Gun Trend Predictions: Exploratory Data Analysis

I. Provision Efficiency Hypothesis Testing

Single Provisions

Under the “Analyzing the Effects of Provisions” portion of my project, I attempt to find gun control provisions that correlate most with reduced gun violence. I examine the differences in gun violence for states that have a certain gun control provision in effect compared with states that don’t have that provision for a given year.
In order to control for sample size and outliers, I used a minimum sample size threshold and discarded provisions where the number of samples in either group was below the threshold. I created bar chart visualizations, but I wanted to conduct a t-test to confirm the differences seen in the bar chart.

The chart below shows the p-values:

Provision	Gun Violence Rates in States Without Provision	Gun Violence Rates in States With Provision	Difference	P-value
nosyg	6.024715	3.466635	2.558080	0.000893
immunity	5.629389	3.030228	2.599161	0.001455
violentpartial	5.359422	3.441462	1.917959	0.024187
statechecksh	5.336470	3.490234	1.846237	0.030354
mcdvdating	5.349055	3.574407	1.774649	0.034754
cap14	5.286320	3.596802	1.689518	0.048442

Here’s a brief description of each provision codename:

nosyg - Use of deadly force is not allowed to be a first resort in public. This is sometimes referred to as a "stand your ground" law.
immunity - No law provides blanket immunity to gun manufacturers or prohibits state or local lawsuits against gun manufacturers.
violentpartial - Firearm possession is prohibited for people who have committed a violent misdemeanor punishable by more than one year of imprisonment.
mcdvdating - All people convicted of a misdemeanor crime of domestic violence are prohibited from possessing firearms.
statechecksh - State conducts separate background checks, beyond the federal NICS check, for handguns.
cap14 - Criminal liability for negligent storage applies to access by children less than 14 years old.

For a better explanation of what each provision means, the codenames can be looked up in the ./data/raw/codebook.xlsx

Pairs of Provisions

I also wanted to see whether a pair of provisions would have a significant correlation with reduced gun violence.

Once again, a t-test is used to get a p-value for each of the differences:

Provision	Gun Violence Rates in States Without Provision	Gun Violence Rates in States With Provision	Difference	P-value
age21handgunsale permith	5.235497	2.698168	2.537328	0.015916
statechecks universalpermith	5.330041	2.913307	2.416734	0.021859
statechecks universalpermit	5.187620	2.913307	2.274313	0.022568
age21handgunsale universalpermith	5.336773	2.982114	2.354659	0.026356
age21handgunsale mcdv	5.443077	3.293261	2.149815	0.028332

Here are the descriptions of the provisions above:
statechecks - State conducts separate background checks, beyond the federal NICS check, for all firearm.
universalpermith - Background checks conducted through permit requirement for all handgun sales (or universal background checks).
universalpermit - Background checks conducted through permit requirement for all firearm sales (or universal background checks).
age21handgunsale - Purchase of handguns from licensed dealers and private sellers restricted to age 21 and older.
mcdv - People convicted of a misdemeanor crime of domestic violence against a spouse, ex-spouse, or cohabitating partner are prohibited from possessing firearms.

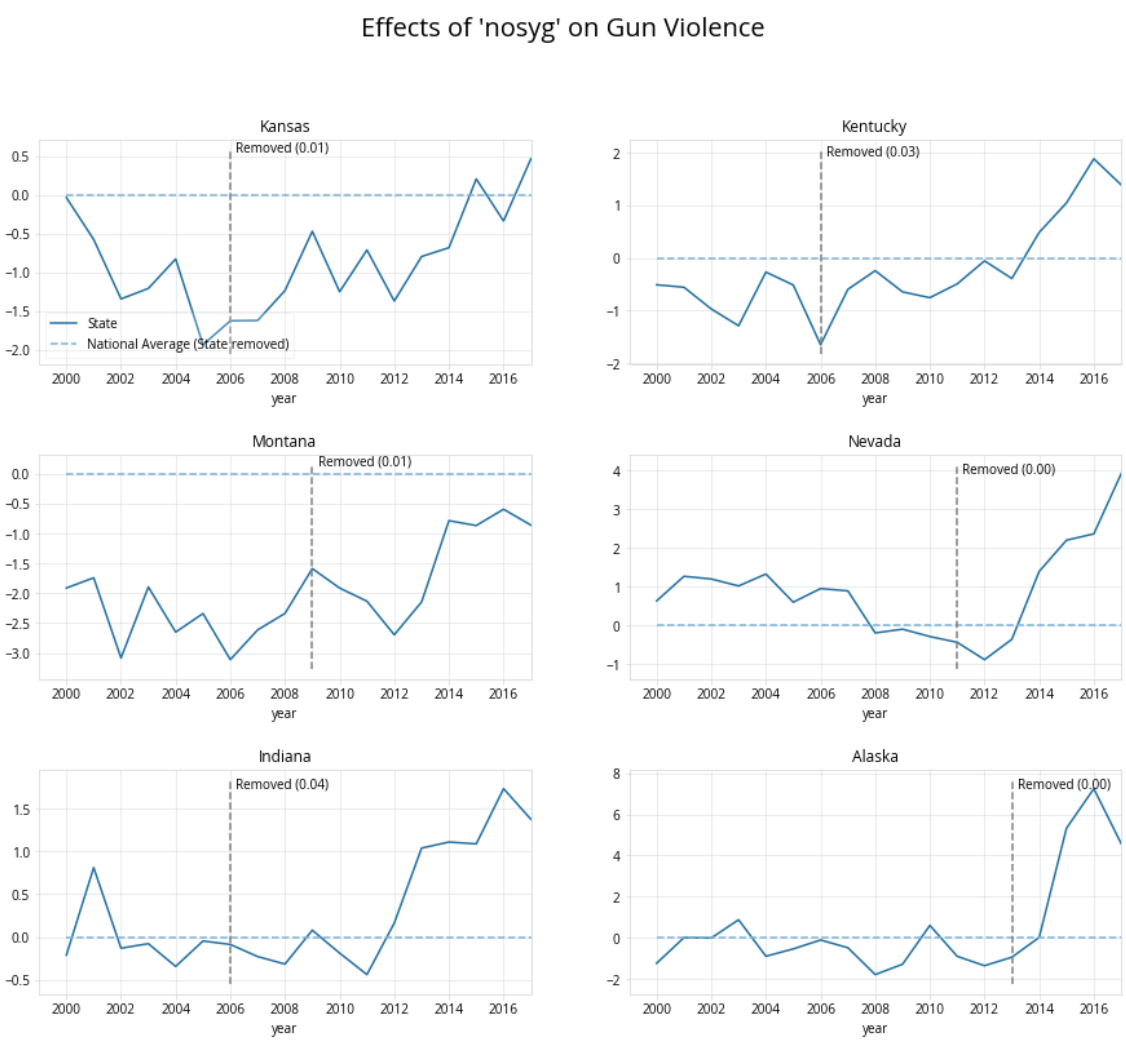
II. Testing for Structural Breaks

We looked at when provisions were either added or removed by a state to see if we could see any differences in gun violence after those changes were made. In order to quantify this change, we did a Chow test for structural break, an f-statistic defined as:

$$\frac{((RSS_{pooled} - (RSS_{before} + RSS_{after})) / k)}{(RSS_{before}+RSS_{after})/(N_{before} + N_{after} - 2k)}$$

where:
RSS_{pooled} - The residual sum of squares from regression on the entire time series
RSS_{before} - The residual sum of squares from regression on the time series before the structural break
RSS_{after} - The residual sum of squares from regression on the time series after the structural break
N_{before} - Number of samples in the time series before the structural break
N_{after} - Number of samples in the series after the structural break
k - Number of parameters estimated during the regression

To visualize this test, each time series is shown with a grey dotted line which indicates a change in provision. Next to it in parentheses is the p-value of the Chow test at that point. For conciceness, only one examples of these tests are shown (more can be found in visualization.ipynb):



This test gives us some confidence in making judgements about observed changes in gun violence rates after a provision is changed. However, there is a limited sample size for the Chow test, and a large number of factors that could affect gun violence. The structural break test results are to be interpreted with this in mind.