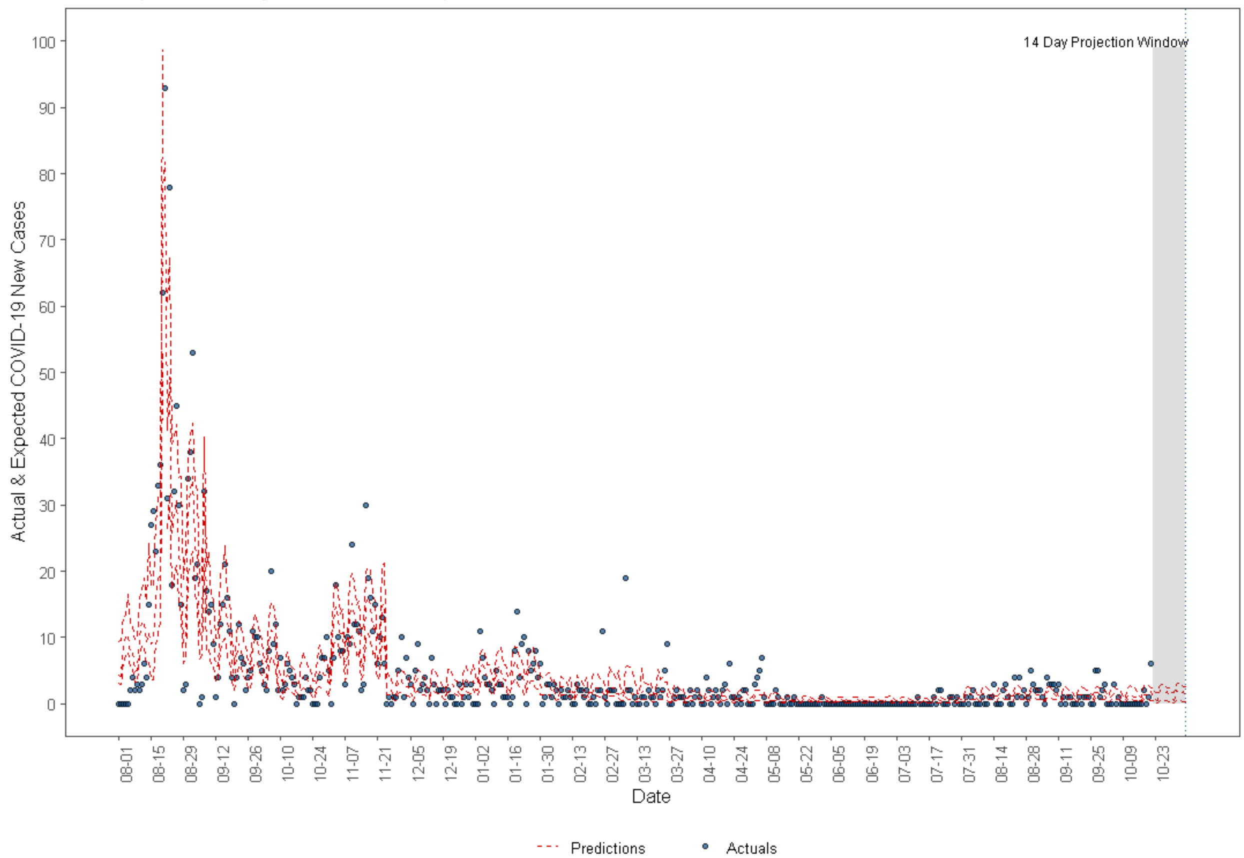
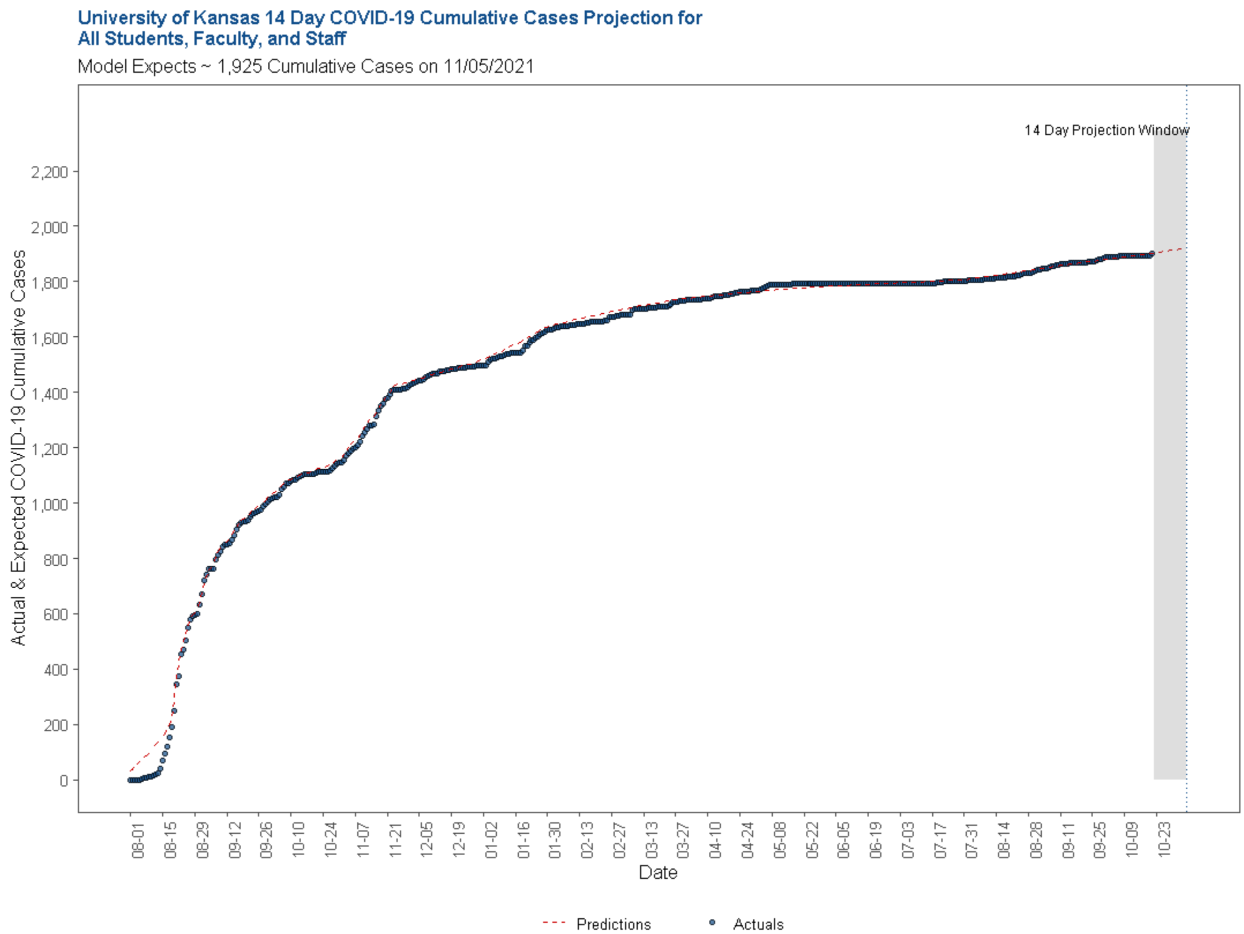
University of Kansas 14 Day COVID-19 New Cases Projection for All Students, Faculty, and Staff

Model Expects an Average of 2 New Cases/Day on 11/05/2021





University of Kansas 14 Day COVID-19 Cumulative Cases Projection for
All Students, Faculty, and Staff
Model Expects ~ 1,925 Cumulative Cases on 11/05/2021
14 Day Projection Window
Actual & Expected COVID-19 Cumulative Cases
2,200
2,000
1,800
1,600
1,400
1,200
1,000
800
600
400
200
0
08-01
08-15
08-29
09-12
09-26
10-10
10-24
11-07
11-21
12-05
12-19
01-02
01-16
01-30
02-13
02-27
03-13
03-27
04-10
04-24
05-08
05-22
06-05
06-19
07-03
07-17
07-31
08-14
08-28
09-11
09-25
10-09
10-23
Date
Predictions
Actuals

Methodology

The goal of the 14 day rolling forecast is to give decision makers a sense for where the growth of COVID-19 cases is likely headed based on all of the most up to date information available. Said another way, the forecast is less a tool for exact prediction on any given day and is instead a tool that helps us think about how case growth might unfold if we push the growth trend 14 days into the future.

The forecast was created with something called an ensemble model. In plain English, an ensemble model is a mathematical way of 'putting multiple heads together' to produce better predictions on average than any single 'head' could produce on its own. From a high level, we built four separate models – sort of like having multiple similar but different opinions on a complex question. Next, we made predictions from each model. Finally, we averaged each model's predictions to produce a median prediction, a minimum prediction and a maximum prediction, which you can see in the graphs as the dashed red lines.

Under the hood, we created a weighted average of the following four model types: generalized linear model (GLM), generalized additive model (GAM), random forest (RF), and conditional inference tree (Tree). Each of the models provides unique information for the ensemble prediction. For example, on one hand, the GLM and GAM models fit smooth lines to the data, much like drawing a line through a set of points on an x-y graph. On the other hand, the RF and Tree models identify date ranges in the data which have meaningfully different trends and fit lines in those subset date ranges.

The four models (GLM, GAM, RF, and Tree) are combined using weights on each model's predictions. The weights applied to each model's predictions are determined through a common measure of predictive error - the root mean squared error (RMSE). The set of weights which produce the lowest RMSE are chosen as the set of weights which produce the forecast we see above.

Lastly, when interpreting the graphs, the wider ranging the actual data is the wider ranging the forecast will be, which makes sense because wider ranging data presents us with more uncertainty. If, for example, reported new cases vary between 0 and 100 over the last 10 days, then the ensemble model will expect a wider range of reported new cases going forward compared to what the ensemble model would expect if instead we recorded between 0 and 10 new cases over the last 10 days.