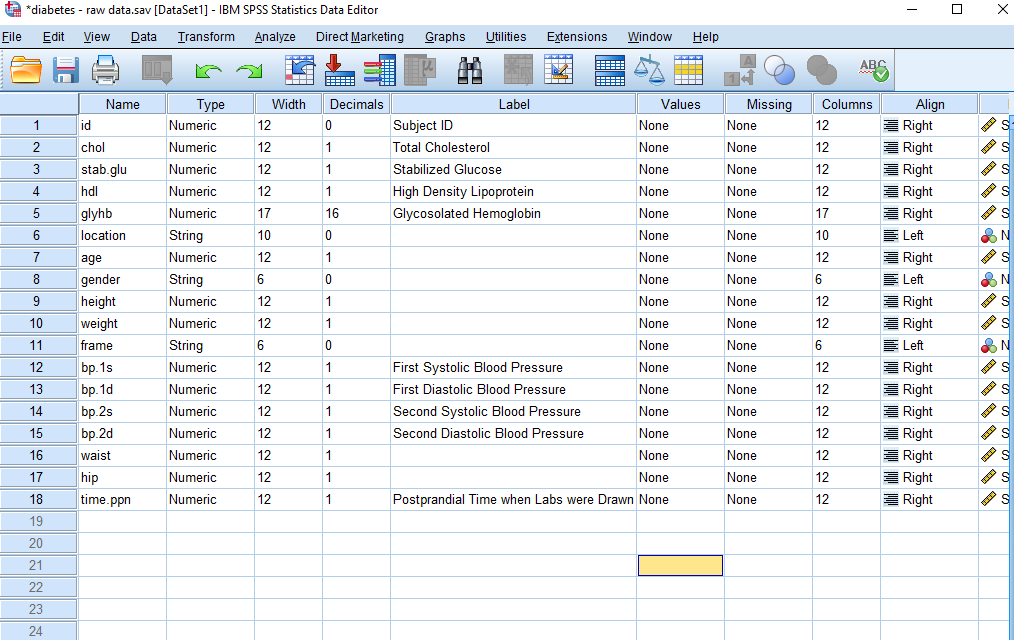
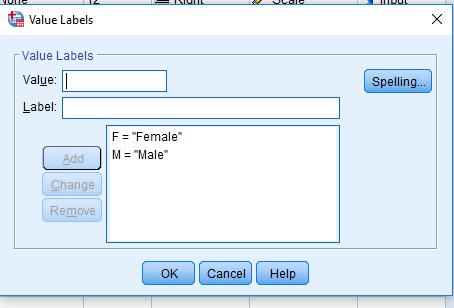
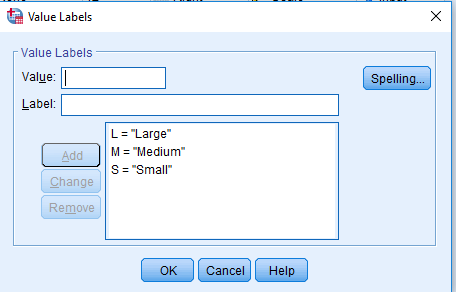
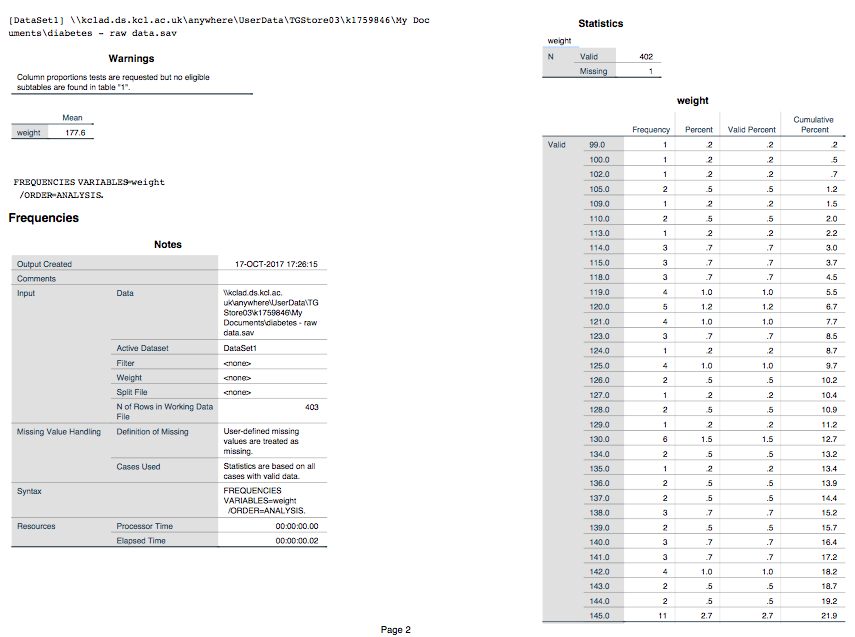
**Statistics Assignment 1 k1759846**

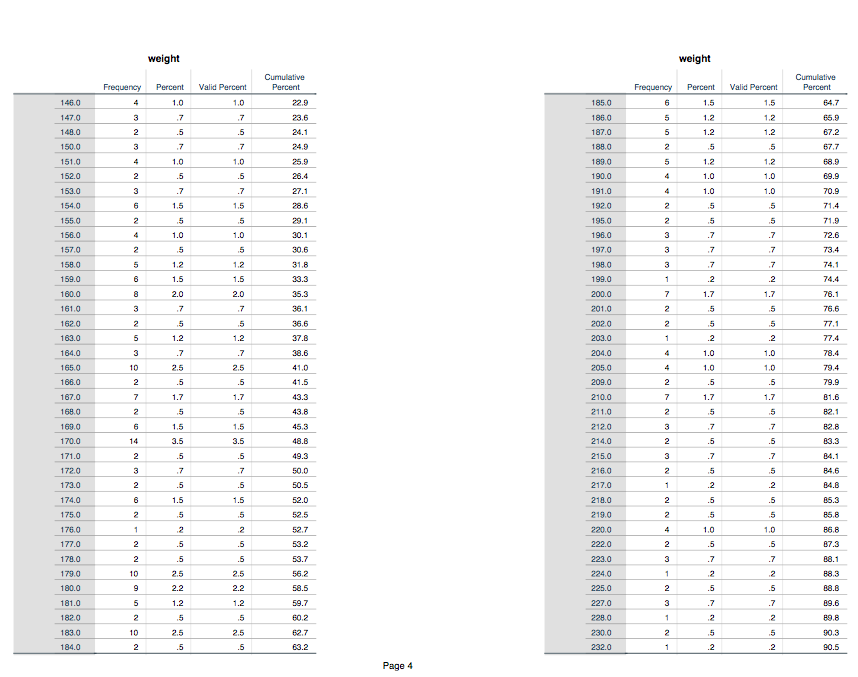
1. I added the labels to the variables as shown. 
2. I added value labels to gender and frame:

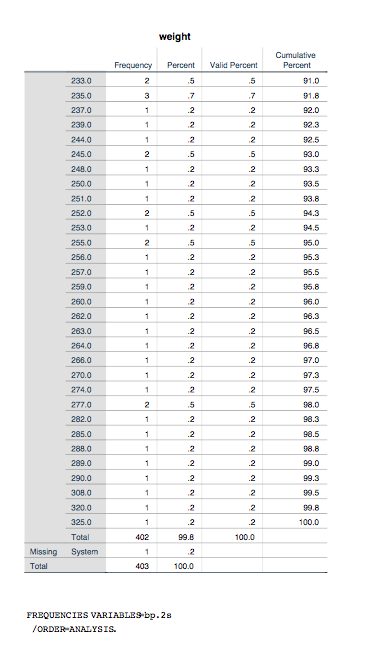




3. 100-76.6 = 23.4% are over 200Lbs.



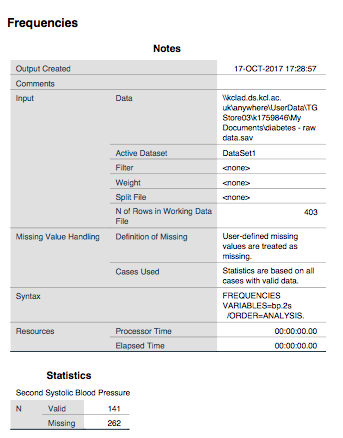


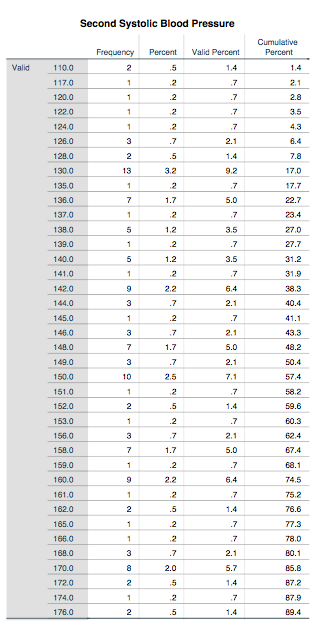


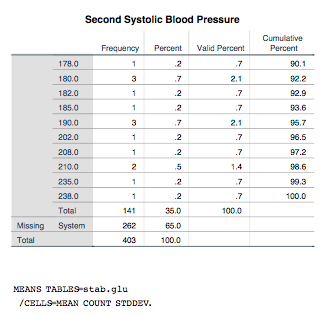
4. 141 people had their blood pressure measurement taken twice; this is the “valid” column; the “missing” refers to the fact that many people do not have data for bp.s2.

This refers to 141/262=53.82% of all people did not have their blood pressure measurement repeated.

In this case, the “percent” column represents the percentage of all of the subjects, even the ones with missing data points for this variable; whereas the “valid percent” represents the percentage of all of the subjects which did have their blood pressure measurement repeated.



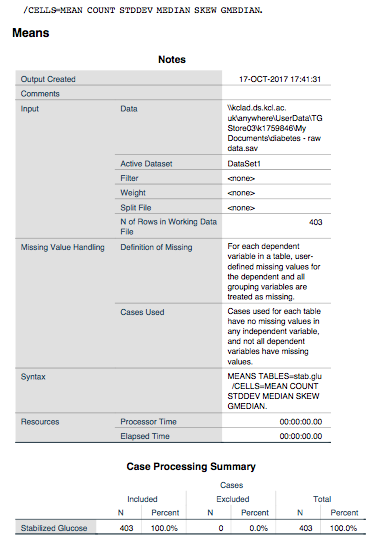


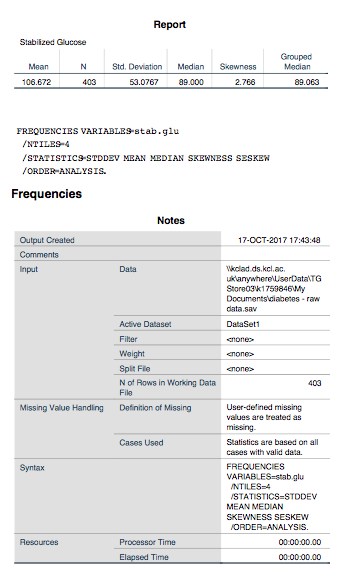


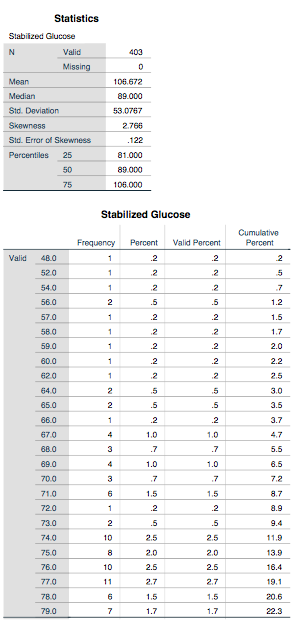
5. The mean for stabilized glucose was 106.672, the standard deviation was 53.0767, the median was 89.000, the lower quartile was 81.000, and the upper quartile was 106.000.

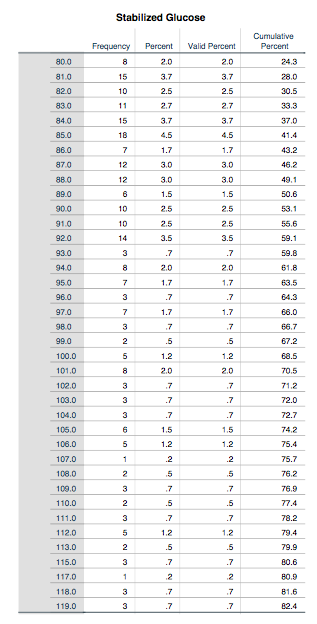
The skewness was 2.766.

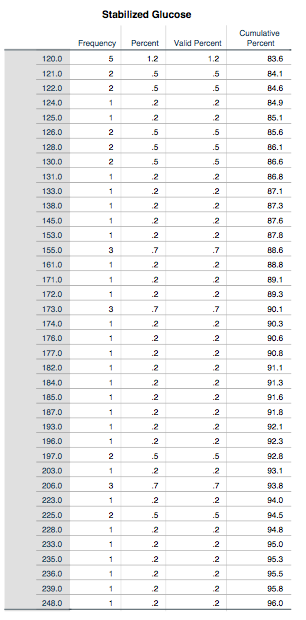
The mean is higher than the upper quartile since although most of the subjects had a stabilized Glucose level of around, or slightly less than 100.00, making the median 89; there are a small number of subjects with extremely high Stabilized Glucose (~400), which skew the mean heavily.

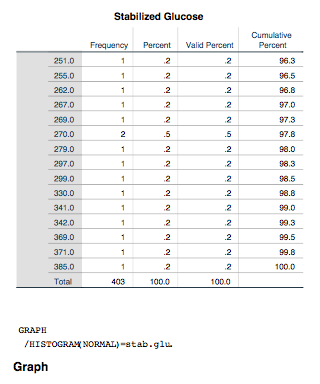


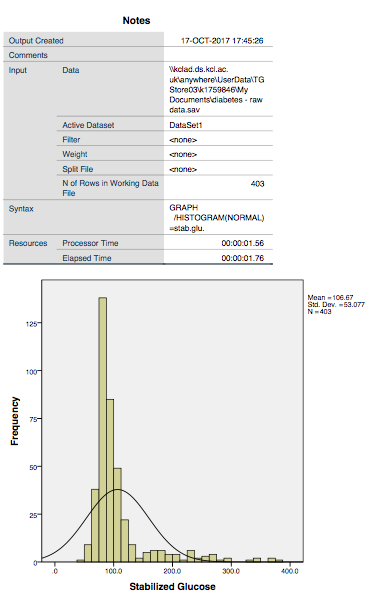


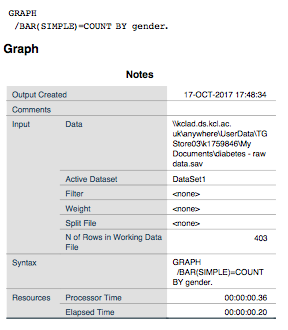


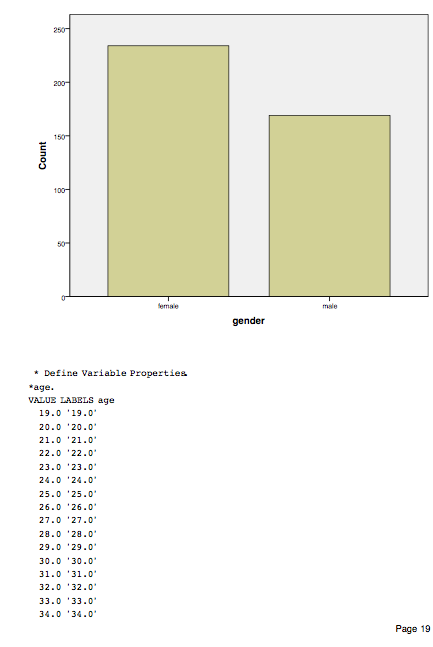




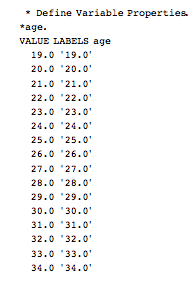


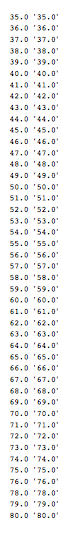


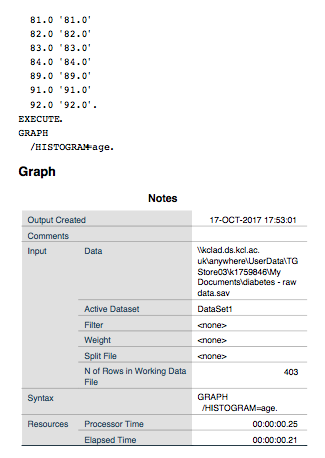
6. 

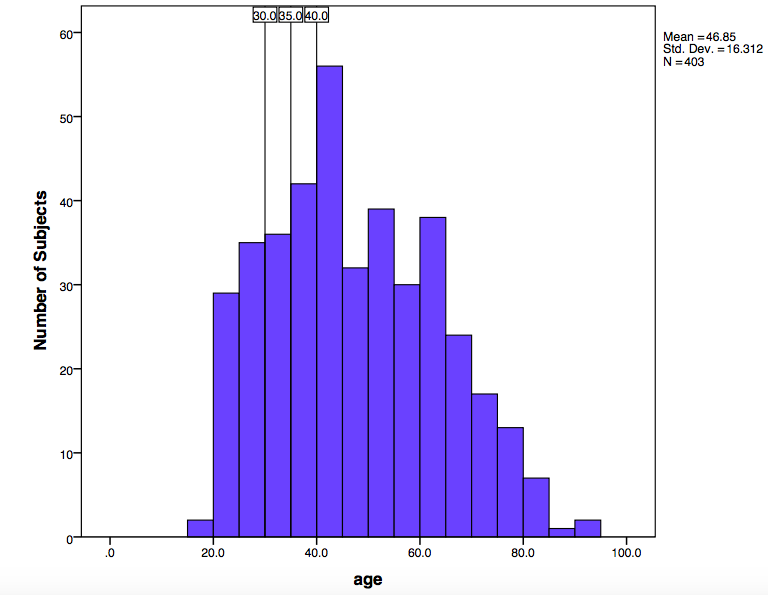


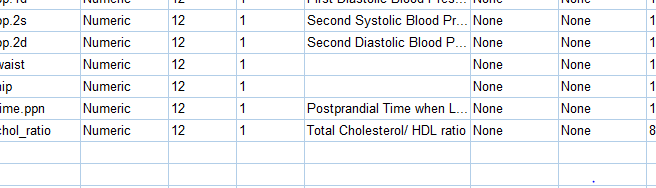
7. From this histogram for the age of subjects, I would summarize that the ages have an almost normal distribution, with a steeper gradient on the younger side. The peak is in the 40.0-45.0 bin. The youngest subject is aged from 15.0-20.0, and the oldest subject is aged between 90.0-95.0 years old. The mean age is 46.85.

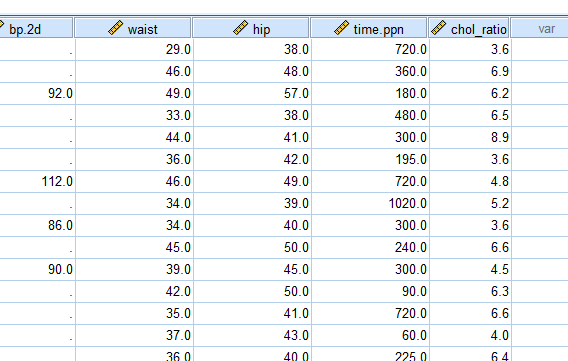


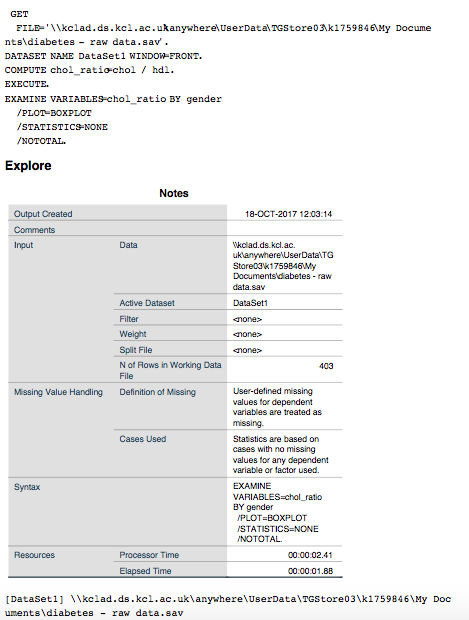


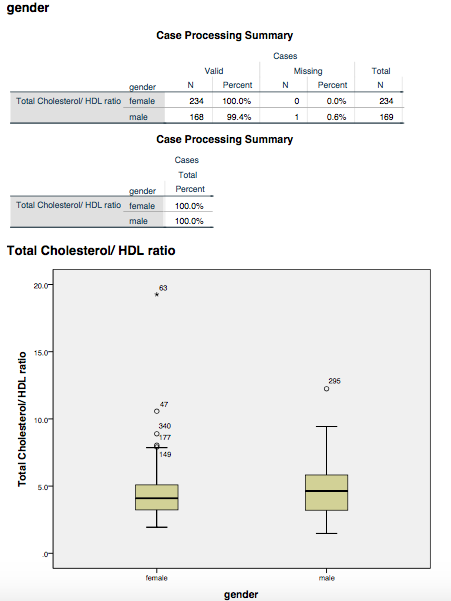




8. 

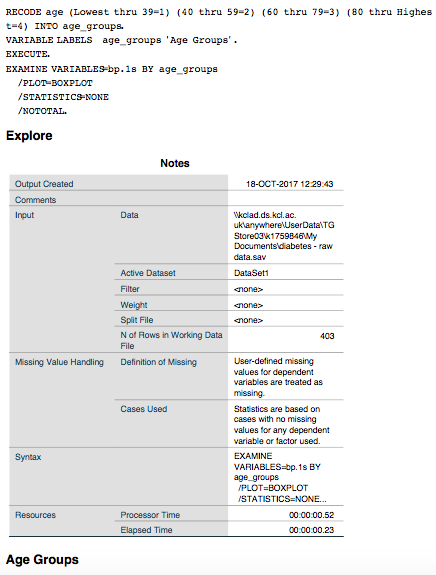


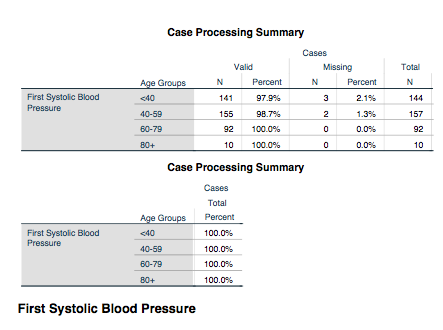


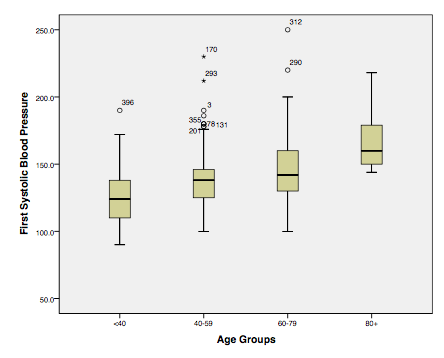
****

From this boxplot I can see that there is a larger range of the Total CHolesterol/HDL ratio in men, which reaches higher than the range for women. The ratio is however similar for women and men, with the average for men only very slightly higher than for women.

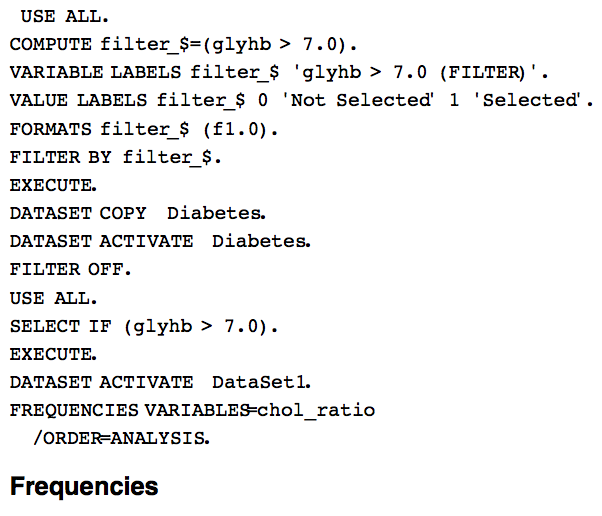
9. From this box plot of First Systolic Blood Pressure split into age groups, you can see that as age increases, so does the blood pressure. The ranges of the blood pressures for each of the ages is similar however.

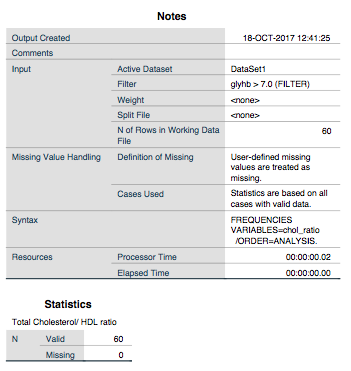


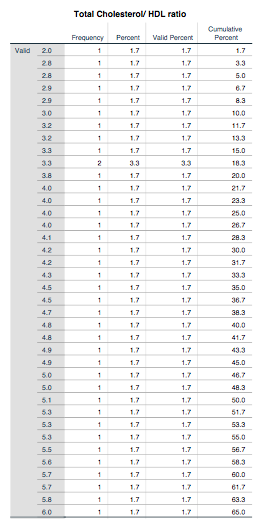


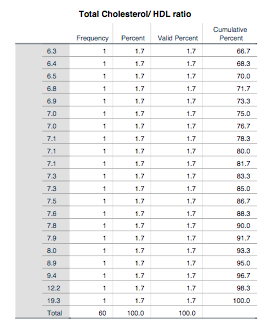


10. 50% of subjects with diabetes had Cholesterol/HDL ratio >5





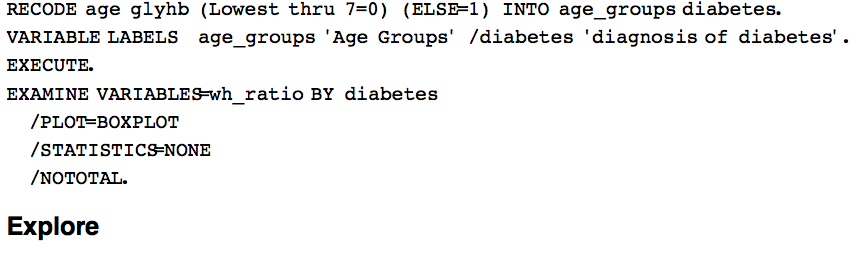


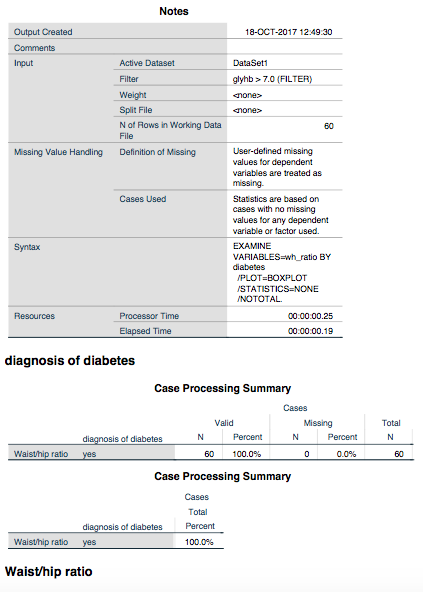


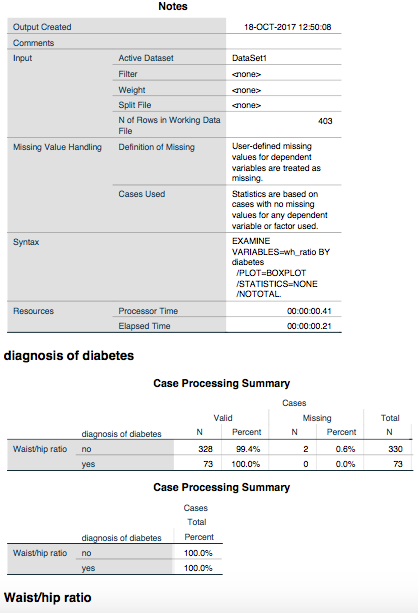
11. 29.png

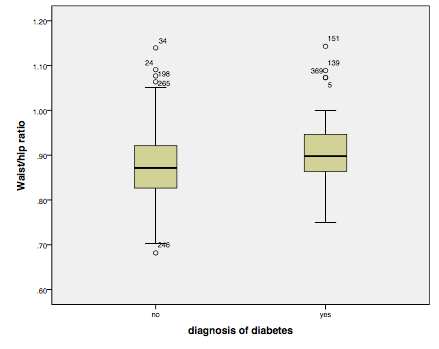
12. 6.PNG

The box plot showed that those with diabetes had a slightly higher waist to hip ratio, however the range of the waist/hip ratio for people without diabetes encapsulated the range for people with diabetes.





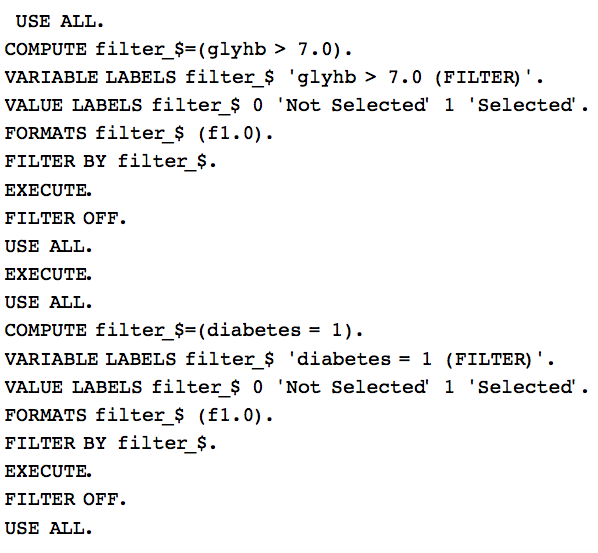


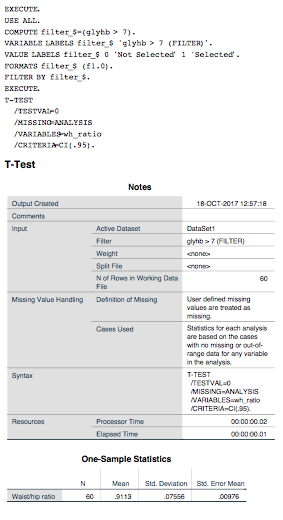


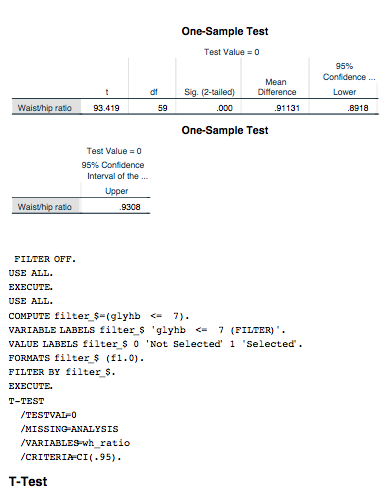
13. The 95% confidence interval for the mean on the waist/hip ratio for people WITH diabetes was 0.8918-0.9308, with the mean being 0.9113. The standard deviation was 0.07556, and the standard error mean was 0.00976. There were 59 degrees of freedom, and 60 people in the sample.

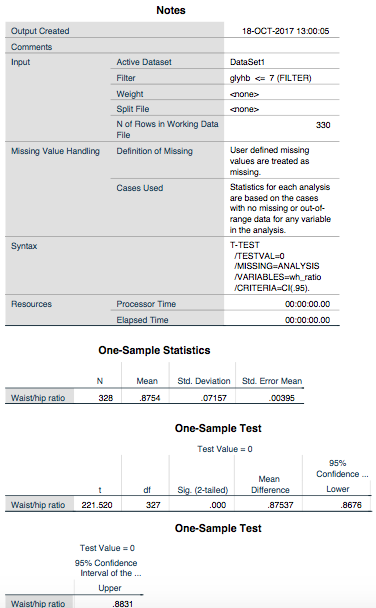
The 95% confidence interval for the mean waist/hip ratio for people WITHOUT diabetes was 0.8676-0.8831, with the mean value being 0.8754. The standard deviation was 0.07157, standard error mean was 0.00395. There were 327 degrees of freedom, and 328 subjects in the sample.

From these results we can conclude that people with diabetes have a higher waist/hip ratio than those without diabetes.









14. From this box plot we can see that the majority of males have a larger range of, and a larger waist/hip ratio than females. We can also see that subjects living in BuckinghLo also have a higher waist/hip ratio than subjects living in Louisa, for both males and females.

