# Tutorial 09 - Variables, functions, data structures, and input/output

# Do you have R installed?

Please see your TA as soon as possible if you do not have R and RStudio installed!

# variable modes or types

Data modes in R:

- ▶ numeric (2, 13.5)
- ▶ integer (1L, 3L - L tells R to store as integer)
- ▶ complex (1+4i)
- ▶ logical (TRUE, FALSE)
- ▶ character ("Stuart")

These data modes are important because certain operations or functions only work for particular data modes.

For example, if you try to do arithmetic with characters (even if it looks like a number) you will get an error or unexpected results.

# When to use what data structure in `R`

To some extent this is based on personal preference or what a particular function in R requires as its input.

There are, however, some clear cases where one data structure is the right one for the job:

- ▶ a list is useful for information that is linked in someway (e.g. from the same analysis or observation), but differs a lot in its "shape" (vectors vs. matrices)

- ▶ a dataframe is probably most often used because it holds tabular data with different data types and is the default data structure used when data is read in from a file

# Terminal tab in RStudio

The Terminal tab in RStudio allows you to interact with bash.

This is helpful because you can use git and interact with github this way.

In OSX you should be all set by clicking on the *Terminal* tab.

In Windows you'll need to run bash (type bash and press Enter) to access the file system and git tools.

Try cloning the Exercise 7 repository using the Rstudio Terminal!

# Challenge - Starting with R

**Don't forget to decompose the problem first**

- ▶ Create a list containing a vector of 5 names and a 2x2 matrix containing the numbers 1 to 4
  - ▶ access the 4th and 5th name from the vector/list in your list/dictionary
  - ▶ access the number in the 1st row and 2nd column of the matrix
- ▶ Load the wages.csv file as a dataframe
  - ▶ access the 15th row of that dataframe with square brackets
  - ▶ find the minimum wage in the entire dataframe
  - ▶ find the gender and education level of the individual that earns the highest wage in the entire data frame

# Challenges 2

- Appending to a vector and data modes
  - Create a vector containing the numbers 10, 15, 32, 64
  - Find the product of the 2nd and 4th numbers in your vector
  - Now add your name to the end of your vector
  - Find the product of the 1st and 3rd numbers in your vector. Did you get an error? Why?
  - Try to solve the error on your own using Google

  - If you are stuck, try googling "how to force a character to numeric in R"

# Challenges 3 - Data requests

Because you have worked with the `wages.csv` data in a past challenge, your boss has asked to you to create some new datasets for co-workers

- ► Create 2 new `.csv` files to send to different co-workers
    - ► The first file should be called `femaleWages.csv` and have all of the same columns as `wages.csv`, but only contain data for females.
    - ► The second file should only have gender and wages columns, but include individuals that have 12 or more years of school.