Modelling Approach so far

- Assumptions of Linear Regression
- Data
  - Understand data
  - Are there external sources
  - Quality of Data
  - Data Cleanup
    - Format
    - Scaling
    - Handling of quantitative & qualitative data
  - Correlations
- Linear Model Selection
  - Selecting predictors
    - Automatic methods
      - Best subset
      - Stepwise selection (forward/backward)
      - Hybrid approach
    - Shrinkage methods
      - Ridge
      - Lasso
    - Dimension reduction
    - Manual - visual / domain expert
- How to choose optimal model
  - C_p, AIC, BIC, Adjusted R^2 - these methods allow us to estimate the "quality of a model"
  - Validation / cross validation
- Assessing Model Accuracy
  - Residual analysis
  - Summary results
- Potential problems
  - Non linearity
  - Correlation of error terms
  - Outliers
  - High leverage points
  - collinearity