

R Projects and RMarkdown

AACC University 2020

Shannon Haymond, PhD

1

A familiar scenario?

	A	B	C	D	E	F	G	H	I	J
1	Subject	Gender	Age_months	Race_ethn	Ft_ngdl	Hgb_gdt	MCV_fL	Fe_ugdL	TIBC_ugdL	Tfsat_pct
2	21046	1	52	1	59	13	78.6	17	346	4.9
3	21121	2	56	1	26	13.1	84.9	86	310	26.1
4	21231	2	49	4	15	12.3	85.3	55	279	29.7
5	21208	1	53	3	53	14.5	84.7	92	331	27.8
6	21222	1	46	4	48	11.4	75	57	278	20.5
7	21235	1	52	3	27	12.5	81.1	48	401	12
8	21251	1	52	5	22	13.1	86.6	103	420	24.5
9	21262	2	47	2	24	12.1	86.6	93	353	16.7
10	21280	1	37	3	14	13.1	84.2	36	356	10.1
11	21354	2	49	4	24	12.1	80.1	15	328	4.6
12	21364	2	50	4	41	11.9	82.8	54	395	13.7
13	21400	2	59	1	33	12.5	87.6	44	346	12.7
14	21405	2	37	1	7	13.1	83.9	76	457	16.6
15	21507	1	37	4	23	12.1	78.8	72	556	15.8
16	21578	2	40	4	70	12.4	79.3	75	335	22.4
17	21633	1	47	2	47	14.1	82.3	94	380	24.7
18	21638	1	49	3	38	13.9	79.9	143	380	37.6
19	21641	1	51	1	25	14	79.6	77	443	17.4
20	21695	1	56	4	24	12.2	77.8	81	337	24
21	21749	1	43	4	10	13.1	76.1	65	475	13.7
22	21827	2	41	1	10	12.5	83.4	69	356	19.4
23	21843	1	36	3	19	12	84.1	128	325	59.4



Exploratory Data Analysis

NHANES Survey 2003-2004 Iron Biomarkers in Children 3-5 years
Shannon Haymond, PhD

Report Date: 2020-01-29 14:28:11

Introduction

This report describes a basic exploratory data analysis of laboratory data from the 2003-2004 NHANES Survey measurements of iron status in children aged 3-5 years old.

NHANES 2003-2004 Laboratory Data Page

Data Description

The data file was received from a collaborator who accessed the NHANES data source files and performed preliminary clean up and filtering to yield complete cases of children 1-3 years old with ferritin and CBC parameters. The dataset includes 295 observations with 10 variables.

Summary Statistics

Statistical summary of factor variables

skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
subject	0	1 FALSE		295	210:1, 211:1, 211:1, 212:1
gender	0	1 FALSE		2	Mal:151, Fem:144
race_ethn	0	1 FALSE		5	Non:103, Non:79, Mex:74, Oth:21

5

5

Is your analysis reproducible?
 Are the steps transparent?
 Can you automate the process?
 Can you scale the process?
 Is the process organized?



6

6

There's a better way!

A	B	C	D	E	F	G	H	I	J	
1	Subject	Gender	Age_months	Race_ethn	Ft_ngdl	Hgb_gdt	MCV_fL	Fe_ugdL	TIBC_ugdL	TfSat_pct
2	21046	1	52	1	59	13	78.6	17	346	4.9
3	21121	2	56	1	26	13.1	84.9	80	310	26.1
4	21231	2	49	1	15	12.3	85.3	55	279	29.7
5	21208	1	53	3	53	14.5	84.7	92	331	27.8
6	21222	1	46	4	48	11.4	75	57	278	20.5
7	21235	1	52	3	27	12.5	81.1	48	401	12
8	21251	1	52	5	22	13.1	86.6	103	420	24.5
9	21262	2	47	2	24	12.1	86.6	93	353	16.7
10	21280	1	37	3	14	13.1	84.2	36	356	10.1
11	21354	2	49	4	24	12.1	80.1	15	328	4.6
12	21364	2	50	4	41	11.9	82.8	54	395	13.7
13	21400	2	59	1	33	12.5	87.6	44	346	12.7
14	21405	2	37	1	7	13.1	83.9	76	457	16.6
15	21507	1	37	4	23	12.1	78.8	72	556	15.8
16	21578	2	40	4	70	12.4	79.3	75	335	22.4
17	21633	1	47	2	47	14.1	82.3	94	380	24.7
18	21638	1	49	3	38	13.9	79.9	143	380	37.6
19	21641	1	51	1	25	14	79.6	77	443	1
20	21695	1	56	4	24	12.2	77.8	81	337	1
21	21749	1	43	4	10	13.1	76.1	65	475	1
22	21827	2	41	1	10	12.2	83.4	69	357	40.4
23	21843	1	36	3	19	12	84.1	128	59.4	

R Notebooks
R Projects

R Studio

Exploratory Data Analysis
 NHANES Survey 2003-2004 Iron Biomarkers in Children 3-5 years
 Shannon Haymond, PhD
 Report Date: 2020-01-29 14:28:11

Introduction
 This report describes a basic exploratory data analysis of laboratory data from the 2003-2004 NHANES Survey measurements of iron status in children aged 3-5 years old.

Data Description
 This data file was received from a collaborator who accessed the NHANES data source files and performed preliminary clean up and filtering to yield complete cases of children 1-3 years old with ferritin and CBC parameters. The dataset includes 295 observations with 10 variables.

Summary Statistics
 Statistical summary of factor variables

skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
subject	0	1 FALSE		295	210:1, 211:1, 211:1, 212:1
gender	0	1 FALSE		2	Mal:151, Fem:144
race_ethn	0	1 FALSE		5	Nan:103, Non:79, Mex:74, Oth:21

7

7

<https://bookdown.org/yihui/rmarkdown/>

R Markdown: The Definitive Guide

Yihui Xie, J. J. Allaire, Garrett Grolemund

2019-12-02



Great resource!

8

What is R Markdown?

R Markdown is an R package and a set of tools that are deeply embedded in RStudio

Specific authoring format that enables documents that combine text, R code, and the output from that code (tables, graphics, etc) in a way that is best for human understanding – *facilitates literate programming*

Any R Markdown document can be used as a Notebook

9

9

3

What is an R Notebook?

Notebooks are a special execution mode for R Markdown documents with chunks that can be executed independently and interactively, with output visible immediately beneath the input and Viewer tab – great for iterating code.

By default, R Markdown enables inline output (notebook) mode

All R Notebooks can be rendered to other R Markdown document types

10

10

R Notebook

```
demo.Rmd x
1 ---  
2 title: "My own test"  
3 output: html_notebook ← yellow arrow  
4 ---  
5  
6 ```{r setup, include=FALSE}  
7 knitr::opts_chunk$set(echo = FALSE)  
8 library(tidyverse)  
9 library(kableExtra)  
10  
11 ...
```

R Markdown

```
demo.Rmd x
1 ---  
2 title: "My own test"  
3 output: html_document ← yellow arrow  
4 ---  
5  
6 ```{r setup, include=FALSE}  
7 knitr::opts_chunk$set(echo = FALSE)  
8 library(tidyverse)  
9 library(kableExtra)  
10  
11 ...
```

11

11

R Projects

A built-in feature in RStudio to help you keep all the files associated with a project together — input data, R scripts, analytical results, figures.

A lifesaver for your future self and collaborators!

12

12

What have we learned so far?

Basic understanding of

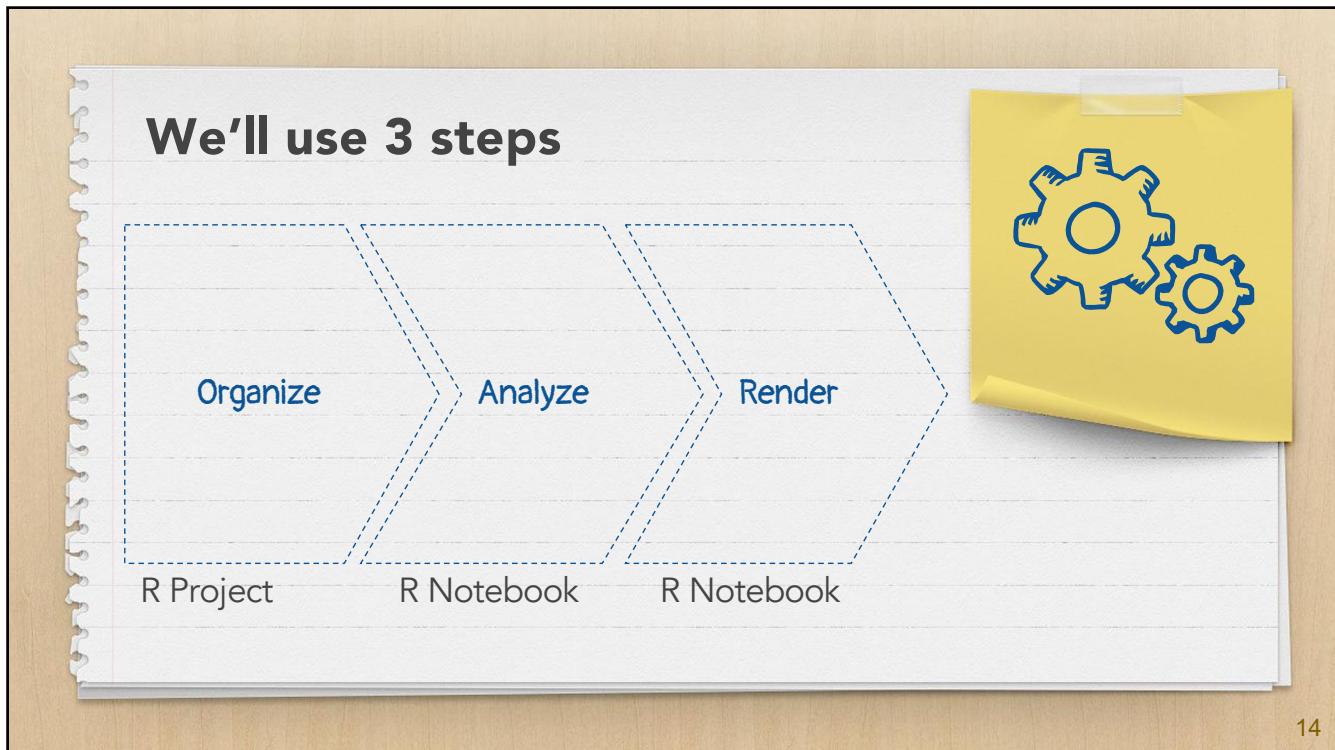
- R vs RStudio – have seen RStudio IDE and some R code
- R Markdown/Notebooks
- R projects



Let's put it all together and go through a very simple workflow for taking structured data, performing basic EDA, and generating a shareable report

13

13



14

14



15

The screenshot shows the RStudio interface. At the top is a dark menu bar with the RStudio logo, followed by a toolbar with several icons. Two yellow arrows point to the first icon on the left, which is a plus sign inside a square, representing the 'New Project' button. To its right is a folder icon. Below the toolbar is a light gray menu bar with tabs: File, Edit, Code, View, Plots, and Session. A search bar labeled 'Go to file/function' is located below the menu bar. To the right of the interface is a yellow sticky note with the text 'Create a Project' written in blue.

1. Use New Project button on toolbar
2. Use New Project command from File tab

16

16

The screenshot shows the RStudio interface with a section titled 'What happens in RStudio?'. Below the title is a numbered list of six items describing the process of creating a project:

1. Creates a project file (.Rproj extension) within project directory
2. Creates hidden directory for project-specific temporary files
3. A new R session is started
4. The current working directory is set to the project directory
5. Loads in whatever was there prior – if options indicate
6. Loads the project and displays name in Projects toolbar.

17

17

hadley RStudio Employee
greg:
It's more reliable and flexible to just restart R any time you want to clean the slate. From RStudio, that's easy enough to do with a keyboard shortcut.

I agree with @greg - it's way better to do this outside of your script than inside of it. Apart from objects and packages there are many other global settings (e.g. options(), par(), environment variables) that won't get cleanly reset. I also highly recommend never saving or loading your workspace:

Options

- General
- Code** (highlighted with a red arrow)
- Appearance
- Pane Layout
- Packages
- Sweave
- Spelling
- GIT/SVN
- Publishing

Default working directory (when not in a project): Browse...

Restore most recently opened project at startup

Restore previously open source documents at startup

Restore RData into workspace at startup

Save workspace to RData on exit: **Never** (highlighted with a red arrow)

Always save history (even when not saving RData)

Remove duplicate entries in history

Use debug error handler only when my code contains errors

Automatically expand trackbacks in error inspector

Default text encoding: UTF-8 Change...

Automatically notify me of updates to RStudio

OK Cancel Apply

That ensures you always get a clean slate when you restart RStudio, and forces you to record all important steps in code.

Start each R session with a fresh environment!

18

18

Files Plots Packages Help Viewer

New Folder Delete Rename More

Home Desktop > Test

Name	Size	Modified
..	0 B	Feb 3,
.Rhistory	205 B	Feb 3,
Test.Rproj		
data		
doc		
figs		
output		
src		

Now all files for this project are together and organized

Set up file system

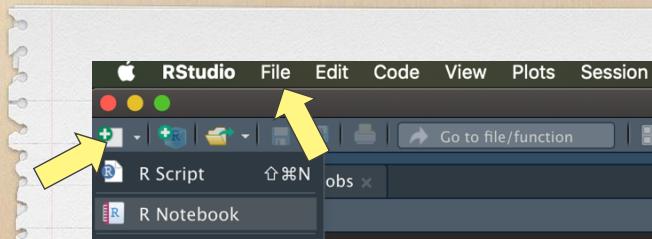
19

19

2. Analyze/Summarize

Next, we'll use R Notebook to analyze/summarize data

20



1. Use New File >> R Notebook button on toolbar
2. Use New File >> R Notebook command from File tab

Create
new R
Notebook

21

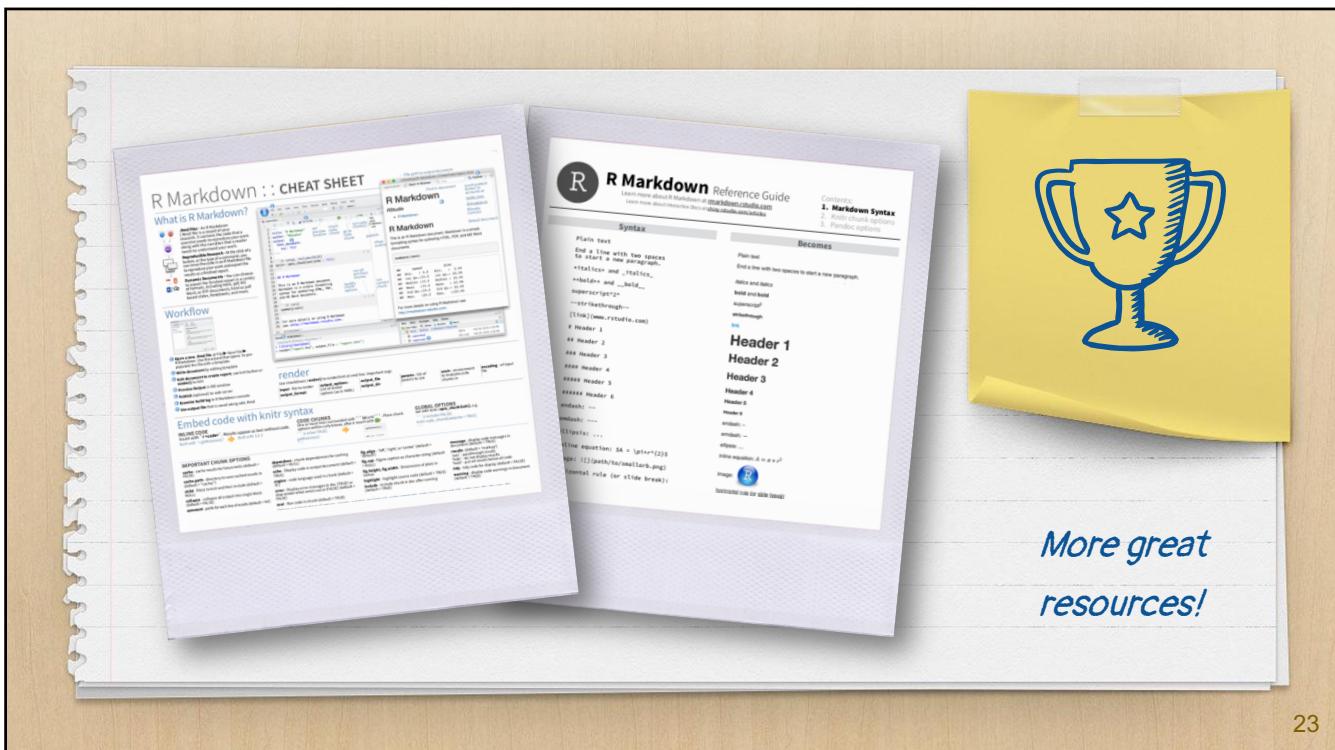
21

```

1 ---
2 title: "R Notebook"
3 output: html_notebook
4 ---
5
6 This is an [R Markdown](http://rmarkdown.rstudio.com) Notebook. When you
7 execute code within the notebook, the results appear beneath the code.
8
9 Try executing this chunk by clicking the *Run* button within the chunk or by
10 placing your cursor inside it and pressing *Cmd+Shift+Enter*.
11
12 ``
13
14 Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by
15 pressing *Cmd+Option+I*.
16
17 When you save the notebook, an HTML file containing the code and output will
18 be saved alongside it (click the *Preview* button or press *Cmd+Shift+K* to
19 preview the HTML file).
20
21 The preview shows you a rendered HTML copy of the contents of the editor.
Consequently, unlike *knit*, *Preview* does not run any R code chunks.
22

```

22



23

Customize the output using the YAML

1. Change the YAML title to your own title
2. Add a YAML option for author

Hint: see page 4 of RMarkdown Reference Guide

24

24

```
1 ---  
2 title: "My Test"  
3 author: Shannon Haymond, PhD  
4 output: html_notebook  
5 ---
```



25

25

Create a new code chunk to define the setup options.

1. Label the chunk, tell R what to do with the chunk when rendering
2. Set options for the document
3. Load necessary packages
4. Execute your code.



Setup chunk goes at the top, so it runs first!

26

26

Add a new code chunk

```
```{r}
```

Note syntax for R chunk

New chunk:  
 \*Insert Chunk\* button on the toolbar or by pressing  
 Cmd+Option+I or Ctrl+Alt+I

Execute code:  
 Press \*play\* button to run entire chunk or select and press Cmd+Enter or Ctrl+Enter

27

27

## Code chunk options

- How code is evaluated and rendered
- How results are displayed
- Features of plots
- Many others – see <https://yihui.org/knitr/options/>

All chunk options have defaults – you only need to worry about setting the options if you want to change from default

28

28

```

7 ```{r setup, include=FALSE}
8 knitr::opts_chunk$set(echo = FALSE)
9 library(tidyverse)
10 library(kableExtra)
11 ...
12 ...
13

```

Chunk label

Prevents code and results for  
this chunk from appearing in the  
finished file (default = TRUE)

Do not display code in output  
document (default = TRUE) –  
here for all chunks

29

29

## Add your own text and create a new code chunk to read in your data.

1. Write some text that describes your report.
2. Create a new code chunk and label it.
3. Tell R what to do with the chunk when rendering.
4. Write the code to read in your data.
5. Execute your code.

30

30

```
14 This notebook was made using RMarkdown and some of its very useful features.
15 |
16
17```{r data, include = FALSE}
18 my_data <- read_csv(file = "data/NHANES_FeMarkers_3to5y.csv")
19 ````
```

31

31

14

# Formatting Text

Inline text formatting requires specific syntax

<https://rstudio.com/wp-content/uploads/2015/03/rmarkdown-reference.pdf>

The screenshot shows the R Markdown Reference Guide. It includes a table comparing 'Syntax' (left) with 'Becomes' (right). Examples include:

- Plain text: End a line with two spaces to start a new paragraph.
- Italics: `_italics_` becomes *italics*.
- Bold: `++bold++` and `__bold__` become **bold** and **bold**.
- Superscript: `superscript2` becomes superscript<sup>2</sup>.
- Strikethrough: `--strickthrough--` becomes ~~strickthrough~~.
- Link: `[link](www.rstudio.com)` becomes [link](http://www.rstudio.com).
- Header 1: `## Header 1` becomes 

# Header 1

.
- Header 2: `### Header 2` becomes 

## Header 2

.
- Header 3: `#### Header 3` becomes 

### Header 3

.
- Header 4: `##### Header 4` becomes 

#### Header 4

.
- Header 5: `##### Header 5` becomes 

##### Header 5

.
- Header 6: `##### Header 6` becomes 

###### Header 6

.
- Endash: `--` becomes –.
- Emdash: `---` becomes —.
- Ellipsis: `...` becomes ....
- Equation: `A = \pi r^2` becomes inline equation:  $A = \pi r^2$ .
- Image: `` becomes
- Horizontal rule: `---` becomes horizontal rule (or slide break).
- Block quote:

  - Unordered list: `* item 1` becomes • item 1.
  - Ordered list: `1. item 1` becomes 1. item 1.

- Table:

Table Header	Second Header
Table Cell	Cell 2
Cell 3	Cell 4

32

# Formatting Tables

Requires additional packages/functions

- knitr::kable for basic tables
- kableExtra for advanced styling

[https://cran.r-project.org/web/packages/kableExtra/vignettes/awesome\\_table\\_in\\_html.html](https://cran.r-project.org/web/packages/kableExtra/vignettes/awesome_table_in_html.html)

33

33

## Make a basic stats summary table with a bold title.

1. Use syntax for text to make the table's title bold.
2. Create a new code chunk and label it.
3. Write the code to create your summary table and format it.
4. Execute your code.

34

34

```
Table 1. Summary Statistics
```{r table}  
knitr::kable(summary(my_data)) %>%  
  kable_styling()  
...````
```

words = words

Function calls for styling table

Function call to make the summary stats table

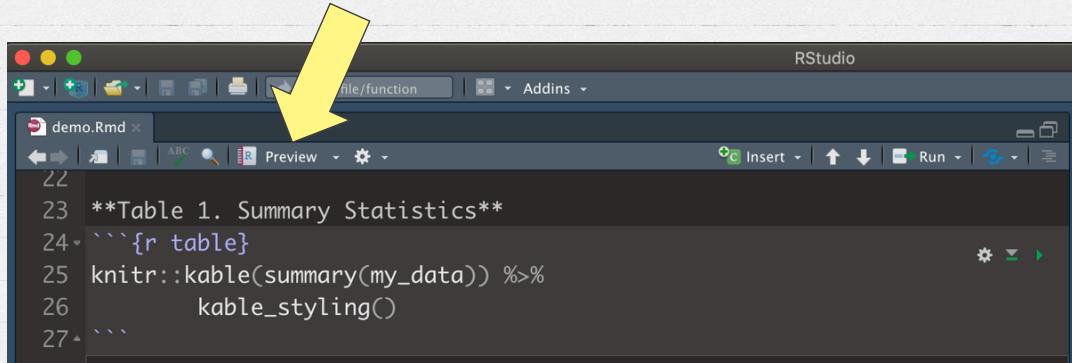
35

35

16

Preview your document

Using an R Notebook allows us to preview the output of our document.



The screenshot shows the RStudio interface with a dark theme. A yellow arrow points to the 'Preview' button in the top toolbar. The code editor window displays the following R Markdown code:

```
22
23 **Table 1. Summary Statistics**
24 ``{r table}
25 knitr::kable(summary(my_data)) %>%
26   kable_styling()
27 ``
```

36

36

Use syntax to write inline code that summarizes results from the table.

1. Use inline code syntax to calculate the number of rows in the data table and include this within a sentence.

37

37

Inline code syntax:
`r code here`

26 There are `r nrow(my_data)` observations in this dataset. The mean age is `r round(mean(my_data\$Age_months),0)` months.



Use backtick, not
single quote!

38

38

Create a scatterplot of TfSat_pct and Fe_ugdL with an italicized title.

1. Use syntax for text to italicize the plot's title.
2. Create a new code chunk and label it.
3. Write the code to create your plot.
4. Execute your code.

39

39

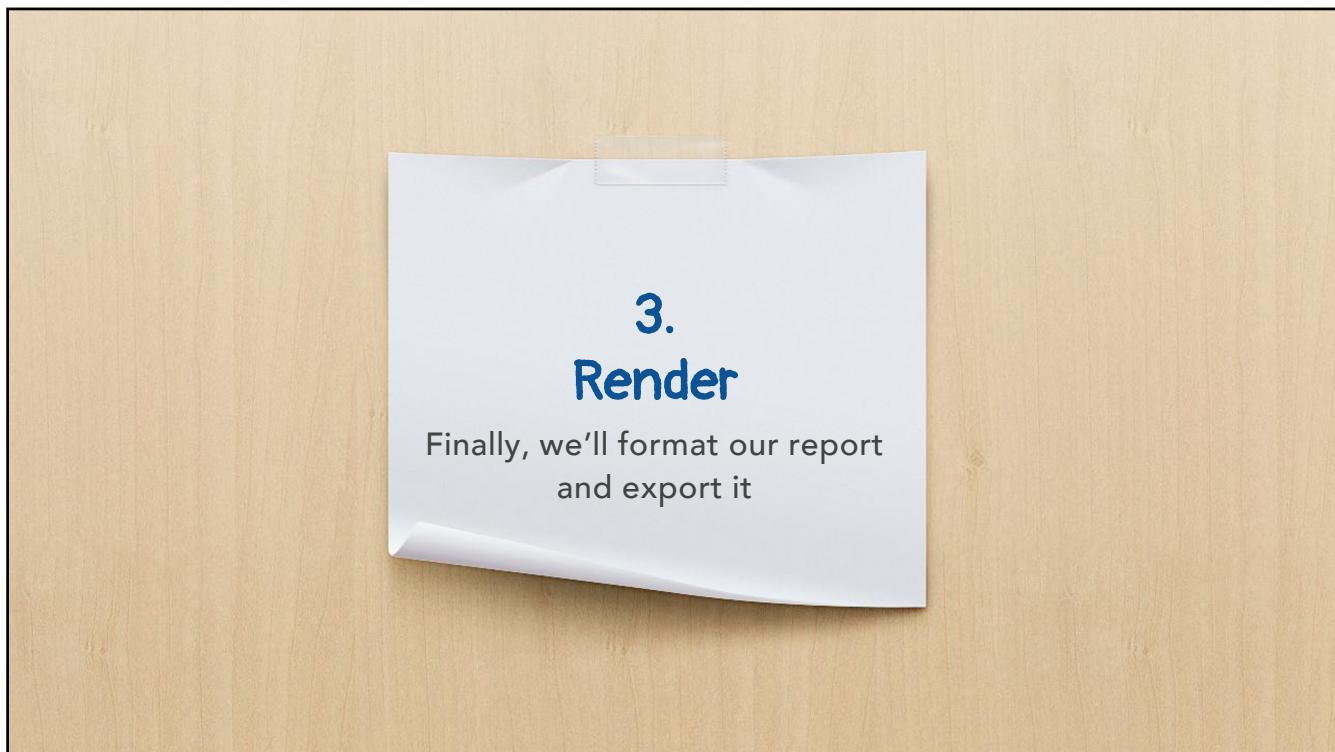
```
28 *Figure 1. Transferrin vs Iron*
29 ````{r plot}
30 ggplot(my_data, aes(x = Fe_ugdL, y = TfSat_pct)) +
31   geom_point()
32 ````
```

words = words

```
ggplot(data = <DATA>, aes(x = <X_DATA>, y = <Y_DATA>)) +
<GEOM_FUNCTION>
```

40

40



41

When you render, R Markdown

1. runs the R code, embeds results and text into .md file with knitr
2. then converts the .md file into the finished format with pandoc



Set a document's default output format in the YAML header:

```

---
output: html_document
---
# Body
  
```

42

output value	creates
<code>html_document</code>	html
<code>pdf_document</code>	pdf (requires Tex)
<code>word_document</code>	Microsoft Word (.docx)
<code>odt_document</code>	OpenDocument Text
<code>rtf_document</code>	Rich Text Format
<code>md_document</code>	Markdown
<code>github_document</code>	Github compatible markdown
<code>ioslides_presentation</code>	ioslides HTML slides
<code>slidy_presentation</code>	slidy HTML slides
<code>beamer_presentation</code>	Beamer pdf slides (requires Tex)

43

R Markdown from R Studio

Gallery

Check out the range of outputs and formats you can create using R Markdown.

<https://rmarkdown.rstudio.com/gallery.html>

44

44

RStudio File Edit Code View Plots Session Build

demo.Rmd x

```
1 -- R Preview Notebook
2 ti Knit to HTML
3 ou Knit to PDF
4 -- Knit to Word
5 
6 . Knit with Parameters...
7 kr Knit Directory cho = FALSE}
8 li Clear Knitr Cache...
9 library(kableExtra)
10 ...
11 ...
```

ABC Preview > ABC

Export our Report

45

45

HTML Reports

- Customization
- Distribution
- Interactivity

My own test

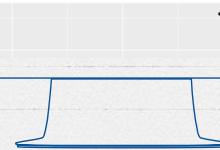
This notebook was made using RMarkdown and some of its very useful features.

Table 1. Summary Statistics

Subject	Gender	Age_months	Race_ethn	Ft_ngdL	Hgb_gdL	MCV_fL	Fe_ugdL	TIBC_ugdL	TISet_pct
Min.:21048	Min.:1,000	Min.:36,00	Min.:1,000	Min.:5,00	Min.:10,00	Min.:83,70	Min.:11	Min.:283,0	Min.:3,00
1st Qu.:23626	1st Qu.:1,000	1st Qu.:42,00	1st Qu.:1,500	1st Qu.:18,00	1st Qu.:12,20	1st Qu.:79,80	1st Qu.:53	1st Qu.:328,0	1st Qu.:14,90
Median:26349	Median:1,000	Median:48,00	Median:3,000	Median:27,00	Median:12,80	Median:82,90	Median:73	Median:352,0	Median:20,20
Mean:26150	Mean:1,488	Mean:48,32	Mean:2,929	Mean:31,27	Mean:12,70	Mean:82,37	Mean:76	Mean:354,4	Mean:21,68
3rd Qu.:28440	3rd Qu.:2,000	3rd Qu.:55,00	3rd Qu.:4,000	3rd Qu.:38,00	3rd Qu.:13,15	3rd Qu.:85,05	3rd Qu.:95	3rd Qu.:378,5	3rd Qu.:27,65
Max.:31085	Max.:2,000	Max.:65,00	Max.:5,000	Max.:148,00	Max.:15,00	Max.:91,20	Max.:288	Max.:498,0	Max.:71,30

There are 295 observations in this dataset. The mean age is 48 months.

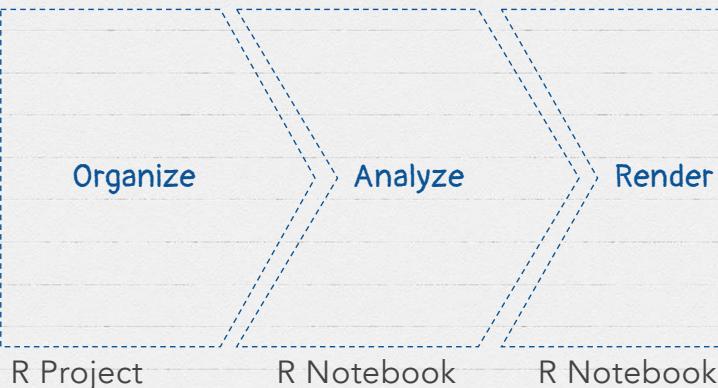
Figure 1. Transferrin vs Iron



46

46

We did it!



47

47

What's Next?

Other output formats,
customization, and interactivity

48



CODE ▾

Text formating

Horizontal lines

Chapter auto numbering

Skip a line

Center an image

White space around img

Footer and header

Space before title

Figures caption

Custom caption

Equations

2 figures in 2 columns

Several columns

Pimp my RMD: a few tips for R Markdown

by Yan Holtz - 10 December 2018 -

R markdown creates interactive reports from R code. This post provides a few tips I use on a daily basis to improve the appearance of output documents. In any case, an unavoidable resource is the Rstudio documentation.

<https://holtzy.github.io/Pimp-my-rmd/>

49

49

The screenshot shows a website with a dark background featuring a close-up of typewriter keys spelling out 'R', 'S', 'A', 'D', 'T', 'H', 'E', 'M', 'P', 'U', 'L', 'C', 'O', 'N', 'G'. The title 'R Markdown Theme Gallery' is prominently displayed in white. The navigation bar includes links for HOME, ABOUT, CONTACT, PUBLICATIONS, TEACHING, and BLOGROLL. Below the title, it says 'Posted by Andrew Zieffler on Thursday, April 12, 2018'.

<https://www.datadreaming.org/post/r-markdown-theme-gallery/>

50

50

The screenshot shows the flexdashboard for R website with a blue header bar containing 'flexdashboard for R', 'Home', 'Using', 'Shiny', 'Layouts', 'Examples', and a link to 'https://rmarkdown.rstudio.com/flexdashboard/'. The main content area features the heading 'flexdashboard: Easy interactive dashboards for R' and a bulleted list of features:

- Use [R Markdown](#) to publish a group of related data visualizations as a dashboard.
- Support for a wide variety of components including [htmlwidgets](#); base, lattice, and grid graphics; tabular data; gauges and value boxes; and text annotations.
- Flexible and easy to specify row and column-based [layouts](#). Components are intelligently re-sized to fill the browser and adapted for display on mobile devices.
- [Storyboard](#) layouts for presenting sequences of visualizations and related commentary.
- Optionally use [Shiny](#) to drive visualizations dynamically.

Below the text are three examples of flexdashboards:

- M&P Training Data**: A heatmap visualization showing training data for various players.
- gauge, geom_text, geom_point**: A dashboard with four panels: 'gauge chart categorical point', 'gauge smooth linear regression', 'geom_point with geom_smooth', and 'Exponentially Smoothing State with state_sensors'.
- Gene Expression Biplotting**: A dashboard with a heatmap titled 'Heatmap: Membrane data matrix for 900 experiments with 762 samples' and a line plot titled 'Time series: Gene expression over time'.

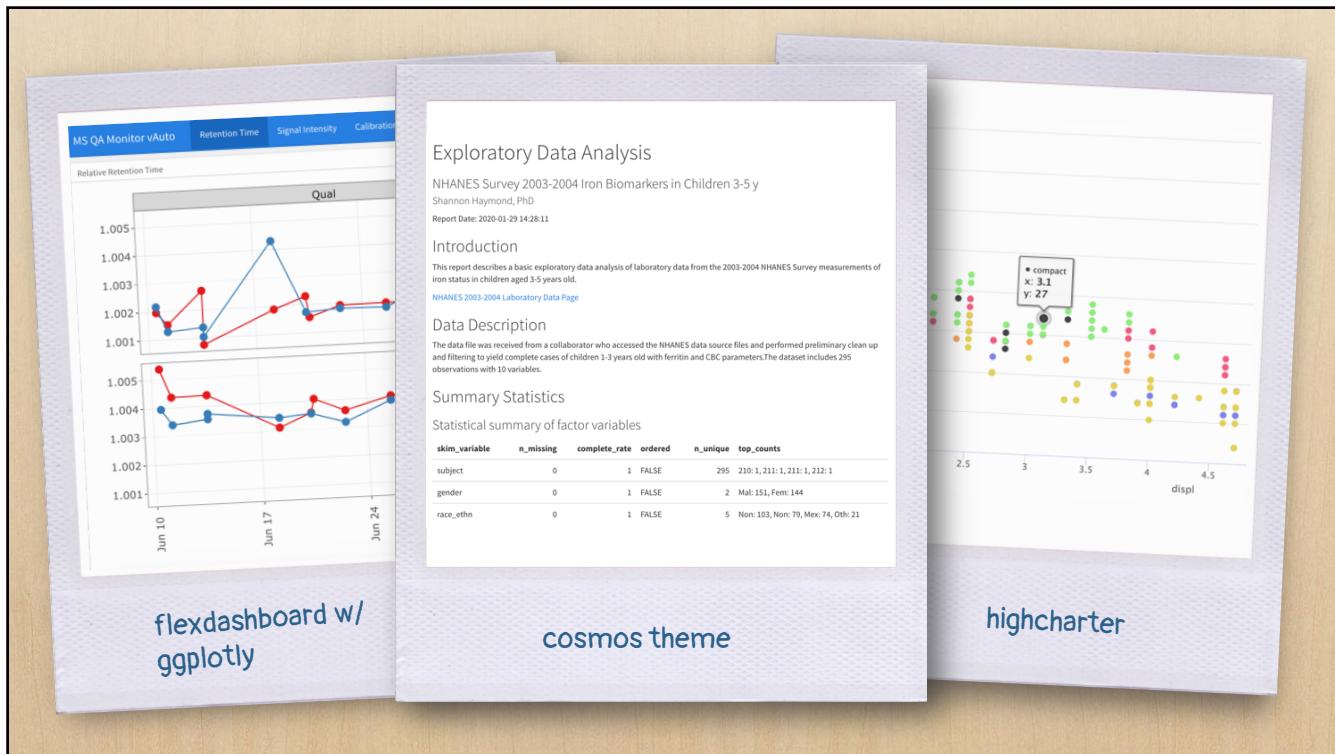
51

The screenshot shows the Plotly website's "Basic Charts" section. The left sidebar includes links for Quick Start, Getting Started, Is Plotly Free?, Cheat Sheet, Figure Reference, User Guide, Use Offline, ggplot2 integration, Dash for R, Examples, and Plotly Fundamentals. The main content area features five examples: Scatter and Line Plots (a scatter plot with a regression line), Line Plots (multiple lines with markers), Bar Charts (a bar chart with a color gradient), Pie Charts (a pie chart with percentages), and More Basic Charts (a large blue arrow pointing right). The URL <https://plot.ly/r/> is visible at the top right.

52

The screenshot shows the HIGHCHARTER website's homepage. The left sidebar lists links for Welcome, Showcase, API, hchart Function, Shortcuts, Themes, Shiny, Highcharts, Highstock, Highmaps, Plugins, and Docs. The main content area displays several charts: a large decorative chart spelling out "HIGHCHARTER" with a dotted pattern, a scatter plot of "displ" vs "hwy" data points, and a line chart titled "Diversity in STAR WARS movies". The URL <http://jkunst.com/highcharter/> is visible at the top right.

53



54



55