# transform

[this file](http://www.dartistics.com/data/gadata_example_2.csv)

**web\_data <- read.csv("gadata\_example\_2.csv", stringsAsFactors = FALSE)**

Q1 : Calculate the average pageviews per day by channel using the summarise()function in the dplyr package.

**group\_by(web\_data, deviceCategory, channelGrouping) %>% summarise(mean(pageviews))**

Q2 : Add two columns to **web\_data** using the mutate() function: **pvs\_per\_session** and **bounce\_rate.**

**web\_data <- mutate(web\_data, pvs\_per\_session = pageviews / sessions, bounce\_rate = bounces / entrances)**

Q3 : Calculate the average bounce rate by **deviceCategory** (be careful with the order of your calculations!).

**group\_by(web\_data,deviceCategory) %>%summarise(entrances = sum(entrances), bounces = sum(bounces)) %>% mutate(bounce\_rate = bounces / entrances)**

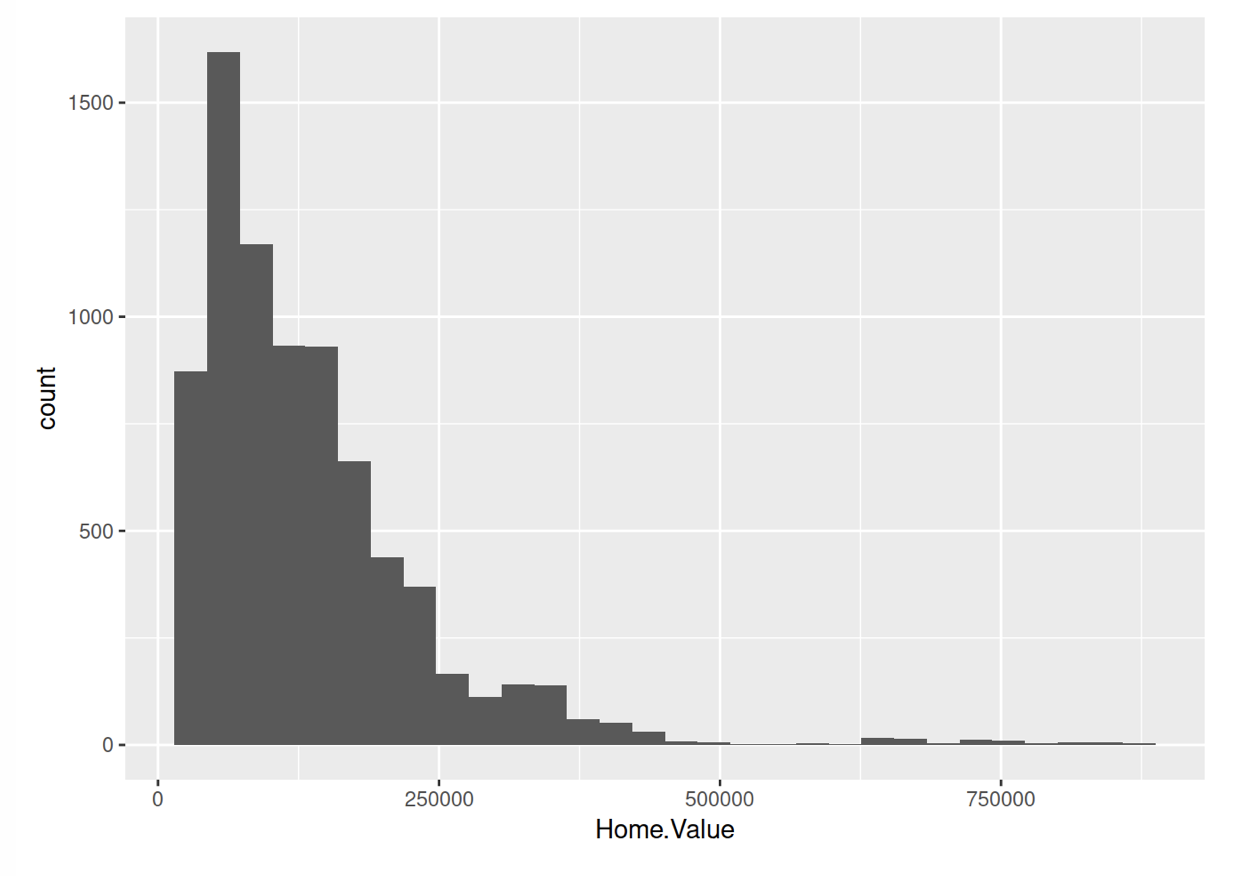
# ggplot2

Download the datasets from :

https://drive.google.com/open?id=1JythvpkgoSP64qB8Jx48lltqobVuUZA0

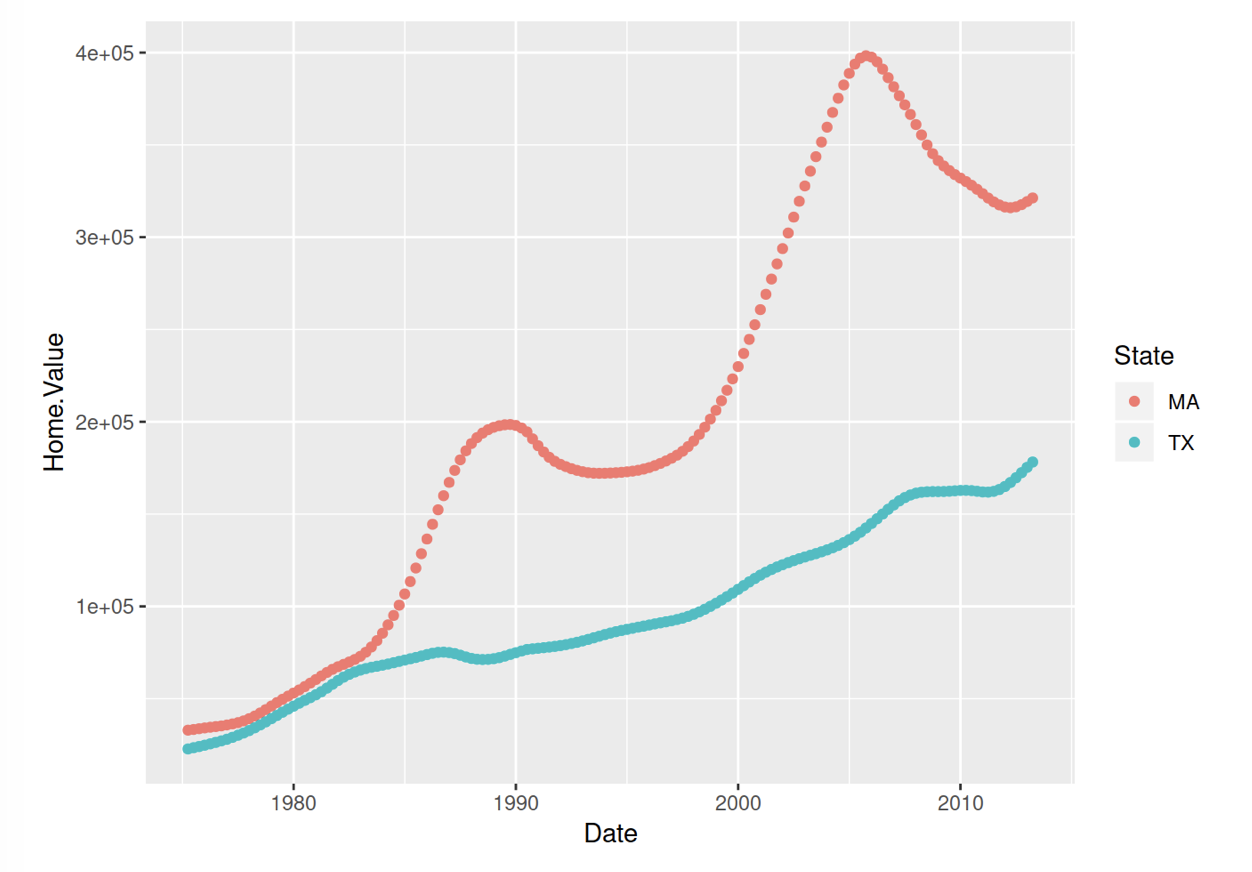
housing <- read\_csv("dataSets/landdata-states.csv")

1. Housing value histogram



**library(ggplot2)**

**ggplot(housing, aes(x = Home.Value)) + geom\_histogram()**

1. Scatterplot for housing with selected states (Hint : %in%)

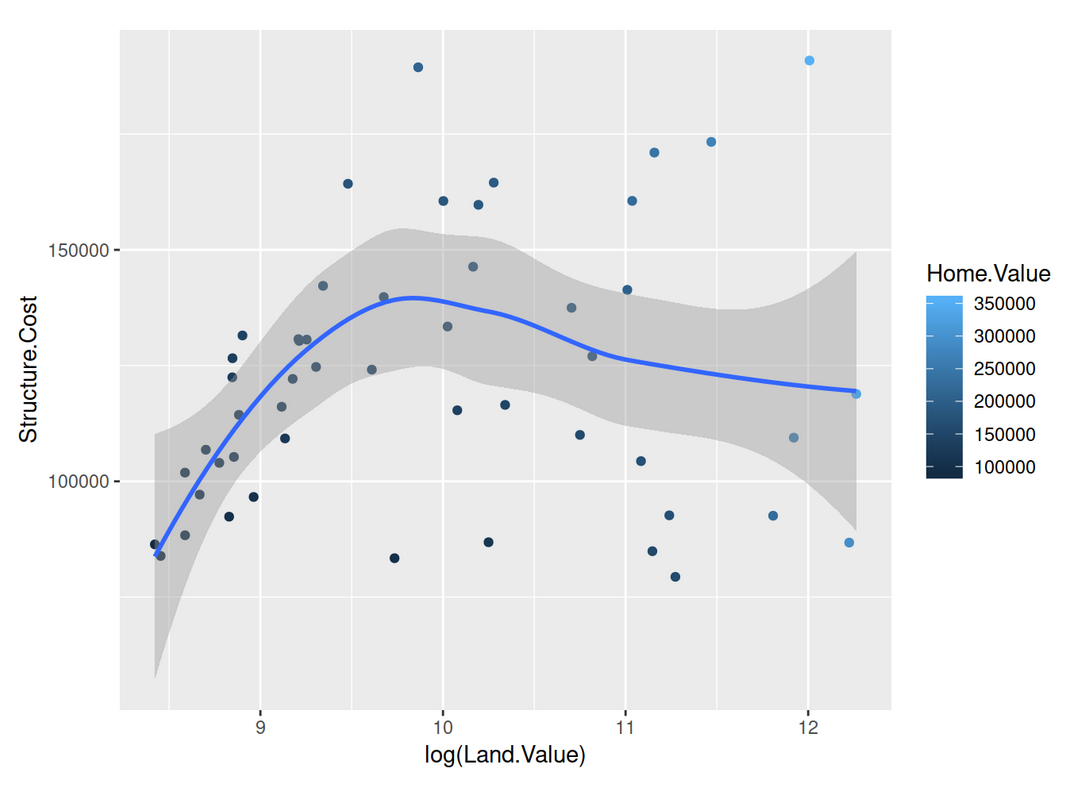
**ggplot(filter(housing, State %in% c("MA", "TX")), aes(x=Date, y=Home.Value, color=State))+ geom\_point()**

Following problems are used on a smaller housing data set. ->

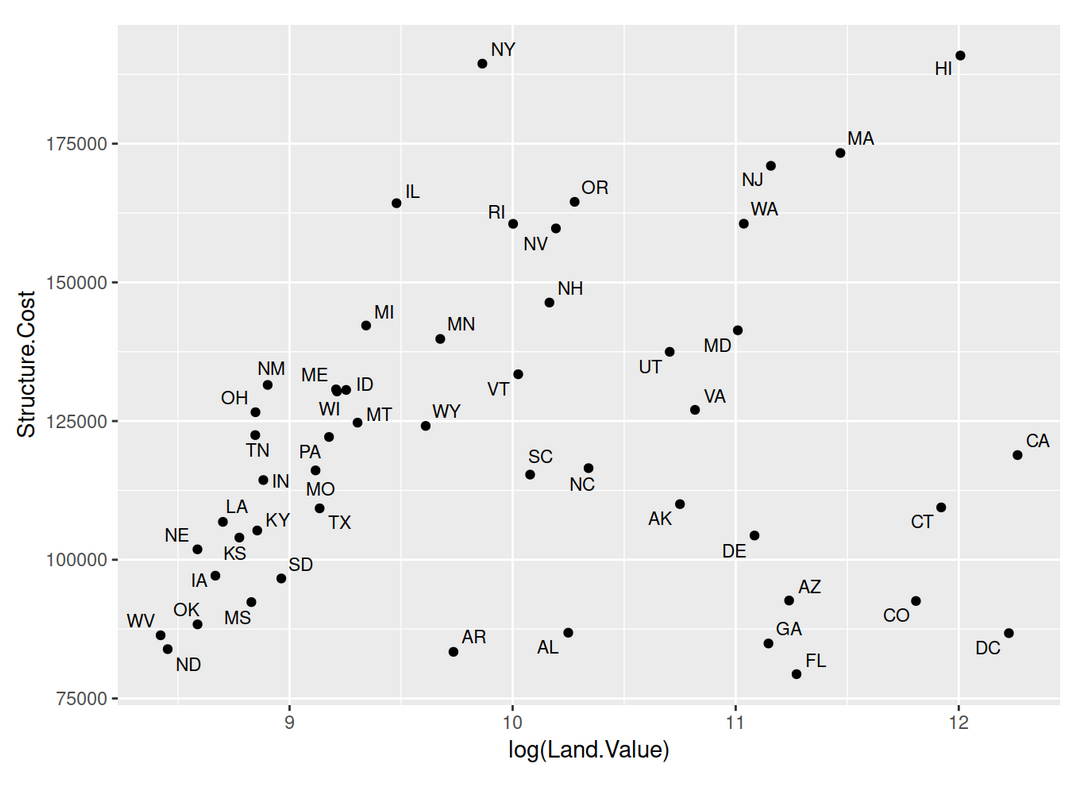
hp2001Q1 <- filter(housing, Date == 2001.25)

Q3-5 should be solved using the above dataset.

1. Having 2 geoms

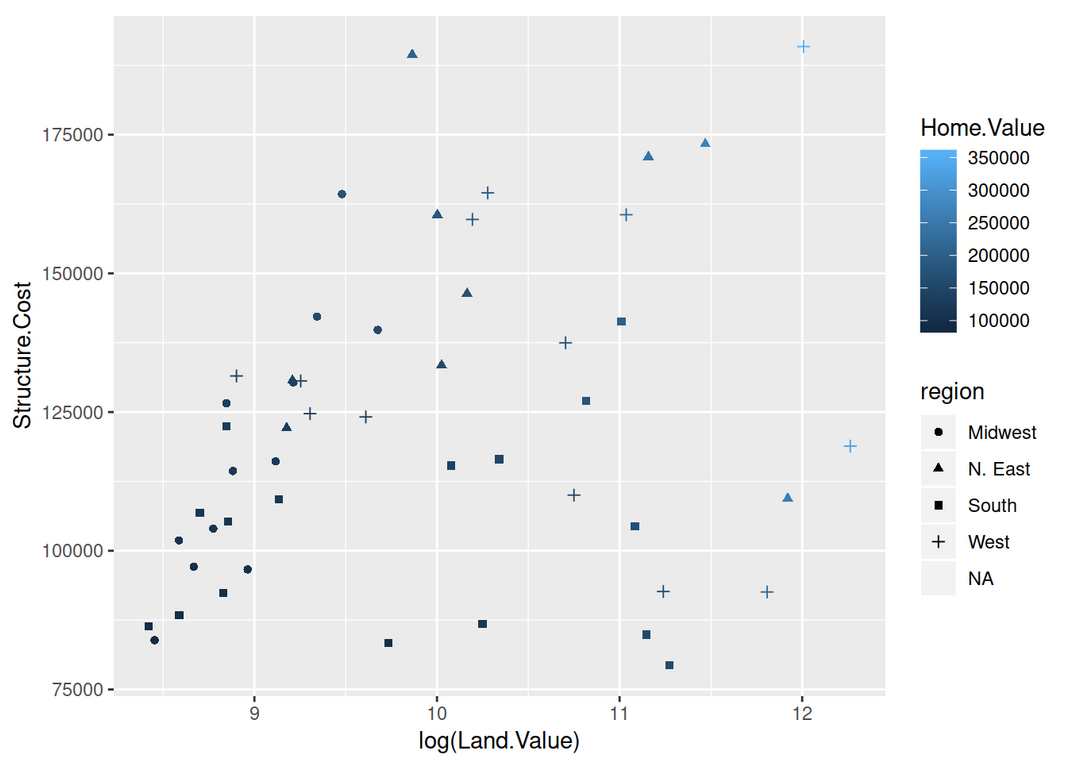


1. Labeling points (geom\_text)



**p1 + geom\_text(aes(label=State), size = 3)**

1. Aesthetics



**p1 + geom\_point(aes(color=Home.Value, shape = region))**

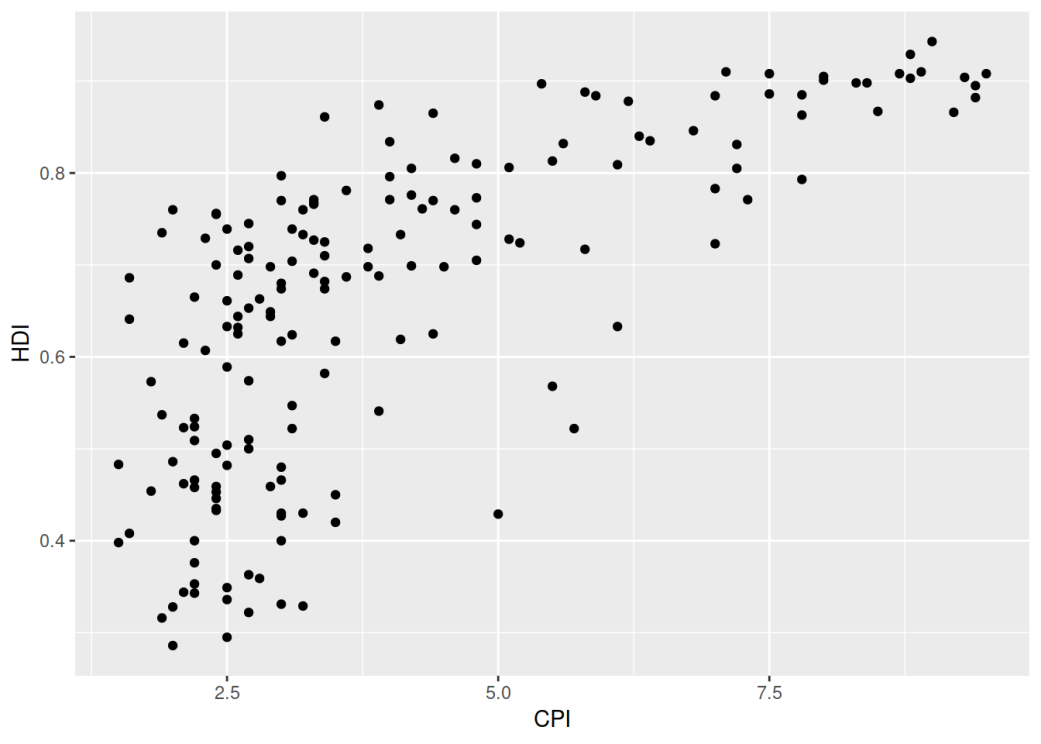
Exercise II

Solve the following problems with the following exercises :

dat <- read\_csv("dataSets/EconomistData.csv")

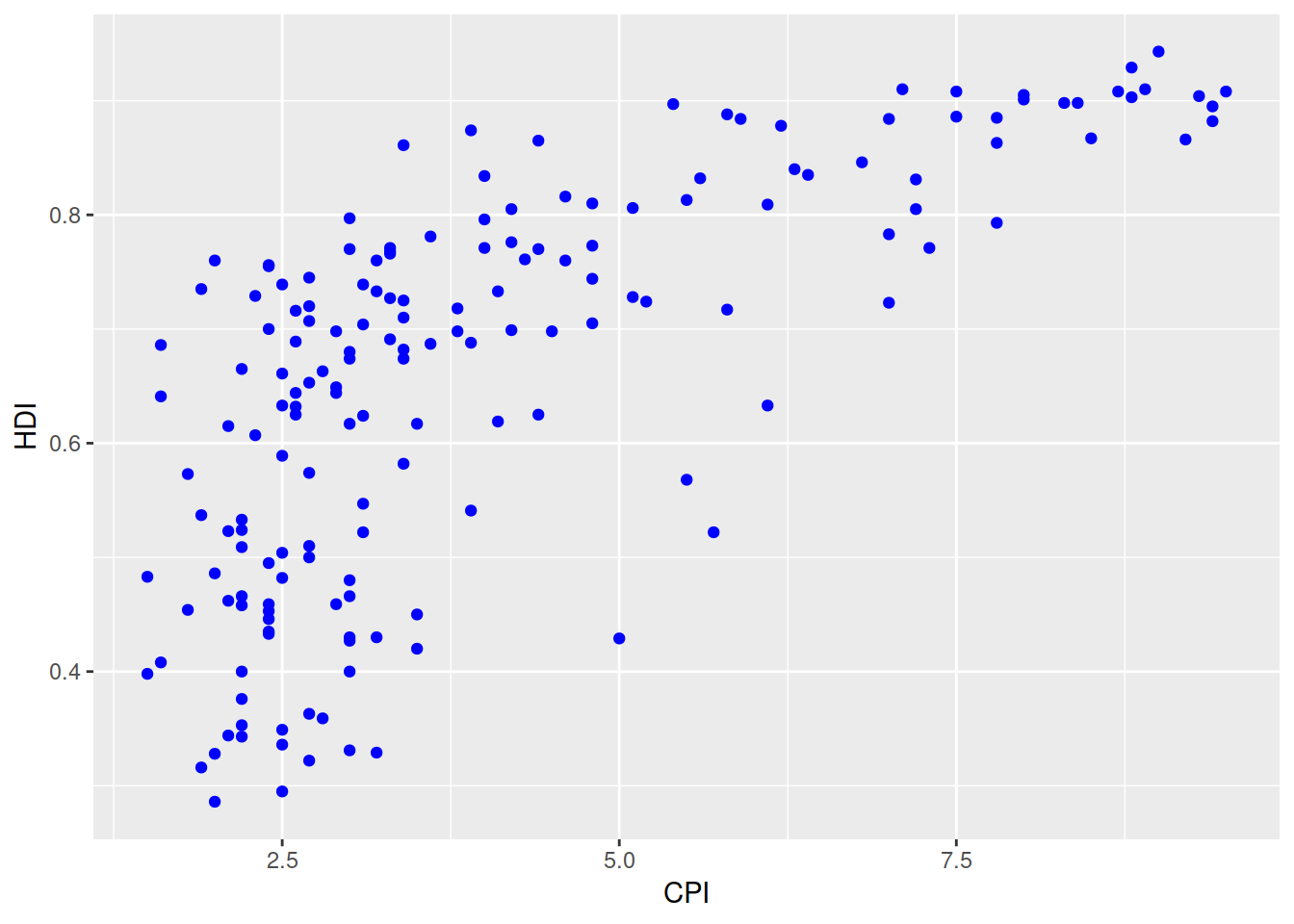
1. Create a scatter plot with CPI on the x axis and HDI on the y axis.

**ggplot(dat, aes(x = CPI, y = HDI)) + geom\_point()**



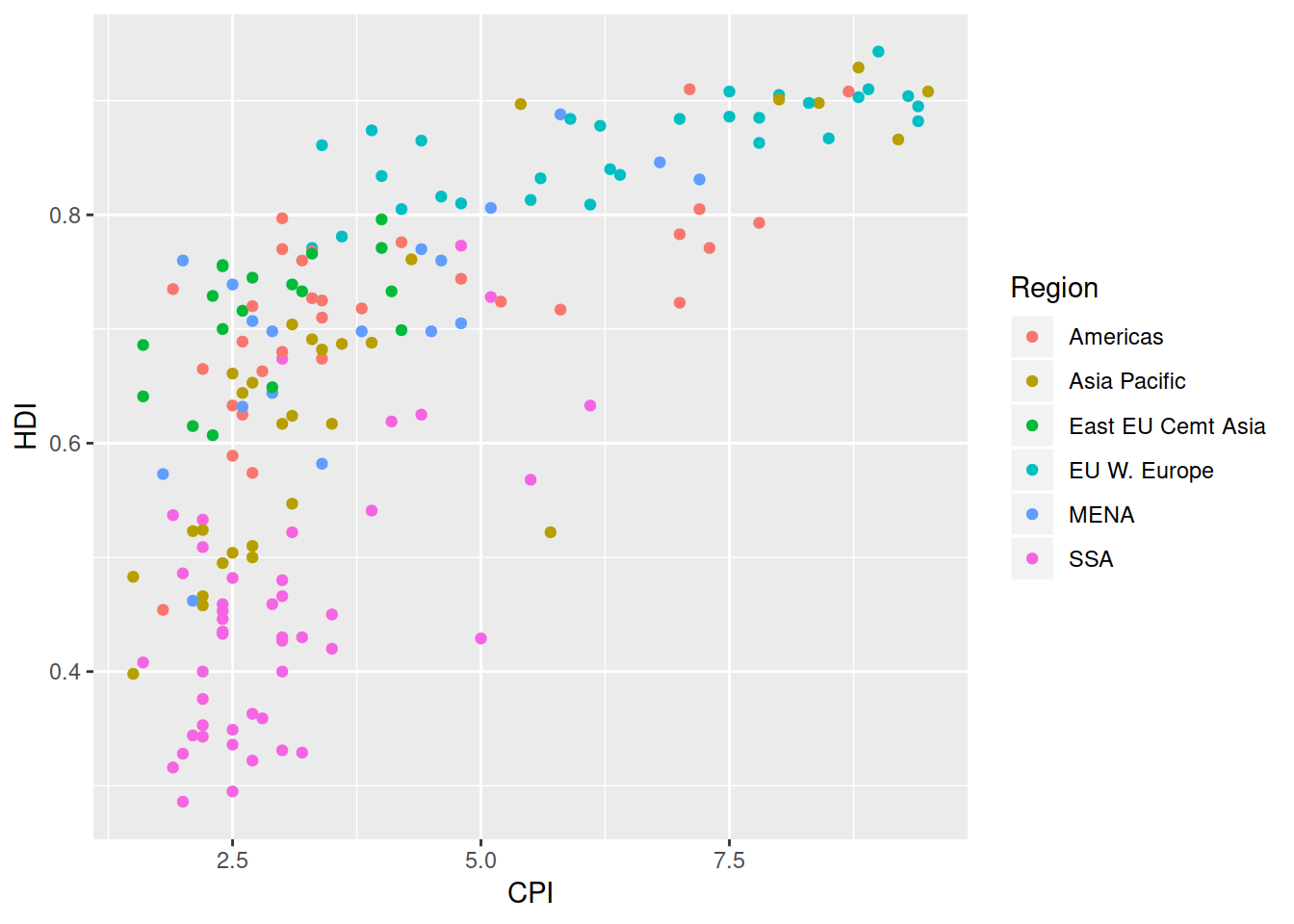
1. Color the points blue.

**ggplot(dat, aes(x = CPI, y = HDI)) + geom\_point(color = "blue")**



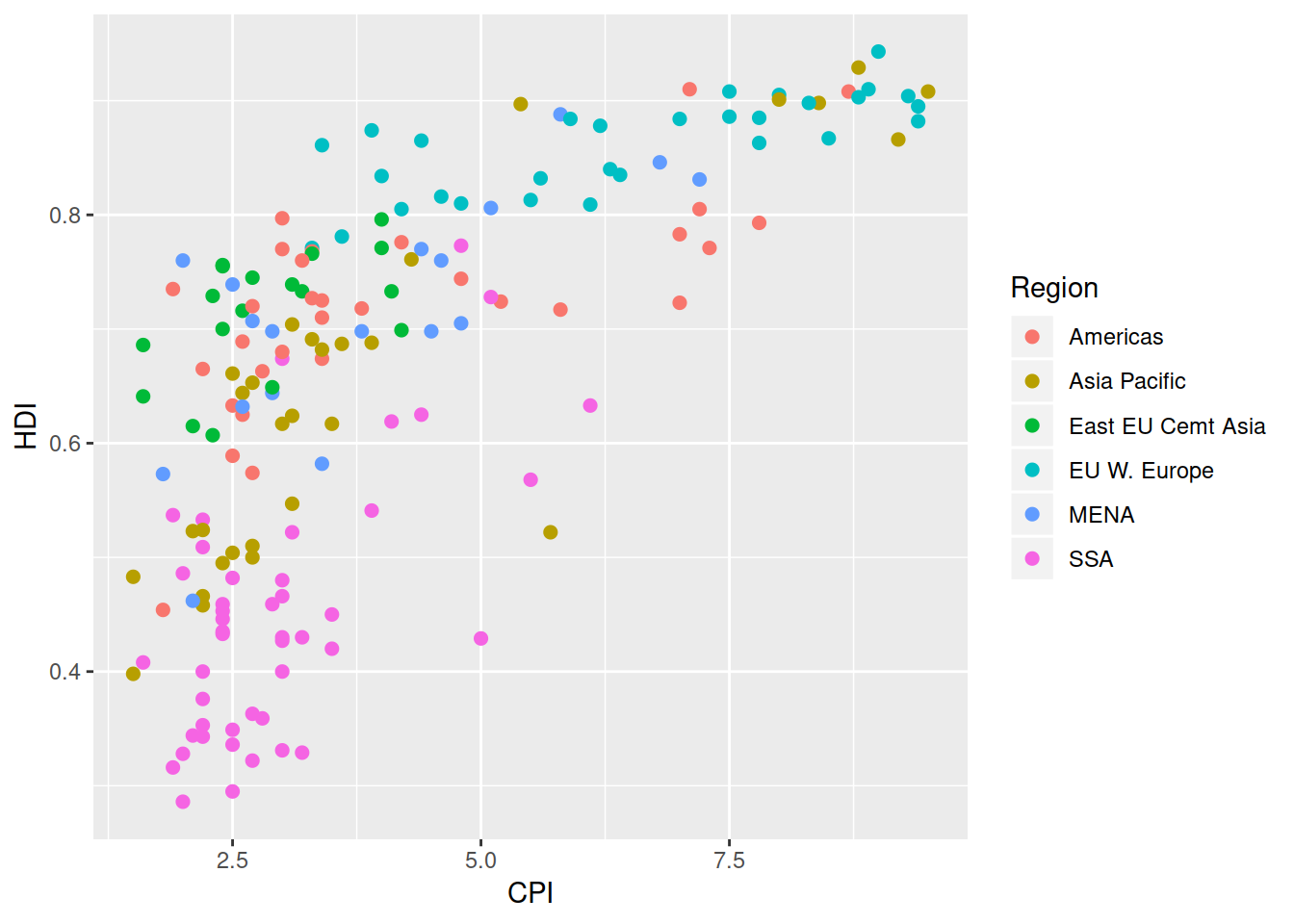
1. Map the color of the the points to Region.

**ggplot(dat, aes(x = CPI, y = HDI)) + geom\_point(aes(color = Region))**



1. Make the points bigger by setting size to 2

**ggplot(dat, aes(x = CPI, y = HDI)) + geom\_point(aes(color = Region), size = 2)**



1. Map the size of the points to HDI.Rank

**ggplot(dat, aes(x = CPI, y = HDI)) + geom\_point(aes(color = Region, size = HDI.Rank))**

