What to do after you select the article?

- Get the raw data for the method reimplementation
  - Can you physically download it?
  - Is it readable, not too large, …
- Re-implement the method
  - Preferably in Jupyter/Python, is easier to evaluate, but anything is fine
- Mitigation
  - If you have problems, let us know, and discuss it with us
  - It might be possible that you cannot re-implement everything, but we need to know in advance why and what
- Write this up in article form
  - We will have question sessions in case you have problems (see schedule)
  - Submission date 30/05, 23:59, with project code etc. (git rep also OK)

Article

- Introduction
  - What is your text about?
- Materials & methods
  - Which data and methods?
- Results
  - Describe the results you obtained
- Discussion
- References

Article

- Introduction
  - Start with generic context
  - Then zoom in: why is your approach necessary and relevant within this context?
  - Finish with a brief description of your method and what you found out

  - Read background information!
  - Brief, 1 page (font size 11, normal margins)

Article

- Materials and methods
  - Describe the datasets you are using
    - Both their content and where you got them
  - Describe the methods you implemented to get your results
    - Describe separate steps required (reproducibility!)
    - Overview figures can help.
  - 2-3 pages

Article

- Results
  - Describe your results
    - This can be in relation to:
      - Different input/validation data
      - Different method parameterisations, …
  - Explain what the results mean
  - Use tables for numbers (do not list in the text)
  - Figures for distributions, relationships, … (easier to understand than text)
  - 2-3 pages

Article

- Discussion
  - Which issues did you identify, and which problems did you encounter?
  - What is different about your approach (and the results you get) in comparison to the original method? Why?
  - What are advantages/disadvantage of each method?
  - 1 page

Article

- Final article about 6-8 pages (excl. References)
- No plagiarism!
- Many guides with background info available…
  - http://abacus.bates.edu/~ganderso/biology/resources/writing/HTWsections.html
  - http://www.nature.com/scitable/ebooks/english-communication-for-scientists-14053993/writing-scientific-papers-14239285

Presenting results

Typical data flow in bioinformatics

| Experiment | Acquisition | Raw Data | Analysis | Processed Data | Visualization | Presentation | Interpretation | Finding |
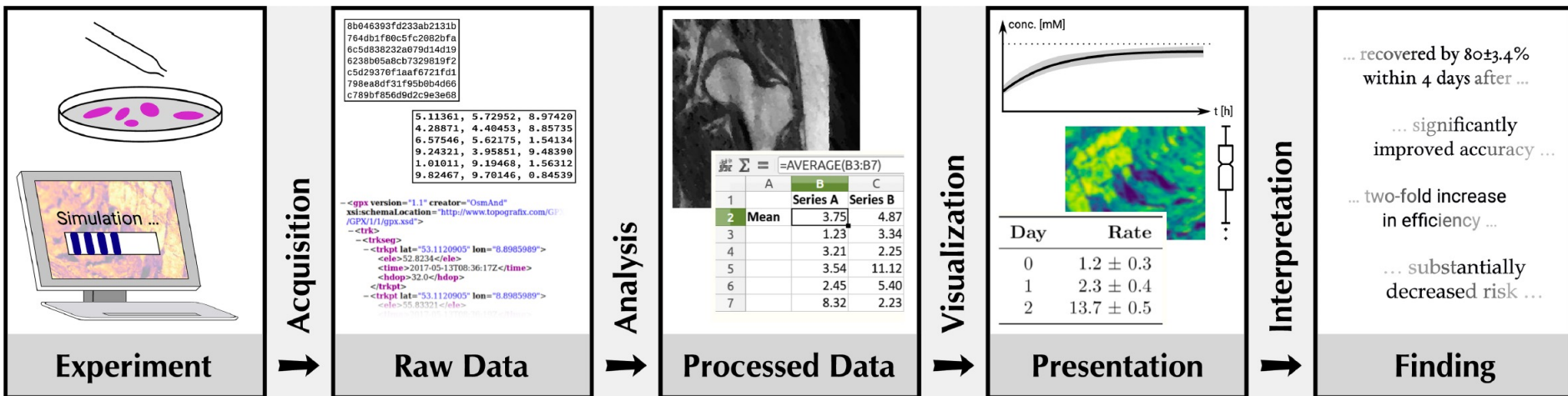
To discuss in relation to data/results

- Bias
  - Accuracy and precision

- Data distributions

- How to fit data

- Data visualization

- Data integration

Data bias

Data bias

- https://study.com/academy/lesson/bias-in-statistics-definition-examples.html

Data bias

Accuracy and precision

Data distributions

Negatively Skewed

Mean Median Mode

Symmetric (Not Skewed)

Mean
Median
Mode

Positively Skewed
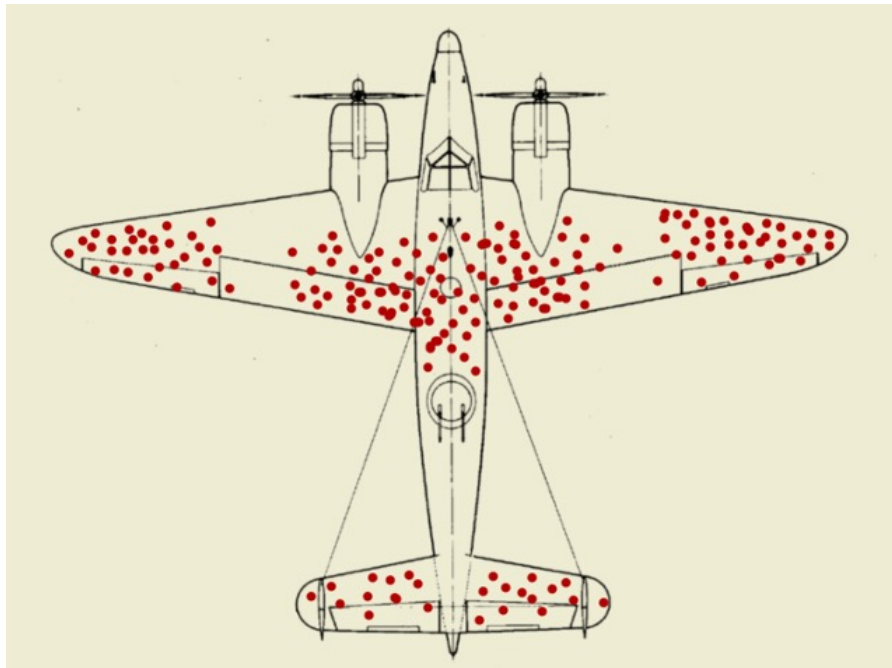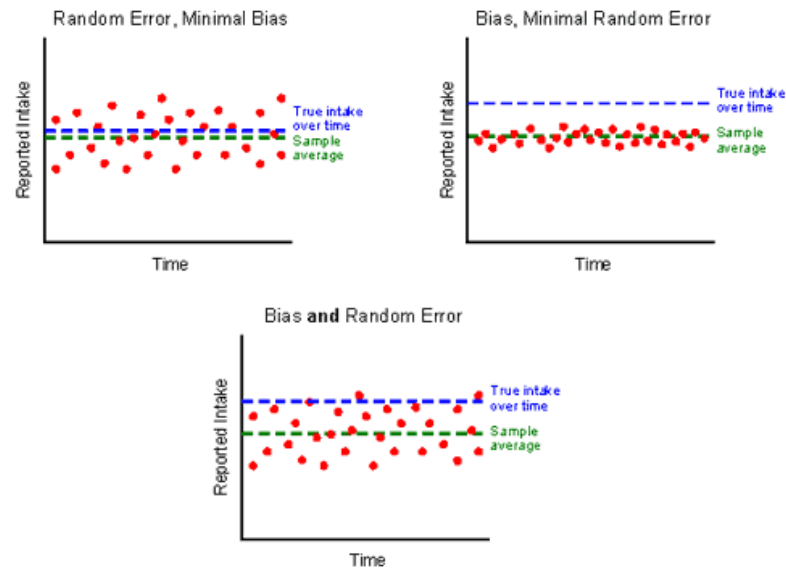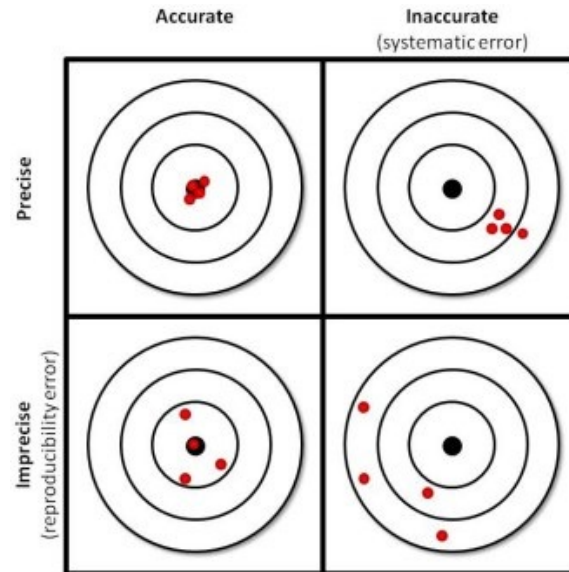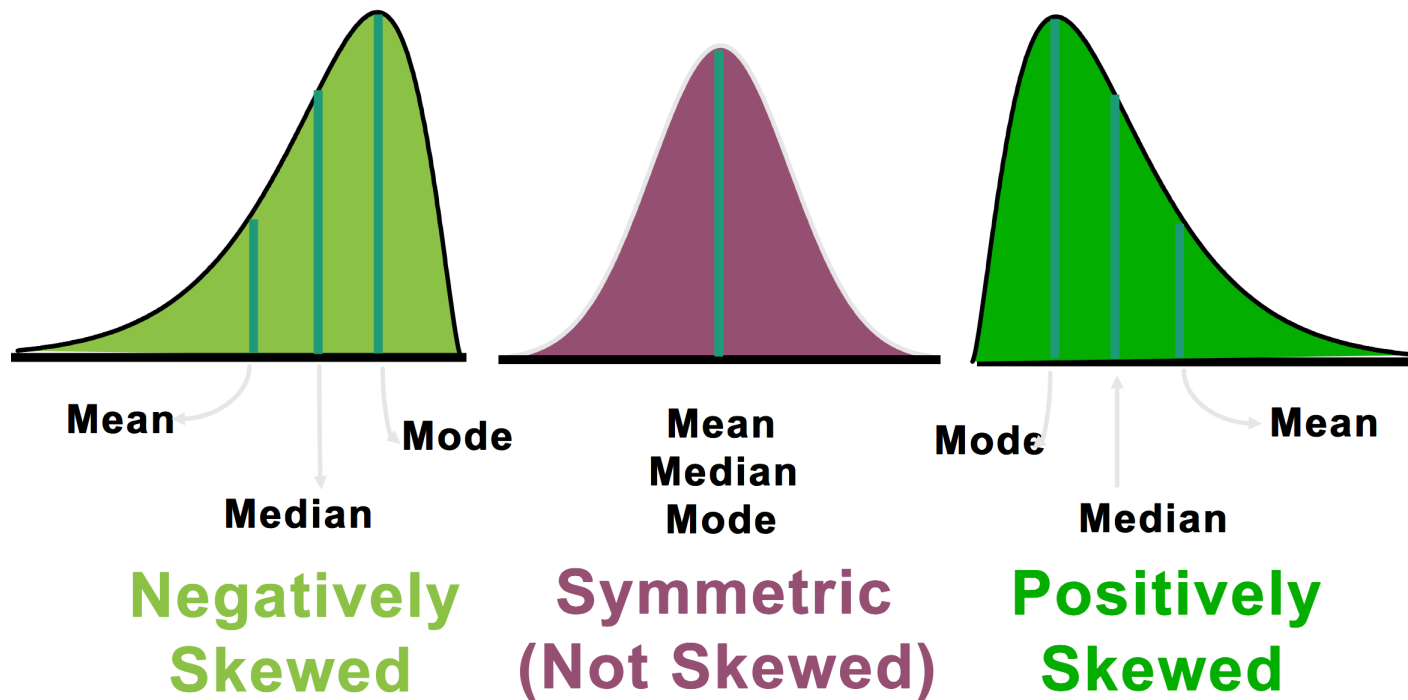
Mode Median Mean

Data distributions

Table 2.1 Two Data Sets

Data Set I:	40	38	42	40	39	39	43	40	39	40
Data Set II:	46	37	40	33	42	36	40	47	34	45

Mean? Median? Mode?

Figure 2.10 *Dot Plots of Data Sets*
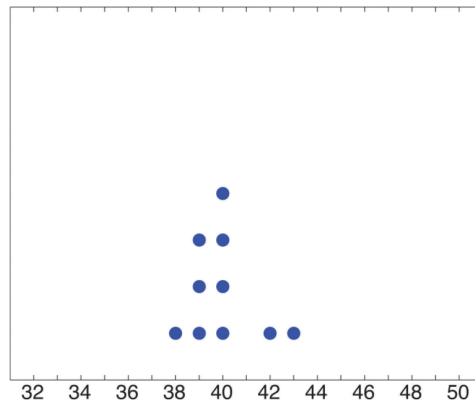
(a) Set I
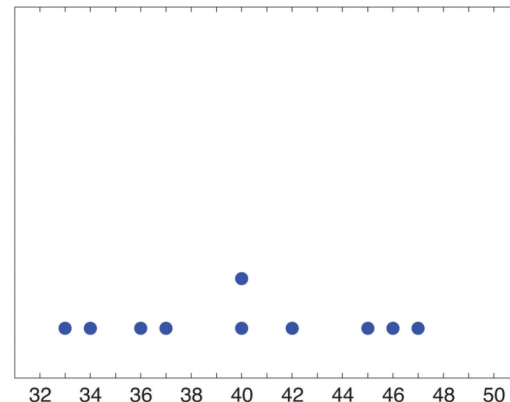
(b) Set II

Data distributions
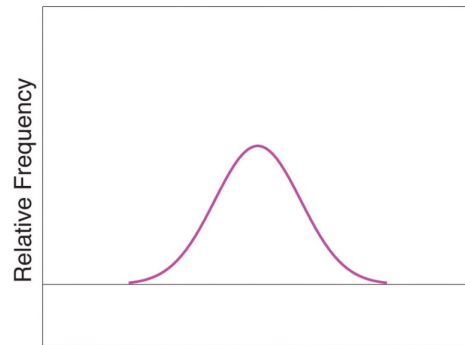
(a) Two Identical Sets
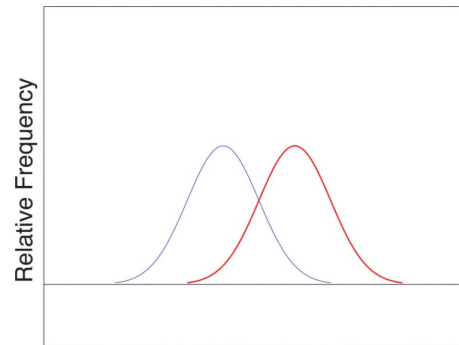
(b) Locations Differ
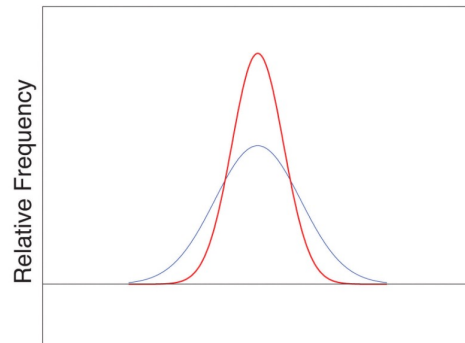
(c) Variabilities Differ
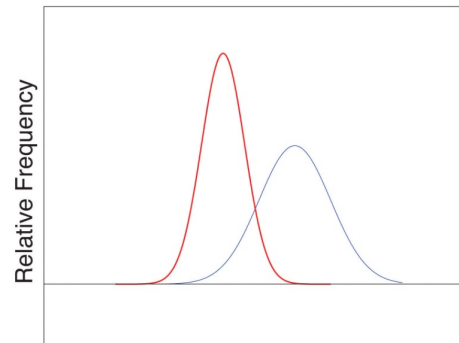
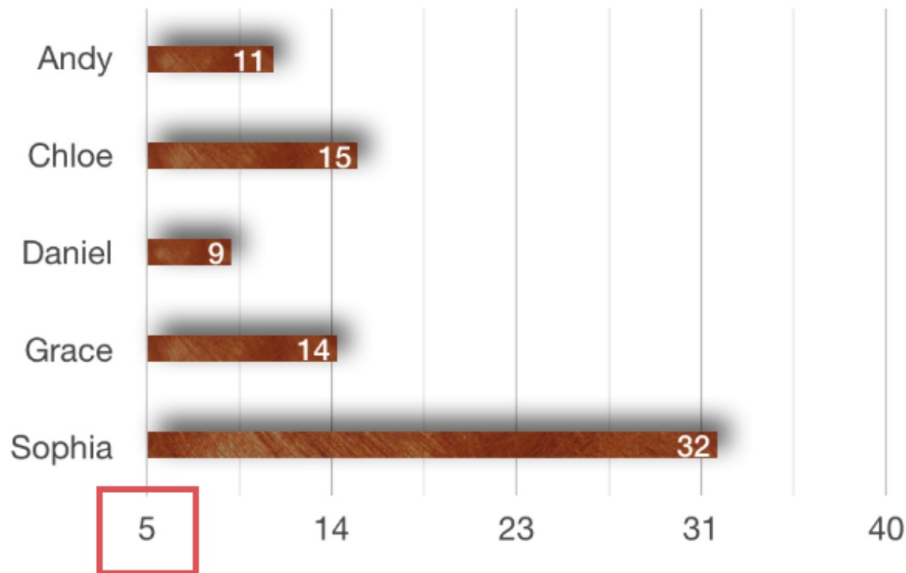(d) Locations and Variabilities Differ

Fitting data
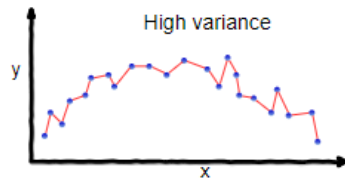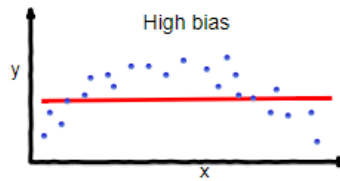
| overfitting | underfitting | Good balance |
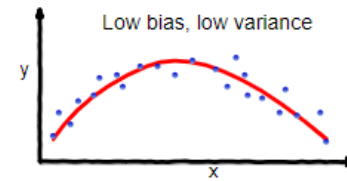
Fitting data
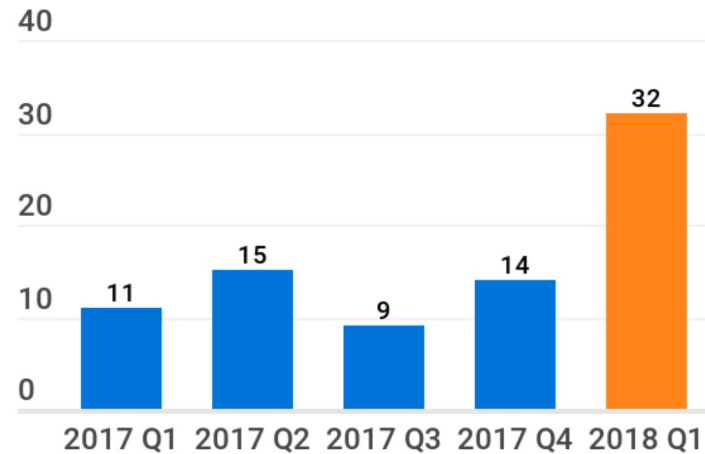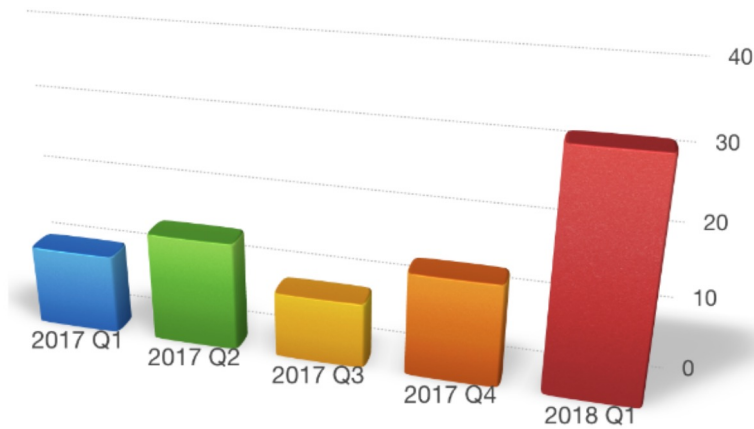

R² = 0.06

REXTHOR, THE DOG-BEARER
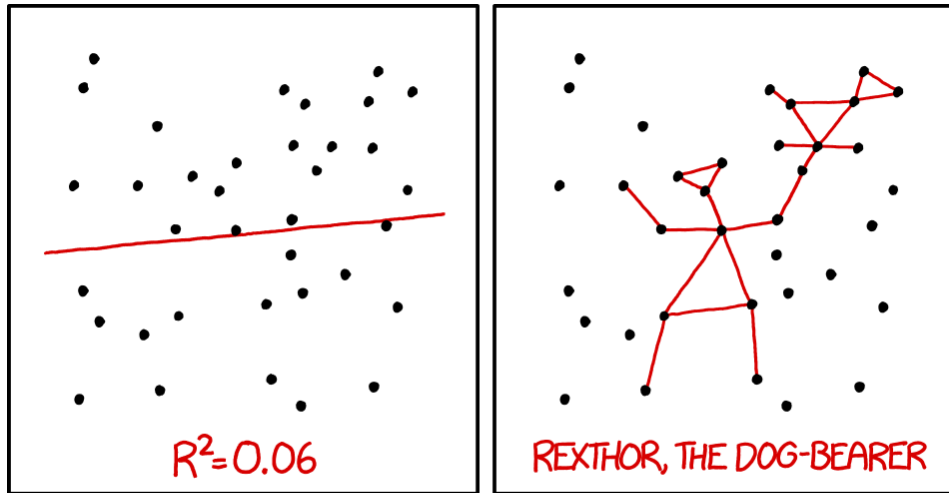
I DON'T TRUST LINEAR REGRESSIONS WHEN IT'S HARDER TO GUESS THE DIRECTION OF THE CORRELATION FROM THE SCATTER PLOT THAN TO FIND NEW CONSTELLATIONS ON IT.

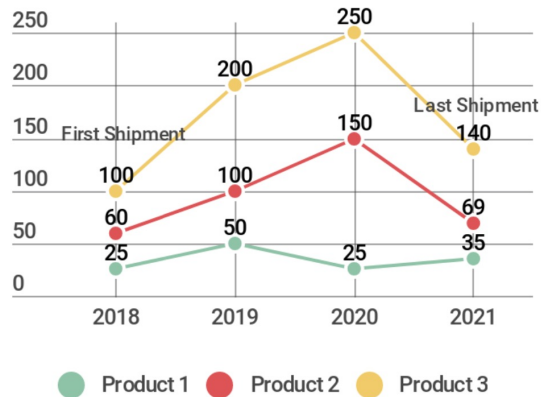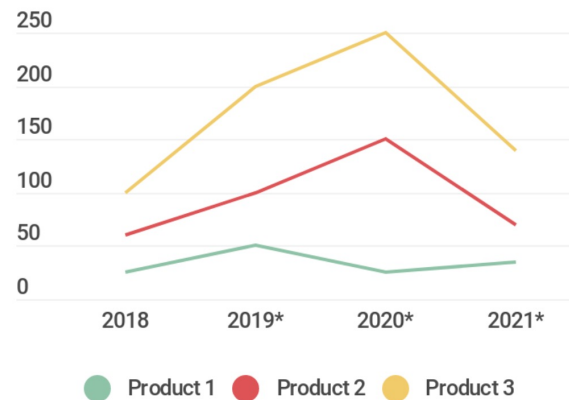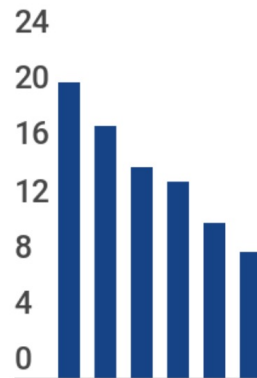# Data visualisation
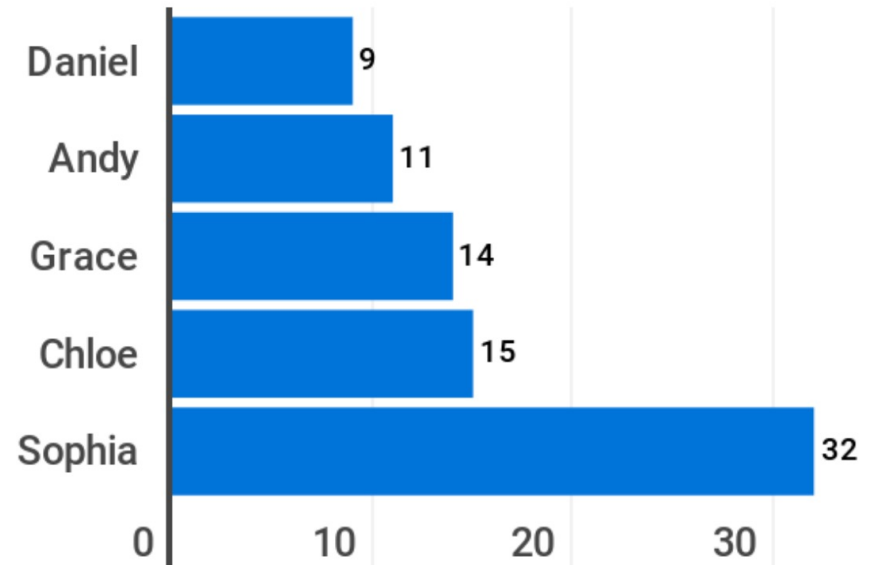
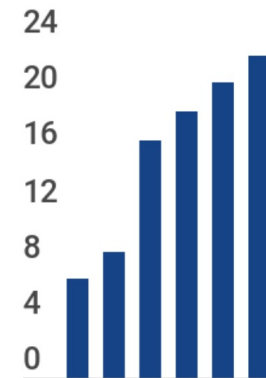# Data visualisation
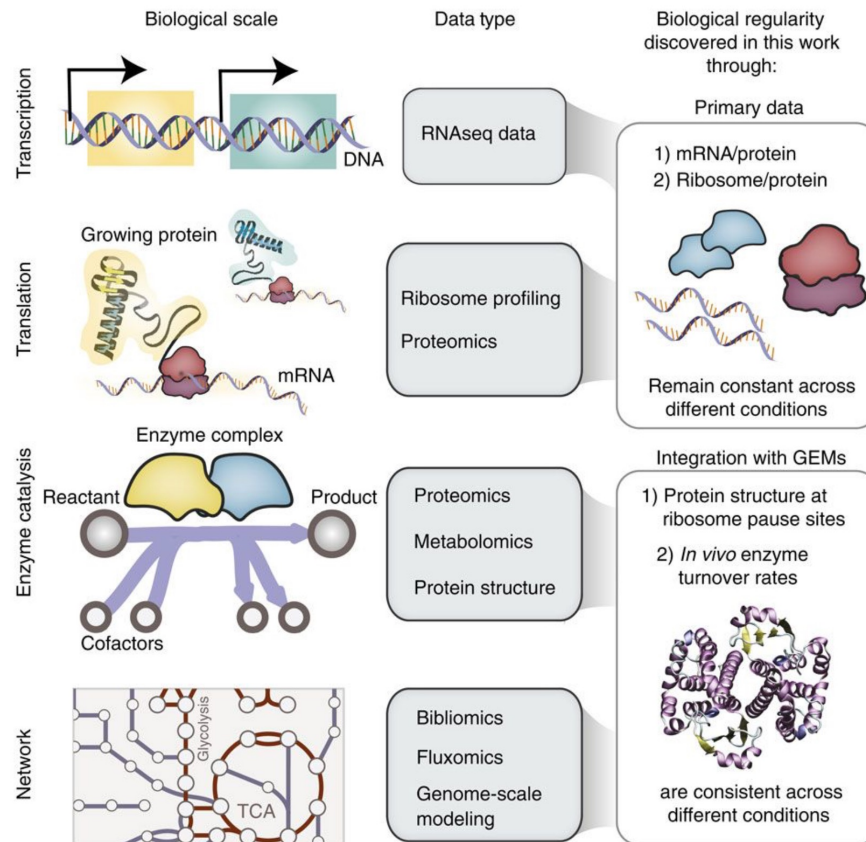
Data visualisation

Data visualisation

Data integration

Data management

- Findable, Accessible, Interoperable, Reusable (FAIR)
- Data management plan (DMP)