

descriptive statistics 2: graphs

Adam Okulicz-Kozaryn

`adam.okulicz.kozaryn@gmail.com`

this version: Wednesday 25th September, 2024 07:48

outline

misc

intuition

howto: some theory (Tufte)

the good, the bad, and the ugly

outline

misc

intuition

howto: some theory (Tufte)

the good, the bad, and the ugly

outline

misc

intuition

howto: some theory (Tufte)

the good, the bad, and the ugly

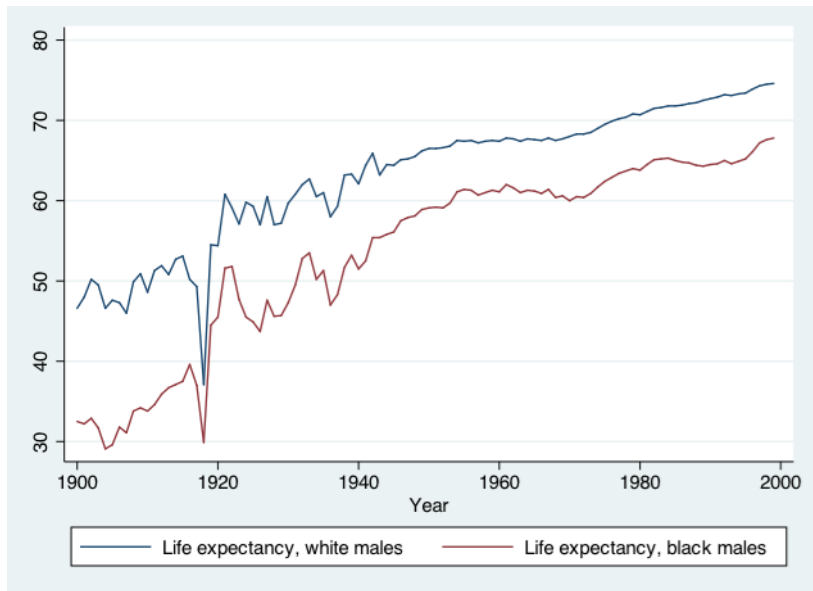
graphics and descriptive statistics

- i was emphasizing importance of understanding your data
- to visualize/understand data, graphics are the best!
- data are numbers, usually many
 - use graphics to comprehend those many numbers
 - (if few numbers, like 3 or so, just say them)

tell a story

- can tell a whole story with a single graph
 - just make it little complex and sophisticated
 - the key is idea, not statistics!
 - tell us something surprising, something overlooked
 - examples follow

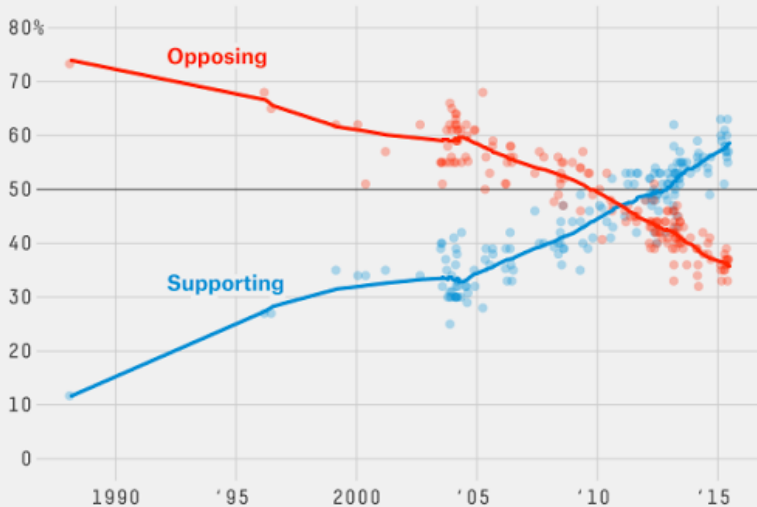
wh/bl lexp: 1918 dip, dropping swings, closing gap



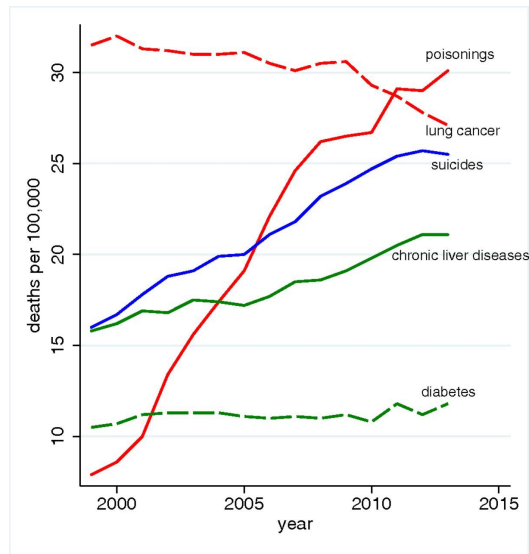
monotonic. but sped up in 05!

Steady Trend On Gay Marriage Views

Percentage opposing and supporting in national, live-interviewer polls since 1988



whites killing themselves recently



<http://www.pnas.org/content/112/49/15078.full>

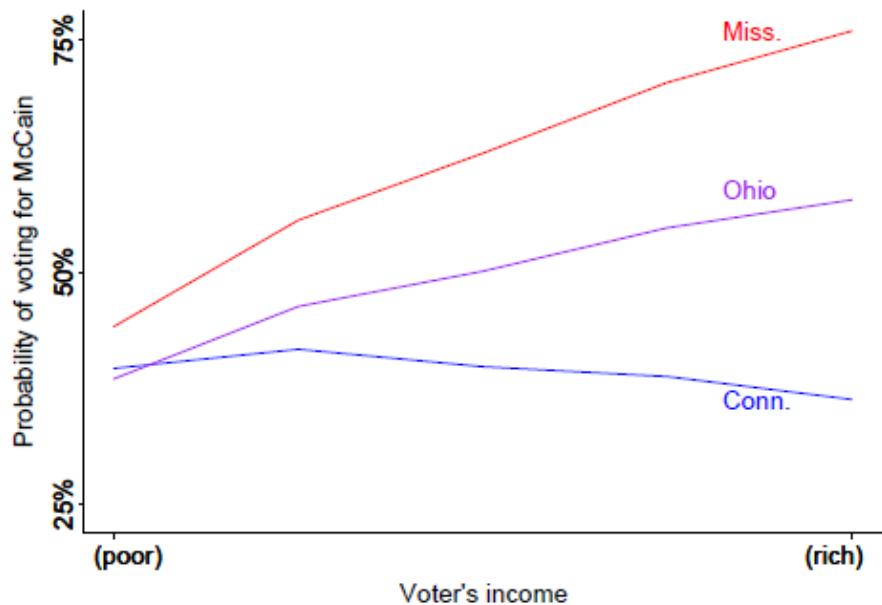
the us healthcare cost v life expectancy

- of rich countries, the us by far spends most on healthcare, and has lowest lexp
- https://www.google.com/search?q=us+healthcare+spending+and+life+expectancy+compared+to+other+countries&sca_esv=02c44965d6d4b280&sca_upv=1&rlz=1CAHWSA_enUS1125US1125&udm=2&biw=1536&bih=695&sxsrf=ADLYWILuSpqaTV9dar2rx1LXymkaga9HXQ%3A1727264656423&ei=kPfzZtLHGY_9ptQP40je6Qc&ved=0ahUKEwjSz52igt6IAxWPvokEHWCON30Q4dUDCBE&uact=5&oq=us+healthcare+spending+and+life+expectancy+compared+to+other+countries&gs_lp=Egxnd3Mtd2l6LXNlcniAARnVzIGh1YWx0aGNhcmUgc3BlbmE

red state, blue state, rich state, poor state

- even the whole book can be based on a single graph!
- <https://www.amazon.com/Red-State-Blue-Rich-Poor/dp/0691143935>
- can tell a complicated story with a graph!
- the richer you are, the more likely you are to vote ...
- 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's income

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



interested in inequality ?

- a book using many graphs
- most of the argument rests on graphs

<https://www.amazon.com/Spirit-Level-Equality-Societies-Stronger/dp/1608193411>

outline

misc

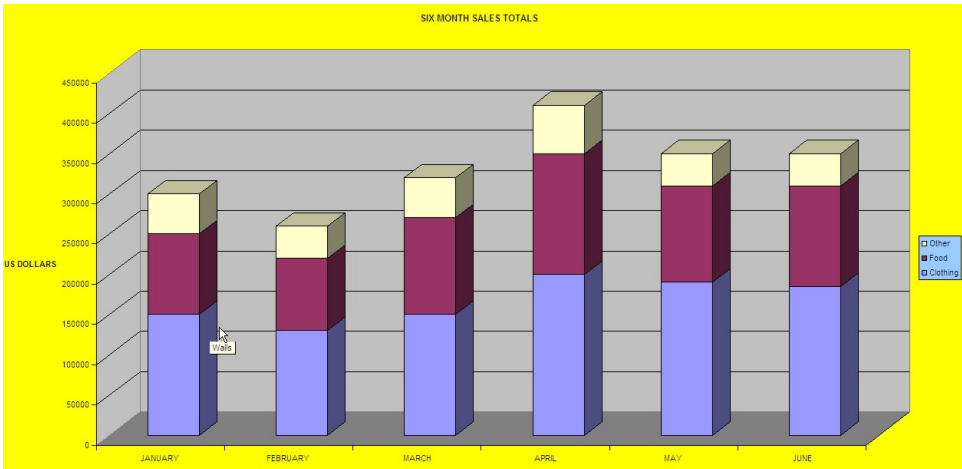
intuition

howto: some theory (Tufte)

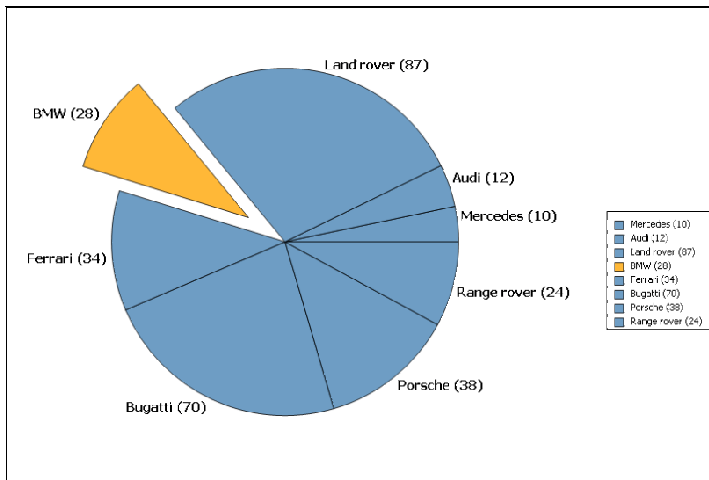
the good, the bad, and the ugly

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
- Tufte (multiple) <http://www.edwardtufte.com/tufte/>
- Kosslyn “Clear and to The Point” <http://www.amazon.com/Clear-Point-Psychological-Principles-Presentations/dp/0195320697>



Exploded Pie Slice Chart



other chartjunk examples

- <https://www.google.com/search?q=fancy+charts&tbm=isch>
- 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



Source: OECD

*Or latest available

not chartjunk (the economist)

Time spent eating and sleeping

2006, minutes per day



Source: OECD

outline

misc

intuition

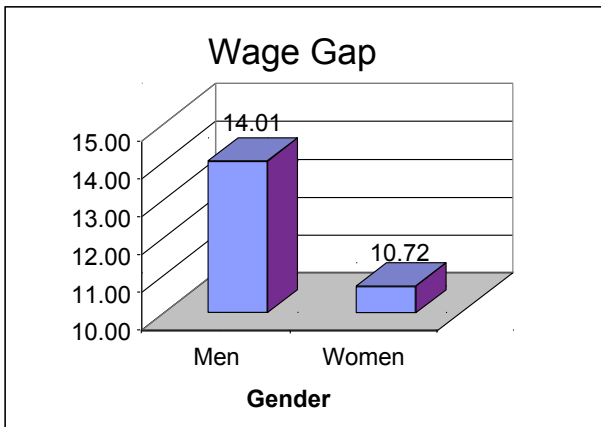
howto: some theory (Tufte)

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

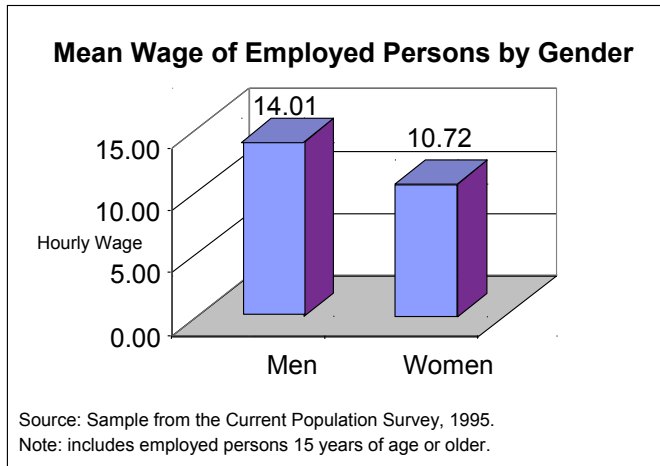
the ugly 1



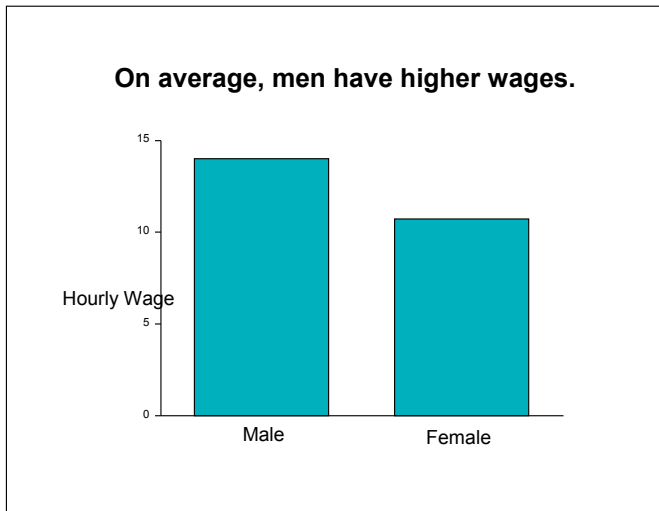
problems

- 3d
- scale not at 0 (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

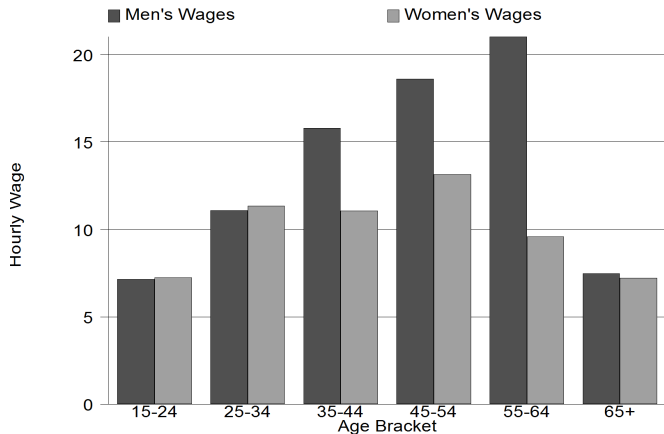


the good (there is virtue insimplicity; still...)



the good (..it's revealing to dig deeper)

Controlling for Age Changes the Picture



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 !

- at the very minimum you should discuss contradicting results in a footnote/appendix
- mention also major limitations somewhere

conclusion

- know your data
- do it by producing descriptive statistics, esp graphs
- for more see Tufte and Kosslyn

to build story get related data from various sources

- eg: Haddonfield median hh income 180k
<https://www.economist.com/briefing/2015/01/22/an-hereditary-meritocracy>
- Camden's 36k
<https://www.census.gov/quickfacts/camdencitynewjersey>
- Yet, Camden spends 1.5 or 2 per kid of Haddonfield
spends! <http://csinj.org/schoolspending>

telling story with vis, examples

- <http://www.economist.com/news/britain/21641283-unnervingly-poor-children-seem-fare-better-poor>
- <http://www.economist.com/news/briefing/21640316-children-rich-and-powerful-are-increasingly-wel>
- this is kind of story (need to expand it more) you can tell in your final project: use multiple graphs to paint some story
- <https://www.nytimes.com/2014/05/18/magazine/who-gets-to-graduate.html> and <https://mileskorak.com/2016/12/04/how-the-great-gatsby-curve-got-its-name/> and <https://s4.ad.brown.edu/projects/diversity/Data/data.htm>

telling story with vis, examples

- opinion on how US educational system produces inequality: <https://www.theatlantic.com/business/archive/2017/08/universities-inequality-fighters/538566/>



- vast majority of Americans do not drink much
- but those who do drink much, they drink a lot!

<https://www.washingtonpost.com/news/wonk/wp/2014/09/25/>

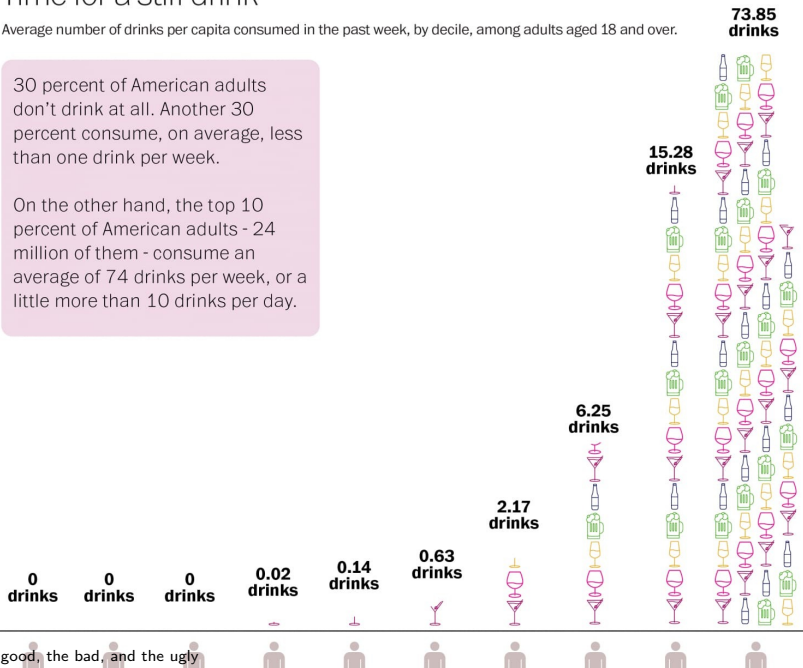
think-you-drink-a-lot-this-chart-will-tell-you

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.



wrap-up

- end every class discussing what we covered and quick look at next week
- end with a review Q&A,
- give some examples (essp in pub pol and pub adm) for concepts covered
- students will discuss concepts from the class
-
- quick look at next class

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.