Data Visualization

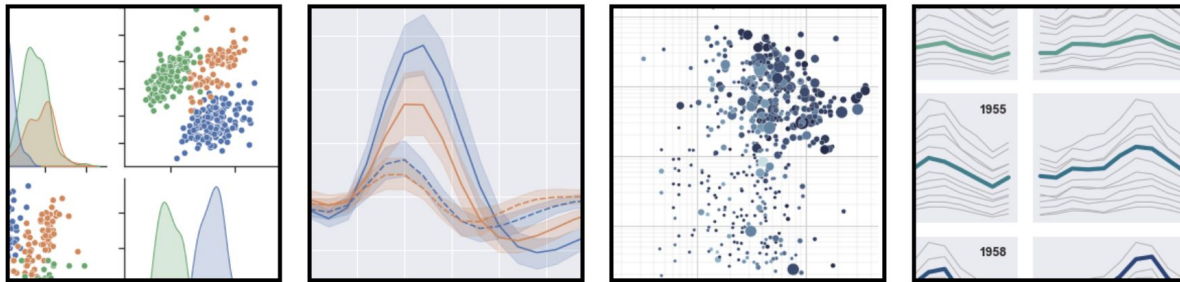COGS 108 Spring 2025
Jason Chen
Week 4

xic007@ucsd.edu

OH: Thu 3-5 pm

Discussion slides and materials adapted from Ruby Ying & previous quarter
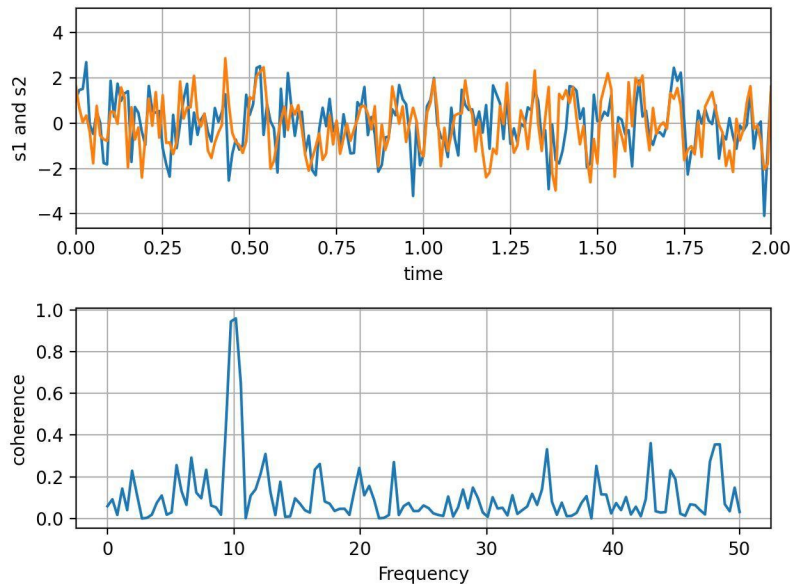
Due dates

- Project Review: Wednesday Apr 23 @ 11:59 PM
- D3: Friday Apr 25 @ 11:59 PM

Pyplot and seaborn

Matplotlib (plt)

- A plotting library for Python
- Makes static, animated, and interactive visualizations in Python.
- Usually imported under the plt alias

There are so many ways to make a same plot!

- All of these do the same thing:
  - Line 1: plt.hist(df['income'], 25)
  - Line 2: df['income'].hist(bins=25)
  - Line 3: df.hist('income', bins=25)

- In Python, most image-based plots created using Matplotlib (plt), e.g., plt.hist, plt.bar, plt.plot, etc.

- Pandas gives shortcuts for matplotlib plots. Lines 2 and 3 are shortcuts for line 1.

Seaborn

- Makes common statistical charts easy to create, like bar plots with confidence intervals.
- Again, seaborn is really just a bunch of shortcuts for matplotlib.
- Usually imported under the sns alias
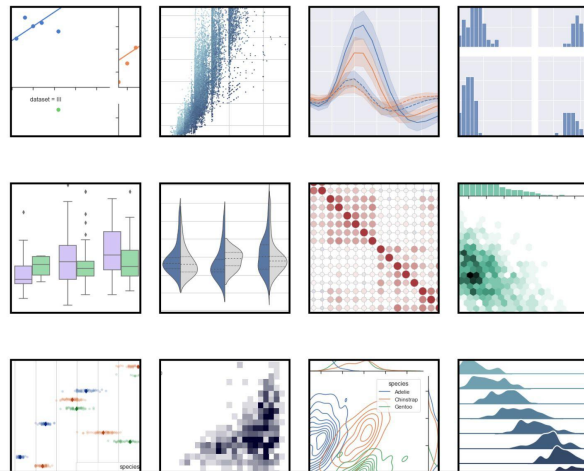
https://seaborn.pydata.org/tutorial.html
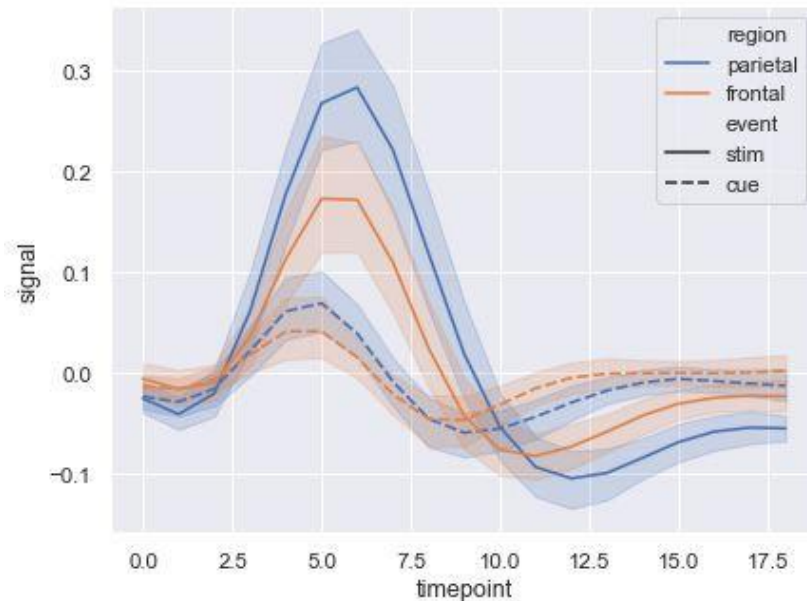
Tips for Data Visualization

- Google is your best friend.
- E.g. "stacked histogram python", "scatter plot sns", "remove legend seaborn"
- Documentation is your friend: https://seaborn.pydata.org/api.html
- Half the time is getting the data in the right format- clean your data

Example gallery

https://seaborn.pydata.org/examples/index.html

It's time to …

- Go over some demos of D3!

Import all the necessary libraries

- `numpy` (`np`)
- `pandas` (`pd`)
- `matplotlib.pyplot` (`plt`)
- `seaborn` (`sns`)
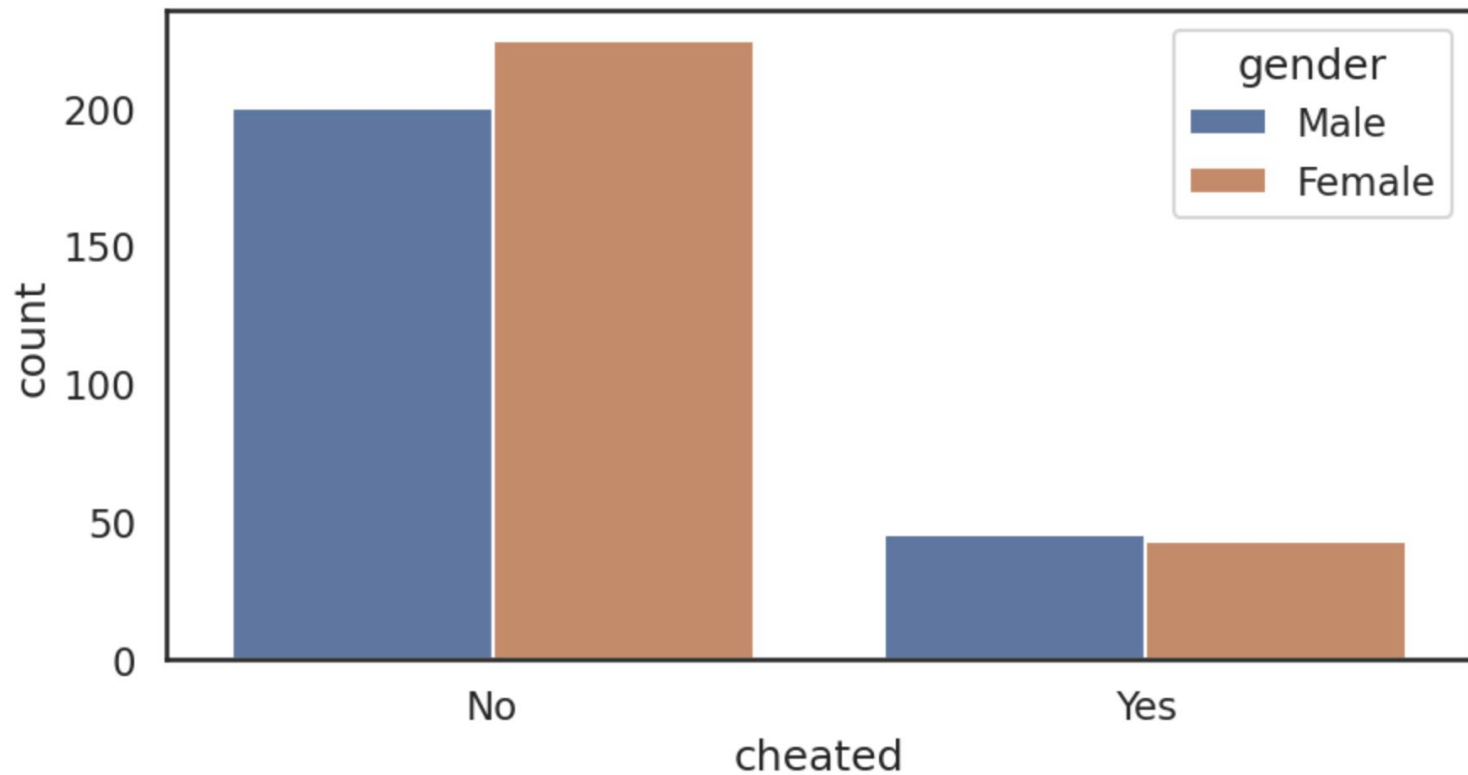
```
[1]:  # YOUR CODE HERE
      import numpy as np
      import pandas as pd
      import matplotlib.pyplot as plt
      import seaborn as sns
```

```
[2]:  assert np
      assert pd
      assert plt
      assert sns
```
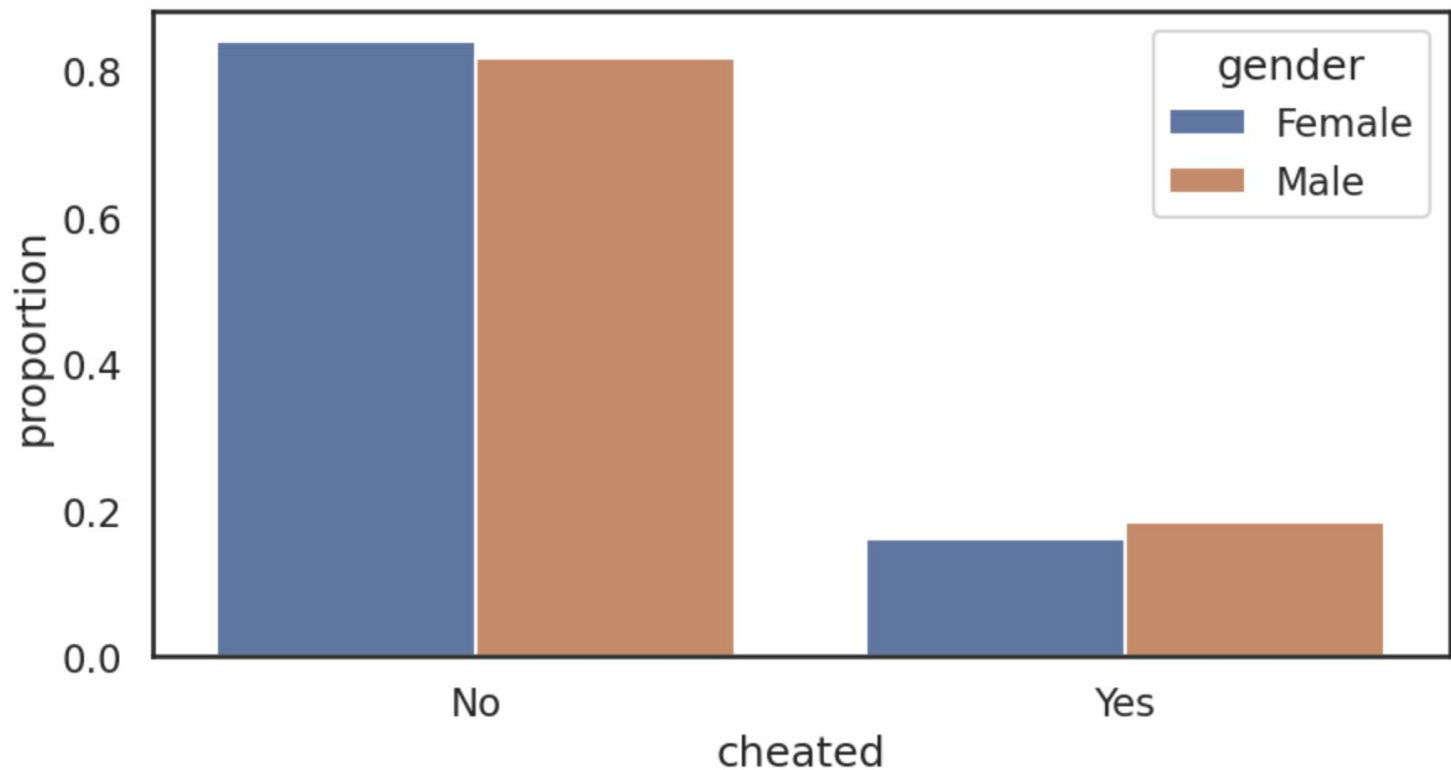
GOT IT?

2. PART I: CHEATING (plot_cheated)

2. PART I: CHEATING (prop_df)

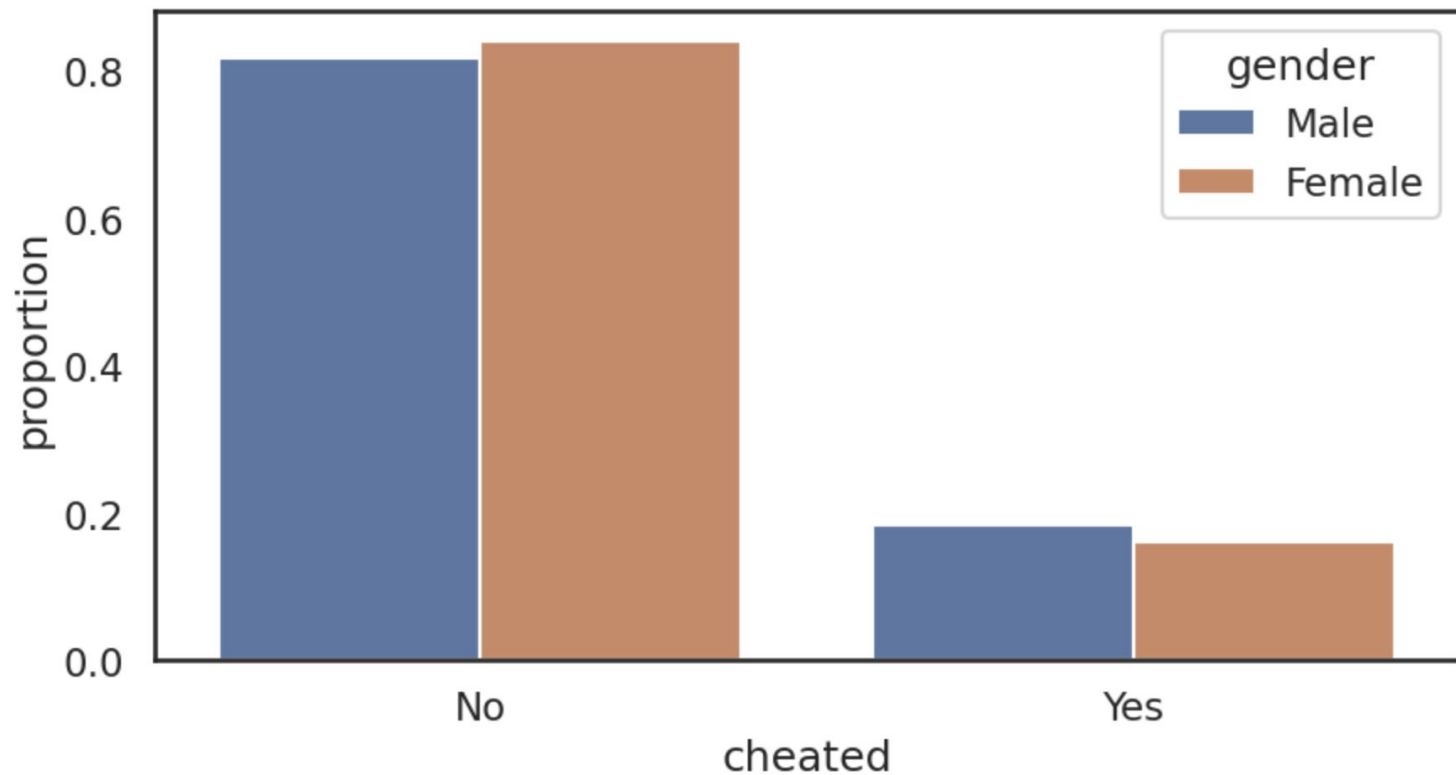|	gender	cheated	proportion
0	Female	No	0.84
1	Female	Yes	0.16
2	Male	No	0.82
3	Male	Yes	0.18

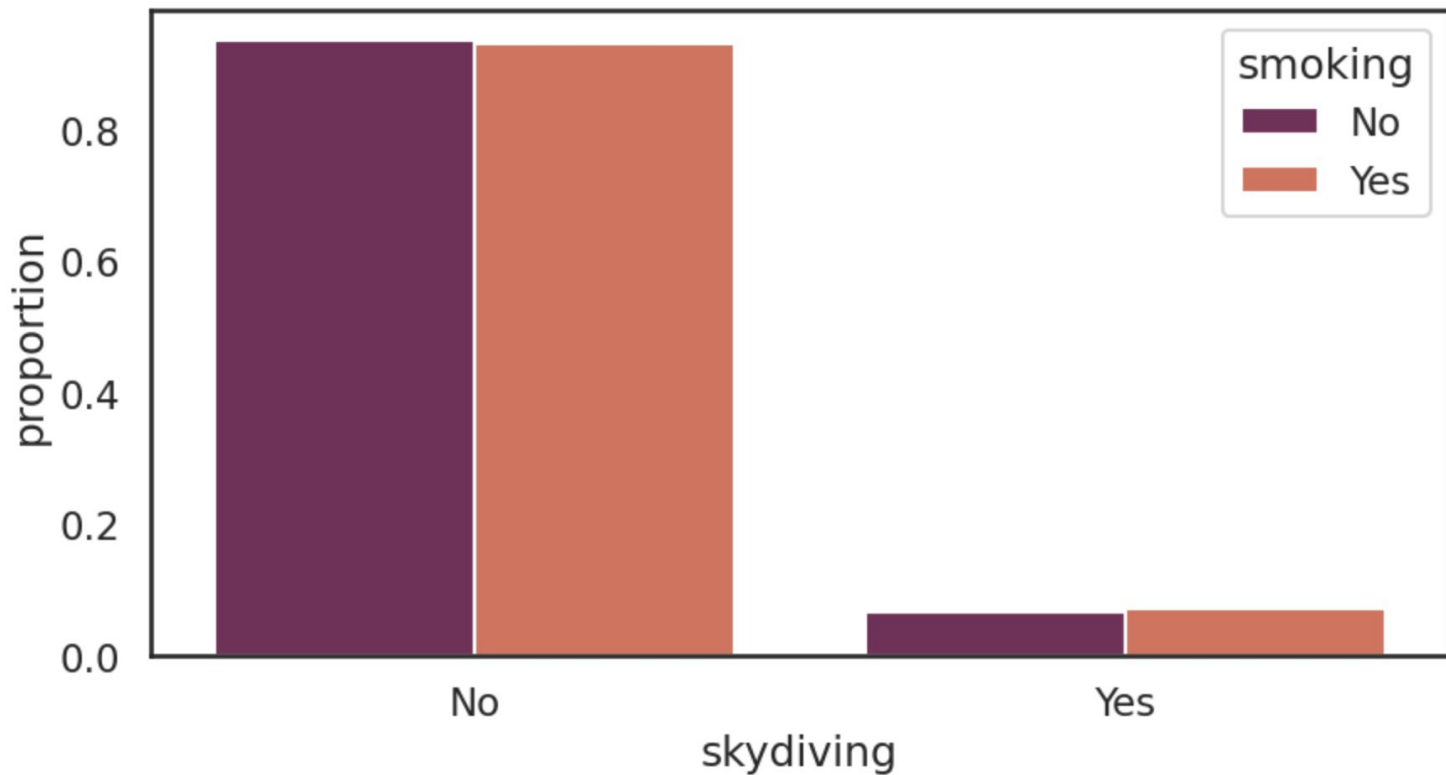I need to make this data into a pretty and easy to read visualization?!?
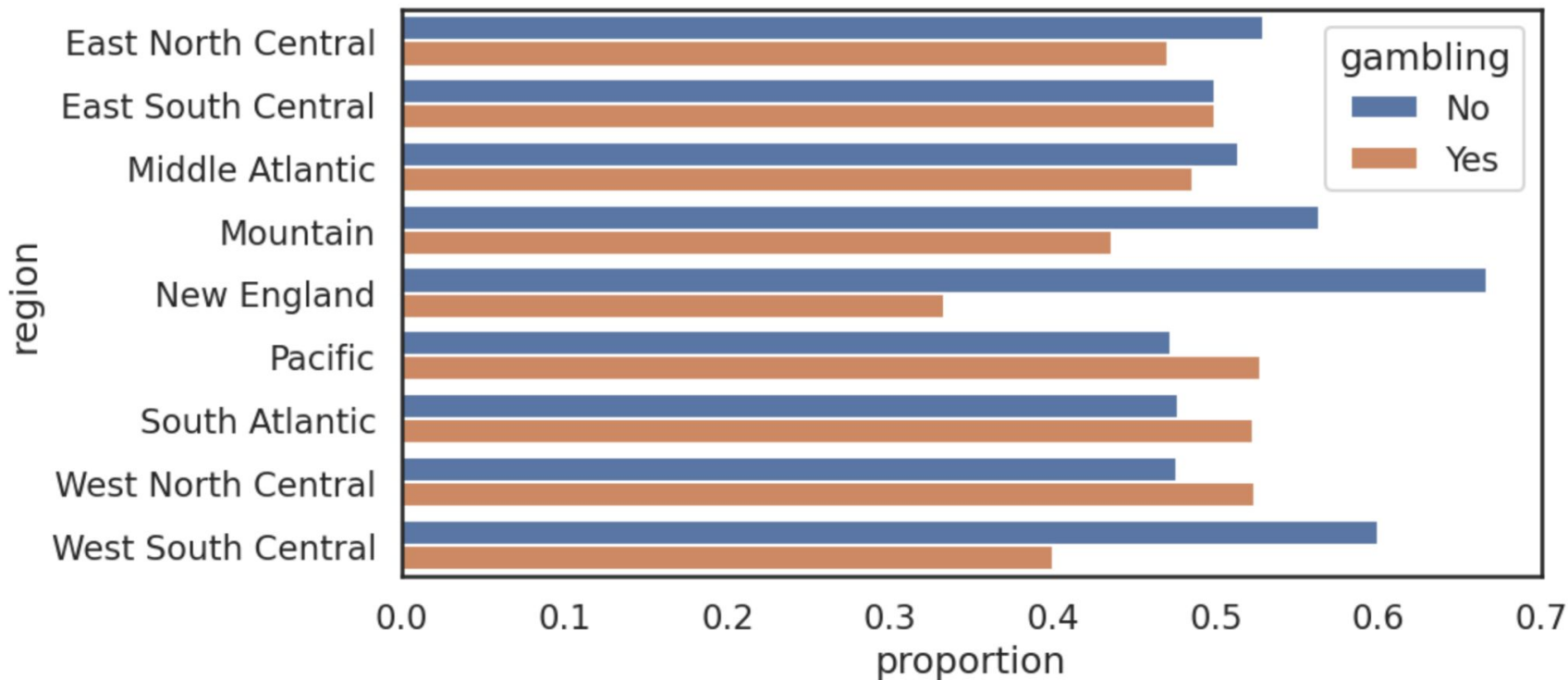
2. PART I: CHEATING (plot_proportion)

2. PART I: CHEATING (plot_reorder)

3. PART II: RISKY BEHAVIOR (plot_skydiving)

4. PART III: GAMBLING (plot_gambling)

Section Materials

https://github.com/JasonC1217/COGS_108_B03-B04_Sp25/tree/master

or:

https://tinyurl.com/4d8wx3ne

SCAN ME

THANKS!

Questions on EdStem or office hours

Office hours: Thu, 3-5 PM