R Review Exercises

Exercises to Review R

1. Vectors

- Establish a vector of numbers from 1 to 99 in steps of 15. Hint: use the `seq()` function
- Add a character item to your vector, something like "foo"
- What is the type of your vector after adding the item?
- Coerce the vector back to a double. What happens to the item you added?

2. Importing data

- read in the Influenza_Vaccination_Coverage data (inside the Influenza sub-folder). You can find the original data here:
- https://data.cdc.gov/Flu-Vaccinations/Influenza-Vaccination-Coverage-for-All-Ages-6-Mont/vh55-3he6
- After you read in the data, examine the top and tail of the data.
- Summarize the data values.
- Summarize the data structure.

3. Data verbs

- In the flu vaccine dataset, how many unique values of "Geography" are there?
- In the same data, how many rows are there of the Season unique values? How many rows of unique Months?
- Filter the records down to Colorado: how many rows are there pertaining just to Colorado (as a Geography entry)?
- Filter to Colorado and group on Vaccine. How many rows are there per Vaccine type in Colorado?

4. Cleaning column names to summarize

We'd like to do some data summaries with the Vaccination data set, but some of the column names and data types are not ideal.

- rename Geography Type to geo_type

- rename other columns that have spaces and punctuation. We like the `janitor` package for this task– it has a function called `clean_names()`. Read the help on the function, and then apply the function to your tibble
- we'd like to summarize our data using the "estimate_percent" column, but its data type may not be correct. Check the data types of the tibble. Convert the data type of the estimate column.
  - There are many ways you can accomplish this. Find the "parse_number()" function and apply it to this column

5. Summarizing data

- Now try to produce the following summaries. These may take a bit of time to figure out.
  - Filter to States/Local Areas
  - Filter to the non-combined vaccine types (filter *out* "Any vaccine...")
  - Filter to dimension type "Age"
  - Summarize the average vaccination rate by vaccine type
  - Summarize the averate vaccination rate by vaccine type by state
  - Filter to one state and present vaccination rates by vaccine type and age group
- Bonus: Can you pivot the results to present the data in an effective way?

6. Extended: Select Certain Age Groups

Let's inspect and organize the Age groups.

First, display the unique levels of the Age dimension.

Find the "Years at High Risk" categories. Can you detect these programmatically? Hint: use `stringr's str_detect` to filter on string patterns.

7. Extended: Summarize Vaccination Rate for High Risk across States

Building on what you produced above, compare average vaccination rates ("Any Influenza Vaccination") for "18-64 Years at High Risk" across the states Colorado, Wyoming, and Montana. If you have additional time, continue to explore and summarize the data.