```
> # Weld Experiment Data Set#
>
> # A detailed dataset description #
> library(readxl)
> data <- read_excel("/Users/zalawadi/Library/Mobile Documents/com-apple-CloudDocs/STAT505/11_WeldExperiment dataset.xlsx", sheet = "Experiment Meta Data") #Read the Excel Sheet
>
> # Structure and data type
> ?class
starting httpd help server ... done
> class(data)
[1] "tbl_df"     "tbl"        "data.frame"
> #data types of each column
> str(data)
tibble [41 × 13] (S3: tbl_df/tbl/data.frame)
 $ Test_Specimen          : chr [1:41] "#" "A1" "A2" "A3" ...
 $ Material_THK           : chr [1:41] "in" "0.25" "0.25" "0.25" ...
 $ Leg_Size               : chr [1:41] "in" "0.25" "0.25" "0.25" ...
 $ Throat_Size            : chr [1:41] "in" "0.17674999999999999" "0.17674999999999999" "0.17674999999999999" ...
 $ Weld_Length            : chr [1:41] "in" "5.9375" "5.9375" "6" ...
 $ Effective_Weld_Area    : chr [1:41] "in^2" "1.0494531249999999" "1.0494531249999999" "1.0605" ...
 $ Tensile_Strength       : chr [1:41] "psi" "70000" "70000" "70000" ...
 $ Transverse_Theo_Failure: chr [1:41] "lbf" "73461.718749999985" "73461.718749999985" "74235" ...
 $ Shear_Theo_Failure     : chr [1:41] "lbf" "22038.515624999996" "22038.515624999996" "22270.5" ...
 $ Actual_Failure         : chr [1:41] "lbf" "4947.49" "6054.02" "6383.93" ...
 $ MacroEtch              : chr [1:41] "Pass/Fail" "P" "P" "P" ...
 $ Break_Fusion           : chr [1:41] "Pass/Fail" "F" "P" "P" ...
 $ Equation_Factor        : num [1:41] NA 0.0673 0.0824 0.086 0.069 ...
> #measurment formats
> ?sapply
> sapply(data,class)
       Test_Specimen             Material_THK                 Leg_Size              Throat_Size
         "character"              "character"              "character"              "character"
         Weld_Length      Effective_Weld_Area         Tensile_Strength Transverse_Theo_Failure
         "character"              "character"              "character"              "character"
  Shear_Theo_Failure           Actual_Failure                MacroEtch             Break_Fusion
         "character"              "character"              "character"              "character"
     Equation_Factor
           "numeric"
> #character is categorical and numeric is numeric
> #categorical columns
> names(data)[sapply(data, function(x) is.character(x) | is factor(x))]
Error: unexpected symbol in "names(data)[sapply(data, function(x) is.character(x) | is factor"
> names(data)[sapply(data, function(x) is.character(x) | is.factor(x))]
 [1] "Test_Specimen"           "Material_THK"            "Leg_Size"
 [4] "Throat_Size"             "Weld_Length"             "Effective_Weld_Area"
 [7] "Tensile_Strength"        "Transverse_Theo_Failure" "Shear_Theo_Failure"
[10] "Actual_Failure"          "MacroEtch"               "Break_Fusion"
> #numeric columns
> names(data)[sapply(data, is.numeric)]
[1] "Equation_Factor"
> #dimensions
> nrow(data)
[1] 41
> ncol(data)
[1] 13
> dim(data)
[1] 41 13
> #granularity
> ?unique
> sapply(data, function (x), length(unique(x)))
Error: unexpected ',' in "sapply(data, function (x),"
> sapply(data, function (x) length(unique(x)))
       Test_Specimen             Material_THK                 Leg_Size              Throat_Size
                  41                        3                        8                        8
         Weld_Length      Effective_Weld_Area         Tensile_Strength Transverse_Theo_Failure
                   4                       11                        2                       11
  Shear_Theo_Failure           Actual_Failure                MacroEtch             Break_Fusion
                  11                       40                        3                        3
     Equation_Factor
                  40
> #will help with charting, higher number means different data values, lower is usually groups
> ?grepl()
> names(data)[grepl("date|time|year|geo|location", names(data, ignore.case = TRUE))]
Error in names(data, ignore.case = TRUE) :
  2 arguments passed to 'names' which requires 1
> names(data)[grepl("date|time|year|geo|location", names(data), ignore.case = TRUE))]
Error: unexpected ')' in "names(data)[grepl("date|time|year|geo|location", names(data), ignore.case = TRUE))"
> names(data)[grepl("date|time|year|geo|location", names(data), ignore.case = TRUE)]
character(0)
> #no geo/temporal data detected
>
> #Research Q1: How accurate are the theoretical predictions to the actual failure?
> #Research Q2: How consistent is the weld performance across different weld setups, such as material thicknesses and leg size?
> Research Q3: What factors best predict weld failure strength?
Error: unexpected symbol in "Research Q3"
> #Research Q3: What factors best predict weld failure strength?
>
> #Variables:
> #Research Q1: How accurate are the theoretical predictions to the actual failure?
> Response variable: Actual_Failure - This is the real measured weld failure load, it's the outcome were trying to predict or explain
Error: unexpected symbol in "Response variable"
> Explanatory variables: Transverse_Theo_Failure and Shear_Theo_Failure - represent estimates of failure strength based on engineering principles
Error: unexpected symbol in "Explanatory variables"
> #Research Q1: How accurate are the theoretical predictions to the actual failure?
> #Response variable: Actual_Failure - This is the real measured weld failure load, it's the outcome #were trying to predict or explain
> #Explanatory variables: Transverse_Theo_Failure and Shear_Theo_Failure - represent estimates of #failure strength based on engineering principles
>
> #Research Q2: How consistent is the weld performance across different weld setups, such as material thicknesses and leg size?
> #Response Variable: Actual_Failure - This is the real measured weld failure load, it's the outcome were trying to predict or explain
> #Explanatory variables: Material_THK and Leg_Size - physical characteristics of the weld setup, changes in these parameters affect weld behavior
>
> #Research Q3: What factors best predict weld failure strength?
> #Response Variable: Actual_Failure - This is the real measured weld failure load, it's the outcome were trying to predict or explain
> #Explanatory Variables: Material_THK, Leg_Size, Throat_Size, Weld_Length, Effective_Weld_Area, Tensile_Strength, MacroEtch, Break_Fusion, Equation_Factor - these variables (physical dimensions + quality checks) might explain failure strength
>
```