Introduction to Data Science
SKKU University, Summer 2015

Prof. Jevin D. West
University of Washington
Lecture 12 – July 14, 2015

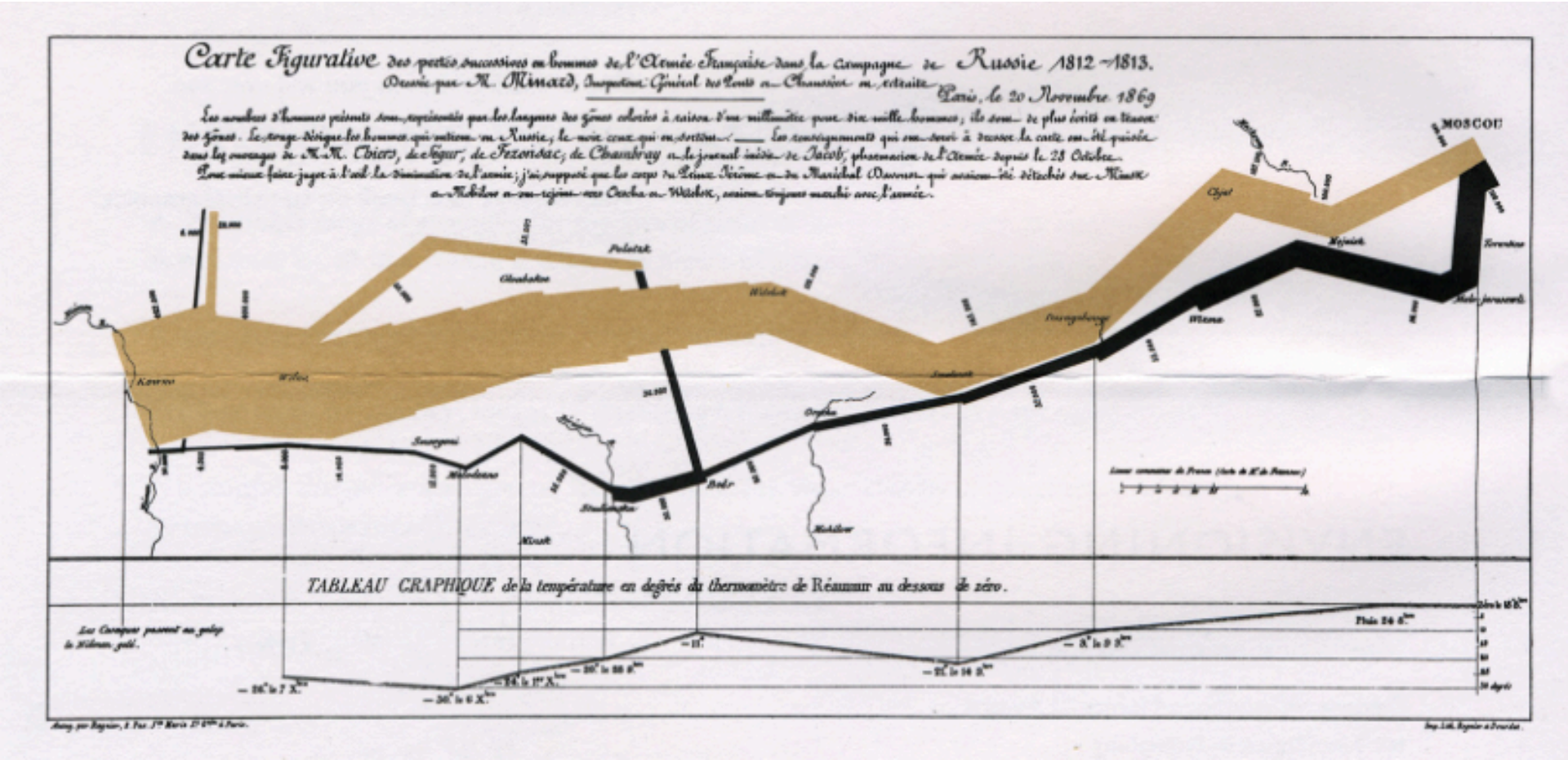
Agenda

- 9:30 – 11:30 Final Paper Prep
- 11:30 – 12:00 Questions

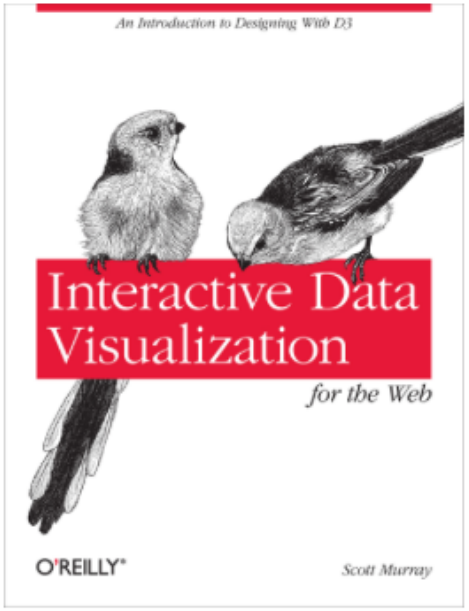
Week 2

- ~~Tuesday, July 14: Quiz #3~~
- **DUE** Wednesday, July 15 at the beginning of class: Final Paper
- **DUE** Thursday, July 16: Assignment #3
- **DUE** Friday, July 17 at the beginning of class: Final Presentation, make sure to turn in presentation on your group repository

Edward Tufte

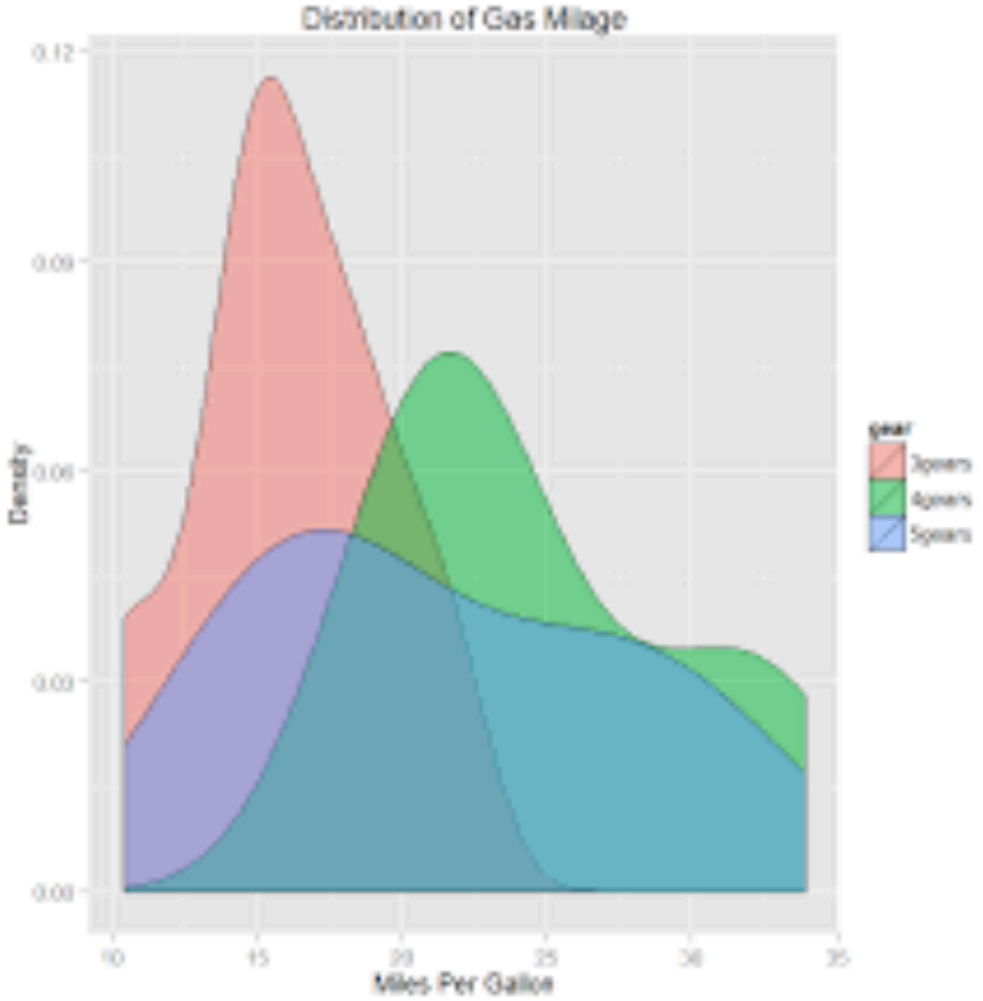


Data-Driven Documents

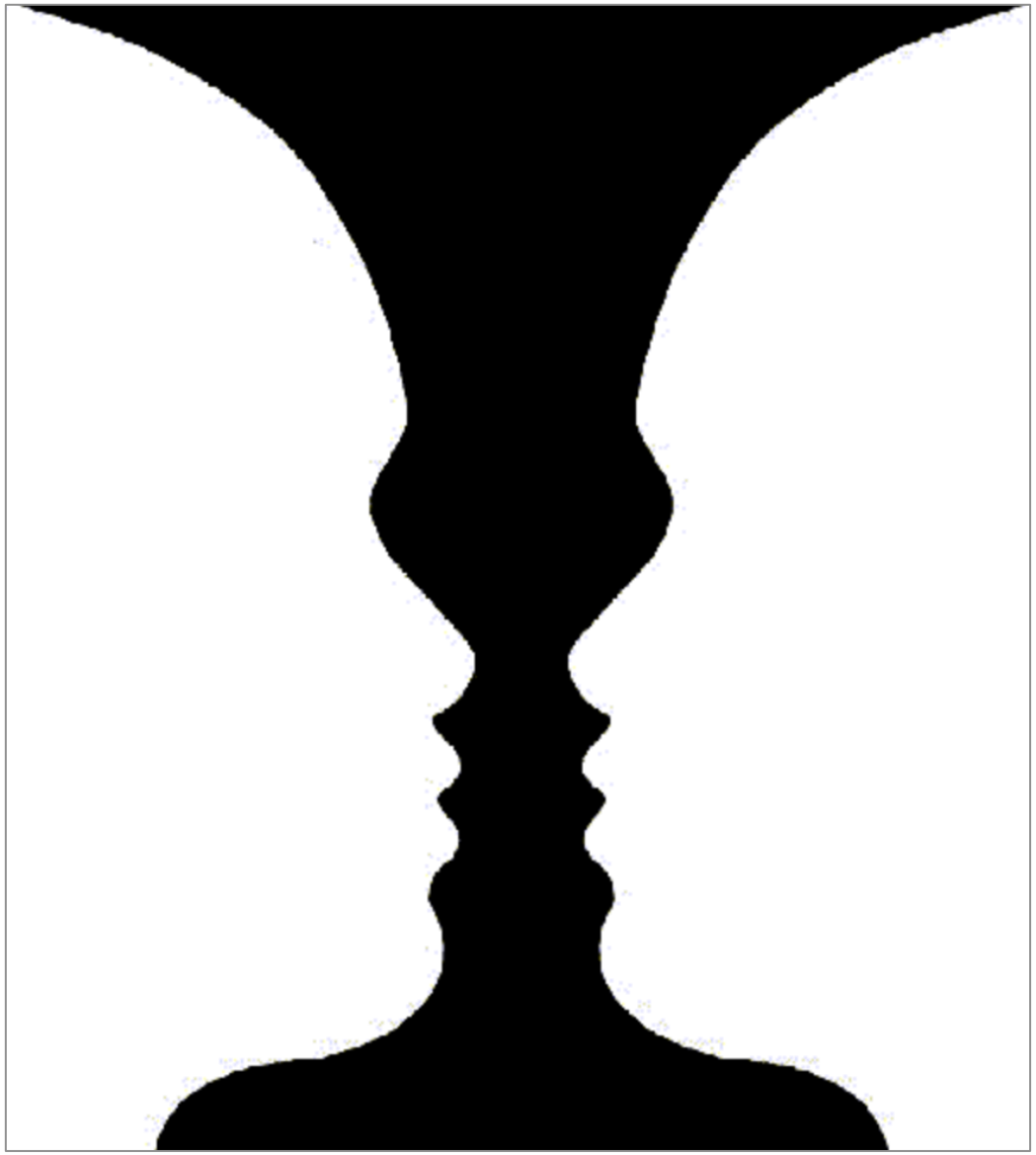


- Elijah Meeks just wrote a D3 book
- Scott Murray wrote a great (free) book on D3 from a designer's perspective
- D3 newsletter: Cool stuff every week
- Interactive Data Visualization: http://chimera.labs.oreilly.com/books/1230000000345

ggplot in R

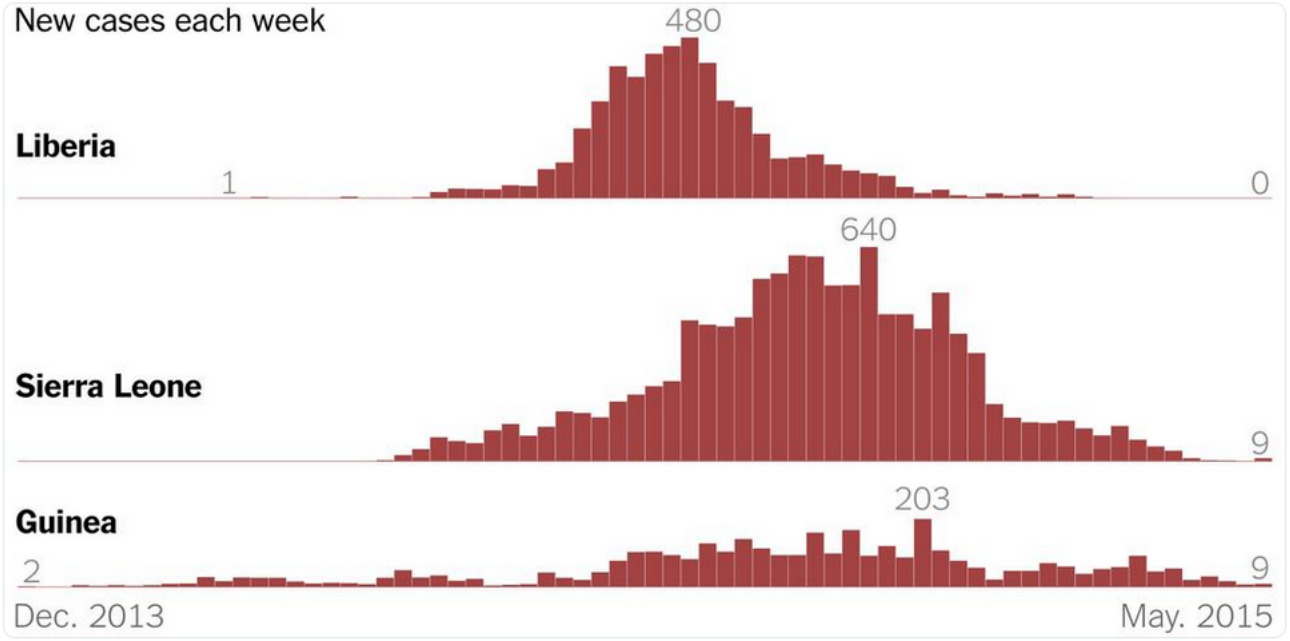

Distribution of Gas Milage
0.12
0.09
0.06
0.03
0.00
Density
10
15
20
25
30
35
Miles Per Gallon
gear
3gears
4gears
5gears

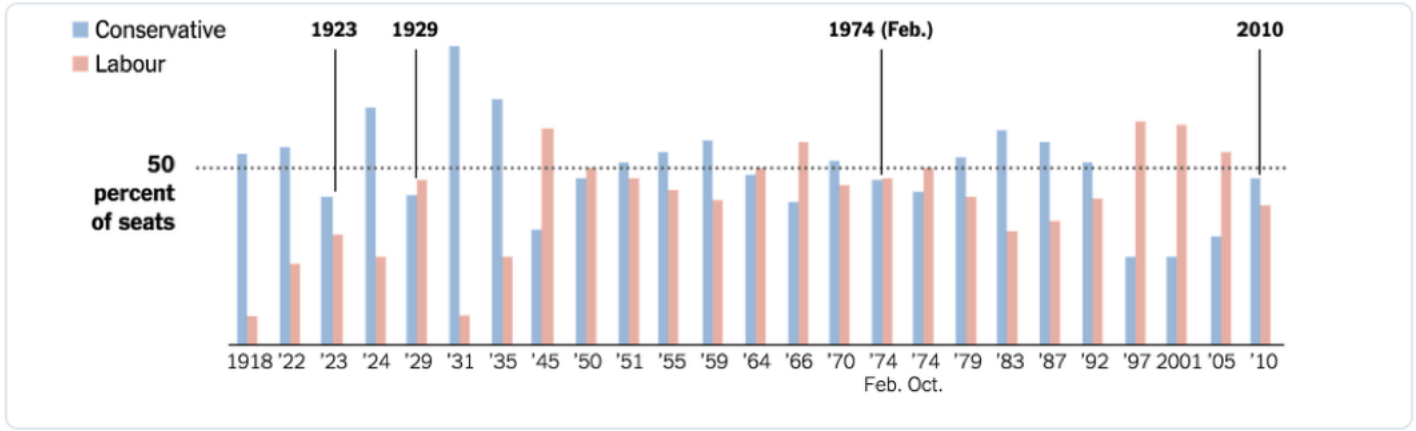
Visual Perception



Gridlines

- Gridlines should always be labeled.
- Corollary: One should not use grid lines on any axis that lacks a scale.
- If one does use grid lines, they should be ONLY be used to demarcate increments of the dependent variable.
- Grid lines should be subtle elements of the background, and not draw the eye from the foreground elements.

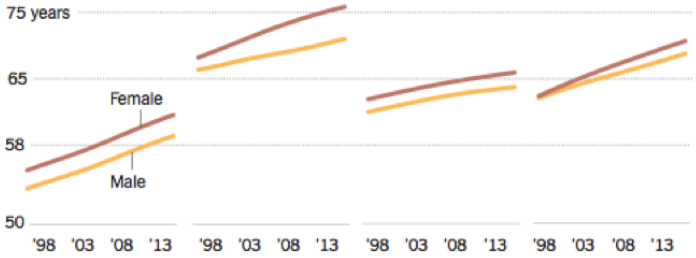


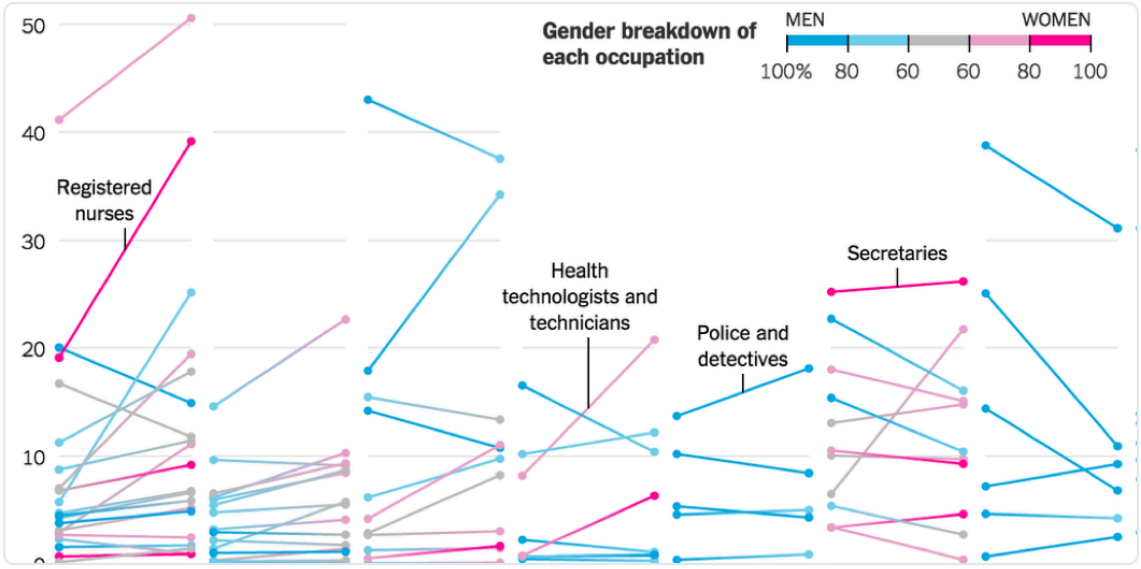


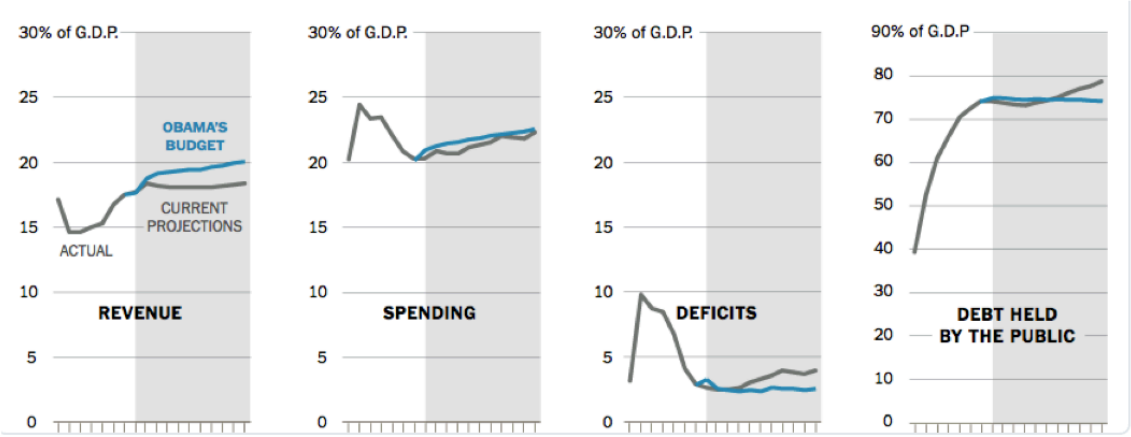
Acknowledgement: slides and discussions around gridlines with Carl Bergstrom

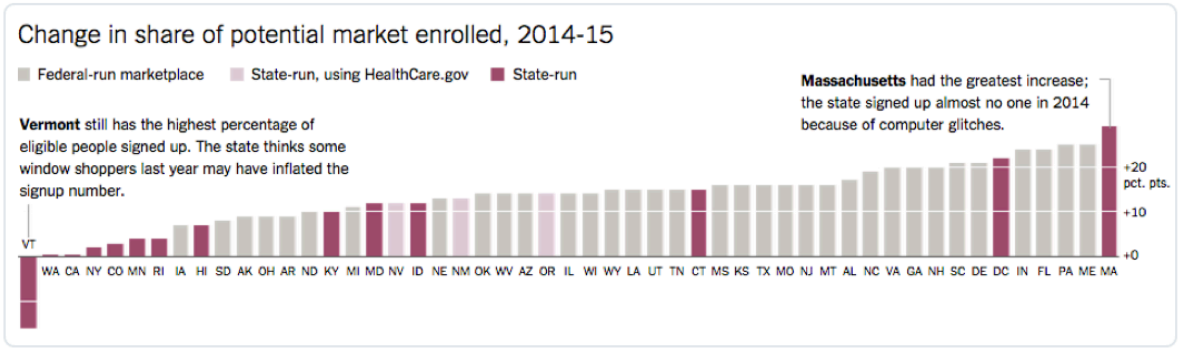
Life Expectancy

Worldwide, women are generally expected to live longer than men. This is true in Afghanistan as well. But life expectancy in Afghanistan is still well behind other countries in the region.



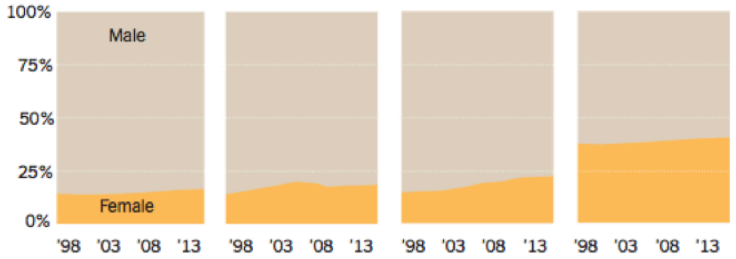


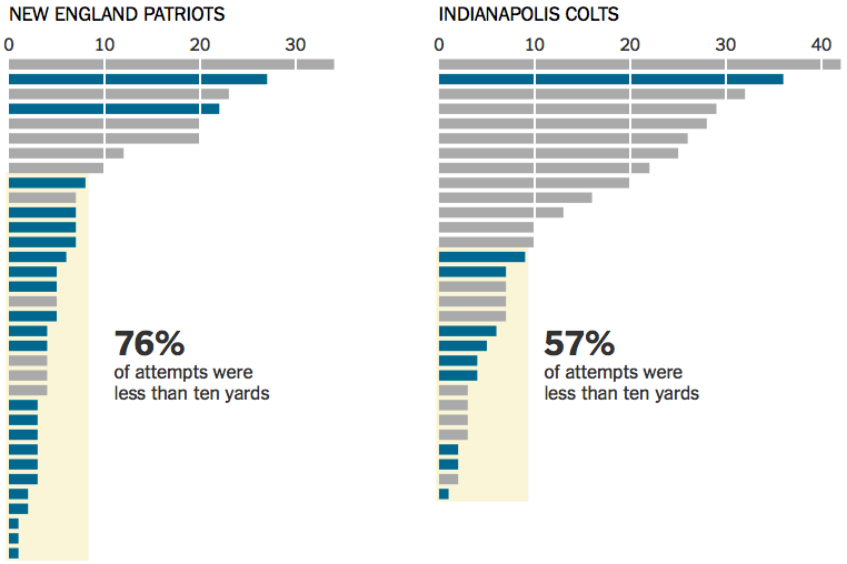


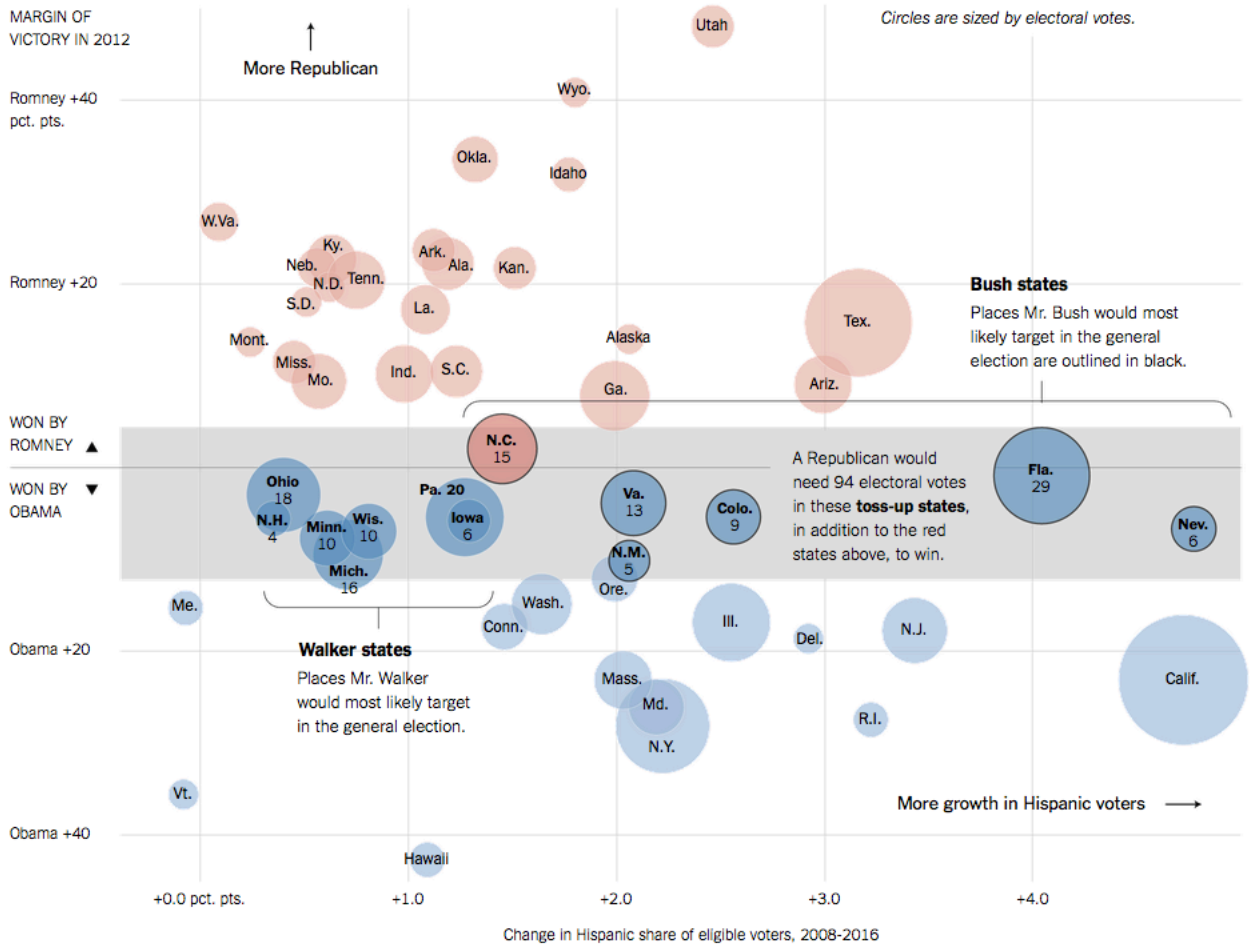
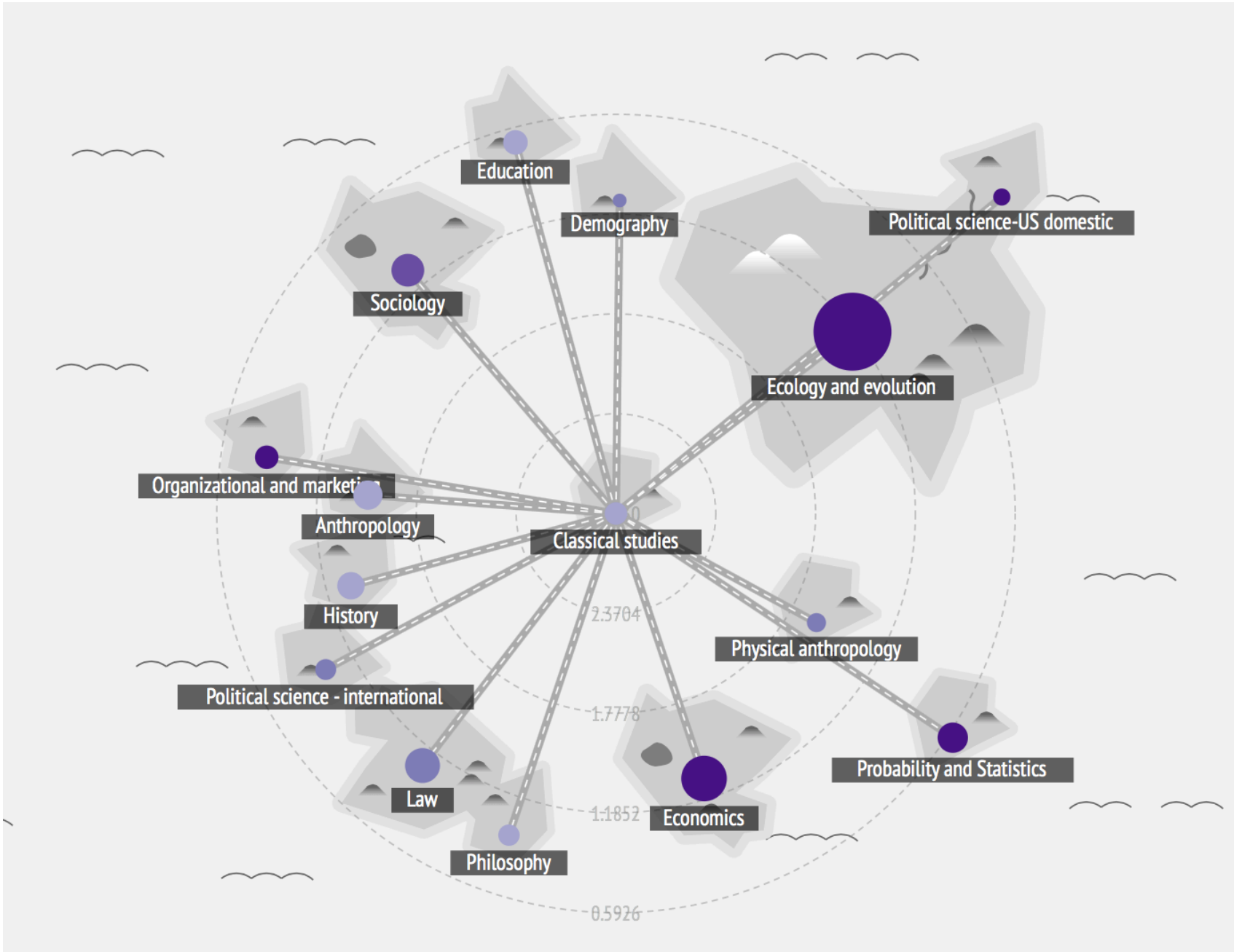


Labor Market Composition

Women have yet to find their footing in Afghanistan's labor market. Woman make up just 16 percent of the country's labor force, slightly less than Iran, at 18 percent, and far less than Bangladesh, at 40 percent.






MARGIN OF VICTORY IN 2012
More Republican
Romney +40 pct. pts.
Romney +20
WON BY ROMNEY ▲
WON BY OBAMA ▼
Obama +20
Obama +40
Circles are sized by electoral votes.
Utah
Wyo.
Okla.
Idaho
W.Va.
Ky.
Neb.
N.D.
Tenn.
Ark.
Ala.
Kan.
S.D.
La.
Mont.
Miss.
Mo.
Ind.
S.C.
Alaska
Ga.
Tex.
Ariz.
N.C. 15
Ohio 18
N.H. 4
Minn. 10
Wis. 10
Mich. 16
Pa. 20
Iowa 6
Va. 13
Colo. 9
N.M. 5
Ore.
Fla. 29
Nev. 6
Me.
Wash.
Conn.
Ill.
Del.
N.J.
Mass.
Md.
N.Y.
R.I.
Calif.
Vt.
Hawaii
Bush states
Places Mr. Bush would most likely target in the general election are outlined in black.
A Republican would need 94 electoral votes in these toss-up states, in addition to the red states above, to win.
Walker states
Places Mr. Walker would most likely target in the general election.
More growth in Hispanic voters
+0.0 pct. pts.
+1.0
+2.0
+3.0
+4.0
Change in Hispanic share of eligible voters, 2008-2016

The utility of metaphors in navigating large corpora