**Navigating datasets in R**

* Data matrices in R are 2-dimensional, with **observations or subjects in the rows**, and **variables in the columns**. Each column contains a name, which is the name of the variable. Let's say your 2D data matrix is stored in an object called “data.”
  + To look at all the variable names in your data matrix: **names(data)**
  + To access or print out a particular variable (e.g. age) in your data matrix, use the $ sign: **data$age**
  + **Important: names in R are case-sensitive** (so data$Age won’t work)
* You can also access parts of your data matrix using brackets [] and numbers (this is called subsetting). The syntax for subsetting datasets is **data[rows, columns]**. In other words, commas separate the 2 dimensions of the matrix (rows from columns), with rows first and columns second.
  + To select all values of a dimension (i.e. you are NOT subsetting rows or NOT subsetting columns), simply leave a blank in the appropriate place (i.e. either before or after the comma).
    - To print out the whole *first column*: data[,1]
    - To print out the *first ten rows*: data[1:10,]
  + Can subset both rows and columns at the same time:
    - To print out the first ten rows and columns 11-15: data[1:10,11:15]
* You can use logical statements in combination with [] to subset your dataset. A logical statement usually involves ==, !=, >, <, etc., and returns TRUE or FALSE. You will use this if you are only investigating a subset of your observations (e.g. only male participants).
  + To select all variables, but only males (data$SEX==2) in the dataset:data[data$SEX==2,]. This selects all columns (variables) for every row (subject) that has a value of 2 in variable SEX. **Make sure the right dimensions are in the []!!**
  + To select only values of cigarettes per day that *are not NA*: data$S3AQ3B1[!is.na(data$S3AQ3B1)]