

Project 1 Report - Investigation of the Human Freedom Index
Scott Barnett, Meagan Colley, Jim Comas, Eden Dadi

Motivation & Summary

The Human Freedom Index is a project that attempts to quantify human freedom -- identified in the report as "the absence of coercive constraint" -- on a per-country basis by measuring a wide variety of factors. We wanted to test the relationships between some of these factors to get a better idea of what drives freedom around the world. Data for the years 2008-2016 was taken from Kaggle and the HFI's website at cato.org and imported into Jupyter Notebook in .csv format. We used a variety of python libraries to build the project including Pandas, NumPy, SciPy, Folium, and Matplotlib.

Questions and Analysis

Our dataset was divided into two categories (Personal and Economic Freedom) which were each comprised of 49 subcategories. Combing through the data for these metrics, we were able to identify four potentially interesting relationships that we wanted to examine more closely. The first relationship we wanted to analyze was that between the Personal Freedom Scores and Economic Freedom Scores. We hypothesized that the two are likely related (H_0 = no relationship). To test the hypothesis we created a scatter plot which visualized a definite positive relationship. We then ran a regression that returned a p-value of $3.92 * 10^{-155}$ which rejected the null hypothesis. To gain further insight we binned the personal freedom score data into 10 quantiles (2 of which contained no data). With that data we created a bar graph. At a glance the bar graph seems to contradict our model. Regardless of quantile, the data still indicated higher economic freedom score.

The second question looked at the relationship between money growth and personal freedom. We hypothesized that having a higher score in money growth would correlate positively with personal freedom (H_0 = there is no relationship between the two variables). To analyze this relationship we created a scatter plot with money growth on the x-axis and personal freedom score on the y-axis. At a glance the plot shows a positive correlation between the two. To further validate this assumption we performed an independent t-test. The test returned a very low p-value of $2.64 * 10^{-284}$, indicating that the null hypothesis should be dismissed and that there is a significant relationship.

Our third question studied the relationship between human freedom and internet expression freedom. We hypothesized that higher rates of personal internet expression would correlate with higher scores of human freedom (H_0 = there is no relationship). The p value was low when we initially fit the scatter plot, but increased after further manipulation. The t-test result was statistically significant (statistic=19.637, $\alpha = 6.524 * 10^{-77}$), meaning we can conclude that personal internet expression increases when human freedom increases.

The last question involved the relationship between the perception of business-government corruption (e.g., bribes and "dark money") and personal freedom. We hypothesized that opinion of government integrity would correlate positively to personal freedom (H_0 = there is no relationship). To determine this effect, we cut the data into quartiles of equal size and created scatter plots for the first and fourth quartiles. We then ran a regression for each plot and took the p-value of each model to determine its significance. We found that both models were statistically significant ($p = 0.023$ for the top quartile and $2.579 * 10^{-7}$ for the bottom, $\alpha = 0.05$) but very weak ($r^2 = 0.019$ for the top quartile and 0.094 for the bottom), suggesting that there may be no practical significance for the relationship.

Conclusion

Our work on this project unearthed four interesting correlations between components of human freedom that could each have been an individual project in their own right. Further statistical analyses such as multiple regressions or clustering the data may be used to uncover future trends. Within the scope of this project, however, our focus was primarily on attempting to qualify basic relationships within subsets of the data and using our findings to determine whether any basis exists for future research.