La Plateforme
La grande école du numérique pour tous

Botanica iris

Derrière chaque pétale, une vérité statistique.

EXPLORATORY DATA ANALYSIS

Presented by: Redha & Rooney

10/06/2025

Supervisor: AKRAM

Outline

1. Project Objective & Approach
2. Data Structure & Preparation
3. Descriptive Statistics
4. Correlation & Visualization
5. Outlier Detection
6. Conclusion & Insights

1. Project Objective & Approach

Objective: Analyze the Iris dataset to understand the structure, relationships, and differences between three species of flowers.

Approach: Use R for descriptive statistics, visualization, and detection of outliers/patterns.

2. Data Structure & Preparation

The dataset contains :

150 flowers from three species (setosa, versicolor, virginica).

Variables: Sepal.Length, Sepal.Width, Petal.Length, Petal.Width

(all numeric), plus Species (category).

No missing or duplicate values detected.

Distribution of species:50 Setosa, 50 Versicolor, 50 Virginica

3. Descriptive Statistics

Means, Medians, Quartiles, Ranges, and Standard Deviations for each species and variable.

```
Descriptive statistics by species:
> print(stats_long)
# A tibble: 12 × 10
   Species    Variable      mean    sd median   min   max    Q1    Q3   IQR
   <fct>      <chr>        <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
 1 setosa     Petal.Length 1.46  0.174   1.5    1     1.9  1.4   1.58 0.175
 2 versicolor Petal.Length 4.26  0.470   4.35   3     5.1  4     4.6  0.600
 3 virginica  Petal.Length 5.55  0.552   5.55   4.5   6.9  5.1   5.88 0.775
 4 setosa     Petal.Width  0.246 0.105   0.2    0.1   0.6  0.2   0.3  0.1
 5 versicolor Petal.Width  1.33  0.198   1.3    1     1.8  1.2   1.5  0.3
 6 virginica  Petal.Width  2.03  0.275   2      1.4   2.5  1.8   2.3  0.5
 7 setosa     Sepal.Length 5.01  0.352   5      4.3   5.8  4.8   5.2  0.400
 8 versicolor Sepal.Length 5.94  0.516   5.9    4.9   7    5.6   6.3  0.7
 9 virginica  Sepal.Length 6.59  0.636   6.5    4.9   7.9  6.22  6.9  0.675
10 setosa     Sepal.Width  3.43  0.379   3.4    2.3   4.4  3.2   3.68 0.475
11 versicolor Sepal.Width  2.77  0.314   2.8    2     3.4  2.52  3    0.475
12 virginica  Sepal.Width  2.97  0.322   3      2.2   3.8  2.8   3.18 0.375
>
```

3. Descriptive Statistics

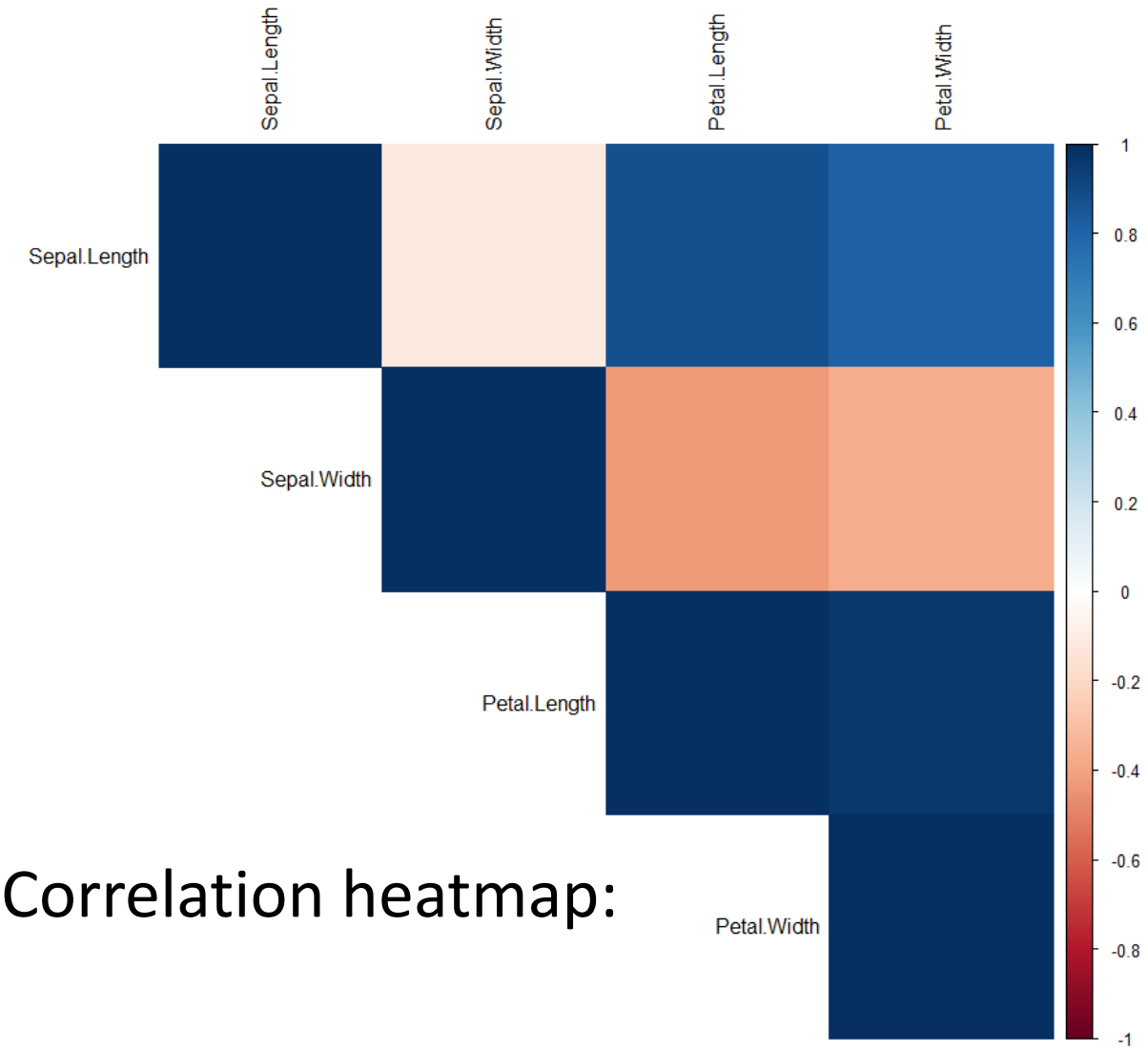
Means, Medians, Quartiles, Ranges, and Standard Deviations for each species and variable.

Variable	mean	sd	median	min	max	Q1	Q3	IQR
<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
Sepal.Length	5.84	0.828	5.8	4.3	7.9	5.1	6.4	1.3
Sepal.Width	3.06	0.436	3	2	4.4	2.8	3.3	0.5
Petal.Length	3.76	1.77	4.35	1	6.9	1.6	5.1	3.5
Petal.Width	1.20	0.762	1.3	0.1	2.5	0.3	1.8	1.5

4. **Correlation** & Visualization

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
Sepal.Length	1.00	-0.12	0.87	0.82
Sepal.Width	-0.12	1.00	-0.43	-0.37
Petal.Length	0.87	-0.43	1.00	0.96
Petal.Width	0.82	-0.37	0.96	1.00

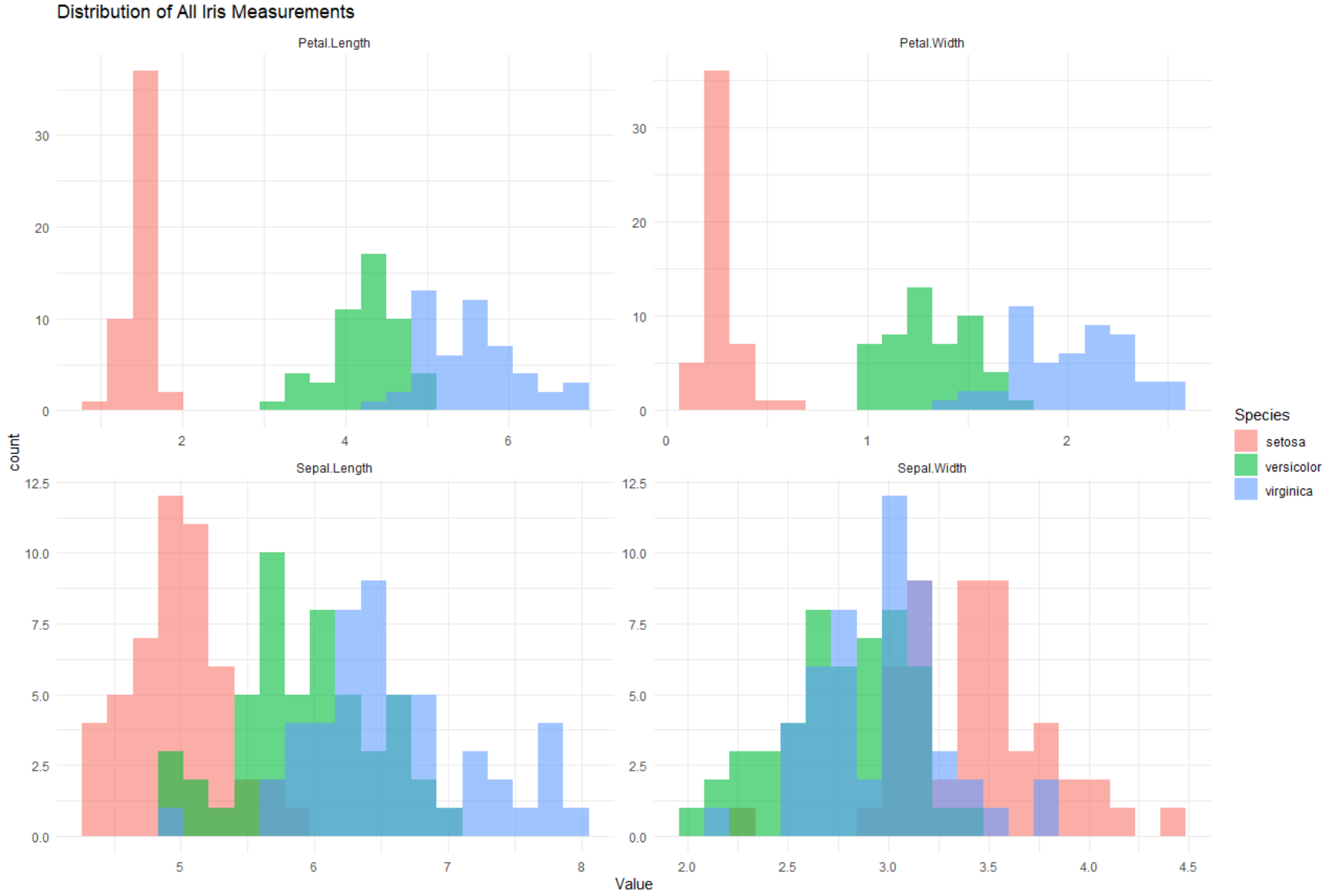
4. **Correlation** & Visualization



Correlation heatmap:

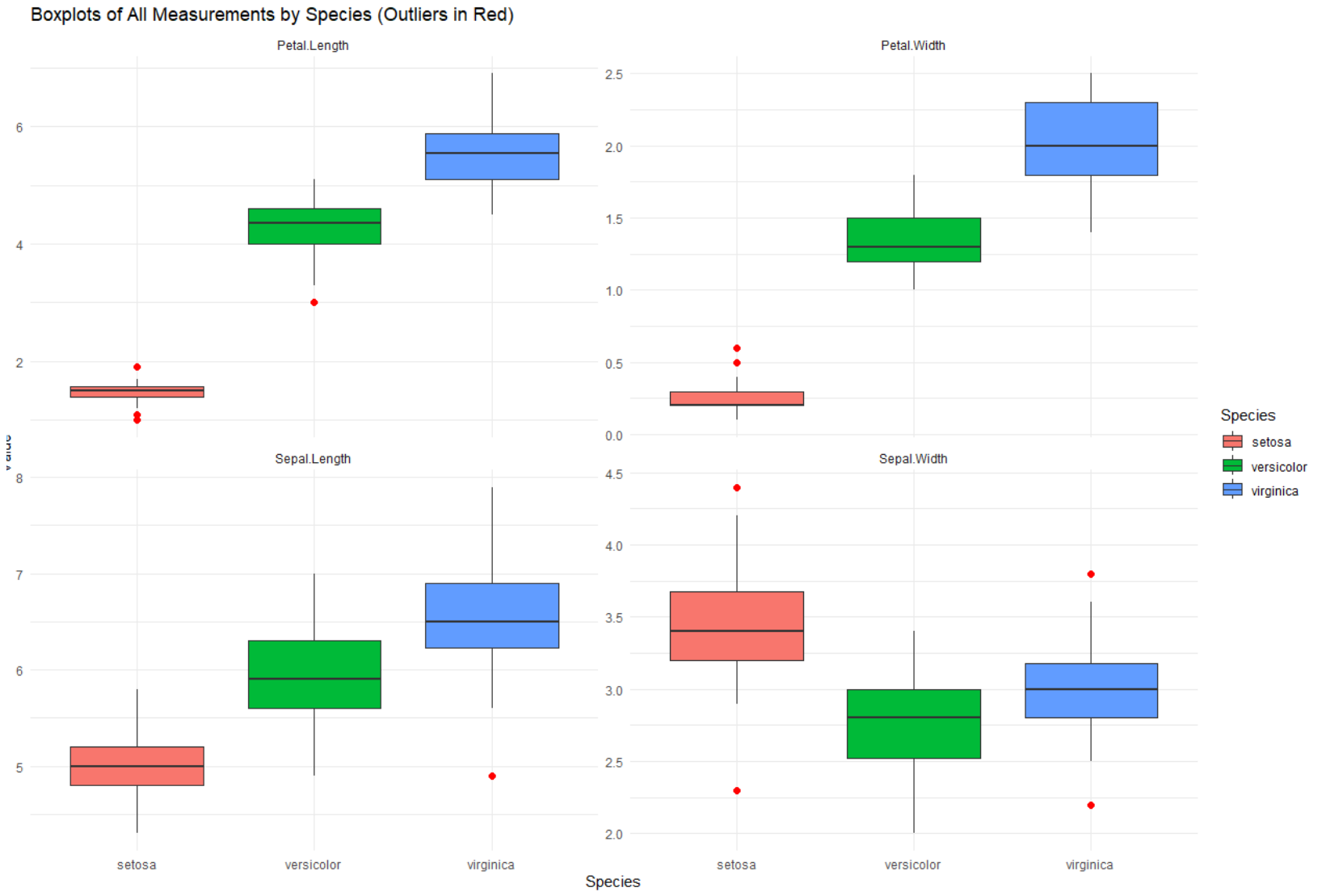
Correlation matrix:
Interpretation:
Petal.Length and Petal.Width are strongly correlated (r ≈ 0.96).

4. **Correlation** & Visualization



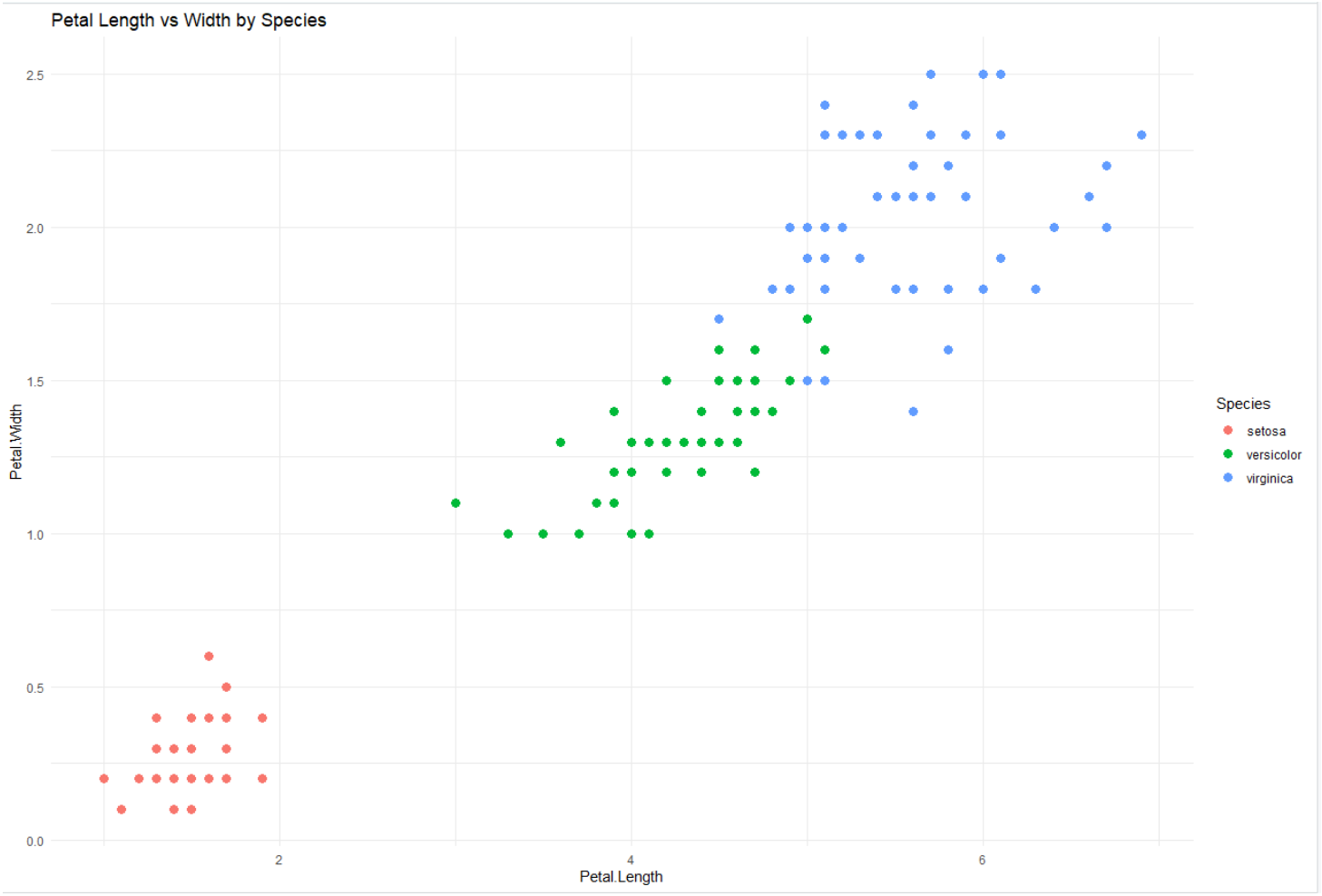
a. Histograms (All Variables, by Species)

4. **Correlation** & Visualization



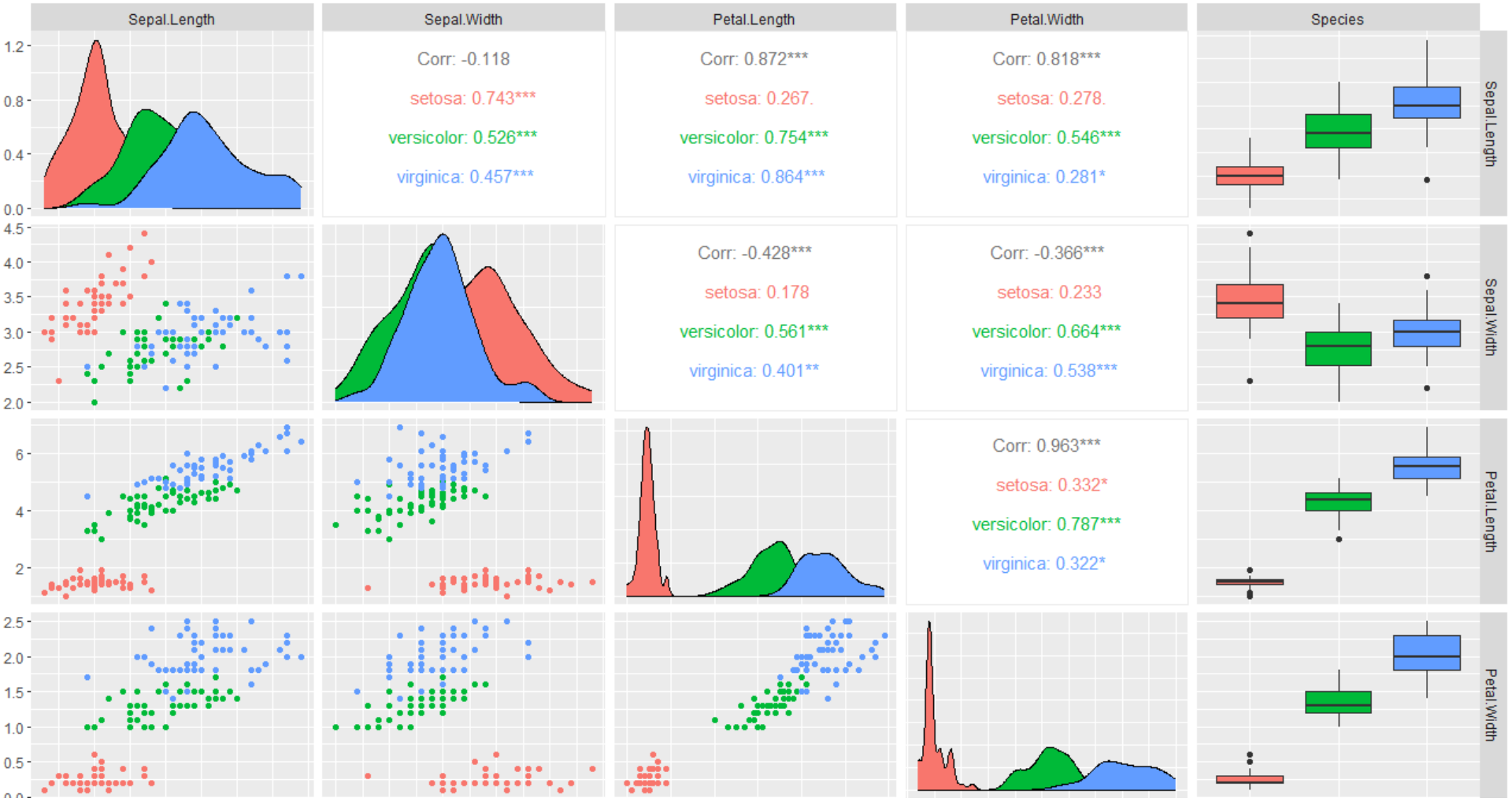
b. Boxplots (All Variables, by Species)

4. **Correlation** & Visualization



c. Scatter Plot: Petal.Length vs Petal.Width

4. **Correlation** & Visualization



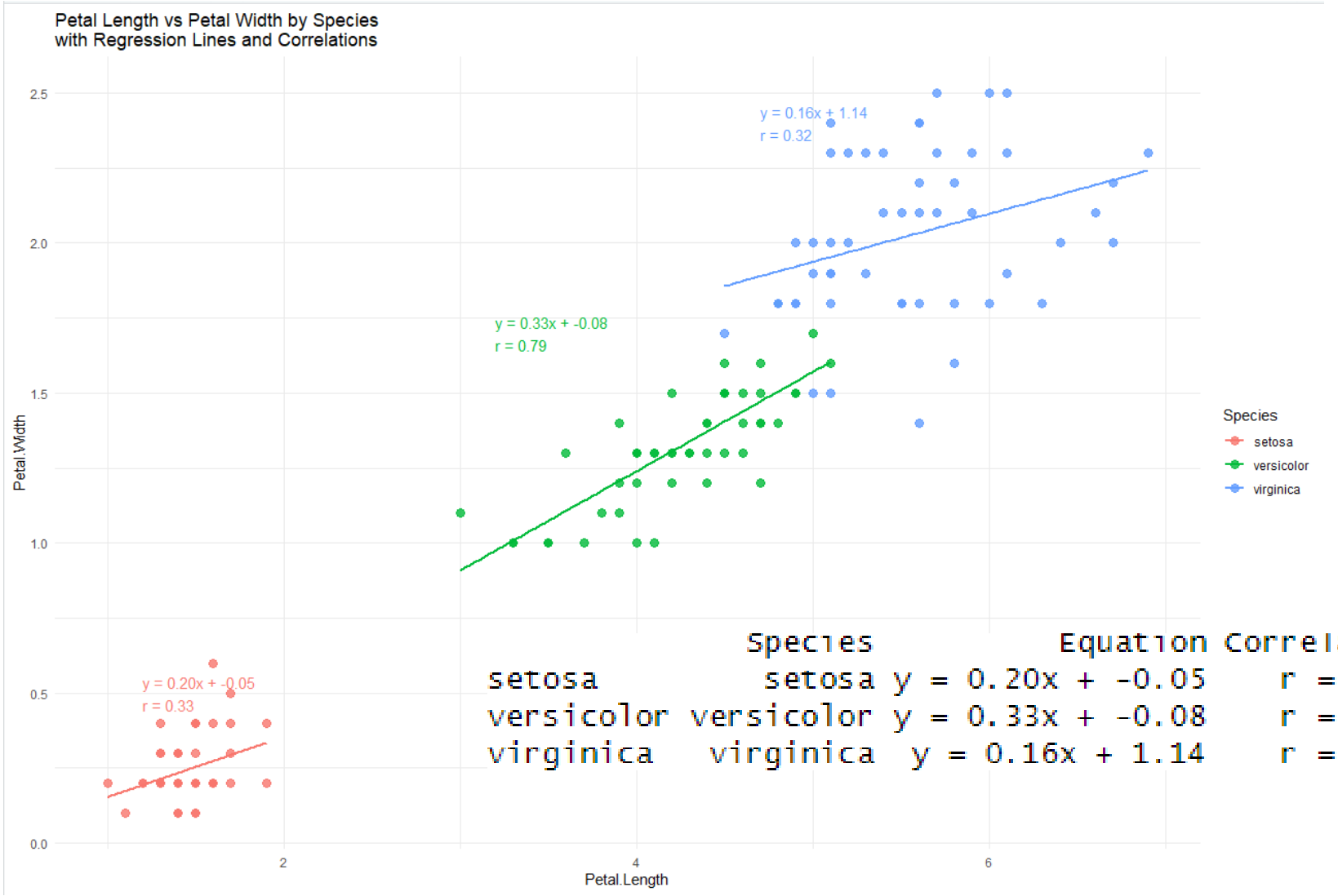
d. Pairwise Plots (GGally)

6. Outlier Detection

<fct>	<int>	<int>
setosa	0	2
versicolor	0	0
virginica	1	3

Outlier rule: Points outside 1.5*IQR from the first/third quartile.

7. Can We Predict Species?



Species	Equation	Correlation
setosa	setosa y = 0.20x + -0.05	r = 0.33
versicolor	versicolor y = 0.33x + -0.08	r = 0.79
virginica	virginica y = 0.16x + 1.14	r = 0.32

8. Conclusions

Findings:
Data is clean, with few outliers.
Petal measurements are highly informative.
Setosa is easily distinguishable.
Some overlap between Versicolor and Virginica, but overall good separation.
With R2 0.79 versicolor is the most predictible
Recommendation:
Petal.Width and Petal.Length are the best predictors for species.