PINO TROGU

COLLEGE OF LIBERAL & CREATIVE ARTS

SAN FRANCISCO STATE UNIVERSITY

Survey questions (before)

DATA? TA-DA! TIPS FOR BETTER DATA VISUALIZATIONS

DATA SCIENCE WORKSHOP

WEDNESDAY, FEBRUARY 20, 2019, 4:00 PM TO 6:00 PM

ROOM 260, CREATIVE ARTS BUILDING

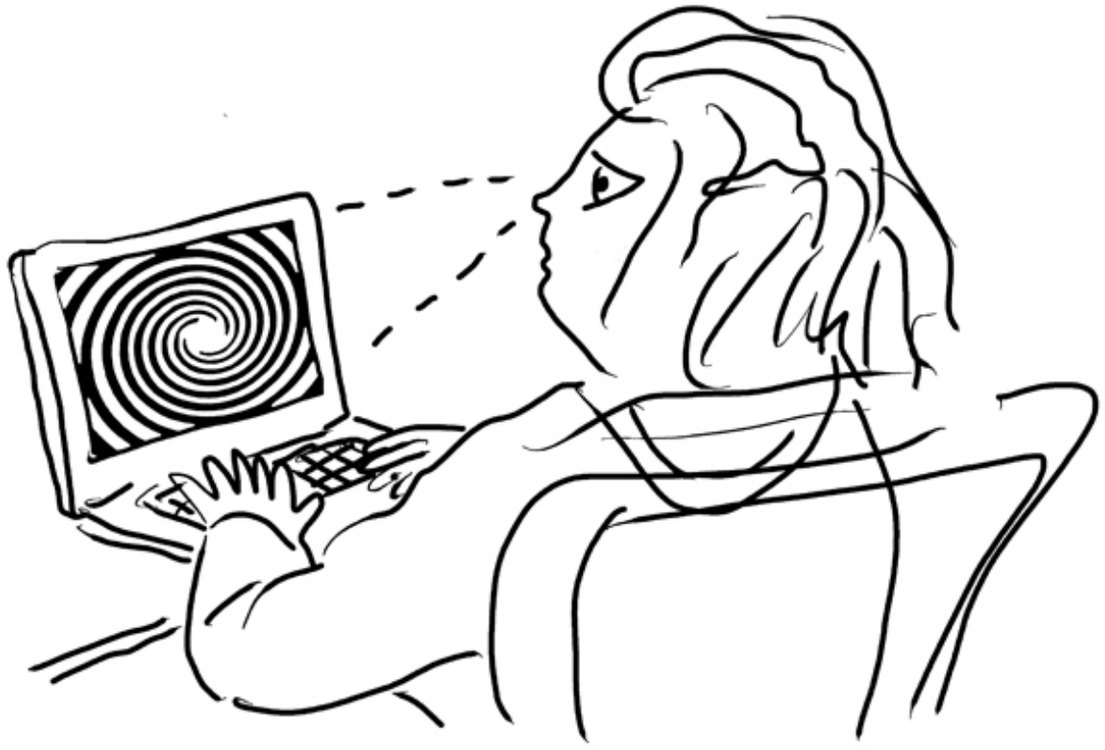
go to last slide

RULES AND EXAMPLES (1–16)

01 Use pencil and paper

Small Handbook of Information Design:
16 Principles for Better Data Visualizations

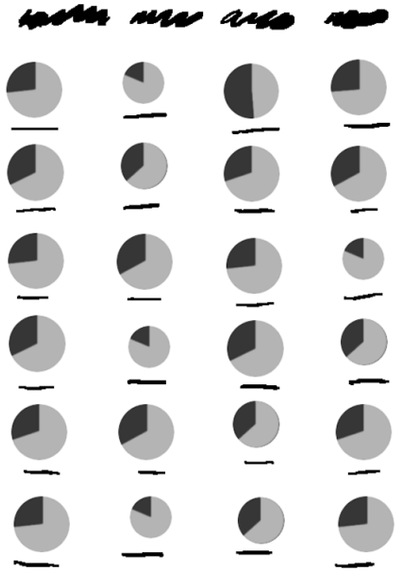
YES



NO

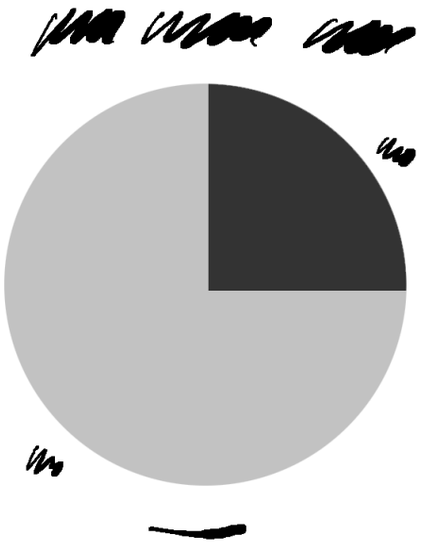
02 Content is first

year	bracket	fed	state	fica	property	sales	corporate	overall
1980	1	0.02306	0.0041	0.0535	0.04467	0.04923	0.02754	0.20209
1980	2	0.07927	0.01276	0.07368	0.03457	0.03149	0.03084	0.2626
1980	3	0.11445	0.01925	0.08513	0.02995	0.02506	0.03103	0.30488
1980	4	0.13588	0.02287	0.08294	0.02908	0.02172	0.03288	0.32537
1980	5	0.1597	0.02654	0.07491	0.03004	0.01917	0.03649	0.34685
1980	6	0.18233	0.02888	0.064	0.02951	0.01799	0.04049	0.3632
1980	7	0.19944	0.03128	0.05037	0.03149	0.01624	0.04944	0.37827
1980	12	0.2324	0.02985	0.03618	0.04056	0.01072	0.0812	0.43091
1980	15	0.29113	0.02653	0.01098	0.0557	0.00529	0.09999	0.4896
1981	1	0.02909	0.00493	0.05851	0.04387	0.04713	0.02476	0.20829
1981	2	0.0851	0.01346	0.07865	0.03441	0.03055	0.02758	0.26975
1981	3	0.12037	0.02009	0.09156	0.0296	0.0241	0.02747	0.31318
1981	4	0.14412	0.02362	0.0905	0.02932	0.02088	0.0296	0.33803
1981	5	0.16341	0.02646	0.08242	0.03047	0.01846	0.03337	0.3546
1981	6	0.18578	0.02785	0.07007	0.03065	0.01704	0.03879	0.3702
1981	7	0.21417	0.02886	0.05839	0.02967	0.01536	0.04134	0.38779
1981	12	0.24037	0.03015	0.03826	0.04005	0.01089	0.06632	0.42604
1981	15	0.2817	0.01948	0.01349	0.04287	0.00553	0.08502	0.44809



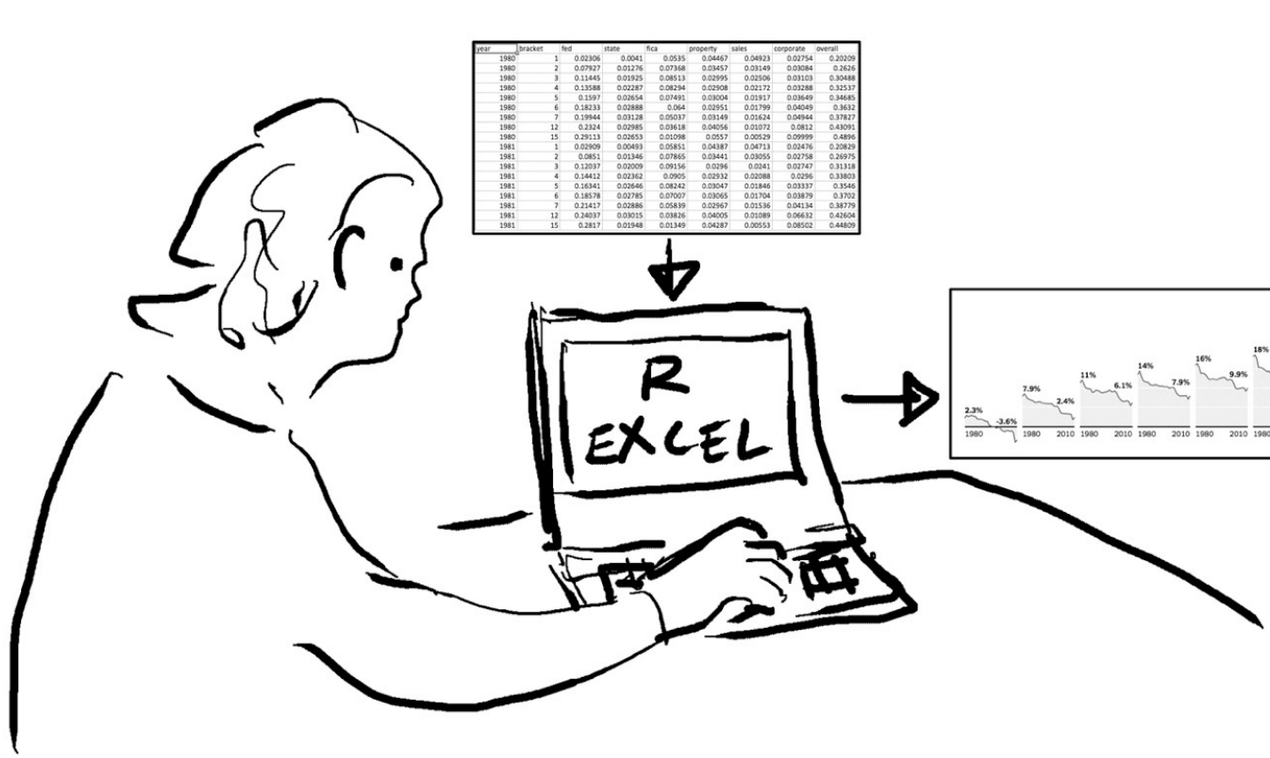
YES

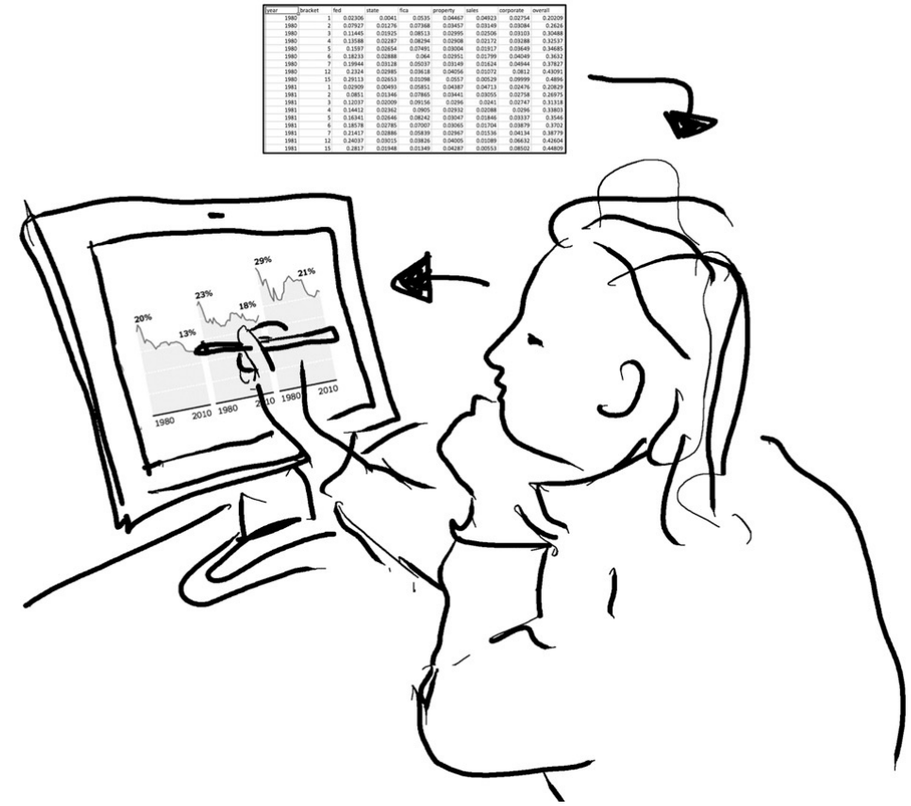
year	bracket	fed
1980	1	0.02306



NO

03 Do not draw graphs by hand

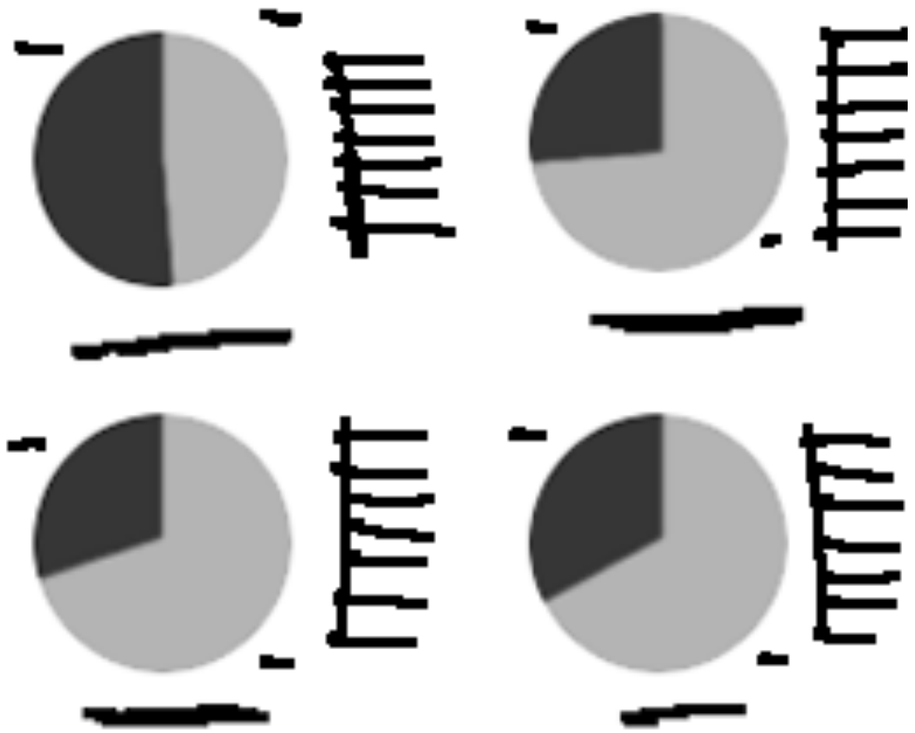




YES

NO

04 Do not enlarge numbers



55% BLA BLA

27% BLA BLA

30% BLA BLA

35% BLA BLA

YES

NO

05 Use words, not just images

http://www.nytimes.com/interactive/2012/09/17/science/driving-safety-in-fits-and-starts.html

the annotation layer

By HANNAH FAIRFIELD
SEPT. 17, 2012

05 Use words, not just images

Driving Safety, in Fits and Starts

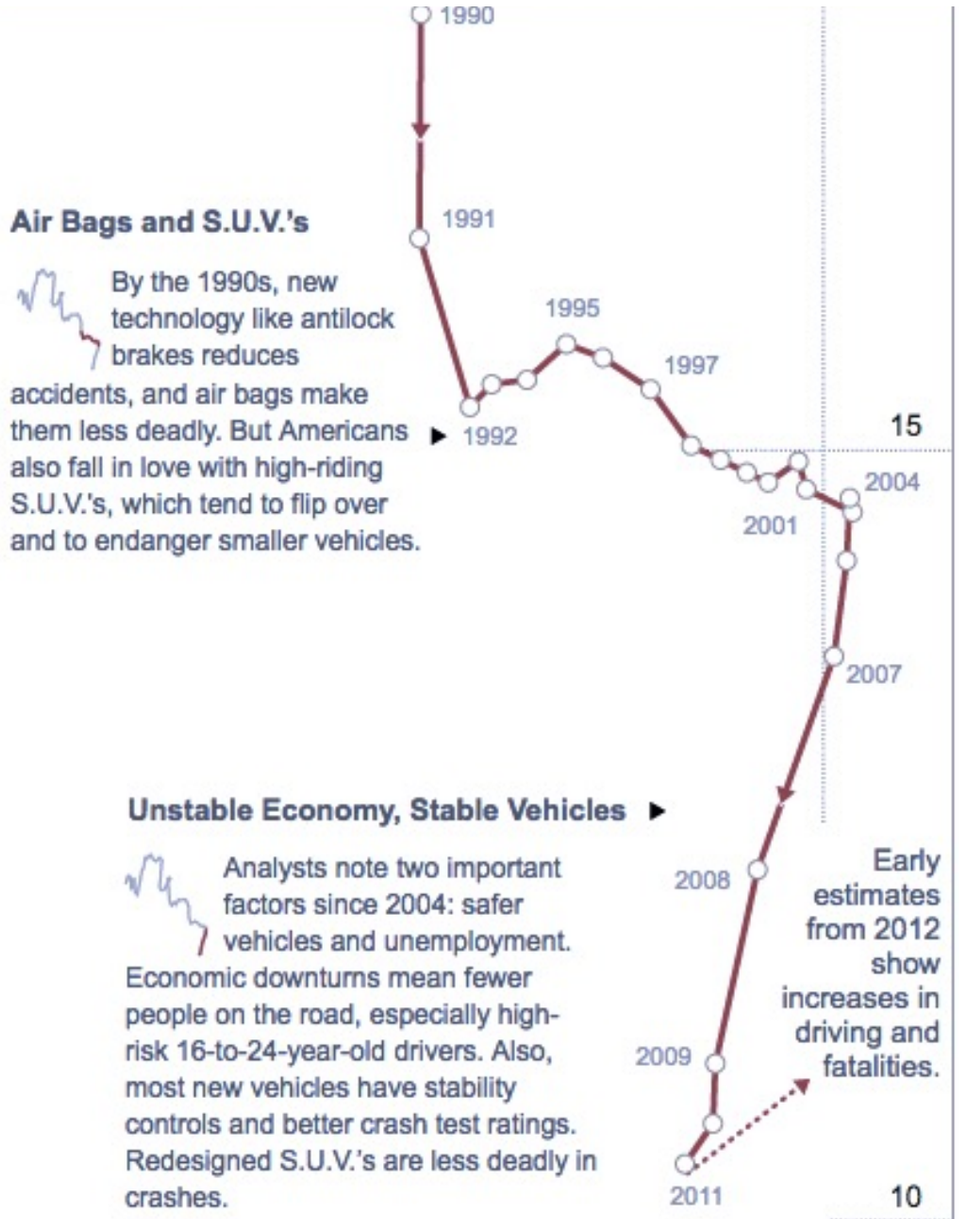
AMERICANS drive a staggering number of miles — close to three trillion every year, according to the government. (That is half a light-year, or 120 million trips around the world.) And although traffic accidents remain a major public safety problem, the biggest killer of people ages 5 to 34, vehicle travel is far safer than it was a few decades ago.

Several factors appear to account for the sharp decline in fatalities. Technology (like anti-lock brakes and air bags) and road behavior (like wearing seat belts and driving sober) have both improved greatly since 1950. Americans almost always drive more each year than the previous one — at least until recently, when the recession curtailed road habits. And the auto fatality rate has been decreasing since the 1960s, when cars with massive engines carried their unbuckled passengers on primarily two-lane roads.

The safety data is usually charted as deaths per miles traveled. But what happens when the metrics are teased apart, and familiar data is charted in an unfamiliar way?

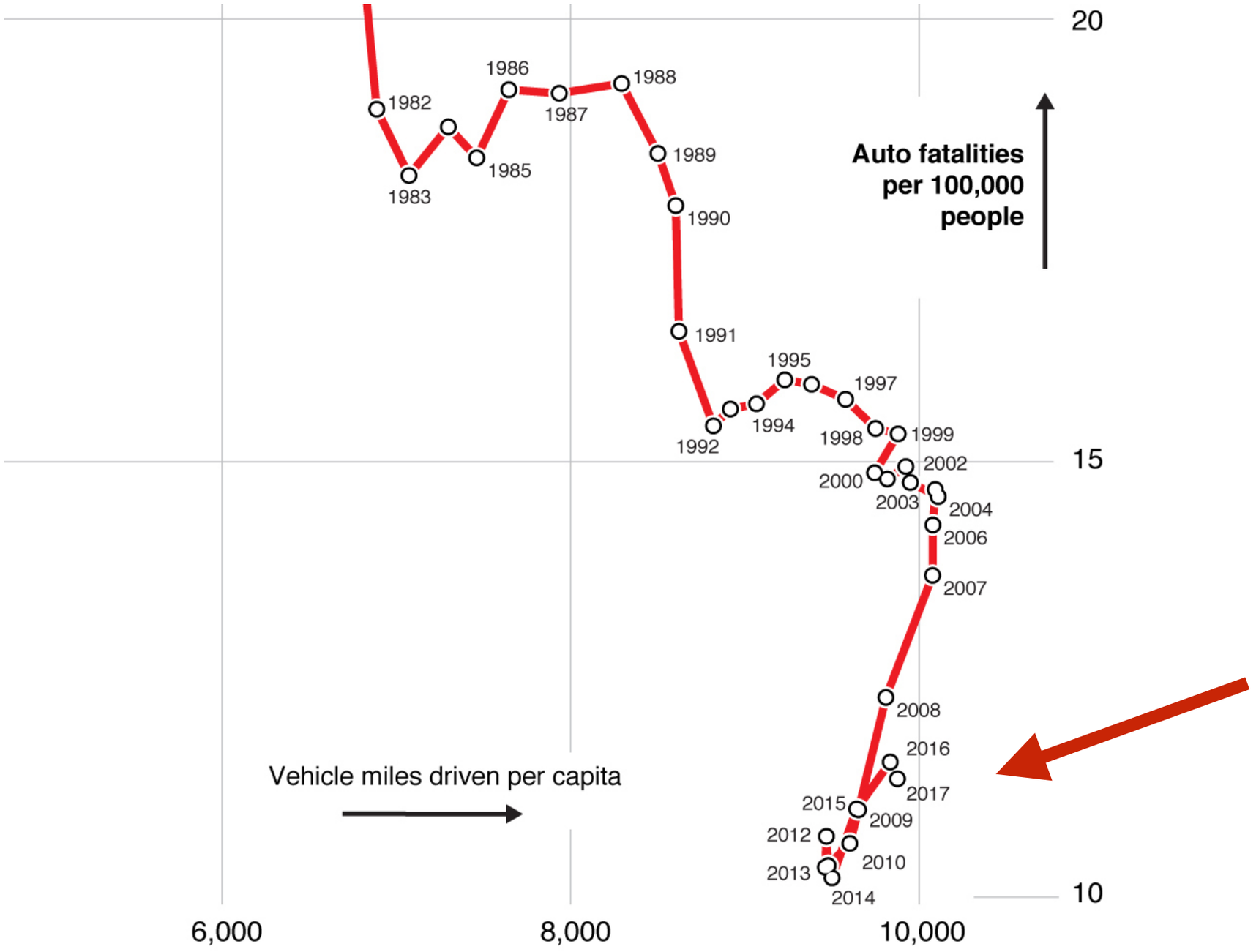
Plotting the two most important variables against each other — miles traveled versus deaths per 100,000 population — yields a pattern that looks like a plateau followed by a steep drop. It evokes the theory of punctuated equilibrium, proposed by the paleontologists Stephen Jay Gould and Niles Eldredge, which suggests that instead of continuous gradual evolution, change occurs abruptly after periods of virtual standstill.

"You see fatalities drop after a breakthrough in new technologies or behaviors, and then plateau until the next one," said David L. Strickland, administrator of the National Highway Traffic Safety Administration. "It takes time for new safety technologies to work their way into the whole fleet of cars on the road."



the annotation layer

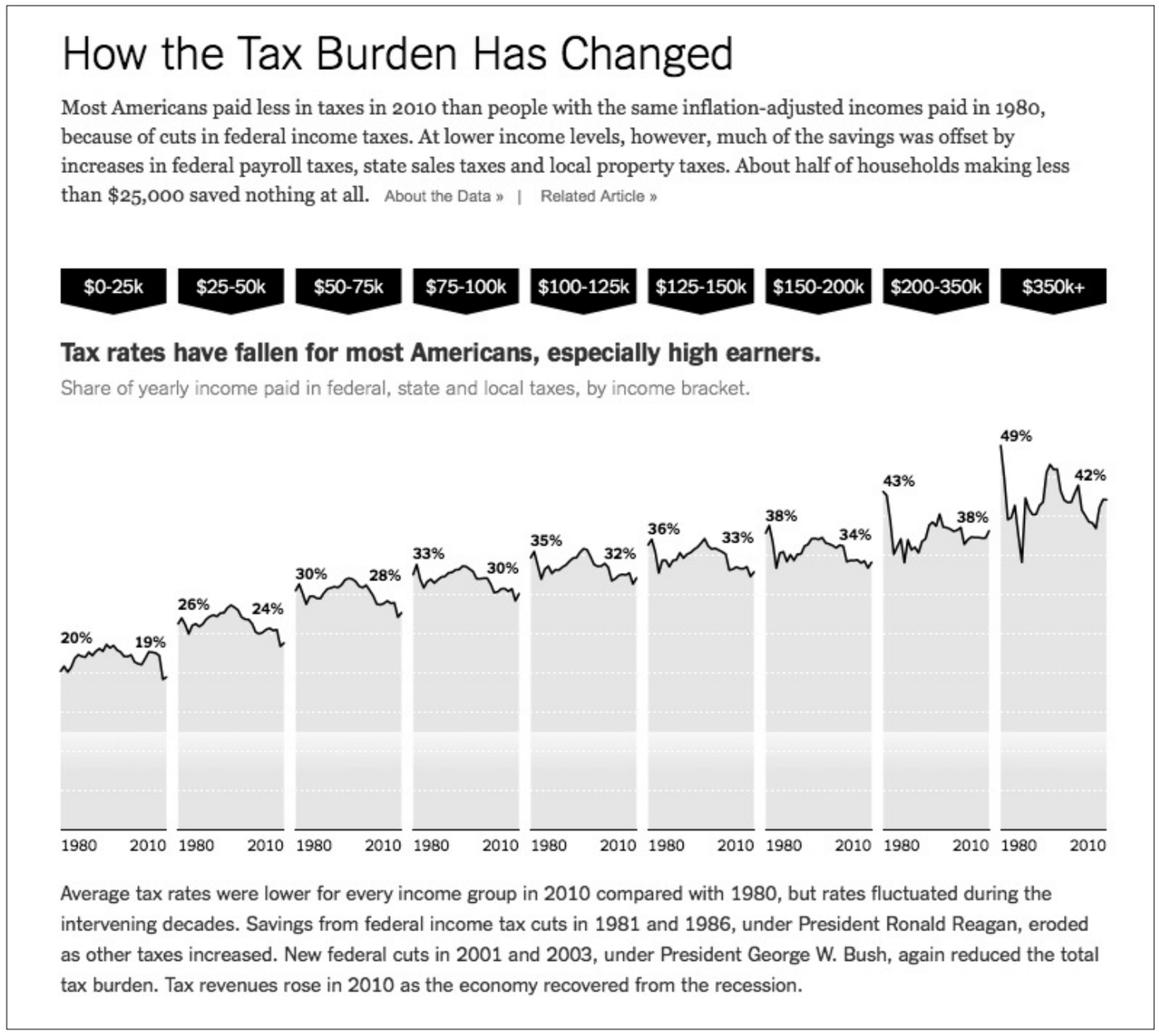
05 Use words, not just images

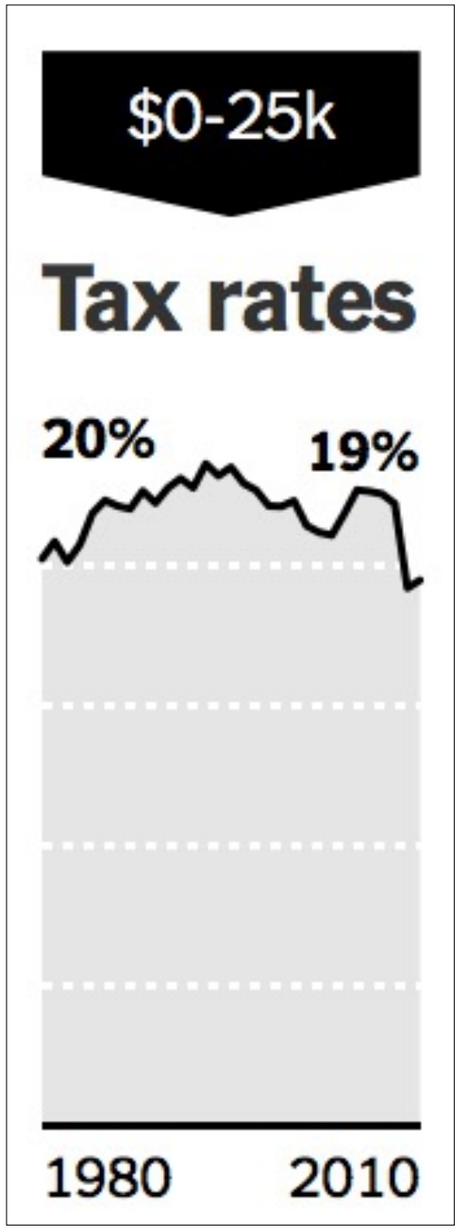


the annotation layer

http://online.sfsu.edu/trogu/523/2018/tutorials/nyt_driving_safety/

06 Use small multiples



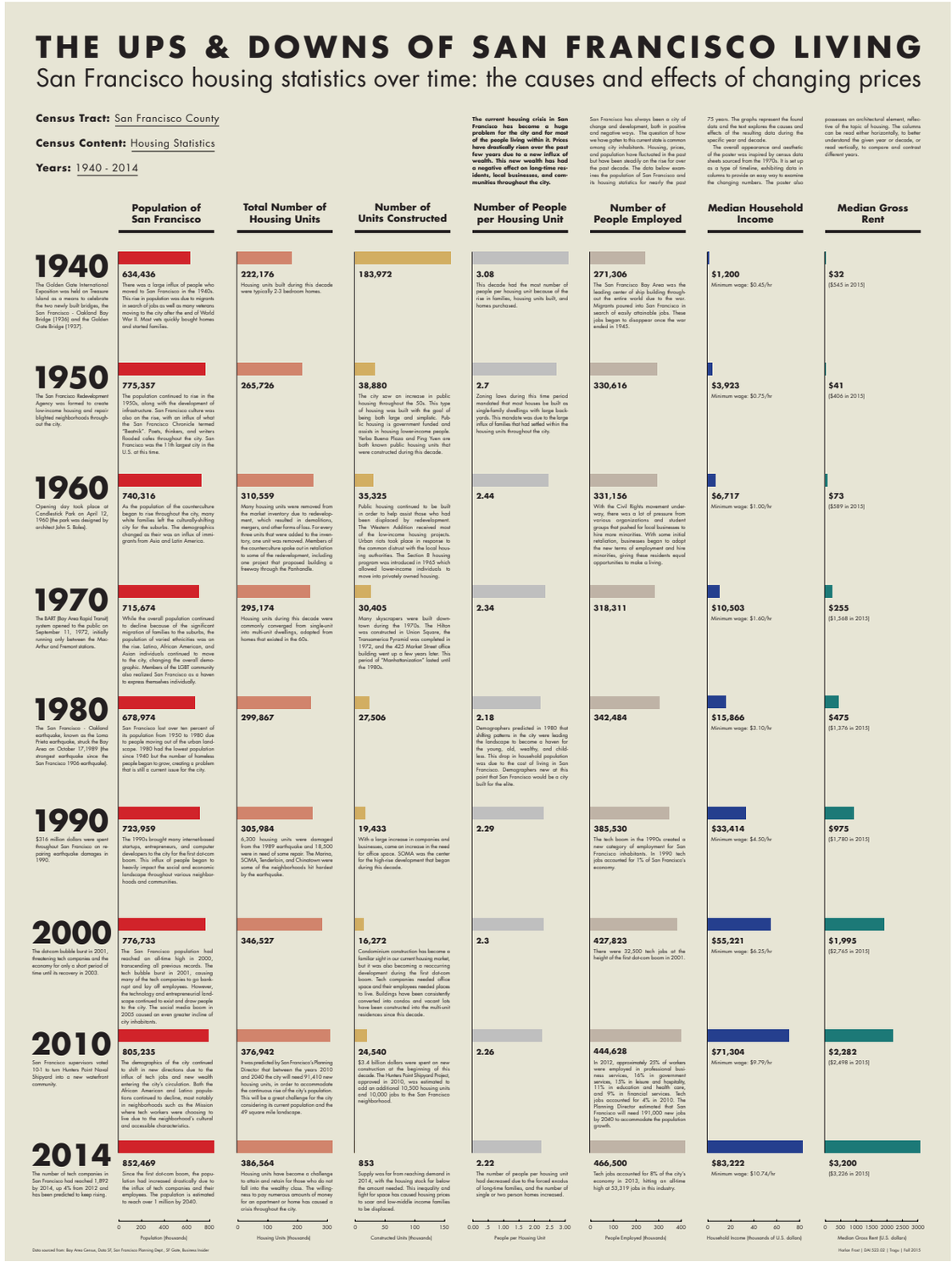


http://www.nytimes.com/interactive/2012/11/30/us/tax-burden.html

YES

NO

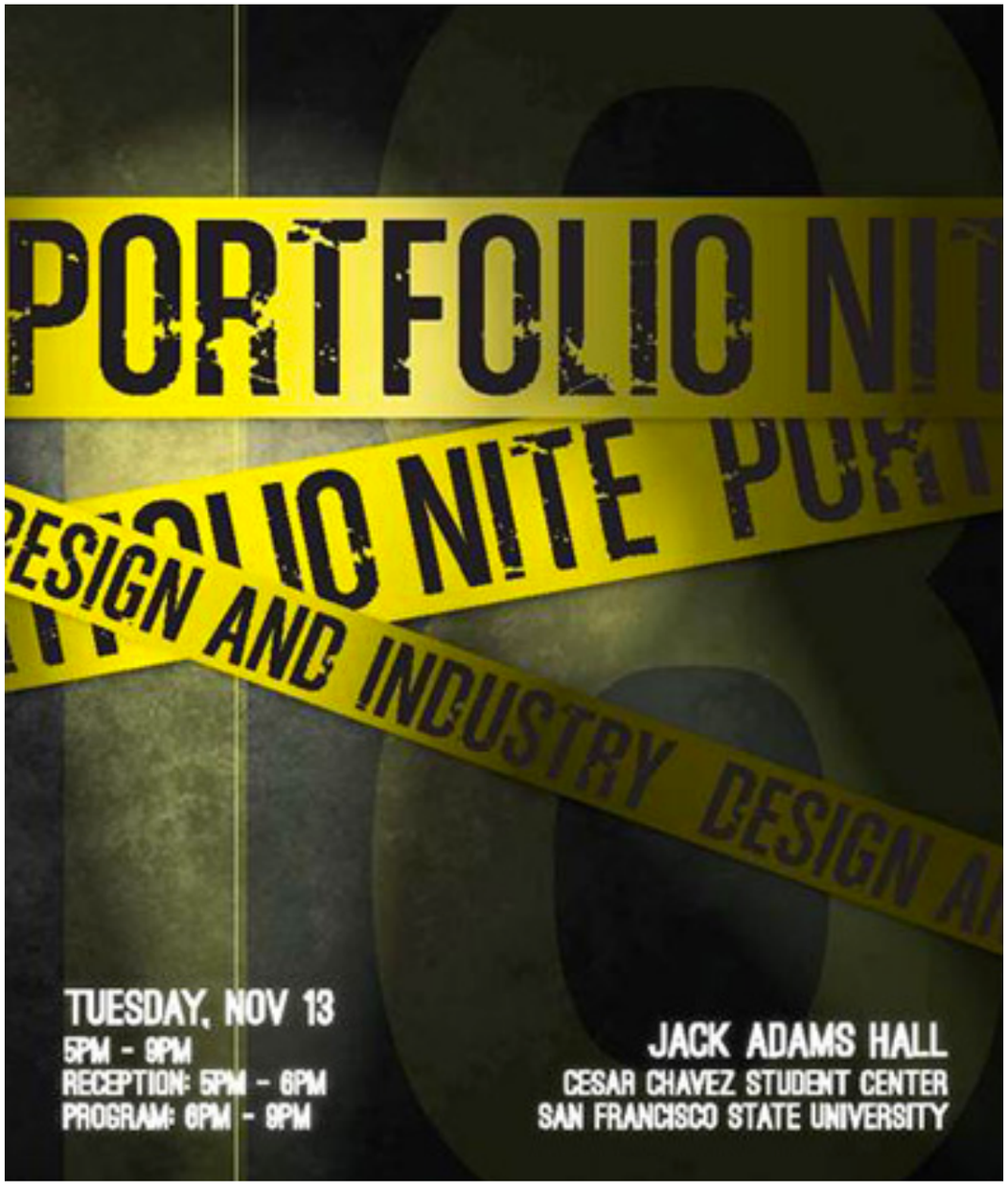
06 Use small multiples



unixlab.sfsu.edu/~trogu/523/2015/posters_30x40/san_francisco_housing_frost.pdf

http://523informationdesign.blogspot.com

07 Do not bungle the meaning



SFSU DAI AIGA Chapter

NO

08 Do not create op-art

Do not make op-art (optical art) effects by using bold condensed sans-serif fonts where the strokes are the same width as the counters in the font and also the width of the spacing between the letters. This creates a very annoying, vibrating checkerboard effect. Do not use solid backgrounds, boxes, thick borders, or arbitrary bold type. If you are using solid backgrounds throughout, invert the whole image to see if it's better with the opposite values. On a Mac, use control-option-command-8 to instantly invert the colors of your screen on the computer. See if it would be better the other way around (black type on white background). If nothing is gained by the solid fills, then get rid of them.

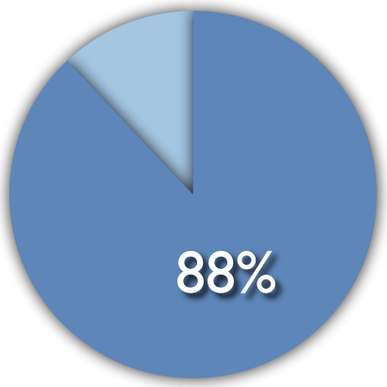
NO

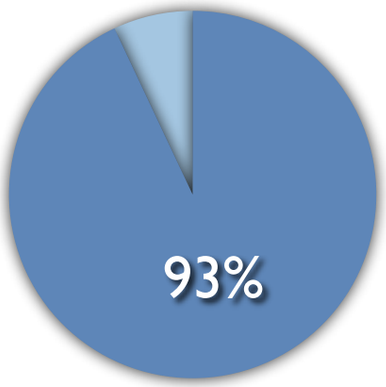
08 Do not create op-art

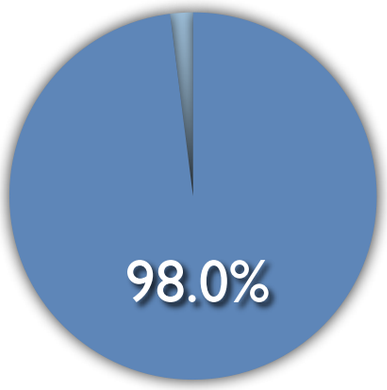
Do not make op-art (optical art) effects by using bold condensed sans-serif fonts where the strokes are the same width as the counters in the font and also the width of the spacing between the letters. This creates a very annoying, vibrating checkerboard effect. Do not use solid backgrounds, boxes, thick borders, or arbitrary bold type. If you are using solid backgrounds throughout, invert the whole image to see if it's better with the opposite values. On a Mac, use control-option-command-8 to instantly invert the colors of your screen on the computer. See if it would be better the other way around (black type on white background). If nothing is gained by the solid fills, then get rid of them.

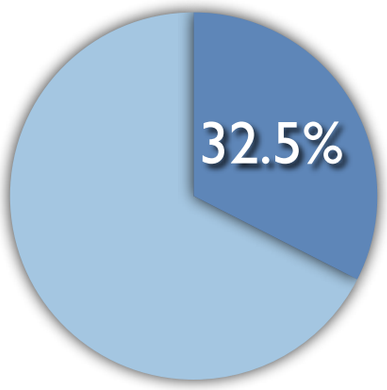
NO

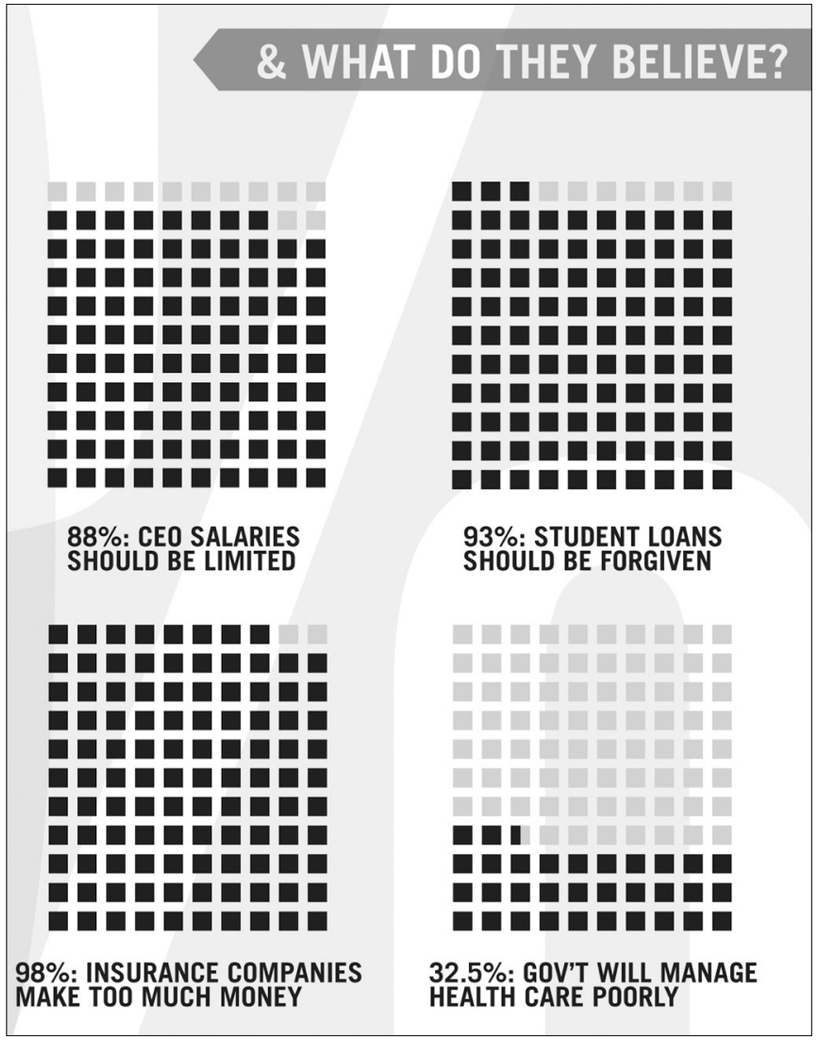
09 Do not use little dots for numbers











Student project: J.L.

YES

NO

09 Do not use little dots for numbers

My problem is that

I have been persecuted

by an integer.

GEORGE A. MILLER

MAGICAL NUMBER SEVEN, 1956

09 Do not use little dots for numbers

4-3-4-6-5-9-6-2-3

“chunks”

434-65-9623

09 Do not use little dots for numbers

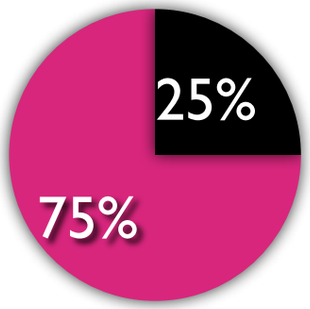
1/3

YES

NO

09 Do not use little dots for numbers

(chunk data)



3/4

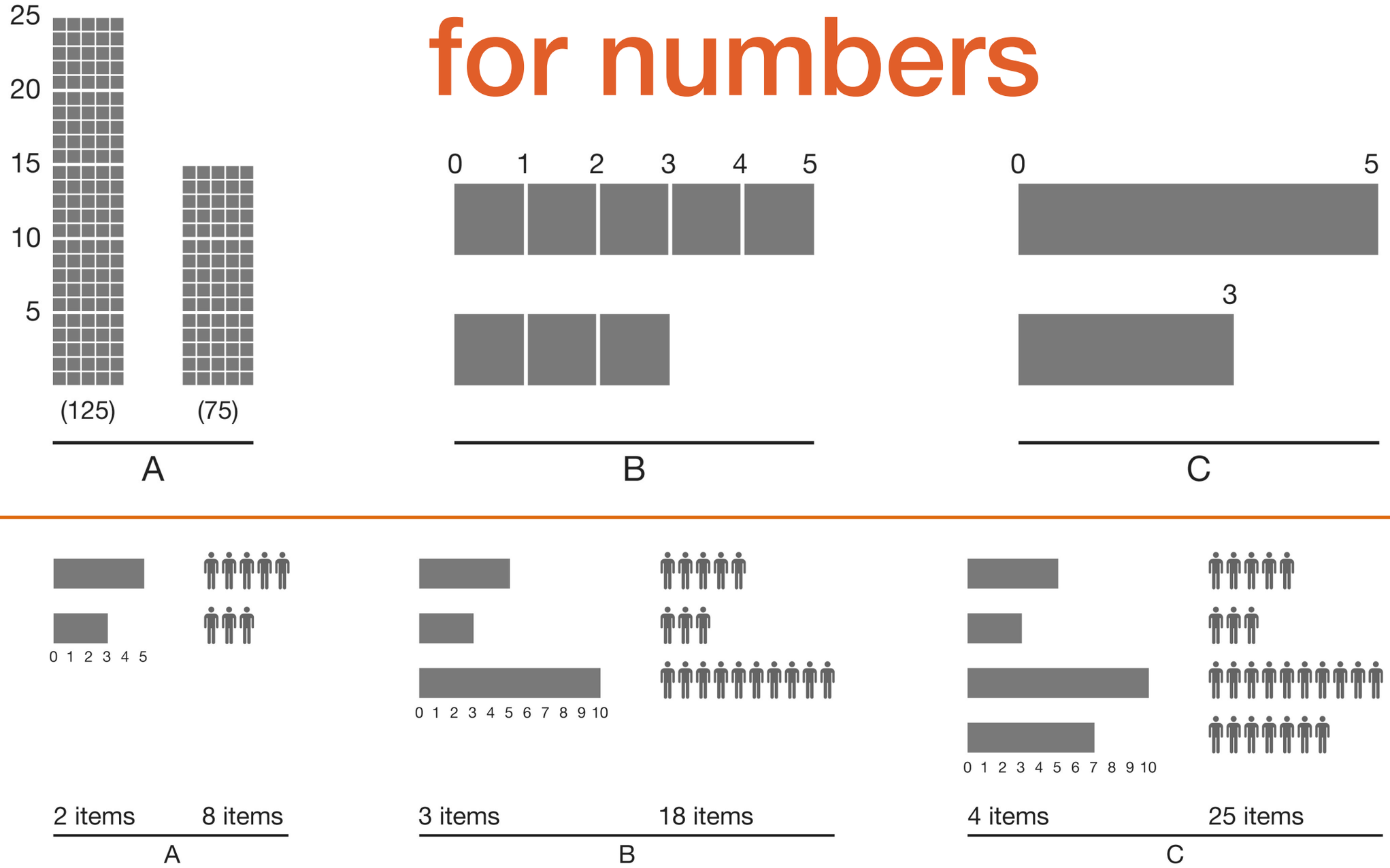
YES

(do not un-chunk data)

Student project: T. B.

NO

09 Do not use little men for numbers

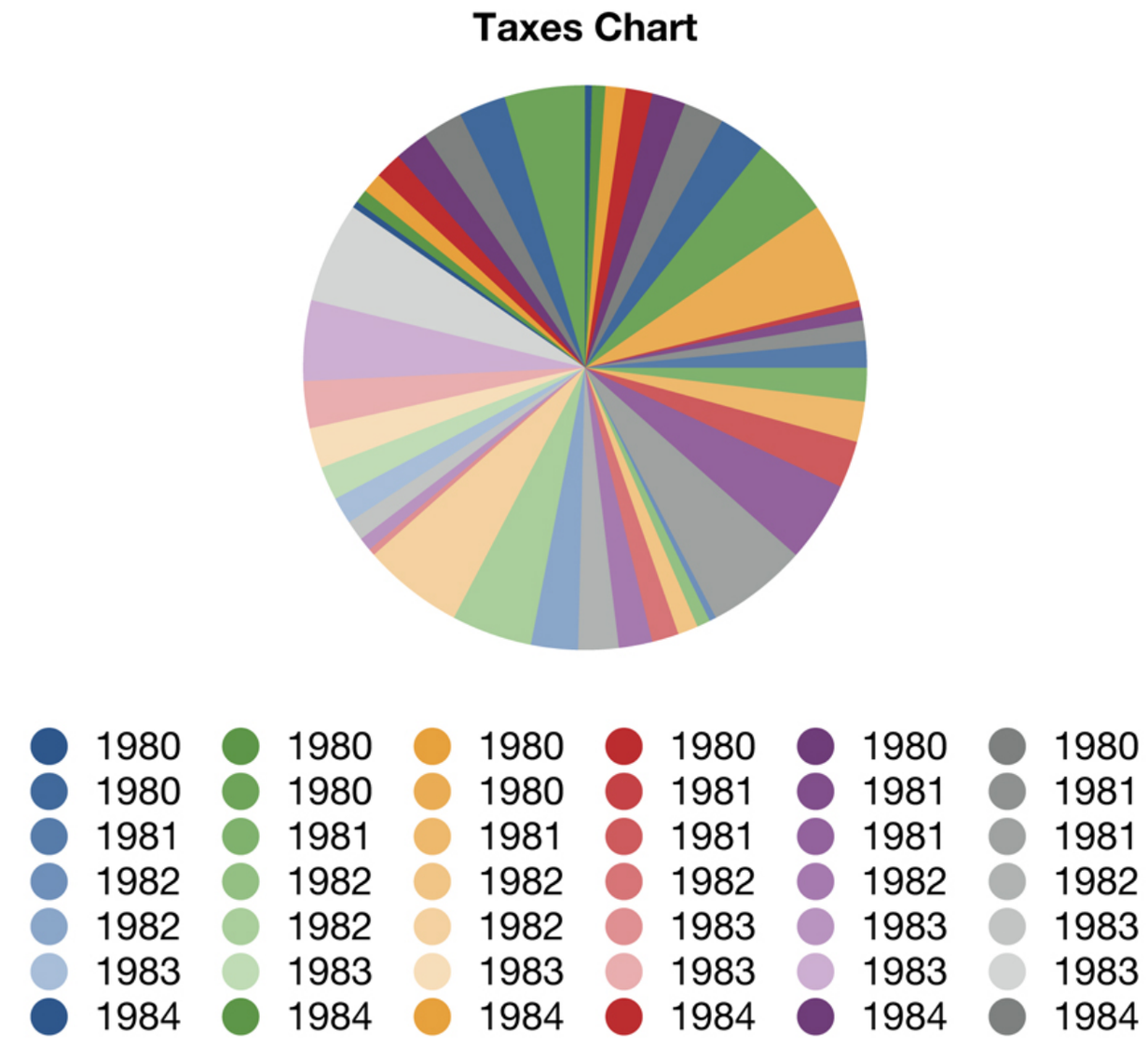


Trogu, Pino. "Counting But Losing Count:
the Legacy of Otto Neurath's Isotype Charts"
Visible Language, 52.2, 2018

http://static.trogu.com/documents/articles/neurath/VisibleLanguage-52-2-8-2018-p82-109-Trogu.pdf

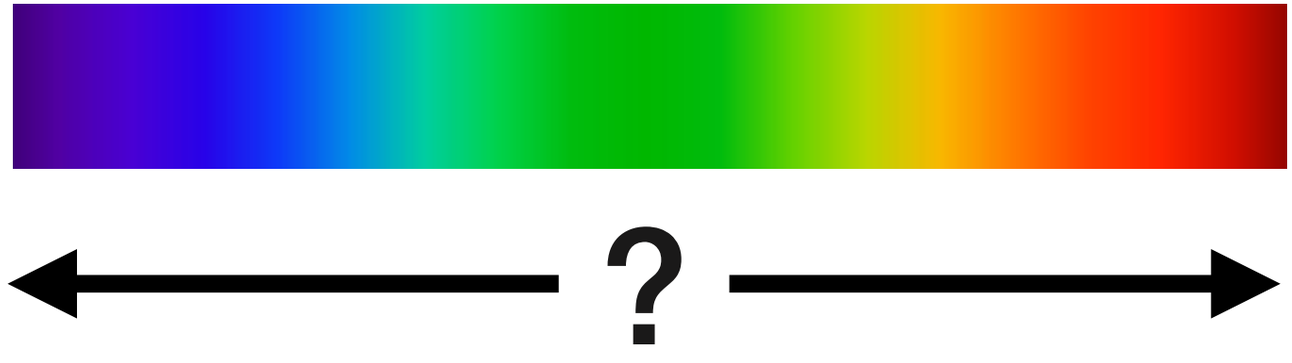
YES | NO

10 Do not use colors (to be memorized, or for sorting)



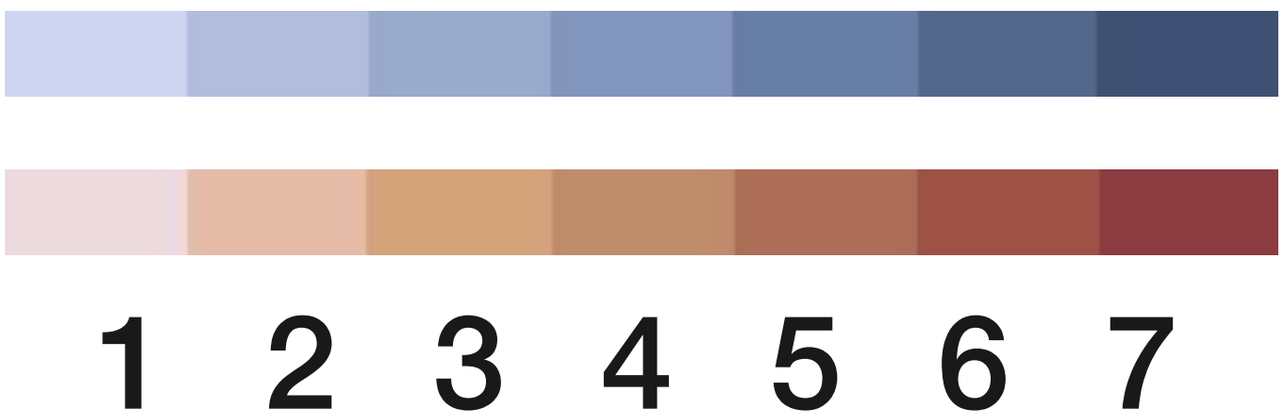
NO

10 Do not use colors (for ordered data)



NO

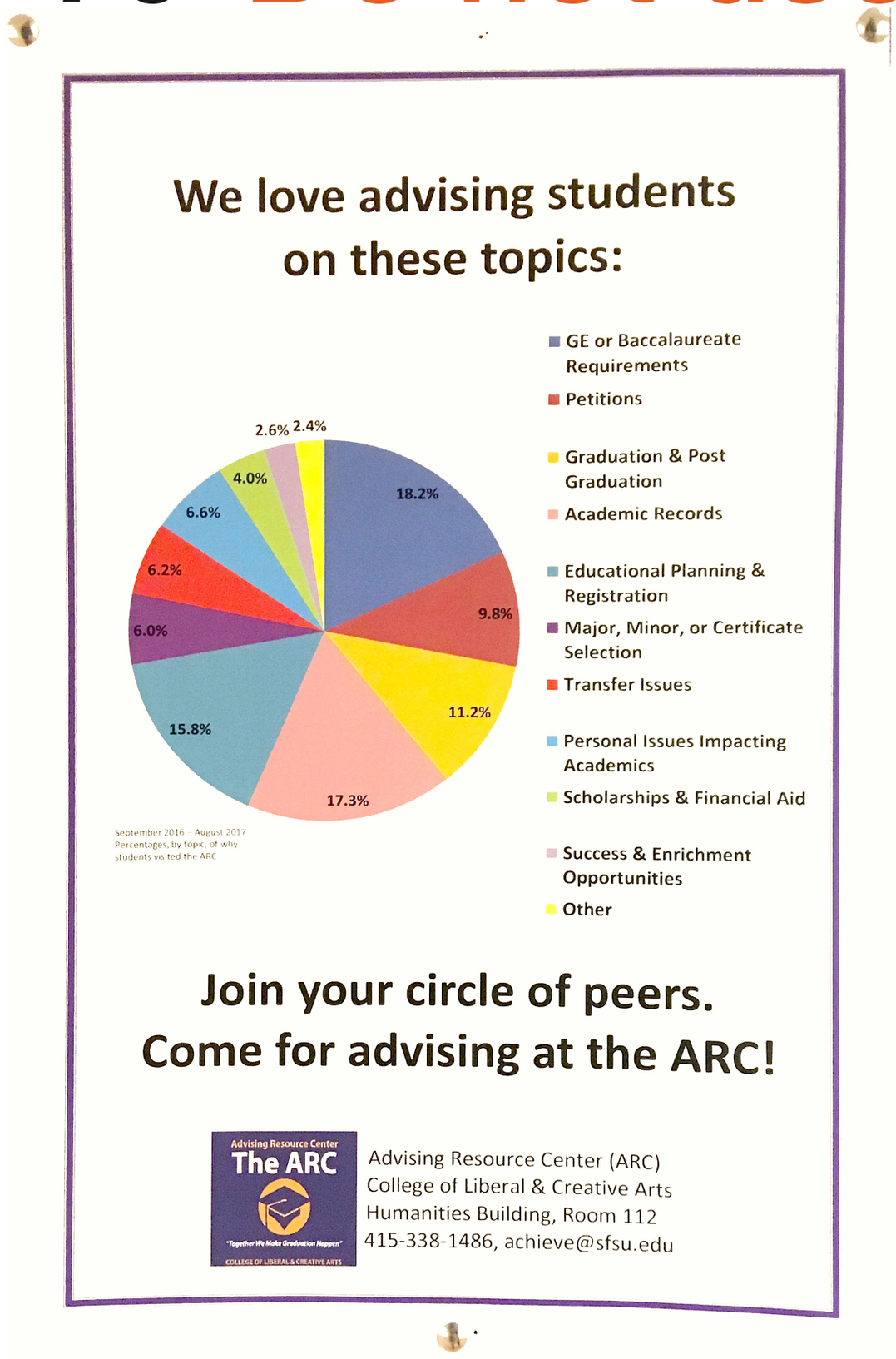
no intrinsic order



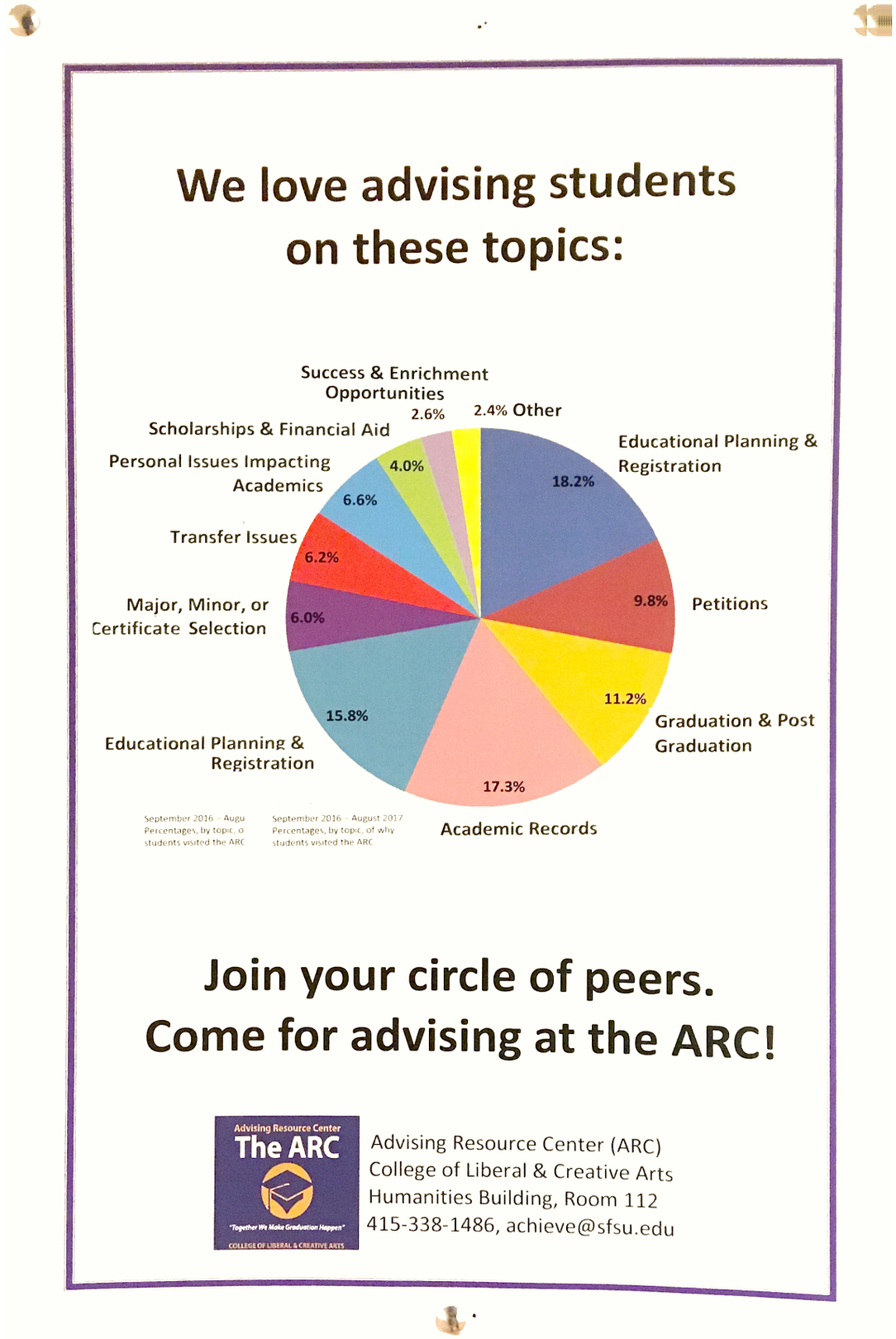
YES

10 Do not use colors

NO



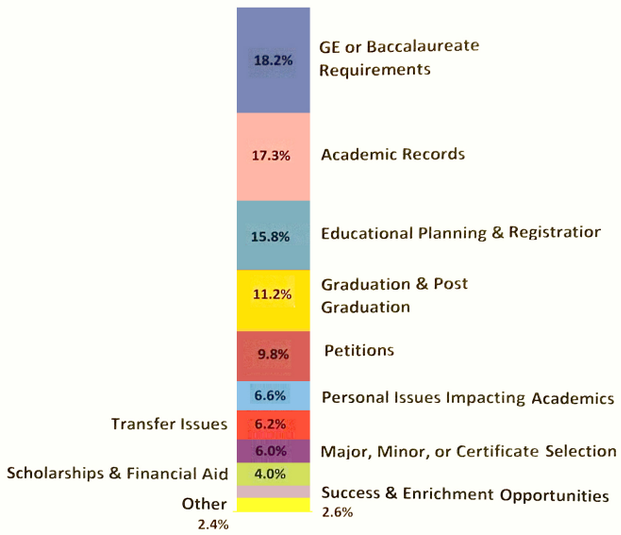
10 Do not use colors



Omit legend whenever possible.

10 Do not use colors

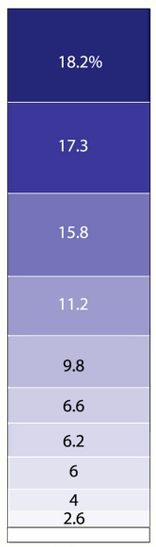
We love advising students on these topics:



Join your circle of peers.
Come for advising at the ARC!

Advising Resource Center (ARC)
College of Liberal & Creative Arts
Humanities Building, Room 112
415-338-1486, achieve@sfsu.edu

We love advising students on these topics:

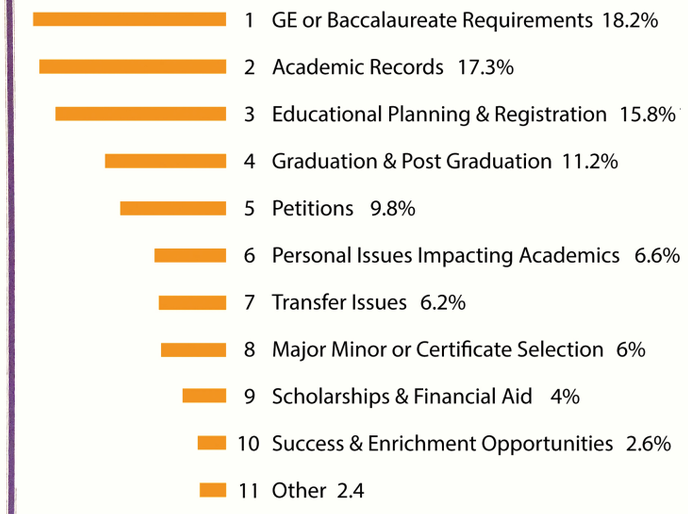


1 GE or Baccalaureate Requirements
2 Academic Records
3 Educational Planning & Registration
4 Graduation & Post Graduation
5 Petitions
6 Personal Issues Impacting Academics
7 Transfer Issues
8 Major Minor or Certificate Selection
9 Scholarships & Financial Aid
10 Success & Enrichment Opportunities
11 Other 2.4

Join your circle of peers.
Come for advising at the ARC!

Advising Resource Center (ARC)
College of Liberal & Creative Arts
Humanities Building, Room 112
415-338-1486, achieve@sfsu.edu

We love advising students on these topics:

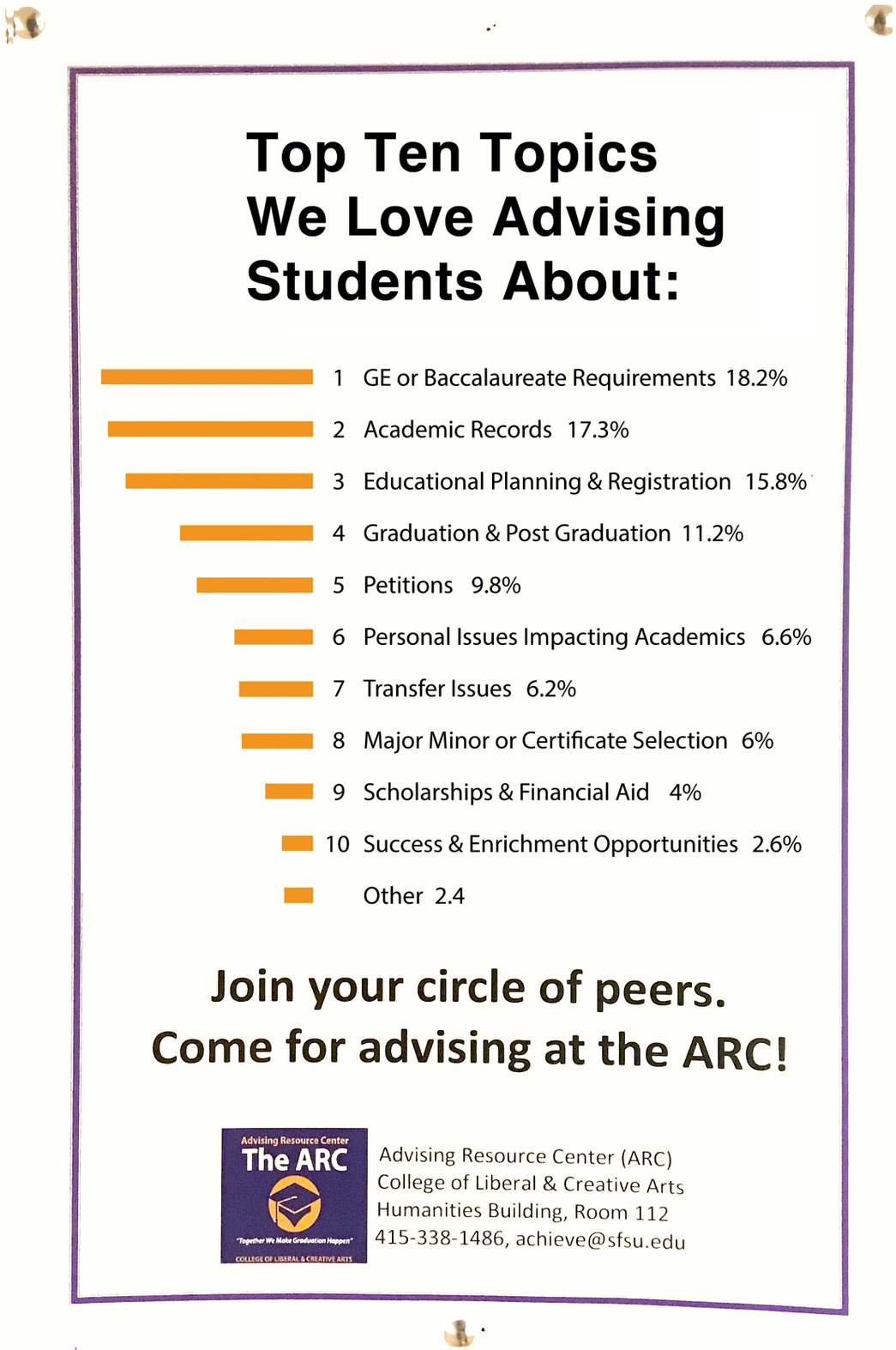


1 GE or Baccalaureate Requirements 18.2%
2 Academic Records 17.3%
3 Educational Planning & Registration 15.8%
4 Graduation & Post Graduation 11.2%
5 Petitions 9.8%
6 Personal Issues Impacting Academics 6.6%
7 Transfer Issues 6.2%
8 Major Minor or Certificate Selection 6%
9 Scholarships & Financial Aid 4%
10 Success & Enrichment Opportunities 2.6%
11 Other 2.4

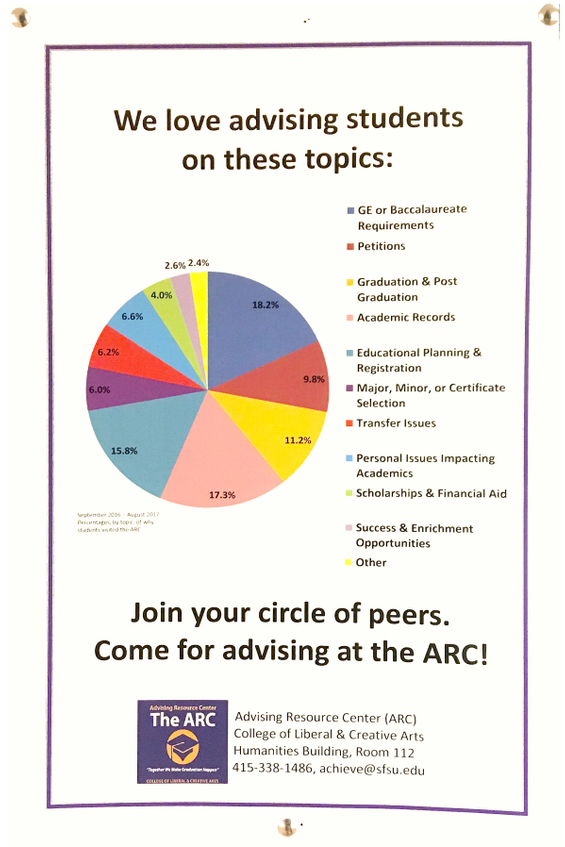
Join your circle of peers.
Come for advising at the ARC!

Advising Resource Center (ARC)
College of Liberal & Creative Arts
Humanities Building, Room 112
415-338-1486, achieve@sfsu.edu

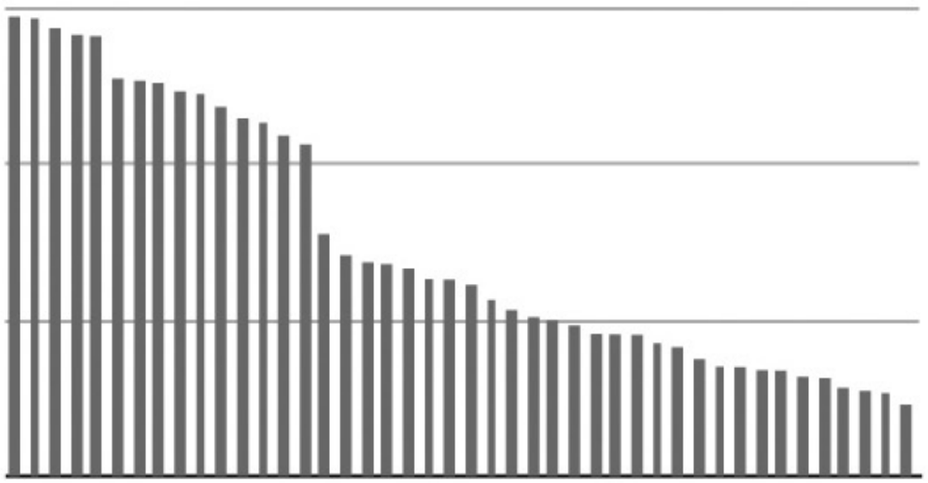
10 Do not use colors



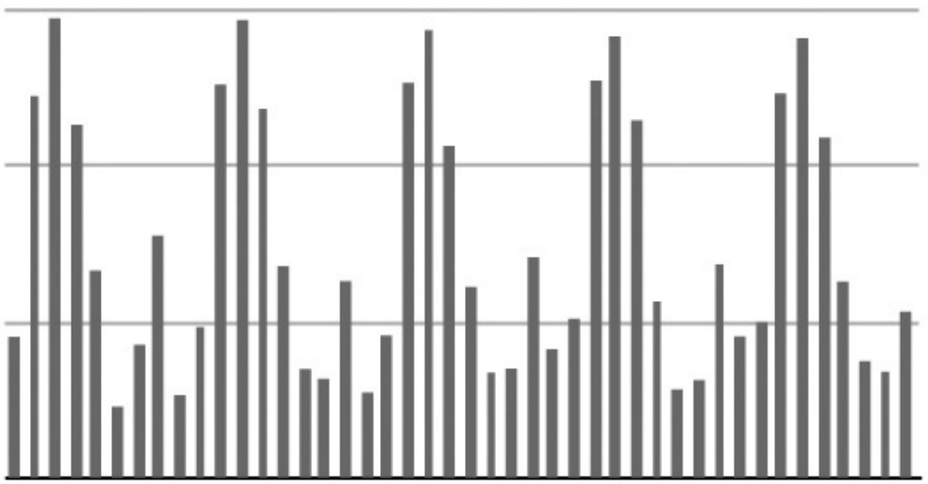
Sometimes a word is worth a thousand colors.



11 Sort by value, not by category (alphabetical)

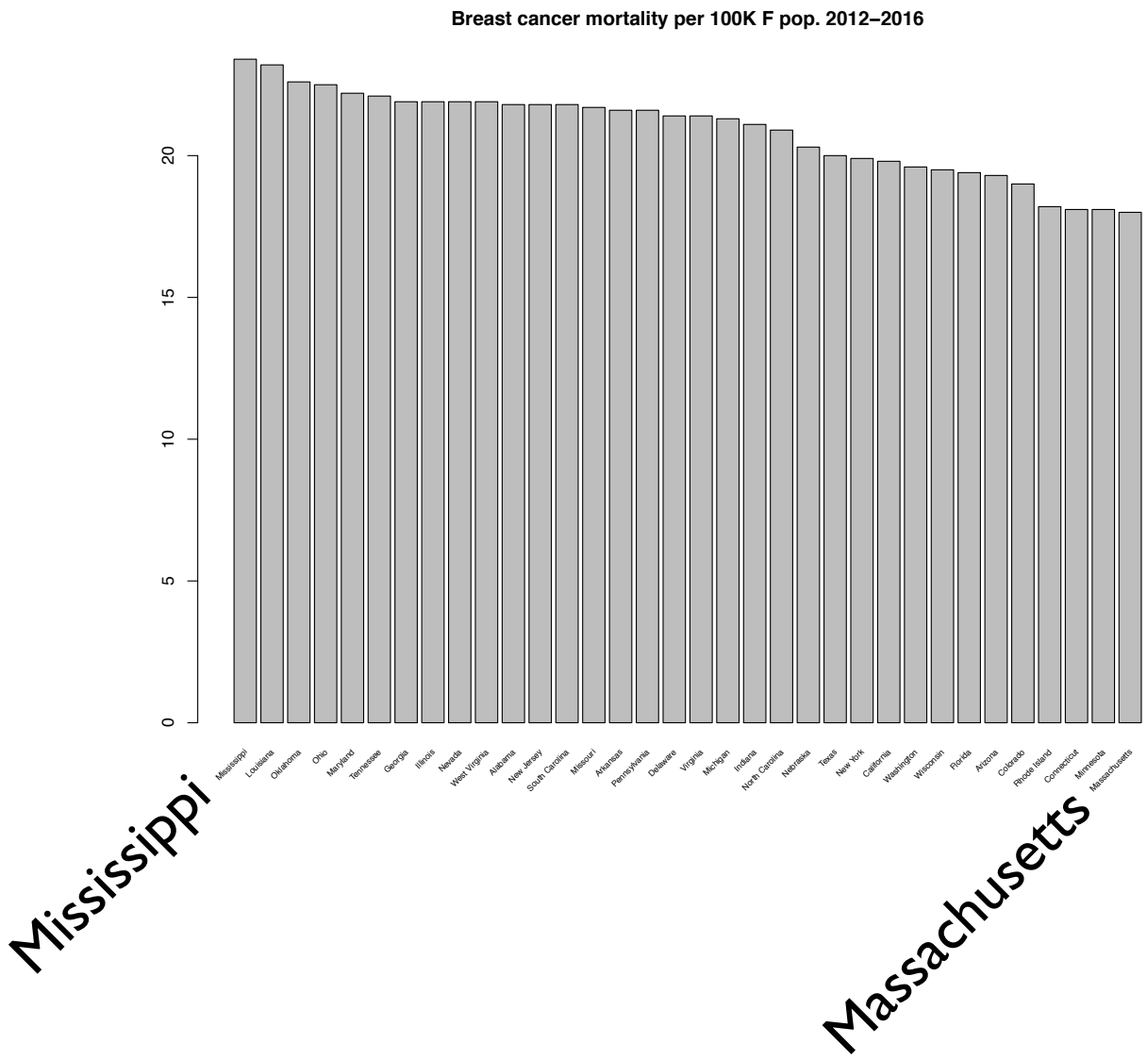


YES

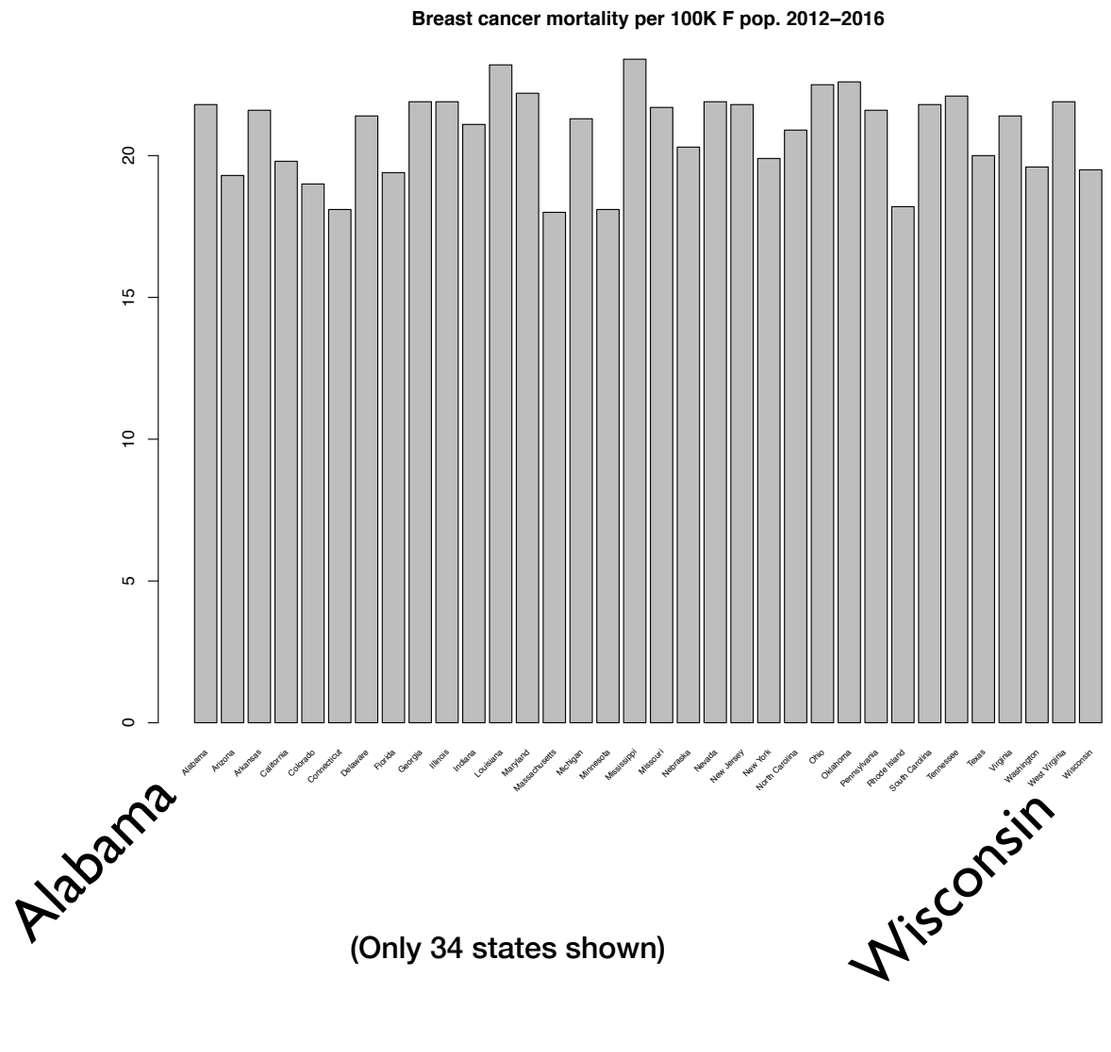


NO

11 Sort by value, not by category (alphabetical)



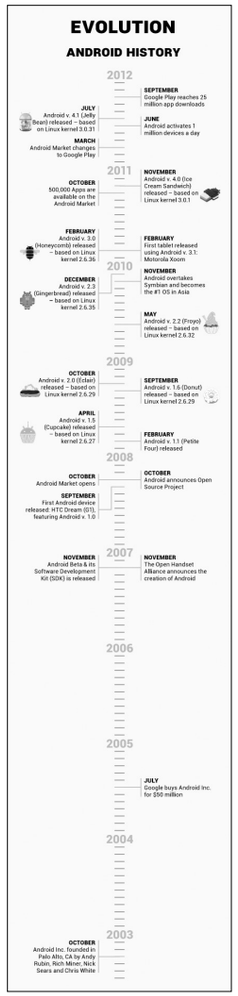
YES

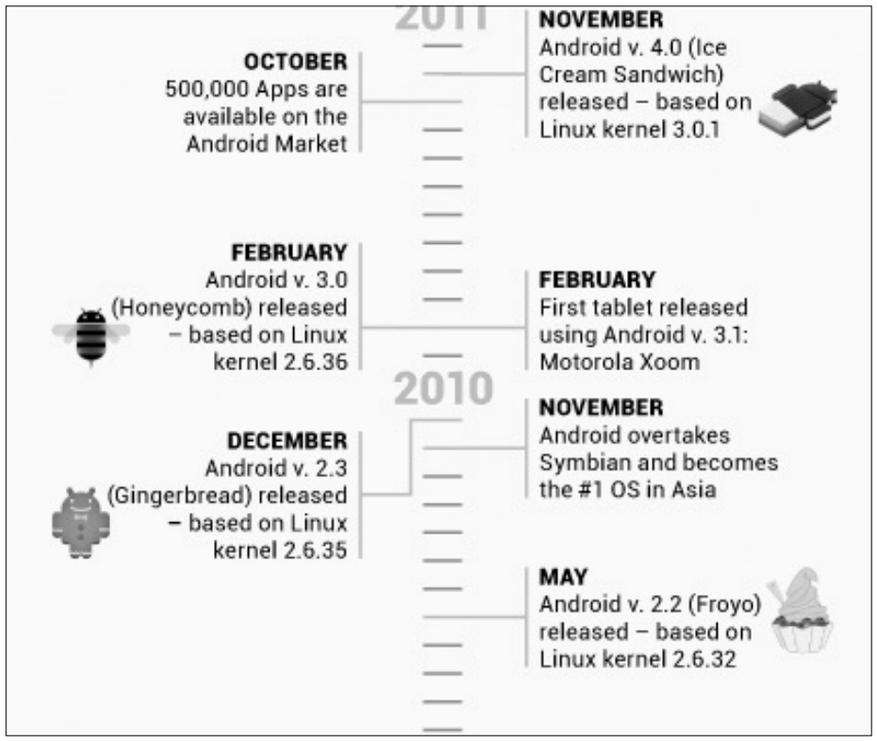


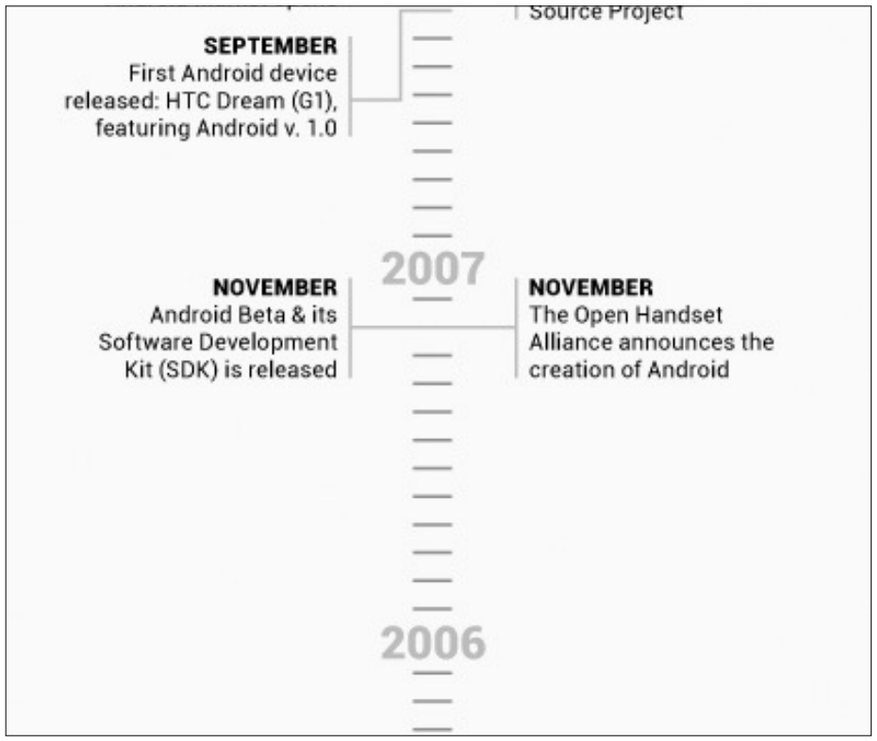
(Only 34 states shown)

NO

12 Equally space time intervals in timelines



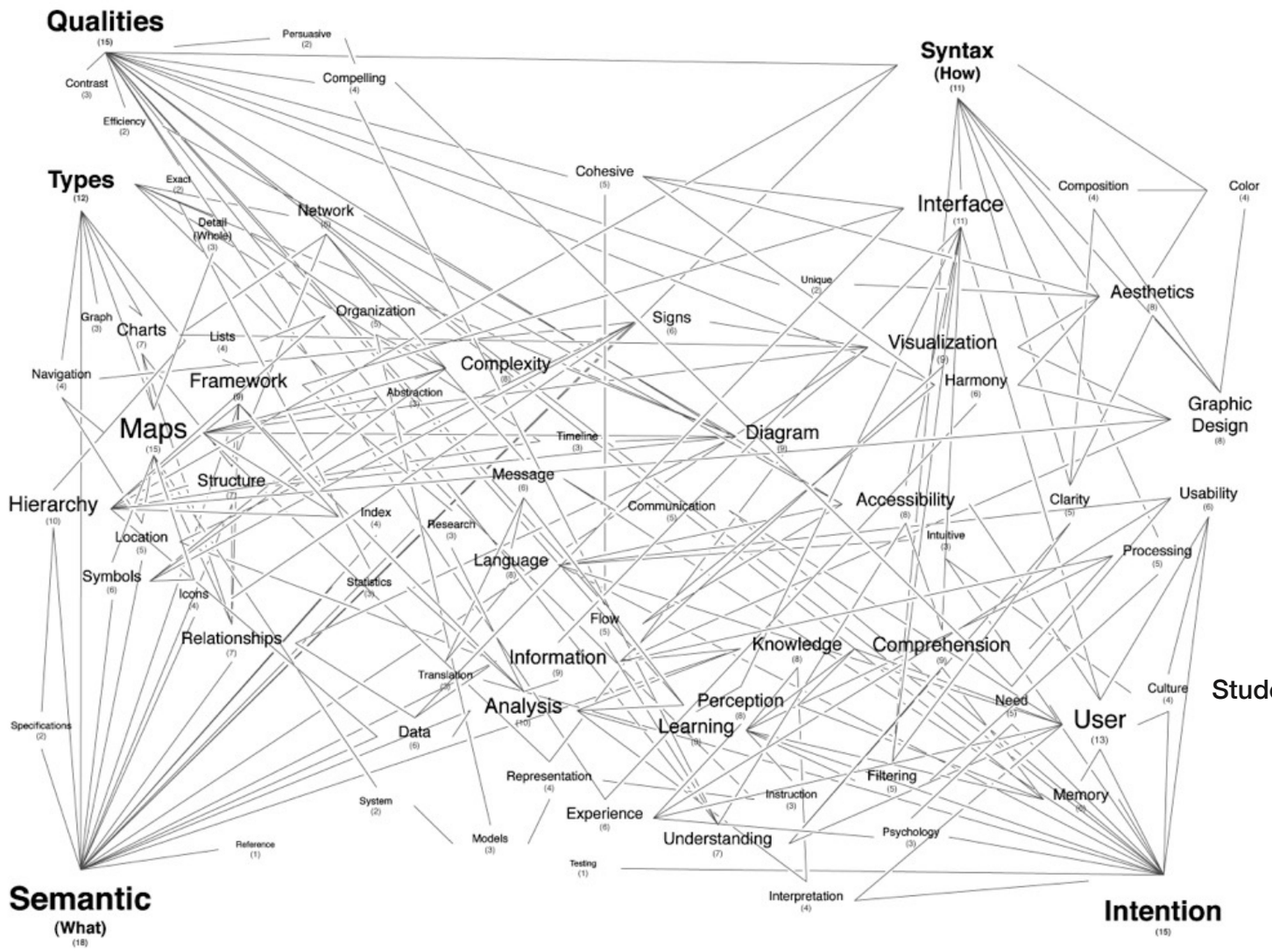




Student project: J.C.

YES

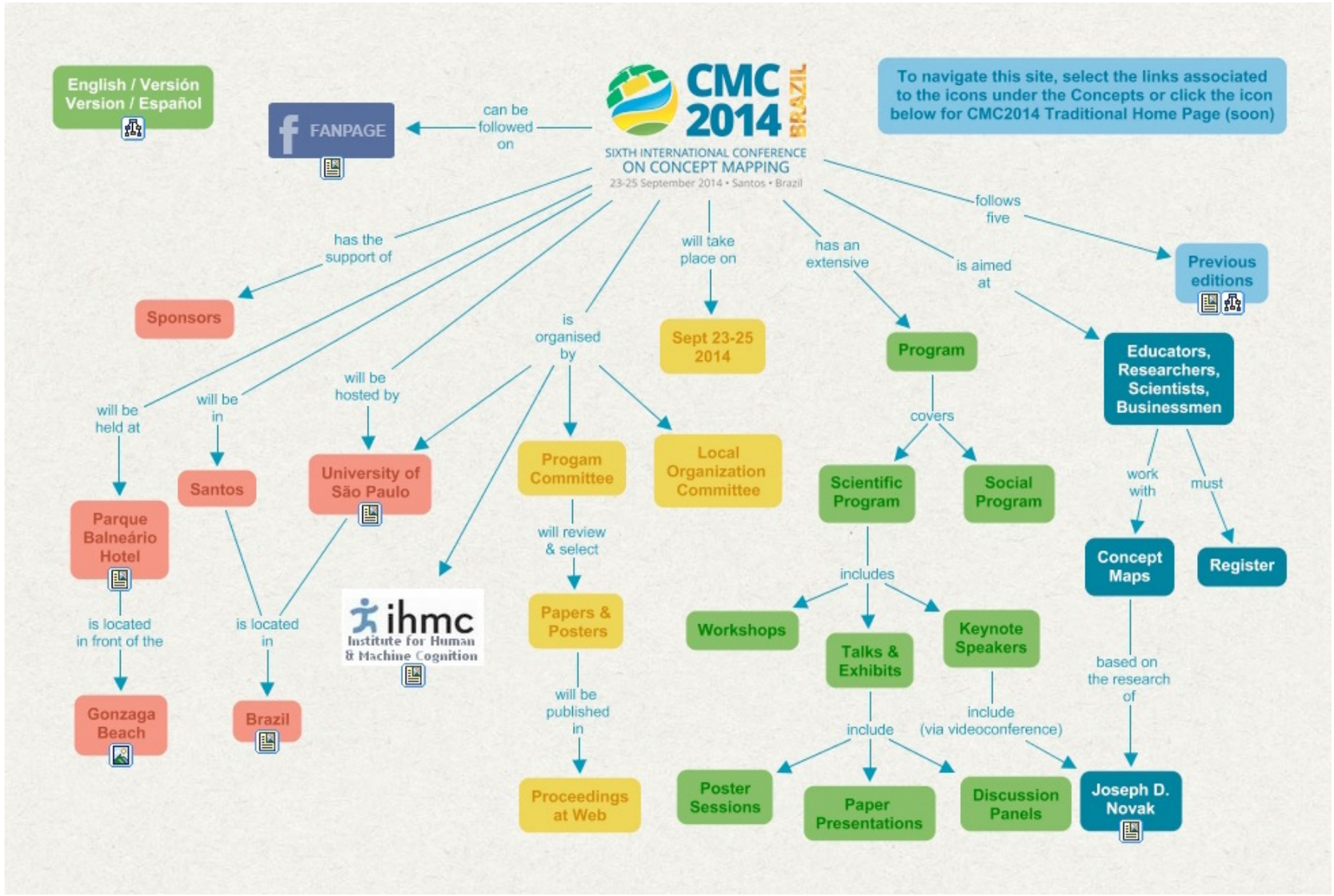
13 Avoid meaningless concept maps



Student project: N.M.

NO

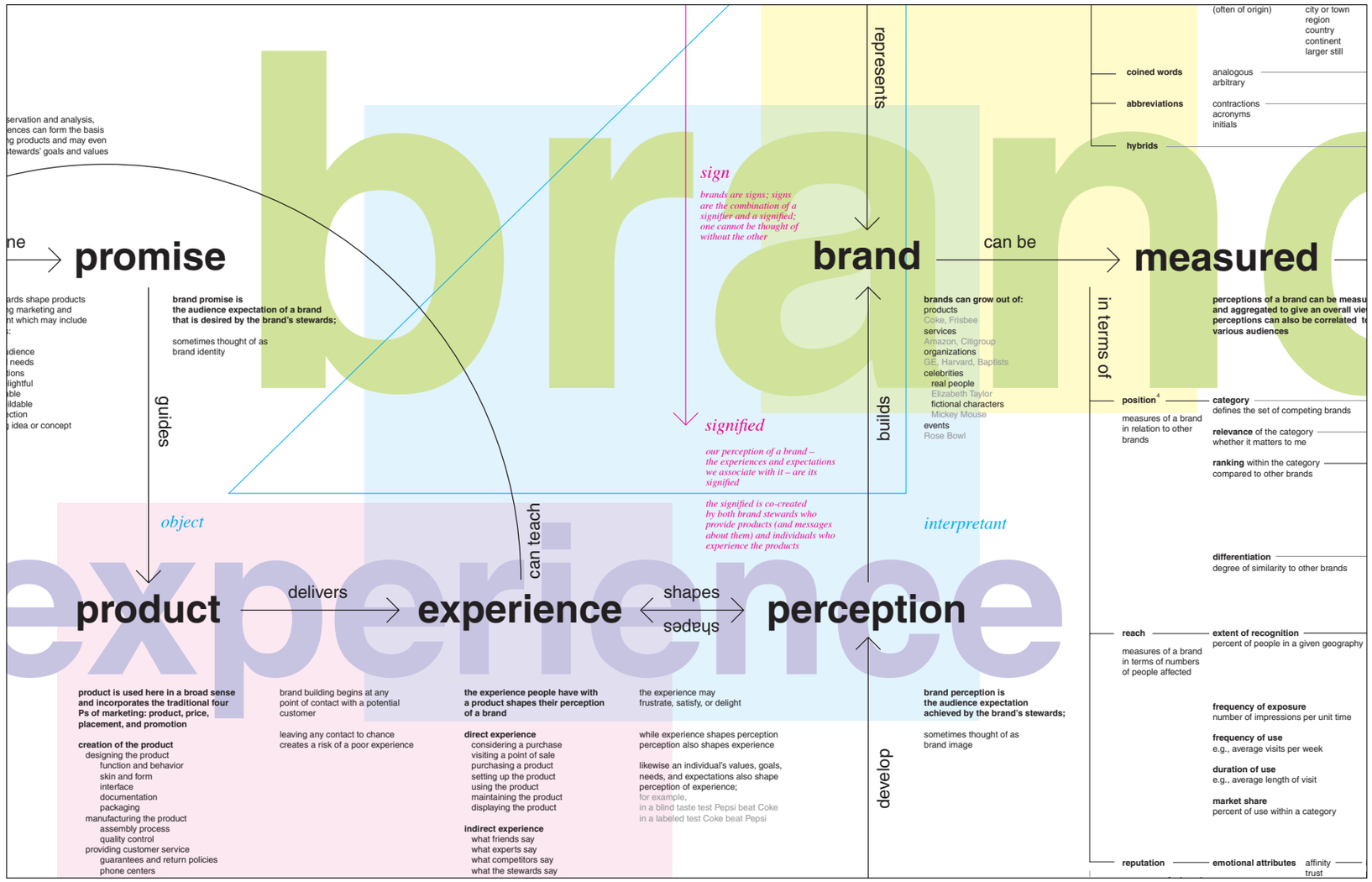
13 Avoid meaningless concept maps



http://cmc.ihmc.us/

NO

13 Avoid meaningless concept maps



http://www.dubberly.com/concept-maps/a-model-of-brand.html

NO

13 Avoid meaningless concept maps

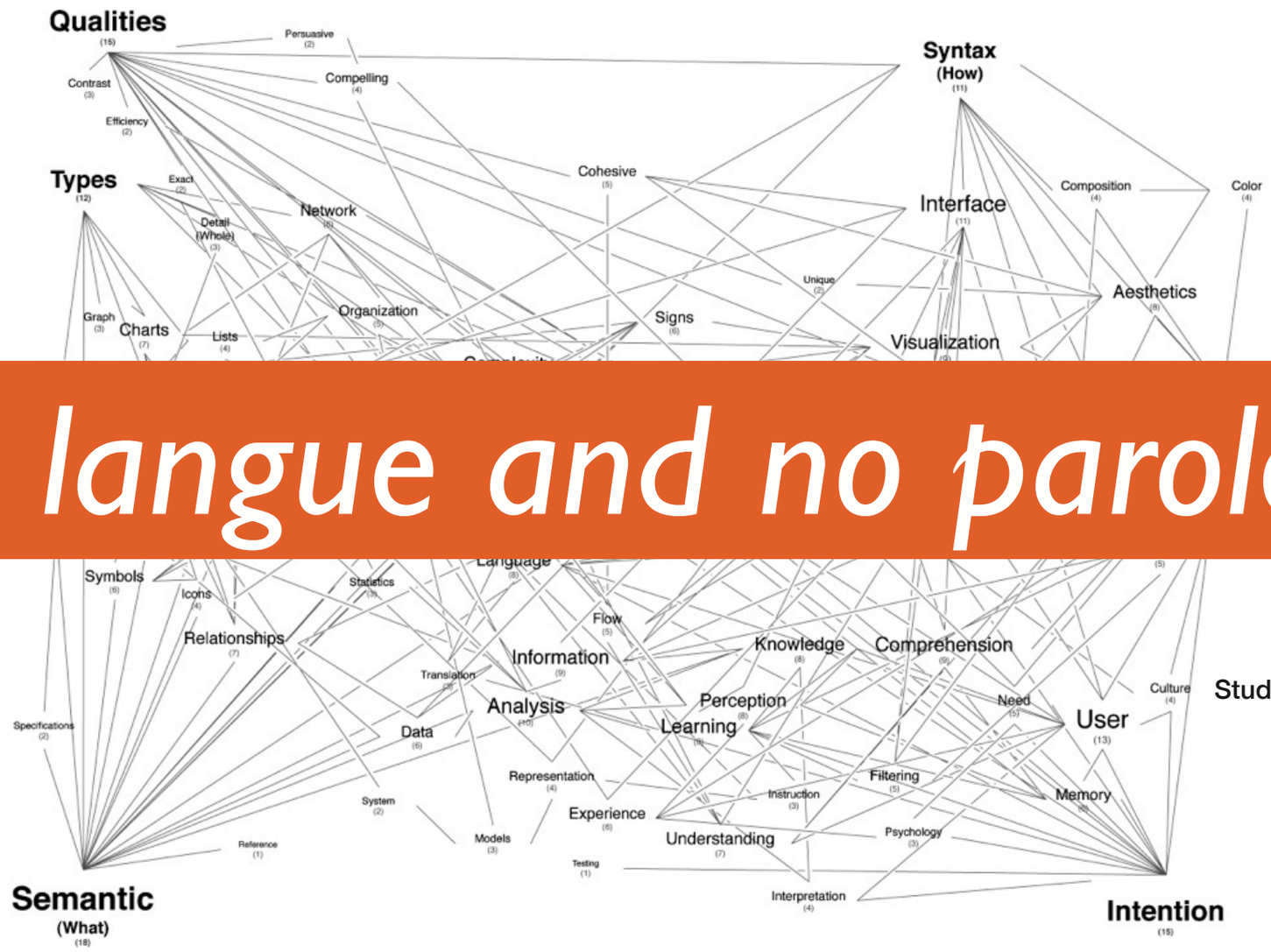
y important sensible and tactile attributes, another, more
uction of the reader's mental image of the book. This is
rds, **the fixed sequence of words making up sentences,**
ragraphs and so on, until the full complete book is

13 Avoid meaningless concept maps

important distinction
between
langue & parole

FERDINAND DE SAUSSURE
LINGUISTICS, 1906-11

13 Avoid meaningless concept maps

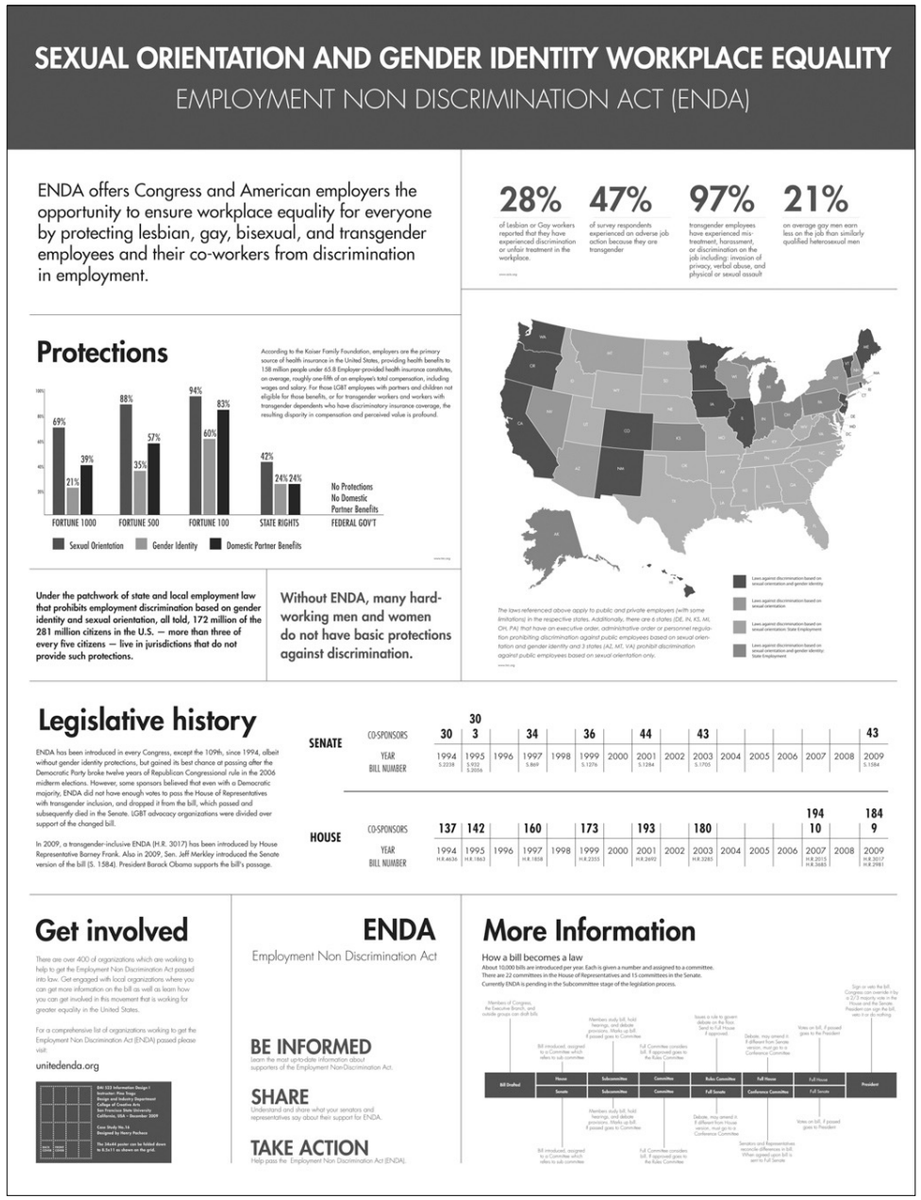


all langue and no parole

Student project: N.M.

NO

14 You can use small type



Student project: H.P.

YES

15 Do not screen type (for print)

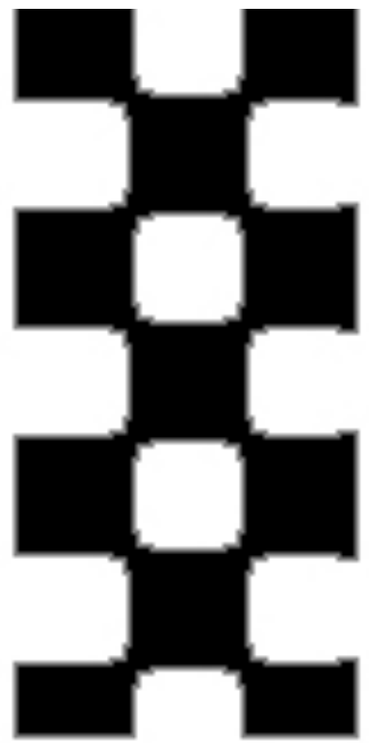
Additional instructions, examples, and videos related to the assignments in this list may be found on iLearn. Many of the examples shown here are for reference only, you will be drawing your own objects and views as appropriate. 1 Milk carton orthographic 5pts Draw the six views of the milk carton using light lines for the projection lines between the views. Use darker lines for the actual object lines. Label each view: front, top, right side, bottom, left side, back. Draw a border around the sheet (1/2″ from the edge) and a 3/4″ title block at the bottom. Include the drawing name, drawing number, your name and the date. Draw thin guidelines for your lettering and labels. This is a freehand drawing. Practice keeping the lines straight, parallel, and even in thickness and value.

Additional instructions, examples, and videos related to the assignments in this list may be found on iLearn. Many of the examples shown here are for reference only, you will be drawing your own objects and views as appropriate. 1 Milk carton orthographic 5pts Draw the six views of the milk carton using light lines for the projection lines between the views. Use darker lines for the actual object lines. Label each view: front, top, right side, bottom, left side, back. Draw a border around the sheet (1/2″ from the edge) and a 3/4″ title block at the bottom. Include the drawing name, drawing number, your name and the date. Draw thin guidelines for your lettering and labels. This is a freehand drawing. Practice keeping the lines straight, parallel, and even in thickness and value.

NO

15 Do not screen type (for print)

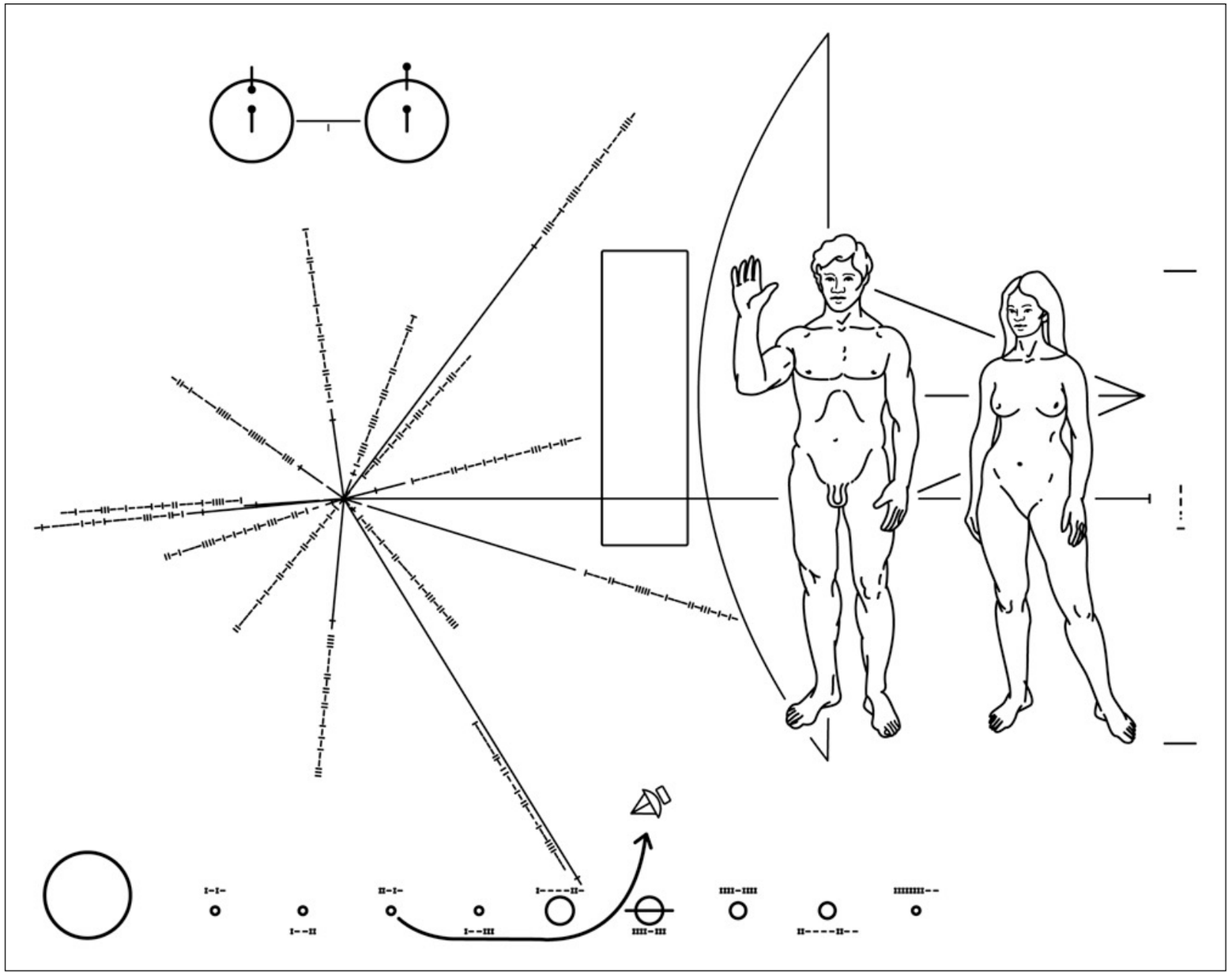
50% black (gray)
(as seen on screen)



50% black (gray)
(as printed on paper)

NO

16 Psychology of perception



http://en.wikipedia.org/wiki/Pioneer_plaque

16 Psychology of perception

Papers:

Counting But Losing Count:
the Legacy of Otto Neurath's Isotype Charts
bit.ly/2JlfkUo

The Four-Second Window
bit.ly/Ve2mph

The Double Constraints of Convention and Cognition in Successful Graphic Design
bit.ly/12zLinL

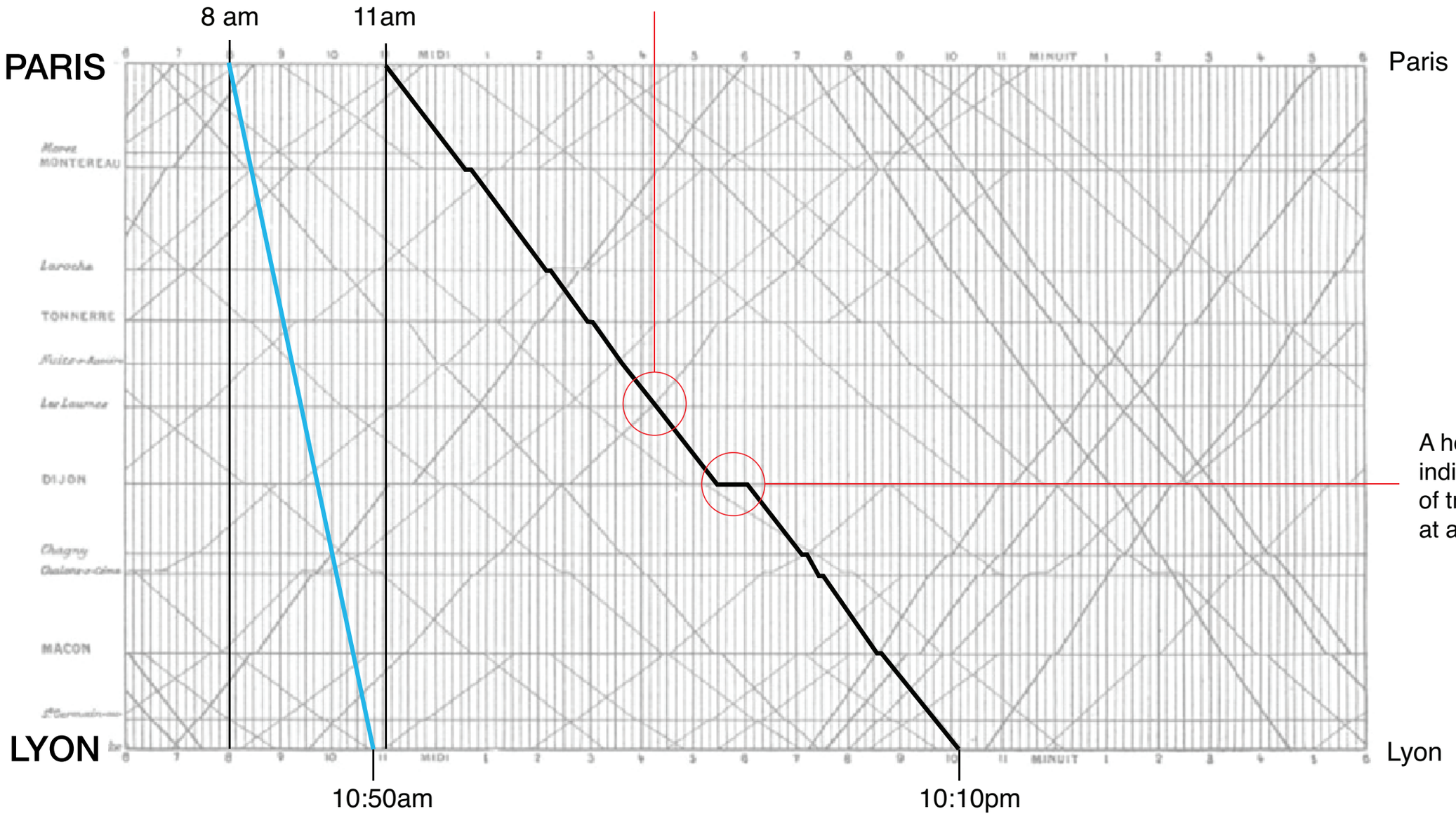
OTHER EXAMPLES:

LINE
BAR
TREEMAP
SCATTER

Line

The intersection of
two lines locates the time and
place that trains going in opposite
direction pass each other.

Graphical train schedule
Paris-Lyon, 1885

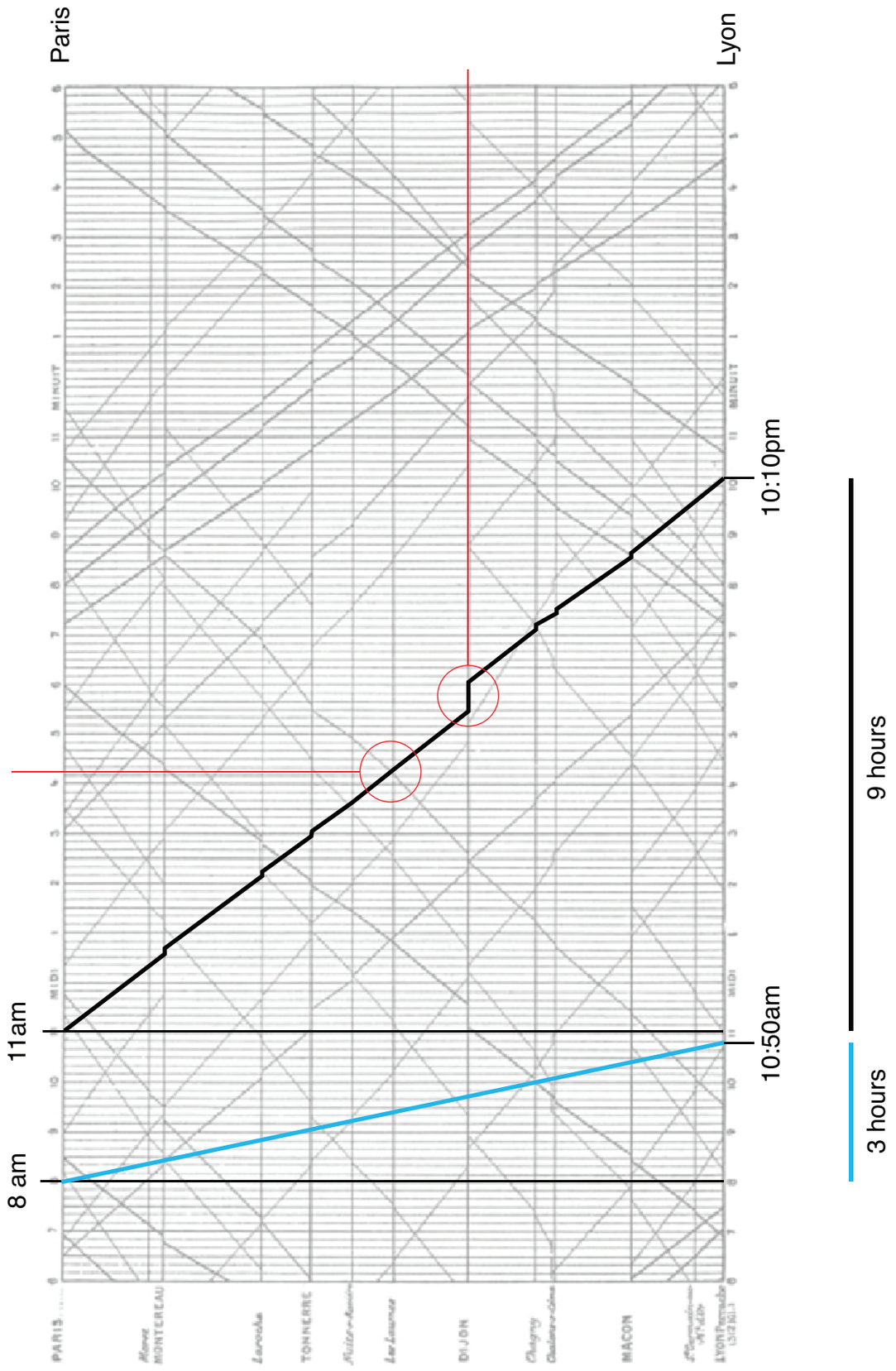


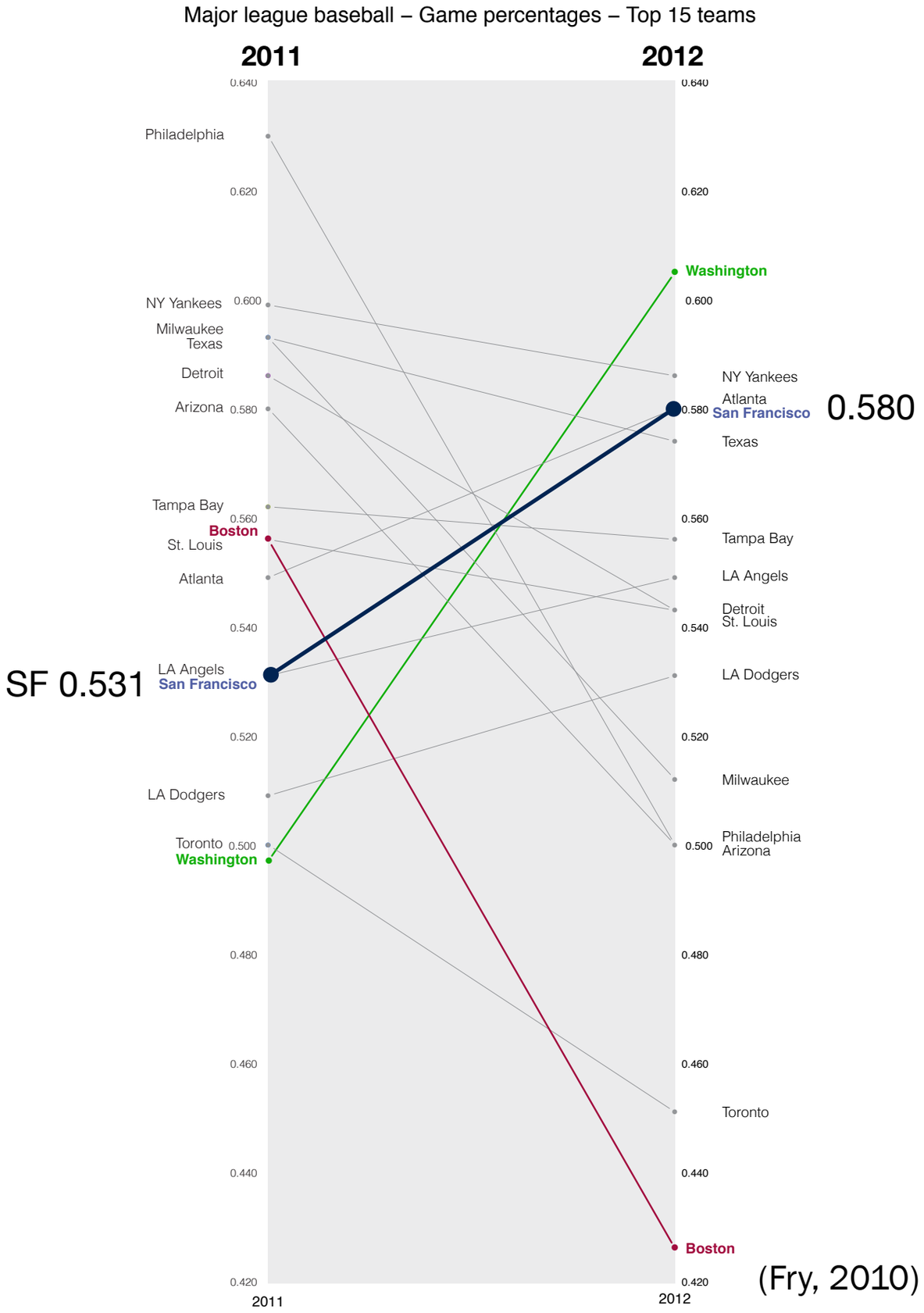
A horizontal line
indicates length
of train stop
at a station.

The cities between Paris and Lyon are spaced proportionally according to the distance between them. Arrival and departures are located along the horizontal. The slope of the lines reflect the speed of the train: the more vertical the line, the faster the train. Compare an express train which took 9 hours to complete the trip in 1885 (black line) with a TGV (*train à grand vitesse*) which took 3 hours in 1981 (blue line).

Graphic representation of a train schedule showing rate of travel along the route from Paris to Lyon. (The method is attributed to the French engineer Ibry)- Etienne-Jules Marey (1830-1904), France.

E.J. Marey, *La Méthode Graphique* (Paris, 1885). p. 20

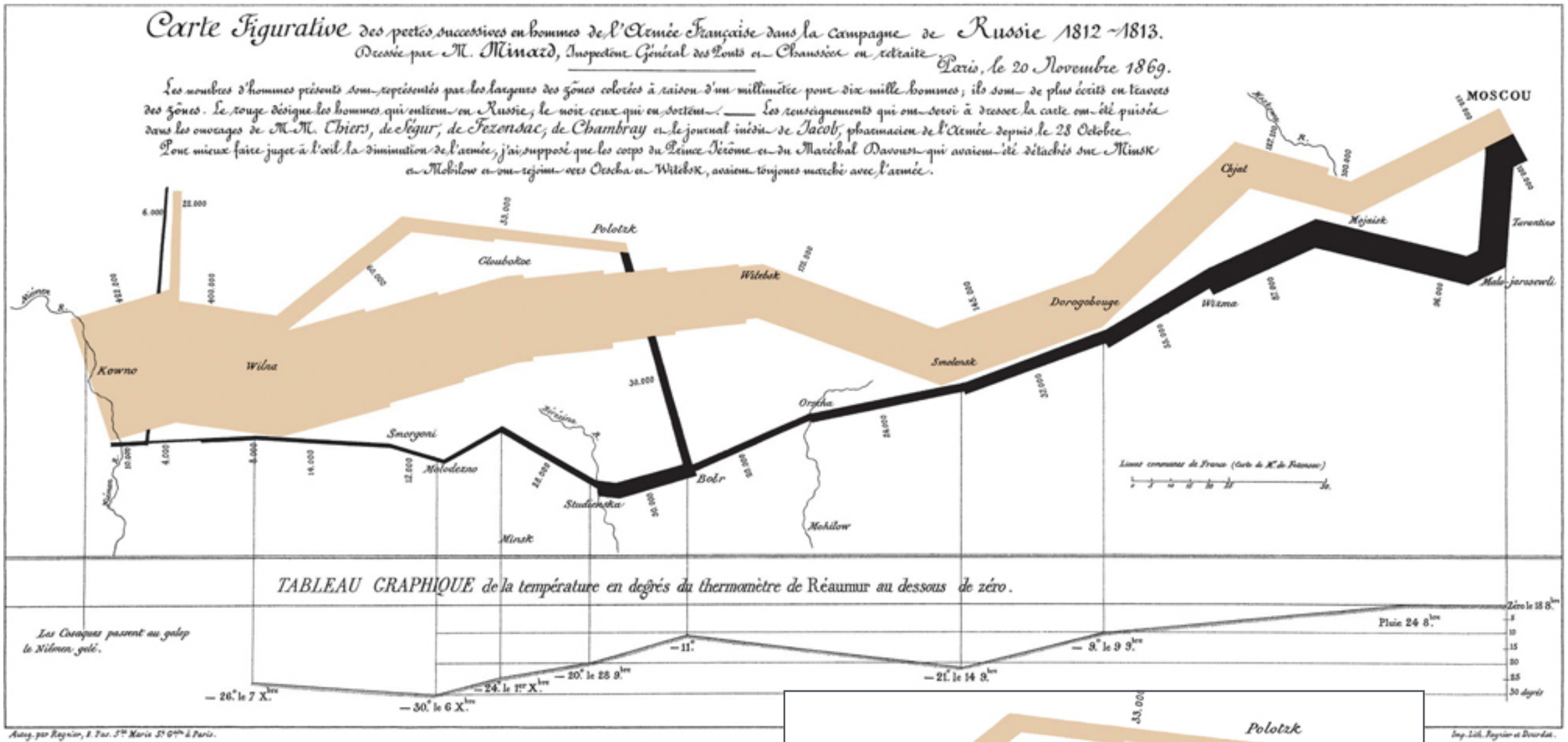

Paris
Lyon
10:10pm
11am
10:50am
8 am
9 hours
3 hours

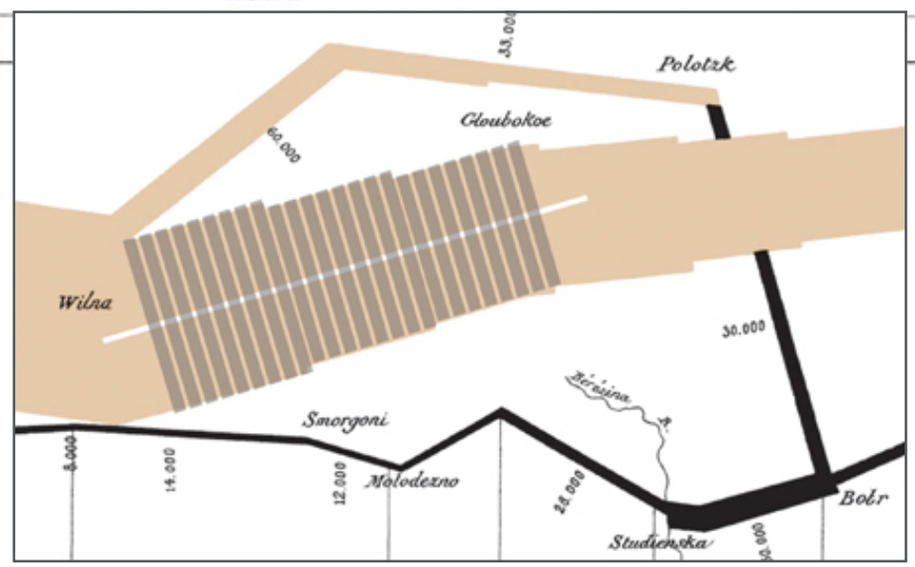

Major league baseball – Game percentages – Top 15 teams
2011
2012
0.640
0.620
0.600
0.580
0.560
0.540
0.520
0.500
0.480
0.460
0.440
0.420
Philadelphia
NY Yankees
Milwaukee
Texas
Detroit
Arizona
Tampa Bay
Boston
St. Louis
Atlanta
LA Angels
San Francisco
LA Dodgers
Toronto
Washington
SF 0.531
Washington
NY Yankees
Atlanta
San Francisco
0.580
Texas
Tampa Bay
LA Angels
Detroit
St. Louis
LA Dodgers
Milwaukee
Philadelphia
Arizona
Toronto
Boston
0.580
2011
2012

(Fry, 2010)

Line/Bar

Charles Minard 1781-1870





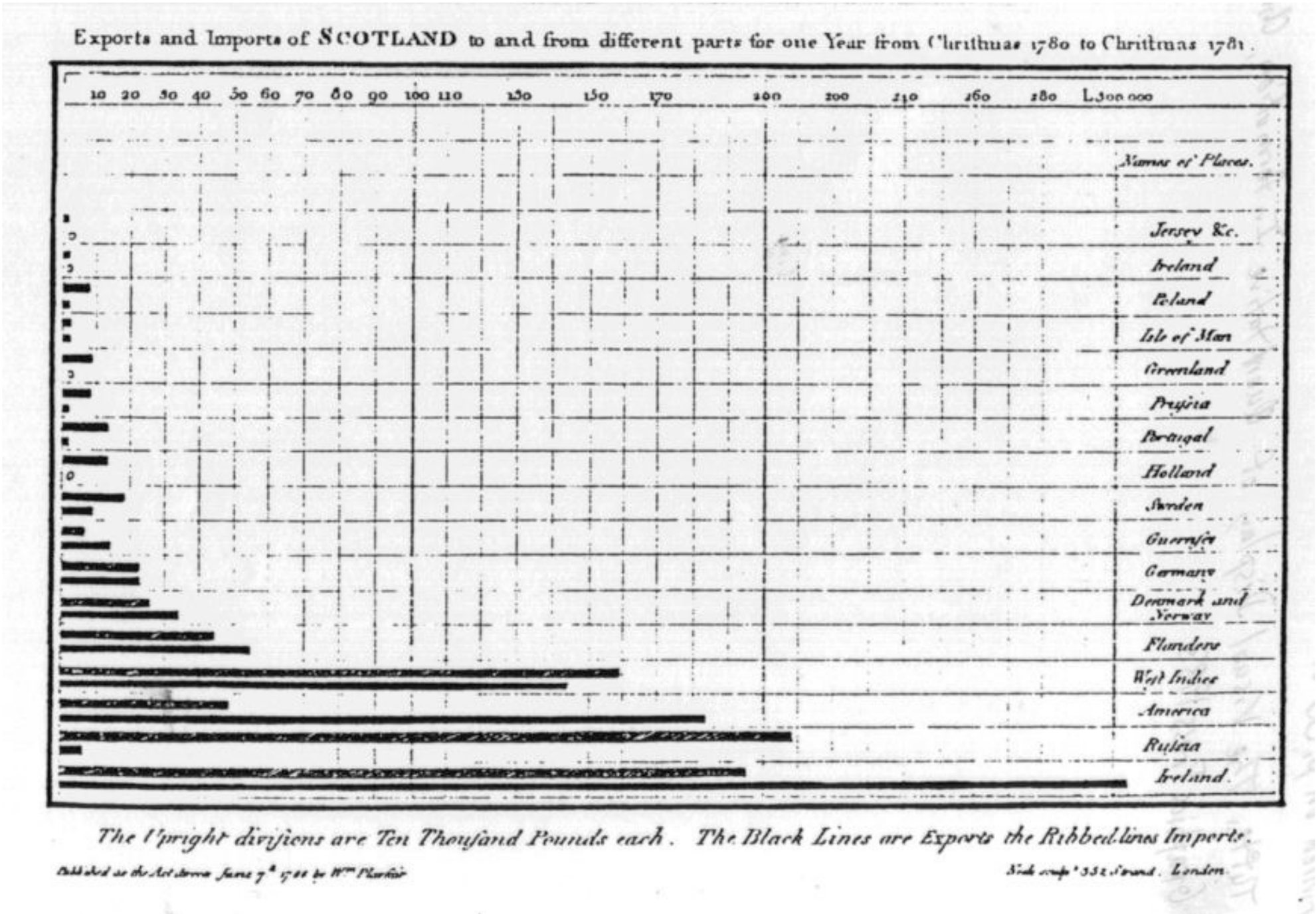
Minard's flow map graphic of Napoleon's March on Moscow (called “the best graphic ever produced” by Tufte).
The “thickness” of the line indicates the size of the army (from 442,000 in 1812 to 10,000 in 1813. See Tufte, p.40

http://www.math.yorku.ca/SCS/Gallery/re-minard.html

(Trogu, 2013)

Bar

William Playfair 1759-1823



First known bar chart (1786). Import and export to and from Scotland in 1781 for 17 countries.

Bar in D3

https://d3js.org

https://bost.ocks.org/mike/

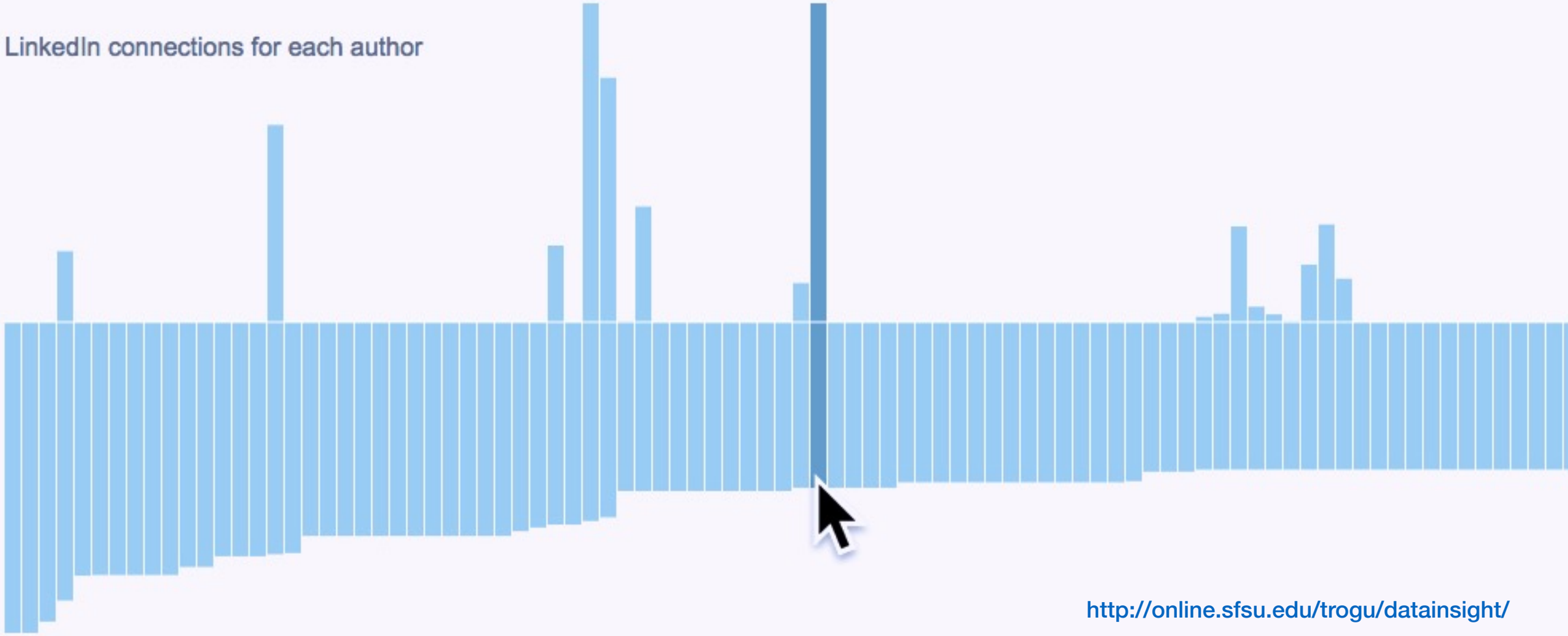
Academia is an Iceberg

Authors of the most read papers in biology who are also on LinkedIn

Move the cursor over the bars to see specific information about each author.

Jun Wang
Readers: **455**
LinkedIn connections: **500**
http://www.linkedin.com/pub/jun-wang/3/9aa/98a
BUT! There are **7** matching results from LinkedIn...

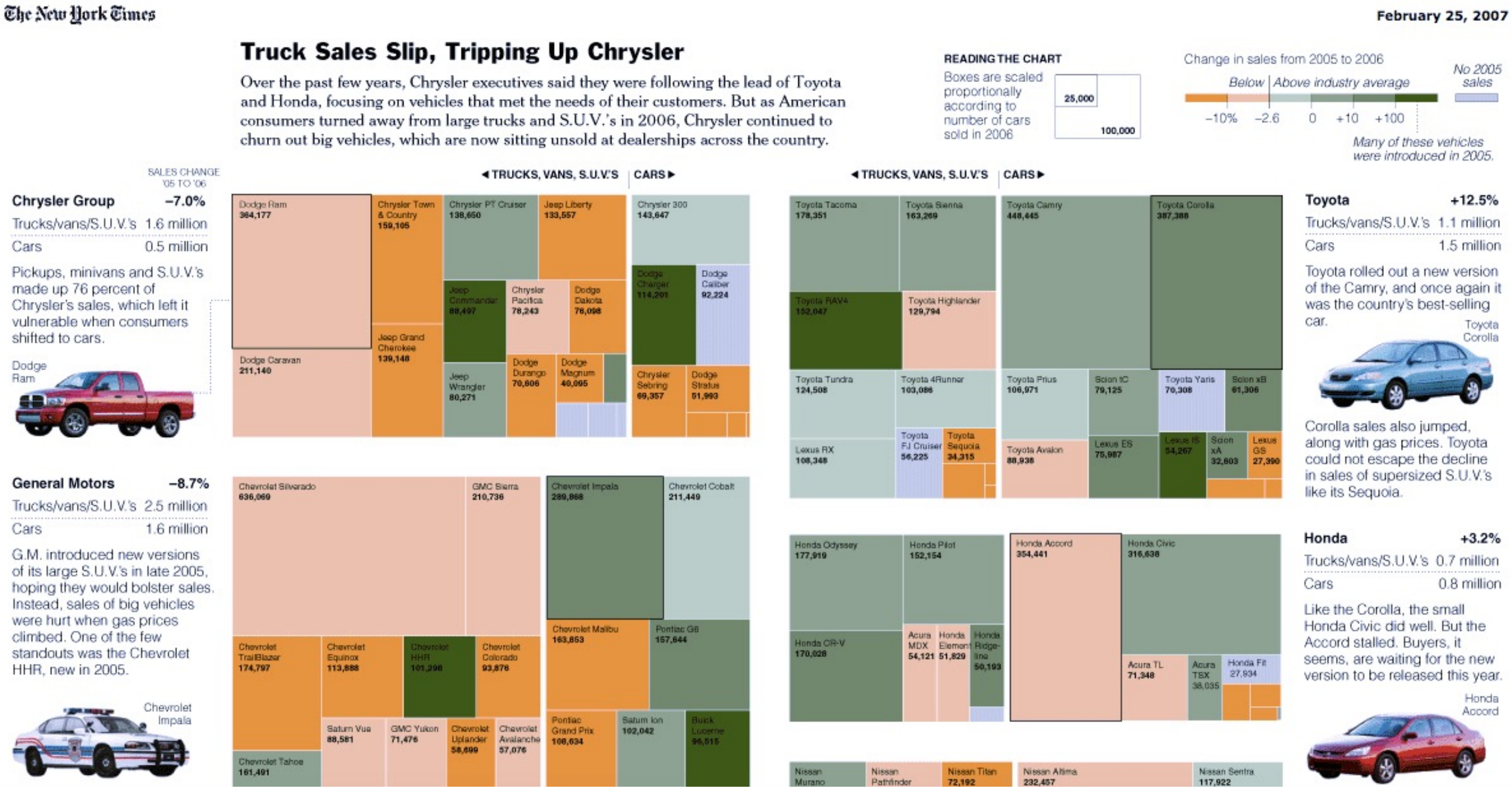
LinkedIn connections for each author



http://online.sfsu.edu/trogu/datainsight/

(Caviglia, Gunn & Trogu, 2011)

Treemap



(Amand Cox, The New York Times, 2007)

https://archive.nytimes.com/www.nytimes.com/imagepages/2007/02/25/business/20070225_CHRYSLER_GRAPHIC.html

Coordinates system

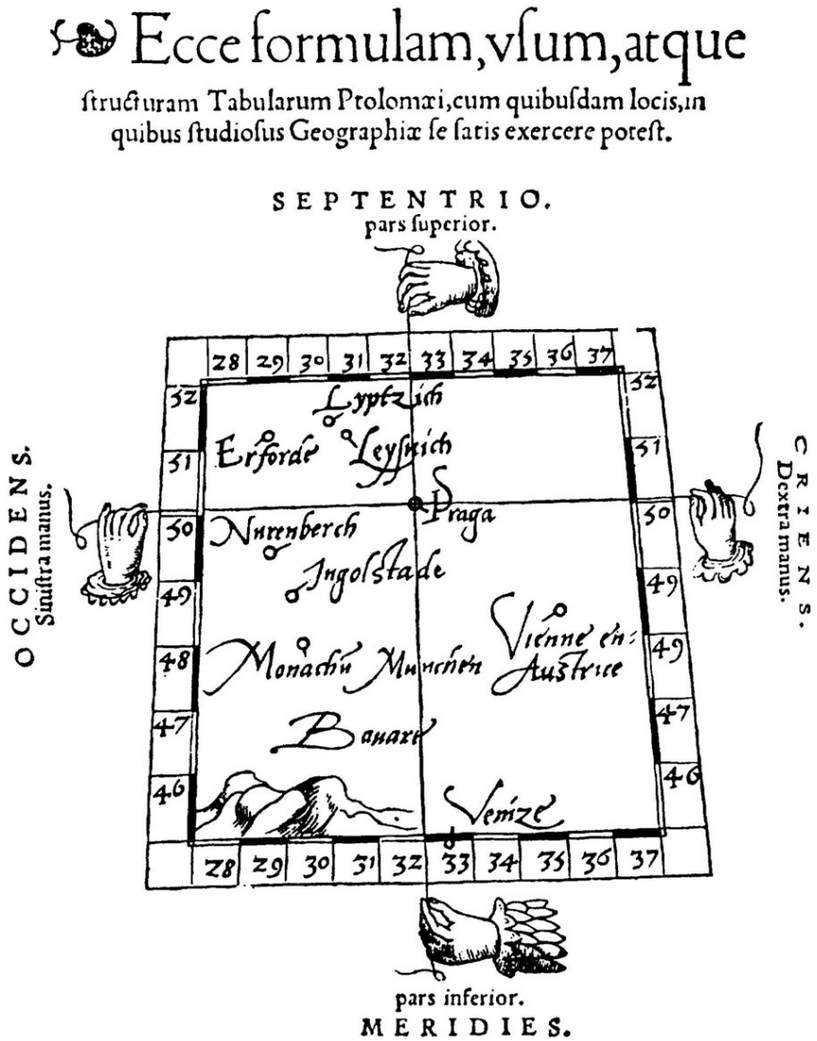
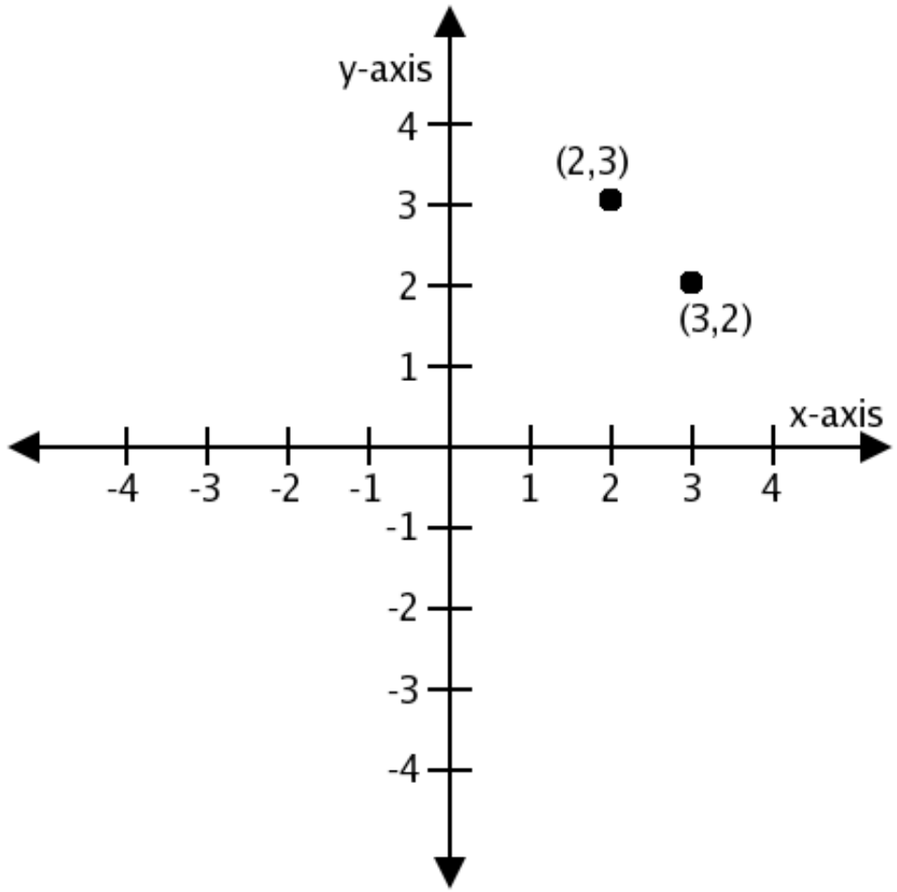


Illustration from Apianus' 1546 edition of *Cosmographia*.

Descartes 1596-1650



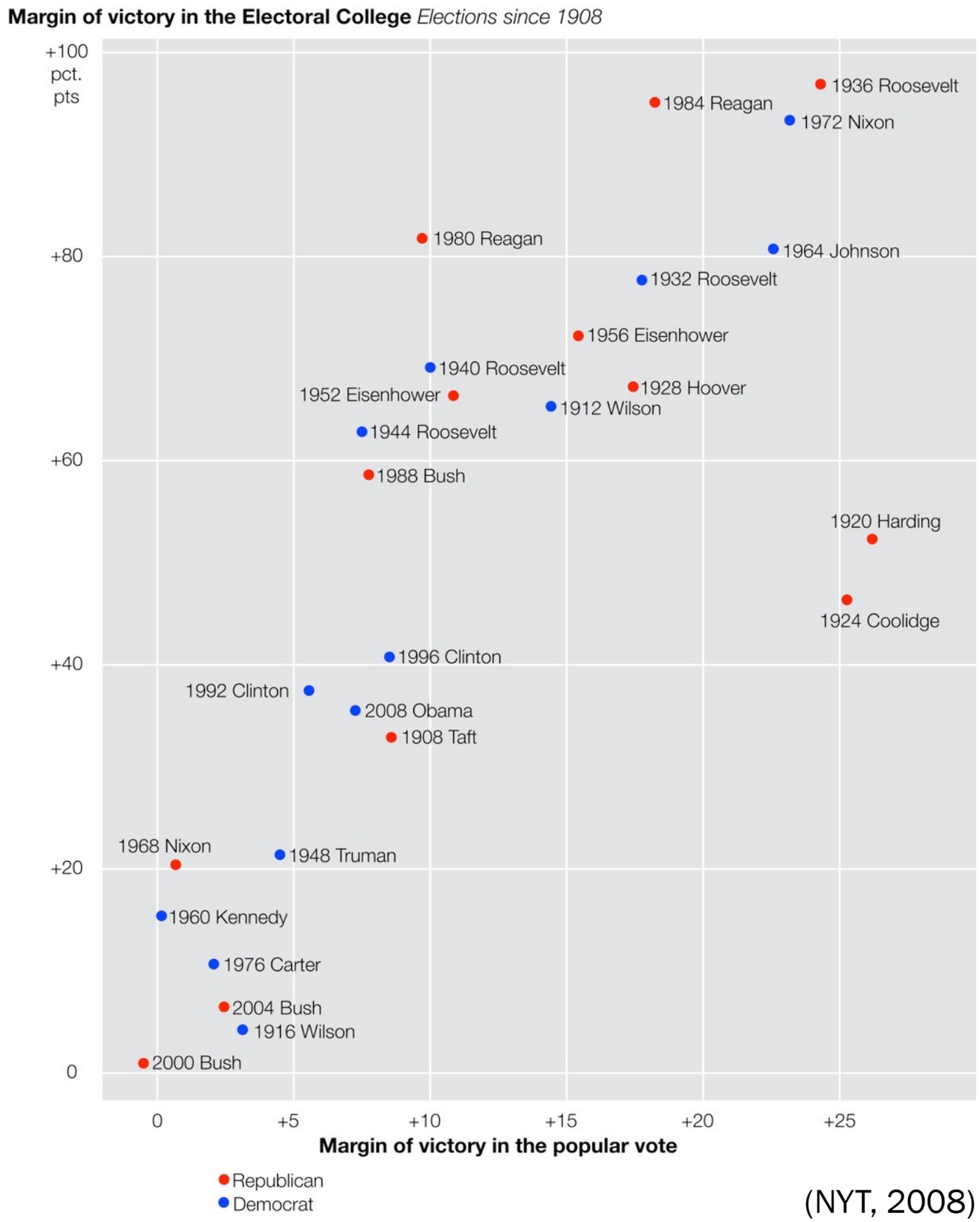
Cartesian coordinates system (1637)

Scatterplot

More information about US presidential elections.
[Election map](). (NY Times)

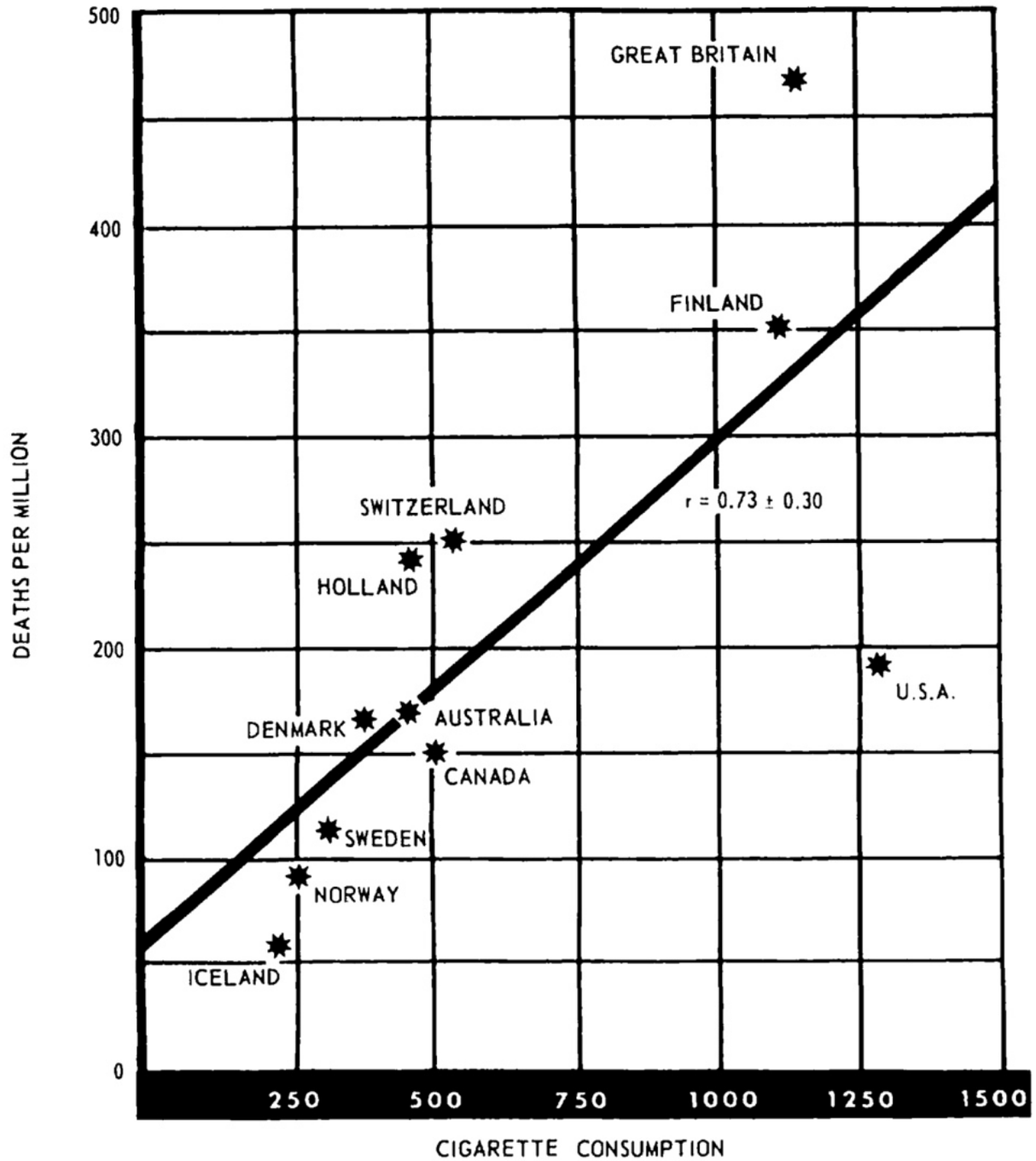
Tabular data used for the scatterplot at left (margin of victory in percentage points).

Winner	pop.vote (%)	Elec.College (%)
2008 Obama	7.27	35.6
2004 Bush	2.46	6.5
2000 Bush	-0.51	1
1996 Clinton	8.51	40.8
1992 Clinton	5.56	37.6
1988 Bush	7.72	58.6
1984 Reagan	18.21	95.2
1980 Reagan	9.74	81.8
1976 Carter	2.06	10.6
1972 Nixon	23.15	93.5
1968 Nixon	0.7	20.4
1964 Johnson	22.58	80.6
1960 Kennedy	0.17	15.6
1956 Eisenhower	15.4	72.4
1952 Eisenhower	10.85	66.4
1948 Truman	4.48	21.5
1944 Roosevelt	7.5	62.8
1940 Roosevelt	9.96	69.2
1936 Roosevelt	24.26	97
1932 Roosevelt	17.76	77.8
1928 Hoover	17.41	67.2
1924 Coolidge	25.22	46.3
1920 Harding	26.17	52.2
1916 Wilson	3.12	4.4
1912 Wilson	14.44	65.3
1908 Taft	8.53	33



(NYT, 2008)

Scatterplot



Report of the Advisory Committee to the Surgeon General. *Smoking and Health* (Washington. D.C., 1964), p. 176; based on R. Doll, "Etiology of Lung Cancer," *Advances in Cancer Research*, 3 (1955). 1-50.

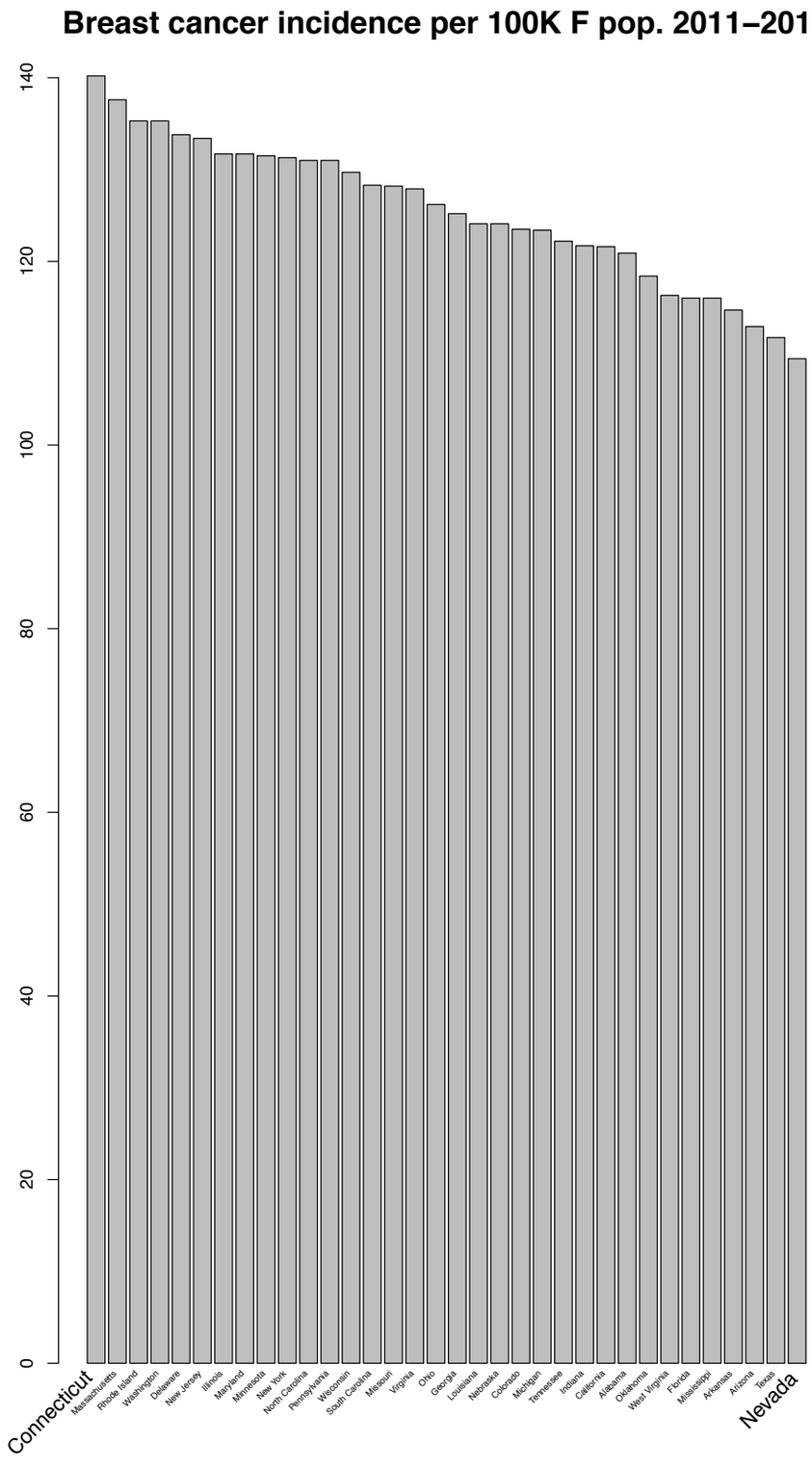
From *The Visual Display of Quantitative Information* Tufte, p.46-47.

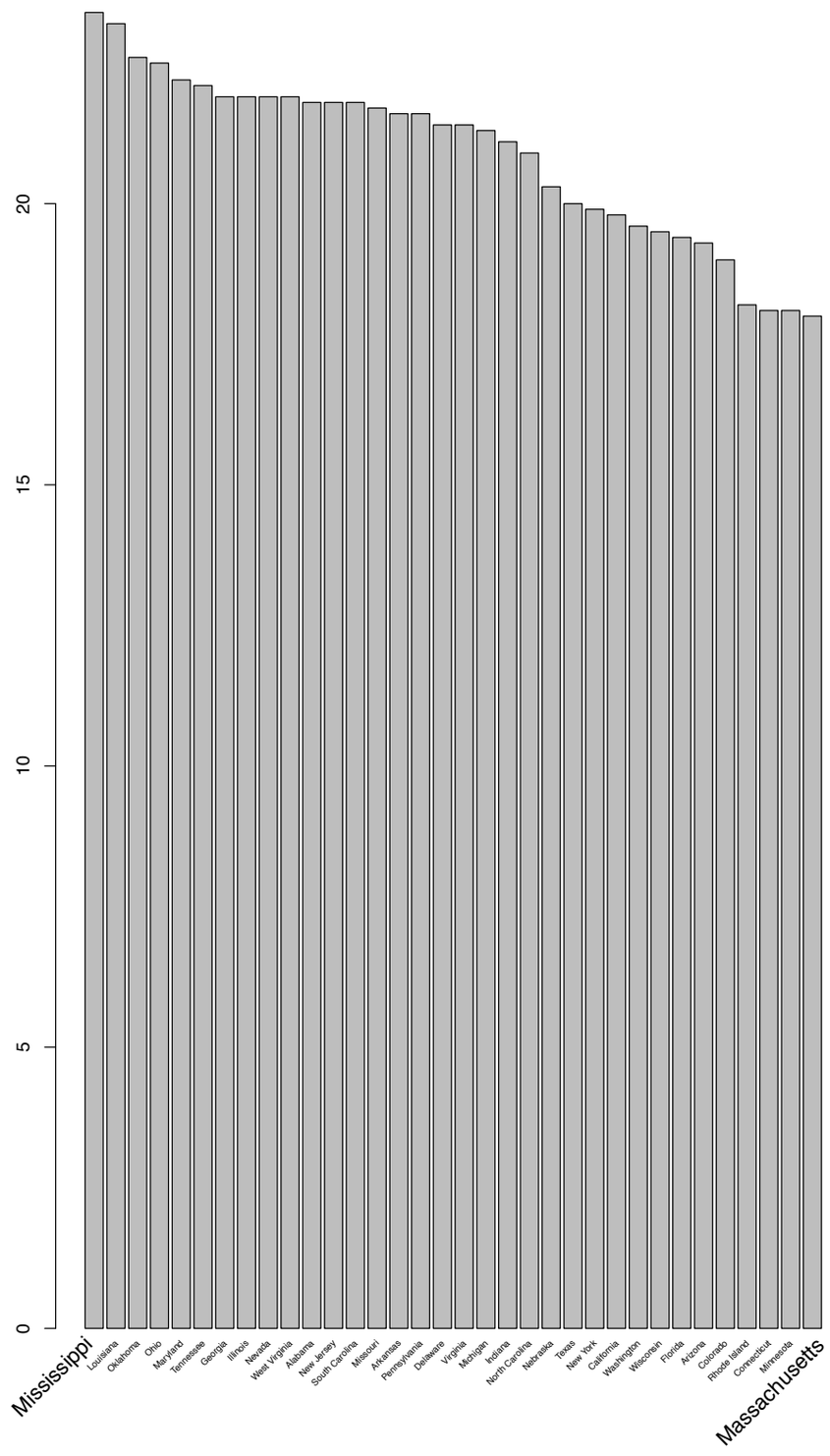
Scatterplot

Hans Rosling – *Factfulness*

https://www.gapminder.org https://www.gapminder.org/tools/#$chart-type=bubbles https://www.ted.com/talks/hans_rosling_at_state

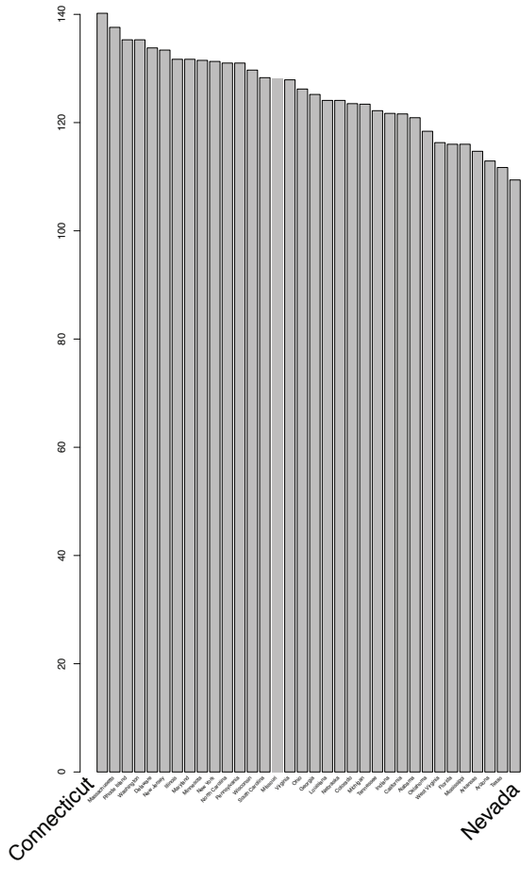
Bars to Scatterplots





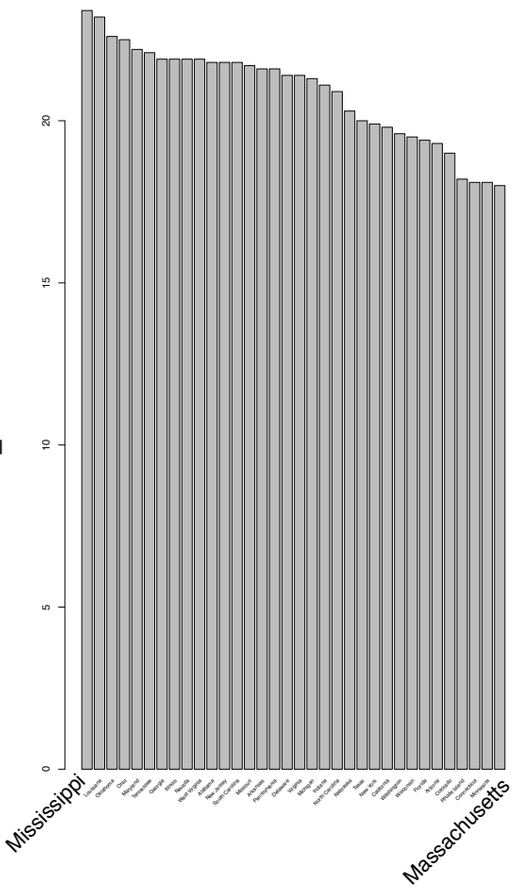
Bars to Scatterplots

Breast cancer incidence per 100K F pop. 2011–2015



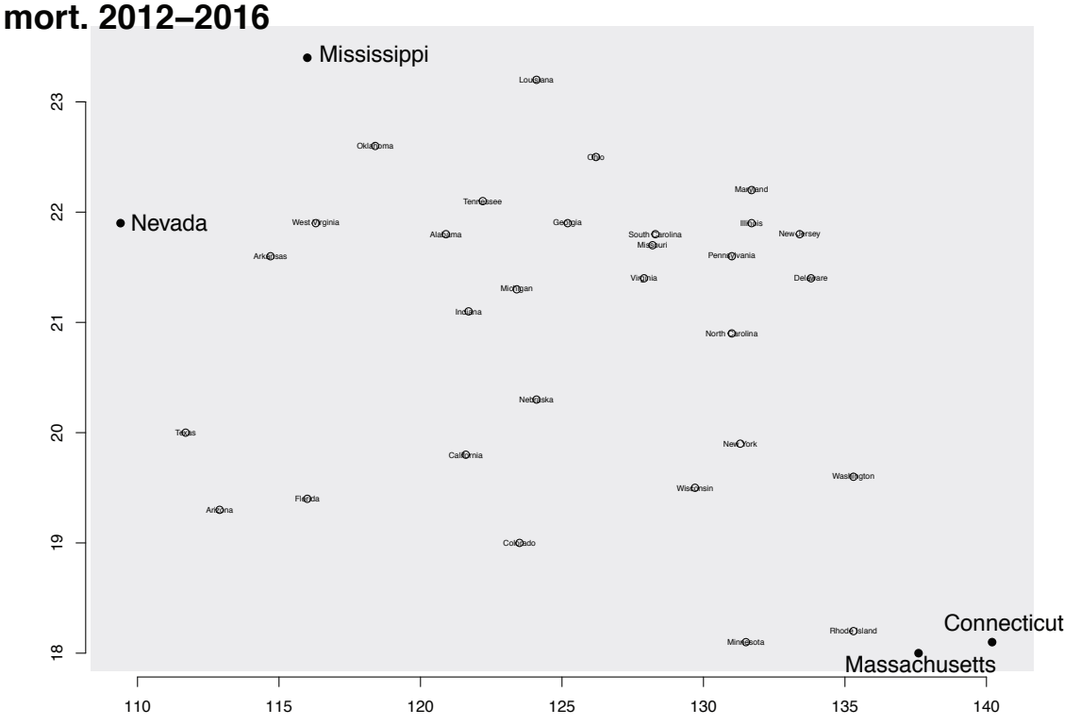
+

Breast cancer mortality per 100K F pop. 2012–2016



=

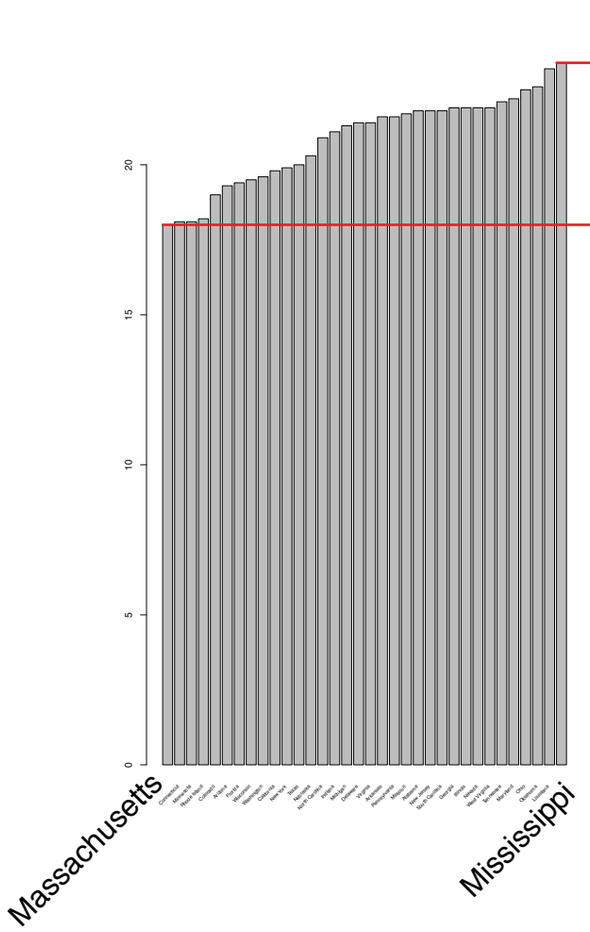
Breast cancer mort. 2012–2016

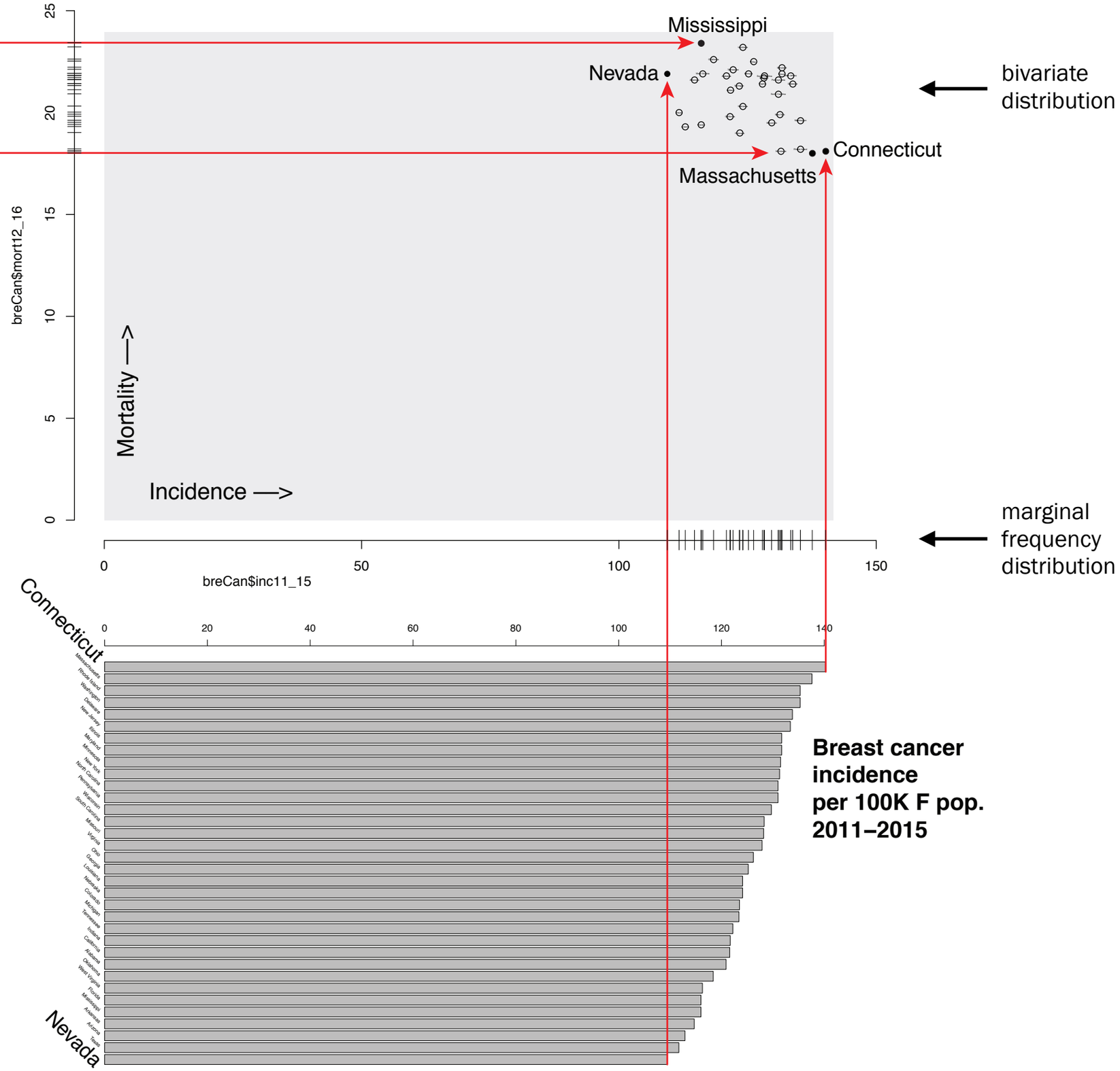


(Only 34 states shown)

Source: National Cancer Institute

Breast cancer mort. per 100K F. pop 2012–2016

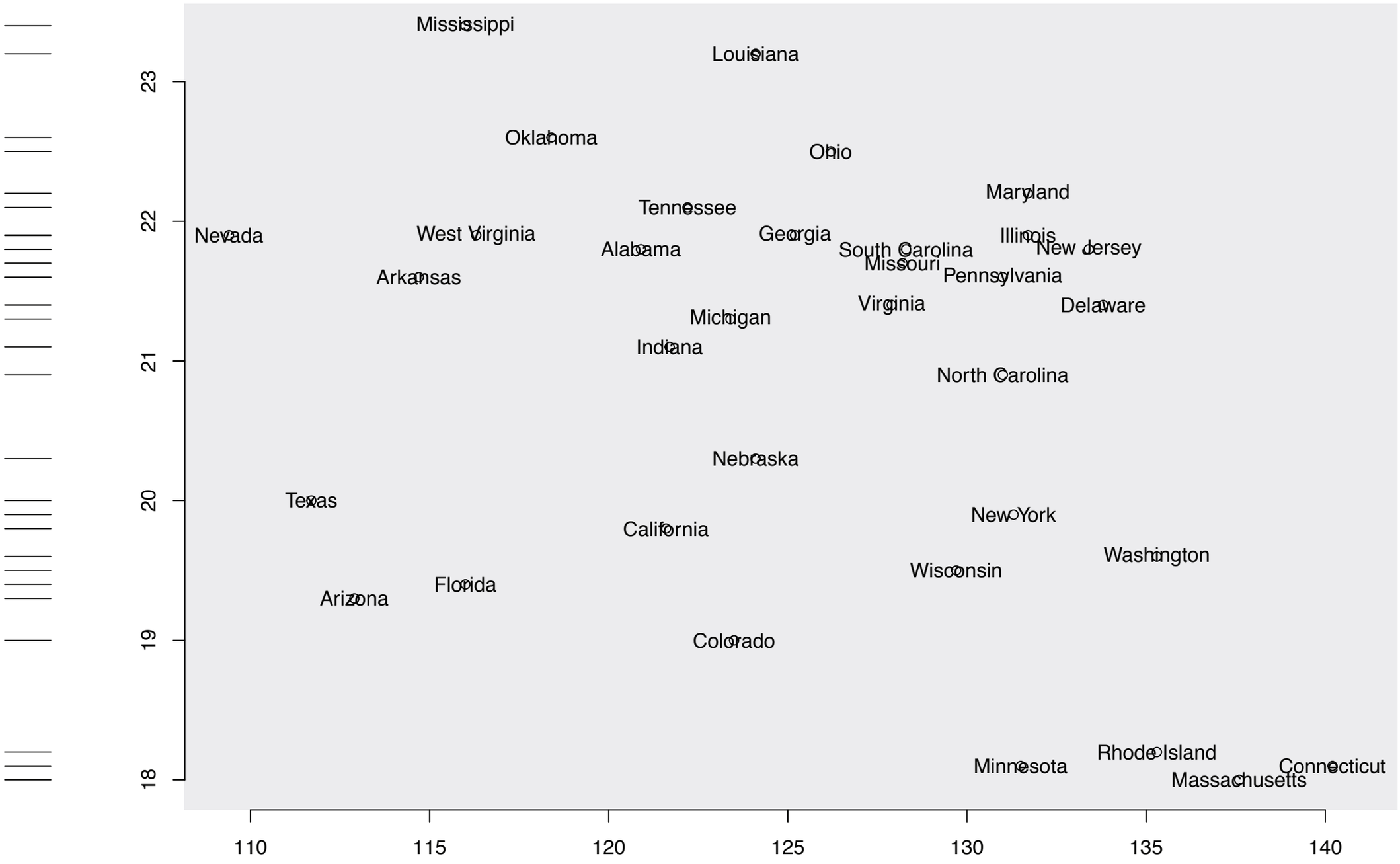




Breast cancer incidence per 100K F pop. 2011–2015

(Only 34 states shown)
Source: National Cancer Institute

Breast cancer mort. 2012–2016

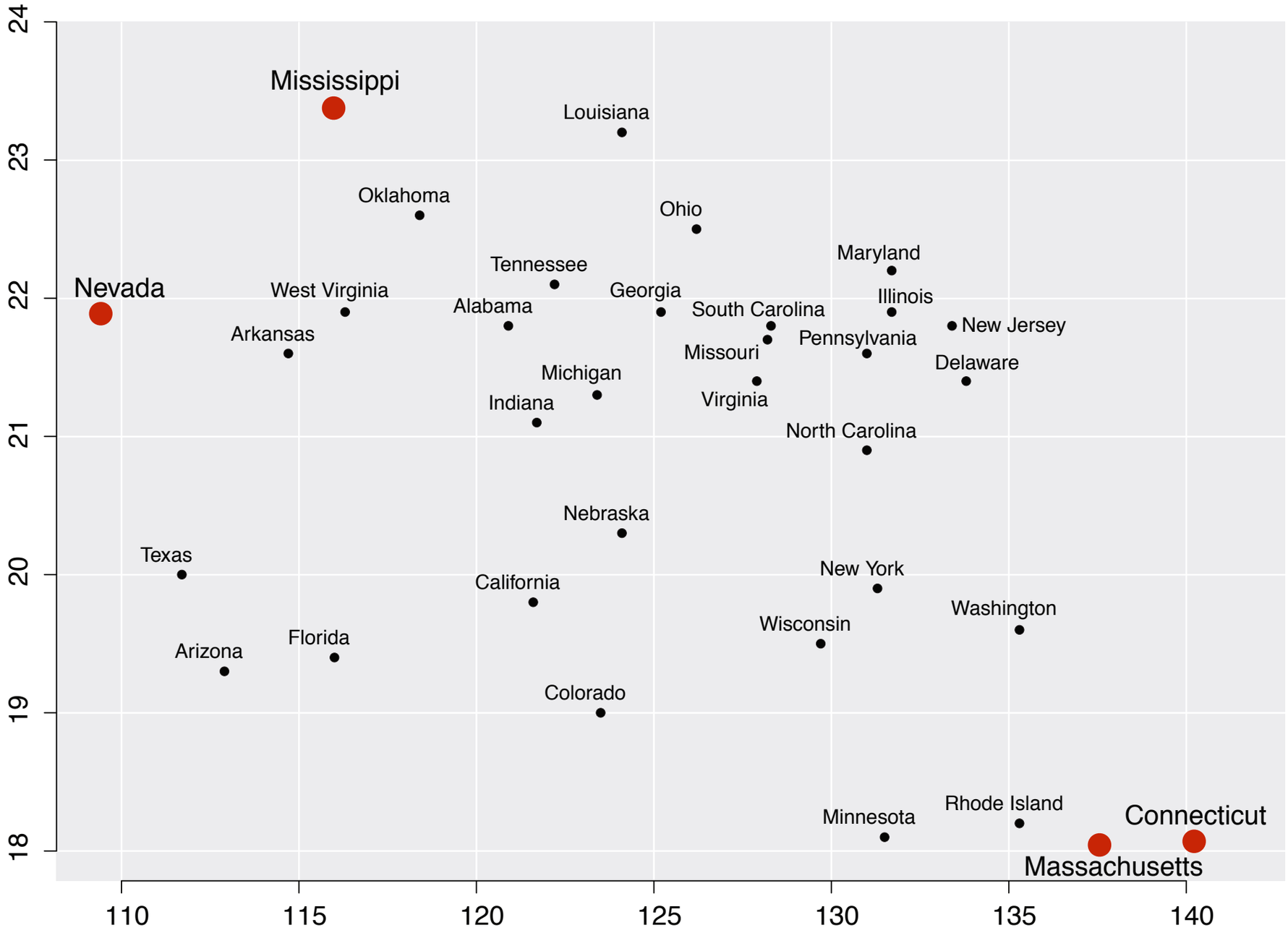


Source: National Cancer Institute

Breast cancer inc. 2011–2015

(Only 34 states shown)

Breast cancer mortality
per 100K F. pop.
2012–2016



Source: National Cancer Institute

Breast cancer incidence
per 100K F. pop. 2011–2015

(Only 34 states shown)

Thank You!

Pino Trogu
trogu@sfsu.edu

Survey questions (after)

Small Handbook of Information Design:
16 Principles for Better Data Visualizations

Booklet and online version
of small handbook:
bit.ly/1bhDU4y

go to first slide