

A Tale of Emotions: NRC Analysis on COVID-19 Narratives/Tweets



Group 2: Elizabeth Kim, Gurnoor Singh, Fayren
Chaerunnissa, and Jennifer Osei



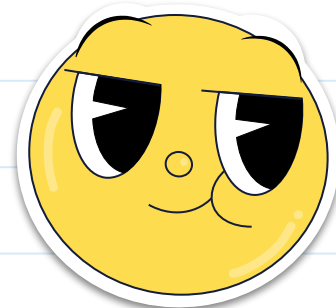
Introduction:

- The pandemic was isolating and frustrating to many, with the internet being the main place people could find community.
 - Many took to social media, specifically **Twitter**, to state their opinions about the pandemic.
- We aimed to use this information to determine user sentiments using the **NRC Emotion Lexicon**.
 - List of 5,636 English words and their associations with 8 basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and 2 sentiments (negative and positive).
- Can use the results to:
 - Quickly implement preventative measures in at-risk areas
 - Create informational campaigns/Increase public awareness
 - Dedicate resources to high-need areas



Overview of Data:

- **Kaggle** dataset that collected **tweets** that were tagged with “**#covid19**”.
 - When: Collected from **July 24, 2020** to **August 30, 2020**
 - How: Used Twitter API and a Python script, by running a query for “#covid19” on a daily basis.
- Information used:
 - User’s Location (US State)
 - Date of the Tweet
 - Text of the Tweet



Example from Dataset:

user_name	user_locati...	user_desc...	user_creat...	# user_follo...	# user_friends	# user_favo...	✓ user_verifi...	📅 date	📝 text
Tom Basile 🇺🇸	New York, NY	Husband, Father, Columnist & Commentator. Author of Tough Sell: Fighting the Media War in Iraq. Bush...	2009-04-16 20:06:23	2253	1677	24	True	2020-07-25 12:27:17	Hey @Yankees @YankeesPR and @MLB - wouldn't it have made more sense to have the players pay their re...

Steps for NRC Analysis:

```
import nltk
from nrclex import NRCLex
import re
import pandas as pd
```

```
text = data['text']
```

```
# Word Tokenization
tokens_list = text.apply(nltk.word_tokenize)
```

```
# Get rid of symbols and special characters from each tweet
```

```
def remove_special_characters(tokens):
    pattern = r'^[a-zA-Z0-9\s]\'
    tokens = [re.sub(pattern, '', token) for token in tokens]
    tokens = [token for token in tokens if token]
    return tokens
```

```
# Apply the remove_special_characters function to each list of tokens
```

```
tokens_list_cleaned = tokens_list.apply(remove_special_characters)
```

```
# Sentiment Analysis using NRC Lexicon
```

```
def get_sentiment_score(tokens):
    text = " ".join(tokens)
    return NRCLex(text).affect_frequencies
```

```
# Apply the get_sentiment_score function to each list of tokens
```

```
sentiment_scores = tokens_list_cleaned.apply(get_sentiment_score)
```

```
# Create a new DataFrame for sentiment analysis results
```

```
sentiment_df = pd.DataFrame(sentiment_scores.tolist(), index=data.index)
```

```
# Combine the new DataFrame with the original data
```

```
data_with_sentiment = pd.concat([data, sentiment_df], axis=1)
```

```
# Display the data with sentiment analysis results
```

```
print(data_with_sentiment.head())
```

Data Cleaning & Pre-Processing

Sentiment Analysis with NRC

Output:

```
user_name      user_location \
0      Vjoe      astroworld
1  Tom Basile    New York, NY
2 Time4fisticuffs  Pewee Valley, KY
3  ethel mertz  Stuck in the Middle
4  DIPR-J&K     Jammu and Kashmir

user_description      user_created \
0  wednesday addams as a disney princess keepin i...  2017-05-26 05:46:42
1  Husband, Father, Columnist & Commentator. Auth...  2009-04-16 20:06:23
2  #Christian #Catholic #Conservative #Reagan #Re...  2009-02-28 18:57:41
3  #Browns #Indians #ClevelandProud #[] #Cavs ...  2019-03-07 01:45:06
4  Official Twitter handle of Department of Inf...  2017-02-12 06:45:15

user_followers  user_friends  user_favourites  user_verified \
0             624           950           18775           False
1             2253          1677              24             True
2             9275          9525           7254           False
3             197           987           1488           False
4            101009          168             101           False

date      text \
0  2020-07-25 12:27:21  If I smelled the scent of hand sanitizers toda...
1  2020-07-25 12:27:17  Hey @Yankees @YankeesPR and @MLB - wouldn't it...
2  2020-07-25 12:27:14  @diane3443 @wdunlap @realDonaldTrump Trump nev...
3  2020-07-25 12:27:10  @brookbanktv The one gift #COVID19 has give me...
4  2020-07-25 12:27:08  25 July : Media Bulletin on Novel #CoronaVirus...

...      anger anticip      trust  surprise  positive  negative  sadness \
0  ...  0.000000      0.0  0.000000  0.000000  0.000000  0.500000  0.000000
1  ...  0.000000      0.0  0.285714  0.000000  0.428571  0.000000  0.000000
2  ...  0.166667      0.0  0.000000  0.166667  0.166667  0.166667  0.166667
3  ...  0.000000      0.0  0.142857  0.142857  0.285714  0.000000  0.000000
4  ...  0.000000      0.0  0.000000  0.000000  0.000000  0.000000  0.000000

disgust      joy  anticipation
0  0.500000      0.000000      NaN
1  0.000000  0.142857  0.142857
2  0.166667  0.000000      NaN
3  0.000000  0.285714  0.142857
4  0.000000  0.000000      NaN

[5 rows x 24 columns]
```

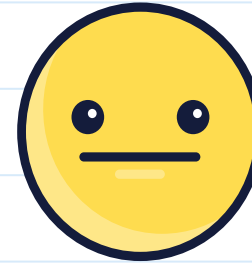
Given 9 Sentiment Scores for each Tweet:

- Joy
- Trust
- Anticipation
- Surprise
- Positive
- Negative
- Anger
- Sadness
- Disgust

Sentiment Analysis:



Negative



Neutral



Positive

Steps to Aggregate and Visualize the Data:



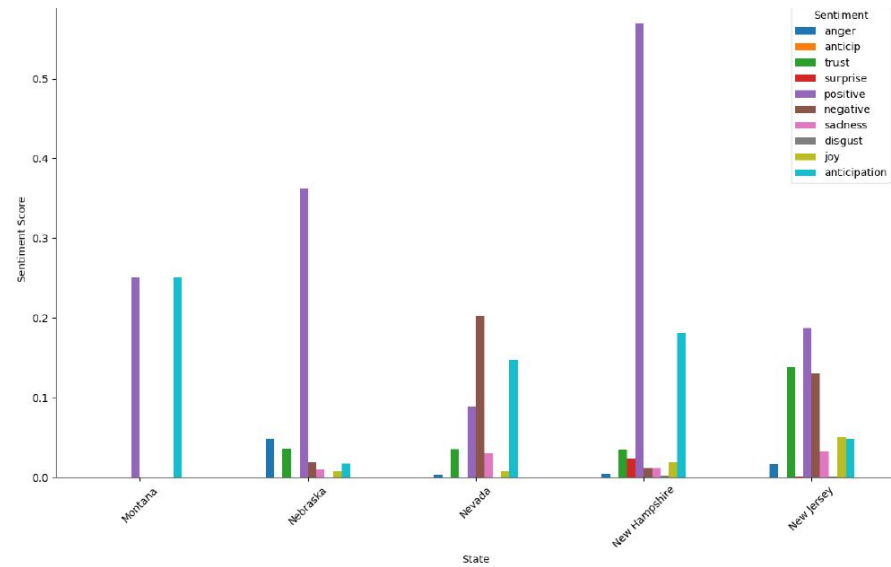
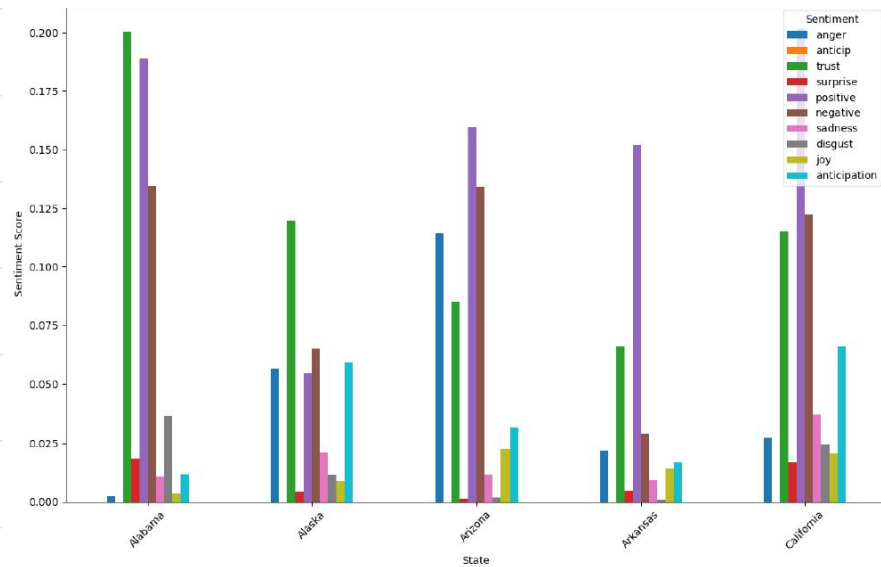
Aggregation Process:

1. Define the location of each user (thousands)
2. Clean the column of each user
3. Identify each which locations are in the US
4. Classify each location according to a given state
 - a. Do this by looking at towns, cities, major parks, and central hubs in each state
 - b. Aggregate the state rows
5. Take the mean of the sentiments for each state
 - a. Assess the data by plotting bar graphs





Bar Graphs:

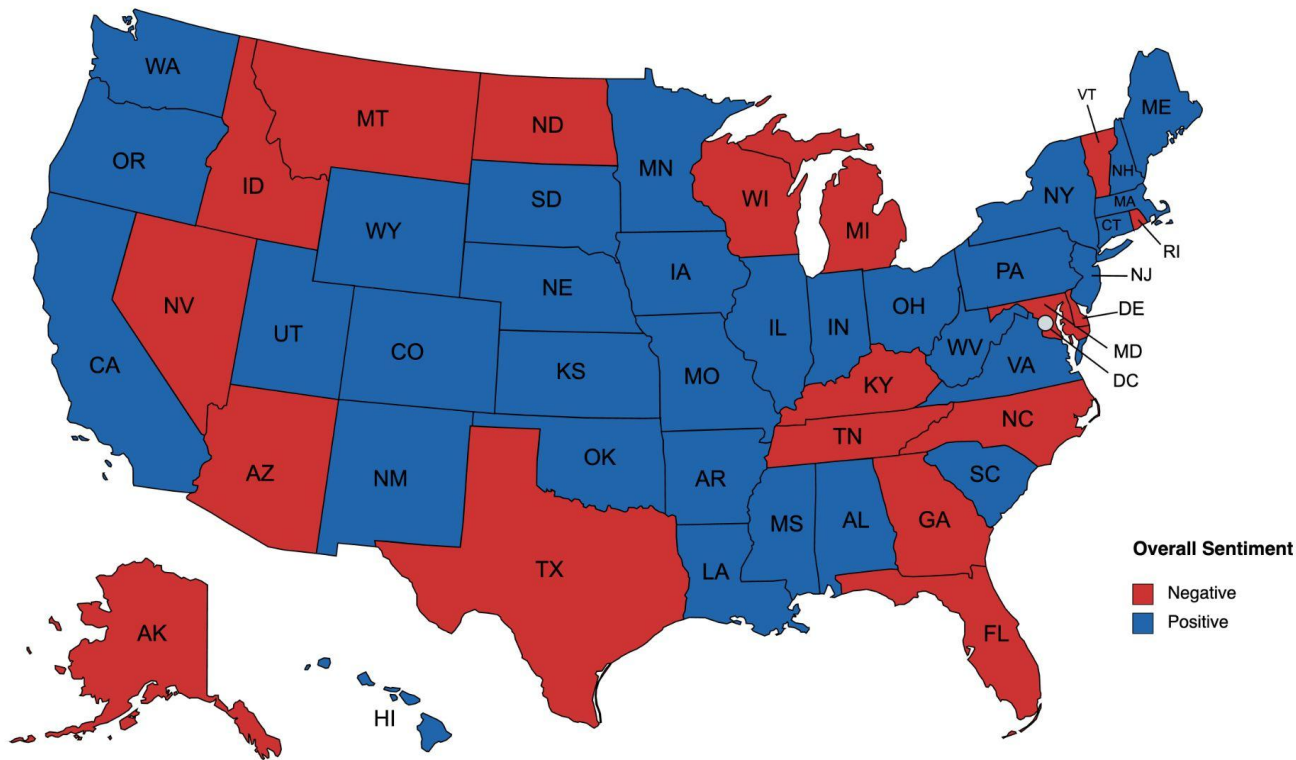




Additional Aggregation and Visualization:

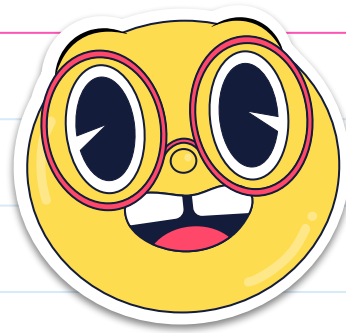
1. Define which sentiments are positive vs negative:
 - a. Positive: joy, trust, positive
 - b. Negative: sadness, disgust, anticipation, surprise, anger, and negative
2. Aggregate the two sentiment groups and calculate a sum for each state
3. Define whether each state is overall positive or negative by comparing the two scores
4. Plot the States using a geojson file for clear analysis







Performance Measures:

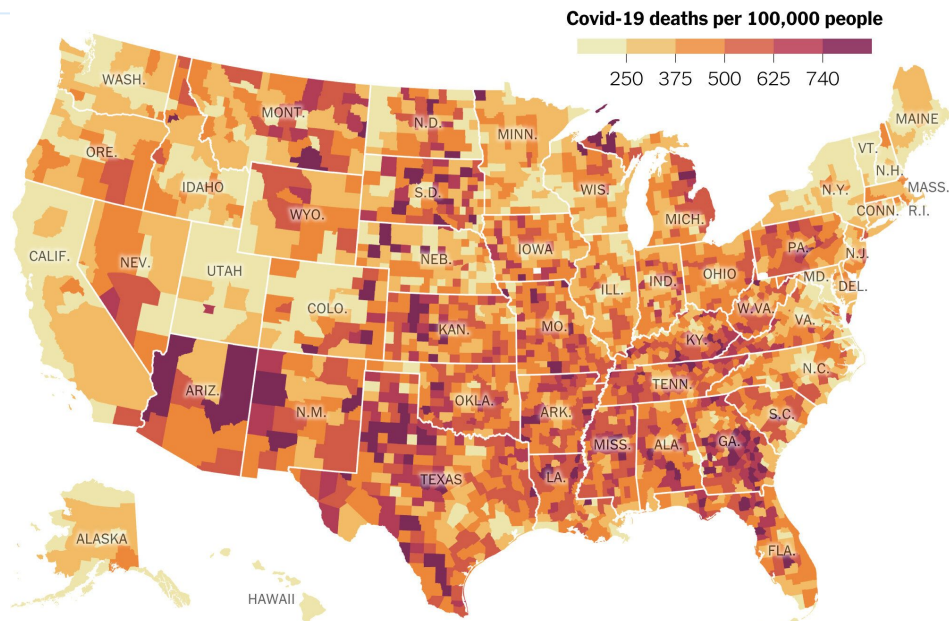
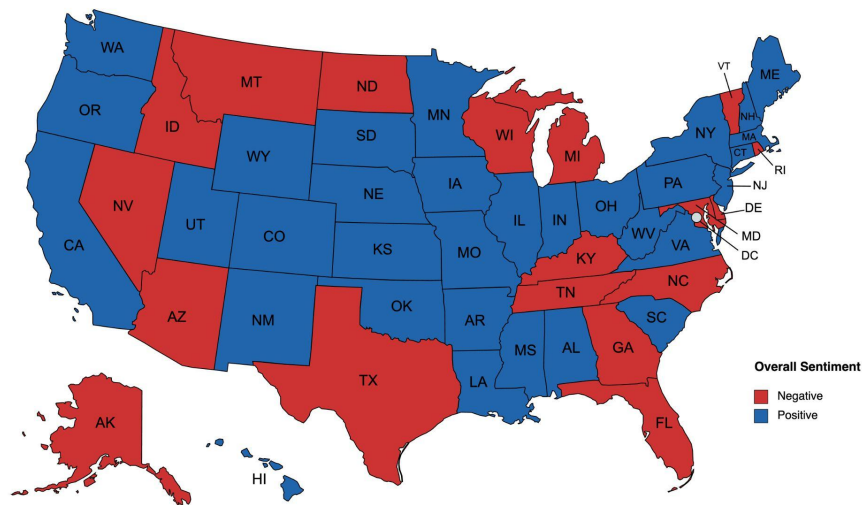


- Checked against Covid-19 data maps:
 - Where there were a lot of death/illness
 - Pre-existing health conditions
 - Where there were stricter enforcement of/people most followed regulations
 - Ex: Can you travel or not?
- This substantiates the sentiments of each user by region
 - If something is more common in one area → fuels a specific emotion and sentiment more



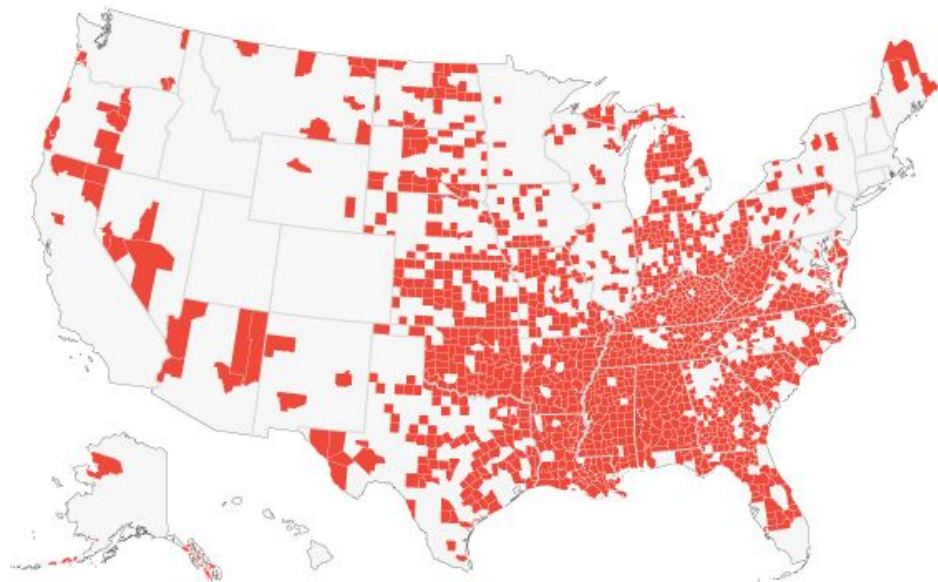
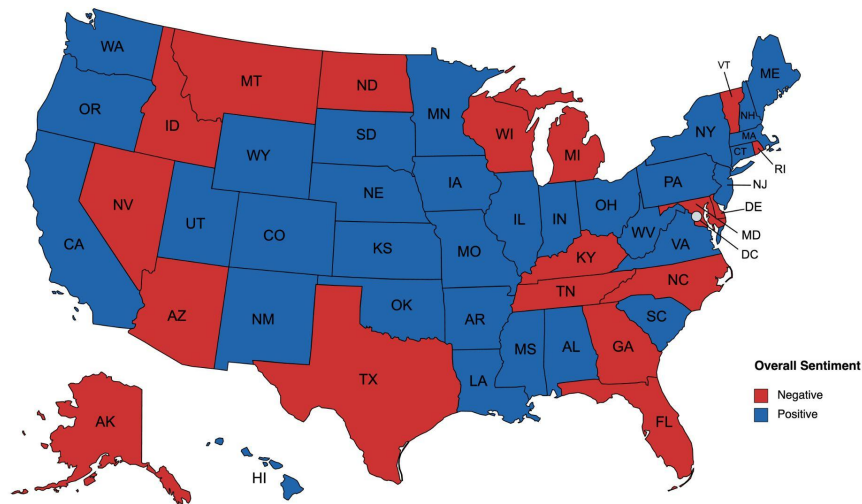
Map of Covid-19 Deaths:

- More deaths → Negative sentiment
 - **Georgia, Arizona, and Texas**



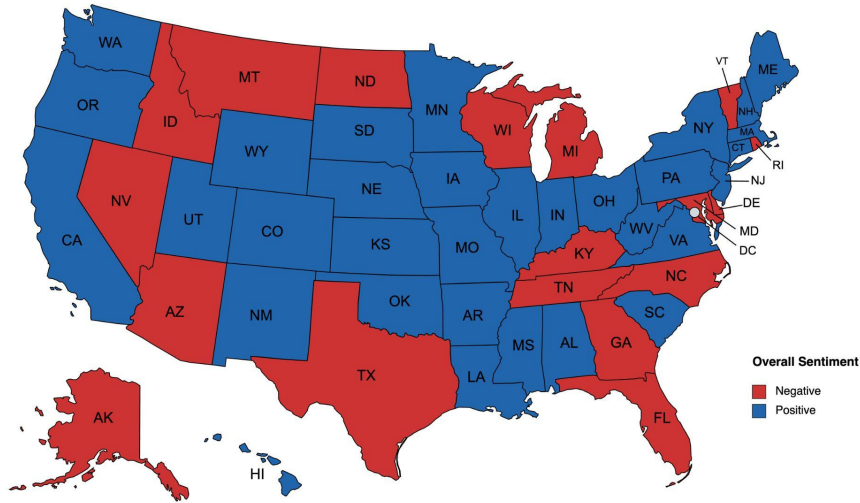
Map of Where Chronic Health Conditions and Coronavirus Could Collide:

- **Georgia, Kentucky, and Tennessee** are more susceptible to this → Increased deaths → Negative sentiment



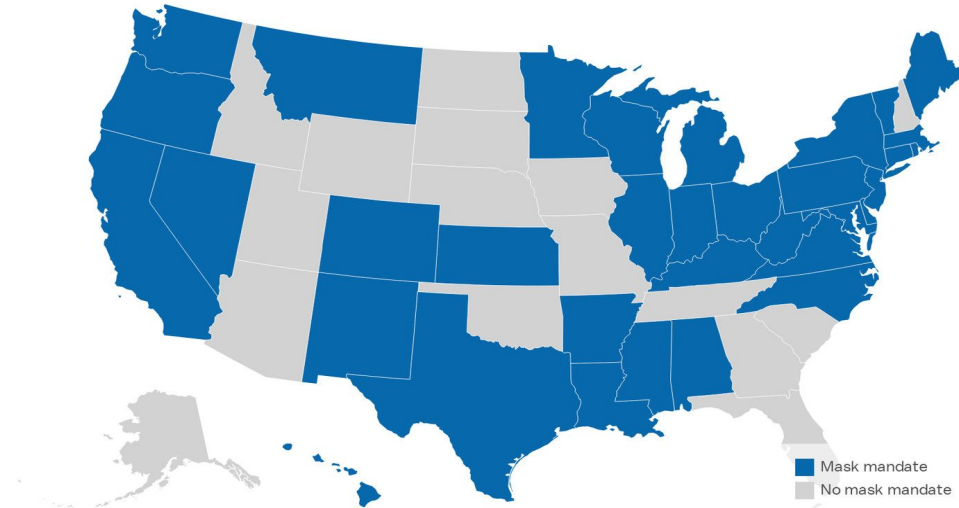
Map of Where Masks are Mandated:

- Require masks → Impede on quality of life → Negative Sentiment
 - Nevada, North Carolina, Kentucky, Maryland, and Delaware**



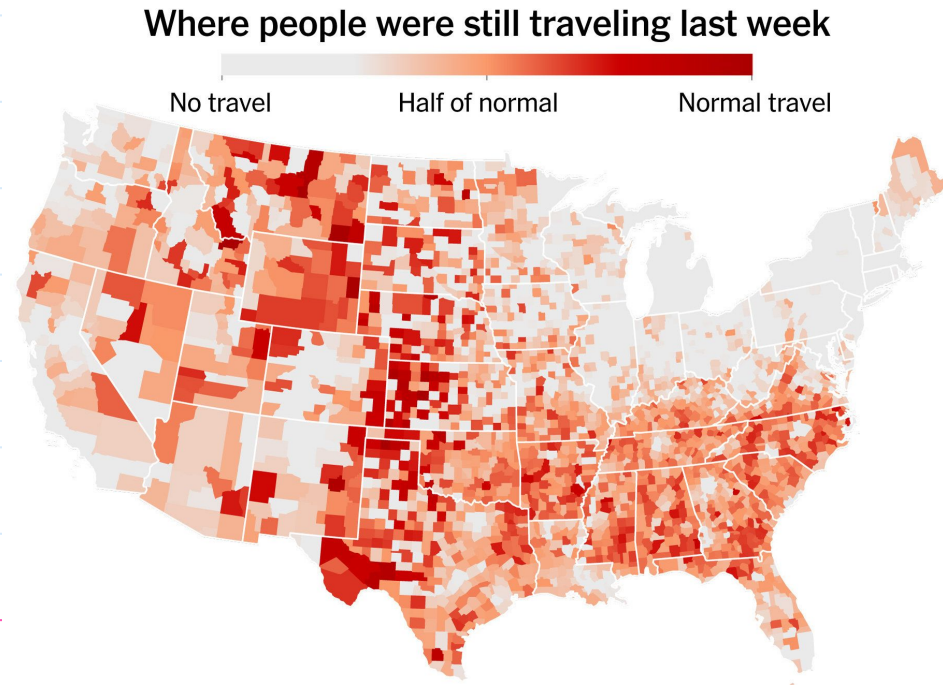
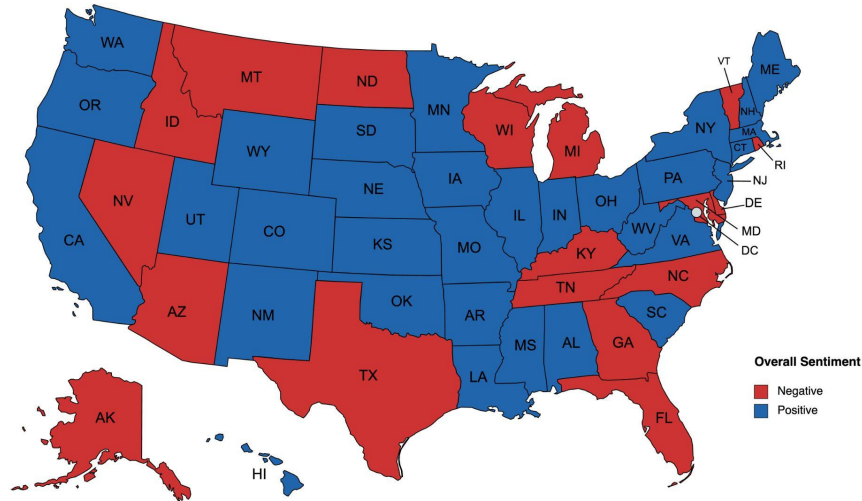
Most states mandate a mask in public

Whether a state requires wearing a mask in public as of August 10



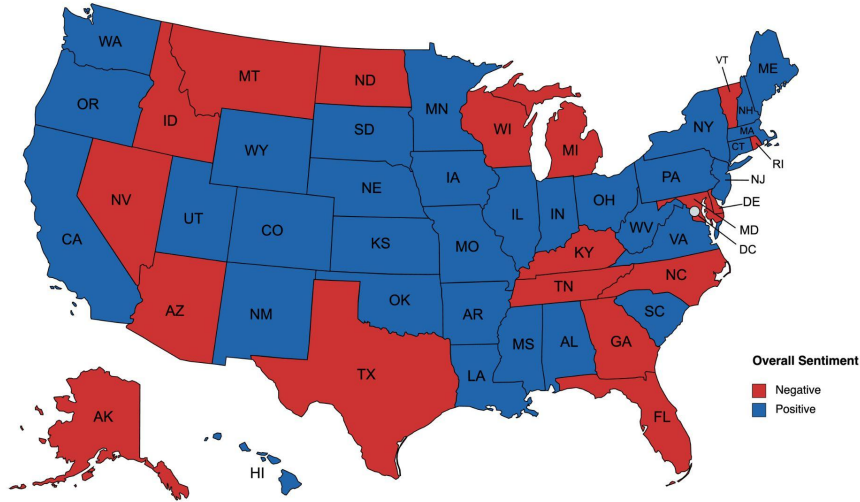
Map of Where People Still Traveled:

- Traveling → Spread of Covid → Anger about contracting → Negative Sentiment
 - **Texas, North Carolina, and Montana**
- Ability to travel → Positive sentiment
 - **Nebraska, Kansas, and Wyoming**

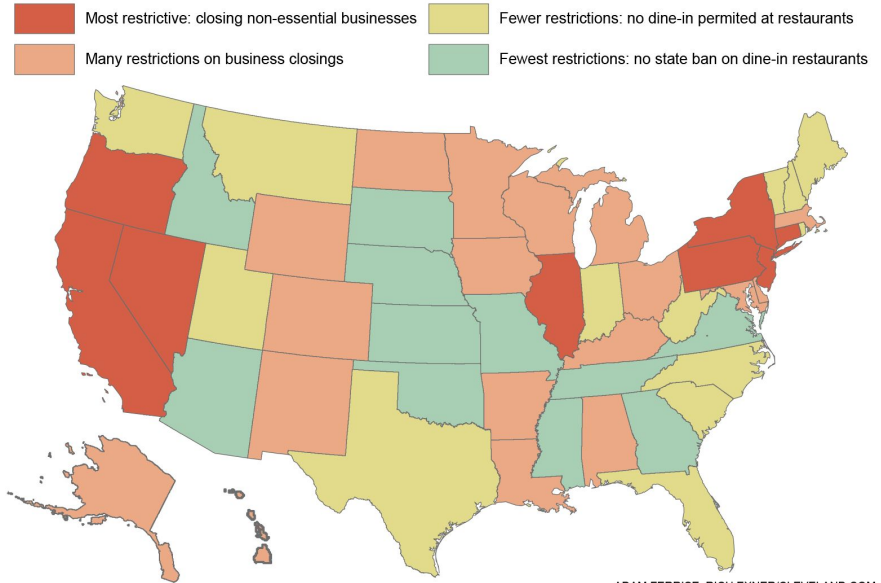


Map of Business Restrictions:

- Mississippi, Missouri, and Virginia had the least restrictions → Positive sentiment

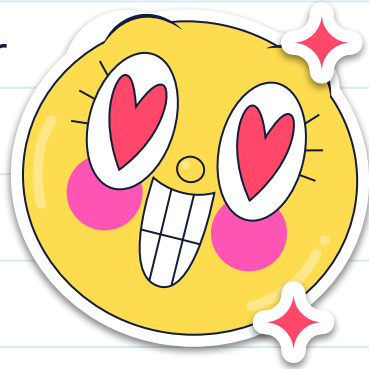


Coronavirus restrictions across the United States



What did we like least and best about the project?

- **We MOST liked...**
 - The flexibility to choose our own dataset that is most interesting to the whole team
 - How applicable/translatable it can be to the real world, our jobs, and school projects
 - How prepared the class made us for our projects
- **We LEAST liked...**
 - A midway check-in point to make sure we are on the right track



Conclusion:

- More positive outlook on Covid-19 than we initially presumed
 - **Trust**
- Secondary Emotions: **Anger, Anticipation, Sadness**
 - Stems from uncertainty and hindered quality of life
- This data can be useful for future pandemics and preventing the spread of them
 - Direct more resources towards and educate states that had more negative sentiments
 - Government can create preventative measures and share information that integrates the most common emotions
 - Remediating these sentiments/emotions → appease population → more likely to follow regulations → prevent spread of future epidemic



References:

- Adam Ferrise, cleveland. com. (2020, March 21). *50 states of coronavirus: How every state in the U.S. has responded to the pandemic*. cleveland.
<https://www.cleveland.com/metro/2020/03/50-states-of-coronavirus-how-every-state-in-the-us-has-responded-to-the-pandemic.html>
- Gamio, L., Lutz, E., & Sun, A. (2023, May 11). *As emergency ends, a look at Covid's U.S. death toll*. The New York Times. <https://www.nytimes.com/interactive/2023/05/11/us/covid-deaths-us.html>
- Glanz, J., Carey, B., Holder, J., Watkins, D., Valentino-devries, J., Rojas, R., & Leatherby, L. (2020, April 2). *Where America didn't stay home even as the virus spread*. The New York Times.
<https://www.nytimes.com/interactive/2020/04/02/us/coronavirus-social-distancing.html>
- Lopez, G. (2020, August 11). *America's uniquely bad covid-19 epidemic, explained in 18 maps and charts*. Vox.
<https://www.vox.com/future-perfect/21353986/coronavirus-covid-pandemic-usa-america-maps-charts-data>
- MultiState. (n.d.). *Covid-19 policy tracker*. MultiState. <https://www.multistate.us/issues/covid-19-policy-tracker>
- Popovich, N., Singhvi, A., & Conlen, M. (2020, May 18). *Where chronic health conditions and coronavirus could collide*. The New York Times.
<https://www.nytimes.com/interactive/2020/05/18/us/coronavirus-underlying-conditions.html>
- Preda, G. (2020, August 30). *Covid19 tweets*. Kaggle. <https://www.kaggle.com/datasets/gpreda/covid19-tweets>