SRC R Session 3

Yiqing Xu

Bellwether counties and the 2020 Election

- "[c]ounties where a majority of voters have supported the election winner in several consecutive elections"
- Of the 19 counties that voted for the eventual winner in every presidential election from 1980 to 2016, Biden defeated Trump in only one.
- Is this anomalous? Why it may not be? If it were, what might be indications?

1. Load data and take a look

```
load("election2020.RData")
head(d)
```

```
##    county state bellwether      dem_16      dem_20 total_20
## 1 Autauga    AL          0 -0.48996917 -0.44418437    27770
## 2 Baldwin    AL          0 -0.57160111 -0.53762343   109679
## 3 Barbour    AL          0 -0.05568822 -0.07663054    10518
## 4    Bibb    AL          0 -0.55153646 -0.57727983     9595
## 5  Blount    AL          0 -0.80909020 -0.80002175    27588
## 6 Bullock    AL          0  0.50743100  0.49859094     4613
```

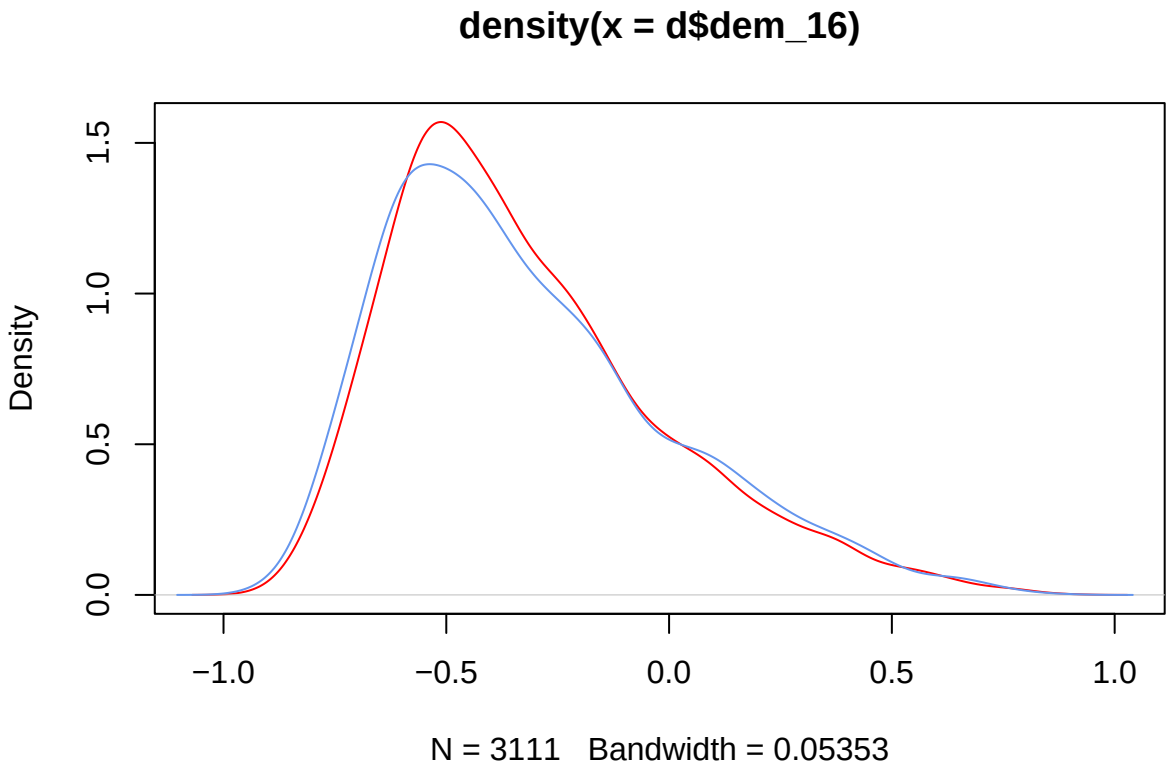
```
dim(d)
```

```
## [1] 3111    6
```

2. Ploting the distributions of democratic winning margin

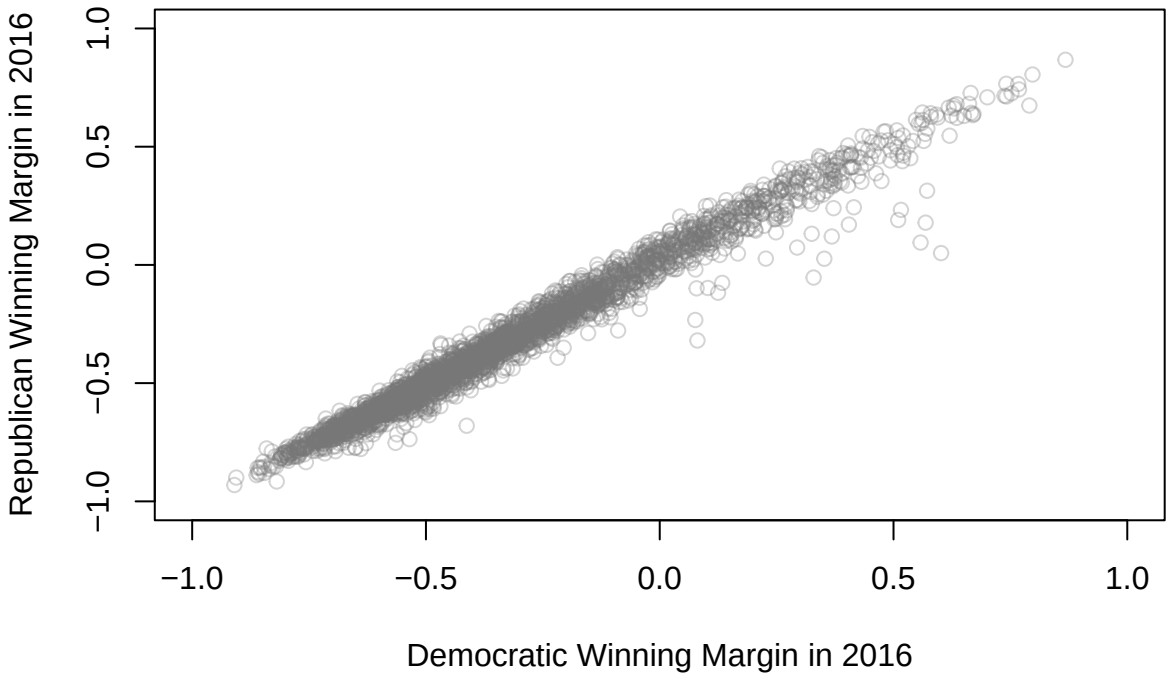
Plot the distributions of democratic winning margin in 2016 and 2020 in one plot. What do you find? Hint: overlay one density plot on the top of the other.

```
plot(density(d$dem_16), col = "red")
lines(density(d$dem_20), col = "cornflowerblue")
```



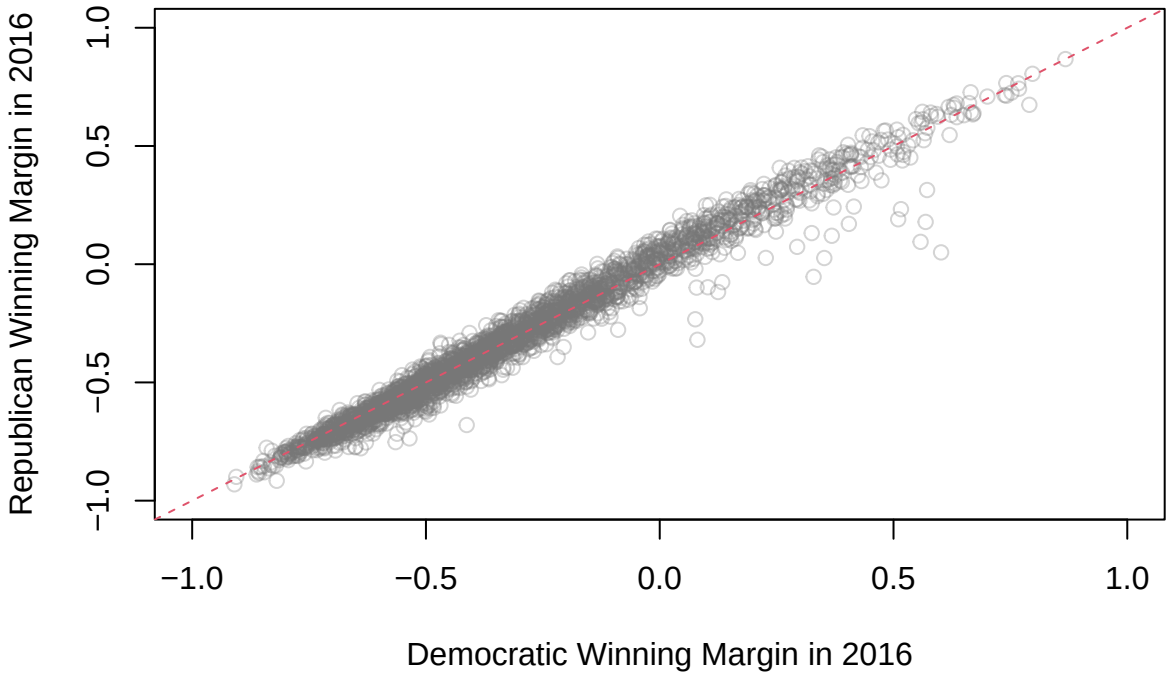
3. Draw a scatterplot of democratic winning margin in 2016 and 2020

```
plot(x = d$dem_16, y = d$dem_20, xlim = c(-1, 1), ylim = c(-1, 1),
     col = "#77777750", xlab = "Democratic Winning Margin in 2016",
     ylab = "Republican Winning Margin in 2016")
```



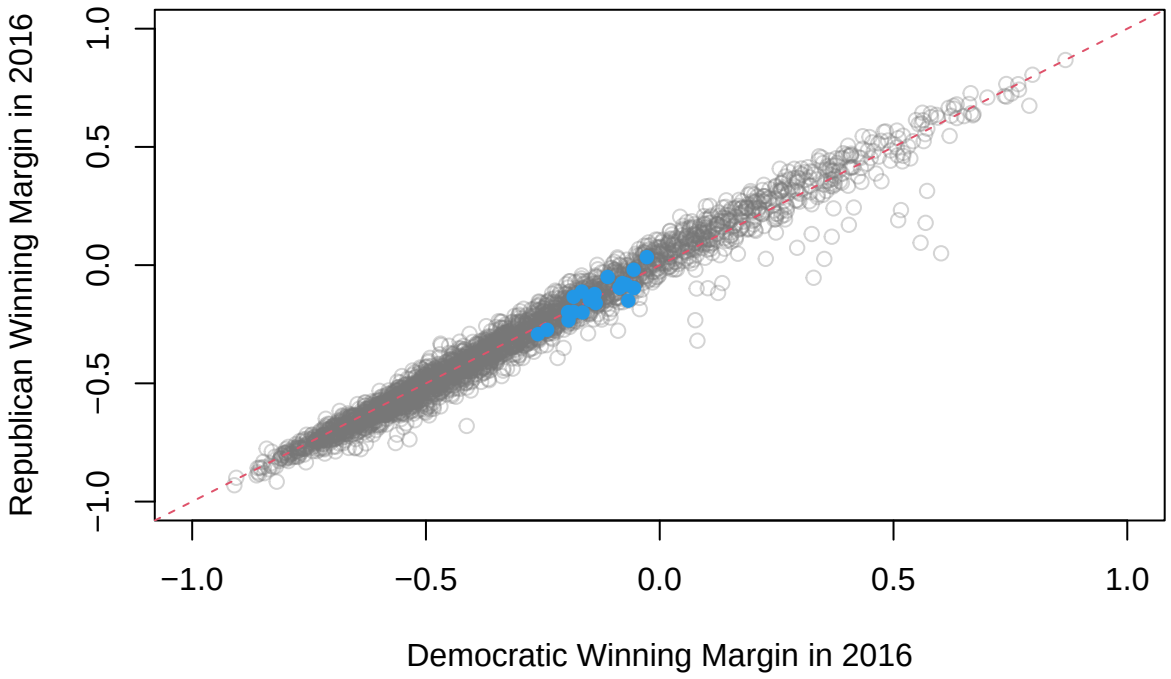
4. add a 45-degree line

```
plot(x = d$dem_16, y = d$dem_20, xlim = c(-1, 1), ylim = c(-1, 1),
     col = "#77777750", xlab = "Democratic Winning Margin in 2016",
     ylab = "Republican Winning Margin in 2016")
# Create a 45 degree line with a red color and dashed type
abline(a = 0, b = 1, col = 18, lwd = 1, lty = 2)
```



5. Mark the "bellwether" counties on the plot

```
plot(x = d$dem_16, y = d$dem_20, xlim = c(-1, 1), ylim = c(-1, 1),
     col = "#77777750", xlab = "Democratic Winning Margin in 2016",
     ylab = "Republican Winning Margin in 2016")
# Create a 45 degree line with a red color and dashed type
abline(a = 0, b = 1, col = 18, lwd = 1, lty = 2)
points(x = d$dem_16[d$bellwether==1], y = d$dem_20[d$bellwether==1],
       col = 12, pch = 16) # pch 16 sets the shape of the points to filled points
```



6.1. Fit a regression using non-bellwether counties and report your finding

Hint: `reg <- lm(y ~ x, data = d); summary(reg)`

```
# To run a regression use the lm() function.
# The first argument passed to the function is the dependent variable.
# Then you have to use the '~' symbol, followed by the independent variables.
# You also have to specify the dataset you are using.
reg <- lm(dem_20 ~ dem_16, data = d)
summary(reg) # Summarize the regression
```

```
##
## Call:
## lm(formula = dem_20 ~ dem_16, data = d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.57899 -0.02477  0.00190  0.02861  0.15069
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.009150   0.001312   6.976 3.71e-12 ***
## dem_16      1.029980   0.002973 346.471  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05091 on 3109 degrees of freedom
## Multiple R-squared:  0.9748, Adjusted R-squared:  0.9747
## F-statistic: 1.2e+05 on 1 and 3109 DF,  p-value: < 2.2e-16
```

6.2 Test if this relationship is different for bellwether counties

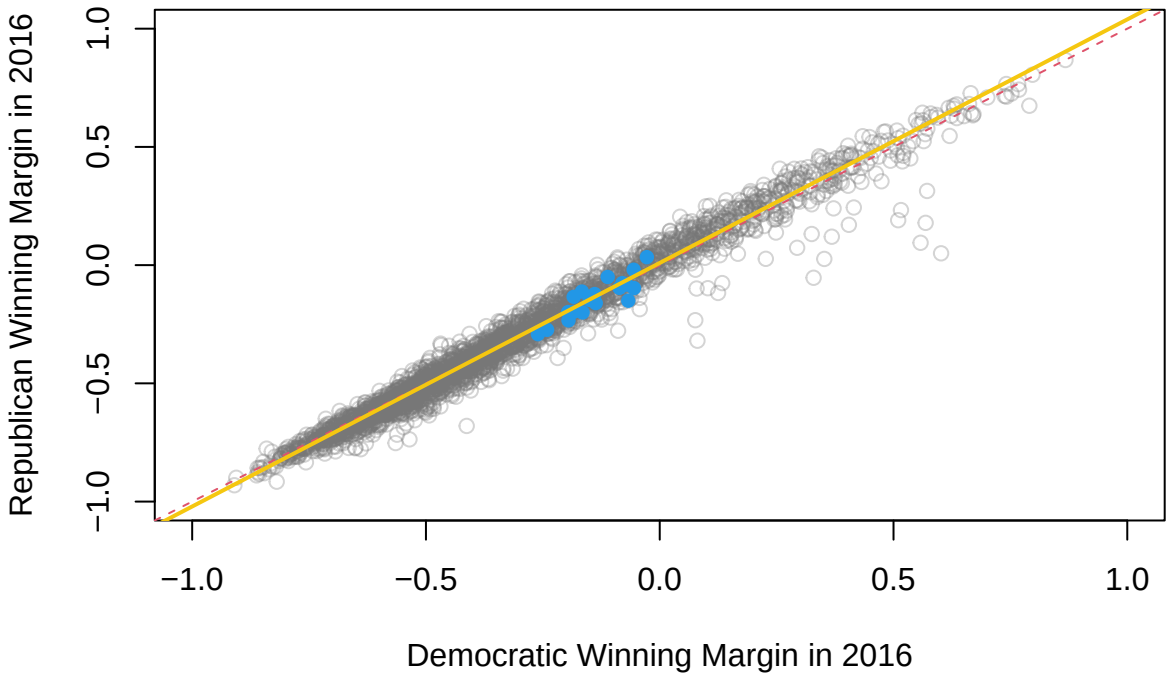
```
bw = which(d$bellwether==1)
reg_nonbell <- lm(dem_20 ~ dem_16, data = d[-bw,]) # remove bellwether counties
summary(reg_nonbell)
```

```
##
## Call:
## lm(formula = dem_20 ~ dem_16, data = d[-bw, ])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.57910 -0.02473  0.00219  0.02854  0.15062
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.009214   0.001319   6.988  3.4e-12 ***
## dem_16      1.030043   0.002980 345.646  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05097 on 3090 degrees of freedom
## Multiple R-squared:  0.9748, Adjusted R-squared:  0.9748
## F-statistic: 1.195e+05 on 1 and 3090 DF,  p-value: < 2.2e-16
```

7. Add the regression line to the plot

Hint: `abline(reg)`

```
plot(x = d$dem_16, y = d$dem_20, xlim = c(-1, 1), ylim = c(-1, 1),
     col = "#77777750", xlab = "Democratic Winning Margin in 2016",
     ylab = "Republican Winning Margin in 2016")
abline(a = 0, b = 1, col = 18, lwd = 1, lty = 2)
points(x = d$dem_16[d$bellwether==1], y = d$dem_20[d$bellwether==1],
       col = 12, pch = 16)
abline(reg_nonbell, col = 15, lwd = 2)
```



8. Predicting winning margin for "bellwether counties"

Using your regression model and the democratic winning margin in 2016 in "bellwether counties", predict the democratic winning margin in 2020 for those counties. Hint: extract the regression coefficients first using `reg$coefficients`.

```
coefs <- reg_nonbell$coefficients # store coefficients
beta0 <- coefs[1] # intercept
beta1 <- coefs[2] # coefficient on dem_16

predict <- beta0 + beta1*d$dem_16[bw]
```

9. Number of Trump wins for bellwether counties

Based on your model, how many of the 19 counties are predicted to have a Trump majority in 2020?

```
sum(predict < 0) # Democratic wins are indicated by values higher than 0.
```

```
## [1] 19
```

```
length(predict)
```

```
## [1] 19
```

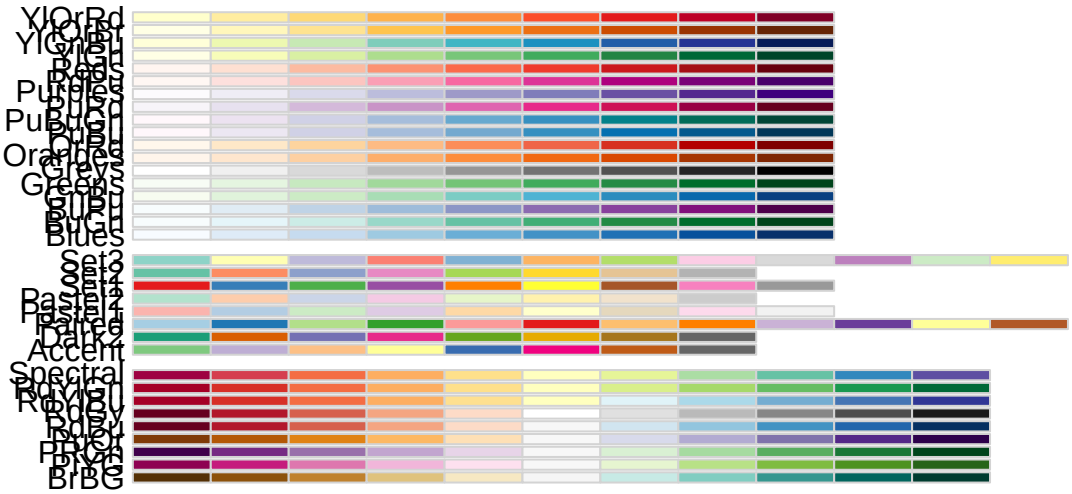
10. Calculate the number of Clinton-Biden countries and the number of total votes

```
q1 <- which(d$dem_16 >= 0 & d$dem_20>= 0) # Clinton-Biden counties
q1.n <-length(q1)
# Sum total votes and divide by a million to get neater number
q1.v <- sum(d$total_20[q1])/1e6
```

11. Beautify your plot;

Consider the following elements: - Change the size of each dot based on the number of total votes in that county. Hint: use the "cex" option - Add text in each of the four quadrant - Add grid to mark the four quadrant - Adjust the symbol and color for the data points - Add background color to the four quadrant

```
# install.packages("RColorBrewer")
library(RColorBrewer) # Get package for color palettes
display.brewer.all()  # Display all color palettes
```




```
mycol<-brewer.pal(4,"Pastel1") # Choose "Pastel1" palette

q1 <- which(d$dem_16 >= 0 & d$dem_20>= 0) # Clinton-Biden counties
(q1.n <- length(q1)) # Number of Clinton-Biden wins
```

```
## [1] 474
```

```
(q1.v <- sum(d$total_20[q1])/1e6) # Total votes
```

```
## [1] 84.46198
```

```
q2 <- which(d$dem_16 < 0 & d$dem_20>= 0) # Trump-Biden counties
(q2.n <- length(q2))
```

```
## [1] 63
```

```
(q2.v <- sum(d$total_20[q2])/1e6)
```

```
## [1] 9.805069
```

```
q3 <- which(d$dem_16 < 0 & d$dem_20 < 0) # Trump-Trump counties
(q3.n <- length(q3))
```

```
## [1] 2559
```

```
(q3.v <- sum(d$total_20[q3])/1e6)
```

```
## [1] 63.50516
```

```
q4 <- which(d$dem_16 >= 0 & d$dem_20 < 0) # Clinton-Trump counties
(q4.n <- length(q4))
```

```
## [1] 15
```

```r
(q4.v <- sum(d$total_20[q4])/1e6)
```

```
## [1] 0.395393
```

```r
plot(1, xlim = c(-1,1), ylim = c(-1, 1), type = "n",
     xlab = "Democratic Vote Margin, 2016",
     ylab = "Democratic Vote Margin, 2020")
# Add quadrants for the four different types of counties
polygon(x = c(0, 2, 2, 0), y = c(0, 0, 2, 2), col = paste0(mycol[1],"50"))
polygon(x = c(-2, 0, 0, -2), y = c(0, 0, 2, 2), col = paste0(mycol[2],"50"))
polygon(x = c(0, 2, 2, 0), y = c(0, 0, -2, -2), col = paste0(mycol[3],"50"))
polygon(x = c(-2, 0, 0, -2), y = c(0, 0, -2, -2), col = paste0(mycol[4],"50"))
box()
abline(v = 0, h = 0, col = 1, lty = 3) # Add grid lines
abline(v = seq(-1, 1, 0.25), h = seq(-1, 1, 0.25), col = "#77777750")
points(d$dem_16, d$dem_20, col = "#33333350",
       cex = sqrt(d$total_20/max(d$total_20)*10)) # Make point size depend on number of votes
points(d$dem_16, d$dem_20, col = "#77777750",
       cex = sqrt(d$total_20/max(d$total_20)*10), pch = 16)
bw <- which(d$bellwether == 1)
points(d$dem_16[bw], d$dem_20[bw], col = "white", pch = 16, cex = 1)
points(d$dem_16[bw], d$dem_20[bw], col = 2, pch = 16, cex = 0.8) # Include bw counties
abline(a = 0, b = 1, col = 2, lty = 3) # 45-degree line
text(0.5, 0.8, paste0("Clinton-Biden\n","(",q1.n," counties,\n",round(q1.v,1)," M voters)"))
text(-0.5, 0.8, paste0("Clinton-Biden\n","(",q2.n," counties,\n",round(q2.v,1)," M voters)"))
text(-0.5, -0.8, paste0("Clinton-Biden\n","(",q3.n," counties,\n",round(q3.v, 1)," M voters)"))
text(0.5, -0.8, paste0("Clinton-Biden\n","(",q4.n," counties,\n",round(q4.v,1)," M voters)"))
```