

Teaching computing using git and GitHub

TLMSCO, Sept 2020

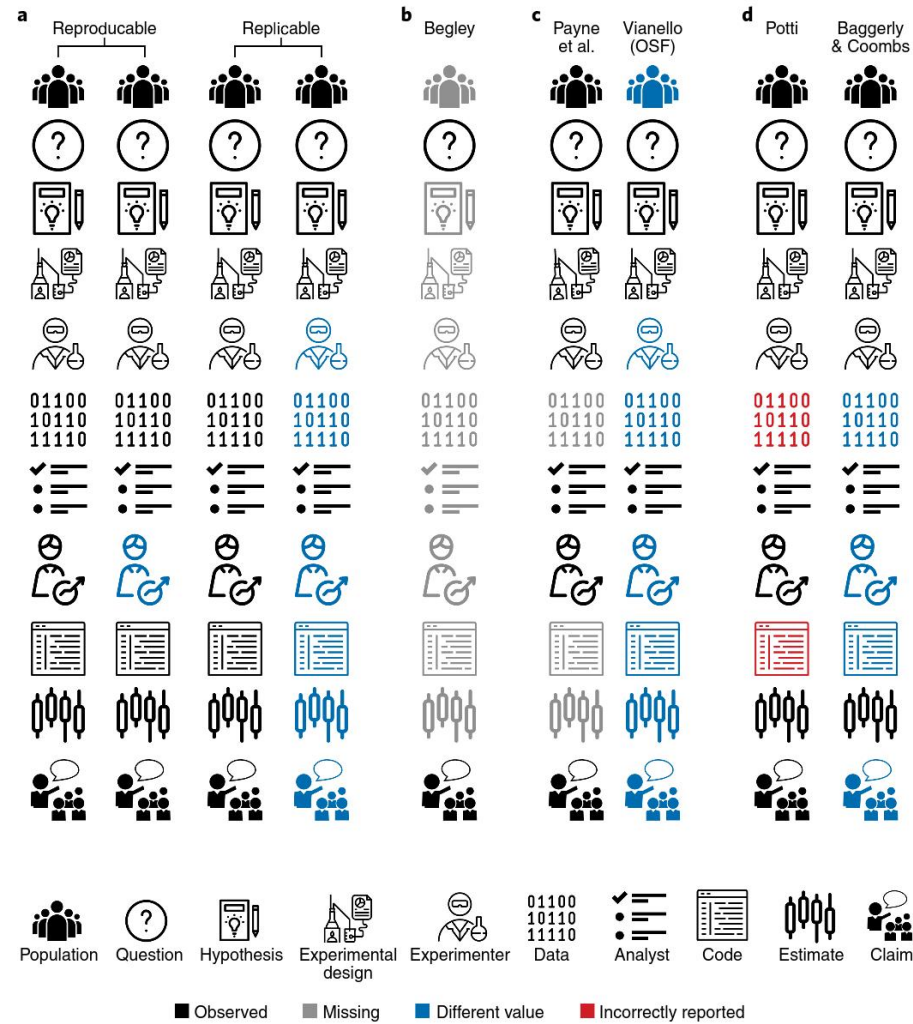
bit.ly/tlmsco_rundel

Dr. Colin Rundel

Univ of Edinburgh

Teaching Reproducible Workflows

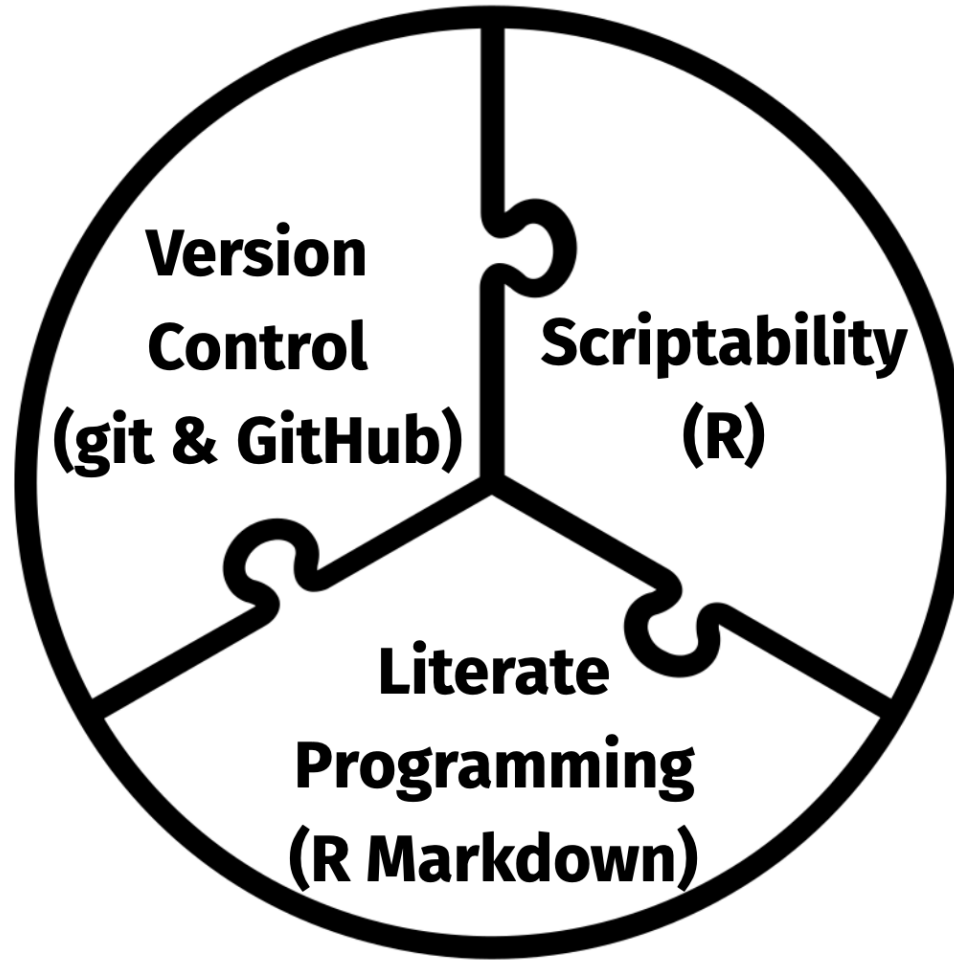
Reproducible vs Replicable



Reproducibility in practice

- Can you recreate the tables and figures reproducible from the code and data?
- Does the code actually do what you think it does?
- In addition to what was done, is it clear *why* it was done? (e.g. how were hyper / tuning parameters chosen?)
- Can the code be used for other data?
- Can you hand the code off to someone else and expect it to work?

Core pieces




Context

- I am the course organizer for Math 11176 - Statistical Programming
 - Course with ~200 Maths MSC students enrolled
- 100% coursework, multiple marked assignments (individual and team based)
- For each assignment we distribute:
 - Instruction document
 - Template Rmd for solutions
 - Data and other support files
- Need to collect:
 - Completed template Rmd
 - Rendered output (pdf, html, md, etc.)


GitHub Organization

- 1 organization / course
- Students are added (anonymously) members of the organization
- 1 template repository / assignment
- 1 private repository / assignment / (team | individual)
- Automate the distribution, collection, and feedback using GitHub's API (`ghclass`)

GitHub Organization





[Pull requests](#) [Issues](#) [Codespaces](#) [Marketplace](#) [Explore](#)





Statistical Programming (Math 11176)


✉ colin.rundel@ed.ac.uk


 **Repositories** 778

 Packages

 People 194


 Teams 190



 Projects


 Settings



Pinned repositories


Customize pinned repositories



 **exercises** ⋮


 R  2



 **hw0** Template ⋮

 R  2


 **hw1** Template ⋮

 R  2

 **hw2** Template ⋮


 R  1







Template Example - hw1


 [statprog-s1-2019](#) / [hw1](#) Template Wat

[Code](#) [Issues](#) [Pull requests](#) 1 [Actions](#) [Projects](#) [Wiki](#) [Security](#) [Insights](#) [Settings](#)

[master](#) [2 branches](#) [0 tags](#) [Go to file](#) [Add file](#) [Code](#) [Use this template](#)

 **rundel** Delete hw1_whitelist.R 479fd4c now 🕒 16 commits

 .github/workflows	Deleting file: .github/workflows/style.yml	3 months ago
 .gitignore	workflows	12 months ago
 README.md	Update README.md	11 months ago
 fizzbuzz.png	Initial	12 months ago
 hw1.Rmd	Fix typo	12 months ago
 hw1.Rproj	Initial	12 months ago

README.md 

Statistical Programming - Homework 1

Due by 11:59 pm on Friday, October 4th.

ghclass

An R package that enables instructors to automate the management of courses on GitHub.

Key features:

- Repository creation, mirroring, updating, collecting, etc.
- Organization management (members, teams, etc.)
- Summary statistics (e.g. commits) by repo or over the org
- Many other common tasks (issues, PR, etc.)

For more details see the package website - [**https://rundel.github.io/ghclass/**](https://rundel.github.io/ghclass/)

Creating a team assignment

```
org_create_assignment(  
    org = "ghclass-demo",  
    repo = c("hw01-team01", "hw01-team01", "hw01-team02", "hw01-team02"),  
    user = c("ghclass-anya", "ghclass-bruno", "ghclass-celine", "ghclass-diego"),  
    team = c("hw01-team01", "hw01-team01", "hw01-team02", "hw01-team02"),  
    source_repo = "statprog-s1-2019/hw1"  
)
```

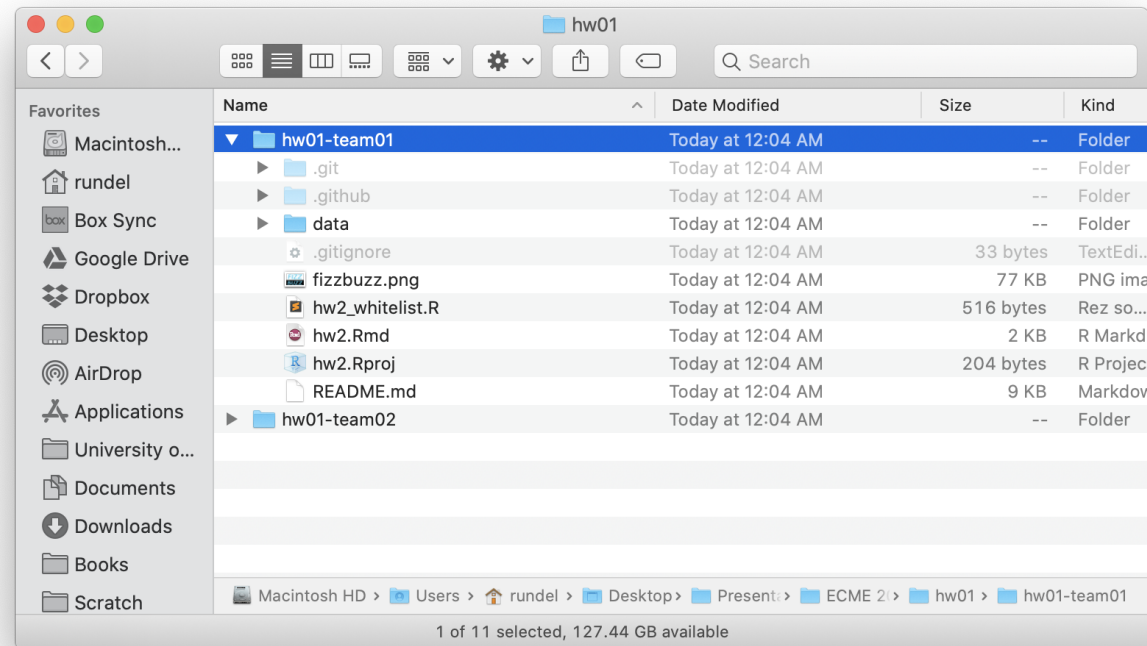
```
## ✓ Mirrored repo 'statprog-s1-2019/hw1' to repo 'ghclass-demo/hw01-team01'.  
## ✓ Mirrored repo 'statprog-s1-2019/hw1' to repo 'ghclass-demo/hw01-team02'.  
## ✓ Created team 'hw01-team01' in org 'ghclass-demo'.  
## ✓ Created team 'hw01-team02' in org 'ghclass-demo'.  
## ✓ Added user 'ghclass-anya' to team 'hw01-team01'.  
## ✓ Added user 'ghclass-bruno' to team 'hw01-team01'.  
## ✓ Added user 'ghclass-celine' to team 'hw01-team02'.  
## ✓ Added user 'ghclass-diego' to team 'hw01-team02'.  
## ✓ Added team 'hw01-team01' to repo 'ghclass-demo/hw01-team01' with 'push' access.  
## ✓ Added team 'hw01-team02' to repo 'ghclass-demo/hw01-team02' with 'push' access.
```

Collecting student work

```
local_repo_clone(repo = org_repos(org = "ghclass-demo", "hw01-"),  
                  local_path = "hw01")
```

```
## ✓ Cloned 'ghclass-demo/hw01-team01'.
```

```
## ✓ Cloned 'ghclass-demo/hw01-team02'.
```



Contributor statistics


```
repo_contributors(repo = "statprog-s1-2019/hw02-lab01-team03") %>%  
  mutate(username = LETTERS[1:4]) %>%  
  arrange(desc(commits))
```

```
## # A tibble: 4 x 3  
##   repo                                username commits  
##   <chr>                                <chr>      <int>  
## 1 statprog-s1-2019/hw02-lab01-team03 D             8  
## 2 statprog-s1-2019/hw02-lab01-team03 B             5  
## 3 statprog-s1-2019/hw02-lab01-team03 C             5  
## 4 statprog-s1-2019/hw02-lab01-team03 A             3
```

```
repo_contributors(repo = "statprog-s1-2019/hw02-lab01-team10") %>%  
  mutate(username = LETTERS[12+1:5]) %>%  
  arrange(desc(commits))
```

```
## # A tibble: 5 x 3  
##   repo                                username commits  
##   <chr>                                <chr>      <int>  
## 1 statprog-s1-2019/hw02-lab01-team10 Q            17  
## 2 statprog-s1-2019/hw02-lab01-team10 P             9  
## 3 statprog-s1-2019/hw02-lab01-team10 O             5  
## 4 statprog-s1-2019/hw02-lab01-team10 M             1  
## 5 statprog-s1-2019/hw02-lab01-team10 N             1
```

Automated feedback

 [statprog-s1-2020](#) / [hw0](#) Template

Watch 0


Star 0







Fork 0


[Code](#) [Issues](#) [Pull requests](#) [Actions](#) [Projects](#) [Wiki](#) [Security](#) [Insights](#) [Settings](#)


master 1 branch 0 tags


[Go to file](#) [Add file](#) [Code](#) [Use this template](#)


 **rundel** Modified content 5c75932 3 minutes ago 🕒 24 commits

 .github/workflows	only run on Rmd change	16 days ago
 .gitignore	Initial commit	16 days ago
 README.md	Modified content	3 minutes ago
 fizzbuzz.png	Initial commit	16 days ago
 hw0.Rmd	Update hw0.Rmd	16 days ago
 hw0.Rproj	Initial commit	16 days ago

README.md 

 Check Allowed Files passing


 Check RMarkdown Renders failing

 Check Rmd Structure failing

Statistical Programming - Homework 0

About

No description, website, or topics provided.

 [Readme](#)

Releases

No releases published
[Create a new release](#)

Packages

No packages published
[Publish your first package](#)

Automated feedback

statprog-s1-2020 / hw0

Template

Watch 0

Star 0

Fork 0

<> Code

! Issues

🔗 Pull requests

▶ Actions

📁 Projects

📖 Wiki

🛡 Security

📈 Insights

⚙ Settings

Workflows

New workflow

All workflows

Check Rmd Structure

Check RMarkdown Renders

Check Allowed Files

All workflows

Filter workflows

48 results

Event Status Branch Actor

✓ only run on Rmd change

Check Allowed Files #14: Commit 90d11db pushed by rundel

master

📅 16 days ago

🕒 49s

...

✓ use the new check version

Check Allowed Files #13: Commit 2c453df pushed by rundel

master

📅 16 days ago

🕒 48s

...

✗ use the new check version

Check Rmd Structure #2: Commit 2c453df pushed by rundel

master

📅 16 days ago

🕒 50s

...

✗ Merge branch 'master' of github.com:statprog-s1-20...

Check Rmd Structure #1: Commit 518e08e pushed by rundel

master

📅 16 days ago

🕒 57s

...

✓ Merge branch 'master' of github.com:statprog-s1-20...

Check Allowed Files #12: Commit 518e08e pushed by rundel

master

📅 16 days ago

🕒 51s

...

✗ Update hw0.Rmd

testing stuff #12: Commit 9d5ead1 pushed by rundel

master

📅 16 days ago

🕒 50s

...

Automated feedback

statprog-s1-2020 / hw0 Template

Watch 0 Star 0 Fork 0

<> Code Issues Pull requests Actions Projects Wiki Security Insights Settings

×

use the new check version

master 2c453df

Re-run jobs

✓ Check Rmd Structure
on: push

1

×

test

Check Rmd Structure / test
failed 16 days ago in 33s

Search logs

▶ ✓ Set up job

1s

▶ ✓ Initialize containers

28s

▶ ✓ Checkout

2s

▼ ✗ Check Structure

1s

1 ▶ Run rmd = parsermd::parse_rmd("https://raw.githubusercontent.com/statprog-s1-2020/hw0/master/hw0.Rmd")

11 Loading required package: utils

12 Tracing function "install.packages" in package "utils"

13

14 * The following document elements were unmodified from the template:

15 • Section "Task 1 – Implementation" has a code chunk named 'fizzbuzz' which

16 has not been modified.

17 • Section "Task 2 – Validation & Testing" has a code chunk named

18 'good_inputs' which has not been modified.

19 • Section "Task 2 – Validation & Testing" has a code chunk named

20 'bad_inputs' which has not been modified.

21

22 **##[error]Process completed with exit code 1.**

▶ ✓ Post Checkout

1s

▶ ✓ Stop containers

0s

▶ ✓ Complete job

0s

Related ongoing work

- Peer evaluation (Mine Cetinkaya-Rundel and Therese Anders)
- Simplifying the automated feedback process:
 - `checklist` - R package for simplifying automated checks
<https://github.com/rundel/checklist>
 - `parsermd` - R package for programmatic interaction with R markdown documents
<https://rundel.github.io/parsermd/>

Additional Resources

- **Happy Git and GitHyb for the useR**

Jenny Bryan, Jim Hester

- **Excuse me, do you have a moment to talk about version control?**

Jenny Bryan (2018), *The American Statistician*.

- **Using GitHub Classroom To Teach Statistics**

Jacob Fiksel, Leah Jager, Johanna Hardin, and Margaret Taub (2019), *Journal of Statistics Education*.

- **Implementing version control with Git as a learning objective in statistics courses**

Matthew Beckman, Mine Çetinkaya-Rundel, Nicholas Horton, Colin Rundel, Adam Sullivan, Maria Tackett (2020), *Journal of Statistics Education (in review)*

- **Teaching Statistics and Data Science Online Workshops**

Mine Çetinkaya-Rundel, Colin Rundel (2020), *Centre for Statistics Online Workshop*

Thank you!



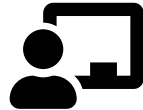
rundel@gmail.com
colin.rundel@ed.ac.uk



@rundel



bit.ly/tlmsco_rundel



statprog-s1-2019.github.io/