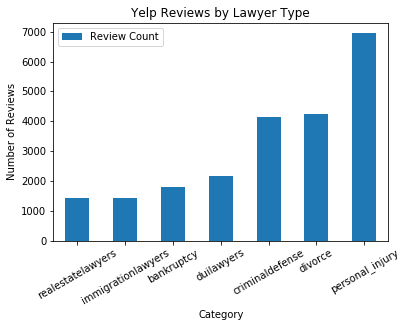
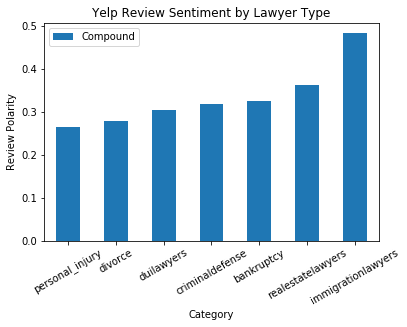
**What types of lawyers are Yelped about most?**

The Yelp API segments lawyers within professional services. Lawyers are then categorized based on type of law they practice. The API distinguishes 21 different types of law practice. Cities were then chosen for the study randomly to achieve a representative sampling of the US population. Finally, reviews were totaled by category. This analysis focused on categories with at least 1000 reviews to ensure a sufficiently large review dataset. Only 7 categories met this threshold: person injury, divorce, DUI lawyers, criminal defense, bankruptcy, real estate, and immigration.

According to the data, personal injury lawyers are the most reviewed, with 6,942 total reviews. Divorce and criminal defense rank second and third, with 4,232 and 4,148 reviews respectively. Clients likely have stronger feelings about personal injury lawyers given the circumstances requiring this type of representation. This is perhaps one explanation for the higher review count. Furthermore, personal injury law may be more commonly practiced than other types of law, leading to the higher review count. However, more research would be needed to confirm if this is true.



The study then analyzed the text of the reviews to determine the sentiment people have towards each type of lawyer. This was measured using the Vader Sentiment average compound score per category. Immigration lawyers had the most positive language in their reviews, followed by real estate, and bankruptcy lawyers. While personal injuries have the most reviews, its reviews are less positive on average compared to the other categories.



**Where do people Yelp most about lawyers?**

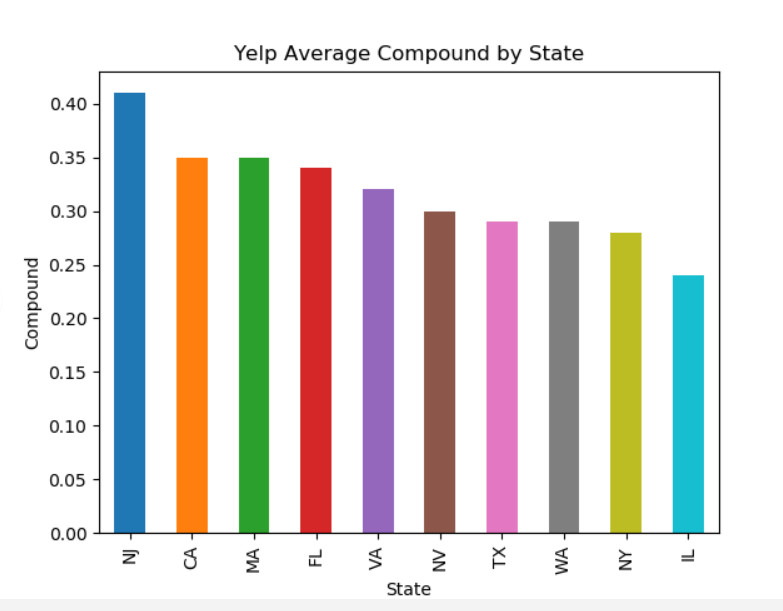
This study selected cities at random to gather Yelp reviews for the dataset. Below is a map of where the reviews came from. All locations are equally weighted. Cities included in the dataset follow the population areas of the country. After cleaning the data, the cities were plotted to determine a representative spread across the country. It appears to follow closely with the US population centers.



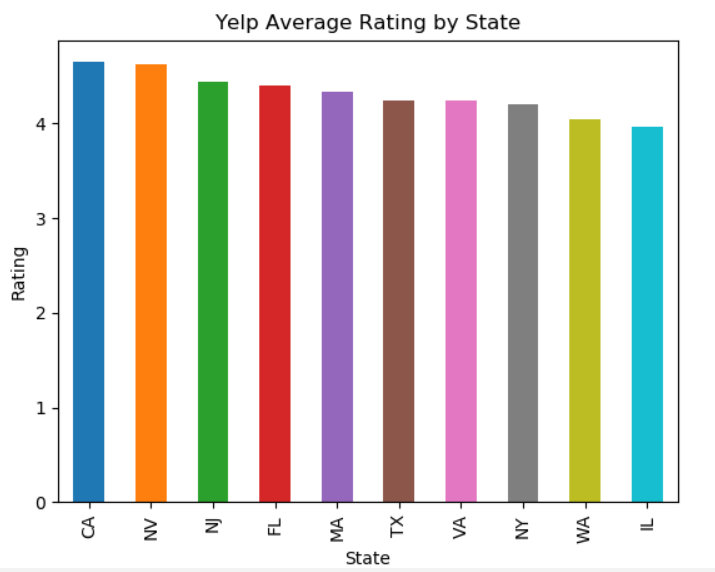
We then measured how many reviews were coming from each city. Like the city plot, total review count also follows population centers, as is expected. Most reviews are coming from urban areas along the coasts, while the center of the country has far fewer reviews.



We then ranked the states by review count to analyze the review sentiment and ratings. This study only focused on top 10 states with the most reviews. The data does show some variability within the states with respect to average sentiment and average Yelp rating. Looking at review sentiment by state across all categories, New Jersey has the most positive reviews, with California and Massachusetts falling slightly less positive. Illinois reviews had the lowest sentiment according to the data.

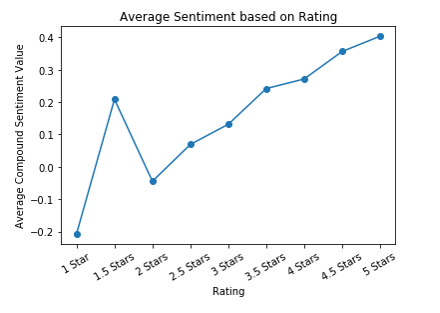


Looking at the Yelp star rating, California has the highest average lawyer rating, followed closely by Nevada. Again, Illinois has the lowest average rating to compliment their low review sentiment.



**Is there a relationship between review sentiment and star rating?**

Yelp allows users to rate business from 1 star to 5 stars. This study looked to see if users were providing ratings that reflected the reviews. To measure this relationship, we calculated the average review sentiment for each rating across all categories. There is a clear positive relationship. 5 star ratings have a higher review sentiment and 1 star rating have a lower review sentiment. 1.5 star ratings have similar sentiment to 3.5 star ratings. We believe this is due to a limitation with the API. Only the first 166 characters are shown in the review, therefore the sentiment could be skewed if the negative comments came after the 166th character. Nevertheless, users are generally leaving reviews that accurately reflect the ratings given.



To further confirm this relationship, we plotted the top 5 categories as well. With the exception of a few outliers, the data is clusters around the overall average as one would expect.

