



# STAT 3010 EXCEL PROJECT

Rebekah Sander

## Abstract

This report provides an exploratory data analysis of the Student Survey data for University of California Davis. The data was collected on a sample of 173 students and surveyed different key variables.

9/19/2022

---

**Getting to Know the Dataset:**


---

1.

Table 1: Data Dictionary of Analysis Variables for University of California Davis Student Survey				
VARIABLE	LABEL	GENEREAL TYPE	SPECIFIC TYPE	MEASUREMENT UNITS
SEX	Biological Sex	Categorical		M= male, F= female
SEAT	Seat location in classroom	Categorical		Front, Back , Middle
LIBARTS	Liberal art major	Categorical		LibArt, NonLib
TV	average time watching tv	Quantitative		hours per week
COMPUTER	average time on a computer	Quantitative		hours per week
SLEEP	average amount of sleep	Quantitative		hours per night
ALCOHOL	consumed alcohol	Quantitative		beverages per week
HEIGHT	height	Quantitative		inches
MOMHEIGHT	height of the mom	Quantitative		inches
DADHEIGHT	height of the dad	Quantitative		inches
EXERCISE	average amount of exercise	Quantitative		hours per week
GPA	current grade point average	Quantitative		points

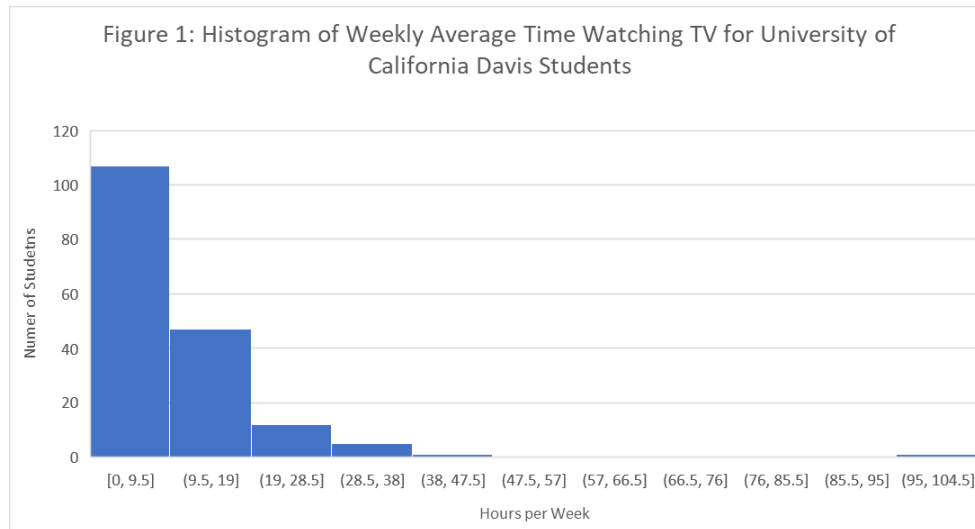
---

**Univariate Quantitative Analysis:**


---

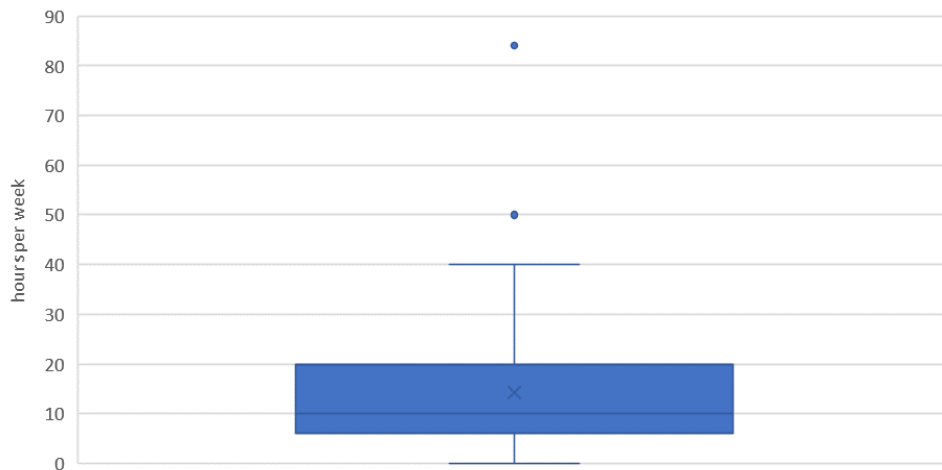
Table 2: Descriptive Statistics for Quantitative Survey Responses									
STATISTIC	TV	COMPUTER	SLEEP	ALCOHOL	HEIGHT	MOMHEIGHT	DADHEIGHT	EXERCISE	GPA
Mean	8.9	14.3	6.9	4.1	66.8	63.4	69.1	4.5	2.9
Median	6.0	10.0	7.0	1.0	66.0	63.0	69.0	3.0	3.0
Standard Deviation	10.4	11.5	1.7	8.0	3.9	3.1	3.7	4.5	0.6
Interquartile Range	10.5	14.0	2.0	4.5	5.5	4.0	5.0	4.0	0.8
Range	100.0	84.0	10.0	55.0	20.0	26.0	23.0	30.0	3.5
Minimum	0.0	0.0	2.0	0.0	57.0	54.0	55.0	0.0	0.5
Quartile 1	2.0	6.0	6.0	0.0	64.0	61.0	67.0	2.0	2.5
Quartile 3	12.5	20.0	8.0	4.5	69.5	65.0	72.0	6.0	3.3
Maximum	100.0	84.0	12.0	55.0	77.0	80.0	78.0	30.0	4.0
Sample Size	173.0	171.0	173.0	167.0	171.0	170.0	167.0	172.0	164.0

2.

**3. Graphics**

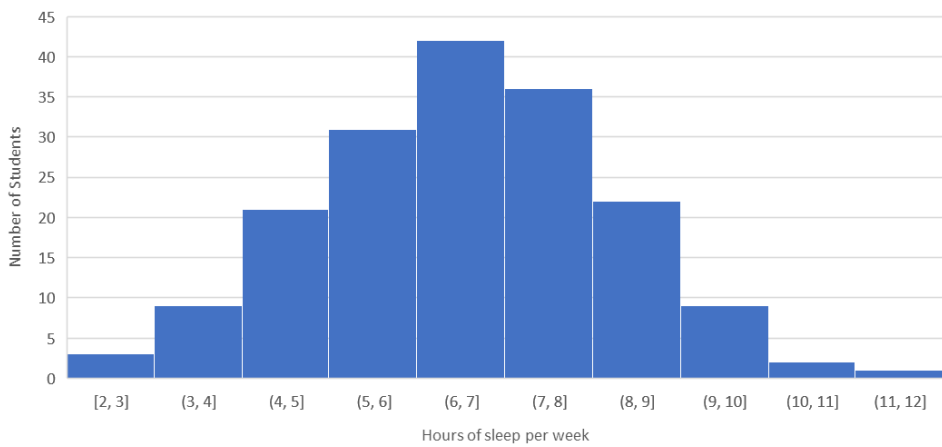
a.

Figure 2: Box Plot of Average Time Spent on a Computer for University of California Davis Students



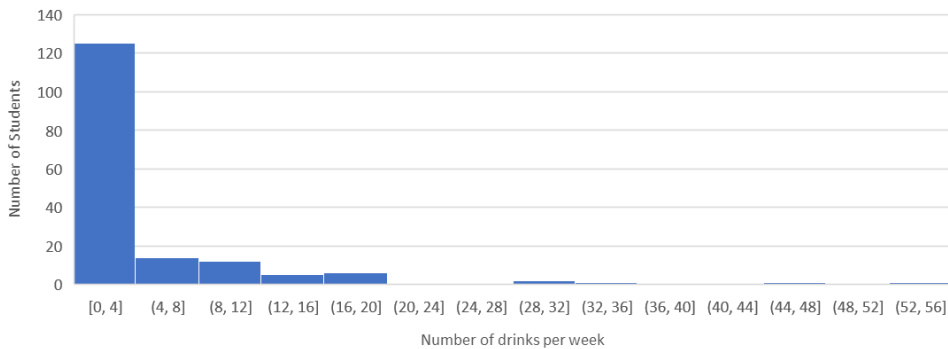
b.

Figure 3: Histogram of Weekly Average of Sleep for University of California Davis Students



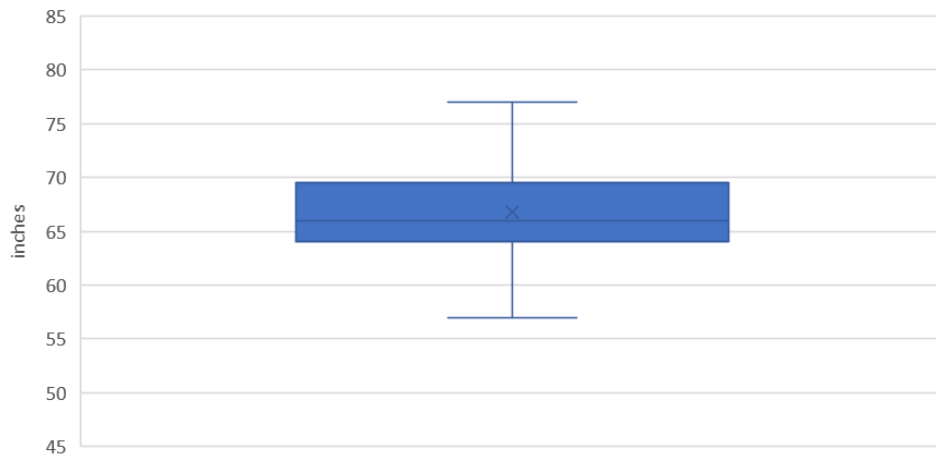
c.

Figure 4: Histogram of Average number of Alcoholic drinks for University of California Davis Students



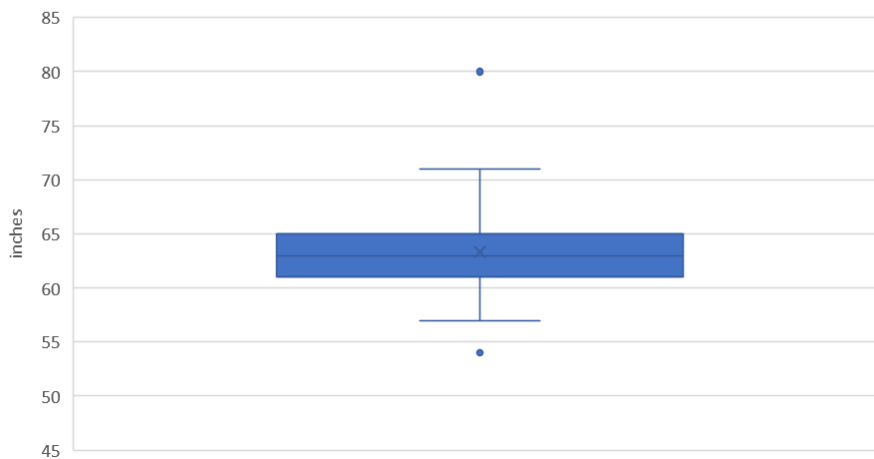
d.

Figure 5: Box Plot of Average Height for University of California Davis Students



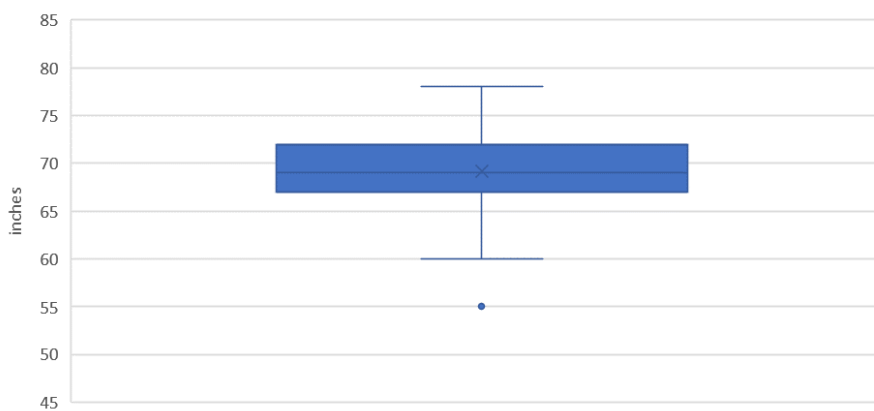
e.

Figure 6: Box Plot of Average Height for University of California Davis Students' Mom

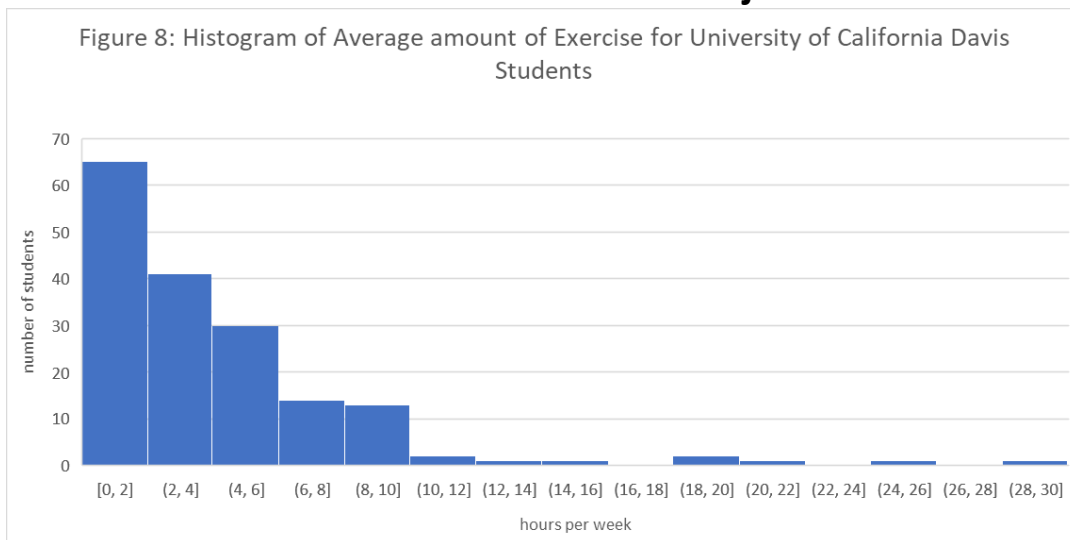


f.

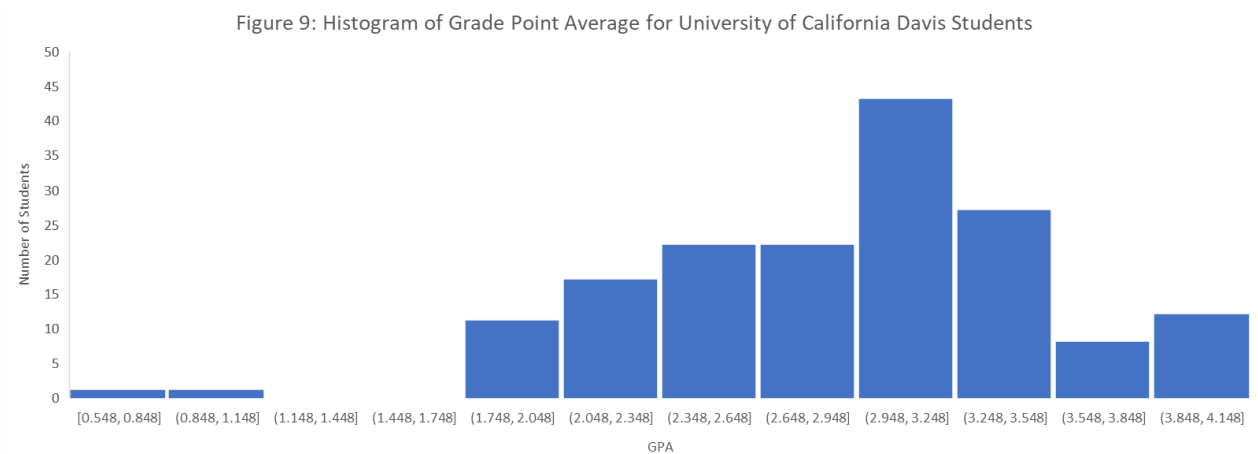
Figure 7: Box Plot of Average Height for University of California Davis Students' Dad



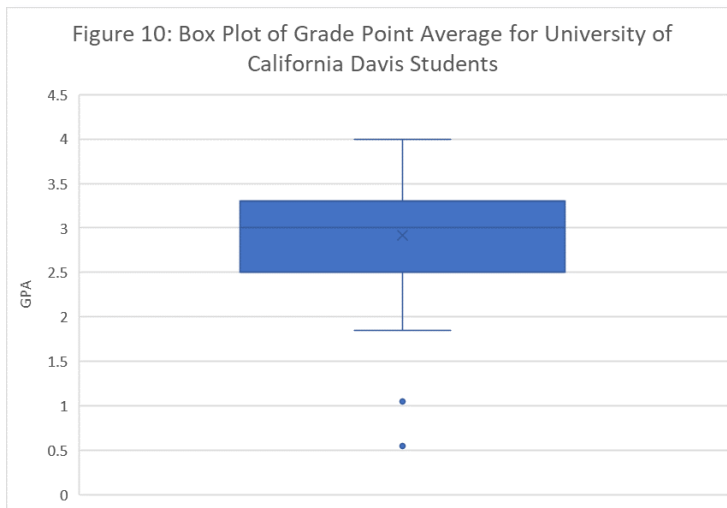
g.



h.



i.



4. Respectively,

- For the weekly average time spent watching tv, the median is the best representation of central tendency since the data is unimodal and skewed right. Because there is an outlier in the data, the interquartile range is the best representation of dispersion. There is a potential outlier of a student watching between 95-104.5 hours of TV per week. There are no missing values for this variable.
- For the weekly average time spent on the computer, the median is the best representation of central tendency since the data is skewed right. Because there is an outlier in the data, the interquartile range is the best

representation of dispersion. There is a potential outlier of a student spending between 80-90 hours on the computer per week. There is 1% of values missing for this variable.

- c. For the weekly average time spent watching tv, the mean is the best representation of central tendency since the data is unimodal and symmetric. Because there is no outliers in the data, the standard deviation is the best representation of dispersion. There are no missing values for this variable.
- d. For the weekly average number of alcoholic drinks, the median is the best representation of central tendency since the data is unimodal and skewed right. Because there are outliers in the data, the interquartile range is the best representation of dispersion. There are potential outliers of students drinking between 28-56 drinks per week. There is 3% of values missing for this variable.
- e. For the average height of a student, the mean is the best representation of central tendency since the data is symmetric. Because there is no outlier in the data, the standard deviation is the best representation of dispersion. There is 1% of values missing for this variable.
- f. For the average height of a student's mom, the median is the best representation of central tendency since the data is slightly skewed right. Because there are outliers in the data, the interquartile range is the best representation of dispersion. There are potential outliers of a student's mom's height at 80 inches and another potential outlier between 50 and 55 inches. There is 2% of values missing for this variable.
- g. For the average height of a student's dad, the median is the best representation of central tendency since the data is slightly skewed left. Because there are outliers in the data, the interquartile range is the best representation of dispersion. There is a potential outlier of a student's dad's height at 55 inches. There is 2% of values missing for this variable.
- h. For the weekly average time spent exercising, the median is the best representation of central tendency since the data is skewed right. Because there are potential outliers in the data, the interquartile range is the best representation of dispersion. There are potential outliers of a student spending between 18-30 hours exercising per week. There is 1% of values missing for this variable.
- i. For the grade point average of students, the median is the best representation of central tendency since the data is unimodal and skewed left. Because there are potential outliers in the data, the interquartile range is the best representation of dispersion. There is a potential outlier of a student having between a 0.55-1.15 grade point average. There is 5% of values missing for this variable.

---

### Univariate Categorical Analysis:

---

5.

Table 3: Frequency Distribution of Gender for UCDavis Students (n=173)			
	Frequency	Relative Frequency	Percent
Female	94	0.54	54%
Male	79	0.46	46%
Total	173	1.00	100%

a.

b.

	Frequency	Relative Frequency	Percent
Back	37	0.21	21%
Front	41	0.24	24%
Middle	93	0.54	54%
No Response	2	0.01	1%
Total	173	1.00	100%

c.

	Frequency	Relative Frequency	Percent
LibArts	25	0.14	14%
NonLib	148	0.86	86%
Total	173	1.00	100%

6.

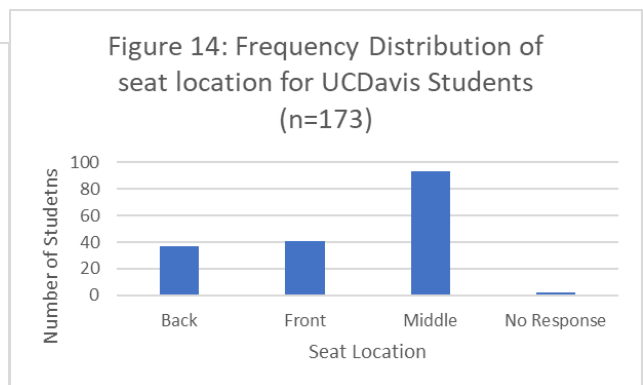
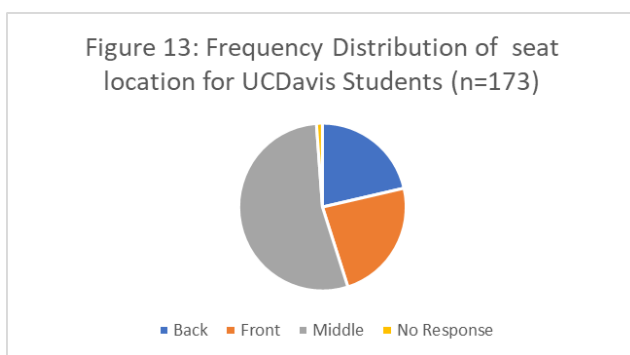
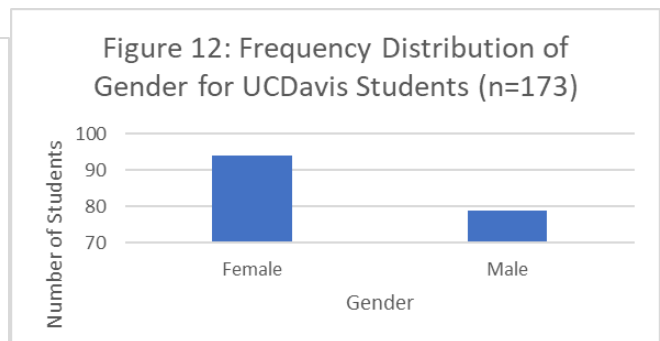
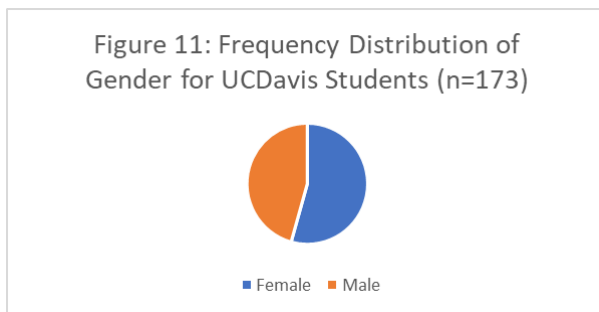
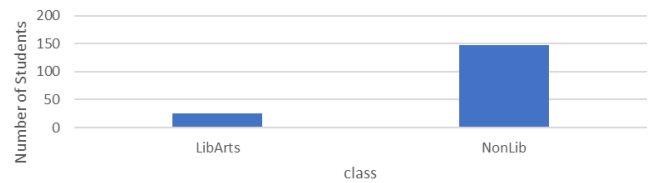




Figure 15: Frequency Distribution of Students in Liberal Arts for UCDavis Students (n=173)



Figure 16: Frequency Distribution of Students in Liberal Arts for UCDavis Students (n=173)



c.

7. Respectively,

- The gender for UCDavis Students is about equally distributed with the mode being 54% female.
- The distribution of the location of where UCDavis students sit has a mode of sitting in the middle of the classroom. Sitting in the front and sitting in the back combined is about equal to the number of students sitting in the middle and 1% of students did not indicate.
- The distribution of UCDavis students that are Liberal Arts majors being significant. The mode lies with non-liberal arts major at 66%.

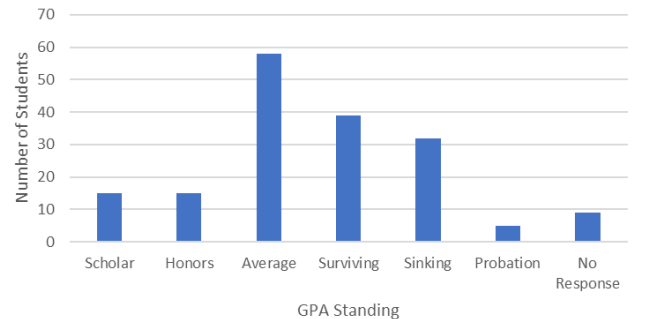
---

**Bivariate Analysis 2 Categorical:**


---

**Table 6: Frequency Distribution of GPA Standing of UC Davis Students**

	Frequency	Relative Frequency	Percent
Scholar	15	0.09	9%
Honors	15	0.09	9%
Average	58	0.34	34%
Surviving	39	0.23	23%
Sinking	32	0.18	18%
Probation	5	0.03	3%
No Response	9	0.05	5%
<b>Total</b>	<b>173</b>	<b>1.00</b>	<b>100%</b>

**Figure 17: Bar Chart of GPA Standing of UC Davis**

- 8.
9. The distribution of GPA standing of UC Davis Students has a clear majority in the “average” category.
- 10.

**Table 7: Contingency Table of GPA Standing of UC Davis Students**

GPA Standing	Sex		Total
	male	female	
Scholar	4	11	15
Honors	9	6	15
Average	26	32	58
Surviving	16	23	39
Sinking	18	14	32
Probation	2	3	5
No Response	4	5	9
<b>Total</b>	<b>79</b>	<b>94</b>	<b>173</b>

**Table 8: Row Percent Table of GPA Standing of UC Davis Students**

GPA Standing	Sex		Total
	male	female	
Scholar	5%	12%	17%
Honors	11%	6%	18%
Average	33%	34%	67%
Surviving	20%	24%	45%
Sinking	23%	15%	38%
Probation	3%	3%	6%
No Response	5%	5%	10%
<b>Total</b>	<b>100%</b>	<b>100%</b>	<b>200%</b>

a.

**Table 9: Contingency Table of GPA Standing of UC Davis Students**

GPA Standing	Seat Location				Total
	Back	Front	Middle	No Response	
Scholar	2	7	6	0	15
Honors	3	4	8	0	15
Average	11	14	32	1	58
Surviving	8	8	23	0	39
Sinking	10	5	16	1	32
Probation	0	0	5	0	5
No Response	3	3	3	0	9
<b>Total</b>	<b>37</b>	<b>41</b>	<b>93</b>	<b>2</b>	<b>173</b>

b.

**Table 10: Row Percent Table of GPA Standing of UC Davis Students**

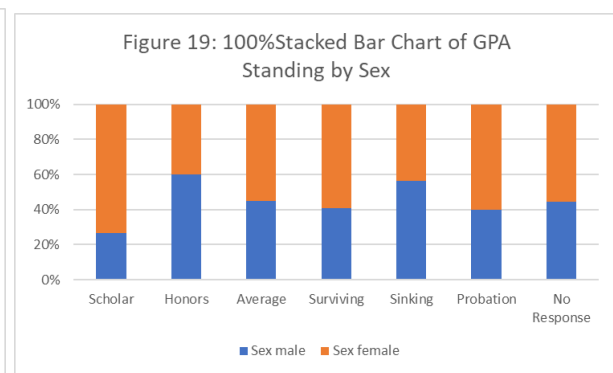
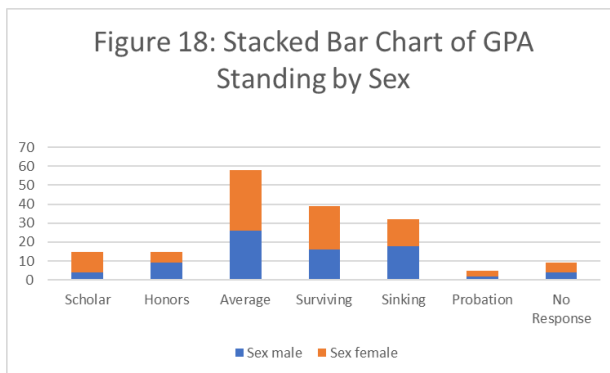
GPA Standing	Seat Location				Total
	Back	Front	Middle	No Response	
Scholar	5%	17%	6%	0%	29%
Honors	8%	10%	9%	0%	26%
Average	30%	34%	34%	50%	148%
Surviving	22%	20%	25%	0%	66%
Sinking	27%	12%	17%	50%	106%
Probation	0%	0%	5%	0%	5%
No Response	8%	7%	3%	0%	19%
<b>Total</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>400%</b>

c.

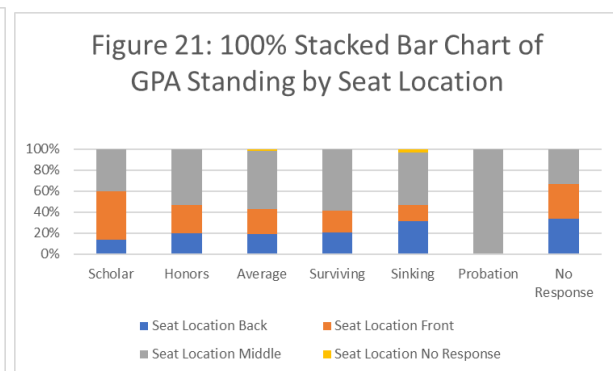
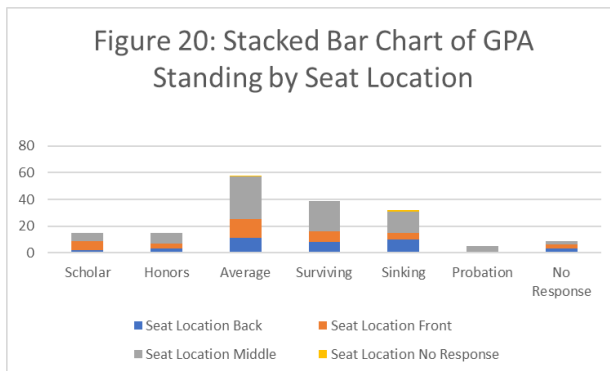
GPA Standing	Class		Total
	LibArts	NonLib	
Scholar	2	13	15
Honors	4	11	15
Average	8	50	58
Surviving	6	33	39
Sinking	4	28	32
Probation	0	5	5
No Response	1	8	9
Total	25	148	173

GPA Standing	Class		Total
	LibArts	NonLib	
Scholar	13%	87%	100%
Honors	27%	73%	100%
Average	14%	86%	100%
Surviving	15%	85%	100%
Sinking	13%	88%	100%
Probation	0%	100%	100%
No Response	11%	89%	100%
Total	93%	607%	700%

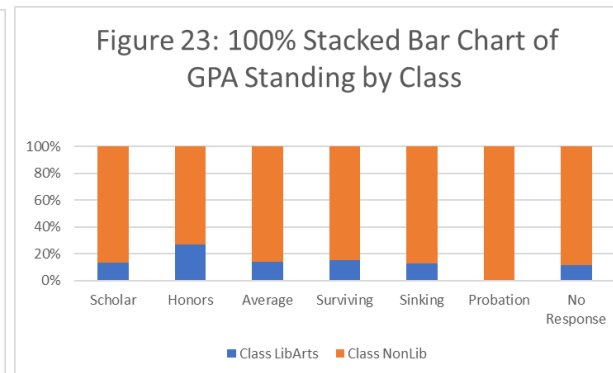
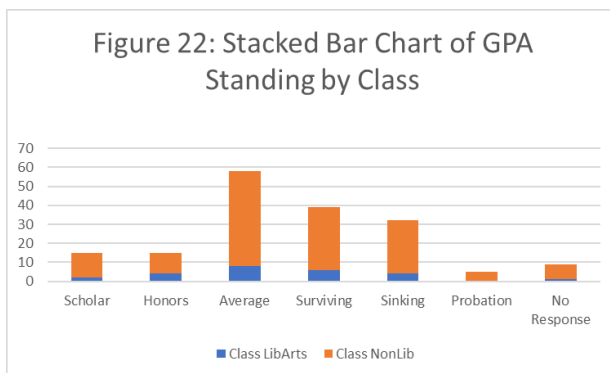
11.



a.



b.



c.

12. The row percent tables and the 100% stacked bar charts do the best at representing the relationships between the categorical data and the target GPA because we may have more values in one category

and we have to account for that. It is clear to observe from these tables that being a female and sitting at the front of the classroom may be more likely to result in a higher GPA.

---

### Bivariate Analysis 1 Cat 1 Quant:

---

13.

Table 13: Stratified Analysis of The Mean and Median GPA by Sex		
GENDER	MEAN GPA	MEDIAN GPA
Female	2.93	3.00
Male	2.90	3.00

a.

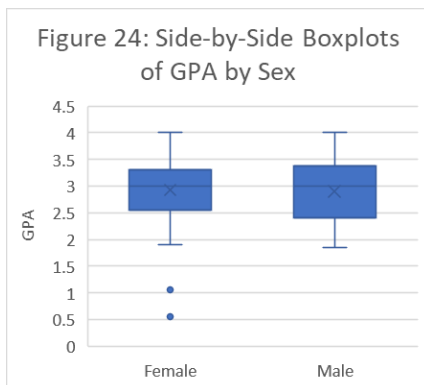
Table 14: Stratified Analysis of The Mean and Median GPA by Seat Location		
SEAT	MEAN GPA	MEDIAN GPA
Front	3.1169737	3.10
Middle	2.8502	3.00
Back	2.875	2.89

b.

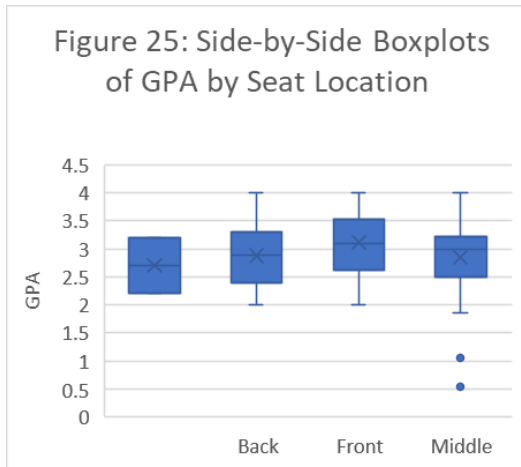
Table 15: Stratified Analysis of The Mean and Median GPA by Class		
CLASS	MEAN GPA	MEDIAN GPA
Lib Arts	3.0304167	3.05
NonLib	2.8955929	3.00

c.

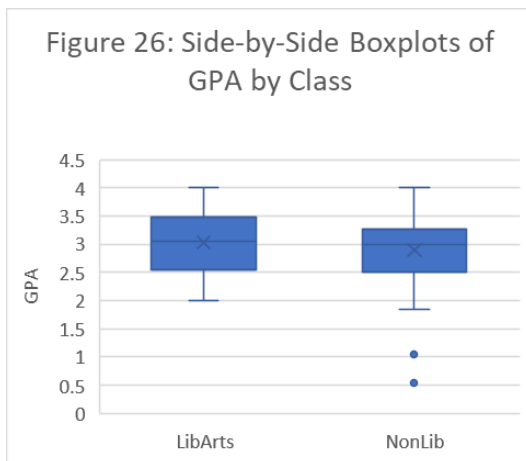
14.



a.



b.



c.

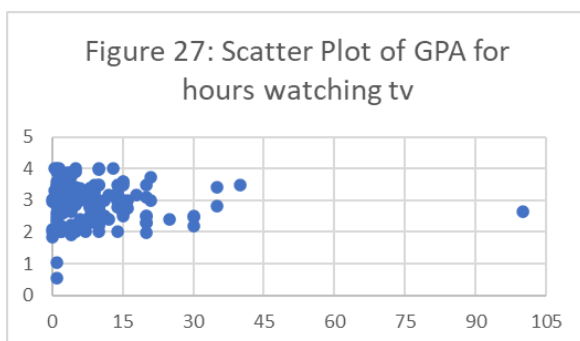
15. When getting more specific while observing the mean and median GPA for the three categorical variables, I first notice all of the outliers. We saw in the quantitative statistics that there were outliers so now we can see where they are occurring and with what group.

---

### Bivariate Analysis 2 Quant:

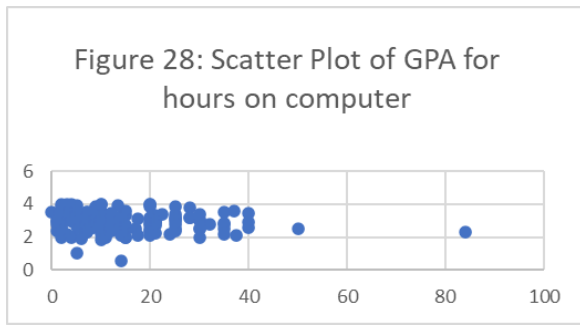
---

16.

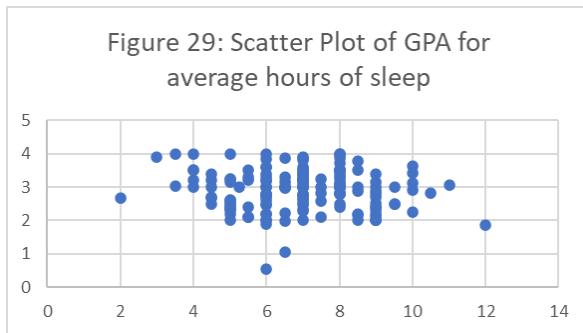


a.

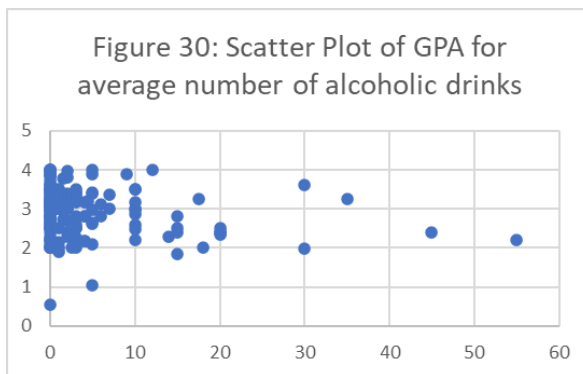
b.



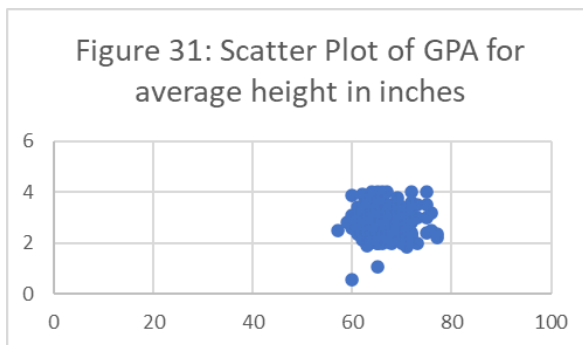
c.



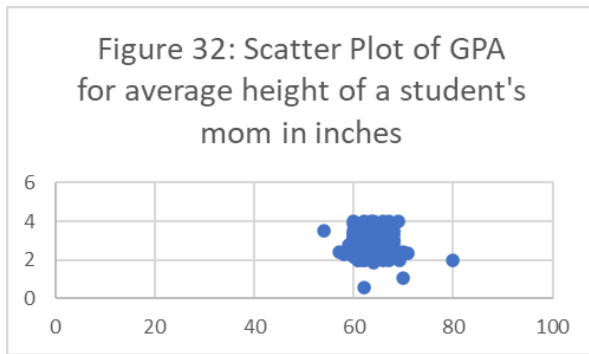
d.



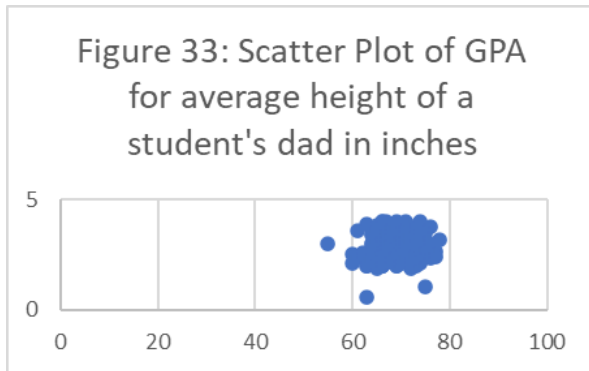
e.



f.



g.



h.

