

Data Visualization & Design

Week 5

This week in **visualization**...

1. Tableau **Review**
2. **Ethics** of Data, Data Visualization, & Digital Storytelling
3. Critique
4. Data **Cleaning**

1. Tableau **Review**
2. **Ethics** of Data, Data Visualization, & Digital Storytelling
3. Critique
4. Data **Cleaning**

- Pivot tables
- Importing messy headers
- Hiding and showing categories
- Interactivity

1. Tableau **Review**
2. **Ethics** of Data, Data Visualization, & Digital Storytelling
3. Critique
4. Data **Cleaning**

Example—

“U.S. Gun Deaths” by Periscopic

U.S. GUN DEATHS IN

2013 2010

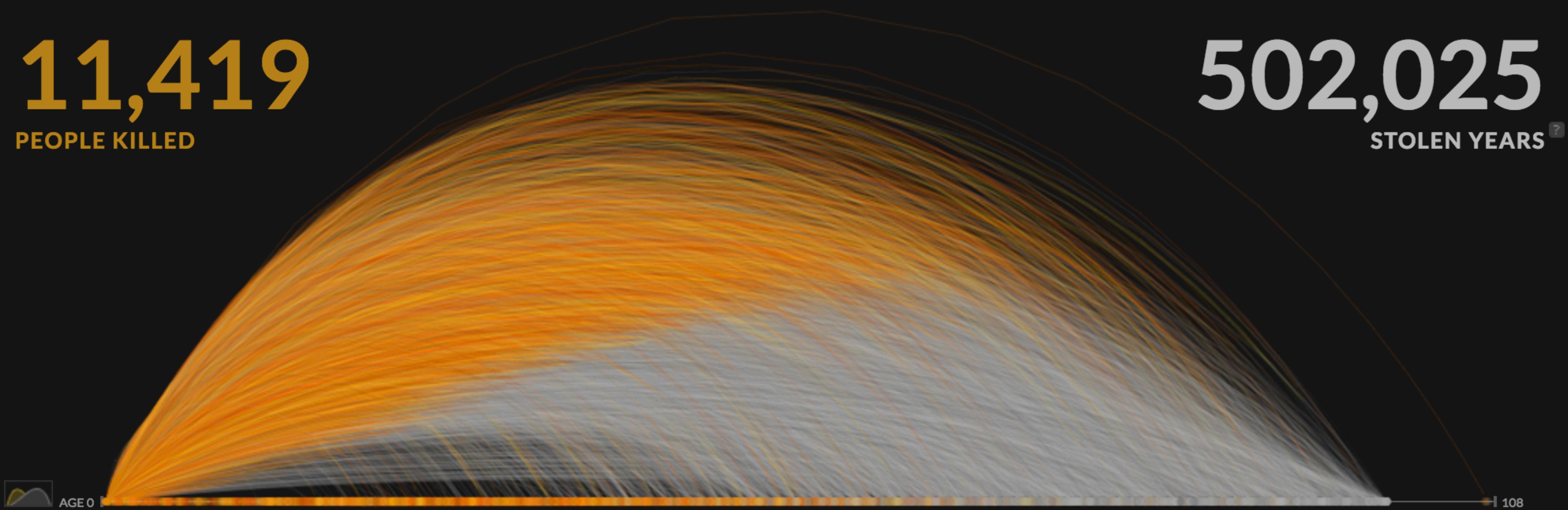
DECEMBER

11,419

PEOPLE KILLED

502,025

STOLEN YEARS 



AGE 0

108



Dataset **Construction** & Data **Analysis**

Before data becomes a visualization, it exists as a series of values that **purports to represent the world** in a formularized way.

One can collect (or construct) data about **anything**.

Data frequently takes the form of information about humans, on topics both seemingly **harmless** and potentially **harmful**.

Movie genre preferences in Europe and Asia

vs.

Immigration status of U.S. residents in a given neighborhood

Data Humanism

- We have a responsibility toward representing the information we have as accurately and transparently as we can.
- When it comes to analyzing or visualizing information that represents different aspects of the lives of individual people, this responsibility becomes even more pressing.
- Errors or carelessness in representation can give rise to more serious consequences.

RD 101: Responsible Data Principles



GUIDES / JANUARY 24, 2018



1. Power Dynamics

- "The least powerful actors in any situation are often the first to see unintended consequences of data collected about them."
- Always keep in mind that **a dataset represents the priorities of the party that had the means to collect that data**, and that just because it exists does not mean that the subjects represented willingly agreed to be represented.
- In a humanitarian crisis, for example, those collecting data about people experiencing the crisis have far more power than the people about whom the data is being collected. How might this affect the way the information is portrayed?

2. Diversity and Bias

- Data has blind spots.
- Consciously striving for diversity in methods, perspectives, and approaches sets the groundwork for thorough and resilient research efforts.
- Recognizing that a dataset is *not* objective (despite how formulaic it seems) is the first step to opening up a space for different voices and sources of information.
- **Do not take data at face value;** question where it comes from and why, and what it might be missing (or excluding).

3. Unknown Unknowns

- We can't *predict* the effect a project will have on its subject matter or its intended audience.
- Anticipating its potential consequences and thinking through ways of mitigating the negative ones are good places to start.

4. Precautionary Principle

- **Just because we can** use data in a certain way, doesn't necessarily mean we *should*.
- Consider what effect our representation of the world can have on the subjects we're representing (ex. Movie genres; race and income).
- Rather than point to a prescriptive way of handling data-driven research, this principle encourages us to question the implications of the decisions we make at every step.

5. Thoughtful Innovation

- The impulse to come up with a data-driven solution to an existing problem is enticing: when data is comprised of *actual numbers*, it seems to imply a similarly computable 'solution'.
- When creating something new, thoughtfulness and consideration are just as important as speed and sound equations.
- **Questions to ask:**
 - Is the solution you propose the only way to solve the problem you want to tackle?
 - Is the data you are working with reliable and representative enough to serve as material?

Data Representation

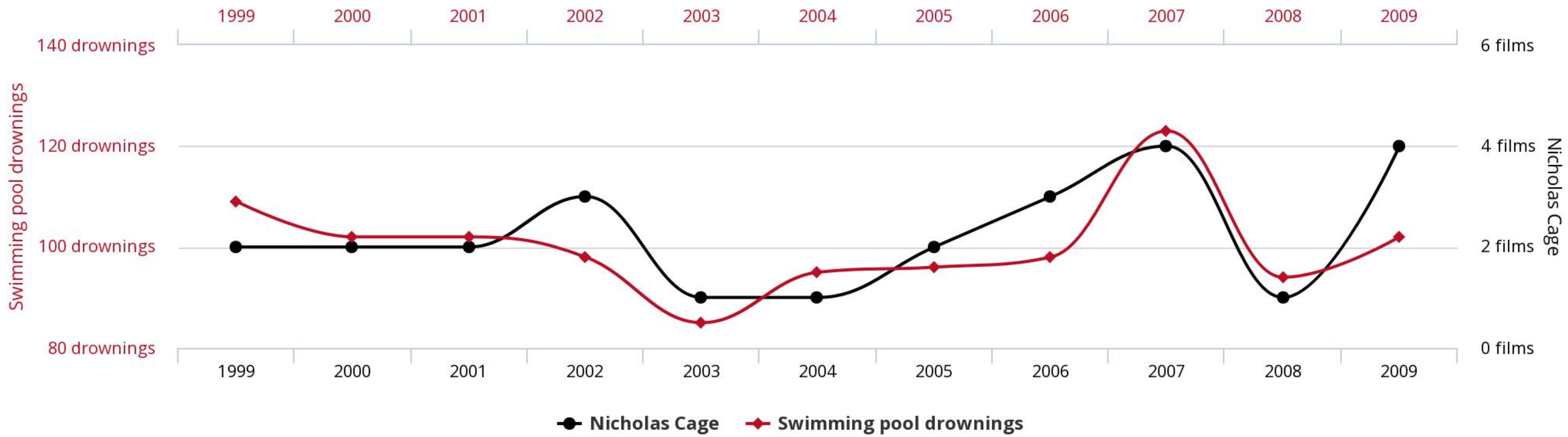
Much of twentieth century thinking about statistical graphics has been preoccupied with the question of how some amateurish chart might fool a naive viewer.

– Edward Tufte

Number of people who drowned by falling into a pool

correlates with

Films Nicolas Cage appeared in



This makes sense!

- Well-constructed visualizations make complex information seem simple.
- Clean charts, beautiful maps, and detailed infographics combine what looks like rigorous, objective research with polished design.
- Data graphics evoke a sense of **implicit authority**, in part due to the regimented format associated with data as a medium.
- Tufte is right—**charts can spread *misinformation*** just as easily as they can spread information.

However, there is **something else** at work when we consider the data graphic as an artifact.

While charts and graphs are well-positioned to distort information for a viewer, **why do we expect more fidelity from a data graphic** than we might from, for example, a political cartoon?

*Vision in this technological feast becomes unregulated gluttony; all perspective gives way to infinitely mobile vision, which no longer seems just mythically about the god-trick of **seeing everything from nowhere**, but to have put the myth into **ordinary practice**.*

– Donna Haraway

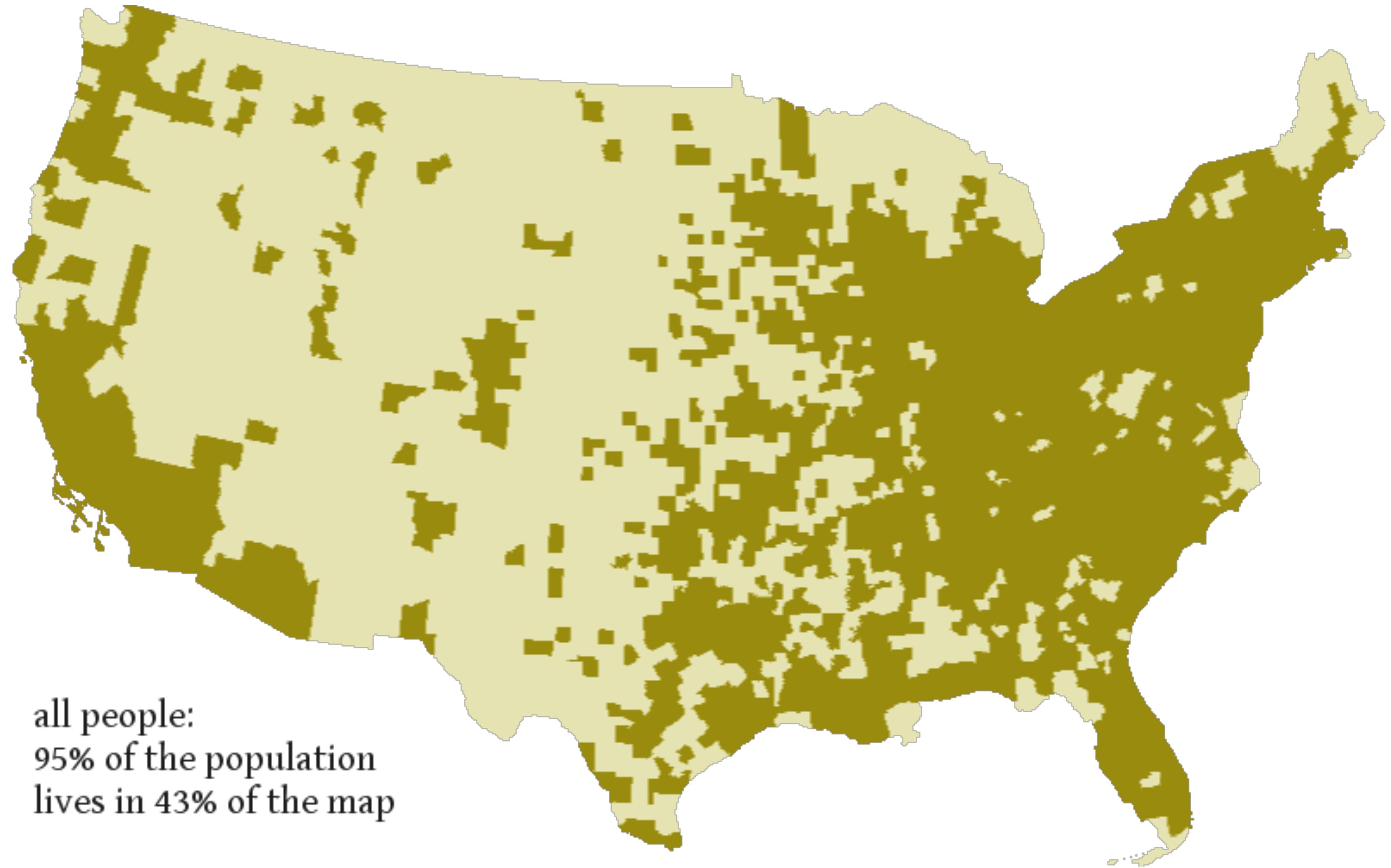


- To Haraway, the "god-trick" is at work in visualizations that grant the viewer an **impossible, disembodied perspective of the world**.
- This kind of perspective makes its subject matter seem like it **simply came into being**, as an effect of the data behind it—the same data that was structured and collected by an individual or an organization, and rendered by a designer in an evocative portrayal.
- Seeming authority that arises from designer's transition from creator to 'translator.'

Authorship displaced by **apparent
automatism.**

Joshua Tauberer—

“How that map you saw on FiveThirtyEight silences minorities, and other reasons to consider a cartogram”



all people:
95% of the population
lives in 43% of the map

“...for every pixel that represents a white person,
only 0.53 pixels represent a racial minority.”

Small & Impactful Considerations

- Always include a *legend* that describes your visualization method and the **units you are using**
- Always list your *data source(s)*
- Always *credit* any externally-sourced graphics or images
- Always list *yourself* as the author

1. Tableau **Review**
2. **Ethics** of Data, Data Visualization, & Digital Storytelling
3. Critique
4. Data **Cleaning**

1. Tableau **Review**
2. **Ethics** of Data, Data Visualization, & Digital Storytelling
3. Critique
4. Data **Cleaning**

The Quartz guide to bad data –

Issues your source should solve VS.
Issues you should solve

Handy **Data Cleaning** Excel Functions

A1

X

fx | Movie,Genre,Production Budget (millions),Box Office (millions),ROI,Rating IMDB

Before anything else—

MAKE A COPY!!!!

1. Delete all formatting

- Remove colors, highlights, fonts, character styles, hyperlinks, and comments
- **Steps**
 - Select the data you want to parse
 - Go to **Edit > Clear > Clear Formats** (or others as needed)

2. Remove duplicates

- Either **Highlight** or **Delete** duplicate data
- **To highlight**
 - Select the data and go to **Home > Conditional Formatting > Highlight Cells Rules > Duplicate Values**
 - Specify the formatting and all the duplicate values get highlighted.
- **To delete**
 - Select the data and Go to **Data > Table Tools > Remove Duplicates**
 - If your data has headers, ensure that the checkbox at the top right is checked.
 - Select the Column(s) from which you want to remove duplicates and click OK
- If deleting data, make sure to copy the dataset to another location to leave the original intact.

3. Parse data using text to columns

- Depending on the format you import, all the data you're working with might show up clustered within a single cell.
- This function enables you to reformat chunks of data into a manageable spreadsheet format.
- **Steps**
 - Select the data you want to parse
 - Go to **Data > Text to Columns**
 - Select the data type (usually “Delimited”—fields separated by commas or other characters)
 - Select delimiter, the character that separates your data
 - Select data format and destination cell

4. Select and treat all blank cells

- Blank spaces should *not necessarily* be populated with zeroes.
 - You may want to fill this spaces with an “N/A”, “--”, or something else.
- **Steps**
 - Select the entire dataset
 - Click **Edit > Go To > Special...**
 - Select “Blanks”
 - Click “OK”
 - Type the value you want to populate these cells with and press **Control + Enter** so the value appears in every cell

5. Convert numbers stored as text into numbers

- Sometimes, importing CSV or database files into Excel results in numbers that are imported as *strings*
- Also, sometimes numbers are preceded by an apostrophe (‘) to ensure they are treated as text—this can cause problems with calculations
- **Steps** (one way to get around this)
 - In any blank cell, type 1
 - Select the cell where you typed 1, and press **Control + C**
 - Select the cell or range that you want to convert to numbers
 - Select Paste > Paste Special
 - Select **Multiply** (under “Operations”)
 - Click OK

6. Highlight errors

- Highlighting errors enables you to delete, reformat, or manually highlight detected cells.
- **Steps**
 - Select the entire dataset
 - Click Find & Select > Go To Special...
 - Select “Formulas” and uncheck all options except “Errors”
 - Click “OK”

7. Get rid of extra spaces and non-printable characters

- Trailing spaces and extra room between labels can be difficult to spot and troubleshoot manually.
- **TRIM()**—Takes text as input. Strips leading and trailing spaces as well as additional spaces between words. Leaves only one space between words.
- **CLEAN()**—Removes non-printable characters.

8. Change text to Lower/Upper/Proper case

- When you inherit working files, name formatting and capitalization are often inconsistent
- **LOWER()** – Converts all text to lower case.
- **UPPER()** – Converts all text to UPPER CASE.
- **PROPER()** – Converts all text to Proper Case.

9. Spell check

- Use it.

(CLEANUP)

-