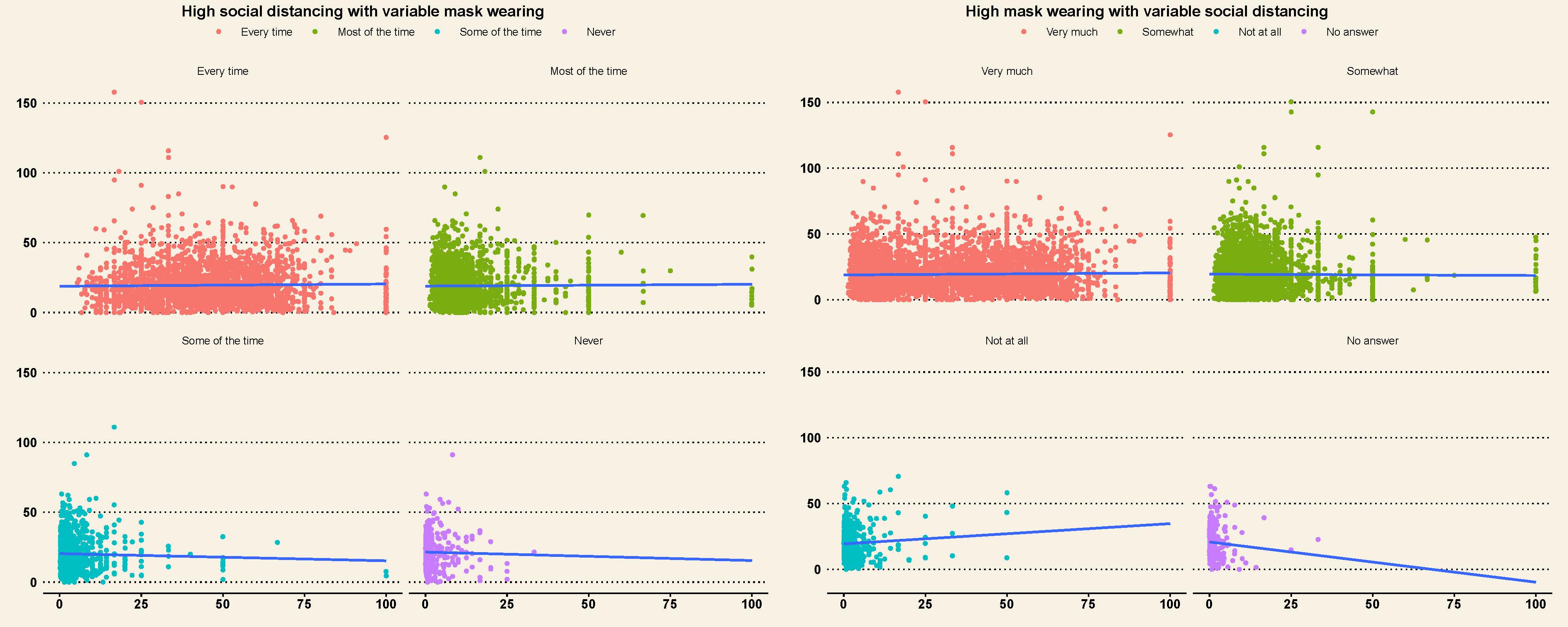
Hackathon Write Up

For the hackathon, I decided to try my hand at regression analysis across the Johns Hopkins COVID dataset, the Survey Monkey data, and the ACS data.

I started by downloading each data set and determining the variables of note. A controversial topic in discourse today, as vaccinations ramp up and new cases plateau, is the use of masks on a day-to-day basis. Two variables in the Survey Monkey data seemed to touch on this. Specifically, the variables asking about how often a respondent would wear a mask and how often the respondent would adhere to social distancing guidelines.

The dataset was rich with possible analysis, but given the time constraints, I chose a relatively simple question to ask: what works best to decrease the outbreak of COVID cases in a county, mask wearing, or social distancing? To do this, I calculated percentages of respondents by county and compared those to the cases per 1000 people (as of 9/30/2020). I then created two datasets to compare, one where we looked at respondents who reported high social distancing, but varying mask wearing and those who reported high mask wearing with variable social distancing. I predicted that social distancing would be the larger factor in COVID transmission.



The graph above shows the linear regressions done using geom\_smooth(method = lm). For the most part, the slopes of our regression lines are relatively stable with an inverse effect between percentage of respondents and covid cases. I was surprised to see that counties where more people responded that they never wear masks but do socially distance a lot had a lower case count per 1000 people.

Graphical user interface

Description automatically generated with low confidence

The graph above confirms my hypothesis as we see a positive slope in the regression analysis of a high mask wearing survey population that does not socially distance while other populations remain relatively flat.

For policy purposes, as vaccinations are distributed and as warm weather approaches, this data could be used to dive more into transmission mitigation methods centered around providing fresh air and space to people. Some recommendations could be to invest more money in small businesses that would allow them to use the outdoors to attract clientele. This is especially a challenge in high-density cities, but ordinances allowing restaurants to take up road space could help provide the necessary space to keep transmission low.

While I was able to come up with some recommendations from my analysis, I would like to take a deeper dive into the data and spend more time comparing other variables, especially those revolving around polarization. The Fox News and CNN viewership divide as well as the ideology divide would have been another interesting aspect to look, and would have helped me loop in the 2020 election data as well.