

# **An Analysis of New York State's DUIs as Compared to the Prevalence of Liquor Licenses in each County**

by Andrew (aoe3), Gabe (gjp74) and Ishaan (ib223)

The objectives of our project were two-fold:

1. To analyze the relationship between income, population and the number of liquor licenses across the different counties in New York State.
2. To understand whether there is a direct relationship between number of liquor licenses and number of impaired driving incidents in the different counties of New York.

Our hypotheses were:

1. Low Income counties will have a higher proportional of liquor licenses.
2. Counties with a higher number of liquor licenses will have a higher number of impaired driving incidents.

## **Data Sources**

- The data for population and for per capita income was obtained from Wikipedia. [https://en.wikipedia.org/wiki/New\\_York\\_locations\\_by\\_per\\_capita\\_income](https://en.wikipedia.org/wiki/New_York_locations_by_per_capita_income)
- The data for number of liquor licenses was obtained from the NYS government data website. <https://data.ny.gov/Economic-Development/Liquor-Authority-Quarterly-List-of-Active-Licenses/hrvs-fxs2>
- The data for the number of impaired driving incidents was obtained from the New York State and County Traffic Safety Data Reports for 2014. <http://www.safeny.ny.gov/14data/datapack14.htm>
- A Json was found online for the New York state counties, and permission was obtained from the owner to use it. A screenshot of the email has been attached at the bottom of this report.

## Mapping

From the Json we plotted 3 data maps and 2 comparison maps, which were used to create maps for our first objective: analyzing the relationship between income, population and the number of liquor licenses in the different counties in New York State.

The 3 data maps were to display the income, population and number of liquor licenses in the different counties.

- For the income map, we used an exponential scale (`Math.pow()` function), in order to highlight differences in income levels. Prior to implementing the scale, there was too soft a difference between many counties in NYS, while the city area was heavily highlighted, due to the discrepancies in income.
- For the population map, we multiplied each county's population value by 2, so as to better highlight the areas with higher populations as compared to more rural areas.
- For the liquor licenses map, we once again scaled the number of licenses in each county by multiplying by 2, so as to better highlight differences.

The 2 comparison maps were used to display the ratio of number of liquor licenses to population and income respectively, in all the different counties.

- For the first comparison map, we plotted the number of liquor licenses per 100 residents in each county. This scale was chosen due to our observation that the ratio of licenses to people was 1:100 in Hamilton, New York. It was the highest ratio of any county, and we wanted to highlight it. Therefore, we multiplied the value for number of licenses in each county by 100.
- For the second comparison map, we plotted number of liquor licenses vs. income levels. For this map, we had to scale the liquor license values by 100 in order to see the differences in hue, as the populations of each county were far greater than the number of licenses in each county. Had we left the number of licenses alone for each county, the map would look very pale.

For all five of these maps, we used a color scale to display the numerical value that different shades of the gradient corresponded to. These can be viewed on the right of each graph, and will assist the reader with understanding what the different color gradients on the map represent. For the JSON, we had to apply a translation and a scale to position each map appropriately.

The last 3 graphs were to fulfill our second objective: analyzing the relationship between number of liquor licenses and number of impaired driving incidents in the different counties.

As mentioned earlier, the income data was taken from Wikipedia and the liquor license data was taken from the NYS open data website. We split the liquor licenses into two categories: on-premise licenses and retail licenses. The purpose for this split is that we wanted to see if different population or income demographics would dictate a difference in the makeup of liquor licenses. We also wanted to be able to see if there was a different effect on impaired driving tickets given different types of liquor licenses. We thought on-premise licenses would be related to a greater amount of impaired driving tickets because, in this case, drivers need to drive home after consuming.

License and income data was imported from CSV files. A series of lookup functions was generated to get access to different parts of the datasets. The datasets were then combined to display different information in either absolute or relative ways. The scatter plots used data from the Json file. We had hard-coded the information regarding the amount of impaired while driving tickets given to drivers in specific counties into the Json. This data was combined with the liquor license data mentioned above to create a plot showing the relationship between the amounts of liquor licenses per capita and the amount of tickets per capita.

The scatter plot graphs have the number of liquor licenses on the x-axis and the number of impaired tickets issued on the y-axis. The y-axis is scaled to number of tickets per 1000 individuals in the county, which ranges from about 2-13. The x-axis is also scaled down to the number of liquor licenses per 1000 individuals in the county; the max value of the x-axis is different for each graph so that the trends can be seen. Finally, the size of the points is scaled to the log of the population divided by 1000. This allows for a small difference in the size of the different circles to show the different total population in each county. We have ignored counties with a population less than 30,000 people, because even a very small amount of liquor licenses or tickets issued skewed the results and made the correlation seem much stronger than it is in reality.

## Analysis

The number of liquor licenses seems to be uncorrelated with per capita income, as the data maps for liquor licenses per county and income levels per capita appear vastly different. This is reinforced by the fact that the comparison map for liquor licenses versus per capita income (Comparison Map 1) looks almost exactly the same as the original liquor licenses map, which implies that the number of liquor licenses does not depend on the income level of the county. ***This varies from our initial hypothesis that low-income counties will have a larger concentration of liquor licenses.***

The number of liquor licenses seems to be largely proportional to the population of each county, as most counties are the same shade of pink in our Comparison Map 2. This tells us that the number of liquor licenses issued per 100 residents is mostly constant across New York State. Hamilton County shows up as being the highest in terms of the proportion of liquor licenses to population - this was not evident by any of the three data graphs! This is due to the estimated population of Hamilton being roughly 4,800 while the number of active liquor licenses is 46, equating to about 1 alcohol-serving establishment per 100 people

The first scatter plot shows a slight positive correlation between number of liquor licenses and number of impaired driving tickets. ***This supports our second hypothesis.*** Scatter Plot 2 shows that the correlation is even stronger between number of on-premise liquor licenses and number of impaired driving tickets per county, which is again something we had hypothesized. Scatter Plot 3 shows that the co-relation is not as strong between number of retail liquor licenses and number of impaired driving tickets.

## Permission email from the Json owner

Re: Permission to use your JSON for New York State Counties Map



Inbox x



**Mark DiMarco** <mark.dimarco@gmail.com>

to Andrew

Mar 1 (6 days ago)



This message may not have been sent by: mark.dimarco@gmail.com [Learn more](#) [Report phishing](#)

Hi Andrew,

You have my permission to use those resources. If possible, please send me a link to your finished project, I'd love to have a look.

Best of luck,  
Mark DiMarco

On Tuesday, March 1, 2016, Andrew Eccles <[ae3@cornell.edu](mailto:ae3@cornell.edu)> wrote:

Hello Mr. DiMarco,

My name is Andrew Eccles. I am a student in a data visualisation course at Cornell University (INFO 3300). I am writing to you, requesting permission to use the JSON for the counties in New York State, found here: <http://bl.ocks.org/markmarkoh/8717334>

This is for a class project, where my group will be comparing the number of liquor stores per capita in each county with the number of recorded DUI traffic stops per capita in each county. I only need the JSON, as I am required to write my own code to draw the map. I will cite your JSON in both my .js file (as a comment) and in the .html file (header/paragraph/footer; your choice as to where).

Thank you for your time,

Andrew Eccles

Cornell University Class of 2017  
College of Arts & Sciences  
Information Sciences