9750 Final Project

2024-11-23

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  
write.csv(final\_data, "C:/Users/Taylor's Home/OneDrive - The City University of New York/Fall 2024/STA 9750 Tools/STA 9750 Final Project/Global Data/Final Data.csv")

# Perform a multiple ANOVA test  
anova\_model = aov(homicide\_rate ~ urban\_population\_growth + pop\_growth + internet\_users + gdp + refugees, 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 0.289 0.5918   
## pop\_growth 1 33 33.4 0.323 0.5709   
## internet\_users 1 507 507.2 4.895 0.0284 \*  
## gdp 1 55 55.4 0.535 0.4656   
## refugees 1 119 119.2 1.151 0.2851   
## Residuals 152 15750 103.6   
## ---  
## 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: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