

# JMM 2018 Minicourse Feedback

Put your questions and comments here. Mimic the style below; and feel free to comment on other people's "posts".

## **How do I use this feedback page?**

Just type a question or comment. Questions in bold, answers in plain text.

## **Where do I find the online resources for this minicourse?**

R': The most important link is <http://rpruim.github.io/talks/jmm-2018/> -- I'll put links to other things on that page.

## **What are you most interested in learning about R/RStudio/teaching with R?**

- Can you show how to import data? R': *It's pretty easy, but yes.*
- 

## **How difficult / computationally expensive is it to run my own RStudio server for my students?**

RMA: You can get an open source version for free if you're an educational institution. It might be a good way to start and will be useful for students who want to do this after they graduate (and leave your server system). It gets them used to finding open source solutions, because it doesn't come with customer support. See <https://www.rstudio.com/products/RStudio/#Server>

**Wrangling:** What was the thing you said about "tidy" and "wrangling"? Was it a reference to a resource? Or a method? Found it: <https://www.tidyverse.org/>

R': Don't let me forget to handout the data wrangling cheat sheet (and some other cheat sheets) on Saturday.

**Bookdown:** A reference: <https://bookdown.org/yihui/bookdown/>

## **Do you ever ask the two big questions replacing "computer" with "statistics"? (e.g. what do you want "statistics" to do? ...)**

R': I haven't, but it might work.

## How do you teach students to deal with typing errors and the sometimes obscure R error messages?

R': I demonstrate typing errors and how to deal with them constantly (whether I mean to or not). When I get error messages as a result, I help interpret them for students when this makes sense. ("object not found" is the most common error, and it is pretty easy to explain to them what the means.)

## Is there a command to show all shapes, all line styles etc, like with colors?

R': Not that I know of, but it would be easy to do it yourself (since the shapes are just numbers).

```
gf_point(1 ~ 1:25, shape = 1:25, color = "red", fill = "skyblue", size = 5)
```



## Why teach both extractor methods, rather than just the %>% syntax?

R': I think it is reasonable for you to pick one of the two and run with it. Whether it makes sense to use `%>%` as your primary way depends on what other things are in your toolkit.

## PLEASE write the article about Excel damaging your brain!

R': We'll see.

**Tutorials:** Do you have to have a server version of R to be able to create the tutorials?

R': We serve them to you on our RS Connect server. But they can also be run directly inside of RStudio or embedded into R packages. More on some of this stuff on Saturday.

## Can I mix in ggplot2 with ggformula?

R': Yes. The plots produced by `ggformula` are `ggplot2` plots. The main reason to do this might be if there is something I neglected to add to `ggformula`. But in general, mixing and matching will probably lead to ugly, herky-jerky code.

## What is your book you mentioned?

R': *Foundations and Applications of Statistics: An Introduction Using R*

- <http://bookstore.ams.org/amstext-28> (second edition)
- <http://bookstore.ams.org/amstext-13> (first edition)

**Datasets:** I want to do this work on my own computer (I don't use an R server). I have loaded *ggformula* and *mosaic*, but the dataset *Births1978* is not loaded. Where do I find the datasets that you used today so that I can load them?

*The data is in the dataset **Births** in the library **mosaicData** package -- you can get **Births1978** using:*

```
Births1978 = Births[Births$year == 1978,]  
/ MVJ
```

R': Here is another way:

```
require(mosaicData)  
Births1978 <- Births %>% filter(year == 1978)
```

(The tutorial basically does this in a hidden R cell.)

Most of the data sets used in this tutorial are in the *mosaicData* package. (NHANES is in the *NHANES* package.)

## Can we get access to the RMD for the tutorials?

R': Let me think about that.

**MVJ:** My main reason for the question is that - as I showed you on day 1 - I can't get past the first page on any of the tutorial sections.

## How do we import data from Excel/csv/SPSS?

R': Come on Saturday to find out.

## Could you remind us how students share information with each other when on an R server?

MVJ: Top right in the Project menu there is an option «Share project»

R': If you (or someone who can help you with this) has root access, you can also set up folders that you can write to and students can read from and put symbolic links to these folders in their accounts. That's a pretty convenient way to share data, stubs of RMarkdown files, etc. but it require some work on the underlying linux server. (Shared projects is much easier, but everyone has equal access to everything in the project.)

### **Do students in an introductory class get confused about projects in R?**

R': I don't always introduce them to students as early as I do to faculty (because faculty typically have multiple projects going and students do not -- at first). But by the time I introduce them, they don't find them very confusing. (With students, I typically introduce this when they do extended projects or work in groups.)

### **How do you load packages at the System level in an R server?**

R': You can run arbitrary R code in the system startup files. For more about what happens at startup, type `?Startup`. (Users can also create user-level startup files.)

### **When you close the browser is the environment of a project automatically saved?**

R': The short answer is yes. The longer answer is that it depends on (a) what you mean by "environment" and how you have things configured. (You can opt in or out of some things.)

### **When asking IT to install R server, how much space do they need to plan for? Will the data used by the students be stored on the server during use? Are there space limitations? If my IT ask for a contact that they can talk to about these questions that I can't answer, where should I direct them?**

R': I'll try to track down the specs of the server we used at JMM (and that our students use). It supports multiple classes per semester -- probably 50 - 150 students depending on the term. Sometimes we have two sections of ~30 meeting at the same time, so ~60 is likely our maximum concurrent user load, and typically it is less than that.

Here are the specs of the server we used at JMM:

- CPU: 4 cores - Intel X5670 @ 2.93Ghz
- Memory: 32GB
- Hard Drive: 200 GB - SSD/HD auto-tiering

Unless you use really large data sets, space for data is typically the least of your problems in running a server. (Hard drives are cheap these days.) RAM is the more important resource since R typically loads the data it is working with into RAM, and each user needs enough RAM for their own copy of the data. So if you have extra money to spend, spend it on RAM.

### **In a more advanced class would you have students build some packages so that they understand what is under the hood?**

R': I have not done that because I don't teach any courses where that is a primary learning objective. If I taught a statistical programming course, I would include it there. We have had summer research students create packages.

Creating packages is not that hard, so if that is one of your learning objectives, don't be scared off.

**If you open a RStudio on your laptop are all of those packages listed? Can I just click on them and install? Do I have to reinstall every time?**

R': Installation is once *per machine/account*. Loading (actually attaching is the more technical term) must be done once *per session*. (Remember, each RMarkdown document compiles in its own session -- a new session each time you knit the file.)

On your laptop, you will see the packages installed on your machine. If you haven't installed any yourself, the list will be short -- just the packages that come automatically with the installation of R. Installing packages from CRAN is super easy, just hit the install button and follow the prompts.

**You may have addressed this, but do I have to store all data files locally? More specifically, can I store them in a cloud-based folder (e.g. Dropbox) for users to access?**

R': CSV files (and some other types of files as well) can be read from a URL. When **import dataset** you will see that the first blank is labeled File/URL and you can provide either one.

**Once you've made an RMarkdown then how do you save it as a document?**

R': It already is a document (an Rmd). Each time you hit the knit button, the Rmd file is saved and some additional files are created (HTML, Word, or PDF, depending on your choice).