

Workflow Pipelines with Orange3 (and R)

What is a workflow?

- 'the sequence of steps involved in moving from the beginning to the end of a working process' Merriam-Webster dictionary
- A series of ordered processes
- Replicable
- Clear steps

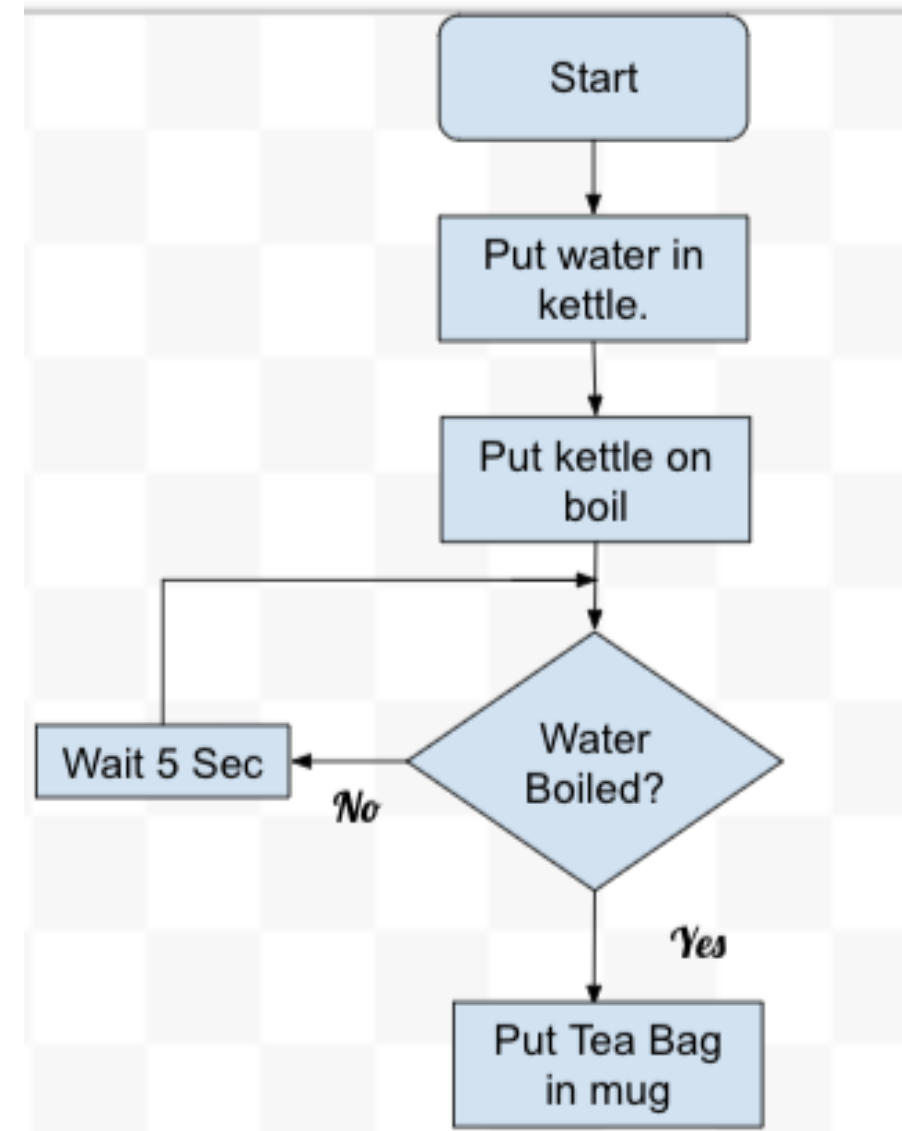


Types of pipeline

- Lots of things could be considered as workflow pipelines
- E.g.
 - Recipes
 - Code
 - Research?
 - Thesis?
- Visually, we can compare pipelines to flow diagrams



Tea making: a pipeline



QMLHS as a workflow pipeline?

- Putting it all together:

"Approaching Humanities Problems Quantitatively"

"Text Mining using only Microsoft Word and Excel"

"Getting Texts from the Right Sources and Preparing Them for Analysis"

"Introduction to Statistics for Humanists"

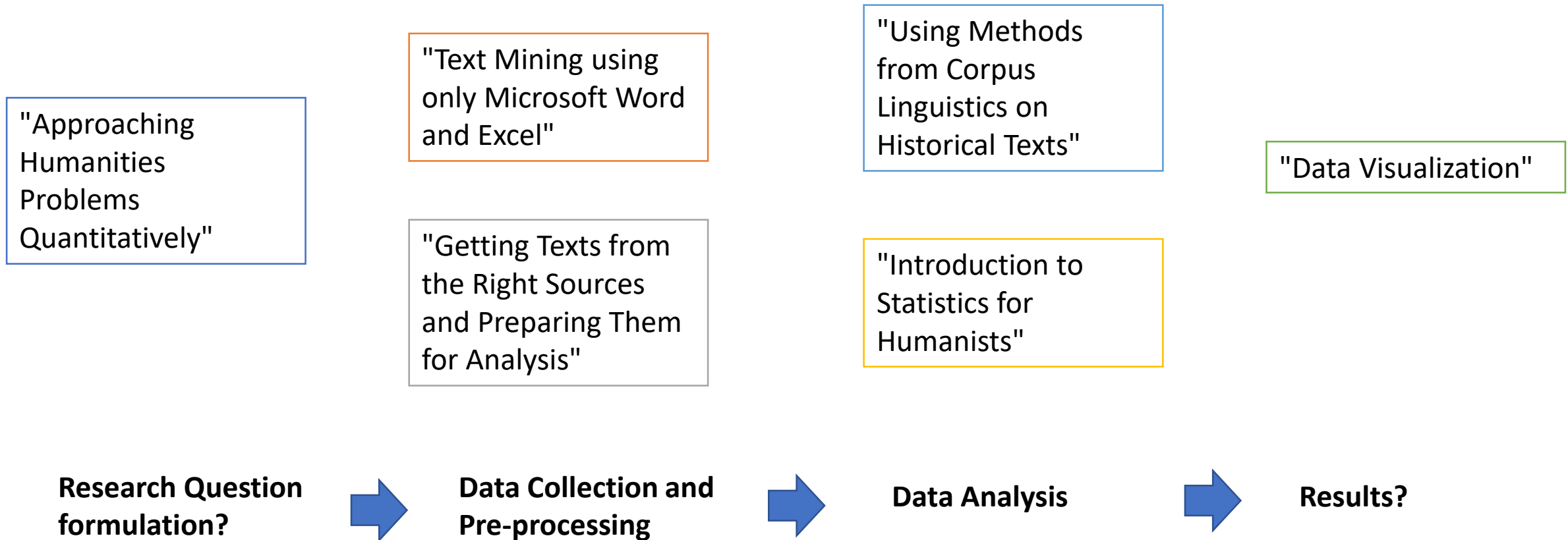
"Using Methods from Corpus Linguistics on Historical Texts"

"Introduction to the Programming Language Python"

"Data Visualization"

QMLHS as a workflow pipeline?

- Putting it all together:



An example from my own work:

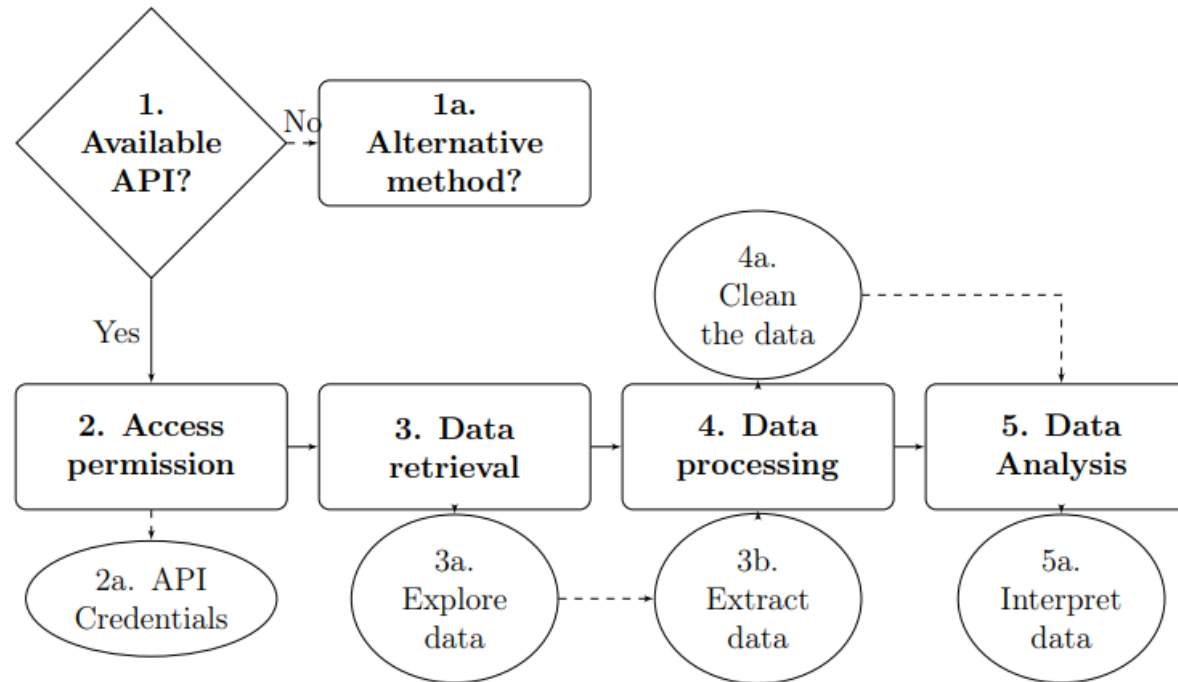
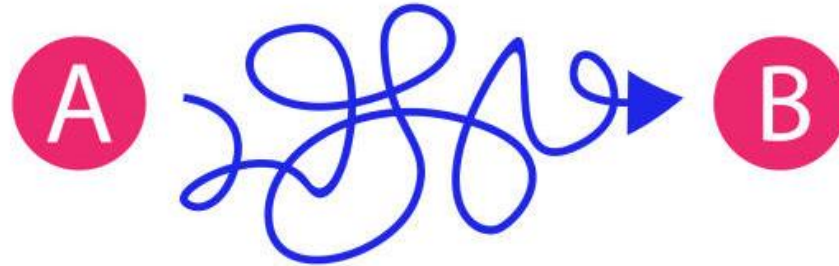


Figure 3.1: OED API workflow implemented in this study. Each rectangular node represents a stage of the process with the circular nodes depicting the exploratory processes or considerations at each stage.

Mental shift?

- Reality of research/process



- Communication/
Creation of a
pipeline



Quantitative approaches to research

- Need to be transparent
- Need to be replicable
- Need to be principled
- Need to be motivated

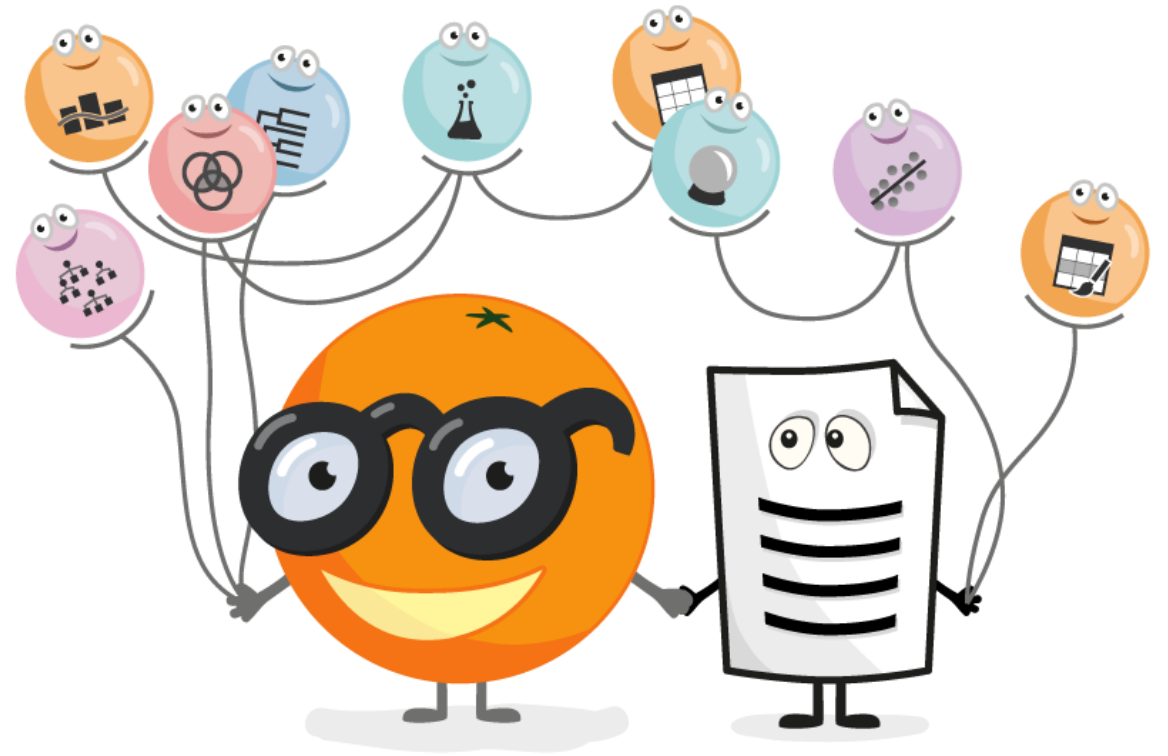
The biggest lie I tell myself



**"I don't need
to write that down,
I'll remember it"**

Orange3

- Visual interface/toolbox for data analysis
- Create workflows by dragging and dropping the stages into the workspace
- Benefit for beginners in being interactive
- Ability to add python scripts into the workflow



Hands On!

R



- A programming language
 - Specialises in statistics
- Can be installed via a variety of IDEs
- In principle, is the same as Python/other languages
 - Idea is we take an input and create the desired output
- Like Orange3 – but goes ‘under the hood’ for more flexibility/options
- RMarkdown is a very useful tool (like Jupyter Notebooks)

