descriptive statistics 2: graphs

Adam Okulicz-Kozaryn
adam.okulicz.kozaryn@gmail.com

this version: Tuesday 13th February, 2018   09:55

outline

misc

intuition

howto: some theory (Tufte)

the good, the bad, and the ugly

<u>outline</u>

misc

intuition

howto: some theory (Tufte)

the good, the bad, and the ugly

ps2

◇ let's quickly discuss the right answer

ps1

$\diamond$ most of you had pretty good ideas

$\diamond$ important–strat woring on them asap!

$\cdot$ just start writing!

$\cdot$ (can do free-writing, dump all the ideas on paper, worry later about organization)

$\diamond$ give me early drafts, discuss with

$\diamond$ again, 2 keys to succes: start early, ask questions!!

$\diamond$ keep ps1 with my comments, and see me one-on-one

$\cdot$ and lets discuss it more!

ps1–last year

◇ nobdy discussed specific data and literature

◇ narrow down, be focused on sth specific,

· can always make it more elaborate if time and resorces permit

◇ be specific, eg how would you measure 'fairness'

· measurment is the key!

◇ email me about data!

◇ great to kill 2 birds with one stone: internship, etc

· and study something you are passionate about!

outline

misc

## intuition

howto: some theory (Tufte)

the good, the bad, and the ugly

graphics and descriptive statistics

$\diamond$ i was emphasizing importance of understanding your data

$\diamond$ graphics are a great way to visualize/understand data

$\diamond$ data are numbers, usually many and in a matrix

$\cdot$ graphics is a great tool to allow humans to comprehend those many numbers

$\diamond$ have graphs in final presentation

$\diamond$ again, quality matters not quantity!

$\diamond$ eg for final presentation: several great graphs would do!

tell a story

$\diamond$ can tell a whole story with a graph

· just make it little complex and sophisticated

· the key is idea, not statistics!

· tell us something surprising,
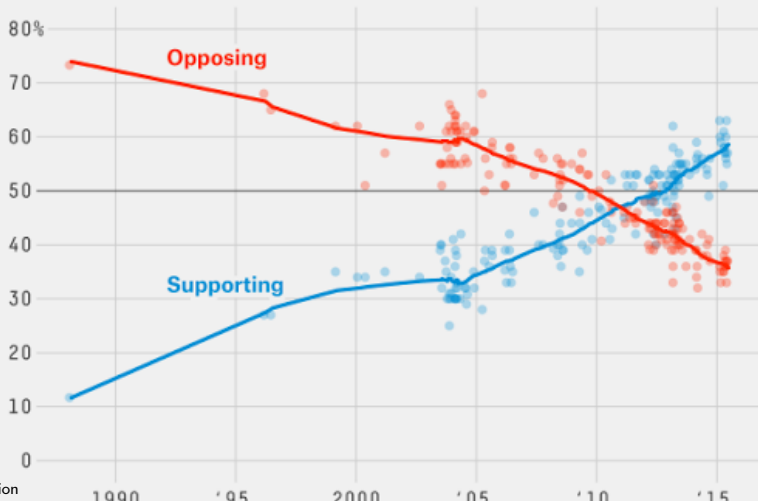 something overlooked

· examples follow

wh/bl lexp: 1918 dip, droping swings, closing gap

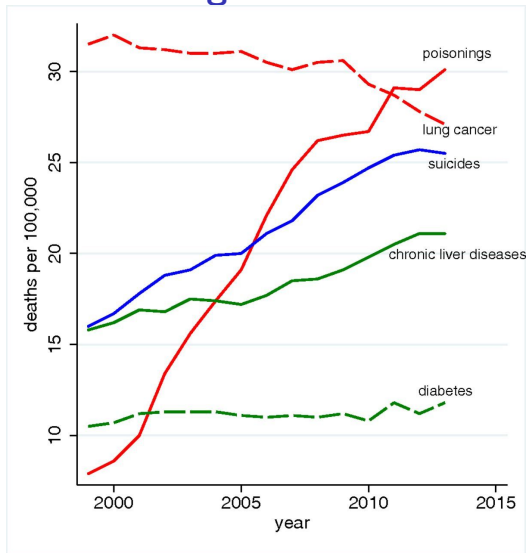Life expectancy, white males — Life expectancy, black males

◇
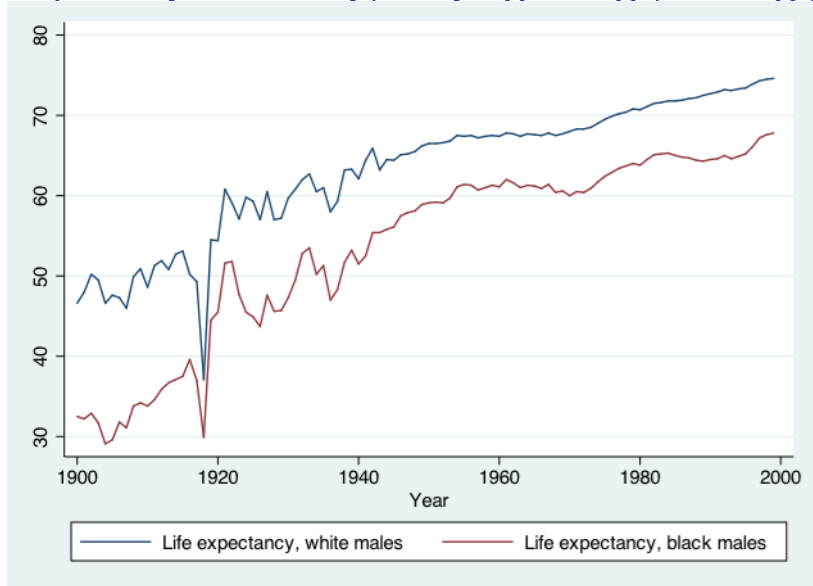
monotonic, but sped up in 05!

Steady Trend On Gay Marriage Views
Percentage opposing and supporting in national, live-interviewer polls since 1988

## whites killing themselves recently

red state, blue state, rich state, poor state

◇ even the whole book can be based on a graph

◇ http://amzn.com/0691143935

◇ can tell a complicated story with a graph!

◇ we know that the richer you are, the more likely you are to vote for ...

◇ but the richer the area, the weaker the relationship

◇ in other words, poor folks just vote Democrat everywhere

◇ rich, on the other hand, vary depending on area income

McCain vote by income in a poor, middle-income, and rich state

Miss.

Ohio

Conn.

Probability of voting for McCain

75%

50%

25%

(poor)

(rich)

Voter's income
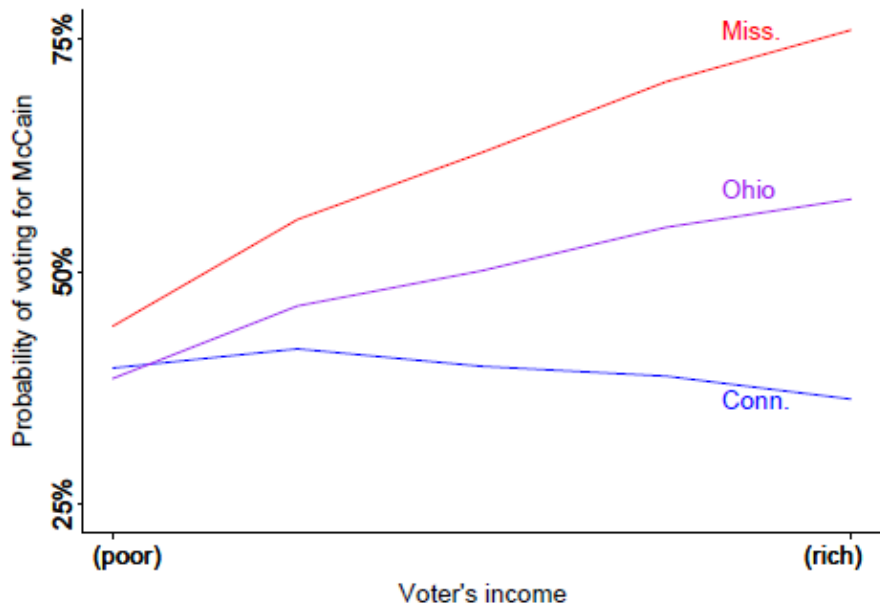
interested in inequality ?

$\diamond$ a book using many graphs

$\cdot$ most of the argument rests on graphs

http://amzn.com/1608193411

<u>outline</u>

misc

intuition

howto: some theory (Tufte)

the good, the bad, and the ugly

links

$\diamond$ Tufte (multiple) `http://www.edwardtufte.com/tufte/`
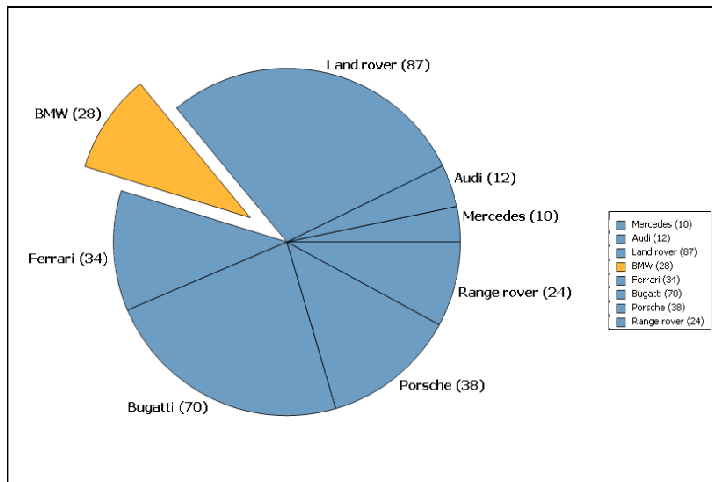
$\diamond$ Kosslyn "Clear and to The Point"

`http://www.amazon.com/Clear-Point-Psychological-`

`Principles-Presentations/dp/0195320697`

avoid visual clutter

◇ all parts of graph should be meaningful

◇ good practices:

· do not use shades

· do not use fancy colors

· do not use any decoration

chartjunk
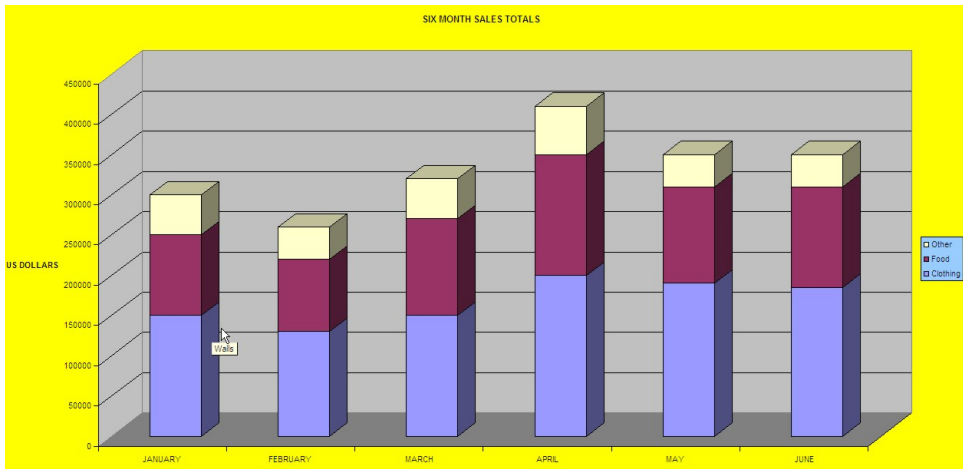
chartjunk

Exploded Pie Slice Chart

other chartjunk examples

$\diamond$ https://www.google.com/search?q=fancy+charts&tbm=isch

$\diamond$ it's all over the place–eg popular media showing oil production as barells instead of bars in a chart

· is the height of barrel meaningful or its volume or what?

not chartjunk (the economist)

Extra leisure time enjoyed by men compared with women
2006*, minutes per day

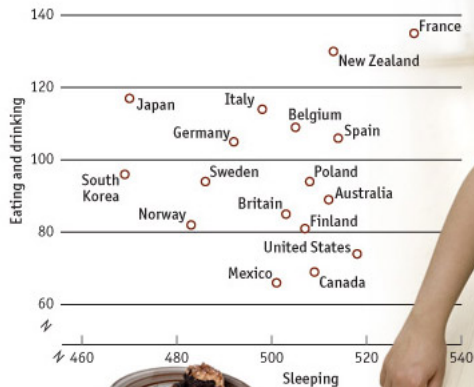|	0	20	40	60	80
Italy					
Poland					
Mexico					
Spain					
Belgium					
United States					
France					
South Korea					
Britain					
Turkey					
Australia					
Finland					
Canada					
Germany					
Sweden					
Japan					
New Zealand					
Norway					

Source: OECD *Or latest available

not chartjunk (the economist)

Time spent eating and sleeping
2006, minutes per day

Eating and drinking

140 — France
New Zealand
120 — Japan Italy Belgium Spain
Germany
100 — South Korea Sweden Poland Australia
Norway Britain Finland
80 — United States
Mexico Canada
60 —

460 480 500 520 540
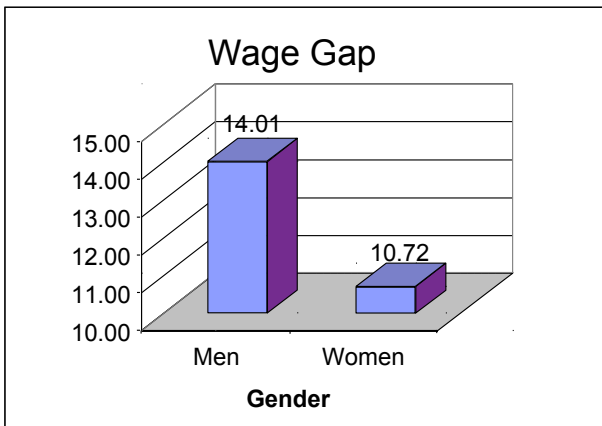Sleeping

Source: OECD

<u>outline</u>

misc

intuition

howto: some theory (Tufte)

the good, the bad, and the ugly

intuition

◇ there are bad graphs: chart junk

◇ yet there are ugly graphs: graphs that not only confuse but also lie
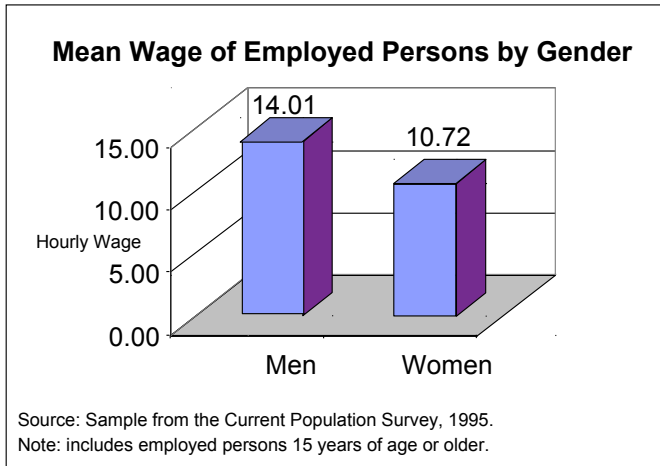
the ugly 1

Wage Gap

15.00	14.01
14.00	
13.00	
12.00	10.72
11.00	
10.00	

Men   Women

Gender

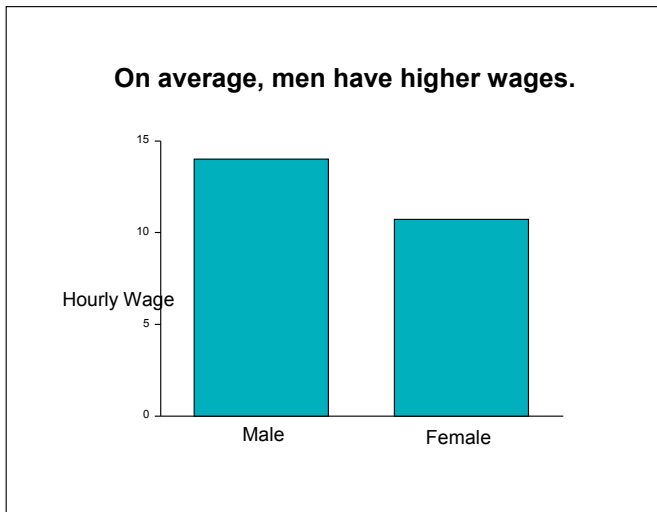### problems

◇ 3d

◇ scale not at 0 (it depends how you describe it in words, too)

◇ what is Y axis' measurement

◇ caption exactly explaining data used: eg part-time, when, etc

◇ data source

the chart junk

Mean Wage of Employed Persons by Gender

Source: Sample from the Current Population Survey, 1995.
Note: includes employed persons 15 years of age or older.

the good (there is virtue insimplicity, but)

On average, men have higher wages.

Hourly Wage

the good (it is revealing to dig deeper)

Controlling for Age Changes the Picture

■ Men's Wages   ■ Women's Wages

Hourly Wage

Age Bracket: 15-24, 25-34, 35-44, 45-54, 55-64, 65+

other misleading examples

◇ you have time series data and you only show a period of time that supports your story

◇ most things can be measured in many ways...

· you just present one measure that supports your story, eg:

· Obama's state of the union address: unempl is down

· but did not mention that lab force participation is down

· and underemployment is up

· not just Obama, most politicians present only data supporting their story

don't be misleading !

$\diamond$ at the very minimum you should discuss contradicting results in a footnote/appendix

$\diamond$ mention also major limitations somewhere

conclusion

◇ know your data

◇ you can understand it by producing descriptive statistics

◇ again for more see Tufte and Kosslyn

if time: discuss alcohol consumption

$\diamond$ but first discuss 2 assigned economist articles

$\diamond$ vast majority of Americans do not drink much

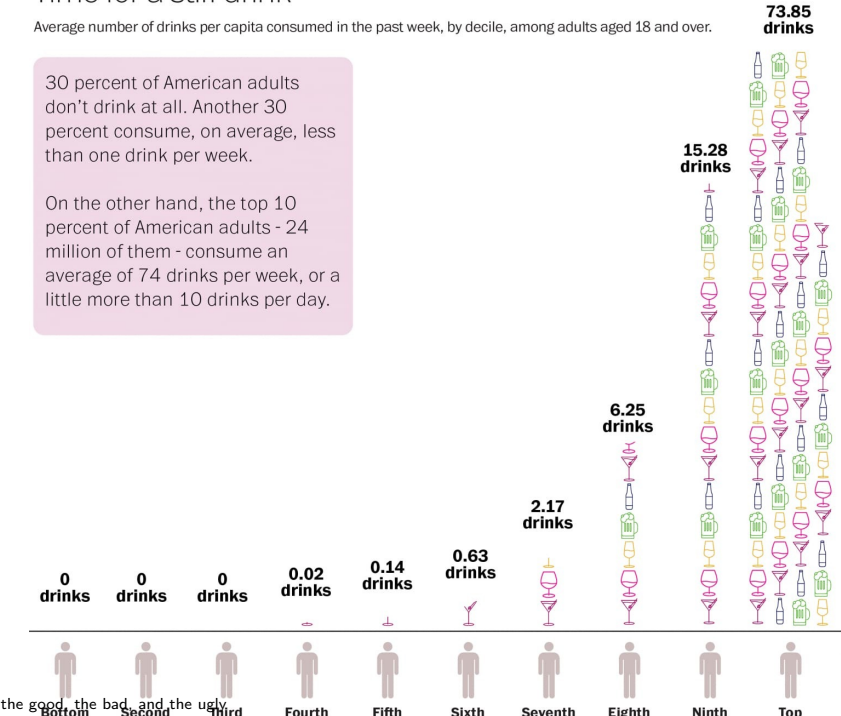$\diamond$ but those that do drink much, they drink a lot!

https://www.washingtonpost.com/news/wonk/wp/2014/09/25/think-you-drink-a-lot-

this-chart-will-tell-you

Time for a stiff drink

Average number of drinks per capita consumed in the past week, by decile, among adults aged 18 and over.

30 percent of American adults don't drink at all. Another 30 percent consume, on average, less than one drink per week.

On the other hand, the top 10 percent of American adults - 24 million of them - consume an average of 74 drinks per week, or a little more than 10 drinks per day.

0 drinks — Bottom
0 drinks — Second
0 drinks — Third
0.02 drinks — Fourth
0.14 drinks — Fifth
0.63 drinks — Sixth
2.17 drinks — Seventh
6.25 drinks — Eighth
15.28 drinks — Ninth
73.85 drinks — Top

the good, the bad, and the ugly

bibliography I

OKULICZ-KOZARYN, A. (2011a): "City Life: Rankings (Livability) Versus Perceptions (Satisfaction)," Social Indicators Research, 110, 433–451.

——— (2011b): "Europeans Work to Live and Americans Live to Work (Who is Happy to Work More: Americans or Europeans?)," Journal of Happiness Studies, 12, 225–243.

——— (2013): "Cluttered writing: adjectives and adverbs in academia," Scientometrics, 1–3.

——— (2016): "Unhappy metropolis (when American city is too big)," Cities.

OKULICZ-KOZARYN, A. AND J. M. MAZELIS (2016): "More Unequal In Income, More Unequal in Wellbeing," Social Indicators Research.

WHEELAN, C. (2013): Naked statistics: stripping the dread from the data, WW Norton & Company.