

COVID-19 Forecast Similarity Analysis for Hospitalizations

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

07/20/2021

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.

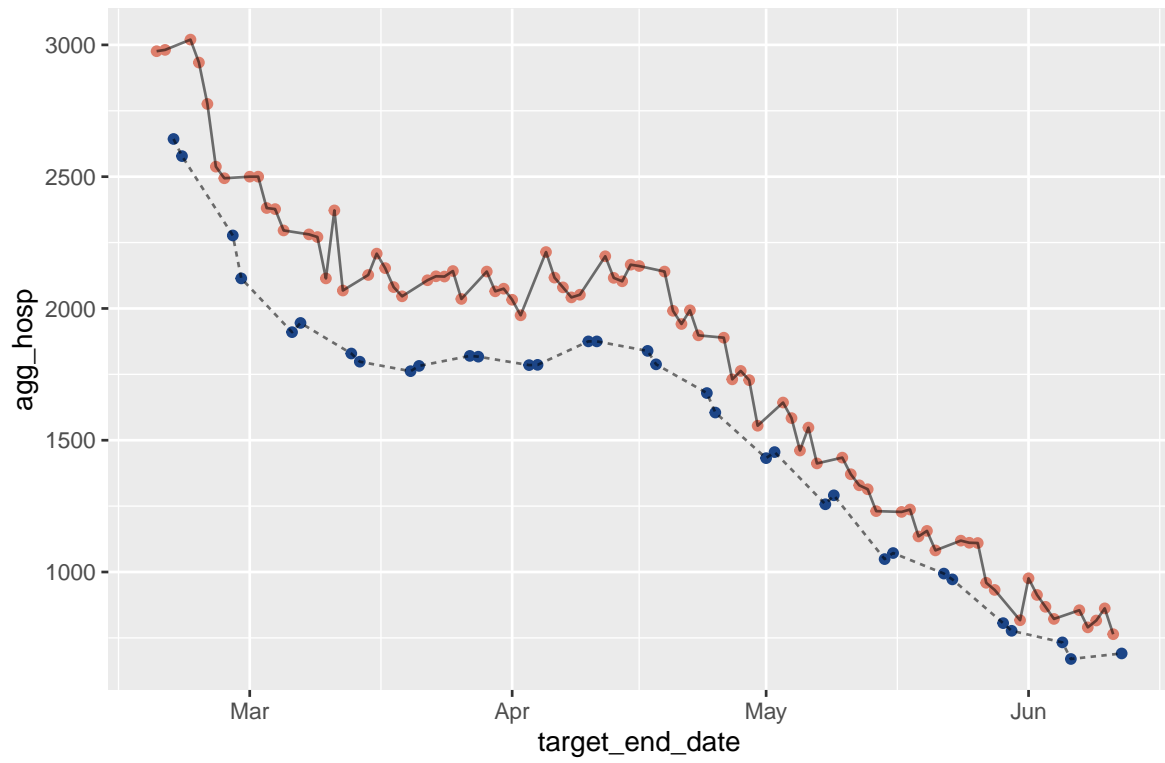
The pairwise approximated Cramer's distances are calculated for the models that have complete submissions for all targets, all probability levels, and no missing forecasts between January 28th, 2020 and June 10th 2021. We aggregate results for the five locations that have the highest number of number of COVID-19 hospitalizations during this period as well as the five locations with the lowest number for the same date range.

There are nine models that fulfilled the criteria for both the five high count and five low count locations for Thursday forecasts. There are ten that fulfill such criteria for the Saturday forecasts, but we perform the analysis on only the overlapping nine models.

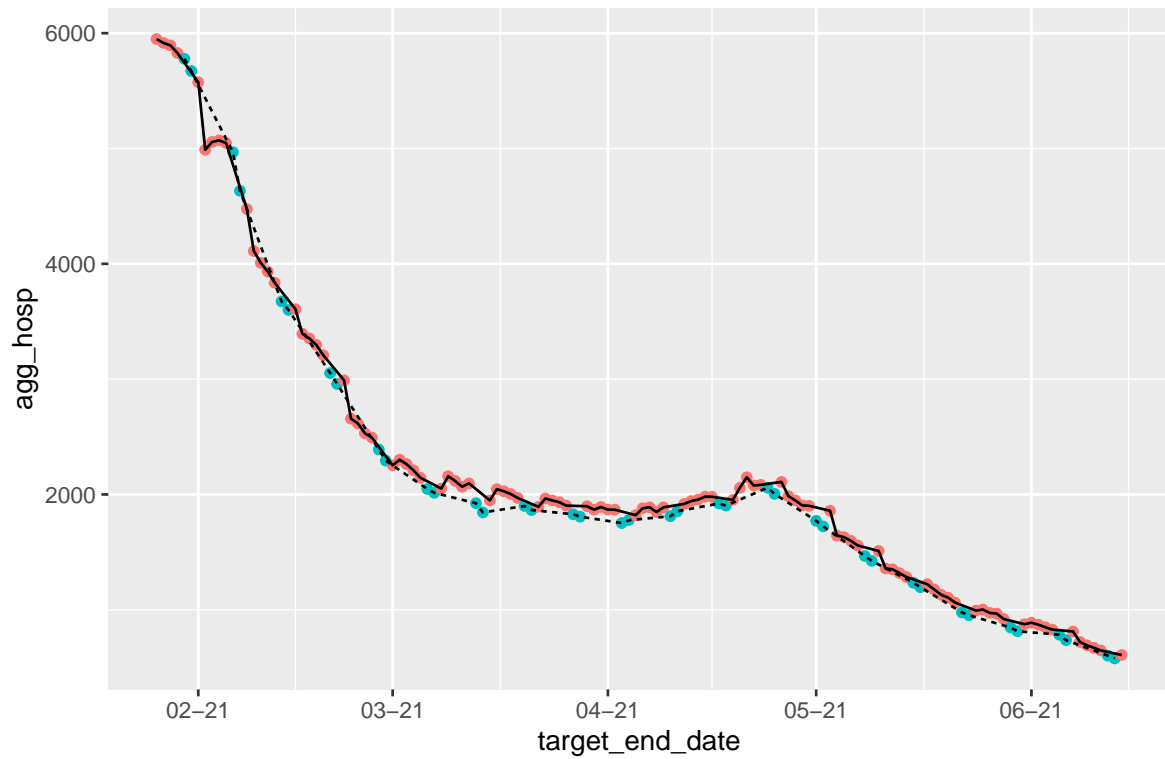
Day of the Week Effects

We can plot point forecasts to see if any of the models incorporate day of the week effects (i.e. do forecasts certain days of the week specifically show lower/higher hospitalizations compared to other days).

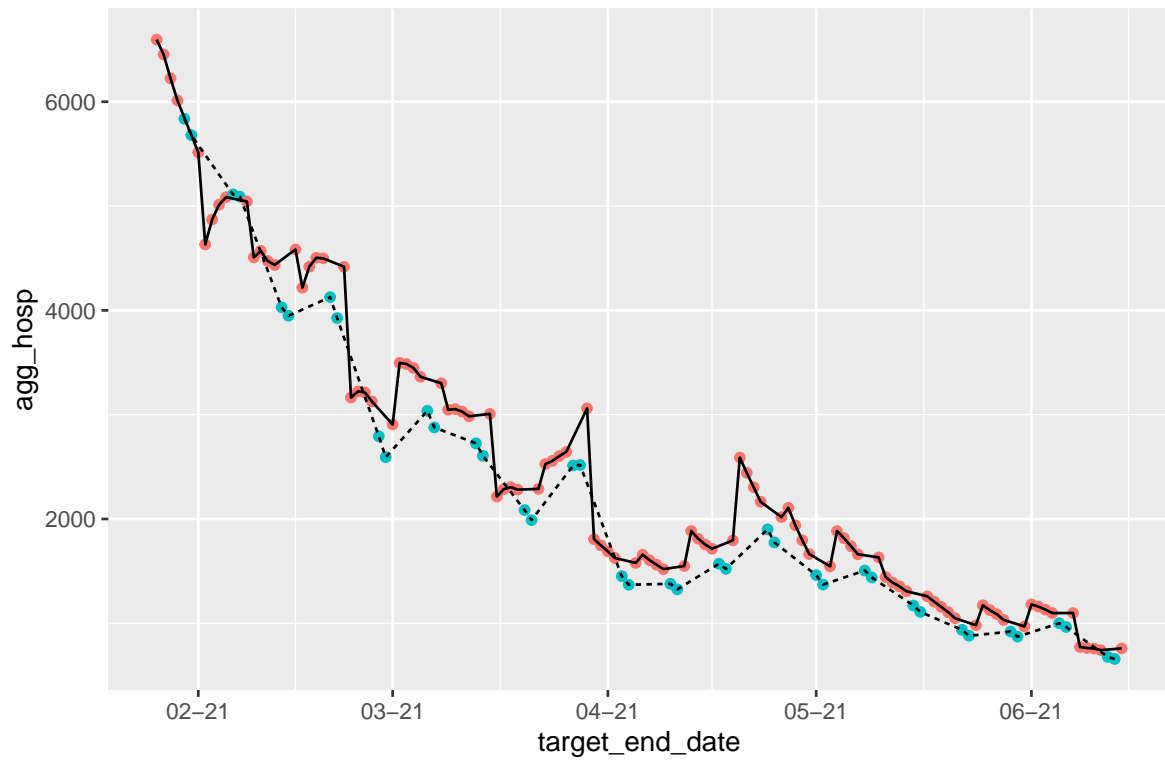
Truth Data



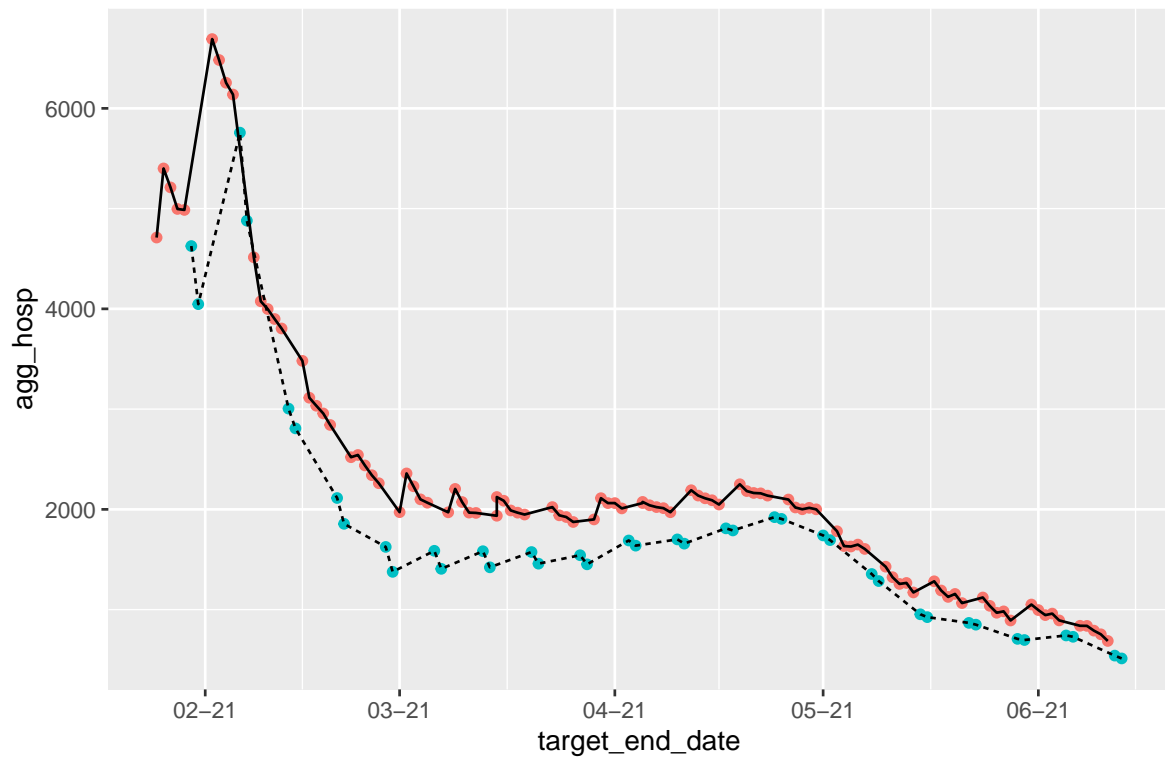
COVIDhub-ensemble Forecasts



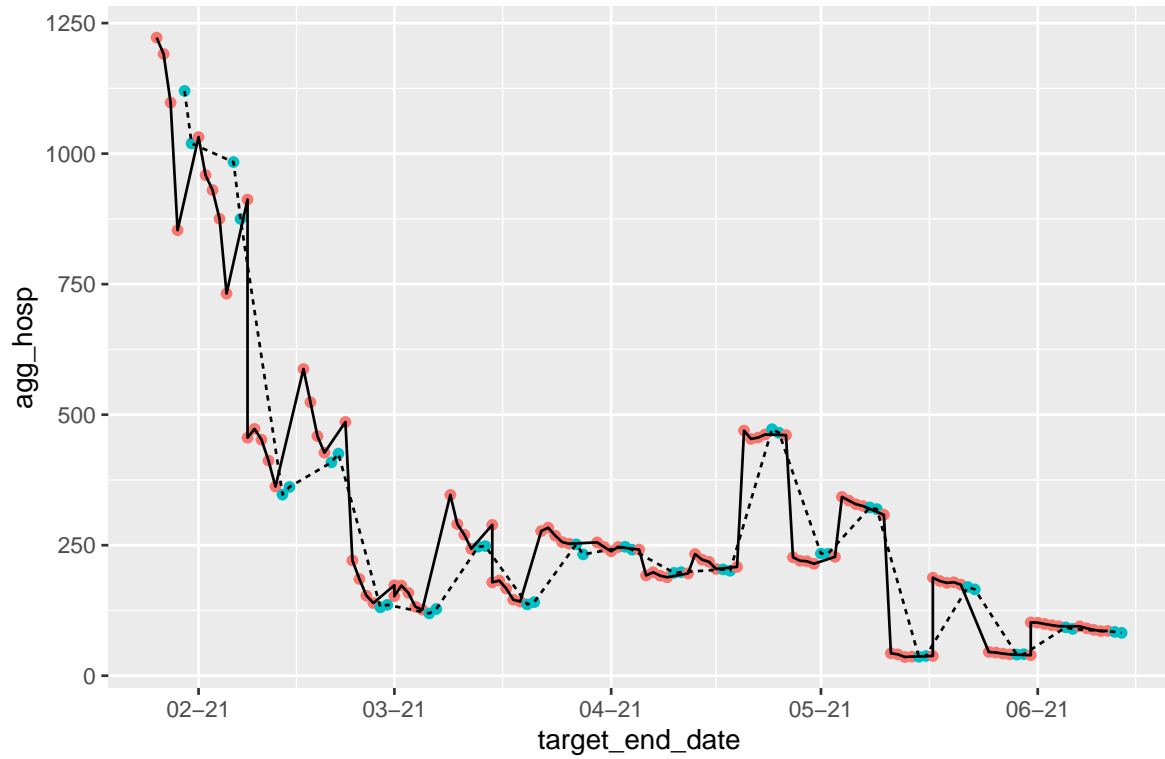
JHUAPL-Bucky Forecasts



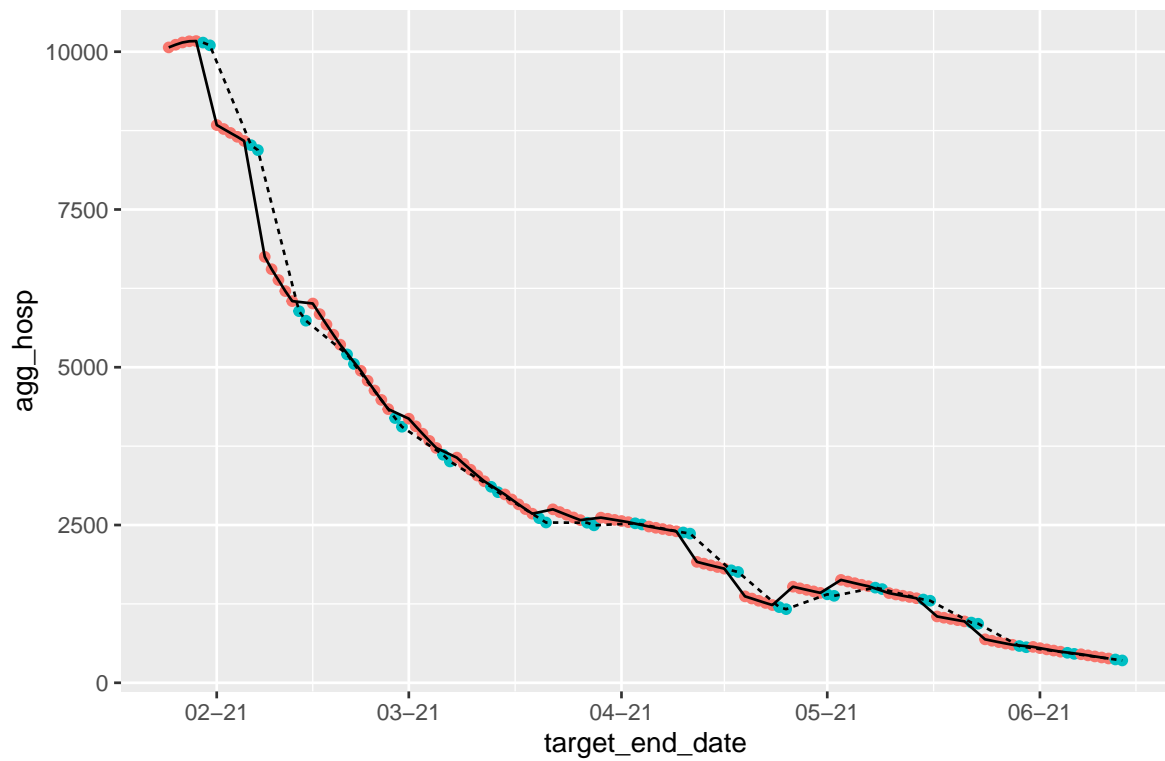
JHUAPL-Gecko Forecasts



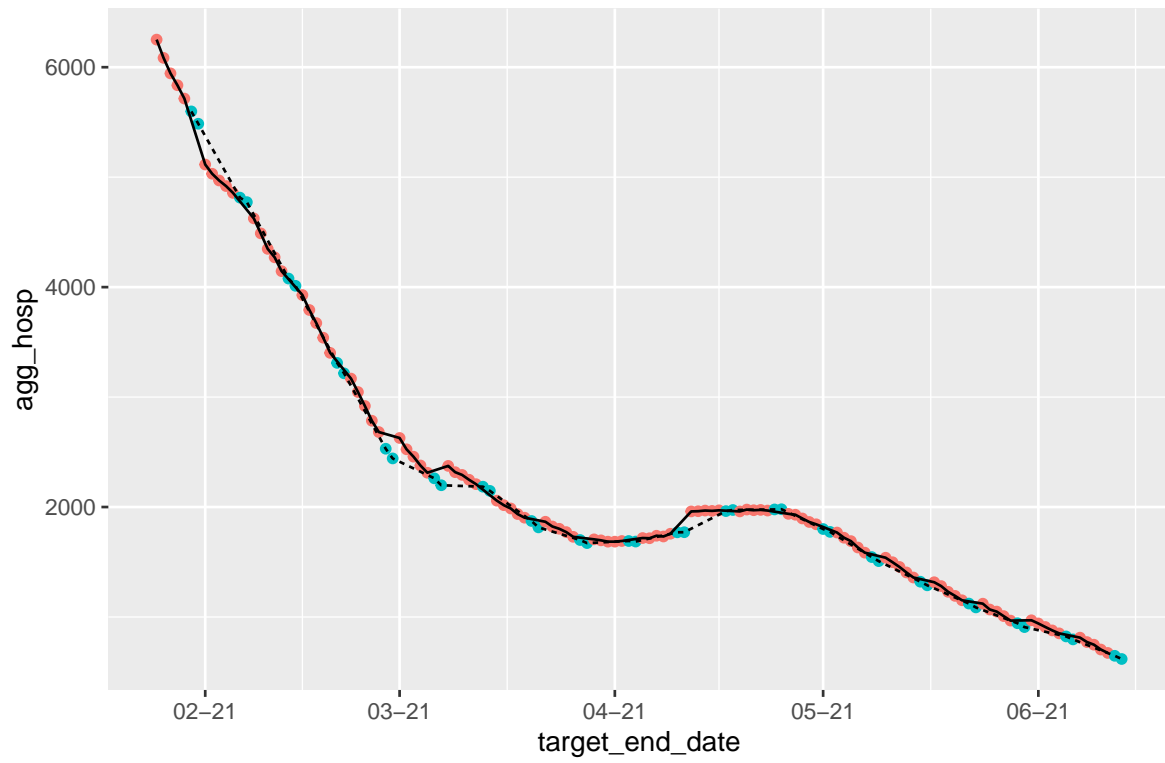
Google_Harvard-CPF Forecasts



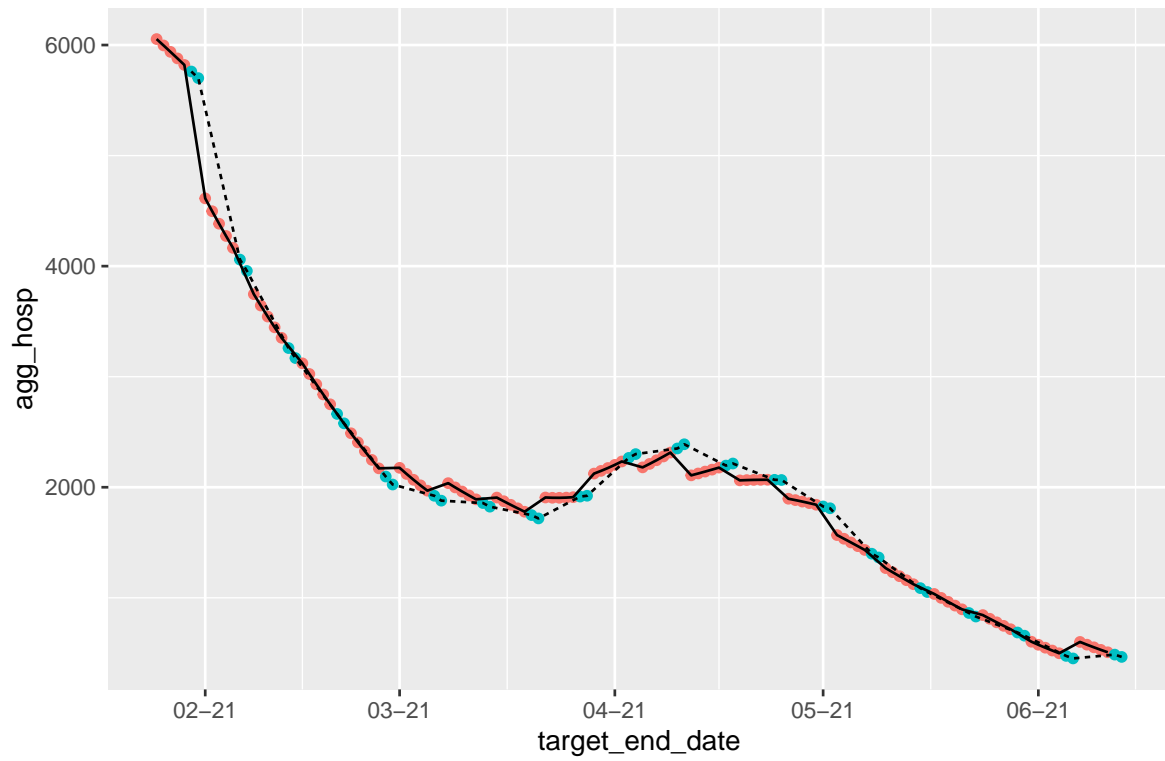
Covid19Sim-Simulator Forecasts



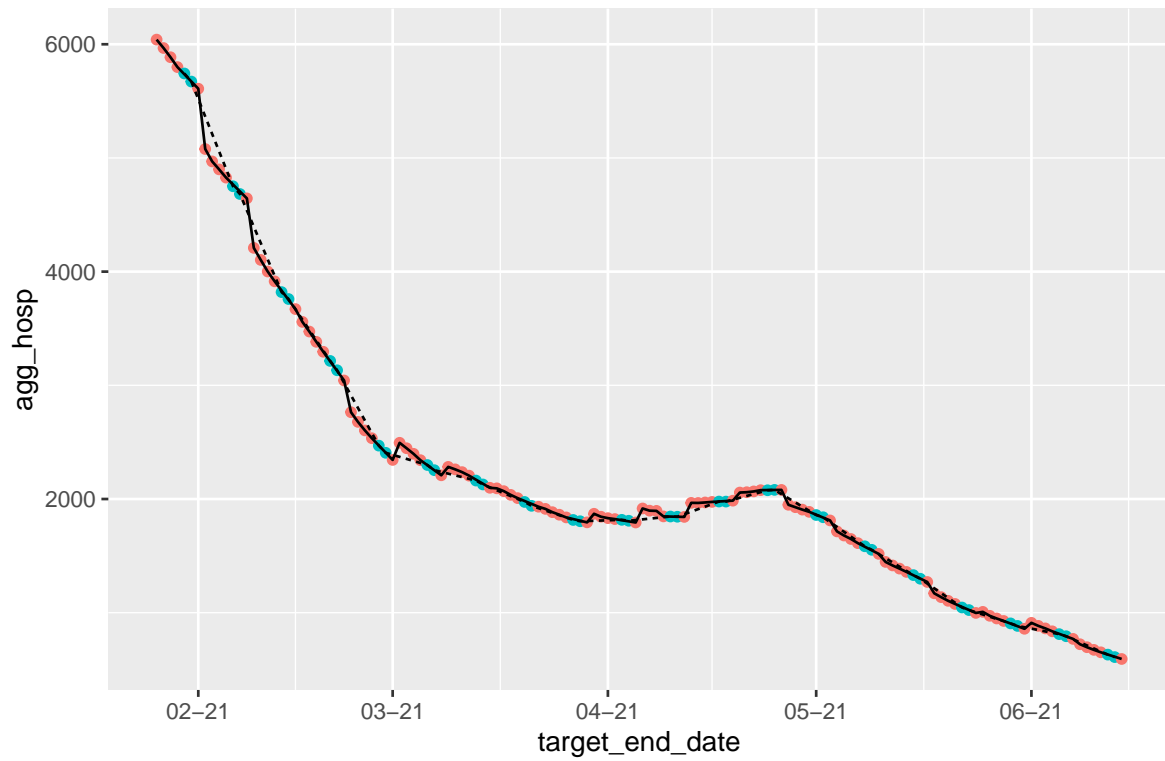
CU-select Forecasts



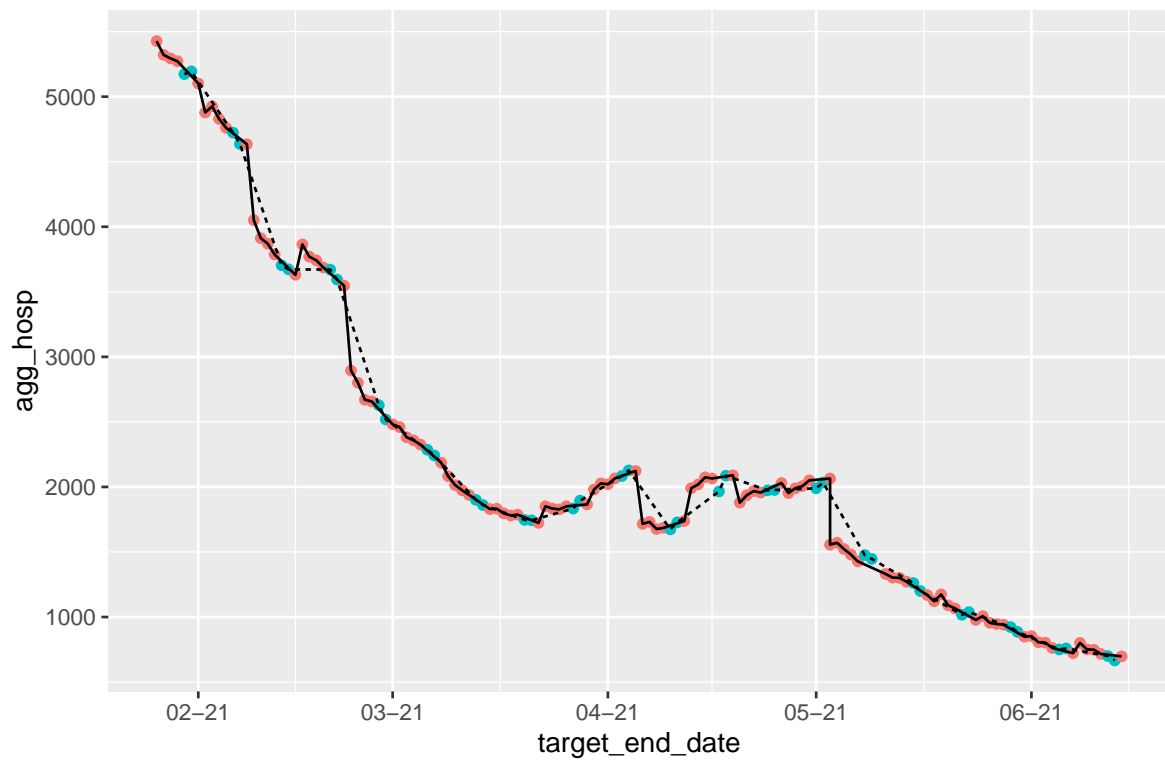
Karlen-pypm Forecasts



JHUAPL-SLPHospEns Forecasts

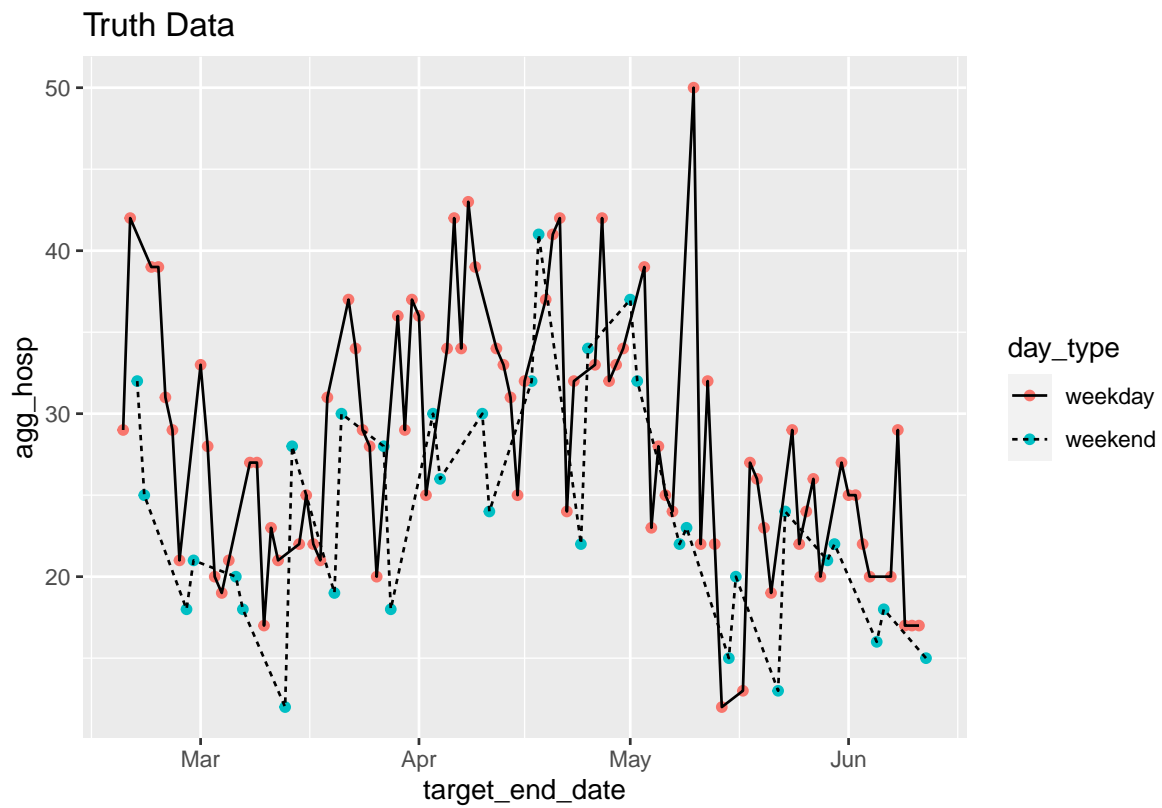


MOBS-GLEAM_COVID Forecasts

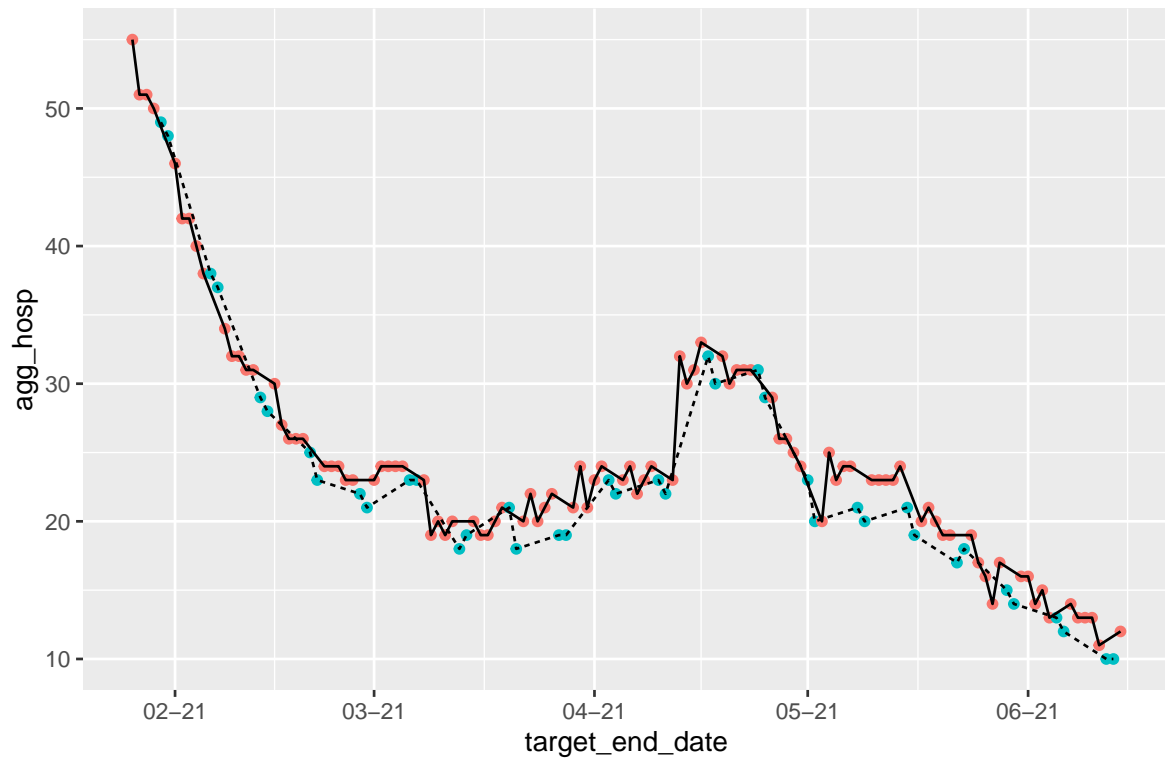


Truth data has always shown day of the week effects (weekends always have lower inc hosp) but the models

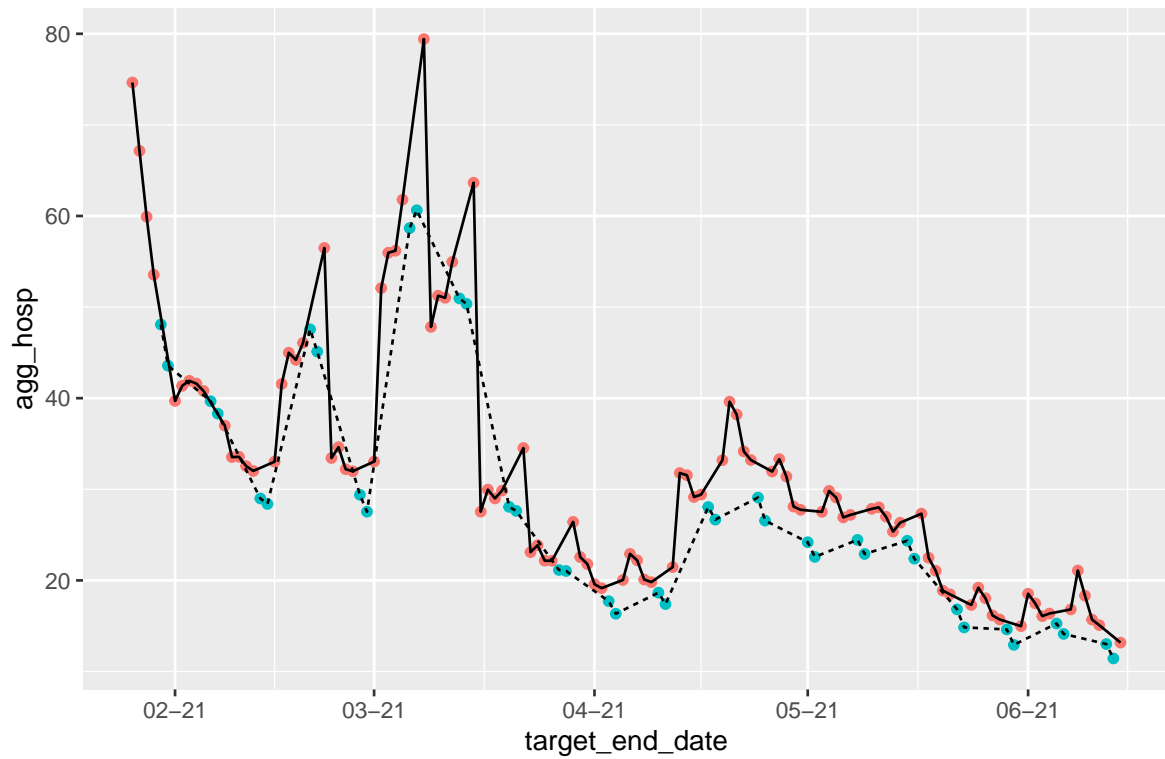
seemed to predict lower inc hosp on Tuesdays for the end of December/beginning of January, then most seem to switch to lower weekend end inc hosp sometime during 2021.



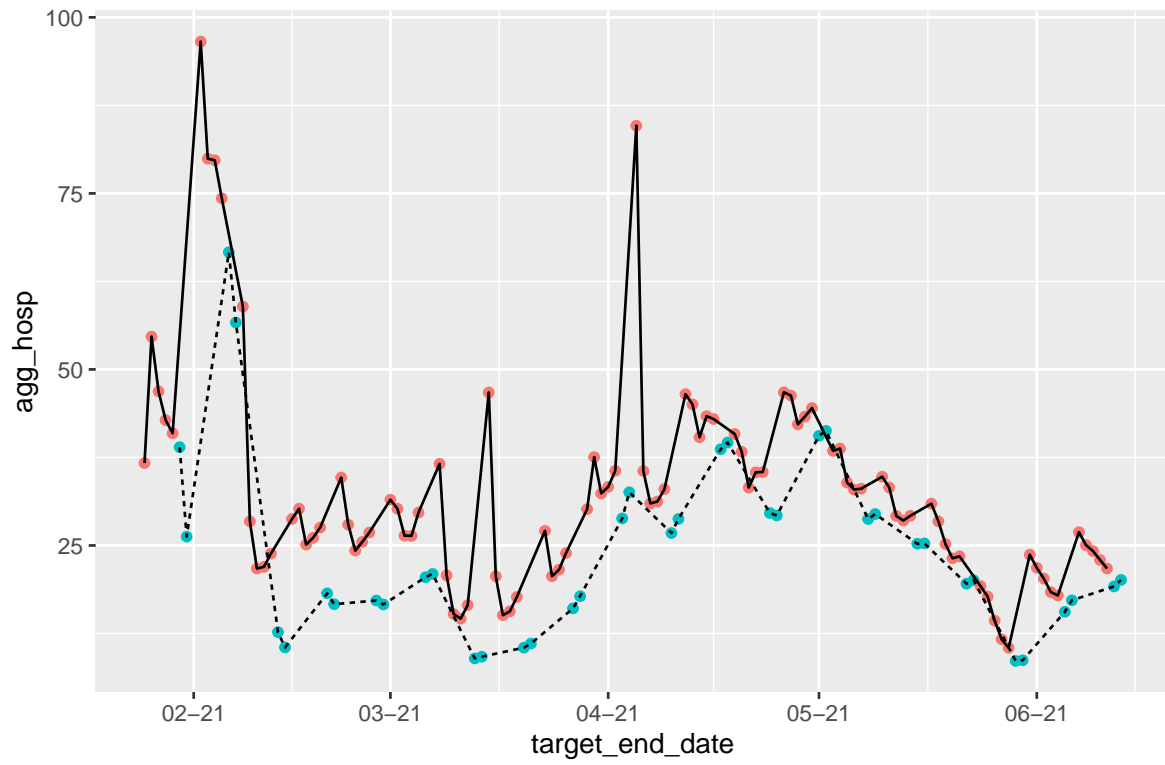
COVIDhub-ensemble Forecasts



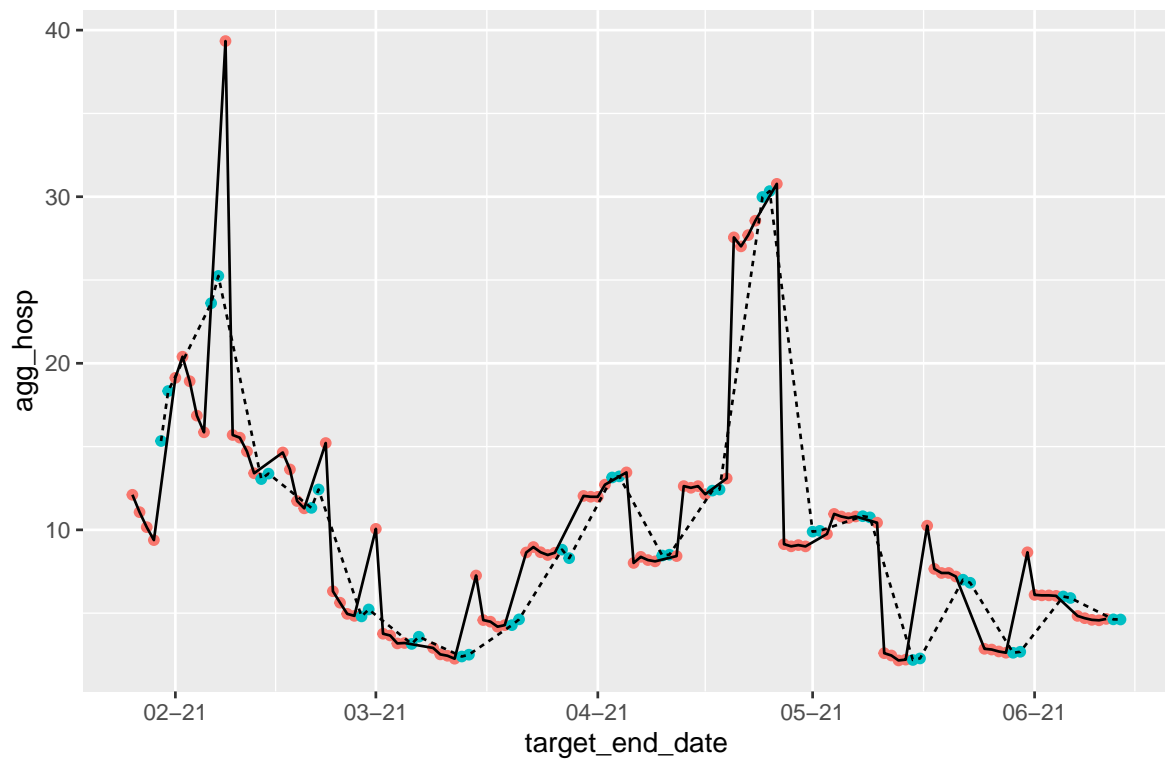
JHUAPL-Bucky Forecasts



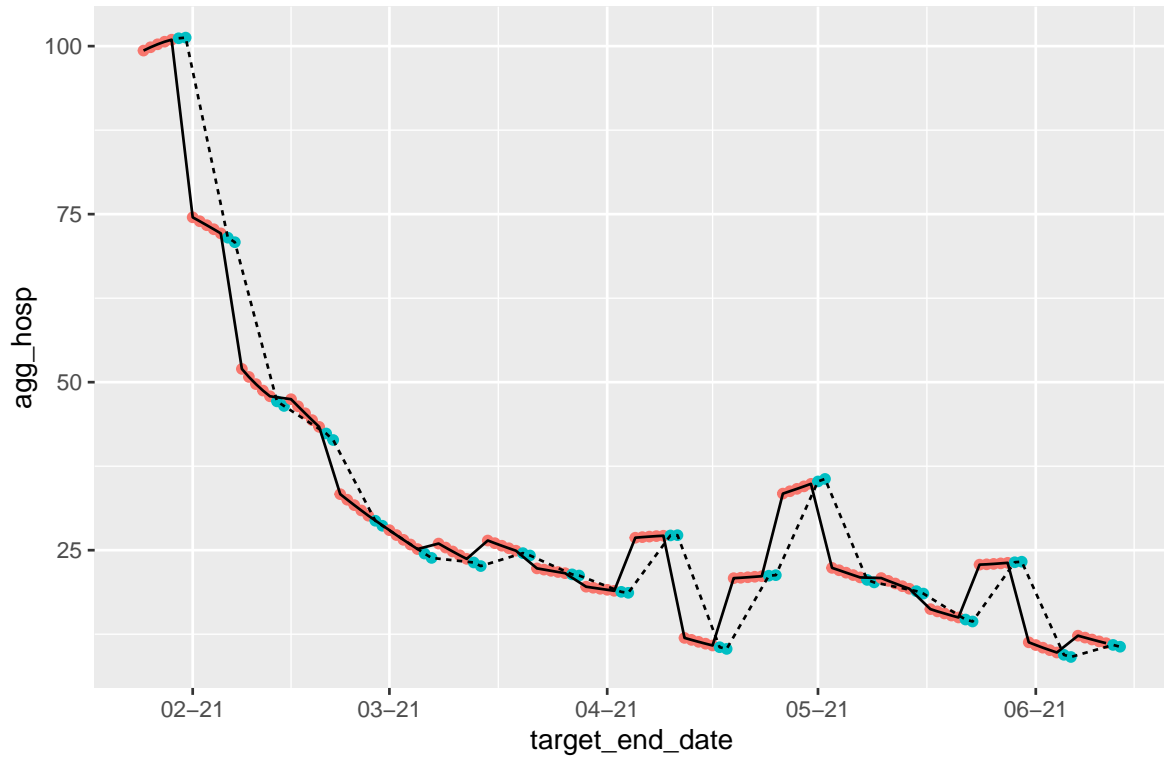
JHUAPL-Gecko Forecasts



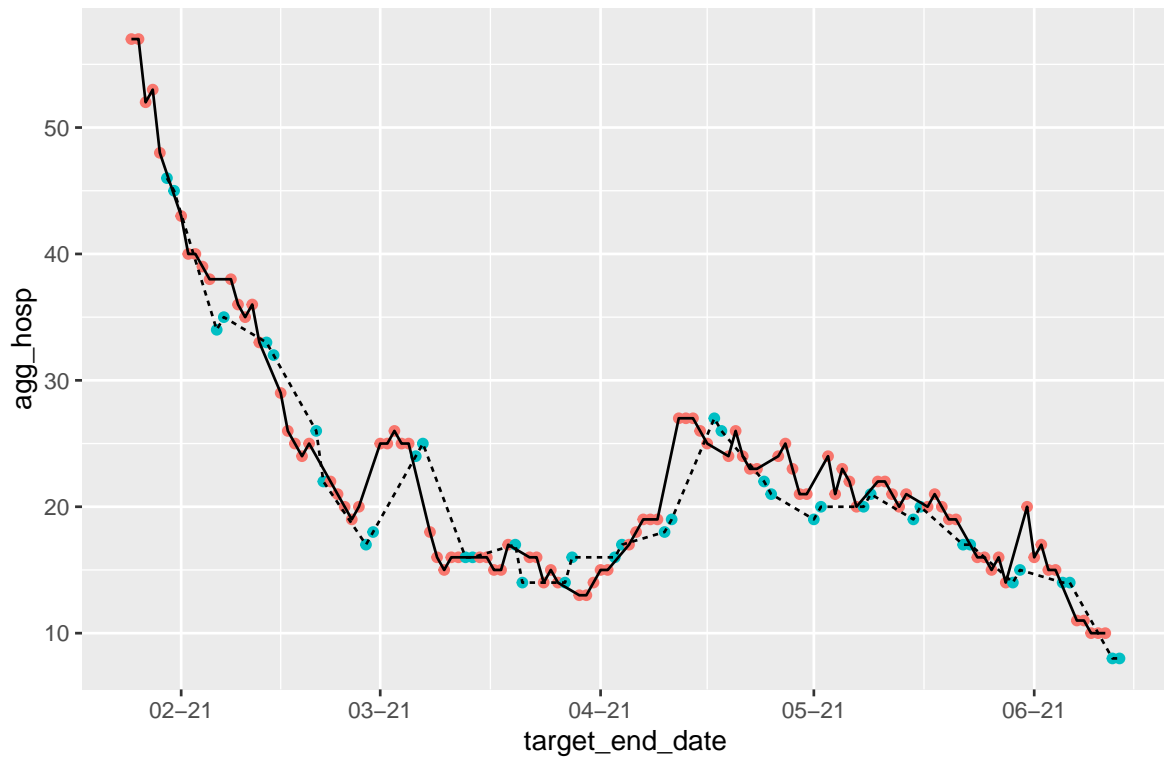
Google_Harvard-CPF Forecasts



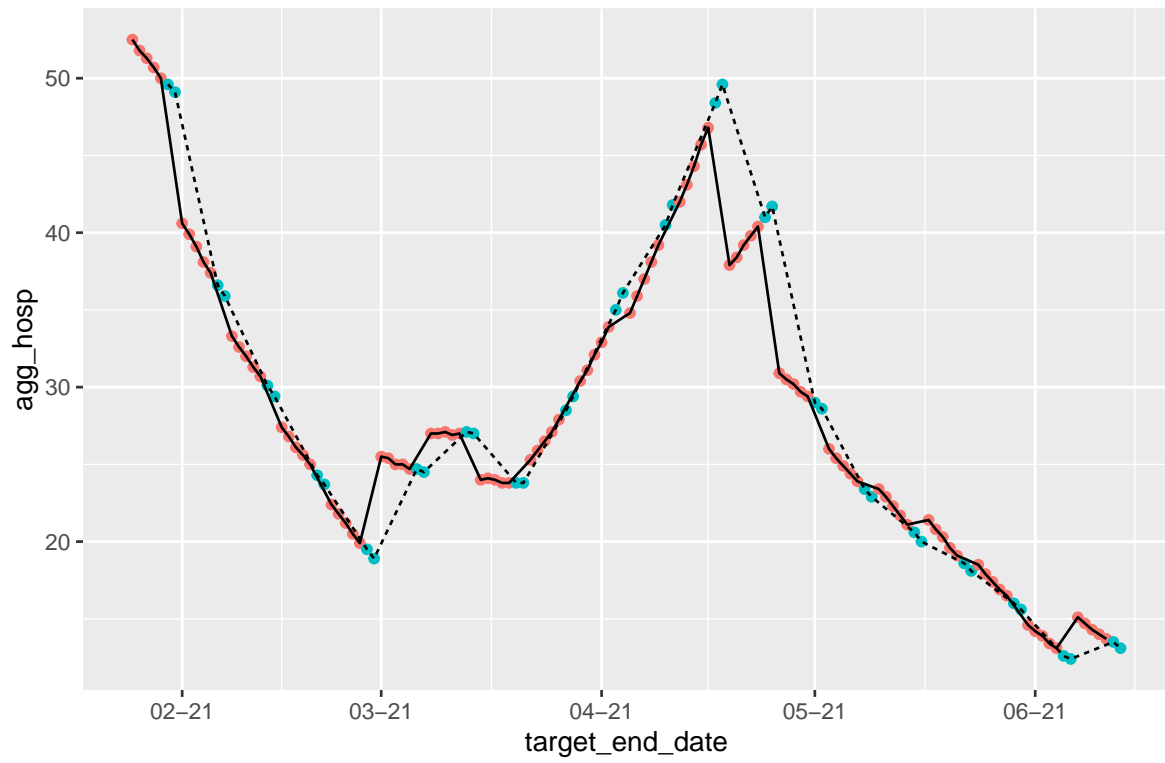
Covid19Sim-Simulator Forecasts



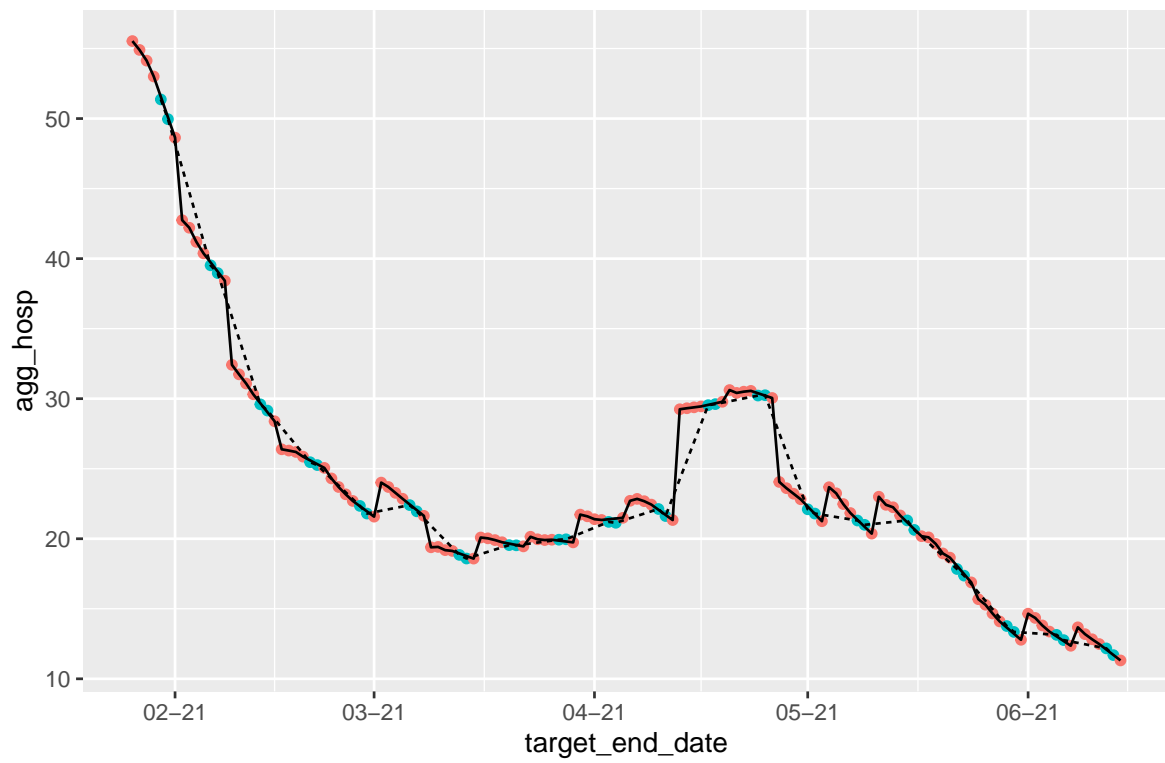
CU-select Forecasts



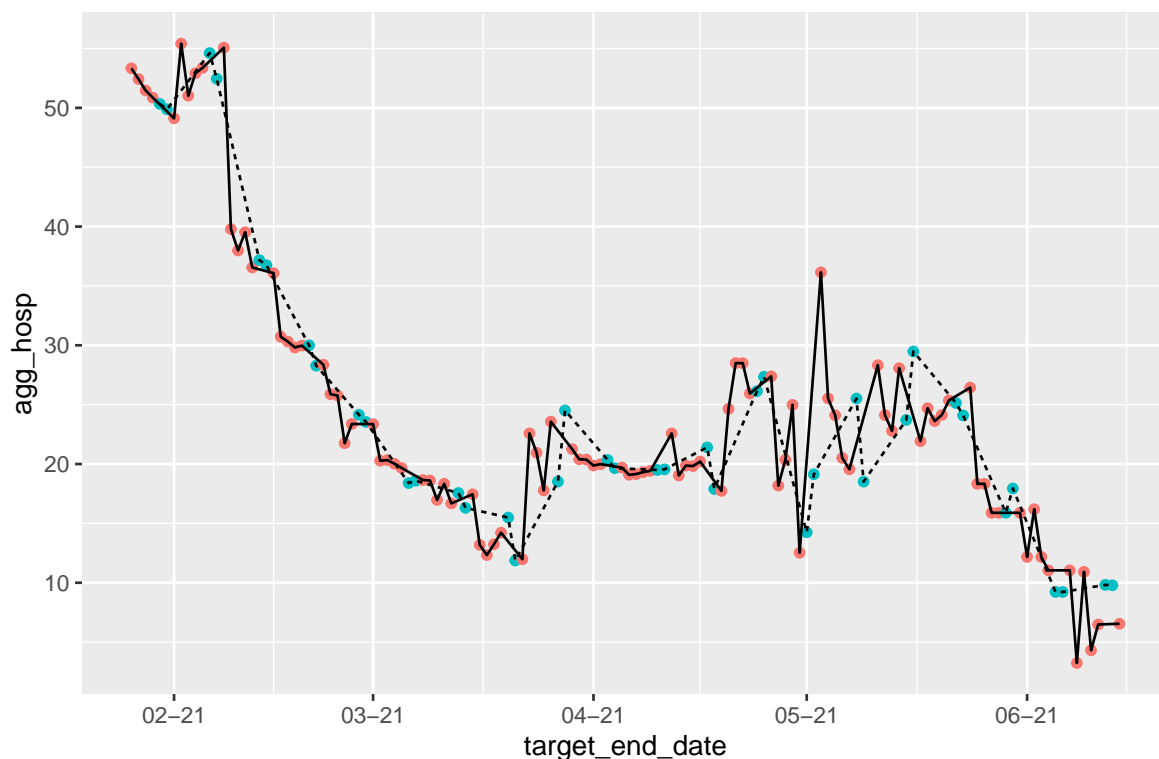
Karlen-pypm Forecasts



JHUAPL-SLPHospEns Forecasts



MOBS-GLEAM_COVID Forecasts



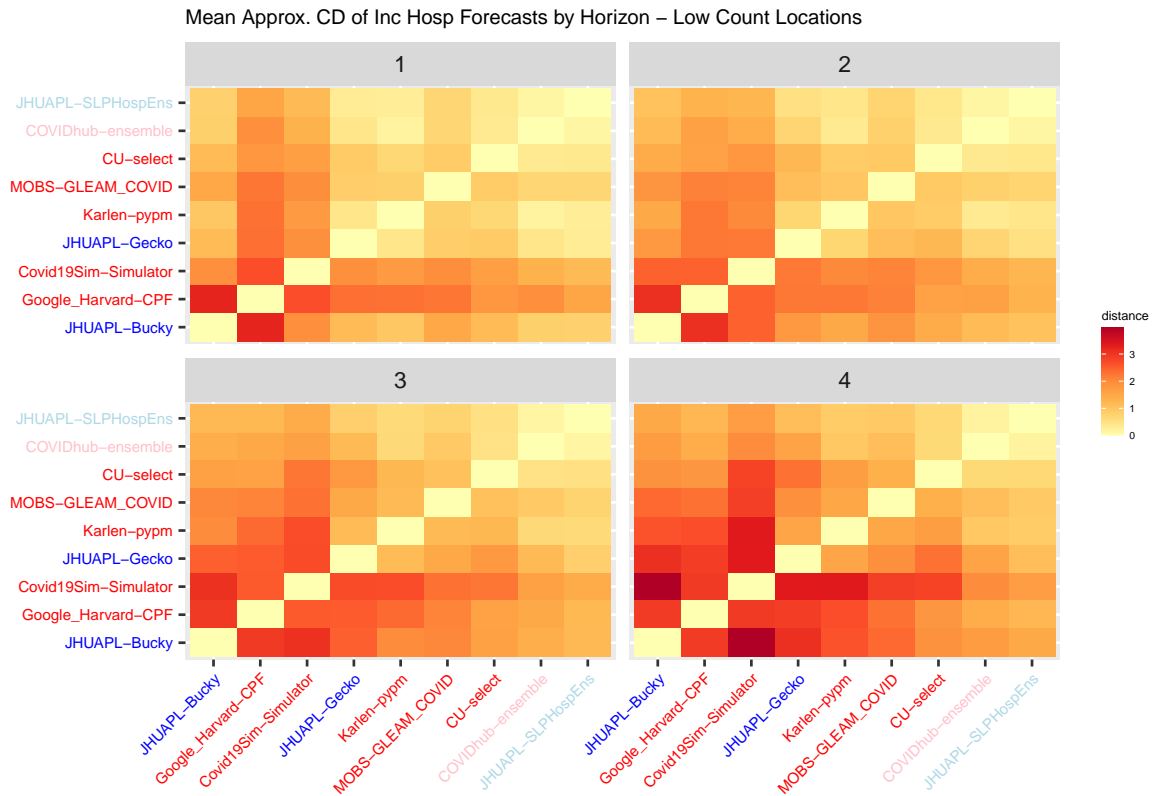
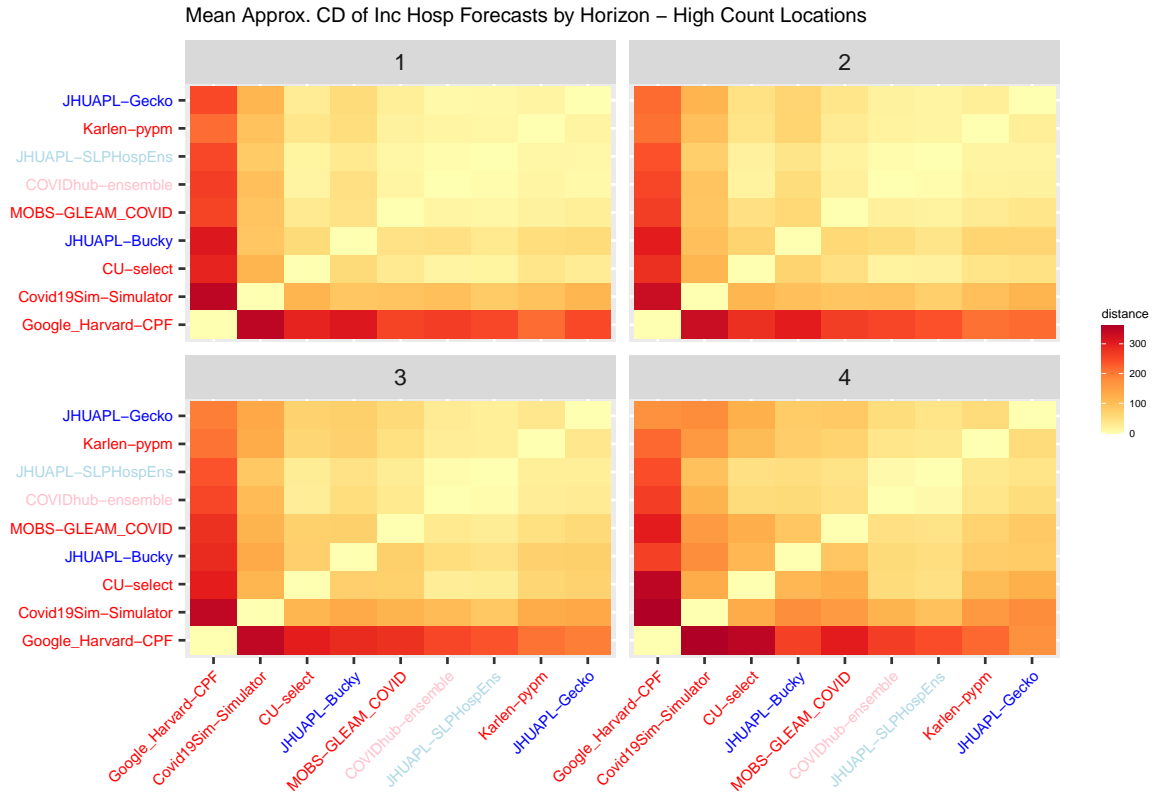
The low hosp location truth data doesn't present such a clear pattern as to day of the week effects that suggest that weekends or weekdays generally have less inc hosp values. However, most of the models seem to switch to incorporating a day of the week effect at about the same time as that for the high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

Model	Day of Week Effect
Covid19Sim-Simulator	FALSE
COVIDhub-ensemble	ENS incl
CU-select	FALSE
Google_Harvard-CPF	FALSE
JHUAPL-Bucky	TRUE
JHUAPL-Gecko	TRUE
JHUAPL-SLPHospEns	ENS excl
Karlen-pypm	FALSE
MOBS-GLEAM_COVID	FALSE

Weekday Analysis

This 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.

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.



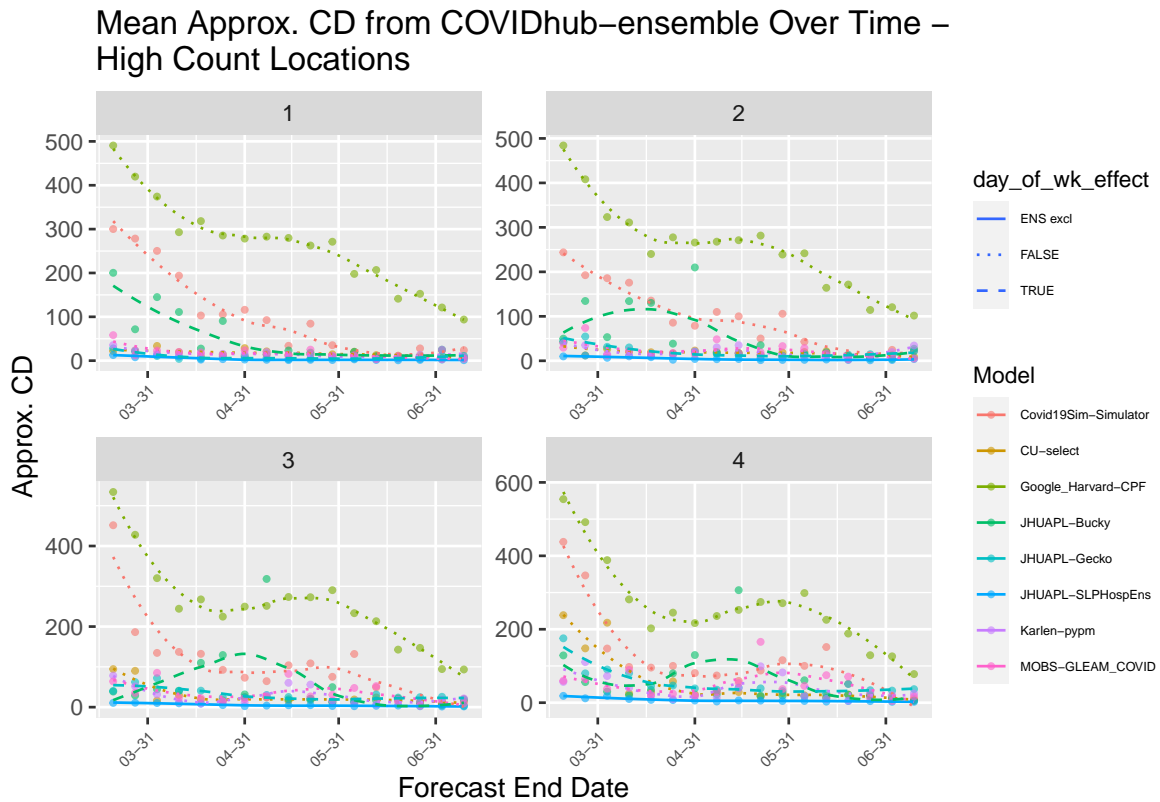
Google_Harvard-CPF is generally the least similar to the other models for both the high count and low

count locations. This is true across all horizons for the high count locations, but only true for the one and two week horizons of the low count locations. Covid19Sim-Simulator, JHUAPL-Bucky, and JHUAPL-Gecko show similar Cramer's Distances at three and four week horizons at the low count locations. However, it is important to note the small scale observed for the low count locations may explain why three models have such similar Cramer's Distances.

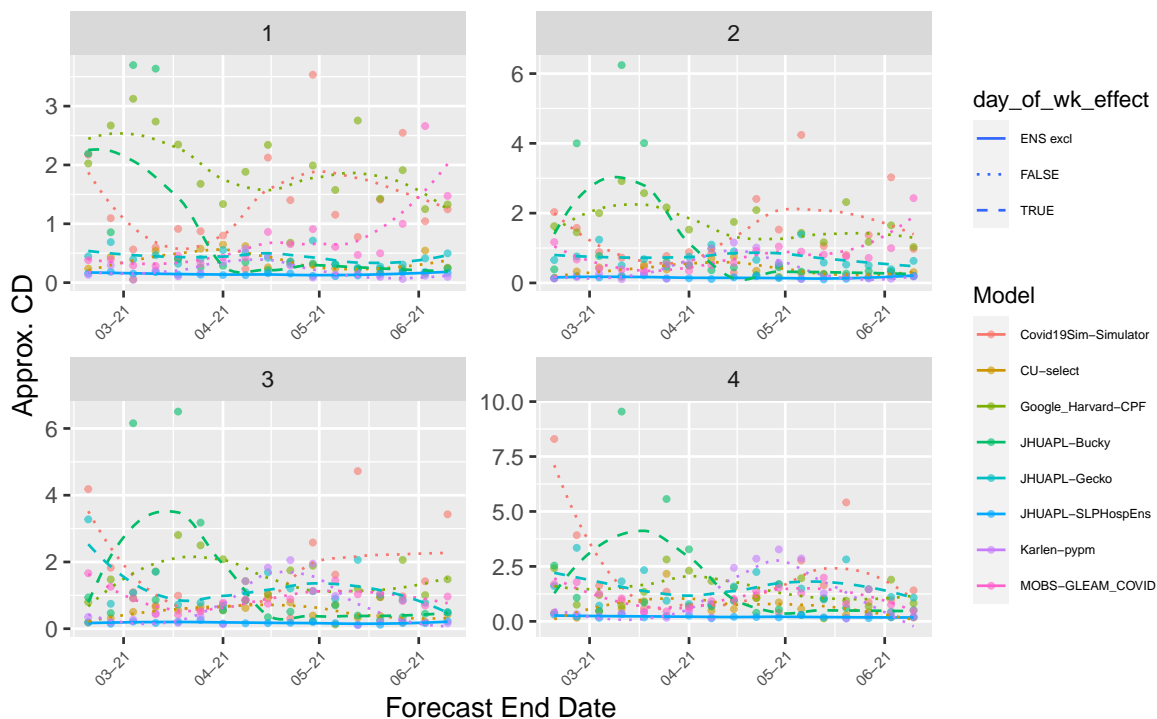
Covid19Sim-Simulator is the second least similar model for the high count locations, but its Cramer's Distance is much smaller than that of Google_Harvard-CPF. Of note is similar differences between models at all horizons for high count locations, unlike the results shown for inc cases and inc deaths which show substantial differences between models as horizon length increases.

For low count locations, models with day of the week effect tend to have higher Cramer's Distances, but there doesn't seem to be much of a pattern for high count locations.

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

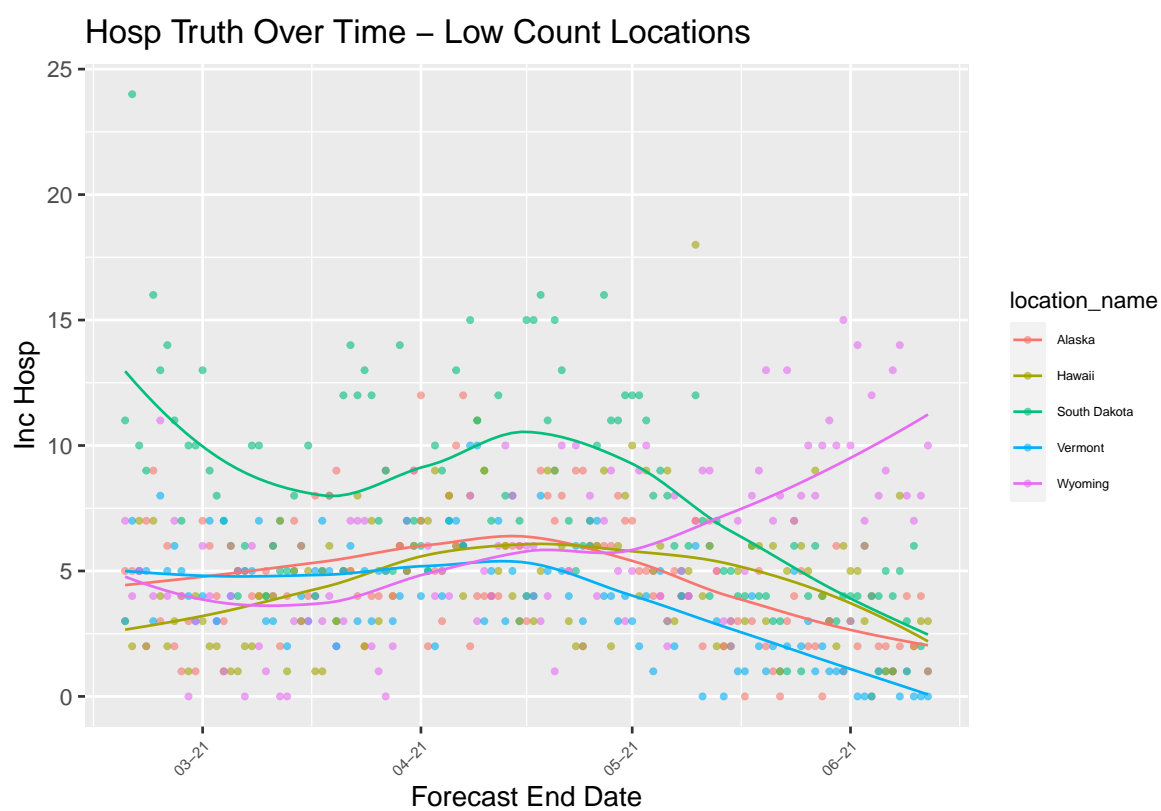
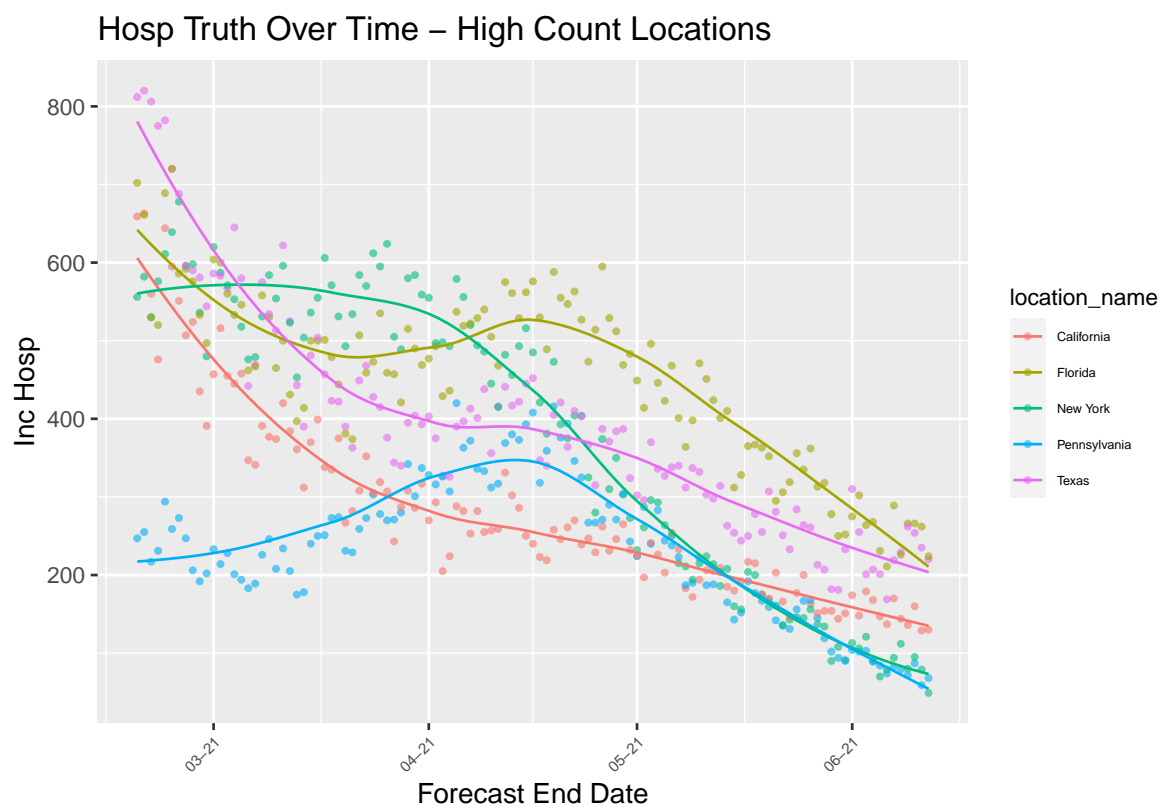


Mean Approx. CD from COVIDhub-ensemble Over Time – Low Count Locations



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

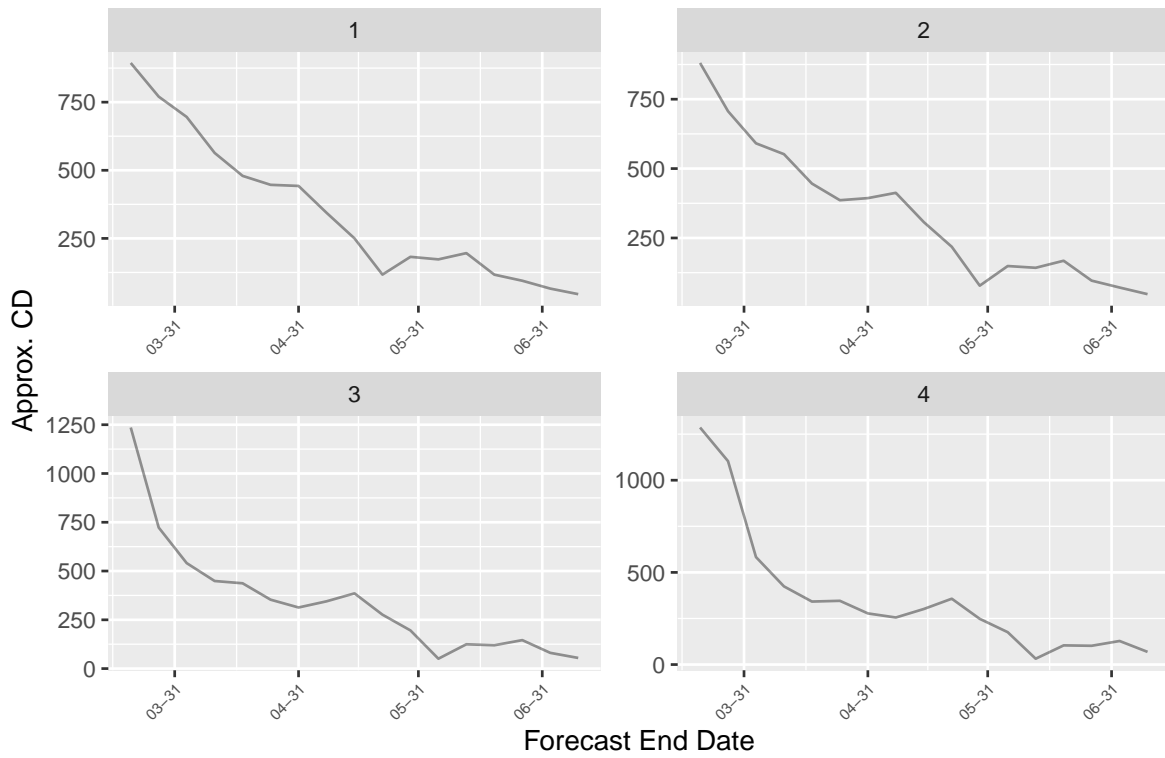
Whether models incorporate a day of the week effect does not seem to have an impact on how much the model differs from the ensemble, nor as to when it differs greatly.



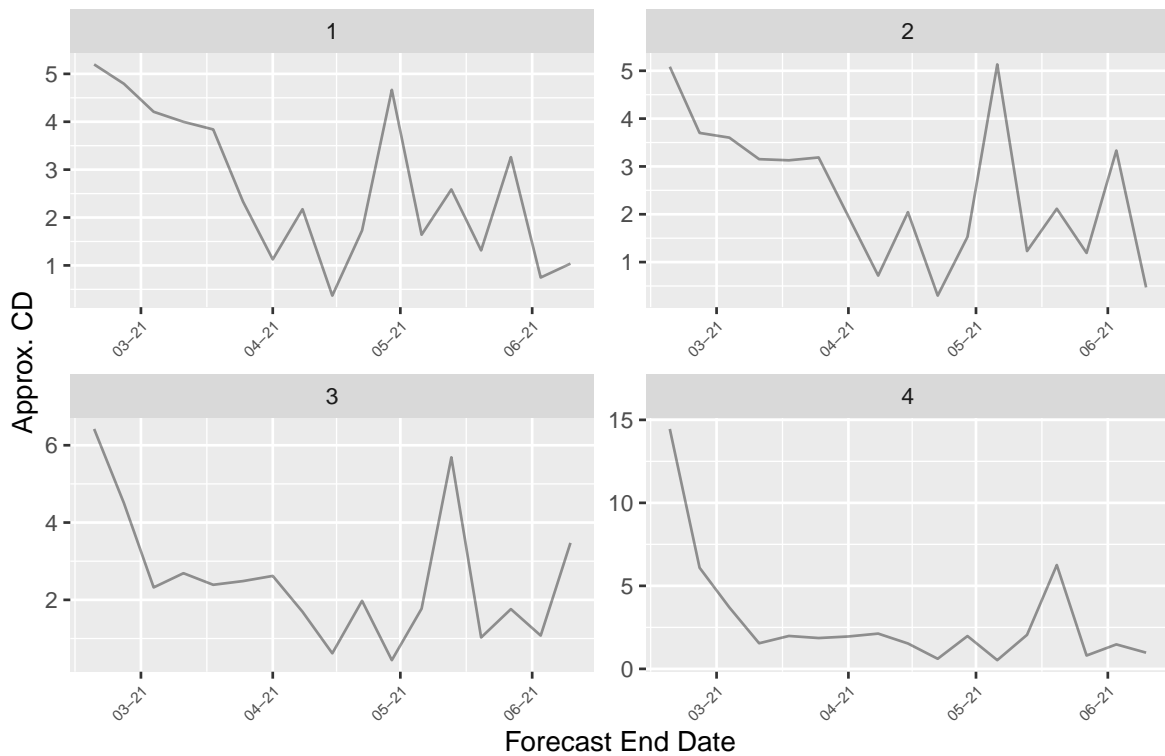
It seems that Google_Harvard-CPF and Covid19Sim-Simulator's differences from the ensemble model follow

the trends shown by the truth data.

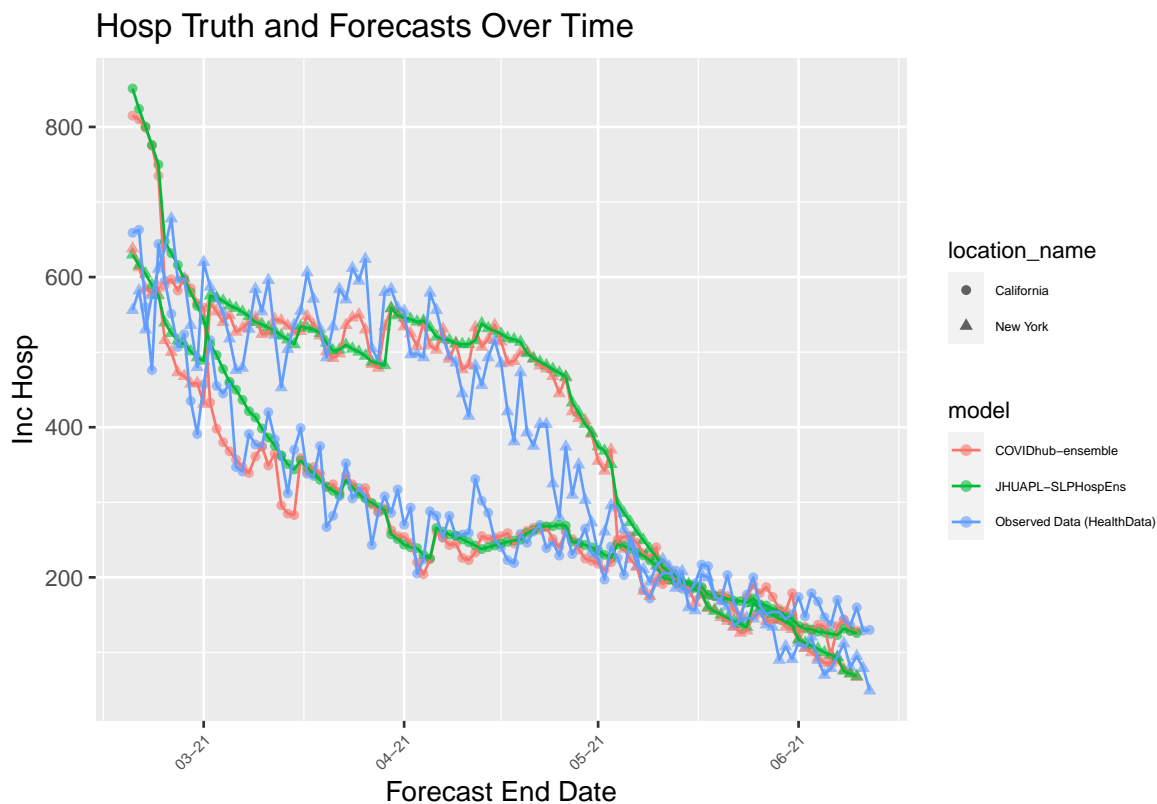
Approx CD over Time – Covid19Sim–Simulator and Google_Harvard–CPF



Approx CD over Time – Covid19Sim–Simulator and Google_Harvard–CPF



These plots indicate that the difference between the models at the high count locations seems to shrink over time. Since the low count locations approx CD is so low, it is difficult to draw conclusions from the plots.



We can also cluster the distances using hierarchical clustering.

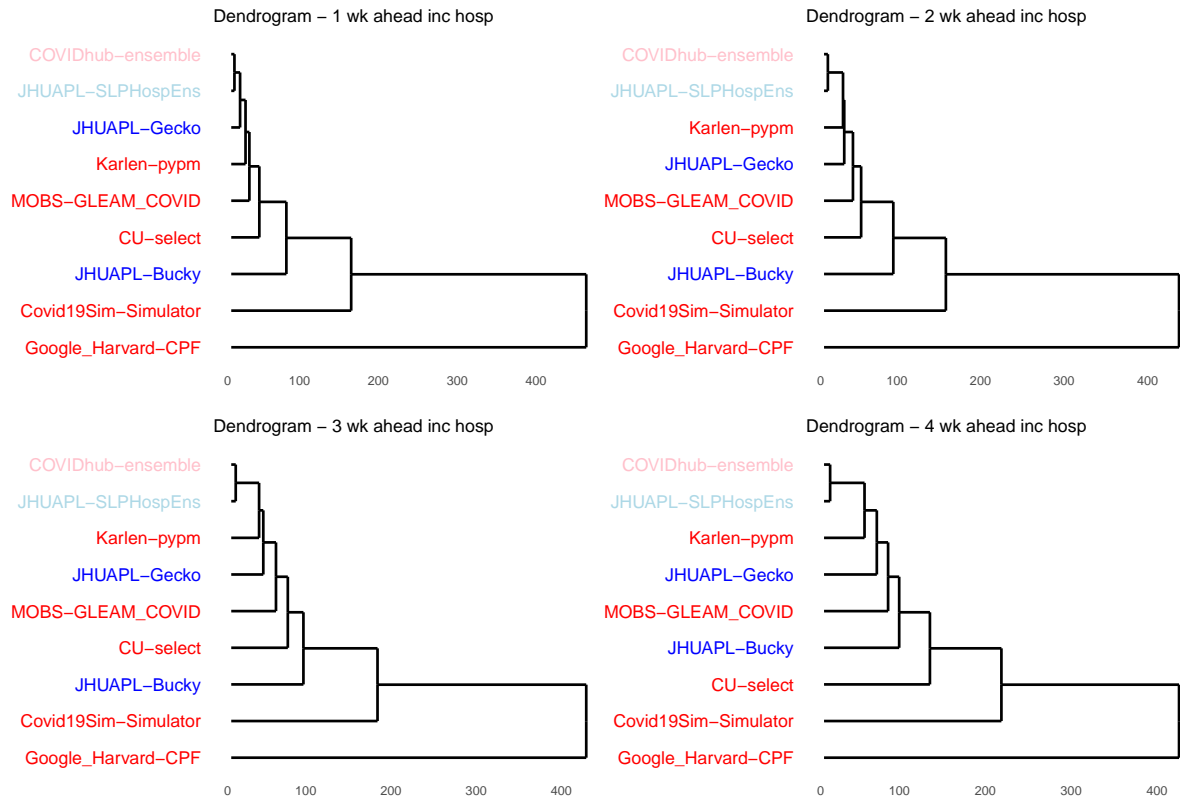


Figure 1: High Hospitalization Count Locations

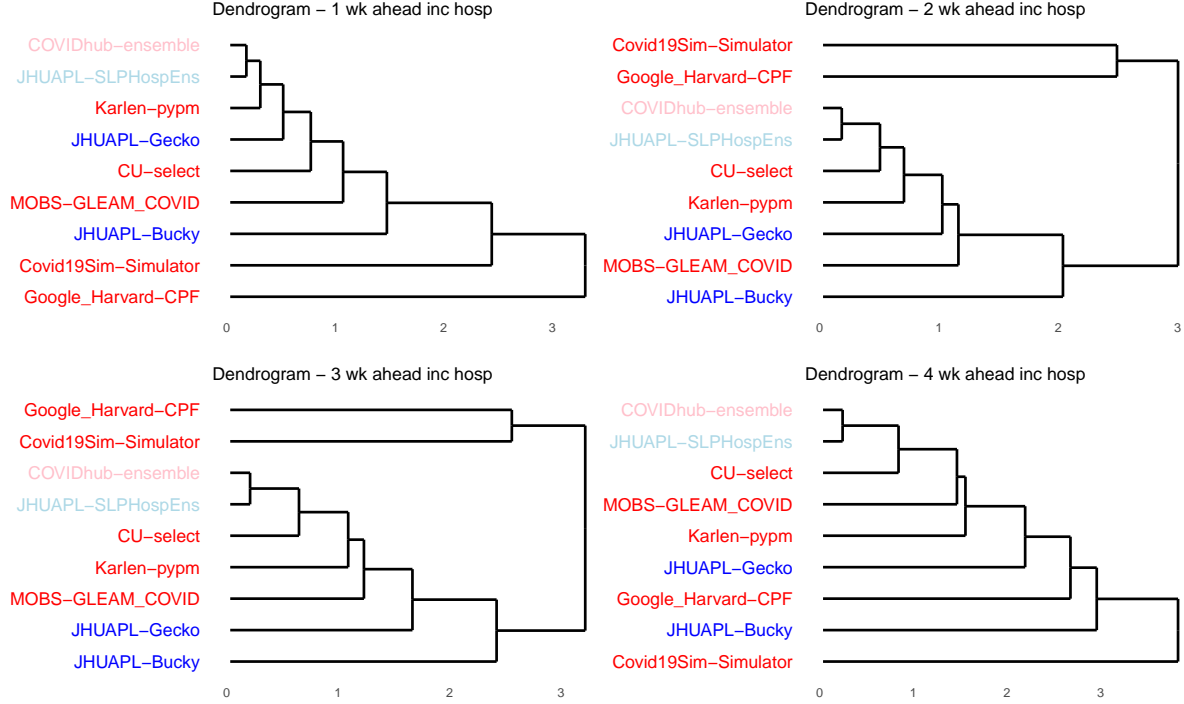


Figure 2: Low Hospitalization Count Locations

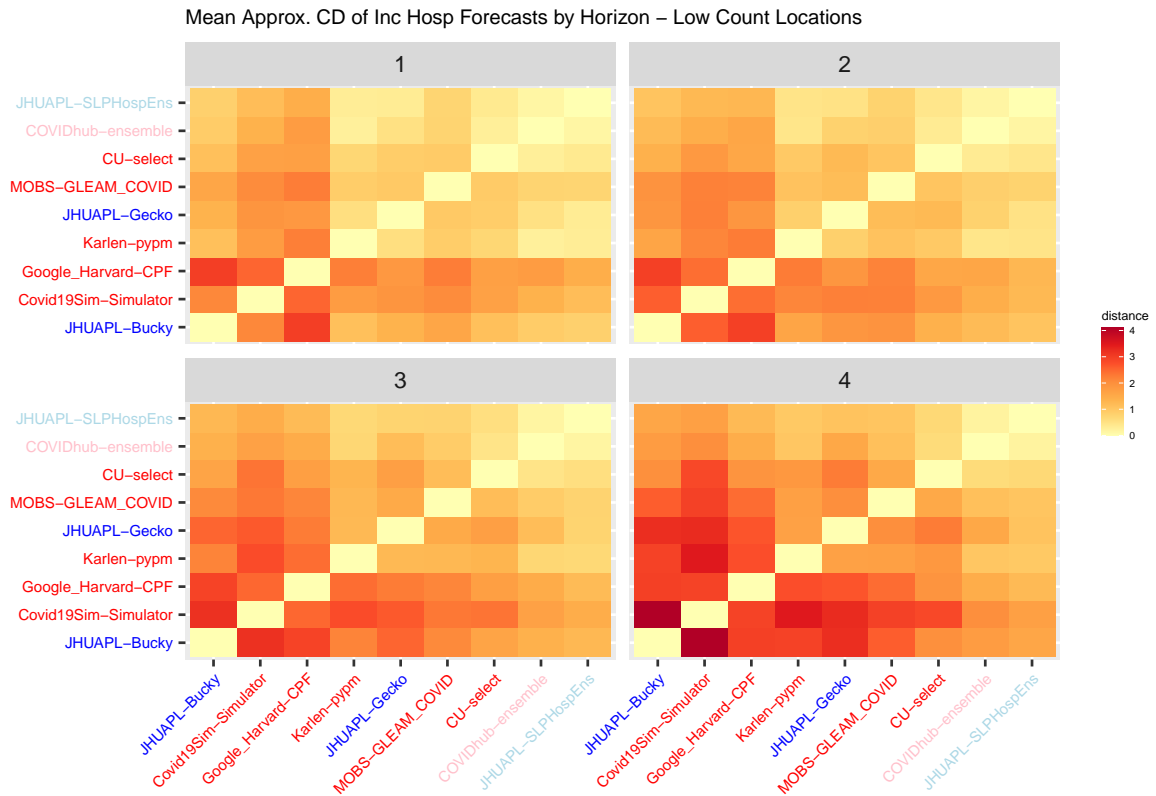
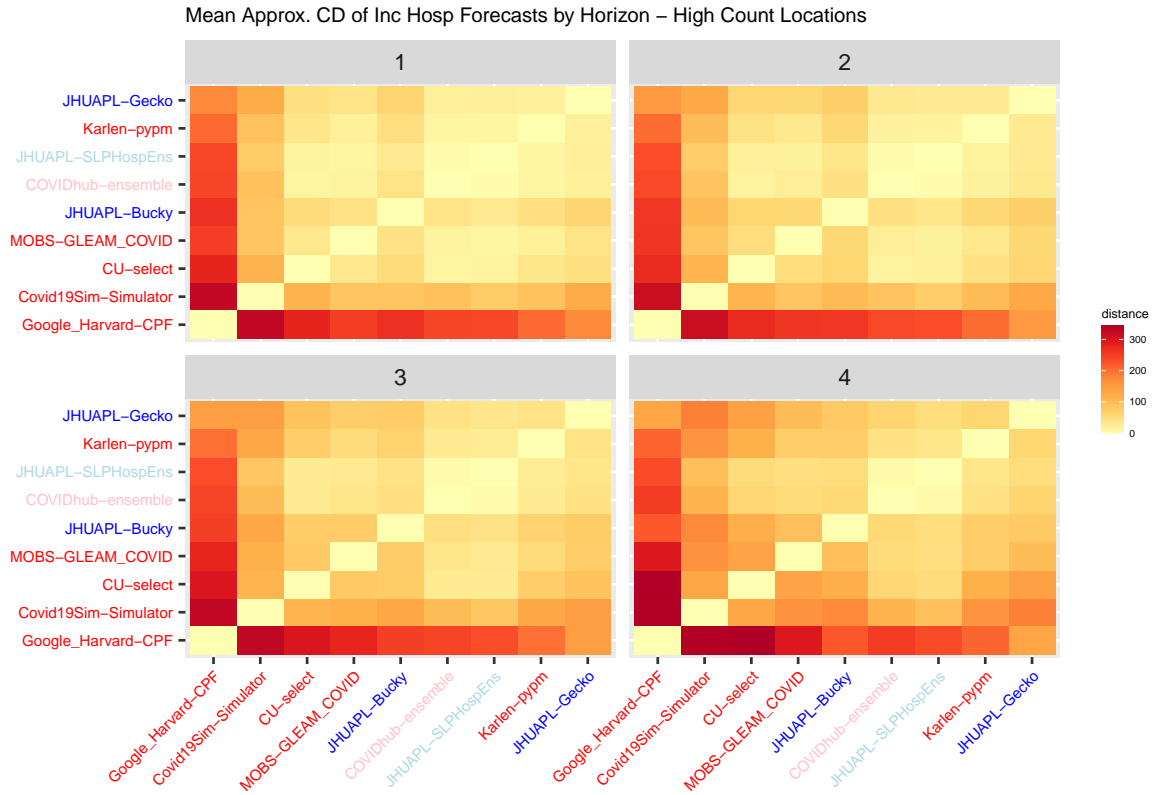
For both high count and low count locations, all of the dendrograms show that Google_Harvard-CPF is generally the most dissimilar to the other models, and substantially so for the high count locations, followed by Covid19Sim-Simulator. However, the scale for differences in Cramer Distance for the low count locations is very small, which may be a result of low hospitalizations, which may explain why Google_Harvard-CPF and Covid19Sim-Simulator show similar Cramer’s Distances for low count locations. We can also see that the ensembles are the most similar among models.

Overall, it seems that Google_Harvard-CPF is consistently the most dissimilar from the other models by a substantial amount, followed by Covid19Sim-Simulator, across almost all horizons for both high-count and low count regions. For Thursday forecasts, it seems models without day of the week effects tend to be more dissimilar from the COVIDhub-ensemble model than models without day of the week effects.

Weekend Analysis

Now we look at forecasts with a target end date on a Saturday to see if day of the week effects change model similarity. We expect forecasts for weekends to show the impact of day of the week effects more strongly than weekdays.

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.



Similarly to the Thursday forecasts, Google_Harvard-CPF is generally the least similar to the other models

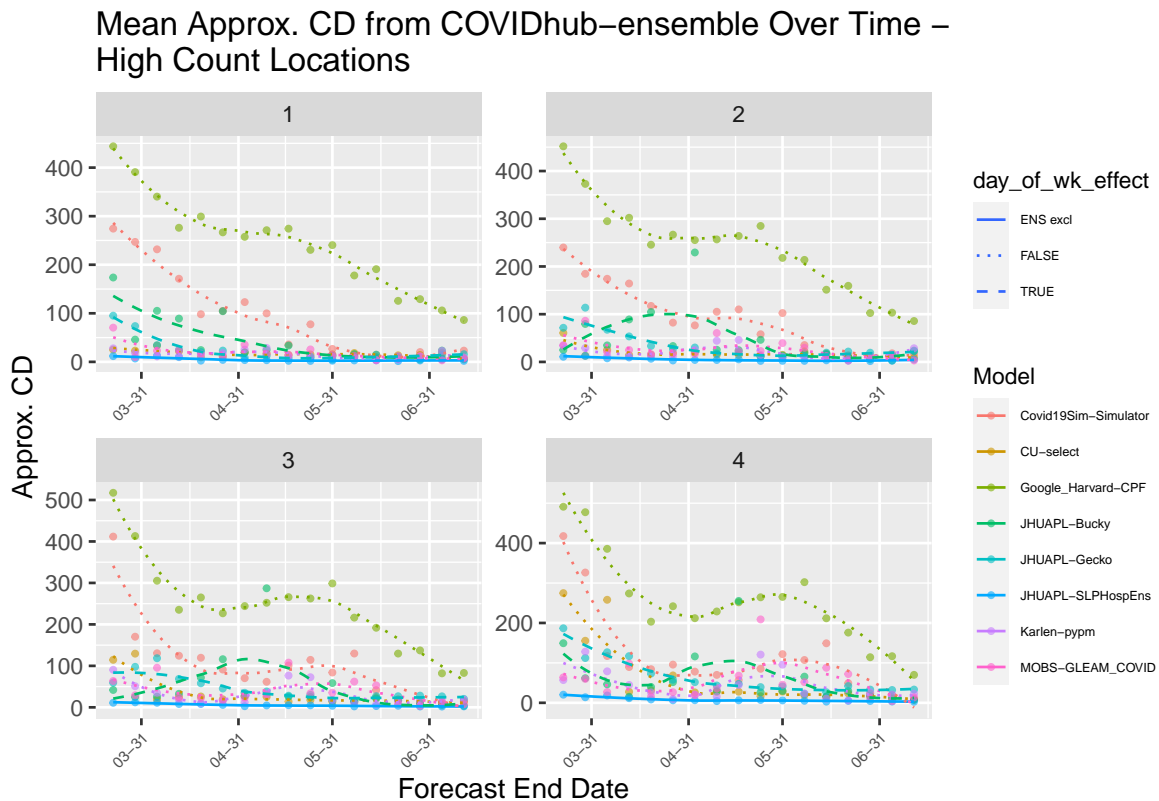
for both the high count and low count locations. This is true across all horizons for the high count locations, but only true for the one and two week horizons of the low count locations. Covid19Sim-Simulator is the most dissimilar for the three and four week horizons of the low count locations. However, it is important to note the small scale observed for the low count locations may explain why this model has a higher Cramer's Distance at longer horizons rather than a true pattern.

Covid19Sim-Simulator is the second least similar model for the high count locations, but its Cramer's Distance is much smaller than that of Google_Harvard-CPF. Of note is similar differences between models at all horizons for high count locations, unlike the results shown for inc cases and inc deaths which show substantial differences between models as horizon length increases.

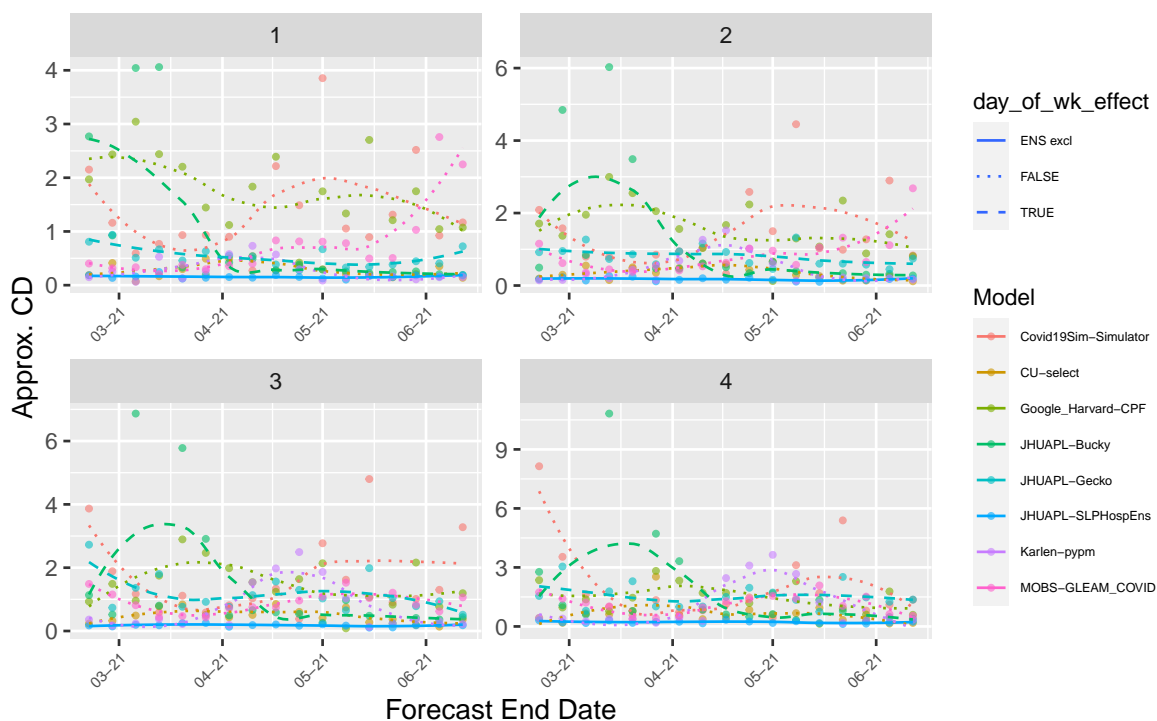
For high count locations, models without day of the week effect tend to have higher Cramer's Distances, but it is unclear if there is a pattern for low count locations.

The Saturday forecasts show similar but slightly different heat maps to the Thursday forecasts.

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



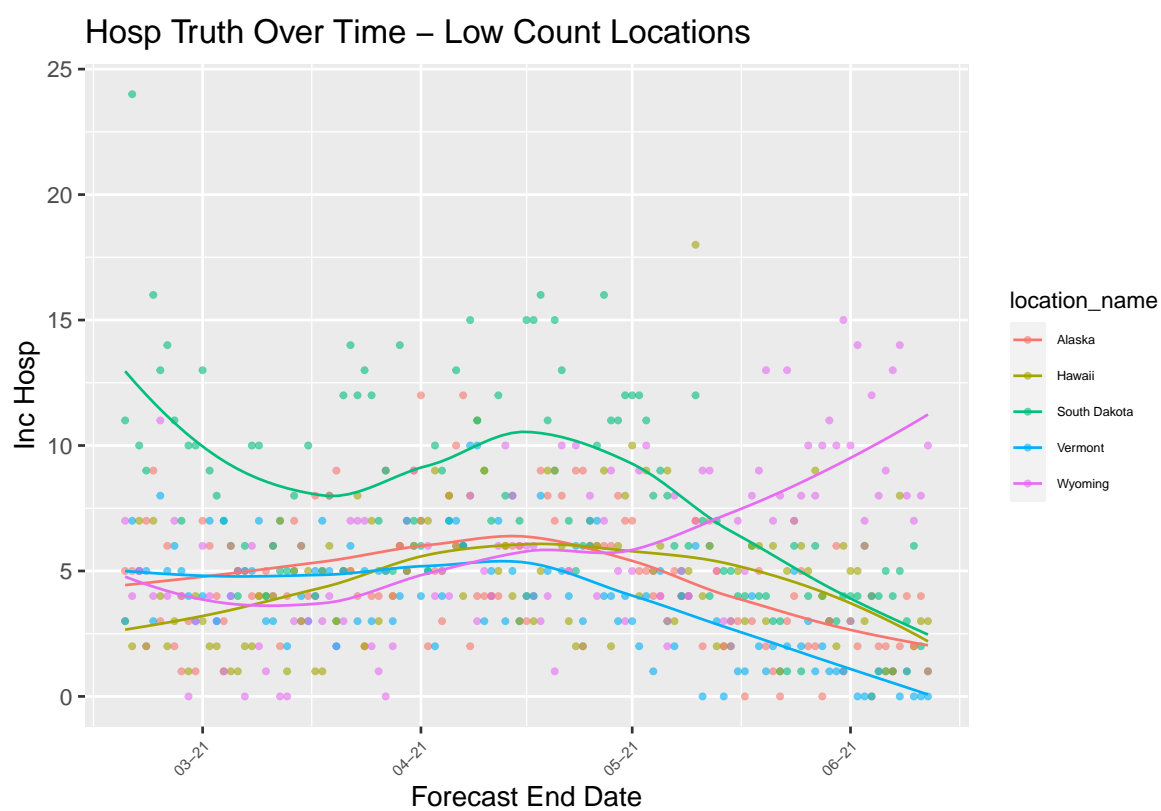
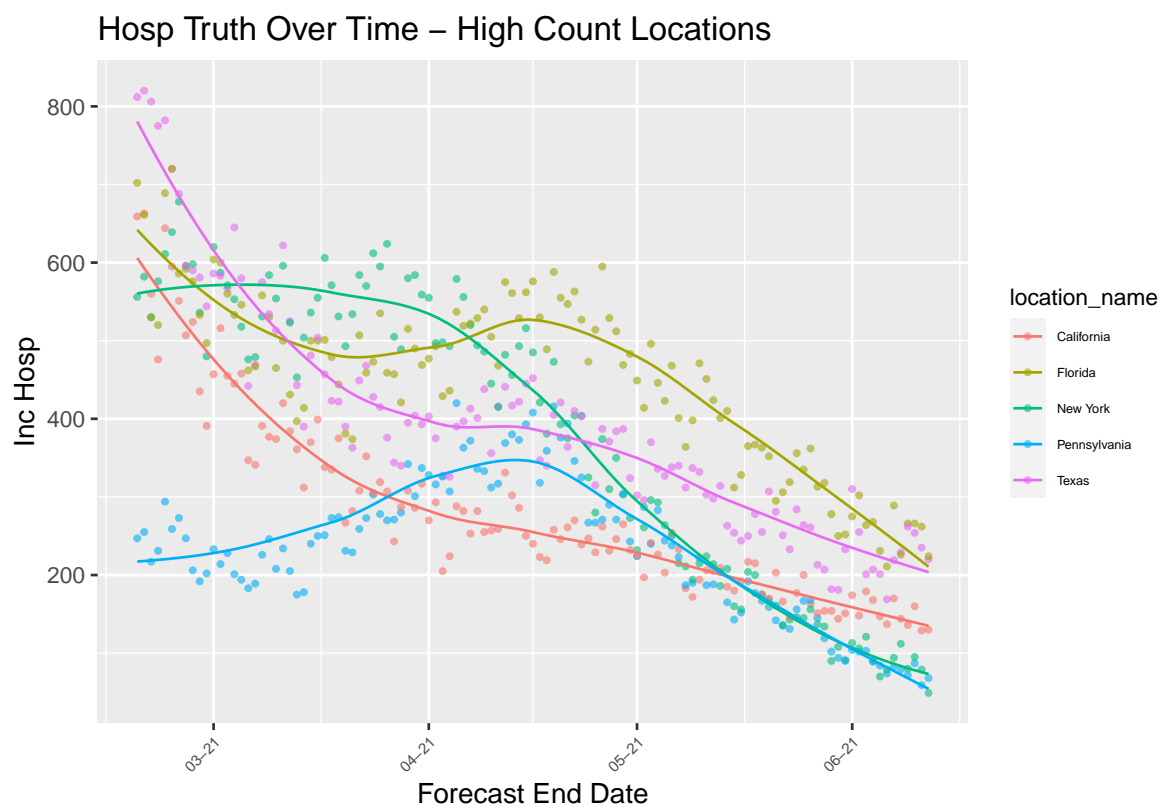
Mean Approx. CD from COVIDhub-ensemble Over Time – Low Count Locations



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

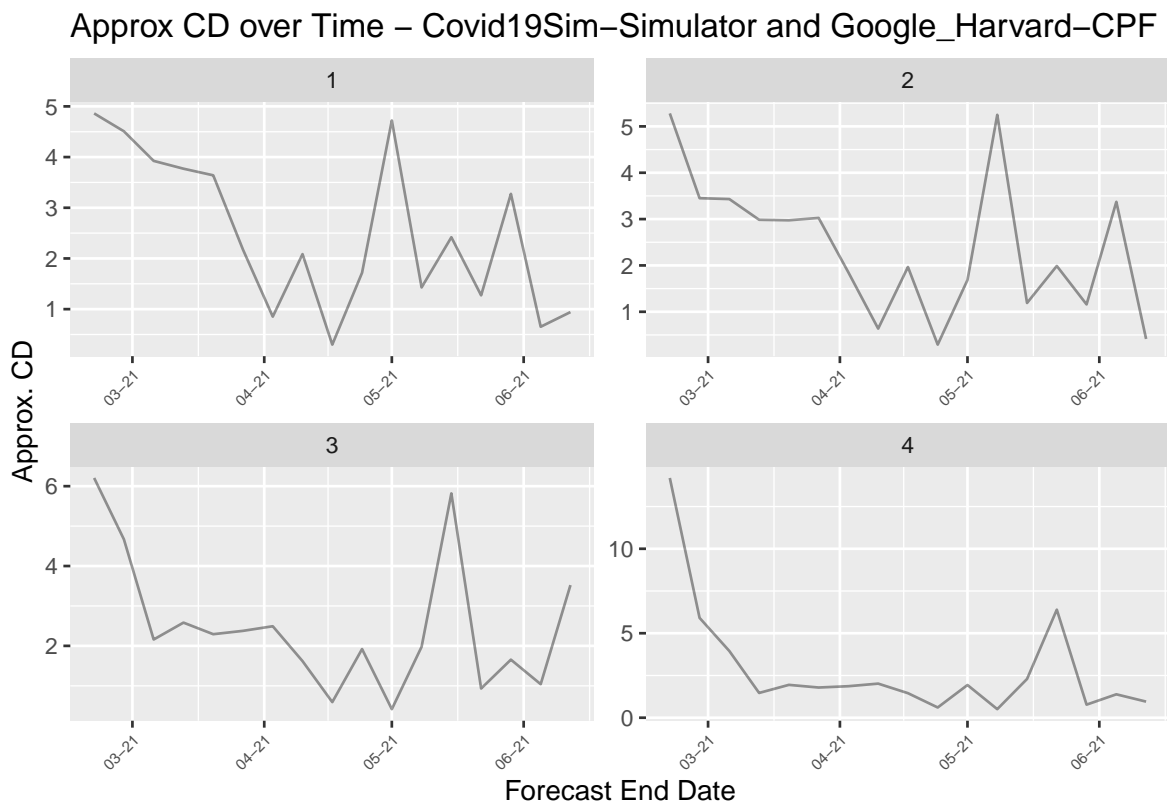
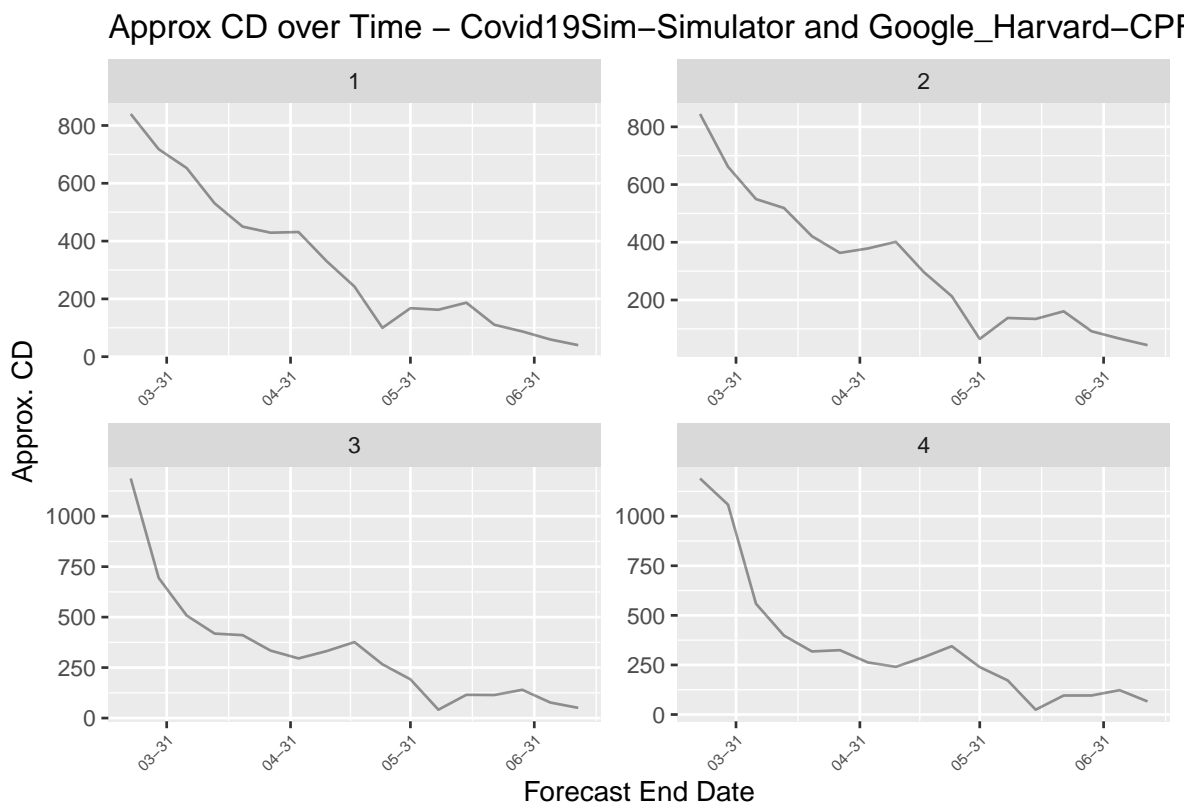
Whether models incorporate a day of the week effect does not seem to have an impact on how much the model differs from the ensemble, nor as to when it differs greatly.

These scatterplots are nearly the same as the ones shown above.



Like with the Thursday forecasts, it seems that Google_Harvard-CPF and Covid19Sim-Simulator's differ-

ences from the ensemble model follow the trends shown by the truth data.



Like with the Thursday forecasts, these plots indicate that the difference between the models at the high count locations seems to shrink over time. Since the low count locations approx CD is so low, it is difficult to draw conclusions from the plots.

We can also cluster the distances using hierarchical clustering.

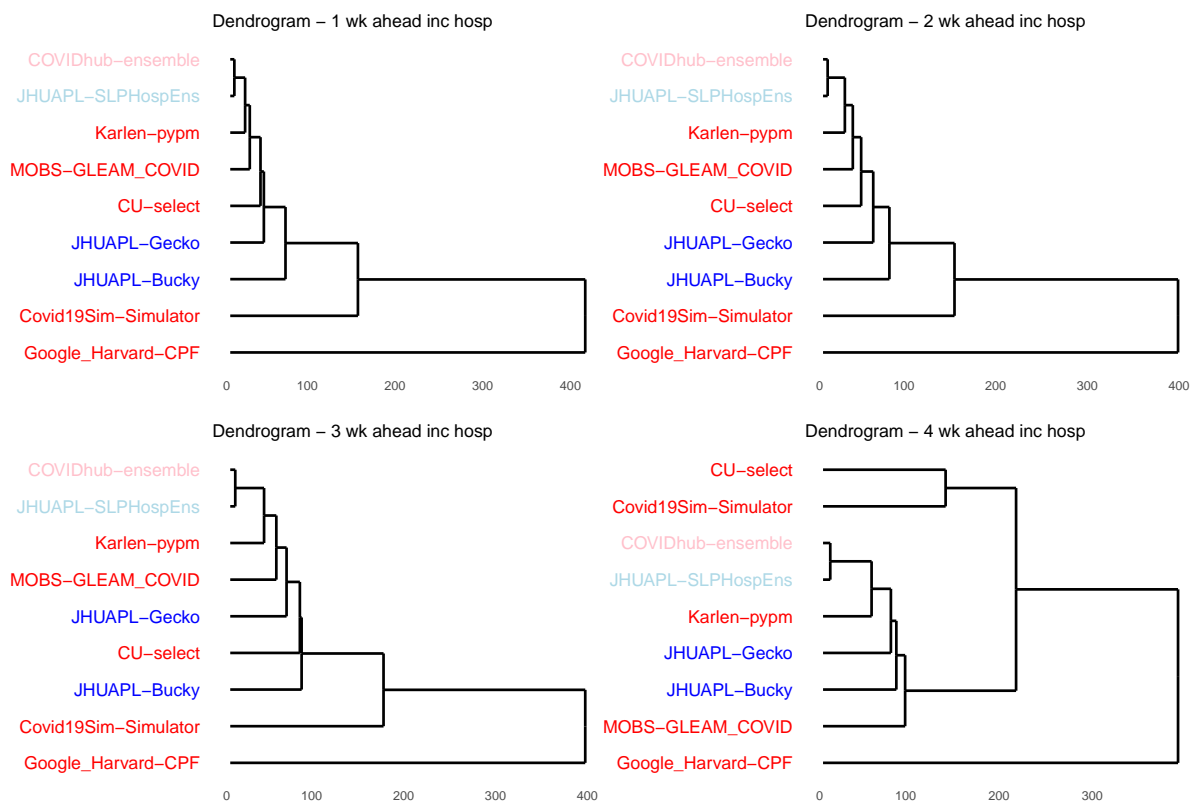


Figure 3: High Hospitalization Count Locations

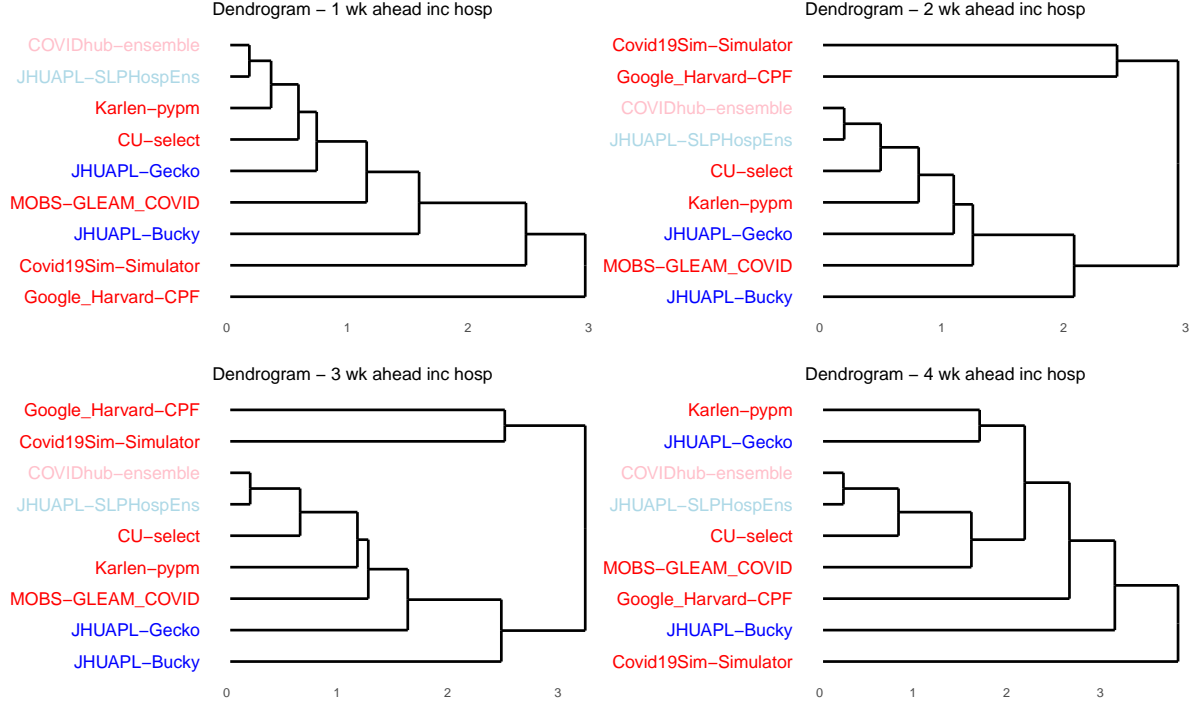


Figure 4: Low Hospitalization Count Locations

For both high count and low count locations, all of the dendrograms show that Google_Harvard-CPF is generally the most dissimilar to the other models, and substantially so for the high count locations, followed by Covid19Sim-Simulator. The one and two week horizons for the high count locations have practically identical dendrograms. The three week horizon one is very similar as well, save for the ordering of CU-select, JHUAPL-Gecko, and JHUAPL-Bucky, but CU-select and JHUAPL-Bucky have almost negligible difference in their Cramer's Distances. However, the shape of the dendrogram changes at the four week horizon for high count locations such that Covid19Sim-Simulator and CU-select are a separate branch together, even though these two models are farther apart for other horizons.

The dendrograms for low count locations are different at each horizon. The one week horizon plot is similar to those for one to three week horizons of high count locations. Meanwhile, the two and three week horizon plots are resemble each other. However, the four week horizon dendrogram is different. At the same time, the scale for differences in Cramer Distance for the low count locations is very small, which may be a result of low hospitalizations, which may explain why there is so much variation in the dendrograms across horizons.

For Saturday forecasts, Google_Harvard-CPF is consistently the most dissimilar from other models, followed by Covid19Sim-Simulator, across almost all horizons for both high-count and low count regions. This is the same as for Thursday forecasts. For Saturday forecasts, the heat maps seem to indicate models without day of the week effects tend to be more dissimilar from the COVIDhub-ensemble model but the dendrograms show less conclusive results.

Day of the week effects may lead to a slight difference between models, but it does not seem to be a significant factor in explaining their differences.