

# E+ ADSEE

0: SETUP

# Overview

You will be using the R language to make your life easier when manipulating and modelling data.

An **IDE** (Integrated Development Environment) allows you to write, run and manage your code.

**Rstudio** is the recommended IDE for this course as it has many useful features and a large community of users and is well documented.

**Orange3** is a drag and drop tool for data mining. This is useful for quickly reviewing moderately sized datasets.

**Butter** allows you to create a custom data pipeline which includes looking at the validity of custom dictionaries. – [How to contribute](#)

# Rstudio Setup

You will need to install the R language and then R studio. You may then wish to configure.

There are many installation instructions on the Internet that may change over time. Try here first:

1. Installing R: <https://www.datacamp.com/community/tutorials/installing-R-windows-mac-ubuntu>
2. Rstudio: <https://www.dummies.com/programming/r/how-to-install-and-configure-rstudio/>
3. [https://www.reed.edu/data-at-reed/software/R/r\\_studio\\_pc.html](https://www.reed.edu/data-at-reed/software/R/r_studio_pc.html)

# Orange3 and Butter Setup

GUI tools for drag and drop data pipelines.

## Installation Instructions:

- <https://orangedatamining.com/download/#windows>
- <https://www.butter.tools/download/>

# HINTS

- There is vast amounts of excellent documentation on the Internet

From within Rstudio or R

- `?command` returns help for the command
- `??command` searches the Internet about the command
- The help option within Rstudio points to useful documentation
- If you code then try and follow conventions and keep your coding short and well commented

# The code

In this course you do not have to program. However, we have programming examples in R.

1. The **.Rmd** files allow you to run the code from R studio
2. The **.nb.html** files are saved output and results. Even if you do not run the code, you can see the results with comments.
3. An instructor can run the code from R studio and share their screen when video conferencing.
4. Packages such as ggplot2 dependencies e.g. which version of a package is run will be mentioned at the end of the .nb.html files. If code does not run then this is the first place to review.

# Resources

- Rstudio educational site: <https://education.rstudio.com/learn/beginner/>
- Introductory video: <https://www.youtube.com/watch?v=mcYcjH-1giM>
- An introduction to R: <https://intro2r.com/the-aim-of-this-book.html>
- Coding conventions: <https://www.r-bloggers.com/2014/07/consistent-naming-conventions-in-r/>

To read a sample of online books, visit the following sites

1. R with github - <https://happygitwithr.com/>
2. R markdown - <https://bookdown.org/yihui/rmarkdown/>
3. R graphics cookbook - <https://r-graphics.org/>
4. Mastering software development - <https://bookdown.org/rdpeng/RProgDA/>

There is much free online support

1. [R with github](#)
2. [R markdown](#)
3. [R graphics cookbook](#)
4. [Mastering software development](#)
5. [Unit testing](#)
6. [R packages](#)
7. [Graphics via plotly](#)
8. Book on [visualisations with R](#).
9. Another excellent book on [visualisations](#)
10. Hands on programming with [R](#)
11. [R shiny](#)
12. [Feature selection from your data](#)
13. [R graphics](#)
14. R for [Data Science](#)
15. [R statistics](#)
16. [Text mining](#)
17. ML with [Caret](#)
18. [Statistical thinking](#)
19. Not for the faint hearted [Regression testing](#)
20. Open books from [Springer](#)