Scientific Reproducibility in Biology Research

Day 3

Terry Neeman

Australian National University

2023

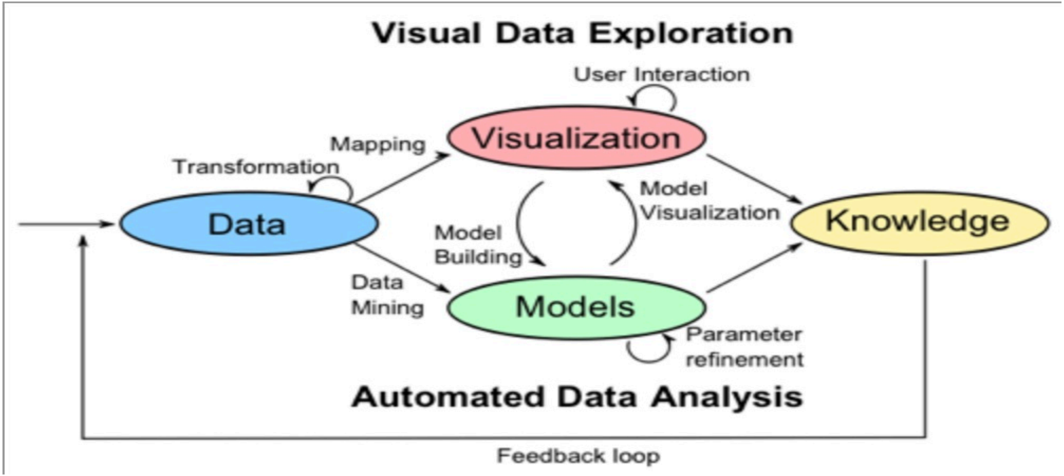
Goal: Create Reproducible and Transparent Workflows

- 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

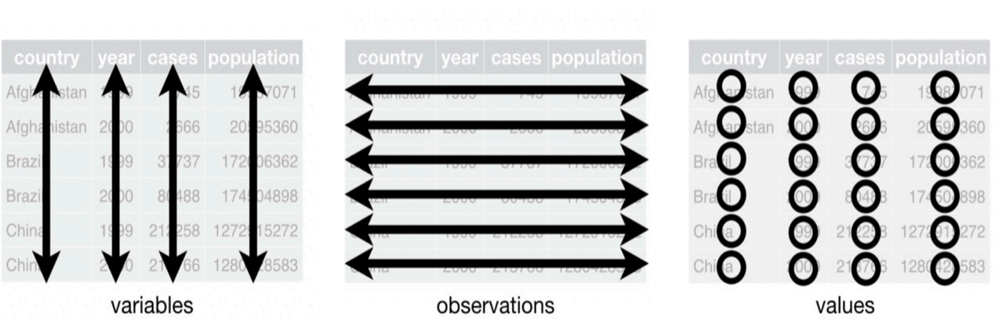
- FAIR data
  - **F**indable, **A**ccessible, **I**nteroperable, **R**euseable
  - Meta-data
    - Experimental Design
    - Data Dictionary
- FAIR software

Automated processes



Tidy Data: making data easy to work with

https://r4ds.had.co.nz/tidy-data.html#tidy-data-1



Tidy Data: making data easy to work with

- Is this tibble “tidy”?

```
## # A tibble: 6 x 4
##   country      year  cases population
##   <chr>       <int>  <int>      <int>
## 1 Afghanistan  1999    745   19987071
## 2 Afghanistan  2000   2666   20595360
## 3 Brazil       1999  37737  172006362
## 4 Brazil       2000  80488  174504898
## 5 China        1999 212258 1272915272
## 6 China        2000 213766 1280428583
```

Tidy Data: making data easy to work with

- Is this tibble “tidy”?

```
## # A tibble: 12 x 4
##    country      year type            count
##    <chr>       <int> <chr>           <int>
##  1 Afghanistan  1999 cases             745
##  2 Afghanistan  1999 population   19987071
##  3 Afghanistan  2000 cases            2666
##  4 Afghanistan  2000 population   20595360
##  5 Brazil       1999 cases           37737
##  6 Brazil       1999 population  172006362
##  7 Brazil       2000 cases           80488
##  8 Brazil       2000 population  174504898
```

Tidy Data: making data easy to work with

- Is this tibble “tidy”?

```
## # A tibble: 6 x 3
##   country      year rate
## * <chr>       <int> <chr>
## 1 Afghanistan  1999 745/19987071
## 2 Afghanistan  2000 2666/20595360
## 3 Brazil       1999 37737/172006362
## 4 Brazil       2000 80488/174504898
## 5 China        1999 212258/1272915272
## 6 China        2000 213766/1280428583
```

Tidy Data: making data easy to work with

- How about separating into 2 tibbles?

```
table4a
```

```
## # A tibble: 3 x 3
##   country     `1999` `2000`
## * <chr>        <int>  <int>
## 1 Afghanistan    745   2666
## 2 Brazil       37737  80488
## 3 China       212258 213766
```

```
table4b
```

```
## # A tibble: 3 x 3
##   country         `1999`     `2000`
## * <chr>            <int>      <int>
## 1 Afghanistan   19987071   20595360
## 2 Brazil       172006362  174504898
## 3 China       1272915272 1280428583
```

Pivoting data: pivot_longer() and pivot_wider()

```
table4a_tidy <- table4a %>%
pivot_longer(cols = c(`1999`,`2000`), names_to = "year",values_to = "cases
table4a_tidy
```

```
## # A tibble: 6 x 3
##   country     year   cases
##   <chr>       <chr>  <int>
## 1 Afghanistan 1999     745
## 2 Afghanistan 2000    2666
## 3 Brazil      1999   37737
## 4 Brazil      2000   80488
## 5 China       1999  212258
## 6 China       2000  213766
```

Pivoting data: pivot_longer() and pivot_wider()

```
table4b_tidy <- table4b %>%
pivot_longer(cols = c(`1999`,`2000`), names_to = "year",values_to = "popula
table4b_tidy
```

```
## # A tibble: 6 x 3
##   country     year  population
##   <chr>       <chr>      <int>
## 1 Afghanistan 1999    19987071
## 2 Afghanistan 2000    20595360
## 3 Brazil      1999   172006362
## 4 Brazil      2000   174504898
## 5 China       1999  1272915272
## 6 China       2000  1280428583
```

Merging data

```
left_join(table4a_tidy,table4b_tidy, by = c("country","year"))
```

```
## # A tibble: 6 x 4
##   country     year   cases population
##   <chr>       <chr>  <int>      <int>
## 1 Afghanistan 1999     745   19987071
## 2 Afghanistan 2000    2666   20595360
## 3 Brazil      1999   37737  172006362
## 4 Brazil      2000   80488  174504898
## 5 China       1999  212258 1272915272
## 6 China       2000  213766 1280428583
```

What we'll cover today

- Using Rmarkdown
- Review Day 2 assignment
- Data wrangling
  - pivot_longer(), pivot_wider()
  - collapsing factor levels
- Explore rat alcohol exposure paper