Business Analytics & Predictive Modeling

Guide:

GitHub Integration with RStudio

Stefan Lessmann

Johannes Haupt

Agenda

Tutorial: GitHub Integration with RStudio

- What is GitHub?
- Configure Git
- Configure RStudio
- Glossary
- Work on your project

Introduction

Git is an open source program for tracking changes in text files. This is the program that does the work.

GitHub is a web-based social and user interface based on Git for easy version control, collaboration and code sharing.

Why do I need GitHub?

- **Lets everybody work on the code, locally and at the same time**
- **Helps you solve potential version conflicts**
- **Gives you a space to share your projects**
- **All the cool people do it and researchers, too**

Introduction

- This guide will help you to set up GitHub in RStudio
- Why? Because then you can do everything inside RStudio
- In case you need further help:
  - https://www.r-bloggers.com/rstudio-and-github/
  - http://rogerdudler.github.io/git-guide/
  - https://help.github.com/articles/good-resources-for-learning-git-and-github/

Configure GitHub

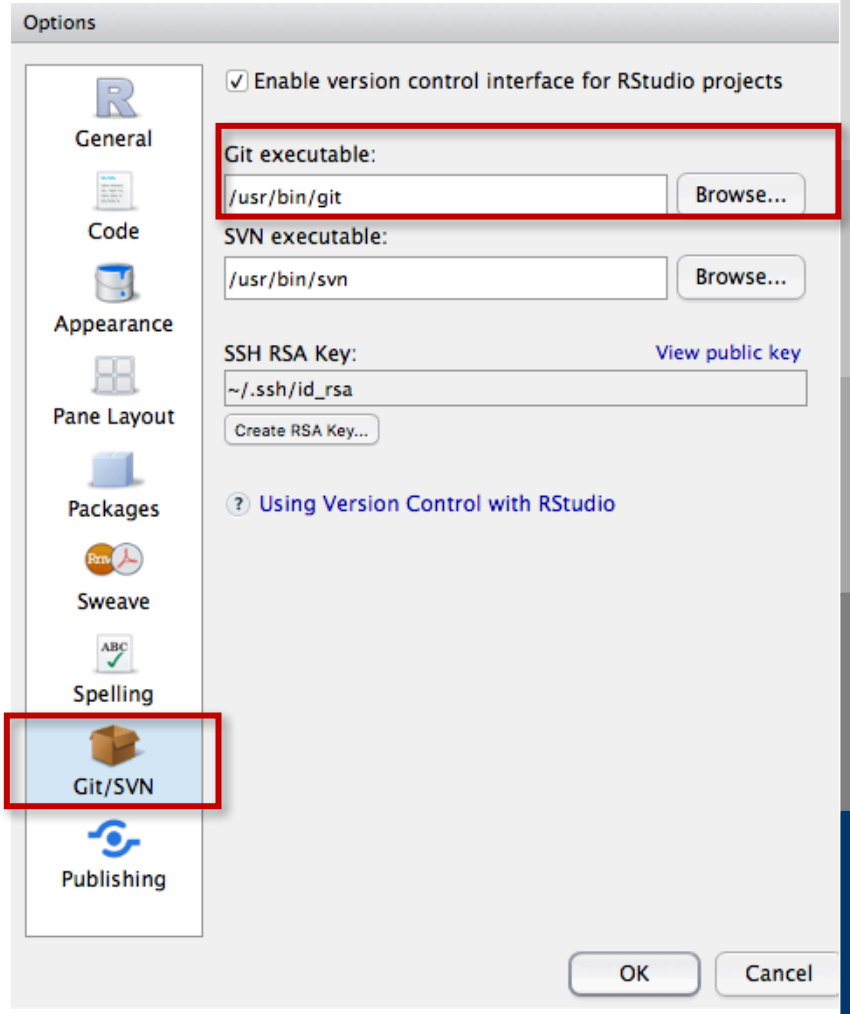
- Create a GitHub account on github.com . Choose and remember your username.
- Download Git: https://www.git-scm.com
- Install Git
- Open your Console (Windows) or Terminal (OS X / Linux)
- Type and run the following two commands to set up your user name and email address

```
git config --global user.name “Your Username”
git config --global user.email “your.email@example.com”
```

- Detailed introduction on git-scm.com (Getting started)

Configure RStudio

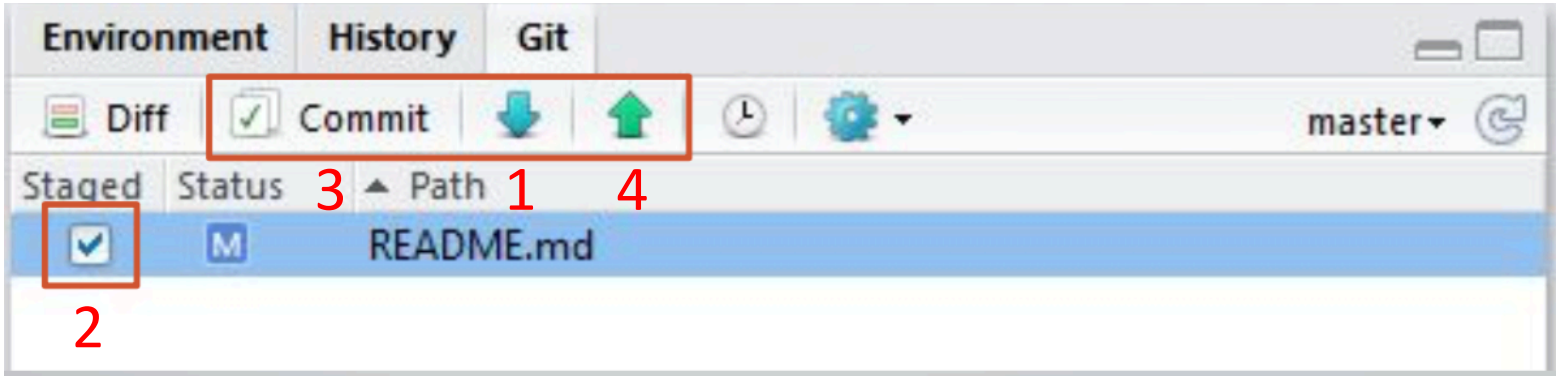
- **Ensure the path is where you saved Git**
  - RStudio: Tool → Global Options → Git/SVN
  - If the path to the Git executable is not correct, search for 'git' on your computer. Make sure to choose the file in folder 'bin'
  - Press OK and restart RStudio



Glossary

- **Repository**
  - This is the online folder/box where all the code for your project is saved on GitHub
  - Other people can see, download and contribute to the code
- **Clone**
  - Download a repository from GitHub to work from it on your computer
- **Pull**
  - Check for the newest code changes and download them
  - Merge the new code into the code on your computer and solve any conflicts one by one
- **Commit**
  - Select code you want to upload to GitHub and write a short comment about what you changed
- **Push**
  - Upload the committed code from your computer to GitHub

Work on your project



- **Workflow in RStudio:**

1. **Always** pull the newest code **before** you start making changes

☐ Work on the code and make changes

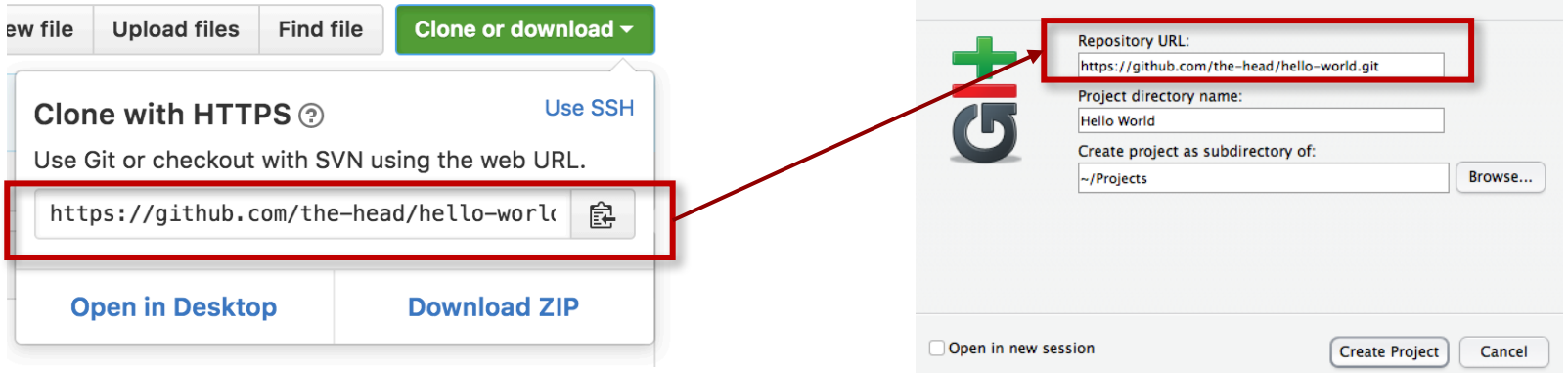
2. Select the files that you want to upload to GitHub
3. Commit the code and write a short comment about changes
4. Push the code to GitHub

Work on your project

- **Best-practice and hints:**
  - Break the project up into small, manageable files, e.g. helper function definitions or steps in the pipeline
  - Make sure to communicate about what is being done in real-life/on a messenger *and* in the commit comments
  - If something breaks or is deleted you can set files back to any version in the past

Work on your project

- **Clone your GitHub repository**
  - RStudio: File → New Project → Version Control, select Git
  - Copy your repository URL from GitHub into the corresponding field in RStudio
  - Choose a file directory where you would like to save it on your computer



Configure GitHub

- Click the link provided on the course page
- Look for a repository with your group's name and join it, if available
- The first group member to click will have to create a repository with name Group#, where # is your group number
- Only your group members and lecturers will be able to see the code you create and upload to this repository
- Detailed introduction on

  https://youtu.be/-52quDR2QSc?t=36