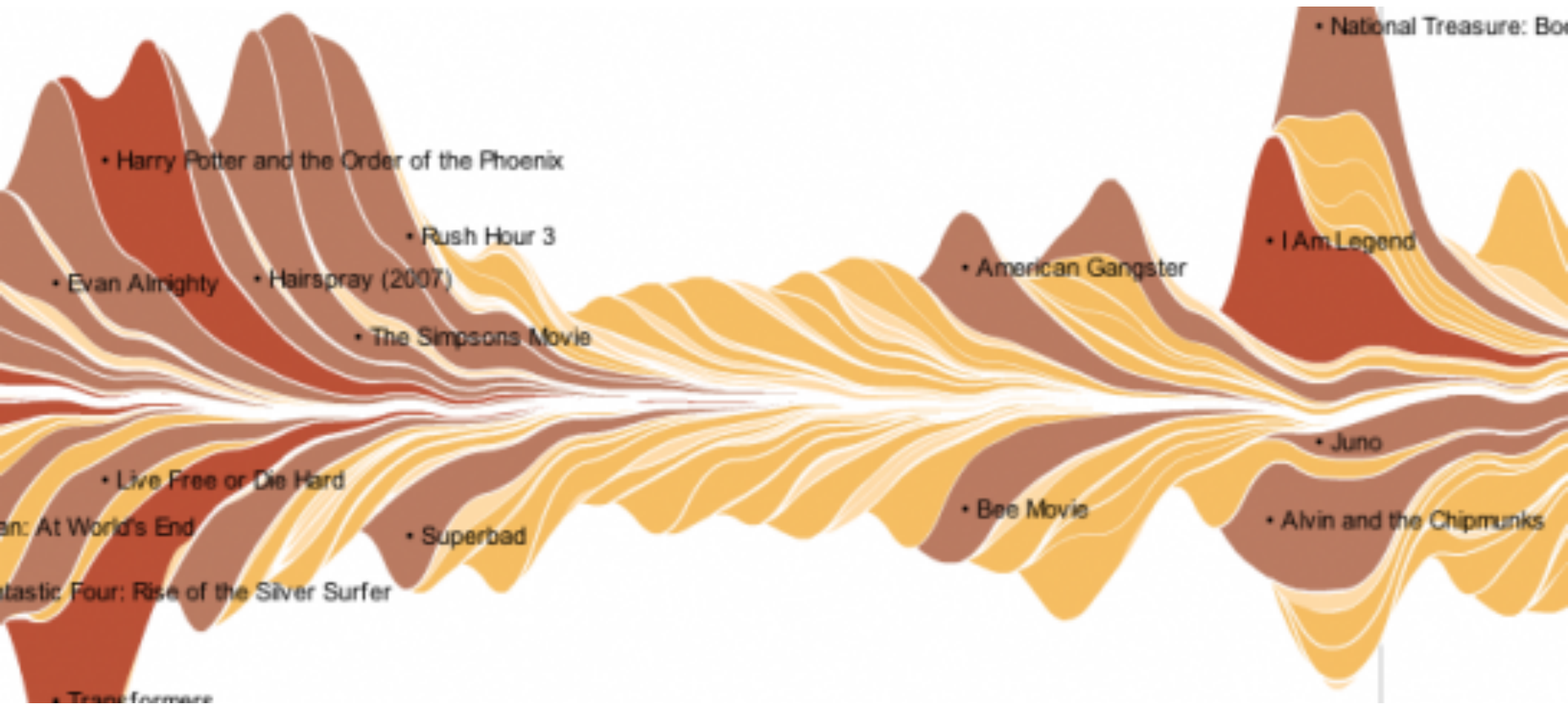
- A pretty presentation of a bad model
- Misleading picture of classification as “decision”

Decision Tree: The Obama-Clinton Divide

In the nominating contests so far, Senator Barack Obama has won the vast majority of counties with large black or highly educated populations. Senator Hillary Rodham Clinton has a commanding lead in less-educated counties dominated by whites. Follow the arrows for a more detailed split.



#3. Box office streamgraphs



Lee Byron

- Yau: “You can see Oscar contenders attract a smaller audience than the holiday and summer blockbusters and kind of slowly build an audience.”

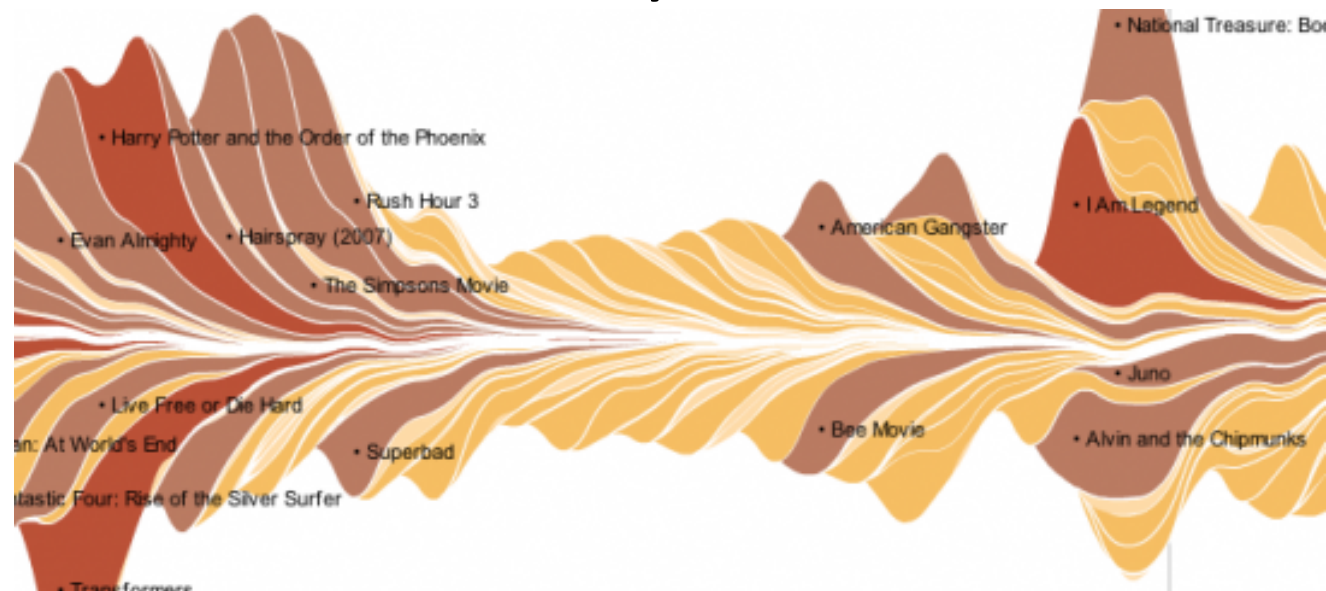
- Our view:

- Huh?

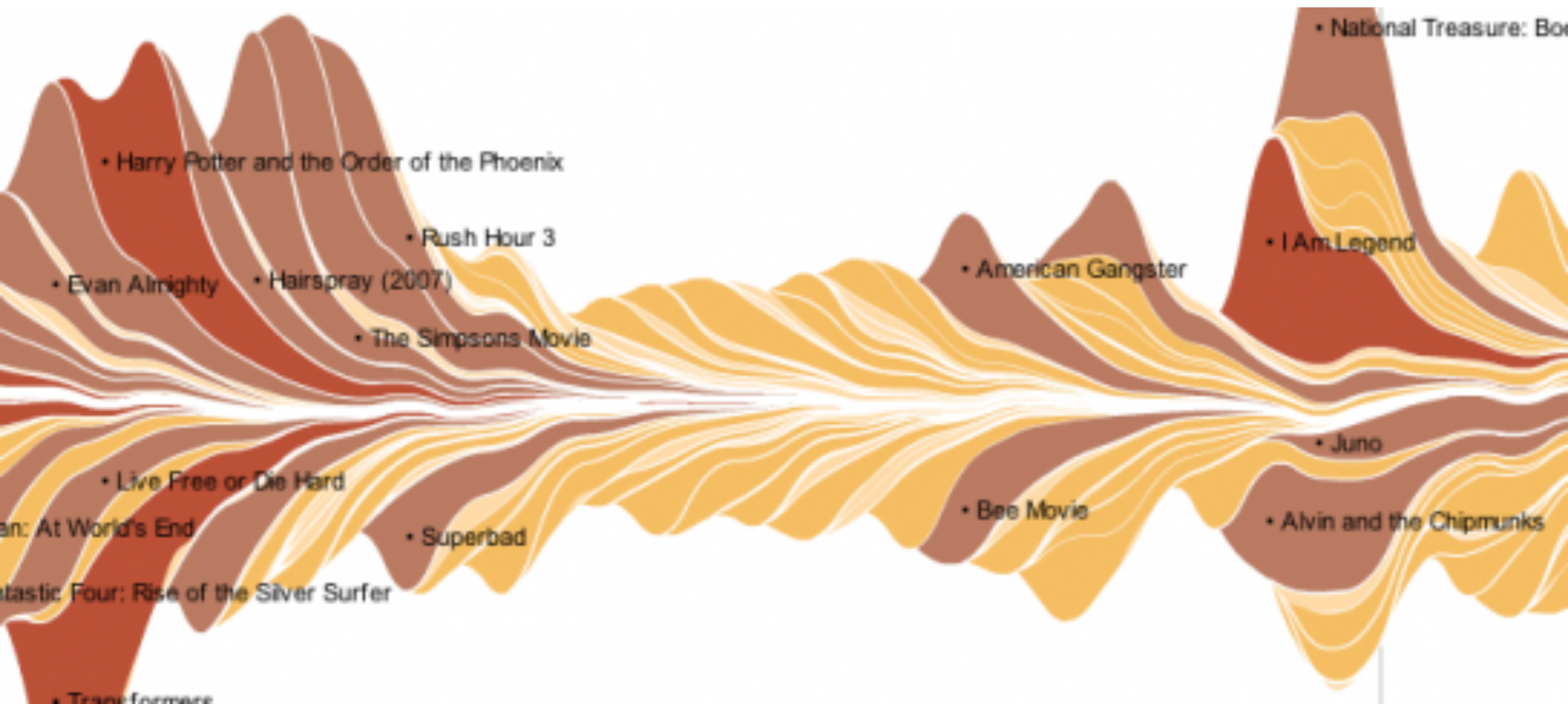
- Better to display as two plots:

(1) total sales over time, (2) trajectories for individual movies

- Again, graph as puzzle



- Yau: “Discussion burst out across the Web . . . that I am convinced would not have come about if instead of a Streamgraph, they used say, a **stacked bar chart**.” [emphasis added]



“5 Best Data Visualizations”: our view

- Eye-catching graphics
- State-of-the art methods in stat and comp sci
- No attempt to achieve the traditional goals of statistical graphics (communication, discovery)

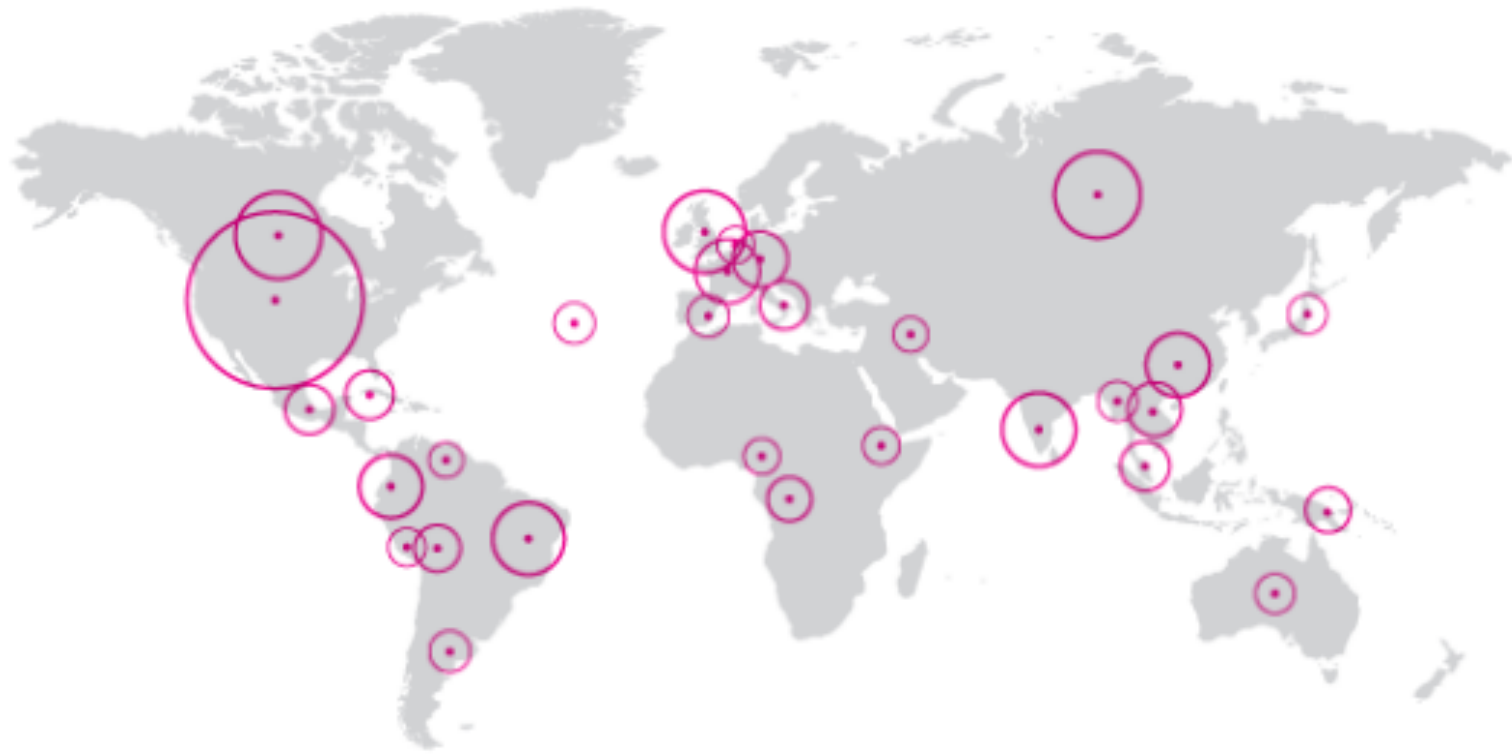
Our resolution of the puzzle

- Statisticians:
 - Graphics as alternative to data tables and numerical data reductions (p-values, etc.)
- Graphic designers:
 - Information visualization as alternative to photos, cartoons, perfume, etc.
- How to get best of all worlds?
 - Shneiderman (1996):
Overview, zoom and filter, details on demand

Winner of *Guardian* newspaper's Visualization Contest

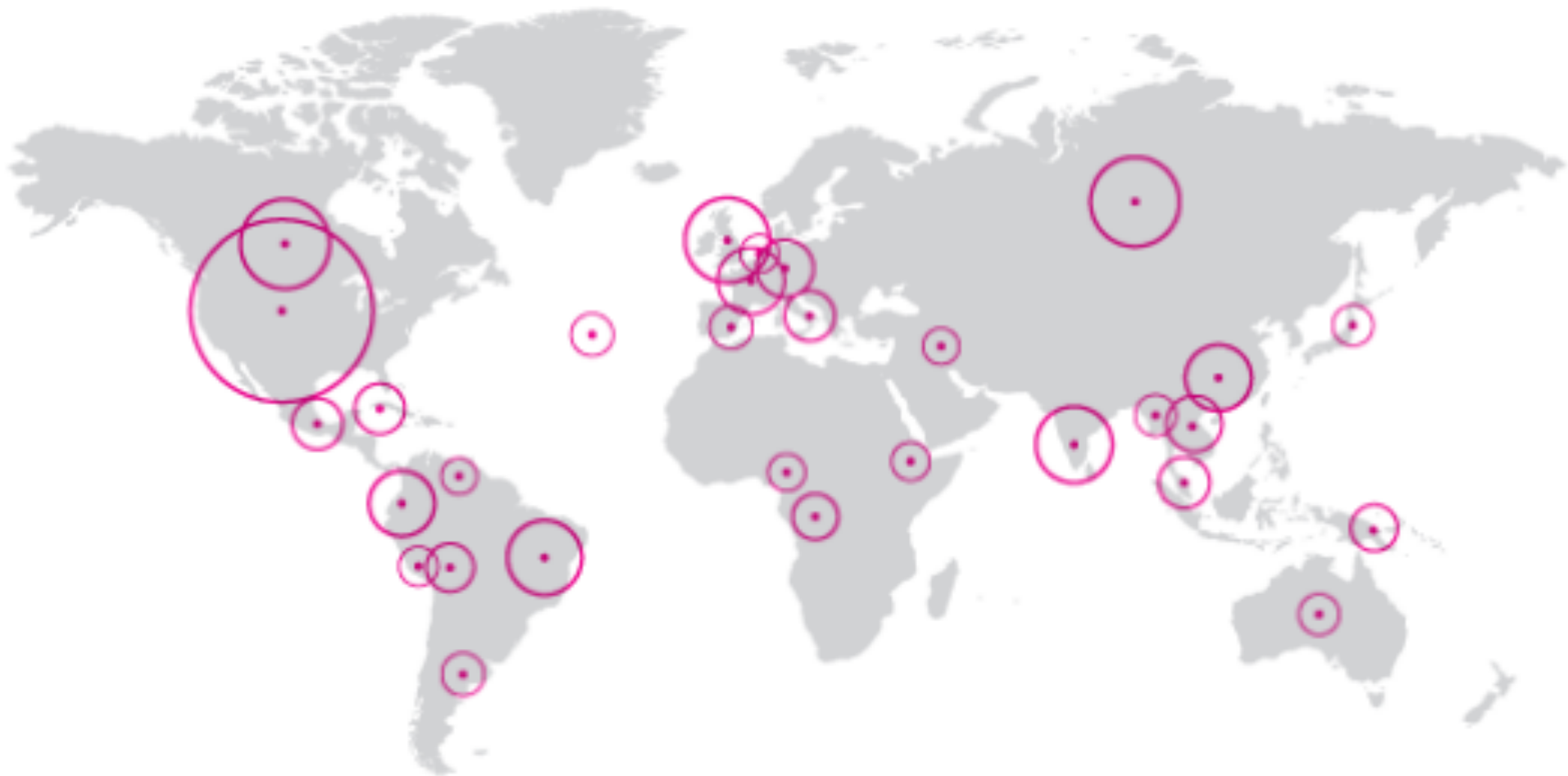
Final Destination

Density of fatal accidents 1942-2009



David McCandless

- Our view:
 - Display looks clean and efficient but isn't!
 - Analogy to some modern architecture



Florence Nightingale's coxcomb

<http://www.Florence-Nightingale-Avengeing-Angel.co.uk/Coxcomb.htm>

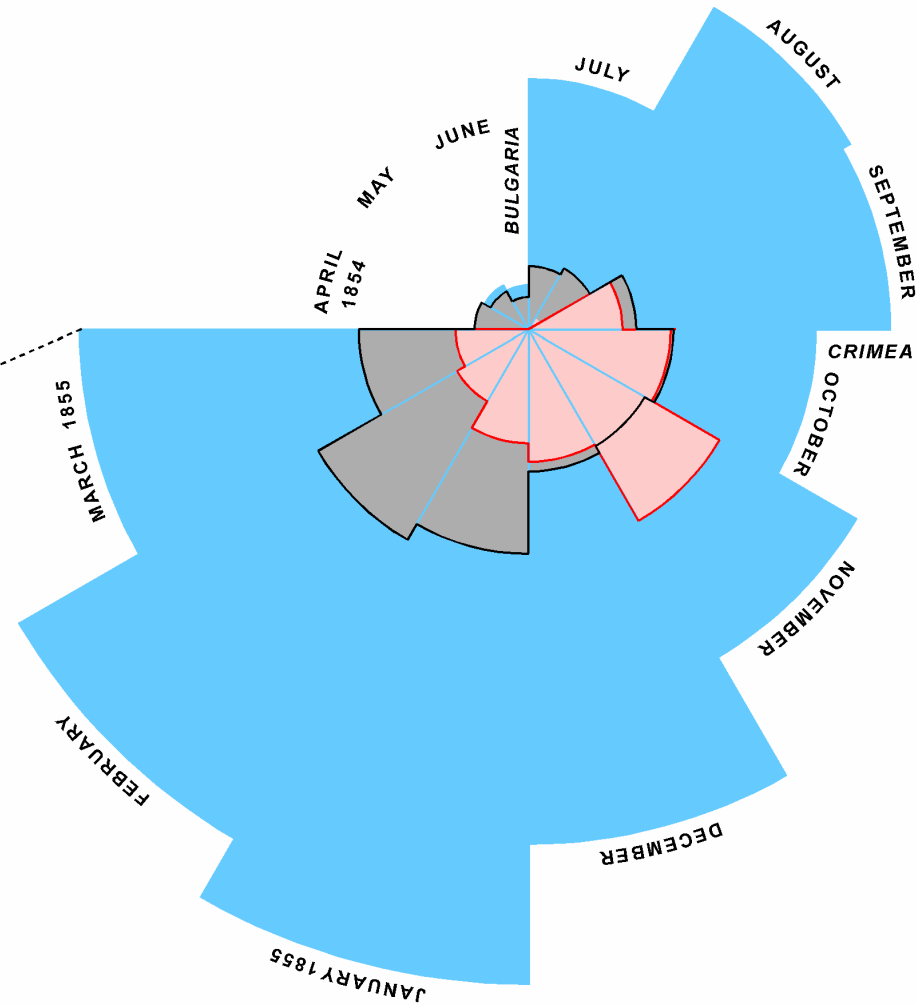
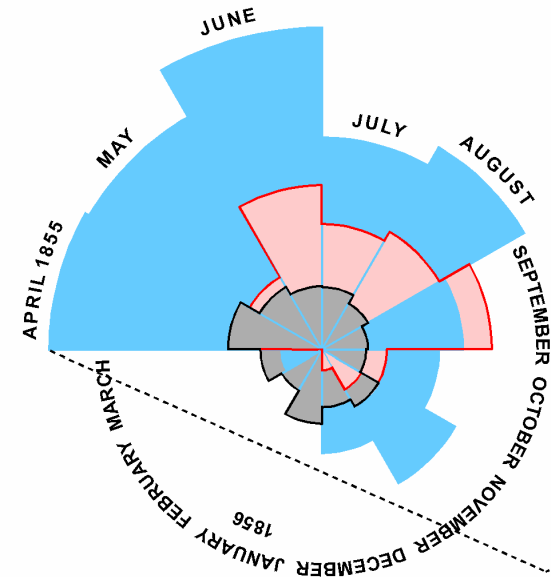
DIAGRAM OF THE CAUSES OF MORTALITY IN THE ARMY IN THE EAST.

2.

APRIL 1855 TO MARCH 1856.

1.

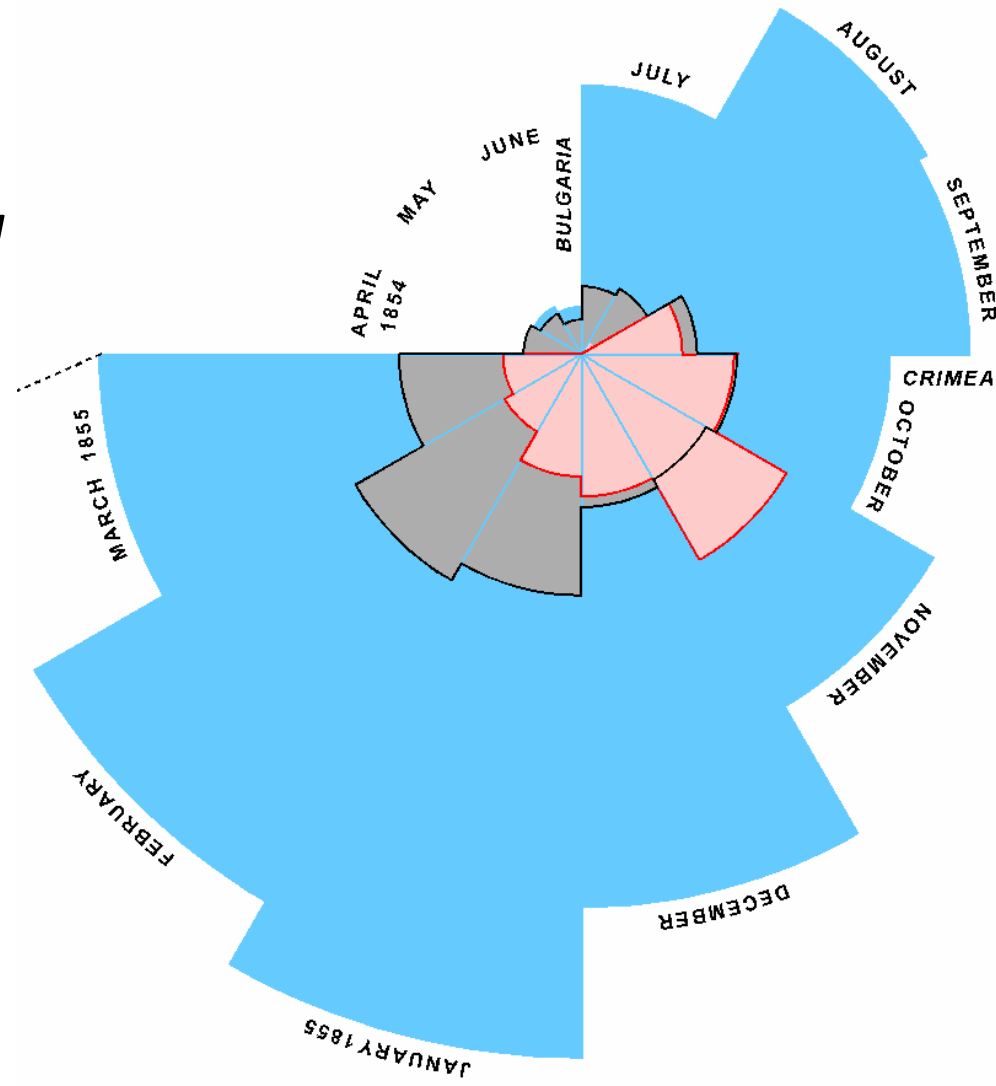
APRIL 1854 TO MARCH 1855.



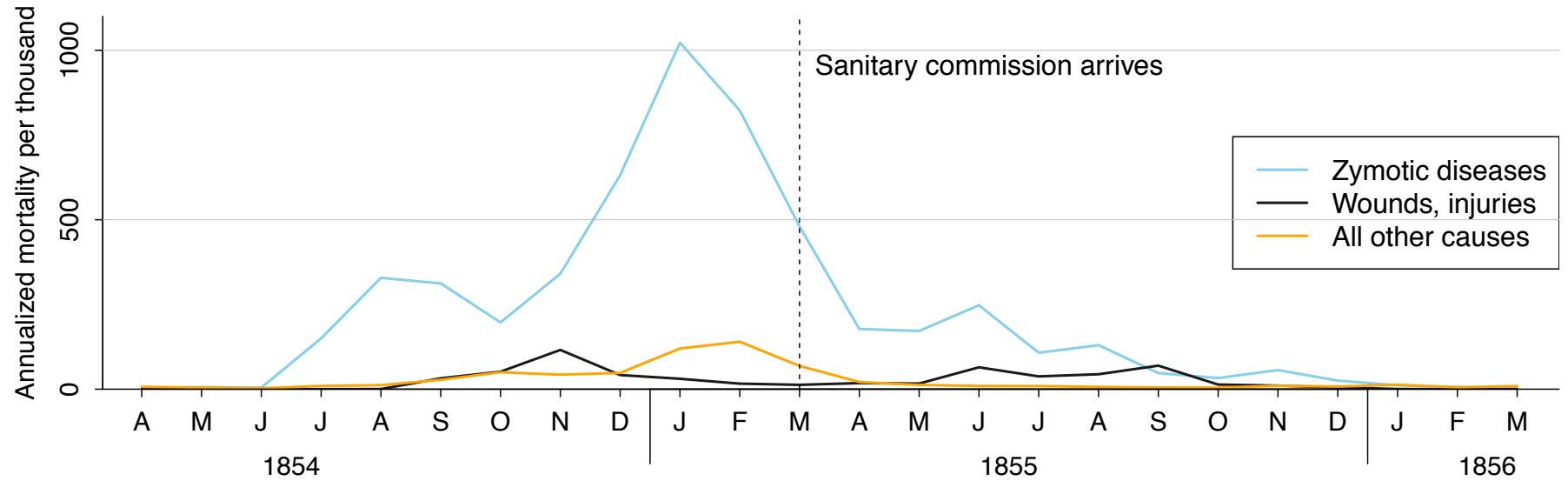
The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex
The blue wedges measured from the centre of the circle represent area for area the deaths from Preventable or Mitigable Zymotic Diseases, the red wedges measured from the centre the deaths from wounds, & the black wedges measured from the centre the deaths from all other causes
The black line across the red triangle in Nov' 1854 marks the boundary of the deaths from all other causes during the month
In October 1854, & April 1855, the black area coincides with the red, in January & February 1856, the blue coincides with the black
The entire areas may be compared by following the blue, the red & the black lines enclosing them

- Our view:

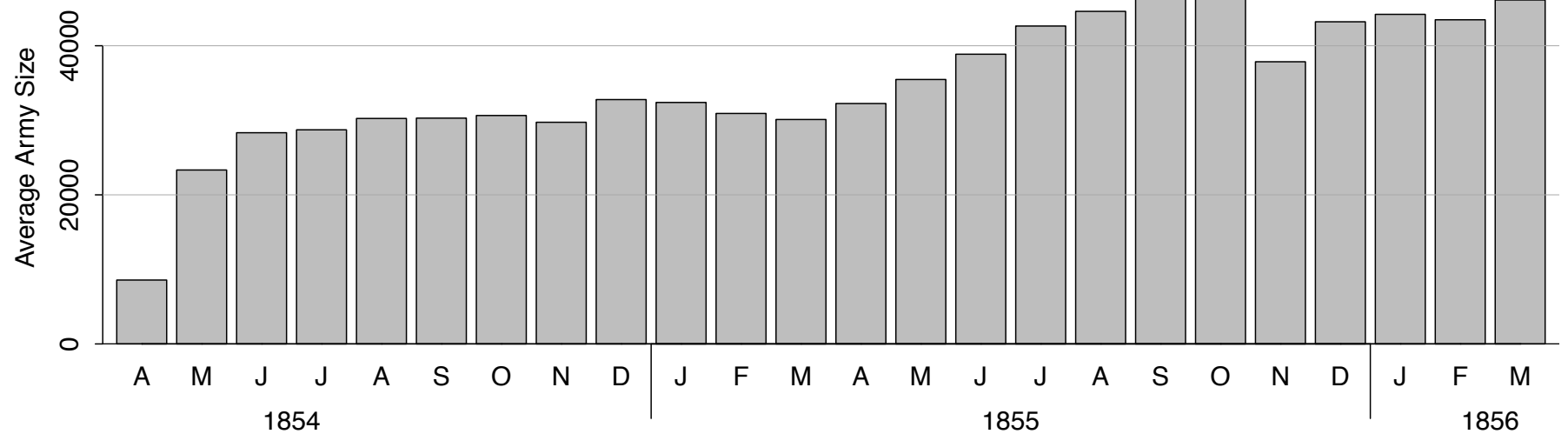
- Excellent *infographic*—it's attractive, grabby, thought-provoking
- Graph as puzzle
- Not a good *statistical graphic* does not push to deeper understanding
- “Clock plot” as dead end



Mortality rates in the Crimean War from April 1854 to March 1856



British Army Size in the Crimean War from April 1854 to March 1856

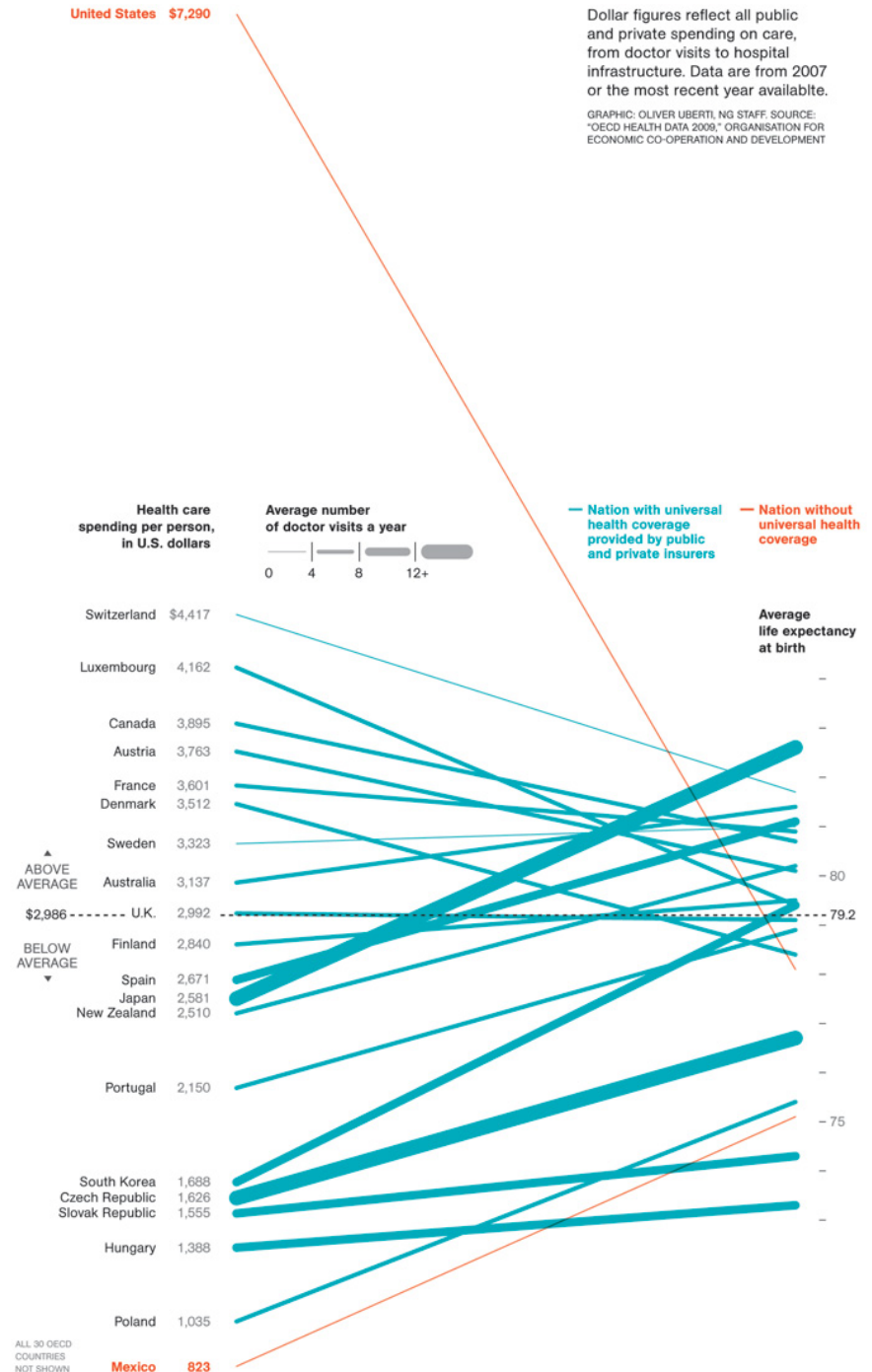


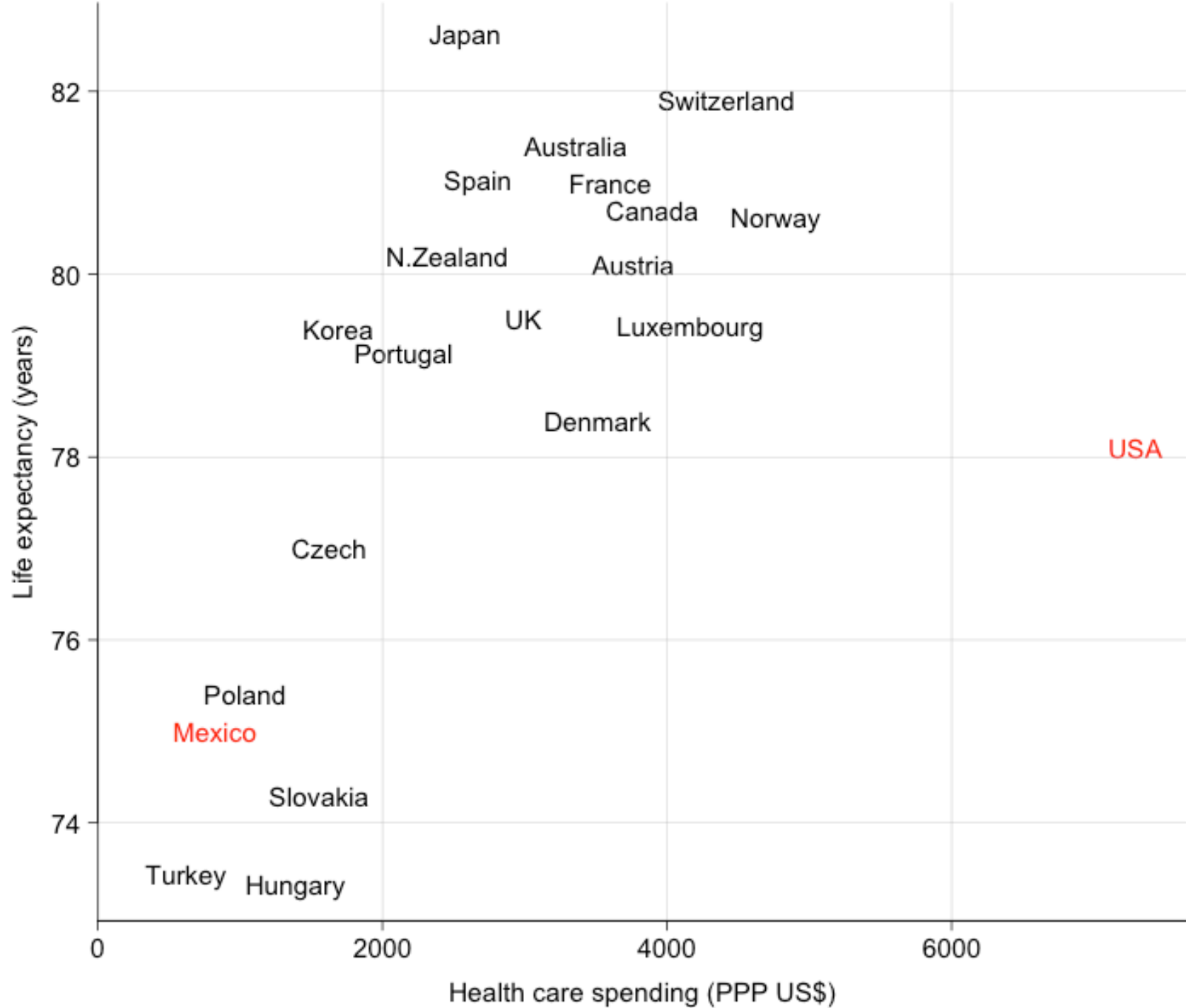
Health spending

(Oliver Uberti,
National Geographic)

- Our view:

- Image is dramatic but doesn't convey data well
- Arbitrary scaling
- Try a scatterplot





Part 3: Software as friend or enemy

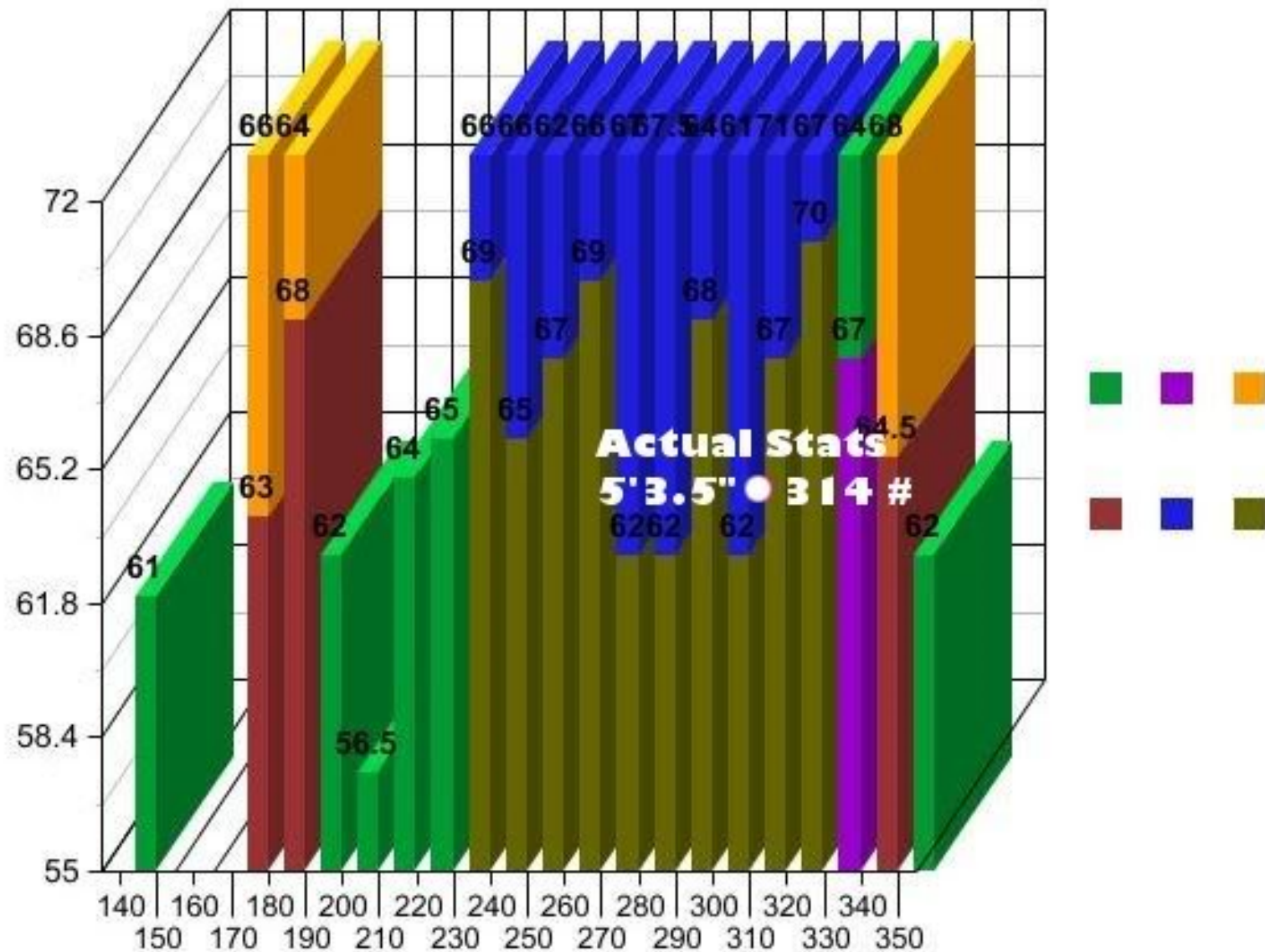
- Beautiful possibilities . . .
- But defaults can kill
 - Bar graphs and pie charts as standards
 - And it can get much worse . . .

Guess her height and weight

From blog of Kate Harding



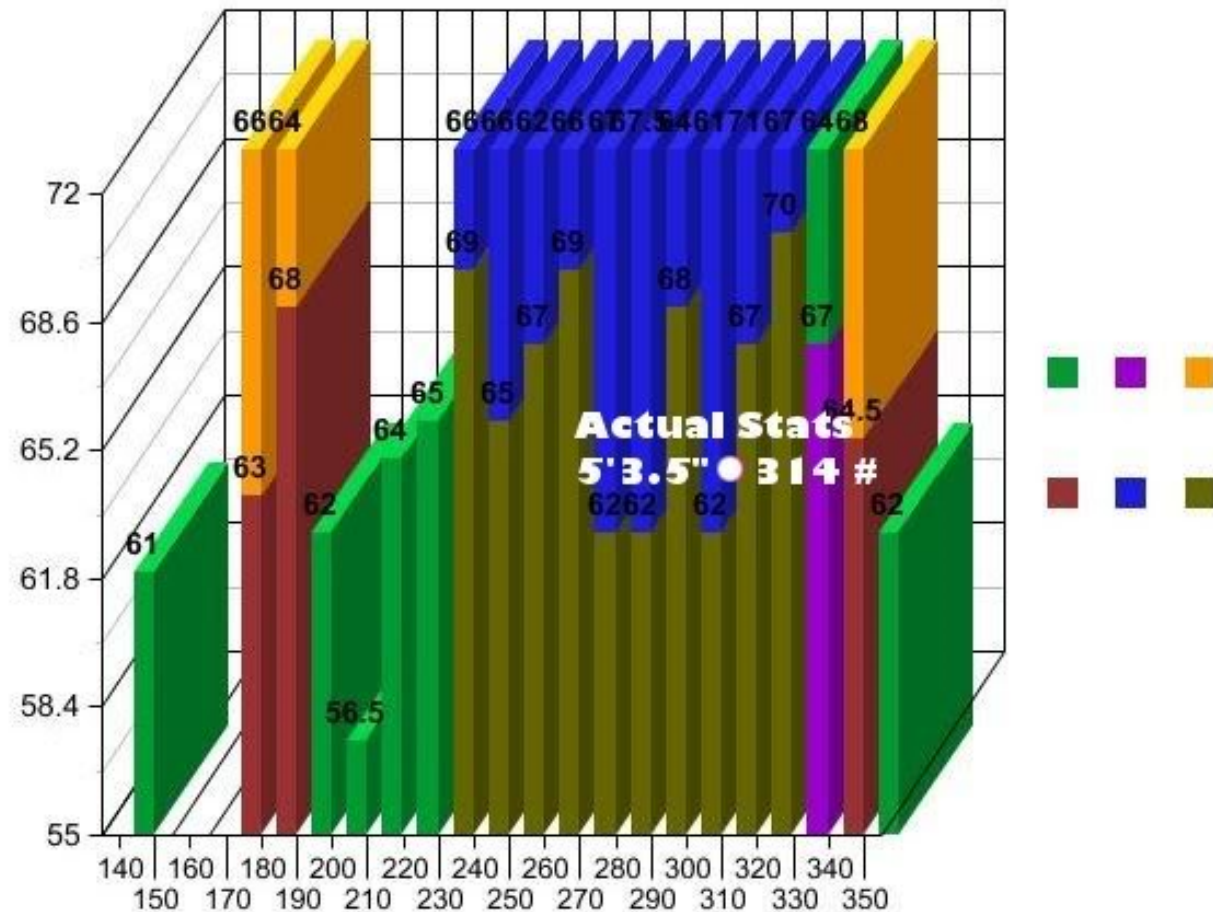
Distribution of the guesses



- Kate Harding: “To be honest, I’m not even sure how you're supposed to read this graph, and I'm the one who made it.”

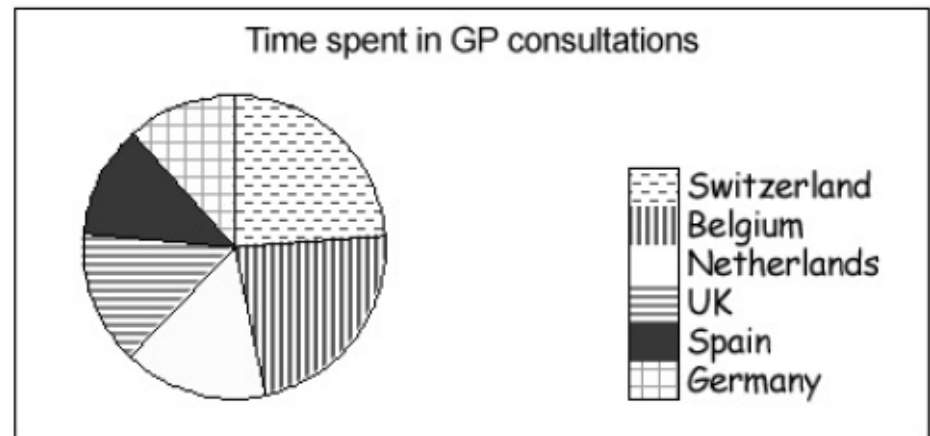
- **Our view:**

- We blame the software!



- Our point is not to mock
- We must communicate with *authors of software* as well as *creators of graphs*
- To paraphrase Keynes:

Practical researchers, who believe themselves quite exempt from any graphical influence, are usually the slaves of some defunct software design



- 4) Which two countries give their patients the most time?
- 5) Which two countries give their patients the least time?
- 6) What colour is the UK slice?
- 7) Which country gives their patients about the same amount of time as the UK?

) Now check your answers with those on the answer sheet.

Part 4: A positive example

Baby Name > ✕

● Both ○ Boys ○ Girls

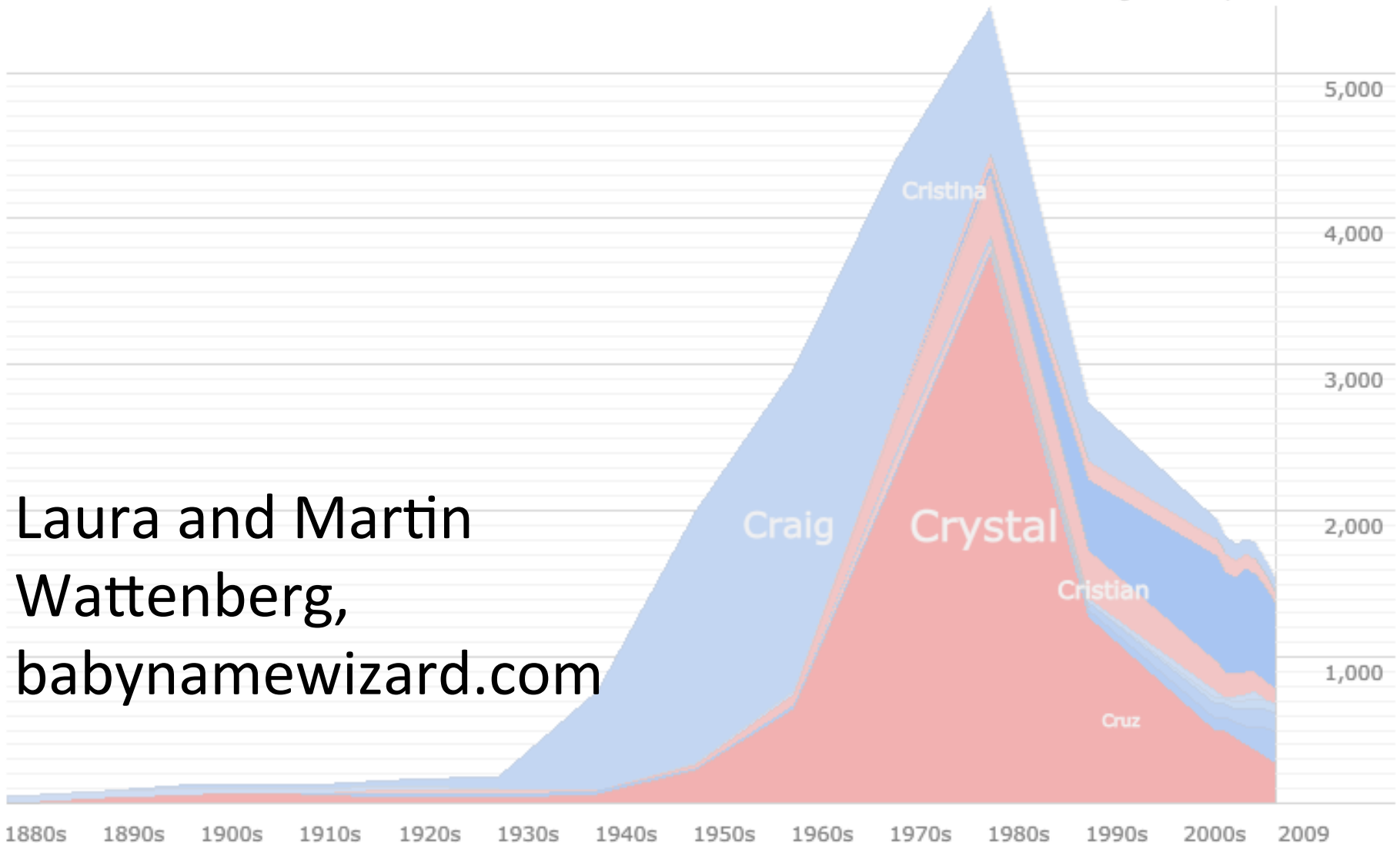
2009 rank: boys

1000	500	100	25	1
------	-----	-----	----	---

girls

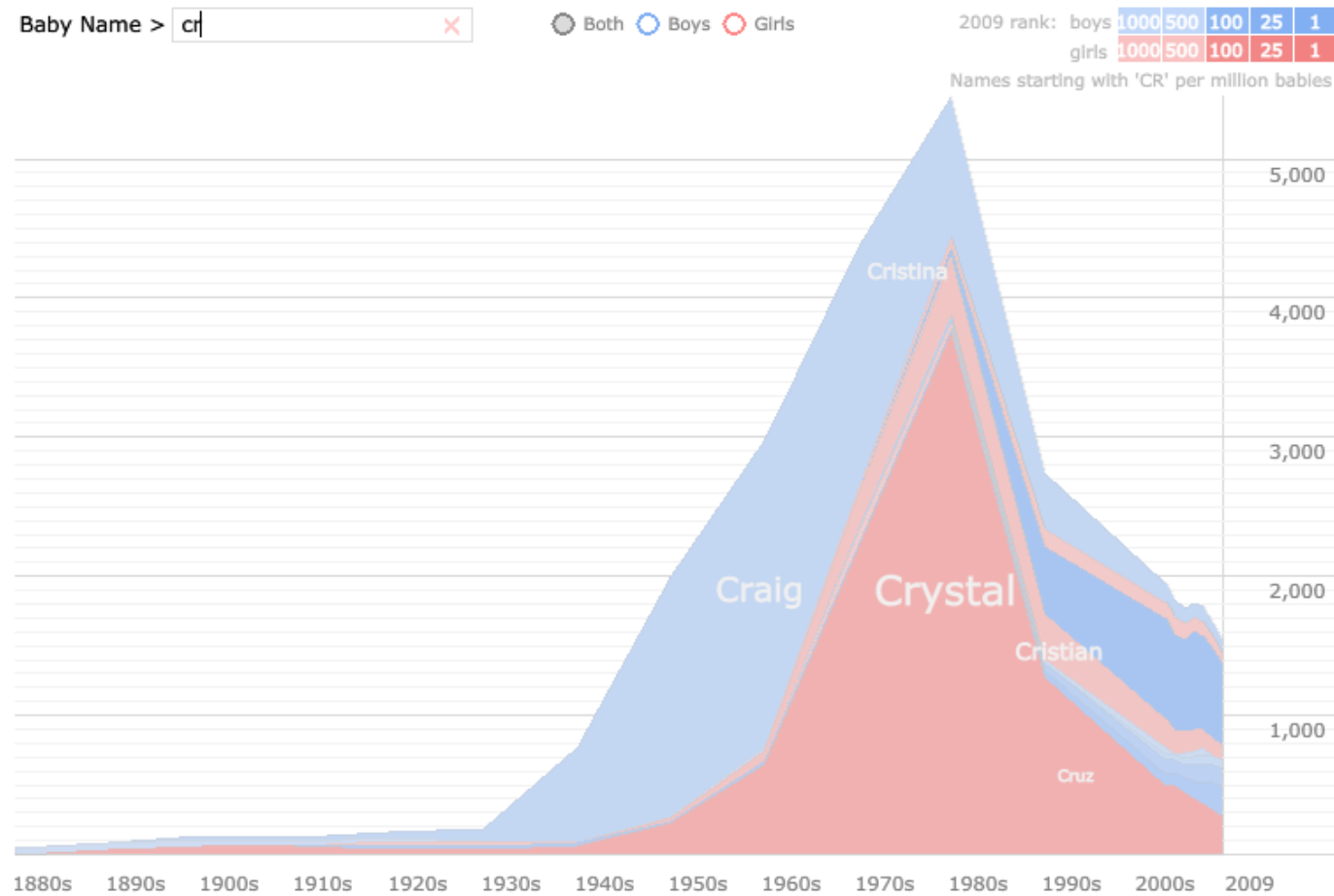
1000	500	100	25	1
------	-----	-----	----	---

Names starting with 'CR' per million babies

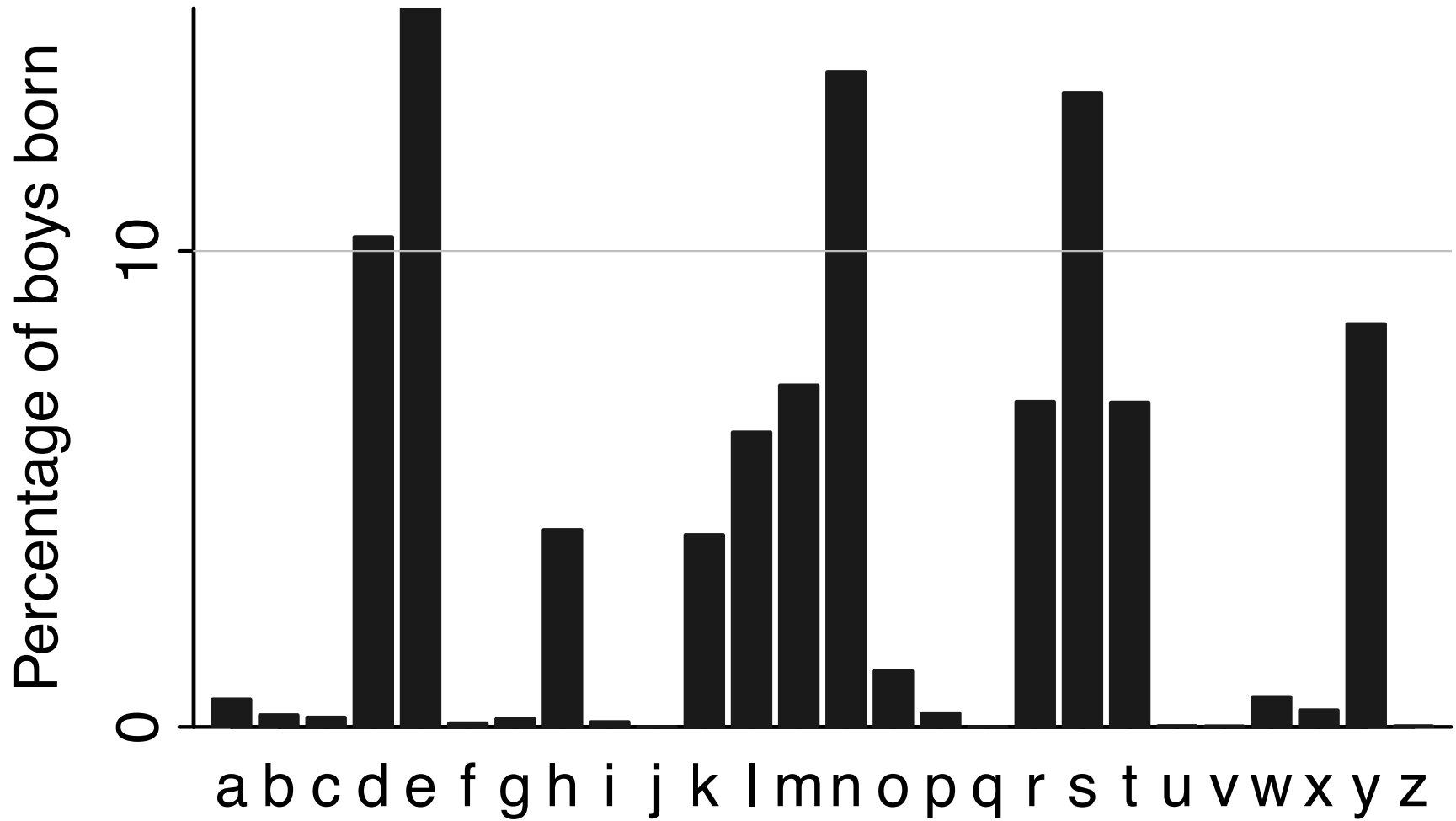


Laura and Martin
Wattenberg,
babynamewizard.com

- Excellent as infovis *and* as statistical graphics
 - Colors are informative rather than decorative
 - Axes go to zero and are labeled clearly but gently
 - Melding of static and interactive

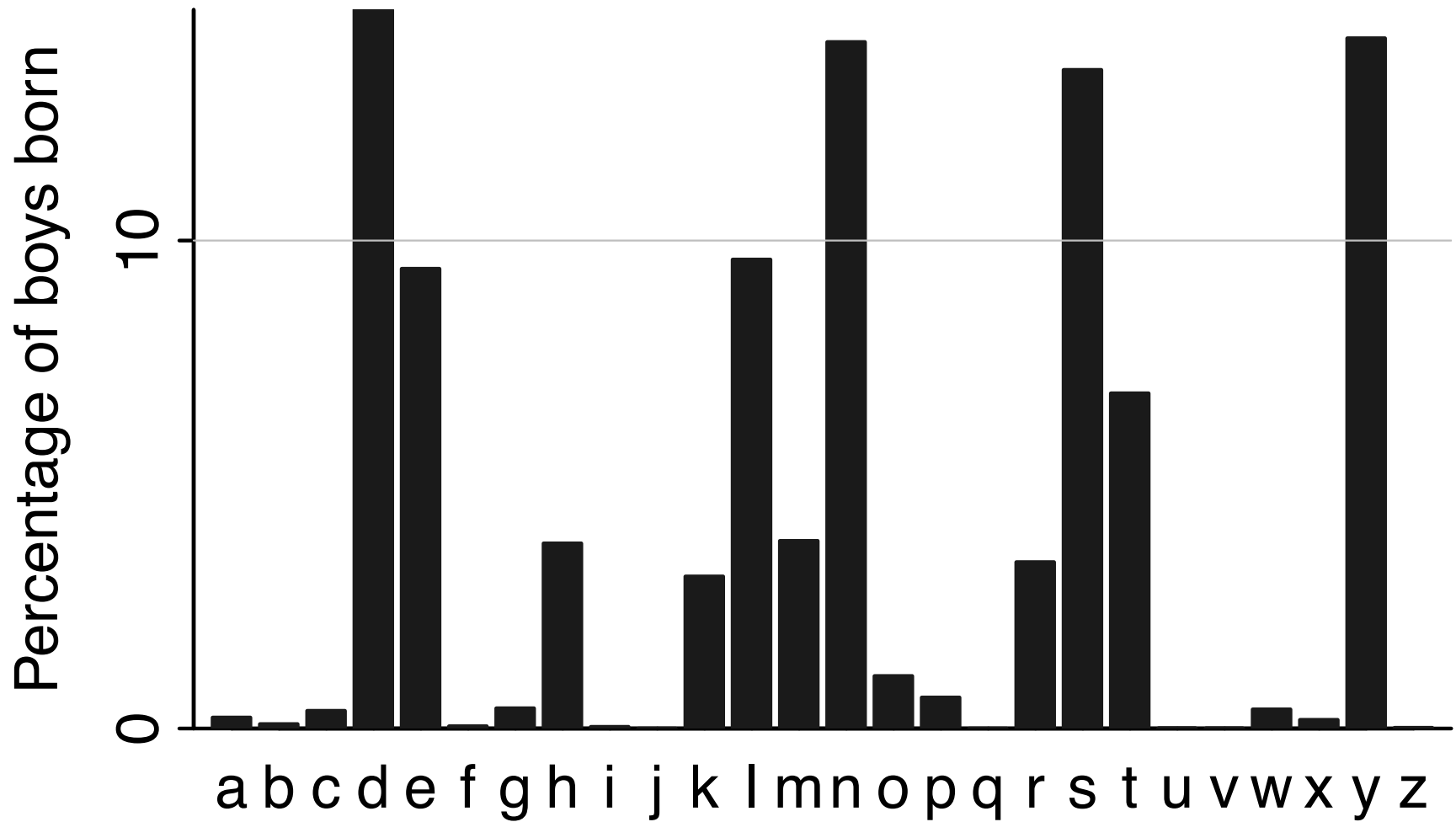


Last letter of boys' names in 1900



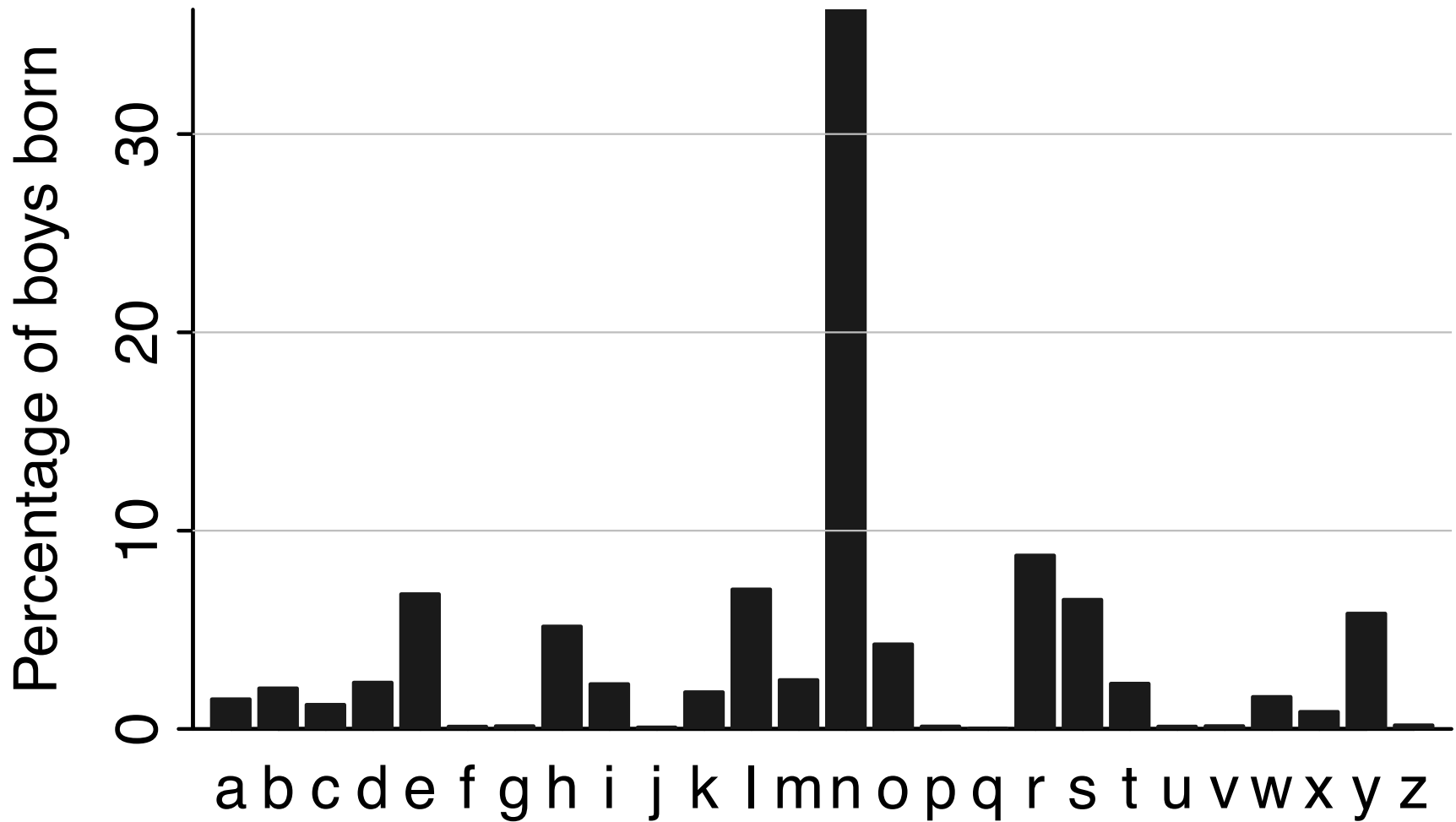
John, James, George, Charles, Edward, ...

Last letter of boys' names in 1950



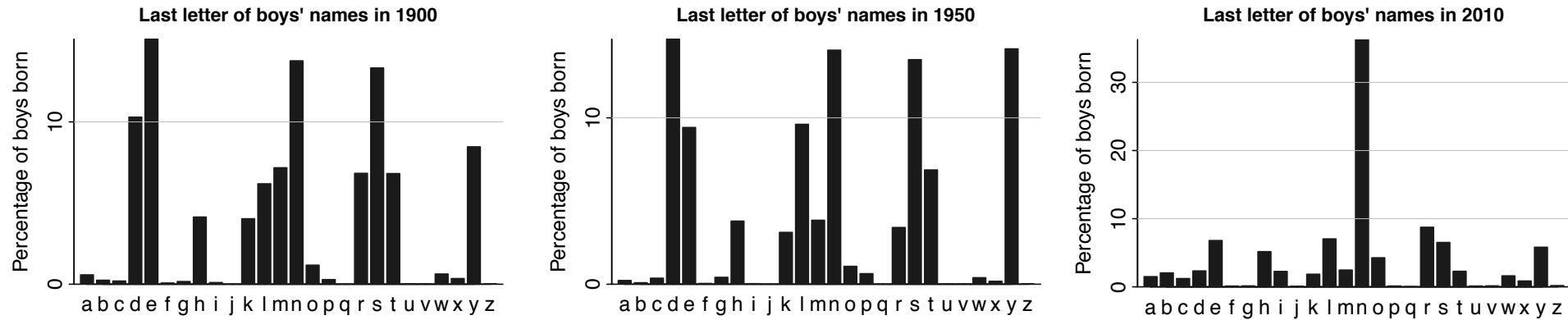
Michael, David, Thomas, Larry, . . .

Last letter of boys' names in 2010



Ethan (#2), Jayden (4), Aiden (9), Mason (12), Logan (17), Benjamin (22), Ryan (23), Jackson (25), John (26), also 27, 28, 29, 31, 32, 36, 37, 40, ...

The trend in last letters of boys' names



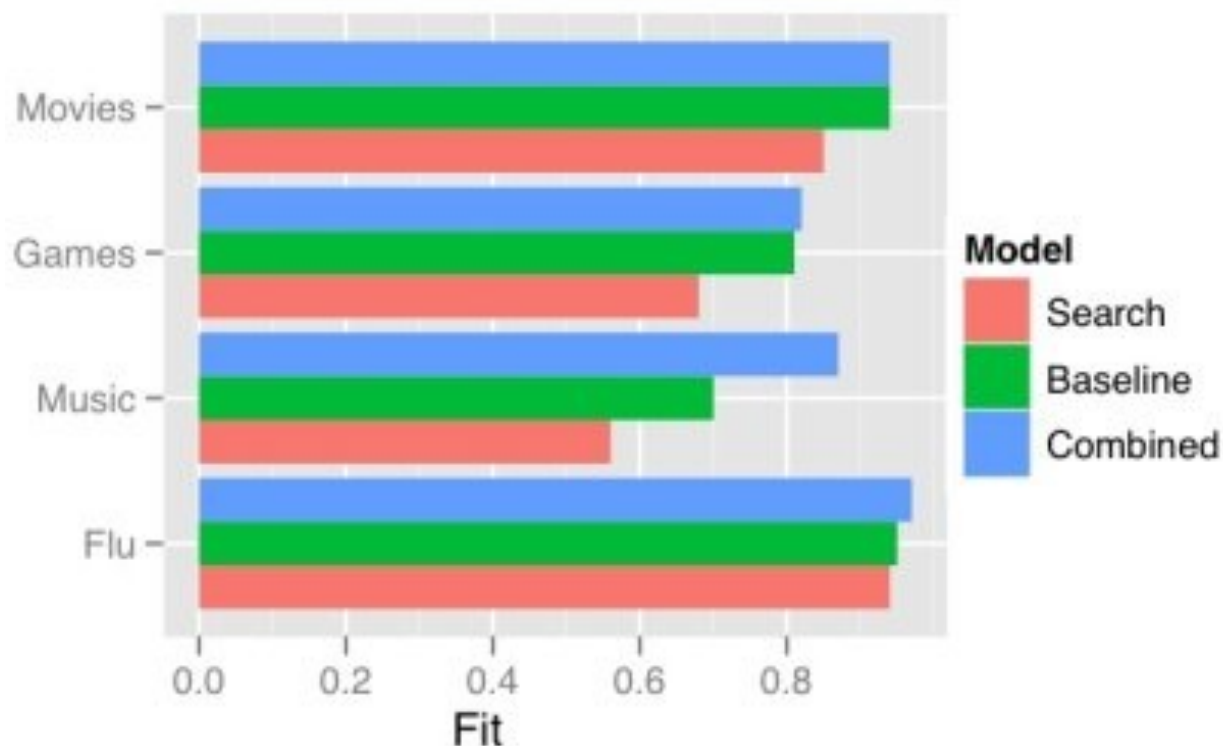
- The long tail . . . and the paradox of freedom

Part 5: Some practical tips

- Line plots and small multiples
- But not just “sparklines”
- Avoid the graphical equivalent of the data dump
- Don’t try to cram everything into one plot
- Combine graphics with text
 - A picture plus 1000 words is worth more than two pictures or 2000 words

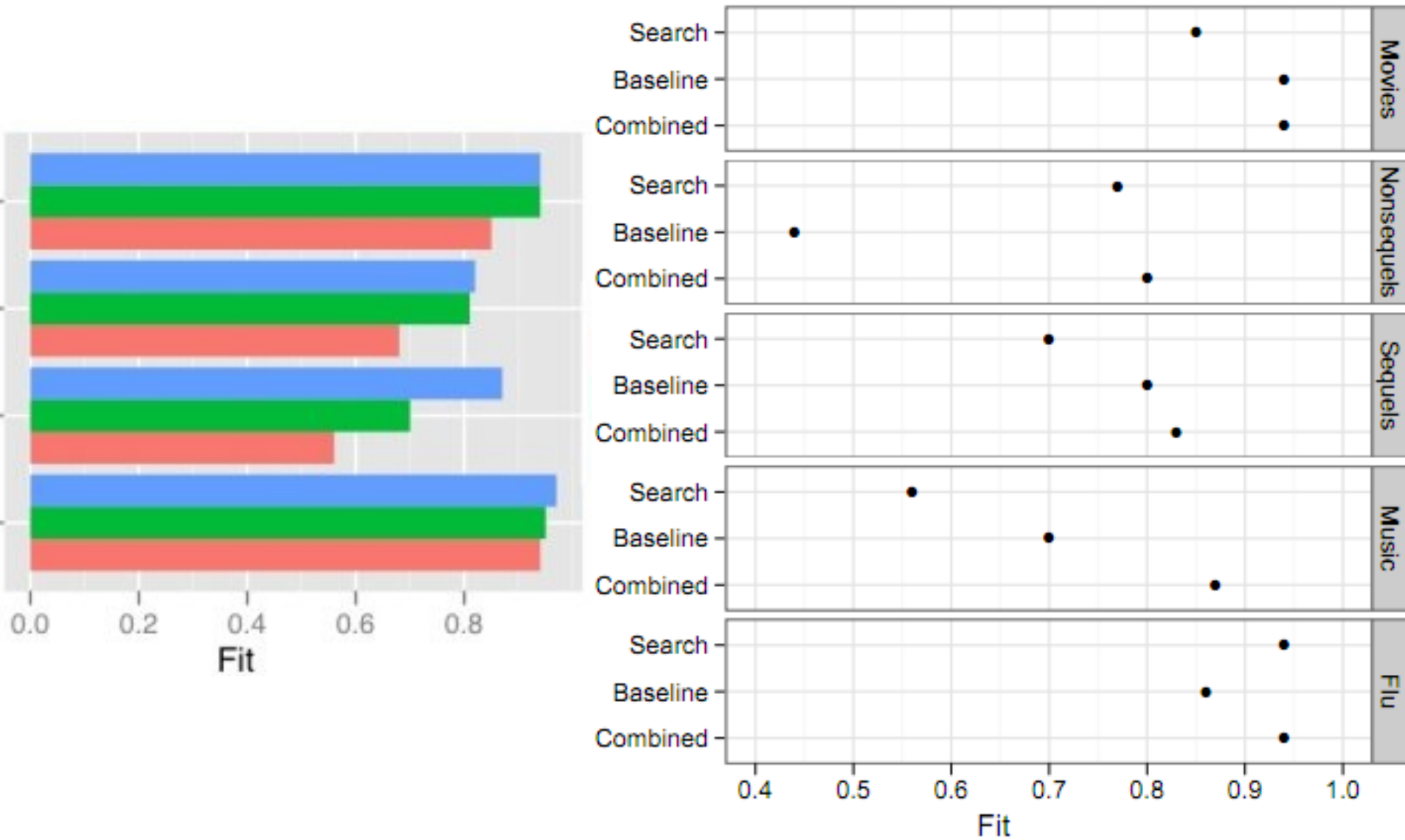
Cleveland's principle

- Always ask: What is the comparison?
- Example: an analysis from marketing

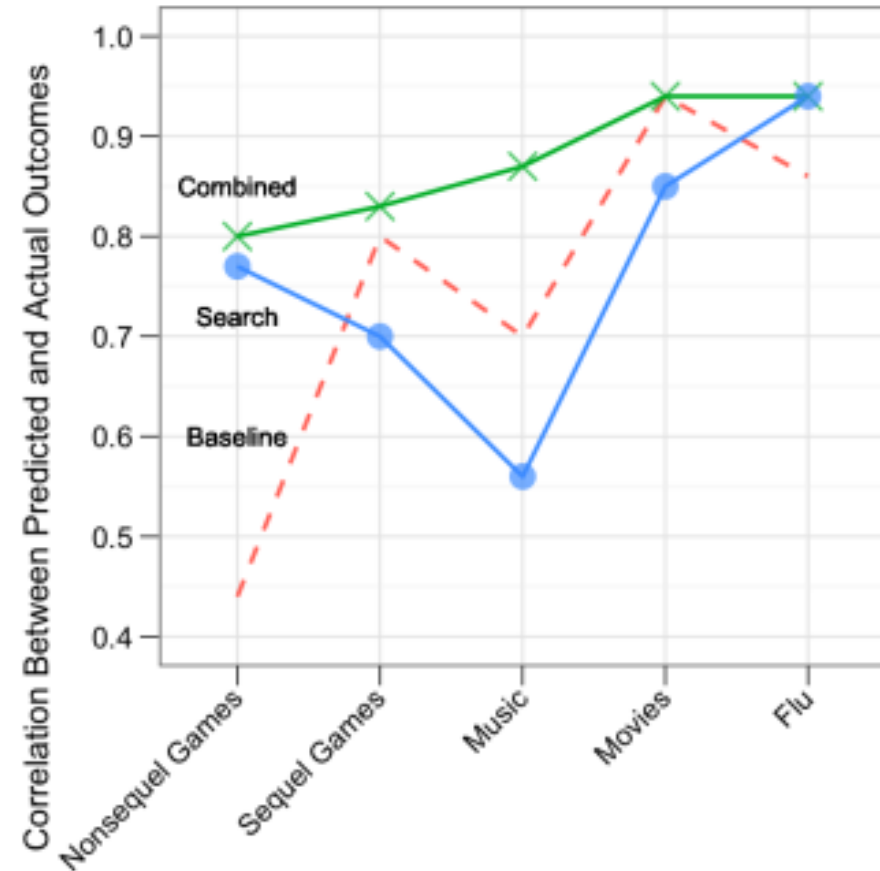
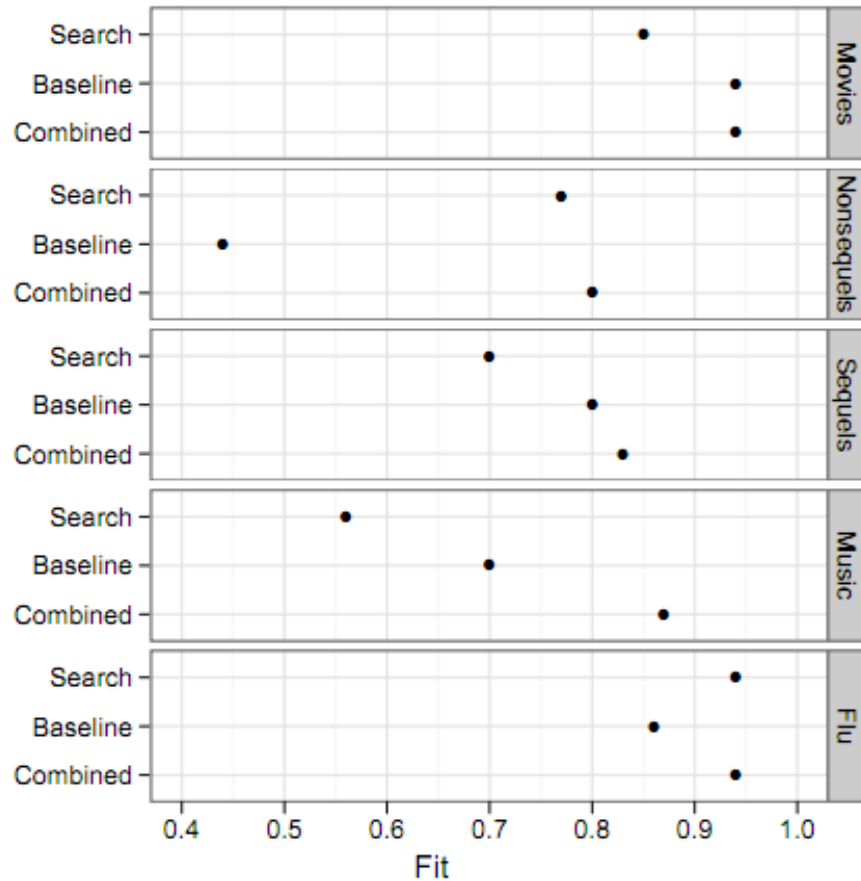


Sharad Goel,
Yahoo Research

Improvement?



Parallel coordinate plot is better

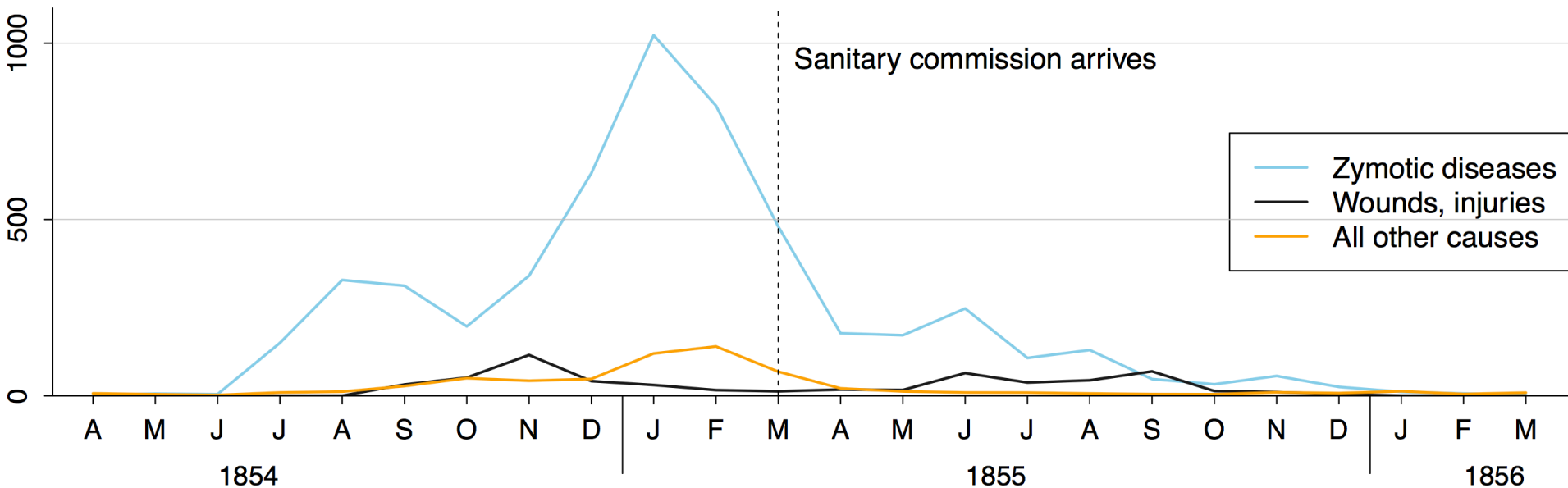


- Consider the **comparisons** you can make

Who do you want to be?

- George Orwell
- Martin Amis
- Pieter Breughel
- Scott Adams
- Chris Ware
- ?

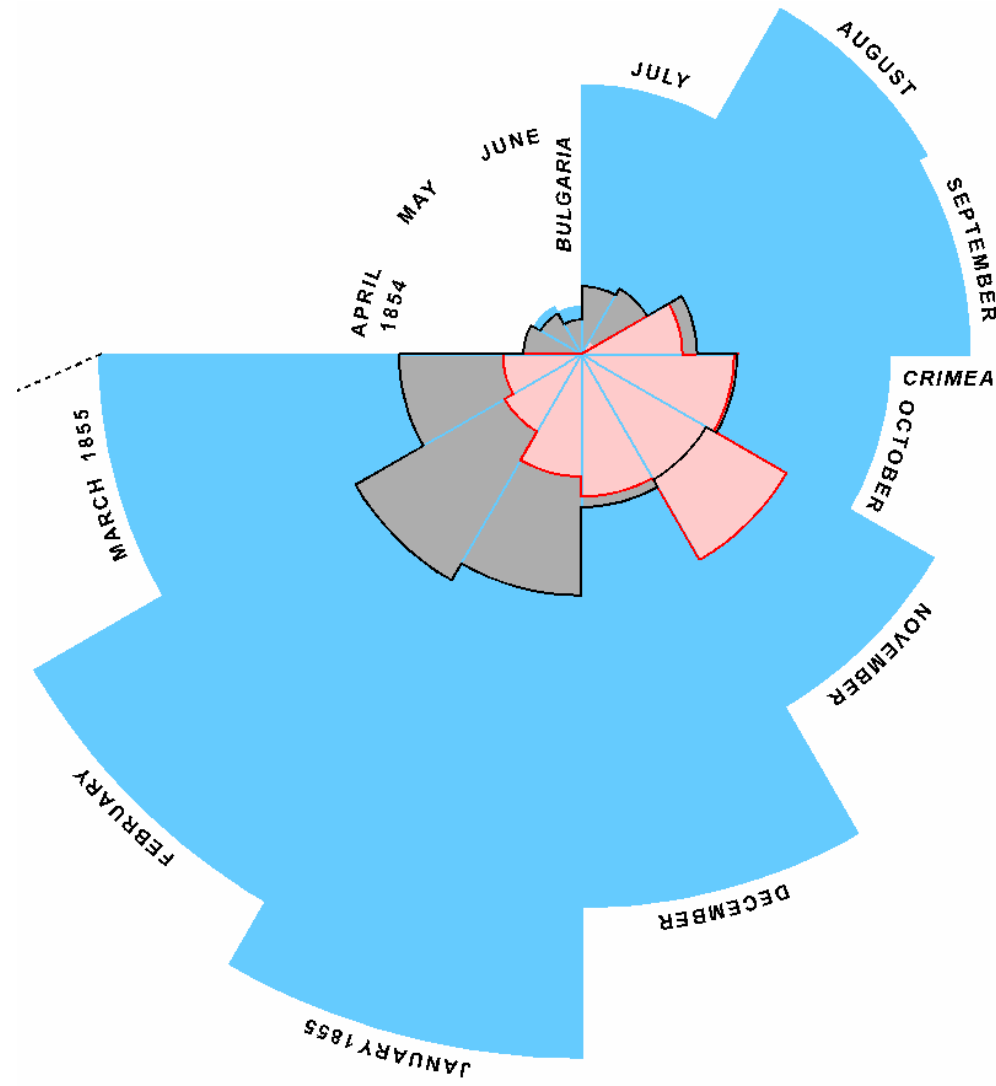
Summary (1): Statistical graphics



- Displaying patterns clearly
(Expected and unexpected)
- Alternative to tables and numerical summaries
- Comparisons

Summary (2): Infovis

- Engages the viewer
- Puzzles and the joy of recognition
- Data vis instead of or cartoon
- One single display
- The correspondence principle





Conclusions

- Infovis and stat graphics can work together
- But we should respect different goals and be open about tradeoffs
 - Puzzles vs. clarity
 - Originality vs. focus on comparisons
 - ...
- Multiple displays