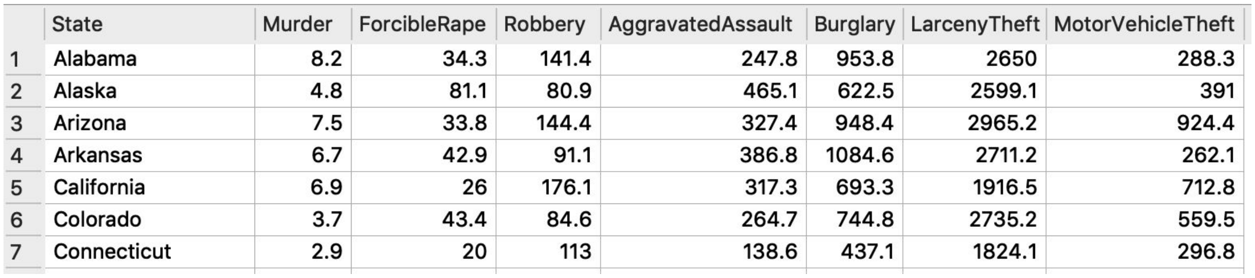
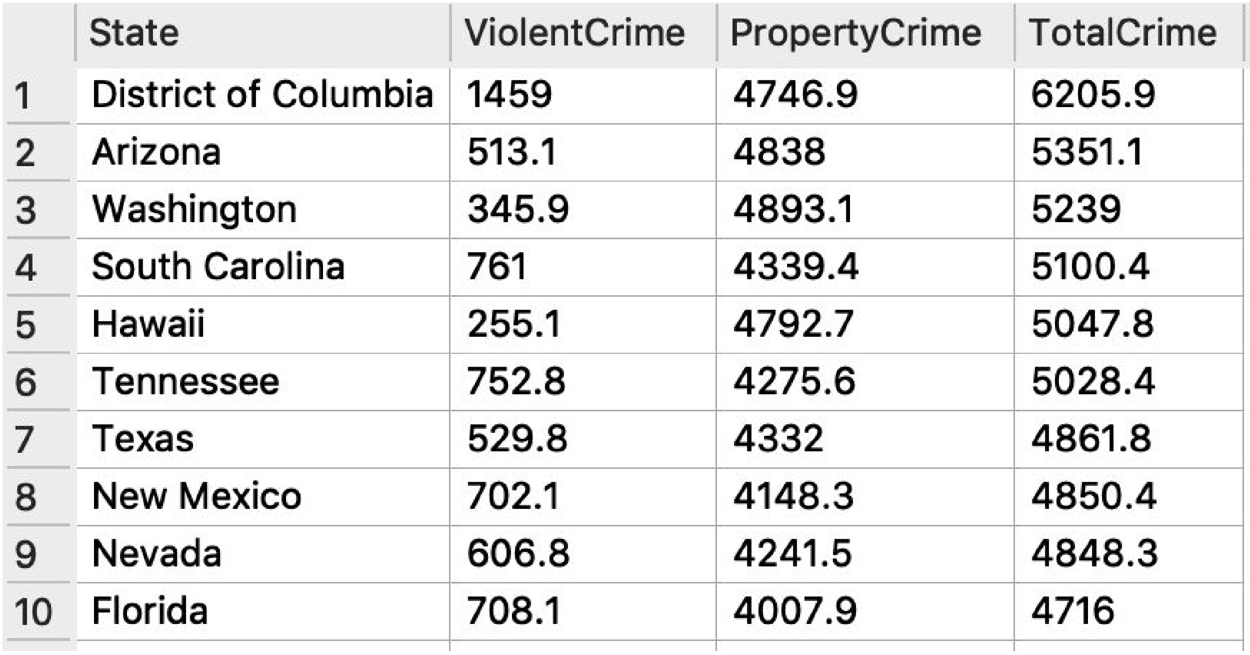
ORIE 3120: Practical Tools for Operations Research, Data Science, and Machine Learning
Spring 2020
Homework #1
Due Date: 2/5/19 (Wednesday) 2:30pm.

The database “CrimeRates2005.sqlite” has a simple structure. It consists of a single table, “CrimeRatesByState2005”. The table is populated with the crime rates per 100,000 people in the population by state for various crime categories. These data were downloaded from http://www.infochimps.com/datasets/crime-rates-by-state-2004-and-2005-and-by-type-2005-cleaned-up-v/downloads/131044. A sampling of rows is:



	State	Murder	ForcibleRape	Robbery	AggravatedAssault	Burglary	LarcenyTheft	MotorVehicleTheft
1	Alabama	8.2	34.3	141.4	247.8	953.8	2650	288.3
2	Alaska	4.8	81.1	80.9	465.1	622.5	2599.1	391
3	Arizona	7.5	33.8	144.4	327.4	948.4	2965.2	924.4
4	Arkansas	6.7	42.9	91.1	386.8	1084.6	2711.2	262.1
5	California	6.9	26	176.1	317.3	693.3	1916.5	712.8
6	Colorado	3.7	43.4	84.6	264.7	744.8	2735.2	559.5
7	Connecticut	2.9	20	113	138.6	437.1	1824.1	296.8

1. Write a view, called “Q01CrimeRateSummaryByState” that results in the following fields for each state and the District of Columbia:



	State	ViolentCrime	PropertyCrime	TotalCrime
1	District of Columbia	1459	4746.9	6205.9
2	Arizona	513.1	4838	5351.1
3	Washington	345.9	4893.1	5239
4	South Carolina	761	4339.4	5100.4
5	Hawaii	255.1	4792.7	5047.8
6	Tennessee	752.8	4275.6	5028.4
7	Texas	529.8	4332	4861.8
8	New Mexico	702.1	4148.3	4850.4
9	Nevada	606.8	4241.5	4848.3
10	Florida	708.1	4007.9	4716

where

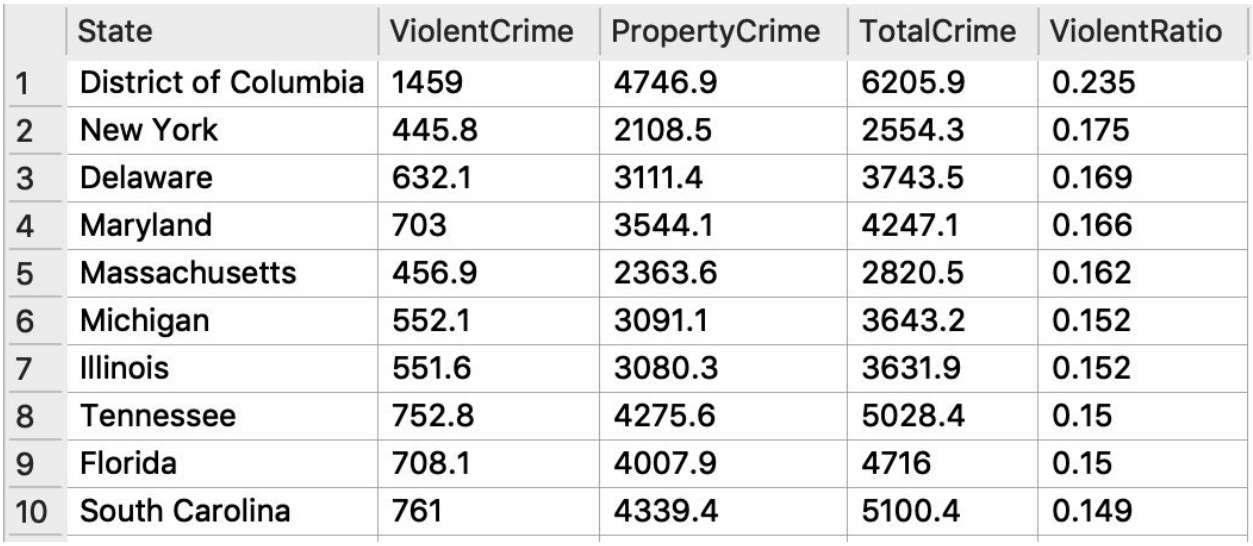
- “ViolentCrime” is the sum of “Murder,” “ForcibleRape,” “Robbery,” and “AggravatedAssault;”
- “PropertyCrime” is the sum of “Burglary,” “LarcenyTheft,” and “MotorVehicleTheft.”

- "TotalCrime" is the sum of all crimes in the state.
- The results are sorted in descending order of total crime.

Round off the results to the nearest tenth, using the function "Round(x,n)" with n (the number of digits of precision after the decimal point) set to 1.

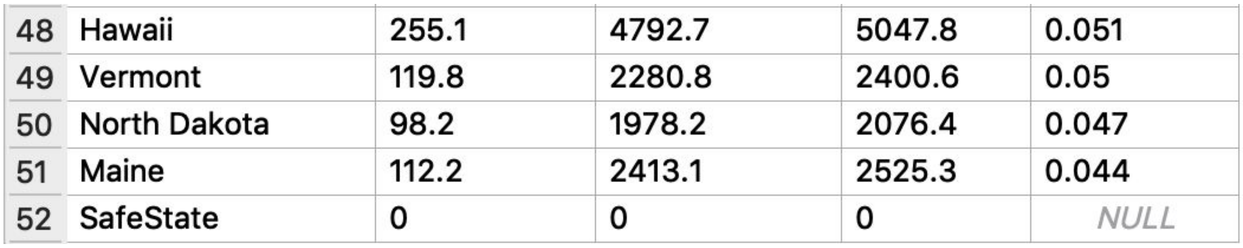
2. Write a second view, called "Q02ViolentRatio", that is based on Q01CrimeRateSummaryByState and adds an extra field, ViolentRatio, computed as ViolentCrime/TotalCrime when TotalCrime is nonzero and then rounded to the nearest thousandth. The view's result should be sorted in decreasing order of ViolentRatio. If there is a state with no crime, then your query should include this state in your result with ViolentRatio set to NULL.

The result should be:



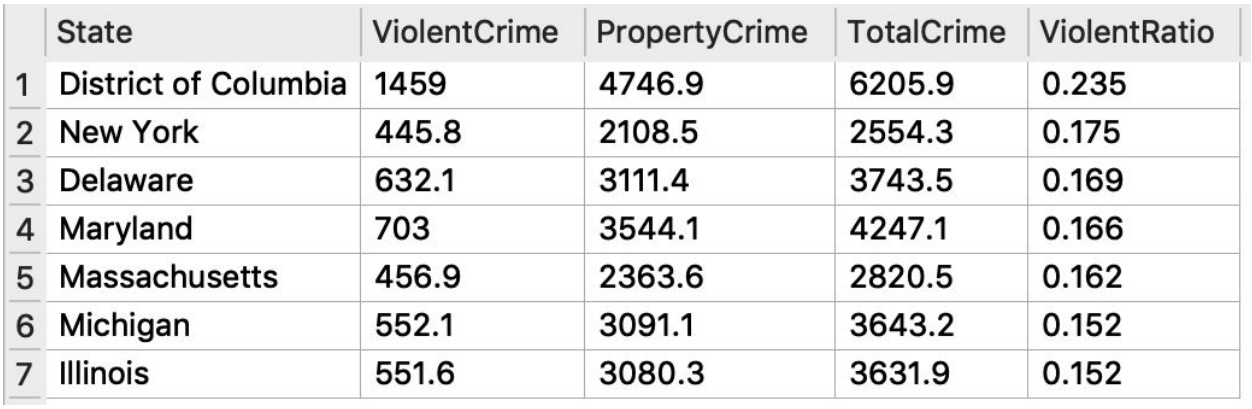
	State	ViolentCrime	PropertyCrime	TotalCrime	ViolentRatio
1	District of Columbia	1459	4746.9	6205.9	0.235
2	New York	445.8	2108.5	2554.3	0.175
3	Delaware	632.1	3111.4	3743.5	0.169
4	Maryland	703	3544.1	4247.1	0.166
5	Massachusetts	456.9	2363.6	2820.5	0.162
6	Michigan	552.1	3091.1	3643.2	0.152
7	Illinois	551.6	3080.3	3631.9	0.152
8	Tennessee	752.8	4275.6	5028.4	0.15
9	Florida	708.1	4007.9	4716	0.15
10	South Carolina	761	4339.4	5100.4	0.149

To test that your query provides the correct result when TotalCrime is 0, add a row to your table for a state called "SafeState" with no crime, and then re-run your query. It should appear at the end and be shown like this:



48	Hawaii	255.1	4792.7	5047.8	0.051
49	Vermont	119.8	2280.8	2400.6	0.05
50	North Dakota	98.2	1978.2	2076.4	0.047
51	Maine	112.2	2413.1	2525.3	0.044
52	SafeState	0	0	0	NULL

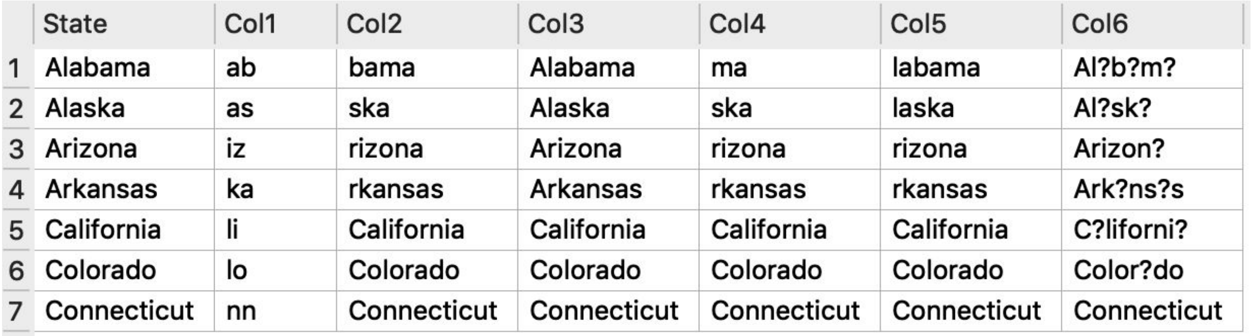
3. Write a third view, called Q03RelativelyViolentStates that returns ViolentCrime, PropertyCrime, TotalCrime, and ViolentRatio for those only those states whose ViolentRatio is strictly larger than 0.15. Your query result should look like this:



	State	ViolentCrime	PropertyCrime	TotalCrime	ViolentRatio
1	District of Columbia	1459	4746.9	6205.9	0.235
2	New York	445.8	2108.5	2554.3	0.175
3	Delaware	632.1	3111.4	3743.5	0.169
4	Maryland	703	3544.1	4247.1	0.166
5	Massachusetts	456.9	2363.6	2820.5	0.162
6	Michigan	552.1	3091.1	3643.2	0.152
7	Illinois	551.6	3080.3	3631.9	0.152

4. String Manipulation in SQL

Using CrimeRates2005.sqlite, write a query that produces these records. It can produce these records in any order, and it can also produce other records.



	State	Col1	Col2	Col3	Col4	Col5	Col6
1	Alabama	ab	bama	Alabama	ma	labama	Al?b?m?
2	Alaska	as	ska	Alaska	ska	laska	Al?sk?
3	Arizona	iz	rizona	Arizona	rizona	rizona	Arizon?
4	Arkansas	ka	rkansas	Arkansas	rkansas	rkansas	Ark?ns?s
5	California	li	California	California	California	California	C?liforni?
6	Colorado	lo	Colorado	Colorado	Colorado	Colorado	Color?do
7	Connecticut	nn	Connecticut	Connecticut	Connecticut	Connecticut	Connecticut

There are multiple ways to write this query. Please just provide one.

You may wish to use these string functions in your query: TRIM, LTRIM, RTRIM, SUBSTR, REPLACE. You can read about how these functions work in the sqlite documentation at https://www.sqlite.org/lang_corefunc.html.

5. Sorting and Comparing Strings

Modify your query for question 4 so that it only produces **only** those records shown in the query result for #4, and is guaranteed to produce them in the **order** shown.

When ordering and comparing two different strings, SQL uses ‘lexicographic order’: Strings beginning with ‘A’ are strictly smaller than those beginning with ‘B’, those beginning with ‘B’ are smaller than those beginning with ‘C’, etc. If the strings have the same first character, then comparison moves to the second character. If there is no second character then the shorter string is considered smaller. If the second character is the same, then comparison moves to the third character etc.