DATA SCIENCE FOR ECONOMISTS

ECON 220 LAB

Jafet Baca-Obando

Week 7, Data visualization – 10/10/2025

Outline

01 Scatterplots with **seaborn**

02 Bar charts

Importing required libraries and dataset

```
# %pip install seaborn
```

✓ 0.0s Python

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns # New package
```

✓ 6.2s Python

```
# Find working directory
import os
path = os.getcwd()
print(path)
```

✓ 0.0s Python

Load the data

```
# Import data
data = pd.read_csv('college.csv')

# First 10 rows
data.head(10)
```

✓ 0.0s Open 'data' in Data Wrangler Python

	# id	name	city
0	102669	Alaska Pacific University	Anchorage
1	101648	Marion Military Institute	Marion
2	100830	Auburn University at Montgomery	Montgomery
3	101879	University of North Alabama	Florence
4	100858	Auburn University	Auburn
5	100663	University of Alabama at Birminghan	Birmingham
6	101480	Jacksonville State University	Jacksonville
7	102049	Samford University	Birmingham
8	101709	University of Montevallo	Montevallo
9	100751	The University of Alabama	Tuscaloosa

10 rows x 17 cols 10 per page Page 1 of 1

Do the data types make sense?

```
# Data's information
data.info()
```

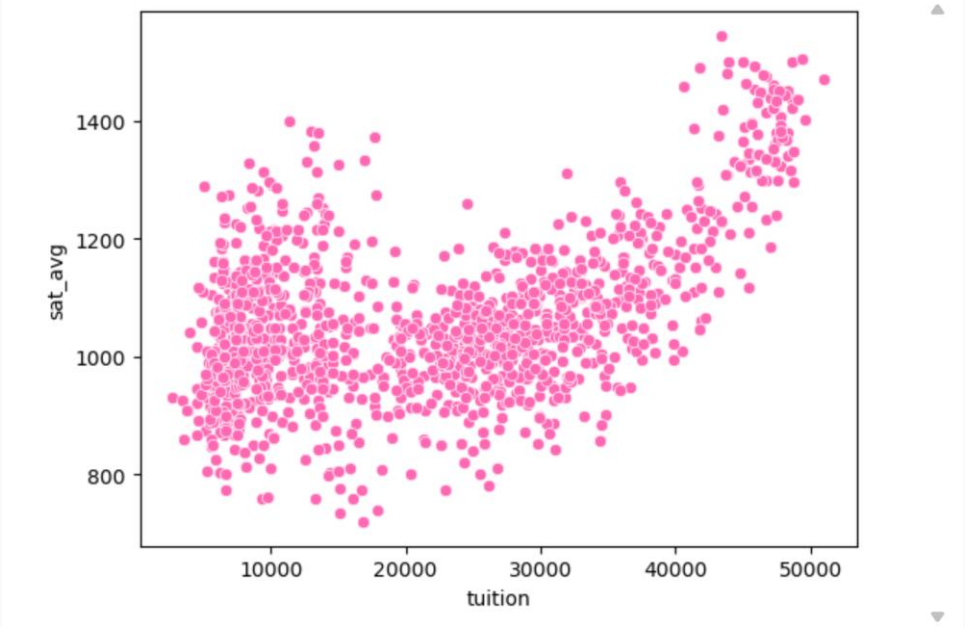
✓ 0.0s Python

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1269 entries, 0 to 1268
Data columns (total 17 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   id                  1269 non-null   int64
 1   name                1269 non-null   object
 2   city                1269 non-null   object
 3   state               1269 non-null   object
 4   region              1269 non-null   object
 5   highest_degree      1269 non-null   object
 6   control             1269 non-null   object
 7   gender              1269 non-null   object
 8   admission_rate      1269 non-null   float64
 9   sat_avg             1269 non-null   int64
 10  undergrads          1269 non-null   int64
 11  tuition             1269 non-null   int64
 12  faculty_salary_avg  1269 non-null   int64
 13  loan_default_rate   1267 non-null   float64
 14  median_debt         1269 non-null   float64
 15  lon                 1269 non-null   float64
 16  lat                 1269 non-null   float64
dtypes: float64(5), int64(5), object(7)
memory usage: 168.7+ KB
```

Scatterplots with **seaborn**

```
# Plot correlation between tuition and SAT score
scatter = sns.scatterplot(data, x='tuition', y='sat_avg', color='hotpink')
```

✓ 0.3s Python

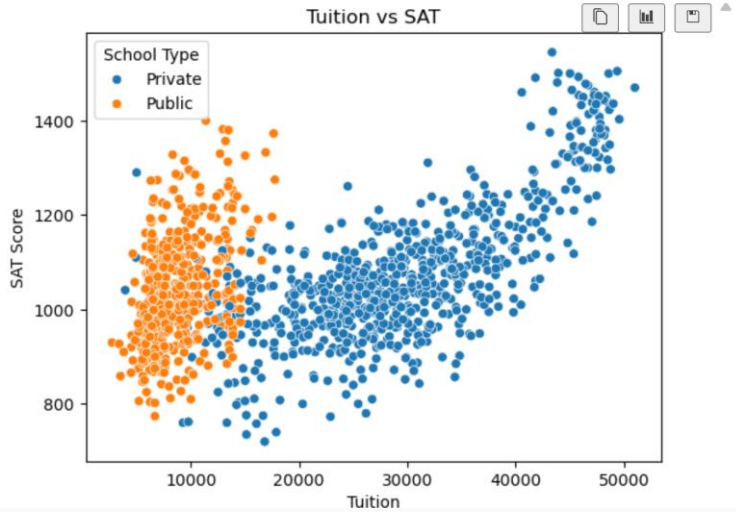


Tuition versus SAT by school type

```
scatter = sns.scatterplot(data, x='tuition', y='sat_avg', hue='control')
# Change title and labels
scatter.set(title = 'Tuition vs SAT', xlabel = 'Tuition', ylabel = 'SAT Score')
# Modify legend: because seaborn is based on matplotlib, so use plt
plt.legend(title='School Type')
```

✓ 0.3s Python

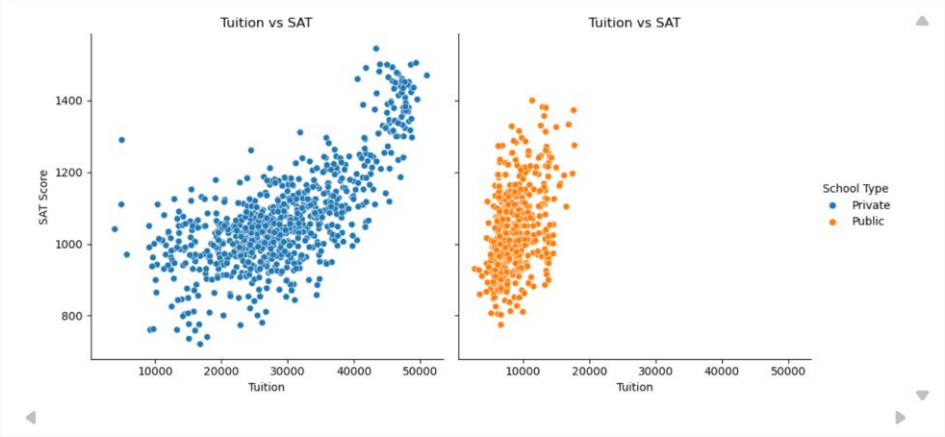
```
<matplotlib.legend.Legend at 0x24d2df55340>
```



Facet grid

```
# Split scatterplot into two: use "relplot()"
scatter = sns.relplot(data, x='tuition', y='sat_avg', hue='control',
col='control');
# Change title and labels
scatter.set(title = 'Tuition vs SAT', xlabel = 'Tuition', ylabel = 'SAT
Score');
# Modify legend for two plots side by side:
scatter._legend.set_title('School Type')
# plt.legend() can only modify legend within 1 plot
```

✓ 0.7s Python

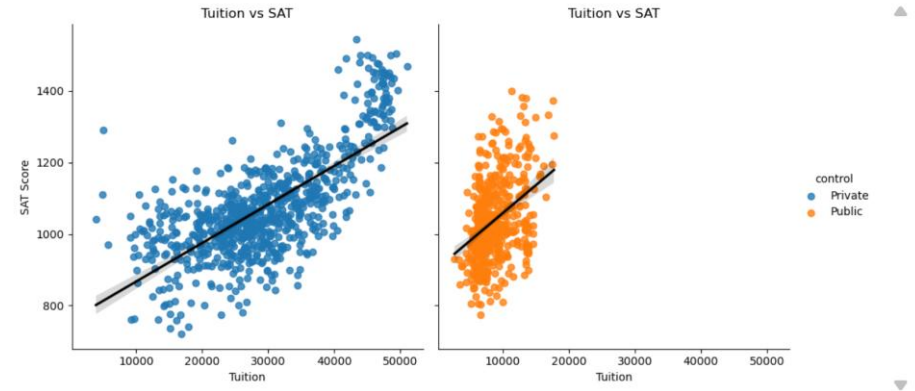


Trend lines

```
# Add linear line: use "lmplot()"
scatter = sns.lmplot(data, x='tuition', y='sat_avg', hue='control',
col='control', line_kws={'color':'black'})
# Change title and labels
scatter.set(title = 'Tuition vs SAT', xlabel = 'Tuition', ylabel = 'SAT
Score').add_legend()
# Use .add_legend() to bring back the legend
```

✓ 0.9s Python

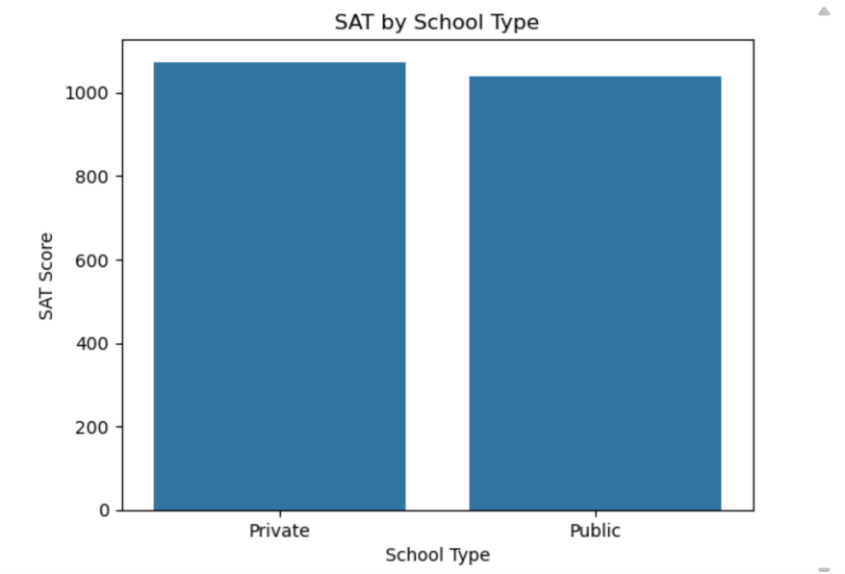
```
<seaborn.axisgrid.FacetGrid at 0x24d2ffb5700>
```



Bar charts with **seaborn**

```
# Barplot to compare average SAT score between public and private universities
bar = sns.barplot(data, x='control', y='sat_avg', errorbar=None);
bar.set(title = 'SAT by School Type', xlabel = 'School Type', ylabel = 'SAT
Score');
```

✓ 0.1s Python

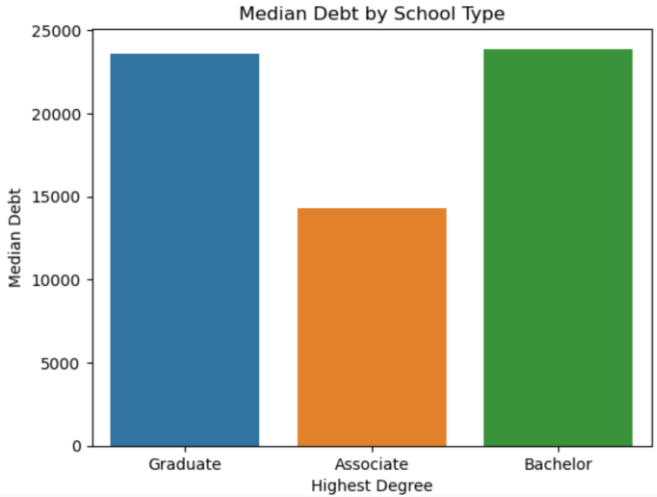


Bar charts with **seaborn**

```
bar = sns.barplot(data, x='highest_degree', y='median_debt',
hue='highest_degree', errorbar=None)
bar.set(title = 'Median Debt by School Type',
        xlabel = 'Highest Degree', ylabel = 'Median Debt')
```

✓ 0.2s Python

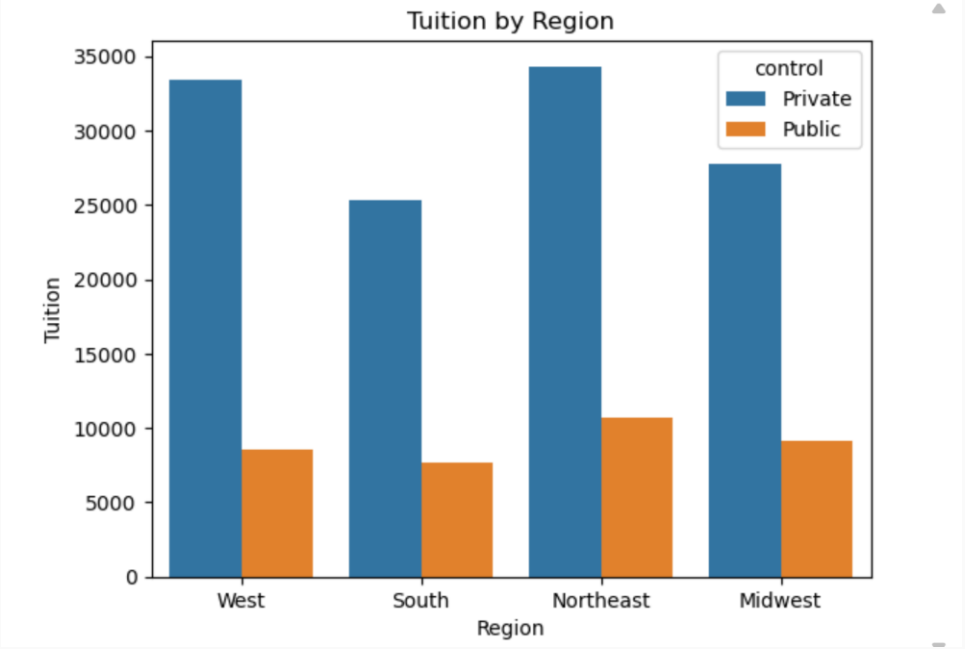
```
[Text(0.5, 1.0, 'Median Debt by School Type'),
 Text(0.5, 0, 'Highest Degree'),
 Text(0, 0.5, 'Median Debt')]
```



Bar charts with **seaborn**

```
bar = sns.barplot(data, x='region', y='tuition', hue='control', errorbar=None);
bar.set(title = 'Tuition by Region', xlabel = 'Region', ylabel = 'Tuition');
```

✓ 0.2s Python



Recap

- We created scatterplots and bar plots using seaborn.
- We analyzed correlations in college data.

To-do list

- **Complete Data Exercise 4**
  - Upload Jupyter notebook (.ipynb file) and HTML file on **October 12**
- **Complete Data Exercise 5**
  - Upload Jupyter notebook (.ipynb file) and HTML file on **October 19**