Data Visualization Project

Loading Data:

In this section, we are going to load the libraries and data as needed in order to perform The Economist data graph as made on the internet.

```
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.3.2

library(data.table)

## Warning: package 'data.table' was built under R version 3.3.2

library(ggthemes)

## Warning: package 'ggthemes' was built under R version 3.3.2

df <- fread('Economist_Assignment_Data.csv', drop=1)
summary(df)

##    Country             HDI.Rank           HDI               CPI        

##  Length:173         Min.   :  1.00   Min.   :0.2860   Min.   :1.500  

##  Class :character   1st Qu.: 47.00   1st Qu.:0.5090   1st Qu.:2.500  

##  Mode  :character   Median : 96.00   Median :0.6980   Median :3.200  

##                     Mean   : 95.28   Mean   :0.6581   Mean   :4.052  

##                     3rd Qu.:143.00   3rd Qu.:0.7930   3rd Qu.:5.100  

##                     Max.   :187.00   Max.   :0.9430   Max.   :9.500  

##     Region         
##  Length:173        
##  Class :character  
##  Mode  :character  
##                    
##                    
##                    

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
```

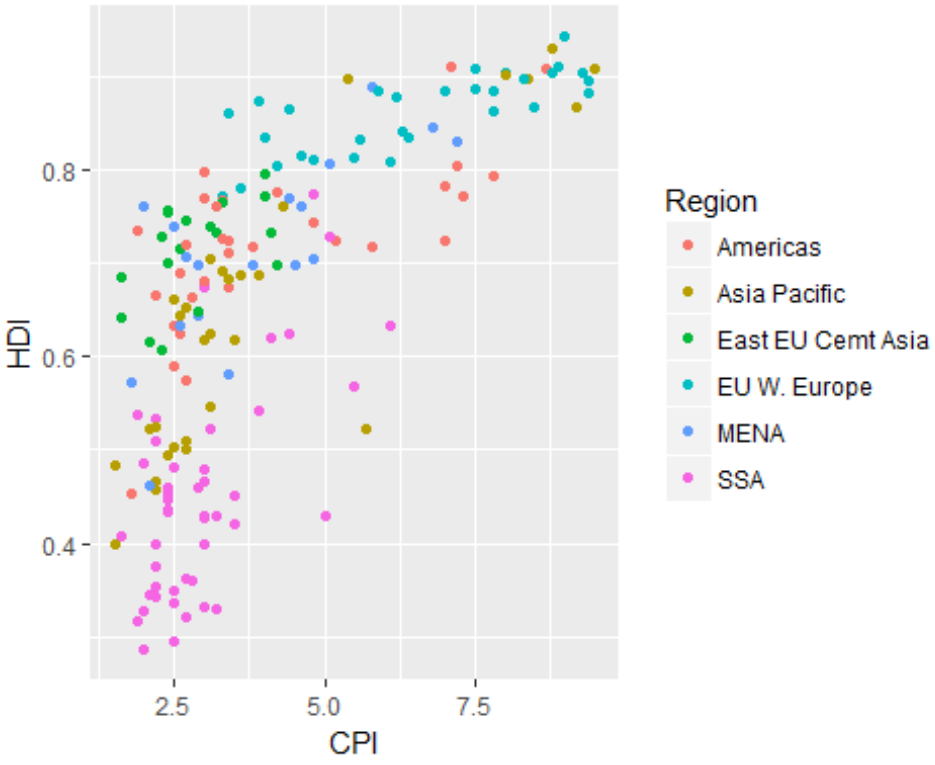
```
tail(df)
```

```
##       Country HDI.Rank   HDI CPI            Region
## 1: Uzbekistan      115 0.641 1.6 East EU Cemt Asia
## 2:    Vanuatu      125 0.617 3.5      Asia Pacific
## 3:  Venezuela       73 0.735 1.9          Americas
## 4:      Yemen      154 0.462 2.1              MENA
## 5:     Zambia      164 0.430 3.2               SSA
## 6:   Zimbabwe      173 0.376 2.2               SSA
```

Creating a Scatterplot:

Based on the information that is given, we are going to create a scatterplot based on the region that is given.

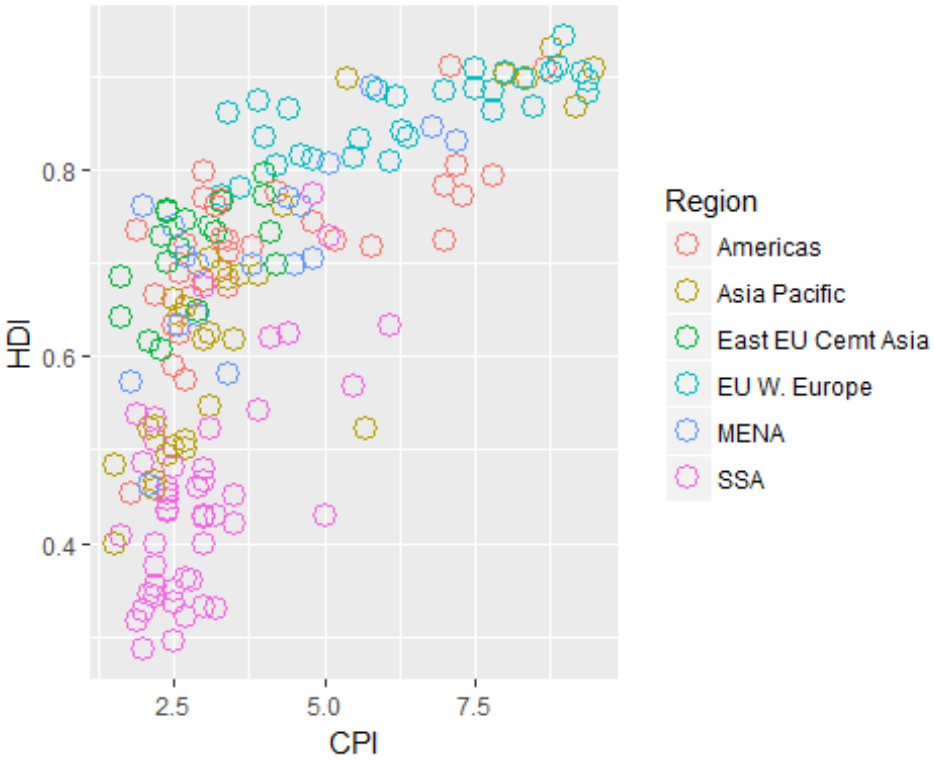
```
plot <- ggplot(df,aes(x=CPI,y=HDI,color=Region)) +
geom_point(aes(color=factor(Region)))
print(plot)
```



Linear Models:

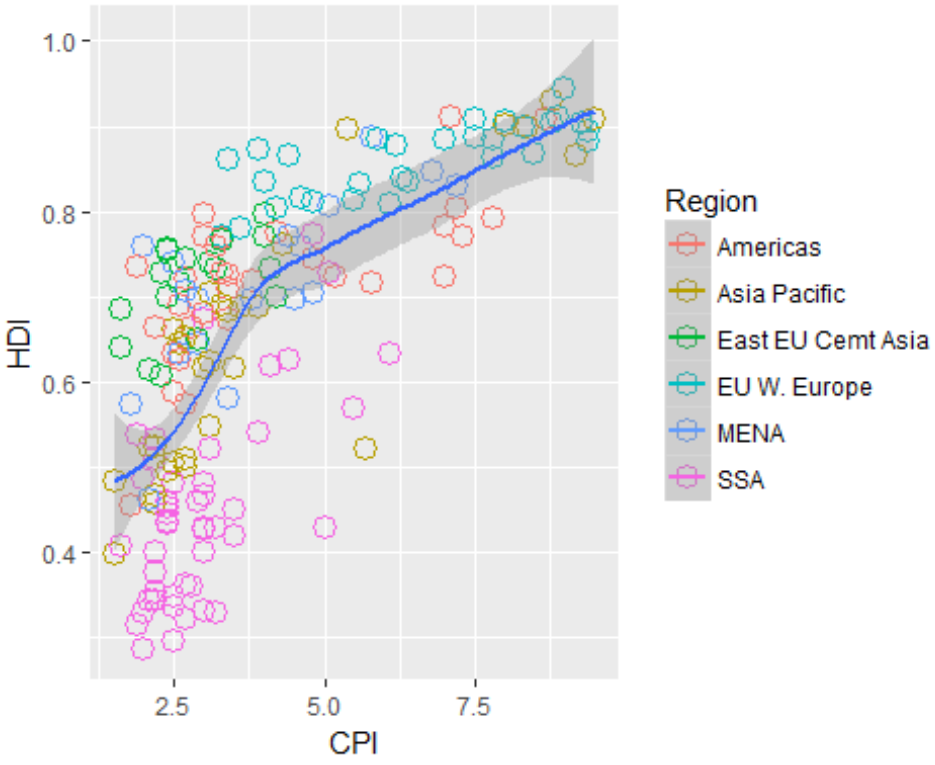
In this section, we are going to use different trend lines to determine the API and HDI differences.

```
plot <- ggplot(df,aes(x=CPI,y=HDI,color=Region)) +
geom_point(size=4,shape=1)
print(plot)
```

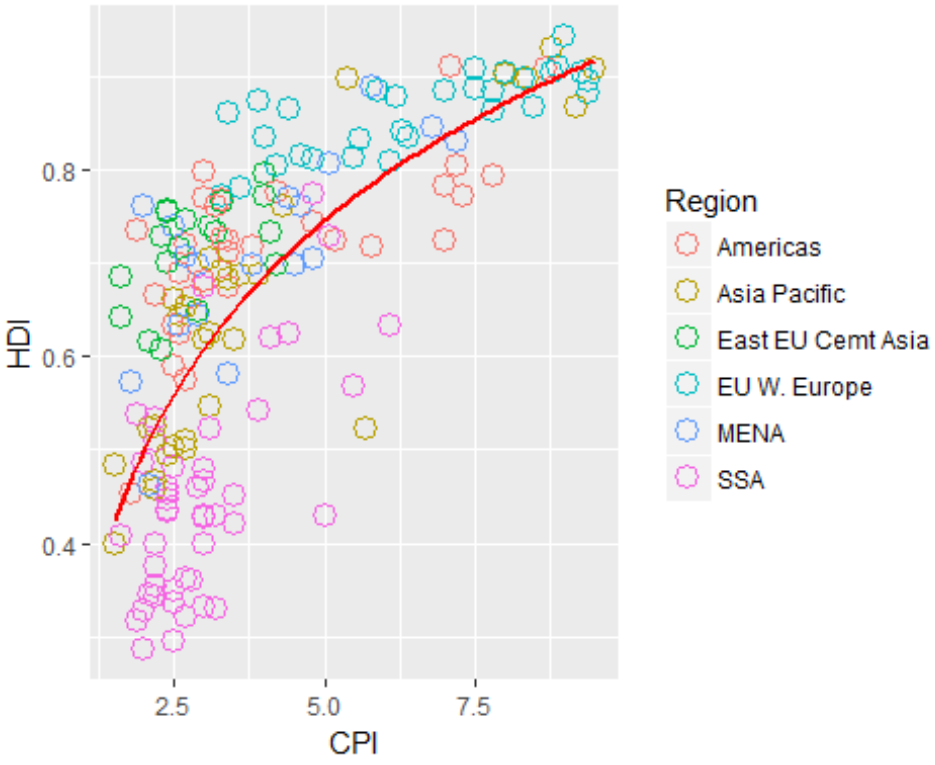



```
plot2 <- plot + geom_smooth(aes(group=1))
print(plot2)

## `geom_smooth()` using method = 'loess'
```



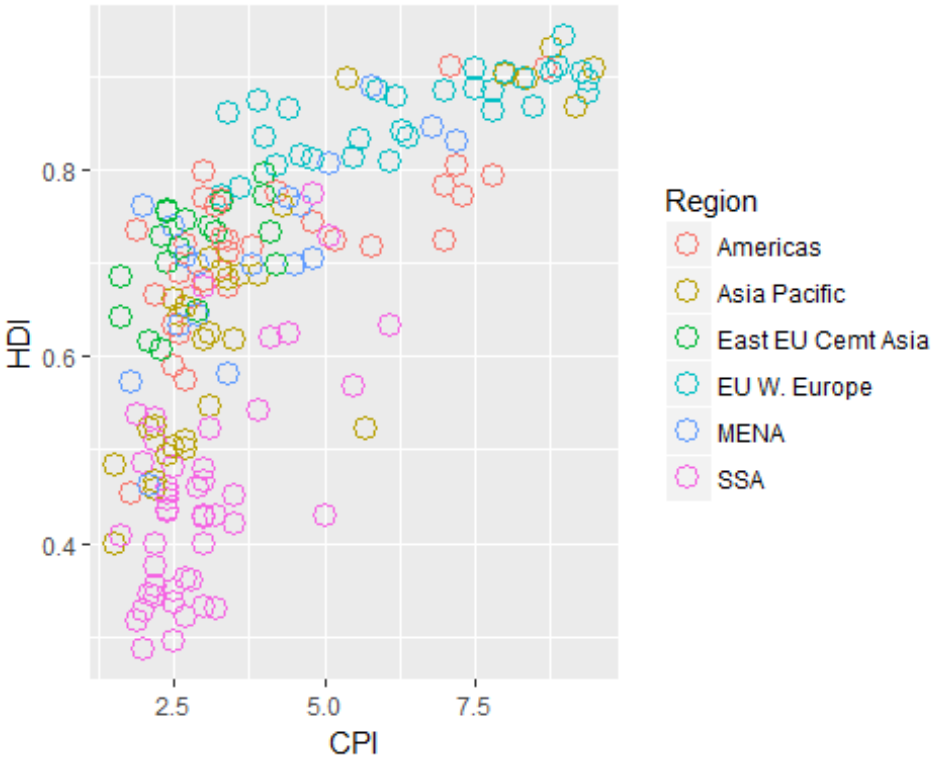

```
plot2 <- plot + geom_smooth(aes(group=1),method='lm', formula =
y~log(x),se=F,color='red')
print(plot2)
```



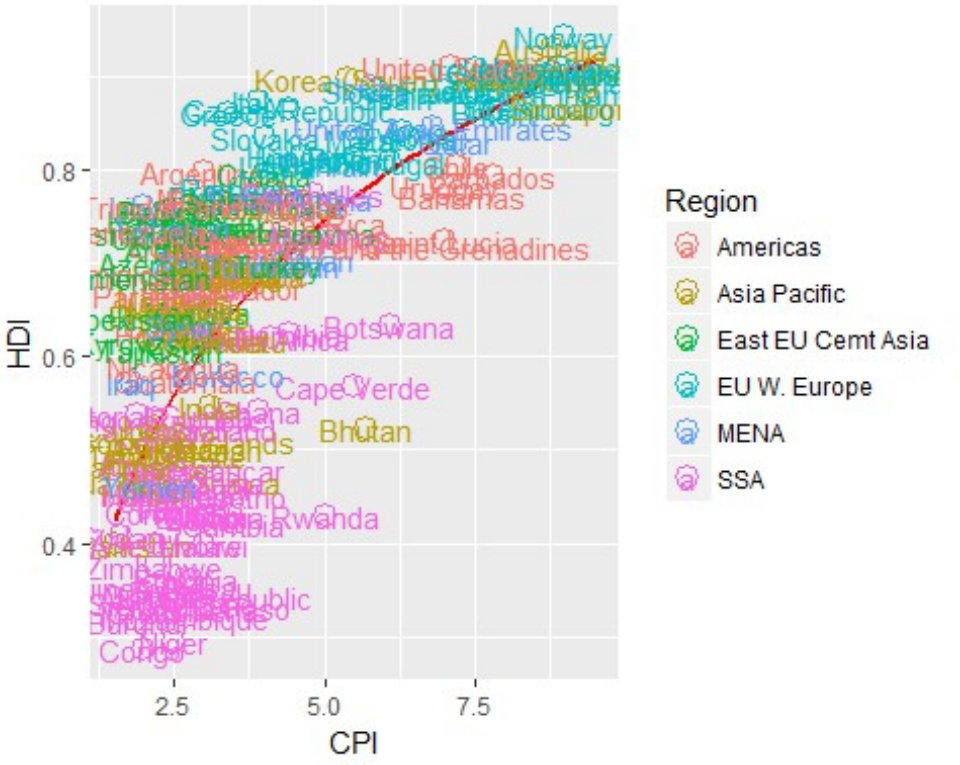
Adding Text:

For this section, we are going to add the labels of the countries that are represented in the graph. With the result, the labels have overlapped in the graph.

```
plot <- ggplot(df,aes(x=CPI,y=HDI,color=Region)) +
geom_point(size=4,shape=1)
print(plot)
```




```
plot2 <- plot + geom_smooth(aes(group=1),method='lm', formula =
y~log(x),se=F,color='red')
plot3 <- plot2 + geom_text(aes(label=Country))
print(plot3)
```



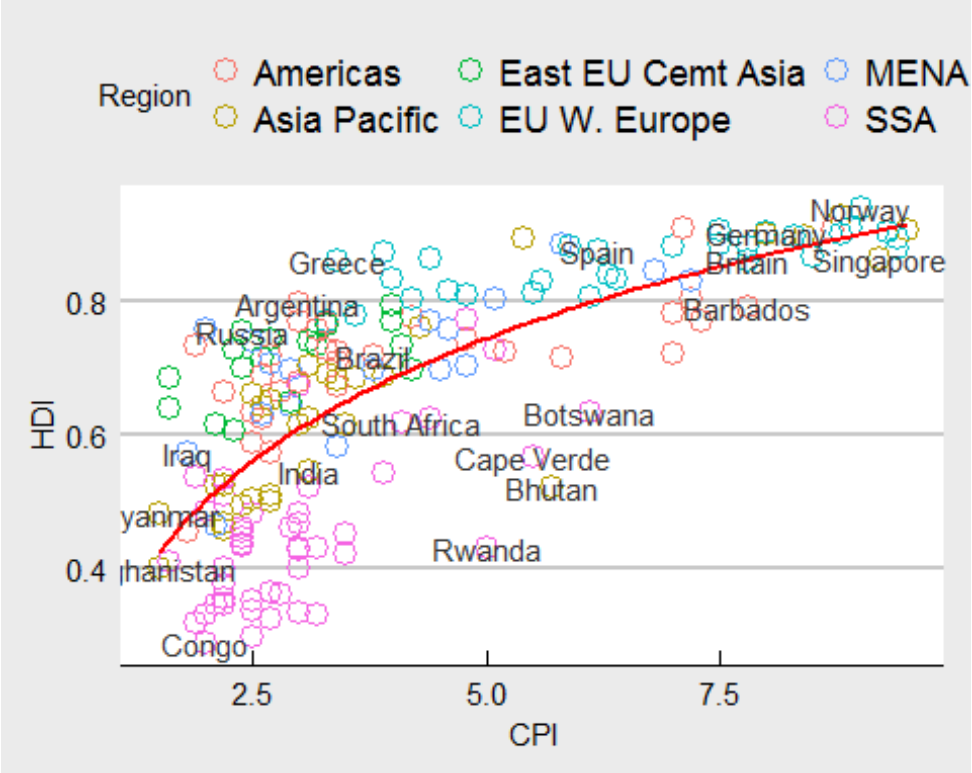
Subsetting the Labels:

Continuing from the previous section, we are only going to pick a select group of countries that will be displayed in the plot we created earlier.

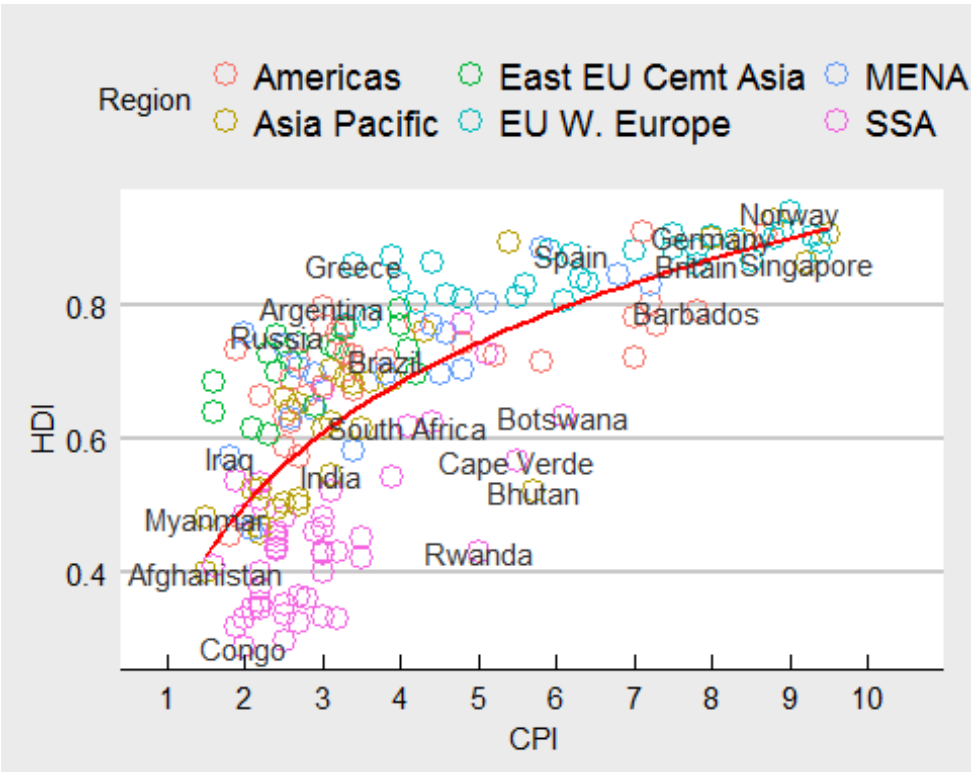
```
pointsToLabel <- c("Russia", "Venezuela", "Iraq", "Myanmar", "Sudan",
                   "Afghanistan", "Congo", "Greece", "Argentina",
"Brazil",
                   "India", "Italy", "China", "South Africa", "Spain",
                   "Botswana", "Cape Verde", "Bhutan", "Rwanda",
"France",
                   "United States", "Germany", "Britain", "Barbados",
"Norway",
                   "Japan","New Zealand", "Singapore")

plot3 <- plot2 + geom_text(aes(label = Country), color = "gray20",
                 data = subset(df, Country %in%
pointsToLabel),check_overlap = TRUE)

print(plot3 + theme_economist_white())
```




```
plot4 <- plot3 + scale_x_continuous(limits=c(.9,10.5),breaks = 1:10)
print(plot4 + theme_economist_white())
```

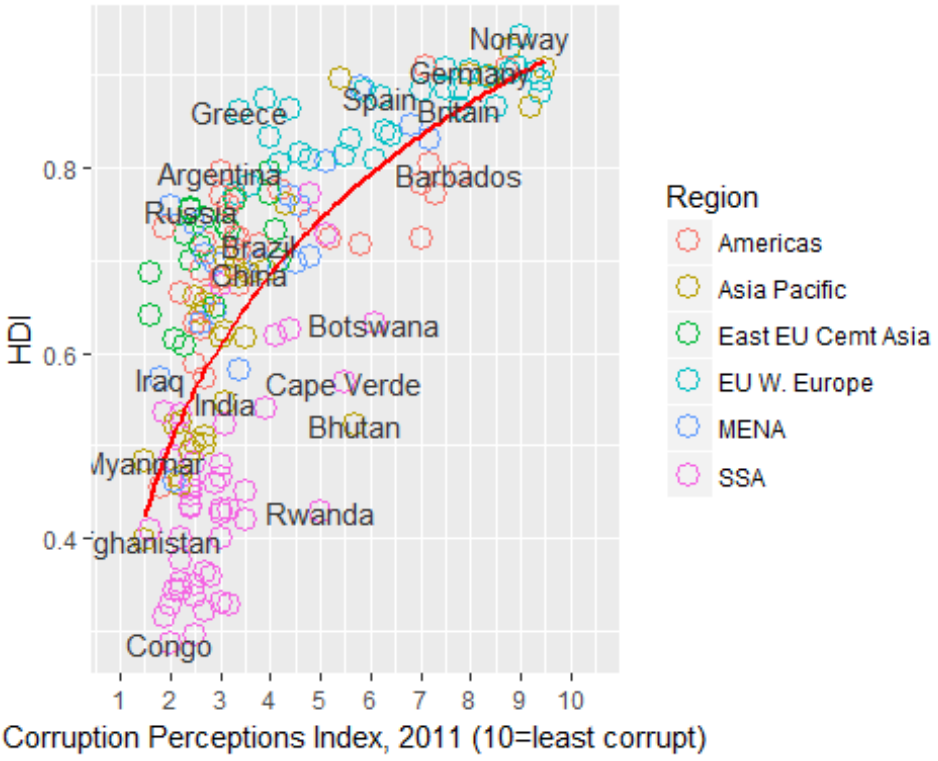


Adding a Title to the Graph and Fixing the Aesthetics on the X and Y Axis

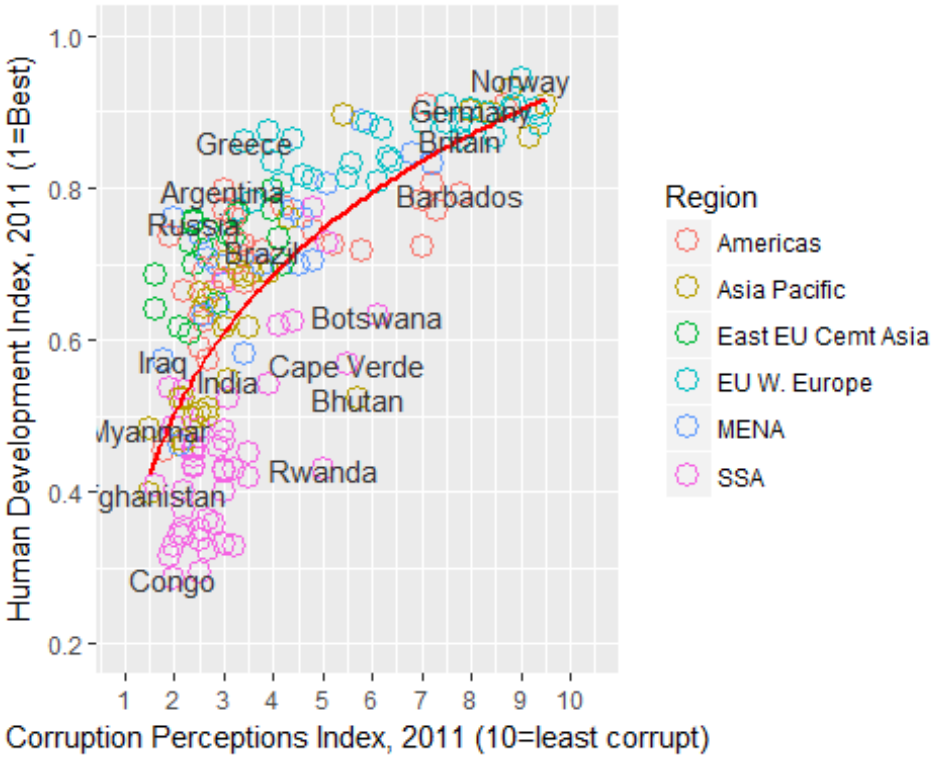
```
plot5 <- plot4 + scale_x_continuous(name = "Corruption Perceptions
Index, 2011 (10=least corrupt)",limits = c(.9, 10.5),breaks=1:10)

## Scale for 'x' is already present. Adding another scale for 'x',
which
## will replace the existing scale.

plot5
```

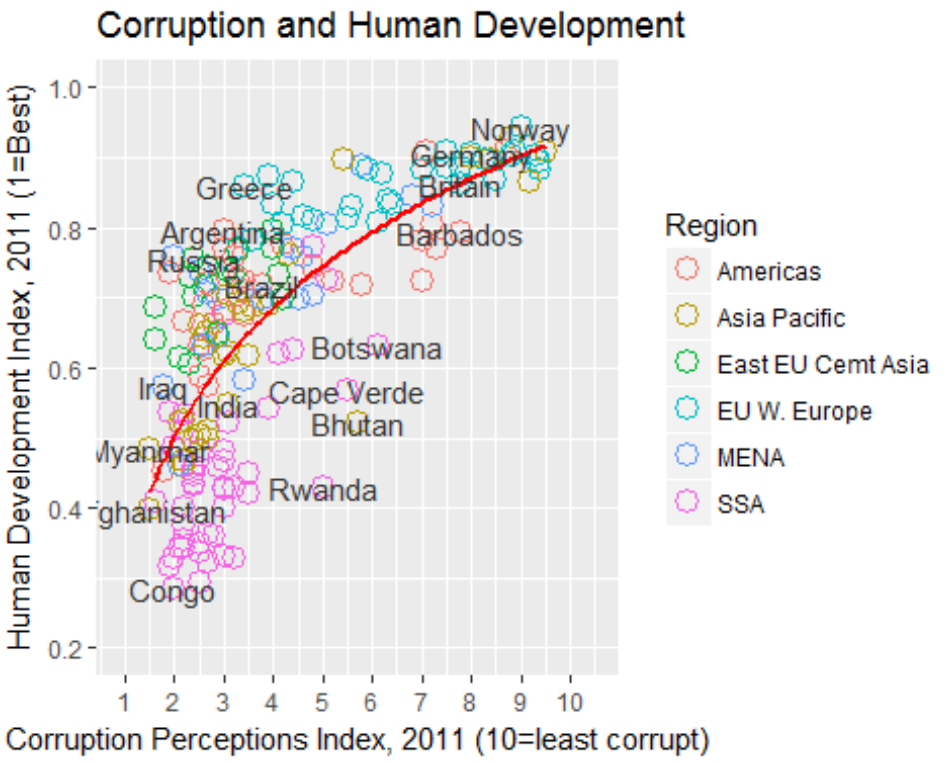



```
plot6 <- plot5 + scale_y_continuous(name = "Human Development Index,
2011 (1=Best)",limits = c(0.2, 1.0))
plot6
```



Adding a Title and Theme to Finish Off the Look of the Graph

```
plot6 + ggtitle("Corruption and Human Development")
```




```
plot6 + theme_economist_white()
```