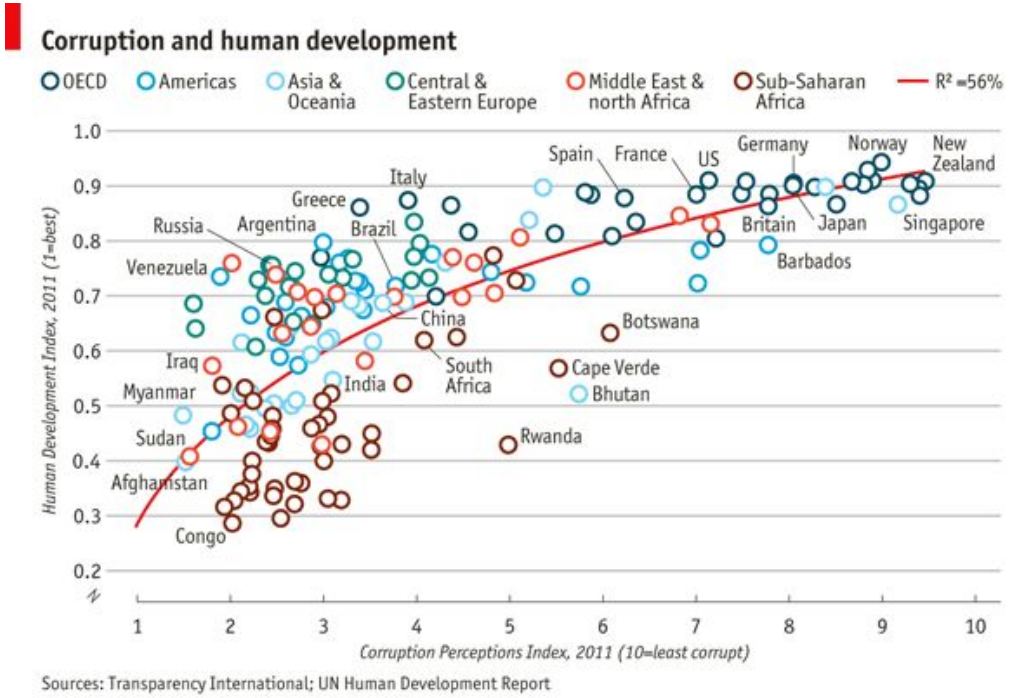
Reproduce a graph from The Economist using ggplot2

Exercise

Reproduce this graph from the Economist



https://www.economist.com/blogs/dailychart/2011/12/corruption-and-development

Exercise

https://www.rstudio.com/wp-content/uploads/2015/03/ggplot2-cheatsheet.pdf

http://ggplot2.tidyverse.org/reference/

LIBRARIES

- ggplot2
- data.table
- purrr
- magrittr

Exercise

Import, transform the data

```
###### IMPORT DATA
# read data from economist_data.csv
dt <- fread('economist_data.csv')

###### TRANSFORM AND TIDY
# check structure
str(dt)
# we have and unnecessary index column (V1)
dt[,V1:=NULL]
# Country and Region must be factors
dt[,c('Country','Region') := map(.SD,as.factor), .SDcols = c('Country','Region')]
# change column names, I want them to be lower case and with _ instead of .
setnames(dt,colnames(dt),c('country','hdi_rank','hdi','cpi','region'))
```

country	hdi	cpi	region
Afghanistan	0.398	1.5	Asia Pacific
Albania	0.739	3.1	East EU Cemt Asia
Algeria	0.698	2.9	MENA
Angola	0.486	2.0	SSA
Argentina	0.797	3.0	Americas
Vanuatu	0.617	3.5	Asia Pacific
Venezuela	0.735	1.9	Americas
Yemen	0.462	2.1	MENA
Zambia	0.430	3.2	SSA
Zimbabwe	0.376	2.2	SSA

Exercise

Import, transform the data

```
###### UNDERSTAND THE TABLE
# number of countries (173)
dt[,uniqueN(country)]
# number of regions (6)
dt[,uniqueN(region)]
# number of countries per region
dt[,uniqueN(country),by=region]
# quick numeric descriptions
dt[,c('hdi','cpi'),with=F] %>% summary()
```

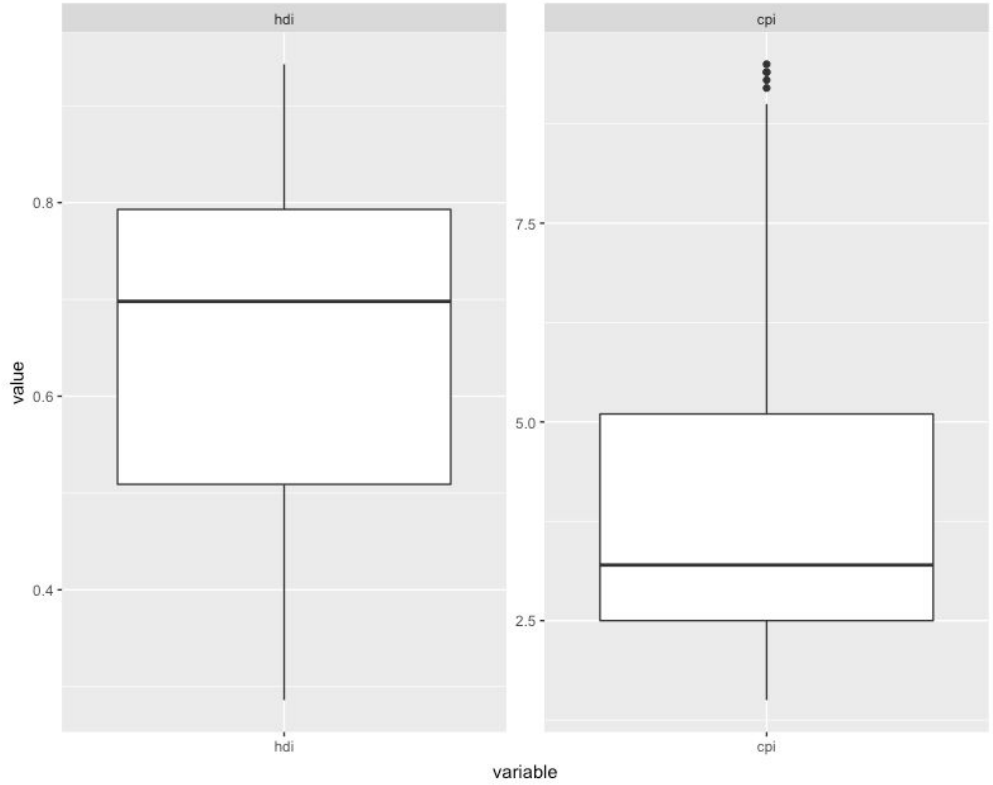
```
            region V1
        Asia Pacific 30
East EU Cemt Asia 18
               MENA 18
                SSA 46
           Americas 31
      EU W. Europe 30
```

```
      hdi              cpi
 Min.   :0.2860   Min.   :1.500
 1st Qu.:0.5090   1st Qu.:2.500
 Median :0.6980   Median :3.200
 Mean   :0.6581   Mean   :4.052
 3rd Qu.:0.7930   3rd Qu.:5.100
 Max.   :0.9430   Max.   :9.500
```

Exercise

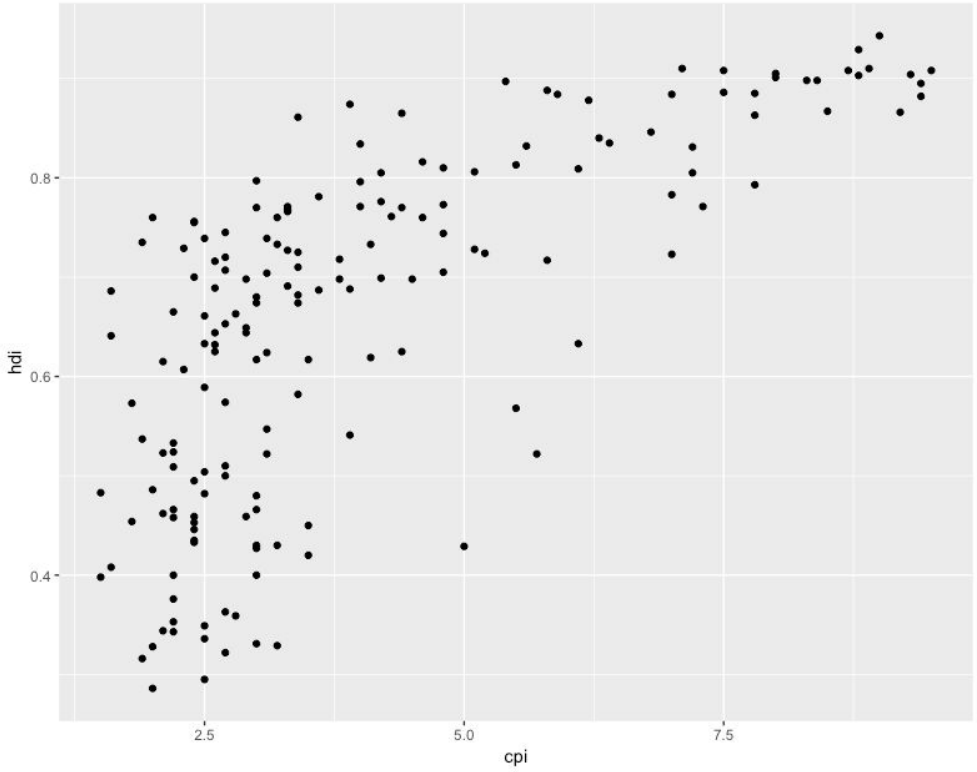
Visual description for numeric variables (quick boxplot)

```
# quick visual description
dt %>% melt(measure.vars = c('hdi','cpi')) %>%
  ggplot(aes(y=value,x=variable)) +
  geom_boxplot() +
  facet_wrap(~variable,scales='free')
```



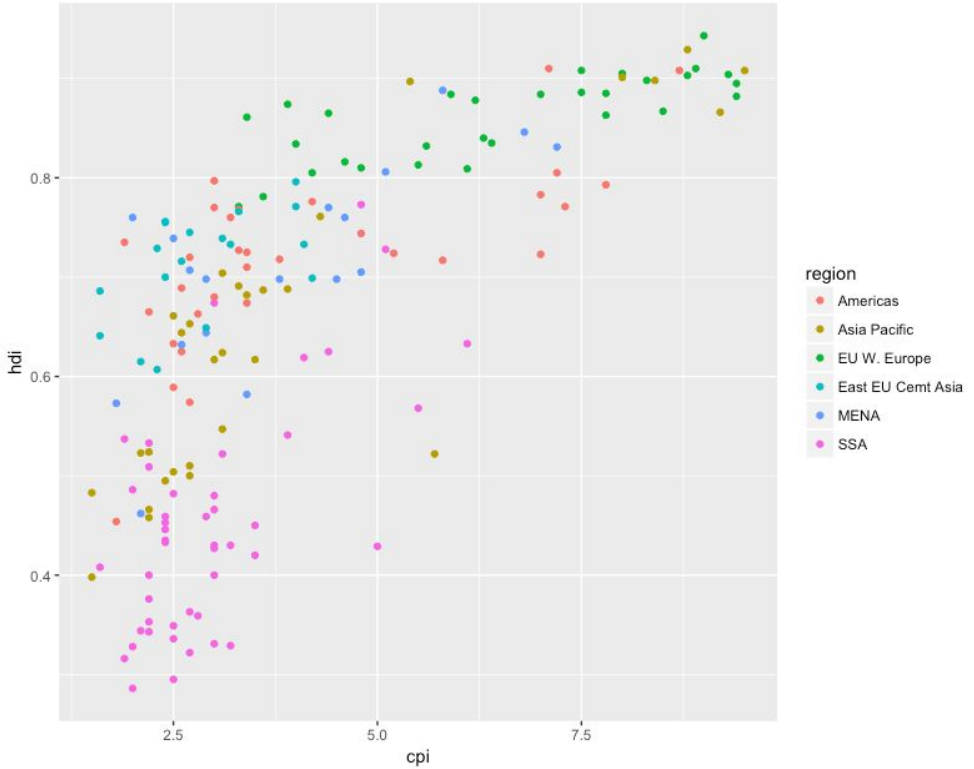
Exercise

1. **Plot this basic scatter plot**



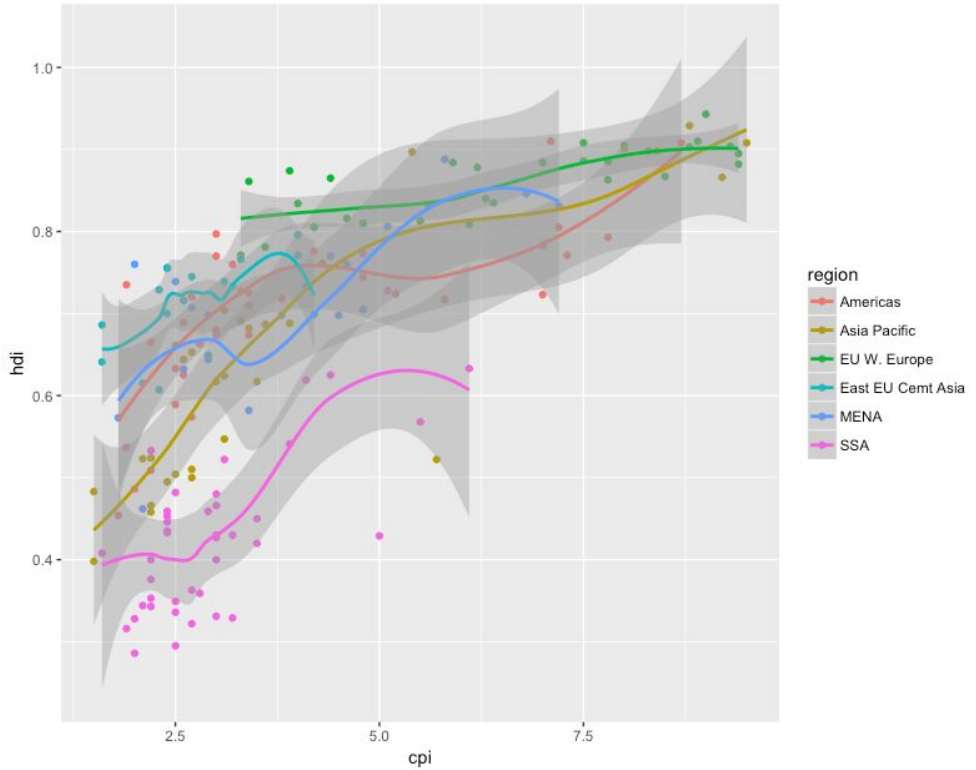
Exercise

2. Add region dimension using color aesthetic



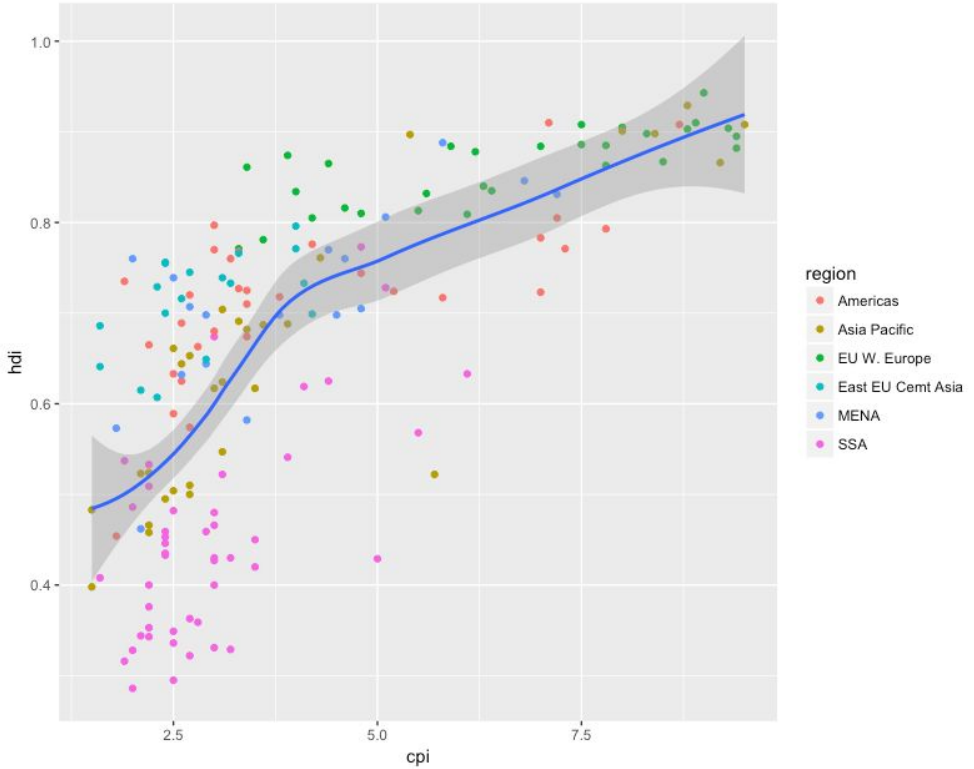
Exercise

3. Add a smooth model



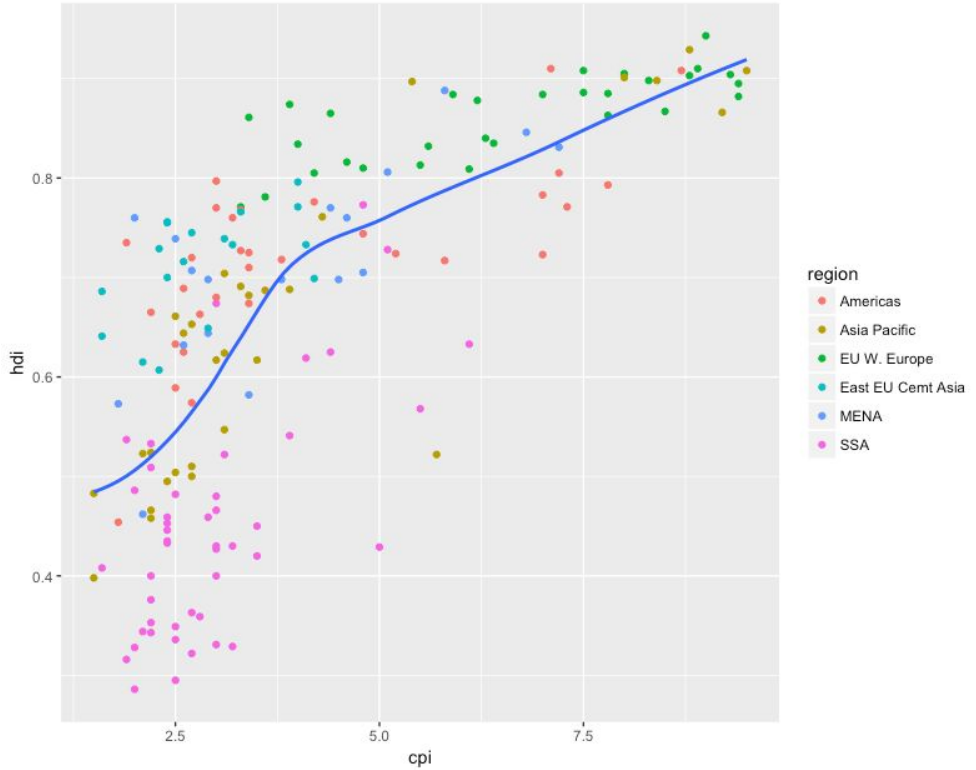
Exercise

3. Change relation between aesthetics and geometrics to plot just one model



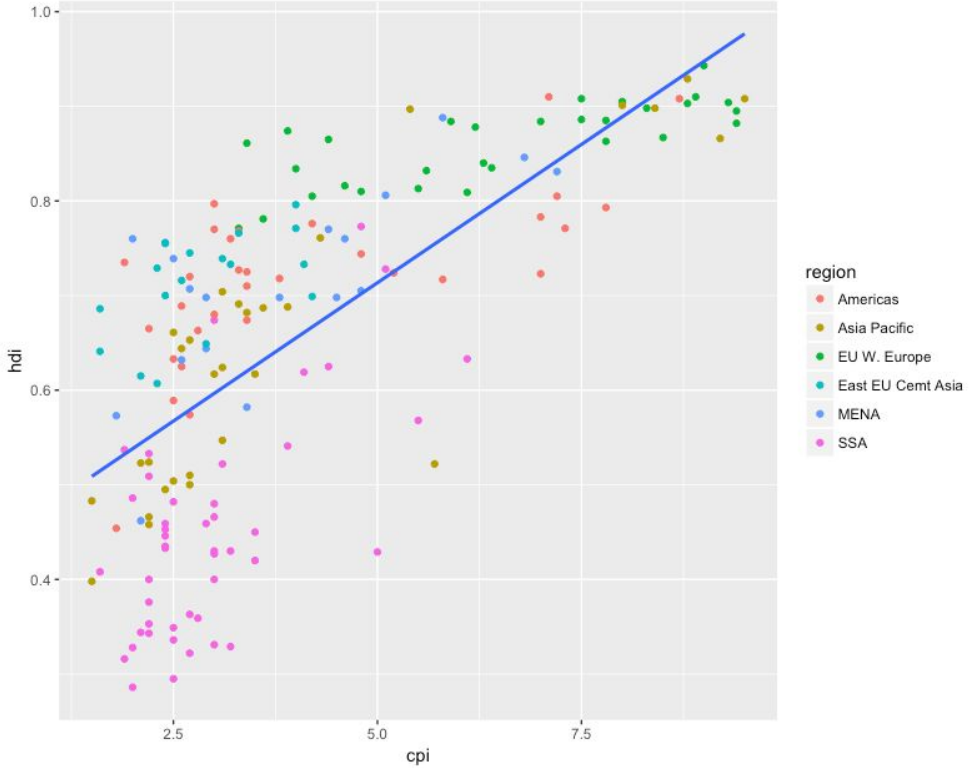
Exercise

5. Remove standard error (se) from the model



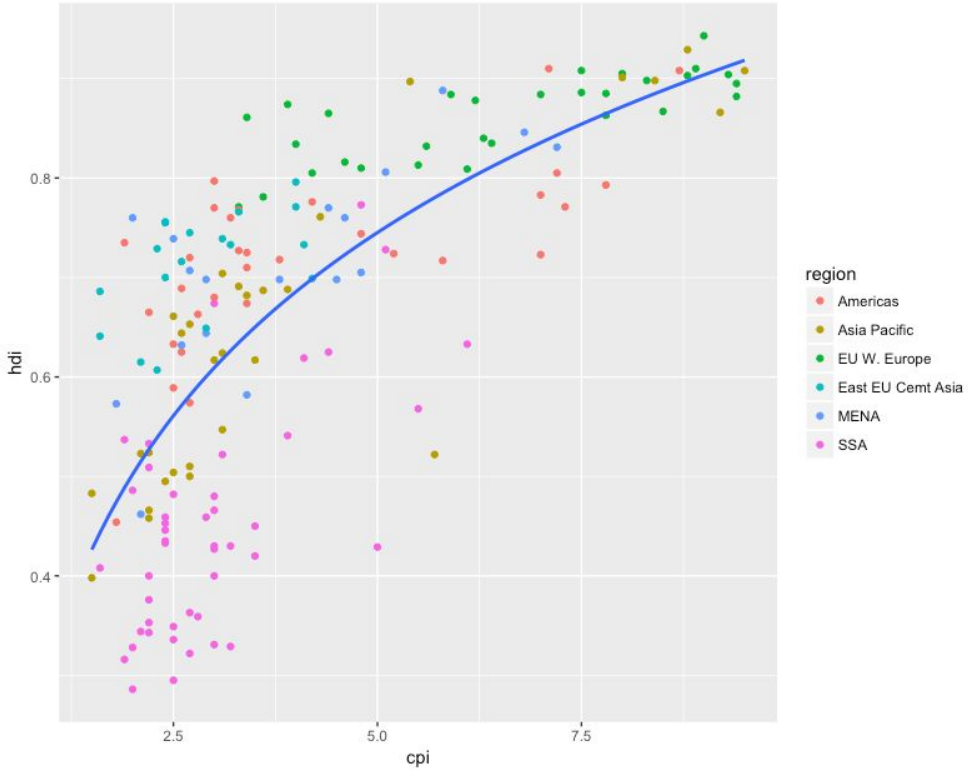
Exercise

6. Change the default method of the model (change it to lm)



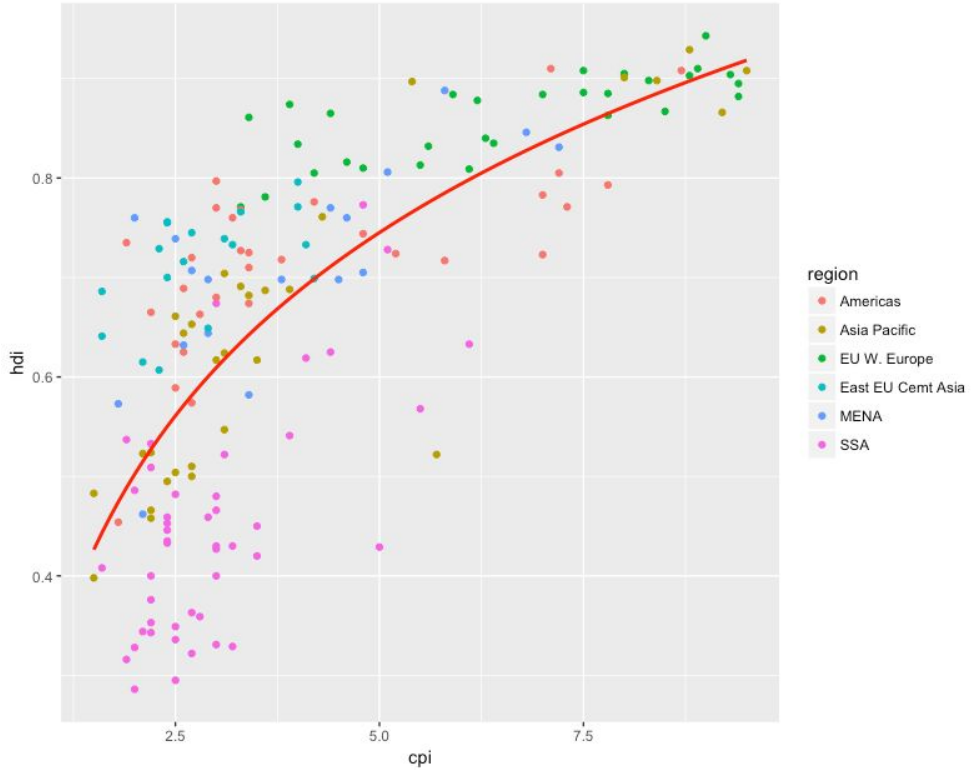
Exercise

7. Change the default formula of the model to y ~ x + log(x)



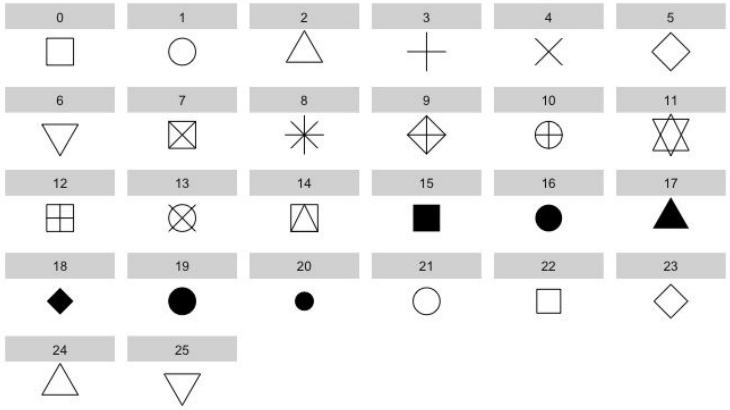
Exercise

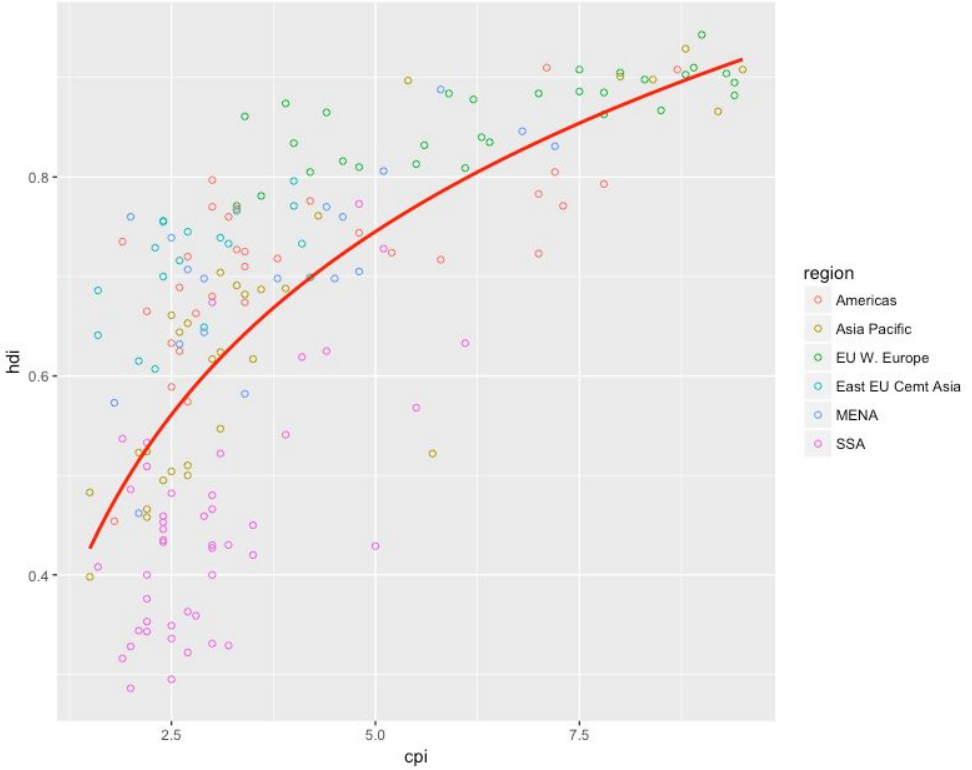
8. Change the default color of the model line (change it to red)



Exercise

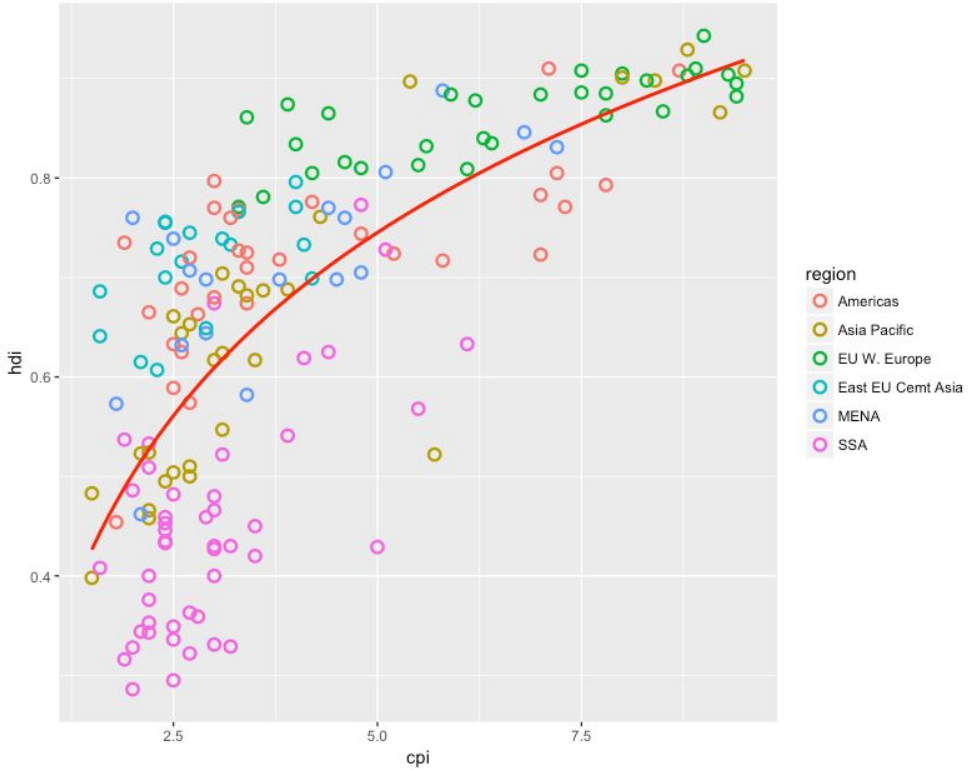
9. Change the shape of points. We want use open points (shape 1)





Exercise

10. Change the size (to 2.5) and the stroke (to 1.25) of points



Exercise

11. Add a new aesthetic to the model called linetype and force it to be ‘r2’

Exercise

12. Save the object into a variable (for instance the variable gg)

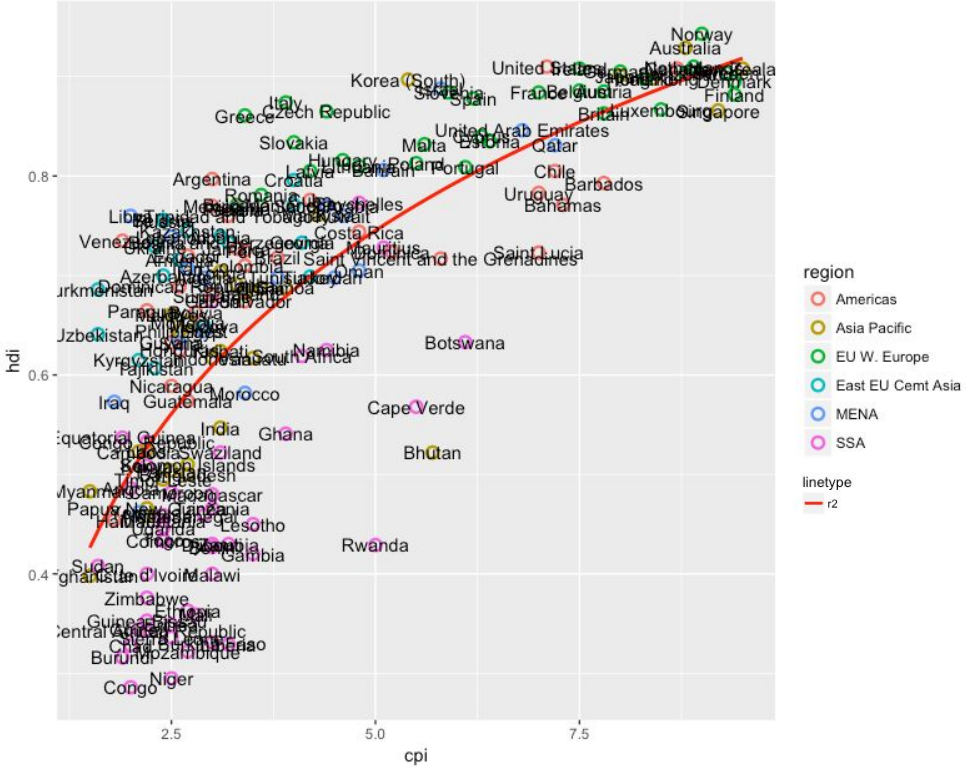
Exercise

13. Labelling points (create a vector with the names of the countries to label)

```
countries_to_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", "Japan",
                       "New Zealand", "Singapore")
```

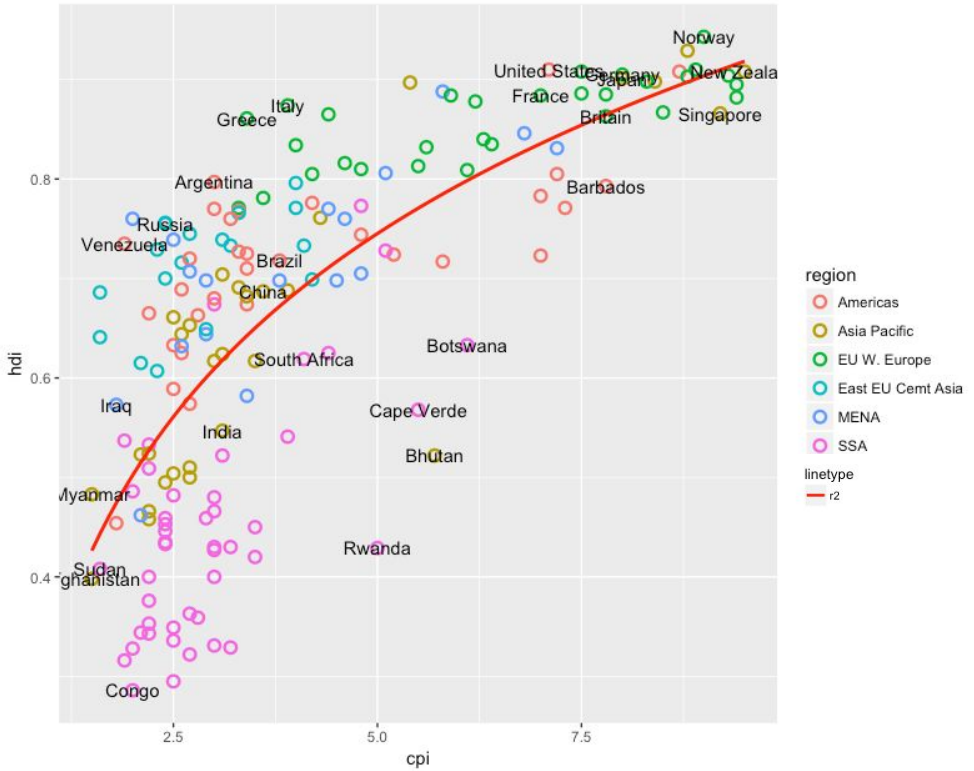
Exercise

14. Add labels for all countries



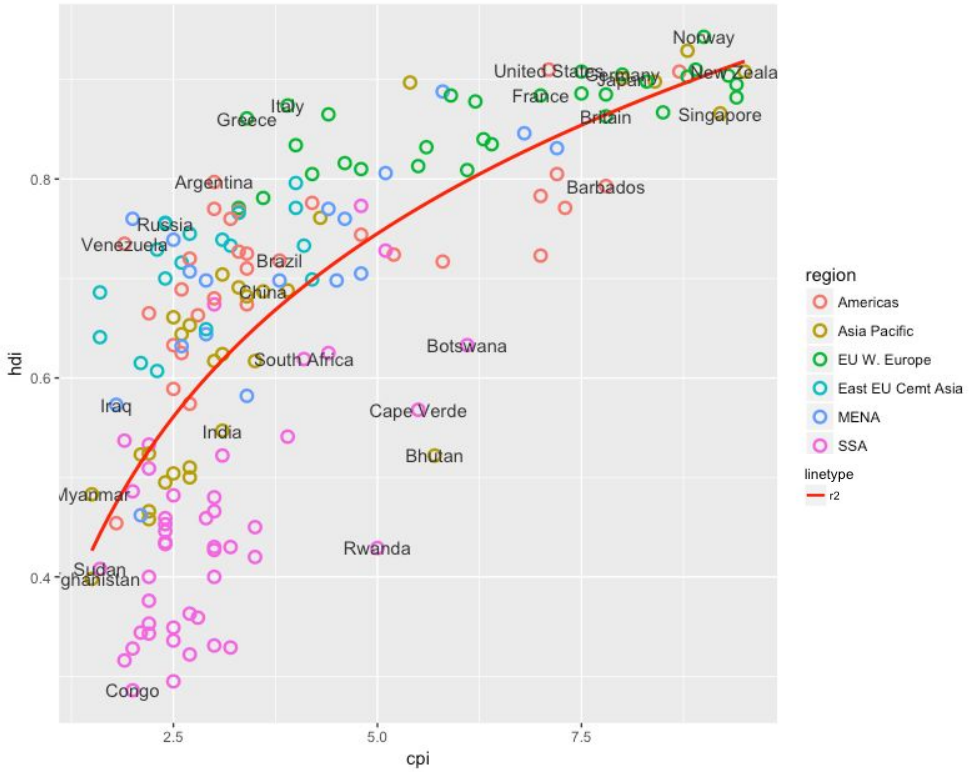
Exercise

15. Add labels for countries in countries_to_label (hint use subset function)



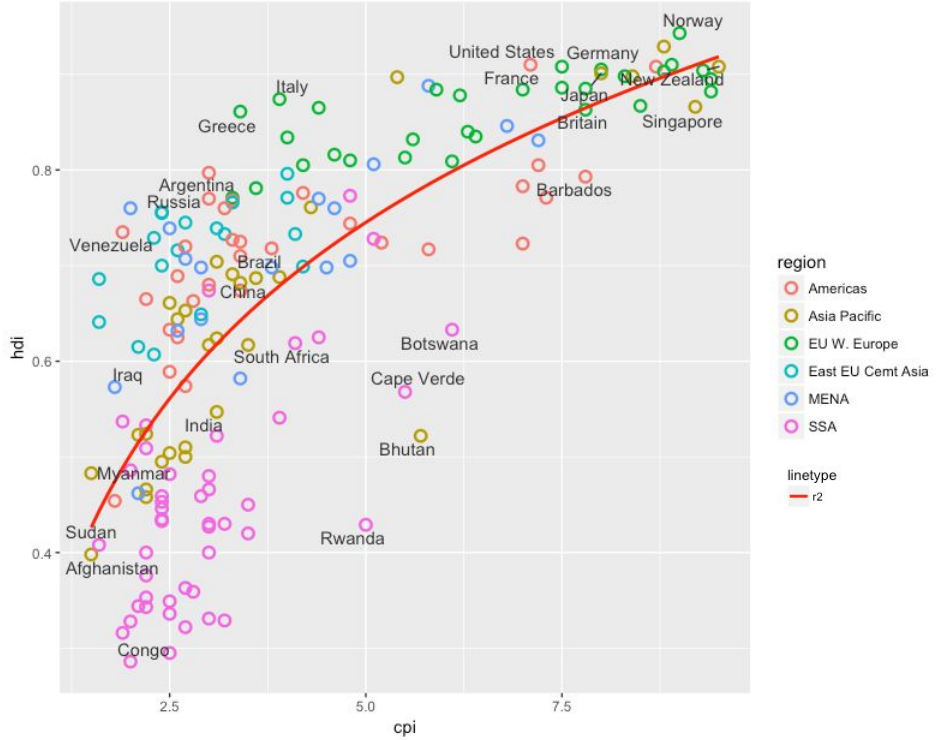
Exercise

16. Change the default color of the text (change it to grey20)



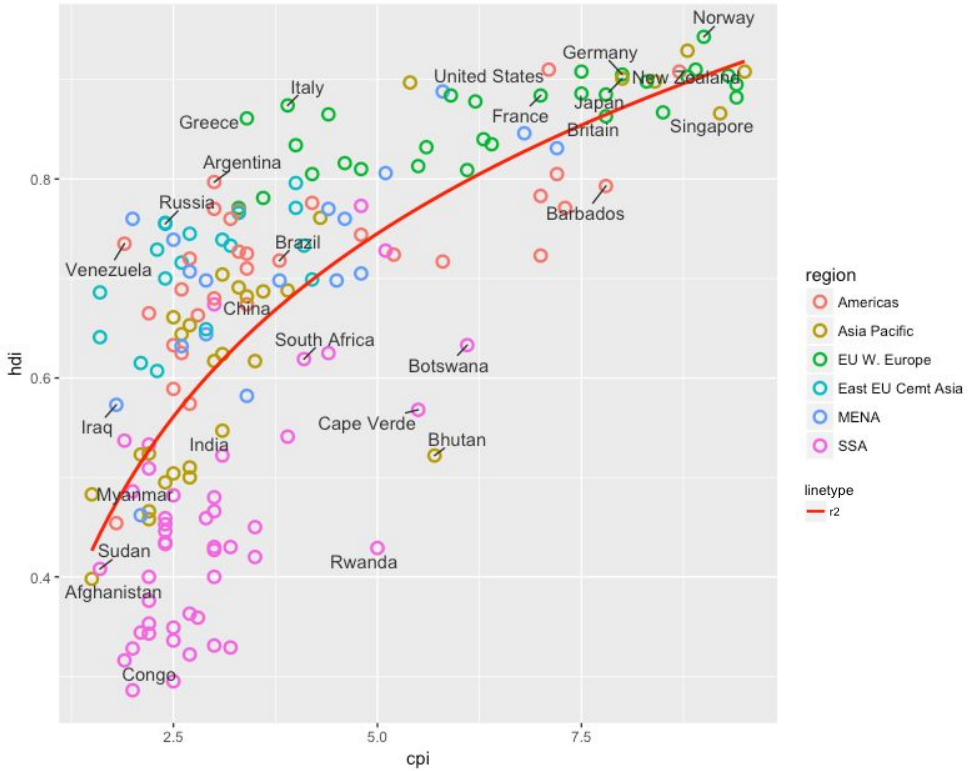
Exercise

17. Use a ggplot extension (ggrepel) to ad link lines between the name and the point (load ggrepel library and change geom_text to geom_text_repel)



Exercise

18. Add an aesthetic to *geom_text_repel* called force and force it to be 10.



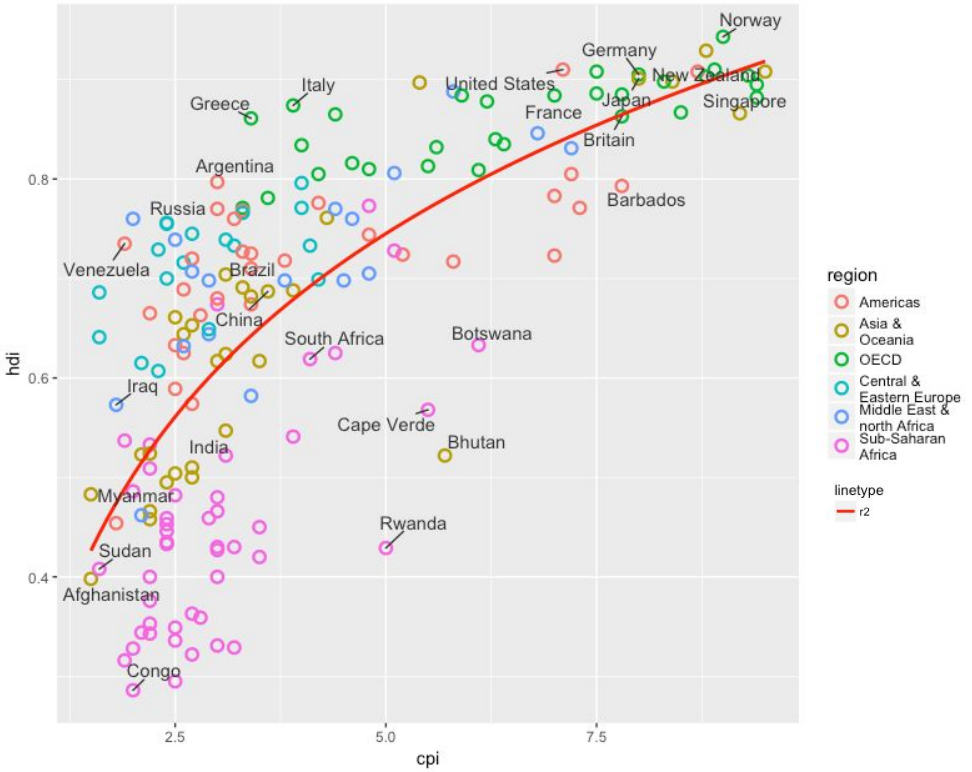
Exercise

19. Change the names of the levels of the factor variable “region”

```
factor_new_levels <- list("EU W. Europe" = "OECD",
                          "Americas" = "Americas",
                          "Asia Pacific" = "Asia &\nOceania",
                          "East EU Cemt Asia" = "Central &\nEastern Europe",
                          "MENA" = "Middle East &\nnorth Africa",
                          "SSA" = "Sub-Saharan\nAfrica")
factor_new_levels_vector <- unlist(factor_new_levels)
new_levels <- unname(factor_new_levels_vector[levels(dt$region)])
levels(dt$region) <- new_levels
```

Exercise

20. Create again the same plot with the new dataset (loot at the legend)



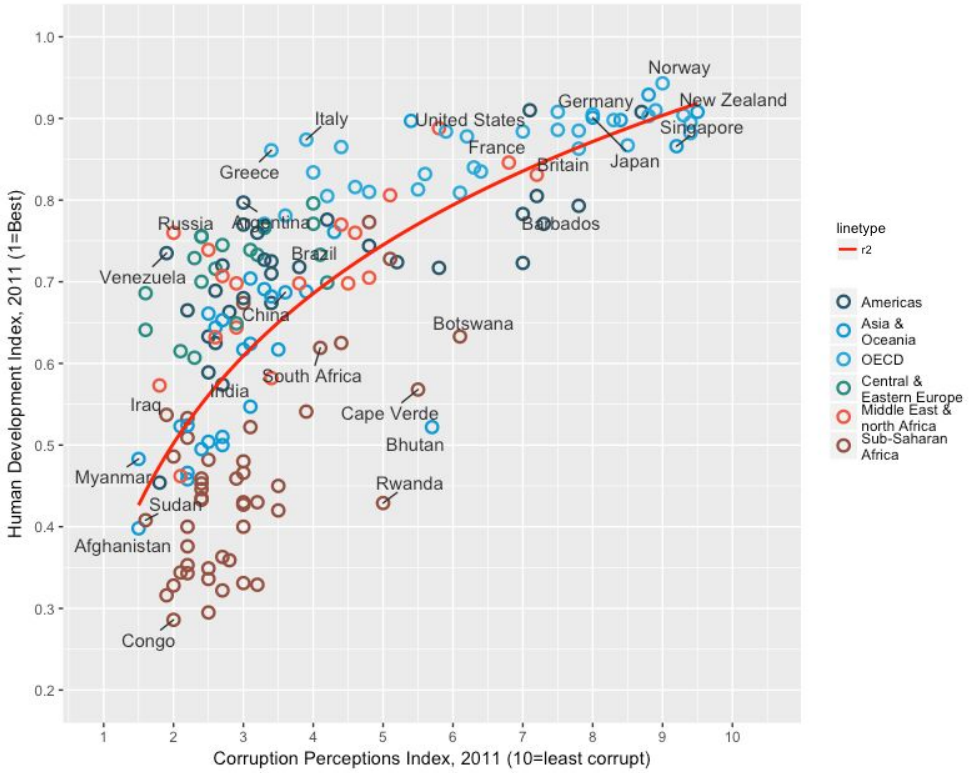
Exercise

21. Change aesthetics scales (x, y and color scales)

```
x_scale <- scale_x_continuous(name = "Corruption Perceptions Index, 2011 (10=least corrupt)",
                              limits = c(.9, 10.5),
                              breaks = 1:10)
y_scale <- 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))
color_scale <- scale_color_manual(name = "",
                                  values = c("#24576D","#099DD7",
                                             "#28AADC","#248E84",
                                             "#F2583F","#96503F"))
```

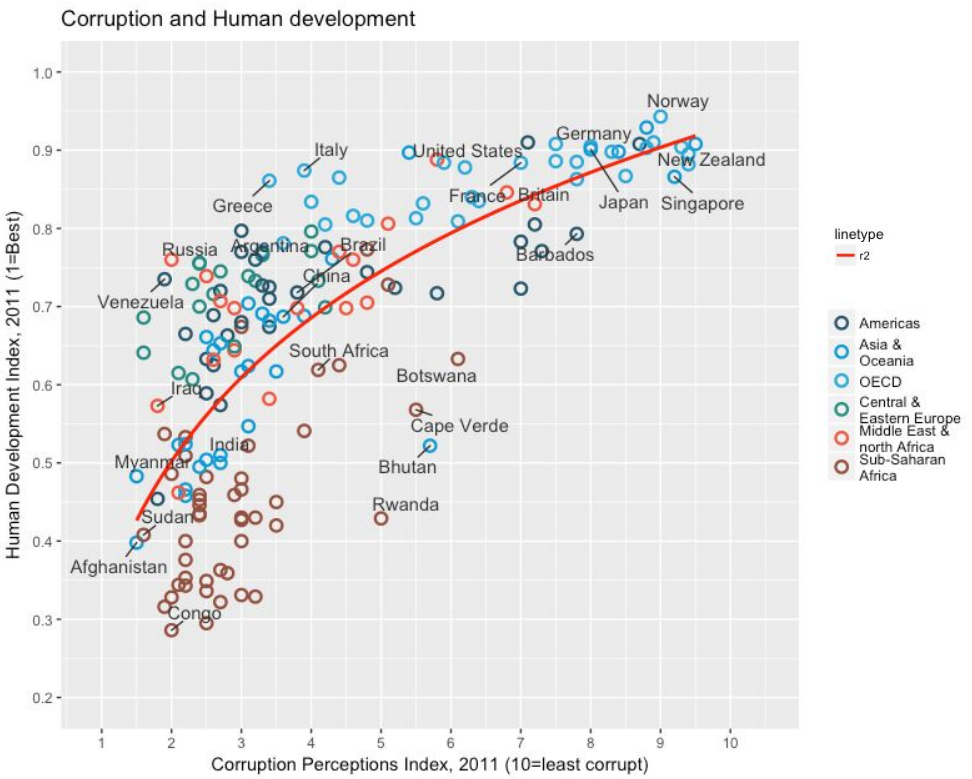
Exercise

22. Apply new scales.



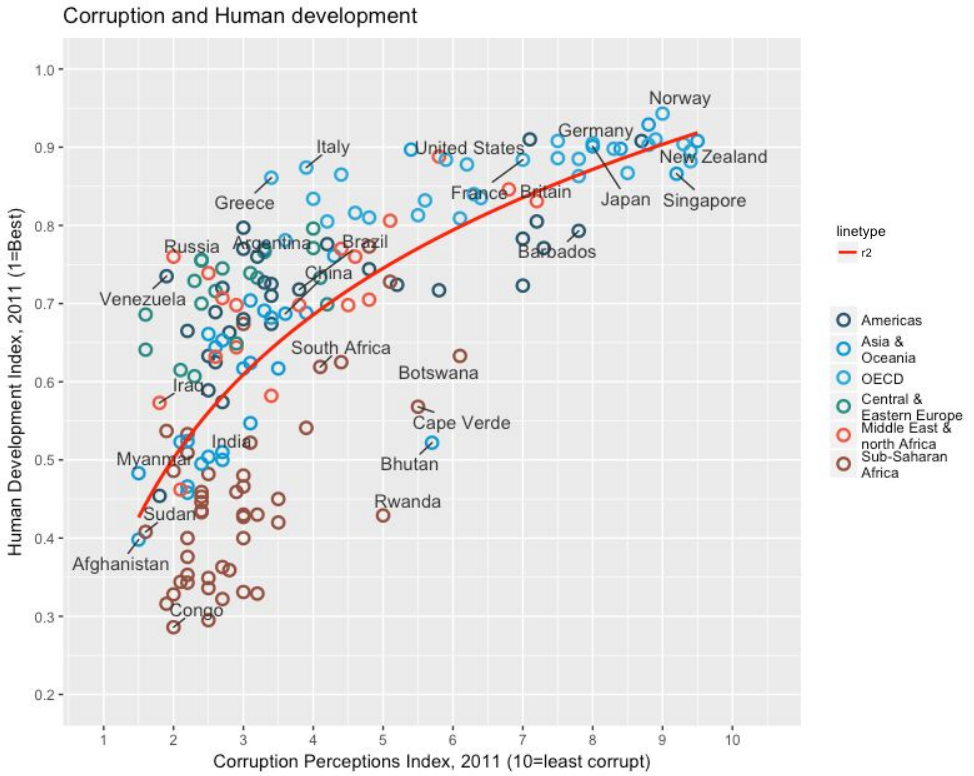
Exercise

23. Add a title



Exercise

24. Save everything made in gg variable



Exercise

25. Create a custom theme based on theme_minimal (save it in economist_theme)

```
theme_minimal() + # start with a minimal theme and add what we need
  theme(text = element_text(color = "gray20"),
        legend.position = c("top"), # position the legend in the upper left
        legend.direction = "horizontal",
        legend.justification = 0.1, # anchor point for legend.position.
        legend.text = element_text(size = 11, color = "gray10"),
        axis.text = element_text(face = "italic"),
        axis.title.x = element_text(vjust = -1), # move title away from axis
        axis.title.y = element_text(vjust = 2), # move away for axis
        axis.ticks.y = element_blank(), # element_blank() is how we remove elements
        axis.line = element_line(color = "gray40", size = 0.5),
        axis.line.y = element_blank(),
        panel.grid.major = element_line(color = "gray50", size = 0.5),
        panel.grid.major.x = element_blank()) -> economist_theme
```

Exercise

26. Create the title for linetype scale

```r
lm <- summary(lm(hdi ~ cpi + log(cpi), data = dt))
mr_2 <- paste0(format(lm$r.squared, digits = 2), "%")
lm
```

```
Call:
lm(formula = hdi ~ cpi + log(cpi), data = dt)

Residuals:
     Min       1Q   Median       3Q      Max
-0.31615 -0.07136  0.01637  0.07576  0.25789

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.31859    0.04020   7.925 2.86e-13 ***
cpi          0.00125    0.01991   0.063  0.95003
log(cpi)     0.26115    0.08862   2.947  0.00366 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1222 on 170 degrees of freedom
Multiple R-squared:  0.5213,    Adjusted R-squared:  0.5157
F-statistic: 92.56 on 2 and 170 DF,  p-value: < 2.2e-16
```

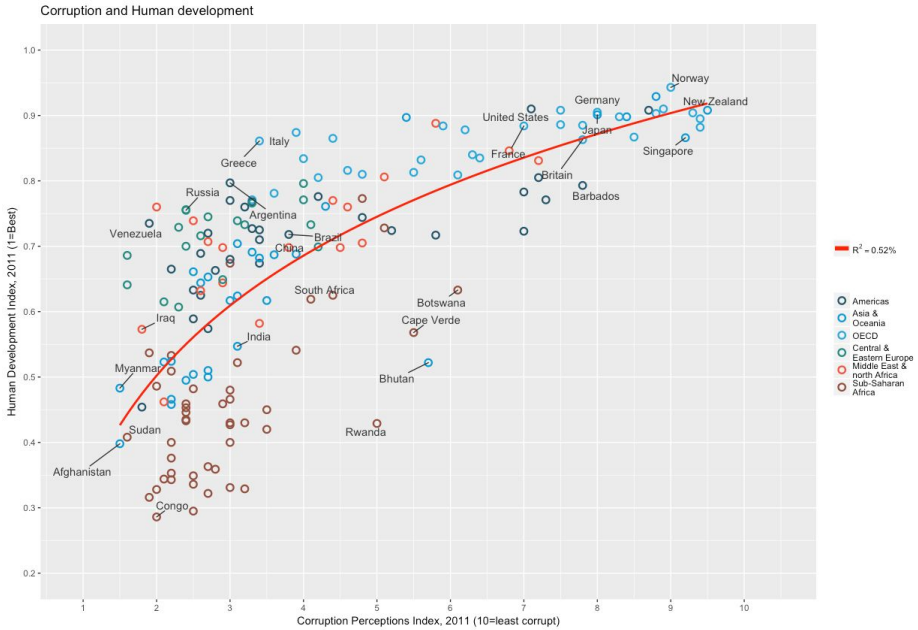
Exercise

27. Add two more scales (alpha scale and linetype scale)

```
alpha_scale <-  scale_alpha_discrete(range = c(0, 1),
                     guide = FALSE)
linetype_sclae <-
  scale_linetype(name = "",
                 breaks = "r2",
                 labels = list(bquote(R^2==.(mr_2))),
                 guide = guide_legend(override.aes = list(linetype = 1, size = 2, color = "red")))
```

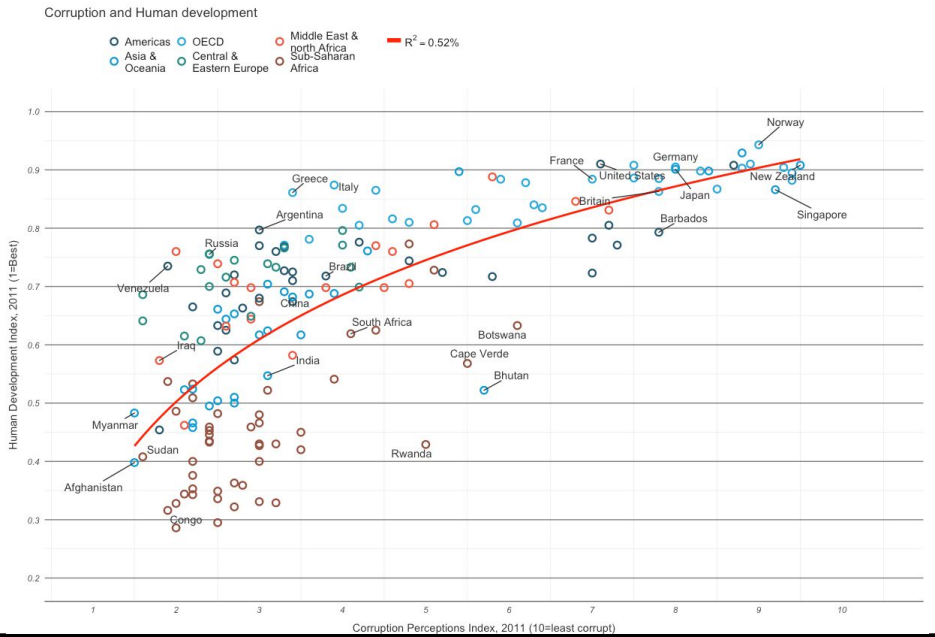
Exercise

28. Add the two new scales in gg object



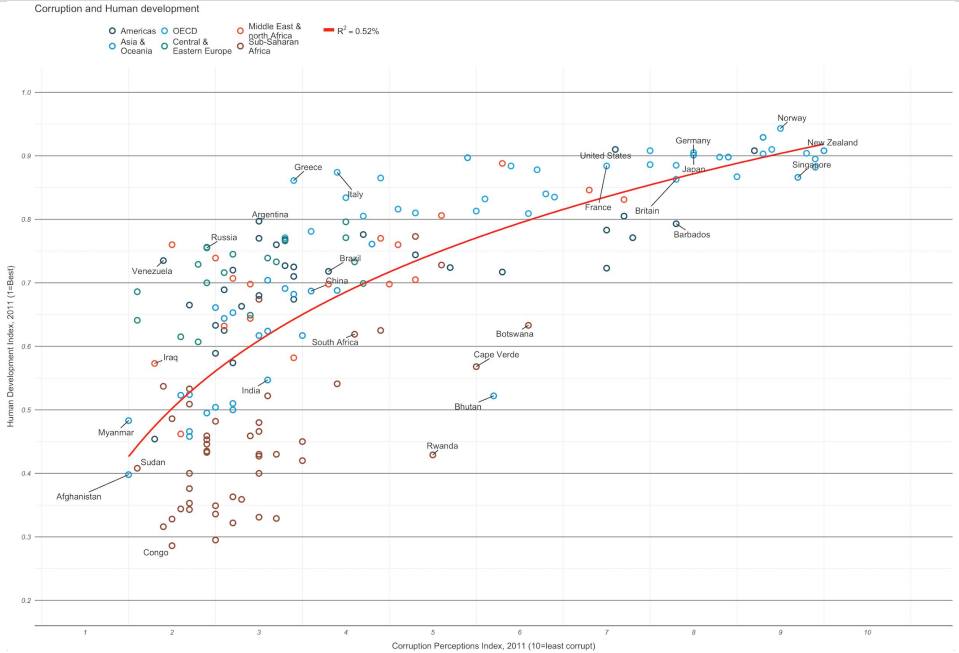
Exercise

29. Add the economist theme



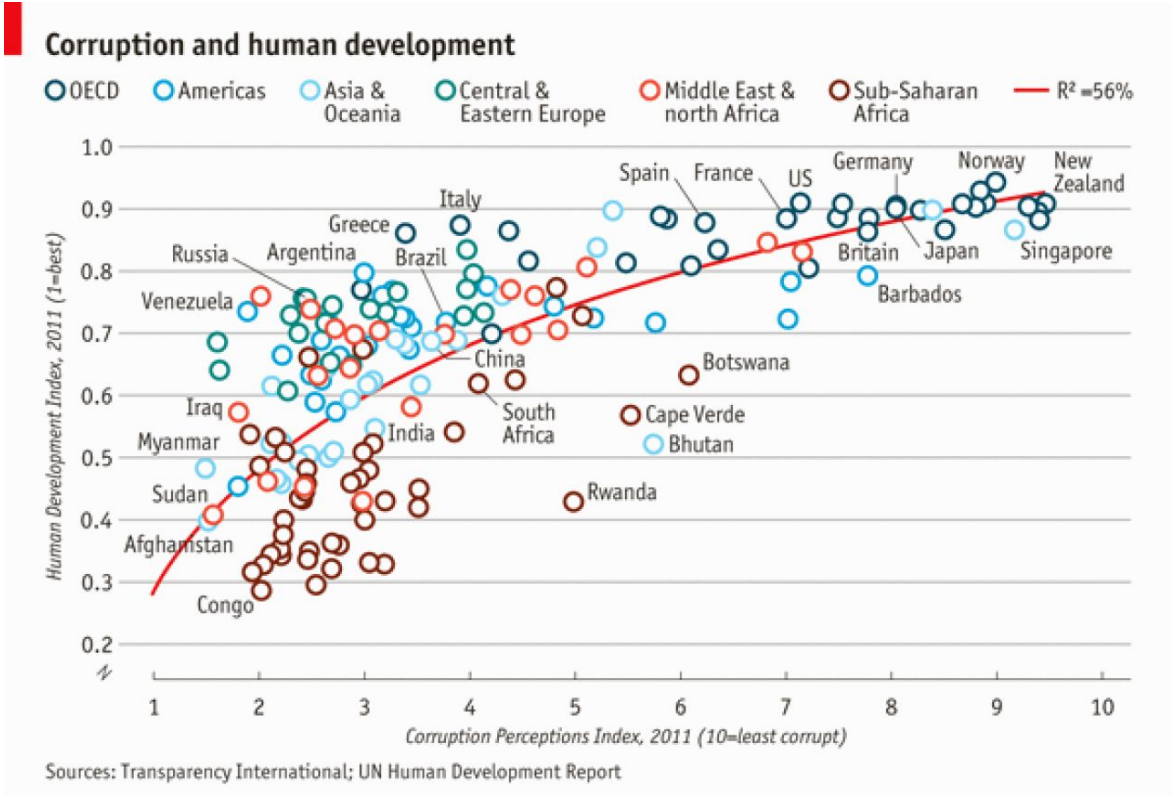
Exercise

30. Save in a png (hint use ggsave function)



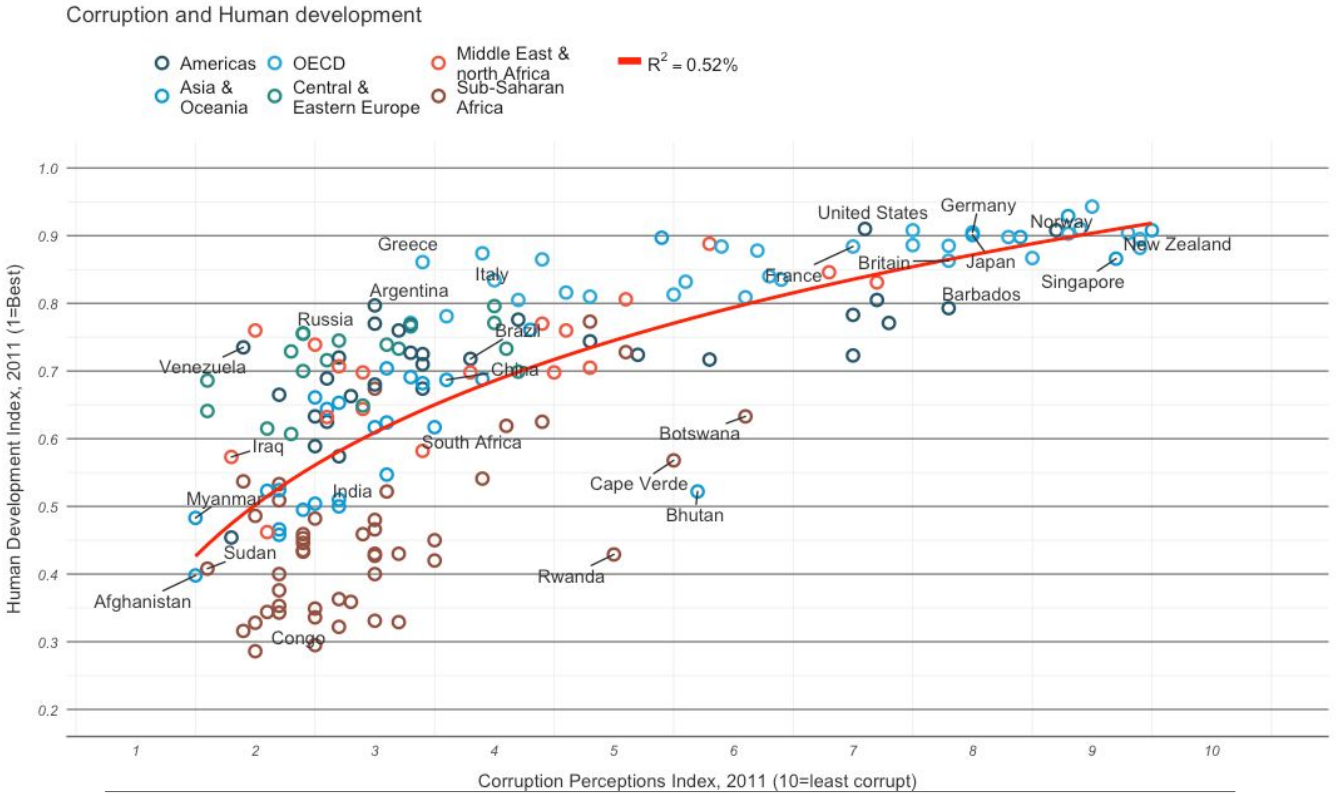
Comparison

ORIGINAL



Comparison

GGPLOT



END