

Covid-19 Study

Data Analysis and Modeling

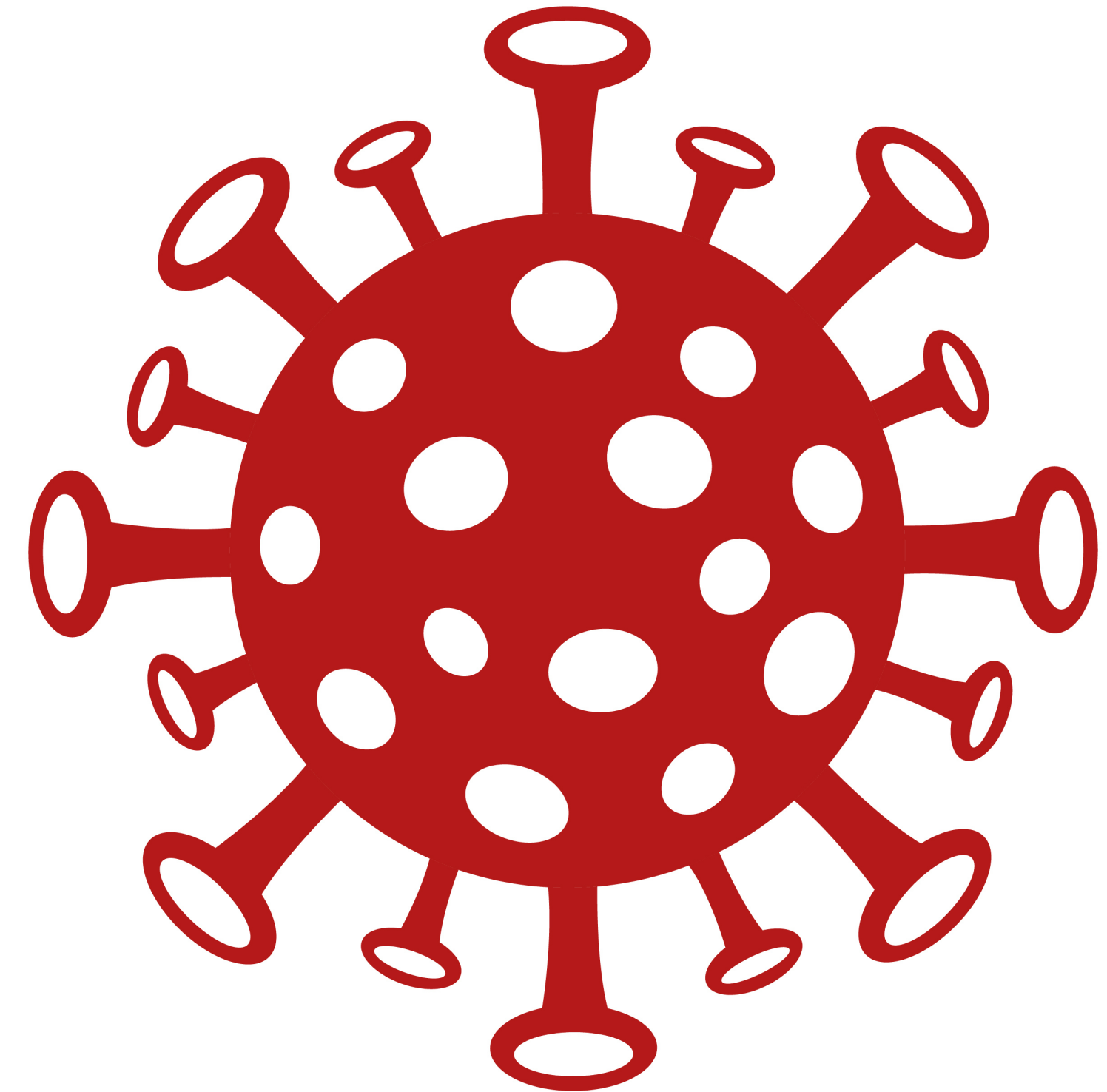
JR Garvin

February 16, 2024

Modern Day Plague

What do we do?

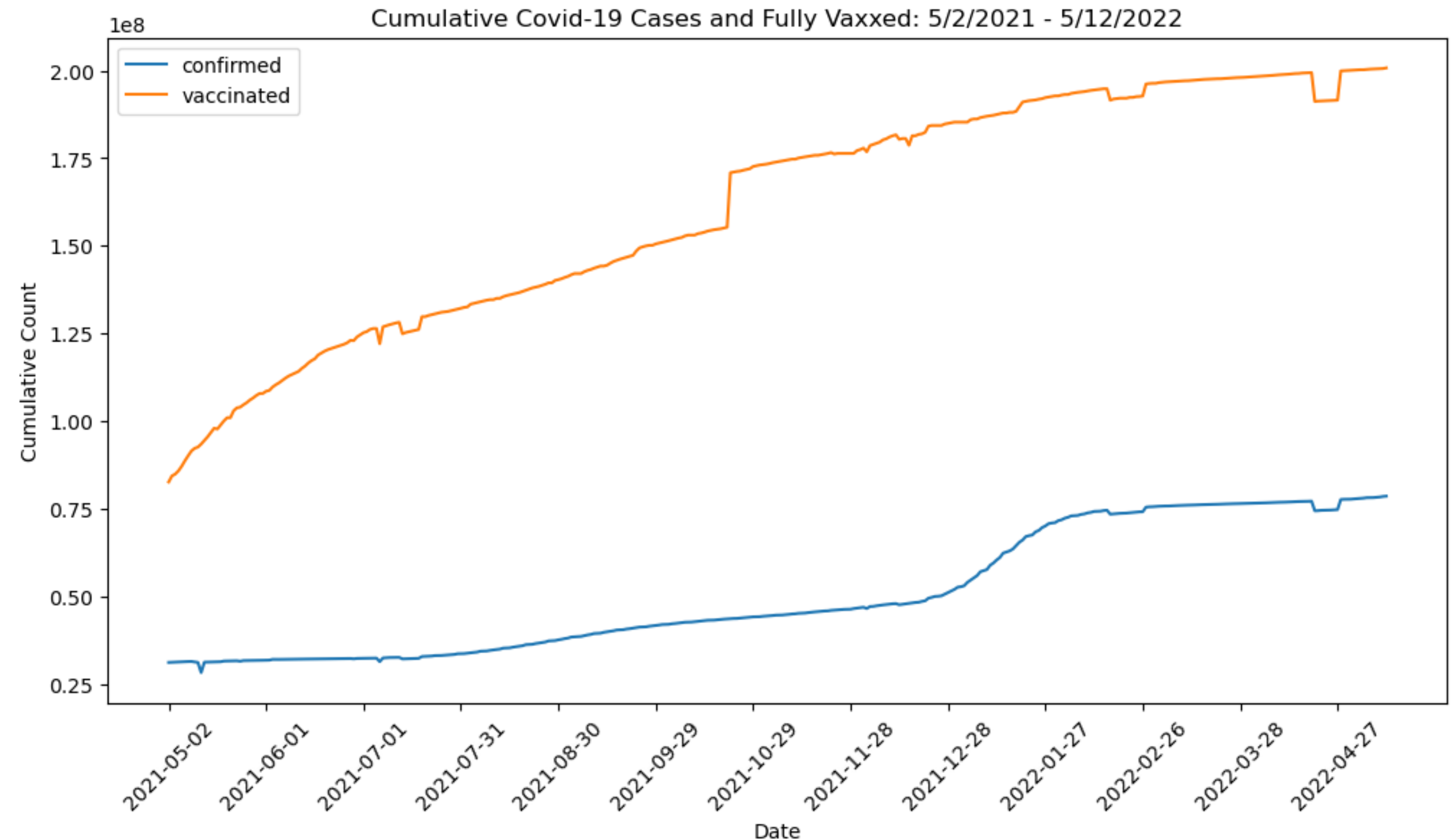
- Wuhan 2019 - World Wide Present
- 5th Deadliest
- China's "zero covid" vs. Sweden's "let it rip"
- **Follow the Data**



Methodology

Flatten the Curve

- Collect Data
- Tools for Tomorrow
 - Identify contributing factors
- Model the spread
- Forecast



Identify Contributing Factors

Inferential Analysis

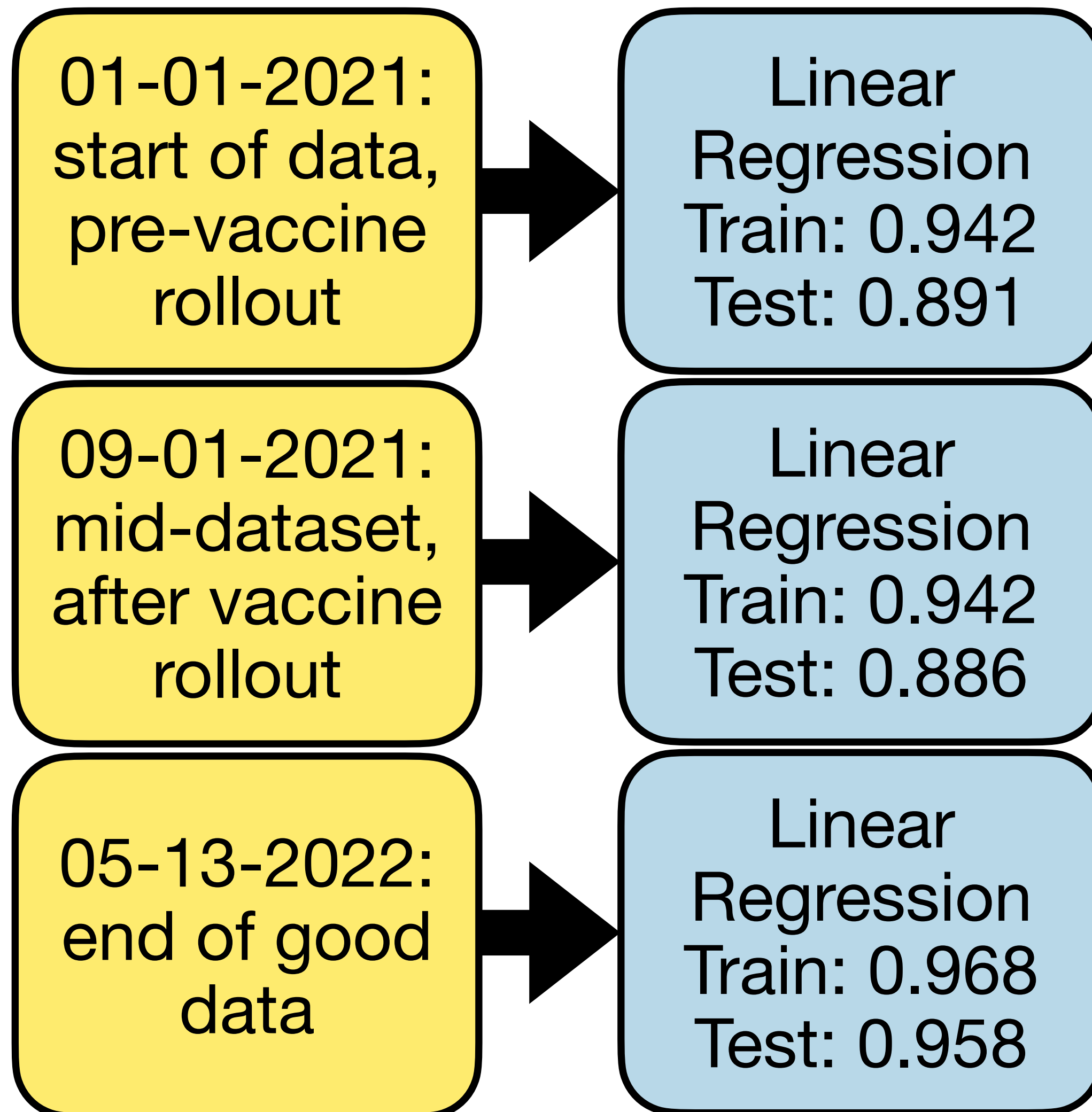
01-01-2021:
start of data,
pre-vaccine
rollout

09-01-2021:
mid-dataset,
after vaccine
rollout

05-13-2022:
end of good
data

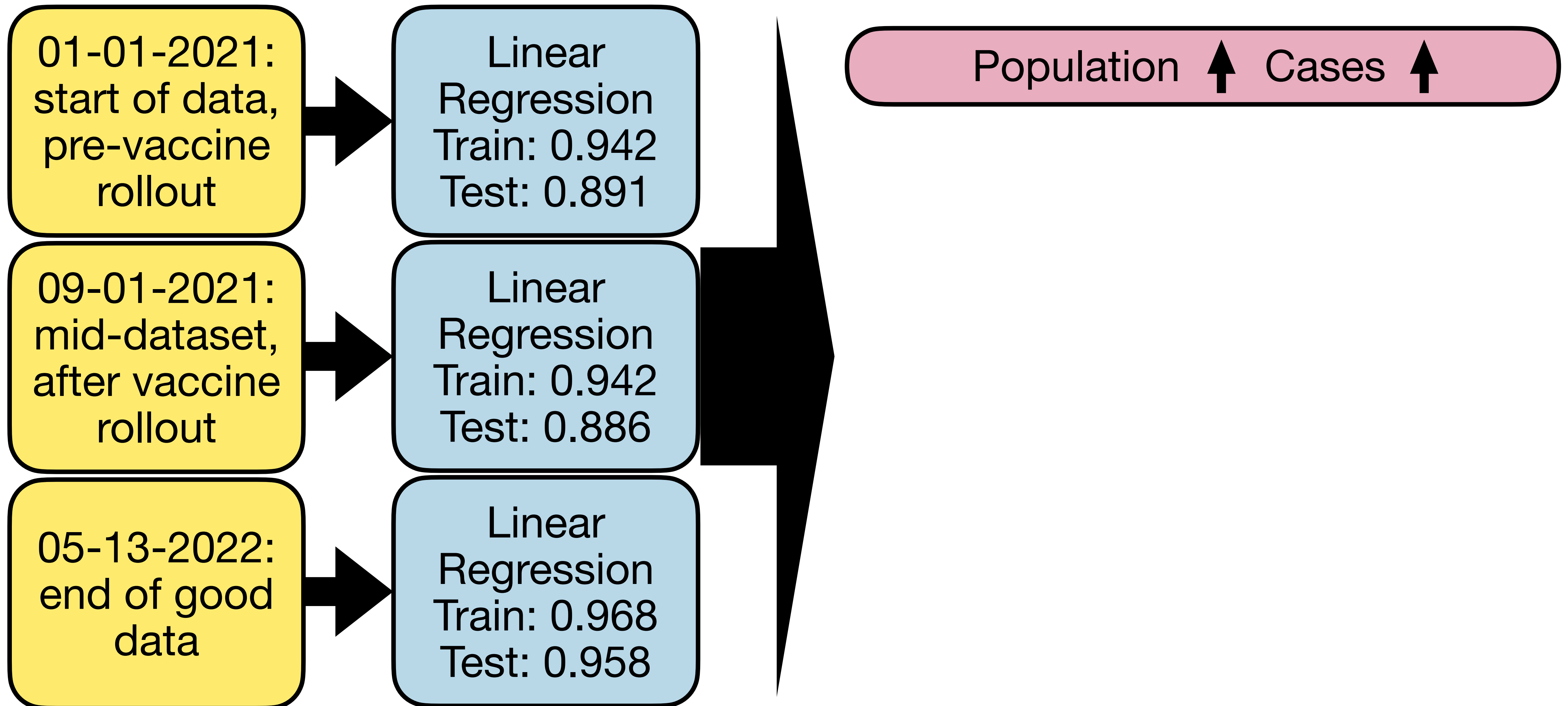
Identify Contributing Factors

Inferential Analysis



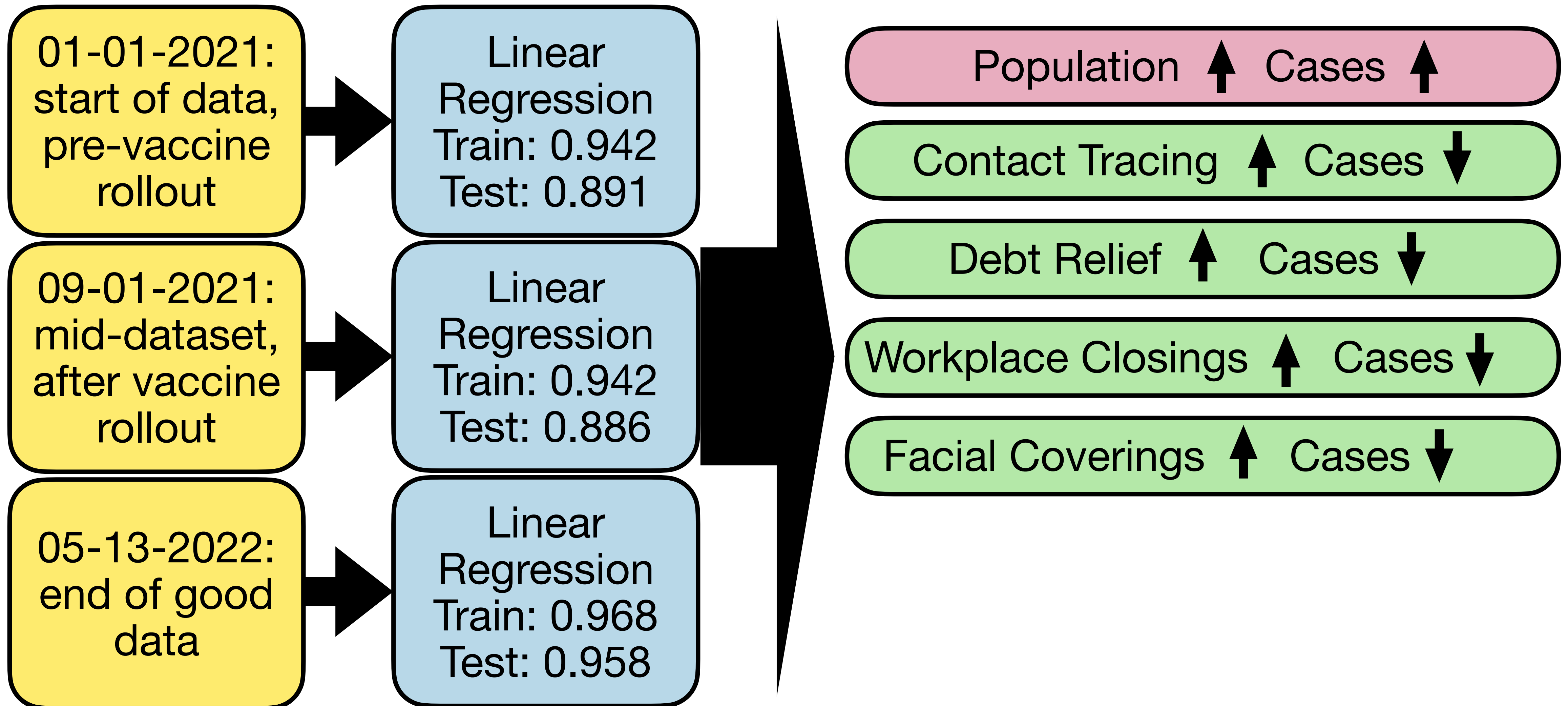
Identify Contributing Factors

Inferential Analysis



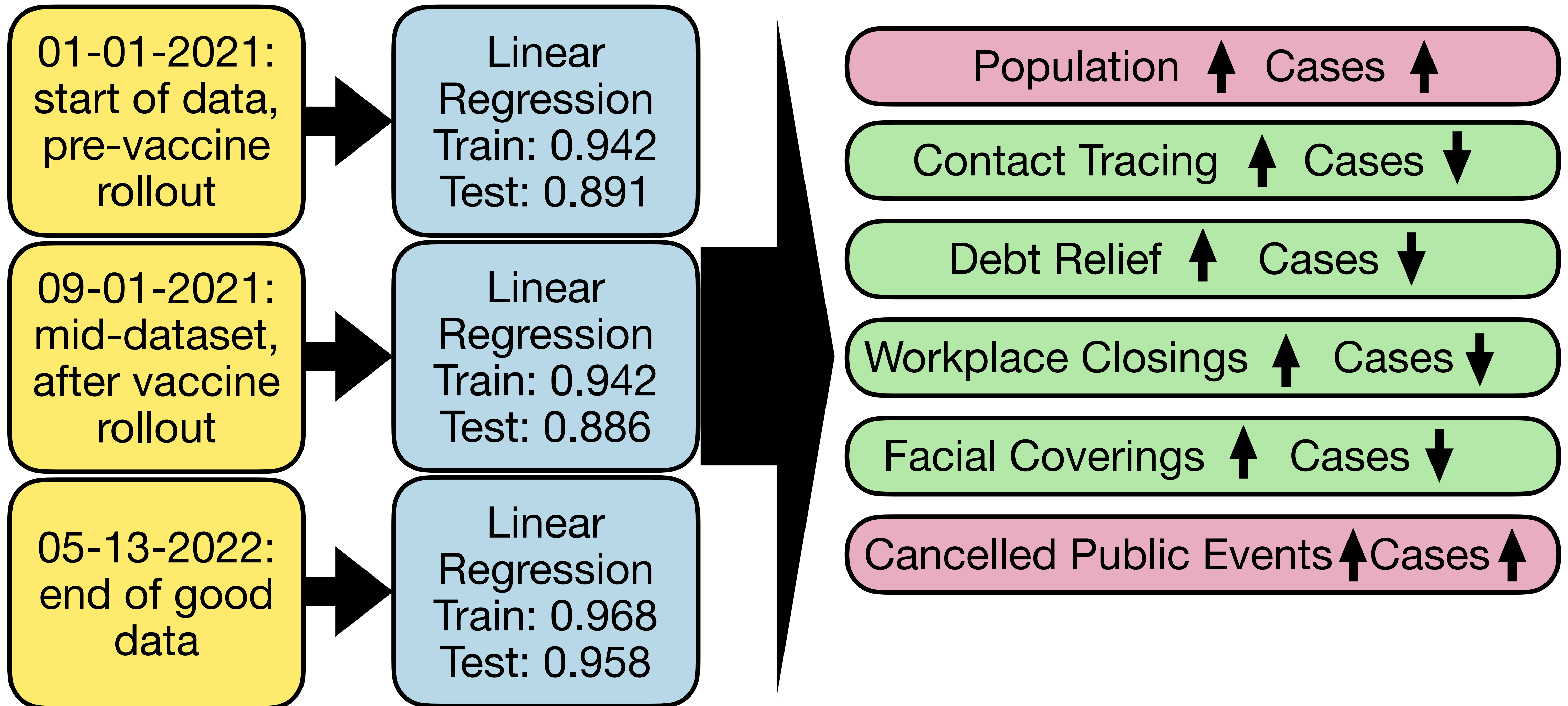
Identify Contributing Factors

Inferential Analysis



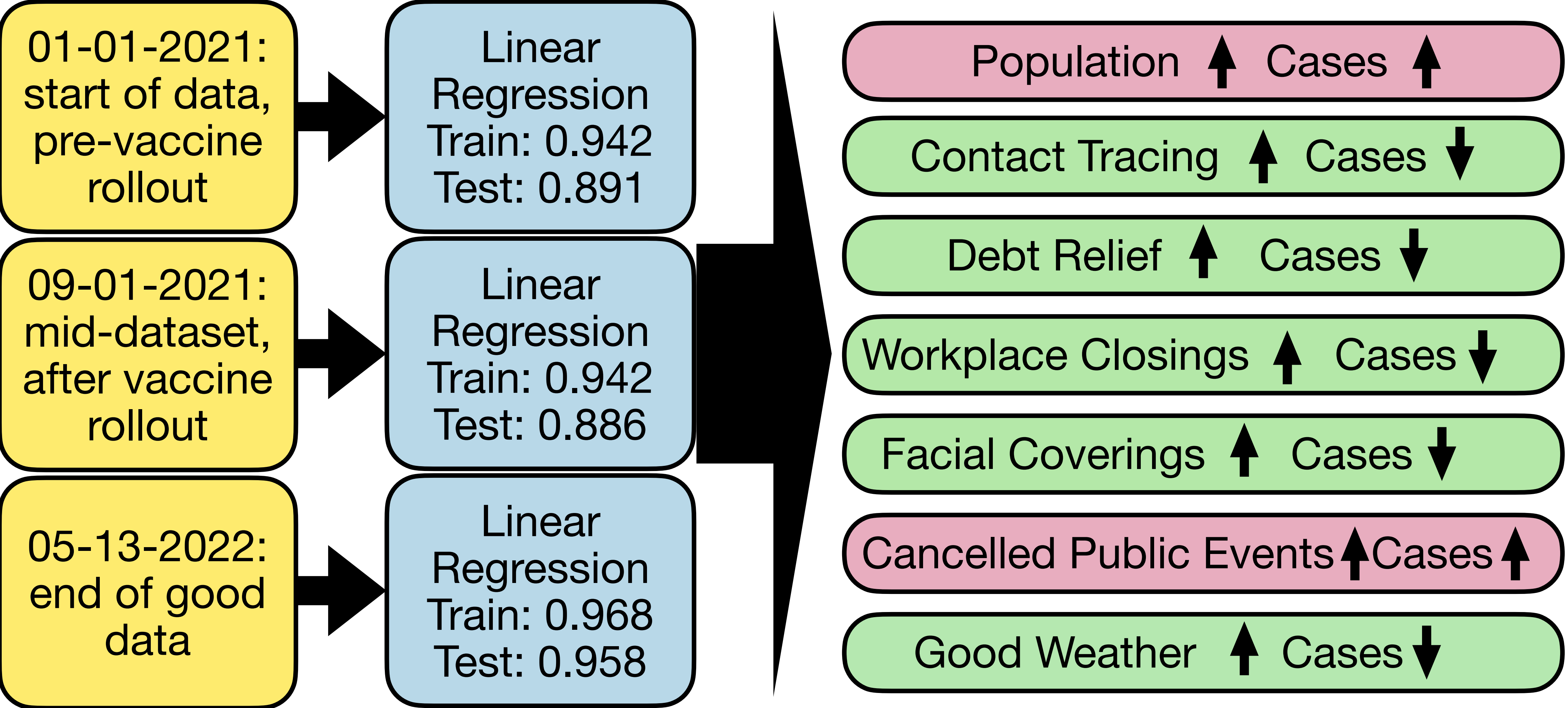
Identify Contributing Factors

Inferential Analysis



Identify Contributing Factors

Inferential Analysis



Model the Spread

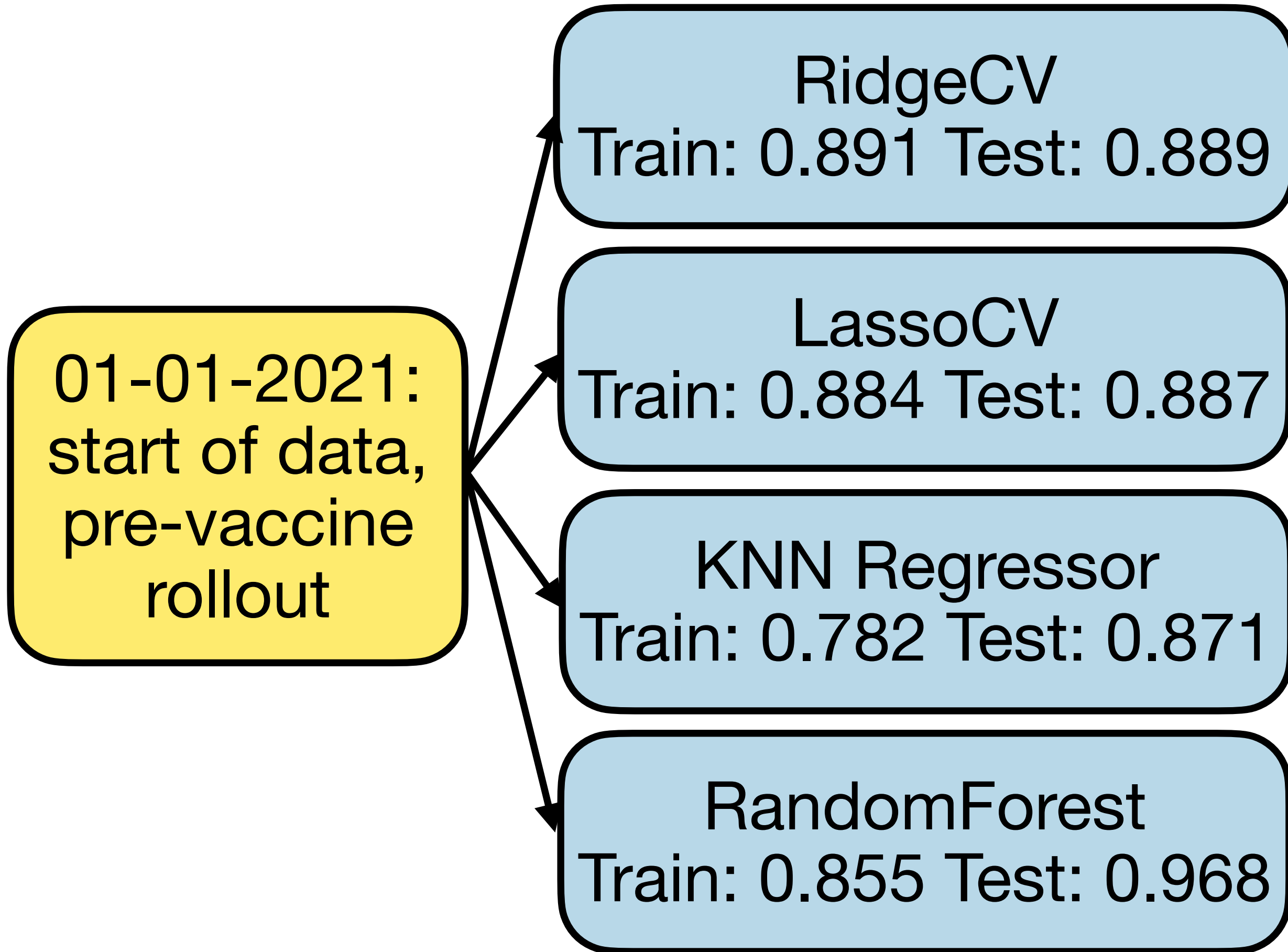
Correct/Substantiate other Sources

01-01-2021:
start of data,
pre-vaccine
rollout

Model the Spread

Correct/Substantiate other Sources

01-01-2021:
start of data,
pre-vaccine
rollout



```
graph LR; A[01-01-2021:  
start of data,  
pre-vaccine  
rollout] --> B[RidgeCV  
Train: 0.891 Test: 0.889]; A --> C[LassoCV  
Train: 0.884 Test: 0.887]; A --> D[KNN Regressor  
Train: 0.782 Test: 0.871]; A --> E[RandomForest  
Train: 0.855 Test: 0.968];
```

RidgeCV
Train: 0.891 Test: 0.889

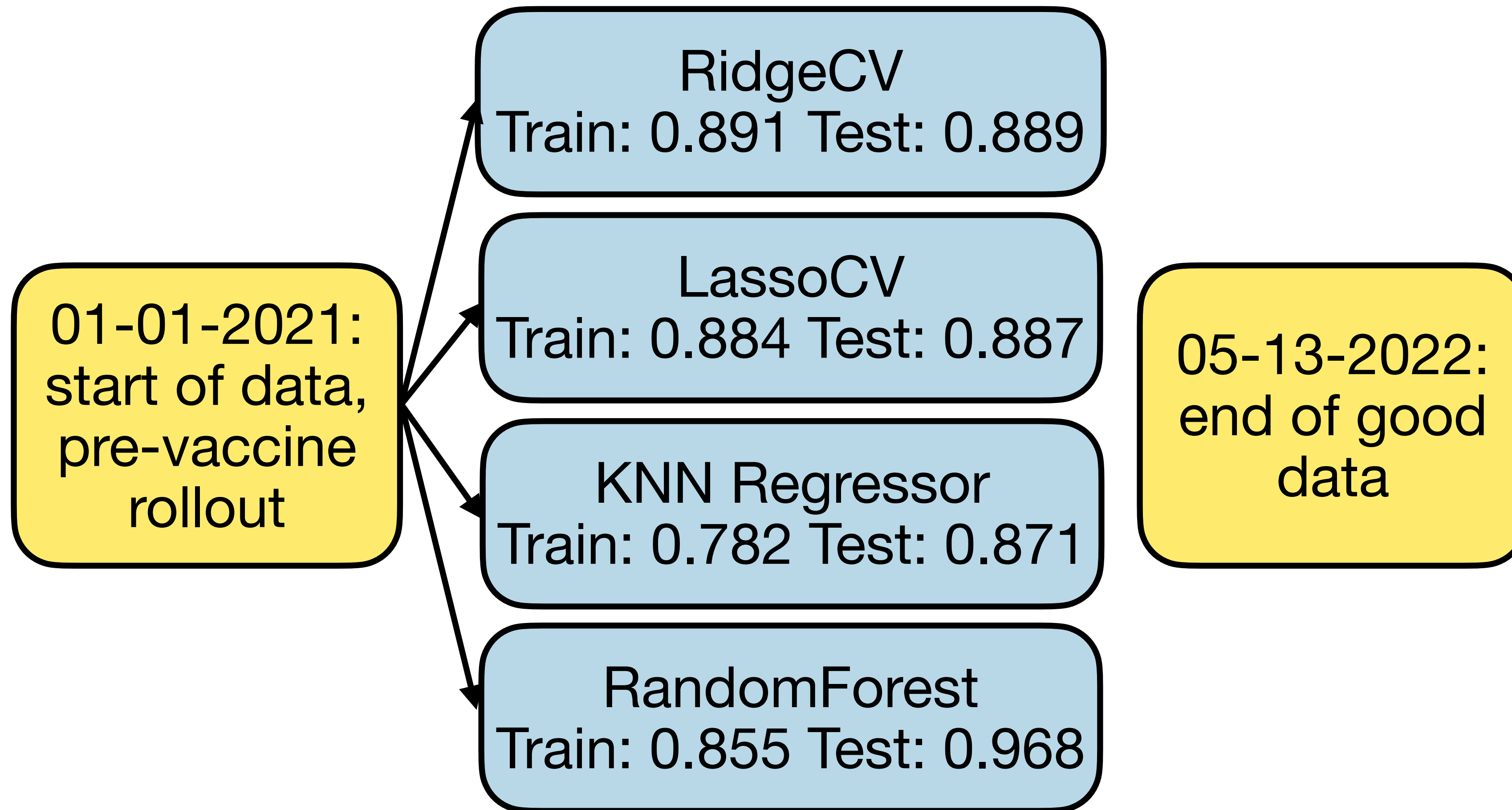
LassoCV
Train: 0.884 Test: 0.887

KNN Regressor
Train: 0.782 Test: 0.871

RandomForest
Train: 0.855 Test: 0.968

Model the Spread

Correct/Substantiate other Sources



Model the Spread

Correct/Substantiate other Sources

01-01-2021:
start of data,
pre-vaccine
rollout

RidgeCV
Train: 0.891 Test: 0.889

LassoCV
Train: 0.884 Test: 0.887

KNN Regressor
Train: 0.782 Test: 0.871

RandomForest
Train: 0.855 Test: 0.968

05-13-2022:
end of good
data

RidgeCV
Train: 0.959 Test: 0.958

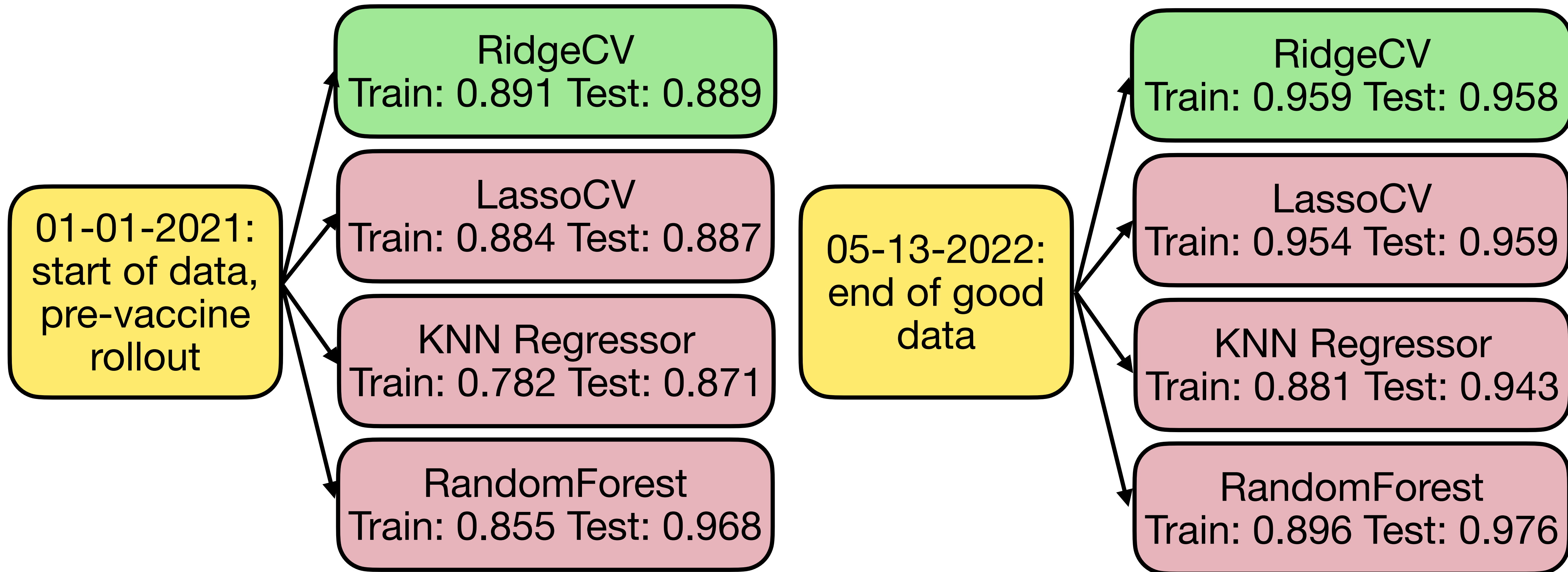
LassoCV
Train: 0.954 Test: 0.959

KNN Regressor
Train: 0.881 Test: 0.943

RandomForest
Train: 0.896 Test: 0.976

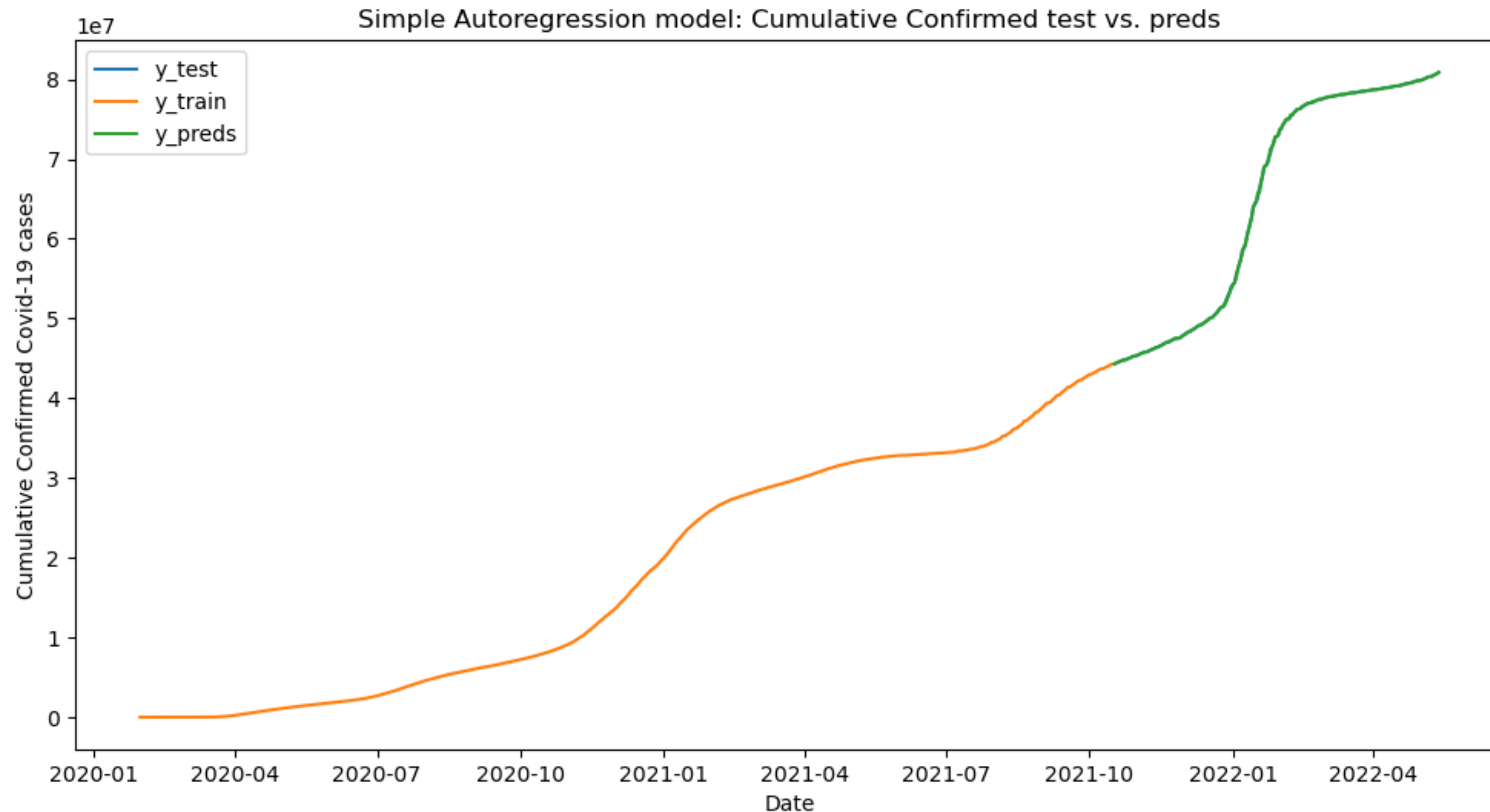
Model the Spread

Correct/Substantiate other Sources



Forecast

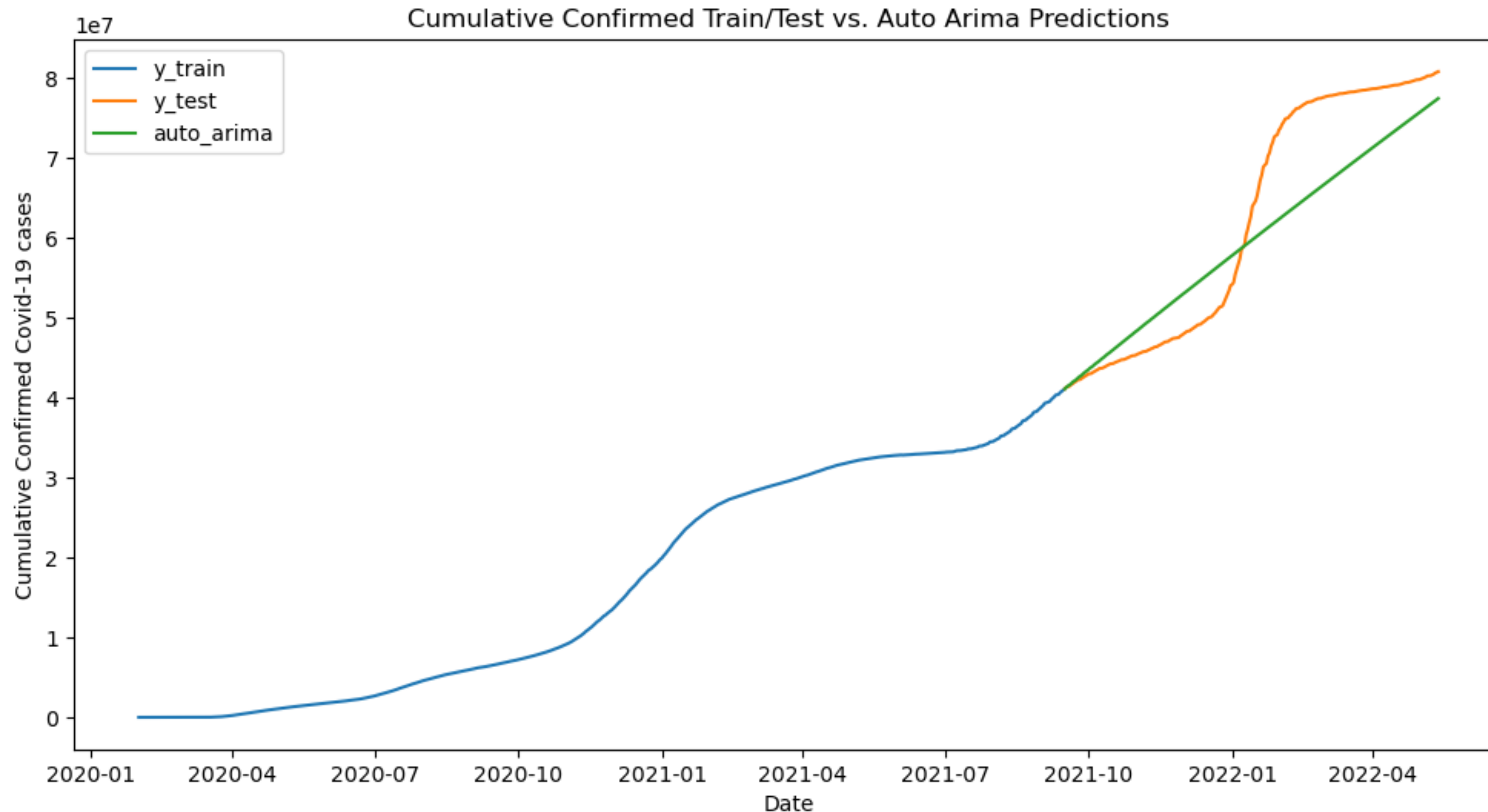
Proactive Planning



- Simple 10 lag Autoregression
- Short Term Planning

Forecast

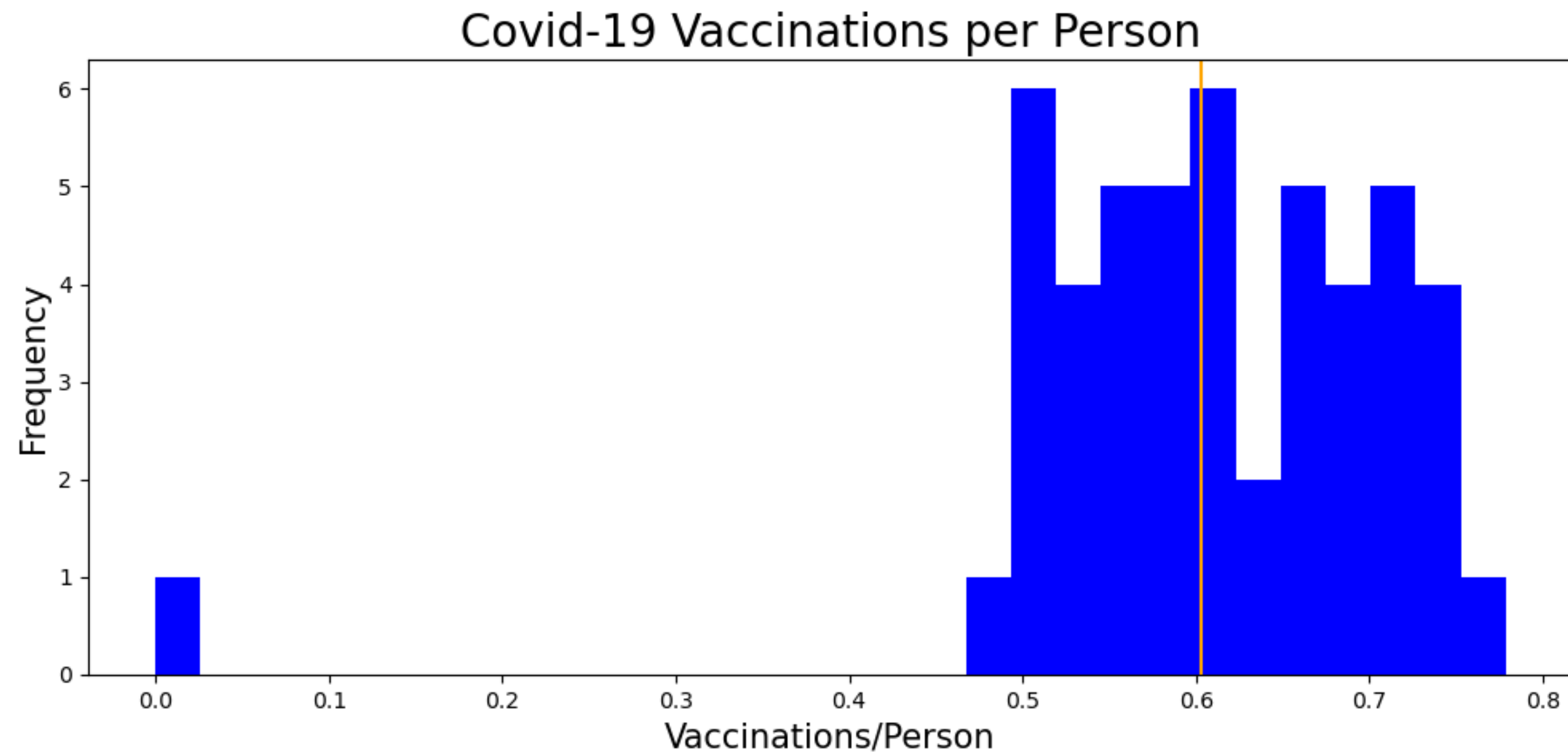
Proactive Planning



- auto-arima
optimized ARIMA
- Longer term
planning
- Non-seasonal,
non-stationary
dataset

Why Does it Matter

Regaining Trust



Why Does it Matter

Regaining Trust

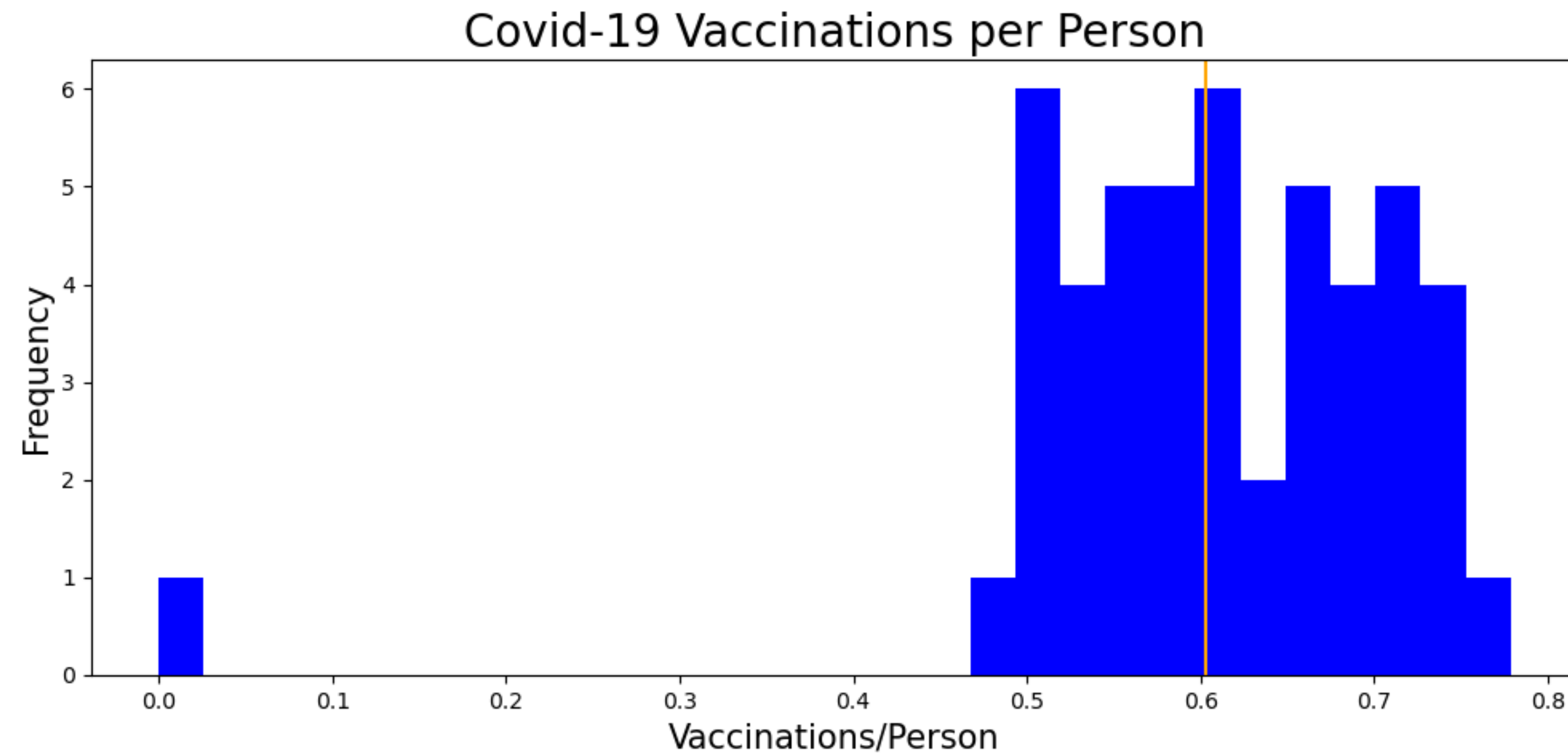
CT 0.78

MD 0.75

NY 0.75

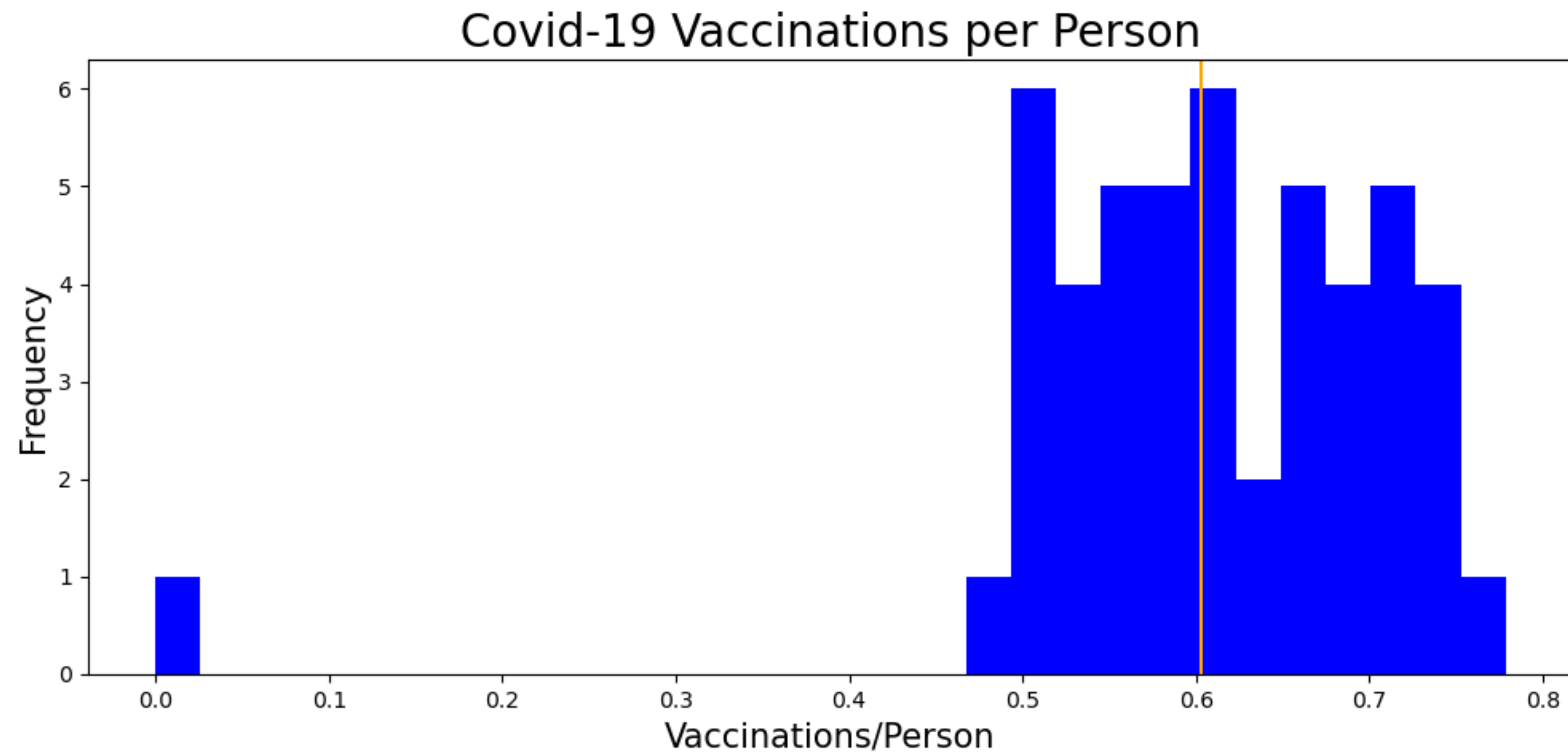
RI 0.74

WA 0.72



Why Does it Matter

Regaining Trust



CT 0.78

MD 0.75

NY 0.75

RI 0.74

WA 0.72

MS 0.51

GA 0.50

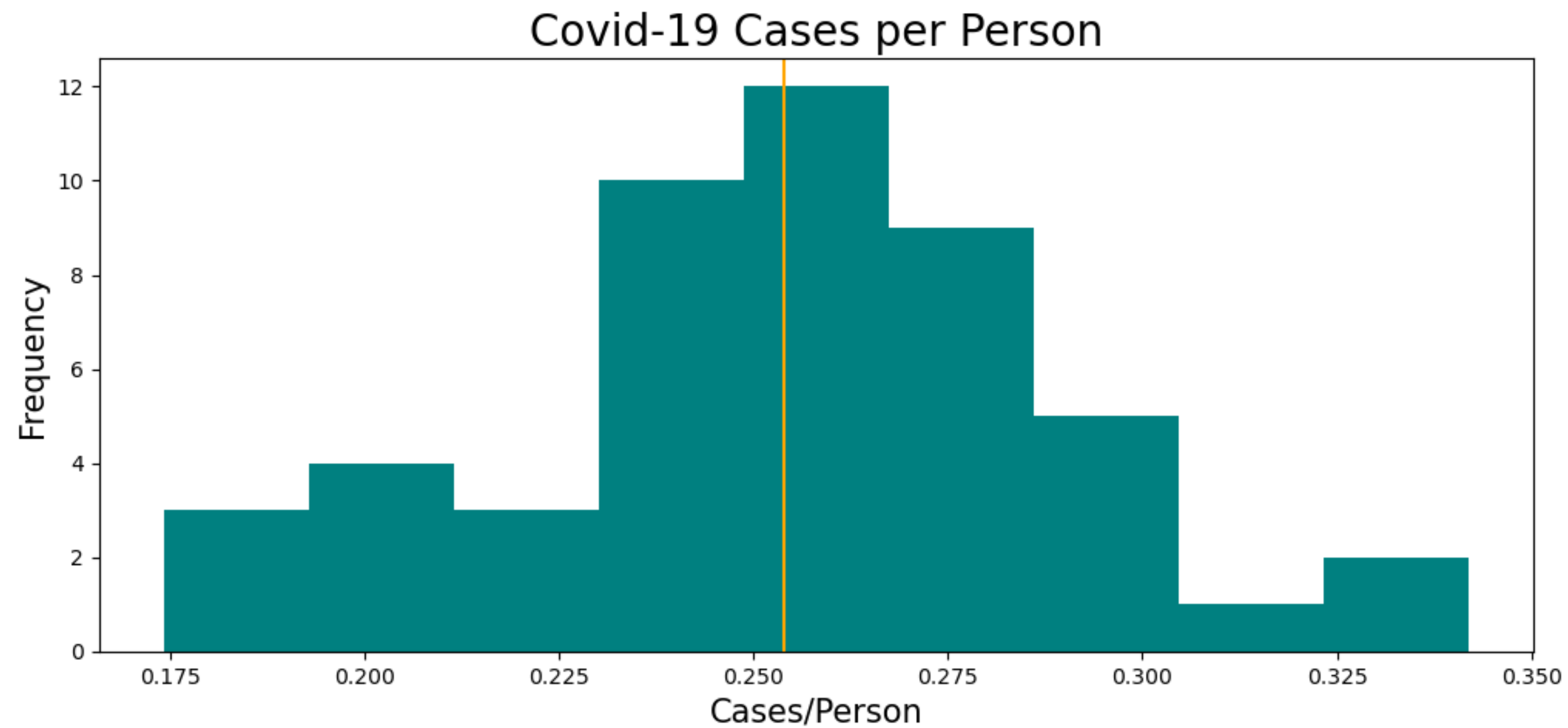
AR 0.50

WY 0.50

AL 0.48

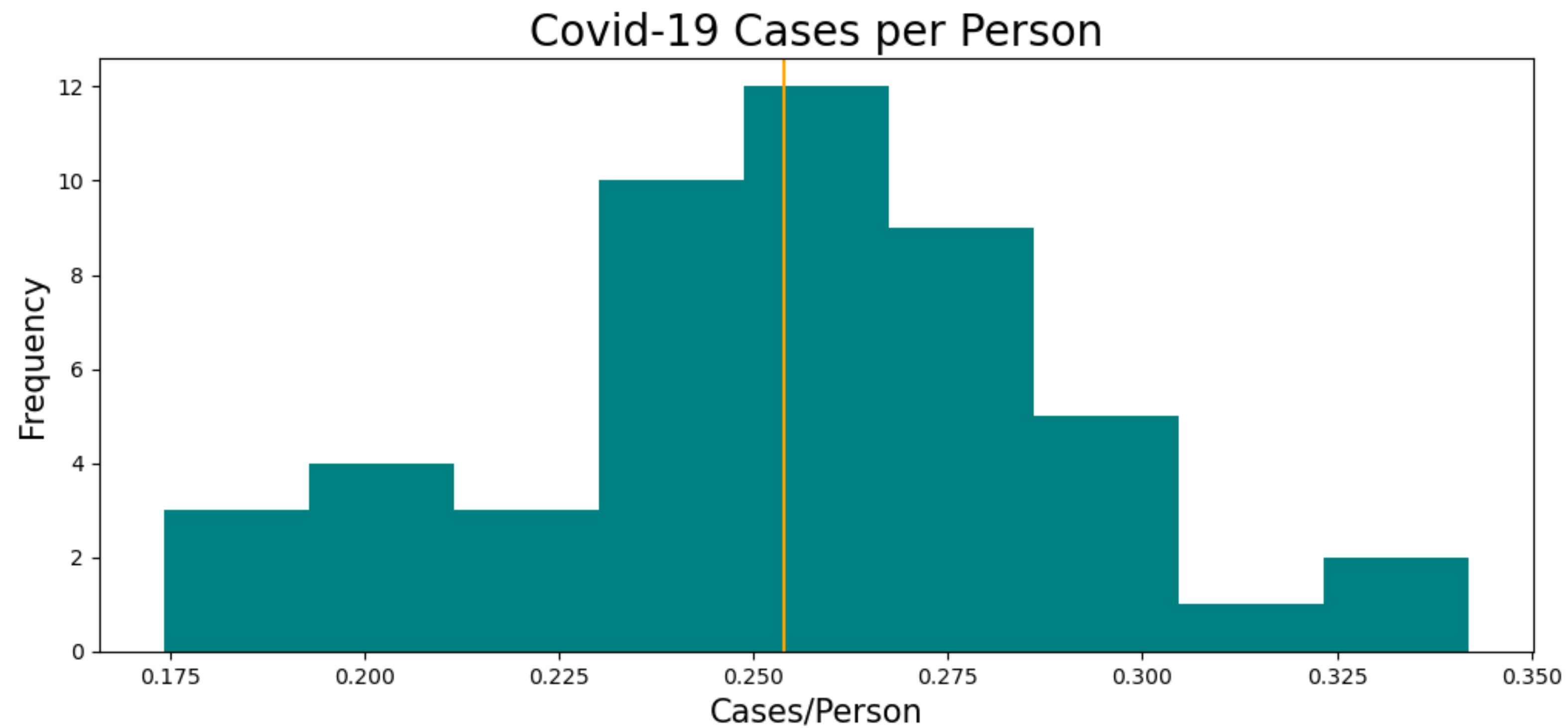
Why Does it Matter

Regaining Trust



Why Does it Matter

Regaining Trust



AK 0.34

RI 0.33

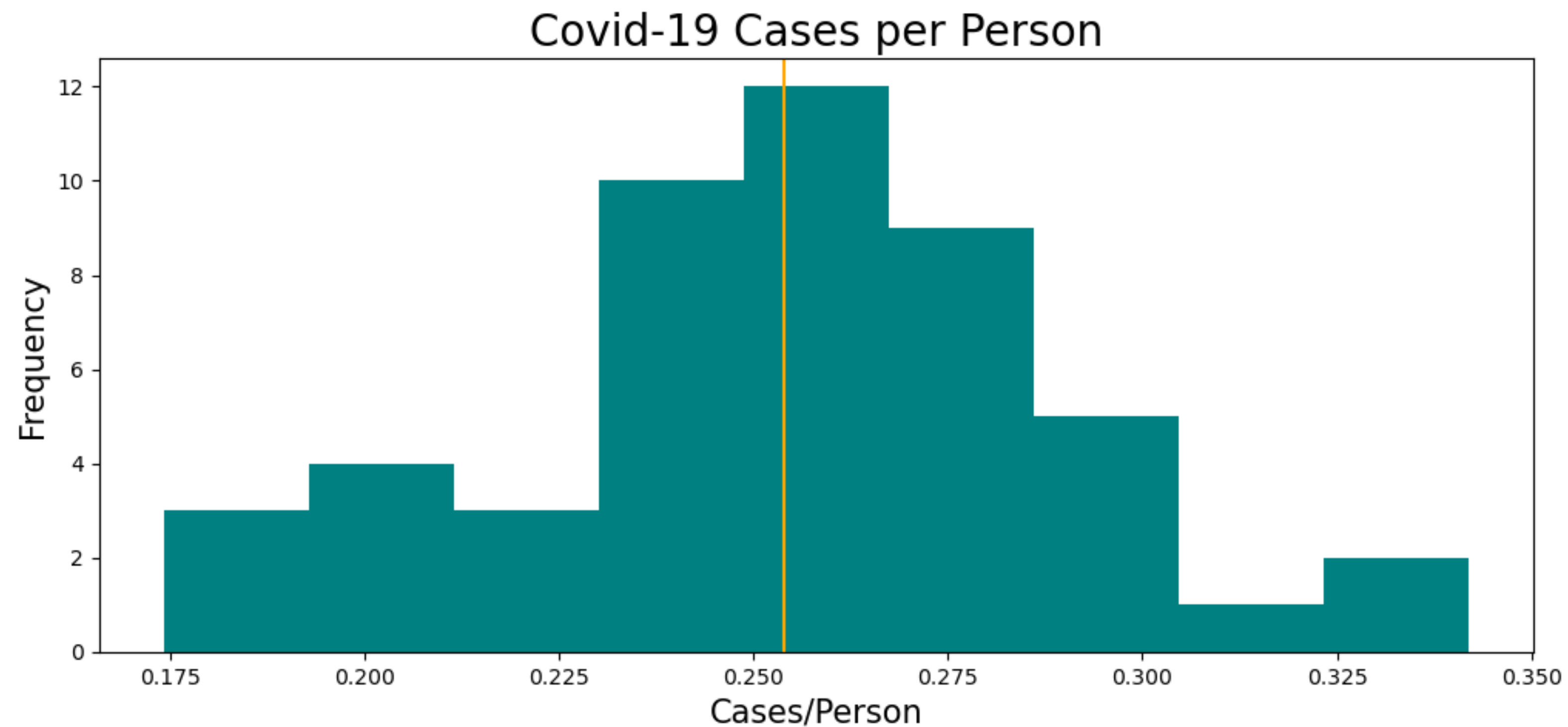
ND 0.32

KY 0.30

UT 0.30

Why Does it Matter

Regaining Trust



AK 0.34

RI 0.33

ND 0.32

KY 0.30

UT 0.30

WA 0.21

VA 0.20

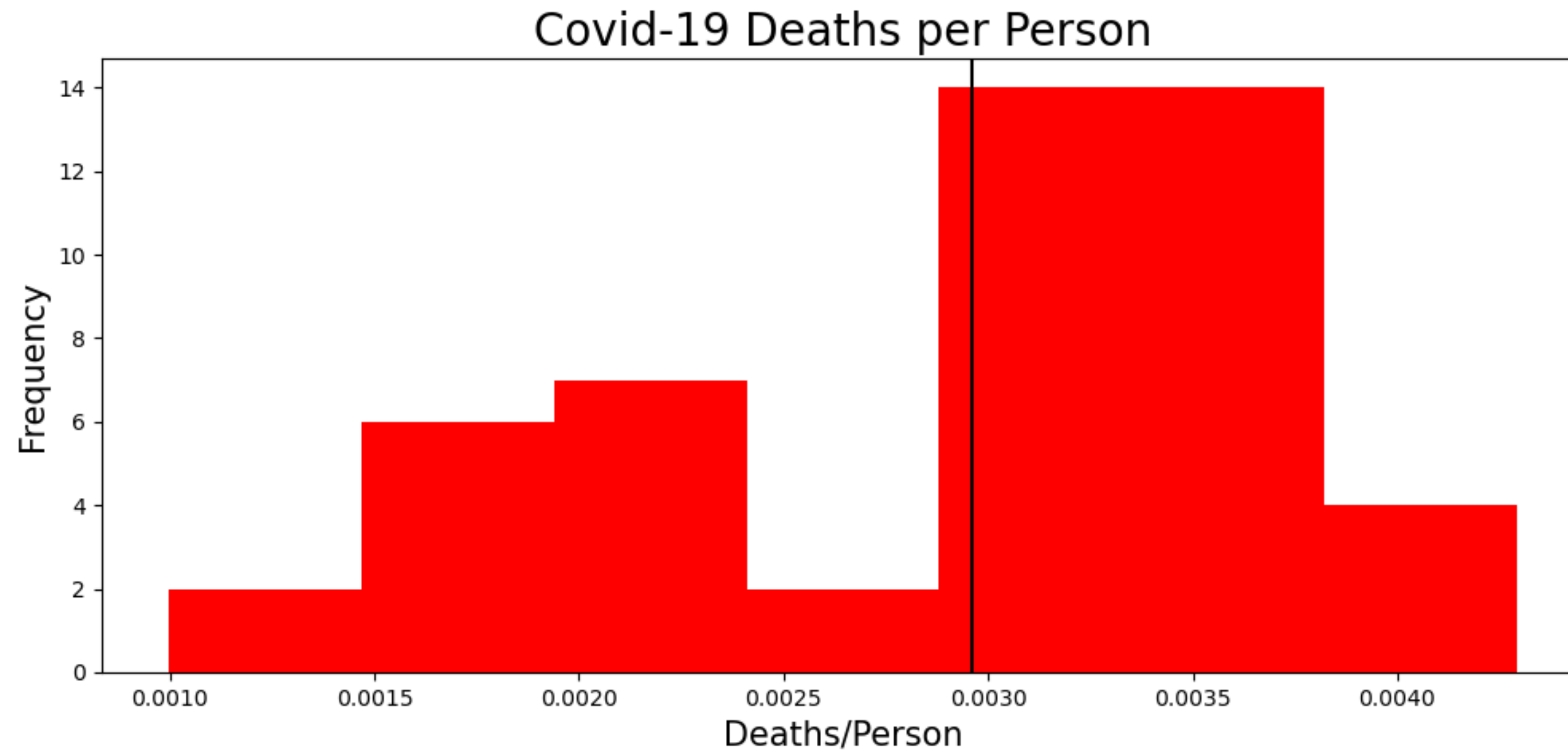
VT 0.20

OR 0.18

MD 0.17

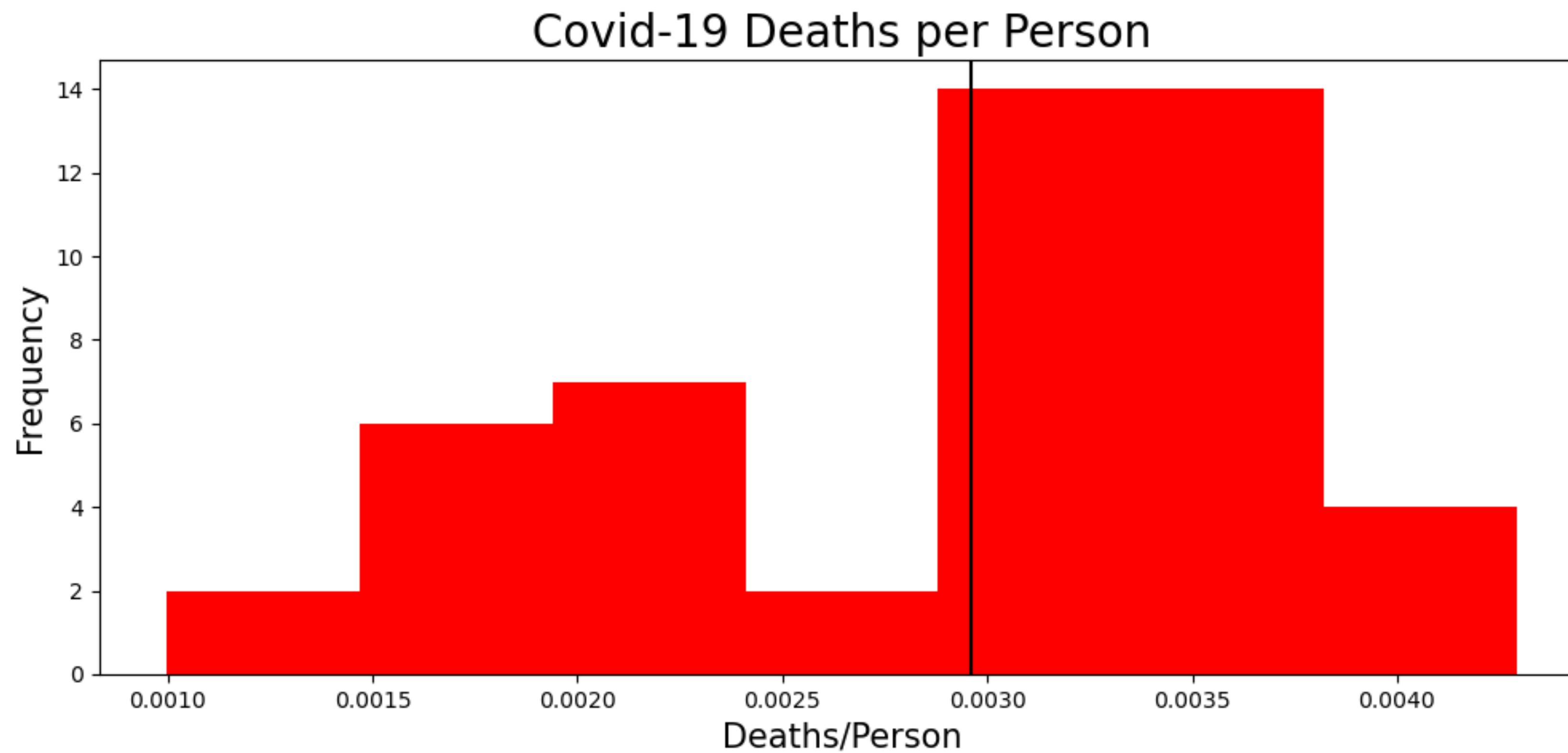
Why Does it Matter

Regaining Trust



Why Does it Matter

Regaining Trust



AZ 0.0043

MS 0.0042

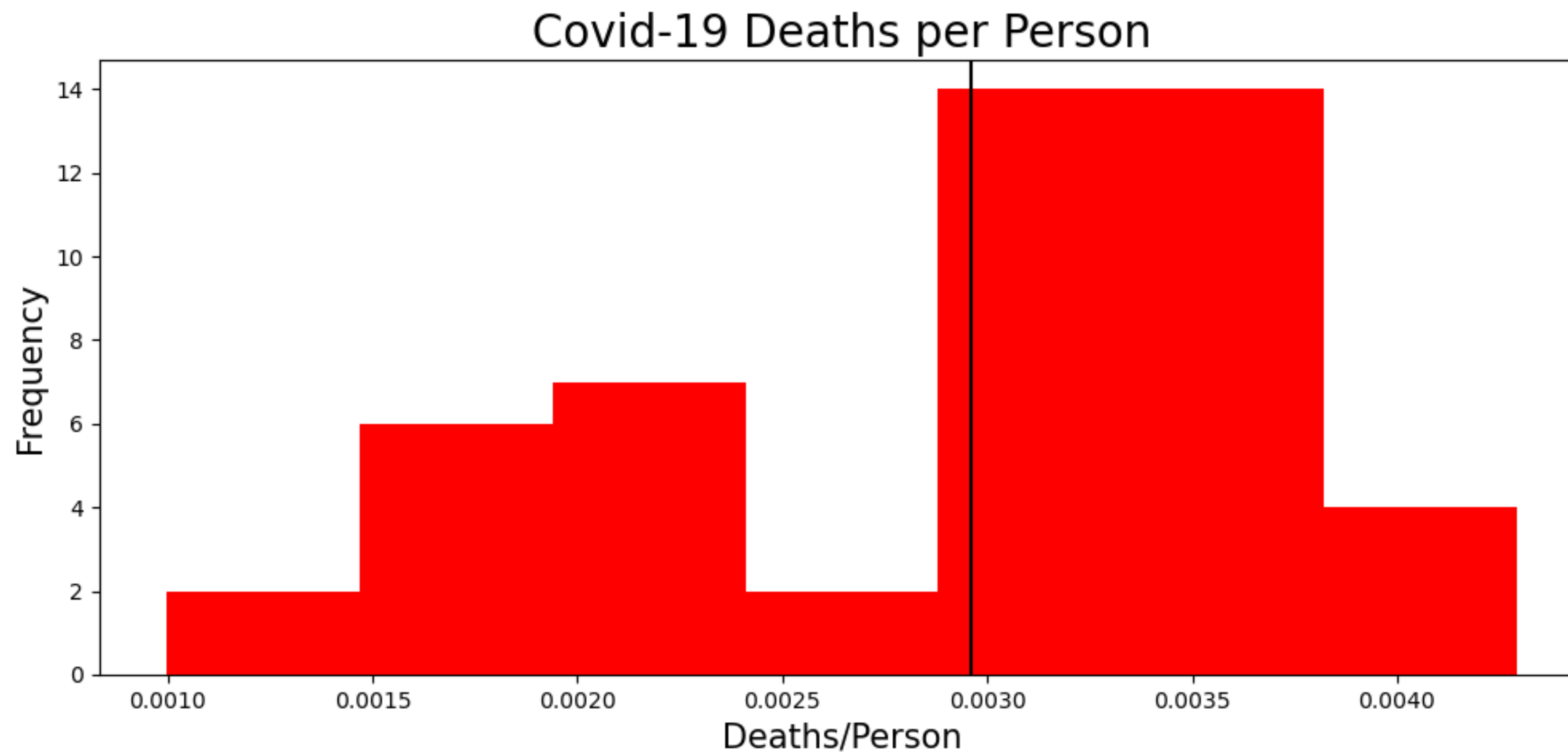
AL 0.0040

TN 0.0038

AR 0.0038

Why Does it Matter

Regaining Trust



AZ 0.0043

MS 0.0042

AL 0.0040

TN 0.0038

AR 0.0038

OR 0.0018

WA 0.0017

AK 0.0017

UT 0.0015

VT 0.0010

Conclusion

- Identify Factors
- Model for correction/
substantiation
- Forecasting
- Recommend dashboard to
compile tools in one place

