

# **Covidcast: Forecasting Aids for Delphi**

August 11, 2021

Taha Bouhoun, Michelle Lee, Shilaan Alzahawi

# Covidcast: Forecasting Aids for Delphi



**Taha Bouhoun**

Minerva University  
*DSSG Fellow*



**Michelle Lee**

Columbia University  
*DSSG Fellow*



**Shilaan Alzahawi**

Stanford University  
*Technical mentor*



**Balasubramanian  
Narasimhan**

Stanford University  
*Faculty mentor*



**Daniel McDonald**

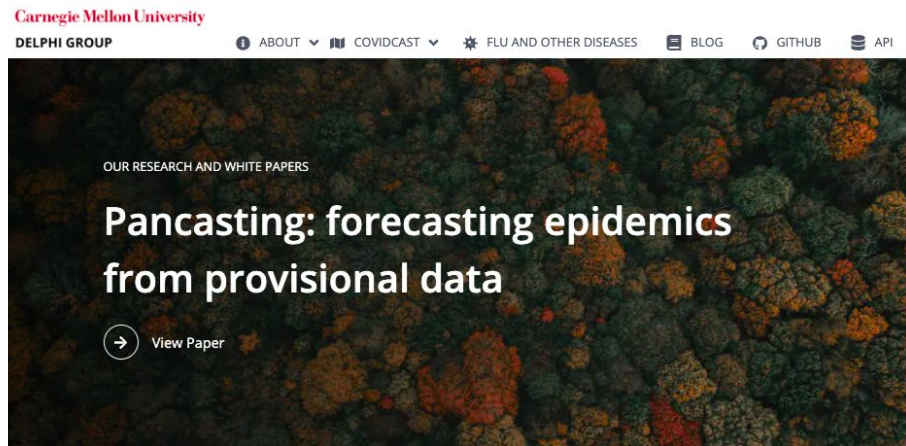
University of British  
Columbia, Canada  
*Faculty mentor*

# Delphi Research Group



The Delphi Research Group at Carnegie Mellon University and is **one of the two influenza forecasters in the United States**

- The group's goal is to develop the theory and practice of **epidemiological forecasting**
- Prior to COVID-19, the group also worked on forecasting for influenza, dengue, and norovirus



# Delphi Covidcast

Since March 2020, the Delphi research group has maintained **the largest public repository of real-time indicators of COVID-19 activity**, through a public API.

Every Monday, the Delphi Covidcast generates forecasts of cumulative COVID-19 cases and deaths in the U.S. These predictions are reviewed by the team and sent to the CDC COVID-19 Forecast Hub

**Target Variable**

- ☒ Incident Deaths
- ☐ Incident Cases
- ☐ Hospital Admissions

**Scoring Metric**

- ☒ Weighted Interval Score
- ☐ Spread
- ☐ Absolute Error
- ☐ Coverage

**Y-Axis Score Scale**

- ☐ Log Scale
- ☐ Scale by Baseline Forecaster

**Forecasters**

Type a name or select from dropdown

COVIDhub-baseline  
COVIDhub-ensemble

Some forecasters may not have data for the chosen location or scoring metric.

**Forecast Horizon (Weeks)**

☒ 1 ☒ 2 ☐ 3 ☐ 4

**Location**

Totaled Over States\*

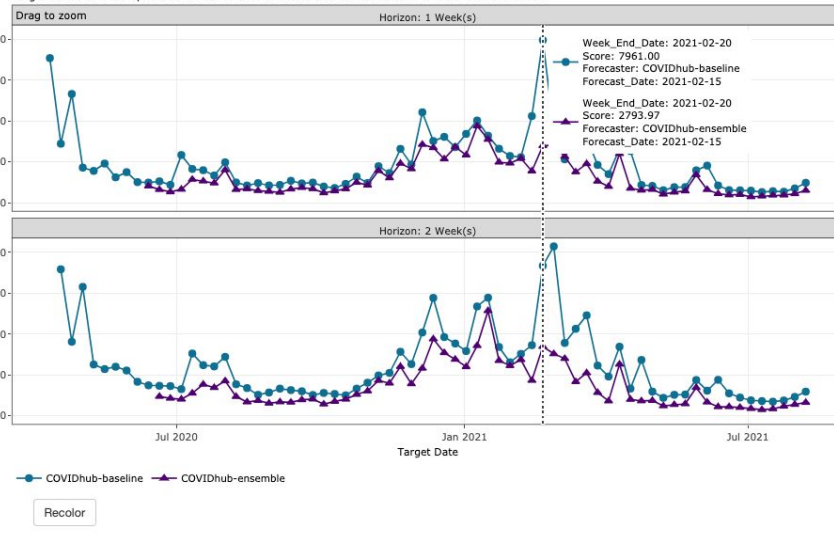
[Download CSV](#)

About

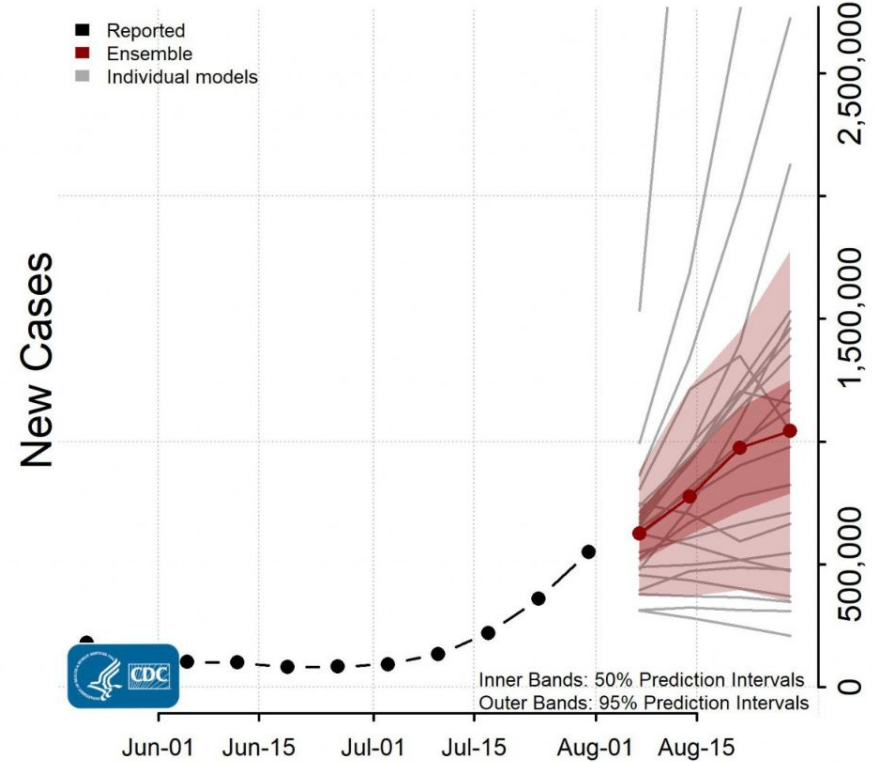
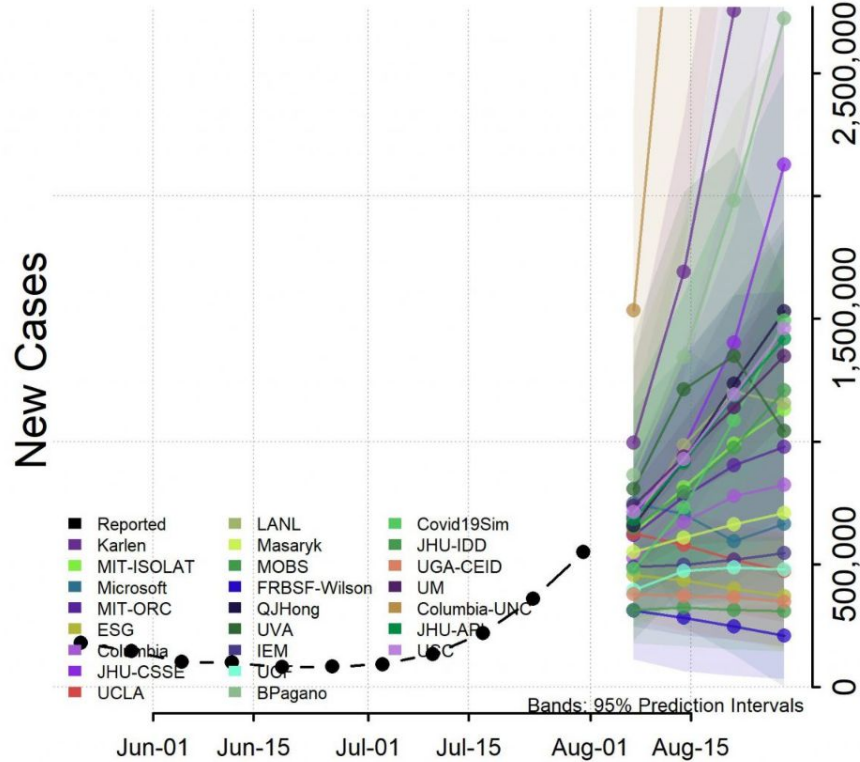
Evaluation Plots

## Weighted Interval Score

Target Variable: Deaths, Location: Totaled over all states and territories common to these forecasters\*



# National Forecast



## What DELPHI would like to know

- How does our (Delphi's Covidcast) forecaster do compared to others?
- Assess new forecasters before they are deployed
- Are there periods of time that we do much worse or better?
- Are there areas of improvement we need to focus on?

**Our project involves the creation of a report that answers these questions**

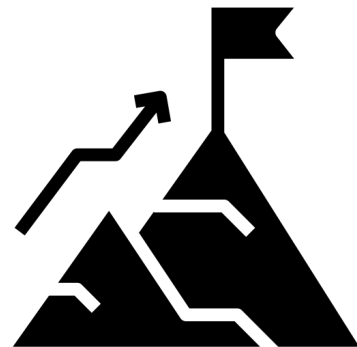
# Goals & Deliverables

## Our Goal

- Develop tools for comparing and evaluating COVID forecasters

## Our Deliverables

- An interactive parameterized report that evaluates and compares the performance of several COVID-19 forecasters for cases, deaths, and hospitalizations
  - Along with the report, the user can download the underlying report-specific data
- The user can automatically generate a report according to their chosen parameters
  - The number of *epi-weeks ahead* that the forecasts are made
  - The specific *forecasters* to compare to
  - Whether to use a *colorblind-safe* palette for generating the plots
- A GitHub repository with fully documented code and vignettes



# Outcomes of Interest & Data Sources

- **Covid-19 cases**  
Number of daily confirmed cases reported by state and local health authorities
- **Covid 19- deaths**  
Official figures of death due to COVID-19 as confirmed by health authorities
- **Covid-19 hospitalizations**  
Daily Covid-19 related hospital admissions, estimated from health authorities' aggregated statistics and patient data



# Metrics to Evaluate Forecasting Performance

- **Weighted Interval Score (WIS)**

A proper score that combines a set of prediction interval scores. A smaller WIS indicates better performance

- **Coverage**

An estimate of the probability that a forecaster's 80% interval correctly includes the actual value

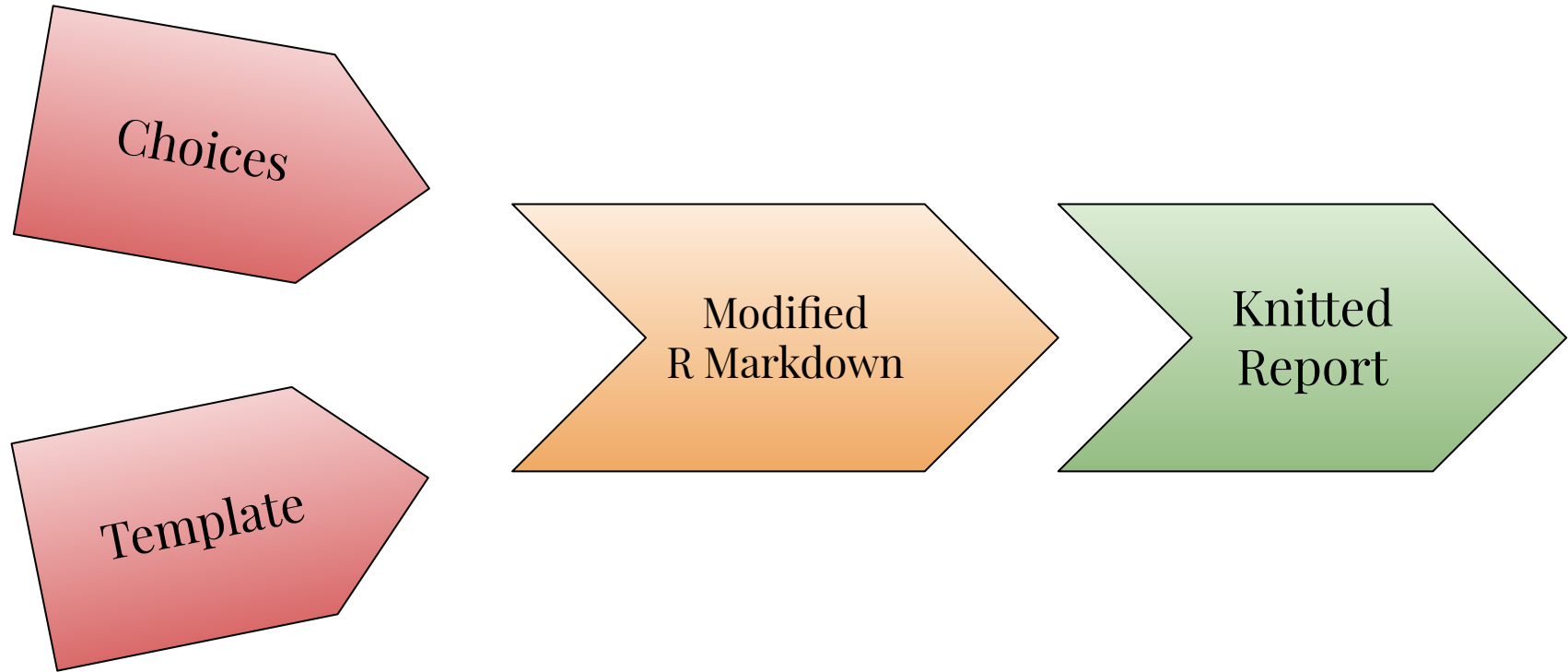
- **Absolute Error**

The difference between the actual value and the point forecast

# Limitations of the Current Report

- Too specific
- API based:
  - Slow in knitting in R studio
  - Unable to run reports in case of API problems
- Unable to personalize
- Visually unappealing (many plots, colors of the graphs)

# Project Architect: Generalizable Report



# Functionality

**Step 1** : Generate (*cases, hospitalizations, or deaths*) report with chosen parameters

Knit with Parameters

Forecasters:

COVIDhub-ensemble COVIDhub-baseline CMU-TimeSeries Karlen-pypm  
CU-select MIT-Cassandra COVIDhub-trained\_ensemble

Primary forecaster:

CMU-TimeSeries

colorblind\_palette

Weeks

1 2 3 4

☐ printcode

Cancel

Knit

## Evaluation of COVID-19 Forecasters

[Code](#)

### Parametrized Report

07 August 2021

#### Abstract

This notebook is a template for evaluating COVID forecast submissions from COVIDhub. After inputting a set of parameters (forecasters, COVID signals, etc), the template yields a comprehensive report on the predictions of COVID forecasters as well as their performance compared to the ground truth. The visualizations generated by the template offer an intuitive way to compare the accuracy of forecasters across all US states.

### Retrieving Forecast Data

Every week, forecasters submit their predictions to COVID-19 ForecastHub. In this report, we rely on an AWS bucket that contains the estimates of a handful of signals (e.g., COVID death, cases, hospitalization, etc). Furthermore, the AWS server stores an array of evaluation metrics of these forecasts (e.g., Absolute Error, Weighted Interval Score, and 80% Coverage). Alternatively, the data can be retrieved from the publicly accessible [covidcast](#) and [covidval](#) APIs.

[Code](#)

#### The target forecast dates are:

2021-05-10, 2021-05-24, 2021-06-07, 2021-06-21

#### The template will compile data of the following forecasters:

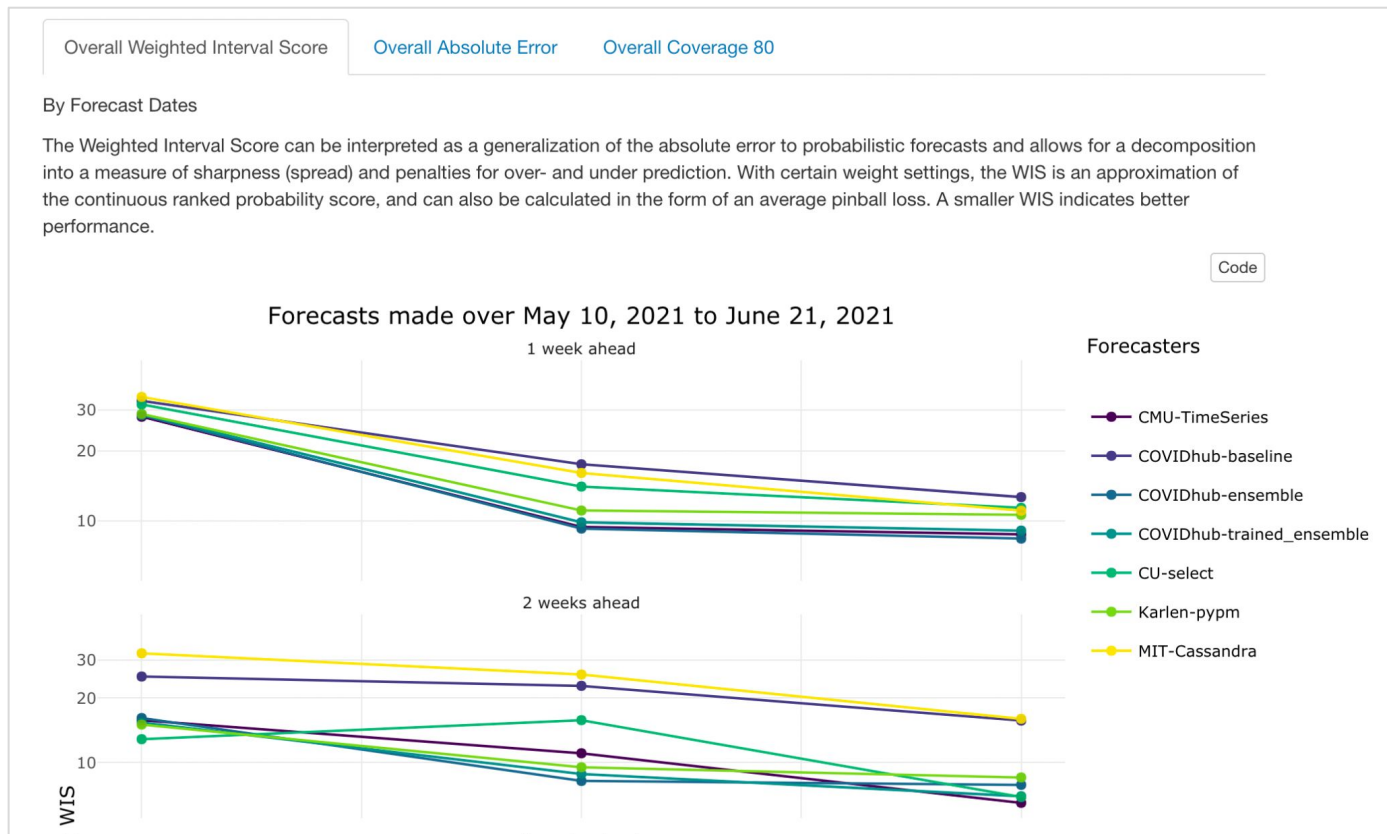
COVIDhub-ensemble, COVIDhub-baseline, CMU-TimeSeries, Karlen-pypm, CU-select, MIT-Cassandra, COVIDhub-trained\_ensemble.

#### The primary forecaster:

CMU-TimeSeries

# Functionality

## Step 2: Explore interactive graphs in tabs



# Functionality

## Step 3 (optional): Download underlying data

To promote the flexibility to replicate the report, the data used in this report can be easily downloaded as a CSV file. By doing so, the user can generate customized plots or even include their own forecaster.

 Download Predictions Evaluation

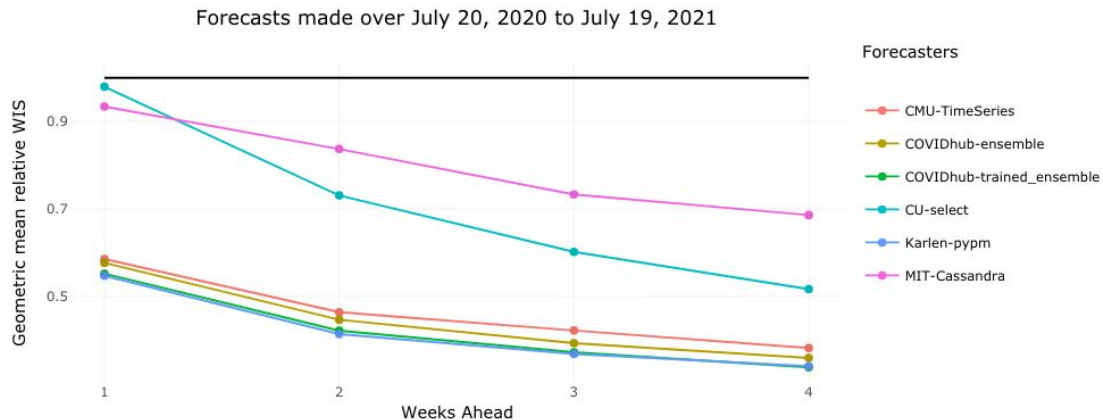
Code

 Download Raw Predictions

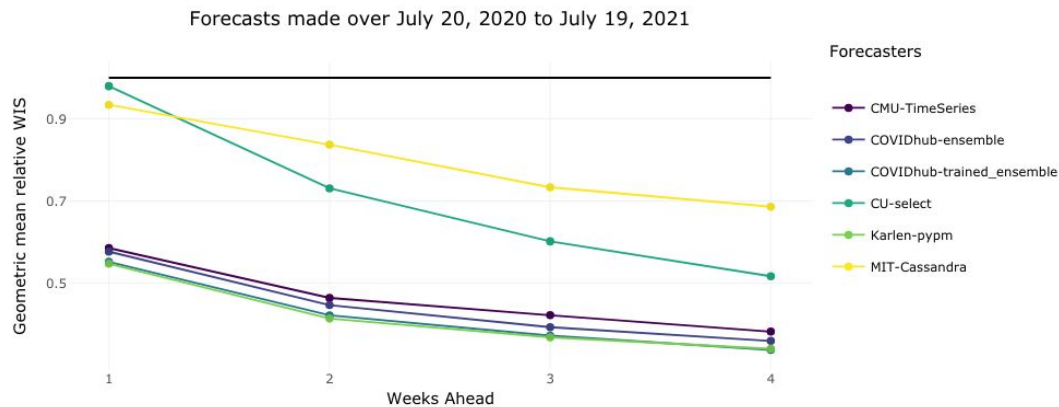
Code

# Colorblind-Safe Mode

Before:



After:



# Identifying Discrepancies and Performances

## Maps

To contextualize the forecast evaluations, the following tabs illustrates the performance of COVID forecasts across all US states over forecast dates and weeks ahead. Note that the results are scaled by population.

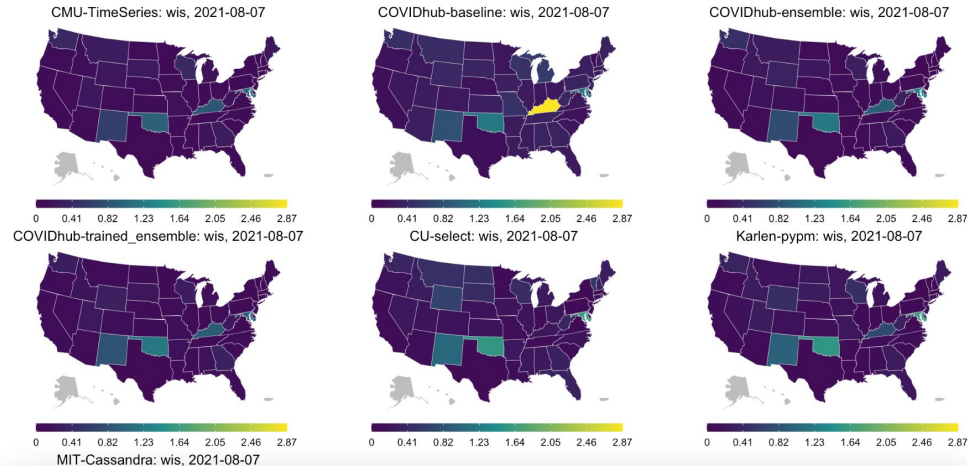
[Code](#)

Mean Weighted Interval Score

Mean Absolute Error

Mean Coverage 80

The Weighted Interval Score can be interpreted as a generalization of the absolute error to probabilistic forecasts and allows for a decomposition into a measure of sharpness (spread) and penalties for over- and under prediction. With certain weight settings, the WIS is an approximation of the continuous ranked probability score, and can also be calculated in the form of an average pinball loss. A smaller WIS indicates better performance.

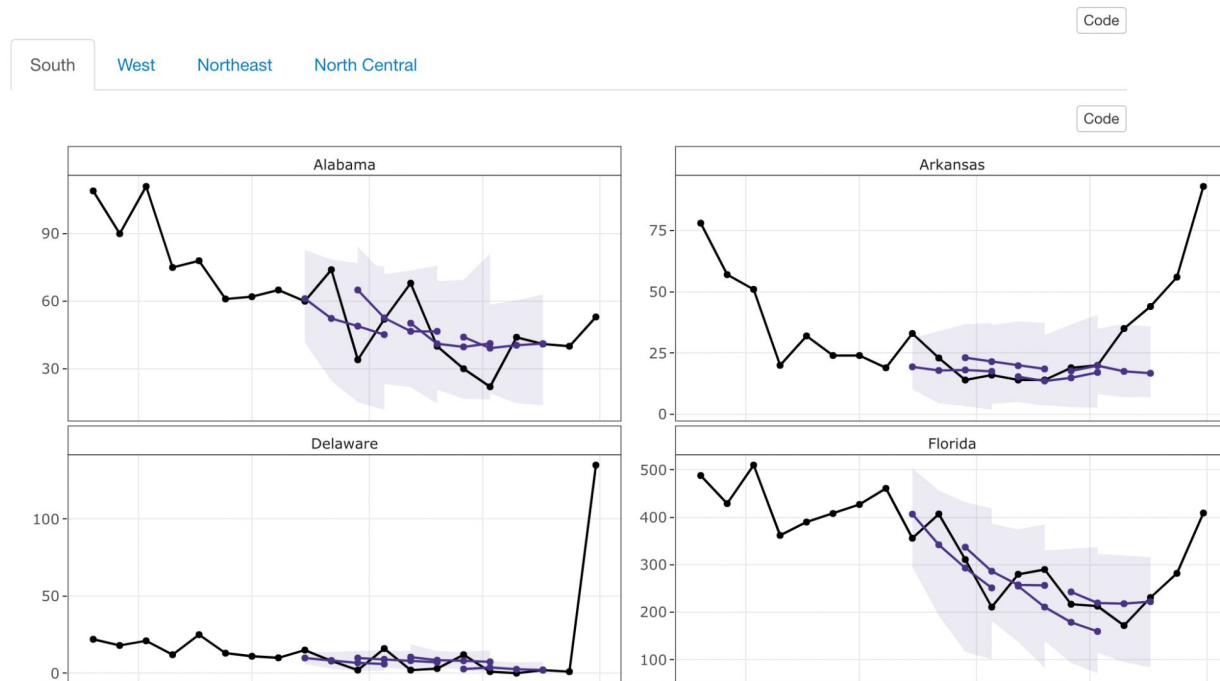
[Code](#)



# A look inside...

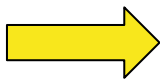
## Trajectory plots

The following plots show the predictions of the **CMU-TimeSeries** forecaster along with the confidence interval for each of the US states. The forecasts project 1, 2, 3, 4 weeks ahead.



## Limitations of the Original Report

1. Too specific
2. API based:
  - a. Slow in knitting in R studio
  - b. Unable to run reports in case of API problems
3. Unable to personalize
4. Visually unappealing (many plots, colors of the graphs)



## Solutions

1. Use parameters and helper functions that can change the markdown parameters
2. Allow download of preformatted data from AWS bucket and prediction data frame (avoid API call)
3. Add better interactivity to plots
4. Organize the plots into tabs for easier navigation

## Project Artifacts

- Templated markdown files
- Auxiliary R scripts for manipulating markdowns and generating reports
- Example reports
- A GitHub repository with fully documented code and vignettes

## Future Directions

- Shiny app that generates the report with the click of a button
- County-specific forecaster performance
- Docker solution for batch generation of reports

# Thank you to our mentors and the DELPHI team !



**Shilaan Alzahawi**

*Technical mentor*



**Balasubramanian  
Narasimhan**

*Faculty mentor*



**Daniel McDonald**

*Faculty mentor*



**Stanford** | Data Science

**Carnegie Mellon University**  
**DELPHI GROUP**