MAD – Data Analysis & Biostatistics in R

Introduction - Basics

James R. Hunter, Ph.D.

DIPA, EPM, UNIFESP

4 de setembro de 2020

Section 1

Introduction to Course

What is Our Objective?

- Learn **practical** data analysis
  - Run a study from beginning to end
- Use R Language
- Learn (or refresh) basic biostatistics

Professor James Hunter

- Professor Afiliado, DIPA
- PhD., Retrovirology Laboratory, DIPA, UNIFESP
- Post-Doc, HIV Cure Project
- Prior career em business consulting and urban planning
  - Consulting & University Teaching
- Focus on Statistics and Quantitative Methods since 1973
- Work with R since 2010

Contact with the Professor

- email: jameshunterbr@gmail.com
- Twitter: @jimhunterbr
- cel: 11-9-5327-5656
- Office Hours:
    - Thurs. 14h - 16.30h
    - EP2, Rua Pedro de Toledo 669, 6th Andar Fundos

Philosophy

- The only way to learn a computer language is to write it
- The more code you write, the easier it will be
- Solve practical problems with R code

Don't Panic. . .

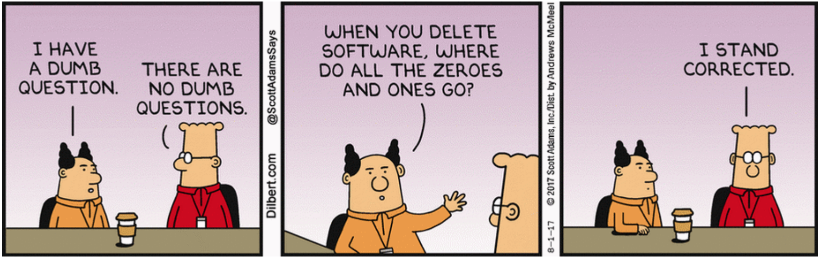
Questions

- Ask a lot of questions
- If you have a doubt, some of your classmates have it as well
- **There are NO dumb questions**

Carl Sagan on Dumb Questions

- Astrophysicist who wrote and hosted the original **Cosmos**
- Book: **The Demon-Haunted World: Science as a Candle in the Dark**
  There are naive questions, tedious questions, ill-phrased questions, questions put after inadequate self-criticism. But every question is a cry to understand the world. ***There is no such thing as a dumb question****.*

Always a Second Point of View


I HAVE A DUMB QUESTION.
THERE ARE NO DUMB QUESTIONS.
WHEN YOU DELETE SOFTWARE, WHERE DO ALL THE ZEROES AND ONES GO?
I STAND CORRECTED.
Dilbert.com @ScottAdamsSays
© 2017 Scott Adams, Inc./Dist. by Andrews McMeel 8-1-17

Resource for Questions – Piazza

- New cloud site for class questions
- Sign up: https://piazza.com/unifesp.br/fall2020/infectomad1
- Questions to prof or to each other

English??

- Why are we doing this course in English?
  - If we are in a Brazilian university?
  - If the prof speaks Portuguese? (Sim, ele fala)
- The language of science is English
  - For better or worse – it's reality
  - Publications, even Brazilian ones – English
- Like programming, the best way to perfect your English is to use it
- Course is about data analysis, not English
  - Any question can be in Portuguese
  - All submissions can be in Portuguese
- If I speak too quickly, let me know even during the class.

How Much Math Do I Need?

- What you learned in secondary education is enough
- No calculus
- Sums (Σ), logarithms and exponents
- Equation for a straight line

$$y = b_0 + b_1 x$$

Section 2

Information and Knowledge

"We are drowning in information, but we are starved for knowledge". – John Naisbitt[1]

[1] Although most frequently ascribed to futurologist John Naisbitt, this quote has many fathers and mothers. Taken here from Danielle Navarro, **Learning statistics with R: A tutorial for psychology students and other beginners**, 2020, http://compcogscisydney.org/learning-statistics-with-r

Why Do We Need Statistics and Data Analysis?

- Can we see the things we study? NO
  - Viruses, bacteria, cells, nucleotides, proteins
- Machines that produce the genomic data we study are probabilistic
  - Term "*calling* bases" - suggestion of error
- Natural process of cellular or viral replication - error prone
- Human responses to disease, drugs, treatments
  - High level of uncertainty and variability
  - Natural differences between people

Statistics Helps Find Underlying Truths

- Develop set of rules to process the information we receive
  - Script
- Form conclusions that others can understand, agree or disagree
- As post-grad students, you need to be able conduct basic analyses
  - More advanced models and methods need a specialist

Necessary Skill for All Scientists

- Understand the statistics you read in papers and books
- Separate important from unimportant
- Separate true from false
- "Call Bullshit"[2] when you are being deceived

- Result: we need probabilistic means to find underlying truths

[2] CT Bergstrom & JD West, **Calling bullshit: the art of skepticism in a data-driven world**, New York: Random House, 2020.

Four Pillars of MAD

1. Basic concepts of biostatistics
2. Organization, cleaning and practial analysis of data
3. Computational and programming tools that support data manipulation and analysis
4. Workflow necessary to execute projects

Work for the Course

- Group Project
  - Group of 2 - 4 people
  - Project based on **your** data
    - Data for a thesis, dissertation
    - Data for a project in laboratory
  - If you need data, we can find some together
  - Start with messy data (ie, real data)
    - Clean and analyze it
  - Projects topics and data sources decided before **1 October**
- Homework Assignments
  - 3 Assignments
    - Can be worked on in groups
- Participation
  - Questions/Comments/Piazza

Course Submissions

- All homeworks and projects **must** be submitted in pdf format prepared as an R Markdown document.
  - I will not accept raw program files nor raw R Markdown documents.
  - You have to get your programs to work.
- We will talk about R Markdown later today.
- All submissions emailed to Prof. Hunter

Section 3

R As Data Manipulation and Analysis Tool

CRAN: The Comprehensive R Archive Network

- An educational NGO that is the owner of the mother code for R
- Official source for copies of the base software and packages

R is a system for statistical computation and graphics. It consists of a language plus a run-time environment with graphics, a debugger, access to certain system functions, and the ability to run programs stored in script files.

History of R

- Based on a statistical programming language ("S")
  - S developed by Bell Labs in 1976
  - Still exists as commercial product
- R developed by Ross Ihaka and Robert Gentleman in 1995 in New Zealand
- Active community of developers and users
- More than 16,000 additional packages available in CRAN's repository
  - Many useful for biological analysis
  - Bioconductor – another 2,000 packages
  - Many others scattered around various sources

Virtues of R for Data Analysis

- Analyze via programs vs. clicking buttons
  - Control the sequence and options of operations in your analysis
- Programs will keep doing the same thing every day
  - No surprises because you clicked a button that changed your analysis
  - Only call for those options you understand
- Keeping a record of how you got the answer
  - Not just a record of the the answer
- **FREE** No cost, ever!
  - No stupidly expensive “student” version
  - Don’t need “cracked” copies of software

Reproducibility Crisis

- Being able to reproduce analyses over time and in different labs
- Most articles cannot be reproduced
- Nature's Reproducibility Checklist
  Workflows based on point-and-click interfaces, such as Excel, are not reproducible. Enshrine your computations and data manipulation in code.[3]
- R and Python trumps Excel, Graphpad and friends

[3] Perkel. Challenge to Scientists Nature 584, no. 7822 (2020).

Is R Hard to Learn?

- If you have never programmed before, all computer languages are hard at first
- R much easier than most
- Initial Steps
    - Specify vectors and data frames
    - Execute statistical and mathematical functions
- Today you will be writing code!
- R gets hard when you start to write your own procedures
    - When you can't find them in the packages

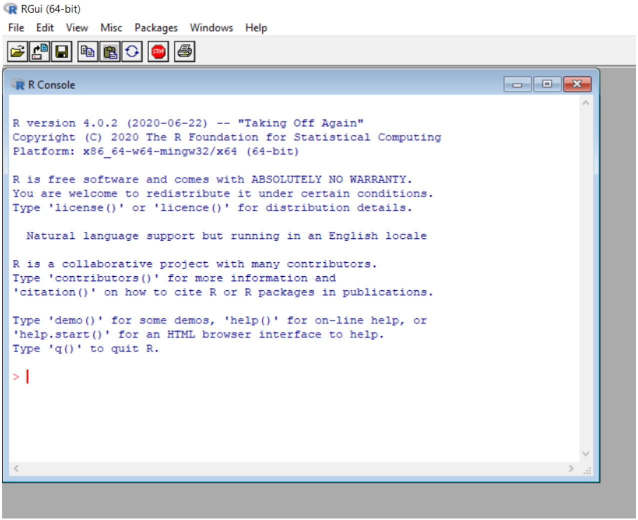
What You Need to Commit to

- Invest time in the course between classes
- Install the software (R and RStudio) on your laptops
- Read the material that suggested here and in my book
- Try out one of the basic R courses on the internet (recommended)
  - Get a second approach to the same material

RStudio – Sophisticated Communication with R

- Integrated Development Environment (“IDE”) for R
- Available since 2010
- Home of the *Tidyverse*
- Where you will do your R work
- Also **FREE**

R Has a (Useless) Graphic User Interface (“GUI”)



```
R version 4.0.2 (2020-06-22) -- "Taking Off Again"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> 
```

R & Python

- Python - another very popular language
  - Based on similar concepts to R
  - High-level interpreted language
- Launched in 1991
  - Guido van Rossum of the Netherlands
  - Name comes from English comedy group, “Monty Python’s Flying Circus”
  - Not the snake species.
- Weaker than R in statistics
  - Need commands from various modules to do basic stat operations
    - ⋆ Numpy, Pandas

Section 4

Course Resources

Course Files and Materials

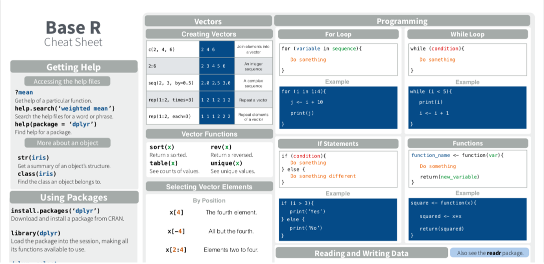
- Stored on GitHub in course repository
  - Data for exercises and lectures
  - Chapters of my text
  - Other files of interest
- https://github.com/jameshunterbr/MAD-Infecto-2020

Key Readings

- **MAD – Data Analysis & Biostatistics in R** by Prof. Hunter
  - A text on the subject of the course in preparation
  - Will provide more detail of what I cover in classes {MAD - Materia de Analise de Dados}
- Statistics texts
  - Diez, Barr & Cetinkaya-Rundel, **OpenIntro Statistics 4**
  - Navarro, D. **Learning statistics with R: A tutorial for psychology students and other beginners**
- Basic R Books
  - Wickham & Grolemund, **R for Data Science**
  - Ismay & Kim, **Statistical Inference via Data Science: A moderndive into R and the Tidyverse**
  - Irizzary, **Introduction to Data Science**

RStudio Cheat Sheets

Series of 1 and 2 page summaries for a number of key packages of R functions



Base R
Cheat Sheet

Getting Help

Accessing the help files

`?mean`
Get help of a particular function.

`help.search('weighted mean')`
Search the help files for a word or phrase.

`help(package = 'dplyr')`
Find help for a package.

More about an object

`str(iris)`
Get a summary of an object's structure.

`class(iris)`
Find the class an object belongs to.

Using Packages

`install.packages('dplyr')`
Download and install a package from CRAN.

`library(dplyr)`
Load the package into the session, making all its functions available to use.

Vectors

Creating Vectors

`c(2, 4, 6)`	2 4 6	Join elements into a vector
`2:6`	2 3 4 5 6	An integer sequence
`seq(2, 3, by=0.5)`	2.0 2.5 3.0	A complex sequence
`rep(1:2, times=3)`	1 2 1 2 1 2	Repeat a vector
`rep(1:2, each=3)`	1 1 1 2 2 2	Repeat elements of a vector

Vector Functions

`sort(x)`
Return x sorted.

`rev(x)`
Return x reversed.

`table(x)`
See counts of values.

`unique(x)`
See unique values.

Selecting Vector Elements

By Position

`x[4]`	The fourth element.
`x[-4]`	All but the fourth.
`x[2:4]`	Elements two to four.

Programming

For Loop

```
for (variable in sequence){

   Do something

}
```

Example

```
for (i in 1:4){

   j <- i + 10

   print(j)

}
```

While Loop

```
while (condition){

   Do something

}
```

Example

```
while (i < 5){

   print(i)

   i <- i + 1

}
```

If Statements

```
if (condition){
   Do something
} else {
   Do something different
}
```

Example

```
if (i > 3){
   print('Yes')
} else {
   print('No')
}
```

Functions

```
function_name <- function(var){

   Do something

   return(new_variable)
}
```

Example

```
square <- function(x){

   squared <- x*x

   return(squared)
}
```

Reading and Writing Data

Also see the **readr** package.

Online Courses

- edX - Harvard courses on R in data science taught by Irizzary
- Coursera - Johns Hopkins courses on R and R in biomedical applications
- Utrecht University (Netherlands) - Introduction to R and data
- Coursera - Duke University - sequence of R courses by Cetinkaya-Rundel

All excellent

Sites about R

- R Bloggers (https://www.r-bloggers.com/)
- Tidyverse (https://www.tidyverse.org/learn/)
- Stack Overflow (https://stackoverflow.com/questions/tagged/r)
- Twitter (#rstats)

R and RStudio Help Systems

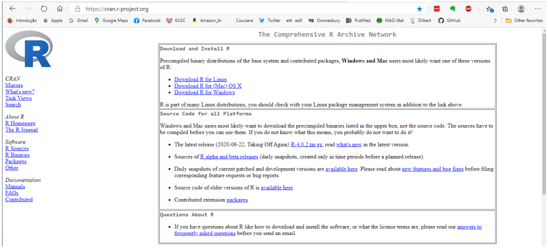
- Very complete
- Every function (command) has a help screen
- Written by geeks for geeks
  - Explanations sometimes opaque
  - Especially error messages
- Last resort: copy error message and Google it
  - Someone, somewhere has not understood the same thing that troubles you

Section 5

Installing the Software

Install R

- Found on following site:
  - https://cran.r-project.org/



The Comprehensive R Archive Network

CRAN
Mirrors
What's new?
Task Views
Search

About R
R Homepage
The R Journal

Software
R Sources
R Binaries
Packages
Other

Documentation
Manuals
FAQs
Contributed

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- Download R for Linux
- Download R for (Mac) OS X
- Download R for Windows

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

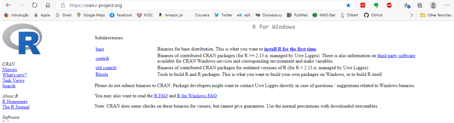
- The latest release (2020-06-22, Taking Off Again) R-4.0.2.tar.gz, read what's new in the latest version.
- Sources of R alpha and beta releases (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are available here. Please read about new features and bug fixes before filing corresponding feature requests or bug reports.
- Source code of older versions of R is available here.
- Contributed extension packages

Questions About R

- If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.

From Initial Screen (Windows)

1. Click on the link “Download R for Windows”
2. On the next screen, click on “base”
   - Mac skips this step



R for Windows

CRAN
Mirrors
What's new?
Task Views
Search

About R
R Homepage
The R Journal

Software

Subdirectories:

base	Binaries for base distribution. This is what you want to **install R for the first time**.
contrib	Binaries of contributed CRAN packages (for R >= 2.13.x; managed by Uwe Ligges). There is also information on third party software available for CRAN Windows services and corresponding environment and make variables.
old contrib	Binaries of contributed CRAN packages for outdated versions of R (for R < 2.13.x; managed by Uwe Ligges).
Rtools	Tools to build R and R packages. This is what you want to build your own packages on Windows, or to build R itself.

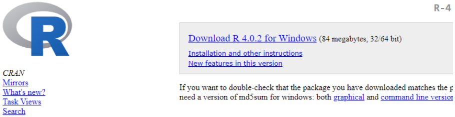
Please do not submit binaries to CRAN. Package developers might want to contact Uwe Ligges directly in case of questions / suggestions related to Windows binaries.

You may also want to read the R FAQ and R for Windows FAQ.

Note: CRAN does some checks on these binaries for viruses, but cannot give guarantees. Use the normal precautions with downloaded executables.

Real Installation

- Click on **Download R 4.0.2 for Windows**



CRAN
Mirrors
What's new?
Task Views
Search

R-4

Download R 4.0.2 for Windows (84 megabytes, 32/64 bit)

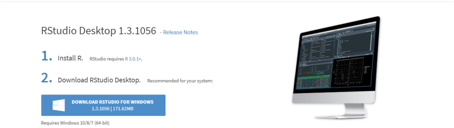
Installation and other instructions
New features in this version

If you want to double-check that the package you have downloaded matches the p
need a version of md5sum for windows: both graphical and command line versior

- Site will download the program on your computer

Installation of RStudio

- Go to site: https://www.rstudio.com/products/rstudio/download/
- Scroll down to a big, *blue* button: “Download RStudio for Windows”
  - Gives version number and size of program



RStudio Desktop 1.3.1056 - Release Notes

1. Install R. RStudio requires R 3.0.1+.

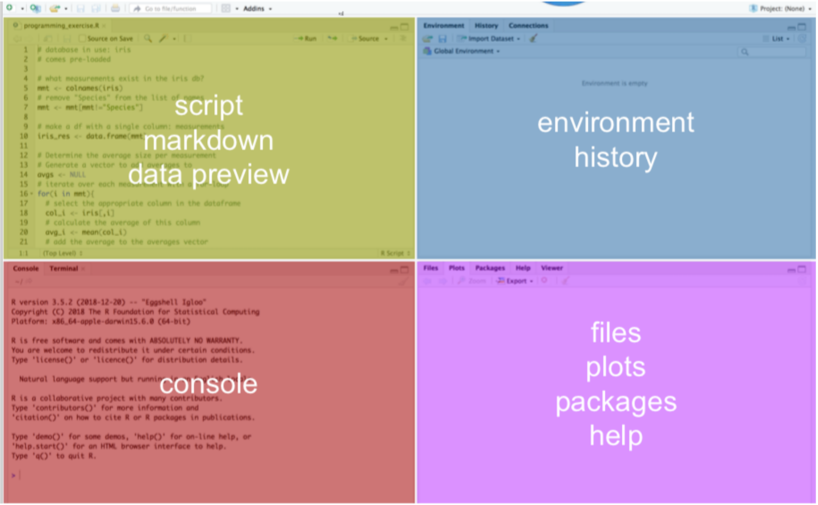
2. Download RStudio Desktop. Recommended for your system:

Requires Windows 10/8/7 (64-bit)

Start RStudio

- On Desktop (or through menus), *double click* on the icon for RStudio
  - Not the icon for R
- RStudio will open
  - R will automatically open within RStudio

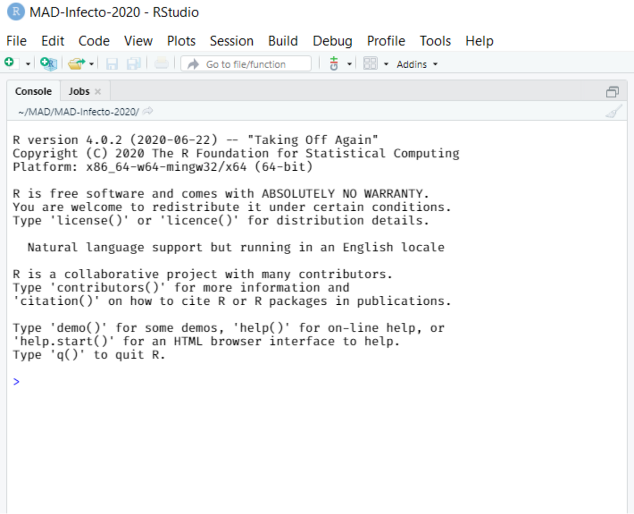
RStudio Screen



```r
# database in use: iris
# comes pre-loaded

# what measurements exist in the iris db?
mmt <- colnames(iris)
# remove "Species" from the list of names
mmt <- mmt[mmt!="Species"]

# make a df with a single column: measurements
iris_res <- data.frame(mmt)

# Determine the average size per measurement
# Generate a vector to add averages to
avgs <- NULL
# iterate over each measurement with a for-loop
for(i in mmt){
  # select the appropriate column in the dataframe
  col_i <- iris[,i]
  # calculate the average of this column
  avg_i <- mean(col_i)
  # add the average to the averages vector
```

```
~/

R version 3.5.2 (2018-12-20) -- "Eggshell Igloo"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

>
```

RStudio Console at Startup – Ready to Rock!



MAD-Infecto-2020 - RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Console
Jobs
~/MAD/MAD-Infecto-2020/

```
R version 4.0.2 (2020-06-22) -- "Taking Off Again"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

>
```

Section 6

Your First Program

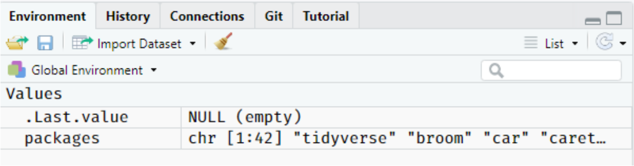
Load Packages

- Most important packages that extend base R
- We will use most during the course
- Simple script

```r
packages <- c("tidyverse", "broom", "car", "caret", "corrr", "data.table",
              "descr", "devtools", "gapminder", "ggpubr", "ggvis", "ggsci",
              "glue", "gmodels", "gt", "here", "Hmisc", "hms", "janitor",
              "jsonlite", "kableExtra","knitr", "lattice", "lubridate",
              "magrittr", "mice", "nortest", "nycflights13",
              "outliers", "palmerpenguins", "pROC", "psych", "RColorBrewer",
              "Rcpp", "readxl", "ROCR", "shiny", "styler", "summarytools",
              "titanic", "triangle", "usethis")

install.packages(packages)
```

What Script Does - Line 1

- Line 1: assignment of set of packages to the name `packages`
  - Uses <- to make the assignment
- Set of packages combined into vector of package names
  - Function `c()` creates a multi-element vector
  - `c()` - *combine* or *concatenate*
  - *vector* - one dimensional matrix
- Elements of `packages` - strings of class *character*
  - Enclosed in quotation marks ("")
- Result of Line 1



```
Values
  .Last.value     NULL (empty)
  packages        chr [1:42] "tidyverse" "broom" "car" "caret…
```

Note: Assignment Operators

- Principal assignment operator: <-
- Discourage use of =
  - **You will confuse it with logical equals ==**
    - Guaranteed! We all do it

What Script Does - Line 2

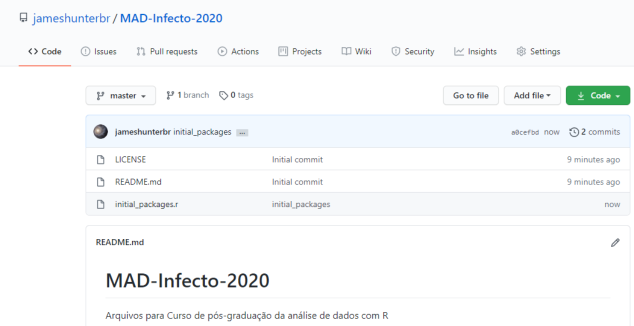
- Installs the packages
  - Goes out to CRAN mirror site
  - Downloads and installs packages
  - Many of the packages have dependencies so will install more packages
  - Dependencies: other packages needed for functions of calling package

Use of Scripts vs. Use of Console

- Write your commands in a script in R Markdown rather than Console
    - You can save your work
- Console is where commands are executed
    - Saving your history from Console more complicated

Where Do I Find Script?

- **GitHub**
  - Public face of version control system called *git*
  - Maintain a clear record of changes to scripts
    - On Computer
    - In remote repository
- GitHub repository for course
  - https://github.com/jameshunterbr/MAD-Infecto-2020



jameshunterbr / MAD-Infecto-2020

jameshunterbr initial_packages ... a0cefbd now 2 commits

LICENSE	Initial commit	9 minutes ago
README.md	Initial commit	9 minutes ago
initial_packages.r	initial_packages	now

README.md

MAD-Infecto-2020

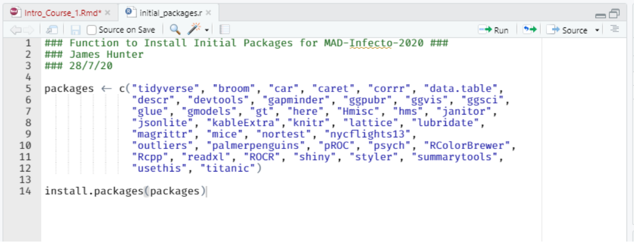
Arquivos para Curso de pós-graduação da análise de dados com R

How Do I Download Script?

- Click on script name: “initial_packages.r”
    - Text file will appear
- Click on **Raw** button
- Right-click on file and save it to your R directory
    - Use “Save Page (As) . . . ” command on pop-up menu

Execute “initial_packages.r”

- Files tab of lower right pane of RStudio
  - Click on initial-packages.r
- Script will open in upper left pane
- Click on Source button in program menu bar
- Follow progress in Console



```r
### Function to Install Initial Packages for MAD-Infecto-2020 ###
### James Hunter
### 28/7/20

packages ← c("tidyverse", "broom", "car", "caret", "corrr", "data.table",
              "descr", "devtools", "gapminder", "ggpubr", "ggvis", "ggsci",
              "glue", "gmodels", "gt", "here", "Hmisc", "hms", "janitor",
              "jsonlite", "kableExtra","knitr", "lattice", "lubridate",
              "magrittr", "mice", "nortest", "nycflights13",
              "outliers", "palmerpenguins", "pROC", "psych", "RColorBrewer",
              "Rcpp", "readxl", "ROCR", "shiny", "styler", "summarytools",
              "usethis", "titanic")

install.packages(packages)
```

Section 7

Basic Operations in R

Use R as a Calculator

```r
5 + 5
```

```
## [1] 10
```

```r
36 * 2500000
```

```
## [1] 90000000
```

```r
5876/35.44320
```

```
## [1] 165.7864
```

```r
2^25    # exponent
```

```
## [1] 33554432
```

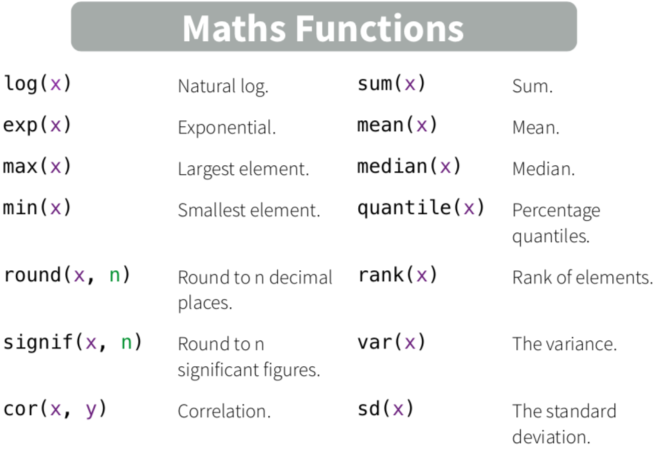
```r
25 * (12 + 27)
```

```
## [1] 975
```

Math Functions in R

Function	What It Does
abs(x)	absolute value of x
sqrt(x)	square root of x
log(x)	natural (Naperian) logarithm of x
exp(x)	natural exponent of x
log10(x)	logarithm base 10 of x
round(x, n)	round x to n decimal places

More Math Functions



Maths Functions

`log(x)`	Natural log.	`sum(x)`	Sum.
`exp(x)`	Exponential.	`mean(x)`	Mean.
`max(x)`	Largest element.	`median(x)`	Median.
`min(x)`	Smallest element.	`quantile(x)`	Percentage quantiles.
`round(x, n)`	Round to n decimal places.	`rank(x)`	Rank of elements.
`signif(x, n)`	Round to n significant figures.	`var(x)`	The variance.
`cor(x, y)`	Correlation.	`sd(x)`	The standard deviation.

Functions at Work

```
abs(-287)
```

```
## [1] 287
```

```
sqrt(9849)
```

```
## [1] 99.24213
```

```
log(377898)
```

```
## [1] 12.84238
```

```
exp(12.84238)
```

```
## [1] 377898.2
```

```
log10(377898)
```

```
## [1] 5.577375
```

```
round(exp(12.84238), 0)
```

```
## [1] 377898
```

Note about `log()` and `exp()`

- In example above, exponent of 12.84238 is 377898.2, not 377898
- R reports 5 decimal places on the screen
    - Internally, it is 12.8423795969182 (13 decimal places)
- We know that $log(x) = e^x$
- We haven't broken any (major) mathematical laws.

```
x <- 377898
y <- log(x) # calculate the log of x and assign it to y
y
```

```
## [1] 12.84238
```

```
exp(y)
```

```
## [1] 377898
```

Comments

- Line 2 of the script above has a comment after it
- Comments start with a hashtag #
    - Everything after it on a line is not interpreted
- Comments remind us what we have done and why we did it
- **Very important**
- Use frequently

Order of Calculation (*PEMDAS*)

Operation	Symbol	Example	PEMDAS
parentheses	()	$5 * (7 + 2) = 45$	P
exponents	ˆ	5ˆ2 $= 25$	E
multiplication	*	$5 * 7 = 35$	M
division	/	$25/5 = 5$	D
addition	+	$5 + 7 = 12$	A
subtraction	-	$5 - 7 =$ -2	S

- If you remove the parentheses from `5 * (7 + 2)`?
- `5 * 7 + 2` $= 37$
- Remember: rules of mathematics don't change because we are using a computer

Assignment

- (name of object) <- (definition of object)
- definition = the values that are the content of the object

Assignment – Styles

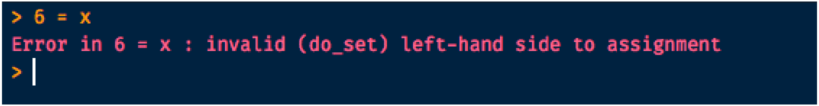
- These work

```
x <- 6

x <- "Hi!"
```

- These work but are not recommended

```
x = 6

6 -> x
```

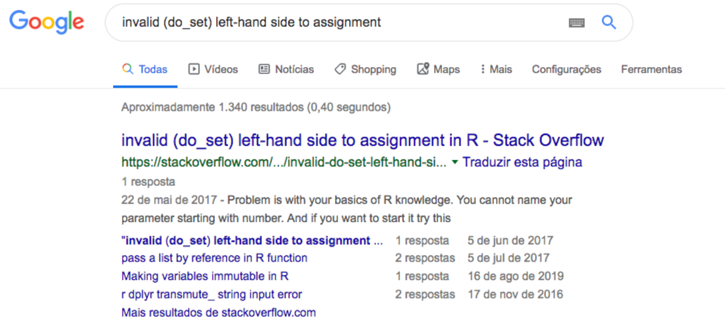
- This produces an error (cannot start a command with a number)



```
> 6 = x
Error in 6 = x : invalid (do_set) left-hand side to assignment
> 
```

What Do You Do When You See a Strange Error Message?

- Google It



invalid (do_set) left-hand side to assignment

Aproximadamente 1.340 resultados (0,40 segundos)

invalid (do_set) left-hand side to assignment in R - Stack Overflow

https://stackoverflow.com/.../invalid-do-set-left-hand-si... ▾ Traduzir esta página

1 resposta

22 de mai de 2017 - Problem is with your basics of R knowledge. You cannot name your parameter starting with number. And if you want to start it try this

"invalid (do_set) left-hand side to assignment ...	1 resposta	5 de jun de 2017
pass a list by reference in R function	2 respostas	5 de jul de 2017
Making variables immutable in R	1 resposta	16 de ago de 2019
r dplyr transmute_ string input error	2 respostas	17 de nov de 2016

Mais resultados de stackoverflow.com

Section 8

Assignment – Variable Names

Assignment – Variable Names

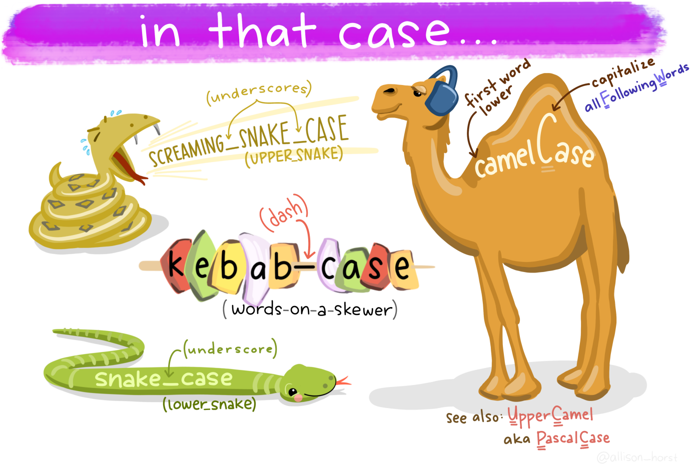
- Primary rules of R

1. Must contain only letters (either upper or lower case), numbers or symbols `.` or `_`.
2. Must start name with a **letter**

Variable Names – Corollaries

- Should not include spaces.
  - “Snake case” overcomes this restriction
    - Connecting words with underscore "_"
- R reserved words cannot be used for variable names
  - Examples: `TRUE`, `FALSE`, `if`, `else`, `for`, `function`
- Variable names case sensitive
  - `Variable` and `variable` are 2 different names
  - Same for `x` and `X`

Cases in R


in that case...
(underscores)
SCREAMING_SNAKE_CASE
(UPPER_SNAKE)
(dash)
kebab-case
(words-on-a-skewer)
(underscore)
snake_case
(lower_snake)
first word lower
capitalize all Following Words
camelCase
see also: UpperCamel
aka PascalCase
@allison_horst

More on Variable Names

- Make them clear and informative
  - x, although popular, is useless as a name

```
## 1st version
peso <- 55  ## Person weighs 55 kg.

## 2nd version
peso_kg <- 55 ## Clearer

## 3rd version, can convert to pounds
peso_lb <- peso_kg * 2.2
peso_lb
```

```
## [1] 121
```

Variable Names – Last Shot

- Make a data dictionary
  - Record of what your variable names are, what kind of data they are and range of values
- Try to keep names as short as possible
- Camel case as alternative to snake case
- If you surround your variable name with single quote (') or backtick (‘), spaces ok
  - `viral load` **illegal**
  - ‘viral load’ **legal**
  - But, don't use this

Section 9

Go to Presentation 2