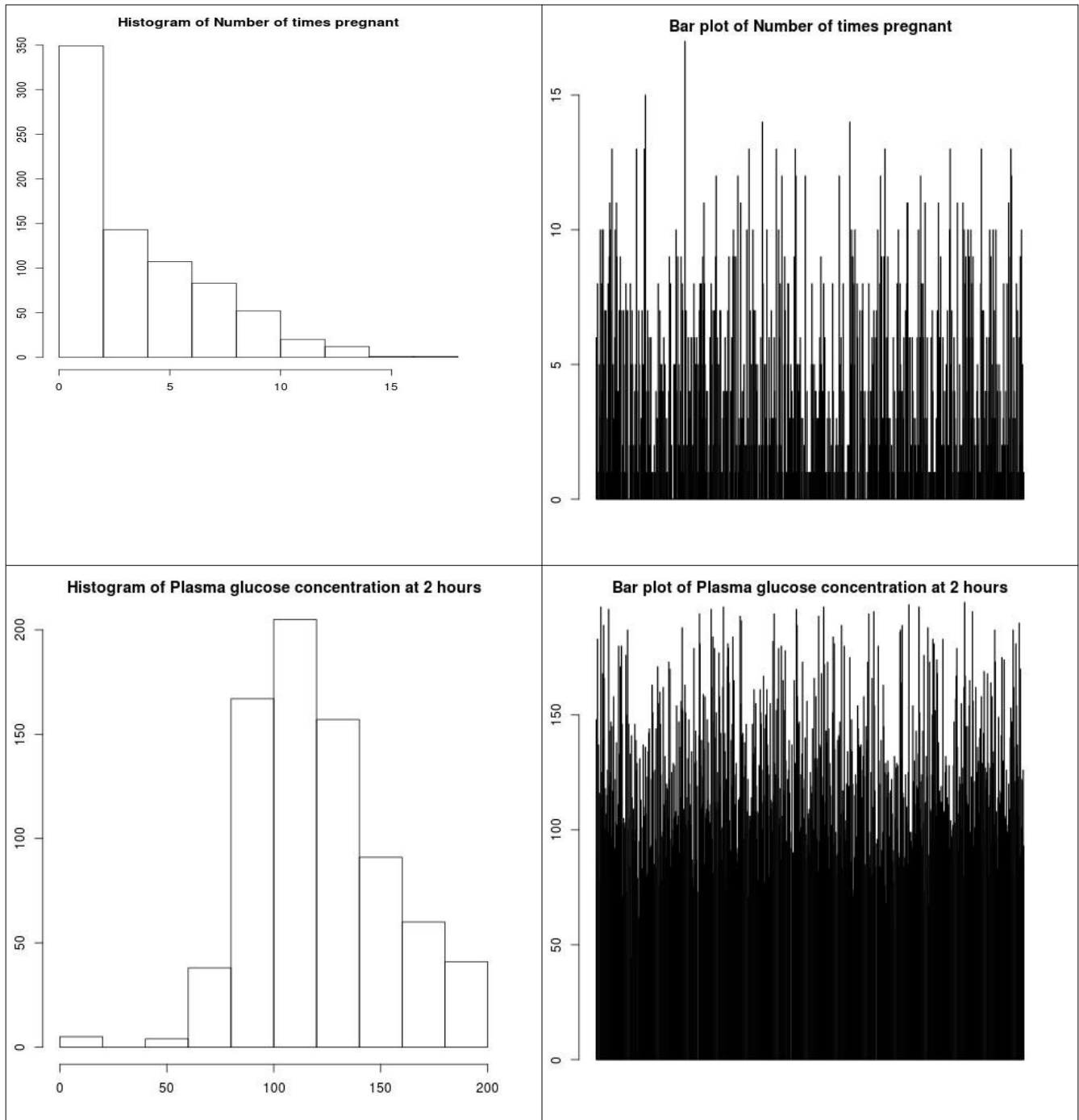


Report.pdf

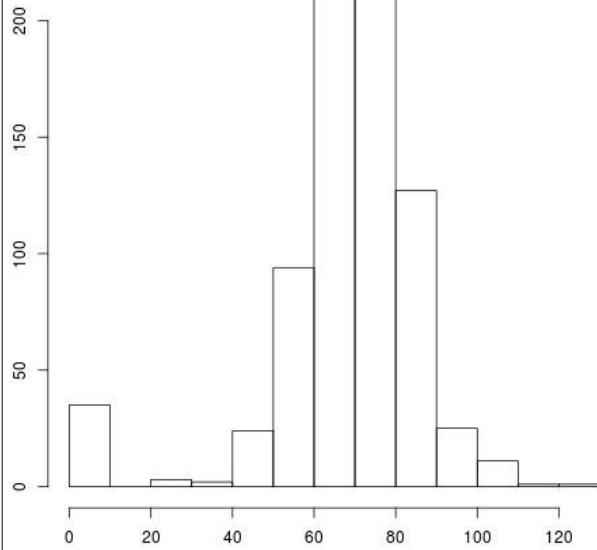
Shane Zabel

exploratory_analysis.R

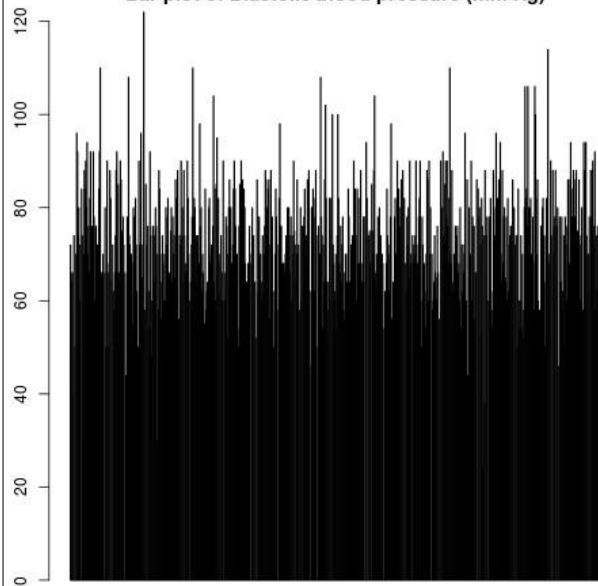
1. Create the following plots: histogram, and barplot.



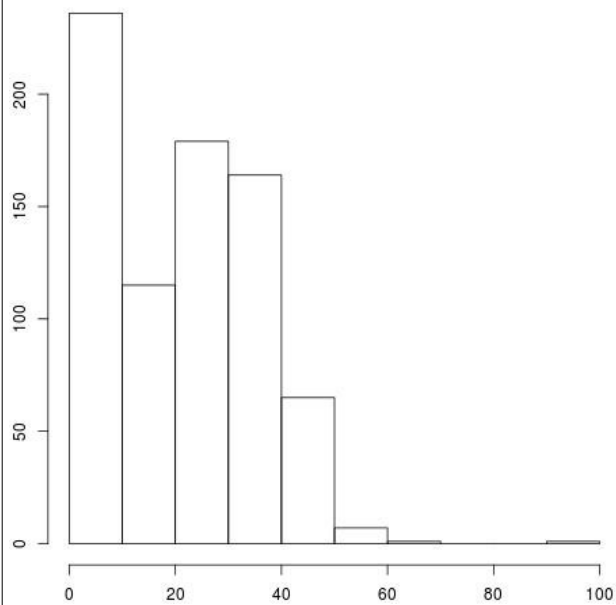
Histogram of Diastolic blood pressure (mm Hg)



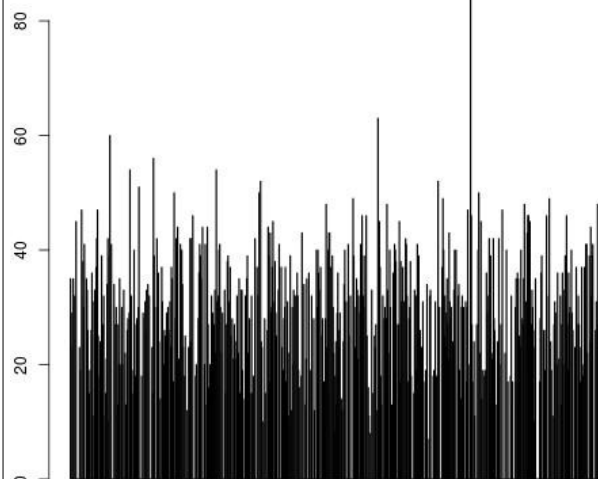
Bar plot of Diastolic blood pressure (mm Hg)



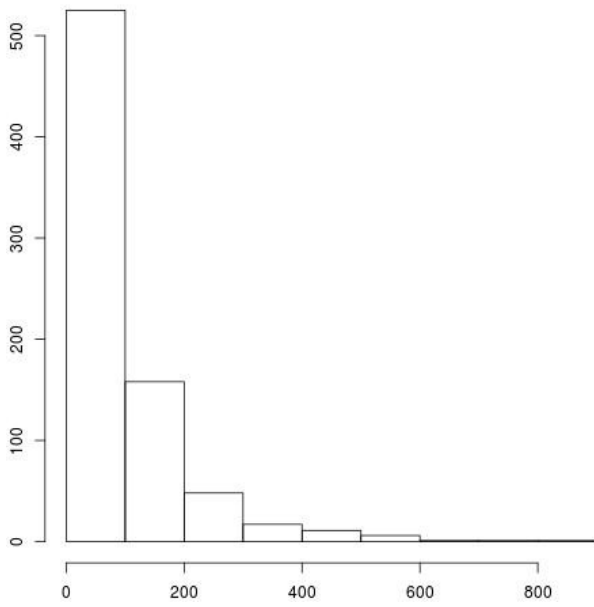
Histogram of Triceps skin fold thickness (mm)



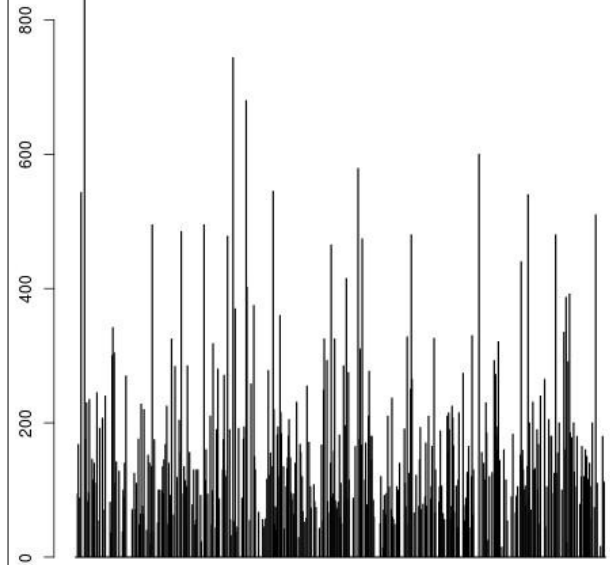
Bar plot of Triceps skin fold thickness (mm)



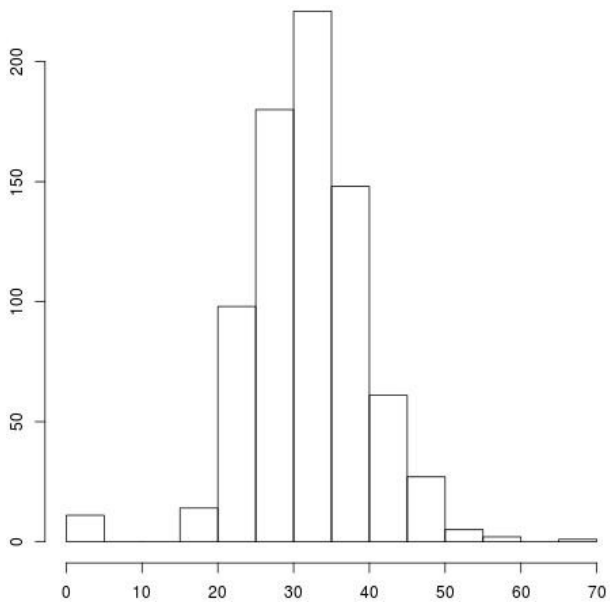
Histogram of 2-Hour serum insulin (mu U/ml)



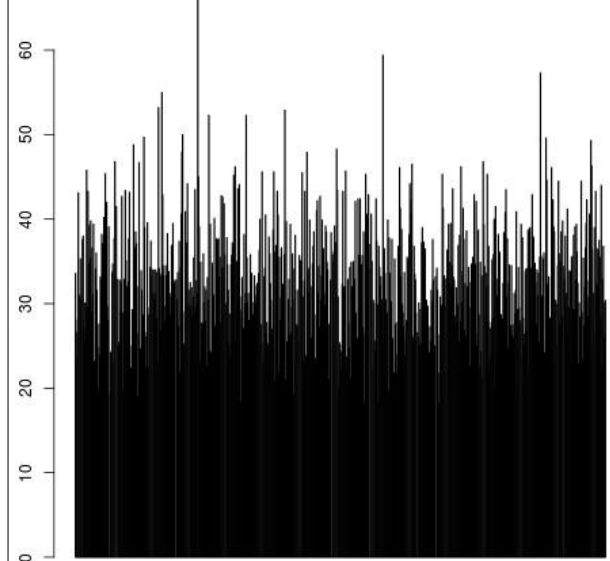
Bar plot of 2-Hour serum insulin (mu U/ml)



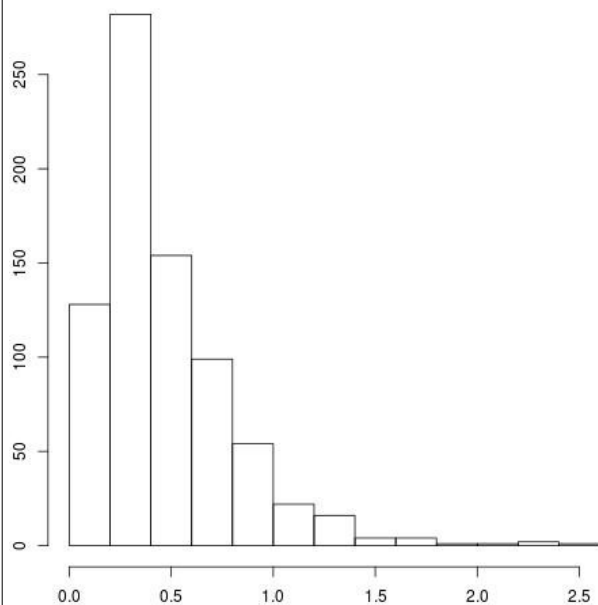
Histogram of Body mass index (weight in kg/(height in m)²)



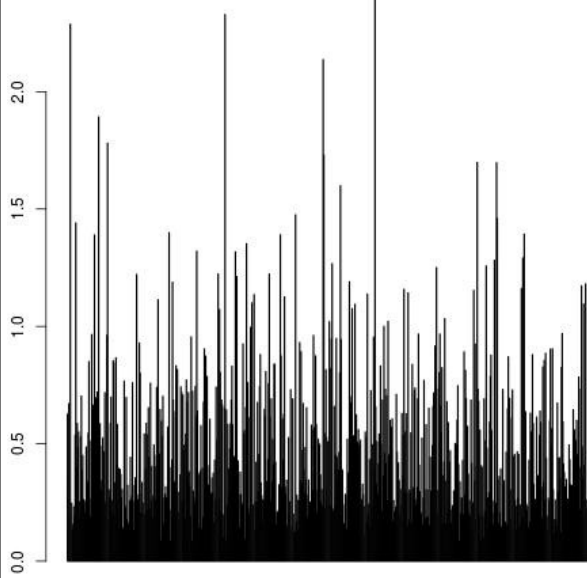
Bar plot of Body mass index (weight in kg/(height in m)²)



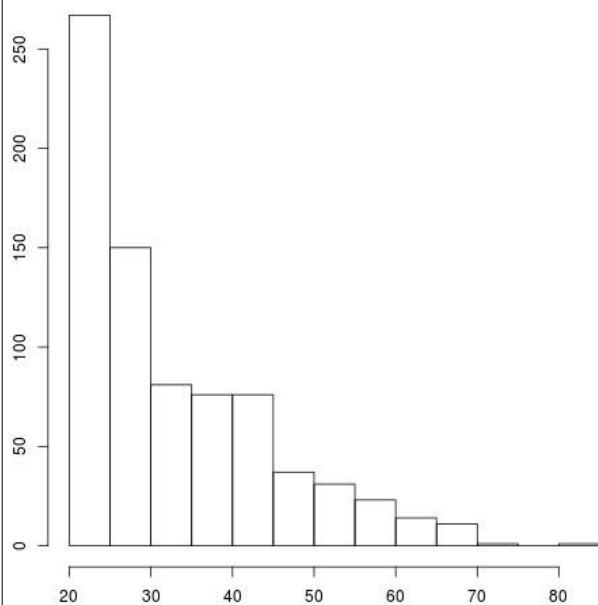
Histogram of Diabetes pedigree function



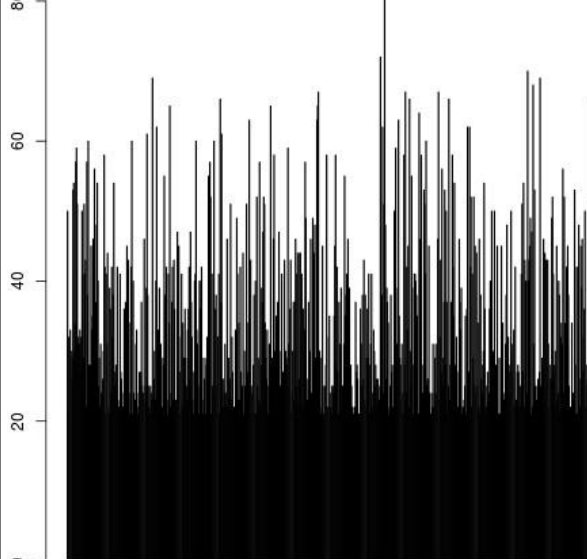
Bar plot of Diabetes pedigree function

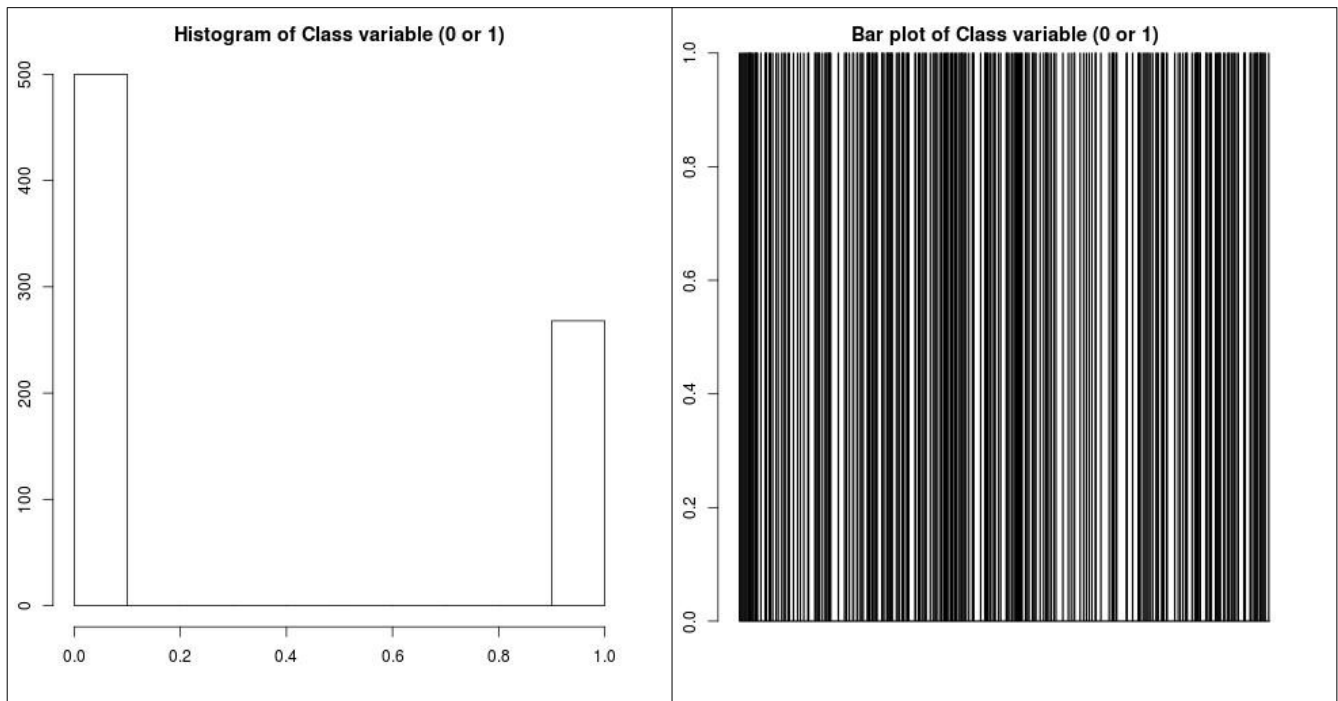


Histogram of Age (years)



Bar plot of Age (years)





1. Write a short note on the distribution of the variables that you observe from the plots. Are they normally distributed?

Based off the histogram plots of the variables it appears some are close to normally distributed and others are not.

The "Body mass index", "Triceps skin fold thickness", and "Diastolic blood pressure" variables all have high values at the extreme left (low) tail. The other variables appear upon inspection to be close to normally distributed.

2. Find the correlation between each of the attributes and the class variable.

Attribute	Correlation with Class Variable
1	0.22189815
2	0.46658140
3	0.06506836
4	0.07475223
5	0.13054795
6	0.29269466
7	0.17384407
8	0.23835598

2. Which attributes seem to have a strong correlation with the output (class) variable?

#1 Attribute 2: "Plasma Glucose Concentration" with a correlation value of 0.467
 #2 Attribute 6: "Body Mass Index" with a correlation value of 0.293

3. *Compute the correlation between all pairs of the 8 attributes. Which two attributes have the highest mutual correlation?*

Attributes 1: "Number of Times Pregnant" and 8 : "Age" have the highest mutual correlation.

Correlation Value	Attribute 1	Attribute 2
0.5443412	1.0000000	8.0000000