Makeover Monday Mini-Project for Data Science
Fall 2018 – Prof. Posner

Purpose: In this mini-project, you are asked to use your data wrangling and visualization skills to reproduce a graph found in the media and suggest improvements. It's also useful to develop skills in working with groups of programmers. You can decide how you want to manage sharing your markdown files (options include splitting up tasks and working independently, sharing via Dropbox or Google or the like, using Github, etc.)

Data Set: The data can be found on the Makeover Monday website. Please get your dataset approved (via email) with me at least one week prior to your presentation. Of note, I recommend you not look at any other visualizations besides the original one that have been created for the data you have chosen. You should NOT replicate someone else's improved visualization, but rather think on your own about how to improve it.

Goal:. Recreate a visualization and offer improvements to it.

1) Identify a visualization and dataset of interest from the Makeover Monday website.
2) Recreate the visualization from the media (to the best of your ability).
3) Suggest an improvement on the visualization. You are welcome, but not required, to include external data.
4) Prepare a 10 minute presentation to the class on what you did.
5) (Encouraged, but not required) Share your visualization on social media. Keep the following in mind if you do so (modified from the Make Over Monday website):

    a) This is not about criticizing the original authors. They are people like you and me and they have feelings. Focus on the data, the charts and the story.
    b) The data set comes from the source article or the source that the article credits. Be mindful that the data is what it is and Makeover Monday is designed to help you practice data visualization.
    c) Use the hashtag #MakeoverMonday on Twitter if you create your own version and would like to share it.
    d) Include a picture of the visualization when you post to Twitter.
    e) Focus on what works, what doesn't work, why those things don't work, and how you made it better.
    f) The purpose of the makeover is to improve on the original visualization. You should try stick to the fields in the data set provided and improve upon the original visualization. However, if supplementing the data helps you tell a better story, go for it!
    g) Give credit to the original data source whenever possible

Grading: Please submit your knitted file and your presentation. Your presentation will be graded according to the project rubric. You will be asked to complete a peer group evaluation as well.

Presentation:
The schedule of presentations is below.

Presentation Date	Students
Oct 10	Katherine Prioli, Thinh Pham
Oct 24	Irina Pogrebivsky, James Fung
Oct 31	Suyin Lee, Andreas Lezis
Nov 7	Subhan Khalid, Matt Huber
Nov 14	Sam Curlee, Sam Giordano
Nov 28	Edwin Lu, Faith Platz
Dec 5	Rio Rinaldi, Felicia Pham, Nick Godshall

Your presentation should generally include the following slides:

1) A title slide
2) A brief introduction to the topic (if the original viz doesn't do that already). This could include why you chose it, if relevant.
3) The original visualization, or perhaps just a subset of it if it's a list (like the Australian salaries). Make sure to cite it appropriately!
4) What you did you process the data (this could be a few slides, depending on the complexity of the task or number of modifications). You should describe what you did in words and are welcome to show snippets of code if you think it's helpful to understand or have a particularly useful way to summarize what you did (since we are all learning to "speak" R). See example of this below, if this is the main visualization/summary for the data. If it's just an intermediary step in data processing or a minor result, it should not be included. If you did a lot of processing, you should choose just the major ones to show us. You should not include your entire markdown file or even entire chunks. If any additional packages were needed, you should state that as well.
   - I calculated the highest (min) rank by year/sex, then calculated the median across the years

```
babynames %>%
  group_by(year, sex) %>%
  mutate(rank = min_rank(desc(prop))) %>%
  group_by(name, sex) %>%
  summarise(score = median(rank)) %>%
  arrange(score)
```

name <chr>	sex <chr>	score <dbl>
Mary	F	1.0
James	M	3.0
John	M	3.0
William	M	4.0
Robert	M	6.0
Michael	M	7.0
Charles	M	9.0
Elizabeth	F	10.0
Joseph	M	10.0
Thomas	M	11.0

5) Your final visualization
6) What were the biggest technical challenges and how you overcame them (if you did)?
7) What would you suggest for further investigation of that data? Feel free to answer this question assuming you had no limitations on what data are available or what data processing skills you possess.