Git: versioning painkiller for the busy agronomist

Antoine Languillaume – PPS Lunch meeting 21-04-21

Credit : Thinkr.fr

Science, Data Analysis, Modeling needs Code

How to keep track of the changes in our code ?

Usual Disclaimer:
This is not a geek-only seminar

Problems: git will solve for you

Problems: git will solve for you

- Keep track of your research project
- Collaborate efficiently with others
- Disseminate your work

Problems: git will solve for you

- Keep track of your research project
- Collaborate efficiently with others
- Disseminate your work

Problems: git will solve for you

- Keep track of your research project
- Collaborate efficiently with others
- Disseminate your work

Keep track of your changes

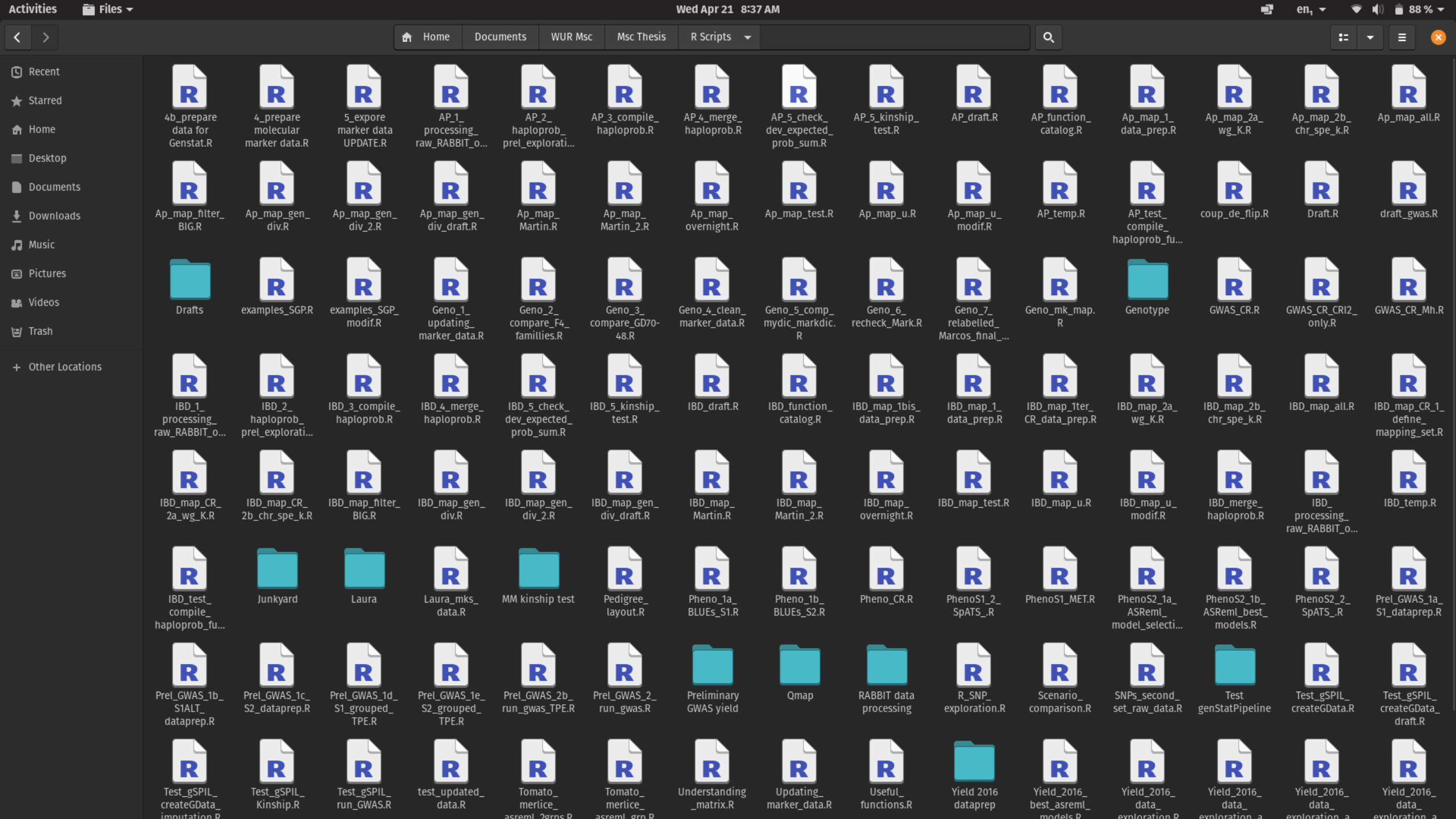
file-1.py

file-1a.py

file-2b.py

file-final.py

file-final-bis3.py


Activities
Files
Wed Apr 21 8:37 AM
en,
88 %
Home
Documents
WUR Msc
Msc Thesis
R Scripts
Recent
Starred
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
Other Locations
4b_prepare data for Genstat.R
4_prepare molecular marker data.R
5_expore marker data UPDATE.R
AP_1_ processing_ raw_RABBIT_o...
AP_2_ haploprob_ prel_explorati...
AP_3_compile_ haploprob.R
AP_4_merge_ haploprob.R
AP_5_check_ dev_expected_ prob_sum.R
AP_5_kinship_ test.R
AP_draft.R
AP_function_ catalog.R
Ap_map_1_ data_prep.R
Ap_map_2a_ wg_K.R
Ap_map_2b_ chr_spe_k.R
Ap_map_all.R
Ap_map_filter_ BIG.R
Ap_map_gen_ div.R
Ap_map_gen_ div_2.R
Ap_map_gen_ div_draft.R
Ap_map_ Martin.R
Ap_map_ Martin_2.R
Ap_map_ overnight.R
Ap_map_test.R
Ap_map_u.R
Ap_map_u_ modif.R
AP_temp.R
AP_test_ compile_ haploprob_fu...
coup_de_flip.R
Draft.R
draft_gwas.R
Drafts
examples_SGP.R
examples_SGP_ modif.R
Geno_1_ updating_ marker_data.R
Geno_2_ compare_F4_ famillies.R
Geno_3_ compare_GD70-48.R
Geno_4_clean_ marker_data.R
Geno_5_comp_ mydic_markdic.R
Geno_6_ recheck_Mark.R
Geno_7_ relabelled_ Marcos_final_...
Geno_mk_map.R
Genotype
GWAS_CR.R
GWAS_CR_CRI2_ only.R
GWAS_CR_Mh.R
IBD_1_ processing_ raw_RABBIT_o...
IBD_2_ haploprob_ prel_explorati...
IBD_3_compile_ haploprob.R
IBD_4_merge_ haploprob.R
IBD_5_check_ dev_expected_ prob_sum.R
IBD_5_kinship_ test.R
IBD_draft.R
IBD_function_ catalog.R
IBD_map_1bis_ data_prep.R
IBD_map_1_ data_prep.R
IBD_map_1ter_ CR_data_prep.R
IBD_map_2a_ wg_K.R
IBD_map_2b_ chr_spe_k.R
IBD_map_all.R
IBD_map_CR_1_ define_ mapping_set.R
IBD_map_CR_ 2a_wg_K.R
IBD_map_CR_ 2b_chr_spe_k.R
IBD_map_filter_ BIG.R
IBD_map_gen_ div.R
IBD_map_gen_ div_2.R
IBD_map_gen_ div_draft.R
IBD_map_ Martin.R
IBD_map_ Martin_2.R
IBD_map_ overnight.R
IBD_map_test.R
IBD_map_u.R
IBD_map_u_ modif.R
IBD_merge_ haploprob.R
IBD_ processing_ raw_RABBIT_o...
IBD_temp.R
IBD_test_ compile_ haploprob_fu...
Junkyard
Laura
Laura_mks_ data.R
MM kinship test
Pedigree_ layout.R
Pheno_1a_ BLUEs_S1.R
Pheno_1b_ BLUEs_S2.R
Pheno_CR.R
PhenoS1_2_ SpATS_.R
PhenoS1_MET.R
PhenoS2_1a_ ASReml_ model_selecti...
PhenoS2_1b_ ASReml_best_ models.R
PhenoS2_2_ SpATS_.R
Prel_GWAS_1a_ S1_dataprep.R
Prel_GWAS_1b_ S1ALT_ dataprep.R
Prel_GWAS_1c_ S2_dataprep.R
Prel_GWAS_1d_ S1_grouped_ TPE.R
Prel_GWAS_1e_ S2_grouped_ TPE.R
Prel_GWAS_2b_ run_gwas_TPE.R
Prel_GWAS_2_ run_gwas.R
Preliminary GWAS yield
Qmap
RABBIT data processing
R_SNP_ exploration.R
Scenario_ comparison.R
SNPs_second_ set_raw_data.R
Test genStatPipeline
Test_gSPIL_ createGData.R
Test_gSPIL_ createGData_ draft.R
Test_gSPIL_ createGData_ imputation.R
Test_gSPIL_ Kinship.R
Test_gSPIL_ run_GWAS.R
test_updated_ data.R
Tomato_ merlice_ asreml_2grps.R
Tomato_ merlice_ asreml_grp.R
Understanding _matrix.R
Updating_ marker_data.R
Useful_ functions.R
Yield 2016 dataprep
Yield_2016_ best_asreml_ models.R
Yield_2016_ data_ exploration.R
Yield_2016_ data_ exploration_a
Yield_2016_ data_ exploration_a
Yield_2016_ data_ exploration_a

There is a better way...

Git allows you to take snapshots of your work

Snapshot == COMMIT

Saving copies

file-1.py

Making commits

file.py

Saving copies

file-1.py

file-1a.py

Making commits

file.py

file.py

Saving copies

file-1.py

file-1a.py

file-2b.py

Making commits

file.py

file.py

file.py

Saving copies

file-1.py

file-1a.py

file-2b.py

file-final.py

Making commits

file.py

file.py

file.py

file.py

Saving copies

file-1.py

file-1a.py

file-2b.py

file-final.py

file-final-bis3.py

Making commits

file.py

file.py

file.py

file.py

file.py

```
...  ...  @@ -99,9 +99,13 @@ aggregate_regional_stats <- function(data_df, regions_sf, buffer_zones_sf, regio
 99   99      vars_ws <- tidyr::pivot_longer(data = data_weighted_mean,
100  100                                     cols = matches("__"),
101  101                                     names_to = c(".value", "crop", "year"),
102       -                                  names_pattern = "(.*)__([a-z]*)__(\\d{4})$")
     102  +                                  names_pattern = "(.*)__([A-z\\.]*)__(\\d{4})$")
103  103      vars_ws$station_id <- as.numeric(vars_ws$station_id)
104  104
     105  +   # Crop names with space are made syntactic when calling .rgyga_st_intersection()
     106  +   # Restore original crop names
     107  +   vars_ws$crop <- gsub("\\.", " ", vars_ws$crop)
     108  +
105  109      return(vars_ws)
106  110    }
```

- Author + email
- Date
- Message

Edmund Blackadder
edmund.blackadder@bbc.uk

19/02/1993 – 3.02 pm

Suppress warnings from extract_area()

Set calling handler sentinel scanning warning and promoting them to error if not corresponding to the matching pattern. This is probably not due to raster::extract() directly but more likely to furrr::future_map() calling raster::extract().

- Author + email
- Date
- Message

Edmund Blackadder
edmund.blackadder@bbc.uk

19/02/1993 – 3.02 pm

Suppress warnings from extract_area()

Set calling handler sentinel scanning warning and promoting them to error if not corresponding to the matching pattern. This is probably not due to raster::extract() directly but more likely to furrr::future_map() calling raster::extract().

- Author + email
- Date
- Message

Edmund Blackadder
edmund.blackadder@bbc.uk

19/02/1993 – 3.02 pm

Suppress warnings from extract_area()

Set calling handler sentinel scanning warning and promoting them to error if not corresponding to the matching pattern. This is probably not due to raster::extract() directly but more likely to furrr::future_map() calling raster::extract().

- Author + email
- Date
- Message
- And …

Edmund Blackadder
edmund.blackadder@bbc.uk

19/02/1993 – 3.02 pm

Suppress warnings from extract_area()

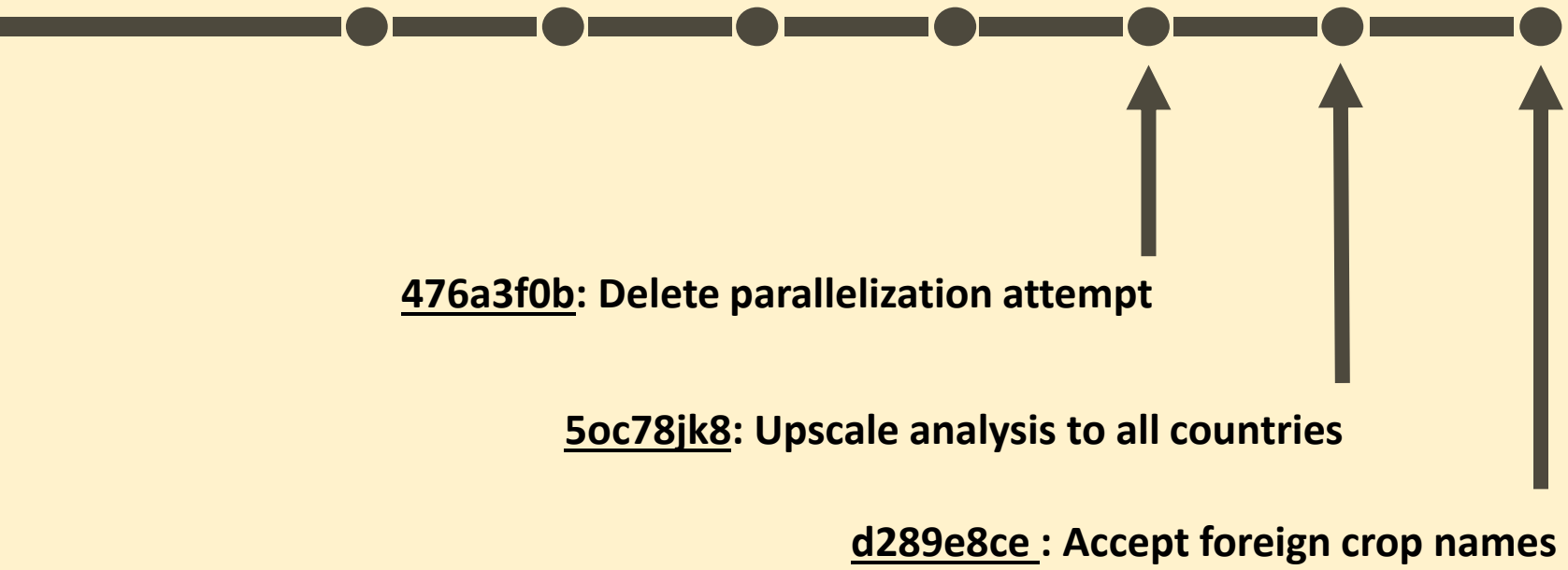
Set calling handler sentinel scanning warning and promoting them to error if not corresponding to the matching pattern. This is probably not due to raster::extract() directly but more likely to furrr::future_map() calling raster::extract().

476a3f0bcb6ae281dc59fa3ba05abbf78c6c197e

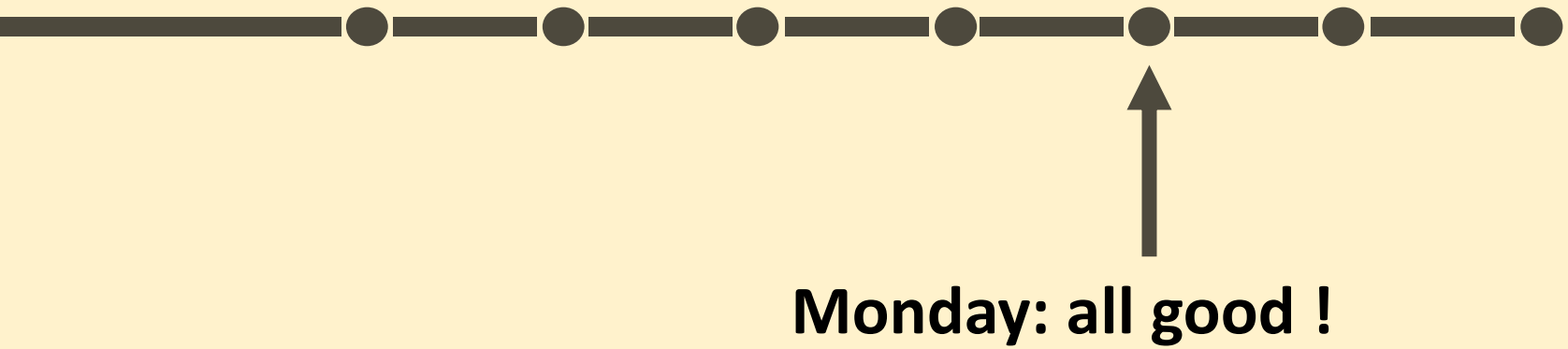
476a3f0bcb6ae281dc59fa3ba05abbf78c6c197e

Commit SHA (or HASH) == unique identifier

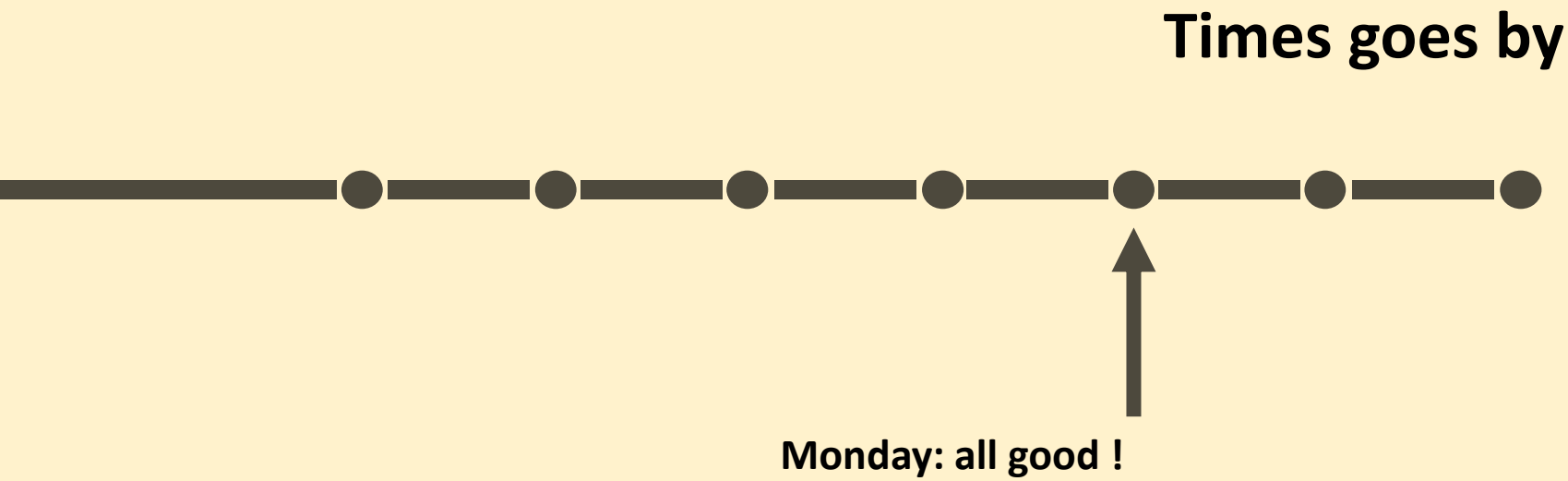
Time travel


476a3f0b: Delete parallelization attempt
5oc78jk8: Upscale analysis to all countries
d289e8ce : Accept foreign crop names

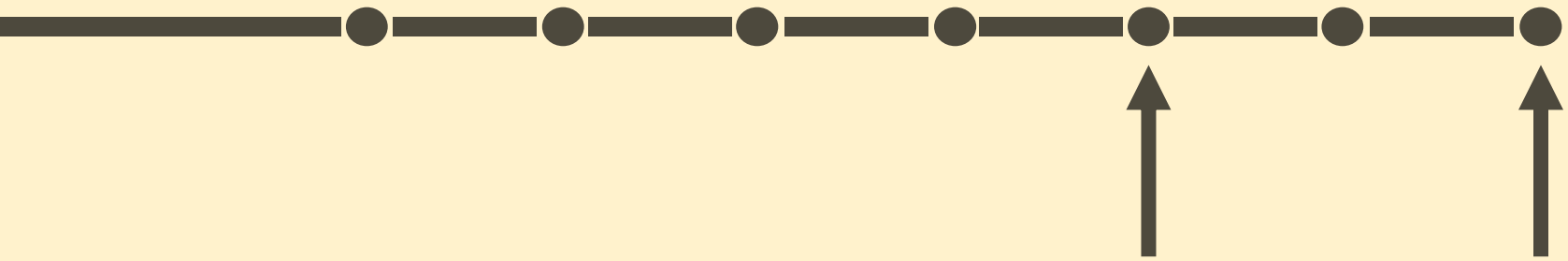
Time travel



Time travel



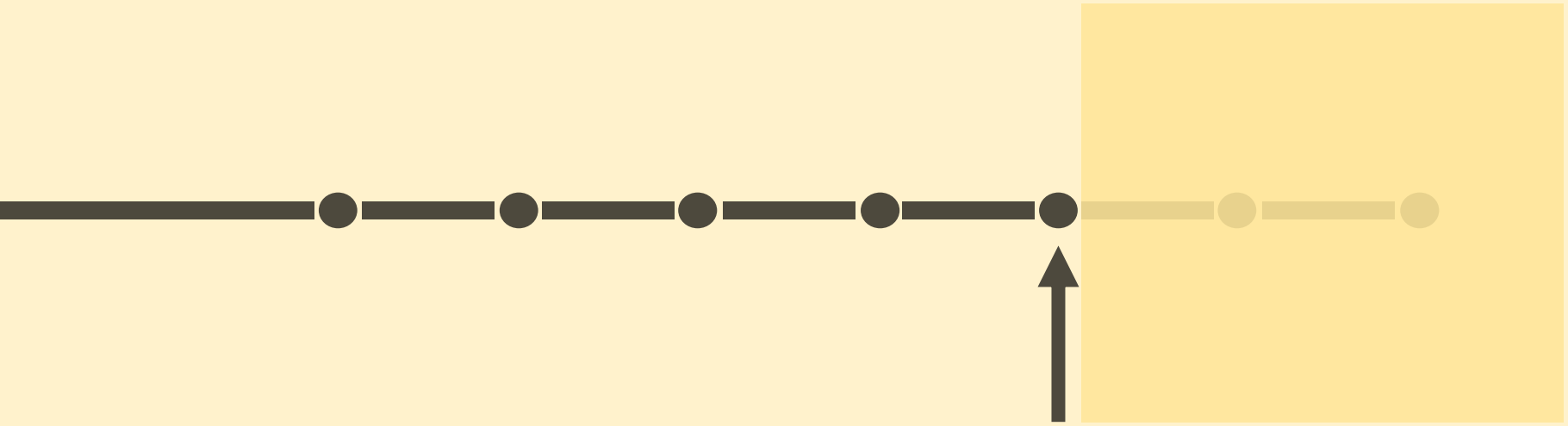
Time travel



Monday: all good !

Wednesday: Meeting with supervisor, Code is broken, I can’t reproduce my results !

Time travel

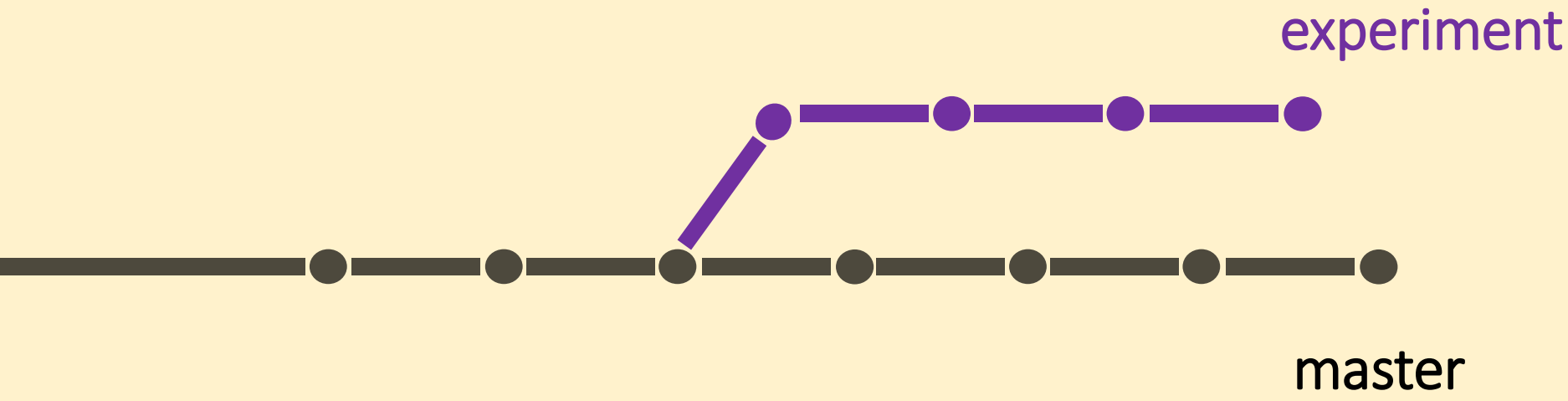


Monday: all good !

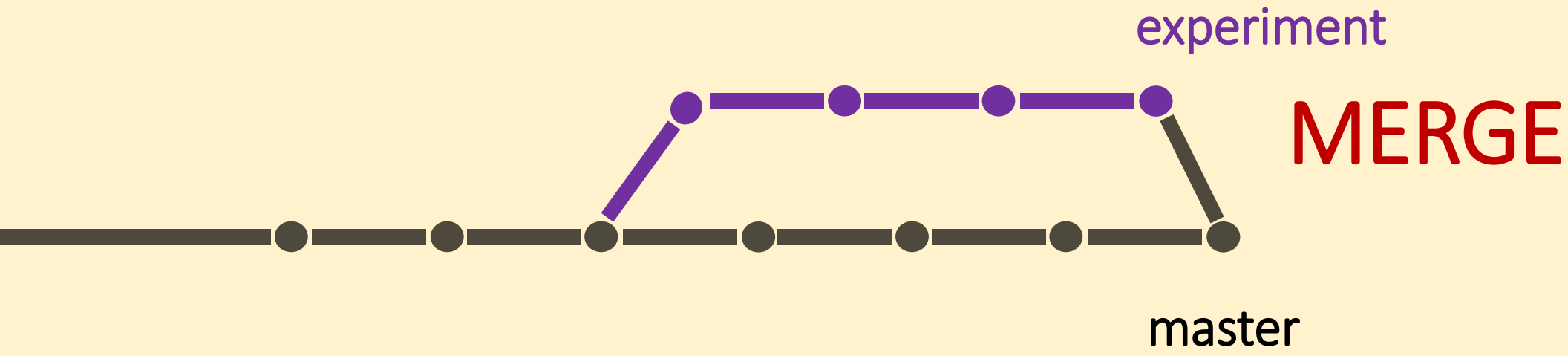
Non-linear development : BRANCH(es)

master

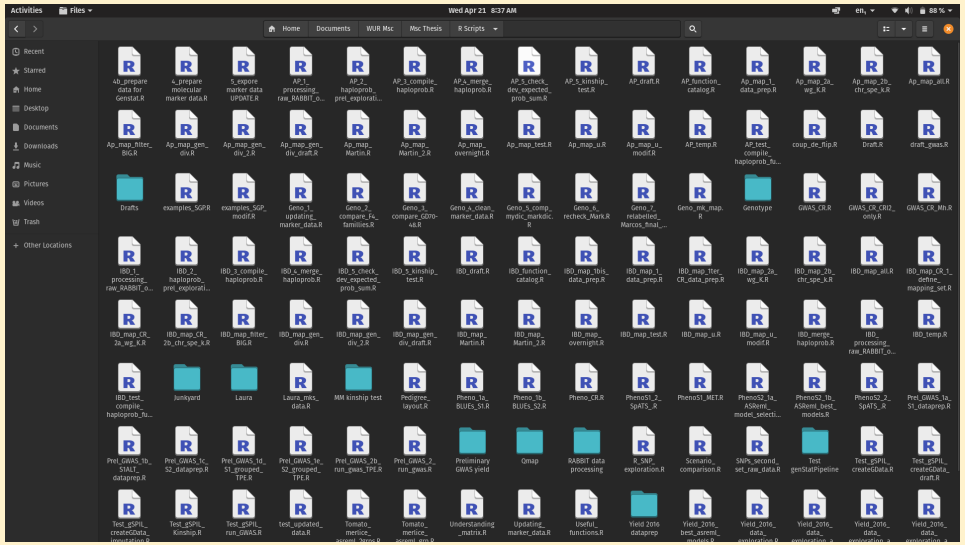
Non-linear development : BRANCH(es)



Non-linear development : BRANCH(es)



Collaborate efficiently with others



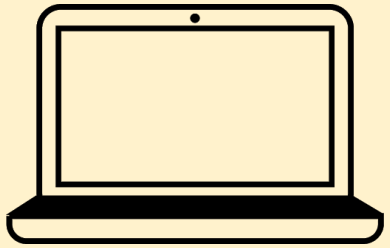
+

X

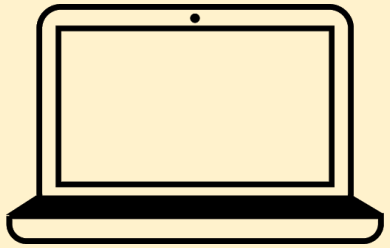
n collaborators

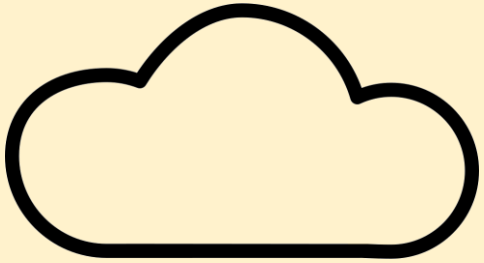
PLEASE NO !

Instead use a REMOTE

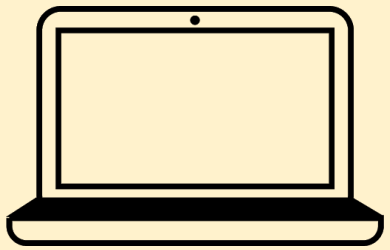


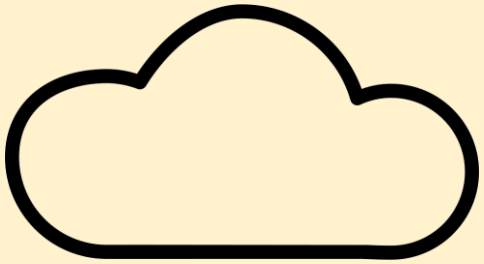
Instead use a REMOTE



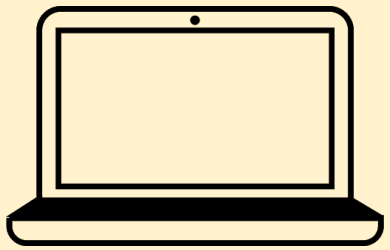


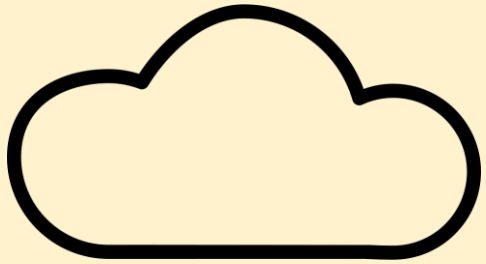
Instead use a REMOTE



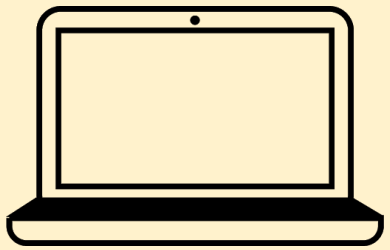


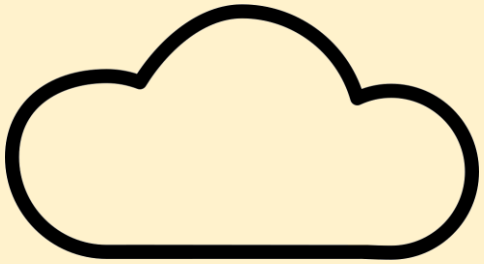
Instead use a REMOTE





Instead use a REMOTE





REPO: github.com/jacinda/save-world.git

Repo owner

CLONE

Repo owner

Commit

Repo owner

RuttjevanOrange01
premie@gov.nl
20/04/2021 – 2.02 pm

Add plans to build dams

We have excellent engineers who can help with rising sea-level.

Repo owner

PUSH

Repo owner

Repo owner

Repo owner

PULL

Repo owner

Disseminate your work

Provided your work is hosted on a public remote:

- Others can use it (no need for a dedicated website)
- Contribute to it

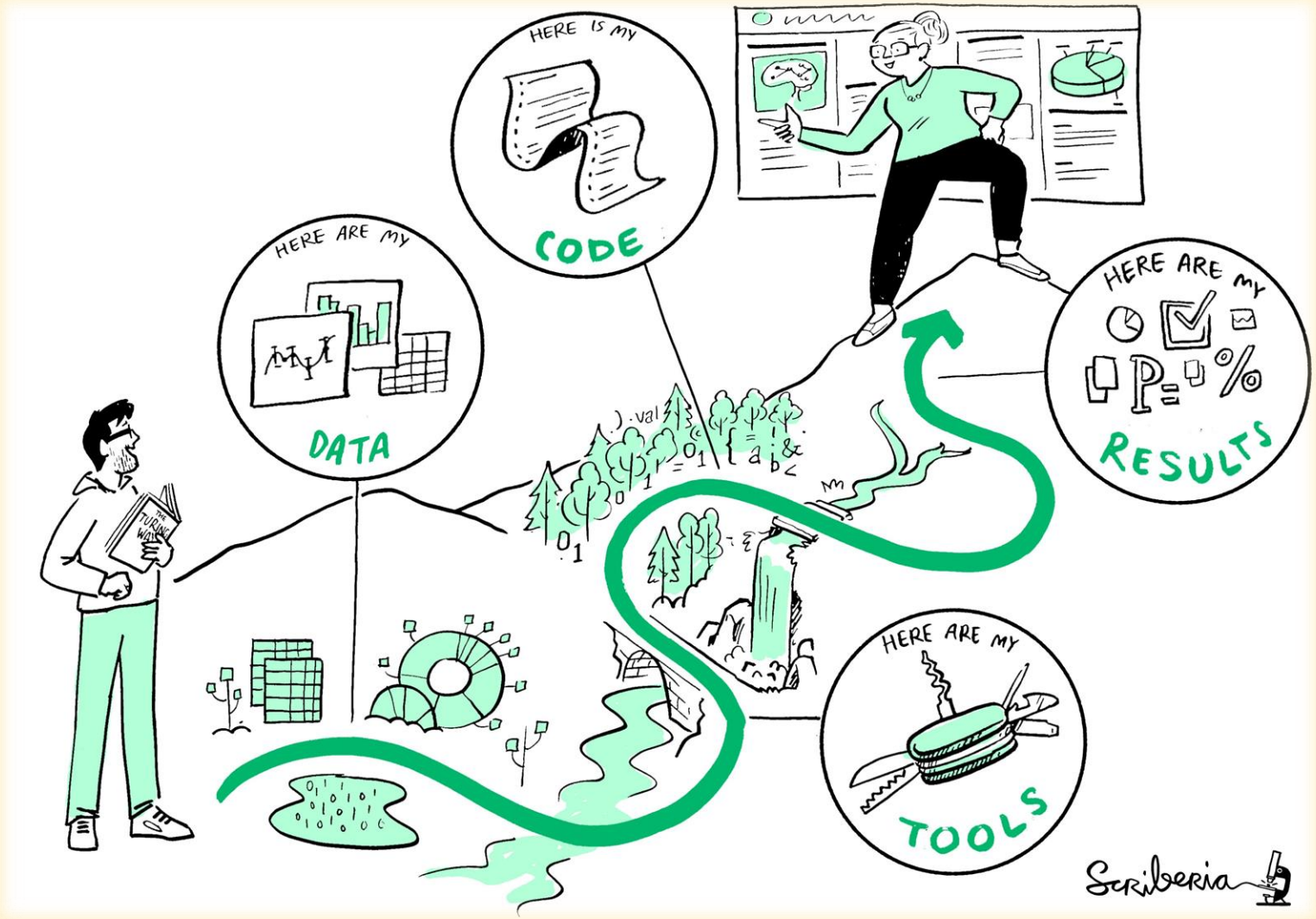
Provided your work is hosted on a public remote:

- Others can use it (no need for a dedicated website)
- Contribute to it

Let’s have a look

Never think you are alone

The Turing way: Guide for Reproducible Research



“Listen young millennial… Your github thingy will not make you Professor, papers will “

5 selfish reasons to work reproducibly

Dr Florian Markowetz
University of Cambridge

SCIENTIFIC DATA

SPRINGER NATURE

Thanks for your indulgence

All credits for slide design goes to @alicebartlett
Speaker Deck: slides