Assignment1_Mar20

Garrett Bullivant

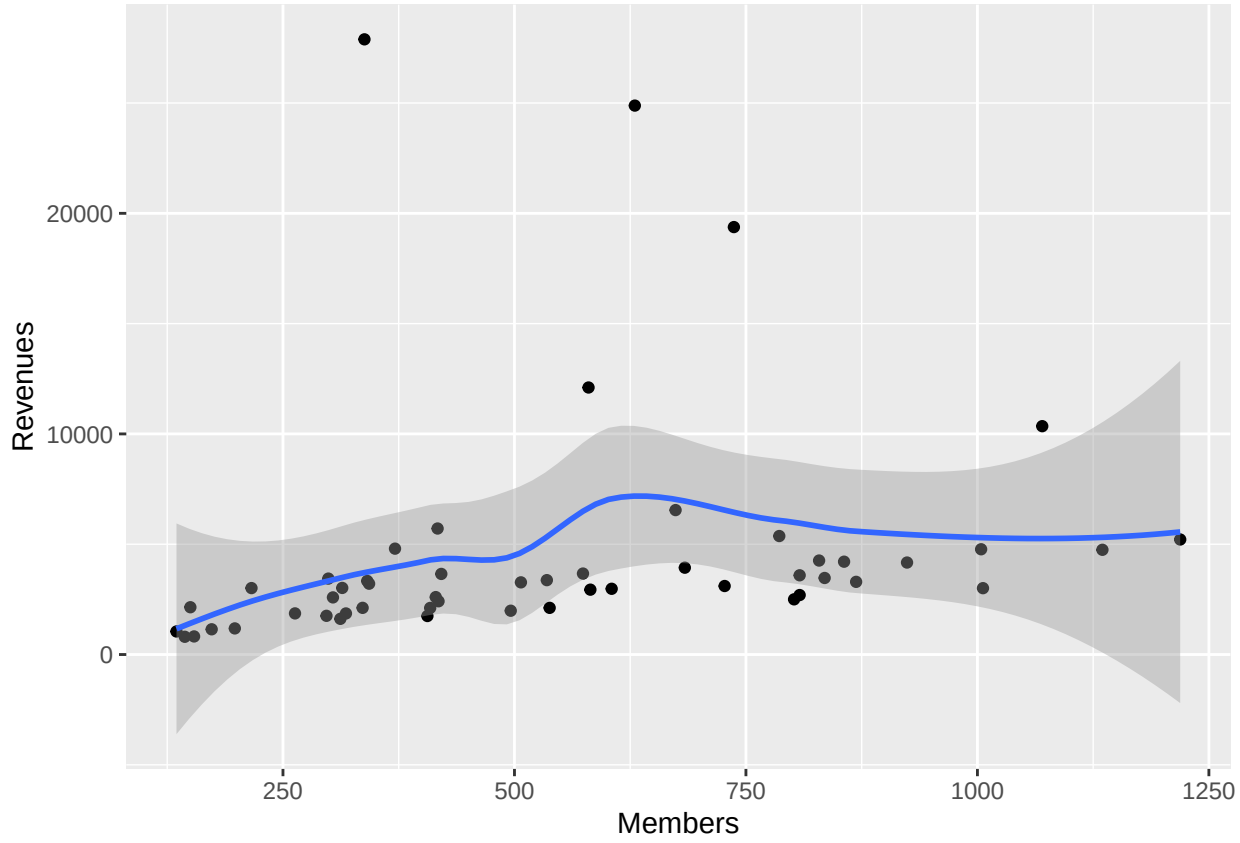
2023-03-20

```
# use a new sample dataset - asasec from the socviz library
library(socviz)

head(asasec)
```

```
##                                Section         Sname Beginning Revenues
## 1      Aging and the Life Course (018)         Aging     12752    12104
## 2     Alcohol, Drugs and Tobacco (030) Alcohol/Drugs     11933     1144
## 3 Altruism and Social Solidarity (047)      Altruism      1139     1862
## 4            Animals and Society (042)       Animals       473      820
## 5             Asia/Asian America (024)          Asia      9056     2116
## 6            Body and Embodiment (048)          Body      3408     1618
##   Expenses Ending Journal Year Members
## 1    12007  12849      No 2005     598
## 2      400  12677      No 2005     301
## 3     1875   1126      No 2005      NA
## 4     1116    177      No 2005     209
## 5     1710   9462      No 2005     365
## 6     1920   3106      No 2005      NA
```

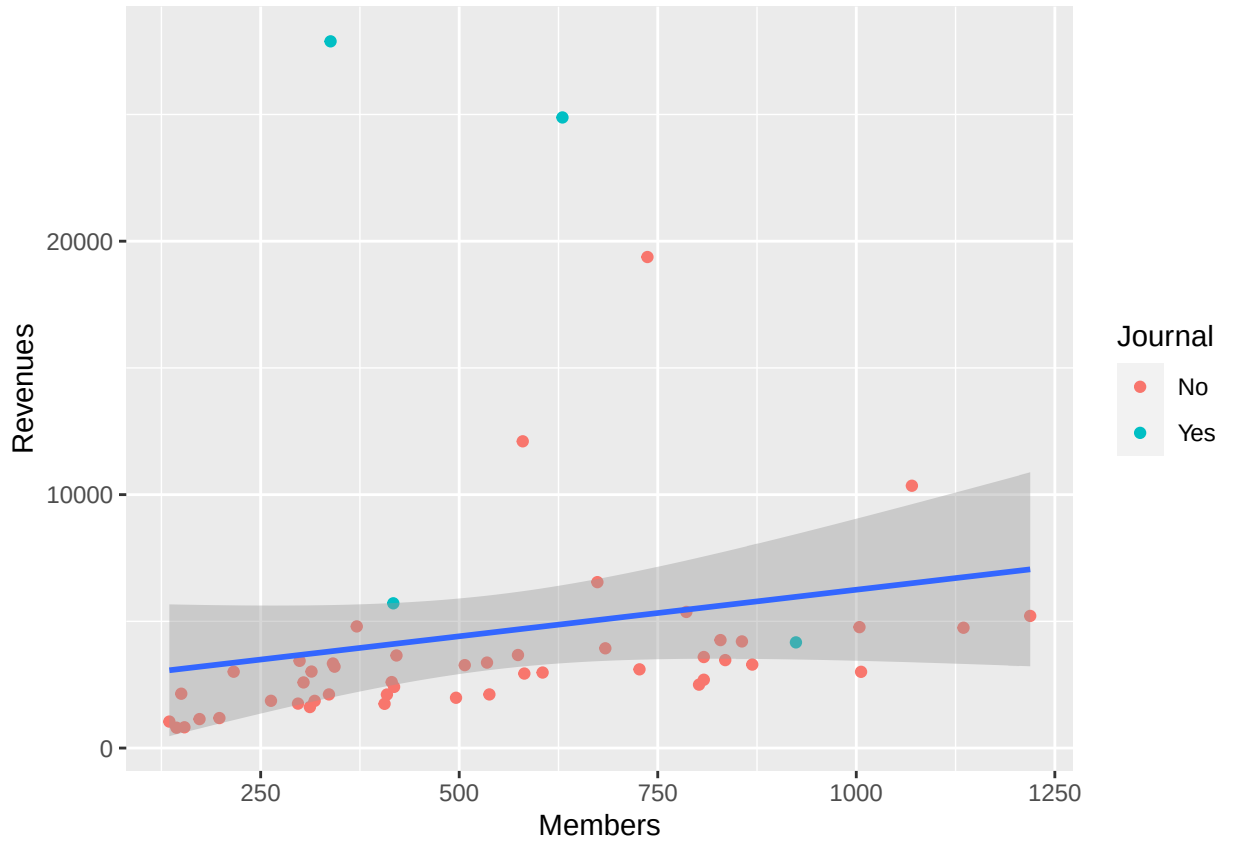
```
# create a scatterplot and smoothed graph comparing membership and revenues for the year 2014
library(ggplot2)
p <- ggplot(data = subset(asasec, Year == 2014 ), mapping = aes(x = Members,
             y = Revenues))
p + geom_point() + geom_smooth()
```

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




```
# add colour to our points based on Journal and switch our geom_smooth to a linear model

p + geom_point(mapping = aes(colour = Journal)) + geom_smooth(method = "lm")
```

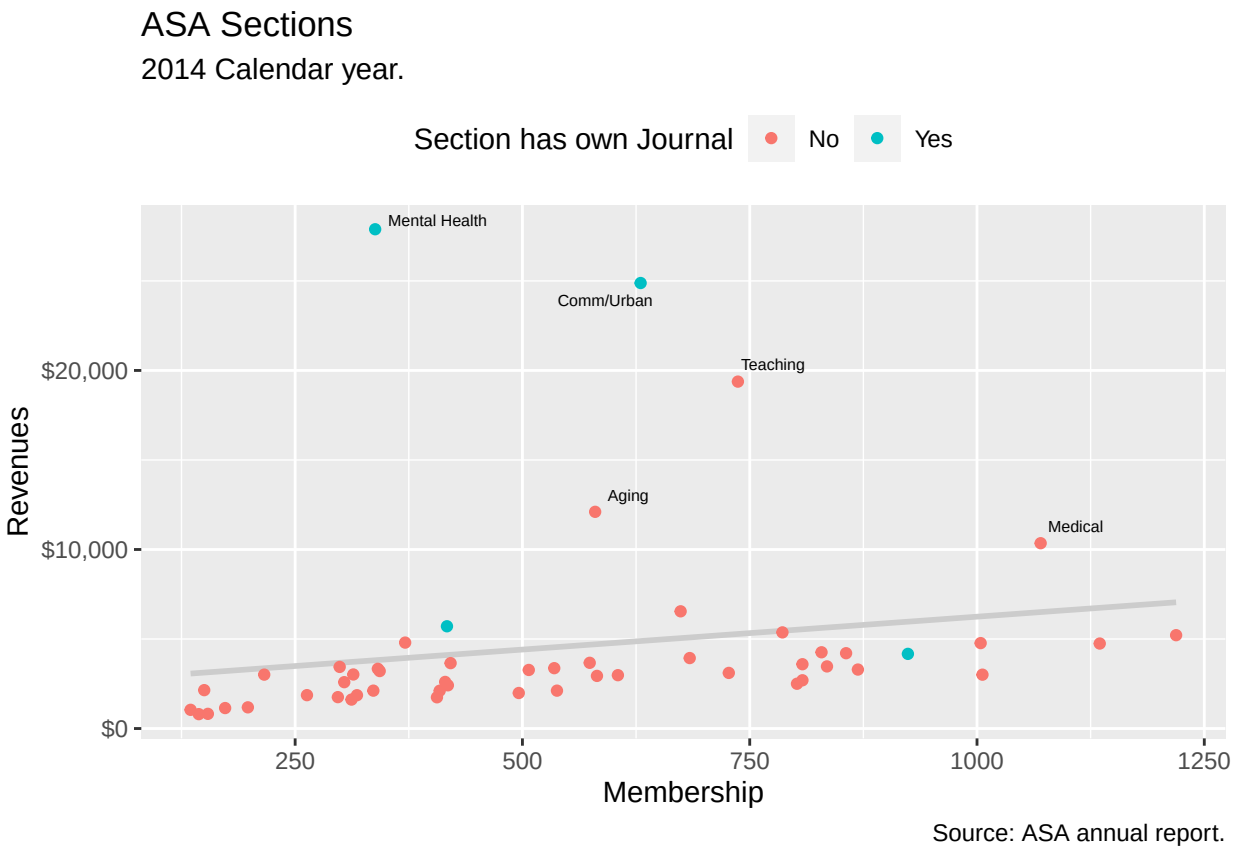
```
## `geom_smooth()` using formula = 'y ~ x'
```



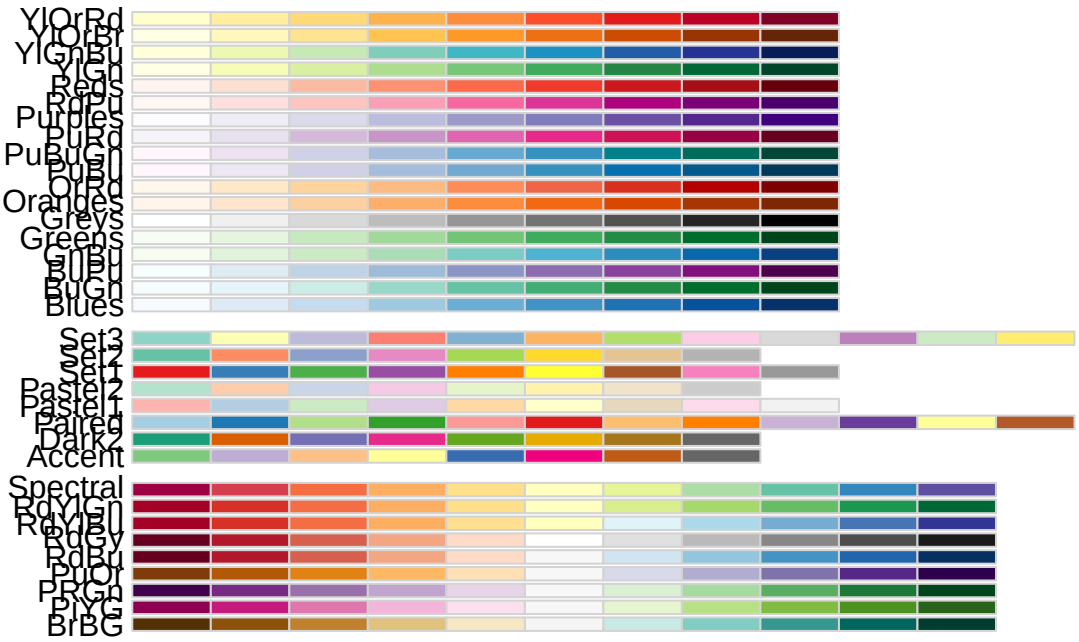

```
library(ggrepel)
# if we want to add more layers to our ggplot, it can be helpful to make intermediate objects
p0 <- ggplot(data = subset(asasec, Year == 2014), mapping = aes(x = Members, y = Revenues, label = Sname

p1 <- p0 + geom_smooth(method = "lm", se = FALSE, color = "gray80") + geom_point(mapping = aes(color = .

p2 <- p1 + geom_text_repel(data = subset(asasec, Year == 2014 & Revenues > 7000), size = 2)

p3 <- p2 + labs(x="Membership",
                y="Revenues",
                color = "Section has own Journal",
                title = "ASA Sections",
                subtitle = "2014 Calendar year.",
                caption = "Source: ASA annual report.")

p4 <- p3 + scale_y_continuous(labels = scales::dollar) +
  theme(legend.position = "top")

p4
```

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

## Warning: The following aesthetics were dropped during statistical transformation: label
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
```

```
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```

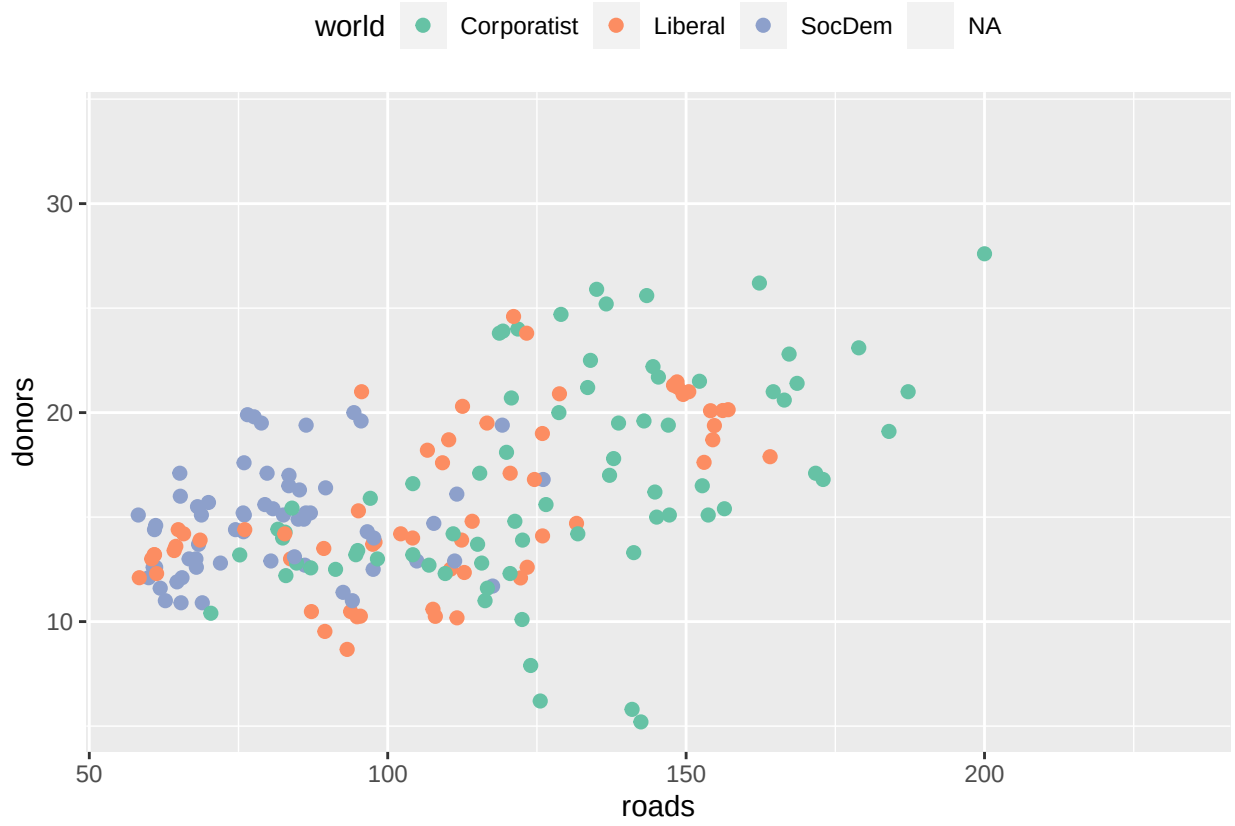



```
# we can observe colour palettes
library(RColorBrewer)
par(mar=c(3,4,2,2))
display.brewer.all()
```




```
# using colour palettes
p <- ggplot(data = organdata, mapping = aes(x = roads, y = donors, color = world))

p + geom_point(size = 2) +
  scale_color_brewer(palette = "Set2") +
  theme(legend.position = "top")
```

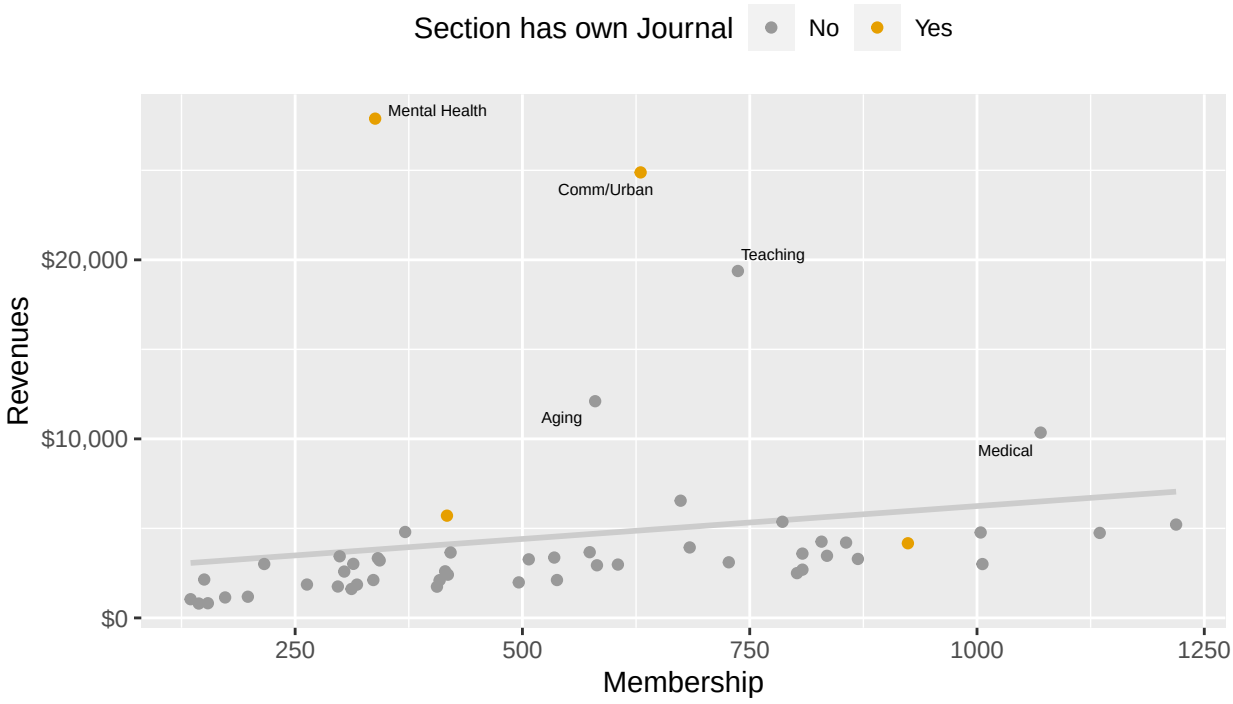
```
## Warning: Removed 46 rows containing missing values (`geom_point()`).
```



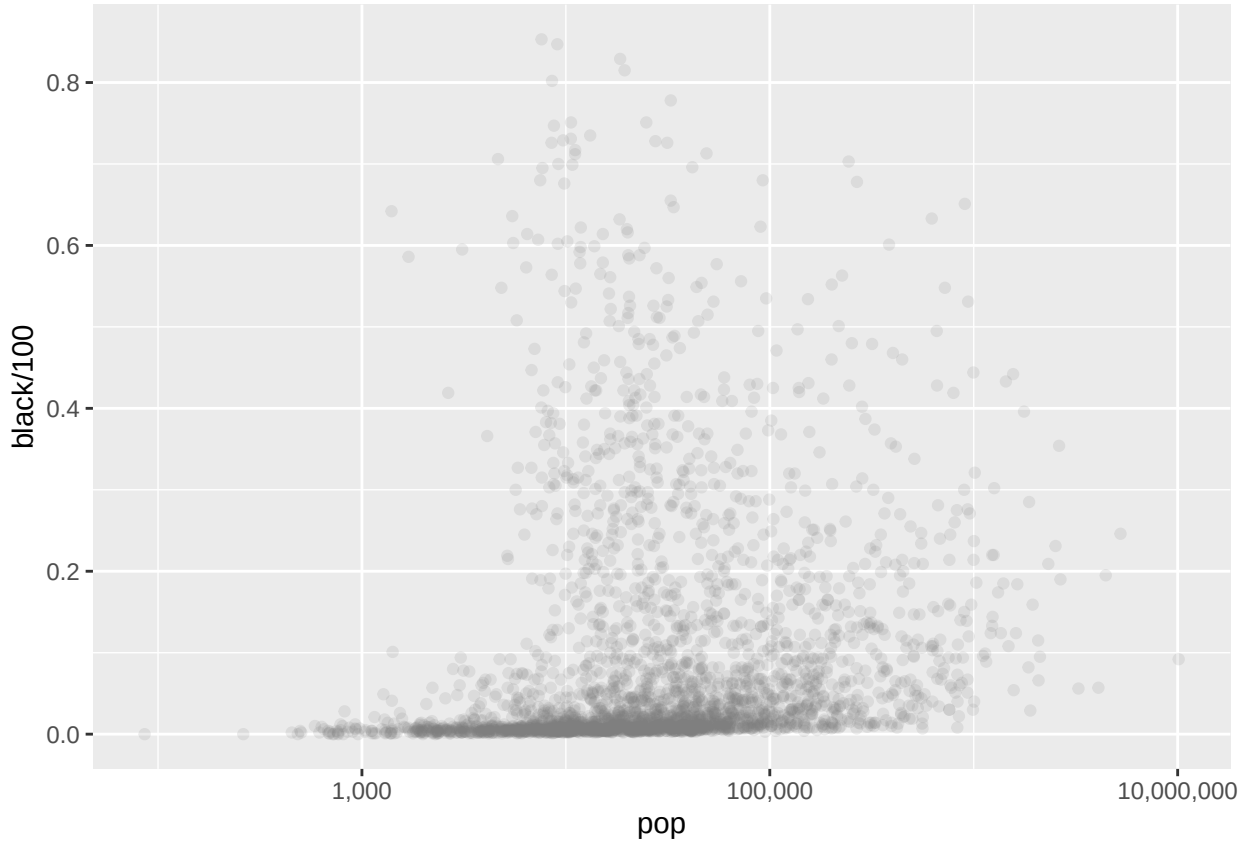

```
# we can also generate our own custom palettes using hexadecimal values or by using colour names known
cb_palette <- c("#999999", "#E69F00", "#56B4E9", "#009E73", "#F0E442", "#0072B2", "#D55E00", "#CC79A7")
p4 + scale_color_manual(values = cb_palette)
```

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

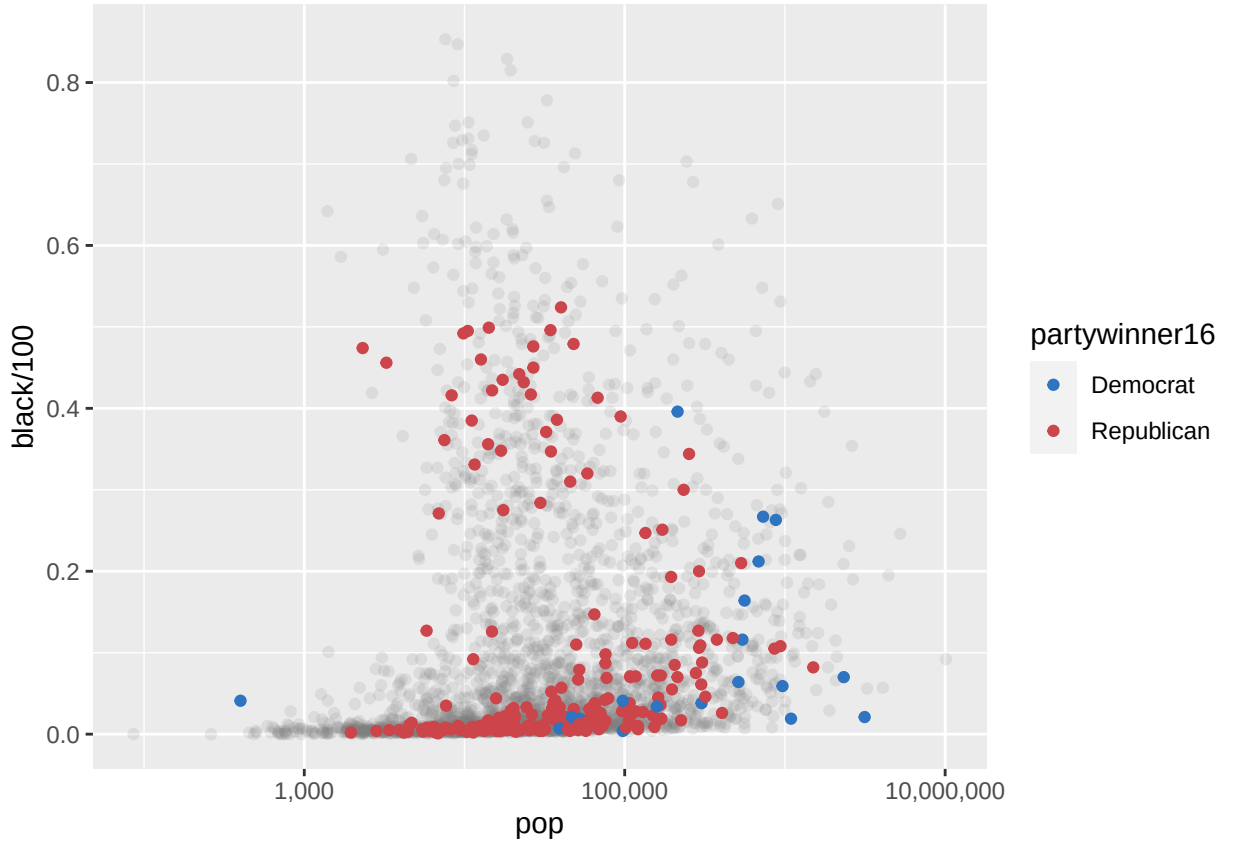
```
## Warning: The following aesthetics were dropped during statistical transformation: label
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```



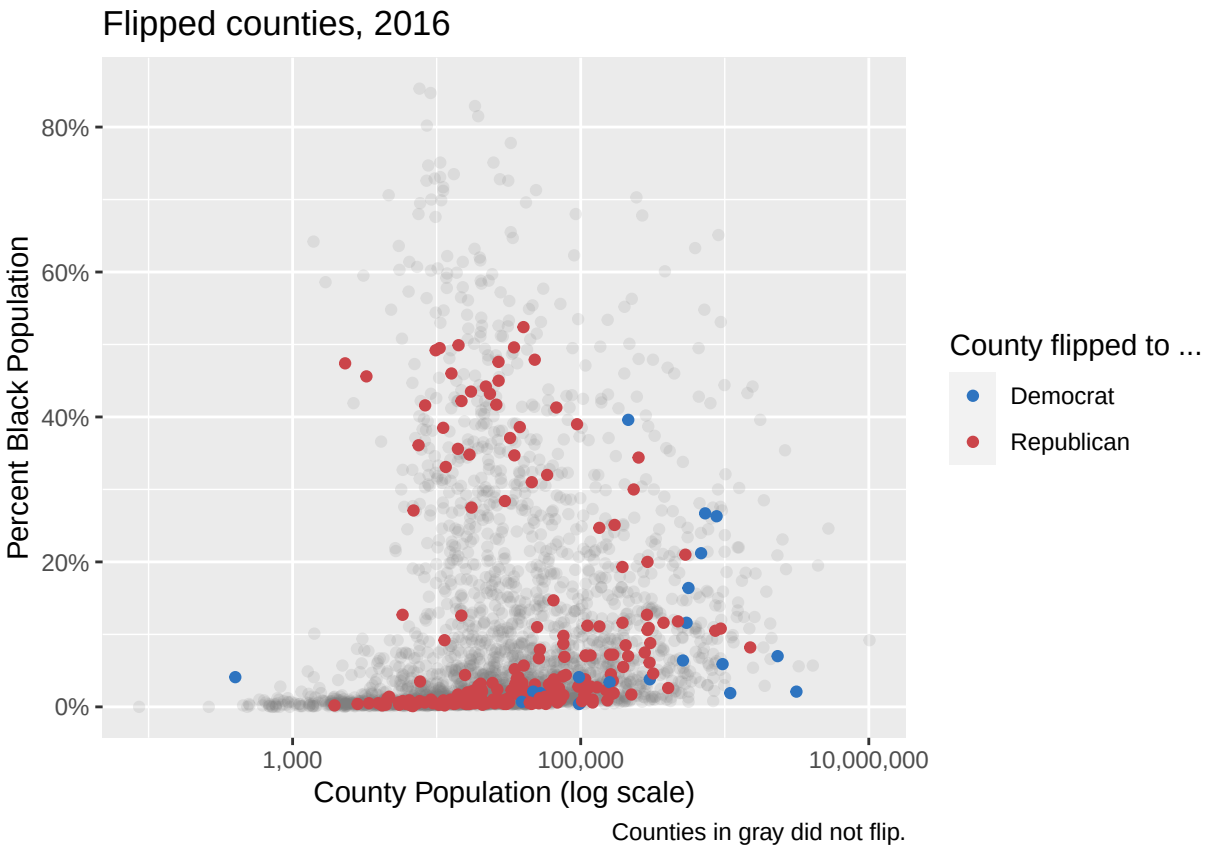

```
 # highlighting data with colour. We define a vector with republican and democratic colour
party_colors <- c("#2E74C0", "#CB454A")

# create our first layer of our ggplot - a scatterplot of counties that did NOT flip parties in the ele
# our x-axis is the population and the y-axis is the percentage of population that is black
p0 <- ggplot(data = subset(county_data, flipped == "No"),
             mapping = aes(x = pop,y = black/100))

p1 <- p0 + geom_point(alpha = 0.15, color = "gray50") +
  scale_x_log10(labels=scales::comma)

p1
```




```
# the second layer of our graph will show counties that DID flip party affiliation in the election
p2 <- p1 + geom_point(data = subset(county_data, flipped == "Yes"), mapping = aes(x = pop, y = black/10
  scale_colour_manual(values = party_colors)
p2
```

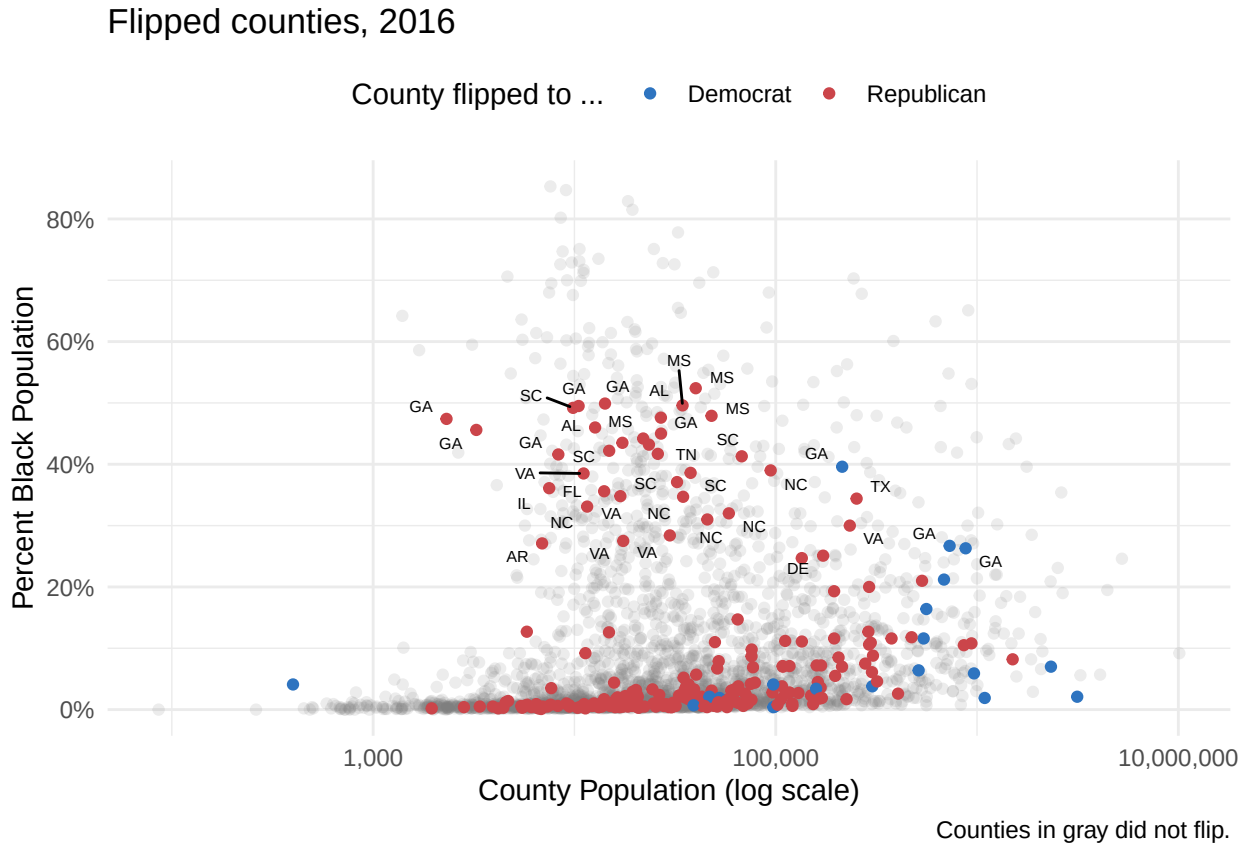



```
# we can add an additional layer that will label our axes, graph, and legend
p3 <- p2 + scale_y_continuous(labels=scales::percent) +
  labs(color = "County flipped to ... ",
       x = "County Population (log scale)",
       y = "Percent Black Population",
       title = "Flipped counties, 2016",
       caption = "Counties in gray did not flip.")
p3
```



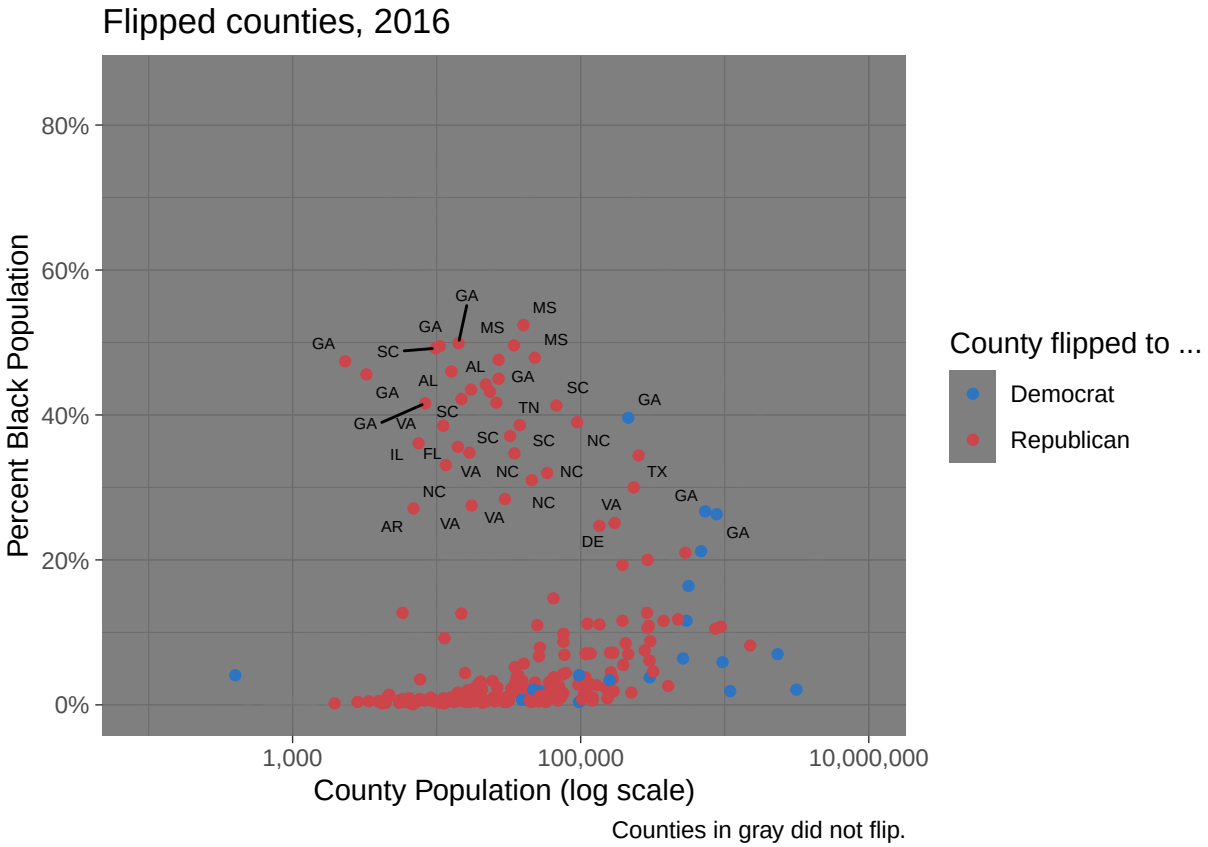

```
# finally we can add another layer that shows flipped counties that have relatively high percentage of
p4 <- p3 + geom_text_repel(data = subset(county_data, flipped == "Yes" & black > 25),
                           mapping = aes(x = pop, y = black/100, label = state), size = 2)

p4 + theme_minimal() + theme(legend.position="top")
```

```
## Warning: ggrepel: 2 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



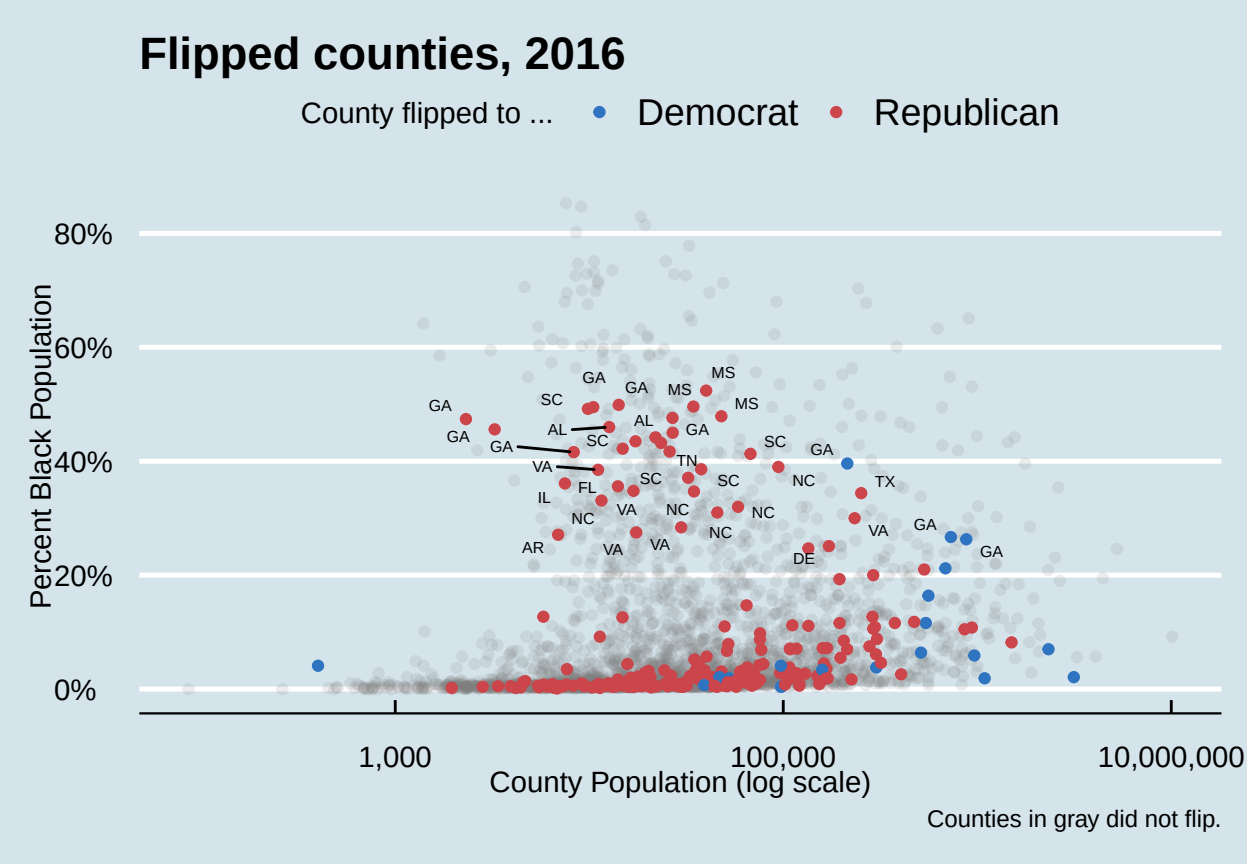

```
# we can alter the theme of our plots by using theme_set()
theme_set(theme_dark())
p4
```

```
## Warning: ggrepel: 3 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```

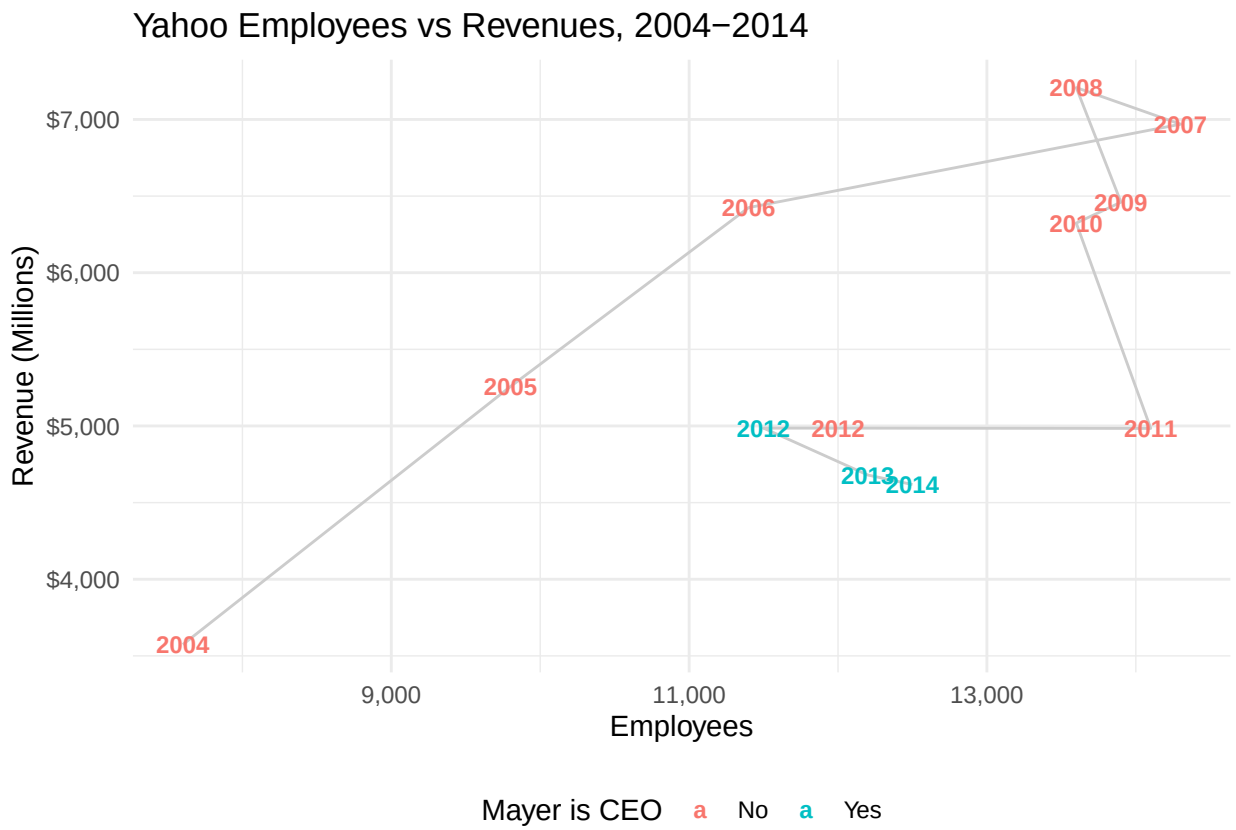



```
# we can load ggthemes library for additional themes
library(ggthemes)
theme_set(theme_economist())

p4
```

```
## Warning: ggrepel: 3 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



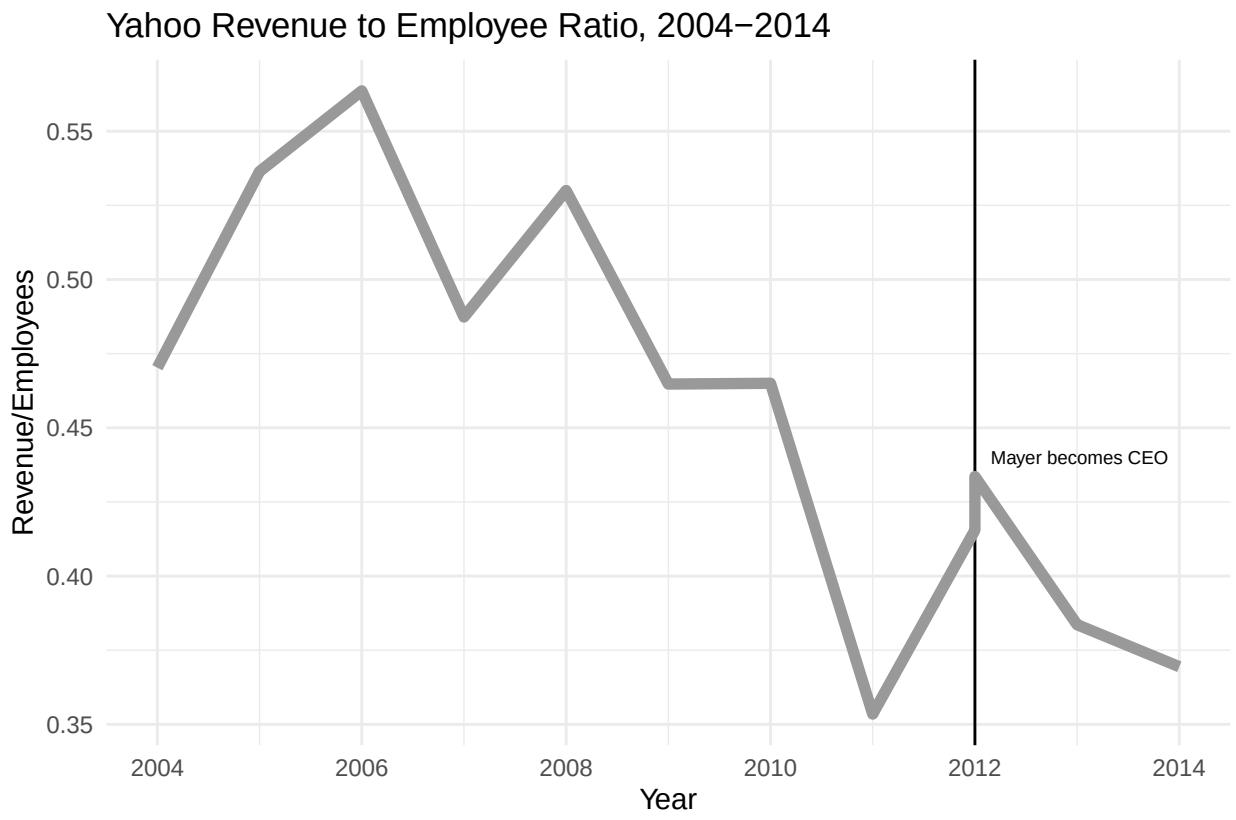

```
# case studies
# case study 1
# i think its a bad visualization because it has 2 y axes. Furthermore, it has a unlabelled red line ru

theme_set(theme_minimal())

# first redrawing of case study 1
# we now have employees and revenue as x and y, instead of both on the y-axis. we instead have the temp

p <- ggplot(data = yahoo,
            mapping = aes(x = Employees, y = Revenue))
p + geom_path(color = "gray80") +
    geom_text(aes(color = Mayer, label = Year),
              size = 3, fontface = "bold") +
    theme(legend.position = "bottom") +
    labs(color = "Mayer is CEO",
         x = "Employees", y = "Revenue (Millions)",
         title = "Yahoo Employees vs Revenues, 2004-2014") +
    scale_y_continuous(labels = scales::dollar) +
    scale_x_continuous(labels = scales::comma)
```




```
# i think this visualization could benefit from different legend placement (on top). Also I think the l
# more succint by removing/renaming the 'a' and 'b'
```

```
# second redrawing of case study 1
# we can put time back on the x-axis and instead make the y-axis a proportion of revenue/employees
p <- ggplot(data = yahoo,
            mapping = aes(x = Year, y = Revenue/Employees))

p + geom_vline(xintercept = 2012) +
    geom_line(color = "gray60", size = 2) +
    annotate("text", x = 2013, y = 0.44,
             label = " Mayer becomes CEO", size = 2.5) +
    labs(x = "Year\n",
         y = "Revenue/Employees",
         title = "Yahoo Revenue to Employee Ratio, 2004-2014")
```

```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use ‘linewidth‘ instead.
```




```
# i think this visualization is very good
```

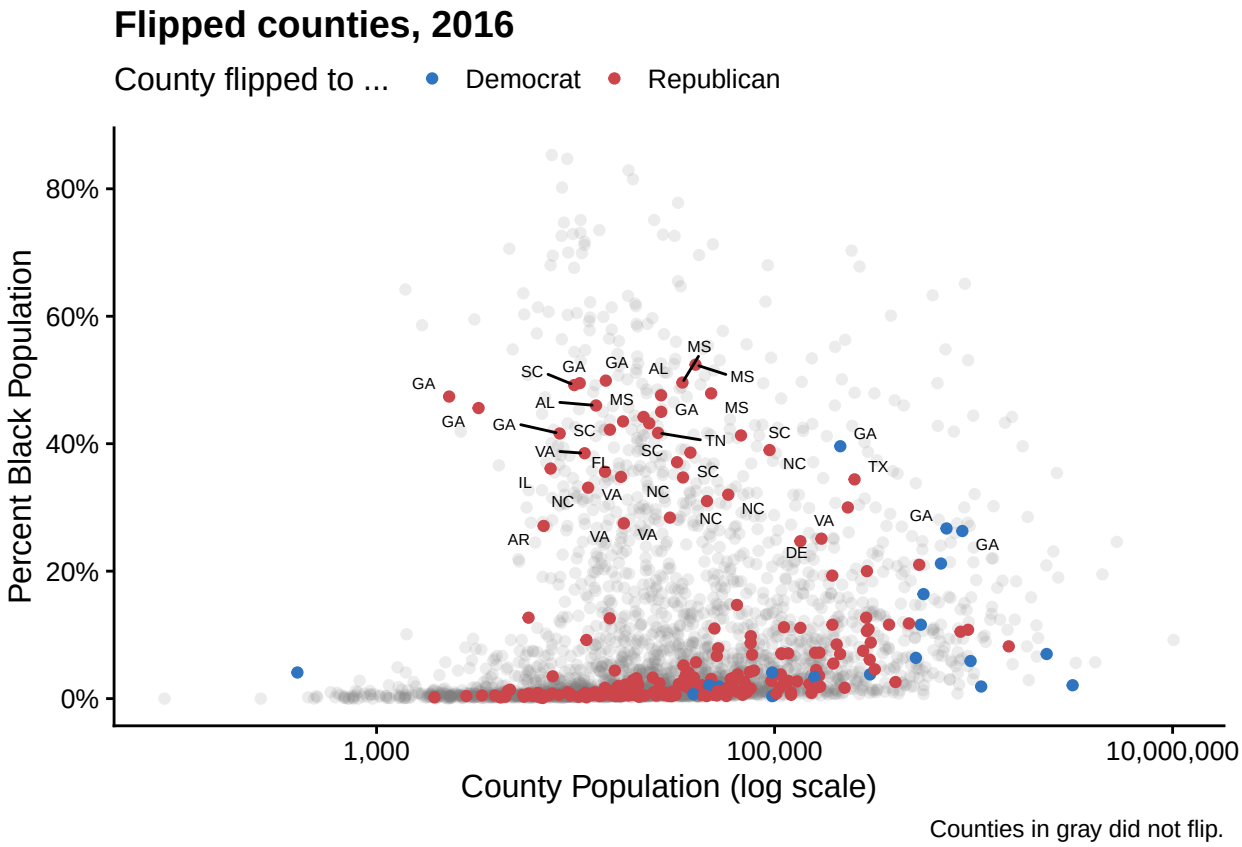
```
# themes are all functions within themselves. The cowplot package contains a theme which makes plots su
library(cowplot)
```

```
##
## Attaching package: 'cowplot'
```

```
## The following object is masked from 'package:ggthemes':
##
##     theme_map
```

```
p4 + theme_cowplot(12) + theme(legend.position="top")
```

```
## Warning: ggrepel: 2 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```


Flipped counties, 2016
County flipped to ...
Democrat
Republican
Percent Black Population
80%
60%
40%
20%
0%
1,000
100,000
10,000,000
County Population (log scale)
Counties in gray did not flip.
MS
SC
GA
AL
TN
FL
VA
NC
IL
AR
DE
TX