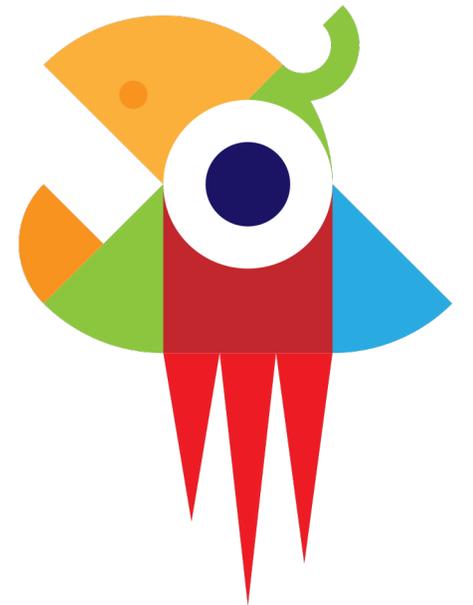


# Text Mining with R



# What is EDA?



# What is EDA?

**EDA stands for Exploratory Data Analysis, which is a method of analyzing and summarizing a dataset in order to better understand its characteristics and properties. EDA involves a range of techniques and tools that can be used to identify patterns, relationships, and anomalies in the data, and to inform the development of hypotheses and models.**



Why is EDA an important part of any data analysis?



# Why is EDA an important part of any data analysis?

**First, EDA can help to identify potential errors or inconsistencies in the data, such as missing values, outliers, or data entry errors. This can help to ensure that the data is of sufficient quality for analysis and can help to prevent inaccurate or misleading results.**



# Why is EDA an important part of any data analysis?

**Second, EDA can provide insights into the distribution and characteristics of the data, including measures of central tendency, variability, and skewness. This can help to identify potential relationships or patterns in the data, as well as potential sources of bias or confounding factors.**



# Why is EDA an important part of any data analysis?

**Third, EDA can help to identify potential outliers or anomalies in the data that may require further investigation or treatment. This can help to ensure that the analysis is based on a more representative sample of the data and can help to prevent inaccurate or misleading results.**



In which conditions is a variable categorical?





# In which conditions is a variable categorical?

**A variable is categorical when it takes on a limited set of values that represent a particular category or group.**



In which conditions is a variable continuous?



# In which conditions is a variable continuous?

**A variable is considered continuous when it can take on any value within a defined range and when the differences between values are uniform and measurable.**



In which conditions can categorical variables vary?



# In which conditions can categorical variables vary?

- 1. Differences in the levels or categories of the variable**
- 2. Changes in the population or sample being studied**
- 3. Changes in the classification or measurement of the variable**
- 4. Differences in the data collection methods or tools**



In which conditions does covariation describe the behavior between variables?



# In which conditions does covariation describe the behavior between variables?

- 1. Correlation:** relationship between two variables that are related, and changes in one variable are systematically related to changes in the other variable.
- 2. Linear relationship:** the change in one variable is proportional to the change in the other variable.
- 3. Statistical significance:** Covariation is considered to be significant when the relationship between the two variables is unlikely to have occurred by chance.



# Why are regular expressions useful?





# Why are regular expressions useful?

- 1. Flexible pattern matching:** Regular expressions allow you to define flexible patterns for searching and matching text.
- 2. Efficient searching:** Regular expressions are designed to be efficient at searching and matching text, especially when dealing with large amounts of data.
- 3. Portable and widely supported:** Regular expressions are supported by many programming languages and tools



How will regular expressions match any part of a string?



# How will regular expressions match any part of a string?

**Regular expressions can be used to match any part of a string using special characters known as "wildcards." Wildcards are characters that can match any character or set of characters in a string.**



# Why is r Markdown important?



# Why is r Markdown important?

- 1. Reproducibility:** R Markdown allows you to create dynamic documents that integrate R code, text, and graphics. This makes it easy to reproduce your analysis and results, as all the code, data, and results are contained in a single document.
- 2. Communication:** R Markdown documents can be used to communicate your analysis and results to others.



# Why is r Markdown important?

3. **Efficiency:** R Markdown can save time and increase efficiency by allowing you to streamline your workflow.
4. **Flexibility:** R Markdown is flexible and can be used for a variety of tasks, including data analysis, data visualization, research reports, and even web applications.
5. **Learning:** R Markdown can also be used as a tool for learning and teaching R.





UNIVERSITÀ DI PISA

université  
de **BORDEAUX**



University  
of Bremen

**ERREQUADRO**

Research over Research



1542

**Universidad**  
Zaragoza

