Miniproject

Poorvi Ashok

2023-02-13

require(dplyr)

## Loading required package: dplyr

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

require(tidyr)

## Loading required package: tidyr

require(purrr)

## Loading required package: purrr

require(ggplot2)

## Loading required package: ggplot2

require(usdm)

## Loading required package: usdm

## Loading required package: sp

## Loading required package: raster

##   
## Attaching package: 'raster'

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

require(knitr)

## Loading required package: knitr

require(sjPlot)

## Loading required package: sjPlot

## Learn more about sjPlot with 'browseVignettes("sjPlot")'.

require(ggeffects)

## Loading required package: ggeffects

require(cowplot)

## Loading required package: cowplot

##   
## Attaching package: 'cowplot'

## The following object is masked from 'package:ggeffects':  
##   
## get\_title

## The following objects are masked from 'package:sjPlot':  
##   
## plot\_grid, save\_plot

The dataset containing selected Environmental, Social and Governance (ESG) variables (mentioned below) for countries was downloaded from the World Bank Sovereign ESG Data Portal and input. Additionally, datasets containing other socio-political variables (mentioned below) for countries were also downloaded from the World Bank DataBank. Climate Change Performance Index (CCPI) data for countries was obtained and input from the CCPI website. The CCPI data was obtained in a pdf format and then converted into a csv format.

World Bank DataBank: <https://databank.worldbank.org/source/environment-social-and-governance?preview=on>. CCPI Website: <https://ccpi.org/download/climate-change-performance-index-2023/>

Variables obtained from World Bank ESG Data Portal: (INSERT) Socio-economic variables obtained from World Bank DataBank: National Income per Capita (US), Population Growth (%), Sex Ratio

#setting working directory  
setwd("/Users/tp/Desktop/Miniproject")  
  
#input ESG data  
data\_t <- read.csv('ESG.csv')  
summary(data\_t)  
str(data\_t)  
  
#input CCPI data  
CCPI <- read.csv('CCPI New.csv')  
summary(CCPI)  
str(CCPI)  
  
#input national income per capita  
NI <- read.csv('National Income Per Capita.csv')  
summary(NI)  
str(NI)  
  
#input GDP per capita   
GC <- read.csv('GDP Per Capita.csv')  
summary(GC)  
str(GC)  
  
#input population growth dataset  
PG <- read.csv('Population Growth.csv')  
summary(PG)  
str(PG)  
  
#input sex ratio dataset  
SR <- read.csv('Sex Ratio.csv')  
summary(SR)  
str(SR)

Data wrangling was performed on the ESG dataset to contain the country name and corresponding values for the chosen ESG variables.

#ESG Data  
#selecting only country name, ESG variable name and 2019 data from ESG dataset  
data <- subset(data\_t, select=-c(Country.Code, Series.Code,   
 X2013..YR2013.:X2018..YR2018.,   
 X2020..YR2020.:X2022..YR2022.))  
summary(data)  
head(data)  
  
#Pivoting the dataframe   
dataf <- pivot\_wider(data, names\_from = 'Series.Name',   
 values\_from = 'X2019..YR2019.')  
summary(dataf)  
head(dataf)  
  
#renaming the colnames  
names1 <- c('Country', 'Welfare\_Aggregate\_Growth', 'GDP\_Growth',   
 'Digital\_Use', 'Labour\_Participation\_Rate', 'Life\_Expectancy',   
 'Literacy\_Rate', 'Ageing\_Population', 'Population\_Density',   
 'Poverty\_Ratio', 'Labour\_Gender\_Ratio', 'School\_Enrollment',   
 'School\_Enrollment\_Gender\_Ratio', 'Unemployment')  
colnames(dataf) <- names1  
head(dataf)  
  
#Removing variables with low number of observations  
dataf <- subset(dataf, select=-c(Welfare\_Aggregate\_Growth, Literacy\_Rate, Poverty\_Ratio))  
summary(dataf)  
head(dataf)

The ESG variables ‘Welfare Aggregate Growth’, ‘Literacy Rate’ and ‘Poverty Ratio’ were found to contain a low number of observations and were hence removed from the analysis.

Data wrangling was performed on the National Income dataset to contain the country name and the corresponding national income values.

#Selecting onyl 2019 values  
NI <- subset(NI, select=c('Data.Source', 'X.61'))  
head(NI)  
  
#Removing top rows (are either empty or not relevant)  
NI <- NI[-(1:4),]  
head(NI)  
  
#renaming the columns  
names3 <- c('Country', 'National\_Income')  
colnames(NI) <- names3  
head(NI)

Data wrangling was performed on the GDP per Capita dataset to contain the country name and the corresponding GDP per Capita values.

#Selecting only 2019 values  
GC <- subset(GC, select=c('Data.Source', 'X.61'))  
head(GC)  
  
#Removing top rows (are either empty or not relevant)  
GC <- GC[-(1:4),]  
head(GC)  
  
#renaming the columns  
names3 <- c('Country', 'GDP\_Capita')  
colnames(GC) <- names3  
head(GC)

Data wrangling was performed on the population growth dataset to contain the country name and the corresponding population growth values.

#Selecting only 2019 values  
PG <- subset(PG, select=c('Country.Name', 'X2019'))  
summary(PG)  
head(PG)  
  
#Changing column names   
names4 <- c('Country', 'Population\_Growth')  
colnames(PG) <- names4  
head(PG)

Data wrangling was performed on the Sex Ratio dataset to contain the country name and the corresponding sex ratio values.

#Selecting only 2019 values  
SR <- subset(SR, select=c('Country.Name', 'X2019..YR2019.'))  
head(SR)  
  
#Changing column names  
names6 <- c('Country', 'Sex\_Ratio')  
colnames(SR) <- names6  
head(SR)

Data wrangling was performed on the CCPI dataset to contain the country name and the corresponding CCPI values. ‘Rank\_Actual’ was chosen instead of ‘Rank’ since it was consistent and started from 1. The choice of starting ‘Rank’ from 4 was made by the CCPI. This was done to reflect the lack of climate change mitigation performance by all countries, deeming none of their CCPI scores high enough to occupy the first 3 ranks.

#Selecting only country, actual rank & score  
#CCPI <- subset(CCPI, select=c('X.1', 'X.2', 'X.3'))  
CCPI <- subset(CCPI, select=c('Actual.Rank', 'Country', 'Score..'))  
summary(CCPI)  
head(CCPI)  
  
#Removing the first 2 rows (Redundant rows)  
#CCPI <- CCPI[-c(1:2),]  
CCPI <- CCPI[-c(1:3),]  
head(CCPI)  
  
#Renaming the column names  
names5 <- c('CCPI\_Rank', 'Country', 'CCPI\_Score')  
colnames(CCPI) <- names5  
head(CCPI)

Post data wrangling the ESG, CCPI, National Income, GDP per Capita, Population Growth and Sex Ratio datasets were merged to form the final dataset. The values were converted to numeric type variable and the data was cleaned to remove all NA values.

#Merging the ESG and CCPI datasets  
Total1<- merge(dataf, CCPI, by.x='Country', all.x=T)  
#Merging Population Growth and Sex Ratio datsets  
Total2 <- merge(SR, PG, by.x='Country', all.x=T)  
#Merging National Income and GDP per Capita Datsets  
Total3 <- merge(NI, GC, by.x='Country', all.x=T)  
#Merging all datsets to create the final datset  
Final <- merge(Total1, Total2, by.x='Country', all.x=T)  
Final <- merge(Final, Total3, by.x='Country', all.x=T)  
summary(Final)  
head(Final)  
str(Final)  
  
#Converting all values in the dataframe to numeric  
Final[,2:17] <- as.data.frame(sapply(Final[,2:17], as.numeric))

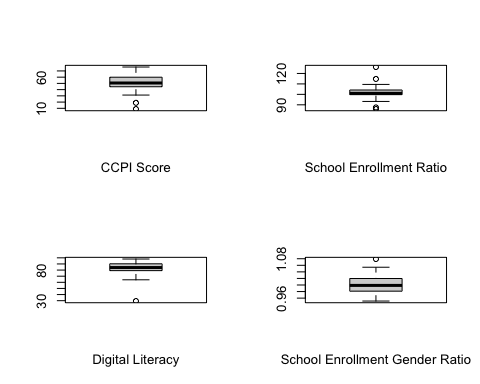
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

str(Final)  
  
#omitting all NAs in the dataframe   
Final <- na.omit(Final)

The final dataset contained the values of 14 socio-economic and demographic variables, and CCPI rank and score for 48 countries.

A linear model was chosen to study the effect of socio-economic and demographic variables on the CCPI scores of the countries. All the variables were checked for outliers.

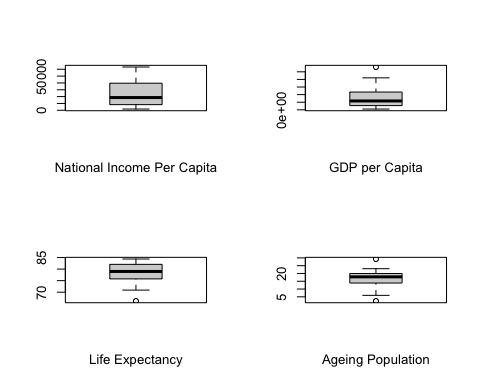
par(mfrow=c(2,2))  
  
#Plotting outliers   
boxplot(Final$CCPI\_Score)  
title(xlab='CCPI Score')  
#No outliers  
#2 outliers  
  
boxplot(Final$School\_Enrollment)  
title(xlab='School Enrollment Ratio')  
#Few outliers (logical)  
  
boxplot(Final$Digital\_Use)  
title(xlab='Digital Literacy')  
#2 outliers (logical)  
  
boxplot(Final$School\_Enrollment\_Gender\_Ratio)  
title(xlab='School Enrollment Gender Ratio')



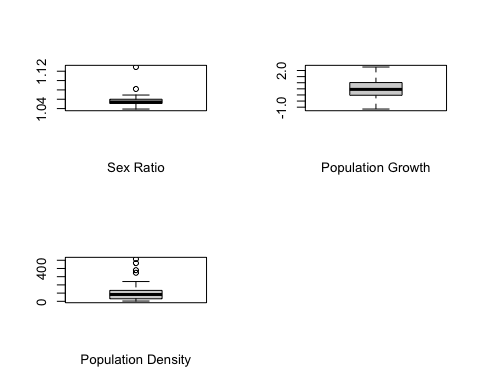
#1 outlier (logical)  
  
boxplot(Final$Labour\_Participation\_Rate)  
title(xlab = 'Labour Force Participation Rate')  
#2 outliers (logical)  
  
boxplot(Final$Unemployment)  
title(xlab = 'Unemployment')  
#4 outliers (logical)  
  
boxplot(Final$Labour\_Gender\_Ratio)  
title(xlab='Gender Ratio of Labour Participation')  
#Few outliers (logical)  
  
boxplot(Final$GDP\_Growth)  
title(xlab='GDP Growth')



#No outliers  
  
boxplot(Final$National\_Income)  
title(xlab='National Income Per Capita')  
#No outliers  
  
boxplot(Final$GDP\_Capita)  
title(xlab='GDP per Capita')  
#1 outlier (logical)  
  
boxplot(Final$Life\_Expectancy)  
title(xlab='Life Expectancy')  
#1 outliers (logical)  
  
boxplot(Final$Ageing\_Population)  
title(xlab='Ageing Population')



#No outliers  
  
boxplot(Final$Sex\_Ratio)  
title(xlab='Sex Ratio')  
#3 outliers (logical)  
  
boxplot(Final$Population\_Growth)  
title(xlab='Population Growth')  
#No outliers  
  
boxplot(Final$Population\_Density)  
title(xlab='Population Density')  
#Few outliers (logical)

 Outliers were found for the variables ‘School\_Enrollment’, ‘Digital\_Use’, ‘School\_Enrollment\_Gender\_Ratio’, ‘Labour\_Participation\_Rate’, ‘Unemployment’, ‘Labour\_Gender\_Ratio’, ‘GDP\_per\_Capita’, ‘Life\_Expectancy’, ‘Sex\_Ratio’, and ‘Population\_Density’. The outliers were checked for abnormalities. All were found to be logical and hence were retained.

The homogeneity of the data was checked for all variables.

while (!is.null(dev.list())) dev.off()  
  
#Checking Homogeneity of Variances   
var(Final$CCPI\_Score)

## [1] 172.5436

var(Final$Digital\_Use, na.rm=T)

## [1] 147.9619

var(Final$School\_Enrollment, na.rm=T)

## [1] 37.93159

var(Final$School\_Enrollment\_Gender\_Ratio, na.rm=T)

## [1] 0.0007715917

var(Final$Labour\_Participation\_Rate, na.rm=T)

## [1] 55.00855

var(Final$Unemployment, na.rm=T)

## [1] 20.65665

var(Final$Labour\_Gender\_Ratio, na.rm=T)

## [1] 196.2944

var(Final$GDP\_Growth, na.rm=T)

## [1] 2.835286

var(Final$GDP\_Capita, na.rm=T)

## [1] 630010927

var(Final$National\_Income, na.rm=T)

## [1] 315769194

var(Final$Life\_Expectancy, na.rm=T)

## [1] 16.61607

var(Final$Ageing\_Population, na.rm=T)

## [1] 31.96966

var(Final$Sex\_Ratio, na.rm=T)

## [1] 0.0001837798

var(Final$Population\_Growth, na.rm=T)

## [1] 0.5391263

var(Final$Population\_Density, na.rm=T)

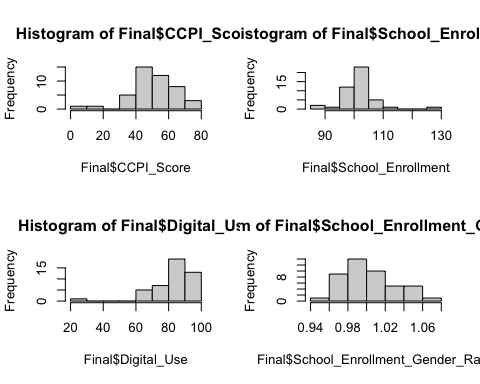
## [1] 14041.07

#Doesn't satisfy conditions of homogeneity   
#Needs to be scaled

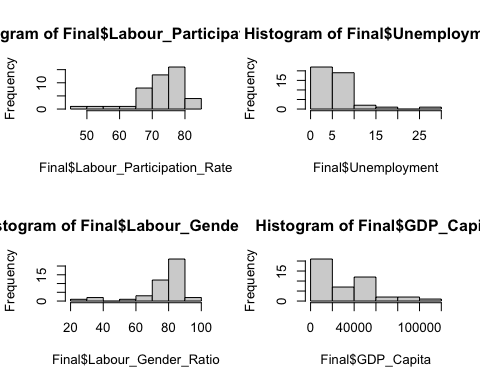
The conditions for homogeneity were not satisfied by the variables. The variables would need to be scaled.

The conditions of normality were checked for the variables.

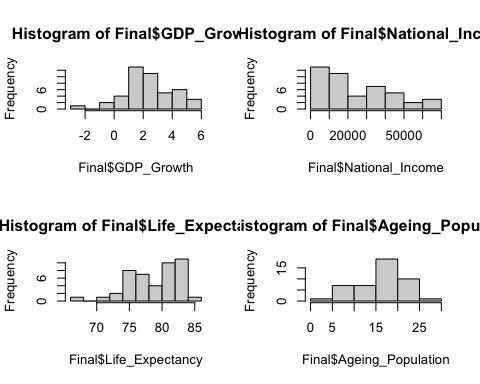
par(mfrow=c(2,2))  
  
#Checking for normal distribution  
  
hist(Final$CCPI\_Score)  
#normal   
  
hist(Final$School\_Enrollment)  
#normal (barring outliers)  
  
hist(Final$Digital\_Use)  
#left-skewed data   
  
hist(Final$School\_Enrollment\_Gender\_Ratio)



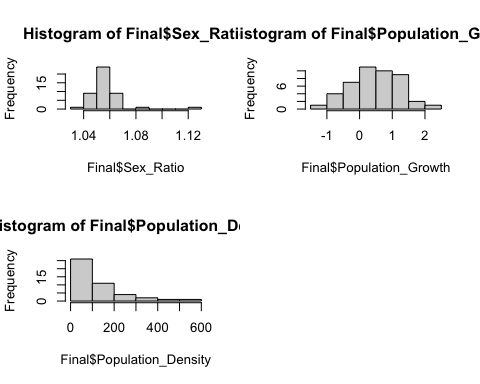
#normal  
  
hist(Final$Labour\_Participation\_Rate)  
#left-skewed  
  
hist(Final$Unemployment)  
#right-skewed  
  
hist(Final$Labour\_Gender\_Ratio)  
#left skewed  
  
hist(Final$GDP\_Capita)



#right skewed  
  
hist(Final$GDP\_Growth)  
#normal   
  
hist(Final$National\_Income)  
#right-skewed  
  
hist(Final$Life\_Expectancy)  
#left skewed  
  
hist(Final$Ageing\_Population)



#normal   
  
hist(Final$Sex\_Ratio)  
#normal (barring outliers)   
  
hist(Final$Population\_Growth)  
#normal   
  
hist(Final$Population\_Density)  
#right skewed

 Variables ‘Digital\_Use’, ‘Labour\_Participation\_Rate’, ‘Labour\_Gender\_Ratio’ and ‘Life\_Expectancy’ were found to be left skewed while variables ‘Unemployment’, ‘GDP\_Capita’, ‘National\_Income’ and ’Population\_Density were found to be right-skewed. The rest of the variables were found to contain a roughly normal distribution. The variables with skewed data should be kept in mind while checking the residual plots of the linear model.

The variables were checked for data points containing 0s.

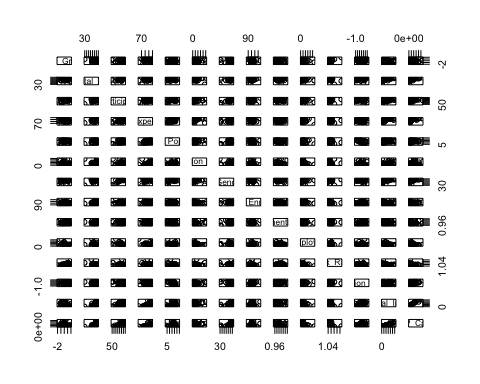
colSums(Final==0)

## Country GDP\_Growth   
## 0 0   
## Digital\_Use Labour\_Participation\_Rate   
## 0 0   
## Life\_Expectancy Ageing\_Population   
## 0 0   
## Population\_Density Labour\_Gender\_Ratio   
## 0 0   
## School\_Enrollment School\_Enrollment\_Gender\_Ratio   
## 0 0   
## Unemployment CCPI\_Rank   
## 0 0   
## CCPI\_Score Sex\_Ratio   
## 0 0   
## Population\_Growth National\_Income   
## 0 0   
## GDP\_Capita   
## 0

There are no 0s in the dataframe.

The collinearity among the explanatory variables was checked using the functions corr and variance inflation factor (VIF).

#Checking for collinearity  
pairs(Final[-c(1,12,13)])



cor(Final[-c(1,12,13)], y=NULL, use='complete.obs')

## GDP\_Growth Digital\_Use  
## GDP\_Growth 1.0000000000 -0.20720640  
## Digital\_Use -0.2072063973 1.00000000  
## Labour\_Participation\_Rate 0.0238803983 0.53866352  
## Life\_Expectancy -0.1320723167 0.57049420  
## Ageing\_Population -0.0356755851 0.29855412  
## Population\_Density -0.0004154503 -0.21195682  
## Labour\_Gender\_Ratio 0.0602128836 0.48818912  
## School\_Enrollment -0.2432022225 0.28442888  
## School\_Enrollment\_Gender\_Ratio 0.0482096279 -0.06653369  
## Unemployment -0.2632311371 -0.17726721  
## Sex\_Ratio 0.5019083569 -0.37683096  
## Population\_Growth -0.0991813826 0.25123276  
## National\_Income -0.2230152265 0.65930429  
## GDP\_Capita -0.1501512937 0.62076595  
## Labour\_Participation\_Rate Life\_Expectancy  
## GDP\_Growth 0.02388040 -0.13207232  
## Digital\_Use 0.53866352 0.57049420  
## Labour\_Participation\_Rate 1.00000000 0.54657069  
## Life\_Expectancy 0.54657069 1.00000000  
## Ageing\_Population 0.59593529 0.61438995  
## Population\_Density -0.09557704 0.19475995  
## Labour\_Gender\_Ratio 0.86021317 0.49161635  
## School\_Enrollment 0.07263707 0.19096192  
## School\_Enrollment\_Gender\_Ratio 0.06869801 0.02079928  
## Unemployment -0.36773422 -0.27929634  
## Sex\_Ratio 0.06348018 -0.06215024  
## Population\_Growth -0.15453211 0.01708315  
## National\_Income 0.51517939 0.75074259  
## GDP\_Capita 0.45007212 0.70554217  
## Ageing\_Population Population\_Density  
## GDP\_Growth -0.03567559 -0.0004154503  
## Digital\_Use 0.29855412 -0.2119568186  
## Labour\_Participation\_Rate 0.59593529 -0.0955770434  
## Life\_Expectancy 0.61438995 0.1947599520  
## Ageing\_Population 1.00000000 0.1721311594  
## Population\_Density 0.17213116 1.0000000000  
## Labour\_Gender\_Ratio 0.58494671 -0.1611621555  
## School\_Enrollment -0.19039914 -0.0424058200  
## School\_Enrollment\_Gender\_Ratio -0.01634249 0.0909750030  
## Unemployment -0.18587499 -0.1791762664  
## Sex\_Ratio -0.06816287 0.0791463873  
## Population\_Growth -0.65658562 -0.0687189457  
## National\_Income 0.39195557 0.1810979132  
## GDP\_Capita 0.31196838 0.1826649692  
## Labour\_Gender\_Ratio School\_Enrollment  
## GDP\_Growth 0.0602128836 -0.24320222  
## Digital\_Use 0.4881891163 0.28442888  
## Labour\_Participation\_Rate 0.8602131673 0.07263707  
## Life\_Expectancy 0.4916163505 0.19096192  
## Ageing\_Population 0.5849467085 -0.19039914  
## Population\_Density -0.1611621555 -0.04240582  
## Labour\_Gender\_Ratio 1.0000000000 0.01851343  
## School\_Enrollment 0.0185134304 1.00000000  
## School\_Enrollment\_Gender\_Ratio 0.1537444308 0.13503556  
## Unemployment -0.0702721900 0.10447007  
## Sex\_Ratio -0.0006855421 -0.12282727  
## Population\_Growth -0.1869067516 0.40275920  
## National\_Income 0.5164424710 0.20125218  
## GDP\_Capita 0.4754140678 0.18363863  
## School\_Enrollment\_Gender\_Ratio Unemployment  
## GDP\_Growth 0.04820963 -0.26323114  
## Digital\_Use -0.06653369 -0.17726721  
## Labour\_Participation\_Rate 0.06869801 -0.36773422  
## Life\_Expectancy 0.02079928 -0.27929634  
## Ageing\_Population -0.01634249 -0.18587499  
## Population\_Density 0.09097500 -0.17917627  
## Labour\_Gender\_Ratio 0.15374443 -0.07027219  
## School\_Enrollment 0.13503556 0.10447007  
## School\_Enrollment\_Gender\_Ratio 1.00000000 -0.02651364  
## Unemployment -0.02651364 1.00000000  
## Sex\_Ratio 0.03009615 -0.18237154  
## Population\_Growth 0.09773795 0.09138224  
## National\_Income -0.03767599 -0.20699512  
## GDP\_Capita -0.02250732 -0.20077837  
## Sex\_Ratio Population\_Growth National\_Income  
## GDP\_Growth 0.5019083569 -0.09918138 -0.22301523  
## Digital\_Use -0.3768309639 0.25123276 0.65930429  
## Labour\_Participation\_Rate 0.0634801835 -0.15453211 0.51517939  
## Life\_Expectancy -0.0621502351 0.01708315 0.75074259  
## Ageing\_Population -0.0681628672 -0.65658562 0.39195557  
## Population\_Density 0.0791463873 -0.06871895 0.18109791  
## Labour\_Gender\_Ratio -0.0006855421 -0.18690675 0.51644247  
## School\_Enrollment -0.1228272676 0.40275920 0.20125218  
## School\_Enrollment\_Gender\_Ratio 0.0300961488 0.09773795 -0.03767599  
## Unemployment -0.1823715366 0.09138224 -0.20699512  
## Sex\_Ratio 1.0000000000 -0.12531436 -0.25068419  
## Population\_Growth -0.1253143553 1.00000000 0.23397158  
## National\_Income -0.2506841865 0.23397158 1.00000000  
## GDP\_Capita -0.2395601344 0.30635111 0.96468592  
## GDP\_Capita  
## GDP\_Growth -0.15015129  
## Digital\_Use 0.62076595  
## Labour\_Participation\_Rate 0.45007212  
## Life\_Expectancy 0.70554217  
## Ageing\_Population 0.31196838  
## Population\_Density 0.18266497  
## Labour\_Gender\_Ratio 0.47541407  
## School\_Enrollment 0.18363863  
## School\_Enrollment\_Gender\_Ratio -0.02250732  
## Unemployment -0.20077837  
## Sex\_Ratio -0.23956013  
## Population\_Growth 0.30635111  
## National\_Income 0.96468592  
## GDP\_Capita 1.00000000

#Calculating VIF  
vif(Final[-c(1,12,13)])

## Variables VIF  
## 1 GDP\_Growth 1.863891  
## 2 Digital\_Use 3.247070  
## 3 Labour\_Participation\_Rate 9.400128  
## 4 Life\_Expectancy 4.928831  
## 5 Ageing\_Population 10.019619  
## 6 Population\_Density 1.636627  
## 7 Labour\_Gender\_Ratio 8.635963  
## 8 School\_Enrollment 1.487977  
## 9 School\_Enrollment\_Gender\_Ratio 1.292067  
## 10 Unemployment 2.559902  
## 11 Sex\_Ratio 1.963050  
## 12 Population\_Growth 6.316458  
## 13 National\_Income 21.543952  
## 14 GDP\_Capita 19.048644

#Dropping National\_Income  
vif(Final[-c(1,12:13,16)])

## Variables VIF  
## 1 GDP\_Growth 1.723578  
## 2 Digital\_Use 3.163210  
## 3 Labour\_Participation\_Rate 9.317087  
## 4 Life\_Expectancy 4.812934  
## 5 Ageing\_Population 10.017520  
## 6 Population\_Density 1.611437  
## 7 Labour\_Gender\_Ratio 8.634815  
## 8 School\_Enrollment 1.485318  
## 9 School\_Enrollment\_Gender\_Ratio 1.284874  
## 10 Unemployment 2.559565  
## 11 Sex\_Ratio 1.961182  
## 12 Population\_Growth 6.275105  
## 13 GDP\_Capita 3.792692

#Dropping Ageing Population  
vif(Final[-c(1,6,12:13,16)])

## Variables VIF  
## 1 GDP\_Growth 1.679770  
## 2 Digital\_Use 3.094370  
## 3 Labour\_Participation\_Rate 7.914834  
## 4 Life\_Expectancy 2.827836  
## 5 Population\_Density 1.517093  
## 6 Labour\_Gender\_Ratio 8.477460  
## 7 School\_Enrollment 1.404100  
## 8 School\_Enrollment\_Gender\_Ratio 1.262686  
## 9 Unemployment 2.251964  
## 10 Sex\_Ratio 1.808052  
## 11 Population\_Growth 1.996513  
## 12 GDP\_Capita 3.764339

#Dropping Labour Gender Ratio  
vif(Final[-c(1,6,8,12:13,16)])

## Variables VIF  
## 1 GDP\_Growth 1.511270  
## 2 Digital\_Use 3.092355  
## 3 Labour\_Participation\_Rate 2.277122  
## 4 Life\_Expectancy 2.822946  
## 5 Population\_Density 1.443725  
## 6 School\_Enrollment 1.384318  
## 7 School\_Enrollment\_Gender\_Ratio 1.098506  
## 8 Unemployment 1.368492  
## 9 Sex\_Ratio 1.805543  
## 10 Population\_Growth 1.758017  
## 11 GDP\_Capita 3.109150

#All good

Choosing a corr > 0.75 to be strong, the variable pairs ‘Labour\_Participation\_Rate’ and ‘Labour\_Gender\_Ratio’; and ‘GDP\_Capita’ and ‘National\_Income’ were found to have a strong correlation. With the consideration of VIF > 5 indicating high correlation, the variables ‘National\_Income’, ‘Ageing\_Population’ and ‘Labour\_Gender\_Ratio’ were found to have a high correlation. Therefore, the variables ‘National\_Income’, ‘Ageing\_Population’ and ‘Labour\_Gender\_Ratio’ were omitted from the linear model.

A simple linear regression was performed to investigate the effects of socio-economic and demographic variables on the CCPI scores of countries. All explanatory variables were scaled. Backwards model selection using AIC was performed to obtain the simplest significant model. The model diagnostic plots were checked.

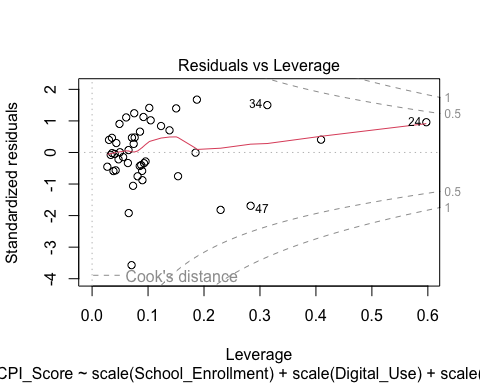
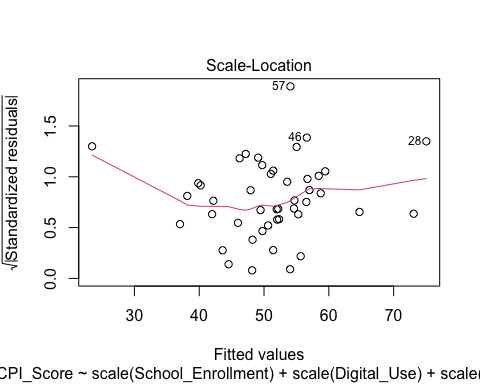
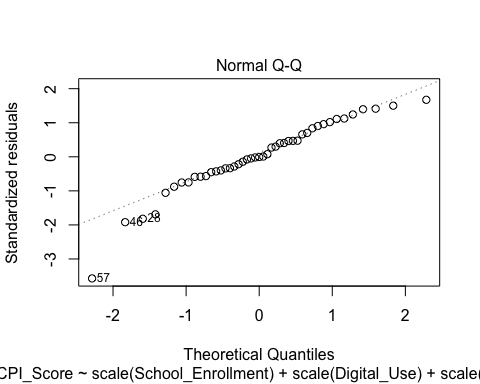
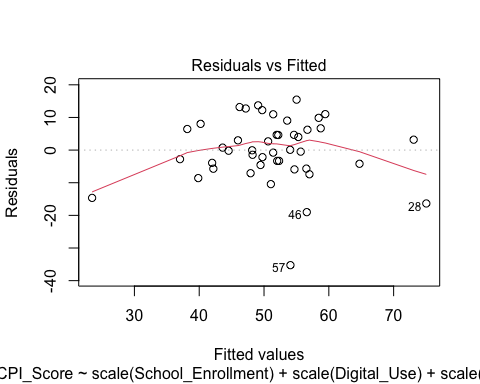
#Upper model  
M1 <- lm(CCPI\_Score ~ scale(School\_Enrollment) + scale(Digital\_Use) +   
 scale(School\_Enrollment\_Gender\_Ratio) +   
 scale(Labour\_Participation\_Rate) + scale(Unemployment) +   
 scale(GDP\_Growth) + scale(GDP\_Capita) + scale(Life\_Expectancy) +   
 scale(Sex\_Ratio) + scale(Population\_Growth) +   
 scale(Population\_Density), data = Final)  
#Backwards model selection  
M2<-step(M1, direction = "backward", scope = list(lower=~1,   
 upper=~scale(School\_Enrollment) + scale(Digital\_Use) +   
 scale(School\_Enrollment\_Gender\_Ratio) +   
 scale(Labour\_Participation\_Rate) + scale(Unemployment) +   
 scale(GDP\_Growth) + scale(GDP\_Capita) + scale(Life\_Expectancy) +   
 scale(Sex\_Ratio) + scale(Population\_Growth) +   
 scale(Population\_Density)))

## Start: AIC=224.12  
## CCPI\_Score ~ scale(School\_Enrollment) + scale(Digital\_Use) +   
## scale(School\_Enrollment\_Gender\_Ratio) + scale(Labour\_Participation\_Rate) +   
## scale(Unemployment) + scale(GDP\_Growth) + scale(GDP\_Capita) +   
## scale(Life\_Expectancy) + scale(Sex\_Ratio) + scale(Population\_Growth) +   
## scale(Population\_Density)  
##   
## Df Sum of Sq RSS AIC  
## - scale(Labour\_Participation\_Rate) 1 0.01 3842.3 222.12  
## - scale(Sex\_Ratio) 1 17.78 3860.1 222.33  
## - scale(Population\_Density) 1 35.71 3878.0 222.54  
## - scale(GDP\_Growth) 1 38.96 3881.3 222.58  
## - scale(School\_Enrollment\_Gender\_Ratio) 1 61.15 3903.5 222.83  
## - scale(Life\_Expectancy) 1 92.05 3934.4 223.19  
## - scale(Unemployment) 1 95.33 3937.6 223.22  
## <none> 3842.3 224.12  
## - scale(Digital\_Use) 1 202.36 4044.7 224.43  
## - scale(GDP\_Capita) 1 474.93 4317.2 227.37  
## - scale(School\_Enrollment) 1 1710.65 5553.0 238.69  
## - scale(Population\_Growth) 1 2006.94 5849.3 241.03  
##   
## Step: AIC=222.12  
## CCPI\_Score ~ scale(School\_Enrollment) + scale(Digital\_Use) +   
## scale(School\_Enrollment\_Gender\_Ratio) + scale(Unemployment) +   
## scale(GDP\_Growth) + scale(GDP\_Capita) + scale(Life\_Expectancy) +   
## scale(Sex\_Ratio) + scale(Population\_Growth) + scale(Population\_Density)  
##   
## Df Sum of Sq RSS AIC  
## - scale(Sex\_Ratio) 1 18.96 3861.3 220.34  
## - scale(Population\_Density) 1 36.94 3879.3 220.55  
## - scale(GDP\_Growth) 1 38.97 3881.3 220.58  
## - scale(School\_Enrollment\_Gender\_Ratio) 1 63.80 3906.1 220.86  
## - scale(Life\_Expectancy) 1 92.89 3935.2 221.20  
## - scale(Unemployment) 1 99.77 3942.1 221.28  
## <none> 3842.3 222.12  
## - scale(Digital\_Use) 1 230.50 4072.8 222.74  
## - scale(GDP\_Capita) 1 499.83 4342.2 225.63  
## - scale(School\_Enrollment) 1 1714.31 5556.6 236.72  
## - scale(Population\_Growth) 1 2349.82 6192.1 241.60  
##   
## Step: AIC=220.34  
## CCPI\_Score ~ scale(School\_Enrollment) + scale(Digital\_Use) +   
## scale(School\_Enrollment\_Gender\_Ratio) + scale(Unemployment) +   
## scale(GDP\_Growth) + scale(GDP\_Capita) + scale(Life\_Expectancy) +   
## scale(Population\_Growth) + scale(Population\_Density)  
##   
## Df Sum of Sq RSS AIC  
## - scale(GDP\_Growth) 1 23.13 3884.4 218.61  
## - scale(Population\_Density) 1 40.23 3901.5 218.81  
## - scale(School\_Enrollment\_Gender\_Ratio) 1 67.88 3929.2 219.13  
## - scale(Unemployment) 1 113.22 3974.5 219.65  
## - scale(Life\_Expectancy) 1 125.74 3987.0 219.79  
## <none> 3861.3 220.34  
## - scale(Digital\_Use) 1 212.28 4073.6 220.75  
## - scale(GDP\_Capita) 1 541.03 4402.3 224.25  
## - scale(School\_Enrollment) 1 1697.70 5559.0 234.74  
## - scale(Population\_Growth) 1 2390.79 6252.1 240.03  
##   
## Step: AIC=218.61  
## CCPI\_Score ~ scale(School\_Enrollment) + scale(Digital\_Use) +   
## scale(School\_Enrollment\_Gender\_Ratio) + scale(Unemployment) +   
## scale(GDP\_Capita) + scale(Life\_Expectancy) + scale(Population\_Growth) +   
## scale(Population\_Density)  
##   
## Df Sum of Sq RSS AIC  
## - scale(Population\_Density) 1 35.01 3919.4 217.02  
## - scale(School\_Enrollment\_Gender\_Ratio) 1 72.73 3957.2 217.45  
## - scale(Unemployment) 1 92.63 3977.1 217.67  
## - scale(Life\_Expectancy) 1 128.42 4012.8 218.08  
## <none> 3884.4 218.61  
## - scale(Digital\_Use) 1 239.32 4123.7 219.30  
## - scale(GDP\_Capita) 1 538.01 4422.4 222.45  
## - scale(School\_Enrollment) 1 1677.66 5562.1 232.77  
## - scale(Population\_Growth) 1 2375.38 6259.8 238.09  
##   
## Step: AIC=217.02  
## CCPI\_Score ~ scale(School\_Enrollment) + scale(Digital\_Use) +   
## scale(School\_Enrollment\_Gender\_Ratio) + scale(Unemployment) +   
## scale(GDP\_Capita) + scale(Life\_Expectancy) + scale(Population\_Growth)  
##   
## Df Sum of Sq RSS AIC  
## - scale(Unemployment) 1 77.33 3996.8 215.90  
## - scale(School\_Enrollment\_Gender\_Ratio) 1 78.90 3998.3 215.91  
## - scale(Life\_Expectancy) 1 110.72 4030.2 216.27  
## <none> 3919.4 217.02  
## - scale(Digital\_Use) 1 414.85 4334.3 219.54  
## - scale(GDP\_Capita) 1 653.82 4573.3 221.96  
## - scale(School\_Enrollment) 1 1701.94 5621.4 231.25  
## - scale(Population\_Growth) 1 2406.15 6325.6 236.56  
##   
## Step: AIC=215.9  
## CCPI\_Score ~ scale(School\_Enrollment) + scale(Digital\_Use) +   
## scale(School\_Enrollment\_Gender\_Ratio) + scale(GDP\_Capita) +   
## scale(Life\_Expectancy) + scale(Population\_Growth)  
##   
## Df Sum of Sq RSS AIC  
## - scale(School\_Enrollment\_Gender\_Ratio) 1 70.25 4067.0 214.68  
## - scale(Life\_Expectancy) 1 146.04 4142.8 215.51  
## <none> 3996.8 215.90  
## - scale(Digital\_Use) 1 442.66 4439.4 218.62  
## - scale(GDP\_Capita) 1 646.11 4642.9 220.64  
## - scale(School\_Enrollment) 1 1851.90 5848.7 231.03  
## - scale(Population\_Growth) 1 2364.18 6360.9 234.81  
##   
## Step: AIC=214.68  
## CCPI\_Score ~ scale(School\_Enrollment) + scale(Digital\_Use) +   
## scale(GDP\_Capita) + scale(Life\_Expectancy) + scale(Population\_Growth)  
##   
## Df Sum of Sq RSS AIC  
## - scale(Life\_Expectancy) 1 128.71 4195.7 214.08  
## <none> 4067.0 214.68  
## - scale(Digital\_Use) 1 494.07 4561.1 217.84  
## - scale(GDP\_Capita) 1 626.50 4693.5 219.13  
## - scale(School\_Enrollment) 1 1956.36 6023.4 230.35  
## - scale(Population\_Growth) 1 2308.98 6376.0 232.91  
##   
## Step: AIC=214.08  
## CCPI\_Score ~ scale(School\_Enrollment) + scale(Digital\_Use) +   
## scale(GDP\_Capita) + scale(Population\_Growth)  
##   
## Df Sum of Sq RSS AIC  
## <none> 4195.7 214.08  
## - scale(GDP\_Capita) 1 521.48 4717.2 217.35  
## - scale(Digital\_Use) 1 661.57 4857.3 218.67  
## - scale(School\_Enrollment) 1 1840.73 6036.4 228.45  
## - scale(Population\_Growth) 1 2223.47 6419.2 231.22

#Model Interpretation  
summary(M2)

##   
## Call:  
## lm(formula = CCPI\_Score ~ scale(School\_Enrollment) + scale(Digital\_Use) +   
## scale(GDP\_Capita) + scale(Population\_Growth), data = Final)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -35.255 -4.563 -0.058 6.454 15.455   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 51.092 1.527 33.465 < 2e-16 \*\*\*  
## scale(School\_Enrollment) 7.239 1.728 4.189 0.00015 \*\*\*  
## scale(Digital\_Use) -5.072 2.019 -2.511 0.01617 \*   
## scale(GDP\_Capita) 4.490 2.014 2.230 0.03145 \*   
## scale(Population\_Growth) -8.039 1.746 -4.604 4.14e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 10.24 on 40 degrees of freedom  
## Multiple R-squared: 0.4473, Adjusted R-squared: 0.3921   
## F-statistic: 8.094 on 4 and 40 DF, p-value: 7.027e-05

#Model diagnostic plots  
plot(M2)



#Residual plots look good   
#QQ (normality plot) looks good  
  
#Fitting a model without scaled variables (for plotting)  
fit <- lm(CCPI\_Score ~ School\_Enrollment + Digital\_Use + GDP\_Capita +   
 Population\_Growth, data = Final)  
  
#Predicted values of CCPI score for School Enrollment (with other effects controlled)  
G1 <- ggpredict(fit, terms = 'School\_Enrollment')  
#plotting using ggpredict (School Enrolment)  
D1 <- ggplot(G1, aes(x, predicted)) + geom\_line(colour = 'sienna1', size=1) +   
 geom\_ribbon(aes(ymin = conf.low, ymax = conf.high), alpha = 0.3) +  
 ggtitle('CCPI Score vs Ratio of School Enrolment') +   
 xlab('Ratio of School Enrolment (%)') +  
 ylab('CCPI Score') + theme(plot.title = element\_text(hjust=0.5, size = 12)) +   
 theme(axis.title = element\_text(size=10)) + ylim(40,70) + xlim(100,110)

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.  
## ℹ Please use `linewidth` instead.

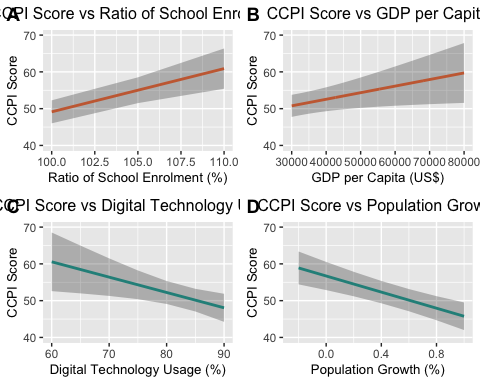
#Predicted values of CCPI score for GDP per Capita (with other effects controlled)  
G2 <- ggpredict(fit, terms = 'GDP\_Capita')  
#plotting using ggpredict (GDP per Capita)  
D2 <- ggplot(G2, aes(x, predicted)) + geom\_line(colour = "sienna1", size=1) +   
 geom\_ribbon(aes(ymin = conf.low, ymax = conf.high), alpha = 0.3) +  
 ggtitle('CCPI Score vs GDP per Capita') + xlab('GDP per Capita (US$)') +  
 ylab('CCPI Score') + theme(plot.title = element\_text(hjust=0.5, size = 12)) +   
 theme(axis.title = element\_text(size=10)) + ylim(40,70) + xlim(30000,80000)  
  
#Predicted values of CCPI score for digital usage (with other effects controlled)  
G3 <- ggpredict(fit, terms = 'Digital\_Use')  
#plotting using ggpredict (GDP per Capita)  
D3 <- ggplot(G3, aes(x, predicted)) + geom\_line(colour = "lightseagreen", size=1) +  
 geom\_ribbon(aes(ymin = conf.low, ymax = conf.high), alpha = 0.3) +  
 ggtitle('CCPI Score vs Digital Technology Usage') +   
 xlab('Digital Technology Usage (%)') +  
 ylab('CCPI Score') + theme(plot.title = element\_text(hjust=0.5, size = 12)) +   
 theme(axis.title = element\_text(size=10)) + ylim(40,70) + xlim(60,90)  
  
#Predicted values of CCPI score for population growth (with other effects controlled)  
G4 <- ggpredict(fit, terms = 'Population\_Growth')  
#plotting using ggpredict (GDP per Capita)  
D4 <- ggplot(G4, aes(x, predicted)) +   
 geom\_line(colour = "lightseagreen", size=1) +   
 geom\_ribbon(aes(ymin = conf.low, ymax = conf.high), alpha = 0.3) +  
 ggtitle('CCPI Score vs Population Growth') +   
 xlab('Population Growth (%)') +  
 ylab('CCPI Score') + theme(plot.title = element\_text(hjust=0.5, size = 12)) +   
 theme(axis.title = element\_text(size=10)) + ylim(40,70) + xlim(-0.25,1)  
  
#combining the plots  
plot\_grid(D1, D2, D3, D4, labels = "AUTO", ncol = 2)

## Warning: Removed 7 rows containing missing values (`geom\_line()`).

## Warning: Removed 607 rows containing missing values (`geom\_line()`).

## Warning: Removed 9 rows containing missing values (`geom\_line()`).

## Warning: Removed 11 rows containing missing values (`geom\_line()`).

 From the simplest significant regression model, School Enrollment, Digital Literacy, GDP per Capita and Population Growth were found to have a significant effect on CCPI Scores. All the model diagnostic plots satisfied the conditions of the linear model. Results further interpreted in the report.