

# EDS 240: Lecture 2.1

*Choosing the right graphic form*

---

Week 2 | January 13<sup>th</sup>, 2024

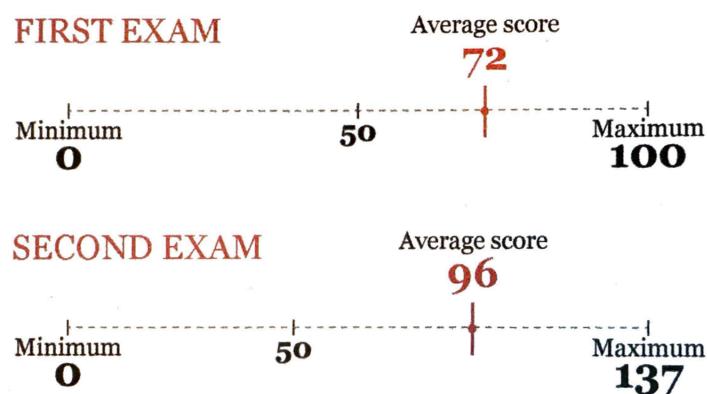
# We understand complex numbers better when they're represented visually

---

“Exams will have a total of 137 points rather than the usual 100. This scoring system has no effect on the grade you get in the course, but it seems to make you happier”

-Richard H. Thaler, economist & professor

- **Early years:** exam graded 0 - 100 with an average score of 72 points = lots of complaints
- **Later years:** exam graded 0 - 137 with an average score of 96 points = very few complaints



# Vision is our most well-developed sense

---

**Mapping data into visual properties is powerful**

## Mapping?

*How values of a variable(s) of interest are represented by visuals (e.g height of bar, shaded region of area plot, color of data points)*

# How do you choose the right graphic form to represent your data?

---

“If I had the answer to that, I’d be rich by now...I have no idea, but I can give you some clues to make your own choices based on what we know about why and how visualization works”

-Alberto Cairo<sup>1</sup>, in his book, *The Truthful Art*

# Exercise: Map data to visual properties

---

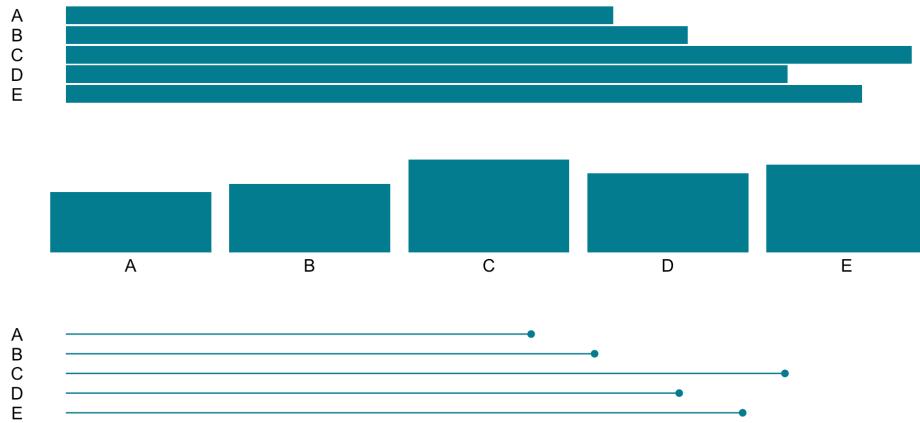
Let's say **you want to compare unemployment rates of 5 countries:** A, B, C, D, E (the actual values here are not important).

How would you **map the unemployment rates to visual properties** in a way that enables your readers to accurately compare values without having to read all the numbers?

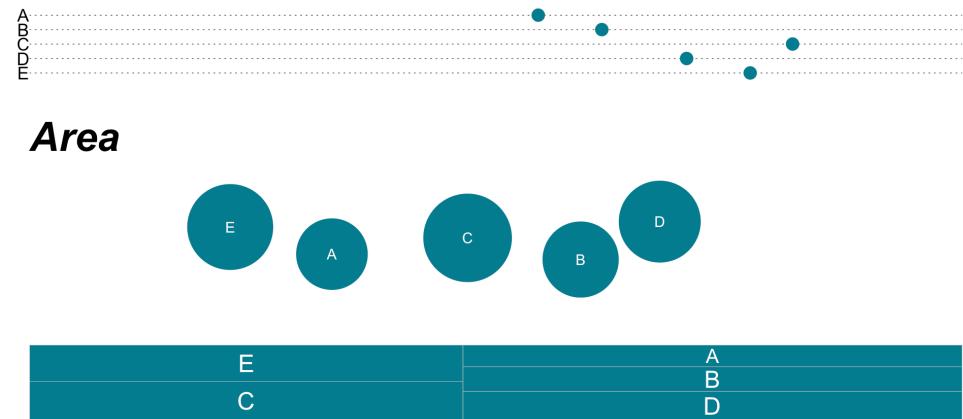
# Different methods of encoding the same data

---

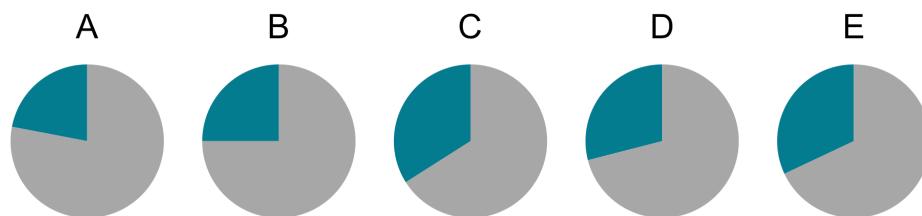
**Length or height**



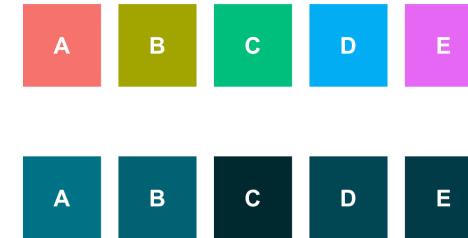
**Position**



**Angle / Area**



**Hue and shade**



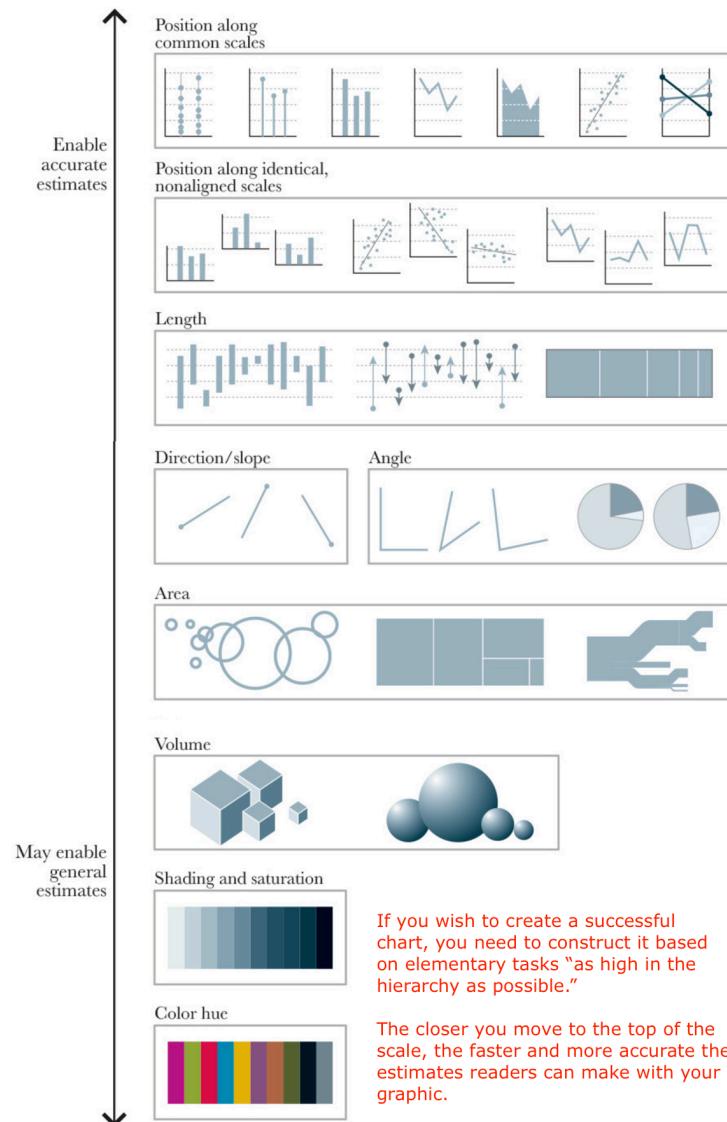
# “Hierarchy of elementary perceptual tasks”

---

William S. Cleveland & Robert McGill (1984) **Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods**, *Journal of the American Statistical Association*, 79:387, 531-554,  
DOI: 10.1080/01621459.1984.10478080

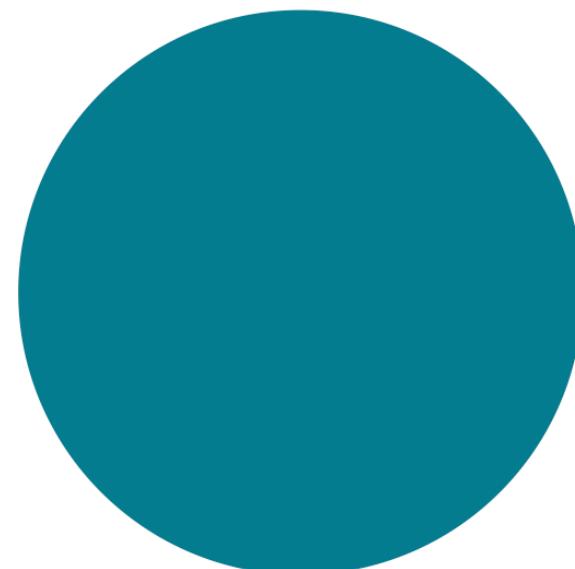
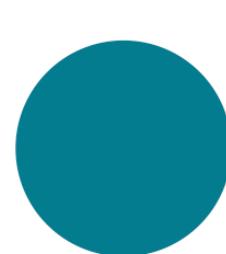
- a viewer performs one or more of these mental-visual tasks (judging position, perceiving angles / areas, etc.) to extract the values of real variables represented on most graphs
- successful charts are constructed based on elementary tasks “as high in the hierarchy as possible”

Alberto Cairo’s recreation of Cleveland & McGill’s  
Hierarchy of Elementary Perceptual Tasks →



# Exercise: How many times bigger is the larger circle?

---



00 : 30

# Exercise: How many times bigger is the larger bar?

---



00 : 30

Both the circles & rectangles differ by a magnitude of 7

---



# Caveats to the hierarchy

---

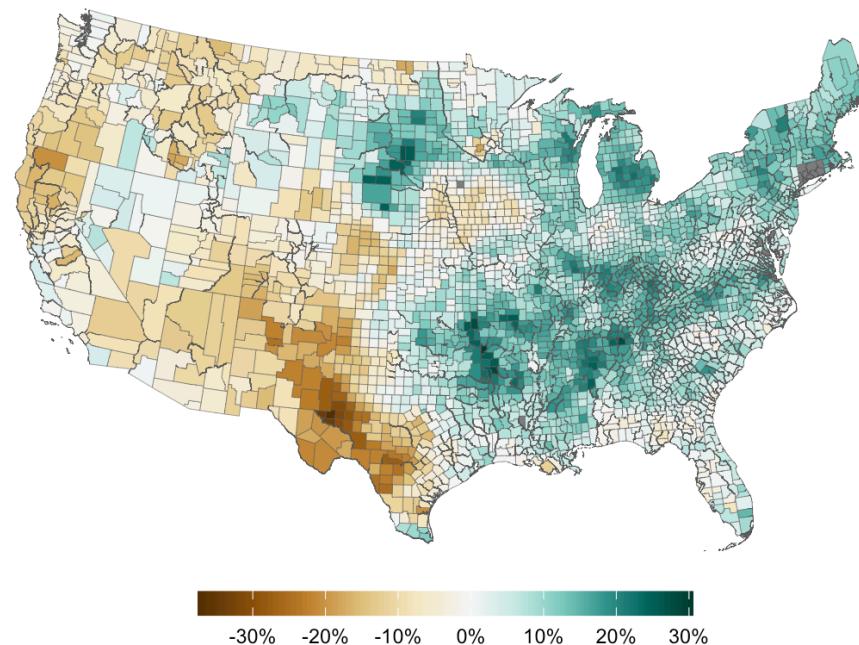
1. Cleveland & McGill only considered *statistical charts*. What about data maps, for example, that rely on area / shading / hue, which fall lower on the hierarchy?

# Caveats to the hierarchy - an example

**Lower scale methods can be appropriate when the goal is to reveal general patterns.** For example, a choropleth map displays divided geographical areas / regions, which are colored in relation to a numeric variable.

5-year precipitation compared to the 20th century average

February 2019 - January 2024



Source: National Centers for Environmental Information

Map created using precipitation data from NOAA's [National Centers for Environmental Information](#).

# Caveats to the hierarchy

---

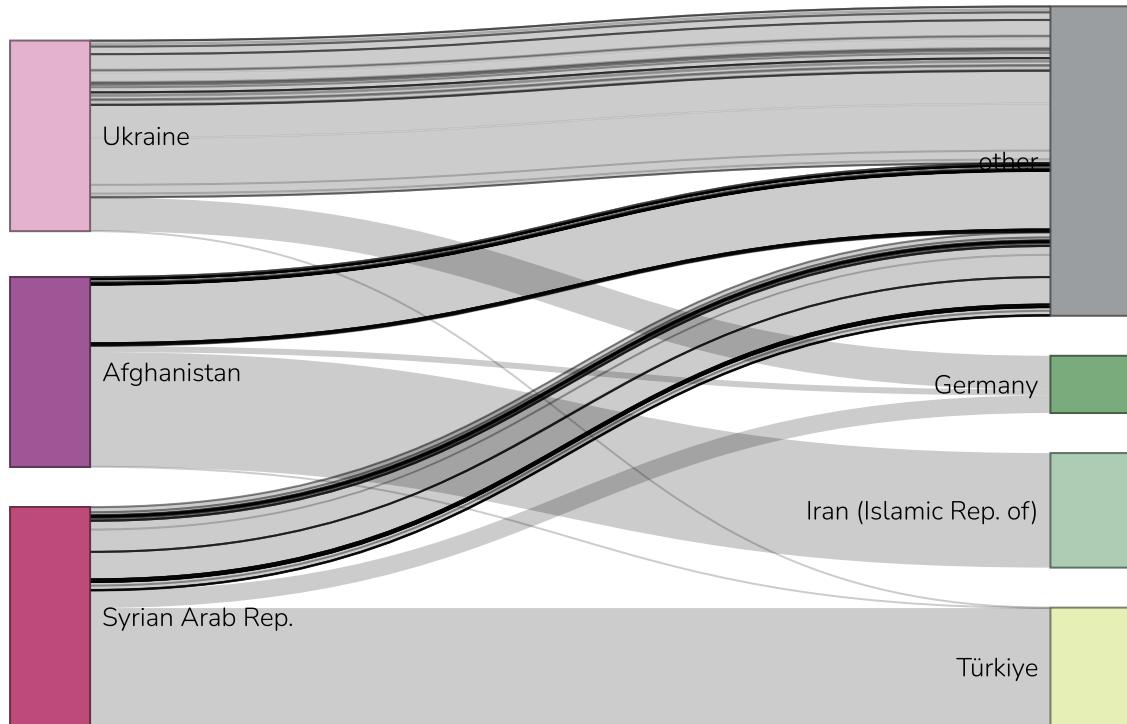
1. Cleveland & McGill only considered *statistical charts*. What about data maps, for example, that rely on area / shading / hue, which fall lower on the hierarchy?
2. No method of choosing a graphic form is perfect! It's important to think critically about your graphic's purpose and how best to represent your data to serve that purpose.

# Caveats to the hierarchy - an example

---

Consider how you might display the same data in the following Sankey diagram, which depicts the flow of refugees in 2022, using graph types from the top of Cleveland & McGill's hierarchy.

**What is the purpose of this chart?**



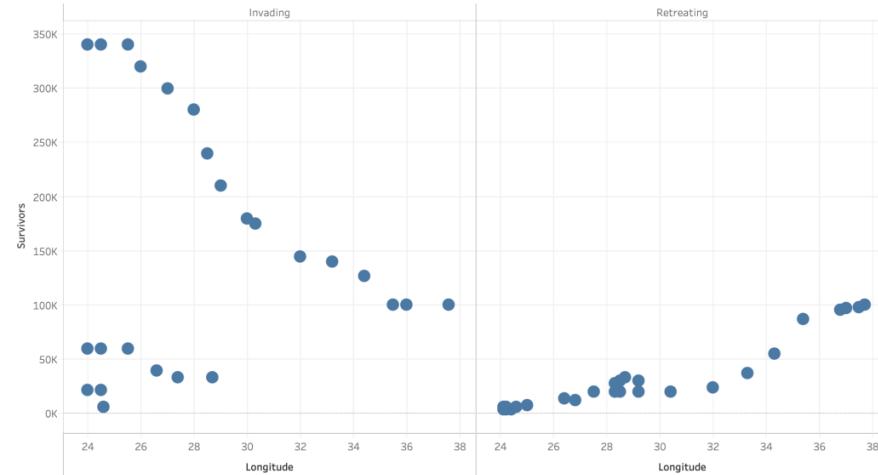
Graphic recreated using the [{networkD3}](#) package following Louise E. Sinks' blog post, [TidyTuesday: Exploring Refugee Flow with A Sankey Diagram](#)

# Unique visualizations are important!

*“Charts are often designed to persuade, educate, and motivate. Designing for serendipitous discovery, educational impact, hedonic response, or changes in behavior is in some cases only tangentially connected with the precision of a particular visualization”*



(a)



(b)

Figure 1: Two re-creations of Charles Minard’s map of the invasion of Russia in Tableau. In (1a) we hew as closely as possible to Minard’s design. In (1b) we attempt to encode all of the same data about the three groups of the invasion force using the most efficient channel: position on a common axis. In which circumstances might we prefer (1a) over (1b)?

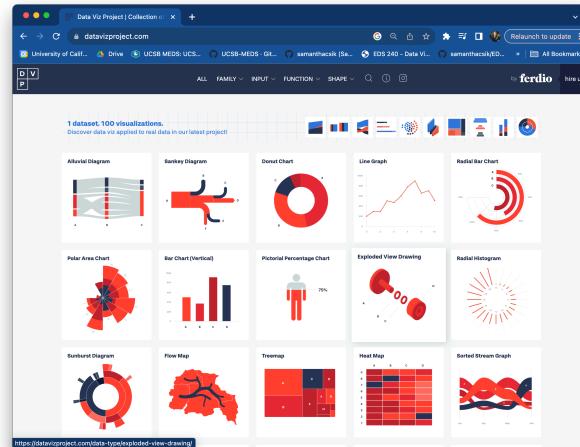
# Tips for choosing the right graphic form

---

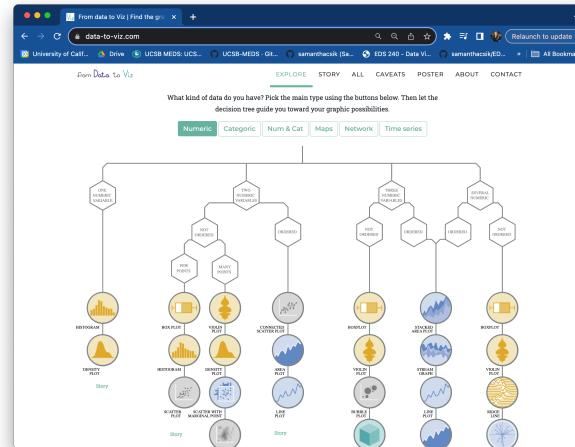
1. **Think about the task(s) you want to enable or message(s) you want to convey.** For example, do you want to compare, see change or flow, reveal relationships or connections, envision temporal or spatial patterns, show big picture trends or allow for comparisons of individual values.
2. **Consider the number of variables and the number of data points, as well as the data types you're working with.** For example, do you have several vs. many data points? How many categorical and/or numeric variables? Are your variables ordered or not ordered? Data types can dictate which graphical form is appropriate.
3. **Try different graphic forms,** especially if you have more than one task to enable or message to convey.
4. **Arrange the components of the graphic** to make it as easy as possible to extract meaning from your graphic quickly.
5. **Test the outcomes of your graphic on others,** particularly on those who are representative of the audience you are trying to reach.

# 1. What task(s) to enable / message(s) to convey

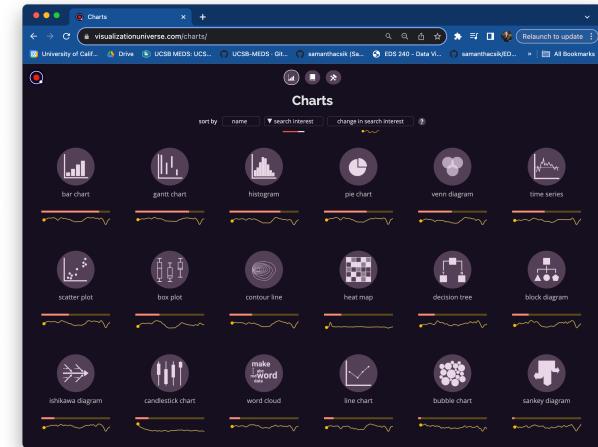
# 2. Number of variables & data points, data types



Data Viz Project displays one small data set 100 different ways



From Data to Viz search graphic types by data type or by function (+ R & Python Graph Gallery)



The Visualization Universe compares most popular graphic forms

## 2. Number of variables & data points, data types

---

### Quantitative data

#### CONTINUOUS

measured data, can have  $\infty$  values within possible range.



I AM 3.1" TALL  
I WEIGH 34.16 grams

#### DISCRETE

OBSERVATIONS CAN ONLY EXIST AT LIMITED VALUES, OFTEN COUNTS.



I HAVE 8 LEGS  
and  
4 SPOTS!

@allison\_horst

**Continuous variables:** temperature (10.6°C, 14.9°C, 8.1°C), rainfall (1.7", 3.3", 9.4")

**Discrete variables:** # of species counted in a region (1, 4, 6), a county's population size (1,578, 10,324, 540,013)

# 2. Number of variables & data points, data types

## Quantitative data

### CONTINUOUS

measured data, can have  $\infty$  values within possible range.



I AM 3.1" TALL

I WEIGH 34.16 grams

### DISCRETE

OBSERVATIONS CAN ONLY EXIST AT LIMITED VALUES, OFTEN COUNTS.



I HAVE 8 LEGS  
and  
4 SPOTS!

@allison\_horst

## Qualitative data

### NOMINAL

UNORDERED DESCRIPTIONS



### ORDINAL

ORDERED DESCRIPTIONS



### BINARY

ONLY 2 MUTUALLY EXCLUSIVE OUTCOMES



@allison\_horst

**Continuous variables:** temperature (10.6°C, 14.9°C, 8.1°C), rainfall (1.7", 3.3", 9.4")

**Discrete variables:** # of species counted in a region (1, 4, 6), a county's population size (1,578, 10,324, 540,013)

**Nominal variables:** gender identity (cisgender, transgender, non-binary), species (dog, cat, bird), land use (residential, parks, agriculture)

**Ordinal variables:** income level (low / middle / high), satisfaction level (unsatisfied, neutral, satisfied)

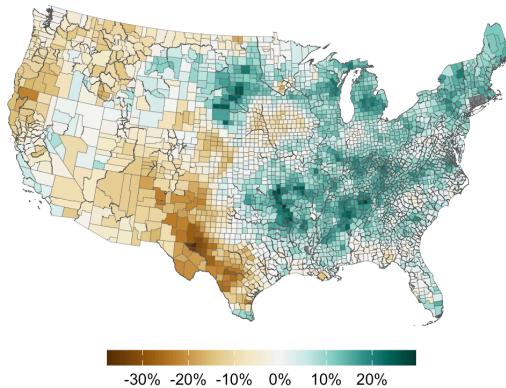
**Binary:** penguin sex (male / female), habitat type (shade / sun)

# 3. Try different graphic forms

---

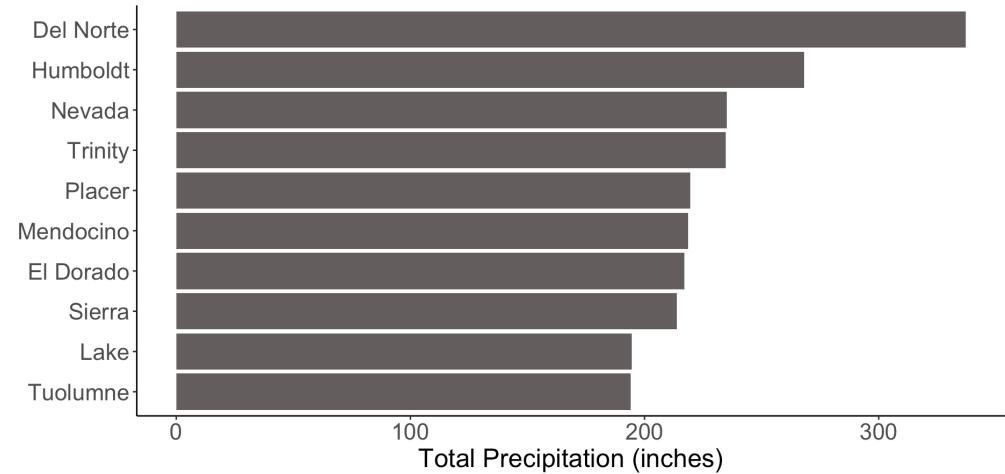
If we want to show both big picture patterns and detailed comparisons, we may **consider including multiple graphic forms in the same visualization.**

5-year precipitation compared to the 20th century average  
January 2019 - December 2023



Source: National Centers for Environmental Information

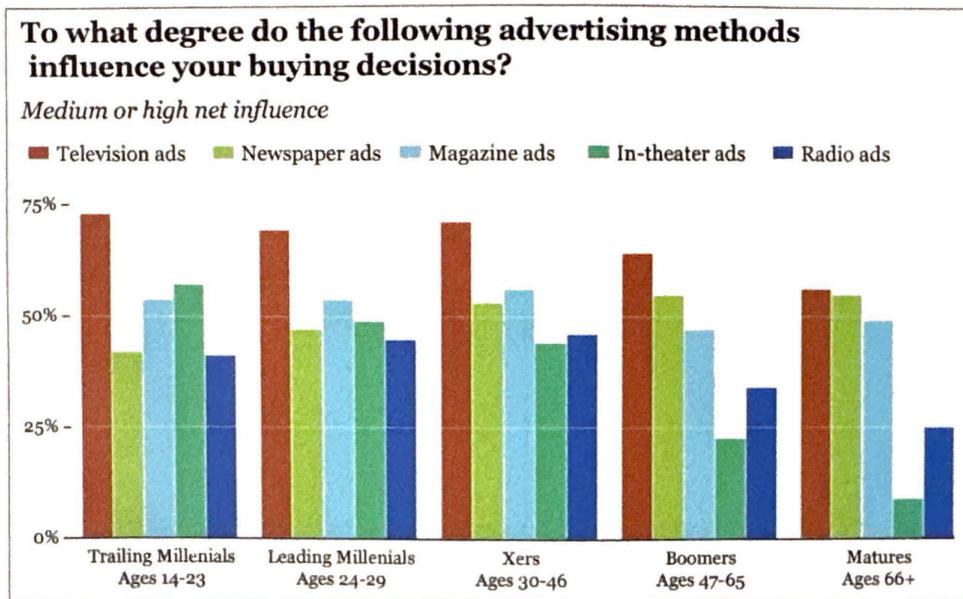
Top 10 California counties to receive the most rainfall, Jan 2019 - Dec 2023



# 4. Arrange components of the graphic

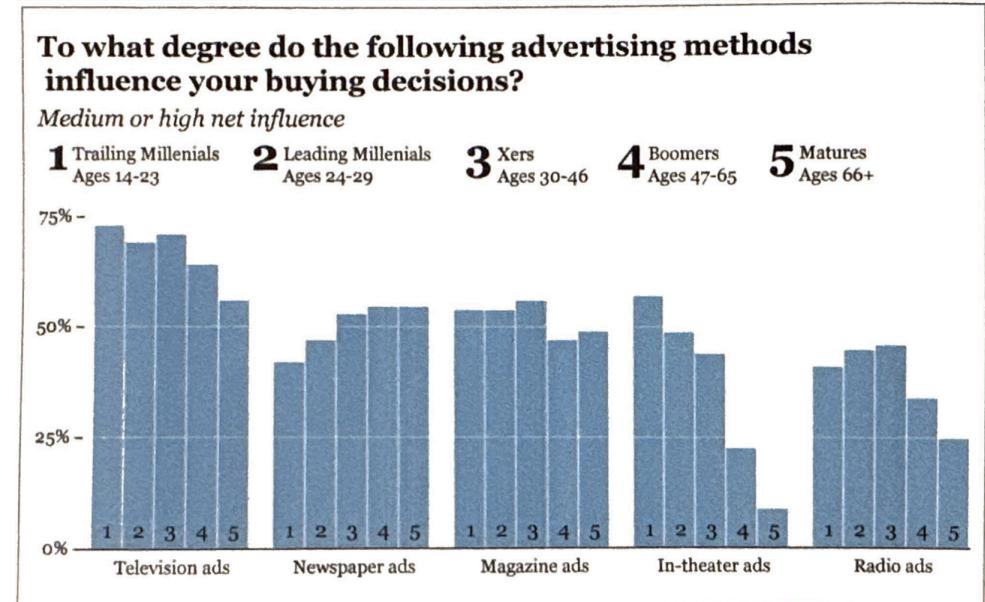
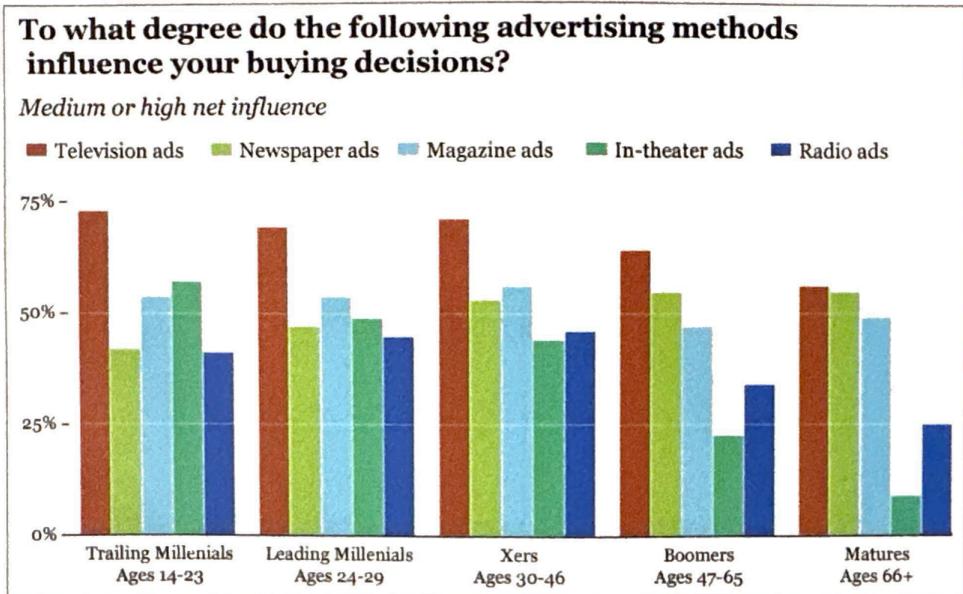
---

How does the influence of **in-theater advertising** change across generations?



# 4. Arrange components of the graphic

How does the influence of **in-theater advertising** change across generations?

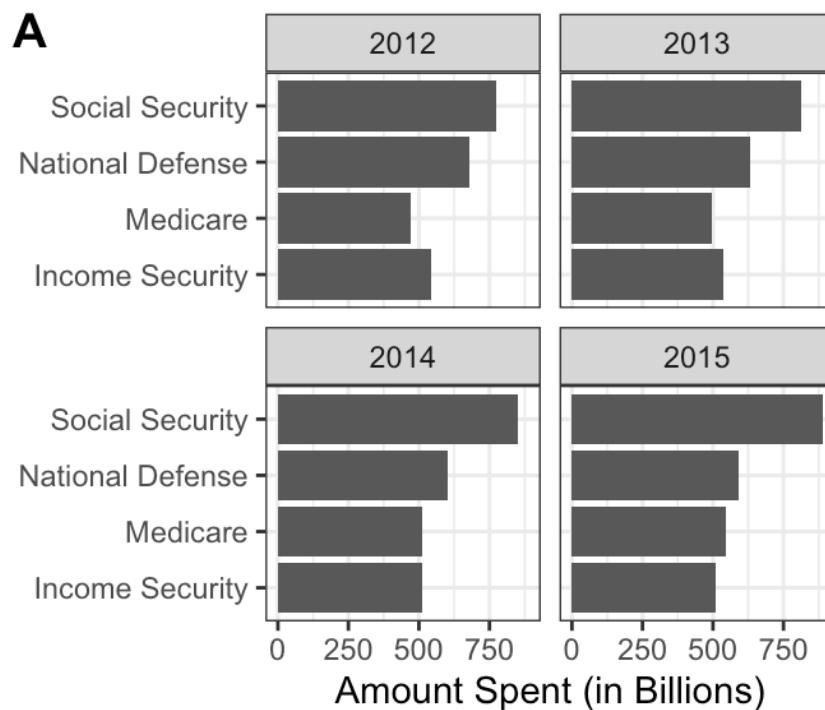


# 4. Arrange components of the graphic

---

**Let's say we're interested in:**

- a. changes in the amount spent on Social Security relative to other major spending categories over time (2012-2015)?

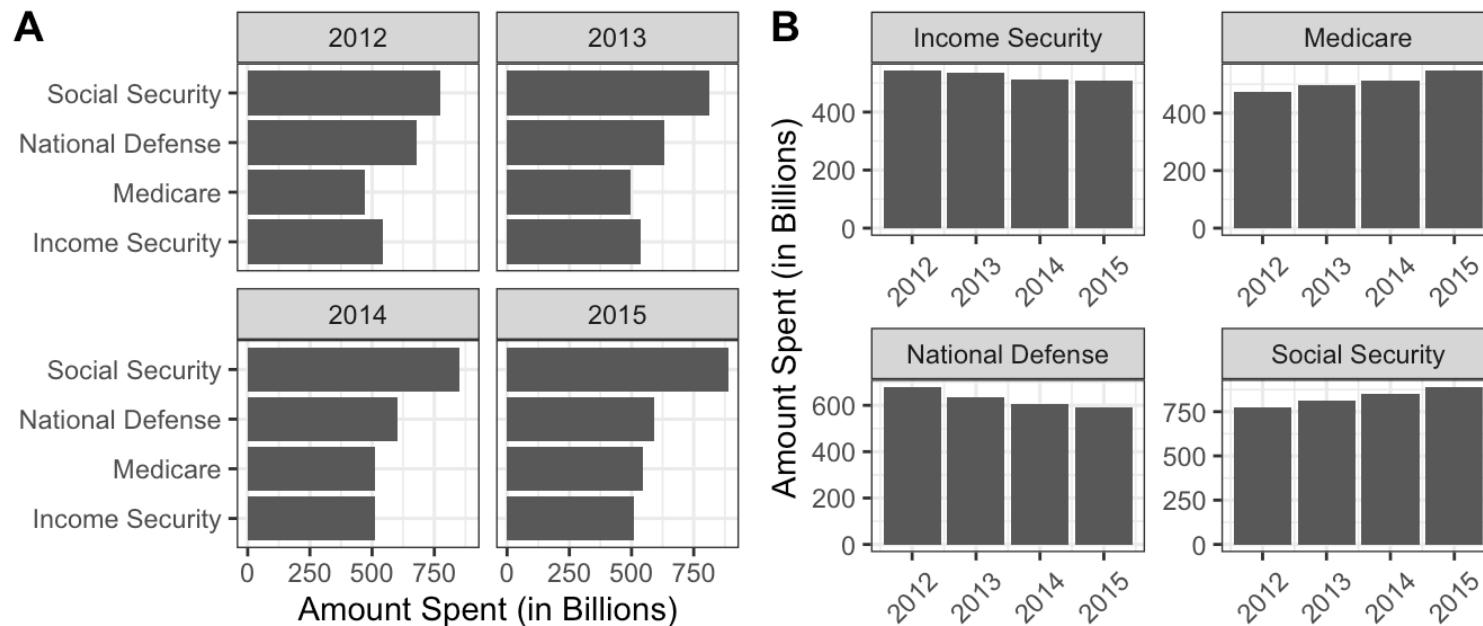


# 4. Arrange components of the graphic

---

**Let's say we're interested in:**

- a. changes in the amount spent on Social Security relative to other major spending categories over time (2012-2015)?
- b. the amount of money spent on Social Security over time (2012-2015)?



# 4. Arrange components of the graphic

---

**Do we want to convey:**

- a. internet usage in 2016? or

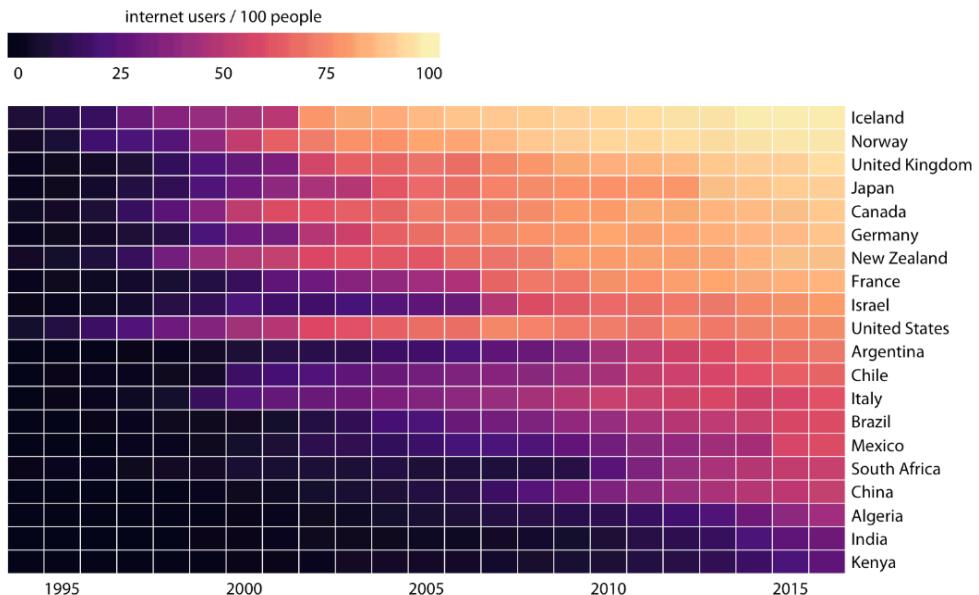


Figure 6.14: Internet adoption over time, for select countries. Color represents the percent of internet users for the respective country and year. Countries were ordered by percent internet users in 2016. Data source: World Bank

# 4. Arrange components of the graphic

Do we want to convey:

- a. internet usage in 2016? or
- b. how early or late adoption of internet relates to current-day usage?

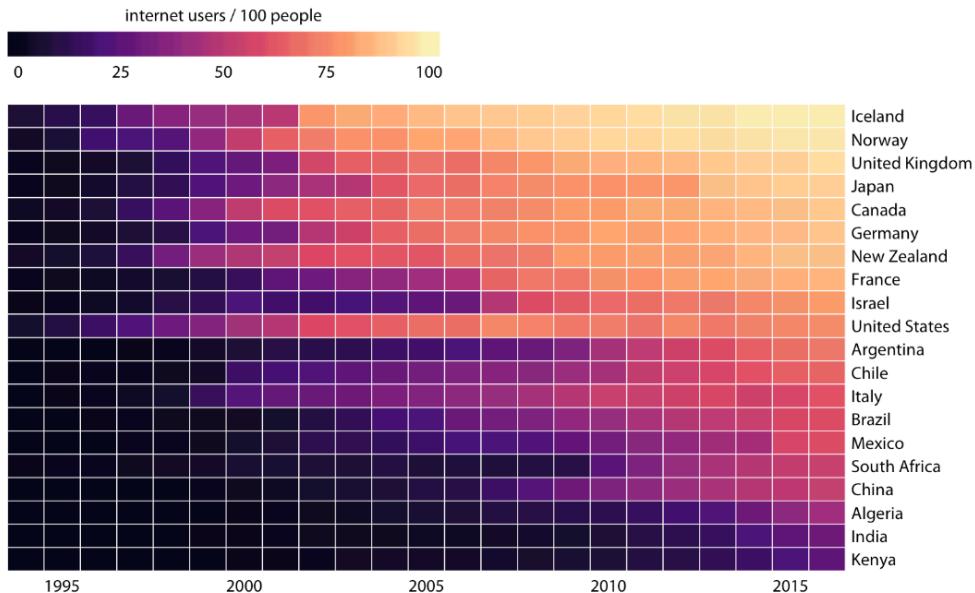


Figure 6.14: Internet adoption over time, for select countries. Color represents the percent of internet users for the respective country and year. Countries were ordered by percent internet users in 2016. Data source: World Bank

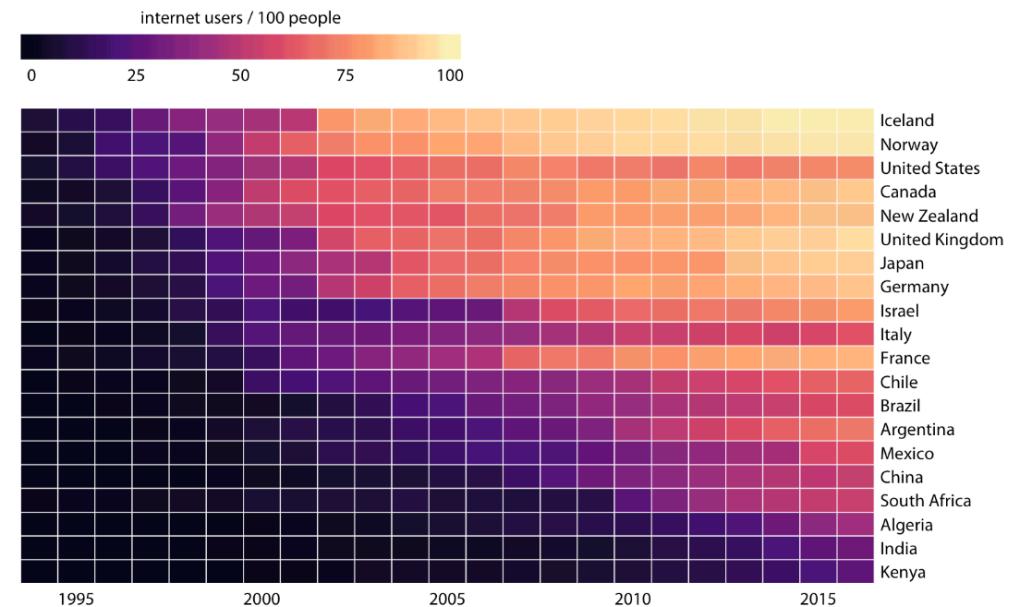
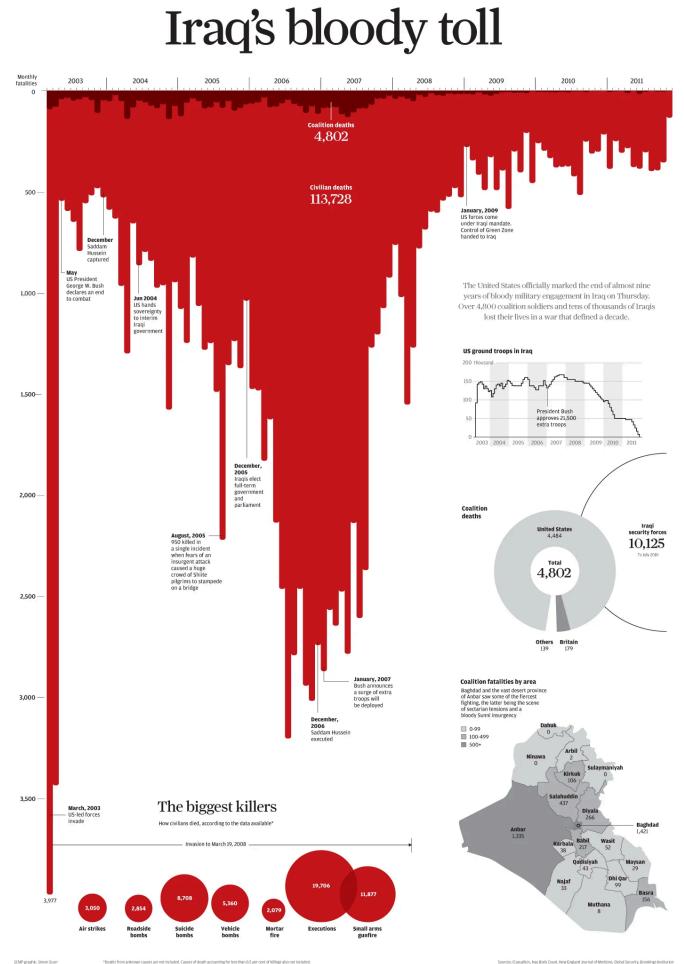


Figure 6.15: Internet adoption over time, for select countries. Countries were ordered by the year in which their internet usage first exceeded 20%. Data source: World Bank

## 5. Test the outcomes of your graphic on others



**To enlarge image (in Chrome), right click on image > Open image in New Tab**

- What is the take home message of this graphic?
  - What is effective? What is confusing?

Source: Iraq's bloody toll, by Simon Scarr

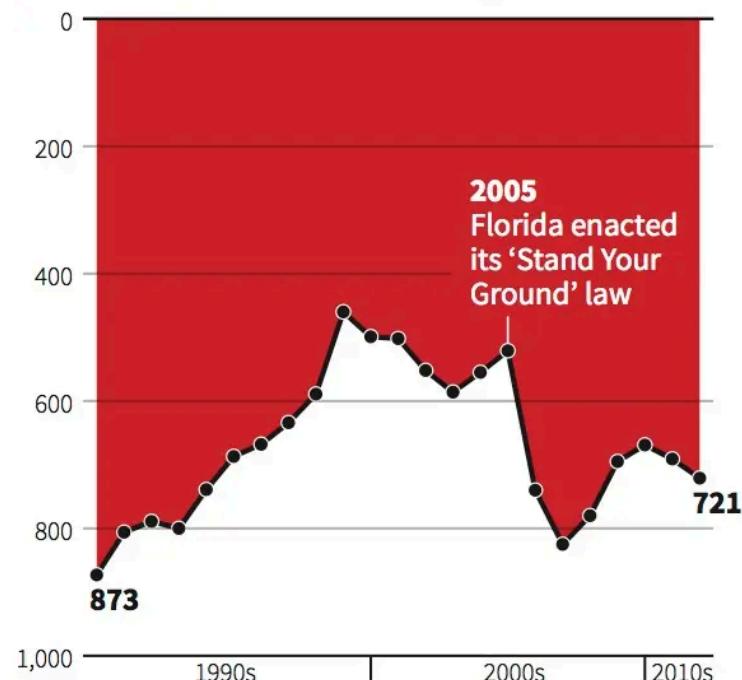
02 : 30

## 5. Test the outcomes of your graphic on others

---

### Gun deaths in Florida

Number of murders committed using firearms



Source: Florida Department of Law Enforcement

C. Chan 16/02/2014

REUTERS

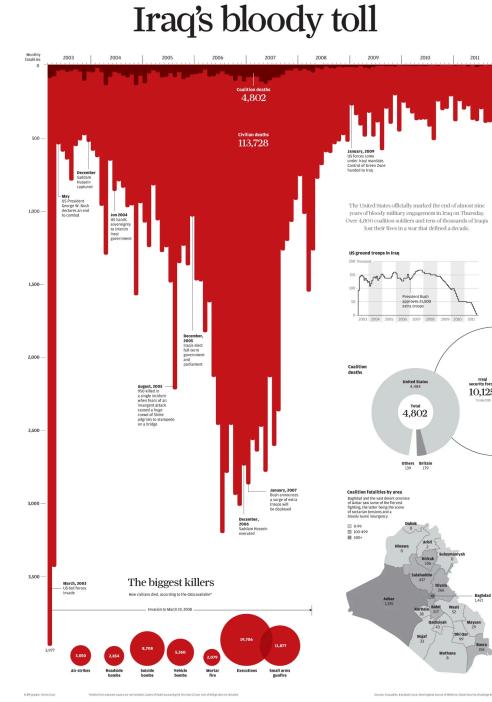
To enlarge image (in Chrome), right click on image > Open image in New Tab

- What is the take home message of this graphic?
- What is effective? What is confusing?

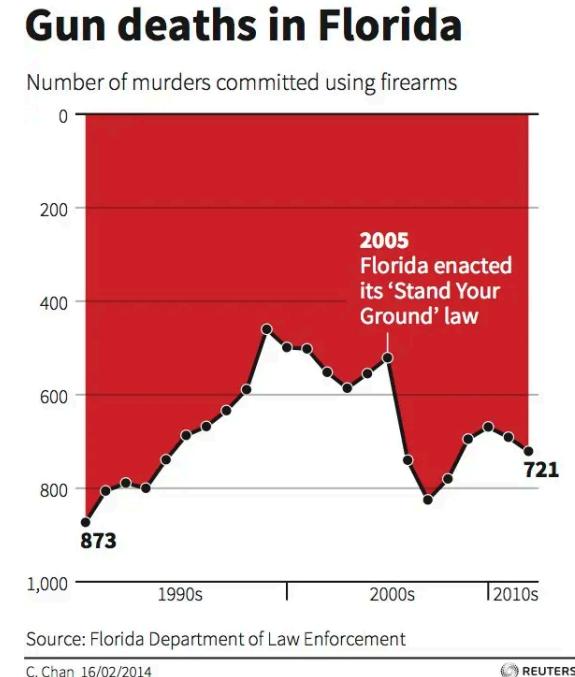
Source: This Chart Shows An Alarming Rise In Florida Gun Deaths After 'Stand Your Ground' Was Enacted.

02 : 30

## 5. Test the outcomes of your graphic on others



Baseline at the *top* is clear, suggesting that bars are *falling* from it. Clear metaphor (dripping blood).



Eyes are drawn to baseline at the *bottom*, on top of which data are sitting. Headline indicates *rise* but visually represented by *falling*. Thick black line makes white area stand out over red (data).

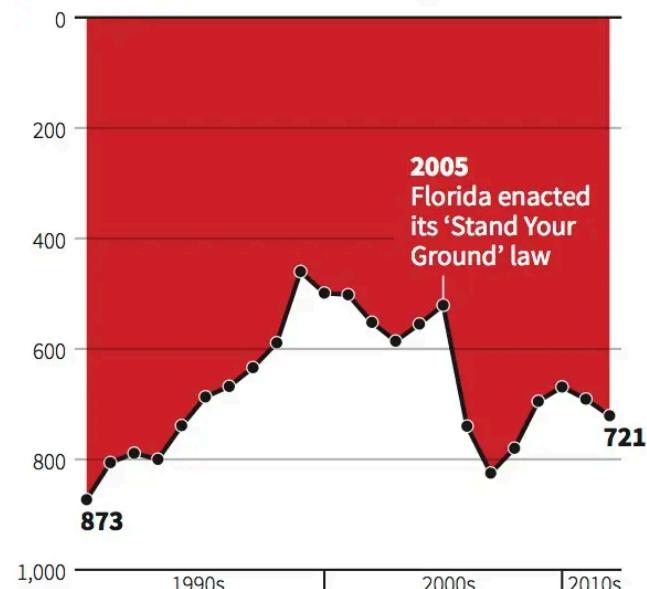
## 5. Test the outcomes of your graphic on others

---

Business Insider [published](#) an updated graphic (originally designed by Reuters), which was submitted by a reader that, **“more clearly shows that gun deaths increased between 2005 and 2007 by flipping the y-axis”**:

## Gun deaths in Florida

Number of murders committed using firearms



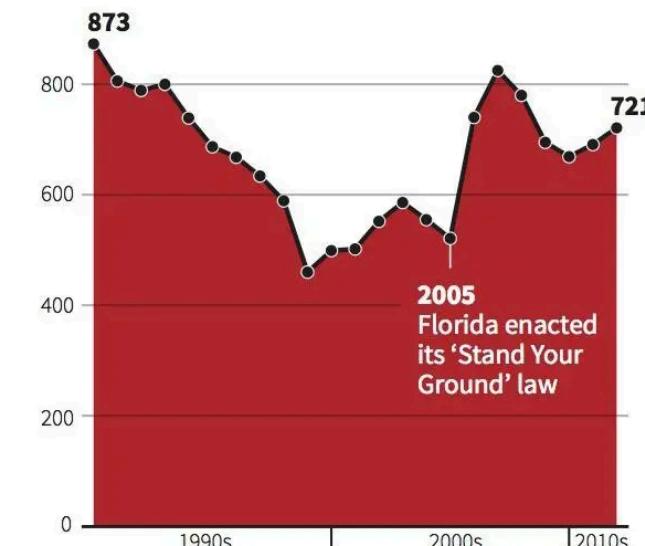
Source: Florida Department of Law Enforcement

C. Chan 16/02/2014

REUTERS

## Gun deaths in Florida

Number of murders committed using firearms



Source: Florida Department of Law Enforcement

 Take a Break

~ *This is the end of Lesson 1 (of 3)* ~

05 : 00