Data Science

CMSC 320

This Lecture

Getting some data.

Before we start...

Before we start...

1. Queuing for office hours.

Before we start...

1. Queuing for office hours.
2. Working together.

Before we start...

1. Queuing for office hours.
2. Working together.
3. Email.

Before we start...

1. Queuing for office hours.
2. Working together.
3. Email.
4. Accommodation Letters.

Queuing

Queuing

1. Thank you for your feedback!

Queuing

1. Thank you for your feedback!
2. After discussing with students/TAs, we're going to use discord, not Quuly.

Queuing

1. Thank you for your feedback!
2. After discussing with students/TAs, we're going to use discord, not Quuly.
3. Allow me to demonstrate.

Queuing

1. Thank you for your feedback!
2. After discussing with students/TAs, we're going to use discord, not Quuly.
3. Allow me to demonstrate.
4. If this does not work for you, let us know! We want to be as flexible as possible.

Working Together

Working Together

Some thoughts on working together for 320. These thoughts only apply to 320.

Email

Email

I get a lot of email

Email

- Never feel shy to email again.
- If you're still shy, email a TA and they'll reach me.

Accommodation Letters

Accommodation Letters

1. For some reason, my ads portal account is borked.

Accommodation Letters

1. For some reason, my ads portal account is borked.
2. While IT is on it, I've lost my patience. Moving to Plan B.

Accommodation Letters

1. For some reason, my ads portal account is borked.
2. While IT is on it, I've lost my patience. Moving to Plan B.
3. If you email me with your letter (i.e. not through the ADS portal) I will respond with affirmation that I have seen your letter.

Accommodation Letters

1. For some reason, my ads portal account is borked.
2. While IT is on it, I've lost my patience. Moving to Plan B.
3. If you email me with your letter (i.e. not through the ADS portal) I will respond with affirmation that I have seen your letter.
4. If you would like further assurance, I will print it out, sign it, and scan it back for you.

Accommodation Letters

1. For some reason, my ads portal account is borked.
2. While IT is on it, I've lost my patience. Moving to Plan B.
3. If you email me with your letter (i.e. not through the ADS portal) I will respond with affirmation that I have seen your letter.
4. If you would like further assurance, I will print it out, sign it, and scan it back for you.
5. I promise this is not my ideal situation, I'm sorry that it's affected how quickly I can turn around these letters.

Data!

The reading teased the notion of 4 ‘kinds’ of data (in roughly two ‘types’):

Data!

The reading teased the notion of 4 ‘kinds’ of data (in roughly two ‘types’):

- Nominal (Categorical)

Data!

The reading teased the notion of 4 ‘kinds’ of data (in roughly two ‘types’):

- Nominal (Categorical)
- Ordinal (Categorical)

Data!

The reading teased the notion of 4 ‘kinds’ of data (in roughly two ‘types’):

- Nominal (Categorical)
- Ordinal (Categorical)
- Interval (Numerical)

Data!

The reading teased the notion of 4 ‘kinds’ of data (in roughly two ‘types’):

- Nominal (Categorical)
- Ordinal (Categorical)
- Interval (Numerical)
- Ratio (Numerical)

Categorical Data: Nominal

Categorical Data: Nominal

- Think ‘finite set’

Categorical Data: Nominal

- Think ‘finite set’
- Marital status, soda flavor, etc.

Categorical Data: Nominal

- Think ‘finite set’
- Marital status, soda flavor, etc.
- Comparison is difficult and nonsensical

Categorical Data: Ordinal

Categorical Data: Ordinal

- Like Nominal data, Ordinal data describes classes or states of things...

Categorical Data: Ordinal

- Like Nominal data, Ordinal data describes classes or states of things...
- But we can provide an order

Categorical Data: Ordinal

- Like Nominal data, Ordinal data describes classes or states of things...
- But we can provide an order
- The lecturer of this class is {boring, neutral, exciting}

Categorical Data: Ordinal

- Like Nominal data, Ordinal data describes classes or states of things...
- But we can provide an order
- The lecturer of this class is {boring, neutral, exciting}
- We have an order but not a mathematical way to measure distance

Numerical Data: Interval

Numerical Data: Interval

- Think: Dates, year in school (i.e. grade level), temperature.

Numerical Data: Interval

- Think: Dates, year in school (i.e. grade level), temperature.
- We have ordering and distance.

Numerical Data: Interval

- Think: Dates, year in school (i.e. grade level), temperature.
- We have ordering and distance.
- What don't we have?

Numerical Data: Ratio

Numerical Data: Ratio

- Everything Interval has, but with a meaningful zero

Numerical Data: Ratio

- Everything Interval has, but with a meaningful zero
- Ratios are meaningful (hence the name)

Numerical Data: Ratio

- Everything Interval has, but with a meaningful zero
- Ratios are meaningful (hence the name)
- Money, distance, volume, etc.

From data to data *representation*

From data to data *representation*

Data structures are important!

From data to data *representation*

Data structures are important! They guide you by limiting the number of appropriate operations

From data to data *representation*

Data structures are important! They guide you by limiting the number of appropriate operations

- What are the appropriate operations for an array?

From data to data *representation*

Data structures are important! They guide you by limiting the number of appropriate operations

- What are the appropriate operations for an array?
- Index, slice, map, reduce, etc.

From data to data *representation*

Data structures are important! They guide you by limiting the number of appropriate operations

- What are the appropriate operations for an array?
- Index, slice, map, reduce, etc.
- What dataset would be appropriate to represent as an array?

From data to data *representation*

Data structures are important! They guide you by limiting the number of appropriate operations

- What are the appropriate operations for an array?
- Index, slice, map, reduce, etc.
- What dataset would be appropriate to represent as an array?
- In what ways could we combine two arrays?

From data to data *representation*

From data to data *representation*

What about multi-dimensional arrays?

From data to data *representation*

From data to data *representation*

What about $\mathbb{N}$-dimensional arrays (i.e. higher-dimensional matrices)

From data to data *representation*

What about $\mathbb{N}$-dimensional arrays (i.e. higher-dimensional matrices)

- This is where Linear Algebra starts to come in handy!

From data to data *representation*

From data to data *representation*

What about...

From data to data *representation*

What about...

- Sets?

From data to data *representation*

What about...

- Sets?
- Maps (a.k.a Dictionaries)?

From data to data *representation*

What about...

- Sets?
- Maps (a.k.a Dictionaries)?
- Tables?

From data to data *representation*

What about...

- Sets?
- Maps (a.k.a Dictionaries)?
- Tables?
- Trees?

From data to data *representation*

What about...

- Sets?
- Maps (a.k.a Dictionaries)?
- Tables?
- Trees?
- Graphs?

Let's get some data!

Let's get some data!

To the REPL!

Any Questions?

Thanks for your time!