

Project 1 Exploratory Data Analysis: Crime and Demographics in Texas Counties

Veronica Remmert

Introduction

The 'health' dataset is the 2019 Country Health Rankings and Roadmaps dataset and describes different health-related statistics in Texas counties. Variables include Texas counties by name, average life expectancy, population size for the county, and average household income. The border status variable categorizes each county as a border or non-border county, defining a county within 100 Kilometers of the U.S. and Mexico border as a border county. The urbanization status variable categorizes each county into rural or urban based on population density and other measures of dense development. The dataset also includes frequent mental distress rate which is defined as experiencing stress, depression, or emotional problems for 14 or more days in the previous 30 days, and motor vehicle and firearm fatality rates. Household income is also broken into average household income for black, white, and hispanic groups in three different variables. Segregation index calculates an index that measures how different the population composition of a subarea is from the population composition of the urban area as a whole. Lastly, the dataset includes a percentage of the number of homeowners in the county. The 'crime' dataset is a collection of information reported by local law enforcement agencies to the Texas Department of Public Safety's Uniform Crime Reporting Section. This 2013 dataset has variables that include county, crime rate, crimes reported, and overall crime ranking. Crime rate is calculated based on crimes in the county for a population of 100,000, such that some counties with small populations have high crime rate based on small population and relatively large number of crimes in the population. Also, crimes reported is the total number of crimes reported in the county. Crime ranking is based on the crime rate per county, where number one is the county with the highest crime rate. Both of these datasets were found on the internet through government websites.

The 'crime' and 'health' datasets are interesting because the environment in which someone lives can impact their health. By looking at the association between health and crime per county, there is a potential that we can more intimately understand if crimes reported and crime rates are correlated to other health-related variables, like life expectancy, mental distress rate, and various mortality rates. I am personally interested in health and crime and how they are correlated because I am interested in social determinants of health how they can impact low-income and historically marginalized communities. Potential associations of this data include urban counties having higher crime rates, rural towns having higher life expectancies, and higher average income being correlated to higher life expectancy per county.

```
library(tidyverse)
library(dplyr)
health=read.csv('health.csv')
crime=read.csv('crime.csv')
names(health)[11]<-"householdincome_black"
glimpse(health)
```

```
## Observations: 254
## Variables: 14
## $ County          <fct> Anderson, Andrews, Angelina, Aransas, Ar...
## $ lifeexpectancy  <dbl> 73.8, 76.8, 76.3, 76.9, 77.0, 72.4, 78.3...
## $ frequentmentaldistressrate <int> 11, 10, 13, 12, 10, 10, 11, 11, 13, 11, ...
```

```
## $ mvmortalityrate      <int> 19, 43, 20, 19, 26, NA, 20, 22, 30, 24, ...
## $ household_income     <int> 42412, 63451, 45318, 46970, 58311, 55337...
## $ householdincome_black <int> 24427, 38359, 28509, NA, NA, NA, 72827, ...
## $ householdincome_hispanic <int> 37635, 66319, 41673, 43302, 34375, 15025...
## $ householdincome_white <int> 48013, 73648, 51631, 45496, 65504, 64737...
## $ segregationindex      <int> 35, 11, 37, 27, 22, NA, 15, 18, NA, 24, ...
## $ firearmfatalitiesrate <int> 14, 14, 13, 20, NA, NA, 17, 19, NA, 10, ...
## $ householdincome_black <int> 71, 74, 66, 74, 82, 81, 74, 74, 72, 85, ...
## $ population           <int> 57741, 17722, 87805, 25572, 8809, 1879, ...
## $ urbanization_status   <fct> Rural, Rural, Rural, Urban, Urban, Urban...
## $ border_status         <fct> Non-Border, Non-Border, Non-Border, Non-...
```

```
glimpse(crime)
```

```
## Observations: 254
## Variables: 4
## $ county      <fct> Anderson, Andrews, Angelina, Aransas, Archer, Armst...
## $ crime_rate   <dbl> 3061.6, 3130.1, 3668.0, 4577.5, 1112.2, 1981.7, 241...
## $ crimes_reported <int> 1787, 521, 3235, 1071, 77, 39, 1152, 520, 90, 283, ...
## $ crime_rank    <int> 67, 65, 42, 14, 211, 147, 103, 160, 198, 192, 121, ...
```

Tidying: Rearranging

```
#Untidying of 'Health' Dataset
```

```
health_untidy <-health %>%
```

```
  pivot_longer(c('firearmfatalitiesrate', 'mvmortalityrate'),
               names_to="mortality types",
               values_to="number of mortalities")%>% glimpse()
```

```
## Observations: 508
## Variables: 14
## $ County      <fct> Anderson, Anderson, Andrews, Andrews, An...
## $ lifeexpectancy <dbl> 73.8, 73.8, 76.8, 76.8, 76.3, 76.3, 76.9...
## $ frequentmental distressrate <int> 11, 11, 10, 10, 13, 13, 12, 12, 10, 10, ...
## $ household_income     <int> 42412, 42412, 63451, 63451, 45318, 45318...
## $ householdincome_black <int> 24427, 24427, 38359, 38359, 28509, 28509...
## $ householdincome_hispanic <int> 37635, 37635, 66319, 66319, 41673, 41673...
## $ householdincome_white <int> 48013, 48013, 73648, 73648, 51631, 51631...
## $ segregationindex      <int> 35, 35, 11, 11, 37, 37, 27, 27, 22, 22, ...
## $ householdincome_black <int> 71, 71, 74, 74, 66, 66, 74, 74, 82, 82, ...
## $ population           <int> 57741, 57741, 17722, 17722, 87805, 87805...
## $ urbanization_status   <fct> Rural, Rural, Rural, Rural, Rural, Rural...
## $ border_status         <fct> Non-Border, Non-Border, Non-Border, Non-...
## $ `mortality types`     <chr> "firearmfatalitiesrate", "mvmortalityrat...
## $ `number of mortalities` <int> 14, 19, 14, 43, 13, 20, 20, 19, NA, 26, ...
```

```
#Tidying of 'Health' Dataset
```

```
health_untidy %>%
```

```
  spread(key="mortality types",value="number of mortalities")%>%
  glimpse()
```

```
## Observations: 254
## Variables: 14
## $ County      <fct> Anderson, Andrews, Angelina, Aransas, Ar...
## $ lifeexpectancy <dbl> 73.8, 76.8, 76.3, 76.9, 77.0, 72.4, 78.3...
```

```
## $ frequentmentaldistressrate <int> 11, 10, 13, 12, 10, 10, 11, 11, 13, 11, ...
## $ household_income <int> 42412, 63451, 45318, 46970, 58311, 55337...
## $ householdincome_black <int> 24427, 38359, 28509, NA, NA, NA, 72827, ...
## $ householdincome_hispanic <int> 37635, 66319, 41673, 43302, 34375, 15025...
## $ householdincome_white <int> 48013, 73648, 51631, 45496, 65504, 64737...
## $ segregationindex <int> 35, 11, 37, 27, 22, NA, 15, 18, NA, 24, ...
## $ householdincome_black <int> 71, 74, 66, 74, 82, 81, 74, 74, 72, 85, ...
## $ population <int> 57741, 17722, 87805, 25572, 8809, 1879, ...
## $ urbanization_status <fct> Rural, Rural, Rural, Urban, Urban, Urban...
## $ border_status <fct> Non-Border, Non-Border, Non-Border, Non-...
## $ firearmfatalitiesrate <int> 14, 14, 13, 20, NA, NA, 17, 19, NA, 10, ...
## $ mvmortalityrate <int> 19, 43, 20, 19, 26, NA, 20, 22, 30, 24, ...
```

#Untidying of 'Crime' Dataset

```
crime1 <-crime %>%
  pivot_longer(c('crimes_reported','crime_rate'),
    names_to='crimes reported and crime rate',
    values_to='values') %>% glimpse()
```

```
## Observations: 508
## Variables: 4
## $ county <fct> Anderson, Anderson, Andrews, Andre...
## $ crime_rank <int> 67, 67, 65, 65, 42, 42, 14, 14, 21...
## $ `crimes reported and crime rate` <chr> "crimes_reported", "crime_rate", "...
## $ values <dbl> 1787.0, 3061.6, 521.0, 3130.1, 323...
```

#Tidying of 'Crime' Dataset

```
crime1 %>%
  spread(key='crimes reported and crime rate',
    value='values') %>%
  glimpse()
```

```
## Observations: 254
## Variables: 4
## $ county <fct> Anderson, Andrews, Angelina, Aransas, Archer, Armst...
## $ crime_rank <int> 67, 65, 42, 14, 211, 147, 103, 160, 198, 192, 121, ...
## $ crime_rate <dbl> 3061.6, 3130.1, 3668.0, 4577.5, 1112.2, 1981.7, 241...
## $ crimes_reported <dbl> 1787, 521, 3235, 1071, 77, 39, 1152, 520, 90, 283, ...
```

To untidy the 'health' dataset, I first used `pivot longer` for the columns 'firearmfatalitiesrate' and 'mvmortalityrate'. I then used 'names_to' code to place the column names into a new column called 'mortality types' and the values in a column called 'number of mortalities'. This process was also recreated for the 'crime' dataset, where the 'crime_rank' and 'crime_rate' column names were placed into 'crime rank and rate' and values were placed into 'values'. This made the data untidy because each variable does not have its own column. This is because with the `pivot_longer` function, I put multiple variables, for example 'firearmfatalitiesrate' and 'mvmortalityrate', into one column. To tidy the datasets, I used the `spread()` function, where the key was the previously named in `pivot_longer` in `names_to=` and the value was the previously named 'values_to='. For example, I used the 'crime rank and rate' name in key, which was named while untidying the dataset, and the 'values' name in value, which was again named while untidying the dataset. This then changed the dataset into a tidy dataset because each variable was now back in a separate column.

Joining/ Merging

```
joined_data<- health%>%full_join(crime,by=c("County"="county"))
glimpse(joined_data)
```

```
## Observations: 254
## Variables: 17
## $ County <fct> Anderson, Andrews, Angelina, Aransas, Ar...
## $ lifeexpectancy <dbl> 73.8, 76.8, 76.3, 76.9, 77.0, 72.4, 78.3...
## $ frequentmentaldistressrate <int> 11, 10, 13, 12, 10, 10, 11, 11, 13, 11, ...
## $ mvmmortalityrate <int> 19, 43, 20, 19, 26, NA, 20, 22, 30, 24, ...
## $ household_income <int> 42412, 63451, 45318, 46970, 58311, 55337...
## $ householdincome_black <int> 24427, 38359, 28509, NA, NA, NA, 72827, ...
## $ householdincome_hispanic <int> 37635, 66319, 41673, 43302, 34375, 15025...
## $ householdincome_white <int> 48013, 73648, 51631, 45496, 65504, 64737...
## $ segregationindex <int> 35, 11, 37, 27, 22, NA, 15, 18, NA, 24, ...
## $ firearmfatalitiesrate <int> 14, 14, 13, 20, NA, NA, 17, 19, NA, 10, ...
## $ householdincome_black <int> 71, 74, 66, 74, 82, 81, 74, 74, 72, 85, ...
## $ population <int> 57741, 17722, 87805, 25572, 8809, 1879, ...
## $ urbanization_status <fct> Rural, Rural, Rural, Urban, Urban, Urban...
## $ border_status <fct> Non-Border, Non-Border, Non-Border, Non-...
## $ crime_rate <dbl> 3061.6, 3130.1, 3668.0, 4577.5, 1112.2, ...
## $ crimes_reported <int> 1787, 521, 3235, 1071, 77, 39, 1152, 520...
## $ crime_rank <int> 67, 65, 42, 14, 211, 147, 103, 160, 198,...
```

The `full_join` function was used to not drop any rows and retain all of the original data. Since both datasets had the 'county' variable and the same number and names of the counties, a `full_join` was used to join the datasets by this common variable since the column appears in both datasets. Since the 'crime' dataset and 'health' dataset had different column names for the county variable and as a result `by=c()` was used to join the columns despite the different column names for the same variable. No cases were dropped because each county in the 'health' dataset were present in the 'crime' dataset, such that the 'health', 'crime', and 'joined_data' datasets all have 254 observations.

Wrangling

```
#Six core dplyr functions
small_joined_data <-joined_data %>%
  filter(lifeexpectancy>77)%>%
  select(County, crimes_reported, lifeexpectancy, householdincome_black, household_income,
         population,crime_rate,
         urbanization_status,border_status) %>%
  arrange(lifeexpectancy) %>%
  group_by(border_status) %>%
  mutate(black_percentageofaveragehouseholdincome=
         (householdincome_black/household_income) *100)

#Different Summary Statistics
small_joined_data %>%
  summarize('number of distinct counties'=n_distinct(County))

## # A tibble: 2 x 2
##   border_status `number of distinct counties`
##   <fct>          <int>
## 1 Border              23
## 2 Non-Border         98

small_joined_data %>%
  summarize('mean population'= mean(population))
```

```
## # A tibble: 2 x 2
##   border_status `mean population`
##   <fct>          <dbl>
## 1 Border          119005.
## 2 Non-Border      219665.
```

```
small_joined_data %>%
  summarize('meancrimerate'=mean(crime_rate))
```

```
## # A tibble: 2 x 2
##   border_status meancrimerate
##   <fct>          <dbl>
## 1 Border          2218.
## 2 Non-Border      2149.
```

```
small_joined_data %>%
  summarize('mean in number of crimes reported'=mean(crimes_reported))
```

```
## # A tibble: 2 x 2
##   border_status `mean in number of crimes reported`
##   <fct>          <dbl>
## 1 Border          3893.
## 2 Non-Border      7515.
```

```
small_joined_data %>%
  summarize('number of distinct life expectancies'=n_distinct(lifeexpectancy))
```

```
## # A tibble: 2 x 2
##   border_status `number of distinct life expectancies`
##   <fct>          <int>
## 1 Border           22
## 2 Non-Border       39
```

```
small_joined_data %>%
  summarize('minimum household income'= min(household_income))
```

```
## # A tibble: 2 x 2
##   border_status `minimum household income`
##   <fct>          <int>
## 1 Border          28377
## 2 Non-Border      35075
```

```
small_joined_data %>%
  summarize('maximum household income'=max(household_income))
```

```
## # A tibble: 2 x 2
##   border_status `maximum household income`
##   <fct>          <int>
## 1 Border          57986
## 2 Non-Border      98442
```

```
small_joined_data %>%
  summarize('median life expectancy'=median(lifeexpectancy))
```

```
## # A tibble: 2 x 2
##   border_status `median life expectancy`
##   <fct>          <dbl>
## 1 Border          79.5
```

```
## 2 Non-Border                                78.5
small_joined_data %>% group_by(urbanization_status) %>%
  summarize(mean(crimes_reported))
```

```
## # A tibble: 2 x 2
##   urbanization_status `mean(crimes_reported)`
##   <fct>                <dbl>
## 1 Rural                403.
## 2 Urban              15350.
```

```
small_joined_data %>% group_by(urbanization_status) %>%
  summarize(min(household_income))
```

```
## # A tibble: 2 x 2
##   urbanization_status `min(household_income)`
##   <fct>                <int>
## 1 Rural              28377
## 2 Urban              35988
```

To create the ‘small_joined_dataset’, I first filtered by groups that had an average life expectancy that was greater than 77 years old using the `filter()` function. I then selected columns including ‘county’, ‘crimes_reported’, ‘lifeexpectancy’, ‘householdincome_black’, ‘household_income’, ‘population’, ‘crime_rate’, ‘urbanization_status’, ‘border_status’ from the full dataset using the `select()` function. I then arranged the dataset by life expectancy using `arrange()` such that the lowest life expectancy was the first group. I then grouped the data by border status using `group_by()`, which in turn grouped the data into either Border or Non-border status groups. Lastly, I created a column using `mutate()` that illustrated the percentage that the average black household income was of the total average household income.

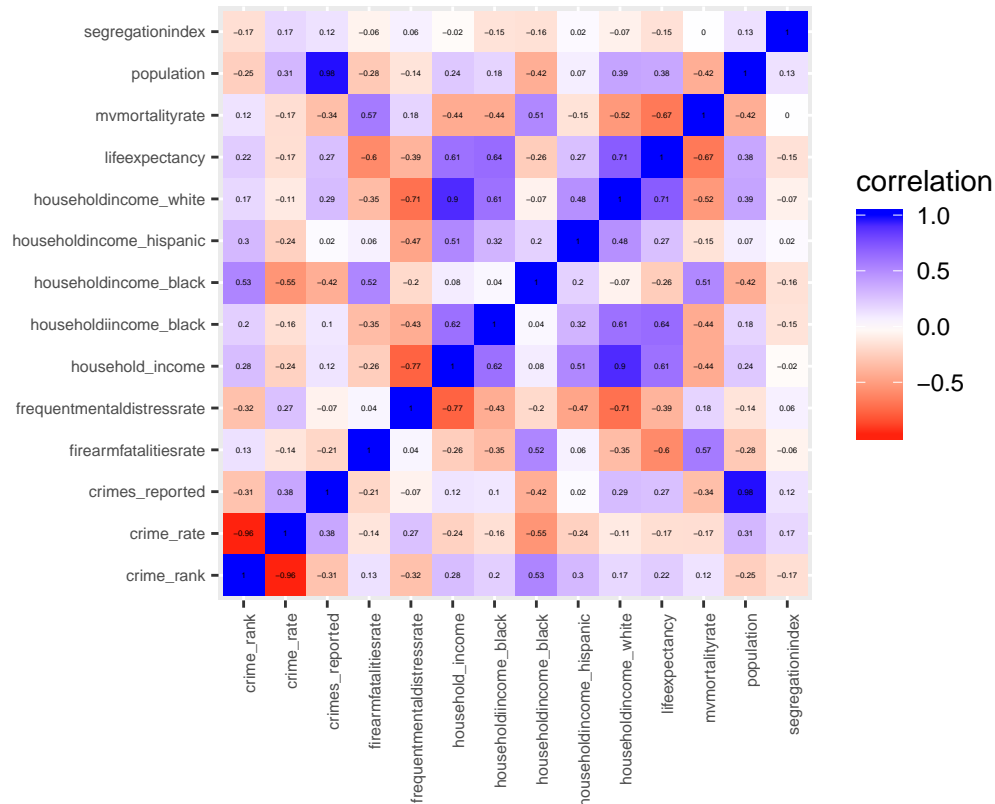
I used the `summarize()` function as a basis for the summary statistics. Using the ‘small_joined_dataset’ I found that the number of distinct counties within the created dataset was 23 border counties and 98 non-border counties. The average population of the border counties was 119,005.5 people, while it was 219,665.3 for non-border counties. The standard deviation in the crime rates reported were 64.50205 in border counties and 74.22479 in non-border counties. The variance in the number of crimes reported was 3,257.130 in border counties and 4,240.635 in non-border counties. There were 22 distinct life expectancies in the border counties, and 39 in the non-border counties. Of the border counties, the minimum average household income was 283,77, while it was 35,075 in non-border counties. The maximum average household income of the border counties in the dataset was 57,986, while it was 98,442 in the non-border counties. The median life expectancy of the border counties in the created dataset was 79.5 years, while the median was 78.5 years in the non-border counties. I also used the `group_by()` function to group the created dataset by urbanization status to determine the mean number of crimes reported per rural and urban area. It was determined that the average was 403.1884 crimes reported in rural counties, and 15,350.2885 crimes reported in urban counties. I used the `group_by()` function again to group the data by urbanization status and saw that the minimum household income of the rural counties in the created data set was 28,377, while it was 35,988 in the urban counties.

Visualizing

```
# Correlation Heat Map
joined_data_nona<-joined_data%>%na.omit%>%select_if(is.numeric)
tidycor<-cor(joined_data_nona)%>%as.data.frame%>%
rownames_to_column%>%
pivot_longer(-1,names_to="name",values_to="correlation")
tidycor%>% ggplot(aes(rowname,name,fill=correlation))+ xlab("")+ylab("")+
geom_tile()+ scale_fill_gradient2(low="red",mid="white",high="blue")+

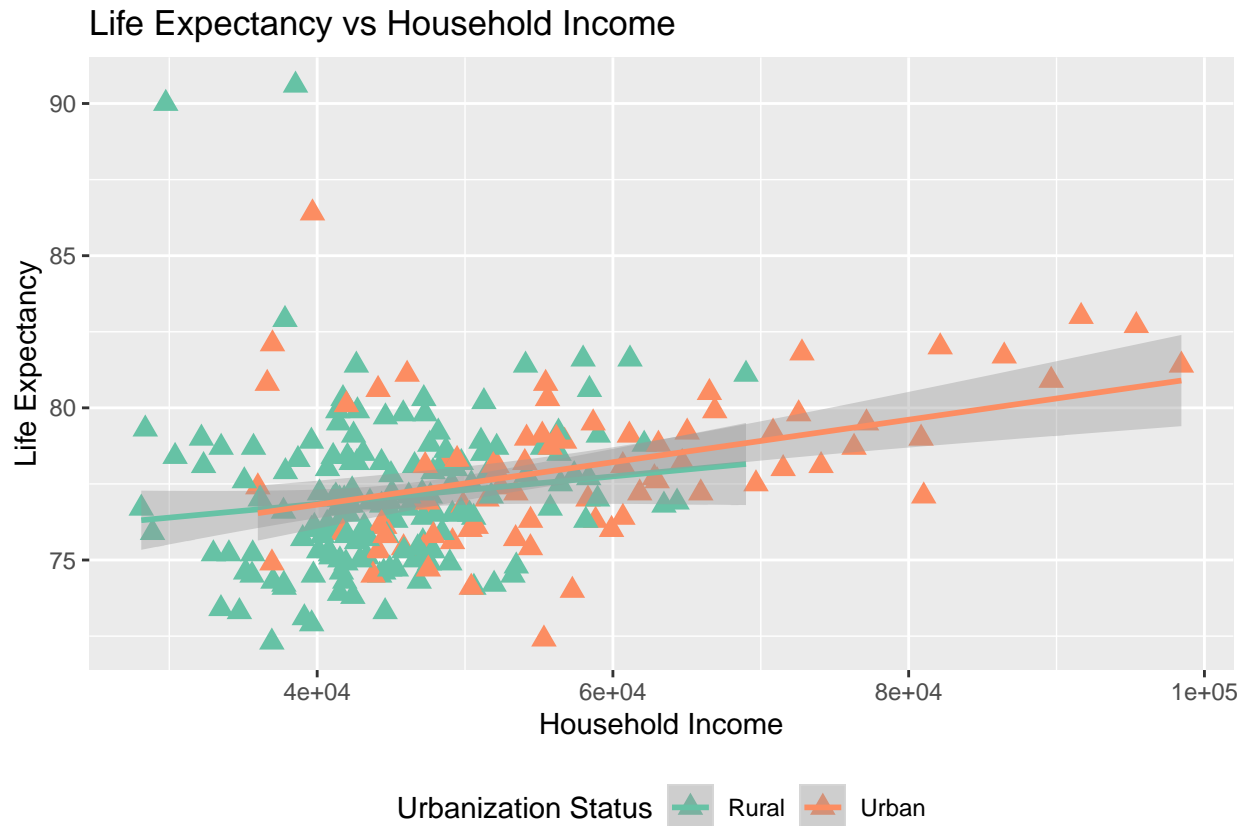
```

```
geom_text(aes(label=round(correlation,2)),color = "black", size = .9)+
theme(axis.text.x = element_text(angle = 90, hjust=1,size=6))+
coord_fixed()+theme(axis.text.y = element_text(hjust=1,size=6))
```



The correlation heat maps reveals what variables are closely correlated, both positively and negatively, and those which are not. Average household income and average household income for white were fairly strongly positively correlated, while the average household income for blacks and Hispanics was less tightly correlated to the average income. Life expectancy was also positively correlated to average household income for whites, such that a higher life expectancy is more likely in counties with higher average income in white populations. The population and crimes reported are also very closely positively correlated, which shows that a larger county is more likely to have more crimes reported. The crime rate and the crimes reported are very negatively correlated, possibly because the crime rate makes a rate based on very little data in small counties and extrapolates it for 100,000 people, which could lead to a high crime ranking in relatively small counties. Interestingly, segregation index is not closely correlated with any of the other variables.

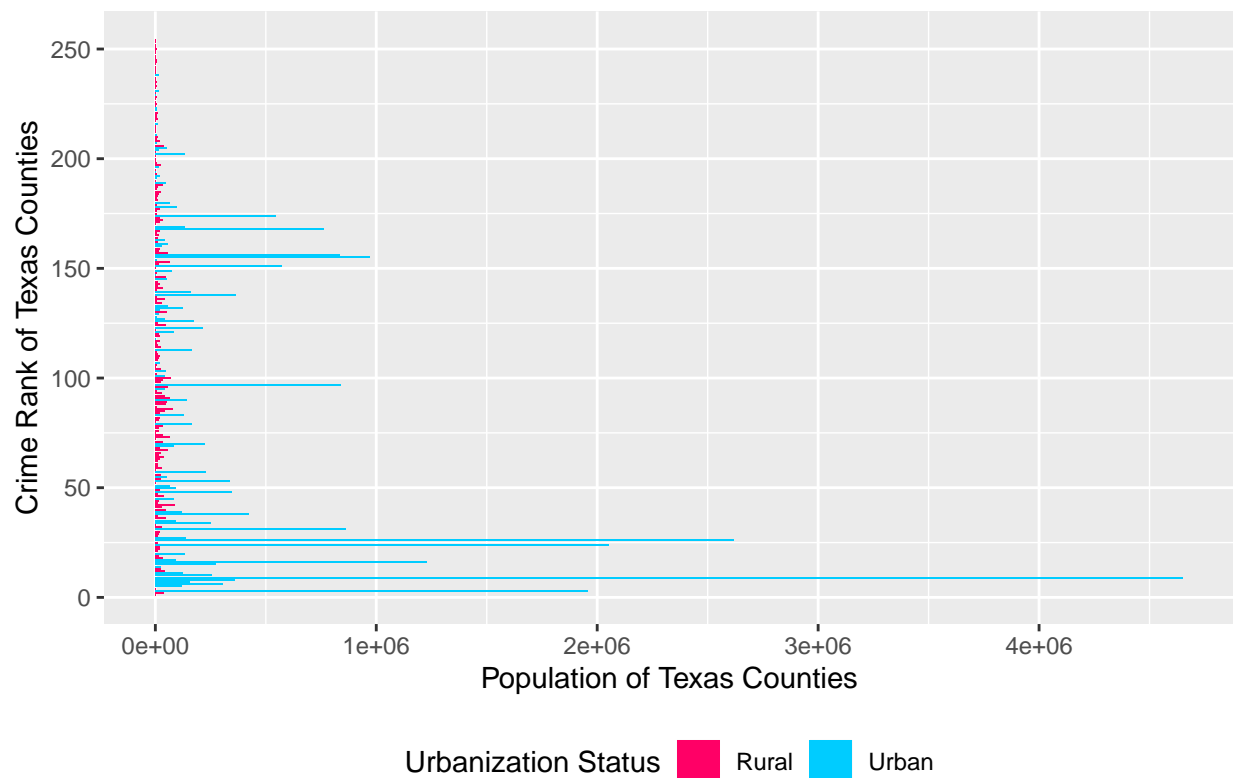
```
# Plot 1: Population versus Number of Crimes Reported in Texas Counties
ggplot(data = joined_data, aes(x = household_income, y = lifeexpectancy,
                               color=urbanization_status))+
  geom_point(size=3,shape='triangle')+
  ggtitle("Life Expectancy vs Household Income") +
  xlab("Household Income")+ylab("Life Expectancy") +
  theme(legend.position = "bottom")+
  geom_smooth(method='lm',formula = y ~ x)+
  labs(colour="Urbanization Status")+ scale_color_brewer(palette = "Set2")
```

The first plot represents the household income versus life expectancy. We can also visualize the difference between urban and rural counties based on the colors of the points. We can see that there is a positive relationship between household income and life expectancy in Texas counties, such that life expectancy increases the household income likely also increases in both rural and urban counties. It appears as though the correlation is slightly stronger in urban counties. However, the association between life expectancy and household income is not extremely strong as the slope of the line is not very steep. This graph also helps visualize that lower income counties are usually rural, while higher income counties are usually rural.

```
#Plot 2: Population versus Crime Ranking of Texas Counties
library(ggplot2)
ggplot(joined_data, aes(x=crime_rank, y=population))+
  coord_flip()+
  geom_bar(aes(fill = urbanization_status), stat="summary", fun.y="mean")+
  theme(legend.position = "bottom")+
  ggtitle('Population versus Crime Ranking of Texas Counties')+
  xlab('Crime Rank of Texas Counties')+ylab('Population of Texas Counties')+
  scale_y_continuous(breaks=seq(0,4000000,1000000))+
  scale_x_continuous(breaks=seq(0,250,50))+
  guides(fill=guide_legend('Urbanization Status'))+
  scale_fill_manual(values=c( "#FF0066", "#00CCFF"))
```


Population versus Crime Ranking of Texas Counties



This plot visualizes the population size versus the crime ranking of Texas counties. It is important to note that the crime rank was created by the database owners as the number of crimes per 100,000 people. However, some of these counties have many less people than 100,000 and therefore the small sampling size impacts the crime rate and therefore the crime ranking. This can be seen for Loving county, which is ranked number 1 and only reported 10 crimes but has a crime rate of 14,706 because it is a county of only 134 people. More generally, we can see that the counties with larger populations usually have a lower crime ranking or have a higher crime rate. Usually counties with small populations have are ranked higher on the crime ranking, meaning that they experience a lower crime rate. Overall, urban counties are usually ranked lower on the crime ranking, meaning they experience higher crime rates, as compared to rural counties.

Dimensionality Reduction

```
#PAM Clustering- Determining Number of Clusters
library(cluster)
pam_dat<-joined_data%>%select(lifeexpectancy,crimes_reported, household_income)
sil_width<-vector()
for(i in 2:10){pam_fit <- pam(pam_dat, k = i)
sil_width[i] <- pam_fit$silinfo$avg.width}
ggplot()+geom_line(aes(x=1:10,y=sil_width))+scale_x_continuous(name="k",breaks=1:10)
```



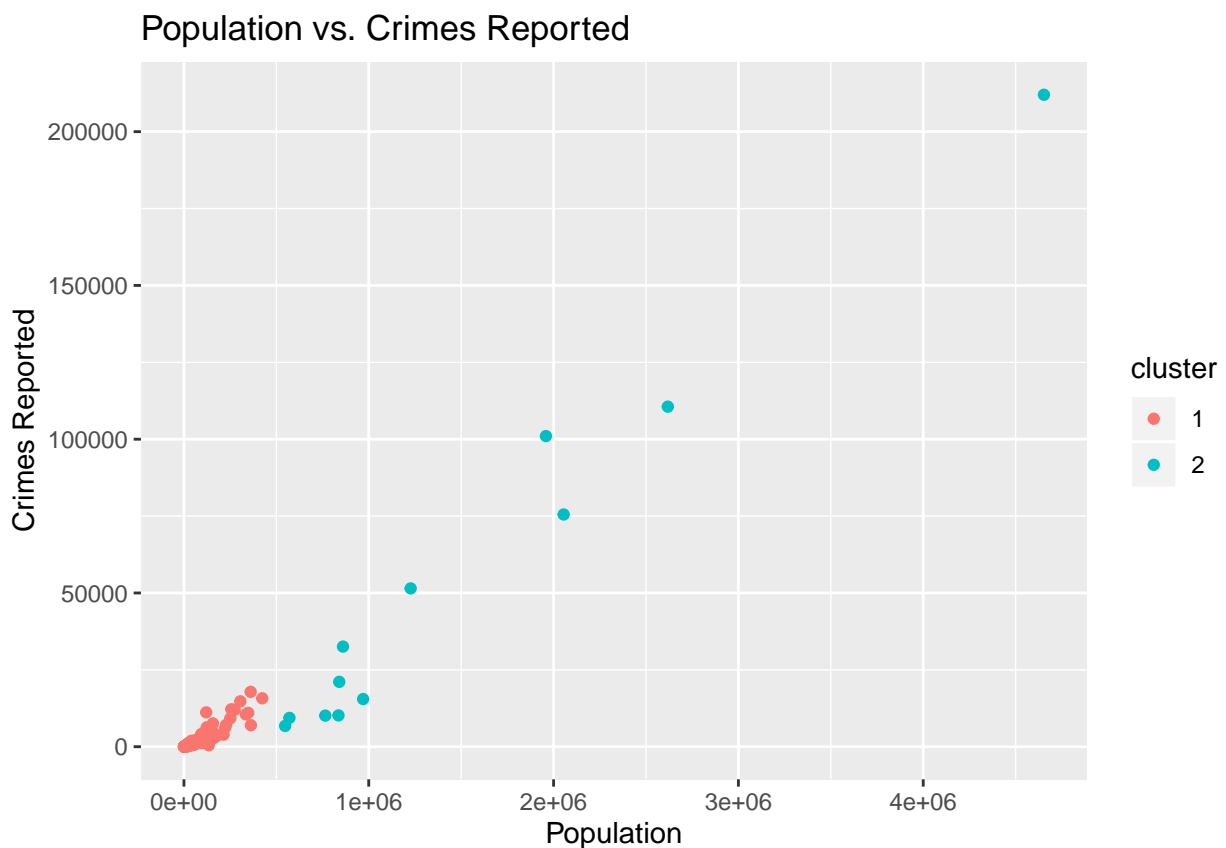
```
##
## Available components:
## [1] "medoids"      "id.med"      "clustering"  "objective"   "isolation"
## [6] "clusinfo"    "silinfo"     "diss"        "call"        "data"

#Silhouette Width
pam2$silinfo$avg.width

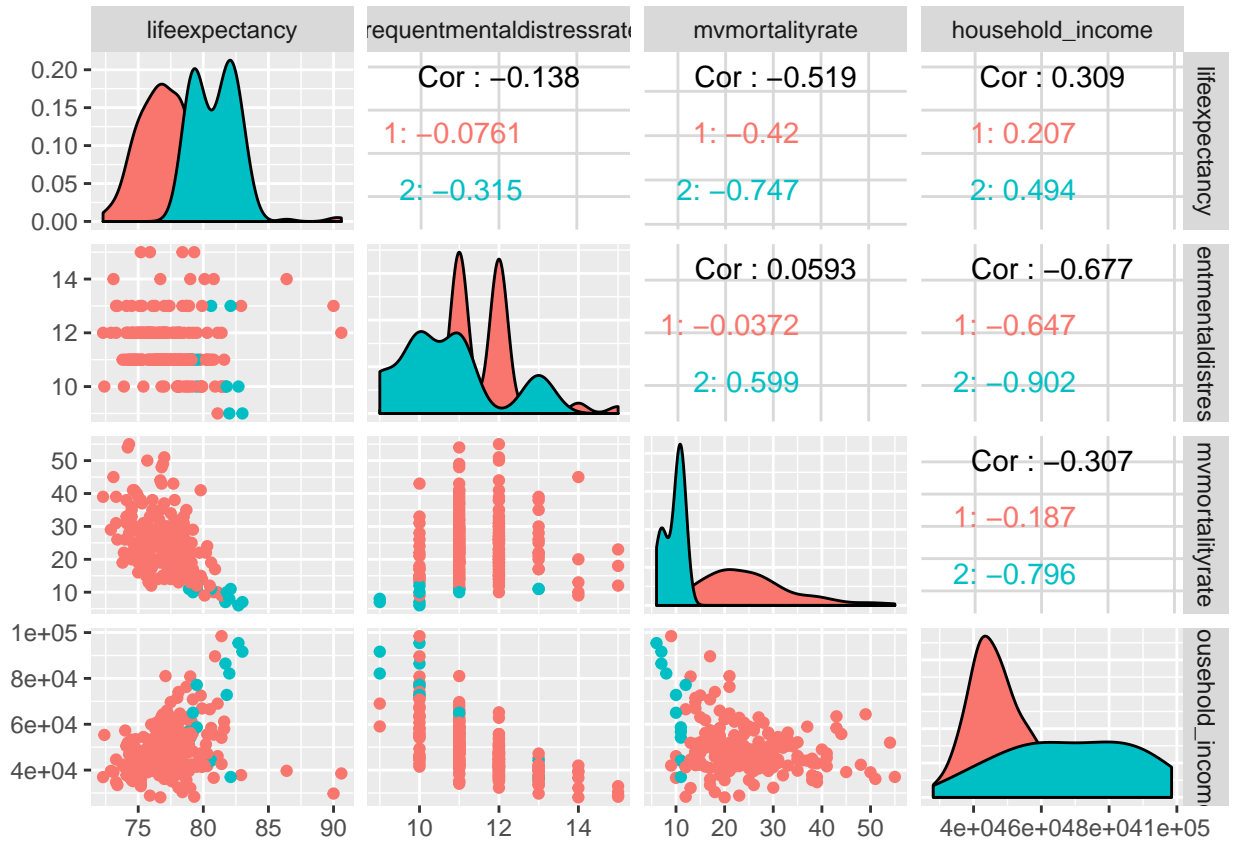
## [1] 0.9067435

final<-joined_data%>%mutate(cluster=as.factor(pam2$clustering))
confmat<-final%>%group_by(urbanization_status)%>%count(cluster)%>%arrange(desc(n))%>%
pivot_wider(names_from="cluster",values_from="n",values_fill = list('n'=0))

#Plot of Clusters: Population vs Crimes Reported
ggplot(final, aes(x=population,y=crimes_reported, color=cluster))+geom_point()+
  ggtitle('Population vs. Crimes Reported') + xlab('Population') + ylab('Crimes Reported')
```



```
#Pairwise Combinations of the 4 Variables
library(GGally)
ggpairs(final, columns=2:5, aes(color=cluster))
```



The Partitioning around Medoids, or PAM, was used to cluster the data into groups. To determine the number of clusters in PAM using the silhouette width, I first created `pam_dat`. We can see from the plot that 2 clusters is probably best for because 2 clusters has the highest silhouette width at approximately 0.49. The overall silhouette width was determined to be 0.9067435. This high silhouette width reveals that the clusters are cohesive and separated, which is very good for PAM and could be because of the large number of counties in the dataset. I then created clusters of population and crimes_reported to see the clusters based on those 2 variables. We can see based on the graph that there are 2 clusters where one cluster was where the population and crimes reported were small and another cluster where the population and crimes reported were higher. Lastly, I visualized all pairwise combinations of the 4 variables running a correlation matrix.

From the correlation visualization of all pairwise combinations of 4 variables we can see correlations between life expectancy, frequent mental distress, motor vehicle mortality rate, and household income both overall and per cluster. The data was colored by clusters to visualize the differences between cluster 1 and cluster 2. We can see that cluster 1 has a low life expectancy, while cluster 2 has a higher life expectancy. The frequent mental distress is not well clustered because the clusters overlap quite greatly. The motor vehicle mortality rate is distinctly clustered where the lowest mortality rate is in cluster 2 and the highest is in cluster 1. In cluster 1, the household income is lower than in cluster 2. Some of the clusters are well separated based on the variable, while others have greater overlap.

The overall correlation between frequent mental distress rate and life expectancy is -0.138, which is not a strong correlation between the variables. There is a very small difference in correlation between clusters, where cluster 1 has a correlation of -0.0761 and cluster 2 has a correlation of -0.315. The motor vehicle mortality rate and life expectancy are somewhat negatively correlated at -0.519 where cluster 1 had a correlation of -0.42 and cluster 2 had a much stronger negative correlation of -0.747. Household income and life expectancy had a correlation of 0.309, which is not very strong, however cluster 2 had a higher positive correlation of 0.494 as compared to cluster 1, which had a correlation of 0.207. Motor vehicle mortality rate has the smallest overall correlation at 0.0593. However, the clusters themselves have very different correlations, where cluster 1 has a negative correlation of -0.0372 and cluster 2 has a positive correlation of 0.599. Household

income and frequent mental distress have the strongest negative correlation of -0.677. Cluster 1 has a negative correlation of 0.647, while cluster 2 has an even stronger negative correlation of -0.902. Household income and motor vehicle mortality rate are not closely correlated at -0.307. The correlation rates between clusters are very different, where cluster 1 has a weak negative correlation of -0.187 and cluster 2 has a strong negative correlation of -0.796.