

## **Data Wrangling**

The question to be answered/addressed in my project is, “Did Americans support a ‘summit’ between U.S. President Donald Trump and North Korean leader Kim Jong Un? And do those opinions vary by geographic region?”

Using Twitter data, I intend to perform sentiment analysis to determine Americans supported a ‘summit’ between U.S. President Trump and North Korean leader Kim Jong Un. First, I needed to authorize a Twitter API client, fetch tweets, and parse the results. As a follow-up visualization, I intend to map the locations of analyzed Twitter users to determine if/how these views vary with geography.

Regarding data cleaning, my Twitter results required little to no cleaning. Because my data was returned in JSON format, I need to ‘unpack’ some of the nested columns. For this, I relied on the `normalize()` function for each of those columns. Additionally, I was primarily interested in original tweets so I removed any retweets (‘RT’). Finally, because the second part of my project involves plotting/mapping users based on the geolocation information contained in the metadata, I also dropped all rows that did not contain this information.