

# INTRODUCTION TO DATA ANALYSIS

Data Analysis for Journalism and Political Communication  
(Spring 2026)

Prof. Bell

# KEY DETAILS

Professor: Nicholas Bell, Ph.D. (he/him)  
nicholasbell@gwu.edu

Office Hours: Tuesdays 5:45 - 6:45pm  
MPA 425

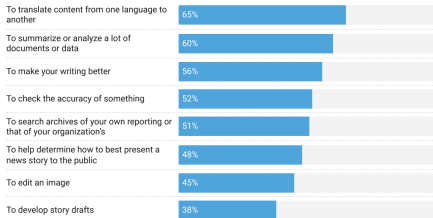
By appointment only. Appointments must be made at least three hours prior to office hours. The scheduling link is <https://tinyurl.com/smpa2152officehours>

If you would like to meet outside of office hours, please email me to schedule an appointment. I try to respond to emails by the end of the next business day (M-F).

# Why Is This Course REQUIRED?

## Journalists use technology for many different purposes

**Question:** In the past year, have you used technology, including AI, in any of the following ways to help you in your work?



*Note: Survey of journalists across 63 countries (N = 433) conducted between October 14, 2024 and December 1, 2024. Percentages factor in attrition. No and Refused/NA responses not shown.*

Chart: Center for News, Technology & Innovation • Source: What It Means to Do Journalism in the Age of AI: Journalist Views on Safety, Technology and Government • Created with Datawrapper

## *Inside the Last-Ditch Hunt by Harris and Trump for Undecided Voters*

Both campaigns are digging through troves of data to find these crucial Americans. They both think many are younger, Black or Latino. The Harris team is also eyeing white, college-educated women.



Listen to this article • 10:16 min [Learn more](#)



Share full article



980



# CHALABI: 3 WAYS TO SPOT A BAD STATISTIC



# 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?

- We rarely get a complete count of everything

# CAN YOU SEE THE UNCERTAINTY?

- We rarely get a complete count of everything
- How do we know that the **sample** we choose is a good representative of the whole **population**?

# CAN YOU SEE THE UNCERTAINTY?

- We rarely get a complete count of everything
- How do we know that the **sample** we choose is a good representative of the whole **population**?
- We will learn how to measure and communicate about the uncertainty inherent in statistics

# CAN YOU SEE THE UNCERTAINTY?

- We rarely get a complete count of everything
- How do we know that the **sample** we choose is a good representative of the whole **population**?
- We will learn how to measure and communicate about the uncertainty inherent in statistics
- Humans do not do well with probability

# CAN YOU SEE THE UNCERTAINTY?

- We rarely get a complete count of everything
- How do we know that the **sample** we choose is a good representative of the whole **population**?
- We will learn how to measure and communicate about the uncertainty inherent in statistics
- Humans do not do well with probability

“There are only five probabilities the average human can handle: 99 percent, one percent, 100 percent, zero, and 50-50. That’s it.”

- Richard Thaler, Nobel Laureate in Economics

# CAN YOU SEE THE UNCERTAINTY?

- We rarely get a complete count of everything
- How do we know that the **sample** we choose is a good representative of the whole **population**?
- We will learn how to measure and communicate about the uncertainty inherent in statistics
- Humans do not do well with probability



# CAN YOU SEE THE UNCERTAINTY?

- We rarely get a complete count of everything
- How do we know that the **sample** we choose is a good representative of the whole **population**?
- We will learn how to measure and communicate about the uncertainty inherent in statistics
- Humans do not do well with probability

**FiveThirtyEight**  
2016 Election Forecast

President  
Updated Nov. 8, 2016

Senate  
Updated Nov. 8, 2016

We're forecasting the election with three models

● Polls-plus forecast

What polls, the economy and historical data tell us about Nov. 8

○ Polls-only forecast

What polls alone tell us about Nov. 8

○ Now-cast

Who would win the election if it were held today

🗳️ National overview

## Who will win the presidency?



### Chance of winning



# CAN WE LOOK BEYOND THE AVERAGES?

- There is always a trade-off between simplicity and precision when working with data

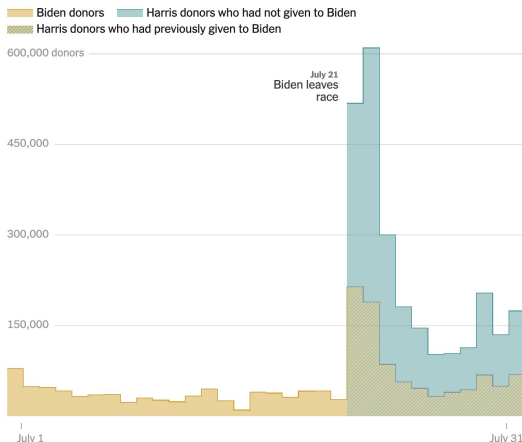
# CAN WE LOOK BEYOND THE AVERAGES?

- There is always a trade-off between simplicity and precision when working with data
- We summarize data to make it easier to comprehend, but we may also lose important context

# CAN WE LOOK BEYOND THE AVERAGES?

## Donors both old and new gave to the newly renamed Harris campaign

Both donors who had given to the Biden re-election campaign and new people who had not previously contributed rushed to donate to the Harris campaign.



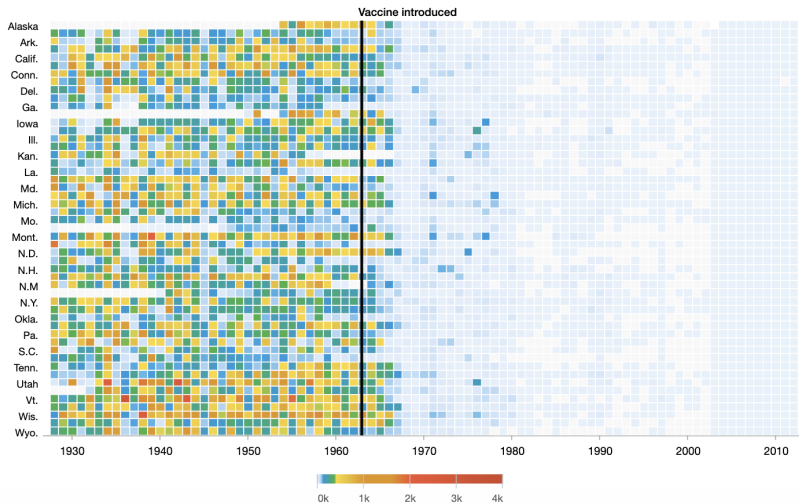
Source: Federal Election Commission • The New York Times

# CAN WE LOOK BEYOND THE AVERAGES?

- There is always a trade-off between simplicity and precision when working with data
- We summarize data to make it easier to comprehend, but we may also lose important context
- We will talk about using data visualization to communicate about data

# CAN WE LOOK BEYOND THE AVERAGES?

## Measles



# CAN WE LOOK BEYOND THE AVERAGES?

- There is always a trade-off between simplicity and precision when working with data
- We summarize data to make it easier to comprehend, but we may also lose important context
- We will talk about using data visualization to communicate about data
- We will also talk about the importance of theory in understanding data, especially correlation vs. causation

## Storks Deliver Babies ( $p = 0.008$ )

### KEYWORDS:

*Teaching:*

*Correlation;*

*Significance;*

*p-values.*

*Robert Matthews*

Aston University, Birmingham, England.

e-mail: rjm@compuserve.com

### Summary

This article shows that a highly statistically significant correlation exists between stork populations and human birth rates across Europe. While storks may not deliver babies, unthinking interpretation of correlation and  $p$ -values can certainly deliver unreliable conclusions.

### ◆ INTRODUCTION ◆

**I**ntroductory statistics textbooks routinely warn of the dangers of confusing correlation with causation, pointing out that while a high correlation coefficient is indicative of (linear) association, it cannot be taken as a measure of causation. Such

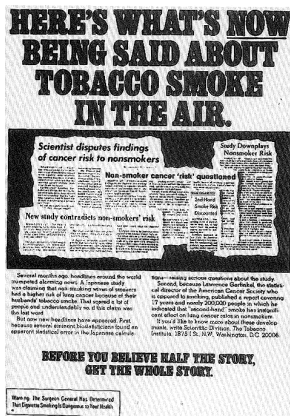
association between storks and the concept of women as bringers of life, and also in the bird's feeding habits, which were once regarded as a search for embryonic life in water (Cooper 1992). The legend lives on to this day, with neonate-bearing storks being a regular feature of greetings cards celebrating births.

# HOW WAS THE DATA COLLECTED?

- Data is not objective – it is generated by humans

# HOW WAS THE DATA COLLECTED?

- Data is not objective – it is generated by humans
- Some data is produced by unscrupulous actors



# HOW WAS THE DATA COLLECTED?

- Data is not objective – it is generated by humans
- Some data is produced by unscrupulous actors
- But most of the time, poor analysis is not nefarious – humans are imperfect

# HOW WAS THE DATA COLLECTED?

- Garbage in = garbage out: no amount of statistical wizardry can compensate for bad data

# HOW WAS THE DATA COLLECTED?

- Garbage in = garbage out: no amount of statistical wizardry can compensate for bad data
- We will spend a lot of time thinking about the **data generating process** and how it can bias our results

## HOW WAS THE DATA COLLECTED?

### Support for mass deportation varies depending on how the question is asked

A sample of different questions asked about deportation this year show significant variation in levels of support — sometimes, even within the same survey.

Pollster	Wording	Support deportations ▼	Do not support deportations
<b>CBS News/YouGov</b> <i>Registered voters, June 5-7</i>	Would you favor or oppose the U.S. government starting a new national program to deport all undocumented immigrants currently living in the U.S. illegally?	<div><div>62%</div></div>	<div><div>38%</div></div>
<b>Marquette Law School</b> <i>Registered voters, October 1-10</i>	Do you favor or oppose deporting immigrants who are living in the United States illegally back to their home countries?	<div><div>58%</div></div>	<div><div>42%</div></div>
<b>ABC/Ipsos</b> <i>US adults, October 4-8</i>	There are at least 11 million undocumented immigrants living in the United States. Would you support or oppose an effort by the federal government to deport all these undocumented immigrants and send them back to their home countries?	<div><div>56%</div></div>	<div><div>43%</div></div>
<b>CNN</b> <i>US adults, January 25-30</i>	If Donald Trump becomes president again, would you favor or oppose him trying to...detain and deport millions of undocumented immigrants?	<div><div>48%</div></div>	<div><div>52%</div></div>
<b>Gallup</b> <i>US adults, June 3-23</i>	Please tell me whether you strongly favor, favor, oppose, or strongly oppose each of the following proposals...Deporting all immigrants who are living in the United States illegally back to their home country	<div><div>47%</div></div>	<div><div>51%</div></div>
<b>Marquette Law School</b> <i>Registered voters, October 1-10</i>	Do you favor or oppose deporting immigrants who are living in the United States illegally back to their home countries even if they have lived here for a number of years, have jobs and no criminal record?	<div><div>40%</div></div>	<div><div>60%</div></div>
<b>Pew Research</b> <i>US adults, April 8-14</i>	Which comes closer to your view about how to handle undocumented immigrants who are now living in the U.S.? (They should not be allowed to stay in the country legally/There should be a way for them to stay in the country legally, if certain requirements are met)		
	If "not be allowed." "Do you think there should be a national law enforcement effort to deport all immigrants who are now living in the U.S. illegally?"	<div><div>33%</div></div>	<div><div>67%</div></div>
	("Support" percentage includes those who say there should be a national deportation effort, "do not support" includes all others)		

# HOW WAS THE DATA COLLECTED?

- Garbage in = garbage out: no amount of statistical wizardry can compensate for bad data
- We will spend a lot of time thinking about the **data generating process** and how it can bias our results
- We will also discuss our ethical responsibilities around data

# HOW WAS THE DATA COLLECTED?

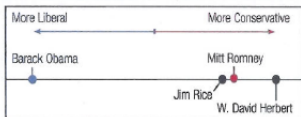


## 2014 Montana General Election Voter Information Guide

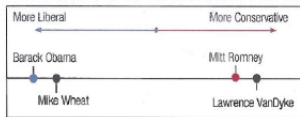
Election Date: November 4, 2014



### Nonpartisan Supreme Court Justice #1 Race



### Nonpartisan Supreme Court Justice #2 Race



For more information on how these figures were created, please see <http://data.stanford.edu/dime>. Please note that this guide is non-partisan and does not endorse any candidate or party. This guide was created as part of a joint research project at Stanford and Dartmouth.

Paid for by researchers at Stanford University and Dartmouth College, 616 Serra Street, Stanford, CA 94305

## Take this to the polls!

# GROUP DISCUSSION

Introduce yourself to your neighbor(s) and take a few minutes to review these additional graphs from Mona Chalabi. Do any of these stand out to you as being good (or bad) examples of our three questions for spotting a bad statistic?

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

# GROUP DISCUSSION

# GROUP DISCUSSION

Introduce yourself to your neighbor(s) and take a few minutes to review these additional graphs from Mona Chalabi. Do any of these stand out to you as being good (or bad) examples of our three questions for spotting a bad statistic?

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

# HOW THE COURSE WILL WORK

Your course grade is calculated as your grade on each of the following course components weighted by:

Attendance	10%
Lab assignments	35%
Class project	20%
Final exam	35%

# LAB ASSIGNMENTS

- You will receive a very basic introduction to the R statistical programming language

# LAB ASSIGNMENTS

- You will receive a very basic introduction to the R statistical programming language
- There are seven in-class Lab Days where you will work on an assignment applying the lessons learned in class to real-world problems using R

# LAB ASSIGNMENTS

- You will receive a very basic introduction to the R statistical programming language
- There are seven in-class Lab Days where you will work on an assignment applying the lessons learned in class to real-world problems using R
- You must watch the video lectures on R prior to the Lab Day class.

# LAB ASSIGNMENTS

- You will receive a very basic introduction to the R statistical programming language
- There are seven in-class Lab Days where you will work on an assignment applying the lessons learned in class to real-world problems using R
- You must watch the video lectures on R prior to the Lab Day class.
- You must turn in your completed lab assignment by 11:59pm the next day.

# LAB ASSIGNMENTS

- You will receive a very basic introduction to the R statistical programming language
- There are seven in-class Lab Days where you will work on an assignment applying the lessons learned in class to real-world problems using R
- You must watch the video lectures on R prior to the Lab Day class.
- You must turn in your completed lab assignment by 11:59pm the next day.
- You may complete these assignments on your own or in collaboration with other students. This means that you may work together to write code and/or solve problems.  
**Do not split up the questions or combine independent work.** Each student must submit an assignment on Blackboard.

# COURSE POLICY ON GENERATIVE AI

This course permits the use of Generative AI on **code** submitted for evaluation without restriction. However, the use of GAI tools for **written text** (e.g., exposition, analysis, etc.) is not permitted.

# CLASS PROJECT

- You will conduct a survey of your fellow GW students

# CLASS PROJECT

- You will conduct a survey of your fellow GW students
- You will have an opportunity to indicate your preference for different roles in the survey process

# CLASS PROJECT

- You will conduct a survey of your fellow GW students
- You will have an opportunity to indicate your preference for different roles in the survey process
- Your grade consists of your successful participation in the process plus a Lab Day assignment which uses the survey data you collect.

# CLASS PROJECT

- You will conduct a survey of your fellow GW students
- You will have an opportunity to indicate your preference for different roles in the survey process
- Your grade consists of your successful participation in the process plus a Lab Day assignment which uses the survey data you collect.
- You must complete the assigned CITI Ethics training to participate in the class project.

# Questions?

*Reminder: There is no class next Monday (MLK, Jr. Day).*

On your notecard, please write:

- 1 Preferred name
- 2 Preferred pronouns
- 3 Year in school and major
- 4 Your background in coding and/or statistics
- 5 One thing you hope to get out of this class