Team 6 Report (word count: 4903)

How does alcohol use appear to affect the prevalence of violence against women?

AUTHORS
Alison Yao
Oscar Bray

AFFILIATIONS
NYU Shanghai
NYU Abu Dhabi

PUBLISHED
Dec. 19, 2021

Introduction

Violence against women (VaW) is an ongoing and universal issue that has only recently been given the attention it deserves within social research and activism. For this paper, we are defining 'violence against women' along the same lines as those used when creating the Violence Against Women Index (VAWI) to measure the phenomenon, which itself is taken from the World Health Organization: "VaW is defined as any act or behavior by partner, family, or community that causes physical, psychological or sexual harm to women" (Cepeda, Lacalle-Calderon, and Torralba 2021, 2). Despite this broad definition, the data that we are focusing on measures the prevalence of VaW by country according to the percentage of women in each population who have experienced physical and/or sexual violence from an intimate partner during their lifetime. Since our paper exclusively investigates prevalence as an indicator, any further mentions of VaW are referring to prevalence as this measurement outlines it (unless stated otherwise).

Our aim for this report was to investigate the potential effects of alcohol use on the prevalence of VaW internationally. We chose to examine alcohol use as a potential factor because it is commonly assumed that alcohol consumption can induce aggressive behavior and a decrease in inhibitions, which would logically make incidents of violence more likely. This assumption is supported by extensive research spanning multiple decades (Jewkes 2002, 1425–26; White 1995, 281–82; *Understanding Violence Against Women* 1996, 54–56) but there is also a consensus that the social role and context of alcohol consumption also plays a role in contributing to violent behavior (*Understanding Violence Against Women* 1996, 55–56; White 1995, 282). Under this paradigm, alcohol is more likely to affect the prevalence of incidents of VaW rather than attitudes or legal standings towards VaW, which is why we chose prevalence as our main indicator and the basis of our definition of VaW.

Moreover, data about alcohol consumption was not only readily available but split across multiple datasets according to gender, which we thought could add an extra dimension to our analysis. By comparing alcohol use and violence along gendered lines, we felt that we were responding to several pre-existing issues within prior research on this topic. Much of the foundational research attempting to explain alcohol's relation to intimate partner violence (IPV) is outdated and culturally homogeneous. Despite claims that the relationship between alcohol use and VaW exists even after controlling for sociodemographic characteristics (*Understanding Violence Against Women* 1996, 55), much research has been based in the United States and hasn't considered the role of different cultural and ethnic attitudes towards alcohol use (White 1995, 282). Additionally, because of the sensitive nature of this topic, there are often problems with sampling in research about alcohol use and domestic violence. Not only is it difficult to ascertain prevalence due to low report rates (which forces researchers to use small sample sizes), but the subjects of this research are often extreme cases since they are sourced from domestic violence shelters, alcoholism support groups or other places where those most affected seek help (White 1995, 283–84). This has the potential to skew any interpretations of results and attempts to create solutions based on the research; it probably also affected the integrity of the prevalence data we used. However, the most interesting flaw in the research to us was the fact that there is a lack of focus on the role of alcohol use among female victims compared to male perpetrators. While alcohol use is associated with increased aggression, it also has the potential to incapacitate victims of violence or cause them to engage in behavior that triggers violent retaliation from perpetrators (White 1995, 284–86). We think that our research on the gendered dimensions of alcohol use and its correlation with domestic violence could provide interesting insights within that realm of study, especially since it uses data from multiple countries.

Theory/Hypotheses

Based on the research cited above, we speculated that there would be a strong positive correlation between alcohol consumption and prevalence of VaW. We thought that there was sufficiently strong evidence of alcohol's role in incidents of VaW, as well as social and biological explanations as to why that was the case.

Therefore, our research question is **how does alcohol consumption appear to reduce or increase violence against women**?

Data

To analyze data in R, we first needed to import the libraries and data we needed and join them in one dataframe.

Import Data

VaW Indicators

Here, we imported libraries and loaded data from a 2019 OECD study of violence against women.

```
library (tidyverse)
library (brms )
library (corrplot)
library (ggpmisc )
library (dagitty )
library (ggdag )
attitude_df = read.csv('./Data/attitude.csv')
law_df = read.csv('./Data/laws.csv')
prevalence_df = read.csv('./Data/prevalence.csv')
```

Covid-19 data from Lab06

We recall from Lab06 that we worked with COVID-19 data from 2019 that listed basic information about countries. In order to examine trends relevant to VaW worldwide, we read the data and selected the 3 columns that we were interested in: country names, abbreviations and continents. We also checked if there are duplicated countries just to be safe.

```
covid_data <- read_csv("https://wzb-ipi.github.io/corona/df_full.csv"
)
covid_subset <- covid_data[c('geoid2', 'country', 'continent'
)]
covid_subset <- covid_subset[!duplicated(covid_subset)
, ]
head(covid_subset)
```

```
# A tibble: 6 × 3
  geoid2 country     continent
  <chr>  <chr>       <chr>
1 ABW    Aruba       America
2 AFG    Afghanistan Asia
3 AGO    Angola      Africa
4 AIA    Anguilla    America
5 ALB    Albania     Europe
6 AND    Andorra     Europe
```

```
# check if there are duplicated countries by running
# dplyr::count(covid_subset, geoid2, sort = TRUE)
# and there are no duplicates
```

We needed to be careful of mismatches because the data we have so far is pulled from two different sources. Here, we checked if every country abbreviation mentioned in the attitude, law and prevalence datasets can also be found in the COVID dataset.

```
check_code <- function(df) {
   for (code in df$LOCATION) {
    if (code %in% covid_subset$geoid2) {
    } else {
      print(paste0("Not Found: ", code))
    }
```

```
  }
}
check_code(        attitude_df)
```

```
[1] "Not Found: TKM"
[1] "Not Found: HKG"
```

```
check_code(        law_df   )
```

```
[1] "Not Found: HKG"
[1] "Not Found: TKM"
```

```
check_code(        prevalence_df)
```

```
[1] "Not Found: HKG"
```

The ideal case is there would not be any output, but here we have two country abbreviations missing, so we added HKG and TKM to the COVID data manually.

```
covid_subset <-        covid_subset %>%      add_row (        geoid2 =          "HKG"     ,
                          country =         "Hong Kong"   ,
                          continent =        "Asia"   )
covid_subset <-        covid_subset %>%      add_row (        geoid2 =          "TKM"     ,
                          country =         "Turkmenistan",
                          continent =        "Asia"   )
colnames(        covid_subset)         <-        c        (        'Country', 'CountryName',
'Continent')
head     (        covid_subset)
```

```
# A tibble: 6 × 3
  Country CountryName Continent
  <chr>   <chr>       <chr>
1 ABW     Aruba       America
2 AFG     Afghanistan Asia
3 AGO     Angola      Africa
4 AIA     Anguilla    America
5 ALB     Albania     Europe
6 AND     Andorra     Europe
```

Alcohol and poverty data for regression analysis

Since we are investigating the relationship between alcohol consumption and prevalence and considering poverty as a potential confounding variable, we acquired data on Total Alcohol Consumption per capita and Poverty headcount ratio at national poverty lines from the World Bank.

```
alcohol   =         read.csv(        './Data/alcohol.csv')
colnames(        alcohol  )        <-        c        (        'CountryName', 'Country',
'Alcohol_pc')
alcohol  <-        alcohol [        c        (        'Country', 'Alcohol_pc')
```

```
]
poverty_headcount_ratio = read.csv('./Data/poverty_cleaned.csv')
poverty_headcount_ratio <- poverty_headcount_ratio[c('Country', 'poverty_headcount_ratio')]
```

Join Data

We chose the columns of interest from the original three dataframes on VaW indicators and combined them using `full_join()`. This function allowed us to see more clearly which country was missing which indicator. Since R automatically drops the 'not applicable' values (NAs) when plotting graphs and doing regression analyses, we kept the NAs in our dataframe at this point.

For the rest of the dataframes, we used a left join because we are only interested in the countries for VaW analysis.

```
# Not sure why this happened, but Oscar's computer modifies the column names, so we first
rename the columns.
colnames(attitude_df) <- c('LOCATION', 'INDICATOR', 'SUBJECT', 'MEASURE', 'FREQUENCY', 'TIME', 'Value', 'Flag.Codes')
attitude_sub <- attitude_df[c('LOCATION', 'Value')]
colnames(attitude_sub) <- c('Country', 'Attitude')

colnames(prevalence_df) <- c('LOCATION', 'INDICATOR', 'SUBJECT', 'MEASURE', 'FREQUENCY', 'TIME', 'Value', 'Flag.Codes')
prevalence_sub <- prevalence_df[c('LOCATION', 'Value')]
colnames(prevalence_sub) <- c('Country', 'Prevalence')

colnames(law_df) <- c('LOCATION', 'INDICATOR', 'SUBJECT', 'MEASURE', 'FREQUENCY', 'TIME', 'Value', 'Flag.Codes')
law_sub <- law_df[c('LOCATION', 'Value')]
colnames(law_sub) <- c('Country', 'Law')

df <- full_join(attitude_sub, prevalence_sub, by = "Country")
df <- full_join(df, law_sub, by = "Country")
df <- left_join(df, covid_subset, by = 'Country')
df <- left_join(df, alcohol, by = 'Country')
df <- left_join(df, poverty_headcount_ratio, by = 'Country')
head(df) # check dataframe
```

```
  Country Attitude Prevalence  Law CountryName Continent Alcohol_pc
1     AUS      3.2       16.9 0.75   Australia   Oceania      10.51
2     CAN      7.8        1.9 0.25      Canada   America       8.94
3     FIN     11.2       30.0 0.75     Finland    Europe      10.78
4     FRA      6.6       26.0 0.25      France    Europe      12.33
5     DEU     19.6       22.0 0.75     Germany    Europe      12.91
```

```
6      HUN       8.7       21.0 0.75     Hungary     Europe      11.35
  poverty_headcount_ratio
1                      NA
2                      NA
3                    12.2
4                    13.6
5                    14.8
6                    12.3
```

Once `df` looked clean and ready, we checked statistical information such as data types and numbers of empty values. We can see that Prevalence has 34 NA's, poverty headcount has 24, Attitude has 11 and Alcohol has 3. This amount should not cause too big of a problem in our analysis.

```
summary(df)
```

```
   Country             Attitude       Prevalence         Law       
 Length:163         Min.   : 0.00   Min.   : 1.90   Min.   :0.250  
 Class :character   1st Qu.: 8.60   1st Qu.:18.30   1st Qu.:0.500  
 Mode  :character   Median :22.05   Median :24.60   Median :0.750  
                    Mean   :27.52   Mean   :28.96   Mean   :0.592  
                    3rd Qu.:42.52   3rd Qu.:35.00   3rd Qu.:0.750  
                    Max.   :92.10   Max.   :85.00   Max.   :1.000  
                    NA's   :11      NA's   :34                     
 CountryName         Continent           Alcohol_pc    
 Length:163         Length:163         Min.   : 0.003  
 Class :character   Class :character   1st Qu.: 2.345  
 Mode  :character   Mode  :character   Median : 5.785  
                                       Mean   : 6.049  
                                       3rd Qu.: 9.332  
                                       Max.   :15.090  
                                       NA's   :3       
 poverty_headcount_ratio
 Min.   : 0.60          
 1st Qu.:15.10          
 Median :23.50          
 Mean   :27.98          
 3rd Qu.:40.00          
 Max.   :76.80          
 NA's   :24             
```

Then, we further visualized the rest of the statistics using histograms, bar charts and scatter plots.

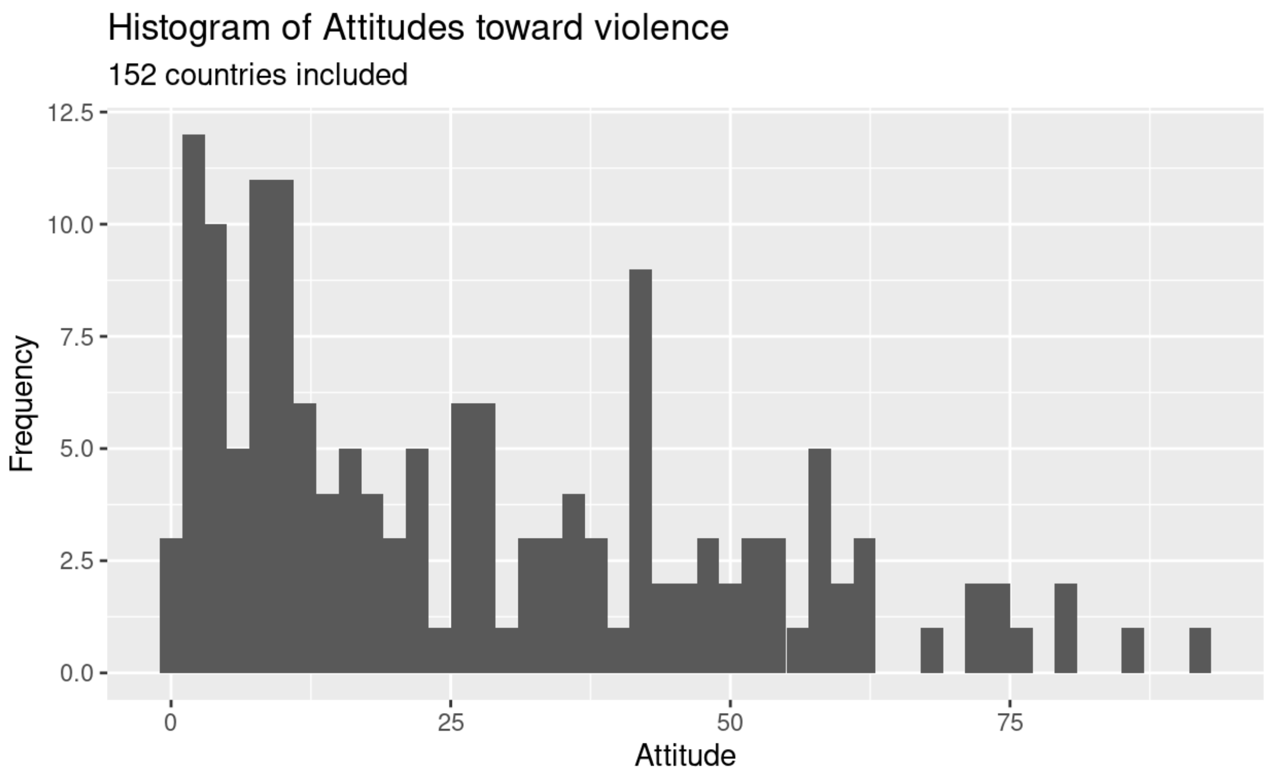
Analysis and Interpretation of Results

Step 1 - Exploratory Analysis

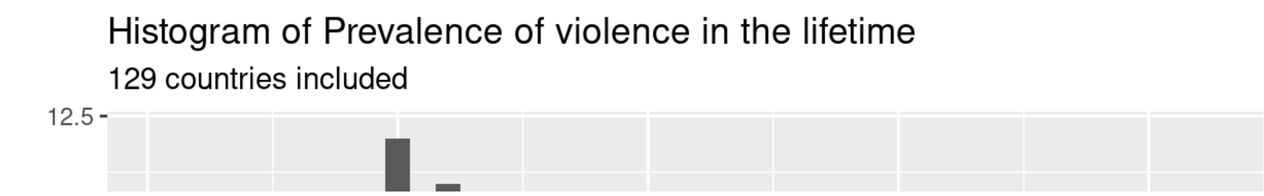
Global Level

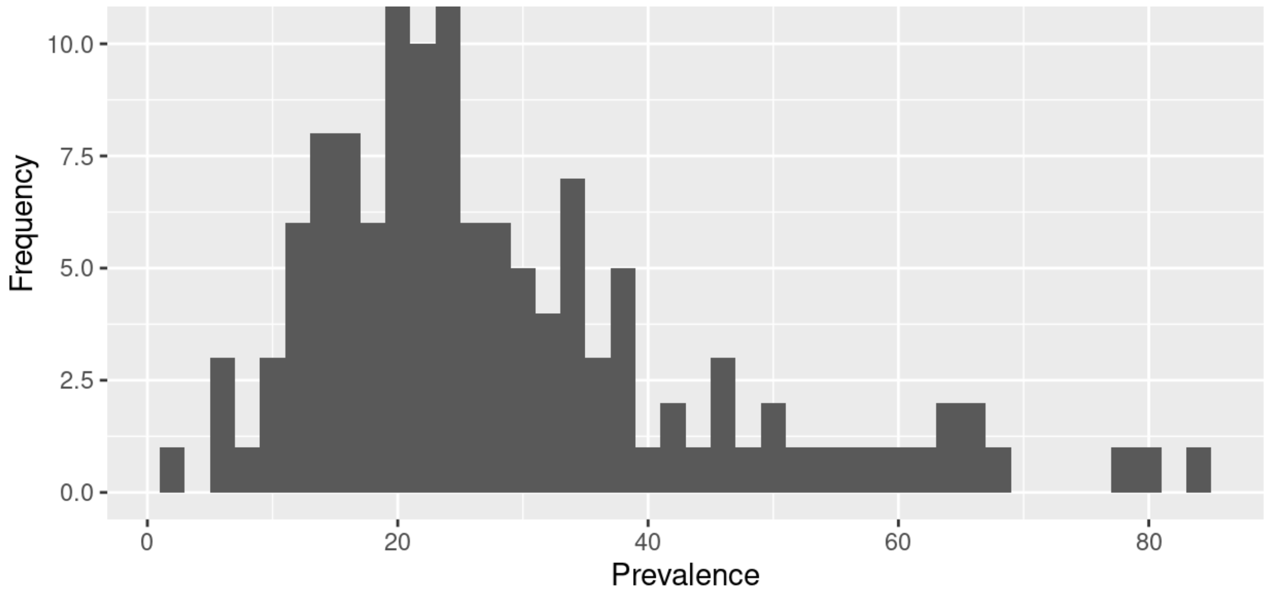
We first visualized the three indicators one by one. We saw that Attitude and Prevalence were both positively skewed, meaning that their median and mean were positioned to the left of the distribution. According to `summary(df)`, their mean and median were around 25, which is low and therefore better, since it means that attitudes towards VaW are largely oppositional and the prevalence is often low.

```
df %>%
  ggplot(aes(x = Attitude)) +
    geom_histogram(binwidth=2) +
    labs(
      title = 'Histogram of Attitudes toward violence',
      subtitle = '152 countries included',
      x = 'Attitude',
      y = 'Frequency'
    )
```



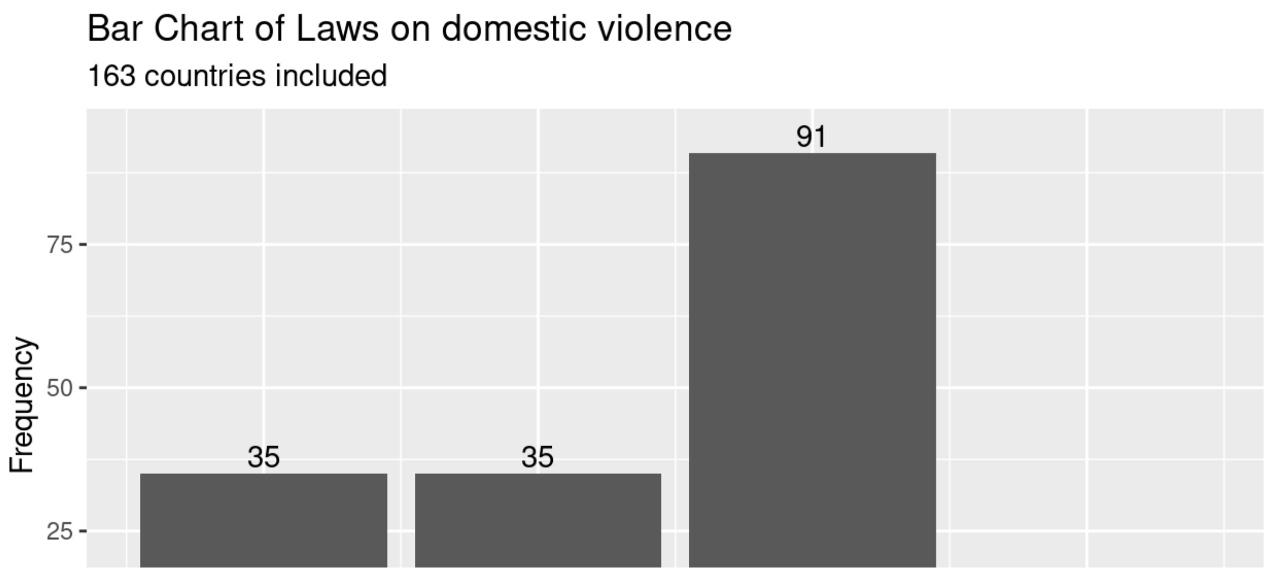

```
df %>%
  ggplot(aes(x = Prevalence)) +
  geom_histogram(binwidth=2) +
  labs(
    title = 'Histogram of Prevalence of violence in the lifetime',
    subtitle = '129 countries included',
    x = 'Prevalence',
    y = 'Frequency'
  )
```

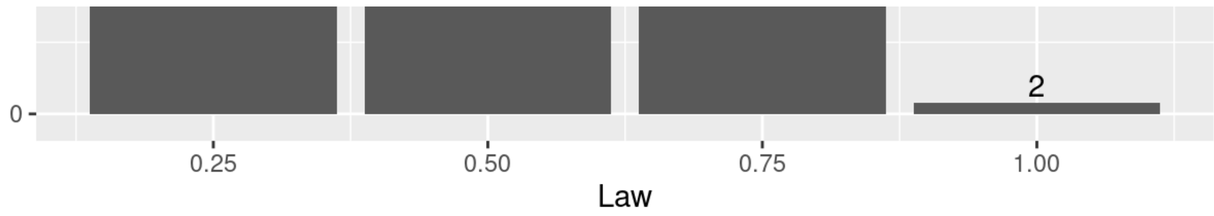




Law, on the other hand, was negatively skewed. Its mean and median were around 0.6 to 0.7, which was worse than Prevalence and Attitude.

```
df %>%
  ggplot ( aes ( x = Law ) ) +
  geom_bar( ) +
  scale_x_continuous( breaks = c ( 0.25 , 0.5 ,
0.75 , 1.00 ) ) +
  geom_text(
      aes ( label= after_stat( count ) ) ,
      stat = 'count' ,
      nudge_y = 3 ,
      va = 'bottom'
  ) +
  labs (
    title = 'Bar Chart of Laws on domestic violence',
    subtitle = '163 countries included',
    x = 'Law' ,
    y = 'Frequency'
  )
```

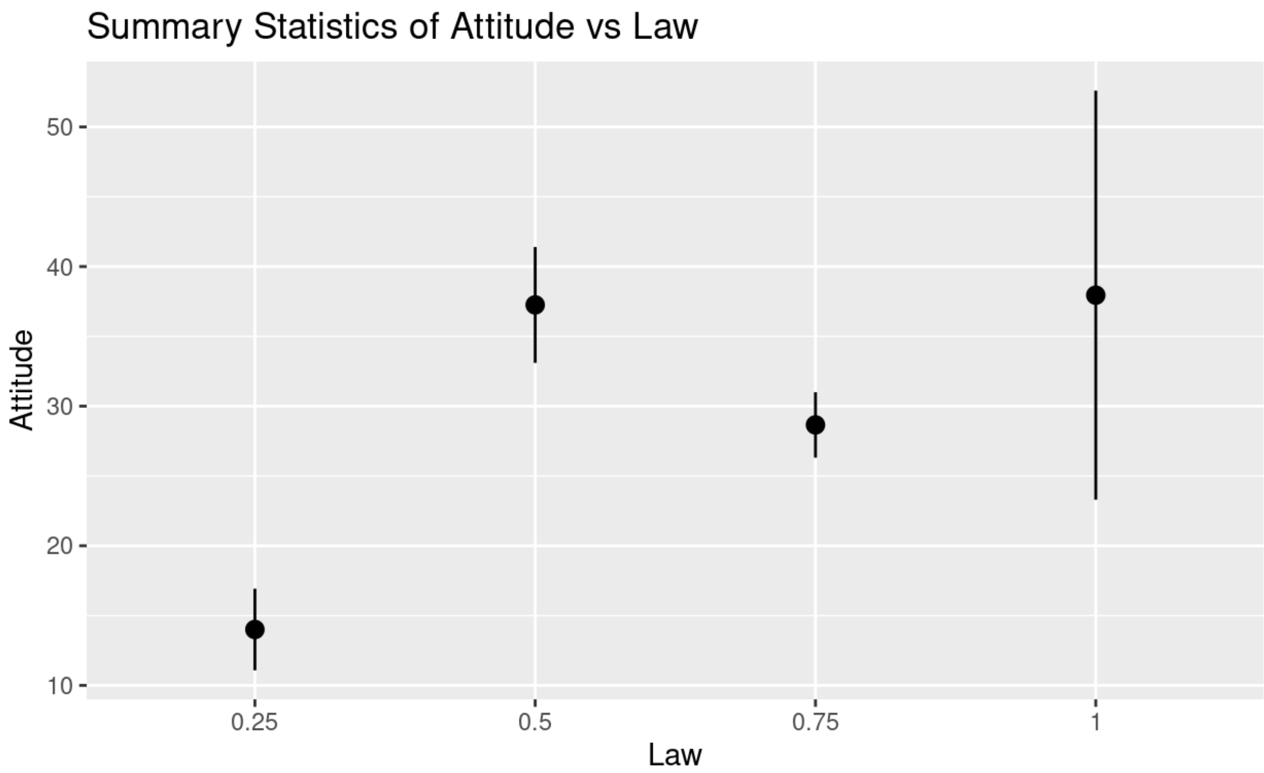




There were also 2 peculiar points we noticed about this indicator:

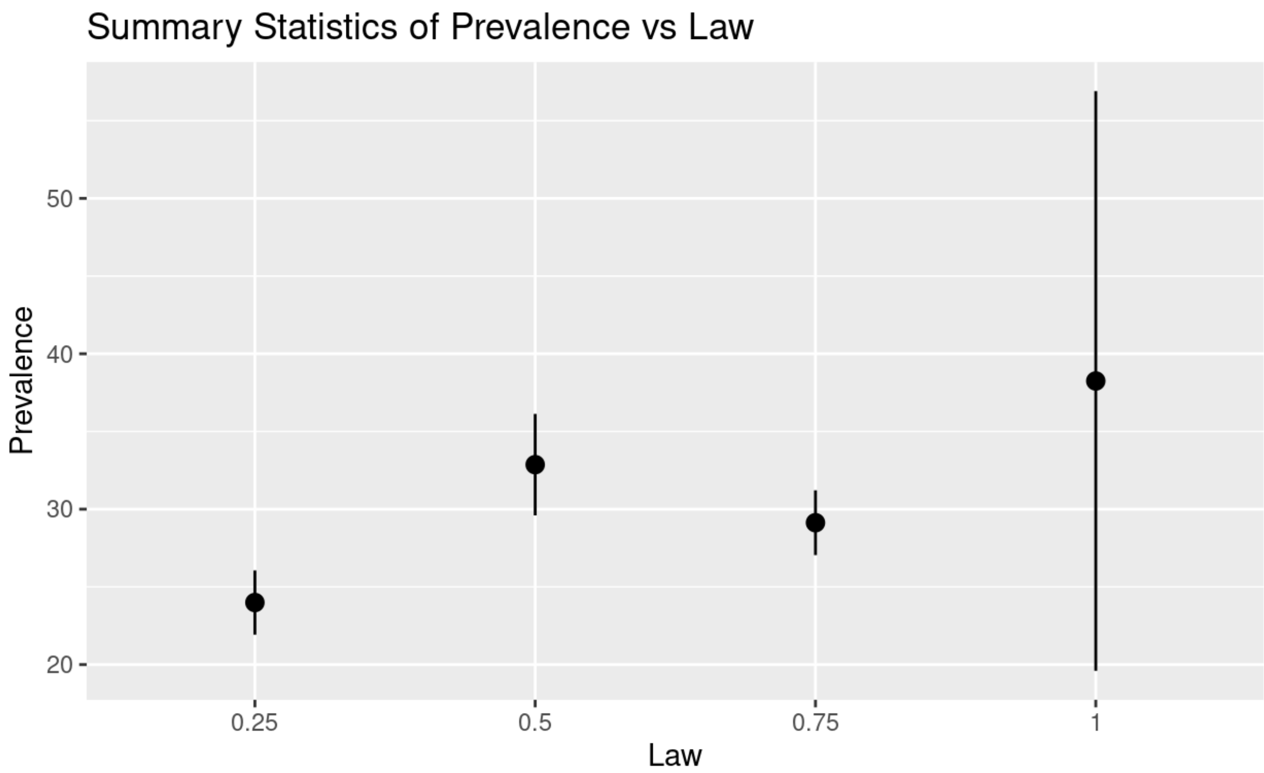
- First, it was categorical rather than continuous. It only had 4 values, so we plotted a bar chart instead of a histogram.
- Second, there was a data imbalance problem because the group where Law = 1.0 only had 2 data points. As demonstrated below, this resulted in a greater degree of uncertainty and a larger error bar.

```
df %>%
  ggplot ( aes ( x = as.factor( Law ) , y = Attitude) ) +
   stat_summary( ) +
  labs (
    title = 'Summary Statistics of Attitude vs Law',
    x = 'Law' ,
    y = 'Attitude'
  )
```




```
df %>%
  ggplot ( aes ( x = as.factor( Law ) , y = Prevalence) ) +
```

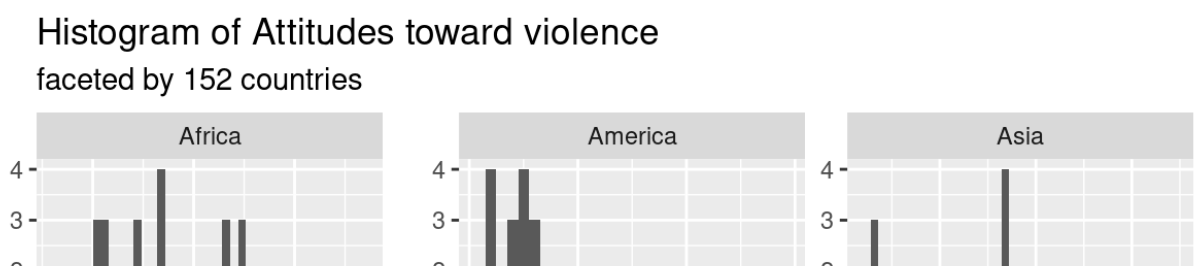
```
  stat_summary() +
  labs(
    title = 'Summary Statistics of Prevalence vs Law',
    x = 'Law',
    y = 'Prevalence'
  )
```

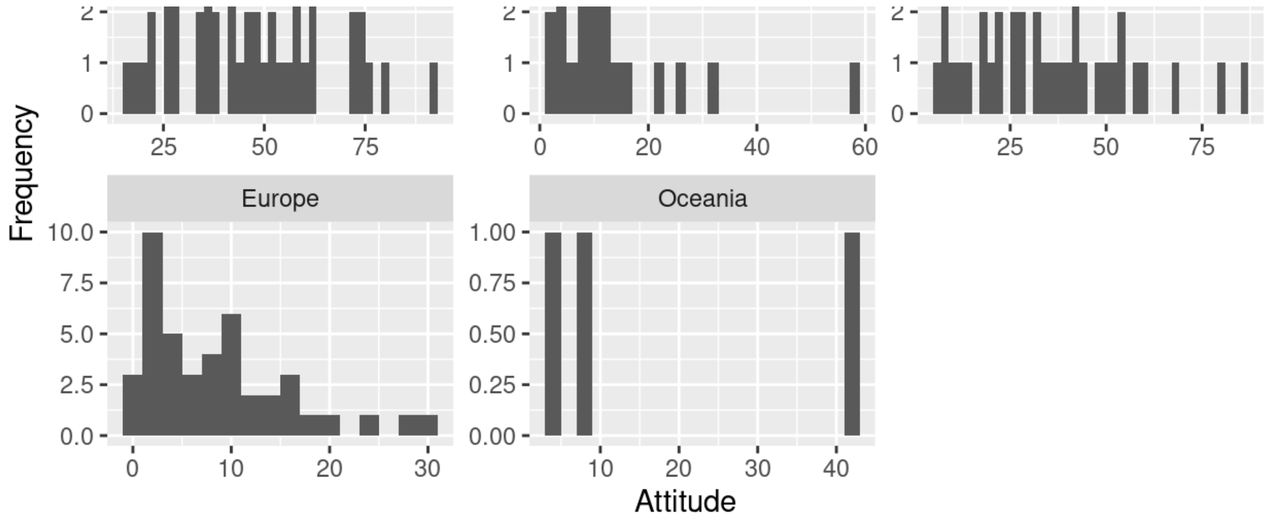


Continent Level

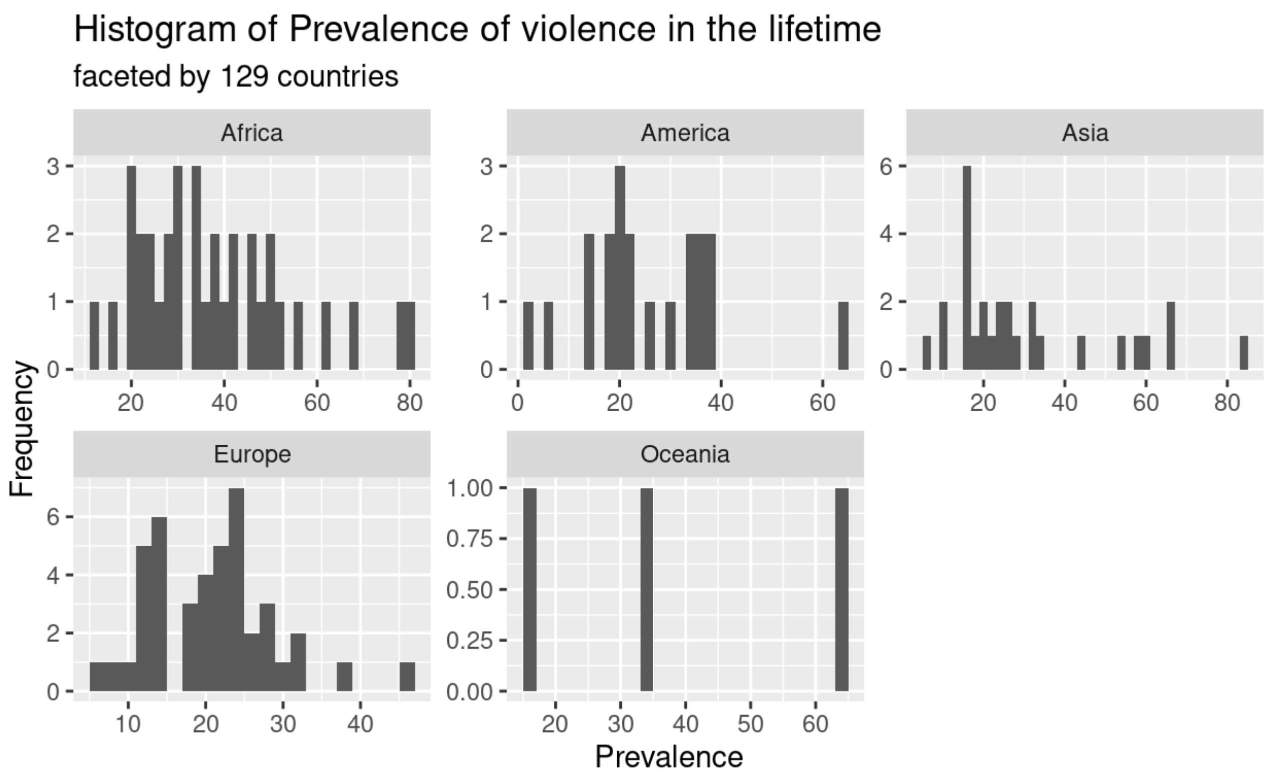
Next, we looked at the three indicators across continent.

```
df %>%
  ggplot(aes(x = Attitude)) +
  facet_wrap(. ~ Continent, scales = "free")
+
  geom_histogram(binwidth=2) +
  labs(
    title = 'Histogram of Attitudes toward violence',
    subtitle = 'faceted by 152 countries',
    x = 'Attitude',
    y = 'Frequency'
  )
```





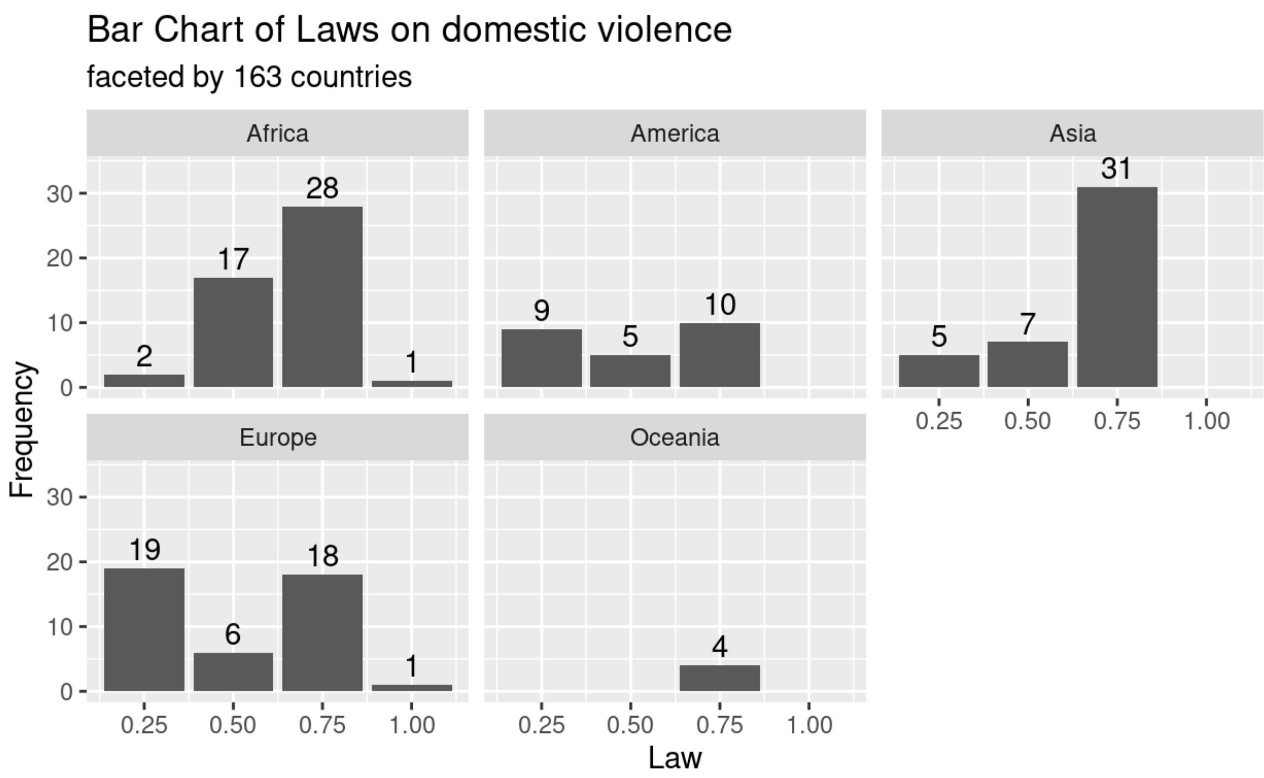

```
df %>%
  ggplot ( aes ( x = Prevalence) ) +
  facet_wrap( . ~ Continent, scales = "free" )
+
  geom_histogram( binwidth= 2 ) +
  labs (
    title = 'Histogram of Prevalence of violence in the lifetime',
    subtitle = 'faceted by 129 countries',
    x = 'Prevalence',
    y = 'Frequency'
  )
```



Again, attitude and prevalence had a similar distribution across continents. Africa and Asia had a wider range, meaning that they had more discrepancy among their countries. America and Europe, in comparison, were clustered to the left and had a lower mean and median, which means their situation regarding VaW was better than the rest of the continents.

Law, on the other hand, had a more interesting distribution.

```
df %>%
  ggplot ( aes ( x = Law ) ) +
  facet_wrap( . ~ Continent) +
  geom_bar ( ) +
  scale_x_continuous( breaks = c ( 0.25 , 0.5 ,
0.75 , 1.00 ) ) +
  geom_text(
      aes ( label= after_stat( count ) ) ,
      stat = 'count' ,
      nudge_y = 3 ,
      va = 'bottom'
  ) +
  labs (
    title = 'Bar Chart of Laws on domestic violence',
    subtitle = 'faceted by 163 countries',
    x = 'Law' ,
    y = 'Frequency'
  )
```

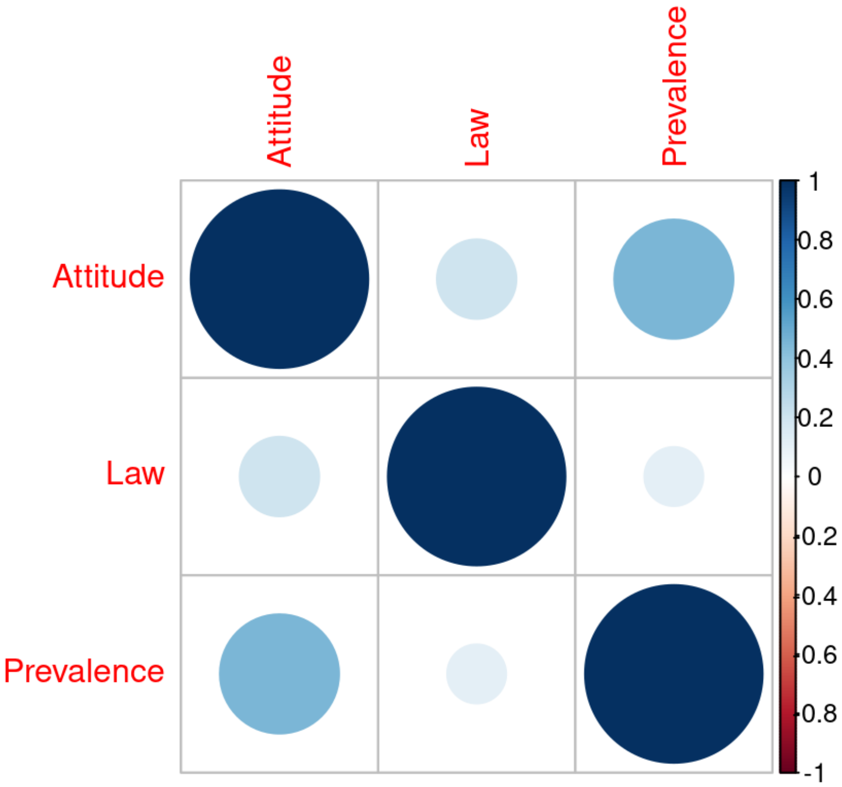


While Africa and Asia were consistent with the global distribution, America and Europe had a bimodal distribution with an almost 50/50 split between 0.25 and 0.75. Therefore, a large proportion of countries in America and Europe either had mostly comprehensive legislation protecting women against violence, or had very little legal protection. A potential explanation for America is that the data does not distinguish between North and South America, so it is possible that North America contributes more to Law = 0.25.

Despite that, we still saw a nice positive correlation between the 2 indicators.

Despite that, we still saw a nice positive correlation between the 3 indicators.

```
cormat <- round(cor(df[c('Attitude', 'Law', 'Prevalence')] %>% drop_na()), 2)
corrplot(cormat,
         # method = 'color',
         # order = 'alphabet'
  )
```



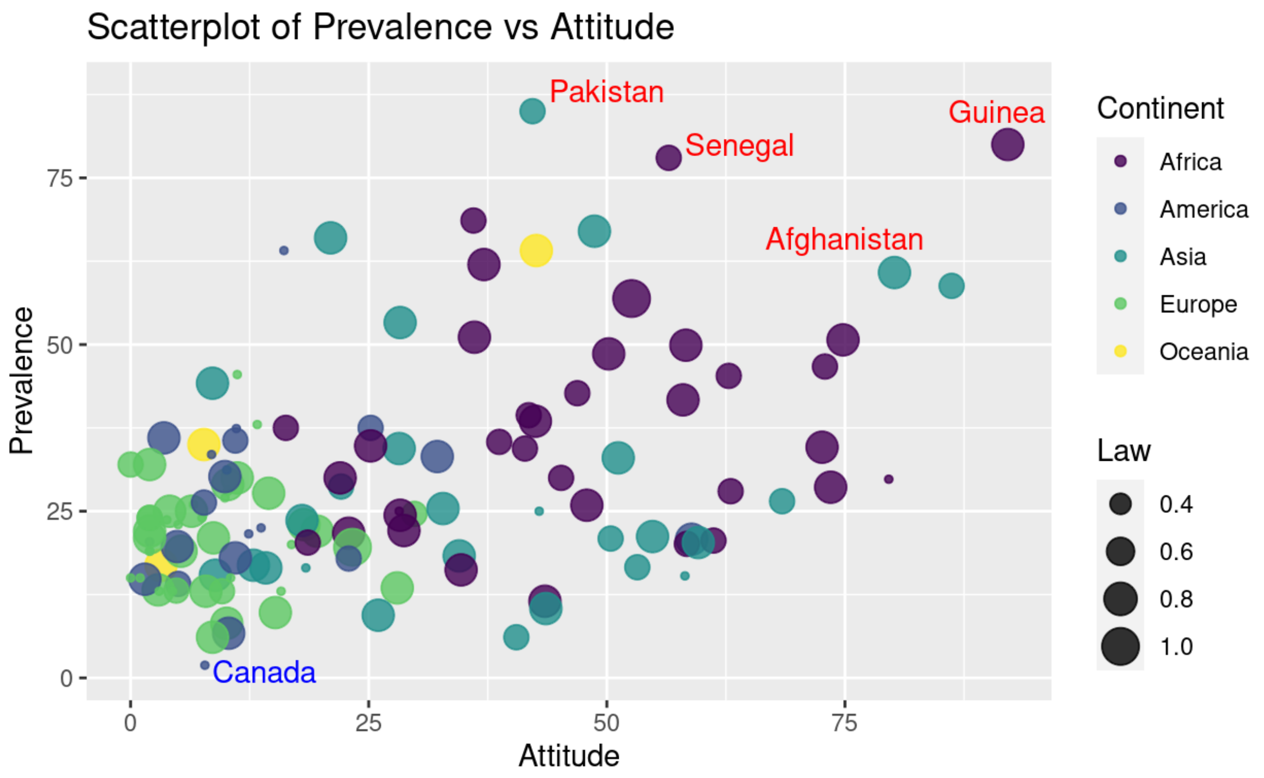
It was not too surprising that Law has a lower correlation with Attitude and Prevalence because it is discrete and due to its bimodal distribution explained above, while Attitude and Prevalence has a much larger correlation.

Country Level

Next, we put all three indicators and the continent information together in one scatterplot. We added some annotations but not too many because the points were clustered in the bottom left corner and too many labels would obscure the data points. We added 5 labels to illustrate that countries like Guinea, Pakistan, Senegal and Afghanistan are doing much worse in terms of VaW, while others like Canada are leading the way in discouraging and preventing VaW.

```
df %>%
  ggplot(aes(x = Attitude,
             y = Prevalence,
             color = Continent,
             size = Law,
             label = CountryName)) +
```

```
  geom_point(        alpha =         0.8       )         +
  annotate(        "text"   , x =        91        , y =        85        , label =
'Guinea' , color=        'red'    )         +
  annotate(        "text"   , x =        50        , y =        88        , label =
'Pakistan', color=        'red'    )         +
  annotate(        "text"   , x =        64        , y =        80        , label =
'Senegal', color=        'red'    )         +
  annotate(        "text"   , x =        75        , y =        66        , label =
'Afghanistan', color=        'red'    )         +
  annotate(        "text"   , x =        14        , y =        1         , label =
'Canada' , color=        'Blue'   )         +
  # geom_text(size = 3,
  #           family = "Times New Roman",
  #           hjust=0,
  #           vjust=-1) +
  labs    (
    title =        'Scatterplot of Prevalence vs Attitude',
    x =        'Attitude',
    y =        'Prevalence',
    color =        'Continent'
  )       +
  scale_colour_viridis_d(        )
```

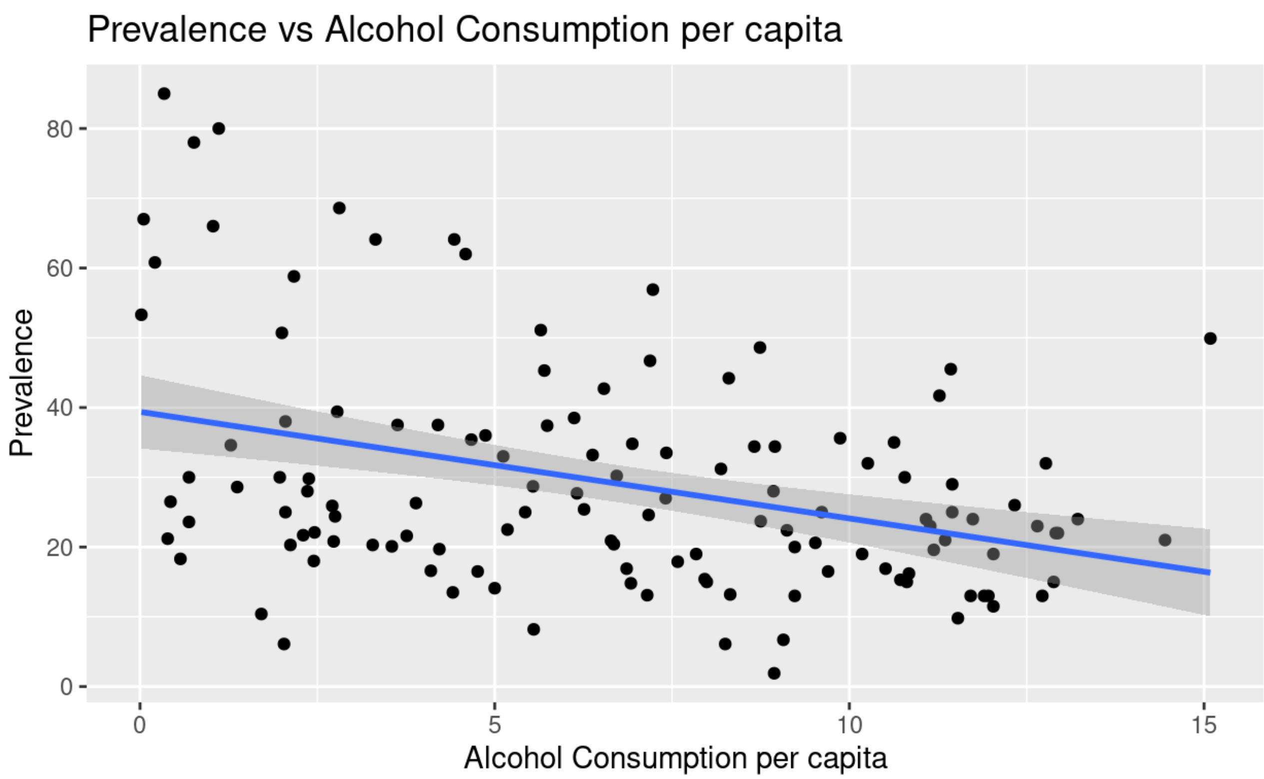


Step 2 - Regression

We were expecting a positive correlation between alcohol consumption and prevalence, but our regression told us the exact opposite.

```
df        %>%
```

```
ggplot(aes(x = Alcohol_pc, y = Prevalence)
) +
  geom_point() +
  stat_smooth(method="lm") +
  labs(
    title = 'Prevalence vs Alcohol Consumption per capita',
    x = 'Alcohol Consumption per capita',
    y = 'Prevalence'
  )
```



We saw a clear negative slope in the scatterplot. To be more precise, we ran a univariate linear regression.

```
reg_uni <- brm(formula = Prevalence ~ Alcohol_pc,
               data= df,
               refresh = 0,
               seed = 123)
# stabilize the outcome for reproducibility
summary(reg_uni)
```

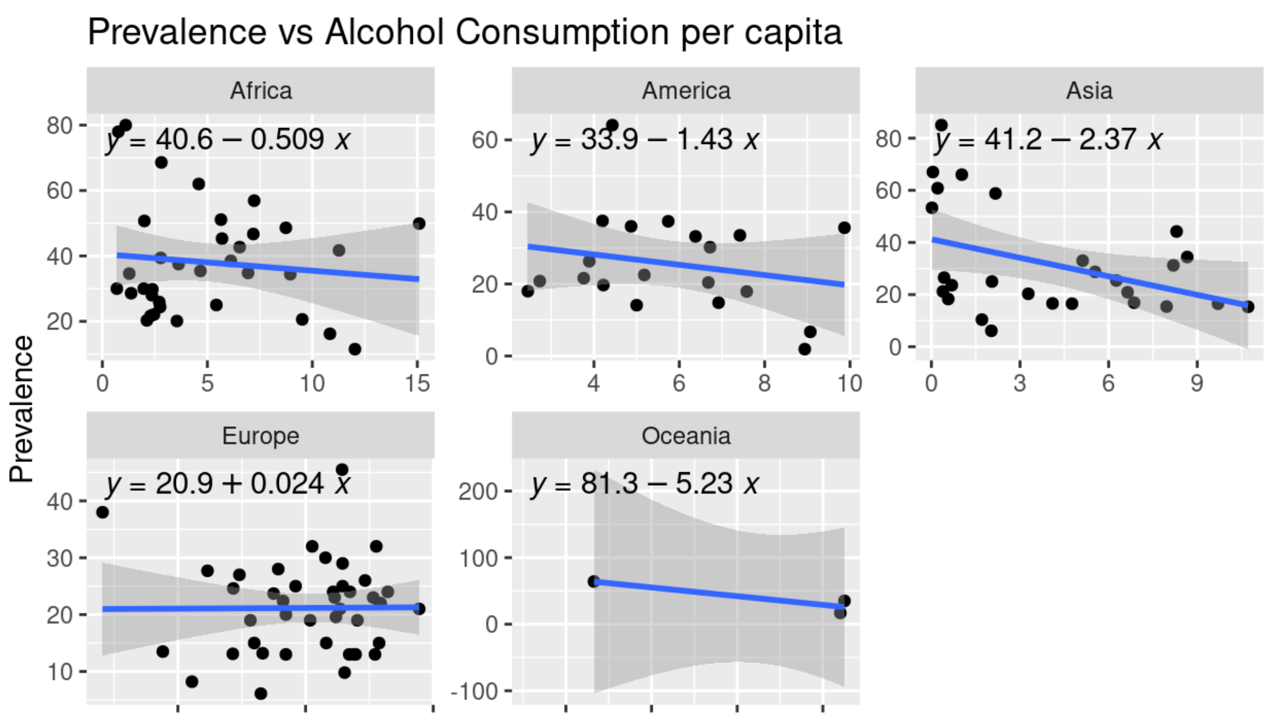
```
 Family: gaussian
  Links: mu = identity; sigma = identity
Formula: Prevalence ~ Alcohol_pc
   Data: df (Number of observations: 128)
  Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup draws = 4000

Population-Level Effects:
           Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
Intercept     39.29      2.65    34.01    44.47 1.00     3744
Alcohol_pc    -1.52      0.35    -2.21    -0.83 1.00     3758
           Tail_ESS
```

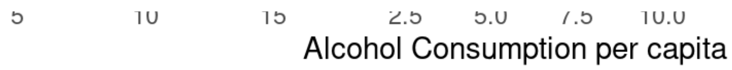
```
          Tail_ESS
Intercept     2797
Alcohol_pc    2639

Family Specific Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma    15.26      0.94    13.52    17.28 1.00     3812     2794

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

Our best estimate shows that a one liter increase in total alcohol consumption was associated with 1.52 percentage point decrease in prevalence. Even with the uncertainty interval ranging from -2.21 to -0.83, we were still confident that the slope was negative. This directly contradicted our hypothesis.

Next, we looked at the scatter plot again, but faceted it by continent, to see if the same trend was true on a continent-level.

```
df %>%
  ggplot(aes(x = Alcohol_pc, y = Prevalence)) +
  geom_point() +
  stat_smooth(method="lm") +
  stat_poly_eq(parse=T, aes(label = ..eq.label..), formula=y ~ x) +
  facet_wrap(. ~ Continent, scales = "free") +
  labs(
    title = 'Prevalence vs Alcohol Consumption per capita',
    x = 'Alcohol Consumption per capita',
    y = 'Prevalence'
  )
```





We can see that except for a slight positive slope in Europe, all slopes are negative. Asia and Oceania seemed to contribute the most to the negative association in the global trend. This led us to question if there might have been a confounding variable causing this spurious association.

We chose poverty as our target and used poverty headcount ratio as the indicator for poverty. This indicator showed the percentage of each country's population that was living below the national poverty line. We thought that an indicator specific to each country was better than a global indicator because it took into account the different definitions of poverty that each country has.

```
reg_multi <- brm ( formula = Prevalence ~ Alcohol_pc
+ poverty_headcount_ratio,
                     data= df ,
                     refresh = 0 ,
                     seed = 123 )
# stabilize the outcome for reproducibility
summary ( reg_multi)
```

```
 Family: gaussian
  Links: mu = identity; sigma = identity
Formula: Prevalence ~ Alcohol_pc + poverty_headcount_ratio
   Data: df (Number of observations: 118)
  Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup draws = 4000

Population-Level Effects:
                        Estimate Est.Error l-95% CI u-95% CI Rhat
Intercept                  26.59      4.26    18.13    35.10 1.00
Alcohol_pc                 -0.96      0.37    -1.65    -0.22 1.00
poverty_headcount_ratio     0.33      0.09     0.16     0.50 1.00
                        Bulk_ESS Tail_ESS
Intercept                   3212     2655
Alcohol_pc                  3767     3002
poverty_headcount_ratio     3275     3185

Family Specific Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma    14.33      0.95    12.61    16.38 1.00     3850     2786

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

We add the poverty headcount and ran the regression again. This time, our best estimate showed that one liter increase in total alcohol consumption was associated with 0.96 percentage point decrease in prevalence, with an uncertainty interval ranging from -1.65 to -0.22.

Compared to the earlier estimate, including poverty headcount ratio reduced the size of the association by roughly a third. Therefore, the original association can be partially explained by the inclusion of poverty headcount ratio as a control variable.

inclusion of poverty_headcount_ratio as a control variable.

Step 3 - Posterior Prediction

We used the multilinear regression model to do a prediction of the prevalence indicator. We chose the data for China because it was missing a prevalence indicator and it is of interest.

```
predict_data <- tibble(Alcohol_pc = 7.05,
                       poverty_headcount_ratio = 0.6)
mod_pred <- posterior_epred(reg_multi,
                            newdata= predict_data)
mod_pred <- as_tibble(mod_pred)
summary(mod_pred$V1)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  10.90   18.24   20.03   20.05   21.73   32.78 
```

The most likely mean is 20.05 and the most likely medium is 20.03. This puts China at around the mean and medium of the global level. A plausible uncertainty interval is from 18.24 to 32.78, which is not too big of an interval.

Step 4 - Investigation of Gender

As established in the literature review, one limitation of prior research into domestic violence and alcohol use is the lack of distinction made between which genders are using alcohol in instances of VaW, and whether alcohol is being used by the victim or the perpetrator. After finding this pattern in the research, we decided to see if the relationships between alcohol use, prevalence of VaW and poverty headcount ratio differed between men and women.

Join Data

First, we imported and joined the new datasets with the dataframe we created so far. We only joined the data from 2018 since it was the most recent and most complete column.

```
MaleAlcoholUse <- read_csv('./Data/MaleAlcoholUse.csv')
head(MaleAlcoholUse)
```

```
# A tibble: 6 × 3
   ...1 `Country Code` `2018`
  <dbl> <chr>           <dbl>
1     1 ABW             NA   
2     2 AFE              8.50
3     3 AFG              0.36
4     4 AFW             11.0 
5     5 AGO             11.0 
```

```
6      6 ALB             11.6
```

```
colnames(MaleAlcoholUse) <- c('X1', 'Country', 'AlcoholM_pc')
MaleAlcoholUse <- MaleAlcoholUse[c('Country', 'AlcoholM_pc')]
```

```
FemaleAlcoholUse <- read_csv('./Data/FemaleAlcoholUse.csv')
head(FemaleAlcoholUse)
```

```
# A tibble: 6 × 3
   ...1 `Country Code` `2018`
  <dbl> <chr>           <dbl>
1     1 ABW            NA
2     2 AFE             1.97
3     3 AFG             0.055
4     4 AFW             2.69
5     5 AGO             3.1
6     6 ALB             2.68
```

```
colnames(FemaleAlcoholUse) <- c('X1', 'Country', 'AlcoholF_pc')
FemaleAlcoholUse <- FemaleAlcoholUse[c('Country', 'AlcoholF_pc')]
```

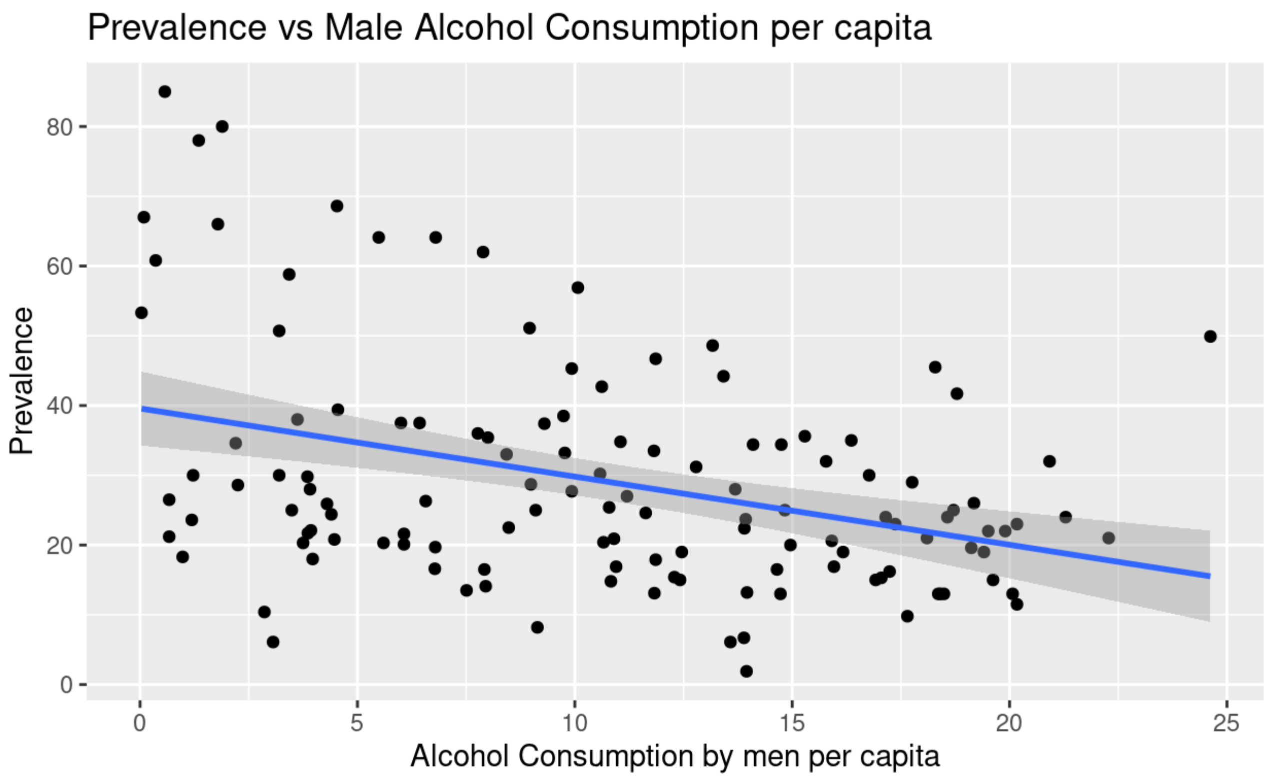
```
df <- left_join(df, FemaleAlcoholUse, by = 'Country')
df <- left_join(df, MaleAlcoholUse, by = 'Country')
head(df)
```

```
  Country Attitude Prevalence  Law CountryName Continent Alcohol_pc
1     AUS      3.2       16.9 0.75   Australia   Oceania      10.51
2     CAN      7.8        1.9 0.25      Canada   America       8.94
3     FIN     11.2       30.0 0.75     Finland    Europe      10.78
4     FRA      6.6       26.0 0.25      France    Europe      12.33
5     DEU     19.6       22.0 0.75     Germany    Europe      12.91
6     HUN      8.7       21.0 0.75     Hungary    Europe      11.35
  poverty_headcount_ratio AlcoholF_pc AlcoholM_pc
1                      NA        5.18       15.96
2                      NA        4.07       13.95
3                    12.2        5.05       16.77
4                    13.6        5.83       19.18
5                    14.8        6.21       19.90
6                    12.3        5.38       18.10
```

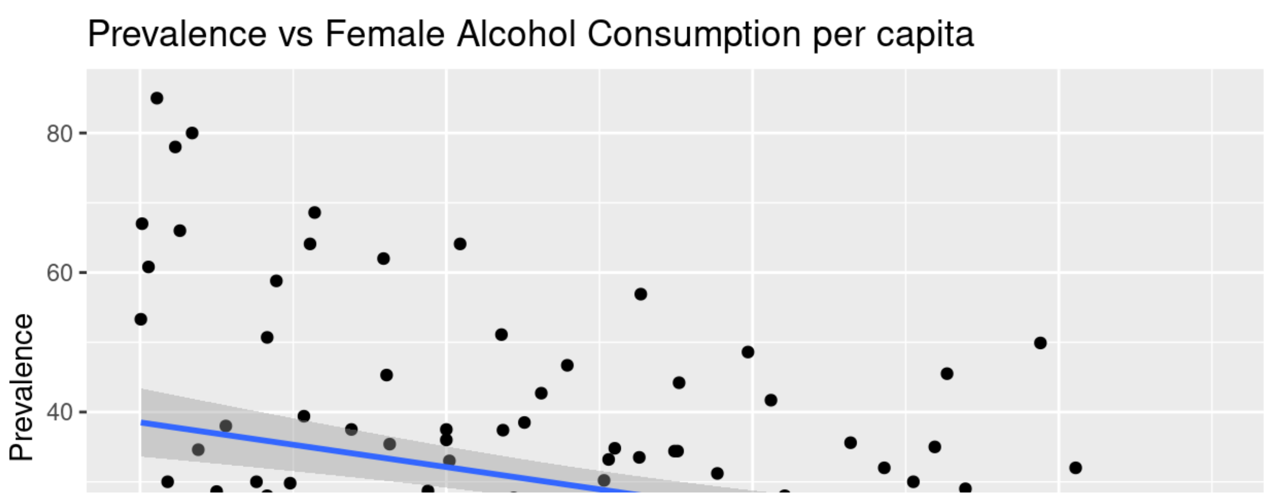
Next, we created scatterplots comparing prevalence and alcohol consumption for both genders.

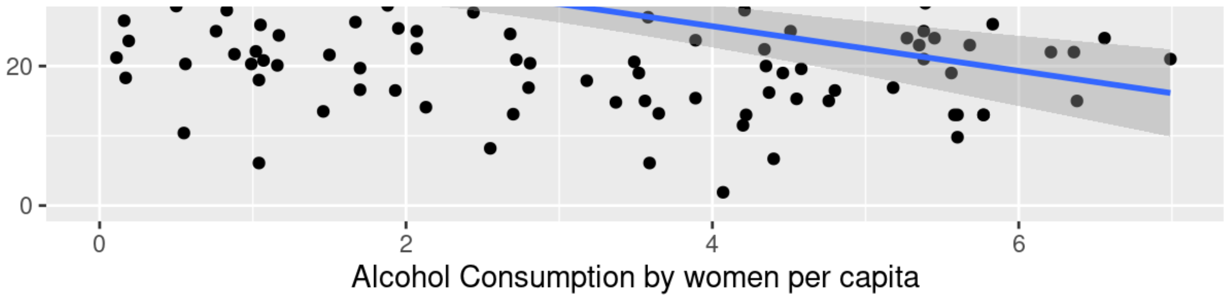
```
df %>%
  ggplot(aes(x = AlcoholM_pc, y = Prevalence)) +
```

```
  geom_point(        )        +
  stat_smooth(        method=        "lm"     )        +
  labs     (
    title =        'Prevalence vs Male Alcohol Consumption per capita',
    x =        'Alcohol Consumption by men per capita',
    y =        'Prevalence'
  )
```




```
df        %>%
  ggplot  (        aes      (        x =        AlcoholF_pc, y =        Prevalence)
)     +
  geom_point(        )        +
  stat_smooth(        method=        "lm"     )        +
  labs     (
    title =        'Prevalence vs Female Alcohol Consumption per capita',
    x =        'Alcohol Consumption by women per capita',
    y =        'Prevalence'
  )
```





Univariate Regressions

As expected, for both genders there is a similar negative correlation to what we found earlier with total alcohol consumption. To further demonstrate this, we ran a univariate regression for both genders.

```
reg_uni3 <- brm ( formula = Prevalence ~ AlcoholM_pc,
                  data= df ,
                  refresh = 0 ,
                  seed = 123 )
# stabilize the outcome for reproducibility
summary ( reg_uni3 )
```

```
 Family: gaussian
  Links: mu = identity; sigma = identity
Formula: Prevalence ~ AlcoholM_pc
   Data: df (Number of observations: 128)
  Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup draws = 4000

Population-Level Effects:
            Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
Intercept      39.55      2.67    34.25    44.68 1.00     3734
AlcoholM_pc    -0.98      0.22    -1.40    -0.55 1.00     3825
            Tail_ESS
Intercept       2808
AlcoholM_pc     3026

Family Specific Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma    15.26      0.95    13.54    17.30 1.00     3541     3005

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

```
reg_uni2 <- brm ( formula = Prevalence ~ AlcoholF_pc,
                  data= df ,
                  refresh = 0 ,
                  seed = 123 )
# stabilize the outcome for reproducibility
summary ( reg_uni2 )
```

```
 Family: gaussian
  Links: mu = identity; sigma = identity
Formula: Prevalence ~ AlcoholF_pc
   Data: df (Number of observations: 128)
  Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup draws = 4000

Population-Level Effects:
            Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
Intercept      38.49      2.47    33.63    43.36 1.00     4284
AlcoholF_pc    -3.20      0.70    -4.55    -1.83 1.00     4231
            Tail_ESS
Intercept       3132
AlcoholF_pc     2926

Family Specific Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma    15.22      0.95    13.52    17.30 1.00     3460     2749

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

Our best estimate shows that for men, a one liter increase in total alcohol consumption was associated with 0.98 percentage point decrease in prevalence. For women, the percentage decrease was 3.20.

Multivariate Regressions

Just as before, this association between more alcohol consumption and lower prevalence of VaW is counter-intuitive so we once again assessed if poverty is a confounding variable by running a multivariate regression analysis.

```
reg3 <- brm ( formula = Prevalence ~ AlcoholM_pc
+ poverty_headcount_ratio,
              data= df ,
              refresh = 0 ,
              seed = 123 )
# stabilize the outcome for reproducibility
summary ( reg3 )
```

```
 Family: gaussian
  Links: mu = identity; sigma = identity
Formula: Prevalence ~ AlcoholM_pc + poverty_headcount_ratio
   Data: df (Number of observations: 118)
  Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup draws = 4000

Population-Level Effects:
                        Estimate Est.Error l-95% CI u-95% CI Rhat
Intercept                  26.77      4.27    18.50    35.16 1.00
AlcoholM_pc                -0.61      0.23    -1.07    -0.17 1.00
poverty_headcount_ratio     0.33      0.09     0.16     0.51 1.00
```

```
poverty_headcount_ratio     0.33      0.09     0.16     0.51 1.00
                        Bulk_ESS Tail_ESS
Intercept                   3212     2611
AlcoholM_pc                 3542     2431
poverty_headcount_ratio     3439     2736

Family Specific Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma    14.31      0.97    12.52    16.34 1.00     4237     3139

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

```
reg4     <-        brm      (        formula =        Prevalence ~         AlcoholF_pc
+         poverty_headcount_ratio,
                      data=        df         ,
                      refresh =        0         ,
                      seed =        123       )
# stabilize the outcome for reproducibility
summary (        reg4      )
```

```
 Family: gaussian
  Links: mu = identity; sigma = identity
Formula: Prevalence ~ AlcoholF_pc + poverty_headcount_ratio
   Data: df (Number of observations: 118)
  Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup draws = 4000

Population-Level Effects:
                        Estimate Est.Error l-95% CI u-95% CI Rhat
Intercept                  25.99      4.13    17.72    33.82 1.00
AlcoholF_pc                -1.94      0.76    -3.43    -0.40 1.00
poverty_headcount_ratio     0.33      0.09     0.16     0.50 1.00
                        Bulk_ESS Tail_ESS
Intercept                   3080     2743
AlcoholF_pc                 3639     2987
poverty_headcount_ratio     3758     3066

Family Specific Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma    14.38      0.99    12.61    16.45 1.00     3939     2651

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

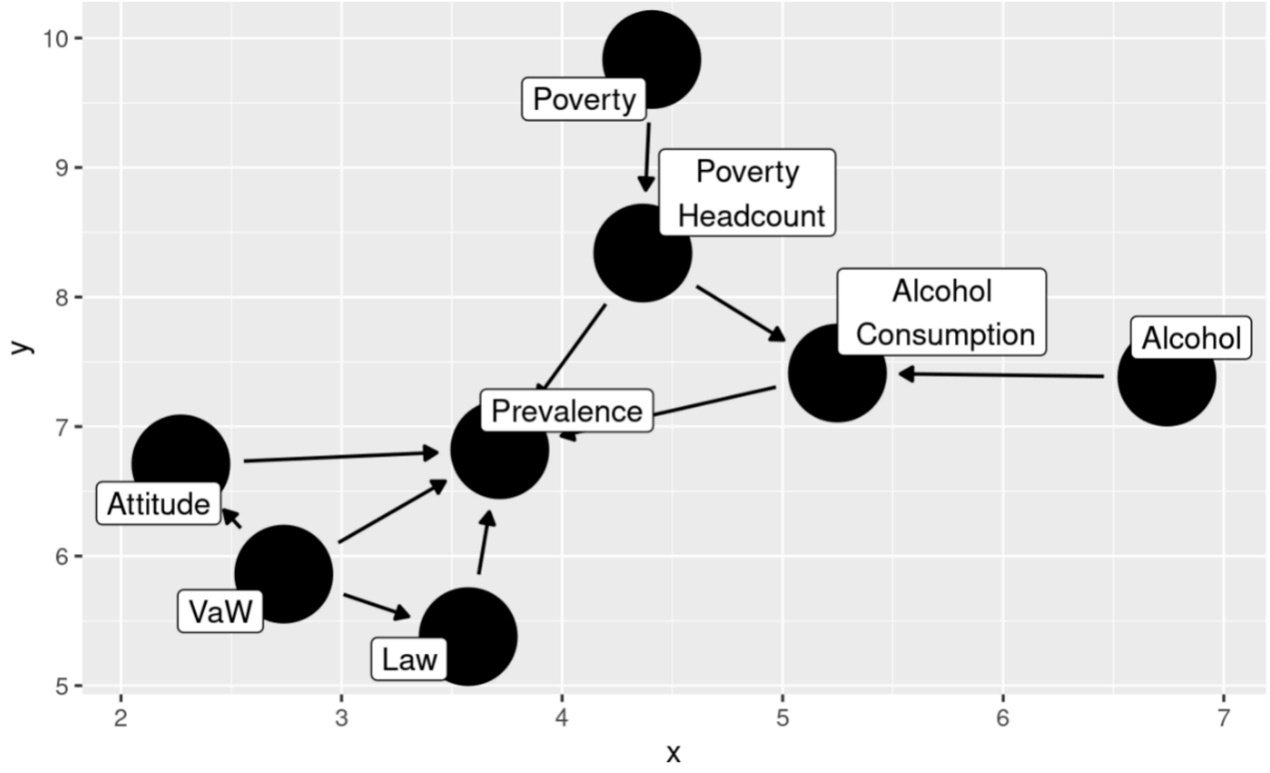
When adding poverty headcount ratio into the regression analyses, the percentage point decrease in prevalence for men went from 0.98 to 0.62, and for women the percentage dropped from 3.20 to 1.93. For both genders, just as with alcohol consumption overall, accounting for poverty headcount ratio reduced the size of the association by roughly a third.

Discussion

Discussion

We drew a causal diagram to summarize the confounding relationships we investigated.

```
vaw_dag <- dagify( Law ~ vaw ,
       Attitude ~ vaw ,
       Prevalence ~ vaw + ph + ac +
Law + Attitude,
       ac ~ a + ph ,
       ph ~ p ,
       labels = c ( "vaw" = "VaW" ,
                   "Law" = "Law" ,
                   "Attitude" = "Attitude",
                   "Prevalence" = "Prevalence",
                   "ph" = "Poverty\n Headcount",
                   "p" = "Poverty",
                   "ac" = "Alcohol\n Consumption",
                   "a" = "Alcohol"
                   ) ,
       latent = "vaw" )

ggdag ( vaw_dag , text = FALSE , use_labels = "label" )
```



As we see it, there are three key layers to this diagram to keep in mind when understanding the causal relationships involved here. Firstly, we start with a major social issue: VaW, alcohol usage and poverty. Secondly, these large-scale problems are narrowed down into indicators that attempt to track and record these phenomena. In this report, these indicators are attitudes and laws, alcohol consumption in liters per capita, and poverty headcount ratio respectively. Finally, we see that all of these aspects narrow down into the prevalence of VaW. This may seem odd given that prevalence has been used as an indicator up until this point, similarly to attitudes and legislation. However, as we stated in the beginning, we focused exclusively on prevalence as an indicator because it served us best in looking at the effects of alcohol, given that alcohol consumption and subsequent acts of violence are small-scale and individualized phenomena, as opposed to the broader indicators of law and attitudes.

We can therefore see that incidences of VaW are affected by four key factors that have been discussed in our research: attitudes towards VaW, the extent to which legislation protects women, the amount of alcohol consumption per capita, and the extent of poverty within a given country. Additionally, poverty is a likely cause of alcohol consumption which would then go on to influence rates of VaW, hence its status within our paper as a confounding variable. While it could be argued that the relationship between alcohol consumption and poverty is bidirectional since alcohol dependence can result in poverty due to an increased risk of unemployment or losing money, even if this is the case, it's still reasonable to assert that poverty is more likely to cause alcohol consumption rather than for alcohol consumption to solely cause poverty. While it's plausible that somebody would be driven into alcoholism by being in an impoverished state, it is harder to argue that an alcohol addiction would be the only causal factor in somebody sinking into poverty. This is why we outlined the causal diagram in this way, and why we feel that poverty interacts with alcohol consumption to a great extent in influencing rates of VaW.

Our initial regression analysis counter-intuitively showed that prevalence decreases as alcohol consumption increases, so we revised our hypothesis and looked for potential confounding variables at play. The fact that there were confounding variables didn't surprise us, given all of the other factors cited within the research that affect alcohol consumption and violent behavior. Once we found that poverty was also a strong potential indicator, we thought that including it in our analysis would give a more holistic picture of the situation regarding alcohol use, and therefore produce regression values that more closely reflected our original hypothesis.

When we selected poverty as a potential confounding variable, we did so because of a similar urge that motivated us to investigate alcohol consumption: it is widely understood that poverty affects the prevalence of violence and substance abuse. Interestingly, contemporary research has mixed perspectives on this matter. Poverty is largely acknowledged as a key factor in IPV but understanding the specifics of why this is the case has proven difficult because theories attempting to explain the relationship (such as: women having more or less financial power, poverty leading to stress that leads to violence and how low-income communities as a whole treat VaW) have just as much research supporting them as contradicting them. The only consistent causal relationship established is a cultural one: the idea that being of a lower economic standing is emasculating, so poorer men are more likely to try and control women in an attempt to assert their power within a household (Jewkes 2002, 1424). However, as with alcohol, the sample sizes in existing research tend to be quite small and localized which highlights the need for a globally scaled analysis. As a result, our analysis fills in this gap in the research by establishing a link between poverty and VaW on an international scale.

Following our further regressions, the negative association between alcohol consumption and prevalence decreased by a third, demonstrating that poverty is a confounding variable. An interesting thing to note is that this result applied across both men and women as individual consumers of alcohol. This is significant because it indicates that despite more men consuming alcohol than women overall, their consumption of alcohol has the same effect as women's on the prevalence of VaW. This further confirms what we discovered when reviewing existing literature on the topic: there needs to be more comparison between alcohol use by perpetrators and by victims of VaW because it seems from our findings that alcohol consumption is significant regardless of gender and power status in a relationship. Further research would result in more specific and targeted interventions while also providing insights into the psychological and biological mechanisms that go into alcohol consumption and subsequent acts of violence.

However, we needed to keep in mind the limitations of this study. Our three VaW indicators were downloaded from a 2019 study, but the latest alcohol data we could obtain was from 2018. Additionally, the latest poverty headcount data ranged from the year 2009 to 2020 for different countries, which is why we manually cleaned the dataset before uploading the CSV file to Rstudio. The lack of more recent data undermines the credibility of our conclusion. The alcohol data were labeled as 'projected estimates' but the World Bank did not disclose how exactly they arrived at these estimates and how far they extend. In terms of the gender differences in alcohol consumption, our analysis hinges on the assumption that all perpetrators of domestic violence are men and all victims are women, which is true in the majority of cases but not always, since sometimes men are victims of female perpetrators, or the abuse dynamics within a relationship are mutual. For the poverty data, we selected the measure of poverty headcount ratio by country, but there were other data sets that measured the percentage of people earning money below specified poverty lines, such as US $3.20 or $1.90. It's possible that using this data instead would have given a more cohesive and unified picture of poverty compared to alcohol consumption (especially since the measure of total alcohol consumption is the same for every country) but we chose the measure that made the most sense for us because we valued the internationalism of our research and wanted the data to reflect that.

Conclusion

By analyzing multiple datasets using visualization and linear regression, we concluded that violence against women is an ongoing and global issue, whose degree various across different continents and countries. Focusing on one indicator of VaW, prevalence of violence in the lifetime, we found that the negative association between prevalence and alcohol consumption, can be partially explained by the inclusion of poverty headcount ratio, an indicator of poverty, as a control variable. As mentioned in our introduction, there is evidence that living in poverty increases the likelihood of VaW and alcohol consumption, so poverty might have been a confounding variable that led to the spurious relationship between VaW prevalence and alcohol consumption.

There are many further avenues for research that our analysis opens up. As mentioned in the introduction, there is also more need for distinguishing between alcohol use by perpetrators versus victims when looking at the relationship between alcohol and IPV, and this should be taken into account by those who collect data on VaW as well as those who analyze it. In terms of the indicators used, there may be different ways of measuring alcohol consumption that are more specific than liter consumption per capita. For example, surveys could be conducted to ask people how much alcohol they drink on average within certain time frames - this is just one example of another potential measure, and it would have its own issues with reliability, but it is at least one other angle from which to view worldwide alcohol consumption. Moreover, one interesting argument we found in our research is that because women's financial independence doesn't necessarily prevent violence against them -

is that because women's financial independence doesn't necessarily prevent violence against them - in fact, when a woman in a household is employed but her partner isn't, she is at greater risk of violence - "economic inequality within a context of poverty is more important than the absolute level of income or empowerment of a man or woman in a relationship" (Jewkes 2002, 1424). This could mean that broader measures of poverty such as the Gini index (a measurement of wealth inequality) might paint a different picture of how poverty affects rates of VaW. Overall, it is especially important to continue considering VaW as a global phenomenon and use globally sourced data in future research.

References

Cepeda, Isabel, Maricruz Lacalle-Calderon, and Miguel Torralba. 2021. "Measuring Violence Against Women: A Global Index." *Journal of Interpersonal Violence*, August, 088626052110374. https://doi.org/10.1177/08862605211037424.

Jewkes, Rachel. 2002. "Intimate Partner Violence: Causes and Prevention." *The Lancet* 359 (9315): 1423–29. https://doi.org/10.1016/s0140-6736(02)08357-5.

Understanding Violence Against Women. 1996. National Academies Press. https://doi.org/10.17226/5127.

White, H R. 1995. "Alcohol and Interpersonal Violence: Fostering Multidisciplinary Perspectives." *Journal of Studies on Alcohol* 56 (1): 123–24. https://doi.org/10.15288/jsa.1995.56.123.