2024 Project Template University of Tirana

Master in Mathematics and Informatics Engineering

Eralda Gjika Dhamo

01-03-2024

Contents

1 Before we start

- Before we start here is the Eralda Gjika **LinkedIn profile** if you want to learn more and get in touch!
- **Link to Eralda Gjika's Github profile** where you may find the R Markdown template of this PDF and many more projects.
- Project **Repository**

2 Introduction to R Markdown for Reporting

R Markdown is a powerful tool that allows you to integrate R code with narrative text, creating dynamic and reproducible documents. With R Markdown, you can seamlessly combine code, results, and text in a single document, making it ideal for generating reports, presentations, articles, and more.

2.1 Getting Started with R Markdown:

1. **Installation**: Ensure that you have R and RStudio installed on your computer. RStudio provides an integrated development environment (IDE) for R, making it easier to work with R Markdown documents.
2. **Creating an R Markdown Document**: In RStudio, go to `File` -> `New File` -> `R Markdown`. You will be prompted to choose a document format; for PDF reports, select `PDF`.
3. **Document Setup**: Once you've selected the document format, you can specify the title, author, and other metadata for your document.
4. **Writing Content**: R Markdown documents use a simple and intuitive syntax for formatting text. You can write narrative text using Markdown syntax, which supports headings, lists, links, and more. Code chunks can be embedded within the document using triple backticks (“‘), allowing you to execute R code directly within the document.
5. **Code Execution**: Code chunks in R Markdown documents are executed sequentially, and the results are displayed inline within the document. You can specify options such as code chunk labels, chunk evaluation options, and chunk output options.
6. **Knitting**: Once you've written your R Markdown document, you can knit it into the desired output format (e.g., PDF) by clicking the `Knit` button in RStudio. This process executes the R code chunks, generates the output, and renders the final document.

2.2 Generating PDF Reports:

1. **Customizing Output**: R Markdown allows you to customize the appearance of your PDF report by specifying document options, including page layout, font size, and margins.
2. **Graphics**: You can include plots and graphics generated from R code directly within your PDF report. R Markdown supports various plotting libraries, including base R graphics, ggplot2, and lattice.
3. **Tables**: Tabular data can be displayed as tables in your PDF report. R Markdown provides flexible options for creating and formatting tables, including data frames, knitr tables, and kableExtra.
4. **Exporting**: Once your R Markdown document is knitted into a PDF report, you can export it as a standalone PDF file for distribution or sharing with others.

3 Tables

3.1 Dealing with tables

Iris is a very known dataset used for many purpose in R. In this template we will use that dataset for data analysis and visualizations. Do you know how easy is to show an image in your report? Just take care of saving it and then call it as in the code.

The following table describes the distribution of our data.We are using **kable** for tables.

Figure 1: Iris Flower

Table 1: Species distribution

Var1	Freq
setosa	50
versicolor	50
virginica	50

4 Individual characteristics

4.1 Number and Description

In this section we will see how to incorporate calculations within text. In R Markdown, you can seamlessly incorporate inline calculations within your text by enclosing the expression within backticks which will render as a text when knitted.

Reporting counts in text:

- **50** Number of individuals for 'setosa' species
- **113** Number of individuals with Petal.Length >= 1.6
- **0** Number of individuals with Sepal.Width >= 4.9

Note: The information contained on this page is for instructional purpose only. You may use it and update based on your needs or knowledge.

4.2 Descriptive Statistics

Table 2: Descriptive statistics

Variable	Statistic	Value
Sepal	Length_mean	5.8433333
Sepal	Length_median	5.8000000
Sepal	Length_sd	0.8280661
Sepal	Length_min	4.3000000
Sepal	Length_max	7.9000000
Sepal	Width_mean	3.0573333
Sepal	Width_median	3.0000000
Sepal	Width_sd	0.4358663
Sepal	Width_min	2.0000000
Sepal	Width_max	4.4000000
Petal	Length_mean	3.7580000
Petal	Length_median	4.3500000
Petal	Length_sd	1.7652982
Petal	Length_min	1.0000000
Petal	Length_max	6.9000000
Petal	Width_mean	1.1993333
Petal	Width_median	1.3000000
Petal	Width_sd	0.7622377
Petal	Width_min	0.1000000
Petal	Width_max	2.5000000

5 Visualizations

5.1 Histogram and Barchart

Below are the histograms for numerical variables and barchart for categorical variable showing the frequencies for each variable. For Figure numbering and title we make sure we write in the chunk options for example: fig.cap="Iris data visualizations".

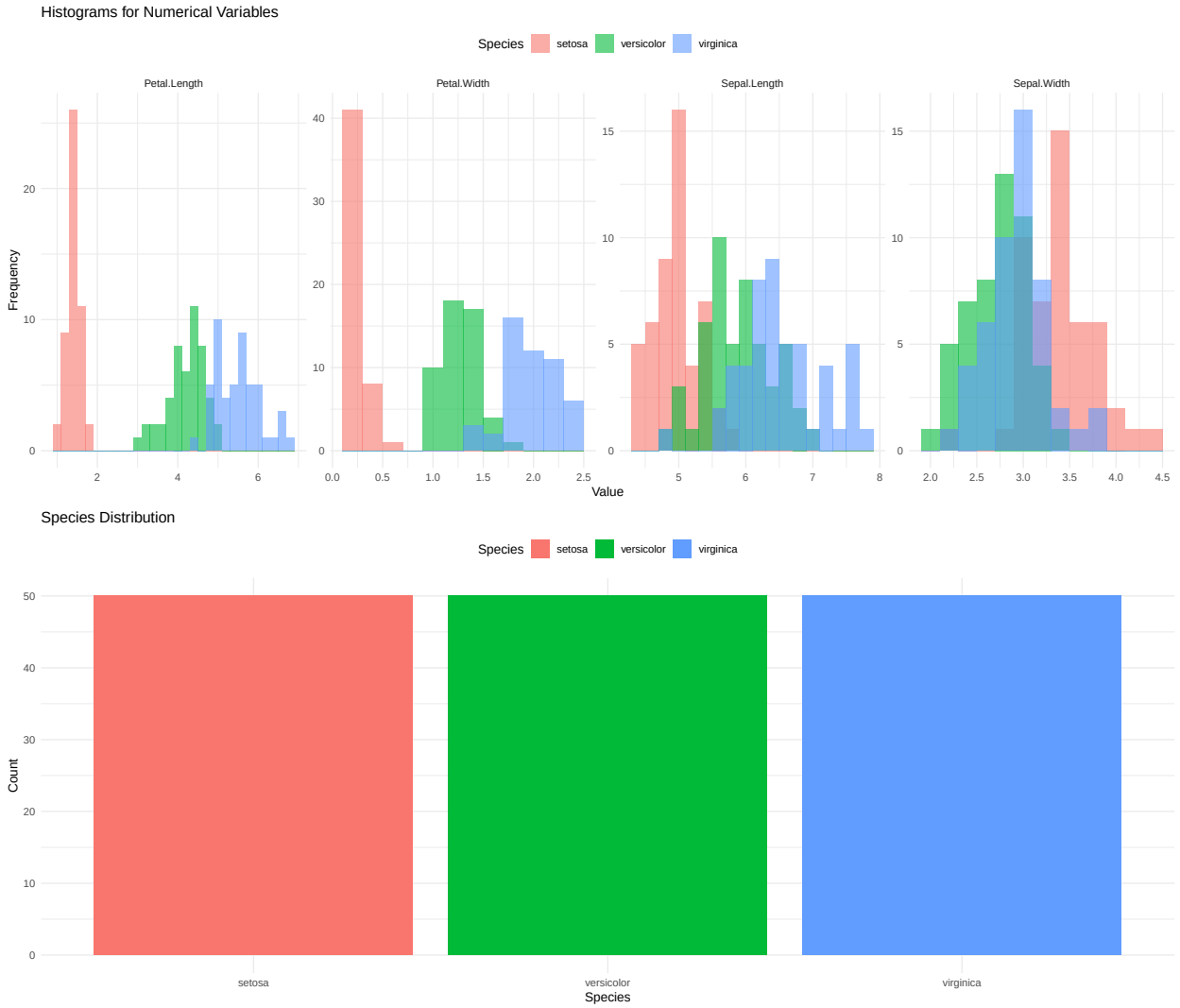

Figure 2: Iris data visualizations

6 Correlations

This code will create a grid of scatterplots for each pair of numeric variables in the Iris dataset along with histograms for each variable's distribution on the diagonal, and correlation coefficients in the upper and lower triangles of the grid. This provides a comprehensive visualization of the relationships and correlations between the numeric variables.

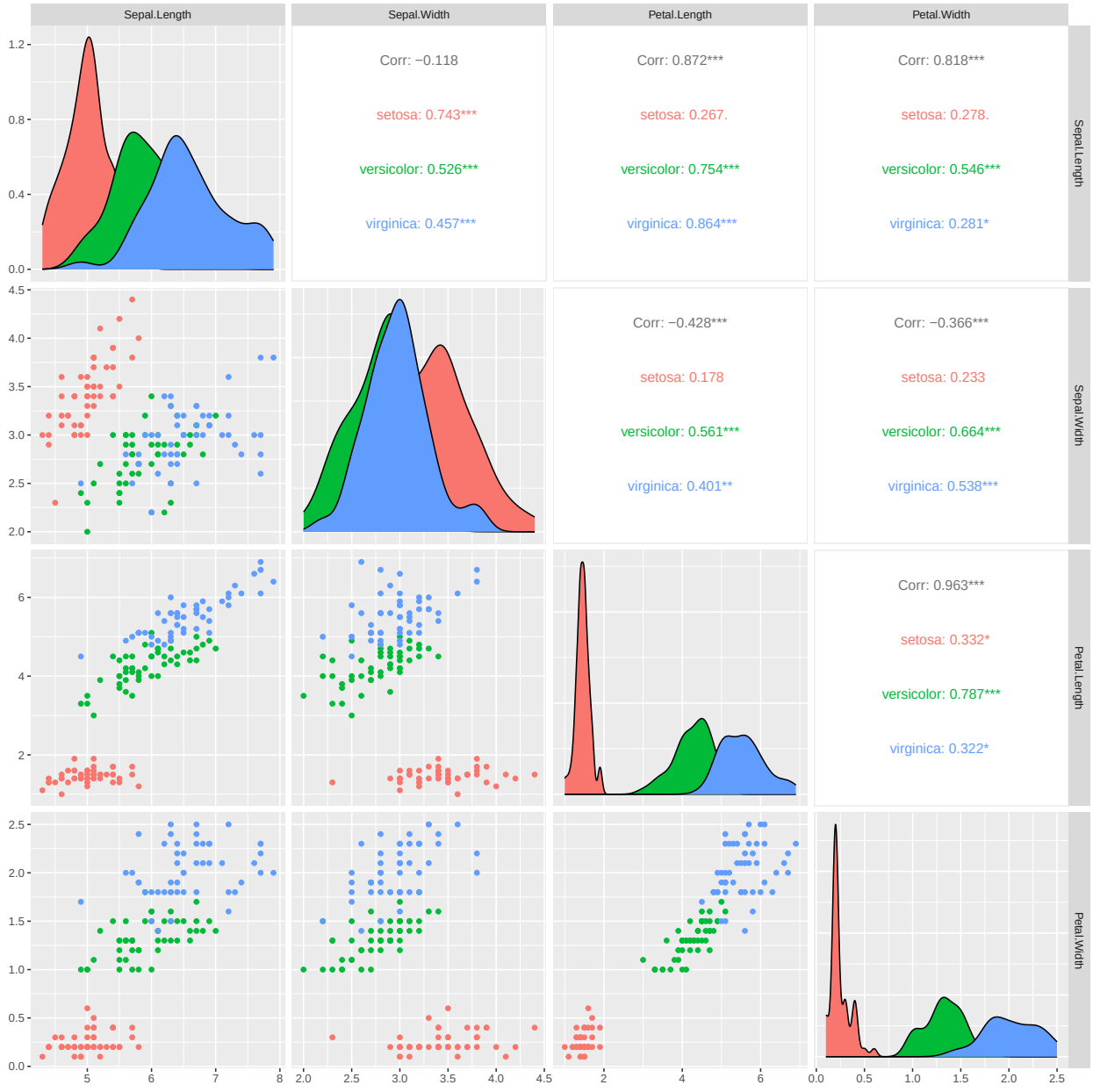

Figure 3: Correlation plot using GGally library

7 Conclusions

Based on the correlation plot generated using the GGally package with the iris dataset, we can draw several conclusions:

1. **Positive Correlation between Variables**:
   - In the scatterplot matrix, we observe positive correlations between some pairs of variables. This is indicated by the upward trend in the scatterplots, where an increase in one variable tends to be associated with an increase in the other variable.
2. **Negative Correlation between Variables**:
   - Conversely, we also notice negative correlations between certain pairs of variables, depicted by a downward trend in the scatterplots. In these cases, an increase in one variable is typically associated with a decrease in the other variable.
3. **Strength of Correlation**:
   - The strength of correlation varies across different pairs of variables. Some pairs exhibit strong correlations, where points cluster tightly around the trend line, indicating a strong linear relationship. In contrast, other pairs show weaker correlations, with points scattered more widely around the trend line.
4. **Outliers and Data Distribution**:
   - We identify outliers in the data, which are points that deviate significantly from the general trend observed in the scatterplots. These outliers could potentially represent anomalous data points or errors in measurement.
   - The histograms on the diagonal of the scatterplot matrix provide insights into the distribution of each individual variable. We observe the shape, spread, and skewness of the distributions, which can help in understanding the data's characteristics.
5. **Species Differentiation**:
   - By coloring the data points based on the species variable, we can discern patterns specific to each species. This facilitates the identification of any species-specific trends or clusters in the data.
6. **Further Analysis**:
   - While the correlation plot provides valuable insights into the relationships between variables, further analysis may be warranted to explore causality or to build predictive models based on these relationships. Techniques such as regression analysis or machine learning algorithms can be employed for deeper exploration and modeling.

8 Summary/Recommendations

When creating an RMarkdown document for reporting purposes, consider the following recommendations to ensure clarity, reproducibility, and professionalism:

1. **Start with a Clear Objective**: Define the purpose of your report and the key questions or insights you aim to address or provide.
2. **Organize Content**: Structure your report logically with clear headings, sections, and sub-sections. This helps readers navigate through the document efficiently.
3. **Include an Introduction**: Provide context for the analysis, briefly outlining the background, objectives, and scope of the report.
4. **Document Your Analysis**: Use code chunks to showcase your R code along with the outputs. This ensures transparency and reproducibility of your analysis.
5. **Use Descriptive Text**: Accompany your code and outputs with descriptive text to explain the analysis, interpretations, and insights. Avoid assuming that readers will understand the code or outputs without explanation.
6. **Visualize Data Effectively**: Utilize appropriate charts, graphs, and tables to represent your data visually. Choose visualization types that effectively convey your message and insights.
7. **Customize Visuals**: Customize visualizations to enhance clarity and aesthetics. This includes labeling axes, adding titles, legends, and color schemes that are consistent and easy to interpret.
8. **Add Interpretation and Insights**: Offer interpretations of the analysis results and provide insights into the implications of the findings. Relate the analysis back to the objectives stated in the introduction.
9. **Include Summary and Conclusion**: Summarize the key findings and conclusions derived from the analysis. Restate the main objectives and discuss how they were addressed.
10. **Provide Recommendations**: If applicable, offer recommendations based on the analysis results. These recommendations should be actionable and relevant to the objectives of the report.
11. **Include References and Citations**: If you've used external sources, datasets, or packages, provide proper references and citations to acknowledge the sources and give credit where due.
12. **Check for Clarity and Consistency**: Review your report for clarity, coherence, and consistency in formatting, language, and presentation style. Ensure that it follows a professional and polished format.
13. **Test and Validate**: Before finalizing the report, thoroughly test and validate your code to ensure accuracy and reproducibility of results.
14. **Share and Collaborate**: Once the report is ready, share it with relevant stakeholders for review and feedback. Collaborate with peers or experts to enhance the quality and credibility of your analysis.

9 Appendix A

In an appendix section of an RMarkdown document, you can include additional information that supports or supplements the main content of your report. Here are some elements you might consider adding to the appendix:

1. **Additional Data**: If your analysis relies on a large dataset or includes multiple datasets, you can provide the raw data or additional datasets in the appendix. This allows readers to access the data used in your analysis for further exploration or validation.
2. **Code Scripts**: Include any additional R scripts or code chunks that were used in the analysis but are not central to the main narrative. This can include data preprocessing steps, exploratory data analysis (EDA), or advanced statistical analyses.
3. **Technical Details**: If your analysis involves complex statistical methods, algorithms, or modeling techniques, you can include detailed technical descriptions or mathematical derivations in the appendix. This provides readers with a deeper understanding of the analytical methods employed.
4. **Supplementary Tables and Figures**: Include supplementary tables, graphs, or visualizations that provide additional insights or details beyond what is presented in the main text. This could include sensitivity analyses, alternative model specifications, or additional results.
5. **Assumptions and Limitations**: Discuss any assumptions made or limitations of the analysis that were not covered in the main text. This helps to provide transparency and context for the interpretation of the results.
6. **Appendices from References**: If your report references external sources, such as academic papers, reports, or datasets, you can include relevant appendices from these sources in your document. Make sure to obtain permission and provide proper citations for any copyrighted material.
7. **Survey Instruments or Questionnaires**: If your analysis involves survey data or qualitative research, you can include the survey instruments, questionnaires, or interview guides used in the appendix. This allows readers to understand the methodology and design of the study.
8. **Data Cleaning and Preprocessing Details**: Provide detailed descriptions of the data cleaning and preprocessing steps undertaken before analysis. This can include handling missing data, data transformations, outlier detection, or variable coding schemes.
9. **Model Specifications**: If your analysis involves statistical models or machine learning algorithms, include detailed specifications of the models used, including parameter estimates, model diagnostics, and goodness-of-fit statistics.
10. **Software and Package Versions**: List the versions of software and R packages used in the analysis to ensure reproducibility. This includes the version of R, RStudio, and any other relevant software tools.

E.Gjika