Introduction to Statistical Software – R language

Asad Haris

5 September, 2019

Overview of R

- R is a programming language built upon the S-language developed at AT&T Bell Laboratories.
- It is a free software environment for statistical computing and graphics.
- To learn more about the history of R see:
    - R-project page: https://www.r-project.org/about.html
    - Wikipedia page: https://en.wikipedia.org/wiki/R_(programming_language)

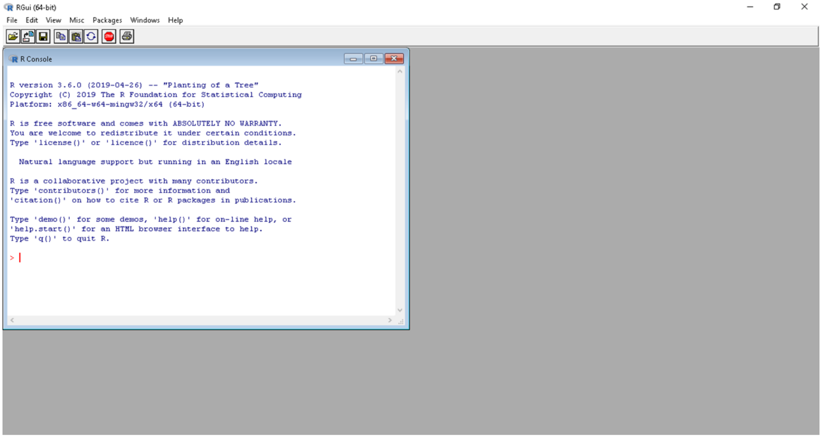
Why R?

- Free, open-source, multi-platform
- Excellent community of users and developers
  - The number of things you can do with a single line of code is continuously growing through new packages (more on this later)
- Has an excellent integrated development environment (IDE): **Rstudio**
  - Makes working with base-R and other add-ons more user friendly
- A programming language that can be used on a spectrum
  - Has tools to do almost point-and-click statistical analysis
  - Flexibility of writing your own code/functions/analysis

Installing base-R

- Available for multiple platforms
  - Windows ($\geq$ Windows 7)
  - OS X (MAC)
  - Linux (ubuntu, debian, redhat, suse)
- Download required files from: https://cran.r-project.org/
  - For MAC, you need to download the .pkg file, open and follow instructions
  - For Windows, you need the .exe file, open and follow instructions

Base R console



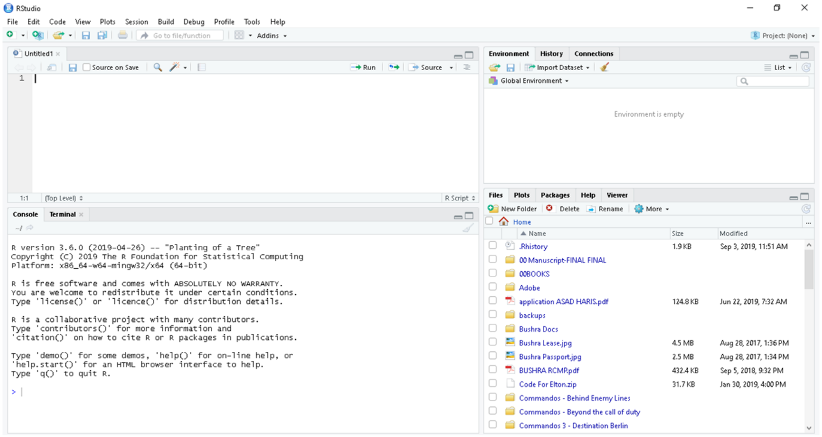
```
R version 3.6.0 (2019-04-26) -- "Planting of a Tree"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

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.

> 
```

Figure 1: Base-R GUI

Installing RStudio

- RStudio is the free R integrated development environment (IDE) that is available from https://www.rstudio.com/products/rstudio/
- It has useful features such as syntax highlighting and tab for suggested code auto-completion.
- It involves four-pane workspace, which better manages multiple R windows for typing commands, storing scripts, viewing command histories, viewing visualizations and more.
- We will work with Rstudio to do things in R for this course.

Rstudio IDE

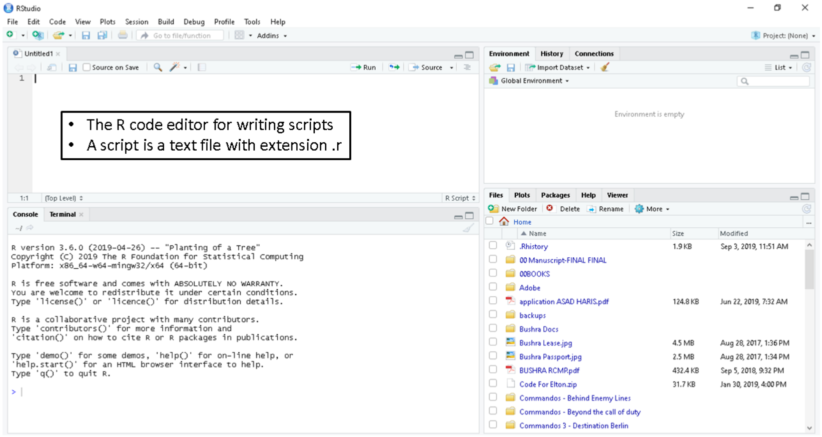


```
R version 3.6.0 (2019-04-26) -- "Planting of a Tree"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

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.

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.

>
```

Name	Size	Modified
.Rhistory	1.9 KB	Sep 3, 2019, 11:51 AM
00 Manuscript-FINAL FINAL		
00BOOKS		
Adobe		
application ASAD HARIS.pdf	124.8 KB	Jun 22, 2019, 7:32 AM
backups		
Bushra Docs		
Bushra Lease.jpg	4.5 MB	Aug 28, 2017, 1:36 PM
Bushra Passport.jpg	2.5 MB	Aug 28, 2017, 1:34 PM
BUSHRA RCMP.pdf	432.4 KB	Sep 5, 2018, 9:32 PM
Code For Elton.zip	31.7 KB	Jan 30, 2019, 4:00 PM
Commandos - Behind Enemy Lines		
Commandos - Beyond the call of duty		
Commandos 3 - Destination Berlin		

Figure 2: Rstudio IDE

Rstudio IDE



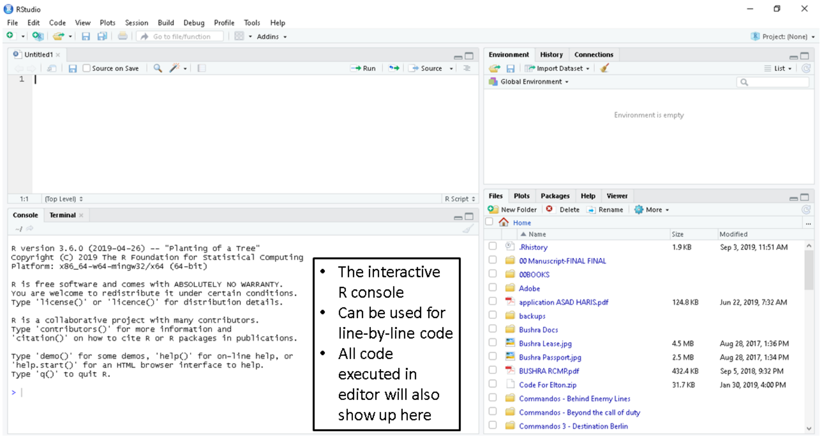
- The R code editor for writing scripts
- A script is a text file with extension .r

```
R version 3.6.0 (2019-04-26) -- "Planting of a Tree"
Copyright (C) 2019 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.

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.

> |
```

Name	Size	Modified
.Rhistory	1.9 KB	Sep 3, 2019, 11:51 AM
00 Manuscript-FINAL FINAL		
00BOOKS		
Adobe		
application ASAD HARIS.pdf	124.8 KB	Jun 22, 2019, 7:32 AM
backups		
Bushra Docs		
Bushra Lease.jpg	4.5 MB	Aug 28, 2017, 1:36 PM
Bushra Passport.jpg	2.5 MB	Aug 28, 2017, 1:34 PM
BUSHRA RCMP.pdf	432.4 KB	Sep 5, 2018, 9:32 PM
Code For Elton.zip	31.7 KB	Jan 30, 2019, 4:00 PM
Commandos - Behind Enemy Lines		
Commandos - Beyond the call of duty		
Commandos 3 - Destination Berlin		

Figure 3: Editor window

Rstudio IDE



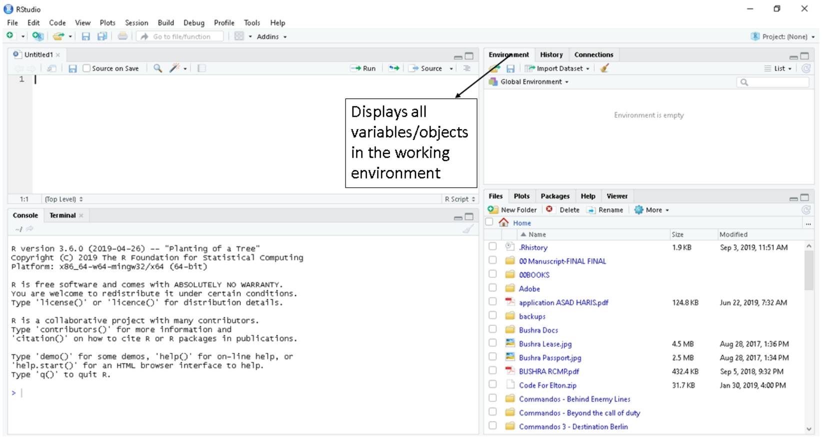
```
R version 3.6.0 (2019-04-26) -- "Planting of a Tree"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

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.

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.

> |
```

- The interactive R console
- Can be used for line-by-line code
- All code executed in editor will also show up here

Name	Size	Modified
.Rhistory	1.9 KB	Sep 3, 2019, 11:51 AM
00 Manuscript-FINAL FINAL		
00BOOKS		
Adobe		
application ASAD HARIS.pdf	124.8 KB	Jun 22, 2019, 7:32 AM
backups		
Bushra Docs		
Bushra Lease.jpg	4.5 MB	Aug 28, 2017, 1:36 PM
Bushra Passport.jpg	2.5 MB	Aug 28, 2017, 1:34 PM
BUSHRA RCMP.pdf	432.4 KB	Sep 5, 2018, 9:32 PM
Code For Elton.zip	31.7 KB	Jan 30, 2019, 4:00 PM
Commandos - Behind Enemy Lines		
Commandos - Beyond the call of duty		
Commandos 3 - Destination Berlin		

Figure 4: Console window

Rstudio IDE



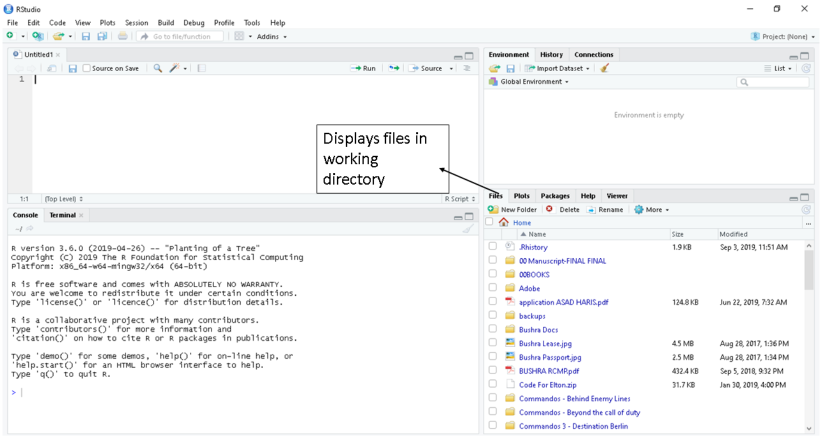
Displays all variables/objects in the working environment

```
R version 3.6.0 (2019-04-26) -- "Planting of a Tree"
Copyright (C) 2019 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.

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.

>
```

Name	Size	Modified
.Rhistory	1.9 KB	Sep 3, 2019, 11:51 AM
00 Manuscript-FINAL FINAL		
00BOOKS		
Adobe		
application ASAD HARIS.pdf	124.8 KB	Jun 22, 2019, 7:32 AM
backups		
Bushra Docs		
Bushra Lease.jpg	4.5 MB	Aug 28, 2017, 1:36 PM
Bushra Passport.jpg	2.5 MB	Aug 28, 2017, 1:34 PM
BUSHRA RCMP.pdf	432.4 KB	Sep 5, 2018, 9:32 PM
Code For Elton.zip	31.7 KB	Jan 30, 2019, 4:00 PM
Commandos - Behind Enemy Lines		
Commandos - Beyond the call of duty		
Commandos 3 - Destination Berlin		

Figure 5: Environment

Rstudio IDE

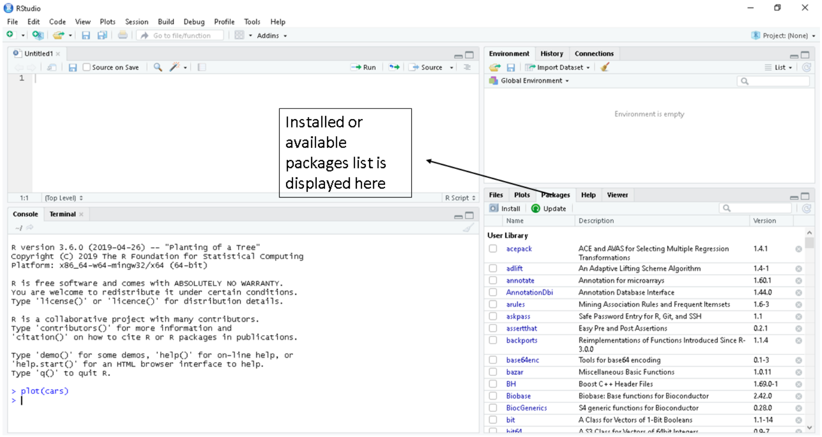


```
R version 3.6.0 (2019-04-26) -- "Planting of a Tree"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

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.

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.

> 
```

Name	Size	Modified
.Rhistory	1.9 KB	Sep 3, 2019, 11:51 AM
00 Manuscript-FINAL FINAL		
00BOOKS		
Adobe		
application ASAD HARIS.pdf	124.8 KB	Jun 22, 2019, 7:32 AM
backups		
Bushra Docs		
Bushra Lease.jpg	4.5 MB	Aug 28, 2017, 1:36 PM
Bushra Passport.jpg	2.5 MB	Aug 28, 2017, 1:34 PM
BUSHRA RCMP.pdf	432.4 KB	Sep 5, 2018, 9:32 PM
Code For Elton.zip	31.7 KB	Jan 30, 2019, 4:00 PM
Commandos - Behind Enemy Lines		
Commandos - Beyond the call of duty		
Commandos 3 - Destination Berlin		

Figure 6: Files

Rstudio IDE



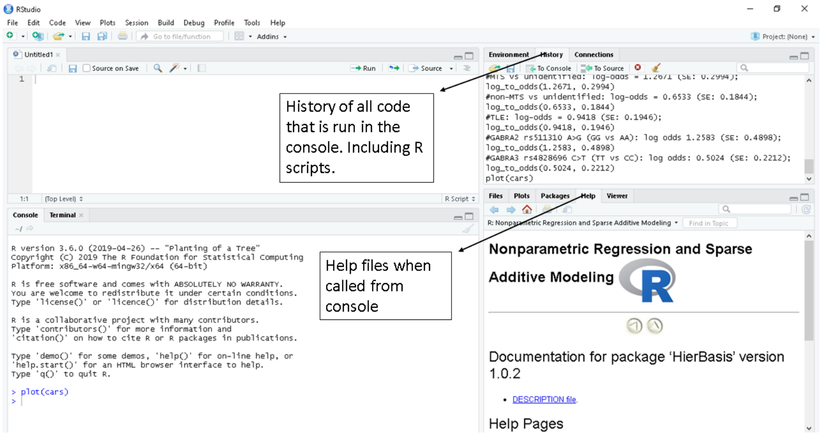
```
R version 3.6.0 (2019-04-26) -- "Planting of a Tree"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

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.

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.

> plot(cars)
>
```

User Library

Name	Description	Version
acepack	ACE and AVAS for Selecting Multiple Regression Transformations	1.4.1
adlift	An Adaptive Lifting Scheme Algorithm	1.4-1
annotate	Annotation for microarrays	1.60.1
AnnotationDbi	Annotation Database Interface	1.44.0
arules	Mining Association Rules and Frequent Itemsets	1.6-3
askpass	Safe Password Entry for R, Git, and SSH	1.1
assertthat	Easy Pre and Post Assertions	0.2.1
backports	Reimplementations of Functions Introduced Since R-3.0.0	1.1.4
base64enc	Tools for base64 encoding	0.1-3
bazar	Miscellaneous Basic Functions	1.0.11
BH	Boost C++ Header Files	1.69.0-1
Biobase	Biobase: Base functions for Bioconductor	2.42.0
BiocGenerics	S4 generic functions for Bioconductor	0.28.0
bit	A Class for Vectors of 1-Bit Booleans	1.1-14
bit64	A S3 Class for Vectors of 64bit Integers	0.9-7

Figure 7: Package list

Rstudio IDE



History of all code that is run in the R console. Including R scripts.

Help files when called from console

```
#MTS vs unidentified: log-odds = 1.2671 (SE: 0.2994);
log_to_odds(1.2671, 0.2994)
#non-MTS vs unidentified: log-odds = 0.6533 (SE: 0.1844);
log_to_odds(0.6533, 0.1844)
#TLE: log-odds = 0.9418 (SE: 0.1946);
log_to_odds(0.9418, 0.1946)
#GABRA2 rs511310 A>G (GG vs AA): log odds 1.2583 (SE: 0.4898);
log_to_odds(1.2583, 0.4898)
#GABRA3 rs4828696 C>T (TT vs CC): log odds: 0.5024 (SE: 0.2212);
log_to_odds(0.5024, 0.2212)
plot(cars)
```

```
R version 3.6.0 (2019-04-26) -- "Planting of a Tree"
Copyright (C) 2019 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.

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.

> plot(cars)
>
```

Nonparametric Regression and Sparse Additive Modeling

Documentation for package ‘HierBasis’ version 1.0.2

- DESCRIPTION file.

Help Pages

Figure 8: Help and History

Rstudio Shortcuts

- Keyboard shortcuts can make coding faster and efficient
- IDE displays list of shortcuts tools-> Keyboard Shortcuts Help or press Alt+Shift+K (Option+Shift+K on Mac)
- Some useful shortcuts to remember:
  - **tab** key for auto compltete code. In console or editor start typing and press **tab** key to display list of suggestions.
  - **Ctrl+Enter (Command+Enter on Mac)** runs the current line of code. If multiple lines selected/highlighted all selected lines will run in the console.
  - **Ctrl+S (Command+S on Mac)** saves the file in editor, good to continously save work incase R or Rstudio crashes (don't worry, it happens to everyone).

First steps in R

- First we want to set a working directory.
- You can use the default (usually where R is installed or $HOME) however the working directory is where any exported results/reports/tables/plots will be saved.
- Having raw data in the working directory also makes importing data easy.

```r
setwd("C:/Users/Asad/Dropbox")
setwd("~/")
setwd("~/work")
```

- In Studio, we can also use Session -> Set Working Directory

Basic calculations in R

- To start, we can use R as a calculator. All simple operations: +, -, *, /, ^

```
4+2
## [1] 6
7-8*(2/3)
## [1] 1.666667
2^4
## [1] 16
```

- R also comes with some built-in math functions

```
abs(-5)
## [1] 5
sin(pi/2)
## [1] 1
log2(10)
## [1] 3.321928
```

Functions and pacakges

- We have already used **functions** in R, i.e. `setwd()`, `abs()`, `log2()` are all examples of functions.
- A function in R (or any programming language) is like a function in math:
  - It is *something* which takes in certain input, performs some operations on this input and returns output
  - We have seen some built-in functions, but we can define our own functions or use functions defined by others.
  - Developers and users can write useful functions and share them as R-packages
- An R-package is a collection of functions, data and compiled code.
  - There are MANY R-packages out there!
  - It is extremely helpful for methods papers to have an accompanying package.
  - For any project/paper making an R-package helps *reproducibility*.

Installing and loading packages

- We have seen that Rstudio shows list of installed packages.
- We can install packages using the Rstudio GUI or in command line:

```
install.packages("the-package-name")

install.packages("ggformula")
install.packages("mosaic")
```

- Once installed we can need to *load* a package to use its functions:

```
load(the-package-name)
```

- Packages need to be re-loaded everytime we start a new session.

Rmarkdown: tools for reproducible research

- Markdown is a *markup language*, a system that can be used to create documents.
  - Purpose of this was to have a simple syntax markup language
  - Rmarkdown, builds upon that and allowing us to incorporate R (and even other programming languages!)
- Simple Rmarkdown can be used for creating simple documents. But it can do much more:
  - Create a variety of formatted documents in multiple formats (pdf, html, etc)
  - Allows users to easily insert tables, lists, math equations
  - Can be used to write papers for journals/conferences
  - Commonly used to write software manuals for R packages
  - Your EPIB 607 assignments
  - These slides were made in Rmarkdown!

Rmarkdown in action

- Easy quickstart: use the template in Rstudio
- Using the simple markdown syntax we can create documents
  - For a quick start: https://rmarkdown.rstudio.com/authoring_basics.html
  - In “Help -> Cheatsheets”, we have a reference guide and a cheatsheet.
- We can include *code chunks*
  - These literally have chuncks of code which are executed when the document is *compiled*
  - We can add options various options for code chuncks which leads to different results (both aesthetically and code output)
- A recent addition to Rstudio is R Notebooks
  - R Notebooks are rmarkdown documents but now the code is executed on-the-fly.
- A simple Rmarkdown document