Spencer Burns, Yanki Saplan

Professor Rigas

DS220

December 14, 2020

Final Project

Introduction

The dataset that we chose was the firearm permits and background checks database. The data analyzes the United States, state by state and we thought it would be a good choice for this project as we have a dataset that has many options to work with.

Firearm permits and background checks Dataset

Dataset: https://github.com/BuzzFeedNews/nics-firearm-background-checks (Links to an external site.)

The data in this repository comes from the FBI's National Instant Criminal Background Check System

The NICS is used by the Federal Firearms Licensees to instantly determine whether a prospective buyer is eligible to buy firearms or explosives. Before ringing up the sale, cashiers call in a check to the FBI or to other designated agencies to ensure that each customer does not have a criminal record or isn't otherwise ineligible to make a purchase. There have been more than 100 million checks over the past decade, which has led to more than 700,000 denials over that timespan. The code in this GitHub repository downloads that PDF, parses it, and produces a spreadsheet/CSV of the data. The data collects data from about the last 20 years, ranging from November 1998 to April 2019.

The dataset we are working with contains these attributes:

- month
- state
- permit
- permit_recheck
- handgun
- long_gun
- other
- multiple
- admin
- prepawn_handgun
- prepawn_long_gun
- prepawn_other
- redemption_handgun
- redemption_long_gun
- redeption_other
- returned_handgun
- returned_long_gun
- returned_other
- rentals_handgun
- rentals_long_gun
- private_sale_handgun
- private_sale_long_gun
- private_sale_other
- return_to_seller_handgun
- return_to_seller_handgun
- return_to_seller_long_gun
- return_to_seller_other_totals

Using DB Browser for SQLite

Here, we go into the data using “firearm,” and are able to SELECT state, totals, and month from the data. This lets us identify totals by each state, month by month, but WHERE is used to only pull data from September of 2020 and September of 2009. This query allows comparison in totals between these two selected months.

```
1   SELECT state, totals, month
2   FROM firearm
3   WHERE month = "2020-09" OR month = "2000-09"
4
```

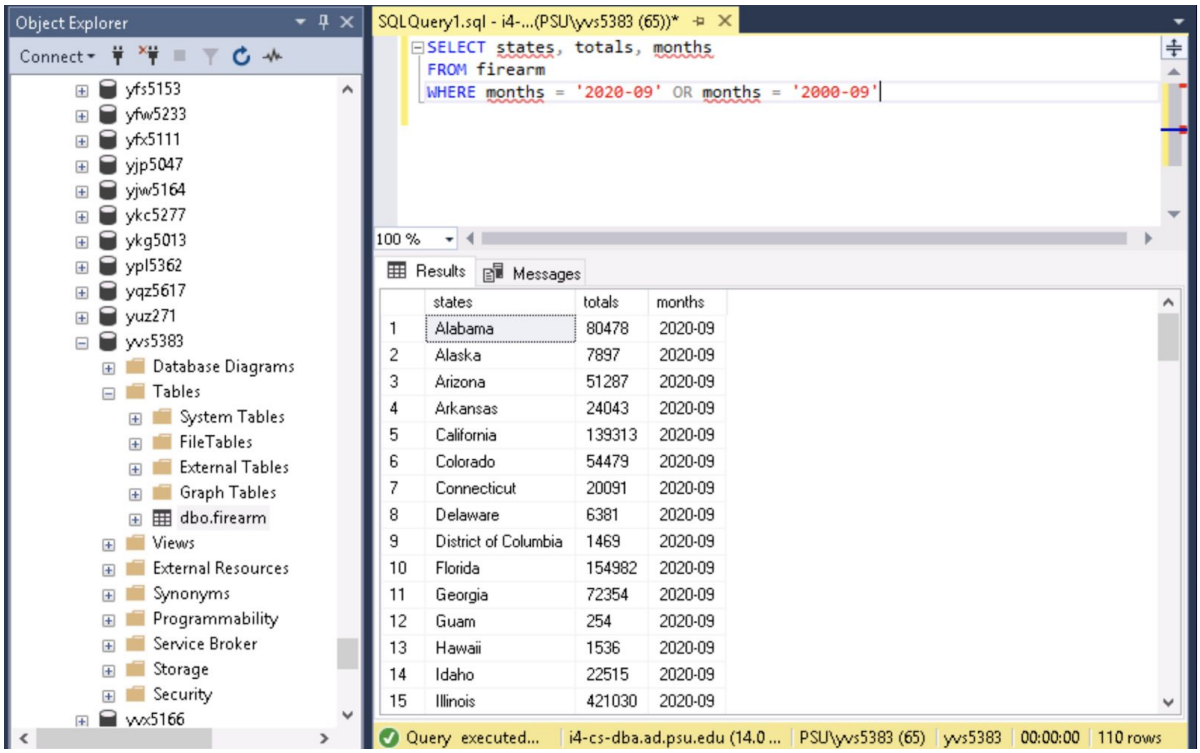
	state	totals	month
1	Alabama	80478	2020-09
2	Alaska	7897	2020-09
3	Arizona	51287	2020-09
4	Arkansas	24043	2020-09
5	California	139313	2020-09
6	Colorado	54479	2020-09
7	Connecticut	20091	2020-09
8	Delaware	6381	2020-09
9	District of Columbia	1469	2020-09
10	Florida	154982	2020-09
56	Alabama	19885	2000-09
57	Alaska	3553	2000-09
58	Arizona	12278	2000-09
59	Arkansas	15732	2000-09
60	California	59837	2000-09
61	Colorado	28974	2000-09
62	Connecticut	6702	2000-09
63	Delaware	1275	2000-09
64	District of Columbia	0	2000-09
65	Florida	21001	2000-09

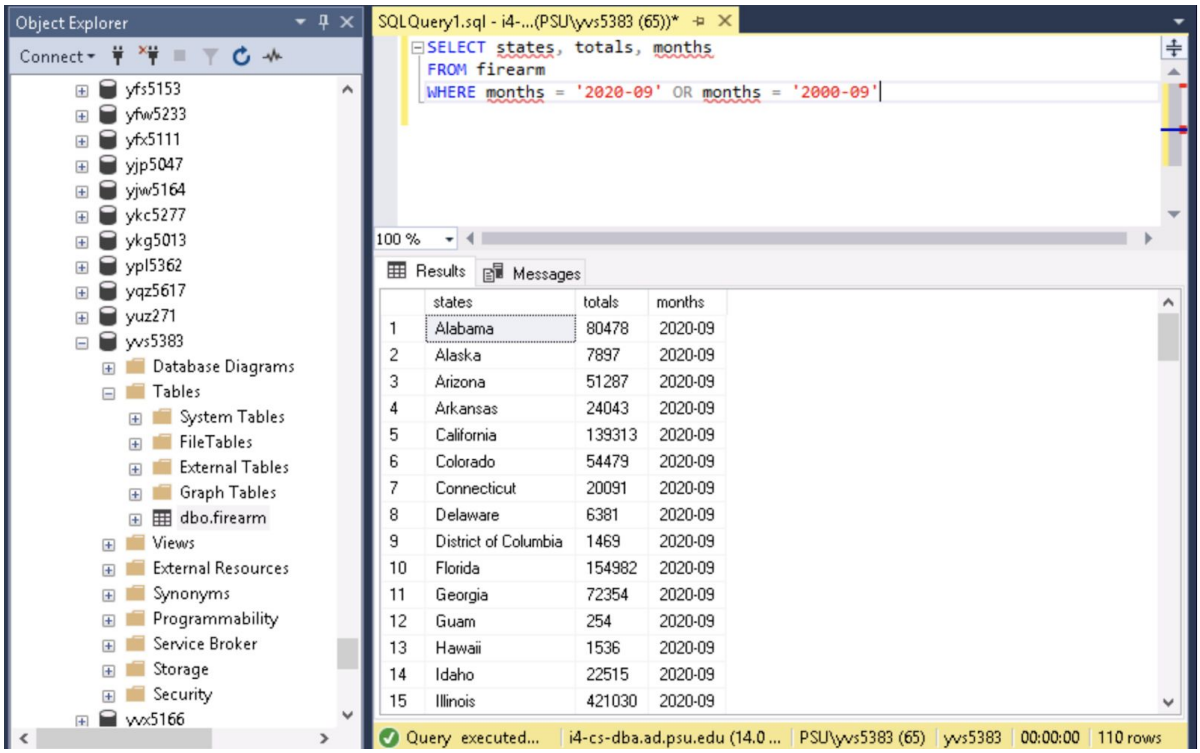
This second query allows to look at different statistics in the data and order them by # of permits in September of 2020.

```
SELECT month, state, permit, permit_recheck, totals
FROM firearm
WHERE month = "2020-09"
ORDER BY permit DESC
```

	state	permit	permit_recheck	month	totals
1	North Carolina	47822	28	2020-09	74810
2	Texas	46746	0	2020-09	178136
3	Pennsylvania	39823	0	2020-09	134626
4	Florida	34496	0	2020-09	154982
5	Illinois	33632	344401	2020-09	421030
6	Alabama	33228	642	2020-09	80478
7	California	32998	0	2020-09	139313
8	Georgia	30356	0	2020-09	72354
9	Michigan	29269	10762	2020-09	101789
10	Minnesota	28892	20113	2020-09	86830
11	Wisconsin	21724	770	2020-09	68257
12	Ohio	19438	106	2020-09	76519
13	Washington	18816	354	2020-09	67228
14	Utah	18425	82234	2020-09	114600
15	Massachusetts	14163	1	2020-09	26145

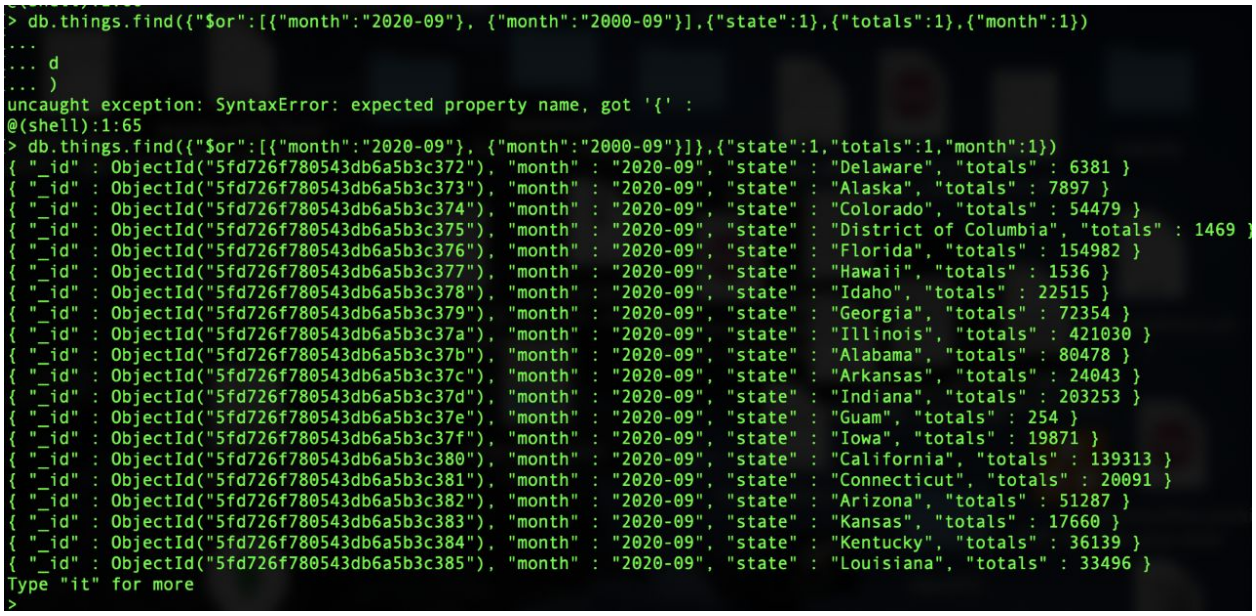
They do return the same results, but they have differences in terms of syntaxing.





Using MongoDB for NoSQL

This query is using .find to call a state's total in only the Septembers of 2000 and 2020, to see the rest of the results,



Uses .find to get the totals,

Introduction or background

The firearms data basically shows a lot of different firearm statistics according to their state and their month/year. The data shows how many individual totals(firearms) and other neat statistics, including permits and permit rechecks. I think that the data is interesting because you're able to see the difference in laws between states. The huge discrepancy in some of the data shows what is required with firearms in some states, and possibly the politics surrounding firearms in each state and their time period.

Dataset

We used the firearm permits and background checks data. The data goes by each state, and the month for that state. The data has over 20 years of data, 12 months for each year, and then every state and American territory. The data has some huge discrepancies, both over time and between states. You can see the certain laws or requirements from each state and that time period by looking at the numbers. Overall, the data is very good for analyzing firearm permits and background checking over the past 20 years and for each state. The data appears to have come from Buzzfeed and I think the only limitation in the dataset is that it doesn't account per capita. The numbers can be slightly misleading due to population size differences, but it does very well with showing possible laws or the political environment of a time period and state.

Methods

First, I downloaded the required files and data including the csv data from the github repositories. After downloading the DB Browser for SQLite, I have imported the csv data into a new database table. The first row had a problem while executing the sql command as the database table inserted the columns as rows. After fixing the mistake SQL execution was working, and I was able to run SQL commands we have decided on.

SQL

firearm		CREATE TABLE "firearm" ("month" TE
month	TEXT	"month" TEXT
state	TEXT	"state" TEXT
permit	INTEGER	"permit" INTEGER
permit_recheck	INTEGER	"permit_recheck" INTEGER
handgun	INTEGER	"handgun" INTEGER
long_gun	INTEGER	"long_gun" INTEGER
other	INTEGER	"other" INTEGER
multiple	INTEGER	"multiple" INTEGER
admin	INTEGER	"admin" INTEGER
prepawn_handgun	INTEGER	"prepawn_handgun" INTEGER
prepawn_long_gun	INTEGER	"prepawn_long_gun" INTEGER
prepawn_other	INTEGER	"prepawn_other" INTEGER
redemption_handgun	INTEGER	"redemption_handgun" INTEGER
redemption_long_gun	INTEGER	"redemption_long_gun" INTEGER
redemption_other	INTEGER	"redemption_other" INTEGER
returned_handgun	INTEGER	"returned_handgun" INTEGER
returned_long_gun	INTEGER	"returned_long_gun" INTEGER
returned_other	INTEGER	"returned_other" INTEGER
rentals_handgun	INTEGER	"rentals_handgun" INTEGER
rentals_long_gun	INTEGER	"rentals_long_gun" INTEGER
private_sale_handgun	INTEGER	"private_sale_handgun" INTEGER
private_sale_long_gun	INTEGER	"private_sale_long_gun" INTEGER
private_sale_other	INTEGER	"private_sale_other" INTEGER
return_to_seller_handgun	INTEGER	"return_to_seller_handgun" INTEGER
return_to_seller_long_gun	INTEGER	"return_to_seller_long_gun" INTEGER
return_to_seller_other	INTEGER	"return_to_seller_other" INTEGER
totals	INTEGER	"totals" INTEGER

New Database | Open Database | Write Changes | Revert Changes | Open Project | Save Project | Attach Database | Close Database

Database Structure | Edit Pragmas | Browse Data | Execute SQL

Table: firearm — Filter in any column

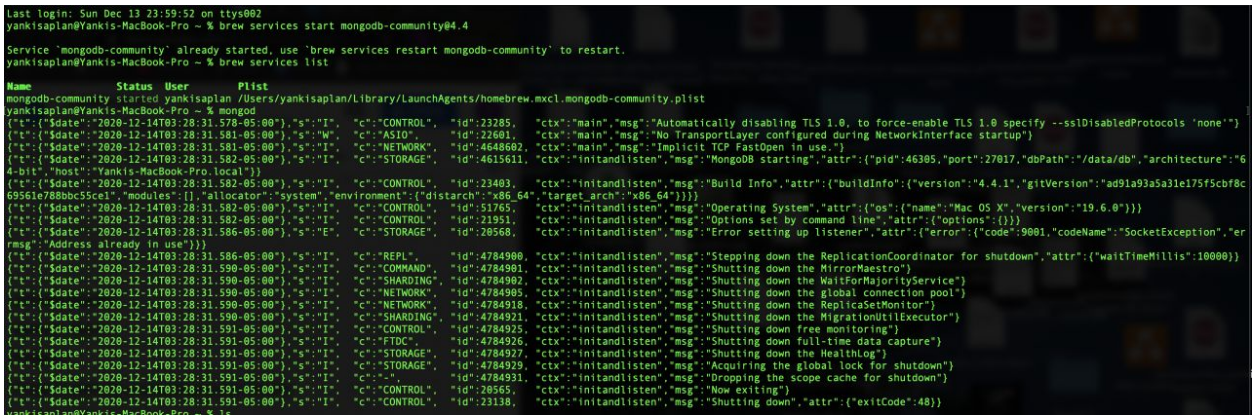
	month	state	permit	permit_recheck	handgun	long_gun	other	multiple	admin	prepawn_handgun	prepawn_long_gun	prepawn_other	redemption_handgun	redemption_long_gun	redemp
1	2020-03	Texas	35688	0	147714	61827	5897	8286	0	138	69	5	9200	5047	
2	2020-06	Florida	20676	0	133285	39111	7341	4846	0	12	4	0	2440	723	
3	2020-06	Texas	41486	0	121926	45845	6857	3968	7	61	39	7	4470	2267	
4	2020-03	Florida	16713	0	117900	38365	5017	6073	0	26	8	0	4912	1602	
5	2015-12	Texas	33163	NULL	107224	80247	4585	4505	0	47	38	1	6643	5168	
6	2020-07	Florida	33826	0	106745	33102	7751	3832	0	13	6	1	2830	767	
7	2012-12	Texas	12389	NULL	98886	108058	3699	5293	0	37	49	0	5761	5767	
8	2020-07	Texas	58415	0	96689	41060	6431	3366	22	76	33	6	5241	2337	
9	2020-04	Texas	38817	0	93980	40921	5668	4490	2	104	70	1	7989	5015	
10	2020-06	Pennsylvania	35753	0	90652	24029	66	0	399	0	0	0	0	0	
11	2015-12	Pennsylvania	44433	NULL	90055	18758	0	0	57	0	0	0	0	0	
12	2013-01	Texas	19093	NULL	89130	60897	3015	3827	0	42	42	0	5484	5338	
13	2015-12	Florida	25044	NULL	86940	45769	4165	3005	85	22	10	0	2637	1093	
14	2020-08	Florida	35826	0	85345	30914	6509	3193	0	9	5	0	2510	750	
15	2020-04	Florida	2258	0	84431	26731	4897	3451	0	9	7	0	4131	1515	
16	2020-05	Texas	41584	0	83977	36456	5393	3142	3	78	42	7	5001	3247	
17	2020-03	California	27792	0	81543	48616	5041	0	0	0	0	0	957	539	
18	2020-03	Pennsylvania	20598	0	81319	22031	30	0	317	0	0	0	0	0	
19	2020-07	Pennsylvania	41217	9	77854	22276	68	0	686	0	0	0	0	0	

NoSQL

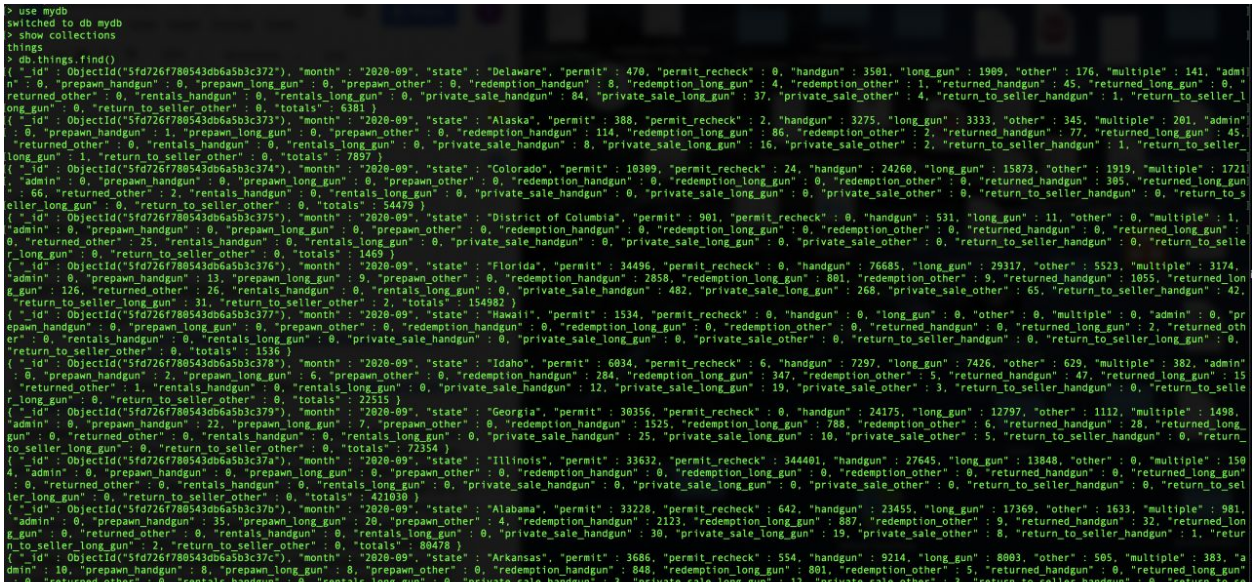
Importing to the MongoDB terminal.

```
yankisaplan@Yankis-MacBook-Pro ~ % cd Desktop
yankisaplan@Yankis-MacBook-Pro Desktop % mongoimport -d mydb -c things --type csv --file nicsfirearm.csv --headerline
2020-12-14T03:48:55.581-0500    connected to: mongodb://localhost/
2020-12-14T03:48:56.039-0500    14465 document(s) imported successfully. 0 document(s) failed to import.
```

An example on running MongoDB.



Using mydb, able to switch to our database and return everything within the database.



Results

I was able to pull out the same output, although I have received a lot of errors at first. Firstly, loading the csv data into my database took some time, even though I reloaded it for 30 minutes. After I solved that problem, there was a problem when I executed my commands, first one was that sql did not understand ‘state’ and ‘month’ as a column, instead executed as an inbuilt function which made me change the column names to ‘states’ and ‘months’. Another error that took me so long to understand was that IST SQL accepts single quotes where my SQLite accepts double quotes. The images can be found at page number 4 and 5.

Discussion

First, both of us both agreed that we like using the DB Browser for SQLite because it felt a lot cleaner and less clunky than other database methods used in the past. This program was very easily navigable and simple to upload our data into. It was also a lot faster and smoother, I would assume because we aren’t “logged into” a PSU computer. I think for circumstances involving the organization of data, the DB Browser would be better due to a simple infrastructure and easy usage once the data is imported.

The obvious differences between the NoSQL tool and the relational tool is the look and how the data is interpreted. I’d argue that the relational tool has a very clean cut look to it that helps look at the data. The SQL tool also had many buttons, or different methods of using the program, it was much simpler to use. The SQL section of the project took a much shorter duration than its NoSQL counterpart. The NoSQL data tool was all coding and a lot less user friendly. Taking a lot more time with this tool and its less-inviting infrastructure sums up the time using MongoDB.

Collaboration report

Both Spencer and Yanki

- On Zoom call while writing the queries and discussing what we need to do
- Figured out what we wanted the queries to be and what we want from the data
- Decided which dataset to use
- Zoom call used for the entire assignment, good communication with each other on what we need to do before the deadlines

Spencer

- Questions about discussion, dataset, and background
- Did first query
- Formatting of the project
- Difference of SQL and NoSQL data tools

Yanki

- Executed first, and second query
- Questions about results and methods
- Downloaded repo from github created the database, columns, and rows
- Executed SQL commands both DB SQLite, IST vlabs, and MongoDB