Big Ideas to help statistics students learn to 'think with data'

Nicholas J. Horton

Department of Mathematics and Statistics
Amherst College, Amherst, MA, USA

Pickard Lecture, September 29, 2016

nhorton@amherst.edu
http://nhorton.people.amherst.edu

Thanks and Acknowledgements

Kudos to many wonderful collaborators:

- Project MOSAIC: Danny Kaplan (Macalester College), Randy Pruim (Calvin College), and Ben Baumer (Smith College)
- Johanna Hardin (Pomona College) and the undergraduate guidelines working group
- Megan Mocko (University of Florida) and Michelle Everson (Ohio State University) and the revised GAISE college report group
- my colleagues at Amherst and the ASA

Thanks and Acknowledgements

Kudos to many wonderful collaborators:

- Project MOSAIC: Danny Kaplan (Macalester College), Randy Pruim (Calvin College), and Ben Baumer (Smith College)
- Johanna Hardin (Pomona College) and the undergraduate guidelines working group
- Megan Mocko (University of Florida) and Michelle Everson (Ohio State University) and the revised GAISE college report group
- my colleagues at Amherst and the ASA
- those listed in Table 2 of Horton and Hardin (TAS 2015) for an incomplete bibliography

(More) thanks and Acknowledgements

- George Cobb, Marcello Pagano, and Nan Laird for their guidance and support on this journey

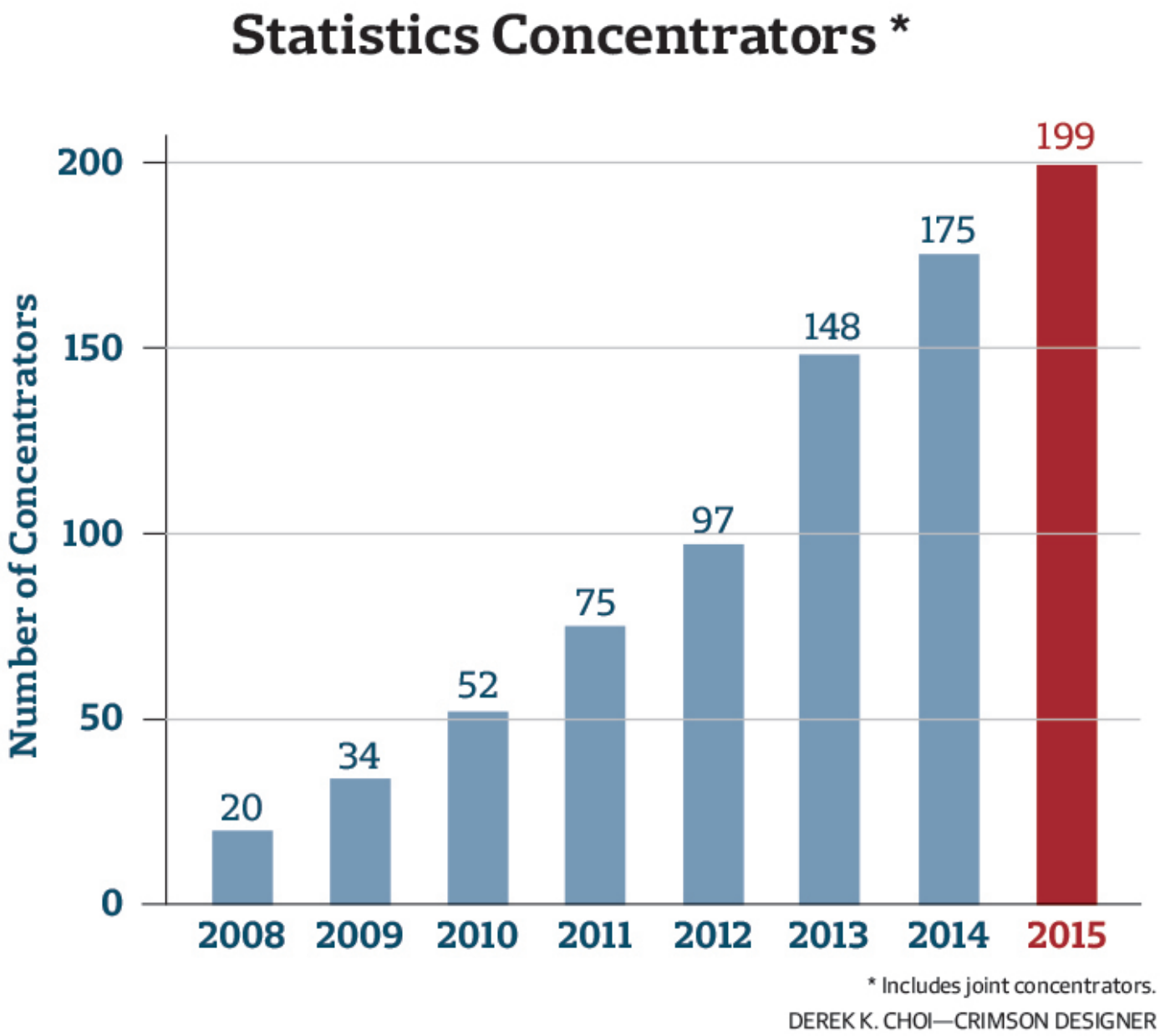
(More) thanks and Acknowledgements

- George Cobb, Marcello Pagano, and Nan Laird for their guidance and support on this journey
- my students

(More) thanks and Acknowledgements

- George Cobb, Marcello Pagano, and Nan Laird for their guidance and support on this journey
- my students
- my parents and family

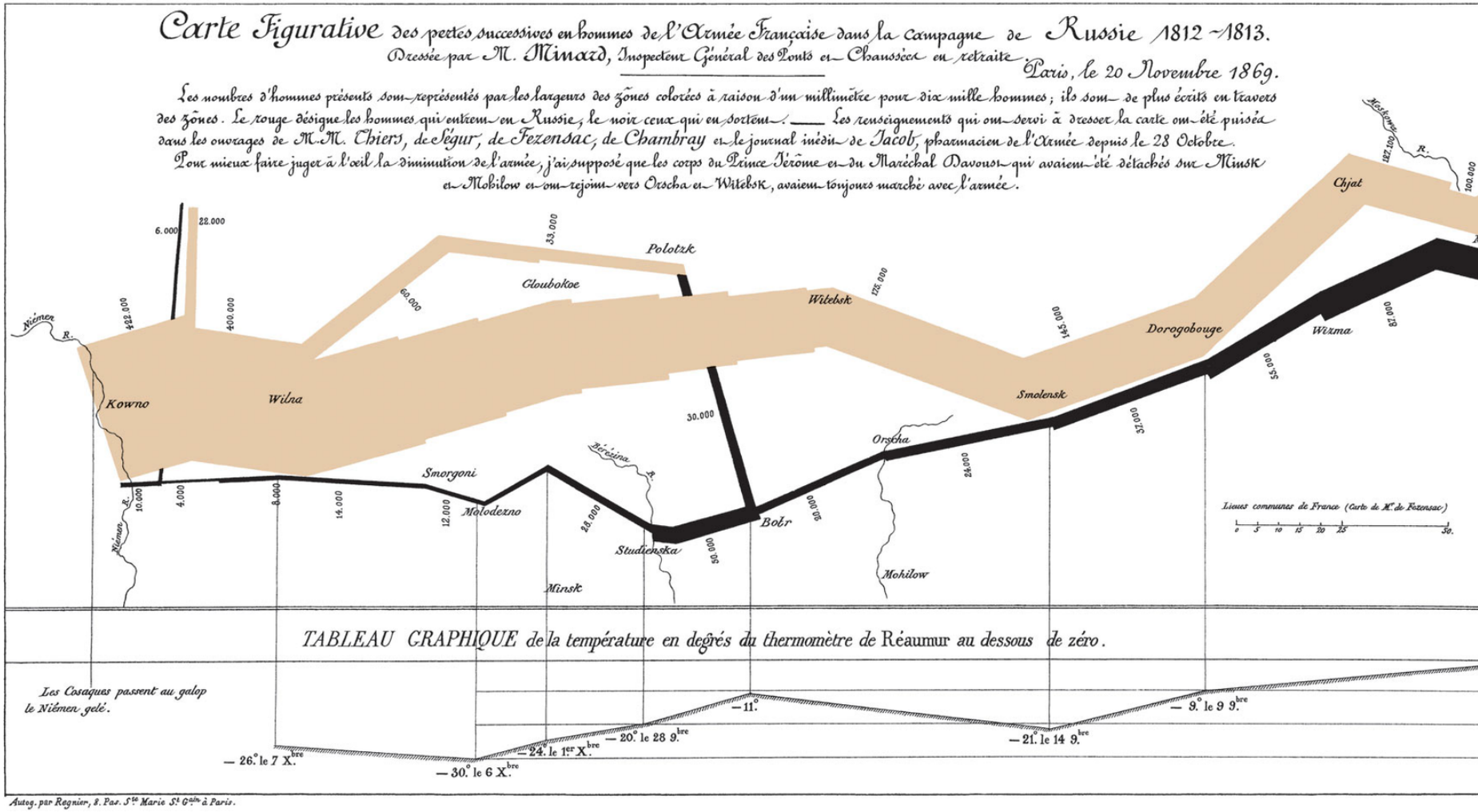
Degrees in statistics at Harvard (Source: Harvard Crimson)



Plan

- a glimpse into the past
- a vision for statistics and statistics education
- some big ideas
- closing thoughts

Minard and Napoleon's campaign (1812)



Statistics circa 1839

A politician, a pastor, a lawyer, a physician, and a poet walk into a bar...

Statistics circa 1839

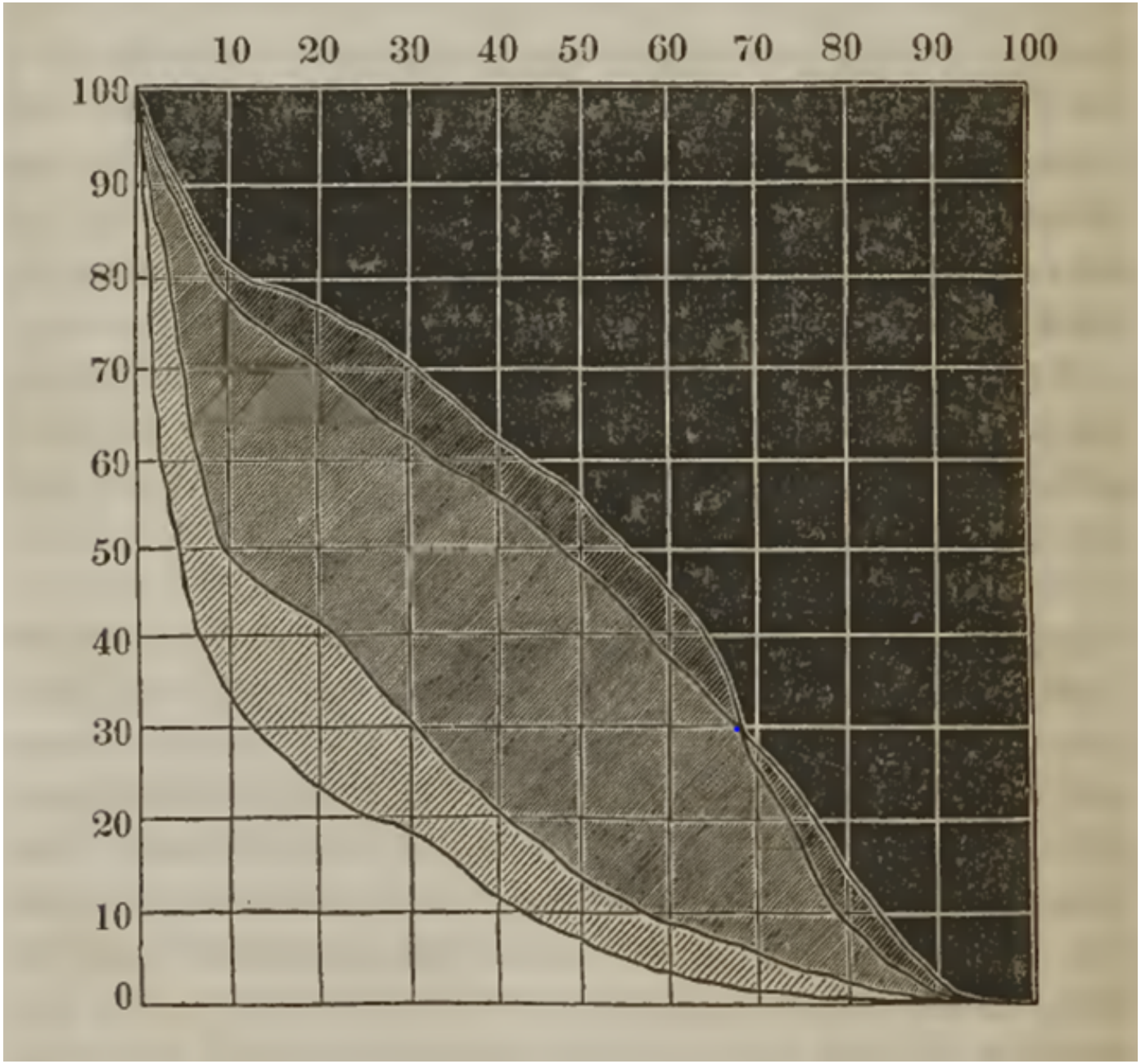
- American Statistical Association (ASA) founded by Lemuel Shattuck (politician) and a former pastor, a lawyer, a physician, and a poet.

Statistics circa 1839

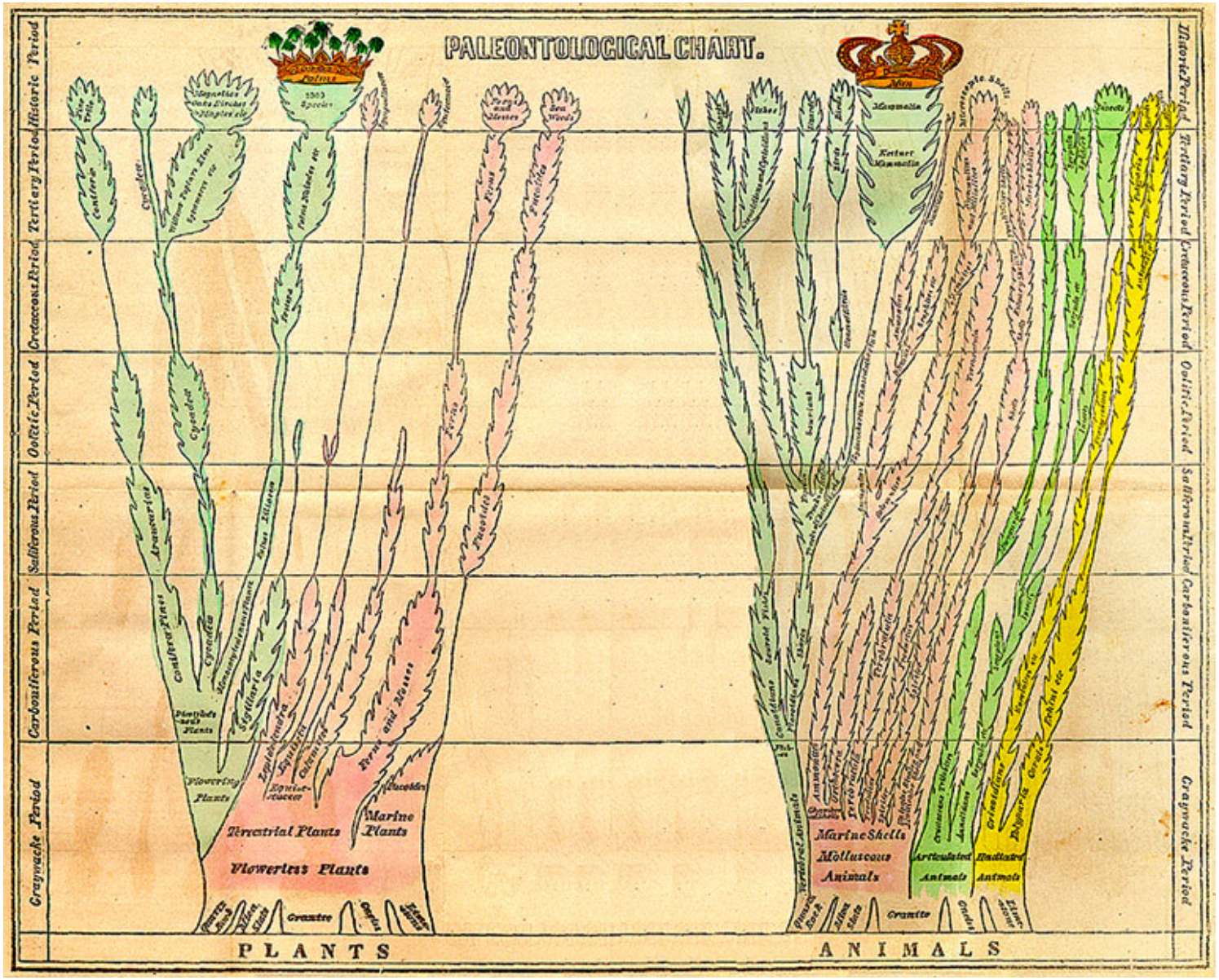
- American Statistical Association (ASA) founded by Lemuel Shattuck (politician) and a former pastor, a lawyer, a physician, and a poet.
- Raymond Pearl (JASA, 1940) described this peculiar group as:

  an odd lot of fish, differing widely from each other in most respects, but all alike in one. Each of them had what the psychiatrists nowadays call a compulsion neurosis impelling him to tinker with numbers and fiddle with figures. Their souls cried out for tabulations in the same way that the prohibitionist of later times yearned for his daily ration of Peruna.

Shattuck and mortality report (1850)



Edward Hitchcock (1840) (early member of ASA)

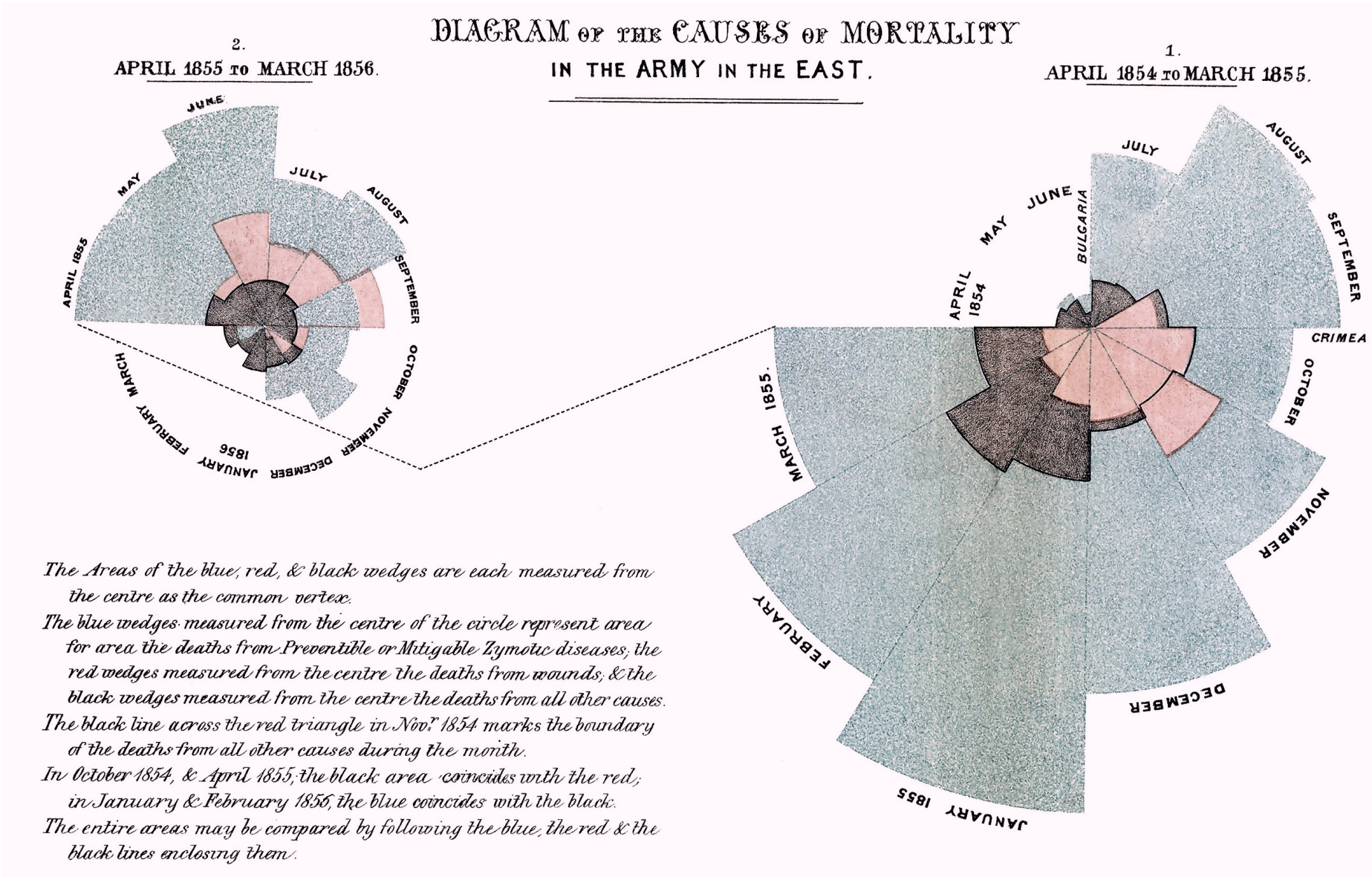


Edward Hitchcock (early member of ASA and President of Amherst College)

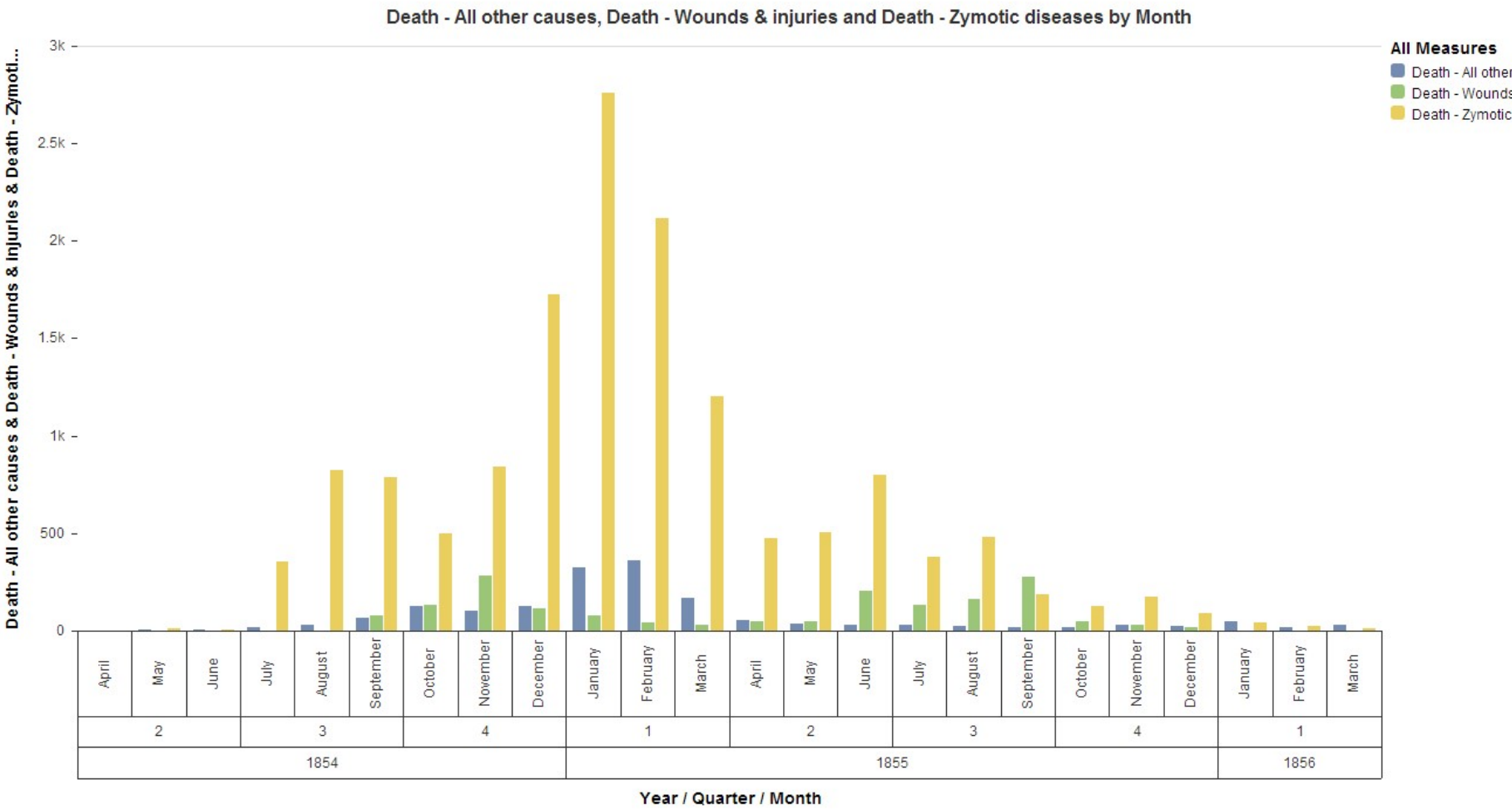
Florence Nightingale (see Utts, *Amstat News* June 2016)

Fast forward a few years across the pond...

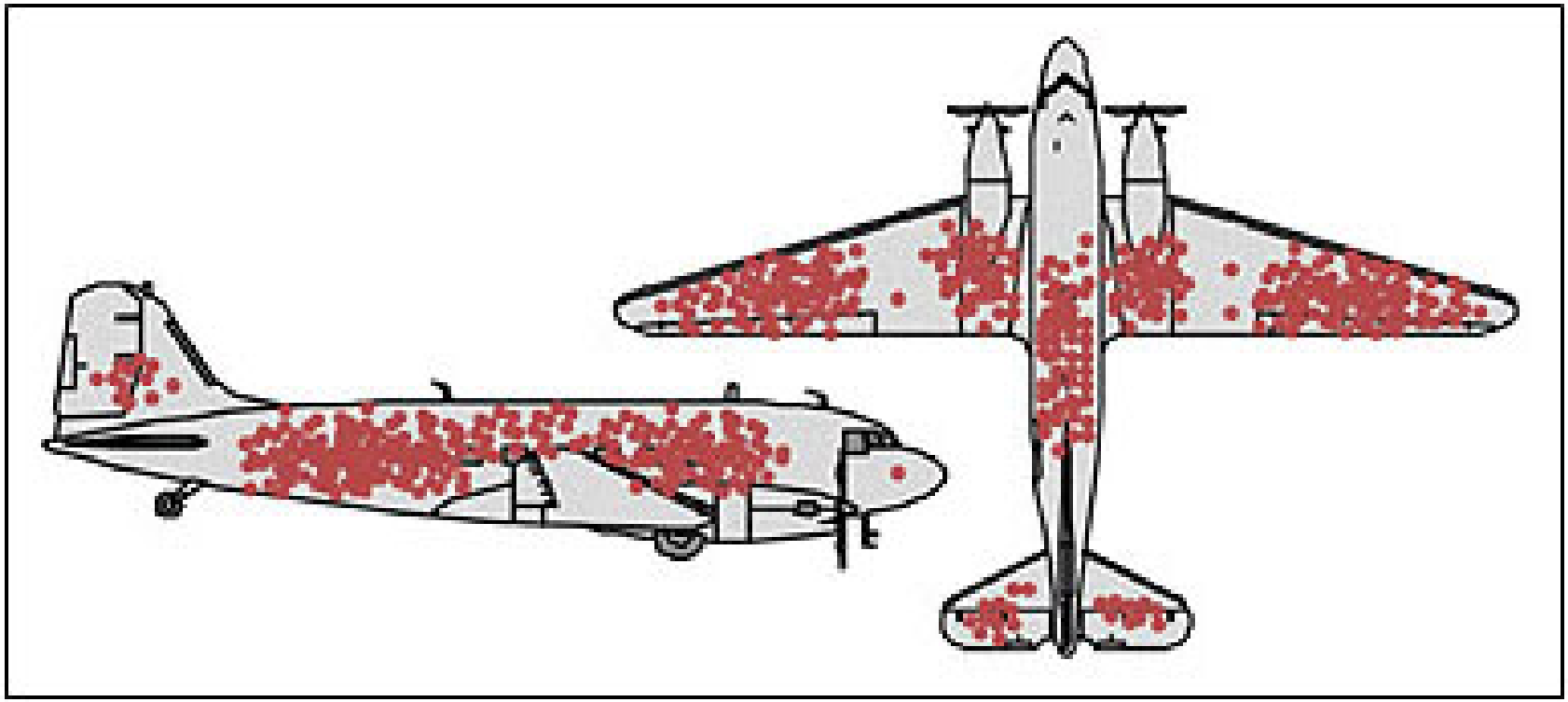
Florence Nightingale (see Utts, *Amstat News* June 2016)



reworked graph: (Source: itelligencegroup.com)



Wald and World War II airplanes (credit to Cameron Moll)



Credit: Cameron Moll

Cobb (1982)

WILL STATISTICS TAKE THE LIBERAL ARTS BY STORM OR BY STEALTH?
REFLECTIONS ON EIGHT YEARS AT MOUNT HOLYOKE COLLEGE

(Presented at the 25th anniversary of the founding of the Department of Statistics, Harvard University, April 16, 1982.)

Cobb (1982)

Students get to do what statisticians do: analyze non-trivial datasets by considering a variety of models, using the imagination and developing their judgment in the process.

Cobb (1982)

I believe it is the use of imagination and judgment that makes our subject appealing. We owe it to our students not to keep that a secret.

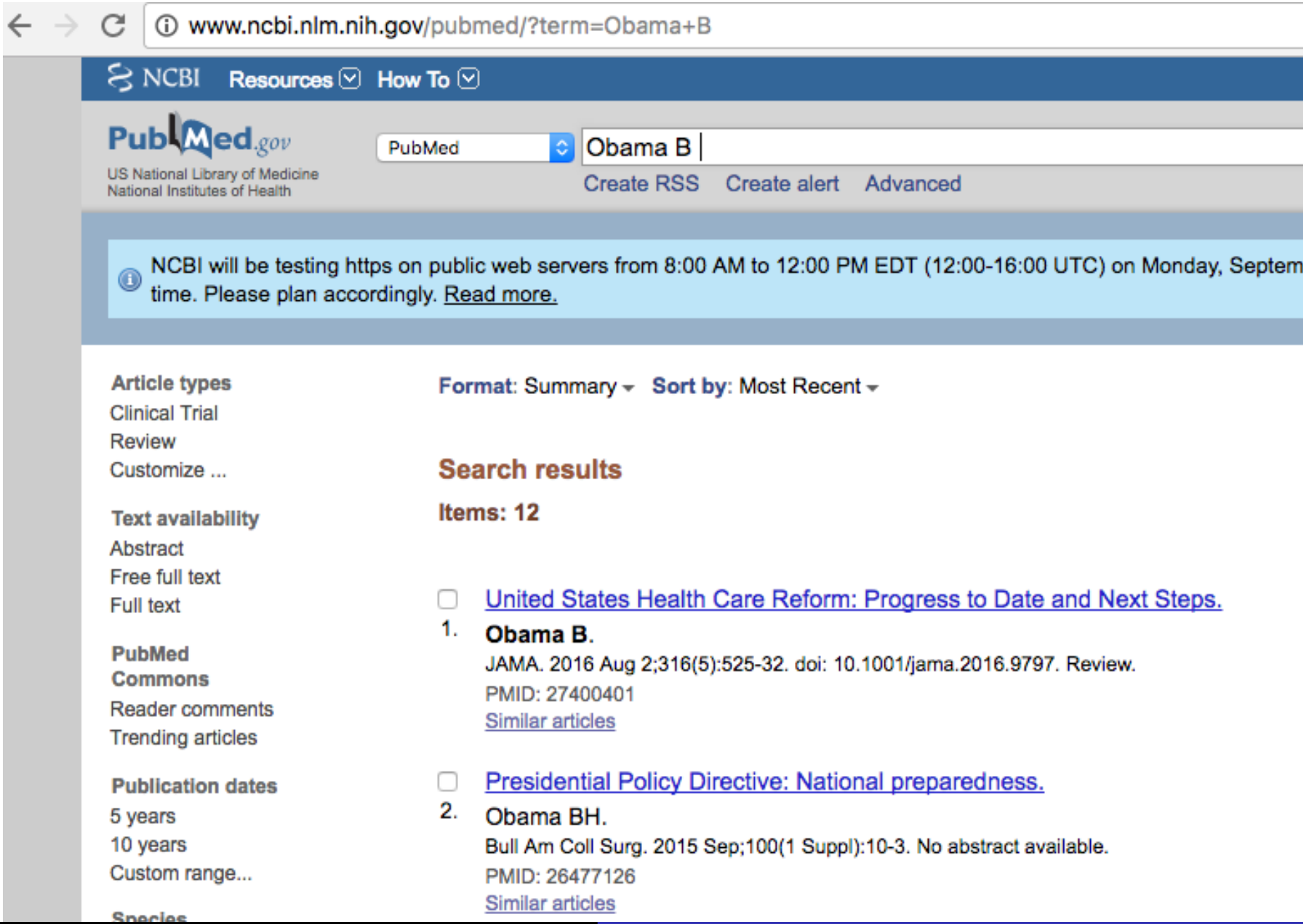
Where does this leave us today?

Obama video keynote for Hadoop/Strata conference:
`https://www.youtube.com/watch?v=vbb-AjiXyh0`

Knowledge of CS?

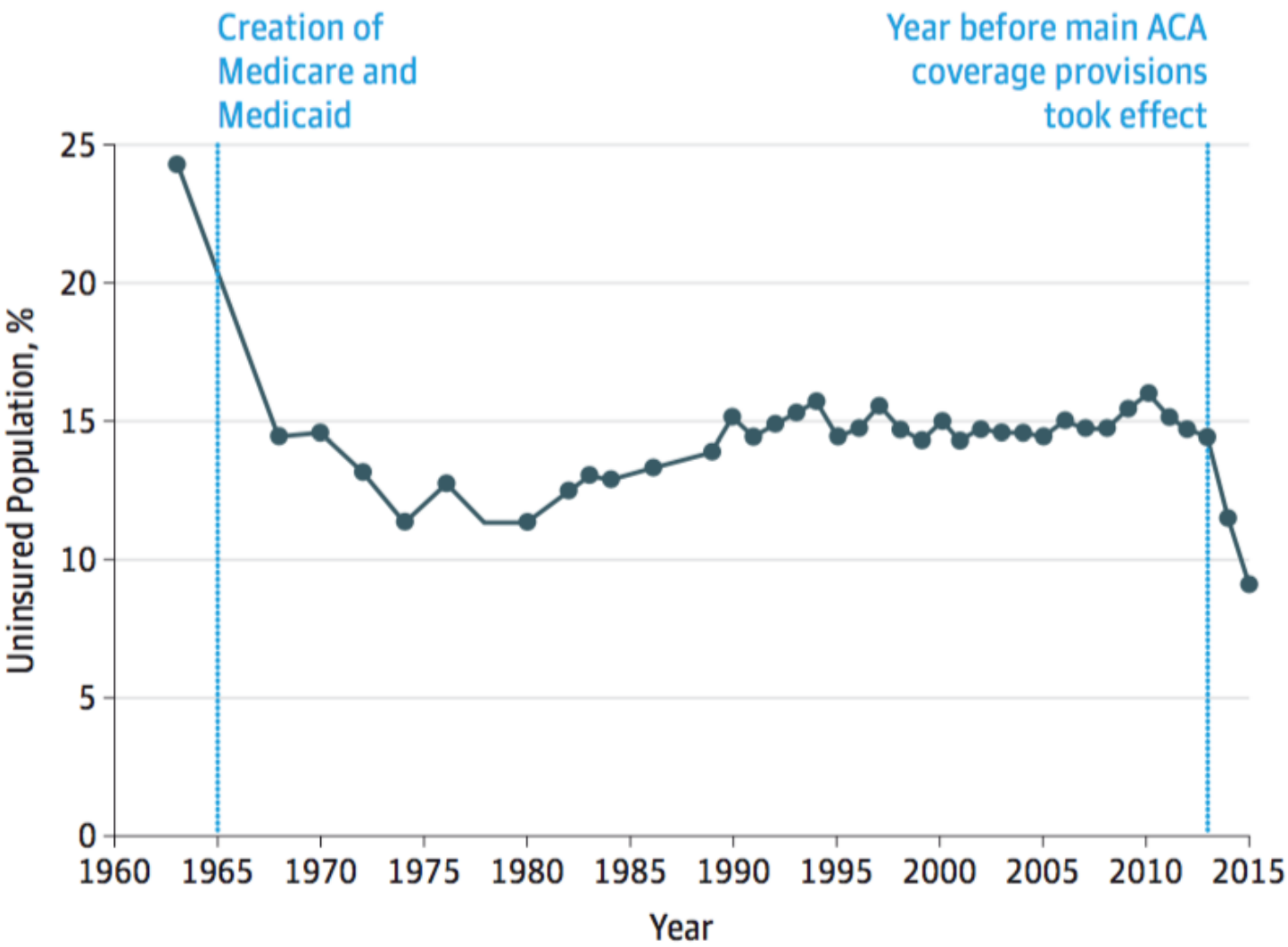
Obama and the bubblesort :
`https://www.youtube.com/watch?v=k4RRi_ntQc8`

Knowledge of stats? Obama's recent publications



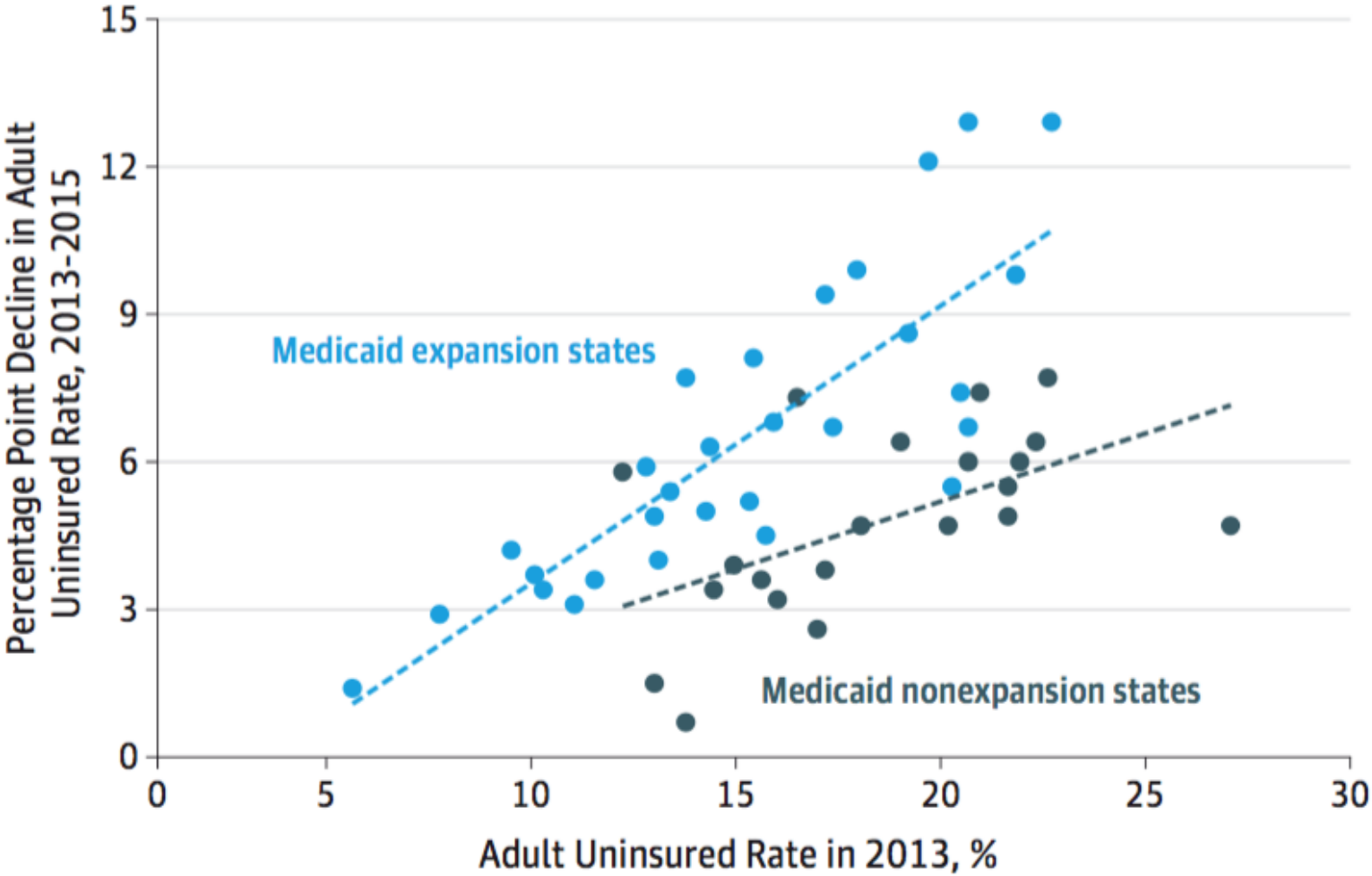
Obama's recent single author JAMA paper

Figure 1. Percentage of Individuals in the United States Without Health Insurance, 1963-2015

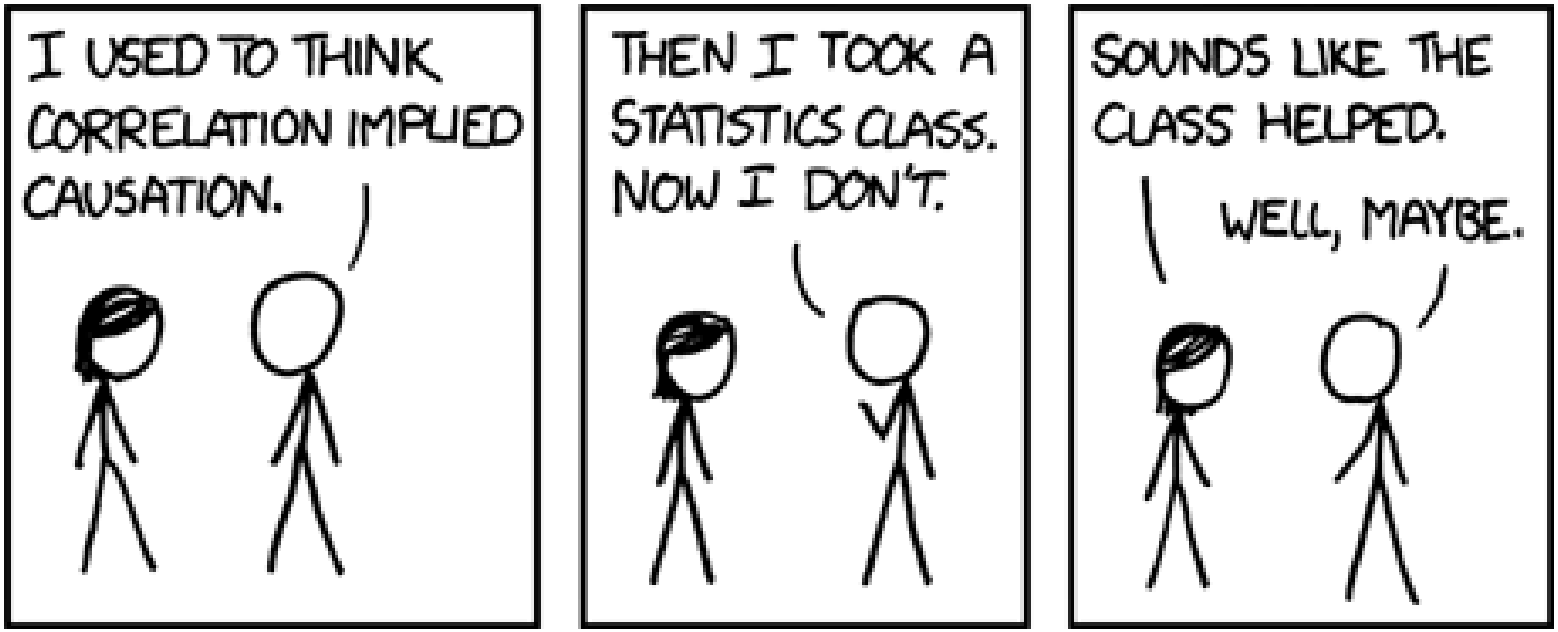


Obama's recent JAMA paper

Figure 2. Decline in Adult Uninsured Rate From 2013 to 2015 vs 2013 Uninsured Rate by State



Is statistics a dirty word? (Source: xkcd.com)



a vision for statistics education (John Bailar)

As academic statisticians, we are missing the boat. We are barking up the wrong tree. ... The kinds of statistics that we teach in undergraduate and especially in graduate programs have almost nothing to contribute to anything that matters. ... Then we wonder why the world passes us by.

a vision for statistics education

The Committee on Applied and Theoretical Statistics (CATS) noted widespread sentiment in the statistical community that upper-level undergraduate and graduate curricula for statistics majors ... are currently structured in ways that do not provide sufficient exposure to modern statistical analysis, computational and graphical tools, communication skills, and the ever growing interdisciplinary uses of statistics.

The growth that statistics has undergone is often not reflected in the education that future statisticians receive. There is a need to incorporate more meaningfully into the curriculum the computational and graphical tools that are today so important to many professional statisticians. There is a need for improved training of statistics students in written and oral communication skills, which are crucial for effective interaction with scientists and policy makers.

A jump back to 1992...

A jump back to 1992...

A jump back to 1992...

previous quotes: CATS Boston workshop in 1992

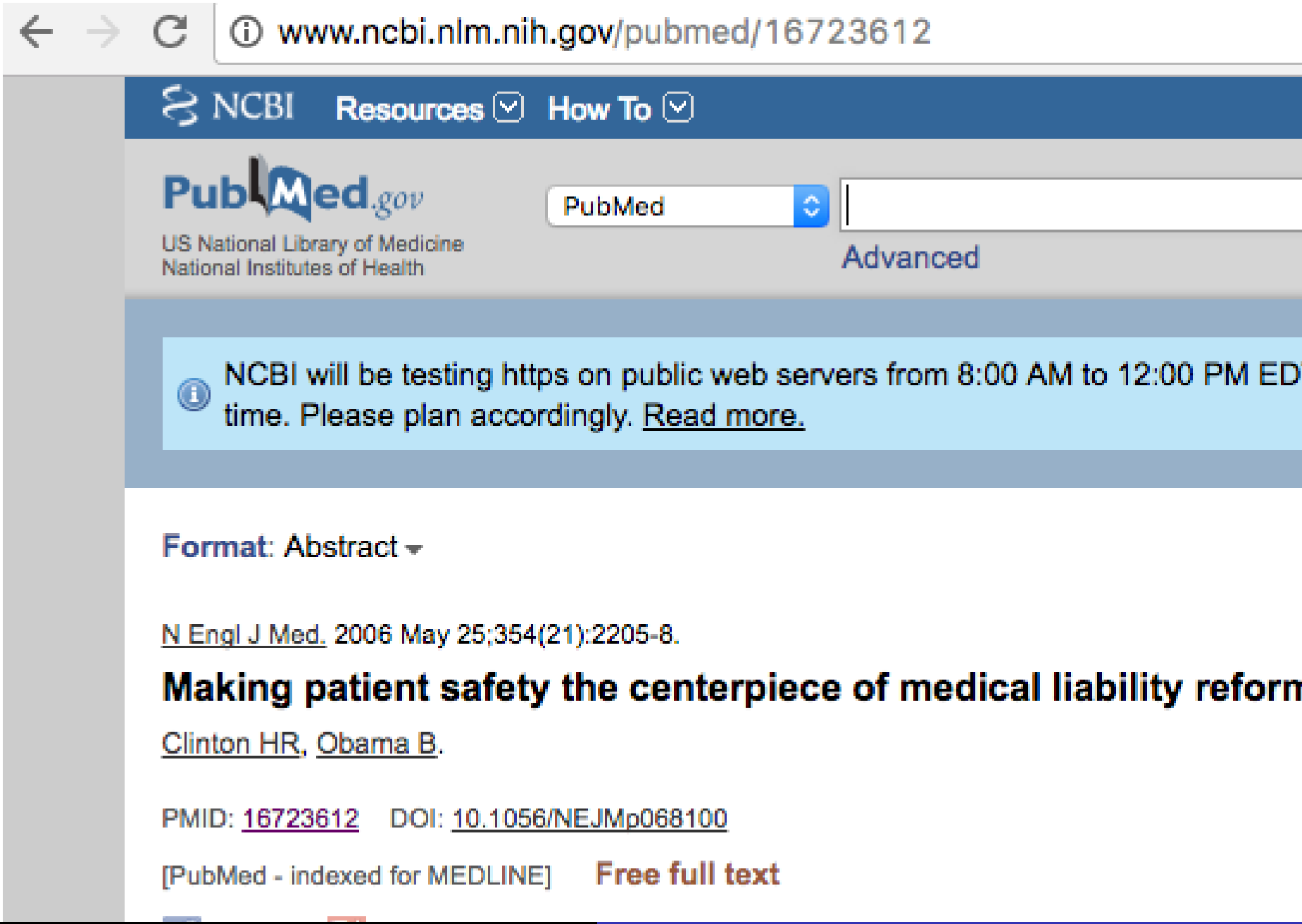
The current curriculum in most statistics departments is, however, entirely too focused on hypothesis testing (Ed Rothman).

previous quotes: CATS Boston workshop in 1992

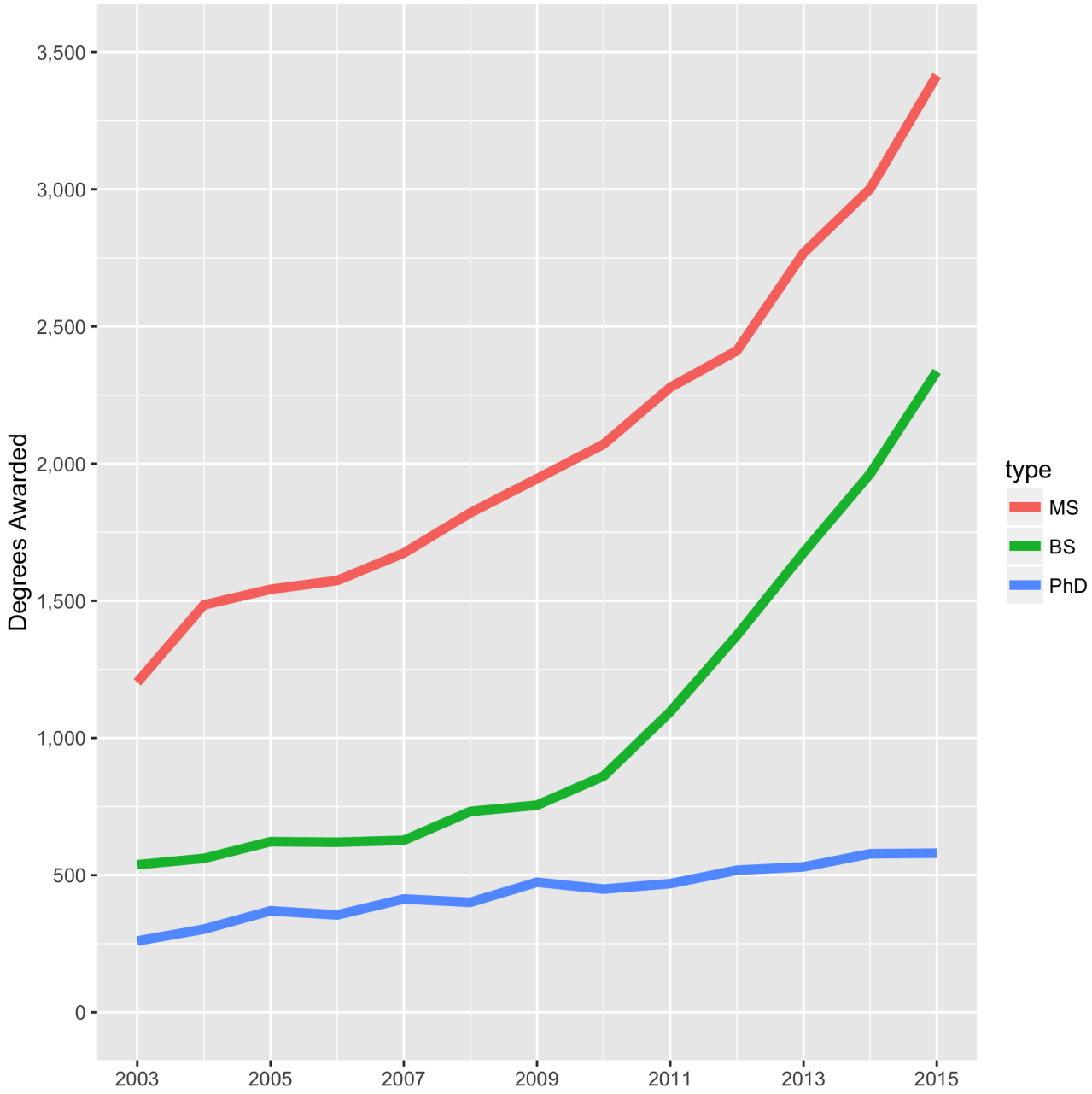
The current curriculum in most statistics departments is, however, entirely too focused on hypothesis testing (Ed Rothman).

We risk being ignored if we do not stay relevant. (Carl Morris)

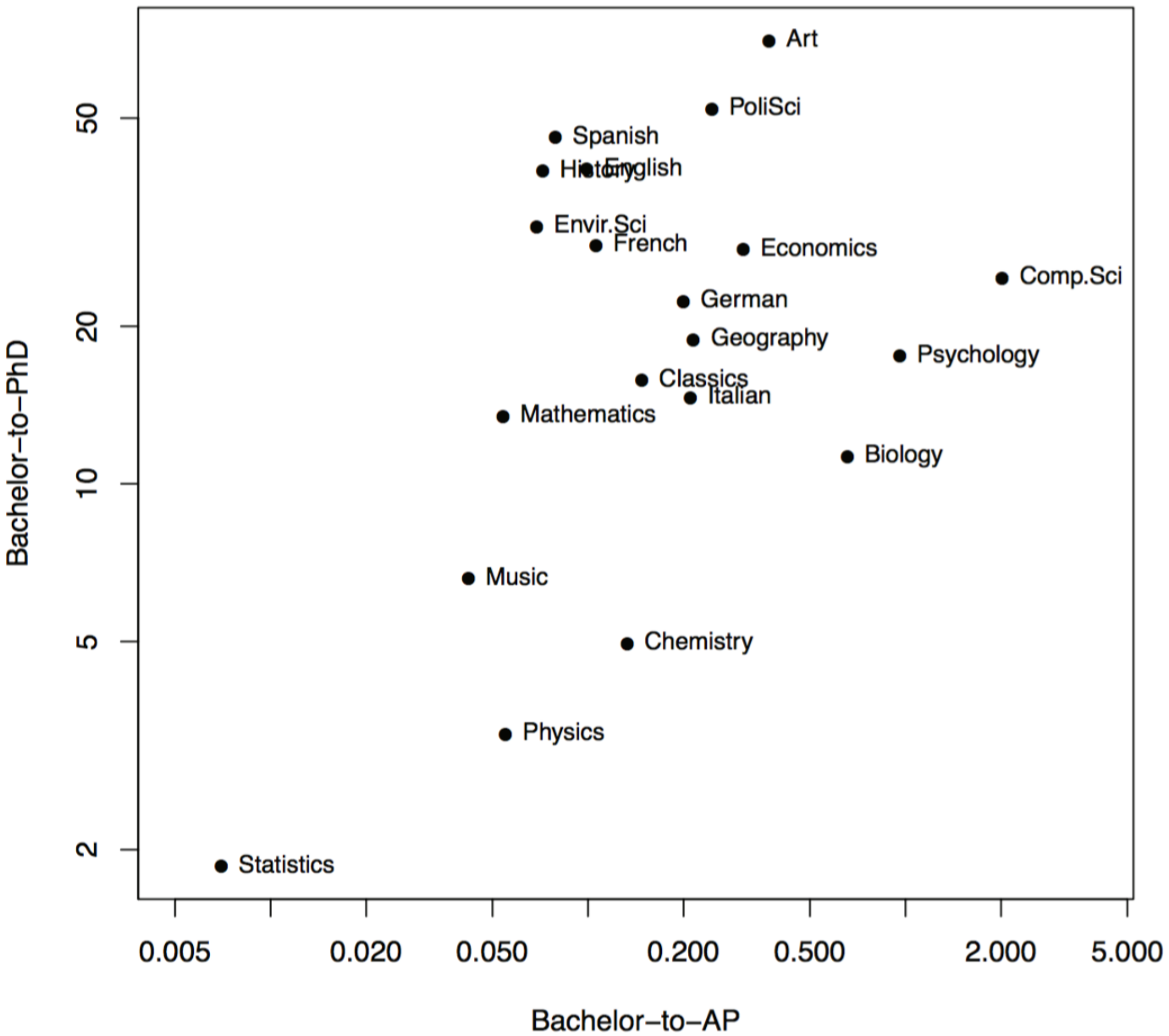
Order of authors?



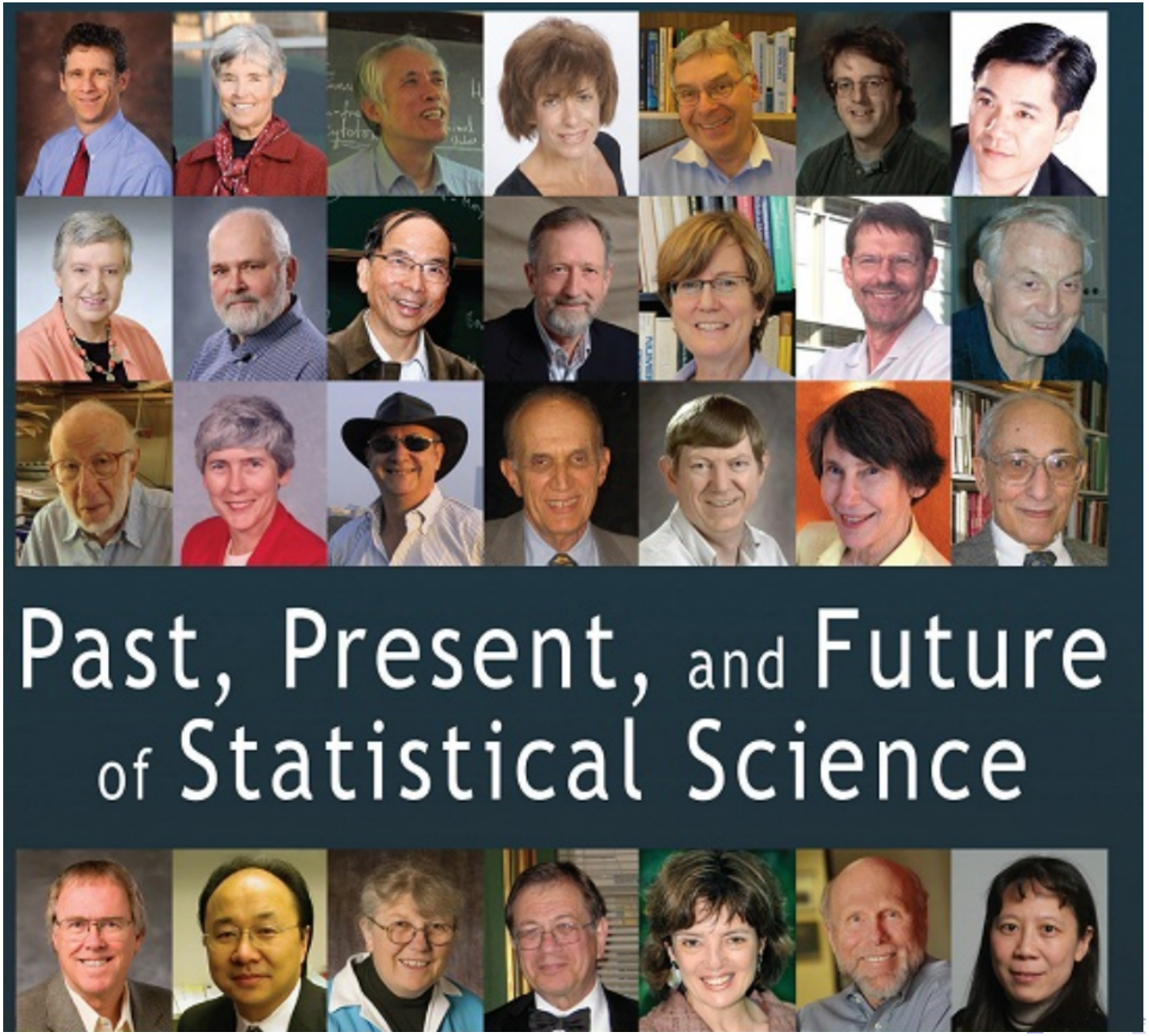
Progress: degrees in statistics over time (Source: IPEDS)



But where are the majors? (Johnstone, COPSS PPF)



Class use of *Past, Present, and Future*



ASA Undergrad Guidelines, 2014

- This is an exciting time to be a statistician.
- The contribution of the discipline of statistics to scientific knowledge is widely recognized with increasingly positive public perception.

ASA Undergrad Guidelines, 2014

We are concerned that many of our graduates do not have sufficient skills to be effective in the modern workforce. Thomas Lumley (personal communication) has stated that our students know how to deal with $n \to \infty$, but cannot deal with a million observations.

If statistics is the science of learning from data, then our students need to be able to "think with data" (as Diane Lambert of Google has so elegantly described).
- Horton and Hardin (TAS, 2015)

Some big ideas

Audience: preaching to the choir...

Caveat: there's nothing new under the sun (ancient Xiao-Li proverb)

1. Increasing importance of data science and computation

Big idea #1: Data science and computation

ASA undergrad guidelines for statistics programs:

- Working with data requires extensive computing skills.
- To be prepared for statistics and data science careers, students need the ability to access and wrangle data in various ways, and the ability to perform algorithmic problem-solving.
- In addition to more traditional mathematical and statistical skills, students should be fluent in higher-level programming languages and facile with database systems.

20 years of R: one of many solutions

The New York Times

Business Computing

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS | OPINION

Data Analysts Captivated by R's Power

Left, Stuart Isett for The New York Times; right, Kieran Scott for The New York Times

R first appeared in 1996, when the statistics professors Robert Gentleman, left, and Ross Ihaka released the code as a free software package.

By ASHLEE VANCE
Published: January 6, 2009

To some people R is just the 18th letter of the alphabet. To others, it's the rating on racy movies, a measure of an attic's insulation or what pirates in movies say.

FACEBOOK
TWITTER

Think, Pair, Share

An analyst wants to calculate the mean `drugrisk` score of subjects in the HELP clinical trial by gender. What's the simplest way to do this in base R? Talk with your neighbor about how you would accomplish this task.

Possible answer

```
> with(HELPmiss, aggregate(drugrisk, by=list(sex),
  FUN=mean, na.rm=TRUE, simplify=TRUE))
  Group.1        x
1    male 1.904762
2  female 1.756757
```

Possible answer

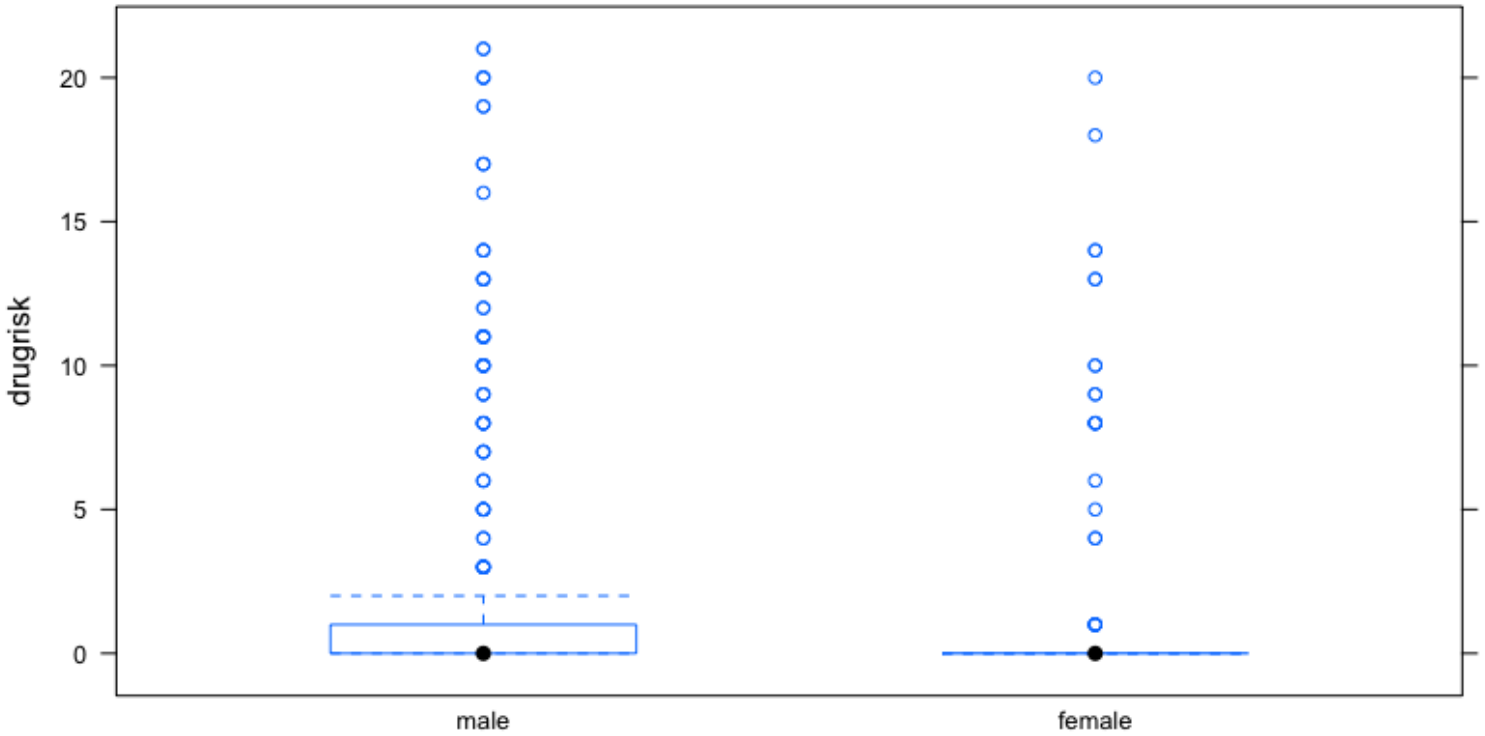
```
> with(HELPmiss, tapply(drugrisk, sex, mean, na.rm=TRUE))
    male   female
1.904762 1.756757
```

My preferred answer

```
> library(mosaic)
> favstats(drugrisk ~ sex, data=HELPrct)
     sex min Q1 median Q3 max mean   sd   n missing
1   male   0  0      0  1  21 1.90 4.37 357       2
2 female   0  0      0  0  20 1.76 4.15 111       0
```

mosaic modeling language (Y ~ X)

```
> bwplot(drugrisk ~ sex, data=HELPmiss)
```



mosaic modeling language (Y ~ X)

```
> lm(drugrisk ~ sex, data=HELPmiss)

Coefficients:
(Intercept)    sexfemale
      1.905       -0.148
```

One simple approach to:

- generate descriptive statistics
- create graphical displays
- fit regression models

Less Volume, More Creativity

Many of the guiding principles of the `mosaic` package reflect the "Less Volume, More Creativity" mantra of Mike McCarthy who had a large poster with those words placed in the "war room" (where assistant coaches decide on the game plan for the upcoming opponent) as a constant reminder not to add too much complexity to the game plan.

A lot of times you end up putting in a lot more volume, because you are teaching fundamentals and you are teaching concepts that you need to put in, but you may not necessarily use because they are building blocks for other concepts and variations that will come off of that … In the offseason you have a chance to take a step back and tailor it more specifically towards your team and towards your players."

Mike McCarthy, Head Coach, Green Bay Packers

Enough R for Intro Stats

Numerical Summaries

These functions have a formula interface to match plotting.

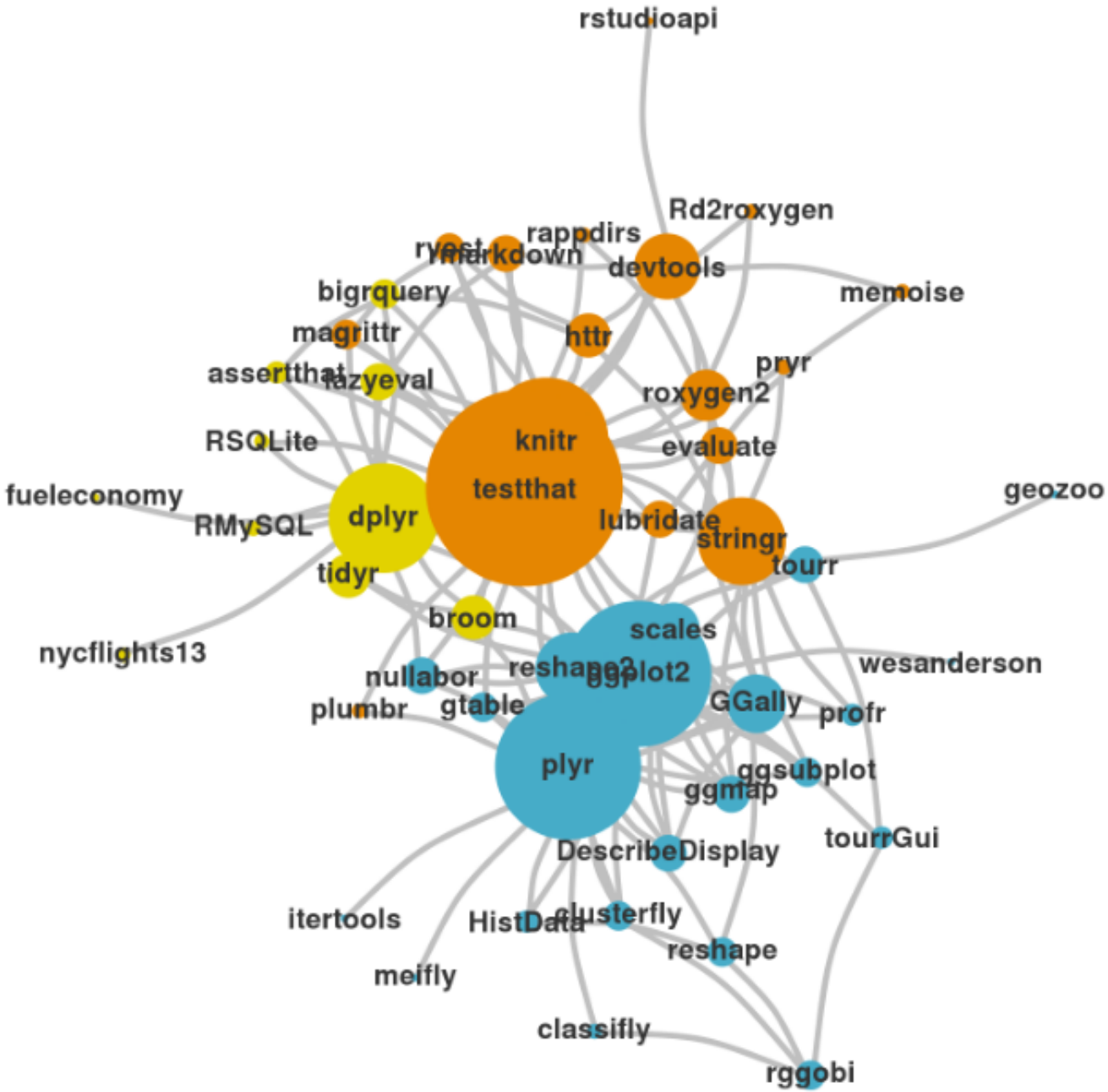
```
favstats()    # mosaic
tally()       # mosaic
mean()        # mosaic augmented
median()      # mosaic augmented
sd()          # mosaic augmented
var()         # mosaic augmented
diffmean()    # mosaic
```

Randomization/Simulation

```
rflip()       # mosaic
do()          # mosaic
sample()      # mosaic augmented
resample()    # with replacement
shuffle()     # mosaic
rbinom()
rnorm()       # etc, if needed
```

Distributions

The advent of the 'tidyverse'



Design goals of tidyverse

- tools that work well together, each one designed for a particular task
- if you don't succeed at first, try, try again (CS prototyping)
  1. `stats::reshape()`
  2. `reshape` package
  3. `reshape2` package
  4. `tidyr` package
- compose simple steps with the pipe (%>%) operator
- connects output from one function to input of another (a la UNIX tools)
- clarifies complex data wrangling workflows

What is the pipe operator?

dplyr::**%>%**

Passes object on left hand side as first argument (or . argument) of function on righthand side.

`x %>% f(y)` *is the same as* `f(x, y)`

`y %>% f(x, ., z)` *is the same as* `f(x, y, z )`

Life without the pipe operator (nested function calls)

```
foo_foo <- little_bunny()
bop_on(
  scoop_up(
    hop_through(foo_foo, forest),
    field_mouse
  ),
  head
)
```

Life with the pipe operator

```
foo_foo %>%
  hop_through(forest) %>%
  scoop_up(field_mouse) %>%
  bop_on(head)
```

Impact of mosaic and the tidyverse

- Small number of simple idioms
- Combine to do powerful operations
- Round off rough edges of R

More big ideas

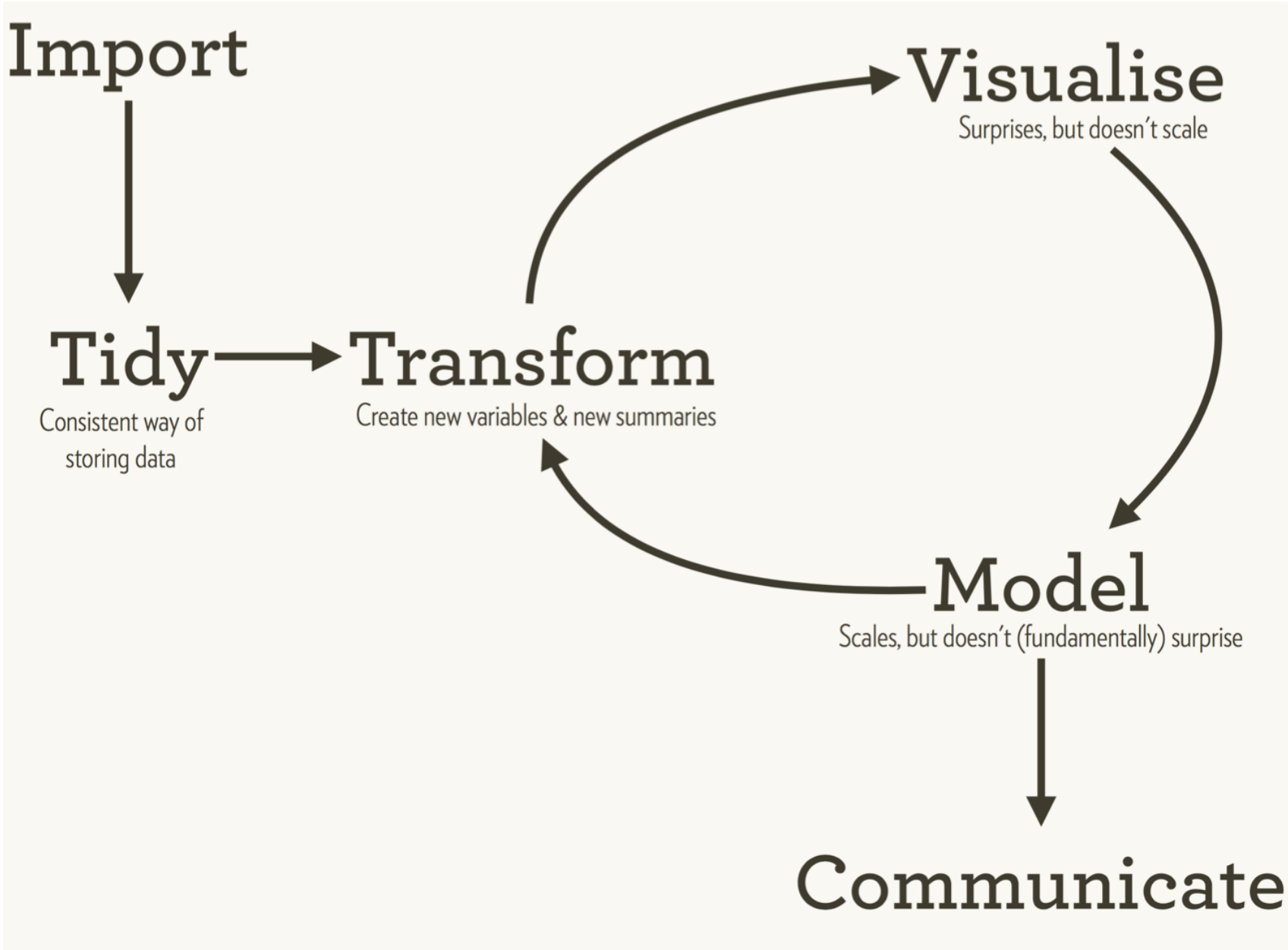
1. Increasing importance of data science and computation
2. Real applications and data

Big idea #2: Real applications and data

ASA undergrad guidelines for statistics programs:

- Data should be a major component of statistics courses.
- Programs should emphasize concepts and approaches for working with complex data and provide experiences in designing studies and analyzing non-textbook data.

Statistics and data analysis cycle (due to Wickham)



Key idioms for dealing with big(ger) data

select:	subset variables
filter:	subset rows
mutate:	add new columns
summarize:	reduce to a single row
group-by:	aggregate
join:	merge tables
gather/spread:	transpose (e.g., wide to tall)

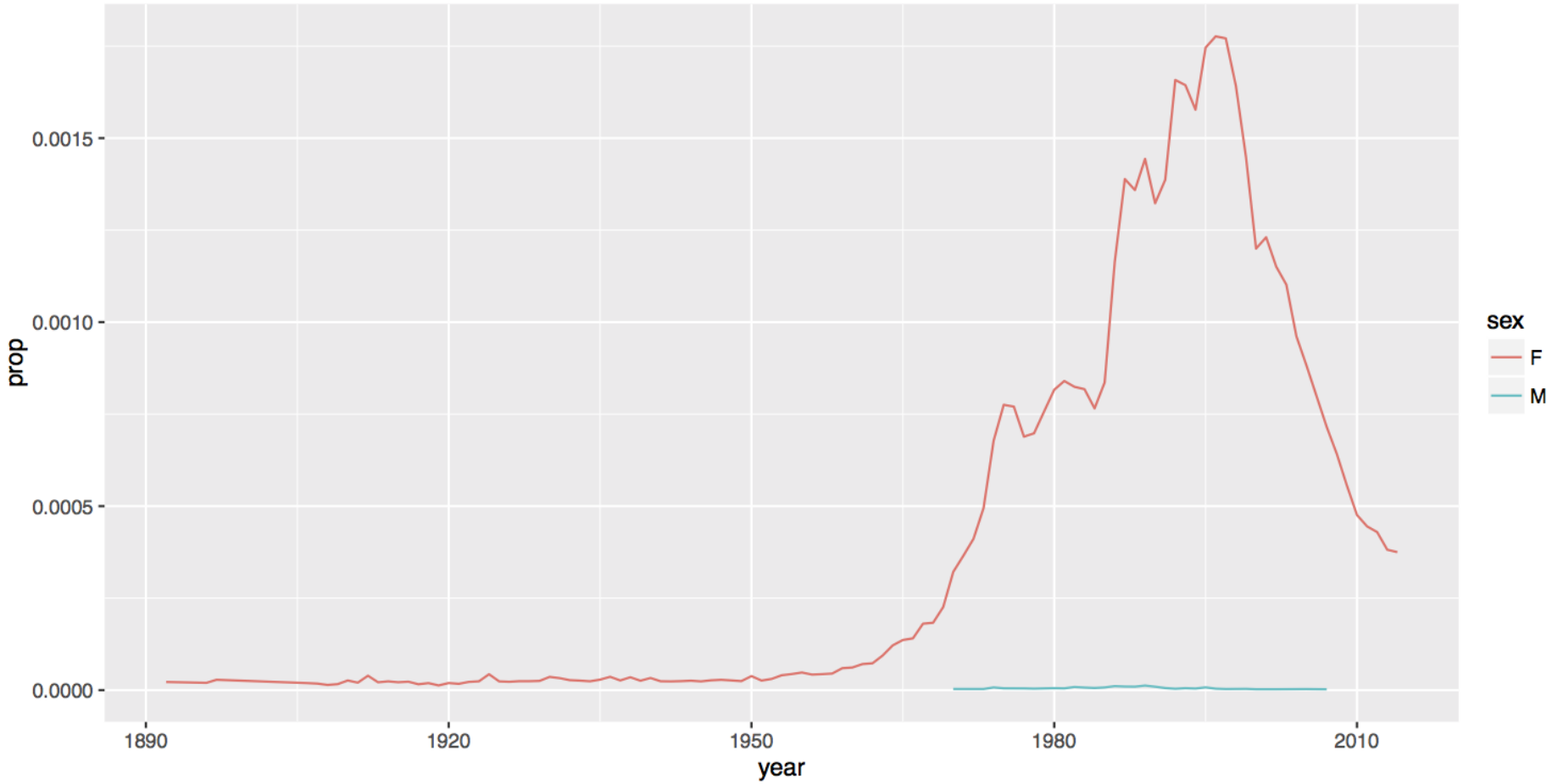
Hadley Wickham, `bit.ly/bigrdata4` and "Building precursors to data science" (CHANCE, 2015, `https://nhorton.people.amherst.edu/precursors`)

Possible answer to earlier question using tidyverse

```
> HELPmiss %>%
    group_by(sex) %>%
    summarise(meanval = mean(drugrisk, na.rm=TRUE))

     sex  meanval
  <fctr>    <dbl>
1   male 1.904762
2 female 1.756757
```

More ‘Variety’ of data (Alexander Hamilton)

More 'Variety' of data (Schuyler sisters and day 1 activity)



Prevalence of `Angelica` as a babyname over time (by gender)

More 'Variety' of data

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Number 302 JUNE, 1963 *Volume 58*

INFERENCE IN AN AUTHORSHIP PROBLEM[1,2]

A comparative study of discrimination methods applied to the authorship of the disputed *Federalist* papers

FREDERICK MOSTELLER
Harvard University
and
Center for Advanced Study in the Behavioral Sciences
AND
DAVID L. WALLACE
University of Chicago

Bigger (medium?) data (more 'Volume')

- use SQL (structured query language) to access databases within dplyr
- NYC Taxis (1.1 billion rides)
- Climate change data
- airline delays

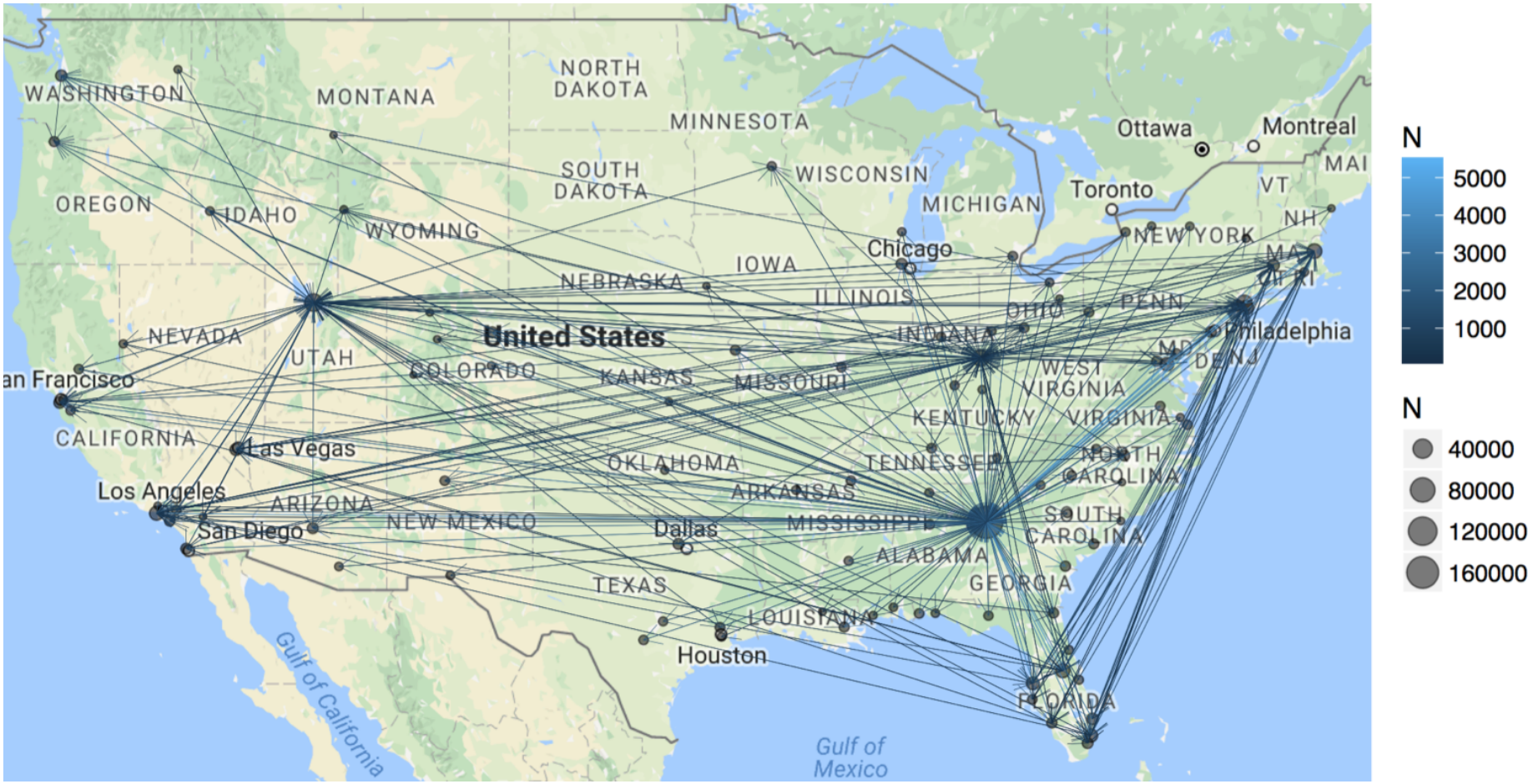
Airline delays: better traveling via data

- Collected by the Bureau of Transportation Statistics since 1987
- All commercial flights within the US (more than 180 million records)
- Easily motivated: have your students been stuck in an airport because a flight was delayed or cancelled (and wondered if they could have predicted it?) (Wickham, JCGS, 2011)
- Details at `http://stat-computing.org/dataexpo/2009`

Data wrangling idioms

- tidy data: each variable in its own column and observation in its row
- focus on Boston area flights (`dplyr::filter()`)
- focus on desired variables (`dplyr::select()`)
- correct odd variable names (`dplyr::rename()`)

Network science and bigger data



More big ideas

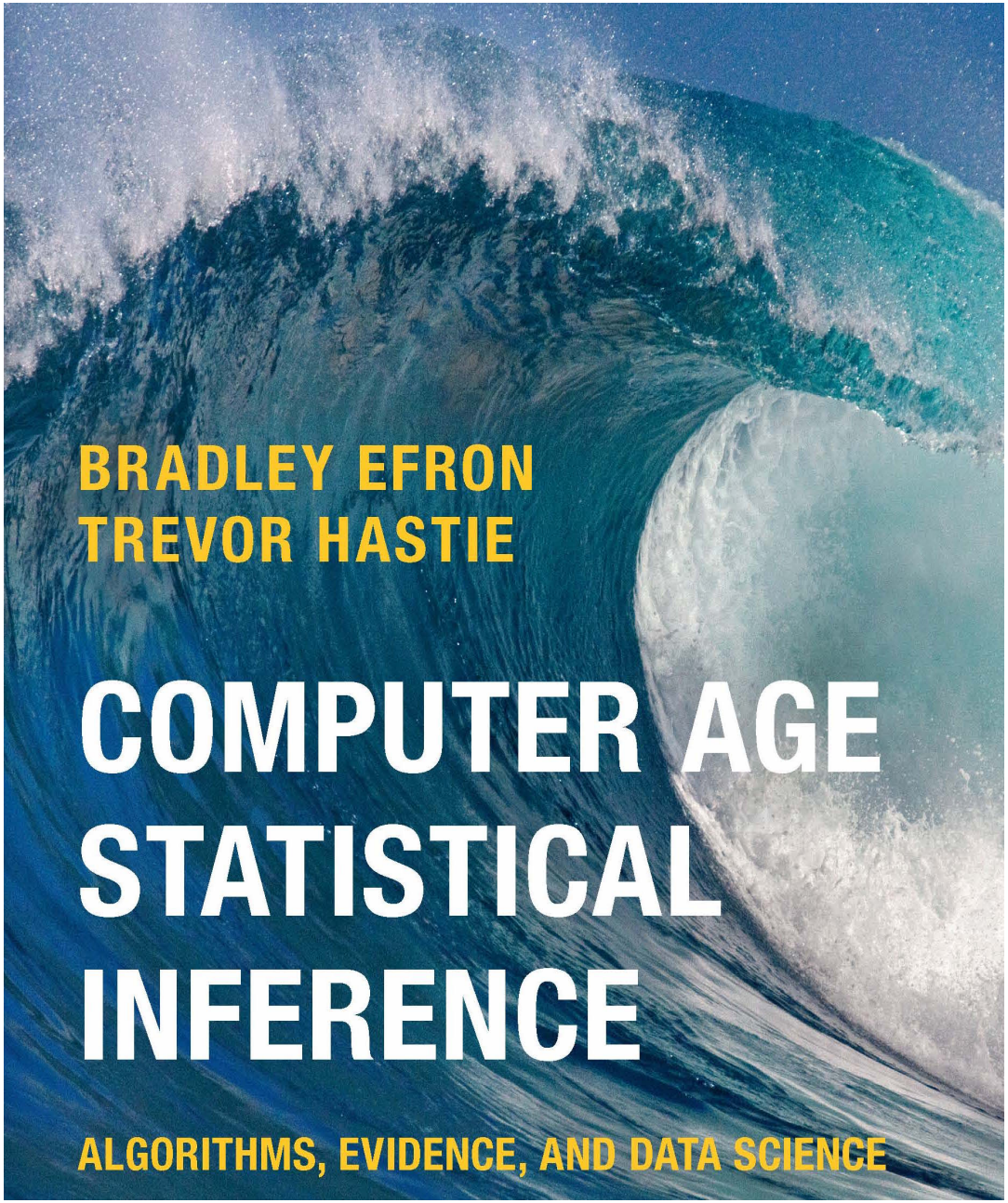
1. Increasing importance of data science and computation
2. Real applications and data
3. Statistical methods and foundations

Big idea #3: Statistical methods and foundations

ASA undergrad guidelines for statistics programs:

- Students require exposure to and practice with a variety of predictive and explanatory models in addition to methods for model building and assessment.
- They must be able to understand issues of design, confounding, and bias.
- They need to know how to apply their knowledge of theoretical foundations to the sound analysis of data.

Computer age statistical inference



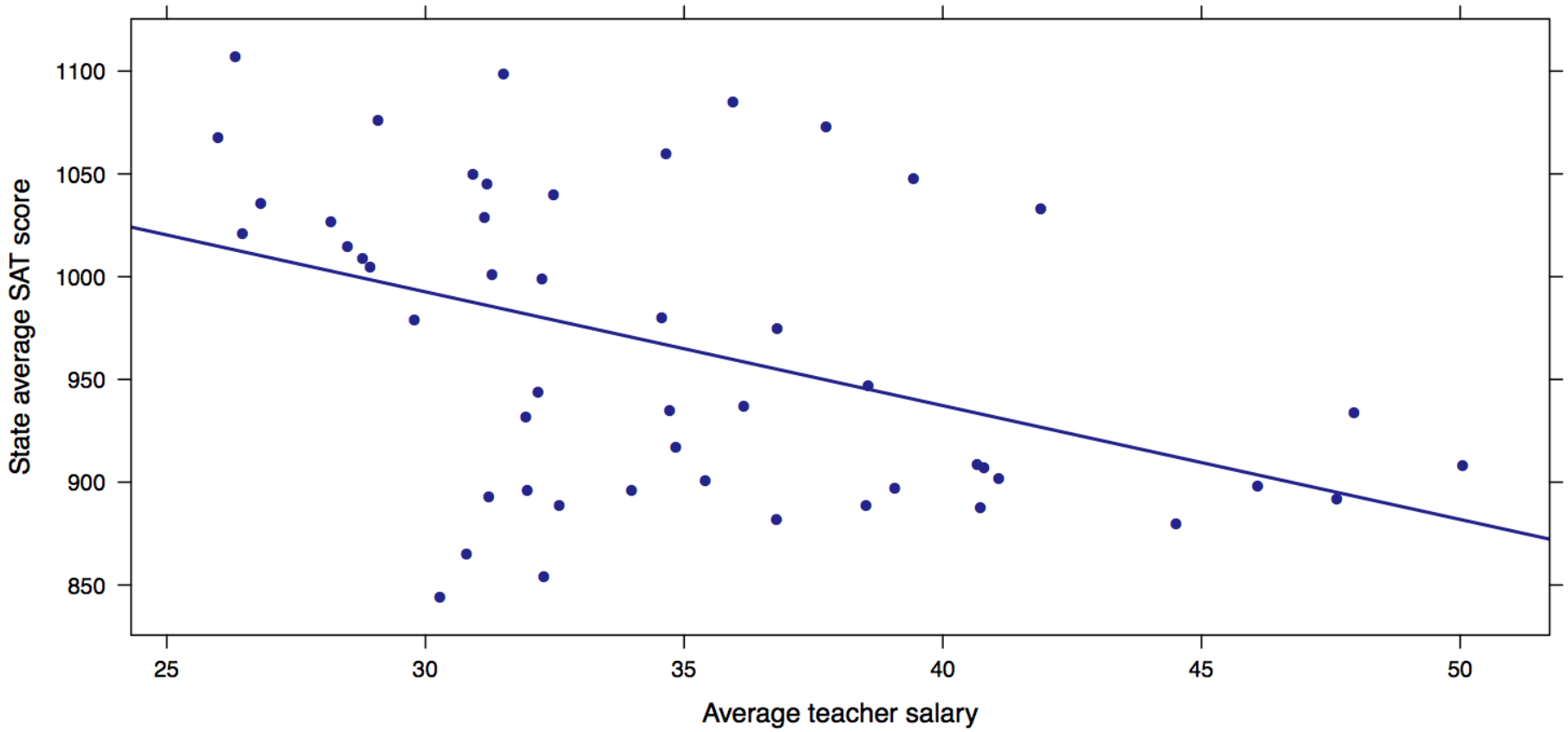
Macalester College DCF course

COMP 110 - DATA AND COMPUTING FUNDAMENTALS

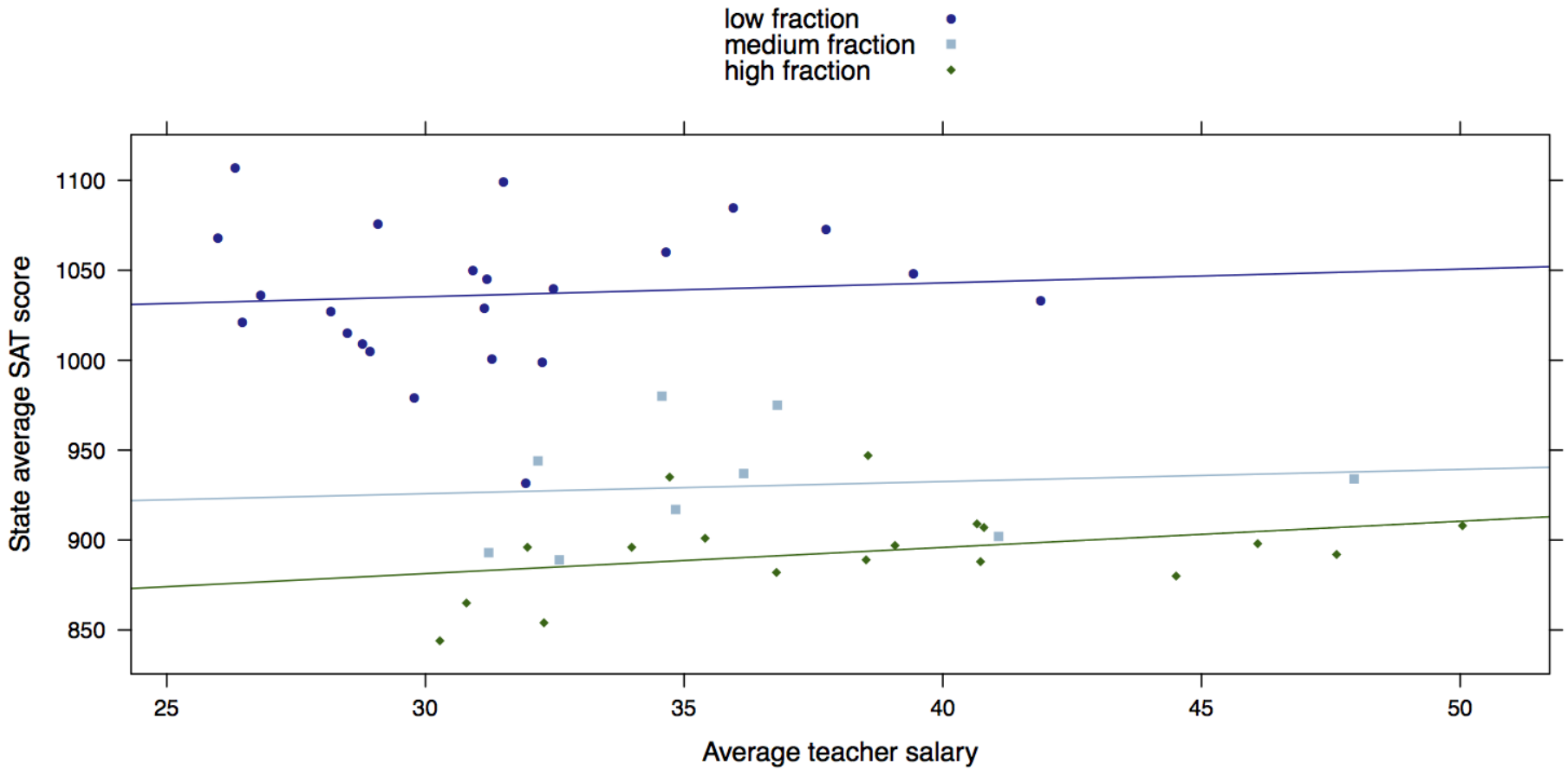
An introduction to the handling, analysis, and interpretation of "big data," the massive datasets now routinely collected in science, commerce, and government. The course is designed to be accessible to all students, regardless of background. Students will become proficient with R, a leading data and statistics computer environment. R skills are in high demand in research, commercial, NGO, and government areas. The course aligns with techniques being used in several courses in the sciences, statistics, and mathematics. (1 Credits)

Includes predictive modeling, databases, dynamic viz...

SAT scores and teacher salaries (state data from 2010)



Multivariate thinking



AP Statistics Vocabulary

Both Sides

confounding

when the levels of one factor are associated with the levels of another factor so their effects cannot be separated

Empirical problem-solving

Setting: Let A, B, and C be independent random variables each independently distributed uniformly in the interval [0,1].

Question: What is the probability that the roots of the quadratic equation given by $Ax^2 + Bx + C = 0$ are real?

Source: Rice Mathematical Statistics and Data Analysis third edition exercise 3.11 (also in first and second editions)

Empirical problem-solving

Setting: Let A, B, and C be independent random variables each independently distributed uniformly in the interval [0,1].

Question: What is the probability that the roots of the quadratic equation given by $Ax^2 + Bx + C = 0$ are real?

Source: Rice Mathematical Statistics and Data Analysis third edition exercise 3.11 (also in first and second editions)

Note: I continue to use this excellent book for my probability and statistical foundations courses

Analytic problem-solving

The distribution of $Y = B^2$ is given by:

$$f(y) = \begin{cases} \frac{1}{2\sqrt{y}} & \text{if } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

The distribution of $W = 4AC$ is given by:

$$f(w) = \begin{cases} -\log(w/4)/4 & \text{if } 0 \leq w \leq 4 \\ 0 & \text{otherwise} \end{cases}$$

Since Y and W are independent, the joint distribution is given by:

$$f(y, w) = \begin{cases} \frac{-\log(w/4)}{8\sqrt{y}} & \text{if } 0 \leq y \leq 1 \text{ and } 0 \leq w \leq 4 \\ 0 & \text{otherwise} \end{cases}$$

The discriminant $B^2 - 4AC$ is non-negative when $Y > W$.

$$\begin{aligned} P(Y > W) &= \int_0^1 \int_0^y f(y, w)\, dw\, dy \\ &= \int_0^1 \int_0^y \frac{-\log(w/4)}{8\sqrt{y}}\, dw\, dy \\ &= \int_0^1 \frac{\sqrt{y}(-\log(y) + 1 + \log(4))}{8}\, dy \\ &= \frac{5 + \log(64)}{36} \approx 0.254413. \end{aligned}$$

Empirical problem-solving

- Answer in the back of the book: 1/9

Empirical problem-solving

- Answer in the back of the book: 1/9
- Straightforward to code in R (or other environments):

```
> numsim <- 1000000
> u1 <- runif(numsim)
> u2 <- runif(numsim)
> u3 <- runif(numsim)
> discrim <- u2^2 - 4*u1*u3
> realroot <- discrim >= 0
> table(realroot)/numsim
FALSE   TRUE
0.7455  0.2545
```

IMPLICATION:

Last of the big ideas

1. Increasing importance of data science and computation
2. Real applications and data
3. Statistical methods and foundations
4. Communication and knowledge transference

Big idea #4: Communication and knowledge transference

ASA undergrad guidelines for statistics programs:

- Students need to be able to communicate complex statistical methods in basic terms to managers and other audiences and to visualize results in an accessible manner.
- They must have a clear understanding of ethical standards.
- Programs should provide multiple opportunities to practice and refine these statistical practice skills and use of analysis cycle.

R Markdown and reproducible analysis

The ability to express statistical computations is an essential skill (Nolan and Temple Lang, TAS 2010)

- R Markdown used as first workflow for introductory statistics students at colleges and universities all over the country
- forms a 'necessary but not sufficient' component of reproducible research
- tightly integrated into RStudio (designed for experts, useful for newbies)

R Markdown and reproducible analysis

example.Rmd

```
1  R Markdown
2  ==========
3
4  This is an R Markdown document. You can embed
   an R code chunk like this:
5
6  ```{r}
7  summary(cars)
8  ```
9
10 You can also embed plots, for example:
11
12 ```{r exPlot, fig.width=4, fig.height=4}
13 plot(cars)
14 ```
15
16
```

RStudio: Preview HTML

Preview: ~/Desktop/example.html | Log | Save As

R Markdown

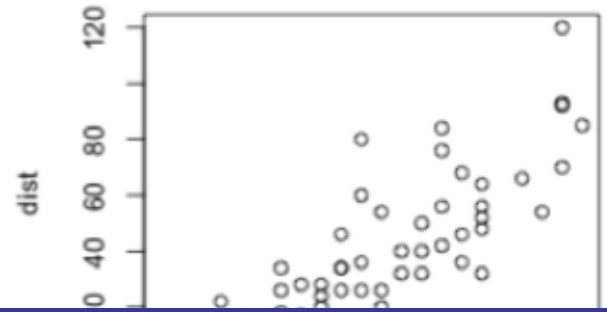
This is an R Markdown document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2
##  1st Qu.:12.0   1st Qu.: 26
##  Median :15.0   Median : 36
##  Mean   :15.4   Mean   : 43
##  3rd Qu.:19.0   3rd Qu.: 56
##  Max.   :25.0   Max.   :120
```

You can also embed plots, for example:

```
plot(cars)
```



Dynamic visualization and Shiny

https://r.amherst.edu/apps/nhorton/sat/

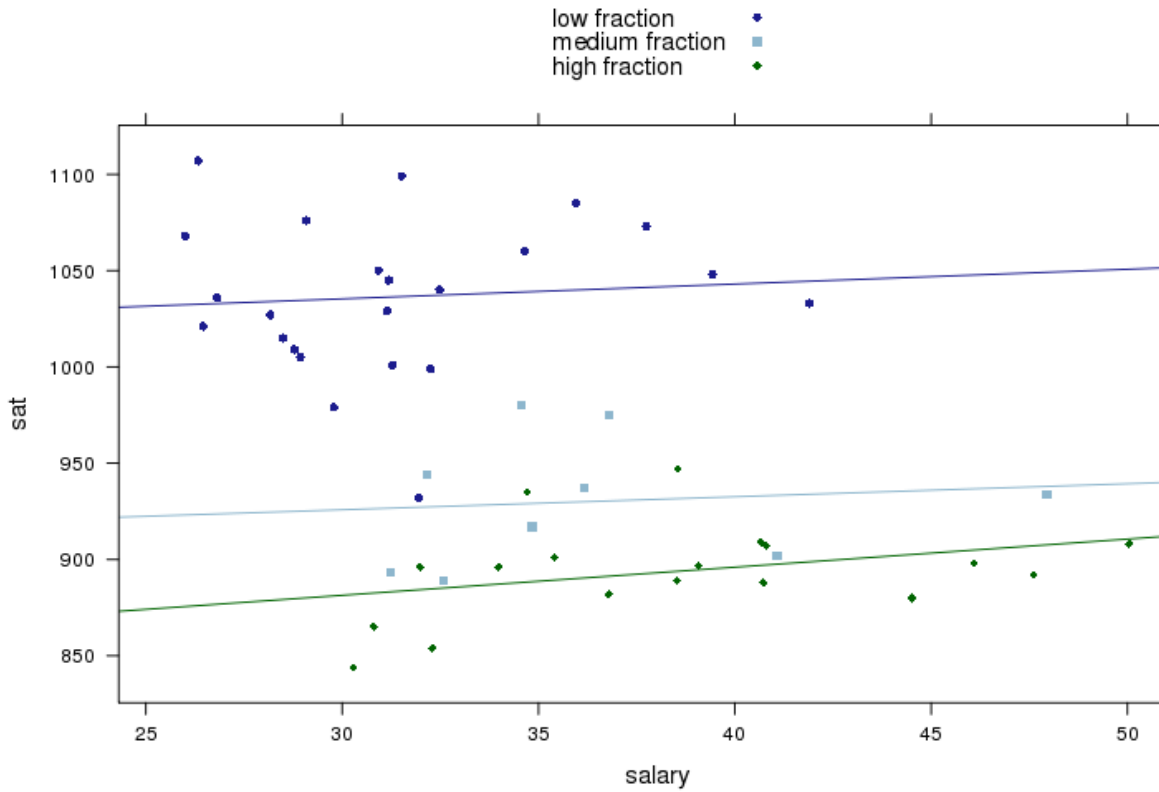
SAT scores and teacher salaries

Stratify by percent taking SAT?

Yes

Yes

No



Dynamic visualization and Shiny

server.R | ui.R

```
shinyServer(function(input, output) {
  output$distPlot <- renderPlot({

    # mosaic setup
    require(mosaic); require(mosaicData)
    trellis.par.set(theme=theme.mosaic())

    # create new variable
    SAT = mutate(SAT, fracgrp = cut(frac,
      breaks=c(0, 22, 49, 81),
      labels=c("low fraction", "medium fraction", "high fraction")))

    # generate the desired plot
    if (input$stratify == "No") {
      xyplot(sat ~ salary, type=c("p", "r"), data=SAT)
    } else {
      xyplot(sat ~ salary, groups=fracgrp, auto.key=TRUE,
        type=c("p", "r"), data=SAT)
    }

    })
  }
)
```

Dynamic visualization and Shiny

server.R | ui.R

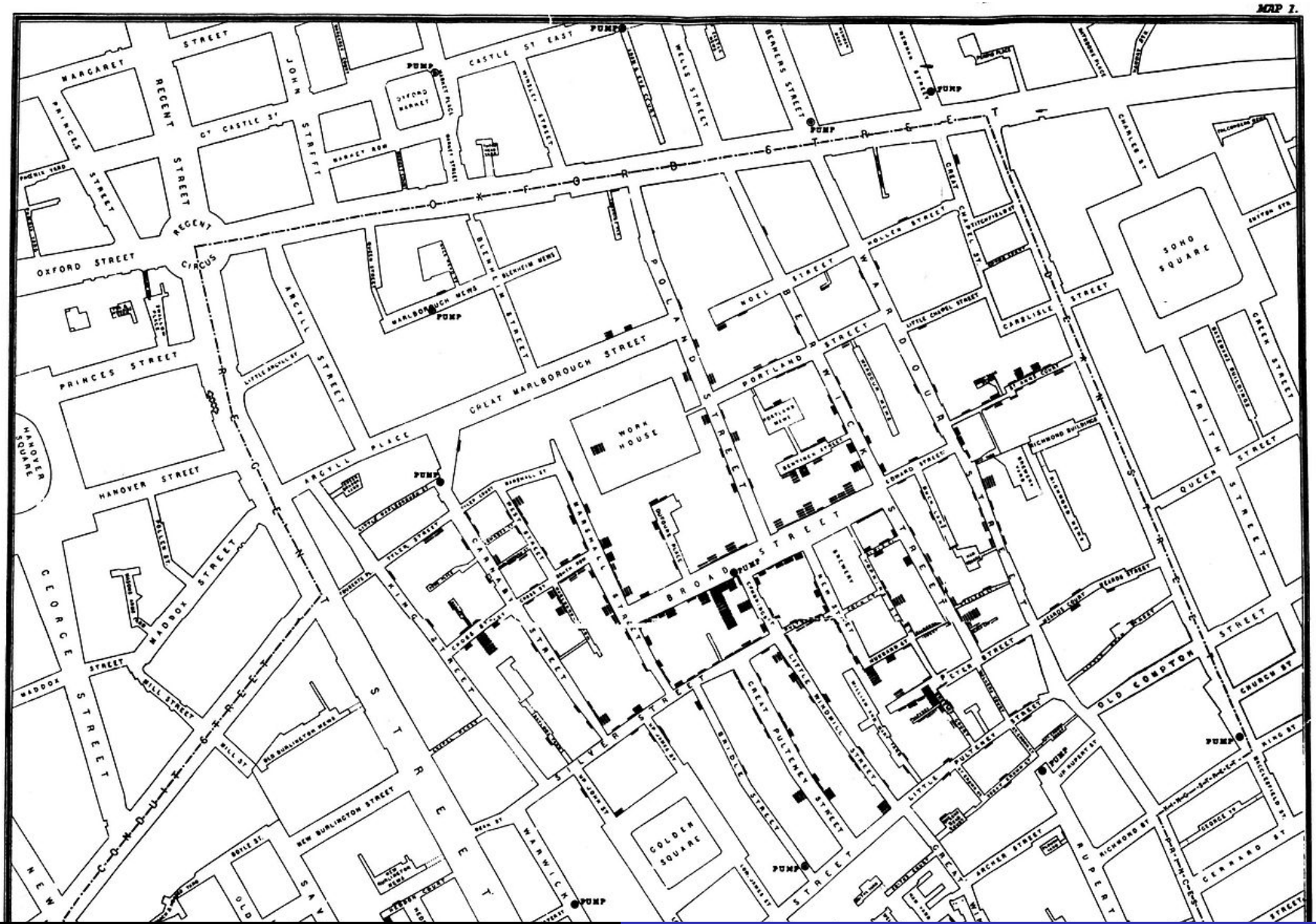
```
library(shiny)

shinyUI(fluidPage(
  # Application title
  titlePanel("SAT scores and teacher salaries"),

  sidebarLayout(
    sidebarPanel(
      selectInput("stratify", "Stratify by percent taking SAT?",
                  choices = c("Yes", "No"), selected="No")),
    # Show a plot of the generated distribution
    mainPanel(plotOutput("distPlot"))
  )
))
```

Interactive maps

```{r}
library(leaflet)
m <- leaflet() %>%
  addTiles() %>%  # Add default OpenStreetMap map tiles
  addMarkers(lng=-71.1191, lat=42.3731,
    popup="The birthplace of Harvard Stat")
m
```

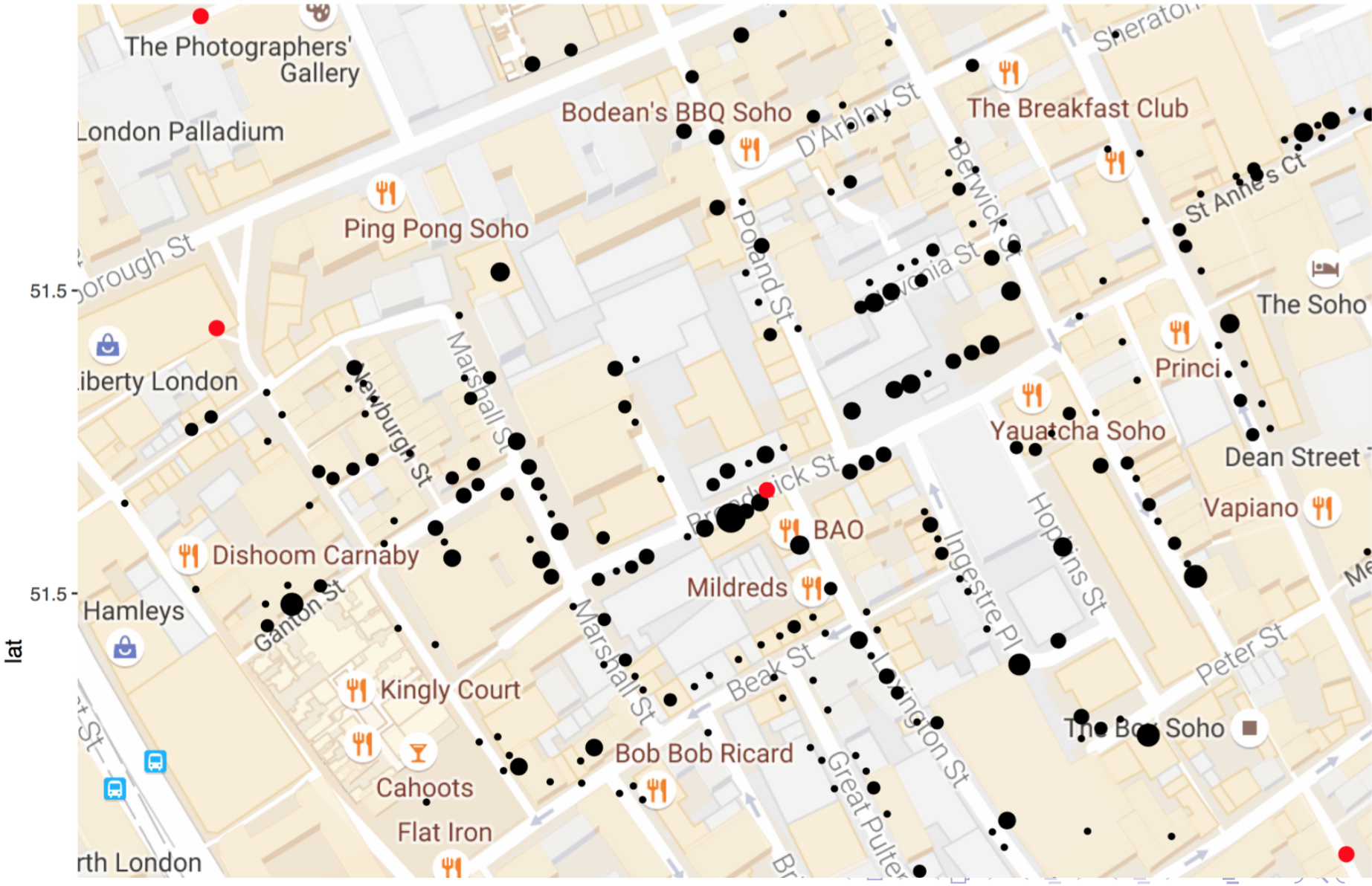
Dynamic visualization

Snow and the cholera epidemic in London...



Dynamic visualization

Snow and the cholera epidemic in London...



Revision control systems

> *Version control is the only reasonable way to keep track of changes in code, manuscripts, presentations, and data analysis projects.*

Karl Broman,
`http://kbroman.org/github_tutorial/pages/why.html`

Revision control systems

> *Version control is the only reasonable way to keep track of changes in code, manuscripts, presentations, and data analysis projects.*

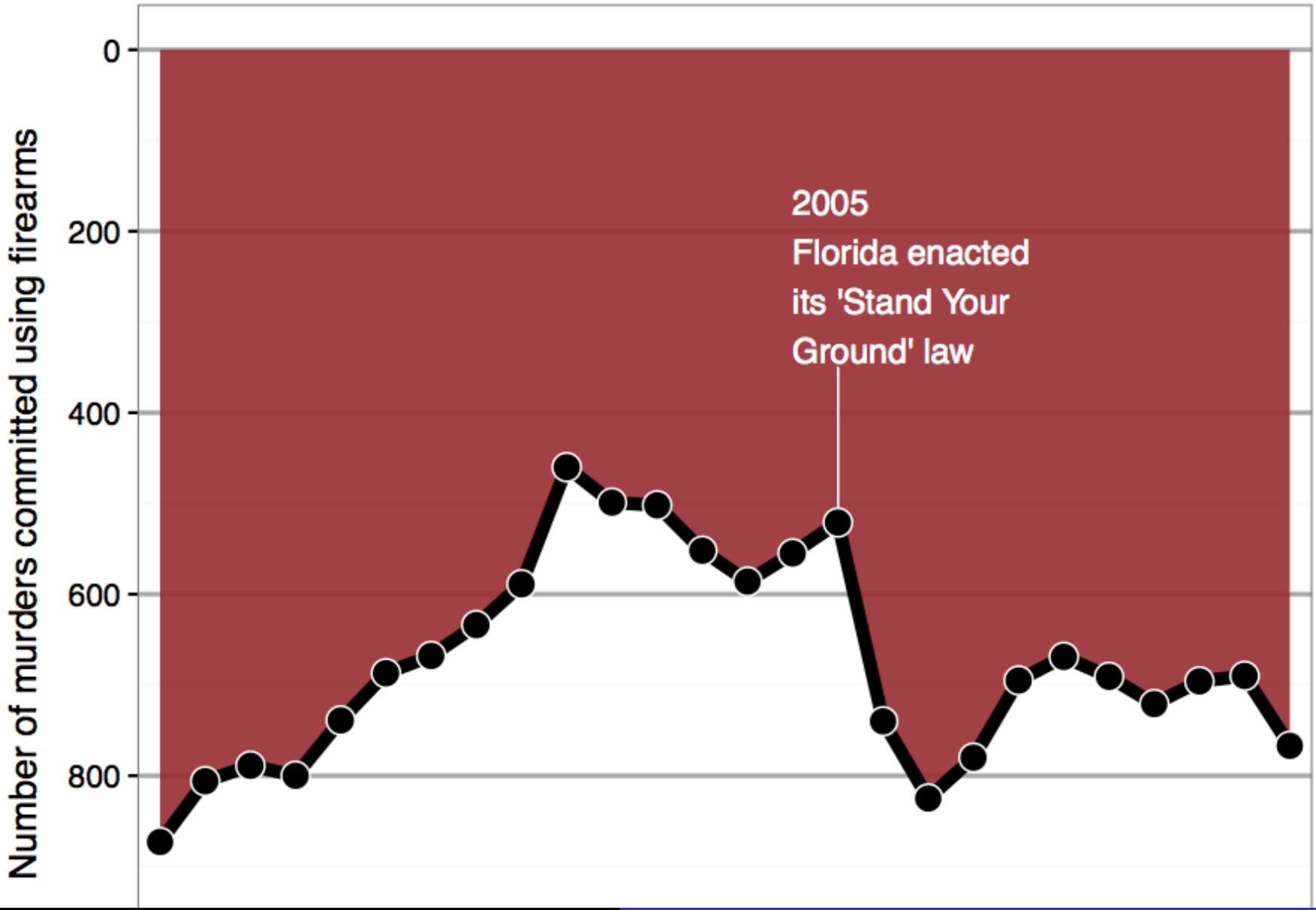
Karl Broman,
`http://kbroman.org/github_tutorial/pages/why.html`

> *If you need to collaborate on data analysis or code development, then all involved should use Git.*

Jenny Bryan, `http://happygitwithr.com`

Ethics

If not now, when?



Challenges and opportunities

Challenges and opportunities

Read the Introducing AP Computer Science Principles video transcript

Computer Science: The New Literacy

Whether it's 3-D animation, engineering, music, app development, medicine, visual design, robotics, or political analysis, computer science is the engine that powers the technology, productivity, and innovation that drive the world. Computer science experience has become an imperative for today's students and the workforce of tomorrow.

The AP Program designed AP Computer Science Principles with the goal of creating leaders in computer science fields and attracting and engaging those who are traditionally underrepresented with essential computing tools and multidisciplinary opportunities.

Big Idea 3: Data and Information

Data and information facilitate the creation of knowledge. Computing enables and empowers new methods of information processing, driving monumental change across many disciplines — from art to business to science. Managing and interpreting an overwhelming amount of raw data is part of the foundation of our information society and economy. People use computers and computation to translate, process, and visualize raw data and to create information. Computation and computer science facilitate and enable new understanding of data and information that contributes knowledge to the world. Students in this course work with data using a variety of computational tools and techniques to better understand the many ways in which data is transformed into information and knowledge.

Challenges and opportunities

Enduring Understandings (Students will understand that …)	**Learning Objectives** (Students will be able to …)
EU 3.1 People use computer programs to process information to gain insight and knowledge.	**LO 3.1.1** Find patterns and test hypotheses about digitally processed information to gain insight and knowledge. [P4]

Challenges and opportunities

LO 3.1.3 Explain the insight and knowledge gained from digitally processed data by using appropriate visualizations, notations, and precise language. [P5]	**EK 3.1.3A** Visualization tools and software can communicate information about data.
	EK 3.1.3B Tables, diagrams, and textual displays can be used in communicating insight and knowledge gained from data.
	EK 3.1.3C Summaries of data analyzed computationally can be effective in communicating insight and knowledge gained from digitally represented information.
	EK 3.1.3D Transforming information can be effective in communicating knowledge gained from data.
	EK 3.1.3E Interactivity with data is an aspect of communicating.

Challenges and opportunities

EU 3.2 Computing facilitates exploration and the discovery of connections in information.

LO 3.2.1 Extract information from data to discover and explain connections or trends. [P1]

Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report 2016

revised GAISE College Report (2016)

1. Teach statistical thinking.
   - Teach statistics as an investigative process of problem-solving and decision-making.
   - Give students experience with multivariable thinking.
2. Focus on conceptual understanding.
3. Integrate real data with a context and purpose.
4. Foster active learning.
5. Use technology to explore concepts and analyze data.
6. Use assessments to improve and evaluate student learning.

Closing thoughts

> *Curriculum unavoidably involves decisions about scarce resources, so curricular innovation cannot escape being political, and of course "all politics is local" (ONeill and Hymel, 1995).*
>
> *Curriculum is political for economic reasons because, averaged over the long term, faculty FTEs and course offerings are at best a zero-sum game. Thus changing curriculum, like moving a graveyard, depends on local conditions: Whose cherished ancestry is uprooted by the change?*

(Cobb 'Mere renovation is too little, too late: we need to rethink our undergraduate curriculum from the ground up' arXiv 2015)

Closing thoughts

- It's never been easier to extract meaning from data (improved tools)
- How do we ensure that statistics remains a vibrant choice for our students?

Big Ideas to help statistics students learn to 'think with data'

Nicholas J. Horton

Department of Mathematics and Statistics
Amherst College, Amherst, MA, USA

Pickard Lecture, September 29, 2016

nhorton@amherst.edu
http://nhorton.people.amherst.edu

Challenges and opportunities

NATURE | NEWS

Statisticians issue warning over misuse of *P* values

Policy statement aims to halt missteps in the quest for certainty.

Monya Baker

07 March 2016

PDF | Rights & Permissions

Misuse of the *P* value — a common test for judging the strength of scientific evidence — is contributing to the number of research findings that cannot be reproduced, the American Statistical Association (ASA) warns in a statement released today[1]. The group has taken the unusual step of issuing principles to guide use of the *P* value, which it says cannot determine whether a hypothesis is true or whether results are important.