

Exploratory Data Analysis2

Prof. Eric A. Suess

September 17, 2018

EDA

The main ideas with EDA are the exploration of the variables.

- ▶ Centers - means, medians
- ▶ Variation - standard deviations, ranges
- ▶ Outliers
- ▶ Missing Values
- ▶ Covariation, Correlation

Two variables

- ▶ Two categorical variables - Contingency Tables, Chi-Square Tests
- ▶ One categorical variable and one Numeric variable - Side-by-side boxplots, T-tests, One-way ANOVA
- ▶ Two numeric variables - Scatterplots, Correlation, Regression, Smoothing

Correlation

What does the correlation coefficient measure?

The correlation coefficient r measures the **strength** and **direction** of the **linear** relationship between two **quantitative/numeric** variables.