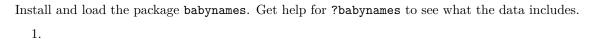
Problem Set 1

ECON 480 — Fall 2021

Due by Class Tuesday September 14

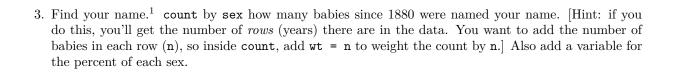
The Popularity of Baby Names



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 for 2017, 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 [or percent, if you made that variable, as I did.] and use geom_col (since you are declaring a specific y, otherwise you could just use geom_bar() and just an x.)				



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

¹If your name isn't in there :(, pick a random name.

a. Find the most common name for boys by year between 1980-2017. [Hint: you'll want to first group_by(year). Once you've got all the right conditions, you'll get a table with a lot of data. You only want to slice(1) to keep just the 1st row of each year's data.]

b. Now do the same for girls.

- 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 in the data. [Hint: first group_by(name).] You may want to create two objects, each with these top 5 names as character elements.
- b. Now make two linegraphs of these 5 names over time, one for boys, and one for girls. [Hint: you'll first want to subset the data to use for your data layer in the plot. First group_by(year) and also make sure you only use the names you found in Part A. Try using the %in% command to do this.]

7. **Bonus** (a challenge!): What are the 10 most common "gender-neutral" names? [This 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.]

Political and Economic Freedom Around the World

For the remaining questions, we'll look at the relationship between Economic Freedom and Political Freedom in countries around the world today. Our data for economic freedom comes from the Fraser Institute, and our data for political freedom comes from Freedom House.

- 8. Download these two datasets that I've cleaned up a bit: [If you want a challenge, try downloading them from the websites and cleaning them up yourself!]
- econ_freedom.csv
- pol_freedom.csv

Below is a brief description of the variables I've put in each dataset:

Econ Freedom

Variable	Description
year	Year
ISO	Three-letter country code
country	Name of the country
ef_index	Total economic freedom index (0 - least to 100 - most)
rank	Rank of the country in terms of economic freedom
continent	Continent the country is in

Pol Freedom

Variable	Description
country C/T	Name of the country Whether the location is a country (C) or territory (T)
year status fh_score	Year Whether the location is Free (F), Partly Free (F) or Not Free (NF) Total political freedom index (0 - least to 100 - most)

Import and save them each as an object using my_df_name <- read_csv("name_of_the_file.csv"). I suggest one as econ and the other as pol, but it's up to you. Look at each object you've created.

9. Now let's join them together so that we can have a single dataset to work with. You can learn more about this in the 1.4 slides. Since both datasets have both country and year (spelled exactly the same in both!), we can use these two variables as a key to combine observations. Run the following code (substituting whatever you want to name your objects):

```
freedom <- left_join(econ, pol, by=c("country", "year")</pre>
```

Take a look at freedom to make sure it appears to have worked.

10.

- a. Make a barplot of the 10 countries with the highest Economic Freedom index score. You may want to find this first and save it as an object for your plot's data layer. Use geom_col() since we will map ef_index to y. If you want to order the bars, set x = fct_reorder(ISO, desc(ef_index)) to reorder ISO (or country, if you prefer) by EF score in descending order.
- b. Make a barplot of the 10 countries with the highest Freedom House index score, similar to what you did for Part A.

11. Now make a scatterplot of Political freedom (fh_score as y) on Economic Freedom (ef_index as x) and color by continent.

12. Save your plot from Question 11 as an object, and add a new layer where we will highlight a few countries. Pick a few countries (I suggest using the ISO code) and create a new object filtering the data to only include these countries (again the %in% command will be most helpful here).

Additionally, **install** and **load** a package called "ggrepel", which will adjust labels so they do not overlap on a plot.

Then, add the following layer to your plot:

This should highlight these countries on your plot.

13. Let's just look only at the United States and see how it has fared in both measures of freedom over time. filter() the data to look only at ISO == "USA". Use both a geom_point() layer and a geom_path() layer, which will connect the dots over time. Let's also see this by labeling the years with an additional layer geom_text_repel(aes(label = year)).