

# investigate-a-dataset-template

May 13, 2020

## 1 Project: FBI Gun Investigation and Data Analysis

### 1.1 Table of Contents

Introduction

Data Wrangling

Exploratory Data Analysis

Conclusions

## Introduction

The source of the data is FBI's National Instant Criminal Background Check System (NICS). This system is used to determine whether a prospective buyer is eligible to buy firearms or explosives. Gun shops call into this system to ensure that each customer does not have a criminal record or isn't otherwise ineligible to make a purchase. The data has been supplemented with state level data from census.gov.

#### About the data:

The NICS data is found in a single sheet of an .xlsx file. It contains the number of firearm checks by month, state, and type. The U.S. census data is found in a .csv file. It contains several variables at the state level. Most variables just have one data point per state (2016), but a few have data for more than one year.

Questions asked: 1. Which state has the highest total gun purchases in April 2000 and April 2010? 2. What is per capita firearm sales for all states in April 2010 vs July 2016?

```
In [420]: # Use this cell to set up import statements for all of the packages that you plan to use
import pandas as pd
import numpy as np
import datetime

# Remember to include a 'magic word' so that your visualizations are plotted
# inline with the notebook. See this page for more:
# http://ipython.readthedocs.io/en/stable/interactive/magics.html
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
sns.set_style('darkgrid')
```

## ## Data Wrangling

**Tip:** In this section of the report, you will load in the data, check for cleanliness, and then trim and clean your dataset for analysis. Make sure that you document your steps carefully and justify your cleaning decisions.

### 1.1.1 General Properties

```
In [421]: # Load your data and print out a few lines. Perform operations to inspect data
#         types and look for instances of missing or possibly errant data.
df_census = pd.read_csv('U.S. Census Data.csv')
df_gun = pd.read_excel('gun_data.xlsx')
```

```
In [422]: # Printing out a few lines of both the datasets
display(df_census.head())
display(df_gun.head())
```

```
      State  Population estimates, July 1, 2016, (V2016) \
0    Alabama                                4863300
1    Alaska                                  741894
2    Arizona                                6931071
3    Arkansas                               2988248
4    California                             39250017
```

```
Population estimates base, April 1, 2010, (V2016) \
0                                4780131
1                                710249
2                                6392301
3                                2916025
4                                37254522
```

```
Population, percent change - April 1, 2010 (estimates base) to July 1, 2016, (V2016) \
0                                0.017
1                                0.045
2                                0.084
3                                0.025
4                                0.054
```

```
Population, Census, April 1, 2010 \
0                                4779736
1                                710231
2                                6392017
3                                2915918
4                                37253956
```

```
Persons under 5 years, percent, July 1, 2016, (V2016) \
0                                0.060
1                                0.073
2                                0.063
```

3	0.064
4	0.063

	Persons under 5 years, percent, April 1, 2010 \
0	0.064
1	0.076
2	0.071
3	0.068
4	0.068

	Persons under 18 years, percent, July 1, 2016, (V2016) \
0	0.226
1	0.252
2	0.235
3	0.236
4	0.232

	Persons under 18 years, percent, April 1, 2010 \
0	0.237
1	0.264
2	0.255
3	0.244
4	0.250

	Persons 65 years and over, percent, July 1, 2016, (V2016) ... \	
0	0.161	...
1	0.104	...
2	0.169	...
3	0.163	...
4	0.136	...

	All firms, 2012	Men-owned firms, 2012	Women-owned firms, 2012 \
0	374153	203604	137630
1	68032	35402	22141
2	499926	245243	182425
3	231959	123158	75962
4	3548449	1852580	1320085

	Minority-owned firms, 2012	Nonminority-owned firms, 2012 \
0	92219	272651
1	13688	51147
2	135313	344981
3	35982	189029
4	1619857	1819107

	Veteran-owned firms, 2012	Nonveteran-owned firms, 2012 \
0	41943	316984
1	7953	56091

2	46780	427582
3	25915	192988
4	252377	3176341

	Population per square mile, 2010	Land area in square miles, 2010	FIPS Code
0	94.4	50645.33	"01"
1	1.2	570640.95	"02"
2	56.3	113594.08	"04"
3	56.0	52035.48	"05"
4	239.1	155779.22	"06"

[5 rows x 66 columns]

	month	state	permit	permit_recheck	handgun	long_gun	other	\
0	2017-09	Alabama	16717.0	0.0	5734.0	6320.0	221.0	
1	2017-09	Alaska	209.0	2.0	2320.0	2930.0	219.0	
2	2017-09	Arizona	5069.0	382.0	11063.0	7946.0	920.0	
3	2017-09	Arkansas	2935.0	632.0	4347.0	6063.0	165.0	
4	2017-09	California	57839.0	0.0	37165.0	24581.0	2984.0	

	multiple	admin	prepawn_handgun	...	returned_other	rentals_handgun	\
0	317	0.0	15.0	...	0.0	0.0	
1	160	0.0	5.0	...	0.0	0.0	
2	631	0.0	13.0	...	0.0	0.0	
3	366	51.0	12.0	...	0.0	0.0	
4	0	0.0	0.0	...	0.0	0.0	

	rentals_long_gun	private_sale_handgun	private_sale_long_gun	\
0	0.0	9.0	16.0	
1	0.0	17.0	24.0	
2	0.0	38.0	12.0	
3	0.0	13.0	23.0	
4	0.0	0.0	0.0	

	private_sale_other	return_to_seller_handgun	return_to_seller_long_gun	\
0	3.0	0.0	0.0	
1	1.0	0.0	0.0	
2	2.0	0.0	0.0	
3	0.0	0.0	2.0	
4	0.0	0.0	0.0	

	return_to_seller_other	totals
0	3.0	32019
1	0.0	6303
2	0.0	28394
3	1.0	17747
4	0.0	123506

[5 rows x 27 columns]

```
In [423]: df_census.set_index('State',inplace = True)
          df_census
```

```
Out[423]:
```

	Population estimates, July 1, 2016, (V2016) \
State	
Alabama	4863300
Alaska	741894
Arizona	6931071
Arkansas	2988248
California	39250017
Colorado	5540545
Connecticut	3576452
Delaware	952065
Florida	20612439
Georgia	10310371
Hawaii	1428557
Idaho	1683140
Illinois	12801539
Indiana	6633053
Iowa	3134693
Kansas	2907289
Kentucky	4436974
Louisiana	4681666
Maine	1331479
Maryland	6016447
Massachusetts	6811779
Michigan	9928300
Minnesota	5519952
Mississippi	2988726
Missouri	6093000
Montana	1042520
Nebraska	1907116
Nevada	2940058
New Hampshire	1334795
New Jersey	8944469
New Mexico	2081015
New York	19745289
North Carolina	10146788
North Dakota	757952
Ohio	11614373
Oklahoma	3923561
Oregon	4093465
Pennsylvania	12784227
Rhode Island	1056426

South Carolina	4961119
South Dakota	865454
Tennessee	6651194
Texas	27862596
Utah	3051217
Vermont	624594
Virginia	8411808
Washington	7288000
West Virginia	1831102
Wisconsin	5778708
Wyoming	585501

Population estimates base, April 1, 2010, (V2016) \

State	
Alabama	4780131
Alaska	710249
Arizona	6392301
Arkansas	2916025
California	37254522
Colorado	5029324
Connecticut	3574114
Delaware	897936
Florida	18804592
Georgia	9688680
Hawaii	1360301
Idaho	1567650
Illinois	12831574
Indiana	6484136
Iowa	3046869
Kansas	2853129
Kentucky	4339344
Louisiana	4533479
Maine	1328364
Maryland	5773786
Massachusetts	6547813
Michigan	9884129
Minnesota	5303924
Mississippi	2968103
Missouri	5988928
Montana	989414
Nebraska	1826334
Nevada	2700691
New Hampshire	1316461
New Jersey	8791953
New Mexico	2059198
New York	19378110
North Carolina	9535688
North Dakota	672591

Ohio	11536727
Oklahoma	3751615
Oregon	3831072
Pennsylvania	12702857
Rhode Island	1052940
South Carolina	4625410
South Dakota	814195
Tennessee	6346298
Texas	25146100
Utah	2763888
Vermont	625741
Virginia	8001041
Washington	6724545
West Virginia	1853011
Wisconsin	5687289
Wyoming	563767

Population, percent change - April 1, 2010 (estimates base) to July 1,

State	
Alabama	0.017
Alaska	0.045
Arizona	0.084
Arkansas	0.025
California	0.054
Colorado	0.102
Connecticut	0.001
Delaware	0.060
Florida	0.096
Georgia	0.064
Hawaii	0.050
Idaho	0.074
Illinois	-0.002
Indiana	0.023
Iowa	0.029
Kansas	0.019
Kentucky	0.022
Louisiana	0.033
Maine	0.002
Maryland	0.042
Massachusetts	0.040
Michigan	0.004
Minnesota	0.041
Mississippi	0.007
Missouri	0.017
Montana	0.054
Nebraska	0.044
Nevada	0.089
New Hampshire	0.014

New Jersey	0.017
New Mexico	0.011
New York	0.019
North Carolina	0.064
North Dakota	0.127
Ohio	0.007
Oklahoma	0.046
Oregon	0.068
Pennsylvania	0.006
Rhode Island	0.003
South Carolina	0.073
South Dakota	0.063
Tennessee	0.048
Texas	0.108
Utah	0.104
Vermont	-0.002
Virginia	0.051
Washington	0.084
West Virginia	-0.012
Wisconsin	0.016
Wyoming	0.039

Population, Census, April 1, 2010 \

State	
Alabama	4779736
Alaska	710231
Arizona	6392017
Arkansas	2915918
California	37253956
Colorado	5029196
Connecticut	3574097
Delaware	897934
Florida	18801310
Georgia	9687653
Hawaii	1360301
Idaho	1567582
Illinois	12830632
Indiana	6483802
Iowa	3046355
Kansas	2853118
Kentucky	4339367
Louisiana	4533372
Maine	1328361
Maryland	5773552
Massachusetts	6547629
Michigan	9883640
Minnesota	5303925
Mississippi	2967297



Missouri	5988927
Montana	989415
Nebraska	1826341
Nevada	2700551
New Hampshire	1316470
New Jersey	8791894
New Mexico	2059179
New York	19378102
North Carolina	9535483
North Dakota	672591
Ohio	11536504
Oklahoma	3751351
Oregon	3831074
Pennsylvania	12702379
Rhode Island	1052567
South Carolina	4625364
South Dakota	814180
Tennessee	6346105
Texas	25145561
Utah	2763885
Vermont	625741
Virginia	8001024
Washington	6724540
West Virginia	1852994
Wisconsin	5686986
Wyoming	563626

Persons under 5 years, percent, July 1, 2016, (V2016) \

State	
Alabama	0.060
Alaska	0.073
Arizona	0.063
Arkansas	0.064
California	0.063
Colorado	0.061
Connecticut	0.052
Delaware	0.058
Florida	0.055
Georgia	0.064
Hawaii	0.064
Idaho	0.068
Illinois	0.060
Indiana	0.064
Iowa	0.064
Kansas	0.067
Kentucky	0.062
Louisiana	0.066
Maine	0.049

Maryland	0.061
Massachusetts	0.053
Michigan	0.058
Minnesota	0.064
Mississippi	0.063
Missouri	0.061
Montana	0.060
Nebraska	0.070
Nevada	0.063
New Hampshire	0.048
New Jersey	0.058
New Mexico	0.062
New York	0.059
North Carolina	0.060
North Dakota	0.073
Ohio	0.060
Oklahoma	0.068
Oregon	0.058
Pennsylvania	0.056
Rhode Island	0.052
South Carolina	0.059
South Dakota	0.071
Tennessee	0.061
Texas	0.072
Utah	0.083
Vermont	0.049
Virginia	0.061
Washington	0.062
West Virginia	0.055
Wisconsin	0.058
Wyoming	0.065

Persons under 5 years, percent, April 1, 2010 \

State	
Alabama	0.064
Alaska	0.076
Arizona	0.071
Arkansas	0.068
California	0.068
Colorado	0.068
Connecticut	0.057
Delaware	0.062
Florida	0.057
Georgia	0.071
Hawaii	0.064
Idaho	0.078
Illinois	0.065
Indiana	0.067

Iowa	0.066
Kansas	0.072
Kentucky	0.065
Louisiana	0.069
Maine	0.052
Maryland	0.063
Massachusetts	0.056
Michigan	0.060
Minnesota	0.067
Mississippi	0.071
Missouri	0.065
Montana	0.063
Nebraska	0.072
Nevada	0.069
New Hampshire	0.053
New Jersey	0.062
New Mexico	0.070
New York	0.060
North Carolina	0.066
North Dakota	0.066
Ohio	0.062
Oklahoma	0.070
Oregon	0.062
Pennsylvania	0.057
Rhode Island	0.055
South Carolina	0.065
South Dakota	0.073
Tennessee	0.064
Texas	0.077
Utah	0.095
Vermont	0.051
Virginia	0.064
Washington	0.065
West Virginia	0.056
Wisconsin	0.063
Wyoming	0.071

Persons under 18 years, percent, July 1, 2016, (V2016) \

State	
Alabama	0.226
Alaska	0.252
Arizona	0.235
Arkansas	0.236
California	0.232
Colorado	0.228
Connecticut	0.211
Delaware	0.215
Florida	0.201

Georgia	0.244
Hawaii	0.216
Idaho	0.260
Illinois	0.229
Indiana	0.238
Iowa	0.233
Kansas	0.246
Kentucky	0.228
Louisiana	0.238
Maine	0.191
Maryland	0.224
Massachusetts	0.202
Michigan	0.221
Minnesota	0.233
Mississippi	0.241
Missouri	0.228
Montana	0.218
Nebraska	0.248
Nevada	0.230
New Hampshire	0.195
New Jersey	0.222
New Mexico	0.236
New York	0.212
North Carolina	0.227
North Dakota	0.233
Ohio	0.225
Oklahoma	0.245
Oregon	0.212
Pennsylvania	0.209
Rhode Island	0.197
South Carolina	0.221
South Dakota	0.246
Tennessee	0.226
Texas	0.262
Utah	0.302
Vermont	0.190
Virginia	0.222
Washington	0.224
West Virginia	0.205
Wisconsin	0.223
Wyoming	0.237

Persons under 18 years, percent, April 1, 2010 \

State	
Alabama	0.237
Alaska	0.264
Arizona	0.255
Arkansas	0.244

California	0.250
Colorado	0.244
Connecticut	0.229
Delaware	0.229
Florida	0.213
Georgia	0.257
Hawaii	0.223
Idaho	0.274
Illinois	0.244
Indiana	0.248
Iowa	0.239
Kansas	0.255
Kentucky	0.236
Louisiana	0.247
Maine	0.207
Maryland	0.234
Massachusetts	0.217
Michigan	0.237
Minnesota	0.242
Mississippi	0.255
Missouri	0.238
Montana	0.226
Nebraska	0.251
Nevada	0.246
New Hampshire	0.218
New Jersey	0.235
New Mexico	0.252
New York	0.223
North Carolina	0.239
North Dakota	0.223
Ohio	0.237
Oklahoma	0.248
Oregon	0.226
Pennsylvania	0.220
Rhode Island	0.213
South Carolina	0.234
South Dakota	0.249
Tennessee	0.236
Texas	0.273
Utah	0.315
Vermont	0.207
Virginia	0.232
Washington	0.235
West Virginia	0.209
Wisconsin	0.236
Wyoming	0.240

Persons 65 years and over, percent, July 1, 2016, (V2016) \

State	
Alabama	0.161
Alaska	0.104
Arizona	0.169
Arkansas	0.163
California	0.136
Colorado	0.134
Connecticut	0.161
Delaware	0.175
Florida	0.199
Georgia	0.131
Hawaii	0.171
Idaho	0.151
Illinois	0.146
Indiana	0.149
Iowa	0.164
Kansas	0.150
Kentucky	0.156
Louisiana	0.144
Maine	0.194
Maryland	0.146
Massachusetts	0.158
Michigan	0.162
Minnesota	0.151
Mississippi	0.151
Missouri	0.161
Montana	0.177
Nebraska	0.150
Nevada	0.150
New Hampshire	0.170
New Jersey	0.153
New Mexico	0.165
New York	0.154
North Carolina	0.155
North Dakota	0.145
Ohio	0.162
Oklahoma	0.150
Oregon	0.168
Pennsylvania	0.174
Rhode Island	0.165
South Carolina	0.167
South Dakota	0.160
Tennessee	0.157
Texas	0.120
Utah	0.105
Vermont	0.181
Virginia	0.146
Washington	0.148

West Virginia	0.188
Wisconsin	0.161
Wyoming	0.150

Persons 65 years and over, percent, April 1, 2010 ... \

State	...
Alabama	0.138 ...
Alaska	0.077 ...
Arizona	0.138 ...
Arkansas	0.144 ...
California	0.114 ...
Colorado	0.109 ...
Connecticut	0.142 ...
Delaware	0.144 ...
Florida	0.173 ...
Georgia	0.107 ...
Hawaii	0.143 ...
Idaho	0.124 ...
Illinois	0.125 ...
Indiana	0.130 ...
Iowa	0.149 ...
Kansas	0.132 ...
Kentucky	0.133 ...
Louisiana	0.123 ...
Maine	0.159 ...
Maryland	0.123 ...
Massachusetts	0.138 ...
Michigan	0.138 ...
Minnesota	0.129 ...
Mississippi	0.128 ...
Missouri	0.140 ...
Montana	0.148 ...
Nebraska	0.135 ...
Nevada	0.120 ...
New Hampshire	0.135 ...
New Jersey	0.135 ...
New Mexico	0.132 ...
New York	0.135 ...
North Carolina	0.129 ...
North Dakota	0.145 ...
Ohio	0.141 ...
Oklahoma	0.135 ...
Oregon	0.139 ...
Pennsylvania	0.154 ...
Rhode Island	0.144 ...
South Carolina	0.137 ...
South Dakota	0.143 ...
Tennessee	0.134 ...

Texas	0.103 ...
Utah	0.090 ...
Vermont	0.146 ...
Virginia	0.122 ...
Washington	0.123 ...
West Virginia	0.160 ...
Wisconsin	0.137 ...
Wyoming	0.124 ...

	All firms, 2012	Men-owned firms, 2012 \
State		
Alabama	374153	203604
Alaska	68032	35402
Arizona	499926	245243
Arkansas	231959	123158
California	3548449	1852580
Colorado	547352	284554
Connecticut	326693	187845
Delaware	73418	38328
Florida	2100187	1084885
Georgia	929864	480578
Hawaii	118454	60409
Idaho	146642	70438
Illinois	1135017	609648
Indiana	479059	253533
Iowa	259121	135382
Kansas	239118	125169
Kentucky	331546	184154
Louisiana	414291	215111
Maine	139570	79243
Maryland	531953	276630
Massachusetts	607664	357158
Michigan	834087	446128
Minnesota	489494	268710
Mississippi	235454	125079
Missouri	491606	257948
Montana	112419	55913
Nebraska	164089	83696
Nevada	227156	114057
New Hampshire	131638	76716
New Jersey	792088	464592
New Mexico	151363	71132
New York	2008988	1139910
North Carolina	805985	435677
North Dakota	68270	37016
Ohio	904814	510078
Oklahoma	327229	174395
Oregon	339305	165691



Pennsylvania	975453	579400
Rhode Island	94642	54651
South Carolina	367726	202446
South Dakota	81314	42418
Tennessee	550453	302249
Texas	2356748	1251696
Utah	251419	132163
Vermont	75827	41270
Virginia	653193	353012
Washington	541522	262650
West Virginia	114435	63112
Wisconsin	432980	236252
Wyoming	62427	30039

	Women-owned firms, 2012	Minority-owned firms, 2012 \
State		
Alabama	137630	92219
Alaska	22141	13688
Arizona	182425	135313
Arkansas	75962	35982
California	1320085	1619857
Colorado	194508	85849
Connecticut	106678	56113
Delaware	23964	14440
Florida	807817	926112
Georgia	376506	371588
Hawaii	44453	74208
Idaho	45121	10592
Illinois	417500	311684
Indiana	162798	61252
Iowa	82345	14707
Kansas	77194	26127
Kentucky	106011	27258
Louisiana	151114	126100
Maine	42067	4339
Maryland	209119	203394
Massachusetts	199210	89967
Michigan	306986	158946
Minnesota	157821	47302
Mississippi	89159	74824
Missouri	162616	61035
Montana	35449	5578
Nebraska	51936	14571
Nevada	82508	71864
New Hampshire	38525	6111
New Jersey	252944	237242
New Mexico	59044	60622
New York	725709	709021

North Carolina	287058	183380
North Dakota	20316	3190
Ohio	306824	122653
Oklahoma	105168	64875
Oregon	123015	41456
Pennsylvania	304803	131512
Rhode Island	30484	14737
South Carolina	131856	83233
South Dakota	23722	4101
Tennessee	195694	105234
Texas	866678	1070392
Utah	76269	24423
Vermont	23417	2354
Virginia	236290	185043
Washington	187677	92807
West Virginia	39065	5777
Wisconsin	133859	40507
Wyoming	19344	4077

	Nonminority-owned firms, 2012	Veteran-owned firms, 2012 \
State		
Alabama	272651	41943
Alaska	51147	7953
Arizona	344981	46780
Arkansas	189029	25915
California	1819107	252377
Colorado	442365	51722
Connecticut	259614	31056
Delaware	54782	7206
Florida	1121749	185756
Georgia	538893	96787
Hawaii	38510	11148
Idaho	130973	12804
Illinois	795129	89110
Indiana	405090	45174
Iowa	236561	25889
Kansas	204562	21610
Kentucky	296155	33208
Louisiana	277676	42211
Maine	131322	15689
Maryland	314902	50976
Massachusetts	499959	58339
Michigan	657237	71861
Minnesota	428716	45582
Mississippi	155094	26789
Missouri	415972	49217
Montana	102746	11486
Nebraska	144122	16693

Nevada	144944	23049
New Hampshire	121297	16119
New Jersey	533808	57996
New Mexico	83857	14096
New York	1248304	137532
North Carolina	603182	86571
North Dakota	62271	6584
Ohio	759569	91316
Oklahoma	249027	36273
Oregon	285028	30918
Pennsylvania	818858	97969
Rhode Island	77042	9904
South Carolina	276269	47987
South Dakota	74228	8604
Tennessee	434025	59379
Texas	1224845	213590
Utah	218826	18754
Vermont	70491	8237
Virginia	450109	76434
Washington	426697	49331
West Virginia	104785	12912
Wisconsin	379934	39830
Wyoming	55397	6470

Nonveteran-owned firms, 2012    Population per square mile, 2010    \

State		
Alabama	316984	94.4
Alaska	56091	1.2
Arizona	427582	56.3
Arkansas	192988	56.0
California	3176341	239.1
Colorado	469524	48.5
Connecticut	281182	738.1
Delaware	60318	460.8
Florida	1846686	350.6
Georgia	800585	168.4
Hawaii	101334	211.8
Idaho	124314	19.0
Illinois	1006885	231.1
Indiana	412543	181.0
Iowa	217779	54.5
Kansas	203401	34.9
Kentucky	282704	109.9
Louisiana	354460	104.9
Maine	115839	43.1
Maryland	462232	594.8
Massachusetts	525667	839.4
Michigan	733517	174.8

Minnesota	419628	66.6
Mississippi	198566	63.2
Missouri	415542	87.1
Montana	93393	6.8
Nebraska	137254	23.8
Nevada	191293	24.6
New Hampshire	108817	147.0
New Jersey	707975	1195.5
New Mexico	128693	17.0
New York	1811544	411.2
North Carolina	684743	196.1
North Dakota	56904	9.7
Ohio	776193	282.3
Oklahoma	273491	54.7
Oregon	288790	39.9
Pennsylvania	837792	283.9
Rhode Island	79837	1018.1
South Carolina	303137	153.9
South Dakota	66219	10.7
Tennessee	469392	153.9
Texas	2057218	96.3
Utah	219807	33.6
Vermont	63317	67.9
Virginia	548439	202.6
Washington	461401	101.2
West Virginia	94960	77.1
Wisconsin	370755	105.0
Wyoming	51353	5.8

Land area in square miles, 2010 FIPS Code

State		
Alabama	50645.33	"01"
Alaska	570640.95	"02"
Arizona	113594.08	"04"
Arkansas	52035.48	"05"
California	155779.22	"06"
Colorado	103641.89	"08"
Connecticut	4842.36	"09"
Delaware	1948.54	"10"
Florida	53624.76	"12"
Georgia	57513.49	"13"
Hawaii	6422.63	"15"
Idaho	82643.12	"16"
Illinois	55518.93	"17"
Indiana	35826.11	"18"
Iowa	55857.13	"19"
Kansas	81758.72	"20"
Kentucky	39486.34	"21"

Louisiana	43203.90	"22"
Maine	30842.92	"23"
Maryland	9707.24	"24"
Massachusetts	7800.06	"25"
Michigan	56538.90	"26"
Minnesota	79626.74	"27"
Mississippi	46923.27	"28"
Missouri	68741.52	"29"
Montana	145545.80	"30"
Nebraska	76824.17	"31"
Nevada	109781.18	"32"
New Hampshire	8952.65	"33"
New Jersey	7354.22	"34"
New Mexico	121298.15	"35"
New York	47126.40	"36"
North Carolina	48617.91	"37"
North Dakota	69000.80	"38"
Ohio	40860.69	"39"
Oklahoma	68594.92	"40"
Oregon	95988.01	"41"
Pennsylvania	44742.70	"42"
Rhode Island	1033.81	"44"
South Carolina	30060.70	"45"
South Dakota	75811.00	"46"
Tennessee	41234.90	"47"
Texas	261231.71	"48"
Utah	82169.62	"49"
Vermont	9216.66	"50"
Virginia	39490.09	"51"
Washington	66455.52	"53"
West Virginia	24038.21	"54"
Wisconsin	54157.80	"55"
Wyoming	97093.14	"56"

[50 rows x 65 columns]

```
In [424]: #Inspecting the data and getting information on the dimensions of the dataframes
display("Census dimensions: " +str(df_census.shape))
display("Gun Dimensions:" + str(df_gun.shape))
```

'Census dimensions: (50, 65)'

'Gun Dimensions:(12485, 27)'

```
In [425]: #Printing concise summary of both the dataframes helping in knowing the datatypes and
df_census.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
Index: 50 entries, Alabama to Wyoming
```

```
Data columns (total 65 columns):
```

#	Column	Non
0	Population estimates, July 1, 2016, (V2016)	50
1	Population estimates base, April 1, 2010, (V2016)	50
2	Population, percent change - April 1, 2010 (estimates base) to July 1, 2016, (V2016)	50
3	Population, Census, April 1, 2010	50
4	Persons under 5 years, percent, July 1, 2016, (V2016)	50
5	Persons under 5 years, percent, April 1, 2010	50
6	Persons under 18 years, percent, July 1, 2016, (V2016)	50
7	Persons under 18 years, percent, April 1, 2010	50
8	Persons 65 years and over, percent, July 1, 2016, (V2016)	50
9	Persons 65 years and over, percent, April 1, 2010	50
10	Female persons, percent, July 1, 2016, (V2016)	50
11	Female persons, percent, April 1, 2010	50
12	White alone, percent, July 1, 2016, (V2016)	50
13	Black or African American alone, percent, July 1, 2016, (V2016)	50
14	American Indian and Alaska Native alone, percent, July 1, 2016, (V2016)	50
15	Asian alone, percent, July 1, 2016, (V2016)	50
16	Native Hawaiian and Other Pacific Islander alone, percent, July 1, 2016, (V2016)	50
17	Two or More Races, percent, July 1, 2016, (V2016)	50
18	Hispanic or Latino, percent, July 1, 2016, (V2016)	50
19	White alone, not Hispanic or Latino, percent, July 1, 2016, (V2016)	50
20	Veterans, 2011-2015	50
21	Foreign born persons, percent, 2011-2015	50
22	Housing units, July 1, 2016, (V2016)	50
23	Housing units, April 1, 2010	50
24	Owner-occupied housing unit rate, 2011-2015	50
25	Median value of owner-occupied housing units, 2011-2015	50
26	Median selected monthly owner costs -with a mortgage, 2011-2015	50
27	Median selected monthly owner costs -without a mortgage, 2011-2015	50
28	Median gross rent, 2011-2015	50
29	Building permits, 2016	50
30	Households, 2011-2015	50
31	Persons per household, 2011-2015	50
32	Living in same house 1 year ago, percent of persons age 1 year+, 2011-2015	50
33	Language other than English spoken at home, percent of persons age 5 years+, 2011-2015	50
34	High school graduate or higher, percent of persons age 25 years+, 2011-2015	50
35	Bachelor's degree or higher, percent of persons age 25 years+, 2011-2015	50
36	With a disability, under age 65 years, percent, 2011-2015	50
37	Persons without health insurance, under age 65 years, percent	50
38	In civilian labor force, total, percent of population age 16 years+, 2011-2015	50
39	In civilian labor force, female, percent of population age 16 years+, 2011-2015	50
40	Total accommodation and food services sales, 2012 (\$1,000)	50
41	Total health care and social assistance receipts/revenue, 2012 (\$1,000)	50
42	Total manufacturers shipments, 2012 (\$1,000)	50

```

43 Total merchant wholesaler sales, 2012 ($1,000) 50
44 Total retail sales, 2012 ($1,000) 50
45 Total retail sales per capita, 2012 50
46 Mean travel time to work (minutes), workers age 16 years+, 2011-2015 50
47 Median household income (in 2015 dollars), 2011-2015 50
48 Per capita income in past 12 months (in 2015 dollars), 2011-2015 50
49 Persons in poverty, percent 50
50 Total employer establishments, 2015 50
51 Total employment, 2015 50
52 Total annual payroll, 2015 ($1,000) 50
53 Total employment, percent change, 2014-2015 50
54 Total nonemployer establishments, 2015 50
55 All firms, 2012 50
56 Men-owned firms, 2012 50
57 Women-owned firms, 2012 50
58 Minority-owned firms, 2012 50
59 Nonminority-owned firms, 2012 50
60 Veteran-owned firms, 2012 50
61 Nonveteran-owned firms, 2012 50
62 Population per square mile, 2010 50
63 Land area in square miles, 2010 50
64 FIPS Code 50
dtypes: float64(31), int64(30), object(4)
memory usage: 25.8+ KB

```

In [426]: *#Printing concise summary of both the dataframes helping in knowing the datatypes and*  
df\_gun.info();

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12485 entries, 0 to 12484
Data columns (total 27 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   month                                12485 non-null  object
1   state                                12485 non-null  object
2   permit                                12461 non-null  float64
3   permit_recheck                        1100 non-null   float64
4   handgun                               12465 non-null  float64
5   long_gun                             12466 non-null  float64
6   other                                 5500 non-null   float64
7   multiple                             12485 non-null  int64
8   admin                                12462 non-null  float64
9   prepawn_handgun                       10542 non-null  float64
10  prepawn_long_gun                      10540 non-null  float64
11  prepawn_other                          5115 non-null   float64
12  redemption_handgun                    10545 non-null  float64
13  redemption_long_gun                   10544 non-null  float64

```

```

14 redemption_other          5115 non-null    float64
15 returned_handgun           2200 non-null    float64
16 returned_long_gun          2145 non-null    float64
17 returned_other             1815 non-null    float64
18 rentals_handgun            990 non-null     float64
19 rentals_long_gun           825 non-null     float64
20 private_sale_handgun        2750 non-null    float64
21 private_sale_long_gun       2750 non-null    float64
22 private_sale_other          2750 non-null    float64
23 return_to_seller_handgun    2475 non-null    float64
24 return_to_seller_long_gun   2750 non-null    float64
25 return_to_seller_other      2255 non-null    float64
26 totals                     12485 non-null   int64
dtypes: float64(23), int64(2), object(2)
memory usage: 2.6+ MB

```

```

In [427]: #Giving a general descriptive statistics of the dataframe
          df_census.describe()

```

```

Out[427]:      Population estimates, July 1, 2016, (V2016) \
count          5.000000e+01
mean           6.448927e+06
std            7.271769e+06
min            5.855010e+05
25%            1.850106e+06
50%            4.559320e+06
75%            7.198768e+06
max            3.925002e+07

```

```

      Population estimates base, April 1, 2010, (V2016) \
count          5.000000e+01
mean           6.163127e+06
std            6.848463e+06
min            5.637670e+05
25%            1.833003e+06
50%            4.436412e+06
75%            6.680362e+06
max            3.725452e+07

```

```

      Population, percent change - April 1, 2010 (estimates base) to July 1, 2016, (
count          50.000000
mean           0.041800
std            0.033811
min           -0.012000
25%            0.016250
50%            0.040500
75%            0.063750

```



max	0.127000
-----	----------

Population, Census, April 1, 2010 \

count	5.000000e+01
mean	6.162876e+06
std	6.848235e+06
min	5.636260e+05
25%	1.833004e+06
50%	4.436370e+06
75%	6.680312e+06
max	3.725396e+07

Persons under 5 years, percent, July 1, 2016, (V2016) \

count	50.000000
mean	0.061600
std	0.006612
min	0.048000
25%	0.058000
50%	0.061000
75%	0.064000
max	0.083000

Persons under 5 years, percent, April 1, 2010 \

count	50.000000
mean	0.065460
std	0.007579
min	0.051000
25%	0.062000
50%	0.065000
75%	0.069750
max	0.095000

Persons under 18 years, percent, July 1, 2016, (V2016) \

count	50.000000
mean	0.22750
std	0.01977
min	0.19000
25%	0.21650
50%	0.22750
75%	0.23675
max	0.30200

Persons under 18 years, percent, April 1, 2010 \

count	50.000000
mean	0.238820
std	0.019075
min	0.207000
25%	0.226750

50%	0.237000
75%	0.248000
max	0.315000

	Persons 65 years and over, percent, July 1, 2016, (V2016) \
count	50.000000
mean	0.156160
std	0.018321
min	0.104000
25%	0.149250
50%	0.156500
75%	0.165000
max	0.199000

	Persons 65 years and over, percent, April 1, 2010 ... \
count	50.00000 ...
mean	0.13292 ...
std	0.01664 ...
min	0.07700 ...
25%	0.12425 ...
50%	0.13500 ...
75%	0.14275 ...
max	0.17300 ...

	Total nonemployer establishments, 2015	All firms, 2012 \
count	5.000000e+01	5.000000e+01
mean	4.855009e+05	5.548918e+05
std	6.076460e+05	6.645229e+05
min	4.814000e+04	6.242700e+04
25%	1.215145e+05	1.478222e+05
50%	3.014300e+05	3.535155e+05
75%	5.096492e+05	5.933612e+05
max	3.206958e+06	3.548449e+06

	Men-owned firms, 2012	Women-owned firms, 2012 \
count	5.000000e+01	5.000000e+01
mean	2.976230e+05	1.971777e+05
std	3.540115e+05	2.482425e+05
min	3.003900e+04	1.934400e+04
25%	7.734775e+04	4.682475e+04
50%	1.951455e+05	1.274355e+05
75%	3.403212e+05	2.066418e+05
max	1.852580e+06	1.320085e+06

	Minority-owned firms, 2012	Nonminority-owned firms, 2012 \
count	5.000000e+01	5.000000e+01
mean	1.585531e+05	3.806367e+05
std	3.017390e+05	3.618627e+05

min	2.354000e+03	3.851000e+04
25%	1.471450e+04	1.310602e+05
50%	6.306350e+04	2.769725e+05
75%	1.301590e+05	4.481730e+05
max	1.619857e+06	1.819107e+06

	Veteran-owned firms, 2012	Nonveteran-owned firms, 2012 \
count	50.00000	5.000000e+01
mean	50502.72000	4.817082e+05
std	52073.87529	5.921109e+05
min	6470.00000	5.135300e+04
25%	15796.50000	1.254088e+05
50%	38051.50000	2.959635e+05
75%	58253.25000	5.116312e+05
max	252377.00000	3.176341e+06

	Population per square mile, 2010	Land area in square miles, 2010
count	50.000000	50.000000
mean	194.962000	70636.887800
std	261.091465	85815.678218
min	1.200000	1033.810000
25%	44.450000	36741.167500
50%	98.750000	53891.280000
75%	209.500000	81225.725000
max	1195.500000	570640.950000

[8 rows x 61 columns]

In [428]: *#Giving a general descriptive statistics of the dataframe*  
df\_gun.describe()

Out[428]:

	permit	permit_recheck	handgun	long_gun \
count	12461.000000	1100.000000	12465.000000	12466.000000
mean	6413.629404	1165.956364	5940.881107	7810.847585
std	23752.338269	9224.200609	8618.584060	9309.846140
min	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	865.000000	2078.250000
50%	518.000000	0.000000	3059.000000	5122.000000
75%	4272.000000	0.000000	7280.000000	10380.750000
max	522188.000000	116681.000000	107224.000000	108058.000000

	other	multiple	admin	prepawn_handgun \
count	5500.000000	12485.000000	12462.000000	10542.000000
mean	360.471636	268.603364	58.898090	4.828021
std	1349.478273	783.185073	604.814818	10.907756
min	0.000000	0.000000	0.000000	0.000000
25%	17.000000	15.000000	0.000000	0.000000
50%	121.000000	125.000000	0.000000	0.000000

75%	354.000000	301.000000	0.000000	5.000000
max	77929.000000	38907.000000	28083.000000	164.000000

	prepawn_long_gun	prepawn_other	...	returned_other	rentals_handgun \
count	10540.000000	5115.000000	...	1815.000000	990.000000
mean	7.834156	0.165591	...	1.027548	0.076768
std	16.468028	1.057105	...	4.386296	0.634503
min	0.000000	0.000000	...	0.000000	0.000000
25%	0.000000	0.000000	...	0.000000	0.000000
50%	1.000000	0.000000	...	0.000000	0.000000
75%	8.000000	0.000000	...	0.000000	0.000000
max	269.000000	49.000000	...	64.000000	12.000000

	rentals_long_gun	private_sale_handgun	private_sale_long_gun \
count	825.000000	2750.000000	2750.000000
mean	0.087273	14.936000	11.602909
std	0.671649	71.216021	54.253090
min	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000
50%	0.000000	0.000000	0.000000
75%	0.000000	2.000000	4.000000
max	12.000000	1017.000000	777.000000

	private_sale_other	return_to_seller_handgun \
count	2750.000000	2475.000000
mean	1.030182	0.402020
std	4.467843	1.446568
min	0.000000	0.000000
25%	0.000000	0.000000
50%	0.000000	0.000000
75%	0.000000	0.000000
max	71.000000	28.000000

	return_to_seller_long_gun	return_to_seller_other	totals
count	2750.000000	2255.000000	12485.000000
mean	0.441818	0.105987	21595.725911
std	1.528223	0.427363	32591.418387
min	0.000000	0.000000	0.000000
25%	0.000000	0.000000	4638.000000
50%	0.000000	0.000000	12399.000000
75%	0.000000	0.000000	25453.000000
max	17.000000	4.000000	541978.000000

[8 rows x 25 columns]

```
In [429]: #Looking into the column names in the dataframe
for col in df_census.columns:
    print(col)
```

Population estimates, July 1, 2016, (V2016)  
 Population estimates base, April 1, 2010, (V2016)  
 Population, percent change - April 1, 2010 (estimates base) to July 1, 2016, (V2016)  
 Population, Census, April 1, 2010  
 Persons under 5 years, percent, July 1, 2016, (V2016)  
 Persons under 5 years, percent, April 1, 2010  
 Persons under 18 years, percent, July 1, 2016, (V2016)  
 Persons under 18 years, percent, April 1, 2010  
 Persons 65 years and over, percent, July 1, 2016, (V2016)  
 Persons 65 years and over, percent, April 1, 2010  
 Female persons, percent, July 1, 2016, (V2016)  
 Female persons, percent, April 1, 2010  
 White alone, percent, July 1, 2016, (V2016)  
 Black or African American alone, percent, July 1, 2016, (V2016)  
 American Indian and Alaska Native alone, percent, July 1, 2016, (V2016)  
 Asian alone, percent, July 1, 2016, (V2016)  
 Native Hawaiian and Other Pacific Islander alone, percent, July 1, 2016, (V2016)  
 Two or More Races, percent, July 1, 2016, (V2016)  
 Hispanic or Latino, percent, July 1, 2016, (V2016)  
 White alone, not Hispanic or Latino, percent, July 1, 2016, (V2016)  
 Veterans, 2011-2015  
 Foreign born persons, percent, 2011-2015  
 Housing units, July 1, 2016, (V2016)  
 Housing units, April 1, 2010  
 Owner-occupied housing unit rate, 2011-2015  
 Median value of owner-occupied housing units, 2011-2015  
 Median selected monthly owner costs -with a mortgage, 2011-2015  
 Median selected monthly owner costs -without a mortgage, 2011-2015  
 Median gross rent, 2011-2015  
 Building permits, 2016  
 Households, 2011-2015  
 Persons per household, 2011-2015  
 Living in same house 1 year ago, percent of persons age 1 year+, 2011-2015  
 Language other than English spoken at home, percent of persons age 5 years+, 2011-2015  
 High school graduate or higher, percent of persons age 25 years+, 2011-2015  
 Bachelor's degree or higher, percent of persons age 25 years+, 2011-2015  
 With a disability, under age 65 years, percent, 2011-2015  
 Persons without health insurance, under age 65 years, percent  
 In civilian labor force, total, percent of population age 16 years+, 2011-2015  
 In civilian labor force, female, percent of population age 16 years+, 2011-2015  
 Total accommodation and food services sales, 2012 (\$1,000)  
 Total health care and social assistance receipts/revenue, 2012 (\$1,000)  
 Total manufacturers shipments, 2012 (\$1,000)  
 Total merchant wholesaler sales, 2012 (\$1,000)  
 Total retail sales, 2012 (\$1,000)  
 Total retail sales per capita, 2012  
 Mean travel time to work (minutes), workers age 16 years+, 2011-2015  
 Median household income (in 2015 dollars), 2011-2015

Per capita income in past 12 months (in 2015 dollars), 2011-2015  
Persons in poverty, percent  
Total employer establishments, 2015  
Total employment, 2015  
Total annual payroll, 2015 (\$1,000)  
Total employment, percent change, 2014-2015  
Total nonemployer establishments, 2015  
All firms, 2012  
Men-owned firms, 2012  
Women-owned firms, 2012  
Minority-owned firms, 2012  
Nonminority-owned firms, 2012  
Veteran-owned firms, 2012  
Nonveteran-owned firms, 2012  
Population per square mile, 2010  
Land area in square miles, 2010  
FIPS Code

```
In [430]: #looking into the column names for the gun dataframe
         for col in df_gun.columns:
             print(col)
```

month  
state  
permit  
permit\_recheck  
handgun  
long\_gun  
other  
multiple  
admin  
prepawn\_handgun  
prepawn\_long\_gun  
prepawn\_other  
redemption\_handgun  
redemption\_long\_gun  
redemption\_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\_long\_gun

```
return_to_seller_other
totals
```

After manually analyzing both the datasets carefully in Excel, data for a few of the states is not very consistent in the datasets we are considering. Some of the states aren't present in the Gun data as compared to the Census data. The states include: Guam, District of Columbia, Puerto Rico, Mariana Islands, Virgin Islands. We can eliminate the data so as to avoid any confusions in the analysis.

```
In [431]: #Removing all the unnecessary state data from the guns data set
          print(df_gun.state.nunique())
```

```
df_gun = df_gun[df_gun.state != 'Guam']
df_gun = df_gun[df_gun.state != 'District of Columbia']
df_gun = df_gun[df_gun.state != 'Puerto Rico']
df_gun = df_gun[df_gun.state != 'Mariana Islands']
df_gun = df_gun[df_gun.state != 'Virgin Islands']
```

55

```
In [432]: df_census.dtypes
```

```
Out[432]: Population estimates, July 1, 2016, (V2016)
Population estimates base, April 1, 2010, (V2016)
Population, percent change - April 1, 2010 (estimates base) to July 1, 2016, (V2016)
Population, Census, April 1, 2010
Persons under 5 years, percent, July 1, 2016, (V2016)

Veteran-owned firms, 2012
Nonveteran-owned firms, 2012
Population per square mile, 2010
Land area in square miles, 2010
FIPS Code
Length: 65, dtype: object
```

```
In [433]: df_gun.dtypes
```

```
Out[433]: month                object
state                object
permit              float64
permit_recheck      float64
handgun             float64
long_gun            float64
other               float64
multiple            int64
admin              float64
prepawn_handgun     float64
```

```

prepawn_long_gun          float64
prepawn_other              float64
redemption_handgun        float64
redemption_long_gun       float64
redemption_other          float64
returned_handgun          float64
returned_long_gun         float64
returned_other            float64
rentals_handgun           float64
rentals_long_gun          float64
private_sale_handgun       float64
private_sale_long_gun     float64
private_sale_other        float64
return_to_seller_handgun  float64
return_to_seller_long_gun float64
return_to_seller_other    float64
totals                    int64
dtype: object

```

```

In [434]: #Check for duplicates in the Guns data
          sum(df_gun.duplicated())

```

```

Out[434]: 0

```

```

In [435]: #Check for duplicates in the Census data
          sum(df_census.duplicated())

```

```

Out[435]: 0

```

```

In [436]: #Checking for NaN values
          df_census.isnull().any().any(), sum(df_census.isnull().any())

```

```

Out[436]: (False, 0)

```

```

In [437]: df_census.isnull().any()

```

```

Out[437]: Population estimates, July 1, 2016, (V2016)
Population estimates base, April 1, 2010, (V2016)
Population, percent change - April 1, 2010 (estimates base) to July 1, 2016, (V2016)
Population, Census, April 1, 2010
Persons under 5 years, percent, July 1, 2016, (V2016)

Veteran-owned firms, 2012
Nonveteran-owned firms, 2012
Population per square mile, 2010
Land area in square miles, 2010
FIPS Code
Length: 65, dtype: bool

```

```

In [438]: df_gun.isnull().any().any(), sum(df_gun.isnull().any())

```



```
Out[438]: (True, 21)
```

```
In [439]: df_gun.isnull().any()
```

```
Out[439]: month                False
          state                False
          permit               True
          permit_recheck       True
          handgun              False
          long_gun             False
          other                True
          multiple             False
          admin                True
          prepawn_handgun       True
          prepawn_long_gun      True
          prepawn_other         True
          redemption_handgun    True
          redemption_long_gun   True
          redemption_other      True
          returned_handgun      True
          returned_long_gun     True
          returned_other        True
          rentals_handgun       True
          rentals_long_gun      True
          private_sale_handgun   True
          private_sale_long_gun  True
          private_sale_other     True
          return_to_seller_handgun True
          return_to_seller_long_gun True
          return_to_seller_other True
          totals                False
          dtype: bool
```

```
In [440]: df_census.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
Index: 50 entries, Alabama to Wyoming
```

```
Data columns (total 65 columns):
```

#	Column	Non
0	Population estimates, July 1, 2016, (V2016)	50
1	Population estimates base, April 1, 2010, (V2016)	50
2	Population, percent change - April 1, 2010 (estimates base) to July 1, 2016, (V2016)	50
3	Population, Census, April 1, 2010	50
4	Persons under 5 years, percent, July 1, 2016, (V2016)	50
5	Persons under 5 years, percent, April 1, 2010	50
6	Persons under 18 years, percent, July 1, 2016, (V2016)	50
7	Persons under 18 years, percent, April 1, 2010	50
8	Persons 65 years and over, percent, July 1, 2016, (V2016)	50

9	Persons 65 years and over, percent, April 1, 2010	50
10	Female persons, percent, July 1, 2016, (V2016)	50
11	Female persons, percent, April 1, 2010	50
12	White alone, percent, July 1, 2016, (V2016)	50
13	Black or African American alone, percent, July 1, 2016, (V2016)	50
14	American Indian and Alaska Native alone, percent, July 1, 2016, (V2016)	50
15	Asian alone, percent, July 1, 2016, (V2016)	50
16	Native Hawaiian and Other Pacific Islander alone, percent, July 1, 2016, (V2016)	50
17	Two or More Races, percent, July 1, 2016, (V2016)	50
18	Hispanic or Latino, percent, July 1, 2016, (V2016)	50
19	White alone, not Hispanic or Latino, percent, July 1, 2016, (V2016)	50
20	Veterans, 2011-2015	50
21	Foreign born persons, percent, 2011-2015	50
22	Housing units, July 1, 2016, (V2016)	50
23	Housing units, April 1, 2010	50
24	Owner-occupied housing unit rate, 2011-2015	50
25	Median value of owner-occupied housing units, 2011-2015	50
26	Median selected monthly owner costs -with a mortgage, 2011-2015	50
27	Median selected monthly owner costs -without a mortgage, 2011-2015	50
28	Median gross rent, 2011-2015	50
29	Building permits, 2016	50
30	Households, 2011-2015	50
31	Persons per household, 2011-2015	50
32	Living in same house 1 year ago, percent of persons age 1 year+, 2011-2015	50
33	Language other than English spoken at home, percent of persons age 5 years+, 2011-2015	50
34	High school graduate or higher, percent of persons age 25 years+, 2011-2015	50
35	Bachelor's degree or higher, percent of persons age 25 years+, 2011-2015	50
36	With a disability, under age 65 years, percent, 2011-2015	50
37	Persons without health insurance, under age 65 years, percent	50
38	In civilian labor force, total, percent of population age 16 years+, 2011-2015	50
39	In civilian labor force, female, percent of population age 16 years+, 2011-2015	50
40	Total accommodation and food services sales, 2012 (\$1,000)	50
41	Total health care and social assistance receipts/revenue, 2012 (\$1,000)	50
42	Total manufacturers shipments, 2012 (\$1,000)	50
43	Total merchant wholesaler sales, 2012 (\$1,000)	50
44	Total retail sales, 2012 (\$1,000)	50
45	Total retail sales per capita, 2012	50
46	Mean travel time to work (minutes), workers age 16 years+, 2011-2015	50
47	Median household income (in 2015 dollars), 2011-2015	50
48	Per capita income in past 12 months (in 2015 dollars), 2011-2015	50
49	Persons in poverty, percent	50
50	Total employer establishments, 2015	50
51	Total employment, 2015	50
52	Total annual payroll, 2015 (\$1,000)	50
53	Total employment, percent change, 2014-2015	50
54	Total nonemployer establishments, 2015	50
55	All firms, 2012	50
56	Men-owned firms, 2012	50

```

57 Women-owned firms, 2012 50
58 Minority-owned firms, 2012 50
59 Nonminority-owned firms, 2012 50
60 Veteran-owned firms, 2012 50
61 Nonveteran-owned firms, 2012 50
62 Population per square mile, 2010 50
63 Land area in square miles, 2010 50
64 FIPS Code 50
dtypes: float64(31), int64(30), object(4)
memory usage: 25.8+ KB

```

```
In [441]: df_gun.info()
```

```

<class 'pandas.core.frame.DataFrame'>
Int64Index: 11350 entries, 0 to 12484
Data columns (total 27 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   month                                11350 non-null  object
1   state                                11350 non-null  object
2   permit                               11348 non-null  float64
3   permit_recheck                       1000 non-null   float64
4   handgun                              11350 non-null  float64
5   long_gun                             11350 non-null  float64
6   other                                5000 non-null   float64
7   multiple                             11350 non-null  int64
8   admin                                11348 non-null  float64
9   prepawn_handgun                      9597 non-null   float64
10  prepawn_long_gun                     9595 non-null   float64
11  prepawn_other                         4650 non-null   float64
12  redemption_handgun                   9600 non-null   float64
13  redemption_long_gun                  9598 non-null   float64
14  redemption_other                     4650 non-null   float64
15  returned_handgun                     2000 non-null   float64
16  returned_long_gun                    1950 non-null   float64
17  returned_other                       1650 non-null   float64
18  rentals_handgun                      900 non-null    float64
19  rentals_long_gun                     750 non-null    float64
20  private_sale_handgun                 2500 non-null   float64
21  private_sale_long_gun                 2500 non-null   float64
22  private_sale_other                   2500 non-null   float64
23  return_to_seller_handgun              2250 non-null   float64
24  return_to_seller_long_gun             2500 non-null   float64
25  return_to_seller_other                2050 non-null   float64
26  totals                               11350 non-null  int64
dtypes: float64(23), int64(2), object(2)
memory usage: 2.4+ MB

```

### 1.1.2 Data Cleaning

```
In [442]: # After discussing the structure of the data and any problems that need to be
#         cleaned, perform those cleaning steps in the second part of this section.
```

```
# We are starting the cleaning process by checking the duplicates and getting rid of them
df_census.drop_duplicates(inplace = True)
sum(df_census.duplicated())
```

```
Out[442]: 0
```

```
In [443]: df_gun.drop_duplicates(inplace = True)
sum(df_gun.duplicated())
```

```
Out[443]: 0
```

```
In [444]: #Now we will fill all the empty / NA values using the fillna function
df_census.fillna('No Records', inplace = True)
df_census.isnull().any()
```

```
Out[444]: Population estimates, July 1, 2016, (V2016)
Population estimates base, April 1, 2010, (V2016)
Population, percent change - April 1, 2010 (estimates base) to July 1, 2016, (V2016)
Population, Census, April 1, 2010
Persons under 5 years, percent, July 1, 2016, (V2016)
```

```
Veteran-owned firms, 2012
Nonveteran-owned firms, 2012
Population per square mile, 2010
Land area in square miles, 2010
FIPS Code
Length: 65, dtype: bool
```

```
In [445]: df_gun.fillna(df_gun.mean(), inplace = True )
df_gun.isnull().any()
```

```
Out[445]: month                False
state                False
permit              False
permit_recheck      False
handgun             False
long_gun            False
other               False
multiple            False
admin               False
prepawn_handgun     False
prepawn_long_gun    False
prepawn_other       False
redemption_handgun  False
```

```

redemption_long_gun      False
redemption_other          False
returned_handgun          False
returned_long_gun         False
returned_other            False
rentals_handgun           False
rentals_long_gun          False
private_sale_handgun       False
private_sale_long_gun      False
private_sale_other         False
return_to_seller_handgun   False
return_to_seller_long_gun  False
return_to_seller_other     False
totals                    False
dtype: bool

```

```
In [446]: df_gun.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
Int64Index: 11350 entries, 0 to 12484
```

```
Data columns (total 27 columns):
```

#	Column	Non-Null Count	Dtype
0	month	11350 non-null	object
1	state	11350 non-null	object
2	permit	11350 non-null	float64
3	permit_recheck	11350 non-null	float64
4	handgun	11350 non-null	float64
5	long_gun	11350 non-null	float64
6	other	11350 non-null	float64
7	multiple	11350 non-null	int64
8	admin	11350 non-null	float64
9	prepawn_handgun	11350 non-null	float64
10	prepawn_long_gun	11350 non-null	float64
11	prepawn_other	11350 non-null	float64
12	redemption_handgun	11350 non-null	float64
13	redemption_long_gun	11350 non-null	float64
14	redemption_other	11350 non-null	float64
15	returned_handgun	11350 non-null	float64
16	returned_long_gun	11350 non-null	float64
17	returned_other	11350 non-null	float64
18	rentals_handgun	11350 non-null	float64
19	rentals_long_gun	11350 non-null	float64
20	private_sale_handgun	11350 non-null	float64
21	private_sale_long_gun	11350 non-null	float64
22	private_sale_other	11350 non-null	float64
23	return_to_seller_handgun	11350 non-null	float64
24	return_to_seller_long_gun	11350 non-null	float64

```

25 return_to_seller_other    11350 non-null float64
26 totals                    11350 non-null int64
dtypes: float64(23), int64(2), object(2)
memory usage: 2.4+ MB

```

In [447]: *#Noticing the names of the column names, we can bring the column names to a single form*

```

df_census.rename(columns = lambda x:x.lower(), inplace = True)
df_census

```

```

Out[447]:
              population estimates, july 1, 2016, (v2016) \
State
Alabama                                                    4863300
Alaska                                                       741894
Arizona                                                     6931071
Arkansas                                                    2988248
California                                                  39250017
Colorado                                                    5540545
Connecticut                                                 3576452
Delaware                                                    952065
Florida                                                     20612439
Georgia                                                     10310371
Hawaii                                                      1428557
Idaho                                                       1683140
Illinois                                                    12801539
Indiana                                                     6633053
Iowa                                                        3134693
Kansas                                                      2907289
Kentucky                                                    4436974
Louisiana                                                   4681666
Maine                                                       1331479
Maryland                                                    6016447
Massachusetts                                               6811779
Michigan                                                    9928300
Minnesota                                                   5519952
Mississippi                                                 2988726
Missouri                                                    6093000
Montana                                                     1042520
Nebraska                                                    1907116
Nevada                                                      2940058
New Hampshire                                              1334795
New Jersey                                                  8944469
New Mexico                                                  2081015
New York                                                    19745289
North Carolina                                             10146788
North Dakota                                               757952
Ohio                                                       11614373

```

Oklahoma	3923561
Oregon	4093465
Pennsylvania	12784227
Rhode Island	1056426
South Carolina	4961119
South Dakota	865454
Tennessee	6651194
Texas	27862596
Utah	3051217
Vermont	624594
Virginia	8411808
Washington	7288000
West Virginia	1831102
Wisconsin	5778708
Wyoming	585501

population estimates base, april 1, 2010, (v2016) \

State	
Alabama	4780131
Alaska	710249
Arizona	6392301
Arkansas	2916025
California	37254522
Colorado	5029324
Connecticut	3574114
Delaware	897936
Florida	18804592
Georgia	9688680
Hawaii	1360301
Idaho	1567650
Illinois	12831574
Indiana	6484136
Iowa	3046869
Kansas	2853129
Kentucky	4339344
Louisiana	4533479
Maine	1328364
Maryland	5773786
Massachusetts	6547813
Michigan	9884129
Minnesota	5303924
Mississippi	2968103
Missouri	5988928
Montana	989414
Nebraska	1826334
Nevada	2700691
New Hampshire	1316461
New Jersey	8791953

New Mexico	2059198
New York	19378110
North Carolina	9535688
North Dakota	672591
Ohio	11536727
Oklahoma	3751615
Oregon	3831072
Pennsylvania	12702857
Rhode Island	1052940
South Carolina	4625410
South Dakota	814195
Tennessee	6346298
Texas	25146100
Utah	2763888
Vermont	625741
Virginia	8001041
Washington	6724545
West Virginia	1853011
Wisconsin	5687289
Wyoming	563767

population, percent change - april 1, 2010 (estimates base) to july 1,

State	
Alabama	0.017
Alaska	0.045
Arizona	0.084
Arkansas	0.025
California	0.054
Colorado	0.102
Connecticut	0.001
Delaware	0.060
Florida	0.096
Georgia	0.064
Hawaii	0.050
Idaho	0.074
Illinois	-0.002
Indiana	0.023
Iowa	0.029
Kansas	0.019
Kentucky	0.022
Louisiana	0.033
Maine	0.002
Maryland	0.042
Massachusetts	0.040
Michigan	0.004
Minnesota	0.041
Mississippi	0.007
Missouri	0.017



Montana	0.054
Nebraska	0.044
Nevada	0.089
New Hampshire	0.014
New Jersey	0.017
New Mexico	0.011
New York	0.019
North Carolina	0.064
North Dakota	0.127
Ohio	0.007
Oklahoma	0.046
Oregon	0.068
Pennsylvania	0.006
Rhode Island	0.003
South Carolina	0.073
South Dakota	0.063
Tennessee	0.048
Texas	0.108
Utah	0.104
Vermont	-0.002
Virginia	0.051
Washington	0.084
West Virginia	-0.012
Wisconsin	0.016
Wyoming	0.039

population, census, april 1, 2010 \

State	
Alabama	4779736
Alaska	710231
Arizona	6392017
Arkansas	2915918
California	37253956
Colorado	5029196
Connecticut	3574097
Delaware	897934
Florida	18801310
Georgia	9687653
Hawaii	1360301
Idaho	1567582
Illinois	12830632
Indiana	6483802
Iowa	3046355
Kansas	2853118
Kentucky	4339367
Louisiana	4533372
Maine	1328361
Maryland	5773552

Massachusetts	6547629
Michigan	9883640
Minnesota	5303925
Mississippi	2967297
Missouri	5988927
Montana	989415
Nebraska	1826341
Nevada	2700551
New Hampshire	1316470
New Jersey	8791894
New Mexico	2059179
New York	19378102
North Carolina	9535483
North Dakota	672591
Ohio	11536504
Oklahoma	3751351
Oregon	3831074
Pennsylvania	12702379
Rhode Island	1052567
South Carolina	4625364
South Dakota	814180
Tennessee	6346105
Texas	25145561
Utah	2763885
Vermont	625741
Virginia	8001024
Washington	6724540
West Virginia	1852994
Wisconsin	5686986
Wyoming	563626

persons under 5 years, percent, july 1, 2016, (v2016) \

State	
Alabama	0.060
Alaska	0.073
Arizona	0.063
Arkansas	0.064
California	0.063
Colorado	0.061
Connecticut	0.052
Delaware	0.058
Florida	0.055
Georgia	0.064
Hawaii	0.064
Idaho	0.068
Illinois	0.060
Indiana	0.064
Iowa	0.064

Kansas	0.067
Kentucky	0.062
Louisiana	0.066
Maine	0.049
Maryland	0.061
Massachusetts	0.053
Michigan	0.058
Minnesota	0.064
Mississippi	0.063
Missouri	0.061
Montana	0.060
Nebraska	0.070
Nevada	0.063
New Hampshire	0.048
New Jersey	0.058
New Mexico	0.062
New York	0.059
North Carolina	0.060
North Dakota	0.073
Ohio	0.060
Oklahoma	0.068
Oregon	0.058
Pennsylvania	0.056
Rhode Island	0.052
South Carolina	0.059
South Dakota	0.071
Tennessee	0.061
Texas	0.072
Utah	0.083
Vermont	0.049
Virginia	0.061
Washington	0.062
West Virginia	0.055
Wisconsin	0.058
Wyoming	0.065

persons under 5 years, percent, april 1, 2010 \

State	
Alabama	0.064
Alaska	0.076
Arizona	0.071
Arkansas	0.068
California	0.068
Colorado	0.068
Connecticut	0.057
Delaware	0.062
Florida	0.057
Georgia	0.071

Hawaii	0.064
Idaho	0.078
Illinois	0.065
Indiana	0.067
Iowa	0.066
Kansas	0.072
Kentucky	0.065
Louisiana	0.069
Maine	0.052
Maryland	0.063
Massachusetts	0.056
Michigan	0.060
Minnesota	0.067
Mississippi	0.071
Missouri	0.065
Montana	0.063
Nebraska	0.072
Nevada	0.069
New Hampshire	0.053
New Jersey	0.062
New Mexico	0.070
New York	0.060
North Carolina	0.066
North Dakota	0.066
Ohio	0.062
Oklahoma	0.070
Oregon	0.062
Pennsylvania	0.057
Rhode Island	0.055
South Carolina	0.065
South Dakota	0.073
Tennessee	0.064
Texas	0.077
Utah	0.095
Vermont	0.051
Virginia	0.064
Washington	0.065
West Virginia	0.056
Wisconsin	0.063
Wyoming	0.071

persons under 18 years, percent, july 1, 2016, (v2016) \

State	
Alabama	0.226
Alaska	0.252
Arizona	0.235
Arkansas	0.236
California	0.232

Colorado	0.228
Connecticut	0.211
Delaware	0.215
Florida	0.201
Georgia	0.244
Hawaii	0.216
Idaho	0.260
Illinois	0.229
Indiana	0.238
Iowa	0.233
Kansas	0.246
Kentucky	0.228
Louisiana	0.238
Maine	0.191
Maryland	0.224
Massachusetts	0.202
Michigan	0.221
Minnesota	0.233
Mississippi	0.241
Missouri	0.228
Montana	0.218
Nebraska	0.248
Nevada	0.230
New Hampshire	0.195
New Jersey	0.222
New Mexico	0.236
New York	0.212
North Carolina	0.227
North Dakota	0.233
Ohio	0.225
Oklahoma	0.245
Oregon	0.212
Pennsylvania	0.209
Rhode Island	0.197
South Carolina	0.221
South Dakota	0.246
Tennessee	0.226
Texas	0.262
Utah	0.302
Vermont	0.190
Virginia	0.222
Washington	0.224
West Virginia	0.205
Wisconsin	0.223
Wyoming	0.237

persons under 18 years, percent, april 1, 2010 \

State

Alabama	0.237
Alaska	0.264
Arizona	0.255
Arkansas	0.244
California	0.250
Colorado	0.244
Connecticut	0.229
Delaware	0.229
Florida	0.213
Georgia	0.257
Hawaii	0.223
Idaho	0.274
Illinois	0.244
Indiana	0.248
Iowa	0.239
Kansas	0.255
Kentucky	0.236
Louisiana	0.247
Maine	0.207
Maryland	0.234
Massachusetts	0.217
Michigan	0.237
Minnesota	0.242
Mississippi	0.255
Missouri	0.238
Montana	0.226
Nebraska	0.251
Nevada	0.246
New Hampshire	0.218
New Jersey	0.235
New Mexico	0.252
New York	0.223
North Carolina	0.239
North Dakota	0.223
Ohio	0.237
Oklahoma	0.248
Oregon	0.226
Pennsylvania	0.220
Rhode Island	0.213
South Carolina	0.234
South Dakota	0.249
Tennessee	0.236
Texas	0.273
Utah	0.315
Vermont	0.207
Virginia	0.232
Washington	0.235
West Virginia	0.209

Wisconsin	0.236
Wyoming	0.240

persons 65 years and over, percent, july 1, 2016, (v2016) \

State	
Alabama	0.161
Alaska	0.104
Arizona	0.169
Arkansas	0.163
California	0.136
Colorado	0.134
Connecticut	0.161
Delaware	0.175
Florida	0.199
Georgia	0.131
Hawaii	0.171
Idaho	0.151
Illinois	0.146
Indiana	0.149
Iowa	0.164
Kansas	0.150
Kentucky	0.156
Louisiana	0.144
Maine	0.194
Maryland	0.146
Massachusetts	0.158
Michigan	0.162
Minnesota	0.151
Mississippi	0.151
Missouri	0.161
Montana	0.177
Nebraska	0.150
Nevada	0.150
New Hampshire	0.170
New Jersey	0.153
New Mexico	0.165
New York	0.154
North Carolina	0.155
North Dakota	0.145
Ohio	0.162
Oklahoma	0.150
Oregon	0.168
Pennsylvania	0.174
Rhode Island	0.165
South Carolina	0.167
South Dakota	0.160
Tennessee	0.157
Texas	0.120

Utah	0.105
Vermont	0.181
Virginia	0.146
Washington	0.148
West Virginia	0.188
Wisconsin	0.161
Wyoming	0.150

persons 65 years and over, percent, april 1, 2010 ... \

State	...
Alabama	0.138 ...
Alaska	0.077 ...
Arizona	0.138 ...
Arkansas	0.144 ...
California	0.114 ...
Colorado	0.109 ...
Connecticut	0.142 ...
Delaware	0.144 ...
Florida	0.173 ...
Georgia	0.107 ...
Hawaii	0.143 ...
Idaho	0.124 ...
Illinois	0.125 ...
Indiana	0.130 ...
Iowa	0.149 ...
Kansas	0.132 ...
Kentucky	0.133 ...
Louisiana	0.123 ...
Maine	0.159 ...
Maryland	0.123 ...
Massachusetts	0.138 ...
Michigan	0.138 ...
Minnesota	0.129 ...
Mississippi	0.128 ...
Missouri	0.140 ...
Montana	0.148 ...
Nebraska	0.135 ...
Nevada	0.120 ...
New Hampshire	0.135 ...
New Jersey	0.135 ...
New Mexico	0.132 ...
New York	0.135 ...
North Carolina	0.129 ...
North Dakota	0.145 ...
Ohio	0.141 ...
Oklahoma	0.135 ...
Oregon	0.139 ...
Pennsylvania	0.154 ...



Rhode Island	0.144	...
South Carolina	0.137	...
South Dakota	0.143	...
Tennessee	0.134	...
Texas	0.103	...
Utah	0.090	...
Vermont	0.146	...
Virginia	0.122	...
Washington	0.123	...
West Virginia	0.160	...
Wisconsin	0.137	...
Wyoming	0.124	...

	all firms, 2012	men-owned firms, 2012	\
State			
Alabama	374153	203604	
Alaska	68032	35402	
Arizona	499926	245243	
Arkansas	231959	123158	
California	3548449	1852580	
Colorado	547352	284554	
Connecticut	326693	187845	
Delaware	73418	38328	
Florida	2100187	1084885	
Georgia	929864	480578	
Hawaii	118454	60409	
Idaho	146642	70438	
Illinois	1135017	609648	
Indiana	479059	253533	
Iowa	259121	135382	
Kansas	239118	125169	
Kentucky	331546	184154	
Louisiana	414291	215111	
Maine	139570	79243	
Maryland	531953	276630	
Massachusetts	607664	357158	
Michigan	834087	446128	
Minnesota	489494	268710	
Mississippi	235454	125079	
Missouri	491606	257948	
Montana	112419	55913	
Nebraska	164089	83696	
Nevada	227156	114057	
New Hampshire	131638	76716	
New Jersey	792088	464592	
New Mexico	151363	71132	
New York	2008988	1139910	
North Carolina	805985	435677	

North Dakota	68270	37016
Ohio	904814	510078
Oklahoma	327229	174395
Oregon	339305	165691
Pennsylvania	975453	579400
Rhode Island	94642	54651
South Carolina	367726	202446
South Dakota	81314	42418
Tennessee	550453	302249
Texas	2356748	1251696
Utah	251419	132163
Vermont	75827	41270
Virginia	653193	353012
Washington	541522	262650
West Virginia	114435	63112
Wisconsin	432980	236252
Wyoming	62427	30039

	women-owned firms, 2012	minority-owned firms, 2012 \
State		
Alabama	137630	92219
Alaska	22141	13688
Arizona	182425	135313
Arkansas	75962	35982
California	1320085	1619857
Colorado	194508	85849
Connecticut	106678	56113
Delaware	23964	14440
Florida	807817	926112
Georgia	376506	371588
Hawaii	44453	74208
Idaho	45121	10592
Illinois	417500	311684
Indiana	162798	61252
Iowa	82345	14707
Kansas	77194	26127
Kentucky	106011	27258
Louisiana	151114	126100
Maine	42067	4339
Maryland	209119	203394
Massachusetts	199210	89967
Michigan	306986	158946
Minnesota	157821	47302
Mississippi	89159	74824
Missouri	162616	61035
Montana	35449	5578
Nebraska	51936	14571
Nevada	82508	71864

New Hampshire	38525	6111
New Jersey	252944	237242
New Mexico	59044	60622
New York	725709	709021
North Carolina	287058	183380
North Dakota	20316	3190
Ohio	306824	122653
Oklahoma	105168	64875
Oregon	123015	41456
Pennsylvania	304803	131512
Rhode Island	30484	14737
South Carolina	131856	83233
South Dakota	23722	4101
Tennessee	195694	105234
Texas	866678	1070392
Utah	76269	24423
Vermont	23417	2354
Virginia	236290	185043
Washington	187677	92807
West Virginia	39065	5777
Wisconsin	133859	40507
Wyoming	19344	4077

	nonminority-owned firms, 2012	veteran-owned firms, 2012 \
State		
Alabama	272651	41943
Alaska	51147	7953
Arizona	344981	46780
Arkansas	189029	25915
California	1819107	252377
Colorado	442365	51722
Connecticut	259614	31056
Delaware	54782	7206
Florida	1121749	185756
Georgia	538893	96787
Hawaii	38510	11148
Idaho	130973	12804
Illinois	795129	89110
Indiana	405090	45174
Iowa	236561	25889
Kansas	204562	21610
Kentucky	296155	33208
Louisiana	277676	42211
Maine	131322	15689
Maryland	314902	50976
Massachusetts	499959	58339
Michigan	657237	71861
Minnesota	428716	45582

Mississippi	155094	26789
Missouri	415972	49217
Montana	102746	11486
Nebraska	144122	16693
Nevada	144944	23049
New Hampshire	121297	16119
New Jersey	533808	57996
New Mexico	83857	14096
New York	1248304	137532
North Carolina	603182	86571
North Dakota	62271	6584
Ohio	759569	91316
Oklahoma	249027	36273
Oregon	285028	30918
Pennsylvania	818858	97969
Rhode Island	77042	9904
South Carolina	276269	47987
South Dakota	74228	8604
Tennessee	434025	59379
Texas	1224845	213590
Utah	218826	18754
Vermont	70491	8237
Virginia	450109	76434
Washington	426697	49331
West Virginia	104785	12912
Wisconsin	379934	39830
Wyoming	55397	6470

nonveteran-owned firms, 2012    population per square mile, 2010    \

State		
Alabama	316984	94.4
Alaska	56091	1.2
Arizona	427582	56.3
Arkansas	192988	56.0
California	3176341	239.1
Colorado	469524	48.5
Connecticut	281182	738.1
Delaware	60318	460.8
Florida	1846686	350.6
Georgia	800585	168.4
Hawaii	101334	211.8
Idaho	124314	19.0
Illinois	1006885	231.1
Indiana	412543	181.0
Iowa	217779	54.5
Kansas	203401	34.9
Kentucky	282704	109.9
Louisiana	354460	104.9

Maine	115839	43.1
Maryland	462232	594.8
Massachusetts	525667	839.4
Michigan	733517	174.8
Minnesota	419628	66.6
Mississippi	198566	63.2
Missouri	415542	87.1
Montana	93393	6.8
Nebraska	137254	23.8
Nevada	191293	24.6
New Hampshire	108817	147.0
New Jersey	707975	1195.5
New Mexico	128693	17.0
New York	1811544	411.2
North Carolina	684743	196.1
North Dakota	56904	9.7
Ohio	776193	282.3
Oklahoma	273491	54.7
Oregon	288790	39.9
Pennsylvania	837792	283.9
Rhode Island	79837	1018.1
South Carolina	303137	153.9
South Dakota	66219	10.7
Tennessee	469392	153.9
Texas	2057218	96.3
Utah	219807	33.6
Vermont	63317	67.9
Virginia	548439	202.6
Washington	461401	101.2
West Virginia	94960	77.1
Wisconsin	370755	105.0
Wyoming	51353	5.8

land area in square miles, 2010    fips code

State		
Alabama	50645.33	"01"
Alaska	570640.95	"02"
Arizona	113594.08	"04"
Arkansas	52035.48	"05"
California	155779.22	"06"
Colorado	103641.89	"08"
Connecticut	4842.36	"09"
Delaware	1948.54	"10"
Florida	53624.76	"12"
Georgia	57513.49	"13"
Hawaii	6422.63	"15"
Idaho	82643.12	"16"
Illinois	55518.93	"17"

Indiana	35826.11	"18"
Iowa	55857.13	"19"
Kansas	81758.72	"20"
Kentucky	39486.34	"21"
Louisiana	43203.90	"22"
Maine	30842.92	"23"
Maryland	9707.24	"24"
Massachusetts	7800.06	"25"
Michigan	56538.90	"26"
Minnesota	79626.74	"27"
Mississippi	46923.27	"28"
Missouri	68741.52	"29"
Montana	145545.80	"30"
Nebraska	76824.17	"31"
Nevada	109781.18	"32"
New Hampshire	8952.65	"33"
New Jersey	7354.22	"34"
New Mexico	121298.15	"35"
New York	47126.40	"36"
North Carolina	48617.91	"37"
North Dakota	69000.80	"38"
Ohio	40860.69	"39"
Oklahoma	68594.92	"40"
Oregon	95988.01	"41"
Pennsylvania	44742.70	"42"
Rhode Island	1033.81	"44"
South Carolina	30060.70	"45"
South Dakota	75811.00	"46"
Tennessee	41234.90	"47"
Texas	261231.71	"48"
Utah	82169.62	"49"
Vermont	9216.66	"50"
Virginia	39490.09	"51"
Washington	66455.52	"53"
West Virginia	24038.21	"54"
Wisconsin	54157.80	"55"
Wyoming	97093.14	"56"

[50 rows x 65 columns]

In [448]: df\_gun

```
Out[448]:
```

	month	state	permit	permit_recheck	handgun	long_gun	\
0	2017-09	Alabama	16717.0	0.000	5734.0	6320.0	
1	2017-09	Alaska	209.0	2.000	2320.0	2930.0	
2	2017-09	Arizona	5069.0	382.000	11063.0	7946.0	
3	2017-09	Arkansas	2935.0	632.000	4347.0	6063.0	
4	2017-09	California	57839.0	0.000	37165.0	24581.0	

...	...	...	...	...	...	...
12480	1998-11	Virginia	0.0	1282.552	14.0	2.0
12481	1998-11	Washington	1.0	1282.552	65.0	286.0
12482	1998-11	West Virginia	3.0	1282.552	149.0	251.0
12483	1998-11	Wisconsin	0.0	1282.552	25.0	214.0
12484	1998-11	Wyoming	8.0	1282.552	45.0	49.0

	other	multiple	admin	prepawn_handgun	...	returned_other \
0	221.0000	317	0.0	15.000000	...	0.000000
1	219.0000	160	0.0	5.000000	...	0.000000
2	920.0000	631	0.0	13.000000	...	0.000000
3	165.0000	366	51.0	12.000000	...	0.000000
4	2984.0000	0	0.0	0.000000	...	0.000000
...	...	...	...	...	...	...
12480	396.0524	8	0.0	5.301969	...	1.130303
12481	396.0524	8	1.0	5.301969	...	1.130303
12482	396.0524	5	0.0	5.301969	...	1.130303
12483	396.0524	2	0.0	5.301969	...	1.130303
12484	396.0524	5	0.0	5.301969	...	1.130303

	rentals_handgun	rentals_long_gun	private_sale_handgun \
0	0.000000	0.000	9.0000
1	0.000000	0.000	17.0000
2	0.000000	0.000	38.0000
3	0.000000	0.000	13.0000
4	0.000000	0.000	0.0000
...	...	...	...
12480	0.084444	0.096	16.4272
12481	0.084444	0.096	16.4272
12482	0.084444	0.096	16.4272
12483	0.084444	0.096	16.4272
12484	0.084444	0.096	16.4272

	private_sale_long_gun	private_sale_other	return_to_seller_handgun \
0	16.0000	3.0000	0.000000
1	24.0000	1.0000	0.000000
2	12.0000	2.0000	0.000000
3	23.0000	0.0000	0.000000
4	0.0000	0.0000	0.000000
...	...	...	...
12480	12.7632	1.1316	0.439556
12481	12.7632	1.1316	0.439556
12482	12.7632	1.1316	0.439556
12483	12.7632	1.1316	0.439556
12484	12.7632	1.1316	0.439556

	return_to_seller_long_gun	return_to_seller_other	totals
0	0.0000	3.00000	32019

1	0.0000	0.00000	6303
2	0.0000	0.00000	28394
3	2.0000	1.00000	17747
4	0.0000	0.00000	123506
...	...	...	...
12480	0.4844	0.11561	24
12481	0.4844	0.11561	361
12482	0.4844	0.11561	408
12483	0.4844	0.11561	241
12484	0.4844	0.11561	107

[11350 rows x 27 columns]

In [449]: *#For dealing with dates, we will bring them to the standard format using the datetime*  
df\_gun.month = pd.to\_datetime(df\_gun['month'], errors='coerce')

df\_gun

Out[449]:

	month	state	permit	permit_recheck	handgun	long_gun	\
0	2017-09-01	Alabama	16717.0	0.000	5734.0	6320.0	
1	2017-09-01	Alaska	209.0	2.000	2320.0	2930.0	
2	2017-09-01	Arizona	5069.0	382.000	11063.0	7946.0	
3	2017-09-01	Arkansas	2935.0	632.000	4347.0	6063.0	
4	2017-09-01	California	57839.0	0.000	37165.0	24581.0	
...	...	...	...	...	...	...	
12480	1998-11-01	Virginia	0.0	1282.552	14.0	2.0	
12481	1998-11-01	Washington	1.0	1282.552	65.0	286.0	
12482	1998-11-01	West Virginia	3.0	1282.552	149.0	251.0	
12483	1998-11-01	Wisconsin	0.0	1282.552	25.0	214.0	
12484	1998-11-01	Wyoming	8.0	1282.552	45.0	49.0	

	other	multiple	admin	prepawn_handgun	...	returned_other	\
0	221.0000	317	0.0	15.000000	...	0.000000	
1	219.0000	160	0.0	5.000000	...	0.000000	
2	920.0000	631	0.0	13.000000	...	0.000000	
3	165.0000	366	51.0	12.000000	...	0.000000	
4	2984.0000	0	0.0	0.000000	...	0.000000	
...	...	...	...	...	...	...	
12480	396.0524	8	0.0	5.301969	...	1.130303	
12481	396.0524	8	1.0	5.301969	...	1.130303	
12482	396.0524	5	0.0	5.301969	...	1.130303	
12483	396.0524	2	0.0	5.301969	...	1.130303	
12484	396.0524	5	0.0	5.301969	...	1.130303	

	rentals_handgun	rentals_long_gun	private_sale_handgun	\
0	0.000000	0.000	9.0000	
1	0.000000	0.000	17.0000	



2	0.000000	0.000	38.0000
3	0.000000	0.000	13.0000
4	0.000000	0.000	0.0000
...	...	...	...
12480	0.084444	0.096	16.4272
12481	0.084444	0.096	16.4272
12482	0.084444	0.096	16.4272
12483	0.084444	0.096	16.4272
12484	0.084444	0.096	16.4272

	private_sale_long_gun	private_sale_other	return_to_seller_handgun \
0	16.0000	3.0000	0.000000
1	24.0000	1.0000	0.000000
2	12.0000	2.0000	0.000000
3	23.0000	0.0000	0.000000
4	0.0000	0.0000	0.000000
...	...	...	...
12480	12.7632	1.1316	0.439556
12481	12.7632	1.1316	0.439556
12482	12.7632	1.1316	0.439556
12483	12.7632	1.1316	0.439556
12484	12.7632	1.1316	0.439556

	return_to_seller_long_gun	return_to_seller_other	totals
0	0.0000	3.00000	32019
1	0.0000	0.00000	6303
2	0.0000	0.00000	28394
3	2.0000	1.00000	17747
4	0.0000	0.00000	123506
...	...	...	...
12480	0.4844	0.11561	24
12481	0.4844	0.11561	361
12482	0.4844	0.11561	408
12483	0.4844	0.11561	241
12484	0.4844	0.11561	107

[11350 rows x 27 columns]

## Exploratory Data Analysis

### 1.1.3 Research Question 1: Which state has the highest total gun purchases in April 2005 and April 2010?

In [450]: *# Use this, and more code cells, to explore your data. Don't forget to add  
# Markdown cells to document your observations and findings.*

*#We will create a new df with all the columns that will be required to perform the analysis*

```
df_gun_q1 = df_gun[['month', 'state', 'totals']]
```

```
df_gun_q1
```

```
Out[450]:
```

	month	state	totals
0	2017-09-01	Alabama	32019
1	2017-09-01	Alaska	6303
2	2017-09-01	Arizona	28394
3	2017-09-01	Arkansas	17747
4	2017-09-01	California	123506
...	...	...	...
12480	1998-11-01	Virginia	24
12481	1998-11-01	Washington	361
12482	1998-11-01	West Virginia	408
12483	1998-11-01	Wisconsin	241
12484	1998-11-01	Wyoming	107

```
[11350 rows x 3 columns]
```

```
In [451]: df_gun_q1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 11350 entries, 0 to 12484
Data columns (total 3 columns):
#   Column  Non-Null Count  Dtype
---  -
0   month   11350 non-null     datetime64[ns]
1   state   11350 non-null     object
2   totals  11350 non-null     int64
dtypes: datetime64[ns](1), int64(1), object(1)
memory usage: 354.7+ KB
```

```
In [452]: #Now lets extract just the April 2000 and April 2010 data
```

```
q1_2005 = df_gun_q1[df_gun_q1['month'] == "2005-04-01"]
display(q1_2005.head())
```

```
q1_2010 = df_gun_q1[df_gun_q1['month'] == "2010-04-01"]
display(q1_2010.head())
```

	month	state	totals
8195	2005-04-01	Alabama	14099
8196	2005-04-01	Alaska	3843
8197	2005-04-01	Arizona	13659
8198	2005-04-01	Arkansas	10027
8199	2005-04-01	California	49679

```
month      state  totals
```

```

4895 2010-04-01    Alabama    20791
4896 2010-04-01     Alaska     6411
4897 2010-04-01    Arizona    16578
4898 2010-04-01    Arkansas    14563
4899 2010-04-01   California    80750

```

```

In [453]: print(q1_2005.describe())
          print('\n')
          print(q1_2010.describe())

```

```

          totals
count      50.000000
mean    12932.920000
std     11216.245617
min       659.000000
25%      5188.750000
50%      9859.000000
75%     16423.750000
max     49679.000000

```

```

          totals
count      50.0000
mean     24517.7400
std     34280.1673
min       963.0000
25%      7197.5000
50%     15242.5000
75%     26335.5000
max     211261.0000

```

```

In [454]: #We will now merge both of the df created into a single df usig merge to do further an

```

```

q1_combined = q1_2005.merge(q1_2010, on = 'state', how = 'outer')
q1_combined.head()

```

```

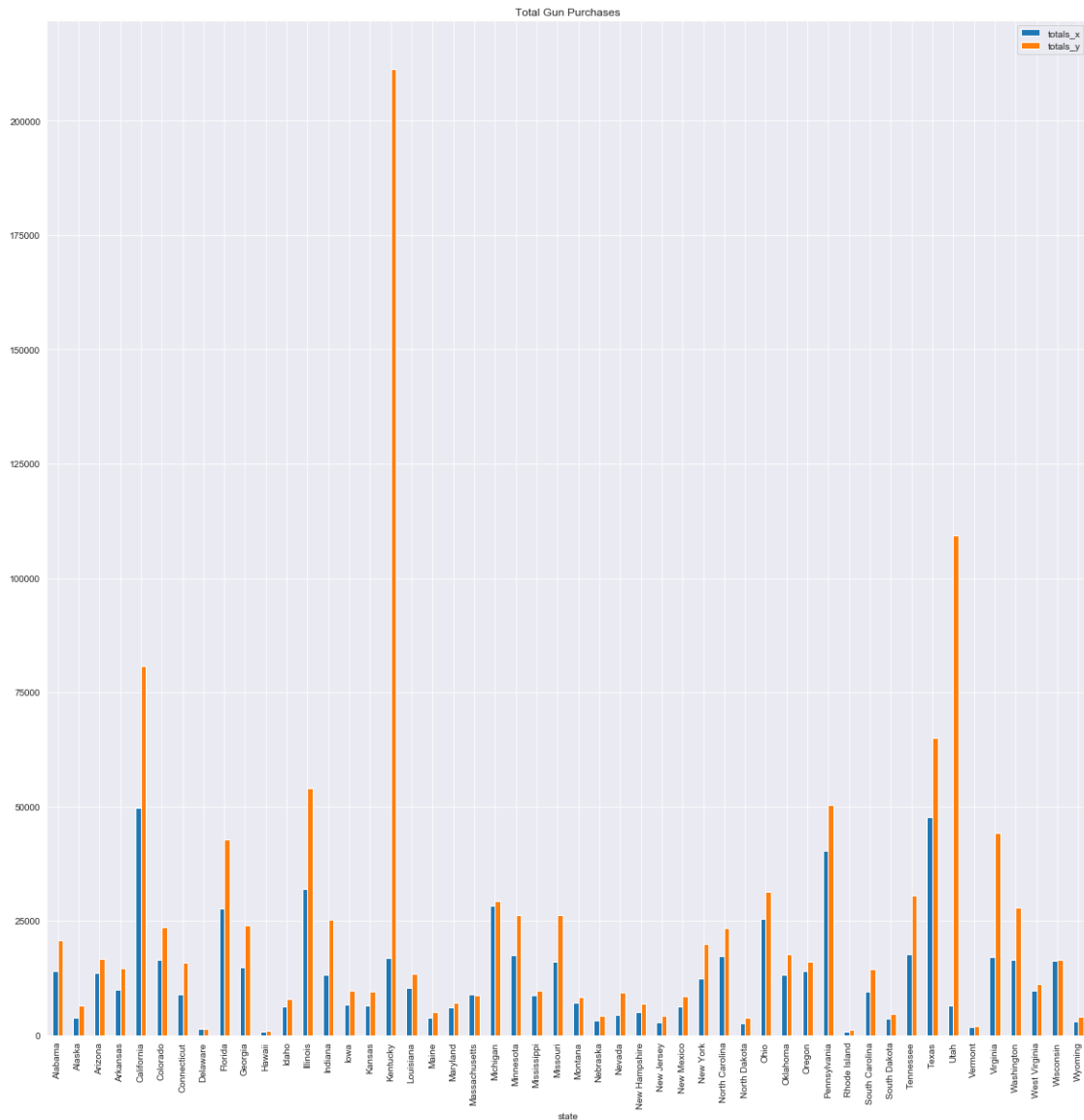
Out[454]:
   month_x  state  totals_x  month_y  totals_y
0  2005-04-01  Alabama    14099  2010-04-01    20791
1  2005-04-01  Alaska     3843  2010-04-01     6411
2  2005-04-01  Arizona    13659  2010-04-01    16578
3  2005-04-01  Arkansas    10027  2010-04-01    14563
4  2005-04-01  California   49679  2010-04-01    80750

```

```

In [456]: q1_combined.plot(x = 'state',title = 'Total Gun Purchases',
                          y = ['totals_x','totals_y'],kind = 'bar', figsize = (20,20));

```



```
In [457]: kent_2005 = q1_combined[q1_combined['state'] == 'Kentucky']['totals_x']
          kent_2010 = q1_combined[q1_combined['state'] == 'Kentucky']['totals_y']

          print('The total number of firearms sold in state of Ketucky in 2000 is:',kent_2005);
          print('The total number of firearms sold in state of Ketucky in 2010 is:',kent_2010);
```

```
The total number of firearms sold in state of Ketucky in 2000 is: 16      16961
Name: totals_x, dtype: int64
The total number of firearms sold in state of Ketucky in 2010 is: 16      211261
Name: totals_y, dtype: int64
```

## Insight:

After looking through the overall firearms sold in the year of 2005 and compared to the year of 2010, there has been a very significant increase in the 5-year span. There has been an increase of 194,300 firearms in 2010 compared to 2005

### 1.1.4 Research Question 2:

**What is per capita firearm sales for all states in April 2010 vs July 2016?**

In [458]: *# To perform analysis on the question above, we will require some data from both the a*

```
df_q2 = df_gun[['state', 'month', 'totals']]
```

```
df_q2 = df_q2[df_q2['month'] == '2016-07-01']
```

```
df_q2
```

```
Out[458]:
```

	state	month	totals
770	Alabama	2016-07-01	48927
771	Alaska	2016-07-01	6793
772	Arizona	2016-07-01	34496
773	Arkansas	2016-07-01	19378
774	California	2016-07-01	190218
775	Colorado	2016-07-01	43094
776	Connecticut	2016-07-01	29755
777	Delaware	2016-07-01	4494
779	Florida	2016-07-01	125208
780	Georgia	2016-07-01	49183
782	Hawaii	2016-07-01	1565
783	Idaho	2016-07-01	12154
784	Illinois	2016-07-01	168227
785	Indiana	2016-07-01	88340
786	Iowa	2016-07-01	11937
787	Kansas	2016-07-01	14140
788	Kentucky	2016-07-01	363085
789	Louisiana	2016-07-01	41063
790	Maine	2016-07-01	7702
792	Maryland	2016-07-01	12228
793	Massachusetts	2016-07-01	20480
794	Michigan	2016-07-01	40142
795	Minnesota	2016-07-01	43368
796	Mississippi	2016-07-01	21907
797	Missouri	2016-07-01	46637
798	Montana	2016-07-01	9869
799	Nebraska	2016-07-01	5429
800	Nevada	2016-07-01	11785
801	New Hampshire	2016-07-01	13068

802	New Jersey	2016-07-01	10074
803	New Mexico	2016-07-01	12219
804	New York	2016-07-01	29513
805	North Carolina	2016-07-01	44123
806	North Dakota	2016-07-01	5470
807	Ohio	2016-07-01	63148
808	Oklahoma	2016-07-01	25946
809	Oregon	2016-07-01	24813
810	Pennsylvania	2016-07-01	86137
812	Rhode Island	2016-07-01	2368
813	South Carolina	2016-07-01	32730
814	South Dakota	2016-07-01	7406
815	Tennessee	2016-07-01	57653
816	Texas	2016-07-01	127207
817	Utah	2016-07-01	17608
818	Vermont	2016-07-01	2674
820	Virginia	2016-07-01	43574
821	Washington	2016-07-01	47887
822	West Virginia	2016-07-01	16791
823	Wisconsin	2016-07-01	38922
824	Wyoming	2016-07-01	4585

In [460]: *#Drop the indexes from the between and get new indexes*  
df\_q2.reset\_index(drop = True ,inplace = True)

df\_q2

Out [460]:

	state	month	totals
0	Alabama	2016-07-01	48927
1	Alaska	2016-07-01	6793
2	Arizona	2016-07-01	34496
3	Arkansas	2016-07-01	19378
4	California	2016-07-01	190218
5	Colorado	2016-07-01	43094
6	Connecticut	2016-07-01	29755
7	Delaware	2016-07-01	4494
8	Florida	2016-07-01	125208
9	Georgia	2016-07-01	49183
10	Hawaii	2016-07-01	1565
11	Idaho	2016-07-01	12154
12	Illinois	2016-07-01	168227
13	Indiana	2016-07-01	88340
14	Iowa	2016-07-01	11937
15	Kansas	2016-07-01	14140
16	Kentucky	2016-07-01	363085
17	Louisiana	2016-07-01	41063
18	Maine	2016-07-01	7702
19	Maryland	2016-07-01	12228

20	Massachusetts	2016-07-01	20480
21	Michigan	2016-07-01	40142
22	Minnesota	2016-07-01	43368
23	Mississippi	2016-07-01	21907
24	Missouri	2016-07-01	46637
25	Montana	2016-07-01	9869
26	Nebraska	2016-07-01	5429
27	Nevada	2016-07-01	11785
28	New Hampshire	2016-07-01	13068
29	New Jersey	2016-07-01	10074
30	New Mexico	2016-07-01	12219
31	New York	2016-07-01	29513
32	North Carolina	2016-07-01	44123
33	North Dakota	2016-07-01	5470
34	Ohio	2016-07-01	63148
35	Oklahoma	2016-07-01	25946
36	Oregon	2016-07-01	24813
37	Pennsylvania	2016-07-01	86137
38	Rhode Island	2016-07-01	2368
39	South Carolina	2016-07-01	32730
40	South Dakota	2016-07-01	7406
41	Tennessee	2016-07-01	57653
42	Texas	2016-07-01	127207
43	Utah	2016-07-01	17608
44	Vermont	2016-07-01	2674
45	Virginia	2016-07-01	43574
46	Washington	2016-07-01	47887
47	West Virginia	2016-07-01	16791
48	Wisconsin	2016-07-01	38922
49	Wyoming	2016-07-01	4585

```
In [473]: #We are adding the Population of July 2016 in the newly created df
df_q2['population_july2016'] = df_census['population estimates, july 1, 2016, (v2016)']

df_q2.head()
```

```
Out[473]:
```

	state	month	totals	population_july2016
0	Alabama	2016-07-01	48927	4863300
1	Alaska	2016-07-01	6793	741894
2	Arizona	2016-07-01	34496	6931071
3	Arkansas	2016-07-01	19378	2988248
4	California	2016-07-01	190218	39250017

```
In [466]: #we are performing the same steps to create a new dataframe for April 2010
df_q2_2010 = df_gun[['state', 'month', 'totals']]

df_q2_2010 = df_q2_2010[df_q2_2010['month']=='2010-04-01']
```

```
df_q2_2010.reset_index(drop=True, inplace=True)
```

```
df_q2_2010
```

```
Out[466]:
```

	state	month	totals
0	Alabama	2010-04-01	20791
1	Alaska	2010-04-01	6411
2	Arizona	2010-04-01	16578
3	Arkansas	2010-04-01	14563
4	California	2010-04-01	80750
5	Colorado	2010-04-01	23609
6	Connecticut	2010-04-01	15922
7	Delaware	2010-04-01	1439
8	Florida	2010-04-01	42794
9	Georgia	2010-04-01	24065
10	Hawaii	2010-04-01	963
11	Idaho	2010-04-01	7814
12	Illinois	2010-04-01	53929
13	Indiana	2010-04-01	25232
14	Iowa	2010-04-01	9720
15	Kansas	2010-04-01	9529
16	Kentucky	2010-04-01	211261
17	Louisiana	2010-04-01	13373
18	Maine	2010-04-01	5073
19	Maryland	2010-04-01	6992
20	Massachusetts	2010-04-01	8748
21	Michigan	2010-04-01	29383
22	Minnesota	2010-04-01	26351
23	Mississippi	2010-04-01	9702
24	Missouri	2010-04-01	26289
25	Montana	2010-04-01	8367
26	Nebraska	2010-04-01	4141
27	Nevada	2010-04-01	9294
28	New Hampshire	2010-04-01	6911
29	New Jersey	2010-04-01	4215
30	New Mexico	2010-04-01	8599
31	New York	2010-04-01	19906
32	North Carolina	2010-04-01	23378
33	North Dakota	2010-04-01	3726
34	Ohio	2010-04-01	31312
35	Oklahoma	2010-04-01	17750
36	Oregon	2010-04-01	16031
37	Pennsylvania	2010-04-01	50249
38	Rhode Island	2010-04-01	1199
39	South Carolina	2010-04-01	14441
40	South Dakota	2010-04-01	4561
41	Tennessee	2010-04-01	30453
42	Texas	2010-04-01	65012



```

43         Utah 2010-04-01 109391
44         Vermont 2010-04-01 2053
45         Virginia 2010-04-01 44137
46         Washington 2010-04-01 27816
47         West Virginia 2010-04-01 11180
48         Wisconsin 2010-04-01 16471
49         Wyoming 2010-04-01 4013

```

```

In [472]: df_q2_2010['population_april2010'] = df_census['population estimates base, april 1, 20
df_q2_2010.head()

```

```

Out[472]:
      state      month  totals  population_april2010
0  Alabama 2010-04-01   20791           4780131
1   Alaska 2010-04-01    6411           710249
2   Arizona 2010-04-01   16578          6392301
3  Arkansas 2010-04-01   14563          2916025
4  California 2010-04-01  80750          37254522

```

```

In [477]: #We will now try to calculate the per capita income percentage by summing all the stat
#and then try showing a graph of the changed from 2010 till 2016

```

```

df_q2['totals_2010'] = df_q2_2010['totals'].values
df_q2['population_april2010'] = df_q2_2010['population_april2010'].values
df_q2

```

```

Out[477]:
      state      month  totals  population_july2016  totals_2010  \
0   Alabama 2016-07-01   48927           4863300          20791
1   Alaska 2016-07-01    6793           741894           6411
2   Arizona 2016-07-01   34496          6931071          16578
3   Arkansas 2016-07-01   19378          2988248          14563
4   California 2016-07-01  190218         39250017          80750
5   Colorado 2016-07-01   43094          5540545          23609
6  Connecticut 2016-07-01   29755          3576452          15922
7   Delaware 2016-07-01    4494           952065           1439
8   Florida 2016-07-01  125208         20612439          42794
9   Georgia 2016-07-01   49183         10310371          24065
10  Hawaii 2016-07-01    1565          1428557            963
11  Idaho 2016-07-01   12154          1683140           7814
12  Illinois 2016-07-01  168227         12801539          53929
13  Indiana 2016-07-01   88340          6633053          25232
14  Iowa 2016-07-01   11937          3134693           9720
15  Kansas 2016-07-01   14140          2907289           9529
16  Kentucky 2016-07-01  363085          4436974         211261
17  Louisiana 2016-07-01   41063          4681666          13373
18  Maine 2016-07-01    7702          1331479           5073
19  Maryland 2016-07-01  12228          6016447           6992
20  Massachusetts 2016-07-01  20480          6811779           8748

```

21	Michigan	2016-07-01	40142	9928300	29383
22	Minnesota	2016-07-01	43368	5519952	26351
23	Mississippi	2016-07-01	21907	2988726	9702
24	Missouri	2016-07-01	46637	6093000	26289
25	Montana	2016-07-01	9869	1042520	8367
26	Nebraska	2016-07-01	5429	1907116	4141
27	Nevada	2016-07-01	11785	2940058	9294
28	New Hampshire	2016-07-01	13068	1334795	6911
29	New Jersey	2016-07-01	10074	8944469	4215
30	New Mexico	2016-07-01	12219	2081015	8599
31	New York	2016-07-01	29513	19745289	19906
32	North Carolina	2016-07-01	44123	10146788	23378
33	North Dakota	2016-07-01	5470	757952	3726
34	Ohio	2016-07-01	63148	11614373	31312
35	Oklahoma	2016-07-01	25946	3923561	17750
36	Oregon	2016-07-01	24813	4093465	16031
37	Pennsylvania	2016-07-01	86137	12784227	50249
38	Rhode Island	2016-07-01	2368	1056426	1199
39	South Carolina	2016-07-01	32730	4961119	14441
40	South Dakota	2016-07-01	7406	865454	4561
41	Tennessee	2016-07-01	57653	6651194	30453
42	Texas	2016-07-01	127207	27862596	65012
43	Utah	2016-07-01	17608	3051217	109391
44	Vermont	2016-07-01	2674	624594	2053
45	Virginia	2016-07-01	43574	8411808	44137
46	Washington	2016-07-01	47887	7288000	27816
47	West Virginia	2016-07-01	16791	1831102	11180
48	Wisconsin	2016-07-01	38922	5778708	16471
49	Wyoming	2016-07-01	4585	585501	4013

	population_april2010
0	4780131
1	710249
2	6392301
3	2916025
4	37254522
5	5029324
6	3574114
7	897936
8	18804592
9	9688680
10	1360301
11	1567650
12	12831574
13	6484136
14	3046869
15	2853129
16	4339344

17	4533479
18	1328364
19	5773786
20	6547813
21	9884129
22	5303924
23	2968103
24	5988928
25	989414
26	1826334
27	2700691
28	1316461
29	8791953
30	2059198
31	19378110
32	9535688
33	672591
34	11536727
35	3751615
36	3831072
37	12702857
38	1052940
39	4625410
40	814195
41	6346298
42	25146100
43	2763888
44	625741
45	8001041
46	6724545
47	1853011
48	5687289
49	563767

In [501]: *#Now let us calculate the capita of 2010 and 2016*

```
capita_2010 = df_q2.totals_2010.sum()/ df_q2.population_april2010.sum()
print('The per capita firearms for April 2010 is:', capita_2010)
```

```
capita_2016 = df_q2.totals.sum()/ df_q2.population_july2016.sum()
print('The per capita firearms for July 2016 is:', capita_2016)
```

The per capita firearms for April 2010 is: 0.003978133320178106

The per capita firearms for July 2016 is: 0.006777933902633841

In [502]: *#To make it easier to perform calculations and analysis, we'll multiply both the variables by 100*

```
capita_2010 = capita_2010*100
```

```
capita_2016 = capita_2016*100

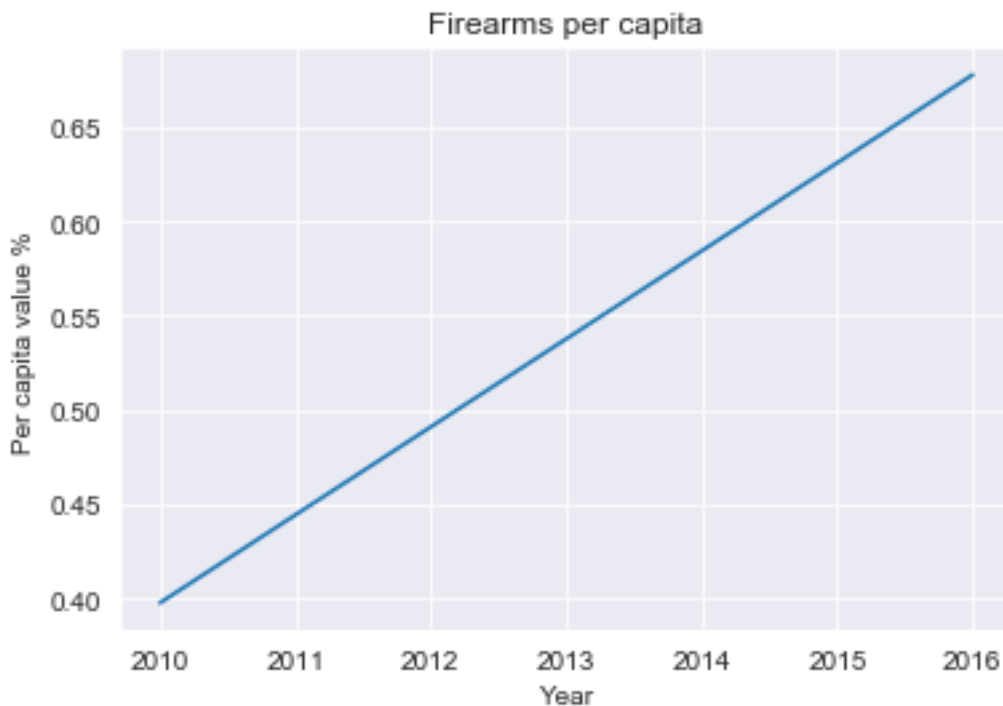
print('New capita_2010:',capita_2010)
print('New capita_2016:',capita_2016)
```

New capita\_2010: 0.3978133320178106

New capita\_2016: 0.677793390263384

In [504]: *#now we plot the graph to note the changes*

```
years = [2010,2016]
per_capita = [capita_2010,capita_2016]
plt.plot(years,per_capita);
plt.title('Firearms per capita');
plt.xlabel('Year');
plt.ylabel('Per capita value %');
```



### Insights:

The graph shows us the increase in per capita ownership of firearms. The per-capita value (percentage value) in April 2010 is 0.38 and it goes on to 0.678 in July 2016.

### ## Conclusions

A few conclusions could be made using the questions asked to the database:

1. After comparing the gun data between the years 2005 and 2010, there has been a significant increase in firearm purchases. We compared the data between April 2005 and April 2010, coming to a conclusion that the state of Kentucky has the highest increase with a number of 194,300 firearms in the 5-year span.
2. The second graph helps us in understanding the increase in per capita ownership of firearms. The per-capita value (percentage value) in April 2010 is 0.38 and it goes on to 0.678 in July 2016.

**Number of things to be considered for the question 2 visual:**

1. This chart is a nation-wide chart and does give us a birds-eye view.
2. At state level this graph's slope will vary and hence the insights cannot be generalized.
3. Also, we are using 2 point of time (i.e. April 2010 and July 2016), whereas we have no information about how this per capita relationship will vary over the 6 years between 2010 and 2016. Again, this insight cannot be generalized.

```
In [1]: from subprocess import call
        call(['python', '-m', 'nbconvert', 'Investigate_a_Dataset.ipynb'])
```

```
Out[1]: 0
```

```
In [ ]:
```