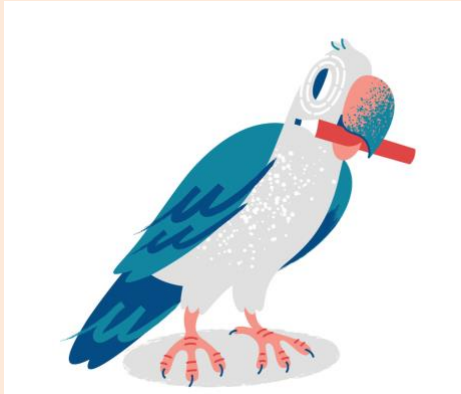


BIOSTATS 640 – Introduction to R
Fall 2023

<https://people.umass.edu/biep640w/webpages/demonstrations.html>



<https://gob.nu/animal-character-designs>

04

Introduction to R Markdown
September 29, 2023

Right click to download R data
hersdata1000.Rdata

Welcome to Lesson 04.

Most of the time, you will be working with R Script and R Markdown files. An **R Script** file is simply a **text file**. That's it. A saved R script file is also a nice repository where you can place fragments of code that you can re-use later. An **R Markdown** file is a **text file + embedded R commands + instructions for output** to a nice looking report. What could be better? These notes are a very rudimentary introduction to the use of R Markdown. Beginners welcome!

Dataset Used: hersdata1000.Rdata

Packages Used: stargazer, summarytools

| | | Page |
|---|--|------|
| 1 | Highlights of Lesson 03– Introduction to Data Visualization using ggplot2 | 2 |
| 2 | Introduction to the HERS Dataset: hersdata.Rdata | 3 |
| 3 | Console <i>versus</i> R Script <i>versus</i> R Markdown | 4 |
| 4 | How to Work with Chunks | 6 |
| 5 | How to Work with R Markdown Markup Language | 11 |
| 6 | Illustration: An R Markdown for Data Inspection | 13 |
| 7 | Some Good Resources. | 16 |
| | | |

1. Highlights of Lesson 03

Introduction to Data Visualization Using ggplot2

| | |
|--|--|
| <p>Tip 1 - Build your graph layer by layer</p> <p>Reminder - The continuation + "tells" R that this command is not yet finished and continues on the next line. For this reason, the + must be at the end of the current line, <i>not</i> at the start of next line.</p> | <p><u>First line:</u> # Execute and correct as needed (must be a dataframe) <code>data = dataframe</code></p> <p><u>Second line is added</u> # Execute BOTH lines 1 and 2 and correct as needed <code>data = dataframe + aes()</code></p> <p><u>Third line is added:</u> # Execute all three lines and correct as needed <code>data = dataframe + aes() + geom_xxx()</code></p> |
| <p>Required (1) dataframe (2) variables (3) type of plot</p> | <ul style="list-style-type: none"> Every ggplot must have these 3 core layers: <code>data = dataframe + aes() + geom_xxx()</code> The continuation + MUST BE at the end of the line (not at the start of the next line) |
| Single discrete variable – Bar Chart | <code>ggplot(data=framingham) + aes(x=sexf) + geom_bar()</code> |
| Single continuous variable – Box and Whisker | <code>ggplot(data=framingham) + aes(x="", y=sbp) + geom_boxplot()</code> |
| Two continuous variables – XY Scatterplot | <code>ggplot(data=temp) + aes(x=bmi, y=sbp) + geom_point(na.rm=TRUE)</code> |

Important Update/correction on handling missing values NA:

For many geom_XXX(), the option na.rm=TRUE does not remove the NA observations before plotting. In order to accomplish this, it is necessary to create a "plot-specific" dataframe that has complete observations only. To do this, use the complete.cases() function. Here is an example

```
myplot_dataframe <- source_dataframe[complete.cases(source_dataframe), ]
```

2. Introduction to the HERS Data

[hersdata1000.Rdata](#)

Source: Hulley S, Grady D, Bush T, Furberg C, Herrington D, Riggs B and Vittinghoff E (1998). Randomized trial of estrogen plus progestin for secondary prevention of heart disease in postmenopausal women. The Heart and Estrogen/progestin Replacement Study. *Journal of the American Medical Association*, **280**(7), 605-613.

In the HERS study, Hulley et al. (1998) sought to determine if exercise, a modifiable behavior, might lower the risk of diabetes in non-diabetic women who were at risk of developing the disease. The question is a complex one because there are many risk factors for diabetes. Moreover, the type of woman who chooses to exercise may be related in other ways to risk of diabetes, apart from the fact of her exercise habit. For example, women who exercise regularly are typically younger and have lower body mass index (BMI); these characteristics also confer a risk benefit with respect to diabetes. Finally, the benefit of exercise may be mediated through a reduction of body mass index. Vittinghoff, Glidden, Shiboski and McCulloch (2005) consider portions of this data in their 2005 text, *Regression Methods in Biostatistics: Linear, Logistic, Survival and Repeated Measures Models* (Springer). Their dataset has n=2,763 observations on 37 variables.


In this lesson, I use this dataset to illustrate the (*highly recommended!*) practice of always inspecting one's data. Here, we consider a random sample of 1000 observations on eight (8) variables in the HERS dataset.

Data Dictionary

| Position | Variable | Variable Label | Type | Codes | Missing data |
|----------|----------|------------------------------------|-----------|--|--------------|
| 1 | HT | Hormone Therapy | numeric | 0 = placebo 1 = hormone therapy | None |
| 2 | age | Age in years | numeric | Range: [44, 79] | None |
| 3 | raceth | Race/ethnicity | character | "1" = White "2" = African American "3" = Other | None |
| 4 | exercise | Exercise at least 3x/week | numeric | 0 = no 1 = yes | None |
| 5 | diabetes | Diabetes | numeric | 0 = no 1 = yes | None |
| 6 | BMI | Body Mass Index, kg/m ² | numeric | Range: [15.49, 49.51] | Yes |
| 7 | glucose | Glucose, mg/dl | numeric | Range: [67, 294] | None |
| 8 | LDL | LDL Cholesterol, mg/dl | numeric | Range: [36.8, 365.2] | Yes |
| | | | | | |

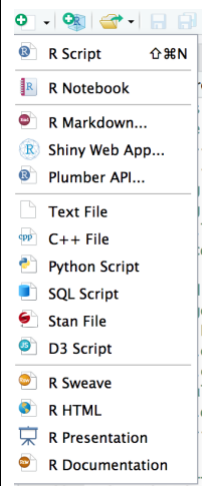
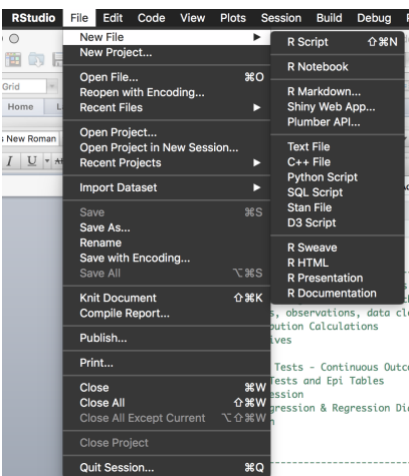
3. Console *versus* R Script *versus* R Markdown

| | Console | R Script | R Markdown |
|------------------|---|--|---|
| At a Glance | Execute | Execute + Save Code | Execute + Save Code + Produce Output |
| What is it | R Studio interface. Here, commands written and executed immediately. | Text editor Here, commands are written. + Utility This utility sends commands to the console for execution. | Chunks R commands are written inside shaded gray chunks. + Utility This utility sends commands to the console for execution. + Text Markup Instead of MS Word, the R Markdown markup language is used to format your narrative (italics, bold, lists, tables, etc.) + Additional Utilities Additional utilities render/knit your work to produce a product in any of several formats (e.g., html, PDF, Word). |
| Where to find it | Console Pane Usually at lower left. <u>How to change the location:</u> tools > global options > tab BASIC > pane layout | Source/Editor Pane Usually at upper left <u>How to change the location:</u> tools > global options > tab BASIC > pane layout | Source/Editor Pane Usually at upper left <u>How to change the location</u> tools > global options > tab BASIC > pane layout |

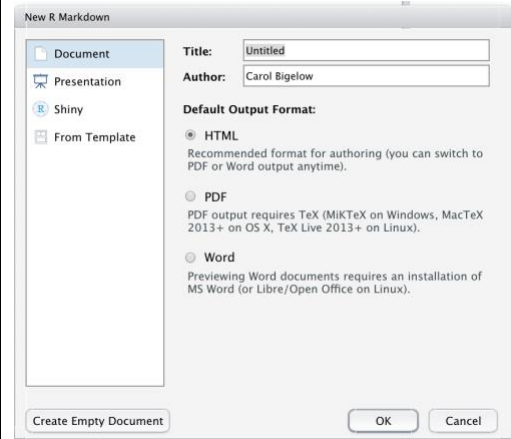
| | Console | R Script | R Markdown |
|----------------------|---|---|---|
| Suggested use | <p>As a <u>calculator</u></p> <p>When learning an R command</p> | <p>To save <u>code fragments</u> as boilers for future work</p> | <p>To archive a <u>full report</u> of text, code, results, graphs, etc.</p> <p><i>ie, for pretty much everything!</i></p> |
| Good to know | <p>Work <u>cannot</u> be saved</p> <p>The command prompt is ></p> <p>Retrieve command </p> <p>+ means your command is incomplete; R is waiting for you to type the rest of it.</p> <p>If you don't know the rest, hit <esc></p> <p><u>How to cancel a command</u> (it might take a while) <control> - c</p> | <p>- Work can be saved as an “.R” file</p> | <p>- Work can be saved as an “.Rmd” file</p> <p>- <u>File must be “self contained”</u> Commands that require external objects or utilities cannot be rendered; e.g.,</p> <p>- View()</p> <p>- install.packages()</p> |

4. How to Work with Chunks

Create a new R Markdown file (2 ways):
(1) From the Source/Editor pane, click on either (+) or, (2) from the toolbar, do File > New File > R Markdown

| The “+” symbol > R Markdown | FILE > NEW FILE > R Markdown |
|--|--|
|  |  |
| | |

Make your choices

| | |
|--|--|
|  | <p>In the box Title: Enter a title of your choice In the box Author: Edit (or not) as you like</p> <p>Under Default Output Format: Leave as is At left, “Document” is highlighted: Leave as is</p> <p>Click OK</p> |
| | |

Delete the pre-supplied R code chunks and text, *but do not delete the header*.

R returns a new R Markdown with some pre-supplied chunks and texts. These are here to help you understand chunks. In practice, you will delete the unwanted chunks and text.

IMPORTANT lines that must NOT be deleted comprise a required header: These "header" lines are at the top and are enclosed between two sets of `---`. They are also called a YAML for “yet another markup language”. Do NOT delete the `---`

```

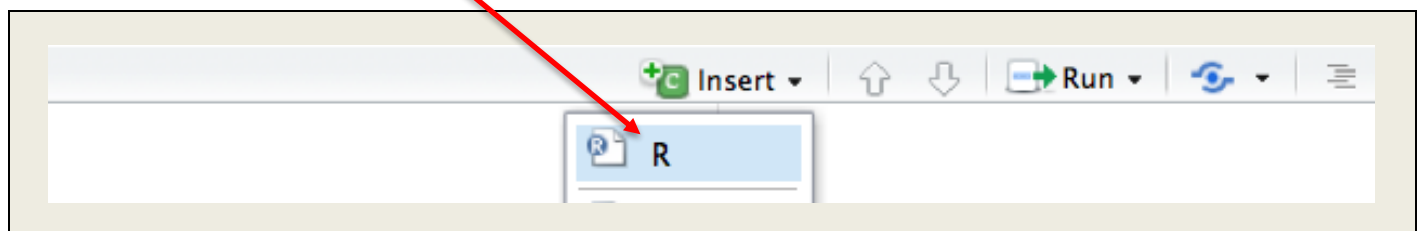
1 ---
2 title: "Untitled"
3 author: "Carol"
4 date: '2023-03-03'
5 output: html_document
6 ^ ---
7
8
9
10
11
12
13
14
15
16
17
18
19

```

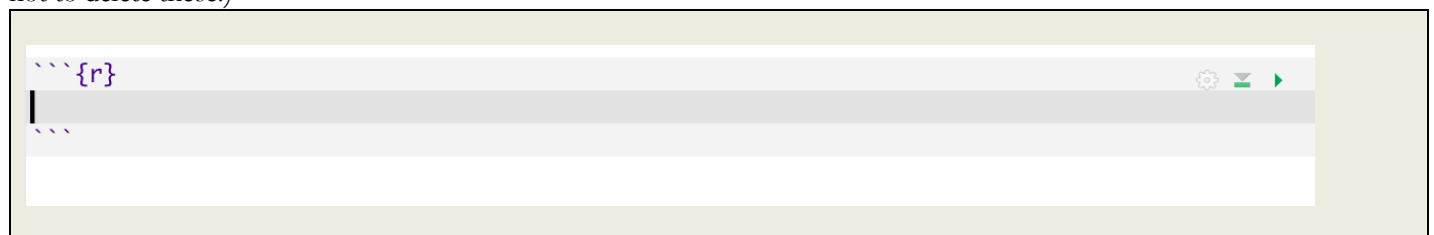
How to Create a New Chunk A chunk is what contains R code in an R Markdown

Click on the little green “insert a chunk” icon at top (on the right).

From the drop down menu, choose R



R returns a blank chunk comprised of: 1) gray shaded box; and 2) ````{r}` at the top; and 3) ````` at the bottom (take care not to delete these!)



Inside the curly brackets `{r }` at the start of your chunk: (1) give your chunk a name (*optional*); and/or (2) supply the options you want (*also optional*), taking care to separate with commas. See example below.

Here are some recommended options.

| Chunk Option | Description |
|----------------------------|---|
| <code>echo=FALSE</code> | Report will show output R code will NOT be shown |
| <code>include=FALSE</code> | Report will NOT show output and R code will NOT be shown So why do this???? . Sometimes, you want to run a chunk as a preliminary to a later chunk and you do NOT want the results of this preliminary chunk to appear in the final report |
| <code>message=FALSE</code> | Nice. This will prevent messages from appearing in the final report. Tip - This is handy when you call a package and you don't want messages to appear in your final report. |
| <code>warning=FALSE</code> | Also nice. This will prevent warnings from appearing in the final report. |
| | |

Example

```
```${r} descriptives, message=FALSE, warning=FALSE`
library(summarytools)

```
```

Key. The name of this chunk is `descriptives`. This is followed by a comma and two options, also separated by a comma

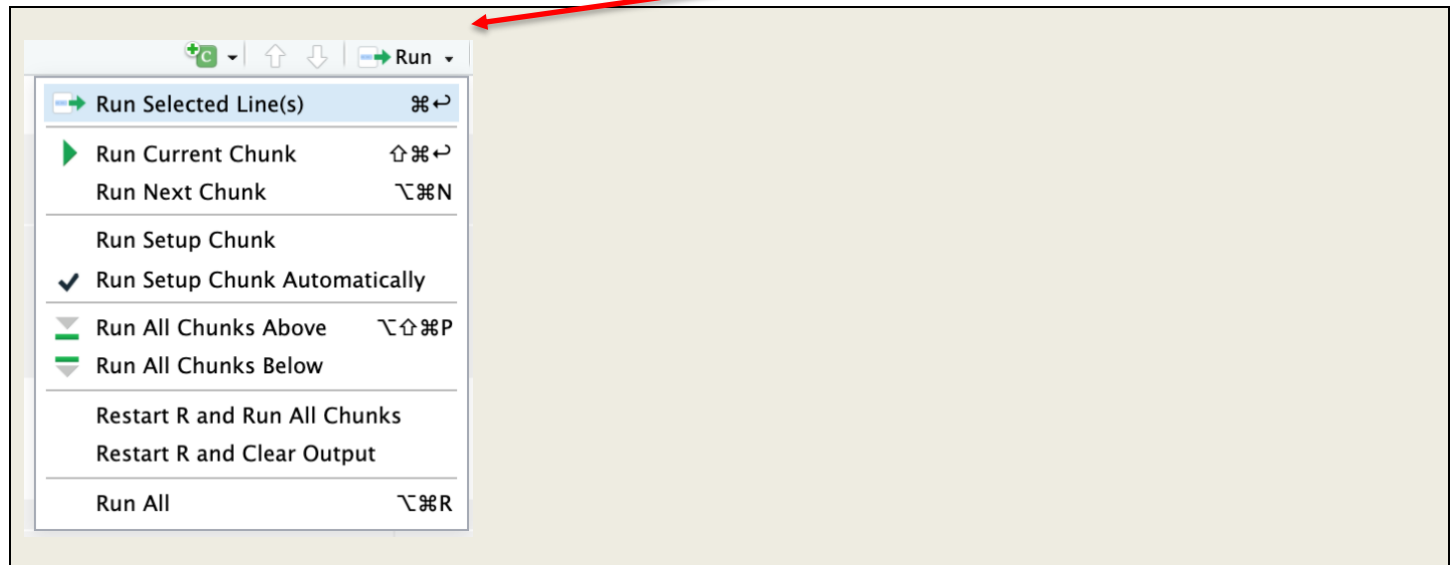
How to Execute a Chunk

There are 2 ways to execute the current chunk:

[Method #1.](#) Click on the little green arrow at the far right



[Method #2.](#) Alternatively, you can use the drop down options under **RUN** at upper right



All done? Render/Knit/Execute/Produce Your Report

IMPORTANT!!! Before you render, be sure to do a final save!

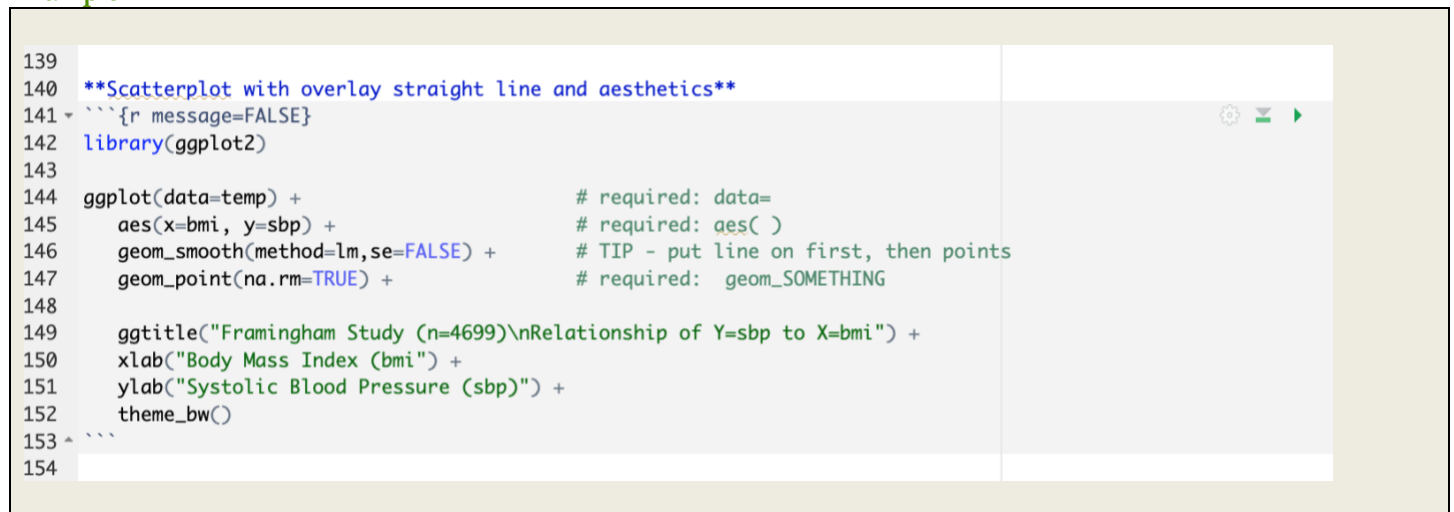
At upper left, click on the icon that looks like a ball of yarn with a knitting needle!



Tips!!

- **Work chunk by chunk.** Focus on each chunk one at a time. Create it, execute, examine the output, fix your mistakes, tweak to your liking, Think "wash, rinse repeat"!
- **Define each chunk to do ONE THING ONLY** (e.g., producing a plot, producing table, creating factors, and so on). This will make your program more readable and much easier to troubleshoot.
- **Preview the looks of your report often as you go along** (like tasting your pasta sauce as you make it).
How to do this: In the **Environment** pane, click on the tab **Viewer**. Now you can periodically knit to HTML; this will produce the output in your viewer for you to review.

Example



5. How to Work with R Markdown Markup Language

How to Markup Your Text Narrative (don't even think of MS Word!)

Tip - Preview your work as you go along! In R Studio, in the environment pane, click on the tab **Viewer**. As you go along you can check your work by knitting to HTML and looking at a preview in the viewer.

| Markup | Syntax |
|-------------------------|--|
| <i>italics</i> | <p><code>* text*</code> <code>_ text_</code></p> <p>Tip – There must be NO SPACE before your text begins</p> |
| bold | <p><code>** text**</code> <code>__ text__</code></p> |
| <i>italics and bold</i> | <p><code>***text***</code> <code>* __ text__ *</code></p> |
| _{subscript} | <p><code>~ text ~</code></p> |
| ^{superscript} | <p><code>^ text ^</code></p> |
| quote | <p><code><RETURN></code> <code>></code> <code>> body of quote</code> <code>></code></p> |
| hyperlink | <p><code>[text name of link] (URL)</code></p> |
| image | <p><code>![text name of image] (path to image)</code></p> |
| inline R code | <p><code>` r YOUR R COMMAND `</code></p> |
| | |

| Markup | Syntax <i>** My personal preference</i> |
|--|--|
| force a line break | ** Just add \ at the end of the line Or, put <u>two</u> trailing blank spaces. Then click RETURN |
| to add 2 blank lines after a sentence (and so on...) | ** Just add \ at the end of the line. Then do 2 more! \ |
| inline equation | $\$equation\$$ |
| equation on its own line, centered | $\$equation\$$ |
| How to produce headers of various sizes first level header (this will be the largest size) second level header (2nd largest size) third level header (... and so on ...) REQUIRED – There MUST be a space before your text begins | # text ## text ### text |
| NUMBERED list 1. etc is level 1 <tab> a. etc is level 2 <tab><tab> i. is level 3 Tip – This relies on use of the tab key | 1. titleoflist a. subitem 1 i. subsub item 1 ii. subsub item 2 b. subitem 2 i. subsub item 1 ii. subsub item 2 |
| UN-numbered list * is level 1 + is level 2 - is level 3 Tips – (1) Each *, +, - must be followed by a space (2) Use the tab key to get to next sublevel | * titleoflist + subitem 1 - subsub item 1 - subsub item 2 + subitem 2 - subsub item 1 - subsub item 2 |

6. Illustration: An R Markdown for Data Inspection

```
#### Introduction\
-----\
File: R Lesson 04 Markdown.Rmd\
Author: Carol\
BIOSTATS 640 Fall 2023 - Learn R Series\
Data used: hersdata1000.Rdata\
Lesson 4 - Introduction to R Markdown\
September 29, 2023\
-----\
\
```

Initialize session

```
``` {r}

setwd("/cloud/project") # Set working directory
getwd() # Check working directory
options(scipen=999) # Turn off scientific notation
rm(list = ls()) # Clear the Decks

```
```

Load data

```
``` {r}

load(file="hersdata1000.Rdata")

```
```

Quick and easy summary of distributions of every variable

```
``` {r}

summary(hersdata1000)

```
```

Inspection 1: Check Data Types

```
``` {r}

hersdata1000 <- as.data.frame(hersdata1000)
str(hersdata1000)

```
```

Preliminary for Later: Create Categorical Variables (factors). Check.

```
``` {r}

hersdata1000$racethf <- factor(hersdata1000$raceth) # convert to factor
hersdata1000$racethf <- factor(hersdata1000$raceth, # create value labels
 levels=c(1,2,3),
 labels=c("White", "African-American", "Other"))

hersdata1000$HTf <- factor(hersdata1000$HT)
hersdata1000$HTf <- factor(hersdata1000$HTf,
 levels=c(0,1),
 labels=c("Placebo", "Hormone Therapy"))

hersdata1000$exercisef <- factor(hersdata1000$exercise)
hersdata1000$exercisef <- factor(hersdata1000$exercisef,
 levels=c(0,1),
 labels=c("No", "Yes"))

hersdata1000$diabetesf <- factor(hersdata1000$diabetes)
hersdata1000$diabetesf <- factor(hersdata1000$diabetesf,
 levels=c(0,1),
 labels=c("No", "Yes"))

str(hersdata1000)

```
```

Inspection 3: Assess Extent of Missing Data and Complete Data

```
``` {r}

Key: Get count of missing values
is.na() produces a 0/1 with 1=TRUE if observation is missing
colSums() sums the 0/1 over all observations in that column
cat("\nNumber of missing values\nby Variable\n")
colSums(is.na(hersdata1000))

Key: Get count of complete observations
!is.na() produces a 0/1 with 1=TRUE if observation is complete
colSums() sums the 0/1 over all observations in that column
cat("\nNumber of complete observations\nby Variable\n")
colSums(!is.na(hersdata1000))

```
```

Inspection 4: Assess Ranges

```
``` {r}

library(stargazer)

hersdata1000 <- as.data.frame(hersdata1000) # peculiar to stargazer
stargazer(hersdata1000,
 type="text",
 summary.stat=c("n","min","max"),
 title="HERS Data (random sample of 1000)")

```
```

Inspection 5: Nice Inspection of Every Variable using function dfSummary() in package {summarytools}

```
``` {r}

library(summarytools) # Be sure you have installed this package first!

print(dfSummary(hersdata1000), method='render')
stview(dfSummary(hersdata1000))

```
```

7. Some Good Resources

VIDEOS

- 1. (*Source: Harold Nelson*)
Introduction to R Markdown 1 ([video, 5:34](#))
- 2. (*Source: Coursera*)
R Markdown Syntax ([video 13:11](#))

CHEAT SHEET

- 1. (*Source: <https://rmarkdown.rstudio.com/lesson-15.HTML>*)
R Markdown ([html](#))