# Tweet Frequency and Sentiment by County Analysis

The goal of this analysis is to look at the frequency and sentiment of tweets relating to human-coyote per county by month. Frequency analysis will help us gage which counties have sufficient data for a more in-depth examination. What is considered "sufficient data" is ambiguous and subjective, so for this analysis I will consider an average of five or more tweets per month as sufficient.

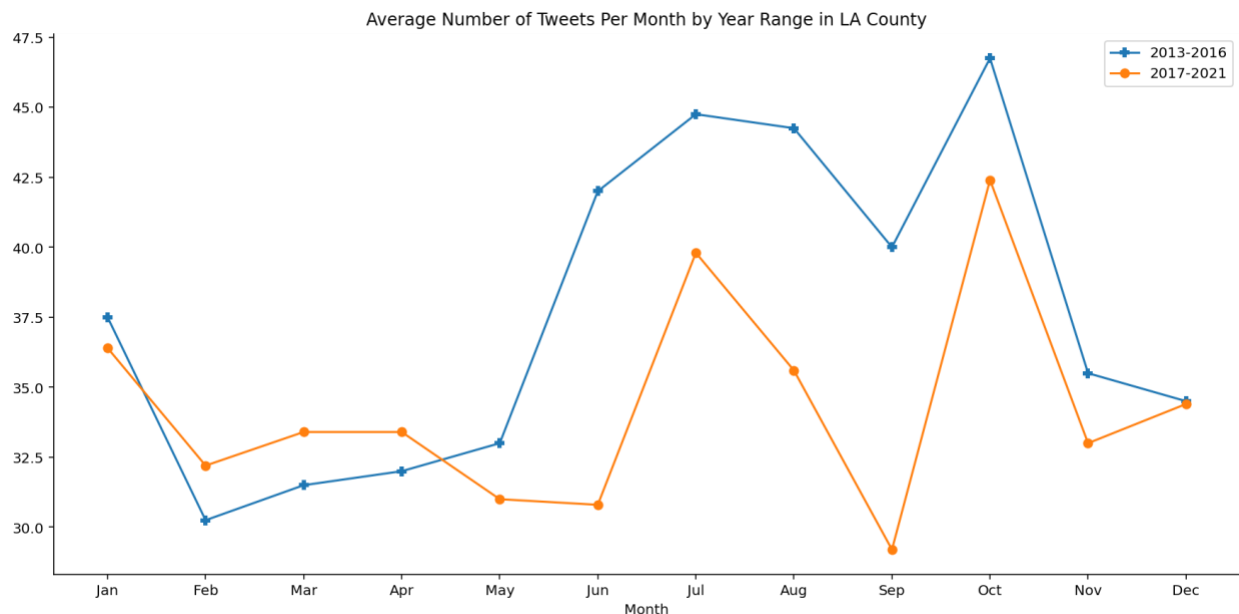Average Monthly Tweet Count by County
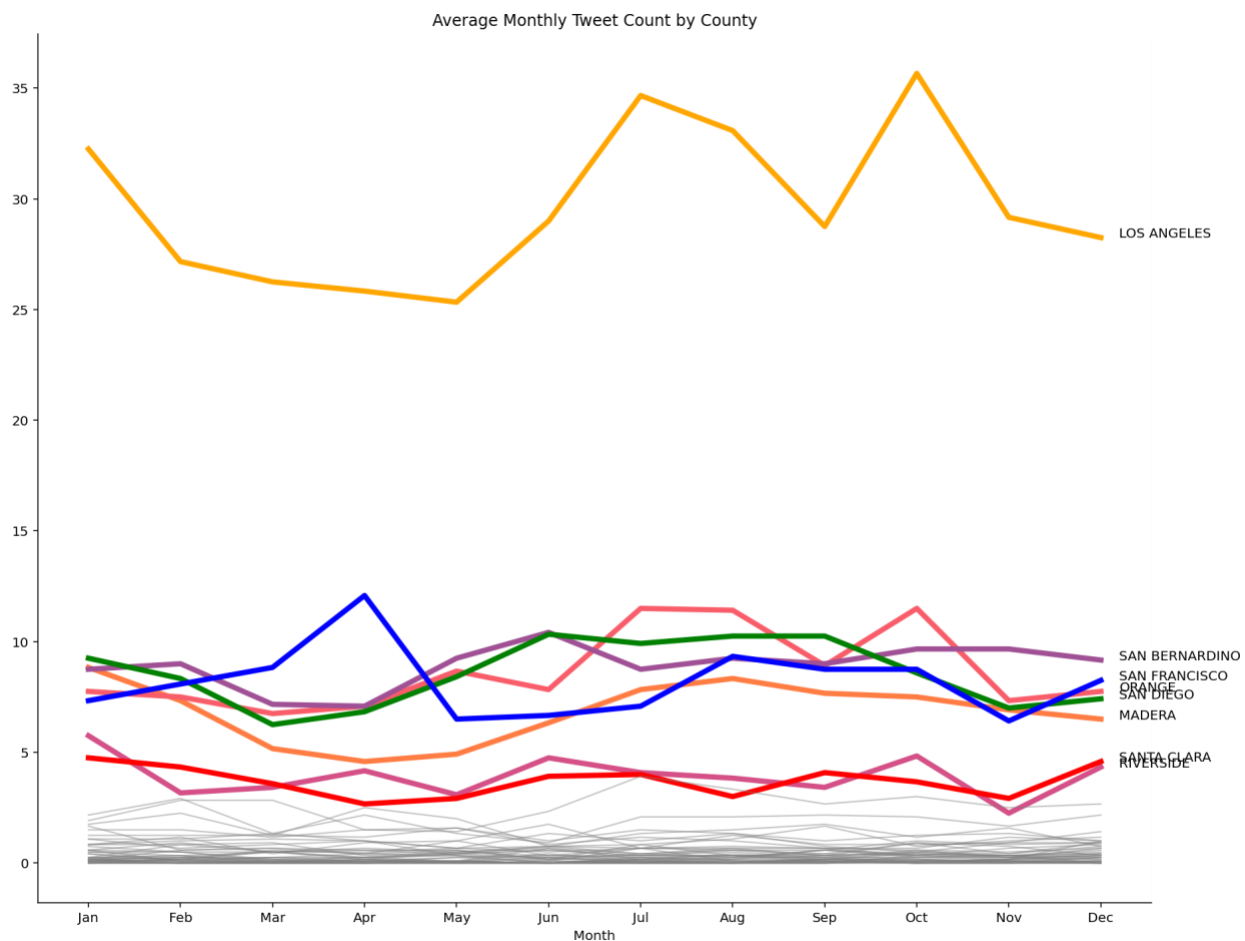
To begin, I grouped the data by county and month to look at the average tweet count per county by month and plotted it above. **Los Angeles** county significantly has more tweets per month than any other county. This is expected since there is quite a lot of human-coyote interactions and a large population to make this data visible.

Other counties worth noting are **San Bernardino, San Francisco, Orange, San Diego, Madera, Santa Clara,** and **Riverside**. These counties all have an average monthly tweet count of five or more and are considered to have "sufficient data" based on my threshold.

Majority of counties do not have sufficient data and are colored in gray. Further analysis with these counties are most likely not feasible due to the small sample size.

I recommend doing a deeper analysis on at least Los Angeles. The counties with a medium monthly count may be feasible but this should be discussed about first.

Frequency Analysis on Los Angeles County

Average Number of Tweets Per Month by Year Range in LA County

Los Angeles had significantly more tweets than other counties with an average monthly tweet count of around 31. I wanted to see if there were any notable patterns from LA county. I started by splitting the data by year ranges: 2013-2016 and 2017-2021. I choose these ranges because I wanted to see if the 2012-2016 California drought could have influenced tweet frequency.

Right off the bat I can see that there were more tweets in May to November for the 2013-2016 years. This could have been due to the drought causing coyotes to be more active in human population centers. It's worth noting that there are typically less tweets the further back in time we go. This is because many people delete their Twitter accounts, and their tweets are removed from the Twitter databases. This means that the 2013-2016 tweets frequency may have potentially been higher than what is shown in the data.

The plot above also indicates some seasonality throughout the year. Frequency is lowest between November to May and greatest in the summer and fall months of June to October. I suspect this is due to coyotes denning during the winter and spring months and being more active during summer and fall.

Average Compound Sentiment Score by Year Range and Month in LA County
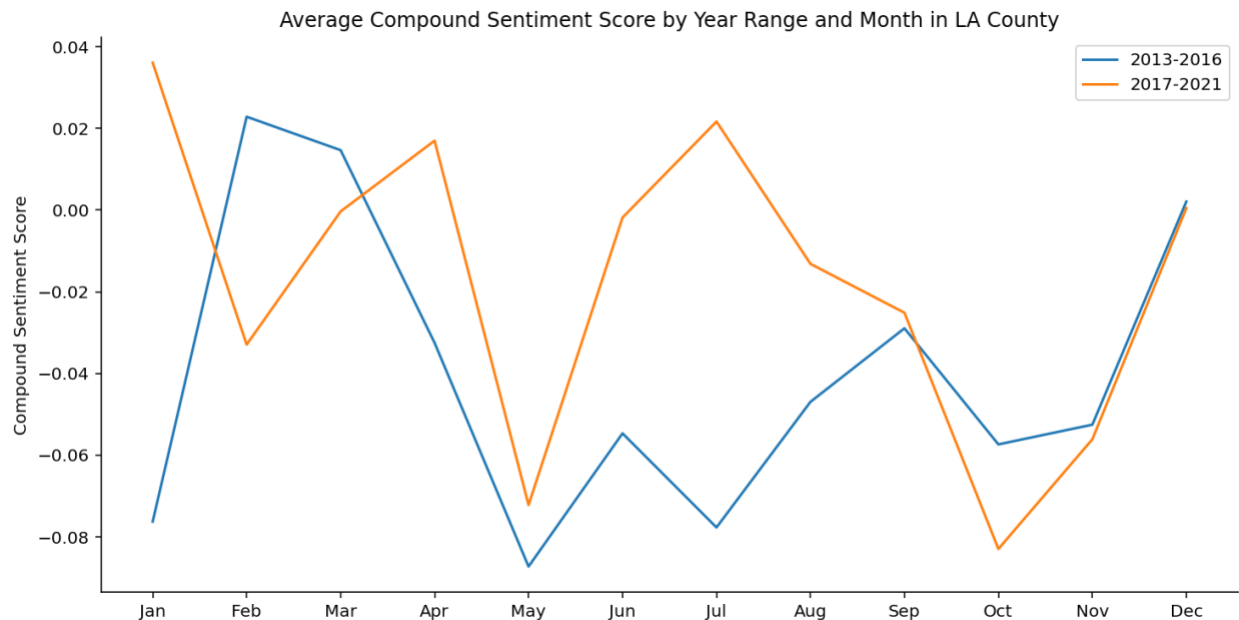
I kept the same year grouping but this time looked at the average compound sentiment score. The data does seem a bit noisy at first but a contrast that caught my attention is the different sentiment between May and September. 2017-2021 had a more positive sentiment score around this time. This may have been because during the drought years (2013-2016) coyotes were more hostile due to food scarcity resulting in more negative incidents like a pet being attacked.

Conclusion

Based on average monthly tweets per county by month, I do recommend that we do a further analysis on Los Angeles County. We should also discuss what it means to have sufficient data and what would minimum sample size should be.

The frequency and sentiment analysis on LA county indicates that the drought could have increased the frequency of human-coyote interactions while lowering the average sentiment. However, further analysis and data is most likely needed to make any concrete conclusions.