DATA STATISTICS

Descriptive statistics
Basic plots

Descriptive Statistics - overall
Diminsions - number of instances (rows), number of columns/attributes

Attribute Statistics (univariate)
- statistics for numeric attributes (mean, variance, etc.)
df.describe()
- Statistics for categorical attributes (histograms, mode, most/least frequent values, percentage, number of unique values)
· Histogram of values df[<attribute>].value_counts()
or seaborn's distplot()
- Target statistics
class distribution df[<target>].value_counts()
or np.bincount(y)

Multivariate Statistics
- Correlation
- contingency tables/cross tabulation

Pandas example

for each numeric feature, can look at density plot, histogram, box plot

2:30 left

Scatterplot - look for relationships between any two variables