

From Data to Insight

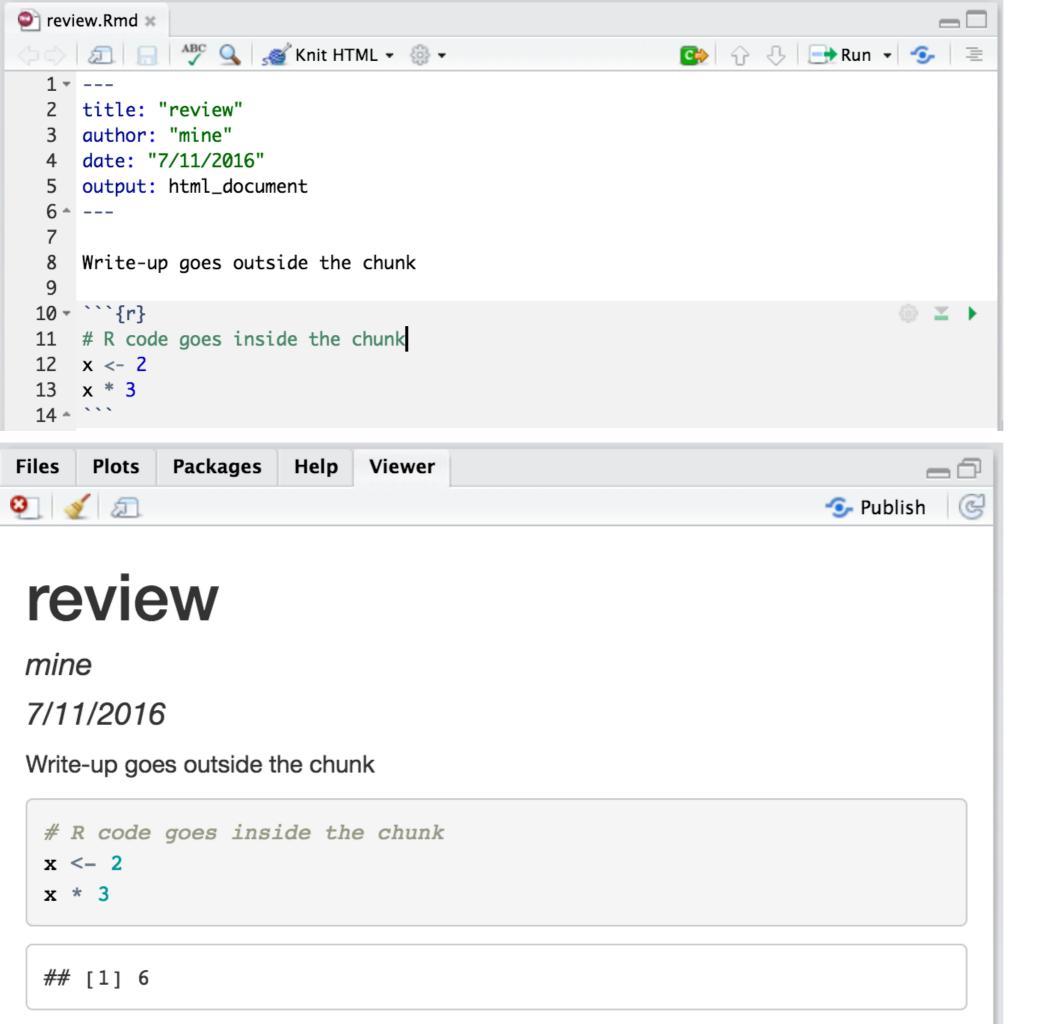
Dr. Çetinkaya-Rundel July 11, 2016

Terminology

- R: statistical programming language
- RStudio: front-end software for R that allows you to organize your files and plot, keeps a history of your command, and provides an environment for creating reports with R Markdown
 - It's much more than that, but for our purposes, this should be a sufficient definition
- R Markdown: Authoring format for dynamic documents including your R code and your write-up

R Markdown

- R code goes in chunks, marked by three backticks and the letter r in curly braces to begin and three backticks to end
 - Within a chunk # is used to mark a comment, any text following this sign on the same line will not get processed as code.
- Interpretations, i.e. your write-up "in English", goes outside of R chunks



← input

← output

Independent environments

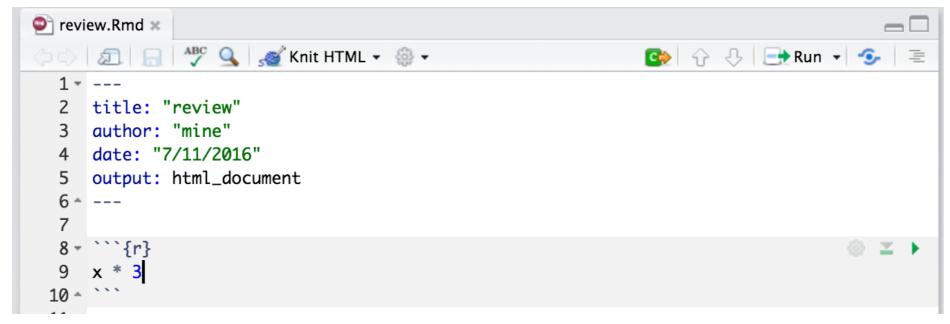
- Your Console uses one working environment
- Your R Markdown document uses a different (independent) working environment
- If you define an object in your Console, but do not define it in your R Markdown document, you will get an error when you try to knit your document saying that the object is not found

```
Console R Markdown *

-/ 

Loading required package: grDevices
Loading required package: stats
Loading required package: utils
Please visit openintro.org for free statistics materials
> x <- 2
> x

[1] 2
> |
```





Deciphering errors

- This is a skill you'll develop over time, so do not get discouraged if initially the errors seem too cryptic
- Approach deciphering what the error is saying methodically you don't need to understand everything printed in the error to figure out what the issue is
 - First see which line of code is causing the error, noting that the error will point you to the first line of the R chunk
 - Go to that chunk to see if you can figure out what the issue is (maybe spelling error?)
 - Read the error further to see if there are other clues like "object not found" or "could not find function" etc.

Common erros in code

- Spelling!
 - Spelling of objects you create as well as spelling of functions
- Non-matching parantheses and quotation marks

ggplot2(+) dplyr(%>%)

- ggplot2: Package we are using for plotting
 - Plots are comprised of layers
 - Layers are separated by +
 - Stylistic requirement: End lines of ggplot2 code with
 - +, move to the next line for the next layer
- dplyr: Package we are using for data wrangling
 - Pipes is comprised of chains
 - Lines of chains are separated by %>%
 - You read a pipe as take the output of the preceding line and use it as the first argument of the next line
 - Stylistic requirement: End lines of dplyr code with %>%, move to the next line for the next step in the chain