Women's Security Around the World

An Analysis of how Women's security changes as a factor of Freedom around the world

Introduction

It is on the basis of gender that access and power are limited and inequality is justified and maintained, and therefore gender issues can play an important role in determining access to resources and control. For this reason, we think that freedom of women might act as a good indicator of the overall freedom in a nation. While gender roles vary across cultures, historically women are always unequal in both economic and political spheres, making gender equity a continuous hot issue to examine cross-nationally. (Caprioli, *Gendered Conflict*)

In this project, we will focus on the relationship that may exist between the level of freedom that women have in any given country compared to the overall evaluation of freedom in the country. Specifically, we will look at selected variables from the Human Freedom Index database for the year 2016 and try to understand what aspects of freedom have the greatest impact on women's freedom in a country. Following this analysis, we will perform a regional analysis of human freedom around the world in combination of women's freedom to understand the possible causes for variation in human freedom and women's freedom around the world.

The study will be based around the following **two guiding questions**:
1.What areas of personal and economic freedom have an impact on women's security in a country?

2.How has Human Freedom and women's security changed over time and what are the regional differences in freedom?

Data

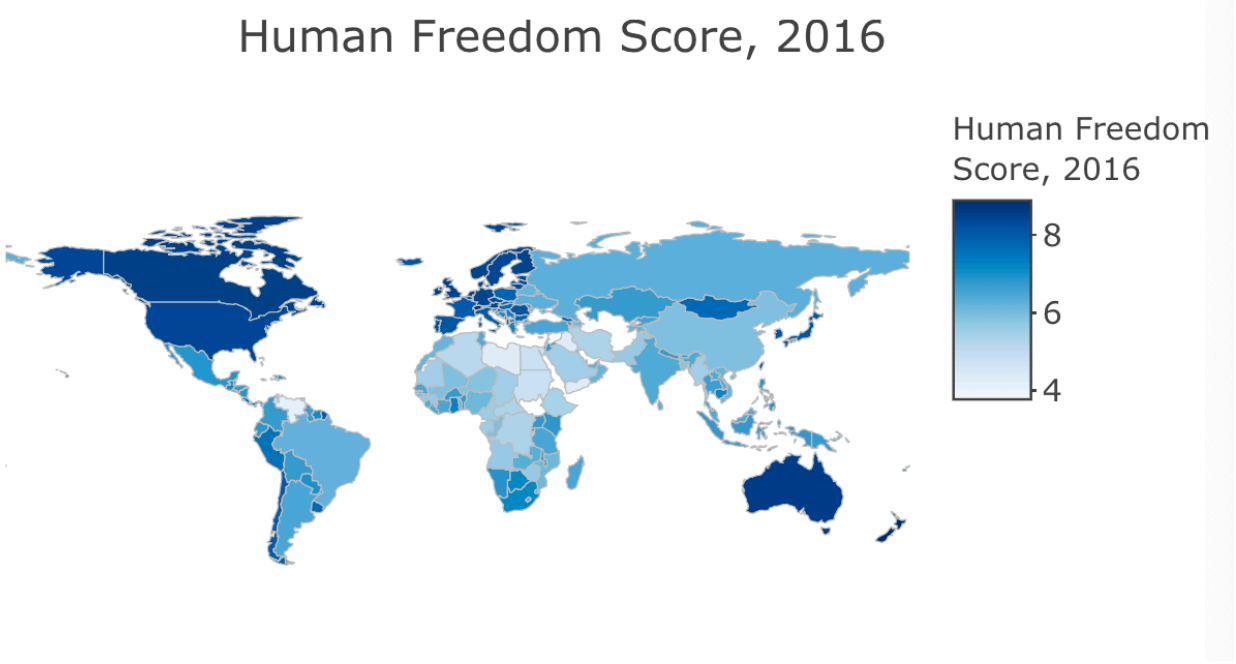
Our project focused on women's security around the world and observing how it related to other freedom variables within different regions. Our main data source came from the 2018 Human Freedom Index from the Cato Institute, however we originally found the dataset looking through Kaggle. The Cato Institute puts out this yearly report to measure the level of freedom they are able to collect data from, where this year the data comprised of 162 countries. To measure the overall freedom of a country the human freedom index uses an extensive set of 79 indicators from 2 areas of that can be broadly categorized as personal or economic freedom. However within each of these broader categories contains more specific divisions of freedom and how it pertains to areas like size of government, access to sound money, religion, rule of law, movement, identity, and safety, among others. What we did know is that each of these those 79 indicators would be weighted accordingly to produce and personal, economic, and

overall freedom score for each of these 163 countries. But what we didn't know is how these indicators could be related to each other.

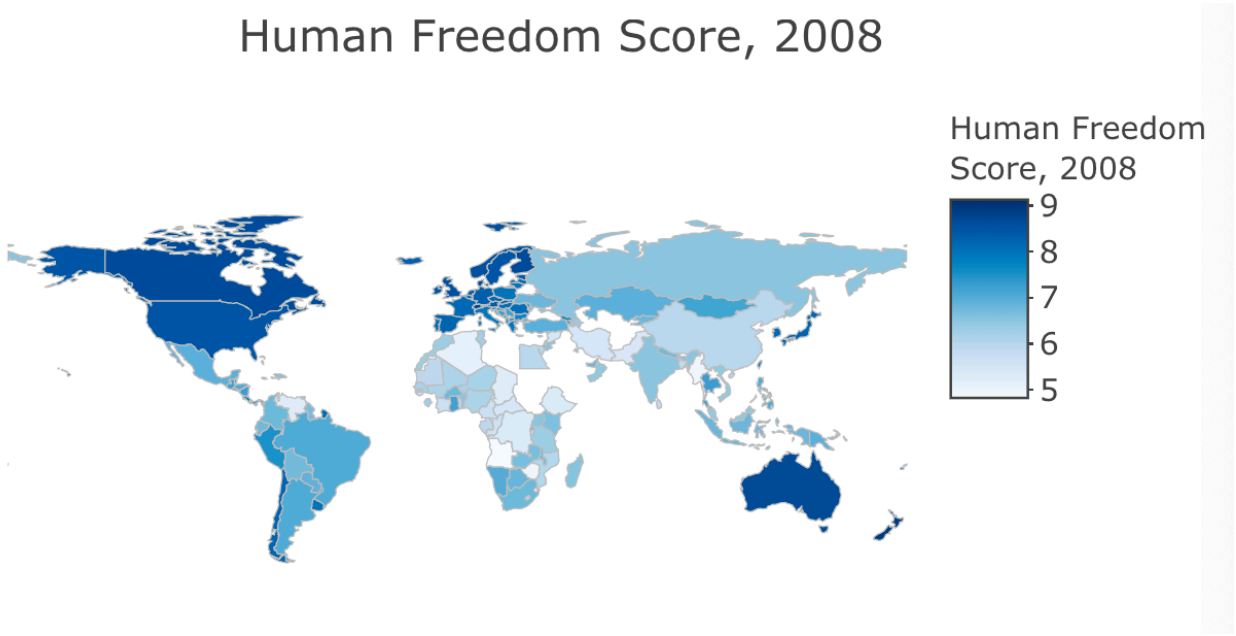
We had to start somewhere when it came to looking at how some of these indicators could be related, so we wanted to choose one variable, or a group of variables that would be ubiquitous throughout all countries and all regions within the human freedom index. We were also interested in thinking about something that could have enough variance throughout these different regions that would be useful to visualize comparisons and improve possible clustering results. What caught our eye as we were perusing through the list of variables was the 7 that encompassed the safety and security of women. These variables were mostly prefaced as being related to safety and security, but also expanded beyond that to incorporate issues like inheritance rights, freedom of movement, genital mutilation, and numbers of missing women. It is for the reasons described above in our introduction that we decided to move forward with focusing our project on what other indicators impact that safety and security of women, and address the other question of how regional differences in women's safety and security have changed over time.

Overlook of Human Freedom

We think it might be helpful to first take a look at the overall human freedom score to have a better understanding of the current situation. Since we have a continuous statistical surface attached to countries, we used plot_geo to make choropleth maps and the latest data we have is from 2016. (30-exploration, chunk choropleth map of Human freedom score 2016)

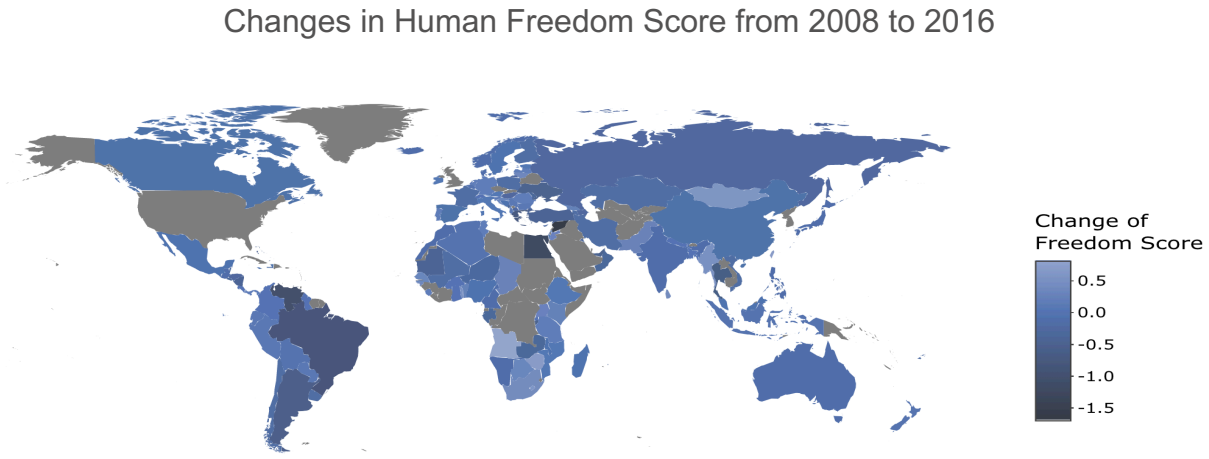


We also decided to visualize the 2008 data to see if we can find some changes here. (30-exploration, chunk choropleth map of Human freedom score 2008)



From these two graphs, we cannot really tell the difference in each country, but we do notice that year 2016 has wider range in human freedom scores. There are countries have human freedom score over 9 in 2008, which didn't happen in 2016.

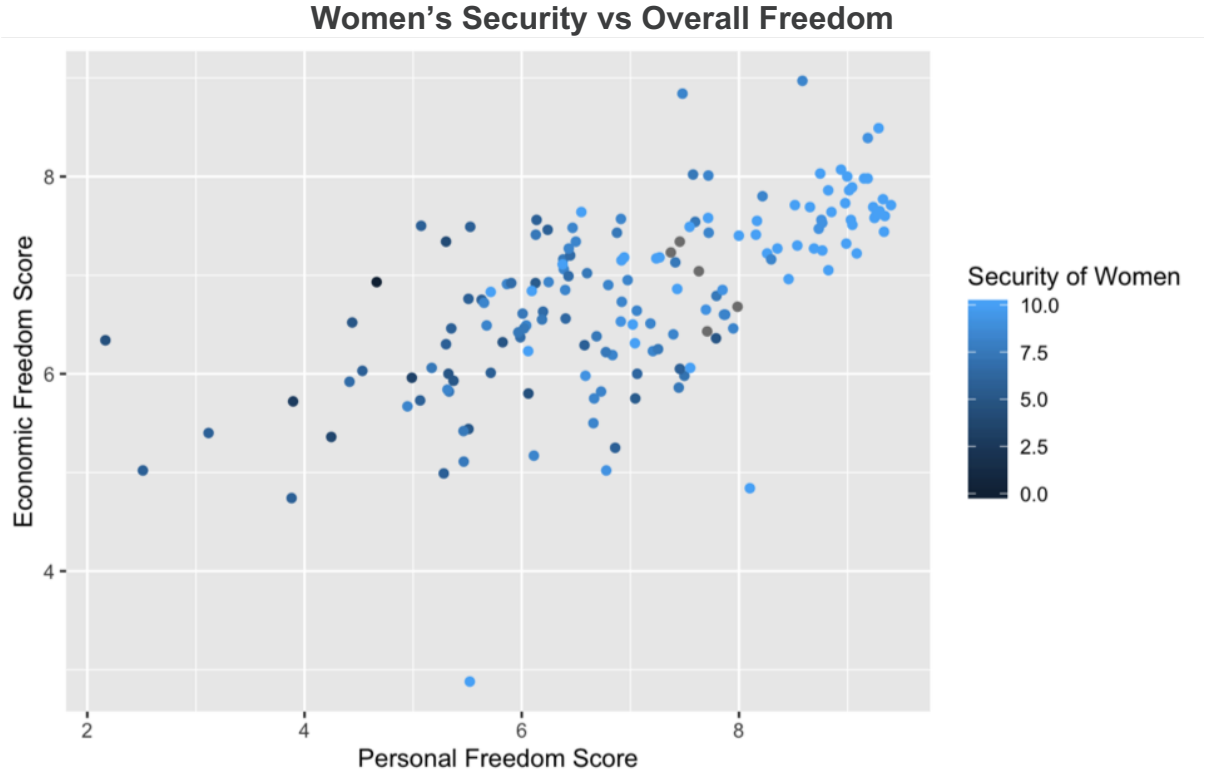
Thus, we did the subtraction between these two years and made a visualization of the changes in human freedom scores over the years. We did the subtraction between these two years. (30-exploration, chunk world map of changes in scores from 2008 to 2016)



As shown from the graph, given the lighter the color is, the higher freedom score the country has, we do see there's a decline in global freedom. Very few countries have positive scores in change. Especially, South America and Egypt have the most noticing decrease.

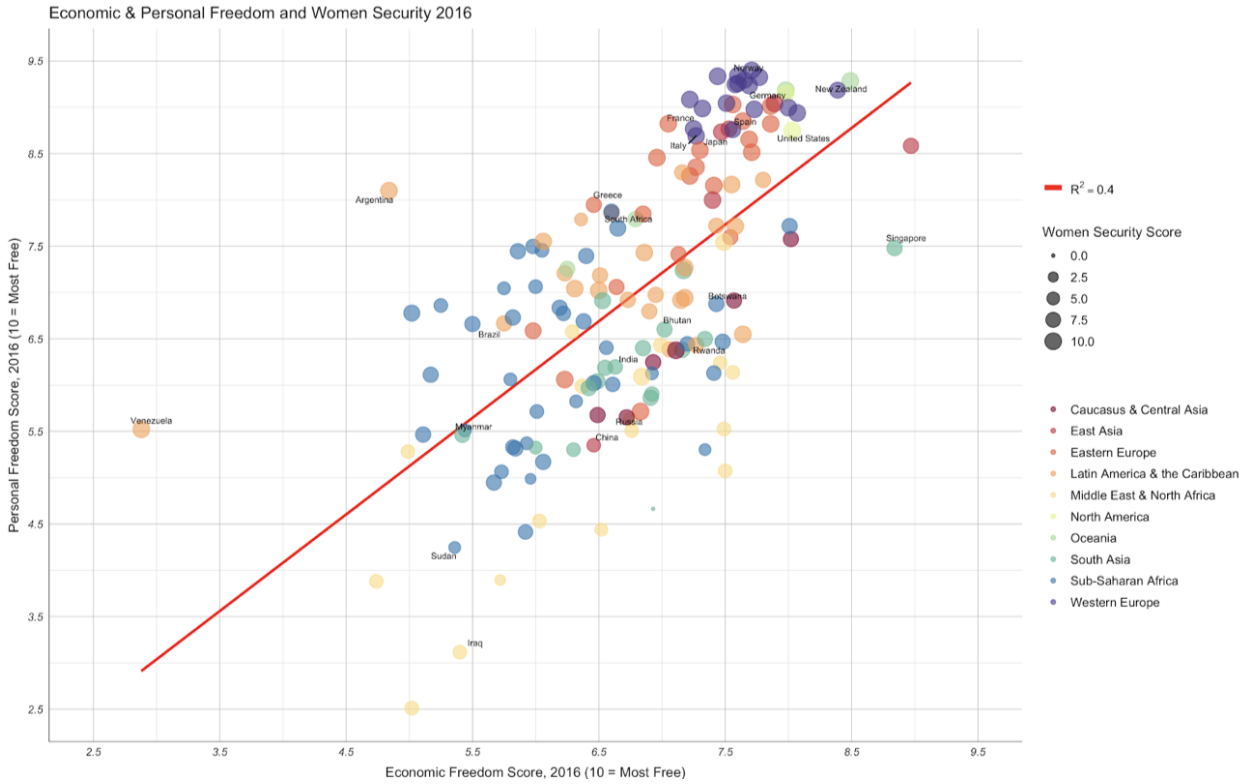
Women's Security and HFI

And when we look at the women related variables from the data set, the women security score seems to be the most aggregated variable. Therefore, we want to see how the security of women might be associated with personal freedom and economic freedom scores. (30-exploration, chunk pf_vs_ef_vs_ws)



A very clear relationship is visible between measures of overall freedom and women's security in any given country. As economic and personal freedom increases, women's security and freedom increases. However, certain countries have high economic freedom but still see low values for women's security, indicating that economic measures of freedom may not be a reliable indicator of the status of women in the nation.

We wanted to look further to see how women security scores vary by regions. Each region is represented by a distinct color. (30-exploration, chunk regional analysis)



There are some interesting groups here as we observed from this graph. Top tier regions are Western Europe, North America and Oceania. Middle East and North Africa have the lowest score. Although the R square (0.4) is relatively low, because there are some outliers, we still assume that women in certain regions share similar security score.

Time Series

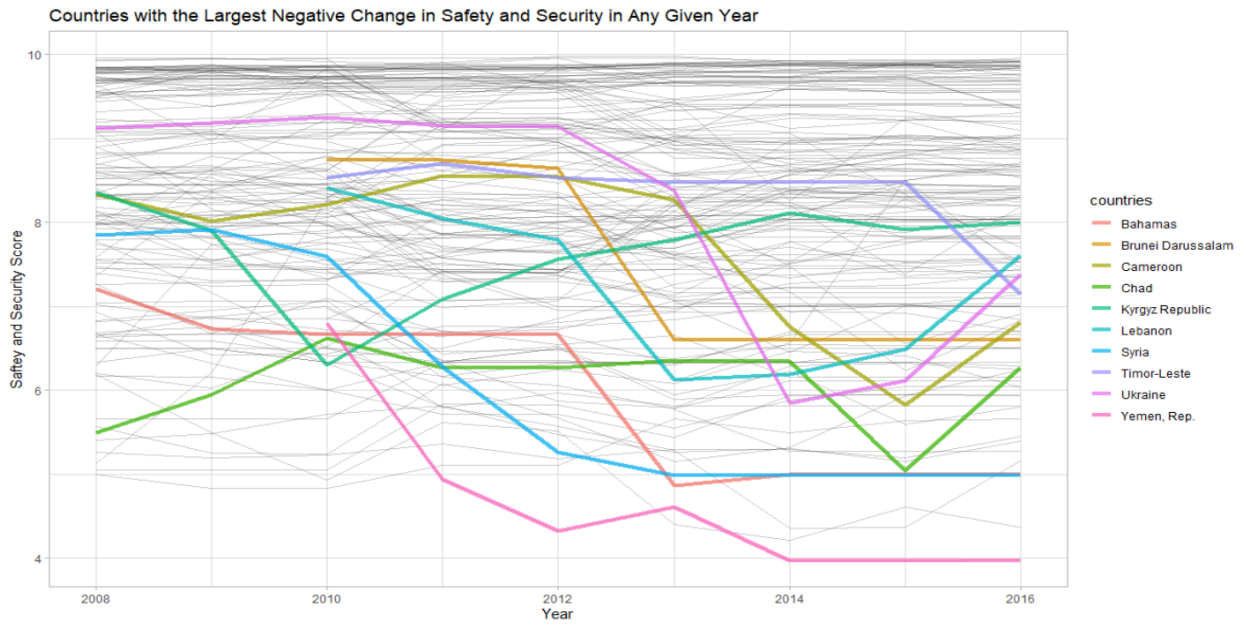
Knowing that we had data ranging back to 2008 and as late as 2016, we felt like we might be able to show some interesting things within the safety and security of the country over time. We understood that there could be more recent data available to us in terms of the safety and security of women within countries, but nothing we could find was as expansive as we wanted it to be in terms of the number of countries that we would need to collect data for. But it was also the case that The Human Freedom Index had consistency in how they were collecting this data and all their variables were nicely produced from this 1-10 scale. By taking data from other sources we were certainly going to lose some of the consistency and thus could possibly see recent seemingly volatile behavior that in reality was not as drastic as it looks or not there at all. We also figured that if The Cato Institute didn't have this data that met their comprehensive standards, we weren't really going to find it anywhere either.

The first thing that we wanted to do was just plot every country's safety and security score over all the years that The Human Freedom Index had data for all on the same plot. We saw examples of this in class in which lightly plotting the line graphs from all

the data points resulted in visually producing the trend that the data was going on. Highlighting specific lines with colors allowed us to pick out how individual data points were going against the overall trend. This was the idea behind these spaghetti plots. We thought we could pick out countries that were outliers in terms of their overall safety and security score, as well as their women's safety and security score, and show just how aberrant their scores over time were. To determine outliers, we thought it would be interesting to look at not just how much their scores changed from our earliest year with data, 2008, to our most recent, 2016, but how the country's scores changed from year-to-year. Essentially what we were trying to do is measure extreme volatility. We mutated the dataframe in the code chunk labeled "add_change" within the "20-feature-engineering.rmd" file. The graphs produced using this dataframe are in the code chunk labeled "largest_change" within the "30-exploration.rmd" file.

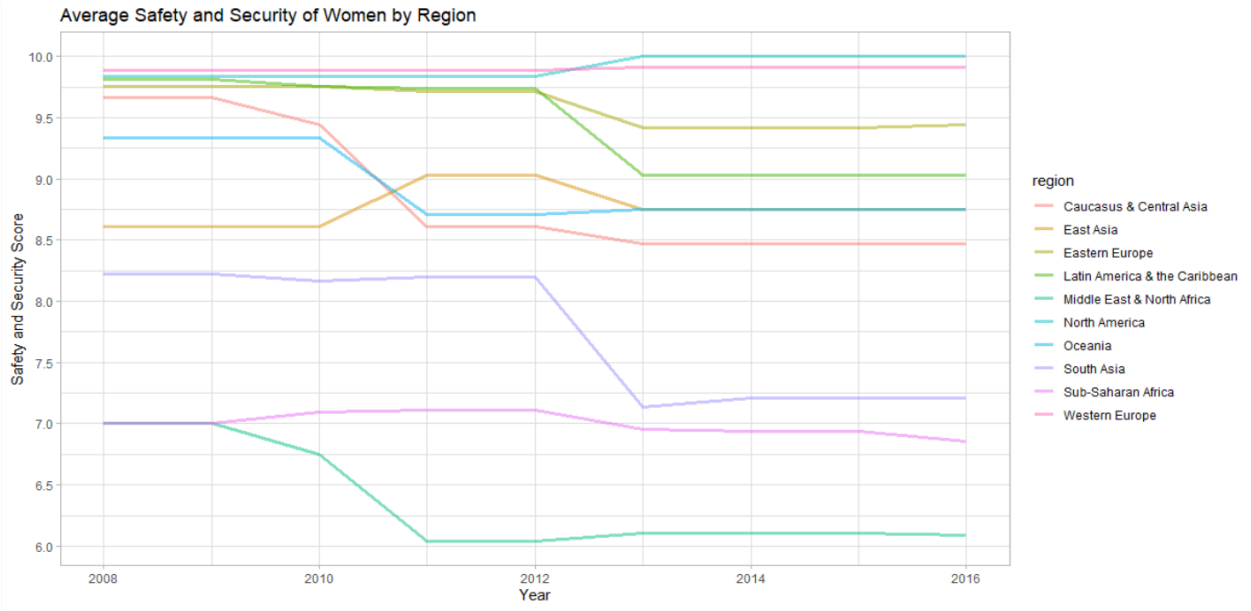
Unfortunately, the visualizations with the women's safety security variable didn't come out as we had hoped. Through our attempts at plotting this variable, we realized that the scores had too much density at specific values on the 0-10 scale. Since women's safety and security was more of a specific variable and not as much of an aggregate of others, like overall safety and security was, the scores within the variable we mostly whole numbers or nice decimals. And a lot of countries were being given similar scores. This meant that when we went to plot the point over time, we were getting a lot of overlapping lines which visually didn't show any variance in the data. In fact, it was extremely confusing to try to determine where the general trend actually was. So, because of this we had to plot the overall safety and security score that had much more variance in its values.

We wanted to see which countries had the largest positive and largest negative change in their safety and security scores in any given year. We would then take the top 10 countries with the greatest change and highlight them on the plot. What we expected to see was a large distinction between the countries that we had determined to be the most volatile in terms of their safety and security and the rest of the countries that we assumed would have exceeding less volatility in their safety and security scores. This turned out to be partially true, but only for certain regions. We didn't expect to see a high level of volatility throughout the majority of the countries, causing the overall trend to really be non-existent. In fact, even evidence of region-based trends within our most volatile countries was hard to come by. What we discovered is that there needs to be an extremely strong force within the region in order to move the safety and security scores of multiple volatile countries together. An example of this is with the countries of Yemen and Syria within the plot, "Countries with the Largest Negative Change in Safety and Security in Any Given Year"



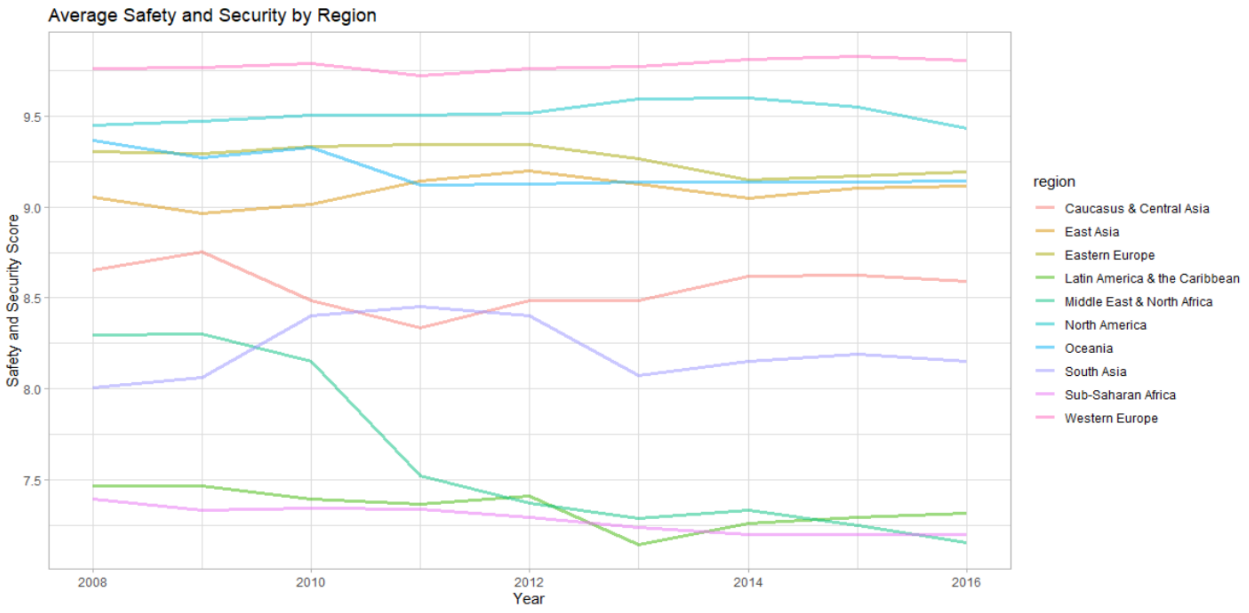
We can see Yemen and Syria at the bottom of our plot highlighted in the pink and blue respectively. Starting in 2010 we see a steep decline that levels out a bit after 2012 in both of these countries. The magnitude of the decline in these countries is nearly identical, causing us to believe that there is a region-based trends affecting these two countries. We can attribute this to the strong foothold that Muslim extremist terrorist group, ISIS, has within this country. But in regard to what we are seeing in other countries and regions around the world, further evidence that general region-based trends have an effect on the volatility of countries with lower safety and security scores is scarce, especially if we are trying to look at trends of more than one year. What we can do however is contrast the countries with high safety and security scores and consistency they show throughout the years with the volatility exhibited in the countries with lower safety and security scores. There is a drastic difference generally between the countries that have a safety and security score above 9.5 in 2008, and those that have scores below 9. Again, we generally see that the lower the safety and security score, the more susceptible that country is to volatility.

We next wanted to look at overall safety and security, and women's safety and security over time by region. From our previous graph we understand that grouping by region does not necessarily reflect the volatility we see in each country, however we still feel like it is a useful method of comparison in order to aggregate that data and eliminate all the noise we found by looking at it from a country level of granularity. And by adding these aggregation features to our data, we were able to actually visualize more of the variance within the mutated regional average women's safety and security variable, than what we saw trying to visualize the women's safety and security of each country.



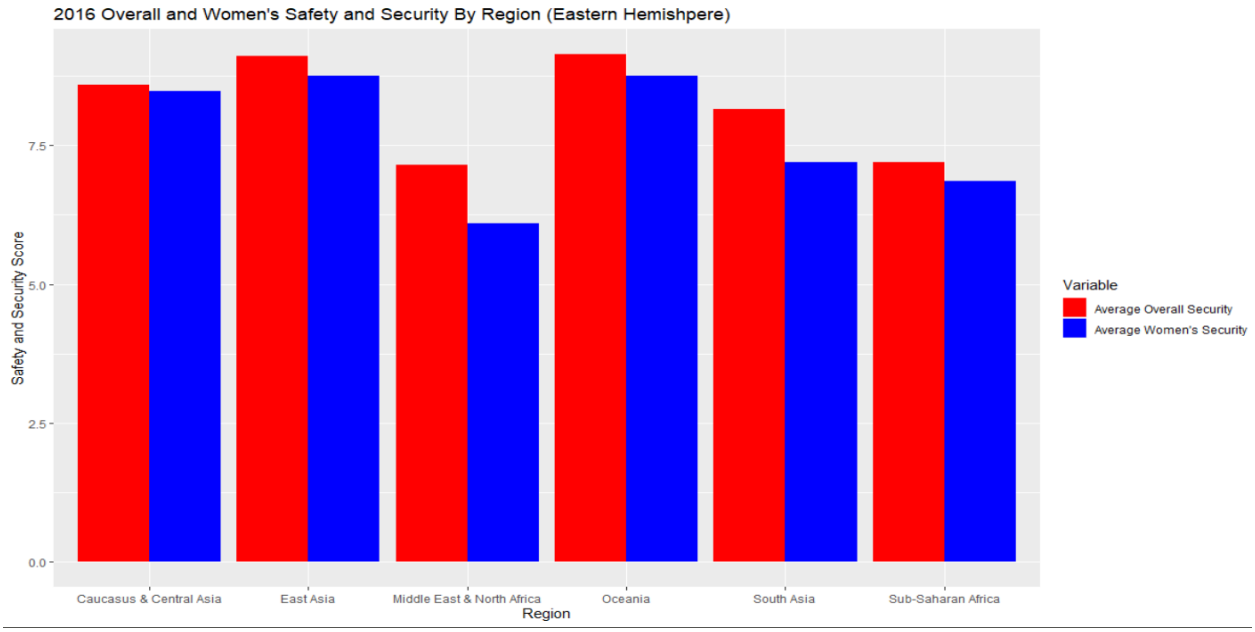
We observe in 2008 that 5 different regions had an average safety and security score above 9.5. This is was we expected to see in regions like North America and Western Europe, but not necessarily for the other regions. However, starting in 2010 we start to see a few regions decline in their average women's safety and security scores. Notably at this time we see the regions of Middle East & North Africa as well as Caucasus & Central Asia. We attributed the decline in these regions scores to the Arab Spring and the regime changes that occurred starting around this time. And moving from 2011-2016 we notice that these regions have not been able to recover from this unrest in terms of their safety and security. South Asia has the largest decline in their score moving from a score near 8.5 in 2012 and falling to just above 7 in 2013. East Asia and North America are the only regions to see their average women's safety and security scores increase from 2008 to 2016, albeit rather minute in the difference.

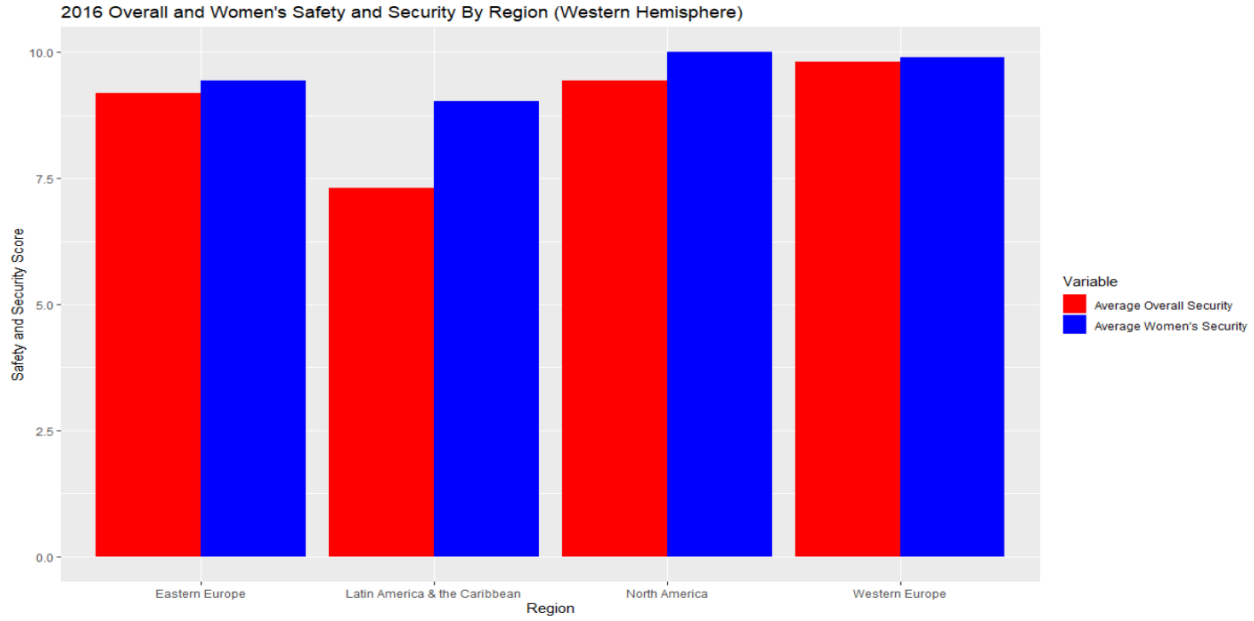
Looking at the overall regional safety and security plots we can discern a few differences from the women's regional safety and security plot. This plot as well as the one above are in the code chunk labeled "region_average" within the "30-exploration.rmd" file.



The first general thing we can point out is that the trend over time for overall safety and security is a lot more consistent in the regions, aside from what we see in the Middle East & North Africa where the decline in 2010 is actually more drastic. But the declines in regions like Central Asia and Oceania are of a less magnitude than what we saw with their average women's safety and security scores. In South Asia we actually see increasing scores of overall safety and security starting in 2009 before the relatively small drop in 2012, if we are comparing this decline to the 1.5 magnitude decrease in the women's safety and security score at this time. These differences between the trends in the region's overall safety and security, and women's safety and security scores help to emphasize the growing inequality between men and women. Oceania, Central Asia, and South Asia exhibit trends that show inequalities between the safeties and securities afforded to each gender is worse in 2016 than it was in 2008.

Now we want to look at how the two variables of average overall and average women's regional safety and security compare to each other in 2016. We will use bar plots to visualize the difference between the two scores. By further breaking up the visualization into two parts, grouping regions in the western and eastern hemispheres together, we can notice one major difference between them. For all intents and purposes, we will consider regions of Western and Eastern Europe as part of the Western Hemisphere. These two plots were created in the code chunk labeled "compare_varaibles" within the "30-exploration.rmd" file.



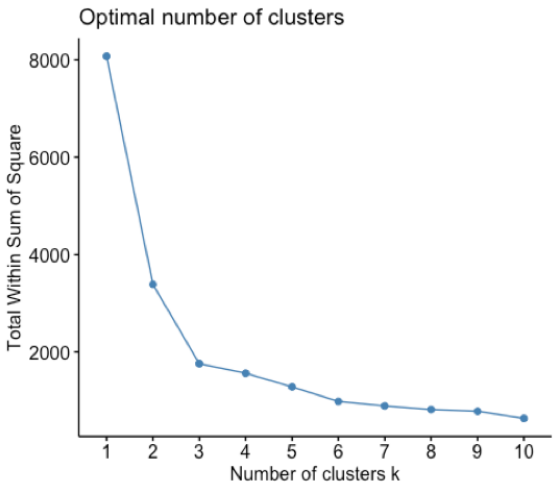


Within the regions of our Eastern Hemisphere, pertaining to Asia, Africa, and Oceania, we similarly see that the average women's safety and security score is lower than the average overall safety and security score in that region. While looking at the regions in the Western Hemisphere and Europe, we see that the average women's safety and security scores are actually higher than the average overall safety and security scores in that region. This makes some sense as for the most part we are seeing this trend in regions where there is more perceived equality among gender. However Latin America & the Caribbean is a bit of an outlier in that its overall average safety and security score is the 3rd lowest of all the regions in our data, yet the region exhibits this same trend and even to a greater extent. Latin America has the biggest difference between their average overall and average women's safety and security scores. So in this case

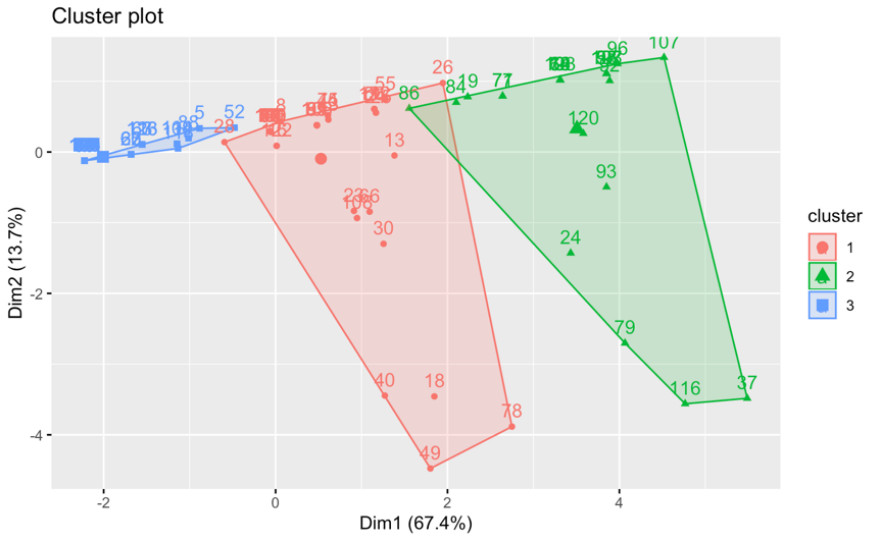
women's safety and security in the region in 2016 is not positively correlated with the overall safety and security nor other aspects of freedom. From this discovery we are led to believe that most of the violence within this region is overwhelmingly happening to men. We are unsure whether this is an overgeneralization of what is going on in the region and whether this is still the case in present day 2019. But the trends that we observed in previous plots show that this drastic difference, between women's safety and security, and overall safety and security, has consistently been the case in Latin America since 2008.

Cluster Analysis - Analyzing Women's Security

Based off the regional distribution of women's security around the world, we decided that a cluster analysis may be useful in determining what patterns may exist in women's security around the world. Performing a cluster analysis would also help us further understand the relationship between each indicator of freedom and women's security in that country. We wanted to explore the relationship between the areas of freedom as described by the Cato Institute within the Human Freedom Index and women's security. We decided to choose some aggregate and some specific variables from the HFI dataset based on our understanding of the importance that those variables could play in altering women's security. As seen by the time series analysis, the data available to us was highly variable by year and region. For this reason, we decided to limit the scope of the data for the cluster analysis to a select year, as the volatility could impact how the clusters are generated.



For our analysis, we decided to perform a cluster analysis for the year 2016, clustering on the 7 variables that were associated with women's security. Using the results from the fviz_nbclust function, we were able to determine and visualize the optimal number of clusters, as seen in the figure on the left. (30-exploratory, chunk 13; kmeans_analysis) We decided to perform our cluster analysis using 3 clusters based on these results. To ensure we were getting the best results, we also generated analyses with 2 and 4 clusters, to find that 3 clusters gave us the optimal outputs. The 3 clusters generated from our analysis are in the figure below:

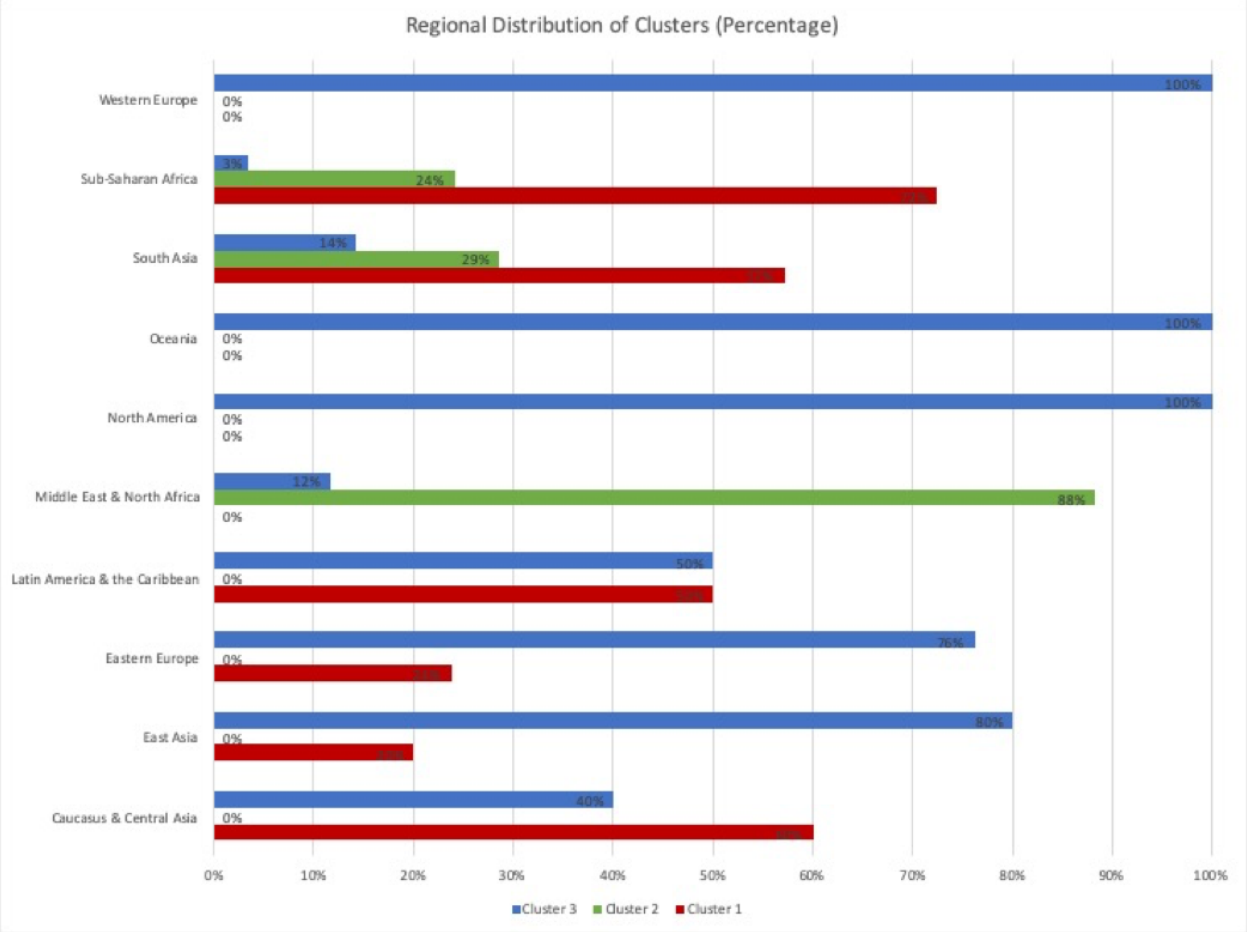


The 3 clusters formed suggest that two out of the three clusters see a lot of variability (clusters 1&2), while cluster 3 is very condensed along the two dimensions. Clusters 1 and 2 also see some overlap in behavior, while cluster 3 may be more homogenous. The output of the resulting k-means analysis is below in figure x. We can see that the between sum of squares / total sum of squares ratio tell us about 78.2% of the variation in data is captured by the clusters, and this was optimal based on trials with more clusters than 3.

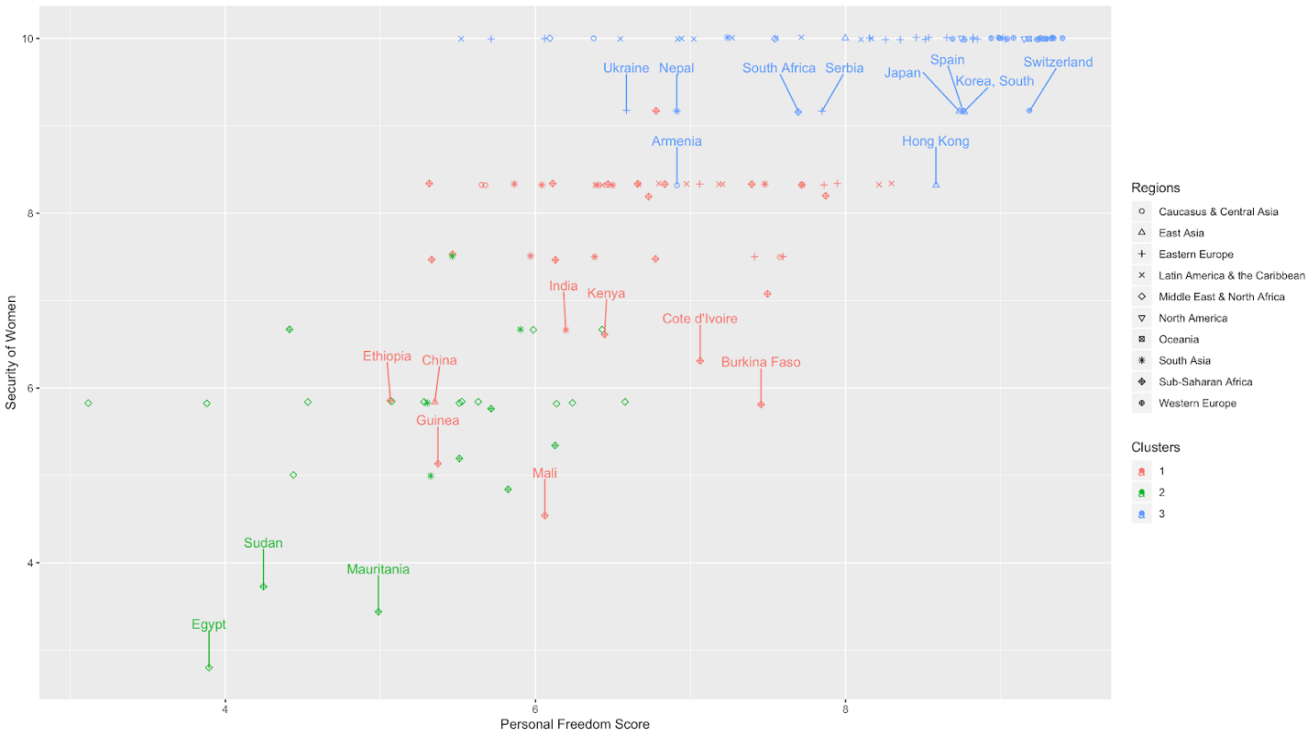
```
K-means clustering with 3 clusters of sizes 48, 26, 58

Cluster means:
  pf_ss_women_fgm pf_ss_women_missing pf_ss_women_inheritance_widows pf_ss_women_inheritance_daughters
1        9.018750            9.010417                      5.3125000                         5.1041667
2        8.684615            7.884615                      0.1923077                         0.1923077
3       10.000000            9.655172                      9.7413793                         9.9137931
  pf_ss_women_inheritance pf_ss_women pf_movement_women
1               5.2083333    7.745833          7.812500
2               0.1923077    5.587179          3.653846
3               9.8275862    9.827586          9.568966

Clustering vector:
  [1] 1 2 1 3 3 3 3 1 2 2 3 3 1 1 1 1 3 1 2 3 1 3 1 2 1 1 3 1 3 1 3 3 3 3 3 3 2 3 3 1 3 3 1 1 3 1 1 3 1 1 1 3 3 3
 [55] 1 1 2 2 3 3 3 3 3 2 3 1 3 2 1 3 2 2 3 3 1 1 2 1 2 1 1 1 3 2 1 2 1 3 3 3 1 2 2 3 2 2 3 1 1 1 3 3 2 3 3 1 2 1
[109] 3 1 3 3 3 3 1 2 3 3 1 2 1 1 2 3 1 3 2 3 3 1 3 1

Within cluster sum of squares by cluster:
[1] 902.7201 557.8034 298.1801
 (between_SS / total_SS =  78.2 %)

Available components:

[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss" "betweenss"    "size"
[8] "iter"         "ifault"
```

With some data wrangling, we were able to determine the percentage of each region captured within the clusters. (30-exploratory, chunk 21; calculate_averages) Using pivot tables in Excel (refer to files hfi_clusters.xlsx and cluster_analysis.xlsx), we were able to extract the number of countries from each region in each cluster and calculated the percentage of the region captured by the cluster. Some clear patterns can be seen in here. Cluster 3 captures 100% of Western Europe, Oceania and North America. The

cluster also captures a large percentage of Eastern Europe and Asia. These regions are primarily comprised of either developed nations or quickly developing nations. Cluster 2 only contains countries in the Middle East and North Africa, South Asia and Sub-Saharan Africa, all regions that are relatively volatile. Cluster 1 contains regions like Central Asia, Latin America, and a large percentage of Sub-Saharan Africa, which are mostly countries that are still developing. Based off this information, we decided to look at the relationships that exist between the indicator for women's security, and select variables from the human freedom index, covering both indicators of personal and economic freedom, within these clusters.



Security of Women and Personal Freedom

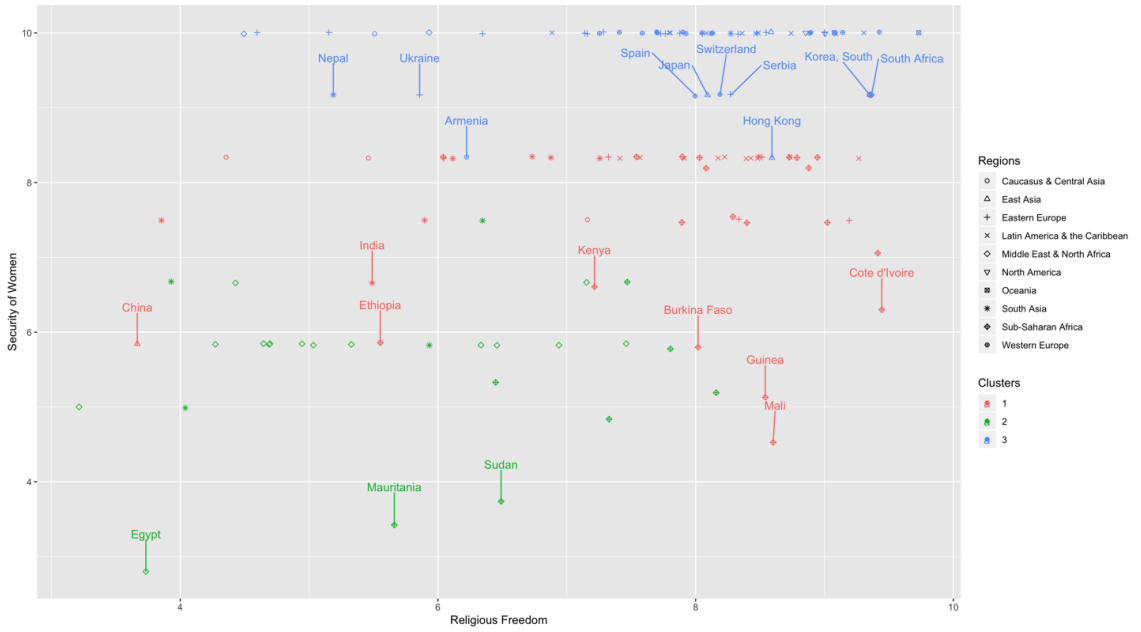


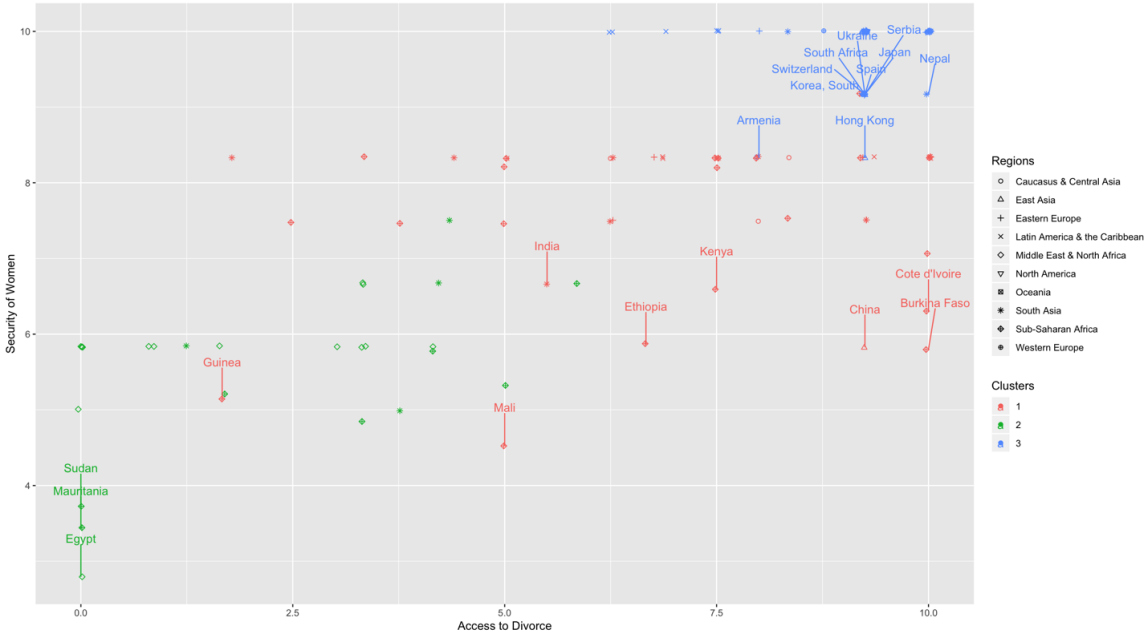
This graphic plots women's Security against the personal freedom score of each country, color coded by the cluster they were divided into (30-exploration, chunk 27; ws_vs_pfscore). The countries labelled are those that have measures of women security less than 1 standard deviation away from the average score of women's securities for that cluster. There is a fairly strong relationship between the two variables. As scores for personal freedom increase, so does the measure for women's security. A clear pattern within the clusters can also be seen, where countries in cluster 2 have the lowest scores of personal freedom and security of women, cluster 1 has moderate scores and cluster 3 has the countries with the highest personal freedom scores and measures of women's security.

Security of Women and Economic Freedom

The graphic above compares the security of women with the economic freedom score. As visible, some relationship similar to that between security and personal freedom also exists here, but to a smaller degree (30-exploration, chunk 24; ws_vs_efscore). Countries like Egypt and Sudan in cluster 2 still have moderate economic freedom scores even though they have the lowest scores for women's security. Along the same lines, there are countries like the ones in the top left corner (Venezuela) that have the highest scores for security but lowest for economic freedom. However, cluster 2 still has the lowest economic scores on average, while cluster 1 has moderate values and cluster 3 has the highest scores on average.

Security of Women and indicators of personal freedom


Security of Women
Religious Freedom
10
8
6
4
4
6
8
10
Nepal
Ukraine
Armenia
Spain
Japan
Switzerland
Serbia
Korea, South
South Africa
Hong Kong
India
China
Ethiopia
Kenya
Burkina Faso
Cote d'Ivoire
Guinea
Mali
Sudan
Mauritania
Egypt
Regions
Caucasus & Central Asia
East Asia
Eastern Europe
Latin America & the Caribbean
Middle East & North Africa
North America
Oceania
South Asia
Sub-Saharan Africa
Western Europe
Clusters
1
2
3

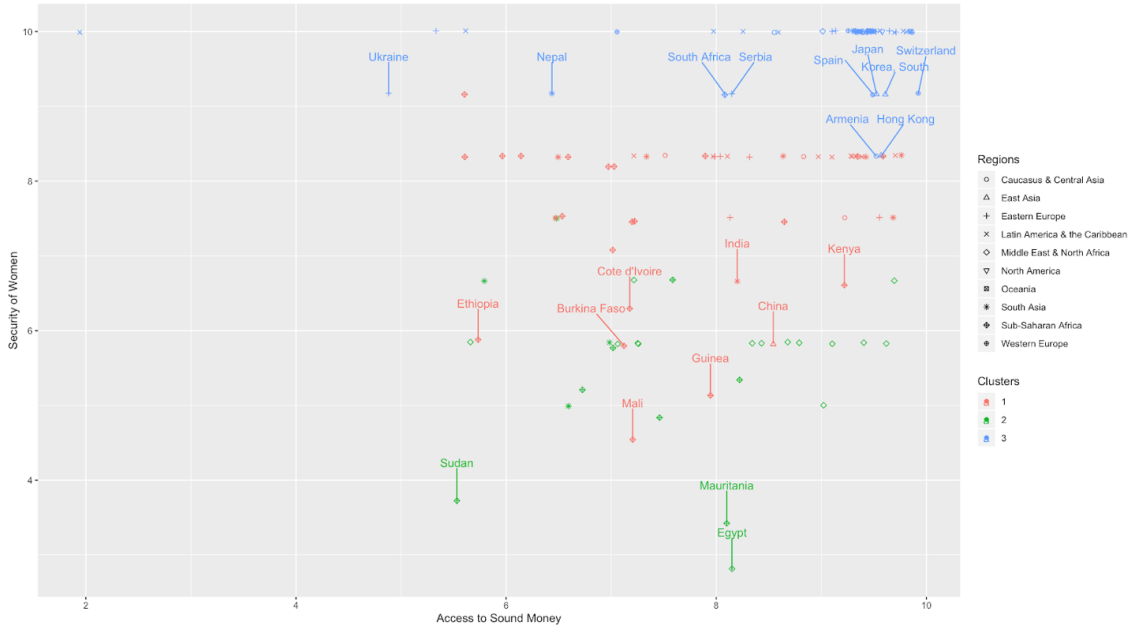

Security of Women
Access to Divorce
10
8
6
4
0.0
2.5
5.0
7.5
10.0
Serbia
Ukraine
South Africa
Japan
Nepal
Switzerland
Spain
Korea, South
Armenia
Hong Kong
India
Kenya
Cote d'Ivoire
Ethiopia
China
Burkina Faso
Guinea
Mali
Sudan
Mauritania
Egypt
Regions
Caucasus & Central Asia
East Asia
Eastern Europe
Latin America & the Caribbean
Middle East & North Africa
North America
Oceania
South Asia
Sub-Saharan Africa
Western Europe
Clusters
1
2
3

We decided to look at two specific indicators of personal freedom; religious freedom (30-exploration, chunk 22; ws_vs_religion) and the ability to access divorce (30-exploration, chunk 26; ws_vs_pfidentity). Based on the countries in each cluster, we thought that religion might play an important role in impacting women's security. Access to divorce was a variable we were very interested in as it directly related to women's security.

The graphic comparing religious freedom and the security of women demonstrates similar trends to those seen between security and scores of personal freedom. As religious freedom increases, the security of women increases. However, we can see some exceptions like China which has one of the lowest scores for religious freedom but has a relatively moderate score for security of women.

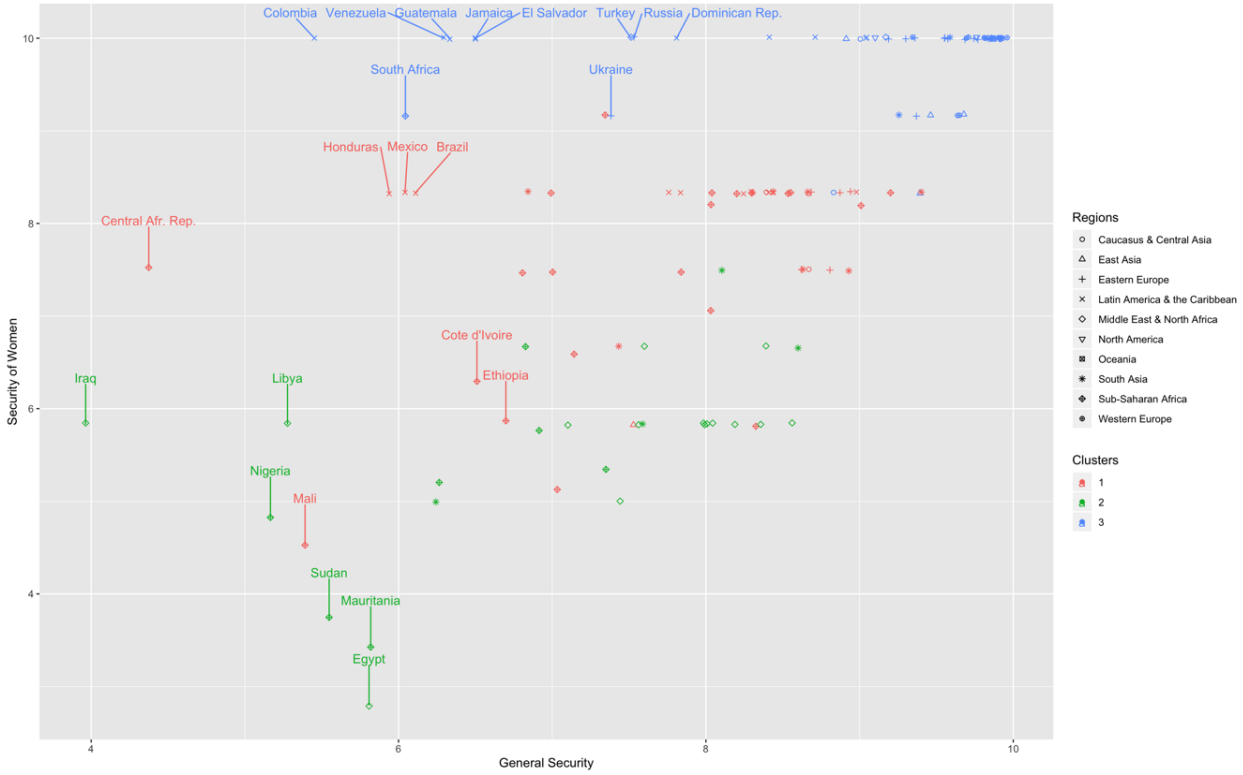
The figure comparing the security of women and their ability to access divorce showed some of the strongest relationships we saw among the variables we explored. As access to divorce increases, so does the security of women, strongly, and this makes sense. A lack of access to divorce in a nation often results in precarious situations for women. Countries that are very conservative such as Egypt and Sudan have extremely low scores of women's security, while more liberal nations like Switzerland have the highest scores of women's security as they have much easier access to divorce.

Security of Women and indicators of economic freedom



Access to sound money was one of the few economic variables that were more related to domestic activities, while most indicators were related to business and trade (30-exploration, chunk 25; ws_vs_efmoney). As access to sound money increases, the security of women increases, but to a lesser degree when compared with variables associated with personal freedom. The relationship mimics that seen when comparing security to the economic freedom score.

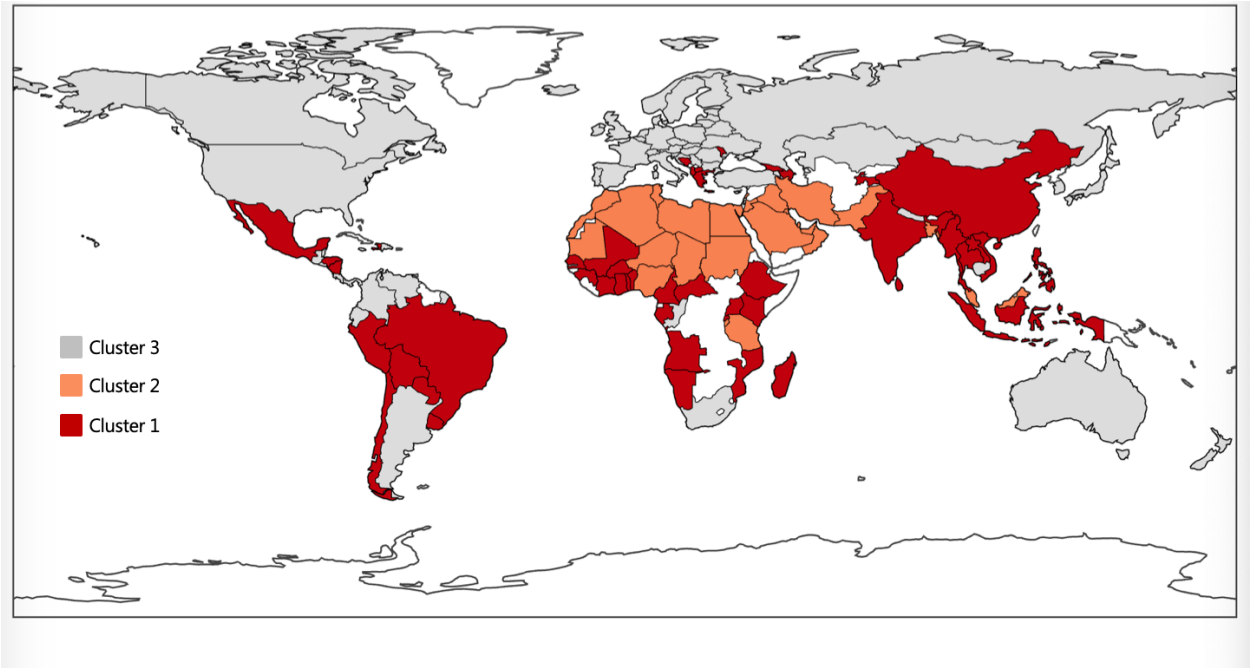
Security of Women and General Security



Things get interesting when we compare the security of women to general security. While overall, we see the intuitive relationship, where in general security increases, scores for women's freedom increases, there is a lot of variation and anomalous data in this plot (30-exploration, chunk 28; ws_vs_general_security). The countries marked in this plot are those which have a score of general security less than one standard deviation from the average of the general security for that cluster. We're unsure why South American nations, such as Venezuela, Colombia and Guatemala have been clustered with all the highly secure nations, that have both high general security and women's security. However, these nations have some of the lowest scores for general security and at the same time have the highest scores for women's freedom. This data is from 2016, so we think that the data we have for women's security might not be updated to reflect the recent volatile events that have been taking place in those nations.

We wanted to see how the clusters reflect on the world map. Firstly, we store the clusters into a matrix. Then we used plot_geo to have an integrated mapping solution. By using the ISO_code as locations from the dataset, we are able to make it an interactive choropleth as well. (30-exploration, chunk clusters on world map)

Global Distribution of Clusters



Here, we globally visualized the clusters to see any geographic variation (30-exploration, chunk 32; clusters on world map). Cluster 1 captures the rapidly industrializing and developing world, including countries like India, China and Brazil. Cluster 2 seems to be localized to north Africa and the middle east and in general more volatile nations. Cluster 3 covers most of the western hemisphere and the more developed nations in Europe and Oceania.

Conclusion

We focused on two main questions to guide our projects on women's security around the world. First, what areas and indicators of personal and economic freedom impacted women's safety and security in a country? Next, how do regional differences in women's safety and security change over time? Based on the cluster analysis, more developed nations with few to no domestic or international disputes saw better scores for women's security; stable nations such as the US, Canada and Australia were clustered together and had the highest values of Women's security. Developing nations such as India,

China and Brazil were clustered together, and had moderate scores for women's security. Troubled and volatile regions in the Middle East and North Africa were clustered together and also saw the lowest scores for women's security.
Lower safety and security scores are associated with more volatility.
Western Hemisphere and Europe saw higher safety and security scores for women than for the country overall, a pattern that needs more exploration to understand the underlying factors controlling this. Part of the reason we see this may be because our data is outdated and does not account for more recent events such as those seen in some South American nations. An analysis with updated data may help us understand this issue better and adding information from additional datasets regarding the kind of issues women face in different countries may help better define the variables for women's security within those nations.

Over time we saw a contrast between the consistency that countries with high safety and security scores showed, and the volatility that countries with lower safety and security showed. Generally, if a country had a score below 9.5 in 2008, they were much more susceptible to volatility over the years than those that had a score higher than 9.5 in 2008. By taking the regional average of the overall and the women's safety and security over time, we noticed that the Middle East & North Africa had the largest decline in both variables, however regions like South Asia, Central Asia, and Oceania saw greater declines in their women's safety and security compared to the overall safety and security. This led us to an overall trend we saw between regions in the Eastern Hemisphere and the regions in the Western Hemisphere, including Europe. These regions in the Eastern Hemisphere had women's safety and security scores lower than their overall scores in 2016, while regions in the Western Hemisphere saw women's safety and security scores higher than their overall safety and security score.

Work Cited

Caprioli, Mary. "Gendered Conflict." *Journal of Peace Research*, vol. 37, no. 1, 2000, pp. 51–68., doi:10.1177/0022343300037001003.

Vasquez, Ian, and Tanja Porcnik. "Human Freedom Index 2018." *Cato Institute*, www.cato.org/human-freedom-index-new.