Data Analysis Summary: GPA scores and hours studied

1. Project Overview:

In this project, I examined GPA scores and the number of hours studied by 193 students from a private university in the US.

2. Methodology:

Throughout my analytical exploration, I engaged in a range of activities that spanned across the use of SQL, Statistics, Excel, Google Docs and Power BI.

2.1. Data Source:

The data originates from a sample of 193 undergraduate students who took an introductory statistics course in 2012 at a private US university.

2.2 Preprocessing:

Preprocessing involved removing invalid data as well as transforming data to make it more suitable for analysis with additional data transformation performed in SQL.

2.3 Exploration:

All data exploration was conducted using SQL. The code is included in my GitHub repository for this project.

2.4 Visualization:

Visualization views were generated using SQL, and these views were then used in Power BI to help visualize key insights in the data sample. Tables were created in Google Docs and box and whisker plots in Excel.

2.5 Descriptive Analysis:

2.5.1 GPA:

Maximum GPA: 4.0 - The highest GPA observed in the dataset is 4.0.

Minimum GPA: 2.6 - The lowest GPA observed in the dataset is 2.6.

Average GPA: 3.6 - The mean GPA across the dataset is 3.6.

Median GPA: 3.6 - The middle value in the dataset.

Lower quartile: 3.4 - The GPA score below which 25% of the scores fall.

Upper quartile: 3.8 - The GPA score below which 75% of the scores fall.

2.5.2 Hours Studied:

Maximum Study Hours: 69 - The highest number of study hours observed in the dataset is 69 hours.

Minimum Study Hours: 1 - The lowest number of study hours observed in the dataset is 1 hour.

Average Study Hours: 15.7 - The mean number of study hours across the dataset is 15.7.

Median Study Hours: 14 - The middle value in the dataset.

Lower quartile: 8 - The number of hours studied below which 25% of hours studied fall.

Upper quartile: 20 - The number of hours studied below which 75% of hours studied fall.

2.5.3 Correlation:

Pearson's Correlation Coefficient: 0.07 - Pearson's correlation coefficient between GPA and study hours is 0.07, suggesting a very weak positive linear relationship between the two variables.

2.6 Frequency Tables:

The tables below show the GPA scores, the number of hours studied, and the number of students for each.

GPA Scores	
GPA Score	**Frequency**
2.0-2.9	4
3.0-3.9	176
4.0	12

Table 1

Hours Studied	
Hours Studied	**Frequency**
1-10	83
11-20	68
21-30	26
31-40	9
41-50	3
51-60	2
61-70	1

Table 2

3. Key Findings:

- 92 student's scored above the average GPA score.
- Majority of the students have a score between 3.0-3.9 with most students having a score of 3.8 (34 students). (Table 1 and Figure 1)
- There is a weak linear relationship between GPA scores and the number of hours studied. (Figure 2)
- Although the correlation coeffienct is only 0.07, it still does indicate **a weak positive correlation** between GPA score and hours studied. (Figure 2)
- Students with a GPA score of 3.7 spend more time studying on average (18.9 hours) compared to anyone else. (Figure 3)
- Most students typically study for 10 hours (33 students). (Figure 4)
- The GPA data is likely to be relatively symmetric and not heavily skewed.
- The hours studied data could show slight positive skewness due to the maximum value being significantly higher than the mean and median, and the lower quartile being considerably smaller than the mean and median.

4. Visualizations:

Figure 1

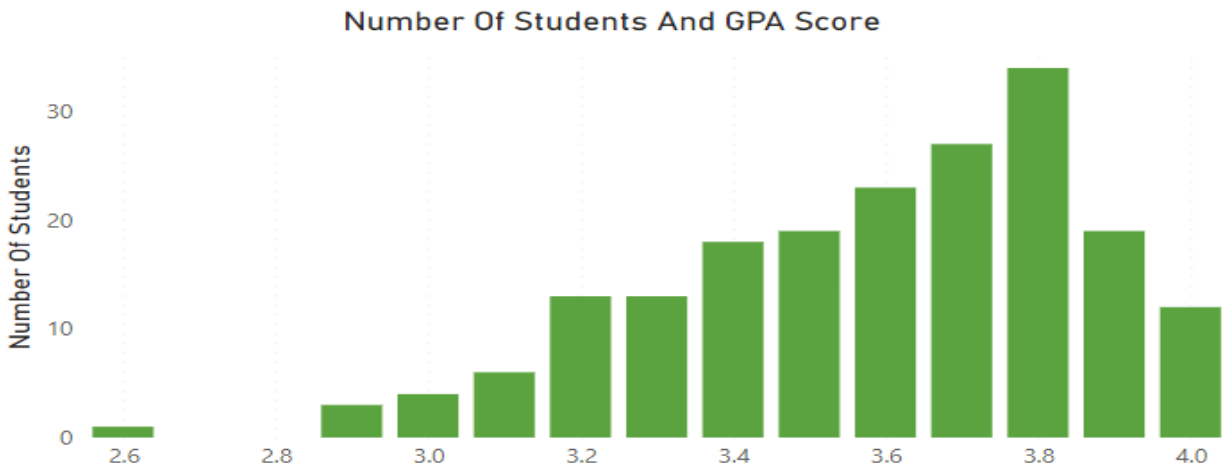


Figure 2

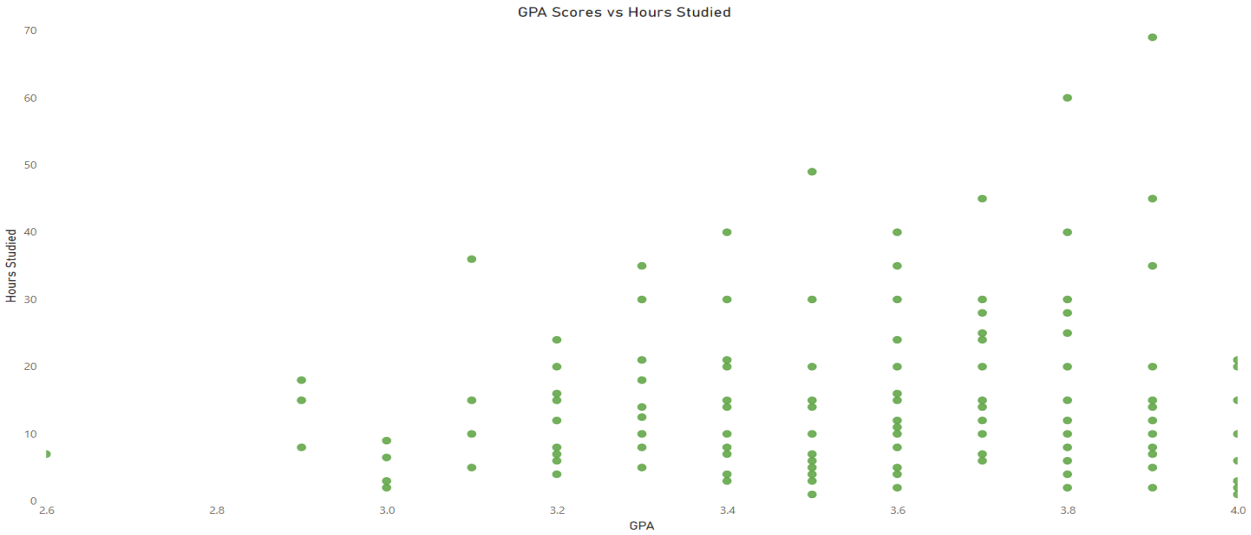


Figure 3

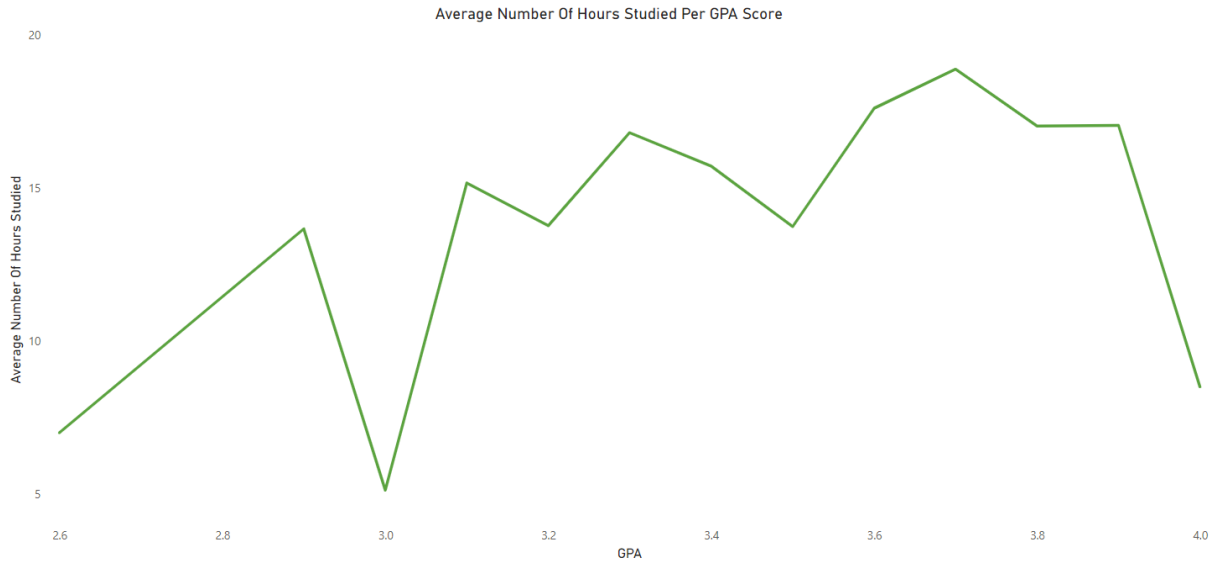


Figure 4

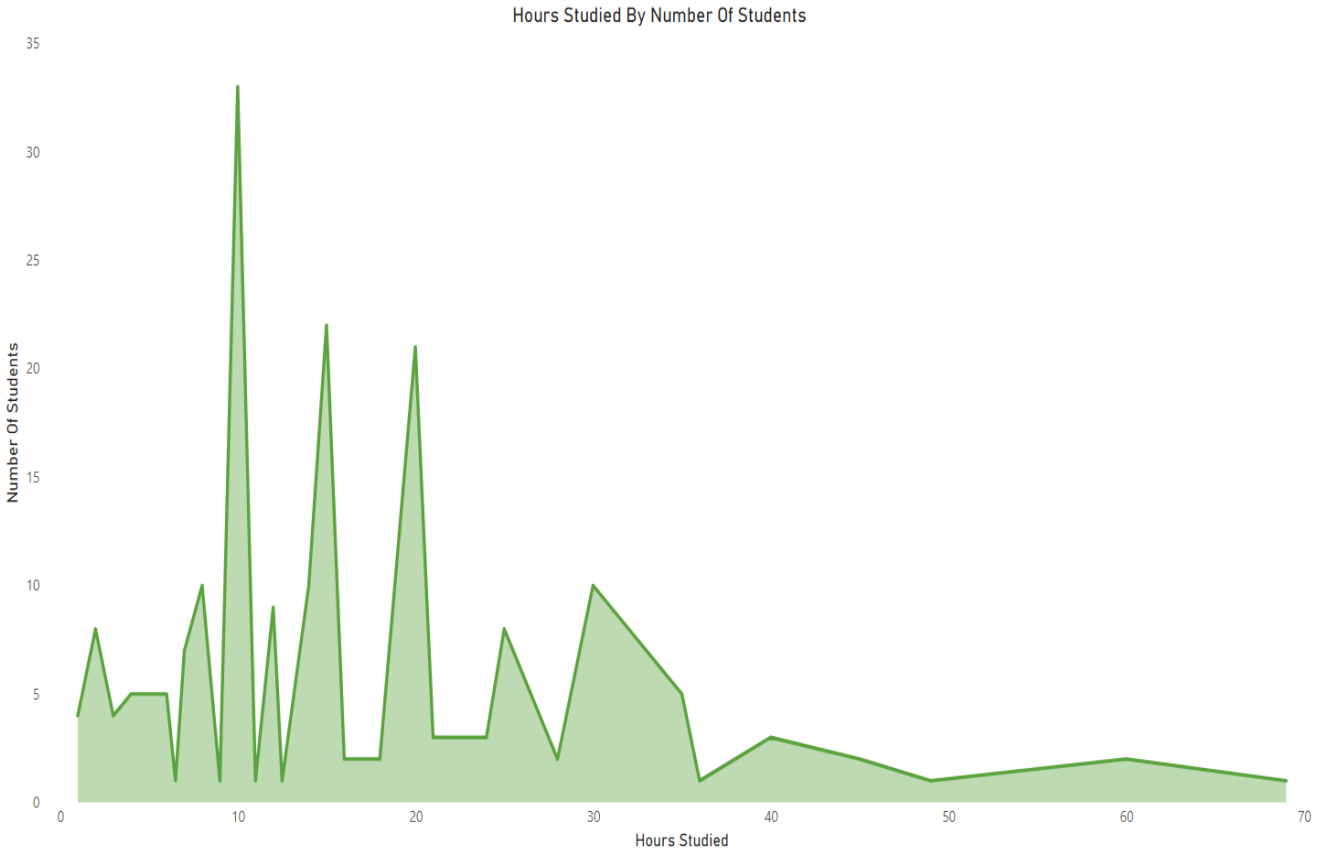
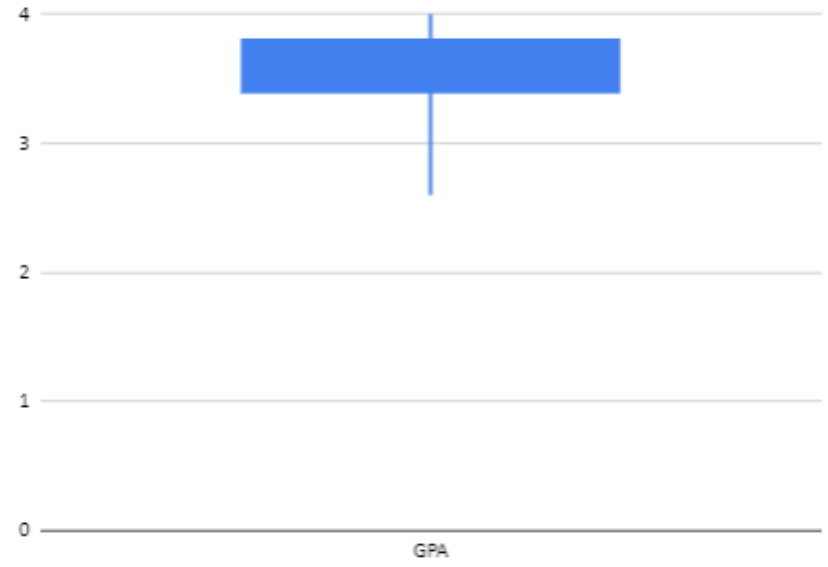
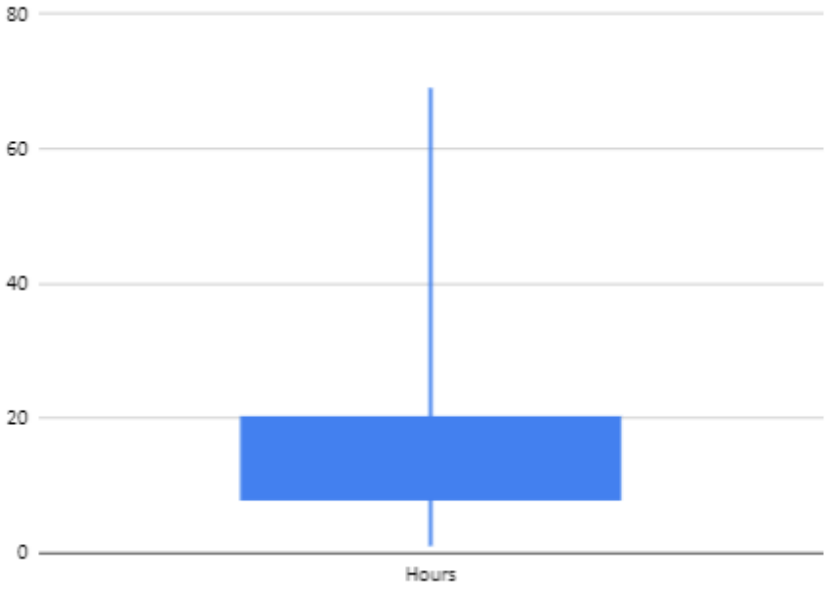


Figure 5



Box and whisker plot for GPA scores

Figure 6



Box and whisker plot for hours studied

5. Remarks:

Data collected from one student was removed due to it being invalid.

6. Data:

https://www.kaggle.com/datasets/joebeachcapital/gpa-study-hours