# Lista 2.4

João Victor Mendes

## Item 1

Com “library(socviz)” temos acesso ao dataset das Eleições Presidenciais dos EUA do ano de 2016;

“library(maps)” é utilizado para criar e modificar mapas;

“library(mapproj)” é usado para ajustar a representação do mapa.

“library(ggthemes)” nos permite usar funções com ‘theme’.

```
library(ggplot2)
library(socviz)
library(maps)
library(mapproj)
library(ggthemes)
library(statebins)
library(tidyr)
library(ggrepel)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

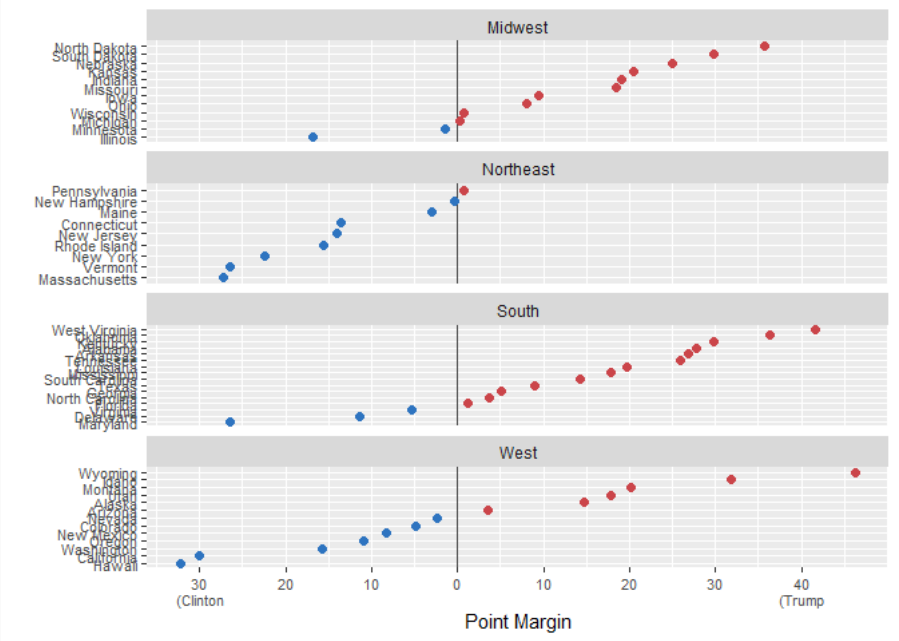
Algumas colunas da base são selecionadas:

```
election %>%
  select(state, total_vote,
         r_points, pct_trump, party, census) %>%
           sample_n(5)
```

| state | total_vote | r_points | pct_trump | party | census |
|---|---|---|---|---|---|
| Connecticut | 1644920 | -13.64 | 40.93 | Democratic | Northeast |
| Minnesota | 2945233 | -1.51 | 44.93 | Democratic | Midwest |
| South Dakota | 370093 | 29.79 | 61.53 | Republican | Midwest |
| Maine | 747927 | -2.96 | 44.87 | Democratic | Northeast |
| California | 14237893 | -29.99 | 31.49 | Democratic | West |

No gráfico inicial, o resultado das eleições é representado por um dotplot.

```
party_colors <- c("#2E74C0", "#CB454A")

p0 <- ggplot(data = subset(election, st %nin% "DC"),
             mapping = aes(x = r_points,
                           y = reorder(state, r_points),
                           color = party))

p1 <- p0 + geom_vline(xintercept = 0, color = "gray30") +
  geom_point(size = 2)

p2 <- p1 + scale_color_manual(values = party_colors)

p3 <- p2 +scale_x_continuous(breaks = c(-30, -20, -10, 0, 10, 20, 30, 40),
                             labels = c("30\n (Clinton", "20", "10", "0", "10", "20", "30",
         "40\n(Trump"))

p3 + facet_wrap(~census, ncol=1, scales = "free_y") +
  guides(color=FALSE) + labs(x = "Point Margin", y = "") +
  theme(axis.text= element_text(size=8))
```




```
us_states <- map_data("state")
head(us_states)
```

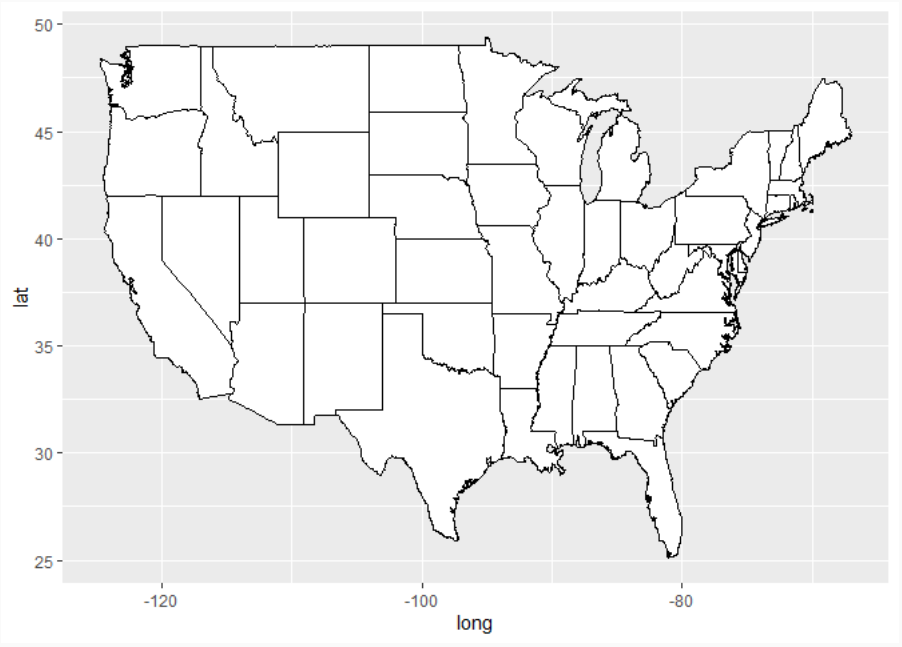
| long | lat | group | order | region | subregion |
|---|---|---|---|---|---|
| -87.46201 | 30.38968 | 1 | 1 | alabama | NA |
| -87.48493 | 30.37249 | 1 | 2 | alabama | NA |
| -87.52503 | 30.37249 | 1 | 3 | alabama | NA |
| -87.53076 | 30.33239 | 1 | 4 | alabama | NA |
| -87.57087 | 30.32665 | 1 | 5 | alabama | NA |
| -87.58806 | 30.32665 | 1 | 6 | alabama | NA |

```
dim(us_states)
```

```
## [1] 15537     6
```

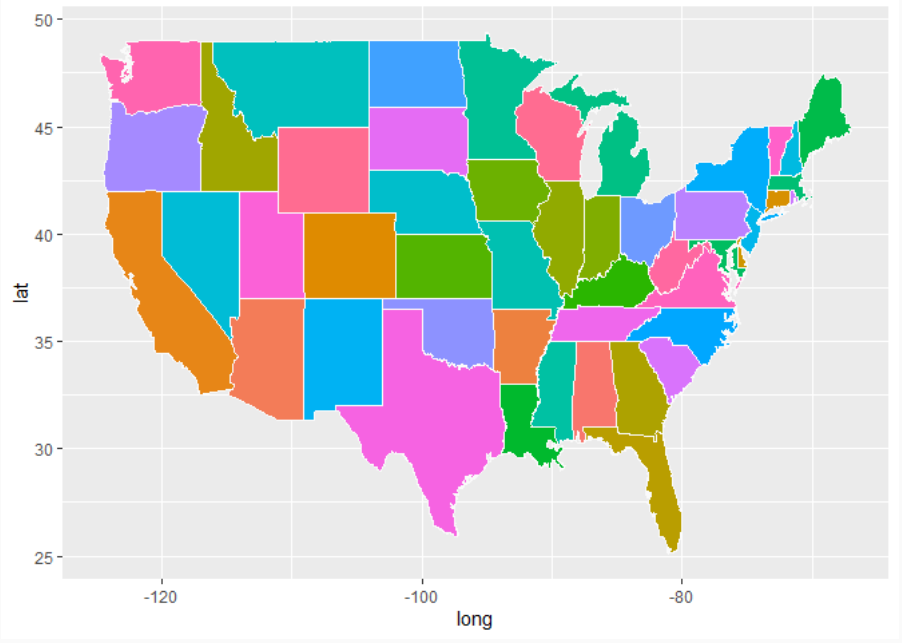
Criado o mapa dos EUA, ainda sem maiores informações além de latitude e longitude;

```
p <- ggplot(data = us_states, mapping = aes(x = long, y = lat, group = group))

p + geom_polygon(fill = "white", color = "black")
```




Adicionando cores aleatórias aos estados e definindo as divisórias como cinza;

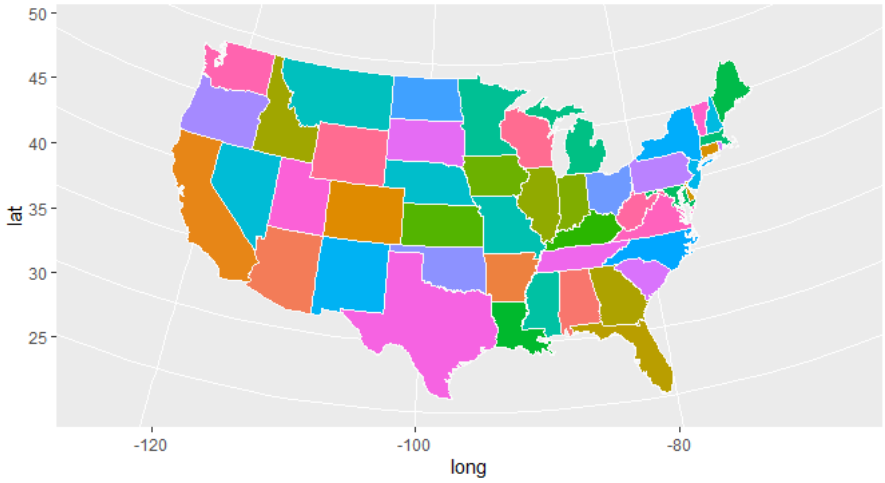
```
p<- ggplot(data = us_states, aes(x = long, y = lat, group = group, fill=region))

p + geom_polygon(color = "gray98", size = 0.1) +
  guides(fill = FALSE)
```




Alternando o formato do mapa para uma representação ajustada em estilo de globo;

```
p <- ggplot(data = us_states,
            mapping = aes(x = long, y = lat,
                          group = group, fill = region))
```

```
p + geom_polygon(color = "gray98", size = 0.1) +
  coord_map(projection = "albers", lat0=39, lat1=45) +
  guides(fill=FALSE)
```

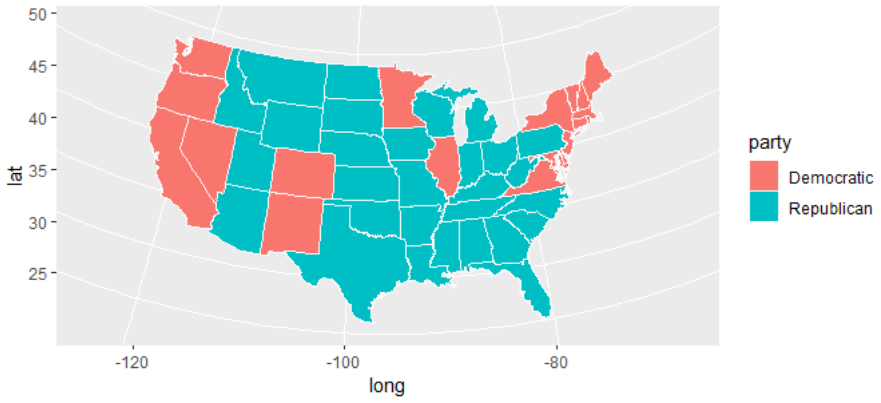



```
election$region <- tolower(election$state)
us_states_elec <- left_join(us_states, election)
```

```
## Joining, by = "region"
```

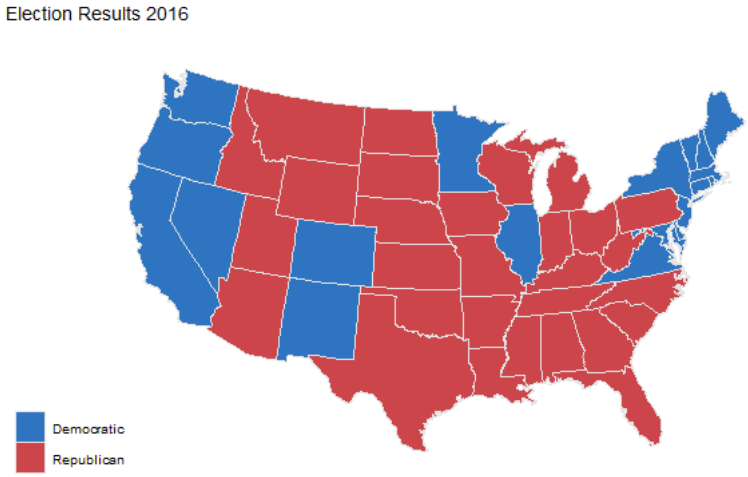
Mapa dividido entre estados que os Republicanos venceram e estados que os Democratas venceram, porém com a cor típica dos partidos estando invertida.

```
p <- ggplot(data = us_states_elec,
            aes(x = long, y = lat,
                group = group, fill = party))

p + geom_polygon(color = "gray98", size = 0.1) +
  coord_map(projection = "albers", lat0 = 39, lat1=45)
```




Cores definidas da maneira correta e título adicionado

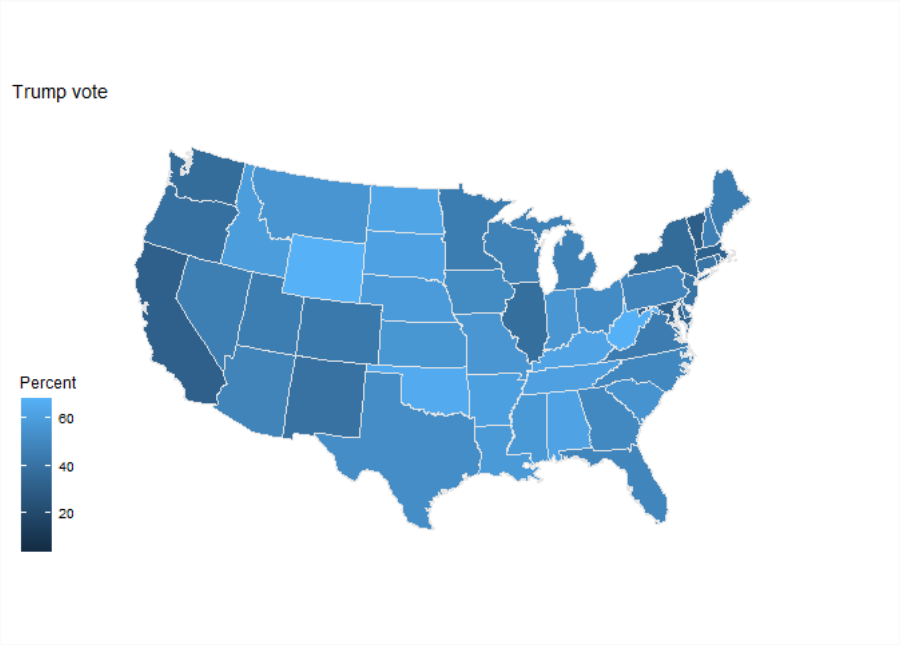
```
p0 <- ggplot(data = us_states_elec,
             mapping= aes(x = long, y = lat,
                          group = group, fill = party))

p1 <- p0 + geom_polygon(color = "gray90", size = 0.1) +
  coord_map(projection="albers", lat0 = 39, lat1=45)

p2 <- p1 + scale_fill_manual(values = party_colors) +
  labs(title = "Election Results 2016", fill = NULL)

p2 + theme_map()
```



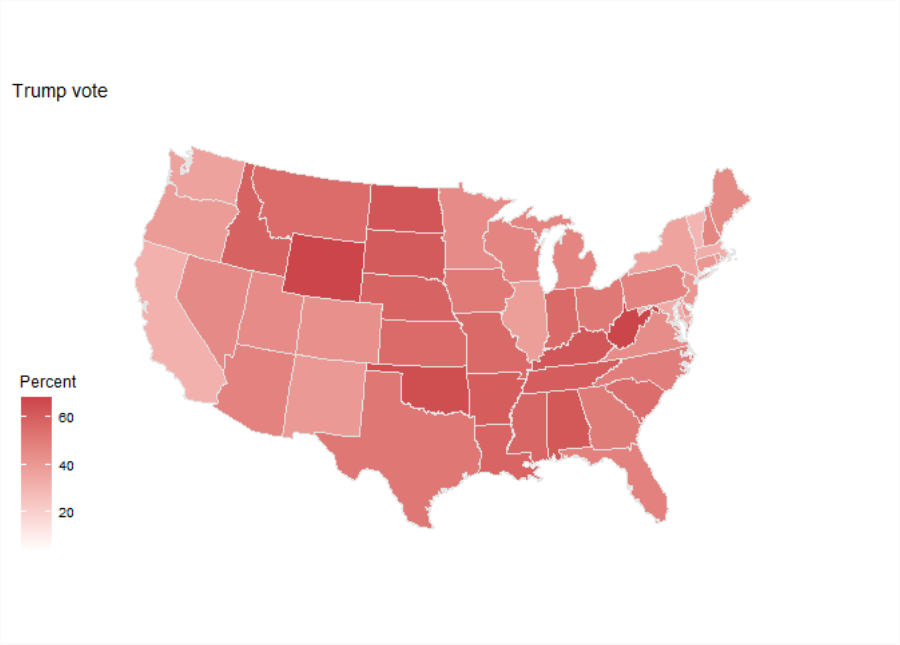

O primeiro gráfico nos entrega uma leitura confusa, representa os votos de trump pela cor dos Democratas e a maior quantidade de votos pela cor mais clara; o segundo gráfico ajusta a representação.

```
p0 <- ggplot(data = us_states_elec,
             mapping = aes(x = long, y = lat, group = group, fill = pct_trump))

p1 <- p0 + geom_polygon(color = "gray90", size = 0.1) +
  coord_map(projection = "albers", lat0 = 39, lat1=45)
```

```
p1 + labs(title = "Trump vote") + theme_map() + labs(fill = "Percent")
```

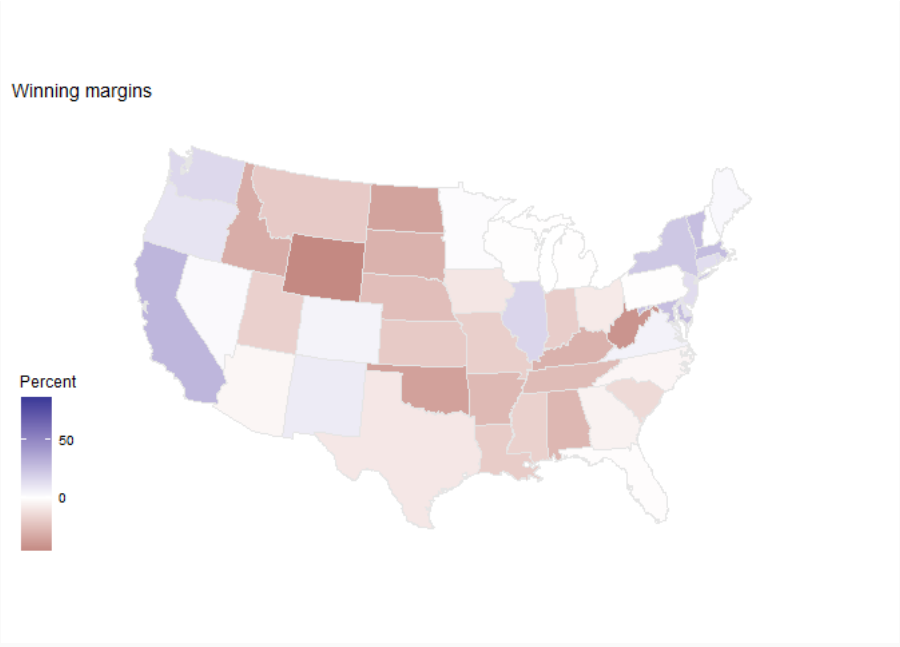



```
p2 <- p1 + scale_fill_gradient(low="white", high="#CB454A") +
  labs(title = "Trump vote")
p2 + theme_map() + labs(fill = "Percent")
```



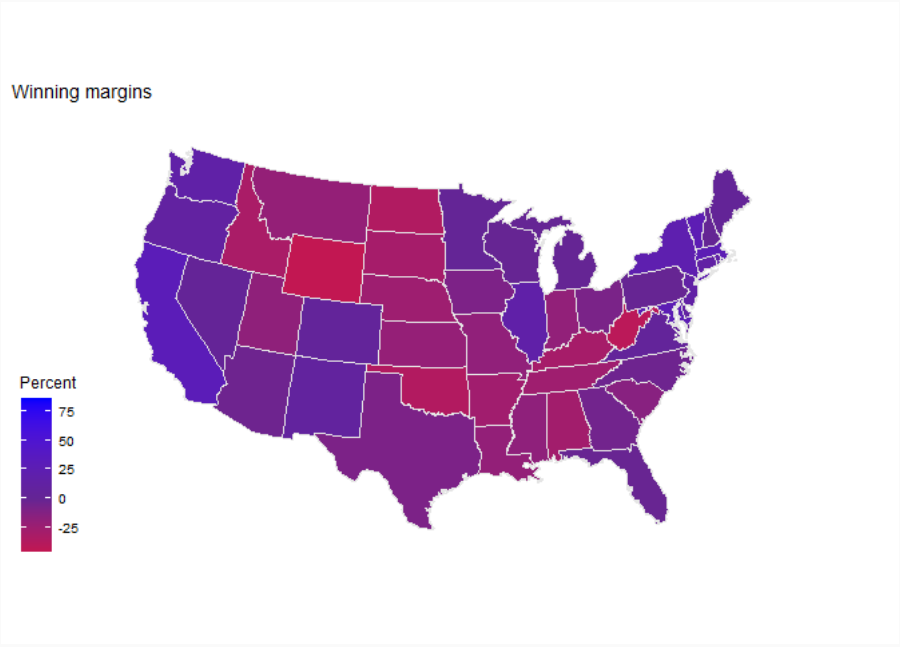

Quanto mais votos para os Democratas mais azul fica o mapa, para os Republicanos, mais vermelho, e quanto mais 50/50, mais branco o mapa;

```
p0 <- ggplot(data = us_states_elec,
             mapping = aes(x=long, y = lat, group = group, fill = d_points))

p1 <- p0 + geom_polygon(color = "gray90", size = 0.1) +
  coord_map(projection = "albers", lat0=39, lat1=45)

p2 <- p1 + scale_fill_gradient2() + labs(title = "Winning margins")
p2 + theme_map() + labs(fill = "Percent")
```



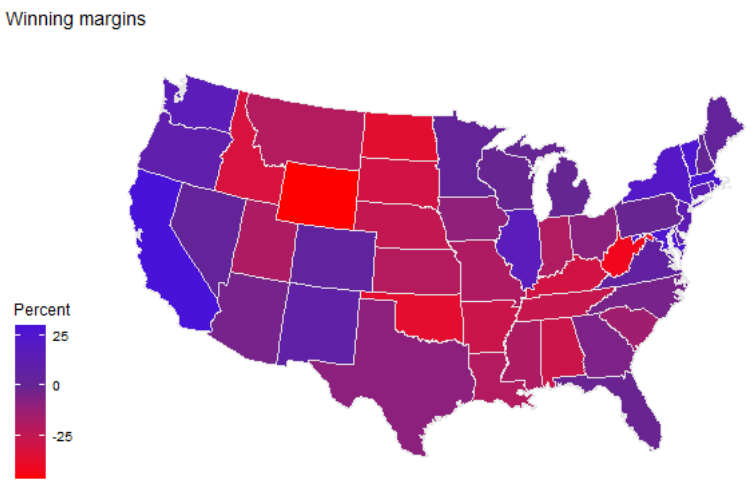

Mesma coisa do interior, porém ao invés de branco é utilizado para os “empates” misturas de azul e vermelho;

```
p3 <- p1+scale_fill_gradient2(low = "red", mid = scales::muted("purple"),
                              high = "blue", breaks = c(-25, 0,25,50,75)) +
  labs(title = "Winning margins")

p3 + theme_map() + labs(fill = "Percent")
```




Porcentagem de diferenças agora é de 25% para cada lado, levando a melhor visualização;

```
p0 <- ggplot(data = subset(us_states_elec,
                           region %nin% "district of columbia"),
             aes(x = long, y = lat, group = group, fill = d_points))

p1 <- p0 + geom_polygon(color = "gray90", size = 0.1) +
  coord_map(projection = "albers", lat0=39, lat1=45)

p2 <- p1 + scale_fill_gradient2(low = "red",
                                mid = scales::muted("purple"),
                                high="blue") +
  labs(title = "Winning margins")
p2 + theme_map() + labs(fill = "Percent")
```




```
county_map %>%
  sample_n(5)
```

| long | lat | order | hole | piece | group | id |
|---|---|---|---|---|---|---|
| -1419601 | -1769243.6 | 13013 | FALSE | 1 | 0500000US02188.1 | 02188 |
| 1097808 | 87669.0 | 88224 | FALSE | 1 | 0500000US26089.1 | 26089 |
| 1788276 | -498854.8 | 183856 | FALSE | 1 | 0500000US54071.1 | 54071 |
| 1046732 | -394546.0 | 52692 | FALSE | 1 | 0500000US17075.1 | 17075 |
| -1427972 | -1947419.9 | 11505 | FALSE | 1 | 0500000US02180.1 | 02180 |

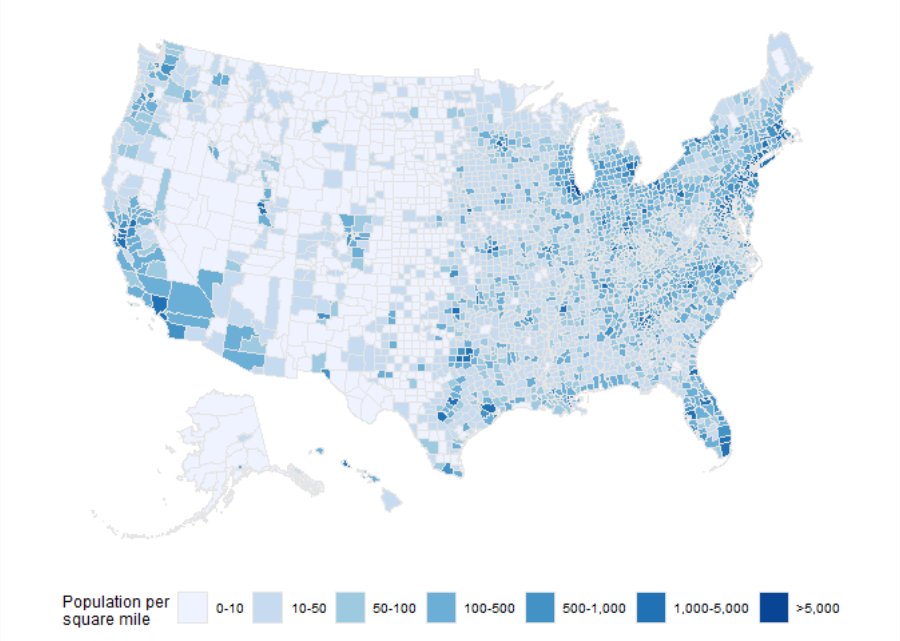
```
county_data %>%
  select(id, name, state, pop_dens, pct_black) %>%
  sample_n(5)
```

| id | name | state | pop_dens | pct_black |
|---|---|---|---|---|
| 21229 | Washington County | KY | [ 10, 50) | [ 5.0,10.0) |
| 51720 | Norton city | VA | [ 500, 1000) | [ 5.0,10.0) |
| 46073 | Jerauld County | SD | [ 0, 10) | [ 0.0, 2.0) |
| 26045 | Eaton County | MI | [ 100, 500) | [ 5.0,10.0) |
| 17151 | Pope County | IL | [ 10, 50) | [ 5.0,10.0) |

Densidade da população:

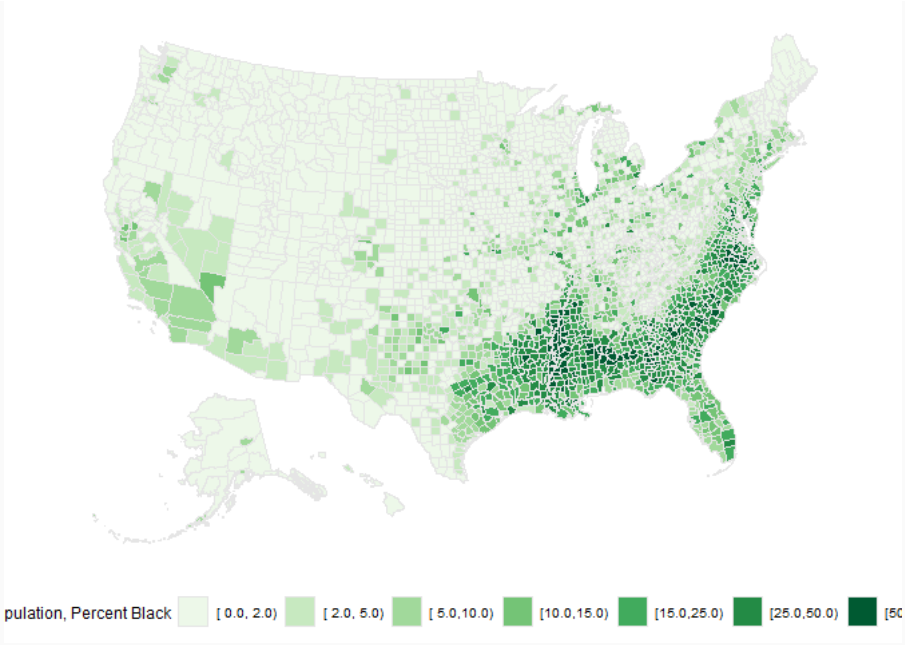
```
county_full <- left_join(county_map, county_data, by = "id")


p <- ggplot(data = county_full,
            mapping = aes(x = long, y = lat,
                          fill = pop_dens,
                          group = group))
```

```
p1 <- p + geom_polygon(color = "gray90", size = 0.05) + coord_equal()

p2 <- p1 + scale_fill_brewer(palette="Blues",
                             labels = c("0-10", "10-50", "50-100", "100-500", "500-1,000",
          "1,000-5,000", ">5,000"))

p2 + labs(fill = "Population per\nsquare mile") +
  theme_map() +
  guides(fill = guide_legend(nrow = 1))+
  theme(legend.position = "bottom")
```




Densidade da população negra:

```
p <- ggplot(data = county_full,
            mapping = aes(x = long, y = lat, fill = pct_black,
                          group = group))

p1 <- p + geom_polygon(color = "gray90", size = 0.05) + coord_equal()

p2 <- p1 + scale_fill_brewer(palette="Greens")

p2 + labs(fill = "US Population, Percent Black") + guides(fill = guide_legend(nrow=1)) +
  theme_map() + theme(legend.position = "bottom")
```




```
orange_pal <- RColorBrewer::brewer.pal(n = 6, name = "Oranges")
orange_pal
```

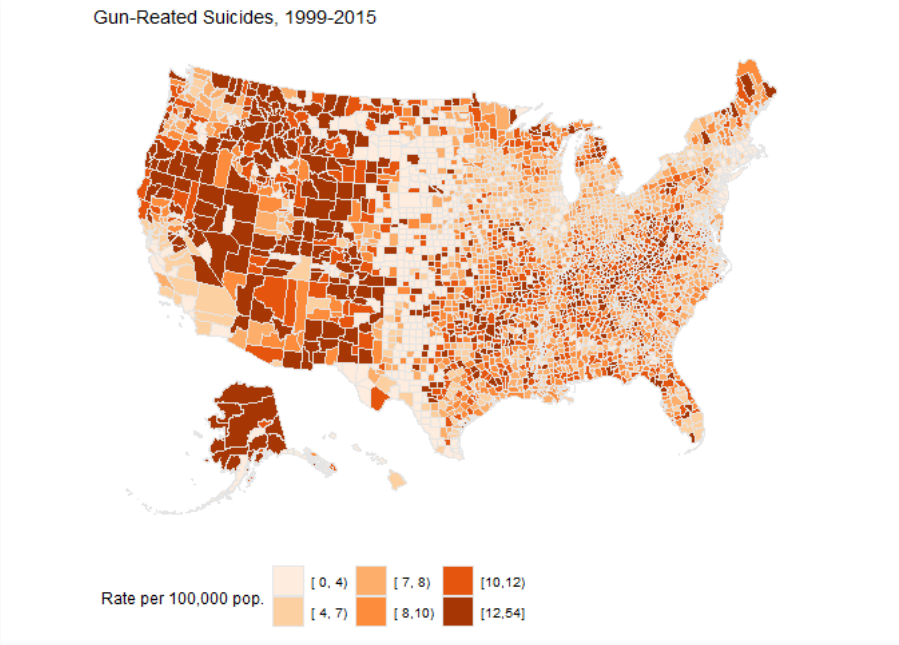
```
## [1] "#FEEDDE" "#FDD0A2" "#FDAE6B" "#FD8D3C" "#E6550D" "#A63603"
```

```
orange_rev <- rev(orange_pal)
orange_rev
```

```
## [1] "#A63603" "#E6550D" "#FD8D3C" "#FDAE6B" "#FDD0A2" "#FEEDDE"
```

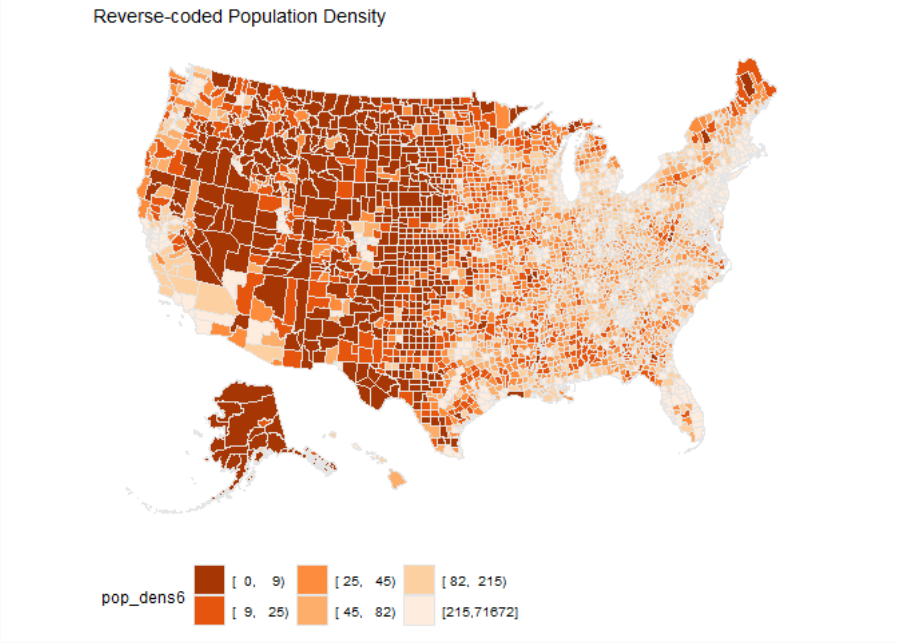
Densidade de suícidios com utilização de armas:

```
gun_p <- ggplot(data = county_full,
                mapping = aes(x = long, y = lat,
                              fill = su_gun6,
                              group = group))

gun_p1 <- gun_p + geom_polygon(color="gray90", size = 0.05) + coord_equal()

gun_p2 <- gun_p1 + scale_fill_manual(values=orange_pal)

gun_p2 + labs(title = "Gun-Reated Suicides, 1999-2015",
              fill = "Rate per 100,000 pop.") +
  theme_map() + theme(legend.position = "bottom")
```



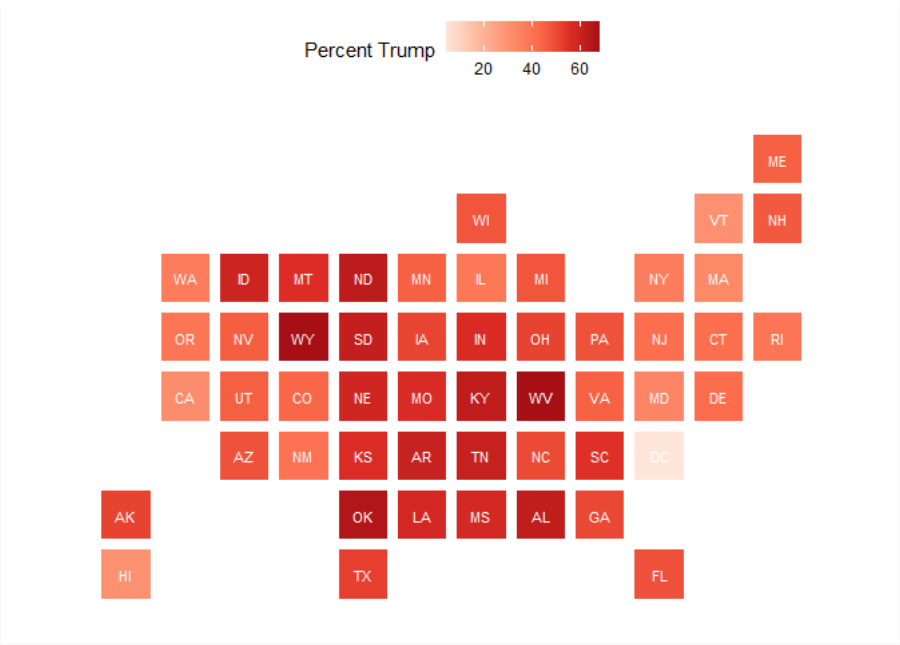

Densidade da população (quanto mais claro mais densa):

```
pop_p <- ggplot(data = county_full, mapping = aes(x = long, y = lat,
                                                  fill = pop_dens6,
                                                  group=group))

pop_p1 <- pop_p + geom_polygon(color = "gray90", size = 0.05) + coord_equal()

pop_p2 <- pop_p1 + scale_fill_manual(values=orange_rev)

pop_p2 + labs(title = "Reverse-coded Population Density",
              fll = "People per square mile") +
  theme_map() + theme(legend.position = "bottom")
```

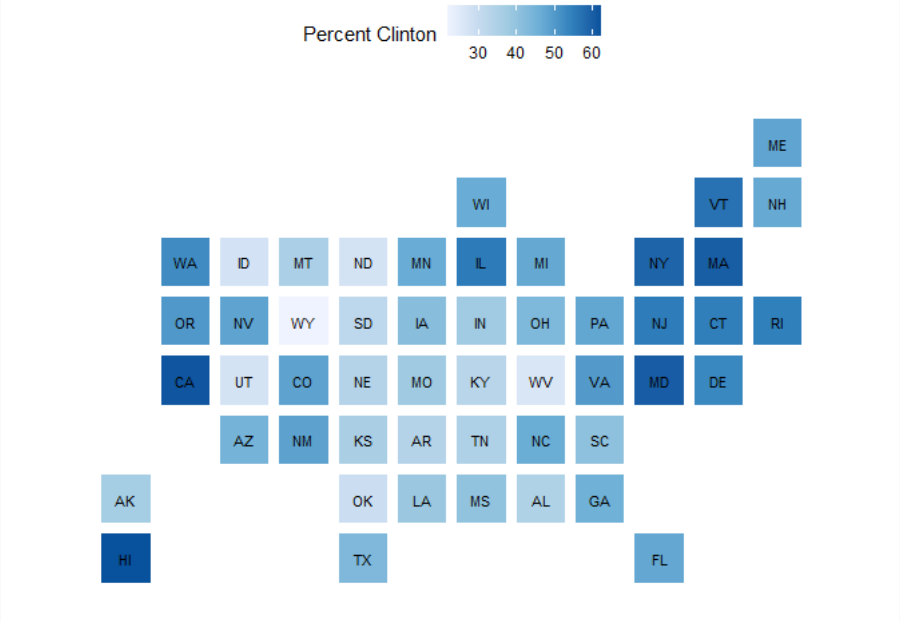



Outro tipo de gráfico utilizando o mapa para representar votos nas eleições;

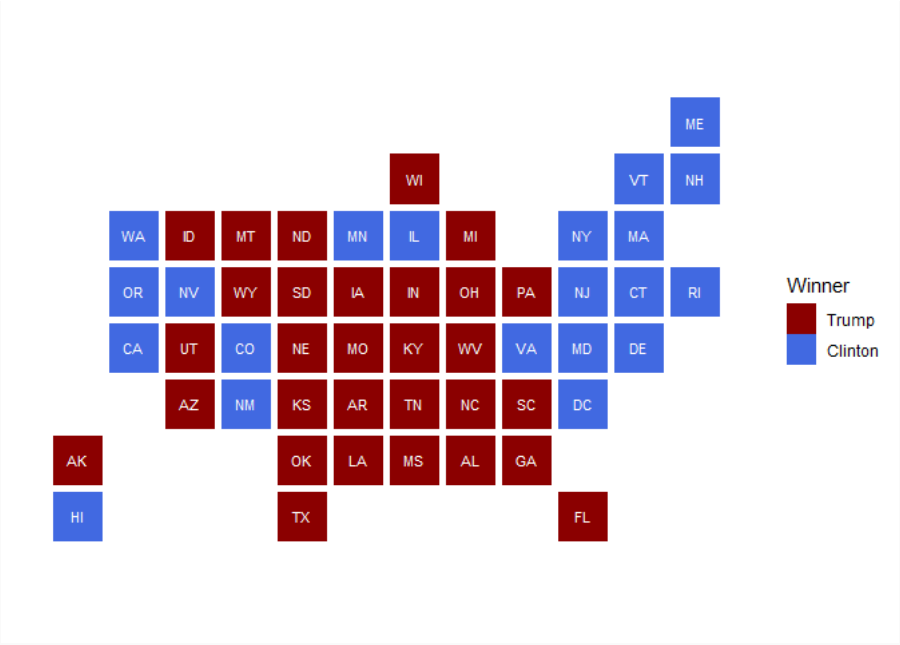
```
statebins_continuous(state_data = election, state_col = "state",
                     text_color = "white", value_col = "pct_trump",
                     brewer_pal = "Reds", font_size=3,
                     legend_title = "Percent Trump")
```



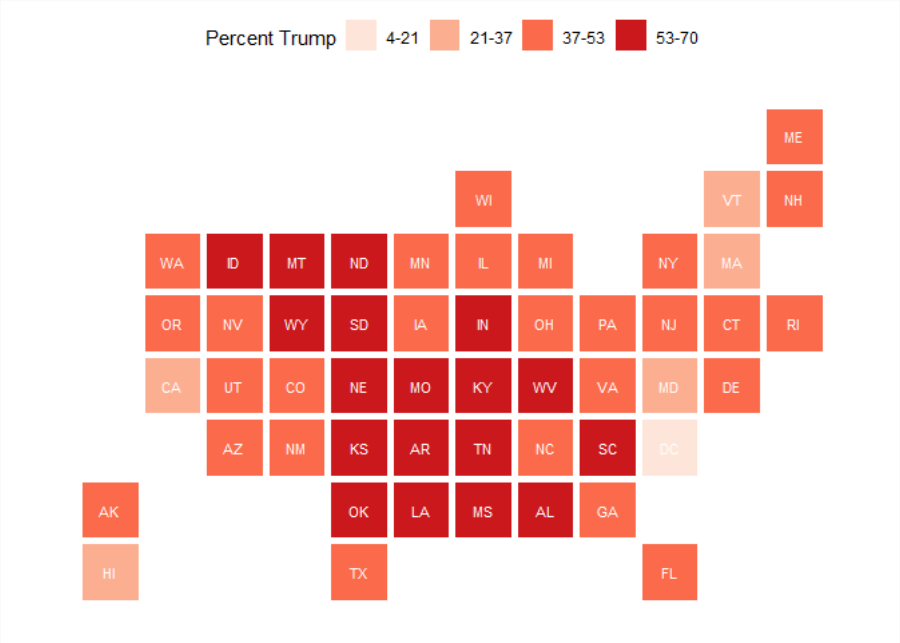

```
statebins_continuous(state_data = subset(election, st %nin% "DC"),
                     state_col = "state",
                     text_color="black", value_col = "pct_clinton",
                     brewer_pal="Blues", font_size = 3,
                     legend_title="Percent Clinton")
```




```
election <- election %>% mutate(color = recode(party, Republican = "darkred", Democratic =
        "royalblue"))

statebins_manual(state_data = election, state_col = "st",
                 color_col = "color", text_color = "white",
                 font_size = 3, legend_title="Winner",
                 labels=c("Trump", "Clinton"), legend_position = "right")
```




```
statebins(state_data = election,
          state_col = "state", value_col = "pct_trump",
          text_color = "white", breaks = 4,
          labels = c("4-21", "21-37", "37-53", "53-70"),
          brewer_pal="Reds", font_size = 3, legend_title="Percent Trump")
```



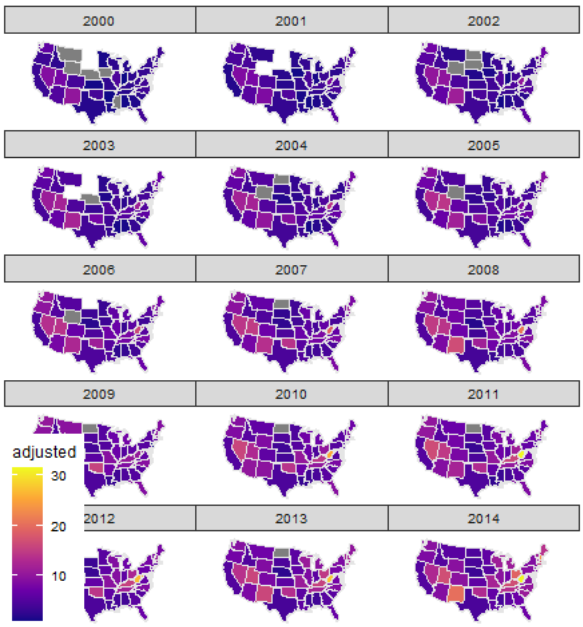

```
opiates$region <- tolower(opiates$state)
opiates_map <- left_join(us_states, opiates)
```

```
## Joining, by = "region"
```

Mortes causadas por narcóticos

```
p0 <- ggplot(data = subset(opiates_map, year >1999),
             mapping = aes(x = long, y = lat,
                           group = group,
                           fill = adjusted))
```

```
p1 <- p0 + geom_polygon(color = "gray90", size = 0.05) +
  coord_map(projection = "albers", lat0=39, lat1=45)

p2 <- p1 + scale_fill_viridis_c(option = "plasma")

p2 + theme_map() + facet_wrap(~year, ncol=3)
```

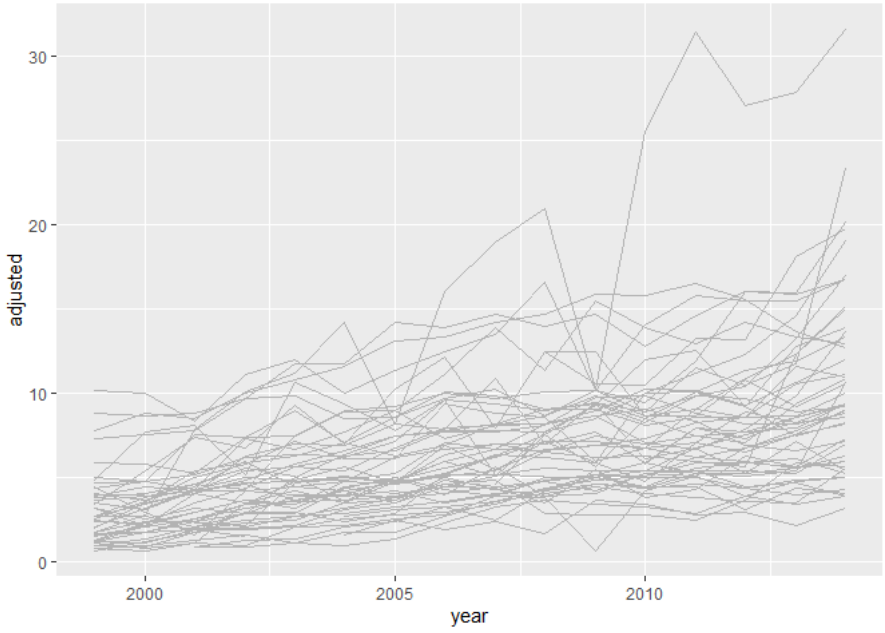



```
theme(legend.position = "bottom",
      strip.background = element_blank()) +
  labs(fill = "Death rate per 100,000 population",
       title = "Opiate Related Deaths by State, 2000-2014")
```

```
## List of 4
##  $ legend.position : chr "bottom"
##  $ strip.background: list()
##   ..- attr(*, "class")= chr [1:2] "element_blank" "element"
##  $ fill            : chr "Death rate per 100,000 population"
##  $ title           : chr "Opiate Related Deaths by State, 2000-2014"
##  - attr(*, "class")= chr [1:2] "theme" "gg"
##  - attr(*, "complete")= logi FALSE
##  - attr(*, "validate")= logi TRUE
```

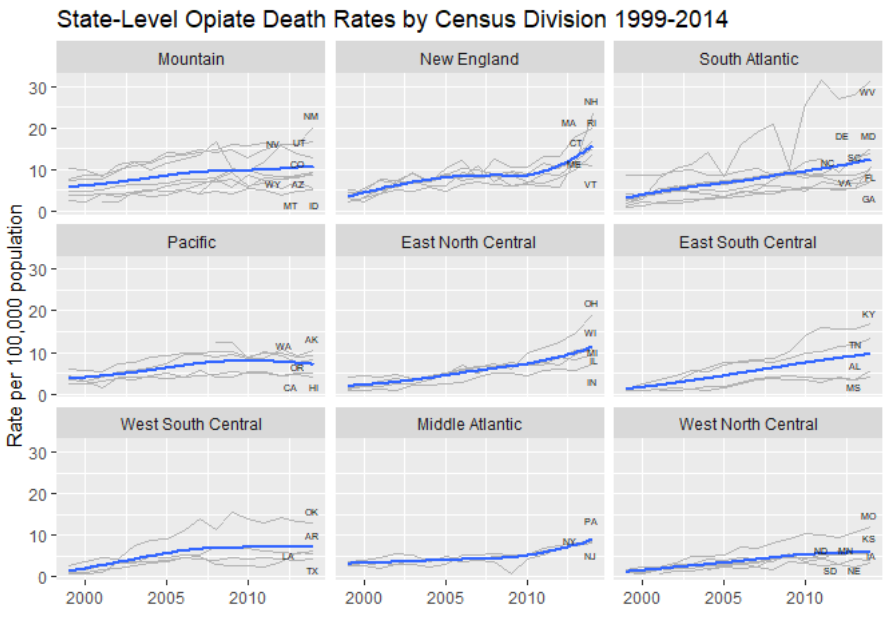
Outro estilo para representar os mesmos dados, porém não é possível identificar a qual estado cada linha refere;

```
p <- ggplot(data = opiates, mapping = aes(x=year, y = adjusted, group = state))
p + geom_line(color = "gray70")
```



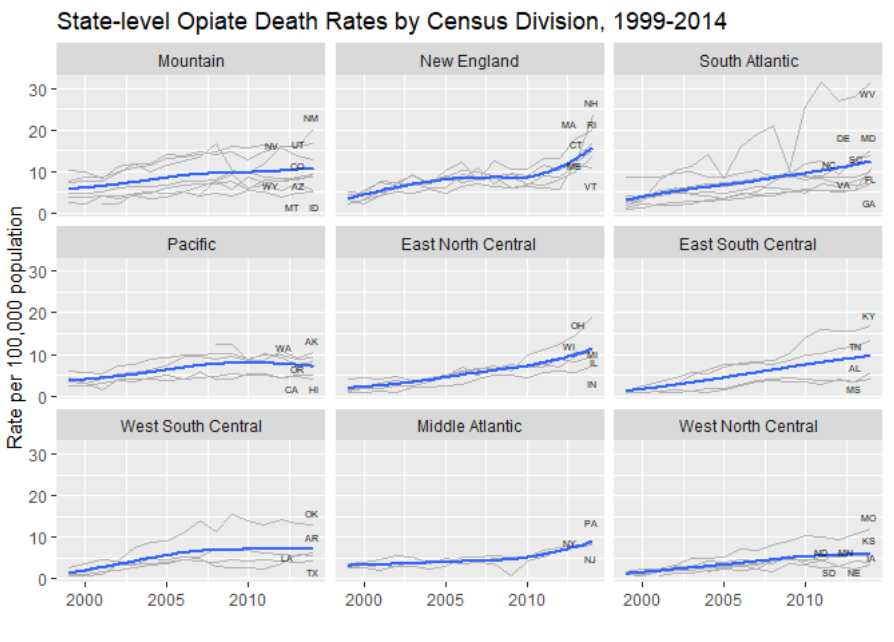

```
p0 <- ggplot(data = drop_na(opiates, division_name),
             mapping=aes(x = year, y = adjusted))

p1 <- p0 + geom_line(color = "gray70",
                     mapping = aes(group=state))

p2 <- p1 + geom_smooth(mapping = aes(group = division_name),
                       se = FALSE)

p3 <- p2 + geom_text_repel(data = subset(opiates,
                                         year == max(year) & abbr !="DC"),
                           mapping = aes(x = year, y = adjusted, label = abbr),
                           size = 1.8, segment.color = NA, nudge_x = 30) +
  coord_cartesian(c(min(opiates$year),
                    max(opiates$year)))

p3 + labs(x = "", y = "Rate per 100,000 population",
          title = "State-Level Opiate Death Rates by Census Division 1999-2014") +
  facet_wrap(~reorder(division_name, -adjusted, na.rm = TRUE), nrow = 3)
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



```
p3 + labs(x = "", y = "Rate per 100,000 population",
          title = "State-level Opiate Death Rates by Census Division, 1999-2014") +
  facet_wrap(~reorder(division_name, -adjusted, na.rm=TRUE), nrow=3)
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




# Item 2

Seguindo a referência deixada para o trabalho, reproduzi os heatmaps listados na mesma;

```
library(plotrix)
library(RColorBrewer)
library(stringr)
library(gplots)
```

```
##
## Attaching package: 'gplots'
```

```
## The following object is masked from 'package:plotrix':
##
##     plotCI
```

```
## The following object is masked from 'package:stats':
##
##     lowess
```

```
m <- read.csv("measles_lev1.csv",header=T,stringsAsFactors=F,skip=2)

head(m)
```

| YEAR | WEEK | ALABAMA | ALASKA | ARIZONA | ARKANSAS | CALIFORNIA | COLORADO | CONNECTICUT | D |
|---|---|---|---|---|---|---|---|---|---|
| 1928 | 1 | 3.67 | - | 1.90 | 4.11 | 1.38 | 8.38 | 4.50 | 8 |

| YEAR | WEEK | ALABAMA | ALASKA | ARIZONA | ARKANSAS | CALIFORNIA | COLORADO | CONNECTICUT | D |
|---|---|---|---|---|---|---|---|---|---|
| 1928 | 2 | 6.25 | - | 6.40 | 9.91 | 1.80 | 6.02 | 9.00 | 7 |
| 1928 | 3 | 7.95 | - | 4.50 | 11.15 | 1.31 | 2.86 | 8.81 | 1 |
| 1928 | 4 | 12.58 | - | 1.90 | 13.75 | 1.87 | 13.71 | 10.40 | 4 |
| 1928 | 5 | 8.03 | - | 0.47 | 20.79 | 2.38 | 5.13 | 16.80 | 5 |
| 1928 | 6 | 7.27 | - | 6.40 | 26.58 | 2.79 | 8.09 | 17.76 | 3 |

```
str(m)
```

```
## 'data.frame':    3952 obs. of  53 variables:
##  $ YEAR                : int  1928 1928 1928 1928 1928 1928 1928 1928 1928 1928 ...
##  $ WEEK                : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ ALABAMA             : chr  "3.67" "6.25" "7.95" "12.58" ...
##  $ ALASKA              : chr  "-" "-" "-" "-" ...
##  $ ARIZONA             : chr  "1.90" "6.40" "4.50" "1.90" ...
##  $ ARKANSAS            : chr  "4.11" "9.91" "11.15" "13.75" ...
##  $ CALIFORNIA          : chr  "1.38" "1.80" "1.31" "1.87" ...
##  $ COLORADO            : chr  "8.38" "6.02" "2.86" "13.71" ...
##  $ CONNECTICUT         : chr  "4.50" "9.00" "8.81" "10.40" ...
##  $ DELAWARE            : chr  "8.58" "7.30" "15.88" "4.29" ...
##  $ DISTRICT.OF.COLUMBIA: chr  "-" "-" "-" "4.18" ...
##  $ FLORIDA             : chr  "0.21" "0.49" "0.42" "0.91" ...
##  $ GEORGIA             : chr  "1.17" "5.96" "-" "8.65" ...
##  $ HAWAII              : chr  "-" "-" "-" "-" ...
##  $ IDAHO               : chr  "-" "0.45" "0.45" "-" ...
##  $ ILLINOIS            : chr  "0.50" "0.77" "0.61" "0.81" ...
##  $ INDIANA             : chr  "1.34" "2.71" "1.71" "4.11" ...
##  $ IOWA                : chr  "0.16" "-" "-" "3.51" ...
##  $ KANSAS              : chr  "0.81" "1.35" "1.41" "1.14" ...
##  $ KENTUCKY            : chr  "3.08" "1.99" "5.26" "5.49" ...
##  $ LOUISIANA           : chr  "1.89" "3.00" "2.33" "4.02" ...
##  $ MAINE               : chr  "4.52" "7.40" "6.78" "9.41" ...
##  $ MARYLAND            : chr  "10.87" "15.47" "21.43" "22.67" ...
##  $ MASSACHUSETTS       : chr  "25.66" "28.50" "34.76" "31.28" ...
##  $ MICHIGAN            : chr  "5.68" "7.59" "9.39" "8.66" ...
##  $ MINNESOTA           : chr  "0.31" "0.23" "0.15" "0.12" ...
##  $ MISSISSIPPI         : chr  "-" "-" "-" "-" ...
##  $ MISSOURI            : chr  "1.19" "0.83" "1.69" "1.58" ...
##  $ MONTANA             : chr  "0.18" "0.18" "0.74" "-" ...
##  $ NEBRASKA            : chr  "1.60" "0.29" "0.36" "0.44" ...
##  $ NEVADA              : chr  "-" "-" "-" "-" ...
##  $ NEW.HAMPSHIRE       : chr  "-" "-" "-" "14.53" ...
##  $ NEW.JERSEY          : chr  "3.55" "4.74" "6.68" "6.78" ...
##  $ NEW.MEXICO          : chr  "14.90" "11.06" "14.90" "27.64" ...
##  $ NEW.YORK            : chr  "7.60" "9.65" "8.54" "9.32" ...
##  $ NORTH.CAROLINA      : chr  "47.86" "119.70" "110.90" "131.60" ...
##  $ NORTH.DAKOTA        : chr  "-" "0.15" "1.20" "3.91" ...
##  $ OHIO                : chr  "2.51" "-" "4.86" "4.40" ...
##  $ OKLAHOMA            : chr  "4.86" "2.56" "6.27" "4.74" ...
##  $ OREGON              : chr  "4.91" "4.91" "3.63" "2.24" ...
##  $ PENNSYLVANIA        : chr  "6.97" "8.74" "8.12" "8.39" ...
##  $ RHODE.ISLAND        : chr  "1.18" "0.74" "2.65" "0.15" ...
##  $ SOUTH.CAROLINA      : chr  "42.04" "83.90" "77.46" "64.75" ...
##  $ SOUTH.DAKOTA        : chr  "5.69" "6.57" "2.04" "2.19" ...
##  $ TENNESSEE           : chr  "22.03" "16.96" "24.66" "18.86" ...
##  $ TEXAS               : chr  "1.18" "0.63" "0.62" "0.37" ...
##  $ UTAH                : chr  "0.40" "-" "0.20" "0.20" ...
##  $ VERMONT             : chr  "0.28" "0.56" "1.12" "6.70" ...
##  $ VIRGINIA            : chr  "-" "-" "-" "-" ...
```

```
##  $ WASHINGTON          : chr  "14.83" "17.34" "15.67" "12.77" ...
##  $ WEST.VIRGINIA       : chr  "3.36" "4.19" "4.19" "4.66" ...
##  $ WISCONSIN           : chr  "1.54" "0.96" "4.79" "1.64" ...
##  $ WYOMING             : chr  "0.91" "-" "1.36" "3.64" ...
```

```
table(m$YEAR)
```

```
##
## 1928 1929 1930 1931 1932 1933 1934 1935 1936 1937 1938 1939 1940 1941 1942 1943
##   52   52   52   52   52   52   52   52   52   52   52   52   52   52   52   52
## 1944 1945 1946 1947 1948 1949 1950 1951 1952 1953 1954 1955 1956 1957 1958 1959
##   52   52   52   52   52   52   52   52   52   52   52   52   52   52   52   52
## 1960 1961 1962 1963 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973 1974 1975
##   52   52   52   52   52   52   52   52   52   52   52   52   52   52   52   52
## 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991
##   52   52   52   52   52   52   52   52   52   52   52   52   52   52   52   52
## 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003
##   52   52   52   52   52   52   52   52   52   52   52   52
```

```
table(m$WEEK)
```

```
##
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
## 76 76 76 76 76 76 76 76 76 76 76 76 76 76 76 76 76 76 76 76 76 76 76 76 76 76
## 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52
## 76 76 76 76 76 76 76 76 76 76 76 76 76 76 76 76 76 76 76 76 76 76 76 76 76 76
```

```
m2 <- m %>%
  # convert data to long format
  gather(key="state",value="value",-YEAR,-WEEK) %>%
  # rename columns
  setNames(c("year","week","state","value")) %>%
  # convert year to factor
  mutate(year=factor(year)) %>%
  # convert week to factor
  mutate(week=factor(week)) %>%
  # convert value to numeric (also converts '-' to NA, gives a warning)
  mutate(value=as.numeric(value))
```

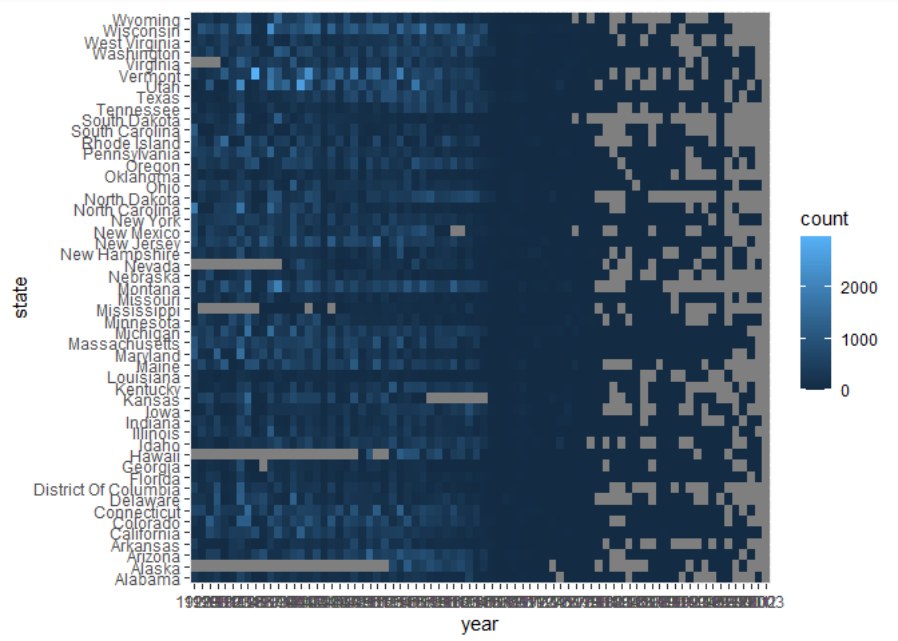
```
# removes . and change states to title case using custom function
fn_tc <- function(x) paste(str_to_title(unlist(strsplit(x,"[.]"))),collapse=" ")
m2$state <- sapply(m2$state,fn_tc)

# custom sum function returns NA when all values in set are NA,
# in a set mixed with NAs, NAs are removed and remaining summed.
na_sum <- function(x)
{
  if(all(is.na(x))) val <- sum(x,na.rm=F)
  if(!all(is.na(x))) val <- sum(x,na.rm=T)
  return(val)
}

# sum incidences for all weeks into one year
m3 <- m2 %>%
  group_by(year,state) %>%
  summarise(count=na_sum(value)) %>%
  as.data.frame()
```

```
## `summarise()` has grouped output by 'year'. You can override using the `.groups` argument.
```

```
#basic ggplot
p <- ggplot(m3,aes(x=year,y=state,fill=count))+
      geom_tile()

#save plot to working directory
ggsave(p,filename="measles-basic.png")
```

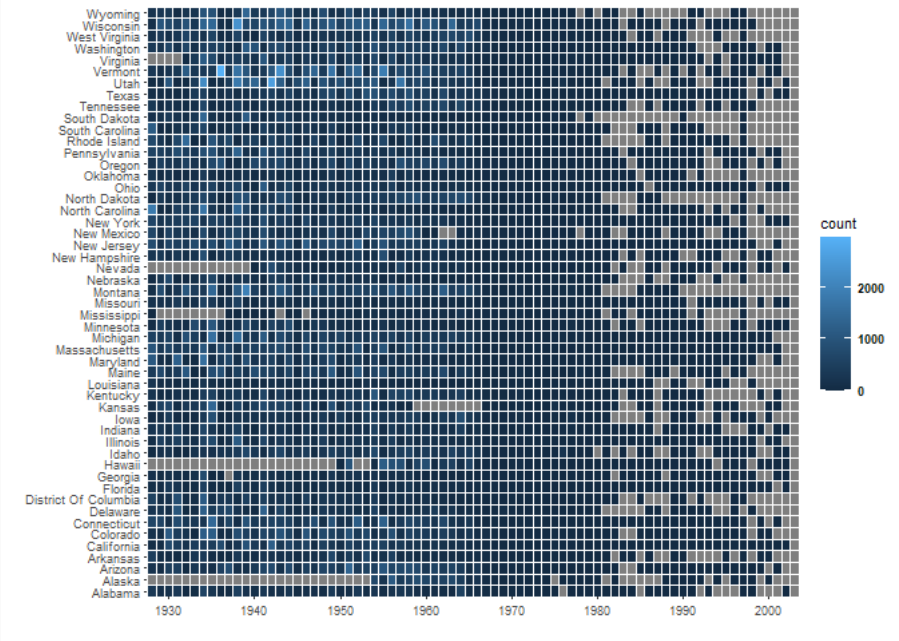
```
## Saving 7 x 5 in image
```

```
p
```



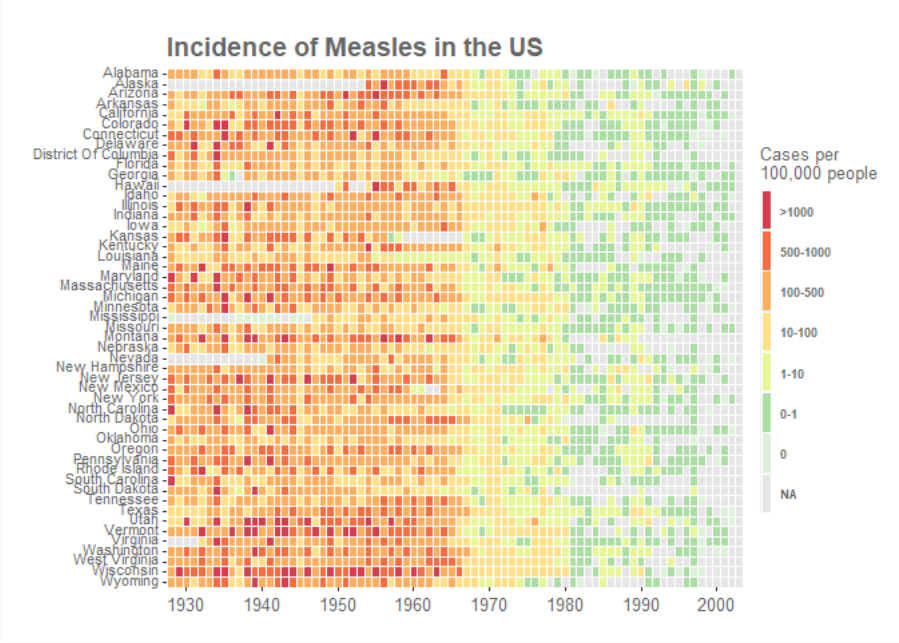

```
p <- ggplot(m3,aes(x=year,y=state,fill=count))+
  #add border white colour of line thickness 0.25
  geom_tile(colour="white",size=0.25)+
  #remove x and y axis labels
  labs(x="",y="")+
  #remove extra space
  scale_y_discrete(expand=c(0,0))+
  #define new breaks on x-axis
  scale_x_discrete(expand=c(0,0),
                   breaks=c("1930","1940","1950","1960","1970","1980","1990","2000"))+
  #set a base size for all fonts
  theme_grey(base_size=8)+
  #theme options
  theme(
    #bold font for legend text
    legend.text=element_text(face="bold"),
    #set thickness of axis ticks
    axis.ticks=element_line(size=0.4),
    #remove plot background
    plot.background=element_blank(),
    #remove plot border
    panel.border=element_blank())

#save with dpi 200
ggsave(p,filename="measles-mod1.png",height=5.5,width=8.8,units="in",dpi=200)
```

```
p
```




```
m4 <- m3 %>%
      # convert state to factor and reverse order of levels
      mutate(state=factor(state,levels=rev(sort(unique(state))))) %>%
      # create a new variable from count
      mutate(countfactor=cut(count,breaks=c(-1,0,1,10,100,500,1000,max(count,na.rm=T)),
                             labels=c("0","0-1","1-10","10-100","100-500","500-
        1000",">1000"))) %>%
      # change level order
      mutate(countfactor=factor(as.character(countfactor),levels=rev(levels(countfactor))))

# assign text colour
textcol <- "grey40"

# further modified ggplot
p <- ggplot(m4,aes(x=year,y=state,fill=countfactor))+
  geom_tile(colour="white",size=0.2)+
  guides(fill=guide_legend(title="Cases per\n100,000 people"))+
  labels(x="",y="",title="Incidence of Measles in the US")+
  scale_y_discrete(expand=c(0,0))+

        scale_x_discrete(expand=c(0,0),breaks=c("1930","1940","1950","1960","1970","1980","1990","2000"))+

        scale_fill_manual(values=c("#d53e4f","#f46d43","#fdae61","#fee08b","#e6f598","#abdda4","#ddf1da"),r
         = "grey90")+
  #coord_fixed()+
  theme_grey(base_size=10)+
  theme(legend.position="right",legend.direction="vertical",
        legend.title=element_text(colour=textcol),
        legend.margin=margin(grid::unit(0,"cm")),
        legend.text=element_text(colour=textcol,size=7,face="bold"),
        legend.key.height=grid::unit(0.8,"cm"),
        legend.key.width=grid::unit(0.2,"cm"),
        axis.text.x=element_text(size=10,colour=textcol),
        axis.text.y=element_text(vjust=0.2,colour=textcol),
        axis.ticks=element_line(size=0.4),
        plot.background=element_blank(),
        panel.border=element_blank(),
        plot.margin=margin(0.7,0.4,0.1,0.2,"cm"),
        plot.title=element_text(colour=textcol,hjust=0,size=14,face="bold"))

#export figure
ggsave(p,filename="measles-mod3.png",height=5.5,width=8.8,units="in",dpi=200)
```

```
p
```


Incidence of Measles in the US
Alabama
Alaska
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
District Of Columbia
Florida
Georgia
Hawaii
Idaho
Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Minnesota
Mississippi
Missouri
Montana
Nebraska
Nevada
New Hampshire
New Jersey
New Mexico
New York
North Carolina
North Dakota
Ohio
Oklahoma
Oregon
Pennsylvania
Rhode Island
South Carolina
South Dakota
Tennessee
Texas
Utah
Vermont
Virginia
Washington
West Virginia
Wisconsin
Wyoming
1930
1940
1950
1960
1970
1980
1990
2000
Cases per 100,000 people
>1000
500-1000
100-500
10-100
1-10
0-1
0
NA