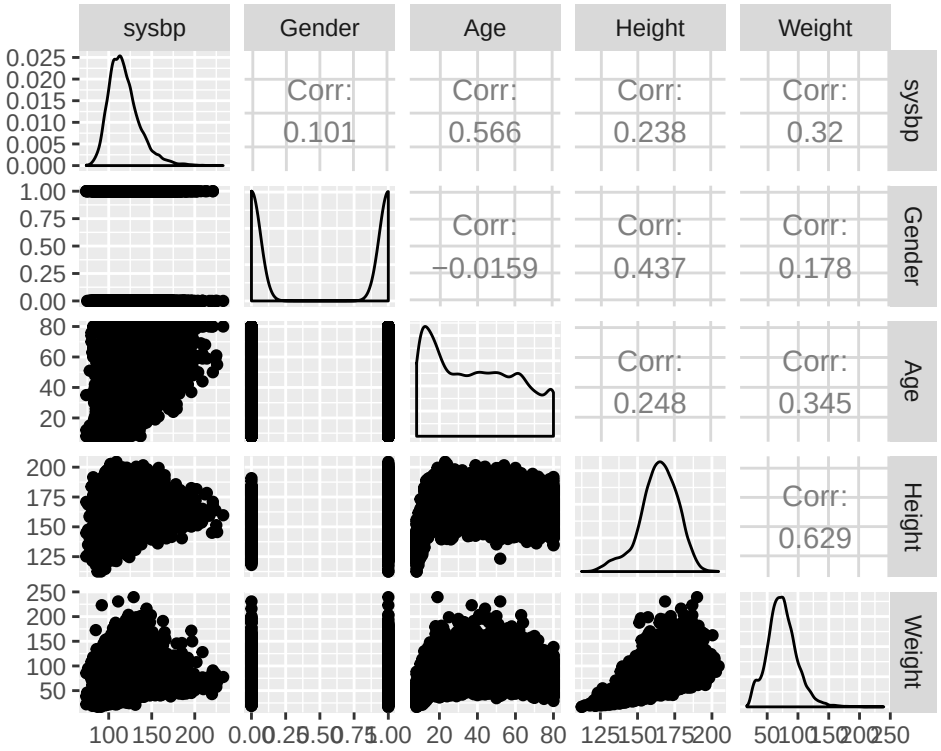
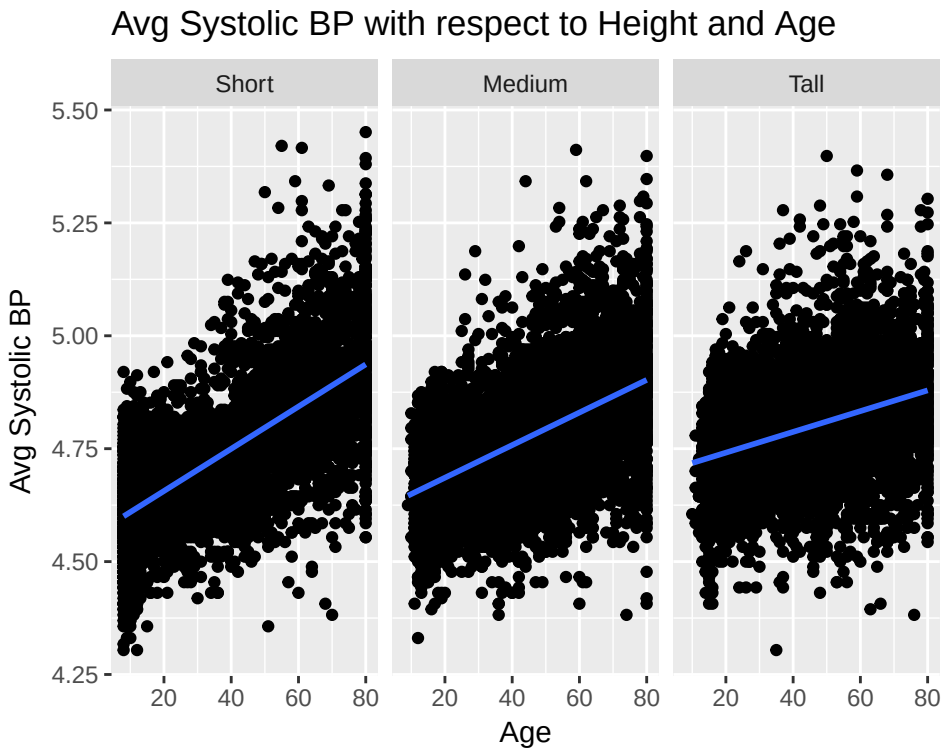
Problem Set 7

Saniya Ambavanekar

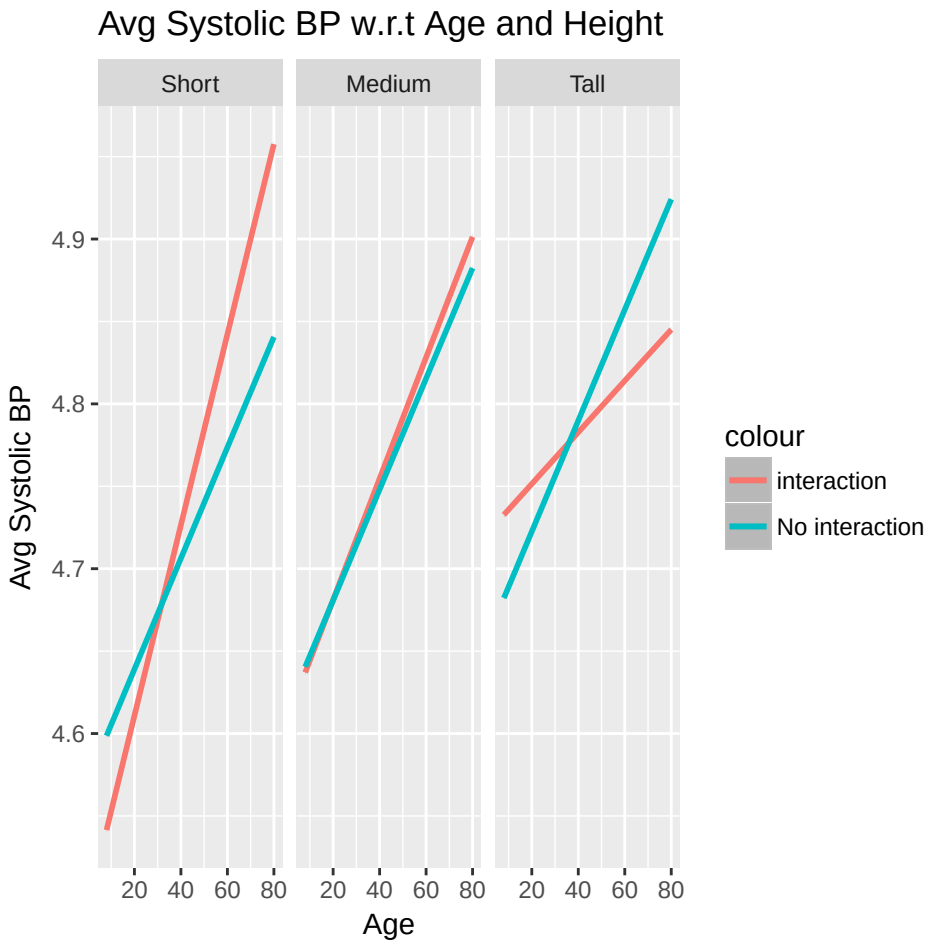
1. Fit a (weighted) model or models to the NHANESraw data to predict average systolic blood pressure (BPSysAve) using some combination of Age, Gender, Height, and Weight as explanatories. Briefly justify your modeling choices. Graph your fitted model (including at least one faceted plot) and explain what the plots tell you.



From ggpairs, the density plot of BPSysAve it shows it is right skewed, hence some transformation like log will be beneficial. We can also see that height, weight and Age are somewhat strongly correlated as compared to gender. Also height and weight are highly correlated to one another, hence we can choose either of them as our explanatory variable.

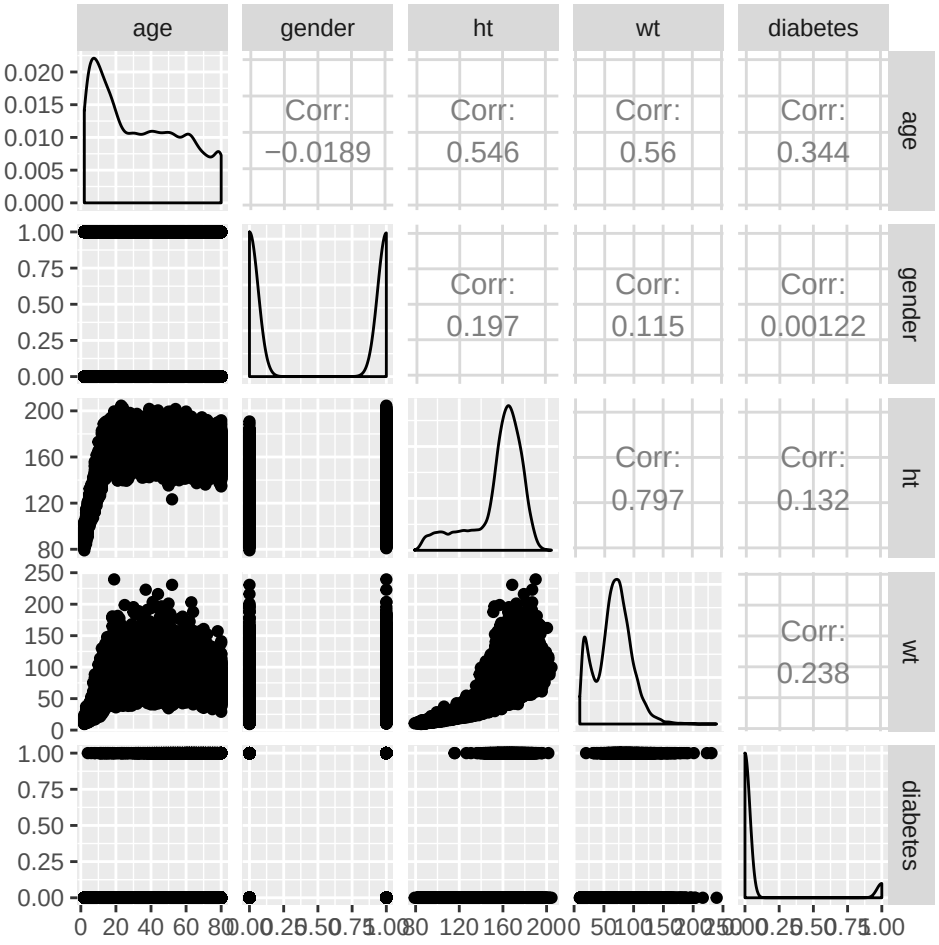


From above plot, we can observe the trend is increasing,there is slight change in slopes from short height to tall height for increase in the age. We can get a better picture, how Height and Age affect the change in Avg Systolic BP by fitting a model.

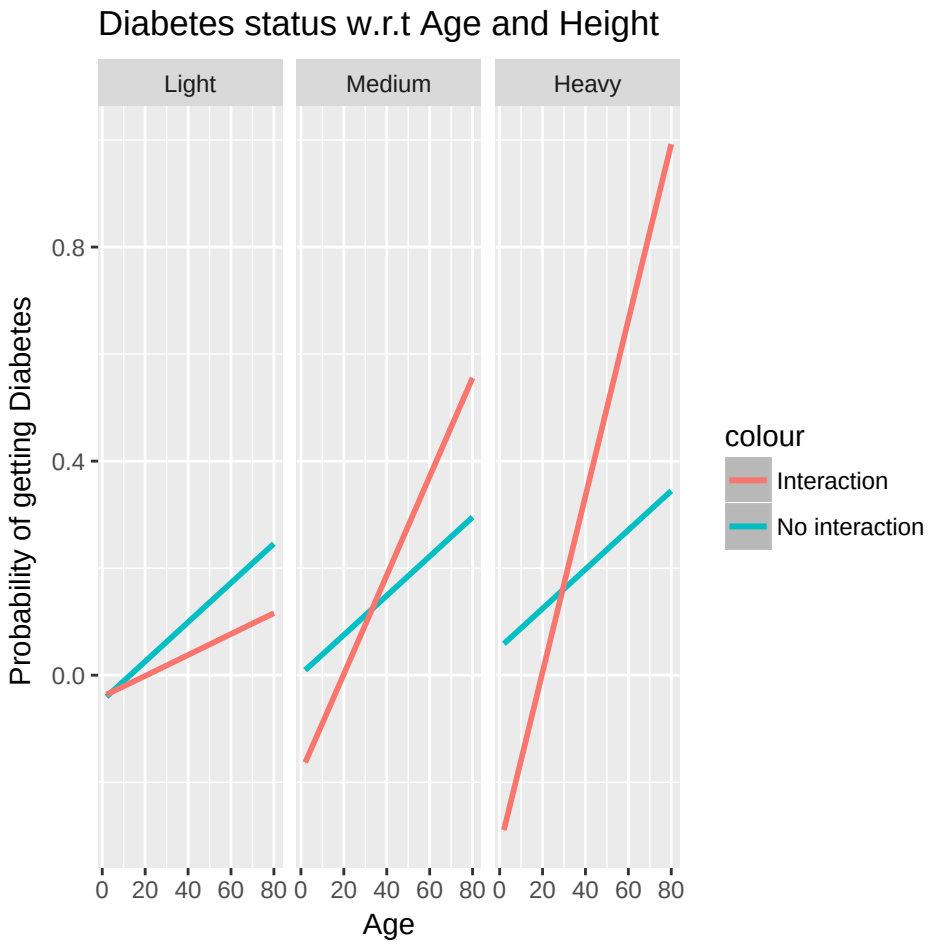


From the above plot, we can observe significant differences in short and tall people and their changes in systolic BP. I will explain for short people where we can see major differences. 1. Short people with age less than 20 with interaction there is less systolic BP as compared to the one with no interaction. 2. Also for the short people as the age increases, with interaction the systolic BP increases as compared to that with no interaction. We can interpret that there is some relation between Age and Height as both of them affect the systolic BP of the people significantly.

2. The NHANESraw data set also contains a Diabetes variable that states whether the individual has diabetes (with some missing values that you may ignore.) Fit a model or models to the NHANESraw data to predict diabetes status using some combination of Age, Gender, Height, and Weight as explanatories. Briefly justify your modeling choices. Graph your fitted model (including at least one faceted plot) and explain what the plots tell you.

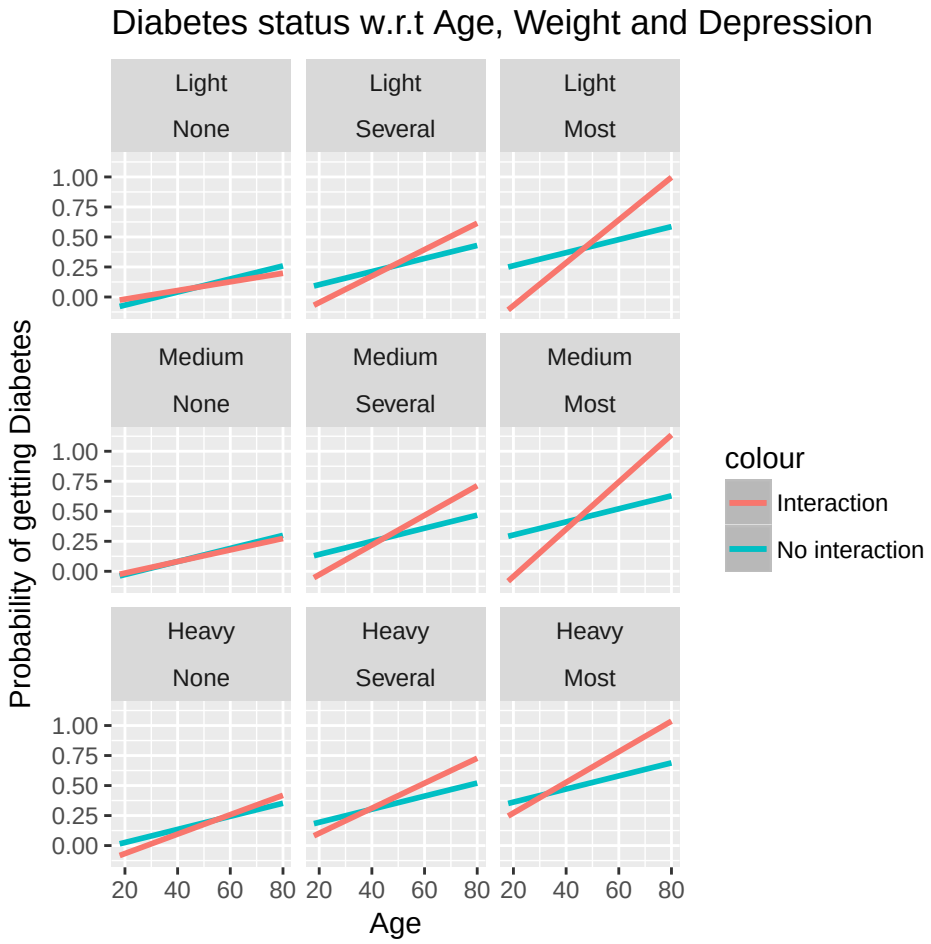


From above ggpairs for diabetes as response variable, we also see that height, weight and age are somewhat strongly correlated to the diabetes. Since it is known that weights play an important role in diabetes, I have considered two explanatory variables age and weight to check its role in diabetes prediction



From above plot, we can see that people who are heavily weighted there is a higher chance of them to get diabetes as the age increases. This is interpreted by the interaction between the age and weight. If there is no interaction between age and weight we get three somewhat parallel lines from which we cannot explain much.

Add one more explanatory variable to your model to predict diabetes status. This must be a different variable to the variables chosen by other members of your final project group. Convince us that adding this variable is actually helpful. Graph your fitted model (including at least one faceted plot) and explain what the plots tell you that your plots for question 2 don't.



Depression plays a role in getting diabetes. But we need to check if weight and depression are correlated or not. Depression has three levels None: Not depressed Several: Depressed for some days Most: Most of Days of year depressed From the above plots, we can observe that most depressed people irrespective of their weights have higher chances of getting diabetes as the age increases.