**Econometrics:**

* Regression model examples: estimating beta, etc.
* We use the information from dependent variables to make a prediction model and explain why the output/independent variable is related to the dependent
  + Regression models – one of the most used ML models
* In econometric – interested more in explaining vs prediction
* Data Wrangling (cleaning) – the process of transforming and mapping raw data into a format suitable for analysis

**Data Wrangling in Excel:**

* Relative, Absolute, Mixed references – important so you can iterate functions across many cells
* Subletting: the process of extracting portions of a dataset that are relevant to the analysis
  + Commonly used to pre-process the data prior to analysis

**Missing Observations:**

* **Two Strategies:** 
  + **Omission**: observations with missing values be excluded from subsequent analysis
  + **Imputation**: missing values be replaced with some reasonable imputed values

**Transforming Categorical Variables:**

* A common technique of transforming categorical variables is to covert them to numerical
* **Dummy Variable:** aka. Indicator or binary variables, commonly used to describe two categories of a variable
  + Assumes 1 for one of the categories and 0 for the other category
    - This is referred to as the reference or the benchmark category
  + If you have more than two categories: general rule – create k-1 dummy variables in the regression analysis, using the last category as a reference