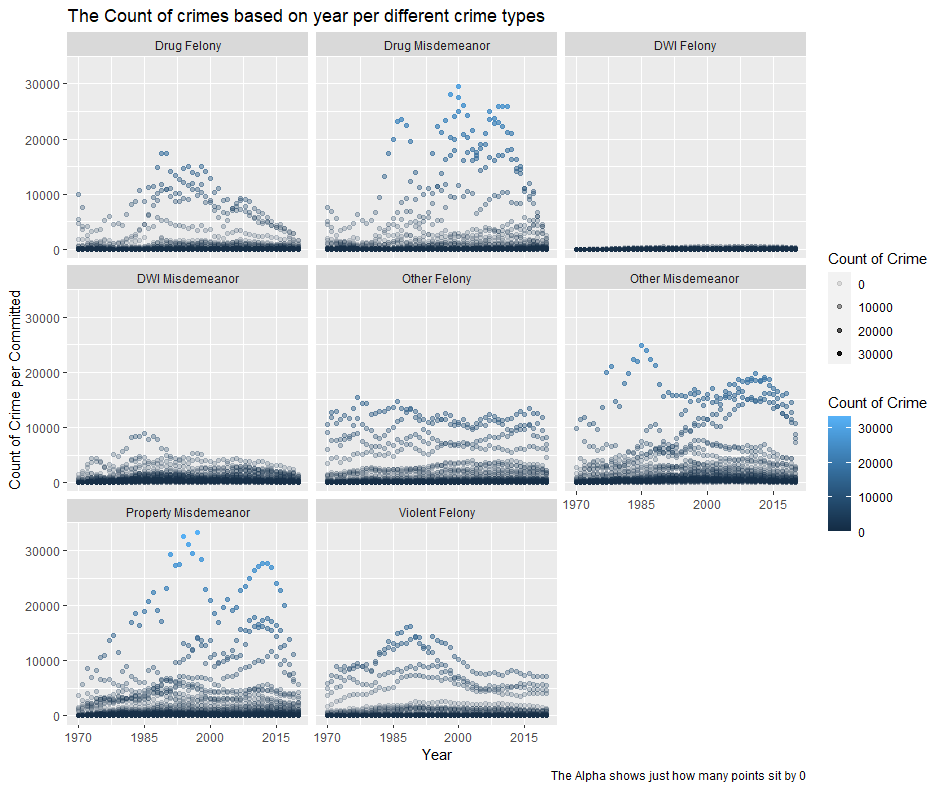
R FINAL

SECTION 1:

GRAPH 1:



In this graph we looked at the count of crimes per each type based on the year. This can be informative as to get a visual on if there are outliers of counts of these crimes and why averages may be skewed on a graph with this much information. It also reminds us of the difference in county size with NYC being a Megacity (a city with population over 10million). We will analyze these counties further later. But it also shows us the difference between the counts of each crime and how many counties have low counts of even the highest counts crimes like Drug and Property crimes.

CODE:

arrests %>%

filter(crimes != "Misdemeanor Total") %>%

ggplot() +

geom\_point(aes(Year, count\_per\_crime,color = count\_per\_crime,alpha = count\_per\_crime)) +

scale\_x\_continuous(limits = c(1970, 2020), breaks = seq(1970, 2020, 15)) +

facet\_wrap(~ crimes)+

labs(title = 'The Count of crimes based on year per different crime types',

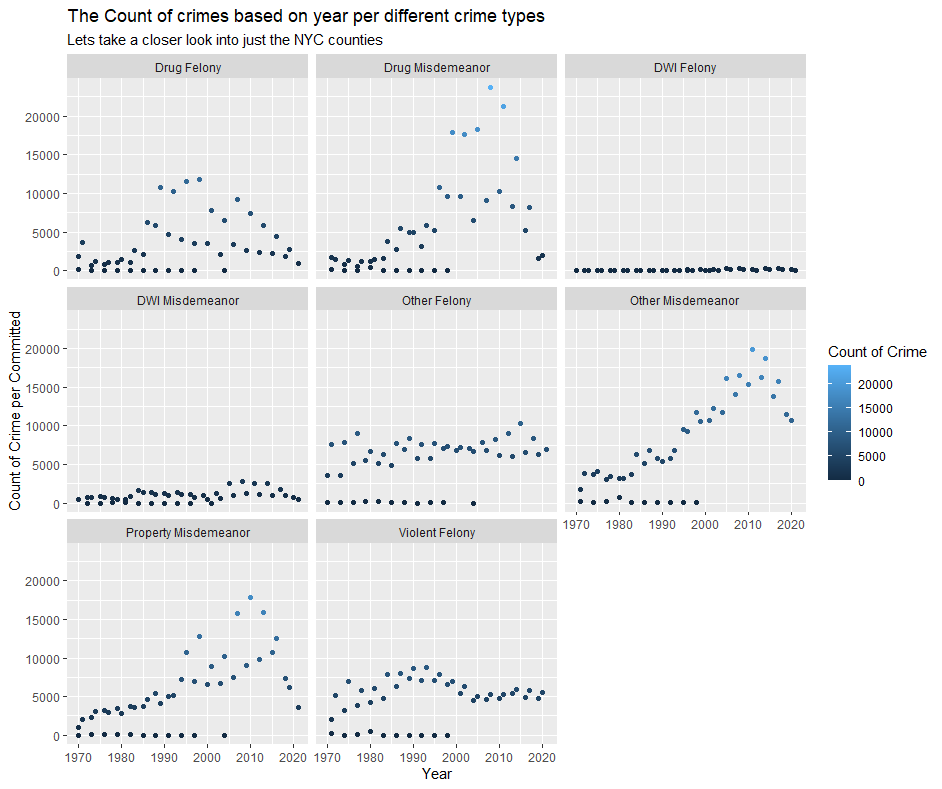
x = 'Year',

y = 'Count of Crime per Committed',

color = 'Count of Crime',

alpha = 'Count of Crime',

caption = 'The Alpha shows just how many points sit by 0')

GRAPH 2:

Lastly, I wanted to investigate the different crimes of each county in NYC. We see that 3 on the dataset worked but there are officially 5. (THANK YOU NY GOVERNMENT, …) But we see that they all follow roughly the same trends where they seem to go in the same motion, taking some of their extreme outliers. Of course, there seems to be other outlier counties, and this could be their other cities and counties that I didn’t realize were NYC counties due to change of naming convention. But this can be very impactful to see if your county is doing well as of recently, like if it has an upward/downward trend in some crimes. One thing I would like to note is the high counts of Other Misdemeanors! I don’t know why this would be other than strict city laws.

CODE:

arrests %>%

filter(crimes != "Misdemeanor Total") %>%

filter(County == c("Bronx","Unknown NYC county","Queens"))%>%

ggplot() +

geom\_point(aes(Year, count\_per\_crime,color = count\_per\_crime)) +

facet\_wrap(~ crimes)+

labs(title = 'The Count of crimes based on year per different crime types',

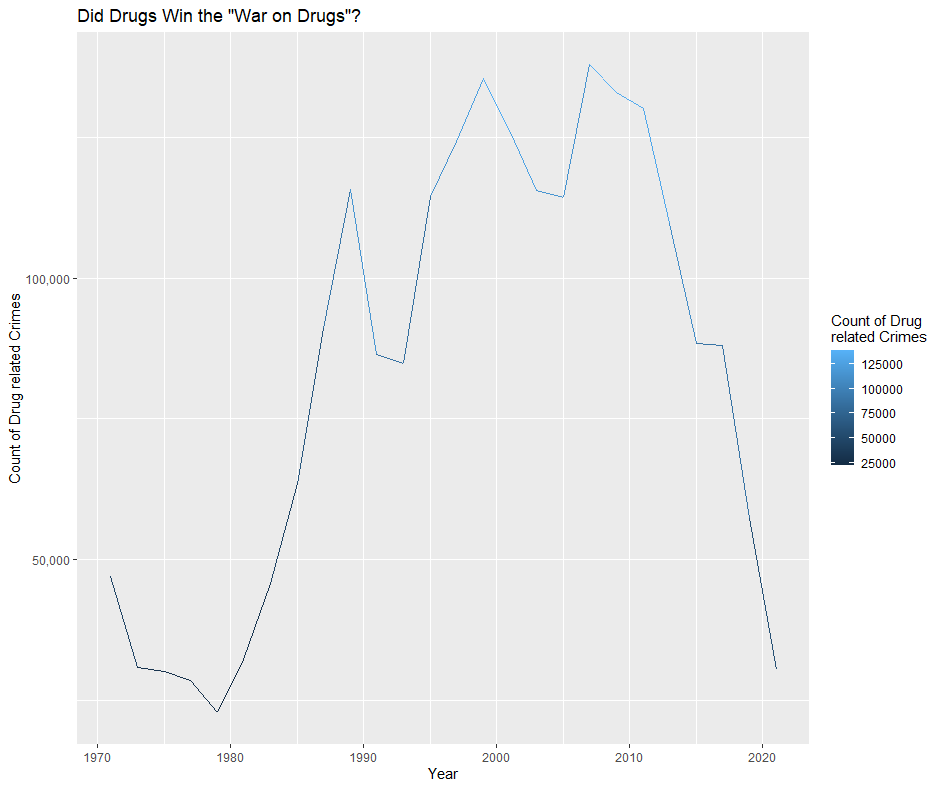
x = 'Year',

y = 'Count of Crime per Committed',

color = 'Count of Crime',

subtitle = 'Lets take a closer look into just the NYC counties')

Graph 3:



I found a large interest in the Drug Crimes and remembered the war on terror… It appears that they are starting to apprehend many people who are committing drug crimes on Felonies or Misdemeanors as the counts went high but seem to start dropping which is great for the state’s residents. Keeping drugs off the street is a very strong tool to keep them safe for there friends and families. But this could also be due to the lowering population as people seem to be moving away from New York.

CODE:

arrests%>%

filter(crimes == c("Drug Misdemeanor","Drug Felony"))%>%

group\_by(Year)%>%

mutate(drug\_user\_caught = sum(count\_per\_crime))%>%

ungroup()%>%

ggplot()+

geom\_line(aes(Year,drug\_user\_caught,color = drug\_user\_caught))+

scale\_y\_continuous(labels = scales::comma\_format())+

labs(title = 'Did Drugs Win the "War on Drugs"?',

x = 'Year',

y = 'Count of Drug related Crimes',

color = 'Count of Drug \nrelated Crimes')

Graph 4:A picture containing text, diagram, screenshot, plot

Description automatically generated

This breaks down all the crimes averages. I think that it’s a good thing to note that we see a large crime spike in all categories in around the 1990s to 2010s. I would assume this is due to lack of policing or lower population like I said before. This appears to be a very political point where they seem to try to limit or grow these numbers artificially. We need to remember that if the police/judges don’t do anything they are not counted where if they care about fining and locking up then they will be counted so this isn’t proving that less people are committing crimes (well it could but I don’t think in these visualizations that’s the direct reason) but that they aren’t caring. This could be a mix of counts going down on judges/police not caring as much and the relaxing of different laws.

CODE:

arrestsmutate%>%

filter(crimes != "Misdemeanor Total")%>%

ggplot()+

geom\_line(aes(Year,avgpercrimes,color = avgpercrimes))+

facet\_wrap(~crimes)+

labs(title = 'Average crime counts based on year per crime',

x = 'Year',

y = 'Average count of crime committed',

color = 'Average count of\ncrime committed')

Graph 5:A picture containing text, screenshot, diagram

Description automatically generated

In this graph we see the average again on the line graph, then we see all the counts on the scatter plot. This is cool to see the amounts go up around the midpoint of their time in the office and back down around the end/beginning. This shows the political cycle at its prime. Make it look good for elections and don’t care much otherwise! The average should stay natural due to this not being a large talking point in political campaigns at a high(governor) level but cared about local office which has high turnover.

CODE:

gov\_arrests %>%

filter(crimes != "Misdemeanor Total") %>%

group\_by(crimes, Year) %>%

mutate(avgcrime = mean(count\_per\_crime)) %>%

ungroup() %>%

ggplot() +

geom\_point(aes(Year, count\_per\_crime,color = Party))+

geom\_line(aes(Year, avgcrime), color = 'Black', size = 1)+

facet\_wrap(~ crimes) +

scale\_x\_continuous(limits = c(1970, 2020), breaks = seq(1970, 2020, 15)) +

scale\_color\_manual(values = c('red', 'blue'),

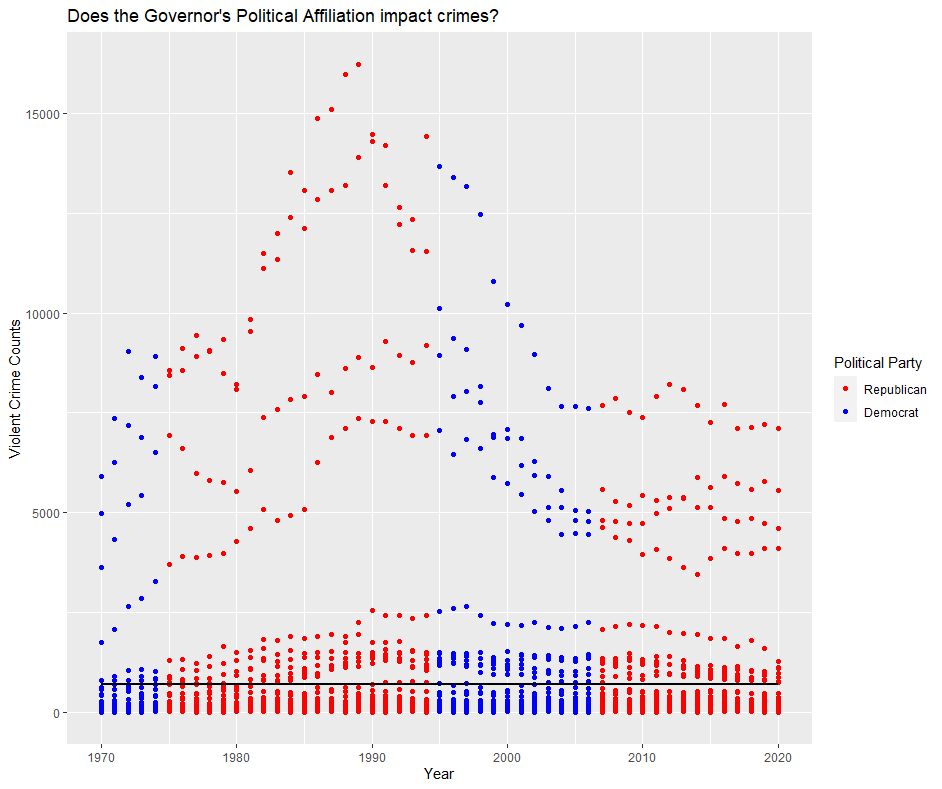
labels = c('Republican', 'Democrat')) +

labs(title = "Does the Governor's Political Affiliation impact crimes?",

x = 'Year',

y = 'Average Crime Counts',

color = 'Political Party')

Graph 6: 

On this graph we look at the total Violent Felonies closer so it’s more visible and we can see the average line better where it appears to the naked eye to go down. There isn’t much more to note that we didn’t see before. I would be curious to see what counties these weird crazy outliers are and if we took them out the line of best fit would stay close to where it is now.

CODE:

## THIS WAS MORE TO MEET A REQUIREMENT OF FILTERING

gov\_arrests %>%

filter(crimes == "Violent Felony") %>%

mutate(avgcrime = mean(count\_per\_crime)) %>%

ggplot() +

geom\_point(aes(Year, count\_per\_crime,color = Party))+

geom\_line(aes(Year, avgcrime), color = 'Black', size = 1)+

scale\_x\_continuous(limits = c(1970, 2020)) +

scale\_color\_manual(values = c('red', 'blue'),

labels = c('Republican', 'Democrat')) +

labs(title = "Does the Governor's Political Affiliation impact crimes?",

x = 'Year',

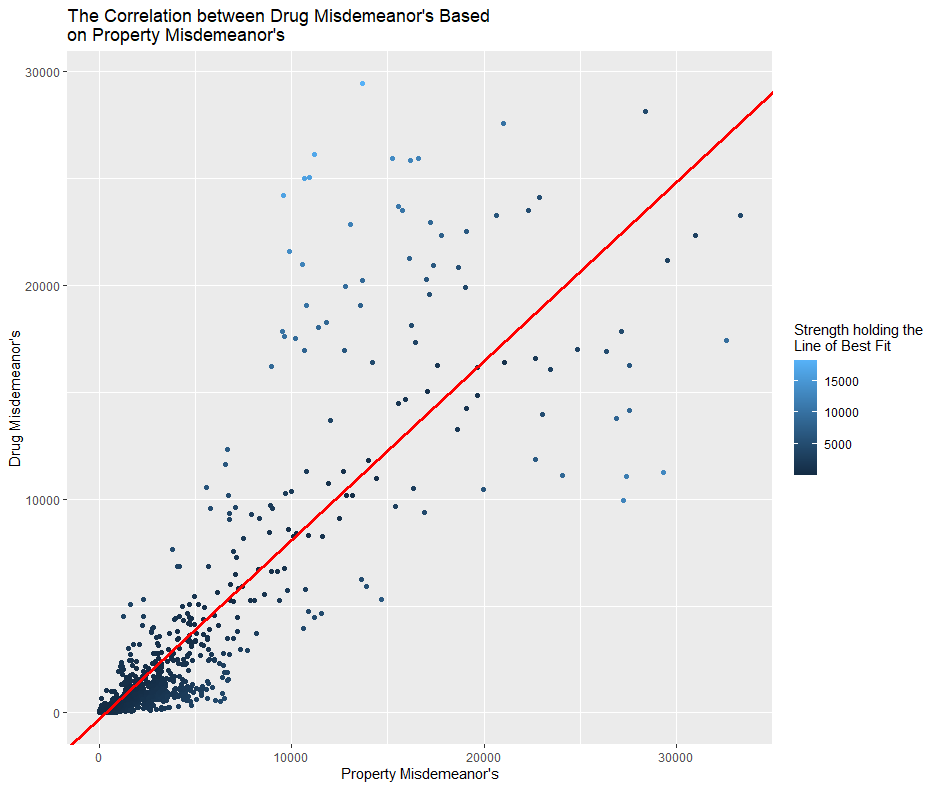
y = 'Violent Crime Counts',

color = 'Political Party')

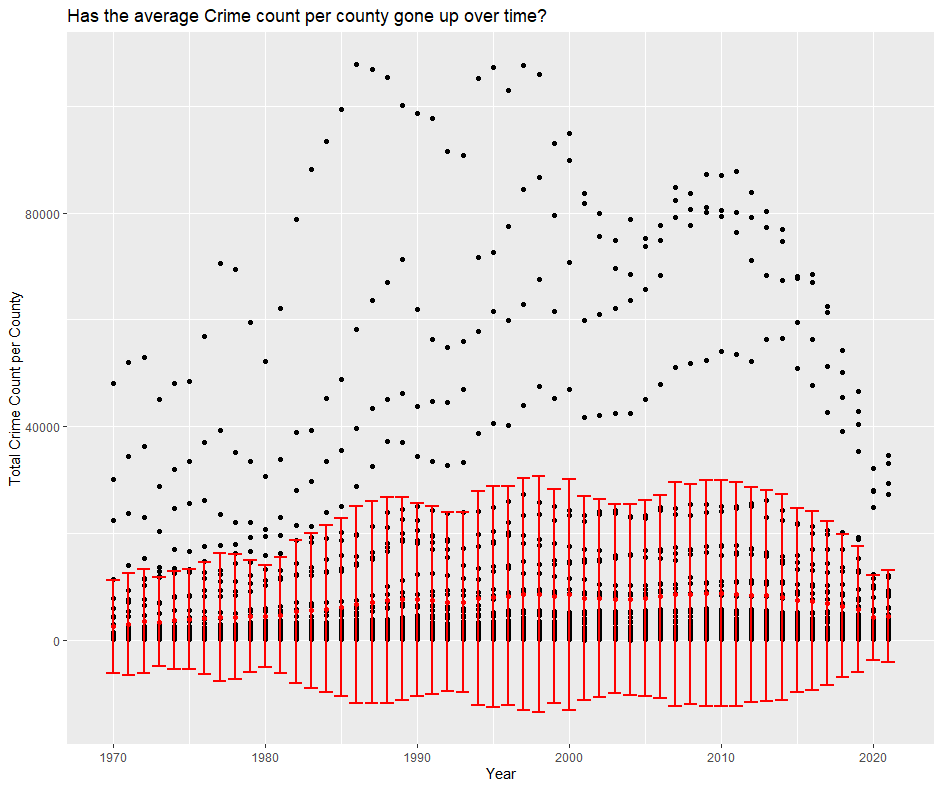
SECTION 2:

Graph 7:A picture containing text, screenshot, diagram

Description automatically generatedA picture containing text, screenshot, diagram, plot

Description automatically generated

I grabbed two different types of crimes and saw that as one goes up the other appears to go up also. Mainly because when I personally think of drugged up people you think of the vandalism they do (thank you YouTube). Thought it was cool so did a line of best fit on a linear model. Though when we start to get far away there doesn’t seem to be a good correlation but seems to follow this line well though it may not be best far. The residuals look fair with close to the h=0 line. They appear to have just as many crazy outliers and aren’t fitting any other model. Though they aren’t tight. I made a histogram with the Standard Deviation of Property Misdemeanor filtering out by 3 and bins is Standard deviation / 5 and wow! It appears they are close to the center!

Graph 8: 

I didn’t love this graph at first because I thought it was just saying nothing, but I think after reviewing the prior ones it says one. Though we see these crazy numbers for some years… There is no difference in crimes committed averages with 95% confidence! Isn’t it amazing to know that New York is staying the same with crime as time moves on as an average? This could be that some counties are getting better as some get worse though so take it with a grain of salt and remember population difference can skew this data.

CODE:

arrestsbefore%>%

group\_by(Year)%>%

mutate(avgcount\_of\_crimes = mean(`Felony Total`+`Misdemeanor Total`))%>%

mutate(sdofcount\_of\_crimes = sd(`Felony Total`+`Misdemeanor Total`))%>%

ungroup()%>%

mutate(upper = avgcount\_of\_crimes + qnorm(p=.975)\*sdofcount\_of\_crimes/sqrt(3))%>%

mutate(lower = avgcount\_of\_crimes + qnorm(p = .025)\*sdofcount\_of\_crimes/sqrt(3))%>%

ggplot()+

geom\_point(aes(Year,`Felony Total`+`Misdemeanor Total`))+

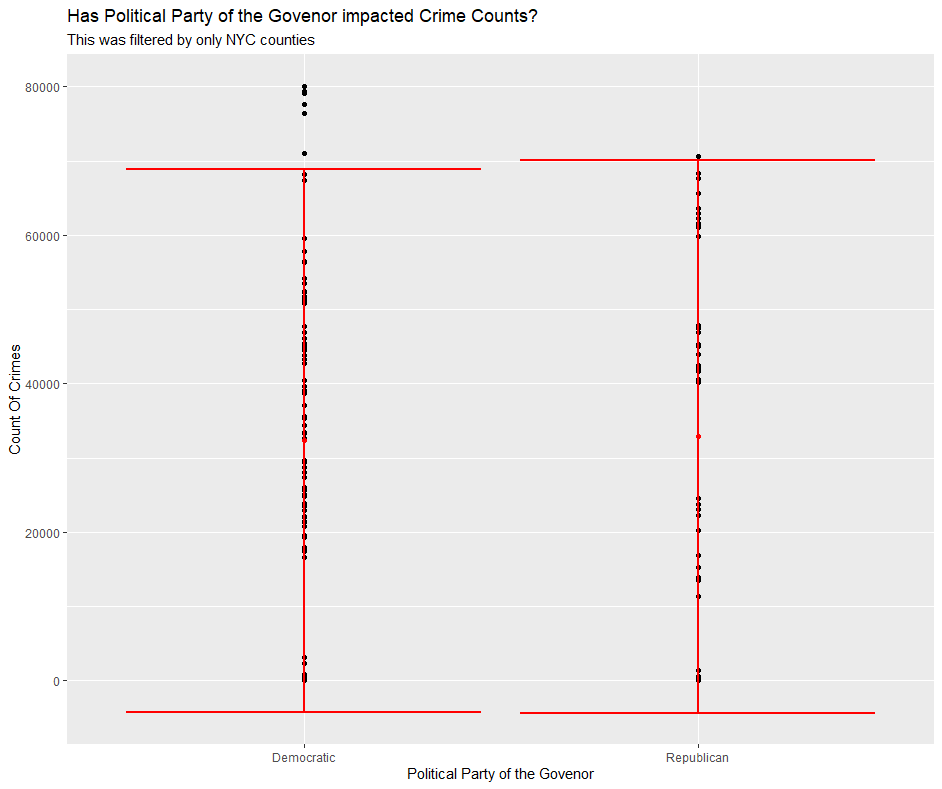
geom\_point(aes(Year,avgcount\_of\_crimes),color = 'red',size = 1.5)+

geom\_errorbar(aes(x = Year , y = avgcount\_of\_crimes , ymin = lower , ymax = upper),color = 'red',size = 1)+

labs(title = 'Has the average Crime count per county gone up over time?',

x = 'Year',

y = 'Total Crime Count per County')

Graph 8 Replacement:A graph of a political party

Description automatically generated with low confidence

I always here that one Political Party says they are better with crime then the other, so I wanted to test this… There appears to be no difference with crime over time with 95% confidence in NYC and New York State as a whole. I thought this was amazing to note. Though this doesn’t mean that each person isn’t better than the other person running but this says that the Party doesn’t mean anything.

CODE:

arrestswithtotaljoin%>%

group\_by(Party)%>%

mutate(avg\_tot\_count = mean(Total))%>%

mutate(sd\_tot\_count = mean(Total))%>%

ungroup()%>%

mutate(upper = avg\_tot\_count + qnorm(p=.975)\*sd\_tot\_count/sqrt(3))%>%

mutate(lower = avg\_tot\_count + qnorm(p = .025)\*sd\_tot\_count/sqrt(3))%>%

ggplot()+

geom\_point(aes(Party,Total))+

geom\_point(aes(Party,avg\_tot\_count),color = 'red',size = 1.5)+

geom\_errorbar(aes(x = Party , y = avg\_tot\_count , ymin = lower , ymax = upper),color = 'red',size = 1)+

labs(title = 'Has Political Party of the Govenor impacted Crime Counts?',

x = "Political Party of the Govenor",

y = "Count Of Crimes")

##Graph8.R.2

arrestswithtotaljoin%>%

filter(County == c("Bronx","Unknown NYC county","Queens"))%>%

group\_by(Party)%>%

mutate(avg\_tot\_count = mean(Total))%>%

mutate(sd\_tot\_count = mean(Total))%>%

ungroup()%>%

mutate(upper = avg\_tot\_count + qnorm(p=.975)\*sd\_tot\_count/sqrt(3))%>%

mutate(lower = avg\_tot\_count + qnorm(p = .025)\*sd\_tot\_count/sqrt(3))%>%

ggplot()+

geom\_point(aes(Party,Total))+

geom\_point(aes(Party,avg\_tot\_count),color = 'red',size = 1.5)+

geom\_errorbar(aes(x = Party , y = avg\_tot\_count , ymin = lower , ymax = upper),color = 'red',size = 1)+

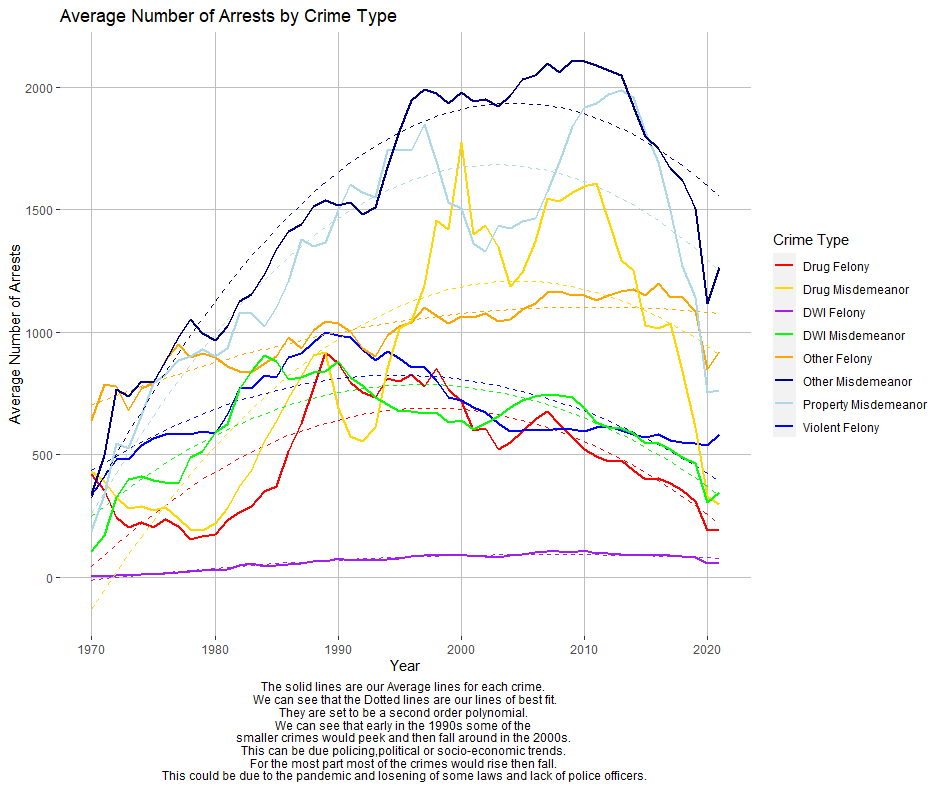
labs(title = 'Has Political Party of the Govenor impacted Crime Counts?',

subtitle = "This was filtered by only NYC counties",

x = "Political Party of the Govenor",

y = "Count Of Crimes")

SECTION 3:

Graph 9:

I don’t know much to say that isn’t seeable by the first look through captions or the title. I used the second order Polynomials as I thought that would represent the data best. I would show the residual plot, but I think it would be rough to have this be 8 pages longer with those graphs. I also used the second order because it fit some of them well and didn’t want to go to a different degree for some and not for others. Also, a residual plot for a line Is kind of weird.

CODE:

##MAKING AVERAGES FOR EACH CRIME

arrestsbefore\_new <- arrestsbefore %>%

group\_by(Year) %>%

mutate(avgdrug\_fel = mean(`Drug Felony`)) %>%

mutate(avgviolent\_fel = mean(`Violent Felony`)) %>%

mutate(avgdwi\_fel = mean(`DWI Felony`)) %>%

mutate(avgother\_fel = mean(`Other Felony`)) %>%

mutate(avgdrug\_mis = mean(`Drug Misdemeanor`)) %>%

mutate(avgdwi\_mis = mean(`DWI Misdemeanor`)) %>%

mutate(avgprop\_mis = mean(`Property Misdemeanor`)) %>%

mutate(avgother\_mis = mean(`Other Misdemeanor`)) %>%

ungroup()

##MAKE THE LINE OF BEST FITS

avgdrug\_fellbf<-lm(avgdrug\_fel~I(Year^2)+Year,data = arrestsbefore\_new)

avgviolent\_fellbf<-lm(avgviolent\_fel~I(Year^2)+Year,data = arrestsbefore\_new)

avgdwi\_fellbf<-lm(avgdwi\_fel~I(Year^2)+Year,data = arrestsbefore\_new)

avgother\_fellbf<-lm(avgother\_fel~I(Year^2)+Year,data = arrestsbefore\_new)

avgdrug\_mislbf<-lm(avgdrug\_mis~I(Year^2)+Year,data = arrestsbefore\_new)

avgdwi\_mislbf<-lm(avgdwi\_mis~I(Year^2)+Year,data = arrestsbefore\_new)

avgprop\_mislbf<-lm(avgprop\_mis~I(Year^2)+Year,data = arrestsbefore\_new)

avgother\_mislbf<-lm(avgother\_mis~I(Year^2)+Year,data = arrestsbefore\_new)

##MAKE IT A COLUMN

arrestsbfe <- arrestsbefore\_new %>%

add\_predictions(avgdrug\_fellbf) %>%

mutate(avgdrug\_fellbfcol = pred) %>%

add\_predictions(avgviolent\_fellbf) %>%

mutate(avgviolent\_fellbfcol = pred)%>%

add\_predictions(avgdwi\_fellbf)%>%

mutate(avgdwi\_fellbfcol = pred)%>%

add\_predictions(avgother\_fellbf)%>%

mutate(avgother\_fellbfcol = pred)%>%

add\_predictions(avgdrug\_mislbf)%>%

mutate(avgdrug\_mislbfcol = pred)%>%

add\_predictions(avgdwi\_mislbf)%>%

mutate(avgdwi\_mislbfcol = pred)%>%

add\_predictions(avgprop\_mislbf)%>%

mutate(avgprop\_mislbfcol = pred)%>%

add\_predictions(avgother\_mislbf)%>%

mutate(avgother\_mislbfcol = pred)

##GRAPH

arrestsbfe%>%

ggplot() +

geom\_line(aes(Year, avgdrug\_fel, color = "Drug Felony"),size = 1) +

geom\_line(aes(Year, avgdrug\_fellbfcol,color = "Drug Felony"),linetype = "dashed",size = .5)+

geom\_line(aes(Year, avgviolent\_fel, color = "Violent Felony"),size = 1) +

geom\_line(aes(Year, avgviolent\_fellbfcol,color = "Violent Felony"),linetype = "dashed",size = .5)+

geom\_line(aes(Year, avgdwi\_fel, color = "DWI Felony"),size = 1) +

geom\_line(aes(Year, avgdwi\_fellbfcol,color = "DWI Felony"),linetype = "dashed",size = .5)+

geom\_line(aes(Year, avgother\_fel, color = "Other Felony"),size = 1) +

geom\_line(aes(Year, avgother\_fellbfcol,color = "Other Felony"),linetype = "dashed",size = .5)+

geom\_line(aes(Year, avgdrug\_mis, color = "Drug Misdemeanor"),size = 1) +

geom\_line(aes(Year, avgdrug\_mislbfcol,color = "Drug Misdemeanor"),linetype = "dashed",size = .5)+

geom\_line(aes(Year, avgdwi\_mis, color = "DWI Misdemeanor"),size = 1) +

geom\_line(aes(Year, avgdwi\_mislbfcol,color = "DWI Misdemeanor"),linetype = "dashed",size = .5)+

geom\_line(aes(Year, avgprop\_mis, color = "Property Misdemeanor"),size = 1) +

geom\_line(aes(Year, avgprop\_mislbfcol,color = "Property Misdemeanor"),linetype = "dashed",size = .5)+

geom\_line(aes(Year, avgother\_mis, color = "Other Misdemeanor"),size = 1) +

geom\_line(aes(Year, avgother\_mislbfcol,color = "Other Misdemeanor"),linetype = "dashed",size = .5)+

labs(title = "Average Number of Arrests by Crime Type",

x = "Year",

y = "Average Number of Arrests",

color = "Crime Type",

caption = "The solid lines are our Average lines for each crime. \nWe can see that the Dotted lines are our lines of best fit.\nThey are set to be a second order polynomial. \nWe can see that early in the 1990s some of the \nsmaller crimes would peek and then fall around in the 2000s. \nThis can be due policing,political or socio-economic trends. \nFor the most part most of the crimes would rise then fall. \nThis could be due to the pandemic and losening of some laws and lack of police officers.") +

scale\_color\_manual(name = "Crime Type",

values = c(

"Drug Felony" = "red",

"Violent Felony" = "blue",

"DWI Felony" = "purple",

"Other Felony" = "orange",

"Drug Misdemeanor" = "gold",

"DWI Misdemeanor" = "green",

"Property Misdemeanor" = "lightblue",

"Other Misdemeanor" = "navy"))+

theme(plot.caption = element\_text(hjust = 0.5),

panel.background = element\_rect(fill = "White"),

panel.grid.major = element\_line(color = 'grey'))

The Governor dataset was made in excel from the information I gave from the website. Most of the Data was the crimes data which was taken from Data.gov.