manipalglobal
Academy of Data Science

Exploratory Data Analysis

by Neelima Vobugari
B.Tech, MBA, Data Science Specialization

manipalglobal

Academy of Data Science

Descriptive Analysis

by Neelima Vobugari

Contents

- Descriptive Analysis
- Data and Datasets
- Quantitative variables and Qualitative variables
- Summarization and aggregation of Data

Descriptive Analysis

- Utilized to portray the essential elements of the data in a study
- Together with basic representation investigation, they frame the premise of for all intents and purposes each quantitative examination of information.
- Used to introduce quantitative portrayals in a sensible frame
- Alternately we may quantify a substantial number of individuals on any measure.
- Unmistakable insights help us to disentangle a lot of data sensibly
- Each elucidating measurement lessens loads of data into an easier outline

Data and dataset

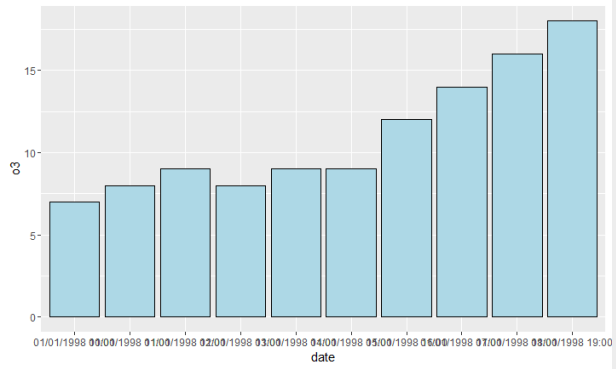
Various dimensionality of the datasets with data are as follows

- Uni-variate : Measurement made with respect to single variable per subject
- Bivariate : Measurement made with respect to two variables per subject
- Multivariate: Measurement made with respect to many variables per subject

Graphical summaries of dataset

Visualization of data gives more insights of data than visualizing in dataset. The visualization plots depends on dimensionality of data and are explained below

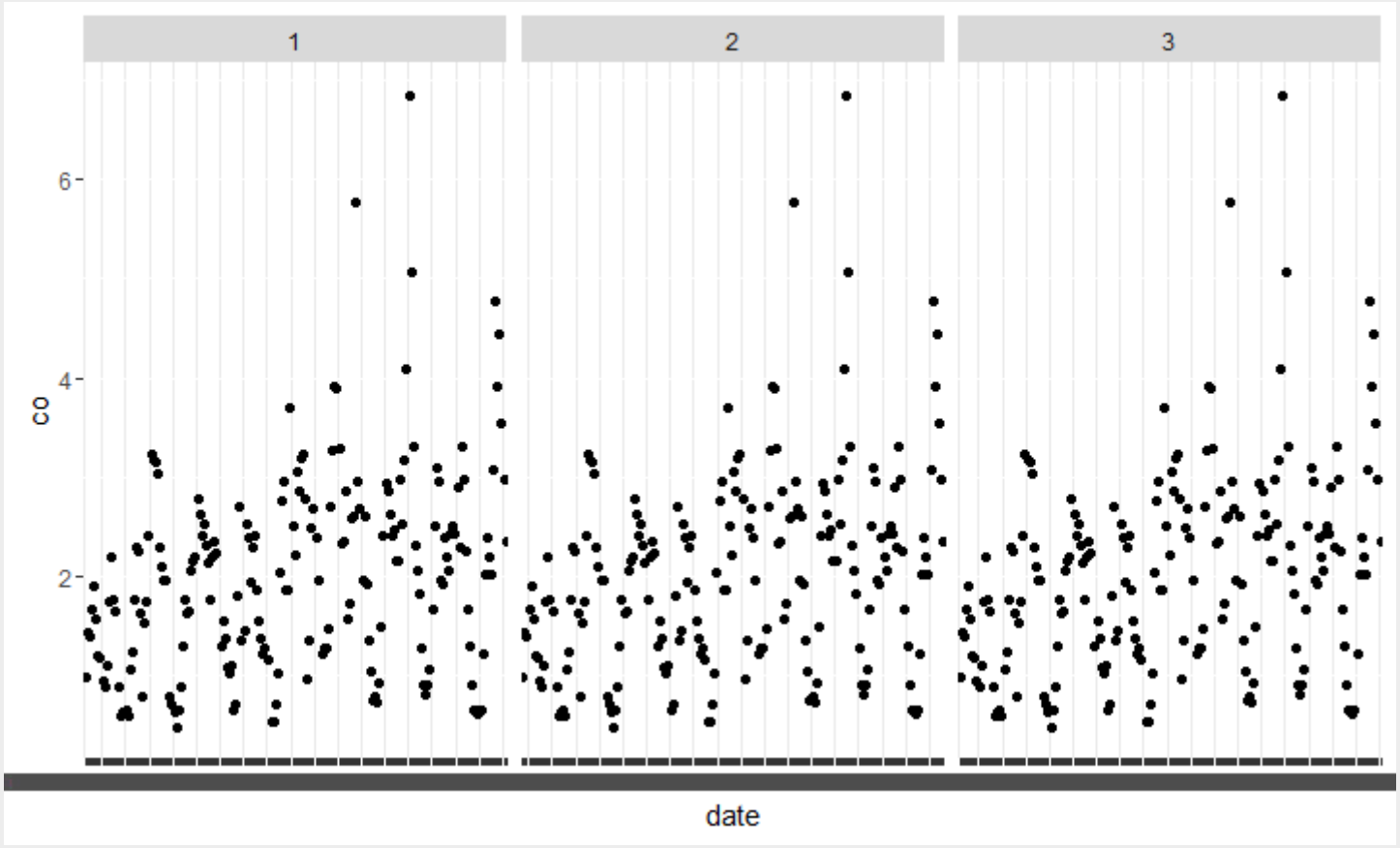
- Uni-variate : plots for these dataset will be representation of variable data in Y axis with increasing index in X axis. The bar plots and histogram plots are used to plot data.



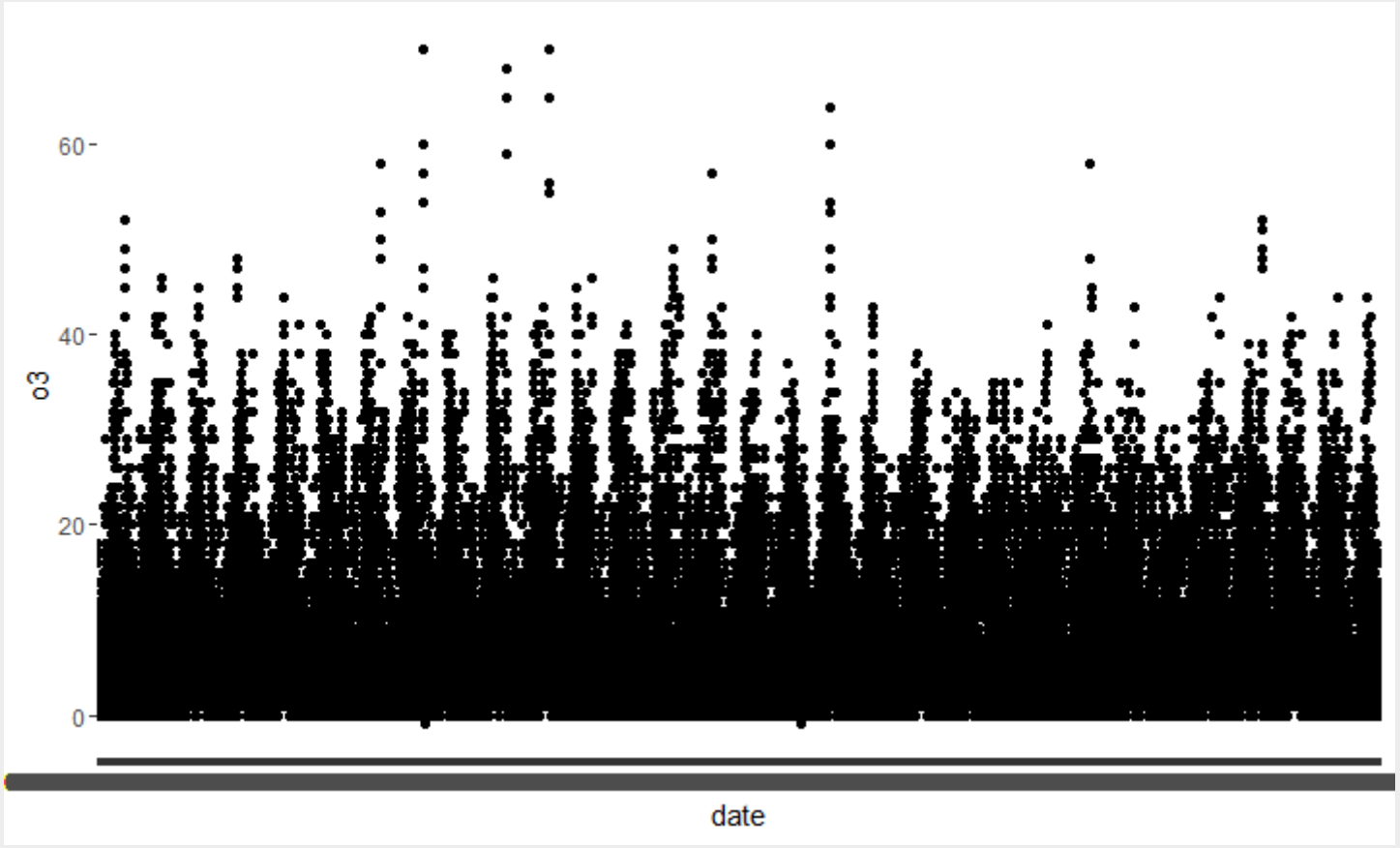
Graphical summaries of dataset

- Bivariate : the visualization of these kinds of plots represent a variable in Y direction with another variable in x direction
- Crosstabs plots are used to display categorical variable with categorical variable
- Box plot is used to display categorical variable with continuous variable
- Scatter plot and stacked box plot is used to display continuous variable with continuous variable

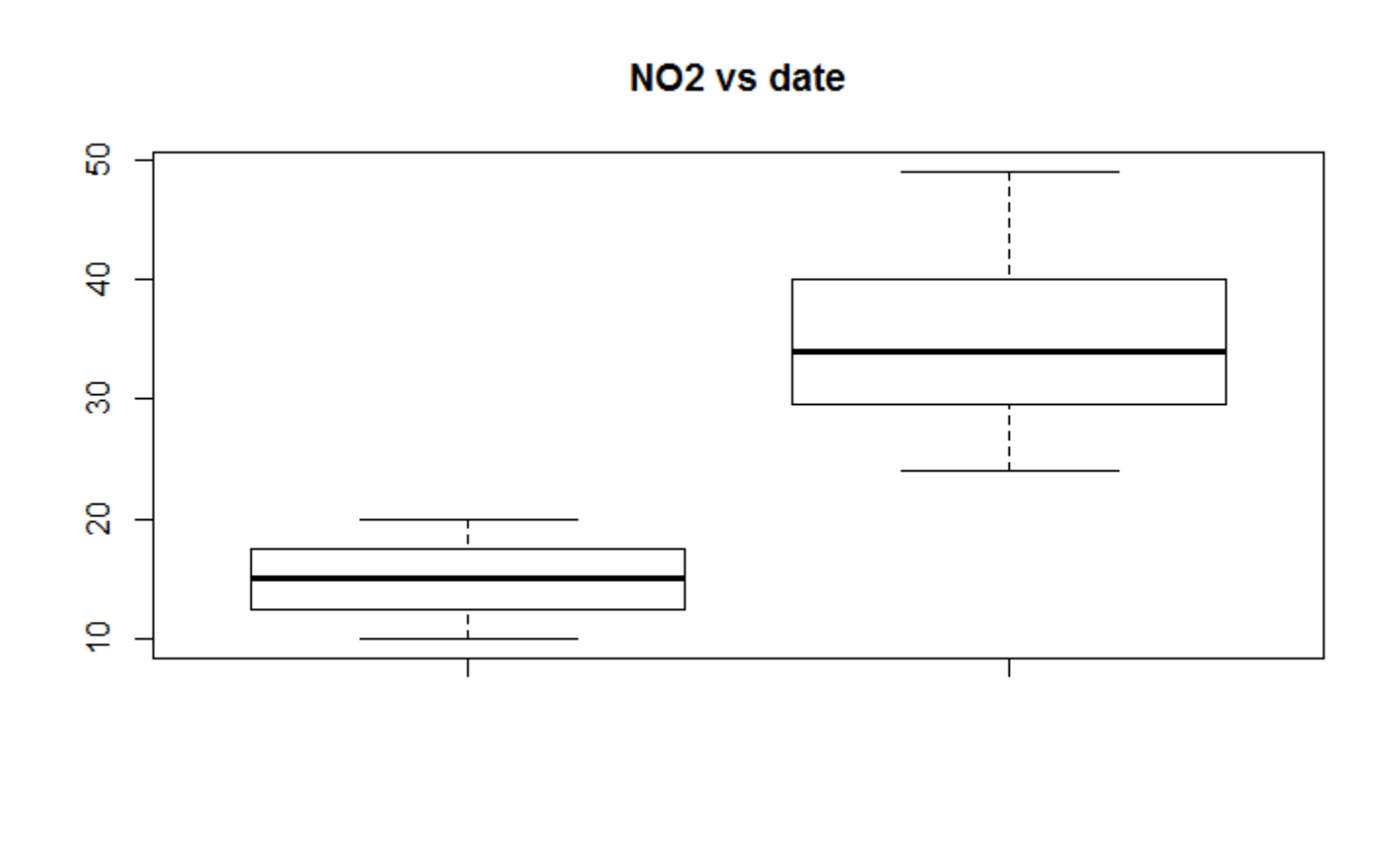
Graphical Summaries of Dataset



Graphical Summaries of Dataset



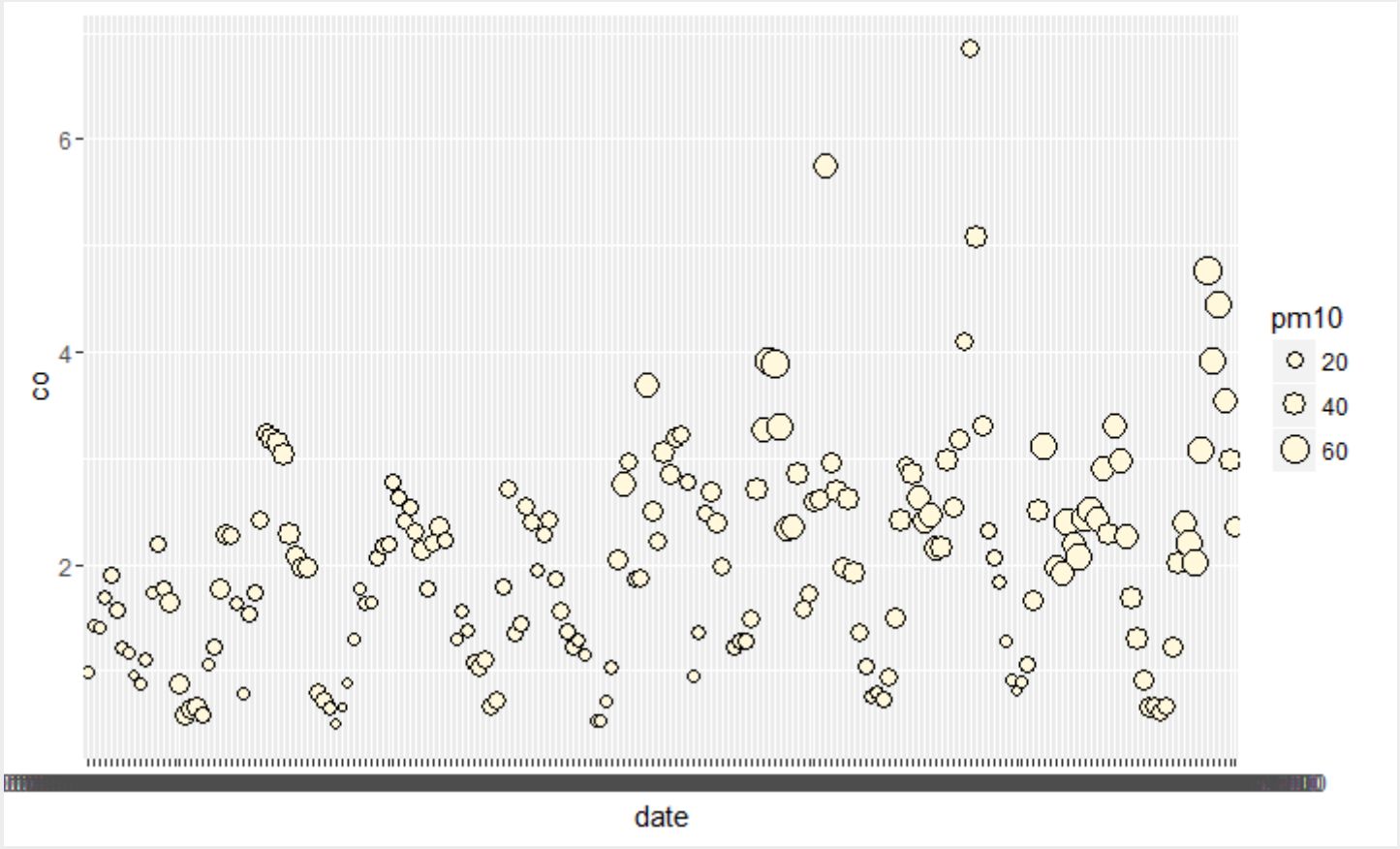
Graphical Summaries of Dataset



Graphical summaries of dataset

- Multivariate : the visualization of these kinds of plots represent a variable with respect to another variable in Y direction with another variable in x direction
- Dimensionality reduction and clustering is suitable method to analyze data
- The plot displayed is showing carbon monoxide with respect to pm10 in y axis against date in x axis

Graphical Summaries of Dataset



Quantitative variables and Qualitative variables

- Quantitative data are measures of qualities or considers and are communicated numbers.
- Quantitative data are information about numeric factors (e.g. what number of; how much; or how frequently).
- Quantitative data are measures of "sorts" and might be spoken to by a name, image, or a number code.
- Quantitative data are information about absolute factors (e.g. what sort)

Quantitative variables and Qualitative variables

- Quantitative and qualitative data give distinctive results, and are frequently utilized together to get a full information of a populace
- For instance, if information are gathered on yearly wage (quantitative), occupation information (subjective) could likewise be accumulated to get more detail on the normal yearly pay for every sort of occupation
- Quantitative and qualitative data can be accumulated from similar information unit relying upon whether the variable of interest is numerical or unmitigated

Summarization and aggregation of Data

- Data Summarization condenses evolutional data included both primitive and determined information, with a specific end goal to make an inferred evolutional data that is general in nature
- Since the data in the data warehouse is of high volume, there should be an instrument to get just the applicable and important data in a less chaotic organization
- Data Summarization gives the ability to give data shoppers sum up perspective of unique main part of information.

Summarization and aggregation of Data

- Data Summarization in expansive multi-dimensional datasets as on account of data warehouse is an exceptionally difficult work
- Data aggregation is any procedure in which data is accumulated and communicated in an outline frame, for purposes, for example, measurable examination
- A typical aggregation intention is to get more data about specific gatherings in light of particular factors, for example, age, calling, or wage

Summarization and aggregation of Data

- The data about such gatherings can then be utilized for Web website personalization to pick substance and promoting prone to speak to an individual having a place with at least one gatherings for which information has been gathered
- Online diagnostic preparing (OLAP) is a straightforward kind of information total in which the advertiser utilizes an internet reporting system to handle the data.

- Thank you