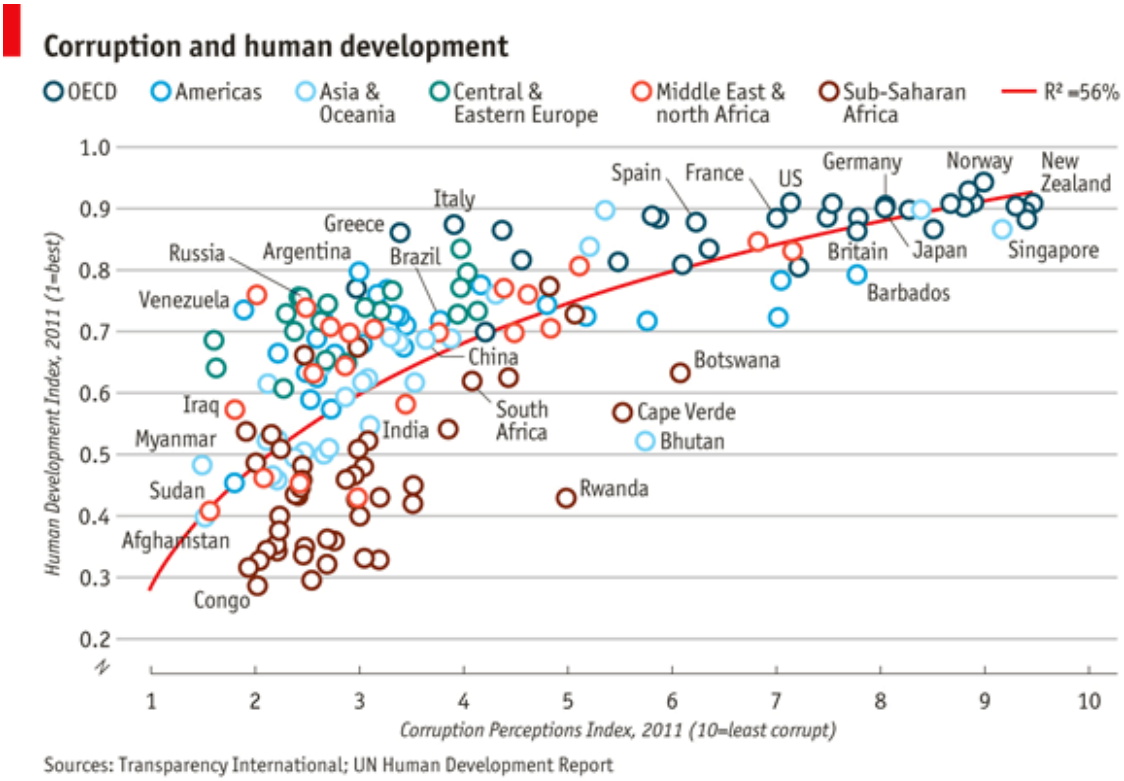
# Data Visualization Project

## Assignment for ggplot2

For this optional assignment we will be recreating this plot from The Economist:




**→ Import the ggplot2 data.table libraries and use fread to load the csv file 'Economist_Assignment_Data.csv' into a dataframe called df**

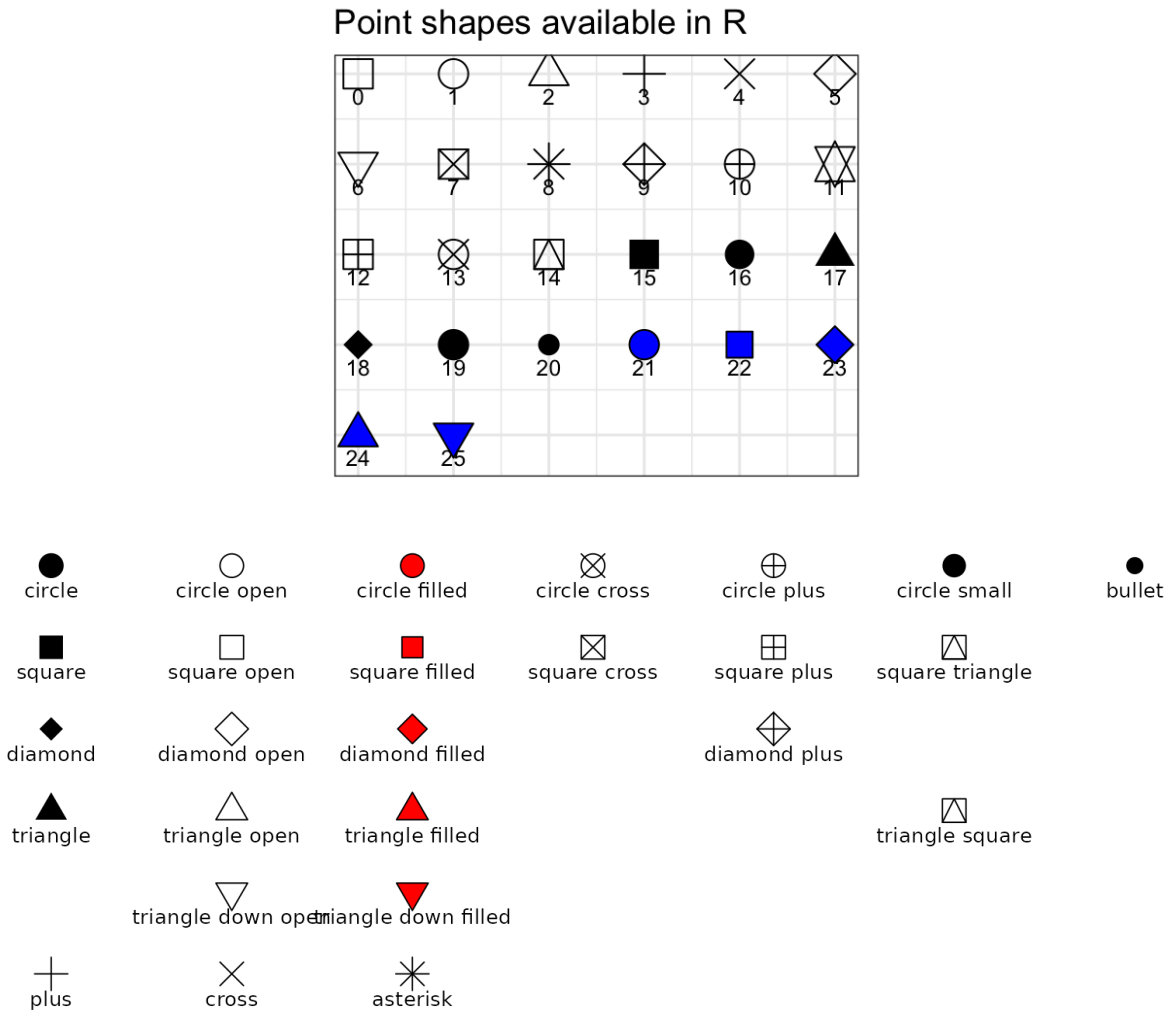
```
library(ggplot2)
library(data.table)

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

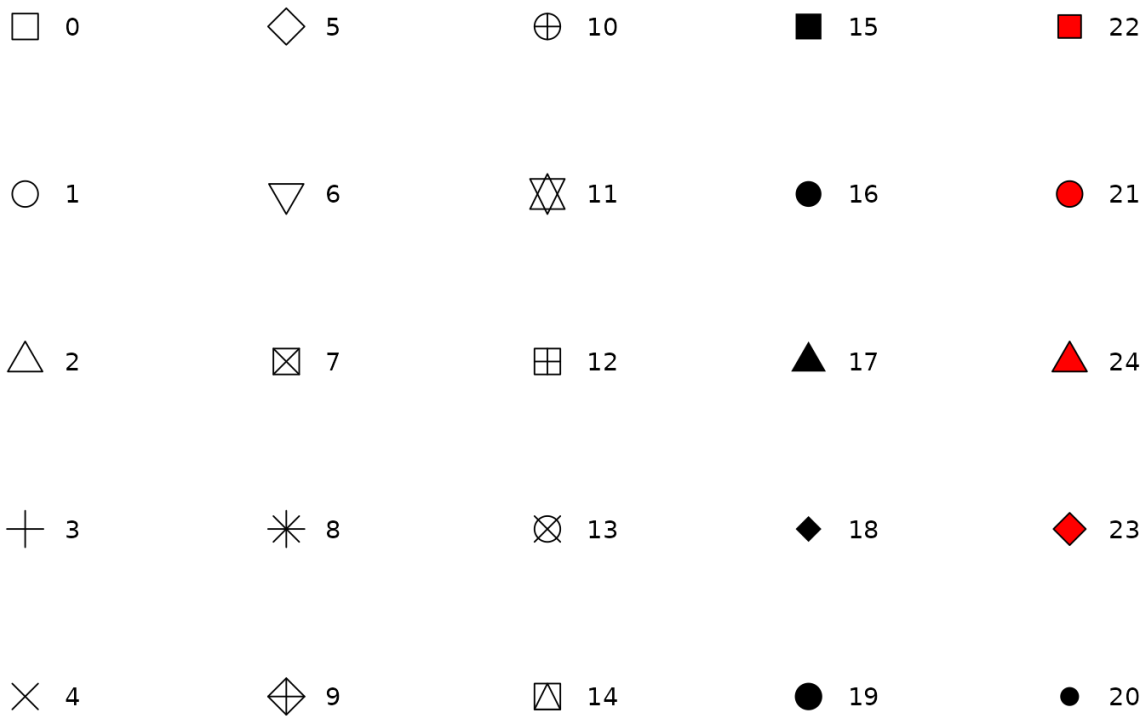
**→ Use ggplot() + geom_point() to create a scatter plot object called *pl*.**

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

→ **Change the points to be larger empty circles.**

▼ Shape






```
pl <- ggplot(data=df,aes(x = CPI,y = HDI))
pl + geom_point(shape = "circle open", size = 2 ,aes(color = Region))
```

→ **Add geom_smooth(aes(group=1)) to add a trend line**

```
pl <- ggplot(data=df,aes(x = CPI,y = HDI))
pl + geom_point(shape = "circle open", size = 2 ,aes(color = Region)) +
  geom_smooth(aes(group=1))
```

→ **Add the following arguments to geom_smooth (outside of aes):**

- method = 'lm'
- formula = y ~ log(x)
- se = FALSE
- color = 'red'

**Assign all of this to pl2**

```
pl2 <- pl + geom_point(shape = "circle open", size = 2 ,aes(color = Region)) +
  geom_smooth(aes(group=1), method = 'lm',
```

```
              formula = y ~ log(x),
              se = FALSE,
              color = 'red')
```

### → Add *geom_text(aes(label=Country))* to pl2

```
pl2 + geom_text(aes(label=Country))
```

```
pointsToLabel <- 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")

pl3 <- pl2 + geom_text(aes(label = Country), color = "gray20",
               data = subset(df, Country %in% pointsToLabel),check_overlap = TRUE)
```

### → Add theme_bw() to your plot and save this to pl4

```
pl4 <- pl3 + theme_bw()
```

### → Add scale_x_continuous()

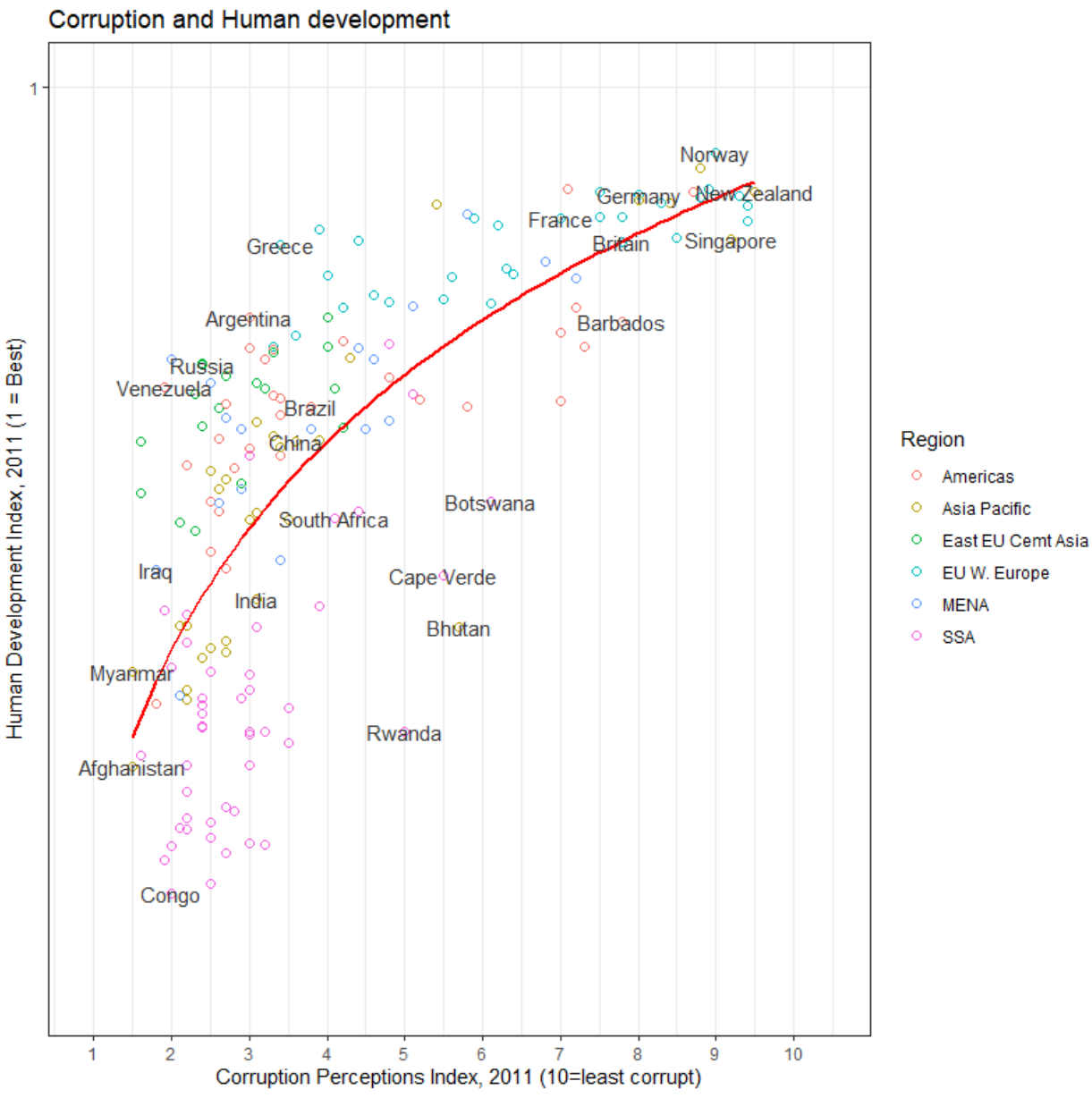
- name = *Same x axis as the Economist Plot*
- limits = *Pass a vector of appropriate x limits*
- breaks = 1:10

```
pl4 + scale_x_continuous(name = "Corruption Perceptions Index, 2011 (10=least corrupt)",
                         breaks = 1:10, limits = c(.9, 10.5))
```

### → Now use scale_y_continuous to do similar operations to the y axis!

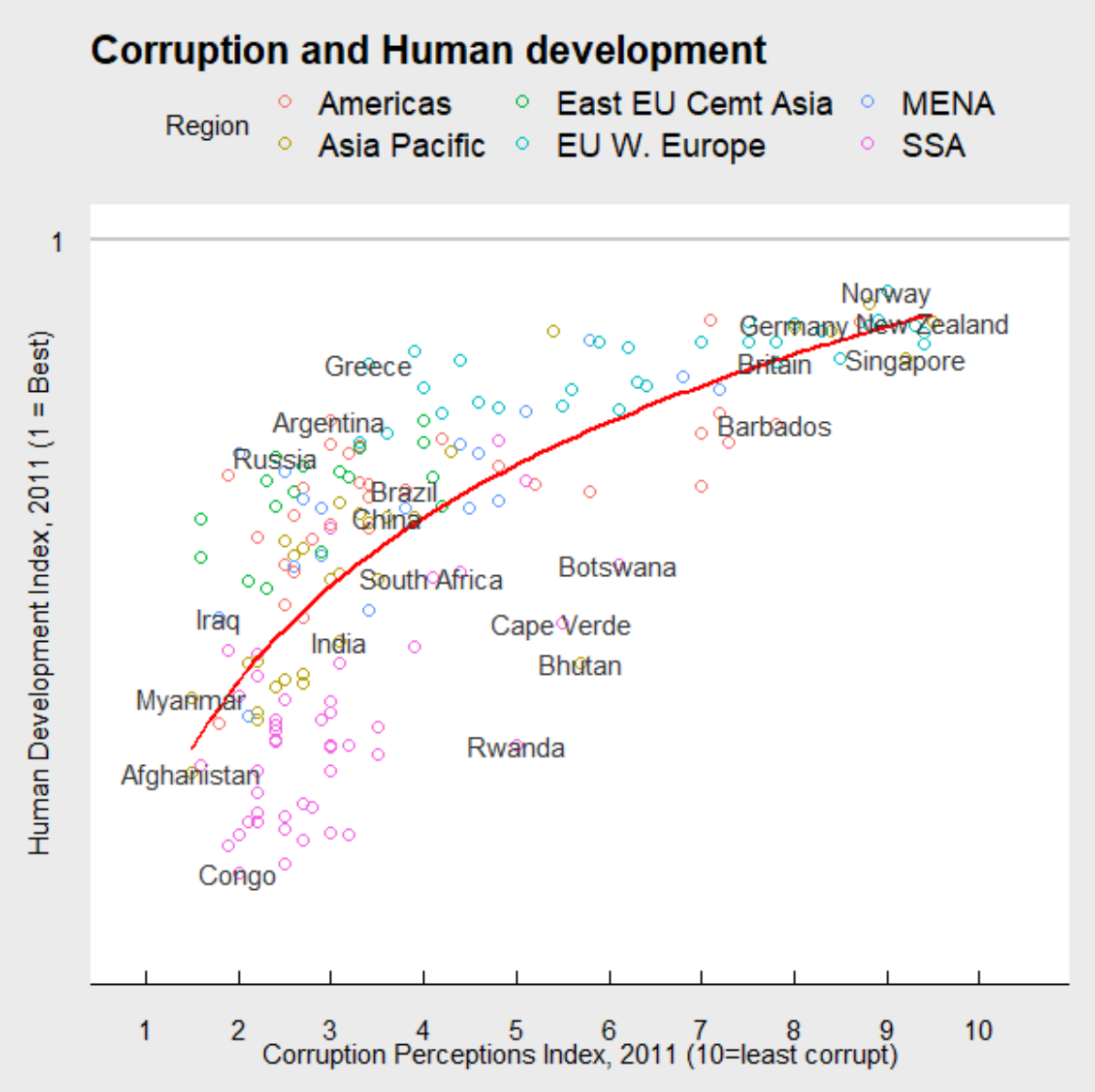
```
pl5 <- pl4 + scale_x_continuous(name = "Corruption Perceptions Index, 2011 (10=least corrupt)",
                         breaks = 1:10, limits = c(.9, 10.5)) +
  scale_y_continuous(name = "Human Development Index, 2011 (1 = Best)",
                     breaks = 1:10, limits = c(0.2, 1.0))
```

```
pl6 <- pl5 + ggtitle("Corruption and Human development")
```




**→ use ggthemes library**

```
library(ggthemes)
pl6 + theme_economist_white()
```


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