Assignment 4: A Rose By Any Other Name

Learning Outcomes

- Study how to concat data sets
- Study how to pivot data sets

Motivation

Every expectant parent in America has a problem provided to them by the government: they must name their child. Naming a child is no simple task. Parents want the name to be both timeless and trendy, satisfy the requirements of relatives, and be something that they won't regret screaming during the child's teenage years. It's a big problem.

When an American child is born, the Social Security Administration (SSA) blesses the occasion in the way governments do best: with a number. This assignment is not about SSNs. It's about names. Ever since the SSA was created back in 1910, they've kept a record of every single baby, their sex, the political state in which they were born, and their name. They've published all of that data on the ssa.gov website. I want you to download the State-specific data. This contains over 100 years of history and is about 20 MB in size. This is a famous data set that always generates a few peer-reviewed papers every year. (John McCain famously joked that his social security number was 2.)

- https://www.ssa.gov/OACT/babynames/limits.html. Make sure that you download the State-specific data.

Upon unzipping the file, you will have 51 text files organized by their two-letter state code (50 states plus the District of Columbia).

Here are each of the columns in each file:

- State (2 letter code)
- Sex (F for female, M for male)
- Year (4 digit number)
- Name
- Count (number of times a baby was born in this state with this name)

Unfortunately there is no header to these files. I provided my own column headers to each of the files required below: "state","sex","year","name","count"

Directions

1. Open 4 different states. The arbitrary requirement here is that three of the states much touch a fourth state. This way we can study a geographic area. You may add your headers manually to each file.
2. Concatenate each of your four data sets into a single datasets.
3. How many females were born in your four selected states? How many males were born? Use a pivot table to create your answer and select the appropriate aggregation method.
4. How many people were born in each state? Again, use a pivot table.
5. What are the top 5 most popular names for any sex? Again, use a pivot table.
6. What are the top 5 most popular female names? (Use a pivot table to answer this.)
7. What are the top 5 most popular male names? (Use a pivot table to answer this.)

Turn it in

Prepare a document of each of your commands and turn it in.