



Examine the Impact of COVID-19 Policies and Restrictions on the Sentiment Score of Tweets

Jenny Chan '22.5 | Data Science Capstone

Research Question:

Does travel restrictions and health policies affects the average sentiment score for COVID-19 related tweets on Twitter?

Hypothesis:

Travel Restrictions and health policies negatively affect the average sentiment score for COVID-19 related tweets on Twitter.

Background:

On March 13, the Travel Ban issued by President Donald Trump prohibits foreign nationals from entering the U.S. if they have been in the Schengen Area within 14 days before their entry. Soon afterwards, the continued rise of COVID-19 cases led to a series of health policies and travel restrictions being issued by the state and federal government, including a nation-wide travel restrictions, states-wide stay-at-home orders and mask mandates.

Twitter is a popular microblogging platform; almost 25% of the US population have a Twitter account. During the pandemic, there were significant numbers of Tweets related to COVID-19. This research focuses on policies changes in the U.S.

Data Collection:

COVID-19 Tweets ID and Sentiment Score (Source: IEEE):

ID and the sentiment score of the COVID-19 related tweets that are geotagged (have a location).

Using a Twitter Hydrator, I extracted the complete information (including the text, date, and location) of the Tweet from the Tweet ID. From there, I converted the location into the name of the country, and filtered only tweets from the U.S.

There were a total of 178049 tweets, between the 926 days from 2020-03-19 to 2022-09-30.

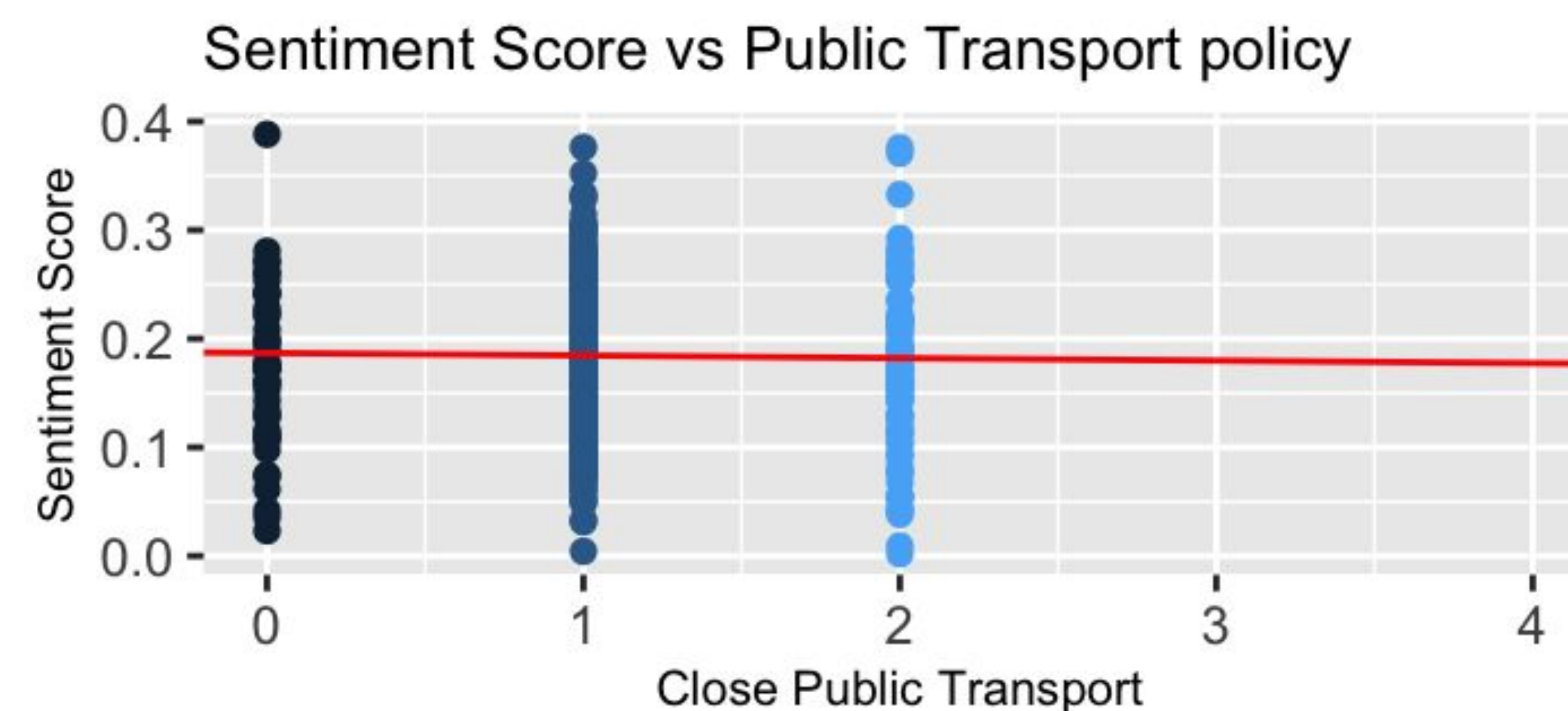
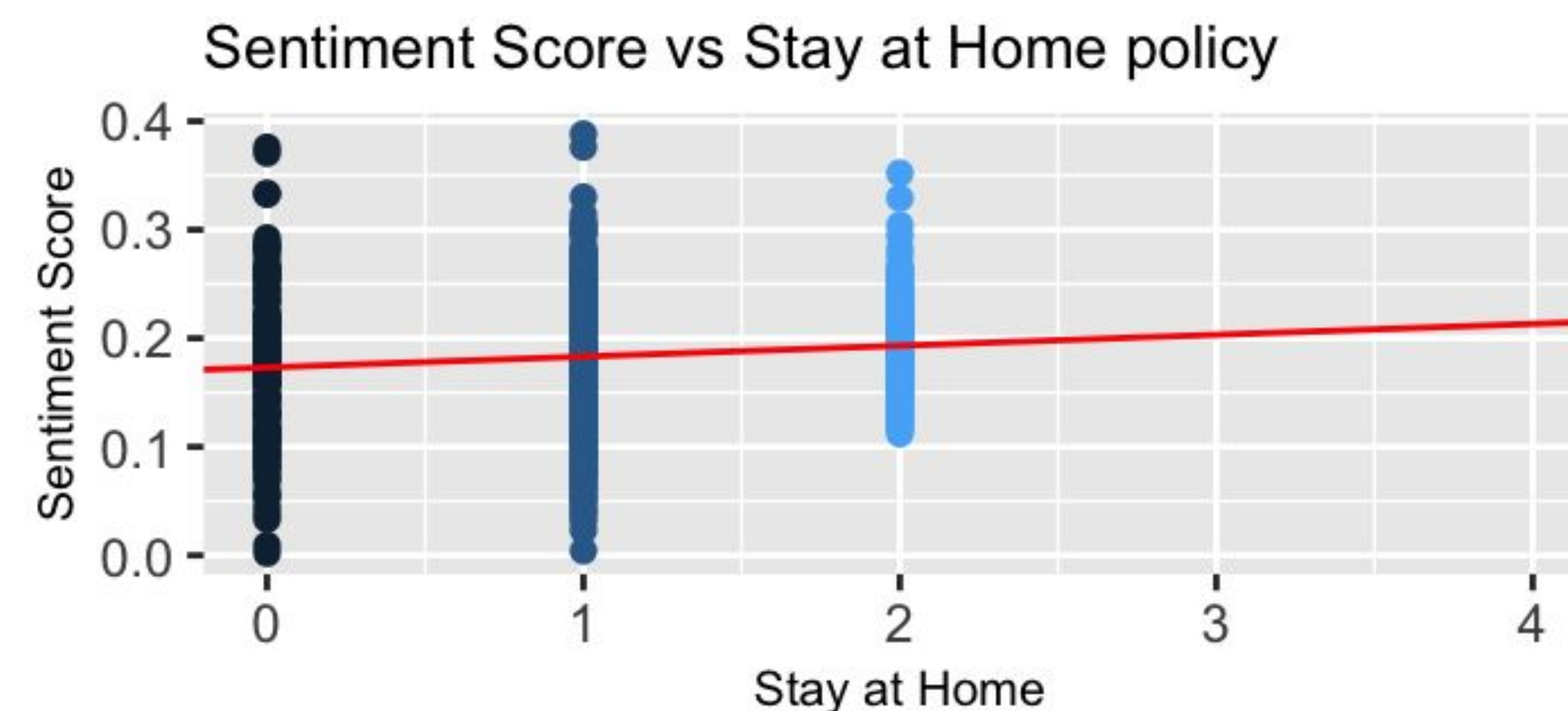
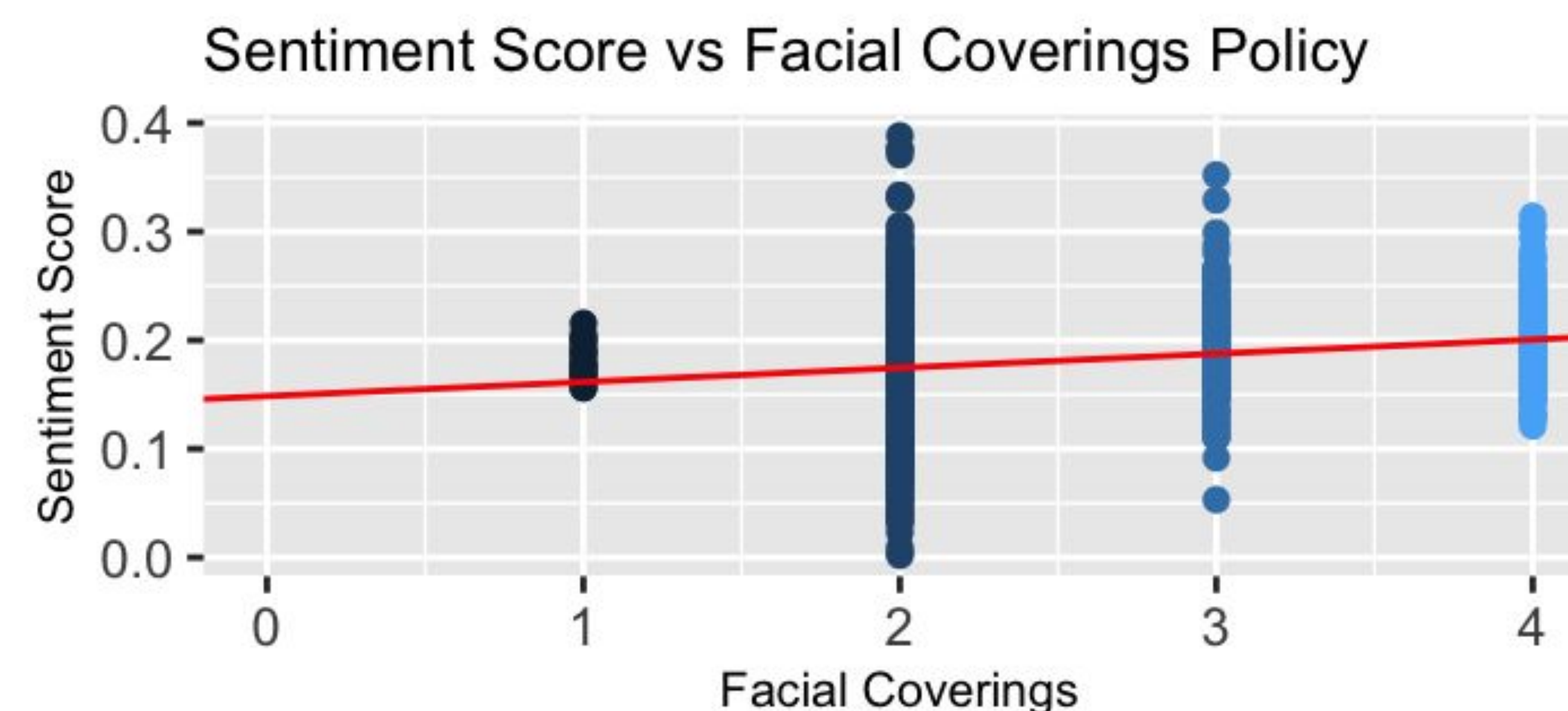
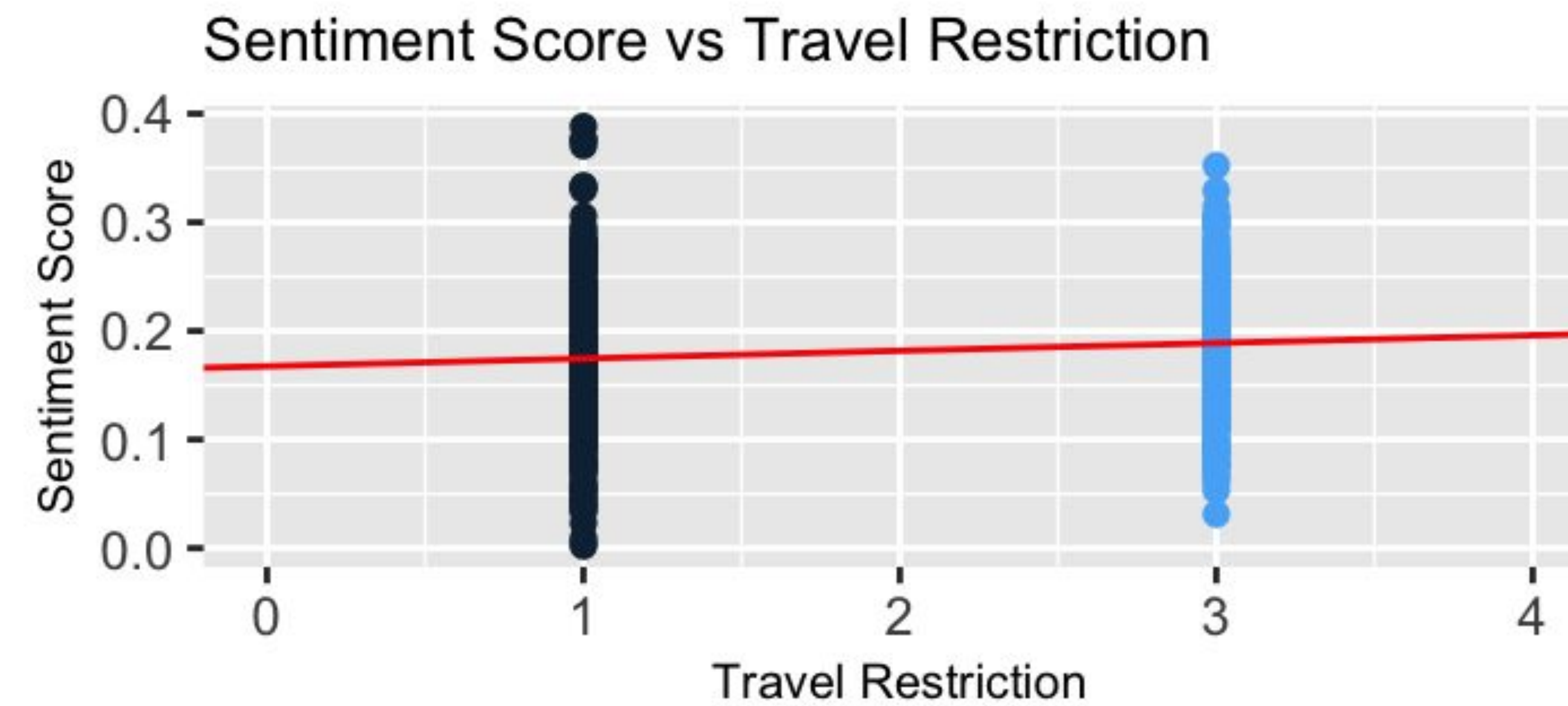
COVID-19 Policies and Restrictions (Source: Oxford):

16 different datasets of panel data on the COVID-19 policies and restrictions across different countries between March 2020 till now (October 2022). Data includes travel restrictions, masking policies, and stay-at-home policies. All datas are ordinal data, ranked from 0 to 4, from no restrictions to tightest restrictions.

For example, for the facial covering policy: 0 (no policies), 1 (Recommended), 2 (Required in some specified shared spaces), 3 (Required in all shared places when social distancing not possible), 4 (Required at all times regardless of location).

I filtered the country code and kept only data on the U.S

Data Visualization:



Data Modeling:

Using a linear regression model, we identify the relationship between COVID-19 policies and restrictions (Independent Variables) on the sentiment score (Dependent Variable) of COVID-19 related tweets on Twitter.

For IV: There 16 different possible independent variables, 3 IVs had no changes over the timeframe. After performing stepwise elimination with a Variance Inflation Factor threshold of 2.5, there are four remaining IVs: 1) Face Covering 2) International Travel Restrictions 3) Closing of public transport 4) Stay-at-home policy.

For DV: In the total of 178049 tweets ID provided, 2843 doesn't contain a sentiment score. When checking the normality assumption for DV, a significant number of tweet (n=54411) has a neutral sentiment score. By removing tweets with neutral sentiment score, we can assume normal distribution for all the Y values.

Preliminary Result:

The slopes for the linear regression model with Travel Restriction ($B1=0.007$), Facial Covering ($B1=0.013$), and Stay-at-home policy ($B1=0.01$) are positive. The slope for the linear regression model with public transport policies is negative, however with a very large p-value ($p=0.629$). All other p-values for the coefficients are <0.05 , indicating there are enough evidence to show that a point will fall on the regression line of the model.

Discussion:

Public tends to have different opinions regarding to policies and restrictions issued by state and federal government. There are definitely less significant effects on the sentiment score for the closing of public transport as the U.S. didn't vary their transport restrictions much over the past 2 years. In general, sentiment score increases and become more positive as travel restrictions, stay-at-home policy, and facial covering policies becomes more strict with the simple single-variable linear regression model.

Data Ethics & Limitations:

TextBlob - NLP library that generated the sentiment score in the dataset. It doesn't recognize Covid-related phrases. "Positive test result" would be classified as a positive phrase, referring to "positive" as an attitude rather than actually referring to a "positive" test result for COVID-19. There's also a significant neutral result for sentiment score.

The reinforcement of some Covid-19 restrictions may be controversial. Even when the sentiment is leaning towards positive, it may not be beneficial for all, especially those who may be most negatively impacted by COVID-19.