GRAD 778: Reproducible Research

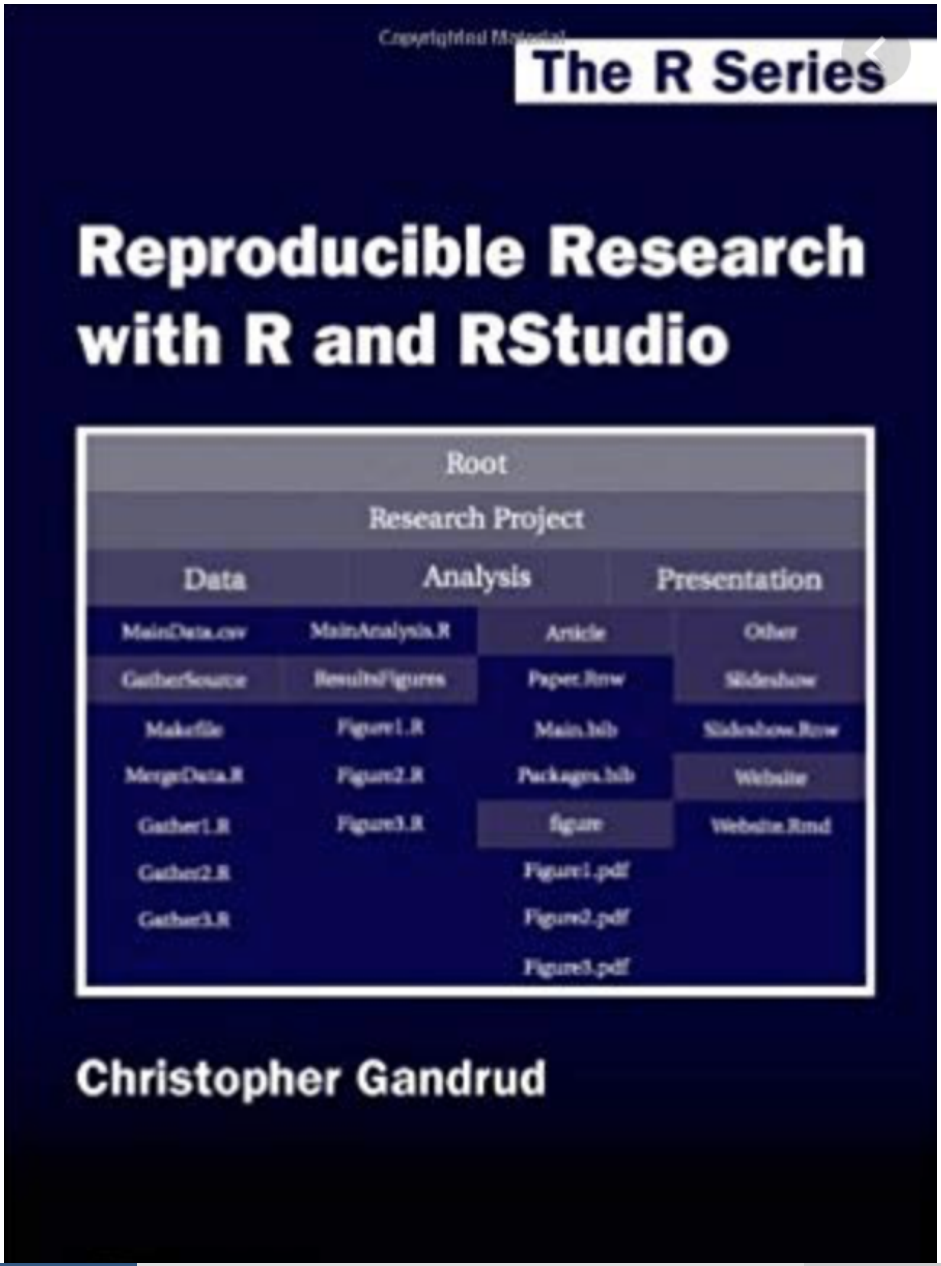
Perry J. Williams

Department of Natural Resources and Environmental Science
University of Nevada, Reno

October 2, 2021

Background & Motivation

Background & Motivation



Background & Motivation

Research is often presented in very abridged packages

- Slide shows
- Journal articles
- Books
- Web sites

Background & Motivation

These presentation documents announce a project's findings

Background & Motivation

These documents are not the research

- These documents are the advertising
- Especially true in computational and statistical sciences

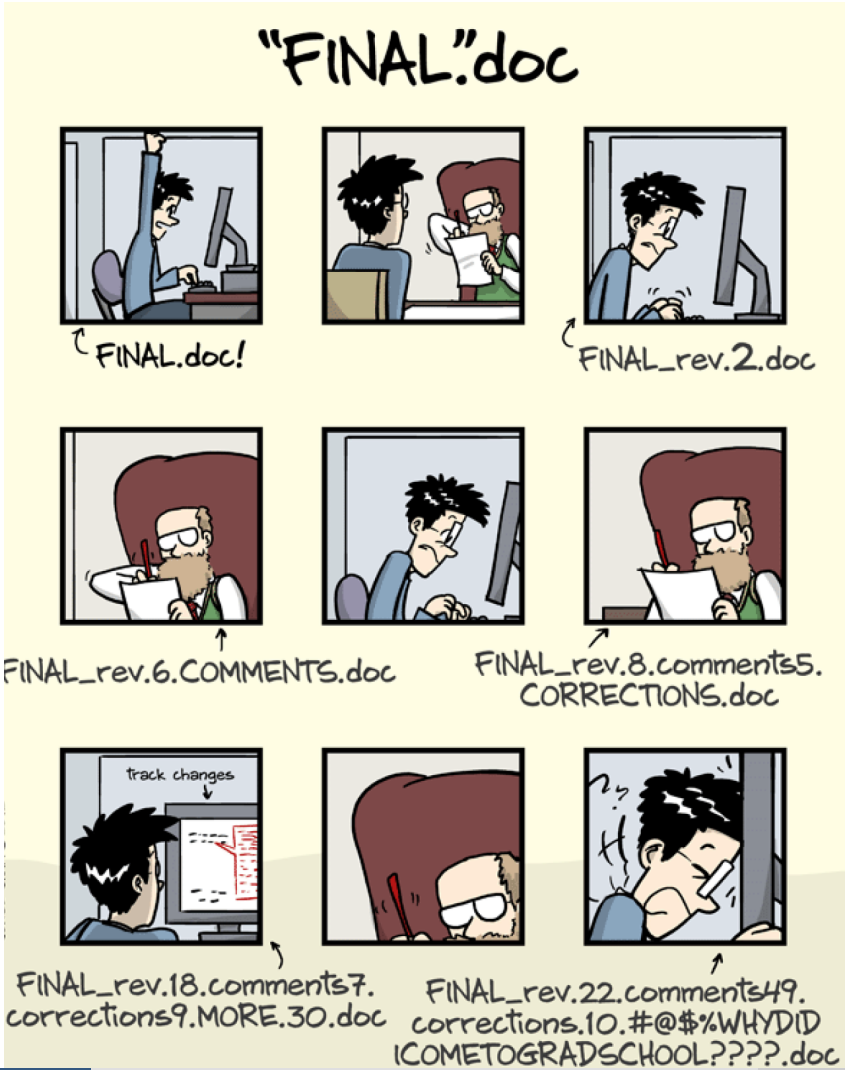
Background & Motivation

The research includes:

- Full software environment
- Code
- Data

Background & Motivation

When we separate the research from its advertisement we make it difficult for other to reproduce our findings



Background & Motivation

This workshop will introduce:

- The tools to dynamically combine research with presentation of findings,
- The `R` statistical language for data analysis,
- the LaTeX mark-up language for documents, slide shows, articles, books, and web-pages,
- the `knitr` package for `R`,
- **RStudio**, a program that brings all of these tools together in one place.

Objective

The objective of this workshop is to:

- Introduce the tools to develop a work-flow to maximize reproducible-ness, collaborations, and research impact.
- Provide templates that can be modified for your own research.

The objective of this workshop is not to:

- Become well-versed in `R`, `RStudio`, LaTeX, or `knitr` - that takes repetition (starting with the basic building blocks that are provided).

Background & Motivation

Additional topics include:

- Version control with Git hub,
- Data gathering,
- R markdown,
- File management,
- Projects in RStudio,
- Using LaTeXto make presentations with Beamer.

All are covered in the book: *Reproducible Research with R and RStudio*

Why R?

- Open Source and free
- Very active development community
- Interfaces with LaTeXor other mark-up languages
- Explicitly write down analyses steps as source code

Why `knitr`?

- Literate programming is a crucial part of reproducible quantitative research
- Highlights `R` code in presentation documents making it easier for readers to follow
- Provides control over inclusion of graphics
- Can cache (save output for later)

Why RStudio?

- Stand alone editor for TEXand Markdown
- Many shortcuts
- Works with C++, CSS, JavaScript, and a few other programming languages
- Integrated with version control of Git and SVN
- Simple compiling of .Rnw files
- **Easier to learn than Emacs or vi!**

What is Reproducible Research?

Research results are replicable if there is sufficient information available for independent researchers to make the same findings using the same procedures (King, 1995, 444).

What is Reproducible Research?

Research results are replicable if there is sufficient information available for independent researchers to make the same findings using the same procedures (King, 1995, 444).

In computational sciences, this means:
The data and code used to make a finding are available and they are sufficient for an independent researcher to recreate the finding.

Getting Started with R

Using R: the Basics

- objects & assignment,
- component selection,
- functions and commands,
- arguments,
- the workspace,
- packages.

Objects & Assignment

- R is an "object-oriented language"
- Objects are analogous to nouns
- "object-oriented:" R is focused on doing actions to objects

Create Objects

```
Number=10
```

To see the contents of our object, type its name.

```
Number
```

```
[1] 10
```

Create Objects

Create a character string

```
Words="Hello World!"
```

To see the contents of our object, type its name.

```
Words
```

```
[1] "Hello World!"
```

Vectors

Create a numeric vector called `foo`

```
foo=c(2,4,6,8,10,12,14,16,18)
```

To see the contents of our object, type its name.

```
foo
```

```
[1]  2  4  6  8 10 12 14 16 18
```

Vectors

Create a vector of strings called `bar`

```
bar=c("the","quick", "brown","fox", "jumps",
"over","the","lazy","dog")
```

To see the contents of our object, type its name.

```
print(bar)
```

```
[1] "the"   "quick" "brown" "fox"   "jumps" "over"  "the"
    "lazy"  "dog"
```

Matrices

Create a 9×2 matrix called `baz.m` using `foo` and `bar`

```
baz.m=cbind(foo,bar)
```

To see the contents of our object, type its name.

```
print(baz.m)
```

```
      foo  bar
 [1,] "2"  "the"
 [2,] "4"  "quick"
 [3,] "6"  "brown"
 [4,] "8"  "fox"
 [5,] "10" "jumps"
 [6,] "12" "over"
 [7,] "14" "the"
 [8,] "16" "lazy"
 [9,] "18" "dog"
```

Data frames

Create a 9×2 data frame called `baz.df` using `foo` and `bar`

```
baz.df=data.frame(foo,bar)
```

To see the contents of our object, type its name.

```
print(baz.df)
```

```
  foo   bar
1   2   the
2   4 quick
3   6 brown
4   8   fox
5  10 jumps
6  12  over
7  14   the
8  16  lazy
9  18   dog
```

```
names(baz.df)
```

```
[1] "foo" "bar"
```

Data frames

Assign row names

```
row.names(baz.df)=c("Row1","Row2","Row3","Row4","Row5",
                    "Row6","Row7","Row8","Row9")
row.names(baz.df)
```

```
[1] "Row1" "Row2" "Row3" "Row4" "Row5" "Row6" "Row7" "Row8"
    "Row9"
```

Component Selection: Sub-scripts

Select rows 3–7 from baz.df

```
baz.df[3:7,]
```

```
     foo   bar
Row3   6 brown
Row4   8   fox
Row5  10 jumps
Row6  12  over
Row7  14   the
```

Select rows 3–7 and column 2 from baz.df

```
baz.df[3:7,2]
```

```
[1] brown fox   jumps over  the
Levels: brown dog fox jumps lazy over quick the
```

```
is.factor(baz.df[3:7,2])
```

```
[1] TRUE
```

Functions and Commands

If objects are the nouns, functions and commands are the verbs

```
mean(baz.df[,1])
```

```
[1] 10
```

```
mean(baz.df[2:7,1])
```

```
[1] 9
```

```
?mean
```

The workspace

Use the `ls()` command to list all objects in your current workspace

```
ls()
```

```
[1] "bar"    "baz.df" "baz.m"  "foo"    "Number" "Words"
```

The workspace

Use the `rm()` command to remove objects in your current workspace

```
rm(baz.m)
```

Save the workspace

Use the `save.image()` command to save your current workspace

```
save.image(file="ShortCourseWorkspace.RData")
```

Clear the workspace

Use the `rm(list=ls())` command to remove ALL objects in your current workspace

```
rm(list=ls())
```

Load the workspace

Use the `load()` command to load your saved workspace

```
load(file="ShortCourseWorkspace.RData")
ls()
```

```
[1] "bar"     "baz.df" "foo"     "Number" "Words"
```

Save a specific object

Use the `save()` command to save an object that was computationally intensive

```
tock=Sys.time()
nr=2000
nc=2000
n=nr*nc
foo.m=matrix(rnorm(n),nr,nc)
bar.m=solve(foo.m)  # O(n^3) complexity
tick=Sys.time()
tick-tock
```

```
Time difference of 7.161771 secs
```

```
save(bar.m,file="MatrixInverse.RData")
rm(bar.m)
load(file="MatrixInverse.RData")
ls()
```

```
 [1] "bar"     "bar.m"   "baz.df" "foo"     "foo.m"   "n"       "
    nc"
 [8] "nr"      "Number" "tick"    "tock"    "Words"
```

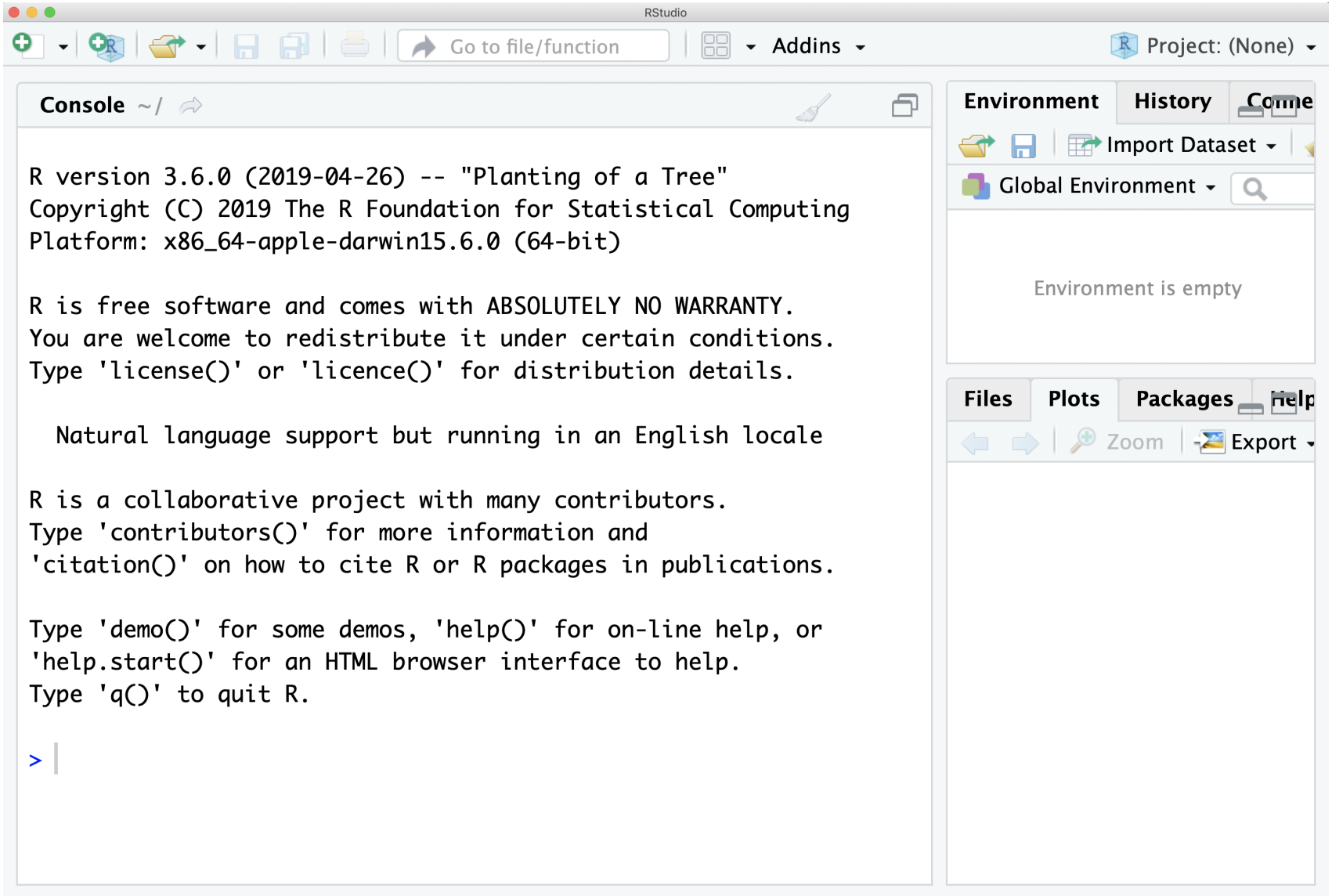
Packages

```
install.packages("Matrix",repos='http://cran.us.r-project.org
    ')
```

```
The downloaded binary packages are in
        /var/folders/2s/9zz_mdbn2652586wqzndfp200000gn/T//
            RtmpUBsElX/downloaded_packages
```

```
library(Matrix)
```

Using RStudio


RStudio
Go to file/function
Addins
Project: (None)
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-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.
>
Environment
History
Import Dataset
Global Environment
Environment is empty
Files
Plots
Packages
Zoom
Export

Using LaTeX

Basic LaTeX command syntax

- Latex command begin with a backslash
- The arguments for Latex commands are written inside of curly braces

```
\documentclass{article}


 \title{My First \LaTeX Document}
 \author{Jane Doe}
 \date{\today}

 \begin{document}
 \maketitle
 Hello world!
 \end{document}
```

My First LaTeX Document

Jane Doe

September 2019

Hello world!

```
\documentclass{article}

\usepackage{lipsum}

\title{My Second \LaTeX Document}
\author{Jane Doe}
\date{September 2019}

\begin{document}
\maketitle
\section{Abstract}
\lipsum[2-4]

\section{Introduction}
\lipsum[2-4]

\section{Methods}
\lipsum[2-4]

\section{Results}
\lipsum[2-4]

\section{Discussion}
\lipsum[2-4]

\end{document}
```

My Second LaTeX Document

Jane Doe

September 2019

1 Abstract

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

2 Introduction

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

3 Methods

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris.

Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

4 Results

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

5 Discussion

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Exercise 1

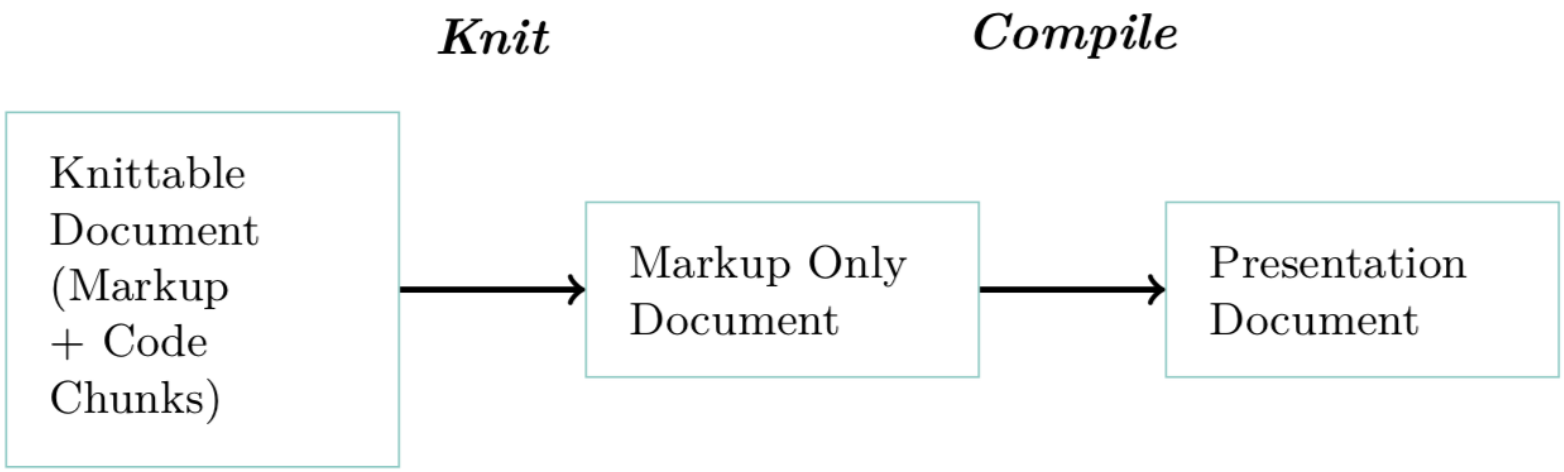
Using LaTeXtemplates

- Navigate to the folder: `/Rep-Res-Workshop/Presentation/LaTeXExercise/`
- Choose one of the templates from the journals Ecography, Journal of Animal Ecology (JAE), Journal of Wildlife Management (JWM), Proceedings of the National Academy of Science (PNAS), or Science.
- Open the `.tex` file in the folder.
- Customize the template to include:
  - Author names from a publication/project you are working on,
  - Affiliations,
  - Sections and sub-sections relevant for your work.
- Open the `.bib` file in the folder.
- Customize the .bib file to include:
  - A reference from your field (go to Google Scholar, find an appropriate article, select the quotation marks below the article, and select "BibTeX" at the bottom of the pop-up window. Copy and past it into the .bib file)
- Add the handle of the reference to the main .tex article using citep{handle} (with a backslash before citep).

Using `knitr`

What `knitr` Does

- `knitr` ties together your presentation of results with the creation of those results
- Choose a mark-up code. We focus on LaTeXin this workshop (an alternative is Markdown)
- Write document with mark-up code, with `R` chunks embedded in mark-up code.
- `knitr` converts `R` chunks to mark-up language (this would be tedious without `knitr`)
- We can then compile final mark-up code using appropriate compiler



knitr LaTeX Example

knitr/rmarkdown Markdown Example

```
\documentclass{article}


\title{My First \texttt{knitr} Document}
\author{Jane Doe}
\date{September 2019}

\begin{document}

\maketitle

\section{Introduction}
Hello world!

\section{Methods}
\subsection{Equation 1}
<<chunk1, echo=TRUE,eval=TRUE>>=
library(knitr)
(output=2+2)
@

\section{Results}
\subsection{Equation 1}
The answer to Equation 1 is \Sexpr{output}


\end{document}
```

My First knitr Document

Jane Doe

September 2019

1 Introduction

Hello world!

2 Methods

2.1 Equation 1

```
library(knitr)
(output=2+2)

## [1] 4
```

3 Results

3.1 Equation 1

The answer to Equation 1 is 4

knitr: elegant, flexible, and fast dynamic report generation with R

Options

Chunk options and package options

2017-02-03

Second `knitr` document

Exercise 2

Create an interactive `knitr` document that:

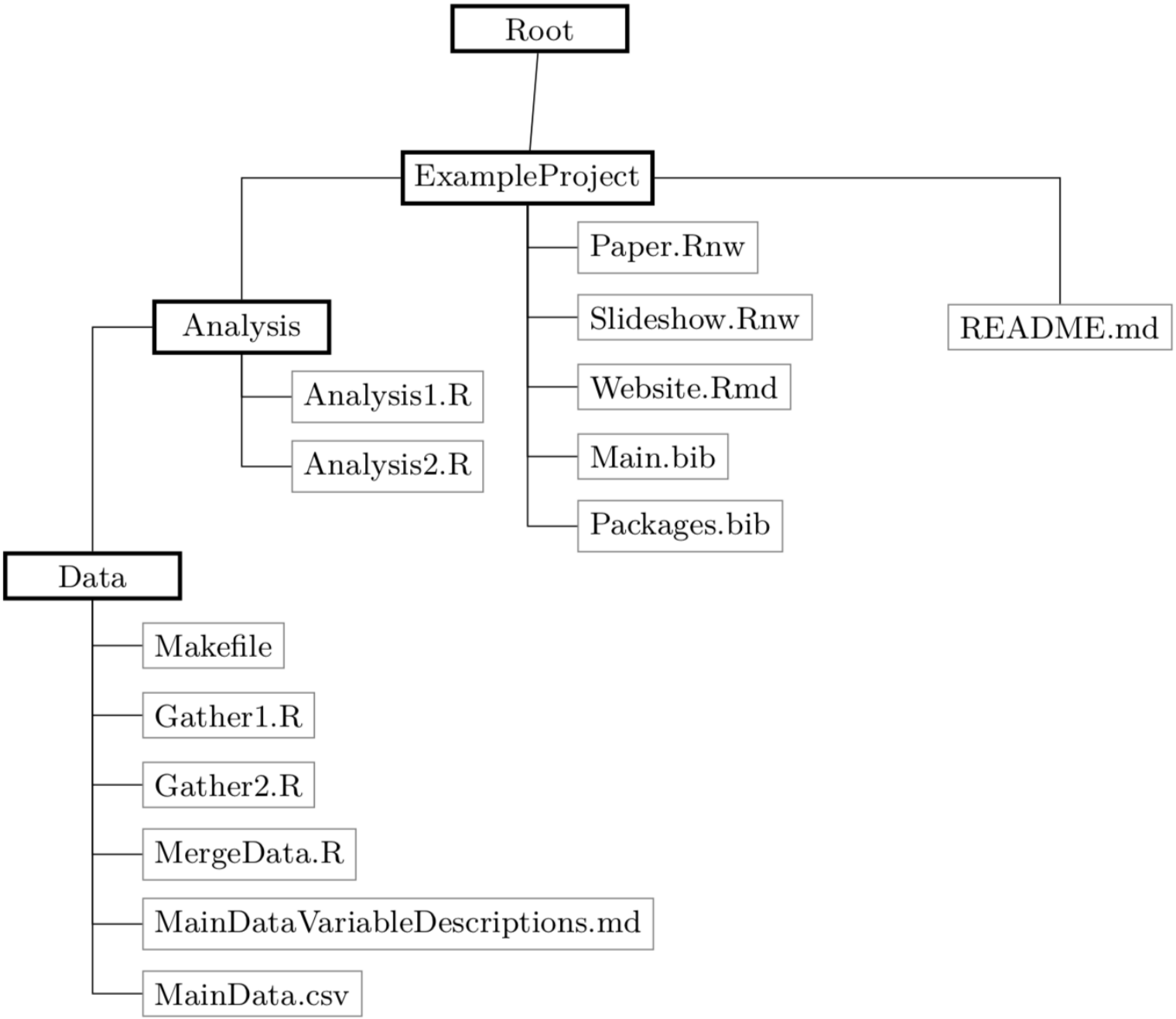
- includes the analysis `MainAnalysis.R` within the `.Rnw` document,
- replaces all static values (e.g., parameter estimates, figure and table numbers) with dynamic values from the incorporated analysis,
- permits you to change the data in the analysis (i.e., remove Alaska), and automatically updates the parameter values.

Organizing Workflow

File Management

Careful file management is crucial for reproducible research

- Explicitly tie your files together,
- Have a plan to organize, store, and make your files available.


Root
ExampleProject
Paper.Rnw
Slideshow.Rnw
README.md
Analysis
Website.Rmd
Analysis1.R
Main.bib
Analysis2.R
Packages.bib
Data
Makefile
Gather1.R
Gather2.R
MergeData.R
MainDataVariableDescriptions.md
MainData.csv

File Management

- If files are well organized and the way they are tied together is clear, replication will be much easier.
- Permits easier changes to analyses.
- Recycle work you have already done.

Is my research reproducible?

What formats are your research documents stored in?

Is my research reproducible?

What formats are your research documents stored in?

- .csv
- .txt
- .pdf
- .html
- .R or .RData

Is my research reproducible?

What formats are your research documents stored in?

- .csv
- .txt
- .pdf
- .html
- .R or .RData

Yes, these are considered "reproducible"

Is my research reproducible?

What formats are your research documents stored in?

- .csv
- .txt
- .pdf
- .html
- .R or .RData

Yes, these are considered "reproducible"

- .doc or .docx
- .sas
- .xls or .xlsx
- any other proprietary file format

Is my research reproducible?

What formats are your research documents stored in?

- .csv
- .txt
- .pdf
- .html
- .R or .RData

Yes, these are considered "reproducible"

- .doc or .docx
- .sas
- .xls or .xlsx
- any other proprietary file format

No, these are not considered "reproducible"

Is my research reproducible?

Is your code linear?

Is my research reproducible?

Is your code linear?

- Clear environment often and at beginning of script
- Each program should focus on one main task or analysis
- Don't rely on manual commenting or uncommenting

```
# Here, I want to know what variables are significant?
#       I will try each variable in turn
lm.out <- lm(weight ~ height, data = trial.data)
remove(lm.out) # clear previous lm.out for each
               #   new lm() definition above

# Is the relationship significant?
#     (If not, clear and try a new variable)
summary(lm.out)
```

Is my research reproducible?

Are your files easily shared with others?

Is my research reproducible?

Are your files easily shared with others?

- Organized directory structure
- Files relatively linked
- Well-documented & commented
- Consistency in coding practices

Is my research reproducible?

Are your files easily shared with others?

- Organized directory structure
- Files relatively linked
- Well-documented & commented
- Consistency in coding practices

"The point of having style guidelines is to have a common vocabulary of coding so people can concentrate on what you are saying, rahter than on how you are saying it." - Google's R Style Guide

styleguide

Google's R Style Guide

R is a high-level programming language used primarily for statistical computing and graphics. The goal of the R Programming Style Guide is to make our R code easier to read, share, and verify.

The Google R Style Guide is a fork of the Tidyverse Style Guide by Hadley Wickham license. Google modifications were developed in collaboration with the internal R user community. The rest of this document explains Google's primary differences with the Tidyverse guide, and why these differences exist.

Syntax

Naming conventions

Google prefers identifying functions with `BigCamelCase` to clearly distinguish them from other objects.

```
# Good
DoNothing <- function() {
  return(invisible(NULL))
}
```

Is my research reproducible?

Do you treat your data as read only?

Is my research reproducible?

Do you treat your data as read only?

- Don't use Excel, etc., to manipulate raw data
- Use an `R` script for data processing
- Process data in one script, then save for loading into subsequent scripts
- When archiving, provide raw data and processing code, not just final tables

Additional Resources

- https://swcarpentry.github.io/r-novice-gapminder/ 02-project-intro/ (RStudio projects)
- https://yihui.name/knitr/ (knitr help)
- https://daringfireball.net/projects/markdown/basics (RMarkdown help)
- https://rpubs.com/alobo/spintutorial (Roxygen)
- http://eriqande.github.io/rep-res-web/ (online reproducible research course)
- https://www.r-bloggers.com/rstudio-and-github/ (Github tutorial)

Thank You!


COMMENT
DATE
CREATED MAIN LOOP & TIMING CONTROL
14 HOURS AGO
ENABLED CONFIG FILE PARSING
9 HOURS AGO
MISC BUGFIXES
5 HOURS AGO
CODE ADDITIONS/EDITS
4 HOURS AGO
MORE CODE
4 HOURS AGO
HERE HAVE CODE
4 HOURS AGO
AAAAAAAA
3 HOURS AGO
ADKFJSLKDFJSDKLFJ
3 HOURS AGO
MY HANDS ARE TYPING WORDS
2 HOURS AGO
HAAAAAAAAANDS
2 HOURS AGO
AS A PROJECT DRAGS ON, MY GIT COMMIT MESSAGES GET LESS AND LESS INFORMATIVE.