Modeling plan

**3/27**

We decided to see if either incidence rates (new cases) or death rates tell us more about people's contact rates and whether they wear masks. We chose to look at the new cases happening right before we collected our data, instead of using all the cases added up from before. We adjusted these numbers by the size of the county's population, making them per 100,000 people, and added these to our model.

Because these rates were highly right-skewed, I logged them to get a more normal distribution.

For the sake of making interpretation easier, I made turned proportion of contacts carried out with a mask into a percentage.

We found that the new cases in the week before we gathered our data were a slightly better predictor than the prior week. I need to update our model to include similar numbers for death rates.

In short, our models now use a new number that shows the new cases in the county for the week before we collected our data. To decide on the specific week, we took the middle date of when we were collecting data and counted 7 days back from there to get our numbers.

**Everything we’ve spoken about as of 03/25**

We’ve decided to break up the paper into two parts: 1. Describing the difference between partisans along the lines of perception, contact rates, and mask usage during the pandemic and how these might be impacted by congressional voter context and 2. Showing how these differences can lead to very different rates of disease contagion and mortality rates in SIR models. For part 2, Ayesha shared the code for a heterogenous mixing model.

For part 1, we are using multivariate regressions to describe the relationship most accurately between political affiliation and health behaviors. Democrats and Republicans generally differ in several key areas such as race, age, gender, level of education, percentage of Hispanics, size of household, and whether they live in densely urbanized areas. However, their likelihood of being employed is similar, though the types of jobs they hold may vary. When members of one party, like Democrats or Republicans, live in districts primarily controlled by the opposite party, their characteristics tend to align more closely with those of the majority party in that area. Therefore, we will control for the following:

1. Race
   1. Hispanic versus not Hispanic
2. Gender
3. age
4. Employment status
5. Rural versus urban
6. Case counts
7. Mask mandates
8. Respondent household size
9. Educational attainment

We will be adding in these variables in three phases. First, a baseline model that includes only exogenous controls that allows us to see the most straightforward “truth” of the relationship. Second, a model that controls for sociodemographic variables. Third, a model that controls for county-level factors.

Given that the results do not change too much when we use the “national” sample versus the pooled sample, we’ve decided to use the pooled sample for the larger sample size. This means we’ll be utilizing weights that make this sample nationally representative (weight\_pooled). We’ll also be controlling for where the sample was collected, the day it was collected, and calendar month.

In part 2, we are using a heterogenous mixing model (SIR variant), designated as groups R and D, to examine the dynamics of disease contagion under varying protective behaviors. Within each group, compartments are further differentiated into "Protected (P)" and "Unprotected (U)" states to simulate the impact of protective measures, such as mask usage and vaccination, on the transmission and infection rates of a disease. Individuals in both groups may adopt protective behaviors through two primary mechanisms: (1) in response to observed deaths within a specific time window, highlighting the role of perceived threat on behavior change; and (2) via contact with other protected individuals, indicating the social influence on the adoption of preventive measures.

This model allows for variations in the rate of contact and the propensity to adopt protective behaviors between the groups. Additionally, the fraction of the population belonging to each group is variable, enabling the exploration of scenarios with differing population compositions of Republicans and Democrats. A key aspect of the model is the inclusion of homophily—the degree to which individuals within a group interact predominantly within their own group versus with those from the other group. However, we do not have information on whether people interact with other partisans (we’ll have to make some assumptions here).

According to Ayesha’s code, an important concept is protective behavior (mask-wearing mostly), which not only reduces the likelihood of transmission and infection by 30% but also varies in adoption based on two mechanisms: the number of deaths observed within a certain timeframe and contact with individuals already practicing protective behaviors. The rate of adopting protective measures differs between the two groups in her code; group A shows a lower threshold (80 deaths) for adopting protective behavior in response to fatalities, while group B requires a higher threshold (100 deaths). We have not yet decided how we will arrive at these thresholds for groups R and D. We’ll also need to decide how to assign differences in waning behavior between the two. Does waning behavior involve use counting contacts in a model where we adjust for prevalence?