Explorative Data Analysis and Visualization

Course Description

Obtaining an overview of the salient characteristics of a data set is one of the core activities at the outset of any data analysis endeavour. The corresponding activities, methods, and techniques are grouped under the term “exploratory data analysis”. During exploratory data analysis, gaining insight into a given data set is often aided by the application of suitable visualization techniques. The utility of visualization, however, does not end at this stage; it is also crucial for communicating analytical outcomes. This course first introduces a set of approaches, tools, and techniques that are useful for exploring data sets. It then takes a thorough look at the subject area of visualization, which is presented in detail by an exposition arc that spans from first principles of visualization to practical implementation to insights into the communication of data science results and findings.

Contents

1. Exploratory Data Analysis
   1. Location and variability
   2. Further exploration of data distribution
   3. Covariance and correlation
2. Data Visualization Principles
   1. Coordinates and axes
   2. Color spaces
   3. Graph types
3. Data Visualization Practice
   1. Amounts, proportions, associations, and distributions
   2. Time series and trends
   3. Geo-spatial data
4. Visualization in Python – Matplotlib and Seaborn
   1. Introduction to PyPlot, Matplotlib, and Seaborn
   2. Basic plots
   3. Geo-spatial plots
5. Communicating Data Science
   1. Unclutter, focus, and capture attention
   2. Lessons from design
   3. Principles of storytelling with data