

Efficient Trouble-Shooting

Haohan Chen

POLI3148 Data Science in PPA (The University of Hong Kong)

Last update: October 11, 2023

Trouble-Shooting Tips

What should you do if things go wrong?

- Identify the problem: Read messages from R
- Locate the problem: Where does it occur?
 - Clear, clear, clear
 - Run your code *from the start* and check where it breaks
- Seek assistance
 - Search engine (Google, Bing)
 - Large Language Model (e.g., *ChatGPT*)
 - Human (e.g., me)

Read Error Messages

- You have not loaded your data.

```
d
```

```
## Error in eval(expr, envir, enclos): object 'd' not found
```

- You have not loaded your package.

```
d <- read_csv("vdem_1984_2022_external.csv")
```

```
## Error in read_csv("vdem_1984_2022_external.csv"): could not find function
```

The First Things to Check

When you start coding, here are the first things to check:

- Have you loaded the data you need?
- Have you loaded the packages you need?

Note:

- *R console* and *Render* are two separate systems.
- That things work in one of them does not guarantee they work in the other.

Other Messages: Loading Packages

When you load a package, the printed output shows you information about the package.

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse

## v ggplot2 3.3.6      v purrr  1.0.1
## v tibble  3.2.1      v dplyr  1.1.1
## v tidyr   1.3.0      v stringr 1.5.0
## v readr   2.1.2      v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflict

## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Other Messages: Loading Data

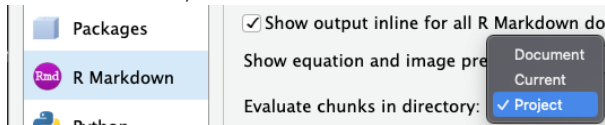
```
d <- read_csv("_DataPublic_/vdem/1984_2022/vdem_1984_2022_external.csv")

## Rows: 6789 Columns: 211
## -- Column specification -----
## Delimiter: ","
## chr      (3): country_name, country_text_id, histname
## dbl      (207): country_id, year, project, historical, codingstart, codingend
## date      (1): historical_date
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Aside: Where to Place Your Data Files

Recap of my previous advice

- Start an R project
- In Rmarkdown, evaluate code chunks in the document from the project directory



- **Open** the R project *every time* you continue working on it
- How to open the R project?
 - File → “Open Project”
 - Click the `.Rproj` file in your R project folder.

Aside: Where to Place Your Data Files

- How to know that you are working from the project?
 - Execute `getwd()` and it is the path of your project
 - DO NOT use `setwd()` to force R to set a working directory *ad hoc*.
- Why use an R project
 - Better organization
 - **Relative path:** However you move your project folder, as long as the *relative* locations of different files remain the same, your code will work.
 - Replicability

Locate the Problem: Clear, clear, clear

What may happen:

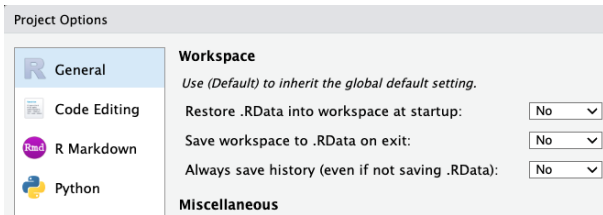
- You get an error message but can't figure why
- You don't get an error message but just can't knit the document.

Why? Your current *Environment* is filled with trash!

Solution: Clear and start over

Locate the Problem: Clear, clear, clear

- Do not save your *Environment* to a .Rdata file. If it is already saved. Delete it!
- Set your project preference so that Rstudio remember not to save your *Environment*.
- Frequently **clear** your Environment and start over
 - to avoid confusion
 - to mimic how your code will run in the *Render* environment.

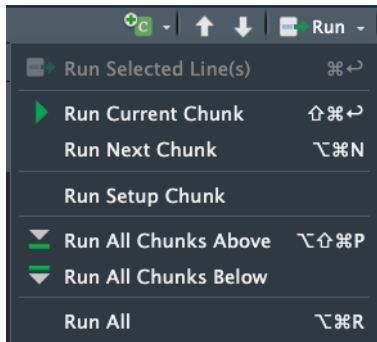


Locate the Problem: Clear, clear, clear

- Make sure your environment is completely empty
- To empty the environment
 - Method 1: Execute `rm(list = ls())` in the *Console*
 - Method 2: Click the broomstick in the *Environment* panel
 - Method 3: Close and restart Rstudio
- Differences between the three methods: 1 and 2 clear loaded *objects* but *not packages*. Loaded packages remain in your R console. 3 clear loaded objects, packages, and free up memory. Consider Method 3 a “deep clean.”

Locate the Problem: Run Your Program

- “**Run All Chunks Above**” and “**Run All**” are handy tools.



- The program will run until error occurs.
- Where the error occurs is where you should double-check carefully.

- “R” + package name (e.g., `tidyverse`) + function’s name
- Include the error message
 - Remove the part that is specific to your case (file name, object name)
 - **Example:** Error in `eval(expr, envir, enclos)`: object ‘`d`’ not found
 - **Example:** Error in `read_csv(“vdem_1984_2022_external.csv”)`: could not find function

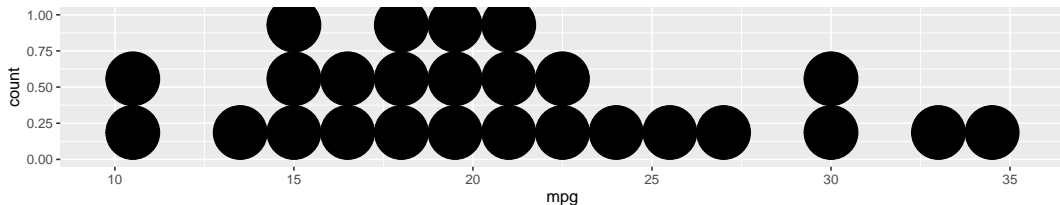
Seek Assistance: Large Language Model

- Similar to search engine: Clearly indicate the programming language, package, function's name, and the error message.
- Doesn't harm to start with pasting your whole chunk of code and ask the model to find mistakes for you.

- Provide a description of the problem using a similar way with the previous approach: package name, function, and error message.
- Provide the part of code where error occurs.
- Preferably, make your code replicable.
 - Make it in a way that the person who tries to assist you can **copy** the code to their R, run it, and see the problem.
 - It uses some data? Include some toy/ example data. It can come from the data you are working on. Or you can use some of R's default datasets.

Example: A (very well-written) question on CampusWire about `geom_dotplot` could have been asked with the following example by R's default dataset `mtcars`:

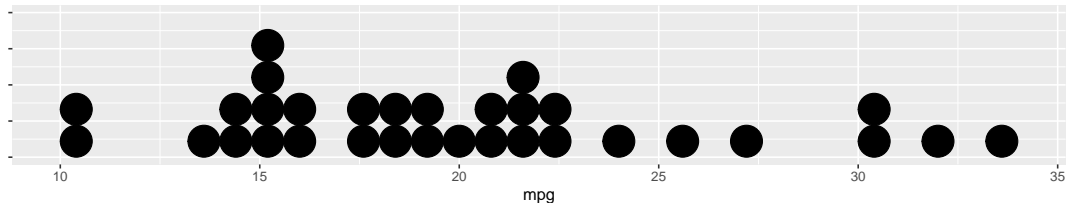
```
ggplot(mtcars, aes(x = mpg)) +  
  geom_dotplot(method="histodot", binwidth = 1.5)
```



Seek Assistance: Human

Then, the person who gives an answer copy the code to their R, suggest changes, and send the code back.

```
ggplot(mtcars, aes(x = mpg)) +  
  geom_dotplot(method="histodot", binwidth = 0.8) +  
  theme(axis.text.y = element_blank(), axis.title.y = element_blank())
```



- The “clean and start over” strategy helps more than you would expect.
- If you don't get any relevant answer, it is rarely the case that the problem you get is unique. It is more likely that you have asked the wrong question.
- Do not code for long hours
 - It will be counter-productive.
 - You will get grumpy and dizzy.

One more thing...

These slides are made with Rmarkdown.

- Check out my code on our course Github repo. Play with it.
- Prefer different themes, colors, and fonts?
https://deic.uab.cat/~iblanes/beamer_gallery/index.html
- A general introduction to making slides with Rmarkdown
<https://rmarkdown.rstudio.com/lesson-11.html>
- A cool theme for Rmarkdown presentation
<https://github.com/yihui/xaringan>