# Investigate_a_Dataset

January 28, 2019

# 1 Project: Investigating Trends in Gun Sales in the US

## 1.1 Table of Contents


## Introduction

The Brady Handgun Violence Prevention Act of 1993 introduced a mandate that Federal background checks be performed prior to the sale of firearms in the US. This led to the launch in 1998 of the FBI National Instant Criminal Background Check System (NICS), a system for conducting background checks to determine a persons eligibility to purchase a firearm. In this project, I will be examining the numbers of NICS background checks that have been performed related to the sales of firearms between January, 2010 and November, 2018. This information will then be compared with information regarding state demographics from the Census Bureau's most recent census information from 2010, 2015, and 2016 to determine if any correlations can be found between state demographics and NICS background checks. I will also be investigating any connection between the instance of mass shootings in the US and nationwide NICS background checks.

The specific questions that will be investigated are:
Is there a correlation between veteran population and NICS background checks?
Is there a correlation between foreign born population and NICS background checks?
Is there a correlation between education level and NICS background checks?
Is there a correlation between economic status and poverty and NICS background checks?
Is there a correlation between mass shootings in the US and NICS background checks?
How many NICS background checks are the result of each category of firearm since 2010?

Disclaimer: The NICS background check information will be used as a measure of firearms sales in the US for this project. However, this NICS information is not necessarily a direct measure of firearms sales as not all NICS checks lead to an actual sale. There are also laws varying from state to state regarding the private sales of firearms so not all sales will result in a NICS check.

```
In [293]: # Importing Packages

          import numpy as np
          import pandas as pd
          import matplotlib.pyplot as plt
          import seaborn as sns
          % matplotlib inline
```

## Data Wrangling
For this project, I used the most up to date FBI NICS data optained from the BuzzFeedNews nics-firearm-background-checks GitHub repository rather than the older version of the data. This data can be found here https://github.com/BuzzFeedNews/nics-firearm-background-checks.

The census data was obtained from https://d17h27t6h515a5.cloudfront.net/topher/2017/November/5a0a55 census-data/u.s.-census-data.csv.

I also included a data set on mass shootings in the US obtained from MotherJones here https://www.motherjones.com/politics/2012/12/mass-shootings-mother-jones-full-data/.

The data files were loaded into dataframes and inspected using a function to display the first few rows, the shape of the dataframe, some descriptive statistics of the data, information about the labels and types, and to determine if any data is missing or duplicated.

### 1.1.1 General Properties

In [294]:
```
# Load CSV files as data frames

df_gun = pd.read_csv('gun_data.csv')
df_census = pd.read_csv('census_data.csv')
df_shootings = pd.read_csv('shootings_data.csv')
```

In [295]:
```
# Transpose census data to match gun data layout

df_census = df_census.transpose();
```

In [296]:
```
# Define function to inspect data frames. Prints first few lines, determines size/shap
# shows descriptive statistics, shows data types, shows missing or incomplete data, ch

def inspect_df(df):
    print('Header:')
    print('{}'.format(df.head()))
    print()
    print('Shape: {}'.format(df.shape))
    print()
    print('Statistics:')
    print('{}'.format(df.describe()))
    print()
    print('Info:')
    print('{}'.format(df.info()))
    print()
    print('Duplicates: {}\n'.format(sum(df.duplicated())))
```

In [297]:
```
# Inspect gun data using inspect_df function.

inspect_df(df_gun)
```

```
Header:
     month      state   permit  permit_recheck  handgun  long_gun   other  \
0  2018-11    Alabama  22477.0            17.0   8408.0   10303.0   357.0
1  2018-11     Alaska    184.0            11.0   2707.0    2827.0   360.0
```

```
2  2018-11     Arizona   4537.0           380.0  12412.0    8886.0  1139.0
3  2018-11    Arkansas   2739.0          1163.0   5739.0   10196.0   300.0
4  2018-11  California  33942.0             0.0  36997.0   30913.0  3606.0

   multiple  admin  prepawn_handgun   ...    returned_other  rentals_handgun  \
0      1153    0.0              3.0   ...               0.0              0.0
1       197    0.0              0.0   ...               0.0              0.0
2       879    0.0              9.0   ...               0.0              0.0
3       914   60.0              9.0   ...               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                  13.0                   10.0
1               0.0                  26.0                   39.0
2               0.0                  26.0                   24.0
3               0.0                  11.0                   14.0
4               0.0                   0.0                    0.0

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

   return_to_seller_other  totals
0                     0.0   45345
1                     0.0    6645
2                     0.0   30385
3                     0.0   24475
4                     0.0  106380

[5 rows x 27 columns]

Shape: (13255, 27)

Statistics:
              permit  permit_recheck        handgun       long_gun  \
count   13231.000000     1870.000000   13235.000000   13236.000000
mean     6944.547729     1981.110160    6188.221005    7816.519946
std     26307.152791    16106.066042    8928.087314    9230.552561
min         0.000000        0.000000       0.000000       0.000000
25%         0.000000        0.000000     931.500000    2107.000000
50%       621.000000        0.000000    3202.000000    5138.000000
75%      4698.000000        3.000000    7604.500000   10418.750000
max    522188.000000   244550.000000  107224.000000  108058.000000

              other      multiple         admin  prepawn_handgun  \
```

```
count   6270.000000  13255.000000  13232.000000     11312.000000
mean     406.253748    275.643606     56.639586         4.942716
std     1317.089840    777.614386    588.479237        11.051790
min        0.000000      0.000000      0.000000         0.000000
25%       20.000000     15.000000      0.000000         0.000000
50%      140.000000    128.000000      0.000000         0.000000
75%      405.000000    307.000000      0.000000         5.000000
max    77929.000000  38907.000000  28083.000000       164.000000

       prepawn_long_gun  prepawn_other      ...        returned_other  \
count      11310.000000    5885.000000      ...           2585.000000
mean           7.669054       0.227867      ...              1.163250
std           16.055155       1.120344      ...              4.463956
min            0.000000       0.000000      ...              0.000000
25%            0.000000       0.000000      ...              0.000000
50%            1.000000       0.000000      ...              0.000000
75%            8.000000       0.000000      ...              0.000000
max          269.000000      49.000000      ...             64.000000

       rentals_handgun  rentals_long_gun  private_sale_handgun  \
count      1760.000000       1595.000000           3520.000000
mean          0.136364          0.124765             20.357955
std           0.965368          0.825477             86.553594
min           0.000000          0.000000              0.000000
25%           0.000000          0.000000              0.000000
50%           0.000000          0.000000              0.000000
75%           0.000000          0.000000              7.000000
max          13.000000         12.000000           1017.000000

       private_sale_long_gun  private_sale_other  return_to_seller_handgun  \
count            3520.000000         3520.000000               3245.000000
mean               16.705966            1.725284                  0.598151
std                68.280997            6.489753                  2.887213
min                 0.000000            0.000000                  0.000000
25%                 0.000000            0.000000                  0.000000
50%                 0.000000            0.000000                  0.000000
75%                 9.000000            1.000000                  0.000000
max               913.000000          111.000000                 70.000000

       return_to_seller_long_gun  return_to_seller_other         totals
count                3520.000000             3025.000000   13255.000000
mean                    0.630114                0.109091   22620.911656
std                     2.594418                0.416270   35420.511907
min                     0.000000                0.000000       0.000000
25%                     0.000000                0.000000    4747.000000
50%                     0.000000                0.000000   12692.000000
75%                     0.000000                0.000000   26384.000000
max                    56.000000                4.000000  541978.000000
```

```
[8 rows x 25 columns]

Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 13255 entries, 0 to 13254
Data columns (total 27 columns):
month                        13255 non-null object
state                        13255 non-null object
permit                       13231 non-null float64
permit_recheck               1870 non-null float64
handgun                      13235 non-null float64
long_gun                     13236 non-null float64
other                        6270 non-null float64
multiple                     13255 non-null int64
admin                        13232 non-null float64
prepawn_handgun              11312 non-null float64
prepawn_long_gun             11310 non-null float64
prepawn_other                5885 non-null float64
redemption_handgun           11315 non-null float64
redemption_long_gun          11314 non-null float64
redemption_other             5885 non-null float64
returned_handgun             2970 non-null float64
returned_long_gun            2915 non-null float64
returned_other               2585 non-null float64
rentals_handgun              1760 non-null float64
rentals_long_gun             1595 non-null float64
private_sale_handgun         3520 non-null float64
private_sale_long_gun        3520 non-null float64
private_sale_other           3520 non-null float64
return_to_seller_handgun     3245 non-null float64
return_to_seller_long_gun    3520 non-null float64
return_to_seller_other       3025 non-null float64
totals                       13255 non-null int64
dtypes: float64(23), int64(2), object(2)
memory usage: 2.7+ MB
None

Duplicates: 0
```

In [298]:
```
# Inspect census data using inspect_df function.

inspect_df(df_census)
```

```
Header:
                                                     0   \
```

```
Fact       Population estimates, July 1, 2016,  (V2016)
Fact Note                                           NaN
Alabama                                       4,863,300
Alaska                                          741,894
Arizona                                       6,931,071

                                                    1  \
Fact       Population estimates base, April 1, 2010,  (V2...
Fact Note                                                NaN
Alabama                                            4,780,131
Alaska                                               710,249
Arizona                                            6,392,301

                                                    2  \
Fact       Population, percent change - April 1, 2010 (es...
Fact Note                                                NaN
Alabama                                                1.70%
Alaska                                                 4.50%
Arizona                                                8.40%

                                      3  \
Fact       Population, Census, April 1, 2010
Fact Note                                NaN
Alabama                            4,779,736
Alaska                               710,231
Arizona                            6,392,017

                                                    4  \
Fact       Persons under 5 years, percent, July 1, 2016, ...
Fact Note                                                NaN
Alabama                                                6.00%
Alaska                                                 7.30%
Arizona                                                6.30%

                                                5  \
Fact       Persons under 5 years, percent, April 1, 2010
Fact Note                                            NaN
Alabama                                            6.40%
Alaska                                             7.60%
Arizona                                            7.10%

                                                    6  \
Fact       Persons under 18 years, percent, July 1, 2016,...
Fact Note                                                NaN
Alabama                                               22.60%
Alaska                                                25.20%
Arizona                                               23.50%
```

```
                                                 7  \
Fact       Persons under 18 years, percent, April 1, 2010
Fact Note                                            NaN
Alabama                                           23.70%
Alaska                                            26.40%
Arizona                                           25.50%

                                                    8  \
Fact       Persons 65 years and over, percent,  July 1, 2...
Fact Note                                               NaN
Alabama                                              16.10%
Alaska                                               10.40%
Arizona                                              16.90%

                                                    9  \
Fact       Persons 65 years and over, percent, April 1, 2010
Fact Note                                               NaN
Alabama                                              13.80%
Alaska                                                7.70%
Arizona                                              13.80%

                          ...                          75  \
Fact                      ...                         NaN
Fact Note                 ...                         NaN
Alabama                   ...                         NaN
Alaska                    ...                         NaN
Arizona                   ...                         NaN

                   76                                              77  \
Fact       Value Flags                                              -
Fact Note          NaN  Either no or too few sample observations were ...
Alabama            NaN                                            NaN
Alaska             NaN                                            NaN
Arizona            NaN                                            NaN

                                                    78  \
Fact                                                 D
Fact Note  Suppressed to avoid disclosure of confidential...
Alabama                                             NaN
Alaska                                              NaN
Arizona                                             NaN

                           79                                      80  \
Fact                        F                                      FN
Fact Note  Fewer than 25 firms  Footnote on this item in place of data
Alabama                   NaN                                     NaN
Alaska                    NaN                                     NaN
Arizona                   NaN                                     NaN
```

```
                      81                                               82  \
Fact                 NaN                                                S   
Fact Note  Not available  Suppressed; does not meet publication standards   
Alabama              NaN                                              NaN   
Alaska               NaN                                              NaN   
Arizona              NaN                                              NaN   

                       83                                                 84  
Fact                    X                                                  Z  
Fact Note  Not applicable  Value greater than zero but less than half uni...  
Alabama               NaN                                                NaN  
Alaska                NaN                                                NaN  
Arizona               NaN                                                NaN  

[5 rows x 85 columns]

Shape: (52, 85)

Statistics:
                0       1      2          3      4      5       6       7  \
count          51      51     51         51     51     51      51      51   
unique         51      51     43         51     23     24      39      33   
top     6,811,779  672591  1.70%  8,791,894  6.40%  6.50%  23.30%  22.30%   
freq            1       1      3          1      6      5       3       3   

             8       9 ...   75           76 77  \
count       51      51 ...  0.0            1  2   
unique      38      35 ...  0.0            1  2   
top     15.00%  13.50% ...  NaN  Value Flags  -   
freq         5       5 ...  NaN            1  1   

                                                       78 79  80  \
count                                                   2  2   2   
unique                                                  2  2   2   
top     Suppressed to avoid disclosure of confidential...  F  FN   
freq                                                    1  1   1   

                   81                                               82 83 84  
count               1                                                2  2  2  
unique              1                                                2  2  2  
top     Not available  Suppressed; does not meet publication standards  X  Z  
freq                1                                                1  1  1  

[4 rows x 85 columns]

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

```
Index: 52 entries, Fact to Wyoming
Data columns (total 85 columns):
0     51 non-null object
1     51 non-null object
2     51 non-null object
3     51 non-null object
4     51 non-null object
5     51 non-null object
6     51 non-null object
7     51 non-null object
8     51 non-null object
9     51 non-null object
10    51 non-null object
11    51 non-null object
12    52 non-null object
13    52 non-null object
14    52 non-null object
15    52 non-null object
16    52 non-null object
17    51 non-null object
18    52 non-null object
19    51 non-null object
20    51 non-null object
21    51 non-null object
22    51 non-null object
23    51 non-null object
24    51 non-null object
25    51 non-null object
26    51 non-null object
27    51 non-null object
28    51 non-null object
29    51 non-null object
30    51 non-null object
31    51 non-null object
32    51 non-null object
33    51 non-null object
34    51 non-null object
35    51 non-null object
36    51 non-null object
37    51 non-null object
38    51 non-null object
39    51 non-null object
40    52 non-null object
41    52 non-null object
42    52 non-null object
43    52 non-null object
44    52 non-null object
45    52 non-null object
```

```
46    51 non-null object
47    51 non-null object
48    51 non-null object
49    51 non-null object
50    52 non-null object
51    52 non-null object
52    52 non-null object
53    52 non-null object
54    51 non-null object
55    51 non-null object
56    51 non-null object
57    51 non-null object
58    51 non-null object
59    51 non-null object
60    51 non-null object
61    51 non-null object
62    51 non-null object
63    51 non-null object
64    51 non-null object
65    0 non-null object
66    1 non-null object
67    0 non-null object
68    1 non-null object
69    2 non-null object
70    0 non-null object
71    1 non-null object
72    2 non-null object
73    2 non-null object
74    2 non-null object
75    0 non-null object
76    1 non-null object
77    2 non-null object
78    2 non-null object
79    2 non-null object
80    2 non-null object
81    1 non-null object
82    2 non-null object
83    2 non-null object
84    2 non-null object
dtypes: object(85)
memory usage: 37.4+ KB
None

Duplicates: 0
```

```
In [299]: # Inspect shootings data using inspect_df function.
```

```
inspect_df(df_shootings)
```

```
Header:
                                case     location     date  shootings
0           Mercy Hospital shooting     Illinois  2018-11          1
1  Thousand Oaks nightclub shooting   California  2018-11          1
2   Tree of Life synagogue shooting  Pensylvania  2018-10          1
3       Rite Aid warehouse shooting     Maryland  2018-09          1
4             T&T Trucking shooting   California  2018-09          1

Shape: (55, 4)

Statistics:
       shootings
count       55.0
mean         1.0
std          0.0
min          1.0
25%          1.0
50%          1.0
75%          1.0
max          1.0

Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 55 entries, 0 to 54
Data columns (total 4 columns):
case         55 non-null object
location     55 non-null object
date         55 non-null object
shootings    55 non-null int64
dtypes: int64(1), object(3)
memory usage: 1.8+ KB
None

Duplicates: 0
```

### 1.1.2 Data Cleaning (Replace this with more specific notes!)

For the NICS data set, I decided to focus only on the data for background checks conducted for sales by Federal licensced firearms dealers. I chose not to use data about permits since the laws regarding permits and licensing vary greatly from state to state and thus are very inconsistent and not very enlightening. I also chose to ignore any background checks conducted by pawn shops or for firearm rentals since I wanted to focus only on sales. Background checks conducted by private sellers was also ignored since the regulations on private sales are very inconsistent in different states and therefor are not very meaningful. I also discarded all NICS data from "non-states" (Washington D.C., Guam, the Mariana Islands, Puerto Rico, and the Virgin Islands) since there is no corresponding data for the locales in the census data. I also trimmed the NICS data to focus only on background checks conducted from January, 2010 to present. After removing all unwanted data, the columns containing the numbers of background checks were converted to integer values only. I also created a new column "total" that shows the combined number of NICS background checks for all types of firearms by state each month as well as a new dataframe that shows the total NICS background checks by state from January, 2010 through November, 2018. I also created a new dataframe grouping combined states background checks by month for comparison with the shootings data.

For the census data, I transposed the dataframe so that the rows match those in the NICS dataframe, discarded the columns that I did not wish to explore, and renamed the remaining columns with simpler titles. I also dropped rows that did not contain any useful information, cleaned the data to remove any commas, dollar signs, or percent signs, and converted the columns to the apprpriate classes (either integers or floats). I also added two new columns, "adult population 2016", and "percent population veterans" to make my analysis easier.

For the shootings data I grouped by month to get shootings per month and renamed the columns. Then I joined the shooting data with the monthly gun data in a new dataframe. I grouped by month, dropped unwanted columns, replaced NaN values with 0's, and sorted by month.

In [300]:
```
# Drop unwanted columns from gun data.

df_gun = df_gun[['month','state', 'handgun', 'long_gun', 'other', 'multiple']]
```

In [301]:
```
# Drop unwanted rows (non-states) from gun data.

non_states = ['District of Columbia', 'Guam', 'Mariana Islands', 'Puerto Rico', 'Virgi
for non_state in non_states:
    df_gun = df_gun[df_gun['state']!= non_state]
```

In [302]:
```
# Trim data from before 2010.
drop_rows = np.r_[5350:12050]
df_gun.drop(df_gun.index[drop_rows], inplace=True)
```

In [303]:
```
# Convert all numbers to int in gun data.

int_columns = list(df_gun.columns[2:])
for col in int_columns:
    df_gun[col] = (df_gun[col].astype(int))
```

In [304]:
```
# Create new column for total NICS background checks by state each month.

df_gun['total'] = df_gun['handgun'] + df_gun['long_gun'] + df_gun['other'] + df_gun['m
df_gun.head()
```

Out[304]:

| | month | state | handgun | long_gun | other | multiple | total |
|---|---|---|---|---|---|---|---|
| 0 | 2018-11 | Alabama | 8408 | 10303 | 357 | 1153 | 20221 |
| 1 | 2018-11 | Alaska | 2707 | 2827 | 360 | 197 | 6091 |

```
2  2018-11     Arizona    12412      8886   1139       879  23316
3  2018-11    Arkansas     5739     10196    300       914  17149
4  2018-11  California    36997     30913   3606         0  71516
```

In [305]:
```
# Create new dataframe for total NICS background checks by state from 2010 to 2018 and
# handguns and long guns.

df_gun_total = df_gun.groupby(['state']).sum()
df_gun_total['percent handguns'] = ((df_gun_total['handgun'] / df_gun_total['total'])
df_gun_total['percent long_guns'] = ((df_gun_total['long_gun'] / df_gun_total['total']
df_gun_total.head()
```

Out[305]:
```
            handgun  long_gun   other  multiple    total  percent handguns  \
state
Alabama     1365431   1293826   42986     65474  2767717             49.33
Alaska       300604    326761   19497     19252   666114             45.13
Arizona     1199361    841046   72772     58213  2171392             55.23
Arkansas     597457    726279   16879     37535  1378150             43.35
California  3770517   3409235  394788         0  7574540             49.78

            percent long_guns
state
Alabama                 46.75
Alaska                  49.05
Arizona                 38.73
Arkansas                52.70
California              45.01
```

In [306]:
```
# Create new dataframe to show the total number of NICS checks by gun type since 2010.

a = sum(df_gun_total['handgun'])
b = sum(df_gun_total['long_gun'])
c = sum(df_gun_total['other'])
d = sum(df_gun_total['multiple'])
df_gun_sums = pd.DataFrame({'type': ['handgun', 'long_gun', 'other', 'multiple'], 'tot
df_gun_sums.head()
```

Out[306]:
```
      total      type
0  54624965   handgun
1  50780105  long_gun
2   2520060     other
3   2063859  multiple
```

In [307]:
```
# Create new dataframe for total NICS background checks by month from 2010 to 2018.

df_gun_monthly = df_gun.groupby(['month']).sum().reset_index()
df_gun_monthly.head()
```

Out[307]:
```
     month  handgun  long_gun  other  multiple   total
0  2010-01   288894    349482   4968     15939  659283
```

```
1  2010-02   359719    387092   5554     18596  770961
2  2010-03   367289    421136   5785     16815  811025
3  2010-04   313739    349245   4366     15411  682761
4  2010-05   274859    290315   3947     12003  581124
```

In [308]:
```
# Drop unwanted columns from census data.

df_census = df_census.iloc[:, np.r_[0, 6, 20:22, 34:36, 47, 49, 62]]
```

In [309]:
```
# Rename remaining columns in census data.

df_census.columns = ['population 2016', 'percent population under 18 2016', 'populatio
                     'percent population foreign born', 'percent population high schoo
                     'percent population bachelors or higher', 'median household incom
                     'percent population in poverty', 'population per sq mile']
```

In [310]:
```
# Drop unwanted rows from census data.

df_census.drop(['Fact','Fact Note'], axis=0, inplace=True)
```

In [311]:
```
# Remove all dollar signs, commas, and percent signs in census data.

df_census.replace(['\$',',','%'],['','',''], inplace=True, regex=True)
df_census.head()
```

Out[311]:
```
           population 2016 percent population under 18 2016  \
Alabama            4863300                            22.60
Alaska              741894                            25.20
Arizona            6931071                            23.50
Arkansas           2988248                            23.60
California        39250017                            23.20

           population veterans percent population foreign born  \
Alabama                 363170                            3.50
Alaska                   69323                            7.40
Arizona                 505794                           13.50
Arkansas                220953                            4.70
California             1777410                           27.00

           percent population high school grad or higher  \
Alabama                                            84.30
Alaska                                             92.10
Arizona                                            86.00
Arkansas                                           84.80
California                                         81.80

           percent population bachelors or higher median household income  \
Alabama                                     23.50                   43623
Alaska                                      28.00                   72515
```

```
Arizona                                            27.50                  50255
Arkansas                                           21.10                  41371
California                                         31.40                  61818

           percent population in poverty population per sq mile
Alabama                                 17.10                   94.4
Alaska                                   9.90                    1.2
Arizona                                 16.40                   56.3
Arkansas                                17.20                     56
California                              14.30                  239.1
```

In [312]:
```
# Convert census data to appropriate data types.

float_nums = np.r_[0:8]
float_cols = list(df_census.columns[float_nums])
for col in float_cols:
    df_census[col] = (df_census[col].astype(float))

int_nums = np.r_[0, 2, 6]
int_cols = list(df_census.columns[int_nums])
for col in int_cols:
    df_census[col] = (df_census[col].astype(int))
```

In [313]:
```
# Add new columns to census data using other columns for calculations.

# Calculate adult population.
df_census['adult population 2016'] = ((((100 - df_census['percent population under 18

# Calculate % population veterans.
df_census['percent population veterans'] = (((df_census['population veterans'] / df_ce

df_census.head()
```

Out[313]:
```
            population 2016  percent population under 18 2016  \
Alabama             4863300                              22.6
Alaska               741894                              25.2
Arizona             6931071                              23.5
Arkansas            2988248                              23.6
California         39250017                              23.2

            population veterans  percent population foreign born  \
Alabama                  363170                              3.5
Alaska                    69323                              7.4
Arizona                  505794                             13.5
Arkansas                 220953                              4.7
California              1777410                             27.0

            percent population high school grad or higher  \
```

```
Alabama                                                    84.3
Alaska                                                     92.1
Arizona                                                    86.0
Arkansas                                                   84.8
California                                                 81.8

            percent population bachelors or higher  median household income  \
Alabama                                       23.5                    43623
Alaska                                        28.0                    72515
Arizona                                       27.5                    50255
Arkansas                                      21.1                    41371
California                                    31.4                    61818

            percent population in poverty population per sq mile  \
Alabama                              17.1                     94.4
Alaska                                9.9                      1.2
Arizona                              16.4                     56.3
Arkansas                             17.2                       56
California                           14.3                    239.1

            adult population 2016  percent population veterans
Alabama                   3764194                         7.47
Alaska                     554936                         9.34
Arizona                   5302269                         7.30
Arkansas                  2283021                         7.39
California               30144013                         4.53
```

In [314]:
```python
# Group shootings data by month and rename columns.

df_shootings = df_shootings.groupby(['date']).sum().reset_index()
df_shootings.columns = ['month', 'shootings']
```

In [315]:
```python
# Join shootings dataframe with monthly gun dataframe, group by month, drop unwanted c
# and sort by month.

df_shootings_monthly = pd.merge(df_shootings, df_gun_monthly, on='month', how='outer')
df_shootings_monthly.groupby(['month'])
df_shootings_monthly = df_shootings_monthly[['month', 'shootings']]
df_shootings_monthly.fillna(0, inplace=True)
df_shootings_monthly.sort_values(['month'], inplace=True)
df_shootings_monthly.head()
```

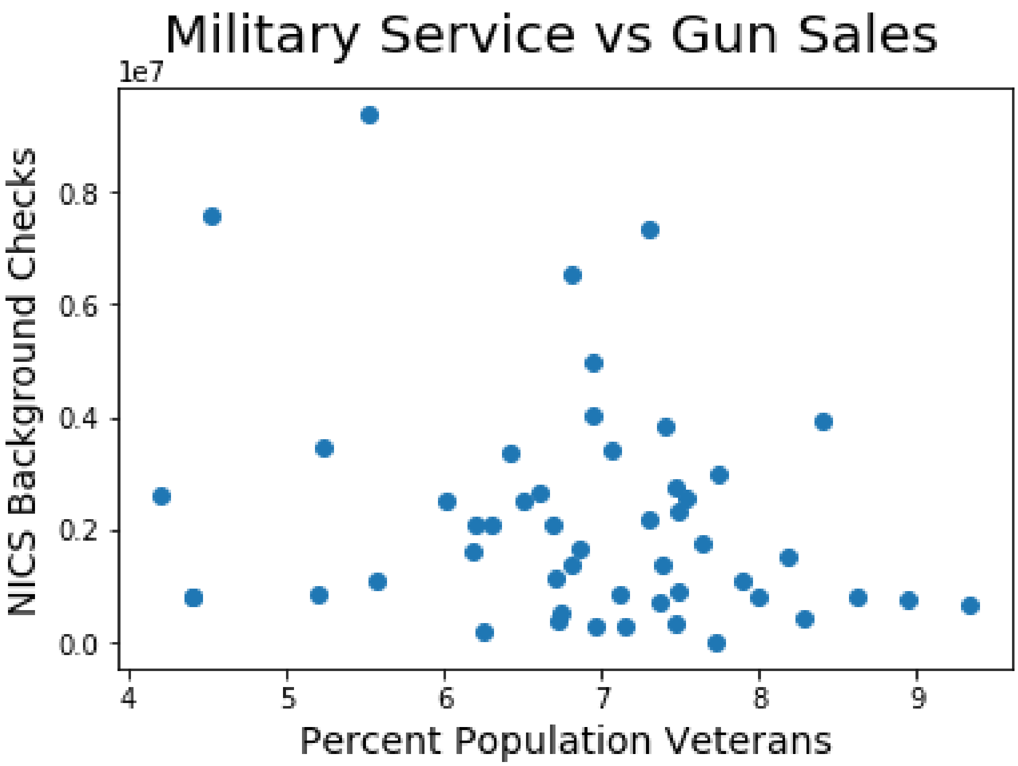
Out[315]:

| | month | shootings |
|---|---|---|
| 43 | 2010-01 | 0.0 |
| 44 | 2010-02 | 0.0 |
| 45 | 2010-03 | 0.0 |
| 46 | 2010-04 | 0.0 |
| 47 | 2010-05 | 0.0 |

## Exploratory Data Analysis

### 1.1.3 Is there a correlation between military service and gun sales?

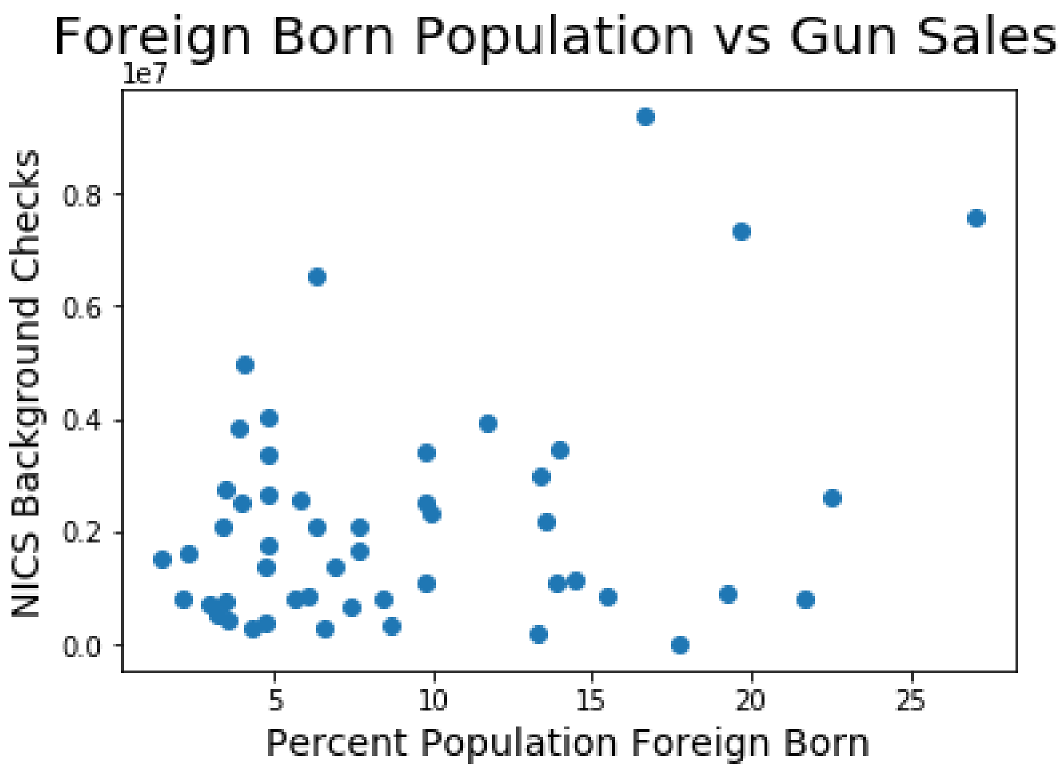
In [316]:
```
# Compare percent veteran population of states with NICS background checks.

fig = plt.figure()
plt.scatter(df_census['percent population veterans'], df_gun_total['total'])
fig.suptitle('Military Service vs Gun Sales', fontsize=20)
plt.xlabel('Percent Population Veterans', fontsize=14)
plt.ylabel('NICS Background Checks', fontsize=14);
```




There does not seem to be much of a correlation between a state's percent of population with military service and NICS checks. The data shows somewhat of a normal distribution with the possibility of a slight inverse relationship.

### 1.1.4 Is there a correlation between amount of foreign born individuals and gun sales?

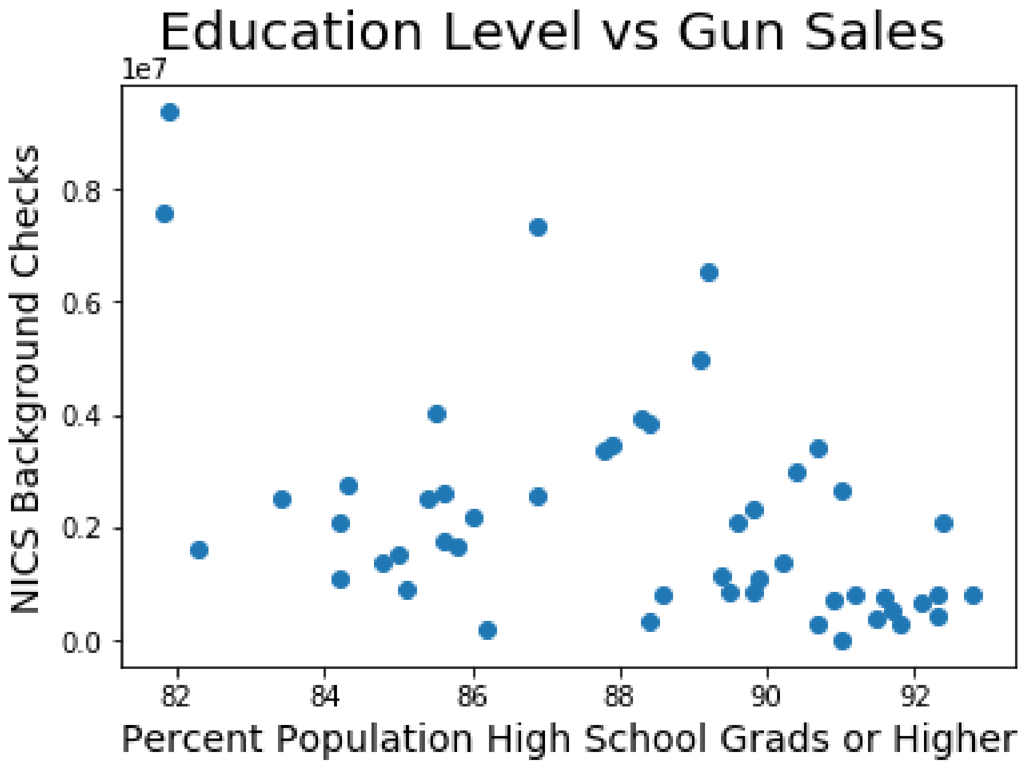
In [317]:
```
# Compare percent foreign born population of states with NICS background checks.

fig = plt.figure()
plt.scatter(df_census['percent population foreign born'], df_gun_total['total'])
fig.suptitle('Foreign Born Population vs Gun Sales', fontsize=20)
plt.xlabel('Percent Population Foreign Born', fontsize=14)
plt.ylabel('NICS Background Checks', fontsize=14);
```




There does not seem to be a correlation between a state's foreign born population and NICS checks, although the 3 states with the highest number of NICS checks do happen to have a high percentage of foreign born citizens.

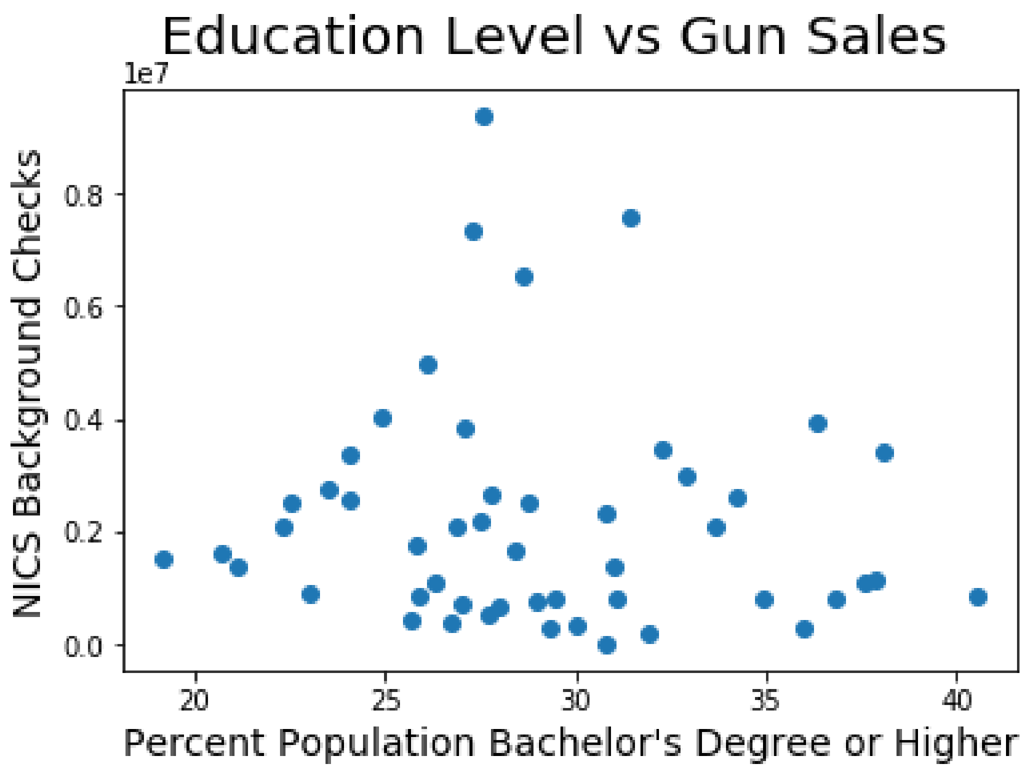
#### 1.1.5 Is there a correlation between education level and gun sales?

```
In [318]: # Compare percent high school graduates with NICS background checks.

          fig = plt.figure()
          plt.scatter(df_census['percent population high school grad or higher'], df_gun_total['
          fig.suptitle('Education Level vs Gun Sales', fontsize=20)
          plt.xlabel('Percent Population High School Grads or Higher', fontsize=14)
          plt.ylabel('NICS Background Checks', fontsize=14);
```




There does not seem to be much of a correlation between a state's percentage of population with a high school education or higher and NICS checks, although the two states with the lowest percentage of high school graduates also happen to have the highest number of NICS checks.

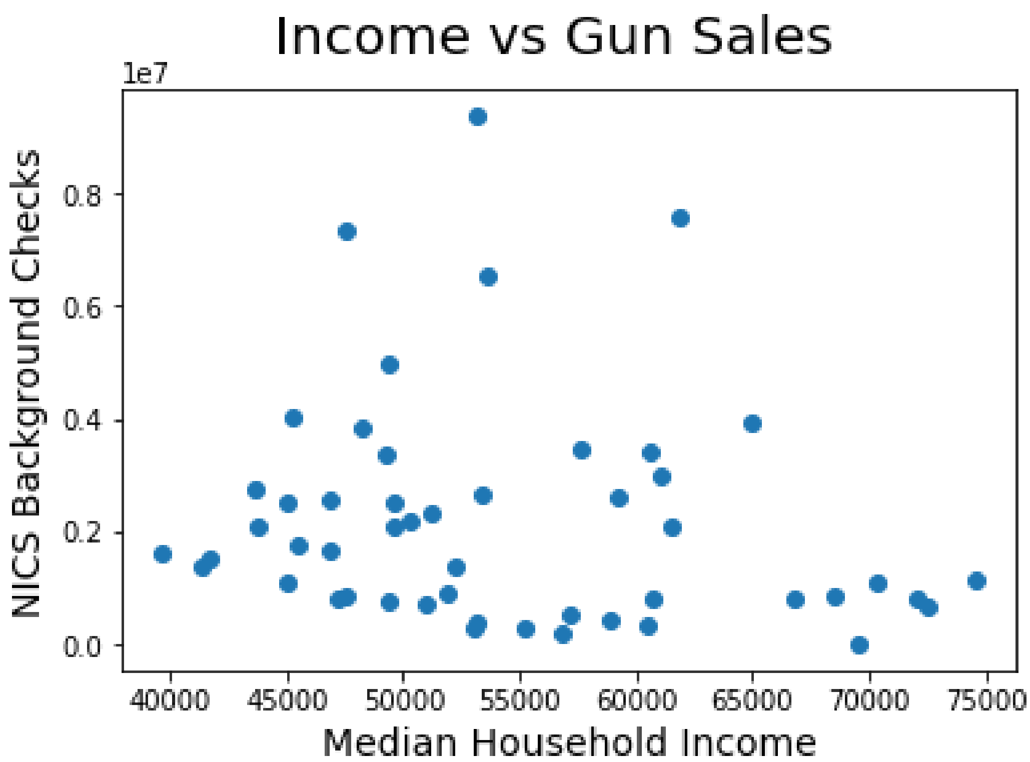
```
In [319]: # Compare percent bachelors or higher with NICS background checks.

          fig = plt.figure()
          plt.scatter(df_census['percent population bachelors or higher'], df_gun_total['total']
          fig.suptitle('Education Level vs Gun Sales', fontsize=20)
          plt.xlabel("Percent Population Bachelor's Degree or Higher", fontsize=14)
          plt.ylabel('NICS Background Checks', fontsize=14);
```



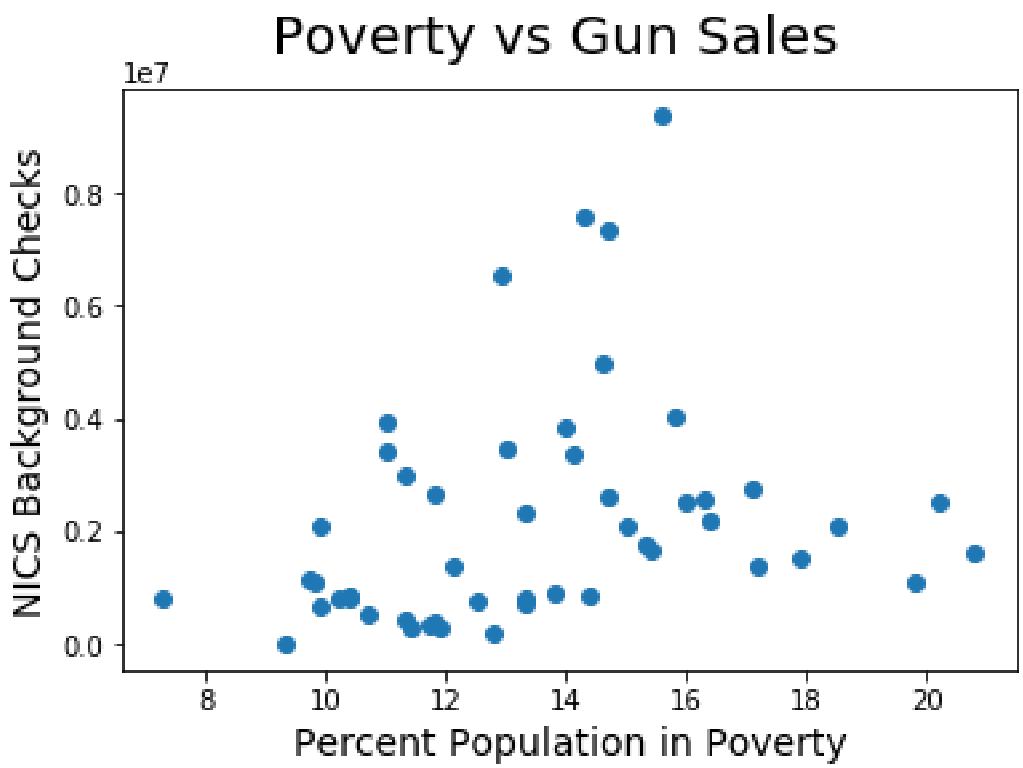

There seems to be no correlation between a state's percentage of population with a college education and NICS checks, the data shows a fairly normal distribution.

#### 1.1.6 Is there a correlation between economic status and gun sales?

In [320]:
```
# Compare household income with NICS background checks.

fig = plt.figure()
plt.scatter(df_census['median household income'], df_gun_total['total'])
fig.suptitle('Income vs Gun Sales', fontsize=20)
plt.xlabel('Median Household Income', fontsize=14)
plt.ylabel('NICS Background Checks', fontsize=14);
```



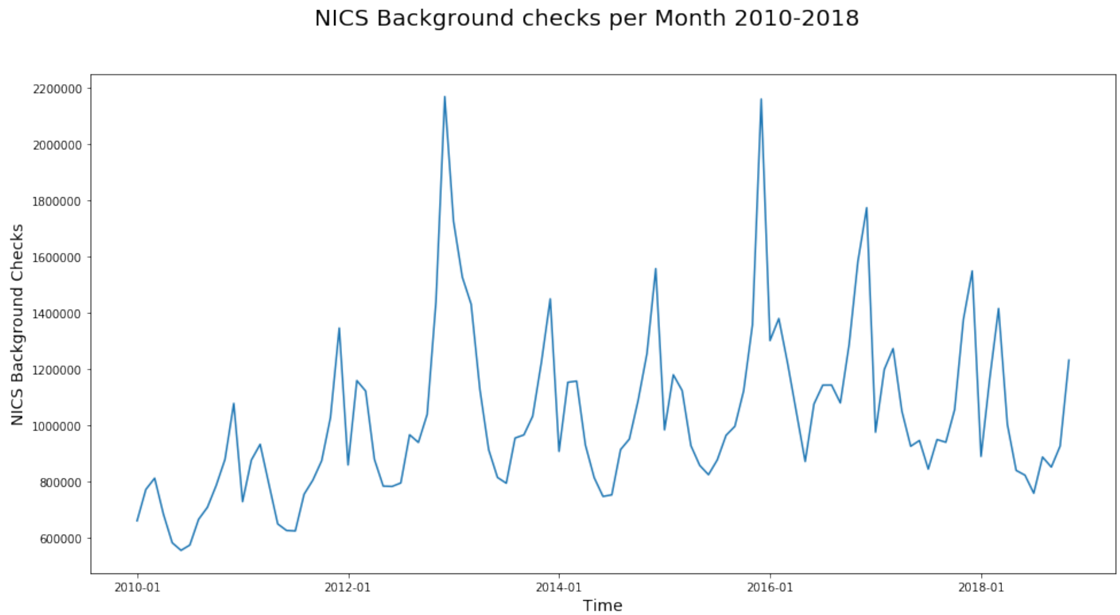

In [321]:
```
# Compare percent of population at poverty level with NICS background checks.

fig = plt.figure()
plt.scatter(df_census['percent population in poverty'], df_gun_total['total'])
fig.suptitle('Poverty vs Gun Sales', fontsize=20)
plt.xlabel('Percent Population in Poverty', fontsize=14)
plt.ylabel('NICS Background Checks', fontsize=14);
```



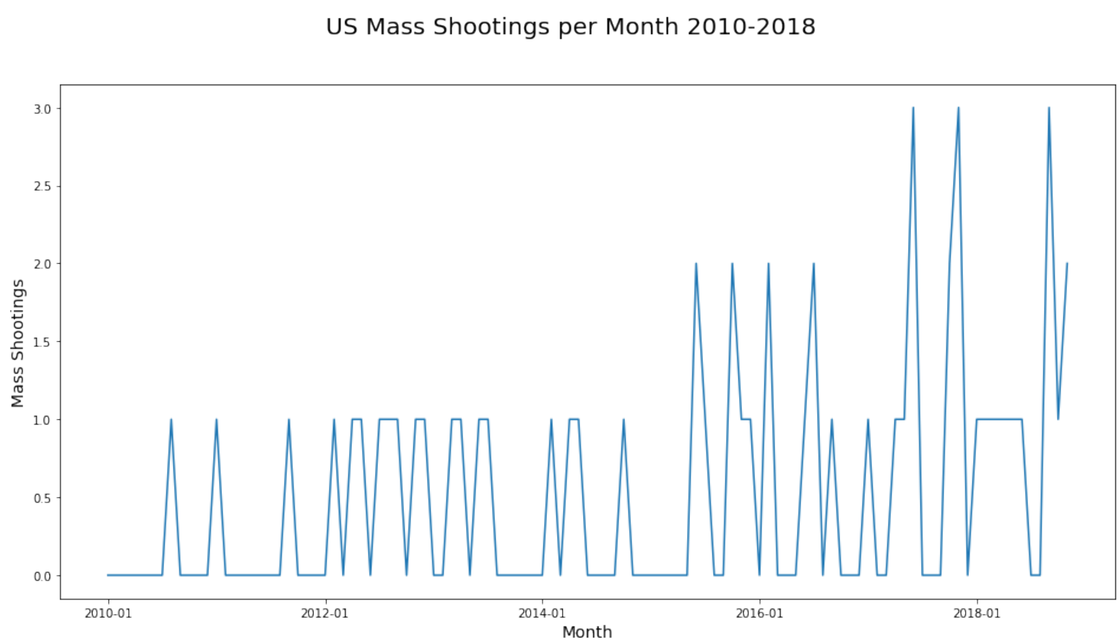

Neither a state's median household income, or the state's percent of population in poverty seem to have any correlation with NICS checks.

### 1.1.7 Are gun sales affected by mass shootings?

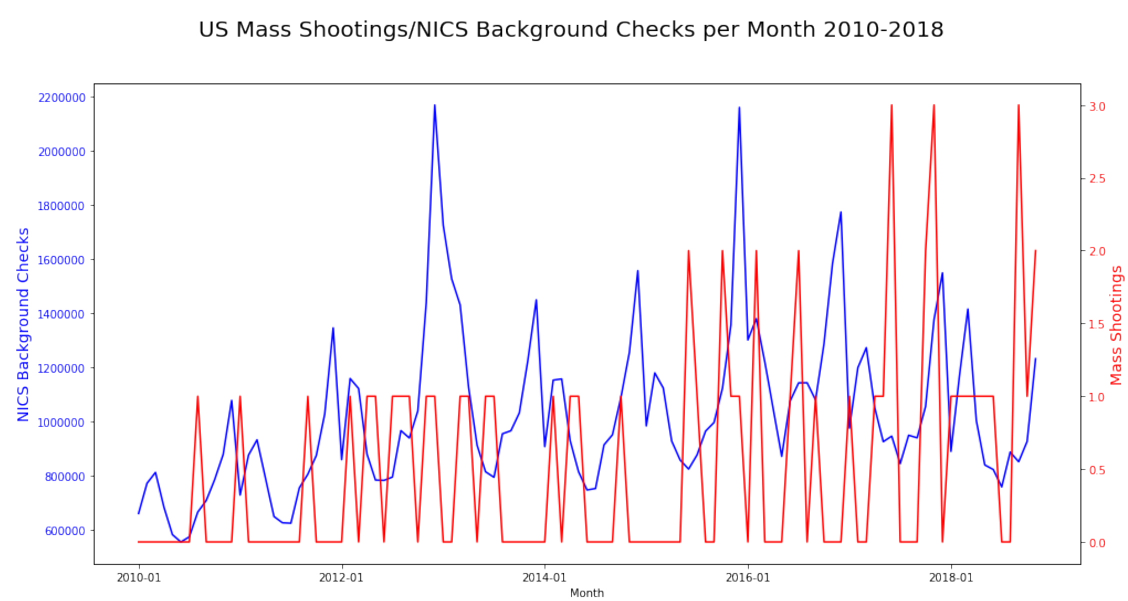
In [322]:
```
# Plot combined NICS background checks per month since 2010.

fig = plt.figure(figsize=(16, 8))
plt.plot(df_gun_monthly['month'],df_gun_monthly['total'])
plt.xticks(['2010-01', '2012-01', '2014-01', '2016-01', '2018-01'], ['2010-01', '2012-
fig.suptitle('NICS Background checks per Month 2010-2018', fontsize=20)
plt.xlabel('Time', fontsize=14)
plt.ylabel('NICS Background Checks', fontsize=14);
```




```
In [323]: # Plot occurence of mass shootings monthly since 2010.

          fig = plt.figure(figsize=(16, 8))
          plt.plot(df_shootings_monthly['month'], df_shootings_monthly['shootings'])
          plt.xticks(['2010-01', '2012-01', '2014-01', '2016-01', '2018-01'], ['2010-01', '2012-
          fig.suptitle('US Mass Shootings per Month 2010-2018', fontsize=20)
          plt.xlabel('Month', fontsize=14)
          plt.ylabel('Mass Shootings', fontsize=14);
```



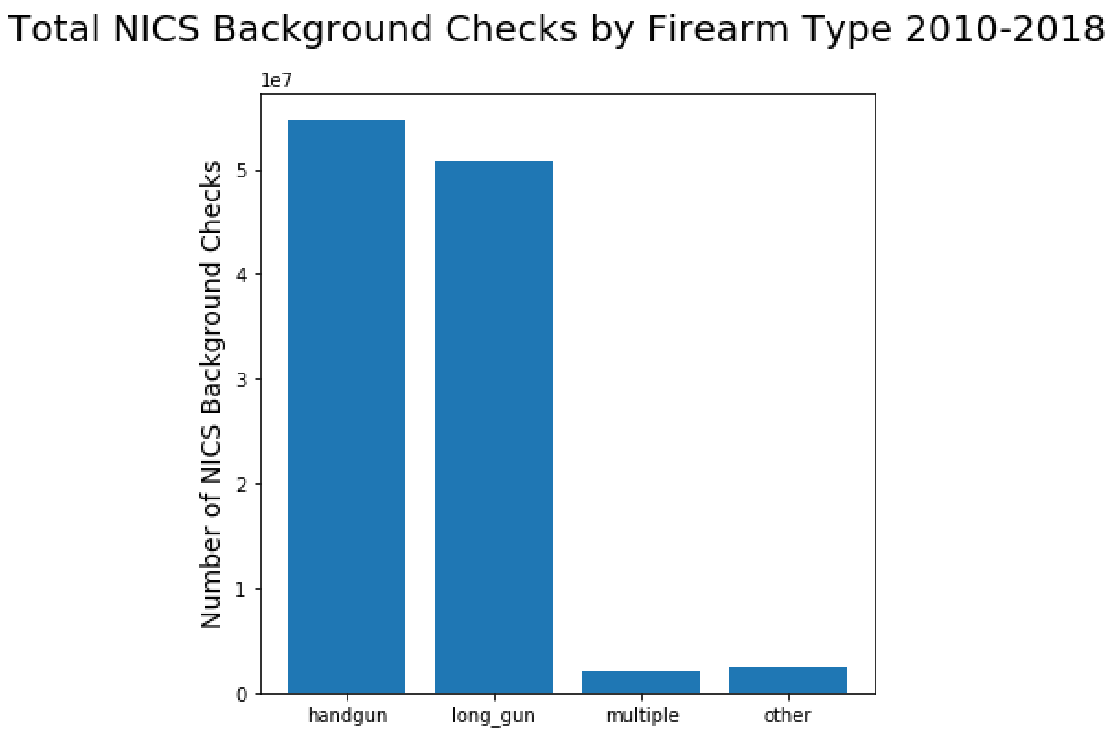
```
In [324]: # Plot NICS monthly background checks data and US monthly mass shootings data together

          fig, ax1 = plt.subplots(figsize=(16, 8))
          fig.suptitle('US Mass Shootings/NICS Background Checks per Month 2010-2018', fontsize=

          color = 'blue'
          ax1.set_xlabel('Month')
          ax1.set_ylabel('NICS Background Checks', color=color, fontsize=14)
          ax1.plot(df_gun_monthly['month'], df_gun_monthly['total'], color=color)
          ax1.tick_params(axis='y', labelcolor=color)

          ax2 = ax1.twinx()

          color = 'red'
          ax2.set_ylabel('Mass Shootings', color=color, fontsize=14)
          ax2.plot(df_shootings_monthly['month'], df_shootings_monthly['shootings'], color=color
          ax2.tick_params(axis='y', labelcolor=color)

          plt.xticks(['2010-01', '2012-01', '2014-01', '2016-01', '2018-01'], ['2010-01', '2012-
```




There often seems to be a reaction among people after mass shootings to move for stricter gun control laws. This in turn seems to lead to people rushing to purchase firearms before any new restrictions can prevent them from doing so. Comparing the instance of mass shootings in the US and the number of NICS checks nationwide monthly since 2010 shows that there may be some truth to this. It is not a perfect correlation, but there does seem to be a tendency for NICS checks to spike shortly after the occurence of mass shootings.

### 1.1.8 How do sales of different types of guns compare since 2010?

```
In [325]: # Make chart of total number of NICS checks related to different gun categories since

          fig = plt.figure(figsize=(6, 6))
          plt.bar(df_gun_sums['type'], df_gun_sums['total'])
          fig.suptitle('Total NICS Background Checks by Firearm Type 2010-2018', fontsize=20)
          plt.ylabel('Number of NICS Background Checks', fontsize=14);
```




The number of checks related to handgun sales vs long gun sales is nearly the same with handguns being slightly greater. The number of checks related to "other" or "multiple" are both quite low compared with handguns and long guns.

## Conclusions

In conclusion, there does not seem to be any significant correlation between any of the state demographics investigated and a state's NICS backgound checks. There does, however, seem to be some truth to the idea that mass shootings tend to lead to more attempts to purchase firearms. Also, the number of NICS background checks related to handgun sales versus long gun sales are nearly equal. These results are merely an observation of correlation though, and do not imply any causation or statistical significance. Also, as noted previously, the data for NICS background checks is not a direct measurement of firearms sales in the US since not all checks result in sales and not all sales have a corresponding NICS check. The fact that the NICS background checks is not a direct measurement of firearms sales is a significant limitation for this analysis since we cannot be sure what number of the NICS checks actually resulted in sales. Another limitation for this analysis is that there are likely to be a significant number of gun sales or exchanges that did not involve a NICS background check due to the sale either being private in a state that does not require checks for private sales, or sales that were not conducted legally.

### 1.2 References

BuzzFeedNews nics-firearm-background-checks GitHub repository https://github.com/BuzzFeedNews/nics-firearm-background-checks

MotherJones Mass Shootings Data https://www.motherjones.com/politics/2012/12/mass-shootings-mother-jones-full-data/

US Census Data https://d17h27t6h515a5.cloudfront.net/topher/2017/November/5a0a554c_u.s.-census-data/u.s.-census-data.csv

Wikipedia National Instant Criminal Background Check System https://en.wikipedia.org/wiki/National_Instant_Criminal_Background_Check_System

In [326]:
```
from subprocess import call
call(['python', '-m', 'nbconvert', 'Investigate_a_Dataset.ipynb'])
```

Out[326]: 0