Explaining Covid-19 Vaccinations in Adults with 2020 Voting and Senior Population

Kale LaMont Perry

State University of New York at Fredonia

30 November, 2021

Abstract

It has been widely reported in recent months that the Covid-19 Pandemic has largely become a ‘pandemic of the unvaccinated’. It is also widely reported that the unvaccinated population skews heavily republican. Additionally, it is widely reported that people age 65 and over are the highest risk age group to get seriously ill and die from Covid-19. This study investigates the explanatory value of votes in the 2020 presidential election and population of seniors (age 65 and over) on actual vaccination rates among adults.

Keywords: Covid-19, pandemic, vaccination rates, 2020 presidential election, seniors

Introduction

I wondered if the risks and experiences of being a senior would override the vaccine hesitancy most common among self-identified Republicans. It has been well established that seniors (people age 65 and over) are more likely to become seriously ill and die from Covid-19 (the disease caused by SARS-CoV-2). Additionally, most seniors have seen first-hand how vaccines have eradicated diseases like polio, measles, mumps, and rubella in their lifetimes.

# Background

Polls have shown that vaccine hesitancy is far more prevalent among self-identified Republicans than other party affiliations/identities. For example, according to KFF COVID-19 Vaccine Monitor, as of October 2021, 31% of Republicans said they would “definitely not” get the vaccine compared to 18% of independents, and 2% of Democrats (Does The Public Want To Get a COVID-19 Vaccine? When?, 2021). The same website reports that people 65 and over are the least vaccine resistant age group with just 10% saying they would “definitely not” get the vaccine compared to 13% for 18 – 29, 21% for 30 - 49, and 18% for 50 – 64-year-olds.

Exit polls from the 2020 presidential election show that approximately 52% of voters age 65 and over voted for Trump (The New York Times, 2021; Igielnik, Keeter, & Hartig, 2021). This indicates that the senior population skews slightly republican. 61.7% of the total number of adults in the data set voted in the 2020 presidential election. Additionally, final vote tallies show that Biden got apx 52% of the total voters to 48% for Trump.

# Initial Hypothesis

My hypothesis is that seniors will have far greater vaccination rates than younger adults. That is, I expect the difference in vaccination rates to be much greater than poll numbers might indicate.

Given their lived experiences with previous vaccine successes and the amount of press regarding the heightened risks they face with Covid-19, it is not surprising that polls show seniors to be the least vaccine hesitant age group. I expect their lack of hesitancy to largely transcend political affiliation or ideology resulting in significantly higher rates of actual vaccinations than younger adults. Additionally, I suspect the risk faced by seniors might bring about slightly higher vaccine rates in younger people who interact with them – such as their adult children.

Further, I expect that significant numbers of younger healthier adults may not be prioritizing getting their vaccine because they assume their risk to be very low. While they may not fall in the “definitely not” group, they may not have actually ‘gotten around’ to getting their vaccinates.

Finally, I expect actual vaccination rates among Democratic voters to be significantly higher than Republican voters. Again, this is based on polls showing Republicans to have much higher rates for vaccine hesitancy.

# Data

Data on vaccinations, population demographics, and votes in the 2020 presidential election were gathered at the county level[[1]](#footnote-1) in order to look for correlations between voting and vaccination and/or percentage of seniors and vaccination.

Data on population demographics was collected from the US Census Bureau website (United States Census Bureau, 2021) in November 2021. The data was compiled by the Census Bureau from 2020 census data and included total population, adult (age 18 and over) population, and senior (age 65 and over) population by county (see Figure 1).

Voting data for the 2020 presidential election was taken from data originally published on Townhall.com but scraped and compiled by tonmcg and published on Github (tonmcg, 2021). This data includes the raw votes for the democratic and republican candidates along with the total votes cast on the presidential ballot. This data is provided at the county level (see Figure 2).

Vaccination data was collected from the Centers for Disease Control and Prevention website (Centers for Disease Control and Prevention, 2021) and has data through November 7, 2021. This data is compiled by the CDC from data submitted by vaccination providers. The data is provided at the county level (see Figure 3).

# Excluded Data

All data for the state of Alaska were excluded from the data set due to the unavailability of voting, demographic, and vaccination data at the same level (county verses precinct etc.).

There were nine counties with no reported adult vaccinations as of November 7, 2021. This was likely a failure to report data rather than an actual zero vaccination rate. Thus, these 9 counties were excluded from the data set. These counties are: Alpine, Inyo, Mariposa, Modoc, Mono, and Plumas Counties California; Kalawao and Kauai Counties Hawaii; and Loving County Texas.

Six counties reported more vaccinated adults that they had adult residents. Some of these might be explained by vaccinations being recorded in the county administered verses the recipient’s county of residence. However, one county (Santa Cruz County Arizona) reported over a 448% vaccination rate for the adults in the county. All six counties were removed from the data set. In addition to the aforementioned Santa Cruz, these were Chattahoochee County Georgia, McKinley County New Mexico, Martin County North Carolina, Presidio County Texas, and Portsmouth City Virgina[[2]](#footnote-2).

When looking at the scatter plot of percent of Biden voters verses vaccination rate, there appears to be some outliers that ‘defy logic’. This could be a data entry error of some other sort of reporting error. While there are other counties that also look a little ‘suspicious’, I am eliminating the six ‘worst offenders’, counties that Biden won that reported less than 10% vaccination rates. These were Hawaii, Honolulu, and Maui Counties in Hawaii and Barnstable, Dukes, and Nantucket Counties in Massachusetts.

# Final Data Set

The final data set contained a total of 3090 counties, parishes, independent cities, and the District of Columbia (see Figure 4).

# Variables

Data for the percent of adult population over the age of 65 was calculated from the Census Bureau data by dividing the raw number of people age 65 and over by the raw number of people age 18 and over. Figure 1 shows a portion of the R dataframe containing the imported data and Figure 5 shows the histogram and statistical summary of the data. The data appears to be normally distributed about a median of 25.5% of the adult population.

Adult vaccination percentage was calculated by dividing the raw number of vaccinated adults from the CDC data by the raw number of people age 18 and over from the Census data. Figure 3 shows a portion of the Excel file containing the downloaded data and Figure 7 shows the histogram and statistical summary of the data. The data appears to be skewed slightly left about a median of 54.1% of the adult population. There were a number of counties that reported suspiciously low vaccination rates. While I believe some of these must be incorrect, I could not justify removing them from the data set.

The percentage of votes cast for Joe Biden for President in the 2020 election was calculated by dividing the raw number of votes for Joe Biden by the raw number of total votes cast in the election. Figure 2 shows a portion of the Excel file containing the downloaded data and Figure 6 shows the histogram and statistical summary of the data. The data appears to be skewed right about a median of 29.9% of the votes cast. This is due to the large number of small, primarily rural counties in which Biden did not do well and the small number of large counties in which Biden did well.

# Methods

R Studio was used to process, clean, model, and visualize the data used in this study.

Demographic data was imported into R directly from the US Census website. Voting data and vaccine data were downloaded to my computer as Excel files. The readxl package in R was used to import data into R from the downloaded files

The tidyverse, ggplot, and dplyr packages in R were used to clean the data set and build histograms, scatterplots, linear models and confidence intervals from the data collected.

The percent of votes for Biden were compared to the adult vaccination percent (see Figure 11). Resulting in the following linear model:

Adult Vaccination Percent = 38.5 + (.47)(Percentage of Votes for Biden)

However, this model had an R2 value of .2826. This means that only 28.26% of the variation in adult vaccine percentage is explained by variation in percent of votes for Biden.

The percent of senior adults was compared to the adult vaccination percent (see Figure11). Resulting in the following linear model:

Adult Vaccination Percent = 60.5 + (-.25)(Percent of Seniors)

However, this model had an R2 value of 0.0096. This means that only 0.96% of the variation in adult vaccine percentage is explained by variation in percent of senior adults.

The percent of senior adults and percent of votes for Biden and the interaction between these factors was compared to the adult vaccination percent (see Figure 13). Resulting in the following linear model:

Adult Vaccination Percent = 39.83 + (-.09)(Percent of Seniors) + (.29)(Percentage of Votes for Biden) + (.008)(Percent of Seniors)(Percentage of Votes for Biden)

However, this model had an R2 and adjusted R2 value of 0.29. This means that only 29% of the variation in adult vaccine percentage is explained by variation in percent of votes for Biden, percent of senior adults, and the interaction between these factors.

# Expectations Based on Polls

Regarding voting data and vaccination data, based on poll data from the KFF, 2% of Democrats and 31% of Republicans said they would “definitely not” get the vaccine. (Does The Public Want To Get a COVID-19 Vaccine? When?, 2021). For the moment, let’s assume the voting population is representative of the overall adult population and we will ignore independents in order to understand the general magnitude. Thus, we would expect a .29% increase in the number of vaccinated adults for every 1% increase in votes for Biden (calculated as (.01 \* .98) – (.01\*.69)). That is to say, ignoring other factors, we expect a linear regression model to yield a result on the order of :

Adult Vaccination Percent = α + (.29)(Percentage of Votes for Biden)

Additionally, if we assumed that the voting population is representative of the overall adult population, based on the 52% of votes being cast for Biden we would expect an overall vaccination rate on the magnitude of 84.1% (calculated as (.98 \* .52) + (.69 \* .48)).

If age is not the primary deciding factor in whether or not someone has actually gotten the vaccine, that is if party affiliation were the only deciding factor in getting or not getting the vaccine, I would expect the overall vaccination rate among seniors to be approximately 60.8%. This is based on polls showing that 52% of seniors voted for Trump and that self-identified Republicans have 31% saying they definitely would not get the vaccine and 48% of seniors voted for Biden and only 2% of Democrats said they would definitely not get the vaccine (calculated as (.52 \* .69) + (.48 \* .98) = .608). Meaning, if age doesn’t play a role, then seniors would have significantly lower vaccination rates than the expected 84.1% of the total adult population (calculated as (.98 \* .52) + (.69 \* .48)).

However, CDC data shows that as of November 7, 2021, only 65.6% of all adults are vaccinated. Clearly, other factors are influencing actual vaccination rates.

Regarding senior data and vaccination data, based on poll data from the KFF, 10% of seniors said they would “definitely not” get the vaccine. (Does The Public Want To Get a COVID-19 Vaccine? When?, 2021). And, according to Census data, seniors make up 21.7% of the adult population. If we assume for the moment that vaccination rates among seniors is independent of vaccination rates of other adults (in order to understand the general magnitude) A 1% increase in percent of seniors in the adult population would be expected to yield a .9% increase in the number of vaccinated adults for every 1% increase in votes for Biden (calculated as .01\* .90). That is to say we expect a linear regression model, ignoring other factors to yield a result on the order of :

Adult Vaccination Percent = α + (.09)(Percentage of Seniors)

If being a senior completely outweighs party affiliation, I would expect the vaccination rate amount seniors to be approximately 90% compared to 85.4% of all adults (calculated as (.217 \* .90) + (.783 \* .48 \* .69) + (.783 \* .52 \* .98) = .854). This is again based on the KFF data on vaccine hesitancy, final vote counts, and Census data showing that seniors make up approximately 21.7% of the total adult population.

We should expect some interaction between percent of seniors and percent of votes for Biden. Additionally, there will surely be other factors at work that we are not considering here. The above are simply to give a rough idea of the range of expected outcomes.

By looking at vaccine data as of November 7th 2021, we have allowed plenty of time for the vast majority of adults to get vaccinated if they choose to. Additionally, starting in September, President Biden announced a series of vaccine mandates that may have increased vaccine participation among the mildly hesitant. Thus, we would have expected actual vaccine rates to nearly mirror the KFF poll data.

# Model Summary: Explaining Vaccinations with Percent of Seniors or Votes for Biden

The two single factor models constructed (see Figure 11 & Figure 12) did result in significant coefficients. Meaning that each model shows a linear relationship between the factor and vaccination rates. When considered individually, the 95% confidence intervals for the coefficients, the F stat and F critical values for these two separate models were as follows:

pct\_voted\_Biden C.I.(0.443, 0.495), F stat = 1216, F critical < 2.2e-16

pct\_adult\_seniors (-0.342, -0.161), F stat = 29.8, F critical < 5.171e-08

While each model does show a significant linear relationship, the R2 values of each model are very low (.28 & .10 respectively) and thus we know these models only explain a small portion of the variation in vaccination rates.

The first of these models shows, with 95% confidence, that every increase of 1% of Biden voters in the voting population correlates to an increase in percent of vaccination rates in an adult population of between .443 and .495. This is a larger effect than poll data on vaccine hesitancy would lead us to believe. However, it is important to keep in mind that, as previously stated, voting behavior does not fully explain the variation in vaccine rates.

It is to be expected that the explanatory value of percent of seniors along would be low since seniors only make up 21.7% of the adult population in our data set. The surprising result here is that the coefficient for percent of seniors is negative. This result shows, with 95% confidence, that every increase of 1% of seniors in the adult population correlates to a decrease in percent of vaccination rates in an adult population of between .161 and .342.

# Model Summary: Explaining Vaccinations with Percent of Seniors and Votes for Biden

Given that each of the two factors considered has a significant linear relationship to the outcome, it seems appropriate to combine them into one model. The model using both percent of seniors and percent of votes for Biden and the interaction of these (see Figure 13) does the best job of explaining the variation in vaccine rates. However, even that model had an R2 and adjusted R2 value of 0.29. This means that only 29% of the variation in adult vaccine percentage is explained by variation in percent of votes for Biden, percent of senior adults, and the interaction between these factors in a linear model.

That linear model is:

Adult Vaccination Percent = 39.83 + (-.09)(Percent of Seniors) + (.29)(Percentage of Votes for Biden) + (.008)(Percent of Seniors)(Percentage of Votes for Biden)

## Other Model Data

F stat = 421.5, F critical < 2.2E-16; F stat >> F critical: thus the model is useful.

Collinearity between Percentage of Votes for Biden and Percent of Seniors = -.313; Collinearity exists but not at a level that is problematic for the model’s validity.

## Hypothesis Test

Ho: βseniors = βvotes = βseniors\*votes = 0; All model terms (percentage of seniors, percentage of votes for Biden in an adult population, and the interaction between these) are unimportant for explaining adult vaccination rates in a linear model.

Ha: At least one βi ≠ 0, At least one model terms (percentage of seniors, percentage of votes for Biden in an adult population, or the interaction between these) are useful for explaining adult vaccination rates in a linear model.

Because α (.05) far exceeds the observed significance level (p <2e-16), the data provided strong evidence that at least one of the model coefficients is non-zero. Therefore, we reject the null hypothesis. The overall model appears to be statistically useful for explaining vaccination rates.

# Conclusion

The linear model explaining vaccinations with percent of seniors and votes for Biden, and the interaction of those factors, is the best of the three models presented (based on R2 and adjusted R2). However, the variation in the factors only explains 29% of the variation in vaccination rates in a linear fashion.

It is possible that a different model might reveal a stronger relationship between these factors and vaccination rates that is significant but non-linear. This study does not explore other such models. This study also doesn’t attempt to uncover other factors that could be added to the linear model to explain more of the variation in vaccination rates.

The most surprising result is that the correlation between percentage of seniors and vaccination rates was negative whether considered as the only factor or in the multi-factor model. This was not expected based on the polling data or my own intuition. Then again, maybe this shouldn’t be surprising as there is a large body of evidence pointing to a lack of candor in polls on controversial subjects.

# References

*Centers for Disease Control and Prevention*. (2021, November 7). Retrieved from COVID-19 Vaccinations in the United States, County: https://data.cdc.gov/Vaccinations/COVID-19-Vaccinations-in-the-United-States-County/8xkx-amqh

*Does The Public Want To Get a COVID-19 Vaccine? When?* (2021). Retrieved from Kaiser Family Foundation: https://www.kff.org/coronavirus-covid-19/dashboard/kff-covid-19-vaccine-monitor-dashboard/

Igielnik, R., Keeter, S., & Hartig, H. (2021, June 30). *Behind Biden's 2020 Victory*. Retrieved from Pew Research Center: https://www.pewresearch.org/politics/2021/06/30/behind-bidens-2020-victory/

The New York Times. (2021). *National Exit Polls: How Different Groups Voted*. Retrieved from The New York Times: https://www.nytimes.com/interactive/2020/11/03/us/elections/exit-polls-president.html

tonmcg. (2021). *Github*. Retrieved from US\_County\_Level\_Election\_Results\_08\_20: https://github.com/tonmcg/US\_County\_Level\_Election\_Results\_08-20/blob/master/2020\_US\_County\_Level\_Presidential\_Results.csv

*United States Census Bureau*. (2021). Retrieved from https://www2.census.gov/programs-surveys/popest/datasets/2010-2020/counties/asrh

Figures

Table

Description automatically generated



Figure 1. Partial view of the R dataframe containing the population demographics data as imported from the US Census website (not all columns were imported and not all rows are shown).

Table

Description automatically generated

Table

Description automatically generated

Figure 2. Partial view of the Excel file containing the voting data downloaded from Github (not all rows are shown).

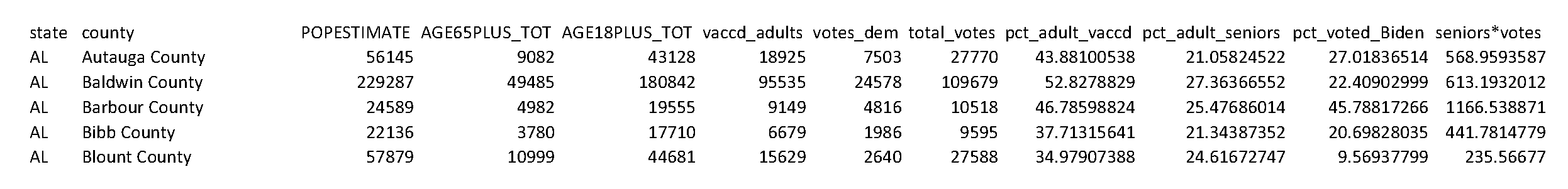
Table

Description automatically generated with low confidence

Table

Description automatically generated

Figure 3. Partial view of the Excel file containing vaccination data downloaded from the CDC website (not all columns or rows are shown).



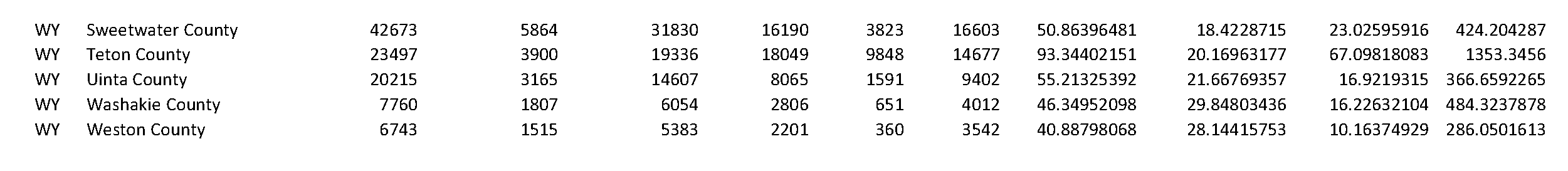


Figure 4. Partial view of the R dataframe containing the final ‘cleaned’ data used for analysis (not all rows are shown).

Chart, histogram

Description automatically generated

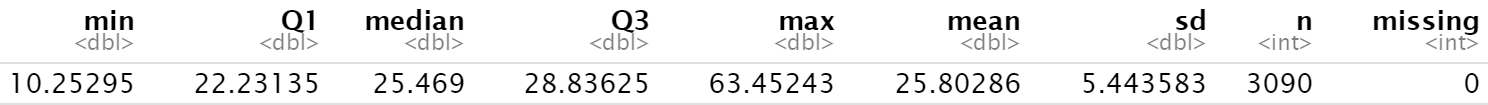


Figure 5. Histogram and statistical summary of counties based on the percentage of the adult population age 65 or older.

Chart, histogram

Description automatically generated

**

Figure 6. Histogram and statistical summary of counties based on the percentage of votes cast for Biden in the 2020 presidential election.

Chart, histogram

Description automatically generated



Figure 7. Histogram of counties based on the percentage of the adult population who have received the Covid-19 vaccine.

Chart, scatter chart

Description automatically generated

Figure 8. Scatterplot (created using ggplot) of counties’ percentage of the adult population age 65 or older vs. the percentage of the adult population who have received the Covid-19 vaccine.

Chart, scatter chart

Description automatically generated

Figure 9. Scatterplot (created using ggplot) of counties’ percentage of votes cast for Biden in the 2020 presidential election vs. the percentage of the adult population who have received the Covid-19 vaccine.

Chart, scatter chart

Description automatically generated

Figure 10. Scatterplot (created using ggplot) of counties percentage of the adult population age 65 or older vs. percentage of votes cast for Biden in the 2020 presidential election.

A screenshot of a computer

Description automatically generated with medium confidence

Text

Description automatically generated

Figure 11. Data for a linear regression model created in R. This model uses the percentage of the adult population age 65 or older as the independent variable and the percentage of the adult population who have received the Covid-19 vaccine as the dependent variable.

A screenshot of a computer

Description automatically generated with medium confidence

Text

Description automatically generated with low confidence

Figure 12. Data a linear regression model created in R. This model uses percentage of votes cast for Biden in the 2020 presidential election as the independent variable and the percentage of the adult population who have received the Covid-19 vaccine as the dependent variable.

A screenshot of a computer

Description automatically generated with medium confidence

Text

Description automatically generated with medium confidence

Text

Description automatically generated

Figure 12. Data a linear regression model created in R. This model uses percentage of votes cast for Biden in the 2020 presidential election, the percentage of the adult population age 65 or older, and the interaction of those factors as the independent variables and the percentage of the adult population who have received the Covid-19 vaccine as the dependent variable.

# The Complete R Notebook

# Vaccine Rates

Kale Perry

library(readxl)  
library(tidyverse)  
library(mosaic)

# create a tibble containing the voting data  
# I wasn't sure the website I pulled this from  
# would remain stable/unchanged  
# so I downloaded the full excel file   
# and will pull from the that file on my drive   
vote1 <- read\_xlsx("votes.xlsx")  
head(vote1)

## # A tibble: 6 x 10  
## state\_name county\_fips county\_name votes\_gop votes\_dem total\_votes diff  
## <chr> <dbl> <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 Alabama 1001 Autauga County 19838 7503 27770 12335  
## 2 Alabama 1003 Baldwin County 83544 24578 109679 58966  
## 3 Alabama 1005 Barbour County 5622 4816 10518 806  
## 4 Alabama 1007 Bibb County 7525 1986 9595 5539  
## 5 Alabama 1009 Blount County 24711 2640 27588 22071  
## 6 Alabama 1011 Bullock County 1146 3446 4613 -2300  
## # ... with 3 more variables: per\_gop <dbl>, per\_dem <dbl>, per\_point\_diff <dbl>

# Create a tibble containing state names and abbreviations  
states <- tibble(abb = state.abb,  
 state\_name = state.name)  
head(states)

## # A tibble: 6 x 2  
## abb state\_name  
## <chr> <chr>   
## 1 AL Alabama   
## 2 AK Alaska   
## 3 AZ Arizona   
## 4 AR Arkansas   
## 5 CA California  
## 6 CO Colorado

# join states and voting data   
# purpose is to get the state abbreviations into the voting tibble  
vote2 <- vote1 %>%  
 left\_join(states, by = "state\_name")  
head(vote2)

## # A tibble: 6 x 11  
## state\_name county\_fips county\_name votes\_gop votes\_dem total\_votes diff  
## <chr> <dbl> <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 Alabama 1001 Autauga County 19838 7503 27770 12335  
## 2 Alabama 1003 Baldwin County 83544 24578 109679 58966  
## 3 Alabama 1005 Barbour County 5622 4816 10518 806  
## 4 Alabama 1007 Bibb County 7525 1986 9595 5539  
## 5 Alabama 1009 Blount County 24711 2640 27588 22071  
## 6 Alabama 1011 Bullock County 1146 3446 4613 -2300  
## # ... with 4 more variables: per\_gop <dbl>, per\_dem <dbl>,  
## # per\_point\_diff <dbl>, abb <chr>

# Clean up some column names and   
# get rid of columns that are not needed  
  
vote <- vote2 %>%  
 mutate(state = abb,  
 county = county\_name)%>%  
 select(state, county, votes\_dem,per\_dem, total\_votes )  
head(vote)

## # A tibble: 6 x 5  
## state county votes\_dem per\_dem total\_votes  
## <chr> <chr> <dbl> <dbl> <dbl>  
## 1 AL Autauga County 7503 0.270 27770  
## 2 AL Baldwin County 24578 0.224 109679  
## 3 AL Barbour County 4816 0.458 10518  
## 4 AL Bibb County 1986 0.207 9595  
## 5 AL Blount County 2640 0.0957 27588  
## 6 AL Bullock County 3446 0.747 4613

# I discovered a difference in how DC was entered in the two DFs  
# In one the state is NA  
vote <- vote %>%  
 mutate(state = ifelse(county == "District of Columbia", "DC", state ))

# I discovered a difference in how Dona Ana County was entered  
# in the 2 dfs  
vote <- vote %>%  
 mutate(county = ifelse(str\_detect(county,"Ana County"), "Dona Ana County", county))

# Create a tibble containing the vaccine data  
vaccine1 <- read\_csv("vaccines.csv")

## Rows: 223193 Columns: 28

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (5): Date, Recip\_County, Recip\_State, SVI\_CTGY, Metro\_status  
## dbl (20): Series\_Complete\_Pop\_Pct, Series\_Complete\_18PlusPop\_Pct, Series\_Com...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

tail(vaccine1)

## # A tibble: 6 x 28  
## Date Recip\_County Recip\_State Series\_Complete\_Pop\_Pct Series\_Complete\_Y~  
## <chr> <chr> <chr> <dbl> <dbl>  
## 1 11/2/2021 Weston County WY 32.2 2233  
## 2 11/3/2021 Weston County WY 32.3 2235  
## 3 11/4/2021 Weston County WY 32.3 2236  
## 4 11/5/2021 Weston County WY 32.4 2242  
## 5 11/6/2021 Weston County WY 32.5 2249  
## 6 11/7/2021 Weston County WY 32.5 2249  
## # ... with 23 more variables: Series\_Complete\_18Plus <dbl>,  
## # Series\_Complete\_18PlusPop\_Pct <dbl>, Series\_Complete\_65Plus <dbl>,  
## # Series\_Complete\_65PlusPop\_Pct <dbl>, Completeness\_pct <dbl>,  
## # Administered\_Dose1\_Recip <dbl>, Administered\_Dose1\_Pop\_Pct <dbl>,  
## # Administered\_Dose1\_Recip\_12Plus <dbl>,  
## # Administered\_Dose1\_Recip\_12PlusPop\_Pct <dbl>,  
## # Administered\_Dose1\_Recip\_18Plus <dbl>, ...

# Clean up some column names and   
# get rid of columns that are not needed  
  
vaccine2 <- vaccine1%>%  
 mutate(county = Recip\_County,  
 state = Recip\_State,  
 vaccd\_adults = Series\_Complete\_18Plus)%>%  
 select(state, county,vaccd\_adults,Series\_Complete\_18PlusPop\_Pct )  
head(vaccine2)

## # A tibble: 6 x 4  
## state county vaccd\_adults Series\_Complete\_18PlusPop\_Pct  
## <chr> <chr> <dbl> <dbl>  
## 1 AK Aleutians East Borough 2143 69.5  
## 2 AK Aleutians East Borough 2143 69.5  
## 3 AK Aleutians East Borough 2159 70   
## 4 AK Aleutians East Borough 2159 70   
## 5 AK Aleutians East Borough 2161 70.1  
## 6 AK Aleutians East Borough 2161 70.1

# Group by each county  
# take the max reported vaccine percentage for each county  
vaccine3 <- vaccine2 %>%  
 group\_by(state,county)%>%  
 summarise(vaccd\_adults = max(vaccd\_adults))

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

head(vaccine3)

## # A tibble: 6 x 3  
## # Groups: state [1]  
## state county vaccd\_adults  
## <chr> <chr> <dbl>  
## 1 AK Aleutians East Borough 2217  
## 2 AK Aleutians West Census Area 3039  
## 3 AK Anchorage Municipality 151357  
## 4 AK Bethel Census Area 9338  
## 5 AK Bristol Bay Borough 739  
## 6 AK Denali Borough 922

# I discovered a difference in how LaSalle Parish   
# was entered in the two DFs  
# In one there was a space, in the other there was not  
vaccine3 <- vaccine3 %>%  
 mutate(county = ifelse(county == "La Salle Parish", "LaSalle Parish", county ))

# Need to make sure the treatment for DC  
# matched what I did in the vote df  
vaccine3 <- vaccine3 %>%  
 mutate(state = ifelse(county == "District of Columbia", "DC", state ))

# Join the voting data and vaccine data into one tibble  
# Filter out rows that don't have any data for vaccines and voting  
dem\_v\_vac <- vaccine3%>%  
 full\_join(vote,by=c("state","county"))%>%  
 filter(vaccd\_adults !=0 & votes\_dem != "NA")  
  
head(dem\_v\_vac)

## # A tibble: 6 x 6  
## # Groups: state [1]  
## state county vaccd\_adults votes\_dem per\_dem total\_votes  
## <chr> <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 AL Autauga County 18925 7503 0.270 27770  
## 2 AL Baldwin County 95535 24578 0.224 109679  
## 3 AL Barbour County 9149 4816 0.458 10518  
## 4 AL Bibb County 6679 1986 0.207 9595  
## 5 AL Blount County 15629 2640 0.0957 27588  
## 6 AL Bullock County 4302 3446 0.747 4613

# The population demographics data  
# is in a separate table for each state  
# All states need to be combined into one tibble  
  
# Importing Alabama data into a tibble  
# as a base tibble for all the different state csv files   
pop\_data<- read\_csv("https://www2.census.gov/programs-surveys/popest/datasets/2010-2020/counties/asrh/CC-EST2020-AGESEX-01.csv")%>%  
 filter(YEAR == "13")%>%  
 select("STNAME", "CTYNAME","POPESTIMATE", "AGE65PLUS\_TOT", "AGE18PLUS\_TOT" )

## Rows: 871 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (5): SUMLEV, STATE, COUNTY, STNAME, CTYNAME  
## dbl (91): YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UNDER5\_MAL...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

head(pop\_data)

## # A tibble: 6 x 5  
## STNAME CTYNAME POPESTIMATE AGE65PLUS\_TOT AGE18PLUS\_TOT  
## <chr> <chr> <dbl> <dbl> <dbl>  
## 1 Alabama Autauga County 56145 9082 43128  
## 2 Alabama Baldwin County 229287 49485 180842  
## 3 Alabama Barbour County 24589 4982 19555  
## 4 Alabama Bibb County 22136 3780 17710  
## 5 Alabama Blount County 57879 10999 44681  
## 6 Alabama Bullock County 9976 1763 7899

# add other states into the pop\_data tibble  
for (i in c("04","05","06","08","09",10:13,15:42,44:51,53:56)){  
 url <- paste("https://www2.census.gov/programs-surveys/popest/datasets/2010-2020/counties/asrh/CC-EST2020-AGESEX-" , i,".csv", sep="")  
 new\_st <- read\_csv(url)%>%  
 filter(YEAR == "13")%>%  
 select("STNAME", "CTYNAME","POPESTIMATE","AGE65PLUS\_TOT", "AGE18PLUS\_TOT" )%>%  
 mutate (POPESTIMATE = as.double(POPESTIMATE),AGE65PLUS\_TOT = as.double(AGE65PLUS\_TOT), AGE18PLUS\_TOT=as.double(AGE18PLUS\_TOT))   
  
 pop\_data <- pop\_data%>%  
 add\_row(new\_st)  
}

## Rows: 195 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (5): SUMLEV, STATE, COUNTY, STNAME, CTYNAME  
## dbl (91): YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UNDER5\_MAL...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 975 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (5): SUMLEV, STATE, COUNTY, STNAME, CTYNAME  
## dbl (91): YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UNDER5\_MAL...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 754 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (5): SUMLEV, STATE, COUNTY, STNAME, CTYNAME  
## dbl (91): YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UNDER5\_MAL...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 832 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (5): SUMLEV, STATE, COUNTY, STNAME, CTYNAME  
## dbl (91): YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UNDER5\_MAL...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 104 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (5): SUMLEV, STATE, COUNTY, STNAME, CTYNAME  
## dbl (91): YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UNDER5\_MAL...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 39 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 13 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 871 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 2067 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 65 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 572 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 1326 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 1196 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 1287 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 1365 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 1560 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 832 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 208 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 312 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 182 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 1079 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 1131 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 1066 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 1495 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 728 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 1209 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 221 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 130 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 273 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 429 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 806 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 1300 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 689 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 1144 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 1001 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 468 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 871 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 65 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 598 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 858 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 1235 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 3302 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 377 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 182 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 1729 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 507 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 715 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 936 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 299 Columns: 96

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (4): SUMLEV, COUNTY, STNAME, CTYNAME  
## dbl (92): STATE, YEAR, POPESTIMATE, POPEST\_MALE, POPEST\_FEM, UNDER5\_TOT, UND...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

tail(pop\_data)

## # A tibble: 6 x 5  
## STNAME CTYNAME POPESTIMATE AGE65PLUS\_TOT AGE18PLUS\_TOT  
## <chr> <chr> <dbl> <dbl> <dbl>  
## 1 Wyoming Sublette County 9856 2146 7752  
## 2 Wyoming Sweetwater County 42673 5864 31830  
## 3 Wyoming Teton County 23497 3900 19336  
## 4 Wyoming Uinta County 20215 3165 14607  
## 5 Wyoming Washakie County 7760 1807 6054  
## 6 Wyoming Weston County 6743 1515 5383

#verify length of pop\_data  
nrow(pop\_data)

## [1] 3113

#compare to length of the previous tibble dem\_v\_vac  
nrow(dem\_v\_vac)

## [1] 3102

# figure out whats missing in dem\_v\_vac   
# \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*  
# setdiff is a wonderful helpful function here  
missing <-   
 setdiff( pop\_data$CTYNAME, dem\_v\_vac$county)  
missing

## [1] "Alpine County" "Inyo County" "Mariposa County"   
## [4] "Modoc County" "Mono County" "Plumas County"   
## [7] "Kalawao County" "Kauai County" "Do<f1>a Ana County"  
## [10] "Loving County"

# NOTE: the first time these checks were performed I discovered a couple of   
# differences between the voting data and vaccine data  
# I choose to go back up and fix those issues earlier in the process  
# see line 46-58 and 85-91 above as examples.  
  
# verify there isn't anything in dem\_v\_vac   
#that is missing in pop\_data  
missing2 <-  
 setdiff(dem\_v\_vac$county, pop\_data$CTYNAME )  
missing2

## [1] "Dona Ana County"

# Most of the rows that appear in our pop\_data   
# but are missing from the dem\_v\_vac data  
# are due to no reported vaccine data for that county  
# I will remove these counties after combining all the data  
vaccine3 %>%  
 filter(county %in% missing)

## # A tibble: 9 x 3  
## # Groups: state [3]  
## state county vaccd\_adults  
## <chr> <chr> <dbl>  
## 1 CA Alpine County 0  
## 2 CA Inyo County 0  
## 3 CA Mariposa County 0  
## 4 CA Modoc County 0  
## 5 CA Mono County 0  
## 6 CA Plumas County 0  
## 7 HI Kalawao County 0  
## 8 HI Kauai County 0  
## 9 TX Loving County 0

# How is Dona Ana listed in the pop\_data?  
# need to figure out what is going on with  
# "Do\xf1a Ana County"  
pop\_data %>%  
 filter(str\_detect(CTYNAME, "Ana"))

## # A tibble: 1 x 5  
## STNAME CTYNAME POPESTIMATE AGE65PLUS\_TOT AGE18PLUS\_TOT  
## <chr> <chr> <dbl> <dbl> <dbl>  
## 1 New Mexico "Do\xf1a Ana County" 221262 36983 168308

# Fix Dona Ana to match porevious  
pop\_data <- pop\_data %>%  
 mutate(CTYNAME = ifelse(str\_detect(CTYNAME,"Ana County"), "Dona Ana County", CTYNAME))

# rerunning missing - Dona Ana County should be fixed  
missing <-   
 setdiff( pop\_data$CTYNAME, dem\_v\_vac$county)  
missing

## [1] "Alpine County" "Inyo County" "Mariposa County" "Modoc County"   
## [5] "Mono County" "Plumas County" "Kalawao County" "Kauai County"   
## [9] "Loving County"

# this leaves only the counties with no vaccine data   
# as the differences - which is good  
  
# make state and county column headers match the voting/vac df  
pop\_data <- pop\_data %>%  
 rename(state\_name = STNAME, county = CTYNAME)  
head(pop\_data)

## # A tibble: 6 x 5  
## state\_name county POPESTIMATE AGE65PLUS\_TOT AGE18PLUS\_TOT  
## <chr> <chr> <dbl> <dbl> <dbl>  
## 1 Alabama Autauga County 56145 9082 43128  
## 2 Alabama Baldwin County 229287 49485 180842  
## 3 Alabama Barbour County 24589 4982 19555  
## 4 Alabama Bibb County 22136 3780 17710  
## 5 Alabama Blount County 57879 10999 44681  
## 6 Alabama Bullock County 9976 1763 7899

# join states and pop\_data   
# purpose is to get the state abbreviations into the voting tibble  
pop2 <- pop\_data %>%  
 left\_join(states, by = "state\_name")  
head(pop2)

## # A tibble: 6 x 6  
## state\_name county POPESTIMATE AGE65PLUS\_TOT AGE18PLUS\_TOT abb   
## <chr> <chr> <dbl> <dbl> <dbl> <chr>  
## 1 Alabama Autauga County 56145 9082 43128 AL   
## 2 Alabama Baldwin County 229287 49485 180842 AL   
## 3 Alabama Barbour County 24589 4982 19555 AL   
## 4 Alabama Bibb County 22136 3780 17710 AL   
## 5 Alabama Blount County 57879 10999 44681 AL   
## 6 Alabama Bullock County 9976 1763 7899 AL

# Clean up and get rid of unwanted columns  
pop <- pop2 %>%  
 mutate(state = abb,pct\_adult\_seniors = 100\*AGE65PLUS\_TOT/AGE18PLUS\_TOT)%>%  
 select(state,pct\_adult\_seniors, county,POPESTIMATE, AGE65PLUS\_TOT,AGE18PLUS\_TOT)  
head(pop)

## # A tibble: 6 x 6  
## state pct\_adult\_seniors county POPESTIMATE AGE65PLUS\_TOT AGE18PLUS\_TOT  
## <chr> <dbl> <chr> <dbl> <dbl> <dbl>  
## 1 AL 21.1 Autauga County 56145 9082 43128  
## 2 AL 27.4 Baldwin County 229287 49485 180842  
## 3 AL 25.5 Barbour County 24589 4982 19555  
## 4 AL 21.3 Bibb County 22136 3780 17710  
## 5 AL 24.6 Blount County 57879 10999 44681  
## 6 AL 22.3 Bullock County 9976 1763 7899

# Need to verify the treatment of DC matches  
pop <- pop %>%  
 mutate(state = ifelse(county == "District of Columbia", "DC", state ))

# Join pop\_data with dem\_v\_vac   
# this puts voting data, demo data, and   
# vaccine data all in one tibble/df  
  
all\_data <- pop %>%  
 full\_join(dem\_v\_vac,by=c("state","county"))  
head(all\_data)

## # A tibble: 6 x 10  
## state pct\_adult\_seniors county POPESTIMATE AGE65PLUS\_TOT AGE18PLUS\_TOT  
## <chr> <dbl> <chr> <dbl> <dbl> <dbl>  
## 1 AL 21.1 Autauga County 56145 9082 43128  
## 2 AL 27.4 Baldwin County 229287 49485 180842  
## 3 AL 25.5 Barbour County 24589 4982 19555  
## 4 AL 21.3 Bibb County 22136 3780 17710  
## 5 AL 24.6 Blount County 57879 10999 44681  
## 6 AL 22.3 Bullock County 9976 1763 7899  
## # ... with 4 more variables: vaccd\_adults <dbl>, votes\_dem <dbl>,  
## # per\_dem <dbl>, total\_votes <dbl>

nrow(all\_data)

## [1] 3113

# need to remove the counties that   
# we previously discovered had  
# not reported vaccine data  
all\_data <- all\_data%>%  
 filter(vaccd\_adults!=0)  
nrow(all\_data)

## [1] 3102

# That looks good  
  
# add a column for the percent of adults  
# reported as vaccinated  
all\_data <- all\_data %>%  
 mutate(pct\_adult\_vaccd = 100\*(vaccd\_adults/AGE18PLUS\_TOT))  
head(all\_data)

## # A tibble: 6 x 11  
## state pct\_adult\_seniors county POPESTIMATE AGE65PLUS\_TOT AGE18PLUS\_TOT  
## <chr> <dbl> <chr> <dbl> <dbl> <dbl>  
## 1 AL 21.1 Autauga County 56145 9082 43128  
## 2 AL 27.4 Baldwin County 229287 49485 180842  
## 3 AL 25.5 Barbour County 24589 4982 19555  
## 4 AL 21.3 Bibb County 22136 3780 17710  
## 5 AL 24.6 Blount County 57879 10999 44681  
## 6 AL 22.3 Bullock County 9976 1763 7899  
## # ... with 5 more variables: vaccd\_adults <dbl>, votes\_dem <dbl>,  
## # per\_dem <dbl>, total\_votes <dbl>, pct\_adult\_vaccd <dbl>

# looking to verify that all data is at the county level  
# or other appropriate level  
# LA has Parish rather than County -   
# so Parish is appropriate for LA  
not\_counties <- all\_data %>%  
 filter( str\_detect(county,"County$", negate = TRUE))%>%  
 filter( str\_detect(county,"Parish$", negate = TRUE))  
nrow(not\_counties)

## [1] 42

# Need to figure out what's going on with  
# the 42 rows that are neither county or parish  
head(not\_counties)

## # A tibble: 6 x 11  
## state pct\_adult\_seniors county POPESTIMATE AGE65PLUS\_TOT AGE18PLUS\_TOT  
## <chr> <dbl> <chr> <dbl> <dbl> <dbl>  
## 1 DC 15.4 District of C~ 712816 89833 583228  
## 2 MD 18.7 Baltimore city 586131 87793 468378  
## 3 MO 18.1 St. Louis city 297645 43829 242343  
## 4 NV 26.6 Carson City 56034 11893 44646  
## 5 VA 15.2 Alexandria ci~ 158726 19760 129672  
## 6 VA 27.9 Bristol city 17329 3845 13776  
## # ... with 5 more variables: vaccd\_adults <dbl>, votes\_dem <dbl>,  
## # per\_dem <dbl>, total\_votes <dbl>, pct\_adult\_vaccd <dbl>

tail(not\_counties)

## # A tibble: 6 x 11  
## state pct\_adult\_seniors county POPESTIMATE AGE65PLUS\_TOT AGE18PLUS\_TOT  
## <chr> <dbl> <chr> <dbl> <dbl> <dbl>  
## 1 VA 26.4 Staunton city 25190 5382 20373  
## 2 VA 20.0 Suffolk city 93913 14322 71672  
## 3 VA 19.4 Virginia Beac~ 451231 68347 352283  
## 4 VA 23.6 Waynesboro ci~ 22741 4127 17506  
## 5 VA 20.2 Williamsburg ~ 15259 2732 13500  
## 6 VA 21.7 Winchester ci~ 27700 4683 21548  
## # ... with 5 more variables: vaccd\_adults <dbl>, votes\_dem <dbl>,  
## # per\_dem <dbl>, total\_votes <dbl>, pct\_adult\_vaccd <dbl>

# With a bit of investigation, I learned  
# that the cities listed above are all "independent"  
# meaning they are not part of any county  
# thus, they would be appropriate to keep in the data set  
  
# cleanup and formatting of the voting percent column  
all\_data <- all\_data %>%  
 rename(pct\_voted\_Biden = per\_dem)%>%  
 mutate(pct\_voted\_Biden = pct\_voted\_Biden\*100)  
head(all\_data)

## # A tibble: 6 x 11  
## state pct\_adult\_seniors county POPESTIMATE AGE65PLUS\_TOT AGE18PLUS\_TOT  
## <chr> <dbl> <chr> <dbl> <dbl> <dbl>  
## 1 AL 21.1 Autauga County 56145 9082 43128  
## 2 AL 27.4 Baldwin County 229287 49485 180842  
## 3 AL 25.5 Barbour County 24589 4982 19555  
## 4 AL 21.3 Bibb County 22136 3780 17710  
## 5 AL 24.6 Blount County 57879 10999 44681  
## 6 AL 22.3 Bullock County 9976 1763 7899  
## # ... with 5 more variables: vaccd\_adults <dbl>, votes\_dem <dbl>,  
## # pct\_voted\_Biden <dbl>, total\_votes <dbl>, pct\_adult\_vaccd <dbl>

# need to verify that there are not  
# any 'BAD' data in our data set  
# namely - check for percentages  
# NOT between zero & 100  
bad\_data <- all\_data%>%  
 filter(pct\_voted\_Biden < 0   
 | pct\_adult\_seniors < 0  
 | pct\_adult\_vaccd < 0  
 | pct\_voted\_Biden >100   
 | pct\_adult\_seniors >100  
 | pct\_adult\_vaccd >100)  
show(bad\_data)

## # A tibble: 6 x 11  
## state pct\_adult\_seniors county POPESTIMATE AGE65PLUS\_TOT AGE18PLUS\_TOT  
## <chr> <dbl> <chr> <dbl> <dbl> <dbl>  
## 1 AZ 26.0 Santa Cruz Co~ 46808 8955 34463  
## 2 GA 6.36 Chattahoochee~ 10551 520 8181  
## 3 NM 18.6 McKinley Coun~ 70824 9483 51068  
## 4 NC 31.2 Martin County 22178 5527 17716  
## 5 TX 33.1 Presidio Coun~ 6508 1565 4729  
## 6 VA 19.9 Portsmouth ci~ 95094 14521 73059  
## # ... with 5 more variables: vaccd\_adults <dbl>, votes\_dem <dbl>,  
## # pct\_voted\_Biden <dbl>, total\_votes <dbl>, pct\_adult\_vaccd <dbl>

# I am not sure how these counties  
# are reporting greater than 100% vaccination rates  
# the numbers slightly over 100% might be   
# explained by people crossing county lines to get their shot  
# the 448% must be a true error  
# regardless - I will delete these locations from the data set  
all\_data <- all\_data%>%  
 filter(pct\_voted\_Biden > 0   
 & pct\_adult\_seniors > 0  
 & pct\_adult\_vaccd > 0  
 & pct\_voted\_Biden <=100   
 & pct\_adult\_seniors <=100  
 & pct\_adult\_vaccd <=100)  
nrow(all\_data)

## [1] 3096

# get fav stats for vaccinations  
favstats(~pct\_adult\_vaccd, data = all\_data)

## min Q1 median Q3 max mean sd n missing  
## 0.03805697 45.86514 54.0994 62.88582 97.78822 53.88723 14.17892 3096 0

# I think my data set is all cleaned up  
# I want to start looking at some visualization  
  
# Create a scatter plot  
# voting vs vaccines  
plot\_vote\_v\_vac <- ggplot(data = all\_data,  
 aes(x = pct\_voted\_Biden, y = pct\_adult\_vaccd ))+  
 geom\_point()+  
 ggtitle("Votes for Biden vs. Vaccined Adults",  
 subtitle = "in Percent of Adults, by US county")+  
 geom\_smooth(method=lm)  
plot\_vote\_v\_vac

## `geom\_smooth()` using formula 'y ~ x'

Chart, scatter chart

Description automatically generated

# There seems to be some outliers  
# counties that Biden won by large margins  
# with extremely low vaccination rate  
# Let's take a look at some of those  
  
outliers <- all\_data%>%  
 filter(pct\_voted\_Biden > 50   
 & pct\_adult\_vaccd < pct\_voted\_Biden/4)  
outliers

## # A tibble: 7 x 11  
## state pct\_adult\_seniors county POPESTIMATE AGE65PLUS\_TOT AGE18PLUS\_TOT  
## <chr> <dbl> <chr> <dbl> <dbl> <dbl>  
## 1 GA 28.4 Hancock County 8494 2050 7223  
## 2 HI 28.6 Hawaii County 203340 45787 160286  
## 3 HI 23.8 Honolulu Coun~ 963826 181138 762530  
## 4 HI 24.9 Maui County 167902 32803 131908  
## 5 MA 37.5 Barnstable Co~ 213164 68379 182176  
## 6 MA 31.4 Dukes County 17461 4512 14366  
## 7 MA 20.0 Nantucket Cou~ 11376 1801 9008  
## # ... with 5 more variables: vaccd\_adults <dbl>, votes\_dem <dbl>,  
## # pct\_voted\_Biden <dbl>, total\_votes <dbl>, pct\_adult\_vaccd <dbl>

# It is my opinon that  
# the 3 outliers from HI  
# and the 3 from MA must be bad data   
# and I will remove them from the data set  
  
# !st Need to get the GA row out of the outliers group  
outliers <- outliers %>%  
 filter(state != "GA")  
  
# Remove the 6 outliers from the data set  
# I was so excited to stumble on this  
# anti-join (which is part of dplyr)  
all\_data <- anti\_join(all\_data, outliers)

## Joining, by = c("state", "pct\_adult\_seniors", "county", "POPESTIMATE", "AGE65PLUS\_TOT", "AGE18PLUS\_TOT", "vaccd\_adults", "votes\_dem", "pct\_voted\_Biden", "total\_votes", "pct\_adult\_vaccd")

nrow(all\_data)

## [1] 3090

# Let's take another look at that scatter plot  
# voting vs vaccines  
plot\_vote\_v\_vac2 <- ggplot(data = all\_data,  
 aes(x = pct\_voted\_Biden, y = pct\_adult\_vaccd ))+  
 geom\_point()+  
 ggtitle("Votes for Biden vs. Vaccined Adults",  
 subtitle = "in Percent of Adults, by US county")+  
 geom\_smooth(method=lm)  
plot\_vote\_v\_vac2

## `geom\_smooth()` using formula 'y ~ x'

Chart, scatter chart

Description automatically generated

# Create a scatter plot  
# %seniors vs %vaccinated  
plot\_65\_v\_vac <- ggplot(data = all\_data,  
 aes(x = pct\_adult\_seniors, y = pct\_adult\_vaccd ))+  
 geom\_point()+  
 ggtitle("Seniors vs. Vaccined Adults",  
 subtitle = "in Percent of Adults by US county")+  
 geom\_smooth(method=lm)  
plot\_65\_v\_vac

## `geom\_smooth()` using formula 'y ~ x'

Chart, scatter chart

Description automatically generated

# I'm curious about that one data point way off to the right  
# I'll just take a look at the ones showing more than 45%  
lots\_o\_seniors <- all\_data %>%  
 filter(pct\_adult\_seniors > 45)  
lots\_o\_seniors

## # A tibble: 5 x 11  
## state pct\_adult\_seniors county POPESTIMATE AGE65PLUS\_TOT AGE18PLUS\_TOT  
## <chr> <dbl> <chr> <dbl> <dbl> <dbl>  
## 1 AZ 49.4 La Paz County 21480 8956 18129  
## 2 FL 46.7 Charlotte Cou~ 194711 80251 172000  
## 3 FL 63.5 Sumter County 139018 82103 129393  
## 4 NM 49.0 Catron County 3623 1567 3195  
## 5 NM 47.9 Harding County 638 268 560  
## # ... with 5 more variables: vaccd\_adults <dbl>, votes\_dem <dbl>,  
## # pct\_voted\_Biden <dbl>, total\_votes <dbl>, pct\_adult\_vaccd <dbl>

# Those actually make sense  
# Sumter County FL is home to The Villages  
# And AZ and NM are popular for retirees too  
  
# Since polls show Trump beating Biden amoung seniors  
# I would expect a negative correlation between  
# percent of seniors and percent of Biden votes  
  
# Create a scatter plot  
# seniors vs voting  
plot\_65\_v\_dem <- ggplot(data = all\_data,  
 aes(x = pct\_adult\_seniors, y = pct\_voted\_Biden ))+  
 geom\_point()+  
 ggtitle("Seniors vs. Votes for Biden",  
 subtitle = "in Percent of Adults by US county")+  
 geom\_smooth(method=lm)  
plot\_65\_v\_dem

## `geom\_smooth()` using formula 'y ~ x'

Chart, scatter chart

Description automatically generated

# Now I want to view fav stats and histograms for each variable  
# looking for anything that jumps out as an anomaly (a zero value)  
# and looking to see if each has a roughly normal distribution  
  
# get fav stats for vaccinations  
favstats(~pct\_adult\_vaccd, data = all\_data)

## min Q1 median Q3 max mean sd n missing  
## 4.747234 45.9016 54.10984 62.90635 97.78822 53.98786 14.00694 3090 0

# create a histogram for vaccination rates  
histo\_vaccd2 <- ggplot(data = all\_data,  
 aes(x = pct\_adult\_vaccd))+  
 geom\_histogram(banwidth = 5)

## Warning: Ignoring unknown parameters: banwidth

histo\_vaccd2

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

Chart, histogram

Description automatically generated

#get fav stats for voting  
favstats(~pct\_voted\_Biden, data = all\_data)

## min Q1 median Q3 max mean sd n missing  
## 3.090909 20.86352 29.88519 42.10355 92.14969 33.12216 15.87332 3090 0

# create histogram for voting  
histo\_votes <- ggplot(data = all\_data,  
 aes(x = pct\_voted\_Biden))+  
 geom\_histogram(banwidth = 5)

## Warning: Ignoring unknown parameters: banwidth

histo\_votes

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

Chart, histogram

Description automatically generated

# The voting data isn't quite normally distributed  
# This is due to a large number of small (rural) counties  
# in which Biden did not perform well  
# and a small number of large counties   
# in which Biden did perform well.  
  
# First attempt linear model  
# using percent of seniors and Biden voters  
# vs adult vaccination rates  
model1 <- lm(pct\_adult\_vaccd ~ pct\_adult\_seniors + pct\_voted\_Biden, data = all\_data)  
summary(model1)

##   
## Call:  
## lm(formula = pct\_adult\_vaccd ~ pct\_adult\_seniors + pct\_voted\_Biden,   
## data = all\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -58.141 -5.152 1.599 7.499 40.800   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 32.68424 1.30489 25.048 < 2e-16 \*\*\*  
## pct\_adult\_seniors 0.19641 0.04116 4.772 1.91e-06 \*\*\*  
## pct\_voted\_Biden 0.49017 0.01411 34.729 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 11.82 on 3087 degrees of freedom  
## Multiple R-squared: 0.2878, Adjusted R-squared: 0.2874   
## F-statistic: 623.8 on 2 and 3087 DF, p-value: < 2.2e-16

# Second attempt linear model  
# using percent of Biden voters  
# vs adult vaccination rates  
model2 <- lm(pct\_adult\_vaccd ~ pct\_voted\_Biden, data = all\_data)  
summary(model2)

##   
## Call:  
## lm(formula = pct\_adult\_vaccd ~ pct\_voted\_Biden, data = all\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -56.821 -5.018 1.486 7.567 40.711   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 38.45145 0.49400 77.84 <2e-16 \*\*\*  
## pct\_voted\_Biden 0.46906 0.01345 34.87 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 11.87 on 3088 degrees of freedom  
## Multiple R-squared: 0.2826, Adjusted R-squared: 0.2823   
## F-statistic: 1216 on 1 and 3088 DF, p-value: < 2.2e-16

confint(model2)

## 2.5 % 97.5 %  
## (Intercept) 37.4828460 39.420057  
## pct\_voted\_Biden 0.4426915 0.495436

# Third attempt linear model  
# using percent of seniors and Biden voters interaction  
# vs adult vaccination rates  
model3 <- lm(pct\_adult\_vaccd ~ pct\_adult\_seniors \* pct\_voted\_Biden, data = all\_data)  
summary(model3)

##   
## Call:  
## lm(formula = pct\_adult\_vaccd ~ pct\_adult\_seniors \* pct\_voted\_Biden,   
## data = all\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -59.706 -5.036 1.628 7.480 40.675   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 39.830423 2.407168 16.547 < 2e-16 \*\*\*  
## pct\_adult\_seniors -0.089049 0.090701 -0.982 0.326285   
## pct\_voted\_Biden 0.286815 0.059305 4.836 1.39e-06 \*\*\*  
## pct\_adult\_seniors:pct\_voted\_Biden 0.008404 0.002381 3.530 0.000421 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 11.8 on 3086 degrees of freedom  
## Multiple R-squared: 0.2907, Adjusted R-squared: 0.29   
## F-statistic: 421.5 on 3 and 3086 DF, p-value: < 2.2e-16

# Get confidence intervals for model 3  
confint(model3)

## 2.5 % 97.5 %  
## (Intercept) 35.110608511 44.55023698  
## pct\_adult\_seniors -0.266889156 0.08879207  
## pct\_voted\_Biden 0.170533481 0.40309594  
## pct\_adult\_seniors:pct\_voted\_Biden 0.003736293 0.01307240

# get correlation for model 3  
# checking for issues with multicollinearity  
model3\_data <- all\_data%>%  
 select(pct\_adult\_vaccd,pct\_adult\_seniors,pct\_voted\_Biden,)  
cor(model3\_data, method = "pearson")

## pct\_adult\_vaccd pct\_adult\_seniors pct\_voted\_Biden  
## pct\_adult\_vaccd 1.00000000 -0.09776418 0.5315649  
## pct\_adult\_seniors -0.09776418 1.00000000 -0.3134117  
## pct\_voted\_Biden 0.53156487 -0.31341172 1.0000000

# A fresh look at the scatter plot   
plot\_65\_v\_vote <- ggplot(data = all\_data,  
 aes(x = pct\_adult\_seniors, y = pct\_voted\_Biden ))+  
 geom\_point()+  
 ggtitle("Seniors vs. Votes for Biden",  
 subtitle = "in Percent of Adults by US county")+  
 geom\_smooth(method=lm)  
plot\_65\_v\_vote

## `geom\_smooth()` using formula 'y ~ x'

Chart, scatter chart

Description automatically generated

```r  
# Forth linear model checking for co linearity between   
# percent of seniors and voting for Biden  
# using percent of seniors  
# vs percent of Biden voters  
model4 <- lm(pct\_voted\_Biden ~ pct\_adult\_seniors, data = all\_data)  
summary(model4)

##   
## Call:  
## lm(formula = pct\_voted\_Biden ~ pct\_adult\_seniors, data = all\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -34.407 -11.297 -3.014 9.516 54.069   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 56.70336 1.31406 43.15 <2e-16 \*\*\*  
## pct\_adult\_seniors -0.91390 0.04983 -18.34 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 15.08 on 3088 degrees of freedom  
## Multiple R-squared: 0.09823, Adjusted R-squared: 0.09793   
## F-statistic: 336.4 on 1 and 3088 DF, p-value: < 2.2e-16

confint(model4)

## 2.5 % 97.5 %  
## (Intercept) 54.126854 59.2798751  
## pct\_adult\_seniors -1.011603 -0.8161953

# Fifth attempt linear model  
# using percent of seniors  
# vs percent vaccd  
model5 <- lm(pct\_adult\_vaccd ~ pct\_adult\_seniors, data = all\_data)  
summary(model5)

##   
## Call:  
## lm(formula = pct\_adult\_vaccd ~ pct\_adult\_seniors, data = all\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -49.211 -8.126 0.059 9.086 42.548   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 60.47877 1.21522 49.768 < 2e-16 \*\*\*  
## pct\_adult\_seniors -0.25156 0.04608 -5.459 5.17e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 13.94 on 3088 degrees of freedom  
## Multiple R-squared: 0.009558, Adjusted R-squared: 0.009237   
## F-statistic: 29.8 on 1 and 3088 DF, p-value: 5.171e-08

confint(model5)

## 2.5 % 97.5 %  
## (Intercept) 58.0960508 62.861495  
## pct\_adult\_seniors -0.3419133 -0.161203

# None of these models explain much about the outcome  
  
# I'm curious if I might learn anything by  
#grouping counties that Biden won vs lost  
# then looking at seniors vs vacc  
# OR  
# grouping above/below median pct of senior  
# and looking at votes vs vacc  
  
# Create new columns for groupings  
  
group\_all\_data <- all\_data %>%  
 mutate(pct\_Biden\_group = ifelse(pct\_voted\_Biden >= 50, "Biden Won", "Biden Lost"), pct\_senior\_group = ifelse(pct\_adult\_seniors <= 25.5,"Below Median", "Above Median"))  
head(group\_all\_data)

## # A tibble: 6 x 13  
## state pct\_adult\_seniors county POPESTIMATE AGE65PLUS\_TOT AGE18PLUS\_TOT  
## <chr> <dbl> <chr> <dbl> <dbl> <dbl>  
## 1 AL 21.1 Autauga County 56145 9082 43128  
## 2 AL 27.4 Baldwin County 229287 49485 180842  
## 3 AL 25.5 Barbour County 24589 4982 19555  
## 4 AL 21.3 Bibb County 22136 3780 17710  
## 5 AL 24.6 Blount County 57879 10999 44681  
## 6 AL 22.3 Bullock County 9976 1763 7899  
## # ... with 7 more variables: vaccd\_adults <dbl>, votes\_dem <dbl>,  
## # pct\_voted\_Biden <dbl>, total\_votes <dbl>, pct\_adult\_vaccd <dbl>,  
## # pct\_Biden\_group <chr>, pct\_senior\_group <chr>

nrow(group\_all\_data)

## [1] 3090

# Create scatter plot of Votes vs Vacc  
# Grouped by percent of seniors  
grouped\_senior\_plot <- plot\_vote\_v\_vac2 +  
 aes(color = group\_all\_data$pct\_senior\_group)  
grouped\_senior\_plot

## `geom\_smooth()` using formula 'y ~ x'

Chart, scatter chart

Description automatically generated

# That doesn't seem insightful  
  
# Create scatter plot of Seniors vs Vacc  
# Grouped by Biden won/lost  
grouped\_vote\_plot <- plot\_65\_v\_vac+  
 aes(color = group\_all\_data$pct\_Biden\_group)  
grouped\_vote\_plot

## `geom\_smooth()` using formula 'y ~ x'

Chart, scatter chart

Description automatically generated

# This clearly shows that, as expected  
# counties that Biden won have higher vaccination  
# rates than counties that Biden lost  
# However, there doesn't apear to be any  
# significant impact based on the percent of seniors  
  
# I need some summary data for my write up  
  
seniors <- all\_data%>%  
 summarize(total\_seniors = sum(AGE65PLUS\_TOT), total\_adults = sum(AGE18PLUS\_TOT), total\_vaccd = sum(vaccd\_adults), total\_votes = sum(total\_votes))%>%  
 mutate (total\_pct\_seniors = 100\*total\_seniors/total\_adults,  
 total\_pct\_vaccd = 100\*total\_vaccd/total\_adults,  
 total\_pct\_voters = 100\*total\_votes/total\_adults)  
head(seniors)

## # A tibble: 1 x 7  
## total\_seniors total\_adults total\_vaccd total\_votes total\_pct\_seniors  
## <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 55148243 254526657 166977488 157139258 21.7  
## # ... with 2 more variables: total\_pct\_vaccd <dbl>, total\_pct\_voters <dbl>

1. Data includes independent cities that are not part of any county, the District of Columbia, and the Parishes of Louisiana. [↑](#footnote-ref-1)
2. Portsmouth Virginia is an independent city and not part of any county. [↑](#footnote-ref-2)