Analysis Supplement: Georgia Gubernatorial Election Data

Noah Woodward

Motivation:
In 2016, I witnessed a voter turned away at a Cobb County polling place on a technicality that was overlooked for me. In the aftermath of the 2018 Georgia gubernatorial election, I sought to identify similar cases of voter suppression in available data. I've included that exploration below and on the following pages.

Background:
I voted in the 2016 presidential election as a resident of Cobb County, Georgia. At my Cobb County polling place, I witnessed a voter turned away because she was in the wrong location. As I saw a volunteer point to the address on the voter's drivers' license, I realized I was soon to be rejected for the same reason. I had recently moved from Fulton County to Cobb County, and I had yet to update my driver's license to reflect my new address. I considered leaving the line to begin driving to a Fulton County polling place, but I decided to stay since I had already invested half an hour in my current line.

Reaching the front of the line, I handed over my license and voter form. The documents I presented were given only a cursory glance and deemed sufficient. The process for the review of my documents differed substantially from the attention paid to the woman in front of me. She was a black woman, and I am a white male. I was allowed to vote and she was not. This experience, along with other reports of voter suppression and the founding of related nonprofits, motivated me to explore available voting data.

Analysis:
- While a voter turned away leaves no trace of their attempt to vote, and not all provisional ballots are eventually counted, data kept on counted provisional ballots can provide an indication of patterns in voter experience.
- Despite similar overall vote totals, Abrams received over 2.5x as many votes via provisional ballots than did Kemp.
- Abrams' high share of the provisional ballot total outpaced her support from in-person and absentee balloting in every major county.
- In counties where substantiated reports of voter suppression took place, Abrams' share of the provisional ballots most exceeded her overall share of the vote.

Conclusion:
Without controlling for other factors that may necessitate casting a provisional ballot, conclusions cannot be drawn directly from this analysis. However, the gap identified between Abrams' share of provisional ballots and her overall share of the vote may provide additional support for reports of voter suppression.

In the following pages, I have included visualizations, mapping, and code supporting these conclusions to demonstrate my interest in using data for the public good in Georgia.

Review of 2018 Georgia Gubernatorial Election Data

Noah Woodward

Data sourced from the Office of the Georgia Secretary of State: Link to Data

A full review of the data, along with code required to produce these visuals, can be found in an accompanying appendix.

Outline:

1) Review of Georgia's largest counties, and provisional ballots counted in those counties
2) Analysis of the relationship between Kemp/Abrams share of provisional and overall ballot totals
3) Review of Abrams' relatively high provisional vote share in large Georgia counties
4) Mapping of Abrams' vote share and relatively high provisional vote share in Metro Atlanta counties

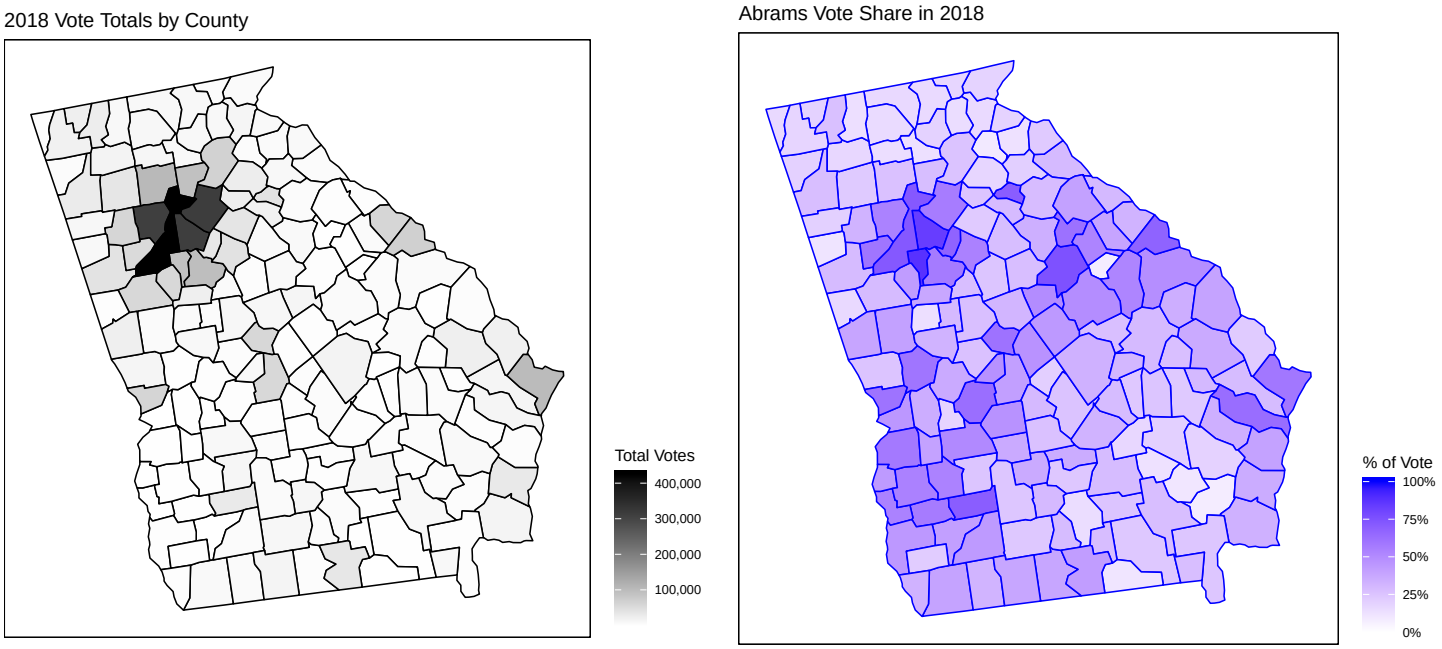
What were the largest counties (by vote totals), and what did the provisional ballot totals look like in those counties?

County	Total	TotalVotesAbrams	TotalVotesKemp	ProvisionalAbrams	ProvisionalKemp
Fulton	423,788	306,589	112,991	1,770	200
Gwinnett	314,918	178,097	132,998	1,518	670
DeKalb	312,741	261,042	48,923	1,617	144
Cobb	311,814	168,767	138,852	1,060	358
Cherokee	106,411	28,047	76,700	33	78
Chatham	103,543	61,059	41,425	145	56
Henry	98,558	56,485	41,364	139	54
Forsyth	93,298	26,092	65,845	18	39
Clayton	92,212	80,971	10,868	349	21
Richmond	70,155	47,531	22,076	94	18

In Georgia, an outsized portion of the vote comes from a few large counties. Despite having 159 counties in total, Georgia's four largest counties (Fulton, Gwinnett, DeKalb, and Cobb) made up over a quarter of the overall vote in 2018. Abrams won 8 of the 10 largest Georgia counties in the '18 election, and it appears that she won a large share of provisional ballots cast in these counties.

Mapping: Where are these counties?

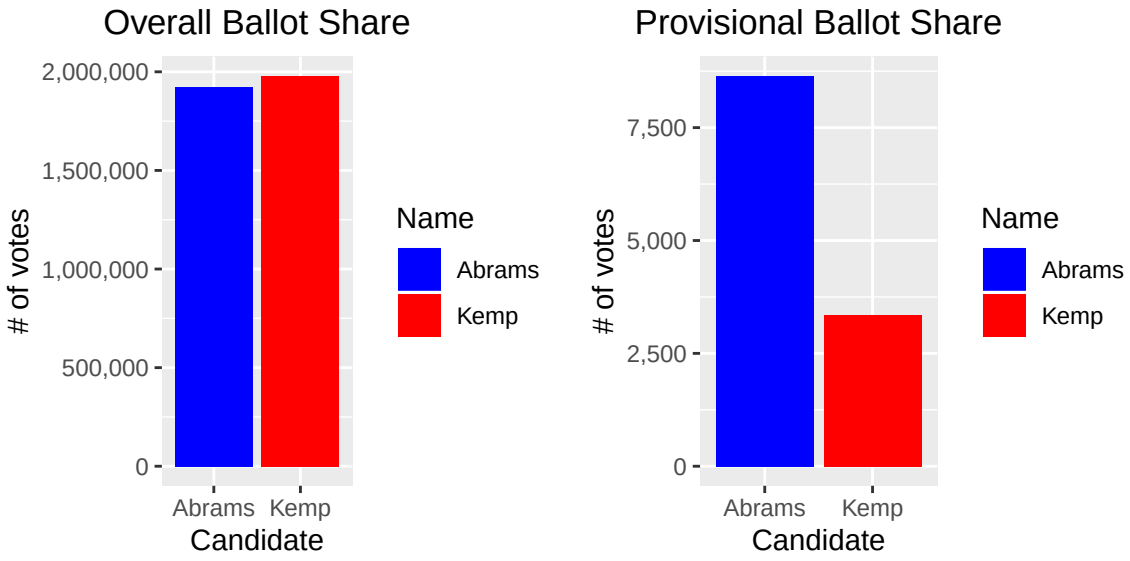
Georgia's four largest counties make up Metro Atlanta, an area Abrams saw her largest majorities in 2018.



In looking at reports of voter suppression in 2018, it follows that provisional ballots cast in Metro Atlanta counties carried by Abrams could be worth a closer look.

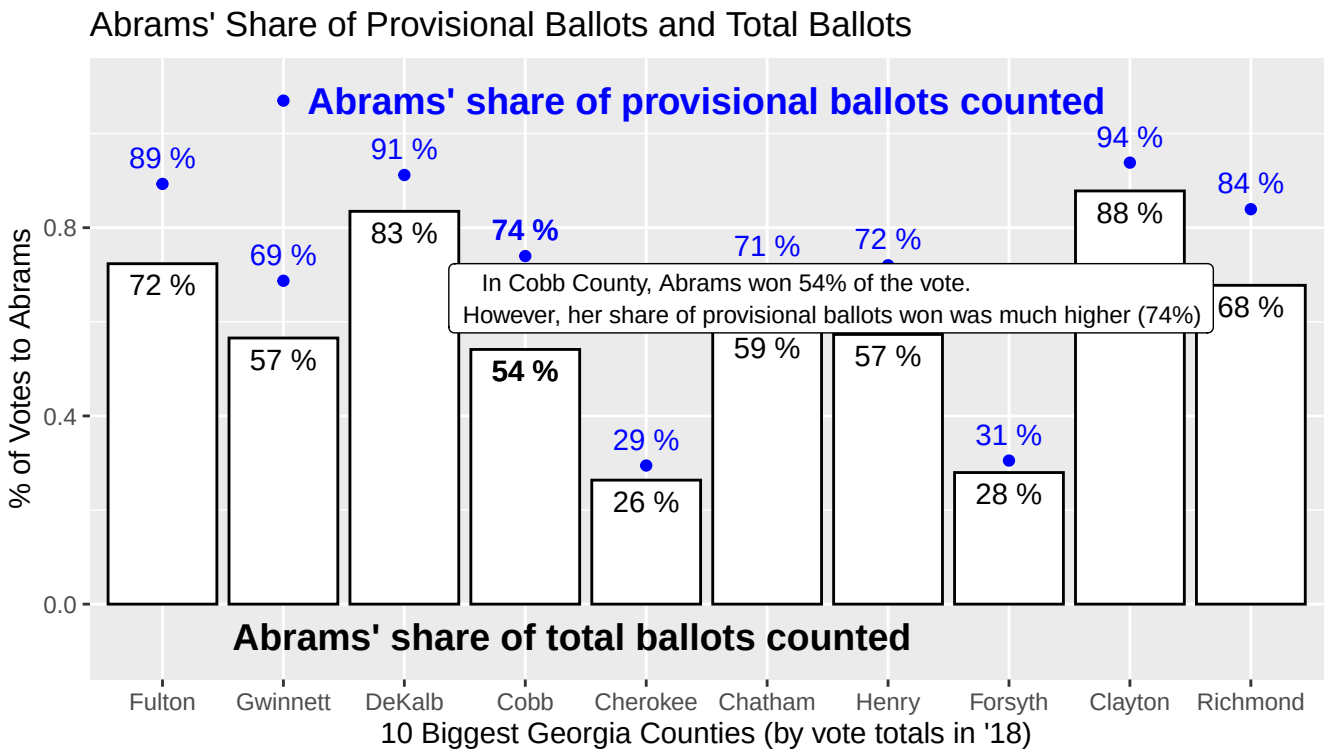
How did Abrams' share of the provisional ballot vote compare to her overall share?

In an election with almost 4 million votes cast, Kemp defeated Abrams by just 54,273 votes. However, Abrams received over 2.5x as many provisional ballot votes as Kemp. Abrams received 8,642 counted votes from provisional ballots (to Kemp's 3,346).



Did Abrams' share of the provisional ballot track her share of the overall ballot?

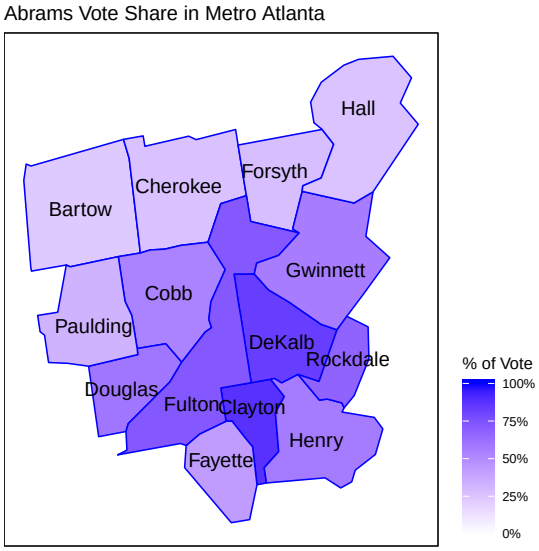
Abrams' share of the provisional vote tracked above her overall vote share in each of Georgia's 10 largest counties:

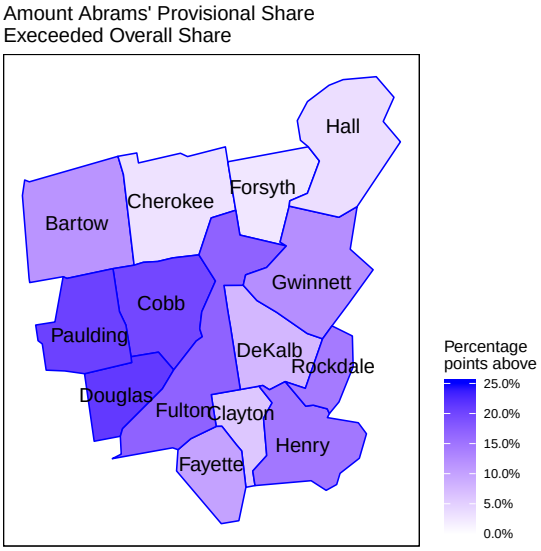


Mapping Metro Atlanta:

Abrams saw her largest majorities won in in Metro Atlanta counties (Fulton, Cobb, Gwinnett, DeKalb, Clayton). She also saw a relatively higher proportion of the provisional ballot won in these same counties, which tracks with reports of voting issues.

Faulty machines were reported in Fulton, Cobb, and Gwinnett counties. Voters also reported long lines in Cobb county after the state closed 12 polling places ahead of the 2018 election.





Code Appendix (R/RMarkdown)

```
#Load required packages in R:
library(rmarkdown)
library(dplyr)
library(usmap)
library(stringr)
library(scales)
library(readxl)
library(ggplot2)
library(knitr)
library(kableExtra)

#Load the data (cleaning below):
dat <- read_excel("Gov_Elect_County_Data_18.xls",skip=2,sheet=3)
```

```
# Renaming columns to appropriate gubernatorial candidate (via labeling in raw excel file)
KempCols <- c("3|4|5|6|7")
AbramsCols <- c("8|9|10|11|12")
MetzCols <- c("13|14|15|16|17")

# Clean up spaces in column names and add candidate names
dat <- dat %>% rename_all(funs(str_replace_all(., c("[.][.][.]|[[:space:]]") ,""))) %>%
          rename_all(funs(str_replace(., MetzCols,"Metz"))) %>%
          rename_all(funs(str_replace(., AbramsCols,"Abrams"))) %>%
          rename_all(funs(str_replace(., KempCols,"Kemp")))

# Convert numerical vote total columns from character to integers
dat <- dat %>% mutate_at(vars(-County),as.integer)
# View snapshot of data
summary(dat)
```

```
##     County          RegisteredVoters  ElectionDayKemp  AbsenteebyMailKemp
##  Length:160         Min.   :   1211   Min.   :   133   Min.   :   15.0
##  Class :character   1st Qu.:   6277   1st Qu.:  1196   1st Qu.:  110.0
##  Mode  :character   Median :  13562   Median :  2413   Median :  229.5
##                     Mean   :  80357   Mean   : 11764   Mean   : 1048.8
##                     3rd Qu.:  31196   3rd Qu.:  5856   3rd Qu.:  481.0
##                     Max.   :6428581   Max.   :941129   Max.   :83907.0
##  AdvanceinPersonKemp ProvisionalKemp   TotalVotesKemp    ElectionDayAbrams
##  Min.   :   185      Min.   :   0.00   Min.   :    350   Min.   :    47.0
##  1st Qu.:  1442      1st Qu.:   1.00   1st Qu.:   2588   1st Qu.:   521.2
##  Median :  3167      Median :   4.00   Median :   5926   Median :   949.0
##  Mean   : 11875      Mean   :  41.83   Mean   :  24730   Mean   : 10591.2
##  3rd Qu.:  6155      3rd Qu.:  13.00   3rd Qu.:  11967   3rd Qu.:  2449.8
##  Max.   :950026      Max.   :3346.00   Max.   :1978408   Max.   :847296.0
##  AbsenteebyMailAbrams AdvanceinPersonAbrams ProvisionalAbrams TotalVotesAbrams
##  Min.   :     7.0     Min.   :    21.0      Min.   :   0.00   Min.   :    107
##  1st Qu.:   110.2     1st Qu.:   509.2      1st Qu.:   0.00   1st Qu.:   1182
##  Median :   213.5     Median :  1060.0      Median :   3.00   Median :   2194
```

```
##  Mean   :  1720.2     Mean   : 11626.6      Mean   : 108.03   Mean   :  24046
##  3rd Qu.:   543.0     3rd Qu.:  2866.2      3rd Qu.:  10.25   3rd Qu.:   5799
##  Max.   :137616.0     Max.   :930131.0      Max.   :8642.00   Max.   :1923685
##  ElectionDayMetz    AbsenteebyMailMetz AdvanceinPersonMetz ProvisionalMetz
##  Min.   :    0.00   Min.   :   0.00    Min.   :    0.00    Min.   :  0.000
##  1st Qu.:   11.75   1st Qu.:   1.00    1st Qu.:    6.00    1st Qu.:  0.000
##  Median :   31.00   Median :   3.50    Median :   19.00    Median :  0.000
##  Mean   :  308.73   Mean   :  25.66    Mean   :  129.76    Mean   :  1.288
##  3rd Qu.:  109.25   3rd Qu.:   8.00    3rd Qu.:   57.25    3rd Qu.:  0.000
##  Max.   :24698.00   Max.   :2053.00    Max.   :10381.00    Max.   :103.000
##  TotalVotesMetz        Total
##  Min.   :    1.0   Min.   :    921
##  1st Qu.:   20.0   1st Qu.:   3991
##  Median :   55.5   Median :   8756
##  Mean   :  465.4   Mean   :  49242
##  3rd Qu.:  172.2   3rd Qu.:  19400
##  Max.   :37235.0   Max.   :3939328
```

```
# Check to ensure "Total" row is the sum of all county rows (it is)
totals_test <- as.data.frame(dat %>% filter(County == "Total:") %>%
      select(-County) ==
      dat %>% filter(County != "Total:") %>% summarise_at(vars(-County), funs(sum)))
totals_test
```

```
##   RegisteredVoters ElectionDayKemp AbsenteebyMailKemp AdvanceinPersonKemp
## 1             TRUE            TRUE               TRUE                TRUE
##   ProvisionalKemp TotalVotesKemp ElectionDayAbrams AbsenteebyMailAbrams
## 1            TRUE           TRUE              TRUE                TRUE
##   AdvanceinPersonAbrams ProvisionalAbrams TotalVotesAbrams ElectionDayMetz
## 1                  TRUE              TRUE             TRUE            TRUE
##   AbsenteebyMailMetz AdvanceinPersonMetz ProvisionalMetz TotalVotesMetz Total
## 1               TRUE                TRUE            TRUE           TRUE  TRUE
```

```
# Remove "Total" row before looking at county-level data
dat <- dat %>% filter(County != "Total:")

# Add "vote share" columns for candidates
dat <- dat %>%
  mutate(KempShare = TotalVotesKemp / (TotalVotesKemp + TotalVotesAbrams + TotalVotesMetz),
         AbramsShare = TotalVotesAbrams / (TotalVotesKemp + TotalVotesAbrams + TotalVotesMetz),
         MetzShare = TotalVotesMetz / (TotalVotesKemp + TotalVotesAbrams + TotalVotesMetz))

# Add "provisional vote share" columns for candidates
dat <- dat %>%
  mutate(ProvShareKemp = ProvisionalKemp / (ProvisionalKemp + ProvisionalAbrams + ProvisionalMetz),
         ProvShareAbrams = ProvisionalAbrams / (ProvisionalKemp + ProvisionalAbrams + ProvisionalMetz),
         ProvShareMetz = ProvisionalMetz / (ProvisionalKemp + ProvisionalAbrams + ProvisionalMetz),
         ProvisionalTotal = ProvisionalKemp + ProvisionalAbrams + ProvisionalMetz)
```

```
dat %>% arrange(desc(Total)) %>%
  select(County, Total, TotalVotesAbrams,TotalVotesKemp,
                      ProvisionalAbrams, ProvisionalKemp) %>%
  top_n(10, Total) %>%
  kable(, format.args = list(big.mark = ",")) %>% kable_styling(latex_options = "striped")
```

County	Total	TotalVotesAbrams	TotalVotesKemp	ProvisionalAbrams	ProvisionalKemp
Fulton	423,788	306,589	112,991	1,770	200
Gwinnett	314,918	178,097	132,998	1,518	670
DeKalb	312,741	261,042	48,923	1,617	144
Cobb	311,814	168,767	138,852	1,060	358
Cherokee	106,411	28,047	76,700	33	78
Chatham	103,543	61,059	41,425	145	56
Henry	98,558	56,485	41,364	139	54
Forsyth	93,298	26,092	65,845	18	39
Clayton	92,212	80,971	10,868	349	21
Richmond	70,155	47,531	22,076	94	18

```
# Prep data for map plotting (join to usmap dataset)
data(countypop)
countypop <- countypop %>% filter(abbr == "GA")
countypop <- countypop %>% mutate_at(vars(county), str_replace_all, pattern = ' County', replacement = "")

# Add column for Abrams' relative share of the provisional vote relative to overall vote
dat <- dat %>% mutate(AbramsRelProv = ProvShareAbrams - AbramsShare)

# Create dataset with county FIPS codes for plotting
mapdf <- inner_join(countypop, dat, by = c("county" = "County"))

# Map plotting
# Where do Georgia's votes come from?
plot_usmap("counties", data = mapdf, values = "Total", include=c("GA"), color="black") +
  scale_fill_continuous( low = "white", high = "black", name = "Total Votes", labels=comma) +
  theme(legend.position = "right") +
  theme(panel.background = element_rect(colour = "black")) +
  labs(title = "2018 Vote Totals by County")

#Abrams share of the vote in 2018
plot_usmap("counties", data = mapdf, values = "AbramsShare", include=c("GA"), color="blue") +
  scale_fill_continuous( low = "white", high = "blue",
  name = "% of Vote", labels=percent, limits = c(0,1)) +
  theme(legend.position = "right") + theme(panel.background = element_rect(colour = "black")) +
  labs(title = "Abrams Vote Share in 2018")
```

```
#Comparison of Kemp/Abrams overall GA vote share to provisional GA vote share (Chart)
totals <- dat %>% summarise_at(vars(ProvisionalKemp, TotalVotesKemp,
                                    ProvisionalAbrams, TotalVotesAbrams), funs(sum))
plot_data <- data.frame(Name = c("Kemp","Abrams"),
                        TotalVotes = c(totals$TotalVotesKemp, totals$TotalVotesAbrams),
                 ProvisionalVotes = c(totals$ProvisionalKemp, totals$ProvisionalAbrams))

ggplot(data=plot_data, aes(x=Name,y=TotalVotes,fill=Name)) + geom_col() +
      ggtitle("Overall Ballot Share") + labs(x="Candidate", y="# of votes") +
      scale_y_continuous(label=comma) + scale_fill_manual(values = c("Abrams"="blue","Kemp"="red")) +
      theme(plot.title = element_text(hjust = 0.5))

ggplot(data=plot_data, aes(x=Name,y=ProvisionalVotes,fill=Name)) + geom_col() +
      ggtitle("Provisional Ballot Share") + labs(x="Candidate", y="# of votes") +
      scale_y_continuous(label=comma) + scale_fill_manual(values = c("Abrams"="blue","Kemp"="red")) +
      theme(plot.title = element_text(hjust = 0.5))
```

```
#Abrams' share of the provisional vote total vs. overall vote total in Georgia's 10 largest counties (Chart)
ggplot(dat=dat %>% top_n(10, Total), aes(x=reorder(County,-Total))) +
  geom_col(aes(y=AbramsShare), fill="white",color="black") +
  geom_point(aes(y=ProvShareAbrams),color="Blue") +
  geom_text(aes(y=ProvShareAbrams, label=paste(round(ProvShareAbrams,2)*100,"%")), vjust=-.75,
    check_overlap = T,color="Blue") +
  geom_text(aes(y=AbramsShare, label=paste(round(AbramsShare,2)*100,"%")), vjust=1.5, check_overlap = T) +
  geom_text(data=dat %>% top_n(10, Total) %>% filter(County=="Gwinnett"),
    aes(y=1.07, label="Abrams' share of provisional ballots counted"),
      size=5,hjust=-.03, color="Blue",fontface="bold") +
  geom_text(data=dat %>% top_n(10, Total) %>% filter(County=="Gwinnett"),
    aes(y=-.07, label="Abrams' share of total ballots counted"), size=5,hjust=.075,fontface="bold") +
  ggtitle("Abrams' Share of Provisional Ballots and Total Ballots")+
    labs(x="10 Biggest Georgia Counties (by vote totals in '18)", y="% of Votes to Abrams") +
  geom_label(data=dat %>% top_n(10, Total) %>% filter(County=="Cobb"),
    aes(y=.65, label=paste("In Cobb County, Abrams won 54% of the vote.\n",
      "However, her share of provisional ballots won was much higher (74%)")), size=3,hjust=.1) +
  geom_text(data=dat %>% top_n(10, Total) %>% filter(County=="Cobb"),
    aes(y=ProvShareAbrams, label=paste(round(ProvShareAbrams,2)*100,"%")),
      vjust=-.75, fontface="bold",color="Blue") +
  geom_text(data=dat %>% top_n(10, Total) %>% filter(County=="Cobb"),
    aes(y=AbramsShare, label=paste(round(AbramsShare,2)*100,"%")), vjust=1.5, fontface="bold") +
  ylim(c(-.1,1.1)) +
  geom_point(data=dat %>% top_n(10, Total) %>% filter(County=="Gwinnett"), colour="blue",
    aes(y=1.07))
```

```
#Abrams share of the vote in Metro Atlanta (Chart)
plot_usmap("counties", data = mapdf, values = "AbramsShare",
           include=c("13135","13121","13067","13089","13057","13117","13113",
                     "13151","13063","13097","13223","13015","13139","13247"),
           labels=TRUE, label_color="black", color="blue") +
  scale_fill_continuous( low = "white", high = "blue",
                         name = "% of Vote", labels=percent, limits = c(0,1)) +
  theme(legend.position = "right") +
  theme(panel.background = element_rect(colour = "black")) +
  labs(title = "Abrams Vote Share in Metro Atlanta")

#Abrams relative share of provisional vs. overall votes (Chart)
plot_usmap("counties", data = mapdf, values = "AbramsRelProv",
           include=c("13135","13121","13067","13089","13057","13117","13113",
                     "13151","13063","13097","13223","13015","13139","13247"),
           labels=TRUE, label_color="black", color="blue") +
  scale_fill_continuous( low = "white", high = "blue",
                         name = "Percentage\npoints above", labels=percent, limits = c(0,.25)) +
  theme(legend.position = "right") +
  theme(panel.background = element_rect(colour = "black")) +
  labs(title = "Amount Abrams' Provisional Share\nExceeded Overall Share")
```