

Creating reproducible reports using R Markdown

C. Tobin Magle
Cyberinfrastructure facilitator
Colorado State University

Outline

- What is literate programming?
- Why is it useful?
- How to use R Markdown to create reproducible reports
 - Text formatting
 - Code chunks
 - Output formats

Literate programming

human readable **text**

+

machine readable **code**

=

Reproducible document

Programs as works of literature

- Developed by Donald Knuth, Stanford University
- **Traditional:** telling a computer what to do
- **Novel:** telling a human what you want the computer to do
- Improves documentation and the programs themselves

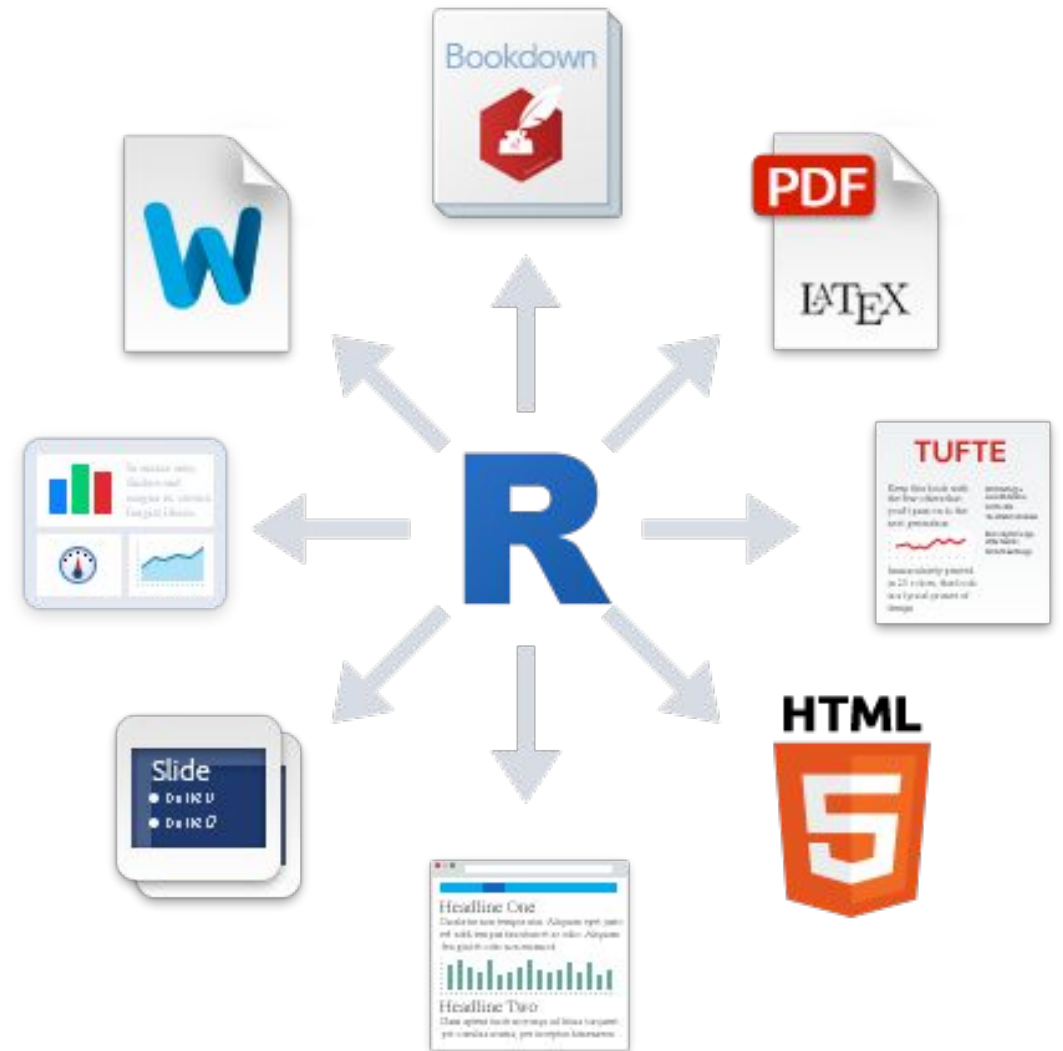
Literate programming in research

- Tailor reports to an audience
- Repeatable and assures reproducibility
- Works well with version control
- Works well with languages used in research:



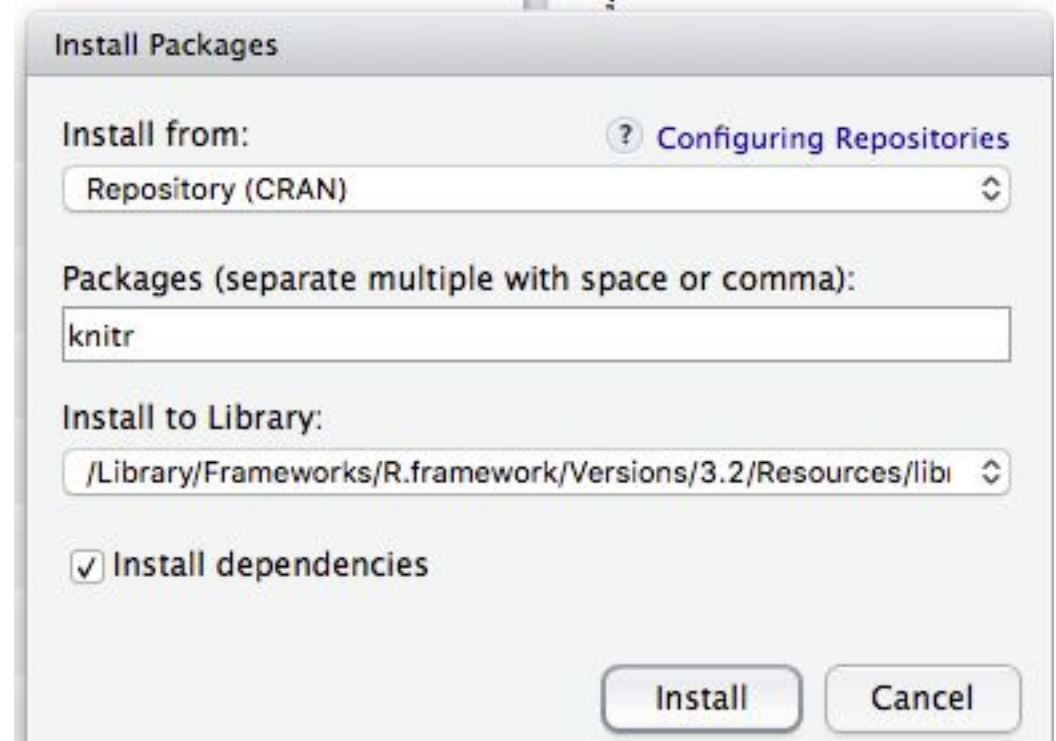
R Markdown

- Weave narrative text and code
 - R
 - Python
 - Bash
 - SQL
- Produce documents in many formats
- Reproducible

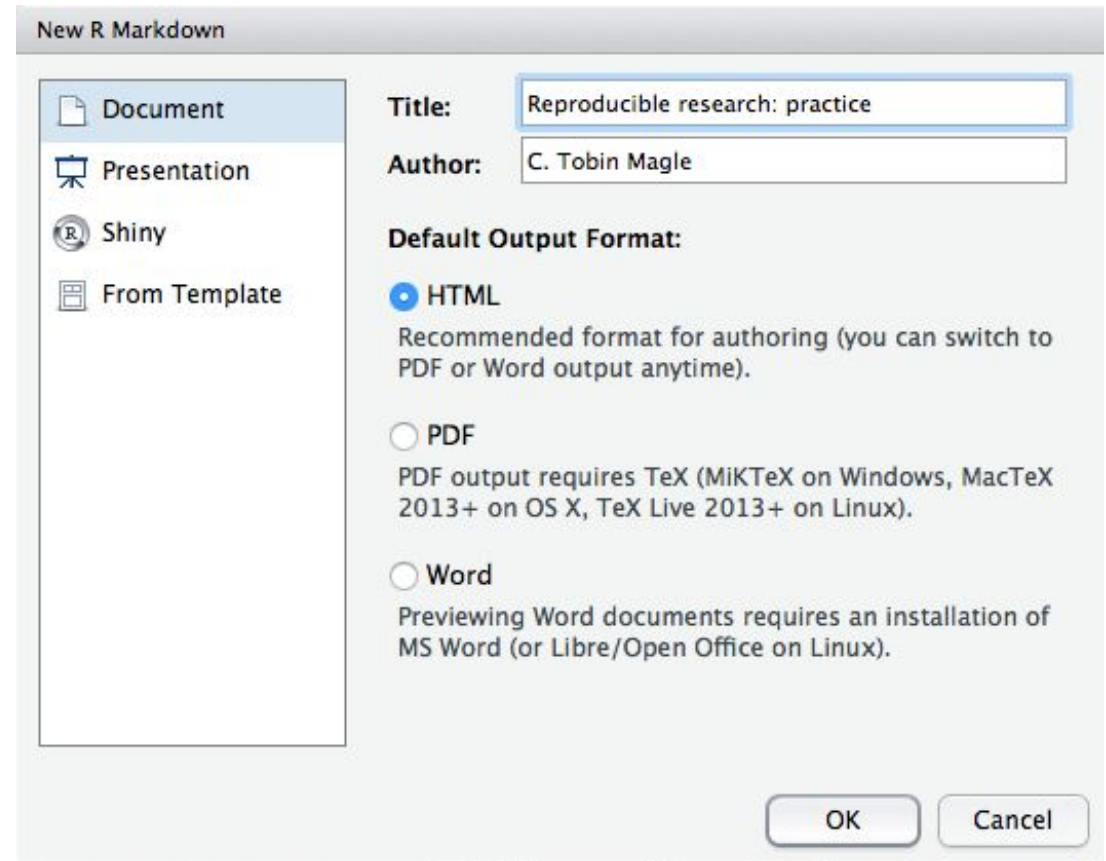
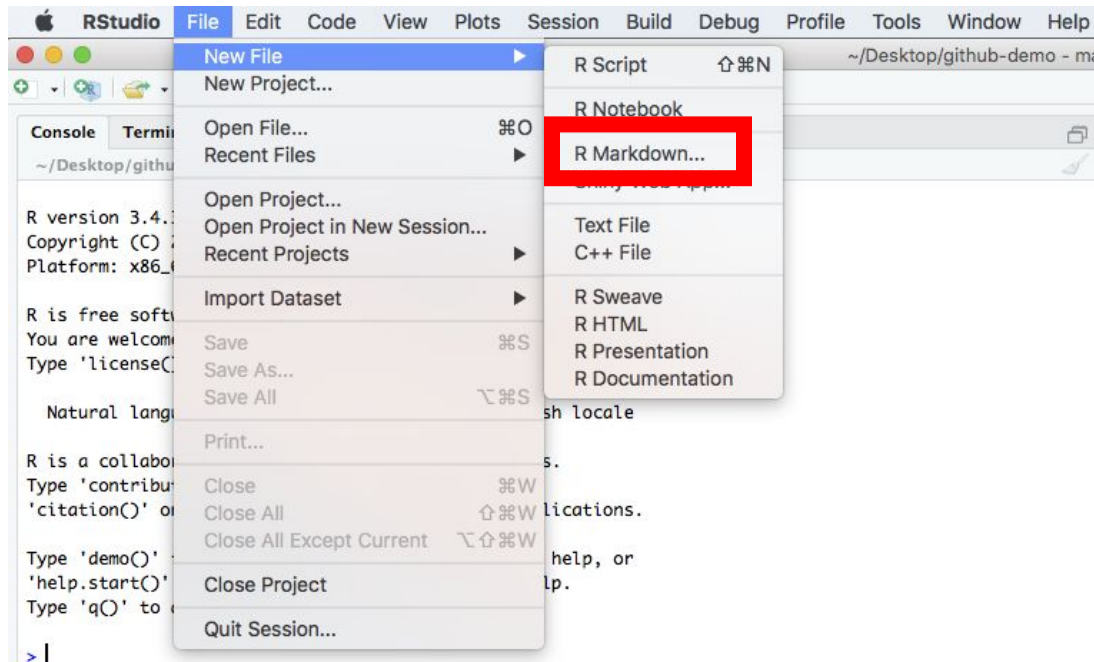


Setup

- Tools > install packages OR `install.packages("rmarkdown")`
- Install TeX if you want create PDFs
- Open R project:



Open/Create a markdown document



Sections of R Markdown

Header

```
---  
title: "Markdown Demo"  
author: "Tobin Magle"  
date: "4/9/2018"  
output: html_document  
---
```

Code chunk

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

Human-readable
text

```
## 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>.

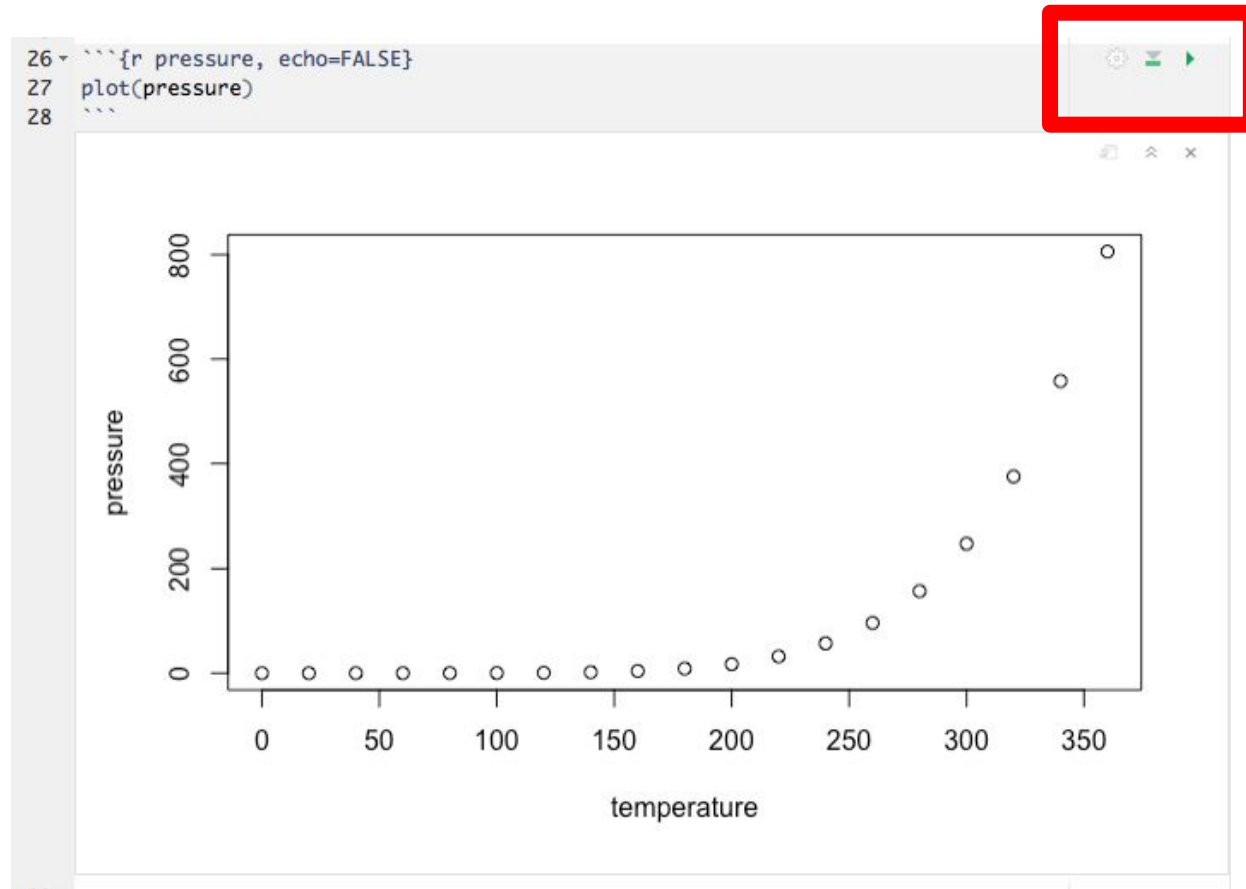
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

Code Block

```
```{r cars}  
summary(cars)
```
```

Notebook interface

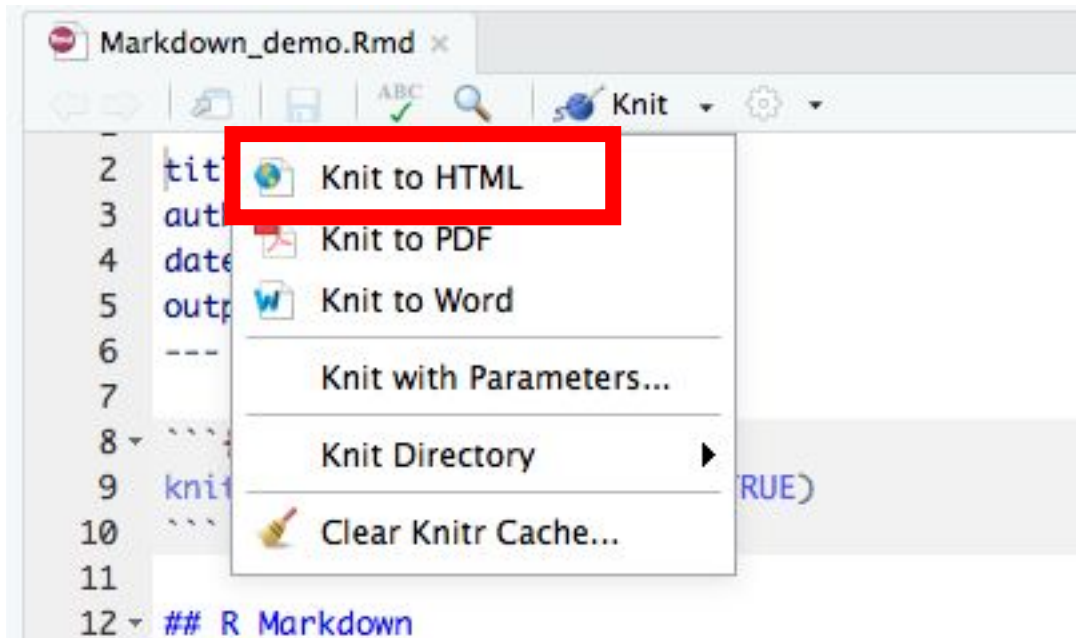
- Press green triangle on a code chunk to run the code and show the output



Knit: from text to document

render(input = "name.Rmd",
output = "html_document")

OR



Markdown Demo

Tobin Magle

4/10/2018

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:



How it works

- knitr package converts the R Markdown document to markdown
- Pandoc converts the markdown document to the final



Exercise 1: render the document

- Knit the document
- Delete everything but the header
- knit again
- What changed? Did you run into any trouble?

Human readable text

- Outline first, code later
- Use markdown rich formatting syntax to
- Cheat sheet:
<https://www.rstudio.com/wp-content/uploads/2015/02/rmarkdown-cheatsheet.pdf>
- # R Markdown – header
 - More # decreases size
- ****knit**** - **bold**
- < > - create hyperlink

<https://rmarkdown.rstudio.com/lesson-8.html>

Exercise 2: write an outline

- Look at the document on the right
- Write an outline of how you'd produce the document
- Use formatting!

Markdown Demo

Tobin Magle

4/10/2018

Small animal survey data

This is a survey of small animals from 24 plots from the years 1977 to 2002. Each row is an animal. Each column is an attribute of that animal. The variables recorded are record_id, month, day, year, plot_id, species_id, sex, hindfoot_length, weight, genus, species, taxa, plot_type.

Sex table

Here is a breakdown of the number of female vs male animals observed in the study

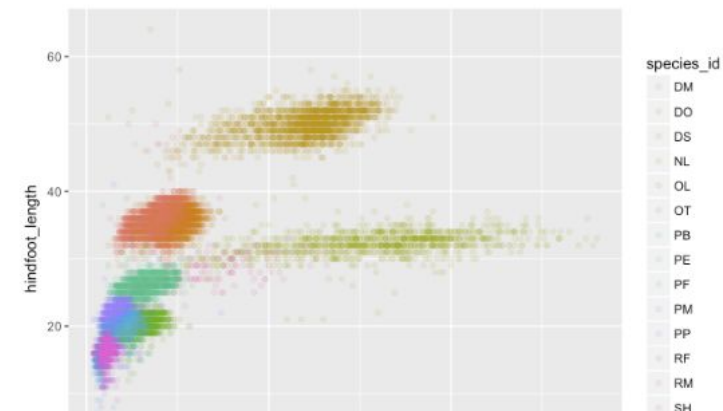
| sex | n |
|-----|-------|
| F | 14466 |
| M | 15997 |

Mean weight

The mean weight of all animals surveyed is 41.8568756.

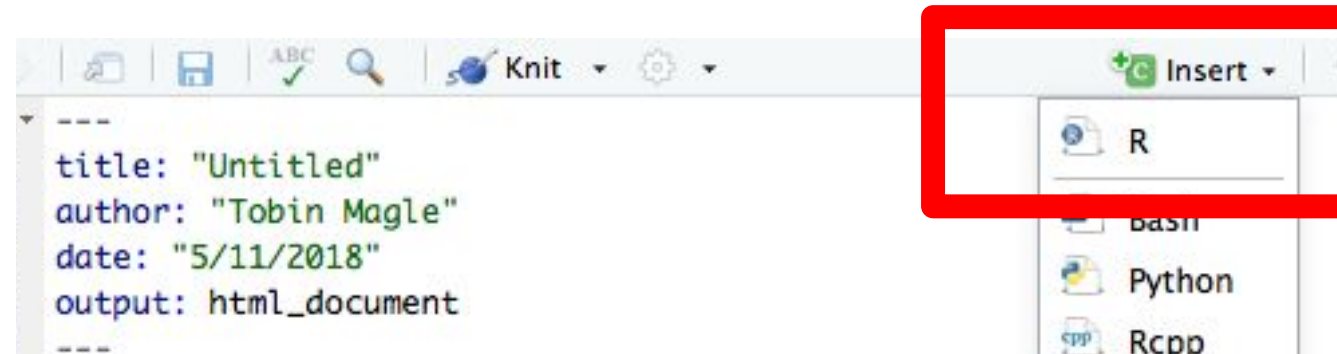
Relationship between hindfoot length and weight

We are interested in the relationship between weight and hindfoot length by species. Here is a graph.



Adding code chunks

- Chunk delimiters
 - beginning: ````${r name}`
`<code>`
 - end: `````
- Shortkey: `Ctrl+Alt+i`
- Editor toolbar



Code chunk options

- **include** = FALSE
 - Hide code and results from document
 - results can still be used later
- **echo** = FALSE
 - hides code, keeps result
- **message** = FALSE - hides messages
- **warning** = FALSE - hides warnings
- **fig.cap** = “ ” - adds figure captions

Exercise 3: code chunk

- insert a code block called graph

```
ggplot(data = surveys_complete,  
       aes(x = weight,  
           y = hindfoot_length,  
           color = species_id)) +
```

- Add the ggplot code

```
geom_point(alpha = 0.1,  
           aes(color = species_id))
```

- Render

- Use options to hide the code

Setting options for all code chunks

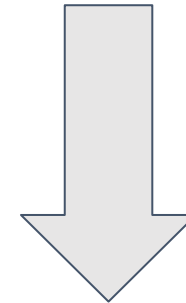
- Set default options for all code chunks
- `knitr::opts_chunk$set()` function
- Put it in your first codeblock
- Must be after you load the rmarkdown package

Tables

- Tables look like console output by default
- `kable()` function makes it pretty

| sex
<fctr> | n
<int> |
|----------------------|-------------------|
| F | 14466 |
| M | 15997 |

2 rows



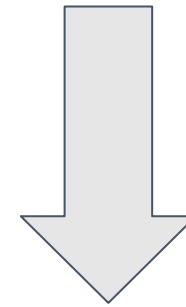
kable

| sex | n |
|------------|----------|
| F | 14466 |
| M | 15997 |

Inline code

- Syntax ``r``
- Include in test for responsive text
- Allows you to write one report and run it over changing datasets
- Example: mean

```
The mean weight of all animals surveyed  
is `r mean(surveys_complete$weight, na.rm=TRUE)`.
```



Render

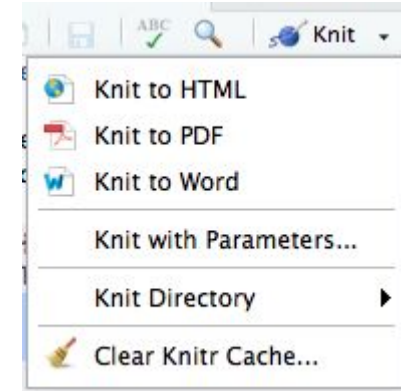
```
The mean weight of all animals surveyed is 41.8568756.
```

Exercise: inline code

- Using inline code, re create the first text box from the example document
- Hints
 - Code for # plots: `nlevels(as.factor(surveys_complete$plot_id))`
 - Code for earliest year: `min(surveys_complete$year)`
 - Code for latest year: `max(surveys_complete$year)`
 - Code for column names: `names(surveys_complete)`

Change output formats

- select output from knit menu
- Specify output in header
- with render



```
---  
title: "Markdown Demo"  
author: "Tobin Magle"  
date: "4/10/2018"  
output: html_document  
---
```

```
library(rmarkdown)  
render("1-example.Rmd", output_format = "word_document")
```

<https://rmarkdown.rstudio.com/lesson-9.html>

Parameters

- Creates custom report without having to change the code chunks
- Add to the header
- Creates a read only list of parameters

```
1 ---
2 title: "Markdown Demo"
3 author: "Tobin Magle"
4 date: "4/10/2018"
5 output:
6   html_document: default
7 params:
8   file: "data/surveys_complete.csv"
9 ---
10
11 ```{r setup, include=FALSE}
12 library(ggplot2)
13 library(dplyr)
14
15 surveys_complete<-read.csv(params$file)
16
17 ```
```


Subset with parameters

- Change the header to add a sex parameter
- Subset by sex using a filter statement in the setup code chunk
- Render

title: "RMD-test"

author: "Tobin Magle"

date: "5/16/2018"

output: html_document

params:

file: data/surveys_complete.csv

sex: 'M'

AND

```
surveys_complete<-surveys_complete%>%
```

```
  filter(sex == params$sex)
```

Parameters in the render function

- params argument

```
render(input = "Markdown_demo.Rmd",  
       output_format = "html_document",  
       params = list( file = "data/surveys_complete.csv",  
                      sex = "M"))
```

- takes a list

- overrides default in the RMD

Parameter user interface

- params = “ask”
- Opens up a web browser
- Select options
- Hit save
- Doc gets rendered

File (default: data/surveys_complete.csv)

Browse...

No file selected

Sex:

M

Cancel

Save

Types of parameter input

| Input Type | Shiny Function |
|------------|----------------------------|
| checkbox | <code>checkboxInput</code> |
| numeric | <code>numericInput</code> |
| slider | <code>sliderInput</code> |
| date | <code>dateInput</code> |
| text | <code>textInput</code> |
| file | <code>fileInput</code> |
| radio | <code>radioButtons</code> |
| select | <code>selectInput</code> |
| password | <code>passwordInput</code> |

Exercise: parameters

- Create a new parameter called weight
- Then create a way to select a range of weight and only report on data in that weight range
- Hint: the slider input has min and max values instead of “choices”

```
---  
title: "My Document"  
output: html_document  
params:  
  minimum:  
    label: "Minimum:"  
    value: 100  
    input: slider  
    min: 0  
    max: 1000  
  region:  
    label: "Region:"  
    value: east  
    input: select  
    choices: [east, west, north, south]  
  data:  
    label: "Input dataset:"  
    value: results.csv  
    input: file  
---
```

Example from:

https://rmarkdown.rstudio.com/developer_parameterized_reports.html#accessing_from_r

Running multiple reports

- **Goal:** run same analysis on many different .csv
- Use a script
 - get list of files
 - loop over files
 - render() for each file
- **Output:** an html doc for each csv

```
library(rmarkdown)
```

```
files<-list.files(path = "data",  
                  pattern=".csv")
```

```
for (file in files){  
  render(input = "Markdown_demo.Rmd",  
         output_format = "html_document",  
         output_file = paste(file, ".html",  
                              sep = ""))  
}
```

Summary

- Literate programming makes reproducible research more machine readable
- R markdown documents facilitate literate programming in RStudio
- R markdown has 3 sections
 - Header: determines output and adds parameters
 - Markdown Text: can include inline code
 - Code chunks: can be customized to mute code or output
- Parameters help you customize reports

Need help?

- Email: tobin.magle@colostate.edu
- Data Management Services website:
<http://lib.colostate.edu/services/data-management>
- R markdown cheatsheet:
<https://www.rstudio.com/wp-content/uploads/2015/02/rmarkdown-cheatsheet.pdf>
- RStudio markdown lesson
<https://rmarkdown.rstudio.com/lesson-1.html>