Investigating Personnel Contributions to United Nations Peacekeeping

By

Corinne Batho-Newton

Jodie Harbour

Djenebou Kaba

Project Supervisor

Eleni Matechou

A Dissertation

School of Mathematics, Statistics and Actuarial Science

MA5957: Year in Data Analytics Project

May, 2021

Table of Contents

1 Introduction

The United Nations (UN) police and military personnel originate from over 100 countries, and serve under the blue flag to carry out orders given by the Security Council. These military operations aim to protect populations against threats and overall make them safer environments. Within this project, data has been explored to identify countries that participate more in peacekeeping operations. Analysis was carried out into why some countries contribute more than others, and the characteristics of these countries were investigated. Furthermore, gender differences within peacekeeping operations are discussed, as well as fatalities. The topic is examined by carrying out different data analytics techniques such as exploratory data analysis, cluster analysis and regression models.

1.1 Motivation and Background

The UN is an international organisation founded in 1945 after World War II and consisting of 193 countries. The UN maintains international peace and security. They promote development, give humanitarian assistance to those in need, uphold international law, protect human rights, and promote democracy. Peacekeeping by the UN is a role run by the Department of Peace Operations. The organisation helps countries torn by conflict to create the conditions for lasting peace, and this topic has been chosen due to personal interest in the organisation and their accomplishments. It has also been chosen due to curiosity of discovering patterns in their data that may highlight characteristics of peacekeeping and trends within the organisation. For example, how the wealth of a country correlates with involvement in peacekeeping. This research is looking to explore how peacekeeping has developed over time and how wealth has contributed to peacekeeping.

Academic research was carried out to explore the background of the chosen topic, to see what themes and conclusions have previously been drawn and how this could aid us with our analysis. Research has shown that some countries are involved in peacekeeping operations for self-interest reasons and political and economic gain (Passmore et al. 2018). This could be helpful for later discussions when analysing trends of countries who contribute to peacekeeping and how the wealth of a country alters its contributions to peacekeeping. Investigations have shown that peacekeeping fatalities are often caused by accident and illness and there is also research into how fatalities affect future peacekeeping operations (Rogers, J., and Kennedy, C., 2014). This article gives a greater insight into the number of fatalities involved with peacekeeping and the effects this causes. Another topic explored was gender equality within personnel contribution in peacekeeping operations. The article by Crawford et al. (2014), discusses why there is an underrepresentation of women within the UN peacekeeping operations and how a country's level of gender equality impacts the amount of female personnel deployed. This raises possible discussions of how gender equality could be related to a country's wealth and the effects this might have. **348**

1.2 Existing Literature

The purpose of this literature review is to give a more critical overview and explore other analysis and research carried out on the chosen topic of exploration into personnel contributions to United Nations Peacekeeping.

Research has shown that a large population base allows for more recruits for peacekeeping operations, suggesting that population plays a part in the amount of personnel contribution a country gives (Gaibulloev et al. 2015). It is also discussed that UN missions deployments increase with the number of coinciding peacekeeping missions, which results in some countries supplying UN peacekeepers as a money-making venture. An empirical study by Passmore et al. (2018) supports this idea of countries participating in UN missions for finance and self-interest reasons. Also, self-interest reasons in UN peacekeeping operations such as financial facility have led to organisational failings. This leads to gaps in traditional peacekeeping operations and quality. As a result, the UN relies on private military and security companies (PMSC) to provide a range of services in peacekeeping operations. The use of these PMSCs is an issue as it potentially undermines the UN legitimacy and creates a blurred state when assessing UN peacekeeping operations (Tkach, B., and Phillips, J., 2019). Private military and security companies could be replaced with member contributions, but there is insufficient evidence for this. Research shows that the UN relies heavily on its funding (approximately 34%) from countries like the USA and China, however it is important to note that this does not equal higher personnel contribution. Also, more financial contributions from these countries equals more power within the organisation. This has potentially become an issue as it has led to the UN falling into a position of deciding whether to listen to its highest donors or to advocate for humanitarianism such as world peace (Dos Santos, G., 2021).

Throughout the literature, there is consistent evidence that personnel contribution to UN peacekeeping is due to several different reasons like self-interest. However, the extent of these studies is limited and they do not delve deeper as required. Therefore, it would be premature to make a solid conclusion with these alone. Perhaps detailed case studies could help further understand. Also, although the paper from Dos Santos was insightful, it does not comprehensively argue this matter.

Overall, it is clear that personnel contributions in peacekeeping are due to several reasons and factors such as larger population and private companies. However, personnel contributions towards peacekeeping are crucial for successful operations, and despite the negatives there are a number of studies that have proven UN peacekeeping to be effective in conflict resolution (Di Salvatore, J., and Ruggeri, A., 2017). Also, these raised issues need to be looked at more comprehensively to fully understand and potentially resolve them in the future. This literature will support the project by enabling a solid and more informed conclusion to be reached. Different analysis techniques such as exploratory data analysis, cluster analysis and statistical analysis will be used to answer these questions with the datasets to see where relations and conclusions can be made. **505**

2 Methodology

2.1 Research Question & Objectives

This research investigates whether personnel contributions to United Nations peacekeeping are influenced by socio-economic or socio-demographic characteristics of the contributing country. For this, the following questions will be addressed:

1. Can socio-economic and socio-demographic variables be used to cluster countries that contribute to UN peacekeeping missions so that patterns can be identified?
    a. Investigate clusters in relation to the total number of contributions per country
    b. Investigate clusters in relation to the percentage of female contributions per country
2. Can socio-economic and socio-demographic variables be used as predictors for the number of personnel that a country will contribute to UN peacekeeping missions?

These questions are interesting as they explore numerous variables, and socio-economic data is easily accessible from a reputable source, the World Bank. A database on World Development Indicators from the World Bank enabled selection of reliable and relevant variables, for the range of countries and dates needed (The World Bank, n.d.). The variables investigated were refugee population (total number of refugees), urban population, armed forces (total armed forces personnel), female labour force (percentage of labor force that is female), female life expectancy (at birth), GDP (Gross Domestic Product, in US dollars), military expenditure (in US dollars), and compulsory education (number of compulsory education years).

The chosen variables give an insight into a range of areas, such as gender equality, economics, and a country's military power. They are factors that are noteworthy to investigate to see how they correlate and whether any have statistically significant relationships to peacekeeping contributions. As discussed earlier, past research into UN peacekeeping contributions implies that financial benefit is one of the main reasons that countries contribute troops to missions. Governments are incentivised to provide personnel for UN missions because they are reimbursed the deployment costs for each person. (Boutton, A., and D'Orazio, V., 2020). However, recently, researchers are concluding that there is more to the story. Coleman and Nyblade (2018) discuss the strong 'peacekeeping for profit' narrative, and raise arguments to suggest that there may be other factors involved, rather than simple financial gain. For these reasons, it is interesting to explore potential factors that could be influencing contributions to UN peacekeeping.

2.2 Data Selection

The primary data explored were the 'Troop and Police Contributions' datasets from the United Nations Peacekeeping open data portal (UN Peacekeeping, n.d., *Troop and Police Contributions*). One dataset, titled 'Rank', consists of a list of 122 countries ranked by the number of total contributions recorded for January in 2021. It was broken down into Female and Male contributions, but other than this it did not provide much detail on the personnel contributed. Table 1 gives more information on the 8 variables included in the Rank dataset. The second dataset, titled 'Gender', gave a more comprehensive overview of contributions to missions from 146 countries since July 2002. It specified the number of personnel contributed by each country per month, and provided information on the type of personnel contributed, their gender, and the mission they were contributed to. This dataset allowed for a much deeper investigation as to how gender and country affect rates of contribution and the type of staff contributed. Table 2 provides further detail on the 9 variables included in the Gender dataset. 533

Table 1: Descriptions on the variables contained within the Rank dataset

Variables (Rank Dataset)	**Description**
ISO Code 3	ISO Code 3 is a 3-letter standardised country code produced by the International Organisation of Standardisation (ISO) and it is used to represent countries or territories around the world (GOV.UK, n.d.).
M49 Code	An M49 Code is a 3-digit number developed and used by the UN to represent different geographical, political, or economic regions (United Nations Statistics Division, n.d.). For this dataset, these were specific to the countries named.
Country Name	Name of the contributing country (a total of 122 different countries included within the dataset).
Rank	A variable for contribution rank of a country, with rank 1 representing the country who had contributed the highest total number of personnel. Within this, some numbers were duplicated to represent ties where several countries had contributed the same number of personnel.
Female Personnel	Total female personnel contributed per country in December 2020 (ranged from 0 to 708).
Male Personnel	Total male personnel contributed per country in December 2020 (ranged from 0 to 6506).
Total Personnel	Total personnel contributed per country in December 2020 (ranged from 1 to 6798).
Date Reported	The monthly reported date, which in this case was 31st December 2020.

Table 2: Descriptions on the variables contained within the Gender dataset

Variables (Gender Dataset)	**Description**
Contribution ID	Unique ID number given to each contribution (different for each country-gender-type-mission-date combination). A total of 152,524 contributions were recorded.
ISO Code 3	ISO Code 3 is a 3-letter standardised country code produced by the International Organisation of Standardisation (ISO) and it is used to represent countries or territories around the world (GOV.UK, n.d.).
M49 Code	An M49 Code is a 3-digit number developed and used by the UN to represent different geographical, political, or economic regions (United Nations Statistics Division, n.d.). For this dataset, these were specific to the countries named.
Country Name	Name of the contributing country (a total of 146 different countries included within the dataset).
Female Personnel	Total female personnel contributed per country per mission (ranges from 0 to 720).
Male Personnel	Total male personnel contributed per country per mission (ranges from 0 to 4392).
Personnel Type	Categories of personnel contributed. This could be one of five types: Staff Officer, Experts on Mission, Troops, Individual Police or Formed Police Units.
Mission Acronym	The acronym for the mission that each personnel contribution was made to (48 missions identified in the dataset, some no longer active).
Date Reported	The date that the contribution was recorded. This ranged from July 2002 until December 2020, however there were some missing months within this range.

To investigate country characteristics in relation to the UN peacekeeping contributions, a new dataset was created that matched World Bank data with the relevant countries for each year between 2010 and 2020 (The World Bank, n.d.). The contributions data contained in this new dataset was the number of male and female personnel contributed per year per country, as well as the total number of personnel contributed. Additionally a percentage of the female contributions column was calculated so the cluster analysis could be conducted. This secondary dataset contained 1,395 cells with missing data. Figure 1 shows where the values are missing for all years, and figures 2 and 3 (page 8) give a breakdown for the missing values in 2010 and 2019.

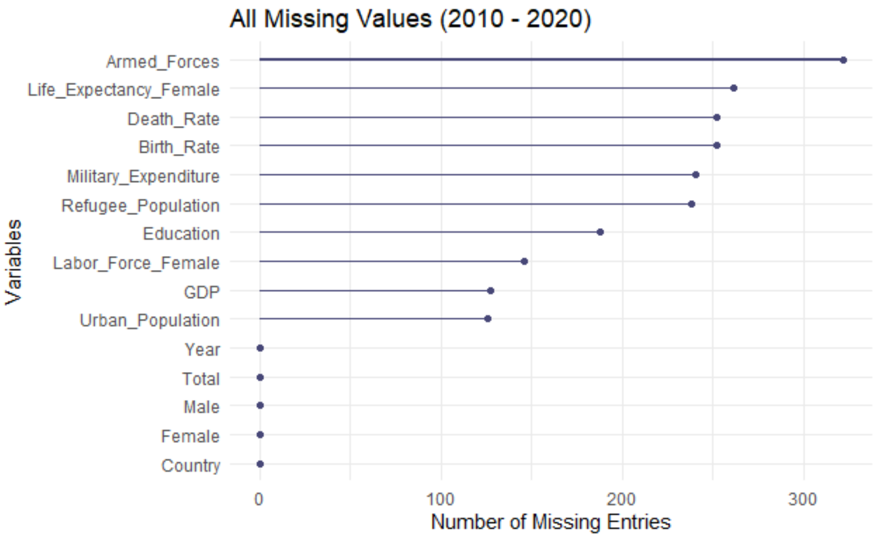


Figure 1: All Missing Values

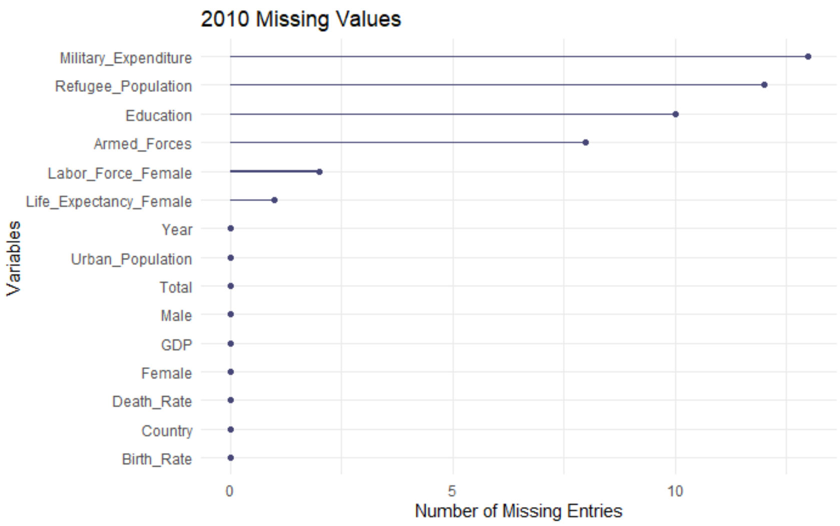


Figure 2: 2010 missing values

As there wasn't more than 30% of the values missing for any variable, an imputation function was used to estimate appropriate values where data was missing, this method was also appropriate for most variables, the values didn't drastically change much each year, so sensible predictions could be made. The only values where predictions couldn't be made were for death rate and birth rate, therefore these columns were dropped from the data set and were not used in the analysis. **194**

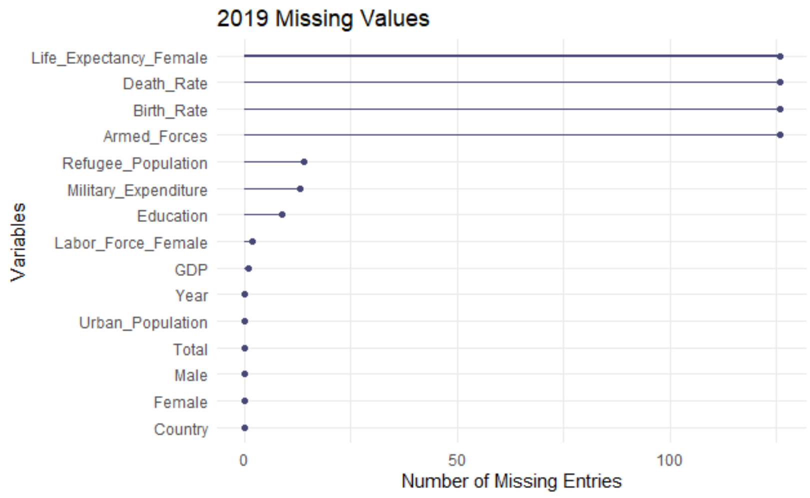


Figure 3: 2019 Missing Values

2.3 Cluster Analysis to Identify Similar Groups of Countries

In order to answer the first research question, cluster analysis methods were used to group countries with similar characteristics. The average variables of the similar countries in each cluster will help to identify how different characteristics correlate with each other to form the different groups. Cluster analysis was carried out in two stages, the first grouped countries by looking at the country characteristic variables against total contributions, and the second stage looking at country characteristics against percentage of female contributions. This meant a percentage of the female contributions column had to be calculated, by dividing the totals column by the female column. Entries of years where a country had made zero contributions were removed from the dataset as no percentage could be calculated.

Cluster analysis was a useful method for identifying the patterns and similarities between a countries' characteristics and the total number of contributions of personnel and the patterns relating to the what percentage of those personnel are female. It also allows for identification of countries with abnormal values and displays the average values for the groups. The K-means clustering algorithm was used especially to find groups which have not been explicitly labelled in the data. Overall, this showed which countries were grouped together according to the characteristics selected beforehand (see section 2.1). The clustering analysis focused on the years 2010 and 2019 so that the dataset set sizes were more manageable, because larger datasets can result in a very long run time. The default euclidean distances method was used for identifying cluster observations and for each analysis carried out, the elbow, silhouette and gap statistic methods were used to determine the optimal number of clusters each time.

For both stages of cluster analysis, the characteristic variables were subgrouped to be separately analysed. All subgroups would contain either the totals column or percentage female column (depending on the section) and compared with 3 other related variables. Clustering is a useful method, as it explores how different countries are grouped as a result of their contributions and socio-economic status. It can help determine whether there are particular variables that strongly affect the groupings, and also shows patterns in the groups, for example wealthier countries could be grouped together. Clustering will indicate which variables are most influential, and it will be interesting to see if it is replicated later in the predictive model. **392**

2.4 Modelling to Identify Predictors of Personnel Contributions

To identify variables that were the strongest predictors of a country's total personnel contributions, a regression model was created. The model will determine which structural characteristics of a country are statistically significant for predicting the number of personnel contributions. The response variable investigated was 'Total', which was the total number of personnel contributed per country per year. One difficulty here was that the contributions to different missions could not be considered, or the ratio of male to female personnel, so this restricted more detailed conclusions being reached. The independent variables tested in the model were the eight socio-economic and socio-demographic variables introduced in section 2.1.

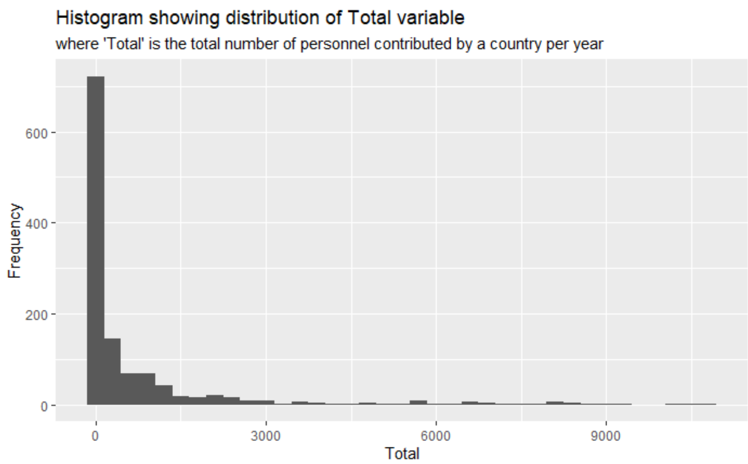


Figure 4: Histogram showing the distribution of the Total responses. There is a large group of zero contributions, and then the distribution tails off to the right.

The response variable is a count variable (positive whole number), so a count regression model was used (Faraway, J., 2006). Initially, a poisson regression model was trialled. There were certain assumptions of the data that need to be checked first to make sure it was suitable for this particular model.

For poisson regression these were: each observation is independent; the distribution of counts follow a poisson distribution; and the mean and the variance of the count (response) variable are equal (Zach, 2019). The observations were all independent as they represented data from different countries, in different years. The response variable ('Total') had a poisson distribution, which is a discrete distribution where there is a high number of zero counts, in this case a high number of incidents where total personnel contribution from a country was zero. Figure 4 shows a histogram of the Total variable. Finally, the mean and variance was checked and this was not equal, which suggested that there could be possible overdispersion when the model was run. However, this was not an immediate problem, as there were other models that could be developed once this was confirmed, such as a negative binomial model. **302**

3 Results

3.1 Exploratory Data Analysis

3.1.1 Location EDA

During exploration of the data, contributions were investigated in relation to where the missions were located. Previous literature on the topic of UN peacekeeping suggested that the location of a contributing country may influence how many personnel a country will contribute. For example, countries that share borders with other countries going through conflict may be more inclined to provide more support if the fighting is negatively affecting their country. Passmore et al. (2018) discuss the negative 'spillover' effects that can occur when conflict expands beyond country borders, for example economic concerns where trade may be impacted. This is a particularly big risk in continents such as Africa. Due to these conclusions, it was important to check location to see if there were any patterns here that may impact later analysis.

To investigate this, current missions with the highest total personnel contributions were selected. Mission and contribution statistics for the 5 largest current missions are shown in table 3. Active missions were chosen as it would allow investigation into countries that are currently contributing. The largest active mission is Monusco, the United Nations Organization Stabilization Mission in the Democratic Republic of the Congo.

Table 3: Location and contribution details for the 5 largest active missions (for personnel contributions).

Mission	Location	Total Personnel Contributions	Top 3 Contributors
Monusco	Democratic Republic of Congo	188,741	India, Pakistan, Bangladesh
Unifil	Lebanon	162,810	Italy, Indonesia, France
Unmiss	South Sudan	111,480	India, Rwanda, Nepal
Minusma	Mali	86,378	Bangladesh, Burkina Faso, Chad
Minusca	Central African Republic	74,783	Rwanda, Pakistan, Cameroon

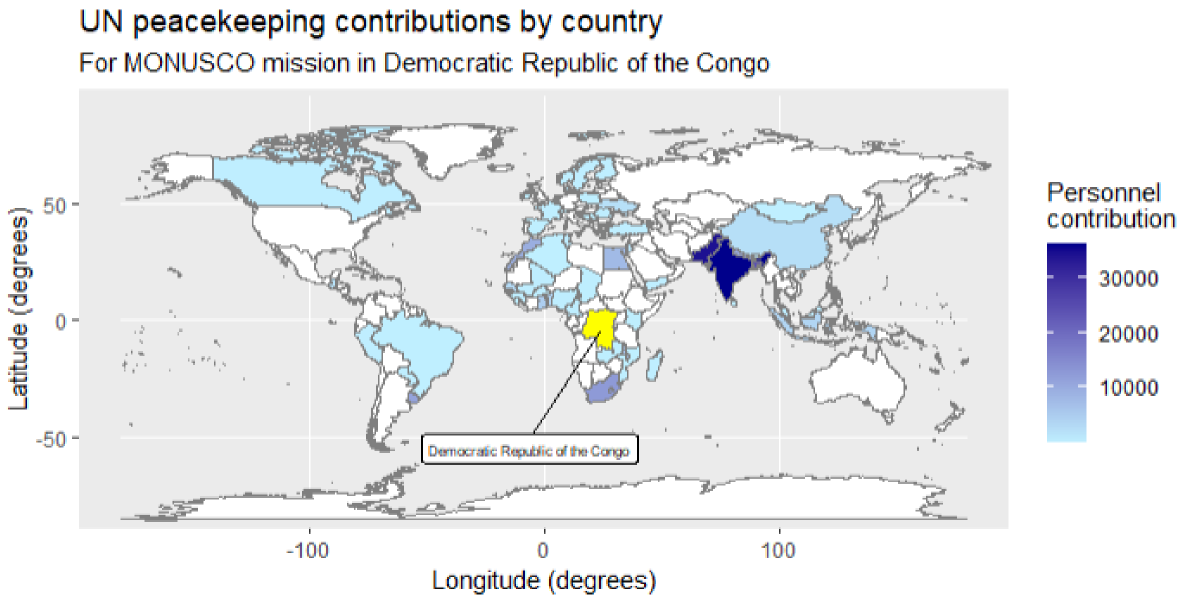


Figure 5: A map showing the contributions to the UN peacekeeping Monusco mission. Mission location (Democratic Republic of the Congo) is highlighted in yellow, and the countries are colour coded depending on size of personnel contribution.

Figure 5 is a map showing contributions to the Monusco mission. The DR Congo is highlighted in yellow, and the countries shaded in blue are those that have contributed to the mission, with dark blue representing the biggest contributors and light blue showing the smallest contributors. Here it is quite clear that the biggest contributors are not close to the mission location. However, there are still a reasonable number of countries close to DR Congo that are contributing. This is expected, and in line with the conclusions drawn by Passmore et al. (2018) as mentioned above. The map (Figure 5) and Table 3 show that the highest contributors are India, Pakistan, and Bangladesh. This is interesting, as these countries are not very close to DR Congo, so it raises questions about their motivations for supplying so many personnel. One suggestion here is that personnel from these countries do not cost a lot to deploy, so the financial compensation that is provided by the UN makes it a profitable scheme for these governments. (Gaibulloev et al. 2015). This is not related to location factors, however it is important to consider as it suggests a possible reason for high contributions from less developed countries. This relationship may become apparent through modelling with the variables chosen in this investigation, for example with GDP.

The next mission looked at was Unifil, the United Nations Interim Force in Lebanon. Figure 6 (page 12) shows the countries that have contributed to this mission, again colour coded by size of contribution. It is clear that a lot more European countries have contributed more strongly to this mission, compared with the contributions to Monusco. These differences indicate support for the location factor, as there is more European involvement, and less African involvement, which is the reverse to what was seen for Monusco. There is strong involvement from countries within the Asian continent that the mission is based in, in particular India and Indonesia. However, as seen earlier with Monusco, it is hard to conclude whether this is due to location or financial gain.

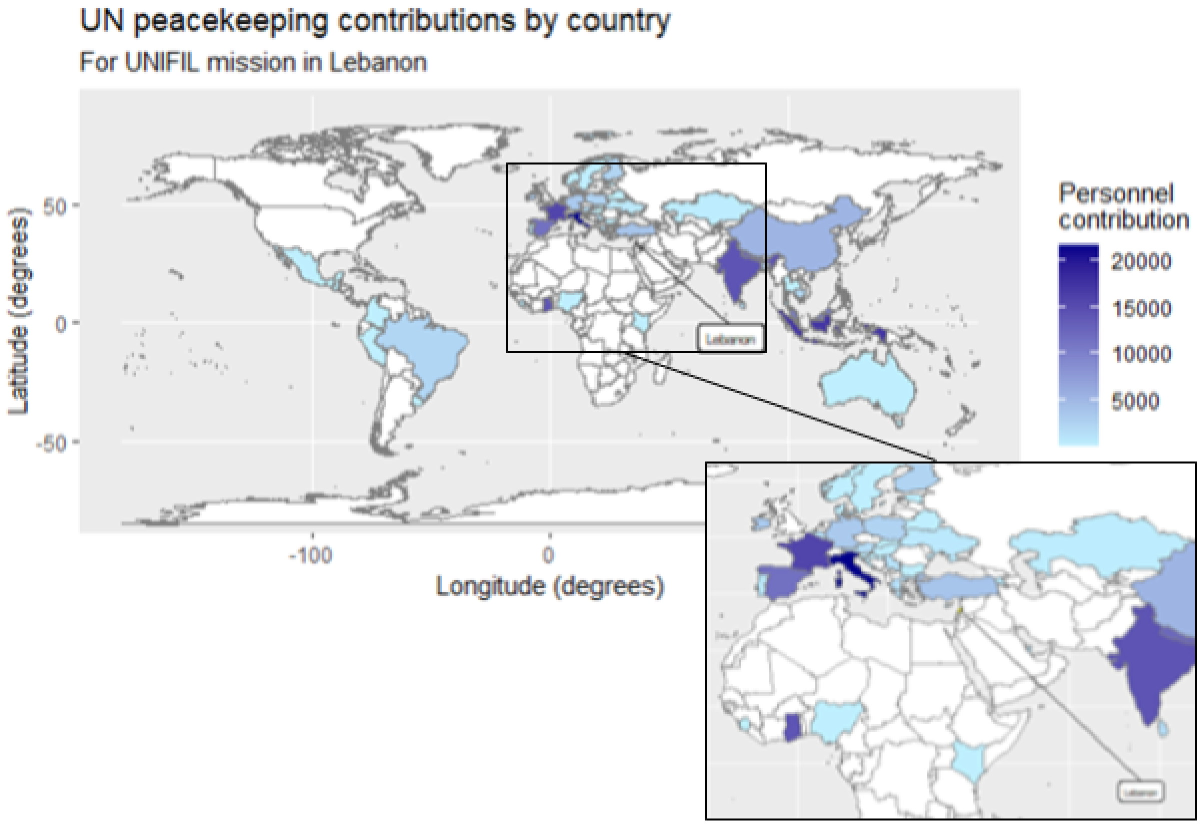


Figure 6: A map showing the contributions to the UN peacekeeping Unifil mission. Mission location (Lebanon) is highlighted in yellow, and the countries are colour coded depending on size of personnel contribution.

Overall, it is apparent that there are potential location and economic factors that are affecting contribution to UN peacekeeping missions. As this research does not consider a variable for location in the statistical analysis, this is something to take into account when the results are discussed. **580**

3.1.2 Personnel Contribution EDA

Figures 7 and 8 (page 13) display personnel type contributions for each separate mission. This analysis was carried out to understand which personnel types took part in which missions, and the extent to which they participated in these missions. The years 2010 and 2020 were selected so that the difference could be compared over a decade. Figures 9 and 10 (page 13) summarise the total mission contributions by personnel type.

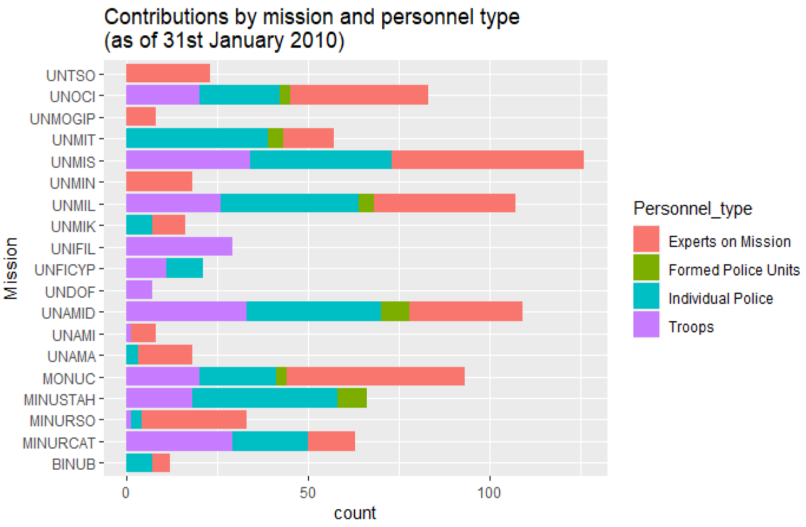


Figure 7: Contributions by mission (2010)

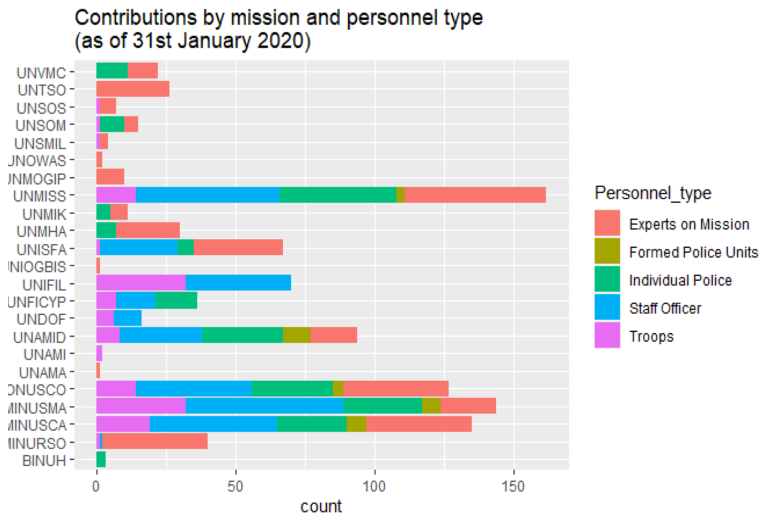


Figure 8: Contributions by mission (2020)

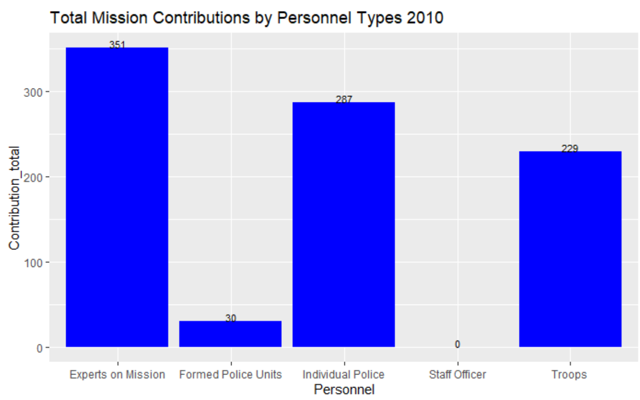


Figure 9: Total mission contribution (2010)

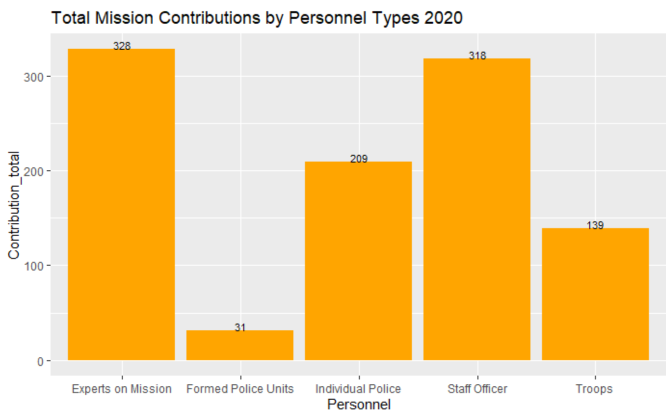


Figure 10: Total mission contribution (2020)

There are several key observations from these two sets of graphs one of them being more missions in 2020 compared to 2010. There were 0 staff officers contributions in 2010 but this changes by 2020 as they start taking part in missions, which increases the number of personnel types taking part in the missions overall, despite a decrease of Troop personnel contributions. Experts on Mission remain the highest personnel contribution in missions from 2010 - 2020. Finally, the most contributed to mission type remains the United States Mission in South Sudan (Unmiss) over the decade.

Percentage of Female Contributions and Socio-Economic Characteristics

In this section, a series of plots show the relation between the percentage of female contributions and the structural characteristics of a country. First, Figure 11 (page 14) shows how the percentage of female contributions has changed between 2010 and 2020. It is clear that the percentage of female personnel has drastically increased in recent years. Figure 12 (page 14) depicts a weak correlation between the percentage of females and GDP, however there is a slight U shape to line, suggesting the percentage of female contributions are higher for much lower GDP countries and higher for the high GDP countries, but the percentage drops for those in the mid range.

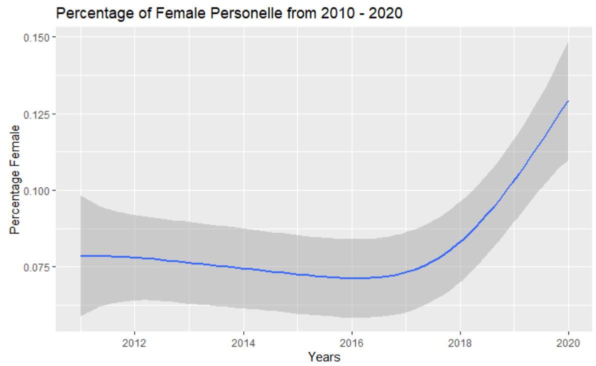


Figure 11: Percentage of Female Personnel from 2010-2020

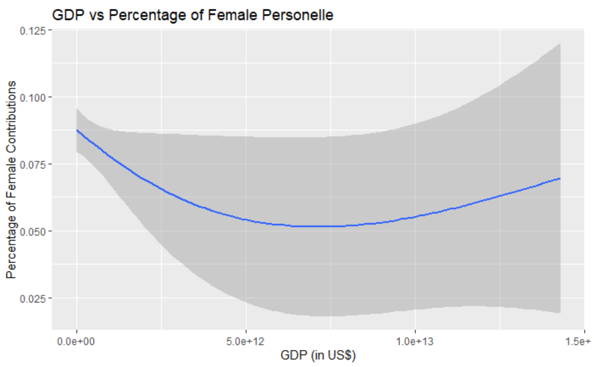


Figure 12: Graph plotting GDP against Percentage of Female Personnel

The line graph in figure 13 plots the percentage of a country's workforce that is female against the percentage of female contributions. It shows a line that is generally positively correlated. Figure 14 shows a negative correlation between a country's military expenditure and the percentage of their peacekeeping contributions that are female. **335**

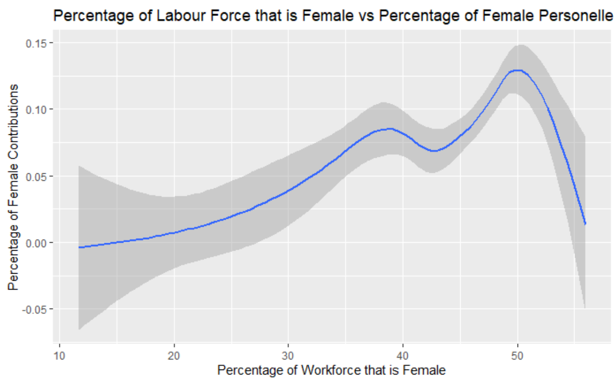


Figure 13: Percentage of female labour force against female personnel

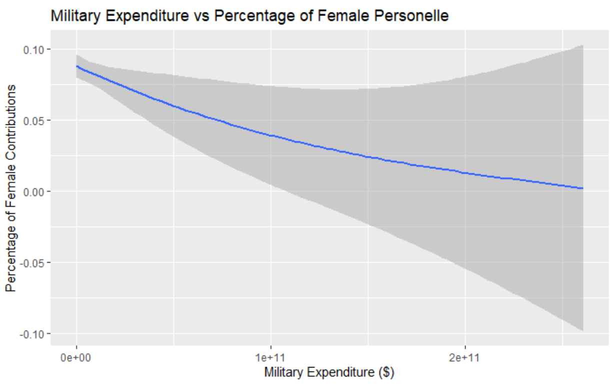


Figure 14: Military Expenditure against percentage of female personnel

3.2 Statistical Analysis

This statistical analysis section describes the results of the two stages of cluster analysis and the regression model. Countries were successfully clustered in both stages, allowing for identification of patterns of variables and what combinations of variables are the most common for high or low total contributions and percentage of female contributions. Statistically significant variables for predicting contributions were identified with the regression model.

3.2.1 Cluster Analysis Results

As outlined in the methodology, two stages of cluster analysis were carried out. The focus variable for stage one was the Totals column and the focus variable for stage two was the percentage female column that had been calculated after data exploration had been done. Both these stages compared the focus variables to subgroups of similar variables to find meaningful patterns in the data. The results of these analyses are outlined below.

Cluster Analysis Stage 1: Totals

This stage of the cluster analysis consisted of grouping countries together depending on the total (female and male) amount of personnel they had. This helped identify patterns that might lead to a country contributing higher or lower levels of total personnels. Two subgroups were selected in this stage which all included the total personnel. As mentioned before, the focus is on two particular years (2010 and 2019), therefore these two years are displayed in separate clusters. Cluster one represents 2010 and cluster two represents 2019.

Subgroup 1: GDP levels against personnel contribution

Subgroup one shows the total personnel contribution against ‘Armed forces’, ‘GDP’ and ‘Military expenditure’. These characteristics are selected to help display the amount of GDP a country has and if that correlates with higher total personnel contribution.

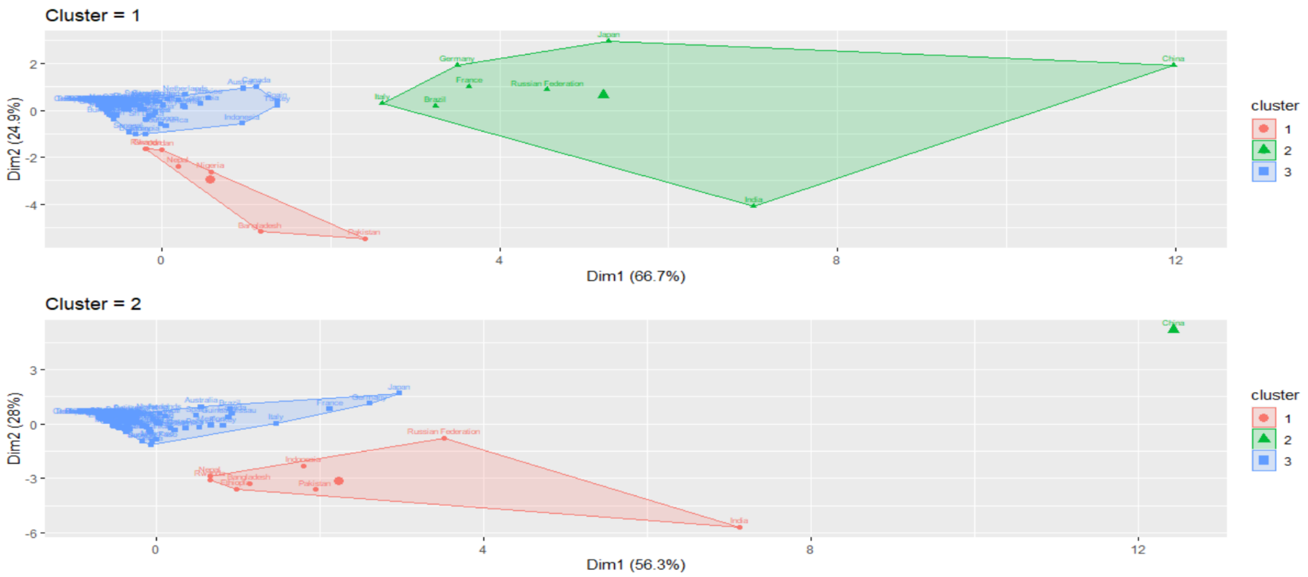


Figure 15: Cluster plots for subgroup 1 (totals)

After carrying out the first cluster analysis there are 3 optimal cluster groups identified for both years. However, figure 15 shows that China and India skewed the dataset as they were much larger in contribution than the other countries. Therefore, they were removed and the analysis was repeated in order to have a more meaningful cluster analysis and see where else there are patterns.

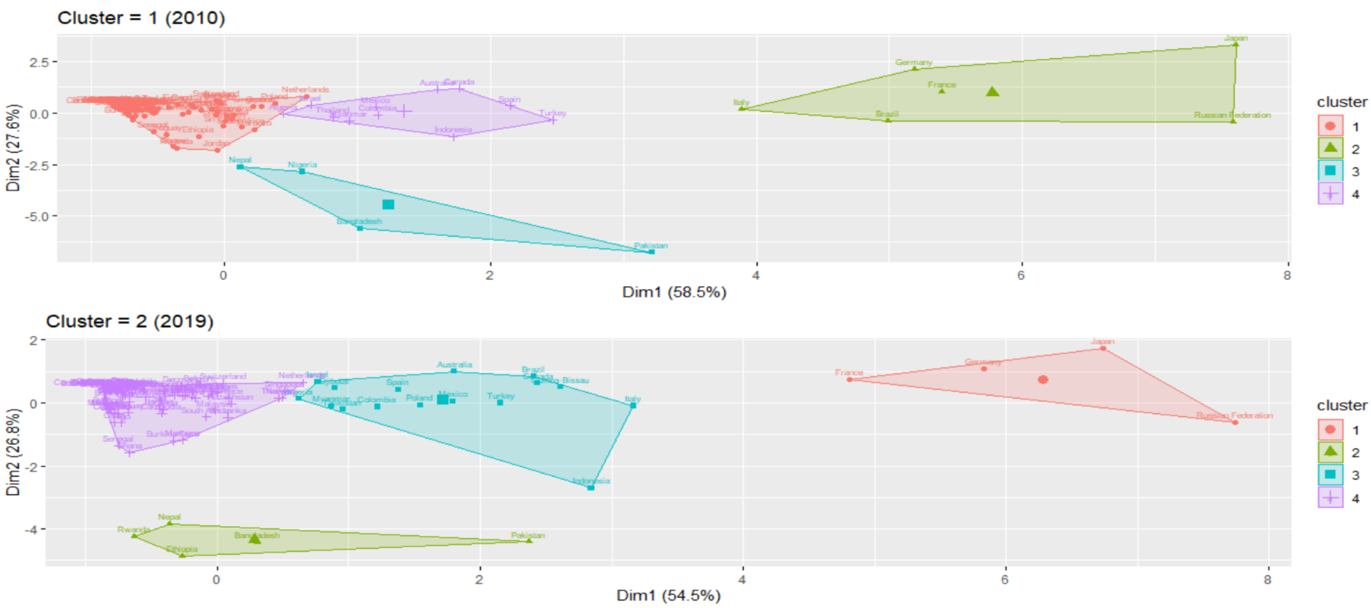


Figure 16: Cluster plots for subgroup 2 (totals)

After carrying out the cluster analysis again, there were four optimal clusters identified for both years as shown in figure 16. This supports the decision to remove China and India, as there are more cluster groups identified this time increasing from three to four. The outcome from the four clusters for 2010 showed that countries with higher total personnel contribution also have higher armed forces and military expenditure however they have lower GDP. The countries included are Bangladesh, Russia and Nigeria. For 2019 it shows that the highest total personnel contributions have average levels of armed forces, but still low GDP and low military expenditure. Countries included are Rwanda, Pakistan and Bangladesh. Overall, this shows that higher GDP countries does not equal higher total personnel. This helps support the previous literature review carried out by Gaberiella (2021). Also, the comparison of the two clusters shows that there is not that much of a difference almost a decade later.

Subgroup 2: Population against total personnel contribution

Gaibulloev et al. (2015) highlighted the fact that higher population levels allow for greater personnel contribution as there are higher recruiting options. Therefore, this cluster analysis explores this idea and examines whether new information can be presented for this idea or further conclusions made. The variables selected are total personnel contribution against 'Refugee Population', 'Urban Population' and 'Education'. These characteristics are also selected to help show if the population and characteristics of the population such as education has an effect on the total personnel contribution a country makes.

For 2010, there were 6 optimal cluster groups identified (see figure 17, page 17). The cluster with the highest total personnel contribution also had the highest refugee population by a significant level, the 3rd highest urban population and second highest compulsory education levels with 11 years as the average. Countries included here are Jordan and Pakistan. The second cluster with the highest total personnel contribution also has a high refugee population, 2nd highest urban population but low years of compulsory education with 7. Countries included are Bangladesh and Nepal. The cluster with the lowest total personnel contribution has a low refugee population, low urban population and 3rd highest compulsory education years with 10. Countries included are Thailand and Sri Lanka. Overall this analysis shows that higher population and in particular higher refugee population does have a correlation with higher total personnel contribution. However, this is not always the case as looking at the table in appendix 1.1 the cluster with the highest urban population has lower total personnel compared to the two highest, lowest refugee population and average compulsory education years with 9. This cluster analysis also shows that compulsory education years do not correlate with total personnel contributions. Refer to appendix 1.1 for further information on this table analysis.

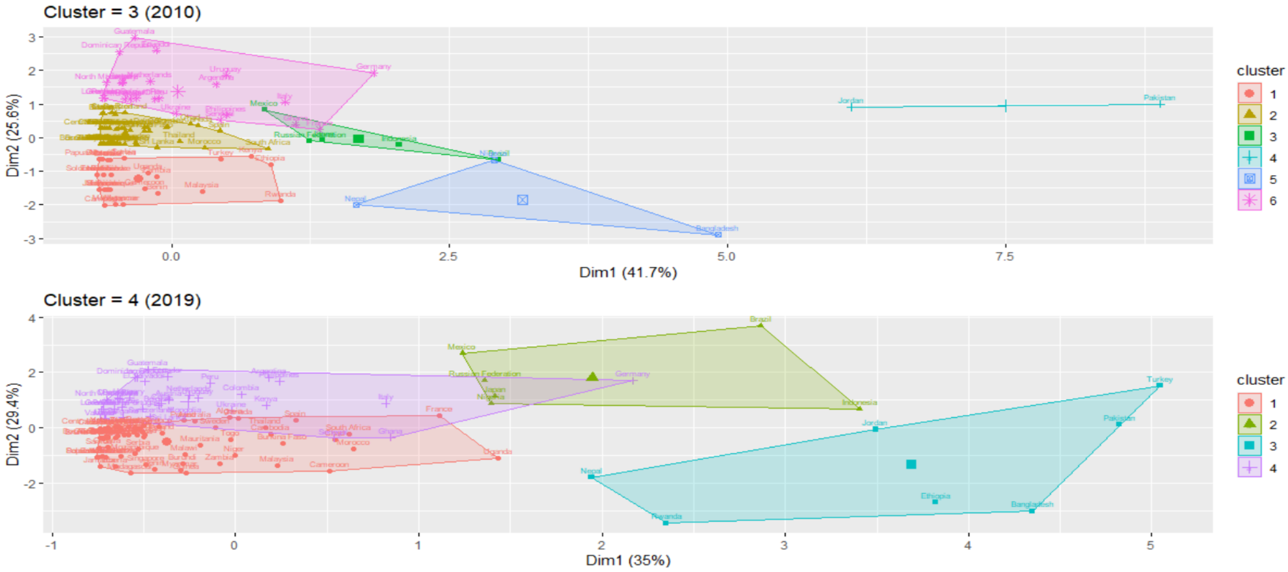


Figure 17: Cluster plots for subgroup 2 (totals)

For 2019 there were 4 optimal clusters identified. The cluster with the highest total personnel contribution also had the highest refugee population, 2nd highest urban population and 9 years of compulsory education which, compared to the other 4, is low. Countries included here are Turkey and Bangladesh. The cluster with the lowest total personnel contribution had a high refugee population, lowest urban population and lowest years of compulsory education. Countries included are Uganda and Niger. This reinforces the idea that higher populations have higher personnel contributions, as mentioned in the existing literature, section 1.2.

Cluster Analysis Stage 2: Percentage Female

Stage 2 of the cluster analysis involved finding groups of similar countries when clustered with the inclusion of the percentage female variable, with the aim of identifying common patterns that may lead to a country contributing a higher or low percentage of female personnel. Three subgroups of data were analyzed in this stage which all included percentage female personnel. Group 1 looked at a general socio-economic overview with GDP, Total contributions and average years of education for the country's citizens. Group 2 looked at military related variables and refugee population, and group 3 looked at female and workforce related columns, as well as urban population.

Sub group 1: General Socio-Economic Overview

The first group of variables clustered were those relating to general socio-economic characteristics. It consisted of GDP, total contributions and average years of education (percentage female was included in all groups as well). For data recorded in 2010, 6 clusters of countries were identified (shown in figure 18). The countries with the lowest percentage of female contributions at 2.3 % were those in cluster 6. The variables consisted of a high GDP and total contributions and an average of 8 years of education, examples of countries include pakistan, india and Nigeria. The cluster with the highest percentage of female contributions is 4 at 65% female, these countries also had the lowest average GDP, the lowest total contributions and an average of 9.7 years of education. Cluster 4 only consisted of 3 countries, Palau, Iceland and Nambia. The abnormally high percentage of female personnel in this cluster can be accounted for by the low levels of total contributions. The full details of the average variables of each cluster are outlined in figure 19.

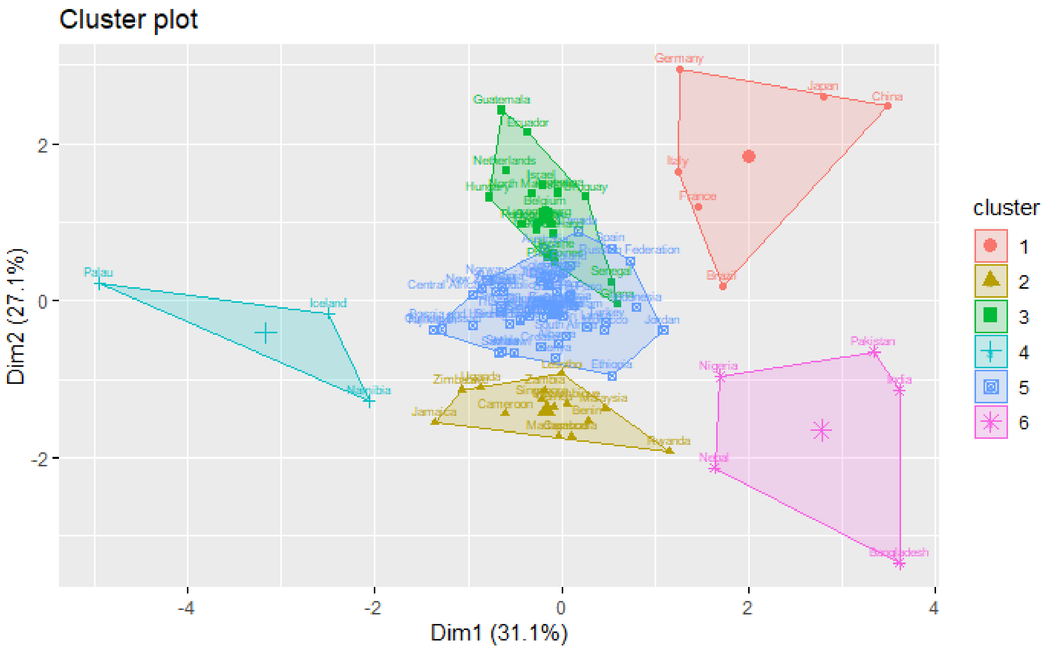


Figure 18: Cluster plot for group 1 with 2010 data

Cluster	Percent Female	GDP	Total	Education
1	2.98 %	3.69e+12	1408	10.2
2	8.8 %	4.64e+10	539	6.1
3	3.9 %	1.87e+11	642	12.4
4	64.9 %	8.38e+09	29	9.7
5	6.46 %	2.25e+11	374	9.2
6	2.28 %	4.70e+11	8197	8.0

Figure 19: Table outlining average values of countries in each cluster for group 1 in 2010

In 2019, 5 clusters were identified. Cluster 1 consisted of China alone, due to it having a much higher GDP than other countries, but not as high of a number for total contributions. It also has a lower percentage of female personnel (at 3.0%) compared to the other high GDP countries in cluster 2 (13%). Countries in this cluster include Germany, Italy, Belgium and Switzerland, this cluster also had the highest years of education. A full table of the average results of group 1 2019 is displayed in appendix A2 and a cluster plot of countries in appendix A3.

Sub group 2: Military

The next group of variables used to cluster countries consisted of refugee population, armed forces and military expenditure. These were grouped together to observe how military related variables affect the percentage of female contributions in a country. With the 2010 data, 6 clusters were identified (see figure 20). The clusters containing the most countries were clusters 2 and 6. Cluster 2 can be characterised by having a high percentage of female personnel (19.0%), a lower military expenditure, armed forces and moderate refugee population. Some examples include Iceland, Australia and Jamaica. Cluster 6 is the majority of countries who have a much lower percentage of female personnel (2.6%), a lower refugee population and low military expenditure. Figure 20 and figure 21 (page 20) show there are countries with more extreme values and these are not in the majority clusters 2 and 6. Notably Palau has a 100% female contribution, China and India in cluster 1 have the highest military expenditure but only 2% female personnel, and Pakistan and Jordan in cluster 4 have 0.3% female contributions.

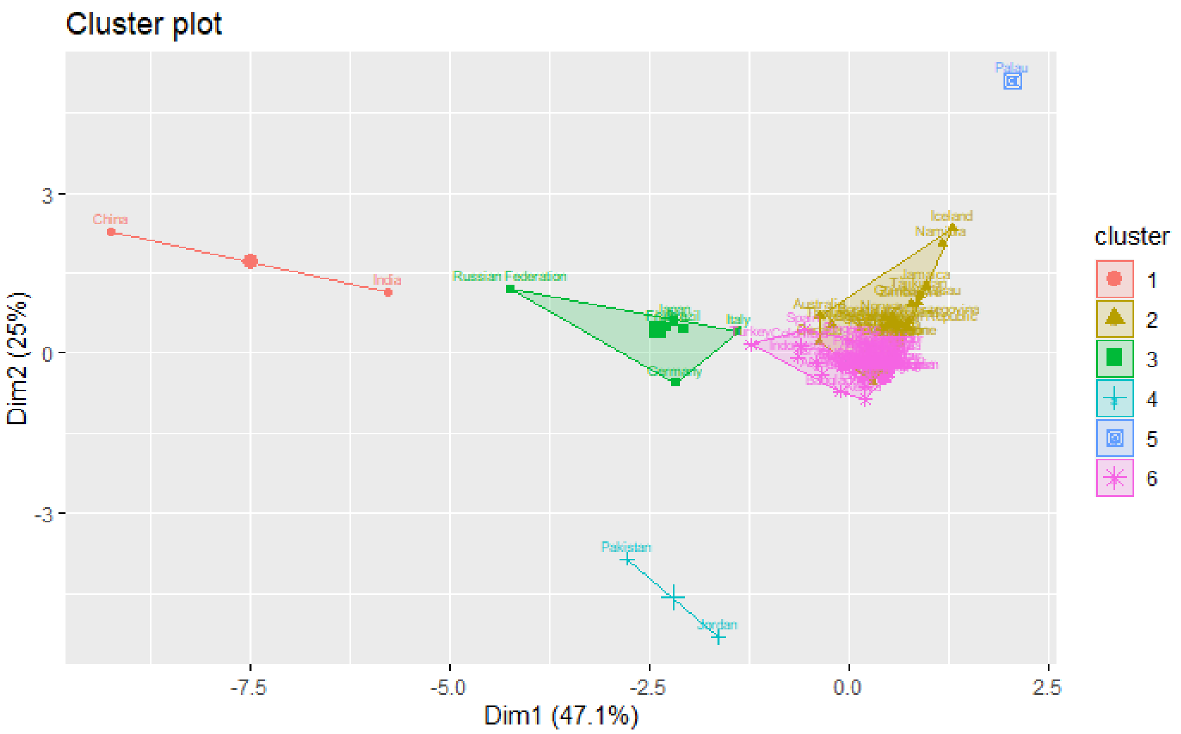


Figure 20: Cluster plot for group 2 with 2010 data

Cluster	Percent Female	Refugee Population	Armed Forces	Military Expenditure
1	2.1 %	242, 899	2, 785, 293	81, 045, 222, 829
2	18.96 %	49, 160	46, 694	3, 305, 210, 348
3	2.7 %	143, 865	559, 445	46, 049, 409, 852
4	0.28 %	2, 146, 979	528, 250	3, 766, 250, 250
5	100 %	287	41, 500	196, 945, 590
6	2.6 %	25, 072	89, 789	2, 506, 727, 710

Figure 21: Table of average values of countries for group 2 in 2010

Cluster analysis of the 2019 data gave 4 clusters, which were similar in structure to those identified in 2010. The Palau point is gone as they made 0 personnel contributions in 2019, and the 2 majority clusters have combined into one at cluster 3 (as shown in figure 22). The average values for the majority cluster are female contributions at 10.7%, low refugee population, and low military expenditure. The low female contributing cluster from 2010 that contained Jordan and Pakistan has grown in 2019 to include 5 more countries, and Russia has joined China and India in cluster 1 due to their increase in military expenditure. Cluster 4 is unique to the 2019 analysis, containing the countries Zimbabwe, Solomon Islands, Japan and Costa Rica. This cluster of countries have very high female contributions at 75%, a low refugee population and a moderate military expenditure. A full table of average cluster values for group 2 in 2019 is given in appendix A4.

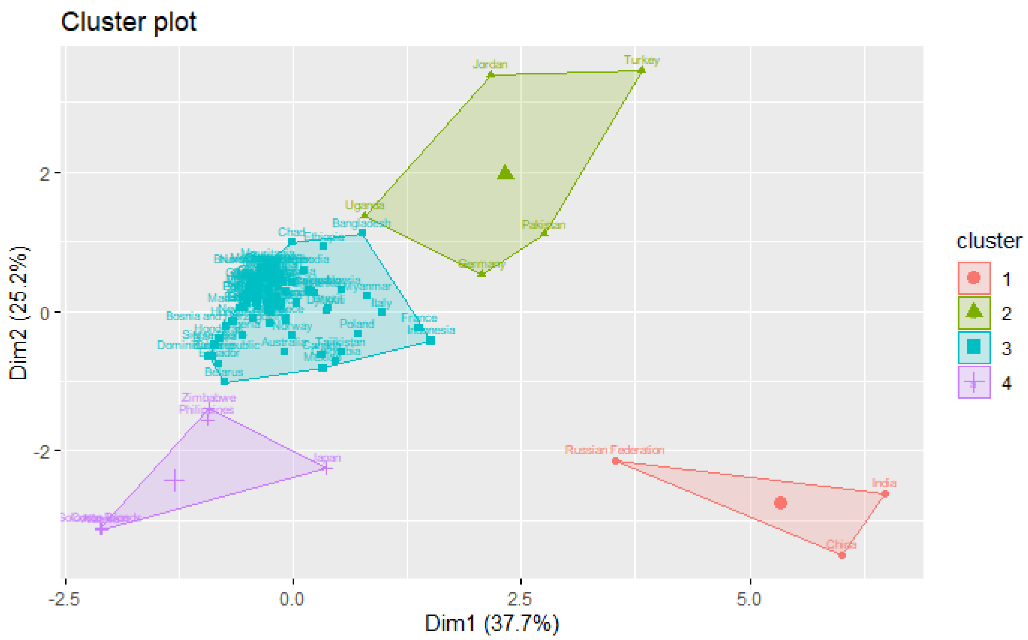


Figure 22: Cluster plot for group 2 with 2019 data

Sub group 3: Female Variables & Urbanisation

The final subgroup analysis with percentage of female contributions contained the urban population size, the percentage of the country's workforce that is female, and the average female life expectancy. 5 clusters were found with the 2010 data (shown in figure 23). The clusters with the highest percentage of workforce being female (clusters 1 and 3, with 47.9% and 47.7% respectively) also had the highest percentage of female contributions (at 64.9 % and 7.7%). The countries in cluster 1 are Namibia, Palau and Iceland, and countries in cluster 3 include Nigeria, Ethiopia and Cambodia.

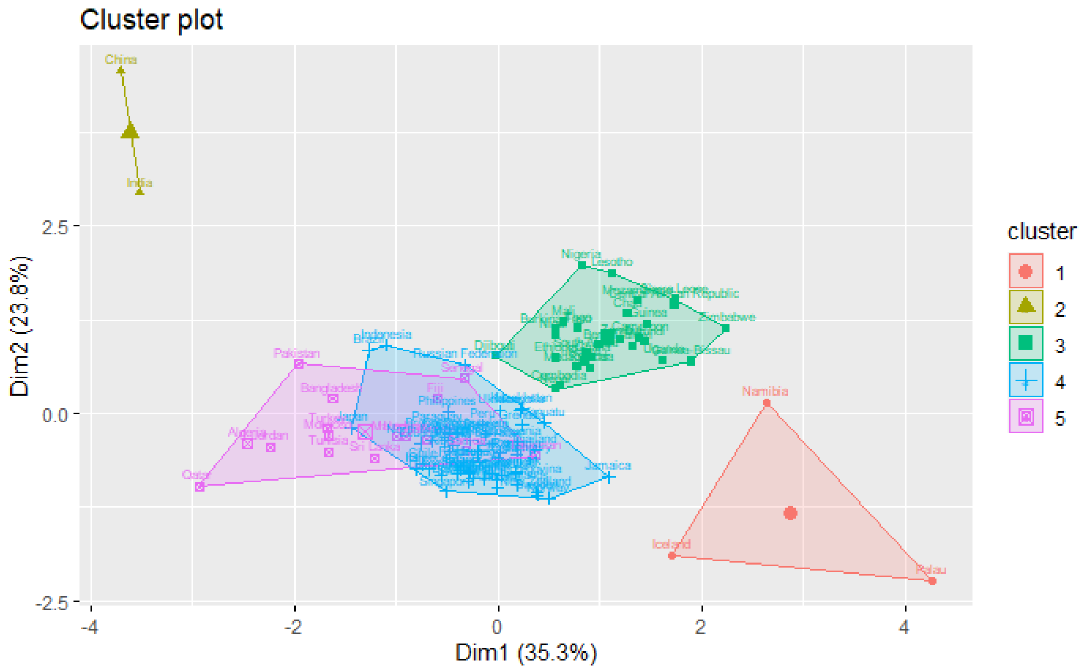


Figure 23: Cluster plot for group 3 in 2010

Cluster	Percent Female	Urban Population	Labour Force Female	Life Expectancy Female
1	64.9 %	397, 609	47.9 %	72.5
2	2.1 %	520, 130, 914	33.6 %	72.1
3	7.7 %	7, 916, 438	47.7 %	58.3
4	5.8 %	20, 426, 087	44.5 %	79.7
5	3.3 %	16, 259, 473	28.3 %	74.1

Figure 24: Table of average values of countries in cluster for group 3 in 2010.

The 2019 data also returned 5 clusters of countries (shown in figure 25, page 22). More clusters contained a higher percentage of female contributions in 2019, with clusters 3, 4 and 5 having 75%, 10.3% and 12.2.% of their contributions as female. These countries also had the highest percentage of their labour force female compared to the other 2 clusters of countries. The cluster with the highest urban population (cluster 1 containing China and India), had the lowest percentage of female contributions. 1776

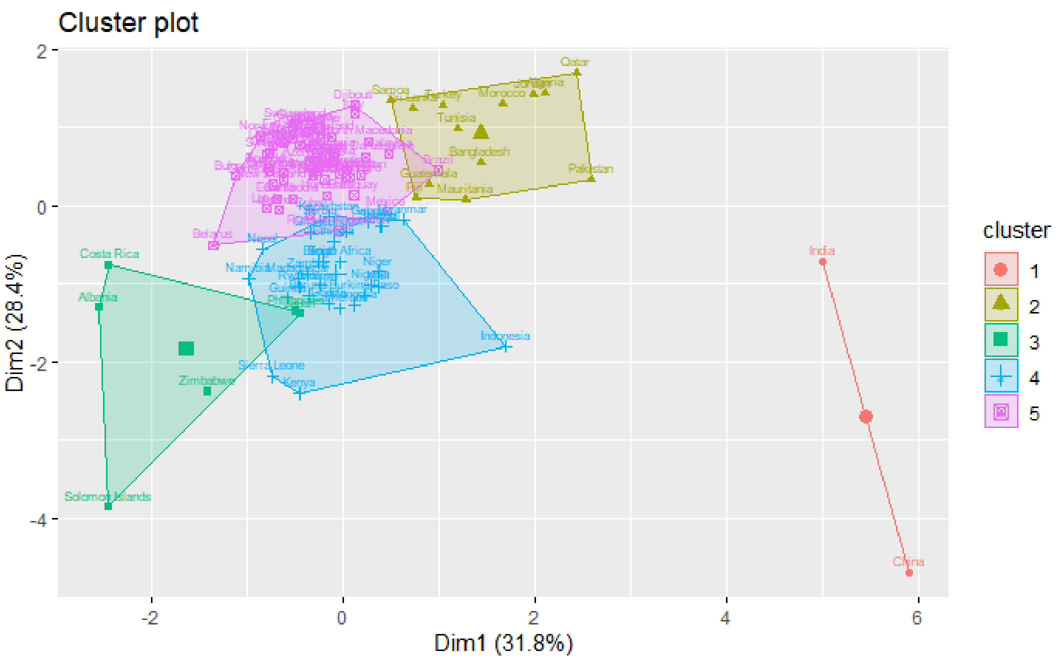


Figure 25: Cluster plot for group 3 with 2019 data

Cluster	Percent Female	Urban Population	Labour Force Female	Life Expectancy Female
1	2 %	6, 569, 82, 745	31.9 %	63.8
2	3.6 %	22, 632, 292	27.2 %	74.6
3	75.6 %	29, 571, 211	44.5 %	69.4
4	10.3 %	15, 924, 142	46.4 %	64.3
5	12.2 %	19, 041, 511	44.9 %	80.4

Figure 26: Table of average values for group 3 in 2019

3.2.1 Regression Modelling

Before building the model, the variables were compared to see if any shared a strong correlation. Figure 27 shows a colour-coded correlation matrix, where dark blue indicates strong positive correlations, and dark red indicates strong negative correlations. These relationships are important to know as they indicate where collinearity may arise between variables, and this can have a detrimental effect on models. It is clear that urban population, armed forces, GDP, and military expenditure are highly correlated, so the final model will likely not contain all of these variables together.

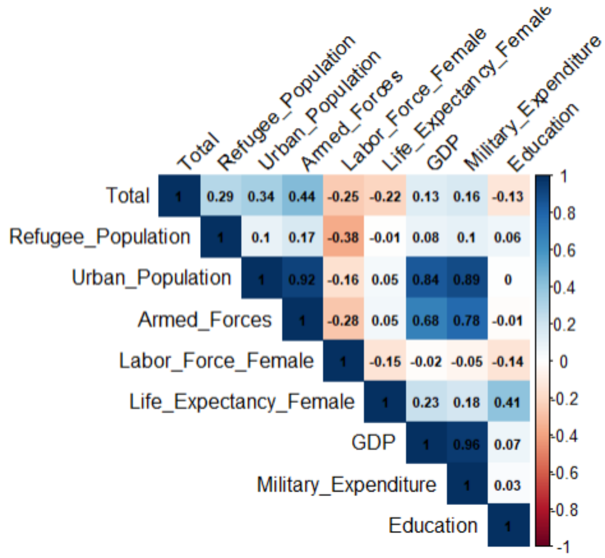


Figure 27: Correlation matrix showing the correlation between each of the socio-economic and socio-demographic variables. Dark blue indicates a strong positive correlation, and dark red indicates strong negative correlation).

Next, the dataset was split into train (dimensions 838 x 9) and test datasets (dimensions 372 x 9), using a 70:30 ratio. This enabled us to train a model and then test the predictions using a second dataset in which the response variable ('Total') had been removed. The first model created was a poisson regression model, and within this all the variables were significant and had p-values less than 0.05. However, the dispersion value was much greater than it should have been (1936.154), which indicated overdispersion. For this type of model, dispersion should be around 1, so the high value indicates the model was not suitable and did not fit the data well. To improve the results, a second model was run using quasi-poisson. The coefficients remain the same in this type of model, however the standard error calculations were more robust and gave better standard errors, which made the p-value significance results more accurate. In this second model, the urban population variable was insignificant (a p-value of 0.11414, which is greater than the significance level of 0.05). This was removed using the backwards elimination method, where the initial model contains all possible variables, and insignificant variables are removed individually. Next, a collinearity check was carried out, and military expenditure was removed as it had a very high value (35.474752) compared to the other variables. After this, the model was run again, and GDP was insignificant (p-value of 0.93063), so it was removed. The final quasi-poisson model was:

$$Total \sim Refugee\ population\ +\ Armed\ forces\ +\ labor\ force\ female\ +\ life\ expectancy\ female\ +\ education$$

Due to the poisson model having a high dispersion, a negative binomial model was created using backwards elimination to see if the dispersion was improved. Negative binomial models are commonly used in cases of overdispersion as they are able to account for an extra parameter that models the overdispersion (UCLA, n.d.). For this model, the variables GDP and armed forces could not be used, as they had high correlation and caused overfitting. After removing all insignificant variables using the p-value significance level (0.05), the final negative binomial model was:

$$Total \sim Refugee\ population\ +\ urban\ population\ +\ life\ expectancy\ female$$

The AIC (Akaike Information Criterion) for this model was much lower in comparison to the poisson regression model, indicating that it was better (AIC value of 10314 compared to 1089549). It also had a much better dispersion value of 1.259064 (where 1 is the ideal).

To further check the suitability of the final model and compare residuals, half normal plots were run. These looked at the deviance residuals for each model to see if they were acceptable. Figure 28 shows the plots for the three models run (poisson, quasi-poisson, and negative binomial). Half normal plots with simulation envelopes are useful for checking the goodness-of-fit for a model, as they provide a confidence interval (envelope) that the residuals should fall inside if the model is suitable (Moral et al. 2017).

The half normal plot for the original poisson model (a) shows that the residuals are very far from where they should be. The simulation envelope is very flat at the bottom of the graph, and this indicates the model is not suitable. The quasi poisson model (b) is more suitable, with almost all of the deviance residuals plotted within the simulation envelope. Finally, the half normal plot for the negative binomial model (c) shows that all the deviance residuals are within the simulation envelope, and this is evidence that the model fits the data well. Overall, the half normal plots suggest that the quasi poisson model would be suitable for prediction. However, the dispersion parameters and AIC values suggest that the negative binomial is still better.

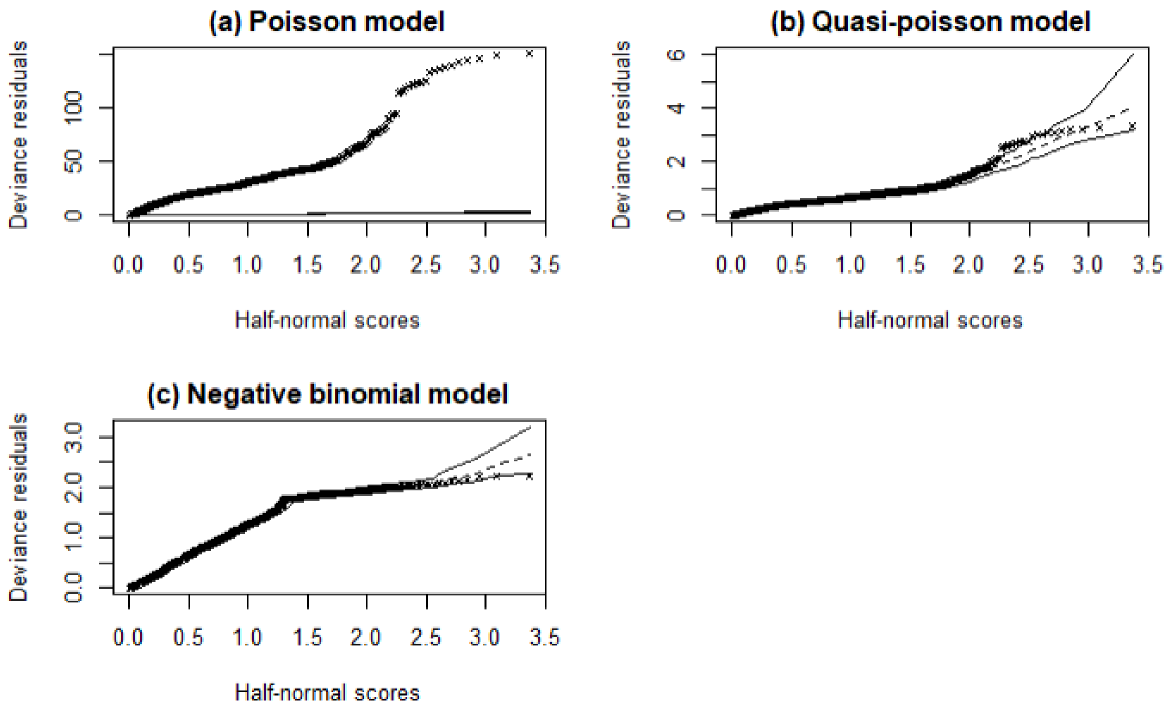


Figure 28: Half normal plots with simulated envelopes. These plot the deviance residuals for each model and indicate whether the model is a good fit.

Table 4 (page 25) provides the coefficients of the variables that were significant in the negative binomial model. Care needs to be taken when interpreting these, as the linear relationship between response variable and independent variables is estimated using log transformed data (Jabeen, H., 2019). Therefore, the relationship is not as straightforward to explain as with a linear model.

Table 4: Negative binomial model coefficients

Variable	**Coefficients**
Refugee population	1.342×10^{-6}
Urban population	7.498×10^{-9}
Female life expectancy	-7.751×10^{-2}

In this case, refugee population and urban population are positively correlated to the response variable, total contribution, and so as these values increase, the contribution total should increase. On the other hand, total contribution is negatively correlated with female life expectancy, so in countries where female life expectancy is high, lower contributions are expected.

After the negative binomial model was trained, it was tested using the test dataset. Contribution predictions were made for all unique country-year combinations in the test dataset, and these were compared with the actual contributions for these same combinations. There were two extremely large predictions, at 125148 and 165785 for China in 2016 and 2018. China appeared to be an outlier in the previous clustering analysis, so this is not surprising. These predictions were removed for the final plot of predictions vs actual contributions in order to see the pattern of the other points properly. Figure 29 shows the results for the predicted contribution values plotted against the actual contribution values.The red line indicates where prediction values = actual values, and for a good model, most of the points would be close to or on this line. This is not the case with this model and there is also no clear pattern, which indicates that the model was not able to predict contribution very well. The smaller values appear to be easier to predict, as there are more values closer to the line in the bottom left corner, however there are still some points where these are massively overpredicted. As the actual contributions increase, they tend

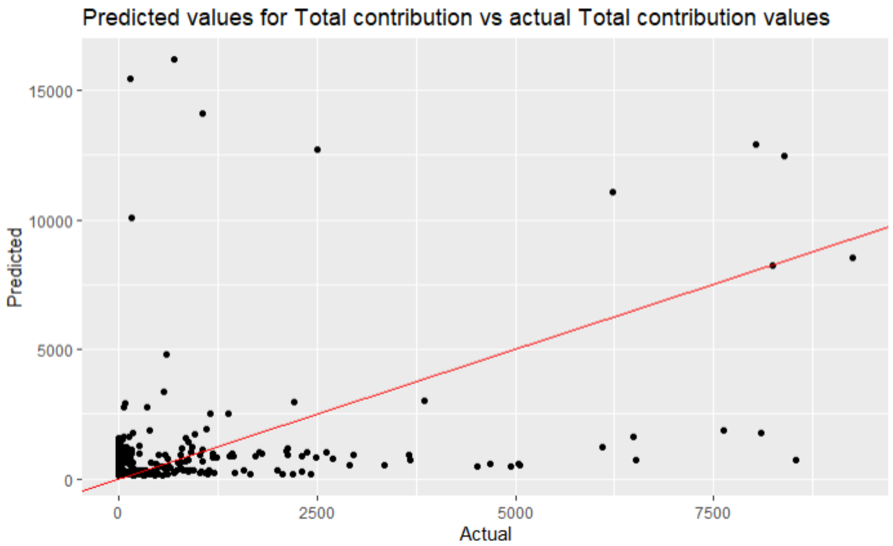


Figure 29: Graph showing the predictions for total contributions vs the actual contributions.

to be underpredicted, as a lot of the higher contributions are below the ideal prediction line.

Something to highlight that is not completely clear in figure 29, is that the lowest predictions for personnel start at 145. This is clearly not the same as what is seen for the actual contributions, where there is a high number of zero counts where countries have contributed zero personnel to UN peacekeeping missions (see figure 4, page 9). **1123**

4 Discussion

The results indicate that there is a correlation between personnel contributions to United Nations peacekeeping and socio-economic or socio-demographic characteristics of the contributing country. The first question being discussed is whether socio-economic and socio-demographic variables can be used to cluster countries that contribute to UN peacekeeping missions. Cluster analysis stage 1 subgroup one shows that higher GDP countries do not cluster with higher personnel, indicating that high GDP does not mean higher personnel contribution. This confirms the claims of previous research mentioned before in section 1.2 (Dos Santos, G., 2021). As for social demographics, cluster analysis stage 1 subgroups show that higher population countries, in particular those with significantly higher refugee populations, are clustered with high total personnel contributions. This indicates that there is a positive correlation between social demographics and personnel contribution. Although, there was one cluster formed where this was not the case (figure 17, cluster 3). However, the majority of the clusters support this conclusion and showed a strong correlation, so it is clear that overall there is a positive correlation between population and personnel contribution. These results also confirm the claims of Khusrav Gaibulloev (2015). Previous research has also shown that there are higher negative spillover coefficients within the UN, showing that contributors' response on whether or not to take part in missions depends on whether they are giving financial or peacekeeping support (Gaibulloev et al. 2015). For financial support, UN missions displayed more publicness, whilst the opposite tends to be true for peacekeeper support (Gaibulloev et al. 2009). This explains why there is lower personnel contribution from higher GDP countries, because they focus on money making personnel deployments rather than contributor-specific gains such as regional stability.

The results from stage 2 of the cluster analysis indicates that socio-economic characteristics can be used to group countries, and this can give insight into the patterns of average values that affect different percentages of female contributions. Sub-group 3 of section 3.5 shows that countries with a higher percentage of females in the workforce contribute a higher percentage of female personnel. It was also indicated that the percentage of female personnel has increased in recent years, with a mean percentage of 7.5% in 2010 and a mean of 14.0% in 2019. This variation is further described by figure 11 (page 14), showing how the percentage has increased over time. UN Peacekeeping has "encouraged and advocated" for more female officers and has set a target of 25% female military staff by 2028 (UN Peacekeeping, n.d., *Women in Peacekeeping*). Encouragement from the UN could be responsible for the increase in the percentage of female peacekeepers shown from 2010 to 2019. Female participation in peacekeeping operations is lower in the early stages of missions, with the uptake increasing once the duration of a mission has increased (Tidblad-Lundholm, K., 2020). It may be possible that 2019 included missions with longer durations, causing the sharper increase in the percentage of female contributions.

Another pattern identified within the cluster analysis is the link between higher military expenditure and a lower percentage of female personnel. As shown in the table of cluster averages in figure 21 (page 20), the countries with higher military expenditure (e.g. China and India from cluster 1 or Pakistan and Jordan in cluster 3) have lower percentages of female contributions. This trend is also demonstrated in the line graph of figure 14 (page 14) Countries with lower military expenditure and lower total contributions contribute higher percentages of female personnel, as demonstrated by countries like Palau from cluster 5 or Iceland from cluster 2. The countries with higher military expenditure were also shown to be the same countries that had a lower percentage of females in the labour force, this is described in figures 25 and 26 (page 22) where it is clearly shown that China and India in cluster 1 have an average female labour force of 31.9%. Therefore, it could be concluded that the higher military expenditure is not directly responsible for a lower percentage of female contributions, but a country with higher military expenditure is more likely to have a lower percentage of females in the labour force, and consequently a lower percentage of female contributions to UN peacekeeping operations. It is important that the number of female personnel involved in operations continues to increase, as female personnel are important for building trust in the communities that peacekeepers are involved with, in particular with the women and children of these communities that the UN missions are aiding (UN Peacekeeping (n.d.). *Women in Peacekeeping*).

The second research question addresses whether socio-economic and socio-demographic variables can be used as predictors for the number of personnel that a country will contribute to UN peacekeeping missions. The regression model analysis showed that three variables were considered significant predictors for contribution to UN peacekeeping. These were: refugee population, urban population, and female life expectancy. Refugee population and urban population had a positive relationship with personnel contribution, which means that as these increased, so did the total personnel contribution. This is logical, as having a higher number of refugees could suggest that the country is close to conflict, and literature has already implied that countries are more likely to contribute if they share a border with a country going through conflict (Passmore et al. 2018). Additionally, larger urban populations are often seen in less developed countries, and these less developed countries are more likely to contribute more to missions as they have cheaper personnel deployment costs so can make more of a profit from UN reimbursements. On the other hand, female life expectancy had a negative relationship with total personnel contribution. This meant that as female life expectancy increased, total personnel contribution decreased. This makes sense, because female life expectancy is generally higher in wealthier countries, where there is good access to better health care and resources which would result in higher life expectancies for the female population. These wealthier countries are less likely to contribute to UN peacekeeping missions as it is not as financially beneficial for them to do so due to higher personnel deployment costs.

In each stage of the analysis, one country whose values were distinct from others was that of China. When conducting the cluster analysis, it was made apparent that China's military expenditure, urban population and GDP were considerably higher than seen in other countries. Also, previous research has shown that China is the second highest contributor to UN funding behind the US, making up approximately 12% of the UN's funding in 2019 (Dos

Santos, G., 2021). However, despite this high expenditure on military and contribution, China is not the top contributor of personnel, ranking 10th in 2010, and 2nd in 2019 below Bangladesh. This reinforces the idea of higher GDP not necessarily equaling higher personnel contribution. **870**

5 Conclusion

In conclusion, investigation into the UN peacekeeping datasets on gender and missions has shown that socio-economic and socio-demographic variables can be used to cluster countries and identify the values that lead to different contributions. This research has highlighted the different types of countries that contribute more total personnel, and how country characteristics correlate to total contributions. The characteristics of countries that contribute higher percentages of female personnel were identified, alongside the different patterns of socio-economic variables which correlate with these percentages. Additionally, the most statistically significant characteristics for predicting total contributions made to UN peacekeeping operations were also identified through regression models.

Some key patterns observed are that countries with higher military expenditures are more likely to contribute a lower percentage of female personnel to peacekeeping operations. This is because countries with higher military expenditure also have a smaller percentage of the workforce as female. Leading to the next finding, countries with a higher percentage of females in the workforce contribute a higher percentage of female personnel. It was also identified that the percentage of female personnel contributed to UN operations has increased over the last 10 years. Variables such as urban population and refugee population are significantly correlated to country contributions, however using these variables to predict contributions appeared to be quite inaccurate and the model would need developing.

There were some limitations to this research. When selecting the socio-economic and socio-demographic variables there were some that had a lot of missing values, so a few of the initial variables chosen had to be replaced with other variables that contained more information. Also, only a few of these characteristics could be selected and investigated in detail so this restricted conclusions slightly. During the research, it was suggested that in years where there were a lot of active missions, personnel contribution could be affected because countries will be contributing to multiple missions. This may result in lower than expected contributions, and could result in a less accurate analysis. A location factor was addressed early on, however there was no measurement available to consider this in the statistical analysis. It is possible that this could be having an effect on contributions, and as it was not measured, any conclusions on the influence of this factor are limited. In future, it would be interesting to find a way to measure this so that a more comprehensive conclusion can be reached.

There were many ideas and topics that couldn't be investigated in this paper due to constraints of the dataset. The information for 2021 was unavailable and there were many missing values for lots of countries for 2020. Also, time constraints had an impact on limiting the number of topics that could be investigated in one research paper. As the research focused on how the socio-economic structure of a country correlated to the total contributions it makes, this paper didn't investigate how the characteristics of a mission and the country it resided in affects the number of contributions that it receives. This area of investigation was visually analysed for two missions (map plots in figure 5 and 6, pages 11 and 12), but investigation into this area was not carried out further as there was no numeric measurement created to allow statistical analysis.

This paper only investigated the statistically significant variables for predicting Total contributions, with further analysis, more models could be built to investigate which variables are statistically significant predictors for a country contributing higher percentages of female personnel. Also, further analysis into building models which could predict the amount of personnel contributions needed would be useful for predicting the number of personnels needed for future missions. Once more data is available for 2021 and other years, it will be important to observe the effects of the COVID-19 pandemic on contributions. On one hand there will be more demand for operations, with more countries in crisis as a result of the pandemics, however, with the possible economic downturn for key peacekeeping contributors, will countries become more self interested post-pandemic or will they still be willing to make personnel contributions to missions helping those more in need.

6 References

Crawford, K., Lebovic, J., and Macdonald, J. (2014). *Armed Forces & Society* [Online], 41, 257-281. Available from: https://journals.sagepub.com/doi/abs/10.1177/0095327x14523416?journalCode=afsa [Accessed 2nd February 2021]

Passmore, T., Shannon, M., and Hart, A. (2018). Rallying the troops: Collective action and self-interest in UN peacekeeping contributions. *Journal of Peace Research* [Online], 55, 366-379. Available from: https://journals.sagepub.com/doi/abs/10.1177/0022343317731152?journalCode=jpra [Accessed 2nd February 2021]

Rogers, J., and Kennedy, C. (2014). Dying for Peace? Fatality Trends for United Nations Peacekeeping Personnel. *International Peacekeeping* [Online], 21, 658-672. Available from: https://www.tandfonline.com/doi/full/10.1080/13533312.2014.972035 [Accessed 2nd February 2021]

UN Peacekeeping (n.d.). Troops and Police Contributions [Online]. *UN Peacekeeping*. Available from: https://peacekeeping.un.org/en/data-troop-and-police-contributions [Accessed 2nd February 2021]

UN Peacekeeping (n.d.). Women in Peacekeeping [Online]. *UN Peacekeeping*. Available from: https://peacekeeping.un.org/en/women-peacekeeping [Accessed 2nd May 2021]

Gaibulloev, K., George, J., Sandler, T., and Shimizu, H. (2015). Personnel contributions to UN and non-UN peacekeeping missions: A public goods approach. *Journal of Peace Research* [Online], 52, 727–742. Available from: https://journals.sagepub.com/doi/10.1177/0022343315579245 [Accessed 1st May 2021]

Di Salvatore, J., and Ruggeri, A. (2017). Effectiveness of Peacekeeping Operations. *Oxford Research Encyclopedias of Politics* [Online]. Available from: https://doi.org/10.1093/acrefore/9780190228637.013.586 [Accessed 1st May 2021]

Tkach, B., and Phillips, J. (2019). UN Organizational and Financial Incentives to Employ Private Military and Security Companies in Peacekeeping Operations. *International Peacekeeping* [Online], 27, 102-123. Available from: https://doi.org/10.1080/13533312.2019.1698296 [Accessed 1st May 2021]

Gabriela Dos Santos (2021). The United Nations and Its Interests: Dollars or Humanitarianism? *Academic Festival 2021* [Online], 55. Available from: https://digitalcommons.sacredheart.edu/cgi/viewcontent.cgi?article=1735&context=acadfest [Accessed 10th May 2021]

Tidblad-Lundholm, K. (2020). When are Women Deployed? Operational Uncertainty and Deployment of Female Personnel to UN Peacekeeping. *International Peacekeeping*

[Online], 27(4), 673–702. Available at: https://www.tandfonline.com/doi/full/10.1080/13533312.2020.1760717 [Accessed 19th May 2021].

Faraway, J. (2006). *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models.* Florida: Taylor & Francis Group, LLC. pg. 55.

Moral, R., Hinde, J., and Demétrio, C. (2017). Half-Normal Plots and Overdispersed Models in R: The hnp package. *Journal of Statistical Software*, 81(10), 10. DOI: 10.18637/jss.v081.i10.

Gaibulloev, K., Sandler, T., and Shimizu, H. (2009). Demands for UN and Non-UN Peacekeeping: Nonvoluntary versus Voluntary Contributions to a Public Good. *Journal of Conflict Resolution*, 53(6), 827-852. Available from: https://journals.sagepub.com/doi/10.1177/0022002709338509.

Jabeen, H. (2019). Tutorial: Poisson Regression in R [Online]. *Dataquest.* Available from: https://www.dataquest.io/blog/tutorial-poisson-regression-in-r/. [Accessed 10th May 2021].

UCLA (n.d.). Negative Binomial Regression: R Data Analysis Examples [Online]. *UCLA: Statistical Consulting Group.* Available from: https://stats.idre.ucla.edu/r/dae/negative-binomial-regression/. [Accessed 10th May 2021].

Zach (2019). A Gentle Introduction to Poisson Regression for Count Data [Online]. *Statology.* Available from: https://www.statology.org/poisson-regression/. [Accessed 10th May 2021].

The World Bank (n.d.). Databank: World Development Indicators [Online]. *The World Bank.* Available from: https://databank.worldbank.org/source/world-development-indicators. [Accessed 10th May 2021].

United Nations Statistics Division (n.d.). Methodology: Standard country or area codes for statistical use (M49) [Online]. *United Nations Statistics Division.* Available from: https://unstats.un.org/unsd/methodology/m49/. [Accessed 10th May 2021].

GOV.UK (n.d.). ISO Country Codes [Online]. *GOV.UK.* Available from: https://www.gov.uk/government/publications/iso-country-codes--2. [Accessed 10th May 2021].

Coleman, K., and Nyblade, B. (2018). Peacekeeping for profit? The scope and limits of 'mercenary' UN peacekeeping. *Journal of Peace Research*, 55(6), 726-741. Available from: https://doi.org/10.1177/0022343318775784.

Boutton, A., and D'Orazio, V. (2020). Buying blue helmets: The role of foreign aid in the construction of UN peacekeeping missions. *Journal of Peace Research*, 57(2), 312–328. Available from: https://doi.org/10.1177/0022343319865929.

A Appendices

A.1 - Table of average results for cluster analysis stage 1, cluster 3, 2010.

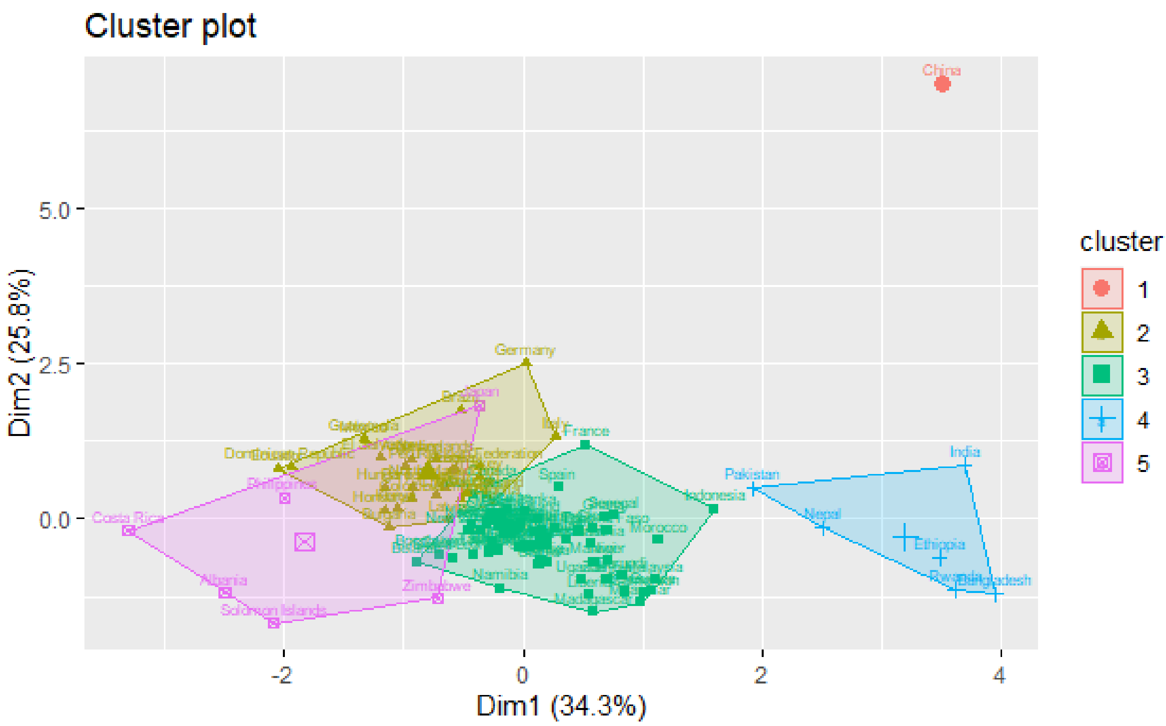
Cluster	Total	Refugee_Population	Urban_Population	Education
1	878.4000	2805.00	119222016	9.200000
2	422.8214	42381.00	6601322	6.785714
3	181.8065	21024.15	6680250	9.580645
4	7226.0000	2146979.00	34522278	11.000000
5	719.3750	47764.04	16510391	12.500000
6	7158.0000	109268.00	39467064	6.666667

High Average Low

A.2 - Table of average results for cluster analysis stage 2, group 1, 2019.

Cluster <dbl>	Percent_Female <dbl>	GDP <dbl>	Total <dbl>	Education <dbl>
1	0.0297300	1.430000e+13	2489.00000	9.000000
2	0.1322072	5.640118e+11	206.79310	12.862069
3	0.0975547	2.322044e+11	505.30303	8.984848
4	0.0421650	5.979847e+11	6183.33333	8.166667
5	0.7555550	9.261852e+11	21.16667	9.666667

A.3 - Cluster plot of countries in stage 2, group 1, 2019.



A.4 - Table of average cluster values of countries in stage 2, group 2, 2019.

Cluster <dbl>	Percent_Female <dbl>	Refugee_Population <dbl>	Armed_Forces <dbl>	Military_Expenditure <dbl>
1	0.06555333	180298.33	1515222.3	132409183372
2	0.05873800	2094269.20	407770.0	16531686802
3	0.10707436	70807.66	103479.1	4320829620
4	0.75555500	3154.50	107166.7	8687522625