1. SYT-Vertiefung

exploratory data analysis (in R and Python)

1.1. Introduction

EDA (exploratory data analysis) is used to understand different data sets by looking at their main characteristics and even plotting them visually to understand them better.

In this document, I would like to discuss the fundamental techniques as well as the concrete implementations of these techniques in R and Python.

Data Analysis is and will get much more important in the future. In 2007 it was assumed that the total quantity of data that was produced until then was about five exabytes ($5 \cdot 10^{18} bytes$) . At that point of time it was more than the estimated number of words spoken by humans. [1]

Because there are some open-source-software-solutions for nearly everything, there is also a solution for data analysis: the R programming language.

> Basically there are three motivations for analyzing data:
>
> - to understand what has happened or what is happening;
> - to predict what is likely to happen, either in the future or in other circumstances we haven't seen yet;
> - to guide us in making decisions.

[1]

In order to analyze our data/the data of our company or something else, we have to begin with the simple sounding first step: understanding our data. And there it is necessary to use exploratory data analysis on every dataset we want to analyze in the future in order to understand it and make decision regarding certain predictions models for example.

When it comes to the size of data

1.2. Milestones

milestone	date
Introduction and Sources complete	17.11.2020
theoretical explanations complete	01.12.2020
simple practical example complete	08.12.2020
two more extensive examples and the comparsion of the complete	22.12.2020

1.3. Sources

[1] Ronald K. Pearson. (n.d.). Exploratory Data Analysis Using R [Book]. In Exploratory Data Analysis Using R. CRC Press. https://doi.org/10.1201/9781315382111

[2] Exploratory data analysis in Python. | by Tanu N Prabhu | Towards Data Science. (n.d.). Retrieved November 17, 2020, from https://towardsdatascience.com/exploratory-data-analysis-in-python-c9a77dfa39ce