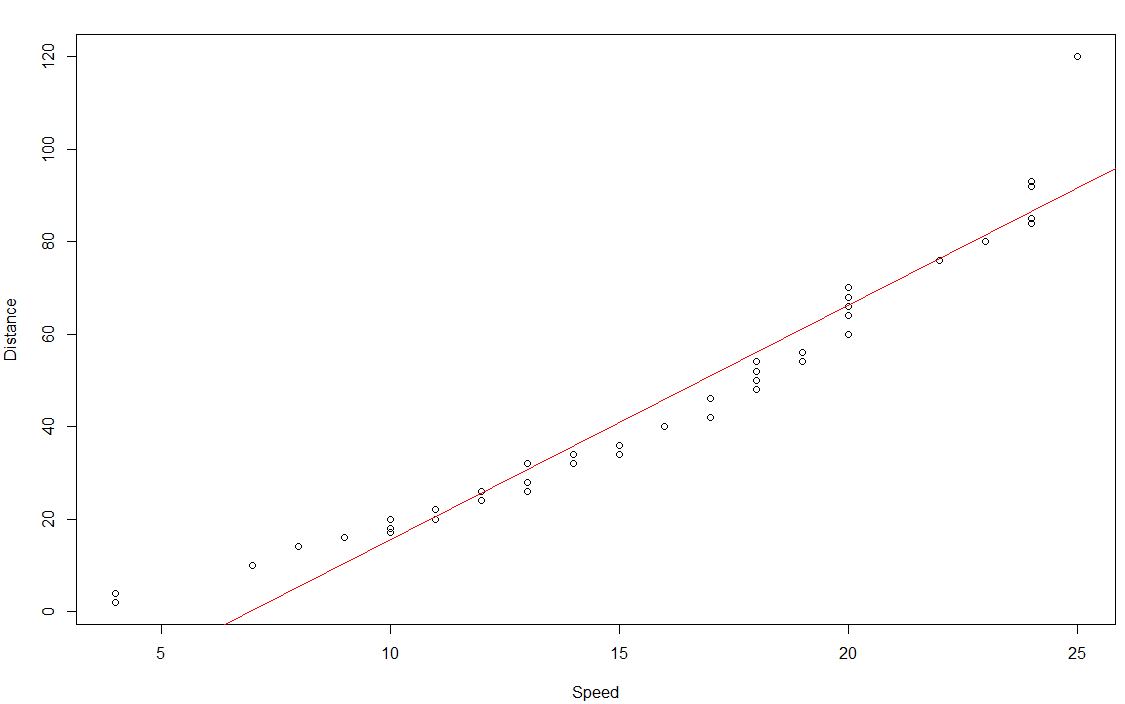
**Course 3 Task 1: Informal Report – Michelle Giniewicz**

**Predictions – R Tutorial**

In the R Tutorial I was able to create predictions concerning how far a certain car can travel based on speed. The model was extremely accurate, with a Multiple R-Squared value of 0.92. The model also showed that there is a statistically significant relationship between Speed and Distance, with a p-value of <2.2e-16. I wanted to learn how to interpret the results of my prediction, since what we were originally provided just predicted the distance based on speed; however, it was hard for me to understand if the predictions were accurate or not. Therefore, I did some research and learned how to plot the model’s regression line against the Speed and Distance of the whole dataset, and it was evident that the model was a good fit for the data:



**Predictions – Find the Errors Task**

In the Find the Errors Task, I was able to create predictions concerning the petal length using the petal’s width. The model was extremely accurate, with a Multiple-R Square value of 0.93. The model also showed that there is a statistically significant relationship between Speed and Distance, with a p-value of <2.2e-16. Additionally, I plotted the model’s regression line against the Petal Width and Length, and it was evident that the model was a very good fit for the data:

Chart, scatter chart

Description automatically generated

**Errors/Warnings and How I Overcame Them**

There were a lot of errors that I encountered in the Find the Errors task:

1. I had an error getting the CSV file and realized that the file name needed to be in quotes.
2. I got a few errors due to spelling mistakes:
   1. ex: in summary(risDataset) or str(IrisDatasets), instead of the correct IrisDataset
   2. ex: trainSizes, instead of the correct trainSize
   3. ex: predictions, instead of the correct prediction
3. There was an error in plotting the histogram of Species, since histograms must use numerical data. To overcome this, I defined numbers for each species (Setosa = 1, Versicolor = 2, and Virginica = 3) and created a new column to map the data in the Species column with the numeric definitions. I was then able to plot the species and realized that the dataset was split evenly between each type of Iris.
4. I received an error where a closed parenthesis was missing, which was a very easy fix, and I didn’t end up getting an error since I made the change prior to running the line of code.
5. I got an error on the Normal Quantile Plot, since only the dataset was included. To fix this, I also needed to include the column name for the code to execute successfully.
6. There was an error when trying to convert Species to numeric data. Since there was no way for it to know how to convert to character data to numeric, it just made all rows in that column ‘NA.’ To overcome this, I had to re-import the data to get the Species data back.
7. While not an error exactly, there was an issue with defining the training dataset. In the POA, it said to use 20%, but I knew this was wrong since the training dataset is supposed to be larger than the test dataset. To overcome this, I used 70%, like we did in the R Tutorial.
8. There was an error in creating the training and test datasets. To overcome this, I made sure to define training\_indices before calling it in my code.
9. There were two sections in the POA that told you to set the seed and then create your training and test sets. I disregarded the second part, since you only need to set the seed and create training and test sets once, not twice (with different seeds).
10. I had an error when creating the Linear Regression Model, based on the syntax that was used in the POA. To resolve this error, I made sure to put in parentheses the following syntax: (X ~ Y, training\_dataset).
11. There was an error when calling the predict function – you have to put a comma between the model name and the dataset you want to use.

**Other Takeaways**

I thought it was very straightforward when installing both R and RStudio – I did not encounter any issues. I also felt that the R Tutorial was extremely helpful, and I would highly recommend it to others. Walking through the tutorial is why I was able to overcome all the errors in the POA task so quickly. The main lesson I learned in this task was to take my time going through the tutorial, especially since I have no previous experience with R or RStudio. By going through step-by-step and making sure that I fully understood each line of code, I was able to easily resolve the errors that I found in the POA. For other employees starting to use R instead of python, I would recommend that they go through the R Tutorial (or something similar) to understand the basics of R and RStudio. I would also recommend they do some research to learn new things they are curious about, like I did with plotting the regression line from the model on the scatterplot to visually see if the model was a good fit.