

# SISBID 2018 Reproducible Research

*Keith Baggerly and Karl Broman*

*2018-07-02*

## Contents

<b>1</b>	<b>Overview</b>	<b>1</b>
<b>2</b>	<b>What You'll Need</b>	<b>1</b>
2.1	Equipment . . . . .	1
2.2	Software . . . . .	1

## 1 Overview

Welcome to the 2018 SISBID Module on Reproducible Research (RR)!

In this course, we'll discuss some of the motivation behind recent efforts to improve the reproducibility of published research. To improve the sharing of code as well as analyses, we'll work through construction of our own R packages, introducing functions which can make our jobs easier. Similarly, we'll introduce Git and GitHub for version control and file sharing.

**PLEASE DOWNLOAD AND/OR INSTALL REQUIRED SOFTWARE, PACKAGES, AND DATA BEFOREHAND!**

This will save time and (possibly) frustration.

## 2 What You'll Need

This course will involve much live demo on our part and experimentation on your part, so you'll need to be able to follow along to get the most out of it.

### 2.1 Equipment

You'll need a laptop with recent versions of the relevant software installed (please bring your power cords too!). You should have write permission to create files and folders on this laptop.

### 2.2 Software

#### 2.2.1 Base

Recent versions of R, RStudio, git, and make are downloadable from CRAN, the RStudio download page, the Git download page, and the GNU Make page respectively.

I'm currently running

- R version 3.5.0 (2018-04-23)
- RStudio 1.1.447
- git version 2.15.2 (Apple Git-101.1)

- GNU Make 3.81

on my MacBook Pro laptop (OS X v10.13.5).

You may already have some of the above tools installed. Macs typically have git and make installed by default. On Windows, make is installed as part of RTools (see below) which we'll use for assembling R packages. Once you have R and RStudio installed (and RTools, if on windows), try going to the Terminal pane (a tab on the Console pane) in RStudio and running

- `which make`
- `which git`
- `make --version`
- `git --version`

If git and make are installed and readily findable, these should tell you where they reside on your machine. We'll be using git with GitHub, which will involve registering for an account on GitHub.

If any of these setups doesn't go quickly and smoothly, *do not panic*, we'll help you with that in Seattle.

A video dealing with git installation on Windows that several people found useful in previous years is [here](#).

We won't be using git on the first day of the module in any event.

## 2.2.2 R Packages

We'll be making use of the following packages (in alphabetic order, my version numbers are shown)

- `devtools` 1.13.5
- `downloader` 0.4
- `GEOquery` 2.47.18 - this is just for parsing one of the datasets I'll use for illustration, and isn't as directly germane to RR per se.
- `here` 0.1
- `knitr` 1.20
- `lattice` 0.20.35 - for example only
- `magrittr` 1.5
- `readr` 1.1.1
- `rmarkdown` 1.9
- `roxygen2` 6.0.1
- `tidyr` 0.8.1

A broader "package of packages" I may use without extensive discussion:

- `tidyverse` 1.2.1

All of the above are available from CRAN, with the exception of `GEOquery`.

`GEOquery` is available from Bioconductor.

## 2.2.3 Windows and RTools

For those of you running Windows machines, you'll also need to install

- Rtools

in order to get new packages you create to compile properly. Jeff Leek gives a slightly more expansive description of installation [here](#).

## **2.2.4 Bonus**

### **2.2.4.1 MikTeX / TeX for pdf output**

RStudio will let us export reports in a variety of formats (e.g., html, docx, md), but some types of output require additional addins. In particular, producing pdf output makes use of pdf<sub>l</sub>atex, which in turn requires that you have some version of TeX installed. The type of distribution to install varies by operating system, so I'd use what the LaTeX page suggests.

The use of pdf<sub>l</sub>atex is discussed a bit more in RStudio posts on Customizing LaTeX Options and Using Sweave and knitr