The course of COVID-19 pandemic in US varies dramatically across regions and time. Popular discussions have emphasized the role of policies, income, age, belief in science, etc. However, the relationship between different factors and case growth is unlikely to be linear (or even parametric) and there are complex interactions between factors which co-determines the pandemic evolution. Clear empirical evidence connecting the complex predictor network to case growth and timely scenario-based simulations are essential to steering policy responses.

In this project, we will provide new evidence on the power of different spatial, temporal and administrative factors in predicting virus transmission curves and provide multiple counterfactual simulations for policy implications. Specifically, we will take three steps to achieve this goal:

1. Prediction of COVID transmission curve.  
   Method: Random forest or Neural network.  
   We will calibrate the disease transmission parameter in Susceptible-Infected-Recovery (SIR) model on a monthly basis. There is a mathematical relationship between the growth rate of COVID cases and R0, the essential input parameter of SIR epidemiological model. We will thus predict the empirically derived COVID growth rate using policies, local socio-demographic factors, political ideology, climate conditions, days since COVID outbreak and their interactions.
2. Counterfactual simulations.   
   We will use the relationship we built between in previous step to simulate different counterfactual scenarios of the virus transmission curve given different policy scenarios (e.g., the timing of stay-at-home orders), various climate conditions (e.g., cold or hot), and heterogeneous local socio-demographic compositions (e.g., share of elderly population), etc. This can serve as an effective tool for policy makers of different jurisdictions to understand the risk and tailor their policy interventions.
3. Predict the onset of second/ third wave.   
   Method: LASSO penalized logistic regressions.   
   The above exercise will provide tools for controlling a given wave of pandemic peak. Yet in US and Europe, we see a second or even third wave which is even more severe than the first wave at the onset of pandemic. There is a desperate need to understand what causes the returns of pandemic and how can we foresee its coming to forestall or mitigate it. For the last step, we will use the policy portfolio, socio-demographic characteristics, mobility by places and weather conditions in the previous month of pandemic resurge to predict the onset of second/ third wave.

Putting together, we wish our predictions using data science approach can better capture the complex interactions between virus transmission moderators and inform the debate on optimal policy responses accounting for the local contexts of each US county.