Working with Strings: Homework

Now let’s practice some of what we learned here using a dataset of names from the US Social Security Administration. The babynames library on CRAN contains a dataset of baby names from 1880 to 2017, pulled from the US Social Security Administration’s web interface. First, we’ll do some basic practice with string manipulation on its own. Then, we’ll integrate the skills from this lesson with dplyr and ggplot.

1. Let’s analyze the set of names given to children born in 2017. To answer this question, you’ll need make a vector that contains every name from 2017. You can do this either with filter() or with pull() from dplyr.
   1. Which vowel is the most common in the first letter of names from 2017?
   2. Do more people name their children names that are longer than 5 letters or shorter than 5 letters?
   3. What are the names that have at least 14 letters? Don’t use str\_length().
   4. What names have more than two vowels in a row? Are there any names with more than three vowels in a row?
   5. What names contain the longest sequential string of consonants?
2. In the most recent year recorded in the dataset (2017), what are the 10 most popular names that start with Z? What about the letter Q?
3. What names from 2017 have the highest number of vowels? How many names end in vowels?
4. The name “Mary” was the most popular girls name from 1880-1946. Track the popularity of the name overtime using a line graph. *Hint: filter for only female instances of Mary for a clearer graph.* What do you notice?
   1. What is the most popular female name in 2017? Track that name’s popularity over time.
5. Peyton is a name that is used for both girls and boys. Plot the popularity of the name by gender. What do you notice?