Forecasting Model Similarity (for inc hosp)

Johannes Bracher, Evan Ray, Nick Reich, Nutcha Wattanachit, Li Shandross

06/28/2021

COVID-19 Forecasting Model Similarity Analysis for 1-4 Week Ahead Incident Hospitalization

In this analysis, we extend the work of Bracher, et. al that evaluates the similarity between Covid-19 models using Cramer's distance. This work was applied to forecasts for incident deaths (inc deaths) and incident cases (inc cases) and here we apply it to incident hospitalizations (inc hosp).

We must first make some adjustments to the inc hosp forecast data, which have a temporal resolution of "day" instead of "week," unlike inc case and inc death. This presents a challenge because the horizons will be different for the same target end date if the forecast dates between two models differ by only a single day. (This is not an issue when the temporal resolution is in terms of weeks, which are defined by epidemiological week, not the number of days between forecast date and target end date.) Thus, we create a new variable called horizon week to solve this issue. This variable counts horizons between 1 and 7 to have a horizon week of 1, horizons between 8 and 14 to have a horizon week of 2, etc. Hence, the analyses used on inc death and inc cases should be able to be applied easily to the inc hosp data.

This initial analysis only examines target end dates for a single day of the week, Thursday, to account for models that include day of the week effects. However, this will be expanded later to more days, perhaps even exploring if performing the analysis on different days of the week leads to different results in terms of model similarity.

The pairwise approximated Cramer's distances are calculated for the models that have complete submissions for all targets, all probability levels, and no missing forecasts between December 17th, 2020 and June 10th 2021. We aggregate results for the five locations that have the highest number of number of COVID-19 hospitalizations during this period as well as the five locations with the lowest number for the same date range.

There are 5 models that fulfilled the criteria for both the 5 locations with highest cumulative deaths and the 5 locations with lowest cumulative deaths.

Model types

Model	Type
Covid19Sim-Simulator	mechanistic
COVIDhub-ensemble	ensemble
JHUAPL-Bucky	neither stats nor mech
JHUAPL-SLPHospEns	ensemble
Karlen-pypm	mechanistic

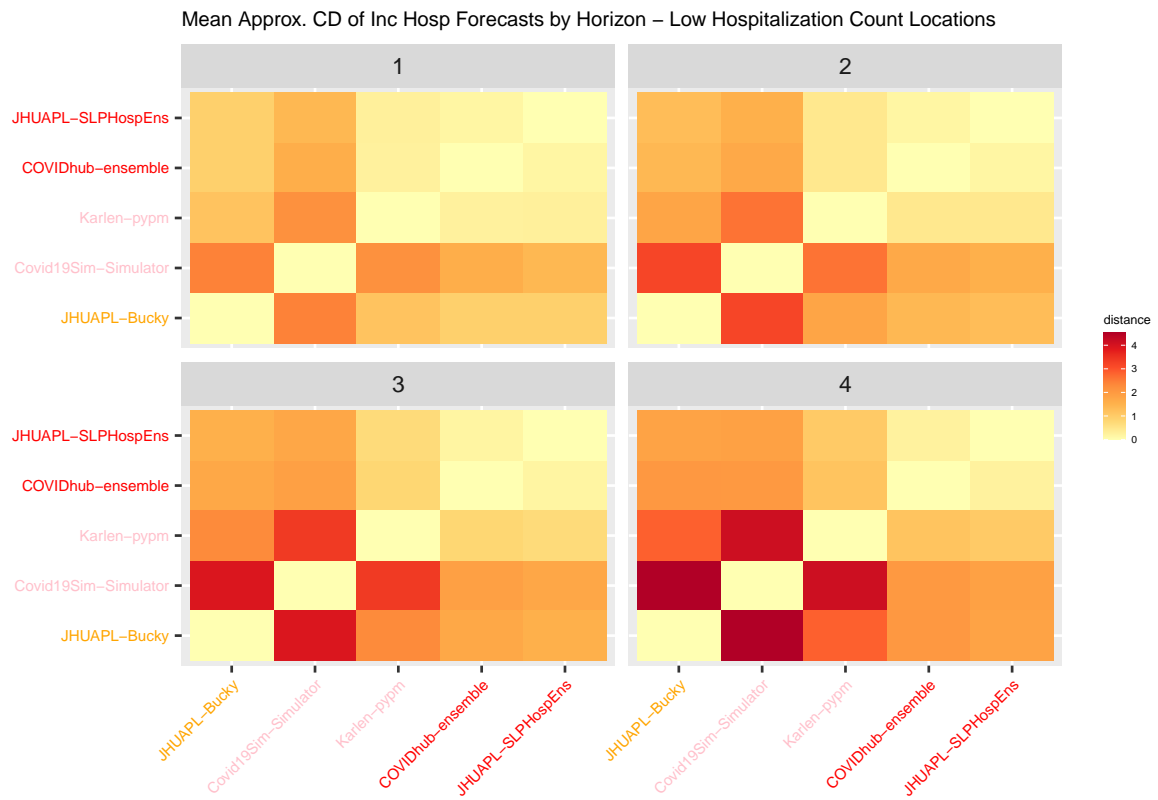**Differences between two approximations (for high count locations only)**

The approximated pairwise Cramer's distances between each forecast and the ensemble are calculated using both types of approximations to check for any large discrepancies between the two methods. The table below shows the averaged approx. CD over all target end dates and all 5 high count locations.

Table 2: Mean approx. CDs relative to the ensemble

Anchor Model	Model	Horizon	Target	CD (uneq)	CD (eq)	Diff
COVIDhub-ensemble	JHUAPL-SLPHospEns	1	inc hosp	6.99	9.51	-2.52
COVIDhub-ensemble	Karlen-pypm	1	inc hosp	25.32	25.58	-0.25
COVIDhub-ensemble	COVIDhub-ensemble	1	inc hosp	0.00	0.00	0.00
COVIDhub-ensemble	JHUAPL-Bucky	1	inc hosp	66.88	59.19	7.69
COVIDhub-ensemble	Covid19Sim-Simulator	1	inc hosp	188.50	164.09	24.41
COVIDhub-ensemble	JHUAPL-SLPHospEns	2	inc hosp	7.86	9.74	-1.89
COVIDhub-ensemble	COVIDhub-ensemble	2	inc hosp	0.00	0.00	0.00
COVIDhub-ensemble	Karlen-pypm	2	inc hosp	30.76	29.29	1.46
COVIDhub-ensemble	JHUAPL-Bucky	2	inc hosp	78.11	67.48	10.63
COVIDhub-ensemble	Covid19Sim-Simulator	2	inc hosp	173.03	148.59	24.44
COVIDhub-ensemble	JHUAPL-SLPHospEns	3	inc hosp	9.59	11.18	-1.59
COVIDhub-ensemble	COVIDhub-ensemble	3	inc hosp	0.00	0.00	0.00
COVIDhub-ensemble	Karlen-pypm	3	inc hosp	44.36	39.81	4.55
COVIDhub-ensemble	JHUAPL-Bucky	3	inc hosp	74.41	63.11	11.30
COVIDhub-ensemble	Covid19Sim-Simulator	3	inc hosp	144.47	124.19	20.27
COVIDhub-ensemble	JHUAPL-SLPHospEns	4	inc hosp	13.40	16.64	-3.25
COVIDhub-ensemble	COVIDhub-ensemble	4	inc hosp	0.00	0.00	0.00
COVIDhub-ensemble	Karlen-pypm	4	inc hosp	69.64	61.76	7.88
COVIDhub-ensemble	JHUAPL-Bucky	4	inc hosp	111.10	93.93	17.17
COVIDhub-ensemble	Covid19Sim-Simulator	4	inc hosp	131.71	114.51	17.20

We notice that Covid19Sim-Simulator has a much higher difference between its equal and unequal Cramer Distances compared to the other models, except for at a four week horizon, where JHUAPL-Bucky has a similarly high difference between its Cramer Distances.

We can visualize the mean approximated pairwise distances across all time points in a heat map shown below. The distance from the model to itself is zero. The $x-$axis is arranged based in an ascending order of the model's approximate pairwise distance from the COVIDhub-ensemble. So, the first model is the model that is most dissimilar (on average) to the ensemble in this time frame.

Mean Approx. CD of Inc Hosp Forecasts by Horizon – High Hospitalization Count Locations

Mean Approx. CD of Inc Hosp Forecasts by Horizon – Low Hospitalization Count Locations
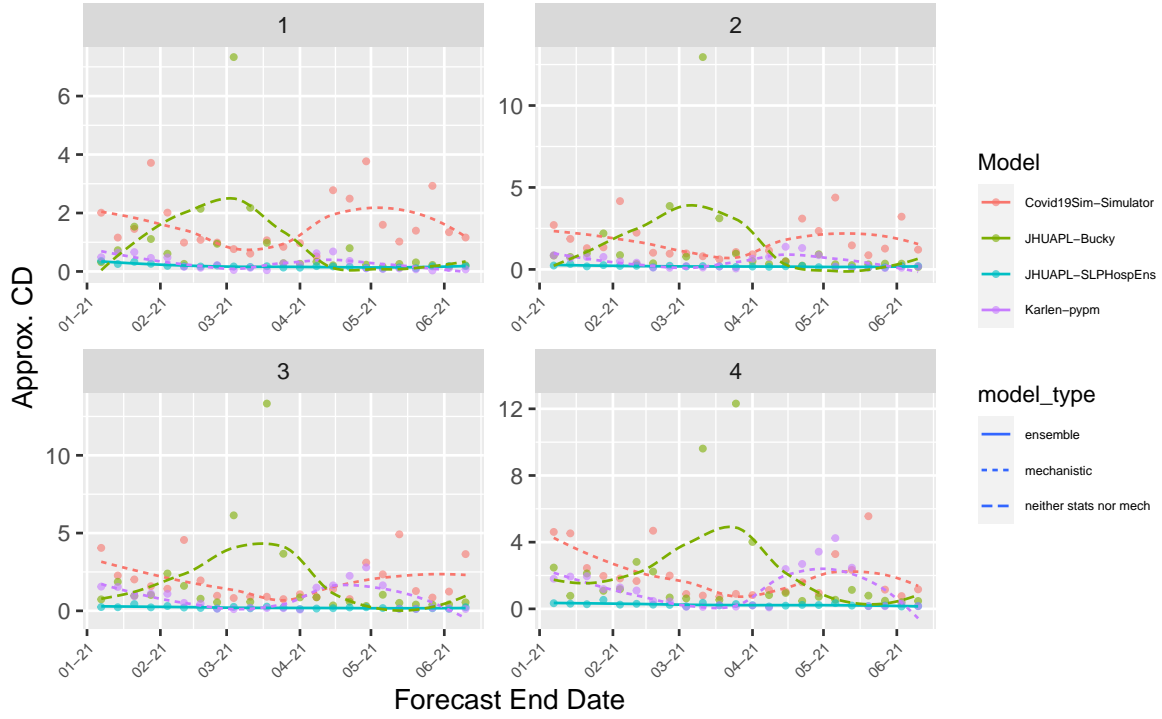
It appears that the Covid19Sim-Simulator is the least similar to the other models for both the high count

and low count locations, across all four horizons. JHUAPL-Bucky is the second least similar to the other models. However, high hospitalization count locations show a greater variation between models, especially at lower horizons. In low count locations, Covid19Sim-Simulator and JHUAPL-Bucky have more similar Cramer Distances. Of note is the greater difference between model at low horizons for high count locations, which is the opposite of the results shown for inc cases and inc deaths.

We can also look at the approximated pairwise distances to see how the models become more similar or dissimilar over time.

Mean Approx. CD from COVIDhub–ensemble Over Time –
High Hospitalization Count Locations
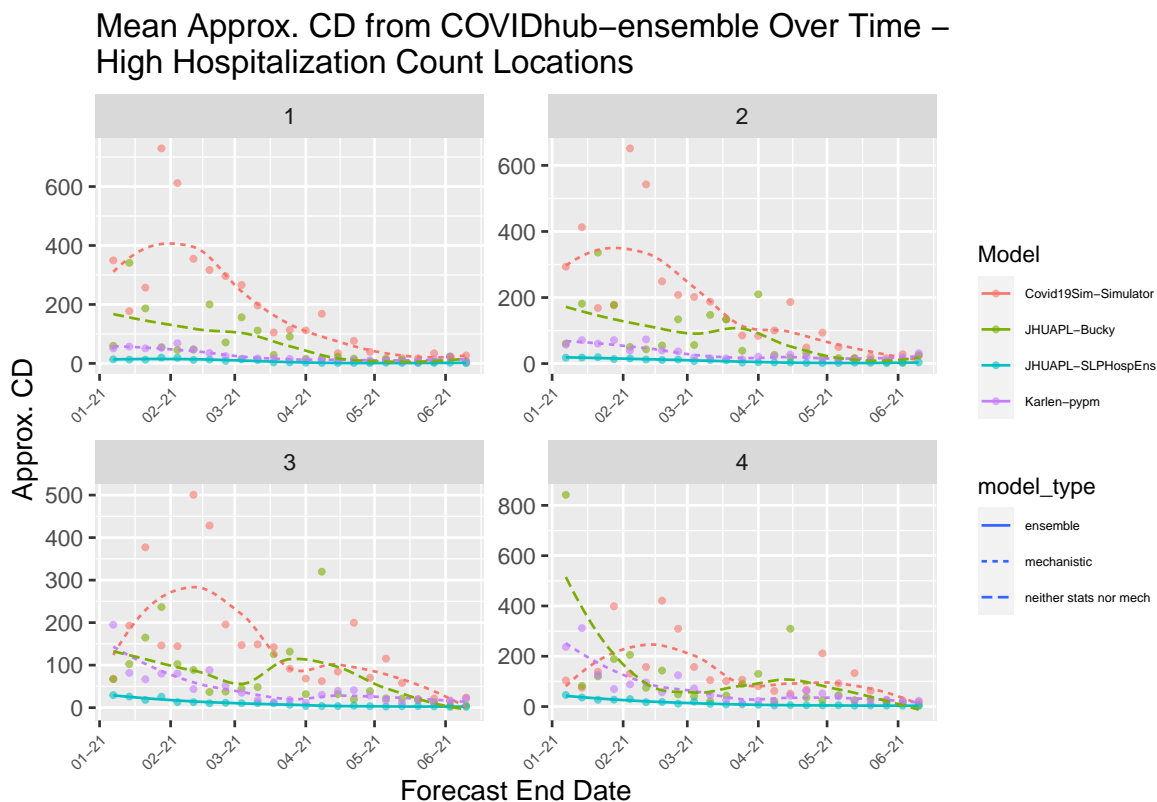
Mean Approx. CD from COVIDhub−ensemble Over Time −
Low Hospitalization Count Locations

The scatterplots show that the Covid19Sim-Simulator and JHUAPL-Bucky models tend to differ from the Covidhub-ensemble model compared to the other two models. This seems to align with the results shown in the heat maps above that show that Covid19Sim-Simulator and JHUAPL-Bucky tend to have the highest mean Cramer's Distance from the other models. In high count locations, Covid19Sim-Simulator is very different from the ensemble model from January until April. However, in low count locations, JHUAPL-Bucky shows a peak in late March/early April, although this peak is not largely different, as the scale is pretty small.

Hosp Truth Over Time – High Count Locations

Hosp Truth Over Time – Low Count Locations
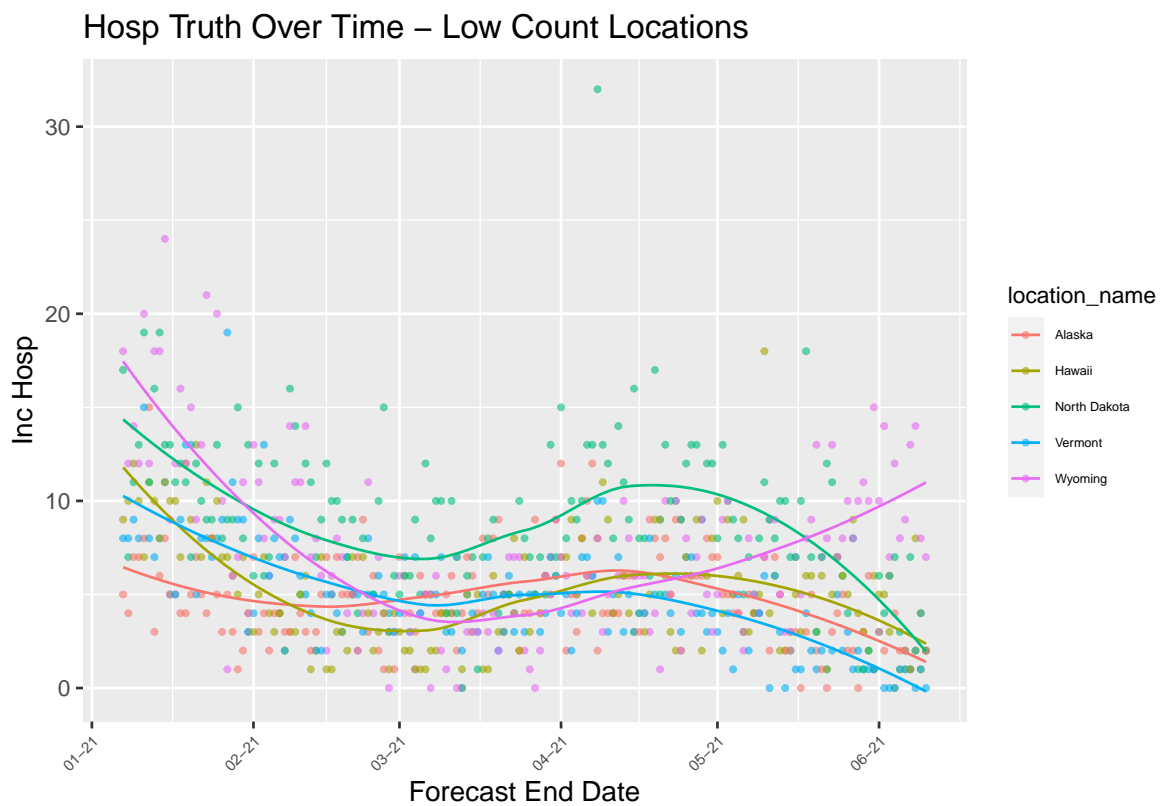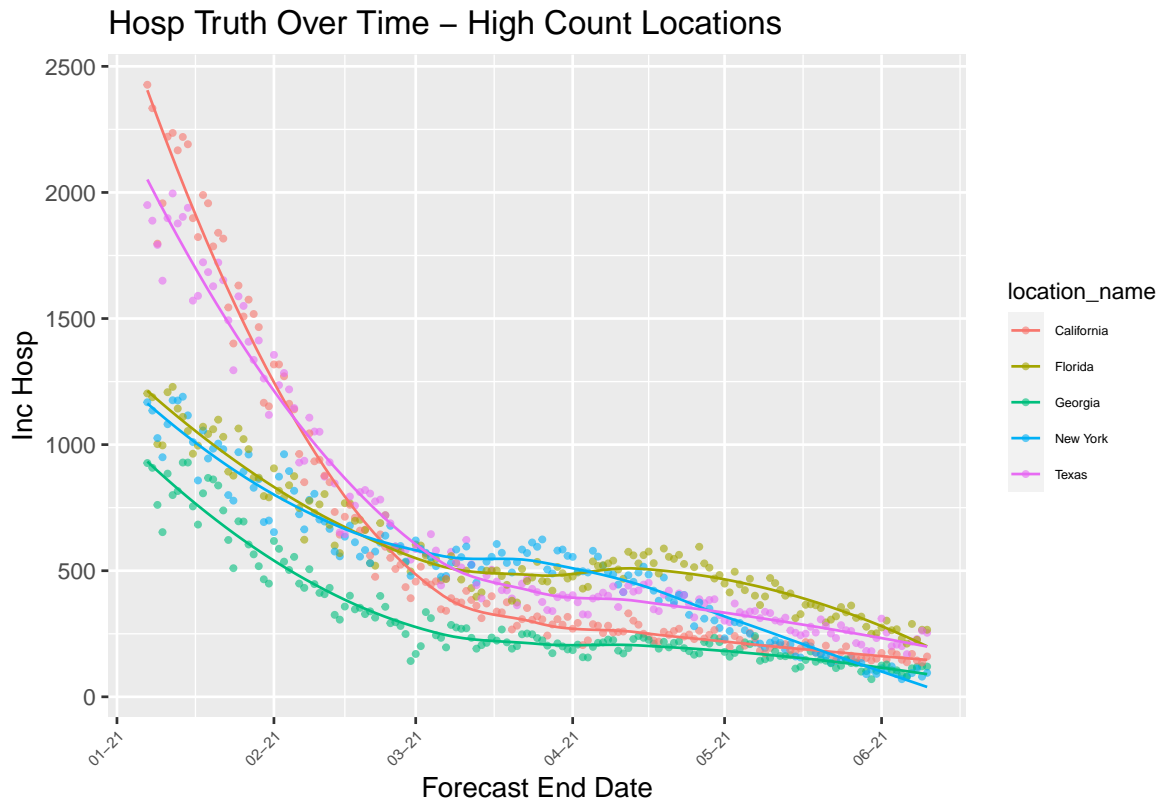
It seems that the Covid19Sim-Simulator's differences from the ensemble model are follow the trends shown

by the truth data.
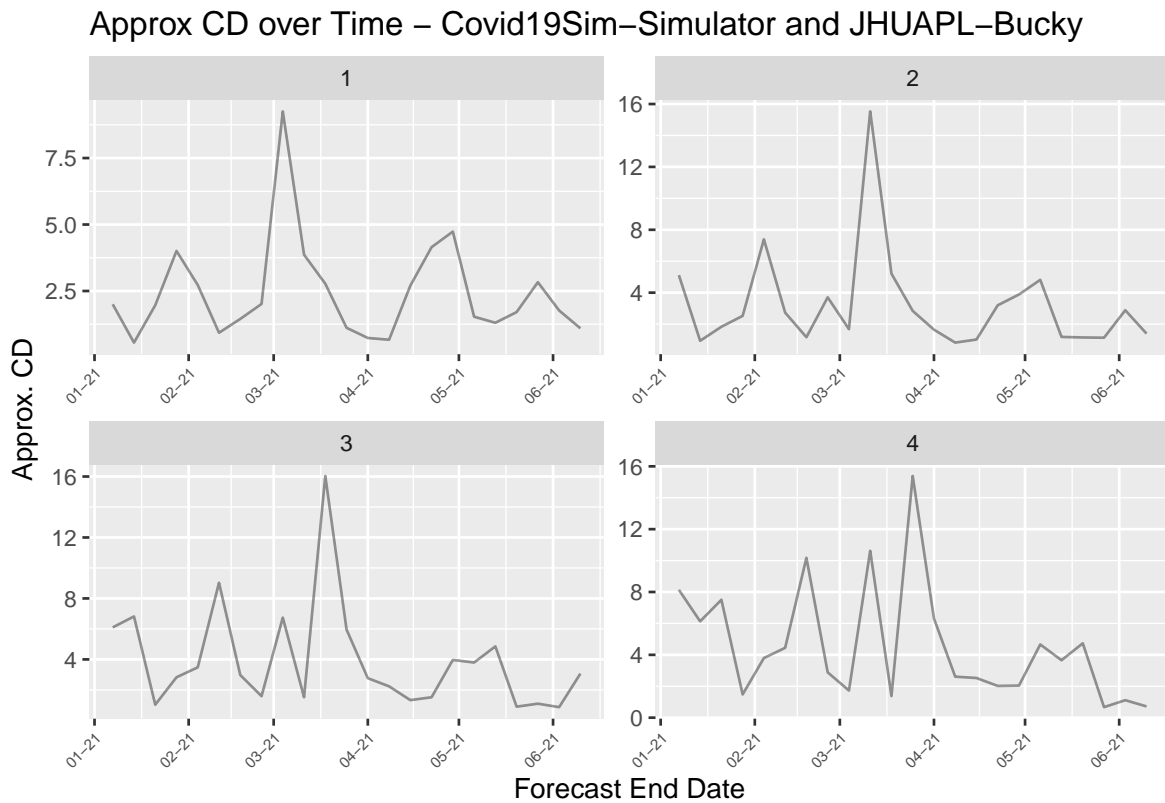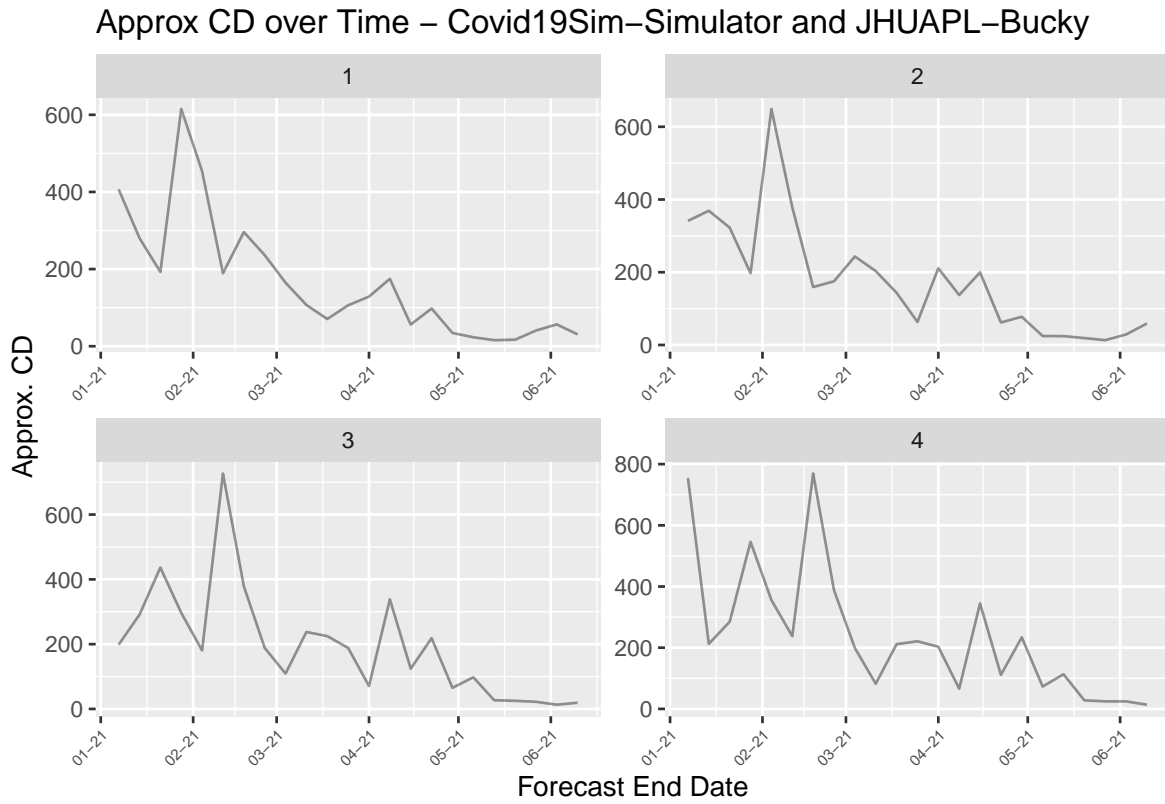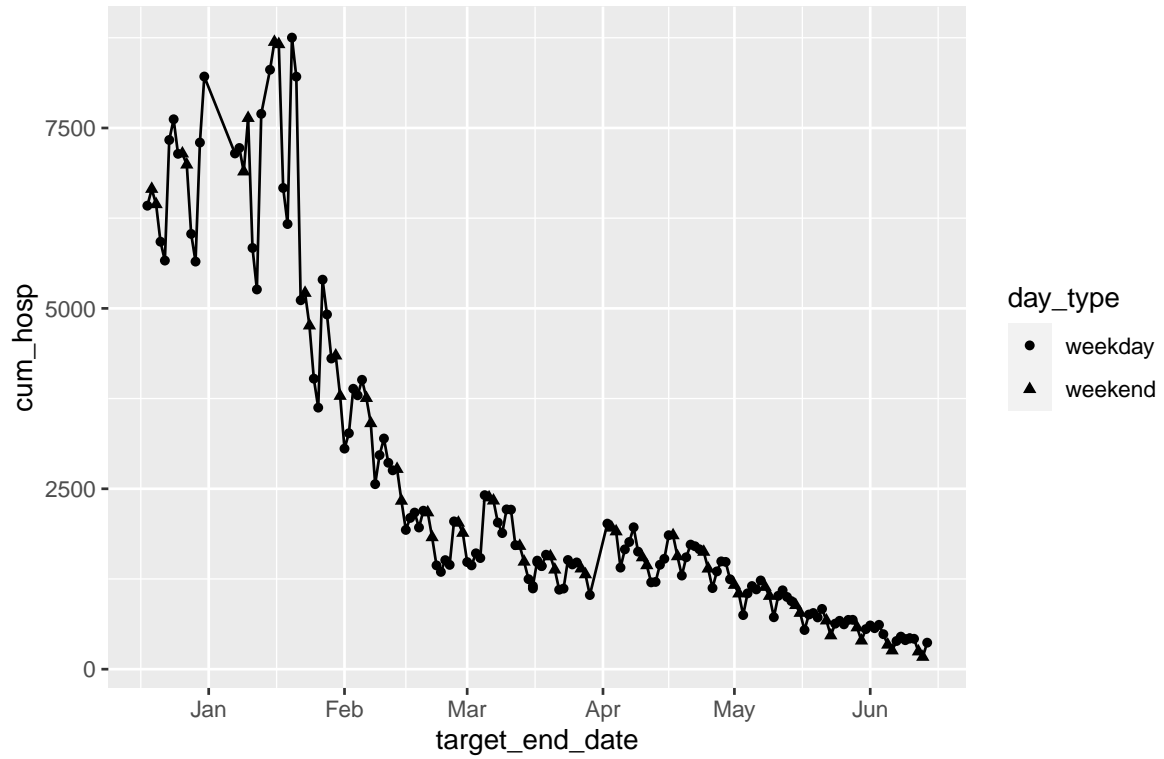
Approx CD over Time – Covid19Sim–Simulator and JHUAPL–Bucky

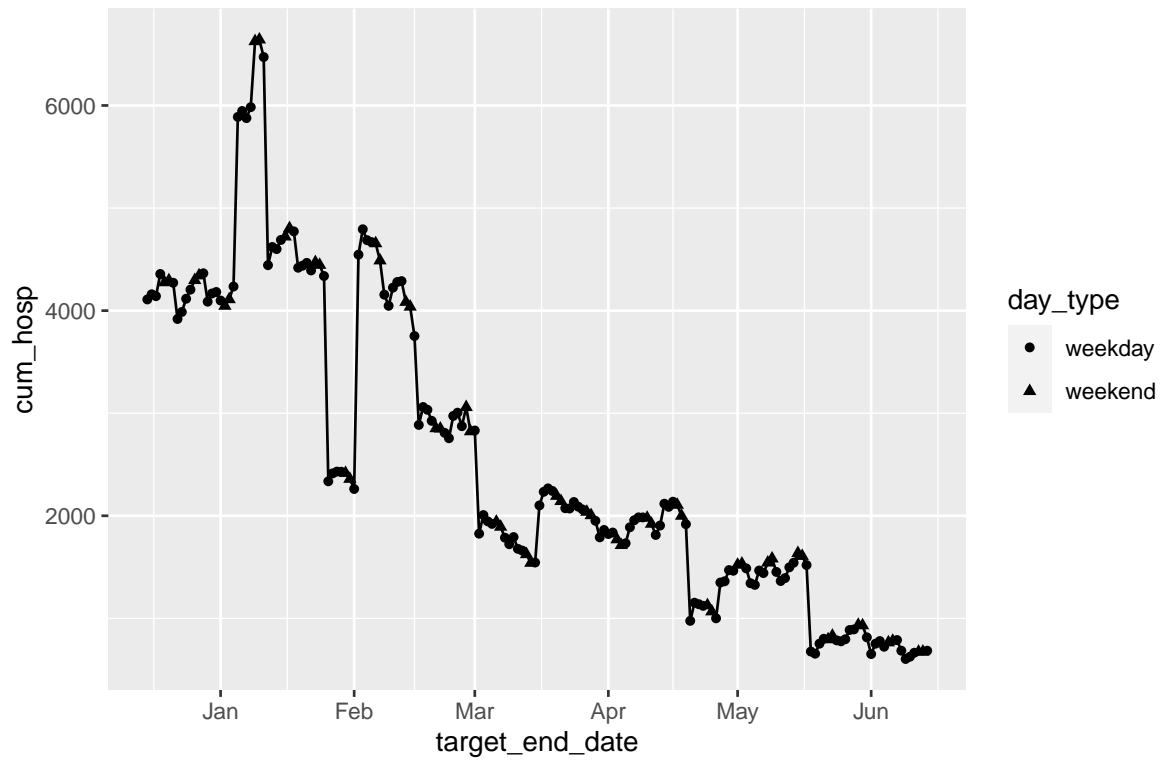Approx CD over Time – Covid19Sim–Simulator and JHUAPL–Bucky
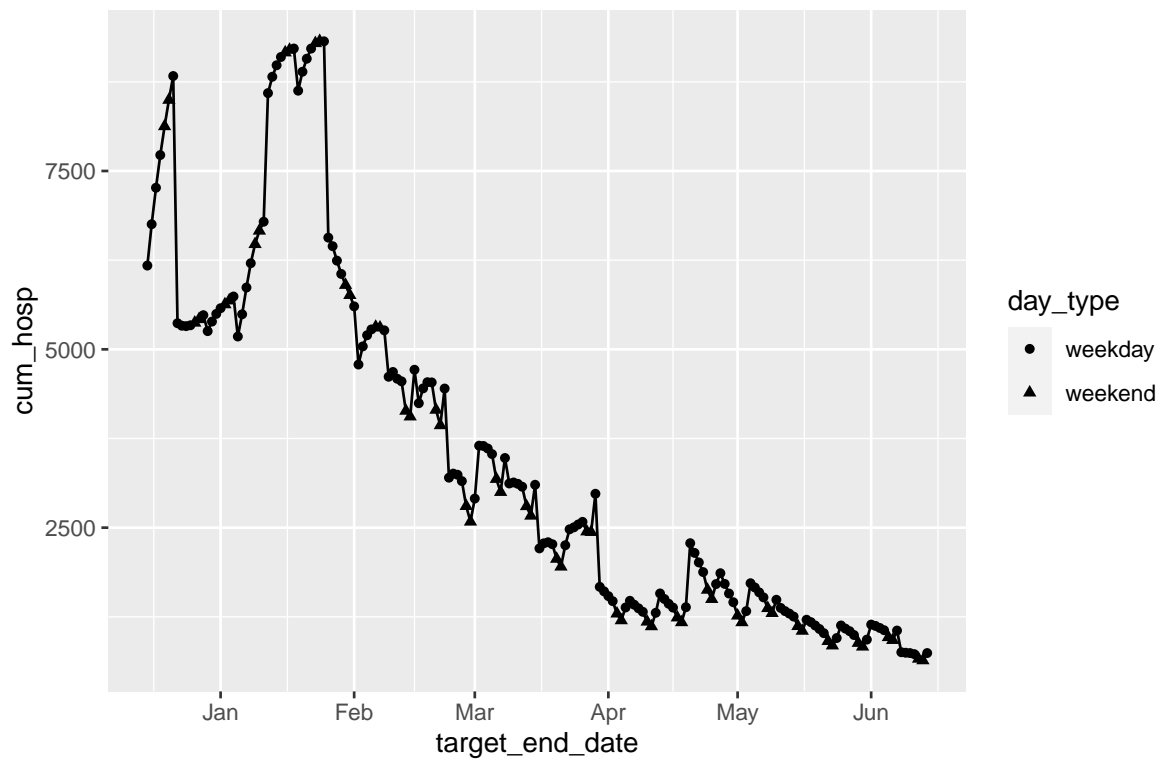
LANL−GrowthRate Forecasts
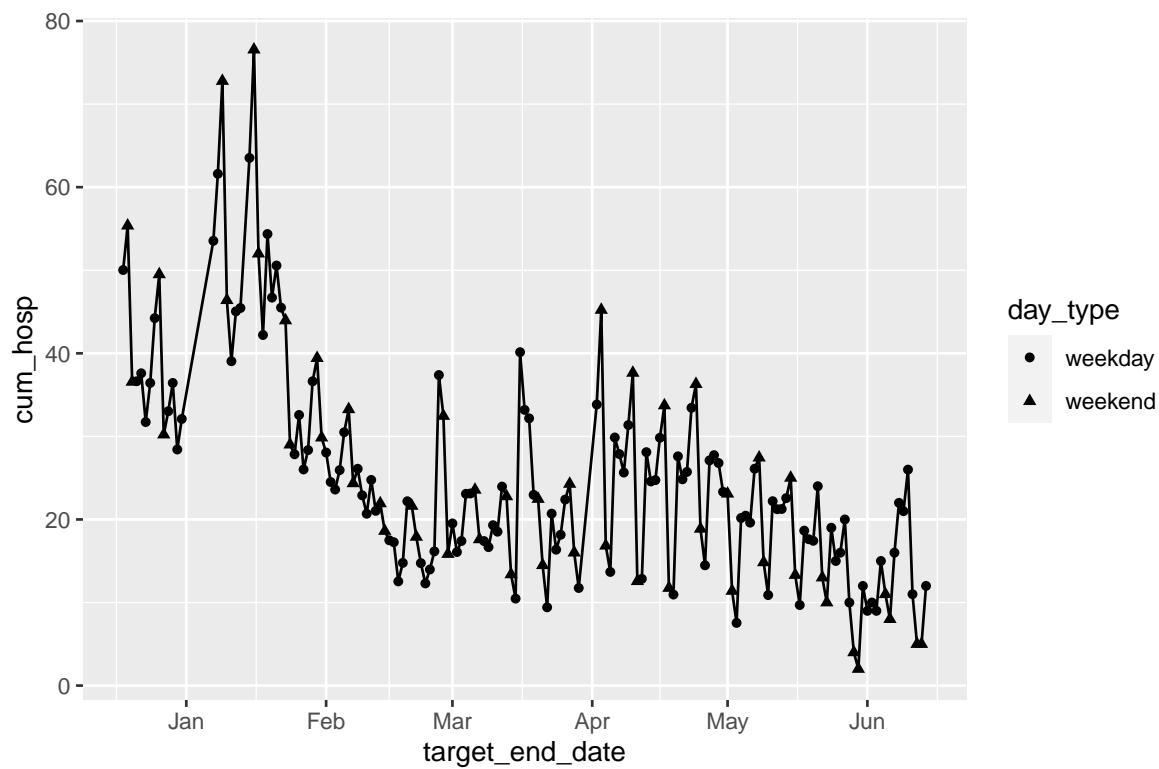

COVIDhub−ensemble Forecasts
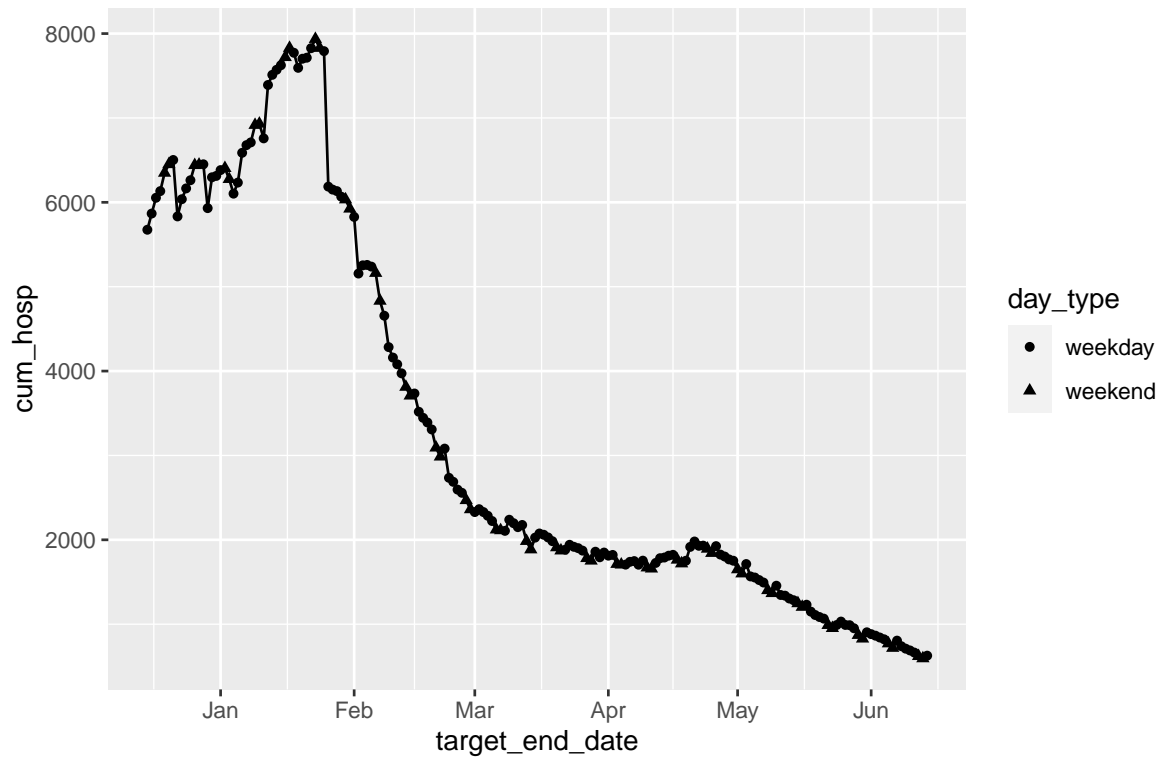
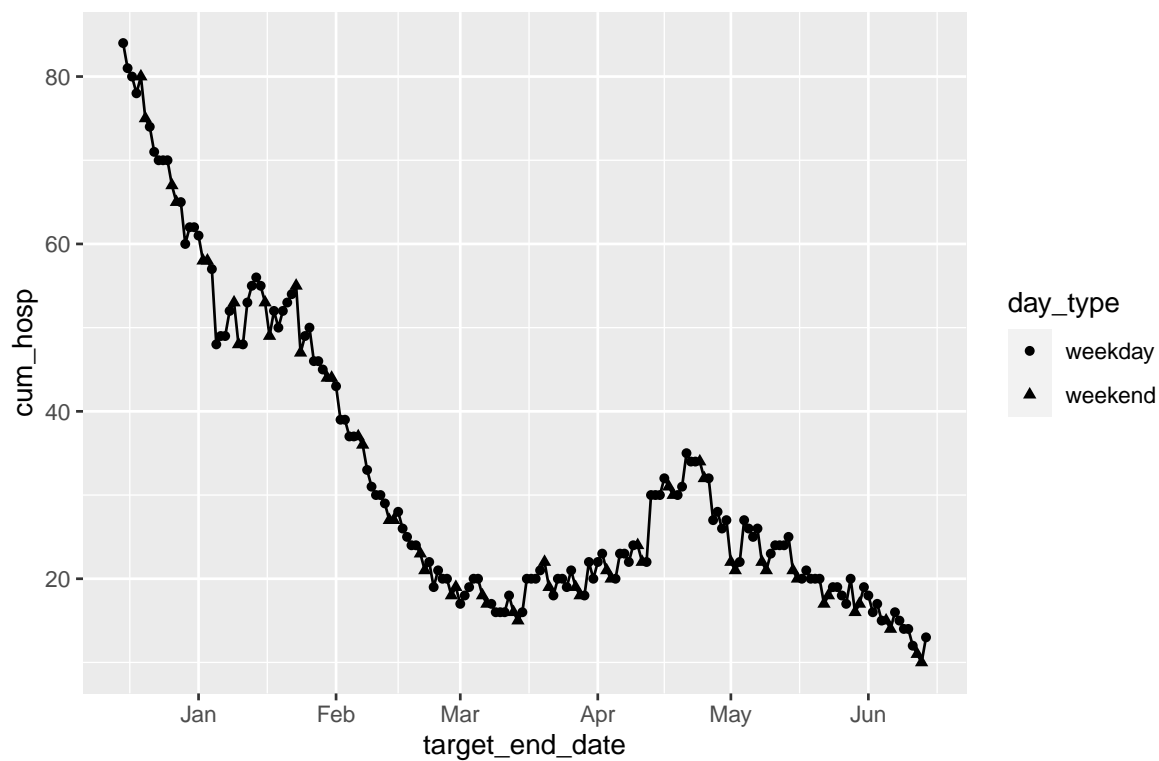GT–DeepCOVID Forecasts

JHUAPL–Bucky Forecasts
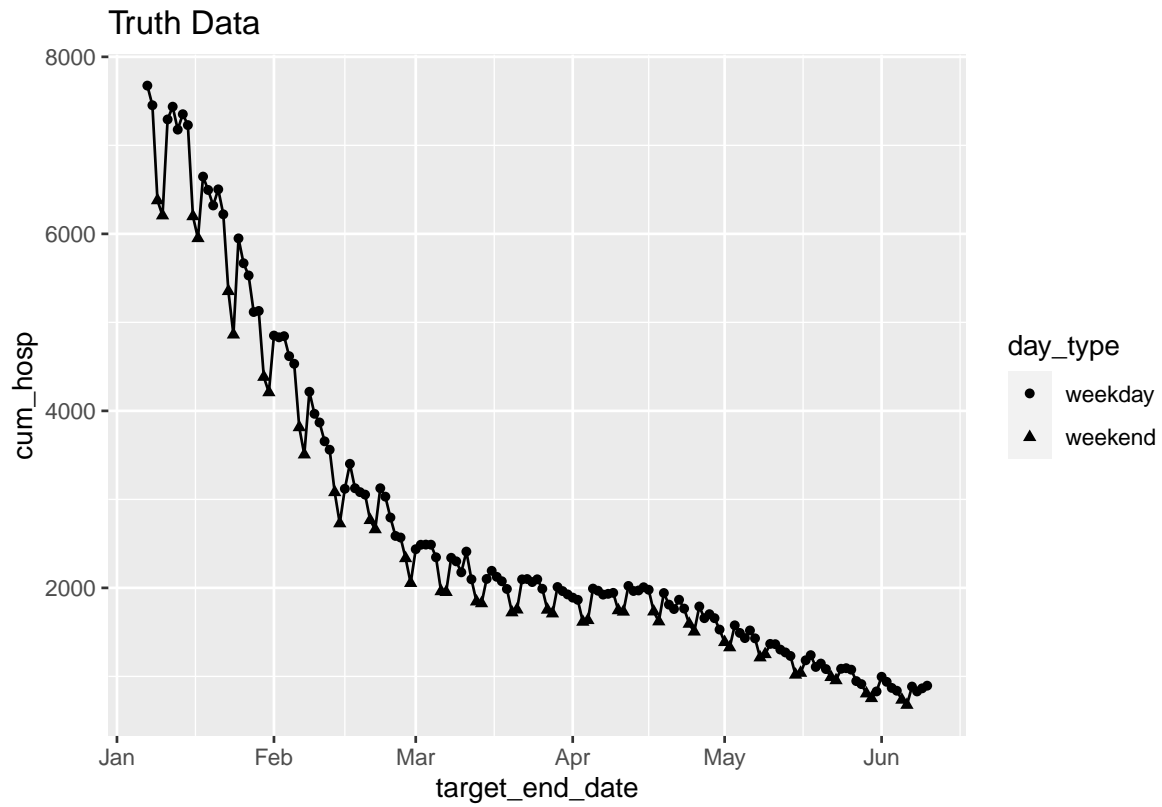
Truth Data

Truth data has always shown day of the week effects (weekends always have lower inc hosp) but the models seemed to predict lower inc hosp on Tuesdays for the end of December/beginning of January, then most seem to switch to lower weekend end inc hosp sometime during 2021.
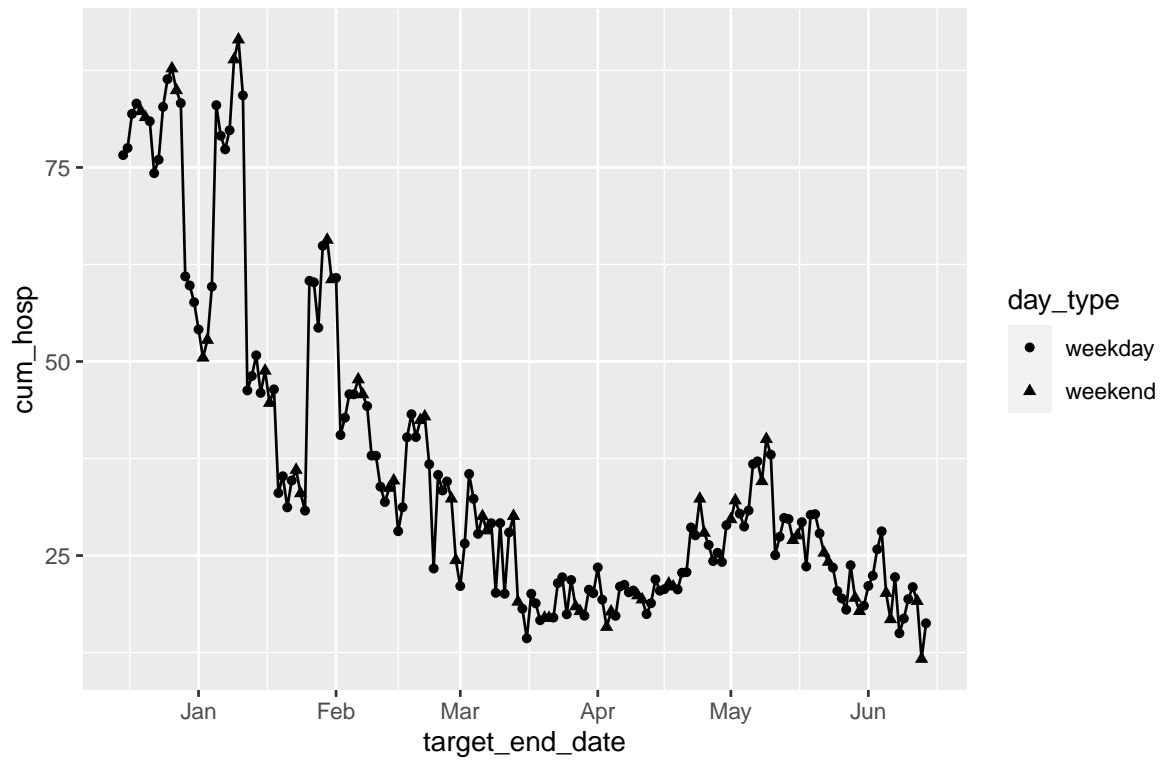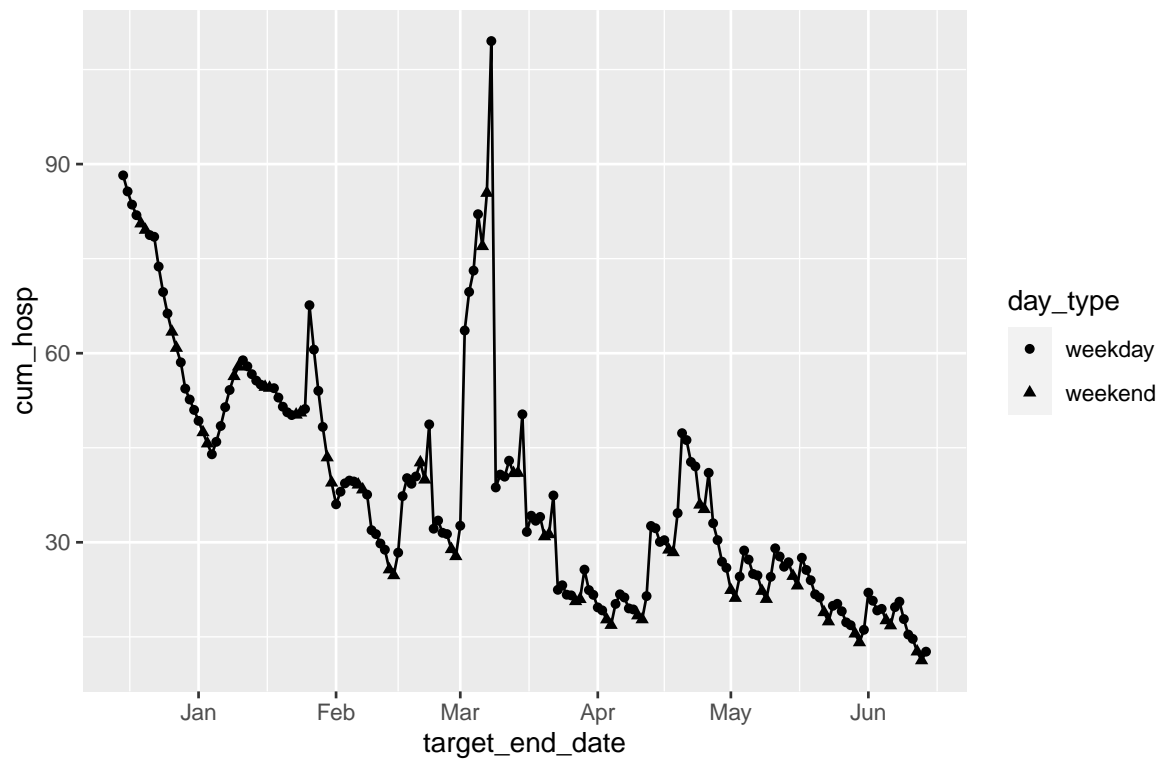
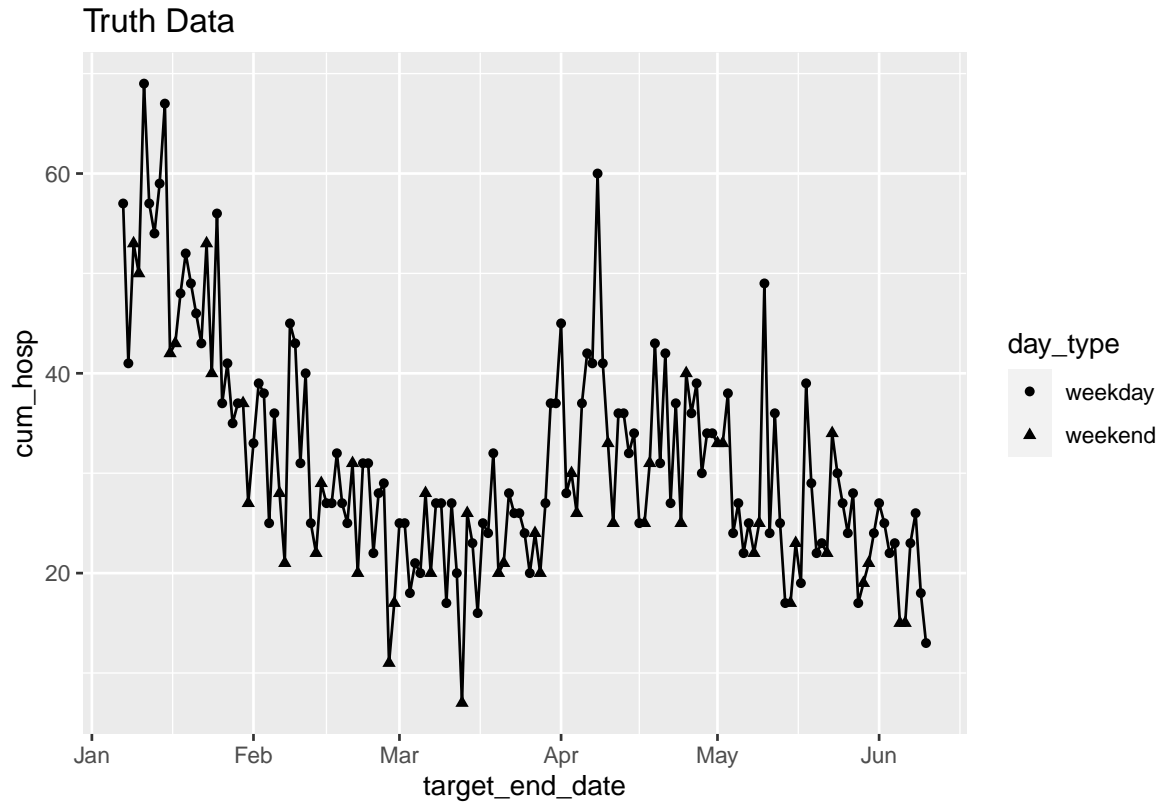LANL−GrowthRate Forecasts

COVIDhub−ensemble Forecasts

GT−DeepCOVID Forecasts

JHUAPL−Bucky Forecasts

Truth Data

The low hosp location truth data doesn't present such a clear pattern as to day of the week effects that suggest that weekends or weekdays generally have less inc hosp values. However, most of the models seem to switch to incorporating a day of the week effect at about the same time as that for the high count values.

Hosp Truth and Forecasts Over Time
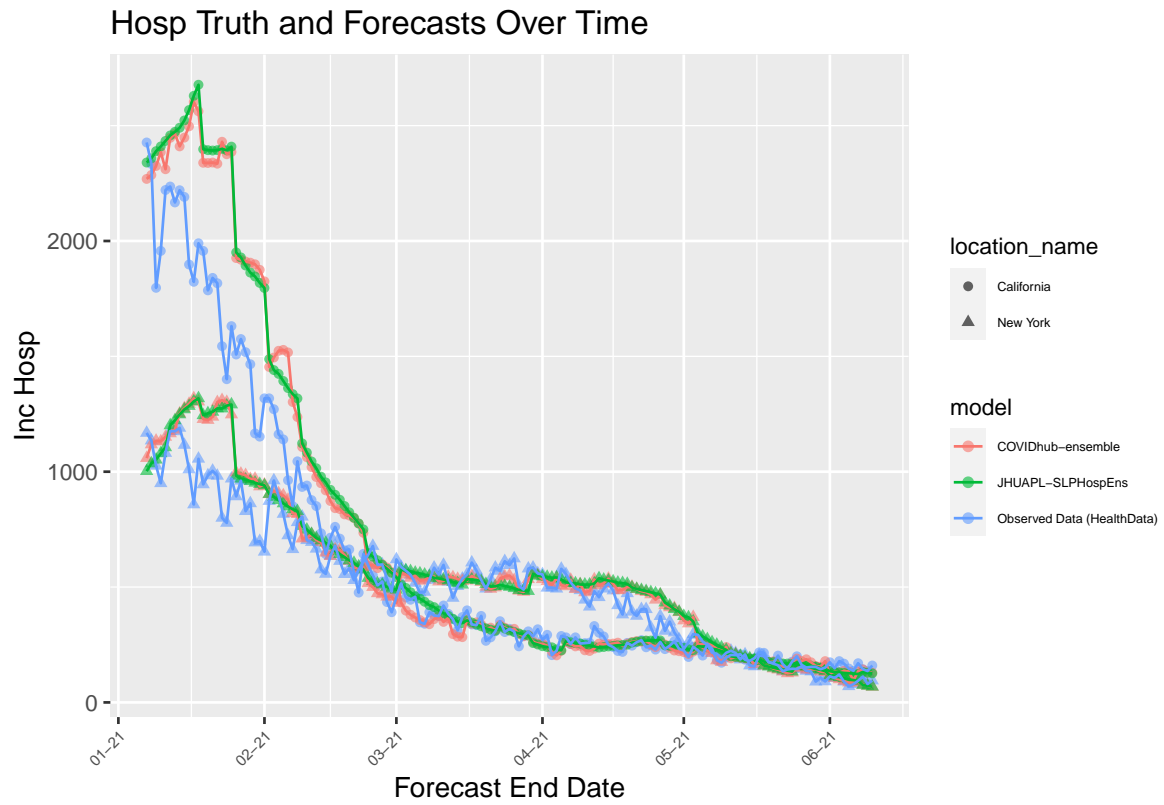
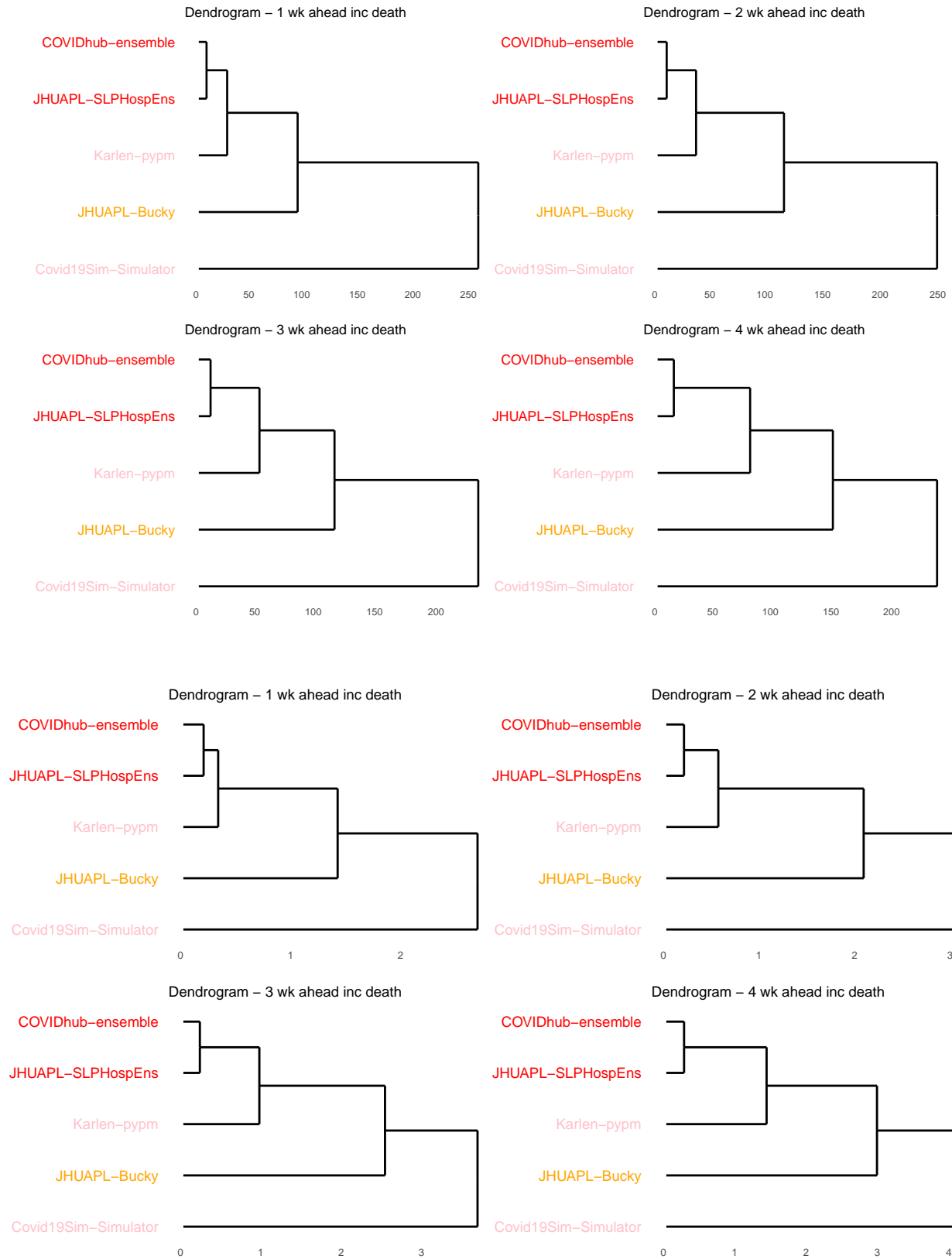We can also cluster the distances using hierarchical clustering.

Figure 1: Low Mortality Count Locations

For both high count and low count locations, all of the dendrograms show that Covid19Sim-Simulator is the most dissimilar to the other models, followed by JHUAPL-Bucky. However, the scale for differences in Cramer Distance for the low count locations is very small, which may be a result of low hospitalizations.

Overall, it seems that Covid19Sim-Simulator and JHUAPL-Bucky are pretty consistently the two models then are the most dissimilar from the other three models, across almost all horizons for both high-count and low count regions.