1. A)

For each gender the distributions of the three measured variables are normal because:
- The median, mode and mean are almost the same for all the variables.
- The points in the probability plot are very close to the straight (probability) line.
- The box plot of the all the three variable are symmetric.
- Shape of the distribution plot is like a bell (bell curve).

1. B)

Means for Female: Length = 136.04, Height = 52.04, Width = 102.58
Means for Female: Length = 113.37, Height = 40.70, Width = 88.29
From the above values it can be observed that the mean values for female is larger than male, so we can say that females are larger than males.

1. C)

Yes the differences observed in the heights, widths, and length of the male and female turtles can be explained by chance:
- In the Q-Q plot for all the three variables, it can be observed that the value of each variable for female in greater than that of male.
- Also the Pr value is negligibly small in the TTest.

1. D)
- In the scatter correlation plots below, for both males and females, we can observe that the height, width and length are positively and strongly correlated to each other.
- For females the correlation for each pair of variables is around ~0.97, which means they are very strongly correlated.
- While for males the correlation is 0.91 for Height-Width and 0.94 for Height-Length and Length-Width, meaning that Height-Length and Length-Width are more strongly correlated than Height-Width.
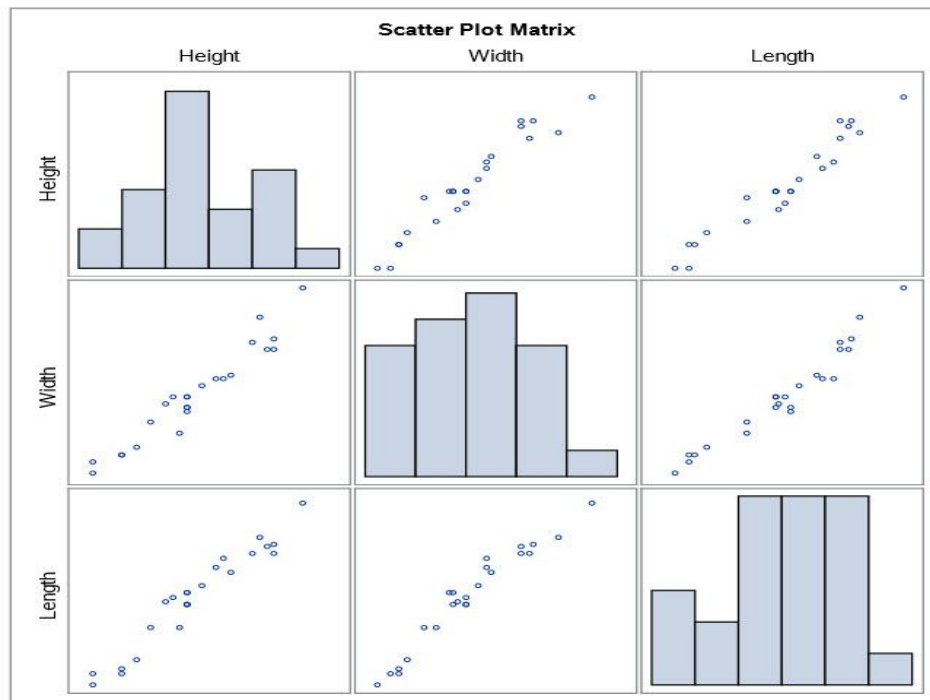
The CORR Procedure

Gender=male

Scatter Plot Matrix
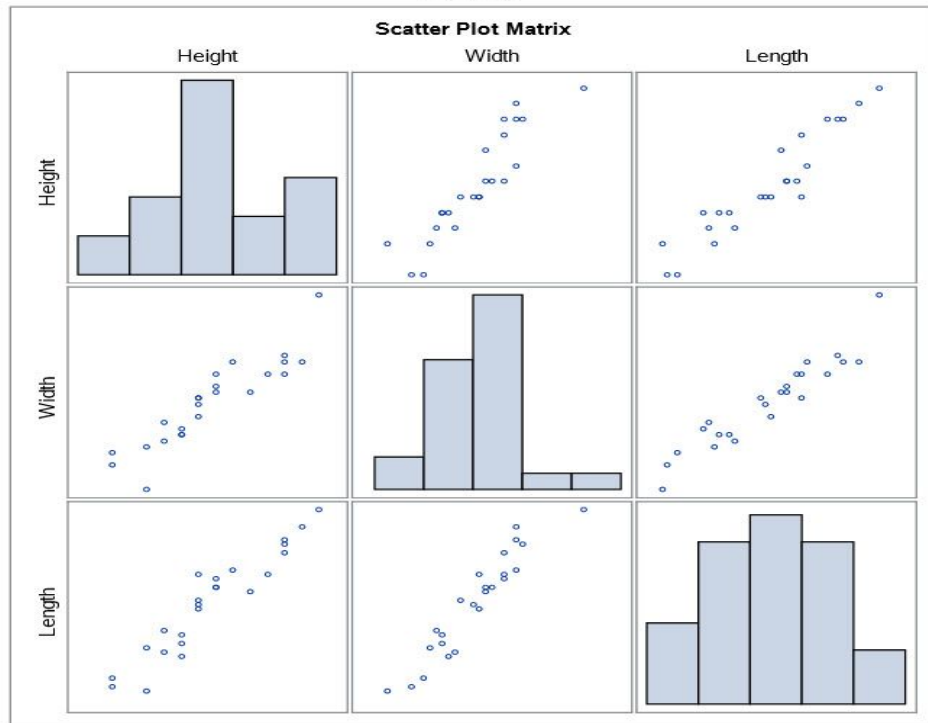
The CORR Procedure

Gender=female

Scatter Plot Matrix

2. A)

- **Midterm**:
  - ○ The numerical values are:
    - ■ Mean: 158.6
    - ■ Median: 164
    - ■ Mode: 171
    - ■ Std Deviation: 18.99
  - ○ Graphical summary:
    - ■ In the distribution plot below we can see that most score are in higher range (greater than 144).
    - ■ There are few outliers in the box plot, in the lower range.
    - ■ In the probability plot, the fit is not good. The data points are not following the straight line, they are curved. We need to fix that.

Distribution and Probability Plot for Quiz
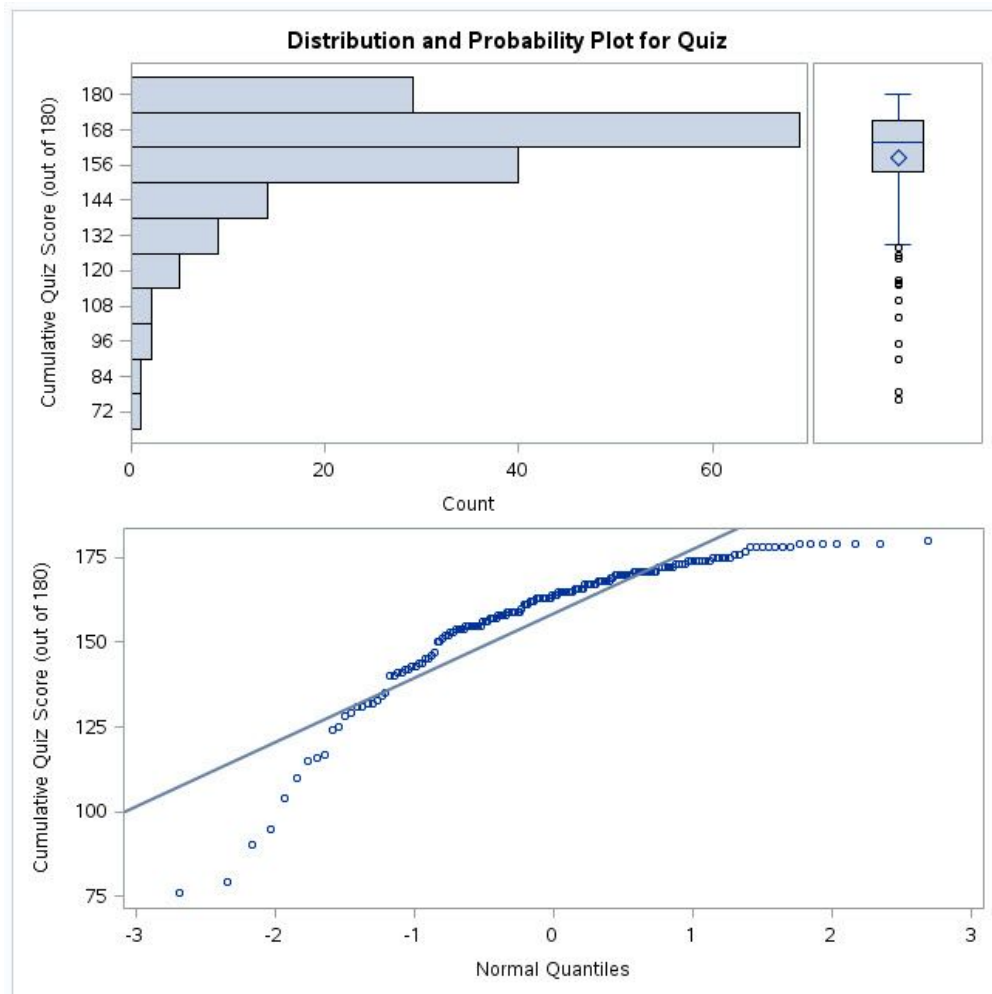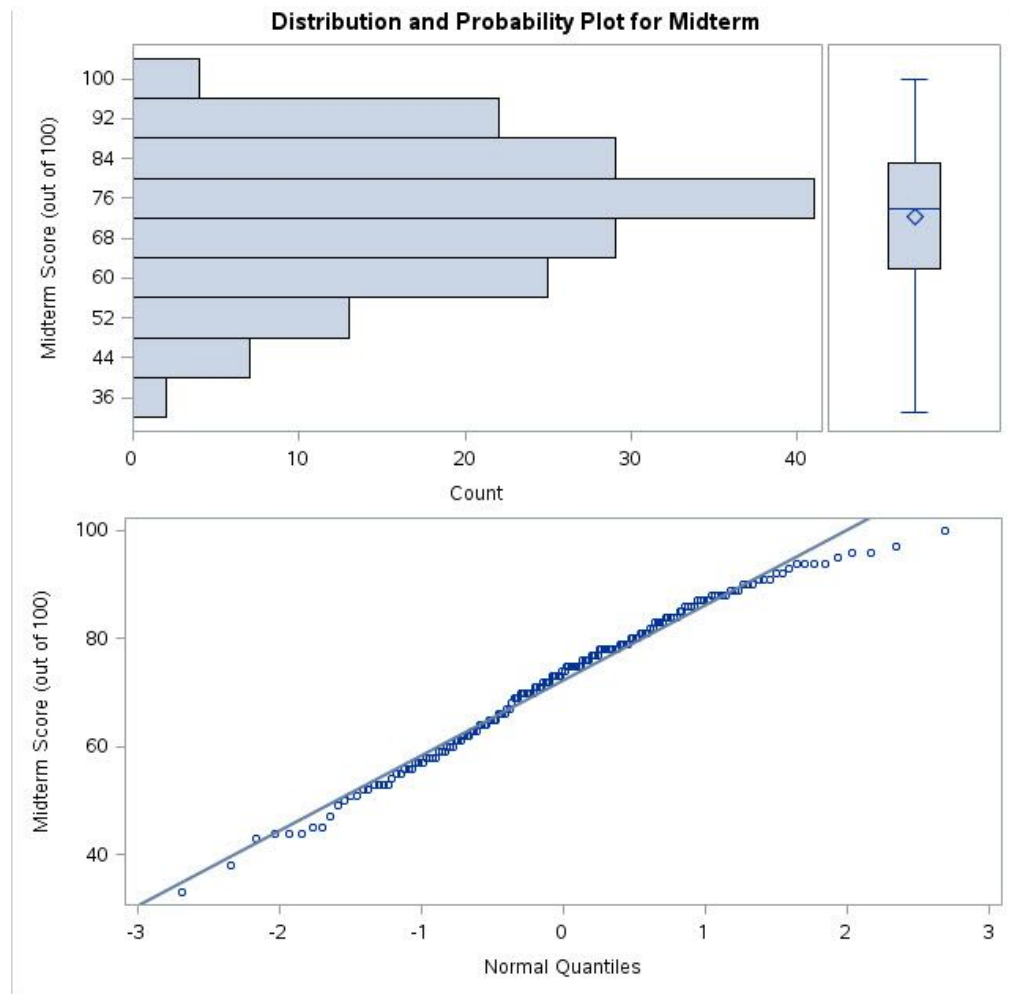
- **Quiz**:
  - The numerical values are:
    - Mean: 72.27
    - Median: 74
    - Mode: 78
    - Std Deviation: 13.9
  - Graphical summary:
    - In the distribution plot below is almost normal, as seen from the bell shaped bar graph.
    - Box plot is also normal with no outliers.
    - Probability plot is normal as well indicated by the data points being close to the normal straight line.


Distribution and Probability Plot for Midterm

- **Final**:
  - The numerical values are:
    - Mean: 137.15
    - Median: 140.5
    - Mode: 122
    - Std Deviation: 34.46
  - Graphical summary:
    - In the distribution plot below is almost normal, as seen from the bell shaped bar graph.
    - Box plot is also normal with a very few outliers.
    - Probability plot is normal as well indicated by the data points being close to the normal straight line.

Distribution and Probability Plot for Final
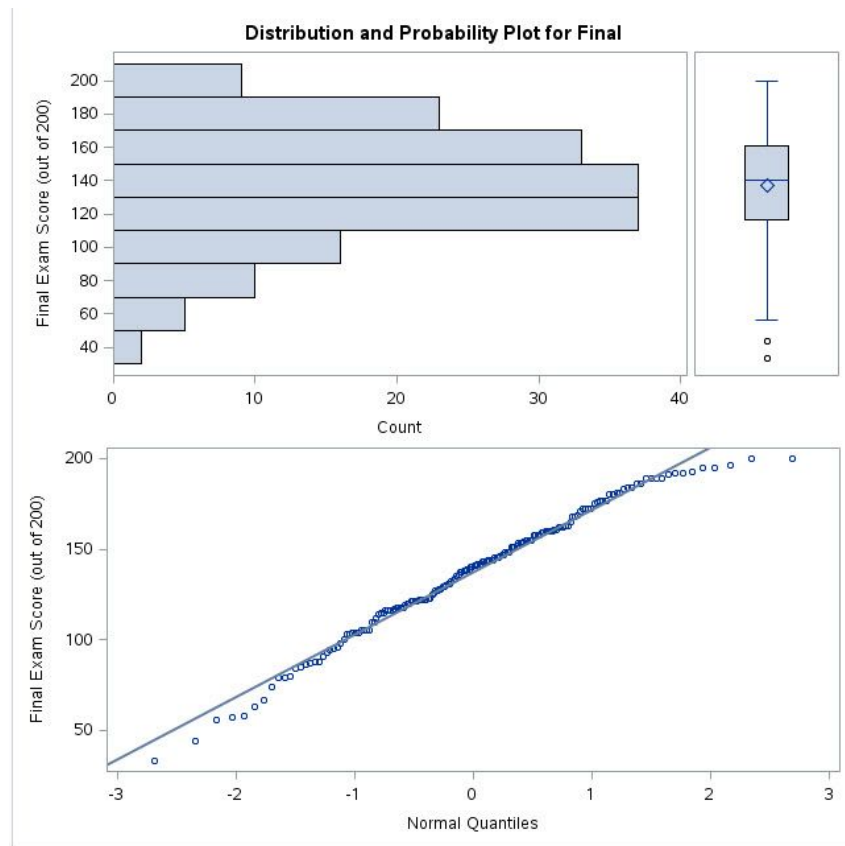
2. B)
I have applied following transformations to the variables to make the data more normal.

- srtFinal = -1*sqrt(200-Final);

- srtMidterm = -1*sqrt(100-Midterm);

- crtQuiz=-1*(180-Quiz)**0.33333;

2. C)
- After applying the above transformations, the data has become more normal, indicated by bell shaped bar graph for distribution.

  The table below shows the correlation matrix for the transformed variables.

Pearson Correlation Coefficients, N = 172 Prob > |r| under H0: Rho=0			
	srtFinal	srtMidterm	crtQuiz
srtFinal Square Root of Final Exam Score	1.00000	0.44337 <.0001	0.63705 <.0001
srtMidterm Square Root of Midterm Score	0.44337 <.0001	1.00000	0.24802 0.0010
crtQuiz Cube root of Quiz Score	0.63705 <.0001	0.24802 0.0010	1.00000

- The pairwise correlation among the transformed variables is on the lower side.
- srtFinal and crtQuiz has highest correlation of 0.63, meaning that quizzes have good impact on final exam scores.
- srtMidterm and crtQuiz has lower correlation, meaning that quizes do not have much impact on midterm scores.
- The low correlations can be visualized from the scatter correlation plot below, in which the dots are scattered all over.
- Though the distributions are normal as seen from the bell shaped bar graphs.

Scatter Plot Matrix
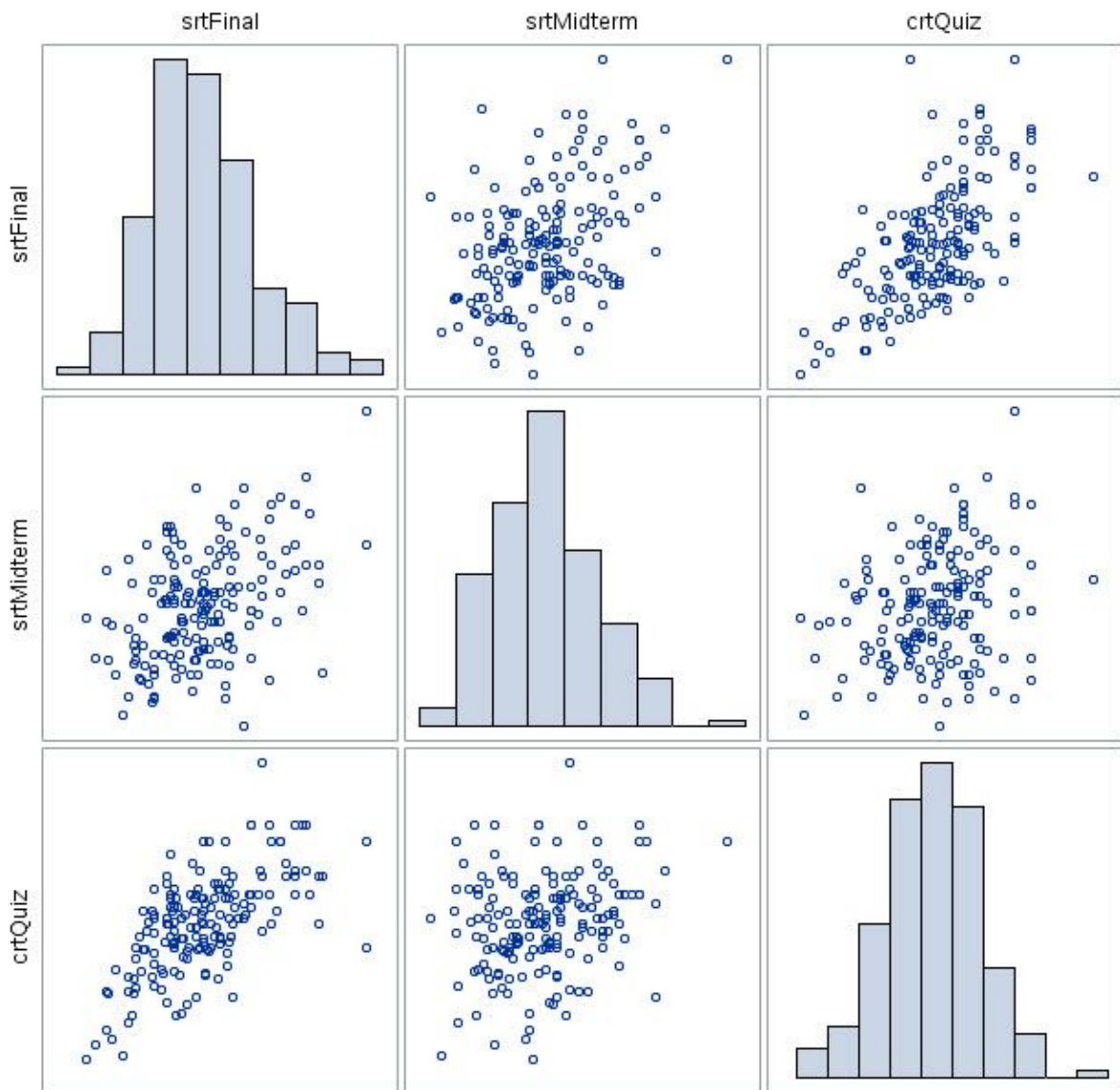
2. D) From the box plots below, we can see that there are subtle differences between the recitation sections.

Distribution of srtFinal by BY Group

Distribution of srtMidterm by BY Group

Distribution of crtQuiz by BY Group
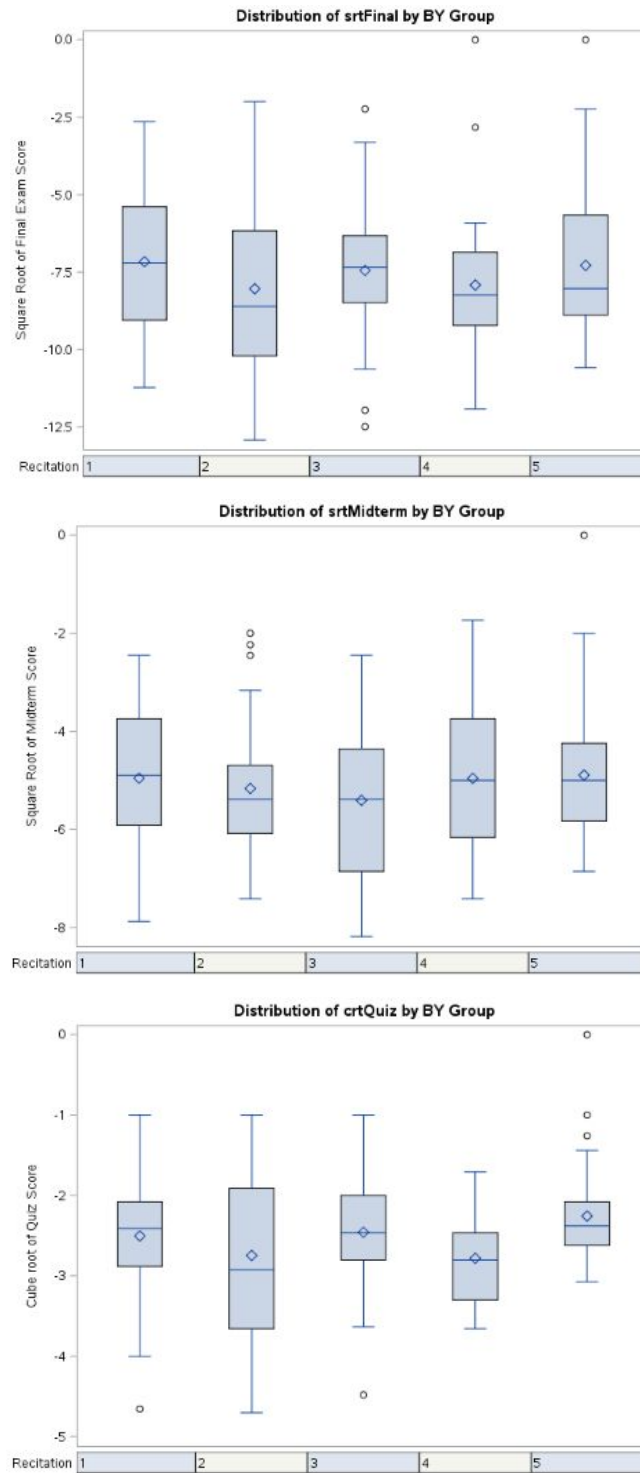
- The box plots for each recitation have different means and medians.
- Box plot for recitation 2 for srtMidterm and recitation 3 for srtFinal have few outliers.
- For recitation 3, mean and median are same for each variable.