

Module 5 R Practice

Angel Waters

2022-06-26

Initial Analysis

The data used for this report was the Iris dataset provided by the base R library.

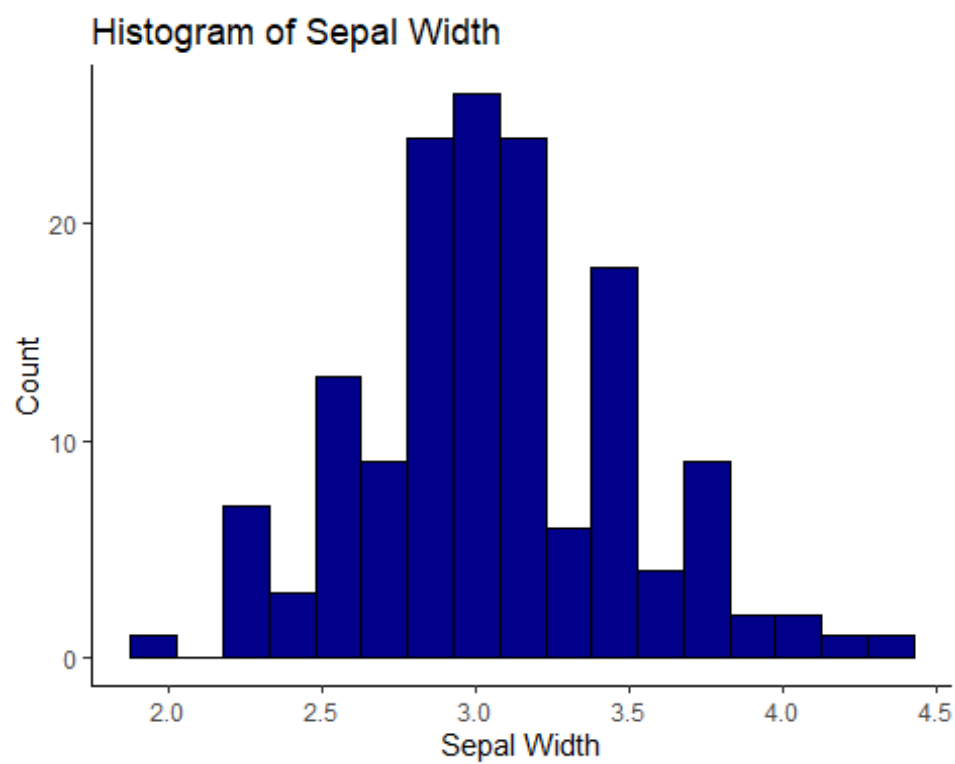
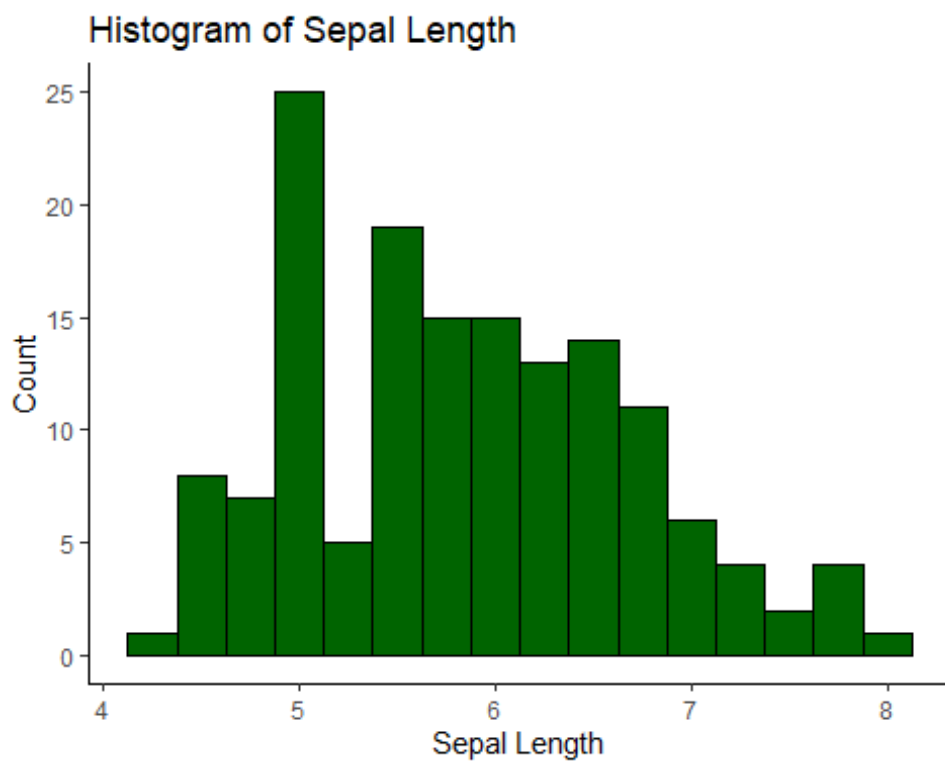
Exploratory Data Analysis

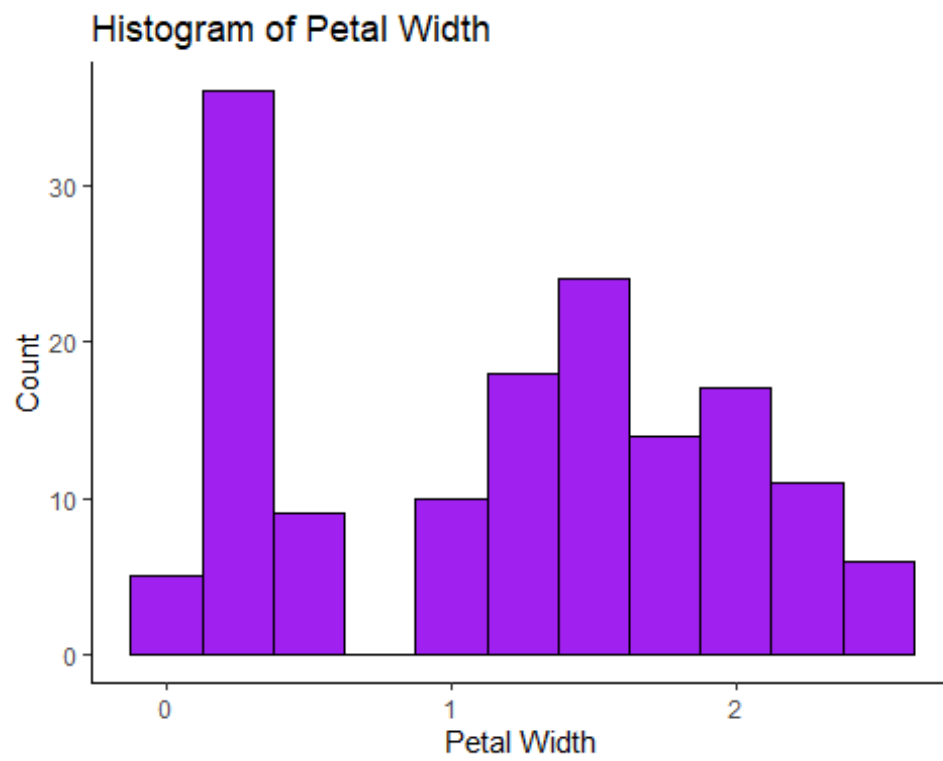
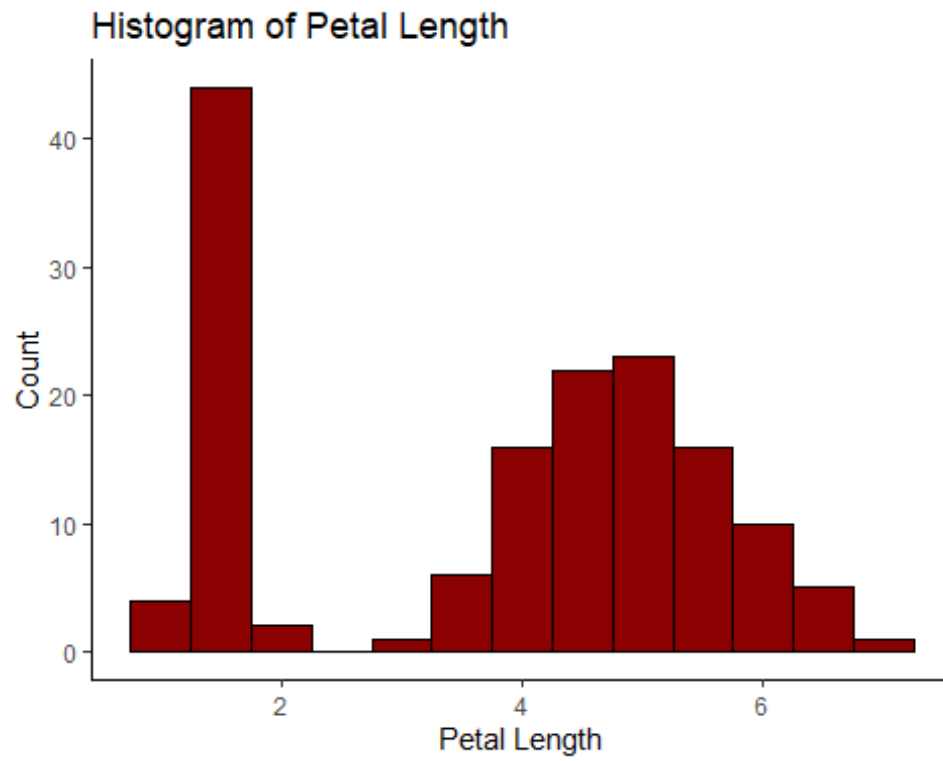
To understand the variables, some exploratory data analysis was conducted. This directed what questions should be asked because the dataset was not collected by the author of this report.

```
## Sepal.Length Sepal.Width Petal.Length Petal.Width
## Min. :4.300 Min. :2.000 Min. :1.000 Min. :0.100
## 1st Qu.:5.100 1st Qu.:2.800 1st Qu.:1.600 1st Qu.:0.300
## Median :5.800 Median :3.000 Median :4.350 Median :1.300
## Mean :5.843 Mean :3.057 Mean :3.758 Mean :1.199
## 3rd Qu.:6.400 3rd Qu.:3.300 3rd Qu.:5.100 3rd Qu.:1.800
## Max. :7.900 Max. :4.400 Max. :6.900 Max. :2.500
## Species
## setosa :50
## versicolor:50
## virginica :50
##
##
##
## Warning: headtail is deprecated. Please use the headTail function

## Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1 5.1 3.5 1.4 0.2 setosa
## 2 4.9 3 1.4 0.2 setosa
## 3 4.7 3.2 1.3 0.2 setosa
## 4 4.6 3.1 1.5 0.2 setosa
## ... ... ... ... <NA>
## 147 6.3 2.5 5 1.9 virginica
## 148 6.5 3 5.2 2 virginica
## 149 6.2 3.4 5.4 2.3 virginica
## 150 5.9 3 5.1 1.8 virginica
```

Because numerical data is being analyzed, histograms of each variable were plotted to verify the shape of the curves (data will need to be relatively normal to complete this analysis).

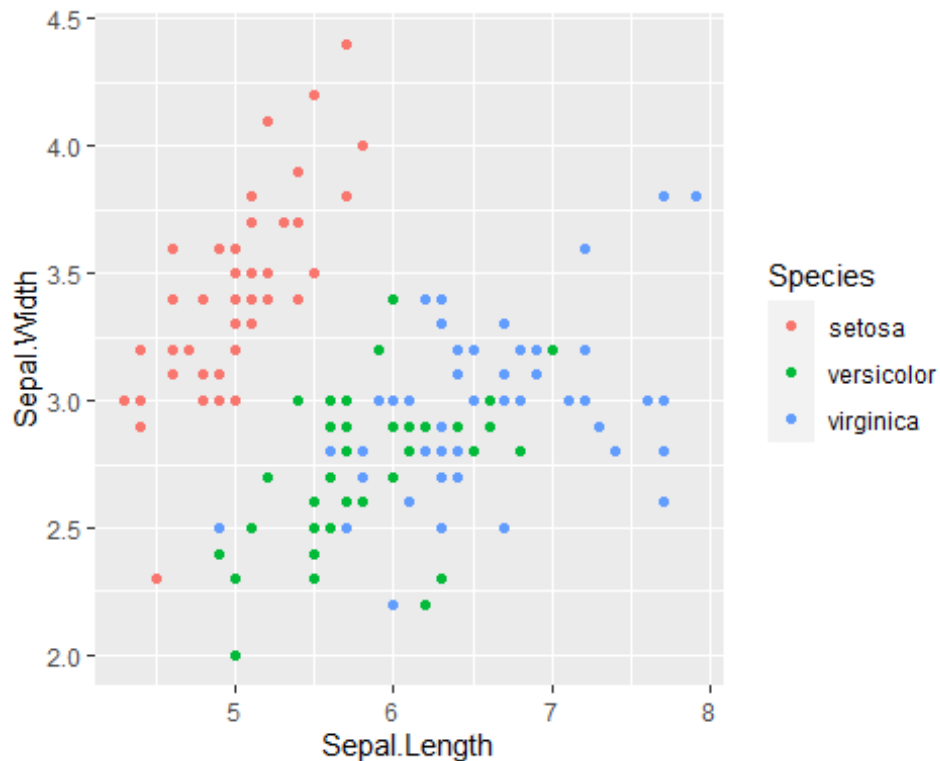




Correlation

For this analysis, the start would be to look at the paired sets of each measured portions of the plant (Sepal or Petal).

Does the length and width of the sepal have a strong correlation?

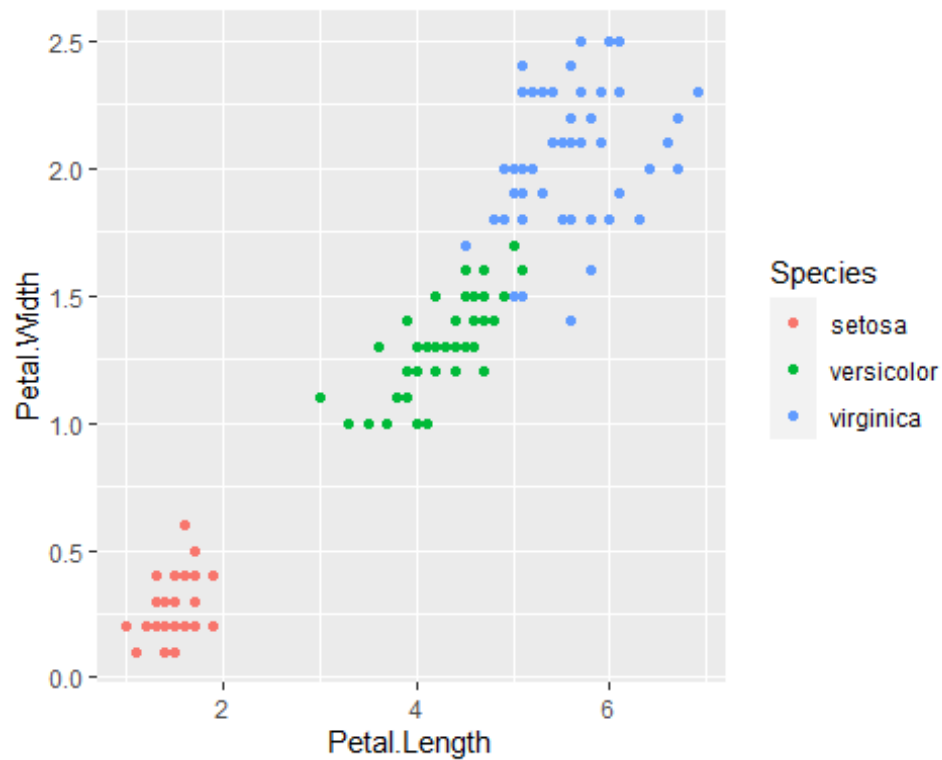


Correlation coefficient was calculated using R functions. The relationship between Sepal width and Sepal length have a weak negative relationship according to its r value.

```
## [1] -0.1175698
```

Does the length and width of the petal have a strong correlation?

An alternative relationship, Petal dimensions, was also analyzed. First the data was plotted.



Correlation coefficient was calculated using R functions. Upon initial analysis, the data for petal dimensions looked promising for a strong positive linear relationship and was further explored.

```
##      x  y data.Species    xy    x2    y2
## 1  1.4 0.2      setosa  0.28  1.96  0.04
## 2  1.4 0.2      setosa  0.28  1.96  0.04
## 3  1.3 0.2      setosa  0.26  1.69  0.04
## 4  1.5 0.2      setosa  0.30  2.25  0.04
## 5  1.4 0.2      setosa  0.28  1.96  0.04
## 6  1.7 0.4      setosa  0.68  2.89  0.16
## 7  1.4 0.3      setosa  0.42  1.96  0.09
## 8  1.5 0.2      setosa  0.30  2.25  0.04
## 9  1.4 0.2      setosa  0.28  1.96  0.04
## 10 1.5 0.1      setosa  0.15  2.25  0.01
## 11 1.5 0.2      setosa  0.30  2.25  0.04
## 12 1.6 0.2      setosa  0.32  2.56  0.04
## 13 1.4 0.1      setosa  0.14  1.96  0.01
## 14 1.1 0.1      setosa  0.11  1.21  0.01
## 15 1.2 0.2      setosa  0.24  1.44  0.04
## 16 1.5 0.4      setosa  0.60  2.25  0.16
## 17 1.3 0.4      setosa  0.52  1.69  0.16
## 18 1.4 0.3      setosa  0.42  1.96  0.09
```

## 19	1.7	0.3	setosa	0.51	2.89	0.09
## 20	1.5	0.3	setosa	0.45	2.25	0.09
## 21	1.7	0.2	setosa	0.34	2.89	0.04
## 22	1.5	0.4	setosa	0.60	2.25	0.16
## 23	1.0	0.2	setosa	0.20	1.00	0.04
## 24	1.7	0.5	setosa	0.85	2.89	0.25
## 25	1.9	0.2	setosa	0.38	3.61	0.04
## 26	1.6	0.2	setosa	0.32	2.56	0.04
## 27	1.6	0.4	setosa	0.64	2.56	0.16
## 28	1.5	0.2	setosa	0.30	2.25	0.04
## 29	1.4	0.2	setosa	0.28	1.96	0.04
## 30	1.6	0.2	setosa	0.32	2.56	0.04
## 31	1.6	0.2	setosa	0.32	2.56	0.04
## 32	1.5	0.4	setosa	0.60	2.25	0.16
## 33	1.5	0.1	setosa	0.15	2.25	0.01
## 34	1.4	0.2	setosa	0.28	1.96	0.04
## 35	1.5	0.2	setosa	0.30	2.25	0.04
## 36	1.2	0.2	setosa	0.24	1.44	0.04
## 37	1.3	0.2	setosa	0.26	1.69	0.04
## 38	1.4	0.1	setosa	0.14	1.96	0.01
## 39	1.3	0.2	setosa	0.26	1.69	0.04
## 40	1.5	0.2	setosa	0.30	2.25	0.04
## 41	1.3	0.3	setosa	0.39	1.69	0.09
## 42	1.3	0.3	setosa	0.39	1.69	0.09
## 43	1.3	0.2	setosa	0.26	1.69	0.04
## 44	1.6	0.6	setosa	0.96	2.56	0.36
## 45	1.9	0.4	setosa	0.76	3.61	0.16
## 46	1.4	0.3	setosa	0.42	1.96	0.09
## 47	1.6	0.2	setosa	0.32	2.56	0.04
## 48	1.4	0.2	setosa	0.28	1.96	0.04
## 49	1.5	0.2	setosa	0.30	2.25	0.04
## 50	1.4	0.2	setosa	0.28	1.96	0.04
## 51	4.7	1.4	versicolor	6.58	22.09	1.96
## 52	4.5	1.5	versicolor	6.75	20.25	2.25
## 53	4.9	1.5	versicolor	7.35	24.01	2.25
## 54	4.0	1.3	versicolor	5.20	16.00	1.69
## 55	4.6	1.5	versicolor	6.90	21.16	2.25
## 56	4.5	1.3	versicolor	5.85	20.25	1.69
## 57	4.7	1.6	versicolor	7.52	22.09	2.56
## 58	3.3	1.0	versicolor	3.30	10.89	1.00
## 59	4.6	1.3	versicolor	5.98	21.16	1.69
## 60	3.9	1.4	versicolor	5.46	15.21	1.96
## 61	3.5	1.0	versicolor	3.50	12.25	1.00
## 62	4.2	1.5	versicolor	6.30	17.64	2.25
## 63	4.0	1.0	versicolor	4.00	16.00	1.00
## 64	4.7	1.4	versicolor	6.58	22.09	1.96
## 65	3.6	1.3	versicolor	4.68	12.96	1.69
## 66	4.4	1.4	versicolor	6.16	19.36	1.96
## 67	4.5	1.5	versicolor	6.75	20.25	2.25
## 68	4.1	1.0	versicolor	4.10	16.81	1.00

## 69	4.5	1.5	versicolor	6.75	20.25	2.25
## 70	3.9	1.1	versicolor	4.29	15.21	1.21
## 71	4.8	1.8	versicolor	8.64	23.04	3.24
## 72	4.0	1.3	versicolor	5.20	16.00	1.69
## 73	4.9	1.5	versicolor	7.35	24.01	2.25
## 74	4.7	1.2	versicolor	5.64	22.09	1.44
## 75	4.3	1.3	versicolor	5.59	18.49	1.69
## 76	4.4	1.4	versicolor	6.16	19.36	1.96
## 77	4.8	1.4	versicolor	6.72	23.04	1.96
## 78	5.0	1.7	versicolor	8.50	25.00	2.89
## 79	4.5	1.5	versicolor	6.75	20.25	2.25
## 80	3.5	1.0	versicolor	3.50	12.25	1.00
## 81	3.8	1.1	versicolor	4.18	14.44	1.21
## 82	3.7	1.0	versicolor	3.70	13.69	1.00
## 83	3.9	1.2	versicolor	4.68	15.21	1.44
## 84	5.1	1.6	versicolor	8.16	26.01	2.56
## 85	4.5	1.5	versicolor	6.75	20.25	2.25
## 86	4.5	1.6	versicolor	7.20	20.25	2.56
## 87	4.7	1.5	versicolor	7.05	22.09	2.25
## 88	4.4	1.3	versicolor	5.72	19.36	1.69
## 89	4.1	1.3	versicolor	5.33	16.81	1.69
## 90	4.0	1.3	versicolor	5.20	16.00	1.69
## 91	4.4	1.2	versicolor	5.28	19.36	1.44
## 92	4.6	1.4	versicolor	6.44	21.16	1.96
## 93	4.0	1.2	versicolor	4.80	16.00	1.44
## 94	3.3	1.0	versicolor	3.30	10.89	1.00
## 95	4.2	1.3	versicolor	5.46	17.64	1.69
## 96	4.2	1.2	versicolor	5.04	17.64	1.44
## 97	4.2	1.3	versicolor	5.46	17.64	1.69
## 98	4.3	1.3	versicolor	5.59	18.49	1.69
## 99	3.0	1.1	versicolor	3.30	9.00	1.21
## 100	4.1	1.3	versicolor	5.33	16.81	1.69
## 101	6.0	2.5	virginica	15.00	36.00	6.25
## 102	5.1	1.9	virginica	9.69	26.01	3.61
## 103	5.9	2.1	virginica	12.39	34.81	4.41
## 104	5.6	1.8	virginica	10.08	31.36	3.24
## 105	5.8	2.2	virginica	12.76	33.64	4.84
## 106	6.6	2.1	virginica	13.86	43.56	4.41
## 107	4.5	1.7	virginica	7.65	20.25	2.89
## 108	6.3	1.8	virginica	11.34	39.69	3.24
## 109	5.8	1.8	virginica	10.44	33.64	3.24
## 110	6.1	2.5	virginica	15.25	37.21	6.25
## 111	5.1	2.0	virginica	10.20	26.01	4.00
## 112	5.3	1.9	virginica	10.07	28.09	3.61
## 113	5.5	2.1	virginica	11.55	30.25	4.41
## 114	5.0	2.0	virginica	10.00	25.00	4.00
## 115	5.1	2.4	virginica	12.24	26.01	5.76
## 116	5.3	2.3	virginica	12.19	28.09	5.29
## 117	5.5	1.8	virginica	9.90	30.25	3.24
## 118	6.7	2.2	virginica	14.74	44.89	4.84

```

## 119 6.9 2.3 virginica 15.87 47.61 5.29
## 120 5.0 1.5 virginica 7.50 25.00 2.25
## 121 5.7 2.3 virginica 13.11 32.49 5.29
## 122 4.9 2.0 virginica 9.80 24.01 4.00
## 123 6.7 2.0 virginica 13.40 44.89 4.00
## 124 4.9 1.8 virginica 8.82 24.01 3.24
## 125 5.7 2.1 virginica 11.97 32.49 4.41
## 126 6.0 1.8 virginica 10.80 36.00 3.24
## 127 4.8 1.8 virginica 8.64 23.04 3.24
## 128 4.9 1.8 virginica 8.82 24.01 3.24
## 129 5.6 2.1 virginica 11.76 31.36 4.41
## 130 5.8 1.6 virginica 9.28 33.64 2.56
## 131 6.1 1.9 virginica 11.59 37.21 3.61
## 132 6.4 2.0 virginica 12.80 40.96 4.00
## 133 5.6 2.2 virginica 12.32 31.36 4.84
## 134 5.1 1.5 virginica 7.65 26.01 2.25
## 135 5.6 1.4 virginica 7.84 31.36 1.96
## 136 6.1 2.3 virginica 14.03 37.21 5.29
## 137 5.6 2.4 virginica 13.44 31.36 5.76
## 138 5.5 1.8 virginica 9.90 30.25 3.24
## 139 4.8 1.8 virginica 8.64 23.04 3.24
## 140 5.4 2.1 virginica 11.34 29.16 4.41
## 141 5.6 2.4 virginica 13.44 31.36 5.76
## 142 5.1 2.3 virginica 11.73 26.01 5.29
## 143 5.1 1.9 virginica 9.69 26.01 3.61
## 144 5.9 2.3 virginica 13.57 34.81 5.29
## 145 5.7 2.5 virginica 14.25 32.49 6.25
## 146 5.2 2.3 virginica 11.96 27.04 5.29
## 147 5.0 1.9 virginica 9.50 25.00 3.61
## 148 5.2 2.0 virginica 10.40 27.04 4.00
## 149 5.4 2.3 virginica 12.42 29.16 5.29
## 150 5.1 1.8 virginica 9.18 26.01 3.24

```

```

##      x  y data.Species
## 1  1.4 0.2      setosa
## 2  1.4 0.2      setosa
## 3  1.3 0.2      setosa
## 4  1.5 0.2      setosa
## 5  1.4 0.2      setosa
## 6  1.7 0.4      setosa
## 7  1.4 0.3      setosa
## 8  1.5 0.2      setosa
## 9  1.4 0.2      setosa
## 10 1.5 0.1      setosa
## 11 1.5 0.2      setosa
## 12 1.6 0.2      setosa
## 13 1.4 0.1      setosa
## 14 1.1 0.1      setosa
## 15 1.2 0.2      setosa
## 16 1.5 0.4      setosa

```


## 17	1.3	0.4	setosa
## 18	1.4	0.3	setosa
## 19	1.7	0.3	setosa
## 20	1.5	0.3	setosa
## 21	1.7	0.2	setosa
## 22	1.5	0.4	setosa
## 23	1.0	0.2	setosa
## 24	1.7	0.5	setosa
## 25	1.9	0.2	setosa
## 26	1.6	0.2	setosa
## 27	1.6	0.4	setosa
## 28	1.5	0.2	setosa
## 29	1.4	0.2	setosa
## 30	1.6	0.2	setosa
## 31	1.6	0.2	setosa
## 32	1.5	0.4	setosa
## 33	1.5	0.1	setosa
## 34	1.4	0.2	setosa
## 35	1.5	0.2	setosa
## 36	1.2	0.2	setosa
## 37	1.3	0.2	setosa
## 38	1.4	0.1	setosa
## 39	1.3	0.2	setosa
## 40	1.5	0.2	setosa
## 41	1.3	0.3	setosa
## 42	1.3	0.3	setosa
## 43	1.3	0.2	setosa
## 44	1.6	0.6	setosa
## 45	1.9	0.4	setosa
## 46	1.4	0.3	setosa
## 47	1.6	0.2	setosa
## 48	1.4	0.2	setosa
## 49	1.5	0.2	setosa
## 50	1.4	0.2	setosa
## 51	4.7	1.4	versicolor
## 52	4.5	1.5	versicolor
## 53	4.9	1.5	versicolor
## 54	4.0	1.3	versicolor
## 55	4.6	1.5	versicolor
## 56	4.5	1.3	versicolor
## 57	4.7	1.6	versicolor
## 58	3.3	1.0	versicolor
## 59	4.6	1.3	versicolor
## 60	3.9	1.4	versicolor
## 61	3.5	1.0	versicolor
## 62	4.2	1.5	versicolor
## 63	4.0	1.0	versicolor
## 64	4.7	1.4	versicolor
## 65	3.6	1.3	versicolor
## 66	4.4	1.4	versicolor

## 67	4.5	1.5	versicolor
## 68	4.1	1.0	versicolor
## 69	4.5	1.5	versicolor
## 70	3.9	1.1	versicolor
## 71	4.8	1.8	versicolor
## 72	4.0	1.3	versicolor
## 73	4.9	1.5	versicolor
## 74	4.7	1.2	versicolor
## 75	4.3	1.3	versicolor
## 76	4.4	1.4	versicolor
## 77	4.8	1.4	versicolor
## 78	5.0	1.7	versicolor
## 79	4.5	1.5	versicolor
## 80	3.5	1.0	versicolor
## 81	3.8	1.1	versicolor
## 82	3.7	1.0	versicolor
## 83	3.9	1.2	versicolor
## 84	5.1	1.6	versicolor
## 85	4.5	1.5	versicolor
## 86	4.5	1.6	versicolor
## 87	4.7	1.5	versicolor
## 88	4.4	1.3	versicolor
## 89	4.1	1.3	versicolor
## 90	4.0	1.3	versicolor
## 91	4.4	1.2	versicolor
## 92	4.6	1.4	versicolor
## 93	4.0	1.2	versicolor
## 94	3.3	1.0	versicolor
## 95	4.2	1.3	versicolor
## 96	4.2	1.2	versicolor
## 97	4.2	1.3	versicolor
## 98	4.3	1.3	versicolor
## 99	3.0	1.1	versicolor
## 100	4.1	1.3	versicolor
## 101	6.0	2.5	virginica
## 102	5.1	1.9	virginica
## 103	5.9	2.1	virginica
## 104	5.6	1.8	virginica
## 105	5.8	2.2	virginica
## 106	6.6	2.1	virginica
## 107	4.5	1.7	virginica
## 108	6.3	1.8	virginica
## 109	5.8	1.8	virginica
## 110	6.1	2.5	virginica
## 111	5.1	2.0	virginica
## 112	5.3	1.9	virginica
## 113	5.5	2.1	virginica
## 114	5.0	2.0	virginica
## 115	5.1	2.4	virginica
## 116	5.3	2.3	virginica

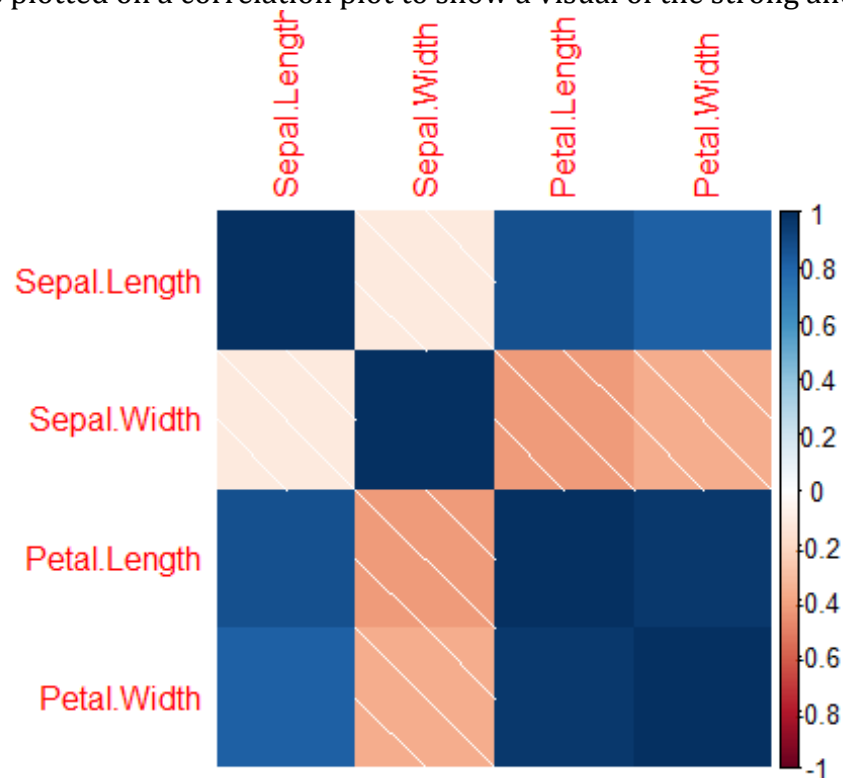
```
## 117 5.5 1.8 virginica
## 118 6.7 2.2 virginica
## 119 6.9 2.3 virginica
## 120 5.0 1.5 virginica
## 121 5.7 2.3 virginica
## 122 4.9 2.0 virginica
## 123 6.7 2.0 virginica
## 124 4.9 1.8 virginica
## 125 5.7 2.1 virginica
## 126 6.0 1.8 virginica
## 127 4.8 1.8 virginica
## 128 4.9 1.8 virginica
## 129 5.6 2.1 virginica
## 130 5.8 1.6 virginica
## 131 6.1 1.9 virginica
## 132 6.4 2.0 virginica
## 133 5.6 2.2 virginica
## 134 5.1 1.5 virginica
## 135 5.6 1.4 virginica
## 136 6.1 2.3 virginica
## 137 5.6 2.4 virginica
## 138 5.5 1.8 virginica
## 139 4.8 1.8 virginica
## 140 5.4 2.1 virginica
## 141 5.6 2.4 virginica
## 142 5.1 2.3 virginica
## 143 5.1 1.9 virginica
## 144 5.9 2.3 virginica
## 145 5.7 2.5 virginica
## 146 5.2 2.3 virginica
## 147 5.0 1.9 virginica
## 148 5.2 2.0 virginica
## 149 5.4 2.3 virginica
## 150 5.1 1.8 virginica

## [1] 0.9628654
```

All numerical relationships were explored and combined into a data table. Under 5 variables were explored for the purpose of analysis and data presentation. More than 5 variables become visually flawed and create complexities that mask the messages trying to be identified

##	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
## Sepal.Length	1.0000000	-0.1175698	0.8717538	0.8179411
## Sepal.Width	-0.1175698	1.0000000	-0.4284401	-0.3661259
## Petal.Length	0.8717538	-0.4284401	1.0000000	0.9628654
## Petal.Width	0.8179411	-0.3661259	0.9628654	1.0000000

Those correlations were plotted on a correlation plot to show a visual of the strong and



weak relationships.

Regression

The petal dimensions were further explored due to the high r value calculated through the correlation analysis. The significance of the relationship needed to be analyzed. This was tested at an significance of 0.05.

```
alpha <- 0.05
null <- "rho = 0"
alt <- "rho neq 0"
#determine significance
cor.test(x=x, y=y,
         alternative="two.sided",
         method="pearson",
         conf.level =0.95)

##
## Pearson's product-moment correlation
##
## data: x and y
## t = 43.387, df = 148, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.9490525 0.9729853
## sample estimates:
```

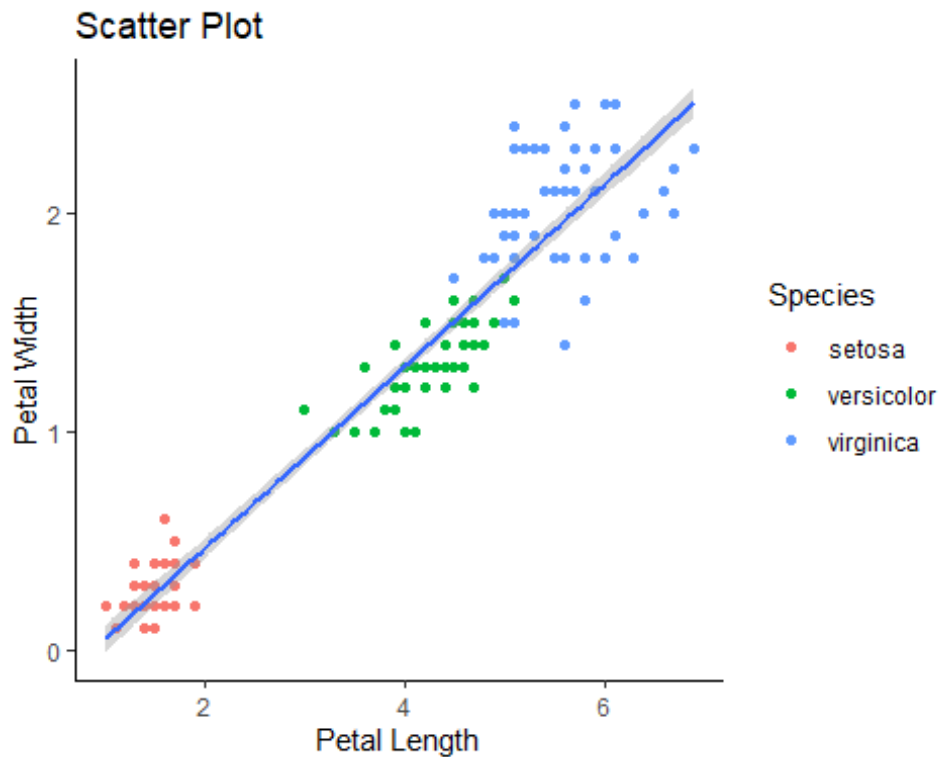
```
##      cor
## 0.9628654
```

The analysis resulted in the relationship being significant and reject the null hypothesis.

Linear Equation

The linear equation coefficients was calculated and plotted against the data.

```
##
## Call:
## lm(formula = Petal.Width ~ Petal.Length, data = data)
##
## Coefficients:
## (Intercept)  Petal.Length
##      -0.3631      0.4158
##
## `geom_smooth()` using formula 'y ~ x'
```

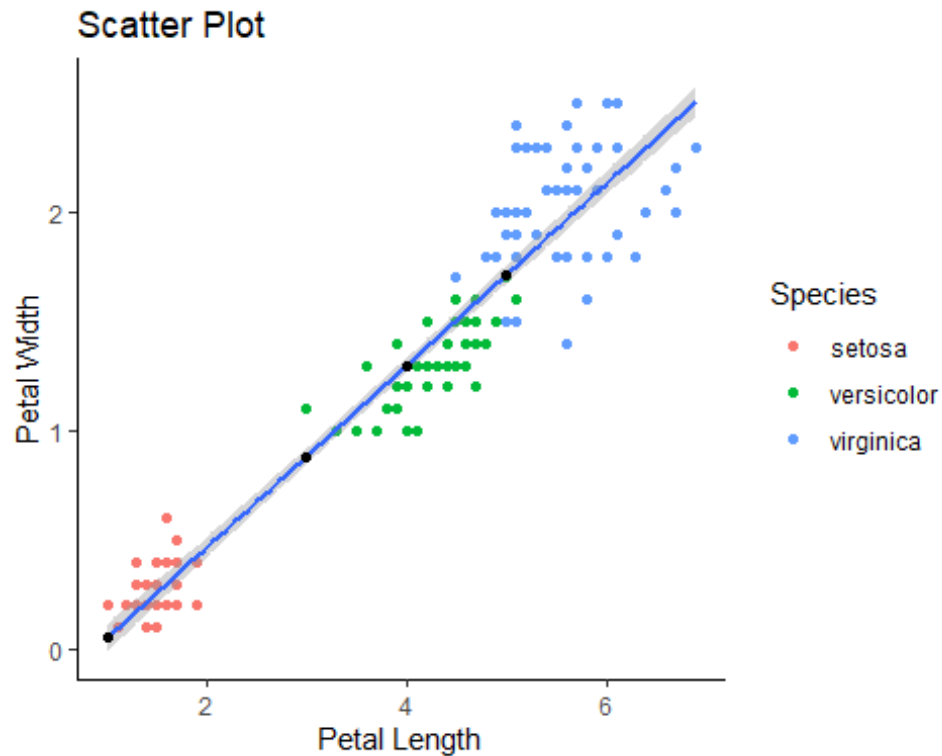


Values were identified and selected to predict the width of the petal given a specific length. These were combined into a data table.

```
##   Petal.Length Petal.Width
## 1             1  0.0526799
## 2             3  0.8841907
## 3             4  1.2999461
## 4             5  1.7157016
```

The final points were overlayed to show where the points would be calculated given the existing data.

```
## `geom_smooth()` using formula 'y ~ x'
```



Conclusion

The regression analysis is a useful tool when it comes to predicting data that doesn't exist or is partially collected. Both correlations and regression analyses utilize existing relationships in the data, however, the regression needs the data to be strongly related in order to predict with some degree of accuracy what the data may look like in the future. Correlation just statistically describes the relationship of two variables in a dataset.