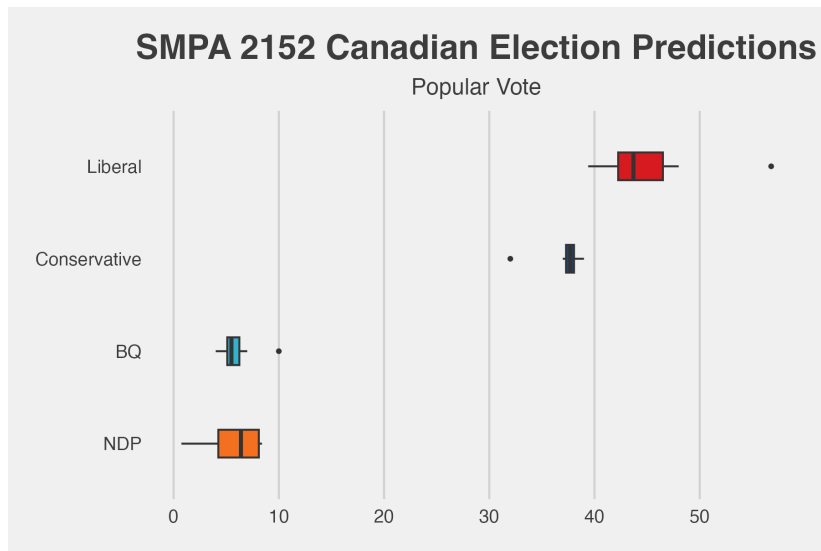


WRAPPING UP

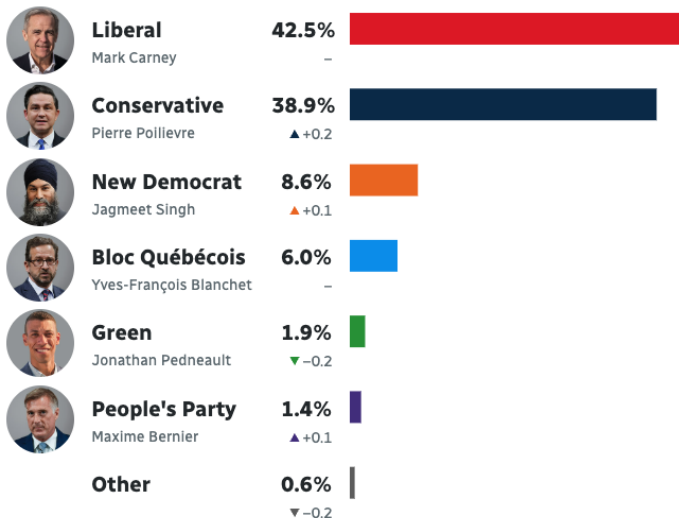
Data Analysis for Journalism and Political Communication
(Spring 2025)

Prof. Bell

CANADIAN ELECTION PREDICTIONS



CANADIAN ELECTION PREDICTIONS



THINK-PAIR-SHARE

- 1 What are the most important ideas that you will take away from this class?
- 2 What burning questions do you still have about data analysis?

CHALABI: 3 WAYS TO SPOT A BAD STATISTIC



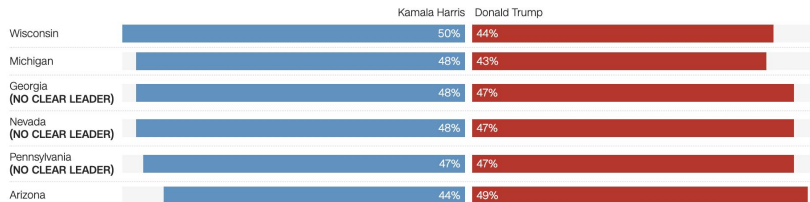
- 1 Can you see uncertainty?
- 2 Can we look beyond the averages?
- 3 How was the data collected?

CAN YOU SEE THE UNCERTAINTY?

- Whenever we take a sample and try to make inferences about a population, there is uncertainty

New polling shows a close race in several key swing states

In a set of CNN polls conducted by SSRS, likely voters say they're currently supporting:



Methodology: The CNN polls were conducted by SSRS from August 23-29 using representative samples of registered voters in Arizona (n=682), Georgia (n=617), Michigan (n=708), Nevada (n=626), Pennsylvania (n=789) and Wisconsin (n=976). Results for likely voters include all registered voters interviewed for the poll weighted by their predicted probability of voting in the 2024 general election.

Results among the full sample in each state have a margin of sampling error of up to ± 4.9 percentage points.

Source: CNN/SSRS polling

Graphic: Ariel Edwards-Levy, CNN

CAN YOU SEE THE UNCERTAINTY?

- Whenever we take a sample and try to make inferences about a population, there is uncertainty
- There is also uncertainty when making causal claims due to confounding and reverse causation



CAN WE LOOK BEYOND THE AVERAGES?

- We need to carefully communicate the context of the data

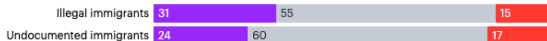
Which of the following groups do you have a favorable opinion of?
Which of the following groups do you have an unfavorable opinion of? (%)

Favorable Neither favorable nor unfavorable Unfavorable

U.S. adult citizens



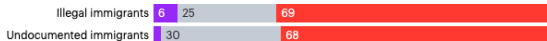
Democrats



Independents



Republicans



YouGov

December 20, 2023 - January 17, 2024

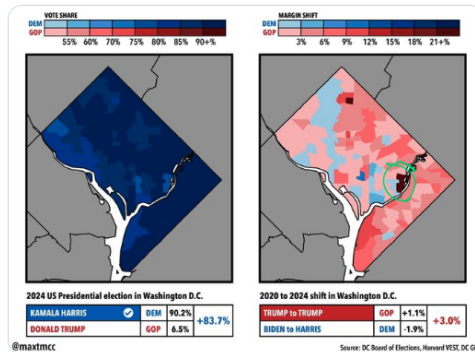
CAN WE LOOK BEYOND THE AVERAGES?

- We need to carefully communicate the context of the data



Jake Rubinstein @Jlrube · Nov 19, 2024

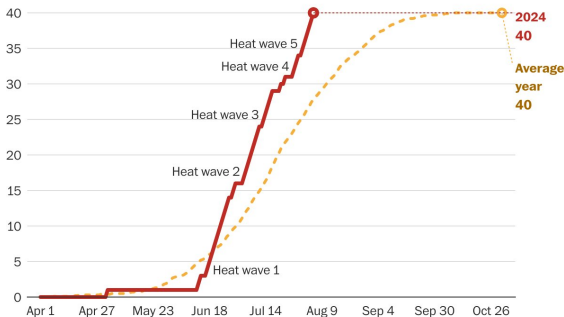
Fun story, this contains the jail with J6 defendants



CAN WE LOOK BEYOND THE AVERAGES?

- We need to carefully communicate the context of the data
- There is both a science and an art to accurately communicating about data using data visualizations

Days reaching at least 90 degrees in Washington



A heat wave is labeled when three or more consecutive days reach or surpass 90 degrees. Includes data through Aug. 5. This chart will be updated regularly throughout the season.

Source: Applied Climate Information System

IAN LIVINGSTON / THE WASHINGTON POST

HOW WAS THE DATA COLLECTED?



Elon Musk   @elonmusk · May 13, 2022



To find out, my team will do a random sample of 100 followers of [@twitter](#).

I invite others to repeat the same process and see what they discover ...

 10K

 11K

 111K



Tesla Owners Silicon...   @teslaow... · May 13, 2022  ...

How are defining random ? How will you select them

 33

 58

 1.2K



Elon Musk   @elonmusk · May 13, 2022



Ignore first 1000 followers, then pick every 10th. I'm open to better ideas.

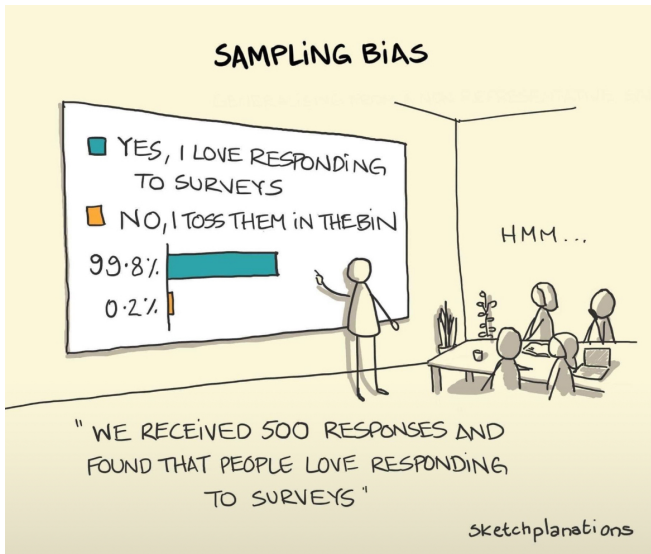
 1.3K

 984

 9.9K



HOW WAS THE DATA COLLECTED?



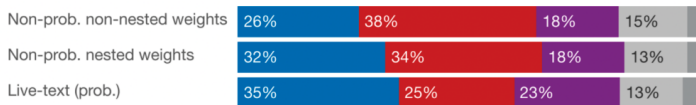
HOW WAS THE DATA COLLECTED?

Young Voters: 2024 Presidential Race



Q. If the 2024 presidential election was today, who would you vote for?

■ Joe Biden ■ Donald Trump ■ Robert F. Kennedy, Jr. ■ I don't know ■ I wouldn't vote

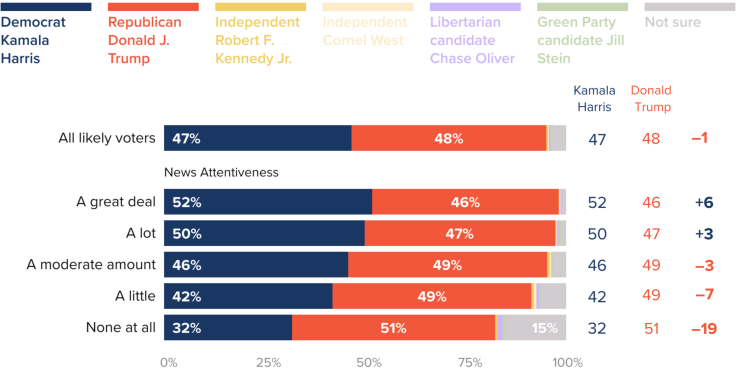


Displaying results among 18-29 likely voters

HOW WAS THE DATA COLLECTED?

Support for Harris Decreases as News Consumption Decreases

If the November 2024 election for U.S. president was being held tomorrow, and these were the candidates, who would you vote for?



Oct 5–Nov 3, 2024 pooled surveys of 13,404 interviews of likely voters

DATA FOR PROGRESS

YOU LEARNED R!

- Base R functions
- Make charts for 3+ variables
- Load data
- Filter data
- Create and change data
- Summarize data
- Pivot data (tidy data)
- Join data
- Weight poll data
- Calculate margins of error
- Conduct t-tests of means
- Conduct t-tests of proportions
- Run regressions
- Create reports