Full Code

2024-12-13

new\_country\_data = country\_data[, c("name","homicide\_rate","gdp", "urban\_population\_growth", "pop\_growth", "internet\_users", "gdp\_per\_capita", "refugees")]  
head(new\_country\_data)

## name homicide\_rate gdp urban\_population\_growth pop\_growth  
## 1 Afghanistan 6.7 20514 4.0 2.5  
## 2 Albania 2.3 15059 1.8 -0.1  
## 3 Algeria 1.4 173757 2.9 2.0  
## 4 Andorra 0.0 3238 -1.7 -0.2  
## 5 Angola 4.8 105902 4.7 3.3  
## 6 Antigua And Barbuda 11.1 1611 0.1 0.9  
## internet\_users gdp\_per\_capita refugees  
## 1 13.5 551.9 2826.4  
## 2 71.8 5223.8 4.3  
## 3 49.0 4114.7 99.5  
## 4 91.6 42051.6 NA  
## 5 14.3 3437.3 70.1  
## 6 76.0 16727.0 0.2

clean\_data\_homcides <- new\_country\_data[!is.na(new\_country\_data$homicide\_rate), ]  
no\_dupes = clean\_data\_homcides%>% distinct(name, .keep\_all = TRUE)  
  
final\_data = no\_dupes

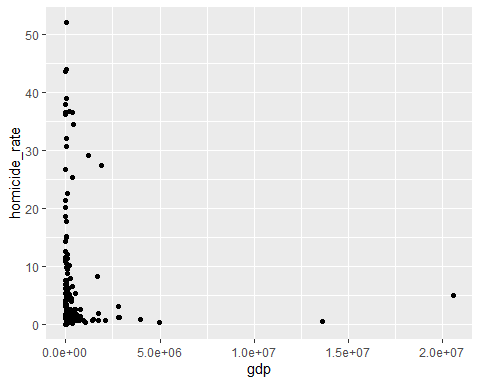
## Including Plots

You can also embed plots, for example:

***Visual analysis using scatterplots:***

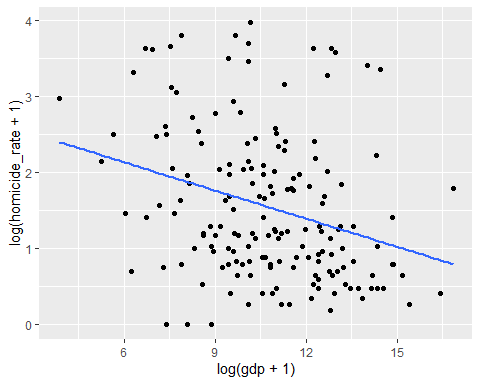
**#1. Homicide Rate vs GDP**

ggplot(final\_data, aes(y = homicide\_rate, x = gdp))+  
 geom\_point()



**Perform log transformation on x and y due to heavy skewness**

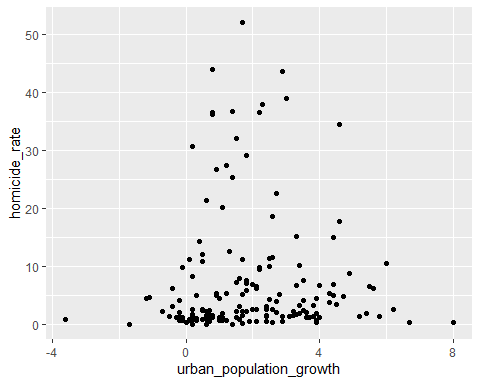
ggplot(final\_data, aes(y = log(homicide\_rate + 1), x = log(gdp + 1)))+  
 geom\_point() +  
 geom\_smooth(method = "lm", se = FALSE)



**Plot seems to indicate an approximate moderate negative linear correlation. Use Pearson’s correlation for testing.**

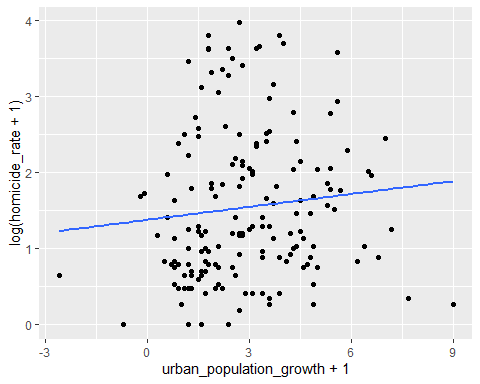
**2. Homicide Rate vs Urban Population Growth**

ggplot(final\_data, aes(y = homicide\_rate, x=urban\_population\_growth))+  
 geom\_point()



**Perform log transformation on y due to nonlinear and non-monotonic relationship**

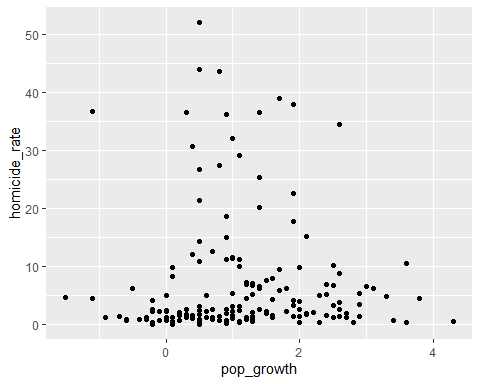
ggplot(final\_data, aes(y = log(homicide\_rate + 1), x=urban\_population\_growth+1))+  
 geom\_point() +  
 geom\_smooth(method = "lm", se=FALSE)



**Plot can be described with a weak positive linear correlation. Use Pearson’s correlation for testing.**

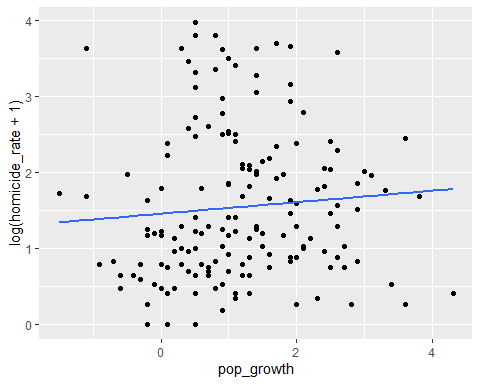
**3. Homicide Rate vs Population Growth**

ggplot(final\_data, aes(y = homicide\_rate, x=pop\_growth))+  
 geom\_point()



**Perform log transformation on y due to nonlinear and non-monotonic relationship**

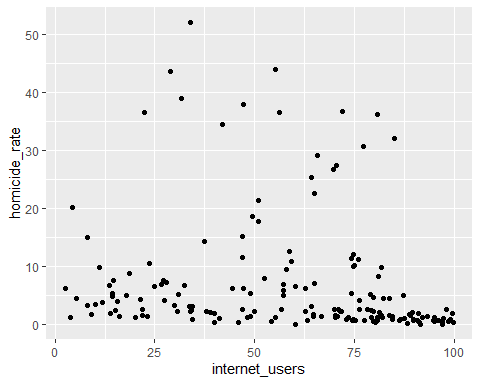
ggplot(final\_data, aes(y = log(homicide\_rate + 1), x=pop\_growth))+  
 geom\_point() +  
 geom\_smooth(method = "lm", se=FALSE)



**Plot can be described with a weak positive linear correlation. Use Pearson’s correlation for testing.**

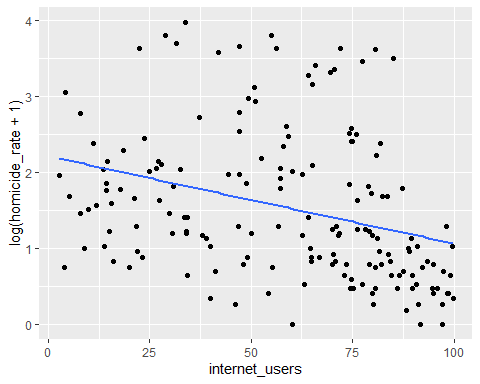
**4. Homicide Rate vs Internet Users Percentage**

ggplot(final\_data, aes(y = homicide\_rate, x=internet\_users))+  
 geom\_point()



**Perform log transformation on y due to nonlinearity and non-monotonic relationship**

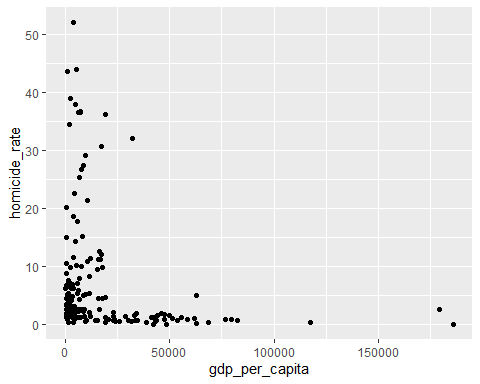
ggplot(final\_data, aes(y = log(homicide\_rate+1), x=internet\_users))+  
 geom\_point() +  
 geom\_smooth(method="lm",se=FALSE)



**Plot can be described with a moderate negative linear correlation. Use Pearson’s correlation for testing.**

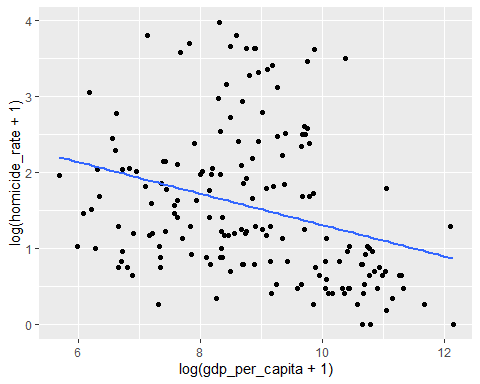
**5. Homicide Rate vs GDP Per Capita**

ggplot(final\_data, aes(y = homicide\_rate, x = gdp\_per\_capita))+  
 geom\_point()



**Perform log transformation on x and y due to heavy skewness**

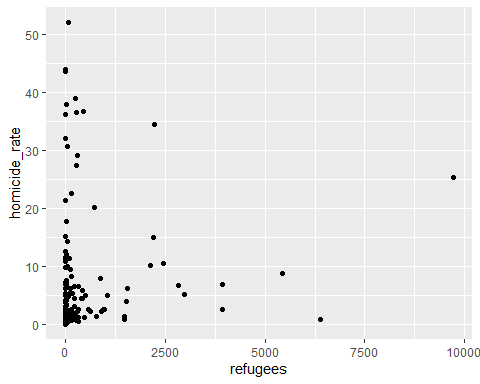
ggplot(final\_data, aes(y = log(homicide\_rate + 1), x = log(gdp\_per\_capita + 1)))+  
 geom\_point() +  
 geom\_smooth(method = "lm", se = FALSE)



**Plot seems to indicate a moderate negative linear correlation. Use Pearson’s correlation for testing.**

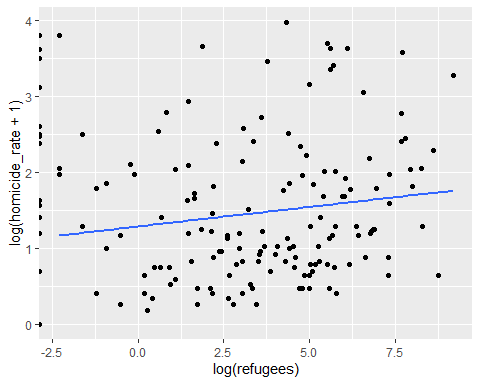
**6. Homicide Rate vs Refugees**

ggplot(final\_data, aes(y = homicide\_rate, x=refugees))+  
 geom\_point()



**Perform log transformation on x and y due to heavy skewness.**

ggplot(final\_data, aes(y = log(homicide\_rate+1), x=log(refugees)))+  
 geom\_point() +  
 geom\_smooth(method="lm",se=FALSE)



**Plot seems to indicate a weak negative linear correlation. Use Pearson’s correlation for testing.**

***Association Analysis***

**1. log Homicide Rate vs log GDP**

associate(log(homicide\_rate + 1)~log(gdp + 1), data=final\_data, seed=9750)

## Association between log(gdp + 1) (numerical) and log(homicide\_rate + 1) (numerical)  
## using 168 complete cases

## Permutation procedure:  
## Value Estimated p-value  
## Pearson's r -0.2951072 0  
## Spearman's rank correlation -0.3068722 0  
## With 500 permutations, we are 95% confident that:  
## the p-value of Pearson's correlation (r) is between 0 and 0.007   
## the p-value of Spearman's rank correlation is between 0 and 0.007   
## Note: If 0.05 is in this range, increase the permutations= argument.  
##   
##   
##   
## Advice: If stream of points is well described by an ellipse, use Pearson's r.  
## Otherwise, as long as stream is monotonic, use Spearman's rank correlation  
## or try logs, e.g. associate( log10(y)~log10(x) )

**There exists a statistically significant correlation, p-value = 0. There is a moderate negative linear correlation, r = -0.2951072**

**2. log Homicide Rate vs Urban Population Growth**

associate(log(homicide\_rate+1)~urban\_population\_growth, data=final\_data, seed=9750)

## Association between urban\_population\_growth (numerical) and log(homicide\_rate + 1) (numerical)  
## using 169 complete cases

## Permutation procedure:  
## Value Estimated p-value  
## Pearson's r 0.1014501 0.186  
## Spearman's rank correlation 0.1816352 0.018  
## With 500 permutations, we are 95% confident that:  
## the p-value of Pearson's correlation (r) is between 0.153 and 0.223   
## the p-value of Spearman's rank correlation is between 0.008 and 0.034   
## Note: If 0.05 is in this range, increase the permutations= argument.  
##   
##   
##   
## Advice: If stream of points is well described by an ellipse, use Pearson's r.  
## Otherwise, as long as stream is monotonic, use Spearman's rank correlation  
## or try logs, e.g. associate( log10(y)~log10(x) )

**There does not exist a statistically significant correlation, p-value = 0.186**

**3. log Homicide Rate vs Population Growth**

associate(log(homicide\_rate+1)~pop\_growth, data=final\_data, seed=9750)

## Association between pop\_growth (numerical) and log(homicide\_rate + 1) (numerical)  
## using 169 complete cases

## Permutation procedure:  
## Value Estimated p-value  
## Pearson's r 0.08182263 0.268  
## Spearman's rank correlation 0.17286409 0.022  
## With 500 permutations, we are 95% confident that:  
## the p-value of Pearson's correlation (r) is between 0.23 and 0.309   
## the p-value of Spearman's rank correlation is between 0.011 and 0.039   
## Note: If 0.05 is in this range, increase the permutations= argument.  
##   
##   
##   
## Advice: If stream of points is well described by an ellipse, use Pearson's r.  
## Otherwise, as long as stream is monotonic, use Spearman's rank correlation  
## or try logs, e.g. associate( log10(y)~log10(x) )

**There is not a statistically significant correlation, p-value = 0.268**

**4. log Homicide Rate vs Internet Users Percentage**

associate(log(homicide\_rate+1)~internet\_users, data=final\_data, seed=9750)

## Association between internet\_users (numerical) and log(homicide\_rate + 1) (numerical)  
## using 167 complete cases

## Permutation procedure:  
## Value Estimated p-value  
## Pearson's r -0.3205548 0  
## Spearman's rank correlation -0.4290509 0  
## With 500 permutations, we are 95% confident that:  
## the p-value of Pearson's correlation (r) is between 0 and 0.007   
## the p-value of Spearman's rank correlation is between 0 and 0.007   
## Note: If 0.05 is in this range, increase the permutations= argument.  
##   
##   
##   
## Advice: If stream of points is well described by an ellipse, use Pearson's r.  
## Otherwise, as long as stream is monotonic, use Spearman's rank correlation  
## or try logs, e.g. associate( log10(y)~log10(x) )

**There exists a statistically significant correlation, p-value = 0. There is a moderate negative linear correlation, r = -0.3205548**

**5. log Homicide Rate vs log GDP Per Capita**

associate(log(homicide\_rate+1)~log(gdp\_per\_capita+1), data=final\_data, seed=9750)

## Association between log(gdp\_per\_capita + 1) (numerical) and log(homicide\_rate + 1) (numerical)  
## using 168 complete cases

## Permutation procedure:  
## Value Estimated p-value  
## Pearson's r -0.3025433 0  
## Spearman's rank correlation -0.3769795 0  
## With 500 permutations, we are 95% confident that:  
## the p-value of Pearson's correlation (r) is between 0 and 0.007   
## the p-value of Spearman's rank correlation is between 0 and 0.007   
## Note: If 0.05 is in this range, increase the permutations= argument.  
##   
##   
##   
## Advice: If stream of points is well described by an ellipse, use Pearson's r.  
## Otherwise, as long as stream is monotonic, use Spearman's rank correlation  
## or try logs, e.g. associate( log10(y)~log10(x) )

**There exists a statistically significant association, p-value = 0. There is a moderate negative linear correlation, r = -0.3025433**

**6. log Homicide Rate vs log Refugee Population**

associate(log(homicide\_rate+1)~log(refugees+1), data=final\_data, seed=9750)

## Association between log(refugees + 1) (numerical) and log(homicide\_rate + 1) (numerical)  
## using 159 complete cases

## Permutation procedure:  
## Value Estimated p-value  
## Pearson's r 0.05412809 0.522  
## Spearman's rank correlation 0.06228297 0.460  
## With 500 permutations, we are 95% confident that:  
## the p-value of Pearson's correlation (r) is between 0.477 and 0.567   
## the p-value of Spearman's rank correlation is between 0.416 and 0.505   
## Note: If 0.05 is in this range, increase the permutations= argument.  
##   
##   
##   
## Advice: If stream of points is well described by an ellipse, use Pearson's r.  
## Otherwise, as long as stream is monotonic, use Spearman's rank correlation  
## or try logs, e.g. associate( log10(y)~log10(x) )

**There is not a statistically significant correlation, p-value = 0.456**

# Perform a multiple ANOVA test  
anova\_model = aov(homicide\_rate ~ urban\_population\_growth + pop\_growth + internet\_users + gdp + refugees + gdp\_per\_capita, data = final\_data)  
  
# Summarize the ANOVA results  
summary(anova\_model)

## Df Sum Sq Mean Sq F value Pr(>F)   
## urban\_population\_growth 1 30 29.9 0.291 0.5905   
## pop\_growth 1 33 33.4 0.325 0.5695   
## internet\_users 1 507 507.2 4.930 0.0279 \*  
## gdp 1 55 55.4 0.539 0.4640   
## refugees 1 119 119.2 1.159 0.2834   
## gdp\_per\_capita 1 216 215.5 2.095 0.1499   
## Residuals 151 15534 102.9   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
## 11 observations deleted due to missingness

library(car)

## Warning: package 'car' was built under R version 4.4.2

## Loading required package: carData

##   
## Attaching package: 'car'

## The following object is masked from 'package:VGAM':  
##   
## logit

## The following object is masked from 'package:dplyr':  
##   
## recode

# Fit a linear model  
linear\_model = lm(homicide\_rate ~ gdp + urban\_population\_growth + pop\_growth + internet\_users + gdp\_per\_capita + refugees, data = final\_data)  
  
# Calculate VIF  
vif\_values = vif(linear\_model)  
print(vif\_values)

## gdp urban\_population\_growth pop\_growth   
## 1.048183 3.067805 3.415295   
## internet\_users gdp\_per\_capita refugees   
## 2.137352 1.654864 1.043676

cor\_matrix = cor(final\_data[, c("gdp", "urban\_population\_growth", "pop\_growth", "internet\_users", "gdp\_per\_capita", "refugees")])  
print(cor\_matrix)

## gdp urban\_population\_growth pop\_growth internet\_users  
## gdp 1 NA NA NA  
## urban\_population\_growth NA 1.0000000 0.8168346 NA  
## pop\_growth NA 0.8168346 1.0000000 NA  
## internet\_users NA NA NA 1  
## gdp\_per\_capita NA NA NA NA  
## refugees NA NA NA NA  
## gdp\_per\_capita refugees  
## gdp NA NA  
## urban\_population\_growth NA NA  
## pop\_growth NA NA  
## internet\_users NA NA  
## gdp\_per\_capita 1 NA  
## refugees NA 1

# Scale numeric variables to standardize them  
numeric\_data = final\_data[, c("gdp", "urban\_population\_growth", "pop\_growth", "internet\_users", "gdp\_per\_capita", "refugees")]  
scaled\_data = scale(numeric\_data)  
  
# Calculate condition number  
condition\_number = kappa(cor(scaled\_data, use = "complete.obs"))  
print(condition\_number)

## [1] 14.65723

library(ggplot2)  
#country\_data = read.csv("country\_data.csv")  
new\_country\_data = country\_data[, c("name","homicide\_rate","gdp", "urban\_population\_growth", "pop\_growth", "internet\_users", "gdp\_per\_capita", "refugees")]  
head(new\_country\_data)

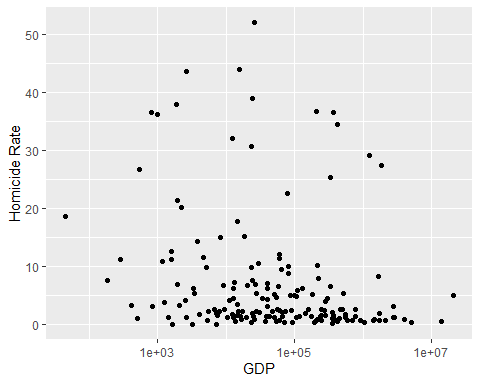
## name homicide\_rate gdp urban\_population\_growth pop\_growth  
## 1 Afghanistan 6.7 20514 4.0 2.5  
## 2 Albania 2.3 15059 1.8 -0.1  
## 3 Algeria 1.4 173757 2.9 2.0  
## 4 Andorra 0.0 3238 -1.7 -0.2  
## 5 Angola 4.8 105902 4.7 3.3  
## 6 Antigua And Barbuda 11.1 1611 0.1 0.9  
## internet\_users gdp\_per\_capita refugees  
## 1 13.5 551.9 2826.4  
## 2 71.8 5223.8 4.3  
## 3 49.0 4114.7 99.5  
## 4 91.6 42051.6 NA  
## 5 14.3 3437.3 70.1  
## 6 76.0 16727.0 0.2

clean\_data\_homcides <- new\_country\_data[!is.na(new\_country\_data$homicide\_rate), ]  
no\_dupes = clean\_data\_homcides%>% distinct(name, .keep\_all = TRUE)  
final\_data = no\_dupes  
names(final\_data)

## [1] "name" "homicide\_rate"   
## [3] "gdp" "urban\_population\_growth"  
## [5] "pop\_growth" "internet\_users"   
## [7] "gdp\_per\_capita" "refugees"

#Regression Analysis  
##Fit a Regression line(HOMICIDE RATE vs GDP)  
M<-lm(homicide\_rate~gdp,data=final\_data)  
##Scatterplot with line(HOMICIDE RATE vs GDP)  
library(ggplot2)  
ggplot(data=final\_data, aes(x = gdp, y = homicide\_rate)) +  
 geom\_point() +  
 scale\_x\_log10() +  
 #geom\_smooth(method = "lm", se = FALSE, col = "red") +  
 labs(x = "GDP", y = "Homicide Rate")

## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom\_point()`).



##Coefficients,standard error,R^2,RMSE,p-value(HOMICIDE RATE vs GDP)  
summary(M)

##   
## Call:  
## lm(formula = homicide\_rate ~ gdp, data = final\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -7.328 -5.949 -4.681 0.174 44.682   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 7.329e+00 8.279e-01 8.853 1.24e-15 \*\*\*  
## gdp -4.053e-07 4.047e-07 -1.001 0.318   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 10.41 on 166 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.006004, Adjusted R-squared: 1.572e-05   
## F-statistic: 1.003 on 1 and 166 DF, p-value: 0.3181

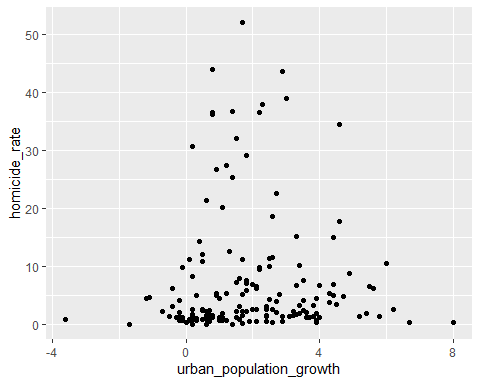
##sum of squared error reduction(HOMICIDE RATE vs GDP)  
anova(M)

## Analysis of Variance Table  
##   
## Response: homicide\_rate  
## Df Sum Sq Mean Sq F value Pr(>F)  
## gdp 1 108.6 108.58 1.0026 0.3181  
## Residuals 166 17977.0 108.30

##Confidence interval for coefficients(HOMICIDE RATE vs GDP)  
confint(M,level=0.95)

## 2.5 % 97.5 %  
## (Intercept) 5.694376e+00 8.963328e+00  
## gdp -1.204372e-06 3.938306e-07

##Fit a Regression line(HOMICIDE RATE vs urban\_population\_growth)  
M<-lm(homicide\_rate~urban\_population\_growth,data=final\_data)  
##Scatterplot with line(HOMICIDE RATE vs urban\_population\_growth)  
library(ggplot2)  
ggplot(data=final\_data, aes(x = urban\_population\_growth, y = homicide\_rate)) +  
 geom\_point() +  
 #scale\_x\_log10() +  
 #geom\_smooth(method = "lm", se = FALSE, col = "red") +  
 labs(x = "urban\_population\_growth", y = "homicide\_rate")



##Coefficients,standard error,R^2,RMSE,p-value(HOMICIDE RATE vs urban\_population\_growth)  
summary(M)

##   
## Call:  
## lm(formula = homicide\_rate ~ urban\_population\_growth, data = final\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -7.799 -5.865 -4.516 0.151 44.951   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.7657 1.1819 5.724 4.71e-08 \*\*\*  
## urban\_population\_growth 0.1666 0.4539 0.367 0.714   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 10.42 on 167 degrees of freedom  
## Multiple R-squared: 0.0008061, Adjusted R-squared: -0.005177   
## F-statistic: 0.1347 on 1 and 167 DF, p-value: 0.7141

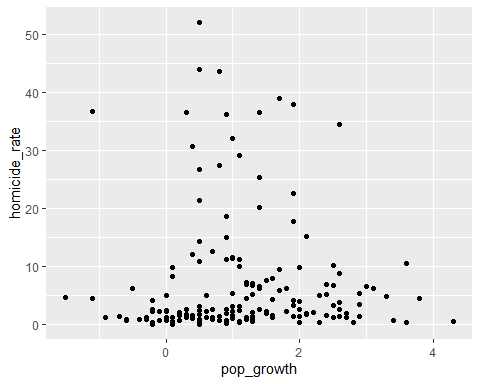
##sum of squared error reduction(HOMICIDE RATE vs urban\_population\_growth)  
anova(M)

## Analysis of Variance Table  
##   
## Response: homicide\_rate  
## Df Sum Sq Mean Sq F value Pr(>F)  
## urban\_population\_growth 1 14.6 14.619 0.1347 0.7141  
## Residuals 167 18121.5 108.512

##Confidence interval for coefficients(HOMICIDE RATE vs urban\_population\_growth)  
confint(M,level=0.95)

## 2.5 % 97.5 %  
## (Intercept) 4.4322196 9.099192  
## urban\_population\_growth -0.7295249 1.062729

##Fit a Regression line(HOMICIDE RATE vs pop\_growth)  
M<-lm(homicide\_rate~pop\_growth,data=final\_data)  
##Scatterplot with line(HOMICIDE RATE vs pop\_growth)  
library(ggplot2)  
ggplot(data=final\_data, aes(x = pop\_growth, y = homicide\_rate)) +  
 geom\_point() +  
 #scale\_x\_log10() +  
 #geom\_smooth(method = "lm", se = FALSE, col = "red") +  
 labs(x = "pop\_growth", y = "homicide\_rate")



##Coefficients,standard error,R^2,RMSE,p-value(HOMICIDE RATE vs pop\_growth)  
summary(M)

##   
## Call:  
## lm(formula = homicide\_rate ~ pop\_growth, data = final\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -7.232 -5.988 -4.577 0.123 44.845   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 7.2101 1.1722 6.151 5.5e-09 \*\*\*  
## pop\_growth -0.1108 0.7548 -0.147 0.883   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 10.42 on 167 degrees of freedom  
## Multiple R-squared: 0.000129, Adjusted R-squared: -0.005858   
## F-statistic: 0.02154 on 1 and 167 DF, p-value: 0.8835

##sum of squared error reduction(HOMICIDE RATE vs pop\_growth)  
anova(M)

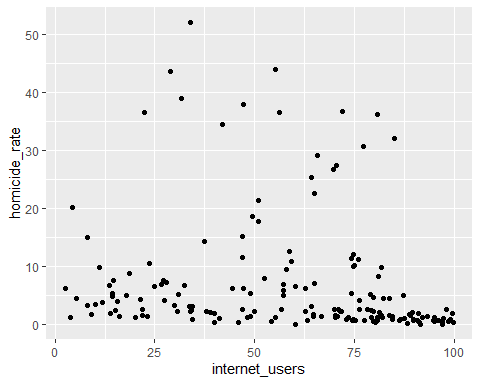
## Analysis of Variance Table  
##   
## Response: homicide\_rate  
## Df Sum Sq Mean Sq F value Pr(>F)  
## pop\_growth 1 2.3 2.339 0.0215 0.8835  
## Residuals 167 18133.8 108.585

##Confidence interval for coefficients(HOMICIDE RATE vs pop\_growth)  
confint(M,level=0.95)

## 2.5 % 97.5 %  
## (Intercept) 4.895917 9.524369  
## pop\_growth -1.600954 1.379396

##Fit a Regression line(HOMICIDE RATE vs internet\_users)  
M<-lm(homicide\_rate~internet\_users,data=final\_data)  
##Scatterplot with line(HOMICIDE RATE vs internet\_users)  
library(ggplot2)  
ggplot(data=final\_data, aes(x = internet\_users, y = homicide\_rate)) +  
 geom\_point() +  
 #scale\_x\_log10() +  
 #geom\_smooth(method = "lm", se = FALSE, col = "red") +  
 labs(x = "internet\_users", y = "homicide\_rate")

## Warning: Removed 2 rows containing missing values or values outside the scale range  
## (`geom\_point()`).



##Coefficients,standard error,R^2,RMSE,p-value(HOMICIDE RATE vs internet\_users)  
summary(M)

##   
## Call:  
## lm(formula = homicide\_rate ~ internet\_users, data = final\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -9.545 -5.244 -3.954 -0.558 43.300   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 10.89886 1.88454 5.783 3.58e-08 \*\*\*  
## internet\_users -0.06505 0.02925 -2.224 0.0275 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 10.31 on 165 degrees of freedom  
## (2 observations deleted due to missingness)  
## Multiple R-squared: 0.0291, Adjusted R-squared: 0.02321   
## F-statistic: 4.945 on 1 and 165 DF, p-value: 0.02753

##sum of squared error reduction(HOMICIDE RATE vs internet\_users)  
anova(M)

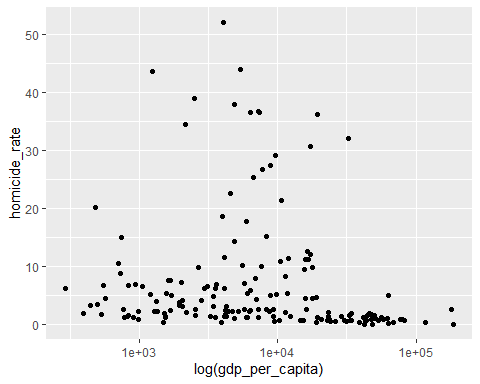
## Analysis of Variance Table  
##   
## Response: homicide\_rate  
## Df Sum Sq Mean Sq F value Pr(>F)   
## internet\_users 1 525.7 525.74 4.9447 0.02753 \*  
## Residuals 165 17543.2 106.32   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##Confidence interval for coefficients(HOMICIDE RATE vs internet\_users)  
confint(M,level=0.95)

## 2.5 % 97.5 %  
## (Intercept) 7.177930 14.619785263  
## internet\_users -0.122809 -0.007290826

##Fit a Regression line(HOMICIDE RATE vs gdp\_per\_capita)  
M<-lm(homicide\_rate~gdp\_per\_capita,data=final\_data)  
##Scatterplot with line(HOMICIDE RATE vs gdp\_per\_capita)  
library(ggplot2)  
ggplot(data=final\_data, aes(x = gdp\_per\_capita, y = homicide\_rate)) +  
 geom\_point() +  
 scale\_x\_log10() +  
 #geom\_smooth(method = "lm", se = FALSE, col = "red") +  
 labs(x = "log(gdp\_per\_capita)", y = "homicide\_rate")

## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom\_point()`).



##Coefficients,standard error,R^2,RMSE,p-value(HOMICIDE RATE vs gdp\_per\_capita)  
summary(M)

##   
## Call:  
## lm(formula = homicide\_rate ~ gdp\_per\_capita, data = final\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8.137 -5.835 -3.625 0.370 43.774   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 8.562e+00 9.353e-01 9.154 < 2e-16 \*\*\*  
## gdp\_per\_capita -8.282e-05 2.921e-05 -2.835 0.00515 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 10.19 on 166 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.04619, Adjusted R-squared: 0.04045   
## F-statistic: 8.039 on 1 and 166 DF, p-value: 0.005147

##sum of squared error reduction(HOMICIDE RATE vs gdp\_per\_capita)  
anova(M)

## Analysis of Variance Table  
##   
## Response: homicide\_rate  
## Df Sum Sq Mean Sq F value Pr(>F)   
## gdp\_per\_capita 1 835.4 835.42 8.0393 0.005147 \*\*  
## Residuals 166 17250.2 103.92   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

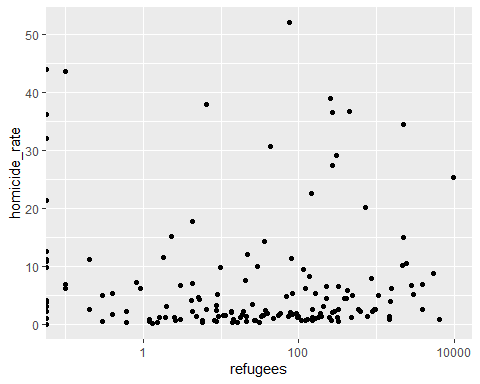
##Confidence interval for coefficients(HOMICIDE RATE vs gdp\_per\_capita)  
confint(M,level=0.95)

## 2.5 % 97.5 %  
## (Intercept) 6.7152827952 1.040836e+01  
## gdp\_per\_capita -0.0001404967 -2.515077e-05

##Fit a Regression line(HOMICIDE RATE vs refugees)  
M<-lm(homicide\_rate~refugees,data=final\_data)  
##Scatterplot with line(HOMICIDE RATE vs refugees)  
library(ggplot2)  
ggplot(data=final\_data, aes(x = refugees, y = homicide\_rate)) +  
 geom\_point() +  
 scale\_x\_log10() +  
 #geom\_smooth(method = "lm", se = FALSE, col = "red") +  
 labs(x = "refugees", y = "homicide\_rate")

## Warning in scale\_x\_log10(): log-10 transformation introduced infinite values.

## Warning: Removed 10 rows containing missing values or values outside the scale range  
## (`geom\_point()`).



##Coefficients,standard error,R^2,RMSE,p-value(HOMICIDE RATE vs refugees)  
summary(M)

##   
## Call:  
## lm(formula = homicide\_rate ~ refugees, data = final\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -11.231 -5.499 -4.235 0.193 45.379   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.5557500 0.8626627 7.599 2.53e-12 \*\*\*  
## refugees 0.0008730 0.0006857 1.273 0.205   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 10.2 on 157 degrees of freedom  
## (10 observations deleted due to missingness)  
## Multiple R-squared: 0.01022, Adjusted R-squared: 0.003914   
## F-statistic: 1.621 on 1 and 157 DF, p-value: 0.2049

##sum of squared error reduction(HOMICIDE RATE vs refugees)  
anova(M)

## Analysis of Variance Table  
##   
## Response: homicide\_rate  
## Df Sum Sq Mean Sq F value Pr(>F)  
## refugees 1 168.7 168.74 1.6209 0.2049  
## Residuals 157 16344.7 104.11

##Confidence interval for coefficients(HOMICIDE RATE vs refugees)  
confint(M,level=0.95)

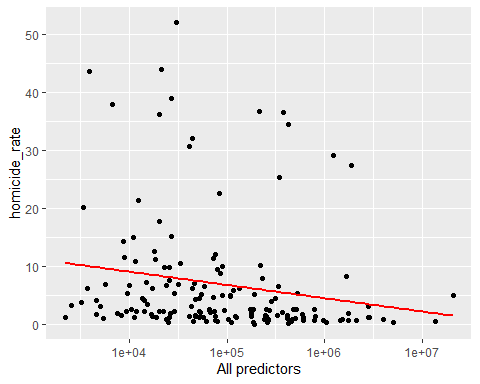
## 2.5 % 97.5 %  
## (Intercept) 4.8518279000 8.259672053  
## refugees -0.0004814254 0.002227474

#Fit a Multilinear Regression Model  
M<-lm(homicide\_rate~gdp+urban\_population\_growth+pop\_growth+internet\_users+gdp\_per\_capita+refugees,data=final\_data)  
  
##Scatterplot with line(HOMICIDE RATE vs all predictors)  
library(ggplot2)  
ggplot(data=final\_data, aes(x = gdp+urban\_population\_growth+pop\_growth+internet\_users+gdp\_per\_capita+refugees, y = homicide\_rate)) +  
 geom\_point() +  
 scale\_x\_log10() +  
 geom\_smooth(method = "lm", se = FALSE, col = "red") +  
 labs(x = "All predictors", y = "homicide\_rate")

## `geom\_smooth()` using formula = 'y ~ x'

## Warning: Removed 11 rows containing non-finite outside the scale range  
## (`stat\_smooth()`).

## Warning: Removed 11 rows containing missing values or values outside the scale range  
## (`geom\_point()`).



##Coefficients,standard error,R^2,RMSE,p-value(HOMICIDE RATE vs all predictors)  
summary(M)

##   
## Call:  
## lm(formula = homicide\_rate ~ gdp + urban\_population\_growth +   
## pop\_growth + internet\_users + gdp\_per\_capita + refugees,   
## data = final\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -11.844 -5.557 -3.442 0.040 42.891   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.030e+01 3.053e+00 3.373 0.000944 \*\*\*  
## gdp -2.636e-07 4.047e-07 -0.651 0.515803   
## urban\_population\_growth 4.411e-01 7.990e-01 0.552 0.581670   
## pop\_growth -1.362e+00 1.396e+00 -0.976 0.330743   
## internet\_users -3.196e-02 4.303e-02 -0.743 0.458763   
## gdp\_per\_capita -5.472e-05 3.781e-05 -1.447 0.149868   
## refugees 6.846e-04 6.967e-04 0.983 0.327385   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 10.14 on 151 degrees of freedom  
## (11 observations deleted due to missingness)  
## Multiple R-squared: 0.05824, Adjusted R-squared: 0.02082   
## F-statistic: 1.556 on 6 and 151 DF, p-value: 0.1637

##sum of squared error reduction(HOMICIDE RATE vs all predictors)  
anova(M)

## Analysis of Variance Table  
##   
## Response: homicide\_rate  
## Df Sum Sq Mean Sq F value Pr(>F)   
## gdp 1 94.7 94.67 0.9202 0.33895   
## urban\_population\_growth 1 22.4 22.42 0.2180 0.64126   
## pop\_growth 1 47.0 47.02 0.4570 0.50004   
## internet\_users 1 461.8 461.83 4.4892 0.03575 \*  
## gdp\_per\_capita 1 235.4 235.43 2.2885 0.13243   
## refugees 1 99.3 99.32 0.9655 0.32738   
## Residuals 151 15534.4 102.88   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##Confidence interval for coefficients(HOMICIDE RATE vs all predictors)  
confint(M,level=0.95)

## 2.5 % 97.5 %  
## (Intercept) 4.265988e+00 1.632905e+01  
## gdp -1.063164e-06 5.359752e-07  
## urban\_population\_growth -1.137437e+00 2.019699e+00  
## pop\_growth -4.120028e+00 1.395931e+00  
## internet\_users -1.169868e-01 5.305912e-02  
## gdp\_per\_capita -1.294214e-04 1.997881e-05  
## refugees -6.919735e-04 2.061102e-03