

# **Exploratory Data Analysis for Ecologists and Biologists**

Thiyanga S. Talagala

2025-04-17

# Table of contents

<b>Preface</b>	<b>5</b>
<b>1 Introduction</b>	<b>6</b>
<b>2 Introduction to Exploratory Data Analysis</b>	<b>7</b>
2.1 What is Exploratory Data Analysis (EDA)? . . . . .	7
2.2 Why EDA is critical before modeling? . . . . .	7
2.3 DA vs Confirmatory Data Analysis (CDA) . . . . .	7
2.4 Examples of questions EDA can answer . . . . .	7
<b>3 Getting to Know Your Data and Making Your Data Ready to Analyse</b>	<b>8</b>
3.1 Types of Data . . . . .	8
3.2 Sacles of Measurements . . . . .	8
3.3 Introduction to tidy data . . . . .	8
3.4 Dos and Don'ts when Entering Data into an Excel Sheet . . . . .	8
<b>4 R Programming Basics</b>	<b>9</b>
4.1 Intalling R . . . . .	9
4.2 Installing RStudio . . . . .	9
4.3 Installing Rtools . . . . .	9
4.4 Installing Packages in R . . . . .	9
4.5 Import Data into R . . . . .	9
4.6 Use factors for categorical variables . . . . .	9
<b>5 Data Wrangling for EDA</b>	<b>10</b>
5.1 Introduction . . . . .	10
5.2 Pivot Data from Long to Wide . . . . .	10
5.3 Pivot Data from Long to Wide . . . . .	10
5.4 Filter Subset of Rows . . . . .	10
5.5 Select Subset of Columns . . . . .	10
5.6 Reorder Rows and Columns . . . . .	10
5.7 Create New Variables . . . . .	10
5.8 Collapse Dataset to Summary Measures . . . . .	10
5.9 Create Group-Wise Calculations . . . . .	10
5.10 Rename Exsisting Variables . . . . .	10

<b>6</b>	<b>Data Profiling for EDA</b>	<b>11</b>
6.1	What is Data Profiling? . . . . .	11
6.2	Why Data Profiling is Important? . . . . .	11
6.3	Structure Analysis . . . . .	11
6.4	Duplicate Analysis . . . . .	11
6.5	Value Distribution . . . . .	11
6.6	Range and Outliers . . . . .	11
6.7	Consistency . . . . .	11
6.8	Relationships Between Variables . . . . .	11
6.9	Missing Value Visualisation . . . . .	11
<b>7</b>	<b>Univariate Exploratory Data Analysis</b>	<b>12</b>
7.1	EDA for Qualitative Variables . . . . .	12
7.2	EDA for Quantitative Variables . . . . .	12
<b>8</b>	<b>Bivariate Exploratory Data Analysis</b>	<b>13</b>
8.1	EDA for Qualitative Variables . . . . .	13
8.2	EDA for Quantitative Variables . . . . .	13
8.3	EDA for Qualitative Variables and Quantitative Variables . . . . .	13
<b>9</b>	<b>Trivariable Exploratory Data Analysis</b>	<b>14</b>
9.1	EDA for Qualitative Variables . . . . .	14
9.2	EDA for Quantitative Variables . . . . .	14
9.3	EDA for Qualitative Variables and Quantitative Variables . . . . .	14
<b>10</b>	<b>High Dimensional(More than 3 Variables) Exploratory Data Analysis</b>	<b>15</b>
10.1	EDA for Qualitative Variables . . . . .	15
10.2	EDA for Quantitative Variables . . . . .	15
10.3	EDA for Qualitative Variables and Quantitative Variables . . . . .	15
<b>11</b>	<b>Spatial Data Visualisation</b>	<b>16</b>
11.1	What is Spatial Data? . . . . .	16
11.2	Data Wrangling for Spatial Data . . . . .	16
11.3	Map Visualisations . . . . .	16
<b>12</b>	<b>Temporal Data Visualisation</b>	<b>17</b>
12.1	What is Temporal Data? . . . . .	17
12.2	Data Wrangling for Temporal Data . . . . .	17
12.3	Time Series Visualisations . . . . .	17
<b>13</b>	<b>Spatio-Temporal Data Visualisation</b>	<b>18</b>
<b>15</b>	<b>Way Forward</b>	<b>20</b>

<b>16 Casestudies in Exploratory Data Analysis</b>	<b>21</b>
16.1 Casestudy 1 . . . . .	21
16.2 Casestudy 2 . . . . .	21
<b>References</b>	<b>22</b>

# 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 Scales 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 Existing 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 Spatial Data Visualisation**

## **11.1 What is Spatial Data?**

## **11.2 Data Wrangling for Spatial Data**

## **11.3 Map Visualisations**



# **12 Temporal Data Visualisation**

## **12.1 What is Temporal Data?**

## **12.2 Data Wrangling for Temporal Data**

## **12.3 Time Series Visualisations**

## **13 Spatio-Temporal Data Visualisation**

**14**

## **15 Way Forward**

## **16 Casestudies in Exploratory Data Analysis**

### **16.1 Casestudy 1**

### **16.2 Casestudy 2**

## References

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