BST 219
Core Principles of Data Science

Lecture 17: Advanced Data Wrangling continued
October 29, 2024

Recipe of the Day!

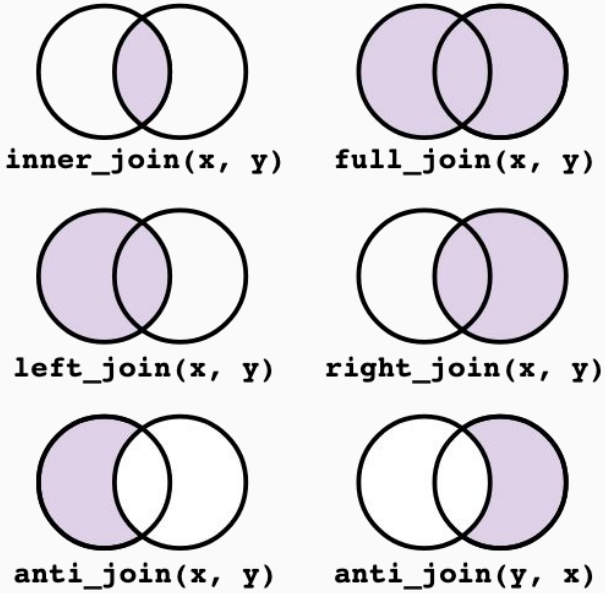
Quinoa stuffed butternut squash with cranberries and kale

When your head is too big for not one, but two costume hats

Agenda

- Announcements
  - Lab this week!
  - Homework 3 is due 11/8
  - Homework 2 grades by 11/4
  - Midterm 11/8 - 11/17

- Continue the advanced data wrangling module
  - Combining tables
  - Dates and times
  - Publishable tables



Special Announcements

- Heather's **11/5** office hour will be moved to **12-1pm**
- Heather's **11/12** office hour will be moved to **12-1pm**
- The 11/14 lecture will be moved to **11/13, 12:30-2pm via Zoom**
- Lecture on **11/26** will be held via **Zoom**
- Heather's office hour on **11/26** will be **Zoom only**

Coding Question of the Day!

Using the **`murders`** dataset from the **`dslabs`** package and the **`state.x77`** dataset that is built in to R, make a scatterplot with the total number of gun murders (**`total`**) on the y-axis and land area in square miles (**`Area`**) on the x-axis. You will need to use the **`left_join`** function to join the **`state.x77`** dataset to the **`murders`** dataset using the **state** column.

Bonus challenge: are there any NAs in the merged dataset? How are they created?

Make sure to run this code first

```r
# Load necessary libraries
library(dplyr)
library(dslabs)
library(ggplot2)

# Load the state.x77 dataset and convert to a data frame
data("state")
state_data <- as.data.frame(state.x77)

# Add the state names to the state.x77 dataset
state_data$state <- rownames(state_data)

head(state_data)

# Load the murders dataset from the dslabs package
data("murders")
```