

Name: Vikram Sahai Saxena

Net ID: vs799

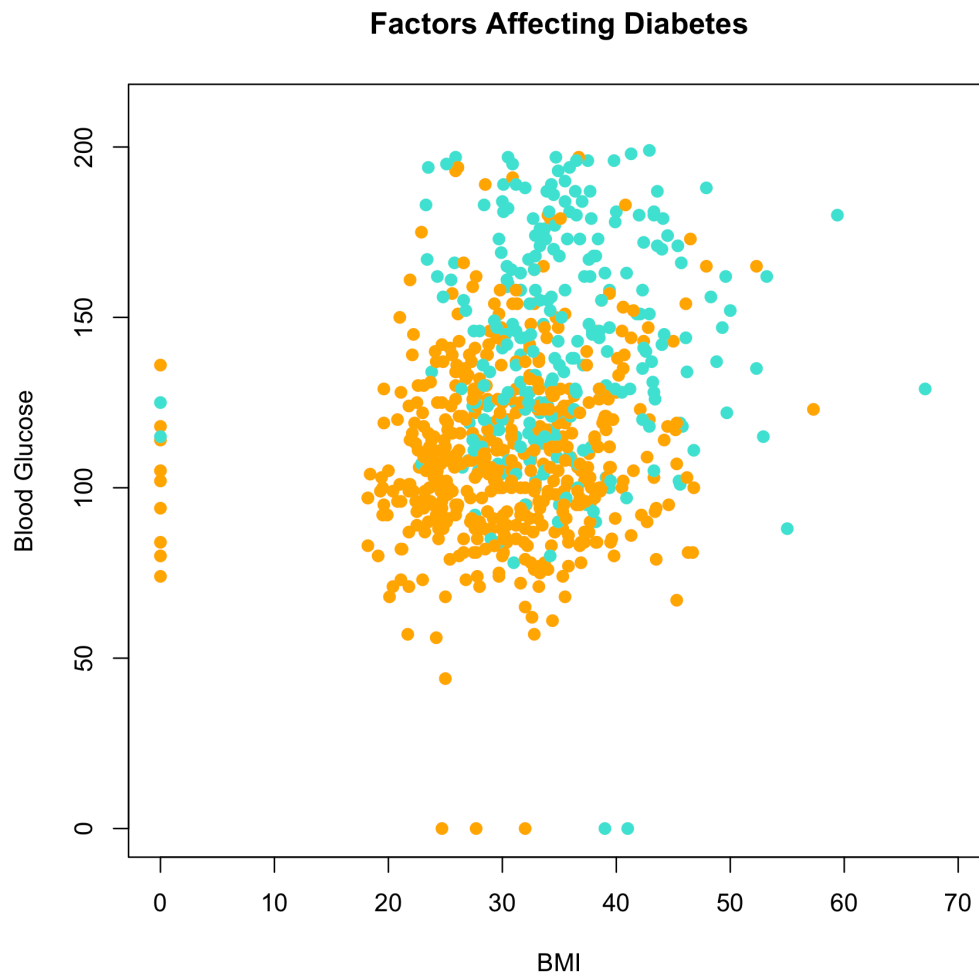
RUID: 219004709

### Module 7 Exercise: Discriminants

**i. The entire R code used when creating the scatter plot in (1)**

```
a<-read.csv("pima-indians-diabetes.csv")  
plot(a$BMI, a$Glu, main="Factors Affecting Diabetes", xlab="BMI", ylab="Blood  
Glucose", xlim=c(0,70), ylim=c(0,210), pch=19, col=ifelse((a$Diabetes == 1),"turquoise",  
"orange"))
```

**ii. Screenshot of the scatter plot created in (1)**

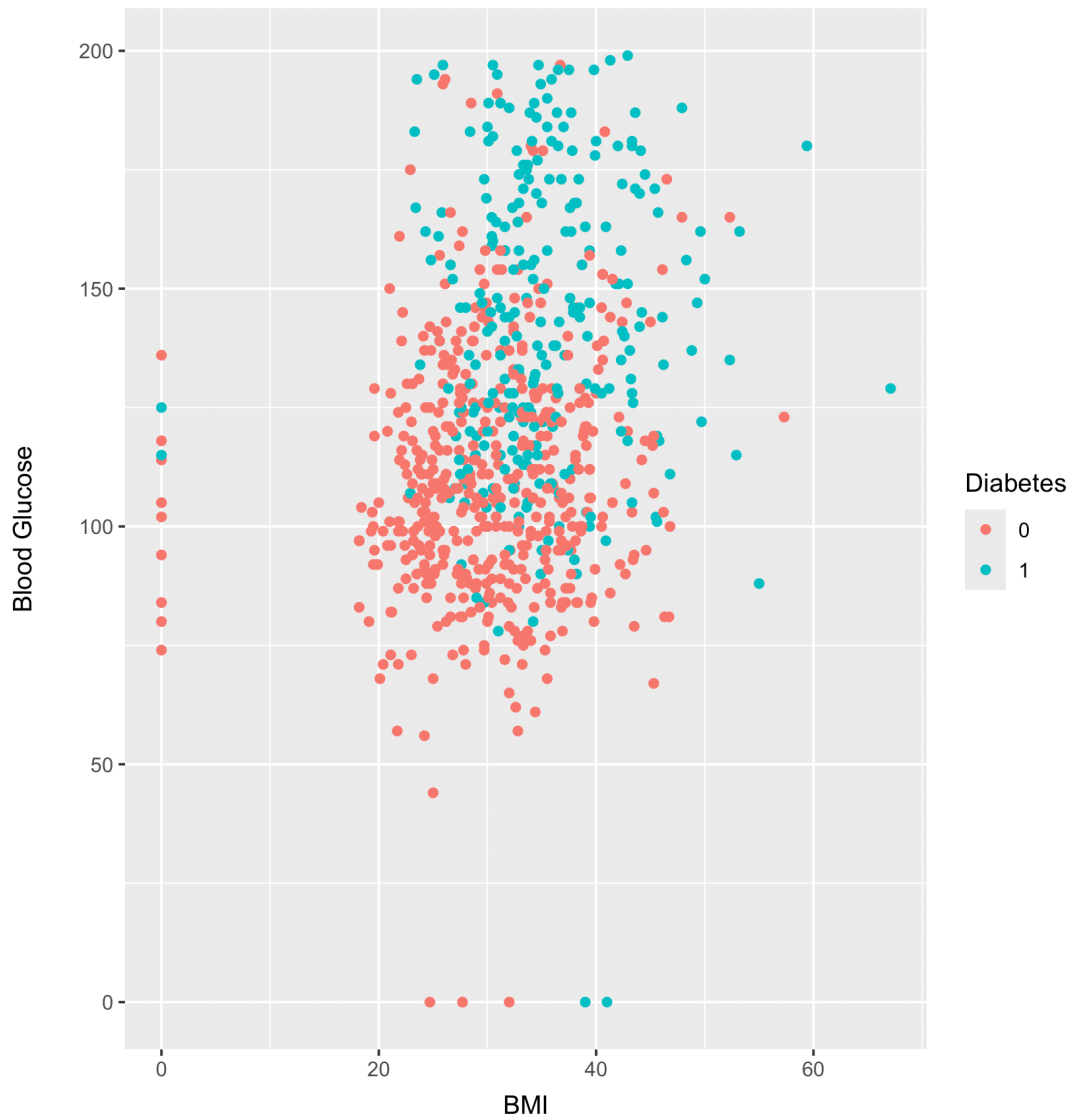


iii. The entire R code used when creating the scatter plot in (2)

```
library(ggplot2)
ggplot(a, aes(x=BMI, y=Glu,
color=factor(Diabetes)))+geom_point()+xlab("BMI")+ylab("Blood
Glucose")+guides(color=guide_legend(title="Diabetes"))+ggtitle("Factors Affecting
Diabetes")
```

iv. Screenshot of the scatter plot created in (2)

Factors Affecting Diabetes



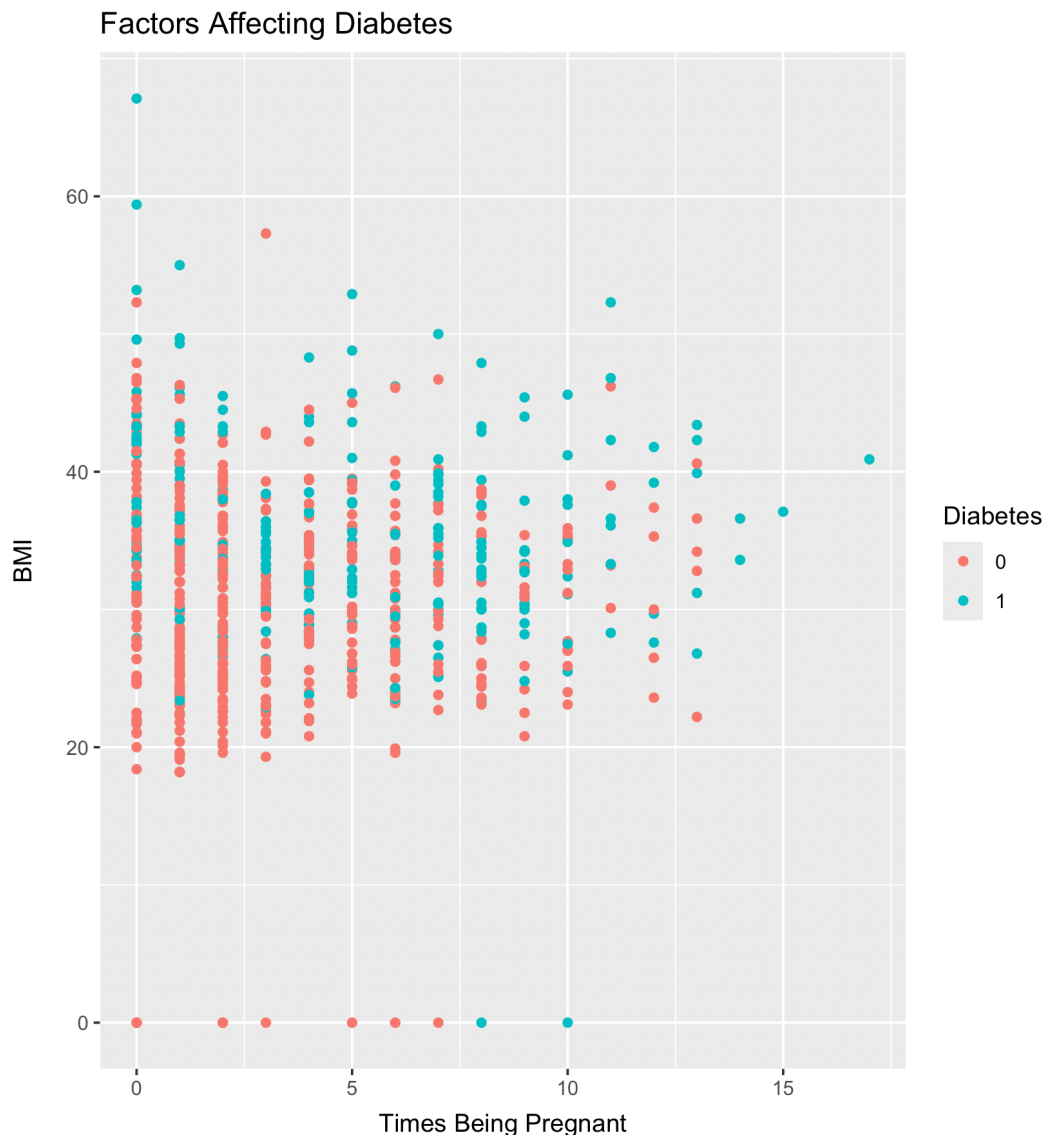
- v. **Whether linear regression analysis can be successfully performed on the above plot of BMI vs. Blood Glucose, to show whether it can successfully determine if a woman is diabetic or not**

Linear regression analysis cannot be performed successfully for the above plot because we have considerable overlap between diabetic and non-diabetic points, making it difficult to draw a line separating the two groups. Also, the outliers in the plot can affect the slope of the regression line, making the linear regression analysis less reliable.

- vi. **The entire R code used when creating the scatter plot in (4)**

```
ggplot(a, aes(x=Preg, y=BMI, color=factor(Diabetes)))+geom_point()+xlab("Times Being Pregnant")+ylab("BMI")+guides(color=guide_legend(title="Diabetes"))+ggtitle("Factors Affecting Diabetes")
```

- vii. **Screenshot of the scatter plot created in (4)**



- viii. **Whether linear regression analysis can be successfully performed on the above plot of Times Being Pregnant vs. BMI, to show whether it can successfully determine if a woman is diabetic or not**

Linear regression analysis is not possible for the above plot because we have considerable overlap between diabetic and non-diabetic points, and drawing a straight line to separate them into two groups is impossible.