Applied Social Data Science - Coding Camp

Marina Schenkel

PhD Candidate

September 4 - 8, 2023

Schedule

- *Monday: Introduction*
- *Tuesday: R basics*
- *Wednesday: R basics + Good practices*
- *Thursday: Python basics*
- **Friday: How to write up and report - LaTeX**

Recap

In the last few days we looked at:

Recap

In the last few days we looked at:

- Some basics of working in R

Recap

In the last few days we looked at:

- ▶ Some basics of working in R
- ▶ Some basics of working in Python

Recap

In the last few days we looked at:

- Some basics of working in R
- Some basics of working in Python
- Some general good practices to follow when doing data science
    - Organising your work
    - Being consistent in things like object naming

Today's class (the final class!)

How to write up and report

Today's class (the final class!)

How to write up and report
- For your own projects

Today's class (the final class!)

How to write up and report

- For your own projects
- For doing your class assignments

Today's class (the final class!)

How to write up and report

- ▶ For your own projects
- ▶ For doing your class assignments

This class is intended to cover the basics and help orient your future learning.

Today's class (the final class!)

How to write up and report

- ▶ For your own projects
- ▶ For doing your class assignments

This class is intended to cover the basics and help orient your future learning.
We will look at:

Today's class (the final class!)

How to write up and report

- ▶ For your own projects
- ▶ For doing your class assignments

This class is intended to cover the basics and help orient your future learning.
We will look at:

- ▶ Jupyter Notebook
- ▶ LaTeX
- ▶ GitHub
- ▶ R Markdown

Jupyter Notebook

Jupyter Notebook

- Jupyter stands for **Ju**lia, **pyt**hon and **r**, the three languages it was initially designed to work with (although it now supports more)

Jupyter Notebook

- ▶ Jupyter stands for **Ju**lia, **pyt**hon and **r**, the three languages it was initially designed to work with (although it now supports more)
- ▶ Like R Markdown, it provides a format for combining code, results and commentary in a single document

Jupyter Notebook

- Jupyter stands for **Ju**lia, **pyt**hon and **r**, the three languages it was initially designed to work with (although it now supports more)
- Like R Markdown, it provides a format for combining code, results and commentary in a single document
- Jupyter runs through your web browser

Jupyter Notebook

- Jupyter stands for **Ju**lia, **pyt**hon and **r**, the three languages it was initially designed to work with (although it now supports more)
- Like R Markdown, it provides a format for combining code, results and commentary in a single document
- Jupyter runs through your web browser
- It contains two components: a kernel (to run code) and a dashboard (to navigate files)

Jupyter Notebook

- ▶ Jupyter stands for **Ju**lia, **pyt**hon and **r**, the three languages it was initially designed to work with (although it now supports more)
- ▶ Like R Markdown, it provides a format for combining code, results and commentary in a single document
- ▶ Jupyter runs through your web browser
- ▶ It contains two components: a kernel (to run code) and a dashboard (to navigate files)

Activity

You already have Jupyter Notebook installed as part of your Anaconda configuration. We will work on turning our script file from Wednesday into a notebook.

R Markdown

- R Markdown is a way of combining your code, its results, and your commentary in a single document.
- It can then be used to create outputs in a variety of formats, including PDFs, Word files, and HTML documents.
- R Markdown can be used to produce reports, presentations, dashboards and more.

LATEX

LaTeX

LaTeX is a free to use software system for typesetting documents. Unlike word processing software such as MS Word - which uses a 'What You See Is What You Get' (WYSIWYG) interface, LaTeX uses plain text and markup language. As with R and Python, LaTeX has a steep learning curve. So why use it?

LaTeX

LaTeX is a free to use software system for typesetting documents. Unlike word processing software such as MS Word - which uses a 'What You See Is What You Get' (WYSIWYG) interface, LaTeX uses plain text and markup language. As with R and Python, LaTeX has a steep learning curve. So why use it?

- ▶ It's free

LaTeX

LaTeX is a free to use software system for typesetting documents.
Unlike word processing software such as MS Word - which uses a
'What You See Is What You Get' (WYSIWYG) interface, LaTeX
uses plain text and markup language. As with R and Python, LaTeX
has a steep learning curve. So why use it?

- ▶ It's free ...But so is Libre Office

LaTeX

LaTeX is a free to use software system for typesetting documents. Unlike word processing software such as MS Word - which uses a 'What You See Is What You Get' (WYSIWYG) interface, LaTeX uses plain text and markup language. As with R and Python, LaTeX has a steep learning curve. So why use it?

- ▶ It's free ...But so is Libre Office
- ▶ It makes very professional looking documents

LaTeX

LaTeX is a free to use software system for typesetting documents. Unlike word processing software such as MS Word - which uses a 'What You See Is What You Get' (WYSIWYG) interface, LaTeX uses plain text and markup language. As with R and Python, LaTeX has a steep learning curve. So why use it?

- ▶ It's free ...But so is Libre Office
- ▶ It makes very professional looking documents ...But with a bit of work you can imitate the 'LaTeX style' in WYSIWYG word processors

LaTeX

LaTeX is a free to use software system for typesetting documents. Unlike word processing software such as MS Word - which uses a 'What You See Is What You Get' (WYSIWYG) interface, LaTeX uses plain text and markup language. As with R and Python, LaTeX has a steep learning curve. So why use it?

- ▶ It's free ...But so is Libre Office
- ▶ It makes very professional looking documents ...But with a bit of work you can imitate the 'LaTeX style' in WYSIWYG word processors
- ▶ It has become omnipresent in many quantitative academic fields

LaTeX

LaTeX is a free to use software system for typesetting documents. Unlike word processing software such as MS Word - which uses a 'What You See Is What You Get' (WYSIWYG) interface, LaTeX uses plain text and markup language. As with R and Python, LaTeX has a steep learning curve. So why use it?

- ▶ It's free ...But so is Libre Office
- ▶ It makes very professional looking documents ...But with a bit of work you can imitate the 'LaTeX style' in WYSIWYG word processors
- ▶ It has become omnipresent in many quantitative academic fields
- ▶ Many data science applications rely on LaTeX to create outputs

Why you need LaTeX

- ▶ R Markdown uses LaTeX 'behind the scenes' to create PDFs. You will therefore need some version of LaTeX installed on your system if you want to create PDFs through R/R Studio.
- ▶ Because R Markdown files can be difficult to debug, in Stats I and II we will be using LaTeX.
- ▶ Overleaf (`https://www.overleaf.com`) (what I use) is an online, collaborative LaTeX editor.
- ▶ TexStudio is an integrated writing environment for creating LaTeX documents. Download at: `https://www.texstudio.org/`.

GitHub Workflow
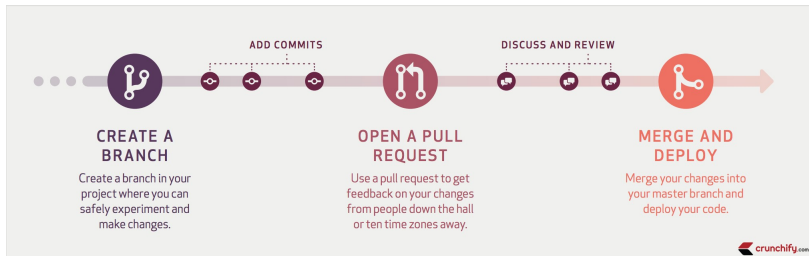
Any Questions?

This concludes Code Camp. Well done on making it through. If you have any additional queries, you can email me at

`schenkem@tcd.ie`