John Rene Lorelli

Foundation of Data Science

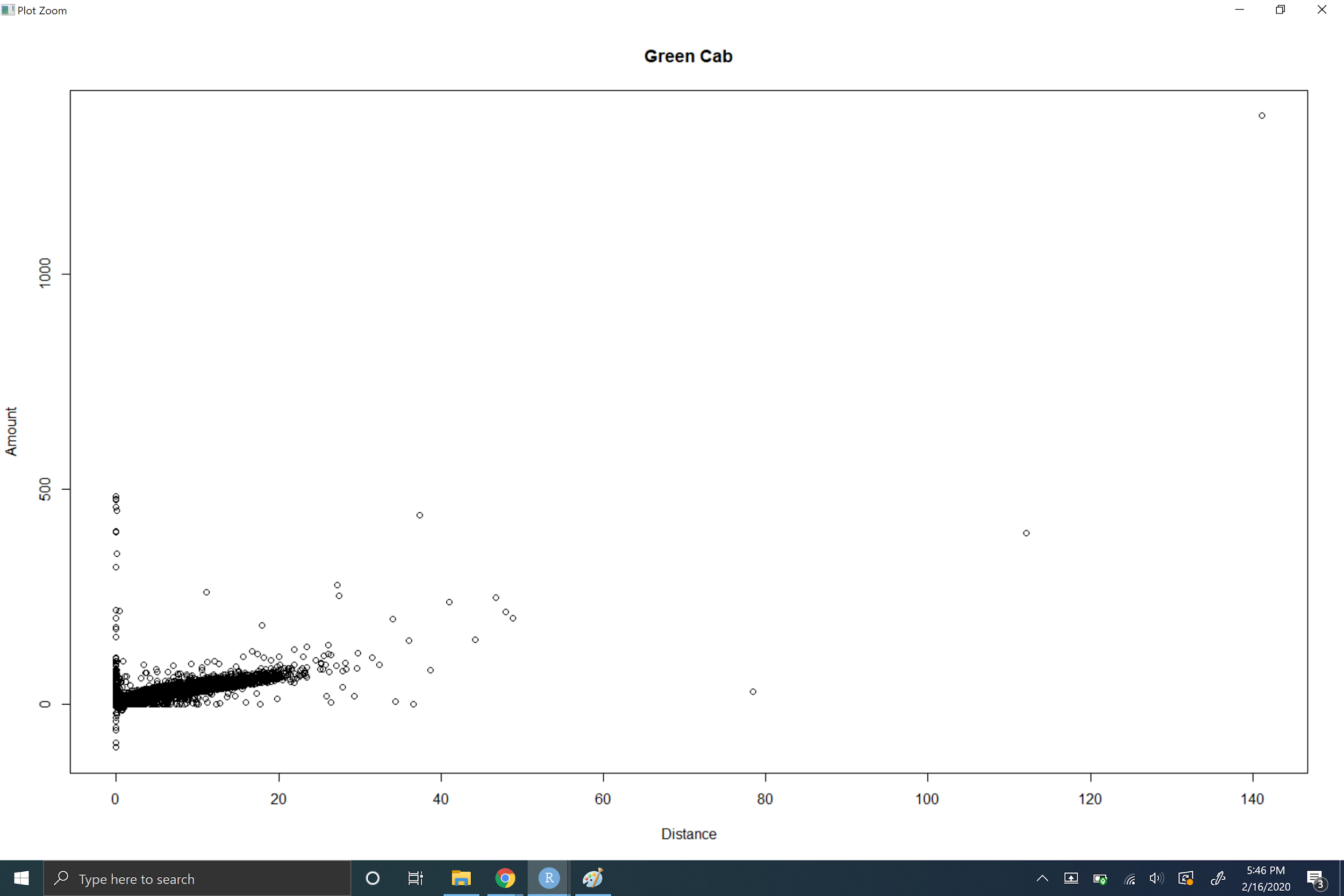
Week 4 Report

M4 Exercise: Scatter Plot

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

* I chose to do this one in RStudio instead of RGUI so I imported greencab.csv into the working directory instead or read command.
* head(greencab)
* gc<-greencab
* plot(gc$Trip\_distance, gc$Total\_amount, xlab="Distance", ylab="Amount", main="Green Cab")

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



1. Your reasoning as to why you selected which variable to be depicted in which axis.

X axis is used for cause and y axis is used for effect. Total amount is dependant on the distance.

1. The message being sent out by this plot.

There seems to be a steady line showing the further the distance the higher the amount.

1. Weaknesses you see in this plot.

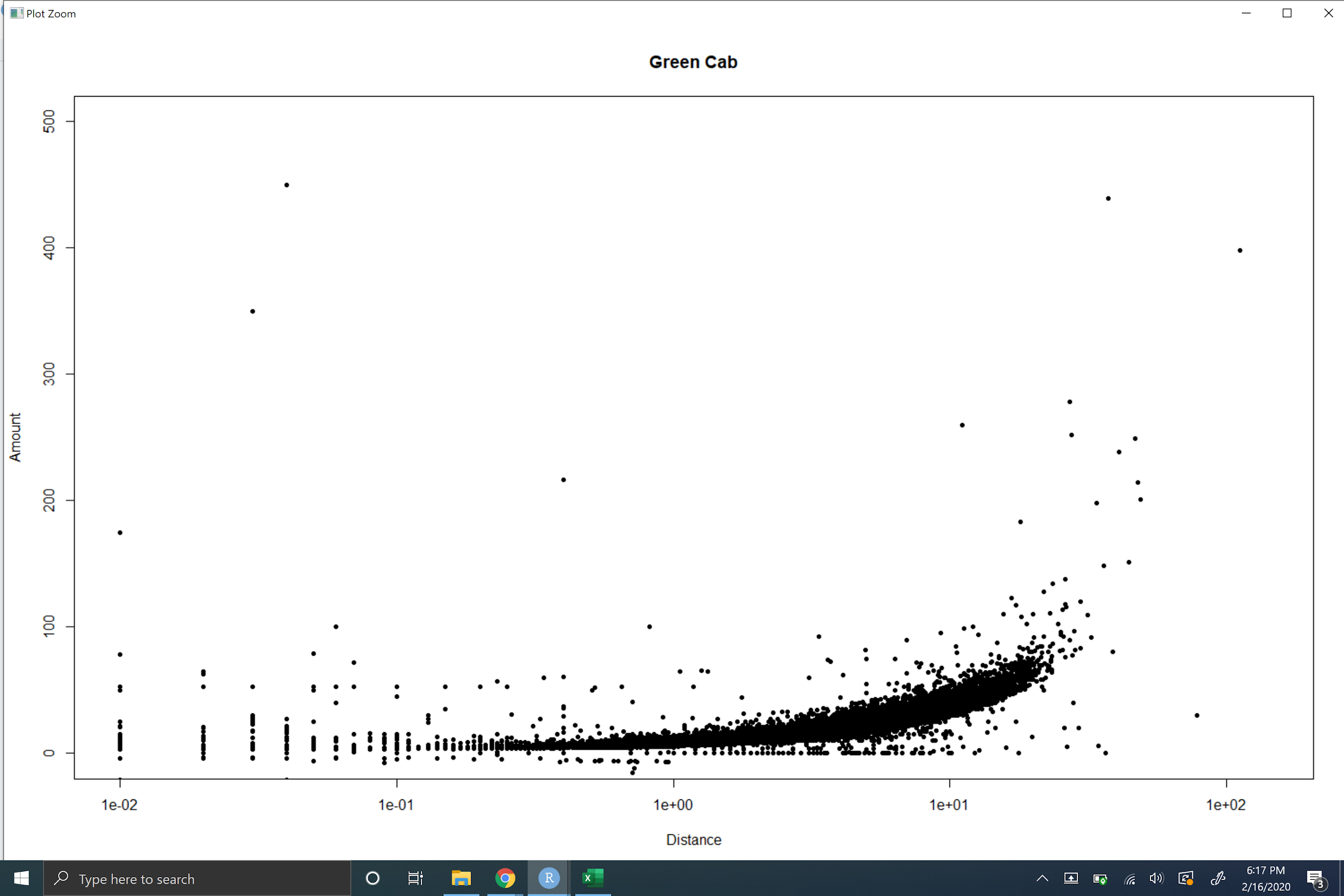
Everything is difficult to read and it is a little difficult to see if there is causality.

1. A list of actions you could incorporate to minimize the identified weaknesses, and emphasize the plot’s message even further.

Limit amount or limit distance to see if it gives better feedback. Add a regression line or curve line to help visually.

1. The entire R code used when creating the scatter plot in (6).

plot(gc$Trip\_distance, gc$Total\_amount, xlab="Distance", ylab="Amount", main="Green Cab", pch=20, log="x", ylim=c(0, 500))

1. Screenshot of the scatter plot created in (6)
2. The entire R code used when creating the scatter plot in (7).

plot(gc$Trip\_distance, gc$Total\_amount, xlab="Distance", ylab="Amount", main="Green Cab", pch=20, log="y", xlim=c(0, 100))

1. Screenshot of the scatter plot created in (7)

