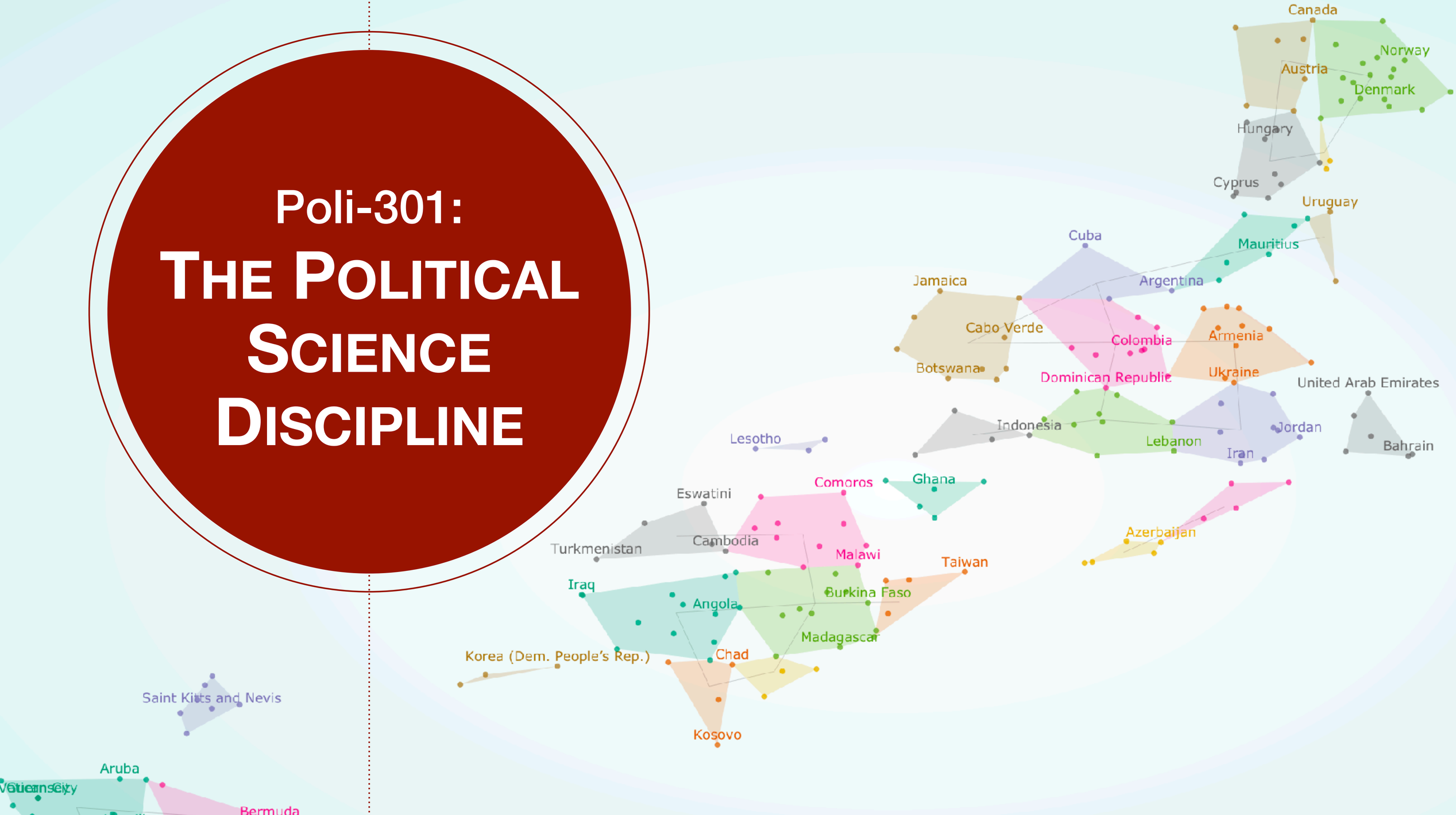
Poli-301:
The Political Science Discipline



TODAY'S AGENDA

1. Why visualize data?
2. Grammar of graphics in ggplot2()
3. The 5 graphs
4. Practice

What's happening with the website...

USC wifi was thinking website was malicious

This is why you could access off wifi but not on wifi

I put a service ticket in

They said it's fixed

We will see….

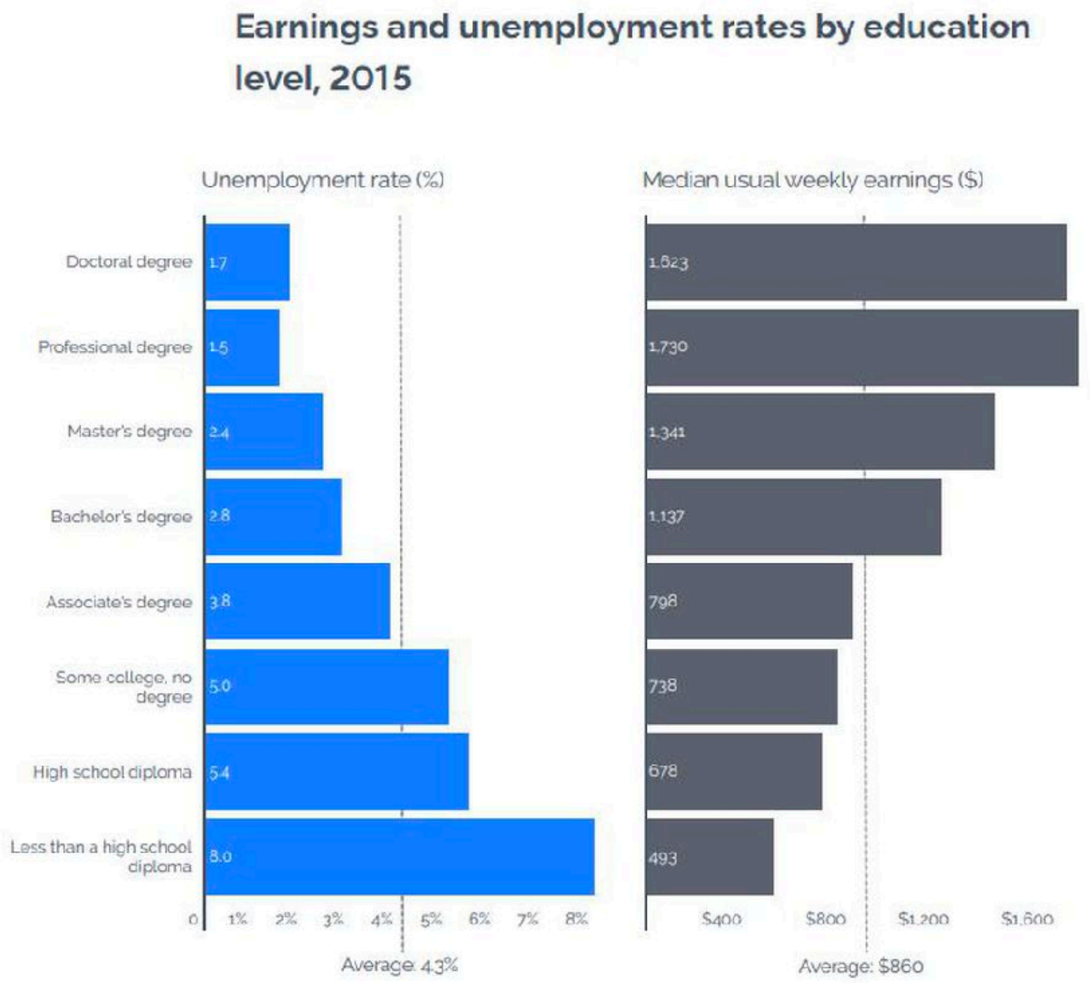
What's better?

EARNINGS AND UNEMPLOYMENT RATES BY EDUCATION LEVEL, 2015.

Education level	Unemployment rate in 2015 (%)	Median weekly earnings in 2015 ($)
Doctoral degree	1.7	1,623
Professional degree	1.5	1,730
Master's degree	2.4	1,341
Bachelor's degree	2.8	1,137
Associate's degree	3.8	798
Some college, no degree	5.0	738
High school diploma	5.4	678
Less than a high school diploma	8.0	493
All workers	4.3	860

Source: Current Population Survey, U.S. Bureau of Labor Statistics

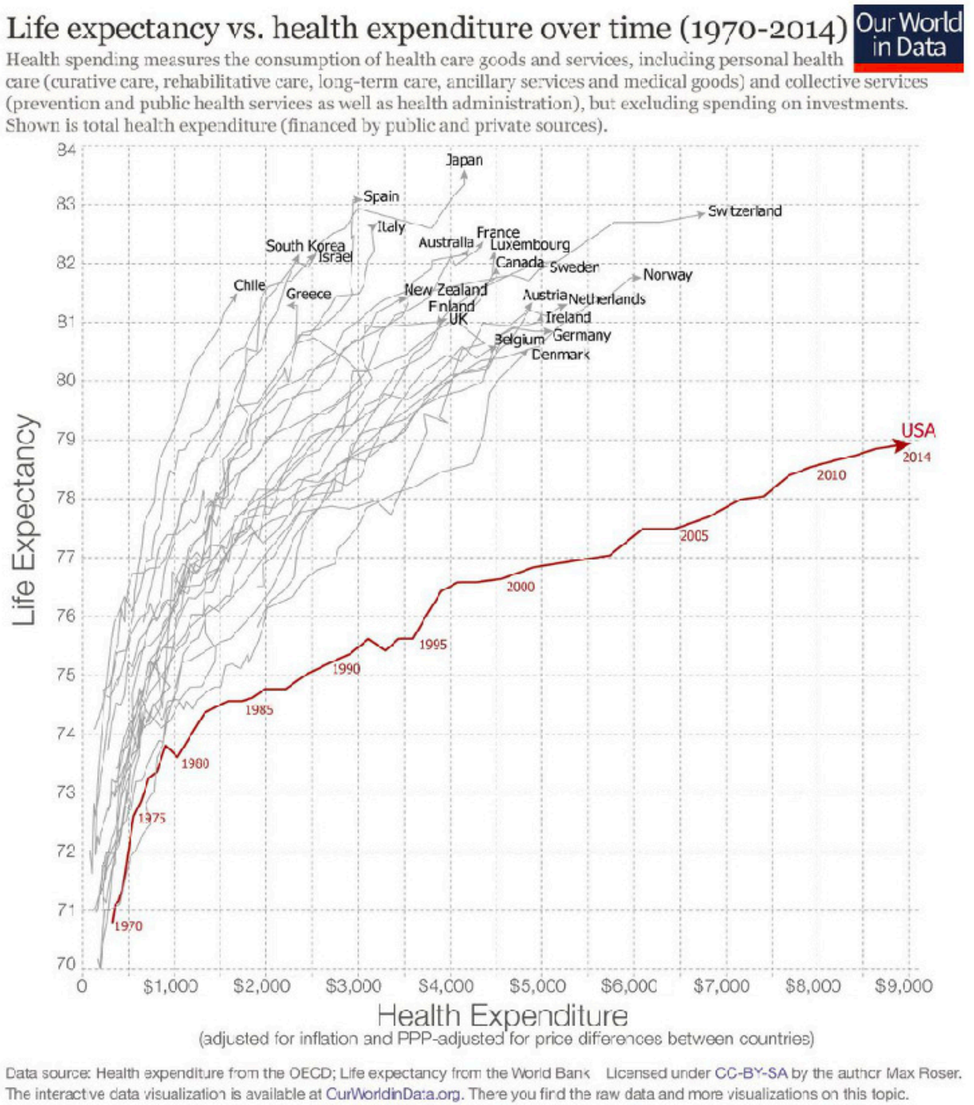
What's better?



Source: Current Population Survey, U.S. Bureau of Labor Statistics.

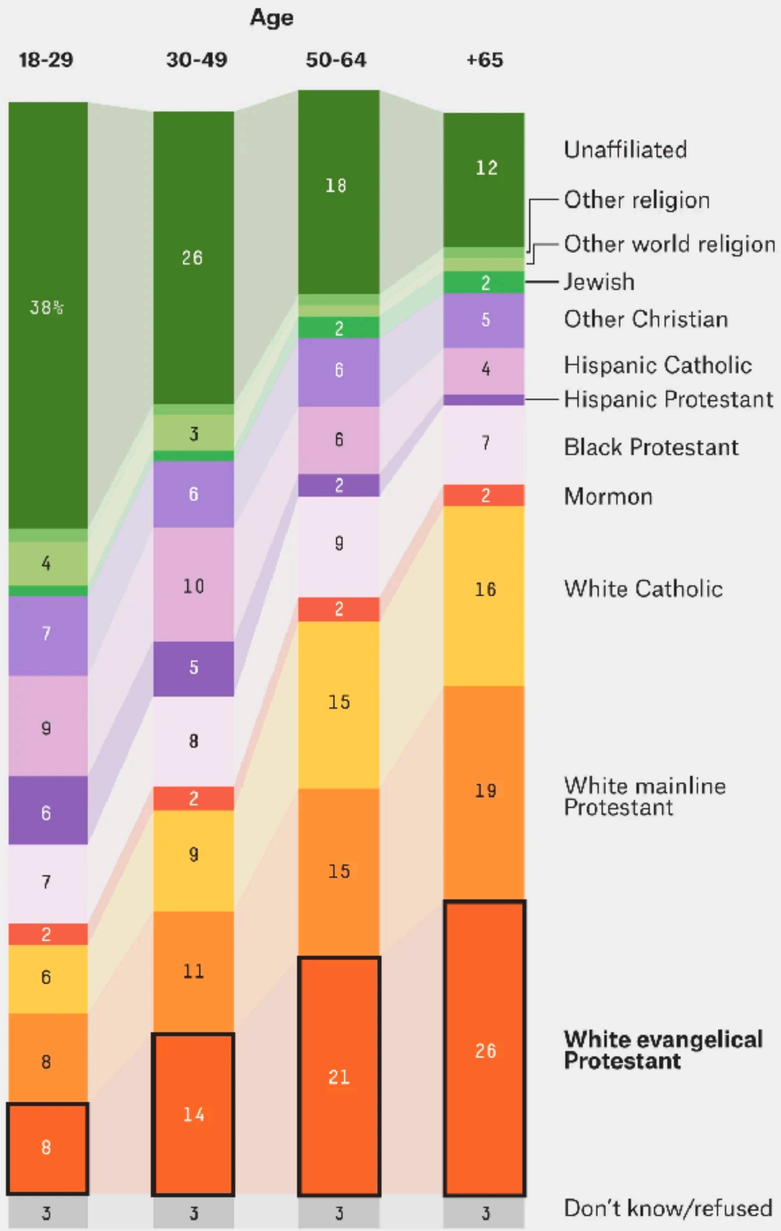
Why visualize?

- Visualizations allow us to **tell stories with data**
- Break data down to highlight **comparisons, contrasts, patterns, outliers**
- Also **not boring** (not trivial!)



A generational shift in religious identity

Share of respondents by age group and religious affiliation



Numbers may not add to 100 percent due to rounding.

FiveThirtyEight

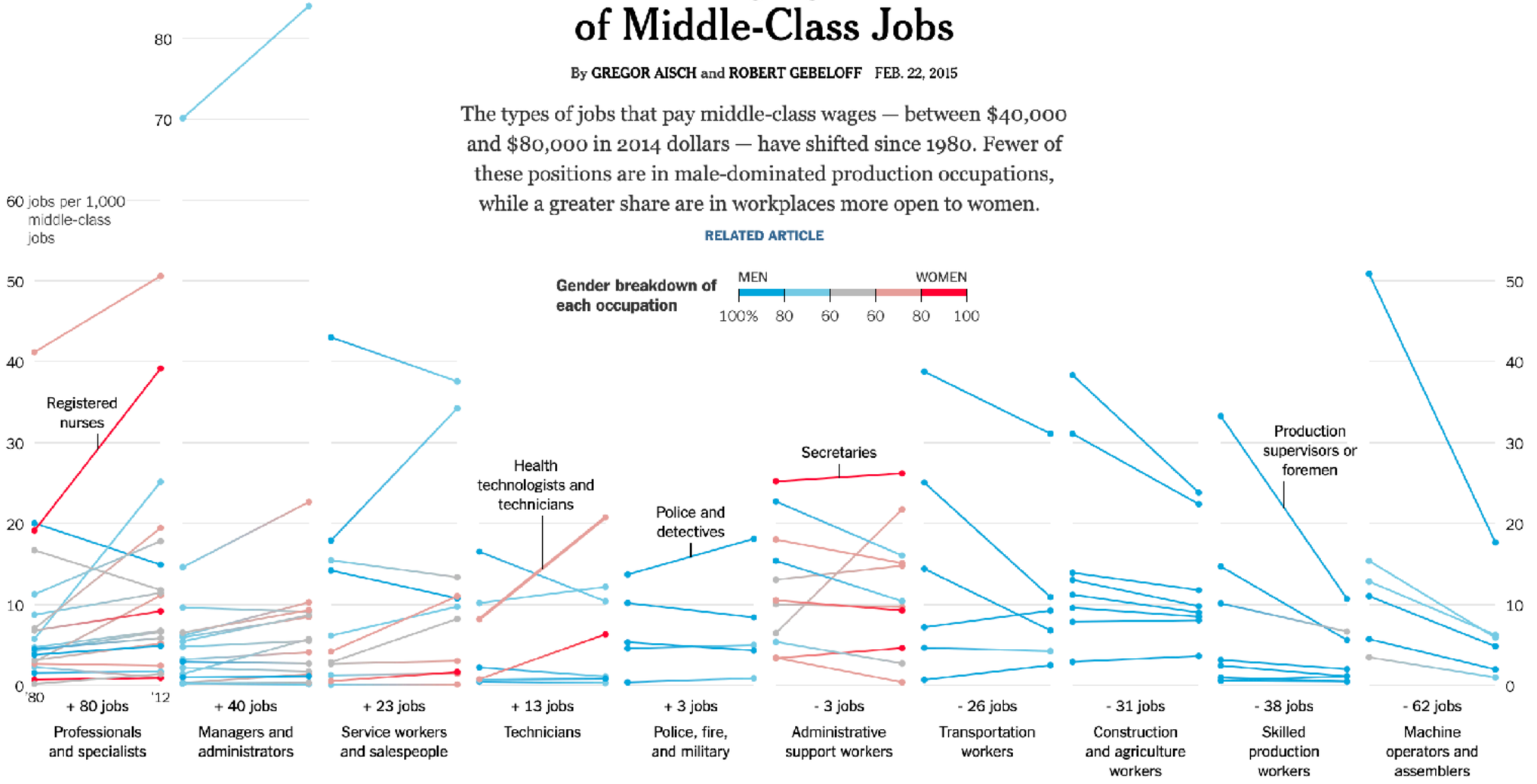
SOURCE: PRRI 2016 AMERICAN VALUES ATLAS

The Changing Nature of Middle-Class Jobs

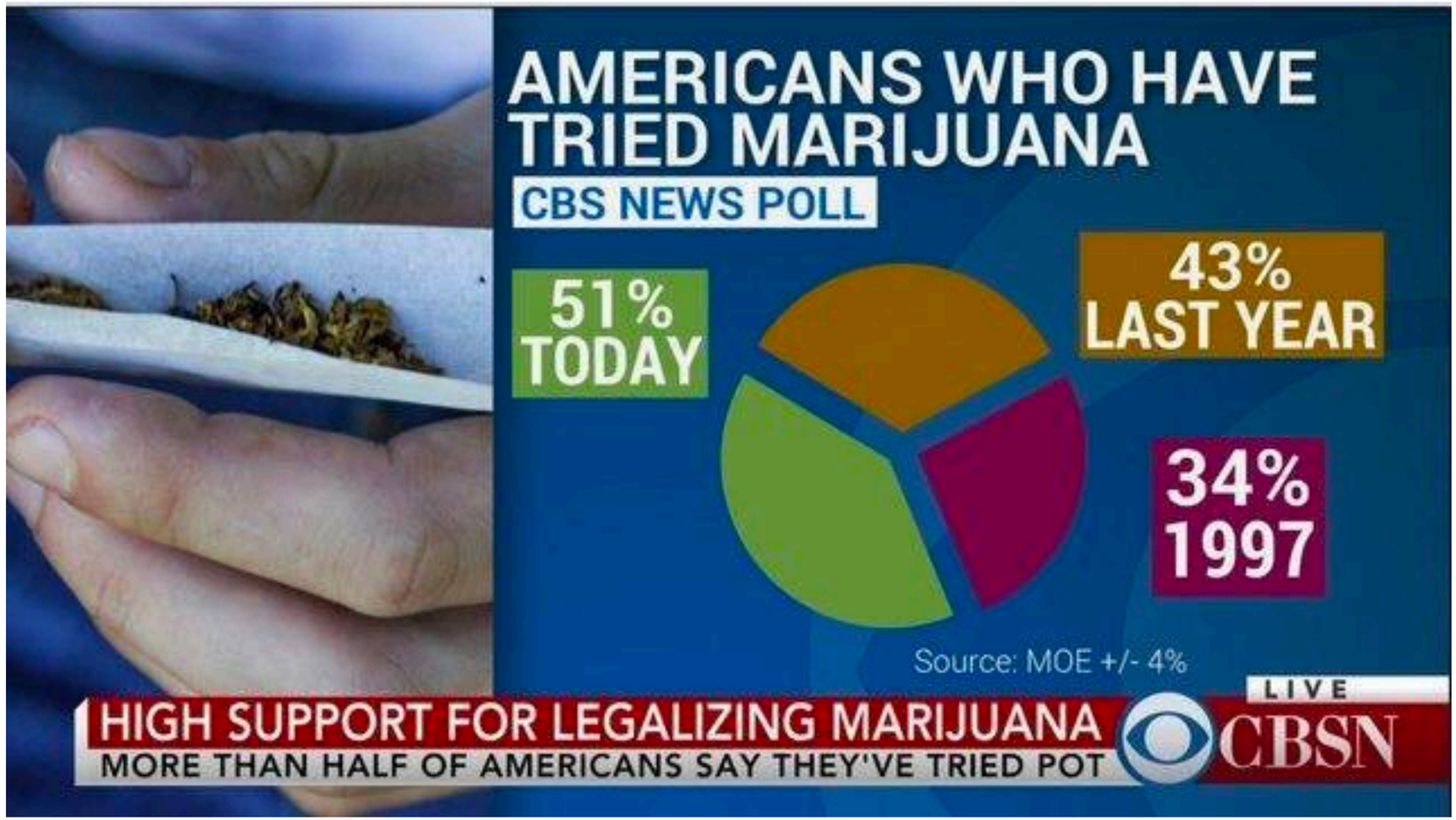
By GREGOR AISCH and ROBERT GEBELOFF FEB. 22, 2015

The types of jobs that pay middle-class wages — between $40,000 and $80,000 in 2014 dollars — have shifted since 1980. Fewer of these positions are in male-dominated production occupations, while a greater share are in workplaces more open to women.

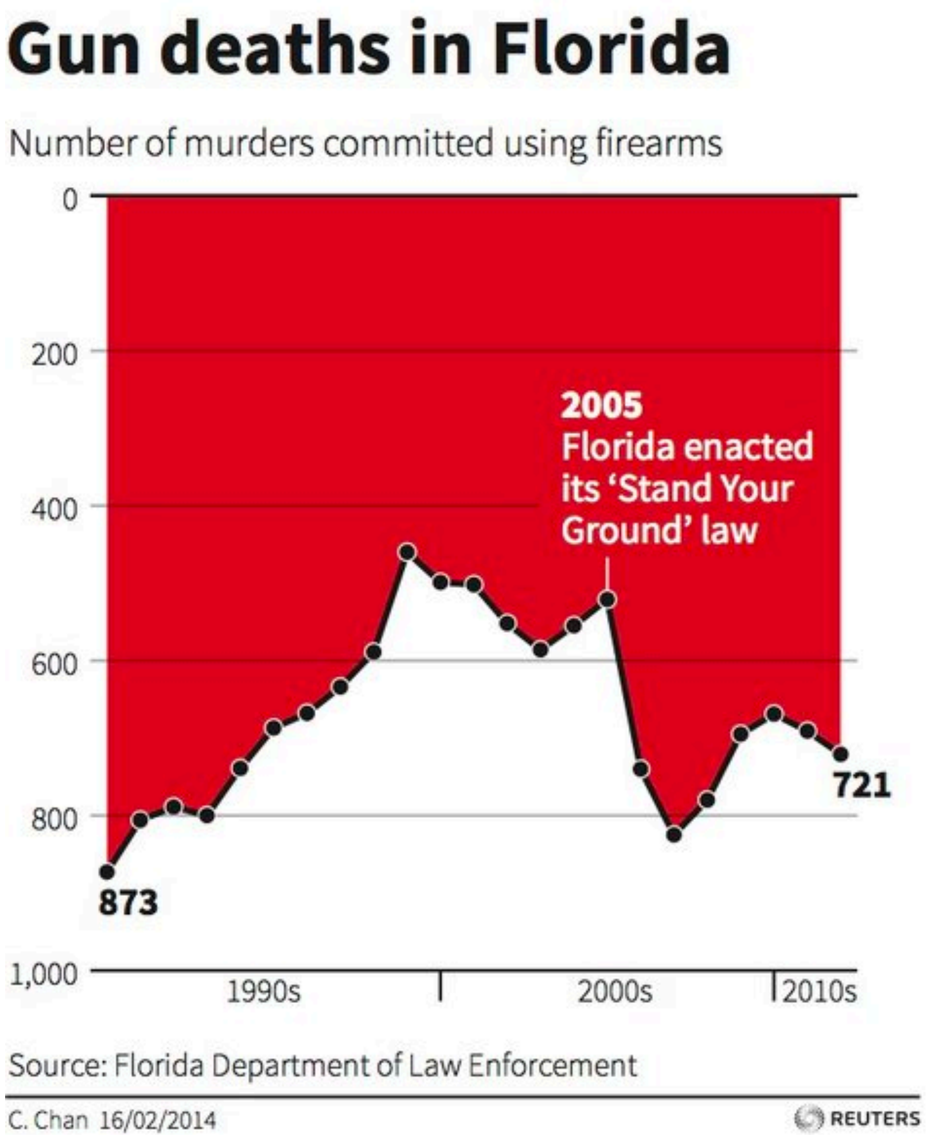
RELATED ARTICLE



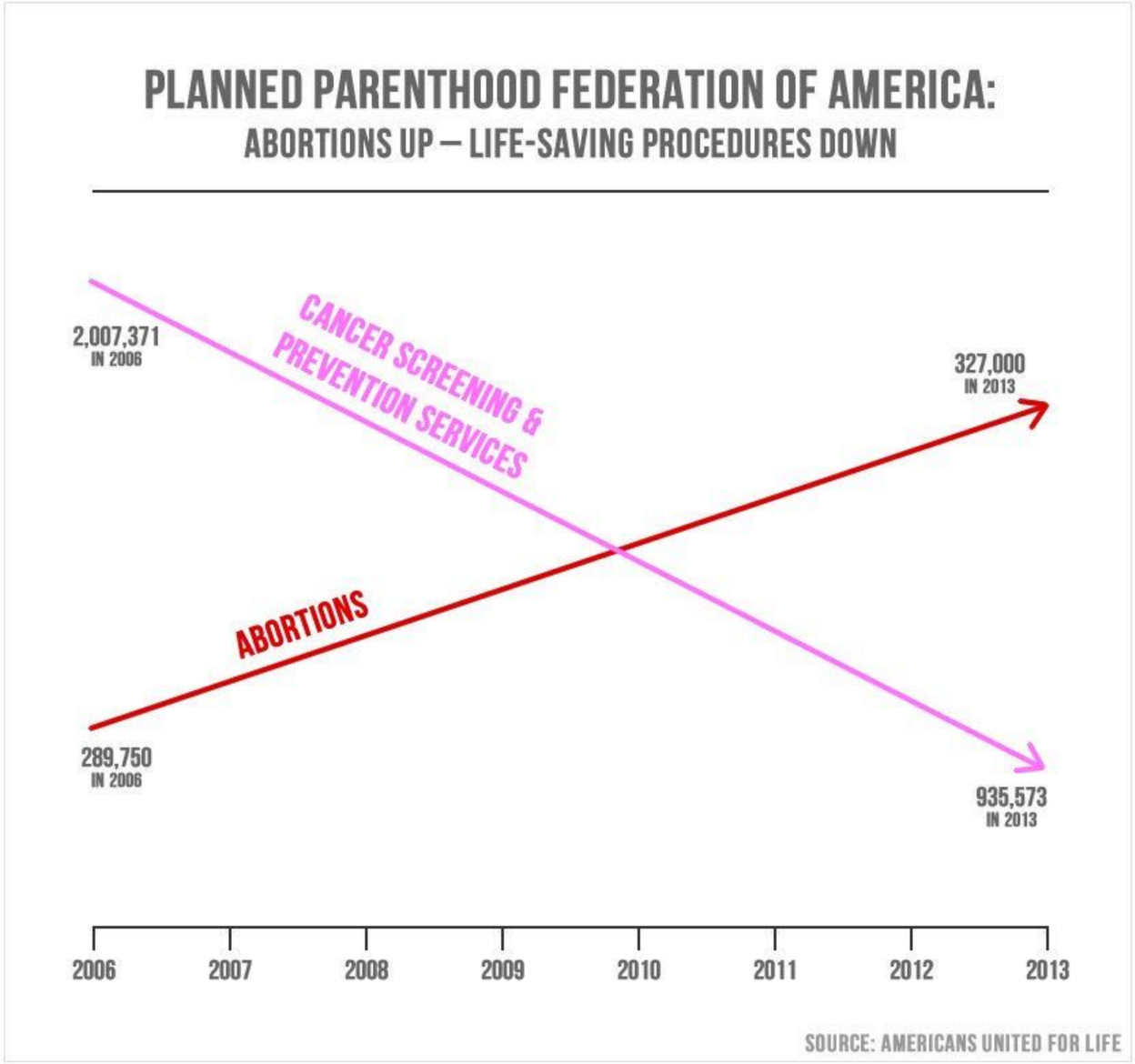
Visualization can also go **terribly** wrong



What's wrong here?



Bad viz on purpose



Good Viz is...

Programming skills and know-how

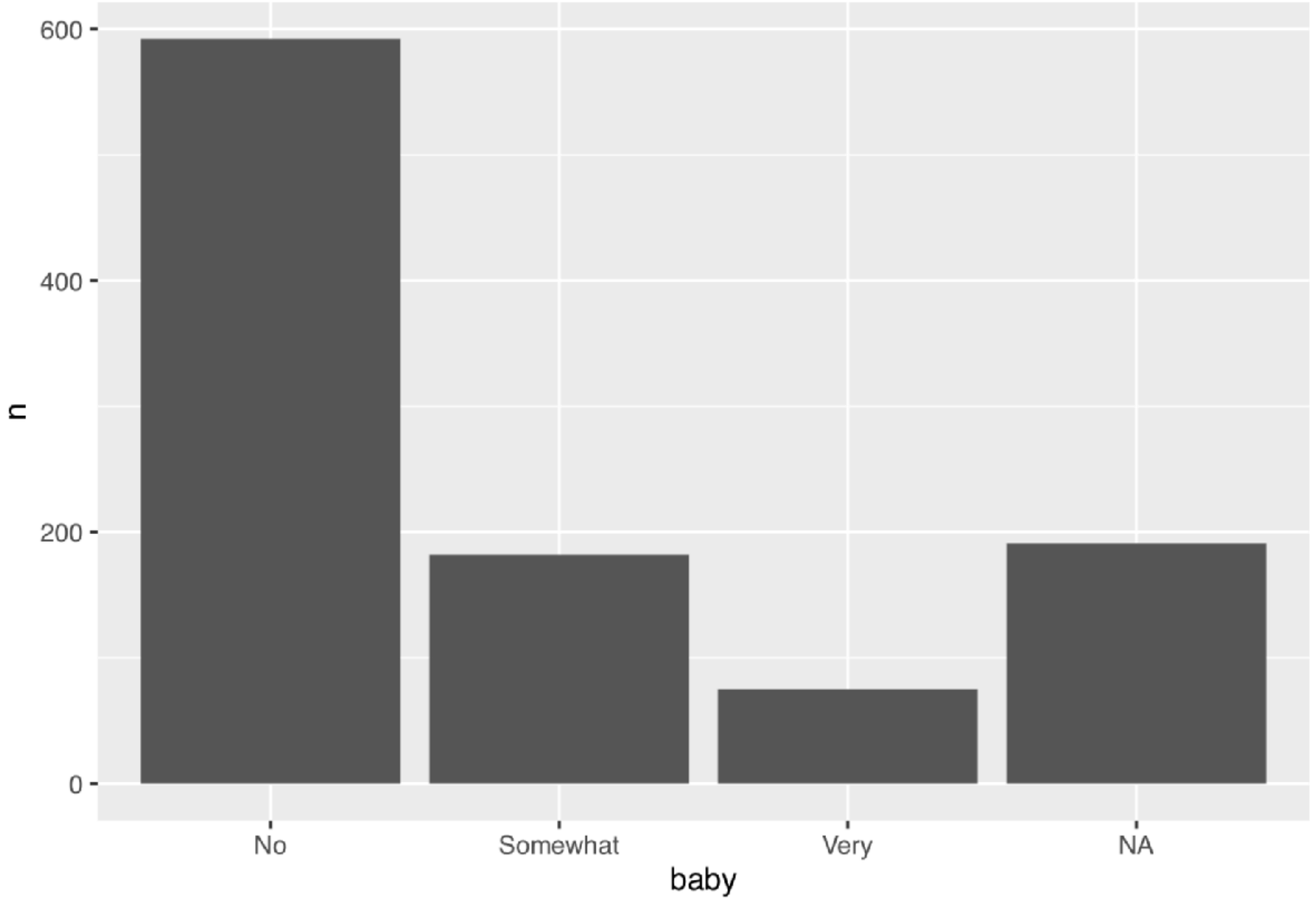
Understanding fit and context

Creativity and “good taste”

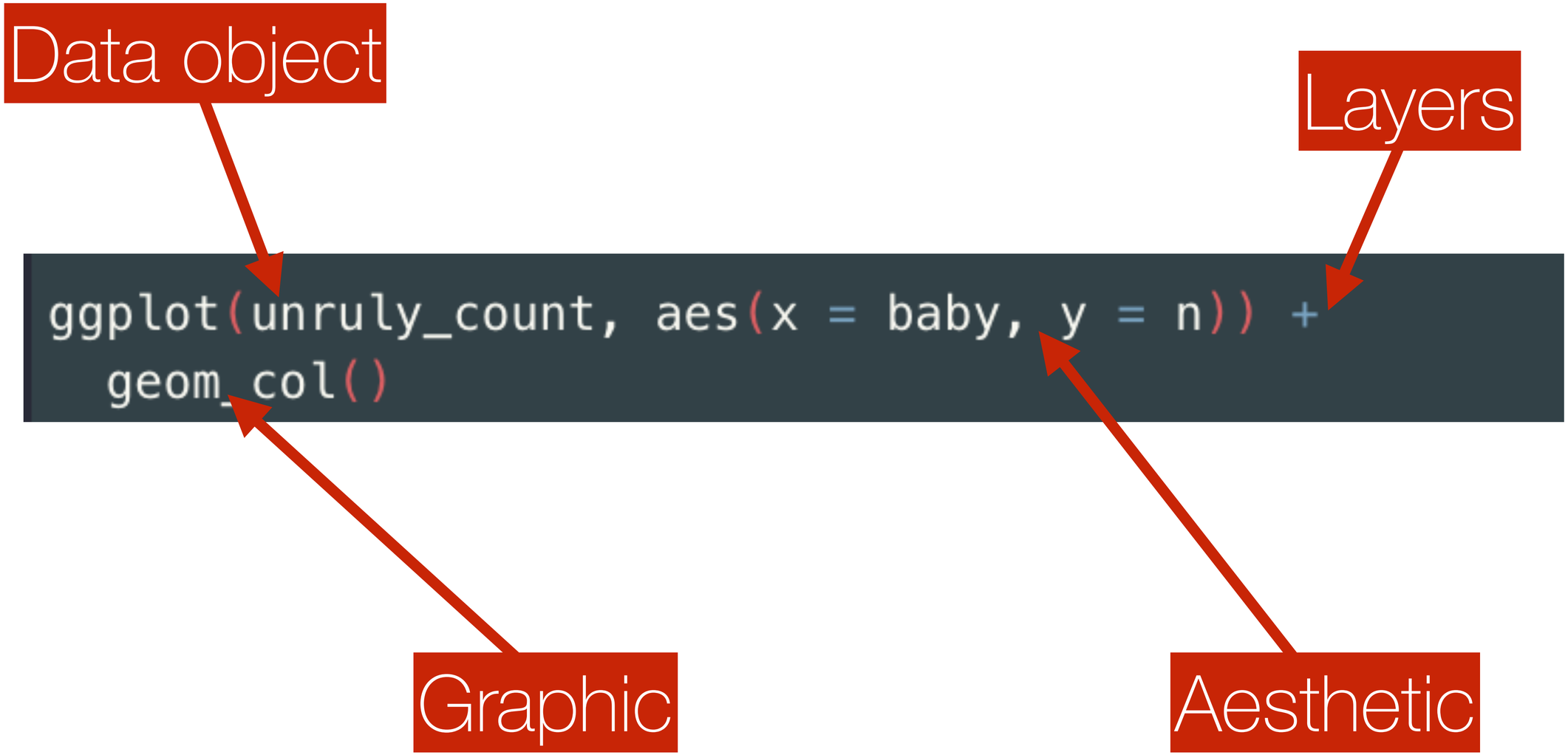
2

Grammar of graphics with ggplot

How rude is it to bring baby on plane?



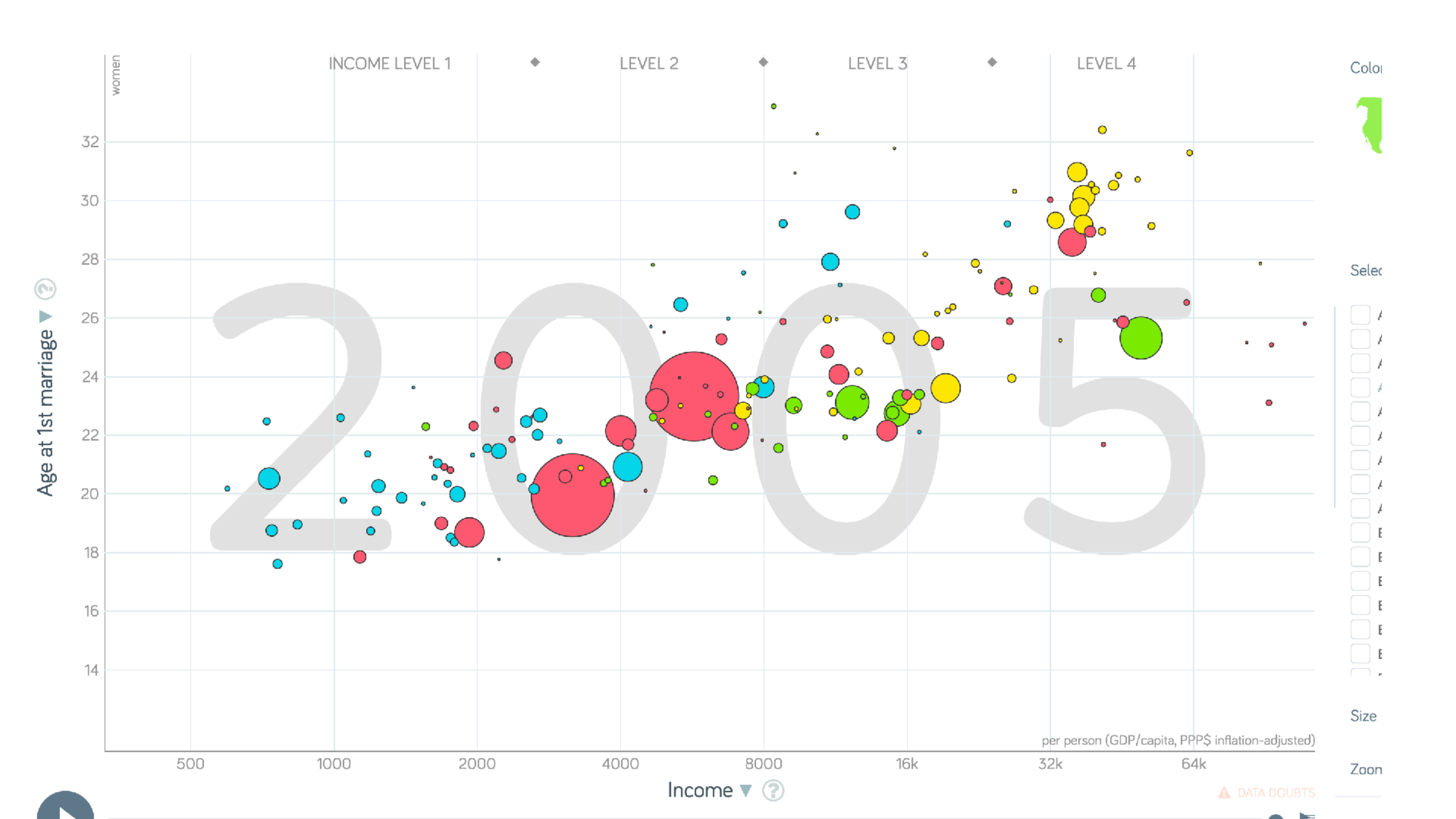
ggplot


Data object
Layers
ggplot(unruly_count, aes(x = baby, y = n)) +
geom_col()
Graphic
Aesthetic

Being explicit is good

But **not** necessary

```
ggplot(data = alaska_flights, mapping = aes(x = dep_delay, y = arr_delay)) +
  geom_point()
```


INCOME LEVEL 1
LEVEL 2
LEVEL 3
LEVEL 4
women
Age at 1st marriage
32
30
28
26
24
22
20
18
16
14
2005
per person (GDP/capita, PPP$ inflation-adjusted)
500
1000
2000
4000
8000
16k
32k
64k
Income
DATA DOUBTS
Color
Select
Size
Zoom

Gapminder data

```
# A tibble: 6 x 6
  country     continent  year lifeExp      pop gdpPercap
  <fct>       <fct>     <int>   <dbl>    <int>     <dbl>
1 Afghanistan Asia       2002    42.1 25268405      727.
2 Albania     Europe     2002    75.7  3508512     4604.
3 Algeria     Africa     2002    71.0 31287142     5288.
4 Angola      Africa     2002    41.0 10866106     2773.
5 Argentina   Americas   2002    74.3 38331121     8798.
6 Australia   Oceania    2002    80.4 19546792    30688.
```

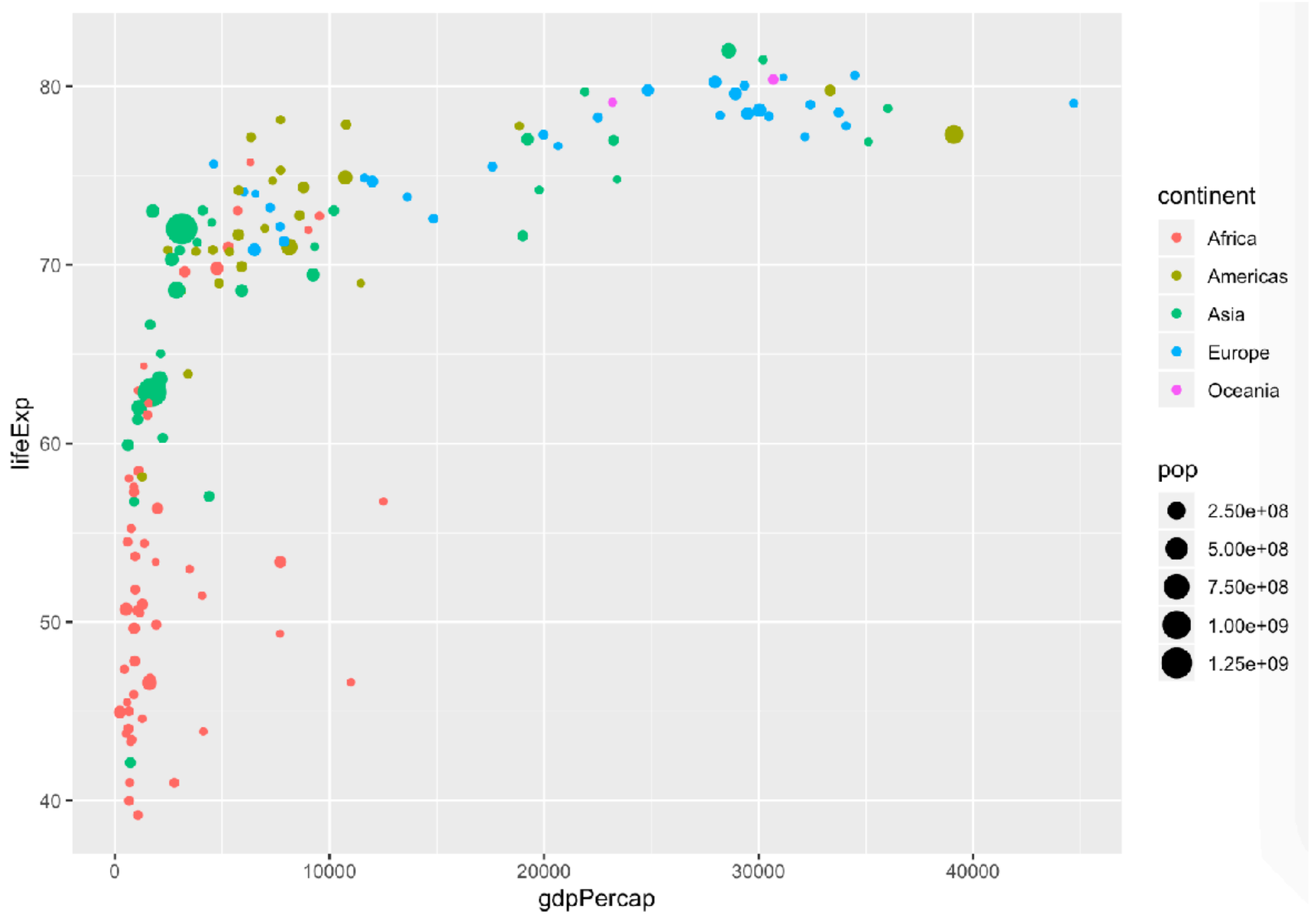
Plotting wealth and health

data	aes()	geom_
Wealth (gdp/cap)	x	point
Health (life expectancy)	y	point
Continent	color	point
Population	size	point

Gapminder: health and wealth

```
ggplot(gapminder_2002, aes(x = gdpPercap, y = lifeExp,
                           size = pop, color = continent)) +
  geom_point()
```

Gapminder: health and wealth

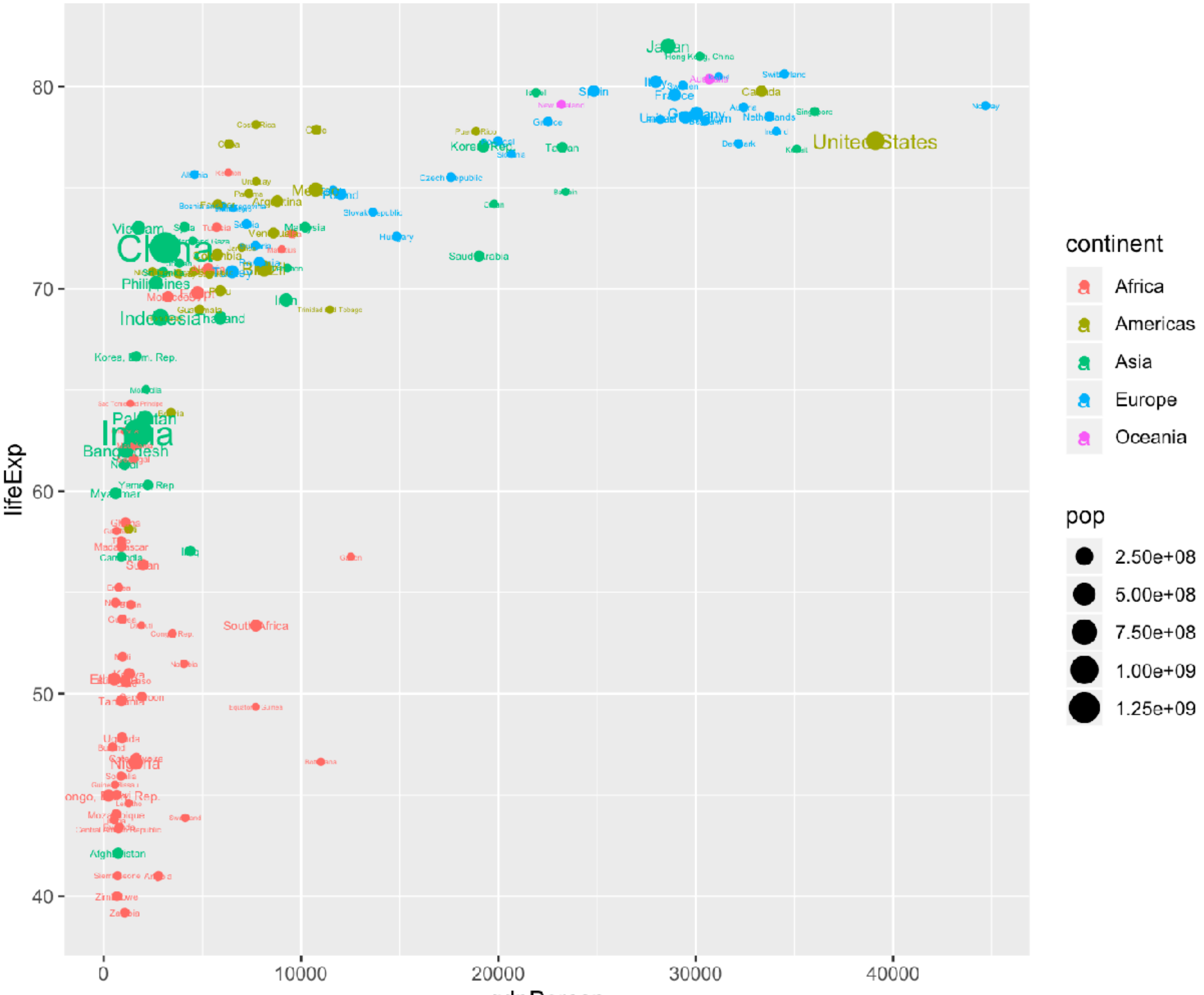


Adding labels

data	aes()	geom_
Wealth (gdp/cap)	x	point
Health (life expectancy)	y	point
Continent	color	point
Population	size	point
Country name	label	text

Adding labels

```
ggplot(gapminder_2002, aes(x = gdpPercap, y = lifeExp,
                           size = pop, color = continent,
                           label = country)) +
  geom_point() +
  geom_text()
```



lifeExp
40
50
60
70
80
0
10000
20000
30000
40000
Japan
United States
Canada
Spain
France
South Africa
Saudi Arabia
Korea, Dem. Rep.
Afghanistan
continent
Africa
Americas
Asia
Europe
Oceania
pop
2.50e+08
5.00e+08
7.50e+08
1.00e+09
1.25e+09

3

The 5 most common graphs

The Big 5

Scatterpots

Linegraphs

Boxplots

Histograms

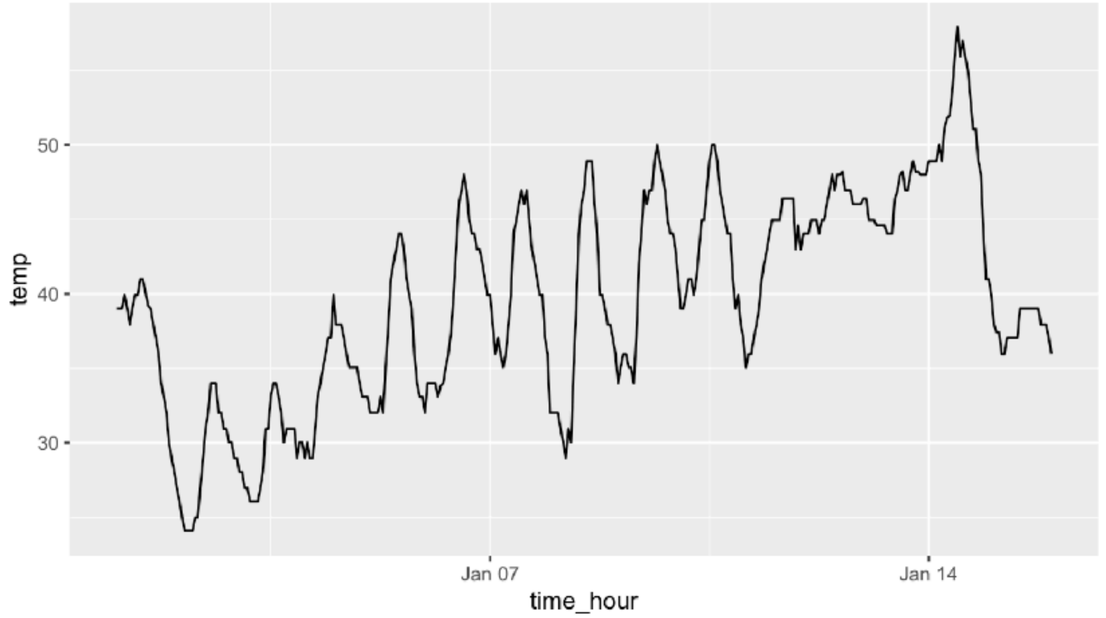
Barplots

~~Piecharts~~

Linegraphs

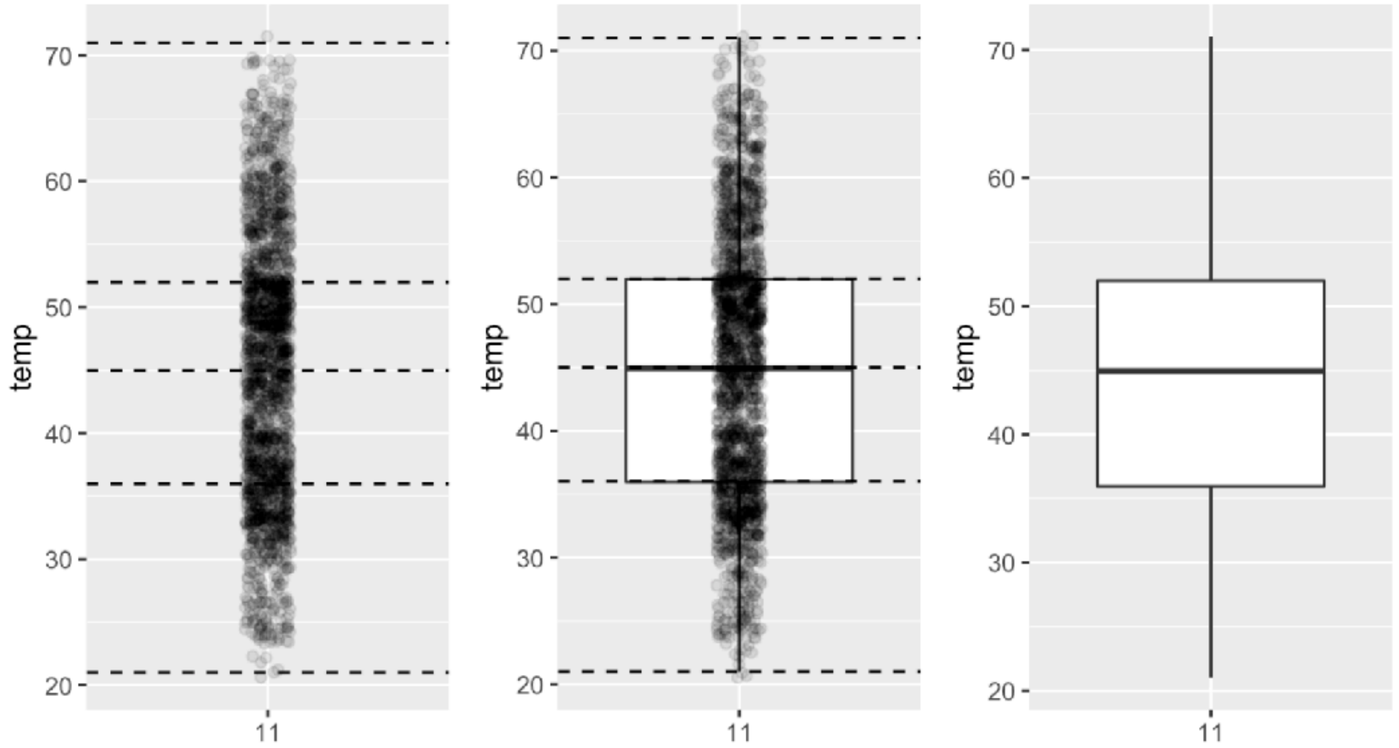
Connect data points to show trend or trajectory

Often used with **time** on the x-axis



Boxplot refresher

Minimum: 21°F
First quartile: 36°F
Median: 45°F
Third quartile: 52°F
Maximum: 71°F

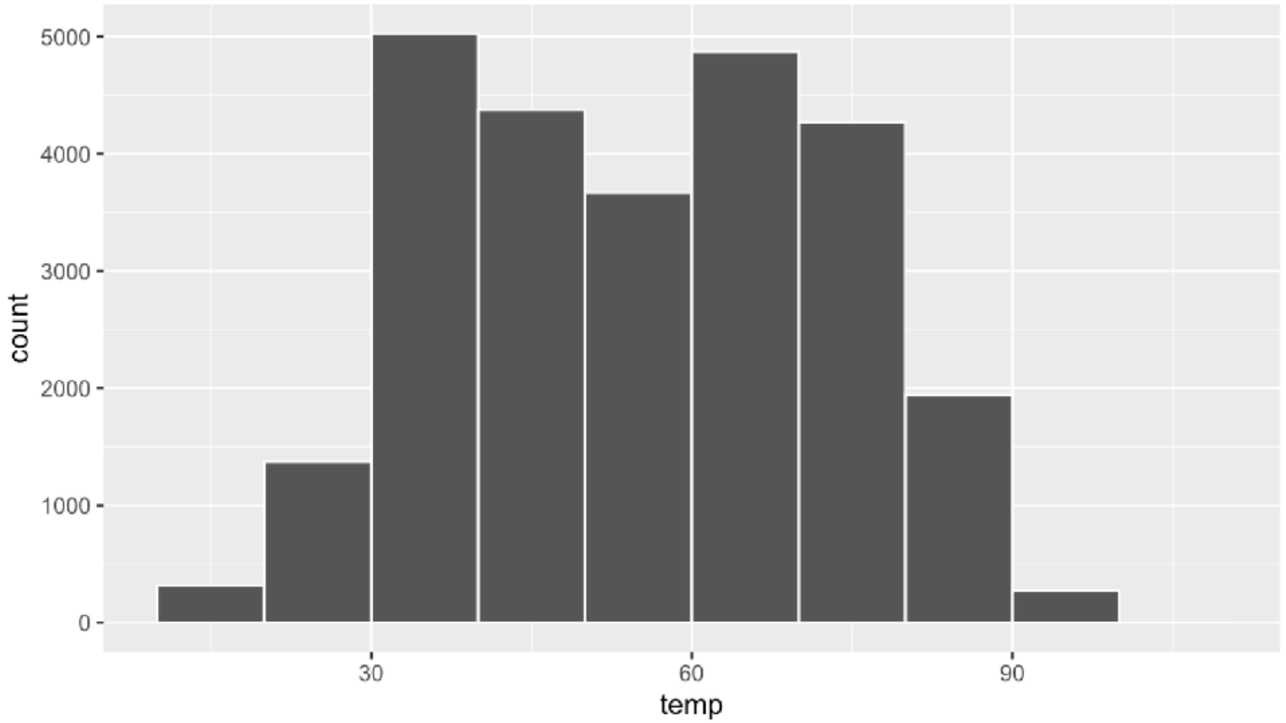


Histograms

Show distribution of one variable

Break up values into **bins** to create groupings

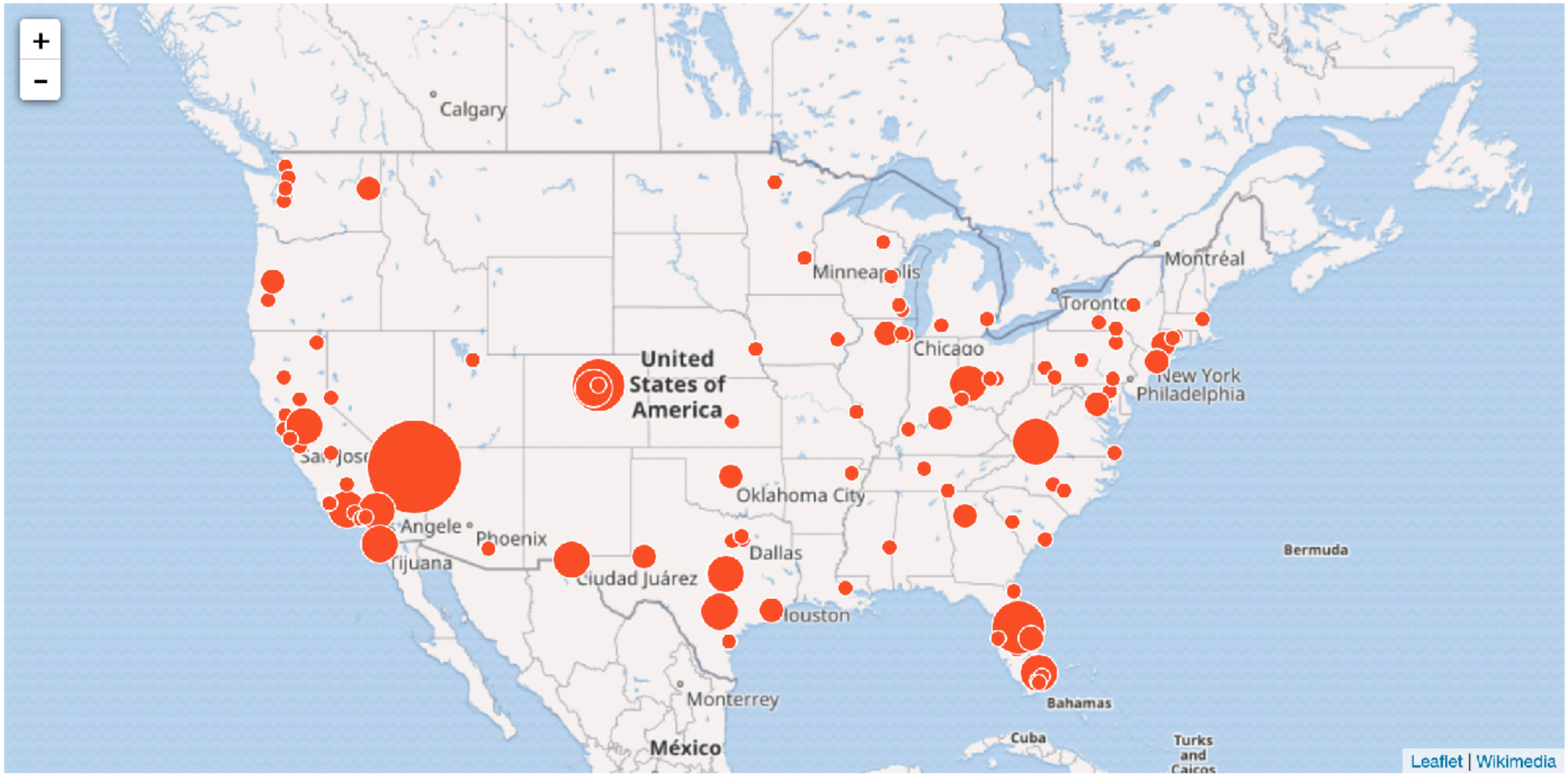
Also used to compare distributions



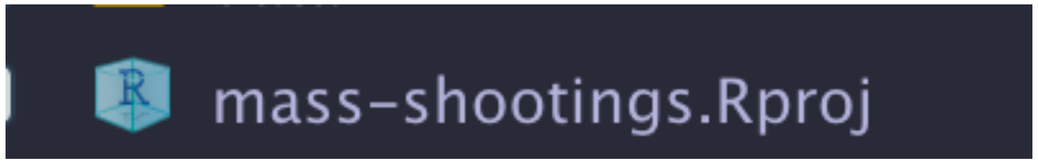
Demo

Mass shootings in the US

US Mass Shootings, 1982-2019: Data From MotherJones



In-class activity

- Download **mass-shootings.zip** from site
- Unzip file somewhere easy to find again (maybe a folder for this class?)
- Double-click on 

- Open **plot-shootings.R**

Practice

What’s least clear to you right now?