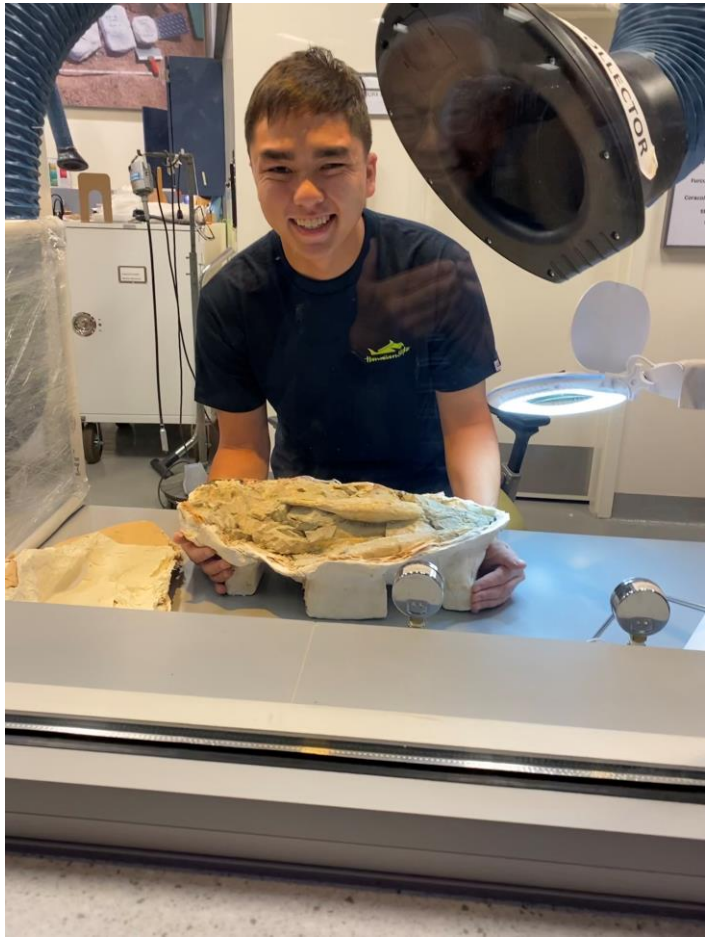
Data Manipulation and Analysis in R

By Travis Oishi

About Me

- Central Shenandoah Health District Epidemiologist
- Jr Epi for Central Shenandoah from Oct 2019 - March 2022
- Limited experience in R when I first started in 2019
- Started coding around Spring of 2020 to improve epi processes
- travis.oishi@vdh.virginia.gov



Type in the chat

- 1: You've been able to use coding in your work routine
- 0: You haven't had the opportunity to use coding in your work routine (But you're interested!)

https://www.shutterstock.com/search/man-raising-hand

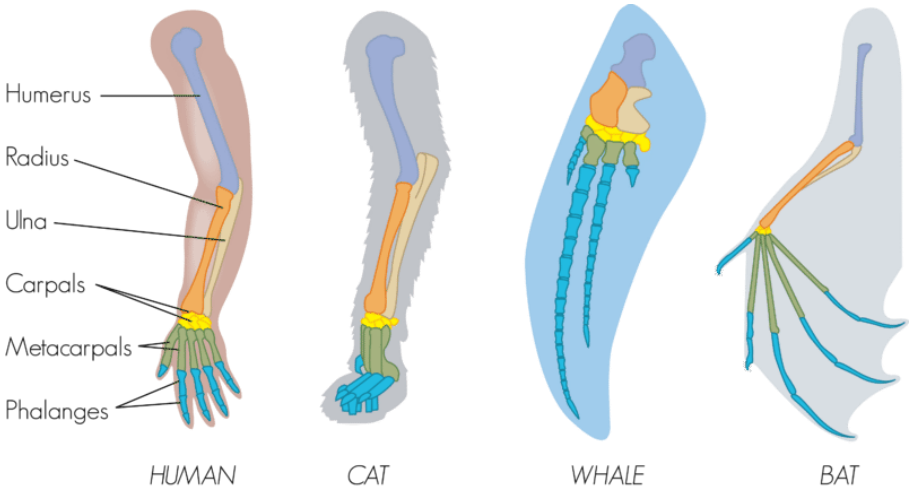
Is coding worth it?

- How long does it take to complete the task manually?
- How frequently is the task performed?
- How long do you think it will take to develop a code?
- How many new techniques will have to be learned?
- Who will this code benefit?

https://www.istockphoto.com/photos/stressed-man-sitting-at-his-desk-in-front-of-computer

Long Term Benefits

- Automate tedious tasks
- Recycle old concepts from previous codes
- New codes might only require learning a few new concepts
- Coding becomes faster



https://www.sanibelseaschool.org/experience-blog/2020/6/24/homology-and-analogy-a-lesson-in-biology

How do you get two datasets to talk to one other?

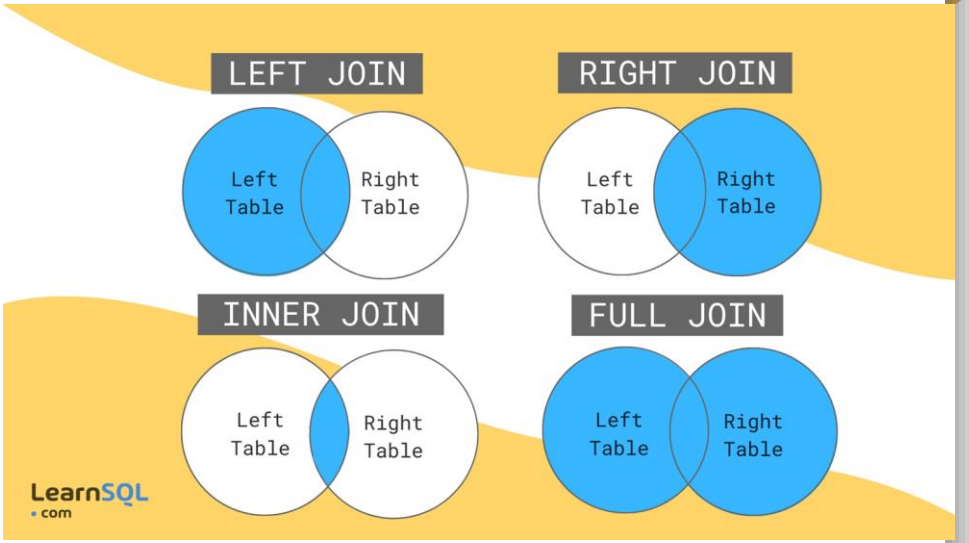
Combining Datasets

- **Joins:** Combine datasets by shared columns
  - Returns rows only when certain conditions are met
- **Binds:** Combine datasets by either their rows or columns
- Make sure the variables match!

https://www.dreamstime.com/royalty-free-stock-image-bad-communication-image1364416

Types of Joins

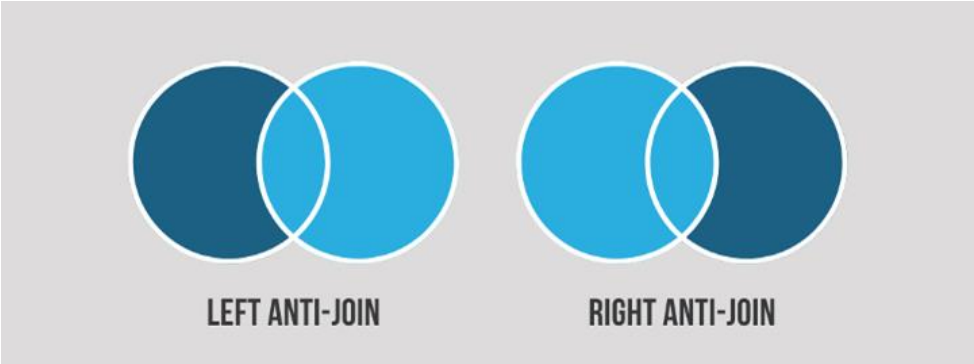
- Left Join: Returns all rows from the **left** table, even if no matching rows have been found in the right table
- Right Join: Returns all rows from the **right** table, even if no matching rows have been found in the left table
- Inner Join: Returns only matching rows found in **both** tables
- Full Join: Returns **all** rows from both tables



https://learnsql.com/blog/learn-and-practice-sql-joins/

Types of Joins cont.

- Anti Join: Returns the rows that are present on one dataset but **missing** in the other
- Good for comparing a new list to an ongoing list

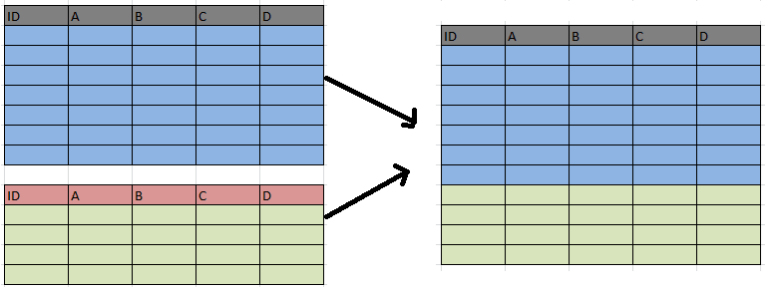


https://towardsdatascience.com/why-is-nobody-talking-about-sql-anti-joins-f970a5f6cb54?gi=1aeaee5e5a37

Binds

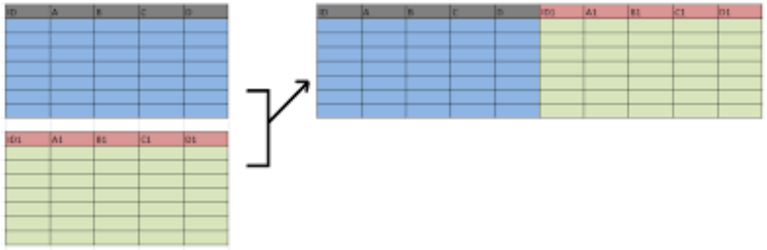
Row Bind

- Merge two datasets together by rows
- Good for adding rows to an existing table
- Adds new rows to the bottom of the table



Column Bind

- Merge two tables based on shared columns
- Similar to joins
- Joins are better (in my opinion)



Data Types

- Character: can contain the alphabets, numbers, and symbols
- Numeric: Number values
- Integer: Negative and positive whole values
- Date: Date values

https://r4ds.had.co.nz/vectors.html

https://www.alamy.com/stock-photo/square-peg-and-round-hole.html

Useful Resources

- https://r4ds.had.co.nz/vectors.html
- https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch8/5214817-eng.htm
- https://dplyr.tidyverse.org/index.html
- https://dplyr.tidyverse.org/reference/mutate-joins.html
- https://dplyr.tidyverse.org/reference/bind.html?q=rbind#null
- https://r4ds.had.co.nz/data-visualisation.html
- https://www.kaggle.com/competitions/titanic/data