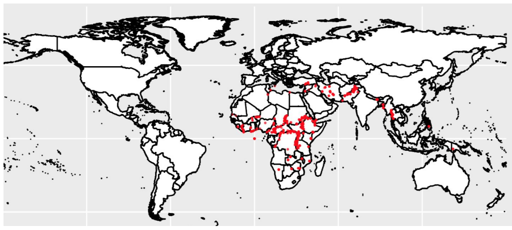
# Accessing Data on Mega-Camps:

## A Policy Recommendation

*Jacquelyn Sieck*



*Red Dots signify the Refugee Camps included in the Spatial Data Analysis*

### Introduction

International bodies have sought to alter refugee camps in search of more humane practices, but fail to account for the rushed nature of camp creation. These attempts to set guidelines are no doubt in response to the protracted conflicts from which refugees flee; a recent World Bank report states that the average length for displacement is 26 years.[1] The following analysis will be formed upon these three policies, and will use data provided by Dr. Kara Ross-Camarena of the University of Chicago to recommend policy to improve existing camps.

### Existing Policy

In response to protracted conflicts, the United Nations High Commissioner for Refugees has laid out multiple policy guidelines on refugee camps. Firstly, the UNHCR *Handbook on Emergencies* notes that "high density camps with very large populations are the worst possible option for refugee accommodation... large camps of over 20,000 people should be avoided." [2] This sets guidelines to try to limit the population of the camps, and was influenced by former UN Secretary General Kofi Annan, who stated that he "strongly urge[s] that refugees be settled at a reasonable distance from the border, in camps of limited size."[3] In my data analysis on violent acts occurring near refugee camps in Pakistan, only 25% of incidents occurred near camps within 50 km of the border. This, in terms of refugee locations in Pakistan, is a success of the aforementioned policy on international borders.

Additionally, UNHCR created an urban refugee policy in 2009 which attempted to ensure that "cities are recognised as legitimate places for refugees to reside and exercise their rights," and responded to the changing nature of refugee camps: half of all refugees now live in cities.[4] However, while some refugees may move into cities, millions of refugees remain in camps and policies must continue to alter and create humane policies. Protracted refugee situations force migrants to remain displaced, and we must create camps which offer both protection and livelihoods for those displaced; post-conflict transformation will be nearly impossible for these populations if they remain safe while they are displaced but do not know how to navigate good governance, taxes, and education.

### Data Analysis

In order to analyze the effectiveness of these policy recommendations, I have used data provided by Dr. Camarena to examine the changing demographics of camps between 2010 and 2018. This data includes both well-established camps formed in 1966 as well as camps created as recently as 2017. This data is not complete, most likely due to constraints by international organizations and the unwillingness of data sharing by host governments. This means that our analysis is bound by the biases and reporting mechanisms through which data on vulnerable populations is shared.

The data which Dr. Camarena provided included 3,035 data points, and each one corresponds to a refugee camp in a specific year. This means that some of the data is on the same refugee camp in a different Fiscal Year. Using just the data just

---

[1] United Nations, Department of Economic and Social Affairs, *Contribution To The Fifteenth Coordination Meeting On International Migration,* UN/POP/MIG-15CM/2017/14 (10 February 2017)

[2] Jeff Crisp, and Karen Jacobsen. 1998. "Refugee Camps Reconsidered" 3 (3): 28

[3] Ibid

[4] UN High Commissioner for Refugees (UNHCR), UNHCR Policy on Alternatives to Camps, 22 July 2014, UNHCR/HCP/2014/9

for 2010, there are 260 camp observations in 42 countries. In 2010, Pakistan had the highest number of camps (74), Afghanistan had the second highest (22), and Sudan had the third highest (20). The population of refugees in 2010 was 4,635,380, with the top host countries being Sudan (1,113,600) and Pakistan (857,735). The UNHCR Global Trends report for 2010 notes that Pakistan, Iran, and Syria hosted the highest refugees. [5] This means that the data which we have been provided is at least in part similar to global refugee trends, since our data includes Pakistan as the country with the most camps and second highest total population.[6] In 2018, however, there were 322 camp observations across 38 countries. The data identifies the top host countries in 2018 as Pakistan (48), Sudan (40), Bangladesh (35), and Ethiopia (35). For camp population in 2018, there was a total of 14,462,543 refugees. These refugees were predominantly in Sudan (2,962,426), Somalia (2,648,000), and Pakistan (2,513,553). In UNHCR's 2018 Global Report, they identify the top host countries as Turkey, Pakistan, Uganda, Sudan, and Germany.[7] This means that our data is also fairly representative for 2018, since we have significant data on refugees in those states.

In terms of this data specifically, the refugee camps are being more evenly dispersed over multiple countries. This is a positive change in comparison to having Pakistan represent nearly thirty percent of refugee camps in 2010. However, since there are fewer camps and more total population, this means that camps are getter larger. This enlargement of the camps contradicts the aforementioned international policies, but speak to the fact that refugee camps are not planned months in advance; they are in direct response to international crises. For most camps, comprehensive planning is not an option; refugees do not flee once a settlement is created, but rather when violence arrives at their door. An example of this was the 2017 mass exodus of Rohingya refugees into Bangladesh in 2017, when nearly 700,000 people were displaced in the span of months.[8] The 2018 UNHCR Global Trends report quotes a man who fled, stating that while "the living conditions in the camps are difficult... the Rohingya feel safe."[9] This means that while size of camps have been emphasized in policy, these large camps continue to exist and to offer safety to those fleeing violence.

**Recommendation**

I propose that instead of focusing on the location of refugee camps in relation to the border, we should work on existing refugee camp reform and create quasi-cities in response to the mass exodus of individuals from their countries of origin. As aforementioned, camps that were formed in 1966 are still in use today. All refugee camps are forced to have underlying structures which allow informal channels of business because of the protracted nature of conflict. I recommend that in response to the protracted nature of conflict, we must focus less on the size of the camps, and rather form small cities with a local government. By creating local institutions, host governments can collect taxes and build infrastructure in the camps that recognize that displacement cannot be treated as temporary. Refugees can stay displaced for up to 26 years, and during this time it is their right to access all the institutions they could have accessed in their country of origin. This means teaching refugees how to navigate good governance, the importance of taxes to improve infrastructure, and education. The creation of these cities will create a foundation on which states can be rebuilt. By offering refugees these most basic necessities, we are cultivating a smooth post-conflict transition for the countries where the refugees will voluntarily repatriate.

[5] "Global Trends 2010". *UNHCR,* 2011. Web, accessed 18 July 2019.

[6] The reason that Sudan is so high in our count is because the camps there are filled predominantly with Internally Displaced Persons instead of refugees, which means that they have a smaller number of refugees as counted by the UNHCR, but a large number of persons of concern.

[7] Global Trends 2018". *UNHCR,* 2019. Web, accessed 18 July 2019.

[8] Global Trends 2018". *UNHCR,* 2019.

[9] Global Trends 2018". *UNHCR,* 2019.

# RefugeesCapstone

## Section 1

```r
camps <- read_csv("dpss-camp.csv")

camps <- camps %>%
  mutate(boy = rsct_5_11_m ,
         girl = rsct_5_11_f ,
         littleb = rsct_0_4_m,
         littleg = rsct_0_4_f ) %>%
  mutate(child = boy + girl + littleb + littleg)

camp <- camps %>%
  mutate(porpkid = child/pop_camp) %>%
  select(year, porpkid)
```

```r
for(i in 2010:2018){
  ggplot (data= subset(camp, year == i), aes(x= porpkid)) +
    geom_histogram(color = "white" , fill = "black" , binwidth = .01) +
    xlab("Proportion of Children Under 12 in Refugee Camps") +
    ylab("Frequency") +
    labs(title = paste("Proportion of Children under 12 in Refugee Camps in ", i))
}
```

## Analysis of the Histograms

These histograms are very different, taking out the constant high number of camps that have 0% children. Most notably, 2015 has incredibly low frequency for the proportion of children in refugee camps. This low frequency stays fairly constant between 2016 and 2018. This is either due to the changing UNHCR policy in 2014 which stated a preference for "alternatives to camps," which may have affected either migration patterns to camps, or forced existing families in the camps to relocate. I think to further explore this, we should compare the camp porportions across all years for all camps. I think it would be interesting to see if when the children become a smaller proportion of the population, if it is the women or men whose numbers increase. Moreover, researchers should examine if there were any changing mechanisms for data collection which may have affected or skewed the number of children counted - i.e., a policy not allowing for the release of any statistics on a child under the age of 5.

## Section 2

```r
campmean <- camp %>%
  group_by(year) %>%
  filter(year == 2010 | year == 2018)

temp <- camps %>%
  mutate(porpkid = child/pop_camp) %>%
```

```r
  select(year, porpkid)


ten <- temp %>%
  filter(year == "2010") %>%
  pull(porpkid)

eighteen <- temp %>%
  filter(year == "2018")  %>%
  pull(porpkid)

t.test(ten, eighteen, mu = 0, conf.level = .95)
```

```
##
##  Welch Two Sample t-test
##
## data:  ten and eighteen
## t = 7.9765, df = 566.32, p-value = 8.374e-15
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.08091651 0.13378513
## sample estimates:
## mean of x mean of y
## 0.3428918 0.2355410
```

This test shows us that the mean proportion of children under the age of 12 in 2010 across all camps was .3428918, while the mean mean proportion of children under the age of 12 in 2018 was 0.2355410. We have a very high degrees of freedom (566.32), an incredibly small p value, and a 95% confidence interval which does not cross zero which means that we can accept our alternative hypothesis that the difference in means that we see between 2010 and 2018 are statistically significant and not random.

```r
men <- camps %>%
  mutate(policychange = ifelse(camps$year > "2013" , "1" , "0")) %>%
  mutate(origins = ifelse(camps$year_o > "2013" , "1" , "0" ))

men <- men %>%
   filter(!is.na(year_o))

men <- men %>%
  mutate(propmen = men$rsct_18_59_m/pop_camp) %>%
  select(year, year_o, pop_camp, rsct_18_59_m, policychange , origins, propmen, nameplaceid, iso , type


regress <- lm(propmen ~ policychange + origins + policychange * origins, data=men)

stargazer(regress, type = 'latex', title = "Effect of the Proportion of Prime Aged Male Refugees" , dig
```

## Analysis of regressions

In this regression, our dependent variable is the proportion of prime aged men. We are regressing this on whether or not the data was collected before or after the policy “Alternatives to Camps” was in place,

Table 1: Effect of the Proportion of Prime Aged Male Refugees

| | *Dependent variable:* |
|---|---|
| | propmen |
| policychange1 | $-0.1^{***}$ |
| | (0.01) |
| origins1 | $-0.03$ |
| | (0.02) |
| policychange1:origins1 | 0.01 |
| | (0.02) |
| Constant | $0.2^{***}$ |
| | (0.004) |
| Observations | 1,337 |
| $R^2$ | 0.1 |
| Adjusted $R^2$ | 0.1 |
| Residual Std. Error | 0.1 (df = 1333) |
| F Statistic | $59.2^{***}$ (df = 3; 1333) |
| *Note:* | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |

whether or not the camp was created after the aforementioned policy, and an interaction between both the policy change and year the camp was created. I did this interaction term because I believe that since both the year the camp was created and the policy change implementation are in the same year, they are inherently correlated and interacting the two will give us a better understanding of the true effect of the policy on the proportion of men.

Moving forward, I think I would like to run a regression divided by each host country, and find out the proportion change of men in each host country and see if there are any country specific changes. I think comparing this to the surrounding countries or the conflicts from which the refugees flee will allow us to see if a specific type of violence or war forces more specific groups to flee (if we look at data before the change) , and to see whether the host country has any specific policies that influence a more substantial change in the proportion of men before and after the policy is implemented. If there are only specific countries whose male proportion is lowered, then we should examine how that host country is implementing the policy in relation to a country whose proportion of men is not altered.

```
countrymen <- men %>%
  group_by(iso)


prop <- lm(propmen ~ policychange + origins + policychange * origins + iso, data= countrymen)

stargazer(prop, type = 'latex', title = "Thefts" , digits = 1, header = FALSE)
```

With this information, I will want to examine the ways in which Ghana, Mozambique, Nepal, Chad, Turkey, and Tanzania since these have the lowest p-values.

```
typemen <- men %>%
  group_by(type)
```

Table 2: Thefts

| | *Dependent variable:* |
|---|---|
| | propmen |
| policychange1 | $-0.1^{***}$ |
| | (0.01) |
| origins1 | $-0.03$ |
| | (0.02) |
| isoBEN | $0.1^{**}$ |
| | (0.1) |
| isoBFA | 0.01 |
| | (0.03) |
| isoBGD | 0.03 |
| | (0.04) |
| isoBWA | $0.2^{***}$ |
| | (0.04) |
| isoCAF | 0.1 |
| | (0.03) |
| isoCMR | 0.02 |
| | (0.02) |
| isoCOD | $-0.02$ |
| | (0.02) |
| isoDJI | 0.02 |
| | (0.03) |
| isoETH | $-0.01$ |
| | (0.02) |
| isoGHA | $0.1^{***}$ |
| | (0.02) |
| isoGIN | 0.03 |
| | (0.04) |
| isoIRN | 0.01 |
| | (0.02) |
| isoIRQ | $-0.02$ |
| | (0.1) |
| isoKEN | $-0.03$ |
| | (0.02) |
| isoLBR | 0.02 |
| | (0.03) |
| isoMOZ | $0.2^{***}$ |
| | (0.04) |
| isoMRT | 0.02 |

```r
typeofcamp <- lm(propmen ~ policychange + origins + policychange * origins + type, data= typemen)

stargazer(typeofcamp, type = 'latex', title = "Thefts" , digits = 1, header = FALSE)
```

Table 3: Thefts

| | *Dependent variable:* |
|---|---|
| | propmen |
| policychange1 | $-0.1^{***}$ |
| | (0.01) |
| origins1 | $-0.03$ |
| | (0.02) |
| typeU | $-0.1^{**}$ |
| | (0.02) |
| typeV | 0.02 |
| | (0.01) |
| policychange1:origins1 | 0.003 |
| | (0.02) |
| Constant | $0.2^{***}$ |
| | (0.005) |
| Observations | 1,337 |
| $R^2$ | 0.1 |
| Adjusted $R^2$ | 0.1 |
| Residual Std. Error | 0.1 (df = 1331) |
| F Statistic | $36.8^{***}$ (df = 5; 1331) |
| *Note:* | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |

In this I decided to look at the type of camp (Urban, Rural, or Various). In this, urban refugees are the only one with a statistical change.

## Section 3

```r
borders <- st_read("ne_10m_admin_0_sovereignty")

glimpse(borders)

spatialcamp <- read_csv("dpss-camp.csv")
```

```r
spatialcamp <- spatialcamp %>%
  filter(!is.na(x_lon))
```

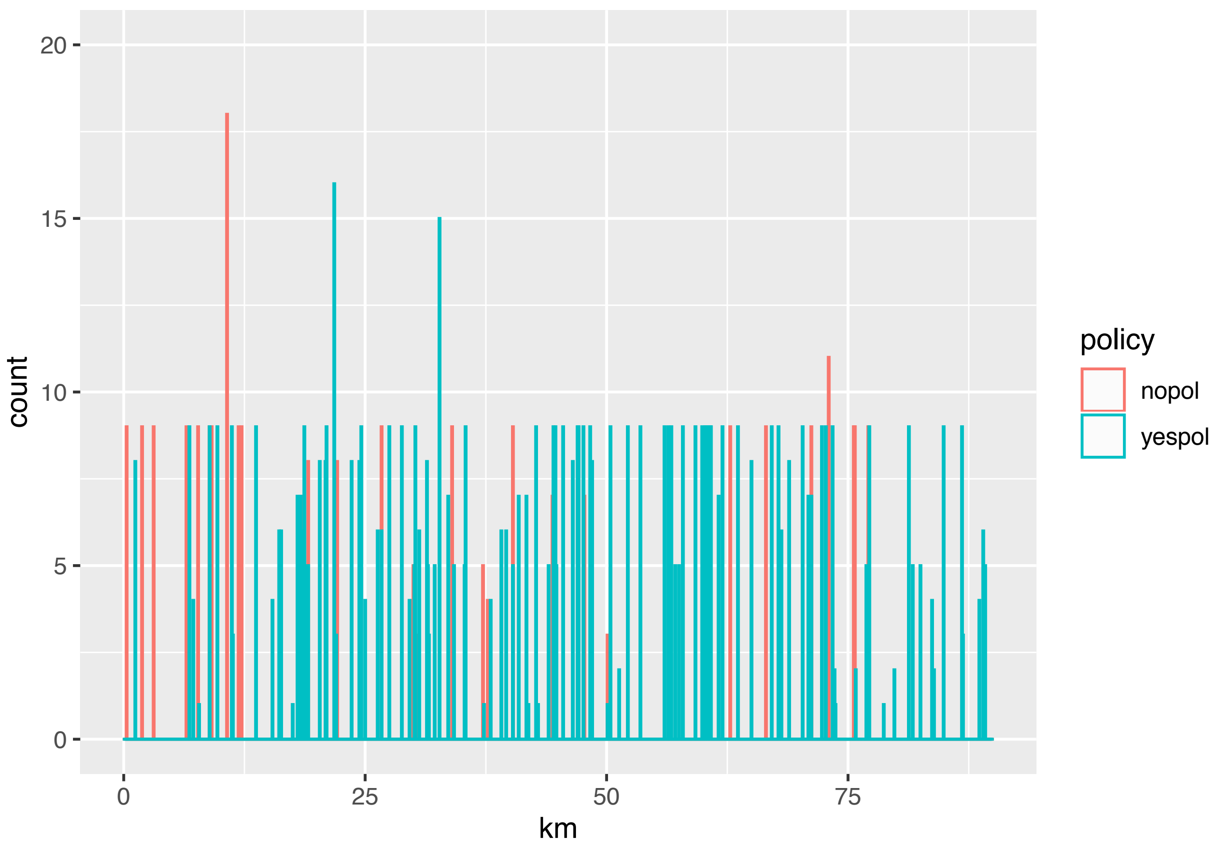
```r
refugeecamps <- st_as_sf(
  spatialcamp,
  coords = c("x_lon", "y_lat") ,
  crs = 4326)


borders$SOVEREIGNT <- as.character(borders$SOVEREIGNT)
```

```r
creation <- read_csv("creation.csv")
```

#This is the code I made to create "creation" –> and then I saved it as a csv.

creation <- tibble () for (i in 1:nrow(refugeecamps)) { intl = refugeecamps[i, ] temp <- borders %>% filter(SOV_A3 != intl$iso)dist < -st_distance(temp, intl)print(temp[which.min(dist), ]$SOVEREIGNT) (min(dist)) print(i) creation <- bind_rows(creation, c(close = (temp[which.min (dist),]$SOVEREIGNT), dista = (min(dist)), nameid = intl$nameplaceid , year = intl$year, yearo = intl$year_o , iso = intl$iso, pop = intl$pop_camp)) }

```r
creation$dista <- as.numeric(creation$dista)

creation <- creation %>%
  mutate(km = dista/1000)

policycreation <- creation %>%
  mutate(policy = ifelse(creation$yearo >= "1999" , "yespol" , "nopol")) %>%
  filter(!is.na(yearo)) %>%
  select(km , policy)

policycount <- policycreation %>%
  count(policy)

ggplot(policycreation , aes(x = km , color = policy)) +
  geom_histogram(fill = "white" , alpha = 0.7, position = "identity" , binwidth = .1) +
  scale_x_continuous( limits = c(0, 90)) +
  scale_y_continuous( limits = c(0, 20))
```




#Analysis of Policy GGPlot

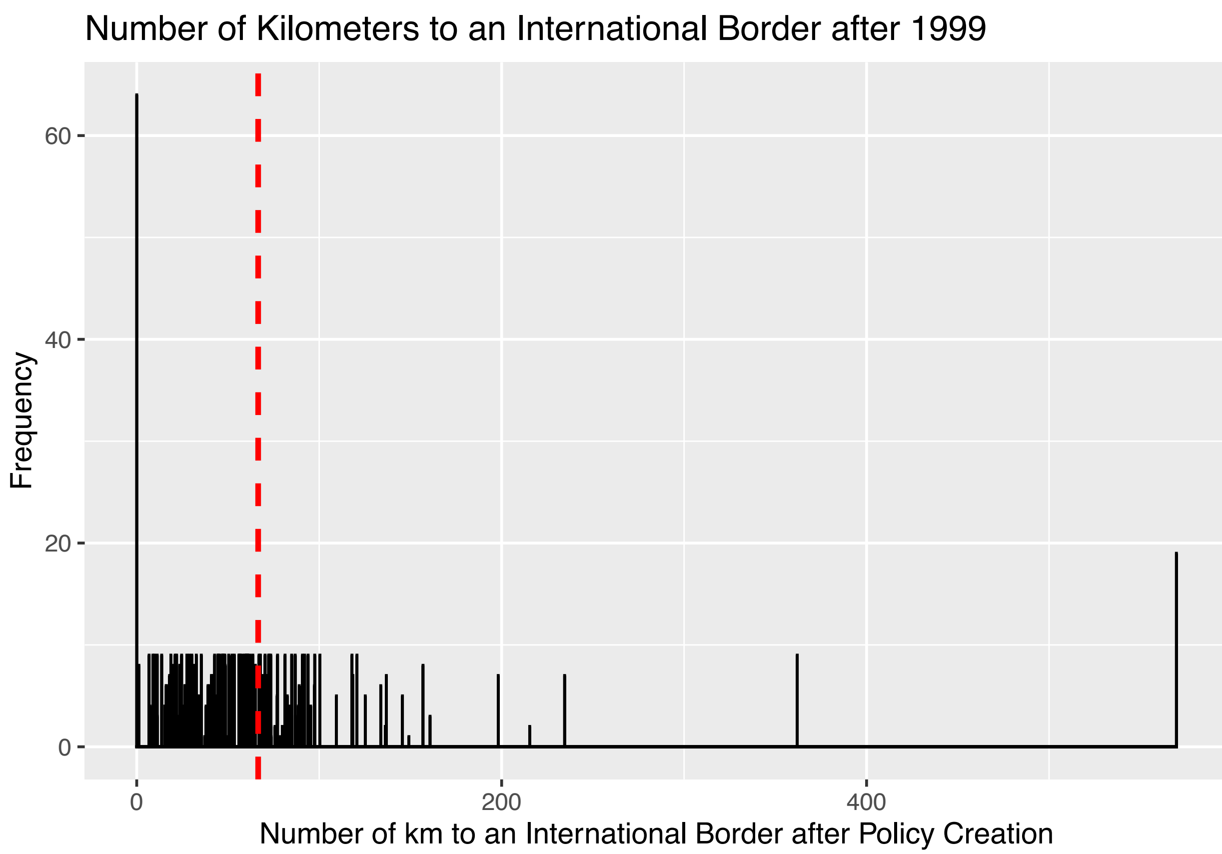
I didn't need to make this graph, but I think it helped put the data into perspective because it demonstrates how much more data we have on the post-policy period in comparision to the pre-policy period.

First of all, there are many closer to the border refugee camps since the creation of this policy. You can see this in the histogram that I created that had both the pre and post policy numbers. However, in the data that we have, we only have 321 refugee camps that were created before the policy was created (pre 1999), and 1016 refugee camps that were created after the policy was created (post 1999). Therefore, I don;t think a valid argument can be made on whether or not refugee camps are closer or further away from the border because we do not have an accurate sample size that would have a more balanced comparision.

```
polyes <- policycreation %>%
  filter(policy == "yespol") %>%
  select(km)

mean(polyes$km)
```

```
## [1] 66.55918
```

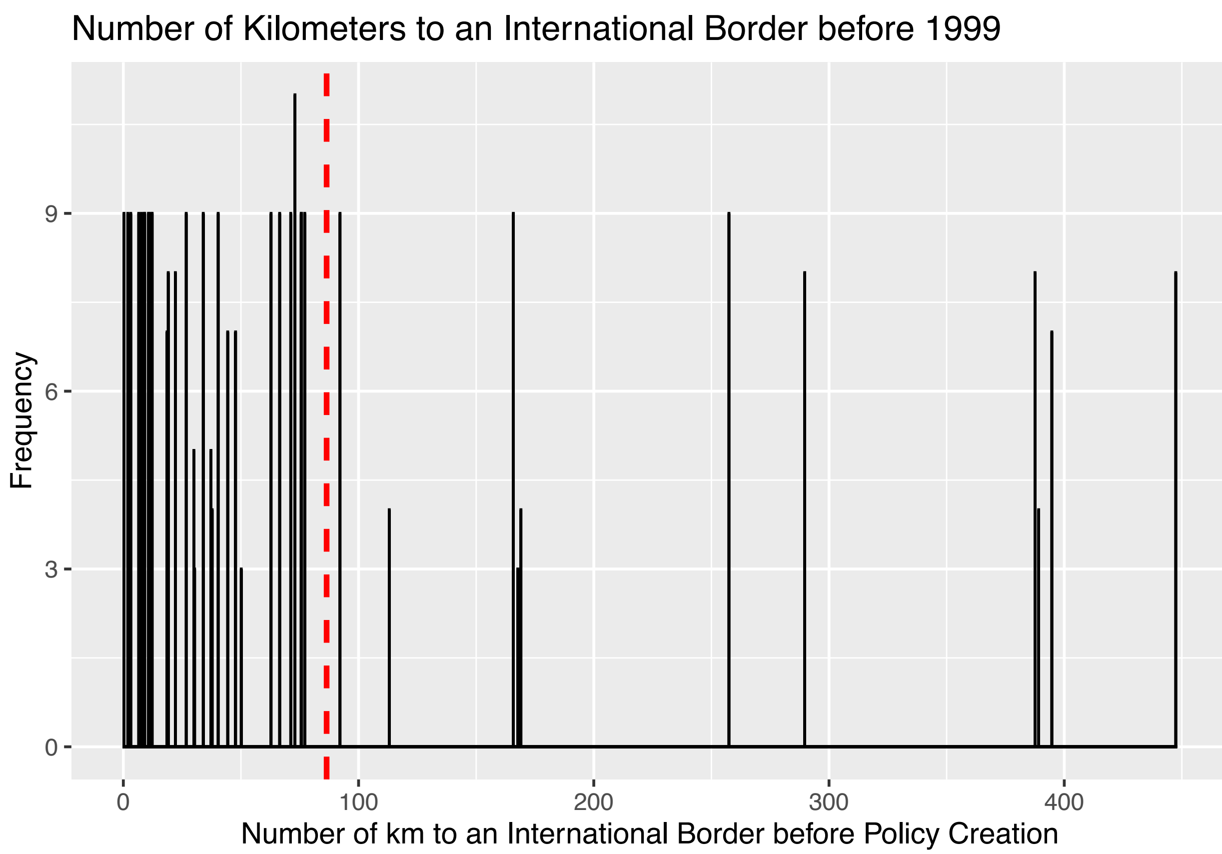
```
ggplot(data = polyes, aes(x = km)) +
  geom_histogram(color = "black" , fill = "black", binwidth = .01) +
 xlab("Number of km to an International Border after Policy Creation") +
  ylab("Frequency") +
  labs(title = "Number of Kilometers to an International Border after 1999") +
   geom_vline(aes(xintercept = mean(km)) , color = "red" , linetype = "dashed" , size = 1)
```



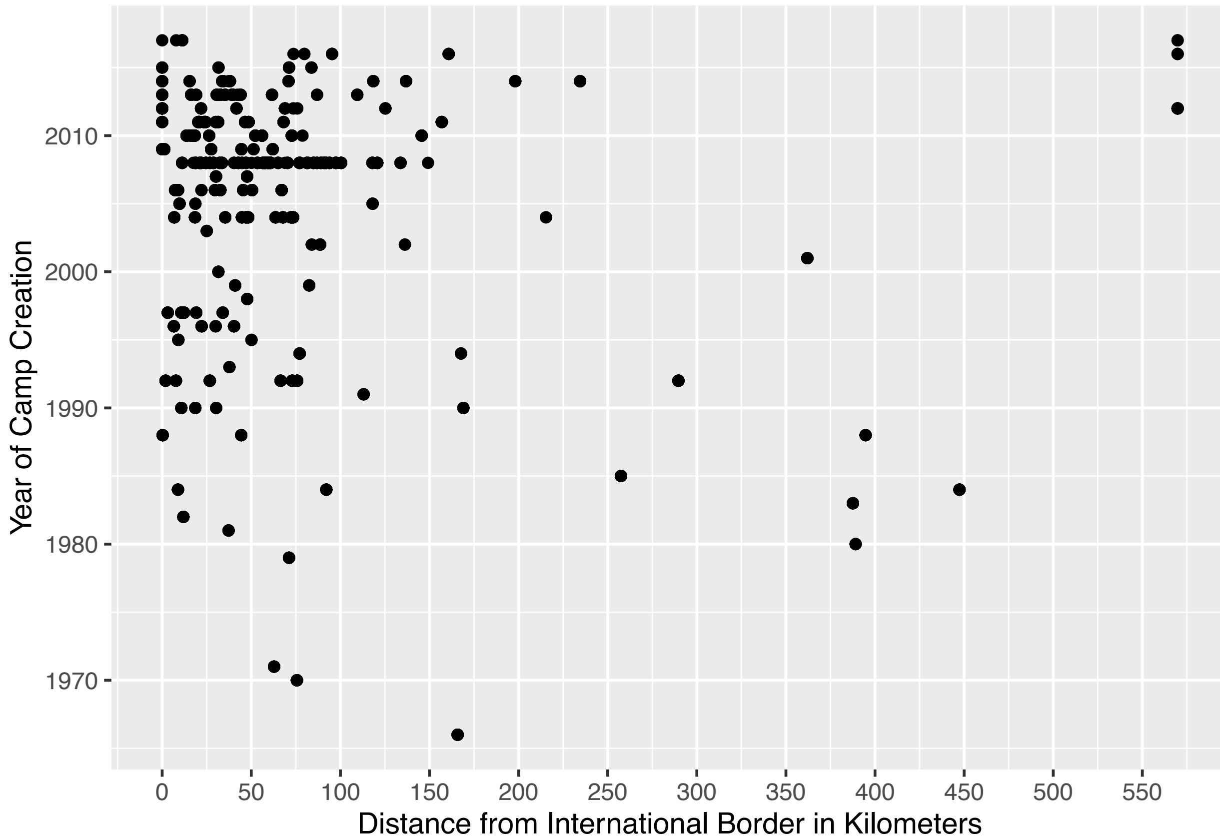

```
polno <- policycreation %>%
  filter(policy == "nopol") %>%
  select(km)

mean(polno$km)
```

```
## [1] 86.41275
```

```
ggplot(data = polno, aes(x = km)) +
  geom_histogram(color = "black" , fill = "black", binwidth = .01) +
  xlab("Number of km to an International Border before Policy Creation") +
  ylab("Frequency") +
  labs(title = "Number of Kilometers to an International Border before 1999") +
  geom_vline(aes(xintercept = mean(km)) , color = "red" , linetype = "dashed" , size = 1)
```




```r
policy <- creation %>%
  mutate(policy = ifelse(creation$yearo >= "1999" , "yespol" , "nopol")) %>%
  filter(!is.na(yearo))

policy <- policy %>%
  select(km, yearo)


ggplot(policy, aes(x= km , y = yearo)) +
  geom_point() +
  ylab("Year of Camp Creation") +
  xlab("Distance from International Border in Kilometers") +
  scale_x_continuous( breaks = c(0 , 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550 , 600))
```

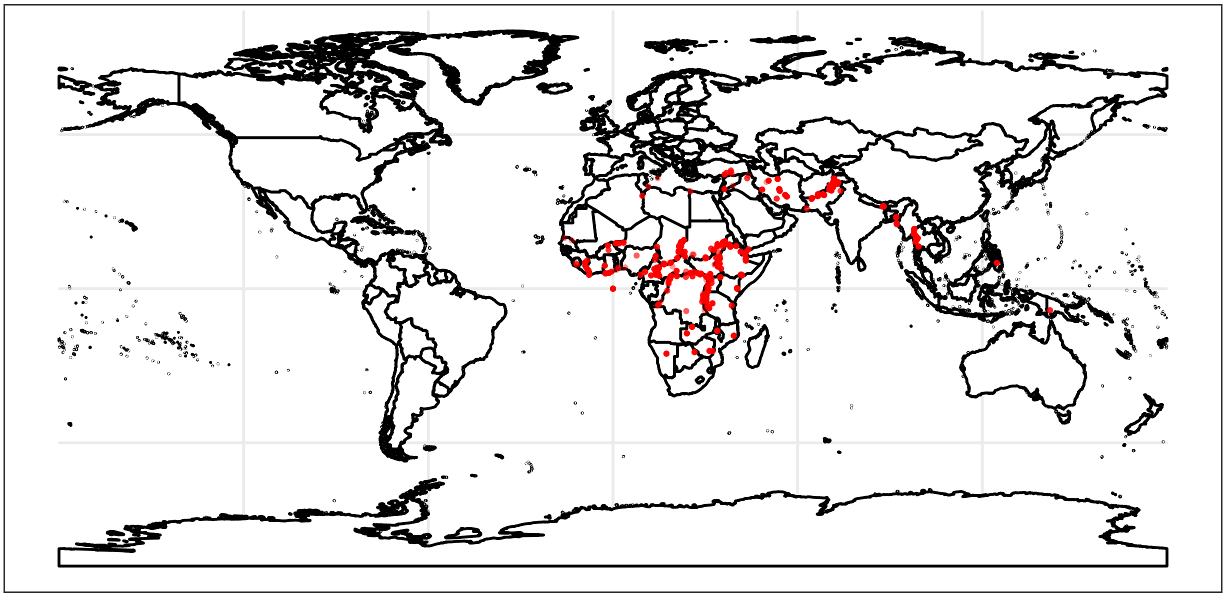



Since 2008, there have been fewer camps created with a more scattered relationship to distance from an international border. A lot of new camps were created in 2008, with a great majority very close to the border. In the past year, for example, only a few camps were created and one of those camps was over 550 kilometers from an international border.

## — a fun little map —

```r
ggplot()  +
  geom_sf(data = borders, color = "black" , fill = "white") +
  geom_sf(data = refugeecamps, color = "red" , fill = "red" , size = .4 , alpha =.4) +
  ggsave("afunlittlegraph.pdf") +
  theme_bw()
```

```
## Saving 6.5 x 4.5 in image
```



```
violence <-read_csv("1997-01-01-2019-07-06.csv")
```

```
violencespatial <- st_as_sf(
  violence,
  coords = c("longitude" , "latitude") ,
  crs = 4326)
```

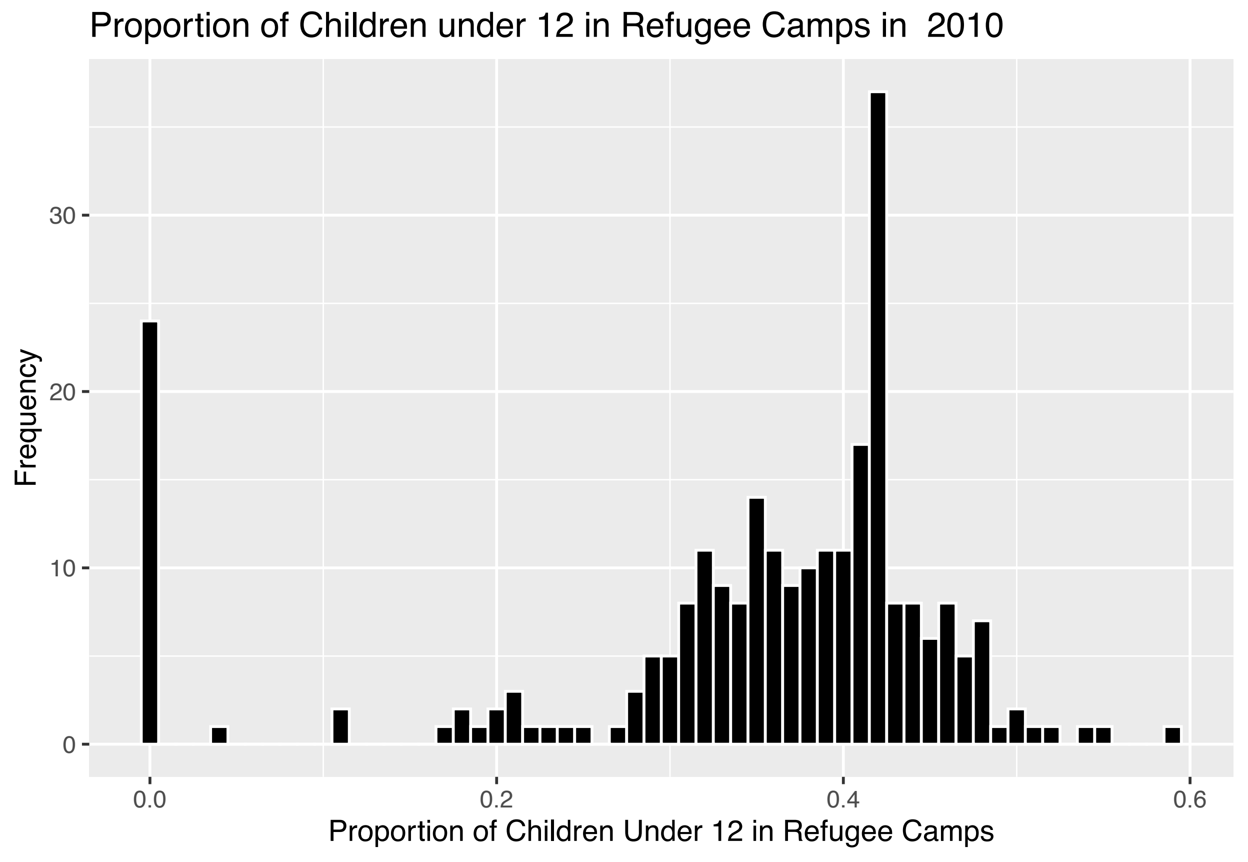
```
country <- camps %>%
  select(iso) %>%
  distinct()

countrycount <- camps %>%
  select(iso, pop_camp, year, pop_concern)

countrycount$pop_camp <- as.numeric(countrycount$pop_camp)


count1 <- countrycount %>%
  filter(countrycount$year == 2010) %>%
  select(iso) %>%
  group_by(iso) %>%
  count("iso")

pop10 <- countrycount %>%
  filter(countrycount$year == 2010) %>%
```

```r
  group_by(iso) %>%
  summarize(frequency = sum(pop_concern))

sum(pop10$frequency)
```

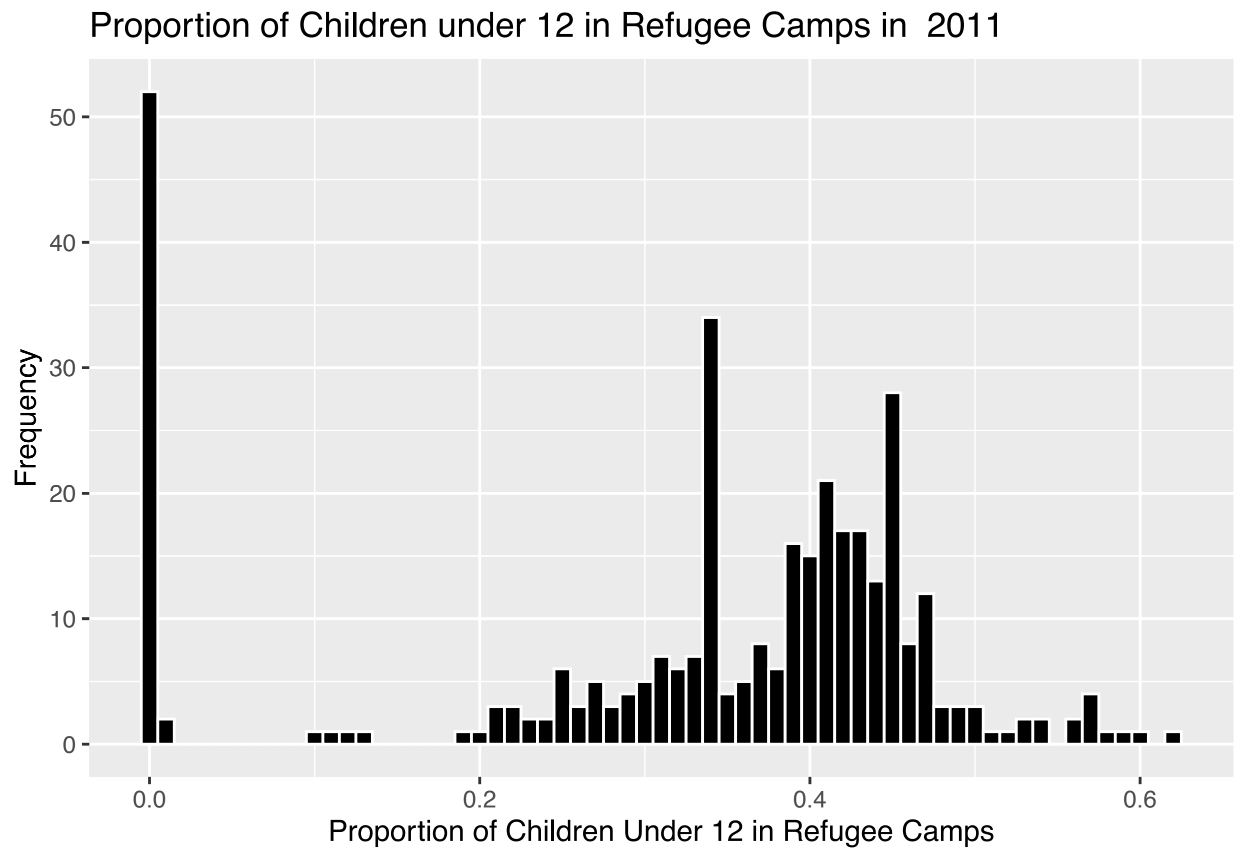
```
## [1] 4635380
```

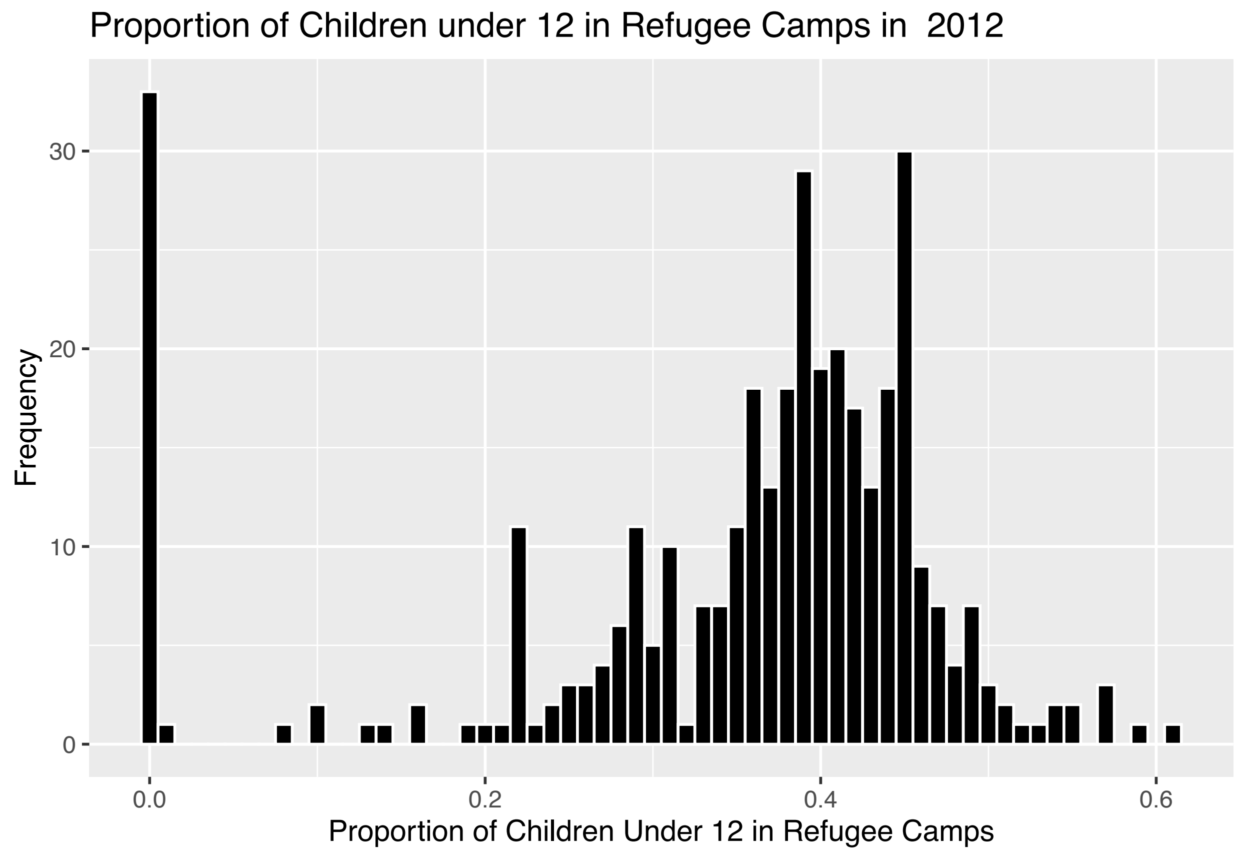
```r
pop18 <- countrycount %>%
  filter(countrycount$year == 2018) %>%
  group_by(iso) %>%
  summarize(frequency = sum(pop_camp))

count18 <- countrycount %>%
  filter(countrycount$year == 2018) %>%
  select(iso) %>%
  group_by(iso)


basicnumbers <- refugeecamps %>%
  group_by(iso) %>%
  count("iso")

sum(basicnumbers$n)
```

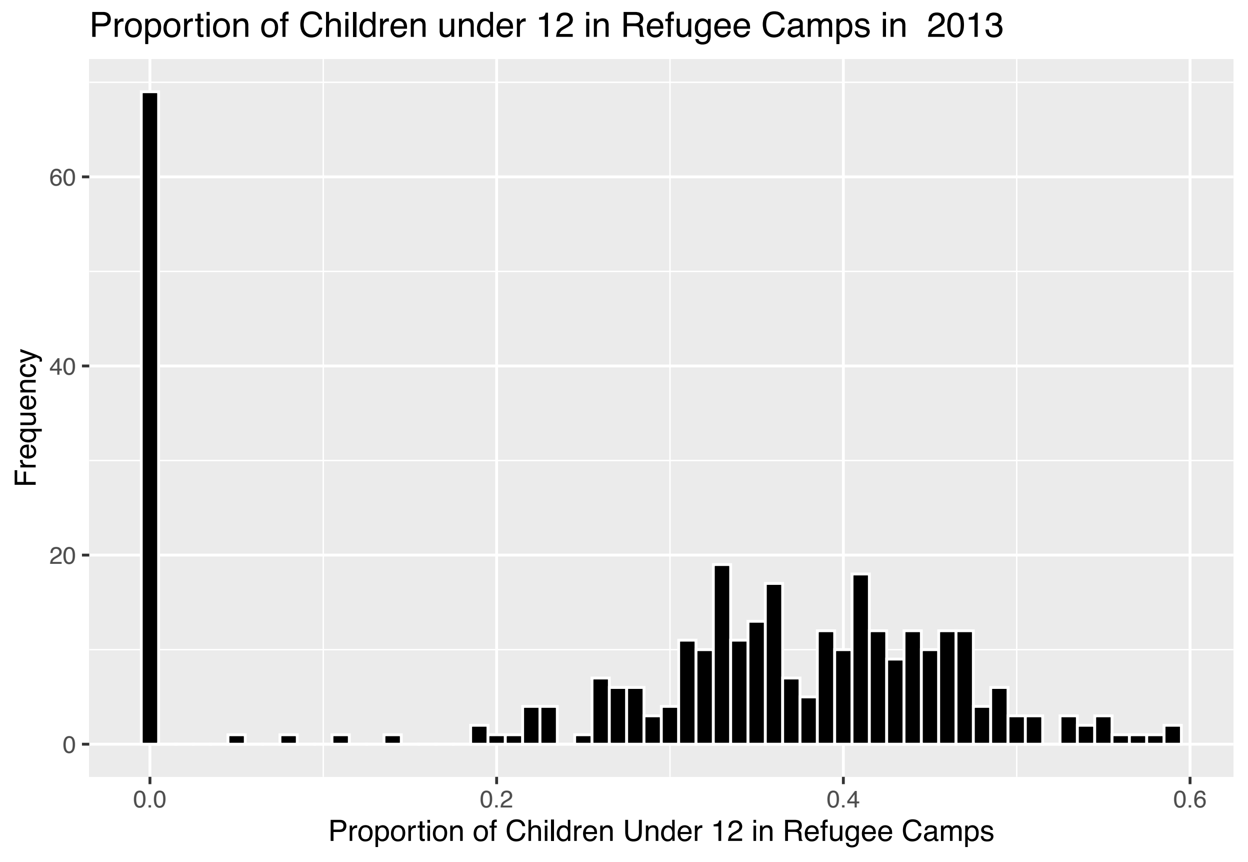
```
## [1] 2025
```



Proportion of Children under 12 in Refugee Camps in 2010
Frequency
30
20
10
0
0.0
0.2
0.4
0.6
Proportion of Children Under 12 in Refugee Camps

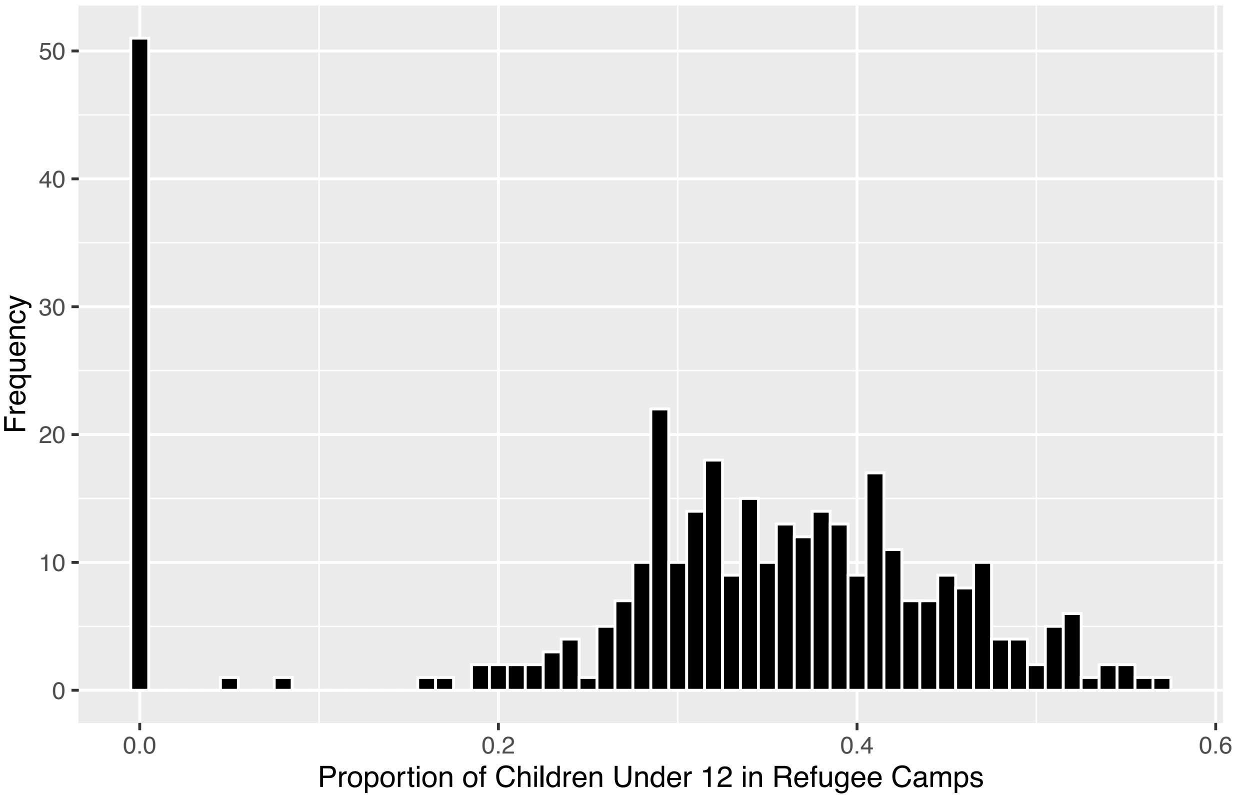

Proportion of Children under 12 in Refugee Camps in 2011
Frequency
0
10
20
30
40
50
0.0
0.2
0.4
0.6
Proportion of Children Under 12 in Refugee Camps


Proportion of Children under 12 in Refugee Camps in 2012
Frequency
0
10
20
30
0.0
0.2
0.4
0.6
Proportion of Children Under 12 in Refugee Camps

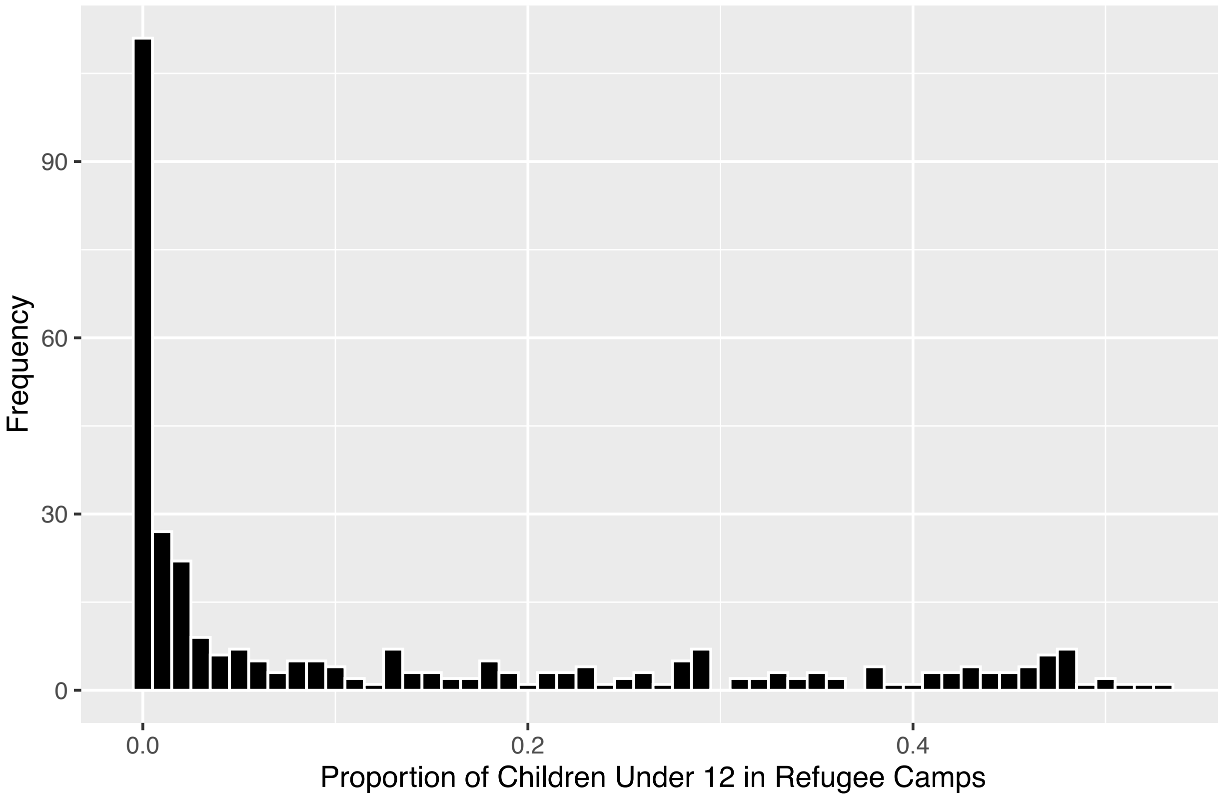


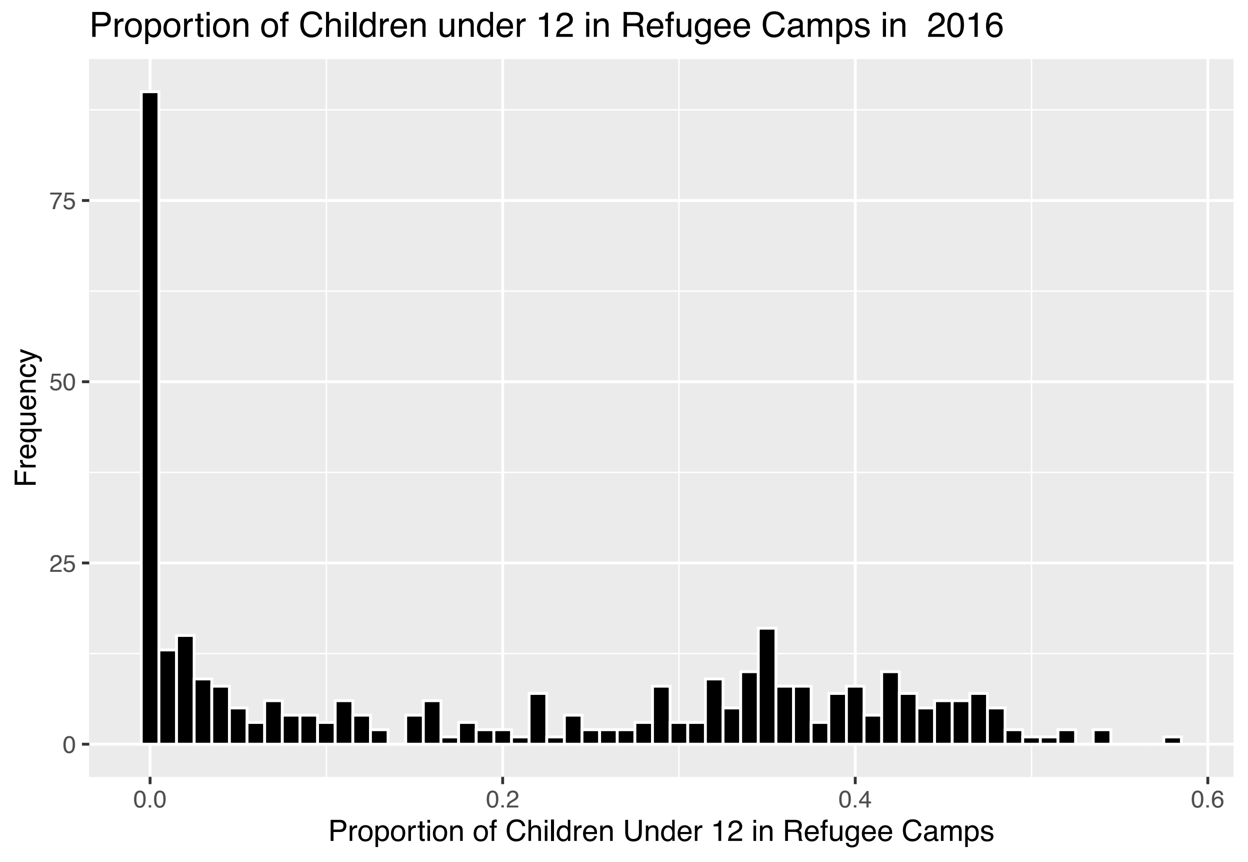
Proportion of Children under 12 in Refugee Camps in 2013
Frequency
60
40
20
0
0.0
0.2
0.4
0.6
Proportion of Children Under 12 in Refugee Camps

Proportion of Children under 12 in Refugee Camps in  2014

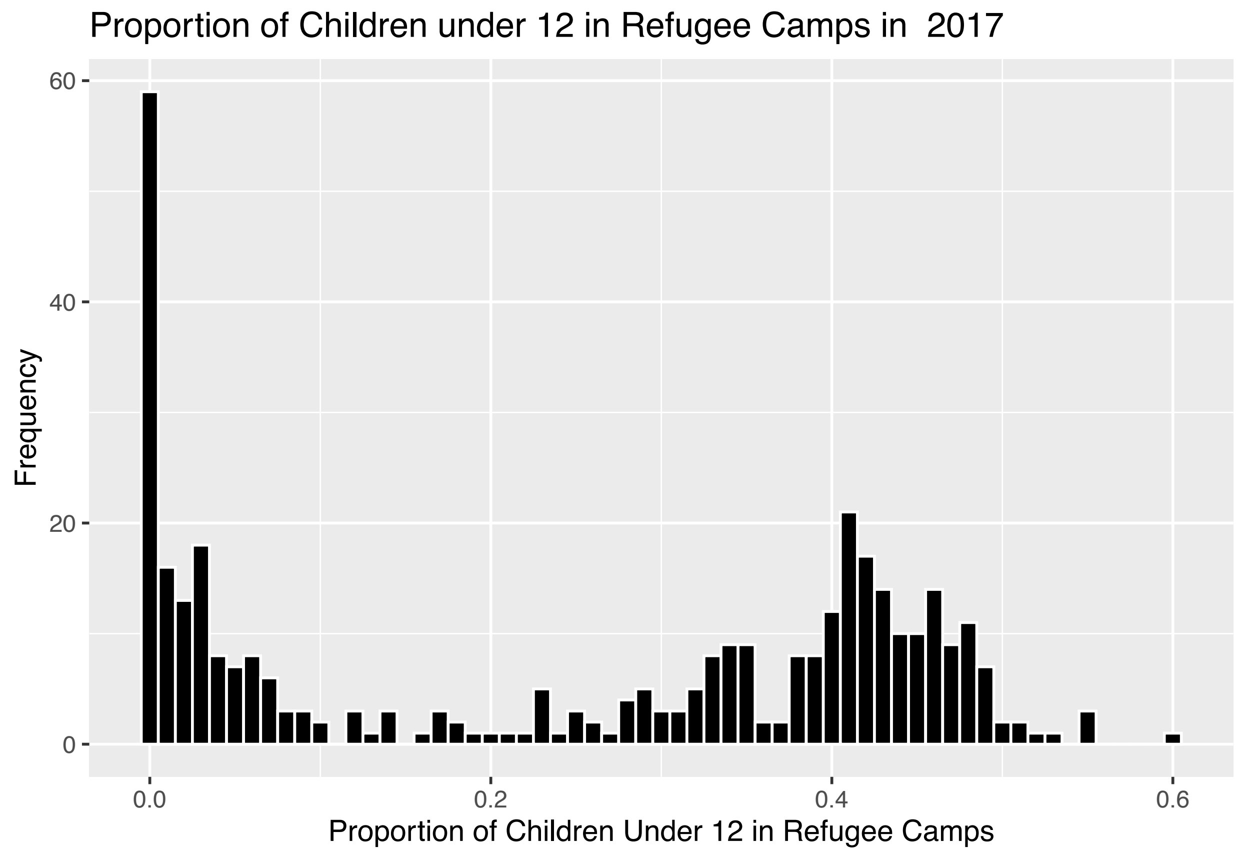


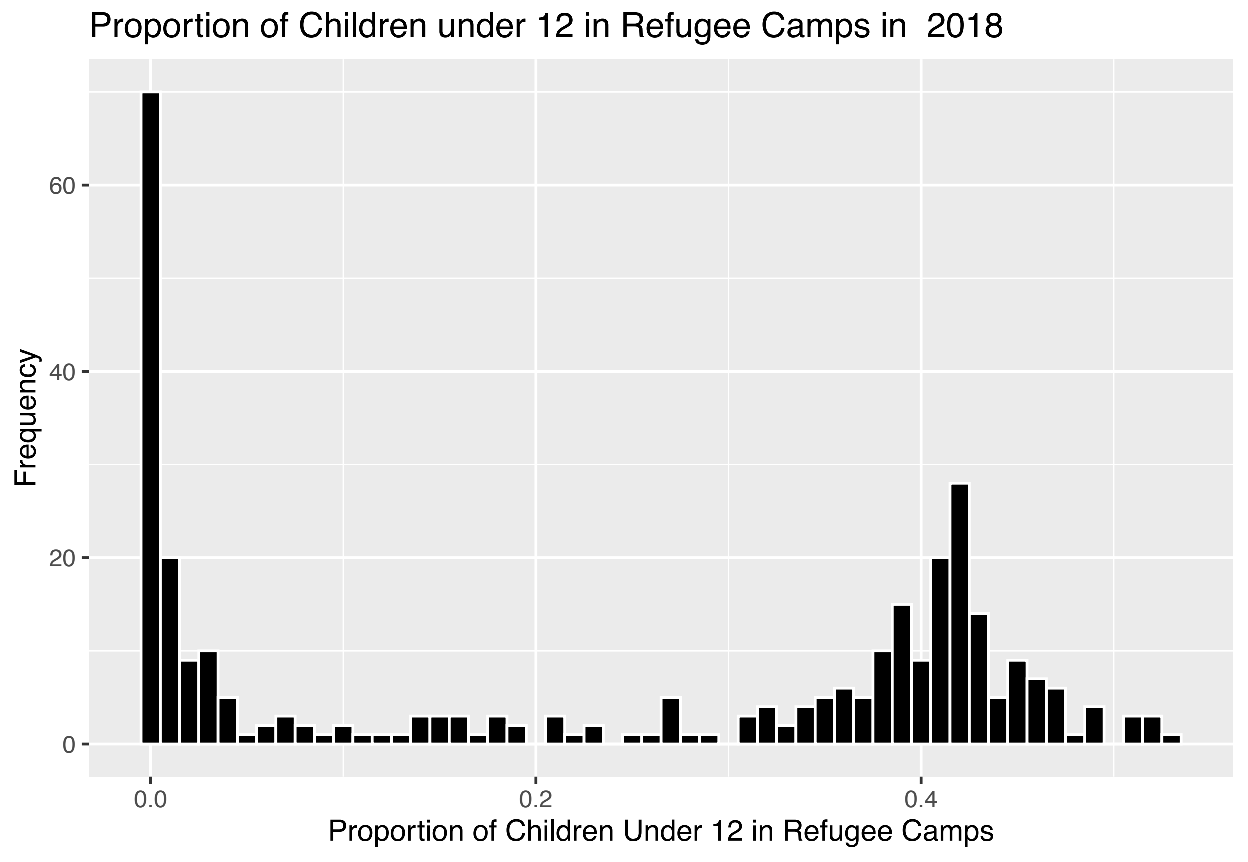
Proportion of Children under 12 in Refugee Camps in  2015





Proportion of Children under 12 in Refugee Camps in 2016
Frequency
75
50
25
0
0.0
0.2
0.4
0.6
Proportion of Children Under 12 in Refugee Camps


Proportion of Children under 12 in Refugee Camps in 2017
Frequency
60
40
20
0
0.0
0.2
0.4
0.6
Proportion of Children Under 12 in Refugee Camps


Proportion of Children under 12 in Refugee Camps in 2018
Frequency
60
40
20
0
0.0
0.2
0.4
Proportion of Children Under 12 in Refugee Camps