

INTRODUCTION TO VERSION CONTROL SYSTEMS

TODAY...

We are going to talk about a life-changing event for me...

It happened in the year 2021...

When I discovered GitHub!

SHORT CHECK

- Raise your hand if your files also look like this

Name	Type	Compressed size
Archive	File folder	
Ingeleverd files Assignment 2	File folder	
Rhistory	RHISTORY File	4
R 20.03.08 - psa_startcode_XP	R File	10
R 20.04.03 - psa_startcode_XP_Improved	R File	11
R 2020.03.01 - Assignment-Part2-DEF_MOW	R File	10
R 2020.03.04 - Assignment-Part2	R File	9
R 2020.03.05 - Part2_PSAres_XP	R File	1
AHEM Assignment v3.2	Adobe Acrobat Document	265
AHEM-assignment	R Project	1

SHORT CHECK

- How do you keep track of changes you make in source code / script?
- Do you have experience with collaborating on a script?
 - How did you prevent conflicts and ensured the possibility to work simultaneously?

PREREQUISITE

- Basic knowledge of the statistical software 'R' and R-studio

LEARNING OBJECTIVES

- After this lecture, you should be able to:
 - Explain what are version control systems
 - Explain why using version control systems
 - Explain the (dis)advantages of using version control systems

VERSION CONTROL – WHAT?

- System to:
 - Systematically track changes made in scripts (and files)
 - Build a history of changes made

VERSION CONTROL – WHY?

- Allows checking adaptations
- Allows to re-use previous versions
- Avoid files' duplication
- Prevent (damages of) deleting the 'wrong' files
- Facilitate collaboration on a project

VERSION CONTROL – HOW?

1. Log table

LOG TABLE

- Keep track manually of:
 - Version number
 - Sometimes: file of each version number
 - Changes made
 - Person who made changes
 - Date of changes
 - Status, eventually (e.g., draft, revised, final)

LOG TABLE EXAMPLES

Title				
Description				
Created By				
Date Created				
Maintained By				
Version Number	Modified By	Modifications Made	Date Modified	Status

Source: The University of Sydney

Table 1: Document Version Control Log

Version	Date	Description of changes and person responsible for making changes
1.0	23/2/2015	Initial draft summary for comments from workshop participants (Craig Sinclair)
1.1	25/2/2015	Revision incorporating feedback from workshop facilitators (Craig Sinclair)
2.0	6/3/2015	Inclusion of submissions from two facilitators (Angus Cook, Amar Varsani) and addition of feedback points raised by 13 survey respondents (Craig Sinclair)

Source: Sinclair et al. 2015

WHAT ARE THE DRAWBACK OF SUCH APPROACH?

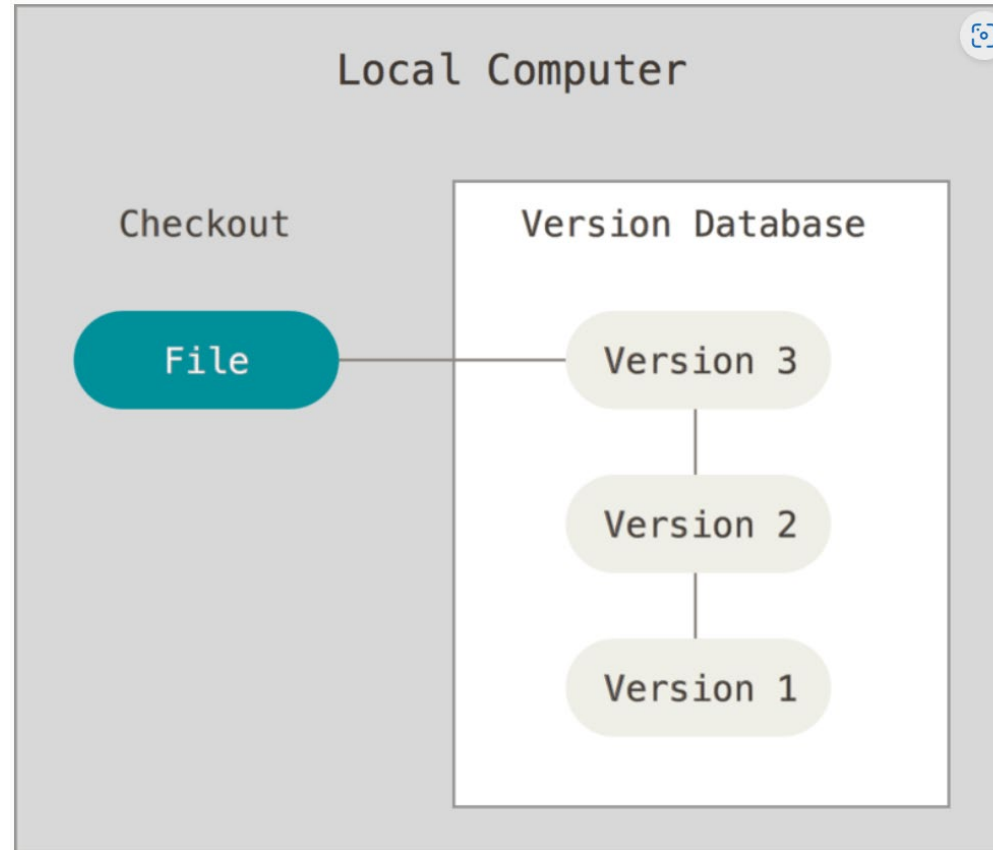
- Time consuming
- Still: save multiple version of a file
 - If not: you cannot go back to a previous version
 - If yes: takes up disc space
- You can still delete wrong files
- Collaborati
 - W... up-to-date file!

Solution?
Let a software do this for you = version control system!

VERSION CONTROL – HOW?

1. Log table
2. Version control systems
 - Local, centralized, **distributed**
 - Softwares, such as '**Git**', Mercurial
 - https://en.wikipedia.org/wiki/List_of_version-control_software

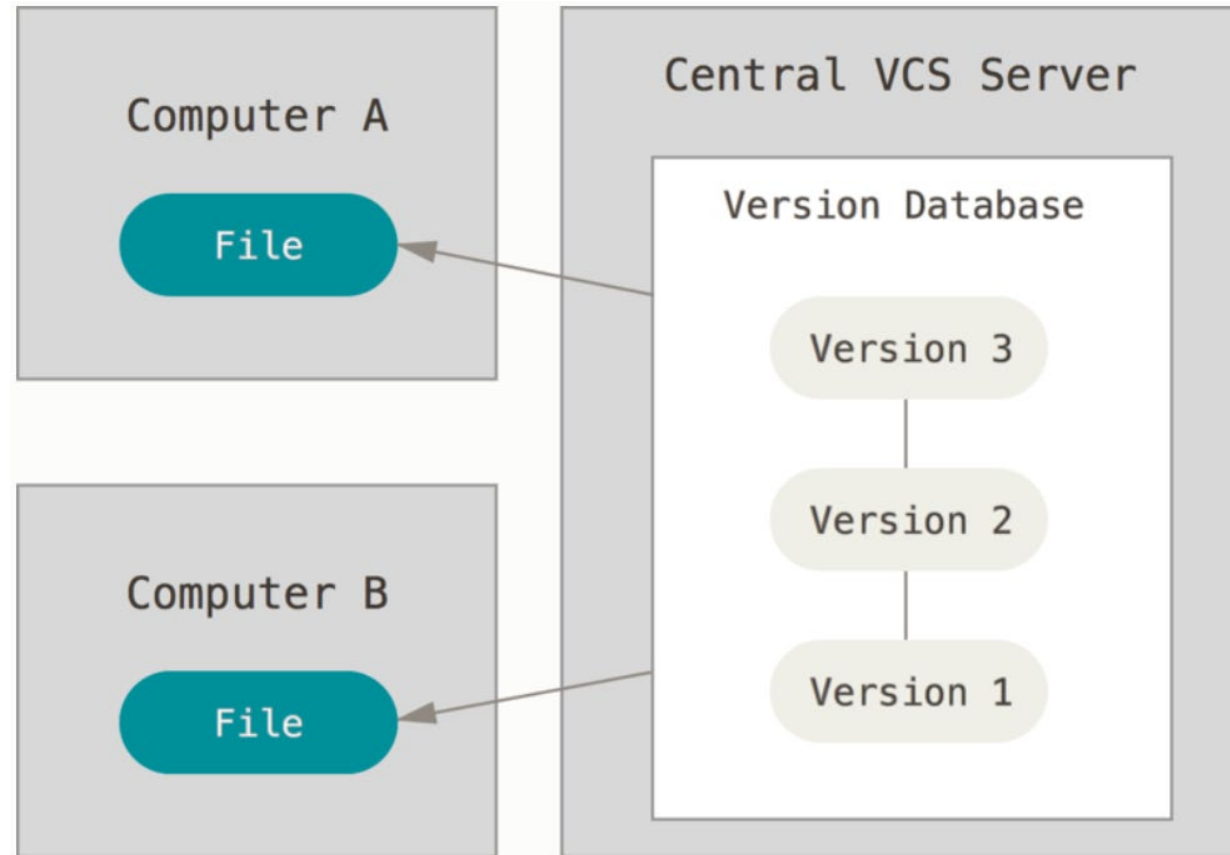
LOCAL VERSION CONTROL SYSTEM



Source: Chacon & Straub

Licence: [CC BY-SA 3.0](https://creativecommons.org/licenses/by-sa/3.0/)

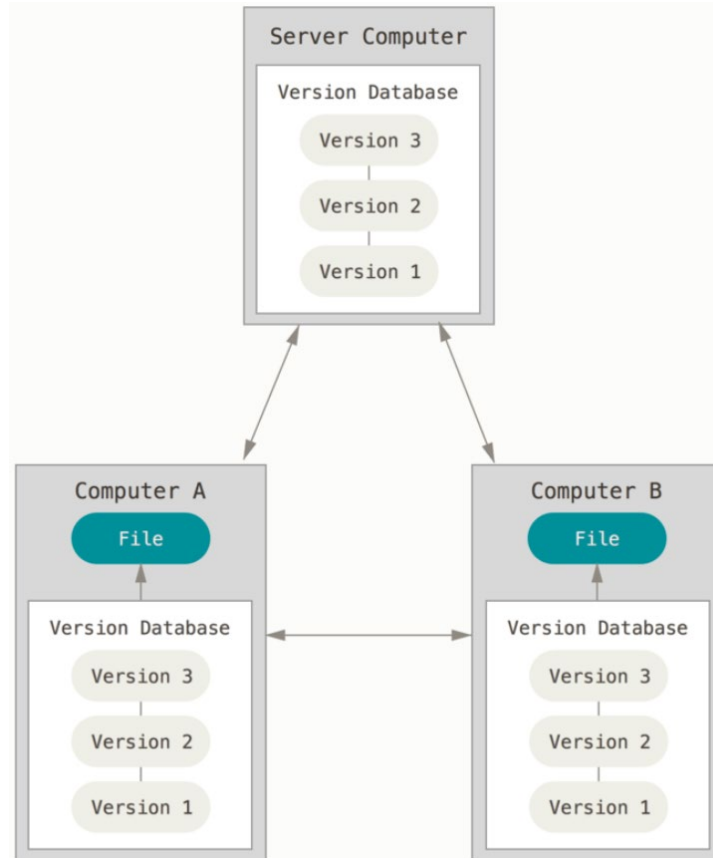
CENTRALISED VERSION CONTROL SYSTEM



Source: Chacon & Straub

Licence: [CC BY-SA 3.0](https://creativecommons.org/licenses/by-sa/3.0/)

DISTRIBUTED VERSION CONTROL SYSTEM



Source: Chacon & Straub

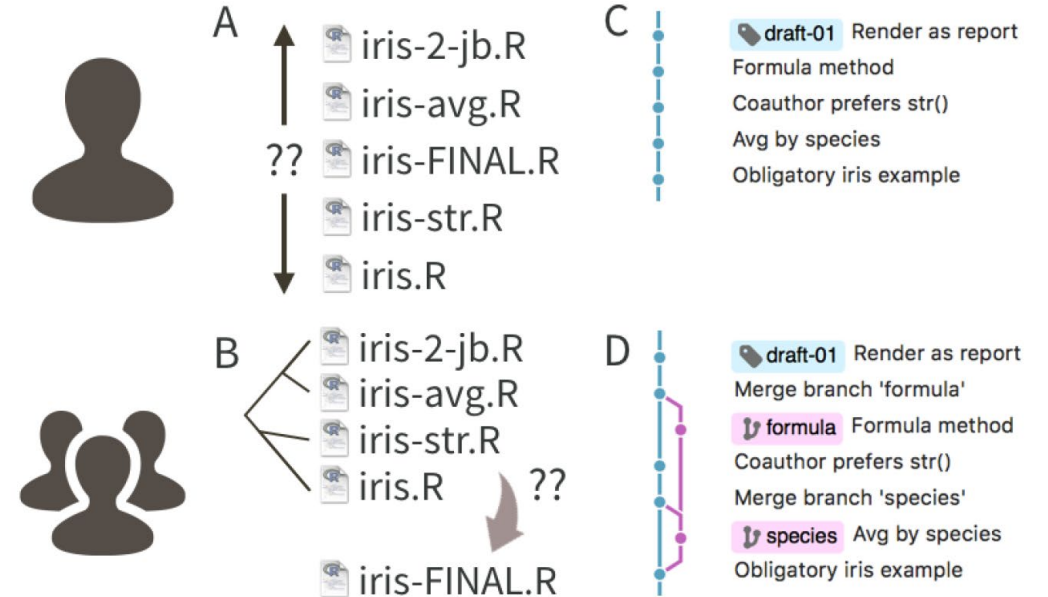
Licence: [CC BY-SA 3.0](https://creativecommons.org/licenses/by-sa/3.0/)

SOFTWARE – GIT

- Git
 - Open-source **distributed** version control system / software
 - Make a local copy of a repository
 - Multiple people can work simultaneously on the same project
 - Allows remote and offline repository / code adaptation
 - Synchronisation with repository on server
 - Integrated in all major script-based development softwares (such as R)
 - Components:
 - repos - repositories
 - branches
 - commits
 - push / pull
 - ...

WHAT IS GITHUB?

- <https://www.youtube.com/watch?v=pBy1zgt0XPc>
- Advantages of Git & GitHub
>> integration in R!
- Difference Git & GitHub
 - Git = software
 - GitHub = web-host Git system
- Alternatives to GitHub:
 - GitLab
 - Codeberg (European)



Source: Bryan – 2017

GIT(HUB) BASICS

- Repos – repositories
 - Contains all the files of your project
 - ‘Equal’ to R-project
- Branch
 - A (local) copy of the (main) repository
- Commit
 - A stamped version of a branch
- Pull
 - Action to ‘download’ a version (commit) of the online repository
- Push
 - Action to ‘upload’ your version (commit) of the repository

HOW DOES IT LOOK LIKE? REPOSITORY

The screenshot shows a web browser window displaying the GitHub repository page for Xa4P/pacheck. The browser's address bar shows the URL https://github.com/Xa4P/pacheck. The repository name is Xa4P / pacheck, and it is marked as Public. The page includes navigation links for Code, Issues, Pull requests, Actions, Projects, Wiki, Security, Insights, and Settings. The repository is currently on the master branch, with 1 branch and 0 tags. It has 237 commits, last updated on Sep 23, 2022. The repository description is "No description, website, or topics provided." The file list includes folders R, analysis, data, figs, man, output, tests, and vignettes, as well as files .Rbuildignore and .gitignore. The right sidebar shows the About section with no description, website, or topics provided, and the Releases section with no releases published.

Xa4P/pacheck Public

Pin Unwatch 2 Fork 0 Star 0

Code Issues Pull requests Actions Projects Wiki Security Insights Settings

master 1 branch 0 tags

Go to file Add file Code

Xa4P unnecessary file c44c798 on Sep 23, 2022 237 commits

R	fix bug message surv curve crossing	3 months ago
analysis	added error PSM and use of load_all()	3 months ago
data	update data	3 months ago
figs	new outputs for paper	4 months ago
man	update documentation	5 months ago
output	formatting	4 months ago
tests	Updated readme, set up for unit test	last year
vignettes	added second HE model	3 months ago
.Rbuildignore	Updated readme, set up for unit test	last year
.gitignore	start more flexible quick check framework	last year

About

No description, website, or topics provided.

Readme

GPL-3.0 license

0 stars

2 watching

0 forks

Releases

No releases published

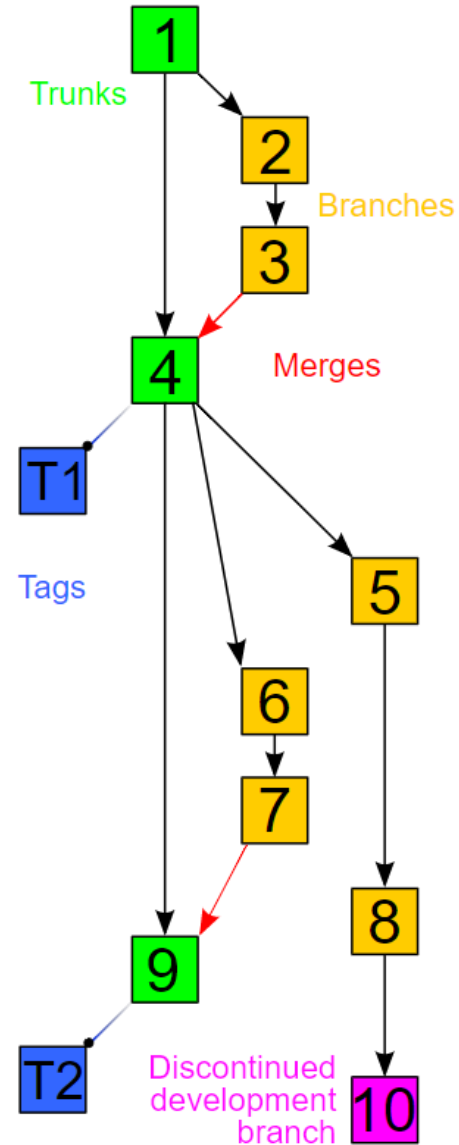
Create a new release

Packages

No packages published

Publish your first package

HOW DOES IT LOOK LIKE? BRANCHES



Source:

https://commons.wikimedia.org/wiki/File:Revision_controlled_project_visualization-2010-24-02.svg

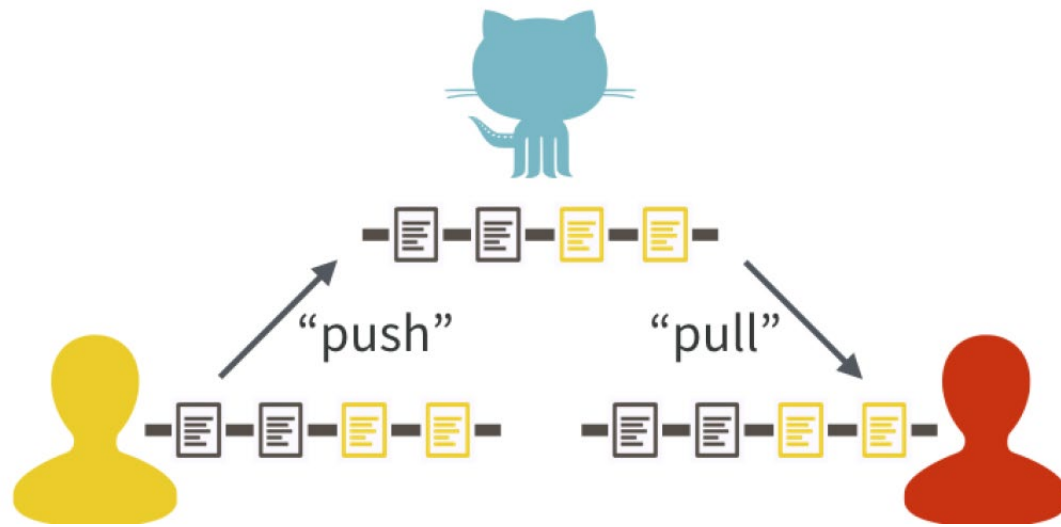
Licence: [CC BY-SA 3.0](https://creativecommons.org/licenses/by-sa/3.0/)

HOW DOES IT LOOK LIKE? COMMIT HISTORY

Commits on Sep 20, 2022

added second HE model Xa4P committed on Sep 20, 2022	05c6747
Increment version number Xa4P committed on Sep 20, 2022	45d90b1
update data Xa4P committed on Sep 20, 2022	3602871
fix bug message surv curve crossing Xa4P committed on Sep 20, 2022	a9577bd
update data Xa4P committed on Sep 20, 2022	69c52aa
increased rate PFs curves Xa4P committed on Sep 20, 2022	db7a519
added psm data with error Xa4P committed on Sep 20, 2022	3f7e652
update data Xa4P committed on Sep 20, 2022	87dda0
corrected wrong parameter definition Xa4P committed on Sep 20, 2022	c2c50b4
changed name parameters otherwise error Xa4P committed on Sep 20, 2022	49cdf0c

GIT(HUB) & COLLABORATION

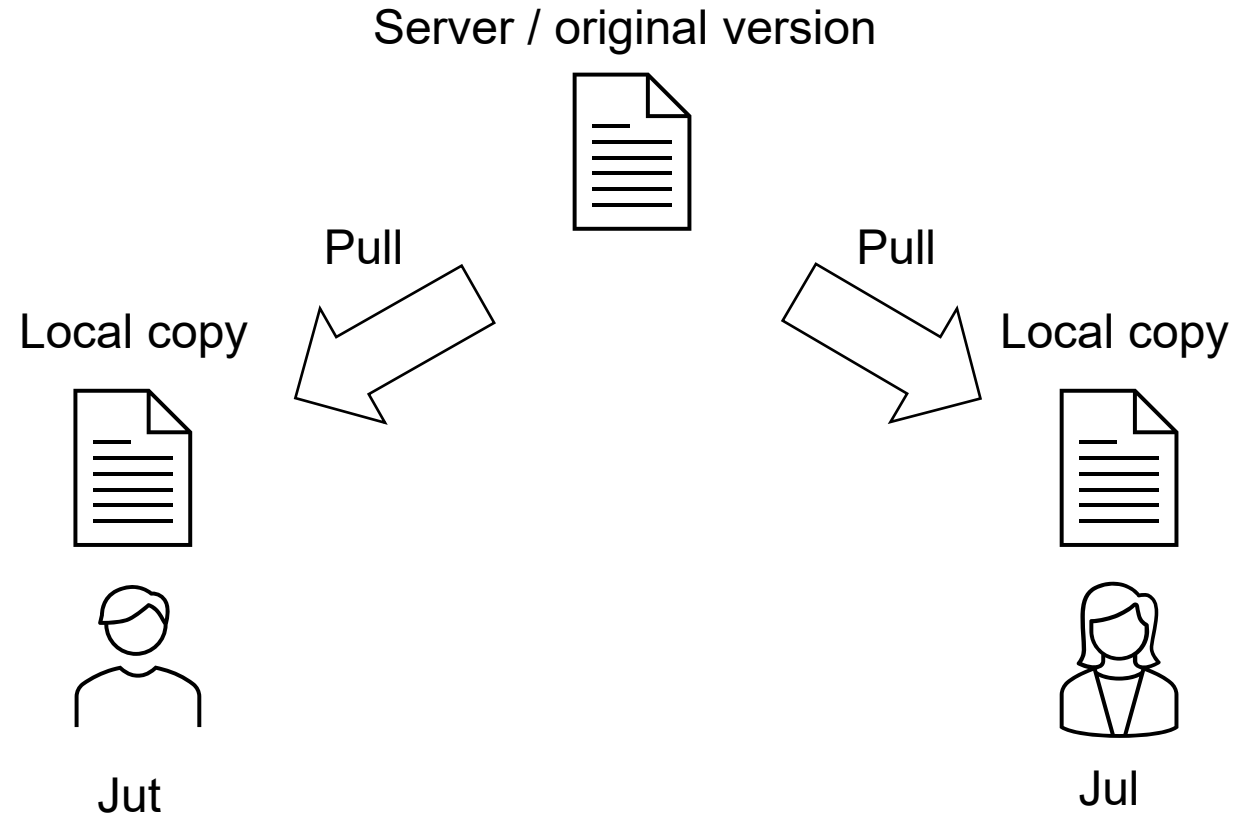


Source: Bryan – 2017

Users of your repo can:

- Download your version of the repo
- Use it
- Raise issues
- Create pull requests
 - Make changes and discussing it before merging in main branch

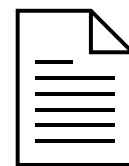
GIT(HUB) & COLLABORATION



GIT(HUB) & COLLABORATION

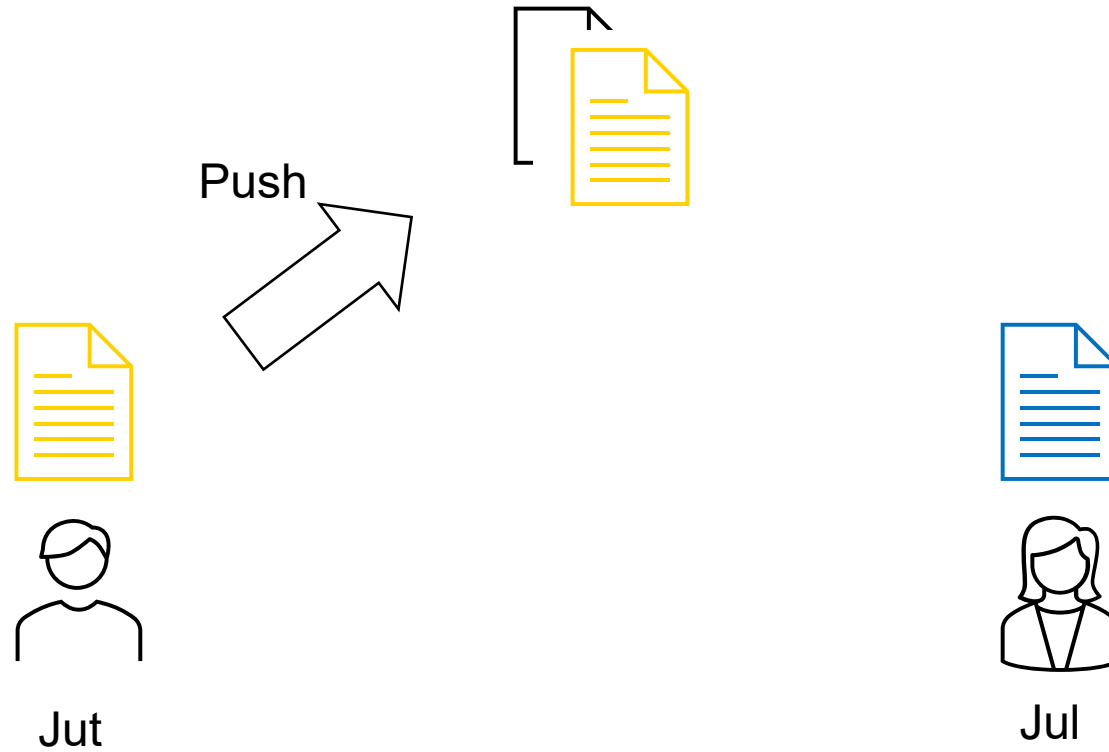


Jut

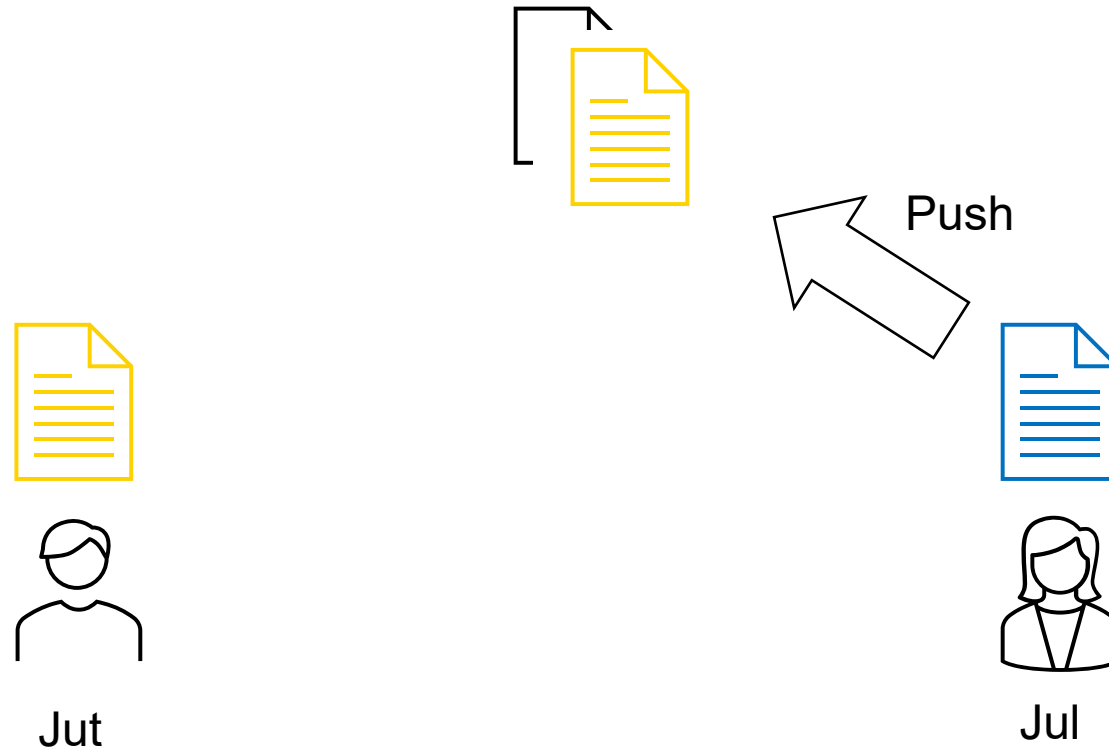


Jul

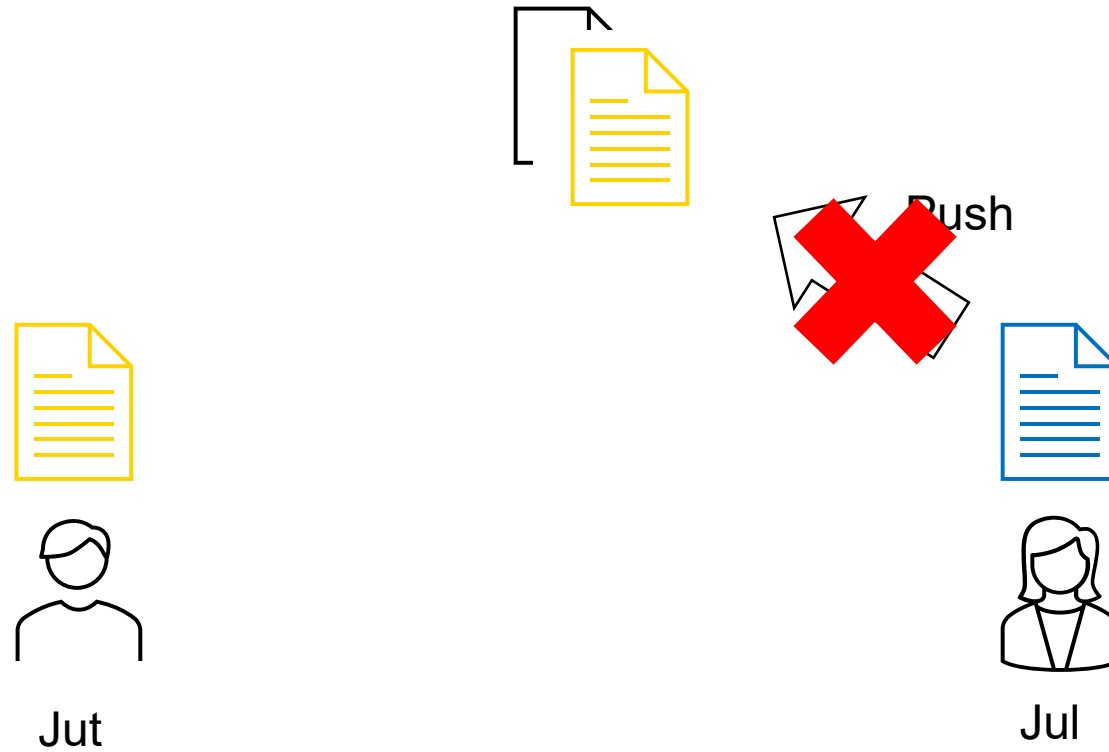
GIT(HUB) & COLLABORATION



GIT(HUB) & COLLABORATION



GIT(HUB) & COLLABORATION



GIT(HUB) & COLLABORATION



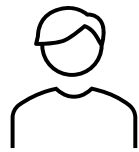
Jut

Resolve conflict



Jul

GIT(HUB) & COLLABORATION

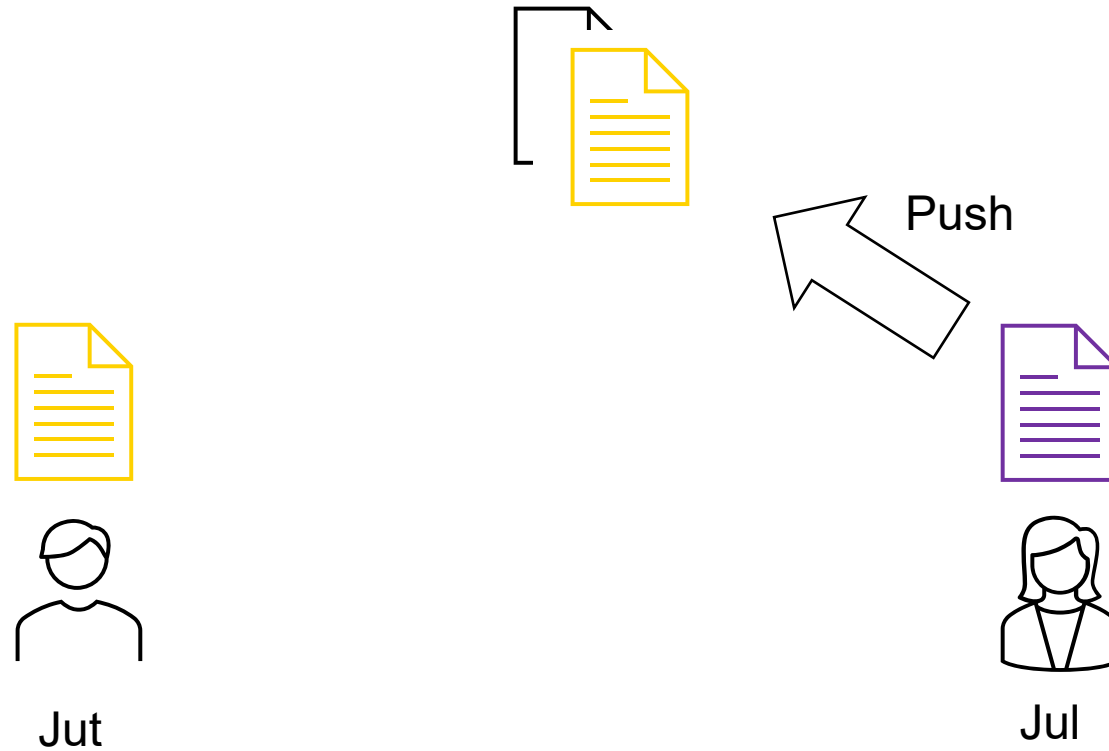


Jut



Jul

GIT(HUB) & COLLABORATION



GIT(HUB) & COLLABORATION



Jut



Jul

WARNING!



Source: <https://sascha-kasper.com/the-bumpy-learning-curve/>

GIT(HUB) & R

1. Create a Github / GitLab / Codeberg / ... account
2. (Install or upgrade R and Rstudio)
3. Install Git (if it is not already)
4. Connect GitHub with R(studio)
5. Preferred approach when working with github:
 - Github first, then R studio

Explanation of the Github – R interaction: Bryan, J. Happy Git and GitHub for the useR: <https://happygitwithr.com/index.html>

WRAP UP

- Reasons why version control system will change YOUR life:
 - No more worries about:
 - Deleting, loosing, or overwriting (CRUCIAL) files
 - File duplication
 - It creates a history of changes
 - You can review these changes
 - You can always go back!
 - It facilitates collaboration
- But... there is a steep learning curve!

ONE FINAL NOTE

- If you use GitHub (or other) for this course, please keep the repository private

ANY QUESTION?



DO IT YOURSELF! START USING GIT(HUB)

- GitHub official starting guide: <https://docs.github.com/en/get-started/quickstart/hello-world>

SOME HE – RELATED EXAMPLES

- DARTH: <https://github.com/DARTH-git>
- IVI: <https://github.com/InnovationValueInitiative>
- Synthea: <https://github.com/synthetichealth/synthea>

RESOURCES

- Introductions to version control system:
 - <https://www.youtube.com/watch?v=zbKdDsNNOhg>
 - <https://www.youtube.com/watch?v=gY2JwRfin1M>
- Git workflow
 - <https://www.youtube.com/watch?v=3a2x1iJFJWc>

RESOURCES

- Chacon, Scott, and Ben Straub. Pro git (2nd editopm) Apress. Accessed at <https://git-scm.com/book/en/v2> on 3 August 2023
- Bryan, J. Excuse me, do you have a moment to talk about version control? (2017). PeerJPreprints. <https://doi.org/10.7287/peerj.preprints.3159v2>
- Github Foundation. GitHub & Git Foundations videos playlist. (2014). <https://www.youtube.com/playlist?list=PL0lo9MOBetEHhfG9vJzVCTiDYcbhAiEqL>
- Microsoft. DevOps resource center: *What is version control?* (2022). <https://learn.microsoft.com/en-us/devops/develop/git/what-is-version-control> accessed on 16-12-2022
- Microsoft. DevOps resource center: *What is git?* (2022). <https://learn.microsoft.com/en-us/devops/develop/git/what-is-git> accessed on 16-12-2022
- Sinclair C et al. Healthy Futures Forum 2015: Regional Health Research Priorities Workshop. (2015). Great Southern Science Council: Albany. [Available from <http://www.greatsouthernsciencecouncil.org.au>]. PDF accessed on 16-12-2022 at https://www.researchgate.net/publication/277511404_Healthy_Futures_Forum_Regional_Health_Research_Priorities_Workshop
- The University of Sydney. Version control document. <https://view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Fwww.library.sydney.edu.au%2Fresearch%2Fdata-management%2Fdownloads%2Fversion-control.docx&wdOrigin=BROWSELINK> accessed on 16-12-2022