**Collaborating on reproducible research using R and github**

*Intro*

Peikert & Brandmeier (2020, preprint)

Primary goals:

* Reporting statistical results consistent with actual statistical results (dynamic report generation)
* Analysis exactly reproduces at a later point in time if the computing platform or software is changed (computational reproducibility)
* Changes at any time are tracked, tagged, and documents while earlier versions of both data and code remain accessible

“Computational reproducibility: Using the same data to get the same results” (Daniel Lakens, 23.11.2020)

**R Markdown (Xie, Allaire, & Grolemund, 2018)**

**= dynamic document generation = literate programming**

* Interweaves human readable code and computed results to eliminate inconsistency errors
* Simple language (similar to Latex) that creates dynamic documents with embedded chunks of R code
* Documents can be exported to documents (docx, rtf, pdf), presentations (ppt, html, pdf), or websites (html) using the knitr package (Xie, 2015, 2019)
* Papers can be written in Markdown. Then the papaja package (Aust & Barth, 2018) offers additional function to enable APA style document formatting and the stargazer package (Hlavac, 2018) provides journal-ready tables and reports of statistical models
* R code is dynamically rendered to content upon the creation of the document
* R code can be given in separate chunks (```{r} ````) or inline (`r …`)
* Supports > 40 different languages

**Git (Chacon & Straub, 2014)**

**= Version control**

* Manages changes to files (data and code) over time so that one can recall specific versions of files later or revert the entire project to a past state
* Tracks changes and allows to “go back in time” –> sequence of “Snapshots” = COMMITS = copy of the state of all tracked files
  + Each commit has a unique identifier (hash code) and a human-readable description (commit massage)
  + Changes between commits can be visually compared
* REPOSITORY = Collection of all snapshots
* Use terminal (Shift + Alt + R) or graphical user interface in R studio
  + Git init = initialize git in current directory
  + Git add ../data/iris.csv . ./R/analysis.R = track specific files
  + Git commit -m “added data and analysis” = track snapshot with comment
  + Git commit -a -m “completed data collection” = add and commit all changes
  + Git log = inspect all changes
  + Git checkout [..hash, e.g. 4372eadsjk] = revert to/examine any previous state (newer states are still saved)
* GitHub = popular serve for sharing materials/collaborating with other researchers via git, but could also be GitLab or else
  + Public or privare respositories
  + Git remote = link remote github repository to local directory
  + Git push = push all changes from local repository to the remote repository
  + Git clone = to downloas git repository on another computer
  + Further tools: web interface to track issues (open/closed/resolved), means to manage and merge multiple, parallel versions of code (branches, merges, pull requests); post-publication platform to discuss manuscripts and their results; everything is backed up on a remote computer, unneeded code can be removed and the removel can be tracked
  + Specific versions can be labelled “submission”, “preprint”, “published”; zenodo.com or figshare.com can create a DOI from such a release
* Language agnostic

**Make (Feldman, 1979)**

**= dependeny management**

* How different parts of an analysis and a corresponding report relate to each other and in what order they need to be executed
* (computational) recipes to create or recreate files
* Makefile = list of recipes
  + Target = name of the recipe
  + Colon = list of dependent targets or files
  + Recipe = List of system commands to create the target
* If any of the dependencies have changed since the last time the target was built, the recipe’s commands are executed to recreate the target file

**Docker (Merkel, 2014)**

**= Containerization**

* Execution of all computer code in a virtual environment that guarantees exact reproduction independent of the host operating system, R version, and installed package versions