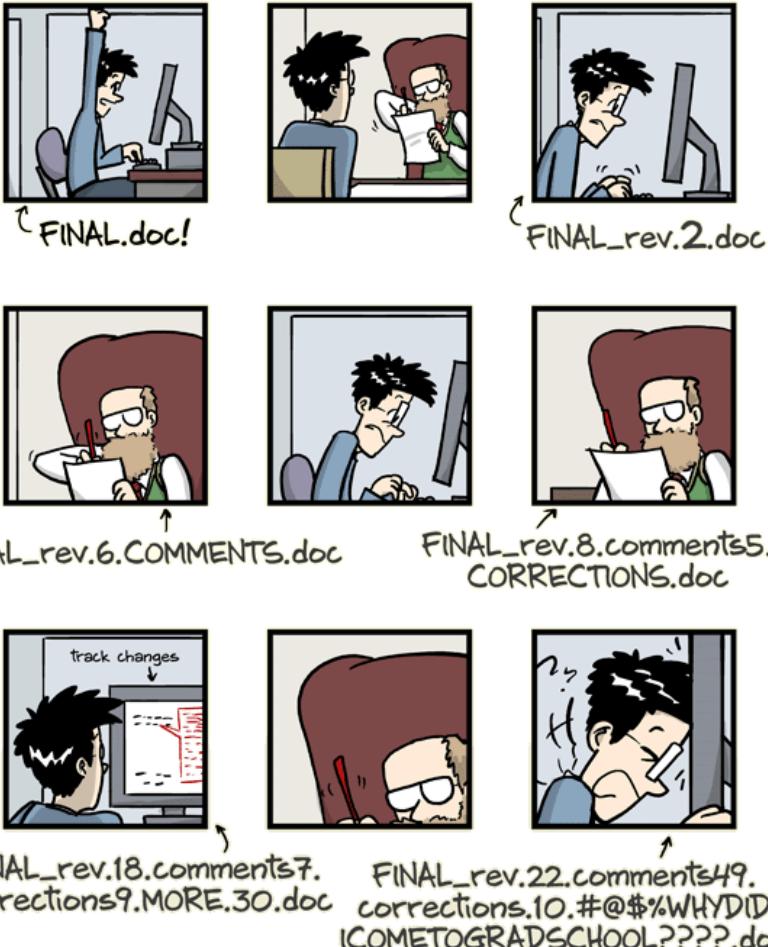


Introduction to R and Git(Hub)

"FINAL".doc



Git: Idea and concept

- distributed version control system
- track and document changes in code
- compare and find differences
- "time travel machine"
- helps to manage changes as smaller tasks
- share, publish, and collaborate on projects

[Intro to Git, for the Social Scientist](#)

[Git for Social Scientists](#)

[Git for Students in the Social Sciences](#)

[AI tool suggesting git command](#)

Git: First and basic steps

```
$ git config --global user.name <your name>
```

```
$ git init <your repository name>
```

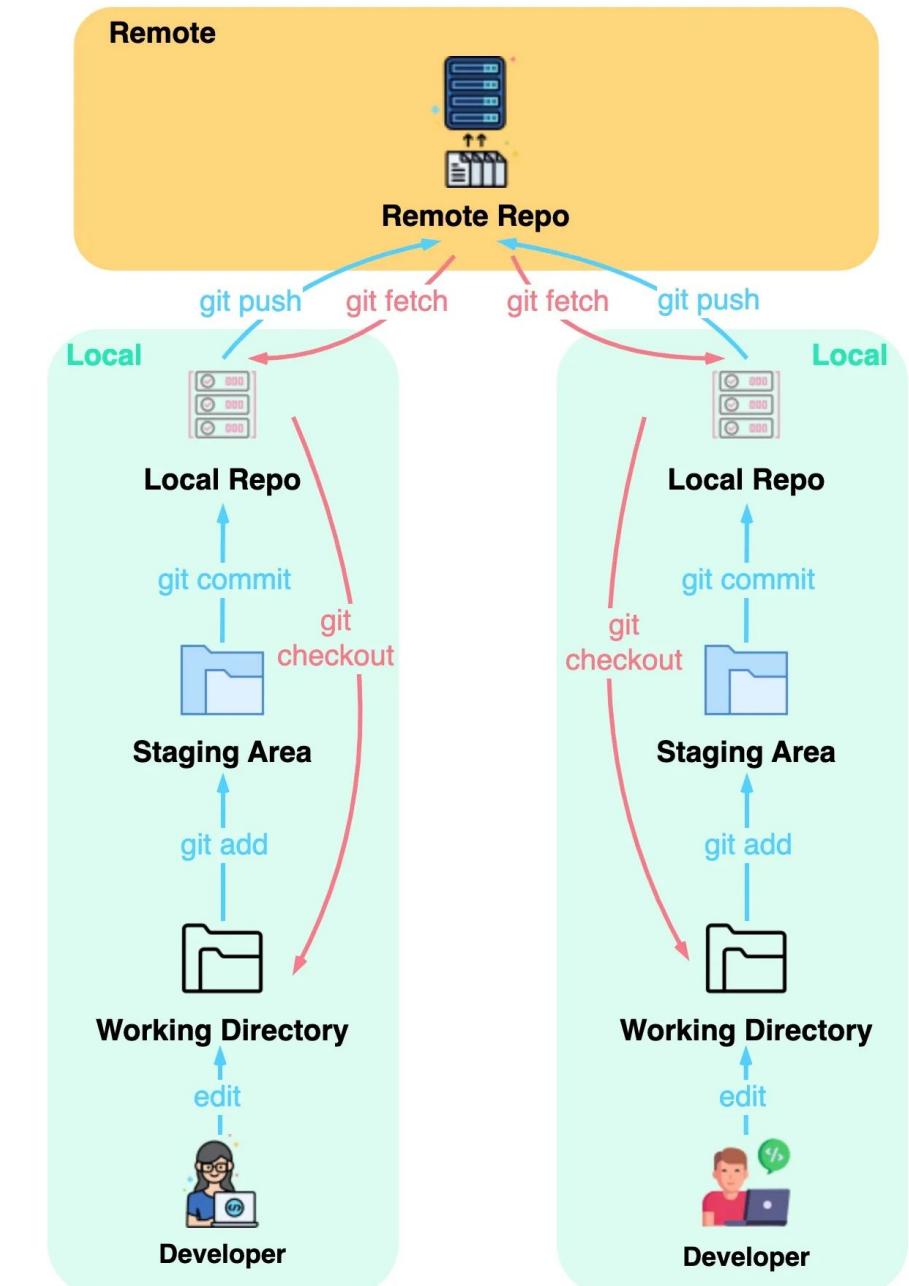
```
$ git status
```

```
$ git add <file-name-1> <file-name-2>
```

```
$ git commit -m "<commit-message>"
```

OR BOTH IN ONE

```
$ git commit -am "<commit-message>"
```



GitHub: Remote and cooperative workflow

```
$ git clone <git-repo-url>
```

```
$ git branch < branch-name>
```

```
$ git checkout <name-of-your-branch>
```

OR BOTH IN ONE

```
$ git checkout -b <name-of-your-branch>
```

```
$ git push -u (<remote> OR origin) <branch-name>
```

```
& git fetch
```

```
& git merge <branch-name>
```

OR BOTH IN ONE

```
$ git pull <remote> (<branch-name>)
```

.gitignore

Why: text file with folders and files (patterns) not to track

What: sensitive data; temp and old files; big data files; (outputs)

-> usually track just plain text working files (e.g. R scripts, LaTeX source, etc.)

How: 2 approaches ([helpful online tool creates content automatically](#))

```
/data/old  
passwords.txt  
*.doc
```

```
/*  
!.gitignore  
!/scripts
```

Code Editor: VSCode

The screenshot shows the Visual Studio Code interface with an R workspace. The left sidebar displays the workspace structure, including attached namespaces, loaded namespaces, and global environment. A code editor window titled "data_analysis.R" contains R code for generating a dataset, fitting a linear model, and plotting residuals. Below the code editor is an "R Plot" window showing a scatter plot of standardized residuals versus leverage, with Cook's distance highlighted. The status bar at the bottom indicates the R version (4.1.3), line number (Ln 14), column number (Col 1), and other settings.

```
data_analysis.R -- rdemo
1 library(data.table)
2
3 set.seed(123)
4 n <- 1000
5 dt <- data.table(id = 1:n)
6 dt[, x1 := rnorm(.N, mean = 0, sd = 2)]
7 dt[, x2 := runif(.N, min = -1, max = 1)]
8 dt[, y := 2 * x1 + x2 + 0.5 * rnorm(.N)]
9
10 model <- lm(y ~ x1 + x2, data = dt)
11
12 summary(model)
13 plot(model)
14
```

```
> rdemo r
R version 4.1.3 (2022-03-10) -- "One Push-Up"
Platform: x86_64-apple-darwin17.0 (64-bit)

> library(data.table)
> set.seed(123)
> n <- 1000
> dt <- data.table(id = 1:n)
> dt[, x1 := rnorm(.N, mean = 0, sd = 2)]
> dt[, x2 := runif(.N, min = -1, max = 1)]
> dt[, y := 2 * x1 + x2 + 0.5 * rnorm(.N)]
> model <- lm(y ~ x1 + x2, data = dt)
> summary(model)

Call:
lm(formula = y ~ x1 + x2, data = dt)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.54376 -0.31646 -0.01093  0.34316  1.61131 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.0007977  0.0155577 -0.051   0.959    
x1          1.9855040  0.0078490 252.963  <2e-16 ***
x2          1.0224384  0.0266384 38.382  <2e-16 ***  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4919 on 997 degrees of freedom
Multiple R-squared:  0.9851, Adjusted R-squared:  0.9851 
F-statistic: 3.298e+04 on 2 and 997 DF, p-value: < 2.2e-16
```

```
> plot(model)
```

Standardized residuals

Residuals vs Leverage

Leverage $\text{lm}(y \sim x1 + x2)$

Cook's distance

Scatter plots below the main plot:

- Top-left: Residuals vs x1
- Top-right: Residuals vs x2
- Bottom-left: Residuals vs $x1^2$
- Bottom-right: Residuals vs $x2^2$

R 4.1.3: 25159 Ln 14, Col 1 Spaces: 2 UTF-8 LF R ⚙

Code Editor: Benefits

- swiss army knife of coding and file management
 - search (and replace) in whole project folder
 - (better) syntax highlighting
 - customizable
- multiple languages supported (e.g. R, Python, LaTeX, Markdown)
- Git(Hub): integration for better workflow
- with R:
 - run multiple R Sessions in parallel
 - script still editable if process busy

Code Editor: HowTo

- <https://code.visualstudio.com/docs/languages/r>
- <https://renkun.me/2019/12/11/writing-r-in-vscode-a-fresh-start/>
- <https://schiff.co.nz/blog/r-and-vscode/>