Stat535_HW4

Jie Wang

October 10, 2017

Read the data:

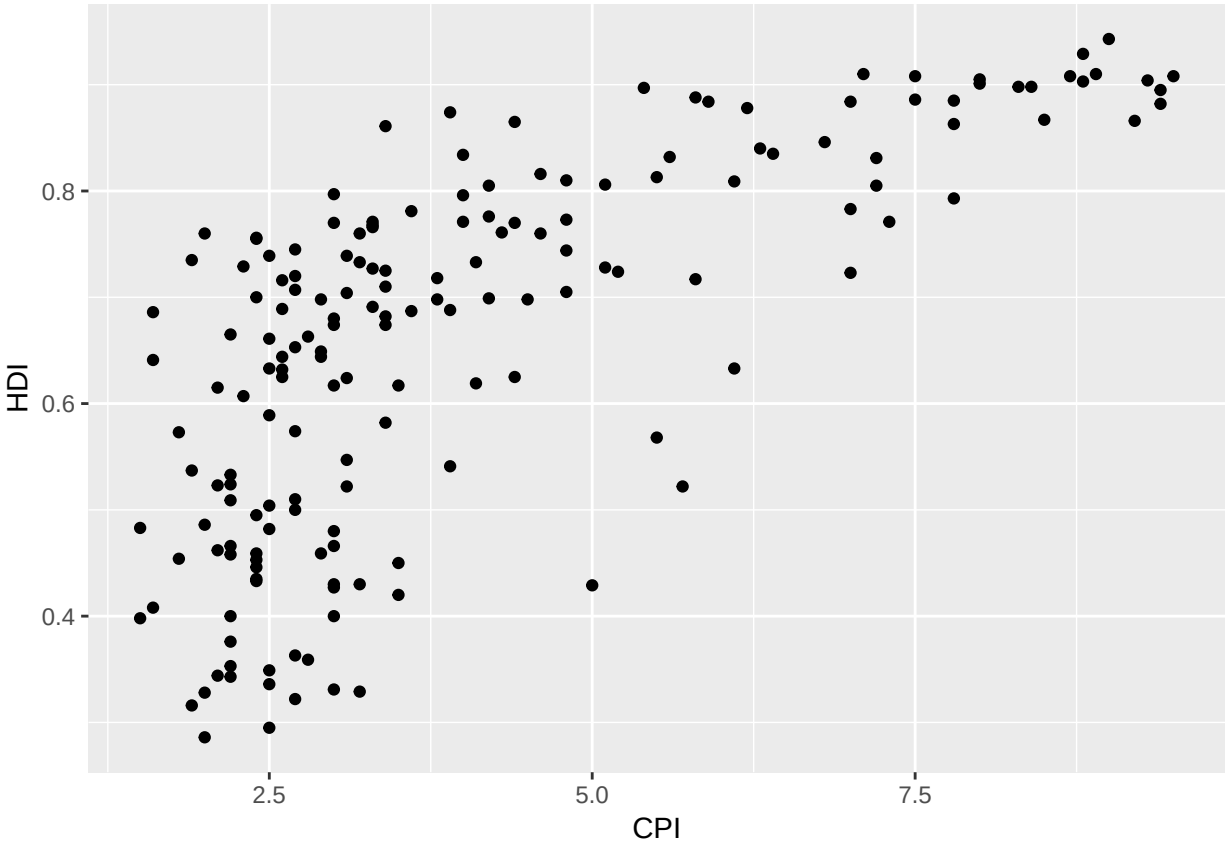
```
rm(list=ls())
dat <- read.csv("Rgraphics/dataSets/EconomistData.csv")
head(dat)
```

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

Exercise I:

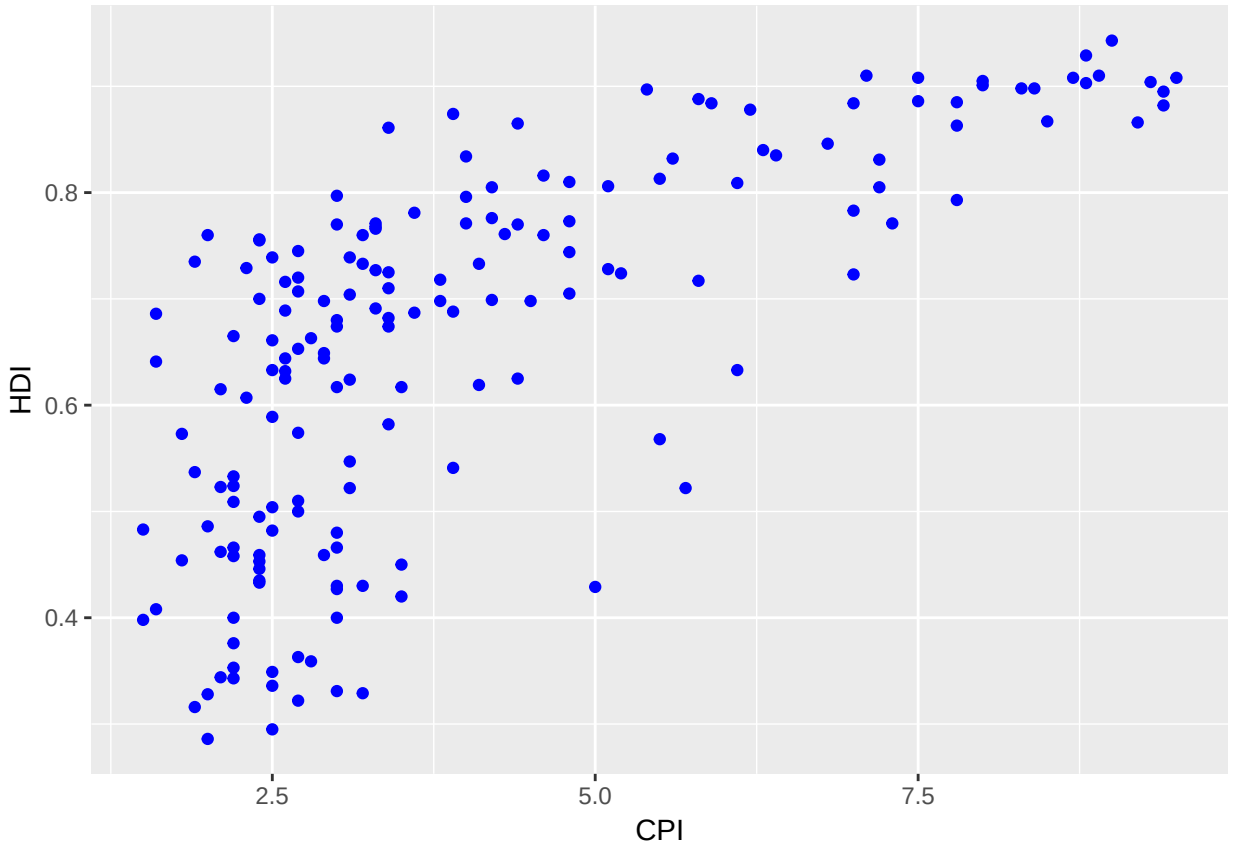
1. Create a scatter plot with CPI on the x axis and HDI on the y axis:

```
library(ggplot2)
ggplot(dat, aes(x = CPI, y = HDI)) + geom_point()
```



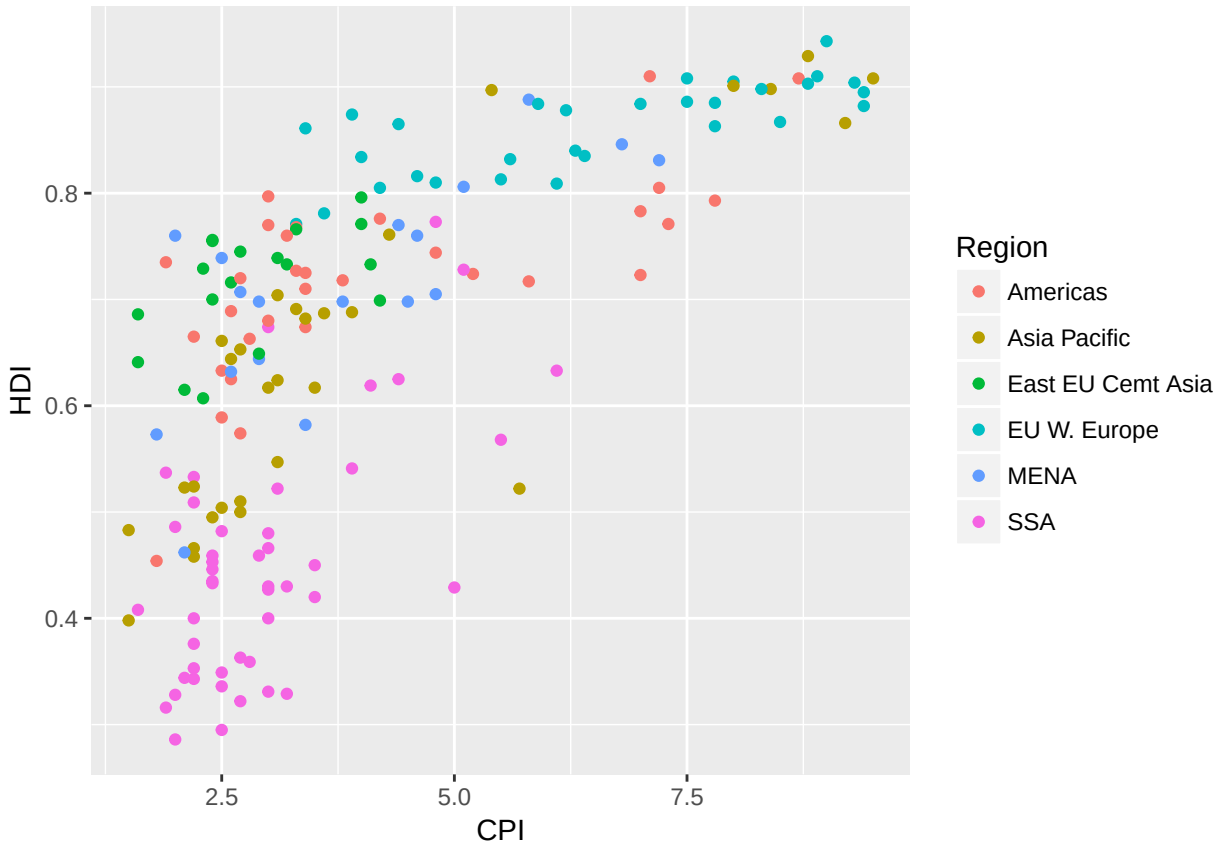
2. Color the points blue:

```
ggplot(dat, aes(x = CPI, y = HDI)) + geom_point(color="blue")
```



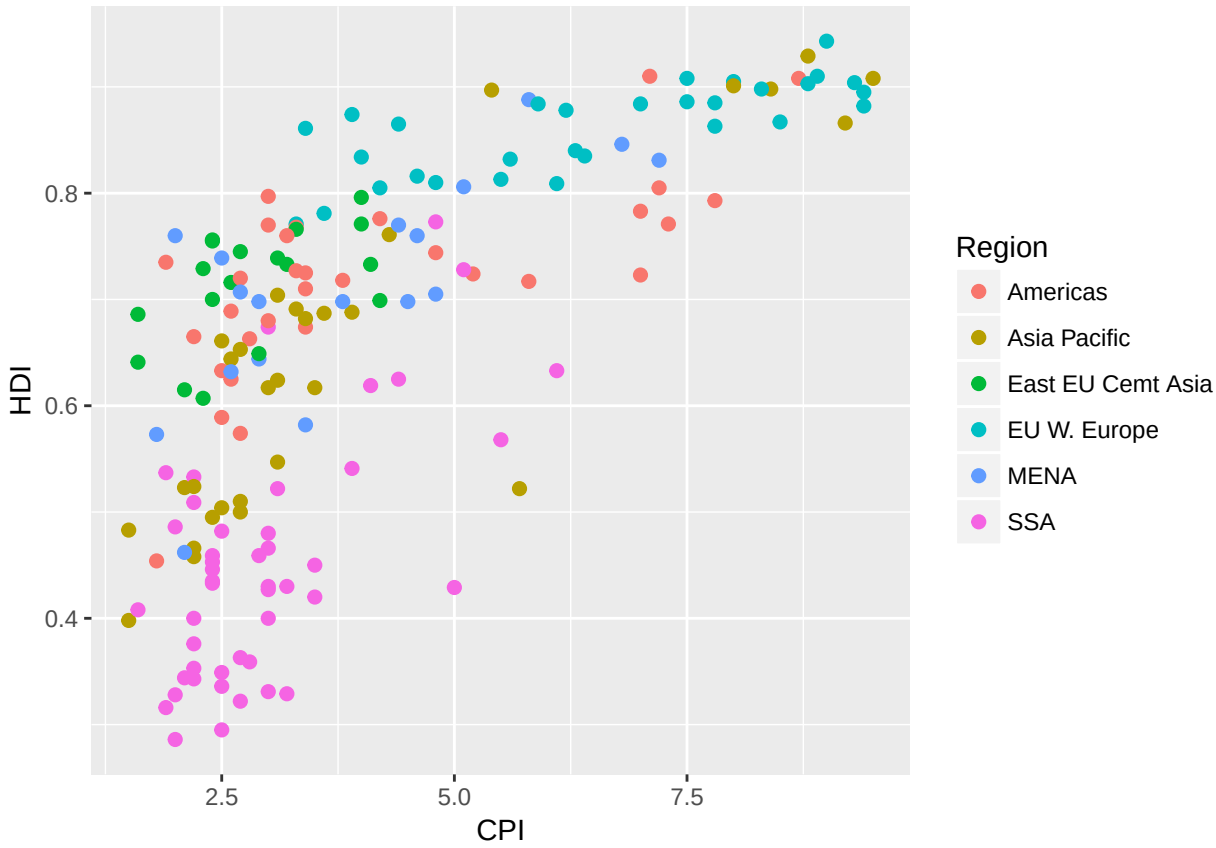
3. Map the color of the the points to Region:

```
ggplot(dat, aes(x = CPI, y = HDI)) + geom_point(aes(color=Region))
```



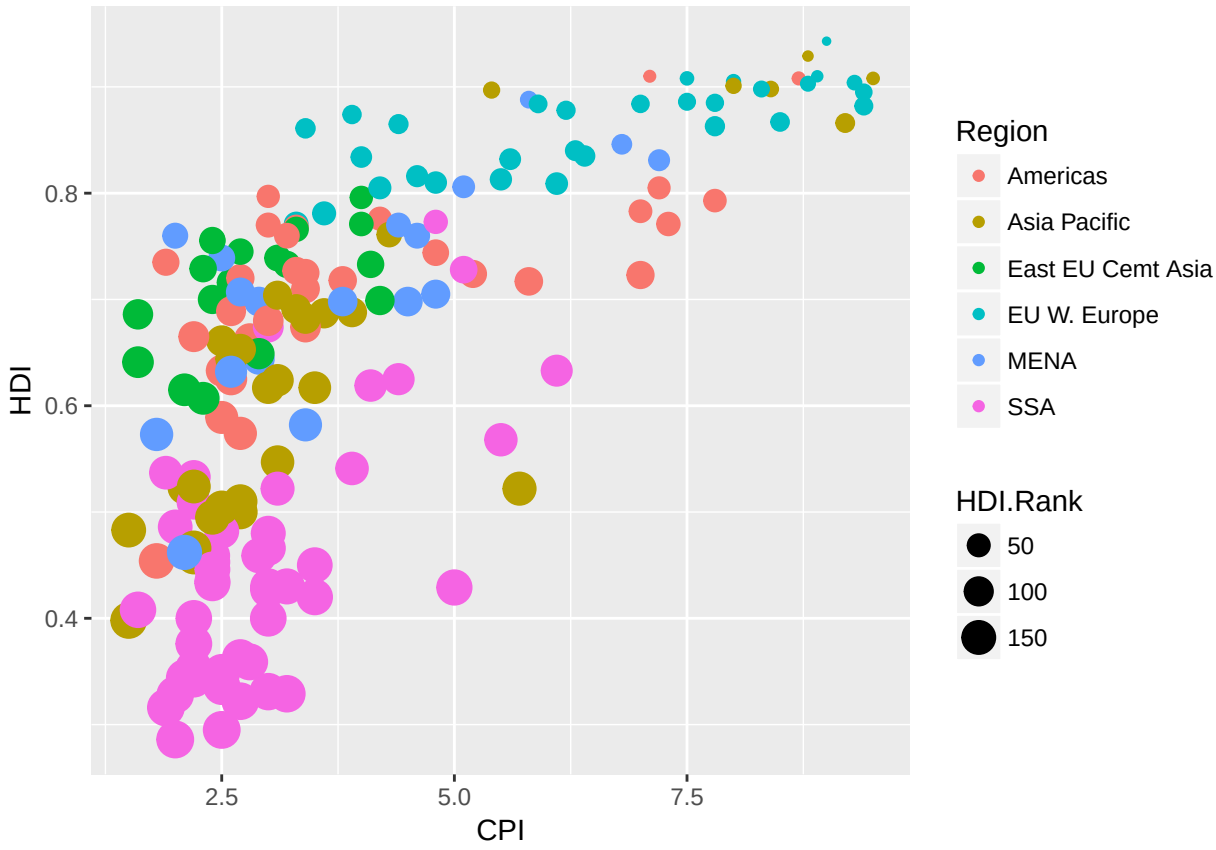
4. Make the points bigger by setting size to 2:

```
ggplot(dat, aes(x = CPI, y = HDI)) + geom_point(aes(color=Region), size=2)
```



5. Map the size of the points to HDI.Rank:

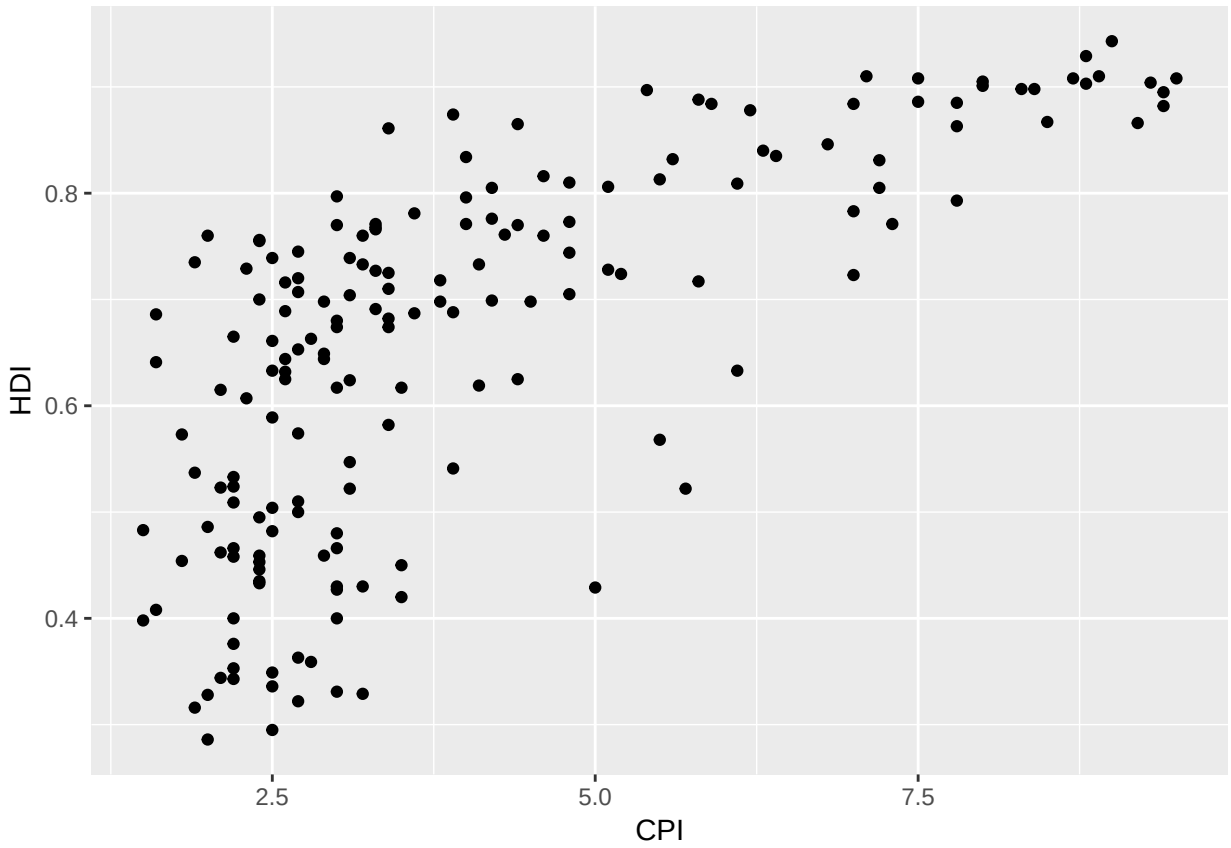
```
ggplot(dat, aes(x = CPI, y = HDI)) + geom_point(aes(color=Region, size=HDI.Rank))
```



Exercise II:

1. Re-create a scatter plot with CPI on the x axis and HDI on the y axis (as you did in the previous exercise):

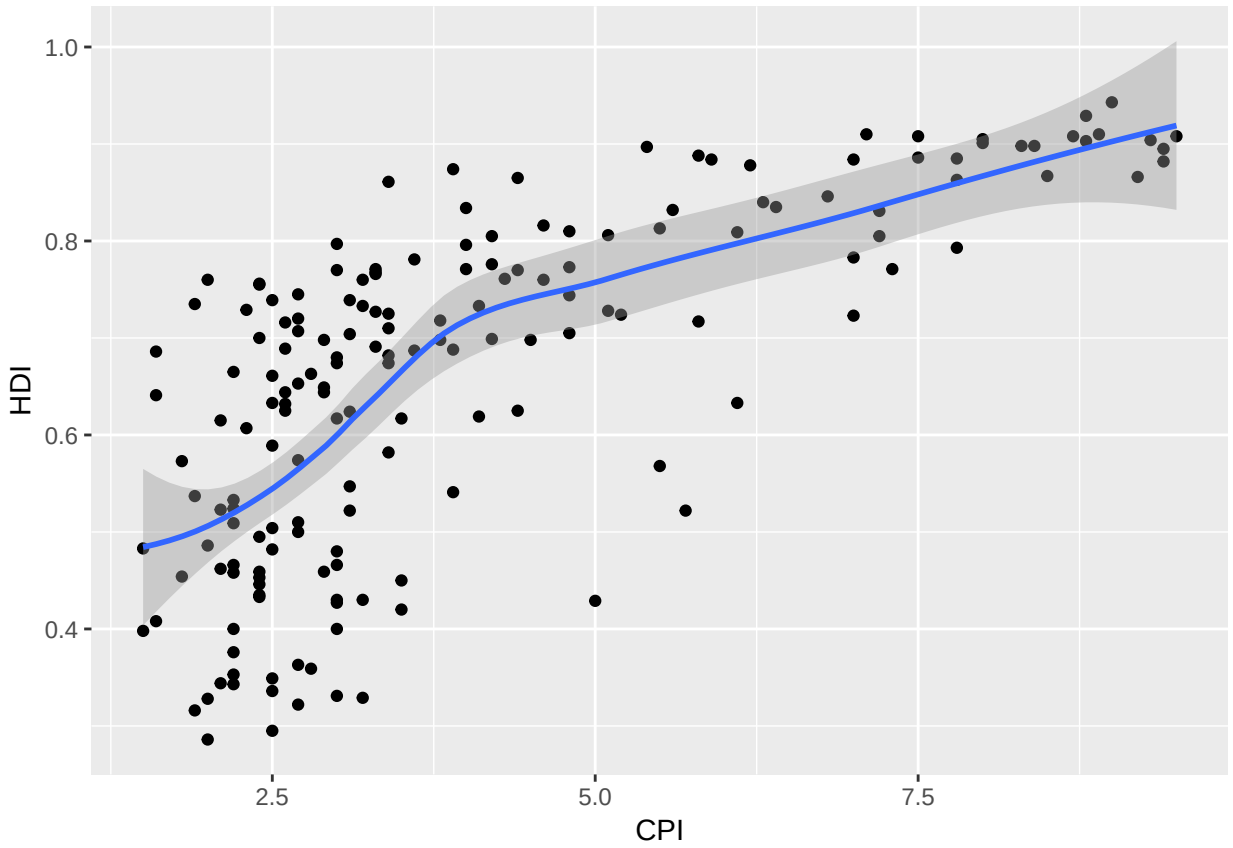
```
ggplot(dat, aes(x = CPI, y = HDI)) + geom_point()
```



2. Overlay a smoothing line on top of the scatter plot using geom_smooth:

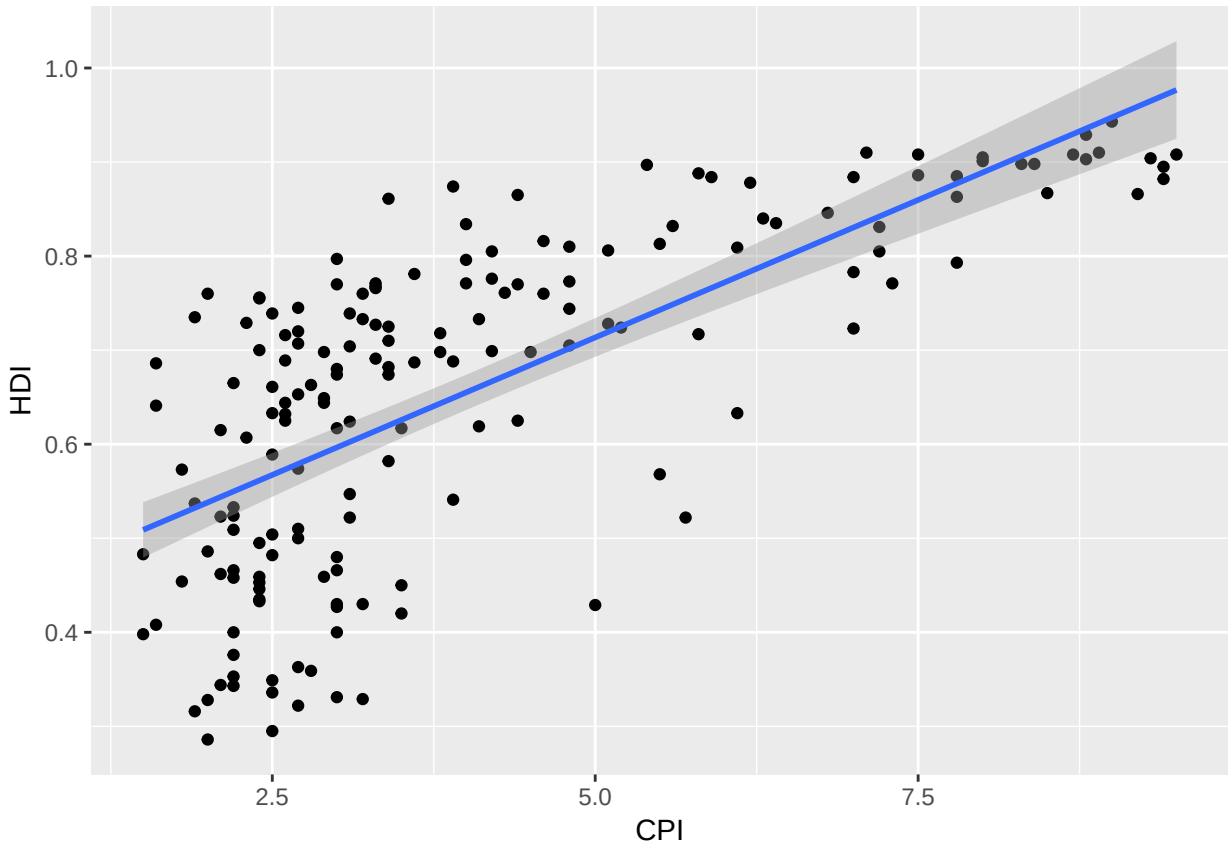
```
ggplot(dat, aes(x = CPI, y = HDI)) + geom_point()+geom_smooth()
```

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



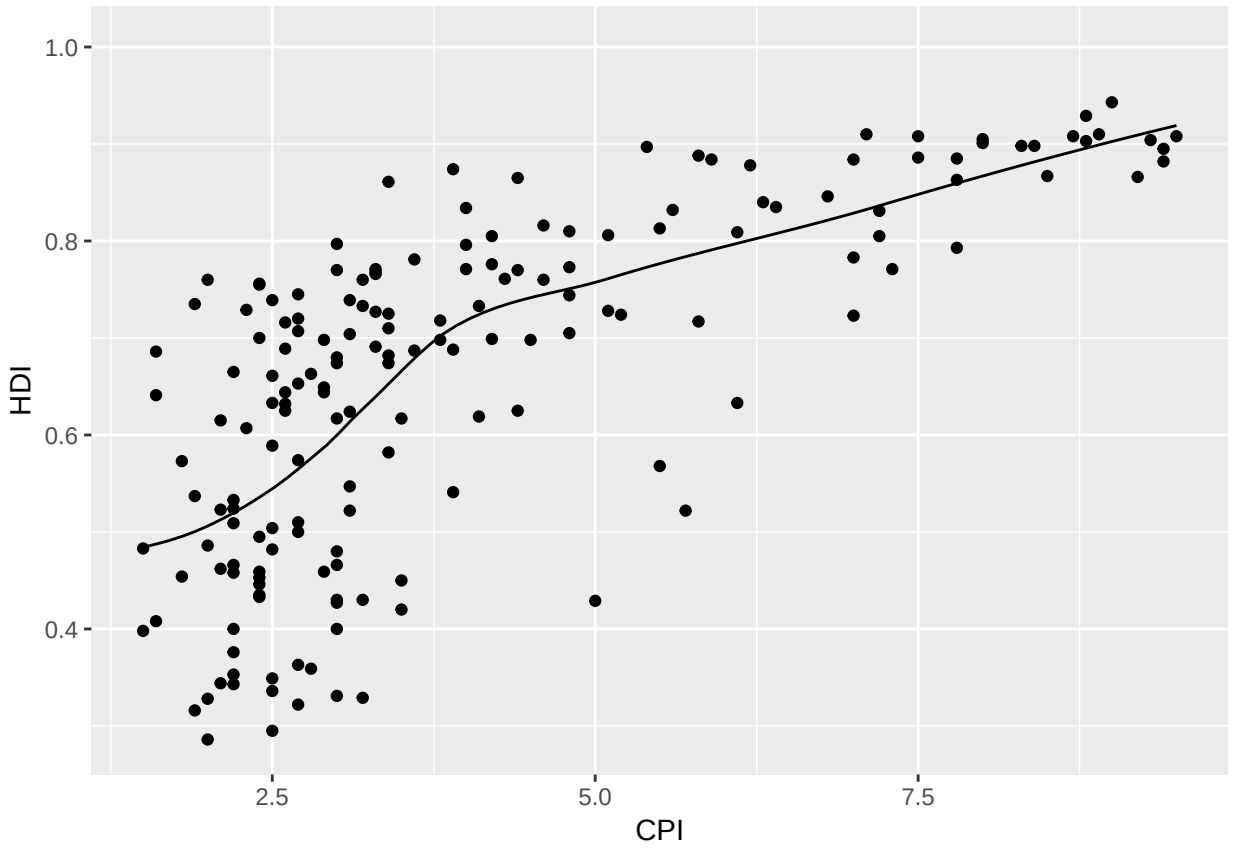
3. Overlay a smoothing line on top of the scatter plot using geom_smooth, but use a linear model for the predictions. Hint: see ?stat_smooth:

```
ggplot(dat, aes(x = CPI, y = HDI)) + geom_point()+geom_smooth(method="lm")
```



4. Overlay a smoothing line on top of the scatter plot using geom_line. Hint: change the statistical transformation.

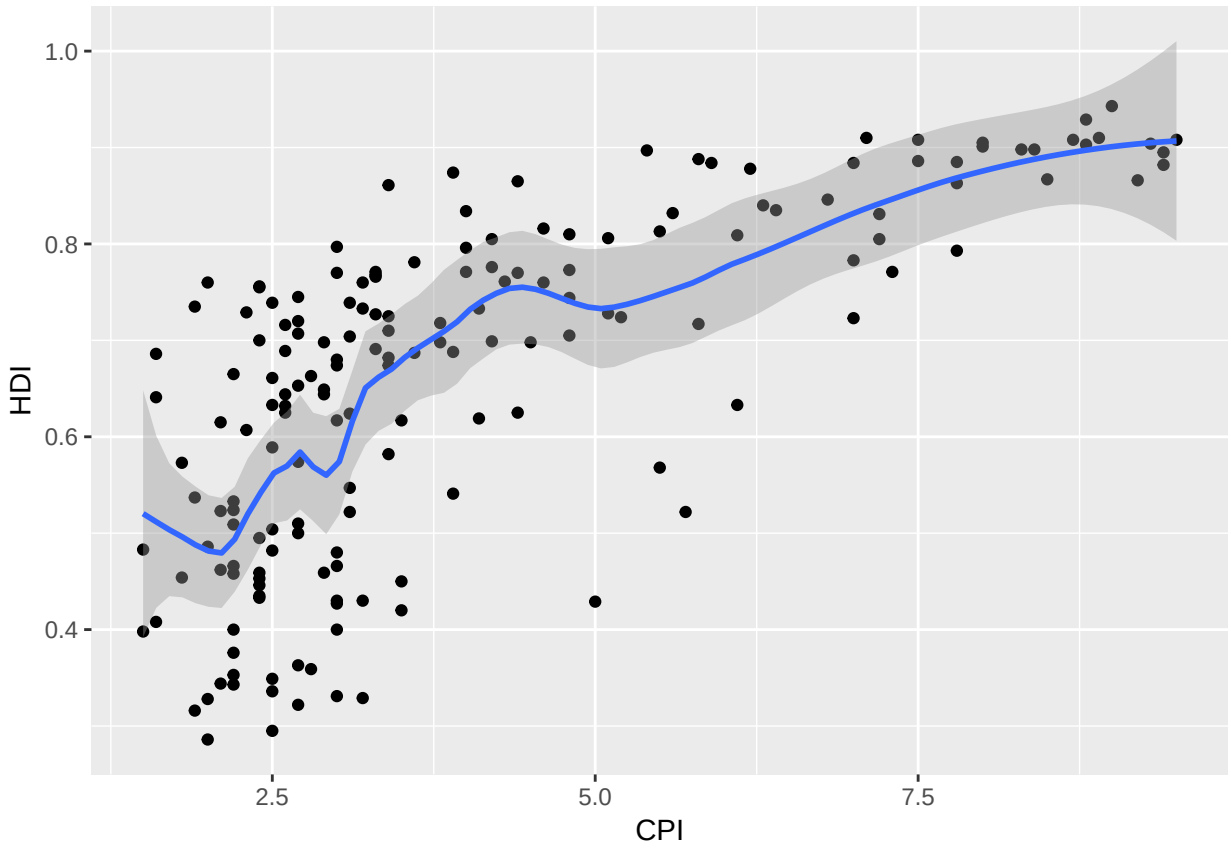
```
ggplot(dat, aes(x = CPI, y = HDI))+geom_point()+geom_line(stat = "smooth", method = "loess")
```



5. BONUS: Overlay a smoothing line on top of the scatter plot using the default loess method, but make it less smooth. Hint: see ?loess.

```
ggplot(dat, aes(x = CPI, y = HDI)) +geom_point() +geom_smooth(span = .3)
```

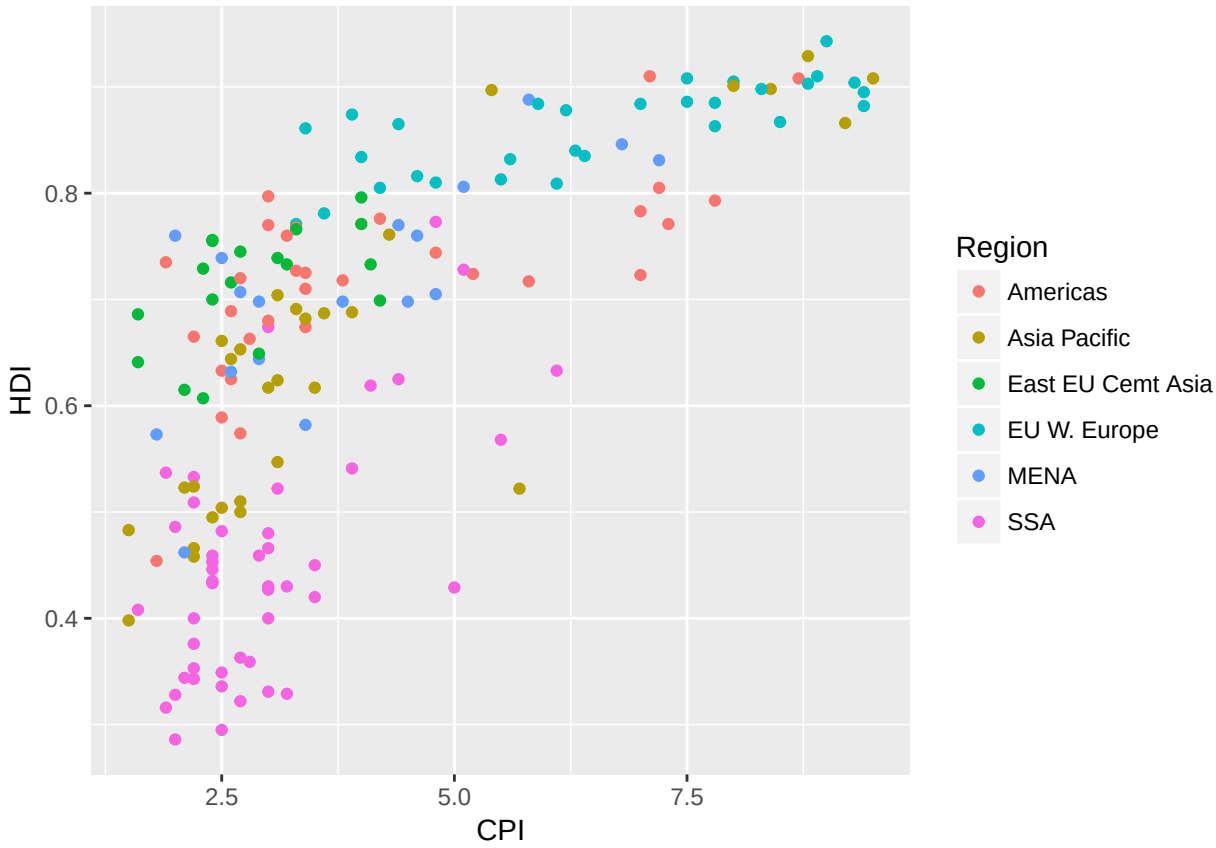
```
## `geom_smooth()` using method = 'loess'
```



Exercise III

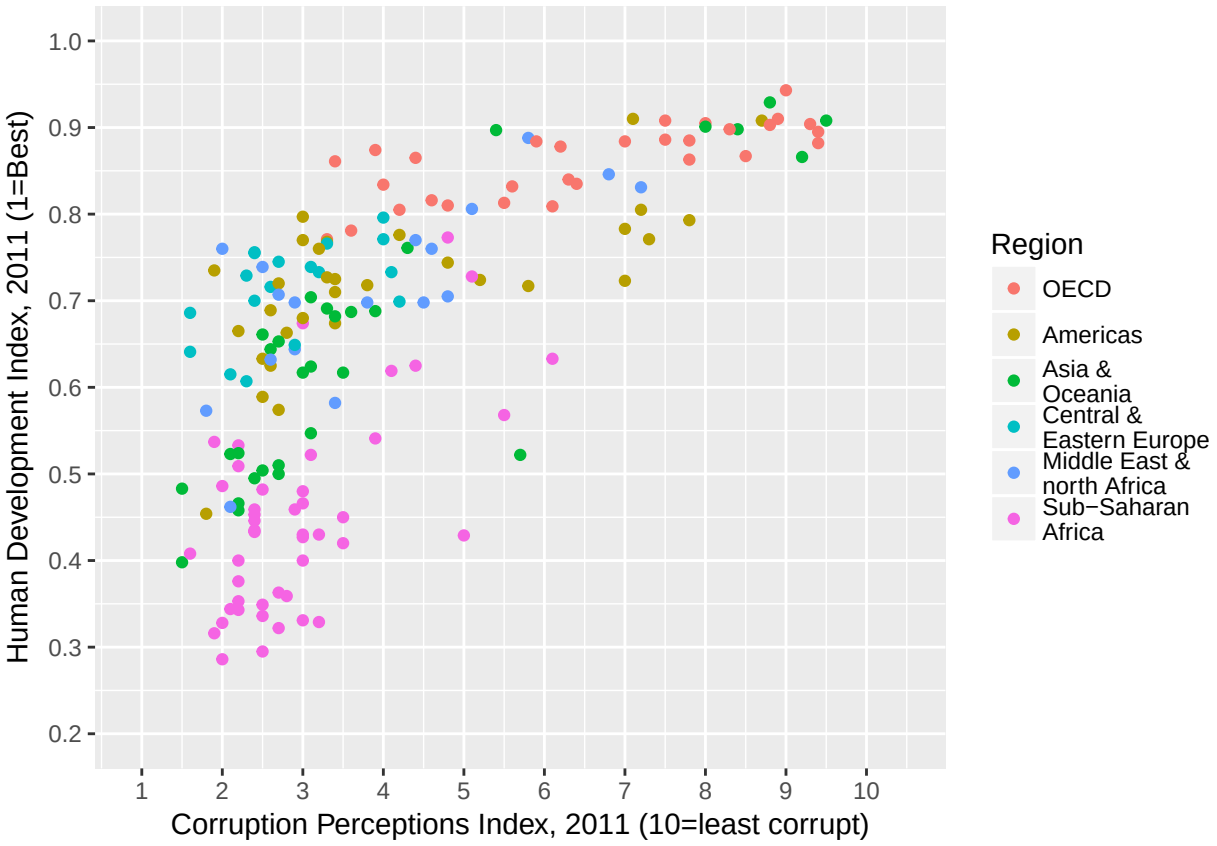
1. Create a scatter plot with CPI on the x axis and HDI on the y axis. Color the points to indicate region.

```
pc1 <- ggplot(dat, aes(x = CPI, y = HDI, color = Region)) +geom_point()
pc1
```



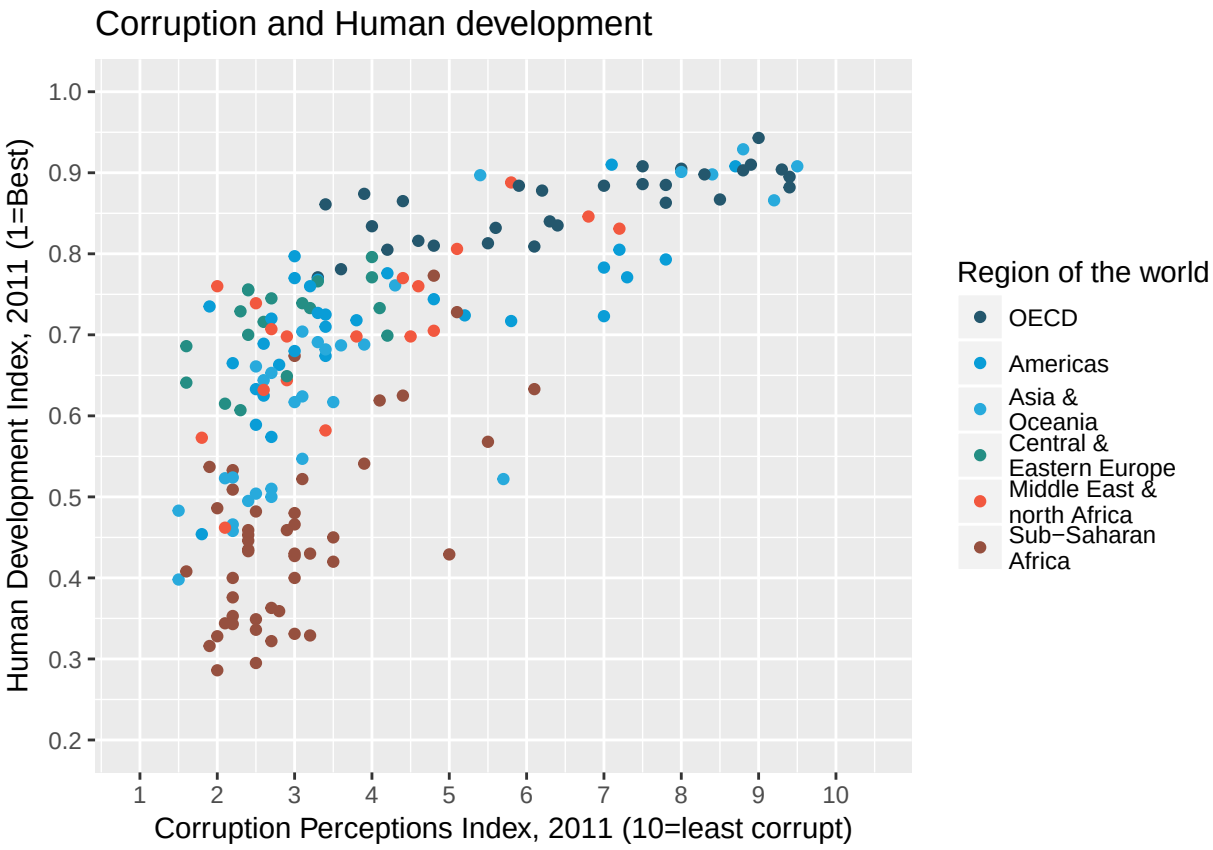
2. Modify the x, y, and color scales so that they have more easily-understood names (e.g., spell out "Human development Index" instead of "HDI").

```
dat$Region <- factor(dat$Region,
                     levels = c("EU W. Europe",
                                "Americas",
                                "Asia Pacific",
                                "East EU Cemt Asia",
                                "MENA",
                                "SSA"),
                     labels = c("OECD",
                                "Americas",
                                "Asia &\nOceania",
                                "Central &\nEastern Europe",
                                "Middle East &\nnorth Africa",
                                "Sub-Saharan\nAfrica"))
pc2 <- ggplot(dat, aes(x = CPI, y = HDI, color = Region)) +geom_point()+
scale_x_continuous(name = "Corruption Perceptions Index, 2011 (10=least corrupt)",
                       limits = c(.9, 10.5),
                       breaks = 1:10) +
scale_y_continuous(name = "Human Development Index, 2011 (1=Best)",
                       limits = c(0.2, 1.0),
                       breaks = seq(0.2, 1.0, by = 0.1))
pc2
```



3. Modify the color scale to use specific values of your choosing. Hint: see ?scale_color_manual.

```
pc3 <- pc2 + scale_color_manual(name = "Region of the world",
                        values = c("#24576D",
                                   "#099DD7",
                                   "#28AADC",
                                   "#248E84",
                                   "#F2583F",
                                   "#96503F")) +
    ggtitle("Corruption and Human development")
pc3
```


Corruption and Human development
Human Development Index, 2011 (1=Best)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
1
2
3
4
5
6
7
8
9
10
Corruption Perceptions Index, 2011 (10=least corrupt)
Region of the world
OECD
Americas
Asia & Oceania
Central & Eastern Europe
Middle East & north Africa
Sub−Saharan Africa