

Pilot Data Cleaner

Daniel K Baissa, Melani Cammett, and Aytug Sasmaz

December 12, 2020

Loading in data

I will start by creating a function to automatically set up the data.

All of this is a work in progress

Now that the function is created, we can use it to make our data. There will be a warning, but it is totally ok for now.

How Many Observations

As of the evening of December 13 Qualtrics recorded the following total of responses:

```
## [1] 942
```

We can use approximately:

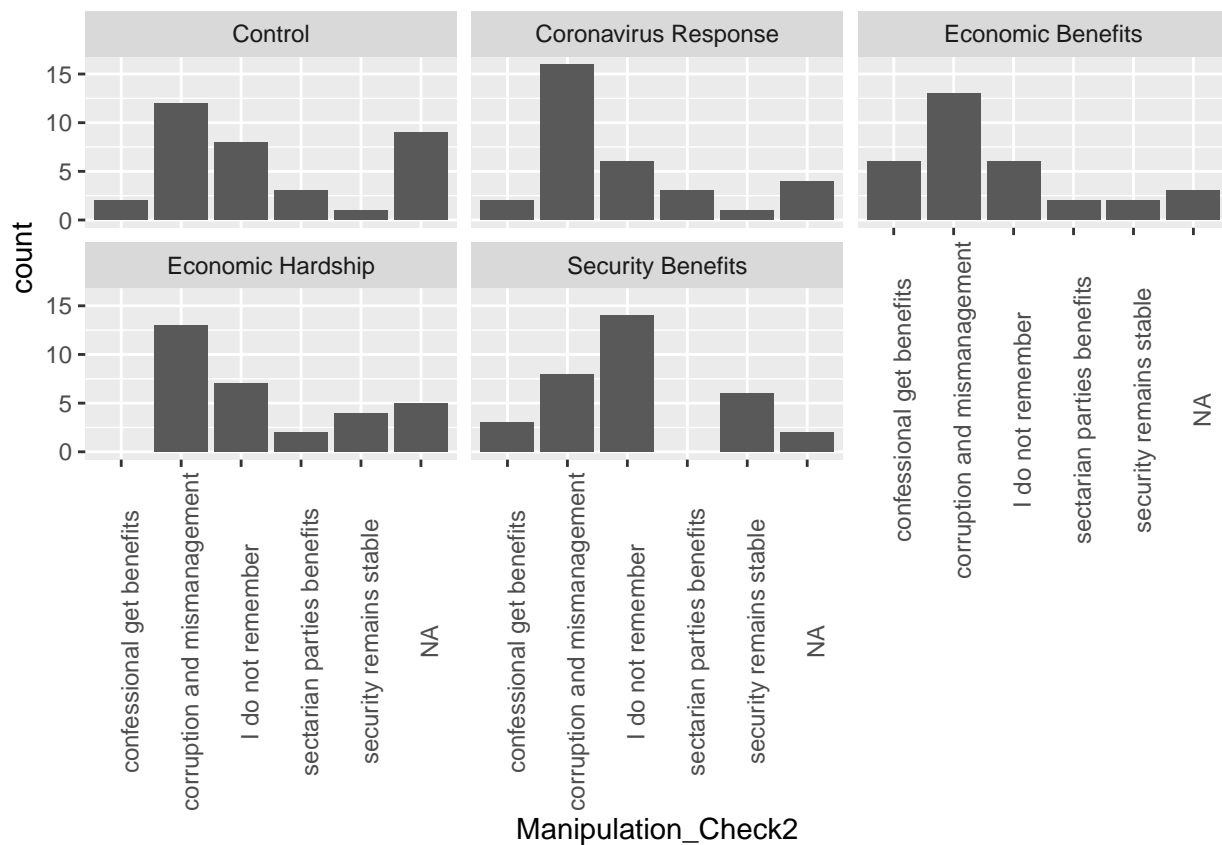
```
## [1] 163
```

Of which, most if not all are 100% complete.

The large discrepancy is because Qualtrics automatically completes the observation after 72 hours.

Manipulation Check

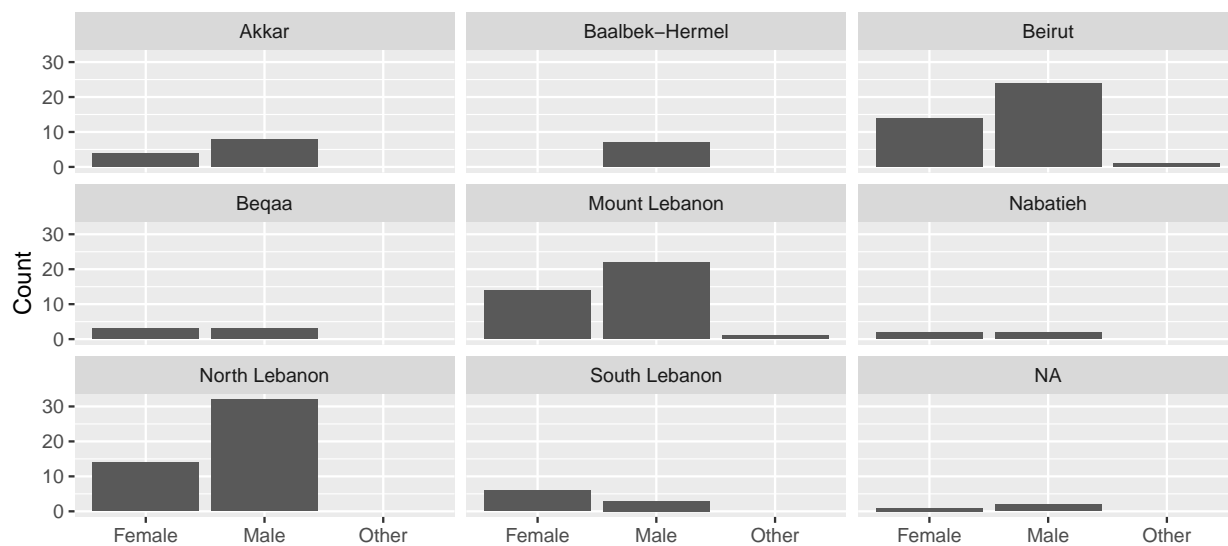
It looks like respondents are heavily biased towards saying “has resulted in corruption and mismanagement of the economy”. It was really high even in the control group



Demographic distribution

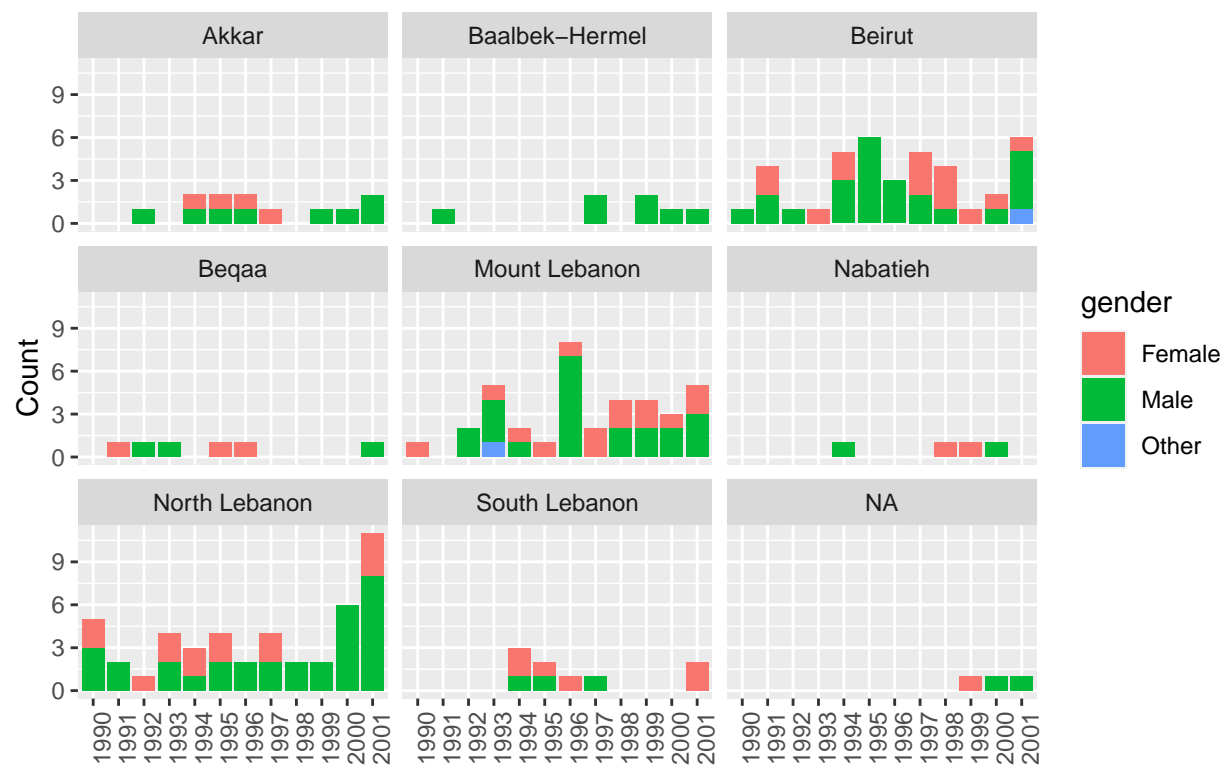
By governorate

Gender Distribution

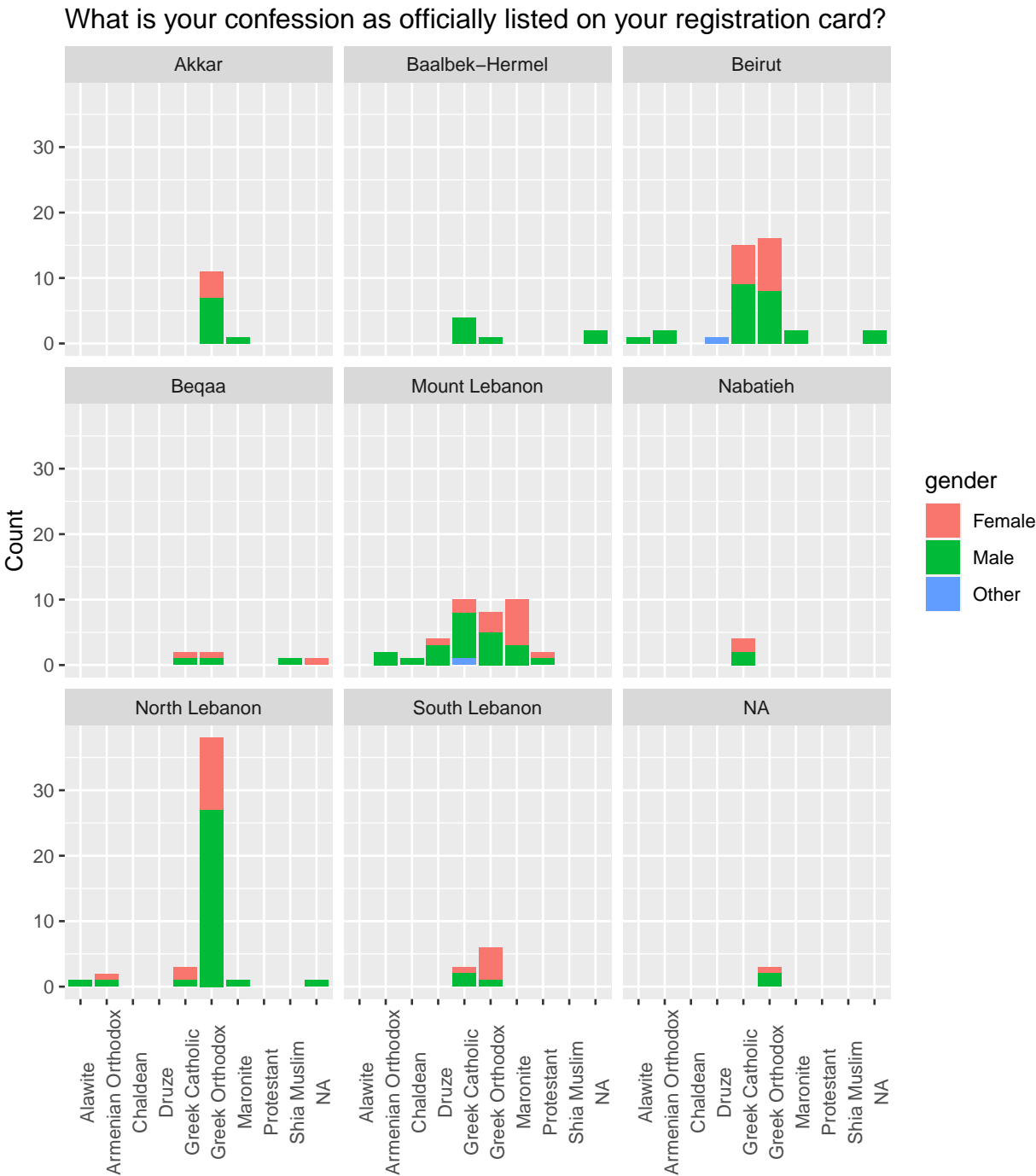


Age demographics

Birth Year Distribution



Religious demographics



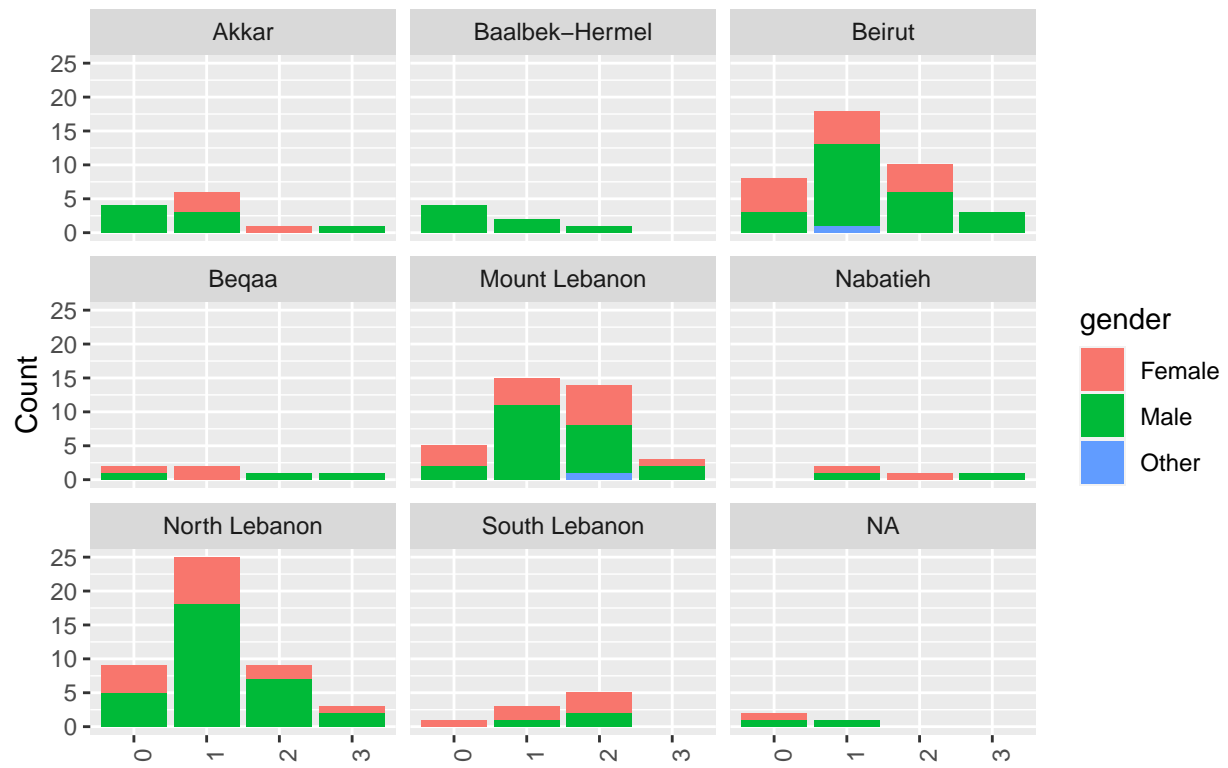
Income distribution

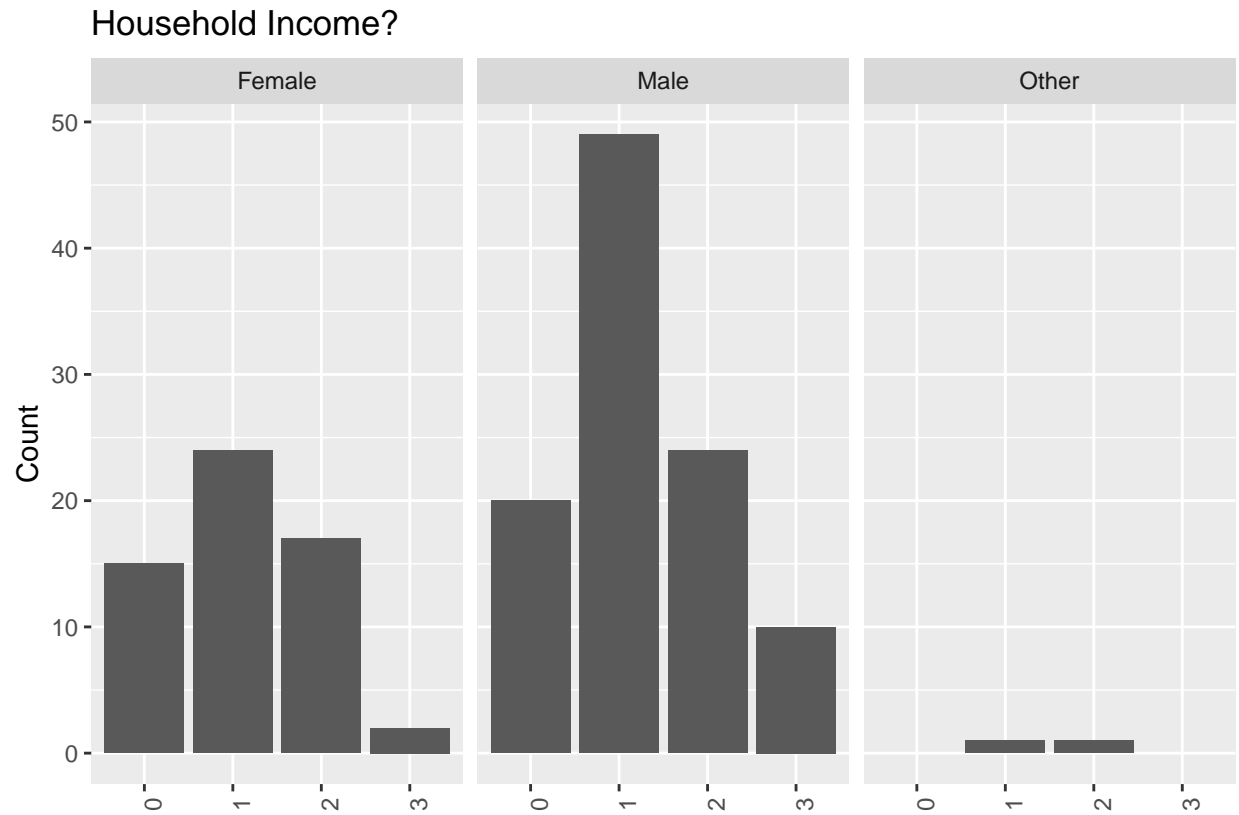
Below are some statements related to your household income. Which of these statements comes closest to describing your household income?

```
summary(as.double(df$dem_income1))
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.00    1.00    1.00   1.19    2.00    3.00
```

Household Income?



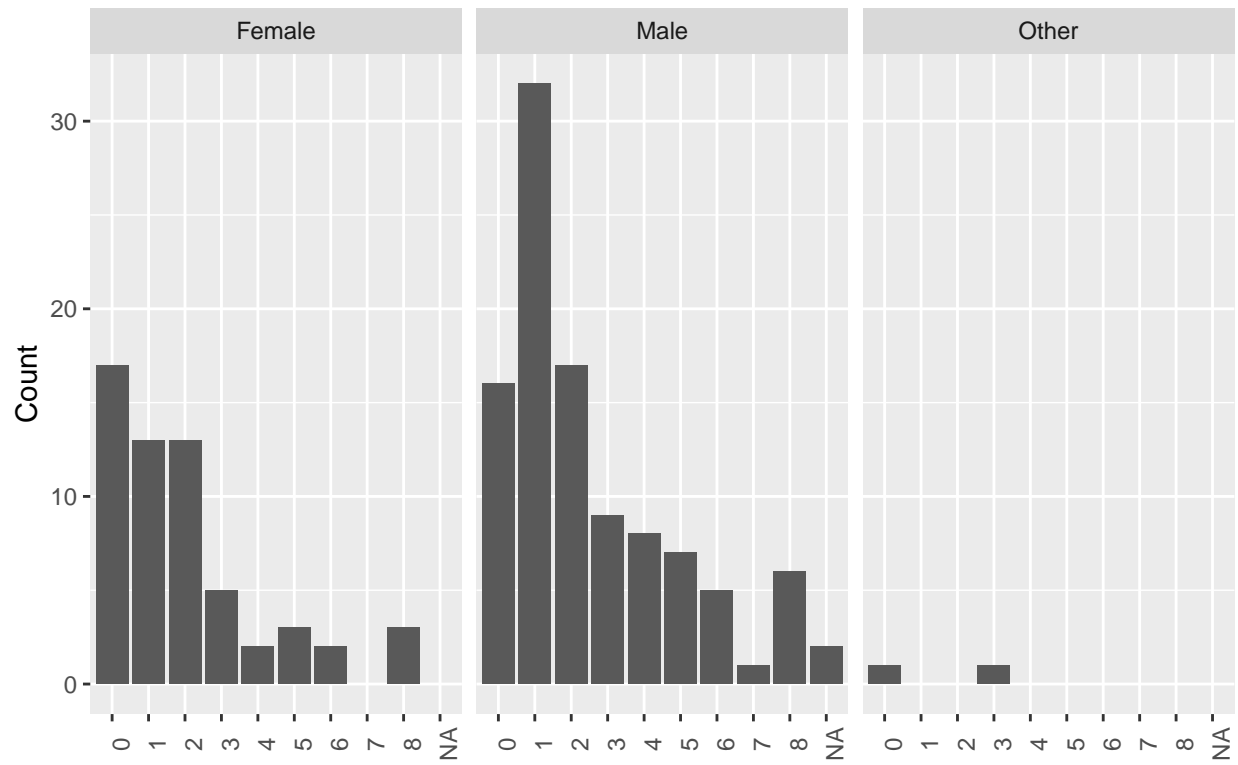


To the best of your knowledge, what is your household's total net income in Lebanese Liras (L.L.) in a typical month?

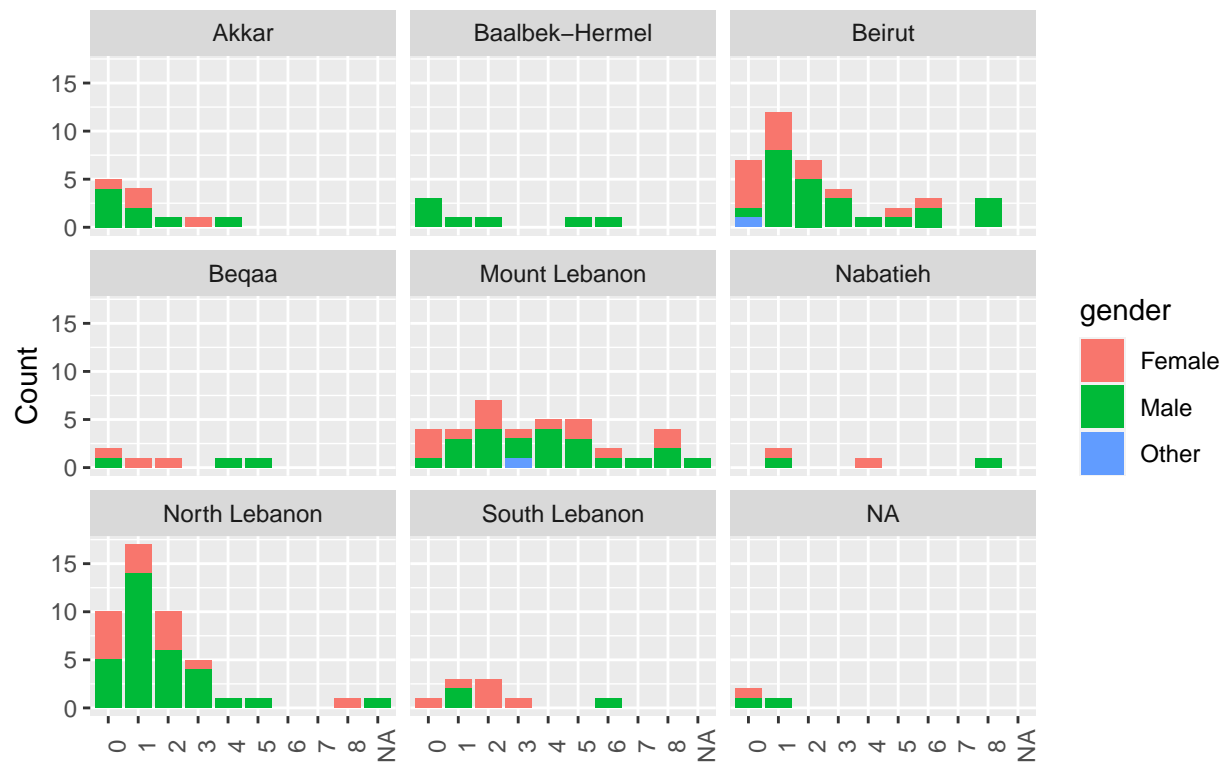
```
summary(as.double(df$dem_income2))
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##    0.000   1.000   2.000   2.242   3.000   8.000     2
```

What is your household's total net income in Lebanese Liras a month?



What is your household's total net income in Lebanese Liras a month?



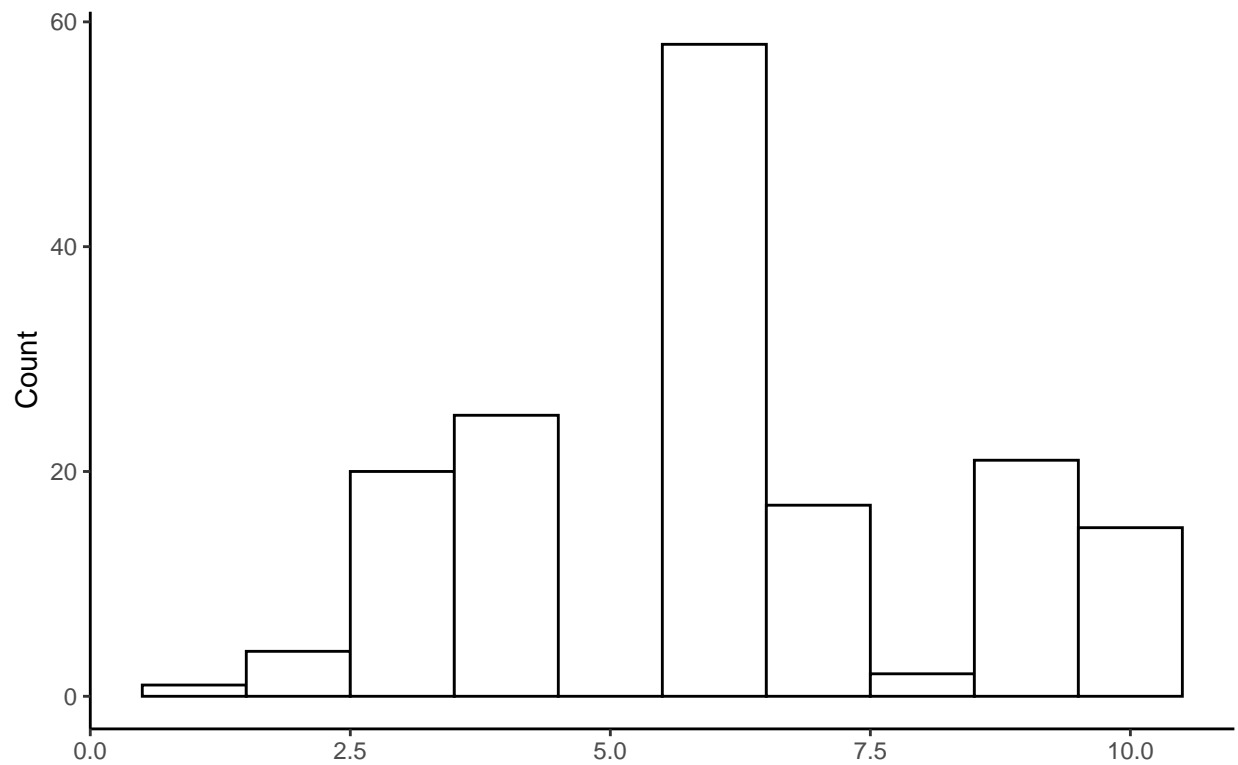
What is the highest level of education that you completed?

```
summary(as.double(df$dem_edu))
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.00   4.00   6.00   6.08   7.00   10.00
```

```
ggplot(df, aes(x = as.double(dem_edu)))+
  geom_histogram(bins = 10, fill="white", color="black")+
  labs(title="What is the highest level of education that you completed?", x="", y = "Count")+
  theme_classic()
```


What is the highest level of education that you completed?



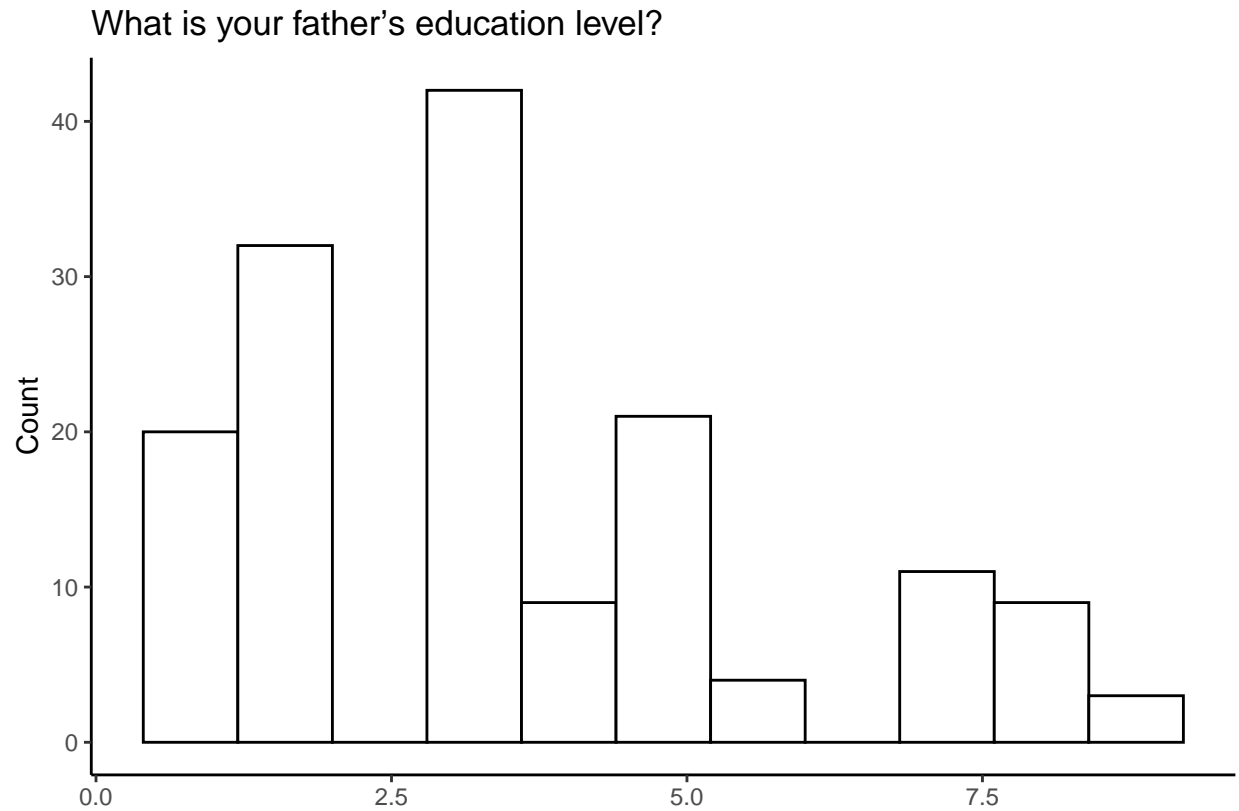
What is your father's education level?

```
summary(as.double(df$dem_fatheredu))
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's  
##      1.000  2.000   3.000   3.649  5.000   9.000    12
```

```
ggplot(df, aes(x = as.double(dem_fatheredu)))+  
  geom_histogram(bins = 11, fill="white", color="black")+  
  labs(title="What is your father's education level? ", x="", y = "Count")+  
  theme_classic()
```

```
## Warning: Removed 12 rows containing non-finite values (stat_bin).
```

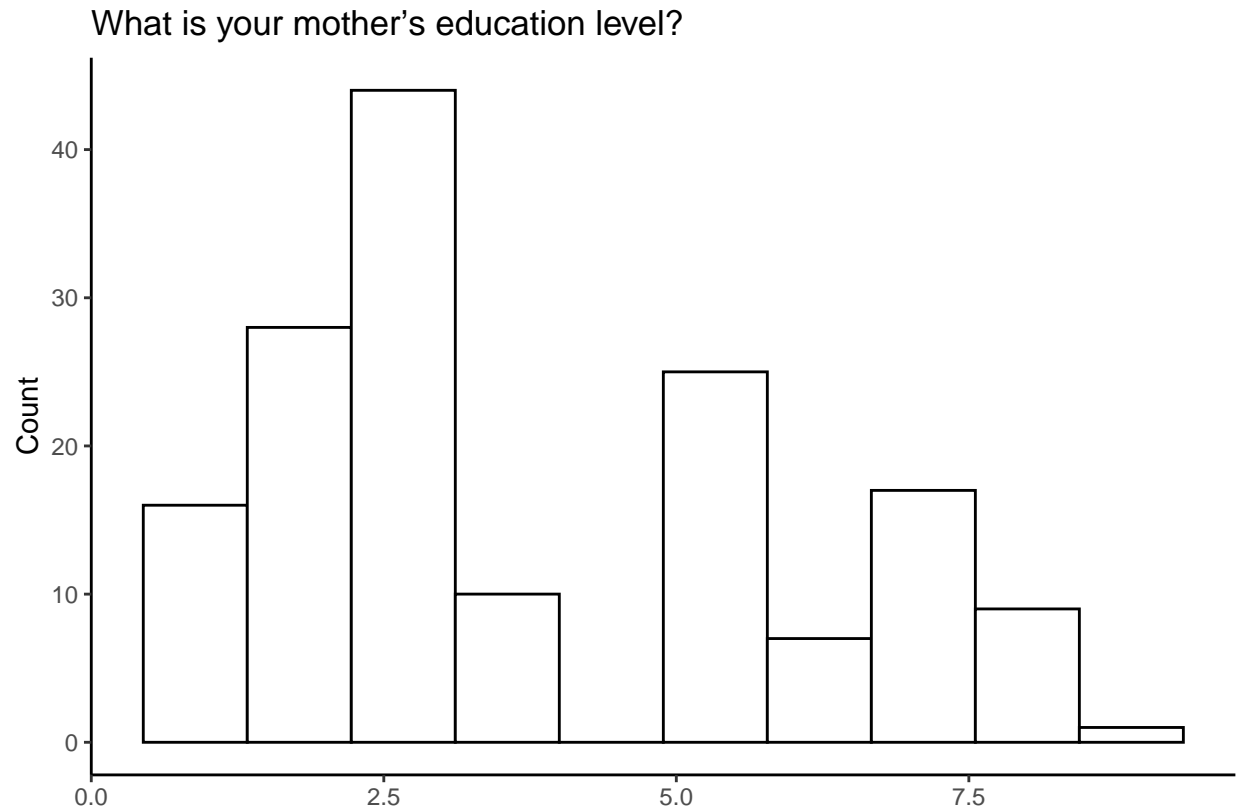


```
summary(as.double(df$dem_motheredu))
```

```
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's  
##   1.000   2.000   3.000   3.892   5.000   9.000     6
```

```
ggplot(df, aes(x = as.double(dem_motheredu)))+  
  geom_histogram(bins = 10, fill="white", color="black")+  
  labs(title="What is your mother's education level?", x="", y = "Count")+  
  theme_classic()
```

```
## Warning: Removed 6 rows containing non-finite values (stat_bin).
```

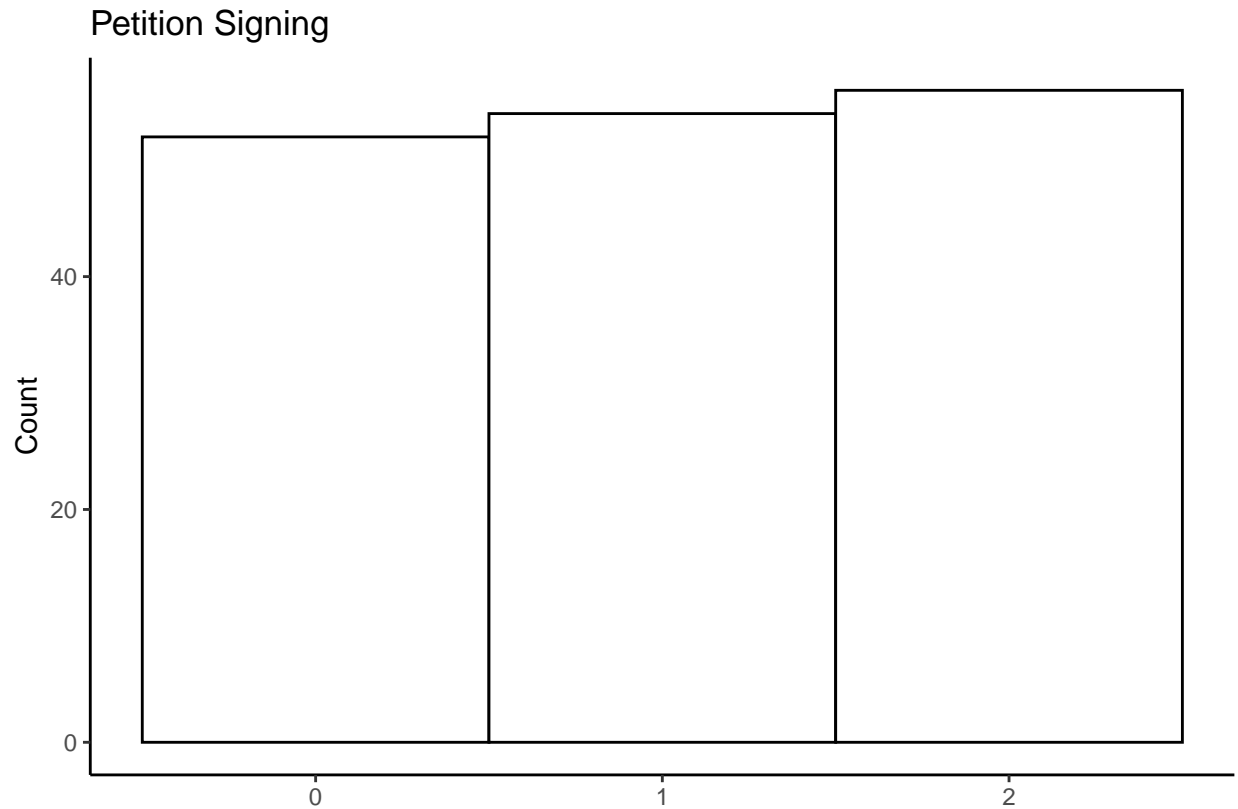


Mother's are better educated than fathers in this sample

Petition Signing distribution

How many people actually signed the petition? Lets find out.

```
## Warning: Removed 1 rows containing non-finite values (stat_bin).
```



Lets do some basic stats on the petition experiment. Here is a simple logistic regression to see if the control group is different from any of the treatments.

```
# colnames(df)
# unique(df$Petition_Experiment_Treatment)

df2 <- df %>%
  filter(Petition_Experiment_Treatment == "ptt_cont_foll" | Petition_Experiment_Treatment == "ptt_treat_eco_foll")

df2$pt_econ_treatment <- 0
df2$pt_econ_treatment[which(df2$Petition_Experiment_Treatment == "ptt_treat_eco_foll")] <- 1
df2$signed <- 0
df2$signed[which(df2$Q144 > 0)] <- 1

test <- glm(signed ~ pt_econ_treatment, data = df2, family = binomial())

summary(test)

##
## Call:
## glm(formula = signed ~ pt_econ_treatment, family = binomial(),
##      data = df2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5829  -1.5043   0.8203   0.8826   0.8826
##
```

```
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      0.9163     0.3742   2.449   0.0143 *
## pt_econ_treatment -0.1744     0.5363  -0.325   0.7451
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 80.970  on 65  degrees of freedom
## Residual deviance: 80.864  on 64  degrees of freedom
## AIC: 84.864
##
## Number of Fisher Scoring iterations: 4
## combined table
(ctable <- cbind(ctable, "p value" = p))
```

```
##              Value Std. Error
## Petition_Experiment_Treatmentptt_treat_bft_foll -0.4939166  0.4581677
## Petition_Experiment_Treatmentptt_treat_cor_foll -0.9289745  0.4618838
## Petition_Experiment_Treatmentptt_treat_eco_foll -0.2264382  0.4748410
## Petition_Experiment_Treatmentptt_treat_sec_foll -0.4178841  0.4569373
## UserLanguageEN                                0.7416721  0.3207150
## 0|1                                           -0.9672841  0.3562132
## 1|2                                           0.4860639  0.3511012
##              t value      p value
## Petition_Experiment_Treatmentptt_treat_bft_foll -1.0780258  0.281022248
## Petition_Experiment_Treatmentptt_treat_cor_foll -2.0112731  0.044296613
## Petition_Experiment_Treatmentptt_treat_eco_foll -0.4768717  0.633453493
## Petition_Experiment_Treatmentptt_treat_sec_foll -0.9145328  0.360436938
## UserLanguageEN                                2.3125580  0.020746951
## 0|1                                           -2.7154638  0.006618302
## 1|2                                           1.3843984  0.166236510
```

The Covid Treatment was the treatment left out for comparison by the model. Lets take a look at that.

```
signif((ctable <- cbind(ctable, "p value" = p)),3)
```

```
##              Value Std. Error t value
## Petition_Experiment_Treatmentptt_treat_cor_foll -0.996     0.481   -2.07
## UserLanguageEN                                1.590     0.519    3.06
## 0|1                                           -0.760     0.394   -1.93
## 1|2                                           0.800     0.402    1.99
##              p value
## Petition_Experiment_Treatmentptt_treat_cor_foll 0.03840
## UserLanguageEN                                0.00224
## 0|1                                           0.05360
## 1|2                                           0.04640
```

Something appears to be going on here, but the N is so small that it could still be chance