

Chapter 1

Exploratory Data Analysis

1.1 Exploratory Data Analysis

Nowadays, most ecological research is done with hypothesis testing and modelling in mind. However, Exploratory Data Analysis(EDA), which uses visualization tools and computer synthetic descriptors, is still required at the beginning of the statistical analysis of multidimensional data, in order to:

- Get an overview of the data
- Transform or recode some variables
- Orient further analyses

1.1.1 Building Intuition about the data

1. Getting domain knowledge It helps to deeper understand the problem
2. Checking if the data is intuitive To be agree with domain knowledge
3. Understanding how the data was generated A it is crucial to set up a proper Validation

1.1.2 Data cleaning

1. Constant features
2. Duplicated features

3. Duplicated rows
4. Check if datasets is shuffled

here graphics!! :D

1.1.3 Data splitting strategy

- 1.

1.1.4 Exploratory data Analysis

1.1.5 Exploratory anonymized data

1.1.6 Validation and overfitting

1.1.7 Validation strategies

1.1.8 Visualizations