## Question

May the 4th (be with you) was a day like any other in the year 2020. Sunny and 72 with a slight breeze, and heavy with despair. The COVID-19 pandemic had slowed the world to a stop. The United States, capital of capitalism, feared this economic downturn most of all. People wanted their freedom. Most of all, they wanted haircuts and buffets. Feeling the pressure from her constituents and the Cheeto-In-Chief, Kansas Governor Laura Kelly announced that Kansas would begin opening for business on May 4, 2020. Slowly, of course. But how, you ask, does the population of Kansas feel about this? Let's investigate Wichita – Kansas's biggest city – for answers.

## Analysis and Methods

North of 20,000 tweets from before May 4th were collected from Wichita and the surrounding 50-kilometer radius, and ditto from May 4th (the first tweet for this period was at 00:01 on the morning of May 4th). Data scraped from these tweets included the ID number of the tweet, the tweeter (their username), the text of the tweet, and the time it was created at. In the end, the ID numbers and the time the tweet were created were only used for sanity checks. The tweets were then fed into two sentiment analysis models. One I created myself, using a database of 1.6 million sentiment-tagged tweets on Kaggle, titled 'Sentiment140'. The other was a popular tool for quick and dirty sentiment analysis called TextBlob. This sentiment classifier is pre-trained on IMDb movie reviews, so it's probably not too accurate for tweets. But that's part of the fun! After both sentiment classifiers spit out their values, the data was exported to R for analysis. The model I created predicted values from 0 to 1 (negative to positive), while TextBlob predicts sentiment from -1 to 1 (again, negative to positive). So, in R, I had to squish the TextBlob sentiments into the 0 to 1 range using a simple transformation. Paired t-tests were run for both models, across both periods of data collection. The reasoning behind the paired t-test was that these were independent observations at two points in time.

## Results

The graphs (see 'Figures and Graphs' section below) show an interesting trend. There is increased left 'skewness' to the density plot for the custom sentiment model, and a visible increase in the density of sentiments near 1. This was backed up by the paired t-test, which gave a significant result ($p < 2.2e\text{-}16$). The TextBlob classifier showed no statistically significant difference between the two time periods, and many of its predicted sentiments were 'neutral' at 0.5.

## Conclusion

This method of looking at a population can be a useful tool in the months ahead. As governors, city mayors, and policy makers in general make big decisions that affect their populations, it will pay to know where that population's collective head is at. Sentiment is a fickle beast, and Twitter even more so. However, by using aggregate data on thousands of individuals, uncertainties can be overcome, and general trends can be uncovered. May the force be with you, Kansas.

## Figures and Graphs