

Lista 5 - Capítulos 4 e 5

Gustavo Tironi

2023-06-01

Setando um tema padrão

Primeiramente, gostaria de destacar que irei apresentar os gráficos já modificados, pois não vou senti-los idênticos ao livro. Então, primeiramente irei definir um tema global, cujo será aplicado em todos os gráficos, caso não seja indicado o contrário.

```
theme_set(theme_classic())
theme_update(legend.position="top")
theme_update(plot.title = element_text(size = 12, face = "bold"))
```

Gráficos do capítulo 4

Aqui, não irei me alongar muito nas explicações, e focarei mais em apresentar os gráficos, pontuando apenas o que eu achar relevante.

Inicialmente, para evitar a repetição de **labs** desnecessários, resolvi alterar o nome das variáveis mais frequentes das *databases*, para assim, os rótulos dos eixos aparecerem de forma correta, sem necessidade de usar o `labs()`. Para isso, usar o comando **rename** do *dplyr*.

```
gapminder <- gapminder::gapminder %>% rename("Ano" = "year",
  "País" = "country",
  "Continente" = "continent",
  "População" = "pop",
  "PIBpercapita" = "gdpPercap",
  "Expvida" = "lifeExp")

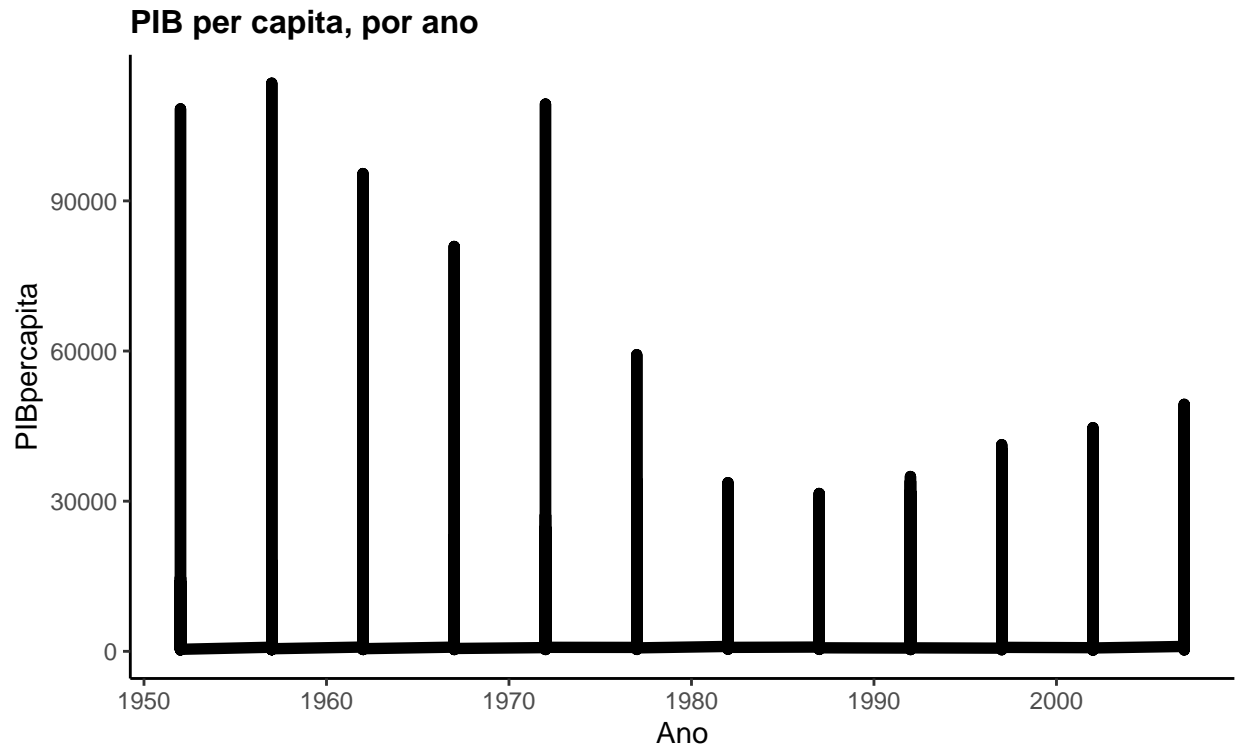
gss_sm <- socviz::gss_sm %>%
  rename("Ano" = "year",
    "Idade" = "age",
    "Crianças" = "childs",
    "Região" = "bigregion",
    "Religião" = "religion")

organdata <- socviz::organdata %>%
  rename("Ano" = "year",
    "País" = "country",
    "Doadores" = "donors",
    "População" = "pop",
    "Estradas" = "roads")
```

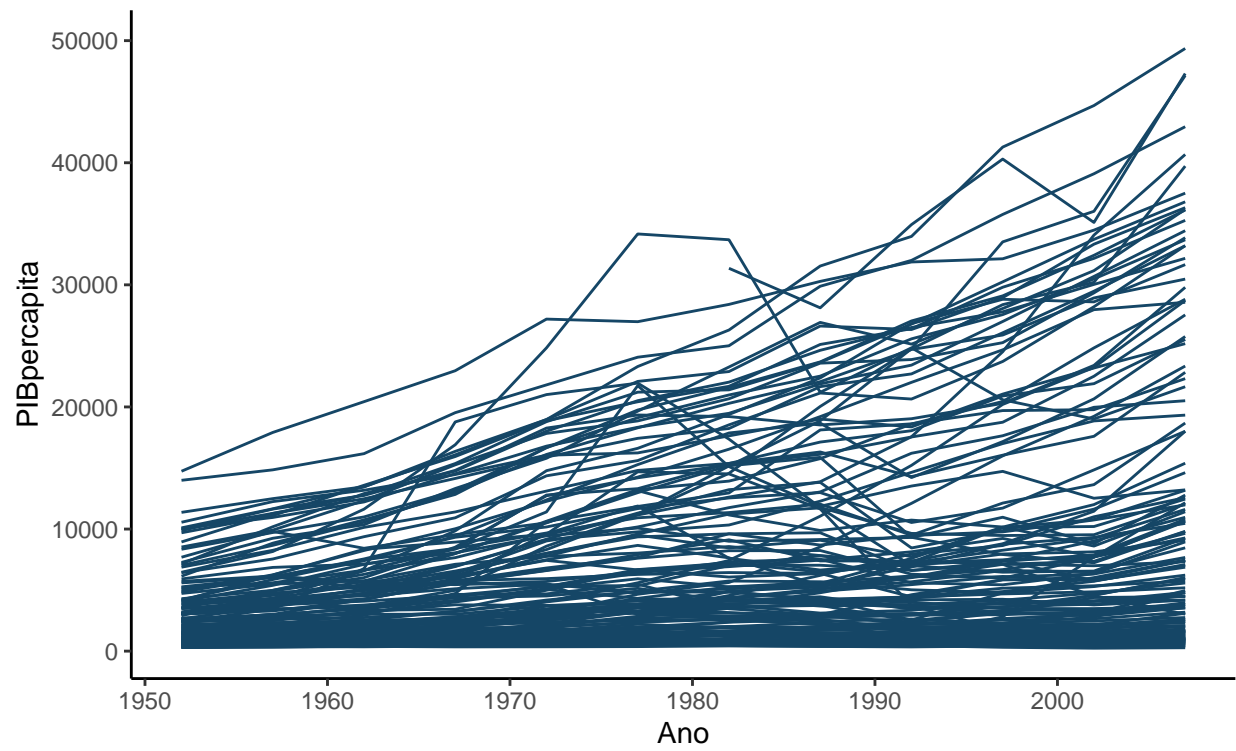
```
midwest <- ggplot2::midwest %>%
  rename("Estado" = "state",
        "Área" = "area")
```

Com isso feito, começarei a apresentar os gráficos construídos.

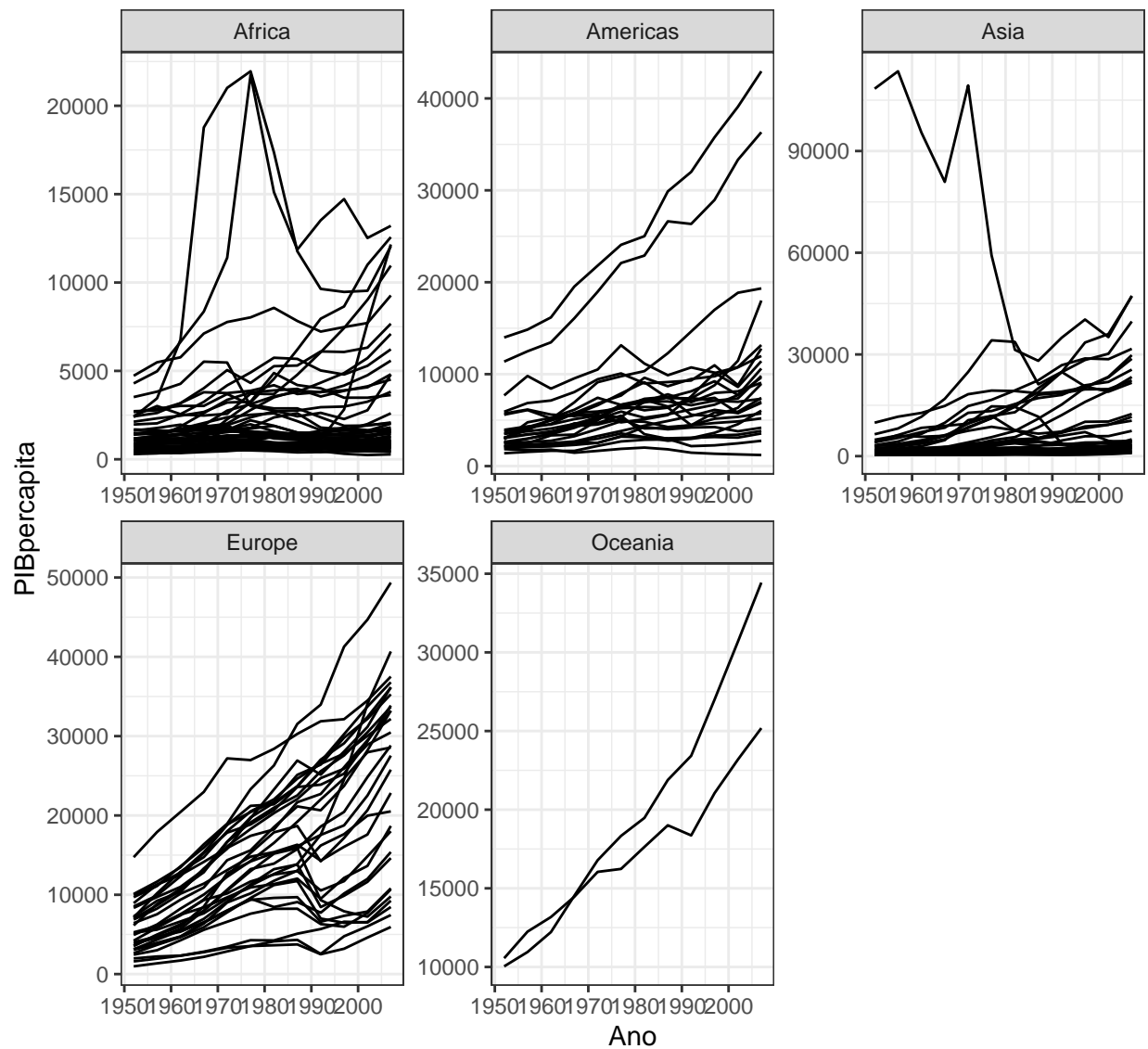
```
ggplot(gapminder, mapping = aes(x = Ano, y = PIBpercapita)) +
  geom_line(size = 2) +
  labs(title = "PIB per capita, por ano")
```



```
ggplot(gapminder, mapping = aes(x = Ano, y = PIBpercapita)) +
  geom_line(aes(group=País), color = "#154666", size = 0.5) +
  scale_y_continuous(limits = c(NA, 50000))
```

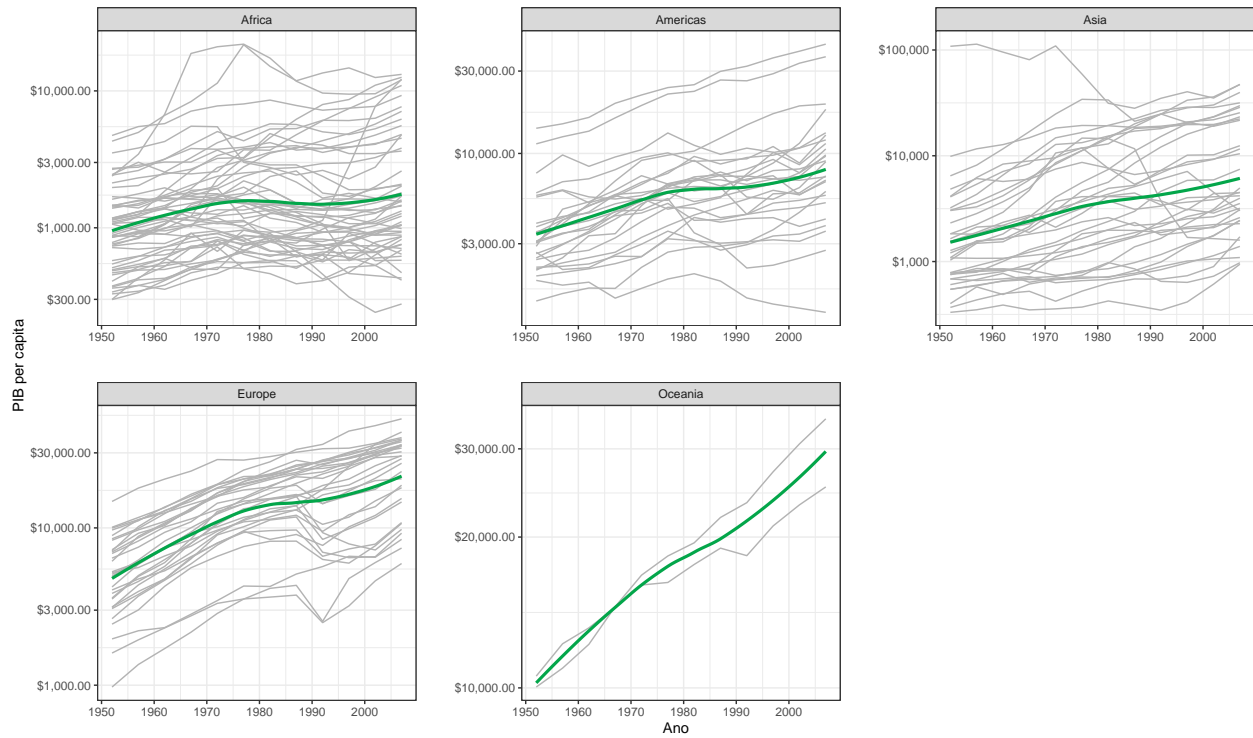


```
ggplot(gapminder, mapping = aes(x = Ano, y = PIBpercapita)) +  
  geom_line(aes(group=País)) +  
  facet_wrap(~Continente, scales = "free") +  
  theme_bw()
```

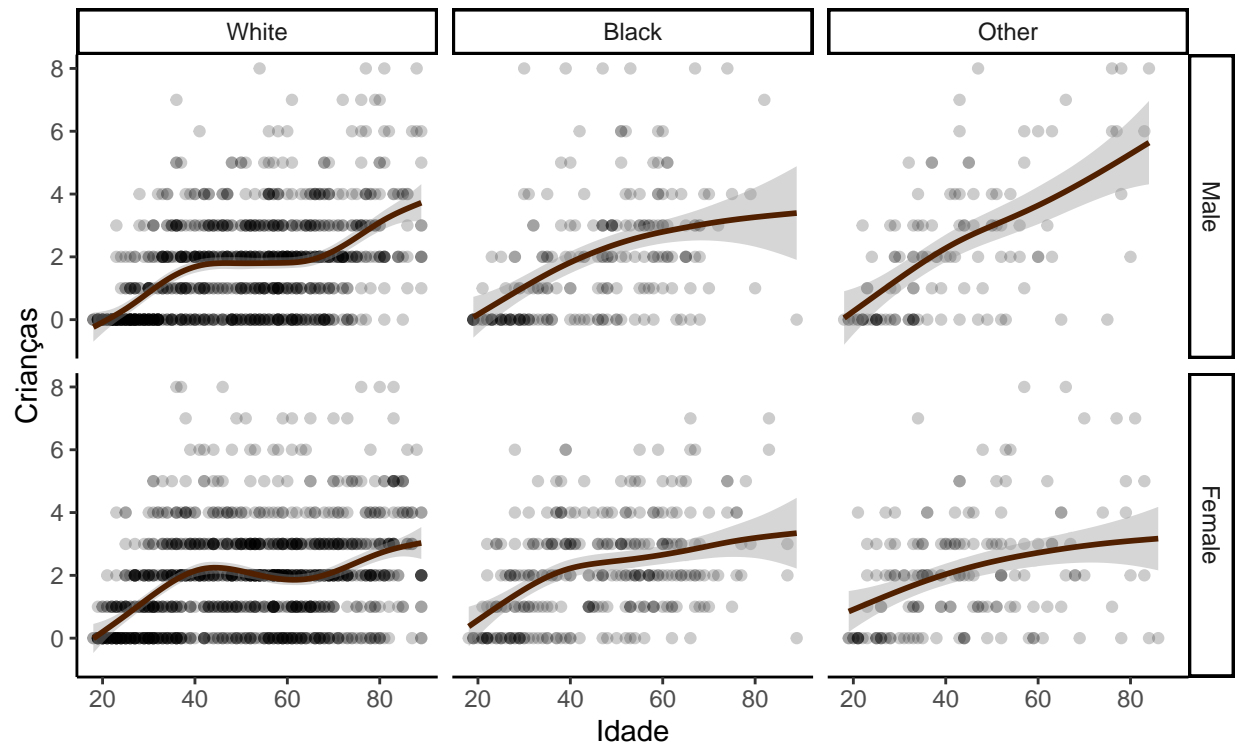


```
ggplot(gapminder, mapping = aes(x = Ano, y = PIBpercapita)) +
  geom_line(color = "gray70", aes(group=País)) +
  geom_smooth(size=1.1, method="loess", se=FALSE, color = "#03A64A") +
  scale_y_log10(labels=scales::dollar) +
  facet_wrap(~ Continente, ncol=3, scale = "free") +
  labs(x = "Ano",
       y = "PIB per capita",
       title = "PIB per capita nos cinco continentes") +
  theme_bw() +
  theme(panel.spacing = unit(2, "lines"), plot.title = element_text(size = 18, face = "bold"))
```

PIB per capita nos cinco continentes



```
ggplot(data = gss_sm, mapping = aes(x = Idade, y = Crianças)) +
  geom_point(alpha=0.2) +
  geom_smooth(color = "#502000") +
  facet_grid(sex~race)
```



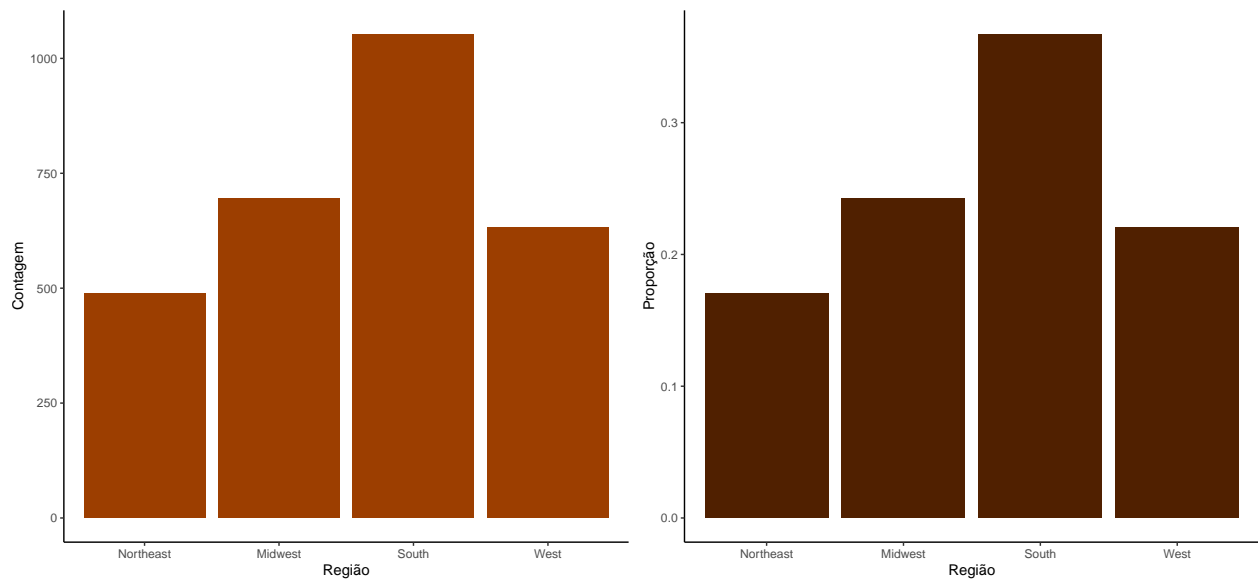
```

p1 <- ggplot(data = gss_sm, mapping = aes(x = Região)) +
  geom_bar(fill = "#9C3E00") +
  labs(y = "Contagem")

p2 <- ggplot(data = gss_sm, mapping = aes(x = Região)) +
  geom_bar(mapping = aes(y=after_stat(prop), group = 1), fill = "#502000") +
  labs(y = "Proporção")

grid.arrange(p1, p2,
              ncol=2, nrow=1)

```



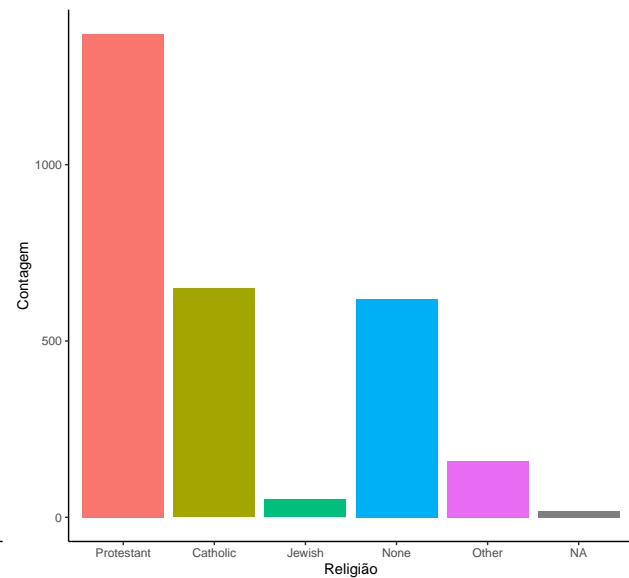
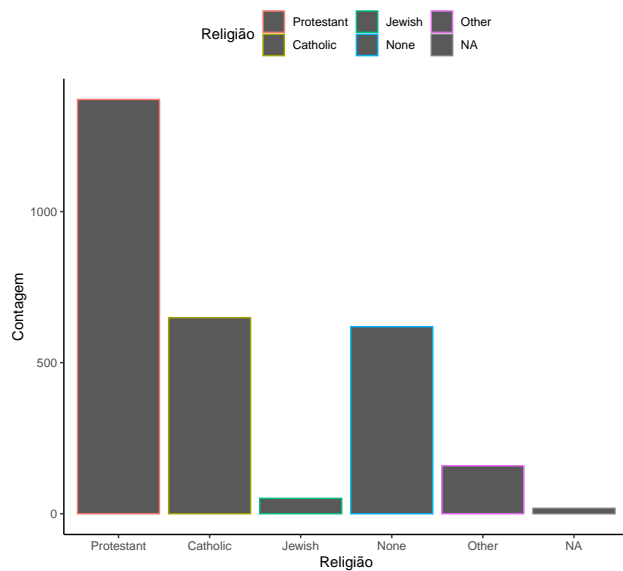
```

p1 <- ggplot(data = gss_sm, aes(x = Religião, color = Religião)) +
  geom_bar() +
  theme() +
  labs(y = "Contagem")

p2 <- ggplot(data = gss_sm, mapping = aes(x = Religião, fill = Religião)) +
  geom_bar() +
  guides(fill = FALSE) +
  labs(y = "Contagem")

grid.arrange(p1, p2,
              ncol=2, nrow=1)

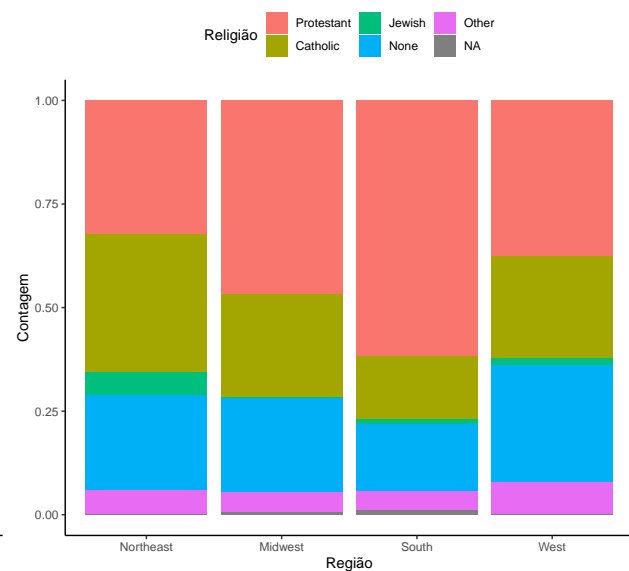
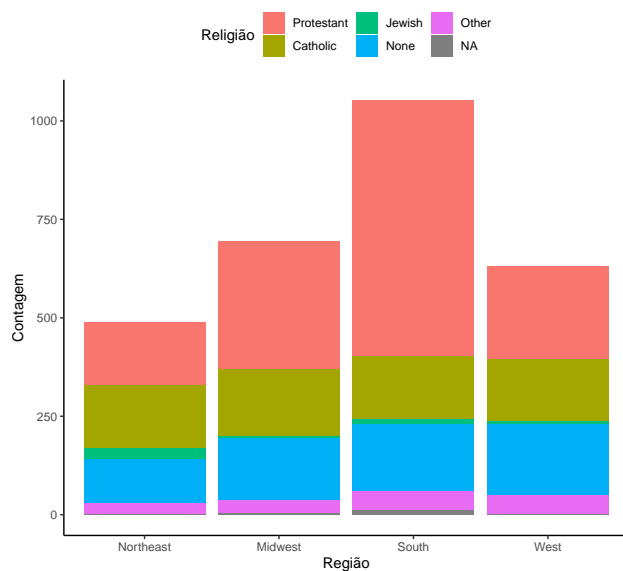
```



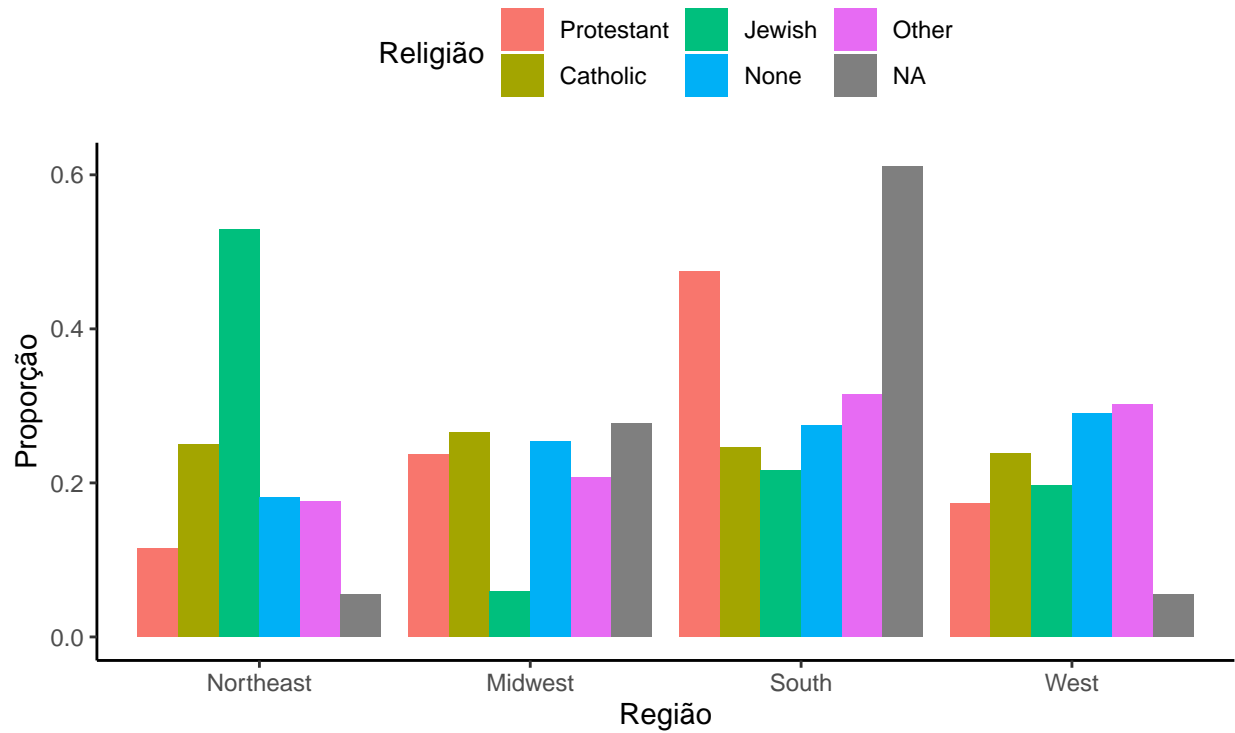
```
p1 <- ggplot(data = gss_sm, mapping = aes(x = Região, fill = Religião)) +
  geom_bar() +
  theme() +
  labs(y = "Contagem")

p2 <- ggplot(data = gss_sm, mapping = aes(x = Região, fill = Religião)) +
  geom_bar(position = "fill") +
  theme() +
  labs(y = "Contagem")

grid.arrange(p1, p2,
  ncol=2, nrow=1)
```

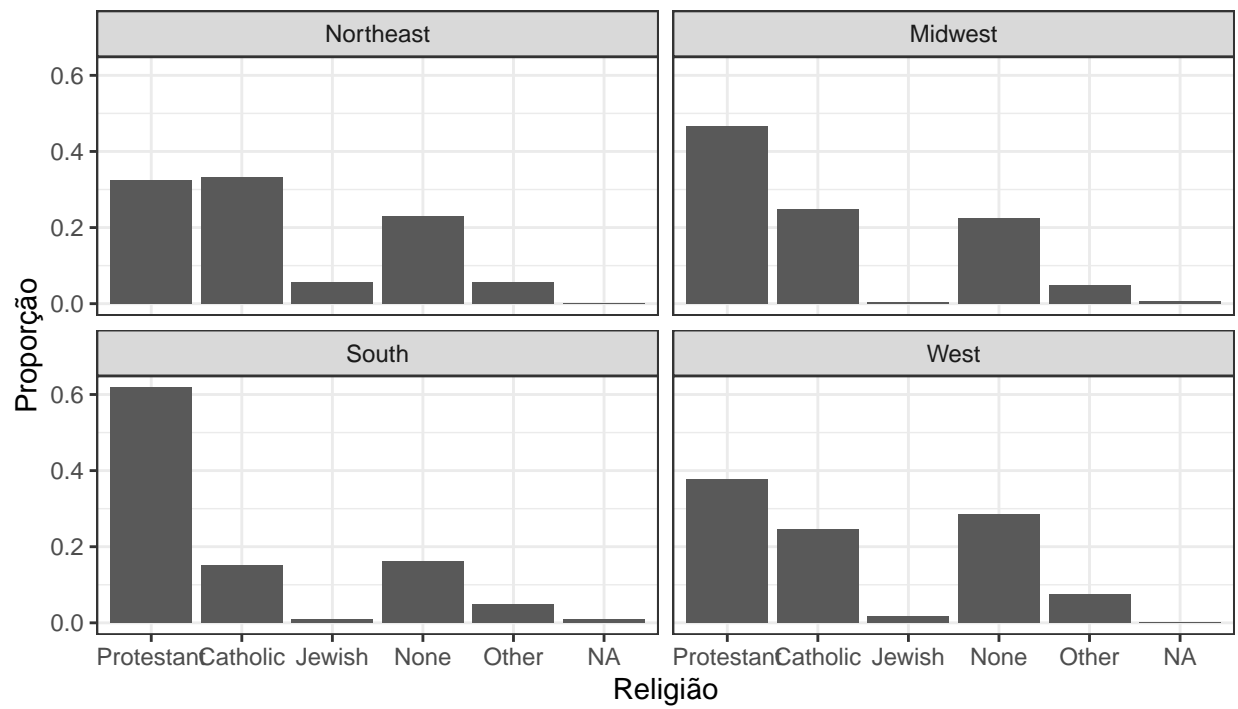


```
ggplot(data = gss_sm, mapping = aes(x = Região, fill = Religião)) +
  geom_bar(position = "dodge", aes(y = after_stat(prop), group = Religião)) +
  labs(y = "Proporção")
```

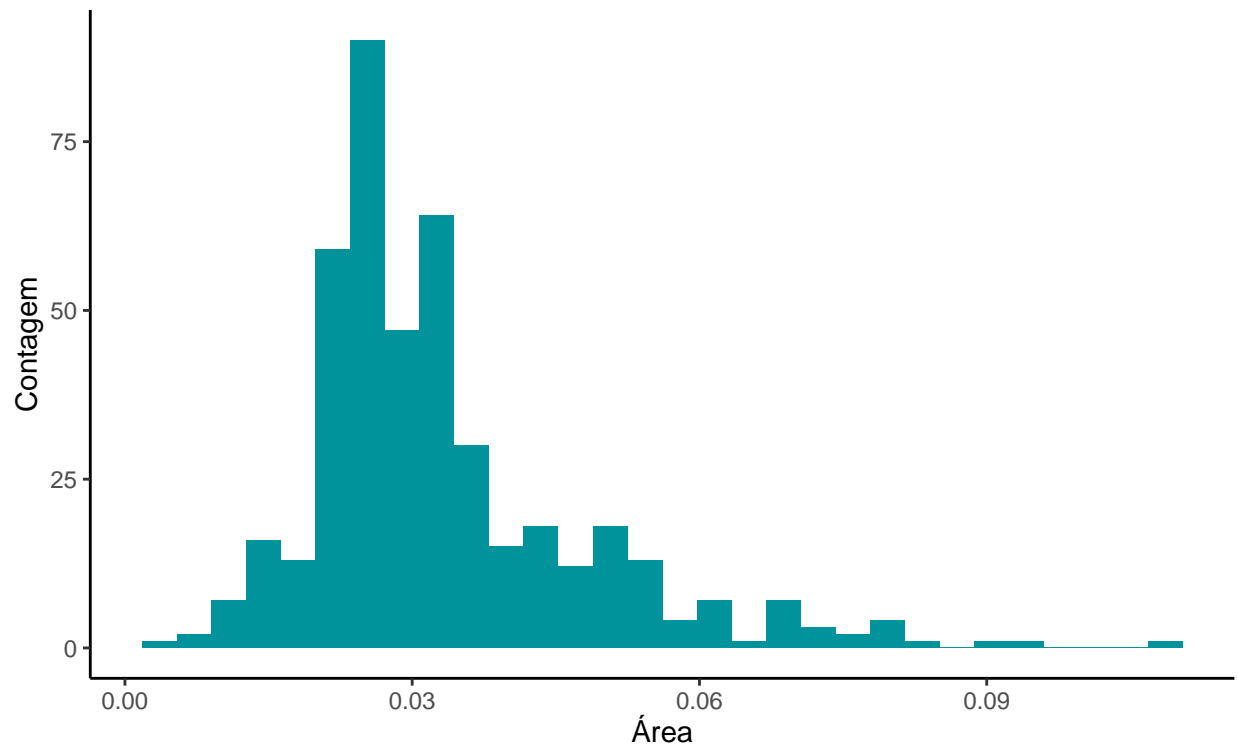


```
ggplot(data = gss_sm, mapping = aes(x = Religião)) +
  geom_bar(position = "dodge", aes(y = after_stat(prop), group = Região)) +
  facet_wrap(~Região, ncol = 2) +
  theme_bw() +
  labs(y = "Proporção") +
  ggtitle("Proporção da religião, por região") +
  theme(plot.title = element_text(size = 12, face = "bold"))
```

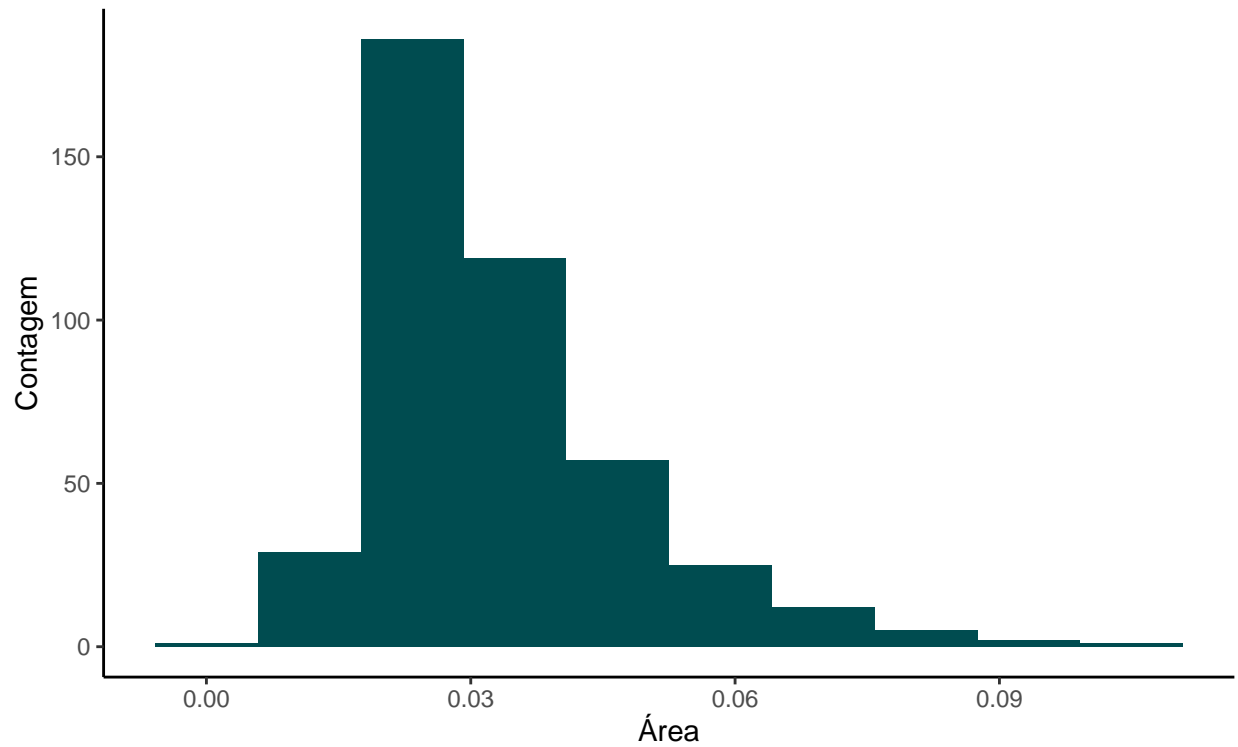

Proporção da religião, por região



```
ggplot(data = midwest, mapping = aes(x = Área)) +  
  geom_histogram(fill = "#00939C") +  
  labs(y = "Contagem")
```

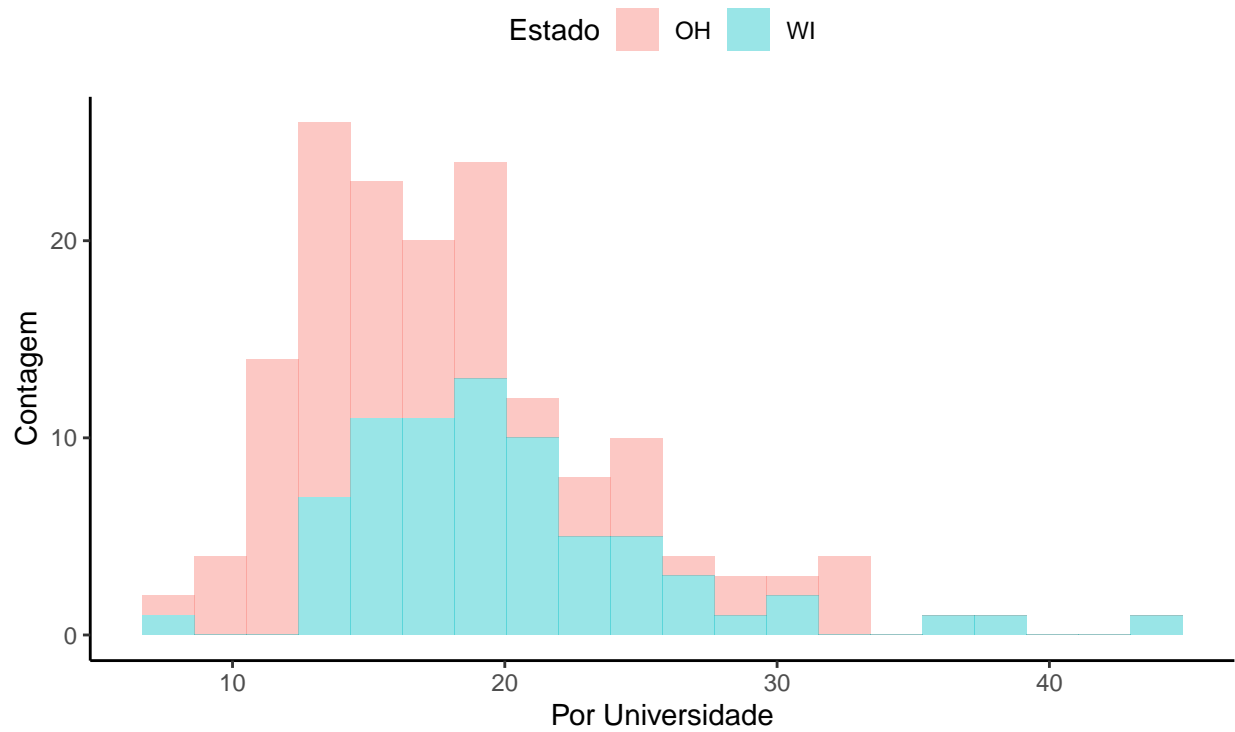


```
ggplot(data = midwest, mapping = aes(x = Área)) +
  geom_histogram(bins = 10, fill = "#004C50") +
  labs(y = "Contagem")
```

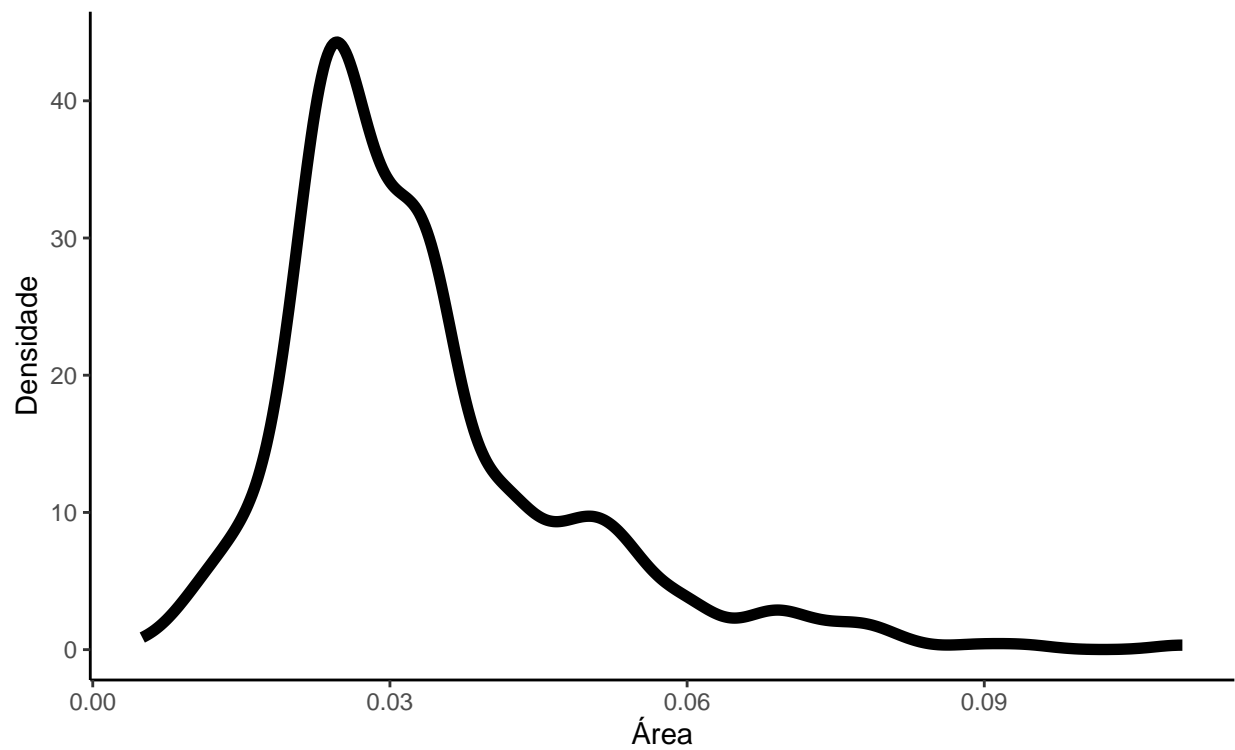


```
oh_wi <- c("OH", "WI")

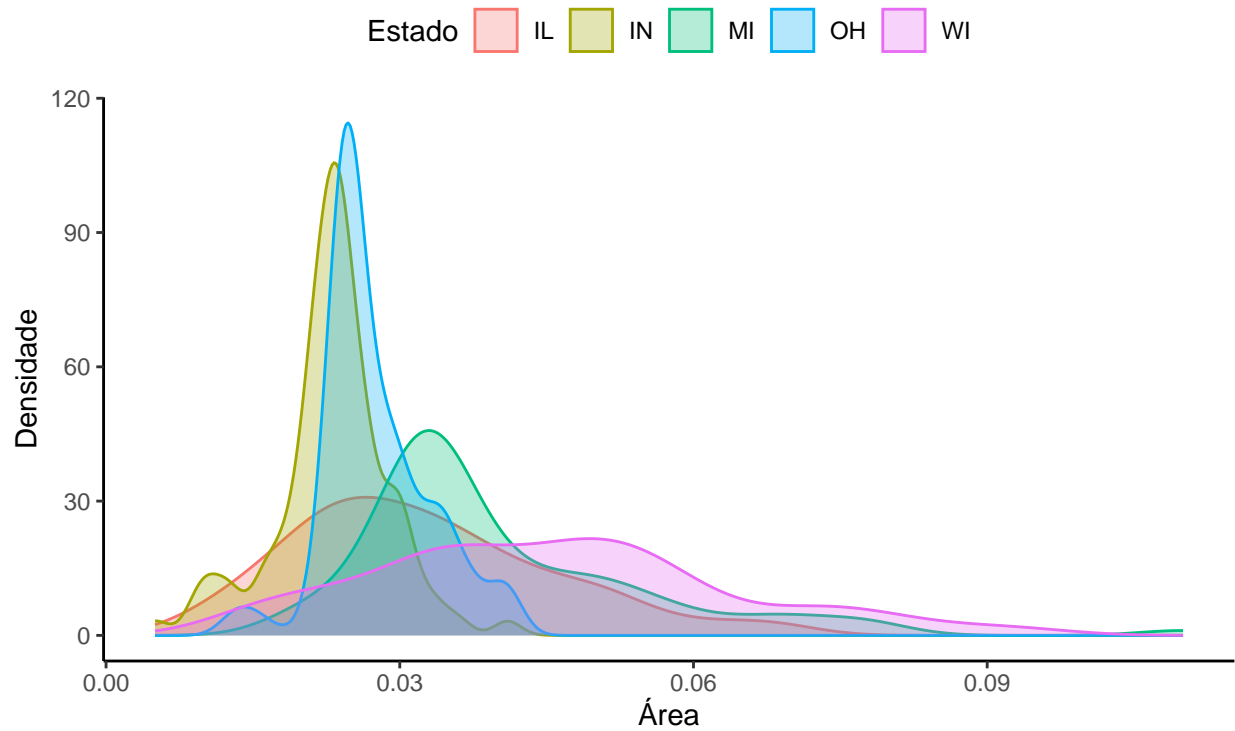
ggplot(data = subset(midwest, Estado %in% oh_wi), mapping = aes(x = percollege, fill = Estado)) +
  geom_histogram(bins = 20, alpha = 0.4) +
  labs(y = "Contagem", x = "Por Universidade")
```



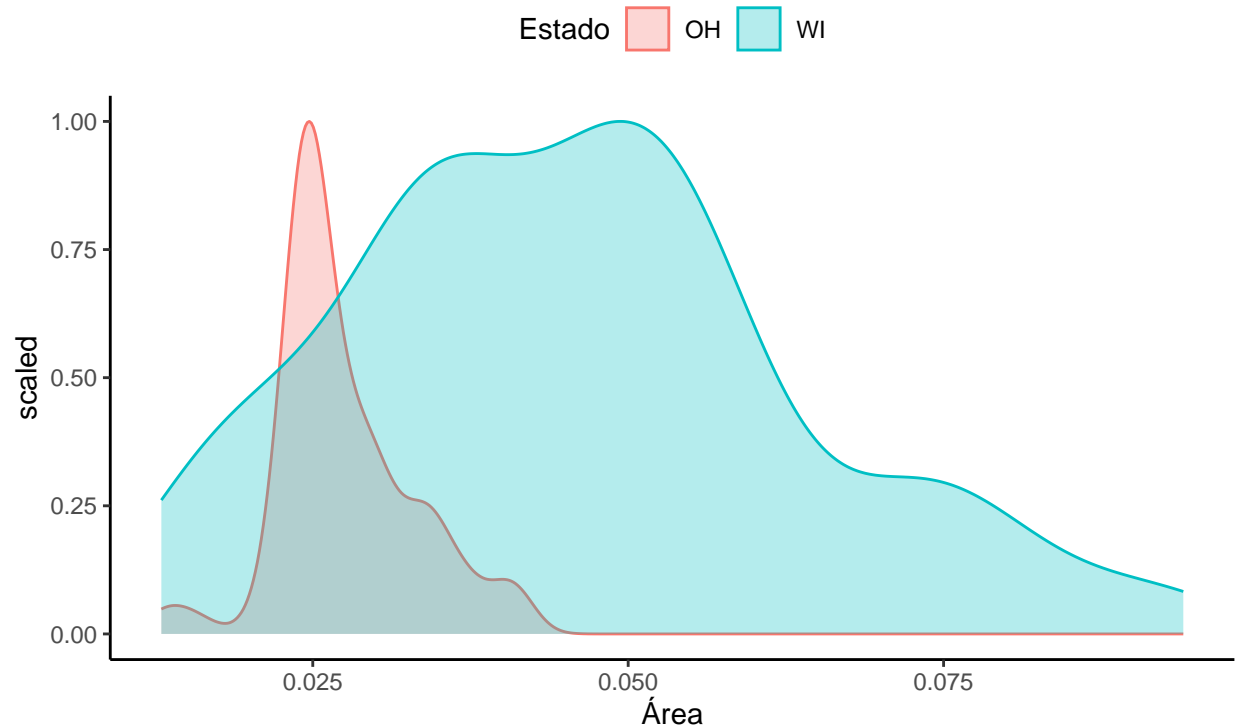
```
ggplot(data = midwest, mapping = aes(x = Área)) +  
  geom_density(size = 2) +  
  labs(y = "Densidade")
```



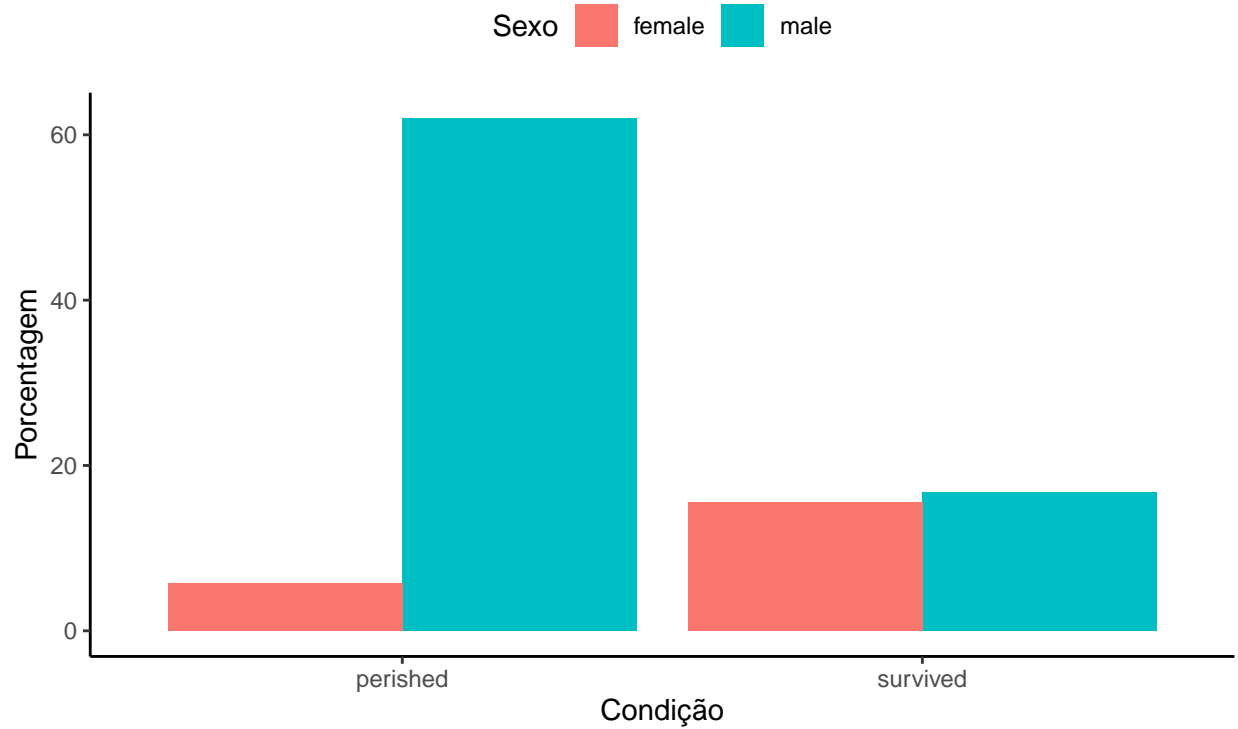
```
ggplot(data = midwest, mapping = aes(x = Área, fill = Estado, color = Estado)) +
  geom_density(alpha = 0.3) +
  labs(y = "Densidade")
```



```
ggplot(data = subset(midwest, Estado %in% oh_wi), mapping = aes(x = Área, fill = Estado, color = Estado)) +
  geom_density(alpha = 0.3, mapping = aes(y = after_stat(scaled)))
```



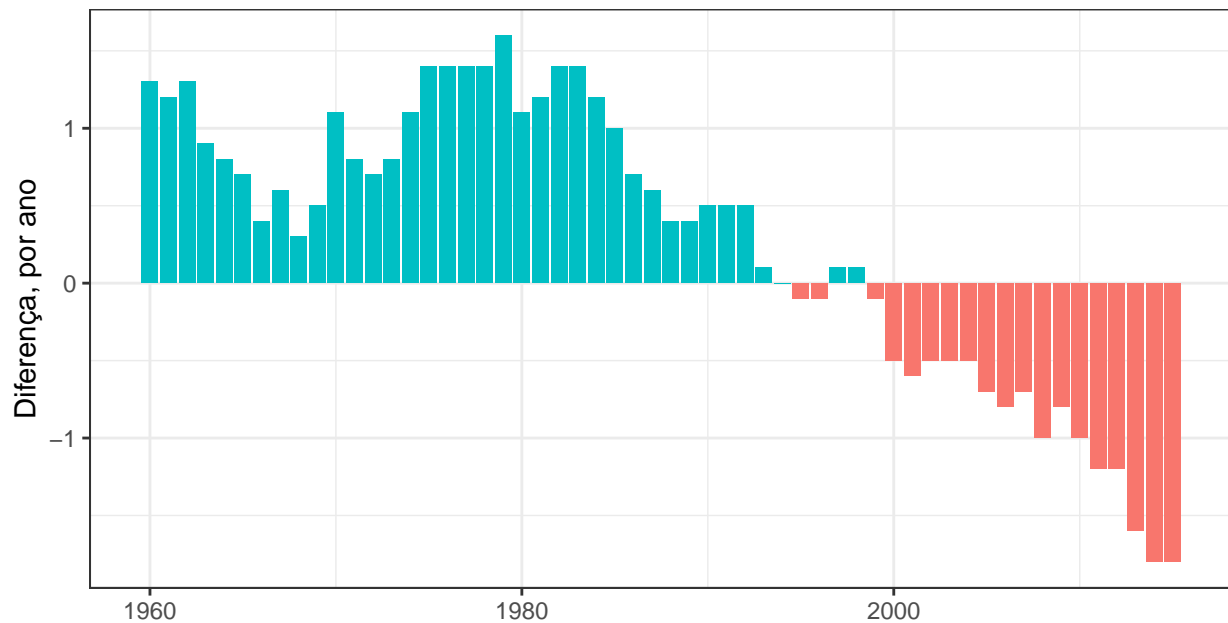
```
ggplot(data = titanic, mapping = aes(x = fate, y = percent, fill = sex)) +
  geom_bar(position = "dodge", stat = "identity") +
  labs(y = "Porcentagem", x = "Condição", fill = "Sexo")
```



```
ggplot(data = oecd_sum, mapping = aes(x = year, y = diff, fill = hi_lo)) +
  geom_col() +
  guides(fill = FALSE) +
  labs(x = NULL,
       y = "Diferença, por ano",
       title = "Diferença na expectativa de Vida nos Estados Unidos",
       subtitle = "Diferença entre a expectativa de vida média dos Estados Unidos e da OECD, 1960-2015",
       caption = "Data: OECD, After a chart por Christopher Ingraham, Washington Post, 27 de Dezembro d
  theme_bw() +
  theme(plot.title = element_text(size = 12, face = "bold"))
```

Diferença na expectativa de Vida nos Estados Unidos

Diferença entre a expectativa de vida média dos Estados Unidos e da OECD, 1960–2015



Data: OECD, After a chart por Christopher Ingraham, Washington Post, 27 de Dezembro de 2017

Gráficos do capítulo 5

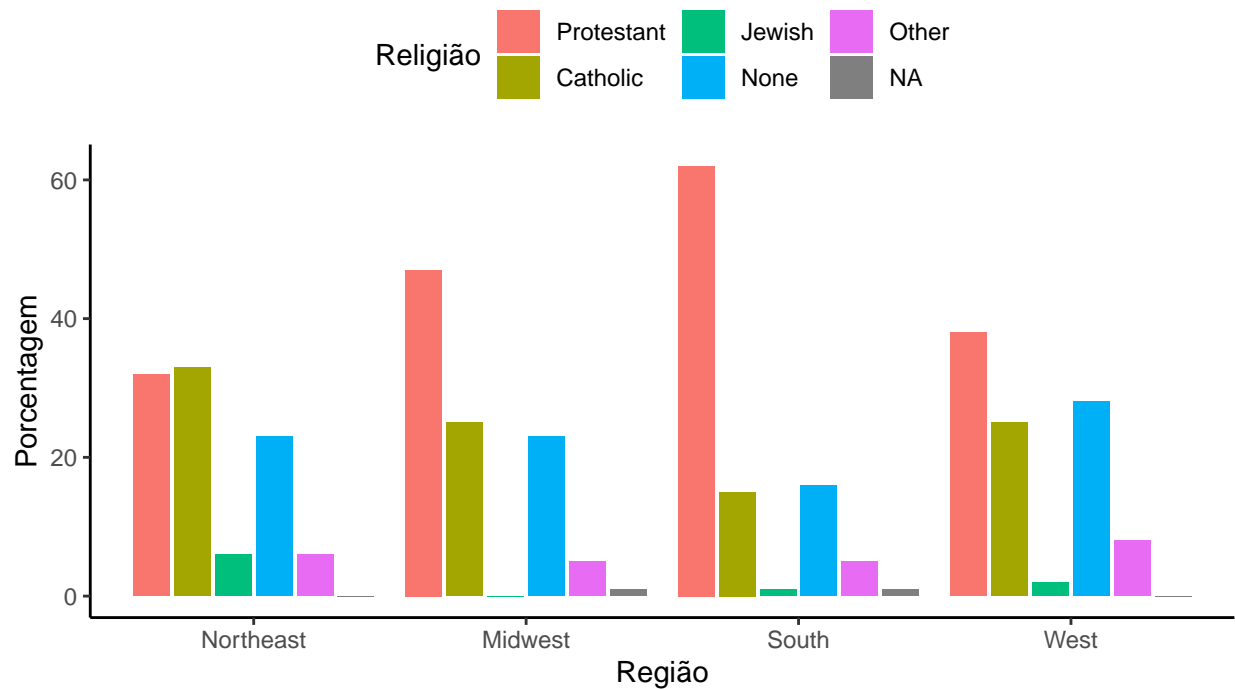
Inicialmente, iremos criar um novo **dataframe** a partir do *gss_sm*. Para isso, usamos as funções `group_by`, `summarize` e `mutate`, conforme pede o livro.

```
rel_by_region <- gss_sm %>%  
  group_by(Região, Religião) %>%  
  summarize(N = n()) %>%  
  mutate(freq = N/sum(N),  
         pct = round((freq*100), 0))
```

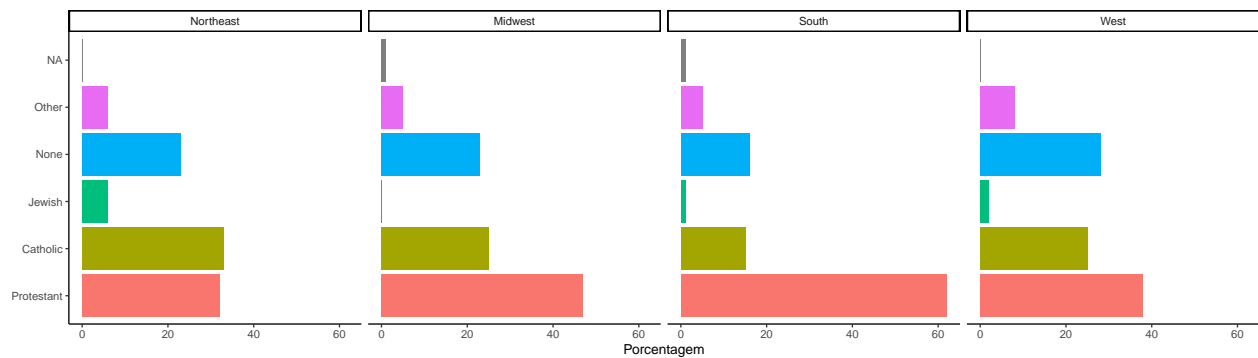
Com o dataframe `rel_by_region` criado, podemos começar a plotar os gráficos.

```
ggplot(rel_by_region, aes(x = Região, y = pct, fill = Religião)) +  
  geom_col(position = "dodge2") +  
  labs(y = "Porcentagem") +  
  ggtitle("Dominancia de cada religião, por região")
```

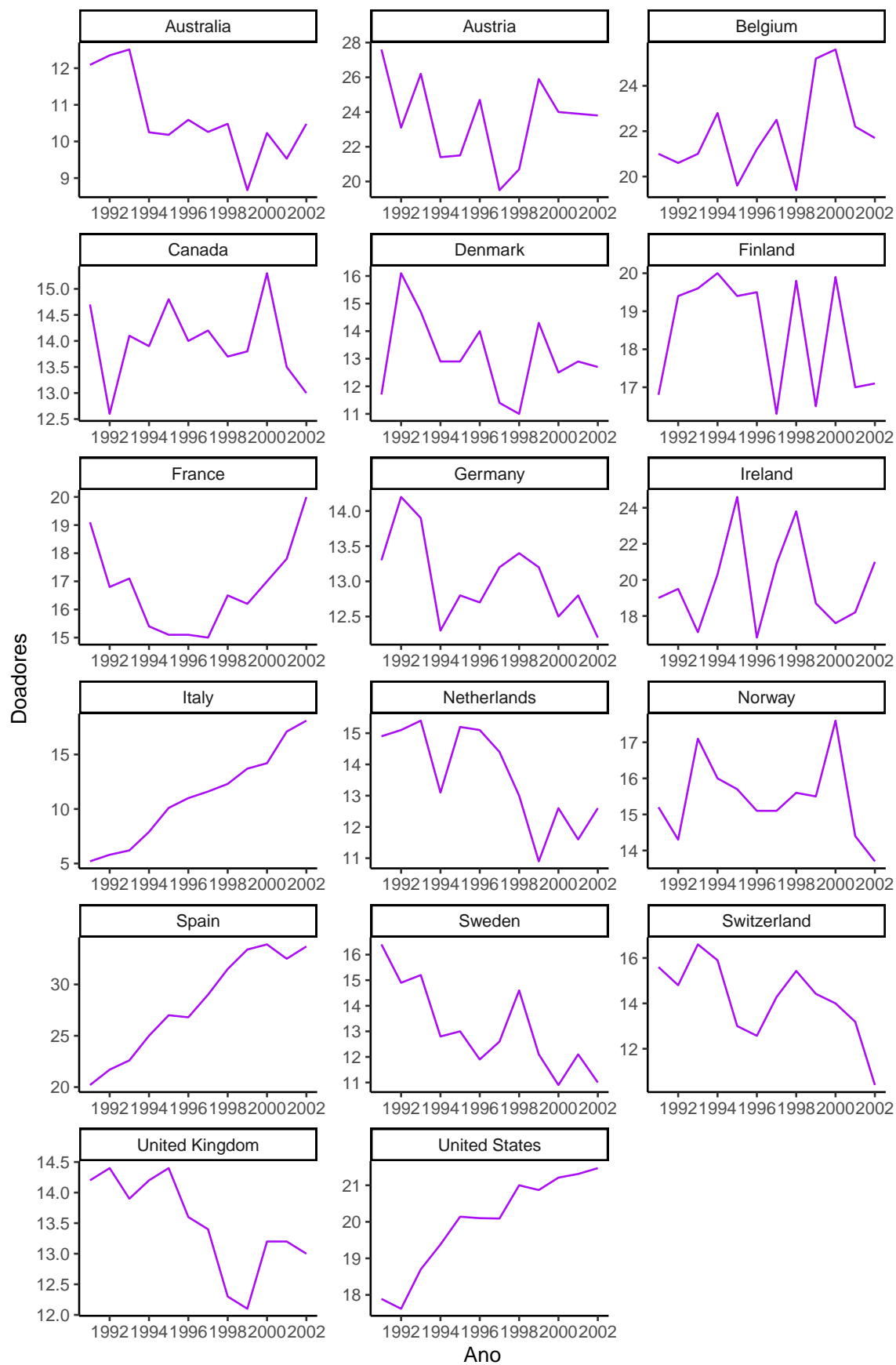
Dominancia de cada religião, por região



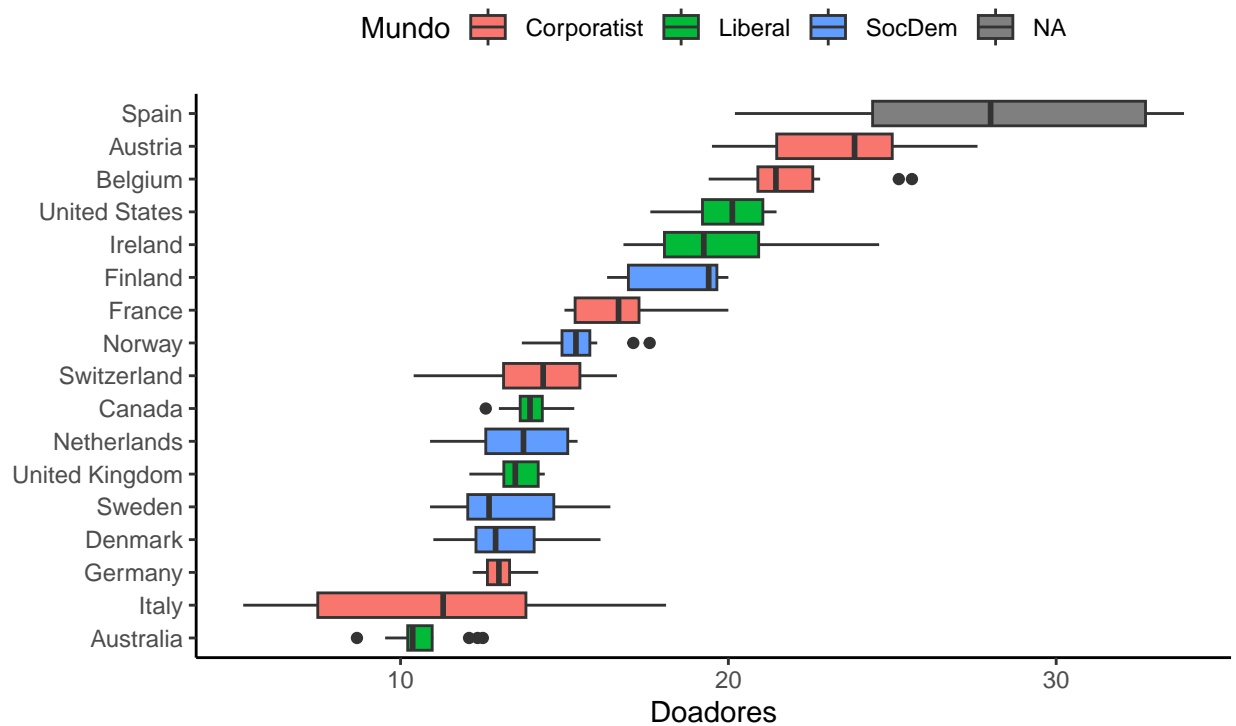
```
ggplot(rel_by_region, aes(x = Religião, y = pct, fill = Religião)) +
  geom_col(position = "dodge2") +
  labs(x = NULL, y = "Porcentagem") +
  guides(fill = FALSE) +
  coord_flip() +
  facet_grid(~Região)
```



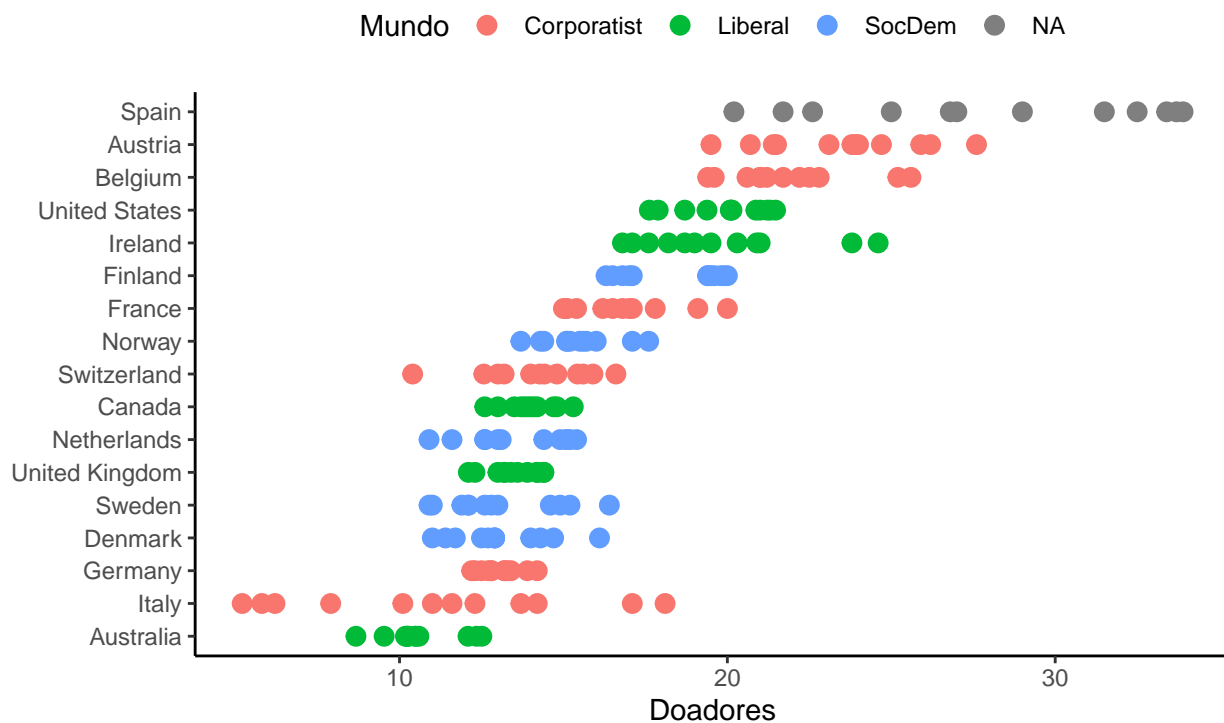
```
ggplot(organdata, mapping = aes(x = Ano, y = Doadores)) +
  geom_line(aes(group = País), color = "#AB0EF5") +
  facet_wrap(~País, ncol = 3, scales = "free")
```




```
ggplot(organdata, mapping = aes(x = reorder(País, Doadores, na.rm = TRUE), y = Doadores, fill = world))
geom_boxplot() +
labs(x = NULL, fill = "Mundo") +
coord_flip()
```



```
ggplot(organdata, mapping = aes(x = reorder(País, Doadores, na.rm = TRUE), y = Doadores, color = world))
geom_point(size = 3) +
labs(x = NULL, color = "Mundo") +
coord_flip()
```

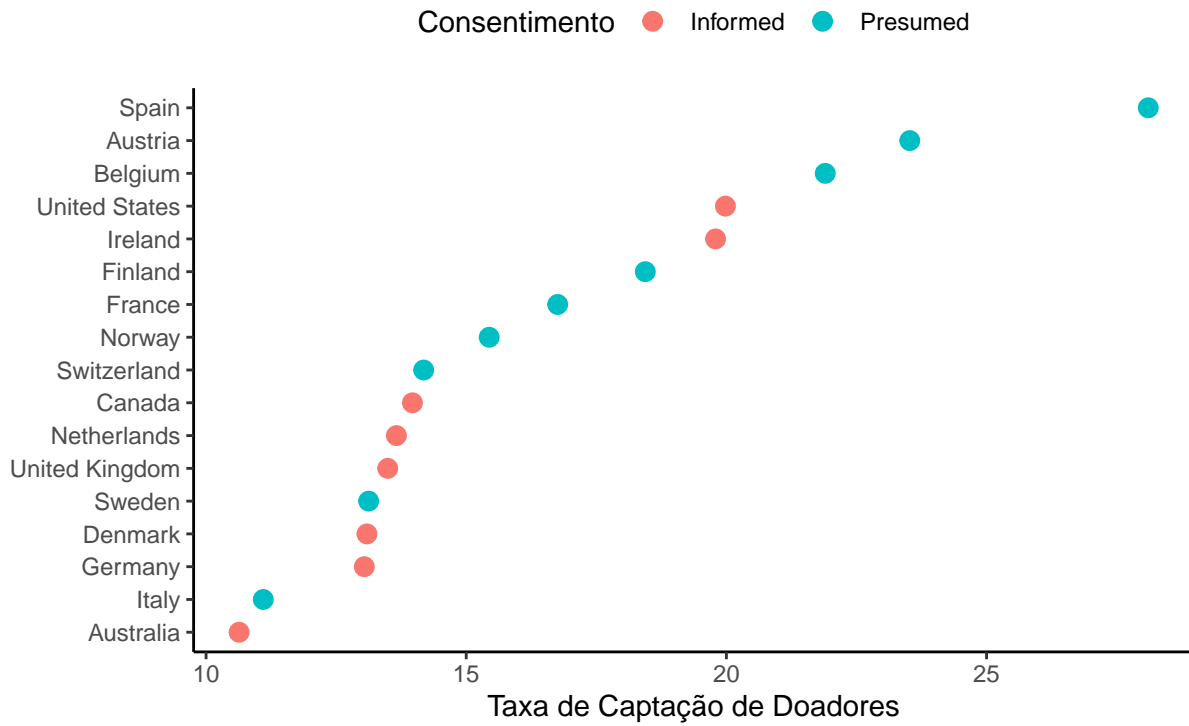


Aqui, criamos outro **dataframe**, dessa vez, a partir do *organdata*.

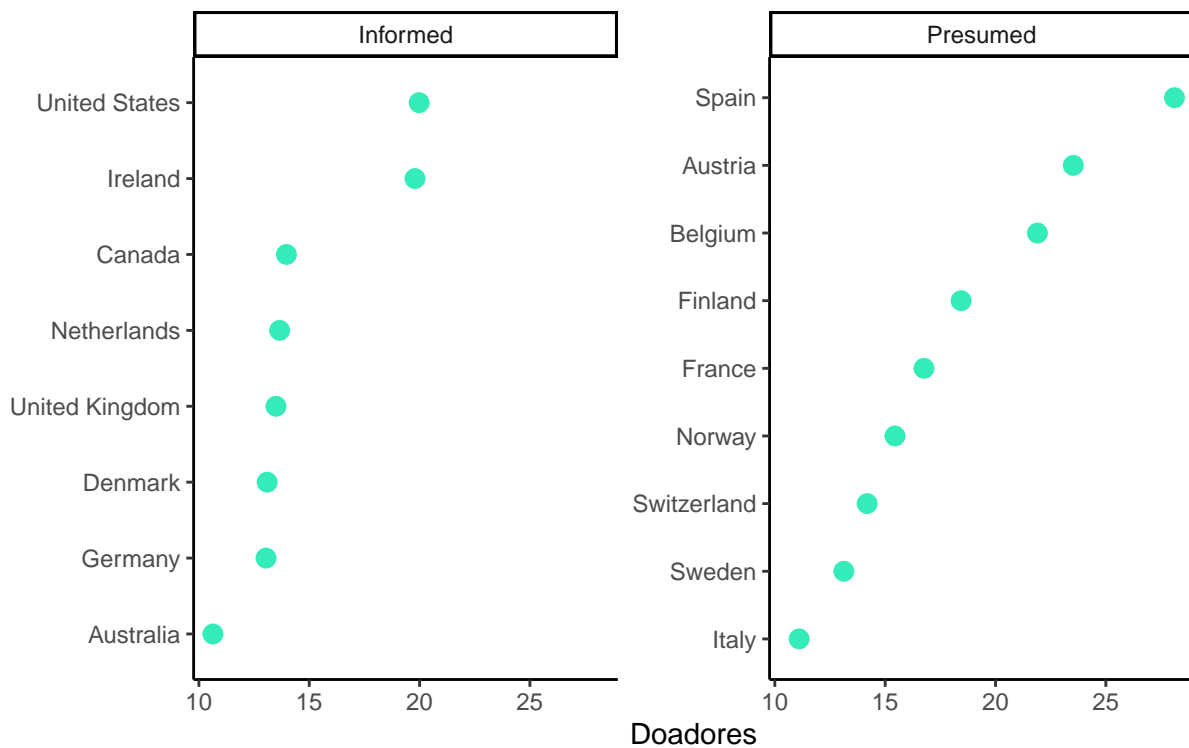
```
by_country <- organdata %>% group_by(consent_law, País) %>%
  summarize_if(is.numeric, funs(mean, sd), na.rm = TRUE) %>%
  ungroup
```

Com isso feito, podemos começar a plotar.

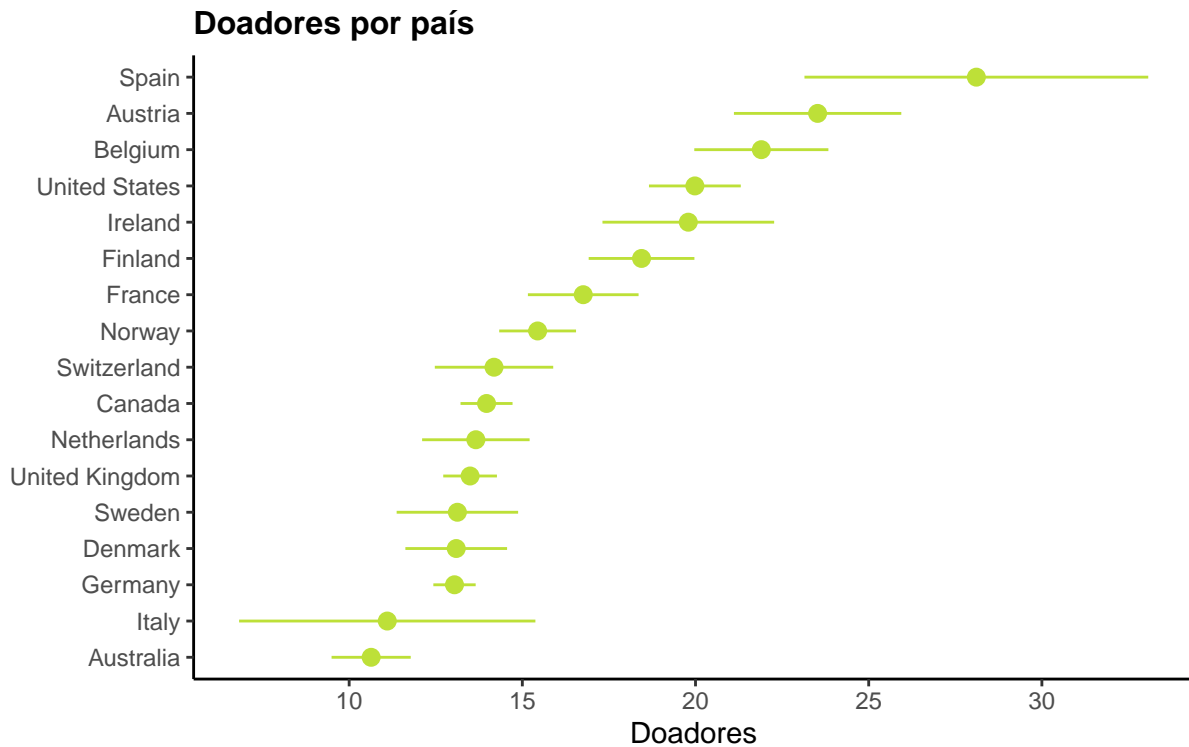
```
ggplot(by_country, mapping = aes(x = Doadores_mean, y = reorder(País, Doadores_mean), color = consent_law)) +
  geom_point(size = 3) +
  labs(x = "Taxa de Captação de Doadores", y = "", color = "Consentimento")
```



```
ggplot(by_country, mapping = aes(x = Doadores_mean, y = reorder(País, Doadores_mean))) +
  geom_point(size = 3, color = "#34EBBA") +
  facet_wrap(~consent_law, scales = "free_y", ncol = 2) +
  labs(x = "Doadores", y = "")
```



```
ggplot(by_country, mapping = aes(x = reorder(País, Doadores_mean), y = Doadores_mean)) +
  geom_pointrange(mapping = aes(ymin = Doadores_mean - Doadores_sd, ymax = Doadores_mean + Doadores_sd),
  labs(x = "", y = "Doadores") +
  coord_flip() +
  theme(legend.position = "none") +
  ggtitle("Doadores por país")
```

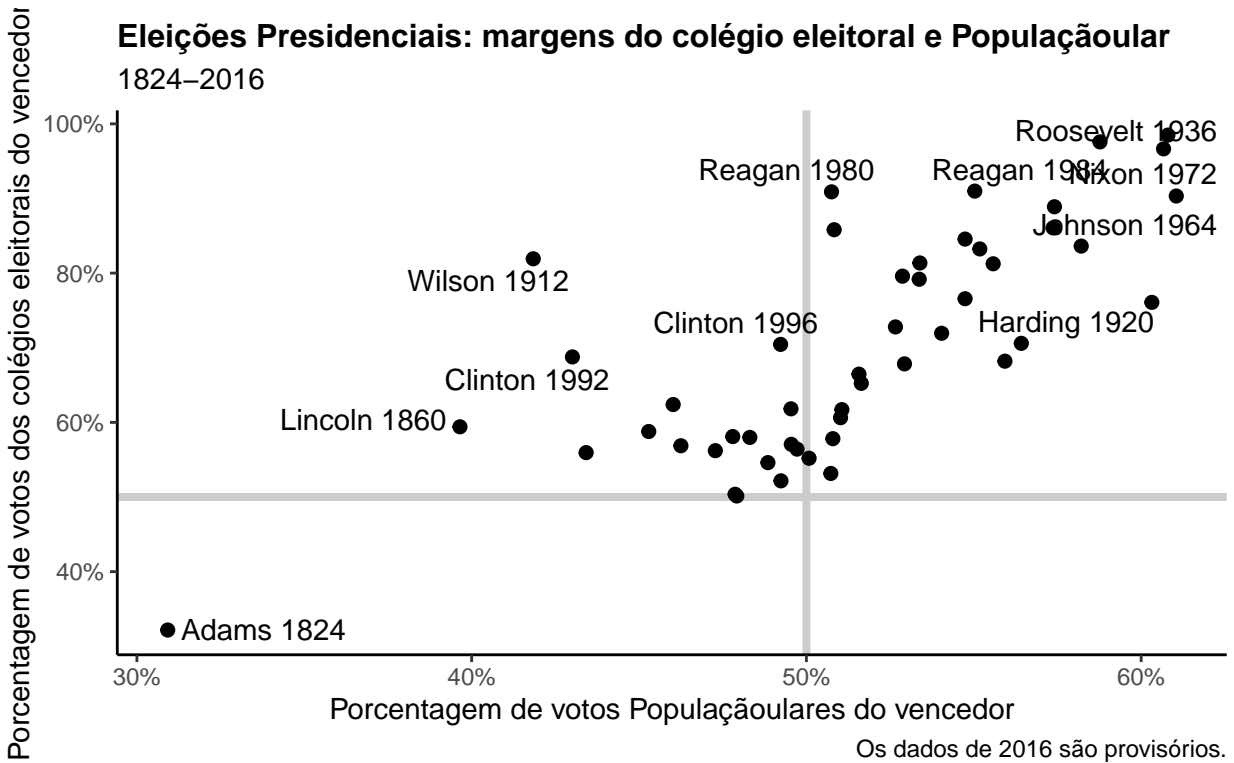


Aqui, usaremos outro métodos para dar nomes no gráfico. Inicialmente, definimos variáveis com os textos que gostaríamos que aparecessem.

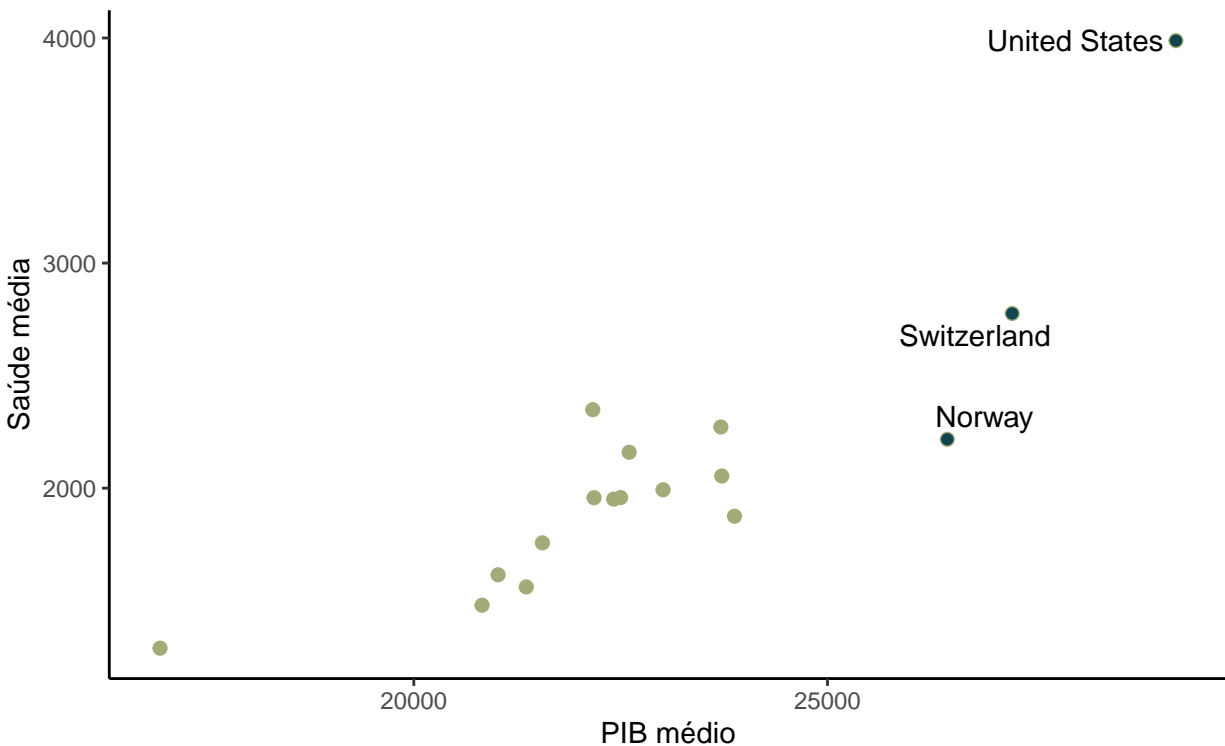
```
p_title <- "Eleições Presidenciais: margens do colégio eleitoral e Populaçãoular"
p_subtitle <- "1824-2016"
p_caption <- "Os dados de 2016 são provisórios."
x_label <- "Porcentagem de votos Populaçãoulares do vencedor"
y_label <- "Porcentagem de votos dos colégios eleitorais do vencedor"
```

Agora, podemos referenciar as variáveis nos **labs**. Isso é importante, pois assim podemos criar gráficos com legendas alteráveis.

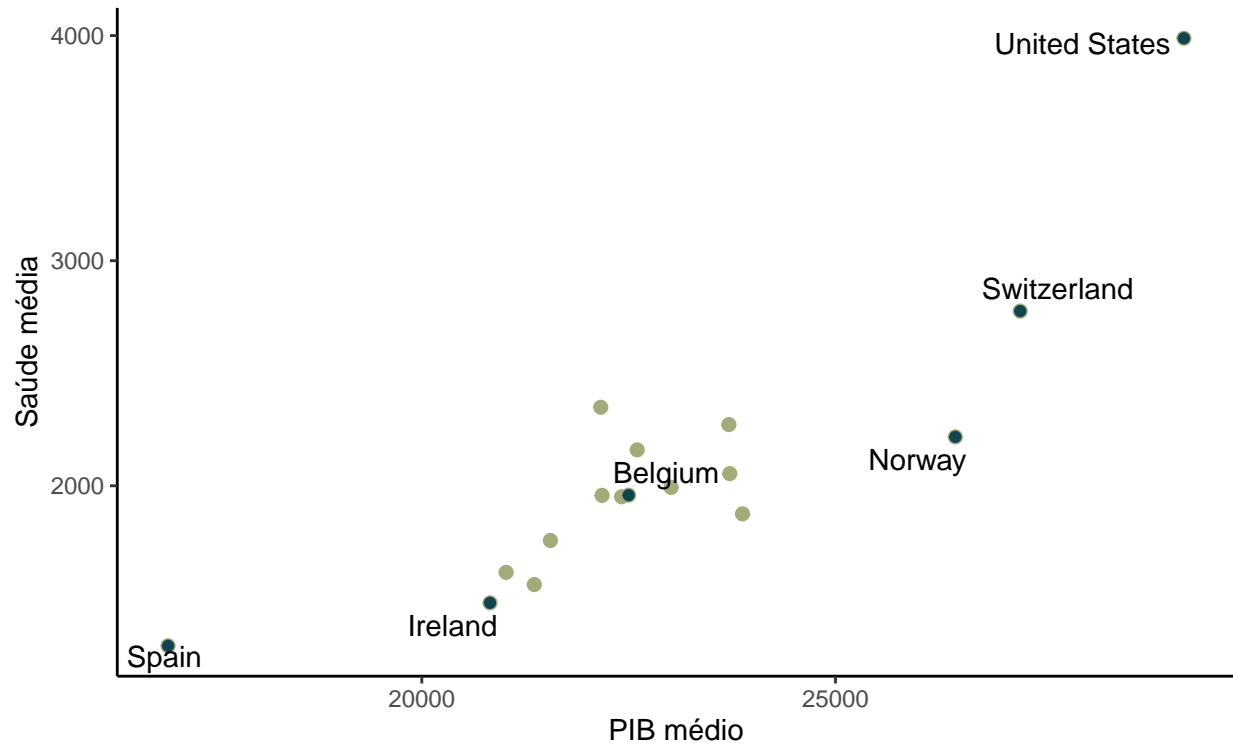
```
ggplot(elections_historic, aes(x = popular_pct, y = ec_pct, label = winner_label)) +
  geom_hline(yintercept = 0.5, size = 1.4, color = "gray80") +
  geom_vline(xintercept = 0.5, size = 1.4, color = "gray80") +
  geom_point(size = 2) +
  geom_text_repel(max.overlaps = 5) +
  scale_x_continuous(labels = scales::percent) +
  scale_y_continuous(labels = scales::percent) +
  labs(x = x_label, y = y_label, title = p_title, subtitle = p_subtitle, caption = p_caption)
```



```
ggplot(by_country, mapping = aes(x = gdp_mean, y = health_mean)) +
  geom_point(size = 2, color = "#A3AB78") +
  geom_point(data = subset(by_country, gdp_mean > 25000), color = "#10454F") +
  geom_text_repel(data = subset(by_country, gdp_mean > 25000), mapping = aes(label = Pa  s)) +
  labs(x = "PIB m  dio", y = "Sa  de m  dia")
```

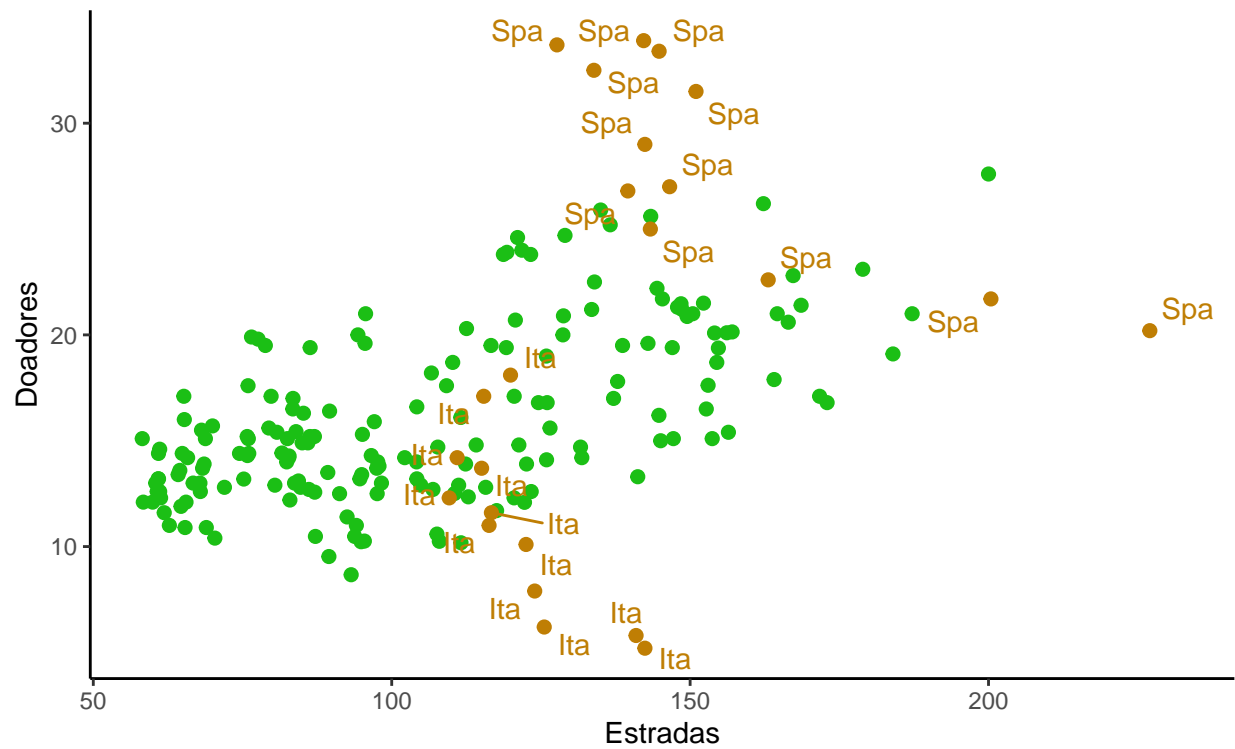


```
ggplot(by_country, mapping = aes(x = gdp_mean, y = health_mean)) +
  geom_point(size = 2, color = "#A3AB78") +
  geom_point(data = subset(by_country, gdp_mean > 25000 | health_mean < 1500 | País %in% "Belgium"), color = "#A3AB78") +
  geom_text_repel(data = subset(by_country, gdp_mean > 25000 | health_mean < 1500 | País %in% "Belgium"),
    labs(x = "PIB médio", y = "Saúde média")
```

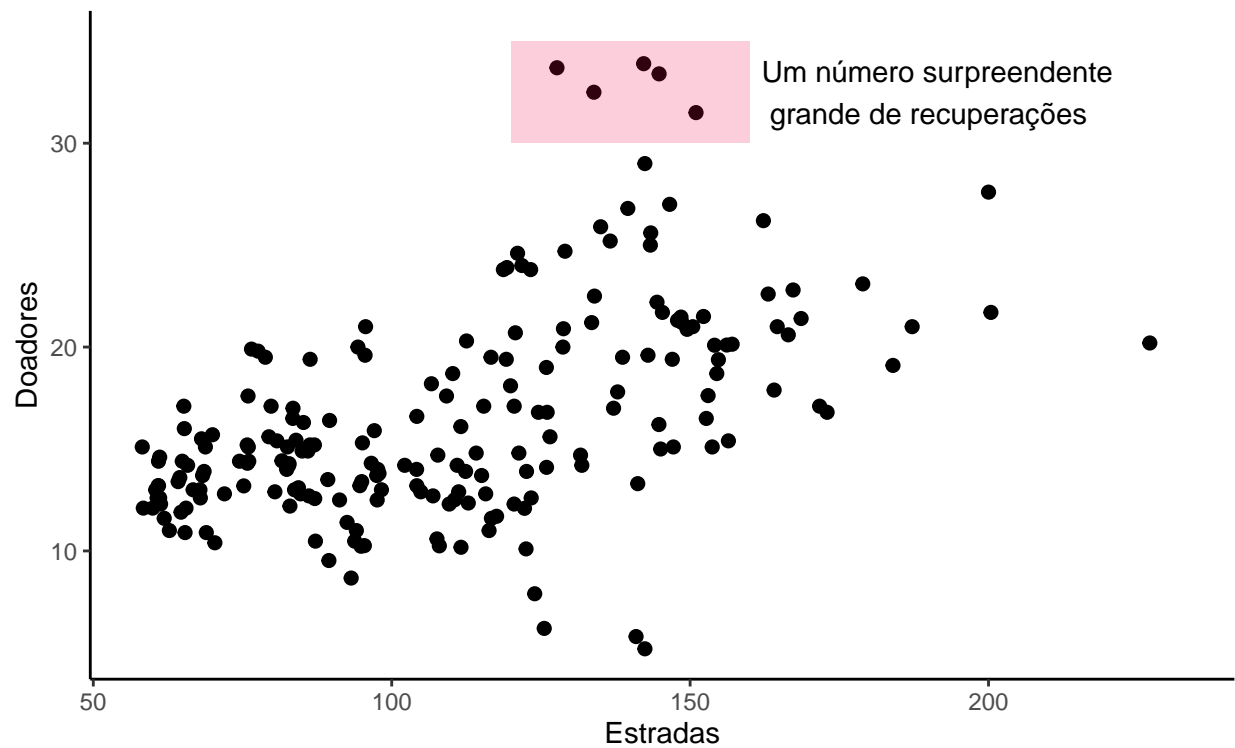


```
organdata$ind <- organdata$ccode %in% c("Ita", "Spa") & organdata$Ano > 1998

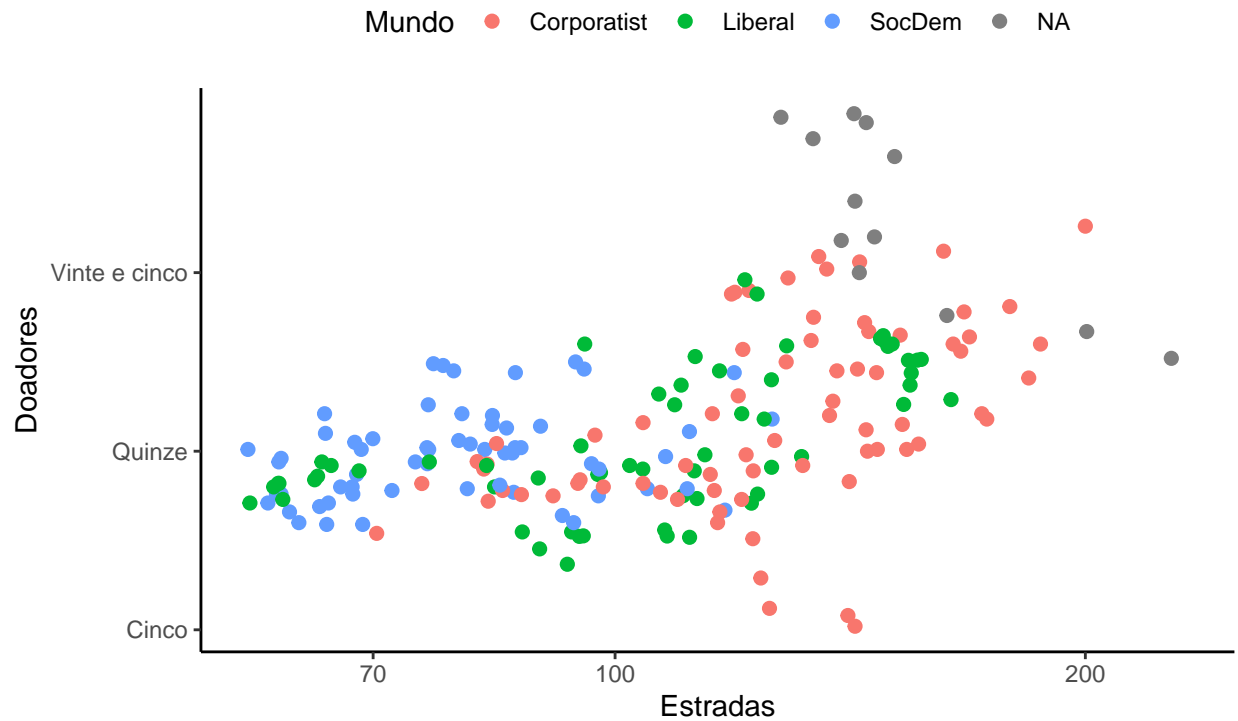
ggplot(organdata, mapping = aes(x = Estradas, y = Doadores, color = ind)) +
  geom_point(size = 2) +
  geom_text_repel(data = subset(organdata, ind), mapping = aes(label = ccode)) +
  guides(label = FALSE, color = FALSE) +
  scale_color_manual(values = c("#1BBF15", "#BF7E04"))
```



```
ggplot(organdata, mapping = aes(x = Estradas, y = Doadores)) +
  geom_point(size = 2) +
  annotate(geom = "rect", xmin = 120, xmax = 160, ymin = 30, ymax = 35, fill = "#F20746", alpha = 0.2) +
  annotate(geom = "text", x = 162, y = 32.5, label = "Um número surpreendente \n grande de recuperações",
```



```
ggplot(organdata, mapping = aes(x = Estradas, y = Doadores, color = world)) +
  geom_point(size = 2) +
  scale_x_log10() +
  scale_y_continuous(breaks = c(5, 15, 25), labels = c("Cinco", "Quinze", "Vinte e cinco")) +
  labs(color = "Mundo")
```



```
ggplot(organdata, mapping = aes(x = Estradas, y = Doadores, color = world)) +
  geom_point(size = 2) +
  labs(x = "Mortes na Estrada", y = "Doadores") +
  guides(color = FALSE) +
  scale_color_manual(values = c("#DFB9AA", "#EBE7CA", "#B9806A", "#575755"))
```