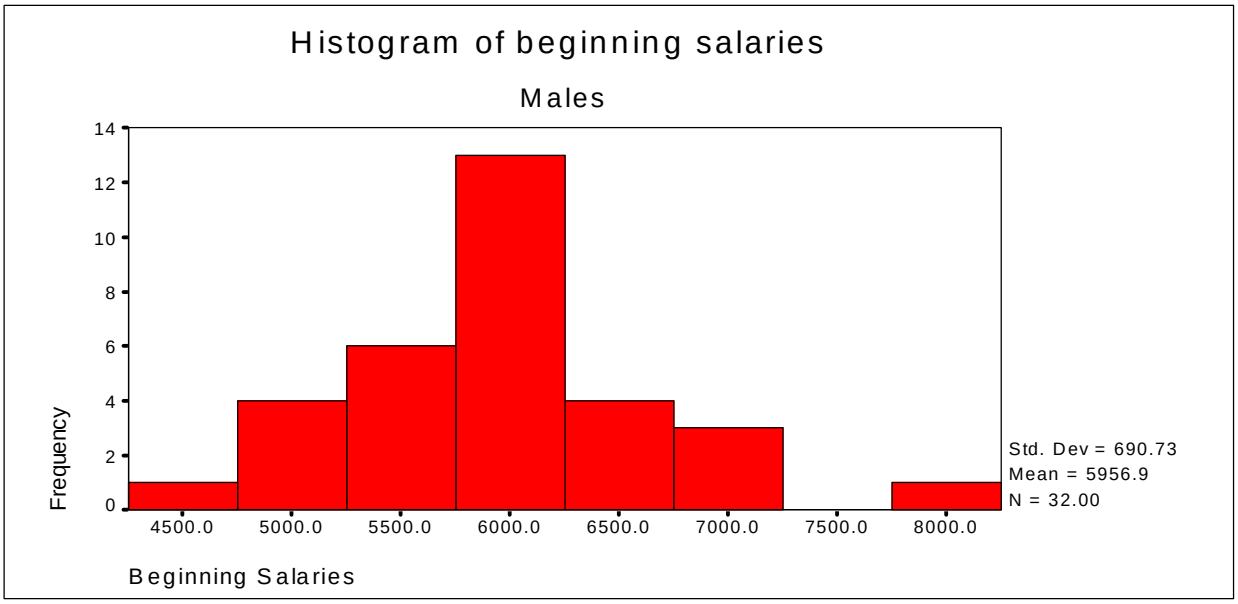
SEX DISCRIMINATION PROBLEM

4. Displaying Relevant Variables

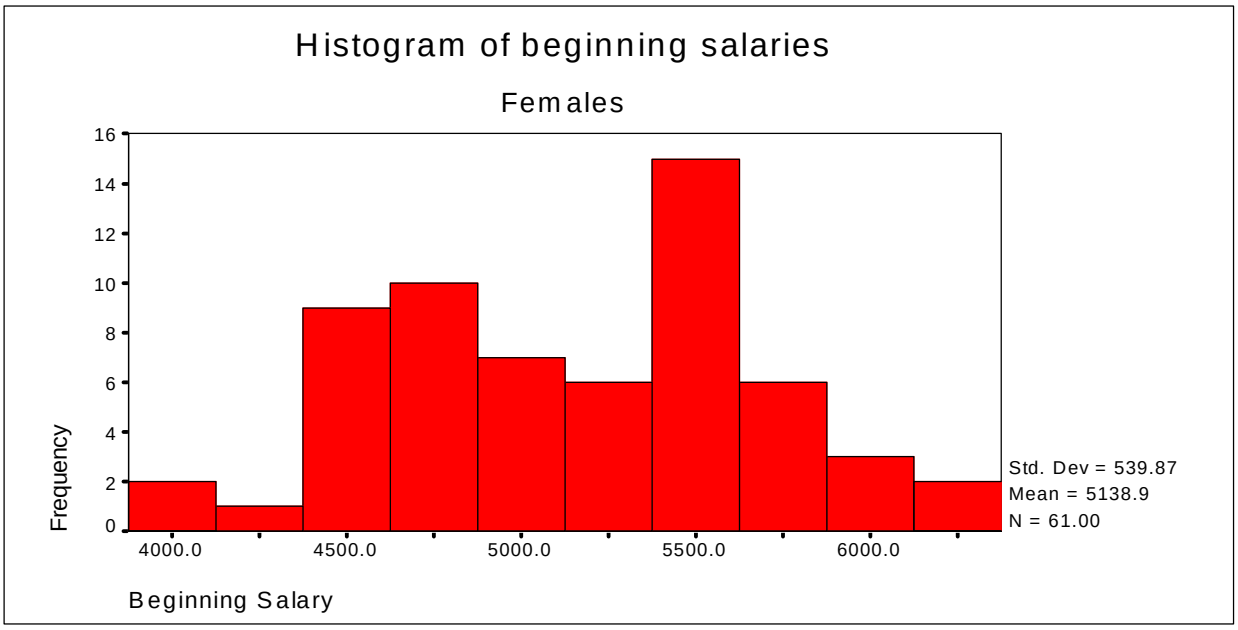
4.1 The frequency histograms for male and female starting salaries.
4.2 The side-by-side boxplots for male and female starting salaries.
4.3 Did the males tend to be younger than the females?
4.4 Did the females tend to be less educated than the males?
4.5 Did the females tend to have less experience prior to employment with the bank than the males?

4.1 First we will obtain the frequency histograms for male and female starting salaries and we describe the shape of each histogram. Then we compare the two histograms in terms of the differences between the starting salaries for males and females.

With the available data SPSS produces the following two histograms of male and female starting salaries.



The histogram of beginning salaries for males is slightly skewed to the left, but generally resembles a symmetric curve. After visual examination of the histogram, you can conclude that the central male starting salary is around $6,000. The spread of salaries about the center is not very large compared to the magnitude of values occurring in the problem. Most salaries are within about $1,000 of the center. There is one male starting salary that is unusually high relative to the rest of the male salaries.

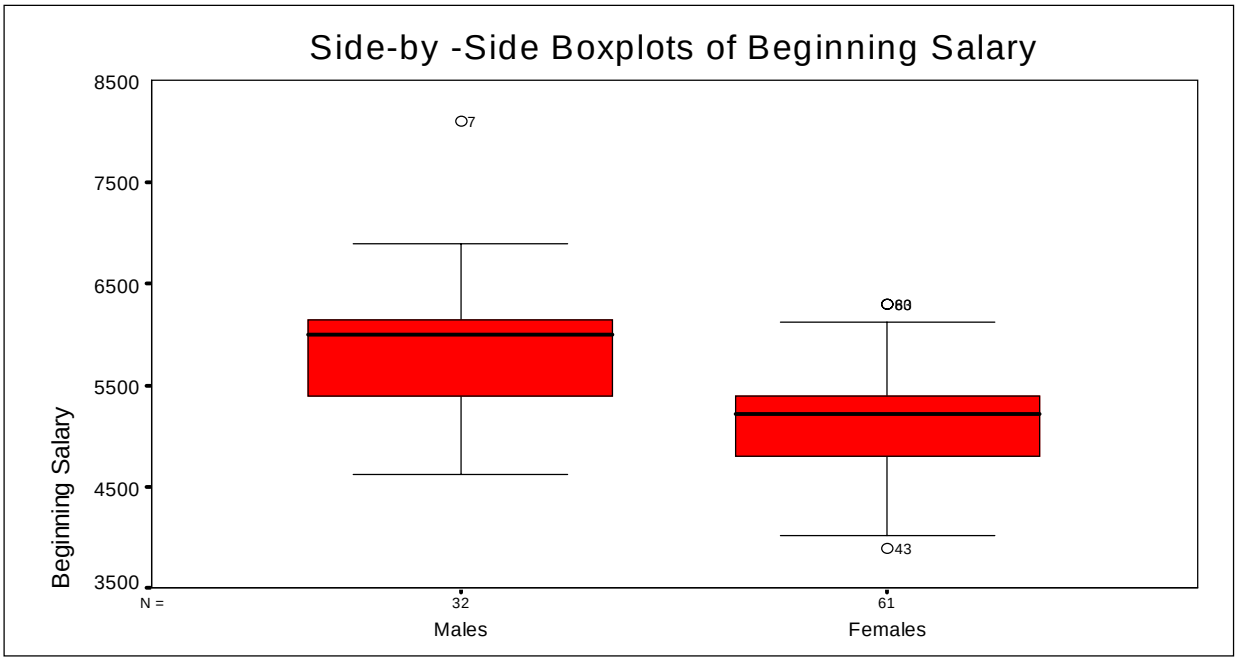


The histogram of beginning salaries for females displays heavy tails on both sides of the distribution. The shape of the histogram displays some departures from symmetry. The bulk of the observations is included in the range from 4500 to 5500. The center of the distribution is at the salary above 5000.

Comparison of both histograms shows that the distribution of male salaries is shifted by several hundred dollars to the right indicating that starting salaries of males are usually higher than the starting salaries of females. The spread of male salaries is larger than the spread of female salaries.

If you examine carefully both data sets, you will notice clusters of identical observations in both male and female data. Under these circumstances, the shapes of both histograms depend significantly on the selection of class intervals. Indeed, redefining the class intervals means shifting a bulk of the data to another class interval and as a consequence changing the appearance of the histogram.

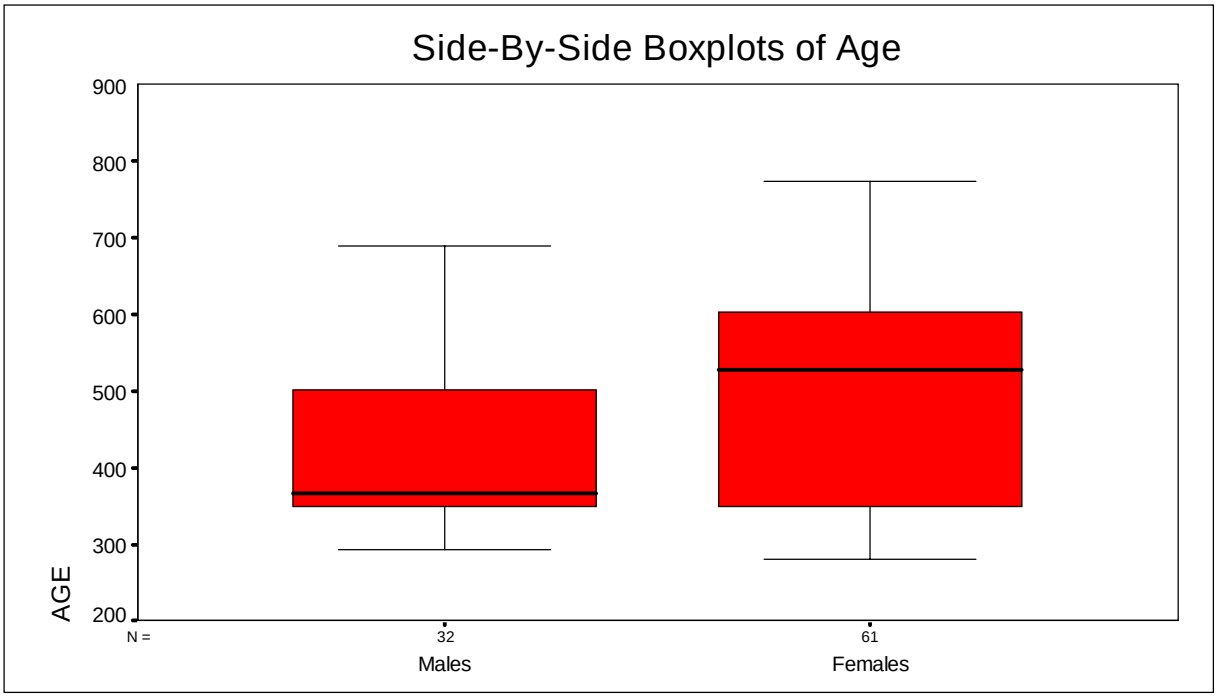
4.2 Now we will use SPSS to obtain the side-by-side boxplots for male and female starting salaries. Moreover, we will identify outliers (if any) by the 1.5*IQR criterion in each distribution.



The side-by-side boxplots confirm our conclusions reached in Section 4.1. The distribution of male starting salaries is shifted up compared to the distribution of female starting salaries. The median male starting salary indicated by the position of the horizontal line within the red box for males is significantly higher than the corresponding value for females. The spread of male observations is larger than The spread of data, represented by the width of the box (interquartile range) is larger for male than female observations.

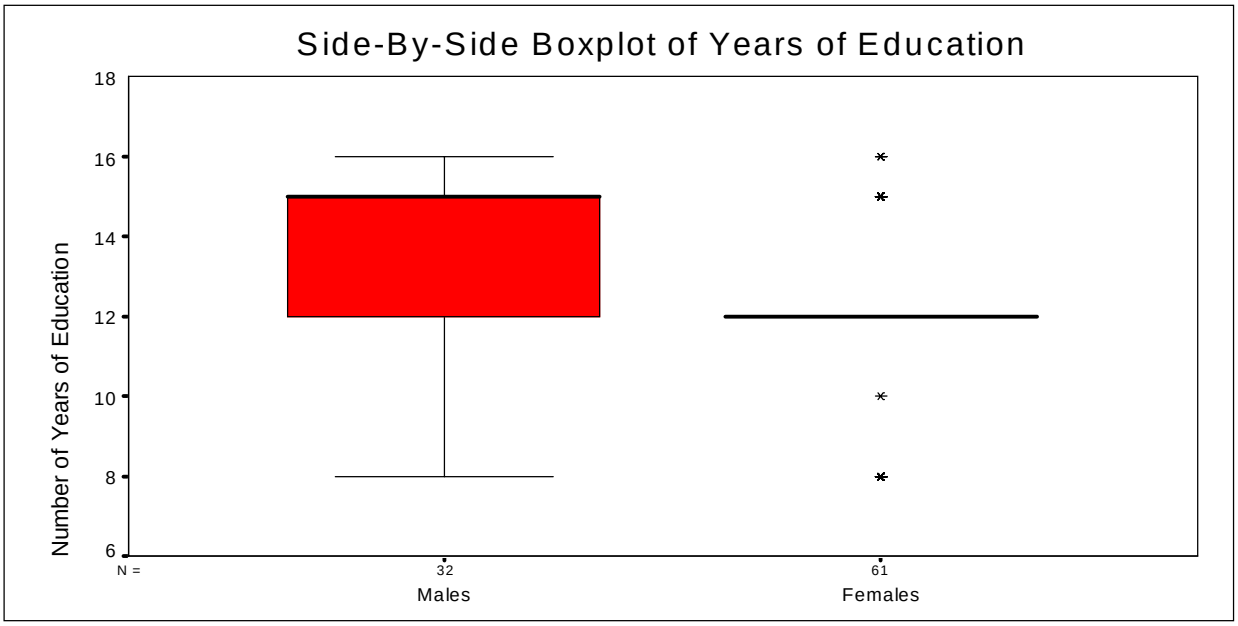
The positions of outliers in the data sets are indicated above or below the whiskers in the boxplots. Thus the seventh observation (O7) in the file is an outlier in the male data. This corresponds to the salary of $8,100. The 43rd and 80th observations are outliers in the female data. This corresponds to $3,900 and $6,300, respectively.

4.3 Did the males tend to be younger than the females? We will answer the question by obtaining the side-by-side boxplots of age for males and females.



It is easy to notice that the male employees tend to be younger than the female employees in the bank. The median age of males is below 400 months (over 30 years old) whereas the median age of females is above 500 months (over 40). Notice much larger spread in the ages of females compared to the spread of ages of males.

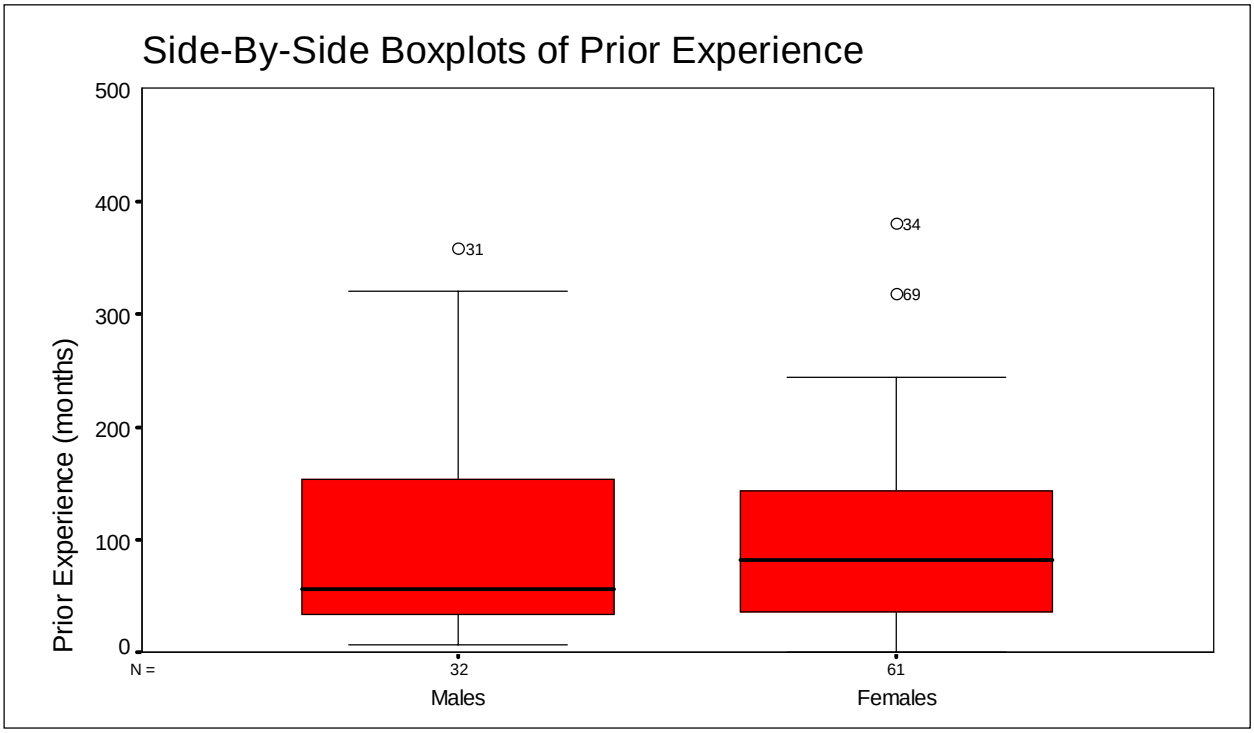
4.4 Did the females tend to be less educated than the males? We will answer the question by producing side-by-side boxplots of number of years of education for males and females.



The males have more years of education, on average, than females. The median number of years of education for males is 15 (upper side of the red box) and the median number of years of education for females is only 12. The boxplot for the females is flat indicating that the interquartile range is zero. The first, the second, and the third quartiles are equal to 12. Most females have 12 years of education. The extreme observations (more than 3*IQR from the end of the box) are marked with asterisks. As the interquartile range is zero, all observations different from 12 are considered extreme observations.

Most of the employees have either 8, 12, or 15 years of education and 10 of the 11 individuals with 8 years of education are females.

4.5 Did the females tend to have less experience prior to employment with the bank than the males? We will answer the question by obtaining side-by-side boxplots of prior experience for males and females.



As you can see, the median number of months of prior experience for females is slightly larger than the corresponding median for males. The spread is similar. The distribution is skewed to the right for the males (look at the position of the median inside the box), and it is approximately symmetric for the females.