Graph Challenge 3

Aryah Rao

2024-09-27

**Instructions**: For this graph challenge, use the florida.csv and florida\_map.csv datasets to create a map of Florida counties that shows the absolute difference in the received versus expected vote *count* for Reform candidate for US president, Pat Buchanan, in 2000 given vote totals for Ross Perot in 1996. The columns lib\_00 and lib\_96 give you these totals, respectively.

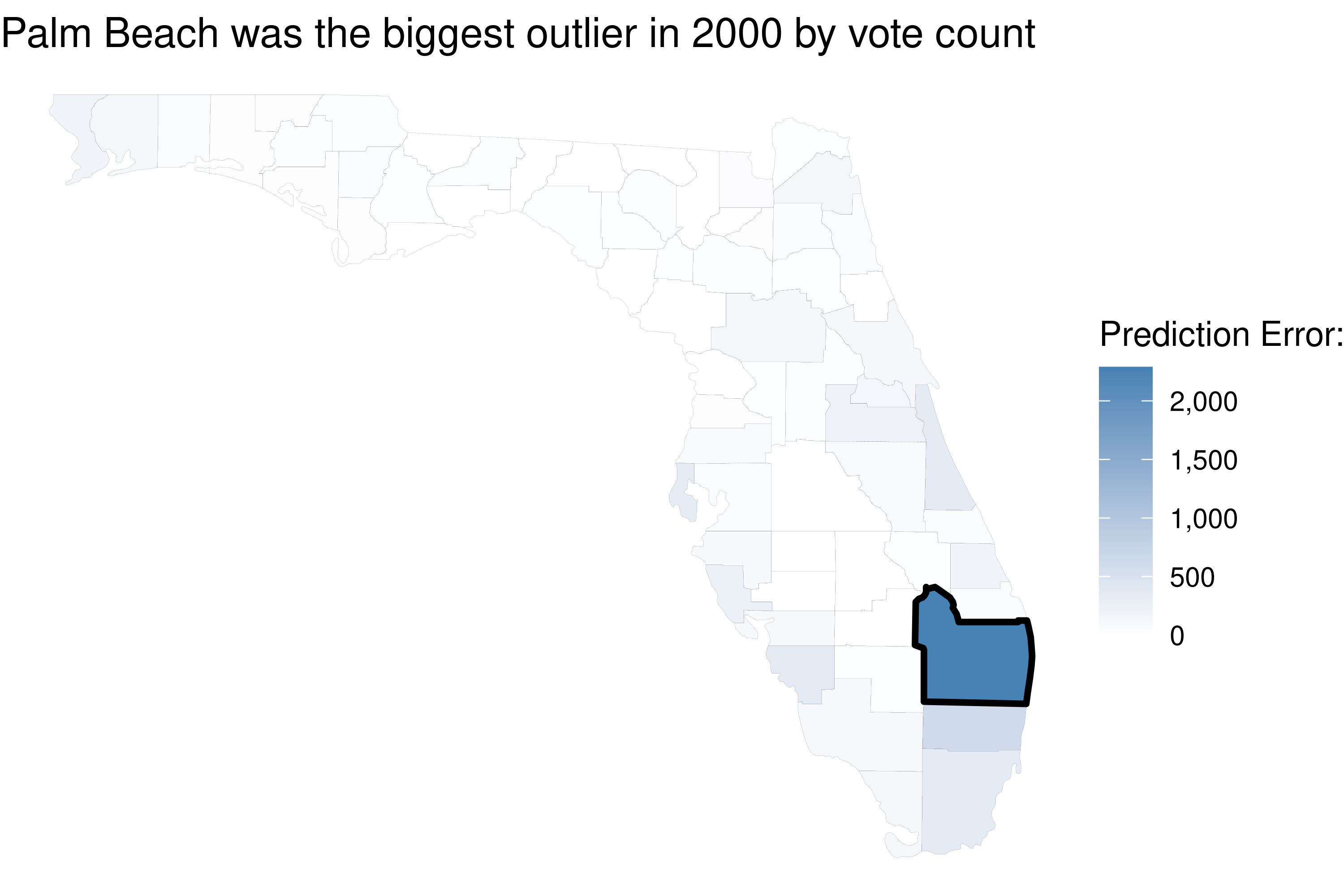
Next, create a similar map, but this time show the absolute difference in the received versus expected vote share received by Buchanan in 2000 given Perot’s vote share in 1996. The columns plib\_00 and plib\_96 give you these totals, respectively.

I’ve written parts of the code for you, but you’ll need to complete the rest of it.

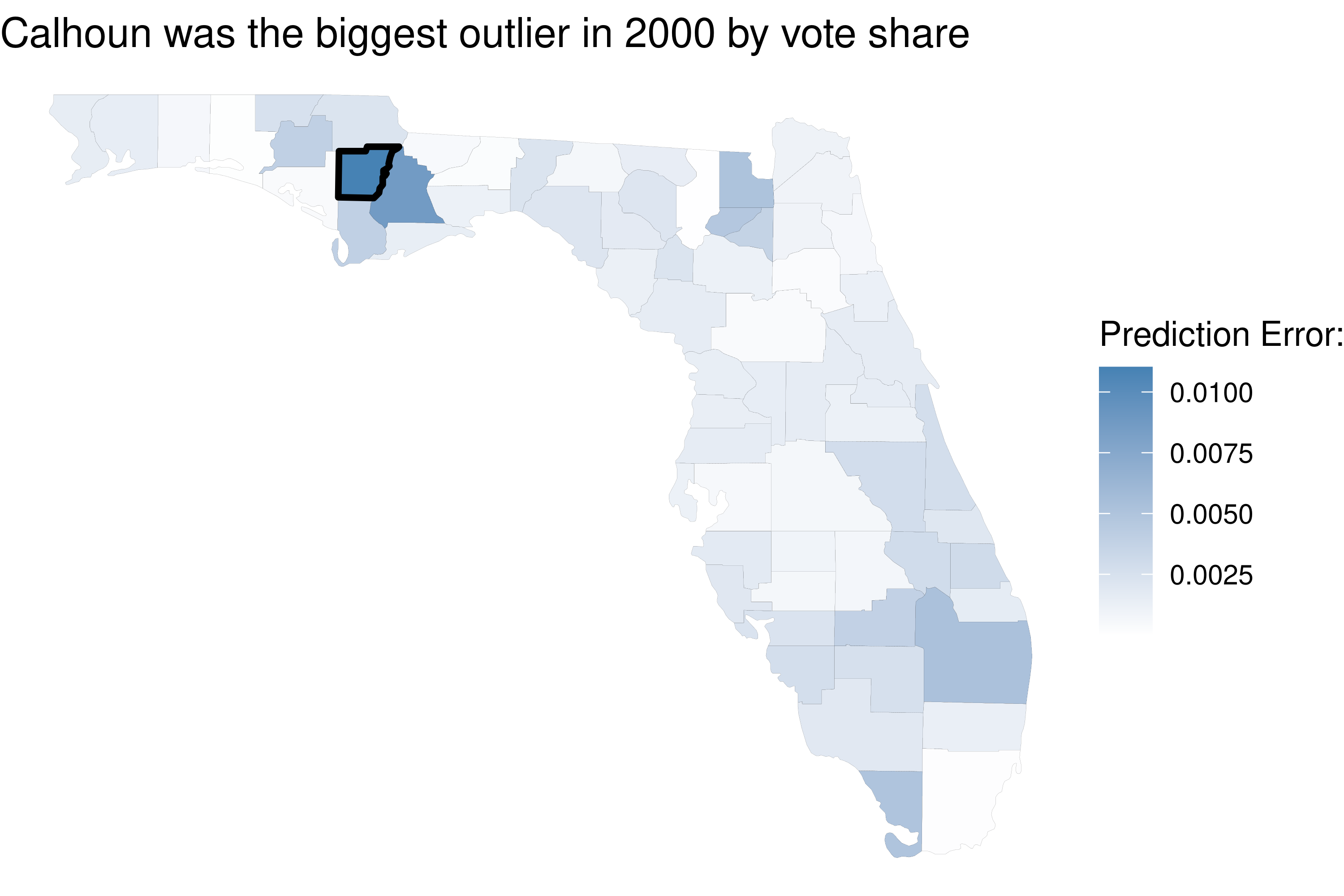
Once you’ve made these figures, provide a short written response to the following questions: (1) ***Why does looking at the total count of the third party vote alter our conclusions about whether Palm Beach is an outlier in the 2000 US Presidential election as compared to looking at the third party vote share of the total vote;*** and (2) ***which way of presenting the data provides a more accurate picture of Palm Beach relative to other counties?***

## Open the tidyverse:  
library(tidyverse)  
  
## Read in the data:  
fl <- read\_csv("../\_data/florida.csv", show\_col\_types = FALSE)  
fl\_map <- read\_csv("../\_data/florida\_map.csv", show\_col\_types = FALSE)  
  
## Write the code to merge the fl and fl\_map data:  
fl\_map = left\_join(fl\_map, fl, by='county')  
  
## Function to get absolute residual difference:  
my\_fun <- function(x1, x2) {  
 lm(x1 ~ x2) |> resid() |> abs()  
}  
  
## Function to return the county that's the biggest outlier:  
biggest\_ol <- function(data, x1, x2) {  
 data |>  
 mutate(x = my\_fun(!!enquo(x1), !!enquo(x2))) |>  
 filter(x == max(x)) |>  
 pull(county) |>  
 unique()  
}

## Get the biggest outlier county:  
biggest\_ol(  
 fl, lib\_00, lib\_96  
) -> outlier  
  
## Make the first map of Florida:  
ggplot(fl\_map) +  
 aes(  
 x = long, # what goes here?  
 y = lat, # what goes here?  
 fill = my\_fun(lib\_00, lib\_96),  
 group = group # do we need to group by something?  
 ) +  
 geom\_polygon(  
 color = "black",  
 size = 0.01  
 ) +  
 geom\_polygon( # this makes the outlier stand out  
 data = . %>% filter(county == outlier),  
 color = "black",  
 size = 1,  
 alpha = 0  
 ) +  
 theme\_void() +  
 scale\_fill\_gradient( # this updates the color gradient  
 low = "white",   
 high = "steelblue",  
 labels = scales::comma  
 ) +  
 labs(  
 title = paste0(  
 outlier,  
 " was the biggest outlier in 2000 by vote count"  
 ),  
 fill = "Prediction Error:"  
 )



## Write code to compute the difference between  
## the observed and expected vote share by county:  
## Get the biggest outlier county (by vote share):  
biggest\_ol(  
 fl, plib\_00, plib\_96  
) -> outlier\_share # Store biggest outlier by vote share  
  
ggplot(fl\_map) +  
 aes(  
 x = long,   
 y = lat,   
 fill = my\_fun(plib\_00, plib\_96),   
 group = group   
 ) +  
 geom\_polygon( # this makes the outlier stand out  
 color = "black",   
 size = 0.01  
 ) +  
 geom\_polygon(   
 data = . %>% filter(county == outlier\_share),  
 color = "black",   
 size = 1,   
 alpha = 0  
 ) +  
 theme\_void() +  
 scale\_fill\_gradient( # this updates the color gradient  
 low = "white",   
 high = "steelblue",  
 labels = scales::comma  
 ) +  
 labs(  
 title = paste0(  
 outlier\_share,   
 " was the biggest outlier in 2000 by vote share"  
 ),  
 fill = "Prediction Error:"  
 )



**Your answer**: (1) Looking at the total count of votes emphasizes the sheer magnitude of the difference, whereas, looking at the vote share accounts for the size of the electorate, giving a clearer picture of how the third-party vote compares relative to the total number of votes cast. While Palm Beach might stand out more by vote count, it may appear less extreme when considering the vote share (percentage). (2) Presenting the data by vote share offers a clearer picture. It adjusts for variations in county size and voter turnout, allowing for a more meaningful comparison across counties. This approach accounts for how unusually large or small the third-party vote was relative to the total votes cast, rather than highlighting the overall number of voters in each county.