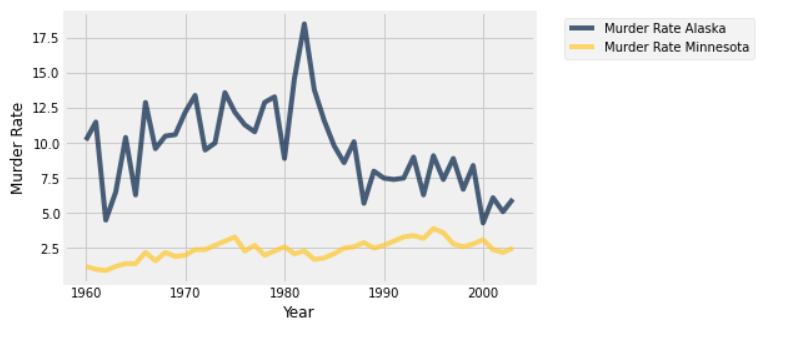
Aditya Pitchuka  
INFO 5502.002 (14676)  
Spring 2020

**Assignment – 5**

Selecting the needed columns from the table and setting format for population.

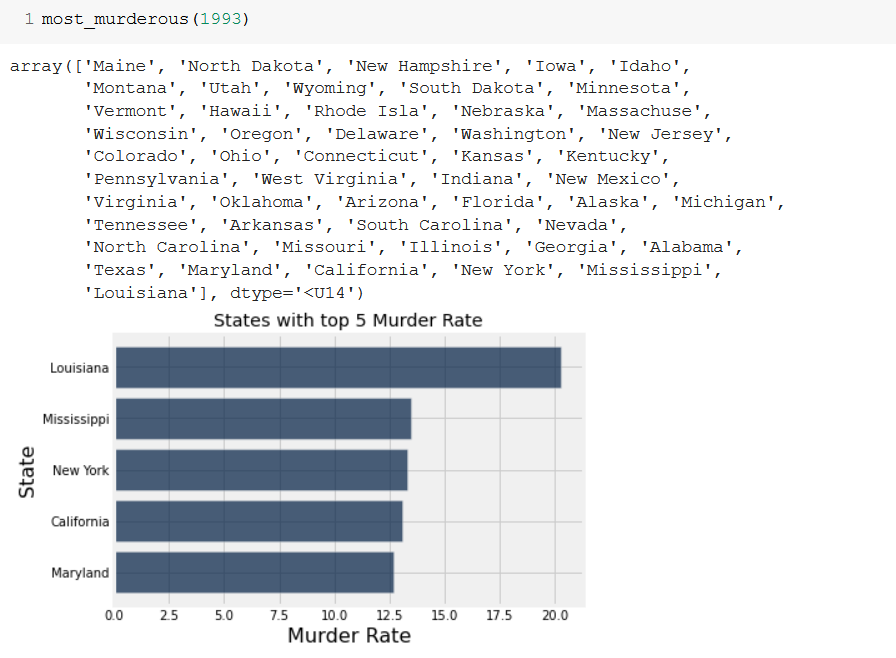


**Q 1.2** Draw a line plot with years on the horizontal axis and murder rates on the vertical axis.  
Include two lines: one for Alaska murder rates and one for Minnesota murder rates. Create this plot using a single call: ak\_mn.plot('Year').

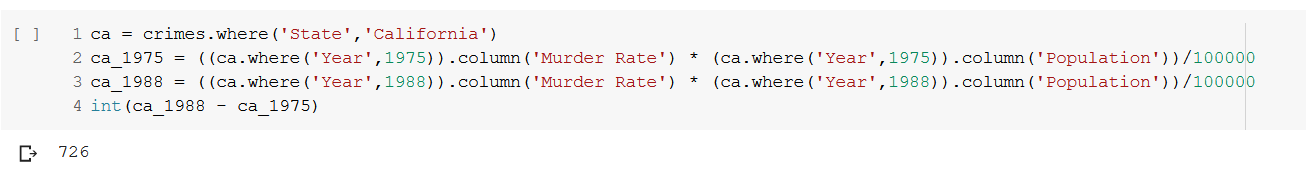


**Q 1.3** Implement the function most murderous, which takes a year (an integer) as its argument.  
It does two things:

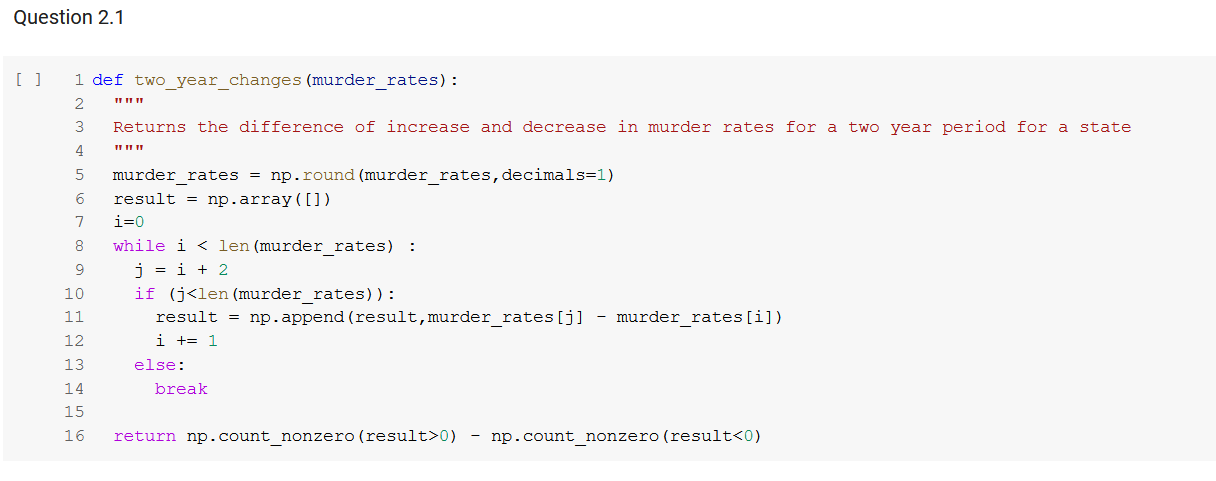
1. It draws a horizontal bar chart of the 5 states that had the highest murder rate in that year.
2. It returns an array of the names of these states in order of increasing murder rate.



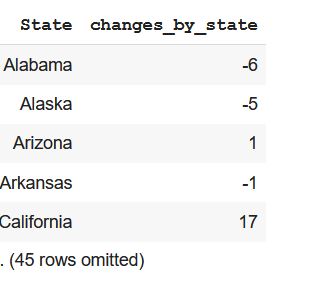
**Q 1.4** How many more people were murdered in California in 1988 than in 1975? Assign  
ca change to the answer.



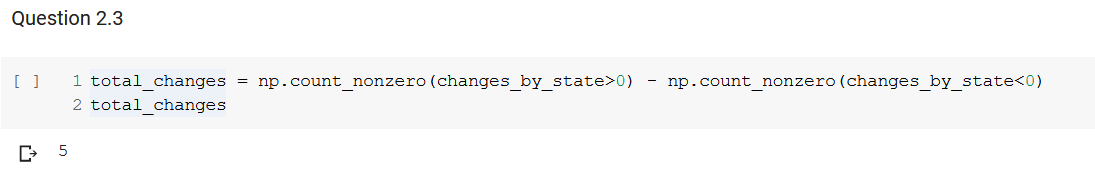
**Q 2.1** Implement the function two-year changes that takes an array of murder rates for a state,  
ordered by increasing year. For all two-year periods (e.g., from 1960 to 1962), it computes and  
returns the number of increases minus the number of decreases.



**Q 2.2** Assign changes by state to a table with one row per state that has two columns: the  
State name and the Murder Rate two year changes statistic computed across all years in our data set  
for that state.



**Q 2.3** Assign total changes to the total increases minus the total decreases for all two-year  
periods and all states in our data set.



**Q 2.4** Set num changes to the number of different two-year periods in the entire data set that  
could result in a change of a state's murder rate. Include both those periods where a change  
occurred and the periods where a state's rate happened to stay the same.



**Q 2.5** Given these null and alternative hypotheses, define a good test statistic.

Null Hypothesis: State murder rates increase and decrease over two-year periods as if \increase" or \decrease" were sampled at random from a uniform distribution, like a fair coin flip.

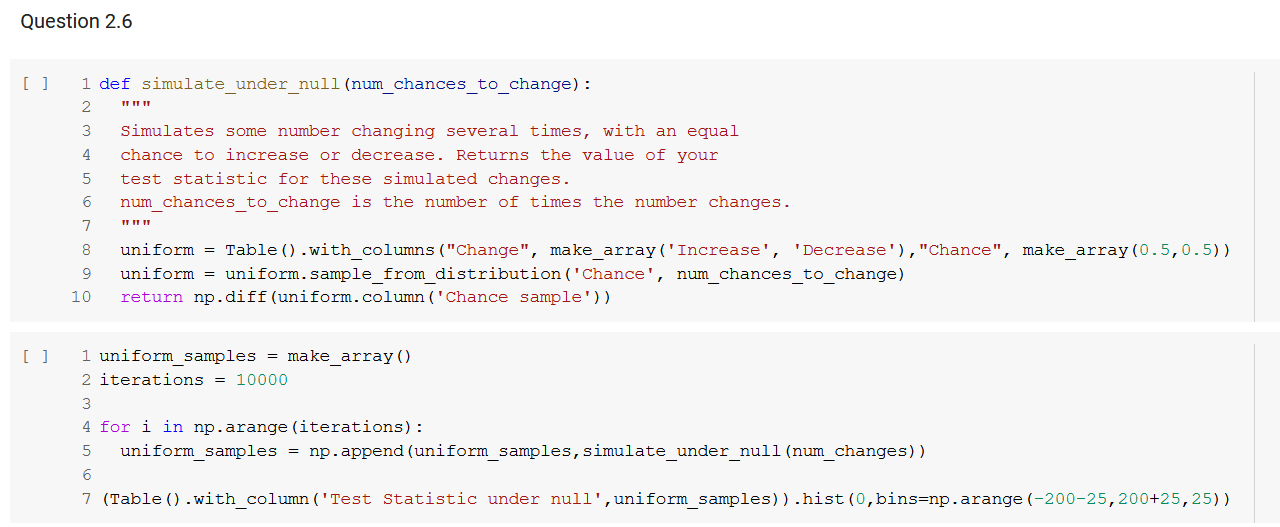
Alternative Hypothesis: State murder rates are either more likely or less likely to increase than decrease over two-year periods.

Test Statistic: Difference of increase and decrease of Murder Rate for a state

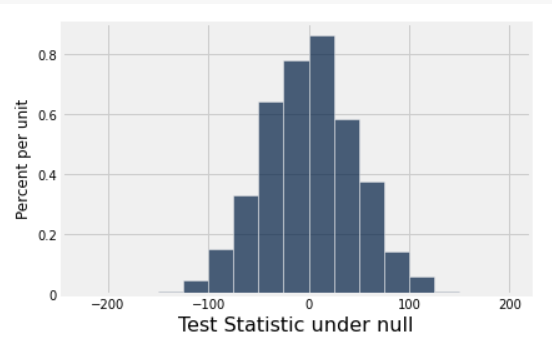
**Q 2.6** Complete the simulation below, which samples num changes increases/decreases at random many times and forms an empirical distribution of your test statistic under the null hypothesis.

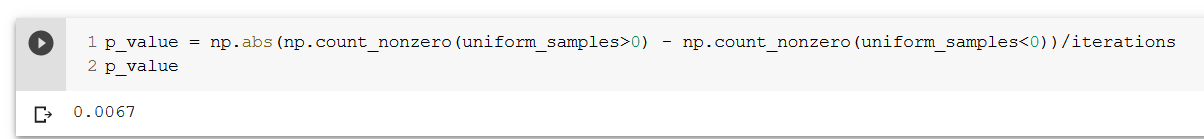
Your job is to

fill in the function simulate under null, which simulates a single sample under the null hypothesis,  
and  
fill in its argument when it's called below.



**Q 2.7** Looking at this histogram, draw a conclusion about whether murder rates basically  
increase as often as they decrease.





We see the left side and the right side of the histogram almost cancel each other.   
And also, we see the p value is about 1% which is 'highly statistically significant'

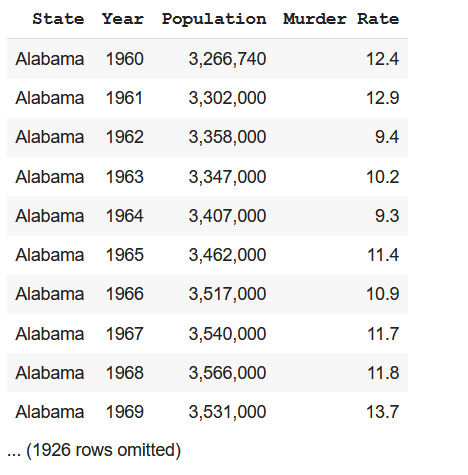
**Q 3.1** Describe this investigation in terms of an experiment. What population are we studying?  
What is the control group? What is the treatment group? What outcome are we measuring?

Population - All the people in the United States  
Control Group - People for whom Death Penalty is in effect  
Treatment Group - People for whom Death Penalty is in effect  
Outcome - Murder Rate

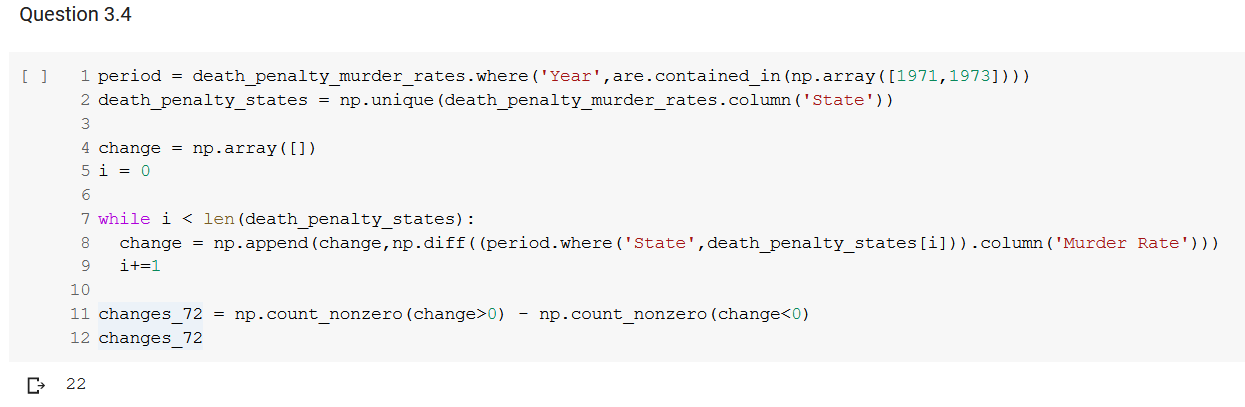
**Q 3.2** We want to know whether the death penalty causes a change in the murder rate. Why  
is it not sufficient to compare murder rates in places and times when the death penalty was in force with places and times when it wasn't?

Some years with low economic growth, high unemployment and other factors may have high crime rates and high murder rates compared to other years. So comparison across years may not indicate towards effect of capital punishment on murder rates.

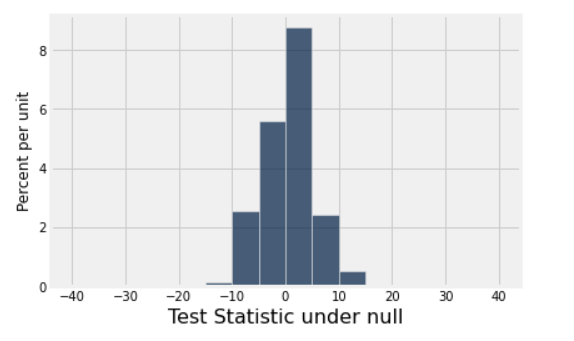
**Q 3.3** Assign death penalty murder rates to a table with the same columns and data as  
murder rates, but that has only the rows for states that had the death penalty in 1971.



**Q 3.4** Assign changes 72 to the value of the test statistic for the years 1971 to 1973 and the  
states in death penalty murder rates.

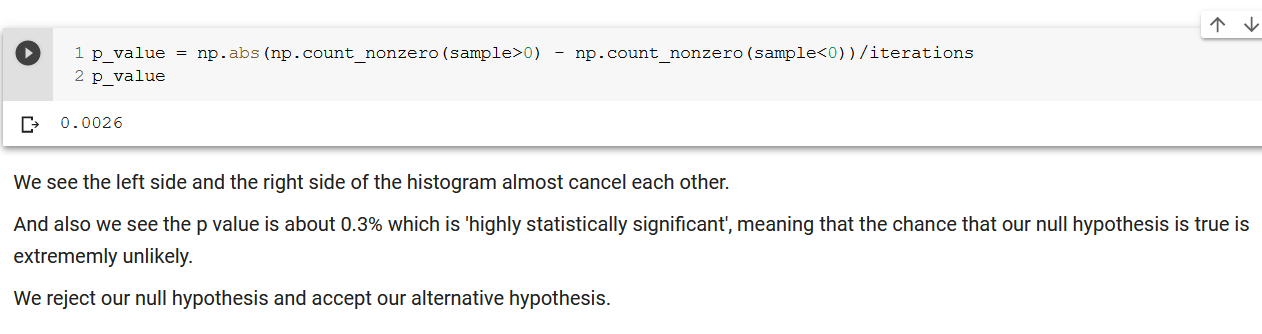


**Q 3.5** Draw an empirical histogram of the statistic under the null hypothesis by simulating  
the test statistic 5,000 times.



**Q 3.6** Complete the analysis as follows:

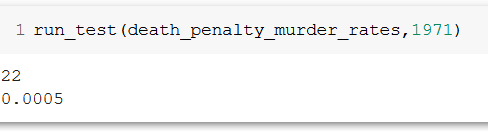
1. Compute a P-value.  
2. Draw a conclusion about the null and alternative hypotheses.  
3. Describe your findings using simple, non-technical language. Be careful not to claim that the  
statistical analysis has established more than it really has.



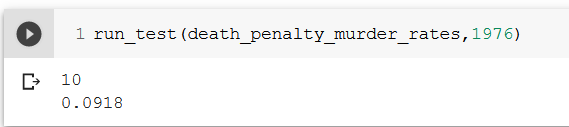
**Q 4.1** Implement run test, which takes the following arguments:

1. A table of murder rates for certain states, sorted by state and year like murder rates, and
2. the year when the analysis starts. (The comparison group is two years later.)

It prints out the observed test statistic and returns the P-value for this statistic under the null hypothesis.



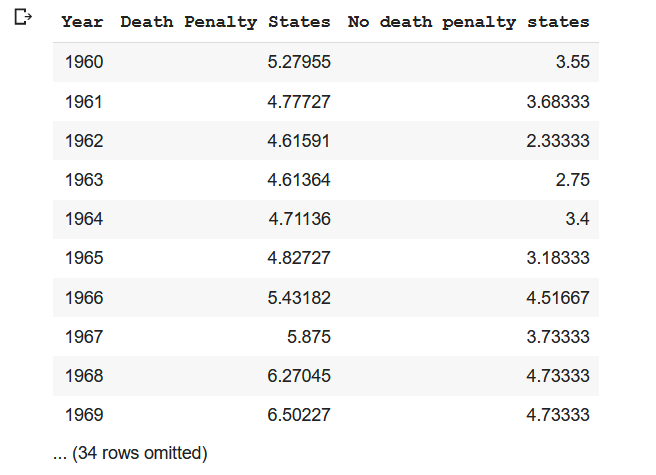
**Q 4.3** Now we've analyzed states where the death penalty went away and came back, as well  
as states where the death penalty was outlawed all along. What do you conclude from the results of the tests we have conducted so far? Does all the evidence consistently point toward one conclusion, or is there a contradiction?



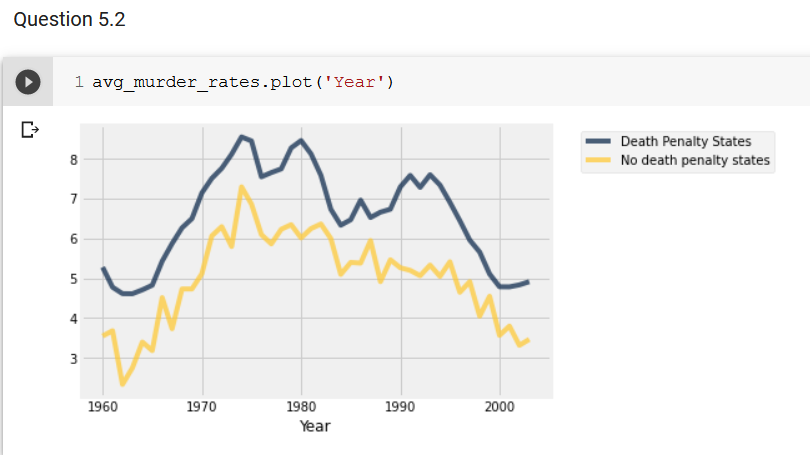
**Q 5.1** Create a table called average murder rates with 1 row for each year in murder rates. It  
should have 3 columns:

1. Year, the year,
2. Death penalty states, the average murder rate of the states that had the death penalty in 1971,  
   and
3. No death penalty states, the average murder rate of the other states.

Average murder rates should be sorted in increasing order by year.



**Q 5.2** Describe in one short sentence a high-level takeaway from the line plot below. Are the  
murder rates in these two groups of states related? average murder rates.plot('Year').



Murder rates for states with no death penalty are consistently lower than the ones with death penalty.   
Also, the rates seem to rise and fall together with a similar trend across all states, both with and without the death penalty.

**Q 5.3** Complete their argument in 2-3 sentences; what features of these plots indicate that  
the death penalty is not an important factor in determining the murder rate? (If you're stuck, read the paper.)

The plot shows that neither the abolishment nor the re-institaing death penalty directly affect murder rates.

Before to the abolishment, murder rates had been increasing steadily from 1960 to 1970.

**Q 5.5** What assumption(s) did we make in Parts 1 through 4 of the project that led us to  
believe that the death penalty deterred murder, when in fact the line plots tell a different story?

When in 1971 when the penalty is lifted in the states, we see our hypothesis supporting alternative. But in 1976 when it is repealed we see supporting null hypothesis again.

It keeps contradicting over the period of data as there are a lot of other external dependencies which is causing this trend.

Therefore, there does not seem to be a causal relationship between increases and decreases to murder rates and whether or not a state has the death penalty.