

Information Visualization Report

Human Freedom vs Suicide Rates

This project used two datasets obtained from kaggle.com to explore relationships between a country's human freedom – measured by their Human Freedom Index (HFI) score – and their suicide rates, population, and gross domestic product (gdp) per capita (Sutter, 2020; Yates, 2018). I joined the two datasets by country and year to determine if there is any link between freedom, number of suicides scaled by population, a country's gdp per capita, and its total population. More specifically my analysis examined the links between a country's population size and its human freedom, female freedom scores and female suicide rates, and important factors that contribute to human freedom. I posited that high female freedom countries will have lower female suicide rates than low female freedom countries and that small population countries will have more freedom than large population countries.

Population Size and Freedom

While the main questions and hypotheses detailed above were explored in the previous data preparation and modeling report, they are further described below with more refined visualizations. Focusing on population size and its relation to human freedom, Figure 1 shows that there is no relationship between a country's population and its human freedom.

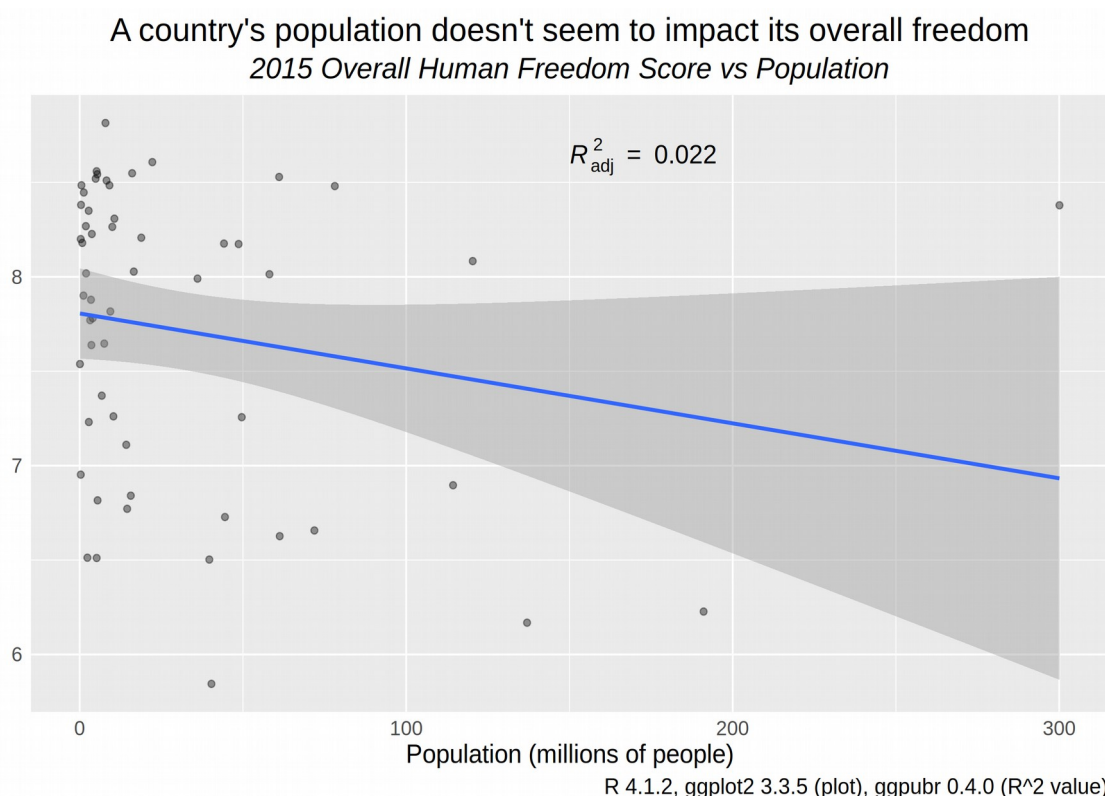


Figure 1: Scatter plot exploring relationship between Freedom and Population Size

Visually, I can see that there is no noticeable trend and this matches the adjusted R^2 value of 0.02, indicating that population explains almost no variance in the freedom score. A linear regressed smoothed line was added to highlight how few points actually fall within its 95% confidence interval (shaded gray cone).

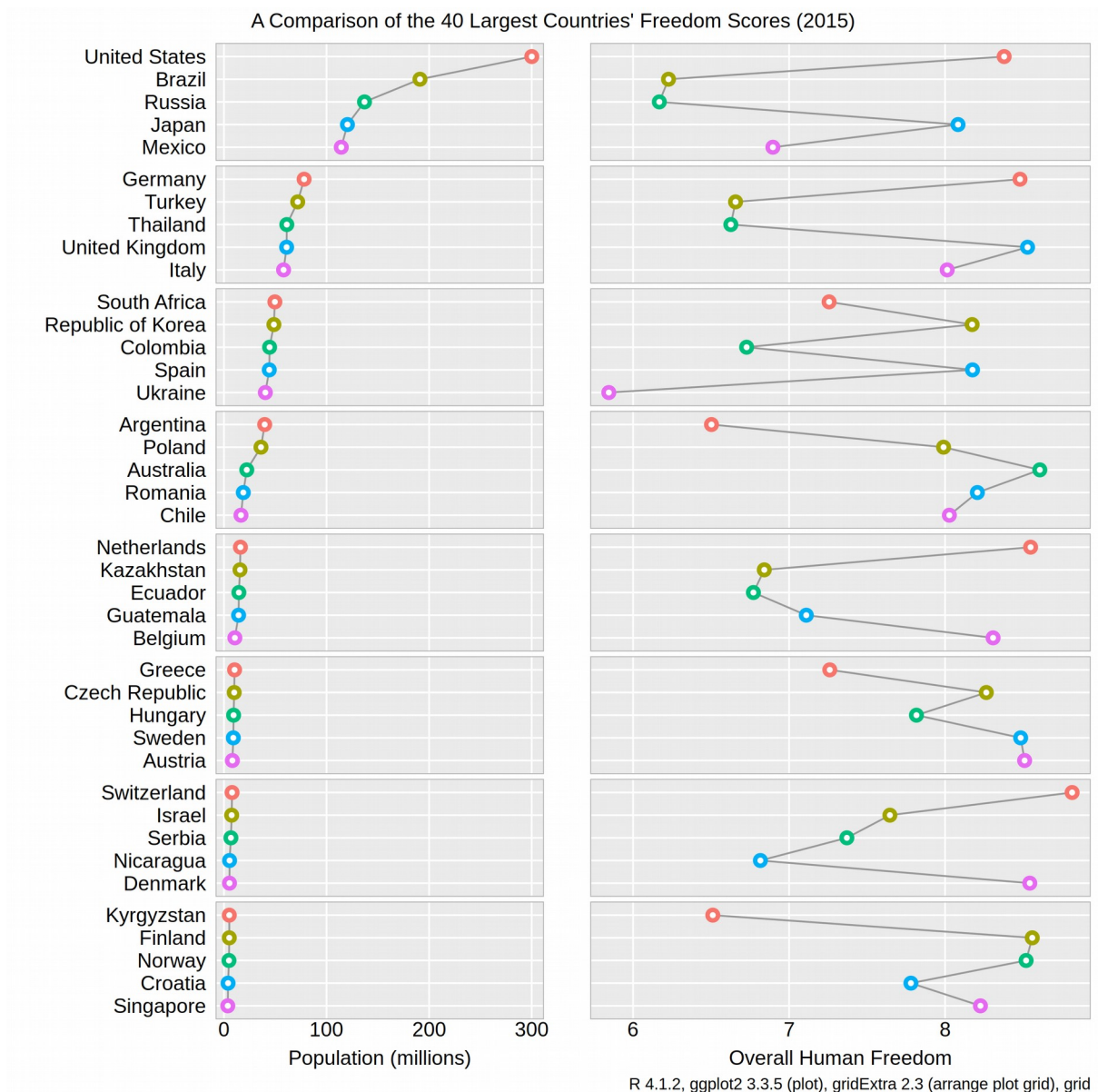


Figure 2: Row-labeled dot plot of Population vs Human Freedom

Another perspective on this relationship can be displayed with a Cleveland Dot Plot, or row-labeled dot plot using R's *gridExtra* and *grid* packages (Chang, 2021). These packages allowed me to arrange two distinct plots on the same grid and save them as one image. I learned about the row-labeled dot plot in an earlier visualization course and thought that could be a new way to visualize the trends between these variables. In Figure 2 the forty largest countries' population and freedom scores are shown alongside each other in two columns with the left column sorted by decreasing population. The freedom scores and population are linked with lines within perceptual groups of five to ease data processing and to accent any sharp changes in values from one country to the next (Carr & Pickle, 2010). While populations decline quickly at the top and level off, freedom scores zig and zag within each subgroup multiple times. There are tradeoffs between the previous scatter plot and this dot plot. I cannot find an individual countries' population and freedom score in a scatterplot (without added annotations), but more data points can be encoded in one plot. The dot plot is defined by one row per country, thus including all 56 would make shrink each country's row to the point where it would be unreadable. Overall, these two plots combined with a very low

R^2 value illustrate that it is very unlikely that the size of a country has a link to its human freedom scores.

Female Freedom and Suicide Rates

Since both the HFI and suicide rates datasets had specific information on female populations, I wanted to see if countries with higher female freedom had lower suicide rates. Figure 3 is my attempt to answer this question visually. The x-axis is an average of five women-related freedom indicators – female genital mutilation, inheritance for widows and daughters, movement, female to female relationships, and divorce. More than half of the countries had an average score of 10! While this made it more difficult to detect a relationship, it seems from a correlation and visual inspection that there is no relationship between female freedom and female suicide rates. I chose to use a series of boxplots instead of a scatterplot as my solution to an overplotting problem. Female freedom scores above 7.5 include a large range of suicide rates from a low of zero all the way to the maximum of twenty-three. Two countries with very similar freedom scores could have wildly different suicide rates.

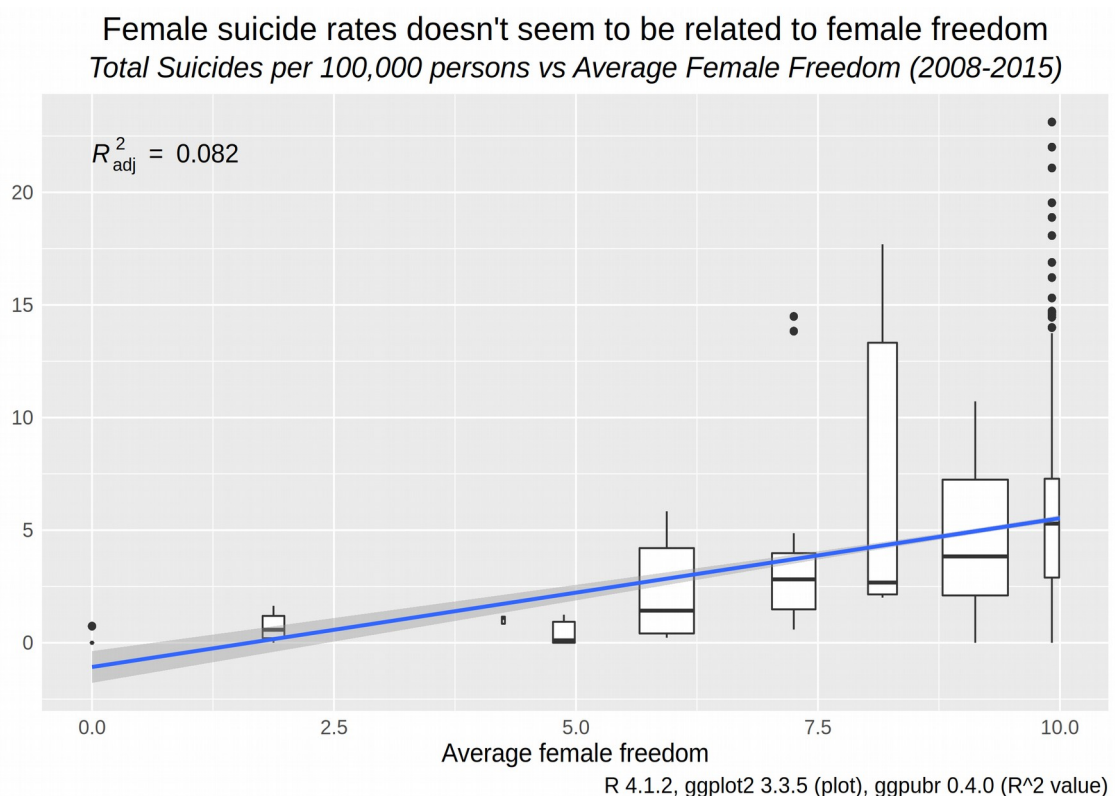


Figure 3: Boxplots showing a lack of relationship between Female Freedom and Female Suicide Rates

Overall Freedom and Suicide Rates

Continuing with the earlier group of forty largest countries, the overall human freedom and suicide rates from 2015 can be compared, after aggregating suicide rates from both genders and across all age ranges. Figure 4 may show a weak positive trend between freedom and suicide rates, which is curious. I expected countries that provided more freedom to their citizens would have lower suicide rates, but the plot weakly gestures at the opposite phenomenon. As I saw during my data exploration process, South Korea topped the suicide

charts and almost looks like an outlier when encoded in this row-labeled plot. While row-labeled dot plots cannot show as many data points as scatterplots and may not show trends as clearly, I find them more visually appealing with some pops of color to link each country to its respective dots in each column.

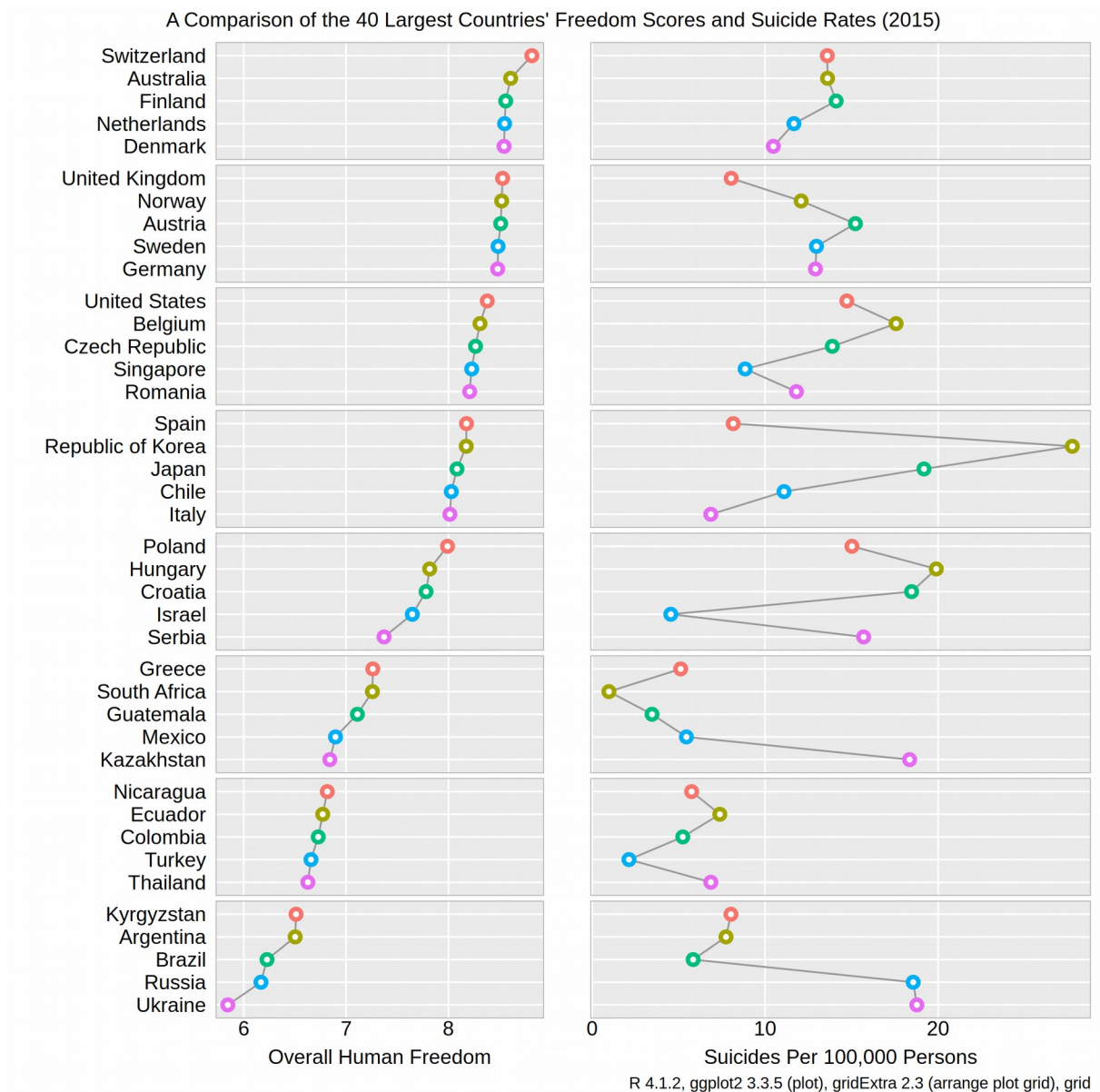


Figure 4: Row-labeled dot plot of Freedom vs Suicide Rates

GDP per Capita vs Freedom

It doesn't come as much of a surprise that a country's wealth is related to its personal and economic freedoms offered to its citizens. The United States has a vast amount of wealth and I enjoy quite a few freedoms that I couldn't in other countries. Thailand, South Africa, and Ukraine have lower freedom scores, but also have much different histories than the US and have much less wealth spread across its citizens – lower gdp per capita. Figure 5 below outlines these two variables and there certainly seems to be a positive trend. Where past plots had inklings of trends, this one definitely stands out. Of course there are still a few countries – Czech Republic, Argentina, and Romania – that buck this trend and contribute more variance to the human freedom column.

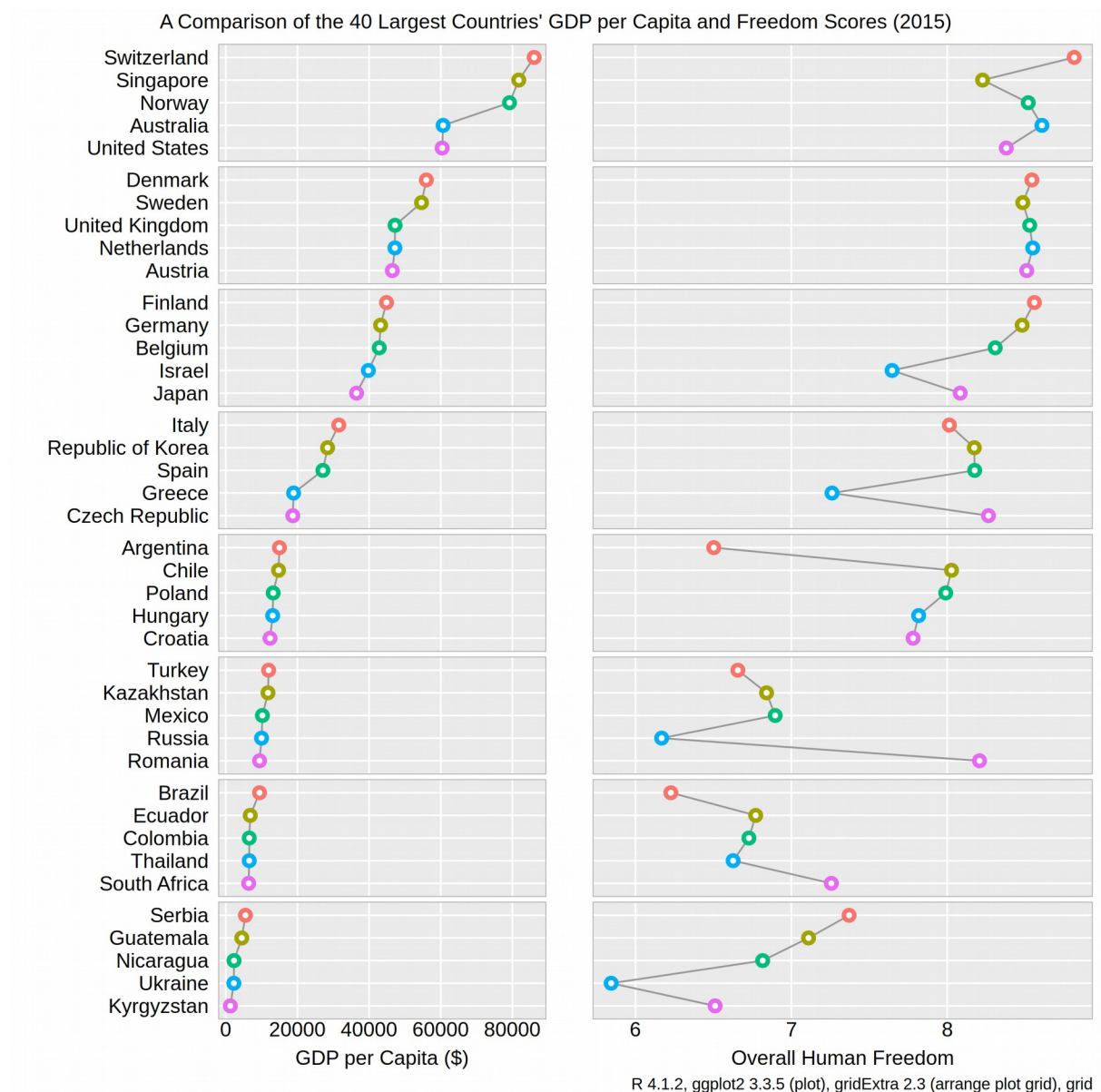


Figure 5: Row-labeled dot plot of GDP per Capita vs Freedom

Time-series Freedom Analysis

I have explored freedom and suicide rates from many angles so far, but have not checked out the time component. Are country's freedom scores trending up, down, or sideways? The final plot in Figure 6 is a line plot of the two freedom subscores – personal and economic – and their combined score, labeled “- Overall -”. I like to remove legends when possible in favor of annotations to keep the readers eyes on the plot and remove the eye ping pong that legends may induce. Three colors from R's *RcolorBrewer* package were used to label each freedom type with their overall score in purple – representing the combination of red and blue colors from the two subscores (Neuwirth, 2014).

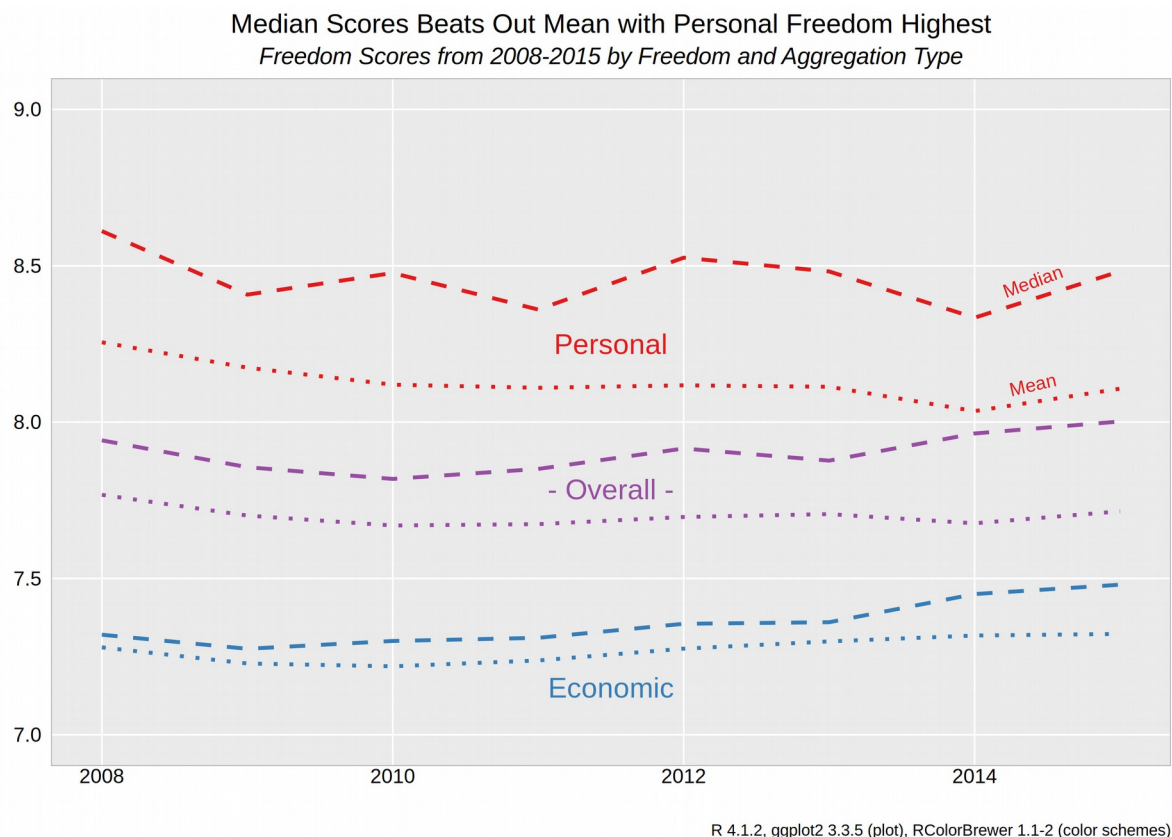


Figure 6: Line plot of 3 types of Freedom Scores Aggregated by Median and Mean

I started with just finding the median of each year's respective freedom score, but then I was curious how different the plot would look if I used the mean instead. This graphic is able to show both by using different line styles and labels to highlight that the median scores are greater than the mean in all cases, indicating a left skew to the data. Personal freedom scores are clearly the highest and economic scores are the lowest on average, leaving their overall score somewhere in between. For the most part, trends are slightly positive for economic freedom while personal freedom has fallen a bit. The separation between median and mean scores has slowly increased and may be due to a few countries with very low freedom scores dragging down the mean without impacting the median. Still the movements are quite small as this plot has been zoomed in to show scores between seven and nine while the full range of scores could include zero to ten. While I don't think it is misleading here, inaccurate axis scaling is one way to twist a graphic's message to fit an alternative narrative.

References

- Auguie, B. (2015). gridExtra: Miscellaneous Functions for "Grid" Graphics. R package version 2.0.0. <http://CRAN.R-project.org/package=gridExtra>
- Carr, D., & Pickle, L. (2010). *Visualizing data patterns with micromaps* (p. 24). London: Chapman & Hall.
- Chang, W. (2021). 3.10 Making a Cleveland Dot Plot | R Graphics Cookbook, 2nd edition. Retrieved 21 November 2021, from <https://r-graphics.org/recipe-bar-graph-dot-plot>
- Chang, W. (2021). 5.5 Dealing with Overplotting| R Graphics Cookbook, 2nd edition. Retrieved 21 November 2021, from <https://r-graphics.org/recipe-scatter-overplot>
- Clark, Z. (2014). How to add whitespace to an RMarkdown document?. Retrieved 21 November 2021, from <https://stackoverflow.com/questions/24425786/how-to-add-whitespace-to-an-rmarkdown-document>
- Kassambara, A. (2020). ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.4.0. <https://CRAN.R-project.org/package=ggpubr>
- Neuwirth, E. (2014). RColorBrewer: ColorBrewer Palettes. R package version 1.1-2. <https://CRAN.R-project.org/package=RColorBrewer>
- R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Saving grid.arrange() plot to file. (2015). Retrieved 21 November 2021, from <https://stackoverflow.com/a/28136155>
- Sievert, C. (2020). Interactive Web-Based Data Visualization with R, plotly, and shiny. Chapman and Hall/CRC Florida, 2020.
- Sutter, G. (2020). The Human Freedom Index. Retrieved 28 October 2021, from <https://www.kaggle.com/gsutters/the-human-freedom-index>
- Vásquez, I., & McMahon, F. (2020). *the Human Freedom Index 2020*. CATO and Fraser Institutes. Retrieved from <https://www.cato.org/sites/cato.org/files/2021-03/human-freedom-index-2020.pdf>
- Wickham, H. (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemond, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686.
- Xie, Y., Allaire, J. J., & Grolemond, G. (2018). *R markdown*. CRC Press.

Yates, R. (2018). Suicide Rates Overview 1985 to 2016. Retrieved 28 October 2021, from <https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016>