

Introduction to R Markdown

Claudia Cabrera MD
CRSP
CWRU

What is R-Markdown?

- Is a tool within RStudio
- Allows you to create documents, presentations, webpages or to style text for Github.
- Combine text with code.
- It saves your document as .Rmd

```

1  ---
2  title: 'LESSON 5: MODIFYING A DATA FRAME'
3  author: "Steven Lewis"
4  date: "July 10, 2017"
5  output:
6    pdf_document: default
7    html_document: default
8  ---
9  ### Using RMarkdown
10
11  ### Text
12
13  Text can be decorated with bold or italics. It is also possible to
14
15  * create [links](http://rmarkdown.rstudio.com/)
16  * include mathematics like  $e=mc^2$  or
17   $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ 
18
19  Be sure to put a space after the * when you are creating bullets and a space
20  creating section headers, but not between $ and the mathematical formulas
21
22  ### Destination formats
23
24  This file can be knit to HTML, PDF, or Word. In RStudio, just select the des
25  and click on `Knit HTML`, `Knit PDF`, or `Knit Word`. Use the dropdown menu
26  change the desired file type.
27
28  ### Documenting file creation
29
30  It's useful to record some information about how your file was created.
31
32  * File creation date: `r sys.Date()`
33  * `r R.version.string`
34  * R version (short form): `r getRversion()`
35  * `mosaic` package version: `r packageversion("mosaic")`
36  * Additional session information
37  ---

```

3:1 # Text ↕



LESSON 5: MODIFYING A DATA FRAME

Steven Lewis

July 10, 2017

Using RMarkdown

Text

Text can be decorated with **bold** or *italics*. It is also possible to

- create links
- include mathematics like $e = mc^2$ or

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

Be sure to put a space after the * when you are creating bullets and a space after # when creating section headers, but **not** between \$ and the mathematical formulas.

Destination formats

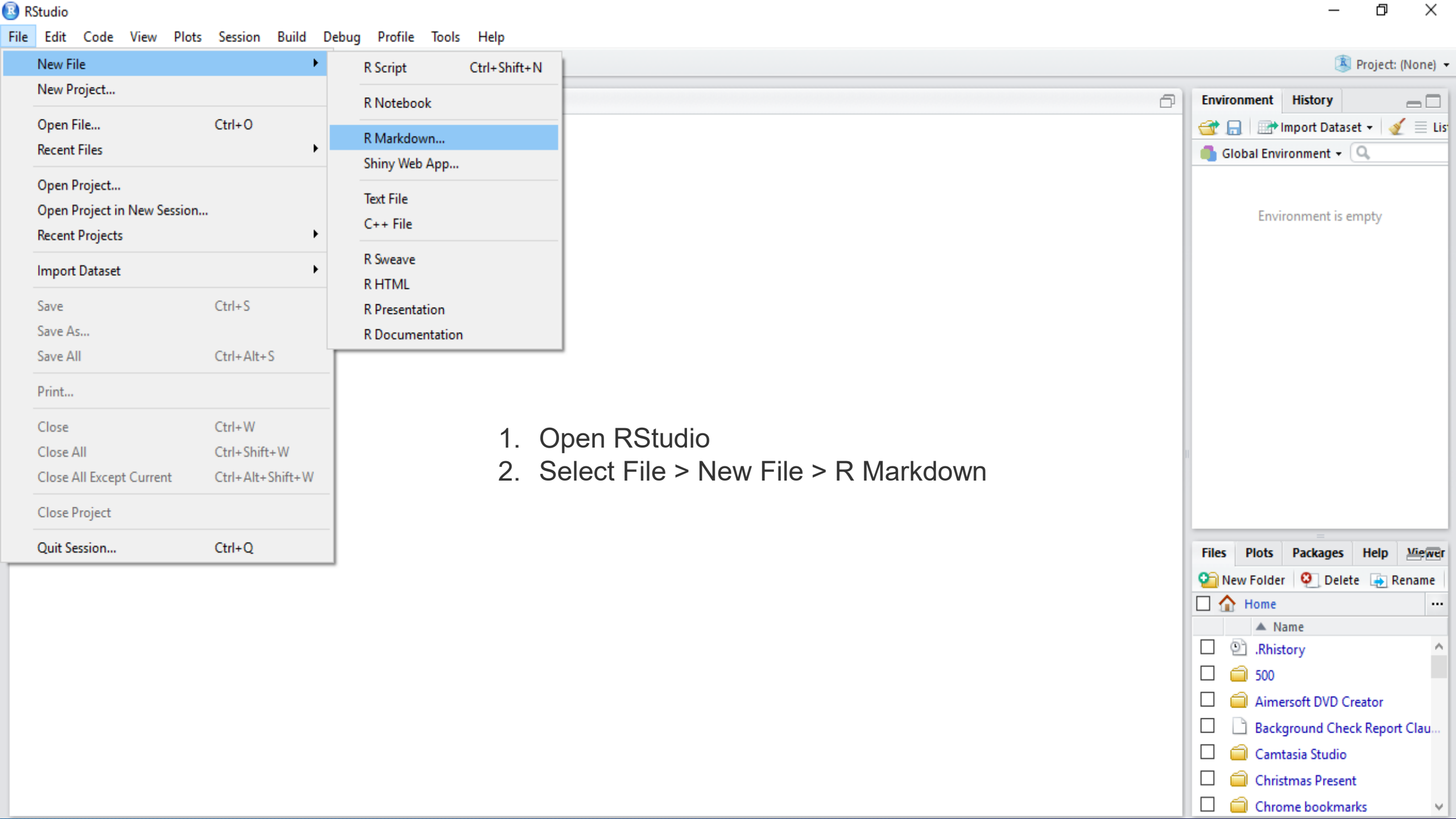
This file can be knit to HTML, PDF, or Word. In RStudio, just select the desired output file type and click on Knit HTML, Knit PDF, or Knit Word. Use the dropdown menu next to that to change the desired file type.

Documenting file creation

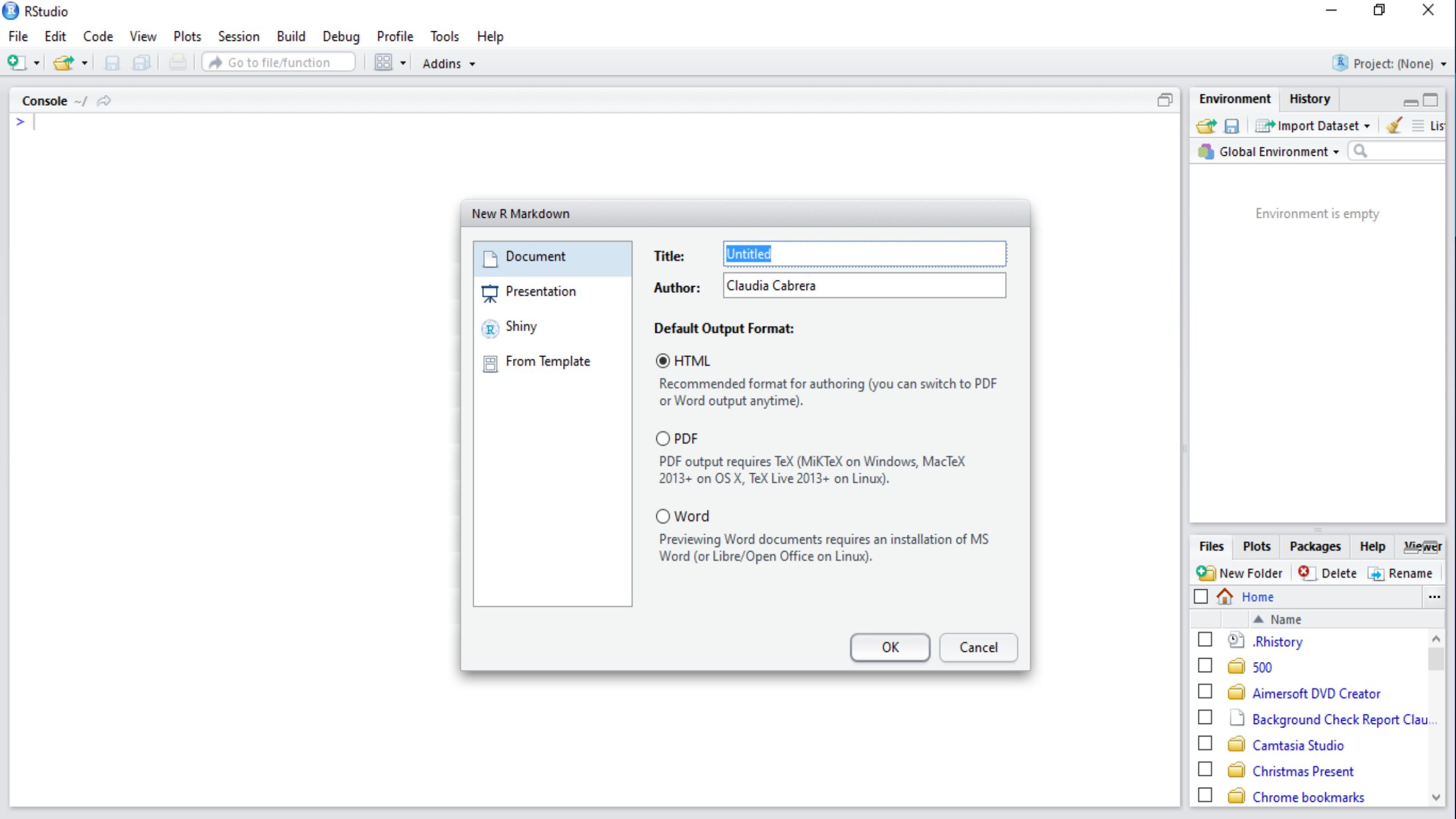
It's useful to record some information about how your file was created.

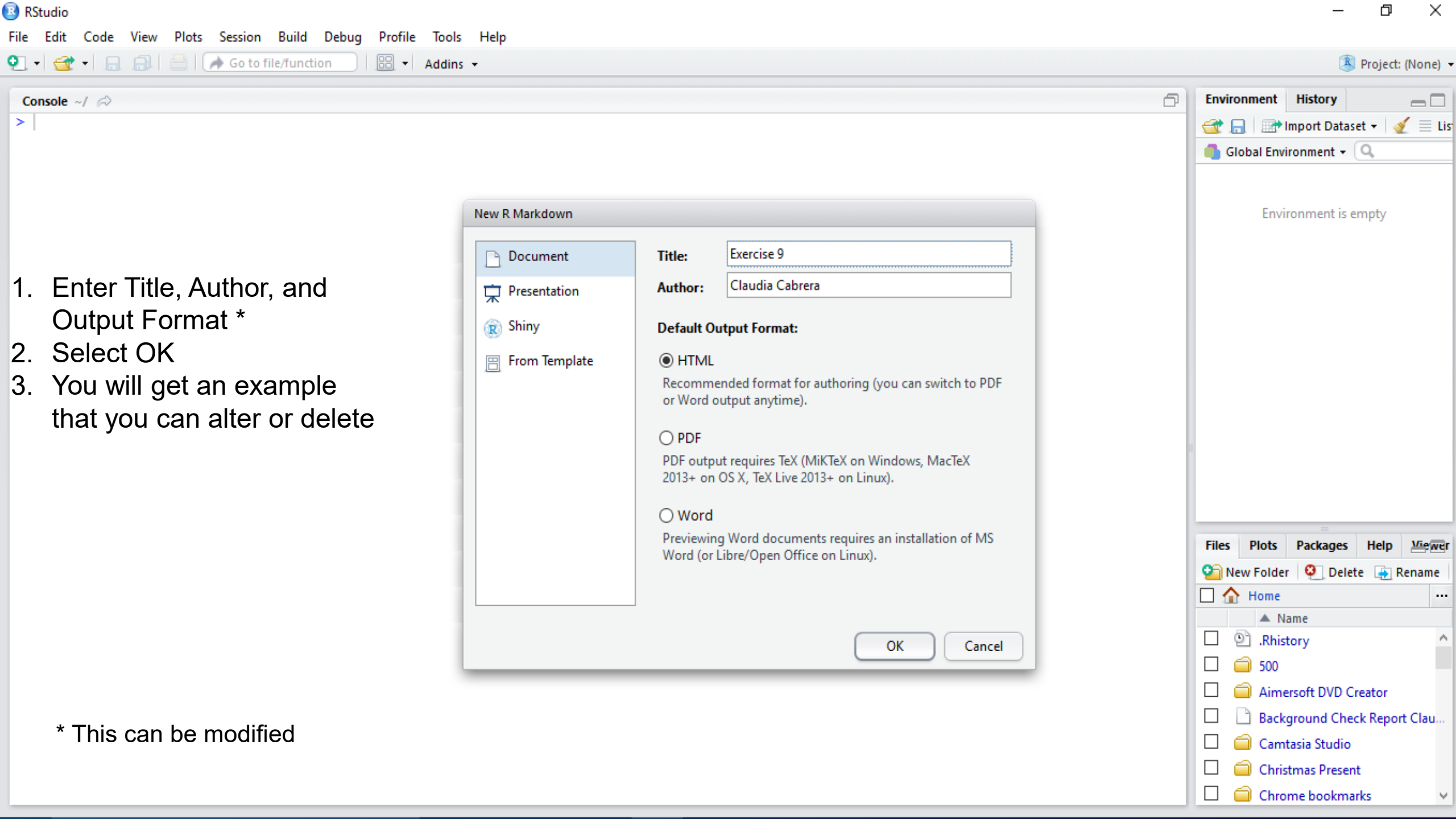
- File creation date: 2017-07-10
- R version 3.4.0 (2017-04-21)
- R version (short form): 3.4.0
- mosaic package version: 0.14.4
- Additional session information

Get Started!



1. Open RStudio
2. Select File > New File > R Markdown












1. Enter Title, Author, and Output Format *
2. Select OK
3. You will get an example that you can alter or delete

* This can be modified

Environment is empty

 New Folder  Delete  Rename

	▲ Name
--	--------

- ☐  .Rhistry
- ☐  500
- ☐  Aimersoft DVD Creator
- ☐  Background Check Report Clau...
- ☐  Camtasia Studio
- ☐  Christmas Present
- ☐  Chrome bookmarks

FileEditCodeViewPlotsSessionBuildDebugProfileToolsHelp

Go to file/function

Addins

Project: (None)

Untitled1*

Knit

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

title: "Exercise 9"

author: "Claudia Cabrera"

date: "July 27, 2017"

output: html_document

```{r setup, include=FALSE}

knitr::opts\_chunk\$set(echo = TRUE)

```

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 code chunks within the document. You can embed an R code chunk like this

```{r cars}

summary(cars)

```

Including Plots

You can also embed plots, for example:

```{r pressure, echo=FALSE}

plot(pressure)

```

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

EnvironmentHistory

Import Dataset

Global Environment

Environment is empty

FilesPlotsPackagesHelpViewer

New FolderDeleteRename

Home

Name

.Rhistory

500

Aimersoft DVD Creator

Background Check Report Clau...

Camtasia Studio

Christmas Present

Chrome bookmarks

2:20

Exercise 9

R Markdown

Console

This is a Chunk!

This is were you will write your code

This is a white space

If you write your code here it will not run*

*With some exceptions

Anatomy of a Chunk

Starting a Chunk

```\${r}```

Closing a Chunk

```

How to create a chunk?

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function

Addins

Project: (None)

Untitled1* x

Knit

Insert

Run

1 ---

2 title: "Exercise 9"

3 author: "Claudia Cabrera"

4 date: "July 27, 2017"

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 within the document. You can embed an R code chunk like this:

17

18 ```{r cars}

19 summary(cars)

20 ```

21

22 ## Including Plots

23

24 You can also embed plots, for example:

25

26 ```{r pressure, echo=FALSE}

27 plot(pressure)

28 ```

29

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

31

R

Python

Rcpp

SQL

Stan

Insert a new R chunk

Environment History

Global Environment

Environment is empty

Files Plots Packages Help Viewer

New Folder Delete Rename

Home

Name

.Rhistory

500

Aimersoft DVD Creator

Background Check Report Clau...

Camtasia Studio

Christmas Present

Chrome bookmarks

5:22 Exercise 9

R Markdown

Console

Code Chunks

- `echo=false`: to hide the R code in the final report
- `results="hide"`: to hide the results/output (figures are shown!)
- `include=false`: to have the chunk evaluated, but neither the code nor its output is displayed
- `warning=false`: suppress any R warnings from being included in the final report
- `message=false`: suppress any R messages from being included in the final report

Code Chunk

FOR FIGURES:

- **fig.height, fig.width:** Specify the height and width of the figure.
- **dpi:** Specifies the pixels per inch. Controls the size of the objects (text, lines, etc.) in your figure.
- **fig.align:** Specify whether your figure appears right, left, or center aligned.

Labeling Code Chunks

- Nothing but letters, numbers, underscores _ and hyphens -
- NO SPACES in Chunk labels
- Every chunk has to have either NO label, or a UNIQUE label (cannot use the same label twice)
- Place a comma after the label if you want to also include a command (like echo = FALSE)
- Why label? So that when your code breaks, you'll know where.

Text in the White Space

Text

- Text can be decorated with **bold** or *italics*.
- It is also possible to :
 - * create [links](http://rmarkdown.rstudio.com/)
 - * include mathematics like $E=mc^2$ or
 - $$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$
- Be sure to put a space after the * when you are creating bullets and a space after # when creating section headers, but **not** between \$ and the mathematical formulas.

Text

Create a new page → \newpage

More than one line between paragraphs →

SECTION HEADERS

#Header 1

##Header 2

###Header 3

####Header 4

Text

- UNORDERED LISTS

- Item
- + Sub-item
- Item
- Item

- ORDERED LISTS

1. Item 1
 2. Item 2
- + Sub-item
3. Item 3

Code Embedded in the White Space

Useful for small pieces of information

It's useful to record some information about how your file was created.

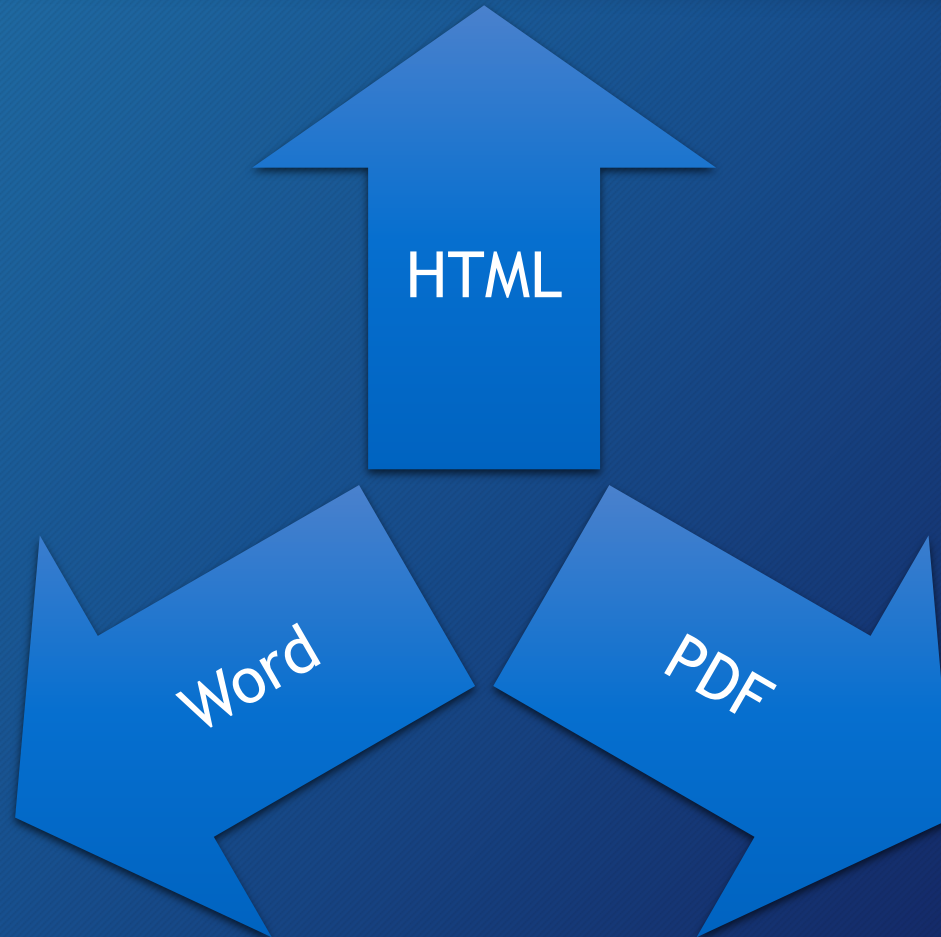
- * File creation date: ``r Sys.Date()``
- * ``r R.version.string``
- * R version (short form): ``r getRversion()``
- * ``mosaic`` package version: ``r packageVersion("mosaic")``
- * Additional session information

Knitting a Document

Knitr

- Is a package in R
- Designed for dynamic report generation.

Destination formats



RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins Project: (None)

R Markdown demo.Rmd

Knit

- Knit to HTML
- Knit to PDF
- Knit to Word
- Knit with Parameters...
- Clear Knitr Cache...

- Select the desired output file type and click on `Knit HTML`, `Knit PDF`, or `Knit Word`.
- Use the dropdown menu next to that to change the desired file type.

```
1 ---
2 tit
3 aut
4 dat
5 out
6 ---
7
8 ---
9 kni
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
15 details on using R Markdown see <http://rmarkdown.rstudio.com>.
16
17 when you click the **knit** button a document will be generated that includes both content as well as the output of any embedded R
18 code chunks within the document. You can embed an R code chunk like this:
19
20 ```{r cars}
21 summary(cars)
22 ```
23
24 ## Including Plots
25
26 You can also embed plots, for example:
27
28 ```{r pressure, echo=FALSE}
29 plot(pressure)
30 ```
31
32 Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.
```

Environment History

Global Environment

Environment is empty

Files Plots Packages Help Viewer

New Folder Delete Rename

Home

Name

- .Rhistory
- 500
- Aimersoft DVD Creator
- Background Check Report Clau...
- Camtasia Studio
- Christmas Present
- Chrome bookmarks

5:22 Exercise 9 R Markdown

Console

LESSON 5: MODIFYING A DATA FRAME

Steven Lewis

July 10, 2017

Using RMarkdown

Text

Text can be decorated with **bold** or *italics*. It is also possible to

- create links
- include mathematics like $e = mc^2$ or

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

Be sure to put a space after the * when you are creating bullets and a space after # when creating section headers, but **not** between \$ and the mathematical formulas.

Destination formats

This file can be knit to HTML, PDF, or Word. In RStudio, just select the desired output file type and click on Knit HTML, Knit PDF, or Knit Word. Use the dropdown menu next to that to change the desired type.

Documenting file creation

It's useful to record some information about how your file was created.

- File creation date: 2017-07-10
- R version 3.4.0 (2017-04-21)
- R version (short form): 3.4.0
- mosaic package version: 0.14.4
- Additional session information

PDF

LESSON 5: MODIFYING A DATA FRAME

Steven Lewis

July 10, 2017

Using RMarkdown

Text

Text can be decorated with **bold** or *italics*. It is also possible to

- create links
- include mathematics like $e = mc^2$ or

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

Be sure to put a space after the * when you are creating bullets and a space after # when creating section headers, but not between \$ and the mathematical formulas.

Destination formats

This file can be knit to HTML, PDF, or Word. In RStudio, just select the desired output file type and click on Knit HTML, Knit PDF, or Knit Word. Use the dropdown menu next to that to change the desired file type.

Documenting file creation

It's useful to record some information about how your file was created.

- File creation date: 2017-08-01
- R version 3.4.0 (2017-04-21)
- R version (short form): 3.4.0
- mosaic package version: 0.14.4
- Additional session information

HTML

LESSON 5: MODIFYING A DATA FRAME

Steven Lewis

July 10, 2017

Using RMarkdown

Text

Text can be decorated with **bold** or *italics*. It is also possible to

- create links
- include mathematics like $e = mc^2$ or

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

Be sure to put a space after the * when you are creating bullets and a space after # when creating section headers, but **not** between \$ and the mathematical formulas.

Destination formats

This file can be knit to HTML, PDF, or Word. In RStudio, just select the desired output file type and click on Knit HTML, Knit PDF, or Knit Word. Use the dropdown menu next to that to change the desired file type.

Documenting file creation

It's useful to record some information about how your file was created.

- File creation date: 2017-08-01
- R version 3.4.0 (2017-04-21)
- R version (short form): 3.4.0
- mosaic package version: 0.14.4
- Additional session information

WORD

For more information

- On using R Markdown → rmarkdown.rstudio.com.
- On Markdown for GitHub → guides.github.com/features/mastering-markdown/

R markdown TIPS

Claudia Cabrera MD

CRSP

CWRU

R markdown

```
library(Hmisc)

## Loading required package: lattice
## Loading required package: survival
## Loading required package: Formula
## Loading required package: ggplot2

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':
##
##   format.pval, round.POSIXt, trunc.POSIXt, units

summary(bmi~sbe, data=analysis)

## bmi      N= 381
##
## +-----+-----+-----+
## |           |      |N| bmi      |
## +-----+-----+-----+
## | sbe       | Long |181| 32.68508|
## |           | Short|200| 30.28000|
## +-----+-----+-----+
## | Overall |      |381| 31.42257|
## +-----+-----+-----+
```

```
library(Hmisc)
```

```
## Loading required package: lattice
```

```
## Loading required package: survival
```

```
## Loading required package: Formula
```

```
## Loading required package: ggplot2
```

```
##
## Attaching package: 'Hmisc'
```

```
## The following objects are masked from 'package:base':
##
##   format.pval, round.POSIXt, trunc.POSIXt, units
```

To get rid of warnings and messages from libraries

- Warning = False
- Message = False

Tip: put your libraries in a different chunk

- Different chunk

R markdown

This 3 ``` shouldn't be here

```
when long when an SD of 20.92 and short when an SD of 17.7.  
104 ▾ ```{r question 8}  
105 agrbmi<- aggregate(analysis["bmi"], by=analysis["sbe"], sd, na.rm=TRUE); agrbmi  
106 agrpackyrs<- aggregate(analysis["packyrs"], by=analysis["sbe"], sd, na.rm=TRUE); agrpackyrs  
107 ```  
108 ▾ ```  
109 All is in blue  
110  
111  
112  
113 ```{r}  
114  
115 ▾ ```  
116
```

Final Document

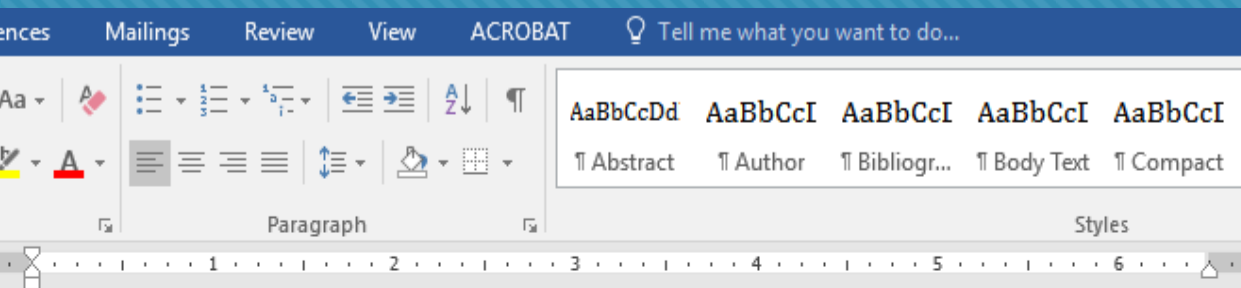
- Put this in another chunk
- Echo = False
- *Unless required

```
setwd("C:/Use[REDACTED]e Drive/R/CRSP 406/DEMOS")
```

1a. load data file

```
load("C:/Use[REDACTED] Drive/R/CRSP 406/DEMOS/DATA/EXPORT/analysis.rdata")
```

Examples on how to organize your documents

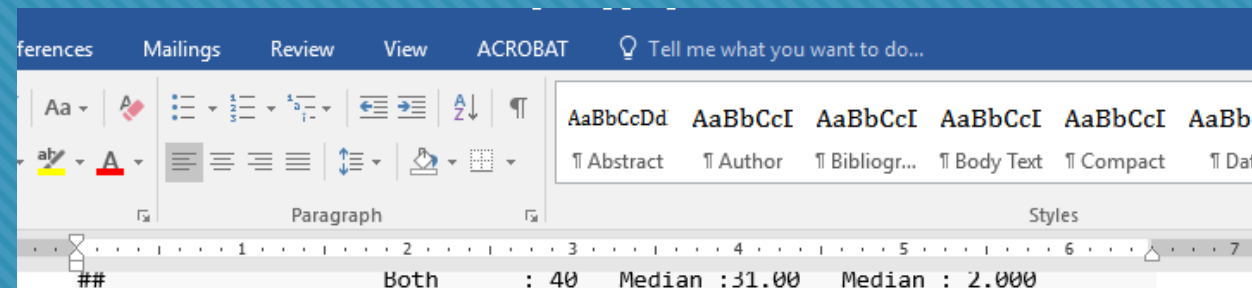


2. Use the summary function to view descriptive statistics for all of the variables within the data frame of 381 patients.

- What is the mean barrett's length?
- What is the mean, median, and range of bmi values?
- What is the mean age at diagnosis?
- How many patients in the sample are female?
- How many patients are Caucasian?
- How many patients in the sample are current smokers?

`summary(analysis)`

```
##      patid      gender      race      charlson      smokstat
## Min.   : 1.0    M:226    Caucasian:291    Min.   :0.00    Current: 92
## 1st Qu.: 98.0    F:155    AA      : 55    1st Qu.:1.00    Never  :171
## Median :194.0                    Hispanic : 21    Median :2.00    Past   :115
## Mean   :194.3                    Other    : 14    Mean   :2.01    NA's   : 3
## 3rd Qu.:292.0                    3rd Qu.:3.00
## Max.   :387.0                    Max.   :9.00
##
##      packyrs      alc.use      alc.abuse      hh.size      ar.dur
## Min.   : 0.00    No :276    No :312    Min.   : 0.000    Min.   : 0.200
## 1st Qu.: 0.00    Yes:105    Yes: 69    1st Qu.: 0.000    1st Qu.: 1.000
## Median : 10.00                    Median : 2.000    Median : 2.000
##
## Mean   : 17.48                    Mean   : 1.882    Mean   : 2.973
## 3rd Qu.: 30.00                    3rd Qu.: 3.000    3rd Qu.: 5.000
```

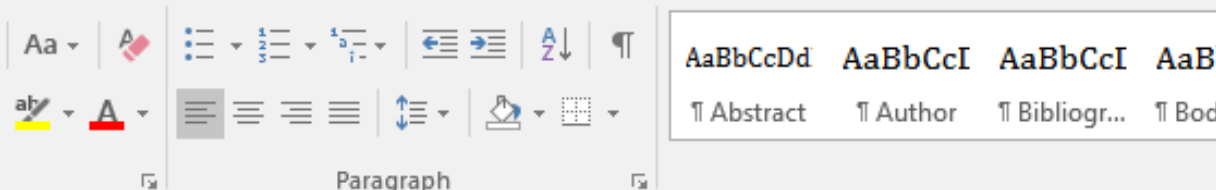


```
##      Both      : 40    Median :31.00    Median : 2.000
##      NA's      : 1    Mean   :31.42    Mean   : 3.582
##      3rd Qu.:35.00    3rd Qu.: 5.000
##      Max.    :56.00    Max.    :20.000
```

```
##      dysplasia      hh      obese      bmi.who
## No      :336      No :172      No :151      Normal      : 43
## Yes     : 28      Yes:209      Yes:230      Overweight   :108
## Undetermined: 17      Moderately Obese :127
##      Severely Obese : 65
##      Very Severely Obese: 38
```

```
##      sbe      age.dx
## long :181      Min.   :21.00
## short:200      1st Qu.:47.00
##      Median :56.00
##      Mean   :55.34
##      3rd Qu.:64.00
##      Max.   :87.00
##      NA's   :1
```

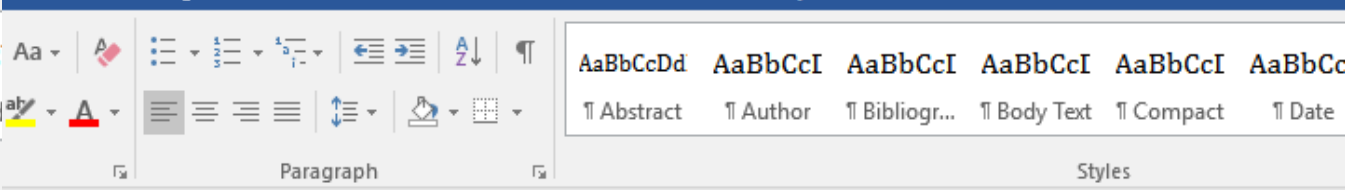
- Answers for question #2:
- 2a) 3.582 cm
- 2b) mean=31.42, median=31, and range 17-56
- 2c) 55.34
- 2d) 155
- 2e) 291
- 2f) 92



2. Use the summary function to view descriptive statistics for all of the variables in the data frame of 381 patients.

`summary(analysis)`

```
##      patid      gender      race      charlson      smok
## Min.   : 1.0      M:226      Caucasian:291      Min.   :0.00      Current
## 1st Qu.: 98.0      F:155      AA      : 55      1st Qu.:1.00      Never
## Median :194.0      Hispanic : 21      Median :2.00      Past
## Mean   :194.3      Other   : 14      Mean   :2.01      NA's
## 3rd Qu.:292.0      3rd Qu.:3.00
## Max.   :387.0      Max.   :9.00
##
##      packyrs      alc.use      alc.abuse      hh.size      ar.c
## Min.   : 0.00      No :276      No :312      Min.   : 0.000      Min.   :
## 1st Qu.: 0.00      Yes:105      Yes: 69      1st Qu.: 0.000      1st Qu.:
## Median :10.00      Median : 2.000      Median :
## Mean   :17.48      Mean   : 1.882      Mean   :
## 3rd Qu.:30.00      3rd Qu.: 3.000      3rd Qu.:
## Max.   :130.00      Max.   :12.000      Max.   :
## NA's    :3
##      ar.freq      ar.type      bmi      barretts
## daily      :339      H2 Blocker: 18      Min.   :17.00      Min.   : 0.
## twice daily: 42      PPI      :322      1st Qu.:27.00      1st Qu.: 1.
##
##      Both      : 40      Median :31.00      Median : 2.
## NA's      : 1      Mean   :31.42      Mean   : 2.
```



```
## Undetermined: 17
## Moderately Obese :127
## Severely Obese   : 65
## Very Severely Obese: 38
##
##      sbe      age.dx
## long :181      Min.   :21.00
## short:200      1st Qu.:47.00
##              Median :56.00
##              Mean   :55.34
##              3rd Qu.:64.00
##              Max.   :87.00
##              NA's    :1
```

- What is the mean barrett's length?
 - 3.582 cm
- What is the mean, median, and range of bmi values?
 - mean=31.42, median=31, and range 17-56
- What is the mean age at diagnosis?
 - 55.34
- How many patients in the sample are female?
 - 155
- How many patients are Caucasian?
 - 291
- How many patients in the sample are current smokers?
 - 92

Question 8

```
aggregate(analysis[c(1,4,6,9,10,13,20)],by=analysis[19],sd,na.rm=TRUE)
```

```
##      sbe      patid charlson  packyrs  hh.size  ar.dur      bmi age.dx  
## 1  long 109.6230 1.686548 26.52844 2.298637 2.767060 5.756663      NA  
## 2  short 114.4805 1.536687 17.71648 2.027053 2.770831 6.473106      NA
```

8a) sd for bmi for short is 6.473106 and for long is 5.756663

8b) sd for pack years does differ across the two groups, for long it's 26.52844, for short is 17.71648

1. Your question
2. Your code
3. Your answer