GETTING STARTED - COURSE INFRASTRUCTURE

Introduction to data science / DSC 105 / Fall 2024

Marcus Birkenkrahe

August 15, 2024

Literate programming in Emacs

- We're going to make excessive use of the Emacs editor and the Org-mode package for literate programming.
- In this session, you're going to get in touch with your inner penguin, install Emacs, R and write your first R code.
- To get started, you don't need to know much, but to get good at using the infrastructure, you need to complete the Emacs tutorial, which you can find at tinyurl.com/emacs-tutor (with videos).

Using Emacs + R in Google Cloud Shell

1. Go to cloud.google.com/shell/ and register with your Lyon Google account. This is completely free for up to 50 hours per week.
2. Open a terminal (aka command-line, aka shell, aka Console).
3. Install Emacs with the following command:

   ```
   sudo apt install emacs
   ```

4. Get the `.emacs` configuration file with the following command:

   ```
   wget -O .emacs http://tinyurl.com/lyon-emacs
   ```

5. To check, start a terminal Emacs with the command:

```
emacs -nw
```

6. The configuration file should be read and used. You may get some errors. The first thing to do for using R with Emacs is to install the ESS ("Emacs Speaks Statistics") package:

   - Inside Emacs, type `ALT-x list-packages RET` ("RET" means 'Enter')
   - In the buffer that should open, type `C-s ess RET`.
   - With the cursor on the line of the ESS package, type `i` (install)
   - Type `x` (execute), and ESS should be installed.
   - Shut Emacs down with the command `CTRL-x CTRL-c`

7. Install R from the command line with the command

   ```
   sudo apt install r-base
   ```

8. You're now ready to run R inside Emacs: you have the editor (Emacs), the programming language (R), and the interface (ESS): start Emacs again with `emacs -nw` from the terminal. Then enter `CTRL-x CTRL-f testR.org RET`

9. An empty buffer named `testR.org` opens. Enter the following at the top of the file:

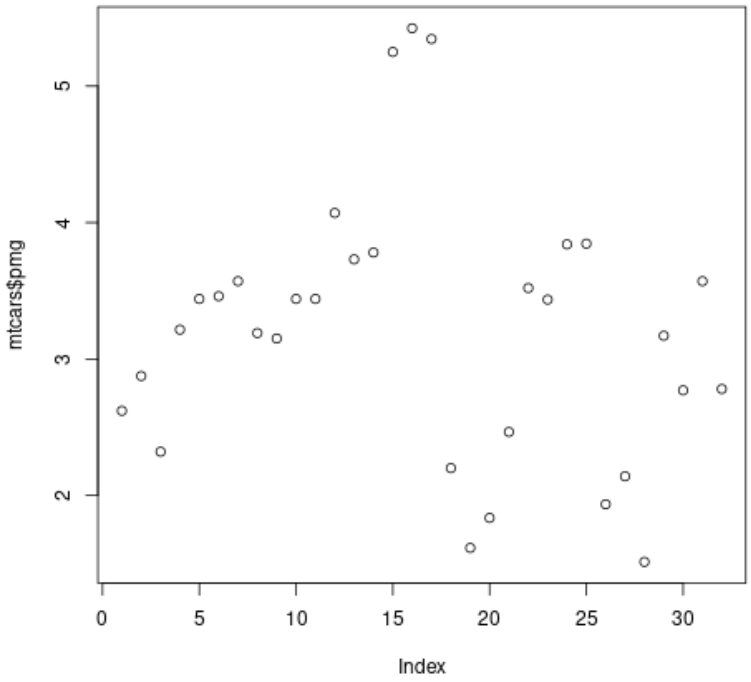
   ```
   * Testing R

     Print the structure of the =mtcars= data frame.

     #+begin_src R
       str(mtcars)
     #+end_src
   ```

   Make a scatterplot of miles-per-gallon vs. weight and store the plot in a file `mtcars.png`.

   ```
   plot(x=mtcars$wt,y=mtcars$pmg)
   ```



10. If everything worked out, close Emacs with `CTRL-x CTRL-c`. You can shut the application down by shutting the browser window down. In your cloud shell, Emacs will remain installed, but R unfortunately will not and you'll have to install it every time you go in.

Summary

1. Literate programming combines documentation, source code, and output in one file. You can tangle the source code and weave the documentation.

2. Google Cloud Shell offers a virtual Linux shell experience. "Shell" is short for a program that accepts Linux commands such as `sudo apt install`, which uses the Ubuntu Linux package manager `apt` to `install` something with super-user (`sudo`) rights.

3. The `wget` program allows you to download files using the Web URL.

4. You can use the Emacs editor in GUI or in terminal (`-nw` or "no windows") mode. In the cloud, only the terminal Emacs will work.

5. To run R programs inside Emacs, you need the Emacs editor, the R programming language, and the ESS package (interface to R).

6. Alternatively, you can also run R commands from the R console, which you open on the command line with the command `R`, and close with `q()`. For example, to plot 100 random numbers:

   ```
   png("myFile.png")
   plot(rnorm(100))
   dev.off()
   ```

7. An Org-mode file is a literate program. Inside Org-mode, code blocks begin with `#+begin_src` and end with `#+end_src`. To run, a code block has to have a language (e.g. R), and the Emacs configuration file has to have been loaded.

8. In Org-mode files, you can run code with `C-c C-c`. In the cloud, you need to leave Emacs to see graphical printout. The header of the code block contains information for Emacs to print graphics to file.

9. You learnt the following R commands already:

   ```
   str
   plot
   ```

10. We will repeat all of this in a systematic fashion - this was only a teaser-trailer to show you (almost) all of the infrastructure. Turns out you can even connect to GitHub from Google Cloud. Alternatives to Emacs + Org-mode are: Google Colaboratory, DataCamp's DataLab, or Kaggle.com (also by Google).