Data Visualization: Crafting Scatterplots in R Programming

James Obanla

2024-02-06

Introduction:

The Corruption Perception Index thus has a deep influence on the HDI of a nation, which in turn assesses parameters like life expectancy, education, and living standards. The countries which have a very low CPI score, such as Congo and Niger, where the extent of corruption is very high, always record lower Human Development Index (HDI) outcomes, thus showing the deleterious link between corruption and human development. In such countries, press freedom is suppressed, allowing corruption to continue unabated, and overall bribes in public and private sectors siphon off the much-needed resources that are direly needed in education and healthcare sectors. Social decay, which is further detrimental to development, is fostered by this depletion of trust in public institutions and misallocation of resources. This relationship stands to be realized by policy makers in a bid to abolish corruption and create lasting prosperity.

```
data<-read.csv(file.choose(), header = TRUE)
head(data)
```

```
##   X     Country HDI.Rank   HDI CPI Region
## 1 1 Afghanistan      172 0.398  15     AP
## 2 2     Albania       70 0.739  31    ECA
## 3 3     Algeria       96 0.698  29   MENA
## 4 4      Angola      148 0.486  20    SSA
## 5 5   Argentina       45 0.797  30    AME
## 6 6     Armenia       86 0.716  26    ECA
```

Data sets Description

The datasets used in this project are from the UNDP, 2023 (https://hdr.undp.org/gender-development-index#/indicies) and Transparency International (https://www.transparency.org/en/cpi/2022/media-kit). The four quantitative data (Human Development Index (HDI), Inequality-adjusted Human Development Index (IHDI, Gender Development Index (GDI) and Corruption Perceptions Index (CPI) were visualized in relation to their respective countries and regions with year 2021 in focus. This project focuses on assessing the relationship between HDI and CPI of countries.

The full meaning of the acronyms used in the variable ‘region’ is defined as follows Americas (AME), Asia Pacific (AP), Eastern Europe & Central Asia (ECA), Western Europe & European Union(WE/EU), Middle Eastern & North Africa (MENA) and Sub-Saharan Africa (SSA).

Attaching, Checking the data features

```
str(data)
```

```
## 'data.frame':    173 obs. of  6 variables:
##  $ X       : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Country : chr  "Afghanistan" "Albania" "Algeria" "Angola" ...
##  $ HDI.Rank: int  172 70 96 148 45 86 2 19 91 53 ...
##  $ HDI     : num  0.398 0.739 0.698 0.486 0.797 0.716 0.929 0.885 0.7 0.771 ...
##  $ CPI     : int  15 31 29 20 30 26 88 78 24 73 ...
##  $ Region  : chr  "AP" "ECA" "MENA" "SSA" ...
```

```
summary(data)
```

```
##        X         Country             HDI.Rank          HDI        
##  Min.   :  1   Length:173         Min.   :  1.00   Min.   :0.2860  
##  1st Qu.: 44   Class :character   1st Qu.: 47.00   1st Qu.:0.5090  
##  Median : 87   Mode  :character   Median : 96.00   Median :0.6980  
##  Mean   : 87                      Mean   : 95.28   Mean   :0.6581  
##  3rd Qu.:130                      3rd Qu.:143.00   3rd Qu.:0.7930  
##  Max.   :173                      Max.   :187.00   Max.   :0.9430  
##       CPI           Region         
##  Min.   :15.00   Length:173        
##  1st Qu.:25.00   Class :character  
##  Median :32.00   Mode  :character  
##  Mean   :40.52                     
##  3rd Qu.:51.00                     
##  Max.   :95.00
```

```
any(is.na(data))
```

```
## [1] FALSE
```

```
any(duplicated(data))
```

```
## [1] FALSE
```

Scatter Plots (HDI as dependent variable against independent variable CPI)

Considering scatter plot to show the relationship that exist between them. The Human Development Index (HDI) serves as a concise assessment of essential aspects of human development: longevity and well-being, access to education, and attainment of a satisfactory standard of living. Meanwhile, the Corruption Perceptions Index (CPI) specifically measures how corrupt each country's public sector is perceived to be, according to experts and business people.

```
data1<- data %>% select(Region,Country,HDI,CPI) %>% na.omit() %>%  mutate(Region=Region %>% as.factor())
head(data1)
```

```
##   Region     Country   HDI CPI
## 1     AP Afghanistan 0.398  15
## 2    ECA     Albania 0.739  31
## 3   MENA     Algeria 0.698  29
## 4    SSA      Angola 0.486  20
## 5    AME   Argentina 0.797  30
## 6    ECA     Armenia 0.716  26
```

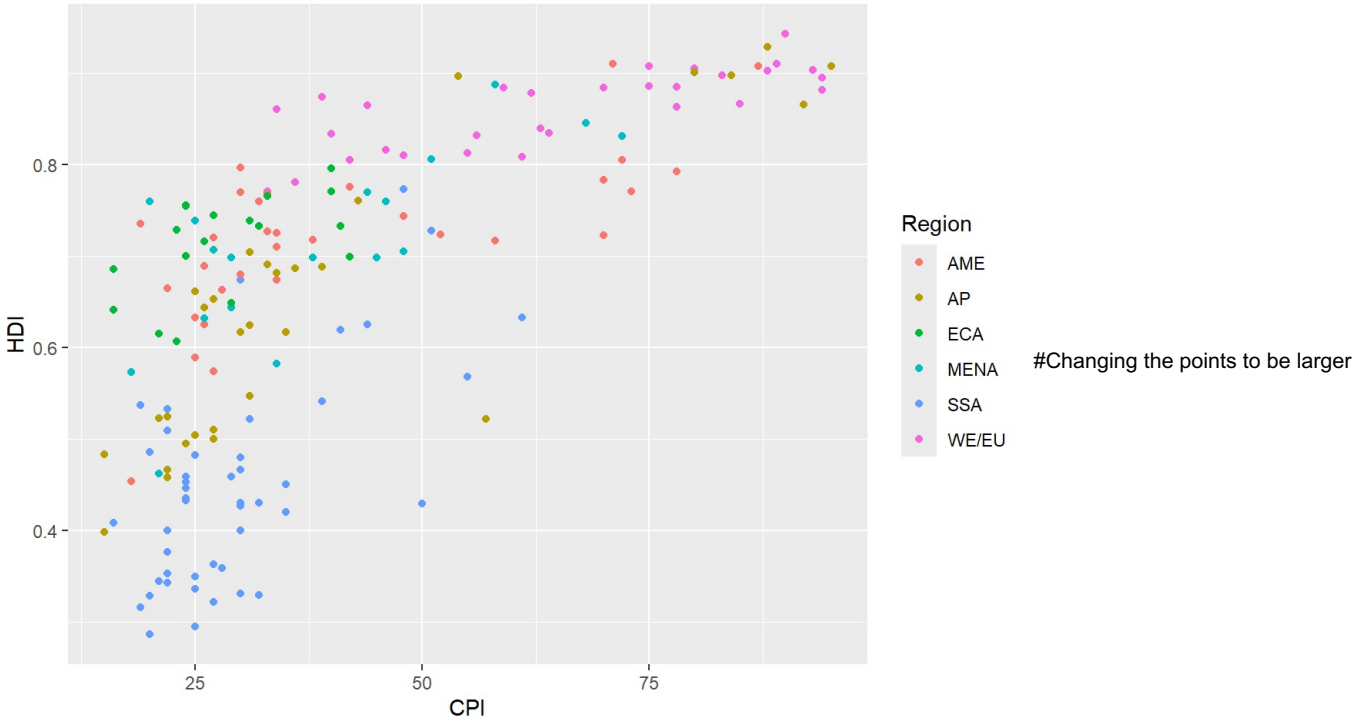
Ensuring the data integrity by checking that the selected variables from the data are clean and good to use

```
any(is.na(data1))
```

```
## [1] FALSE
```

Showing association that exist between HDI and CPI on Scatter Plot

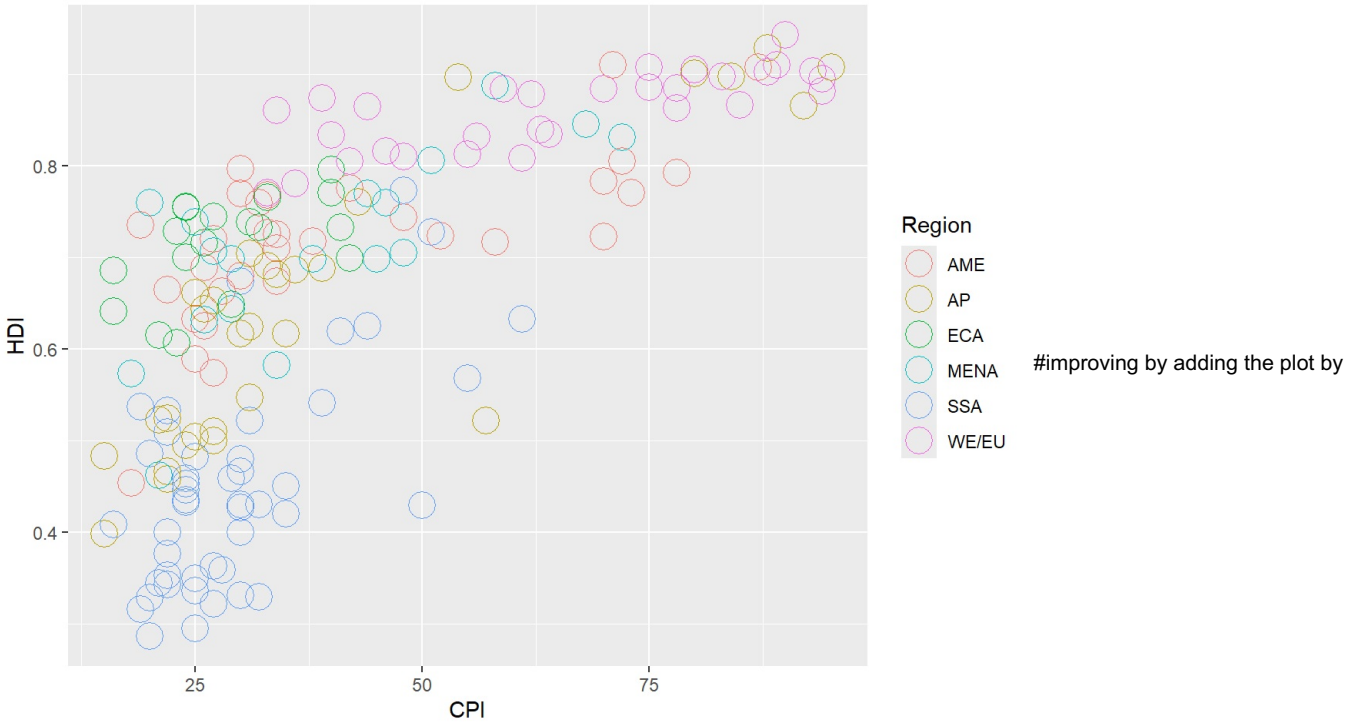
```
data1_sp <- ggplot(data1,aes(x=CPI,y=HDI,color=Region)) + geom_point()
data1_sp
```



#Changing the points to be larger

empty circles for better view of the relationship existing between the two variables, separated by colour (region).#

```
data1_sps <- ggplot(data1,aes(x=CPI,y=HDI,color=Region)) + geom_point(size=6,shape=1.5)
data1_sps
```

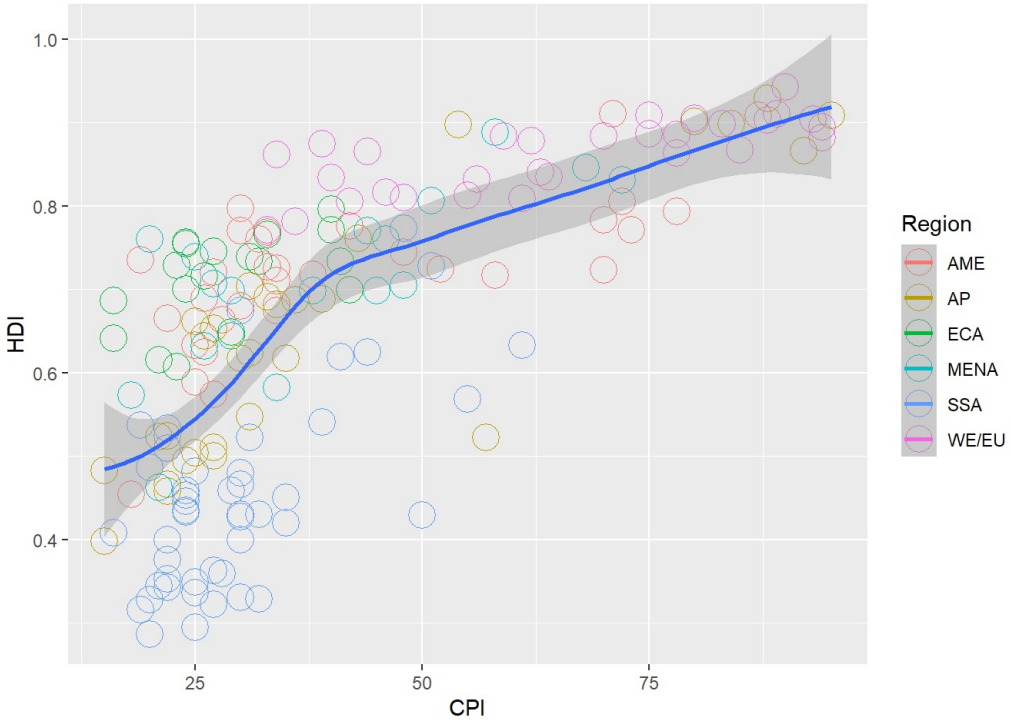


#improving by adding the plot by

showing the trend of the points#

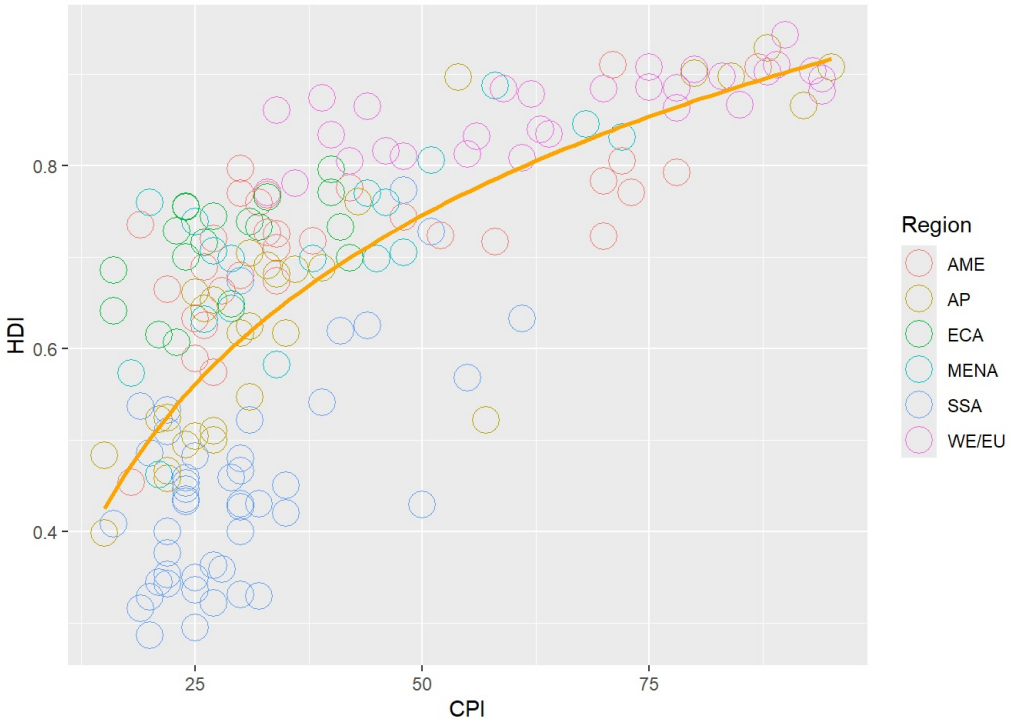
```
data1_spst<- ggplot(data1,aes(x=CPI,y=HDI,color=Region)) + geom_point(size=6,shape=1.5)+ geom_smooth(aes(group=1)
)
data1_spst
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



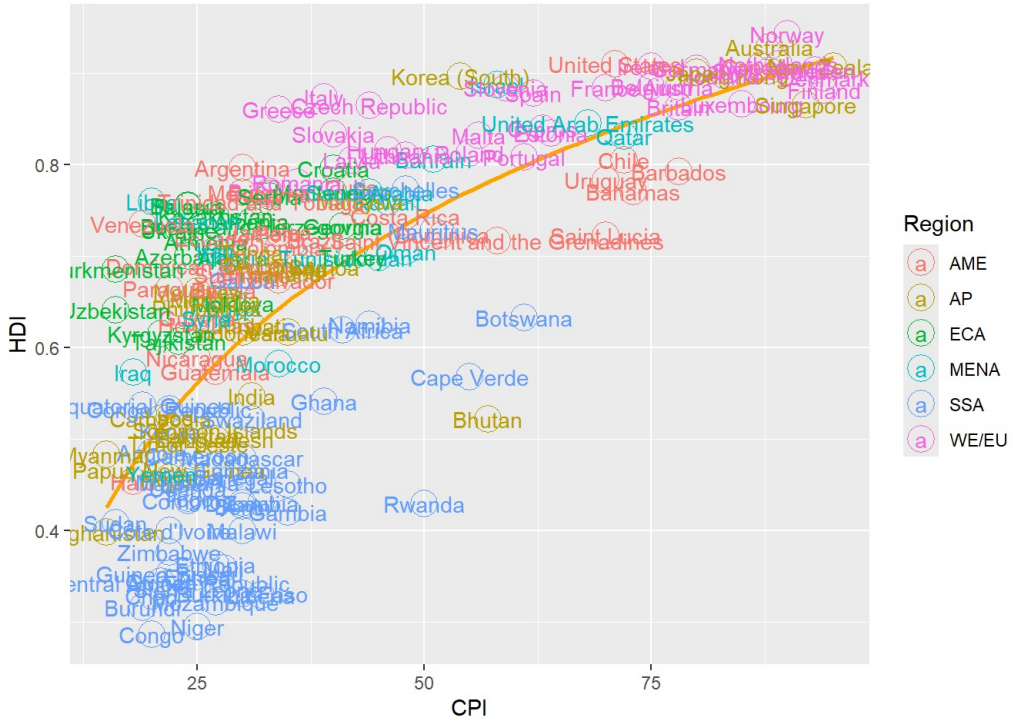
Close observation of the trend in the data set require us to introduce logarithms making linear the plot for clearer understanding of the associations existing among the variables of interest. To this end, addition of few arguments to geom_smooth data1_sps will be made.

```
data1_log_spst<-data1_sps+ geom_smooth(aes(group=1),method ='lm',formula = y~log(x),se=FALSE,color='orange')
data1_log_spst
```



Introducing geom_text to the visual to see how the text will change he face of our graph and possibly further increase information for more insight.

```
data1_log_spst_text <- data1_log_spst + geom_text(aes(label= Country))
data1_log_spst_text
```

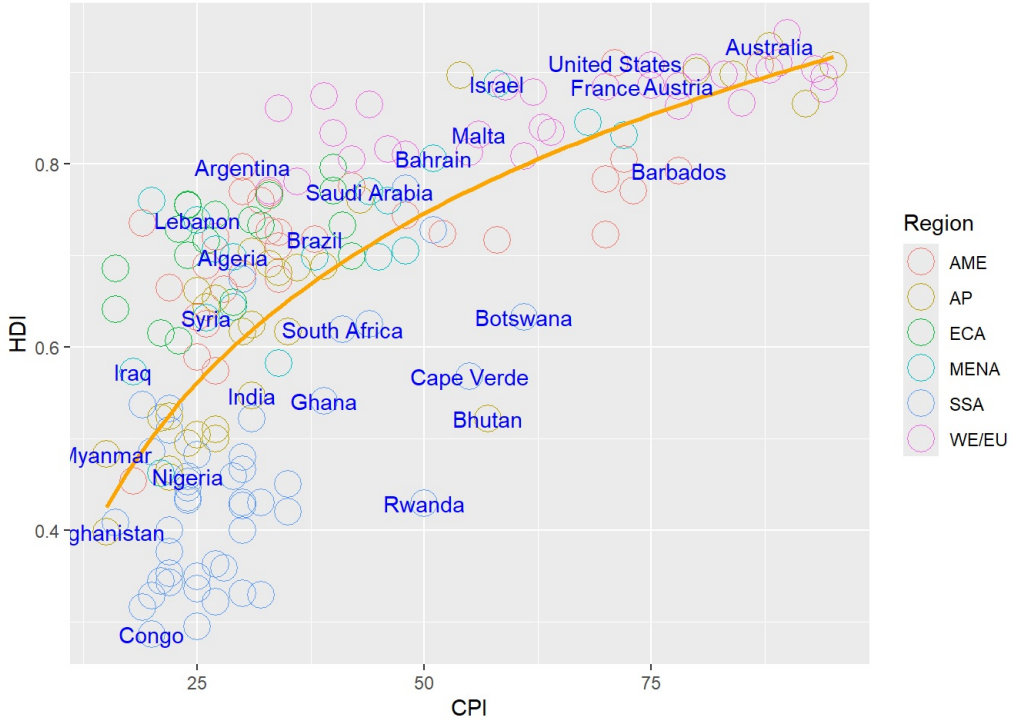


Defining country_label using selected countries.

```
country_label <- c("Nigeria", "Ghana", "Iraq", "Sudan", "Afghanistan", "Congo", "Argentina", "Brazil", "India", "
China", "South Africa", "Russia", "Botswana", "Cape Verde", "Bhutan", "Rwanda", "Barbados", "Norway", "Switzerlan
d" , "Iceland", "Australia", "Denmark", "Sweden", "Ireland", "Germany", "Netherlands", "Finland", "Singapore", "B
elgium" , "New Zealand", "Canada", "Luxembourg", "United Kingdom", "Japan", "United States", "Israel", "Malta", "
Austria", "United Arab Emirates","Spain","France", "Bahrain","Venezuela", "Myanmar", "Saudi Arabia", "Yemen", "Le
banon", "Syria", "Fiji","Algeria", "Iran", "Somalia", "Venezuela")
```

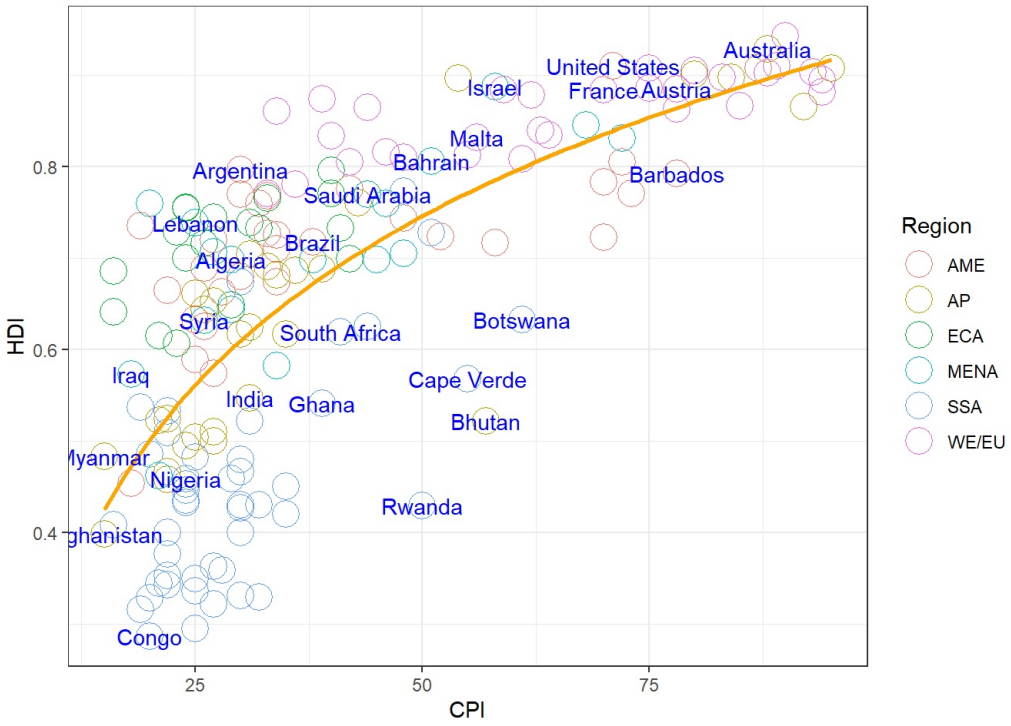
Making the labels clearer by separating the text using selected countries.

```
 data1_log_spst_text1<- data1_log_spst + geom_text(aes(label = Country), color = "blue",data = subset(data1, Coun
try %in% country_label),check_overlap = TRUE)

 data1_log_spst_text1
```

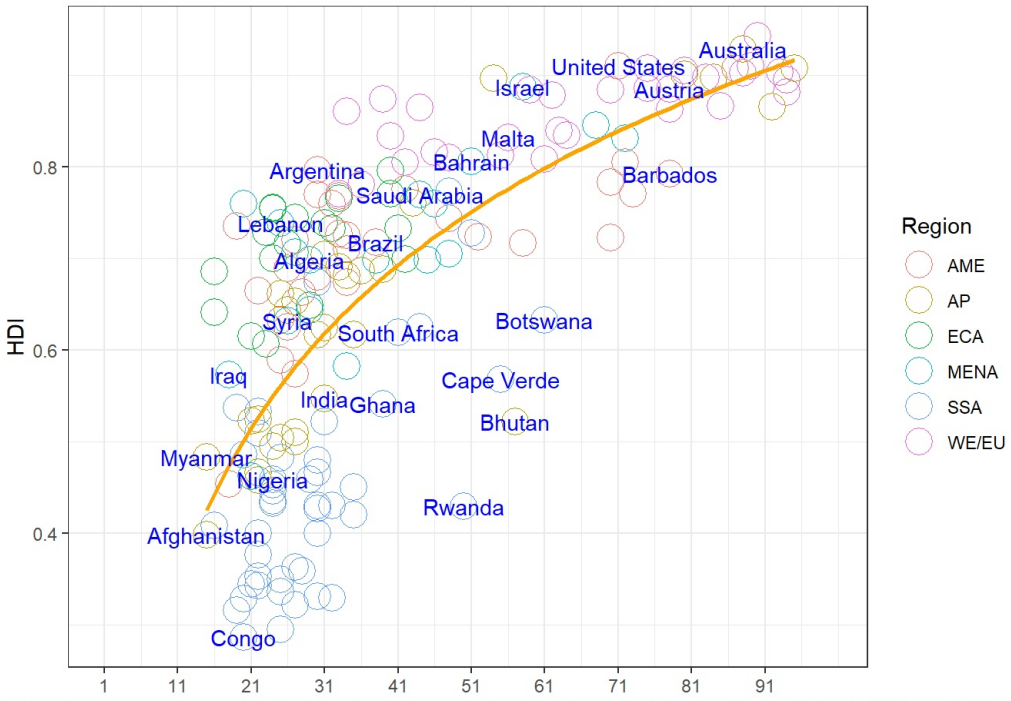


Increasing the features of the graph by adding some labels and a theme to vertical and horizontal axis.

```
data1_log_spst_text2 <- data1_log_spst_text1 + theme_bw()
data1_log_spst_text2
```

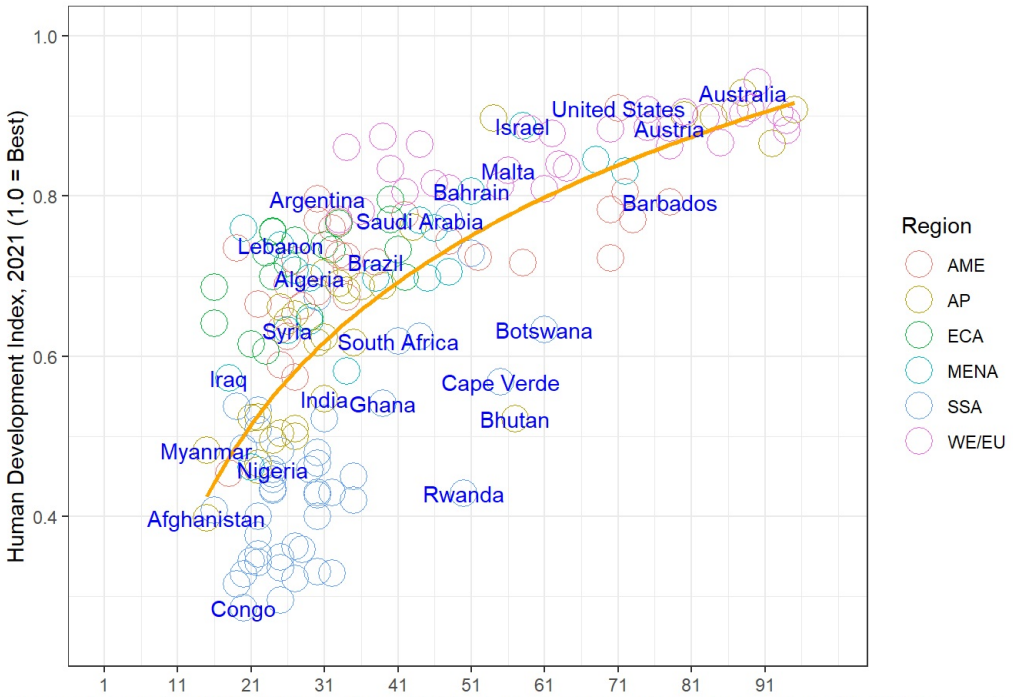



```
data1_log_spst_text3 <- data1_log_spst_text2 + scale_x_continuous(name = "(Corruption Perceptions Index, 2021 ( 0
indicating a highly corrupt country, whereas a country with 100 is very clean)",limits = c(1.0, 100.0), breaks =
seq(1.0, 100.0, 10.0))

data1_log_spst_text3
```



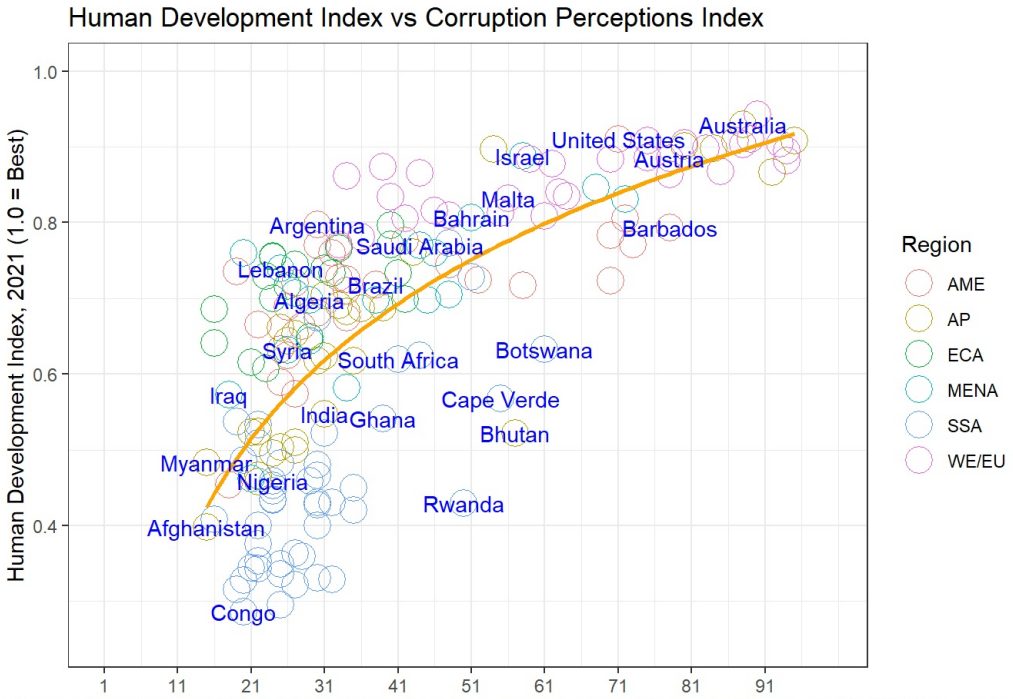
n Perceptions Index, 2021 (0 indicating a highly corrupt country, whereas a country with 100 is very clear

```
data1_log_spst_text4 <- data1_log_spst_text3 + scale_y_continuous(name = "Human Development Index, 2021 (1.0 = Be
st)", limits = c(0.25, 1.0))

data1_log_spst_text4
```



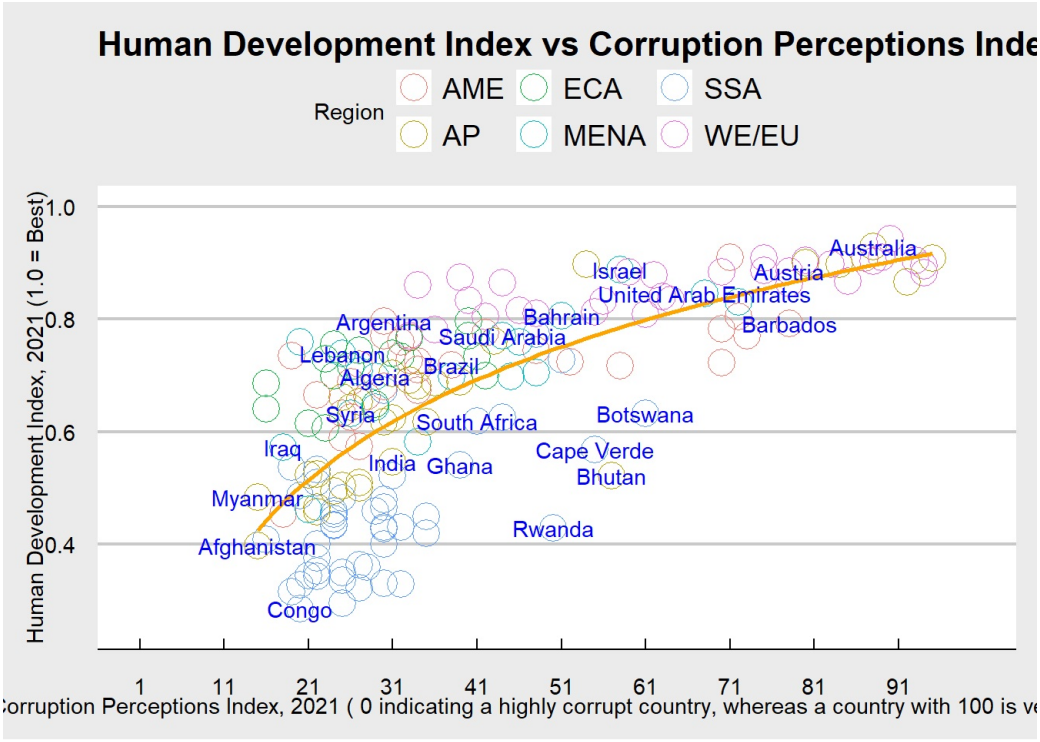
n Perceptions Index, 2021 (0 indicating a highly corrupt country, whereas a country with 100 is very clear

```
data1_log_spst_text5 <-data1_log_spst_text4 + ggtitle("Human Development Index vs Corruption Perceptions Index")
data1_log_spst_text5
```



n Perceptions Index, 2021 (0 indicating a highly corrupt country, whereas a country with 100 is very clear

```
data1_log_spst_text5 + theme_economist_white()
```



Conclusion:

A Scatter plot analysis clearly indicates the positive relationship of Corruption Perception Index (CPI) and Human Development Index(HDI). For instance, Australia and Switzerland — both high on the CPI ranks – also enjoy very good HDI scores which demonstrate effective governance structures as well as a comfortable life standard. And their place on the plot indicates that low corruption levels reduce human development outcomes. These counties have a population with more wealth than some citizens but they invest wisely in their human and social capital.UN-mediators can only provide options, after the initial alternative meetings to break deadlocks. With clear-sighted institutions which deliver public services efficiently these are where important resources end up,(improving health and education especially) decreasing overall instability within the country (HWG). However, Congo and Niger of the low ranker on CPI show an opposite case to adverse effect against HDI. Corruption is perceived to be high in these countries which results into poor governance and wasteful use of resources dampening development efforts.As can be seen in the scatter plot, greater corruption is associated with lower human development (i.e., low provision of basic services and poor quality of life), most notably by those regions located at upper left corner. The right image in the above picture is a stark contrast to that and highlights why improved governance through anti-corruption reforms can make such an immediate difference for countries facing corruption challenges.