



Automating Routine Analyses in R

More info: micom-hub.github.io

Outline

- Introduction
- Case study

If time:

- A note about debugging
- Outputs for your automation! Intro to quarto

Planning for automating a task: questions to ask yourself

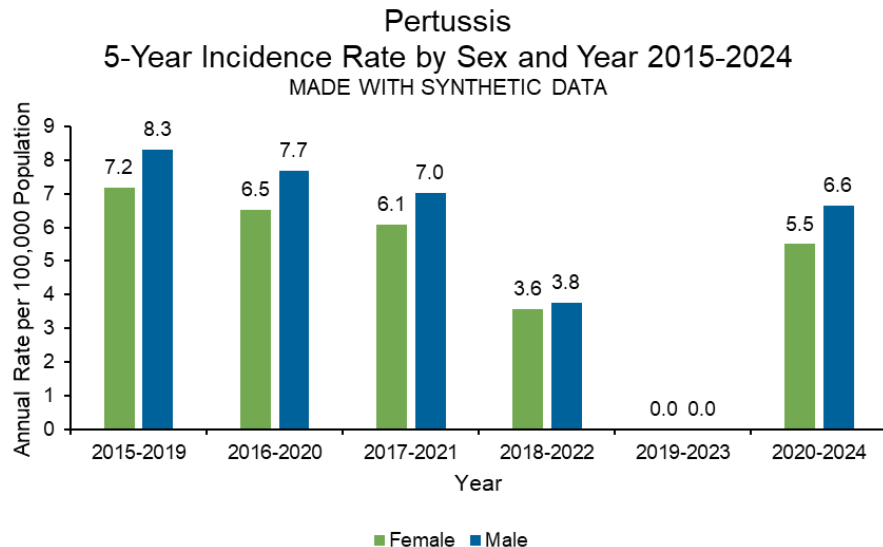
- How consistent is the process you want to automate? Do you do basically the same steps each time?
- Think through and write down the logic you use to do the task—this will be your guide to writing the R script for it!
- What are some checks to implement inside the data? How to make sure that the data and the output are behaving the way you want?
- What is the output you want your script to generate? A figure or table? Or an html report, word doc, pdf, etc?

Additional considerations

- Document, document, document!
 - You will definitely need to update/change the script at some point and future you will thank you for the documentation!
- Data alerts, warnings, or other notes to mark progress through code
- Test the code! Run the old way and the automated way and make sure you get the same answers—once the script is catching human errors you know you're in good shape

Case Study!

Suppose we want to make a summary plot like this from an MDSS pull of pertussis cases:



How would we translate the steps you'd do in Excel and automate them in R?
Let's look at some examples!



From our procedure:

- Copy all data [from the MDSS export] into a `pertussis_cases` tab...



Consider these functions:

- `read_csv()`
- `this.path::here()`
- `mutate()`
- `mdy()`



From our procedure:

- Select the “Insert” section of the ribbon and select “PivotTable” in the “Tables” section.
- In the “PivotTable Fields” sidebar, click and drag **Diagnosis_Date** into the “Rows” box, and ensure “Years” is the top level of the hierarchy.



Consider these functions:

- `cross_join()`
- `tibble()`
- `min()`, `max()`
- `unique()`

- `group_by()`, `summarize()`, `n()`
- `*_join()`, `join_by()`
- `replace_na()`



From our procedure:

- In column H, label a variable as **Female** and use `=SUM()` in the cell below this label to sum the corresponding values for this sex and [the 2015-2019] period. ...
- In two new variables... copy population values corresponding to the sex and final year of each time period



Consider these functions:

- `pivot_wider()`
- `slider::slide_dbl()`
- `pivot_longer()`
- `*_join()`



From our procedure:

- Select the “Insert” section of the ribbon, expand the “Charts” subsection, and insert a “Clustered Column” chart.
- ... use the “Chart Design” section of the ribbon to achieve desired aesthetics



Consider these functions:

- `ggplot()`
- `geom_col()`, `position_dodge()`
- `scale_fill_manual()`
- `geom_text()`
- `theme_classic()`
- ... too many to list here
 - This is where we start googling

Case Study!

Suppose we want to make a table that has each month from 2015-2024 as rows, has each sex as a column, and has the number of cases for that sex in that month as values.

Let's try writing out the steps that we think we'll need to follow before jumping back into RStudio!

Homework

Pick a task that you do on a regular basis and try this process out!

1. Do the task, paying attention to the steps and the logic you're using
2. Write out the steps
3. Convert them into code!

(If you need help with step 3, come join a ML-Support office hour!)

A few more things if time!

How do I debug my code?

- Debugging - crucial skill in programming
- Many ways to do it
- Best way to learn is lots of practice!
- But here are some common methods & useful tips to get you started

How do I debug my code?

The computer just does what you tell it!

Questions to ask yourself:

- What did you expect your code to do and why?
- What did your code do instead and how do you know?

How do I debug my code?

- Use the error messages—if nothing else, they give you a line number
- Don't forget to scroll all the way up to the first error!
- Find where the problem appears to be (note it might not actually be there!)
- Silent errors are the real villains - often your code will run perfectly fine, but things may still be incorrect.

How do I debug my code?

- Follow through the steps of the code as though you are the computer
- Use print statements!
- Hypothesize what the problem could be, then test it

SOME FUN: WHY ARE THEY CALLED BUGS?

9/2
9/9

0800 Antan started
1000 " stopped - antan ✓ { 1.2700 9.037 847 025
1300 1032 HP-MC 2.1304776415 9.037 846 985 correct
033 PRO 2 2.1304776415 4.615925059(-2)
convd 2.1304776415
Relays 6-2 in 033 failed special speed test
in relay 10,000 test.

1400 Started Cosine Tape (Sine check)
1525 Started Multi-Adder Test.

1545  Relay #70 Panel F
(moth) in relay.

1630 Antan started.
1700 closed down.

Relay
2145
Relay 3370



Quarto!

Not sure how likely we are to get here, but: https://epimath.org/wchd-R-workshop/quarto_introduction.html#/title-slide