TUTORIAL 2: WRITING TIDY CODE

C91AR: Advanced Statistics using R

Dr Pete McKenna

2025-01-28

HERIOT WATT UNIVERSITY
SoSS Doctoral Centre

PACKAGES FOR TODAY

```r
# load packages for session
pacman::p_load(tidyverse,
               summarytools,
               nycflights13)
```

SoSS Doctoral Centre

TUTORIAL 2

- Today's lesson is going to include some:
  - General recap
  - Further exploration and applications of the Tidyverse verbs
  - Paired activity

HERIOT WATT UNIVERSITY
SoSS Doctoral Centre

GENERAL RECAP

General Recap, a stern, quirky chap,

Would bellow, “Forget? That’s a tactical trap!”

He drilled through lessons with a comical flair,

“Lost your memory? Try losing a war—if you dare!”

His motto remains, a soldier’s creed:

“Recall or retreat—your life may concede!”

ChatGPT (2025)

HERIOT WATT UNIVERSITY

SoSS Doctoral Centre

ACTUAL RECAP

- **Always** open the `.RProj` file to get the session started, as this will load us onto the correct directory where our data is located.
- Create a Markdown file for the activities using the format “LectureX_lesson_details.rmd”
- Your R Markdown files will contain both text and R code chunks.
- Make sure you use the `#` key to appropriately header your Markdown file.
- Make sure you use `#` inside code chunks to provide further code clarification using comments

HERIOT WATT UNIVERSITY
SoSS Doctoral Centre

SYNTAX RECAP

- `mutate()`: adds new variables or modifies existing variables.
- `select()`: picks (and amends) variables based on their names.
- `filter()`: subsets cases based on specified conditional criteria.
- `count()`: counts the number of occurrences of unique values.
- `summarise()`: reduces multiple values down to a single summary.
- `arrange()`: changes the ordering of the vectors.
- `|>`: the **pipe operator** means 'then do this'
- `<-`: the **assign** operator to create new variables or data

HERIOT WATT UNIVERSITY
SoSS Doctoral Centre

ON THE ASSIGN OPERATOR

- What is the difference between these two chunks of code?

```r
# Sample 1
short_flights <-
  flights |> filter(air_time < 60)

# Sample 2
flights |>
  filter(air_time < 60)
```

- What possible challenges might arise from these two approaches?
- How can these be handled in Markdown?

HERIOT WATT UNIVERSITY
SoSS Doctoral Centre

REMINDER ABOUT TIDY CODE

```r
# Strive for:
short_flights <-
  flights |> filter(air_time < 60)

# Avoid:
SHORTFLIGHTS <- flights |> filter(air_time < 60)
```

SoSS Doctoral Centre

CODING ALONG DEMO WITH `starwars` DATA

- Continue using the R Markdown we created on Monday (related to tidy code)
- The `glimpse` function is a really nice way to examine the structure of your data

```r
# examine data
glimpse(starwars)

# you could also check the related help section
# ?starwars
```

HERIOT WATT UNIVERSITY
SoSS Doctoral Centre

QUICK SUMMARY WITH `summarytools`

- I will sort the installation issue out!

```r
dfSummary(starwars)
```

HERIOT WATT UNIVERSITY
SoSS Doctoral Centre

DATA EXPLORATION: COUNT THE NUMBER OF UNIQUE SPECIES IN THE DATASET

```r
starwars |>
  summarise(n_species = n_distinct(species))
```

HERIOT WATT UNIVERSITY
SoSS Doctoral Centre

DATA FILTERING: FIND ALL *NON-HUMAN CHARACTERS WHO WEIGH LESS THAN 75KG*

```r
starwars |>
  filter(species != "Human",   # != denotes "not including"
             mass    < 75)
```

HERIOT WATT UNIVERSITY
SoSS Doctoral Centre

GROUPING AND SUMMARISING: CALCULATE THE AVERAGE HEIGHT FOR EACH SPECIES

- `group_by` = Group data by one or more variables

```
starwars |>
  group_by(species) |>
  summarise(mean_height = mean(height, na.rm = TRUE)) # na.rm = TRUE deno
```

HERIOT WATT UNIVERSITY
SoSS Doctoral Centre

PARIED ACTIVITY

- **In pairs**, I'd like you to write some tidy code chunks with the `starwars` dataset
- This dataset is a part of the *tidyverse* package, so it's already loaded
- You will need to refer to the **Syntax Recap** above to write your code
- So, in your pairs, it would be worth having someone consult the syntax, whilst the other writes the code in Markdown

HERIOT WATT UNIVERSITY
SoSS Doctoral Centre

MARKDOWN TASKS

- T1: Using `filter`, subset the data so it contains female characters with blue eyes
- T2: Using `summarise` count the number of different planets represented in the vector `homeworld`
- T2: Using `group_by` and `summarise`, calculate the **average height and weight** of each species
- T4: Using `select`, subset the data by `name`, `gender` and `eye_color`
  - You will need to look up how to use `select` for this one

HERIOT WATT UNIVERSITY
SoSS Doctoral Centre

CLEANUP

```r
# Clear data
rm(list = ls())  # Removes all objects from environment

# Clear packages
p_unload(all)  # Remove all contributed packages

# Clear plots
graphics.off()  # Clears plots, closes all graphics devices

# Clear console
cat("\014")  # Mimics ctrl+L
```

SoSS Doctoral Centre

RECAP

- Tidy coding practice
- Complexity of the assign `<-` operator
- *tidyverse verbs*

NEXT WEEK

- Cleaning messy, realistic data

HERIOT WATT UNIVERSITY
SoSS Doctoral Centre

HERIOT WATT UNIVERSITY

SoSS Doctoral Centre