

Schedule

Yesterday

Time	Topic
09:00	Introduction to Reproducible Research
09:45	Data management and project organization
10:15	Fika break
10:45	Master your dependencies - environments and reproducibility - <i>Introduction to the package and environment manager Conda</i> - <i>Practical tutorial: Conda</i>
12:00	Lunch
13:00	Organize your analysis using workflow managers - <i>Introduction to Snakemake</i> - <i>Practical tutorial: Snakemake</i>
16:15	Wrap-up
16:30	Free time!
17:00	Joint departure for dinner
17:30	Dinner at Market

Today

Time	Topic
08:30	Distributing and version tracking your code - <i>Introduction to version control and git</i> - <i>Practical tutorial: Git</i>
12:00	Computational notebooks - <i>Introduction to Jupyter</i> - <i>Practical tutorial: Jupyter</i>
12:00	Lunch
13:00	Reproducible reports - <i>Introduction to R Markdown</i> - <i>Practical tutorial: R Markdown</i>
16:00	Containerization - <i>Introduction to containers</i> - <i>Practical tutorial: Docker</i>
16:00	Wrap-up
16:30	All done!



Managing
dependencies



Managing and executing
analysis workflow



Versioning and
collaborating on code
(and some other files)



Connecting code
and reporting

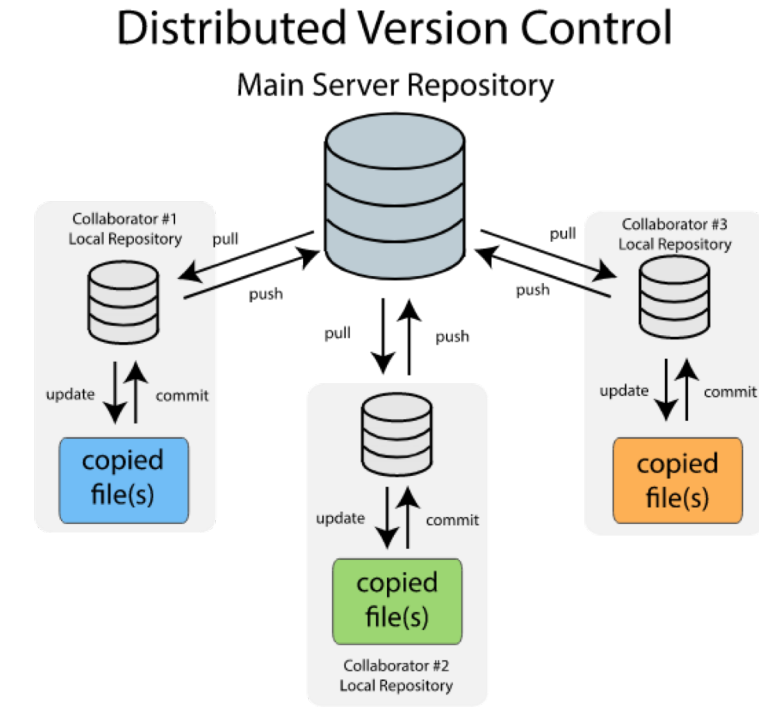
and...



Isolating and exporting
environment

What is Git?

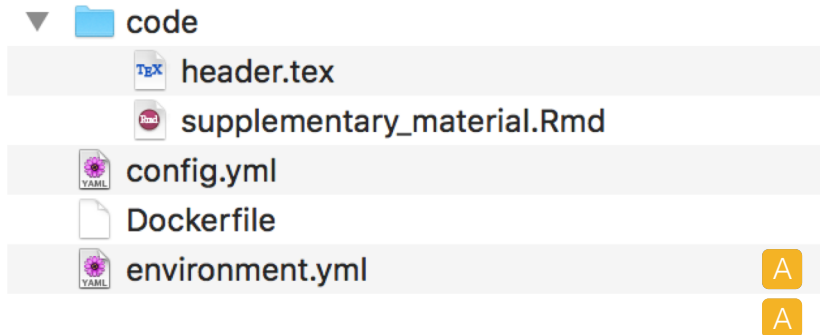
- Widely used system for file version control.
- Keeps copies of files and code from every stages in their lifecycle.
- Kind of like Dropbox, but you decide when each version is saved (and a lot more advanced features).
- Each file has a distinct history with specific incremental changes (each with a reference code).
 - Revert files to previous state.
 - Compare changes over time.
 - See who modified what.
- Makes you fearless.
- Runs on command line, but there also exists GUI and integration in e.g. text editors.
- Mainly for text files, not for binary files or large files.
- Versioning, backup, and sharing!



How does it work in practice?

Nomenclature

Repository	<i>Directory with all files, will include a .git folder</i>
Commit	<i>A specific version of the repository</i>
Push	<i>Upload local changes to remote repository</i>
Pull	<i>Download changes from remote repository</i>



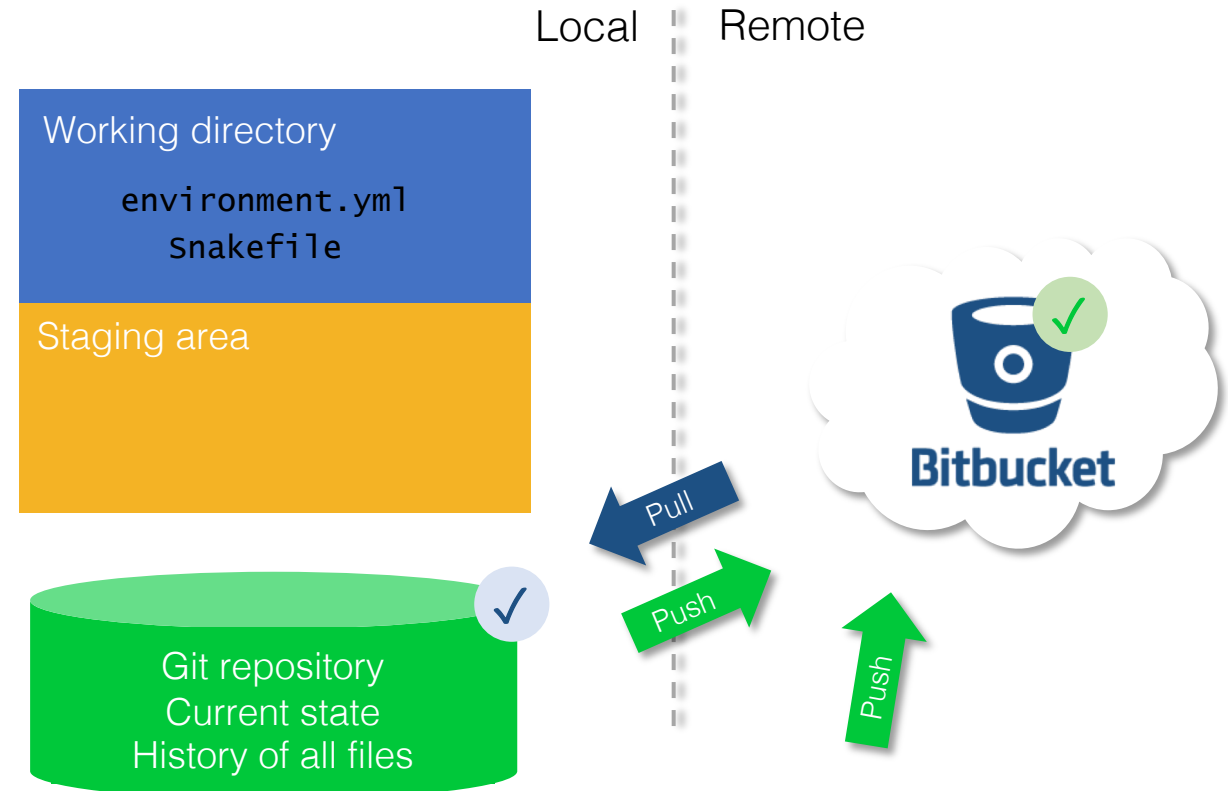
- Edit environment.yml
- Add a new file Snakefile
- `git add environment.yml`
- `git commit -m "Add snakemake 4.4.0"`
- `git add Snakefile`
- `git commit -m "Add Snakefile"`
- `git push`
- `git pull`

git log


```
a2c36bs Add heatmap figure
6152ff6 Format figure label
0abd0cb Update multiqc version
8dhfls8 Add snakemake 4.4.0
kfhs7s6 Add Snakefile
2kd7f0f Fix alignment command
```


git log

```
a2c36bs Add heatmap figure
6152ff6 Format figure label
0abd0cb Update multiqc version
8dhfls8 Add snakemake 4.4.0
kfhs7s6 Add Snakefile
2kd7f0f Fix alignment command
```



 reproducible_resear...

 Overview

 **Source**

 Commits

 Branches

 Pull requests

 Pipelines

 Issues

 Downloads

 Boards

 Settings

SciLifeLab Bioinformatics LTS / Bioinformatics LTS / reproducible_research_course / Source





docker/






















 master ▾  ▾ | reproducible_research_course /





..

code

 Dockerfile	1.1 KB	5 hours ago	split git_jupyter_docker
 Snakefile	6.5 KB	4 hours ago	moved counts table
 config.yml	1.5 KB	5 hours ago	split git_jupyter_docker
 environment.yml	202 B	5 hours ago	split git_jupyter_docker

	Rasmus Ågren	232ffc0	language and spelling in git tutorial	5 hours ago
	varemo	a81f6e8	remove Where to next sections	6 hours ago
	varemo	d8e30df <small>M</small>	Merge branch 'master' of https://bitbucket.org/scilifelab-lts/reproducible_research...	6 hours ago
	varemo	449b4b9	remove todo	6 hours ago
	Rasmus Ågren	bfa0edd <small>M</small>	Merge branch 'master' of https://bitbucket.org/scilifelab-lts/reproducible_research...	7 hours ago
	Rasmus Ågren	851e43e	language and spelling in git tutorial	7 hours ago
	varemo	cc2bbcb	fix figure fonts	7 hours ago
	varemo	19c67d4 <small>M</small>	Merge branch 'master' of https://bitbucket.org/scilifelab-lts/reproducible_research...	7 hours ago
	varemo	3a1806f	fix nicer figure	7 hours ago
	Rasmus Ågren	1487f3b	test anchoring	8 hours ago
	Rasmus Ågren	3397711	split snakemake env for speed	8 hours ago
	varemo	8f265b0	tutorial text	9 hours ago
	varemo	414de52	typo	9 hours ago
	varemo	208e1d9	restructure pages	10 hours ago
	varemo	9fca2f7	remove take down part	10 hours ago
	varemo	831a44e	revert back to project a from project new	10 hours ago
	varemo	3575b9c <small>M</small>	Merge branch 'master' of https://bitbucket.org/scilifelab-lts/reproducible_research...	11 hours ago
	Rasmus Ågren	0e4296a	typo	11 hours ago
	varemo	aeffaa3	restructure pages	11 hours ago
	Rasmus Ågren	14c60cb <small>M</small>	Merge branch 'master' of https://bitbucket.org/scilifelab-lts/reproducible_research...	11 hours ago
	Rasmus Ågren	9795b0f	language and spelling in the conda tutorial	11 hours ago

Diff from  8baf4a1 2017-11-22 ▼ to  d797810 2017-11-23 ▼

 snakemake/Snakefile_mrsa **MODIFIED**

Side-by-side diff

...

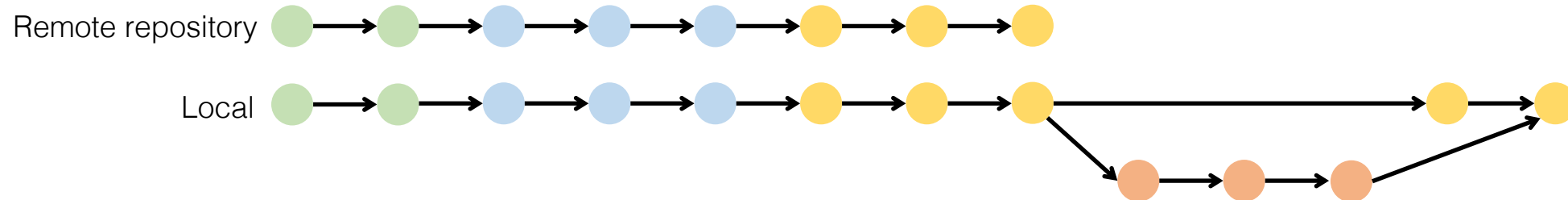
...

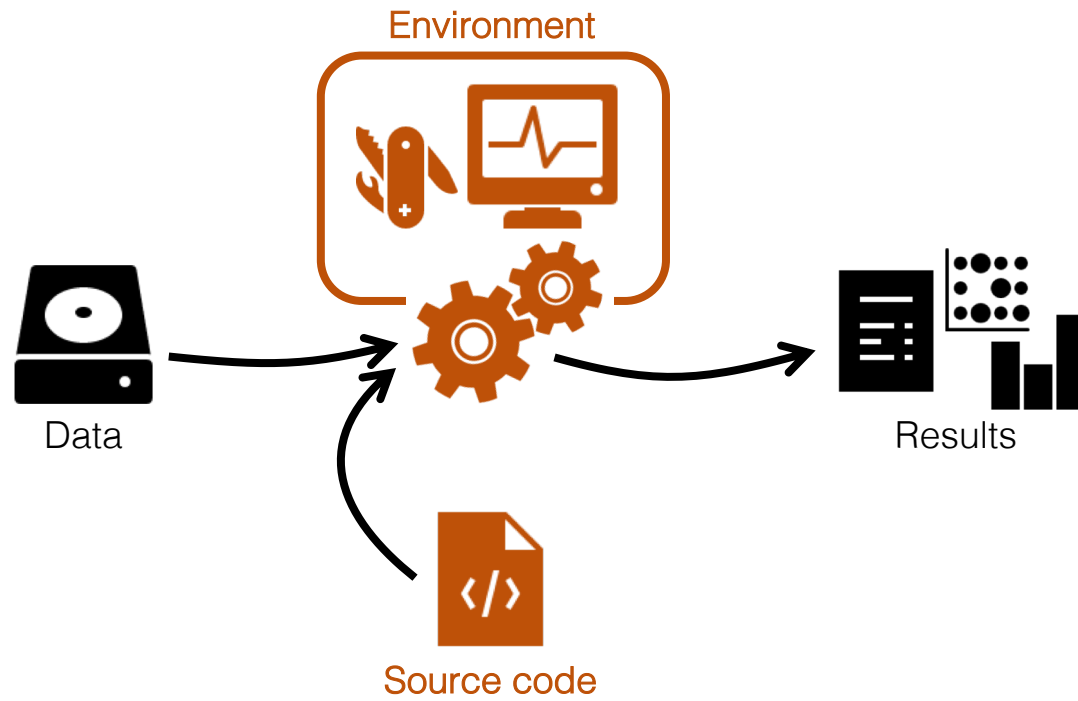
```
156 156
157 157     # Save the count table as a temporary file and then prepend a header line
158 158     # with the sample names
159      - htseq-count --format bam --type gene --idattr gene_id {input.bams} tempfile > tempfile2
      159 + htseq-count --format bam --type gene --additional-attr description --idattr gene_id {input.bams} tempfile > tempfile2
160 160     echo '{input.bams}' | tr ' ' '\t' | cat - tempfile2 > {output}
161 161
162 162     # Remove the temporary files
```

...

During the working day...

- Pull collaborator's latest work to get your local repository up to date.
- Carry on with your work and edit files.
- Commit often!
 - Each commit should be related to a distinct change/addition/task.
 - Write descriptive commit messages.
- Push your changes to the remote repository.
- If you know several people are actively working on the same repository, push and pull often!





```
project
|- doc/
|
|- data/
|   |- raw_external/
|   |- raw_internal/
|   |- meta/
|
|- code/
|- notebooks/
|
|- intermediate/
|- scratch/
|- logs/
|
|- results/
|   |- figures/
|   |- tables/
|   |- reports/
|
|- Snakefile
|- config.yml
|- environment.yml
|- Dockerfile
```