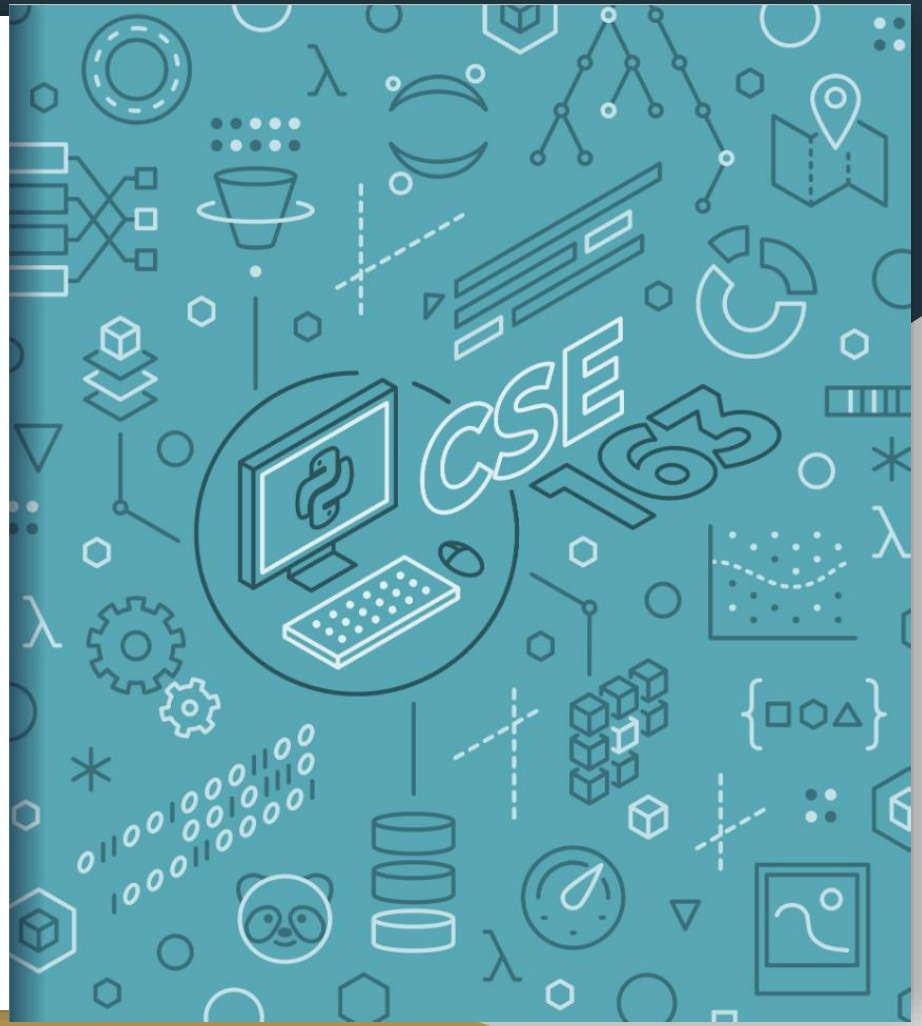
CSE 163
Data Literacy

Adrian Salguero
Spring 2026

💭 Icebreaker (discuss with neighbors): ‘
Best place for ramen in Seattle?
Add to our Slido!

slido.com
#cse163



Announcements

- **Take Home Assessment 3: Pokemon** due Thursday, April 30th at 11:59pm!
- **Peer Reviews for THA 2** due Wednesday, April 29th at 11:59pm!
- **Lesson 12 Canvas Quiz** due tonight at 11:59pm!
- **Reading Assignment 3** due tonight, April 27th at 11:59pm!
- **Project Part 1** due Friday May 1st at 11:59pm on Gradescope!
- **Resub Cycle 2** will be made available today and due tomorrow at 11:59pm!
  - Eligible assignment(s): ***THA 1***

MultiIndex

```
# MultiIndex of (year, month, day)

# returns count for one day
earthquakes.loc[(2016, 7, 27), “count”]

# returns count for all days
earthquakes.loc[:, “count”]

# returns count for multiple days in the same year/month
earthquakes.loc[(2016, 7, slice(10, 15)), “count”]

# returns count for multiple months
earthquakes.loc[(2016, [7, 8], 27), “count”]
```

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

data = pd.read_csv('home/data.csv')

data = data[['name', 'fin_length', 'age']]
data = data.dropna()

sns.relplot(data=data, kind='line', x='age',
                        y='fin_length', hue='name')
plt.title('Shark Ages vs. Fin Length')
plt.xlabel('Age (months)')
plt.ylabel('Fin Length (in)')

plt.savefig('/home/plot.png')
```

Does code tell us anything about data?

suggests >3 columns

suggests missing values

suggests quantitative and categorical variables

Data Context Guiding Questions

- ***Who***: the people represented in the data or responsible for its collection and usage
- ***What***: the information that is represented in the data such as demographics and measurements
- ***When***: the timeframe associated with data collection or represented in the data
- ***Where***: the location of the data, virtual or otherwise
- ***Why***: the stated (or unstated) purpose of collecting the data
- ***How***: the methods used for data collection and storage

Where do we find this context?

- Data dictionaries
- *ReadME.md* or *CHANGELOG.md* files
- Meta datasheets
- Contacting the researchers/data collectors

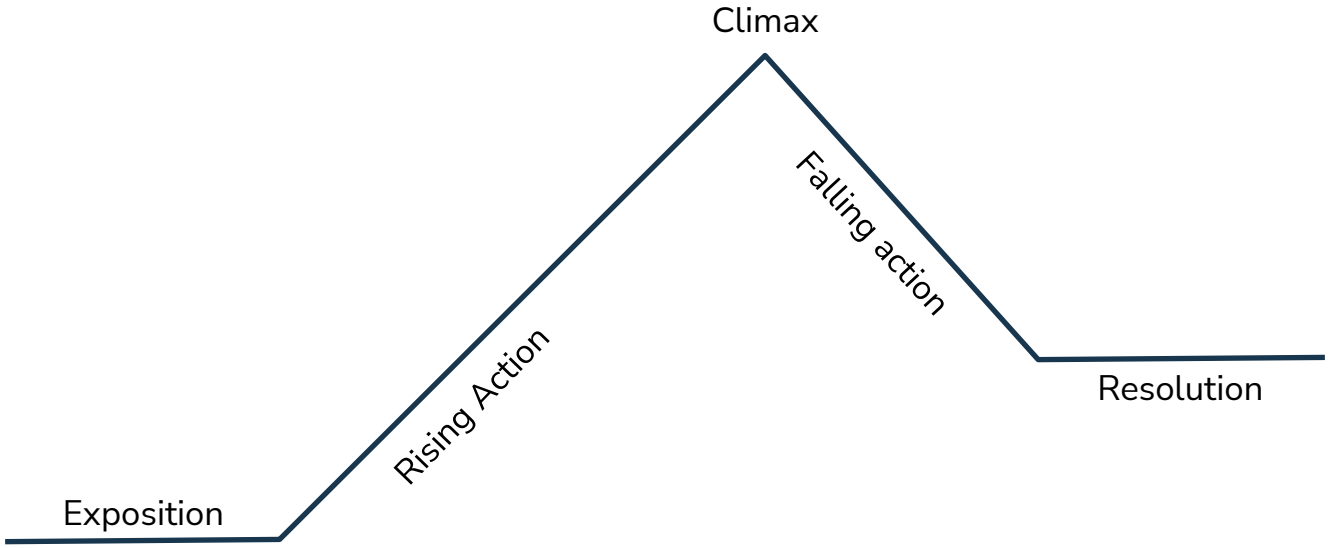
Data Storytelling

Let’s talk about parallel universes…

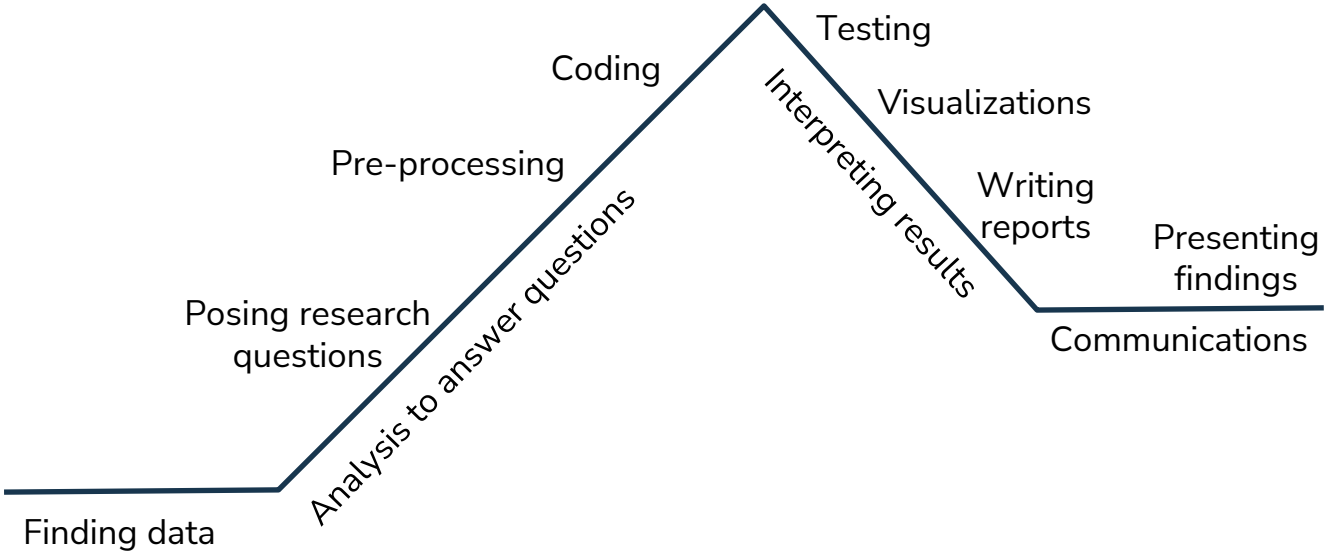
Shark Multiverse...

- A set of bar graphs illustrating different categorical qualities of different shark species
- A YouTube tutorial using this dataset to demonstrate how to drop or replace missing values
- A neural network using all columns to predict whether the shark is endangered or not
- A neural network using all columns to predict shark tail size
- A single string representing the most common shark species in the dataset

Narrative Plot Mountain



Data Story Plot Mountain



Finding Main Characters

- Make sure the questions are relevant to your data, and the data is relevant to your questions.
- Ask questions that do not have simple answers.
- Use your domain knowledge to come up with interesting questions.
- Pacing matters!

Writing Tips

- We care more about ***what*** you have to say rather than ***how*** you say it!
- Keep your writing relevant and to the point!
- Some strategies for answering your research questions:
  - Answer a question, then explain that answer
  - Answer the question, think about a counterexample/counterargument, then refute it
  - Answer the question, pose a follow-up question, then answer the follow-up

Avoiding the Vacuum

- Interpret ***any*** numbers and/or trends in context!
- Do NOT leave free-floating numbers.
- Don’t assume that your reader knows your code as well as you do! (For in-class assignments, you may assume that your reader knows any definitions that we give in the spec.)
- Think about explaining your project or portfolio to someone who is not in this class.

Sentence Level Details

- A few mechanical things to think about:
  - Writing in the first-person (“I/We did X for my/our analysis to find Y” as opposed to “X was used to find Y”)
  - Active voice vs. passive voice (“The results of A proved B” as opposed to “B was proven by A”)
  - Present vs. past-tense
  - Using abbreviations or shorthand instead of full names

Key Takeaways

- Think of how your analysis might be used or interpreted.
- Consider biases in your data and analysis – even the ones that might come from you!
- Consider the impacts, ethics, and consequences of your analysis
- Data science tells a story – who is our “main character” and what do we want to focus on or highlight?