To work on this, I have used an open source data ‘crime\_rates.csv’ and ‘canada.csv’ files from the GitHub (https://github.com/data-8/data8assets/blob/gh-pages/materials/fa16/project/project2/project2.ipynb). Attached is the Python code generated in Jupyter, this code produces the respective tasks mentioned in the ‘Tasks.pdf’ file.

**SECTION 1: MURDER RATES.**

As per the task requirement, I have read the CSV file and considered the 'State', 'Year', 'Population' and 'Murder Rate' columns.

**QUESTION 1.1:**

As we were asked to check the association between the existence of the death penalty in a state in a year and the murder rate in that state and year, following information would help for better analysis.

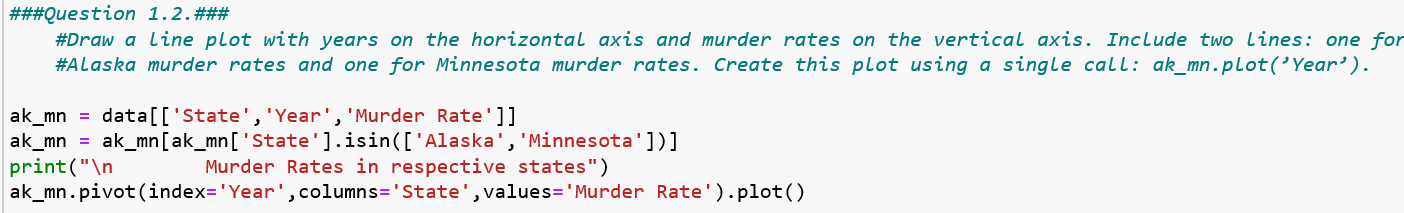
* Information on death penalty: csv file, doesn’t have any information on death penalty.

1. We need the year when the death penalty was introduced
2. We also need the year when the death penalty was abolished.
3. We also need the year when the death penalty was re-introduced.
4. Is the death penalty rule (introduced year/ abolished year) same for all the states in US? If any exemptions or outlawed death penalty to specific areas, we need the list of those states. (as mentioned in 3.2, -Alaska, Hawaii, Maine, Michigan, Wisconsin, and Minnesota)

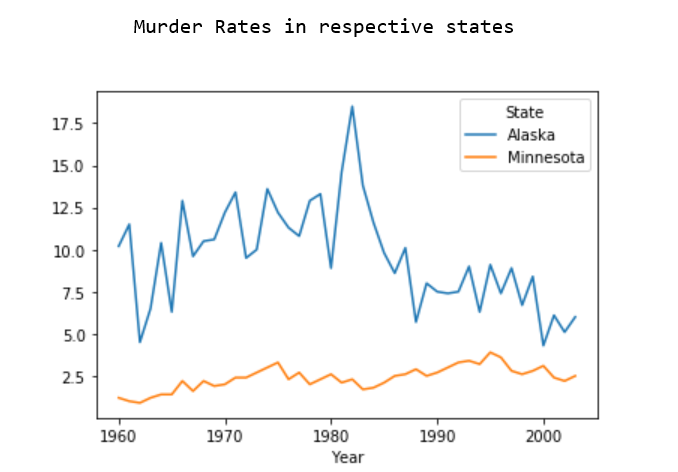
* Several other factors (detailed information given in 3.2) may also influence the murder rates, like, location/place/area, season/time. If we are provided with the external factors which also influence the murder rates, we will be able to control these and can find the cause of death penalty over murder rates.

**QUESTION 1.2:**

As stated, a line plot has been drawn using a single call for Alaska and Minnesota states, with “Murder Rates” on the Y-axis and “Year” on the X-axis.



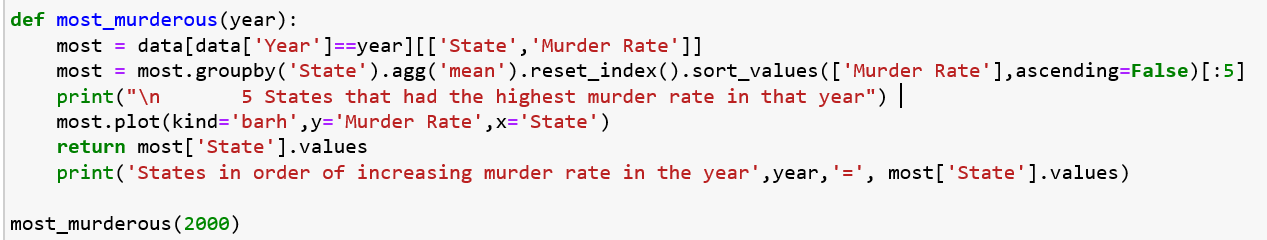
Output:



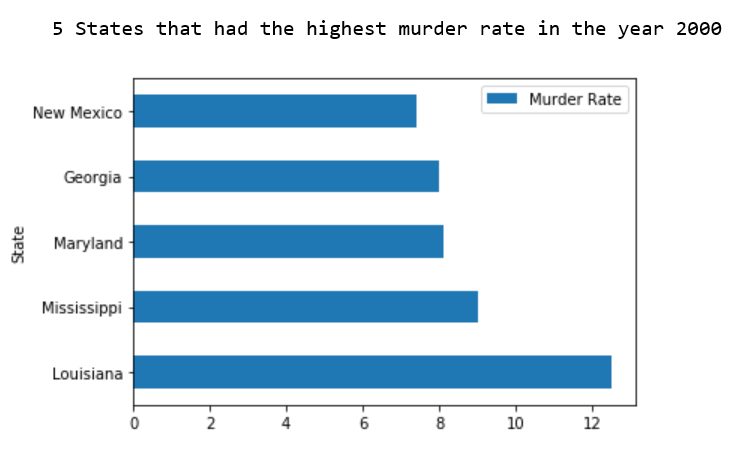
**QUESTION 1.3:**

As stated, “most\_murderous” function with year as its argument has been created to create a

1. horizontal bar chart the 5 states that had the highest murder rate in that year
2. It also returns an array of the names of these states in order of increasing murder rate.



Output #1



Output #2:

Using return option

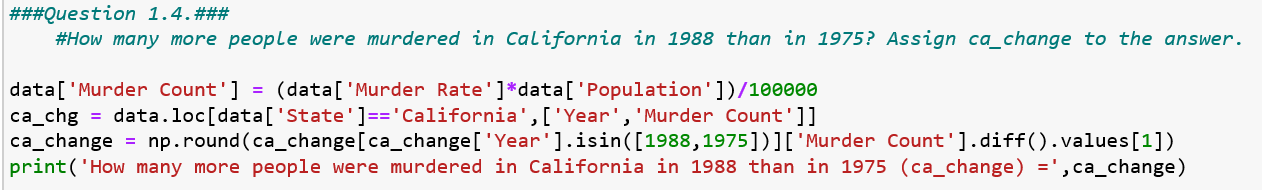


Or using print option, below is the output



**QUESTION 1.4:**

As per the question ca\_change has been calculated to show, how many more people were murdered in California in 1988 than in 1975.



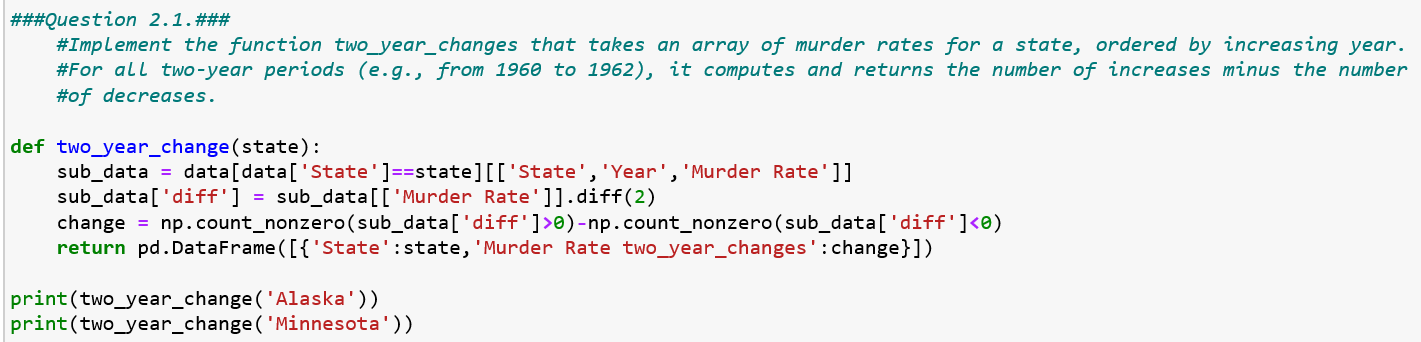
Output:



**SECTION 2: CHANGES IN MURDER RATES**

**QUESTION 2.1:**

As per the requirement, two-year changes that takes an array of murder rates for a state has been created. This function produces the “murder rate” in the respective states. Below are the outputs produced for two states. The “murder rate” values for all the 50 states will be produced in a dataset as per 2.2 question.



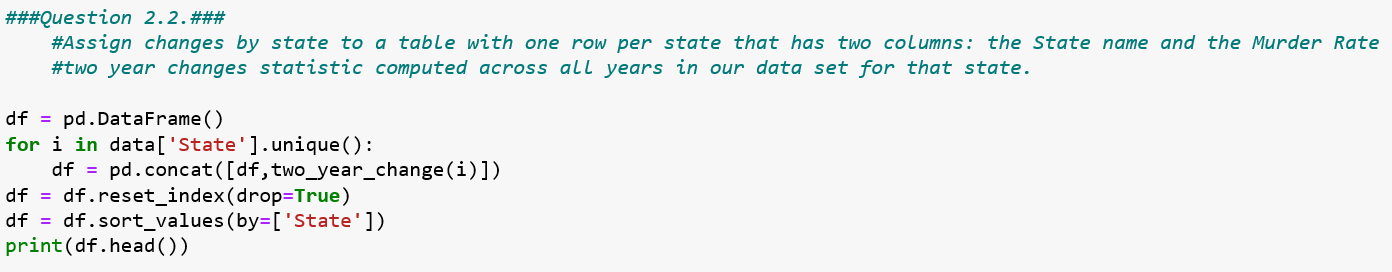
Outputs:



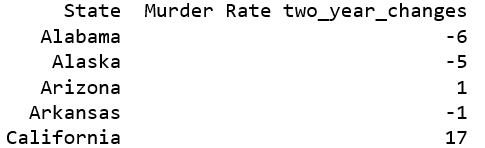


**QUESTION 2.2:**

As per the requirement, a table with one row per state that two columns i.e., State name and the Murder Rate two-year changes across all years has been created.

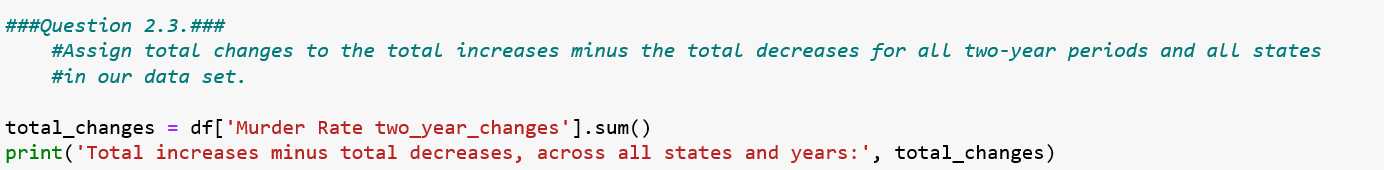


Output:



**QUESTION 2.3:**

As stated, ‘total\_changes’ has been created to show the “total increases minus the total decreases” for all two-year periods and all states in our data set.

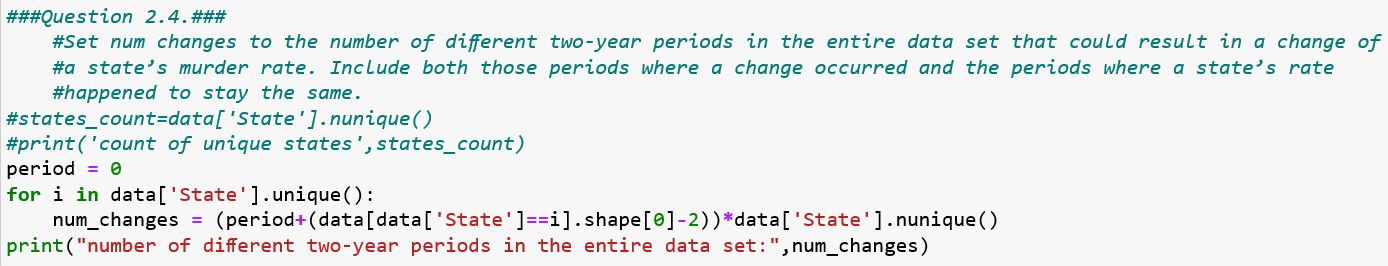


Output:



**QUESTION 2.4:**

“num\_changes” has been created to set the number of different two-year periods in the entire data set.



Output:



**QUESTION 2.5:**

Test Statistic: As larger positive values are in favour of alternative hypothesis and other values in favour of null hypothesis, out test statistic should be one which shows overall increase or decrease in murder rates. Hence, the best test statistic as per our requirements is “no.of increases minus no.of decreases”

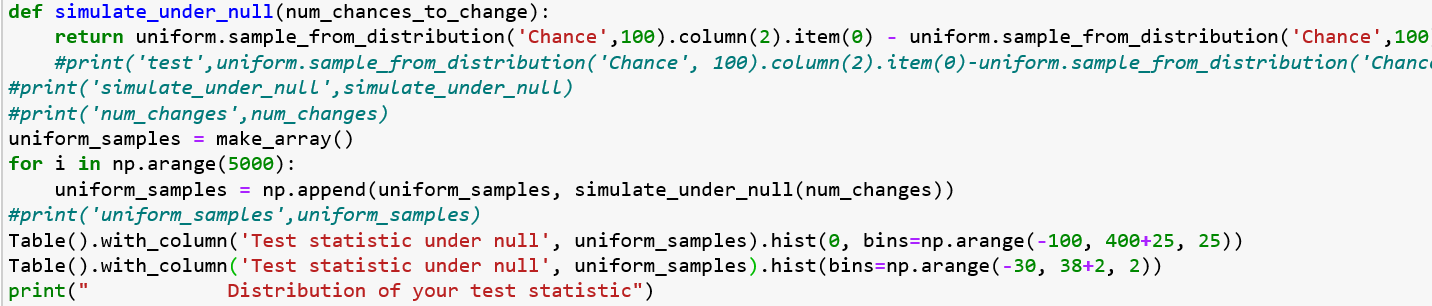
Test-statistic= No.of Increases – No.of Decreases.

**QUESTION 2.6:**

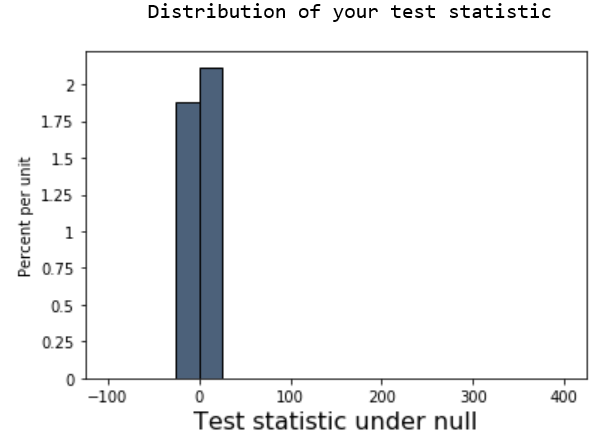
1. simulate\_under\_null function has been created and simulated a single sample under null hypothesis.
2. The argument for simulate\_under\_numm will be the num\_changes in the code below.

I have created two histograms using simulate\_under\_null

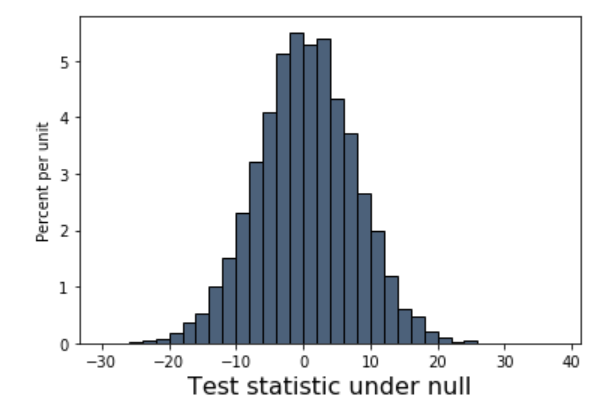
1. One has only two bars, the one on the left side represents the no.of decreases decrease and the one on the right side represents the no.of increases.
2. Other has a normal histogram which shows different bars for different test statistic values.



Output#1:



Output#2:



**QUESTION 2.7:**

From the above (output#1) histogram, the bar towards the left (i.e., on negative side) shows the no.of decreases in the dataset, whereas the bar towards the right (i.e., on positive side) shows the no.of increases in the dataset.

As the no.of decreases are less when compared to the no.of increases, if favours the alternative hypothesis and rejects the null hypothesis, i.e, It favours murder rates are likely to increase than decrease over two-year periods.

**SECTION 3: THE DEATH PENALITY.**

**QUESTION 3.1:**

To investigate the relationship between the death penalty and murder rates, we are taking the benefit of natural experiment. The natural experiment design consists of two different groups, a treatment and a control group. Based on the differences between these groups we can draw a relationship between the death penalty and murder rates.

1. Population: The population for our study is all the states (50) in the USA.

As we are analysing the murder rates before and after the death penalty was abolished, our control and treatment groups will be

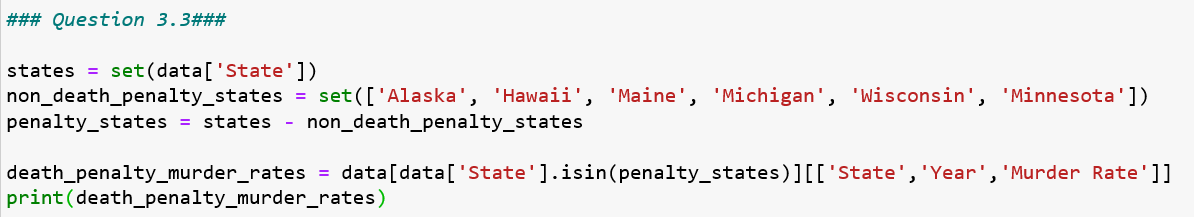
1. Control Group: The states without death penalty for the given years.
2. Treatment Group: The states with death penalty for the given years.
3. Outcome: Increase or decrease in the murder rates of a state during the given years.

**QUESTION 3.2:**

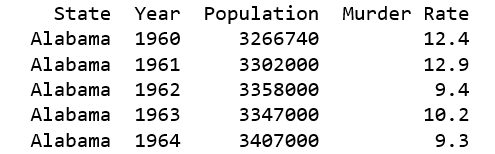
* In order to find out whether the death penalty causes a change in the murder rate, we would need some additional information as stated in 1.1.
* Also we cannot determine, if the death penalty is the major cause for the increase/decrease in the death rates as there are several other factors that may also influence the murder rates, like the area/ places (like rich/ poor, rural/ urban, areas with educated and uneducated people and areas/ places that doesn’t have security system etc.) and also the season/time.
* As mentioned in the above point, in specific areas murder rates can be higher/ lower across all states and all the time. Hence, comparing the murder rates of the states with and without death penalty may give the biased results.
* If we can able to control all other external factors that influence the murder rate (like mentioned in point 2 above), we may be able to decide if Death penalty causes a change in the murder rates.

**QUESTION 3.3:**

As stated, created a death\_penalty\_murder\_rates table to show the 'State', 'Year', 'Population' and 'Murder Rate' columns with only having the rows of the states those are having the death penalty.

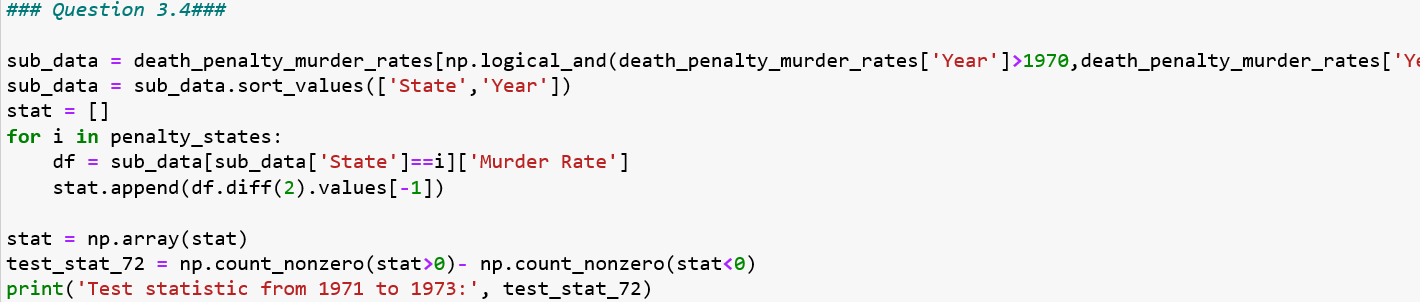


Output:



**QUESTION 3.4:**

As required, changes\_72 value has been calculated which represents value of the test statistic for the years 1971 to 1973.



Output:

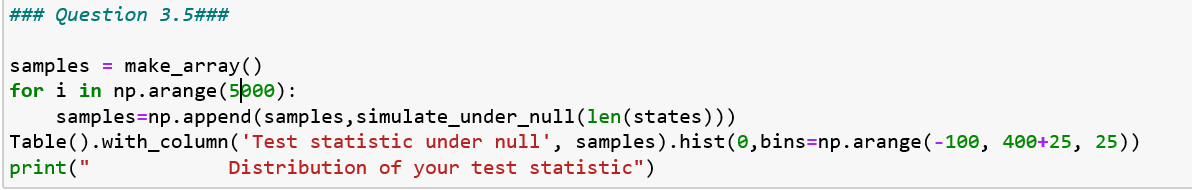


**QUESTION 3.5:**

An empirical histogram of the statistic has been plotted under the null hypothesis by simulating the test statistic 5,000 times.

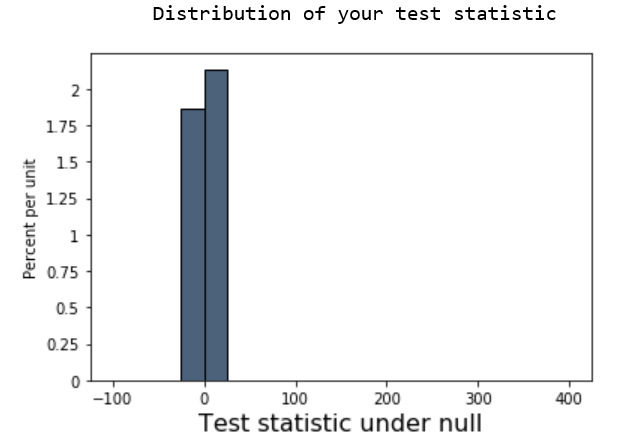
I have created two histograms using simulate\_under\_null

1. One has only two bars, the one on the left side represents the no.of decreases decrease and the one on the right side represents the no.of increases.
2. Other has a normal histogram which shows different bars for different test statistic values.

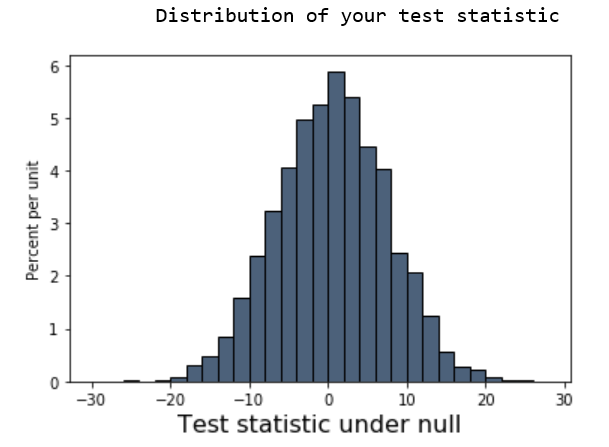




Output#1:



Output#2:



**QUESTION 3.6:**

1. Compute a P-value: P-value has been calculated as follows.



Output:

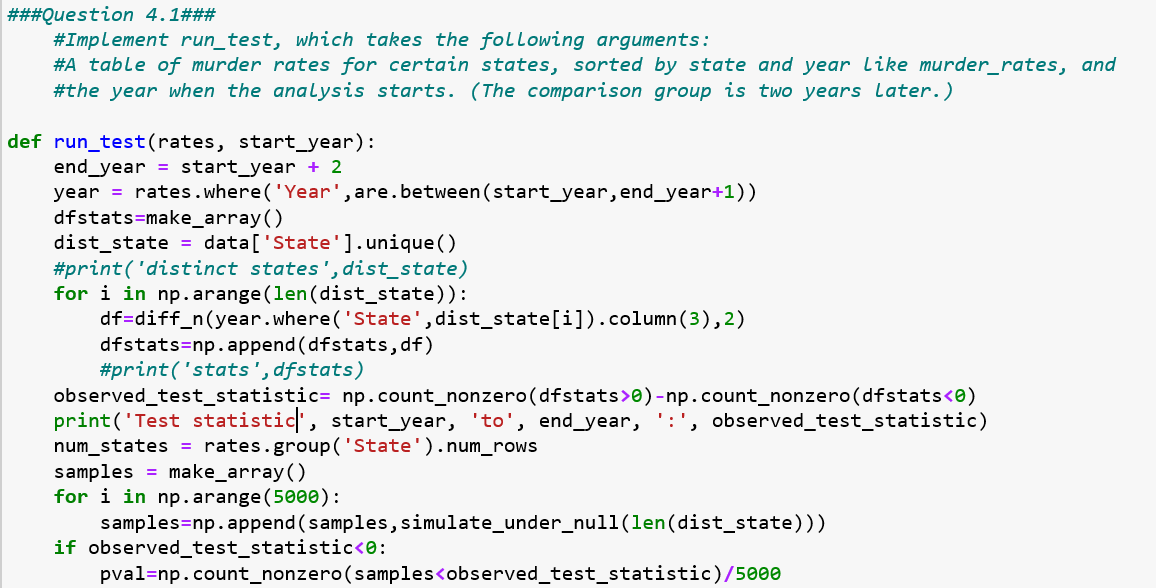


1. Conclusion about the hypotheses: As the P-value is 0.001, which is <1% and is significant, we can reject the null-hypothesis and accept the alternative hypothesis.
2. Findings: By looking at the empirical histograms and the P-value, the murder rates are more likely to increase after the death penalty was abolished. However, we cannot confirm the same as there cab be other factors which can influence the murder rate as stated in the points #1.1 and #3.2 sections.

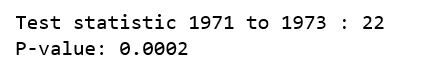
**SECTION 4: FURTHER EVIDENCE.**

**QUESTION 4.1:**

Run\_test function has been created, considering the states which has control group (i.e., states had death penalty) and run the test for the year when the analysis started i.e., 1971 for the period of two years. Hence, we will get the test statistic for the period 1971 to 1973.

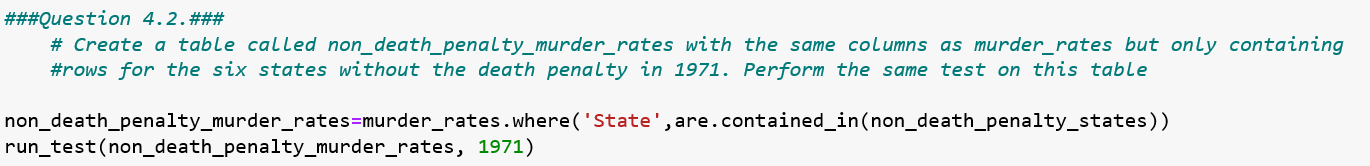


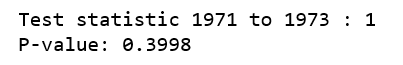
Output:



**QUESTION 4.2:**

In this part, control group states were analysed, i.e, the six states that didn’t had death penalty at all. The analysis was performed for the same years as we did in #4.1 i.e., from 1971 to 1973, to see whether the murder rates has been increased or decreased when compared to previous years.



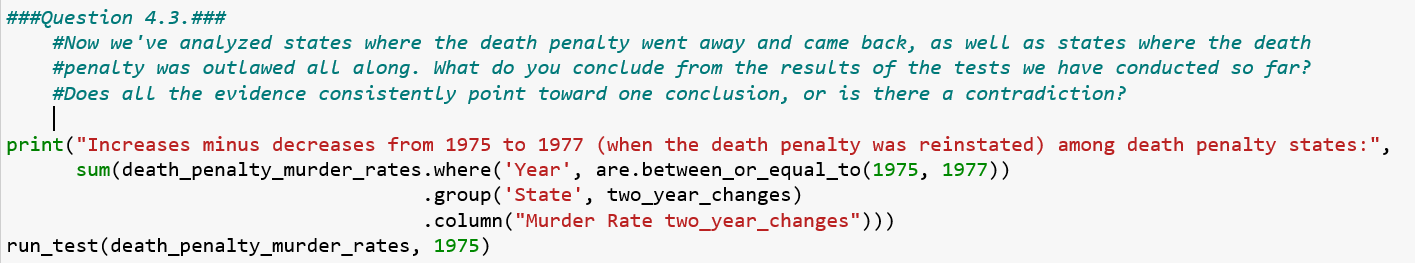


As per the test statistic, we can see the positive result, i.e., on an average for all the 6 states there is an increase of 1 in the murder rate from 1971 to 1973.

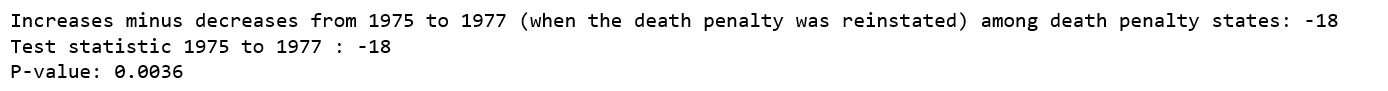
**QUESTION 4.3:**

Before I conclude on the analysis that I have done, I want to show the report for the 44 states (treatment group/ death penalty states) during the death penalty reinstated period (1975 to 1977).

The output shows there is a decrease in the murder rates by -18.



Output:



Conclusion:

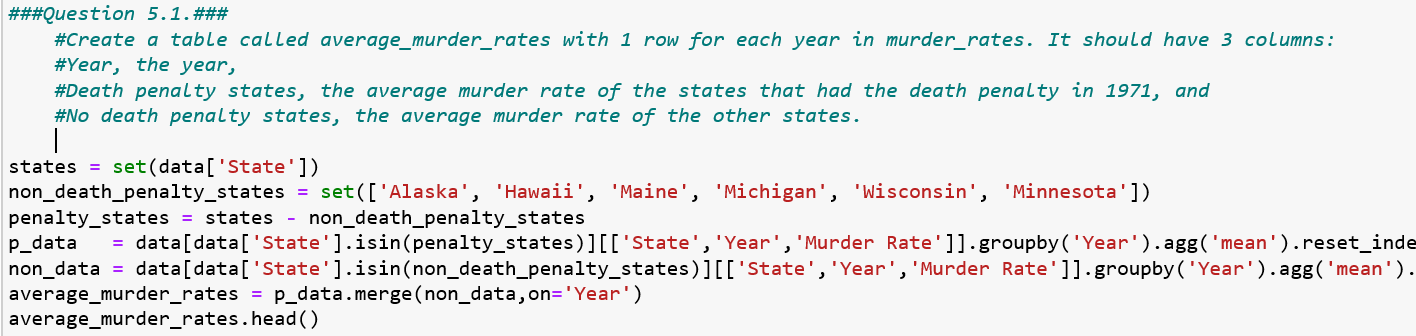
As per the analysis I have done, I have observed that there is an increase in the murder rates when the death penalty was temporarily abolished and a decrease in the murder rates when the death penalty is reinstated. This appears to support the conclusion that the death penalty deters murder.

However, we did not consider other factors like time/ places which may also have influence on the change of murder rates whether they increase or decrease. Though our analysis supports the conclusion that the death penalty deters murder, however, we cannot conclude unless we control the other influencing factors on murder rates like time/places.

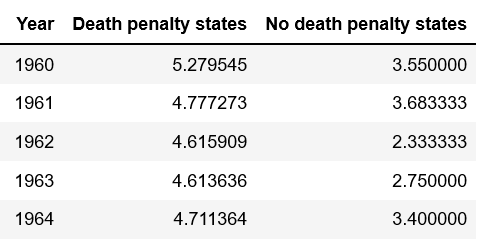
**SECTION 5: VISUALIZATION.**

**QUESTION 5.1:**

average\_murder\_rates table has been created with one row for each year. Apart from the year, two extra columns have been created, to populate the average of murder rates for “Death penalty states” and “No death penalty states”.

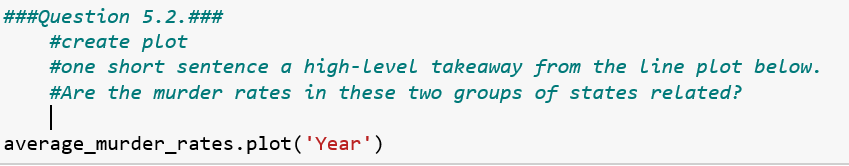


Output:

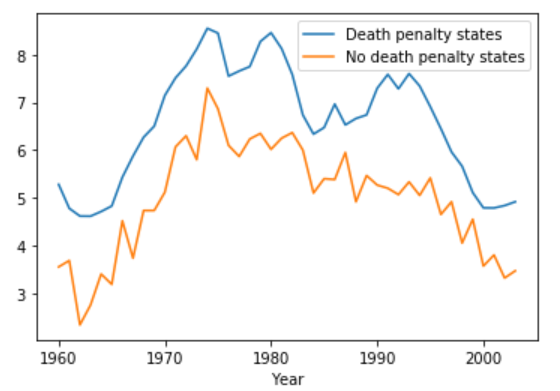


**QUESTION 5.2:**

A plot has been created to analyse the murder rates in the two groups that were created in the table “average\_murder\_rates”.



Output:



Takeaway:

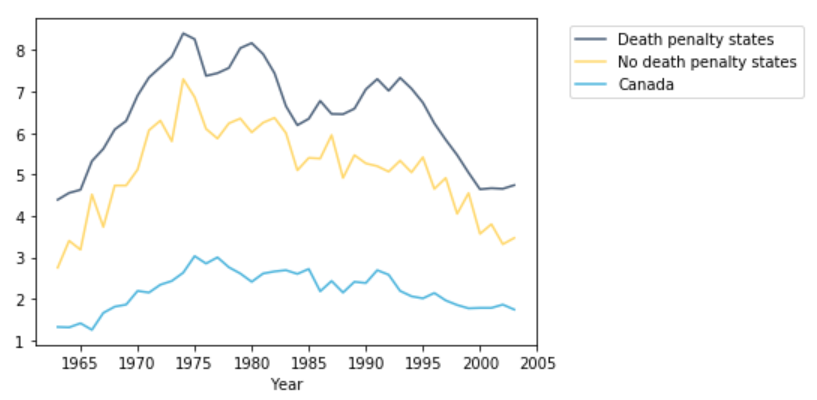
The ‘Death penalty states’ has the same pattern of increases and decreases in murder rates as the ‘No death penalty states’, however, the states with ‘no death penalty’ has the lower murder rates all over the years from 1960 to 2003, when compared to the ‘Death penalty states’.

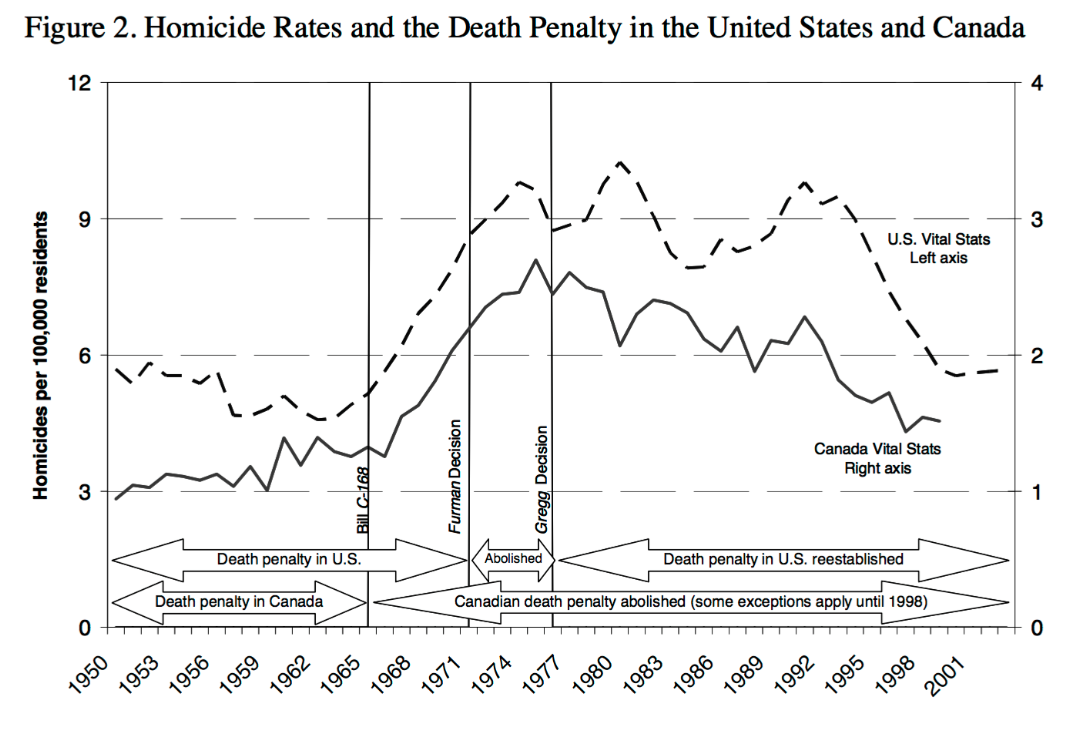
**QUESTION 5.3:**

Argument**:**

* The main feature of the plot is the murder rates pattern for all the 3 groups (US non death penalty states, US death penalty states, Canada) looks same irrespective of the factor death penalty establishment.
* Though, our analysis says that the murder rates from the years (1972 to 1976) have raised once the death penalty is abolished, the line charts clearly showing the gradual increase of death rates from almost 1962 for the ‘death penalty states’ of USA, even when the death penalty is in action, which says, we cannot attribute the death penalty to the increase in the murder rate.
* After the death penalty has reinstated in US, the murder rates seem to be gradually decreasing over time with a steep spike of increase at the years ~1981 and ~1990 and is the same case with Canada as well, although the death penalty has abolished in the year 1967, the murder rate gradually increased and then decreased with the some steeps as the US plot.

In overall, all the 3 groups follow the same pattern of murder rate for the years 1960 to 2003 irrespective of death penalty factor establishment.





**QUESTION 5.5:**

As per the analysis from part 1 to 4,

* The histogram plots, test statistic and P-value of “Death penalty states”, says that the murder rate has increased (Test statistic=22) after the death penalty abolishment (1972).
* When I did the same analysis for the years (1975-1977) after the death penalty reinstated, there is a significant decrease (Test statistic=-18) in the murder rates.
* With the natural experiment, though we controlled the differences among the ‘Death Penalty’ and ‘No death penalty’ states, I find an increase in the murder rates for the ‘Death penalty states’.
* All the above from our analysis supports the death penalty deterred murder.

From the line plot,

* However, When I see the line plots for three different groups (‘Death penalty states’, ‘No death penalty states’, ‘Canada’). All seems to follow the same pattern across all the years from ~1960 - ~2003, irrespective of the death penalty establishment factor. In contrary to our analysis, this clearly shown that the death penalty was not the only determining factor for the change in the murder rates.
* Though our analysis says that the murder rates from the years (1972 to 1976) have raised once the death penalty is abolished, the line charts clearly showing the gradual increase of death rates from almost 1962 for the ‘death penalty states’ of USA, even when the death penalty is in action, which says, we cannot attribute the death penalty to the increase in the murder rate.
* After the death penalty has reinstated in US, the murder rates seem to be gradually decreasing over time with a steep spike of increase at the years ~1981 and ~1990 like other two groups.
* As to Canada, although the death penalty has abolished in the year 1967, the murder rate gradually increased and then decreased with some steeps as the US plot.
* Hence, to determine the influence of death penalty on murder rates, we need to control several other factors (detailed information given in 3.2) like, location/place/area, season/time. If we are provided with the external factors which also influence the murder rates, we will be able to control these and can find the influence of death penalty over murder rates.