An Exploratory Analysis of Refugee Statistics

Bentley Schieckoff and Carlo Didonè

2017-12-14

Introduction

In the past few years, migration has been front-and-centre in international politics. Highlighted by the European refugee crisis in 2015, people are fleeing their homelands now more than ever in search of a better life. Using data from the United Nations Refugee Agency (otherwise known as the UNHCR), we will examine global refugee flows over the past few decades. Furthermore, we will explore the relationship between UN assistance for asylum applicants and the number of minor refugee applicants. The paper will have the following structure:

I. Data Cleaning
II. Exploratory Data Analysis
III. Empirical Analysis
IV. Discussion and Conclusion

I. Data Cleaning

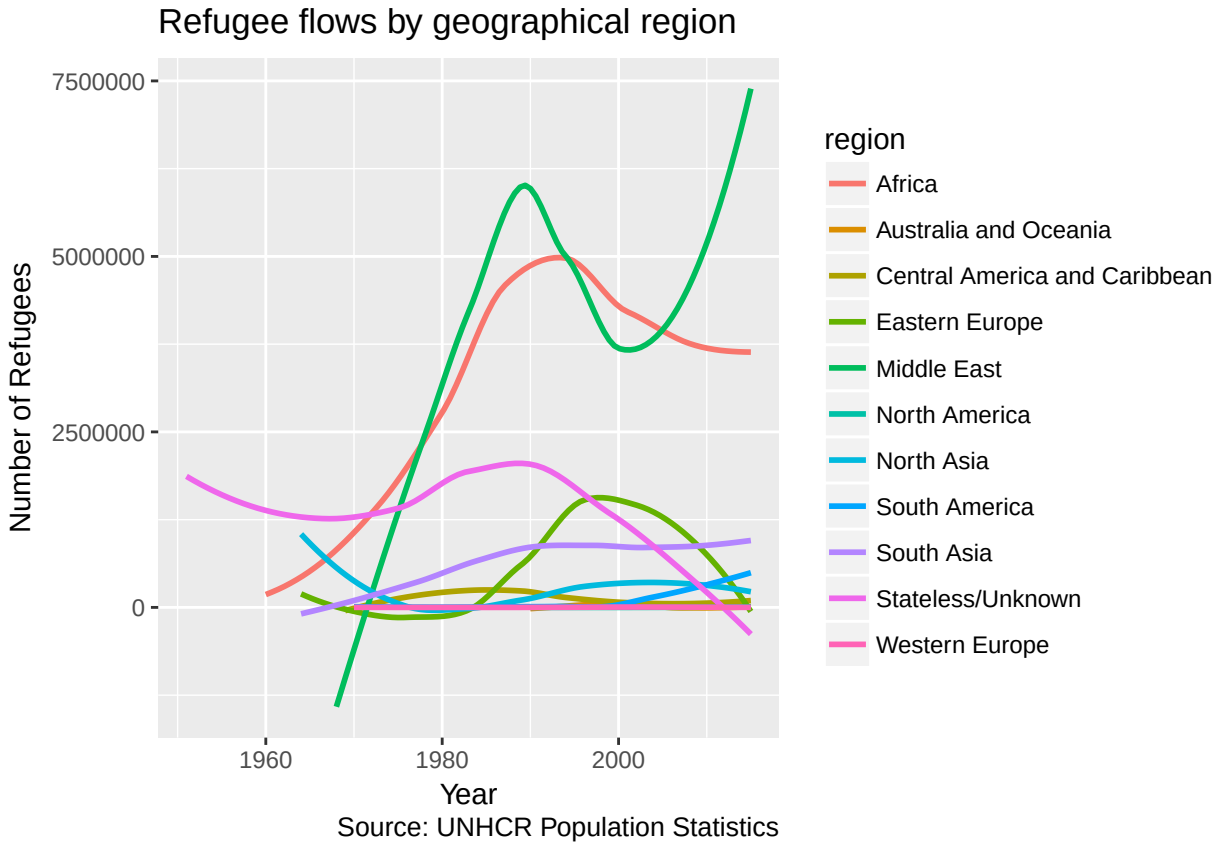
The data provided by the UNHCR can be freely downloaded from their website in comma separated variable form. The tables we chose to work with are: Asylum-Seekers, Time Series and Demographics. This data was not immediately ready to be worked with, since often the "class" of the data did not match the type of data and the variable names are inconvenient to work with. After loading these tables into R, our first task was to rename the variables, drop any unnecessary columns, and coerce the data into "numeric", "date" and "character" classes as appropriate.

One variable that was difficult to deal with was in the Demographics table. This data was broken down into different age groups, for instance 5-17 years old. Some countries entered their data under this variable, while others had it split into sub-groups (5-11 and 12-17). To solve this problem, we had to create a new variable that summed the observations across these three columns. This created a common measure, since countries either entered their data in the 5-17 column or split it into the sub-groups.

II. Exploratory Data Analysis

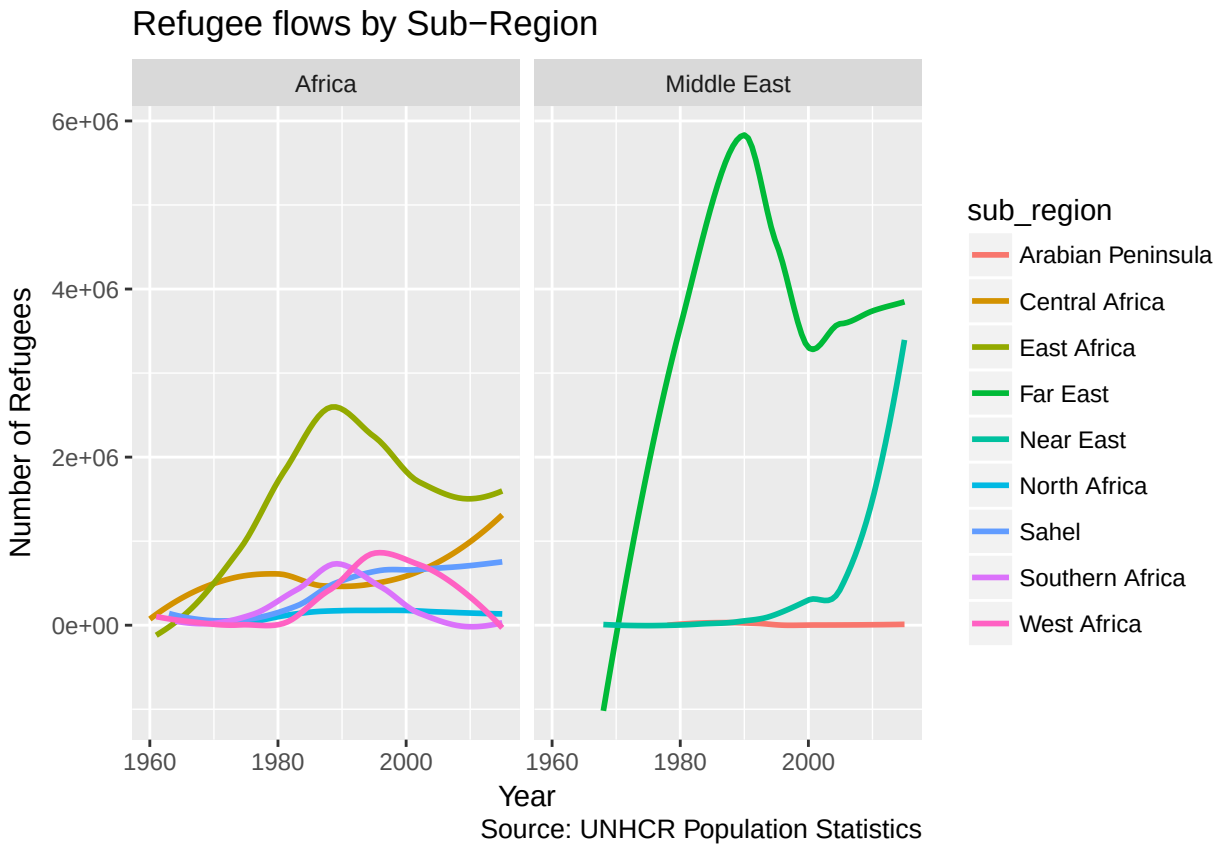
Before conducting any empirical analysis, we wanted to examine the general trends in refugee flows. The Time Series dataset from UNHCR contains information about a variety of persons seeking international protection, including refugees, asylum seekers, internally-displaced persons and so on, reaching back to 1951. This information is displayed by year, country of current residence and country of origin.

To identify where refugee flows are originating, we filtered this data for just statistics on the number of refugees, and generated a geographic "region" variable to sort each country of origin into. We then grouped the refugee statistics by the region of origin and year. The figure below shows how the number of refugees from each global region has changed over time.



This graph shows that many regions have not, in fact, been the source of a high number of refugees in the past few decades. As one might expect, these regions are North America, Western Europe, Australia and Oceania, but also Central America and the Caribbean, North Asia (the post-Soviet states, China, Japan and the Koreas) and South America. You can see that the number of refugees from Eastern Europe climbed during the 1990's, when the region was experiencing military conflict and political turbulence, but it steadily decreases from 1999 on. The two regions sending (by far) the highest number of refugees are Africa and the Middle East. These two regions started with similar patters, both increasing steeply from 1960 to 1990 and dropping off thereafter. Around 2001, the number of refugees from the Middle East begins to radically increase again, no doubt due to the military conflicts and political turbulance that started after September 11, 2001 and has continued to this day.

We wanted to look deeper in the data to focus specifically on these two areas with higher refugee flows. To accomplish this, we sorted countries into sub-regions and grouped the data at this level. The graph below is similar to the first, showing the number of refugees per sub-region, from 1960 until 2016.



We can see that very few refugees come from North Africa during this period. Refugees from Southern and West Africa follow similar trends, rising to a peak during the 1980's and 1990's (respectively), then dropping back to a negligible number. Refugees from the Sahel region and from Central Africa both show a general upwards trend since 1960. The most prominent sub-region in Africa, with by far the highest number of refugees is East Africa. This region includes countries like Somolia, Eritrea, Ethiopia, Rwanda and Uganda, all countries that have struggled with military conflicts, political turbulence and dire living standards throughout the past few decades.

This being said, the overall numbers coming from sub-regions in Africa are quite frankly dwarfed by the number coming from the Middle East. Those coming from the Far East (Afghanistan, Pakistan, Iraq and Iran) are the most numerous of all. In recent years, the number coming from the Near East (Palestine, Syia, Turkey, Lebanon etc.) have spiked sharply. Very few refugees come from the countries of the Arabian Peninsula.

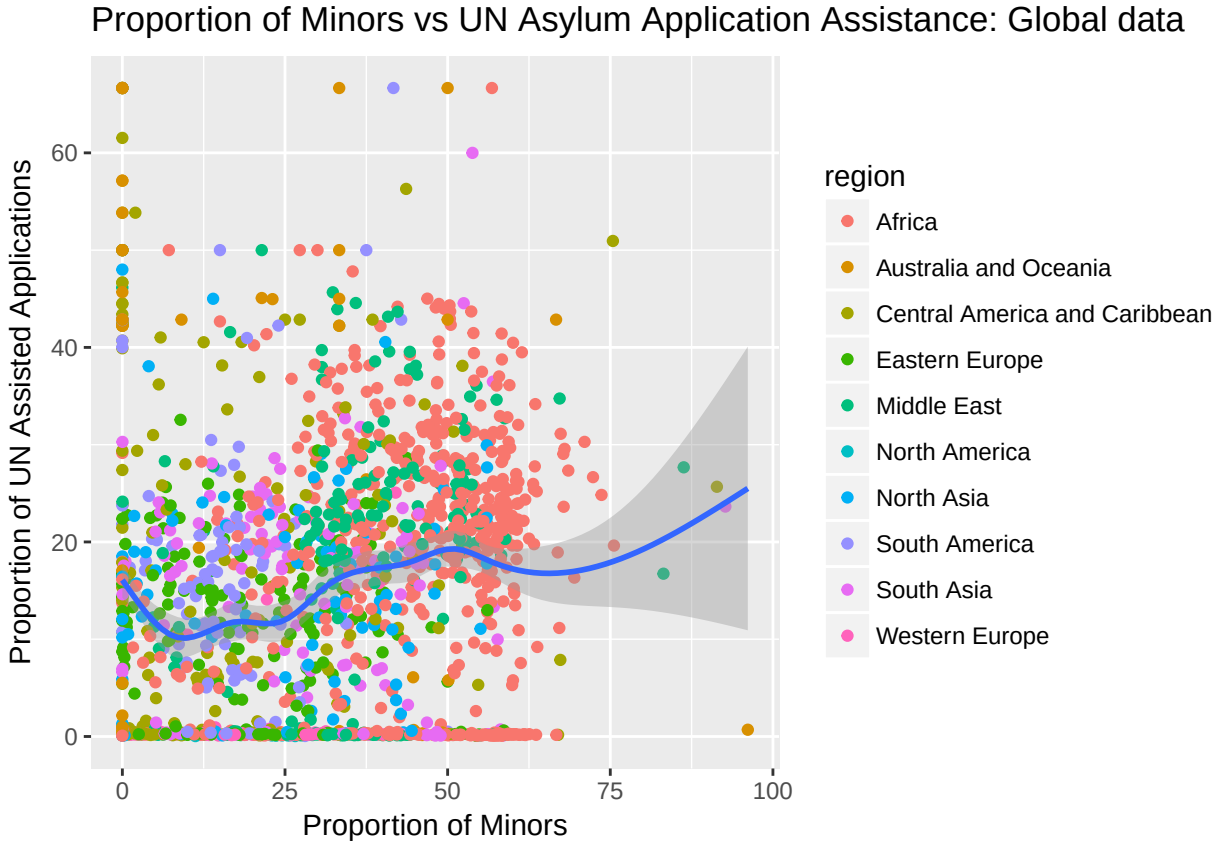
Overall, the global data shows that many refugees are coming from the Middle East and Africa. The question is, where are these asylum seekers going?

III. Empirical Analysis

Our empirical analysis revolves around two assuptions. Firstly, we think that families are most likely to stick close to their country of origin. Long travels may be harder to face for minors. Therefore, minors should compose a large proportion of the refugee populations near areas with a high refugee flow. Moreover, if minors are concentrated in particular areas of the world, we think that the UNHCR would be more active in those areas to assist with applications, since children and teenagers are not as autonomous as adults and families with young children are priorities in the asylum system. Thus, we expect to see a high proportion of minor asylum-seekers in areas close to high-outflow regions of the globe, and to find a positive relation between the proportion of minor asylum seekers in a region and UN assisted applications.

The Asylum Seekers table contains data on the decisions made on asylum applications, based on the country of arrival, of origin and the year. We also have demographic data, which might play an important role here. We therefore merged these two tables, using country of residence and year as keys for the merge. For the purpose of creating an interactive map, we added country codes to the dataset. The map (which can be viewed in the HTML version of the document) shows the proportion of total refugees in each country that are minors.

The map supports our first hypothesis: a higher proportion of minors are located within refugee populations in Africa and the Middle East, areas where most migrants are coming from. Over the years, countries with a consistently high population of minor asylum seekers are Afghanistan, Pakistan, Chad, Somalia and Ethiopia. With relation to our second hypothesis, we wanted to see if this was reflected in a more active role of the UNHCR in assisting with applications. We first explored the data using a scattered plot.



The trend seems to show a positive correlation between these two variables, even if its explanatory power may not be strong. We then ran a regression to check this hypothesis, regressing the proportion of minors in the country on the proportion of UN assisted applications in the same region.

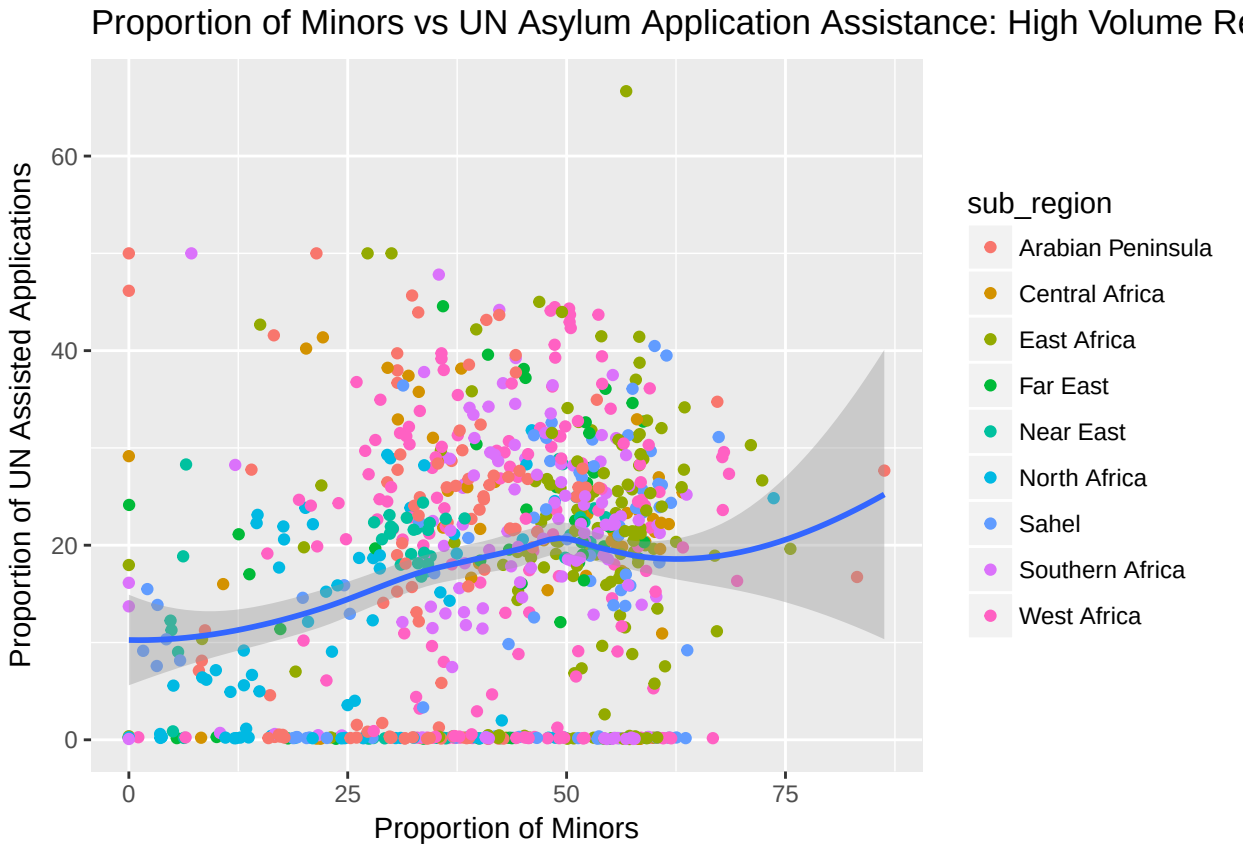
```
##
## Call:
## lm(formula = assisted_pro ~ minor_pro, data = full_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -21.221 -12.484  -0.311   7.854  54.294
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.37316    0.62228  19.884  < 2e-16 ***
```

```
## minor_pro    0.09925    0.01730   5.739 1.13e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.01 on 1659 degrees of freedom
##   (605 observations deleted due to missingness)
## Multiple R-squared:  0.01946,    Adjusted R-squared:  0.01887
## F-statistic: 32.93 on 1 and 1659 DF,  p-value: 1.132e-08
```

The regression results show that the proportion of minor refugees indeed has a positive, statistically significant influence on the proportion of applications assisted by the UN (though the effect size is small).

However, we believed that this effect would be even more pronounced in Africa and the Middle East, areas where we have seen the proportion of minors seeking asylum to be highest. What emerges from the scattered plot above is that these two regions are also the ones with the highest support in applications from the UN. Therefore we ran the same analysis using only observations coming from the Middle East and Africa.

Dividing the observations into two geographic regions helps us also with outliers. Statistical analysis performed on the whole world may be too general, not accounting for the diversified nature of geographical and political structures.



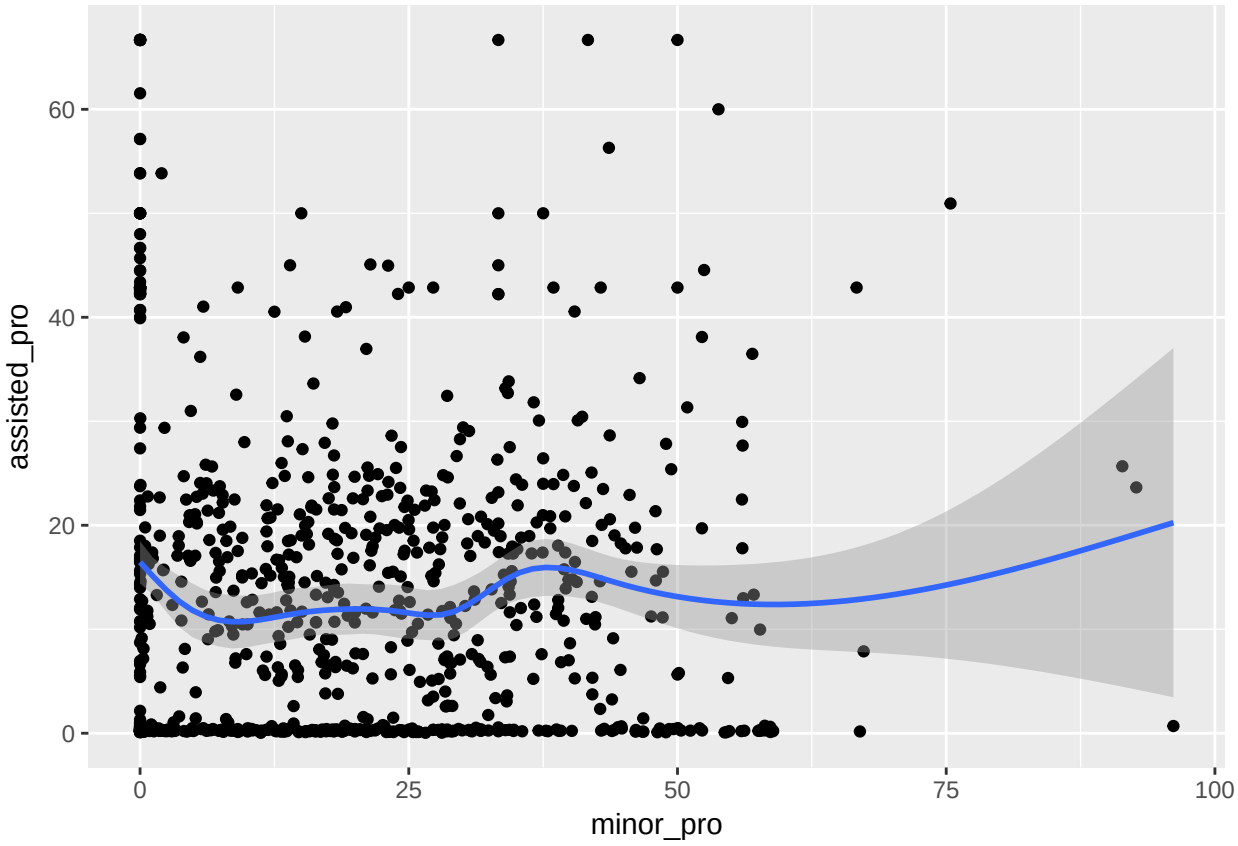
When limiting our data to high-volume regions, the correlation between the proportion of minor refugees and the proportion of UN assisted applications seems to be stronger. In the scattered plot, two regions are noticeable: Northern Africa has a low number of refugees and low support, the opposite can be said for Southern Africa. Outliers do not seem to be an issue. Here are the results of the regression:

```
##
## Call:
## lm(formula = assisted_pro ~ minor_pro, data = high_volume_refugees)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -21.296 -12.588   1.564   8.721  46.767
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.94272    1.27896   8.556  < 2e-16 ***
## minor_pro    0.15767    0.02875   5.485 5.57e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.88 on 793 degrees of freedom
##   (72 observations deleted due to missingness)
## Multiple R-squared:  0.03655,    Adjusted R-squared:  0.03533
## F-statistic: 30.08 on 1 and 793 DF,  p-value: 5.575e-08
```

This second regression produces a significant estimate, whose value is slightly higher than the previous estimate on the global data.

We now checked if this correlation is similar for low-volume regions of the world.




```
##
## Call:
## lm(formula = assisted_pro ~ minor_pro, data = low_volume_refugees)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -13.623 -12.958  -2.750   6.887  53.811
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13.71267    0.75989  18.046   <2e-16 ***
## minor_pro   -0.01715    0.02944  -0.582     0.56
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.78 on 864 degrees of freedom
##   (533 observations deleted due to missingness)
## Multiple R-squared:  0.0003924,  Adjusted R-squared:  -0.0007646
## F-statistic: 0.3392 on 1 and 864 DF,  p-value: 0.5605
```

We can see from the scattered plot that the correlation is not definite. The estimate in the regression confirm this hypothesis, which is negative but not statistically significant.

IV. Discussion and Conclusion

Our exploratory analysis and regression find that the UN is more active in countries where minors make up a high proportion of the migrant population. This effect is particularly true in the Middle East and Africa, where minors are present in high proportions.

One caveat to our analysis is that scattered plots show some heteroskedasticity. We need to consider the possibility of our estimates being slightly biased. No doubt, our models are not comlete, omitting many variables (not included in these datasets) which would impact the proportion of applications in an area that the UN assists with. Controlling for more variables would improve our R-squared and give more precise estamiates.

Moreover, the observations coming from countries with almost no assistance from the UN, or a low proportion of minors, may be a problem to our analysis. We decided to include them nevertheless, to not alter the dataset excessively.