

Introduction to Social Data Analytics

Class 1

Arushi Kaushik

Department of Economics
UCSD

Why take this class?

- Social science is experiencing a revolution

Why take this class?

- Social science is experiencing a **revolution**
- Social data **EXPLOSION** available from:
 - governments
 - non-profits
 - businesses
 - citizens

Why take this class?

- Social science is experiencing a revolution
- Social data EXPLOSION available from:
 - governments
 - non-profits
 - businesses
 - citizens
- Better computing power and statistical techniques to analyze it
- Requires knowledge of social science, computer programming and statistics to be useful!

Why take this class?

- Social science is experiencing a **revolution**
- Social data **EXPLOSION** available from:
 - governments
 - non-profits
 - businesses
 - citizens
- Better **computing power** and **statistical** techniques to analyze it
- Requires knowledge of **social science**, **computer programming** and **statistics** to be useful!
- The short answer: **better career prospects**

Learning objectives and class goals

By the end of this course, you should be able to:

- Analyze data to solve **real world problems**
- Conduct basic operations in **Excel, Stata, and R**
- Identify resources for **further learning**
- Feel **inspired** and **empowered** to work with data

What you can expect of me

- Enthusiasm: I am **passionate** about data analytics

What you can expect of me

- Enthusiasm: I am **passionate** about data analytics
- Timeliness: class will start **on time**, questions answered in 24 hours

What you can expect of me

- Enthusiasm: I am **passionate** about data analytics
- Timeliness: class will start **on time**, questions answered in 24 hours
- Helpfulness: **your success** is my goal and my priority
 - I will happily answer any and all questions
 - My office hour is Wednesdays 2-3 pm at Room No. 124, Economics Building
 - My email address: arkaushi@ucsd.edu

My expectations of you

- Adopt a **growth mindset**: work hard, seek out resources, believe in yourself

My expectations of you

- Adopt a **growth mindset**: work hard, seek out resources, believe in yourself
- **Interrupt** me and ask questions! Especially *when* I make mistakes

My expectations of you

- Adopt a **growth mindset**: work hard, seek out resources, believe in yourself
- **Interrupt** me and ask questions! Especially *when* I make mistakes
- Use **Piazza** for all course content-related questions

My expectations of you

- Adopt a **growth mindset**: work hard, seek out resources, believe in yourself
- **Interrupt** me and ask questions! Especially *when* I make mistakes
- Use **Piazza** for all course content-related questions
- **Actively participate**:
 - in class: lecture will be interactive
 - on Piazza: help your classmates
 - on assignments: working in groups okay (and encouraged) but copying not

My expectations of you

- Adopt a **growth mindset**: work hard, seek out resources, believe in yourself
- **Interrupt** me and ask questions! Especially *when* I make mistakes
- Use **Piazza** for all course content-related questions
- **Actively participate**:
 - in class: lecture will be interactive
 - on Piazza: help your classmates
 - on assignments: working in groups okay (and encouraged) but copying not

Let's start with introductions.

Who am I?

Hi! I'm
Arushi

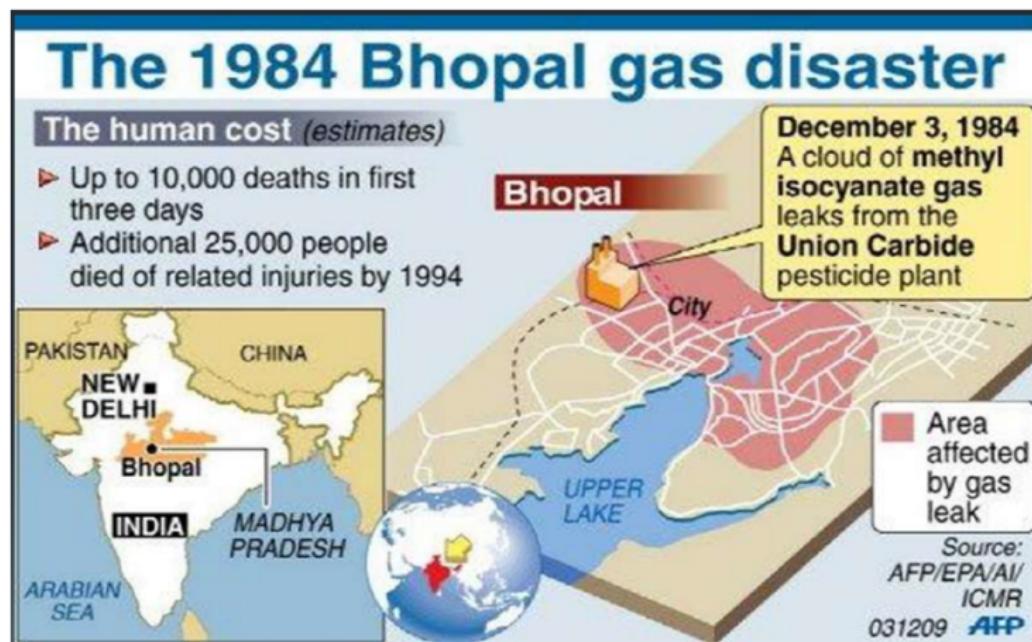


I like Travelling and
painting!

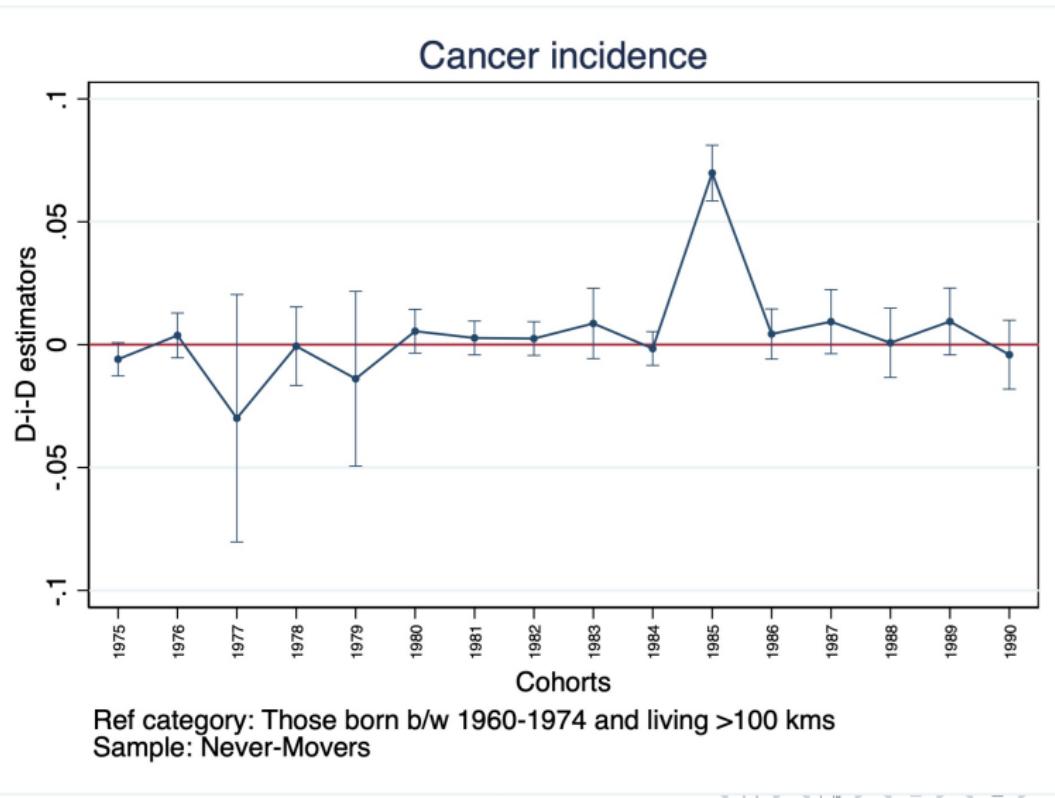
Data analytics in my research

Bhopal Gas Disaster

The worst industrial disaster in India.



Data analytics in my research



Introduce yourself to your neighbor

- Good places to start:
 - Name
 - Major
 - Hometown
- I like to share something **memorable**
- Exchange phone number and/or email

Introduce yourself to your neighbor

- Good places to start:
 - Name
 - Major
 - Hometown
- I like to share something **memorable**
- Exchange phone number and/or email
- I'd like to meet all of you, too!

How do we get started?

Hilary Mason (Chief Data Scientist at Bitly)



What you need to do data science:

- ① A question
- ② A dataset
- ③ Something to analyze it with

1. A question

Good questions could come from:

- Tech Firms/Startups
- Social media
- Government policy
- Academic research
- Political campaigns ... and much more!

2. A dataset

It's easy to find datasets online:

- data.gov
- census.gov/data
- ipums.org
- data.worldbank.org
- registry.opendata.aws
- Our TritonEd page for data used in this course

2. A dataset

It's easy to find datasets online:

- data.gov
- census.gov/data
- ipums.org
- data.worldbank.org
- registry.opendata.aws
- Our TritonEd page for data used in this course

We will begin exploring data next class.

3. Something to analyze data with

How will we work with the data? So many options....

Excel, SAS, SPSS, Stata, R, Python, Matlab, Gauss, C, Java, SciLab...

Our focus: **Excel, Stata, R**

3. Something to analyze data with

How will we work with the data? So many options....

Excel, SAS, SPSS, Stata, R, Python, Matlab, Gauss, C, Java, SciLab...

Our focus: **Excel, Stata, R**

- Why only these three?
 - **Important** for the social sciences at UCSD
 - Highly requested by **employers**

3. Something to analyze data with

How will we work with the data? So many options....

Excel, SAS, SPSS, Stata, R, Python, Matlab, Gauss, C, Java, SciLab...

Our focus: **Excel, Stata, R**

- Why only these three?
 - **Important** for the social sciences at UCSD
 - Highly requested by **employers**
- Why not just focus on one?
 - You will **shift** between softwares in your career
 - We'll focus on the **main takeaways** and demonstrate how they apply to each
 - You will learn how to **further** your knowledge independently

3. Something to analyze data with

Main takeaways:

- How to **import** data
- How to **automate** routines and **reproduce** them
- How to **visualize** data
- How to conduct **regressions**
- How to use **logic** in **if statements**
- How to use **for loops**
- How to write **functions**

All the while better understanding **social science questions!**

The Organizing Framework

① Unit 1: Excel

- Introduction to data structure
- Functions and plotting in Excel

The Organizing Framework

① Unit 1: **Excel**

- Introduction to data structure
- Functions and plotting in Excel

② Unit 2: **Stata**

- Introduction to scripting and reproducibility
- Descriptive statistics
- If statements
- Graphics and regression
- Data wrangling

The Organizing Framework

① Unit 1: Excel

- Introduction to data structure
- Functions and plotting in Excel

② Unit 2: Stata

- Introduction to scripting and reproducibility
- Descriptive statistics
- If statements
- Graphics and regression
- Data wrangling

③ Unit 3: R

- Introduction to scripting in R
- Visualizing data
- If statements and for loops
- Regression and functions
- Data wrangling

Assignments

- Four Problem Sets: (40% of grade)
 - One Excel, one Stata, and two R
 - Turn in commented code that can run
 - List names of people you worked with on first sheet of solutions

Assignments

- **Four Problem Sets:** (40% of grade)
 - One Excel, one Stata, and two R
 - Turn in commented code that can run
 - List names of people you worked with on first sheet of solutions
- **Midterm:** (20%)
 - computer-based assessment of data analysis in Excel and Stata

Assignments

- **Four Problem Sets:** (40% of grade)
 - One Excel, one Stata, and two R
 - Turn in commented code that can run
 - List names of people you worked with on first sheet of solutions
- **Midterm:** (20%)
 - computer-based assessment of data analysis in Excel and Stata
- **Final Project:** (30%)
 - Updates due throughout quarter with problem sets

Assignments

- **Four Problem Sets:** (40% of grade)
 - One Excel, one Stata, and two R
 - Turn in commented code that can run
 - List names of people you worked with on first sheet of solutions
- **Midterm:** (20%)
 - computer-based assessment of data analysis in Excel and Stata
- **Final Project:** (30%)
 - Updates due throughout quarter with problem sets
- **Pre-class exercises:** Due midnight before each class (5%)
 - Will help prepare you for in class activities

Assignments

- **Four Problem Sets:** (40% of grade)
 - One Excel, one Stata, and two R
 - Turn in commented code that can run
 - List names of people you worked with on first sheet of solutions
- **Midterm:** (20%)
 - computer-based assessment of data analysis in Excel and Stata
- **Final Project:** (30%)
 - Updates due throughout quarter with problem sets
- **Pre-class exercises:** Due midnight before each class (5%)
 - Will help prepare you for in class activities
- **Participation:** (5%)
 - Lecture and section attendance required

Course Materials

- Before each class, you will be expected to submit brief exercises
 - The best way to improve your analytics skills is to **practice!**
 - These exercises will prep you for in-class activities

Course Materials

- Before each class, you will be expected to submit brief exercises
 - The best way to improve your analytics skills is to **practice!**
 - These exercises will prep you for in-class activities
- Course materials listed in the syllabus will help with the exercises
 - Readings
 - Youtube videos
 - When we get to R, *A First Course in Quantitative Social Science* by Kosuke Imai

Course Materials

- Before each class, you will be expected to submit brief exercises
 - The best way to improve your analytics skills is to **practice!**
 - These exercises will prep you for in-class activities
- Course materials listed in the syllabus will help with the exercises
 - Readings
 - Youtube videos
 - When we get to R, *A First Course in Quantitative Social Science* by Kosuke Imai
- Datasets, problem sets, and more will be posted on **TritonEd**

Course Materials

- Before each class, you will be expected to submit brief exercises
 - The best way to improve your analytics skills is to **practice!**
 - These exercises will prep you for in-class activities
- Course materials listed in the syllabus will help with the exercises
 - Readings
 - Youtube videos
 - When we get to R, *A First Course in Quantitative Social Science* by Kosuke Imai
- Datasets, problem sets, and more will be posted on **TritonEd**
- Questions on course content should be asked on **Piazza**

Software you will need

- **Excel**

- Install before next class (free campus license)
- Available on most campus computers

- **Stata**

- Normally expensive
- We will provide a campus license

- **R**

- Free!
- I will show you how to download in the second half of the course

Some housekeeping

What does **academic integrity** mean to you?

Some housekeeping

What does **academic integrity** mean to you?

From the Academic Integrity Office: “Cheating occurs when a student attempts to get academic credit in a way that is dishonest, disrespectful, irresponsible, untrustworthy, or unfair.”

Some housekeeping

What does **academic integrity** mean to you?

From the Academic Integrity Office: “Cheating occurs when a student attempts to get academic credit in a way that is dishonest, disrespectful, irresponsible, untrustworthy, or unfair.”

We will take academic integrity seriously.

I want to know about you.

In the spirit of data science....

<https://bit.ly/2q1mUdg>

Remember to submit your pre class exercise (via TritonEd) before next class!