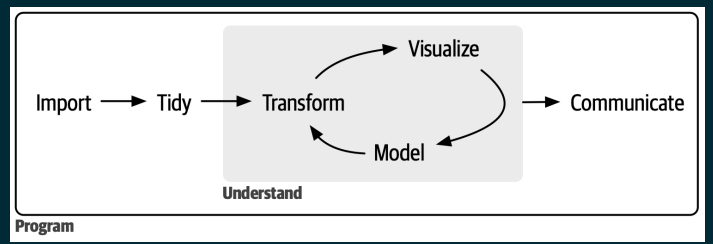


C91AR: ADVANCED STATISTICS USING R

Lecture 1: Welcome to C91AR

Dr Peter E McKenna

2025-01-15

HERIOT WATT UNIVERSITY
SoSS Doctoral Centre

BEFORE WE GET STARTED...

Tell us a little bit about why you want to learn R?

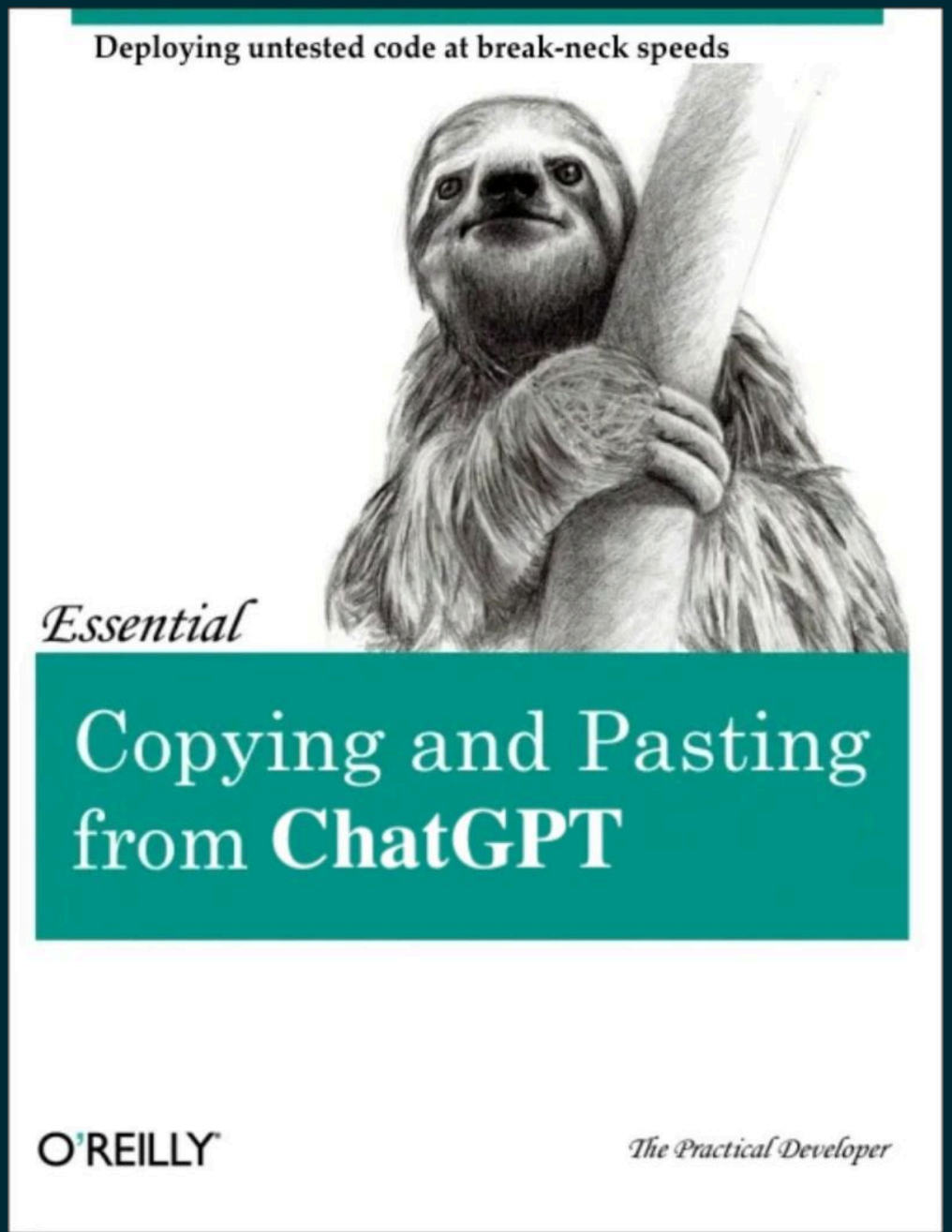
INTRODUCING R

- R is a programming language designed to perform tasks with data.
- It is a popular tool among researchers as it provides a suite of the latest statistical tests.
- It is free and open source, meaning that anyone, anywhere can use it (so long as they have an internet connection).
- R is undergoing continuous development, meaning that it can be used for almost any data task.
- R forums, R social media (e.g., #rstats) and Co-Pilot or ChatGPT offer excellent support.

HERIOT WATT UNIVERSITY
SoSS Doctoral Centre

CHALLENGES

- Learning a programming language takes time.
- A single mistake in your code will throw an error it will not run.
- Because of this, it is guaranteed you will become severely frustrated at times.
- However, as stated, there is lots of support available.
- Think of it like a marathon, not a 100m sprint.



HERIOT WATT UNIVERSITY

SoSS Doctoral Centre

5 LEARNING R TIPS

1. Make mistakes
2. Experiment and try new things
3. Find community (e.g., edinbR on X)
4. Take. Your. Time.
5. Practice on a topic you care about.

COURSE LEARNING OUTCOMES

1. Develop cutting-edge R programming skills for high quality research.
2. Apply R programming techniques to appropriately reconfigure data for analysis.
3. Create visualisations of data trends in R.
4. Create scripts to analyse data in R, including advanced modelling techniques (e.g., data simulation).
5. Interpret the output of simple (e.g., linear regression) and advanced (e.g., multiple regression) statistical analyses performed in R.
6. Compose professionally formatted research reports in R that include code chunks and text summaries.

HERIOT WATT UNIVERSITY
SoSS Doctoral Centre

WHAT ARE THE BENEFITS OF LEARNING A PROGRAMMING LANGUAGE?

According to Duffany (2017) and Smaldone et al. (2022)

- Pick up basic, transferable programming skills.
- Reinforce mathematical concepts.
- Develop your critical thinking and problem solving skills.
- Improve your employment prospects widen to include positions related to Data Science.
- Equip you for the labour market of Industry 4.0.

ANY QUESTIONS?

	I KNOW WHY	I DON'T KNOW WHY
MY CODE RUNS		
MY CODE DOESN'T RUN		

HERIOT WATT UNIVERSITY

SoSS Doctoral Centre

HOW THE COURSE WILL BE DELIVERED

- 1hr Lecture/seminar style session
- 2hr computer labs
- Make sure you have both `R` and `RStudio` installed on your machines before we start.
- These are both available for install on the HWU Software Centre app
- `R` is the software that runs the computation, and `RStudio` is the front end user interface.
- Once you have installed both of these you only need to run `RStudio`
- My teaching philosophy focuses on practical application: the best way to learn R, is to code in R.

HERIOT WATT UNIVERSITY
SoSS Doctoral Centre

SOME THINGS TO BEAR IN MIND

- To perform certain operations in R we have to load different packages, which are kind of like toolboxes.
- This is fairly easy, although you may have to specify the CRAN library.
- If you do, set it to a UK based library (there is one in London and another in Bristol).
- Packages are regularly updated, so R may need time to install updates from time to time.
- Packages can also cause conflicts because they use the same syntax
  - R will notify you about this once the package is installed and loaded.
- *Red text* will often be generated in the R console window
  - Do not be afraid of this. Take your time to read it, as it does not always refers to a code error.
  - Getting used to error messages or warning messages is essential a programming language.

INSTALLING PACKAGES VS LOADING PACKAGES

```
1  # This is what the syntax looks like to install a package
2  install.packages('tidyverse')
3
4  # This is what it looks like to load a package
5  library(tidyverse)
```

- You have to install a package in order to load it
- You only need to install a package once, so typically the install is done using a one off command in the **console** rather than our script (where it install every time we ran the script).

COOL EXAMPLE TO WHET YOUR R APPETITES

- `babynames` dataset
- Full baby name data provided by the SSA. This includes all names with at least 5 uses.
- Five variables: `year`, `sex`, `name`, `n` and `prop` (n divided by total number of applicants in that year, which means proportions are of people of that gender with that name born in that year).

EXPLORE THE DATA A BIT

- The `summarytools` package is better than baseR's summary functions.
- Have a look here for more info.

Code to run `dfSummary` function | Output of `dfSummary`

```
1  # Summarise with the `dfSummary` function from summarytools package
2  dfSummary(babynames_df)
```

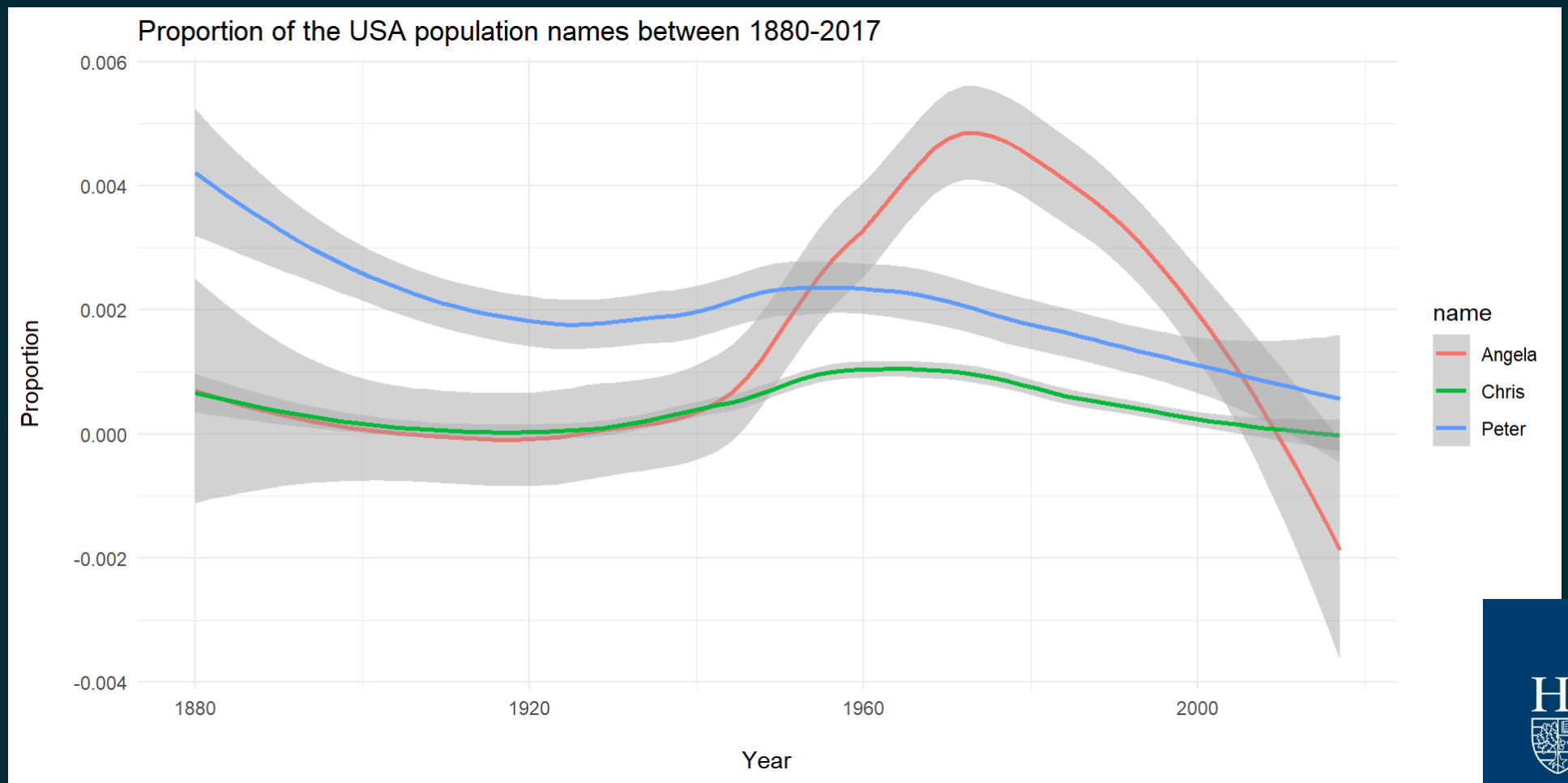
ON THE USE OF PACKAGES

- Anyone using R can develop their own package.
- Some are better than others.
- The one's I include on this course are widely used by scientists across disciplines.
- If you find other packages more suited to your data, by all means use them!

SOME FUN WITH NAMES

```r
# Wrangle a fraction of names dataset from `babynames` dataset
babynames_class <-
  babynames_df |>
  filter(name %in% c("Angela", "Chris", "Peter"))
```

PLOT PROPORTION OF NAMES SPECIFIED ACROSS DATASET YEARS



HERIOT WATT UNIVERSITY
SoSS Doctoral Centre

ADDITIONAL RESOURCES

- If you are keen to learn more about R Programming here is material related to both sessions

  R for Data Science – useful for learning *Tidyverse* syntax

  The R Book – useful for data analysis syntax

  ggplot2 cheatsheet – useful for learning to plot with ggplot2

  Data Wrangling and Tidying cheatsheet – useful for wrangling and tidying syntax of the *Tidyverse*

  R Studio cheatsheets – useful guides for all things R Studio

  Plotting means and error bars – useful guide for plotting means and error bars with

HERIOT WATT UNIVERSITY
SoSS Doctoral Centre

REFERENCES

Duffany, Jeffrey L. 2017. “Application of Active Learning Techniques to the Teaching of Introductory Programming.” *IEEE Revista Iberoamericana de Tecnologias Del Aprendizaje* 12 (1): 62–69. https://doi.org/10.1109/RITA.2017.2658918.

Smaldone, Francesco, Adelaide Ippolito, Jelena Lagger, and Marco Pellicano. 2022. “Employability Skills: Profiling Data Scientists in the Digital Labour Market.” *European Management Journal* 40 (5): 671–84. https://doi.org/10.1016/j.emj.2022.05.005.

HERIOT WATT UNIVERSITY

SoSS Doctoral Centre


HERIOT
WATT
UNIVERSITY
SoSS Doctoral
Centre