John Rene Lorelli

Foundations Data Science

Week 7

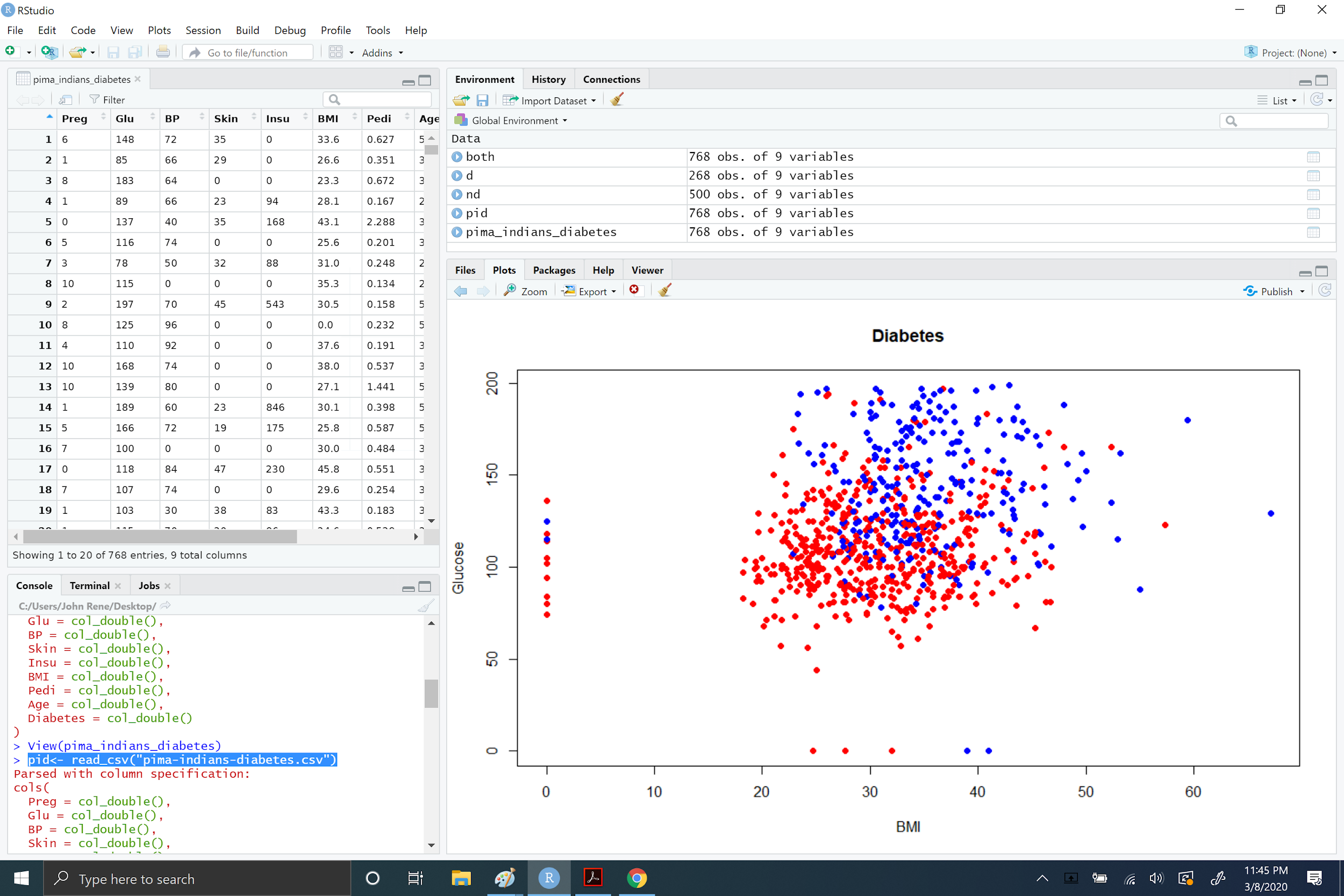
Discriminants

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

pid<- read\_csv("pima-indians-diabetes.csv")

plot(pid$BMI, pid$Glu, main="Diabetes", xlab="BMI", ylab="Glucose", pch=19, col=ifelse((pid$Diabetes == "1"), "blue", "red"))

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

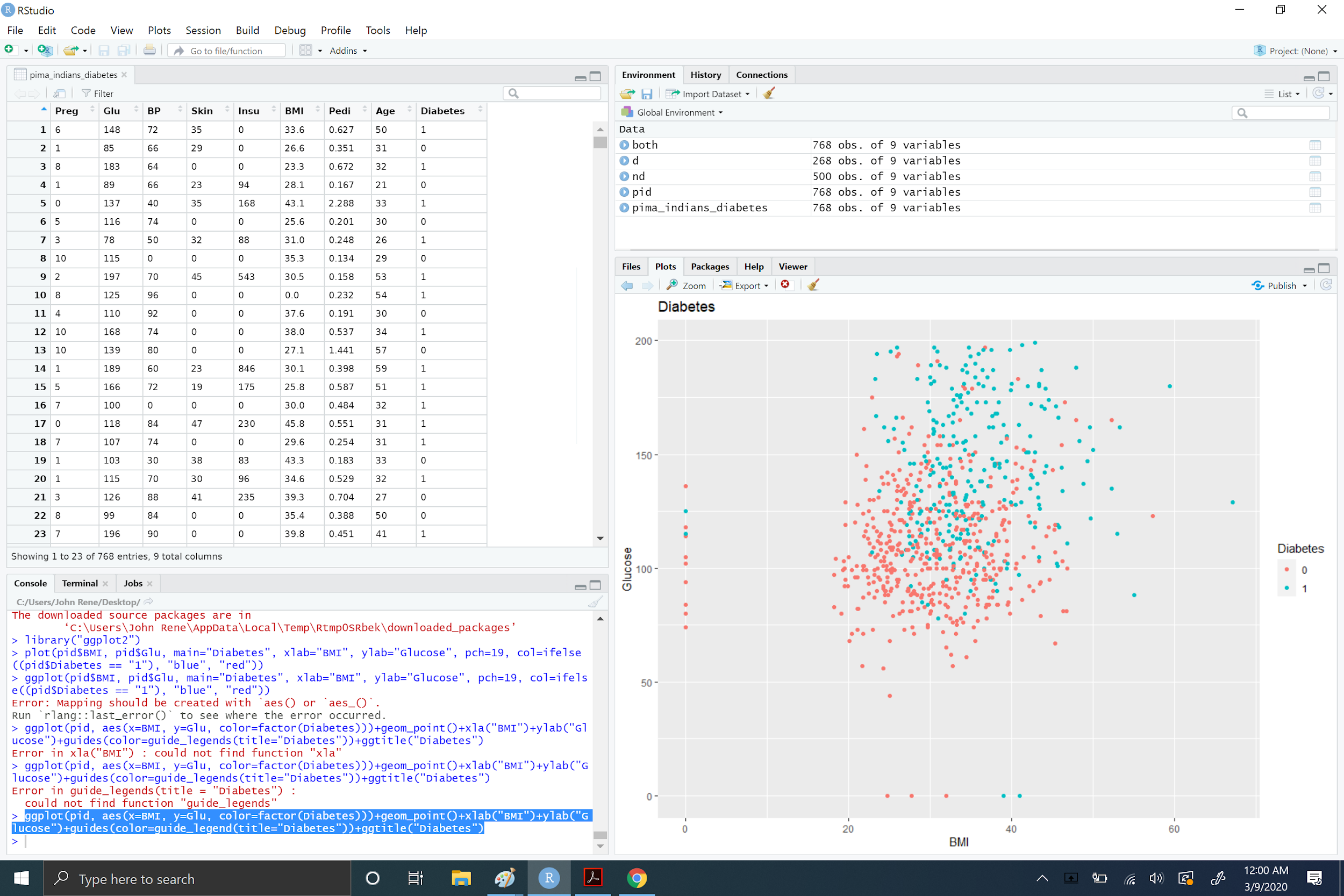


1. The entire R code used when creating the scatter plot in (2).

install.packages("ggplot2")

library("ggplot2")

ggplot(pid, aes(x=BMI, y=Glu, color=factor(Diabetes)))+geom\_point()+xlab("BMI")+ylab("Glucose")+guides(color=guide\_legend(title="Diabetes"))+ggtitle("Diabetes")

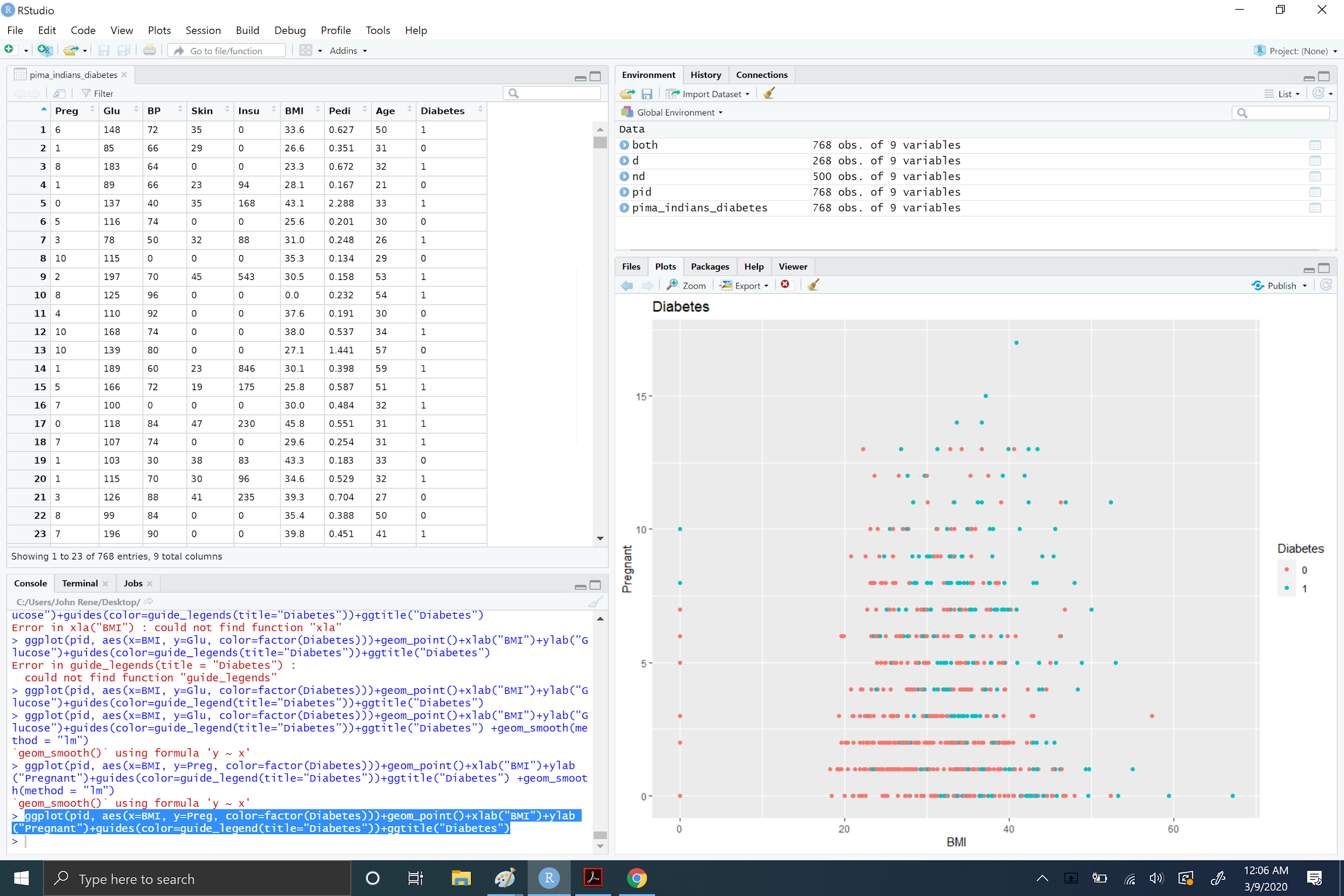
1. Screenshot of the scatter plot created in (2).
2. Whether linear regression analysis can be successfully performed on the above plot of BMI vs. Blood Glucose, to show whether it can successfully determine if a woman is diabetic or not A list of actions you could incorporate to reduce the identified problems/weaknesses.

* We would need min max and mean slope
* We may need some more parameters like certain age or between ages
* I ran it in ggplot using geom\_smooth command but it did not look accurate

1. The entire R code used when creating the pie chart in (4).

ggplot(pid, aes(x=BMI, y=Preg, color=factor(Diabetes)))+geom\_point()+xlab("BMI")+ylab("Pregnant")+guides(color=guide\_legend(title="Diabetes"))+ggtitle("Diabetes")

1. Screenshot of the pie chart created in (4).



1. 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.

I do not think it will be successful I tried the command geom\_smooth and the lines did not make sense.