Errors in the code with Iris Dataset :

Code: Install.packages(readr)
Error Message : Error in install.packages : object 'readr' not found
fix: readr should be within double quotes

Code: IrisDataset <- read.csv(iris.csv)
Error Message : Error in read.table(file = file, header = header, sep = sep, quote = quote, :
object 'iris.csv' not found
Fix: iris.csv should be within double quotes

Code: summary(risDataset)
Error Message: Error in summary(risDataset) : object 'risDataset' not found
Fix: the name is IrisDataset

Code: str(IrisDatasets)
Error in str(IrisDatasets) : object 'IrisDatasets' not found
Fix: Delete the s in the end of the name

Code: plot(IrisDataset$Sepal.Length
Error: unexpected symbol in:
"plot(IrisDataset$Sepal.Length
Fix : add the closing parentheses.

Code : qqnorm(IrisDataset)
Error in FUN(X[[i]], ...) :
only defined on a data frame with all numeric-alike variables
Fix: Convert one of the char column to numeric and use qqnorm function with individual columns instead of the entire dataframe

Code : testSize <- nrow(IrisDataset) - trainSet
Error in FUN(left, right) : non-numeric argument to binary operator
Fix: Typo change trainSet to trainSize

Code : trainSizes
Error: object 'trainSizes' not found
Fix: Typo. Change it to trainSize instead of trainSizes.

Code: LinearModel<- lm(trainSet$Petal.Width ~ testingSet$Petal.Length)
Error: Error in eval(predvars, data, env) : object 'testingSet' not found
Fix : Specify the dataframe and the column names to do the regression.
LinearModel<-lm(Petal.Length~Petal.Width,data=IrisDataset)

Code : #prediction<-predict(LinearModeltestSet)
Error : Error in predict(LinearModeltestSet) :
  object 'LinearModeltestSet' not found
Fix : Change to prediction<-predict(LinearModel)

Prediction on Iris dataset:

Predictions concerning the petal length through Petal Width:

Residuals: The residuals range from -1.335 to 1.394 and the median is at 0.029. This is the error between the prediction of the model and actual results. The median is close to 0.

Coefficients:

Estimate: Used to predict the value of the response variable. For every unit increase in distance the model will produce an increase of **4.6959** units of speed.

Std.Error: The average amount that the estimate varies from the actual value.

R-Squared gives a measurement of what % of the variance in the response variable can be explained by the regression.

Multiple R-Squared : Gives a measurement of what % of the variance in the response variable can be explained by the regression. If there are more predictor variables, the R-squared will typically increase. If there is a large difference between Multiple R-Squared and Adjusted R-Squared, then the model may be overfitting.

Adjusted R-Squared: controls for each additional predictor added(to prevent from overfitting), so it may not increase as you add more variables.

They both are at 92%. So it is a good % to explain the response variable.

p-Value: The p value of Petal Width is 2.2e-16 which is less than 0.05. The probability of rejecting the null hypothesis is greater. It means that there is a statistically significant relationship between petal length and Petal width.

Significance code : Since the p value is in the range [0,0.001], it has a significance code of ****

Prediction on cars dataset:

Distance a car can travel based on the speed:

Residuals: The residuals range from –8.493 to 31.939 and the median is at -1.102. This is the error between the prediction of the model and actual results.

Coefficients:

Estimate: Used to predict the value of the response variable. For every unit increase in petal length the model will produce an increase of **2.229** units of petal width.

Std.Error: The average amount that the estimate varies from the actual value.

R-Squared gives a measurement of what % of the variance in the response variable can be explained by the regression.

Multiple R-Squared : Gives a measurement of what % of the variance in the response variable can be explained by the regression. If there are more predictor variables, the R-squared will typically increase. If there is a large difference between Multiple R-Squared and Adjusted R-Squared, then the model may be overfitting.

Adjusted R-Squared: controls for each additional predictor added(to prevent from overfitting), so it may not increase as you add more variables.

They both are at 92%. So it is a good % to explain the response variable.

p-Value: The p value of Speed is 2.2e-16 which is less than 0.05. The probability of rejecting the null hypothesis is greater. It means that there is a statistically significant relationship between distance and speed.

Significance code : Since the p value is in the range [0,0.001], it has a significance code of ****