Introduction to
Telling Stories with Data

Author: Nicholas G Reich

What are data?

da·tum
/ˈdātəm,ˈdatəm/ ◄))

noun
plural noun: **data**
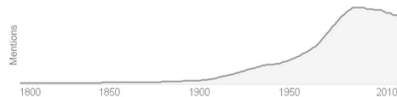
1. a piece of information.
   - an assumption or premise from which inferences may be drawn.

2. a fixed starting point of a scale or operation.

Origin

LATIN	LATIN	
dare	datum	datum
give	something given	*mid 18th century*

mid 18th century: from Latin, literally 'something given,' neuter past participle of *dare* 'give.'

Use over time for: data

Mentions

1800   1850   1900   1950   2010

Data as building blocks

▶ Data are units of information, encoded for processing, e.g. "machine-readable"

▶ Data, like words, can be woven together to create new conceptual understanding

▶ Data are the raw material of the digital economy.

Communicating ideas with evidence
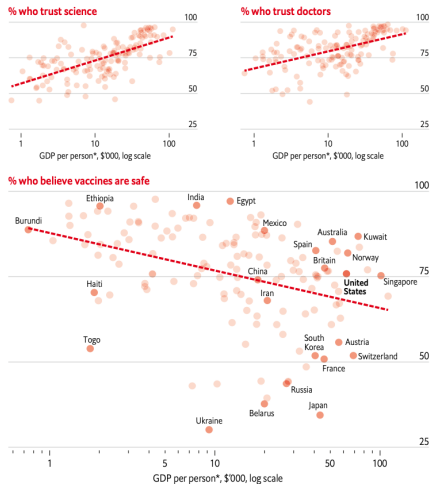
What is a narrative? [From the OED]

An account of a series of events, facts, etc., given in order and with the establishing of connections between them; a narration, a story, an account.

Telling stories with data

- raw material: words, data
- tools: code, software, computers
- finished product: visualization, explanations, inference

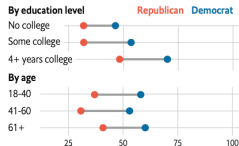→ **Rich countries have more trust in science and doctors, but less in vaccines**

Trust v GDP per person*, 2018, by country

% who trust science

GDP per person*, $'000, log scale

% who trust doctors

GDP per person*, $'000, log scale

% who believe vaccines are safe

GDP per person*, $'000, log scale

United States, % responding "yes"
By political alignment, July-August 2020

"If and when a coronavirus vaccine becomes available, will you get vaccinated?"

By education level   Republican   Democrat
No college
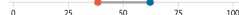Some college
4+ years college

By age
18-40
41-60
61+

"In general, do you think vaccines given to children for diseases like measles are safe?"

"Do you think parents should be required to have their children vaccinated?"

"If and when a coronavirus vaccine is available, will you get your school-aged children vaccinated?"

"Do you believe it would be safe for the country to fully reopen, before a vaccine became available?"

Sources: Wellcome Trust; World Bank; YouGov/The Economist
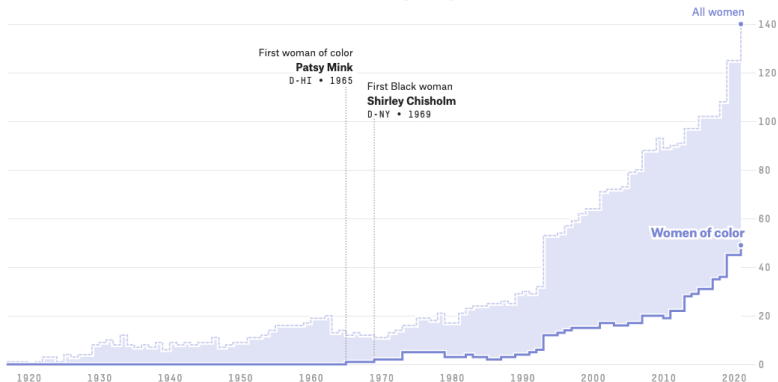*At purchasing-power parity

Women of Color Were Shut Out of Congress For Decades. Now They're Transforming It.

By Meredith Conroy, Amelia Thomson-DeVeaux and Anna Wiederkehr

Illustration by Chelsea Alexander

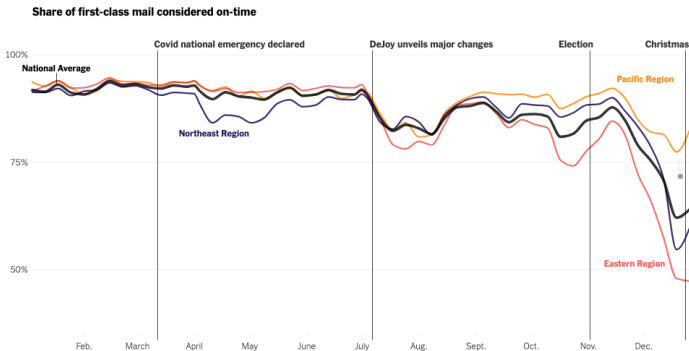**It wasn't until 1965 that a woman of color arrived in Congress**

The number of women serving in Congress since 1917

All women

First woman of color
Patsy Mink
D-HI • 1965

First Black woman
Shirley Chisholm
D-NY • 1969

Women of color

The Postal Service Survived the Election. But It Was Crushed by Holiday Packages.

By Emily Badger, Quoctrung Bui and Margot Sanger-Katz   Jan. 19, 2021

Share of first-class mail considered on-time

Note: Includes three of the seven Postal Service regions. Region definitions can be found here. Source: United States Postal Service

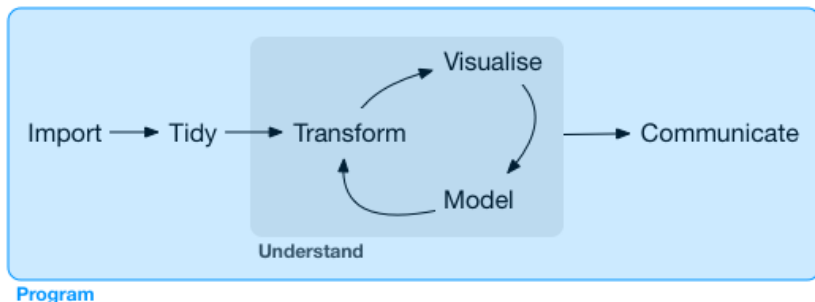# How to tell a story using data

Telling stories with data requires

- a topic for the story you want to tell
- detective work
- creativity, both scientific and artistic
- experimentation with different storylines
- statistical literacy
- good data (good data does not necessarily equal "big data")

Common mistakes in data storytelling

- ▶ not knowing the audience
- ▶ making the story too complex too quickly
- ▶ encouraging correlation to be seen as causation
- ▶ missing the little things (clear axis labels, plot annnotations)
- ▶ not making the results reproducible

A process for data analysis

This is where we are going to spend most of our time this semester. With a bit of a focus on public health applications.