Notes: Tab from the lower right corner of the table makes a new row
For those familiar with Slack, we have an SMU MSDS organization on Ryver (Great way to meet your cohorts and collaborate on assignments): smudatascience.ryver.com

# R Commands:

| Command | Explanation | Example |
|---|---|---|
| args | identify the arguments to a function | args(mean) |
| mean | Average value of arguments | mean(c(1,2,3,4,5)) |
| sd | Standard deviation of arguments | sd(c(4,5,6,7,8)) |
| quantile | Divisions (quantiles) of first argument | quantile(x, probs = c(.05, .95)) |
| rnorm | Generate random nums from norm dist | rnorm(100, sd = 2) |
| summary | Get summary statistics | summary(df) |
| tail | Last 6 rows of Dataframe | tail(df) |
| str | Structure of an object. Shows you the class of variables, # of objects, and # of variables. | str(x) |
| head | First 6 rows of Dataframe | head(df) |
| dim | # of rows and columns in Dataframe | dim(df) |
| read.table and read.csv | Read in a csv file into a variable from your current working directory | ao <- read.csv("appleorange.csv", stringsAsFactors = False, header=FALSE) |
| class | Returns attribute of the variable | class(df) |
| setwd | Set current working directory | setwd("C:\\Sandbox") |
| dir.create() | Create dir in cwd | dir.create("temp") |
| getwd | Get current working directory | |
| ls | List files in cwd | |
| dir | List files in cwd | |
| file.create | Create a file | |
| list.files | List the files in your current working directory | list.files() |
| file.exists | | |
| file.info("test.R")$mode | | |

| | | |
|---|---|---|
| file.rename("test.R", "test2.R") | | |
| file.copy() | | |
| file.path() | | |
| seq | generate a sequence | seq(1:10) |
| length | get length of element | length(df) |
| seq_along | | |
| rep | | |
| paste | | paste(1:3 c("X", "Y", "Z"), sep = ""), (my_char, collapse = " ") |
| LETTERS | | |
| is.na | Logical indicator for missing elements | missing <-df[is.na(df$var) , ] #subset obs with missing var in df |
| names | Get or set character vector names for an object | names(df) <- c('v1', 'v2', 'v3') |
| identical | Test two objects for being exactly equal | |
| dim | Retrieve or set the dimension of an object. | |
| class | Find or set the class of an object (i.e. list, vector, etc) | |
| cbind | Take a sequence of vector, matrix or data-frame arguments and combine by columns or rows | |
| colnames | get column names | colnames(df) |
| abline | adds one or more straight lines through the current plot. | |
| min | Returns the smallest of any numbers | min(2, 550, 8, 95, 65) ## returns 2 |
| | | |
| median | Returns the median of any numbers | median(2, 550, 8, 95, 65) ## returns 65 |
| rm(list=ls()) | Clear your workspace | |
| quantile | produces sample quantiles corresponding to the given probabilities | |
| | | |
| max | Returns the largest of any numbers | max(2, 550, 8, 95, 65) ## returns 550 |

| | | |
|---|---|---|
| submit | | |
| Sys.Date() | returns the system's idea of the current date with and without time. | |
| %% | Modulus (remainders) | 5%%2 (returns 1) |
| barplot | Creates a bar plot with vertical or horizontal bars | barplot(height) |
| hist | computes a histogram of the given data values | hist(x) |
| fivethirtyeight | Website & R-package with datasets to play with | |
| sqldf | Package that allows you to write sql for a data frame | sqldf("select dataframe$column  from dataframe") |
| | | |
| | | |
| | | |
| | | |
| | | |
| sample | takes a sample of the specified size from the elements of x using either with or without replacement | sample(df, floor(length(df)*.7), replace = FALSE) |
| | | |
| foo <- function(x) {<br>  return (x)<br>} | A Basic function definition | |
| Library (Rserve) | Package to link Tableau & R | |
| | | |
| | | |
| | | |
| | | |
| | | |
| subset | Return subsets of vectors, matrices or data frames which meet conditions | > vec <- c(1,3,5,6,7,9,25)<br>> subset(vec, vec < 6)<br>[1] 1 3 5 |
| sum | returns the sum of all the values present in its arguments | sum(1,2,3,4,5) #returns 15<br>sum(1:2, 3:5) # also returns 15 |

| unique | returns a vector, data frame or array like x but with duplicate elements/rows removed | unique(x) |
|---|---|---|
| tabulate | takes the integer-valued vector bin and counts the number of times each integer occurs in it | tabulate(bin, nbins = 5) |
| which.max | Determines the location, i.e., index of the (first) minimum or maximum of a numeric (or logical) vector. | which.min(x) |
| match | match returns a vector of the positions of (first) matches of its first argument in its second. | match(x, table, nomatch = NA_integer_, incomparables = NULL) |
| %in% | Better version of match | x %in% table |
| replicate | is a wrapper for the common use of sapply for repeated evaluation of an expression | |
| | | |
| | | |
| | | |

# Git

| Command | Explanation | Example |
|---|---|---|
| git clone *repo-loc* | | |
| git add *file-name* | | |
| git pull | Fetch data from a remote repository and try to merge it with your commits | |
| git commit -m " *message*" | | |
| git push | Add committed changes to a remote Git repository such as GitHub | |

| | | |
|---|---|---|
| git status | Show the status of a Git repository including uncommitted changes made to files | |
| git del | | |
| git checkout | Checkout a brach | |
| git commit -am "*message*" | Add any files and commit them to your local repository with "message" for tracking reason for changes. | git commit -am "Fixed Bug #23" |
| git branch | Create and delete branch | |
| git fetch | Download objects from the remote (or another) repository | |
| git config --global user.name "Your Name" | | |
| git config --global user.email "your_email@whatever.com" | | |
| git init | To set the directory as a repository | |
| git add . | To add a file to the Git repository | |
| git log | Show a repo's commit history | |
| git log --pretty=oneline | | --max-count=2<br>git log --pretty=oneline --since='5 minutes ago'<br>git log --pretty=oneline --until='5 minutes ago'<br>git log --pretty=oneline --author=<your name><br>git log --pretty=oneline --all |
| | | git log --pretty=format:"%h %ad \| %s%d [%an]" --graph --date=short |
| git checkout master | Go back to master | |
| git tag | Bookmark particularly significant commits | |
| git config --global alias | Create alias / shortcut | **git config --global alias.st** |

| | | |
|---|---|---|
| | | **"status"**<br>- git st - Shortcut for git status<br><br>**git config --global alias.logg "log --graph --decorate --oneline --abbrev-commit --all"**<br>- git logg would display the log files in a decorated graphical style. |
| | Commonly used alias | [alias]<br>      co = checkout<br>      ci = commit<br>      br = branch<br>      df = diff<br>      st = status<br>      dfs = diff --staged<br>      logg = log --graph<br>--decorate --oneline<br>--abbrev-commit --all |
| .gitignore | Not a Git command, but a file we can add to the repository to specify which files/file types Git should ignore | |
| git merge | Mege two or more commits/branches together | |
| git remote add | Add a new remote repository to an existing project | |

**Git**:
To read: https://githowto.com/setup
To Practice CMD: https://try.github.io/levels/1/challenges/1
To Set up R-studio: https://www.r-bloggers.com/rstudio-and-github/

**Regular Expressions:**

http://regexlib.com/CheatSheet.aspx?AspxAutoDetectCookieSupport=1
More advanced:
http://www.rexegg.com/regex-quickstart.html#ref

# Regular Expressions:

| Quantifiers: | | |
|---|---|---|
| +: matches at least 1 times. | | |

| | | |
|---|---|---|
| ?: matches at most 1 times. | | |
| {n}:matches exactly n times. | | |
| {n,}:matches at least n times. | | |
| {n,m}: matches between n and m times. | | |
| value = True returns matches instead of indices | | |
| ^: matches the start of the string. | | |
| $: matches the end of the string. | | |
| . : matches any single character, | | |
| [...] : a character list, matches any one of the characters inside the square brackets. | | |
| [^...] : an inverted character list, similar to [...], but matches any characters except those inside the square brackets. | | |
| \ : suppress the special meaning of metacharacters | | |
| | : an "or" operator, matches patterns on either side of the | | | |

Grepish things:  grep, grepl, regexpr, gregexpr and regexec

***************** NOTES *****************

https://aadrake.com/command-line-tools-can-be-235x-faster-than-your-hadoop-cluster.html

Git error message "please enter a commit message to explain why this merge is necessary"

1. press "i"

2. write your merge message

3. press "esc"

4. write ":wq"

5. then press enter