# Viz Assignment 2

Nick Sun

April 29, 2019

In assignment 1, I created the following stacked bar chart using the `babynames` dataset. In my exploration phase, I identified 200 names that were not uncommon and had an approximately equal ratio of males and females. I grouped these names together as "gender neutral names". The goals of this chart were to:

1. Show how the number of Americans born with gender neutral names has trended over time
2. Show how the gender proportion of Americans with gender neutral names has trended over time
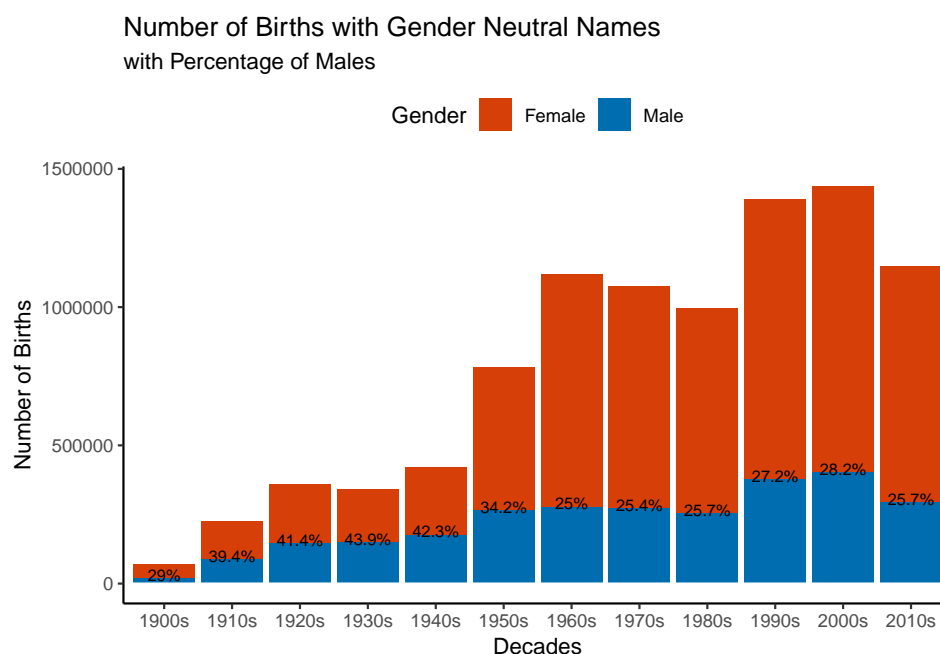
Unsurprisingly, there are a few issues with this plot.

First, one of my goals was to demonstrate how the gender proportion changes over time. It is tough to tell just from looking at each stacked bar if the proportion of females is increasing or decreasing because the area representing each proportion does not start at the same point. According to Cleveland and Mcgill (1984), *position along a common axis* is the easiest perceptual task for numeric data and we can take advantage of this by changing the stacked bar chart to a side-by-side bar chart with `postion=dodge`.
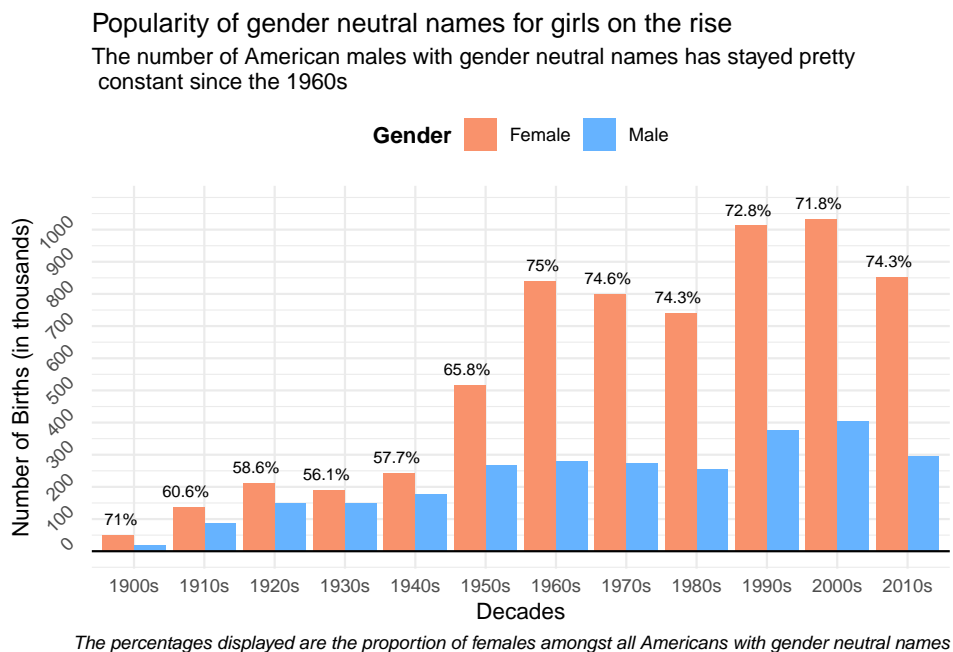
Second, we chose to plot the percentage of males out of the total population of gender neutral named Americans. In terms of writing an interesting conclusion for the plot, we are probably more fascinated by the increasing trend of females with gender neutrals names. Additionally, we put the `geom_text` layer on top of the male bar, making the `geom_text` layer hard to read. We would probably be better served putting it outside of the bars and against a light background. When we create our side-by-side barplot, we can actually place the text on top of the female bar which makes it much easier to read.

Third, the color selection here violates some of our previously discussed color guidelines. These while these bar charts are not as big as area or map plots, it is probably better to use less saturation. Additionally, the red and blue color scheme is tiring on the eyes, and we might be better served picking different hues

altogether. We should make sure to select colors that are also color blind friendly - a lighter orange and blue are good choices according to `colorblindr`.

Lastly, this plot could use polishing. The y-axis labels are cumbersome and could do with some rescaling and tidying up to make it more readable. Lastly, we can also definitely use some guiding gridlines within the plot area itself.

After making these changes, we get this plot which is much nicer to digest!

Popularity of gender neutral names for girls on the rise
The number of American males with gender neutral names has stayed pretty
constant since the 1960s

Gender ▮ Female ▮ Male

The percentages displayed are the proportion of females amongst all Americans with gender neutral names

Now it's much easier to visualize the main story points of the plot: the number of men with gender neutral names has started relatively constant while among females, gender neutral names have taken off in popularity!

Moving the `geom_text` on top of the bar for female births was accomplished by using the `hjust` and `vjust` options and I think really helps the ease of readability and interpretation.

The y-axis scaling was accomplished by using the `labels` option in `scale_y_continuous` and the angling in the `axis.text.y` component of theme. It saves space on the plot and again allows for easier readability.

Lastly, the color selection was done via trail and error with the `colorblindr::cvd_grid` function and some pointers from a "Points of view: Color blindness", B. Wong, Nature Methods **8**, 441, (2011). The palette is friendly to the three main types of color blindness in order to make the plot as accessible as possible.