Data Visualization Project

Chia-Yu Wei

2022-12-14

## Correlation between corruption and development

library(dplyr)

library(tidyr)  
library(ggplot2)  
library(data.table)

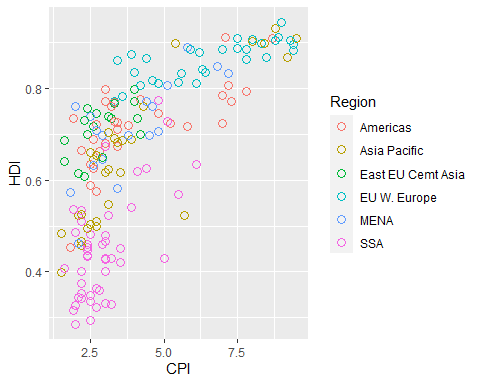
### Import the data from ggplot2

# Name the data as 'df'  
df <- fread("C:/Users/Chiayu/OneDrive/Desktop/Data Science and Machine Learning Bootcamp with R/Project/Economist\_Assignment\_Data.csv",  
 drop = 1)  
head(df)

## Country HDI.Rank HDI CPI Region  
## 1: Afghanistan 172 0.398 1.5 Asia Pacific  
## 2: Albania 70 0.739 3.1 East EU Cemt Asia  
## 3: Algeria 96 0.698 2.9 MENA  
## 4: Angola 148 0.486 2.0 SSA  
## 5: Argentina 45 0.797 3.0 Americas  
## 6: Armenia 86 0.716 2.6 East EU Cemt Asia

* Consumer Price Index (CPI) is a measure of the average change over time in the prices paid by urban consumers for a market basket of consumer goods and services.
* Human Development Index (HDI)

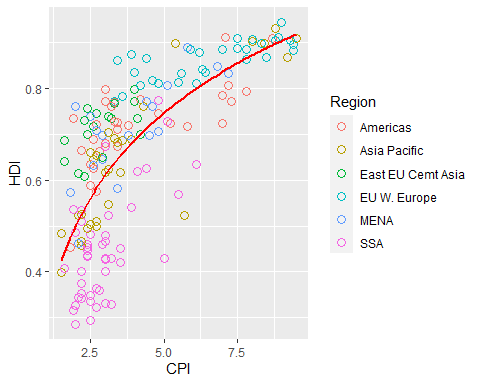
pl <- ggplot(  
 df,  
 aes(x = CPI, y = HDI, color = Region))  
pl2 <- pl +  
 geom\_point(size = 3, shape = 1) +   
 theme\_gray()  
pl2



From the information above, we can know that there is an increasing trend, which means that the counties with higher CPI also get higher HDI. Moreover, we can distinguish the counties into different region base on the color. As we can see, most of counties in SSA region have lower CPI and most counties in the EU W. Europe have higher HDI.

## Now, we’re going to create the linear regression line of the model y ~log(x)

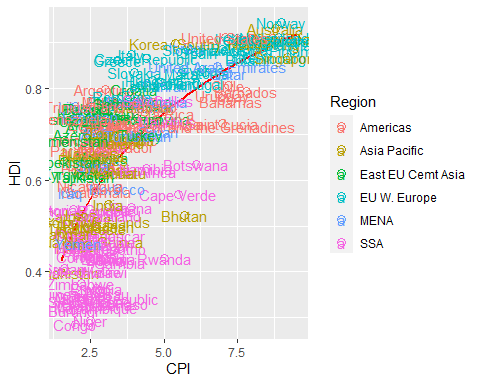
# Make linear regression  
pl3 <- pl2 + geom\_smooth(aes(group = 1),  
 method = 'lm',   
 formula = y ~ log(x),  
 se = FALSE,  
 color = 'red')  
pl3



From the linear regression line, the relation between CPI and HDI is positive.

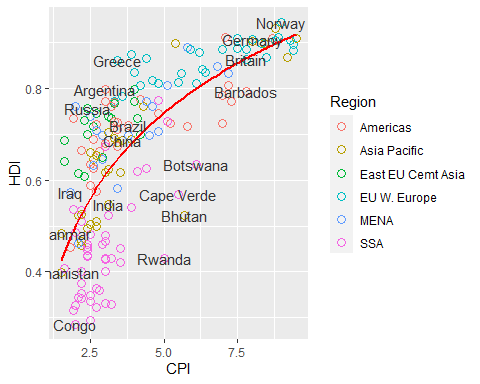
### Using ‘geom\_text’ to add text to the plot.

pl3 + geom\_text(aes(label = Country))



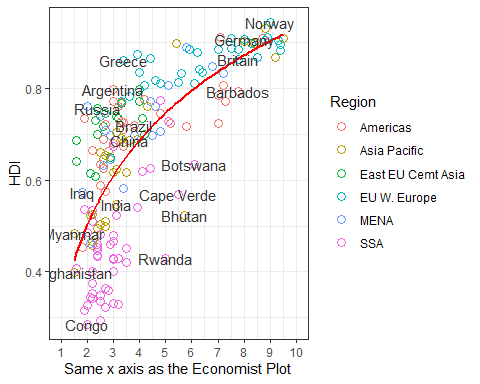
# Label some counties.

Total\_label <- c("Russia", "Venezuela", "Iraq", "Myanmar", "Sudan",  
 "Afghanistan", "Congo", "Greece", "Argentina", "Brazil",  
 "India", "Italy", "China", "South Africa", "Spane",  
 "Botswana", "Cape Verde", "Bhutan", "Rwanda", "France",  
 "United States", "Germany", "Britain", "Barbados", "Norway",  
 "Japan", "New Zealand", "Singapore")  
pl4 <- pl3 + geom\_text(aes(label = Country),  
 color = "gray20",  
 data = subset(df, Country %in% Total\_label),  
 check\_overlap = TRUE)  
pl4



# Add limit of x-axis and y-axis to make the graph better to be read

pl4 + theme\_bw() + scale\_x\_continuous(name = 'Same x axis as the Economist Plot',  
 limits = c(1, 10),  
 breaks = 1:10)



**Reference**

The Economist online (2011) Corrosive corruption – A correlation between corruption and development

<https://www.economist.com/graphic-detail/2011/12/02/corrosive-corruption>