R-Data-Visualization-Project.R

MinghaoLiu

2019-10-29

```
# Data Visualization Project

# load data&library
df <- read.csv('/Users/AncelotLiu/Desktop/R-Course-HTML-Notes/R-for-Data-Science-and-Machine-Learning/Training\ Exercises/Capstone\ and\ Data\ Viz\ Projects/Data\ Visualization\ Project/Economist_Assignment_Data.csv')
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

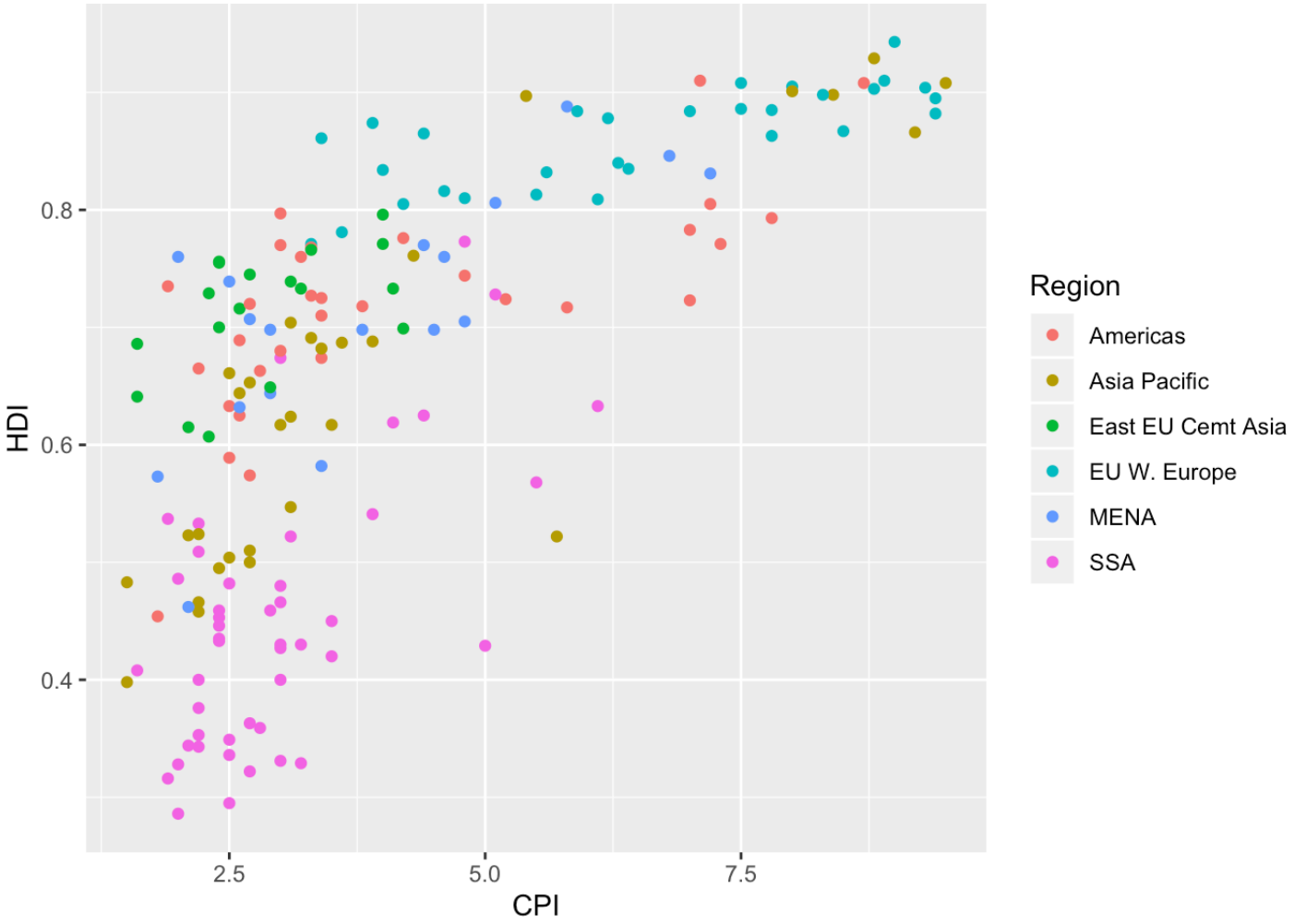
```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

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

```
df <- select(df, -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
```

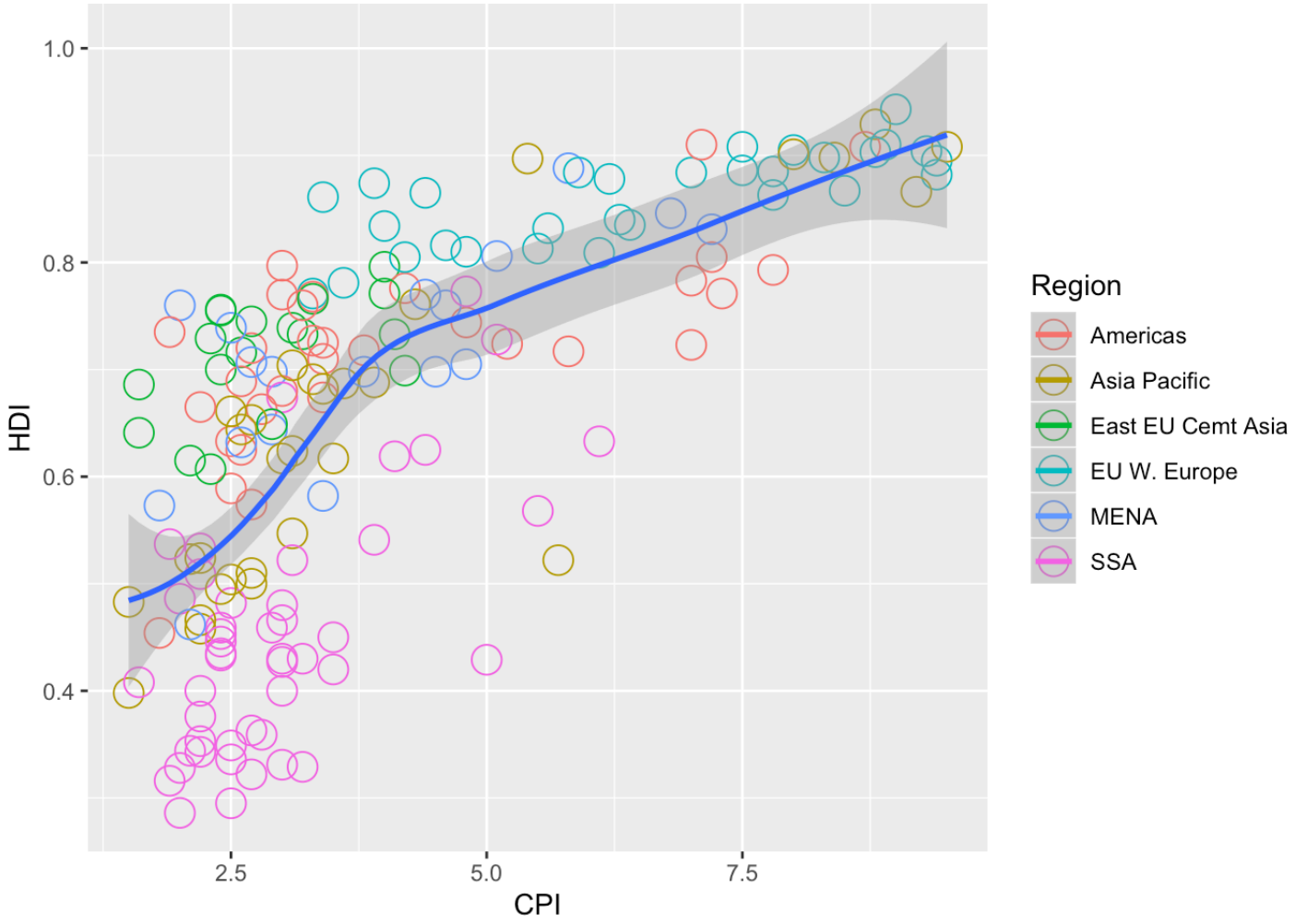
```
# To create a scatter plot of x=CPI and y=HDI

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



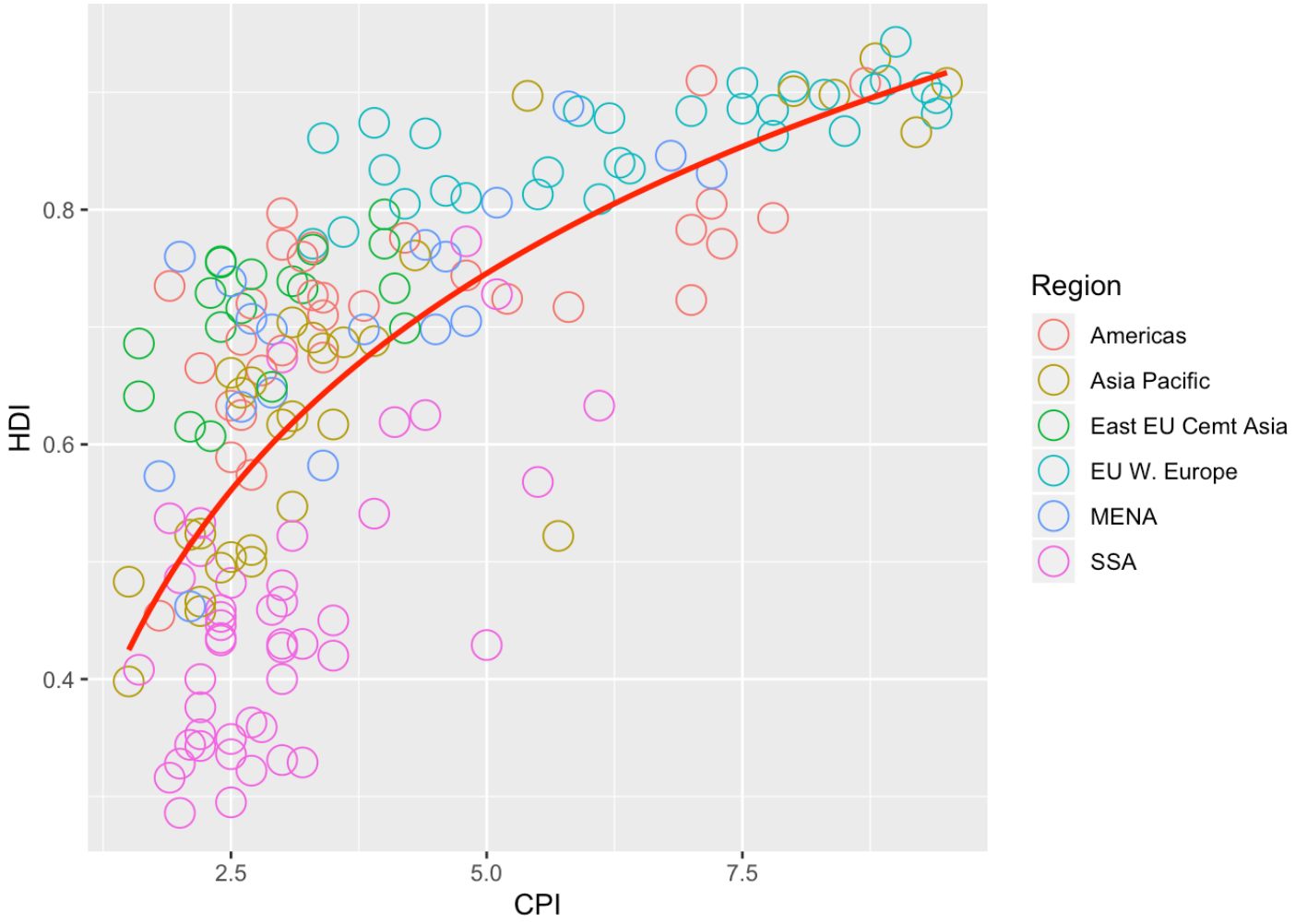

```
# Change the points to be larger empty circles.
# use "shape = " inside the geom_point to change the point shape

pl1 <- ggplot(df, aes(x=CPI, y=HDI, color = Region)) + geom_point(size=5, shape = 1)
pl1
```


```
# Add a trend line

pl2 <- pl1 + geom_smooth(aes(group = 1))
pl2
```

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

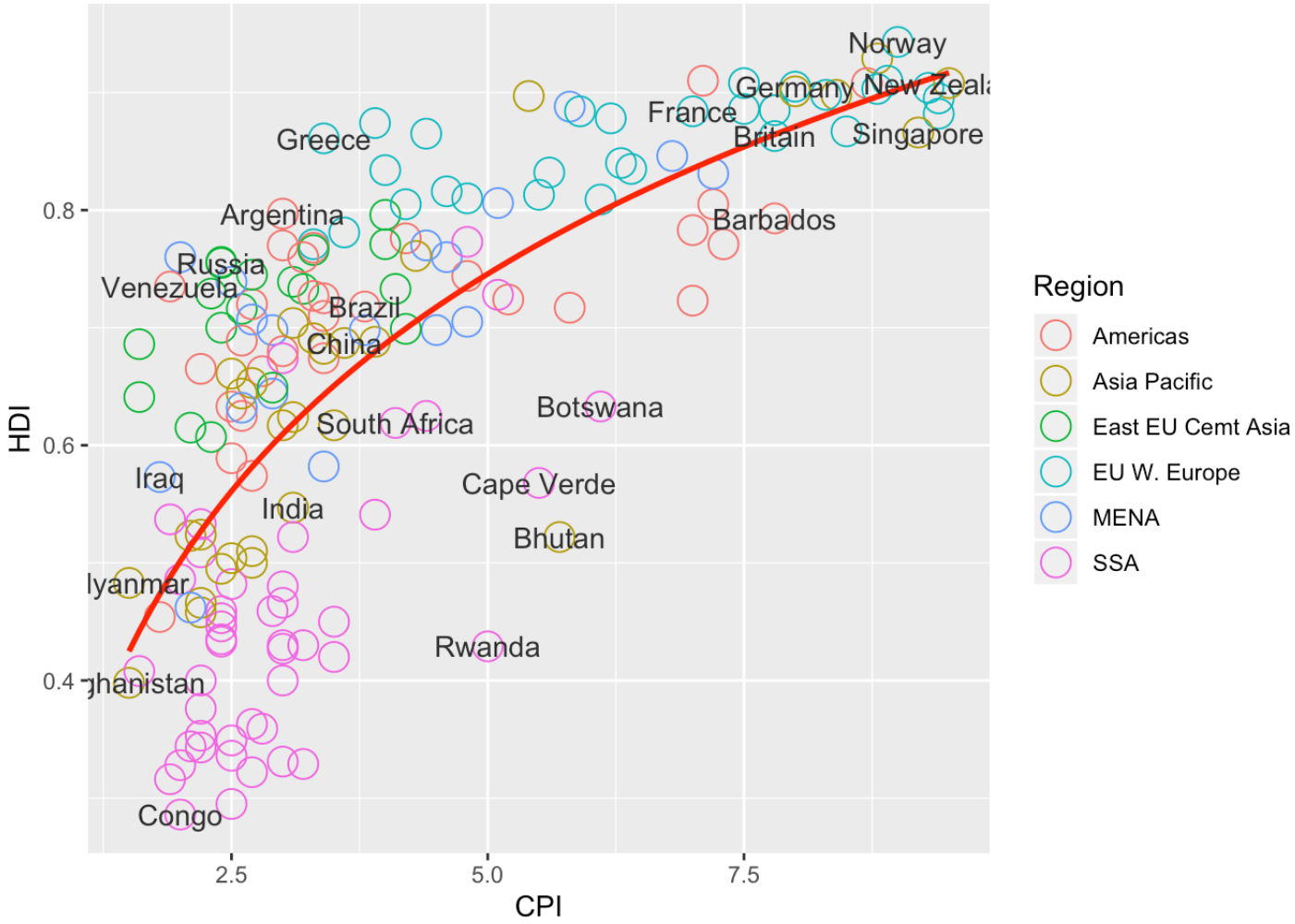



```
# To edit this trend line

pl3 <- pl1 + geom_smooth(aes(group = 1), method = 'lm', formula = y ~ log(x), se = FA
LSE, color = 'red')
pl3
```

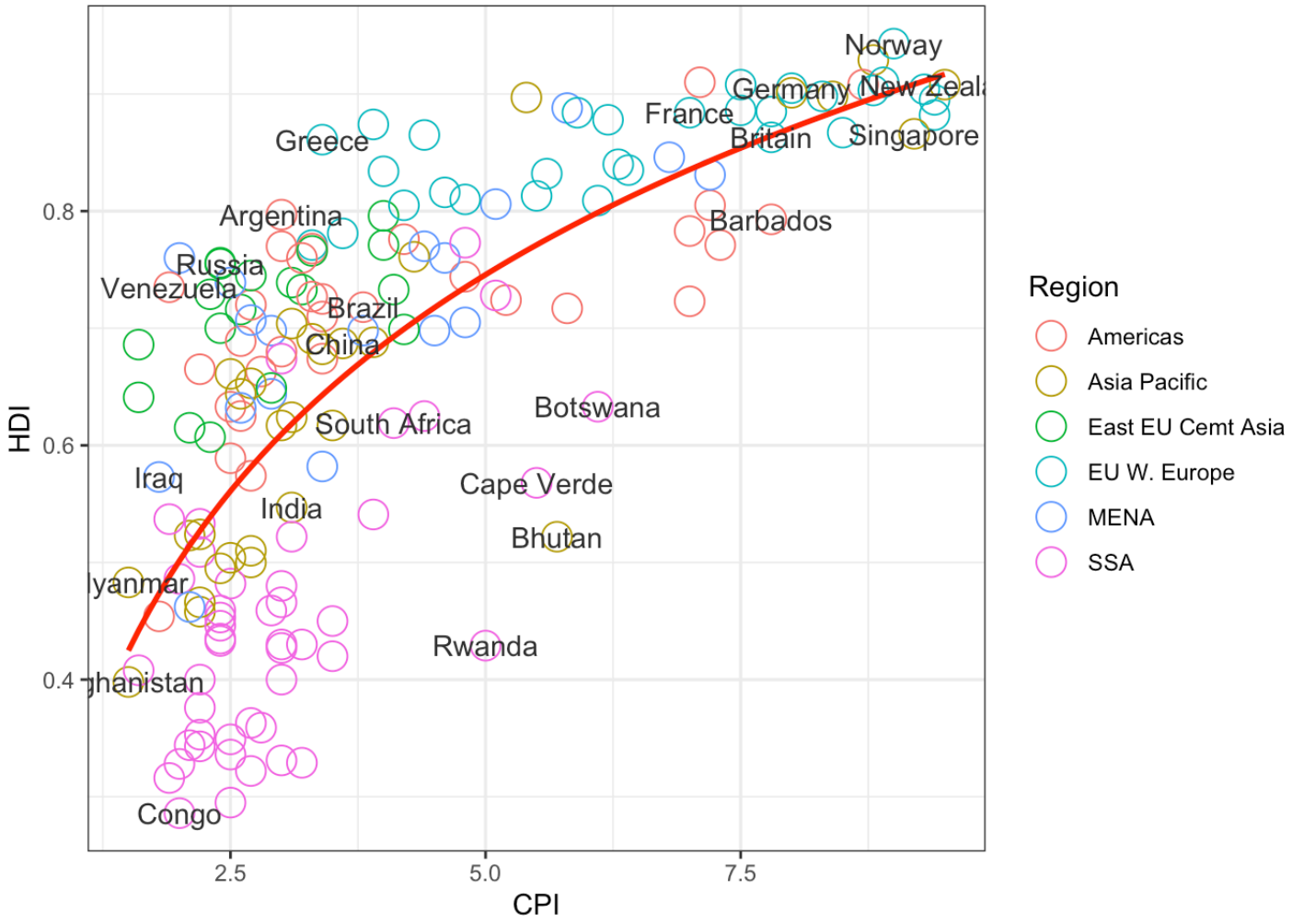



```
# To add text labels on the scatter points

pl4 <- pl3 + geom_text(aes(label = Country))
pl4
```

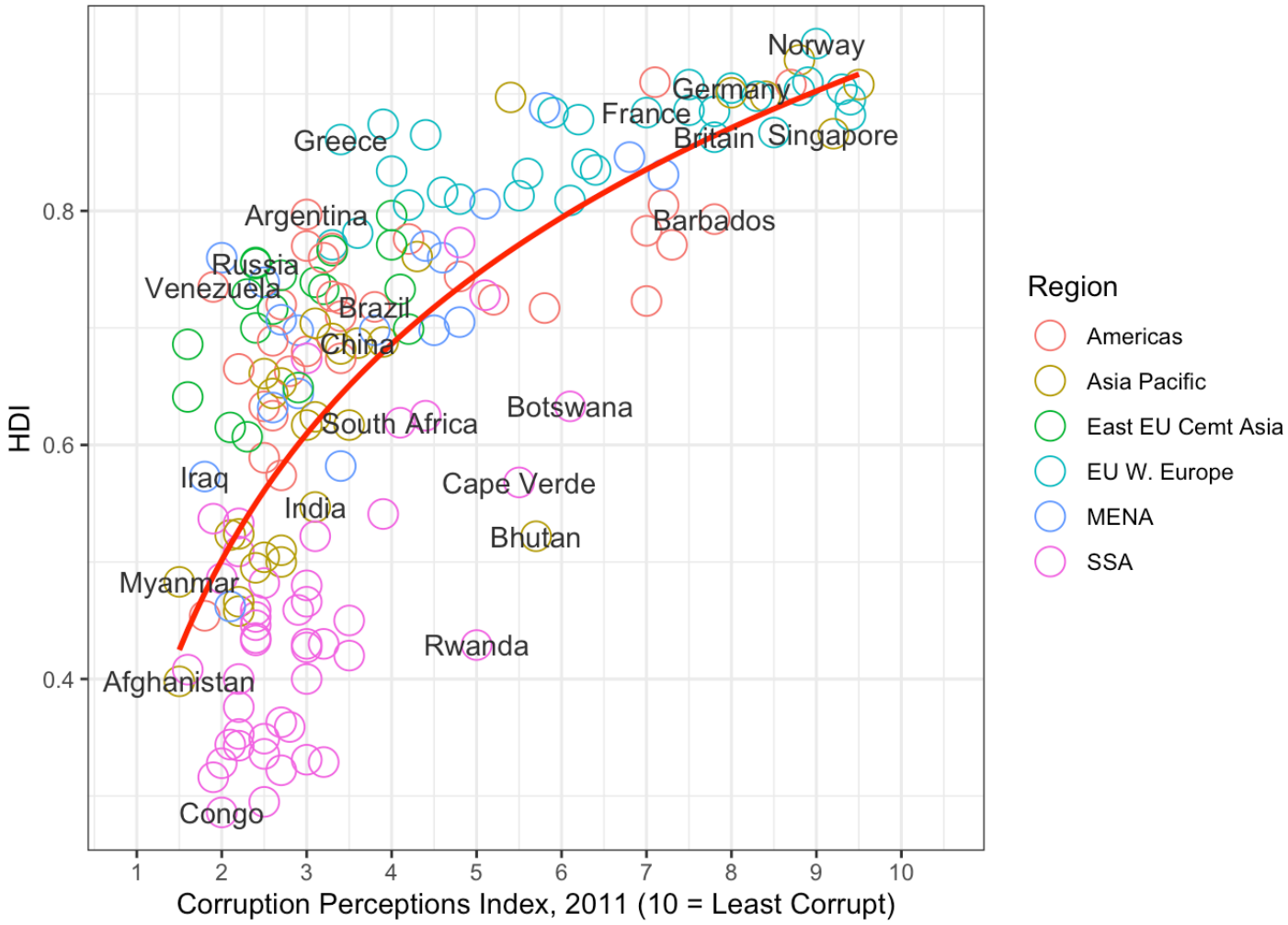

```
# To show only the selected countries label we want

selected_Country_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", "Japa
n",
                    "New Zealand", "Singapore")

pl5 <- pl3 + geom_text(aes(label = Country), color = 'gray20',
                       data = subset(df, Country %in% selected_Country_Label),
                       check_overlap = T)
pl5
```



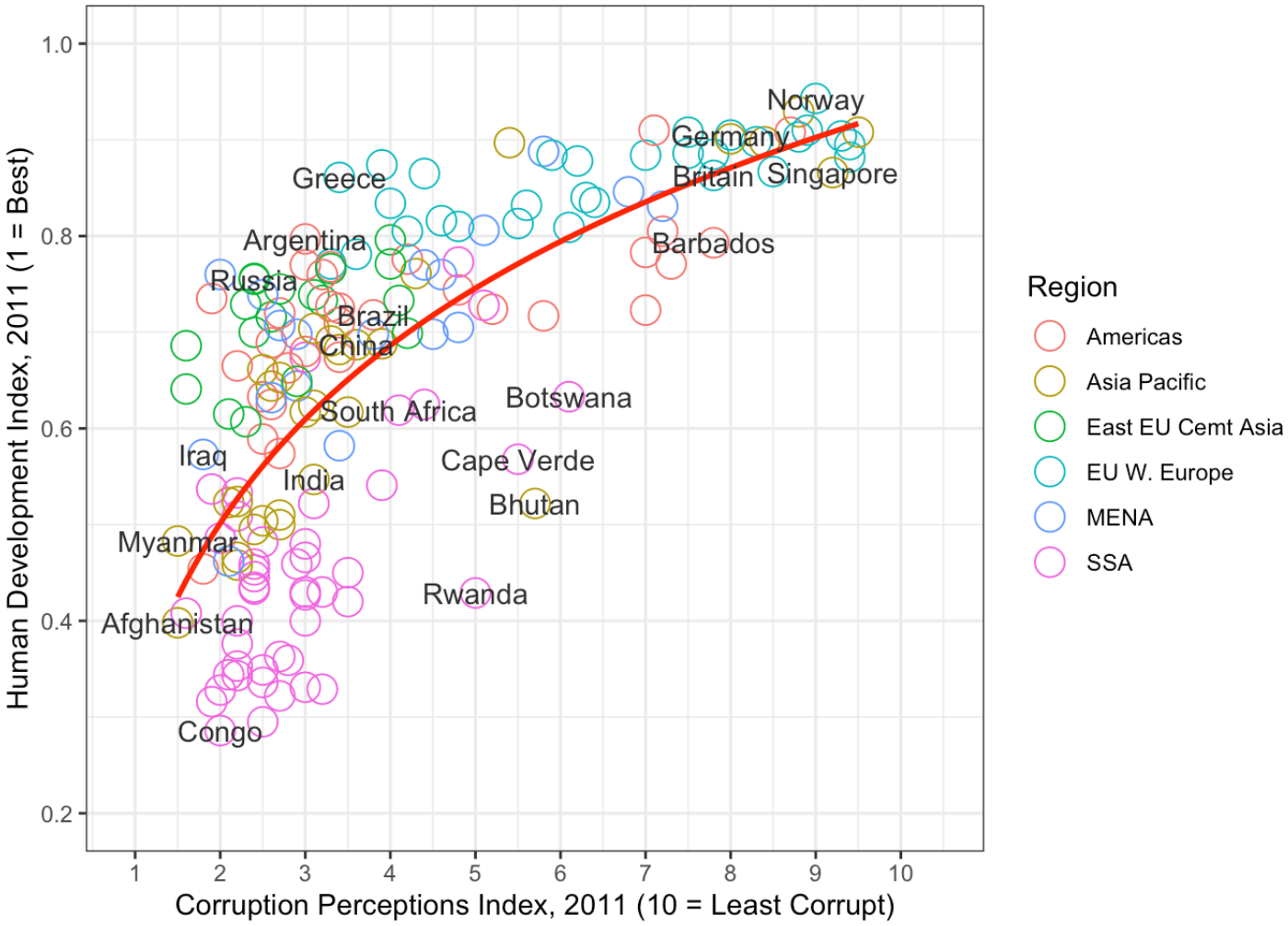

```
# Change theme

pl6 <- pl5 + theme_bw()
pl6
```



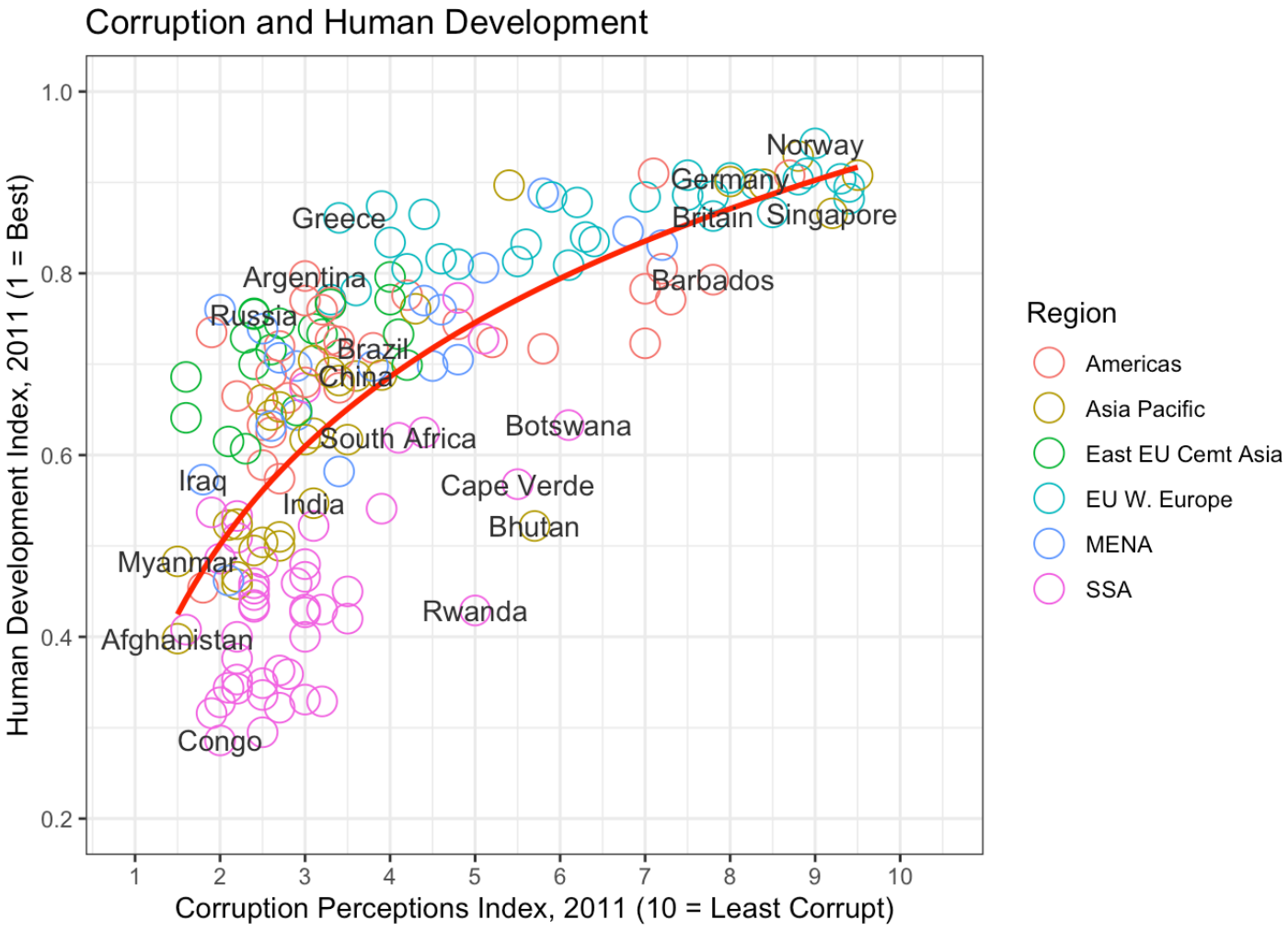

```
# To modify scale of Variable X, CPI, and scale of Variable Y, HDI

pl7 <- pl6 + scale_x_continuous(name = 'Corruption Perceptions Index, 2011 (10 = Leas
t Corrupt)',
                                limits = c(0.9,10.5), breaks = 1:10)
pl7
```




```
pl8 <- pl7 + scale_y_continuous(name = 'Human Development Index, 2011 (1 = Best)',
                                limits = c(0.2,1))
pl8
```




```
# Add a title

pl9 <- pl8 + ggtitle('Corruption and Human Development')
pl9
```


Corruption and Human Development
Human Development Index, 2011 (1 = Best)
Corruption Perceptions Index, 2011 (10 = Least Corrupt)
1.0
0.8
0.6
0.4
0.2
1
2
3
4
5
6
7
8
9
10
Norway
Germany
Britain
Singapore
Greece
Argentina
Barbados
Russia
Brazil
China
Botswana
South Africa
Iraq
Cape Verde
India
Bhutan
Myanmar
Rwanda
Afghanistan
Congo
Region
Americas
Asia Pacific
East EU Cemt Asia
EU W. Europe
MENA
SSA