* The ‘<-‘ is called an assignment operator. You assign a value or name to an object. Objects are vectors, matrices, dataframes, and other structures.
  + Vectors
  + Dataframe
  + Factor
* The data loaded in the workspace is an object called a dataframe. A dataframe is a data structure. It’s a table-like structure with rows and columns (think an Excel spreadsheet).
* Use the $ (dollar sign) to open up a dataframe and create a pull list of the variables in the dataframe. You can then choose a variable to perform a task.
* 2 branches of statistics:

1. Descriptive statistics: “must knows” of the data, describes the data
2. Inferential statistics: infer sample (statistic) to the population (parameter), hypothesis testing

* Common forms of descriptive statistics:
* Summary statistics: summarize and provide information about your sample data. It tells you something about the values in your data set.
  + Measures of location (central tendency) – mean, mode, median.
  + Measures of spread (dispersion) (i.e., how wide is the data) – range, standard deviation, variance.
  + Count, maximum, minimum
* Tables
  + Frequency table – a distribution of a single value or category in the column (variable)
  + Cross tabular – a distribution of two variables
* Plots
  + Histogram – a plot of a continuous variable
  + Bar plot – a plot of the categorical or discrete variable
  + Boxplot – a visual representation of summary statistics of a continuous variable
* What can we find out about the input variables?
  + Create a data dictionary (\*NOTE – table does not include definitions of variables as it should)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Variable Name | Data Type | Data Type in R | Missing values? | How many Unique values? (count) | Minimum | Mean (average) | Maximum |
| Year |  |  |  |  |  |  |  |
| Treatment\_Team |  |  |  |  |  |  |  |
| DAYS |  |  |  |  |  |  |  |
| SEX |  |  |  |  |  |  |  |
| Race |  |  |  |  |  |  |  |
| Ethnicity |  |  |  |  |  |  |  |