# Detecting Changepoints During COVID-19

Dhruv Arora, Priyanka Bijlani, Sharmeelee Bijlani, Divya Pandey, Lakshmi Venkatasubramanian

## INTRODUCTION

Meta's Infrastructure Data Science team has released a package called KATS to analyze timeseries datasets. The KATS package implements multiple changepoint detection algorithms and identifies points in a timeseries which show a sudden or abrupt change. A changepoint is defined as a 'persistent change' rather than an anomaly or an outlier in the timeseries data.

## PROBLEM STATEMENT

The COVID-19 pandemic is an ongoing event in world history that various professionals are trying to comprehend in their respective domains. In this project, we as data scientists, make sense of the events during the pandemic with a machine learning approach of changepoint detection. We have conducted research on timeseries data and applied the KATS package to multiple datasets in various contexts to assess and evaluate the performance of the timeseries algorithms, and to provide recommendations to improve the algorithms from our applied research.

## METHODOLOGY

1. Explore KATS Changepoint Detection Algorithms with TCPD Benchmark Study
   - Cumulative Sum Detector
   - Bayesian Online Changepoint Detector
   - Robust Stat Detector
2. Apply Algorithms to Google Mobility Data
3. Tell the COVID-19 Story through True and Detected Changepoints

## TCPD BENCHMARK STUDY

There are two challenges with changepoint detection problems. The first is using the algorithms to identify changepoints in the dataset since the true label is unknown. The second is identifying the optimal number of changepoints by tuning hyperparameters. We used the Turing Changepoint Datasets benchmark to evaluate the KATS algorithms for their performance on over 30 time series with different characteristics. Each timeseries in the study includes crowdsourced annotations which serve as true changepoints. With a source of truth for changepoints and hyperparameter tuning of the algorithms, we developed an algorithm selection criteria based on characteristics of a timeseries.

### CHANGEPOINT ALGORITHM SELECTION

| Timeseries | CUSUM | BOCP | RS |
|---|---|---|---|
| Multiple CPs | | ✓ | ✓ |
| Seasonality | | ✓ | |
| Outliers | ✓ | | ✓ |
| Few Data Points | ✓ | | ✓ |
| Known CP Direction | ✓ | | |
| No Domain Knowledge | ✓ | ✓ | |

## GOOGLE MOBILITY DATA

There are six categories of global mobility data aggregated at a country, state, and county level, such as retail, transit, and park traffic. The data is completely anonymized from users who have turned on the Location History setting, which is off by default. The data spans from February 2020 to February 2022 and marks the day-to-day percent change compared to a pre-pandemic baseline. We applied the KATS changepoint algorithms to this data to analyze how the algorithms behave in a real data scenario where catalyst dates are known, but the changepoints are unknown. The time when a change is announced may differ from the time when the change is observed in effect.

## CHANGE IN MOBILITY: RETAIL AND RECREATION IN THE UNITED STATES

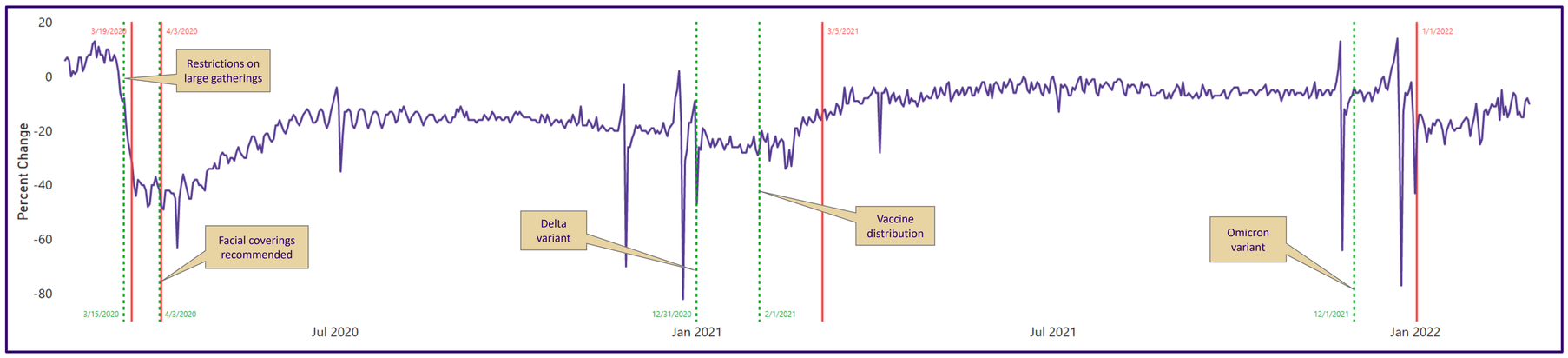


*Figure 1: Catalyst events during COVID-19 are marked by the dotted green lines. Changepoints detected by Robust Stat Detector are marked by solid red lines.*

## THE COVID-19 STORY

While true changepoints provide context on when an event occurred, such as policy change, detected changepoints mark when the change actually takes effect on the ground.

- *Figure 1*illustrates how the Robust Stat Detector translates the policy announcements during COVID-19 into changepoints that impacted retail and recreation in the United States.
- *Figure 2* is an example of how the CUSUM detector captures a drastic change in transit traffic and an expected, subtle changepoint when given an interest window.

## TRANSIT TRAFFIC IN KING COUNTY, WA

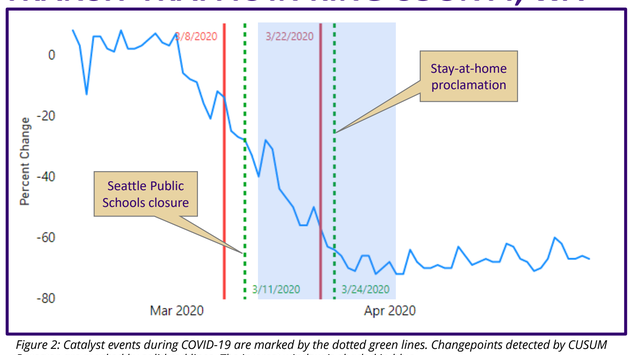


*Figure 2: Catalyst events during COVID-19 are marked by the dotted green lines. Changepoints detected by CUSUM Detector are marked by solid red lines. The interest window is shaded in blue.*

## RECOMMENDATIONS

- For Cumulative Sum Detector: Acceptable threshold values and multiple changepoint detection
- For Bayesian Online Changepoint Detector: Better handling of outliers
- For Robust Stat Detector: Better handling of seasonal data

### ACKNOWLEDGEMENTS

Sponsor: Sourav Chatterjee, Meta | Instructor: Megan Hazen, UW Data Science