Problem Set 1

Ryan Safner ECON 480 - Fall 2019

Instructions

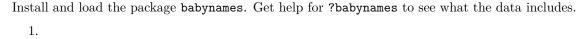
For this problem set, you may type up your answers to each question (save plots as images and include them with your answers) and also turn in a (commented!) R file of commands for each relevant question. You may email or turn in a paper copy. It is not relevant for this particular problem set, but in the future, I will allow handwritten answers for math/theory questions.

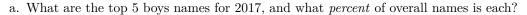
Alternatively, you may download the .Rmd file, do the homework in markdown, and email to me a single knitted html or pdf file (and be sure that it shows all of your code).

To minimize confusion, I suggest creating a new R Project (e.g. hw1) and storing any data and plots in that folder on your computer. See my example workflow.

You may work together (and I highly encourage that) but you must turn in your own answers. I grade homeworks 70% for completion, and for the remaining 30%, pick one question to grade for accuracy - so it is best that you try every problem, even if you are unsure how to complete it accurately.

The Popularity of Baby Names





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



3.	Find your name.	² count by	sex how	many	babies	since	1880	were	named	your	${\rm name.}^3$	Also	add a
	variable for the p	ercent of ea	ch 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.

a. Make a table of the most common name for boys by year between $1980\text{-}2017.^4$

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. I	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

 $^{^5}$ 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: 6
- econfreedom.csv
- freedomhouse2018.csv
- 9. Load them with df<-read_csv("name_of_the_file.csv") and save one as econfreedom and the other as polfreedom. Look at each tibble you've created.
- 10. The polfreedom dataset is still a bit messy. Let's overwrite it (or assign to something like polfreedom2) and select Country. Territory and Total (total freedom score) and rename Country. Territory to Country.
- 11. Now we can try to merge these two datasets into one. Since they both have Country as a variable, we can merge these tibbles using left_join(econfreedom, polfreedom, by="Country")⁷ and save this as a new tibble (something like freedom).

12. Now make a scatterplot of Political Freedom (total)⁸ as y on Economic Freedom (ef) as x and color by continent.

⁶If you want, try downloading them from the websites yourself!

⁷Note, if you saved as something else in question 9., use that instead of polfreedom!

⁸Feel free to rename these!

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 ggrepel9 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!