

# Assignment 1

- Create 1 Mn random numbers using:
  - `num = round( 1e6 * runif(1e6) );` # six represents number of zeros, i.e.  $1e6 = 10^6$
  - Sort these numbers using:
    - `num = sort(num);`
  - Did it take noticeable time? Try with 2 Mn, 5 Mn and 10 Mn. What do you observe?
  - Do the following:
    - `count = 1e6; a = round( count * runif(count) ); start = Sys.time(); a = sort(a); end = Sys.time(); print(end - start);`
      - Do the above for 1 Mn, 2 Mn, ... , 10 Mn and note the time. Make a table out of it and comment on it!
  - Get help on following functions from R console: `mvnrm`, `lm`, `inner_join`, `merge`, `stargazer`.- Figure out from internet how to run system commands in R. I want to know the R function name for doing this. Write the source of your answer.

# Assignment 2

- Your default preloaded R library (called `base`) comes with a number of preloaded datasets. You can check them with `data()`;
- Use one of their datasets called: `USPersonalExpenditure`
  - You can also check for its help using `?USPersonalExpenditure`
- Load it as: `df = USPersonalExpenditure;`
- Create a new dataset showing only the percentage consumption of each category in a particular year. Also add an additional row showing overall consumption in each year.
  - Do it with using for loop
    - You may use one or two loops
  - Your answer should look like the output of below code:
    - `rbind(round( t( t(100*df) / colSums(df) ), 2 ), "Total" = colSums(df));`
  - Try to understand what I did above. You can use internet help if you wish!

- Instead of values, now find the growth rates of each category of consumption for each year
  - Use for loops!
  - Now again, compare your answer to the table: `df[, -1] / df[, -5];`
  - Do they match? What did I do above? Explain!
  - With explanation I mean what does the numbers represent and how the above code is doing that in R.
- Find help (and read) about the following functions:
  - `which`, `head`, `tail`, `rnorm`, `runif`, `cor`, `cov`
- Install `dplyr` package. Figure out how to do that from internet!

# Assignment 3

- Use the NASA dataset
  - `ns = dplyr::nasa;`
  - `ns = as.data.frame(ns);`
- Plot avg pressure for each latitude. Do also for longitude
  - Hint: You will have to `aggregate` data!
  - Do you see any pattern?
- How many `cloudlow` entries are `NA` ? Create a new data frame after excluding these. Repeat the above plots.
- Does the level of `ozone` depend on latitude or longitude? How will you find that? Make multi-color plots with legends.

- Do you expect `temperature` and surface temperature (`surftemp`) to be related? How can you find that?
- Has the average `temperature` increased over the years? What about the average variability in `temperature`? Does that also vary with `year`?
- The month of May shows least variability in both `surftemp` and `temperature`. Am I right or not? Support your answer with data!
- Comment on the relation between `surftemp` and `cloudmid`! Use plots.
- Take a random sample (w/o replacement) of 5000 points of the NASA dataset. Repeat the last 4 questions on this smaller dataset!
- Make fractional year as:
  - `ns$frac_yr = ns$year + (ns$month-1)/12;`
  - Plot average `surftemp` for each `frac_yr`. Do you see a pattern? Comment!