Data Report on Human Data | Final Bioinformatics Bootcamp Project

Smita Sahay

06/30/2022 - 07/04/2022

### Introduction

This is a dataset containing various information about 500 different human subjects.   
These variables include age, gender, height, weight, body mass index (BMI), systolic blood pressure (SBP), and diastolic blood pressure (DBP).   
In this report, we will analyze these variables and search for meaningful relationships in the data.

### Descriptive Statistics of Major Variables

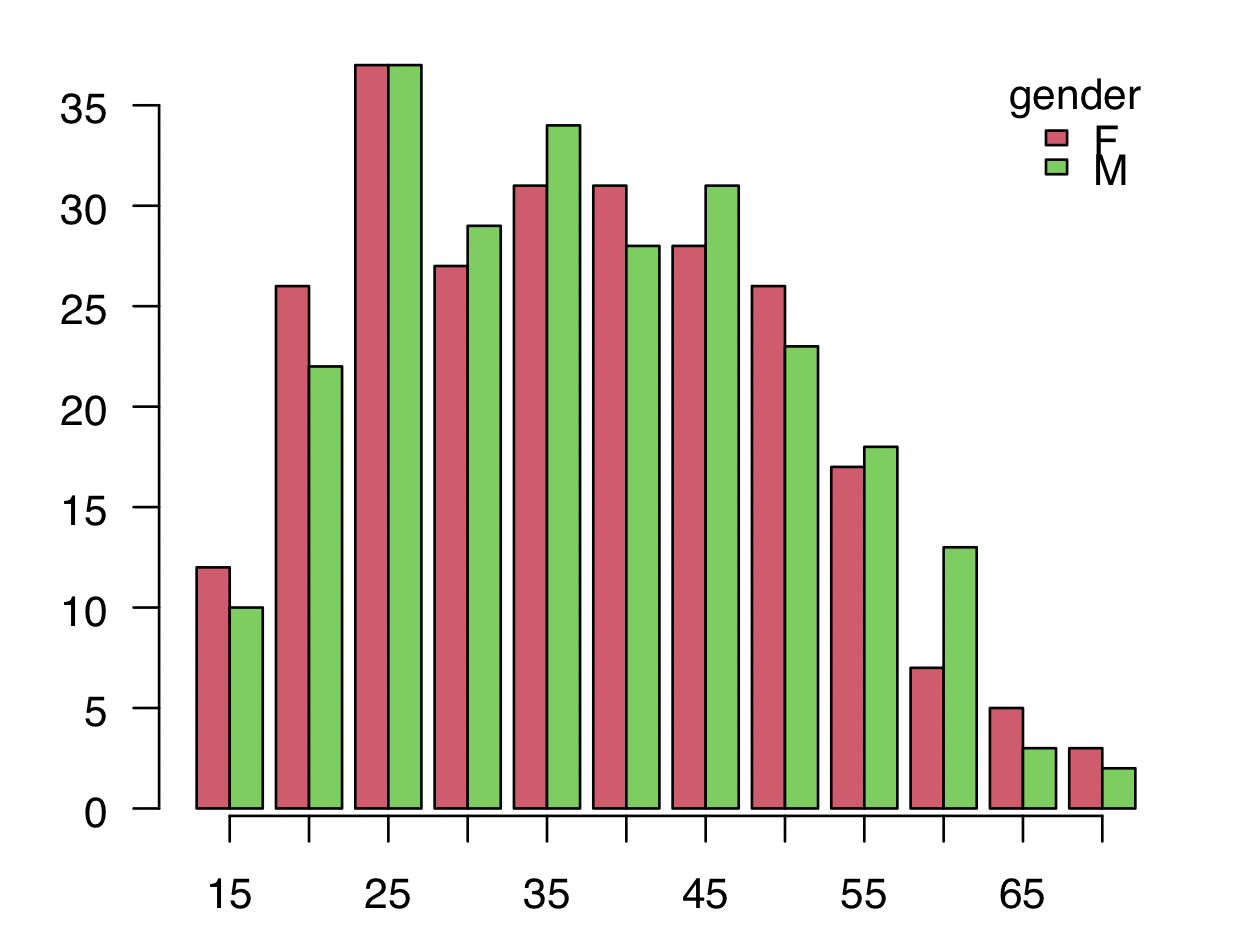


Figure 1: Relationship between Gender and Age of the 500 Subjects in the Dataset.

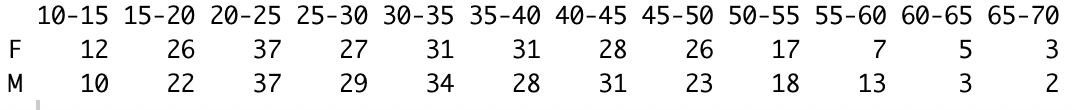


Table 1: Frequency Table with Gender and Age of the 500 Subjects in the Dataset.

From Figure 1 and Table 1, we can see that of the 500 subjects in this dataset, the greatest frequency in terms of age of both males and females is in the 20-25 age group. The least frequency in terms of age of both males and females is in the 65-70 age group.

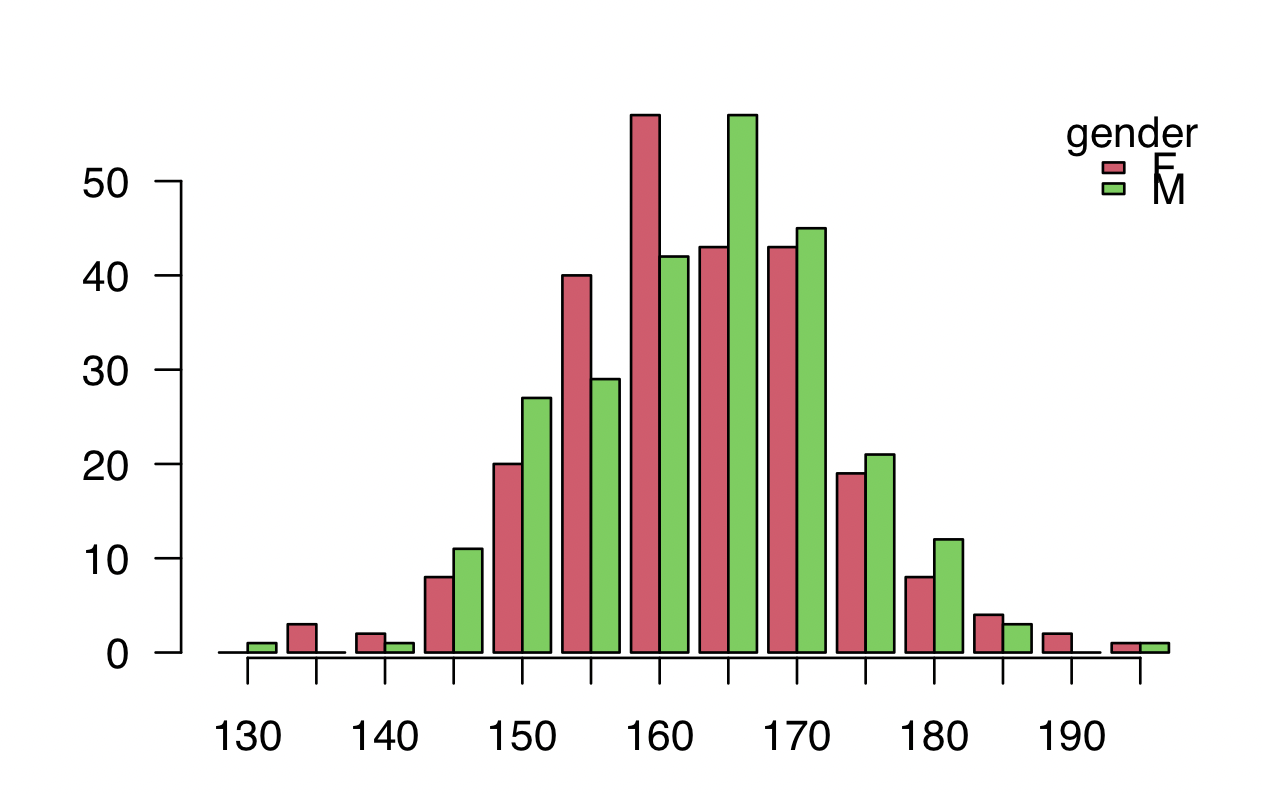


Figure 2: Relationship between Gender and Height of the 500 Subjects in the Dataset.

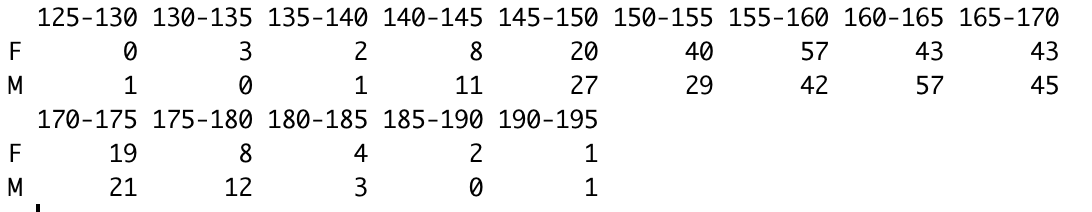


Table 2: Frequency Table with Gender and Height of the 500 Subjects in the Dataset.

From Figure 2 and Table 2, we can see that of the 500 subjects in this dataset, the greatest frequency in terms of height of females is in the 155-160 centimeter group. The greatest frequency in terms of height of males is in the 160-165 centimeter group.The least frequency in terms of height of females is in the 125-130 centimeter group. The least frequency in terms of height of males is in the 185-190 centimeter group.

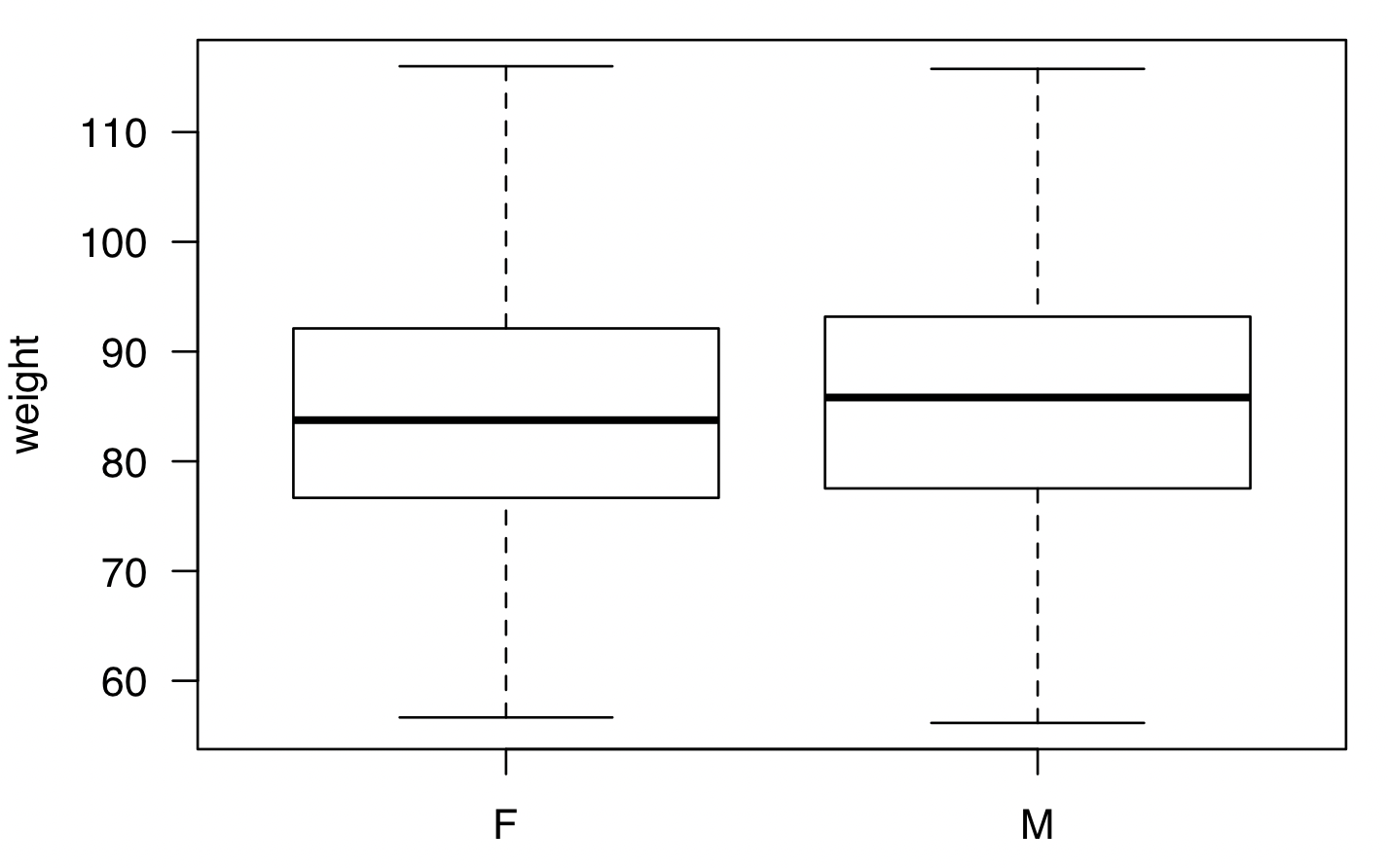


Figure 3: Relationship between Gender and Weight of the 500 Subjects in the Dataset.

From Figure 3, we can see that when comparing the weights of the 500 subjects in this dataset, they are fairly similar between males and females with the average weight being ~85 lbs for females and ~86 lbs for males.

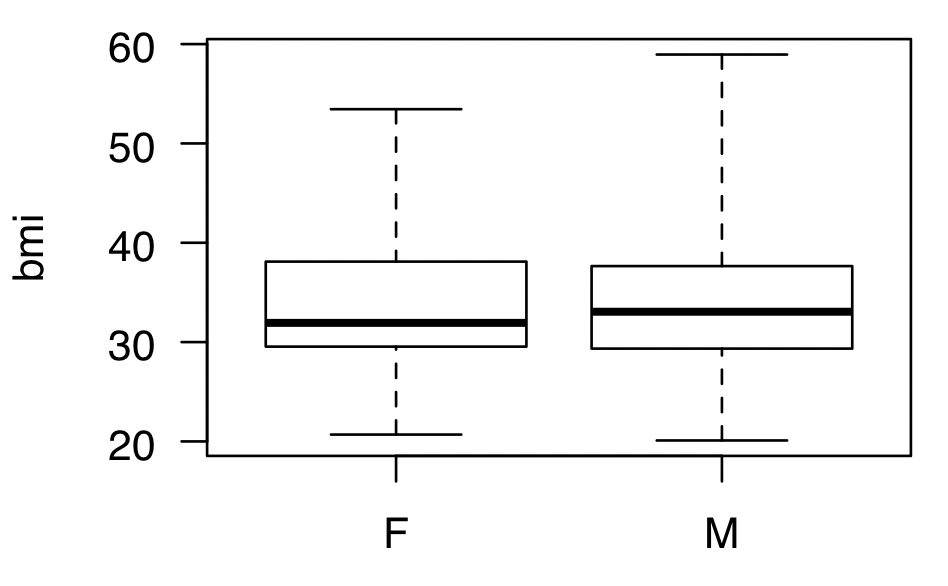


Figure 4: Relationship between Gender and BMI of the 500 Subjects in the Dataset.

From Figure 4, we can see that when comparing the BMIs of the 500 subjects in this dataset, they are fairly similar between males and females with the average BMI being ~31 for females and ~32 for males.

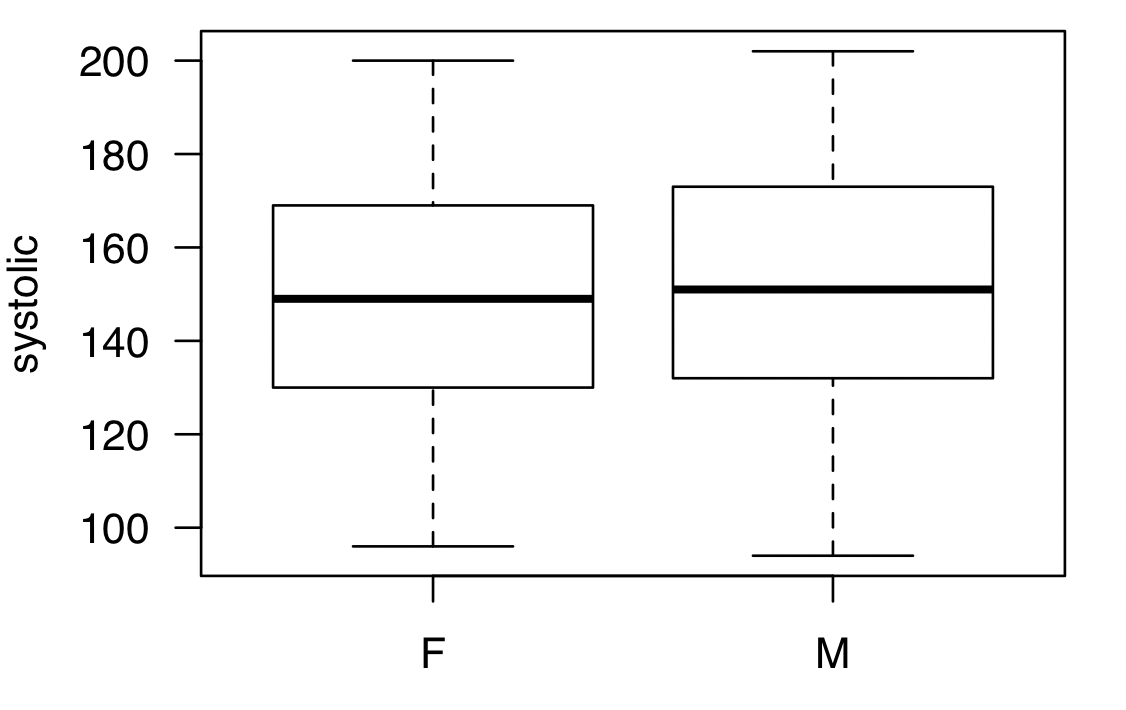


Figure 5: Relationship between Gender and SBP of the 500 Subjects in the Dataset.

From Figure 5, we can see that when comparing the SBPs of the 500 subjects in this dataset, they are fairly similar between males and females with the average SBP being ~150 for females and ~151 for males.

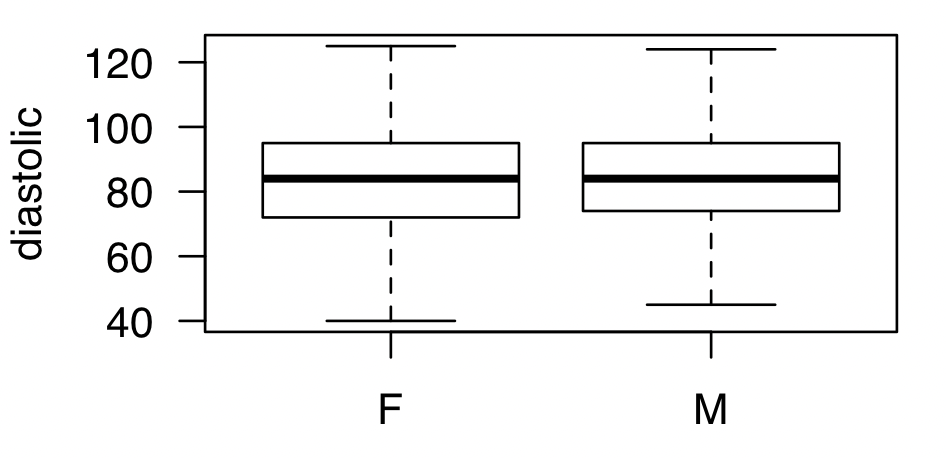


Figure 6: Relationship between Gender and DBP of the 500 Subjects in the Dataset.

From Figure 6, we can see that when comparing the DBPs of the 500 subjects in this dataset, they are fairly similar between males and females with the average DBP being ~83 for both females and males.

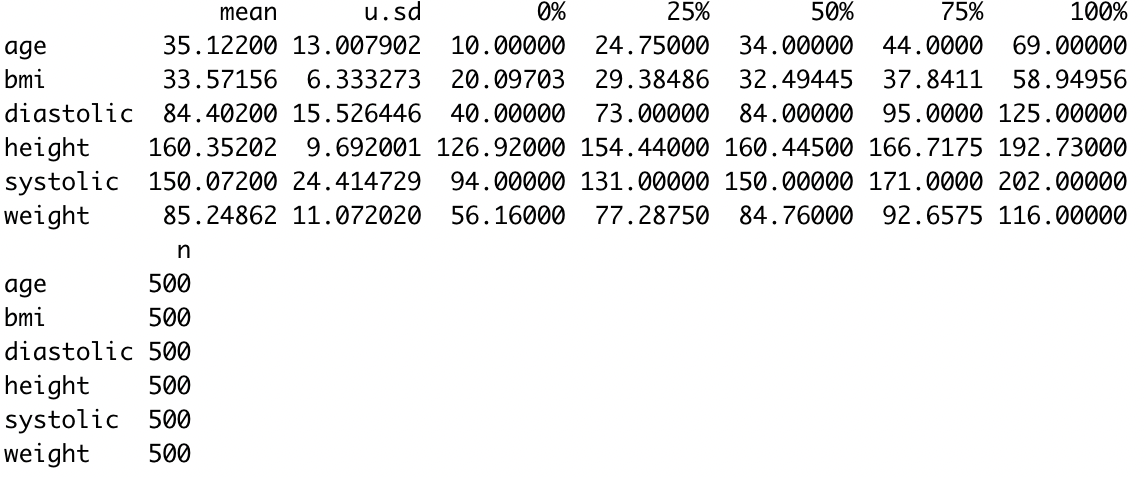


Figure 7: Summary Statistics of the 500 Subjects in the Dataset.

From Figure 7, we see the mean, standard deviation, and quartile values for the different variables on our 500 subjects in the dataset.

### Linear Model Relating BMI with Systolic BP

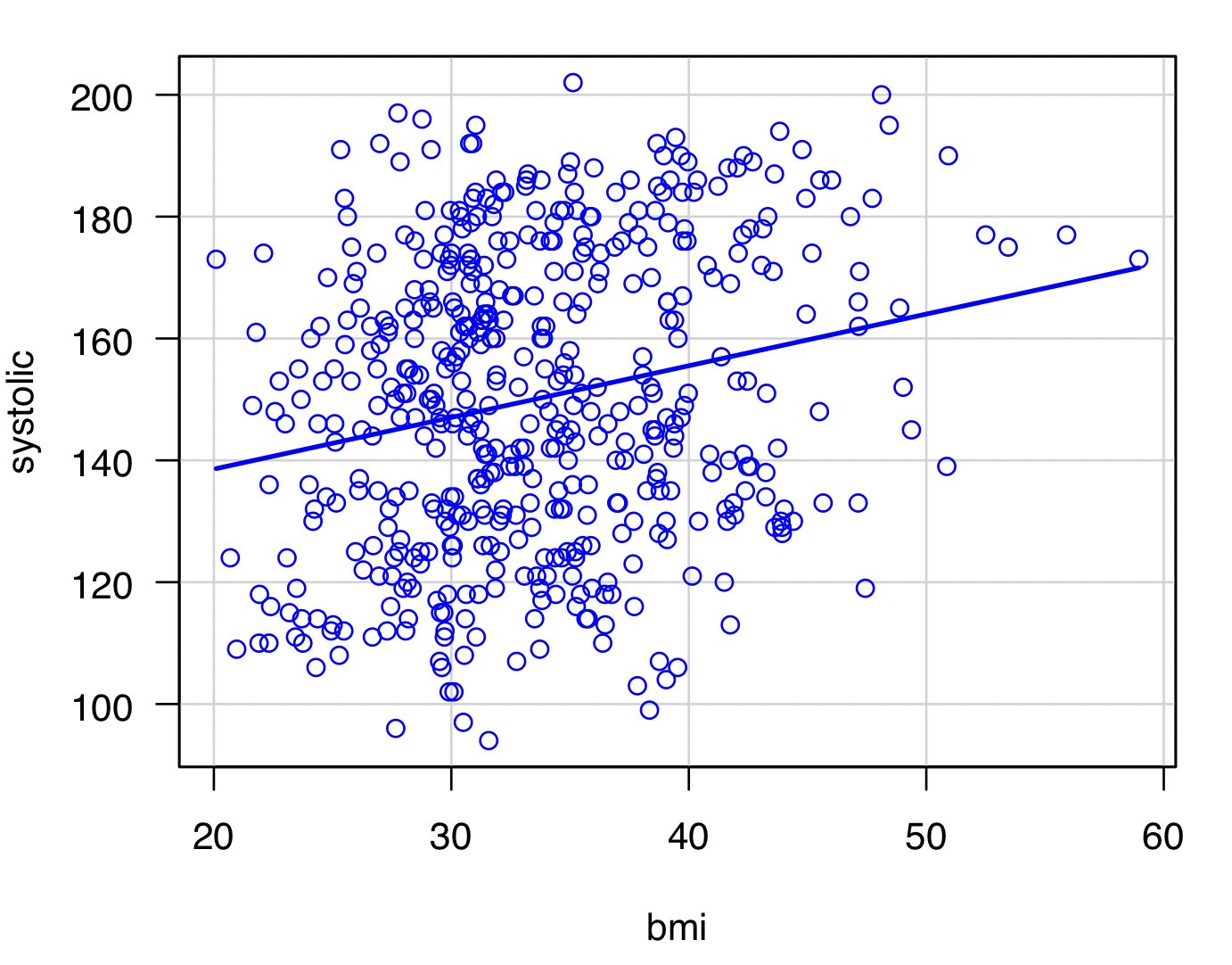


Figure 8: Linear Regression Model of BMI vs SBP of the 500 Subjects in the Dataset.

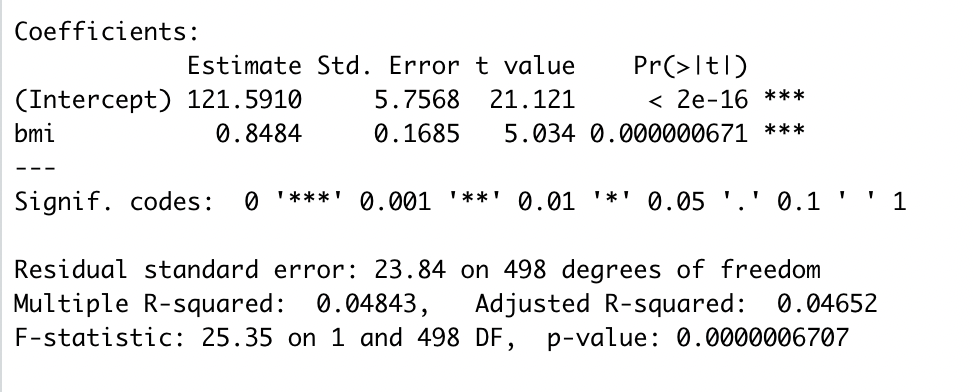


Figure 9: Table of BMI vs SBP of the 500 Subjects in the Dataset.

Based on Figures 8 and 9, the regression line that models the relationship between BMI and SBP for the 500 subjects is SBP = 121.6 + 0.84(BMI). The positive slope value tells us that the data is weakly positively correlated. R^2 = 0.05, meaning that 5% of the variability in SBP can be explained by BMI values.

### Linear Model Relating BMI with Diastolic BP

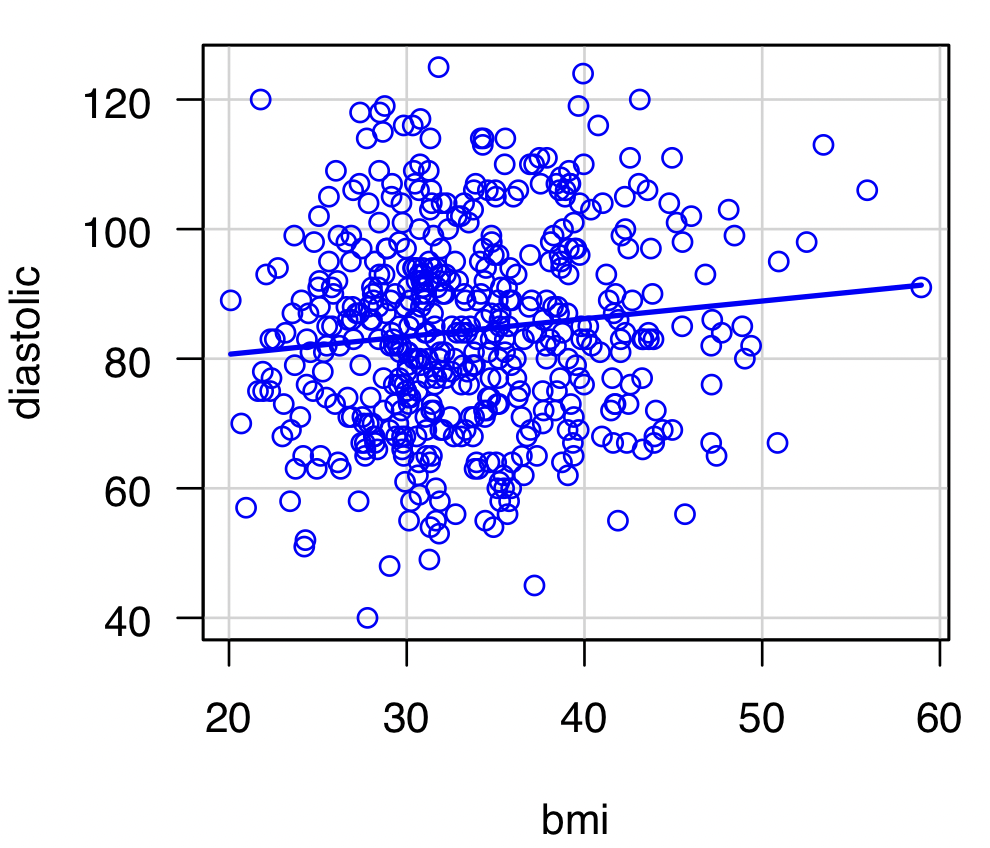


Figure 10: Linear Regression Model of BMI vs DBP of the 500 Subjects in the Dataset.

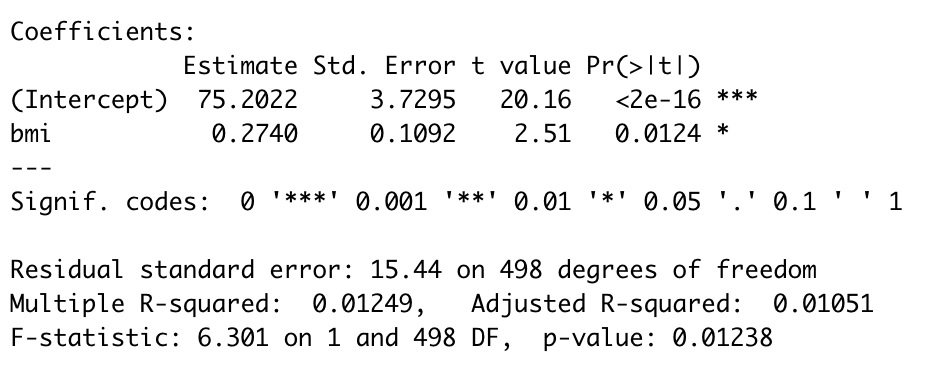


Figure 11: Table of BMI vs DBP of the 500 Subjects in the Dataset.

Based on Figures 10 and 11, the regression line that models the relationship between BMI and DBP for the 500 subjects is DBP = 75.6 + 0.27(BMI). The positive slope value tells us that the data is weakly positively correlated. R^2 = 0.01, meaning that 1% of the variability in DBP can be explained by BMI values.

### Conclusions

From the data presented in this report, we can see that the male and female subjects were well-matched for the various variables presented: age, height, weight, BMI, SBP, and DBP. When comparing BMI vs SBP and BMI vs DBP, we can see that the linear models show a weak positive correlation between the stated variables.