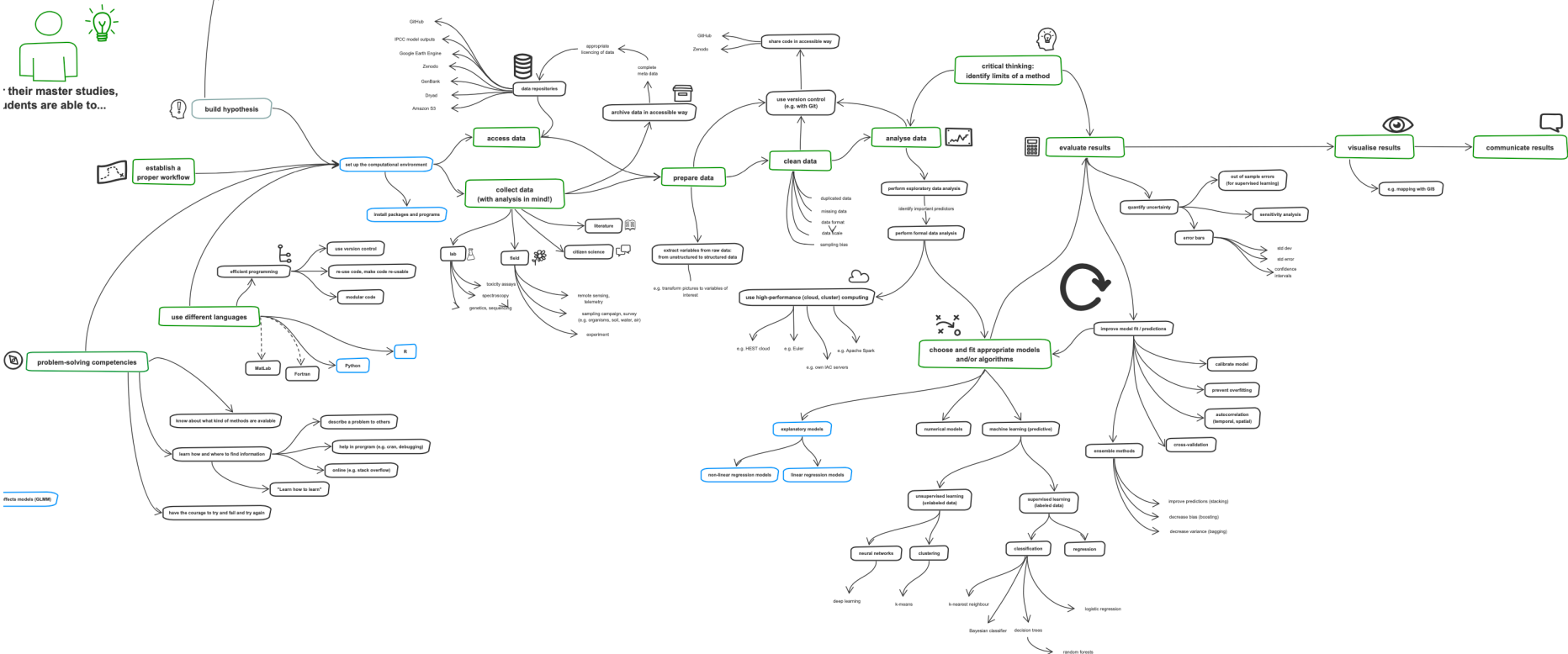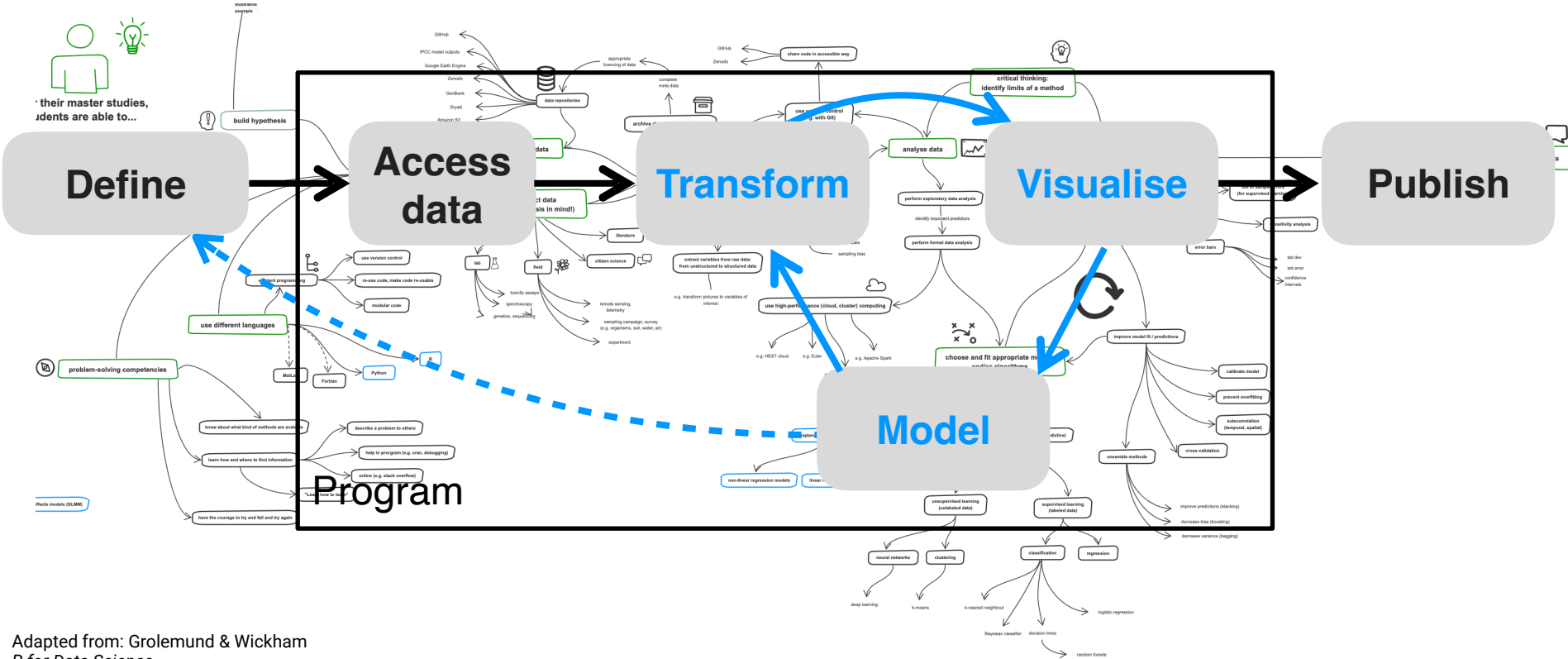# The data science workflow

The data science workflow

Adapted from: Grolemund & Wickham
R for Data Science

• Programming
• R, RStudio, *tidyverse*, *caret*
• Data cleaning

• Grammar of graphics

Define → **Access data** → **Transform** → **Visualise** → **Publish**

Model

Program

• Research data repositories
• APIs
• Data formats

• Open Science
• Reproducibility
• *git*

• Regression and classification
• Supervised machine learning
• Bias-variance trade-off
• Random Forest

Define → Access data → Transform → Visualise → Publish

Program