



Introduction to R Markdown

Saghir Bashir

24th October 2019

<https://github.com/saghirb/Rmarkdown-Intro-Workshop>

Relax

Experiment

Make Mistakes

Learn

Enjoy



Outline

R Markdown Examples

What is R Markdown?

HTML & PDF Documents

Pretty Tables

Summary



Prerequisites

Latest versions of:

1. R: <https://cran.r-project.org>
2. RStudio: <https://www.rstudio.com/products/rstudio>



R Markdown Examples

HTML "Analysis" Document

World Population Data

Exercise Questions

Read in the data

Pre-processing the Data

Exercise Answers

Portugal: 1950 v 2017

World Population

Population by Continent

Population by Country

Portugal: 1950 to 2015

High & Low Female Proportion

Portuguese Females Over Time

Read in the data

First load the `data.table` package (hint: use the `library()` function). Then use the `fread()` function to read in the data file `World-Population.csv` in an object called `un`. Finally look at the `un` dataset using the `str()` function.

```
library(data.table)
un <- fread("World-Population.csv")
un
```

```
##           continent    Location Time AgeGrp AgeGrpStart PopFemale PopMale
## 1:      Asia Afghanistan 1950   0-4          0  661.58  630.04
## 2:      Asia Afghanistan 1950   5-9          5  487.33  516.21
## 3:      Asia Afghanistan 1950 10-14         10  423.33  461.38
## 4:      Asia Afghanistan 1950 15-19         15  369.36  414.37
## 5:      Asia Afghanistan 1950 20-24         20  318.39  374.11
## ...
## 47942: Africa Zambia 2015 80-84        80  23.30  16.10
## 47943: Africa Zambia 2015 85-89        85  9.17   5.84
## 47944: Africa Zambia 2015 90-94        90  2.40   1.30
## 47945: Africa Zambia 2015 95-99        95  0.35   0.16
## 47946: Africa Zambia 2015 100+       100  0.03   0.01
```

```
str(un)
```

```
## Classes 'data.table' and 'data.frame':  47946 obs. of  7 variables:
## $ continent : chr  "Asia" "Asia" "Asia" "Asia" ...
## $ Location  : chr  "Afghanistan" "Afghanistan" "Afghanistan" "Afghanistan" ...
## $ Time      : int  1950 1950 1950 1950 1950 1950 1950 1950 1950 ...
## $ AgeGrp    : chr  "0-4" "5-9" "10-14" "15-19" ...
## $ AgeGrpStart: int  0 5 10 15 20 25 30 35 40 45 ...
## $ PopFemale : num  662 487 423 369 318 ...
```

HTML Website

https://dsup.org

120% ⋮ ⭐

Data Science Unplugged About Resources Events Blog Search



Data Science Unplugged
A community for data science
📍 Lisbon

Follow

1 POSTS	12 TAGS
------------	------------



CATEGORIES

- ▶ blog (5)
- ▶ events (1)

TAGS

- ▶ accessibility (1)
- ▶ critical-thinking (2)
- ▶ data-science (5)
- ▶ events (2)
- ▶ food (1)
- ▶ imposter-syndrome (1)
- ▶ latex (1)
- ▶ learning (4)
- ▶ r-language (3)
- ▶ r-markdown (1)

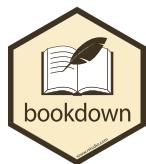
We Are All Imposters

📅 2018-06-05 · 441 WORDS · 3 MINUTE READ 📄 BLOG

🏷️ IMPOSTER SYNDROME · STATISTICS · CRITICAL THINKING · LEARNING

We are all imposters. Let's enjoy it!

Read more



HTML Book



R Markdown: The Definitive Guide

☰ ⌂ A ⌂ i ⌂

[Preface](#)

How to read this book
Structure of the book
Software information and conventions
Acknowledgments
About the Authors
Yihui Xie
J.J. Allaire
Garrett Grolemund

I Get Started

1 Installation

2 Basics

2.1 Example applications

2.1.1 Airbnb's knowledge repository
2.1.2 Homework assignments on GitHub
2.1.3 Personalized mail
2.1.4 2017 Employer Health Benefits Survey
2.1.5 Journal articles
2.1.6 Dashboards at eeloo
2.1.7 Books
2.1.8 Websites

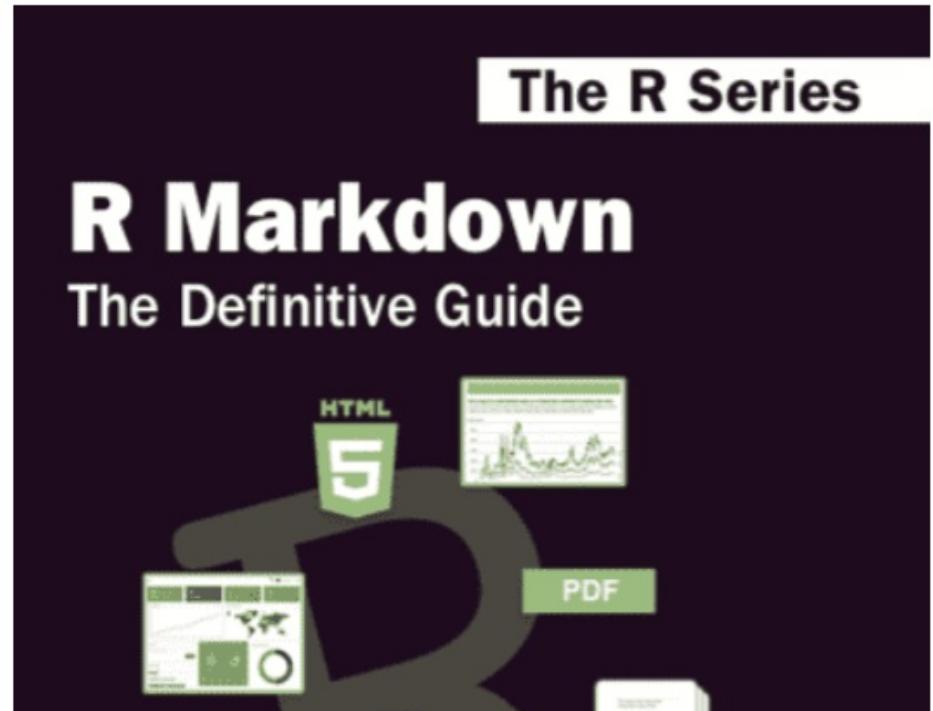
R Markdown: The Definitive Guide

Yihui Xie, J. J. Allaire, Garrett Grolemund

2019-06-03

Preface

Note: This book has been published by Chapman & Hall/CRC. The online version of this book is free to read here (thanks to Chapman & Hall/CRC), and licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.



Twitter icon ⌂

Print icon ⌂

Share icon ⌂

Next page icon ➤

Women in Parliament – data.table

Saghir Bashir

This version was compiled on April 16, 2019

We will use the World Bank's indicator data for "Women in Parliament" as a case study when working with the data.table R package. We will guide you through the geographical and time trends for the percentage of women in national parliaments. We will start by learning about and understanding the raw data, which we will then process ("wrangle") in preparation for some exploratory analysis.

Women in Parliament | World Bank Indicator | data.table | Tinyverse

1. Preface

We present a real-life case study for the data.table¹ package using the World Bank's "Women in Parliament" indicator data. To get the most out of this case-study guide, repeat the examples and do the exercises whilst reading it.

Guide materials. You can download materials for this guide from this link:

- <https://ilustat.com/shared/WiP-rdatatable.zip>

Unzip the file, which contains the data, this guide and an R script exercise file. We advise you to work with "WiP-Exercise.R" file to follow the examples and do the exercises. If you are using RStudio, you can double click on "WiP-dt.Rproj" to get started.

Data limitations. Take caution when interpreting these data, as parliamentary systems vary from country to country, and in some cases over time. Some of the issues to consider include:

- Who has, and who does not have, the right to become a Member of Parliament (MP)?
- How does someone become an MP? Through democratic elections? How is "democratic election" defined?
- What is the real power of MPs and their parliament? Can MPs make a difference?

Data definitions & assumptions.

"**Women**". The definition for "women" is not given, so we will assume that it refers to a binary classification for gender (sex).

"**Country (Region)**". The definition of countries and regions can change over time. (e.g. formation of new countries after conflicts, new member states joining a pre-existing collective). How are these changes reflected in the data? How do they affect the interpretation?

Pro tip. Understand the limitations of your data before anybody else points them out to you.

4. About the data file

Chapter 3

Tables, Graphics, References, and Labels

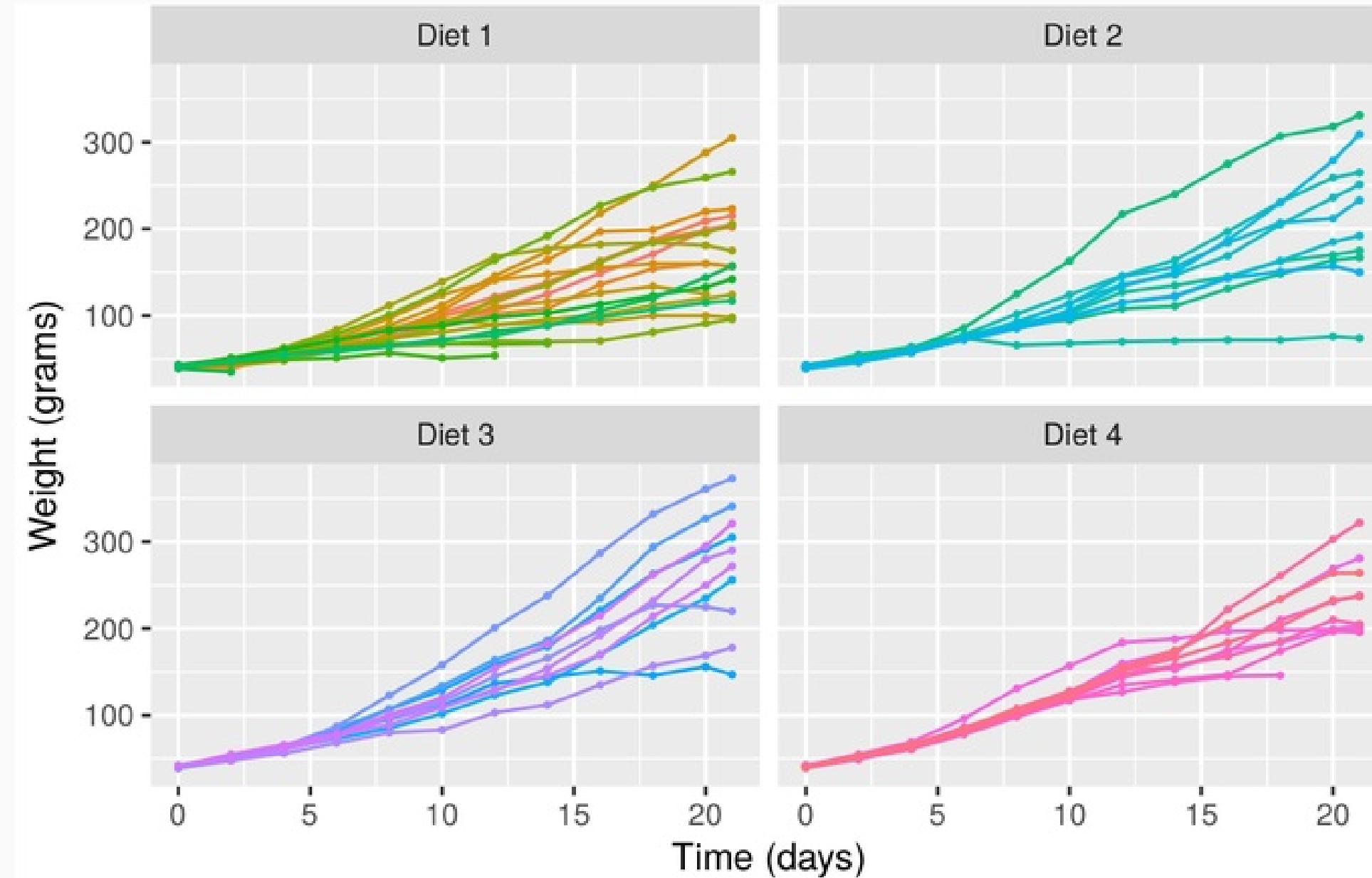
3.1 Tables

In addition to the tables that can be automatically generated from a data frame in **R** that you saw in [R Markdown Basics](#) using the `kable` function, you can also create tables using `pandoc`. (More information is available at <http://pandoc.org/README.html#tables>.) This might be useful if you don't have values specifically stored in **R**, but you'd like to display them in table form. Below is an example. Pay careful attention to the alignment in the table and hyphens to create the rows and columns.

Table 3.1: Correlation of Inheritance Factors for Parents and Child

Factors	Correlation between Parents & Child	Inherited
Education	-0.49	Yes
Socio-Economic Status	0.28	Slight
Income	0.08	No
Family Size	0.18	Slight

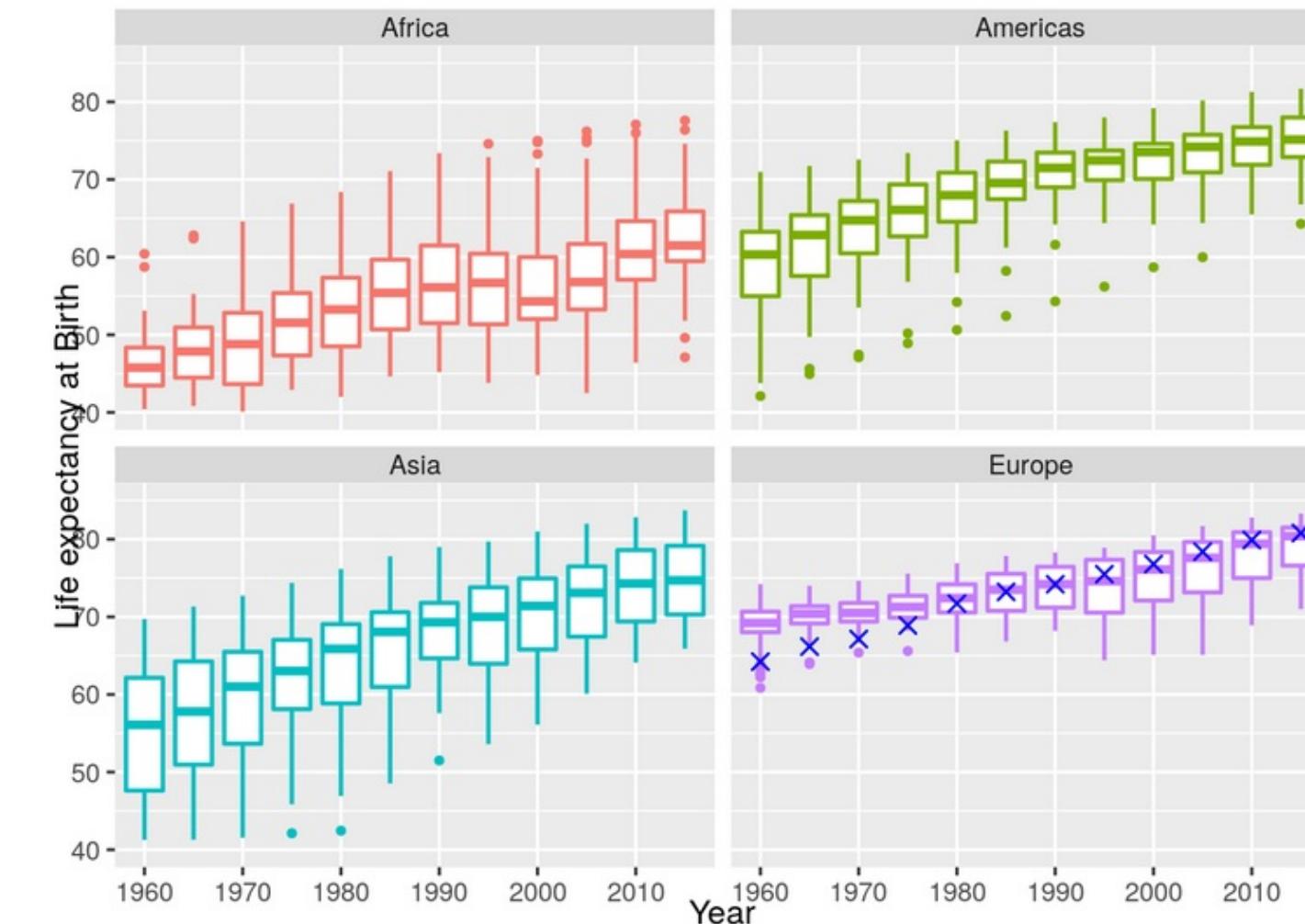
Analysis - By Chick Plot



Life Expectancy - By Continent



```
gm[continent!="Oceania"] %>%
  ggplot(aes(fyear, lifeexp,
             colour = continent)) +
  geom_boxplot(outlier.size = .5) +
  geom_point(data=gmPT,
             shape = 4,
             colour="blue") +
  scale_x_discrete(
    breaks=seq(1960, 2010, 10)) +
  scale_y_continuous(limits=c(40,85)) +
  facet_wrap(~continent) +
  xlab("Year") +
  ylab("Life expectancy at Birth") +
  theme(legend.position = "none")
```



More R Markdown



Formats

- HTML
- PDF
- Word & ODF
- Power Point & ODF

Other

- Dashboards
- Web Applications
- HTML Notebooks
- CVs



What is R Markdown?

What is Markdown?



Special computer language ("Markup")

- Plain text file with special syntax.
- Human readable.
- "Tags" define Structure & elements.
- Converts to HTML, PDF, ...

Markdown Example



```
# Markdown Example  
Paragraphs are separated by a blank line.  
  
## Format text  
Some _italic_, **bold** and `monospace` text.  
  
Bullet list  
+ Apples.  
+ Oranges.  
+ Pears  
  
An [example link](https://github.com/saghirb)..
```

Markdown Example

Paragraphs are separated by a blank line.

Format text

Some *italic*, **bold** and monospace text.

Bullet list

- Apples.
- Oranges.
- Pears

An [example link](https://github.com/saghirb).

What is R Markdown?



A mix of R programming and markdown.

- R code and documentation in one place.
- Great for doing reproducible research.
- Great for collaborating and sharing.
- Converts to HTML, PDF, websites, article, books, ...

R Markdown Example



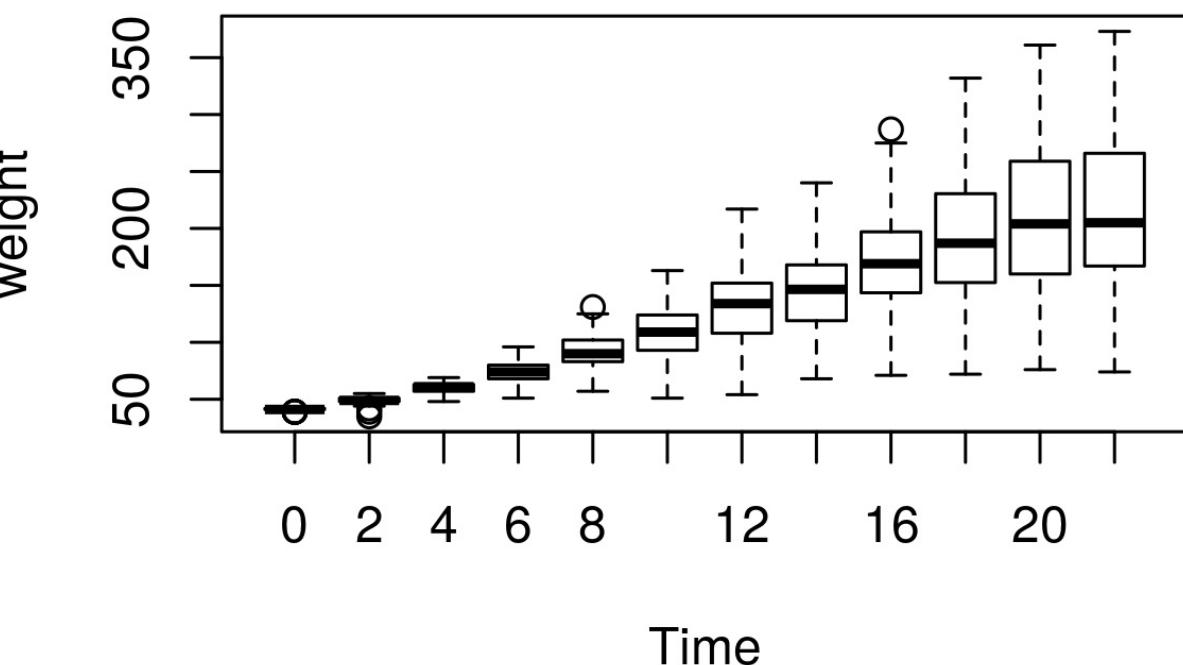
```
--  
title: "R Markdown Demo"  
author: "Saghir Bashir"  
date: "13/06/2019"  
output: html_document  
  
---  
  
_Using_ **R** as a calculator.  
  
```{r Rcalc}  
2+5
8**2
```  
  
Plot the `ChickWeight` data.  
```{r plotCW, echo=FALSE, out.width="80%"}  
with(ChickWeight, boxplot(weight ~ Time))
```
```

Using R as a calculator.

```
(2+5)**2
```

```
## [1] 49
```

Plot the ChickWeight data.



R Markdown Explained - YAML



```
---
title: "R Markdown Demo"
author: "Saghir Bashir"
date: "13/06/2019"
output: html_document
---
Using_ **R** as a calculator.

```{r Rcalc}
2+5
8**2
```

Plot the `ChickWeight` data.

```{r plotCW, echo=FALSE, out.width="80%"}
with(ChickWeight, boxplot(weight ~ Time))
```
```

File extension is .Rmd

- e.g. My-Report.Rmd

Top is YAML header

- Descriptive information.
- Format & Style.

R Markdown Explained - Markdown



```
---
```

```
title: "R Markdown Demo"
author: "Saghir Bashir"
date: "13/06/2019"
output: html_document
```

```
---
```

```
_Using_ **R** as a calculator.
```

```
```{r Rcalc}
2+5
8**2
```

```

```
Plot the `ChickWeight` data.
```

```
```{r plotCW, echo=FALSE, out.width="80%"}
with(ChickWeight, boxplot(weight ~ Time))
```

```

Below the YAML header

- **Markdown code.**

R Markdown Explained - R Chunks



```
---
title: "R Markdown Demo"
author: "Saghir Bashir"
date: "13/06/2019"
output: html_document
---

Using **R** as a calculator.
```

```
```{r Rcalc}
2+5
8**2
```

```

Plot the `ChickWeight` data.

```
```{r plotCW, echo=FALSE, out.width="80%"}
with(ChickWeight, boxplot(weight ~ Time))
```

```

Below the YAML header

- **Markdown code.**
- **R code chunks with:**
 - **unique names.**
 - **chunk options.**

HTML Document

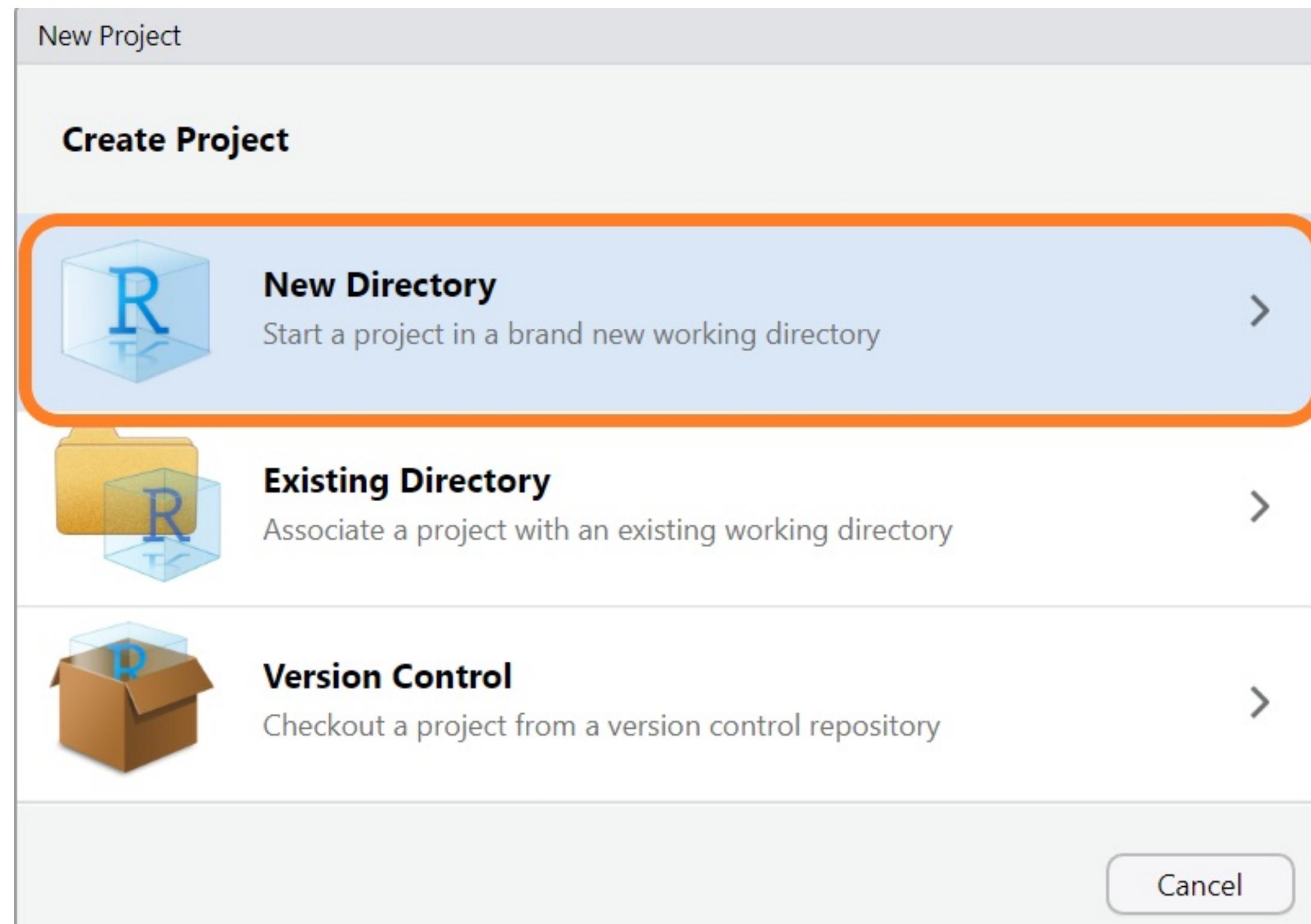
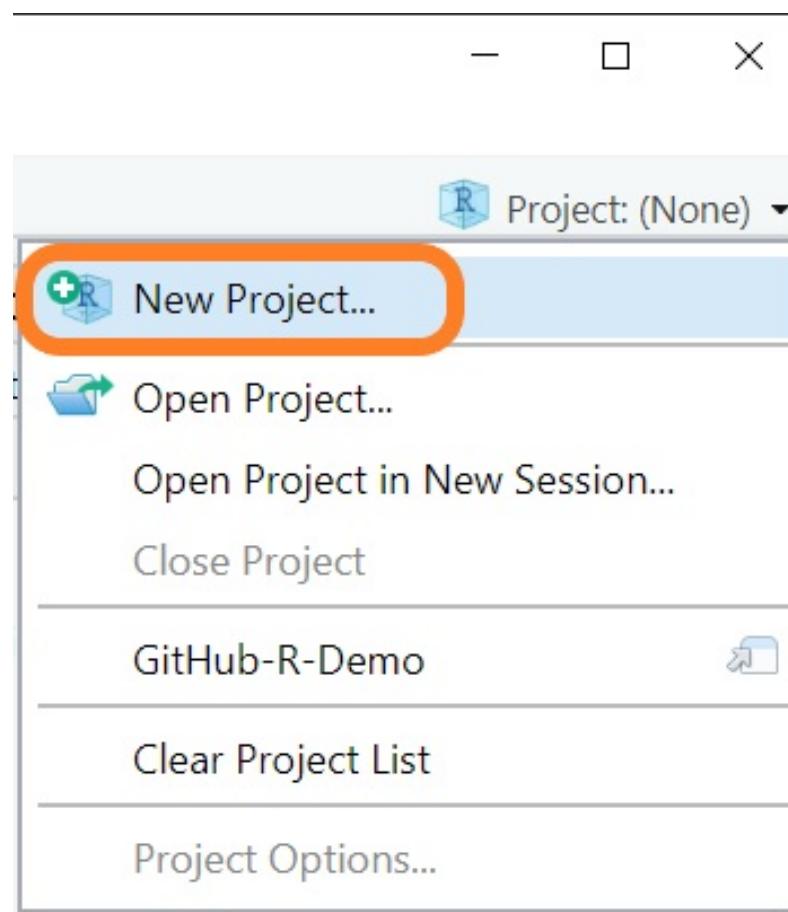
```
install.packages("kableExtra")
```

```
install.packages("rticles")
```

```
install.packages("tinytex")
```

```
tinytex::install_tinytex()
```

Create a New RStudio Project



Select "New Project"

New Project

Back Project Type

-  New Project > Create a new project in an empty directory
-  R Package >
-  Shiny Web Application >
- R Package using Rcpp >
- R Package using RcppArmadillo >
- R Package using RcppEigen >
- R Package using devtools >

Cancel



Define Directory Location

New Project

Back

Create New Project



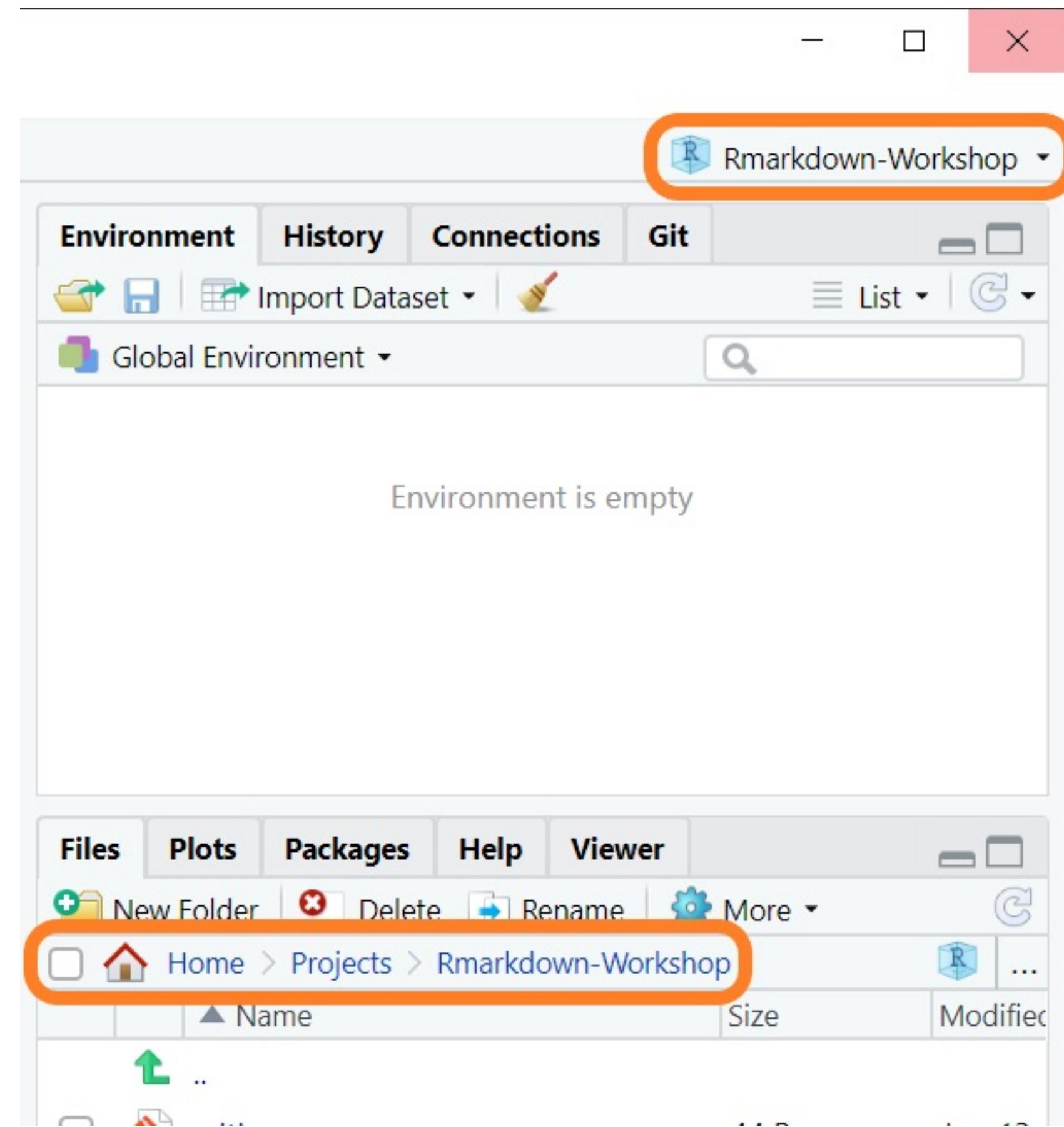
Directory name:
Rmarkdown-Workshop

Create project as subdirectory of:
~/Projects

Create a git repository

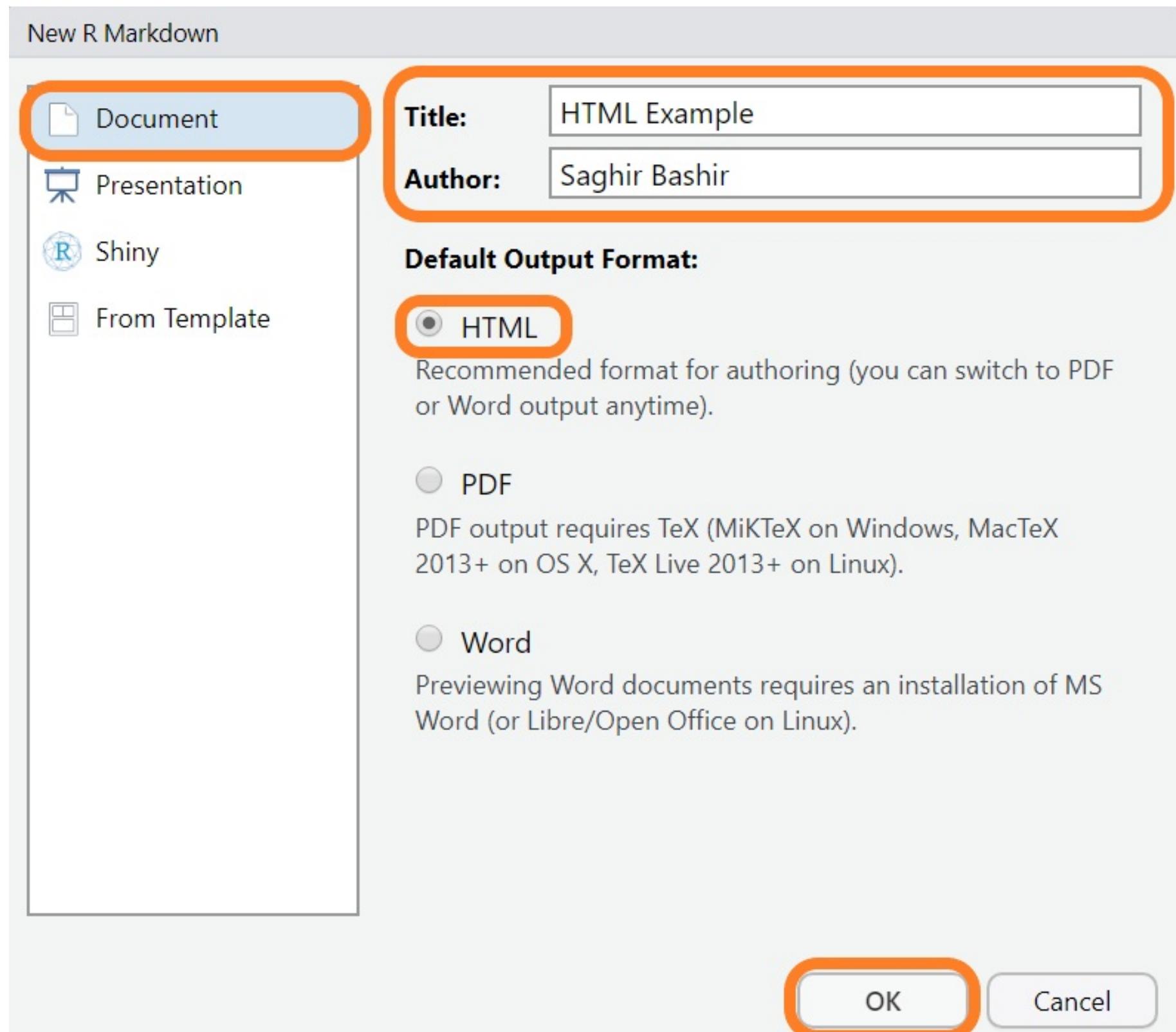
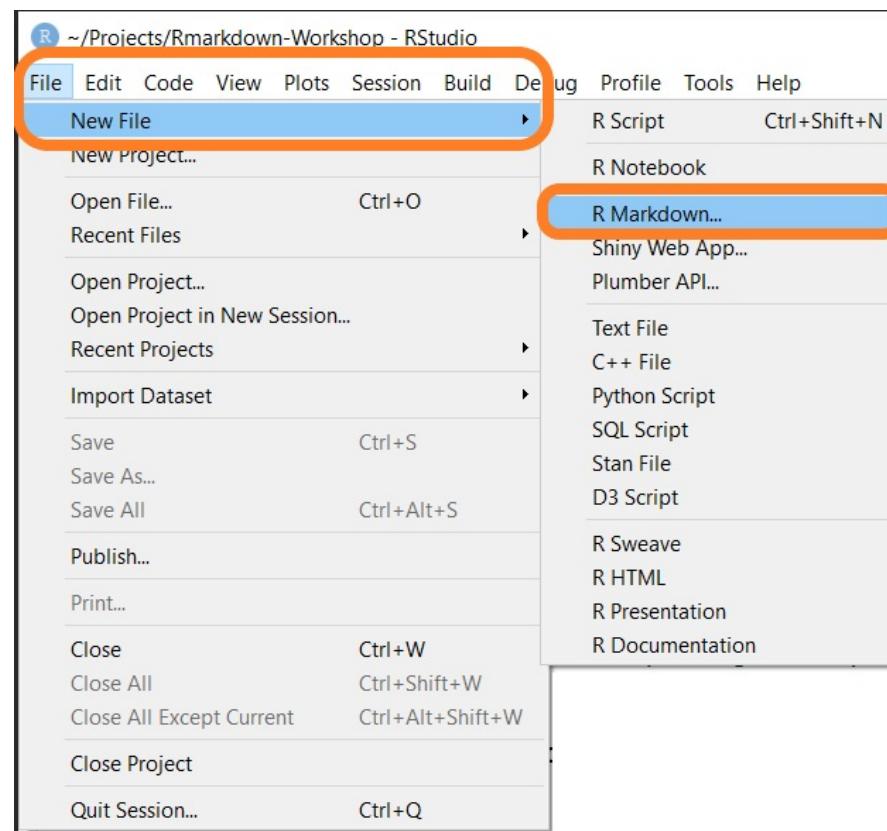
Open in new session

New Project Created





Create R Markdown Document



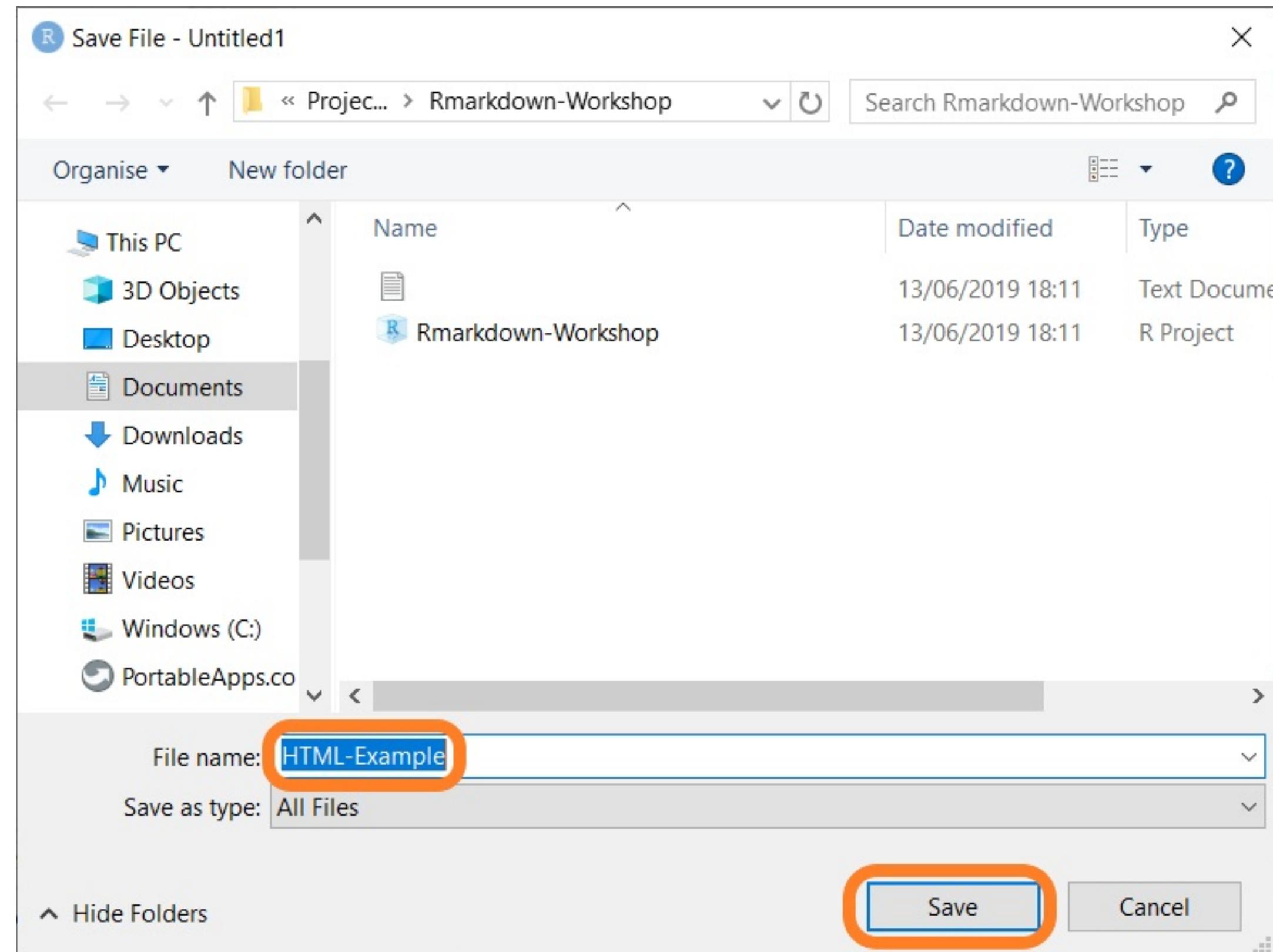
Untitled and Unsaved



The screenshot shows the RStudio interface with the following components:

- Top Bar:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Toolbar:** Includes icons for New, Open, Save, Print, Go to file/function, and Addins.
- Document Editor:** The main workspace where the R Markdown code is written. A file named "Untitled1" is open, highlighted with an orange border. The code includes YAML front matter and R code chunks. Lines 14 and 15 contain explanatory text about R Markdown.
- Environment Tab:** Shows the Global Environment, which is currently empty.
- Files Tab:** Displays the project structure: Home > Projects > Rmarkdown-Workshop. It lists files like ".gitignore" and "Rmarkdown-Workshop.Rproj".
- Console:** Shows the standard R startup message and help text.
- Terminal:** Shows the command `~/Projects/Rmarkdown-Workshop/`.
- Jobs:** Shows the status of any running R processes.

Save as .Rmd



Click "Knit" to Create HTML File



R ~/Projects/Rmarkdown-Workshop - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

HTML-Example.Rmd

Knit

```
1 ---  
2 title: "HTML Example"  
3 author: "Saghir Bashir"  
4 date: "13/06/2019"  
5 output: html_document  
6 ---  
7  
8 ```{r setup, include=FALSE}  
9 knitr::opts_chunk$set(echo = TRUE)  
10  
11  
12 ## R Markdown  
13  
14 This is an R Markdown document. Markdown is a simple formatting syntax  
for authoring HTML, PDF, and MS Word documents. For more details on using  
R Markdown see <http://rmarkdown.rstudio.com>.  
15  
16 When you click the **Knit** button a document will be generated that  
includes both content as well as the output of any embedded R code chunks  
2:1 # HTML Example R Markdown
```

HTML Document - Top

A screenshot of the RStudio interface displaying an HTML document titled "HTML Example".

The document header includes:

- File path: ~/Projects/Rmarkdown-Workshop/HTML-Example.html
- Tab: HTML-Example.html
- Buttons: Open in Browser, Find
- Top right: Publish dropdown, C button

The main content of the document is:

HTML Example

Saghir Bashir
13/06/2019

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed      dist
## Min.   : 4.0   Min.   : 2.00
## 1st Qu.:12.0   1st Qu.: 26.00
## Median :15.0   Median : 36.00
## Mean   :15.4   Mean   : 42.98
## 3rd Qu.:19.0   3rd Qu.: 56.00
## Max.   :25.0   Max.   :120.00
```

Including Plots

HTML Document - Bottom



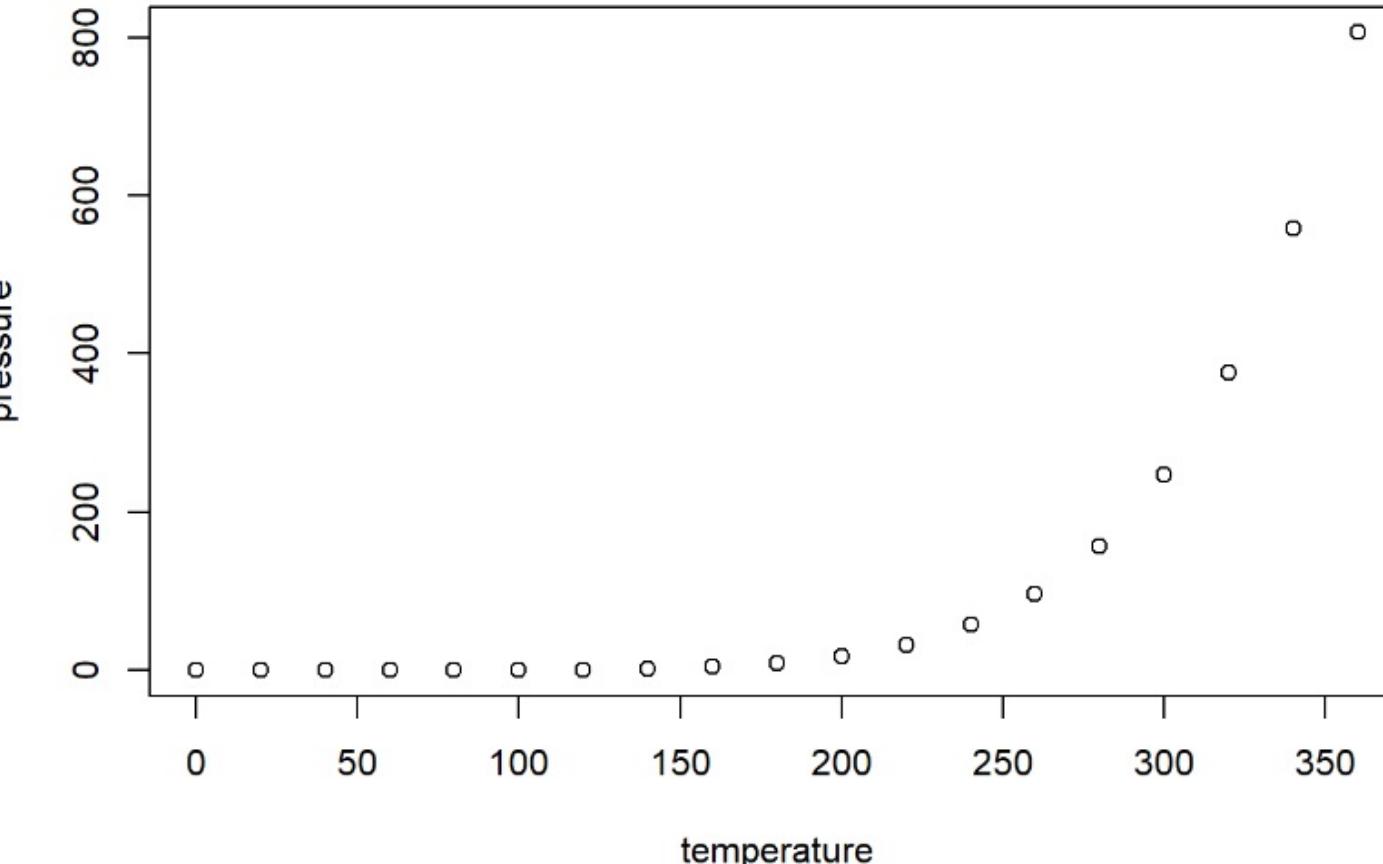
R ~/Projects/Rmarkdown-Workshop/HTML-Example.html

HTML-Example.html | Open in Browser | Find

Max. :25.0 Max. :120.00

Including Plots

You can also embed plots, for example:



A scatter plot showing the relationship between temperature (x-axis) and pressure (y-axis). The x-axis ranges from 0 to 350 with major ticks every 50 units. The y-axis ranges from 0 to 800 with major ticks every 200 units. The data points show a strong positive correlation, starting near (0, 0) and ending at approximately (350, 800).

| temperature | pressure |
|-------------|----------|
| 0 | 0 |
| 25 | 0 |
| 50 | 0 |
| 75 | 0 |
| 100 | 0 |
| 125 | 0 |
| 150 | 0 |
| 175 | 0 |
| 200 | 0 |
| 225 | 0 |
| 250 | 100 |
| 275 | 180 |
| 300 | 250 |
| 325 | 400 |
| 350 | 800 |

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

You Can "Knit" to PDF



R ~/Projects/Rmarkdown-Workshop - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

HTML-Example.Rmd x

Knit to HTML
Knit to PDF (highlighted with an orange oval)
Knit to Word
Knit with Parameters...
Knit Directory
Clear Knitr Cache...

cho = TRUE)

```
1 t Knit to HTML
2 a Knit to PDF
3 d Knit to Word
4 o Knit with Parameters...
5 
6 
7 
8 
9 kn Knit Directory
10 
11 Clear Knitr Cache...
12 ## R Markdown
13 
14 This is an R Markdown document. Markdown is a simple formatting syntax
15 for authoring HTML, PDF, and MS Word documents. For more details on using
16 R Markdown see <http://rmarkdown.rstudio.com>.
17 
18 When you click the **Knit** button a document will be generated that
19 includes both content as well as the output of any embedded R code chunks
20 
21 # HTML Example R Markdown
```



HTML Example

Saghir Bashir

13/06/2019

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed          dist
## Min.   : 4.0   Min.   : 2.00
## 1st Qu.:12.0   1st Qu.: 26.00
## Median :15.0   Median : 36.00
## Mean   :15.4   Mean   : 42.98
## 3rd Qu.:19.0   3rd Qu.: 56.00
## Max.   :25.0   Max.   :120.00
```

Including Plots

You can also embed plots, for example:

Updated Theme

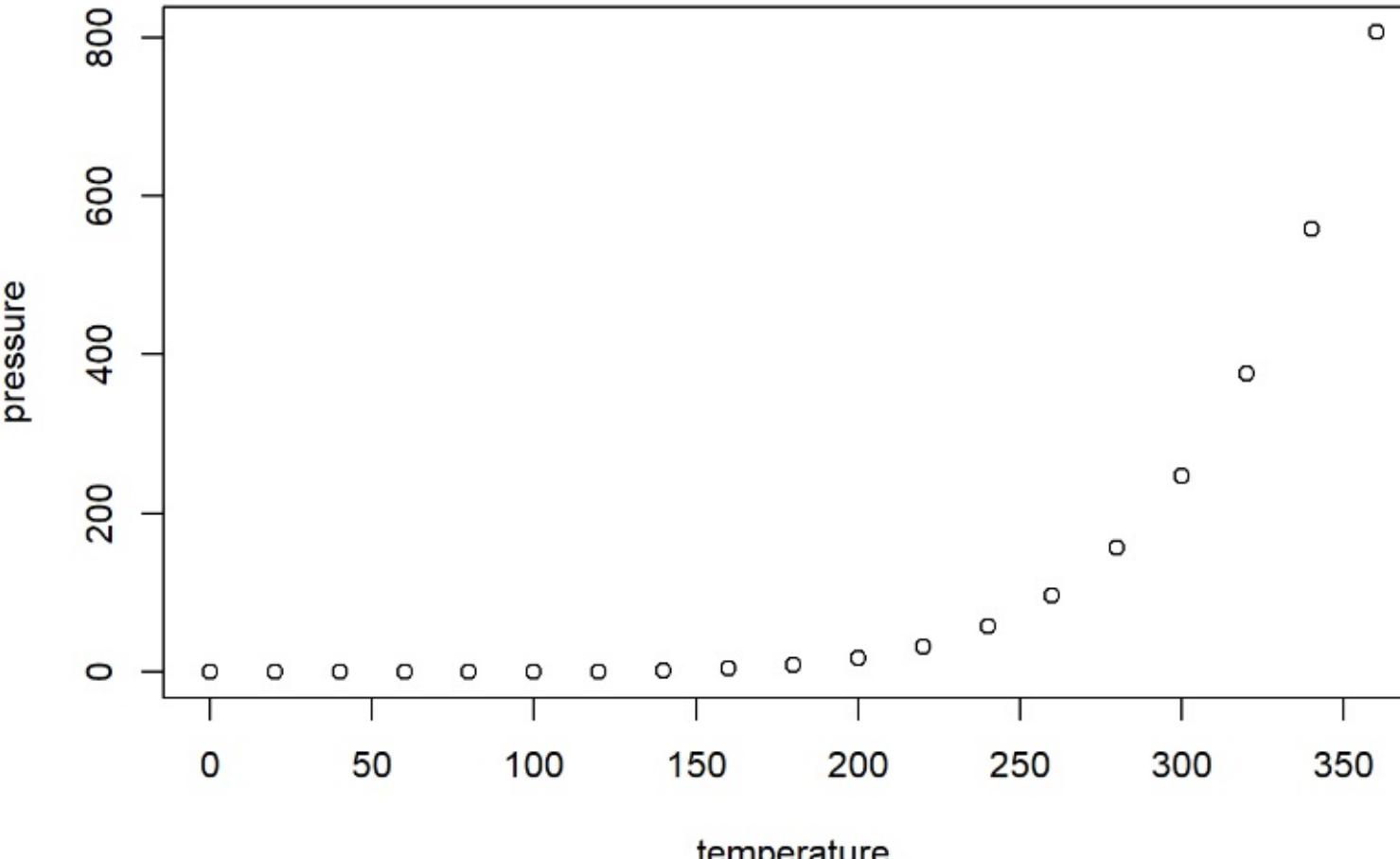
R ~Projects/Rmarkdown-Workshop/HTML-Example.html

HTML-Example.html | Open in Browser | Find

R Markdown

Including Plots

You can also embed plots, for example:



A scatter plot showing the relationship between temperature (x-axis) and pressure (y-axis). The x-axis ranges from 0 to 350 with major ticks every 50 units. The y-axis ranges from 0 to 800 with major ticks every 200 units. The data points, represented by open circles, show a strong positive linear correlation, starting near (0, 0) and ending near (350, 800).

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

R Markdown Cheatsheet



Pandoc's Markdown

Write with syntax on the left to create effect on right (after render)

| Plain text | Plain text |
|---|---|
| End a line with two spaces
to start a new paragraph.
italics and **bold** | End a line with two spaces
to start a new paragraph.
<i>italics</i> and bold |
| `verbatim` code | verbatim code |
| sub/superscript ² ₂ | sub/superscript ² ₂ |
| ~~strikethrough~~ | strikethrough |
| escaped: `\\` | escaped: `\\` |
| endash: --, emdash: --- | endash: --, emdash: --- |
| equation: \$A = \pi * r^2\$ | equation: $A = \pi * r^2$ |
| equation block: | equation block: |
| $E = mc^2$ | $E = mc^2$ |
| > block quote | block quote |
| # Header1 {#anchor} | |
| ## Header2 {#css_id} | |
| ### Header3 {.css_class} | |

Header1

Header 2

- Useful to find R markdown "tags".
- Use it for the exercises.
- Lots of other useful information too.

Source: <https://www.rstudio.com/resources/cheatsheets/>



Pretty Tables

Objective



Before

```
sData  
##   Treatment   Visit   N Mean S.D.  
## 1: Placebo Baseline 182 27.4 3.23  
## 2: Placebo   Final 157 25.6 4.11  
## 3: Active Baseline 179 26.7 3.45  
## 4: Active   Final 145 20.7 4.27
```

After

| Treatment | Visit | N | Mean | S.D. |
|-----------|----------|-----|------|------|
| Placebo | Baseline | 182 | 27.4 | 3.23 |
| | Final | 157 | 25.6 | 4.11 |
| Active | Baseline | 179 | 26.7 | 3.45 |
| | Final | 145 | 20.7 | 4.27 |



Use knitr::kable

```
library(magrittr)
library(knitr)
sData %>%
  kable()
```

| Treatment | Visit | N | Mean | S.D. |
|-----------|----------|-----|------|------|
| Placebo | Baseline | 182 | 27.4 | 3.23 |
| Placebo | Final | 157 | 25.6 | 4.11 |
| Active | Baseline | 179 | 26.7 | 3.45 |
| Active | Final | 145 | 20.7 | 4.27 |

```
library(magrittr)
library(knitr)
sData %>%
  kable(align="lcccc", digits=c(0, 0, 0, 1, 2))
```

| Treatment | Visit | N | Mean | S.D. |
|-----------|----------|-----|------|------|
| Placebo | Baseline | 182 | 27.4 | 3.23 |
| Placebo | Final | 157 | 25.6 | 4.11 |
| Active | Baseline | 179 | 26.7 | 3.45 |
| Active | Final | 145 | 20.7 | 4.27 |

Use kableExtra Package



```
library(magrittr)
library(knitr)
library(kableExtra)
sData %>%
  kable(align="lcccc", digits=c(0, 0, 0, 1, 2)) %>%
  kable_styling(bootstrap_options="striped",
                full_width=TRUE)
```

| Treatment | Visit | N | Mean | S.D. |
|-----------|----------|-----|------|------|
| Placebo | Baseline | 182 | 27.4 | 3.23 |
| Placebo | Final | 157 | 25.6 | 4.11 |
| Active | Baseline | 179 | 26.7 | 3.45 |
| Active | Final | 145 | 20.7 | 4.27 |

```
library(magrittr)
library(knitr)
library(kableExtra)
sData %>%
  kable(align="lcccc", digits=c(0, 0, 0, 1, 2)) %>%
  kable_styling(bootstrap_options="striped",
                full_width=TRUE) %>%
  column_spec(1:5, width = "4em")
```

| Treatment | Visit | N | Mean | S.D. |
|-----------|----------|-----|------|------|
| Placebo | Baseline | 182 | 27.4 | 3.23 |
| Placebo | Final | 157 | 25.6 | 4.11 |
| Active | Baseline | 179 | 26.7 | 3.45 |
| Active | Final | 145 | 20.7 | 4.27 |

Final Table



```
library(magrittr)
library(knitr)
library(kableExtra)
sData %>%
  kable(align="lcccc", digits=c(0, 0, 0, 1, 2)) %>%
  kable_styling(bootstrap_options="striped",
                full_width=TRUE) %>%
  column_spec(1:5, width = "4em") %>%
  collapse_rows(1, valign = "top")
```

| Treatment | Visit | N | Mean | S.D. |
|-----------|----------|-----|------|------|
| Placebo | Baseline | 182 | 27.4 | 3.23 |
| | Final | 157 | 25.6 | 4.11 |
| Active | Baseline | 179 | 26.7 | 3.45 |
| | Final | 145 | 20.7 | 4.27 |

We started with

```
sData
```

```
##   Treatment   Visit   N Mean S.D.
## 1: Placebo Baseline 182 27.4 3.23
## 2: Placebo   Final 157 25.6 4.11
## 3: Active   Baseline 179 26.7 3.45
## 4: Active     Final 145 20.7 4.27
```

Exercise (i)



- 1) Download <https://ilustat.com/shared/RmarkdownWS.zip>
- 2) Unzip and double click "CW-Summary.Rproj" file.
- 3) Create a new Rmarkdown file called "CW-Report.Rmd"
- 4) Recreate "CW-Report-Target.html"

Exercise (ii)



Creating a HTML presentation

- 1) Create a new "Presentation" from R Markdown.
- 2) Select "HTML (ioslides)" style.
- 3) Save as "CW-Slides.Rmd"
- 4) Recreate "CW-Slides-Target.html"

Summary



R Markdown

- R code and documentation in one place.
- Great for doing reproducible research.
- Great for collaborating and sharing.
- A big range of possible outputs.
- The rewards and benefits are big!

Feedback

Please send your feedback or comments via:

- <https://github.com/saghirb/Rmarkdown-Intro-Workshop> or
- <https://twitter.com/ilustat>

Thanks!

**This work is licensed under the
Creative Commons Attribution-Non Commercial 4.0 International License.**

To view a copy of this license, visit

<http://creativecommons.org/licenses/by-nc/4.0/>