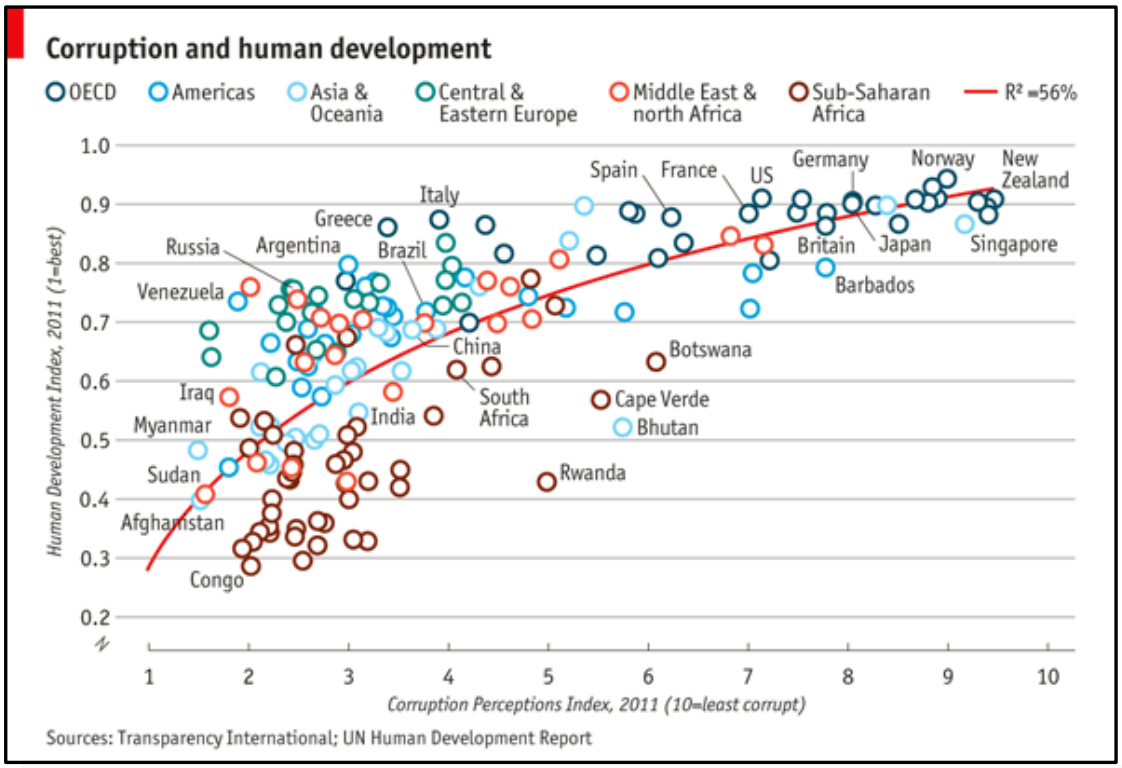
Data Visualization Project with R

In this project I will trying to recreating this plot from [The Economist:](#)




```
library(ggplot2)

library(ggplot2movies)
library(ggthemes)

library(ggrepel)

library(dplyr)

df = read.csv('Economist_Assignment_Data.csv')
df = df[-c(1)]
print(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
```

```
pl = ggplot(df,aes(CPI,HDI)) + geom_point(aes(color =  Region),size
= 5,shape =1)
pl = pl + geom_smooth(aes(group=1),method = 'lm',
                      formula = y ~ log(x),se = FALSE,color = 'red')
```

Data Visualization Project with R

```
points2lable = c("Russia", "Venezuela", "Iraq", "Myanmar", "Sudan",
                 "Afghanistan", "Congo", "Greece", "Argentina",
                 "Brazil", "India", "Italy", "China", "South Africa,
                 "Spain", "Botswana", "Cape Verde", "Bhutan",
                 "Rwanda", "France", "United States", "Germany",
                 "Britain", "Barbados", "Norway", "Japan",
                 "New Zealand", "Singapore")
pl = pl + geom_text(aes(label=Country),data = subset(df,Country %in%
points2lable),check_overlap = TRUE)

pl = pl + theme_bw() + scale_x_continuous(name =
           'Corruption Perceptions Index, 2011 (10=least corrupt)',
            limits = c(.9,10.5),breaks = c(1:10)) +
            scale_y_continuous(name =
           'Human Development Index, 2011 (1=best)',
            limits = c(.2,1.0))
pl = pl + ggtitle('Corruption and Human Development') +
theme_economist_white()
print(pl)
```