

Data Wrangling and Visualization 101 in R

Florian Jaeger

Fall 2020

Contents

1 Preliminaries	1
1.1 Version control	1
1.2 Reproducibility and literate coding	2

1 Preliminaries

1.1 Version control

RStudio makes version control, data backup, and data sharing easy (e.g., via Github.com). To use it, download and install git on your computer. Get a free github.com or bitbucket.com account. You only have to do this once.

Then, for each project, create a new project in RStudio and link it to the remote repository (select “Create project” > “Version control”). You will have to enter a URL for the remote repository, which you get, for example, at github.com under the repository’s main page by clicking the “Clone or download button”.

For step by step instructions, follow these links:

- [Setting up RStudio for version control](#)
- [RStudio help on version control](#)
- [Reverting a file to an earlier version](#)

You can clone this repository from Github. In RStudio:

1. From the file menu (top left), select “New project ...”
2. Select “Version control”
3. Select “Git”
4. Enter “<https://github.com/tfjaeger/BCS206-207-R-lectures.git>” in the top field (for the Repository URL). Select where you want this project to be stored in the last/third field. I usually just go with the default, which for me is the desktop.
5. This should clone the tutorial for Monday from Github.com to your local drive, and open it as a new R project.
6. Once that has happened, check in the Git tab (top right of RStudio window) whether you’re on the right branch of the repository.
7. **This tutorial is the 2020-2021 branch of the git repository.** Right next to the “New Branch” button it should say “2021-2022”. If it instead says “master”, click on “master”. This opens a menu from which you can select “2021-2022”. Please do so.

1.2 Reproducibility and literate coding

R and RStudio support reproducibility oriented literate coding via Sweave and Knitr: lab books, presentations, and papers can weave/knit together data, code, and text. The document you share contains the code needed to create its outputs (figures, tables, etc.). This is achieved by combining latex or R markdown with R code (or, for that matter, code from other programming languages). For an excellent video-based introduction, see this [tutorial on R markdown](#) (see also this [helpful step by step introduction](#)). *This document is R markdown compiled with RStudio's knitr.* If you downloaded the [git repository](#), you should see both the .Rmd and the .pdf file. The latter is created (“knitted”) from the former. The former contains all the R code required to generate the figures, etc. in the PDF file.