

Quality Improvement—Improved with R

Mara Alexeev, MD, MPH

Boston Children's Hospital, Department of Pediatrics

Clinical Fellow, Harvard Medical School

Created Spring 2021: Last updated May 10, 2021

Workshop Website

These slides are intentionally bland to be amenable for printed materials.

To view the slides in fancier formatting go to the [workshop's website](#).¹

All the workshop materials are available on [Github](#).²

The RStudio Cloud materials are available [here](#).³

¹https://maraalexeev.github.io/R_for_Clinical_Informatics/

²https://github.com/MaraAlexeev/R_for_Clinical_Informatics#r-for-clinical-informatics

³<https://rstudio.cloud/project/2465874>

Why should I come to this workshop?

For those who don't know anything about R

- ▶ Learn how your organization can use R
- ▶ Write your first R code!
- ▶ See some highlights of the R ecosystem

Why should I come to this workshop?

For the beginner R user

- ▶ See start-to-finish project management in R
- ▶ Make some gorgeous plots!
- ▶ Get super excited about next steps in your learning process
- ▶ Make R friends
- ▶ Python users welcome too!

Why should I come to this workshop?

For the more advanced R user

- ▶ There is extra bonus material available for the advanced user!
- ▶ Hear how to expand the use of R in your organization
- ▶ Be a cheerleader to folks learning about R
- ▶ Meet other people in Clinical Informatics using R

Wait!?! What about QI?

I didn't forget about quality improvement!

The materials we will go through as we play with R are focused around a simulated quality improvement project to help spark your imagination about how you could use R in your work from project conception to analysis and publication.

Pre-Workshop Preparation

To Know

The workshop will be held on Zoom.

To Do

You'll need the following during the workshop:

- ▶ A computer with internet connection
- ▶ A supported browser⁴
- ▶ An RStudio Cloud account—**free**

⁴<https://support.rstudio.com/hc/en-us/articles/227449447-Supported-browsers-for-RStudio-Connect>

Extras

If you would like to some additional preparation for the workshop, I recommend the following:

- ▶ Learn about [Markdown](#) in 10 minutes!
- ▶ Sign up for a [Github](#) Account—***free***

Schedule

Introduction ~30 minutes

Systems check ~15 minutes

Play ~30 minutes

Wrap Up and Discussion ~20 minutes

What is R?

It's a programming language.

It's software.

It's free!

Why is R amazing?

You can run it locally.

You can run it in the cloud.

You can run it on a high performance computing cluster!

There's a great community around the world using R!!

Why should you add R to your workflow

You can reproduce your work.

You can share every step of your analysis.

You can reuse your code for many projects!

You can use **other's** code for many projects!

What does R look like

```
# Text after a hash is a comment  
# It is not run by the computer  
# Can be used as notes to you (or future you!)
```

```
# Here are a few simple calculations
```

```
1 + 1
```

```
## [1] 2
```

```
2 * 3
```

```
## [1] 6
```

```
4^2
```

```
## [1] 16
```

Text

You can manipulate strings of text.

```
host <- "Mara"

greeting <- "Hello World"

paste0(greeting, " from ", host, "!")

## [1] "Hello World from Mara!"
```

You can import data easily

- ▶ Excel
- ▶ Google Sheets
- ▶ SAS
- ▶ SPSS
- ▶ Stata
- ▶ Lot's more

Here's some data in Excel

You might have data like this?

How do you get it into R?

	A	B	C	D
1	count	time	resident	factor_servic
2	1	16.225	Pediatrics	0.55
3	2	15.675	Pediatrics	0.55
4	3	16.225	Pediatrics	0.55
5	4	11.55	Pediatrics	0.55
6	5	5.775	Pediatrics	0.55
7	6	12.375	Pediatrics	0.55
8	7	6.325	Pediatrics	0.55
9	8	4.33125	Pediatrics	0.55
10	9	8.830556	Pediatrics	0.55
11	10	8.415	Pediatrics	0.55
12	1	54.175	Pediatrics	0.55
13	2	31.35	Pediatrics	0.55

Data Import Examples

Like this!

```
#From Excel
```

```
fake_data <-  
  read_excel("./data/qi_spreadsheet_workshop.xlsx",  
    sheet = "scatter")
```

Tables

You can make tables from your data.

Table 1: Recreating booktabs style table

count	time	resident	factor_service
1	16.225	Pediatrics	0.55
2	15.675	Pediatrics	0.55
3	16.225	Pediatrics	0.55
4	11.550	Pediatrics	0.55
5	5.775	Pediatrics	0.55

Tables, behind the curtains

How did I make that table??

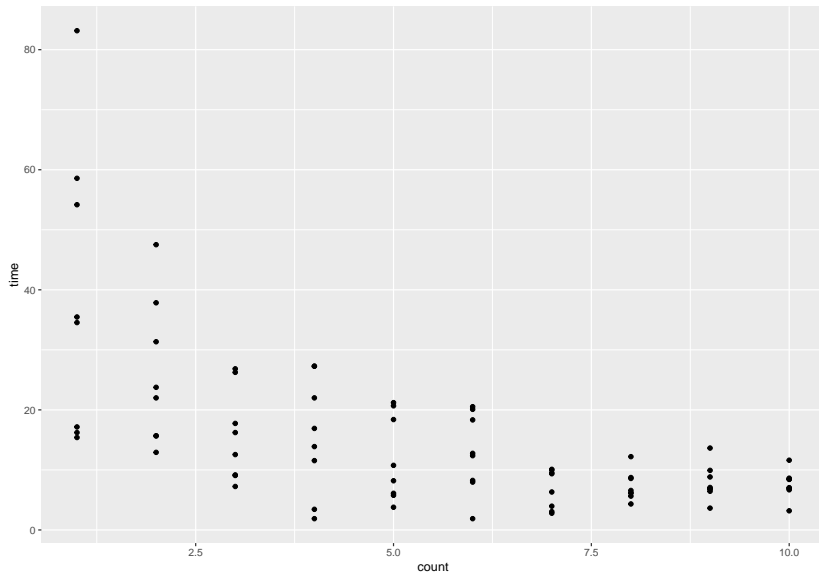
```
head(fake_data, 5) %>%  
  kable(caption = "Recreating booktabs style table",  
        format = "latex",  
        booktabs = T) %>%  
  kable_classic(full_width = F, html_font = "Cambria")
```

Table 2: Recreating booktabs style table

count	time	resident	factor_service
1	16.225	Pediatrics	0.55
2	15.675	Pediatrics	0.55
3	16.225	Pediatrics	0.55
4	11.550	Pediatrics	0.55
5	5.775	Pediatrics	0.55

Plots, simple

You can make plots.

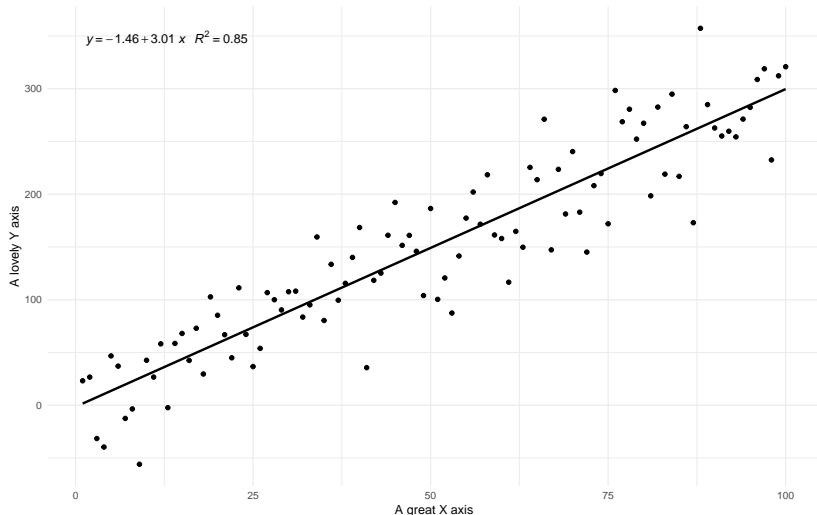


Plots, statistics

You can make plots—with statistical overlay.

What a graph!?!

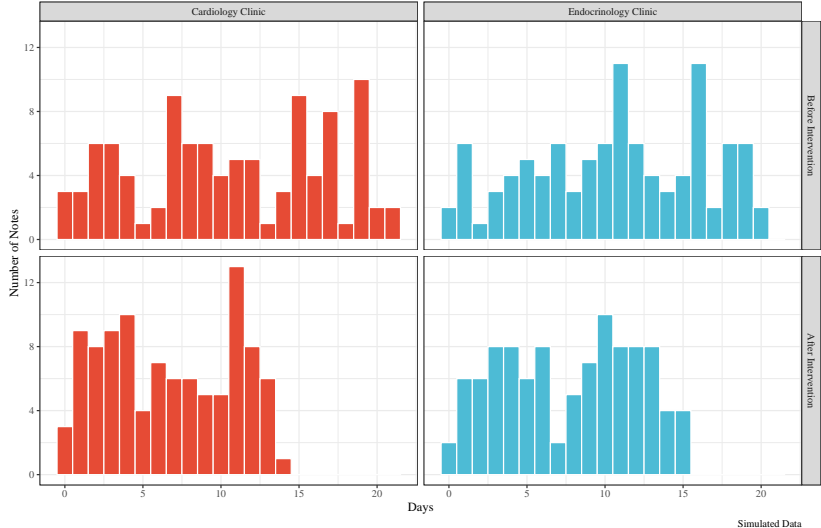
Really swell



This data is simulated

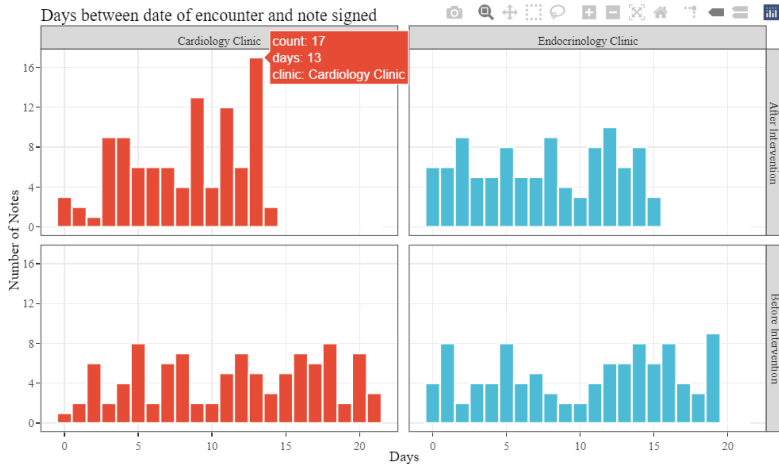
Plots, fancier

Days between date of encounter and note signed
Adult Patients at XYZ Campus



Plots, interactive

Not interactive here because this is a PDF! But here's a picture.
The code to transform the plot is a single function `ggplotly()`!



Code + Text + Figures

With R you can create documents in different formats:

- ▶ Word
- ▶ PDF
- ▶ Slides
- ▶ Websites
- ▶ Blogs
- ▶ Books
- ▶ PNG/JPEG
- ▶ And more!!

ggplot2: Elegant Graphics for Data Analysis

Table of contents

Welcome

Preface to the third edition

Preface to the second edition

Getting started

1 **Introduction**

2 First steps

Layers

Introduction

3 Individual geoms

4 Collective geoms

1 Introduction

1.1 Welcome to ggplot2

ggplot2 is an R package for producing statistical, or data, graphics, based on the Grammar of Graphics,¹ that allows you to make ggplot2 powerful. Rather than being limited to sets or to your specific problem. While the idea of having to learn a there is a simple set of core principles and there are very few the preconceptions that you bring over from using other gra

ggplot2 provides beautiful, hassle-free plots that take care of mean that you can produce publication-quality graphics in s ggplot2's comprehensive theming system makes it easy to d time making your graph look pretty, you can instead focus c

ggplot2 is designed to work iteratively. You start with a layer statistical summarier. This allows you to produce graphics us analysis. This reduces the distance between the plot in your who have not yet developed the structured approach to ana

Figure 1: Book Made with R

Other amazing things

You can mix programming languages together

- ▶ R and Python
- ▶ R and Python and SQL
- ▶ R and Python and SQL and Bash
- ▶ ...

R is flexible!

- ▶ You use R from start to finish with a project
- ▶ You can use it for a single part of your project—eg making plots.
- ▶ You can use it alongside other software like Tableau and REDCap

People know R, or want to!

- ▶ You could hire someone to do R work for you in a day.
- ▶ You probably already have folks at your institution who are R pros!
- ▶ People can upskill and learn R!

Your turn now!

Systems Check

- ▶ Log into Rstudio Cloud
- ▶ If you don't have a link to the materials, message Mara directly in the Zoom chat with your email
- ▶ Raise hand in zoom if you are having trouble
- ▶ Take this [survey](#)
- ▶ If you are an advanced user, open up the `QI_playground.Rmd` and start playing around

Open up the project

Once you are in RStudio Cloud, click on the project called
CIC_2021_QI_R_Workshop



RStudio Cloud

For this workshop we will be using RStudio Cloud.

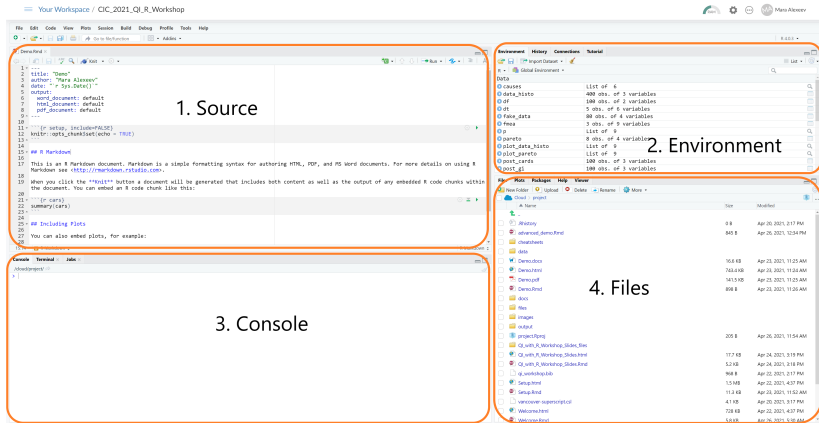
Why? Because there is nothing that you need to download!

You don't **need** to use RStudio to use R, but I do!

RStudio on your computer and RStudio Cloud look very similar.

R and RStudio are free to download to your computer.

RStudio Orientation: Basics



The screenshot displays the RStudio IDE interface with the following components highlighted:

- 1. Source:** The main code editor window showing an R script. The script includes a header with author information, a setup function, and a series of R code chunks for data manipulation and plotting.
- 2. Environment:** The Global Environment pane on the right, showing the loaded data objects: `causes` (list of 6), `data_histo` (400 obs. of 3 variables), `df` (100 obs. of 2 variables), `dt` (5 obs. of 6 variables), `face_data` (100 obs. of 4 variables), `face` (3 obs. of 9 variables), `fp` (list of 8), `pareto` (8 obs. of 4 variables), `plot_data_histo` (list of 9), `plot_pareto` (list of 9), `post_cards` (100 obs. of 3 variables), and `post_g1` (100 obs. of 3 variables).
- 3. Console:** The terminal window at the bottom left, showing the output of the R code execution.
- 4. Files:** The File Manager pane on the bottom right, displaying the file structure of the project, including folders like `data`, `plots`, and `output`, and files like `causes.R`, `data_histo.R`, and `df.R`.

RStudio Orientation: Knit

To knit something in R means to transform it from the raw text and code into a nice output like a PDF or slides. One source material can be knit into many different outputs—from the same source I might make a website, a word document, or a set of slides.

The image shows the RStudio interface with four numbered callouts highlighting key components:

- 1. Source**: Points to the **Knit** button in the Source pane toolbar.
- 2. Environment**: Points to the **Global Environment** pane, which lists objects in the environment.
- 3. Console**: Points to the **Console** pane at the bottom left, used for running code and viewing output.
- 4. Files**: Points to the **Files** pane at the bottom right, which shows the file explorer.

Global Environment

Object	Class	Attributes
data	list	list of 4 variables
data_histo	400 obs.	of 3 variables
df	100 obs.	of 2 variables
df	5 obs.	of 6 variables
df	90 obs.	of 4 variables
fake_data	3 obs.	of 5 variables
fit	list	of 5 variables
fit	8 obs.	of 4 variables
plot_data_histo	list	of 4 variables
plot_pareto	list	of 5 variables
post_cards	100 obs.	of 3 variables
post_fit	100 obs.	of 3 variables

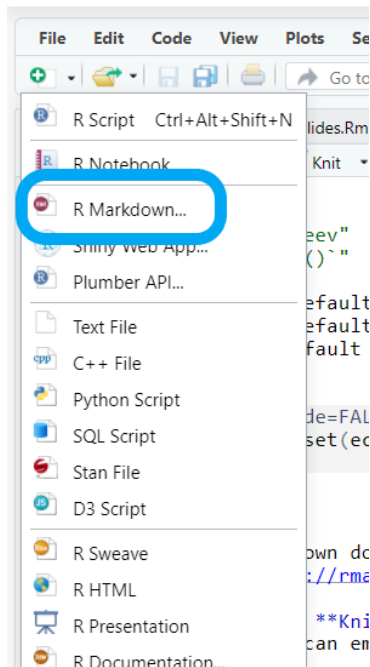
RStudio Orientation: Create a New File

Click there to make a new file.

The screenshot shows the RStudio IDE interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The top toolbar contains icons for creating a new file, opening a file, saving, and running code. The main window is divided into four panels:

- 1. Source:** The top-left panel, containing a code editor with R Markdown syntax. It includes a title, author, date, output type, and a code chunk for generating a summary of cars data.
- 2. Environment:** The top-right panel, displaying the current environment. It lists objects such as data, data_histo, df, of, of, fmea, p, pareto, plot_data_histo, plot_pareto, post_cards, and post_gi, along with their respective data types and dimensions.
- 3. Console:** The bottom-left panel, showing the output of the R code. It displays the result of the `summary(cars)` command.
- 4. Files:** The bottom-right panel, showing the file explorer. It lists files in the current project, including `data`, `data_histo`, `df`, `of`, `fmea`, `p`, `pareto`, `plot_data_histo`, `plot_pareto`, `post_cards`, and `post_gi`.

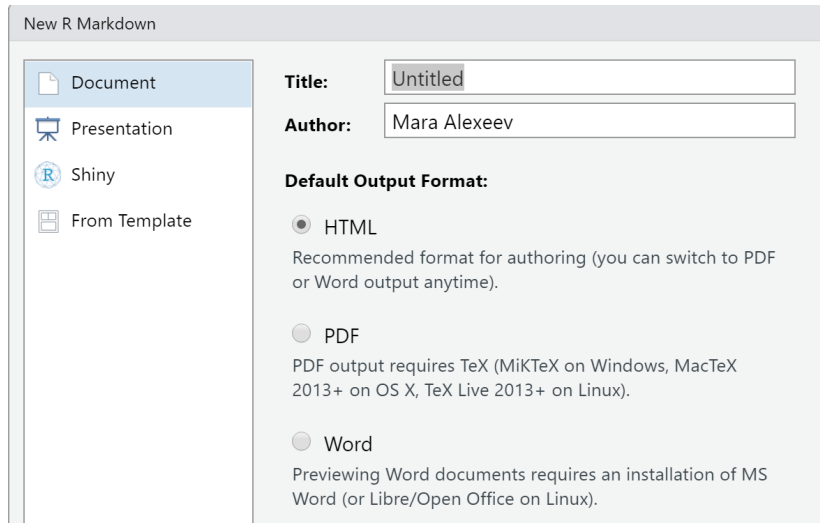
RStudio Orientation: Select R Markdown



Make an R Markdown Document!!

After you have selected R Markdown, you will get a pop up like below.

Title it Hello World! Hit “OK”



The image shows a dialog box titled "New R Markdown". On the left is a sidebar with four options: "Document" (selected), "Presentation", "Shiny", and "From Template". Each option has a corresponding icon. On the right, there are two text input fields: "Title:" with the text "Untitled" and "Author:" with the text "Mara Alexeev". Below these fields is a section titled "Default Output Format:" with three radio button options: "HTML" (selected), "PDF", and "Word". Each option has a descriptive text block below it.

Title:

Author:

Default Output Format:

☒ **HTML**
Recommended format for authoring (you can switch to PDF or Word output anytime).

☐ **PDF**
PDF output requires TeX (MiKTeX on Windows, MacTeX 2013+ on OS X, TeX Live 2013+ on Linux).

☐ **Word**
Previewing Word documents requires an installation of MS Word (or Libre/Open Office on Linux).

View the Rmd

Your new file should look something like this.

```
1 ---
2 title: "Hello world"
3 author: "Mara Alexeev"
4 date: "4/26/2021"
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
15 HTML, PDF, and MS word documents. For more details on using R Markdown see
16 <http://rmarkdown.rstudio.com>.
17
18 When you click the **knit** button a document will be generated that includes both
19 content as well as the output of any embedded R code chunks within the document. You can
20 embed an R code chunk like this:
```

YAML

Code Chunk

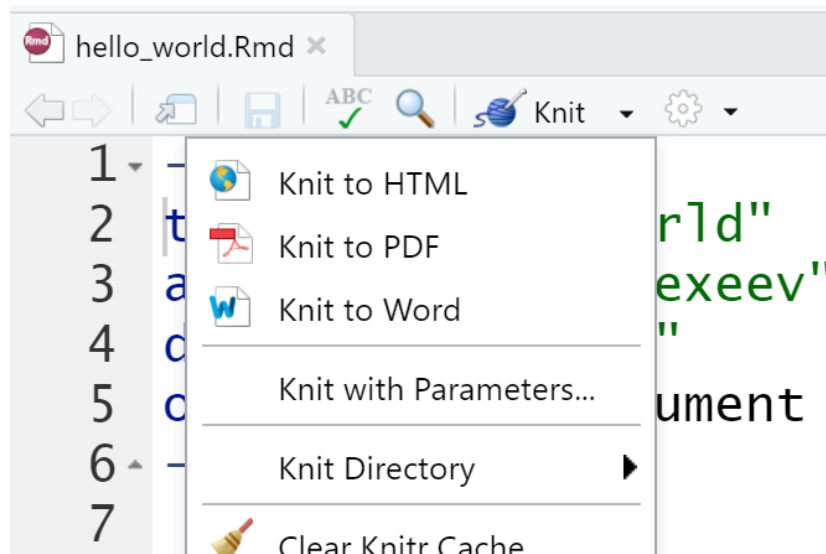
R Markdown Text

Run button

Rmd—What does it all mean?!?

An Rmd file allows you to mix code and text in a single document.

You can then knit an Rmd file to many different outputs.



YAML

The YAML header contains the special instructions on how to create the output document. We won't do much with it here today, but it is a very powerful way to make your Rmd file as bespoke as you want it!

Code Chunks

Code chunks are where the code will go.

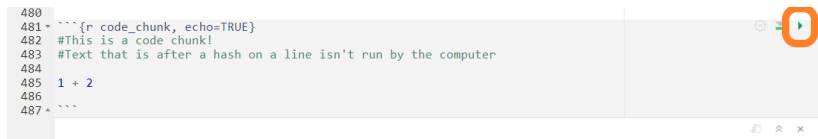
Code chunks have a gray background.

```
#This is a code chunk!  
#Here is a simple calculation  
1 + 2
```

```
## [1] 3
```

You can run a code chunk by pressing the green play button.

```
480  
481 {r code_chunk, echo=TRUE}  
482 #This is a code chunk!  
483 #Text that is after a hash on a line isn't run by the computer  
484  
485 1 + 2  
486  
487 ^
```



Text

The text areas are the white background areas below the YAML header.

You can use R Markdown syntax to generate rich text formatting from simple symbols like the examples below.

Syntax

Plain text

End a line with two spaces
to start a new paragraph.

italics and `_italics_`

****bold**** and `__bold__`

superscript²

~~~~strikethrough~~~~

[link](www.rstudio.com)

# Header 1

## Header 2

## Becomes

Plain text

End a line with two spaces to start a new paragraph.

*italics* and *italics*

**bold** and **bold**

superscript<sup>2</sup>

~~strikethrough~~

[link](#)

# Header 1

## Header 2

# Knit that R Markdown Document!!

Press the knit button.




You'll be asked to save the file; name it `hello_world`



The screenshot displays the RStudio interface with four main panes highlighted by orange boxes and numbered:

- 1. Source:** The top-left pane shows the R Markdown source code. A blue circle highlights the 'Knit' button (a green square with a white 'K') in the toolbar, with a blue arrow pointing to it. The code includes a title, author, date, output type, word document settings, and a code chunk for generating a plot.
- 2. Environment:** The top-right pane shows the current environment. It lists variables such as 'data', 'dt', 'dt\_data', 'p', 'p\_data', 'p\_data\_hist', 'p\_data\_hist\_data', 'p\_data\_hist\_data\_hist', 'p\_data\_hist\_data\_hist\_data', and 'p\_data\_hist\_data\_hist\_data\_hist', along with their types and sizes.
- 3. Console:** The bottom-left pane shows the R console output. It displays the results of the R code execution, including the output of the 'knitr::opts\_chunk\$set' function and the 'summary' function.
- 4. Files:** The bottom-right pane shows the file explorer. It lists files in the current project, including 'Rproj', 'data', 'plots', 'plots\_data', 'plots\_data\_hist', 'plots\_data\_hist\_data', 'plots\_data\_hist\_data\_hist', 'plots\_data\_hist\_data\_hist\_data', and 'plots\_data\_hist\_data\_hist\_data\_hist'.

# View your output

FilesPlotsPackagesHelpViewer



 Publish 

## Hello World

Mara Alexeev

4/26/2021

## 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

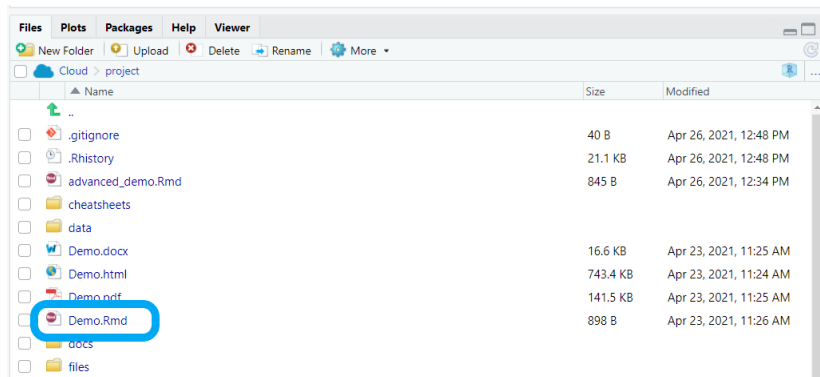
You can also embed plots, for example:



Congratulations!

# Let's start the Demo

Go to your file pane. Open the Demo.Rmd file.



Play!

# Live Demo

During the workshop we will work through the Demo.Rmd.

Ask questions in the Zoom chat or raise your hand.<sup>5</sup>

---

<sup>5</sup>Your Zoom hand!

## Finally, the QI!

After finishing up the Demo.Rmd, we will move on to the QI part.  
Finally!



Open the QI\_playground.Rmd file