Getting Started with R Markdown

Tyler J. Loewenstein   Ryan T. Moore[*]

Wednesday 20th September, 2023 at 15:53

This is a short introduction to using R Markdown on problem sets and labs. Before using this guide, read the *Short Guide to R and RStudio* available at ryantmoore.com/teaching.html.

1 Background

R Markdown is a useful format for integrating plain text information and code into one compiled document. It can be used to create reports, articles, slide presentations, and a variety of other formats for scientific communication. Quarto (`.qmd`) files can be *rendered* or *knitted* to create documents in HTML, PDF, or Microsoft Word. For this class, we recommend a package called `tinytex` to create the PDF files that you will submit to Blackboard. Instructions for installing `tinytex` can be found via at http://www.ryantmoore.org/files/ht/httinytex.pdf.

`.R` files versus `.qmd` files

It can be easy to confuse the two types of files used in RStudio. Both `.R` and `.qmd` files can be used to write and run code, but they will each play a different role when performing data analysis.

A `.R` file will . . .

- Accept only functions and scripts written in R
- Ignore anything that is commented out using the `#` symbol.
- Keep a record of your code as you work
- Allow you to run **one or more lines** of code at a time

[*] Department of Government, American University, Kerwin Hall 228, 4400 Massachusetts Avenue NW, Washington DC 20016-8130. tel: 202.885.6470; fax: 202.885.2967; `rtm` (at) `american` (dot) `edu`; `http://ryantmoore.org`.

A `.qmd` file will ...

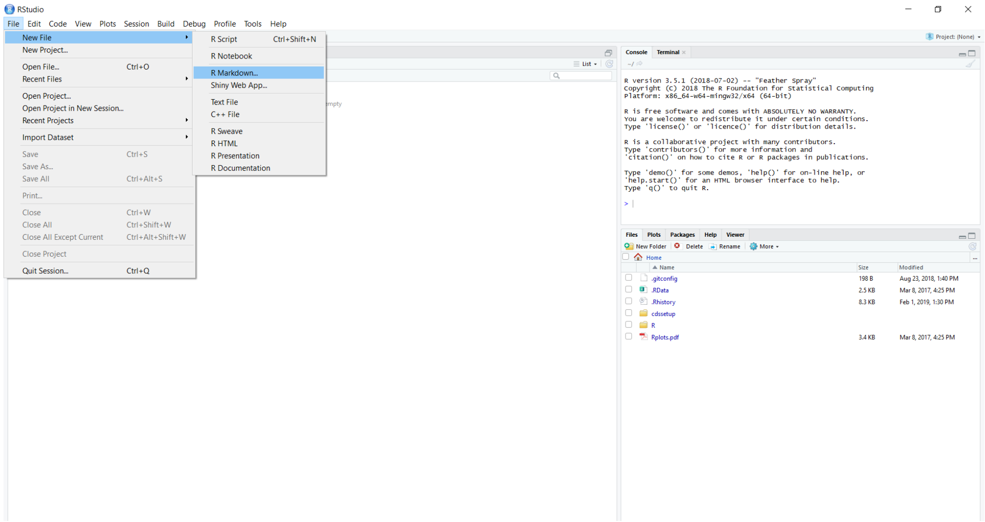
- Accept both code and regular text
- Evaluate R code written in gray chunks (more on this later)
- Allow you to format the document around your code
- Compile **the entire document** before outputting your work

2 Using R Markdown

Now that you know the differences between an `.R` and `.qmd` file, it is important to recognize how they can be integrated into your workflow. This section will illustrate how they can be integrated into your workflow, and even work together.

Creating an `.qmd` file

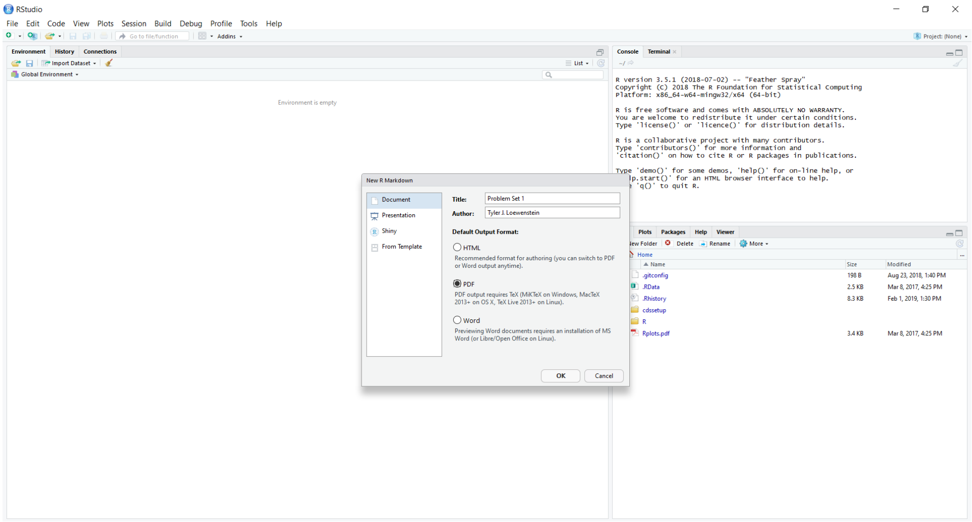
Creating an `.qmd` is similar to creating an `.R` file. To begin, open RStudio and select File > New File > Quarto Document.



```
R version 3.5.1 (2018-07-02) -- "Feather Spray"
Copyright (C) 2018 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
.gitconfig	198 B	Aug 23, 2018, 1:40 PM
.RData	2.5 KB	Mar 8, 2017, 4:25 PM
.Rhistory	8.3 KB	Feb 1, 2019, 1:30 PM
cdssetup		
R		
Rplots.pdf	3.4 KB	Mar 8, 2017, 4:25 PM

RStudio will ask you to give the document a title and select the default output. If you have `tinytex` installed, you should select PDF. Otherwise, choose HTML. Once you have filled in these fields, select OK.



New R Markdown

Document

Presentation

Shiny

From Template

Title: Problem Set 1

Author: Tyler J. Loewenstein

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).

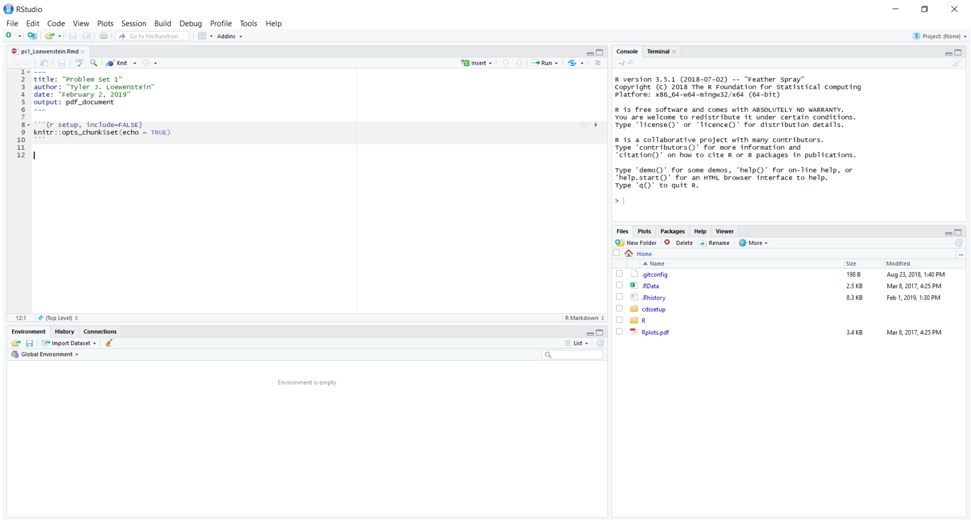
OK Cancel

```
R version 3.5.1 (2018-07-02) -- "Feather Spray"
Copyright (C) 2018 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
lp.start()' for an HTML browser interface to help.
e 'q()' to quit R.
```

Name	Size	Modified
.gitconfig	198 B	Aug 23, 2018, 1:40 PM
.RData	2.5 KB	Mar 8, 2017, 4:25 PM
.Rhistory	8.3 KB	Feb 1, 2019, 1:30 PM
cdssetup		
R		
Rplots.pdf	3.4 KB	Mar 8, 2017, 4:25 PM

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Project: (None)
Environment History Connections
Import Dataset
List
Global Environment
Environment is empty
Console Terminal
Plots Packages Help Viewer
New Folder Delete Rename More
Home

The default .qmd includes some sample text to show new users the kinds of information that can be entered. To remove this, simply delete lines 12 to 31. After you've done that, your file should look like this:



ps1_Loewenstein.Rmd

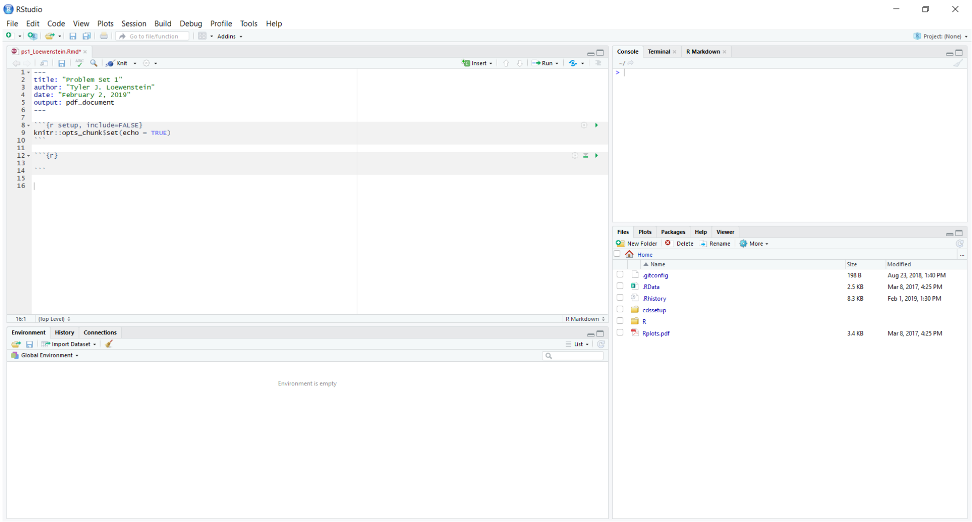
~~~
---
title: "Problem Set 1"
author: "Tyler J. Loewenstein"
date: "February 2, 2019"
output: pdf_document
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```


~~~

```
R version 3.5.1 (2018-07-02) -- "Feather Spray"
Copyright (C) 2018 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
.gitconfig	198 B	Aug 23, 2018, 1:40 PM
.RData	2.5 KB	Mar 8, 2017, 4:25 PM
.Rhistory	8.3 KB	Feb 1, 2019, 1:30 PM
cdssetup		
R		
Rplots.pdf	3.4 KB	Mar 8, 2017, 4:25 PM

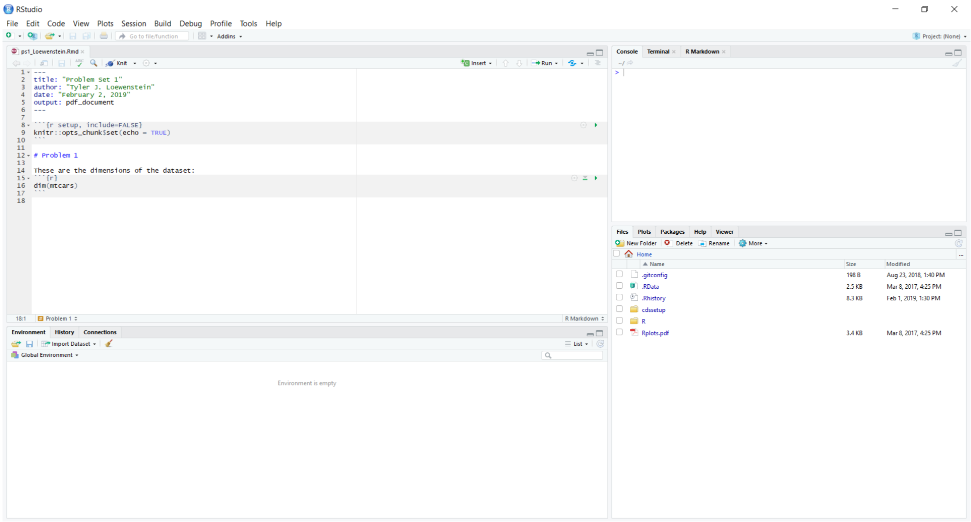
Adding content to an .qmd file

Your new .qmd file will accept both code and regular text. Anything you want R to evaluate (i.e. your code) should be placed inside a code *chunk*. To make a new chunk, select a blank line and type out three backticks followed by curly brackets that contain the letter r. Skip a few lines and type three more backticks to close the chunk. You can also add a chunk by selecting Code > Insert Chunk. Any lines inside a code will appear gray.



~~~
---
title: "Problem Set 1"
author: "Tyler J. Loewenstein"
date: "February 2, 2019"
output: pdf_document
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
~~~

```{r}

```

```

Environment is empty

| Name | Size | Modified |
|---|---|---|
| .gitconfig | 198 B | Aug 23, 2018, 1:40 PM |
| .RData | 2.5 KB | Mar 8, 2017, 4:25 PM |
| .Rhistory | 8.3 KB | Feb 1, 2019, 1:30 PM |
| cdssetup | | |
| R | | |
| Rplots.pdf | 3.4 KB | Mar 8, 2017, 4:25 PM |

Text can be added to your document normally by typing anywhere *outside* of a code chunk. There are numerous ways to customize your `.qmd` using LaTeX, a document preparation system commonly used in social science. To learn more, you can read the *Getting Started with LaTeX* guide via the URL at the top of this guide. For now, you can add a section header by using the # symbol.

Here is an example of how to format the answer to a problem set using these techniques:


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Project: (None)
ps1_Loewenstein.Rmd
Knit
Insert
Run
Console
Terminal
R Markdown
18:1
Problem 1
R Markdown
Environment
History
Connections
Import Dataset
List
Global Environment
Files
Plots
Packages
Help
Viewer
New Folder
Delete
Rename
More
Home

```
---
title: "Problem Set 1"
author: "Tyler J. Loewenstein"
date: "February 2, 2019"
output: pdf_document

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

Problem 1

These are the dimensions of the dataset:
```{r}
dim(mtcars)
```

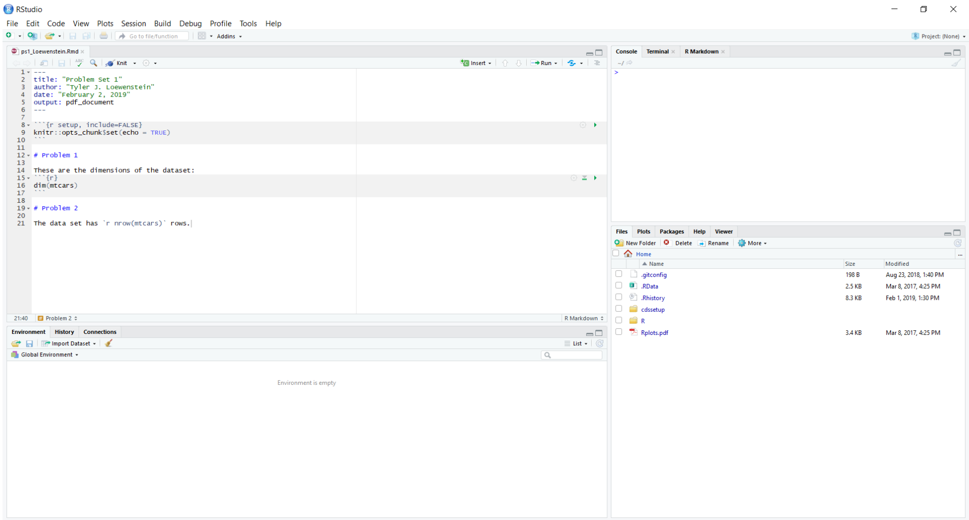
```

Environment is empty

| Name | Size | Modified |
|---|---|---|
| .gitconfig | 198 B | Aug 23, 2018, 1:40 PM |
| .RData | 2.5 KB | Mar 8, 2017, 4:25 PM |
| .Rhistory | 8.3 KB | Feb 1, 2019, 1:30 PM |
| cdssetup | | |
| R | | |
| Rplots.pdf | 3.4 KB | Mar 8, 2017, 4:25 PM |

There is another way to add code into an `.qmd`: inline code. This allows you to write sentences as you would normally, but also integrate the results of your code. This is done by putting your code inside of `` `r` `` instead of a regular chunk. Here's an example:
Instead of typing …

- `The sum of 12 and 45 is: 57.`
```

You can type . . .

- `The sum of 12 and 45 is: `r 12 + 45`.`

Both of which will output . . .

- The sum of 12 and 45 is: 57.

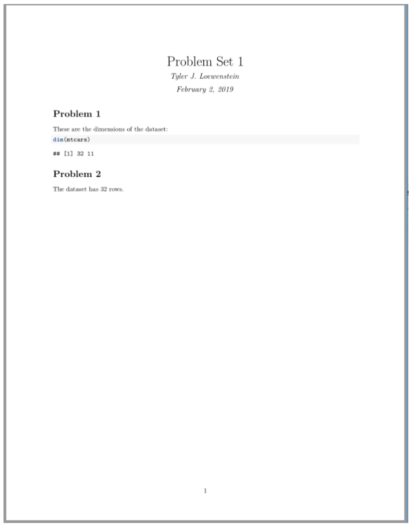


~~~
---
title: "Problem Set 1"
author: "Tyler J. Loewenstein"
date: "February 2, 2019"
output: pdf_document
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
~~~

Problem 1

These are the dimensions of the dataset:
```{r}
dim(mtcars)
```

Problem 2

The data set has `r nrow(mtcars)` rows.
```

| Name | Size | Modified |
|---|---|---|
| .gitconfig | 198 B | Aug 23, 2018, 1:40 PM |
| .RData | 2.5 KB | Mar 8, 2017, 4:25 PM |
| .Rhistory | 8.3 KB | Feb 1, 2019, 1:30 PM |
| cdssetup | | |
| R | | |
| Rplots.pdf | 3.4 KB | Mar 8, 2017, 4:25 PM |

## Compiling a `.qmd` file

The last step in preparing your document is to compile the `.qmd`; this is done through a process called *knitting*. To do this, simply click on the **Render** button when you are ready. Recall that unlike an `.R` file, an `.qmd` will be read and evaluated all at once. If a single line is incorrect or has a missing character, RStudio will produce an error and the document will not be knitted. Here's what a finished PDF document looks like after knitting to PDF:



Problem Set 1

*Tyler J. Loewenstein*

*February 2, 2019*

**Problem 1**

These are the dimensions of the dataset:

```
dim(mtcars)
```

```
## [1] 32 11
```

Problem 2

The dataset has 32 rows.

1
```

With these techniques, you should be able to add both code and narrative response to your `.qmd` file and produce a well-formatted PDF. For problem sets and labs, remember to always include the code used to reach your answer. It's also helpful to use a section header and separate code chunk for each problem.

3 Optimizing Your Workflow

R Markdown is great at creating a professional-looking work product that integrates all of your content. It may be difficult, however, to use an `.qmd` when solving problems and analyzing data. To solve this, we recommend you use both a `.R` and `.qmd` file when starting your analysis. This achieves several goals:

- **Solve problems faster** – You can run your code line-by-line without having to wait for R to compile all of your work.
- **Stay organized** – You can add hundreds of lines to your `.R` without worrying about cluttering your `.qmd`.
- **Archive your process** – You have a record of the work you did, the errors you made, and how you came to the solution.

When working on a new project, think of R Script as your scrap paper and the Quarto file as the final product. Use comments to break up your `.R` file and then copy and paste your work in the `.qmd` after you've solved the problem. This is an iterative process; you should go back to your `.R` file any time you need to test the code in your `.qmd`.

Here are the steps:

1. Open RStudio
2. Create an `.R` file and name it according to your project
3. Create an `.qmd` file with a similar name
4. In your `.R` file, write, test, and edit code for each problem
5. Copy and paste the best code from your `.R` file into your `.qmd` (use a separate code chunk for each problem part)
6. Add text, graphs, and other formatting to your `.qmd`
7. Render your `.qmd`
8. Save your work
9. Repeat steps 4-8 until you've solved every problem!

4 Quarto vs. the Console: Beware the Green Arrow

When you knit a `.qmd` file, R runs the code from the top of the `.qmd` file to the bottom. It does not run other code, and it does **not** look in the Console's workspace. This means your `.qmd` file needs to be entirely self-contained. You need to set the working directory, read the data, create intermediate objects, etc. *within* the `.qmd` file itself.

RStudio provides a small green arrow to the right of your code chunks. When you click the green arrow, R looks in the current workspace for objects, prints the results to the console or plotter, and then reprints them in the confines of your `.qmd`. This behavior is "notebook-like". However, the objects are not really in your `.qmd` file, in the sense that they are not part of the compiled/knit output. When you compile/knit to create an output file, the only code that is run is the code in your `.qmd` file, from top to bottom.

This notebook-like feature is an aspect of RStudio that is not *inherent* in `.qmd` (which could be compiled from a command line outside of RStudio, e.g.). To avoid confusion about the state, we recommend examining your output by knitting your `.qmd` file, rather than by using the green arrow.