Data Analysis Project Final Project

05/04/2024

INTRODUCTION This data analysis project will examine how different factors, such as incarceration, wealth, internet access, business activity and race play a role in Baltimore City's 2020 general election voter turnout. It will also examine how voter turnout has changed in the last 10 years across gubernatorial and general elections as well as over longer periods of time and will compare political parties in the city. This is an especially newsworthy topic ahead of election season this November. I will ask these five questions to guide my findings:

1. Which five Baltimore City community statistical areas had the lowest voter turnout percentage in the 2020 general election? Which five had the highest?
2. Are there correlations between the Baltimore City 2020 general election voter turnout and factors including incarceration rate, business activity, wealth and internet access?
3. How has voter turnout in Baltimore City changed from 2010 to 2020? Which CSA had the biggest difference between the 2012 and 2020 general election?
4. In August 2020, which parts of the city had the most registered democrats? Which had the most republicans? Is there any correlation to race?

THE DATA Note: if you click on the Baltimore Neighborhood Indicators Alliance links, make sure the year is set to 2020 (Compare Years -> 2020 -> Update Your Selections) and the corresponding indicator on the left-hand side is clicked. It doesn't automatically apply these when I copy the link

1. pct_vote_csas - data from the Baltimore City Board of Elections compiled by the Baltimore Neighborhood Indicators Alliance that has the percent of people over the age of 18 who voted in the 2020 general election in each Baltimore City community statistical area. I exported the data from the website as an excel sheet and then saved it as a csv.
2. incarceration_x2020 - A data table scraped from a Prison Policy Initiative website with information about the incarceration rate in each Baltimore City community statistical areas in 2020. It has the community statistical area, the total population, the census population (I used the total population here), the rate per 100,000 and the number in

state prison. Rate per 100,000 is the number of imprisoned people divided by the total CSA population and then multiplied by 100,000. It allows ready comparison of the frequency of imprisonment between each CSA of different population sizes. I webscrape this later.

3. biz - data from InfoUsa compiled by the Neighborhood Indicators Alliance that has the number of businesses per 1,000 residents in each Baltimore City community statistical area. I exported the data from the website as an excel sheet and then saved it in excel as a csv.

4. baltcity_income - data from the US Census Bureau compiled from the Neighborhood Indicators Alliance that has the median household income in each Baltimore City community statistical area. I exported the data from the website as an excel sheet and then saved it as a csv.

5. internet - data from the US Census Bureau compiled by the Neighborhood Indicators Alliance that has the percent of households with no internet at home and community statistical area. I exported the data from the website as an excel sheet and then saved it in excel as a csv.

6. voting_2012 through voting_2022 - data from the Baltimore City Board of Elections compiled by the Neighborhood Indicators Alliance that has the voter turnout rate for each Baltimore City community statistical area. I exported the data from the website as an excel sheet and then saved it in excel as a csv.

7. trends_turnout - for this dataset I found it easiest to take screenshots from two websites and put them into Adobe Acrobat. For 1984 to 2008, I used this site , which has the Baltimore City voter turnout in each general election from 1984 to 2008. From 2012 to 2020, I used this site , which has the voter turnout in Baltimore City from 2012 to 2020. When I say "voter turnout" I mean percent of people of the voting age who voted in the election. For the second website you might have to click around as the settings I put might not save when I copy the URL.

8. maryland_voting - I downloaded this dataset from the U.S. Census Bureau website (table A-5a) and then downloaded it as a CSV. It three tables within the dataset, with rows for every state and then columns for each year from 2020 to 1972 that contain the voting rate for each corresponding state.

9. precinct_voters_x2020 - I used Tabula to get a table from this PDF from the Baltimore City Board of Elections. This data shows voter registration from August 2020, the most recent dataset from 2020. I used this dataset because it had figures for the number of voters registered in each precinct (a more detailed way to show voting figures) and a breakdown of their political parties. This is a citywide registration number taken on August 27, 2020. It contains voting precinct number (rows), number of democrats, number of republicans, and a handful of other political parties, including a column for other and total.

10. demo_voting - I used Tabula to get a table from this PDF from the Maryland Department of Planning. A precinct represents a small area within the city, so although each precinct has one voting place within it, these counts are for the number of people in each precinct by population, not the number of voters. It has information for the congressional/legislatie district a voting precinct is in, the voting precinct number (also called voting district VTD), precinct name, 2020 census and total population, breakdowns by number of people with one race, white alone, black/african american alone, american indian and alaska native alone, asian alone, native hawaiian/pacific islander alone, some other race alone, 2+ races alone, hispanic/latino alone, and total person over 18. Because I'm comparing race to voter registration data, I'm going to use the figure for total people over 18 (the voting age). This dataset has rows for every voting precinct in every county in Maryland, including Baltimore City.

NOTE I decided to use Flourish for all my visualizations instead of R because there's much more flexibility for interactivity and styling. I also wasn't sure if I was supposed to note exactly where I took feedback, but I pretty much overhauled my entire project from Tuesday so hopefully it's clear. All the data is uploaded to Github on my end but there's been a few instances where the files unsave in my folder and I have to redownload them through icloud, but I don't think this will be an issue since I see that all the files are on Github.

SET UP

```
#libraries and settings
library(tidyverse)
library(janitor)
library(formattable)
library(dplyr)
library(rvest)

#turn off scientific notation
options(scipen=999)
```

UPLOAD DATA pct_vote_csas This dataset has a row for all of Baltimore City, which is not one of the official 55 CSAs, so I will remove that row

```
#upload csv and clean names/rows
pct_vote_csas <- read_csv("data/pct_vote_csa_clean.csv")

pct_vote_csas <- pct_vote_csas%>%
  rename(csa = 1, pct_voters = 2)
```

incarceration_x2020

```
#webscrape data table from prison policy initiative website
incarceration_x2020 <- "https://www.prisonpolicy.org/origin/md/2020/baltimore_csa.html"
incarceration_x2020 <- incarceration_x2020 %>%
  read_html() %>%
  html_table()
incarceration_x2020 <- incarceration_x2020[[1]]%>%
  clean_names()%>%
  rename(csa = 1, rate_per_x100000 = 5)%>%
  select(csa, rate_per_x100000)
```

baltcity_income This dataset also has a row for all of Baltimore City, which is not one of the official 55 CSAs, so I will remove that row

```
baltcity_income <- read_csv("data/baltcity_income.csv")%>%
  clean_names()%>%
  rename(csa = community, median_income = x2020_data)%>%
  filter(csa != "Baltimore City")
```

biz This dataset also has a row for all of Baltimore City, which is not one of the official 55 CSAs, so I will remove that row

```
biz <- read_csv("data/smallbiz_csa.csv")

biz <- biz%>%
  clean_names()%>%
  rename(csa = community, num_biz_x1000 = x2020_data)%>%
  filter(csa != "Baltimore City")
```

internet This dataset also has a row for all of Baltimore City, which is not one of the official 55 CSAs, so I will remove that row

```
internet <- read_csv("data/internet.csv")

internet <- internet%>%
  clean_names()%>%
  rename(csa = community, pct_no_internet = x2020_data)%>%
  filter(csa != "Baltimore City")
```

voting_2010 through voting_2020 These datasets also have a row for all of Baltimore City, which is not one of the official 55 CSAs, so I will remove that row

voting_2010

```
#voter turnout in the 2010 election
voting_2010 <- read_csv("data/voting_2010.csv")%>%
  clean_names()%>%
  rename(csa = community, pct_voters = x2010_data)%>%
  filter(csa != "Baltimore City")
```

voting_2012

```
#voter turnout in the 2012 election
voting_2012 <- read_csv("data/voting_2012_clean.csv")%>%
  clean_names()%>%
  rename(csa = community, pct_voters = x2012_data)%>%
  filter(csa != "Baltimore City")
```

voting_2014

```
#voter turnout in the 2014 election
voting_2014 <- read_csv("data/voting_2014.csv")%>%
  clean_names()%>%
  rename(csa = community, pct_voters = x2014_data)%>%
  filter(csa != "Baltimore City")
```

voting_2016

```
#voter turnout in the 2016 election
voting_2016 <- read_csv("data/voting_2016.csv")%>%
  clean_names()%>%
  rename(csa = community, pct_voters = x2016_data)%>%
  filter(csa != "Baltimore City")
```

voting_2018

```
#voter turnout in the 2018 election
voting_2018 <- read_csv("data/voting_2018.csv")%>%
  clean_names()%>%
  rename(csa = community, pct_voters = x2018_data)%>%
  filter(csa != "Baltimore City")
```

voting_2022

```
#voter turnout in the 2022 election
voting_2016 <- read_csv("data/voting_2016.csv")%>%
  clean_names()%>%
  rename(csa = community, pct_voters = x2016_data)%>%
  filter(csa != "Baltimore City")
```

trends_turnout

```
trends_turnout <- read_csv("data/trends_turnout.csv")

trends_turnout <- trends_turnout%>%
  rename(pct_turnout = 2)
```

maryland_voting This one needs a lot of cleaning. It gives me three tables in one dataset with one for elections from 2020 to 2004, one for elections from 2000 to 1984 and one from elections from 1980 to 1972. Because Maryland only has a breakdown of Baltimore City data from 1984 on, I'll just be using the first two tables. I'm going to seperate the two tables and then rejoin them. Note: these tables will have 51 ones because it includes a separete count for Washington, DC.

```
#upload data
maryland_voting <- read_csv("data/maryland_voter_turnout_historic.csv")

#change this column name to make join easier later
maryland_voting <- maryland_voting%>%
  rename(state = 1)

#make two tables into separate datasets and rejoin
placeholder1 <- maryland_voting%>%
  slice(7:57)%>%
  rename("2020" =2, "2016" = 4, "2012" = 6, "2008" = 8, "2004" = 10)

placeholder2 <- maryland_voting%>%
  slice(75:125)%>%
  rename("2000" = 2, "1996" =4, "1992" = 6, "1988" = 8, "1984" = 10)

#join to make cleaner table - I'm going to use the total reported voters instead of just the
maryland_voting_clean <- inner_join(placeholder1, placeholder2, by = "state")%>%
  select("2020", "2016", "2012", "2008", "2004", "2000", "1996", "1992", "1988", "1984", sta

#filter to just maryland
```

```
maryland_voting_clean <- maryland_voting_clean%>%
  filter(state == "Maryland")%>%
  select(-state)

#switch rows with columns
maryland_cleanest_voting <- t(maryland_voting_clean)

#rename columns
colnames(maryland_cleanest_voting) <- c("pct_turnout")
```

precinct_voters_x2020

```
#upload csv
precinct_voters_x2020 <- read_csv("data/cleaned_precinct_x2020.csv")

#remove rows with NAs and get rid of extra column copied from PDF
precinct_voters_x2020 <- precinct_voters_x2020 %>%
  filter(complete.cases(.))%>%
  slice(-111)%>%
  clean_names()%>%
  select(precinct, dem, rep, total)
```

demo_voting

```
#voting by demographic 2020
demo_voting <- read_csv("data/demo_voting_x2020_finalll.csv", locale=locale(encoding="latin1"
  clean_names
#delete top seven rows for worcester copied from the page before and column three filled wit
demo_voting <- demo_voting[-c(1:7), -3]%>%
  slice(-1)%>%
  rename(precinct = voting_district_vtd, precinct_pop = total_population, black_aa = black_o
  select(precinct, precinct_pop, black_aa)
```

Q1: Which five Baltimore City community statistical areas had the lowest voter turnout percentage in the 2020 general election? Which five had the highest?

```
#use slicemax/min to make new tables with top and bottom five csas by voter turnout
pct_vote_csas%>%
  slice_min(pct_voters, n = 5)
```

```
# A tibble: 5 x 2
  csa                                     pct_voters
  <chr>                                        <dbl>
1 Southwest Baltimore                           41.1
2 Madison/East End                              41.6
3 Brooklyn/Curtis Bay/Hawkins Point             42.7
4 Sandtown-Winchester/Harlem Park               44.1
5 Poppleton/The Terraces/Hollins Market         46.1
```

```
#Southwest Baltimore: 41.1%, Madison/East End: 41.6%, Brooklyn/Curtis Bay/Hawkins Point: 42.7

pct_vote_csas%>%
  slice_max(pct_voters, n = 5)
```

```
# A tibble: 5 x 2
  csa                                    pct_voters
  <chr>                                       <dbl>
1 Greater Roland Park/Poplar Hill              80.1
2 Mount Washington/Coldspring                  77
3 South Baltimore                              75
4 North Baltimore/Guilford/Homeland            74.6
5 Medfield/Hampden/Woodberry/Remington         72.9
```

```
#Greater Roland Park/Poplar Hill: 80.1%, Mount Washington/Coldspring: 77%, South Baltimore: 7
```

Let's visualize this! I created a custom map in flourish by uploading a geojson file with the Baltimore City CSA boundaries, which I found through Open Baltimore . I then uploaded my csv for pct_vote_csas to fill in the data points.

```
cat('<div class="flourish-embed flourish-map" data-src="visualisation/17796037"><script src="
```

Looking at this visualization, it's clear the areas with the highest voter turnout are in the middle areas of Baltimore, especially in the Inner Harbor area stretching north to Homeland and Guilford, with east and west Baltimore showing significantly less voter turnout. Sixteen CSAs had a voter turnout of less than 50%, and they were all concentrated in similar areas.

Q2: Are there correlations between the Baltimore City 2020 general election voter turnout and factors including incarceration rate, business activity, wealth, internet access and race?

Let's compare all these factors in the five CSAs with the highest and lowest voter turnout rates

I'll start with incarceration rate and voter turnout

```
#join incarceration_x2020 and pct_vote_csas datasets
incarceration_voting <- inner_join(incarceration_x2020, pct_vote_csas, by = "csa")

incarceration_voting %>%
  slice_min(pct_voters, n = 5)
```

```
# A tibble: 5 x 3
  csa                                    rate_per_x100000 pct_voters
  <chr>                                  <chr>                 <dbl>
1 Southwest Baltimore                    2,223                  41.1
2 Madison/East End                       2,528                  41.6
3 Brooklyn/Curtis Bay/Hawkins Point      1,204                  42.7
4 Sandtown-Winchester/Harlem Park        2,563                  44.1
5 Poppleton/The Terraces/Hollins Market 1,623                  46.1
```

```
incarceration_voting %>%
  slice_max(pct_voters, n = 5)
```

```
# A tibble: 5 x 3
  csa                                   rate_per_x100000 pct_voters
  <chr>                                 <chr>                 <dbl>
1 Greater Roland Park/Poplar Hill       14                     80.1
2 Mount Washington/Coldspring           102                    77
3 South Baltimore                       97                     75
4 North Baltimore/Guilford/Homeland     93                     74.6
5 Medfield/Hampden/Woodberry/Remington 225                    72.9
```

```
#Incarceration rate per 100,000 people in five areas of lowest voter turnout:
#Southwest Baltimore: 2,223, Madison/East End: 2,528, Brooklyn/Curtis Bay/Hawkins Point: 1,20

#Incarceration rate per 100,000 people in five areas of highest voter turnout:
#Greater Roland Park/Poplar Hill: 14, Mount Washington/Coldspring: 102, South Baltimore: 97,
```

Let's visualize this!

```
cat('<div class="flourish-embed flourish-map" data-src="visualisation/17809409"><script src="
```

This visualization shows that there is a correlation between incarceration rate and voter turnout. Looking at the map of voter turnout next to the map of incarceration rates, it's clear that the two are generally inverses of on another, with areas in the north and middle of Baltimore having a low incarceration rate and high voter turnout and east and west Baltimore having a higher incarceration rate and low voter turnout.

Another visualization. This time, a scatterplot

```
cat('<div class="flourish-embed flourish-scatter" data-src="visualisation/17795149"><script s
```

This scatterplot shows the negative relationship between incarceration rate and voter turnout. As the rate of incarceration per 100,000 people increases, the percent of registered voters decreases.

Just to be really sure, let's compute some correlation coefficients. Values close to 1 or -1 indicate a strong positive or negative correlation, while values closer to zero indicate a weak correlation.

This is not something I would put in an actual article since it's a more complex concept for readers to digest, but it's useful for checking my work and determining whether these factors are worth investigating at all.

Correlation between incarceration rate and voter turnout.

```
#convert all values to numeric for calculation
incarceration_voting$rate_per_x100000 <- gsub("[, ]", "", incarceration_voting$rate_per_x1000
incarceration_voting$rate_per_x100000 <- as.numeric(incarceration_voting$rate_per_x100000)

#compute correlation coefficient
cc_incarceration <- cor(incarceration_voting$pct_voters, incarceration_voting$rate_per_x1000

cc_incarceration
```

```
[1] -0.8162851
```

A correlation coefficient of -0.82 indicates a strong negative correlation, meaning areas of lower incarceration rates likely have higher voter turnout rates.

Now let's look at whether there's a correlation between wealth and voter turnout

```
#join baltcity_income and pct_vote_csas
wealth_voting <- inner_join(pct_vote_csas, baltcity_income, by = "csa")

wealth_voting %>%
  slice_min(pct_voters, n = 5)
```

```
# A tibble: 5 x 3
  csa                                     pct_voters median_income
  <chr>                                        <dbl>         <dbl>
1 Southwest Baltimore                           41.1        29768.
2 Madison/East End                              41.6        37303.
3 Brooklyn/Curtis Bay/Hawkins Point             42.7        32599.
4 Sandtown-Winchester/Harlem Park               44.1        26690.
5 Poppleton/The Terraces/Hollins Market         46.1        23374.
```

```r
wealth_voting %>%
  slice_max(pct_voters, n = 5)
```

```
# A tibble: 5 x 3
  csa                                    pct_voters median_income
  <chr>                                       <dbl>         <dbl>
1 Greater Roland Park/Poplar Hill              80.1       125869.
2 Mount Washington/Coldspring                  77          86715.
3 South Baltimore                              75         124827.
4 North Baltimore/Guilford/Homeland            74.6       114247.
5 Medfield/Hampden/Woodberry/Remington         72.9        72757.
```

```r
#median household income in areas of lowest voter turnout (rounded to nearest dollar):
#Southwest Baltimore: $29,768, Madison/East End: $37,303, Brooklyn/Curtis Bay/Hawkins Point:

#median household income in areas of highest voter turnout:
#Greater Roland Park/Poplar Hill: $125,869, Mount Washington/Coldspring: 86,715, South Baltin
```

Let's visualize this one, too

```r
cat('<div class="flourish-embed flourish-map" data-src="visualisation/17824036"><script src="
```

Looking at this map shows that areas of higher median household income are concentrated in the middle of Baltimore City and in the northern areas. This is a similiar pattern to the voter turnout rate map.

Another visualization

```r
cat('<div class="flourish-embed flourish-scatter" data-src="visualisation/17824265"><script s
```

This scatterplot shows the positive relationship between median household income and voter turnout. As the median household income increases, the percent of registered voters decreases.

Correlation between median household income and voter turnout

```
cc_wealth <- cor(wealth_voting$pct_voters, wealth_voting$median_income, method = "pearson", u

cc_wealth
```

```
[1] 0.814721
```

A correlation coefficient of 0.81 indicates a strong positive correlation between voter turnout and median household income.

Number of businesses per 1,000 people and voter turnout

```
#join biz and pct_vote_csas dataset
biz_voting <- inner_join(biz, pct_vote_csas, by = "csa")

biz_voting %>%
  slice_min(pct_voters, n = 5)
```

```
# A tibble: 5 x 3
  csa                                   num_biz_x1000 pct_voters
  <chr>                                         <dbl>      <dbl>
1 Southwest Baltimore                            18.4       41.1
2 Madison/East End                               19.2       41.6
3 Brooklyn/Curtis Bay/Hawkins Point              13.4       42.7
4 Sandtown-Winchester/Harlem Park                10.4       44.1
5 Poppleton/The Terraces/Hollins Market          14.2       46.1
```

```
biz_voting %>%
  slice_max(pct_voters, n = 5)
```

```
# A tibble: 5 x 3
  csa                                  num_biz_x1000 pct_voters
  <chr>                                        <dbl>      <dbl>
1 Greater Roland Park/Poplar Hill               33.5       80.1
2 Mount Washington/Coldspring                   31.4       77
3 South Baltimore                               21.9       75
4 North Baltimore/Guilford/Homeland             16.3       74.6
5 Medfield/Hampden/Woodberry/Remington          35.4       72.9
```

```
#number of businesses per 1,000 people in five areas of lowest voter turnout:
#Southwest Baltimore: 18.4, Madison/East End: 19.2, Brooklyn/Curtis Bay/Hawkins Point: 13.4, 

#number of businesses per 1,000 people in five areas of highest voter turnout:
#Greater Roland Park/Poplar Hill: 33.5, Mount Washington/Coldspring: 31.4, South Baltimore: 
```

There doesn't seem to be a significant correlation between number of businesses per 1,000 residents and voter turnout. While looking at just the top 5 top and bottom areas of high/low voter turnout seems like areas of lower voter turnout have less businesses, there are quite a few outliers, including Oldtown/Middle East, which was one of the lowest voter turnout areas, but had the second highest amount of businesses per 1,000 residents. This could be for a variety of reasons. For one, many of the areas that have high voter turnout are in wealthier, suburban areas with less concentration of businesses.

Correlation between number of businesses per 1000 residents and voter turnout

```
#compute correlation coefficient
cc_biz <- cor(biz_voting$pct_voters, biz_voting$num_biz_x1000, method = "pearson", use = "co

cc_biz
```

```
[1] -0.02123877
```

There is little to no correlation between the amount of businesses and voter turnout, as indicated by a correlation coefficient of -0.02

Percent of households without internet and voter turnout

```
#join internet and pct_vote_csas dataset
internet_voting <- inner_join(internet, pct_vote_csas, by = "csa")

internet_voting %>%
  slice_min(pct_voters, n = 5)
```

```
# A tibble: 5 x 3
  csa                                    pct_no_internet pct_voters
  <chr>                                            <dbl>      <dbl>
1 Southwest Baltimore                               27.6       41.1
2 Madison/East End                                  16.7       41.6
3 Brooklyn/Curtis Bay/Hawkins Point                 30.4       42.7
4 Sandtown-Winchester/Harlem Park                   38.6       44.1
5 Poppleton/The Terraces/Hollins Market             25.1       46.1
```

```r
internet_voting %>%
  slice_max(pct_voters, n = 5)
```

```
# A tibble: 5 x 3
  csa                                      pct_no_internet pct_voters
  <chr>                                              <dbl>      <dbl>
1 Greater Roland Park/Poplar Hill                      3.4       80.1
2 Mount Washington/Coldspring                          8.9       77
3 South Baltimore                                      3.5       75
4 North Baltimore/Guilford/Homeland                    6.1       74.6
5 Medfield/Hampden/Woodberry/Remington                 9.1       72.9
```

```r
#Percentage of homes without internet access in the five areas of lowest voter turnout:
#Southwest Baltimore: 27.6%, Madison/East End: 16.7%, Brooklyn/Curtis Bay/Hawkins Point: 30.4

#Percentage of homes without internet access in the five areas of highest voter turnout:
#Greater Roland Park/Poplar Hill: 3.4%, Mount Washington/Coldspring: 8.9%, South Baltimore: 3
```

Let's do another visualization

```r
cat('<div class="flourish-embed flourish-scatter" data-src="visualisation/17824515"><script s
```

This scatterplot shows the negative relationship between percent of households without internet access and voter turnout. As the percent of households without internet increases, the percent of registered voters decreases.

Correlation between percent of households without internet access and voter turnout

```r
cc_internet <- cor(internet_voting$pct_voters, internet_voting$pct_no_internet, method = "pea

cc_internet
```

```
[1] -0.7095833
```

```r
#A correlation coefficient of -0.71 indicates a strong negative correlation between voter tur
```

Q3: How has voter turnout in Baltimore City changed from 2010 to 2020? Which CSA had the biggest difference between the 2012 and 2020 general election?

Let's look at community statistical areas first

```
#rename row from "&" to "and" to match up with the rest of the voting data used later
pct_vote_csas$csa <- gsub("&", "and", pct_vote_csas$csa)

#make one dataset for combined 2012 and 2020
voting_x2012_x2020 <- inner_join(voting_2012, pct_vote_csas, by = "csa")

voting_x2012_x2020 <- voting_x2012_x2020 %>%
  rename(x2012_turnout = 2, x2020_turnout = 3) %>%
  mutate(pct_change = ((x2020_turnout - x2012_turnout) / x2012_turnout) * 100)

voting_x2012_x2020%>%
  slice_max(pct_change)
```

```
# A tibble: 1 x 4
  csa                            x2012_turnout x2020_turnout pct_change
  <chr>                                  <dbl>         <dbl>      <dbl>
1 Orangeville/East Highlandtown           31.5          51.9       64.8
```

```
#Orangeville/East Highlandtown had the biggest change, with an almost 65% increase in voter t
```

I'm going to make one big dataset to visualize this

```
voting_all <- inner_join(voting_2010, voting_2012, by = "csa") %>%
                inner_join(voting_2014, by = "csa") %>%
                inner_join(voting_2016, by = "csa") %>%
                inner_join(voting_2018, by = "csa") %>%
                inner_join(pct_vote_csas, by = "csa")

voting_all <- voting_all%>%
  rename(x2010_turnout = 2, x2012_turnout = 3, x2014_turnout = 4, x2016_tunout = 5, x2018_tu
```

This is easiest to see through visualizations. Press play to see change overtime between gubernational and general elections in the last 10 years

```
cat('<div class="flourish-embed flourish-map" data-src="visualisation/17832534"><script src="
```

Gubernatorial elections tend to get less voter turnout than general elections, which is unsurprising as that is a well-documented phenomenon. But it seems that 2020 has the highest voter turnout out of the 10-year period. It also seems that the "Black Butterfly White L" phenomenon becomes more pronounced over time

Let's draw even more conclusions by looking at Baltimore City as a whole. For this I'm just going to compare general election turnouts.

```
cat('<div class="flourish-embed flourish-chart" data-src="visualisation/17832645"><script src
```

For this chart, I uploaded my trends_turnout csv to flourish. Looking at a wider spread of data shows that voter turnout has seen an overall decline between 1984 and 2020 in Baltimore City.

Q4: In August 2020, which parts of the city had the most registered democrats? Which had the most republicans?

```
#add new column for percent of whole
precinct_voters_x2020 <- precinct_voters_x2020%>%
  mutate(pct_dem = (dem / total)*100)%>%
  mutate(pct_rep = (rep / total)*100)

precinct_voters_x2020%>%
  slice_min(pct_dem)
```

```
# A tibble: 1 x 6
  precinct   dem   rep total pct_dem pct_rep
  <chr>    <dbl> <dbl> <dbl>   <dbl>   <dbl>
1 23-003    1030   412  1939    53.1    21.2
```

```
#the precinct with the lowest percent of democrats still had 53% in August 2020, which means

precinct_voters_x2020%>%
  slice_max(pct_rep)
```

```
# A tibble: 1 x 6
  precinct   dem   rep total pct_dem pct_rep
  <chr>    <dbl> <dbl> <dbl>   <dbl>   <dbl>
1 26-013     411   183   773    53.2    23.7
```

```
#the precinct with the highest percent of republicans only had 23.7% in August 2020
```

Let's see if there's a correlation with the amount of Black residents who are over the voting age of 18

```r
#add a column for percent Black residents of whole population in precinct
demo_voting <- demo_voting%>%
  mutate(pct_black = (black_aa / precinct_pop)*100)
```

Voting precincts don't make much sense without a visualization. I uploaded the geojson for 2020 voting precincts on the Open Baltimore site I got the first geojson. This map has a toggle which allows users to switch from concentration of Black residents over the age of 18 by precinct and concentration of democrats

```r
cat('<div class="flourish-embed flourish-map" data-src="visualisation/17835622"><script src="
```

There were higher concentrations of registered democrats in east and west Baltimore City in August 2020 than in the north or south. There was also a significantly higher concentration of Black residents over the age of 18 in east and west baltimore.

Let's look at republicans now

```r
cat('<div class="flourish-embed flourish-map" data-src="visualisation/17835183"><script src="
```

There are not many republicans in Baltimore City. Those that are reside in the middle and more rural parts of the city. Note: sometimes this graphic doesn't load right away on the page so you might have to reload and it should pop up.