Descriptive Data Analysis (DDA)

DDA is a basic tool for understanding your data

DDA is used throughout all stages of data analytics.

Types of Data

Be aware of the type of data that you have:

- categorical versus continuous

– categorical: nominal versus ordinal

– continuous: interval versus ratio

- structured versus unstructured

and for numerical values, be aware of

- accuracy

- precision

- significant digits

In Week 4 we will cover these topics (not covered in Week 3):

structured versus unstructured

numerical data, and issues

- accuracy

- precision

- significant digits

Describe Data Quantitatively

Describe the data and the data distribution for each feature in the dataset

- central tendency: mean, median, mode

- variation: standard deviation, inter-quartile range (IQR)

- outlier values and extreme values

- skew

- kurtosis

You want descriptions that are *robust* to presence of outliers

Visualize Data

Univariate and bivariate visualization

Visualization as bar charts, histograms, box-plots, violin plots, scatter plots