R: A Hitchhikers Guide to Reproducible Research

- Message in a bottle

Brendan Palmer,

Clinical Research Facility - Cork &

School of Public Health

@B_A_Palmer

CRF-C
HRB Clinical Research Facility Cork

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

School of
Public Health

Daily overview

09.30   **Practical**
        - Catch up on Day 1 and walkthroughs

10.30   **Coffee break**

11.00   **Lecture/Practical**
        - Introduction to the core tidyverse packages, demonstrations and walkthroughs

13.00   **Lunch break**

14.00   **Practical session**
        - Worksheet

15.00   **Lecture/Practical**
        - Lecture and demonstration of ggplot2 for data analysis

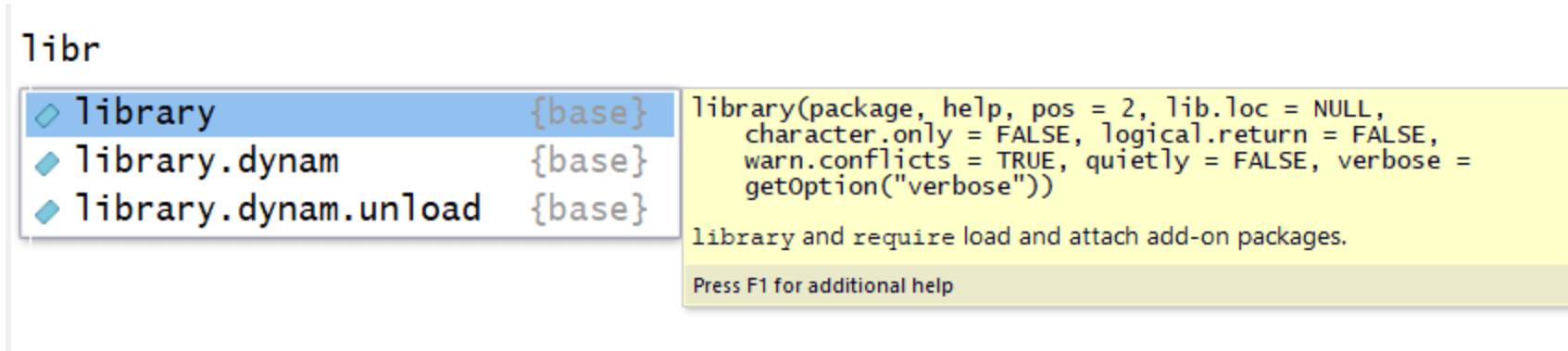**15.30**   **Practical session**
- Worksheet
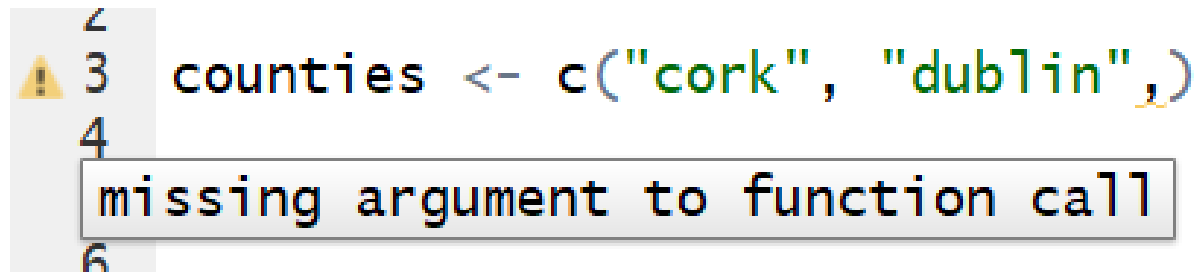
16.30   **Finish**

Rewind

- We took a dive into base R code on Day 1

- Open Day_1/scripts/07_practice_worksheet_solutions.R

- Any burning questions?

Levels of help within RStudio

1. Let R help you write your code using Tab

```
libr
```

```
library              {base}    library(package, help, pos = 2, lib.loc = NULL,
library.dynam        {base}        character.only = FALSE, logical.return = FALSE,
library.dynam.unload {base}        warn.conflicts = TRUE, quietly = FALSE, verbose =
                                   getOption("verbose"))

                               library and require load and attach add-on packages.

                               Press F1 for additional help
```

2. Hover over the error symbols to identify what the error is

```
⚠ 3  counties <- c("cork", "dublin",)
  4
     missing argument to function call
  6
```

```
> counties <- c("cork", "dublin",)
Error in c("cork", "dublin", ) : argument 3 is empty
>
```

Levels of help within RStudio

3. Watch out for capitalisation and naming errors as the code might run, but all the arguments supplied may return "FALSE"

```
> counties <- c("cork", "dublin")
> cork_dublin_df <- hse_df %>%
+                       filter(county %in% counties)
>

> ncol(cork_dublin_df)
[1] 3
> nrow(cork_dublin_df)
[1] 0
> unique(hse_df$county)
 [1] "Carlow"    "Cavan"     "Clare"      "Cork"      "Donegal"    "Dublin"     "Galway"
 [8] "Kerry"     "Kildare"   "Kilkenny"   "Laois"     "Leitrim"    "Limerick"   "Longford"
[15] "Louth"     "Mayo"      "Meath"      "Monaghan"  "Offaly"     "Roscommon"  "Sligo"
[22] "Tipperary" "Waterford" "Westmeath" "Wexford"   "Wicklow"
>

> unique(hse_df$county) %in% counties
 [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[15] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
>
```

Levels of help within RStudio

4. `> ?gather`

   - If you're unsure about a function, place the question mark before it to retrieve help documentation

   - describes the various arguments to the function

   - provides useful examples to guide you through common operations

Walkthrough example script

- 01_correct_the_errors

Package Cheatsheets

- Cheatsheets are available for many commonly used packages

- Day_2 > docs > Cheatsheet-tidyverse_for_beginners.pdf

Package Vignettes

- A vignette is a long-form guide to your package

- Before R version 3.0, vignettes were standard pdfs
  - The development of R Markdown has made vignette building
    and navigation more accessible

- A vignette should divide functions into useful categories,
  and demonstrate how to coordinate multiple functions to
  solve problems (but this may not always the case)

- You can see all the installed vignettes with;
  browseVignettes()

Package Webpages

- Many packages are one offs;
  - developed by individuals/labs to solve specific problems
    - once funding expires, package development ends
      - may get released but never updated

- Popular packages are more stable and updated regularly;
  - new versions released periodically
    - dedicated webpages

ggplot2 part of the tidyverse 3.2.1

Navigating RStudio – some useful tips

- Open the script 02_rstudio_ide_tricks.R