

Problem Set 1

ECON 480 — Fall 2020

Due by 11:59 PM Sunday September 6, 2020

The Popularity of Baby Names

Install and load the package `babynames`. Get help for `?babynames` to see what the data includes.

1.

a. What are the top 5 boys names for 2017, and what *percent* of overall names is each?

b. What are the top 5 *girls* names, and what *percent* of overall names is each?

2. Make two barplots, of these top 5 names, one for each sex. Map aesthetics `x` to `name` and `y` to `prop`¹ and use `geom_col` (since you are declaring a specific `y`, otherwise you could just use `geom_bar()` and just an `x`.)

¹Or `percent`, if you made that variable, as I did.

3. Find your name.² `count` by `sex` how many babies since 1880 were named your name.³ Also add a variable for the percent of each sex.

4. Make a line graph of the number of babies with your name over time, colored by `sex`.

²If your name isn't in there, pick a random name.

³Hint: if you do this, you'll get the number of *rows* (years) there are in the data. You want to add the number of babies in each row (`n`), so inside `count`, add `wt=n` to weight the count by `n`.

5.

a. Make a table of the most common name for boys by year between 1980-2017.⁴

b. Now do the same for girls.

⁴Hint: once you've got all the right conditions, you'll get a table with a lot of data. You only want to slice the 1st row for each table.

6. Now let's graph the evolution of the most common names since 1880.
 - a. First, find out what are the top 10 *overall* most popular names for boys and for girls. You may want to create two vectors, each with these top 5 names.
 - b. Now make two **linegraphs** of these 5 names over time, one for boys, and one for girls.

7. **Bonus (hard!)**: What are the 10 most common “gender-neutral” names?⁵

⁵This is hard to define. For our purposes, let's define this as names where between 48 and 52% of the babies with the name are Male.

13. Let's do this again, but highlight some key countries. Pick three countries, and make a new tibble from `freedom` that is only the observations of those countries. Additionally, *install* and *load* a packaged called `ggrepel`⁹ Next, redo your plot from question 11, but now add a layer: `geom_label_repel` and set its `data` to your three-country tibble, use same `aesthetics` as your overall plot, but be sure to add `label = ISO`, to use the ISO country code to label.¹⁰

14. Make another plot similar to f, except this time use GDP per Capita (`gdp`) as `y`. Feel free to try to put a regression line with `geom_smooth()`!¹¹ Those of you in my Development course, you just made my graphs from Lesson 2!

⁹This automatically adjusts labels so they don't cover points on a plot!

¹⁰You might also want to set a low `alpha` level to make sure the labels don't obscure other points!

¹¹If you do, be sure to set its `data` to the full `freedom`, not just your three countries!