Introducing RStudio and GitHub

Dr. Maria Tackett

08.28.19

Click for PDF of slides

Reproducible data analysis

Reproducibility checklist

What does it mean for a data analysis to be "reproducible"?

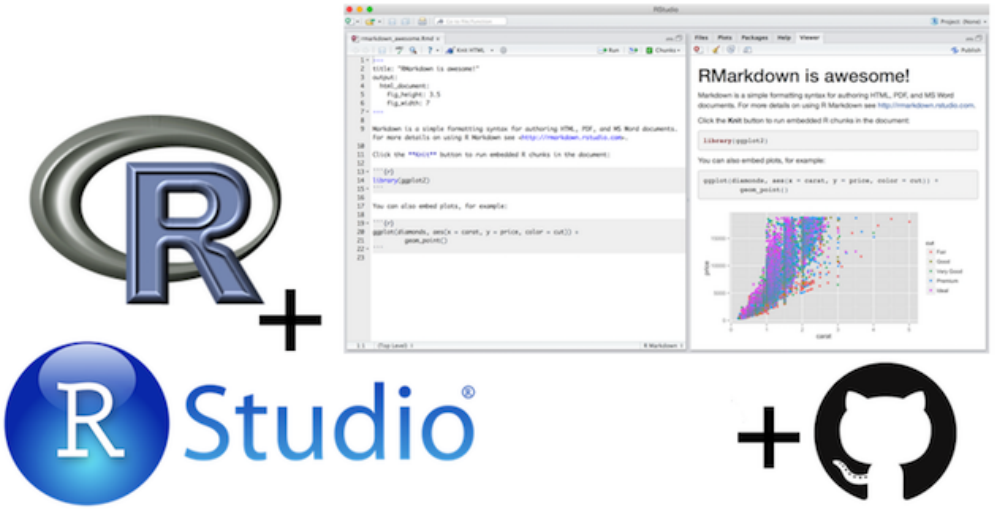
Near-term goals:

- Are the tables and figures reproducible from the code and data?
- Does the code actually do what you think it does?
- In addition to what was done, is it clear **why** it was done? (e.g., how were parameter settings chosen?)

Long-term goals:

- Can the code be used for other data?
- Can you extend the code to do other things?

Toolkit



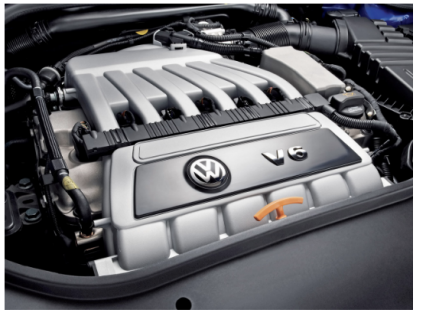
- Scriptability → R
- Literate programming (code, narrative, output in one place) → R Markdown
- Version control → Git / GitHub

R and RStudio

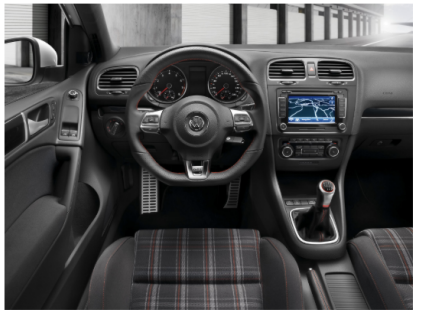
What is R/RStudio?

- R is a statistical programming language
- RStudio is a convenient interface for R (an integrated development environment, IDE)
- At its simplest:*
  - R is like a car's engine
  - RStudio is like a car's dashboard

R: Engine



RStudio: Dashboard



*Source: Modern Dive

R essentials (a short list)

- **Functions** are (most often) verbs, followed by what they will be applied to in parentheses:

```
do_this(to_this)
do_that(to_this, to_that, with_those)
```

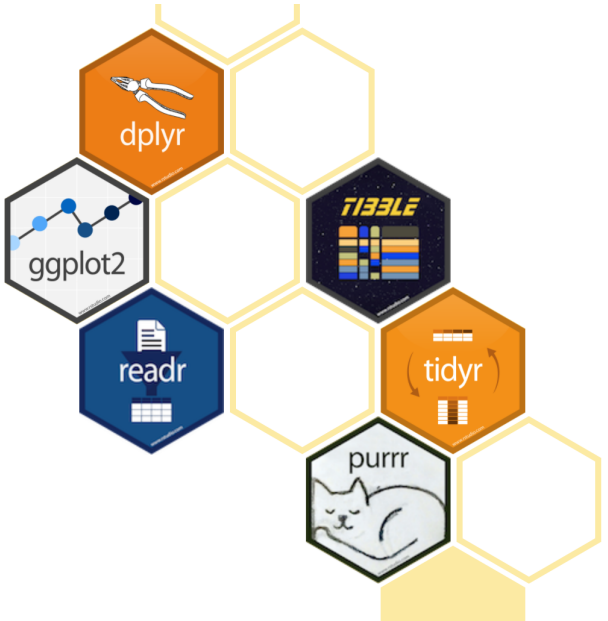
- **Columns** (variables) in data frames are accessed with `$`:

```
dataframe$var_name
```

- **Packages** are installed with the `install.packages` function and loaded with the `library` function, once per session:

```
install.packages("package_name")
library(package_name)
```

tidyverse



tidyverse.org

- The tidyverse is an opinionated collection of R packages designed for data science.
- All packages share an underlying philosophy and a common grammar.

R Markdown

R Markdown

- Fully reproducible reports -- the analysis is run from the beginning each time you knit
- Simple Markdown syntax for text
- Code goes in chunks, defined by three backticks, narrative goes outside of chunks

R Markdown tips

Resources

- R Markdown cheat sheet
- Markdown Quick Reference:
  - `Help -> Markdown Quick Reference`

Remember: The workspace of the R Markdown document is separate from the console

Workspace vs. console

- Run the following in the console.

```
x <- 2
x * 3
```

- Then, open a new R Markdown document. Add the following chunk in your R Markdown document and knit.

```
x * 3
```

What happens? Why the error?

How will we use R Markdown?

- Every assignment / lab / project / etc. is an R Markdown document
- You'll always have a template R Markdown document to start with
- The amount of scaffolding in the template will decrease over the semester