week8-lab-homework

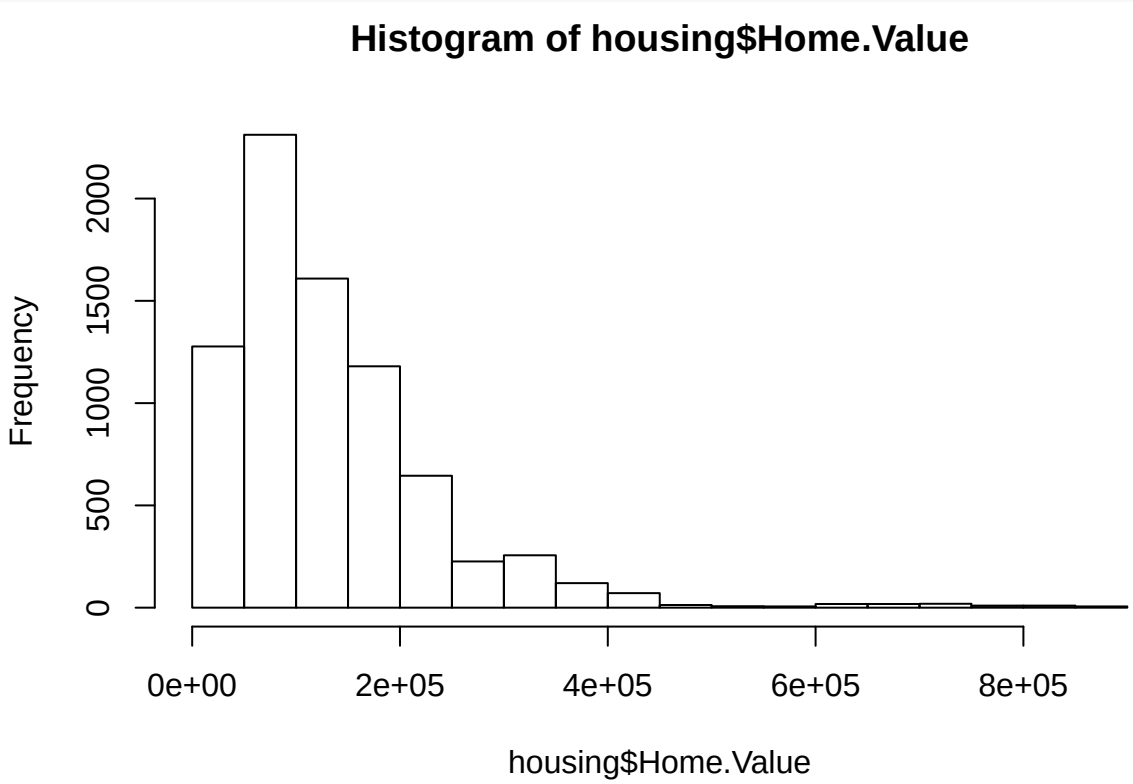
Haily Kil

March 1, 2017

```
housing <- read.csv("~/Desktop/eeb-177/Lab-Work/exercise-8/Rgraphics/dataSets/landdata-states.csv")
head(housing[1:5])
```

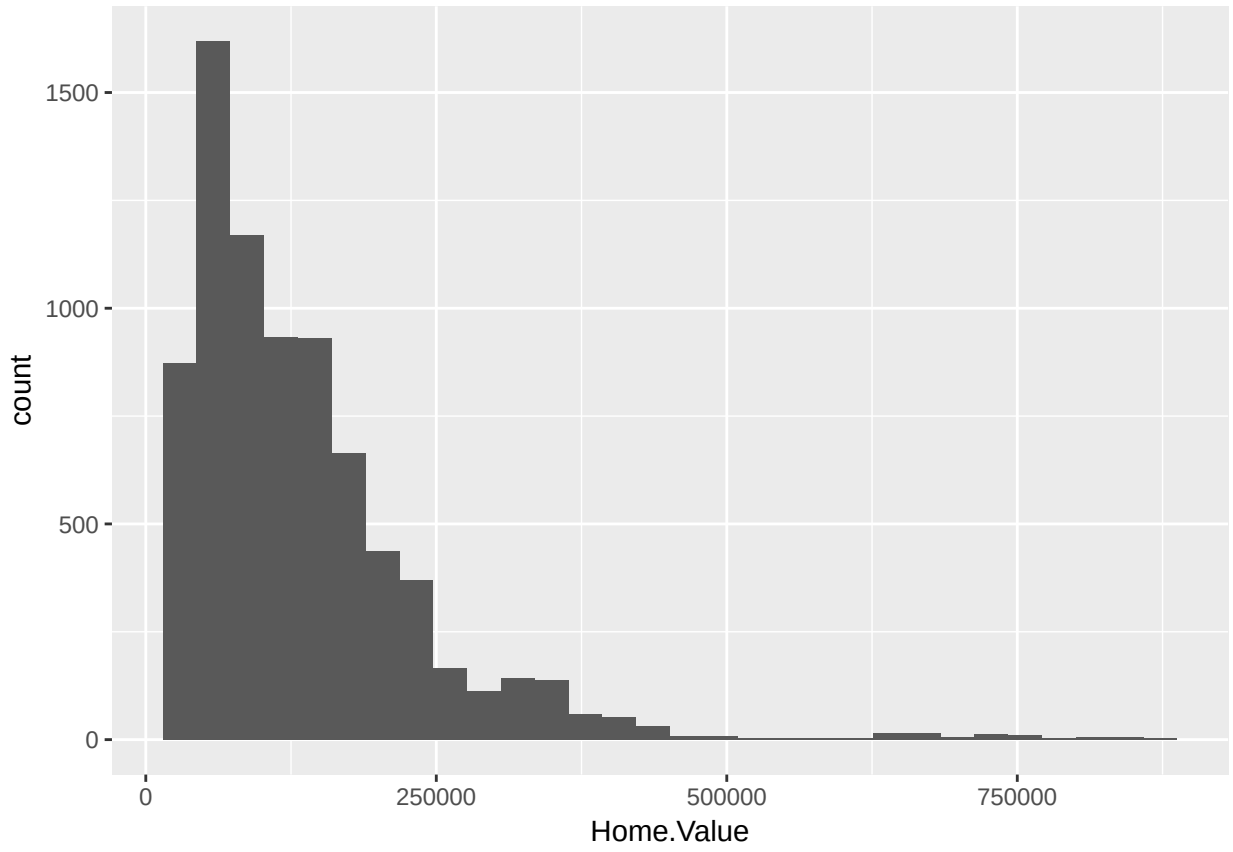
```
##   State region    Date Home.Value Structure.Cost
## 1    AK   West 2010.25     224952          160599
## 2    AK   West 2010.50     225511          160252
## 3    AK   West 2009.75     225820          163791
## 4    AK   West 2010.00     224994          161787
## 5    AK   West 2008.00     234590          155400
## 6    AK   West 2008.25     233714          157458
```

```
hist(housing$Home.Value)

library(ggplot2)
```

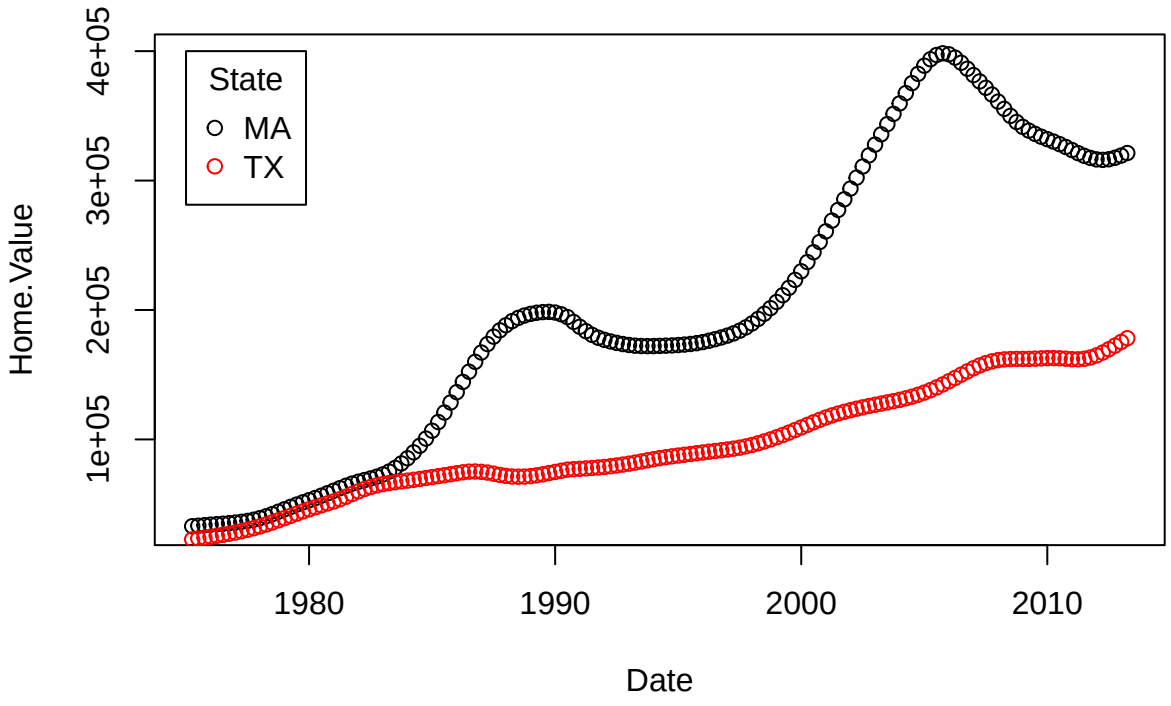



```
ggplot(housing, aes(x = Home.Value)) + geom_histogram()
```

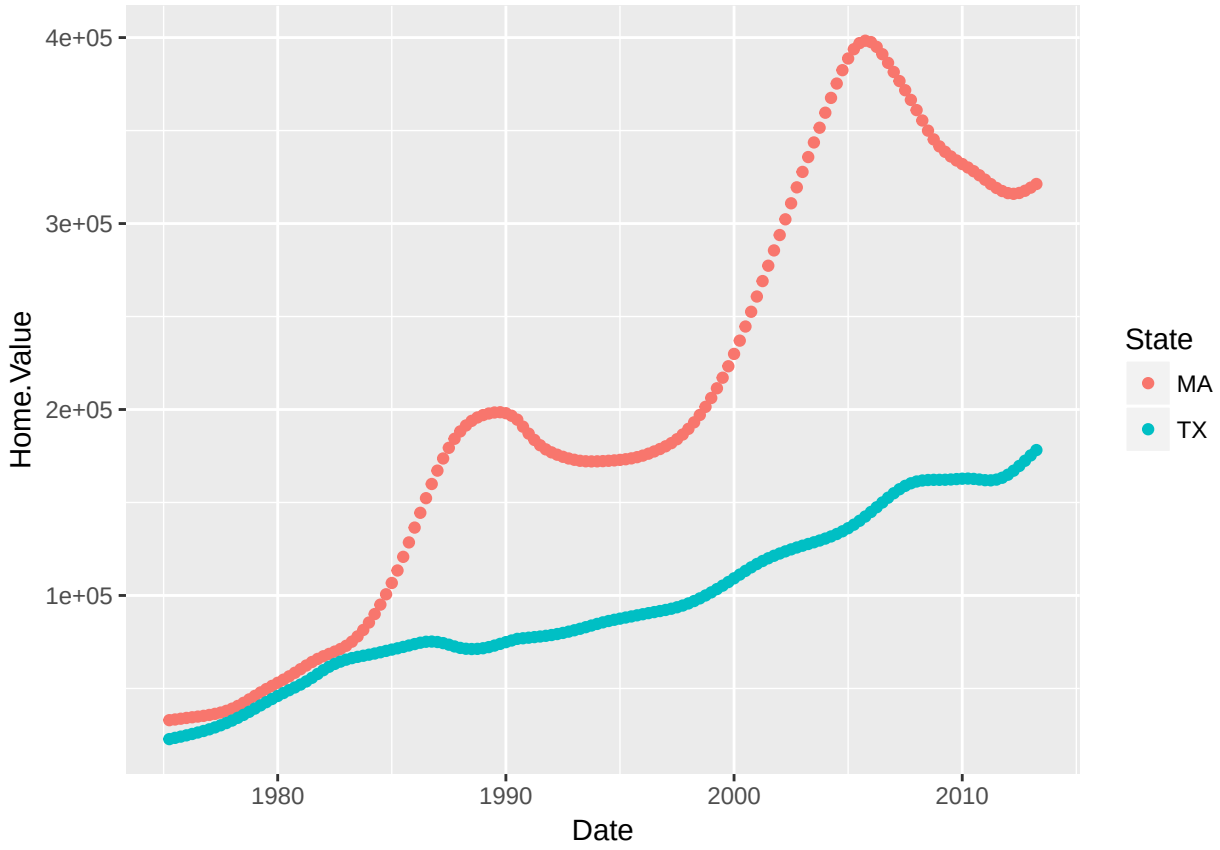
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



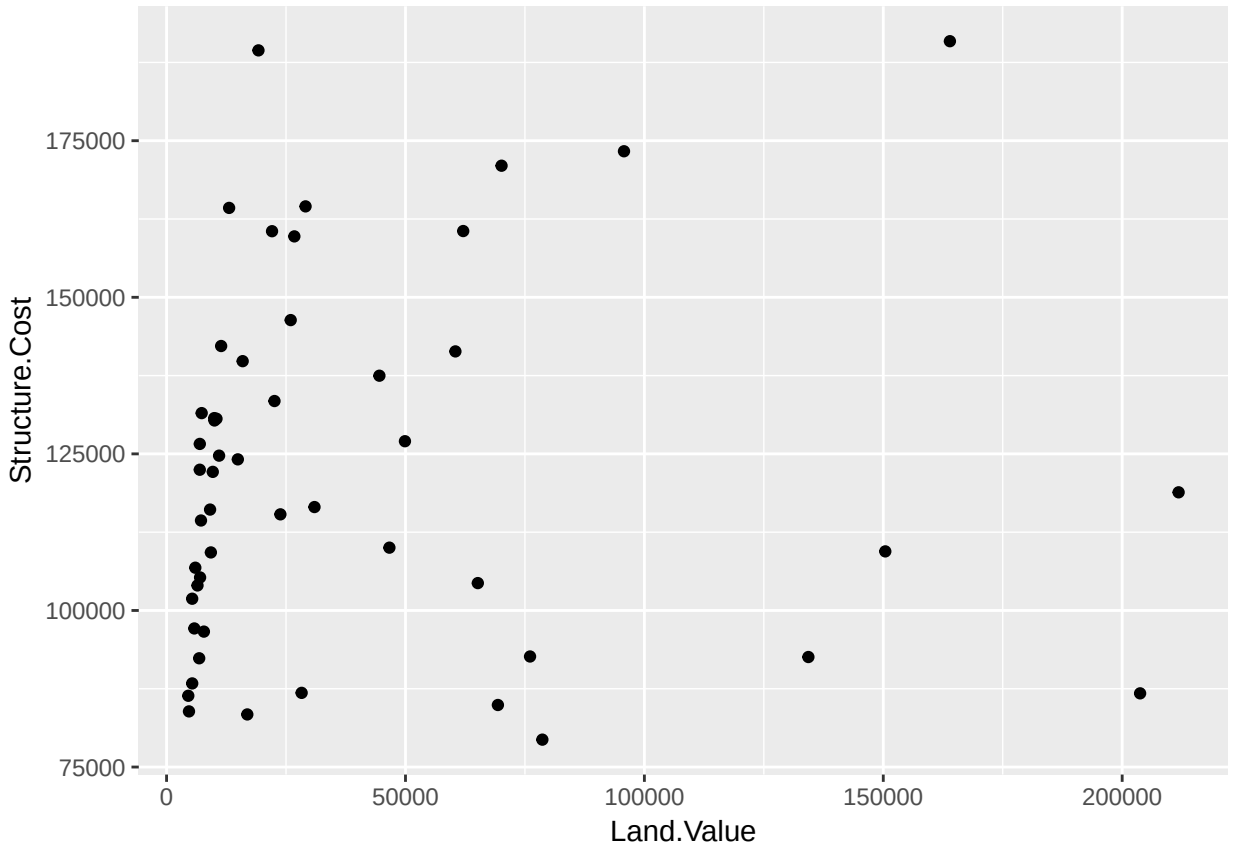

```
plot(Home.Value ~ Date, data=subset(housing, State == "MA"))
points(Home.Value ~ Date, col = "red", data=subset(housing, State == "TX"))
legend(1975, 400000, c("MA", "TX"), title="State", col=c("black", "red"), pch=c(1, 1))
```



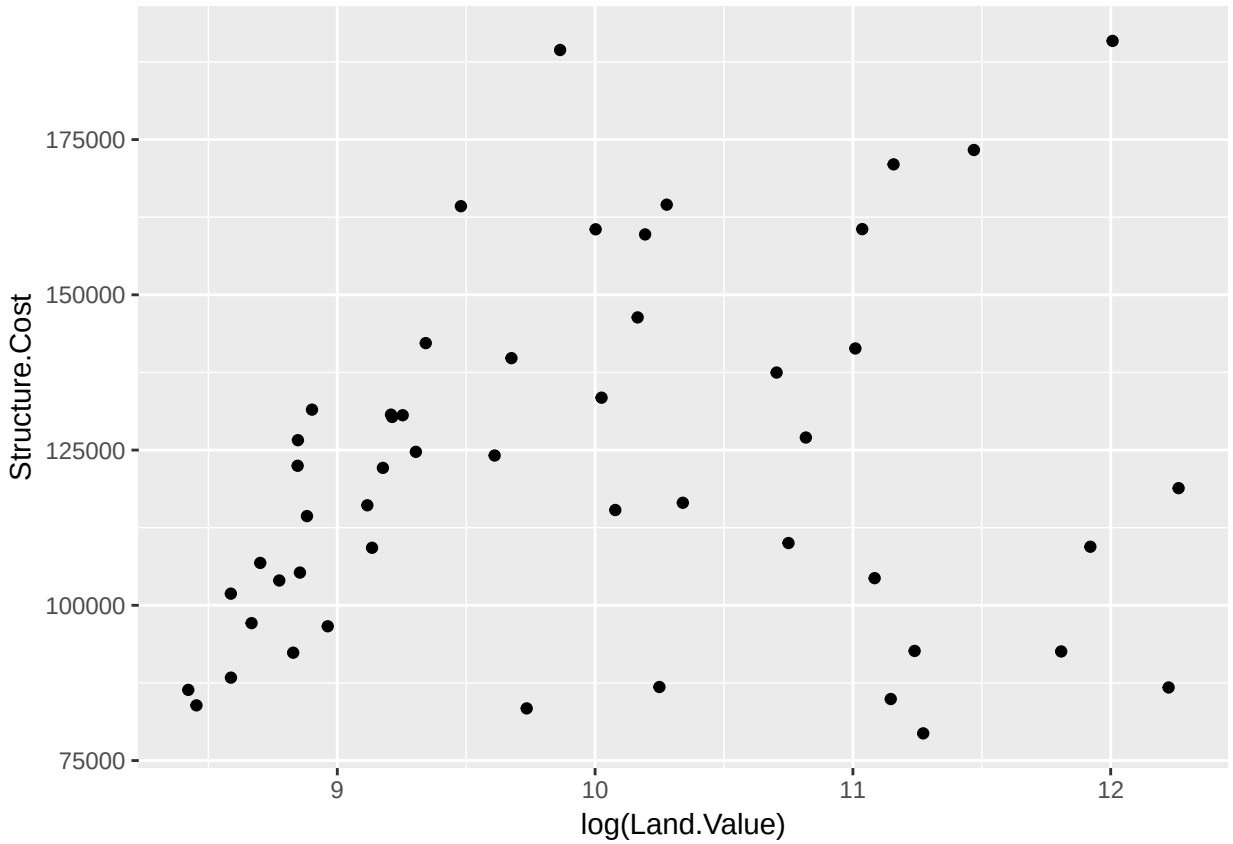
```
ggplot(subset(housing, State %in% c("MA", "TX")), aes(x=Date, y=Home.Value,color=State))+ geom_point()
```



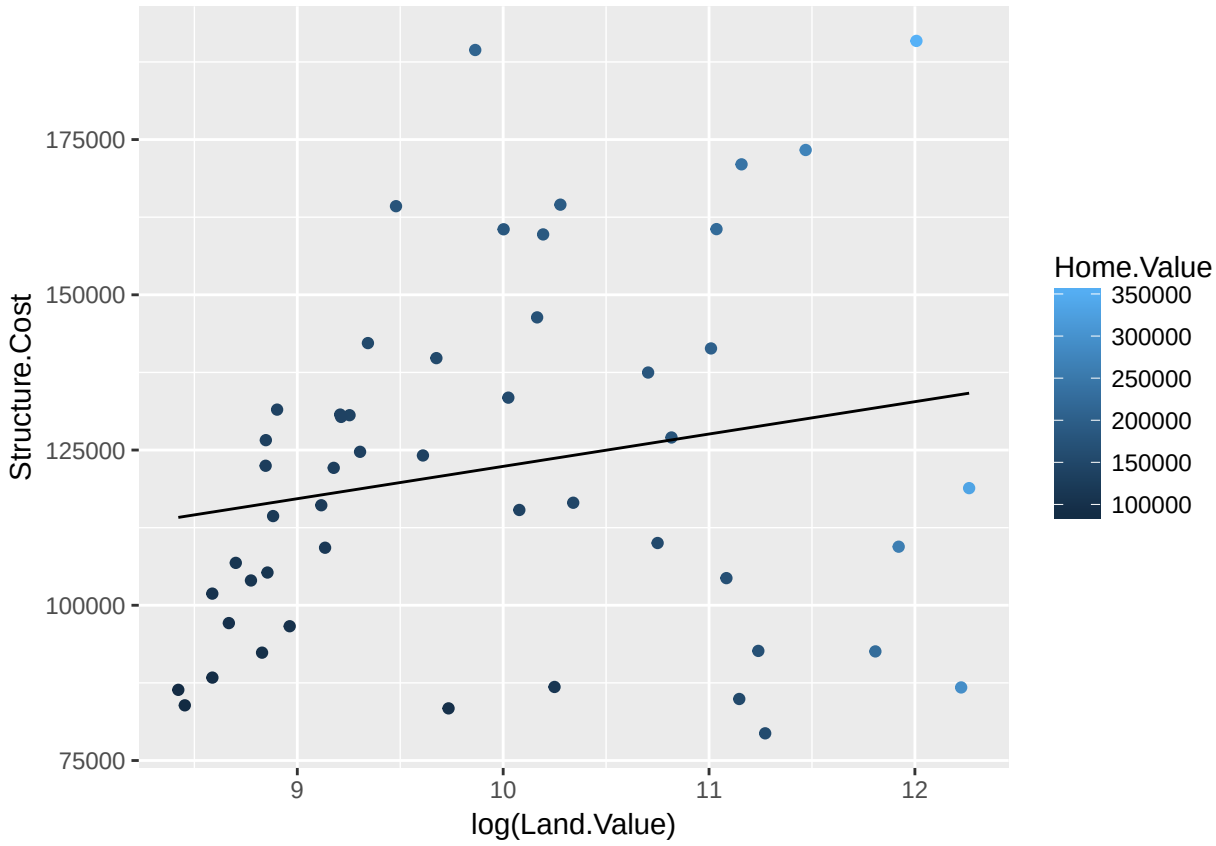

```
help.search("geom_", package = "ggplot2")

hp2001Q1 <- subset(housing, Date == 2001.25)
ggplot(hp2001Q1, aes(y = Structure.Cost, x = Land.Value)) + geom_point()
```




```
ggplot(hp2001Q1, aes(y = Structure.Cost, x = log(Land.Value))) + geom_point()
```

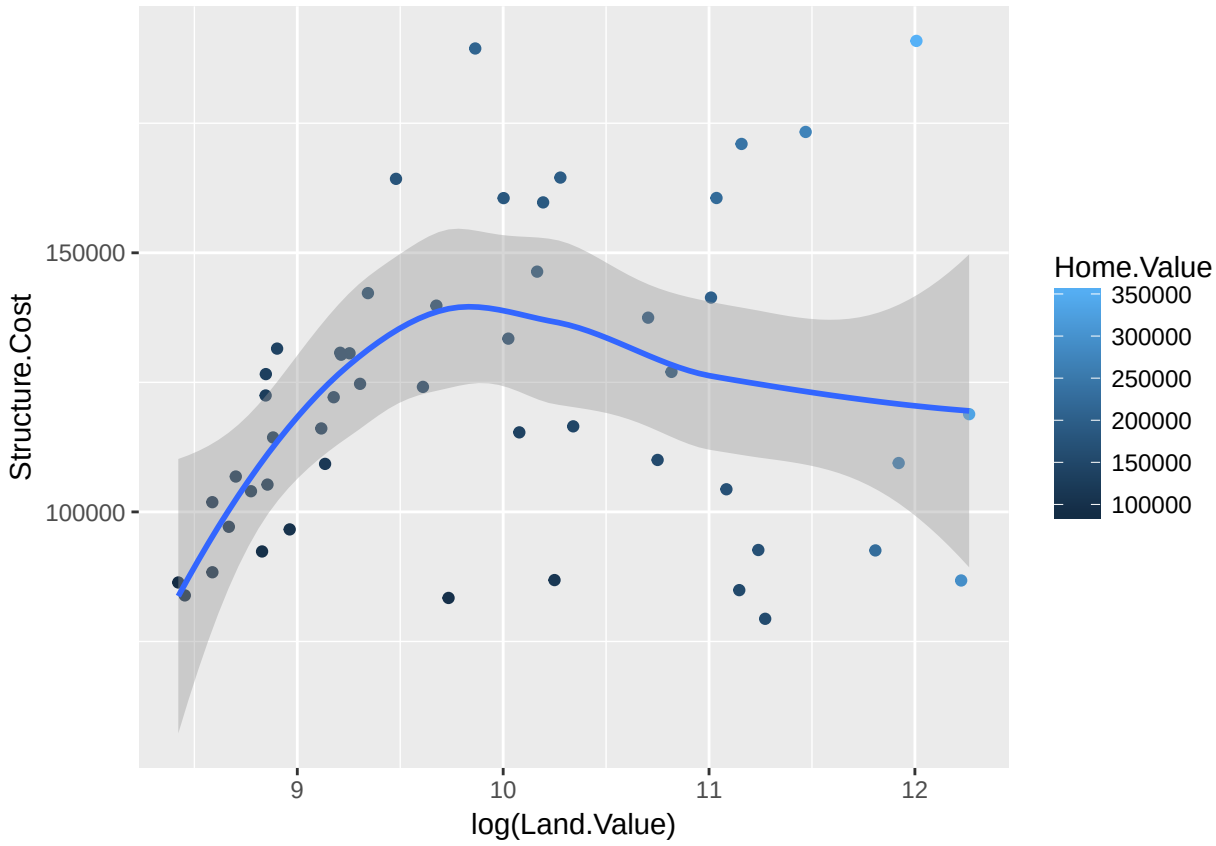



```
hp2001Q1$pred.SC <- predict(lm(Structure.Cost ~ log(Land.Value), data = hp2001Q1))

p1 <- ggplot(hp2001Q1, aes(x = log(Land.Value), y = Structure.Cost))

p1 + geom_point(aes(color = Home.Value)) +
  geom_line(aes(y = pred.SC))
```

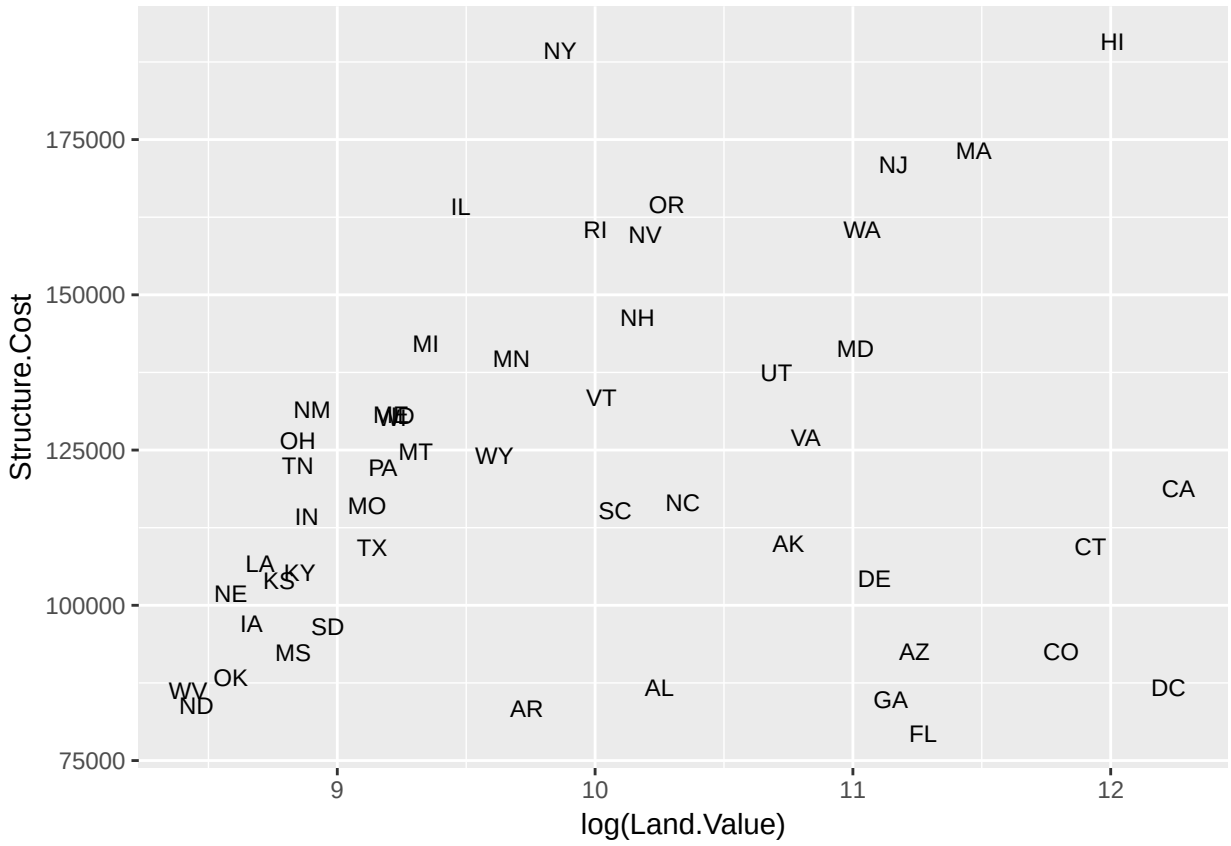



```
p1 + geom_point(aes(color = Home.Value)) + geom_smooth()
```

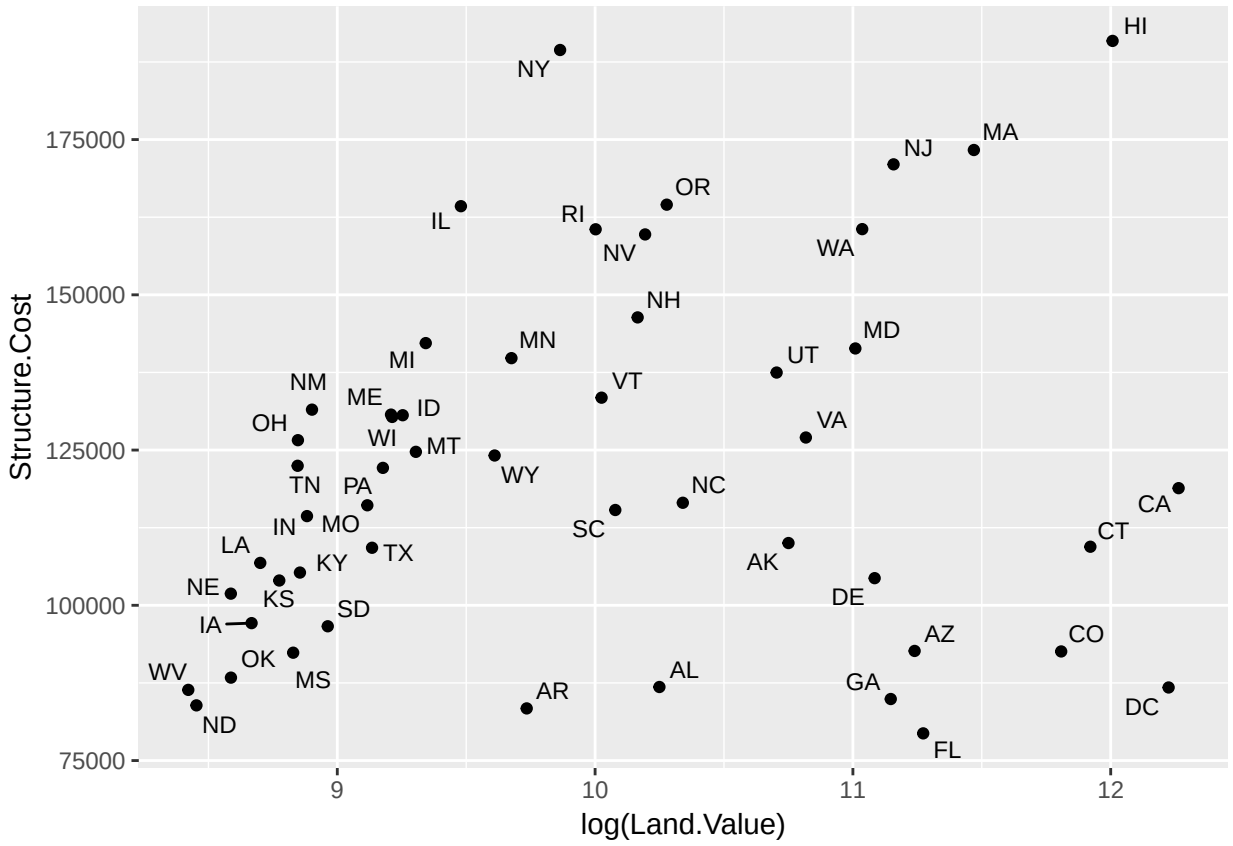
```
## `geom_smooth()` using method = 'loess'
```



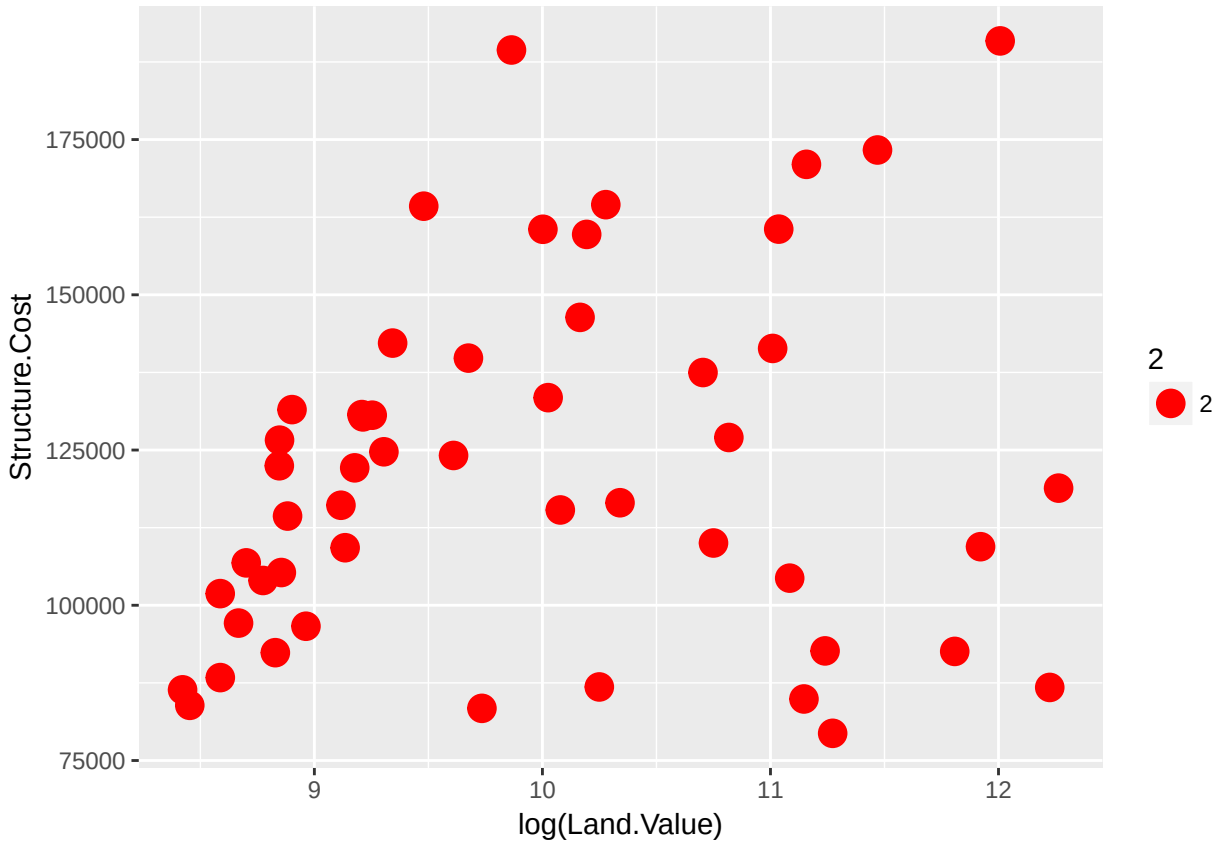

```
p1 + geom_text(aes(label=State), size = 3)
```




```
library("ggrepel")
p1 +geom_point() + geom_text_repel(aes(label=State), size = 3)
```

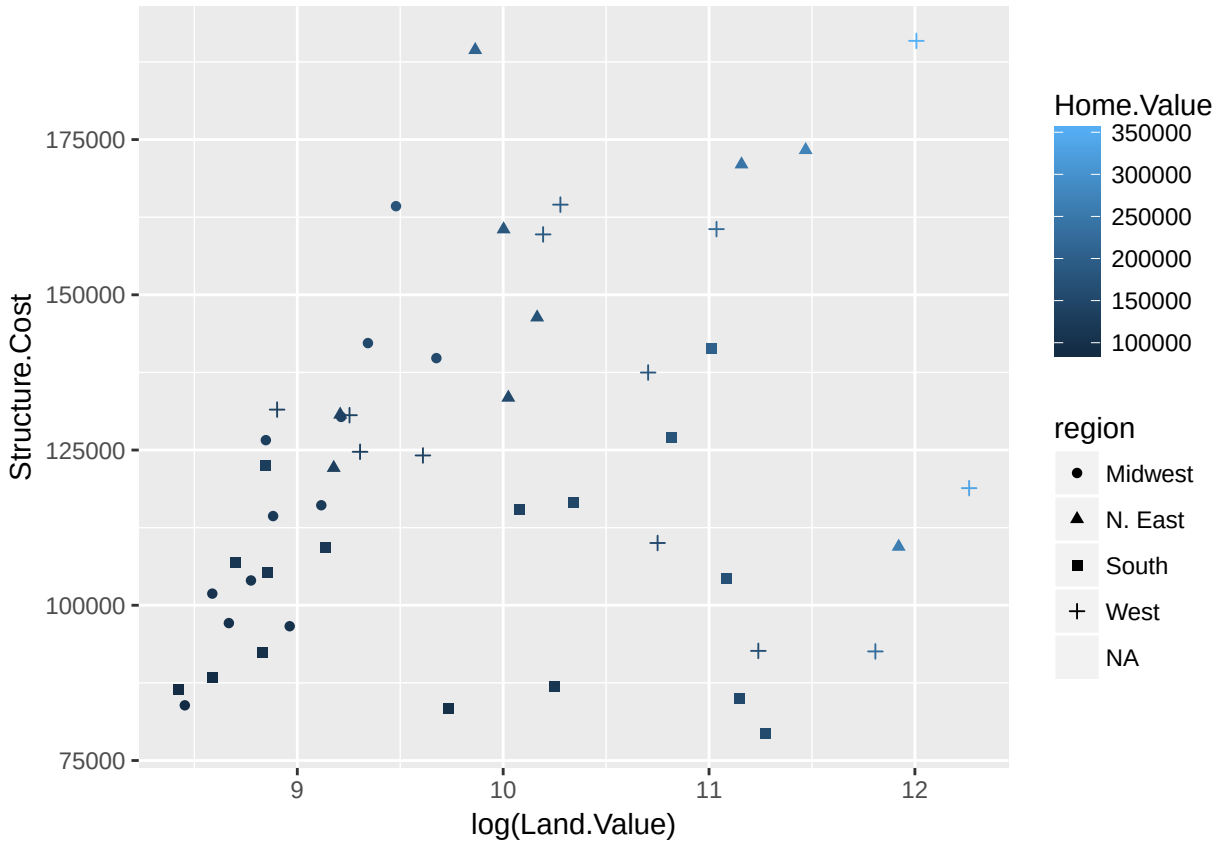



```
p1 + geom_point(aes(size = 2), color="red")
```




```
p1 + geom_point(aes(color=Home.Value, shape = region))
```

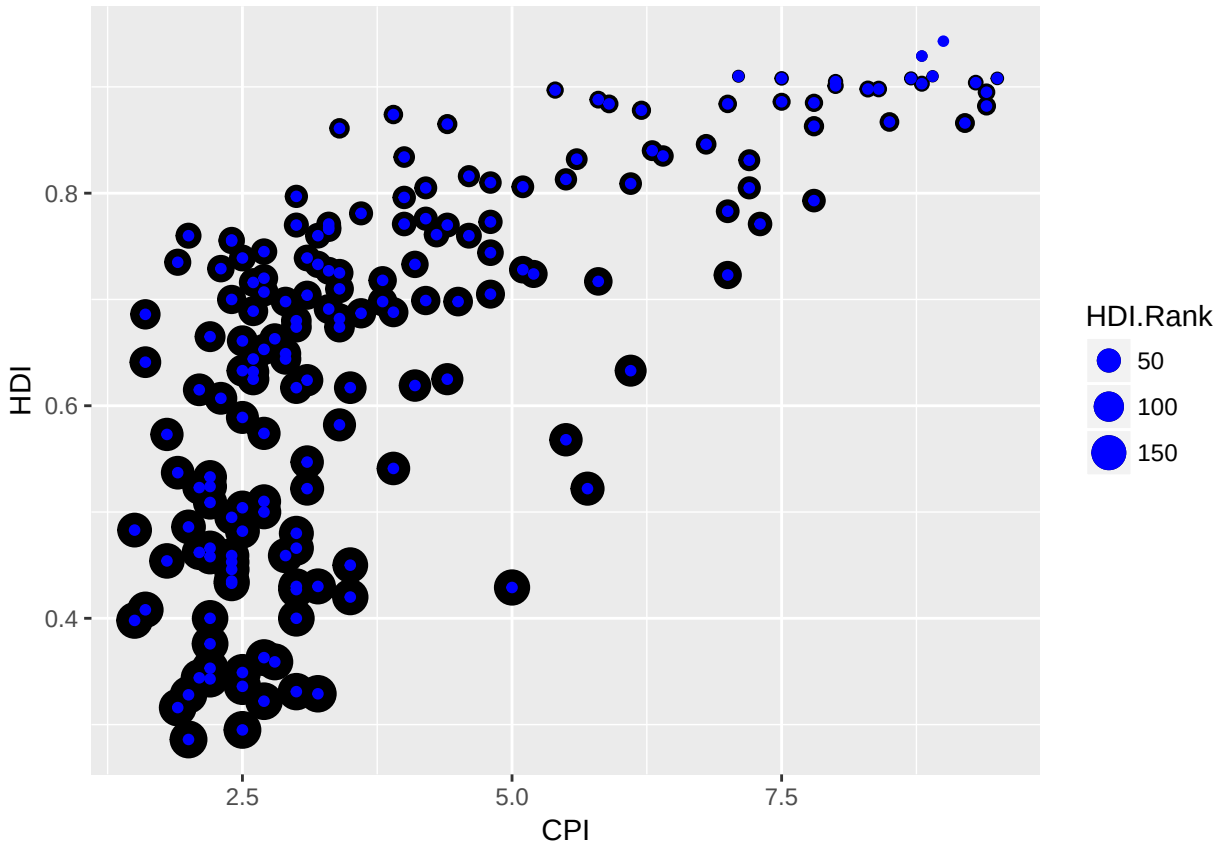
```
## Warning: Removed 1 rows containing missing values (geom_point).
```



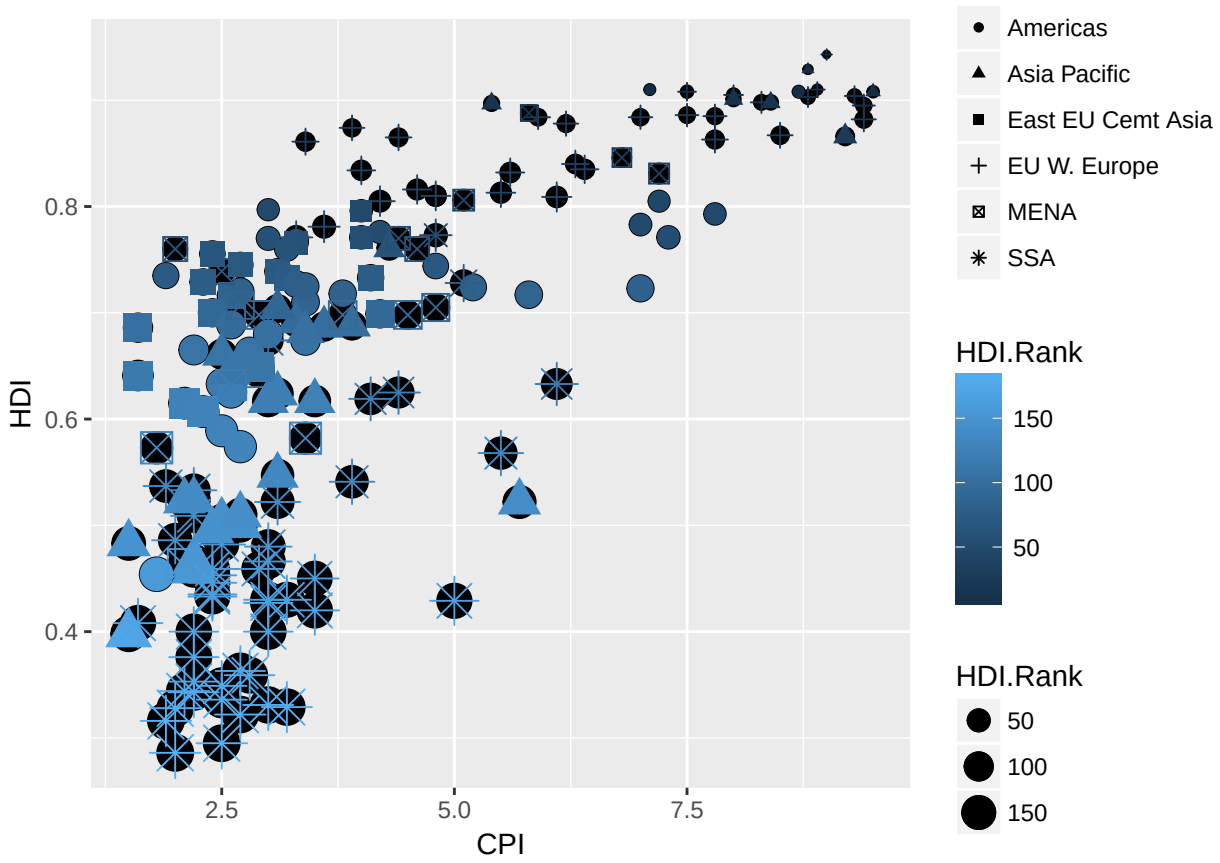

```
dat <- read.csv("~/Desktop/eeb-177/Lab-Work/exercise-8/Rgraphics/dataSets/EconomistData.csv")
head(dat)
```

```
##   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
```

```
p2 <- ggplot(dat, aes(x = CPI, y = HDI, size = HDI.Rank)) + geom_point()

p2 +  geom_point(aes(size = 2), color="blue")
```

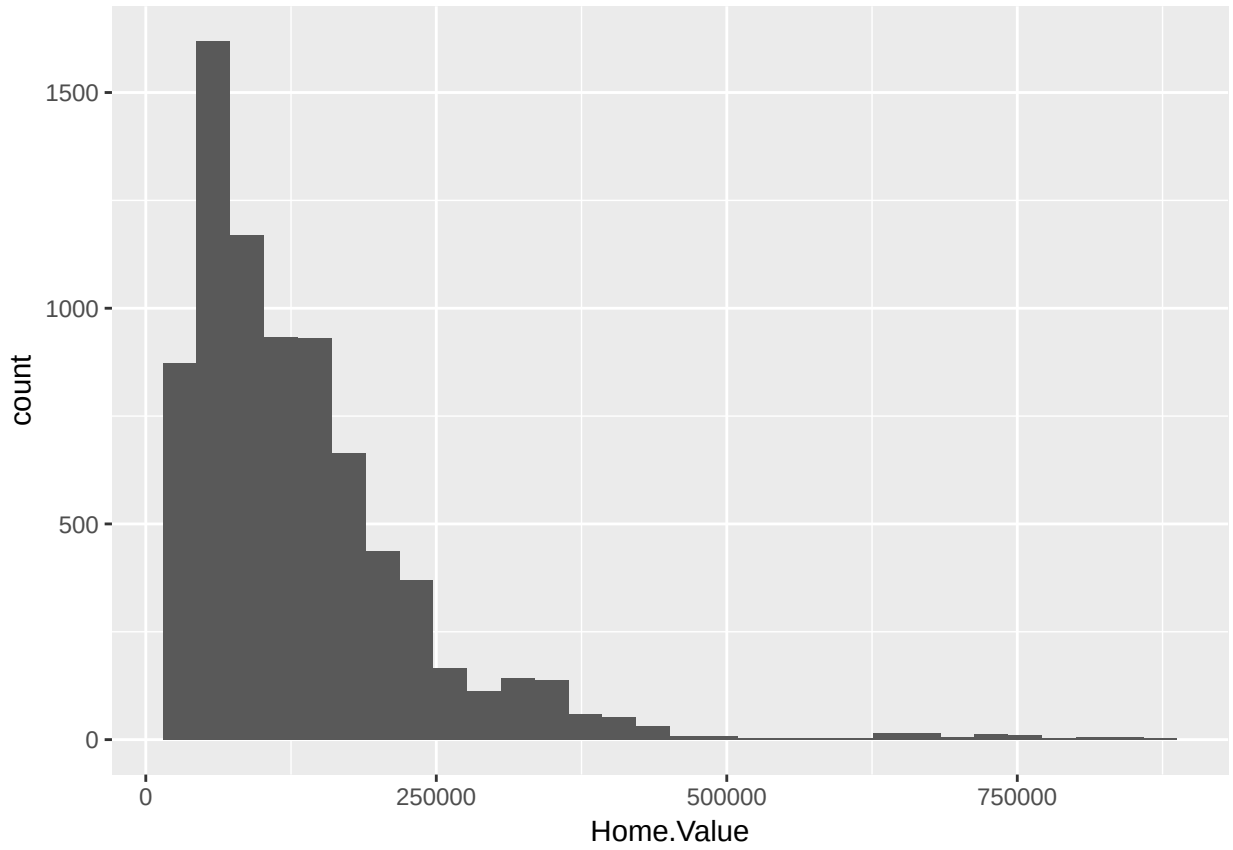



```
p2 + geom_point(aes(shape = Region, color=HDI.Rank))
```

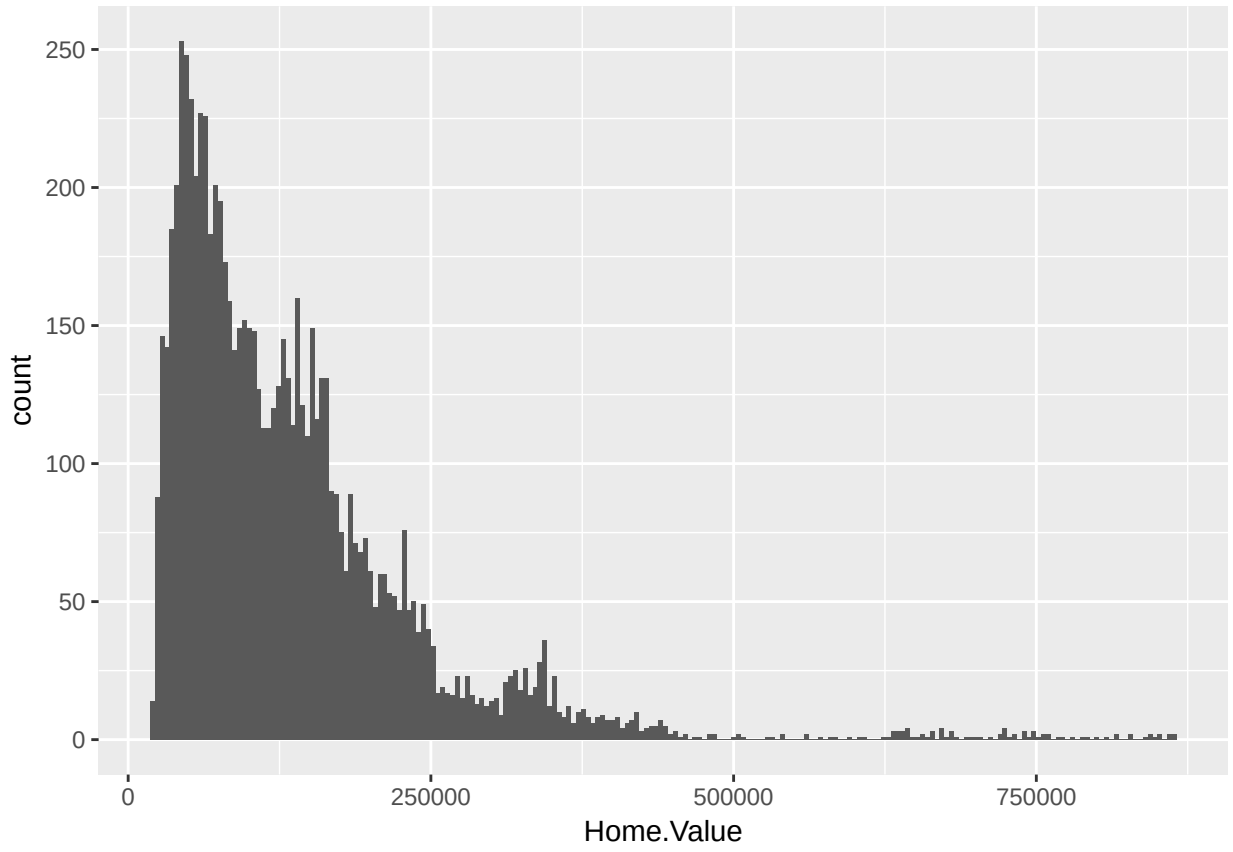



```
housing <- read.csv("~/Desktop/eeb-177/Lab-Work/exercise-8/Rgraphics/dataSets/landdata-states.csv")
p2 <- ggplot(housing, aes(x = Home.Value))
p2 + geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



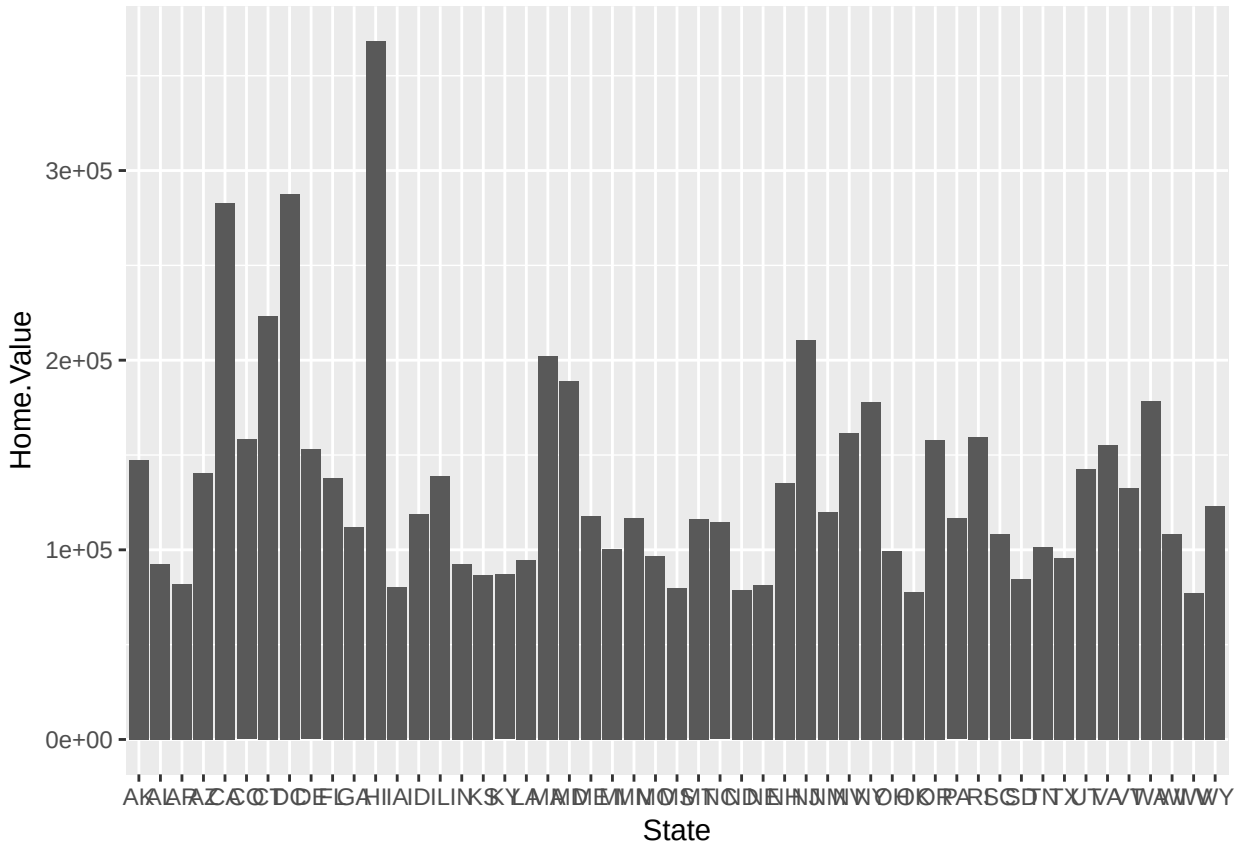

```
p2 + geom_histogram(stat = "bin", binwidth=4000)
```



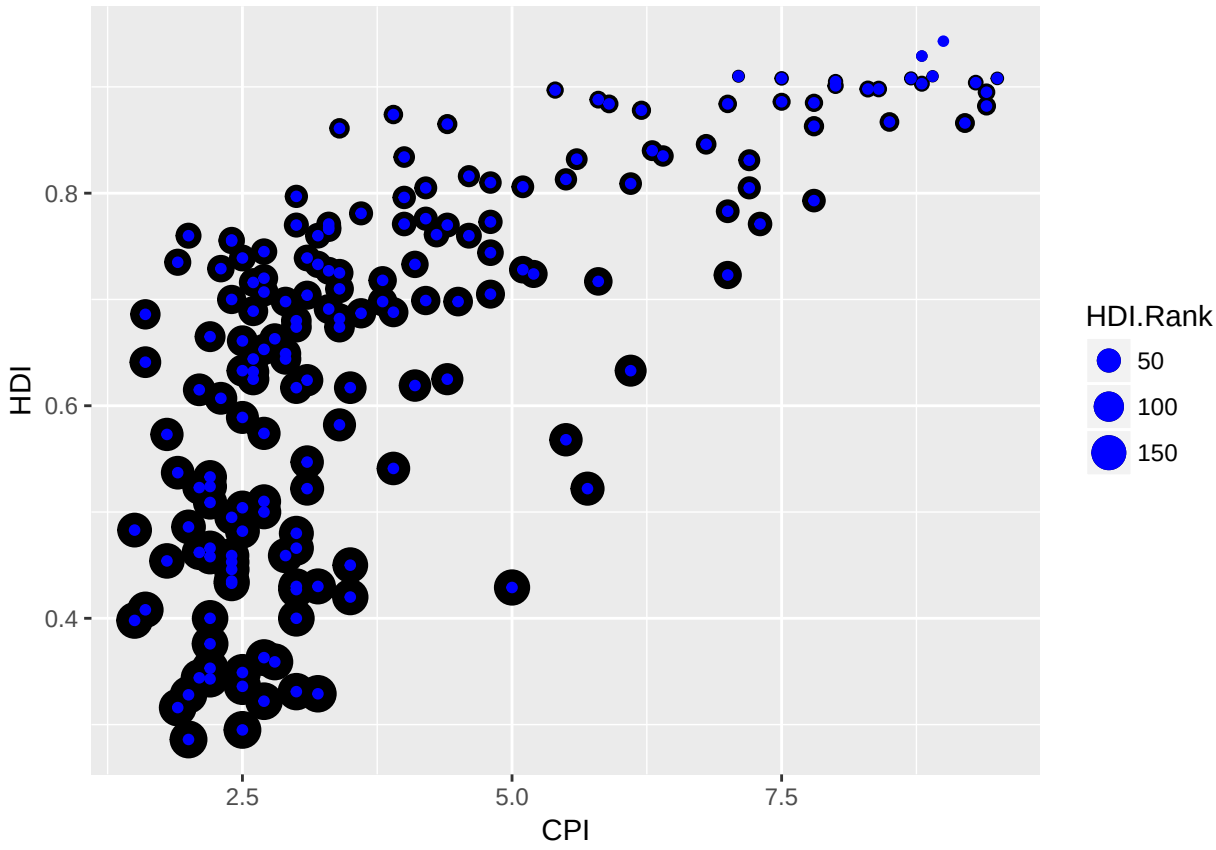

```
housing.sum <- aggregate(housing["Home.Value"], housing["State"], FUN=mean)
rbind(head(housing.sum), tail(housing.sum))
```

```
##    State Home.Value
## 1     AK  147385.14
## 2     AL   92545.22
## 3     AR   82076.84
## 4     AZ  140755.59
## 5     CA  282808.08
## 6     CO  158175.99
## 46    VA  155391.44
## 47    VT  132394.60
## 48    WA  178522.58
## 49    WI  108359.45
## 50    WV   77161.71
## 51    WY  122897.25
```

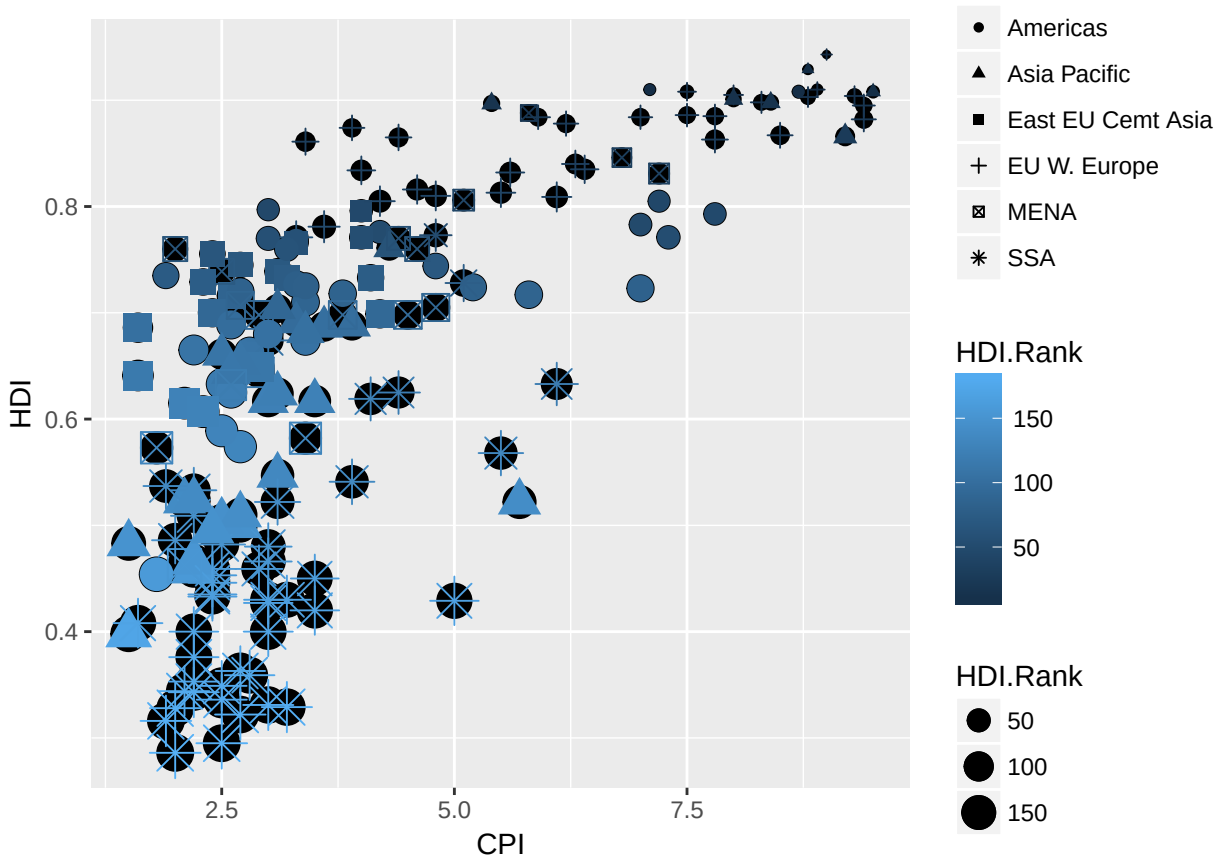
```
ggplot(housing.sum, aes(x=State, y=Home.Value)) +
  geom_bar(stat="identity")
```




```
dat <- read.csv("~/Desktop/eeb-177/Lab-Work/exercise-8/Rgraphics/dataSets/EconomistData.csv")

p2 <- ggplot(dat, aes(x = CPI, y = HDI, size = HDI.Rank)) + geom_point()

p2 +  geom_point(aes(size = 2), color="blue")
```

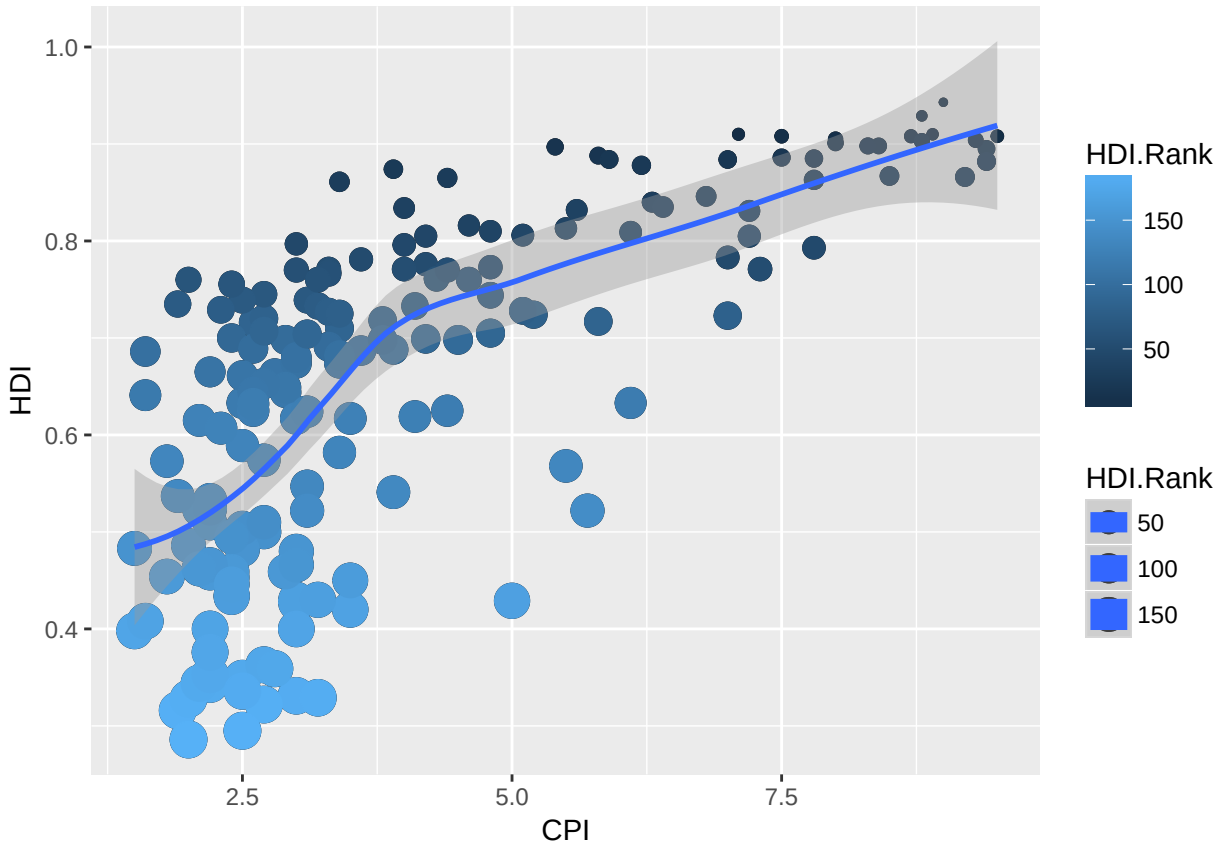



```
p2 + geom_point(aes(shape = Region, color=HDI.Rank))
```



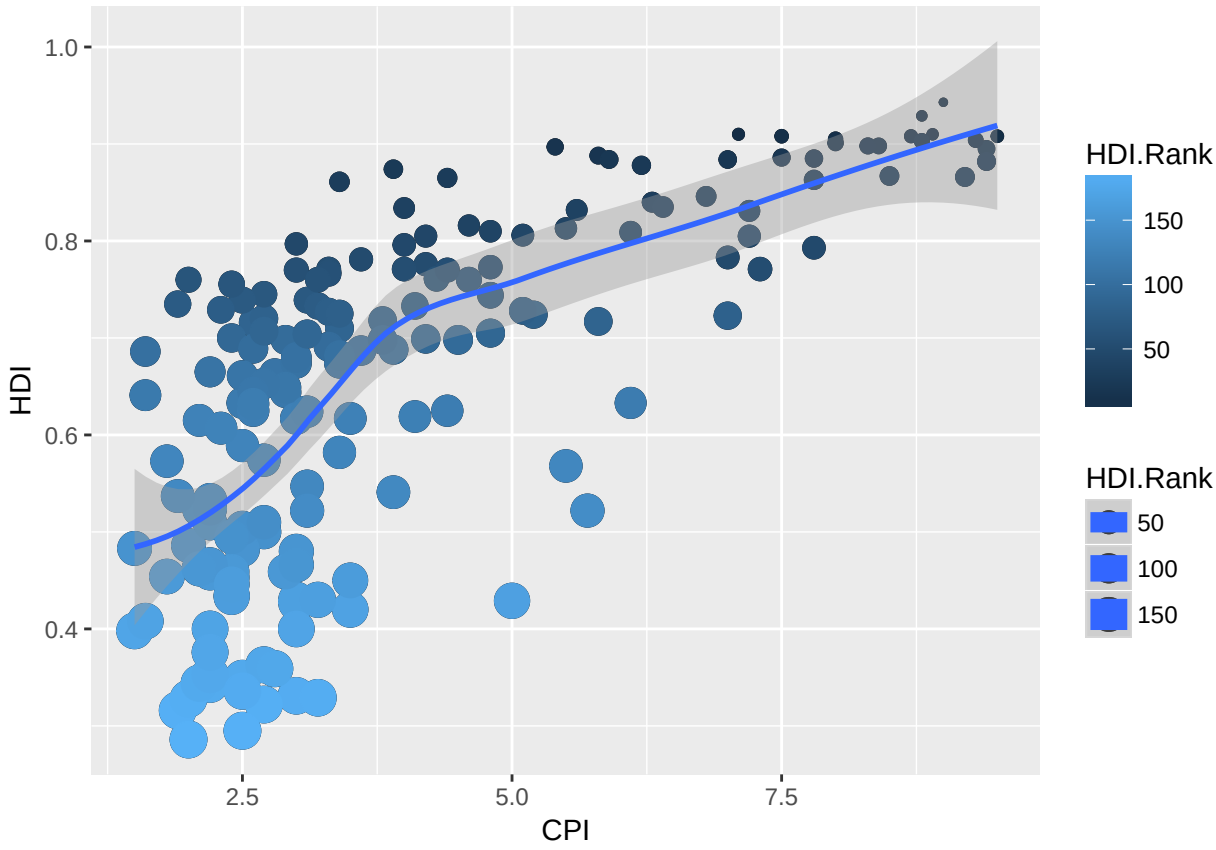

```
p2 + geom_point(aes(color = HDI.Rank)) + geom_smooth()
```

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

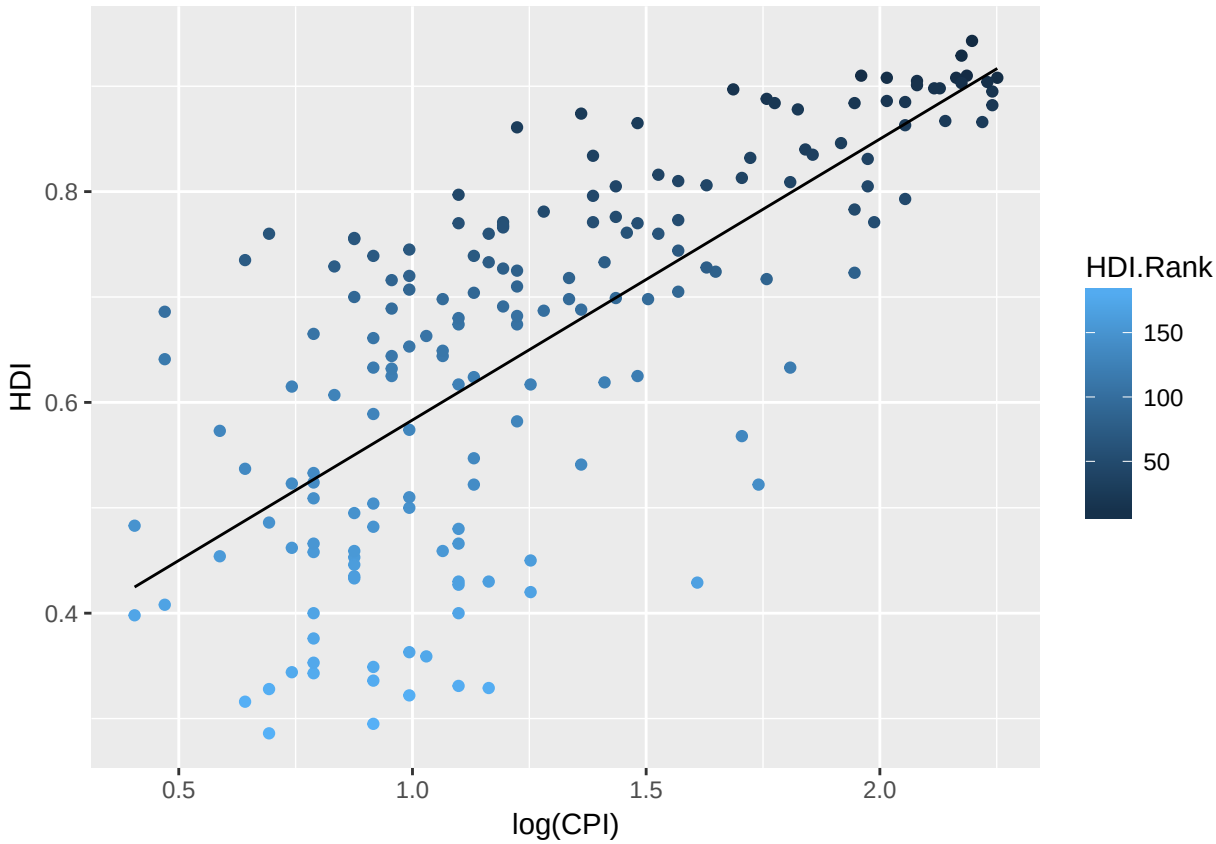



```
p2 + geom_point(aes(color = HDI.Rank)) + stat_smooth()
```

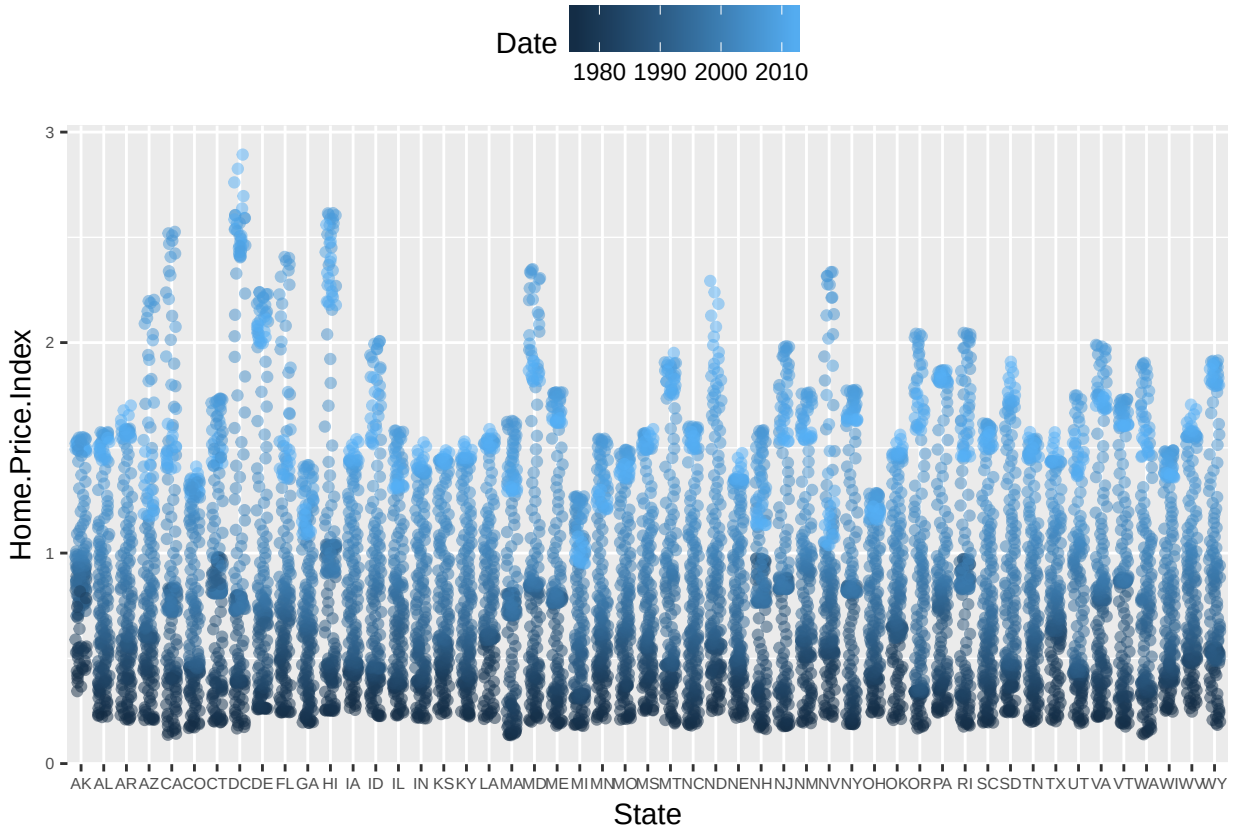
```
## `geom_smooth()` using method = 'loess'
```



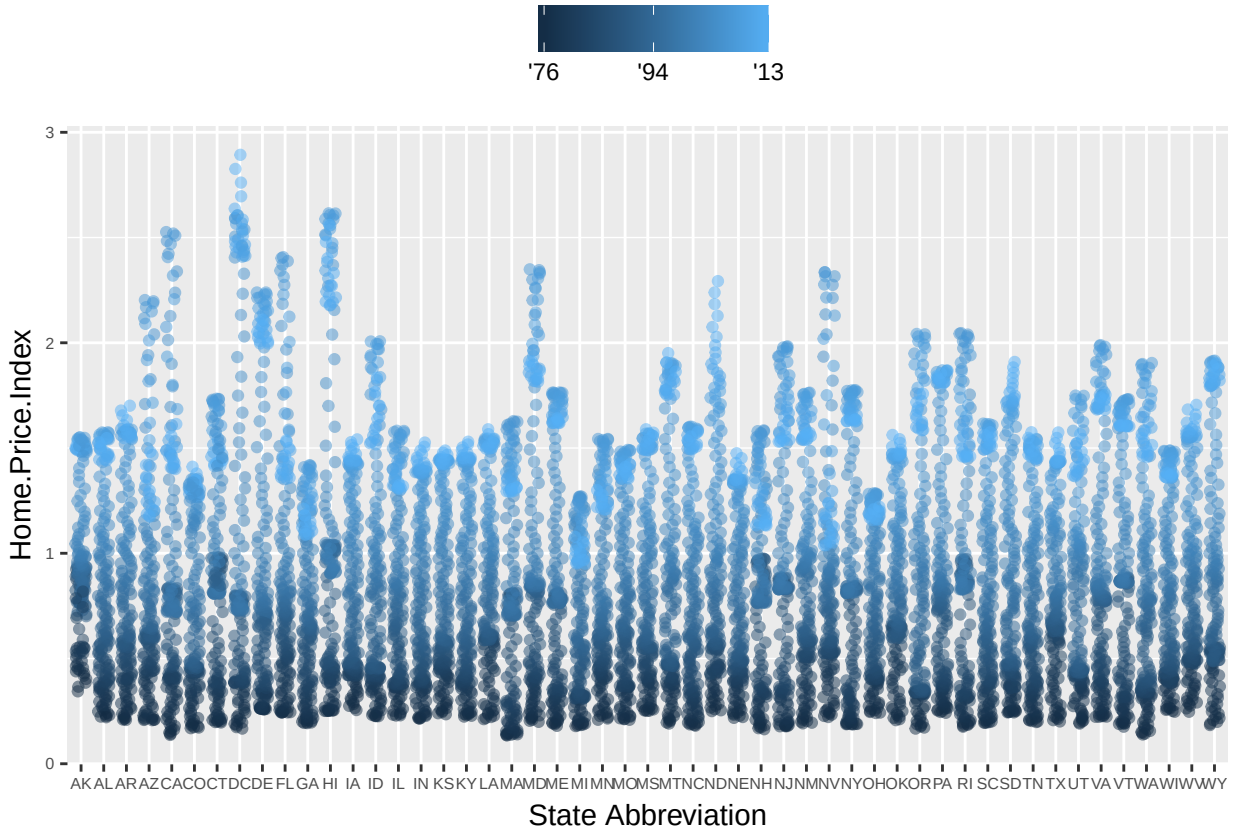

```
dat$pred.SC <- predict(lm(HDI ~ log(CPI), data = dat))
p2 <- ggplot(dat, aes(x = log(CPI), y = HDI))
p2 + geom_point(aes(color = HDI.Rank)) + geom_line(aes(y = pred.SC))
```



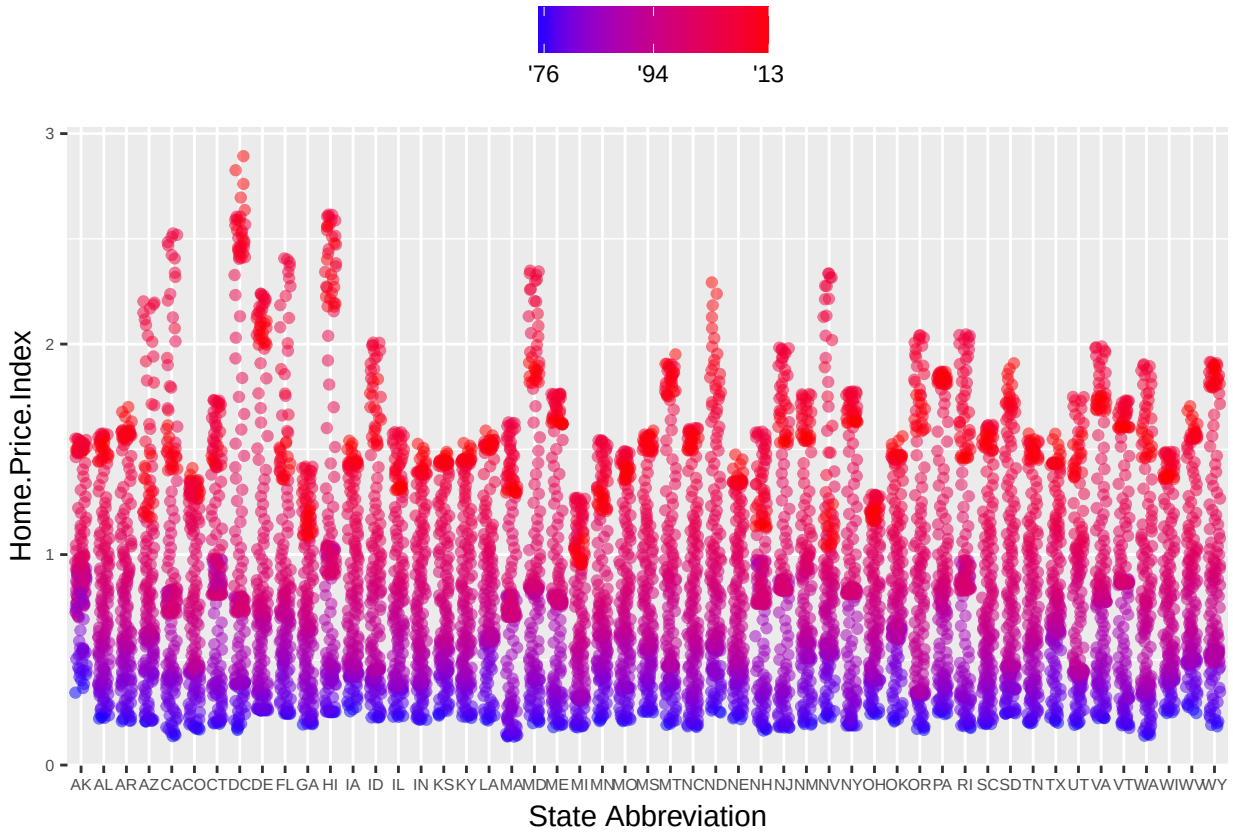

```
p3 <- ggplot(housing,
             aes(x = State, y = Home.Price.Index)) + theme(legend.position="top", axis.text=element_tex
(p4 <- p3 + geom_point(aes(color = Date), alpha = 0.5, size = 1.5, position = position_jitter(width = 0
```



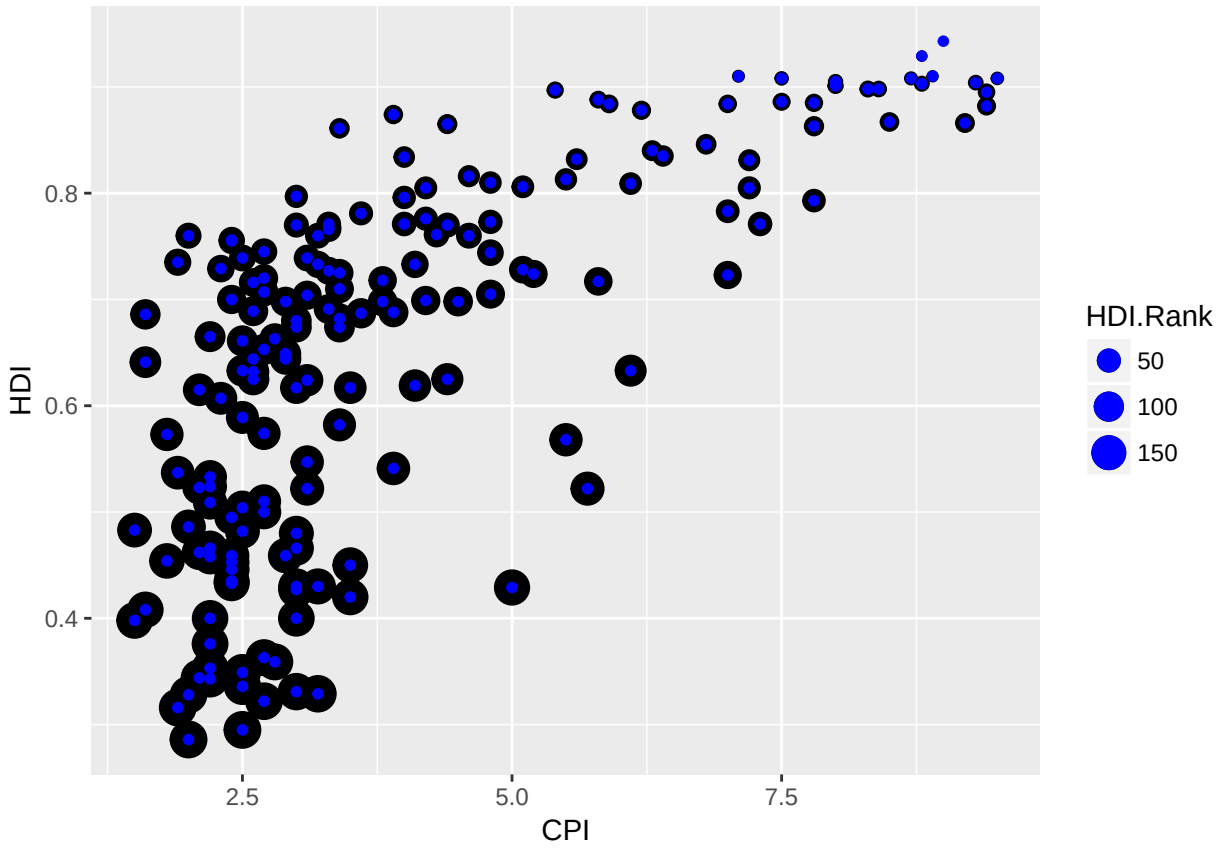

```
p4 + scale_x_discrete(name="State Abbreviation") +  scale_color_continuous(name="", breaks = c(1976, 199
```



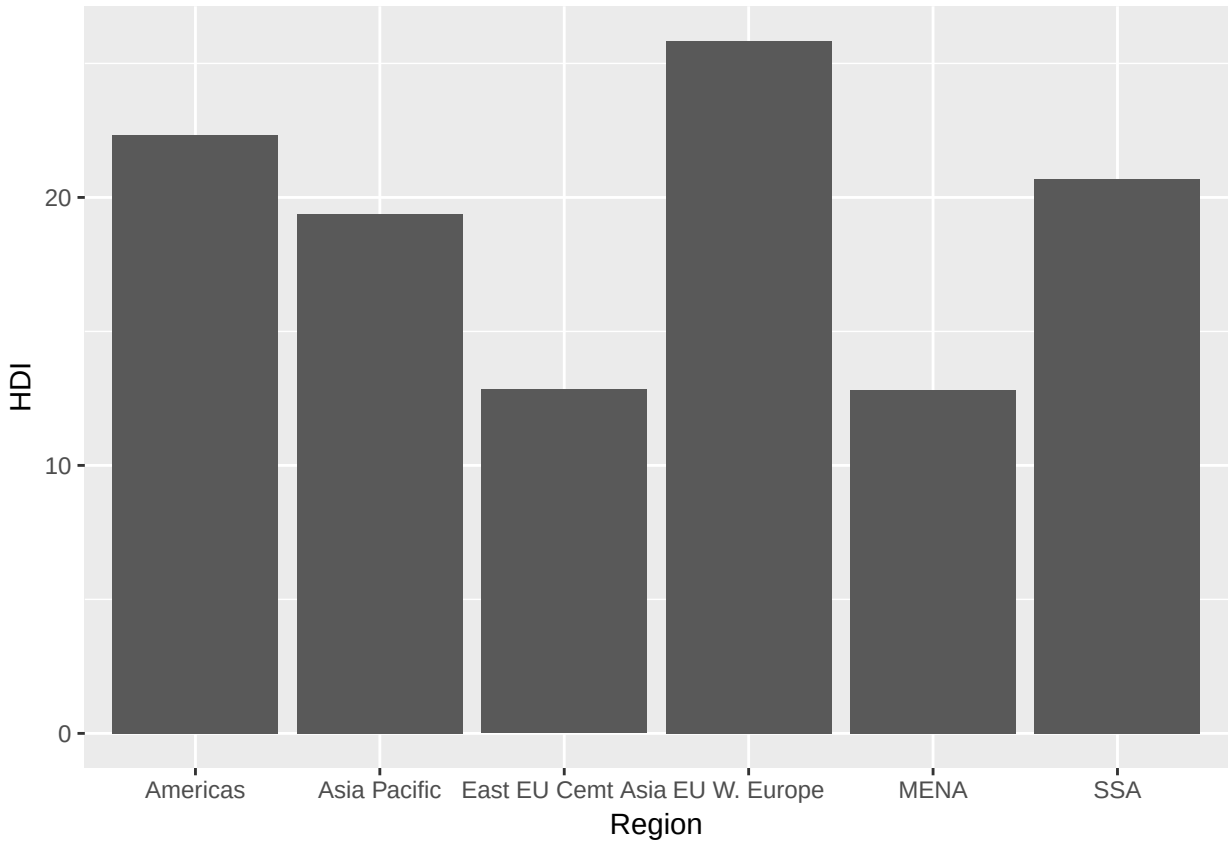
```
p4 + scale_x_discrete(name="State Abbreviation") + scale_color_continuous(name="", breaks = c(1976, 199
```



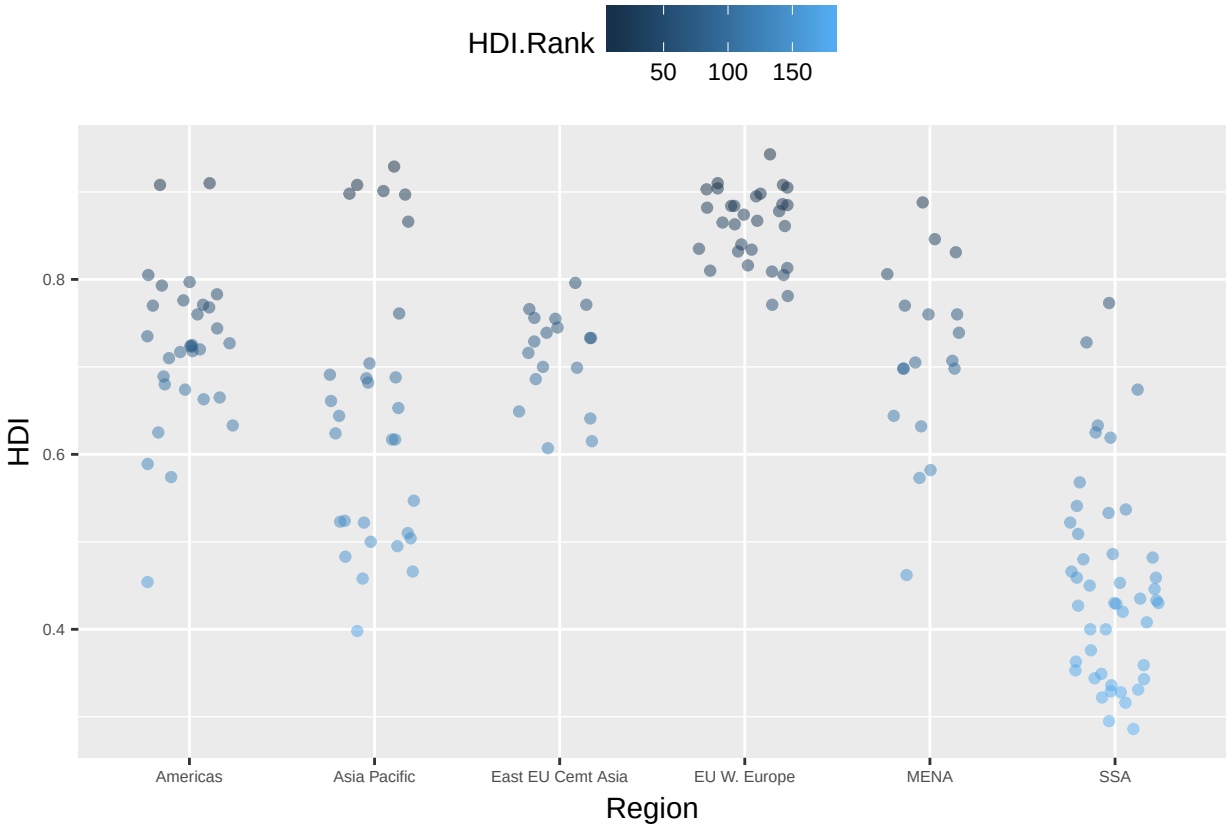

```
dat <- read.csv("~/Desktop/eeb-177/Lab-Work/exercise-8/Rgraphics/dataSets/EconomistData.csv")

p2 <- ggplot(dat, aes(x = CPI, y = HDI, size = HDI.Rank)) + geom_point()
p2 +  geom_point(aes(size = 2), color="blue")
```




```
ggplot(dat, aes(x=Region, y=HDI)) + geom_bar(stat="identity")
```




```
p3 <- ggplot(dat, aes(x = Region, y = HDI)) + theme(legend.position="top", axis.text=element_text(size =
(p4 <- p3 + geom_point(aes(color = HDI.Rank), alpha = 0.5, size = 1.5, position = position_jitter(width
```


HDI.Rank
50
100
150
HDI
0.8
0.6
0.4
Americas
Asia Pacific
East EU Cemt Asia
EU W. Europe
MENA
SSA
Region