Jiří Moravec

Escape Excel

A quick introduction to programming, R, git and reproducibility

What is this about?

- Project organization – files, folders and data
- Introduction to script-based analyses
- Introduction to pipelines
- Some R and git

What is this NOT about?

- R workshop
- Robust tutorial to programming
- Extensive tutorial to git
  (most of the time, you are using only 3 commands anyway)

Showing you a direction!

This way!

Showing you a direction!

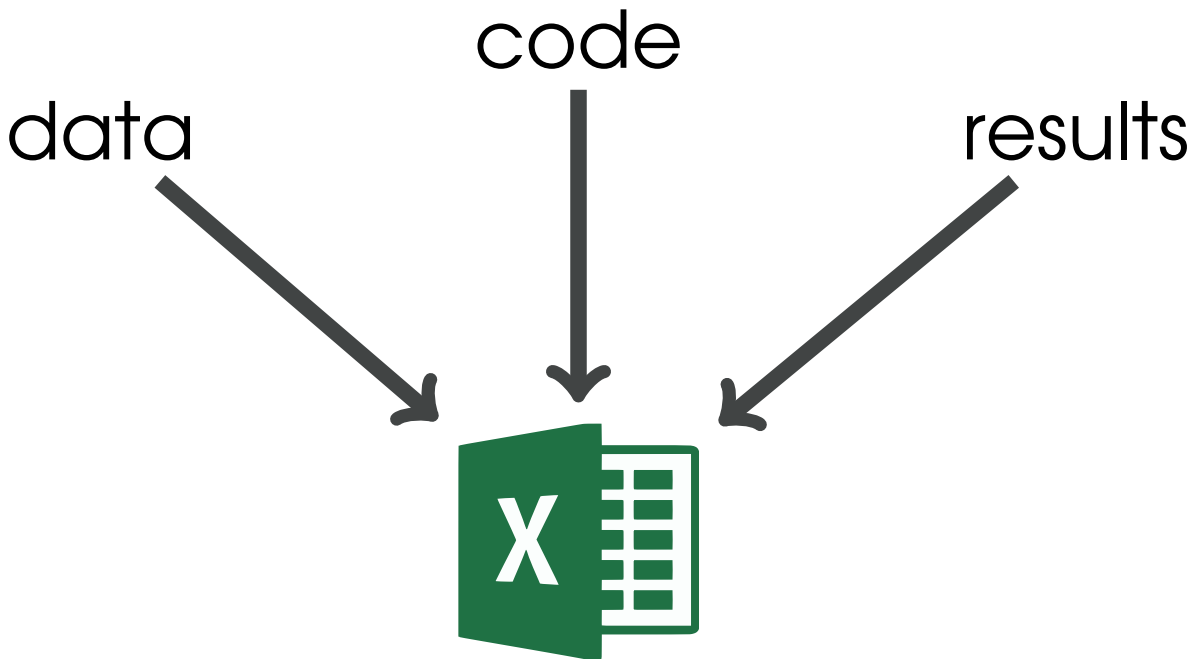
Program:

1. Motivation
2. Project organization
3. Pipelines and script-based analyses
4. Introduction to programming
5. git

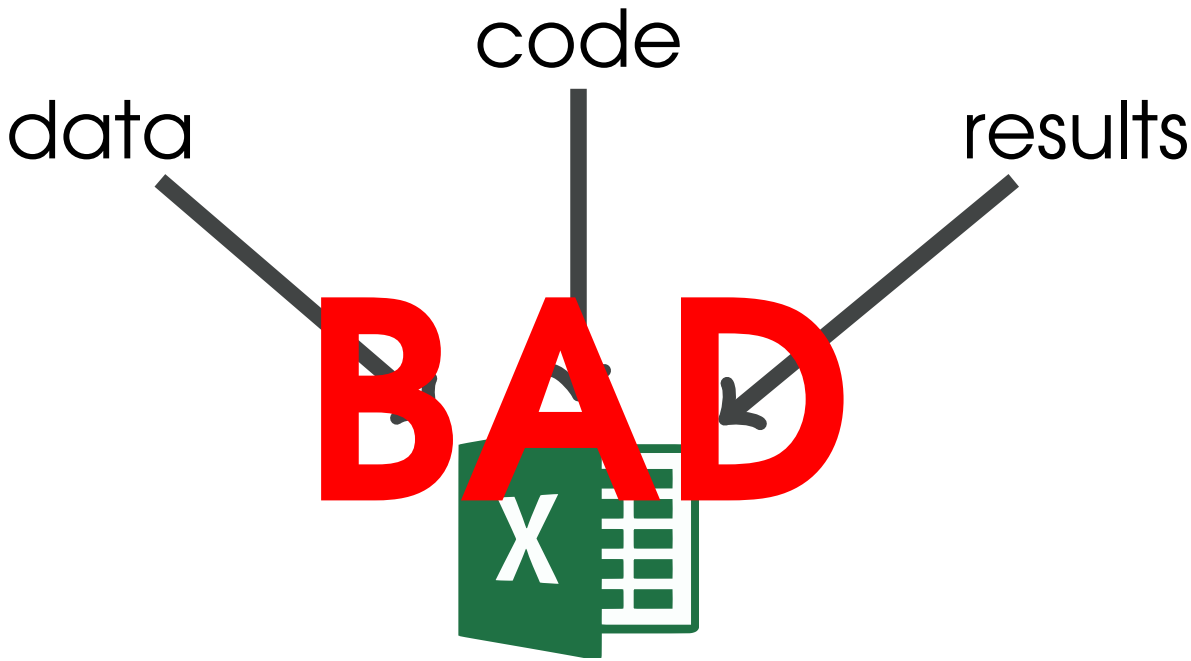
Program:

1. Motivation
2. Project organization
3. Pipelines and script-based analyses
4. Introduction to programming
5. git

Motivation:



Motivation:



Motivation:

analysis.xls

Motivation:

analysis_final.xls

Motivation:

Collaborator_analysis_final.xls

Motivation:

Collaborator_analysis_finalFinal.xls

Motivation:

Motivation:


Data
Scripts
Results
Intermediaries
Documentation


makefile
run.sh
README.md

Program:

1. Motivation
2. Project organization
3. Pipelines and script-based analyses
4. Introduction to programming
5. git

Project organization

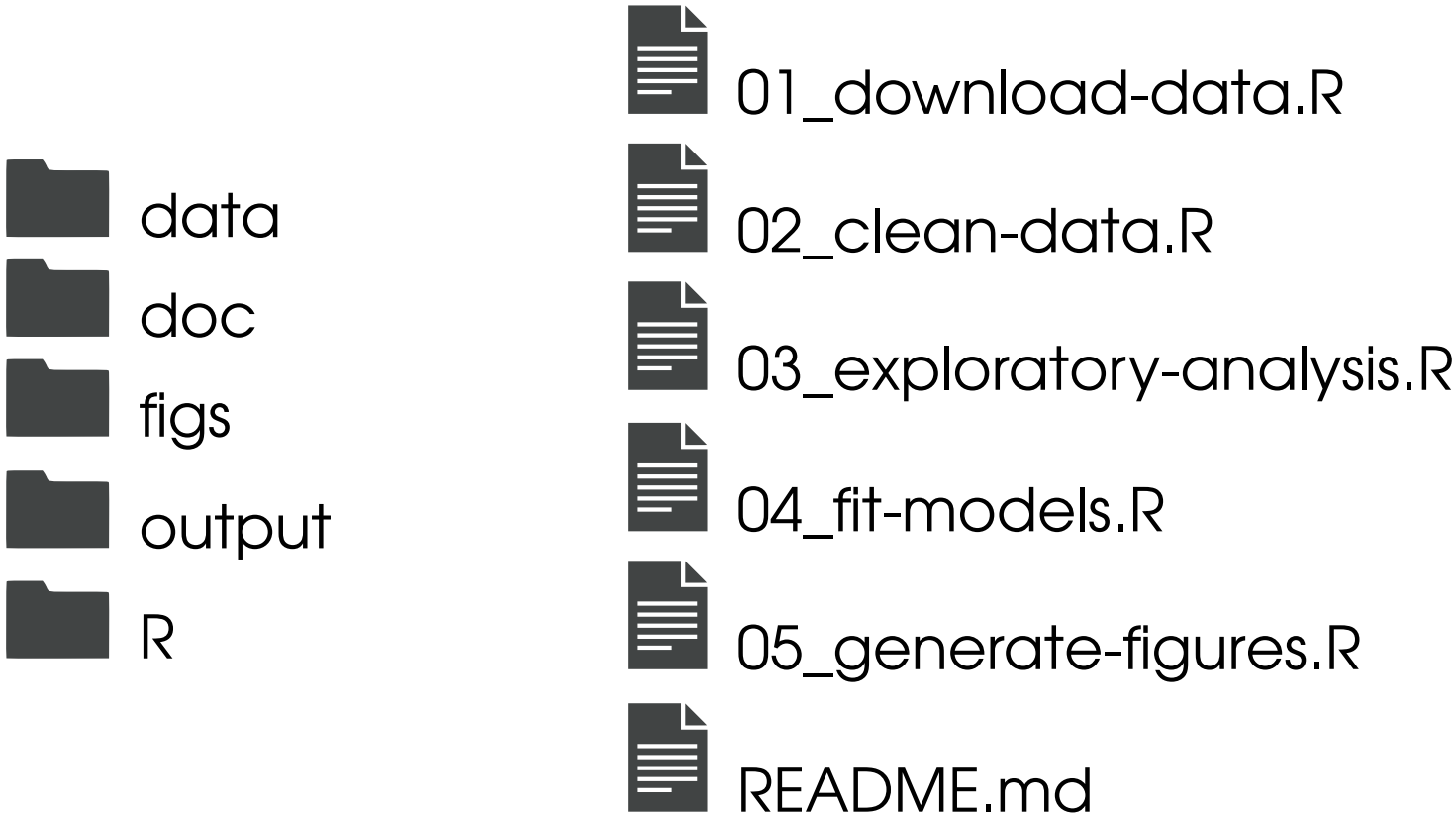
Make sure you have:

1. a single folder for entire project
2. reasonable naming scheme
3. good folder structure

Project organization

Project organization

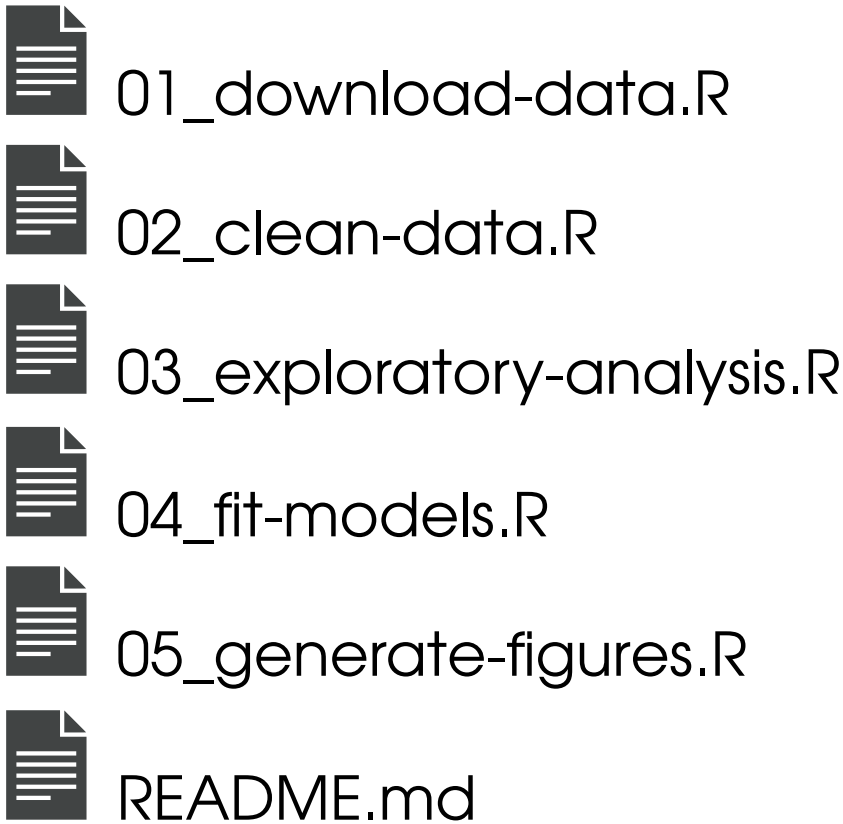
From *A Guide to Reproducible Code in Ecology and Evolution*



Project organization

From *A Guide to Reproducible Code in Ecology and Evolution*

Notice the names!

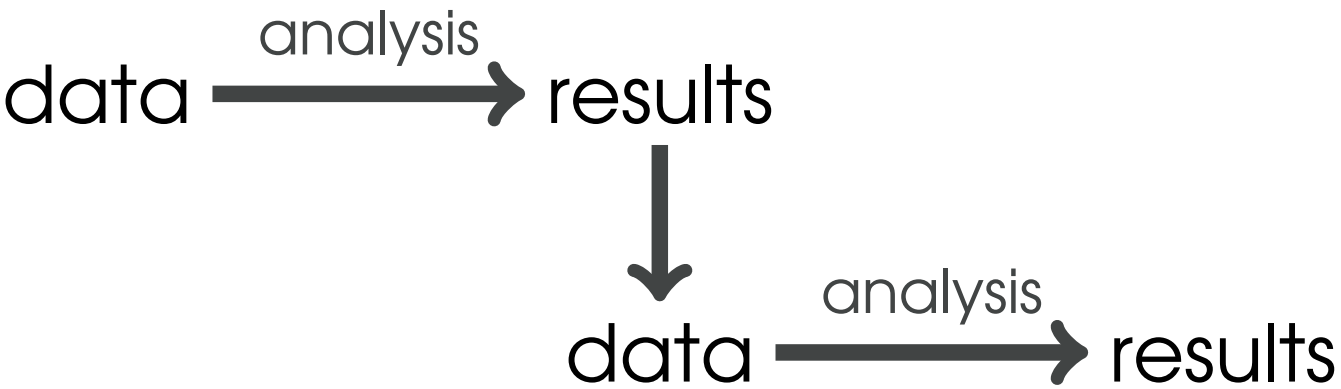


Program:

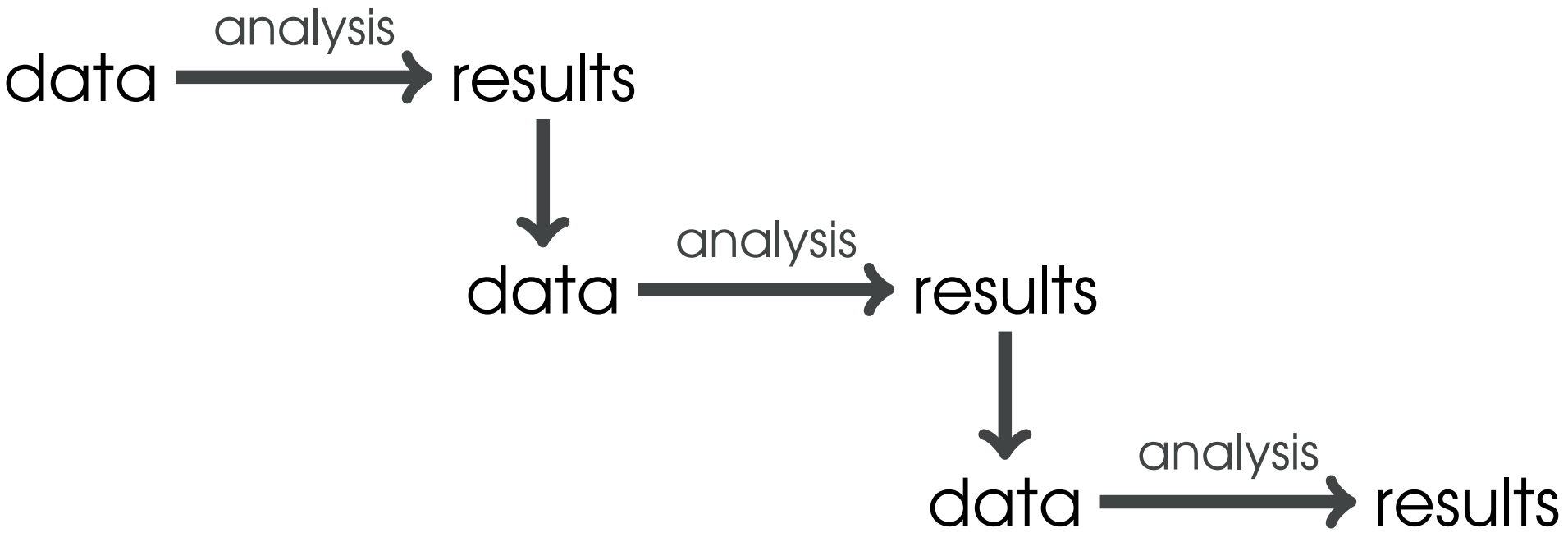
1. Motivation
2. Project organization
3. Pipelines and script-based analyses
4. Introduction to programming
5. git

Pipelines

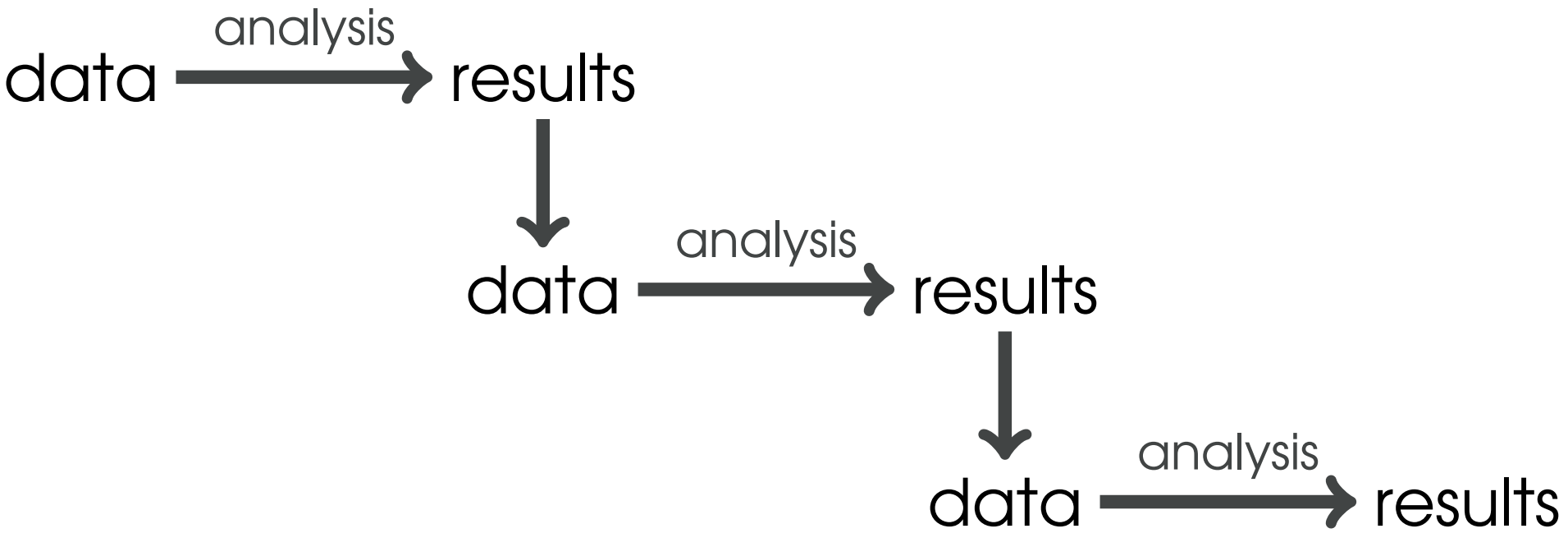
Pipelines



Pipelines

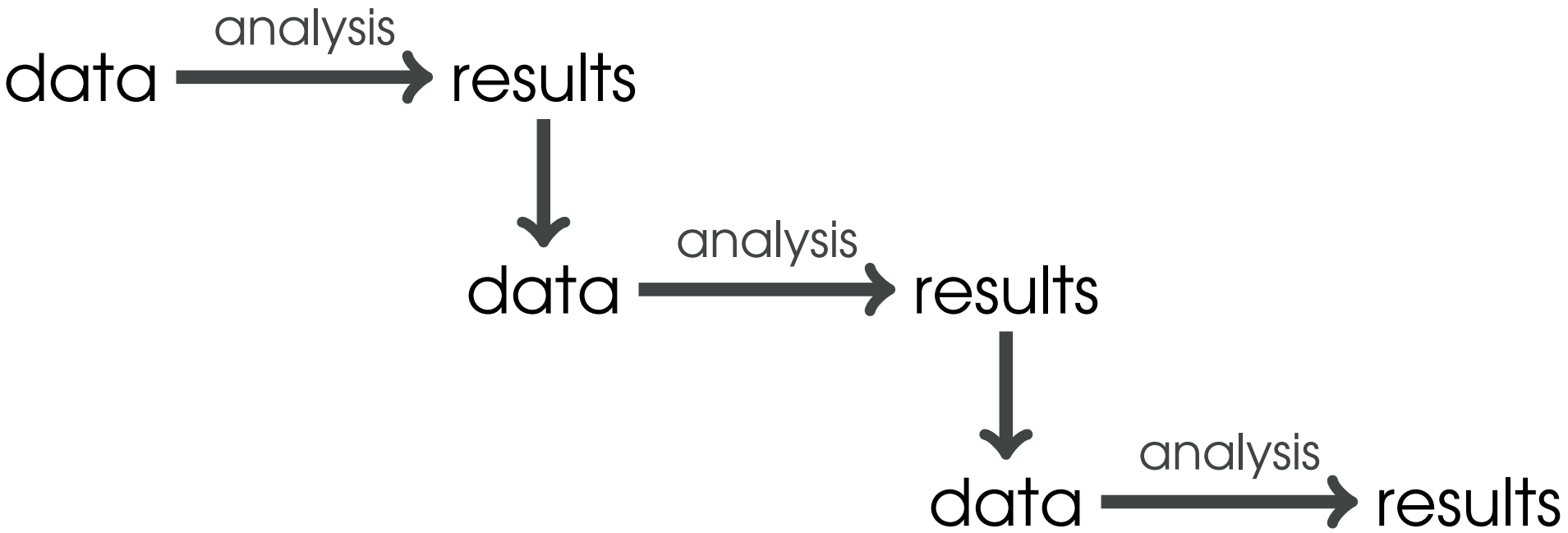

data
analysis
results
data
analysis
results
data
analysis
results

Pipelines



Functions, Scripts, smaller or bigger analyses

Pipelines



Functions, Scripts, smaller or bigger analyses

Script-based analyses

- repeatability
- documented process
- (including any data modification and filtering)
- smaller effort to modify
- can be repeated by just running the run script

Program:

1. Motivation
2. Project organization
3. Pipelines and script-based analyses
4. Introduction to programming
5. git

Introduction to programming

Why R?

- easy to get data into R
- closer to statistics than other programming languages
- graphical capabilities
- huge library of packages for essentially anything
- FREE and OPEN SOURCE

Introduction to programming

Helpful tips:

- Scripts:
  - self-contained
  - doing one thing well rather than many things badly
- Functions:
  - self-contained
  - small
  - use them instead of repeating code
  - use them for a well-defined tasks and concepts
  - documentation and comments
- Data:
  - text human readable files
  - not binary proprietary formats
  - modify with scripts
  - never overwrite

Helpful tips (cont):

- Paths:
  - use relative rather than absolute
  - relative `data\my_data.csv`
  - absolute: `c:\my_projects\data\my_data.csv`
- Figures:
  - generate with scripts
  - easier to modify
  - easier to change once data or analysis changes

Live coding excercise!

Program:

1. Motivation
2. Project organization
3. Pipelines and script-based analyses
4. Introduction to programming
5. git

Git is:

- version control system
- track all changes in text files (data, code)
- tracks where the changes were done
- FREE and OPEN-SOURCE

Git commands:

- `git init`
- `git clone <path or url>`
- `git add file1 file2` or `git add -A`
- `git commit -m "Information about commit"`
- `git push remote_repository branch`
- `git pull remote_repository branch`
- `git status`

Sources and guides

R

- Start with: `https://www.statmethods.net/r-tutorial/index.html`
- Continue with: `http://tryr.codeschool.com`
- Advanced R: `http://adv-r.had.co.nz/`

git

- Start with: `https://guides.github.com/introduction/git-handbook/`
- Documentation: `https://git-scm.com/docs`
- Book: `https://git-scm.com/book/en/v2`

A Guide to Reproducible Code in Ecology and Evolution

- `https://tinyurl.com/reproducibleCode`
- original link

I want your feedback!