Data Analysis and Visualization in R

Instructor: Philip (Pip) Grylls

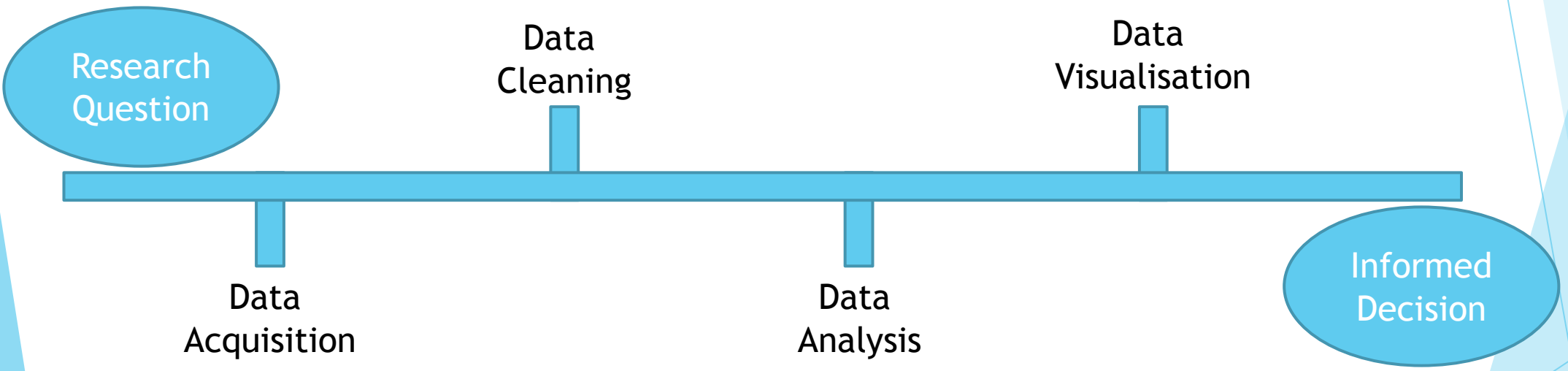
Helpers: Philly Broadbent

Resources and Acknowledgements

- Training material for this module:
  - https://southampton-rsg.github.io/data-analysis-and-visualisation/
- Based on the Data Carpentry curriculum:
  - https://datacarpentry.org./
- The Carpentries worldwide community:
  - Teaching foundational computational and data analysis skills to researchers
  - Software, Data and Library Carpentry
  - https://carpentries.org

Typical ‘data’ journey

Aspects of a ‘data analytics pipeline’



Learning Objectives

- The working environment
- The basics of R
- Loading Data into R

Working environment

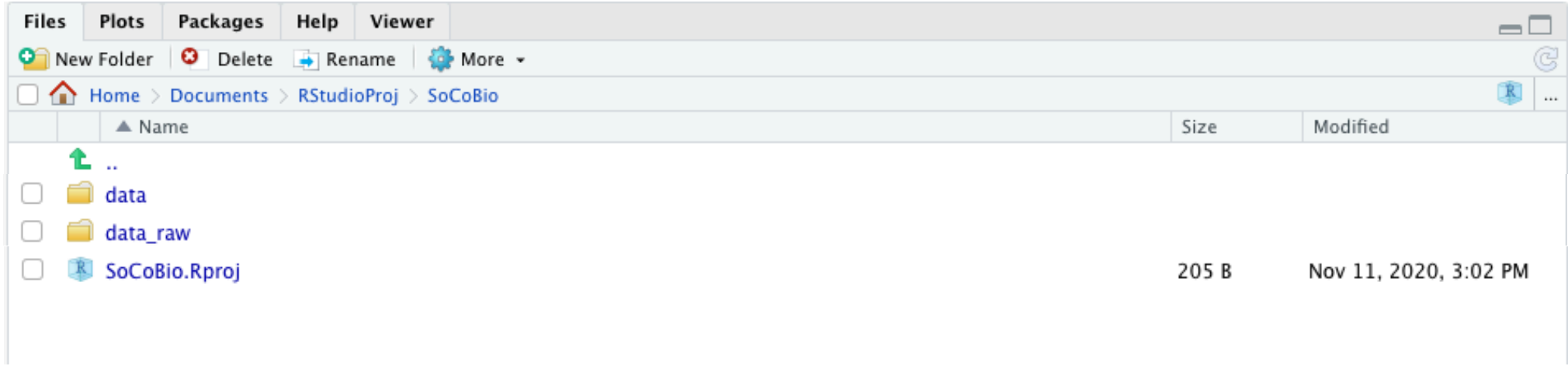
What is RStudio

- RStudio is an integrated development environment (IDE)

  It contains:

  - An editor: A place to edit scripts and texts.
  - Consoles: Where commands and scripts can be run.
  - Environment inspector: Shows loaded modules, current data, and other relevant system information.
  - A system file viewer/help: Shows the file structure, generated plots, info about r programs e.t.c.

Working environment

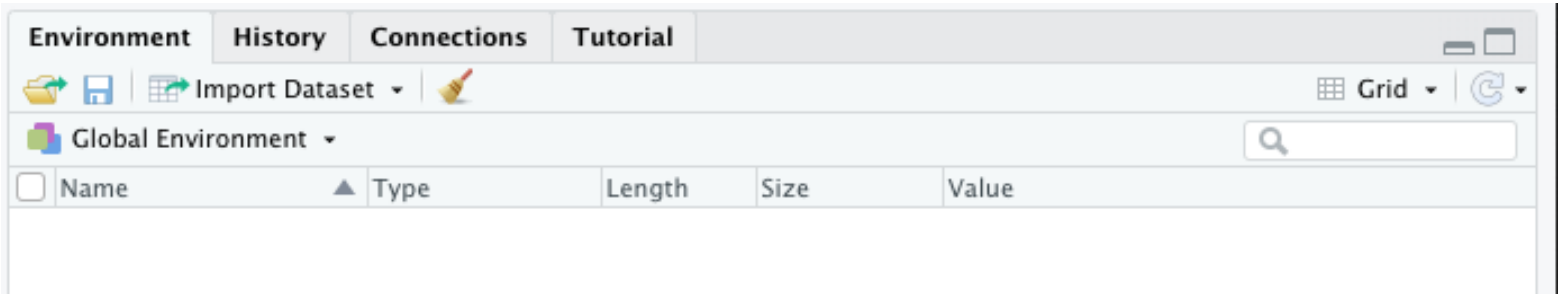


Main directory: This is where you will keep scripts and have sub-directories for keeping raw data and data outputs.

.Rprog file: A good indicator of where your 'Main directory' is.

Working environment

Console/Terminal/Jobs:
Run individual code,
access the terminal (bash shell),
and running jobs.



Environment/History/Connections/Tutorial:
A file system viewer, a history of input commands,
A place to manage connections to external data, and a place to access tutorials.
Other programs may add other tabs here, for example RStudio will pick up git version control.

The basics of R

What is R

- "R is a free software environment for statistical computing and graphics." – *r-project.org*
  - R can be used to load, clean, and analyse data.
  - R can be used to create graphics flexible enough for exploration of data which look professional enough to publish in a journal.
- R is open source
  - Anyone can view, and suggest changes to the core of R
  - We (the R community) can check that the programs operate as intended.

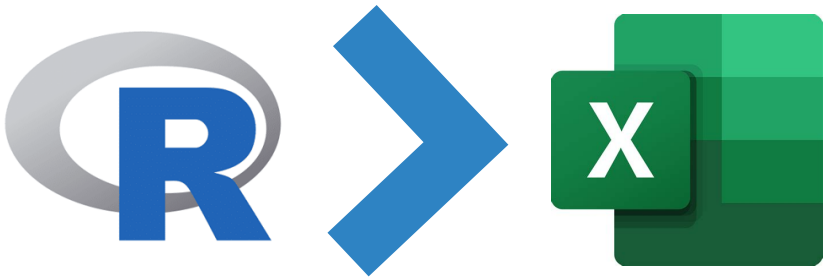
The basics of R

We need to learn:

- How to create objects.
- How to inspect and name objects.
- What comments are and why/how to write them.
- What functions are and why we use them.
- What vectors are and how to use them.

Loading data into R

Why R?



- I know all of these powerful spreadsheet tools why should I use R?
  - R can load **LARGE** amounts of data.
  - R scripts can be re-run using the same or more data to reproduce or reenforce data.
- Where do I put my data?
  - Load CSVs into data frames.
  - Inspect data frames to get views of data too large for spreadsheeting tools.
  - Extract and format portions of the data frame using indexing.

Good luck and have fun.

Remember to use the teams groups to help each other and reach out to us for help!