Vanessa Trujillo
Stats 381
Project 1
February 12, 2018

Introduction: The purpose of this assignment is to take sets of data corresponding to different demographics: Gender, Career Sectors, Company sizes, Salaries, and Savings, to determine whether or not there were any correlations between these factors that determine a conclusion or spark a speculation. This data is compared and presented against each other to help us get a better understanding and a better analyzation of possible trends. All the data presented here has been analyzed and processed through Microsoft Excel.

Results:

1.)

(a)

Company	
Mean	2.821643
Standard Error	0.050075
Median	3
Mode	4
Standard Deviation	1.118596
Sample Variance	1.251258
Kurtosis	-1.20286
Skewness	-0.43033
Range	3
Minimum	1
Maximum	4
Sum	1408
Count	499
Confidence Level(95.0%)	0.098385

Gender	
Mean	1.532
Standard Error	0.022337
Median	2
Mode	2
Standard Deviation	0.499475
Sample Variance	0.249475
Kurtosis	-1.99143
Skewness	-0.12865
Range	1
Minimum	1
Maximum	2
Sum	766
Count	500
Confidence Level(95.0%)	0.043887

Salaries	
Mean	82764.68
Standard Error	1299.844
Median	83990.5
Mode	43992
Standard Deviation	29065.39
Sample Variance	8.45E+08
Kurtosis	-1.07152
Skewness	0.062446
Range	117410.6
Minimum	27225.93
Maximum	144636.5
Sum	41382342
Count	500
Confidence Level(95.0%)	2553.841

Sector	
Mean	3.268
Standard Error	0.077177
Median	3
Mode	2
Standard Deviation	1.725727
Sample Variance	2.978132
Kurtosis	-1.34922
Skewness	0.215149
Range	5
Minimum	1
Maximum	6
Sum	1634
Count	500
Confidence Level(95.0%)	0.151632

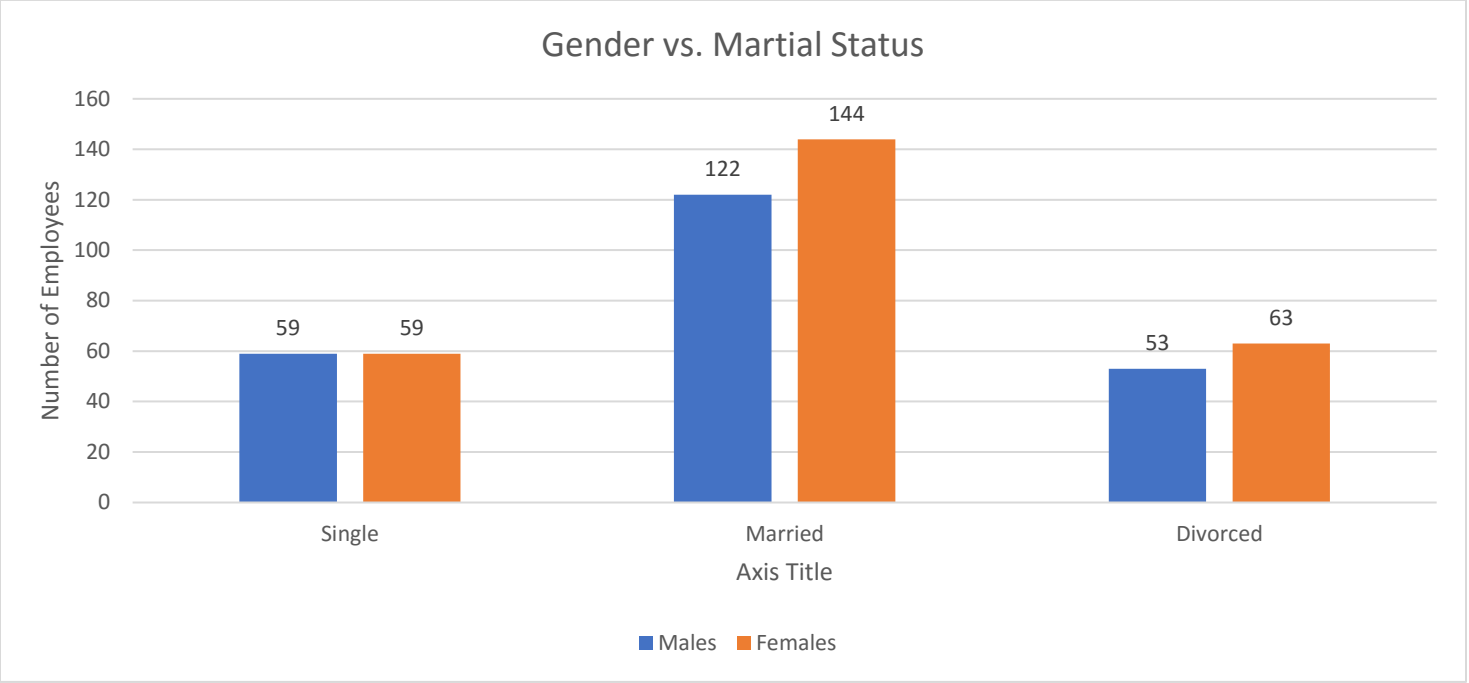
Age	
Mean	44.196
Standard Error	0.584896
Median	46
Mode	35
Standard Deviation	13.07867
Sample Variance	171.0517
Kurtosis	-0.99887
Skewness	-0.11605
Range	50
Minimum	17
Maximum	67
Sum	22098
Count	500
Confidence Level(95.0%)	1.149162

Savings	
Mean	51698.23
Standard Error	5123.081
Median	37477.5
Mode	#N/A
Standard Deviation	114555.6
Sample Variance	1.31E+10
Kurtosis	4.8372
Skewness	0.048245
Range	941940
Minimum	-466828
Maximum	475112
Sum	25849114
Count	500
Confidence Level(95.0%)	10065.47

For the variables that correspond with the Company size, Sector, Martial Status, and Gender, there does not appear to be any outliners. The variables that correspond with, Salaries, Savings, and Age do appear to have outliers that influence the mean causing it the data to be skewed far right or left.

1.)

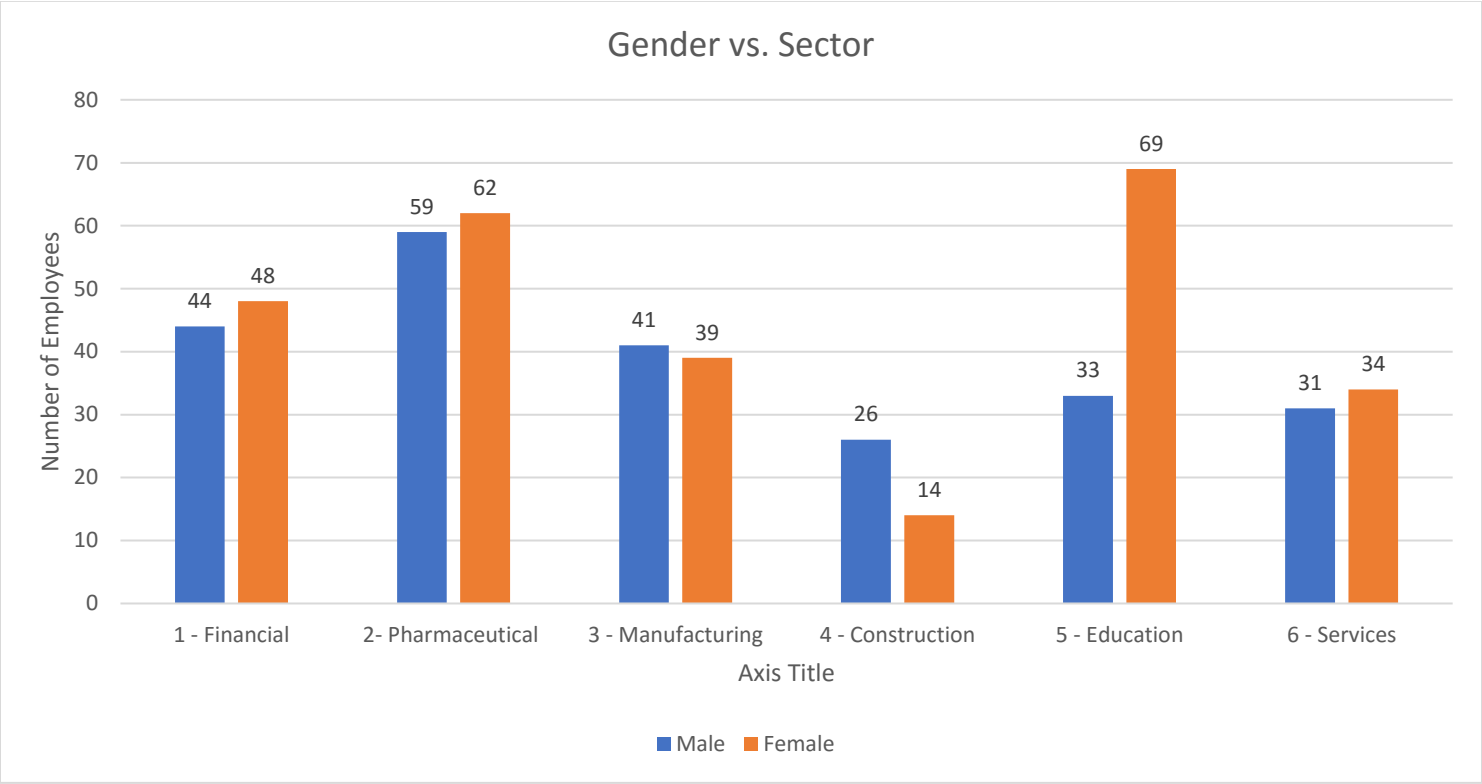
(b)



From the data displayed here we can easily see that there are more women than men. Even with the genders not exactly equal it appears males and females are evenly distributed among the three Marital Status.

1.)

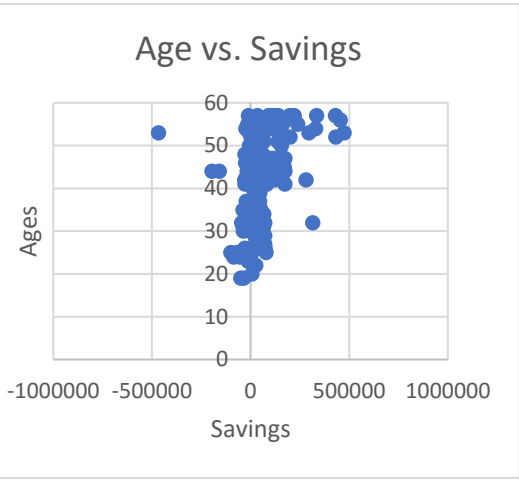
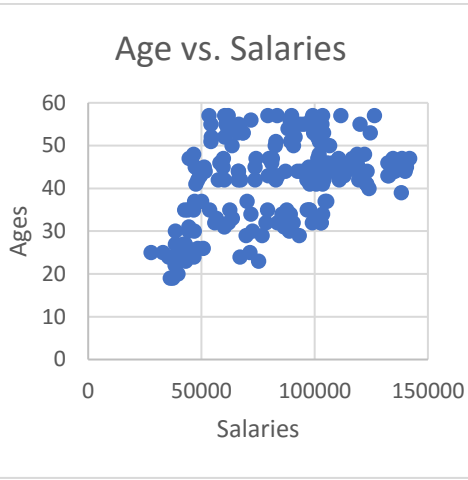
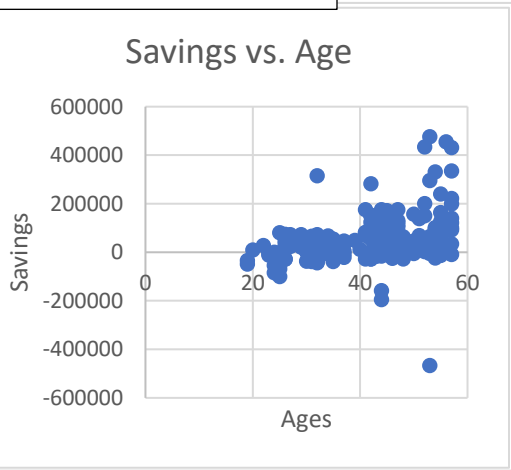
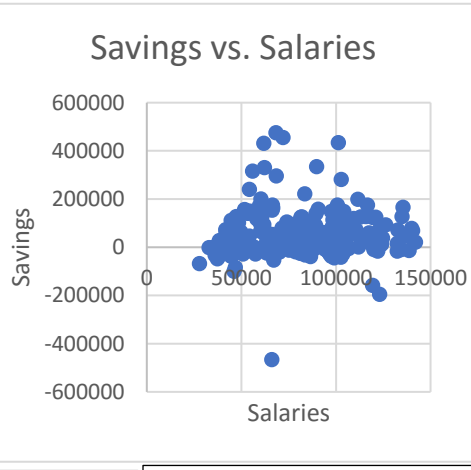
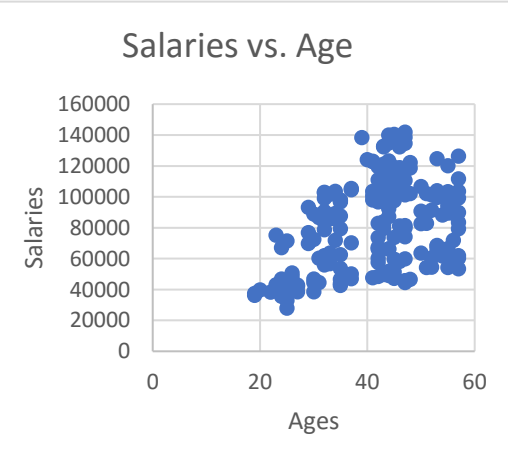
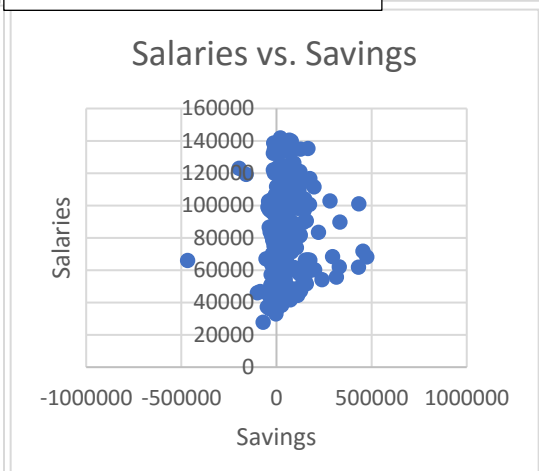
(c)



Both female's and males are closely distributed for each of the sectors, Finance, Pharmaceutical, Manufacturing, and Services. Although appears to be a large gap between number of Males and Females for Construction and Education. More men are working in Construction, while a lot more women are working in Education.

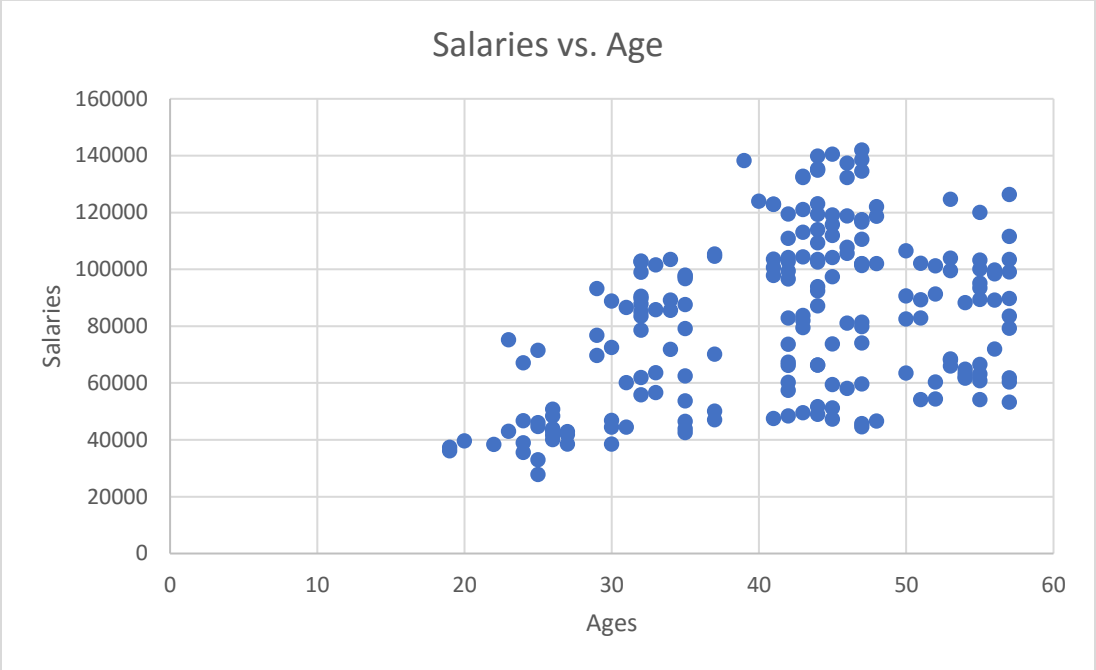
1.)

(d)

AGE		
	**SAVINGS**	
		**SALARY**

2.)

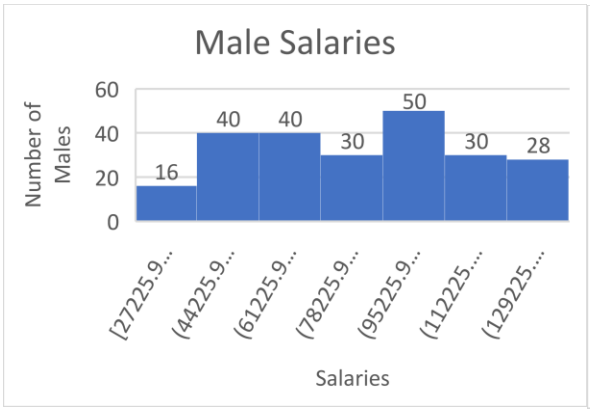
(a)

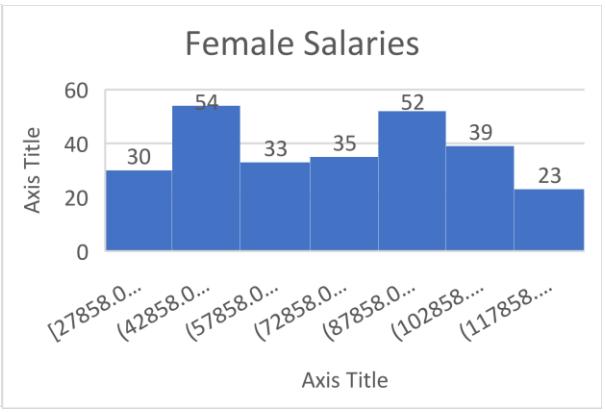


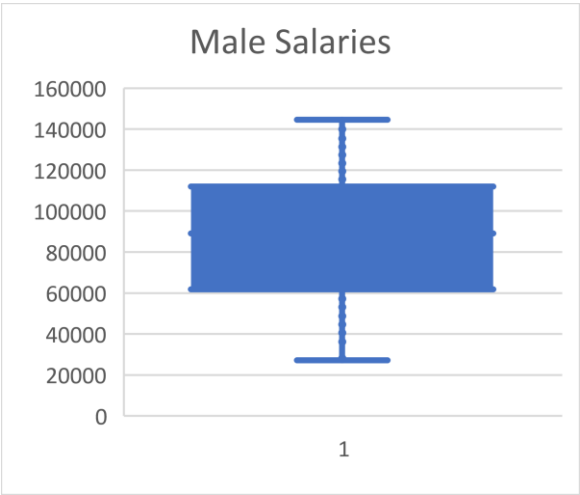
The Wedge shape of this scatter plot graph suggests that the older a person is the more money they are suggested to make, along there being a limit, after the age of 50 the income seem to be steadily decreasing. The model seems logically correct because usually starting salary are low and many “starters” begin their careers in their twenties. As they get older and gain experience their income increase over the years. For the decreasing trend people tend to work less as they age, so this trend makes sense here as well.

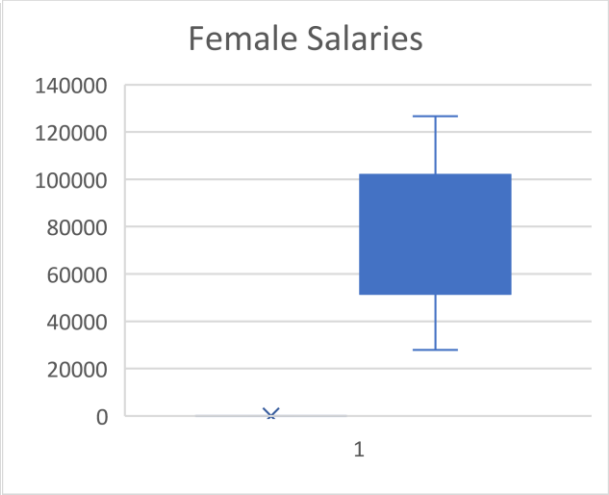
3.)

(a)





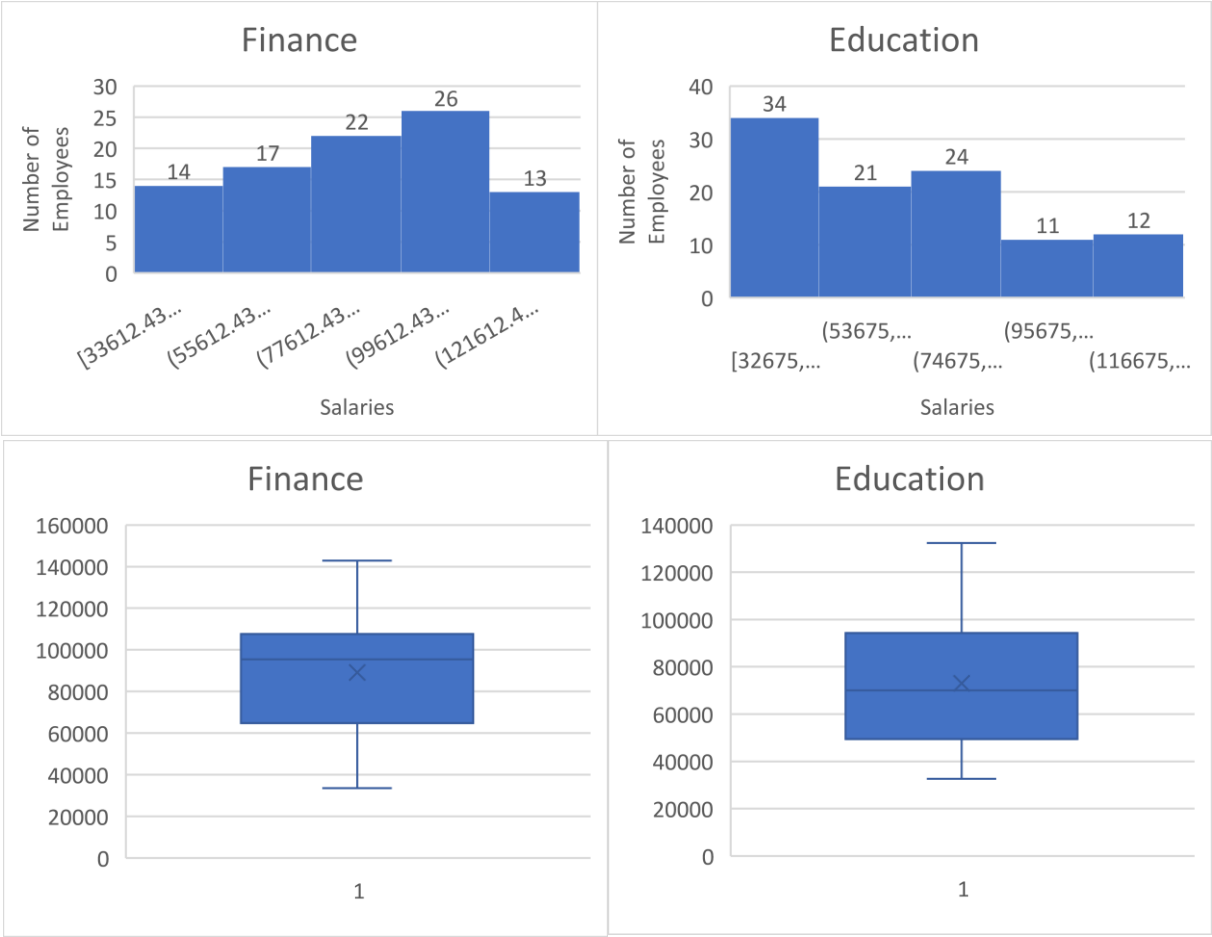




By examining the histograms, I am able to determine that the number of men is less towards the beginning. The number of men and women are closely distributed towards the center. Towards the end of the graph the number of men is higher than the number of women. By examining both of the boxplots we can see that the median for men's salary is higher than women's salaries.

3.)

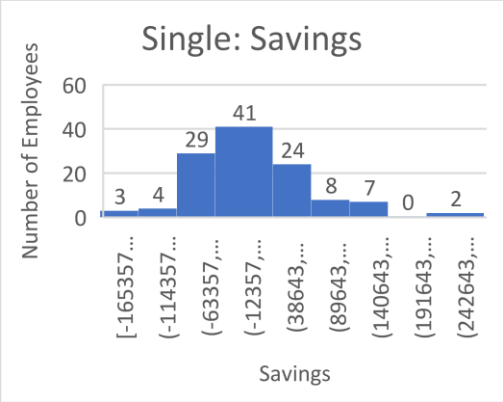
(b)

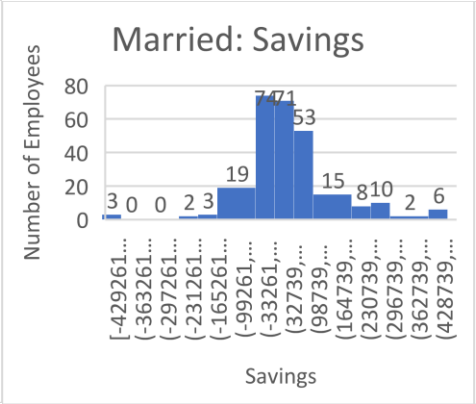


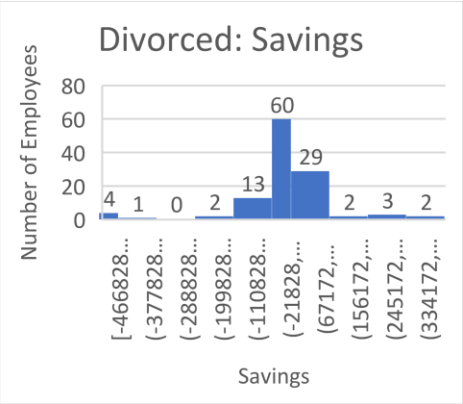
It appears as though there are 2 employees working in Education for every 1 employee in Finance. While there are more people working in the Education sector, there seems to be a greater income for those who work in Finances, where the median in Finances make around $80,000 annually and the median in Education makes $74,000 annually.

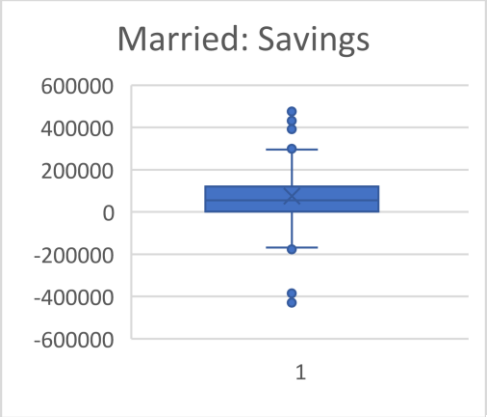
3.)

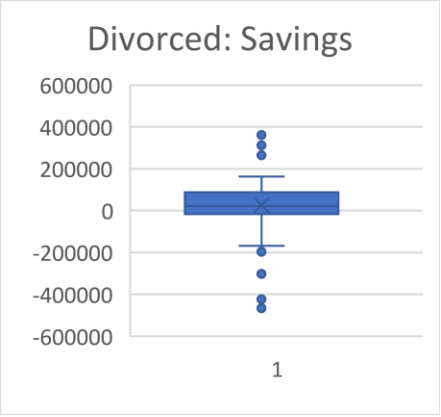
(c)











Based off of the information presented in the histograms, married people have the most savings, followed by single people and divorcees respectfully.

3.)

(d)

The distributional implications for the single and divorcee seem normal while the data for the people who are married is skewed to the left. The boxplots indicate that there are outliers in each of the data sets collected. There also seems to be outliners for single: 1 top outline (UIND:485) and 2 bottom(UIND:428,325), Married: 6 top (UIND:**108**,**260**,324,401,468,479) and 3 bottom(UIND:382,364,394), Divorced: 3 top (UIND:488,312,438,440) and 4 bottom (UIND:26,467,222,**318**,**102**), based on the boxplots. A large majority of the outliners were females that worked in Medium sized companies, while 4 of them (bolded) were men.

4.)

(a)

	Male Salaries	Female Salaries
1 - Financial	97272.73	81639.82
2- Pharmaceutical	92744.74	78598.69
3 - Manufacturing	77712.82	82186.92
4 - Construction	90024.64	89198.27
5 - Education	78048.66	70566.26
6 - Services	89063.68	77999.69

	Male Savings	Female Savings
1 - Financial	48215.07	69043.46
2- Pharmaceutical	61466.98	58772.79
3 - Manufacturing	56453.93	3197.462
4 - Construction	80055.62	82614.71
5 - Education	56392.33	32360.43
6 - Services	38409.58	64153.59

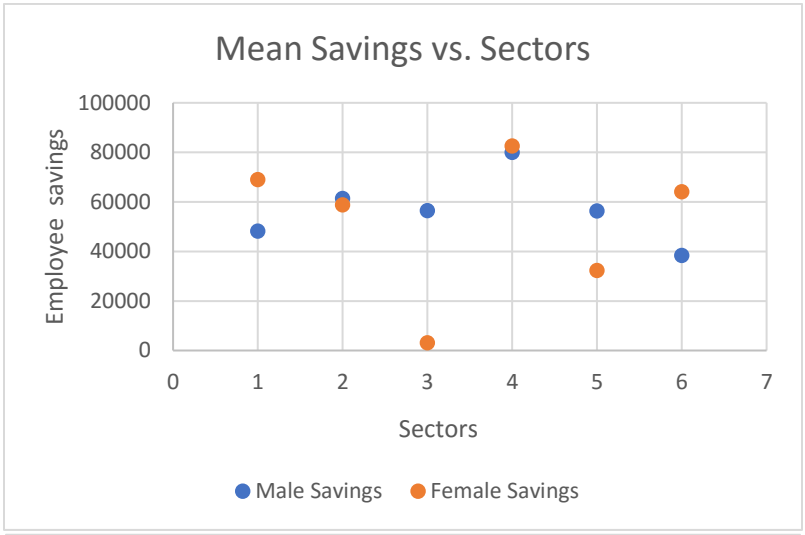
4.)

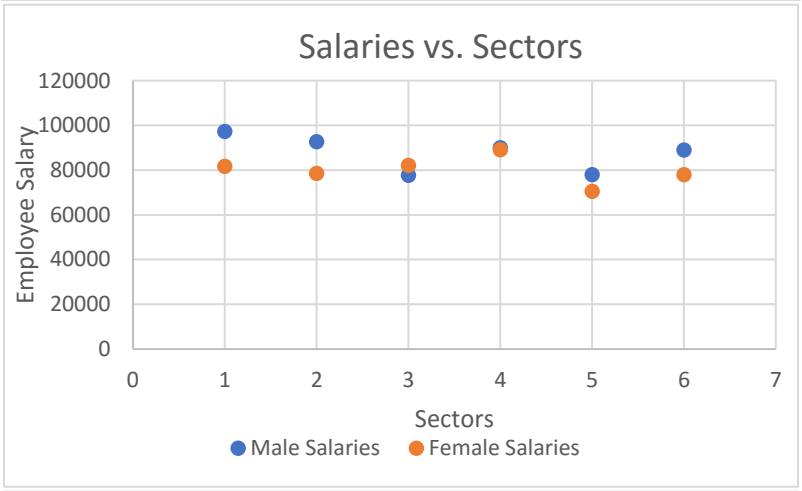
(b)

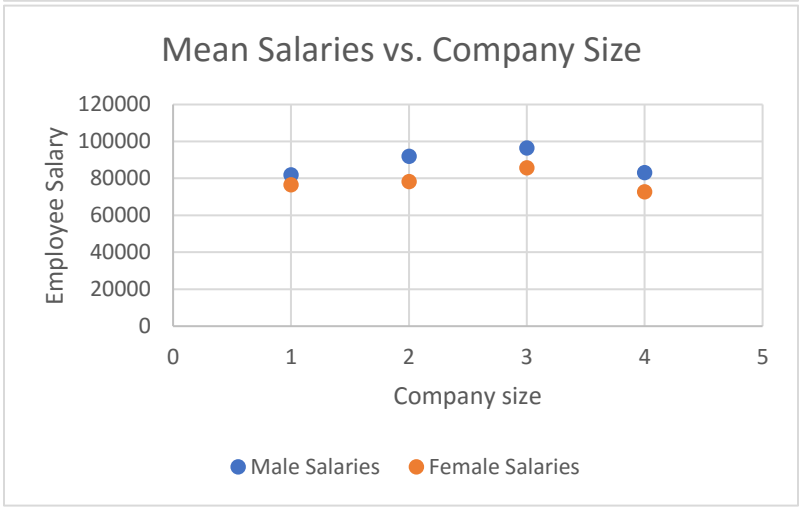
	Male Salaries	Female Salaries
1 - Small	82025.85	76654.65
2 - Small to Medium	92115.2	78359.6
3 - Medium	96492.8	85811.14
4 - Large	83269.92	72862.62

4.)

(c)







One of the biggest questions that rise from these data models is why is there such diversity in savings through out the sectors when the mean salaries between company sizes and sectors are nearly the same throughout? Based off these data models it appears that women get paid overall less then male employees, while there also seems to be evidence that teachers get paid the least in comparison to the other sectors.

Conclusions

1. The study of this given data was successful, we were able to make many conclusions based of evidence, such as supporting the claim that teachers do make less in comparison to other career paths.
2. Some plausible reasons that explain the Success or failure of the original study maybe the fact that we have a limitation on the sample we are analyzing. This is something that can account for both Success and failure of the original study, analyzing a sample that does not extend to a good sizable amount can cause large fluctuation in our models, while it can also be successful because it allows us to see data in a closer range.
3. Some of the questions I would have would be:

   How large was the population of this sample? How was this data collected? When was this data collected? How conducted the collection of this data? How and who analyzed the data and constructed the data models?
4. When drawing conclusions from observational studies and surveys it is difficult to come a conclusion about a specific person, while that may be a difficult aspect, it also makes it easier to make assumptions about a larger group of people overall. What might be learned from such studies is how large groups tend to make decisions or develop certain traits based off where they fall on a given scale.