

Aaron J. Brown

DSC 640: Data Presentation & Visualization - Winter 2023

WEEKS 11-12 Exercises: Histograms, Box Plots, & Bullet Charts (PYTHON)

IMPORTING LIBRARIES.

```
In [1]: import os
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import plotly.express as px

import warnings
warnings.filterwarnings('ignore')

# nltk.download('example')
# pip install squarify
```

IMPORTING DATA.

```
In [2]: education_df = pd.read_csv('/Users/aaronbrown/Documents/Classwork/DSC 640 - Data Presentation and Visualization/Data/education_data')
education_data = education_df
```

```
In [3]: crime_df = pd.read_csv('/Users/aaronbrown/Documents/Classwork/DSC 640 - Data Presentation and Visualization/Data/crimeratesb')
crime_data = crime_df
```

```
In [4]: hotdog_df = pd.read_excel('/Users/aaronbrown/Documents/Classwork/DSC 640 - Data Presentation and Visualization/Data/hotdog-c')
hotdog_data = hotdog_df
```

```
In [5]: obama_df = pd.read_excel('/Users/aaronbrown/Documents/Classwork/DSC 640 - Data Presentation and Visualization/Data/obama-app')
obama_data = obama_df
```

Removing United States and D.C. rows.

```
In [6]: education_data = education_data.drop([0, 9])
crime_data = crime_data.drop([0, 9])
```

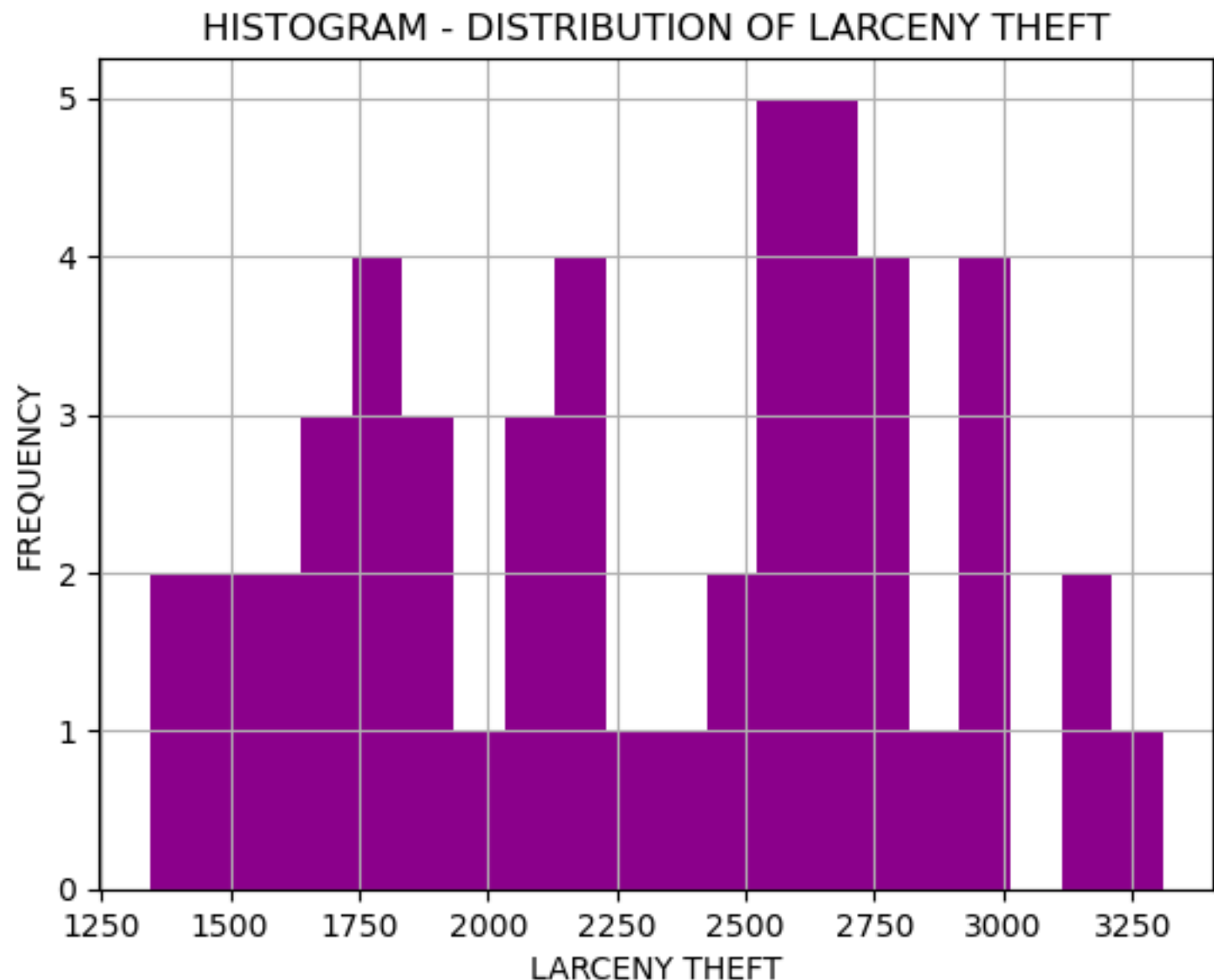
```
In [7]: crime_data['total_crime'] = crime_data['murder'] + crime_data['forcible_rape'] + crime_data['robbery'] + crime_data['aggravated_assault']
crime_data.tail()
```

```
Out[7]:
```

	state	murder	forcible_rape	robbery	aggravated_assault	burglary	larceny_theft	motor_vehicle_theft	total_crime
47	Virginia	6.1	22.7	99.2	154.8	392.1	2035.0	211.1	2921.0
48	Washington	3.3	44.7	92.1	205.8	959.7	3149.5	783.9	5239.0
49	West Virginia	4.4	17.7	44.6	206.1	621.2	1794.0	210.0	2898.0
50	Wisconsin	3.5	20.6	82.2	135.2	440.8	1992.8	226.6	2901.7
51	Wyoming	2.7	24.0	15.3	188.1	476.3	2533.9	145.1	3385.4

PYTHON - GENERATING HISTOGRAM.

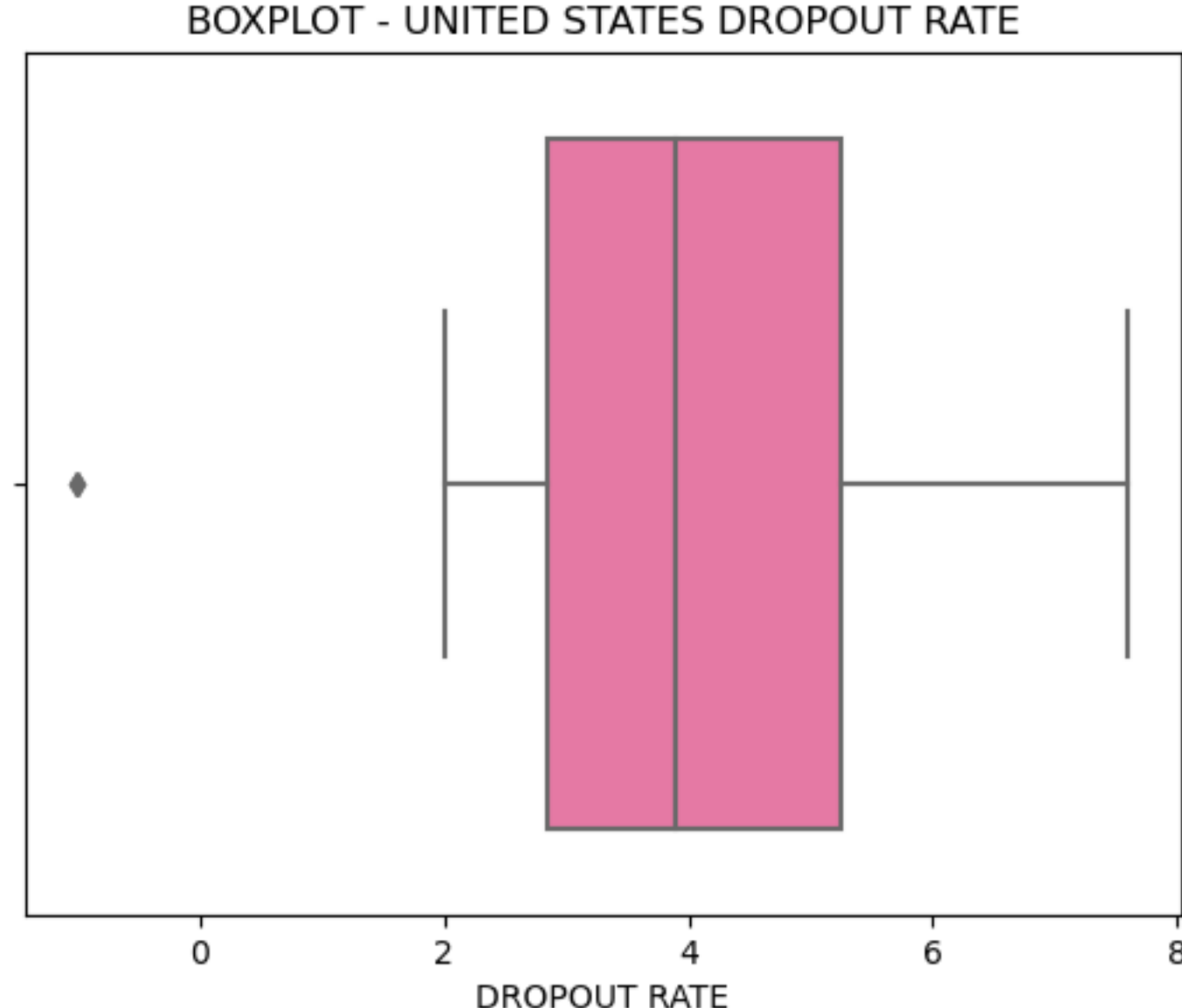
```
In [8]: plt.hist(crime_data['larceny_theft'], bins=20, color='darkmagenta')
plt.title("HISTOGRAM - DISTRIBUTION OF LARCENY THEFT")
plt.xlabel("LARCENY THEFT")
plt.ylabel("FREQUENCY")
plt.grid(True)
plt.show()
```



PYTHON - GENERATING BOXPLOT.

```
In [9]: sns.boxplot(x = 'dropout_rate', data = education_data, palette='RdPu').set(title = 'BOXPLOT - UNITED STATES DROPOUT RATE', x
```

```
Out[9]: [Text(0.5, 1.0, 'BOXPLOT - UNITED STATES DROPOUT RATE'),
Text(0.5, 0, 'DROPOUT RATE')]
```



PYTHON - GENERATING BULLET CHART.

```
In [10]: crime_data2 = crime_df[crime_df["state"].str.strip()=="Georgia"][["state", "burglary"]]
crime_data2['target'] = 900

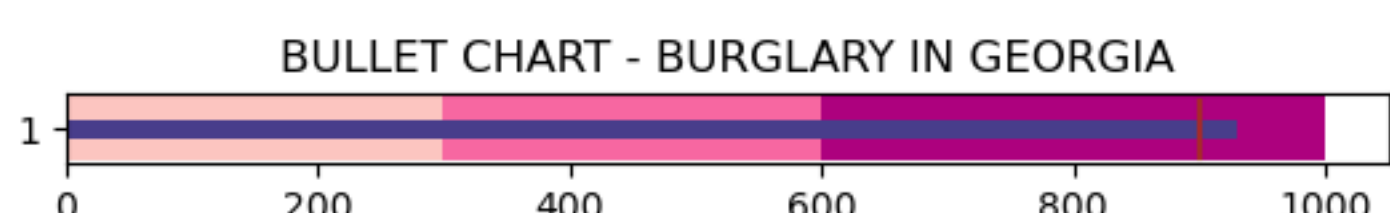
crime_data_tuple = [tuple(x) for x in crime_data2.values][0]
print(crime_data_tuple)

limits = [300, 600, 1000]
palette = sns.color_palette("RdPu", len(limits))
fig, ax = plt.subplots()
ax.set_aspect('equal')
ax.set_yticks([1])
ax.set_title("BULLET CHART - BURGLARY IN GEORGIA")

previous_limit = 0
for idx, lim in enumerate(limits):
    ax.barh([1], lim-previous_limit, left=previous_limit, height=50, color=palette[idx])
    previous_limit = lim

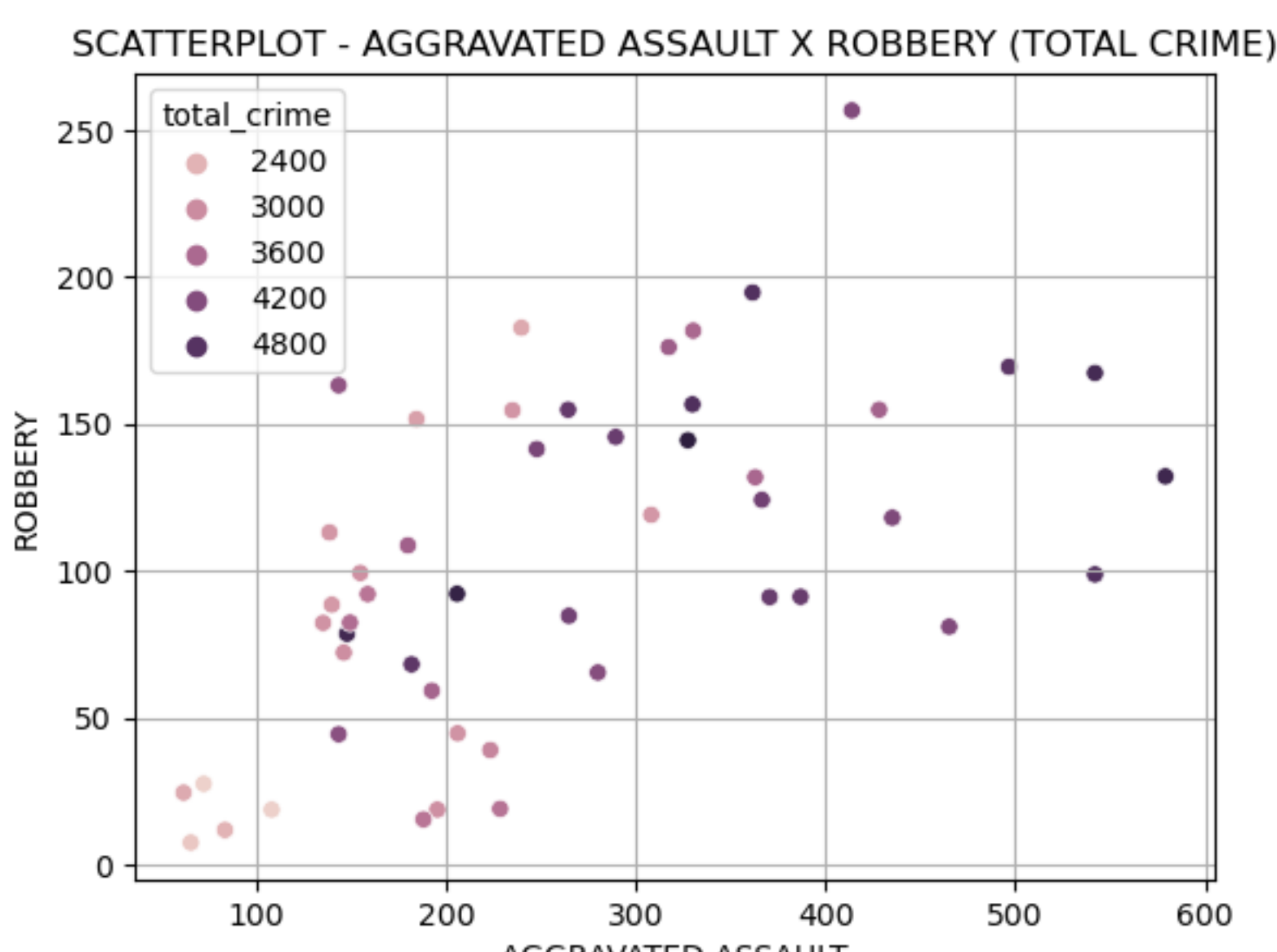
ax.barh([1], crime_data_tuple[1], color='darkslateblue', height=15)
ax.axvline(crime_data_tuple[2], color="brown", ymin=0.10, ymax=0.9)
```

```
('Georgia ', 931.0, 900)
<matplotlib.lines.Line2D at 0x140c01050>
```



PYTHON - GENERATING SCATTERPLOT.

```
In [11]: plt.figure(7)
sns.set_palette("PRGn")
sns.scatterplot(x = crime_data['aggravated_assault'], y = crime_data['robbery'], hue = crime_data['total_crime'])
plt.title("SCATTERPLOT - AGGRAVATED ASSAULT X ROBBERY (TOTAL CRIME)")
plt.xlabel("AGGRAVATED ASSAULT")
plt.ylabel("ROBBERY")
plt.grid(True)
```



References

Choosing Colormaps in Matplotlib:

<https://matplotlib.org/stable/tutorials/colors/colormaps.html>

List of named colors in matplotlib:

https://matplotlib.org/stable/gallery/color/named_colors.html

Seaborn Styling, Color:

<https://www.codecademy.com/article/seaborn-design-ii#>

In []: