Adding to the story: Annotations, maps and interactions

Abhijit Dasgupta, PhD

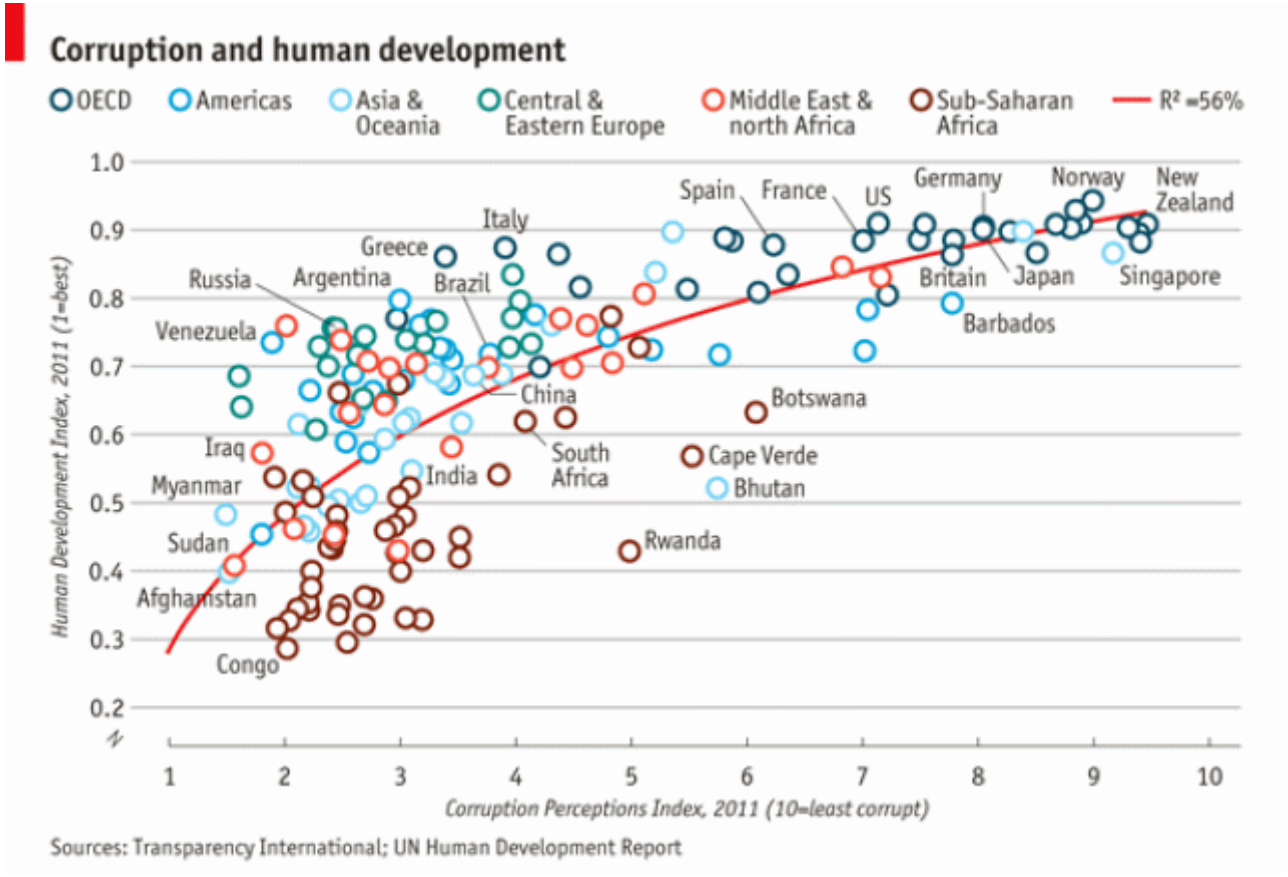
Spring 2019

Annotations

Stand-alone stories

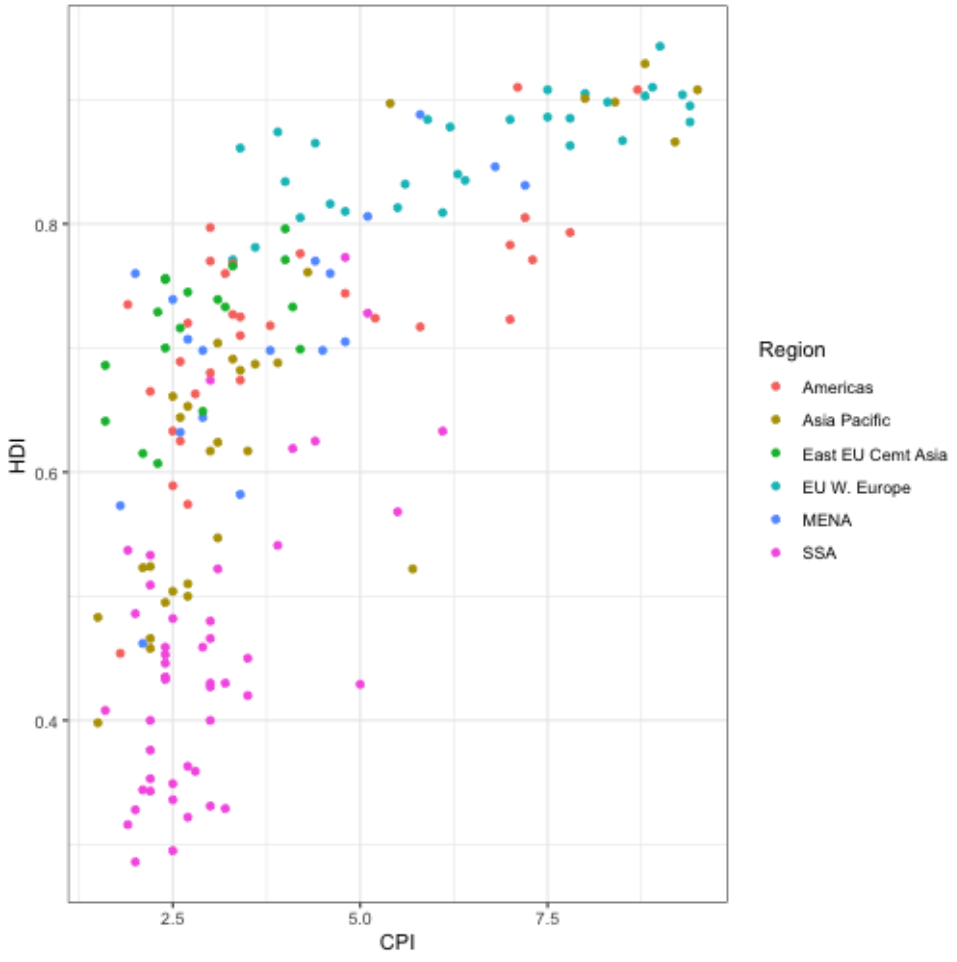
- You would like a data visualization to stand on its own
- Relevant information should be placed on the graph
- However, you need to balance the information content with real estate
    - Don't clutter the graph and make it not readable

An example


Corruption and human development
OECD
Americas
Asia & Oceania
Central & Eastern Europe
Middle East & north Africa
Sub-Saharan Africa
R² =56%
Human Development Index, 2011 (1=best)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
Spain
France
US
Germany
Norway
New Zealand
Italy
Greece
Russia
Argentina
Brazil
Britain
Japan
Singapore
Venezuela
Barbados
China
Botswana
Iraq
South Africa
Cape Verde
India
Myanmar
Bhutan
Sudan
Rwanda
Afghanistan
Congo
1
2
3
4
5
6
7
8
9
10
Corruption Perceptions Index, 2011 (10=least corrupt)
Sources: Transparency International; UN Human Development Report

Reconstructing this annotated graph

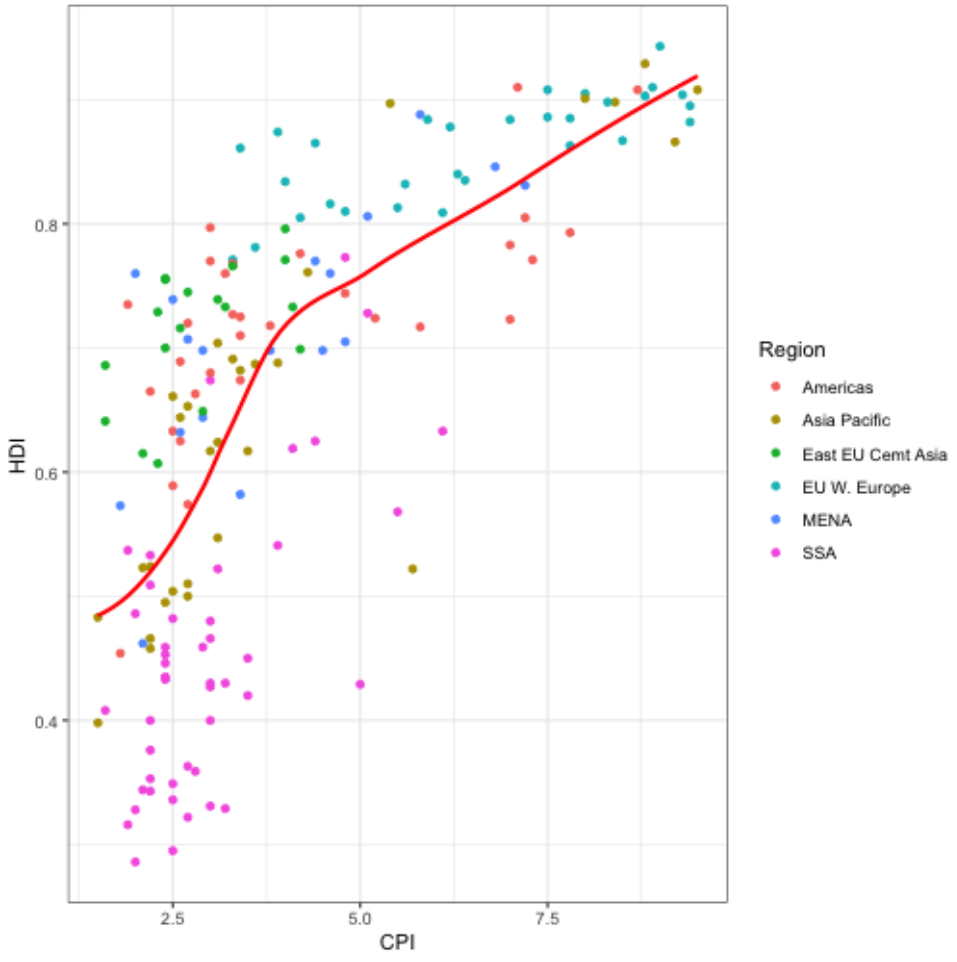
```
library(tidyverse)
econ_data <- rio::import('data/EconomistData.csv')
ggplot(econ_data,
       aes(x = CPI, y = HDI, color=Region))+
  geom_point()
```



Reconstructing this annotated graph

```
ggplot(econ_data,
       aes(x = CPI, y = HDI, color=Region))+
  geom_point() +
  geom_smooth(color='red', se=F)
```

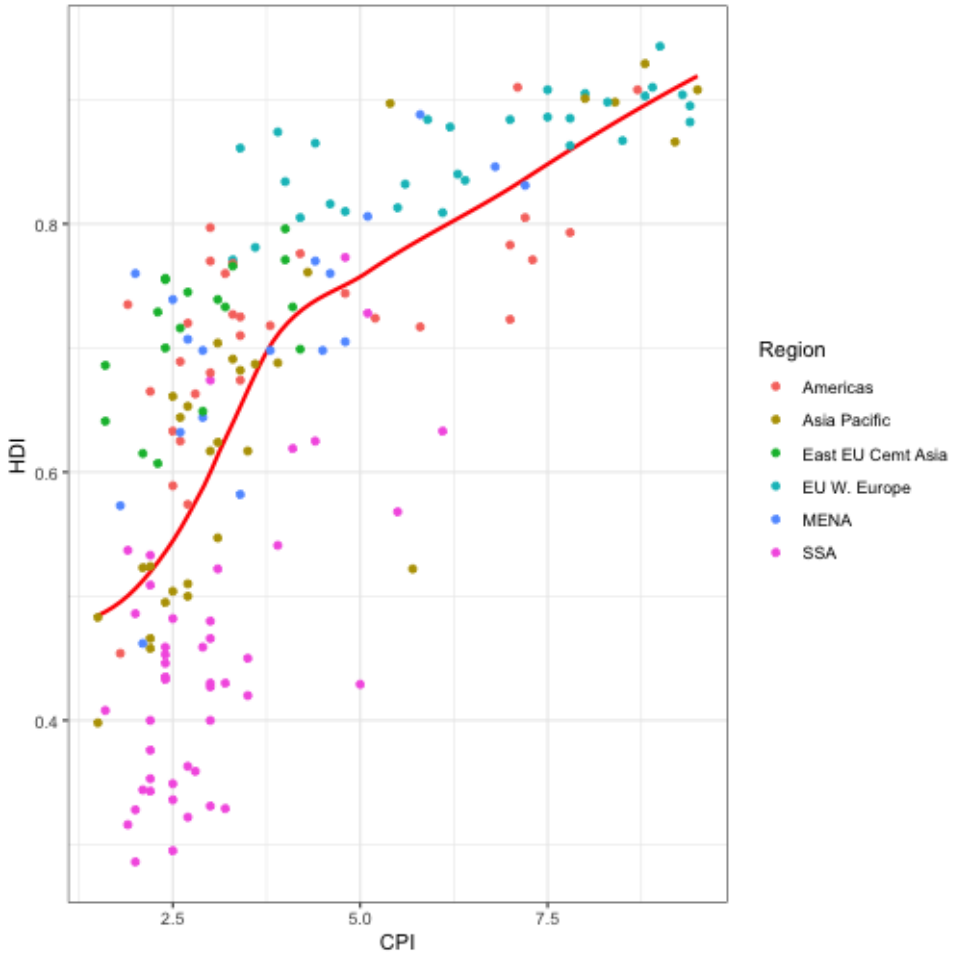
Add a trend line



Reconstructing this annotated graph

```
ggplot(econ_data,
       aes(x = CPI, y = HDI, color=Region))+
  geom_smooth(color='red', se=F) +
  geom_point()
```

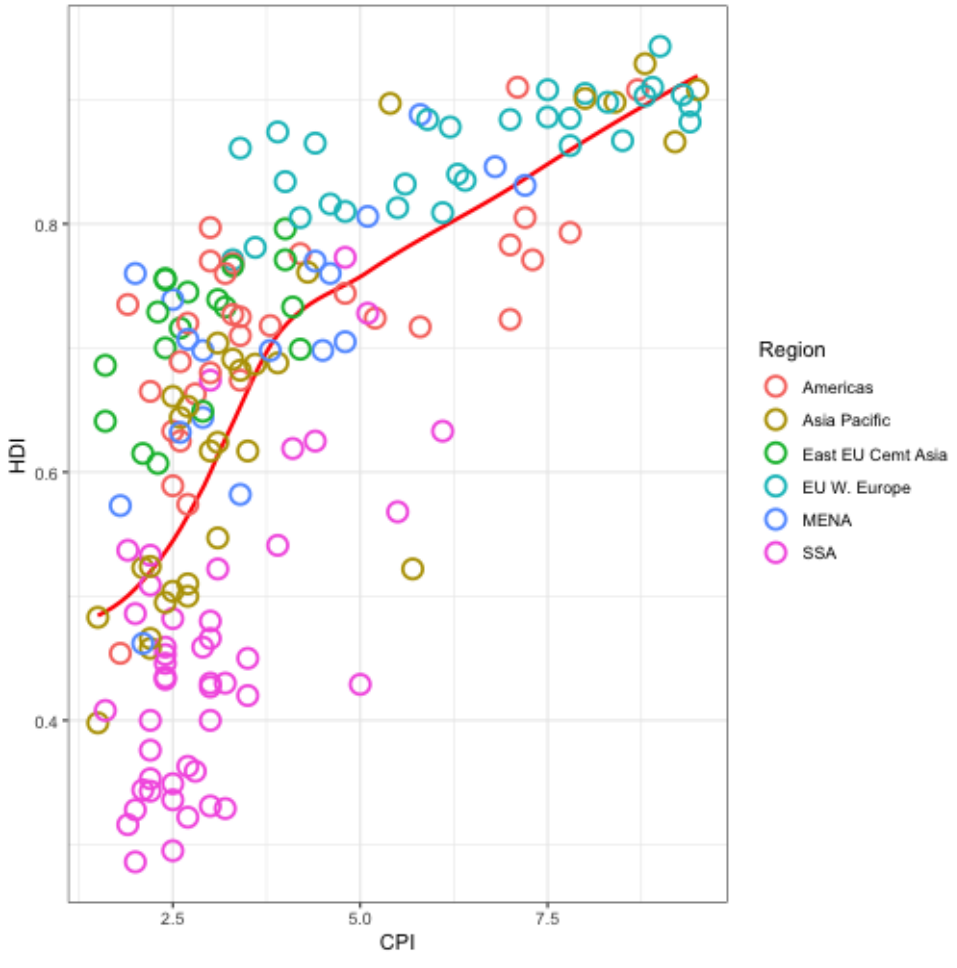
Reverse order so points are above line



Reconstructing this annotated graph

```
ggplot(econ_data,
       aes(x = CPI, y = HDI, color=Region))+
  geom_smooth(color='red', se=F) +
  geom_point(shape = 1, size = 4, stroke=1.25)
```

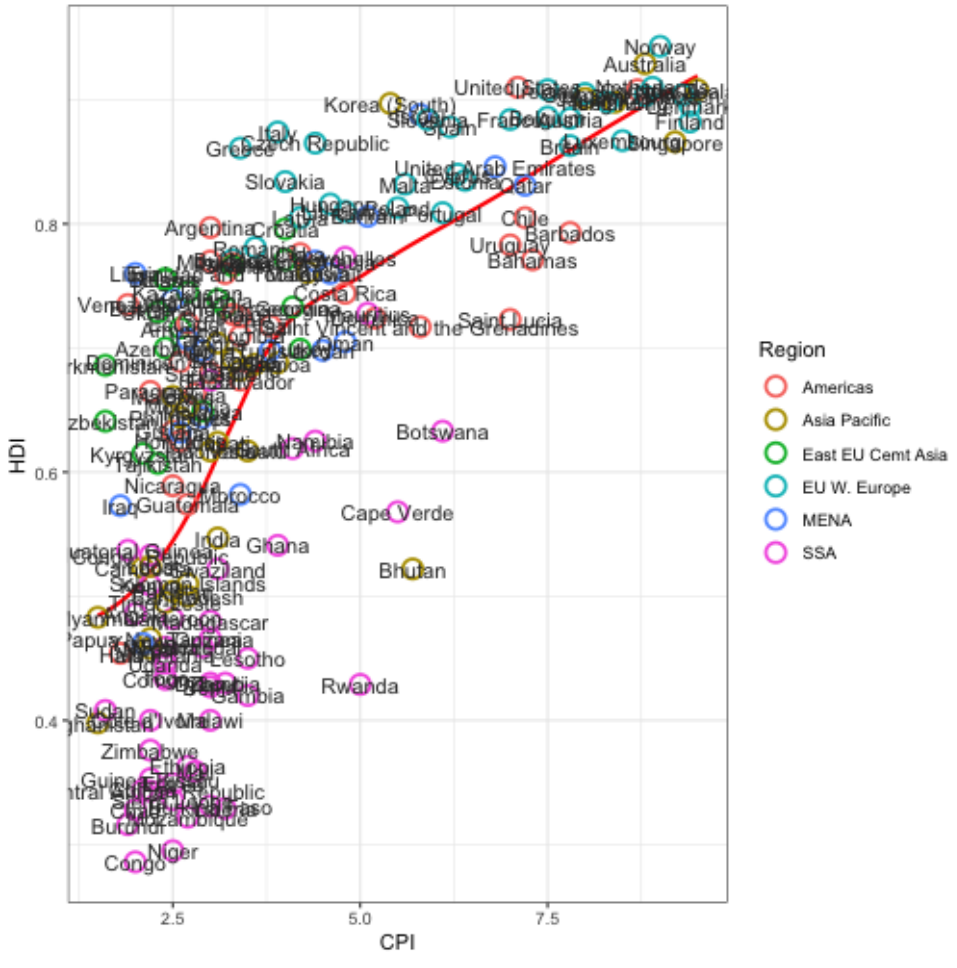
Different shape for points



Reconstructing this annotated graph

```
pointsToLabel <- c("Russia", "Venezuela", "Iraq", "My
                   "Afghanistan", "Congo", "Greece",
                   "India", "Italy", "China", "South
                   "Botswana", "Cape Verde", "Bhutan"
                   "United States", "Germany", "Brita
                   "New Zealand", "Singapore")
ggplot(econ_data,
       aes(x = CPI, y = HDI, color=Region))+
  geom_smooth(color='red', se=F) +
  geom_point(shape = 1, size = 4, stroke=1.25) +
  geom_text(aes(label=Country),
            color = 'gray20')
```

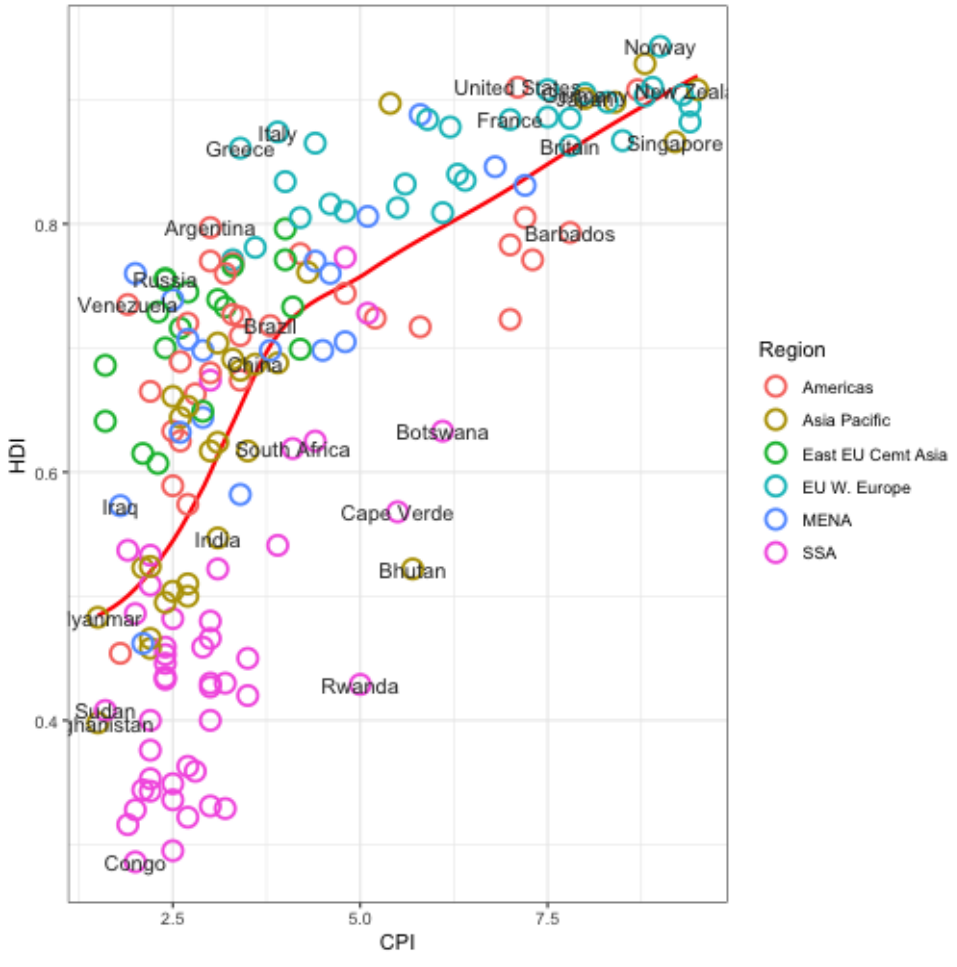
Label countries



Reconstructing this annotated graph

```
pointsToLabel <- c("Russia", "Venezuela", "Iraq", "My
                   "Afghanistan", "Congo", "Greece",
                   "India", "Italy", "China", "South
                   "Botswana", "Cape Verde", "Bhutan"
                   "United States", "Germany", "Brita
                   "New Zealand", "Singapore")
ggplot(econ_data,
       aes(x = CPI, y = HDI, color=Region))+
  geom_smooth(color='red', se=F) +
  geom_point(shape = 1, size = 4, stroke=1.25) +
  geom_text(aes(label=Country),
            color = 'gray20',
            data = econ_data %>%
              filter(Country %in% pointsToLabel))
```

Better, but labels are overlayed on points



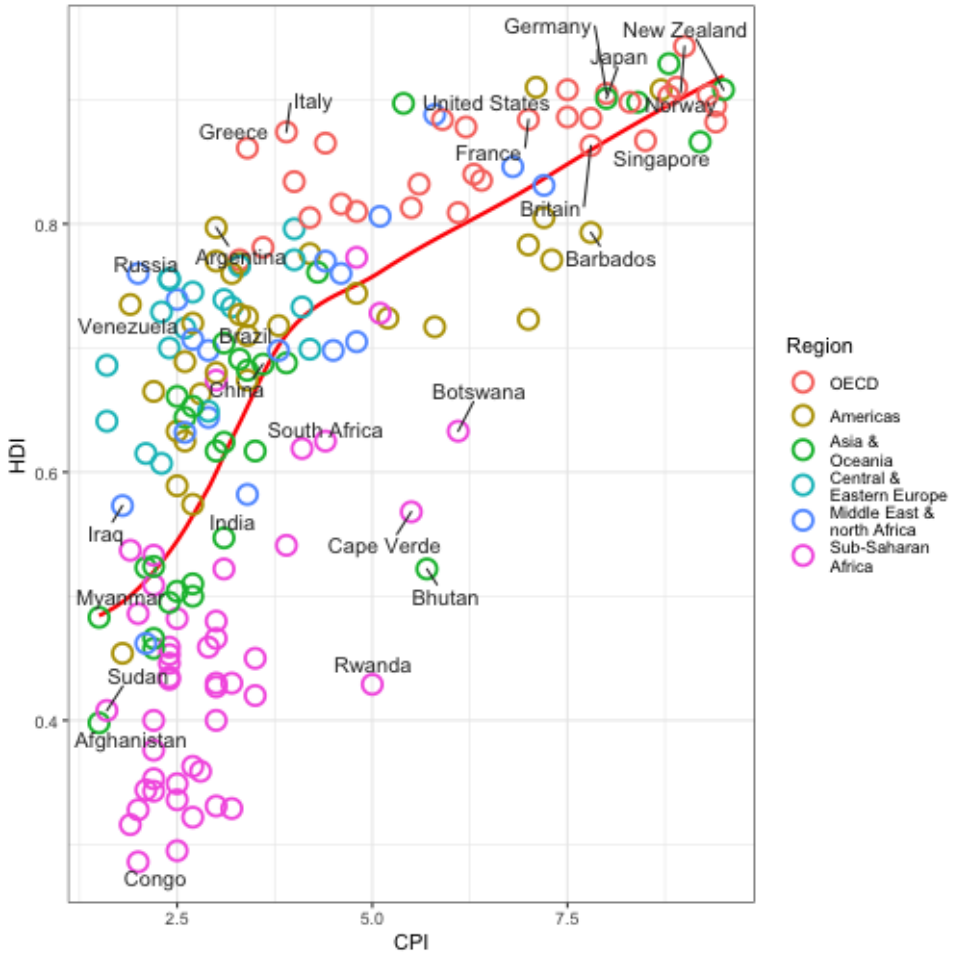
Reconstructing this annotated graph

```
library(ggrepel)
pointsToLabel <- c("Russia", "Venezuela", "Iraq", "My
                   "Afghanistan", "Congo", "Greece",
                   "India", "Italy", "China", "South
                   "Botswana", "Cape Verde", "Bhutan"
                   "United States", "Germany", "Brita
                   "New Zealand", "Singapore")
(plt <- ggplot(econ_data,
       aes(x = CPI, y = HDI, color=Region))+
  geom_smooth(color='red', se=F) +
  geom_point(shape = 1, size = 4, stroke=1.25) +
  geom_text_repel(aes(label=Country),
            color = 'gray20',
            force=20,
            data = econ_data %>%
              filter(Country %in% pointsToLabel)))
```

Reconstructing this annotated graph

Let's re-order the regions

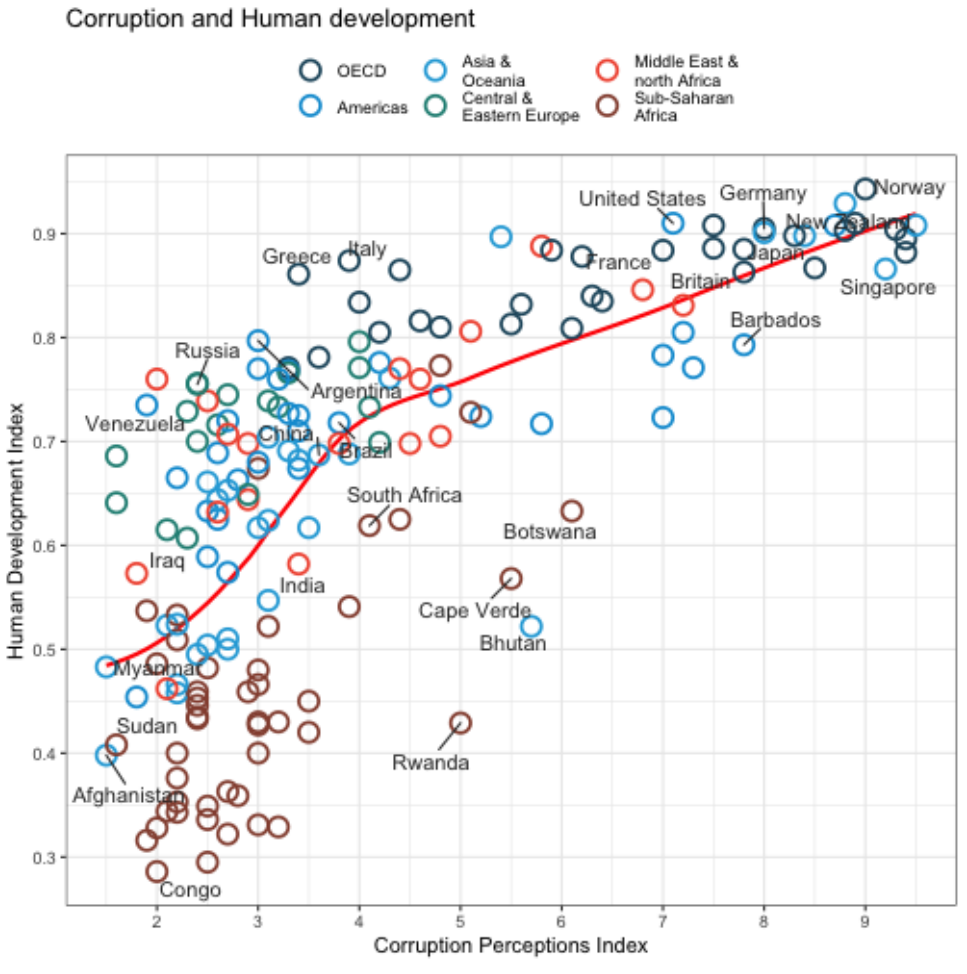
```
econ_data$Region <-
  factor(econ_data$Region,
         levels = c("EU W. Europe",
                    "Americas",
                    "Asia Pacific",
                    "East EU Cemt Asia",
                    "MENA",
                    "SSA"),
         labels = c("OECD",
                    "Americas",
                    "Asia &\nOceania",
                    "Central &\nEastern Europe",
                    "Middle East &\nnorth Africa",
                    "Sub-Saharan\nAfrica"))

plt$data = econ_data
plt
```



Reconstructing this annotated graph

Clean up the graphic

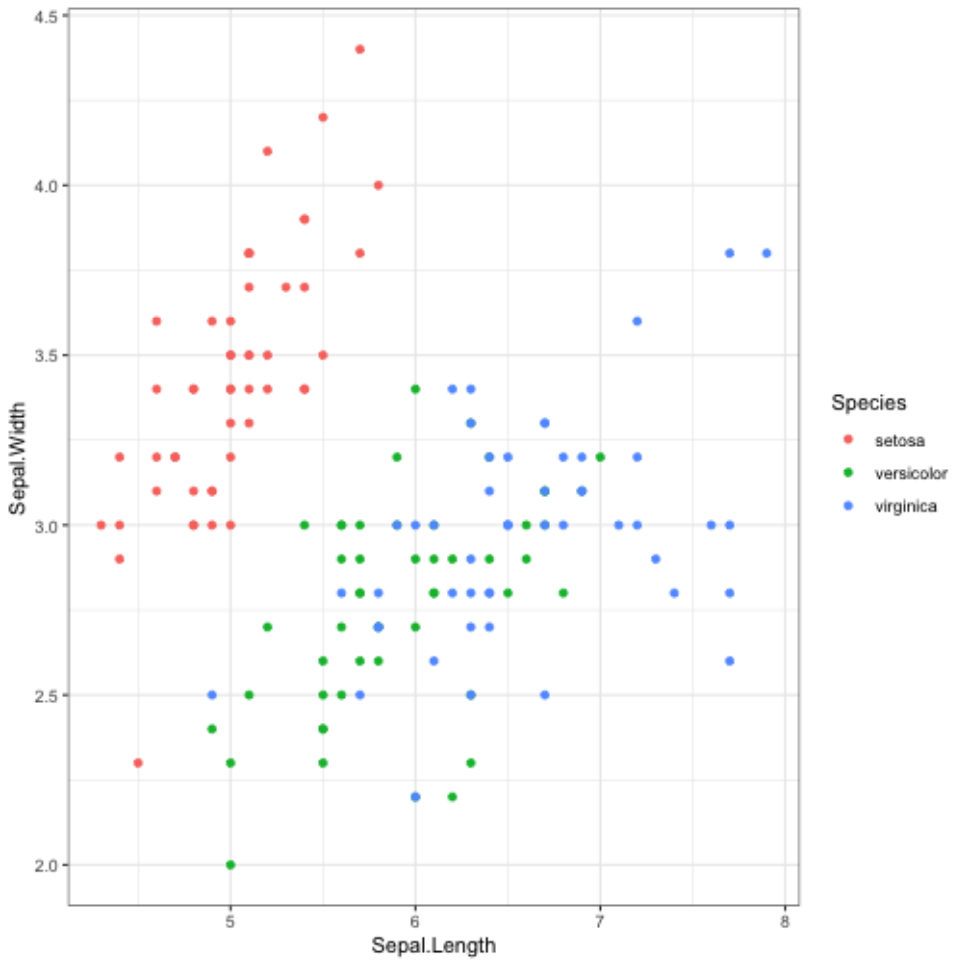
```
plt +
  scale_x_continuous(name = 'Corruption Perceptions I
                     breaks = 1:10) +
  scale_y_continuous(name="Human Development Index",
                     breaks = seq(0.2, 1, by = 0.1))+
  scale_color_manual(name = '',
                     values = c("#24576D",
                                "#099DD7",
                                "#28AADC",
                                "#248E84",
                                "#F2583F",
                                "#96503F")) +
  ggtitle("Corruption and Human development")+
  theme_bw()+
  theme(legend.position='top',
        legend.direction='horizontal')
```



Adding derived statistics to a plot

Adding group means

```
ggplot(iris,
       aes(x = Sepal.Length,
           y = Sepal.Width,
           color = Species))+
  geom_point()+
  theme_bw()
```

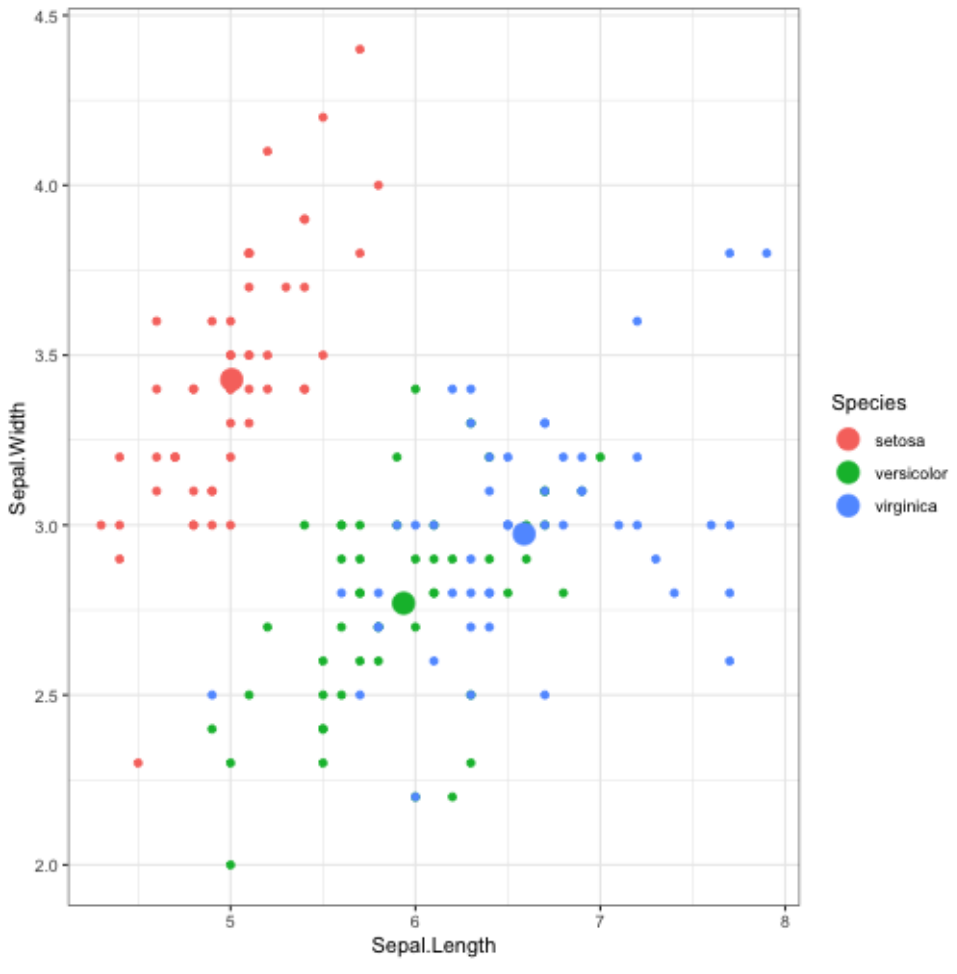


Adding group means

```
means <- iris %>% group_by(Species) %>%
  summarize_at(vars(starts_with('Sepal')),
               mean)
means
```

```
#>  # A tibble: 3 x 3
#>    Species    Sepal.Length Sepal.Width
#>    <fct>             <dbl>       <dbl>
#>  1 setosa             5.01        3.43
#>  2 versicolor         5.94        2.77
#>  3 virginica          6.59        2.97
```

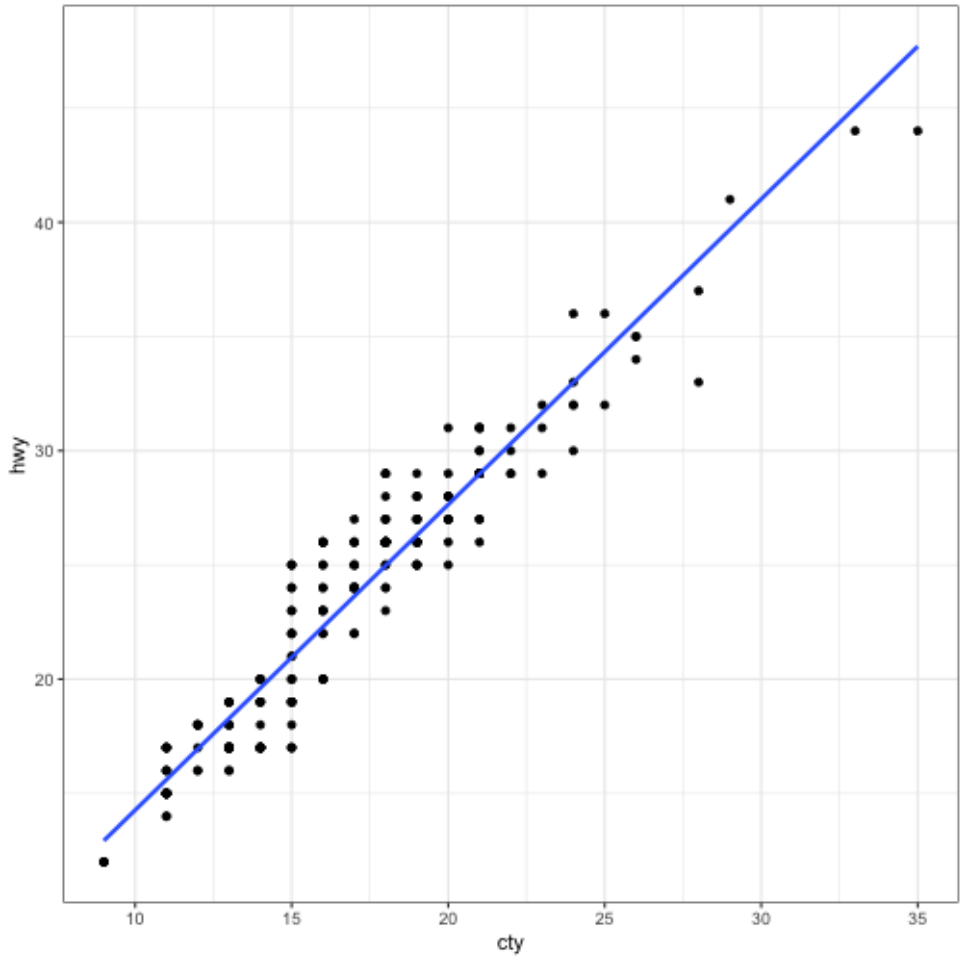
```
ggplot(iris,
       aes(x = Sepal.Length,
           y = Sepal.Width,
           color = Species))+
  geom_point() +
  geom_point(data = means,
             size=5) +
  theme_bw()
```



Adding regression metrics

Regress highway mileage on city mileage (data: mpg)

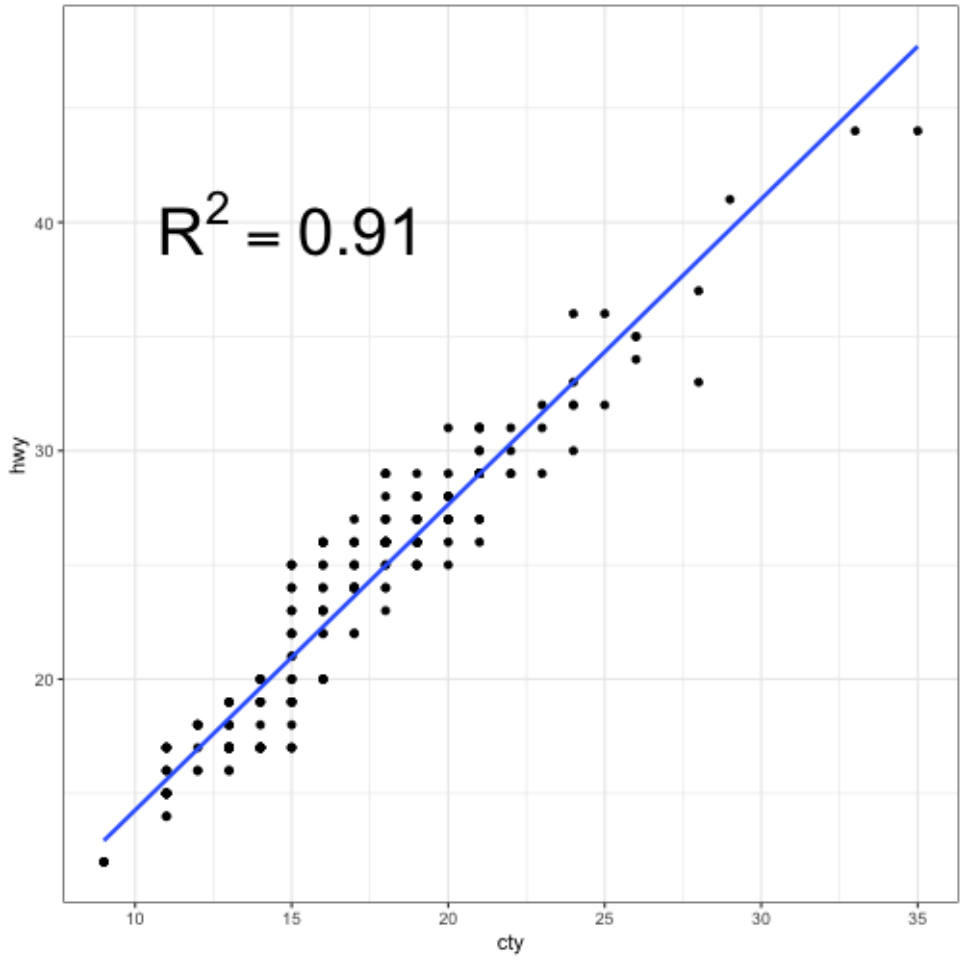
```
mod1 <- lm(hwy ~ cty, data = mpg)
r2 <- broom::glance(mod1) %>% pull(r.squared)

ggplot(mpg,
       aes(x = cty, y = hwy))+
  geom_point() +
  geom_smooth(method = 'lm', se=F) +
  theme_bw()
```



Adding regression metrics

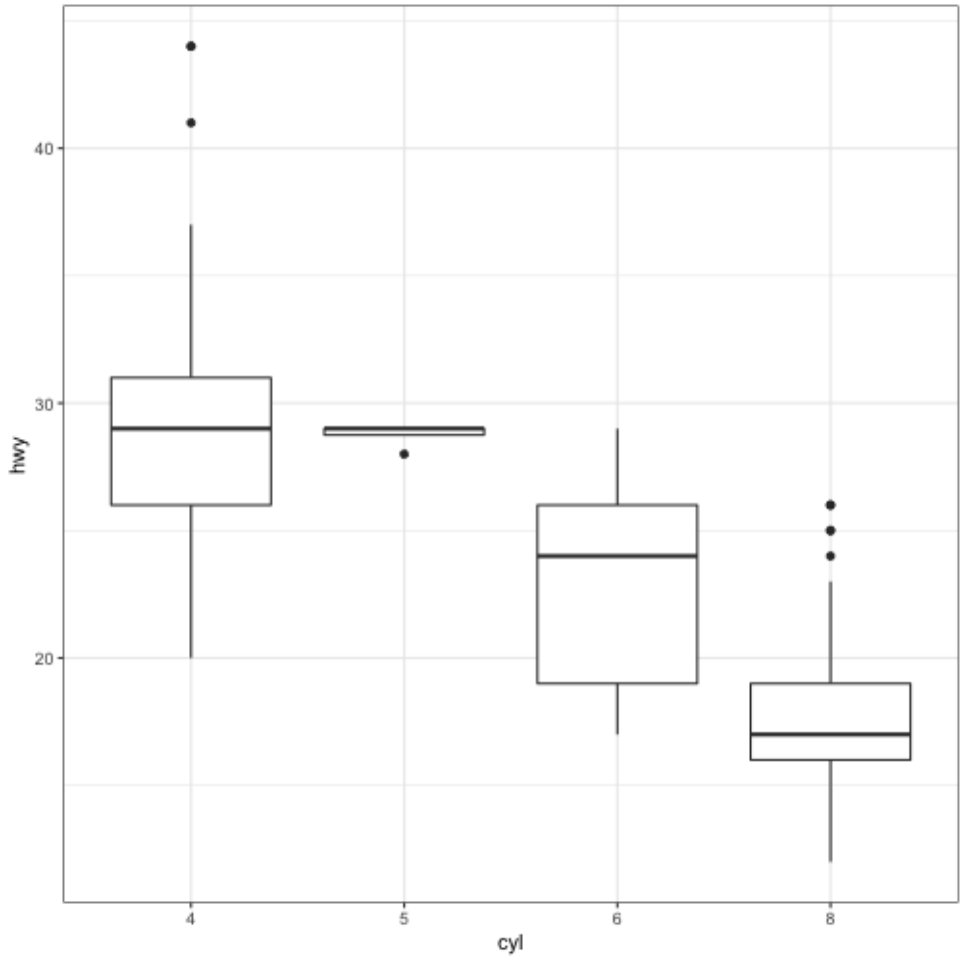
Regress highway mileage on city mileage (data: mpg)

```
mod1 <- lm(hwy ~ cty, data = mpg)
r2 <- broom::glance(mod1) %>% pull(r.squared) %>%
  round(., 2)

ggplot(mpg,
       aes(x = cty, y = hwy))+
  geom_point() +
  geom_smooth(method = 'lm', se=F)+
  annotate(geom='text',
           x = 15, y = 40,
           label=glue::glue("R^2 == {r}",r=r2),
           size=12,
           parse=T) +
  theme_bw()
```



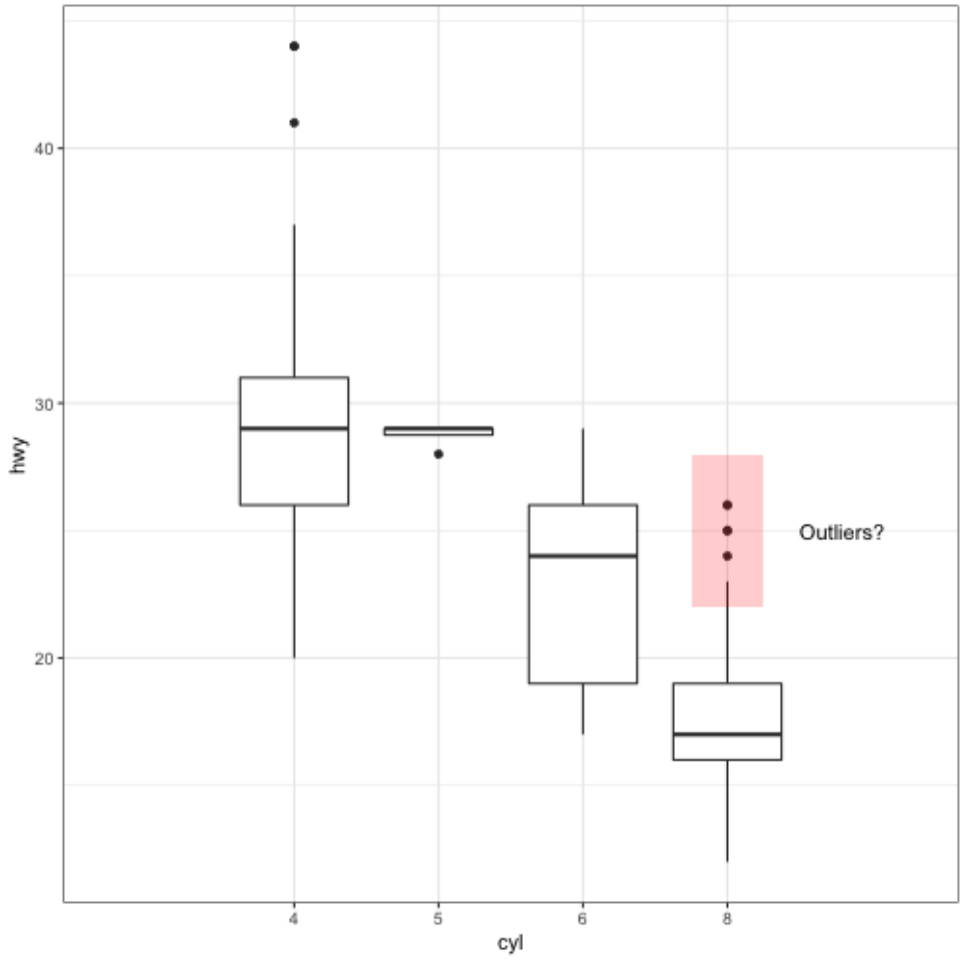
Highlighting regions

```
mpg %>%
  mutate(cyl = as.factor(cyl)) %>%
  ggplot(aes(x = cyl, y = hwy)) +
  geom_boxplot() +
  theme_bw()
```



Highlighting regions

```
mpg %>%
  mutate(cyl = as.factor(cyl)) %>%
  ggplot(aes(x = cyl, y = hwy)) +
  geom_boxplot() +
  theme_bw()+
  annotate(geom = 'rect',
           xmin=3.75,xmax=4.25,
           ymin = 22, ymax = 28,
           fill = 'red',
           alpha = 0.2) +
  annotate('text',
           x = 4.5, y = 25,
           label = 'Outliers?',
           hjust = 0)+
  coord_cartesian(xlim = c(0,5))+
  theme_bw()
```



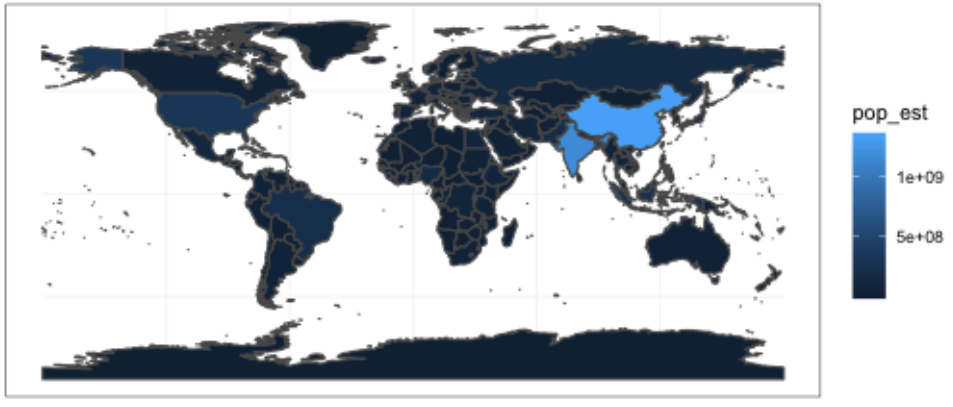
Maps

For maps, we need a couple of new packages.

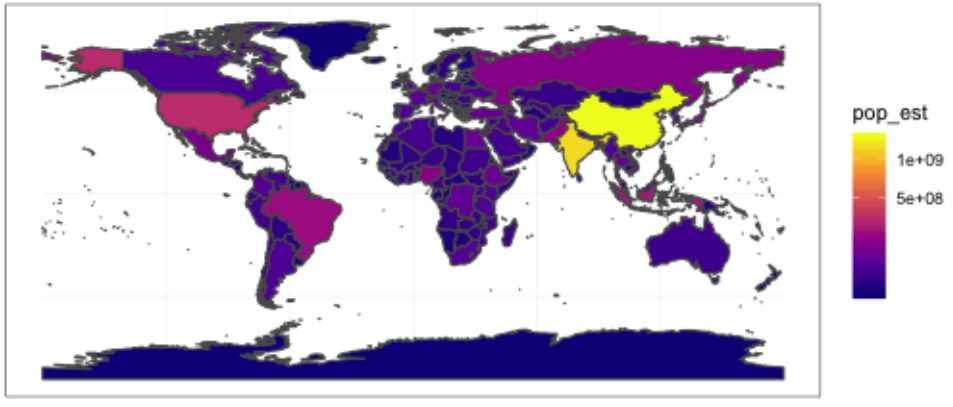
- `sf`: Simple features in R
- `rnaturalearth` & `rnaturalearthdata`: map data

```
library(sf)
library(rnaturalearth)
library(rnaturalearthdata)

world <- ne_countries(scale='medium', returnclass='sf
ggplot(data = world) +
  geom_sf()
```

```
library(sf)
library(rnaturalearth)
library(rnaturalearthdata)

world <- ne_countries(scale='medium', returnclass='sf
ggplot(data = world) +
  geom_sf(aes(fill = pop_est))
```

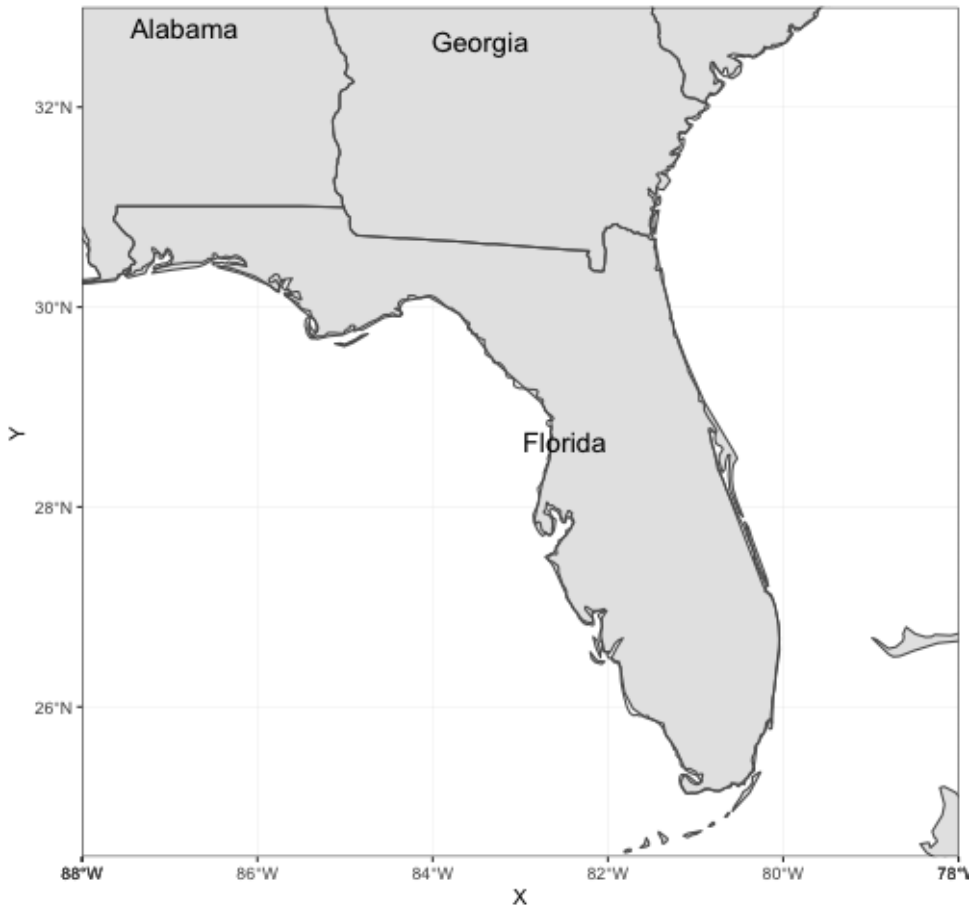


```
library(sf)
library(rnaturalearth)
library(rnaturalearthdata)

world <- ne_countries(scale='medium', returnclass='sf
ggplot(data = world) +
  geom_sf(aes(fill = pop_est))+
  scale_fill_viridis_c(option = 'plasma', trans='sqrt
```



Looking at Florida

```
library(maps)
states <- st_as_sf(map('state', plot = F, fill = T))
  cbind(st_coordinates(st_centroid(states))) %>%
  mutate(ID = str_to_title(ID))

ggplot(data = world)+
  geom_sf() +
  geom_sf(data = states, fill = NA) +
  geom_text(data = states, aes(X, Y, label = ID),
            size = 5) +
  coord_sf(xlim = c(-88, -78), ylim = c(24.5, 33),
           expand = F)
```



Looking at Florida

```
library(maps)
states <- st_as_sf(map('state', plot = F, fill = T))
  cbind(st_coordinates(st_centroid(states))) %>%
  mutate(ID = str_to_title(ID))

ggplot(data = world)+
  geom_sf() +
  geom_sf(data = states, fill = NA) +
  geom_label(data = states, aes(X, Y, label = ID),
             size = 5) +
  coord_sf(xlim = c(-88, -78), ylim = c(24.5, 33),
           expand = F)
```