

Scientific Reproducibility in Biology Research

Day 2

Terry Neeman

Australian National University

2023

Goal: Create Reproducible and Transparent Workflows

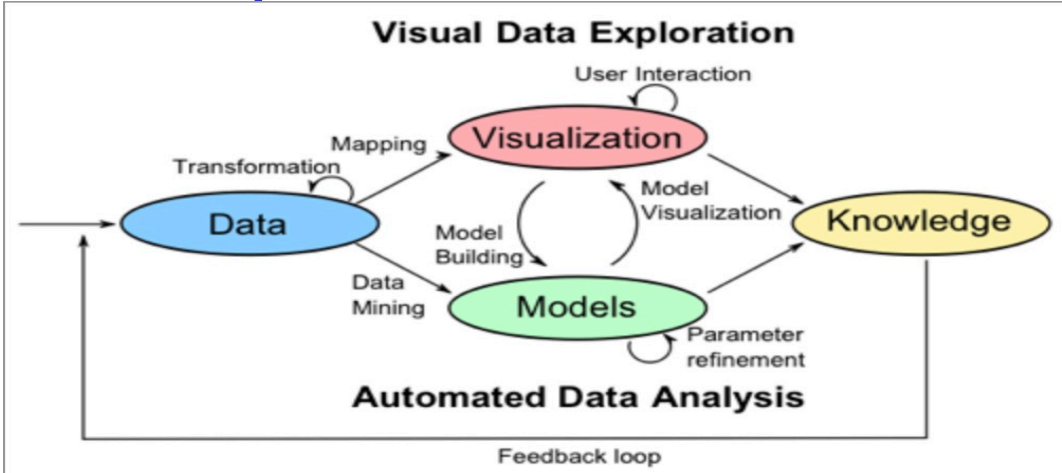
- ▶ Clear directory/file structures
- ▶ FAIR data
- ▶ Good coding practices
 - ▶ well organised (good code hygiene)
 - ▶ well annotated
 - ▶ small code chunks between annotation
- ▶ Automated processes

The FAIR principles

- ▶ FAIR data
 - ▶ Findable, **A**ccessible, **I**nteroperable, **R**euseable
 - ▶ Meta-data
 - ▶ Experimental Design
 - ▶ Data Dictionary
- ▶ FAIR software



Automated processes



Import Raw or “Minimally Manipulated” Data

```
library(tidyverse)
library(readxl)
cancer_df <- read_xlsx("data/Cuzick_2010.xlsx",
                      sheet = 1,
                      skip = 0,
                      range = "A1:D100")
```

What we'll cover today

- ▶ Good coding practice: the “pipe” `%>%`
- ▶ Data wrangling using `dplyr::` (tidyverse)
 - ▶ changing data types
 - ▶ subsetting
 - ▶ filter by rows
 - ▶ select columns
 - ▶ sorting
 - ▶ create new columns
 - ▶ summarise by groups
- ▶ Exploratory analysis using `ggplot2::` (tidyverse)