

Documenting your analyses in R

Mini R-Workshop

Monday, Sept 28, 2015

Why is good documentation useful?

- Replicate your analyses/results
(reproducible research!)
- Share with collaborators
(more reproducible research!)
- Maintain your own sanity, because you like to make your future self much happier

Writing reproducible reports in R with markdown, knitr and pandoc

So you have some code, data and a cool result, now it's time to communicate this with your collaborators (or supervisor). What do you do? In this guide, we want to show you how to write nice, reproducible reports using some of the fantastic, free tools and packages that are now on offer. These tools will help you communicate your science, and hopefully mean that you never copy and paste your R output again.

As a start, let's review the key components to any good analysis:

1. Data
2. Code used to analyse the data
3. Figures and tables generated by the code
4. Text, interpreting the figures and results, and describing the methods.

These elements come together in the form of a report. As scientists, we write many reports, both small and large. Large reports like papers, are rare, but we write smaller reports all the time. These include all the preliminary results, weekly updates, emails with figures, and simply one's own note taking, written during the lifespan of a project. Traditionally, most biologists do stages 2 and 3 in R, then fire up Word or Powerpoint and copy-paste everything for stage 4. That works, but there several downsides to this approach:

- lots of time wasted, plus copy and pasting sucks
- your interpretation is separated from your code

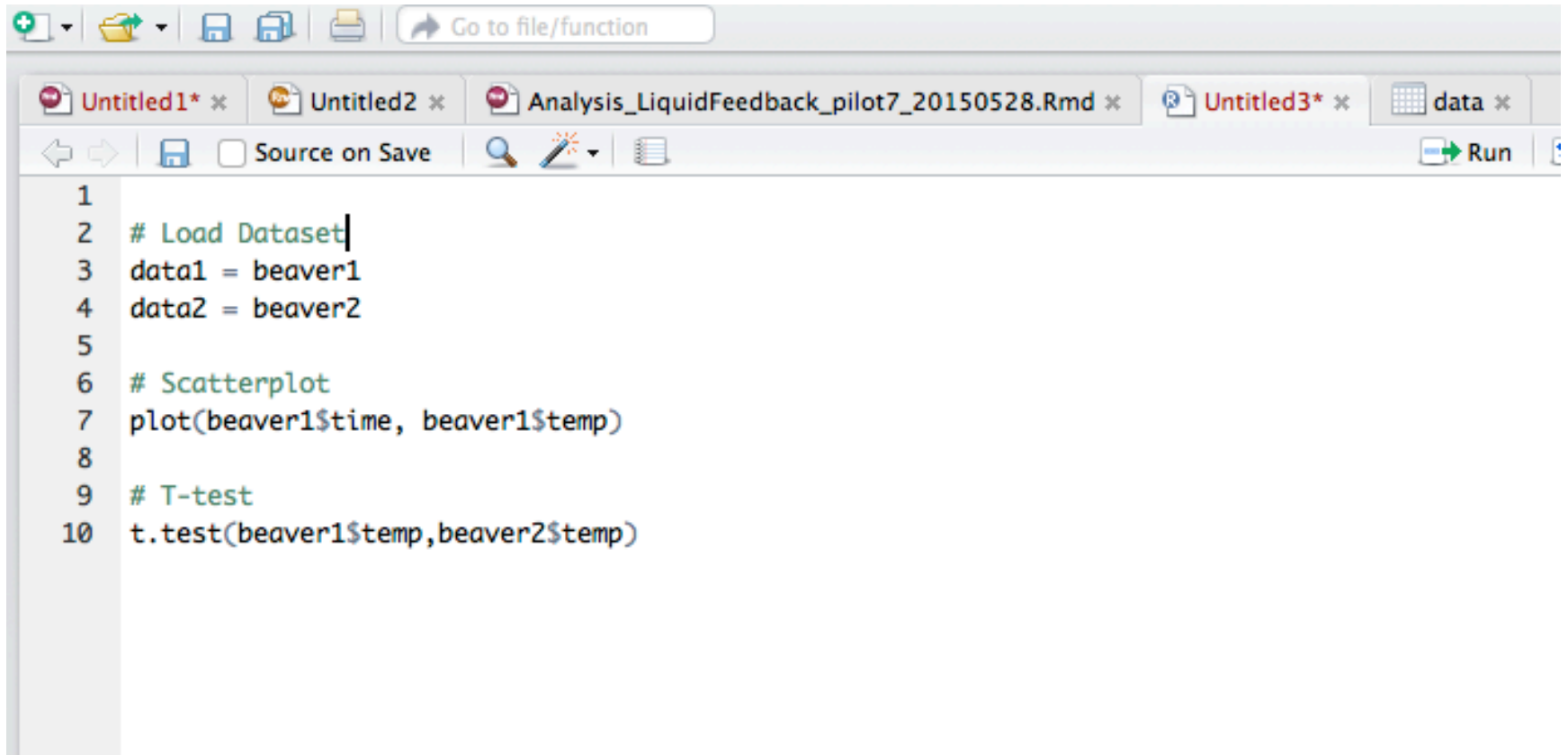
Overview

- Documenting your Code
- Dynamic Reporting
 - R Markdown
 - R Sweave
 - Knitr/Latex

Documenting your code

- You should ALWAYS comment your code. (# = pound key, or “hashtag”)
- People vary in degree of detail, but imagine you were giving your code to a stranger who you will never meet. In theory, they should be able to understand your analyses without you ever talking to them (mostly).

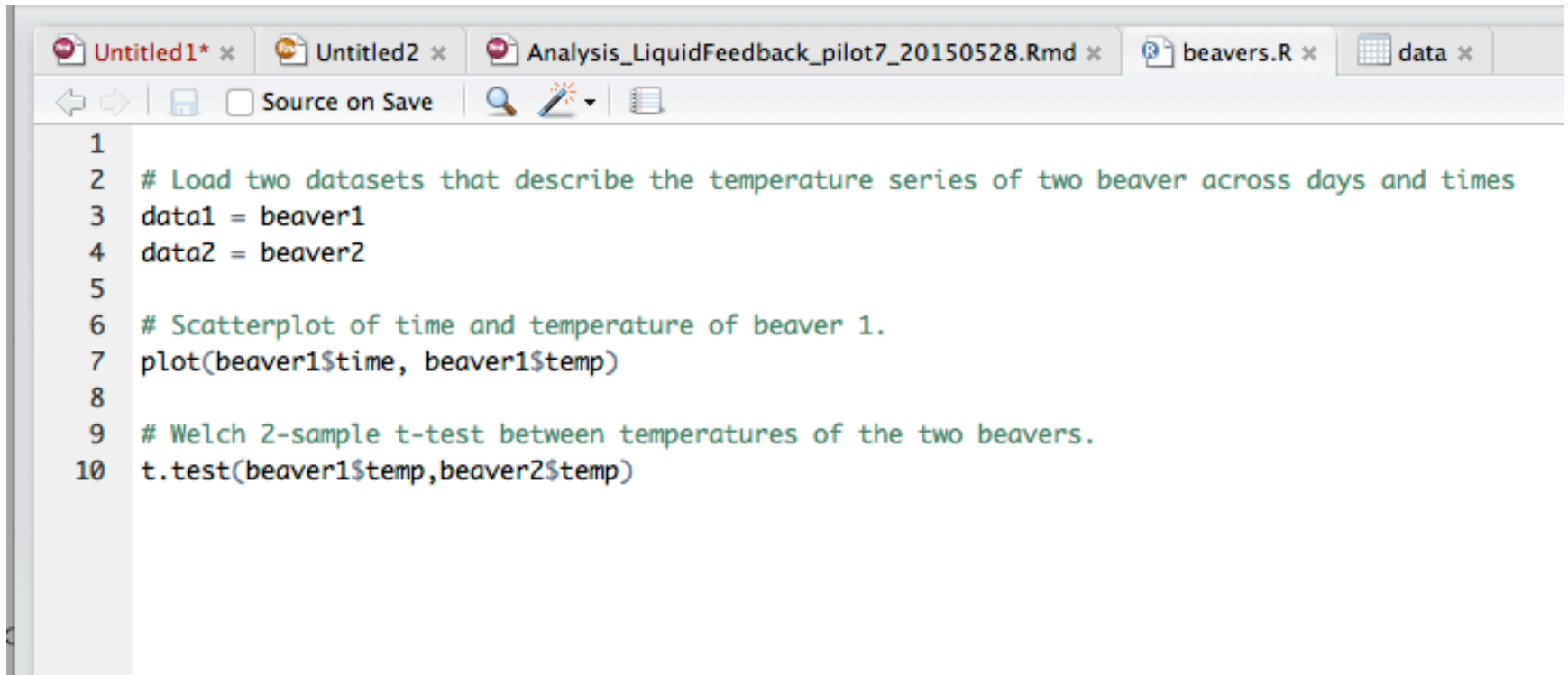
Bad documentation



The image shows a screenshot of the RStudio interface. The top toolbar includes icons for file operations and a search bar labeled "Go to file/function". The tab bar shows several open files: "Untitled1*", "Untitled2*", "Analysis_LiquidFeedback_pilot7_20150528.Rmd", "Untitled3*", and "data". Below the toolbar, there are navigation arrows, a "Source on Save" checkbox, and a "Run" button. The main editor area contains the following R code:

```
1  
2 # Load Dataset  
3 data1 = beaver1  
4 data2 = beaver2  
5  
6 # Scatterplot  
7 plot(beaver1$time, beaver1$temp)  
8  
9 # T-test  
10 t.test(beaver1$temp, beaver2$temp)
```

Good(ish) documentation



```
1  
2 # Load two datasets that describe the temperature series of two beaver across days and times  
3 data1 = beaver1  
4 data2 = beaver2  
5  
6 # Scatterplot of time and temperature of beaver 1.  
7 plot(beaver1$time, beaver1$temp)  
8  
9 # Welch 2-sample t-test between temperatures of the two beavers.  
10 t.test(beaver1$temp, beaver2$temp)
```

Dynamic Reporting

- What is it?
- Wait, I can write AND code in the same document? No way!
- And wait, this creates a PDF report for me to send to my PI / collaborators?

R Markdown

- <http://rmarkdown.rstudio.com/>
- Uses syntax like plain text email

R Sweave / knitr

- <https://support.rstudio.com/hc/en-us/articles/200552056-Using-Sweave-and-knitr>
- Requires LaTeX
- Also, requires knowledge of how to use LaTeX

Learning good habits takes time...

- R is flexible (many options)
- **Bottom Line:** learning how to document well in R is time consuming, but totally worth it