BIOL 3207/6207

19 September 2022

Recap of BIOL 3207/6207

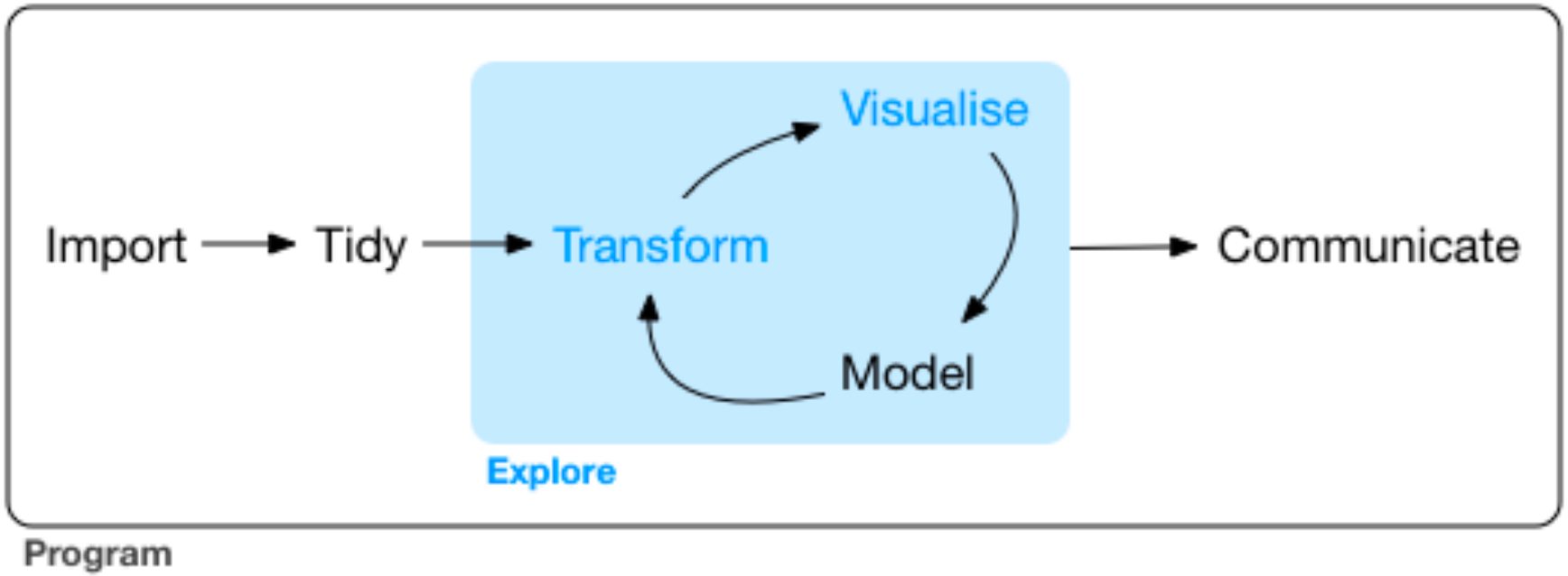
PART 1

2. Null Hypothesis Significance Testing and Exploratory Data Analysis
3. Computational Statistics
4. Understanding and Visualising Variation
5. Data Visualisation and Visual Inference
6. Using Git and GitHub

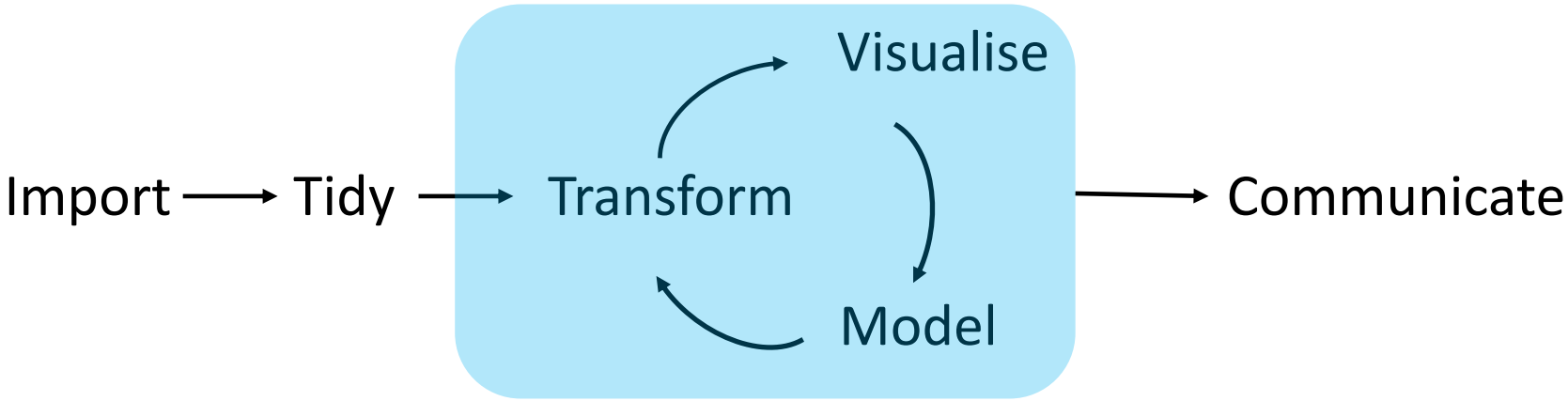
PART 2

7. Data Wrangling
8. Statistical Models
9. Meta-Analysis I
10. Meta-Analysis II
11. Dimension Reduction and Visualisation

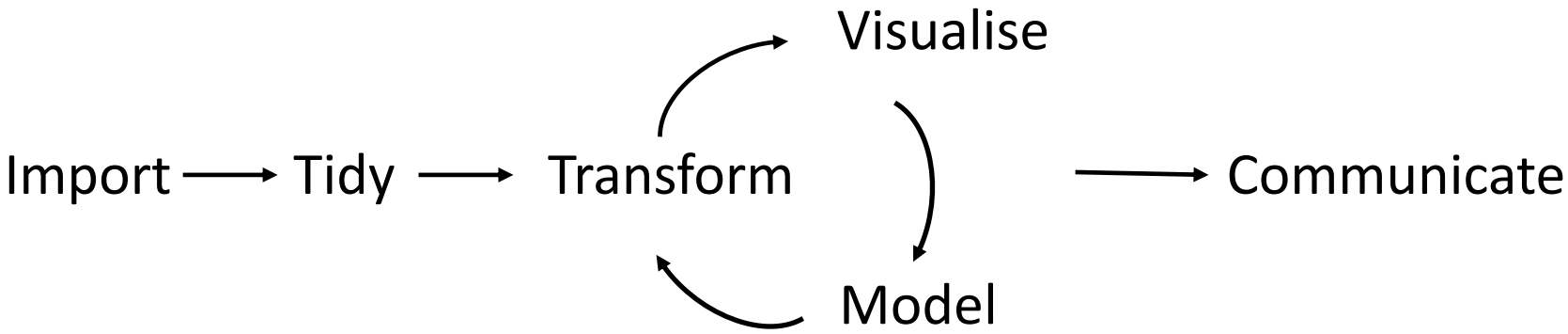
From R4DS, Chapter 2:

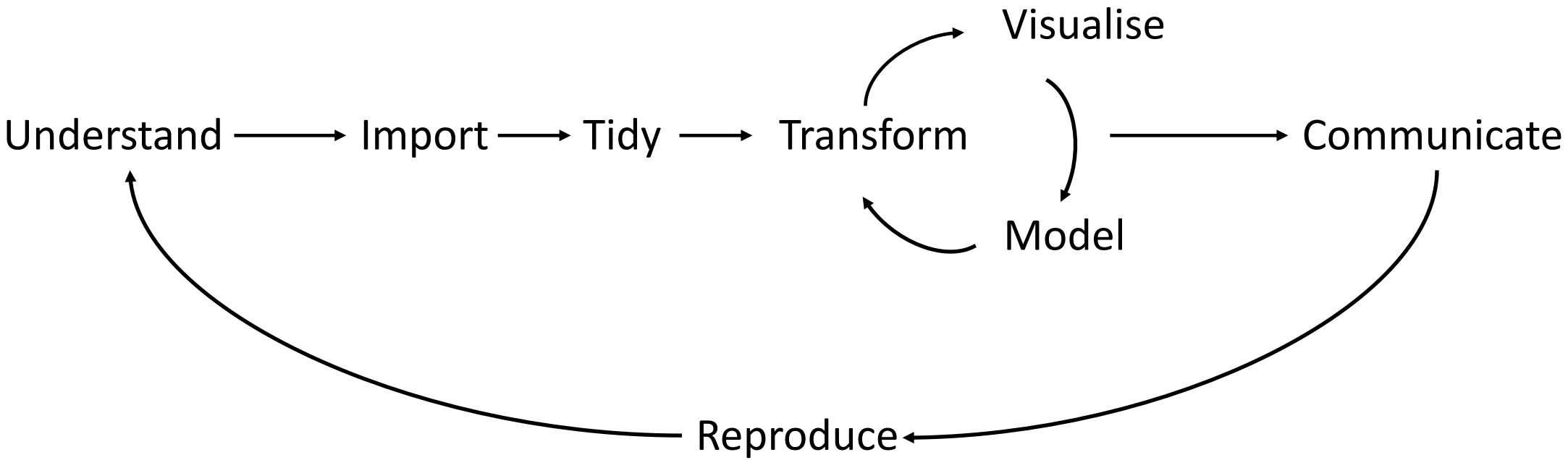


From R4DS, Chapter 2:

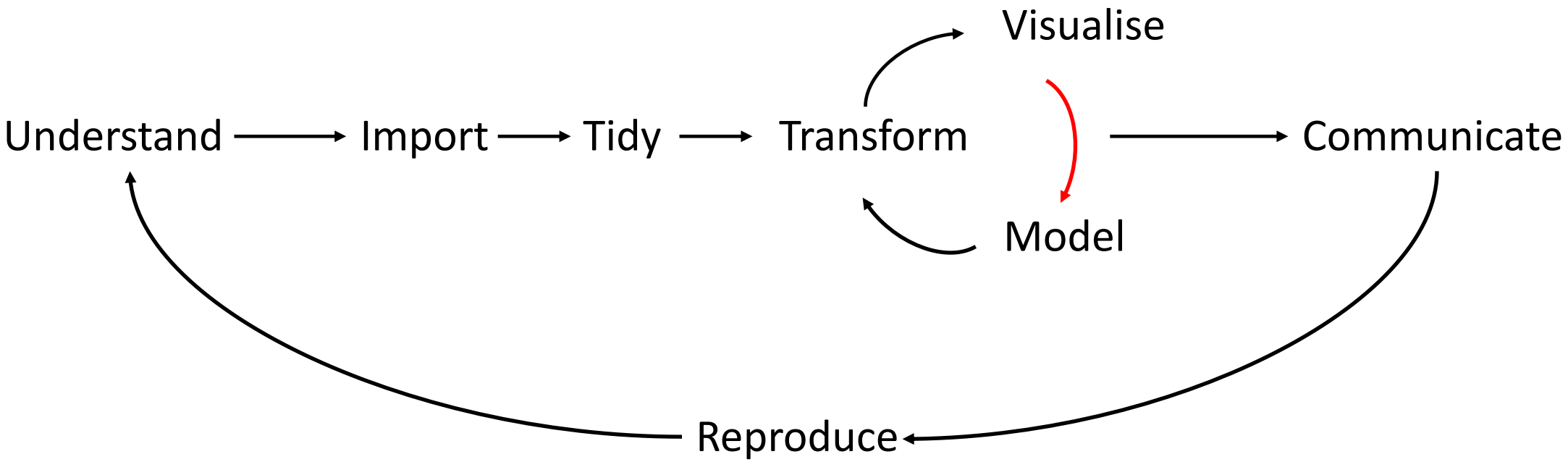


From R4DS, Chapter 2:

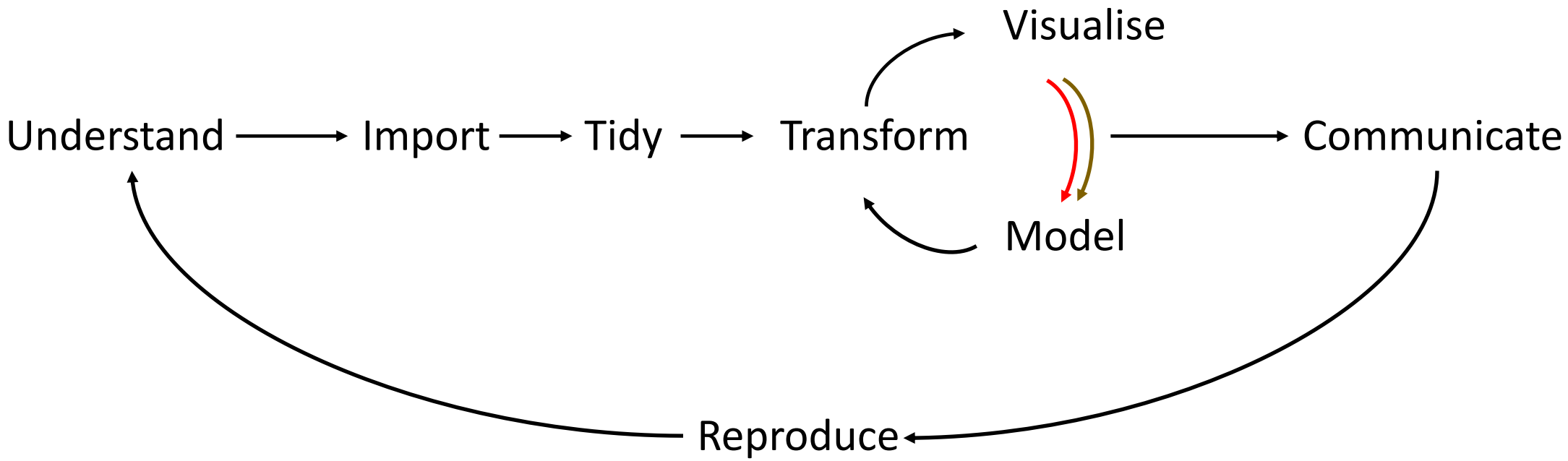



Visualise
Understand
Import
Tidy
Transform
Communicate
Model
Reproduce

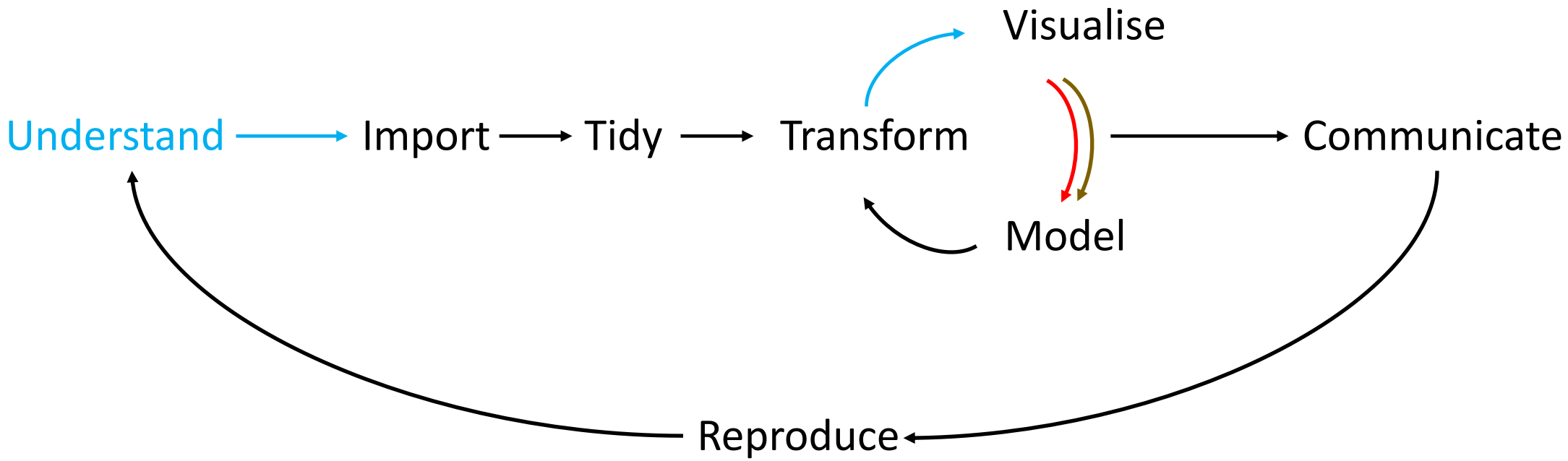
2. Null Hypothesis Significance Testing and Exploratory Data Analysis



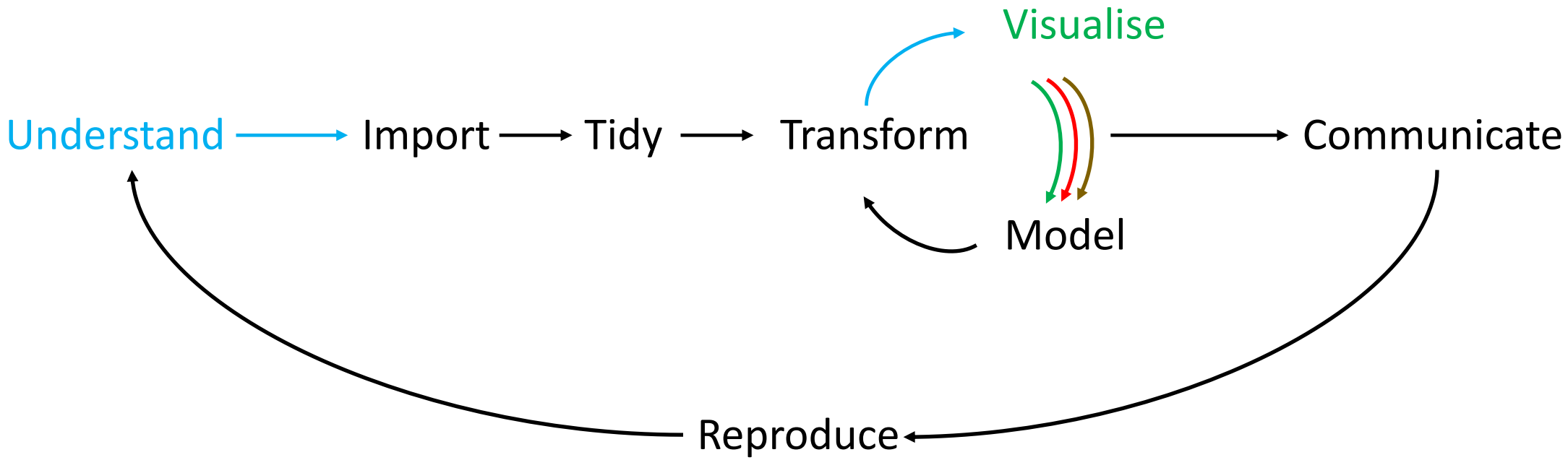
2. Null Hypothesis Significance Testing and Exploratory Data Analysis
3. Computational Statistics



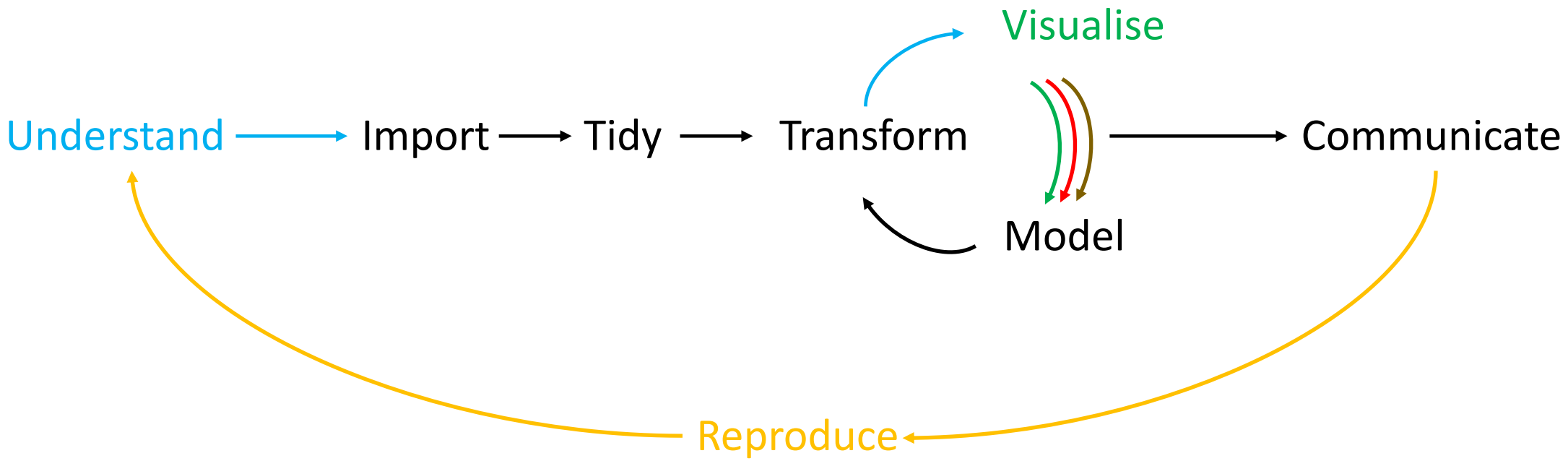
2. Null Hypothesis Significance Testing and Exploratory Data Analysis
3. Computational Statistics
4. Understanding and Visualising Variation



2. Null Hypothesis Significance Testing and Exploratory Data Analysis
3. Computational Statistics
4. Understanding and Visualising Variation
5. Data Visualisation and Visual Inference

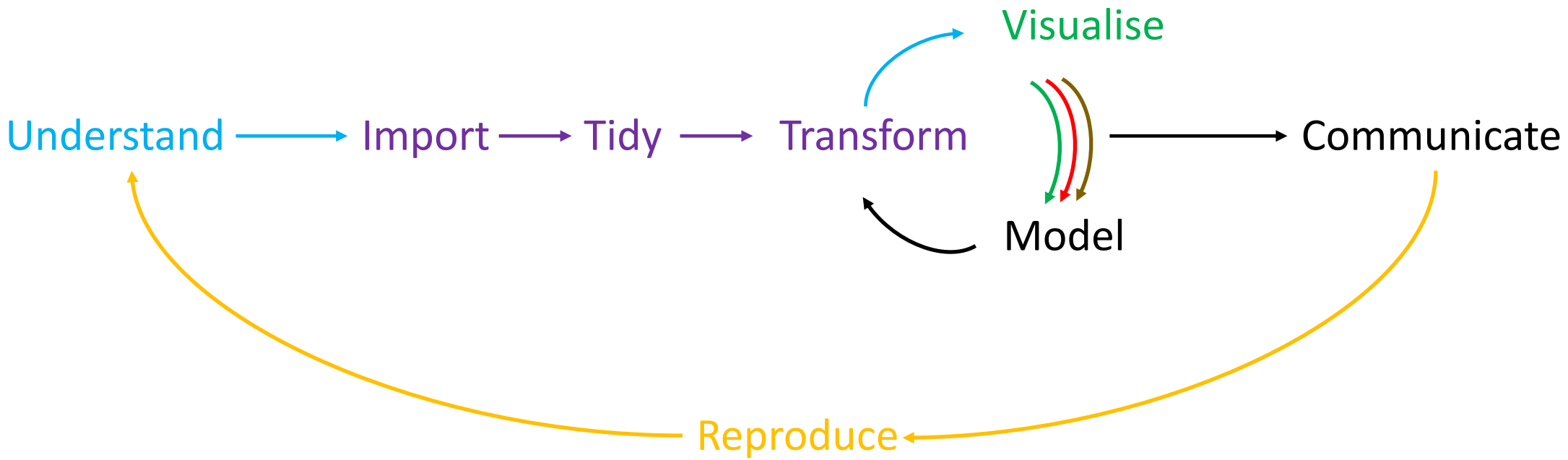


2. Null Hypothesis Significance Testing and Exploratory Data Analysis
3. Computational Statistics
4. Understanding and Visualising Variation
5. Data Visualisation and Visual Inference
6. Using Git and GitHub



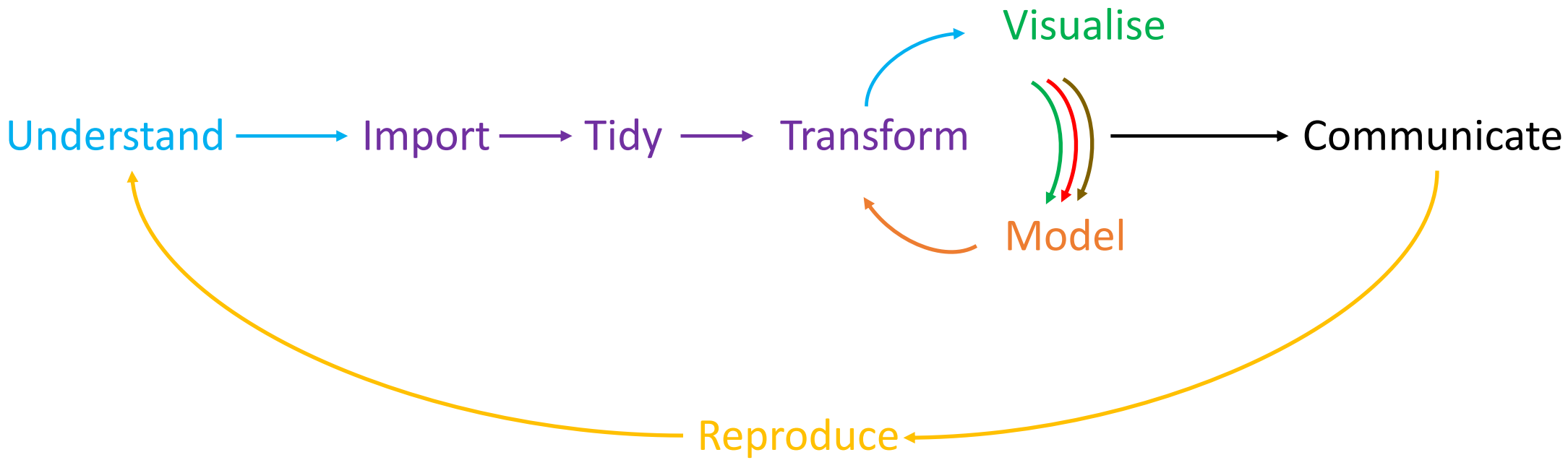
2. Null Hypothesis Significance Testing and Exploratory Data Analysis
3. Computational Statistics
4. Understanding and Visualising Variation
5. Data Visualisation and Visual Inference
6. Using Git and GitHub

7. Data Wrangling **(THIS WEEK)**



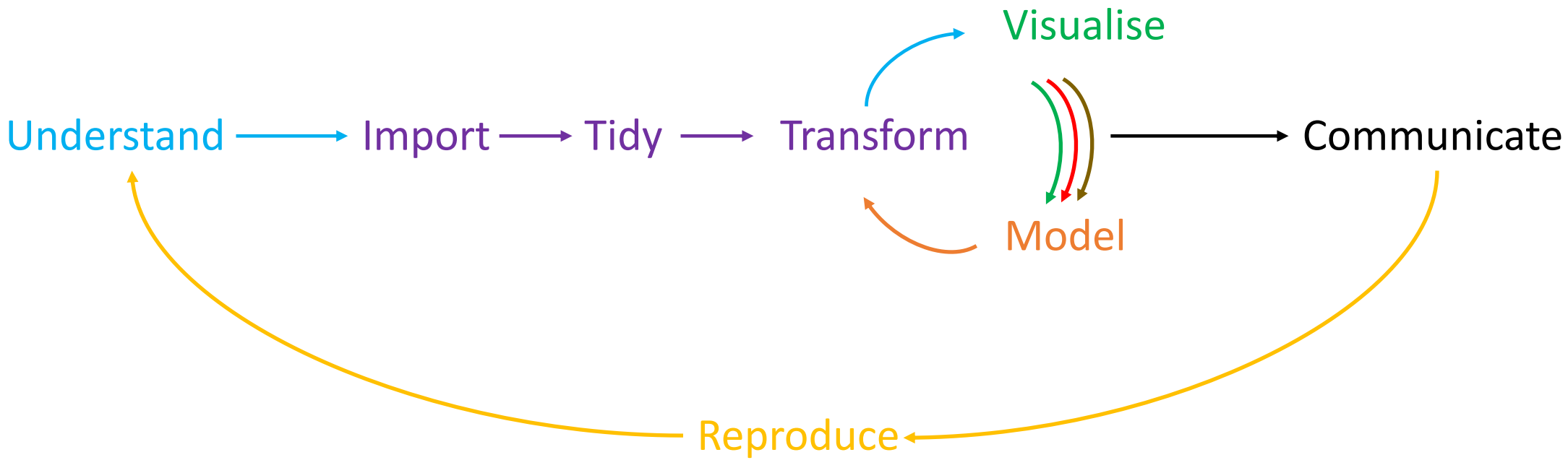
2. Null Hypothesis Significance Testing and Exploratory Data Analysis
3. Computational Statistics
4. Understanding and Visualising Variation
5. Data Visualisation and Visual Inference
6. Using Git and GitHub

7. Data Wrangling **(THIS WEEK)**
8. Statistical Models



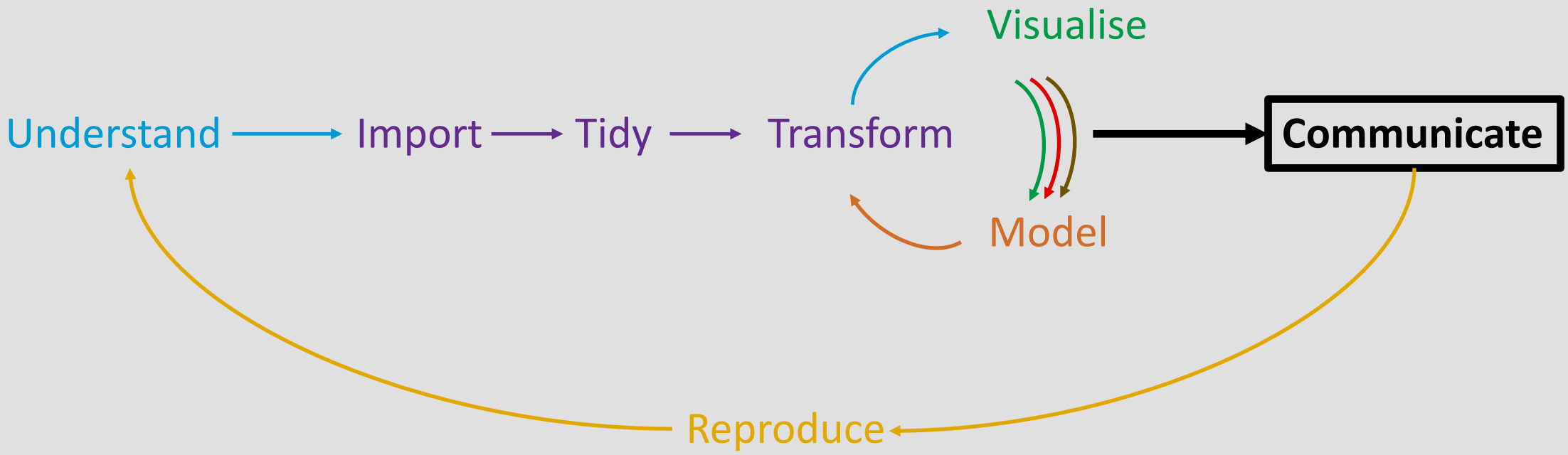
2. Null Hypothesis Significance Testing and Exploratory Data Analysis
3. Computational Statistics
4. Understanding and Visualising Variation
5. Data Visualisation and Visual Inference
6. Using Git and GitHub

7. Data Wrangling **(THIS WEEK)**
8. Statistical Models
9. Meta-Analysis I
10. Meta-Analysis II
11. Dimension Reduction and Visualisation



2. Null Hypothesis Significance Testing and Exploratory Data Analysis
3. Computational Statistics
4. Understanding and Visualising Variation
5. Data Visualisation and Visual Inference
6. Using Git and GitHub

7. Data Wrangling **(THIS WEEK)**
8. Statistical Models
9. Meta-Analysis I
10. Meta-Analysis II
11. Dimension Reduction and Visualisation



ASSIGNMENTS!

Mode and tempo for Part 2

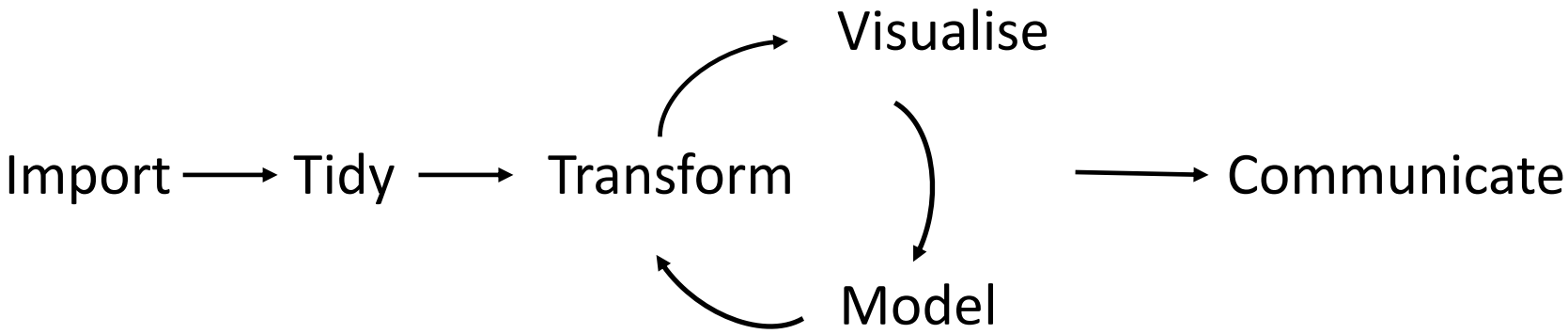
- Tutorial used for short overview lectures and answering questions

- Workshops are as in Part 1 but cover less material, potentially leaving time to work on and receive help on assignments

- Assignment 1 (Digital Detective): Release on 23 Sep, due 7 Oct
- Assignment 2 (Ocean Acidification): Release on 7 Oct, due 28 Oct(?)
- *Assignment 3* (Poster Presentation): Present in class on 28 October

Assignment 1: Digital Detective

- You will be given a dataset and challenged to explore it

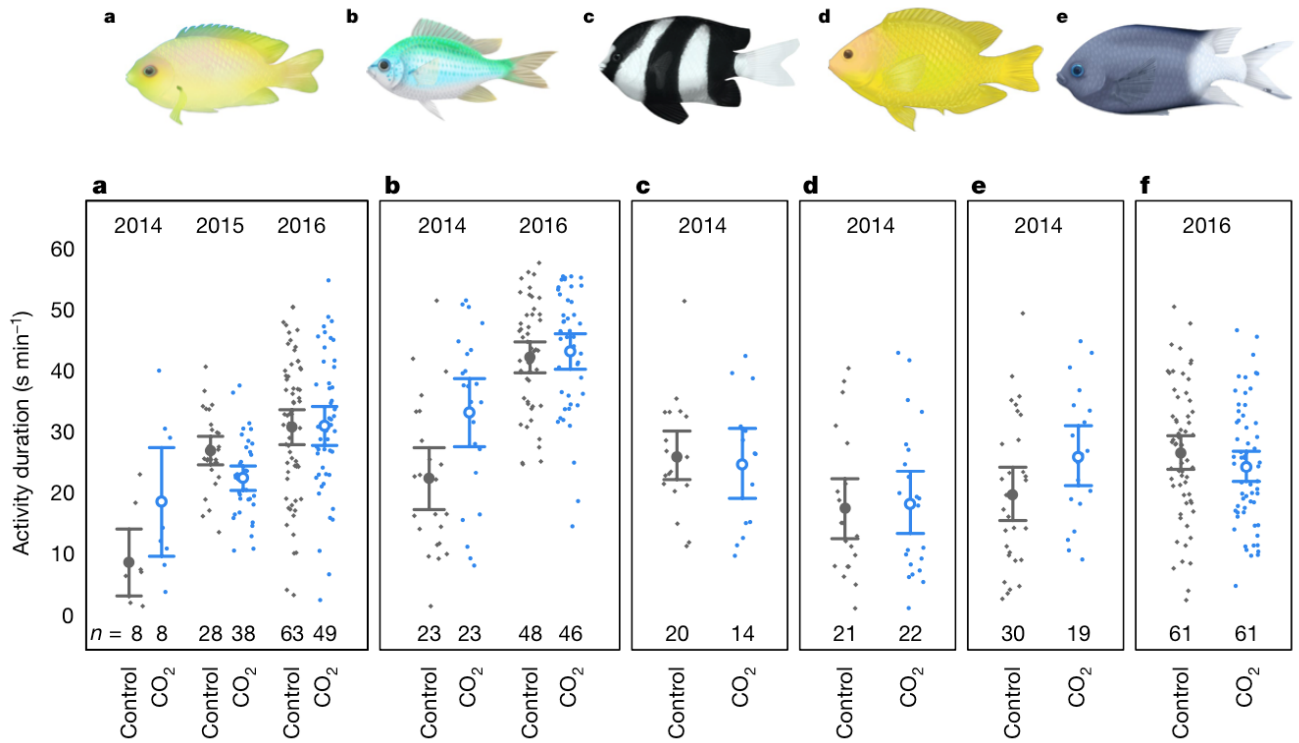


- You will be expected to demonstrate the skills learned in class so far

- You will be required to prepare a report in R Markdown and knit to html

Assignment 2: Ocean Acidification

- Combine your analysis on Clark et al. 2020 into larger meta-analytic dataset
- Conduct meta-analysis to estimate the overall impact ocean acidification has on behaviour
- Understand factors driving effect size variation
- Identify any publication biases
- You will prepare a report in Rmarkdown that should be rendered as an html

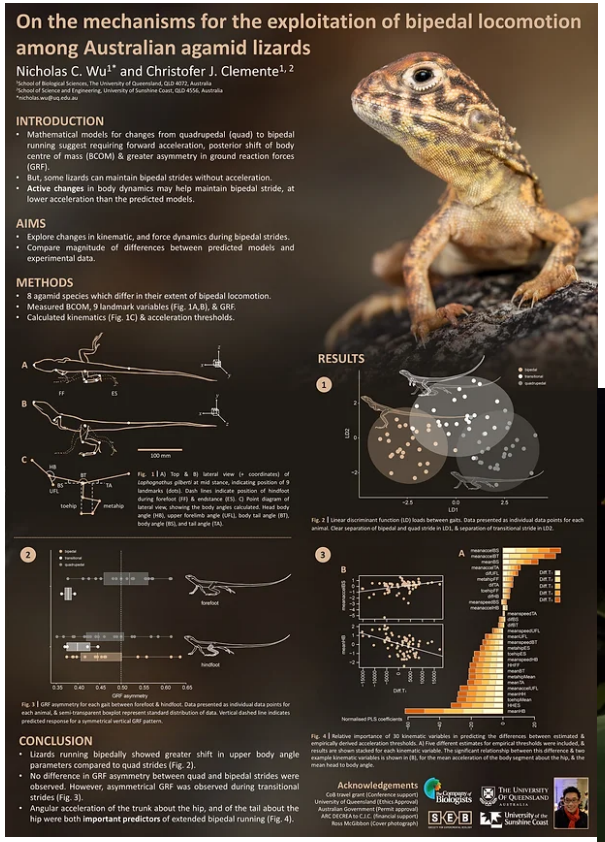


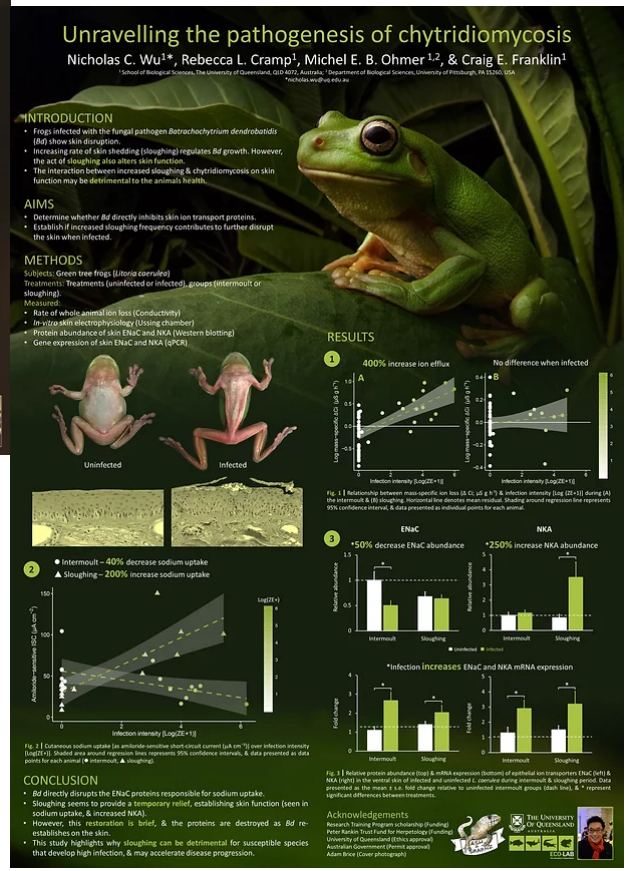
Assignment 3: Poster Presentation

- BIOL6207 (20% of your grade) and HPO for BIOL3207 (If you would like)
- You will prepare a poster on a manuscript of your choice
- Manuscript data ***must*** be available for you to re-analyse

Tentative Assessment….

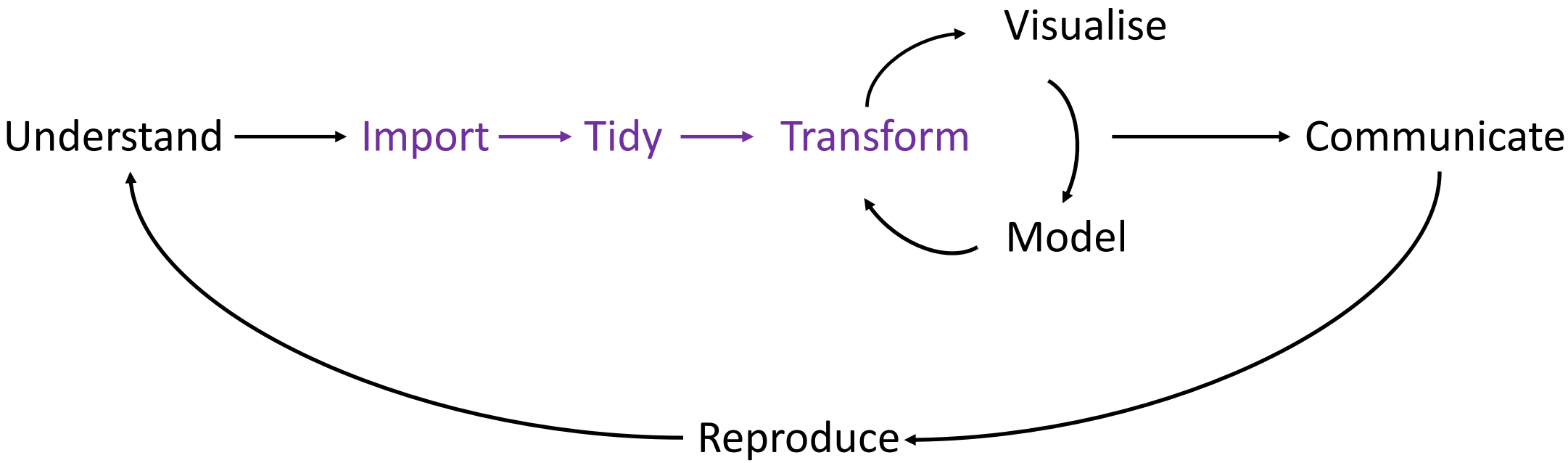
1. **Introduction to the paper (10%)**
    1. Based on your reading, provide a brief summary of the paper in your own words
    2. Identify the knowledge gaps identified by the paper
2. **Clear articulation of the problem / question / aims of the paper (10%)**
    1. Provide a clear statement of the major aims, questions, hypotheses of the paper
3. **A brief summary describing how the data was collected (15%)**
    1. Provide a summary of how the data was collected
    2. If sensible, use a figure to provide an overview of the experimental design
    3. Clearly identify the key variables that are relevant to the question
    4. Provide a brief few statements of the statistical approaches you used to analyse the data and why
4. **Statistical analysis and interpretation of results in relation to the question (20%)**
    1. Statistical analysis is appropriate for the question and variables
    2. Results are correctly and accurate presented and described
    3. Meaning and interpretation of the results in relation to the question(s) are correct
5. **Quality and appeal of figures and their relevance to the question / aims / objectives (25%)**
    1. Figures are labelled correctly
    2. Information rich, but without too much detail
    3. Figures are clearly described by legends
    4. Colour schemes chosen are appropriate
    5. Figures contain useful visualisation of the results in relation to the main aims / questions / objectives
6. **Overall presentation, layout and aesthetics of poster (20%)**
    1. Layout of the poster is visually appealing
    2. Poster does not contain too much text
    3. Figures are appropriately placed and organised
    4. Knowledgeable about the poster, and answers questions about the analysis and paper sufficiently well





Some great examples!!

This Week: Data Wrangling



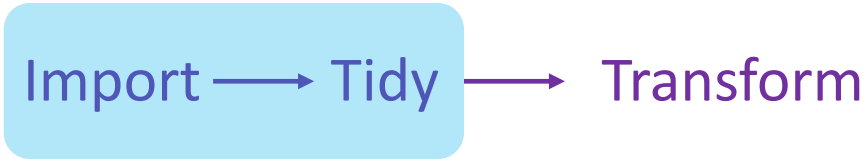
Data wrangling explained

- Data wrangling is the art of getting your data into R in a useful form for visualisation and modelling (see R4DS, Chapter 9)

Import ⟶ Tidy ⟶ Transform

Data wrangling explained

- Data wrangling is the art of getting your data into R in a useful form for visualisation and modelling (see R4DS, Chapter 9)



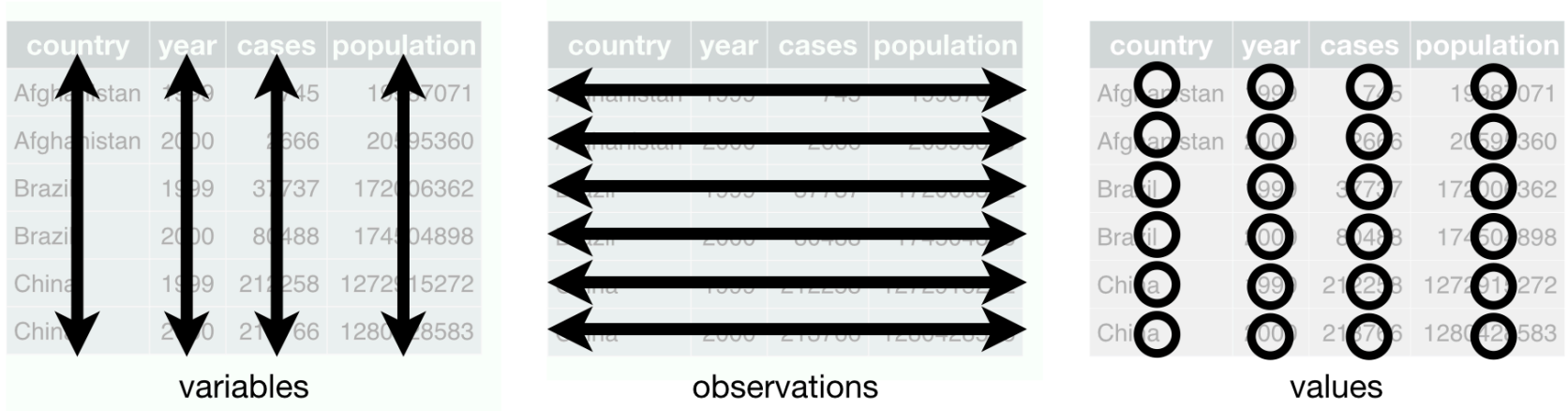
- Tidy data (in R) obeys three interrelated rules:

1. Each variable must have its own column
2. Each observation must have its own row
3. Each value must have its own cell

Tidy data

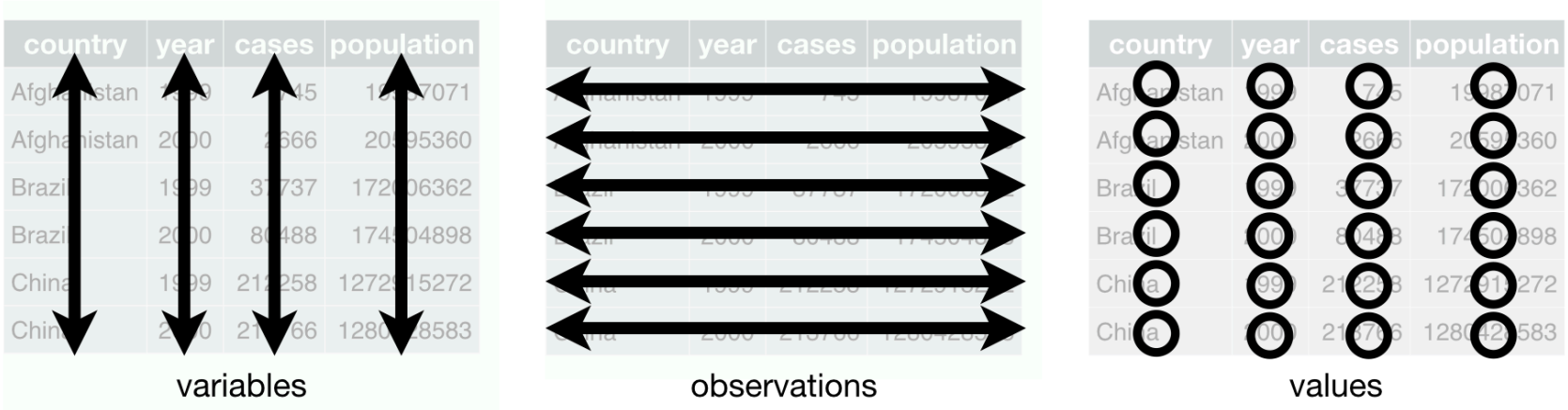
- Tidy data (in R) obeys three interrelated rules:

1. Each variable must have its own column
2. Each observation must have its own row
3. Each value must have its own cell



Tidy data in practice, in R

- Put each dataset in a tibble
- Put each variable in a column



- This is what we've been doing for weeks…

Why bother with tidy data?

- Uniformity
- Reproducibility
- Efficiency
- Functionality

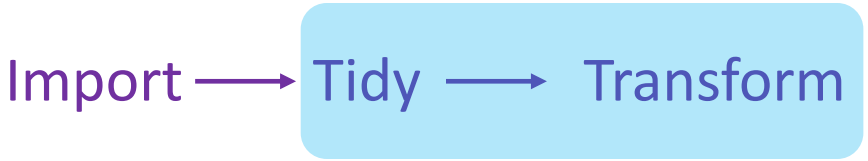
Data wrangling explained

- Data wrangling is the art of getting your data into R in a useful form for visualisation and modelling (see R4DS, Chapter 9)

Import ⟶ Tidy ⟶ Transform

Data wrangling explained

- Data wrangling is the art of getting your data into R in a useful form for visualisation and modelling (see R4DS, Chapter 9)



- Transformation is how we manipulate tidy data to make it usable (for visualisation and modelling)

Transformation

- Emphasis on working with types of data frequently encountered in practice (see R4DS, Chapters 13-16)

    1. Relational data (multiple interrelated datasets)
    2. Strings (regular expressions)
    3. Categorical data (factors)
    4. Dates and times

Friday's workshop

- A guided series of exercises (in R Markdown) on Data Wrangling
- Corresponds to material in Chapters 9-16 of R4DS
- Release of Assignment 1, including detailed rubric
- No more quizzes or pseudo-quizzes!