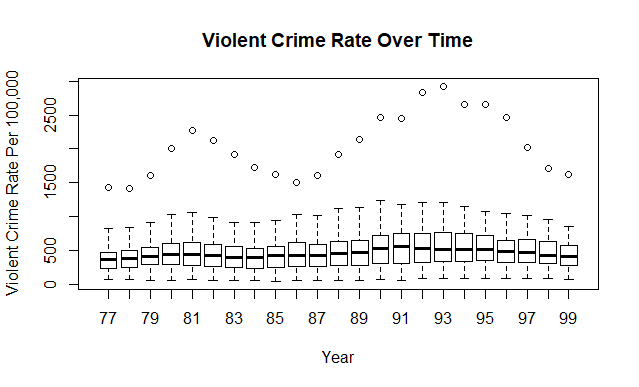
**Guns Project**

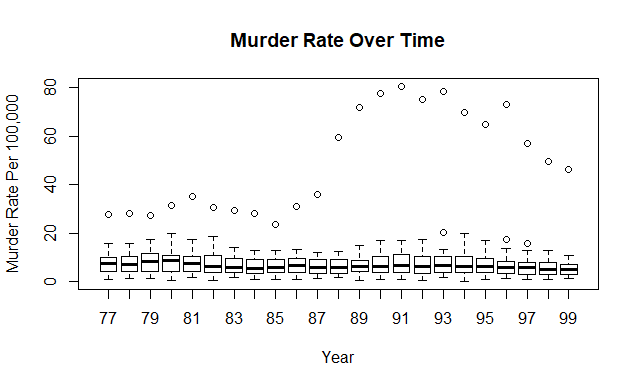
I begin by looking for outlier in each data set when comparing crime rates and incarceration rates over time:

**Violent Crime Rate:**



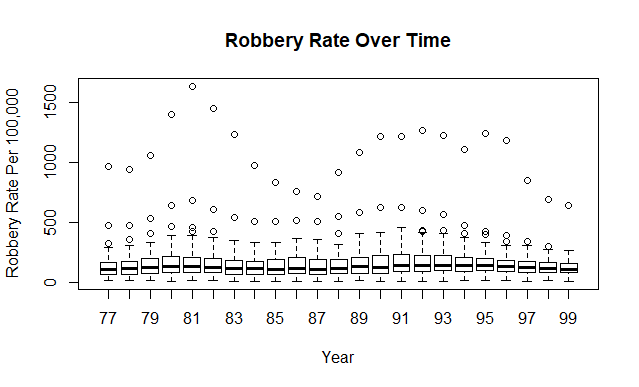
There seems to be 1 significant outlier in each year for the violent crime rate.

**Murder Rate:**



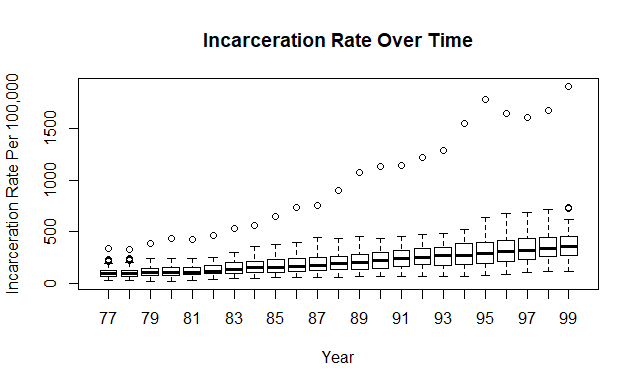
Once again there seems to be 1 significant outlier for each year, with some like 93, 96, and 97 having another outlier that looks to still be pretty close.

**Robbery Rate:**



There are a lot more outliers for robbery rate than other crime rates, with 2-3 outliers within each year.

**Incarc Rate:**



There is one or two outliers in each year.

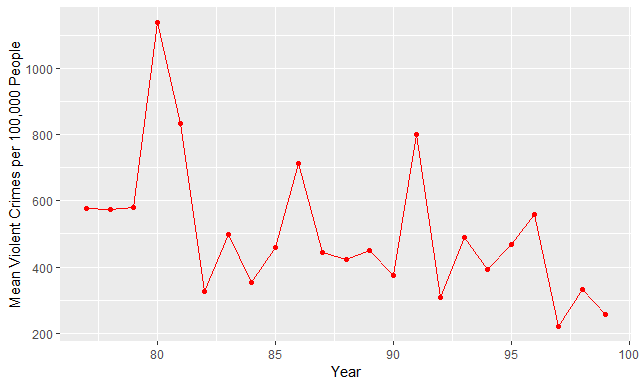
I want to look at the trends of these rates over time. Since each year has multiple values for crime rates, I will need to look at the mean of the crime rates in each year. If I include the outliers when doing this, they will greatly skew the resulting plots.

I have included both plots with outliers, and plots where I use a capping method to replace any value outside the quartile of that year with the 75th percentile value in order to get rid of outliers.

**Trends over Time (Outliers Included):**

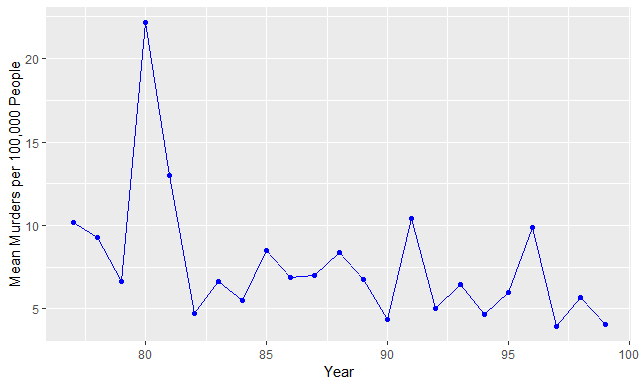
I begin by looking at trends of the variables murder rate, violent crime rate, robbery rate, and incarceration rate over time. Because each of these had multiple values per year, I wanted to take the mean of each year and plot it over time to see general trends:

**Violent Crime Rate:**



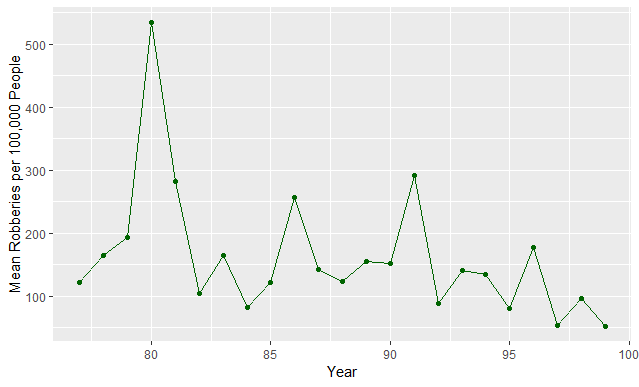
As shown above the average violent crime rate decreased significantly over time

**Murder Rate:**



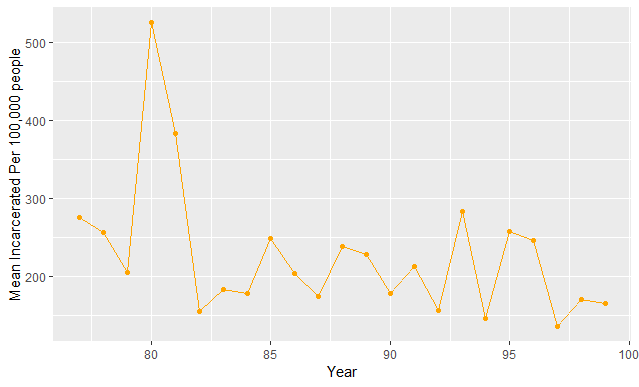
Similarly to violent crime rate, murder rate decreased substantially after 1980, and stayed somewhat consistent before falling again in 1999.

**Robbery Rate:**



Like the other variables, robbery rates fluctuated but tended to decrease over time, especially towards 1999.

**Incarceration Rate:**

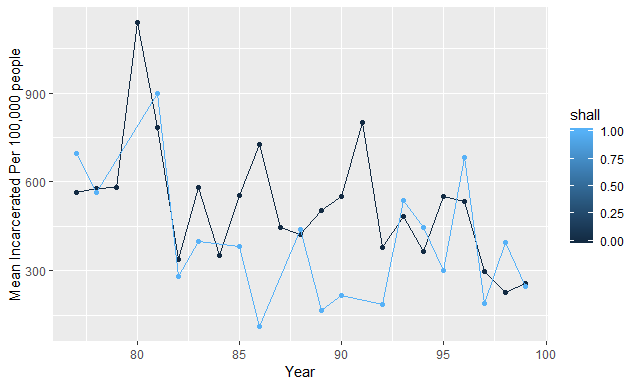


Incarceration rate again decreases over time.

The commonality between all 4 is that by the late 1990’s, the rates had all been lowered significantly from where they where in the late 70s and early 80s.

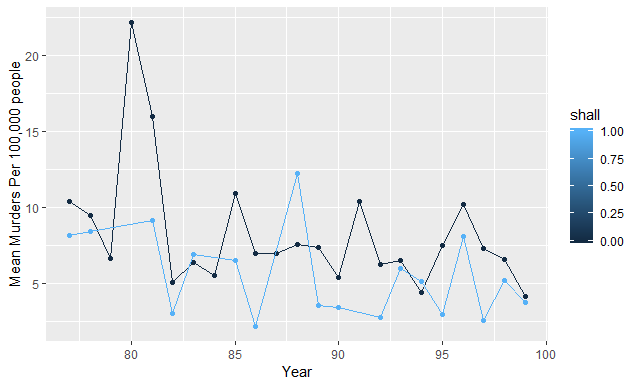
**Separated By Shall Carry Law/Non Shall States**

I then re-did the process above except separated the data between those in state’s with shall laws (shall issue concealed carry licenses if person meets legal requirements) and those without:



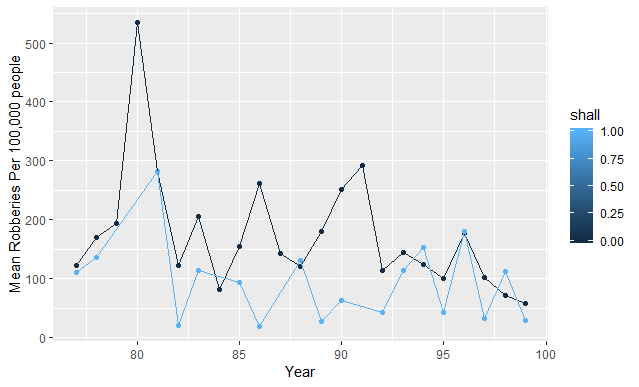
Initially when shall = 1 the average violent crime rate is lower than when shall = 0, however after 1995 the violent crime rate seems to be higher when shall = 1, until approximately 1999 when there seems to be no difference between the two.

**Murder Rate:**



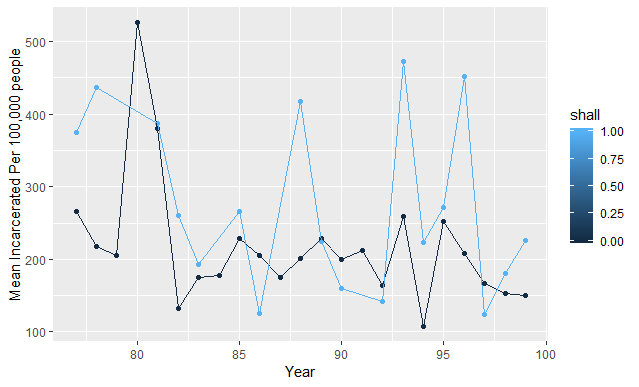
Other than in 1988, the average murder rate seems to be smaller when shall = 1 than when shall = 0.

**Robbery Rate**



When shall = 1, the average robbery rate is lower (other than a few years where the two groups converge) than when shall = 0.

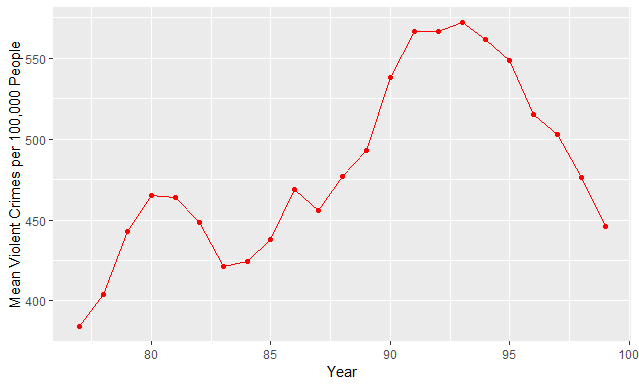
**Incarceration Rate**



After around 1983, the average incarceration rate when shall = 1 is much much higher than when shall = 0. This could possibly be a good explanation why the average crime rates when shall = 1 seem to be lower (it may be better attributed to a higher incarceration rate).

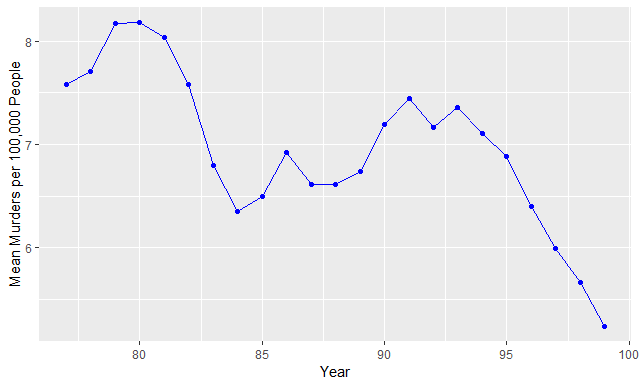
**Trends Over Time (Without Outliers):**

**Violent Crime Rate:**



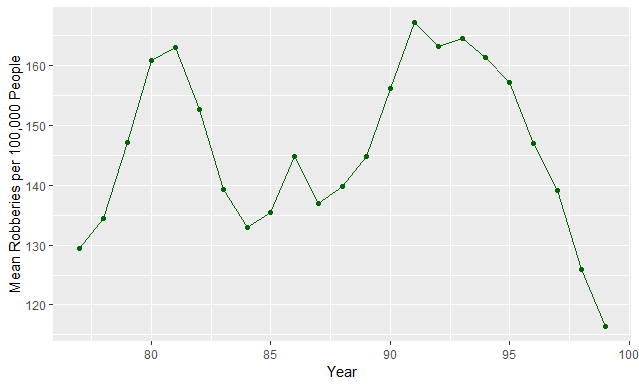
Interestingly this paints a complete opposite picture that violent crime actually increased up until 1993 when it decreased again

**Murder Rate**



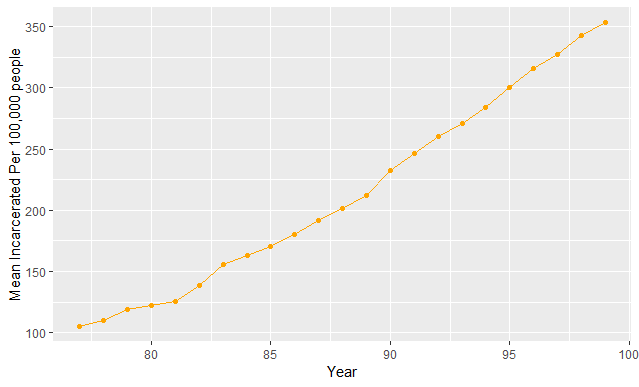
Murder rate seems to decrease significantly over time.

**Robbery Rate:**



Robbery Rate seems to have fluctuated over time. It started increasing through 1977 to 1981, decreased, increased slightly before decreasing again, increased significantly from 1987 to 1993 before falling drastically again.

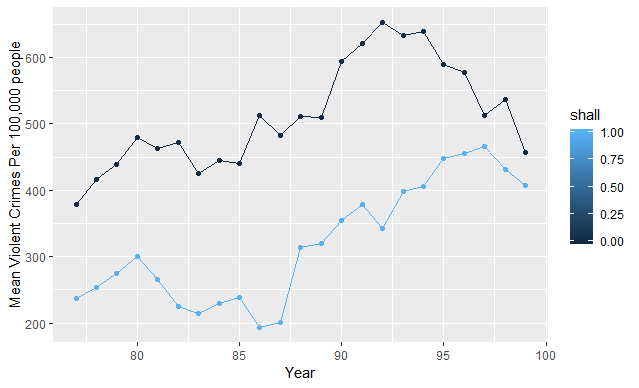
**Incarceration Rates**



Once again incarceration rates seem to increase exponentially over time.

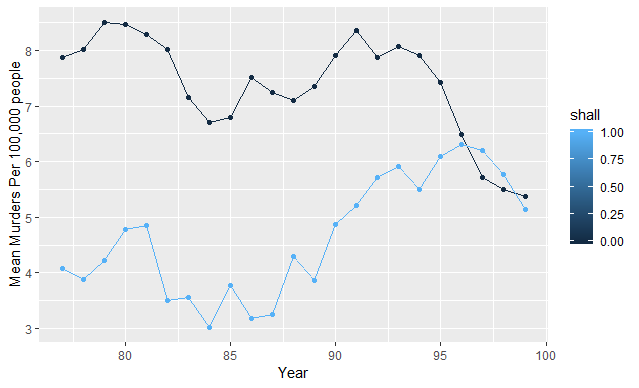
**Shall Comparisons**

**Violent Crime Rate:**



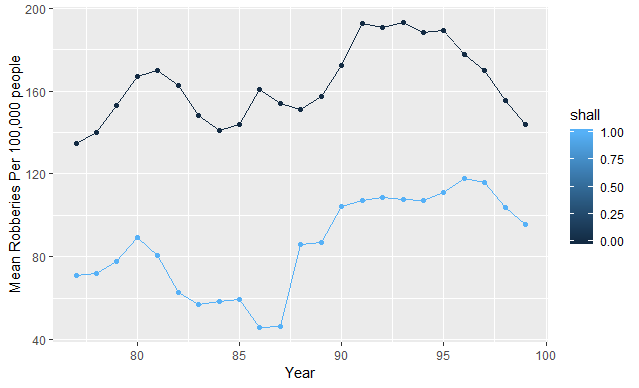
When Shall = 1 the mean violent crime rate tends to be much lower

**Murder Rate**



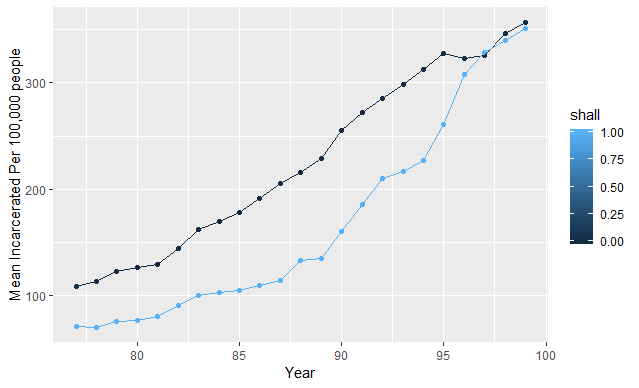
When Shall = 1 the murder rate was much lower UNTIL around 1996 when both states with shall laws and those without had similar murder rates.

**Robbery Rate**



Robbery rate seems to be much lower when Shall = 1

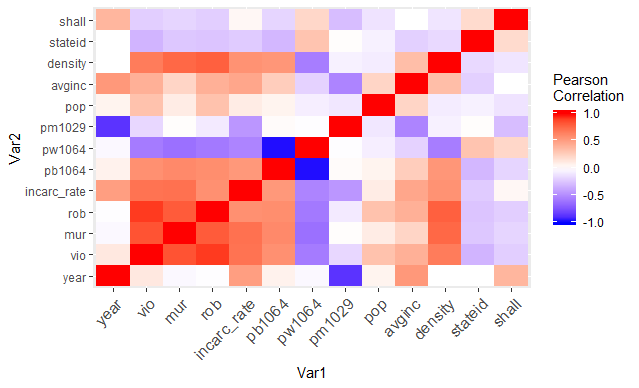
**Incarceration Rate:**



Incarceration Rate seems to contradict the outlier findings because incarc\_rate seems to be lower for Shall = 1 until 1997/1998 when the rates converge.

**Collinearity**

Before I get into the model estimation, I wanted to see the correlations between all of the variables to check if there would be collinearity issues. Below is a correlation heatmap showing the correlations between all of the variables:



Based on the heat map, there are a few correlations that we want to be careful of when estimating our models:

Rob-Vio: Correlation is very close to 1

Mur-Vio: Correlation also very close to 1

Mur-Rob: Correlation close to 1

Rob-Density: Correlation close to 1

Pw1064-pb1064: Perfect correlation of -1.00

Pm1029-year: Correlation very close to -1.00

Do a graph comparing the crime rates for states when shall = 0 and change to shall = 1 to see if all of the states react to the shall law the same way.

**Modeling**

Based on our current theory and the data visualizations we came up with, we came up with the following variables to be included in the model:

Incarc\_rate: Wanted to control for it since it could have a real effect on crime rates

Avginc: We have seen that this could potentially have an impact on crime rates and want to control for it

Density: Same as above, there is strong correlation between density and crime rate so we want to control for it.

Shall: What we are trying to test for.

We did not include race variables because they are perfectly collinear and not that correlated with crime rate.

**Fixed Effects, Random Effects, Or Pooled OLS?**

*The Hausman Test*

The first thing we do is run two models, one fixed effects and one random effects using the variables above and ran with each crime rate variable as the dependent variable. The following is the results of the Hausman test for each of these model pairs:

Hausman Test

data: mur ~ incarc\_rate + avginc + density + shall

chisq = 314.31, df = 4, p-value < 2.2e-16

alternative hypothesis: one model is inconsistent

Hausman Test

data: vio ~ incarc\_rate + avginc + density + shall

chisq = 15.005, df = 4, p-value = 0.004691

alternative hypothesis: one model is inconsistent

Hausman Test

data: rob ~ incarc\_rate + avginc + density + shall

chisq = 3.3634, df = 4, p-value = 0.4989

alternative hypothesis: one model is inconsistent

The Hausman test has a null hypothesis is that the regressors are not correlated with the errors (where random effects models would be appropriate). With the murder and violent crime rate models, the p-value is less than .05 so we would reject the null hypothesis that the regressors are not correlated with the errors, meaning we would need to use the fixed models. For robbery rate, however, the p-value is much higher than .05 which means we would not reject the null hypothesis that the regressors are not correlated with the errors, meaning we would use the random effects model.

Note: We would not want to use Pooled OLS because we are taking the same samples in every year.