Exploratory Data ANalysis for Ecologist and Biologists

Thiyanga S. Talagala

2025-04-17

Table of contents

Preface **4**

1 Introduction **5**

2 Introduction to Exploratory Data Analysis **6**
2.1 What is Exploratory Data Analysis (EDA)? 6
2.2 Why EDA is critical before modeling? 6
2.3 DA vs Confirmatory Data Analysis (CDA) 6
2.4 Examples of questions EDA can answer 6

3 Getting to Know Your Data and Making Your Data Ready to Analyse **7**
3.1 Types of Data 7
3.2 Sacles of Measurements 7
3.3 Introduction to tidy data 7
3.4 Dos and Don'ts when Entering Data into an Excel Sheet 7

4 R Programming Basics **8**
4.1 Intalling R 8
4.2 Installing RStudio 8
4.3 Installing Rtools 8
4.4 Installing Packages in R 8
4.5 Import Data into R 8
4.6 Use factors for categorical variables 8

5 Data Wrangling for EDA **9**
5.1 Introduction 9
5.2 Pivot Data from Long to Wide 9
5.3 Pivot Data from Long to Wide 9
5.4 Filter Subset of Rows 9
5.5 Select Subset of Columns 9
5.6 Reorder Rows and Columns 9
5.7 Create New Variables 9
5.8 Collapse Dataset to Summary Measures 9
5.9 Create Group-Wise Calculations 9
5.10 Rename Exsisting Variables 9

6 Data Profiling for EDA **10**
6.1 What is Data Profiling? . . . 10
6.2 Why Data Profiling is Important? . . . 10
6.3 Structure Analysis . . . 10
6.4 Duplicate Analysis . . . 10
6.5 Value Distribution . . . 10
6.6 Range and Outliers . . . 10
6.7 Consistency . . . 10
6.8 Relationships Between Variables . . . 10
6.9 Missing Value Visualisation . . . 10

7 Univariate Exploratory Data Analysis **11**
7.1 EDA for Qualitative Variables . . . 11
7.2 EDA for Quantitative Variables . . . 11

8 Bivariate Exploratory Data Analysis **12**
8.1 EDA for Qualitative Variables . . . 12
8.2 EDA for Quantitative Variables . . . 12
8.3 EDA for Qualitative Variables and Quantitative Variables . . . 12

9 Trivariable Exploratory Data Analysis **13**
9.1 EDA for Qualitative Variables . . . 13
9.2 EDA for Quantitative Variables . . . 13
9.3 EDA for Qualitative Variables and Quantitative Variables . . . 13

10 High Dimensional(More than 3 Variables) Exploratory Data Analysis **14**
10.1 EDA for Qualitative Variables . . . 14
10.2 EDA for Quantitative Variables . . . 14
10.3 EDA for Qualitative Variables and Quantitative Variables . . . 14

11 Casestudies in Exploratory Data Analysis **15**
11.1 Casestudy 1 . . . 15
11.2 Casestudy 2 . . . 15

References **16**

Preface

This is a Quarto book.

To learn more about Quarto books visit https://quarto.org/docs/books.

1 Introduction

This is a book created from markdown and executable code.

See Knuth (1984) for additional discussion of literate programming.

2 Introduction to Exploratory Data Analysis

2.1 What is Exploratory Data Analysis (EDA)?

2.2 Why EDA is critical before modeling?

2.3 DA vs Confirmatory Data Analysis (CDA)

2.4 Examples of questions EDA can answer

3 Getting to Know Your Data and Making Your Data Ready to Analyse

3.1 Types of Data

3.2 Sacles of Measurements

3.3 Introduction to tidy data

3.4 Dos and Don'ts when Entering Data into an Excel Sheet

4 R Programming Basics

4.1 Intalling R

4.2 Installing RStudio

4.3 Installing Rtools

4.4 Installing Packages in R

4.5 Import Data into R

4.6 Use factors for categorical variables

5 Data Wrangling for EDA

5.1 Introduction

5.2 Pivot Data from Long to Wide

5.3 Pivot Data from Long to Wide

5.4 Filter Subset of Rows

5.5 Select Subset of Columns

5.6 Reorder Rows and Columns

5.7 Create New Variables

5.8 Collapse Dataset to Summary Measures

5.9 Create Group-Wise Calculations

5.10 Rename Exsisting Variables

6 Data Profiling for EDA

6.1 What is Data Profiling?

6.2 Why Data Profiling is Important?

Ensures data quality before analysis or modeling

Helps in detecting errors early

Guides decisions on data cleaning and transformation

Builds trust in data-driven research or decision-making

6.3 Structure Analysis

6.4 Duplicate Analysis

6.5 Value Distribution

6.6 Range and Outliers

6.7 Consistency

6.8 Relationships Between Variables

6.9 Missing Value Visualisation

7 Univariate Exploratory Data Analysis

7.1 EDA for Qualitative Variables

7.2 EDA for Quantitative Variables

8 Bivariate Exploratory Data Analysis

8.1 EDA for Qualitative Variables

8.2 EDA for Quantitative Variables

8.3 EDA for Qualitative Variables and Quantitative Variables

9 Trivariable Exploratory Data Analysis

9.1 EDA for Qualitative Variables

9.2 EDA for Quantitative Variables

9.3 EDA for Qualitative Variables and Quantitative Variables

10 High Dimensional(More than 3 Variables) Exploratory Data Analysis

10.1 EDA for Qualitative Variables

10.2 EDA for Quantitative Variables

10.3 EDA for Qualitative Variables and Quantitative Variables

11 Casestudies in Exploratory Data Analysis

11.1 Casestudy 1

11.2 Casestudy 2

References

Knuth, Donald E. 1984. "Literate Programming." *Comput. J.* 27 (2): 97–111. https://doi.org/10.1093/comjnl/27.2.97.