

An Analysis of US Social Distancing Amidst the Covid-19 Pandemic

In an effort to allow researchers to better understand how the world is social distancing, Google has collected and released data on the mobility of the public during this pandemic, comparing movement to a baseline, defined as the median value, for the corresponding day of the week, during the 5-week period of 1/3–2/6, 2020. Google grouped the mobility of the public into six location categories, including: Retail/Recreation, Grocery/Pharmacy, Parks, Transit Stations, Workplaces, and Residential areas. Our dataset contains mobility data for every US state during three dates: 3/29, 4/5 and 4/11. Our main goals included exploring the differences in social distancing measures in different regions of the country, to determine why these differences exist and explore which factors are most strongly correlated to them.

In order to assist us in data exploration, we calculated a *Social Distancing Score* by normalizing the mobility data. In general, the country distanced when it came to workplaces, transit stations, and retail/recreation, yet increased their activity at parks. Furthermore, while Americans initially avoided grocery stores and pharmacies, they began to return to the supermarkets after weeks of quarantining. The *Social Distancing Score* also enables us to determine which states were the best (Hawaii, New York, New Jersey, Vermont, and Nevada) and worst (Nebraska, Arkansas, Iowa, South Dakota, and Wyoming) at social distancing. Plotting social distancing scores on a map of the country revealed significant regionality when it came to distancing. The northeast, west coast, and gulf coast states were the best when it came to distancing, while the midwest and the deep south were the worst.

In order to better understand the factors that lead to social distancing and its effectiveness, it is imperative that we compare social distancing between states that are on a similar timeline in terms of the virus' spread. In order to control for these differences, we clustered the US by similarities in initial conditions (number of cases on 3/29) and by their population densities. In all clusters, while the cases/capita increased over the three week period, the rate of spread of coronavirus had decreased, corresponding to the "flattening of the curve" that public health officials and countries that are further along in disease progression have promised. Controlling for cluster, those who distanced more initially (on 3/29), saw the greatest reduction in rate of spread. For example, while Rhode Island had a social distancing score in the 80th percentile, they were last in their cluster, and became the only state among all states to experience an increase in the rate of spread. In other words, given their starting point of cases and population density, they needed to be much more diligent in their distancing.

We fit a linear model (shown on the right) to determine what drives certain areas to take distancing more seriously than others. Notably, over 70% of states that had positive social distancing scores had Democratic governors. In all clusters, states governed by a Democrat had a higher median social distancing score than Republican governed states. This may be partially driven by the earlier stay at home orders enacted by Democratic governors. Excluding the six Republican governed states that have yet to declare stay at home orders, Democratic governors enacted stay at home orders an average of three days before their Republican counterparts.

These data give credence to the effectiveness of social distancing while revealing that social distancing is not randomly spread throughout the country, but a consequence of many factors, such as population density, initial cases/capita, governor political affiliation, stay at home orders, and geographical location, that differ between states.

Characteristic	Beta	95% CI ¹	p-value
First Case	-0.08	-0.17, 0.01	0.086
Stay at Home Order	-0.30	-0.46, -0.15	<0.001
Governor			0.049
Democrat	—	—	
Republican	-2.8	-5.6, -0.01	
Cluster	0.99	-0.31, 2.3	0.13

¹ CI = Confidence Interval
Linear regression predicting Social Distancing Scores
Multiple R²: 0.4766