reproducible research

A. Michelle Lawing
Ecosystem Science and Management
Texas A&M University
College Station, TX 77843
alawing@tamu.edu
Announcements

• May 5\textsuperscript{th} is redefined day (go to your Friday classes), so no presentations

• Final Project (2 options) – 10% of final grade
  – Turn in a video of your final project (around three minutes, these will be posted on my webpage)
    (e.g., http://www.kindealabs.com/)
  – Turn in a manuscript for critical review and feedback.
Schedule for today

• My presentation
  – on reproducible research

• Presentation on end to end reproducibility
• Demonstration on end to end reproducibility

• ?BREAK
• ?Tutorial on end to end reproducibility
reproducible research
Cornerstones of the scientific method

- Reproducibility
  Does not make the science right or wrong
- Rigor
  To evaluate reproducibility
- Transparency
  To evaluate reproducibility
- Independent verification
  Allows science to move forward
Getting help

• Post online questions

• Email code to another R user

• Need to be able to get reproducible errors

• This allows someone else to copy and paste your code without changing anything
Four items for Code Share

• R environment

• Required packages

• Data

• Code
R environment

- sessionInfo() #summarizes your R environment

- makes it easy to detect out of date packages

- usually I make a folder for each “project” and open the R script from within that folder – automatically takes me to the correct environment
Packages

- Should be loaded at the top of the script
- Easy to see which ones are needed
- I usually include and comment out the `install.packages()` command to make it easy for installing packages.
Data

• You can use R code to recreate data to output

• `dput()` outputs an ASCII text representation of your data

• `save()` writes an external representation of an R object

• `write()`, `write.table()`, `write.csv`
Code

• write code that is easy to read by others
• use spaces
• make variable names concise, but informative
• use comments
• remove everything that is not related
• the shorter your code, the easier it is to understand
Check for yourself

• Start up a fresh R session
• Paste in your script
Post it

- https://gist.github.com/

- **Gist** is a simple way to **share snippets and pastes** with others. All gists are Git repositories, so they are automatically **versioned, forkable** and usable from Git.
Reproducible Research Task View on CRAN

- [http://cran.r-project.org/web/views/ReproducibleResearch.html](http://cran.r-project.org/web/views/ReproducibleResearch.html)
knitr() and Markdown

- special packages for reproducibility and document creation
- combines many other packages and their functionality
- takes many input languages (e.g., R and Python)
- output formats are optional (LaTeX, HTML, Markdown, ASCII)
Let’s Try

• In Rstudio, go to File > New File > R Markdown...

• Read in Title and Author, click OK

• You will see a new document with .Rmd extension

• Save this untitled document
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:

```
r
summary(cars)
```

You can also embed plots, for example:
R Markdown

You can also embed plots, for example:

```{r, echo=FALSE}
plot(cars)
```

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

• Your example markdown document should open as an html file.
Let’s do it again

• michellelawing.info/Rmarkdown.Rmd

• michellelawing.info/Rmarkdown.html