FEMA Disaster Fund Forecasting

Presented By: Josh, Yetti and Caleb

Background info

The data that we acquired for analysis came from the **Federal Emergency Management Agency,** (FEMA).

The dataset provided detailed information about past disasters in the United States and its affiliated territories dating back to August of 1998.

For the purposes of this project, the data was broken down into a 10 year set (2011-2020) and two different 5 year sets (2016-2020): one with COVID-19 data intact and one with COVID-19 data removed to provide a best and worst case scenario for 2020. The 10 year dataset has COVID-19 data included as well, to compare the forecasted prediction against one with more training data.

[source].

Problem Statement

We are working for a client that is helping FEMA prepare their disaster relief aid for the next year. We have been tasked to present forecasts of funds needed in future disaster recovery scenarios, such that all future funding can be optimized.

Based on past budgeting and past natural disasters , can we optimize a model to predict the future allocated budget for future disasters [in 2021]?

Processing the Data

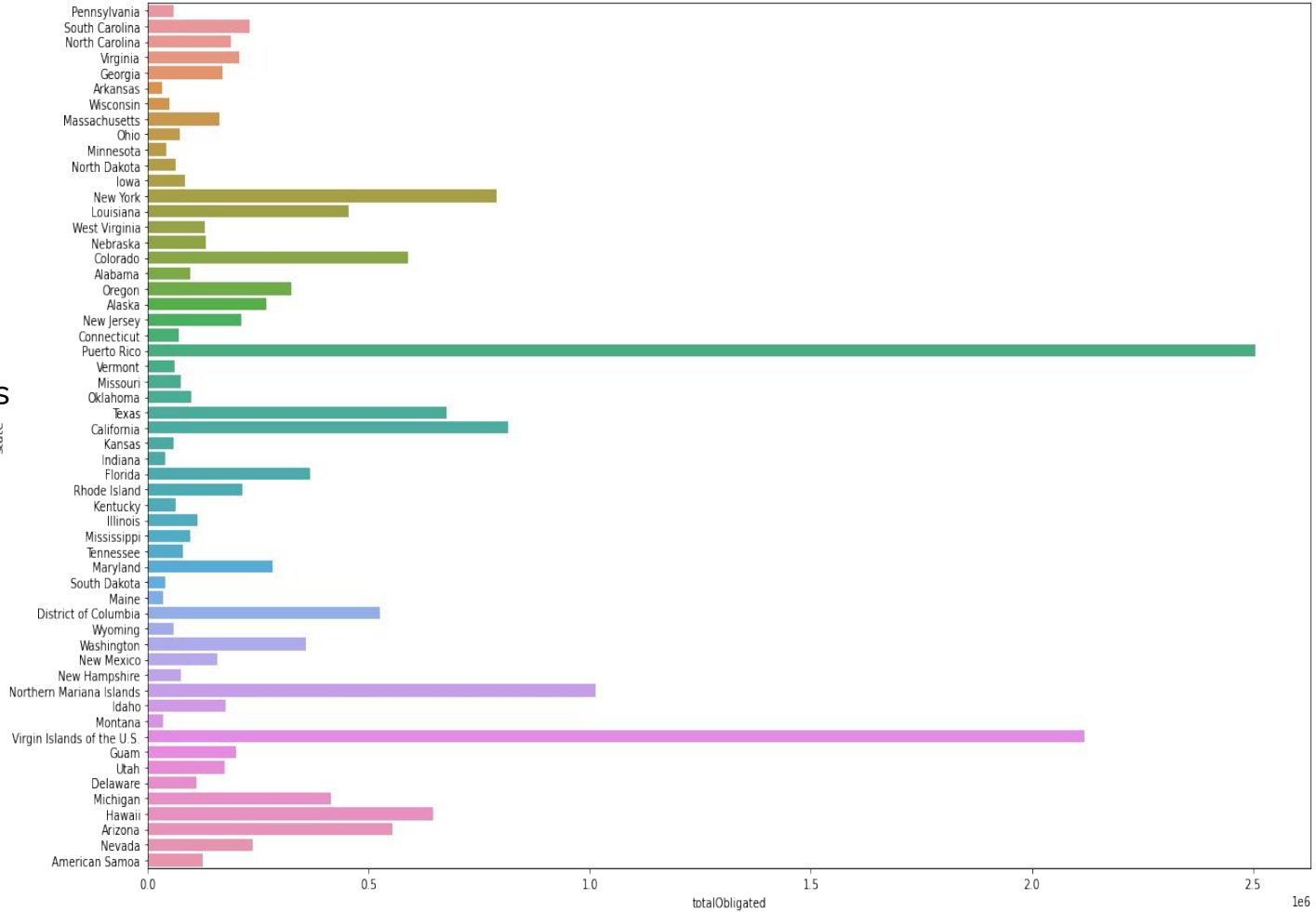
Light processing and cleaning of our data was done in a separate notebook for the entire dataset (nearly 800,000 rows). Below are some methods we used for our model:

- Set date column to index
- Converted the date to “datetime”
- Removed unnecessary features
- Kept and removed Covid data from different datasets

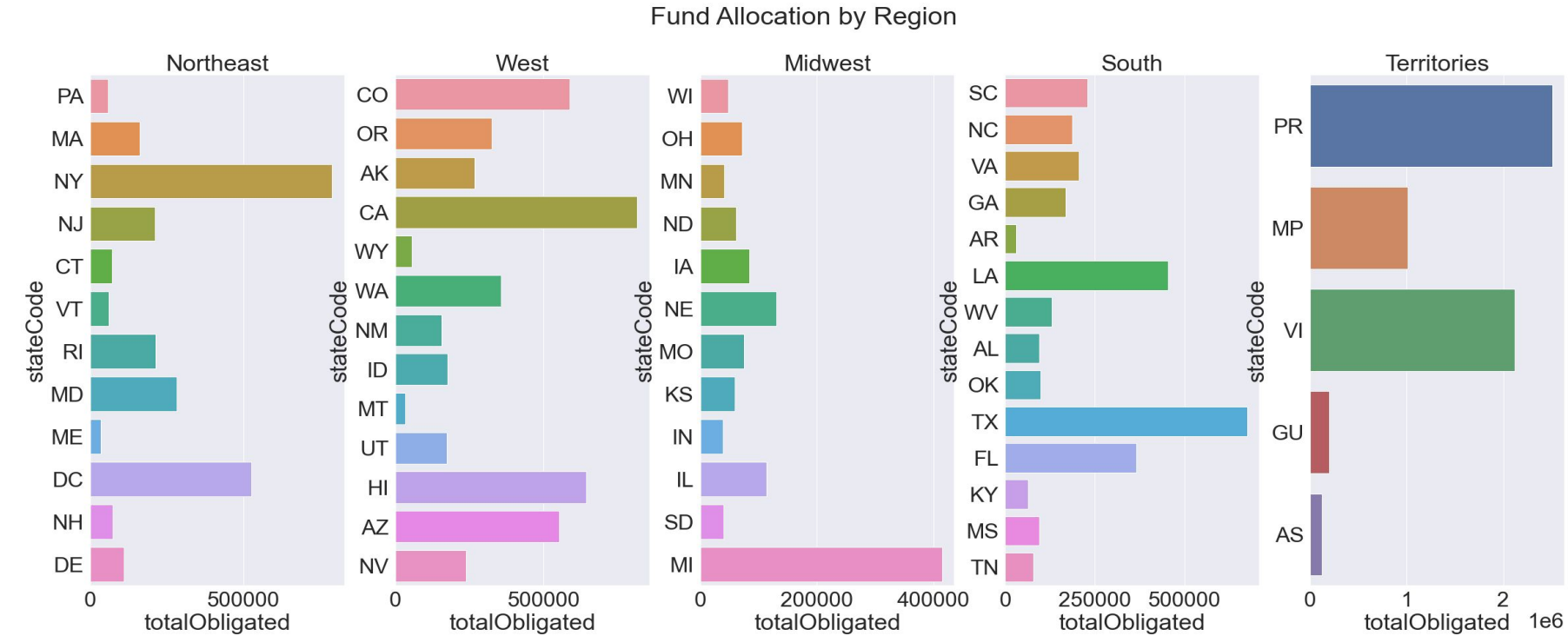
EDA

Funds allocated per State and/or Territory from 2011-2020.

- Puerto Rico and the Virgin Islands are the top 2 FEMA recipients.



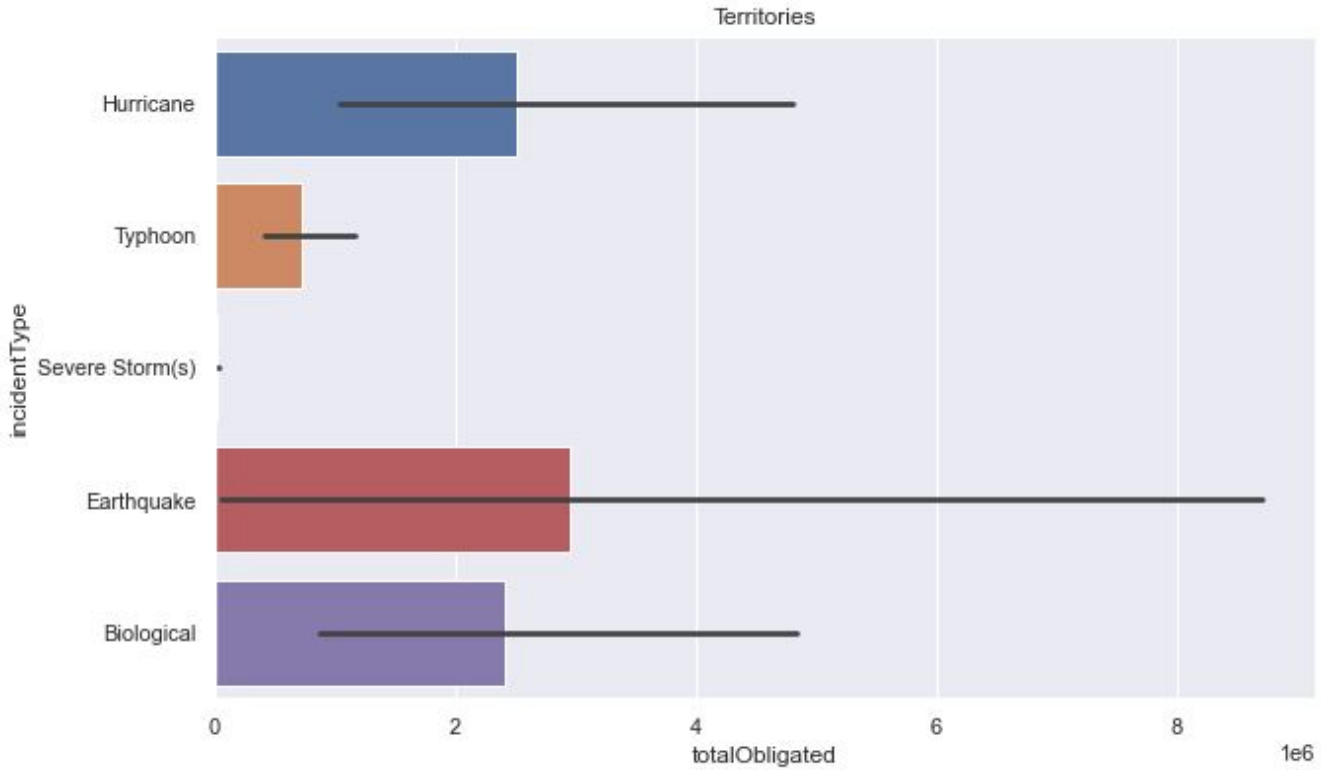
EDA


Fund Allocation by Region
Northeast
West
Midwest
South
Territories
stateCode
totalObligated
PA
MA
NY
NJ
CT
VT
RI
MD
ME
DC
NH
DE
CO
OR
AK
CA
WY
WA
NM
ID
MT
UT
HI
AZ
NV
WI
OH
MN
ND
IA
NE
MO
KS
IN
IL
SD
MI
SC
NC
VA
GA
AR
LA
WV
AL
OK
TX
FL
KY
MS
TN
PR
MP
VI
GU
AS
0
500000
200000
400000
250000
1
2
1e6

EDA

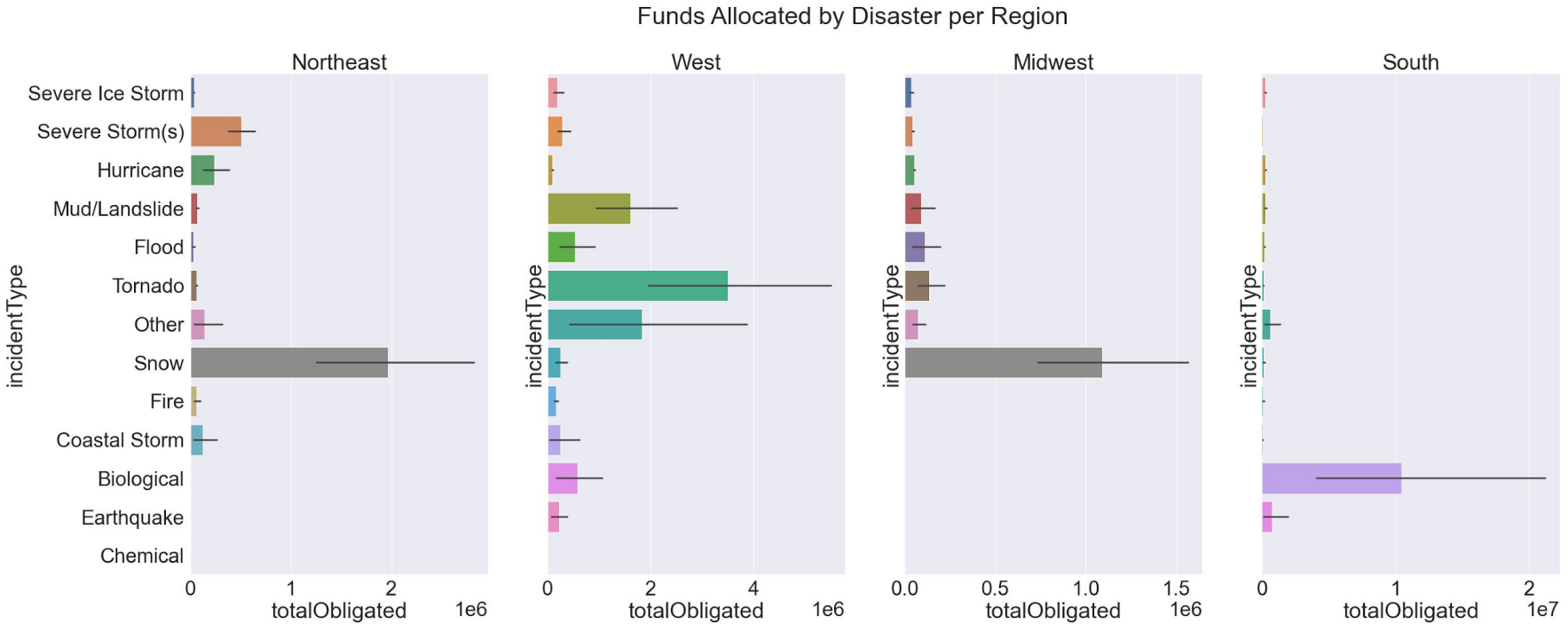
Visual of funds allocated per disaster per region.

We can see why territories is allocated so much money on average.



EDA

The northeast and midwest had highest allocations for snow, and the west had the highest fund allocation for tornados. The south had the highest fund allocation for biological events.



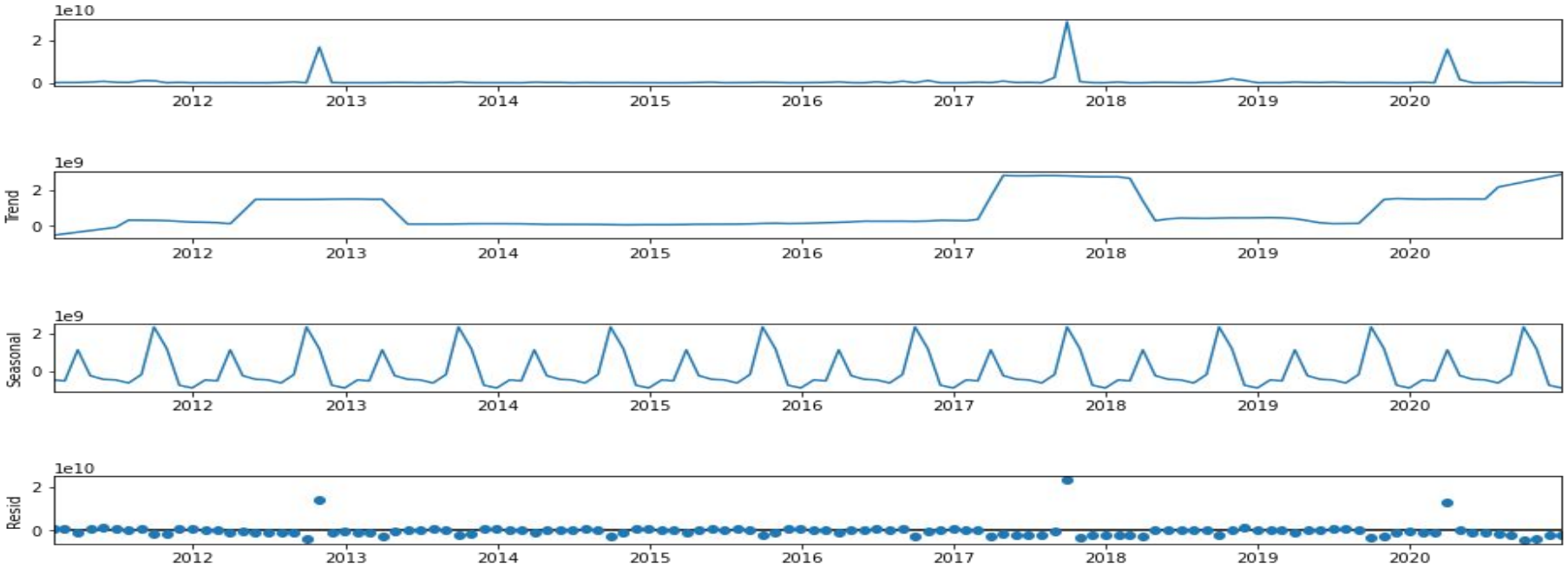
Modeling Process

We used ARIMA (Autoregressive Integrated Moving Average) for our time series modeling.

1. Ran our results through **adfuller.**
2. Plotted our acf (auto-correlation function) and pacf (partial auto correlation)
3. Split our data into train- test-split
4. Integrated a manual gridsearch to run our ARIMA
5. Plotted our predictions
6. Ran ARIMA again
7. Calculated the forecasted budget for 2021

Modeling

Our baseline dataset was the 10 year covid data. We resampled the data by month. Below is our seasonal composition data for this dataset.



Modeling

From the previous slide we see no increasing trend in the data. We see outliers in 2013, 2017 and in 2020.

- 2013 - Hurricane Sandy
- 2017 - Hurricane Maria, Irma, Harvey
- 2020 - COVID!

Obligated funds are relatively stationary outside of those 3 events

Modeling - Minimized AIC (Akaike Information Criteria)

The AIC is a score that is used to determine the best model in a set of models. Lower = better.

Model parameters for all:

- Train-Test-Split
- Test Size = 0.2
- Random State = 42
- Shuffle = False

- Best AIC score for **10 year + Covid**: 4447.31 (ARIMA (0,1,1))
- Best AIC score for **5 year + Covid**: 2223.68 (ARIMA (0,1,1))
- Best AIC score for **5 year no Covid**: 2223.68 (ARIMA (0,1,1))

ADFuller Score

The ADFuller score helps us check for stationarity. Low p-value (< 0.05) is good. Each model had stationarity (as expected)

10_year = Stationarity

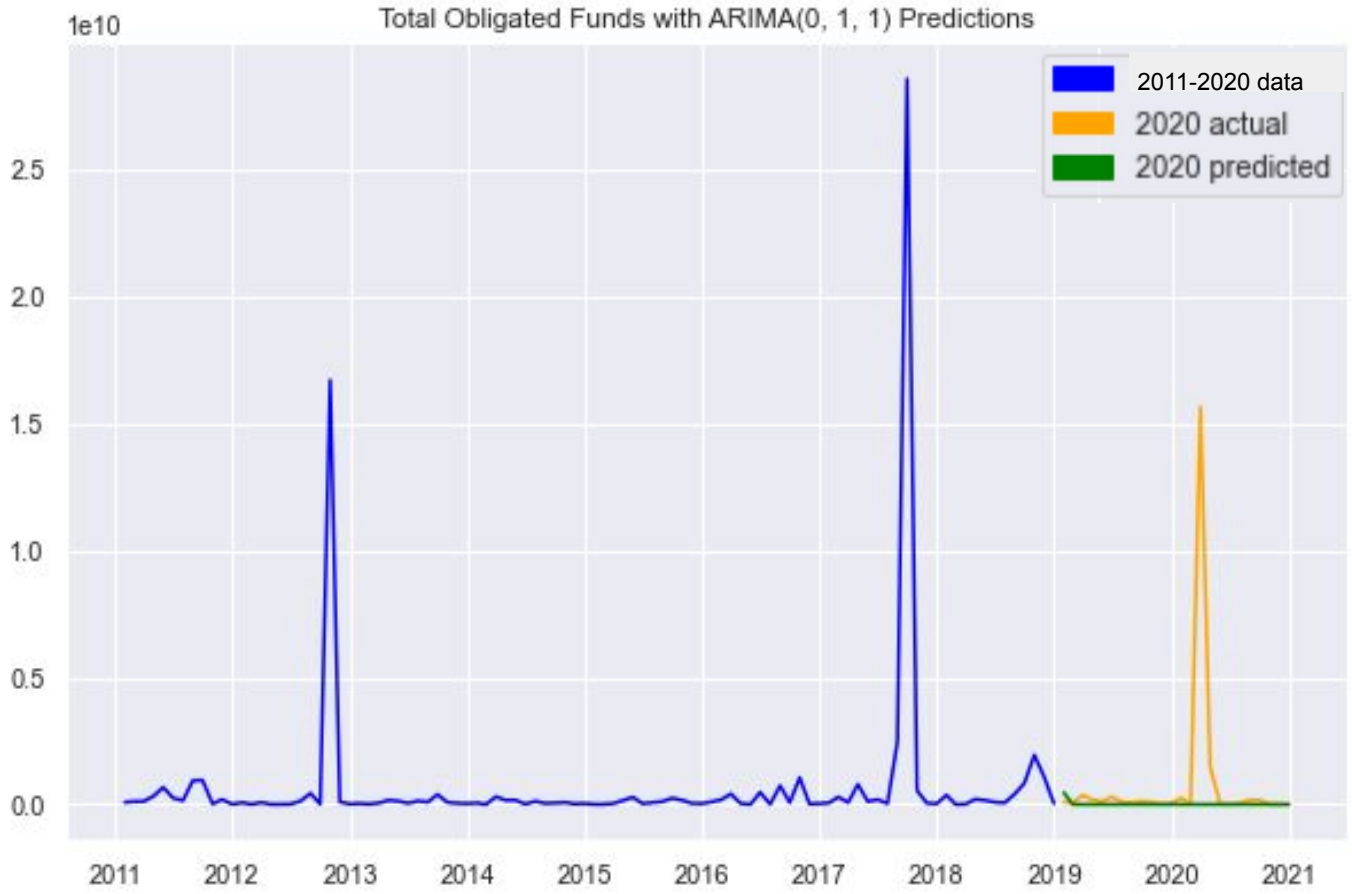
```
ADF Statistic: -10.356153143938634
p-value: 2.4636685565508843e-18
```

5_year (No Covid) = Stationarity

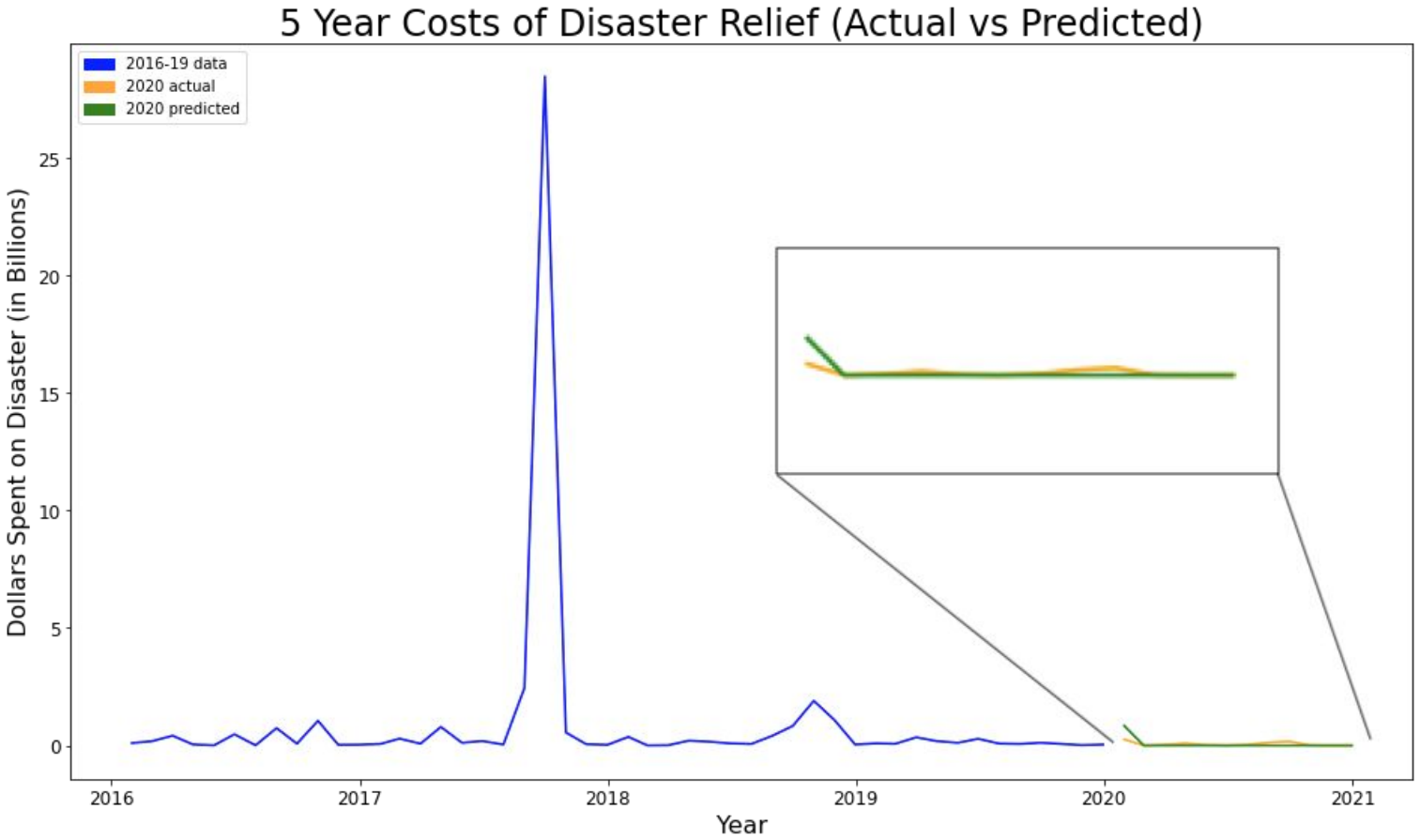
```
Test Statistic   -7.014470e+00
p-value           6.788640e-10
```

5_year + Covid) = Stationarity

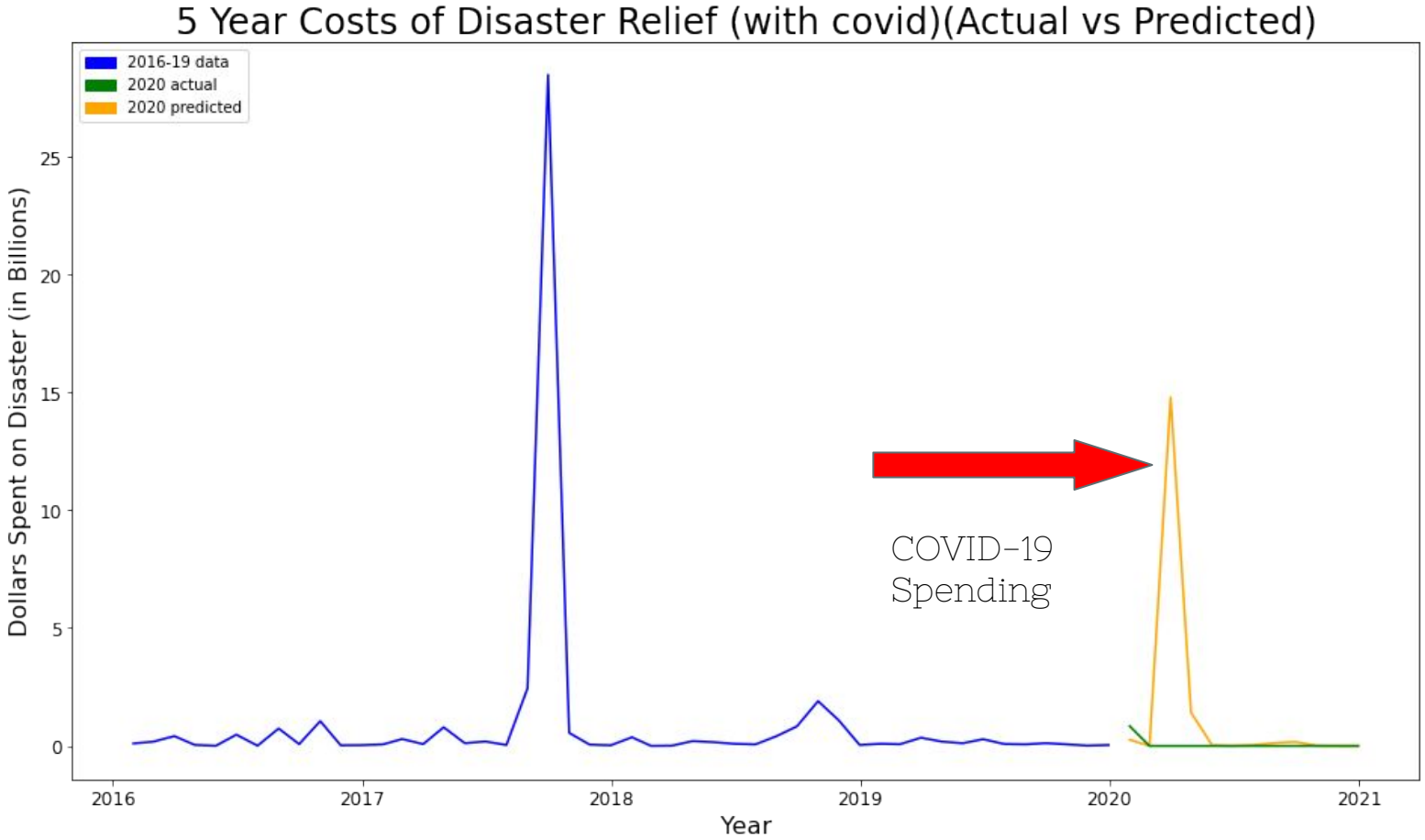
```
Test Statistic   -7.168433e+00
p-value           2.845882e-10
```



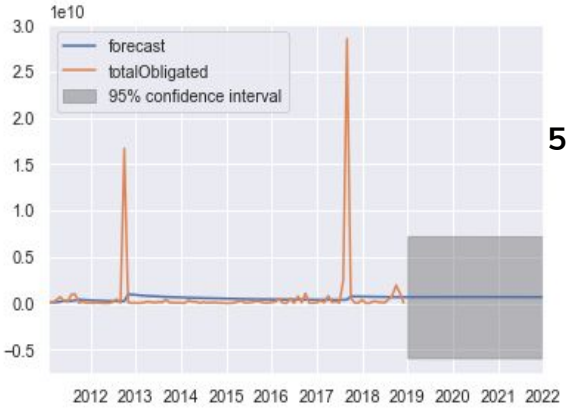
Arima Predictions Plot with 10 Year Data



Arima Predictions Plot with 5 Year Data-No Covid

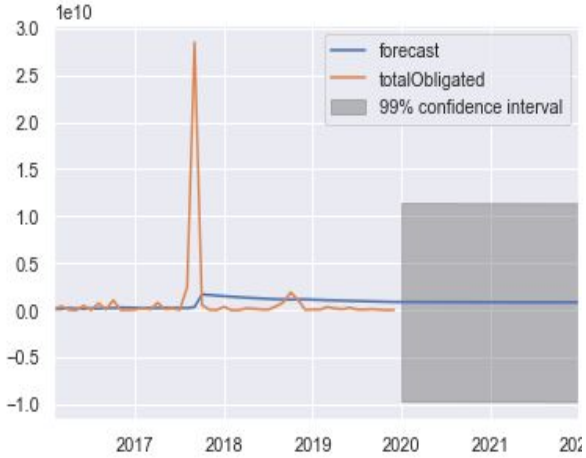


Arima Predictions Plot with 5 Year Data + Covid

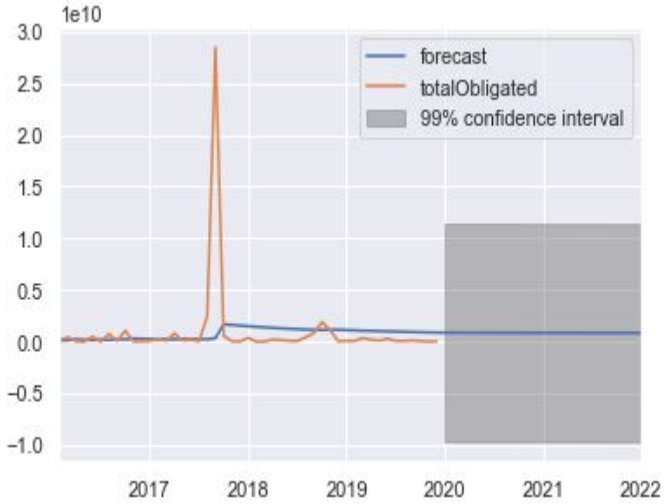


10 year dataset - 1 year forward.

5 year dataset (No Covid) - 1 year forward.



5 year dataset + Covid - 1 year forward.



- Our forecasts into the future for FEMA funding appear relatively similar, with the exception of some minor deviations.

Forecasted 2021 Plots

Results

Our forecasted 2021 budgets for each model:

- 10-year-model: **$8.08B**
- 5-year-no Covid model: **$10.39B**
- 5-year+Covid model: **$10.39B**
  - Covid had very little impact on the budget outcome for the 5 year models.

Yes, we were able to optimize a model that predicts future FEMA fund allocations for 2021, however more optimization could possibly be achieved with the recommendations below.

Conclusions and Recommendations

- The 10 year model projects the lowest needed budget at approximately $8.08B while keeping COVID-19 intact. The shorter models of 5 years both with and without COVID-19 data shows a projected budget of $10.39B.
- Neither 5 year model, presents a “worst case scenario” wherein the nation is impacted by some unforeseen factor. Modeling different compounded disasters could give a better idea of what the budget should look like.
- The comparison between the 10 year and 5 year models under normal circumstances could be an indicator of the effect of climate change, the rate of natural disasters and the associated costs of recovery.
- The final recommendation would be to budget according to the 5 year model, but prepare an emergency plan, in the event that a once-in-a-lifetime event becomes more than that and as the frequency of large scale disasters increases.


Questions?