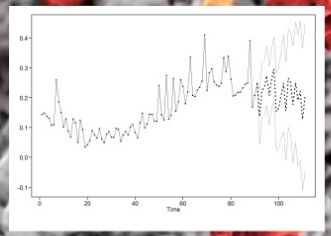
US AND ARIZONA COVID-19 TIME SERIES ANALYSIS AND FORECASTS

Daniel Clark and Gavin Hudgeons
Southern Methodist University



Data Collection and EDA

Univariate Analysis

Multivariate Analysis

Summary

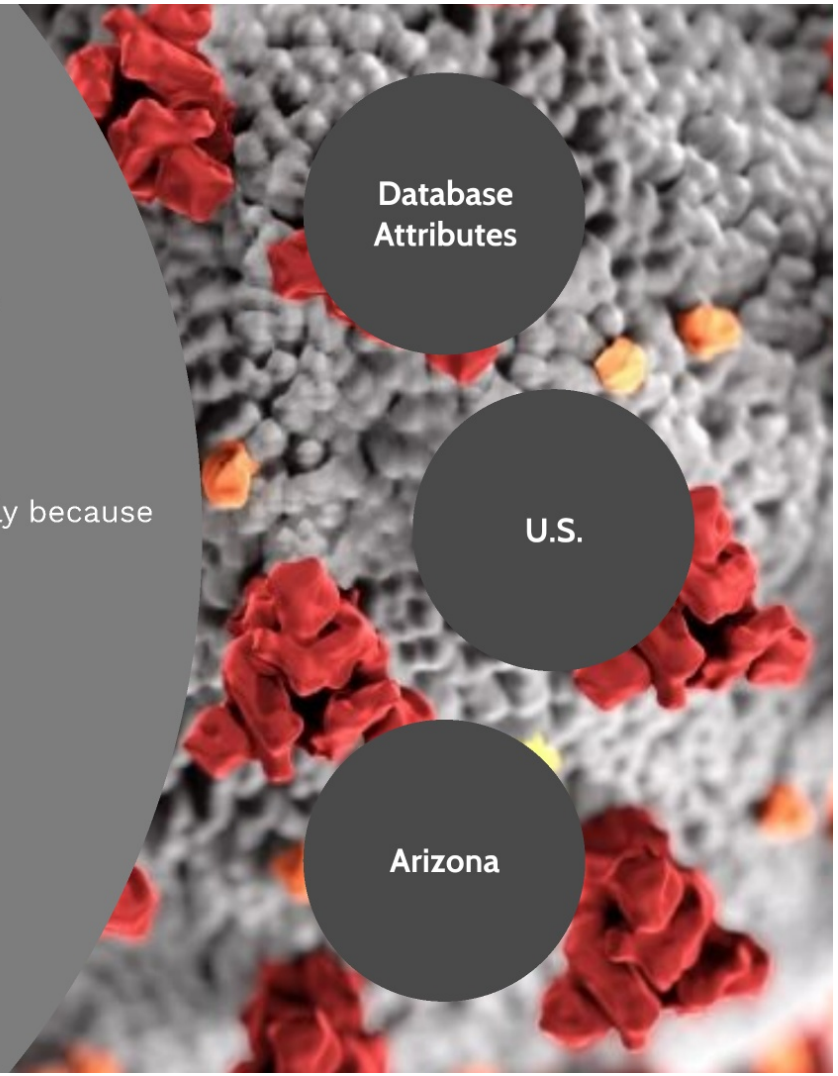
Data Collection and EDA

1. An exploratory data analysis (EDA) was completed and presented. Links to EDA presentations are here:
   - https://www.youtube.com/watch?v=uPOOcZL9o3g&feature=youtu.be
   - https://vimeo.com/443496753/6cb59d1153
2. We explored 2 Databases during the EDA:
   - John's Hopkins
   - Covid Tracking Project

In the end, we used the data from the Covid Tracking Project primarily because of the consistancy in the attributes between tables.

The COVID Tracking Project
AT The Atlantic

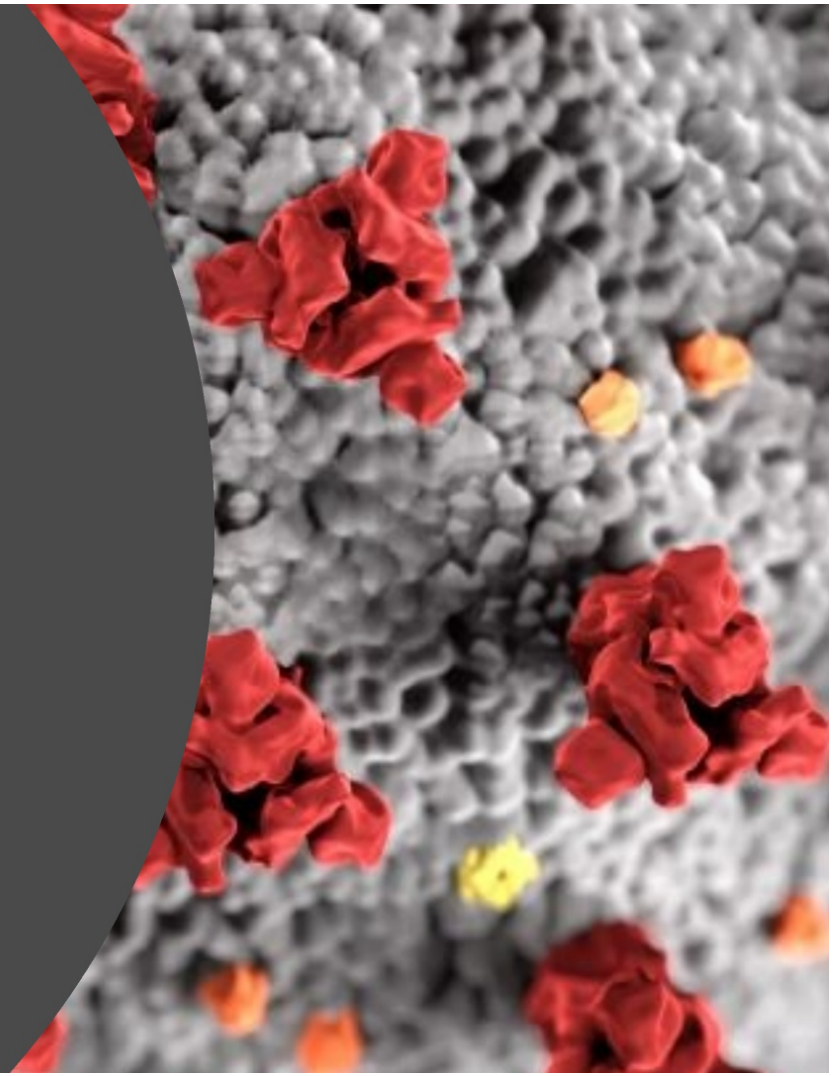
SOURCE: HTTPS://covidtracking.com/data/download
CSV Files Downloaded: U.S. Data and Arizona Data



Database Attributes

```
[1] "date"                    "state"
 [3] "positive"                "negative"
 [5] "pending"                 "hospitalizedCurrently"
 [7] "hospitalizedCumulative"  "inIcuCurrently"
 [9] "inIcuCumulative"         "onVentilatorCurrently"
[11] "onVentilatorCumulative"  "recovered"
[13] "dataQualityGrade"        "lastUpdateEt"
[15] "dateModified"            "checkTimeEt"
[17] "death"                   "hospitalized"
[19] "dateChecked"             "totalTestsViral"
[21] "positiveTestsViral"      "negativeTestsViral"
[23] "positiveCasesViral"      "deathConfirmed"
[25] "deathProbable"           "fips"
[27] "positiveIncrease"        "negativeIncrease"
[29] "total"                   "totalTestResults"
[31] "totalTestResultsIncrease" "posNeg"
[33] "deathIncrease"           "hospitalizedIncrease"
[35] "hash"                    "commercialScore"
[37] "negativeRegularScore"    "negativeScore"
[39] "positiveScore"           "score"
[41] "grade"
```

Date Range: March 4 - July 21



```
 [1] "date"                    "state"
 [3] "positive"                "negative"
 [5] "pending"                 "hospitalizedCurrently"
 [7] "hospitalizedCumulative"  "inIcuCurrently"
 [9] "inIcuCumulative"         "onVentilatorCurrently"
[11] "onVentilatorCumulative"  "recovered"
[13] "dataQualityGrade"        "lastUpdateEt"
[15] "dateModified"            "checkTimeEt"
[17] "death"                   "hospitalized"
[19] "dateChecked"             "totalTestsViral"
[21] "positiveTestsViral"      "negativeTestsViral"
[23] "positiveCasesViral"      "deathConfirmed"
[25] "deathProbable"           "fips"
[27] "positiveIncrease"        "negativeIncrease"
[29] "total"                   "totalTestResults"
[31] "totalTestResultsIncrease" "posNeg"
[33] "deathIncrease"           "hospitalizedIncrease"
[35] "hash"                    "commercialScore"
[37] "negativeRegularScore"    "negativeScore"
[39] "positiveScore"           "score"
[41] "grade"
```

Date Range: March 4 - July 21

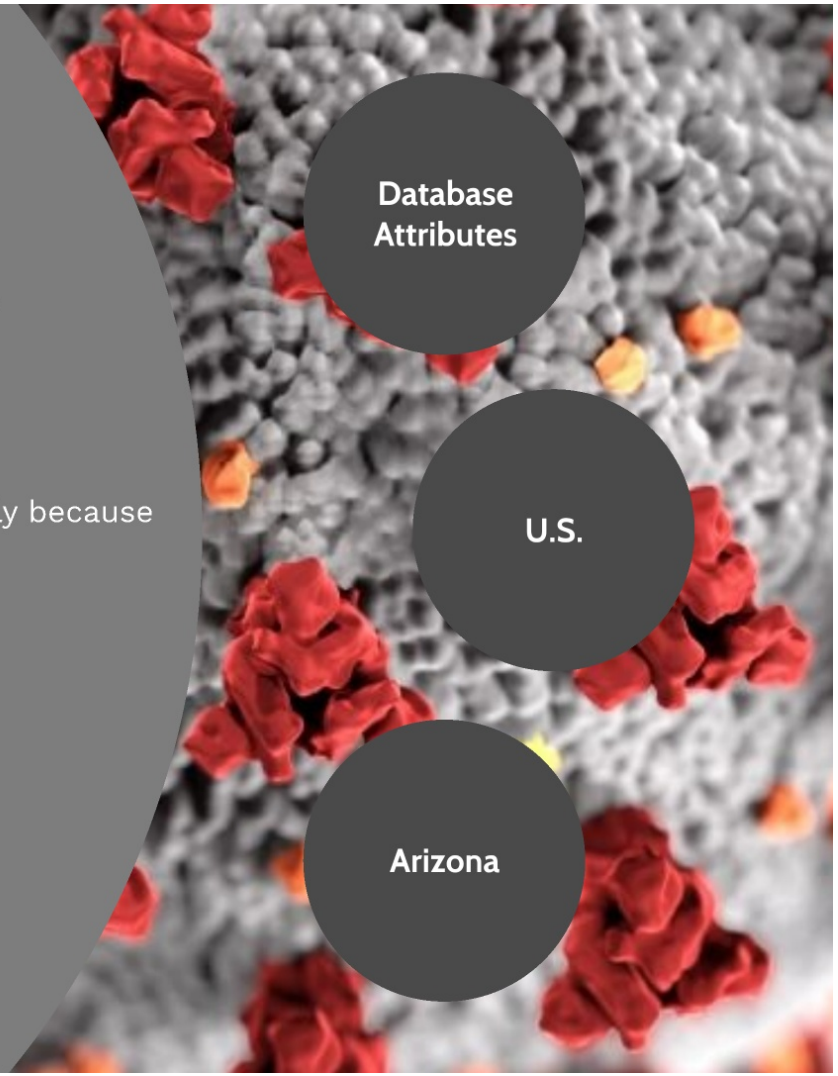
Data Collection and EDA

1. An exploratory data analysis (EDA) was completed and presented. Links to EDA presentations are here:
   - https://www.youtube.com/watch?v=uPOOcZL9o3g&feature=youtu.be
   - https://vimeo.com/443496753/6cb59d1153
2. We explored 2 Databases during the EDA:
   - John's Hopkins
   - Covid Tracking Project

In the end, we used the data from the Covid Tracking Project primarily because of the consistancy in the attributes between tables.

The COVID Tracking Project
AT The Atlantic

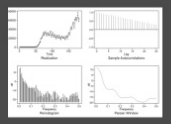
SOURCE: HTTPS://covidtracking.com/data/download
CSV Files Downloaded: U.S. Data and Arizona Data



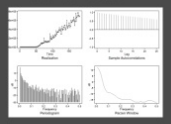
U.S. Data Overview

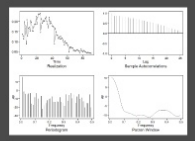
Key Features:
*Date
*Deaths
*Hospitalizations
*New Positive Tests Per Day
*New Negative Tests Per Day
*Total Test Results Per Day

Positive Increase

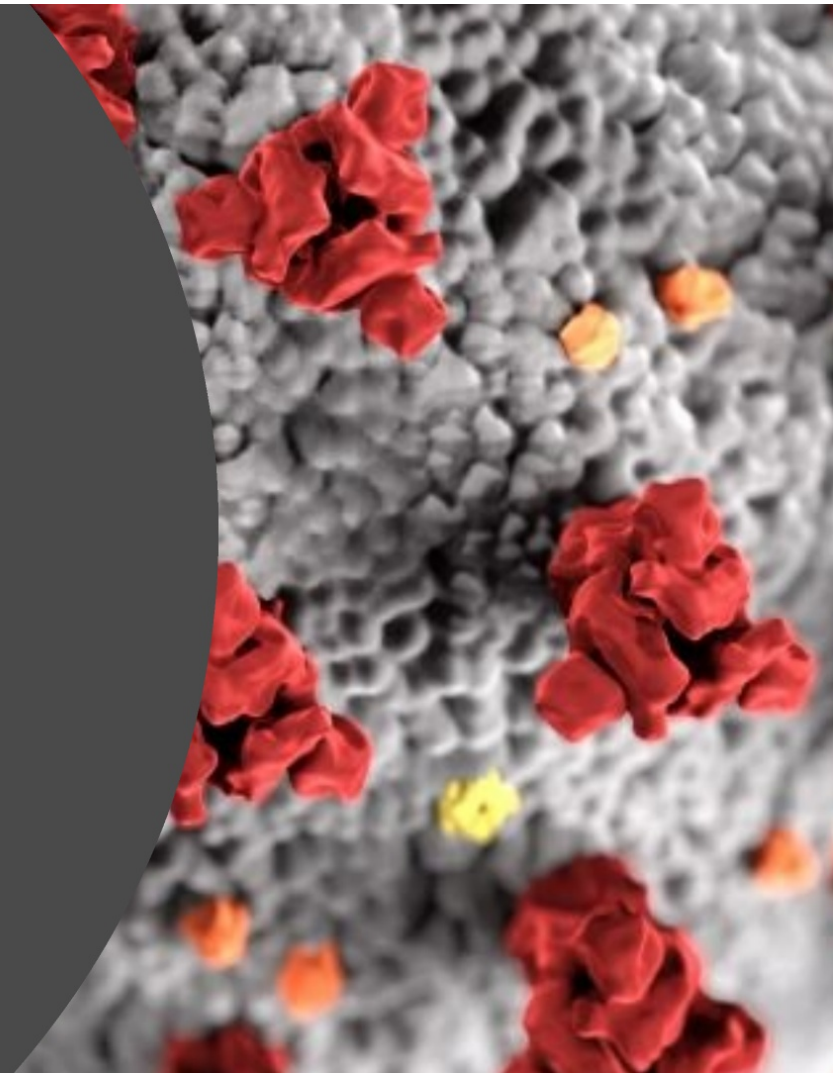


Total Test Increase



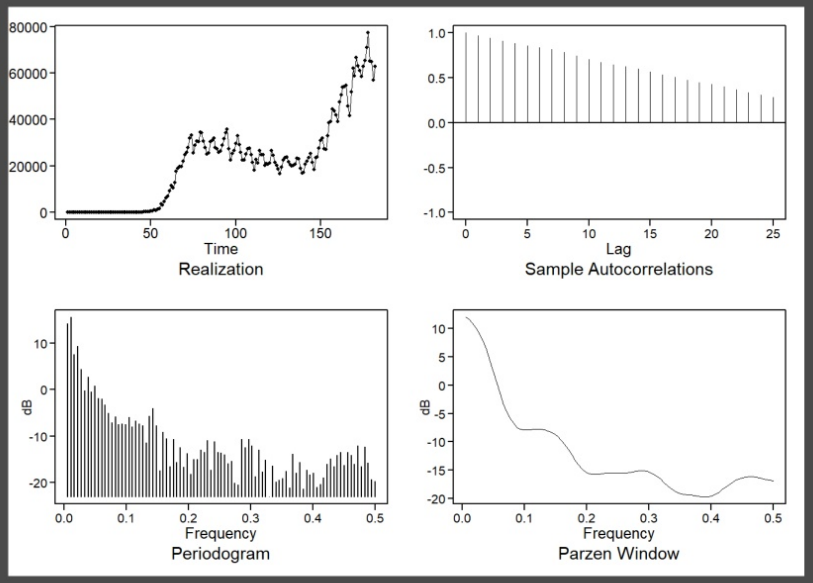


Positive Increase/Total Increase = Positive Percentage

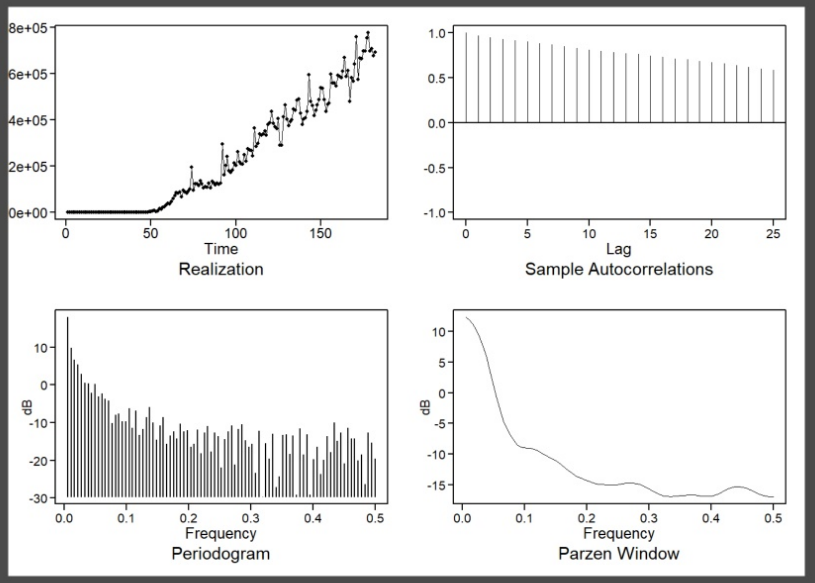


Key Features:
*Date
*Deaths
*Hospitalizations
*New Positive Tests Per Day
*New Negative Tests Per Day
*Total Test Results Per Day

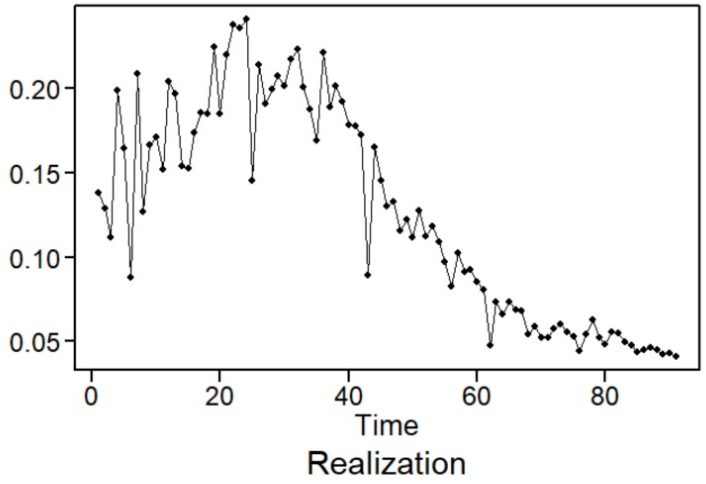
Positive Increase



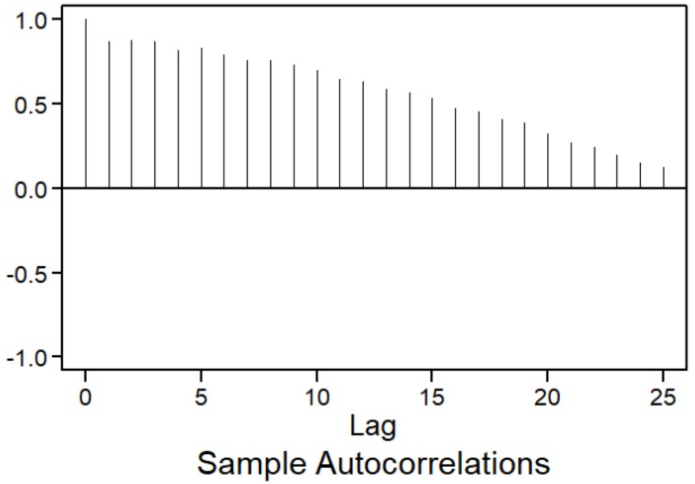
Total Test Increase



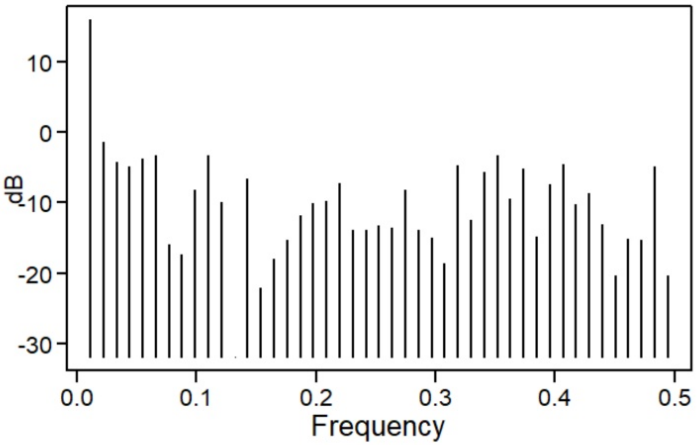
Positive Increase/Total Increase = Positive Percentage



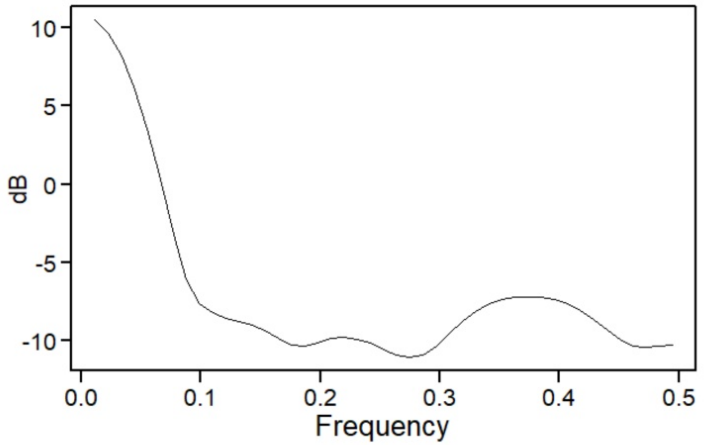
Realization



Sample Autocorrelations



Periodogram



Parzen Window

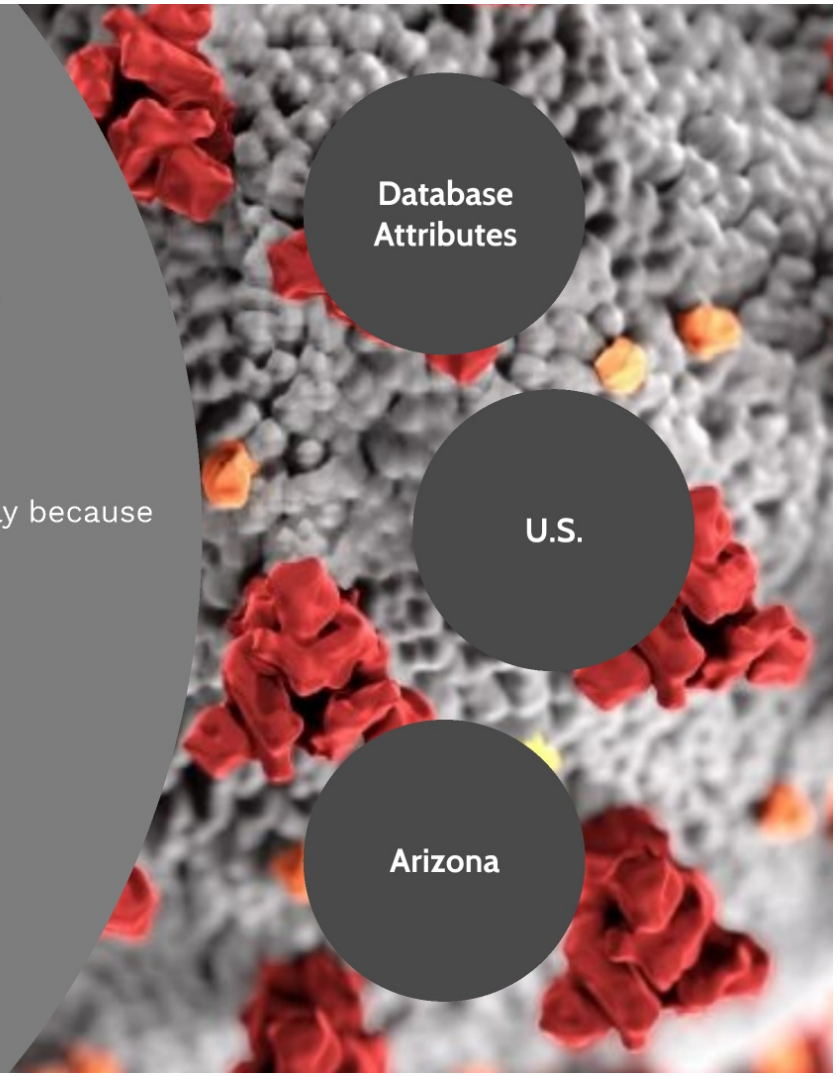
Data Collection and EDA

1. An exploratory data analysis (EDA) was completed and presented. Links to EDA presentations are here:
   - https://www.youtube.com/watch?v=uPOOcZL9o3g&feature=youtu.be
   - https://vimeo.com/443496753/6cb59d1153
2. We explored 2 Databases during the EDA:
   - John's Hopkins
   - Covid Tracking Project

In the end, we used the data from the Covid Tracking Project primarily because of the consistancy in the attributes between tables.

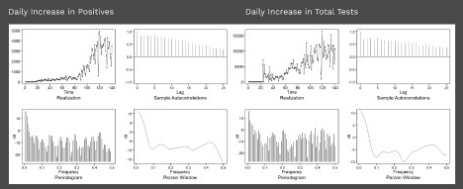
The COVID Tracking Project
AT *The Atlantic*

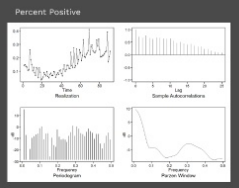
SOURCE: HTTPS://covidtracking.com/data/download
CSV Files Downloaded: U.S. Data and Arizona Data

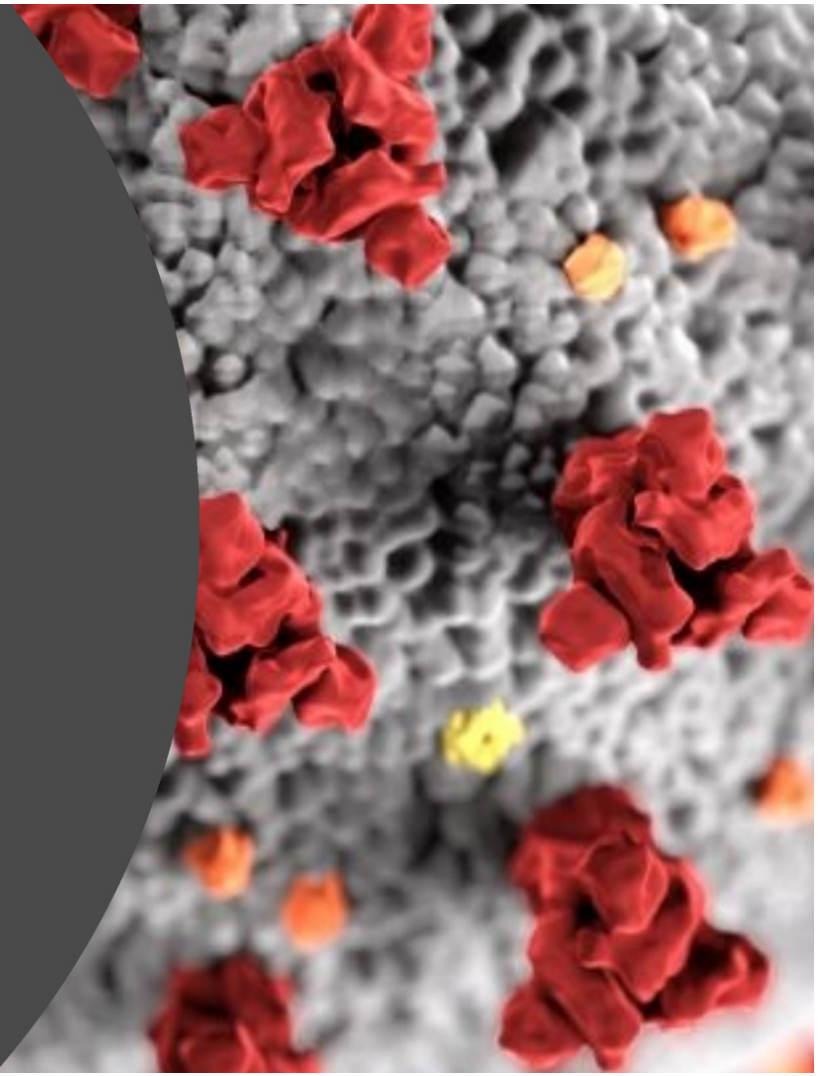


Arizona Data Overview

Same Variables as the US. 180 Observations

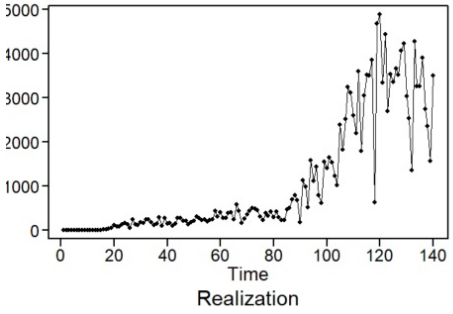


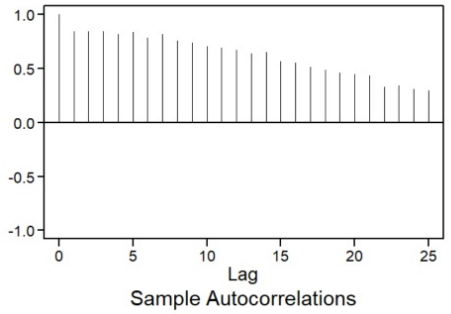


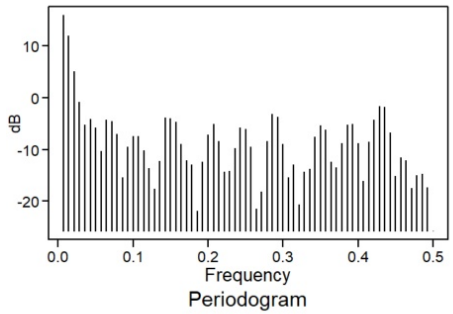


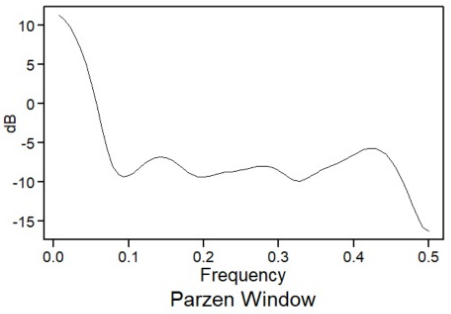
Same Variables as the US. 180 Observations

Daily Increase in Positives

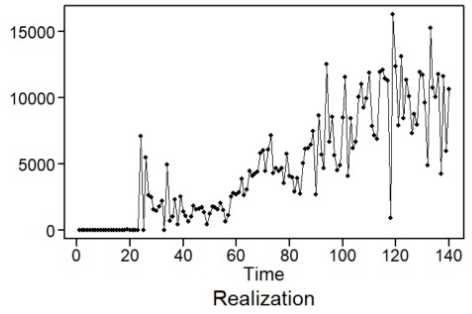


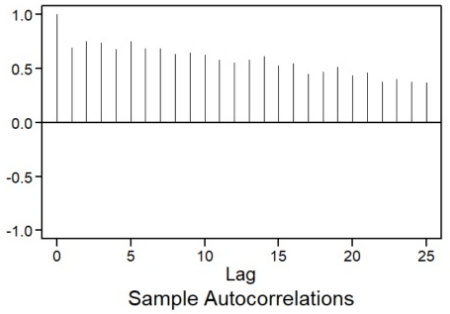


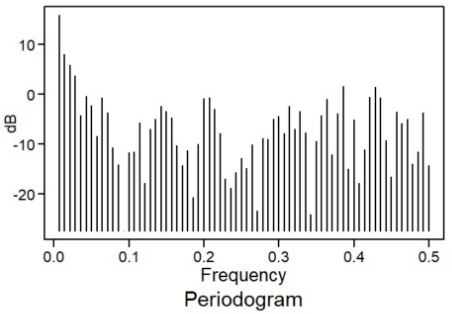


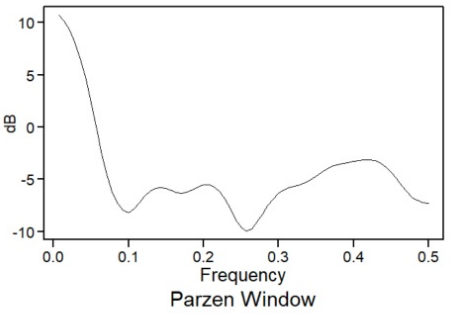


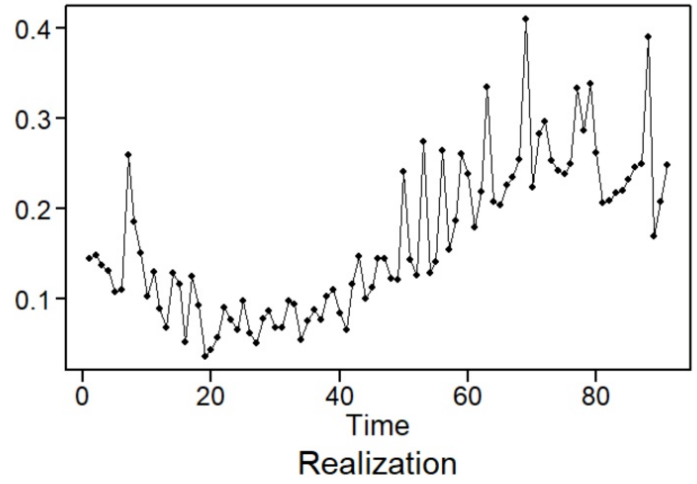
Daily Increase in Total Tests



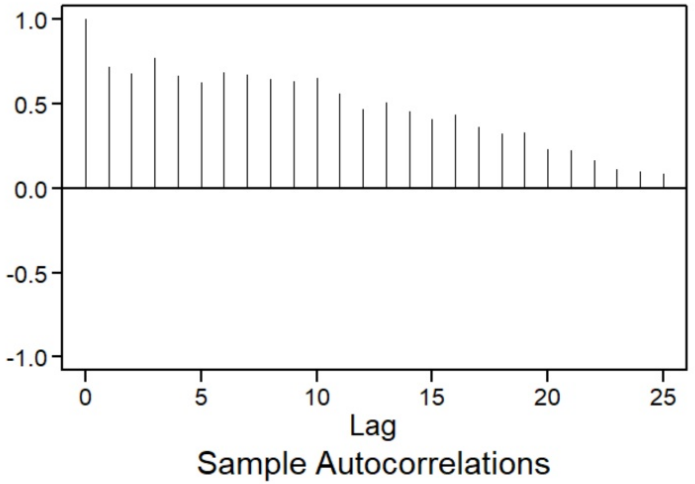




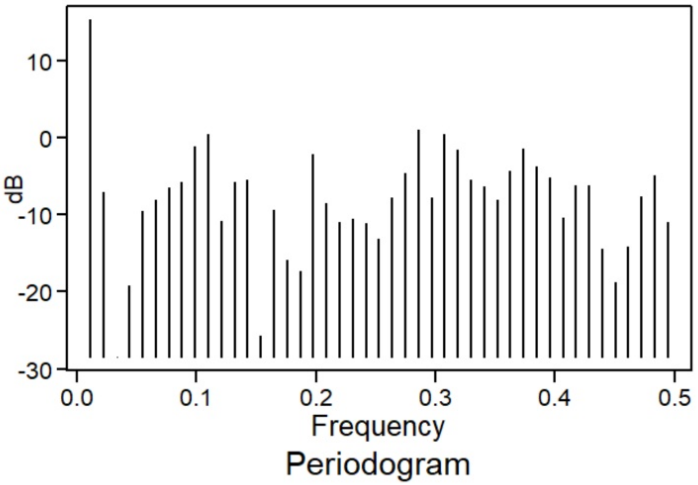




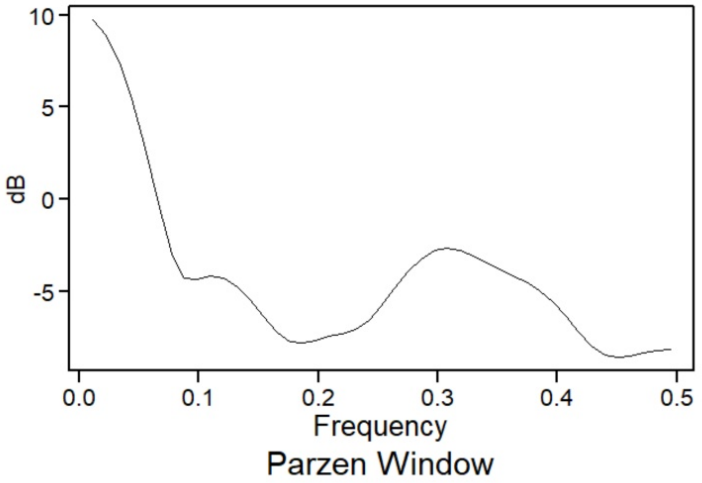
Realization



Sample Autocorrelations



Periodogram



Parzen Window

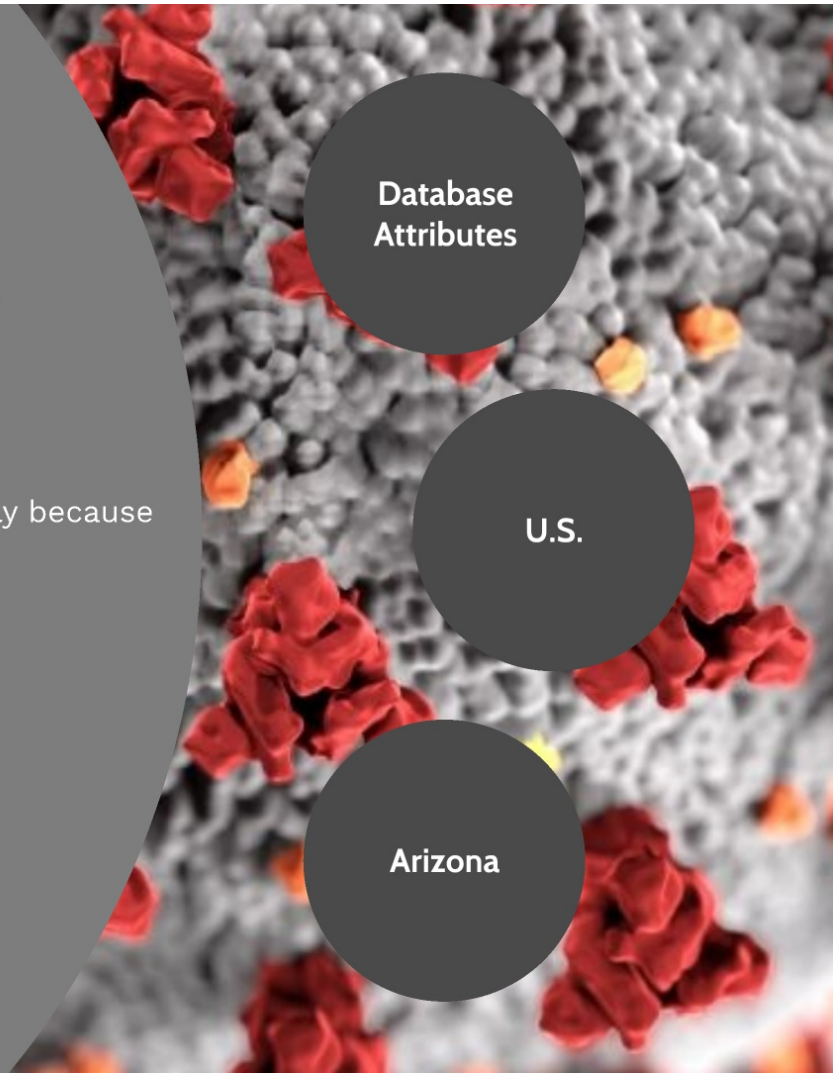
Data Collection and EDA

1. An exploratory data analysis (EDA) was completed and presented. Links to EDA presentations are here:
   - https://www.youtube.com/watch?v=uPOOcZL9o3g&feature=youtu.be
   - https://vimeo.com/443496753/6cb59d1153
2. We explored 2 Databases during the EDA:
   - John's Hopkins
   - Covid Tracking Project

In the end, we used the data from the Covid Tracking Project primarily because of the consistancy in the attributes between tables.

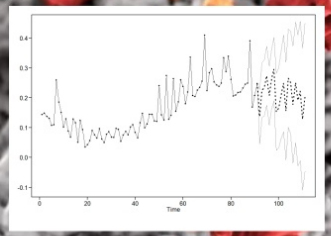
The COVID Tracking Project
AT The Atlantic

SOURCE: HTTPS://covidtracking.com/data/download
CSV Files Downloaded: U.S. Data and Arizona Data



US AND ARIZONA COVID-19 TIME SERIES ANALYSIS AND FORECASTS

Daniel Clark and Gavin Hudgeons
Southern Methodist University



Data Collection and EDA

Univariate Analysis

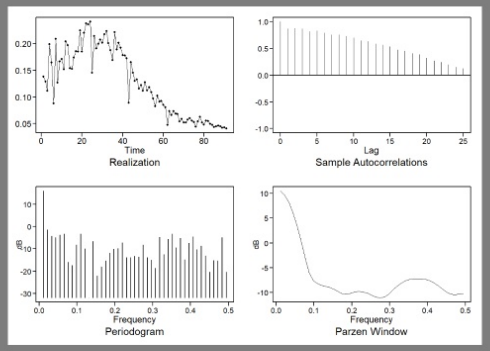
Multivariate Analysis

Summary

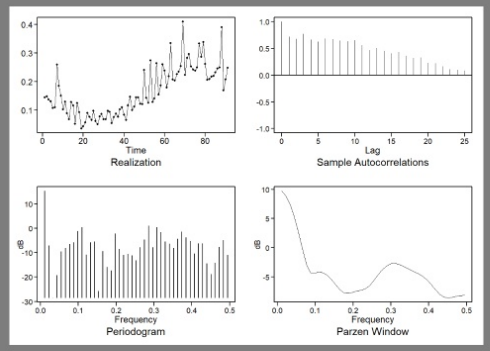
Univariate Analysis

Positive Percentage
Calculated by dividing the daily positive increase by the total test results increase

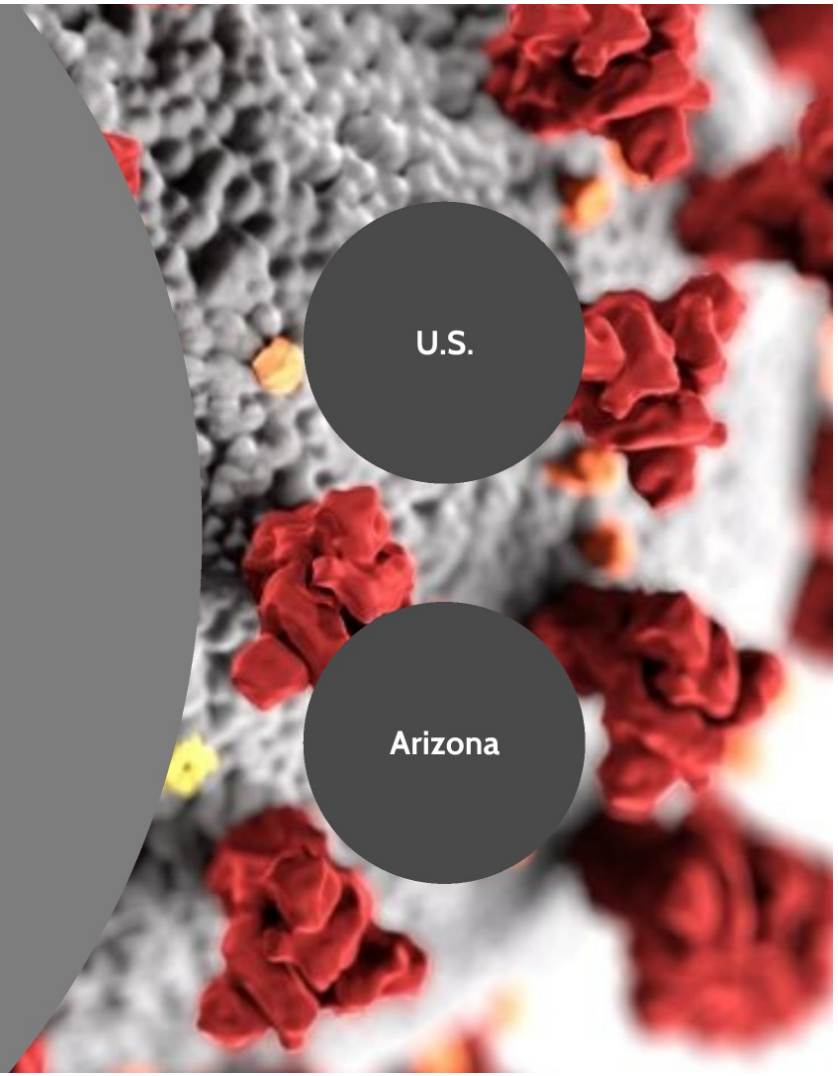
U.S.



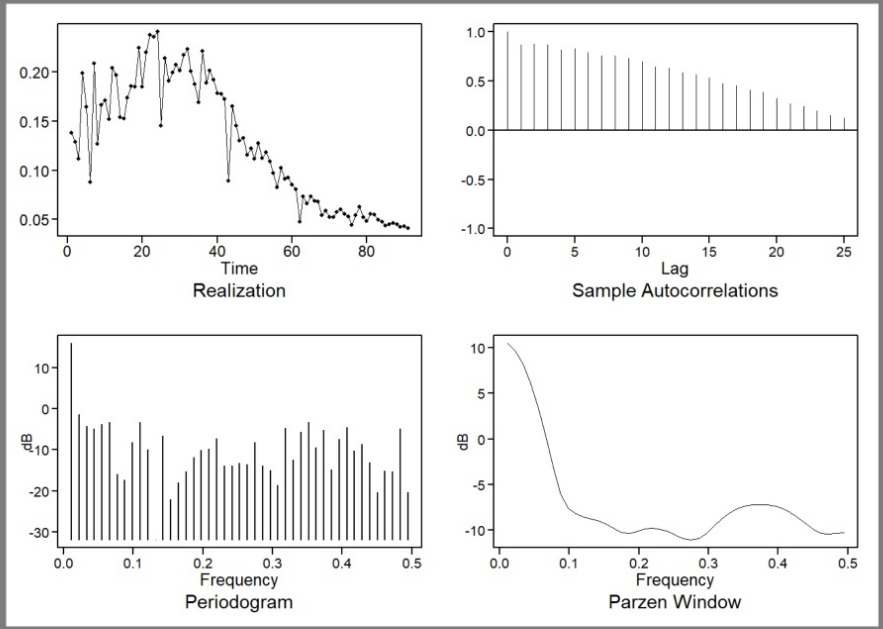
AZ



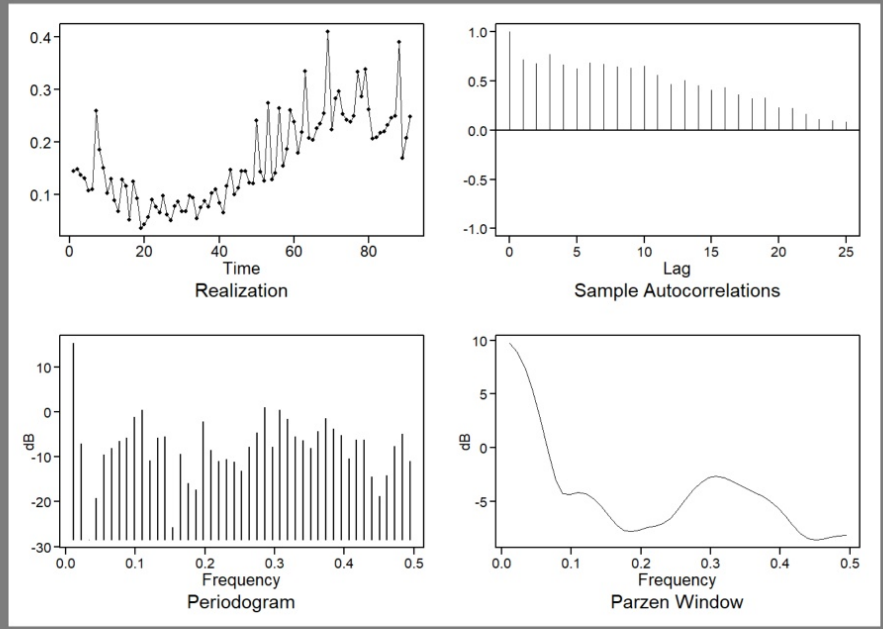
'Percent Positive'
'What percent of total tests given were positive?'



U.S.



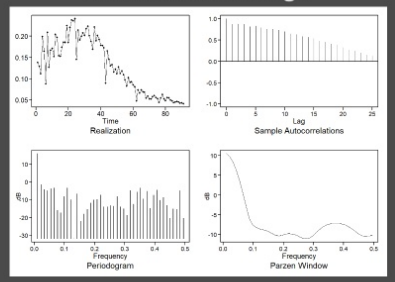
AZ



'Percent Positive'
'What percent of total tests given were positive?'

Positive Percentage
Calculated by dividing the daily positive increase by the total test results increase

Positive Percentage

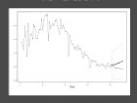


AIC5 selected ARMA(2,2)

(1 - 1.96B + 0.97B^2)(X - 0.13) = (1 -1.185a_t + 0.89a_t^2)

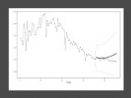
ASE Calculations

10 Back



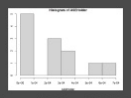
ASE = 0.00034

20 Back

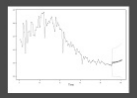


ASE = 0.00026

Rolling Window



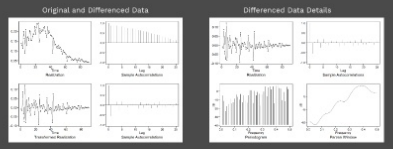
Mean ASE = 0.00023



Forecast: 10 Forward

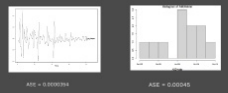
The data was also modeled with non-stationary attributes.

- Non-stationary modeling (1-B)
  - First Difference was Taken
  - AIC recommended an AR(2) from the differenced data. Therefore, an ARIMA(2,1,0) was pursued.

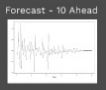


(1 + 0.78B + 0.40B^2)(1 - B) = a_t

ASE Calculations

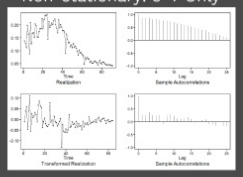


Forecast - 10 Ahead

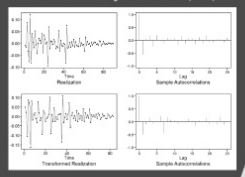


Considered 2 non-stationary options:

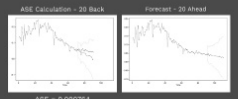
Non-stationary: s=7 Only



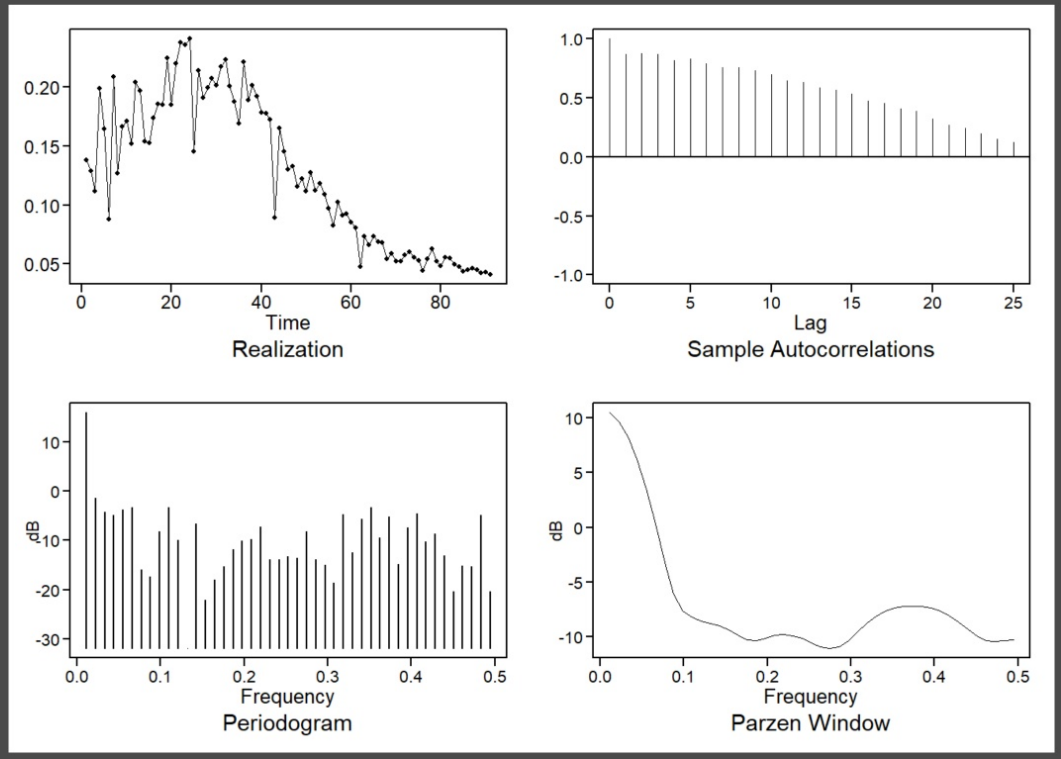
Non-stationary: s=7 and (1-B)



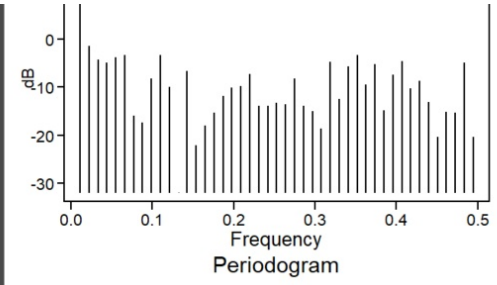
ARUMA(9,1,0)s=7

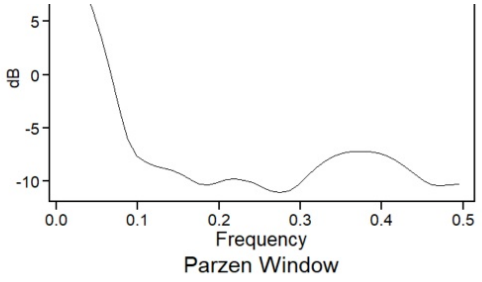


Positive Percentage



AIC5 selected ARMA(2,2)



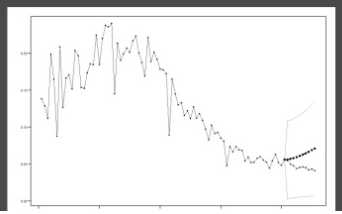


AIC5 selected ARMA(2,2)

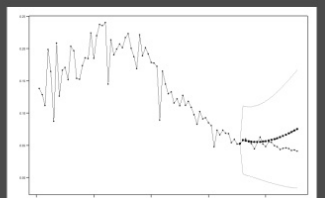
$$(1 - 1.96B + 0.97B^2)(X - 0.13) = (1 - 1.185a_t + 0.89a_t^2)$$

ASE Calculations

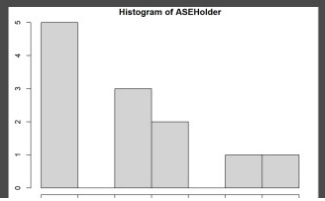
10 Back



20 Back

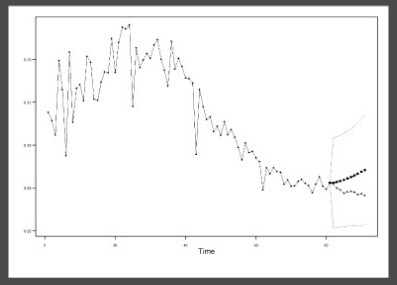


Rolling Window



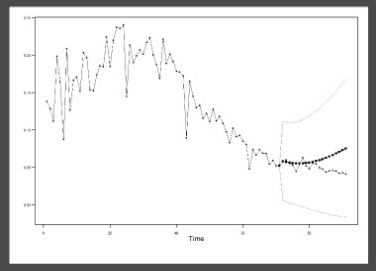
ASE Calculations

10 Back



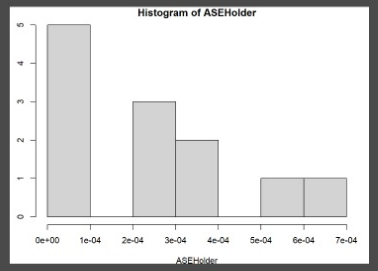
ASE = 0.00034

20 Back

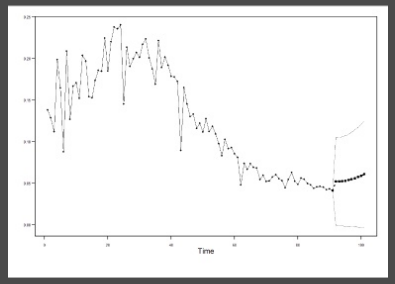


ASE = 0.00026

Rolling Window



Mean ASE = 0.00023

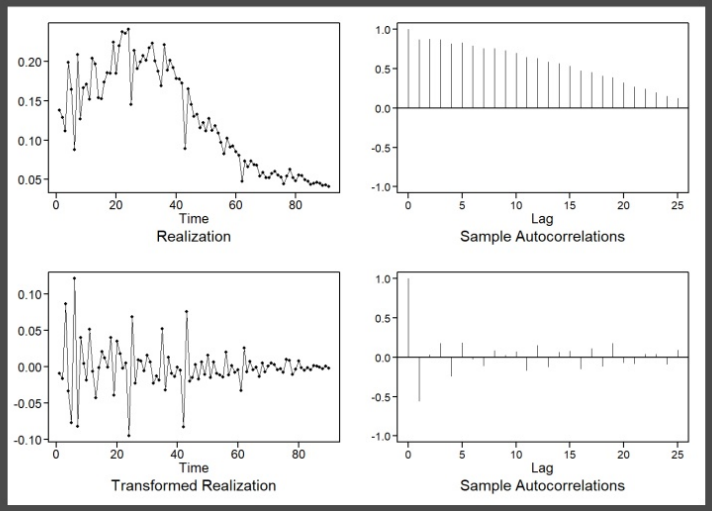


Forecast: 10 Forward

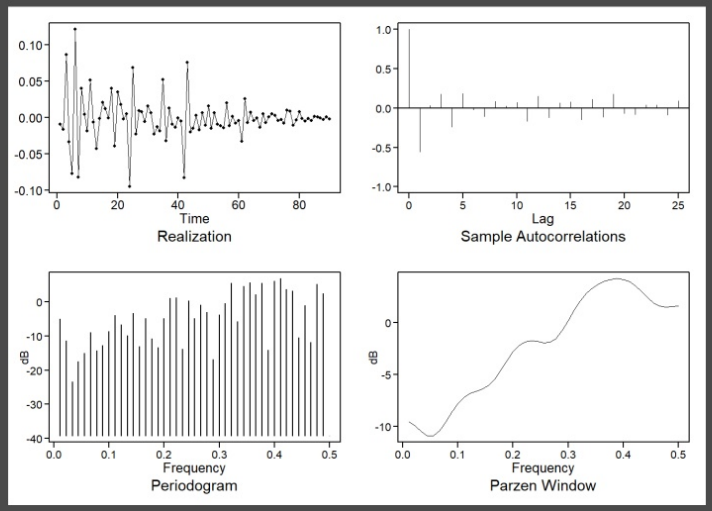
The data was also modeled with non-stationary attributes.

- Non-stationary modeling (1-B)
  - First Difference was Taken
  - AIC recommended an AR(2) from the differenced data. Therefore, an ARIMA(2,1,0) was pursued.

Original and Differenced Data

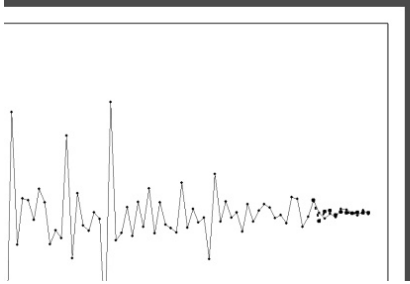


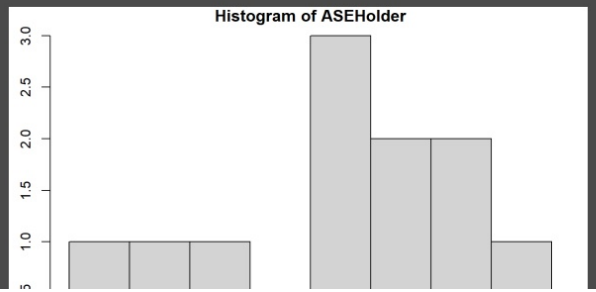
Differenced Data Details



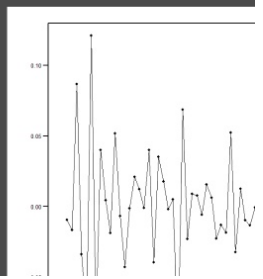
$$(1 + 0.78B + 0.40B^2)(1 - B) = a_t$$

ASE Calculations



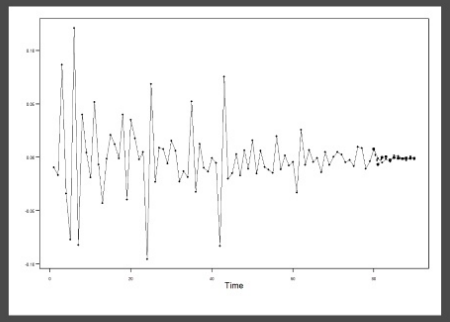


Forecast

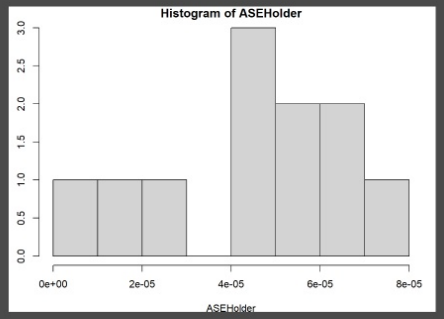


$$(1 + 0.78B + 0.40B^2)(1 - B) = a_t$$

ASE Calculations

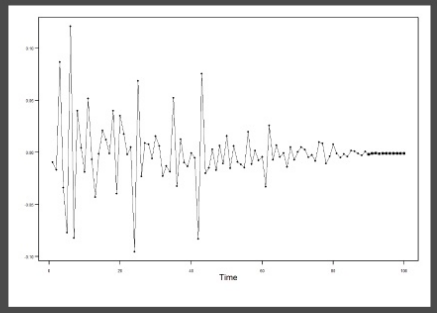


ASE = 0.0000354



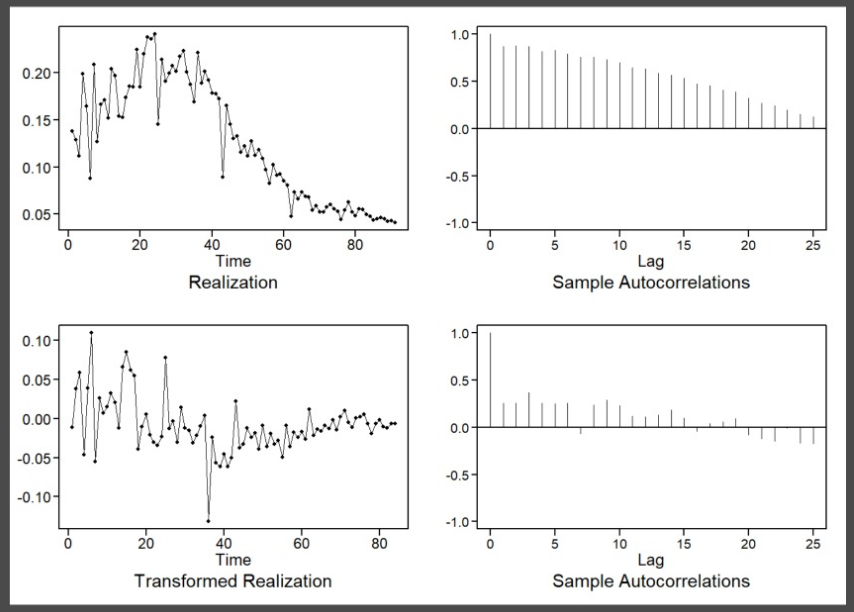
ASE = 0.00045

Forecast - 10 Ahead

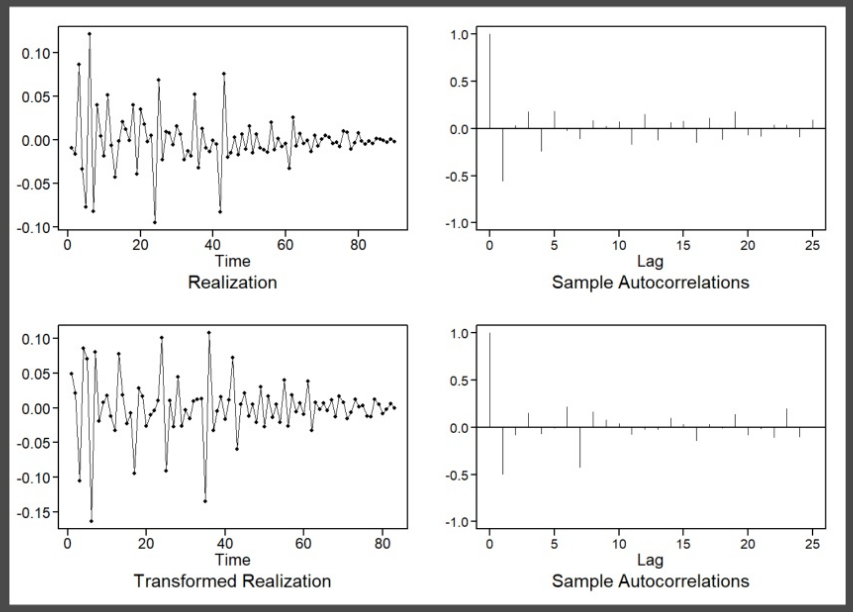


Considered 2 non-stationary options:

Non-stationary: s=7 Only



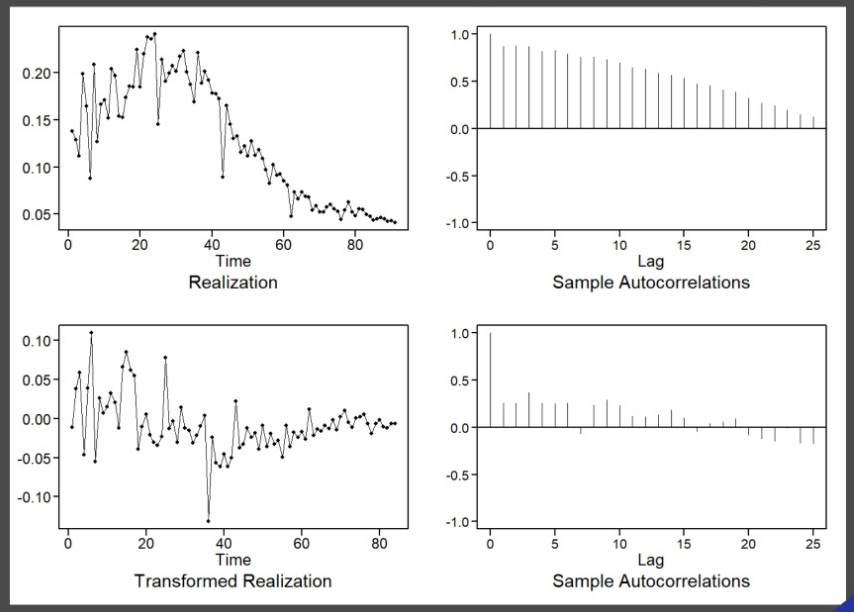
Non-stationary: s=7 and (1-B)



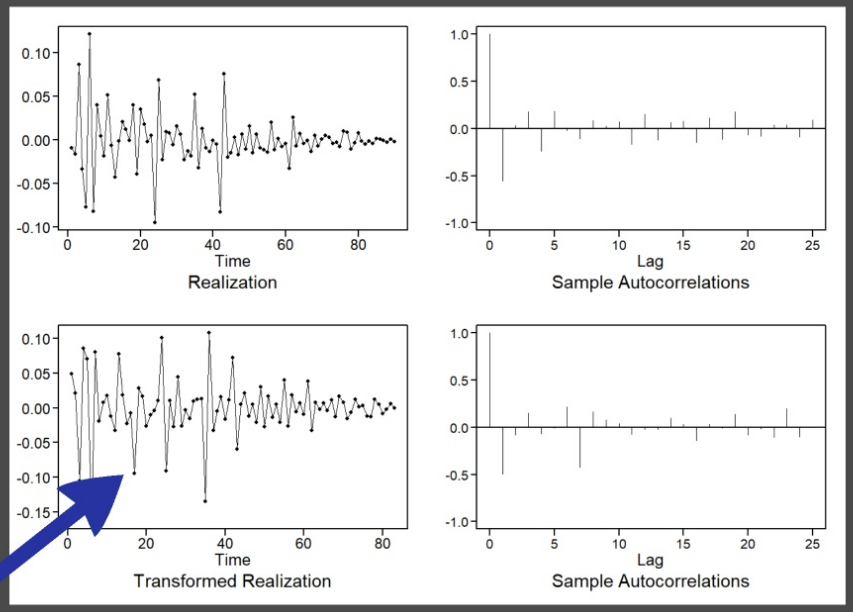
ARUMA(9,1,0)s=7

Considered 2 non-stationary options:

Non-stationary: s=7 Only

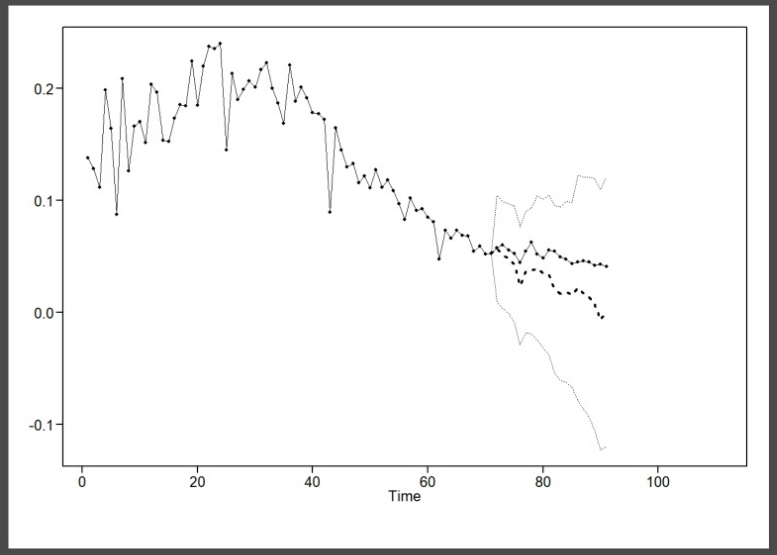


Non-stationary: s=7 and (1-B)



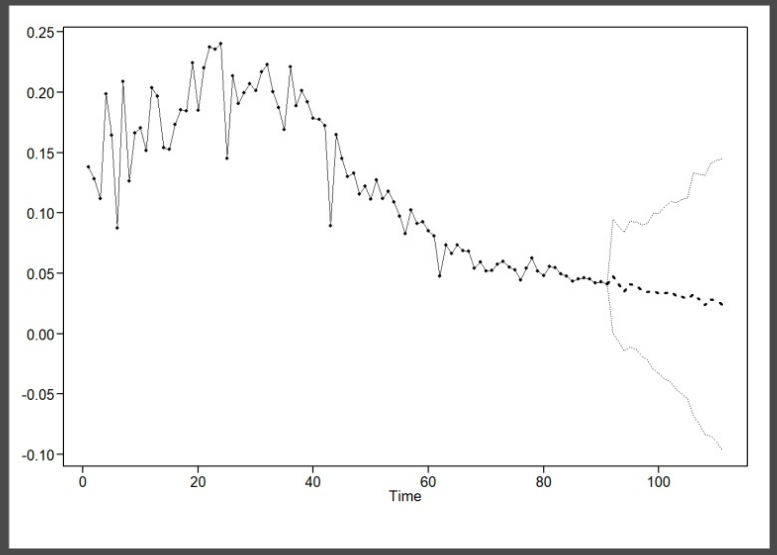
ARUMA(9,1,0)s=7

ASE Calculation - 20 Back



ASE = 0.000764

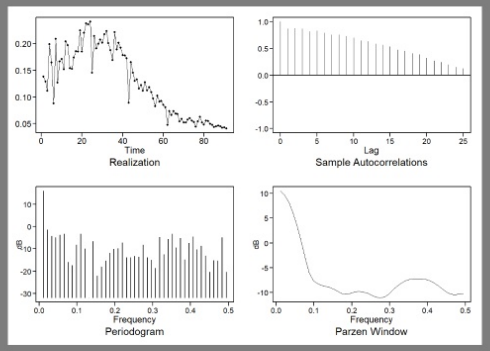
Forecast - 20 Ahead



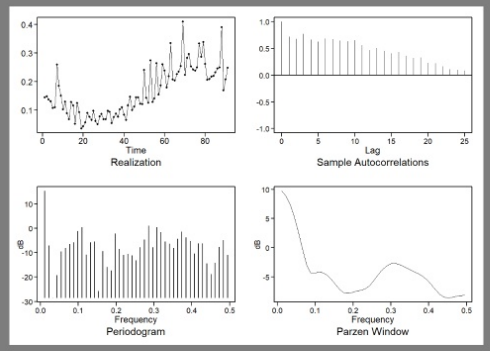
Univariate Analysis

Positive Percentage
Calculated by dividing the daily positive increase by the total test results increase

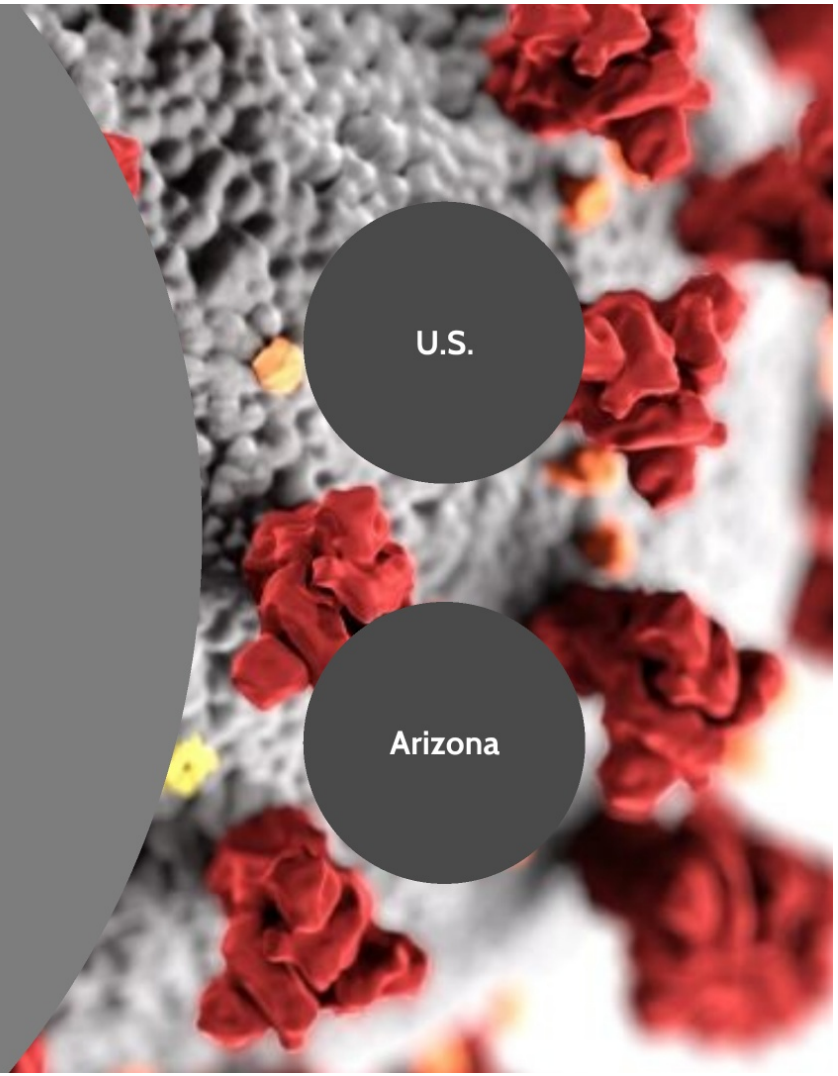
U.S.



AZ

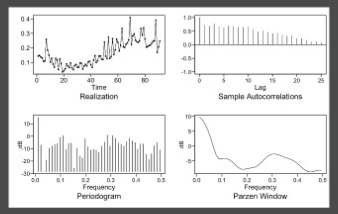


'Percent Positive'
'What percent of total tests given were positive?'



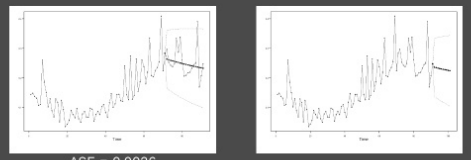
Arizona Data Modeling

Positive Percentage



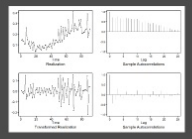
- Model #1
- Stationary Forecast - ARMA(1,1)
- Selected from AIC5

(1 - 0.98B)(X - 0.16) = (1 - 0.73a_t)



ASE = 0.0026

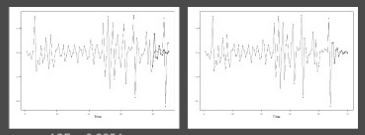
- Model #2
- NonStationary Forecast (1-B)



(1 + 0.65B + 0.50B^2)(1 - B) = a_t

ASE Calculation - 20 behind

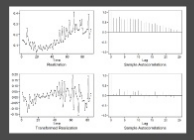
Forecasts - 20 ahead



ASE = 0.0054

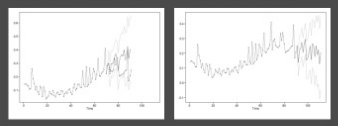
- Model #3
- NonStationary Forecast (1-B) and (s=7)

ARUMA (11,1,0) s= 7



ASE Calculation - 20 behind

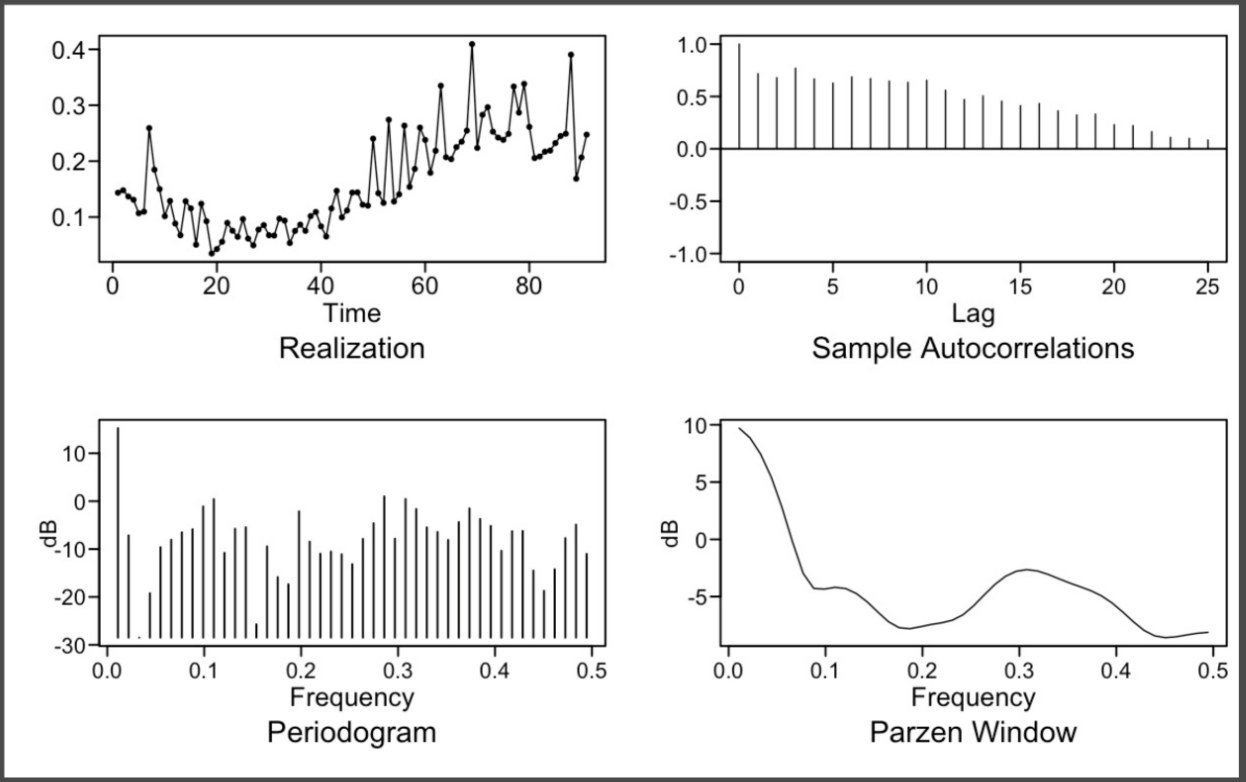
Forecasts - 20 ahead



(1 + 0.93B + 0.78B^2 + 0.45B^3 + 0.56B^4 + 0.55B^5 + 0.22B^6 + 0.44B^7 + 0.23B^8 - 0.02B^9 - 0.49B^10 - 0.36B^11)(1 - B^7)(1 - B) = a_t

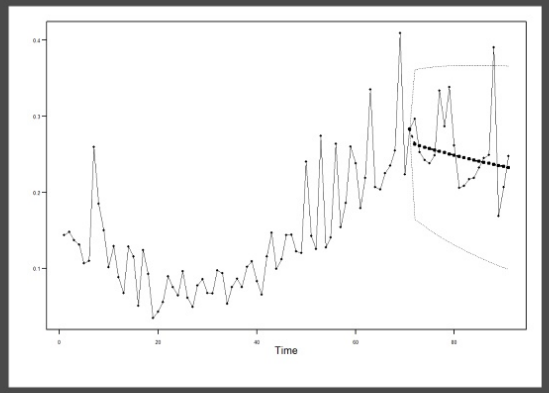
ASE = 0.012

Positive Percentage

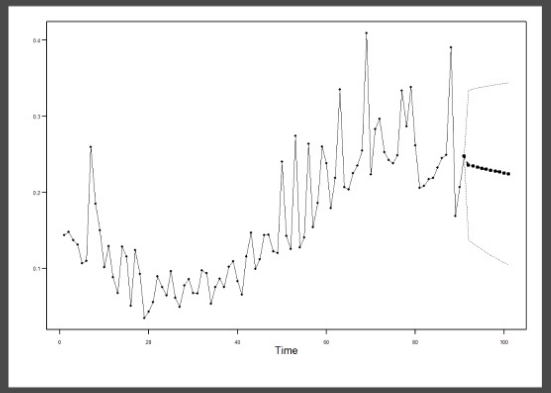


- Model #1
- Stationary Forecast - ARMA(1,1)
- Selected from AIC5

$$(1 - 0.98B)(X - 0.16) = (1 - 0.73a_t)$$

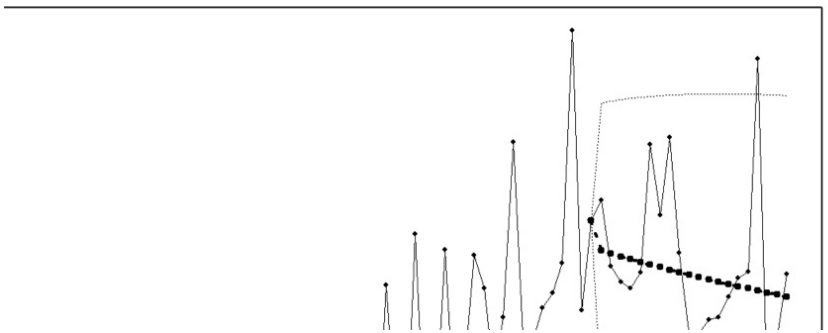


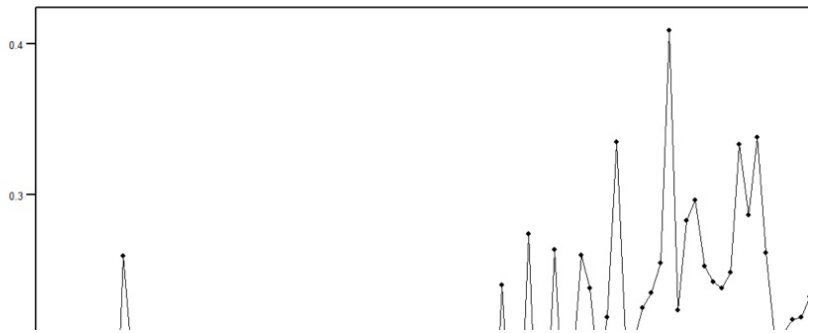
ASE = 0.0026



odel #1
tationary Forecast - ARMA(
elected from AIC5

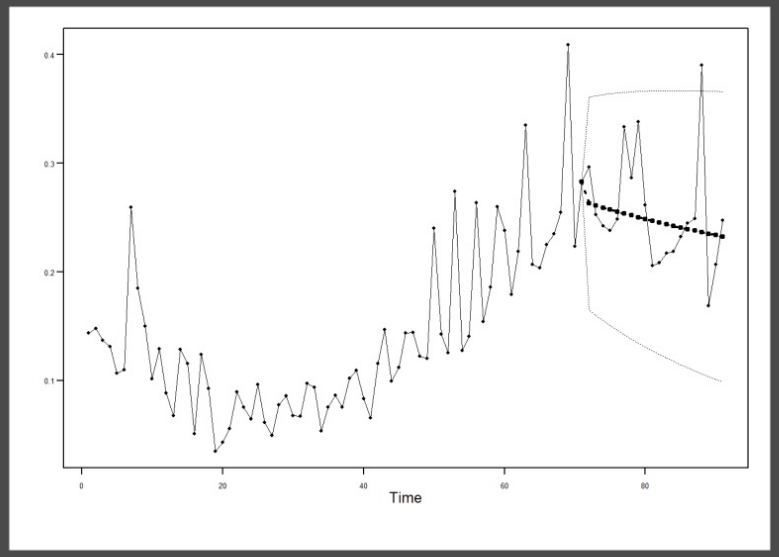
$$(1 - 0.98B)(X - 0.16) = (1 - 0.73a_t)$$



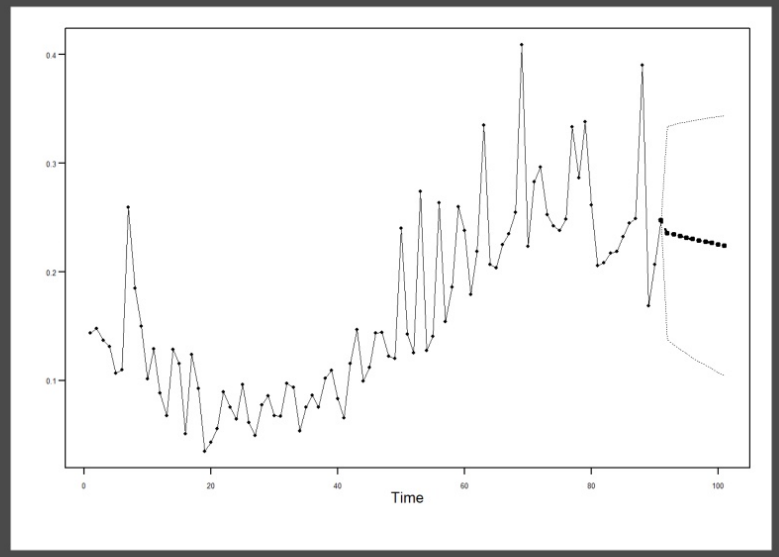


Selected from AIC

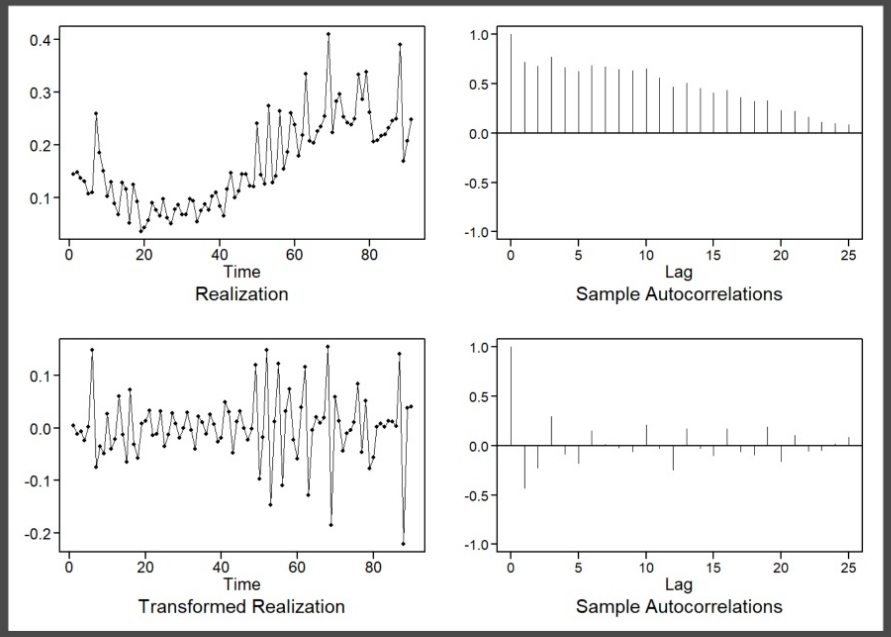
$$(1 - 0.98B)(X - 0.16) = (1 - 0.73a_t)$$



ASE = 0.0026

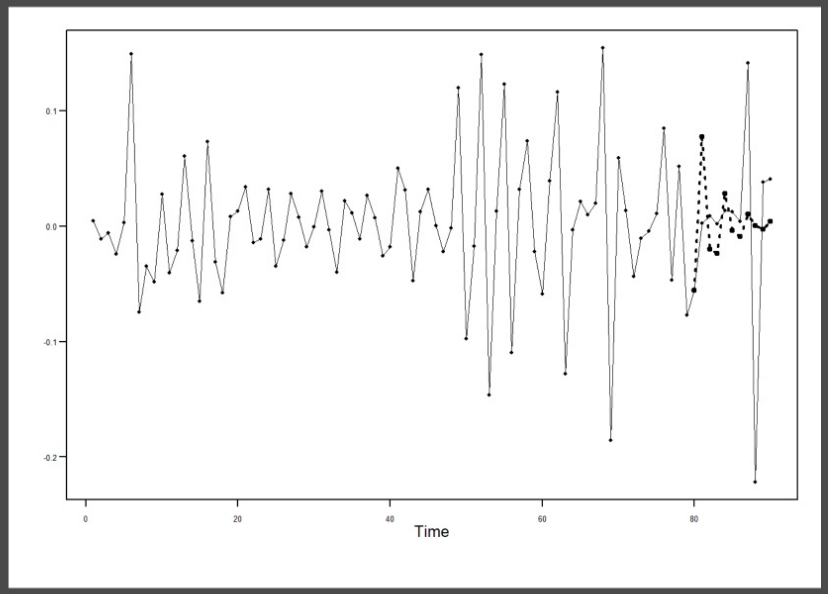


- Model #2
- NonStationary Forecast (1-B)



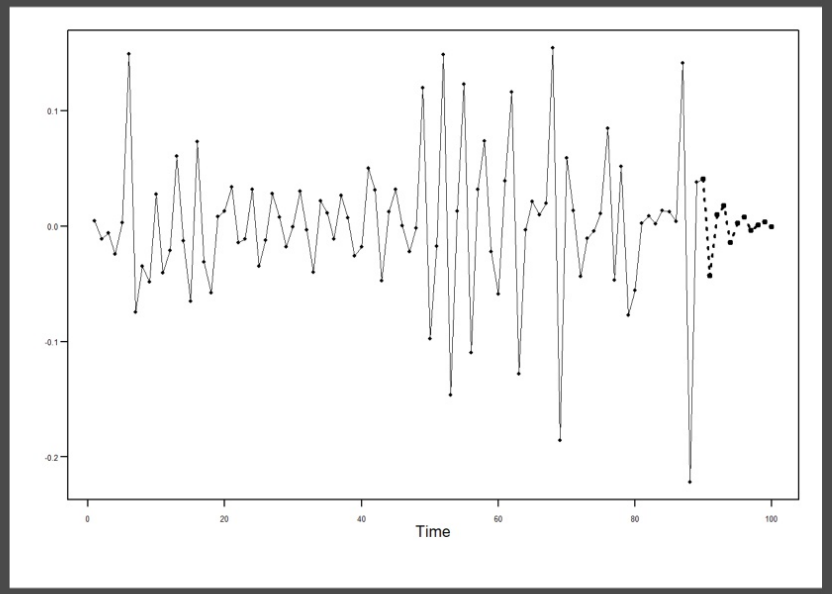
(1 + 0.65B + 0.50B^2)(1 - B) = a_t

ASE Calculation - 20 behind



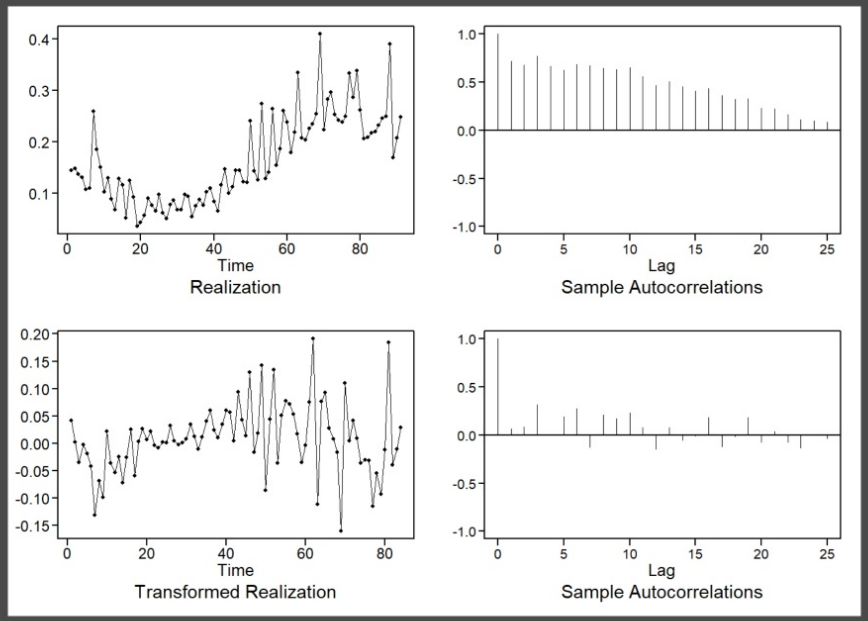
ASE = 0.0054

Forecasts - 20 ahead

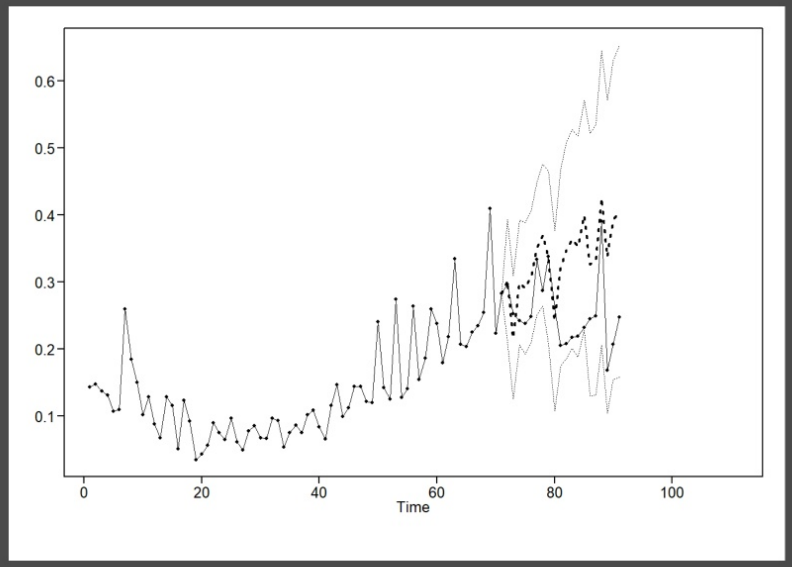


- Model #3
- NonStationary Forecast (1-B) and (s=7)

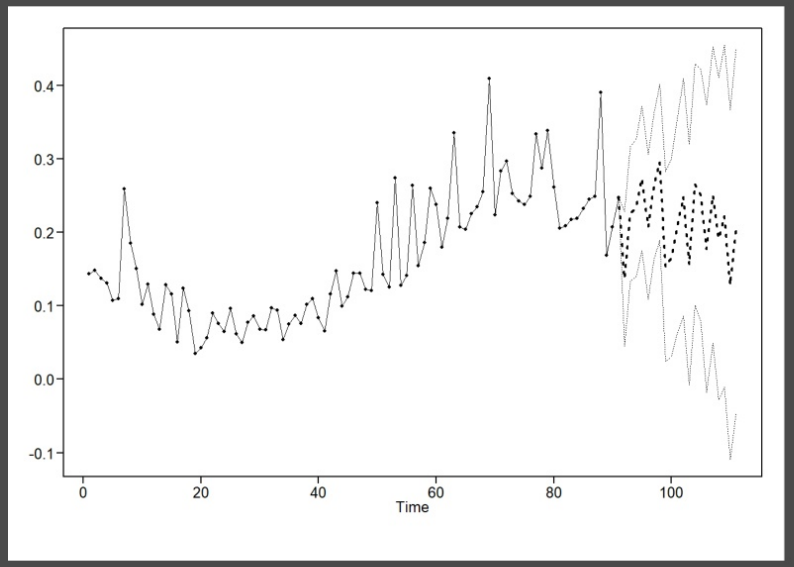
ARUMA (11,1,0) s= 7



ASE Calculation - 20 behind



Forecasts - 20 ahead



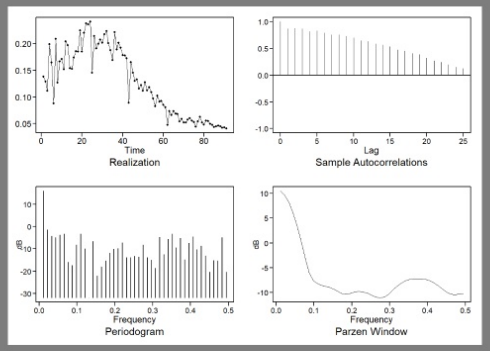
(1 + 0.93B + 0.78B^2 + 0.45B^3 + 0.56B^4 + 0.55B^5 + 0.22B^6 + 0.44B^7 + 0.23B^8 - 0.02B^9 - 0.49B^10 - 0.36B^11)(1 - B^7)(1 - B) = a_t

ASE = 0.012

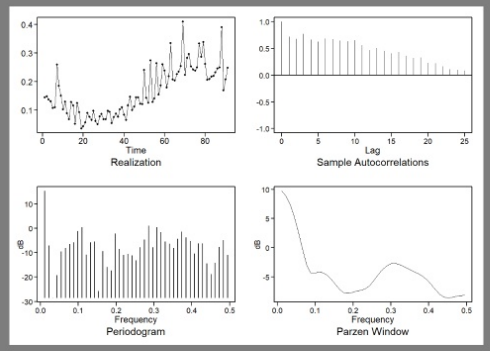
Univariate Analysis

Positive Percentage
Calculated by dividing the daily positive increase by the total test results increase

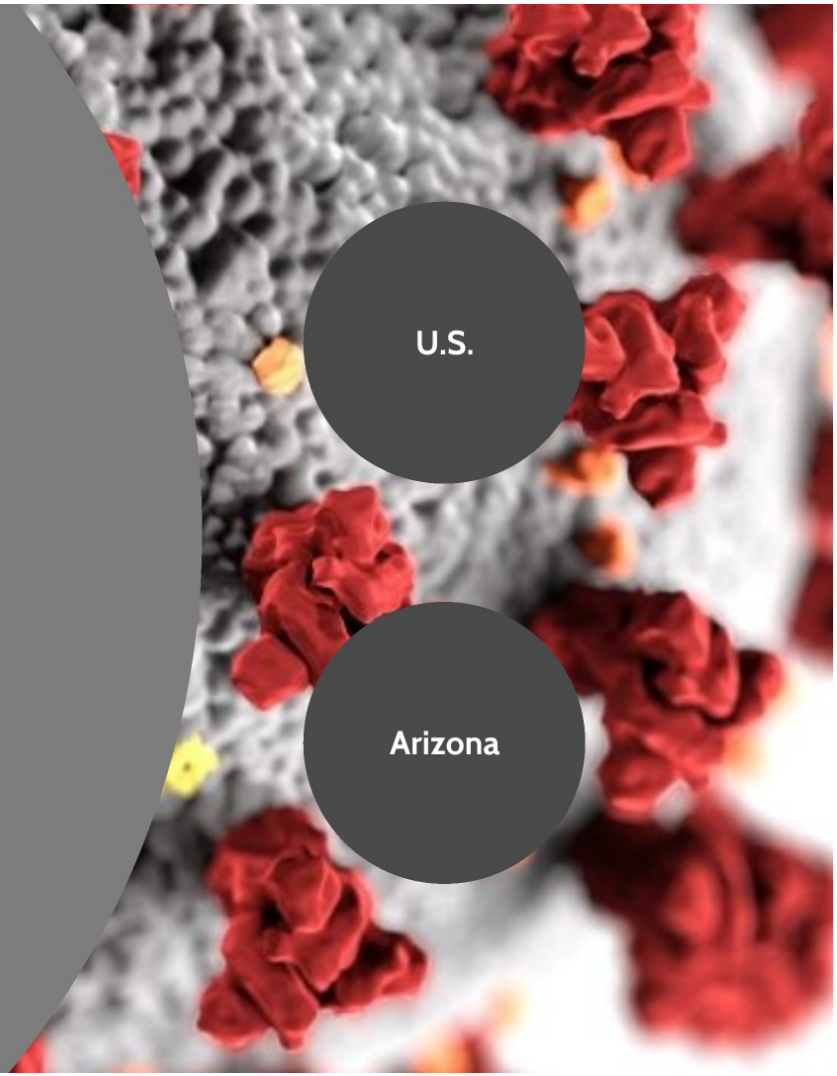
U.S.



AZ

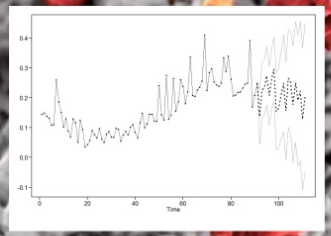


'Percent Positive'
'What percent of total tests given were positive?'



US AND ARIZONA COVID-19 TIME SERIES ANALYSIS AND FORECASTS

Daniel Clark and Gavin Hudgeons
Southern Methodist University



Data Collection and EDA

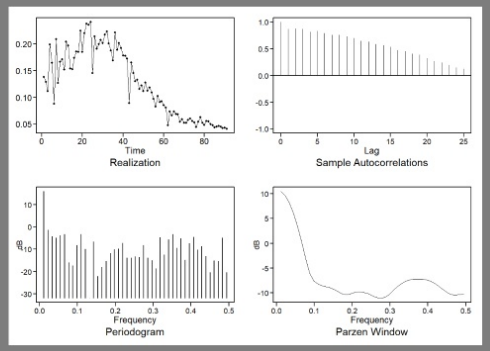
Univariate Analysis

Multivariate Analysis

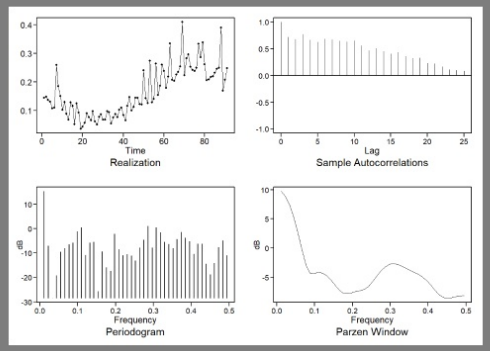
Summary

Multivariate Analysis

U.S.



AZ



'Percent Positive'
'What percent of total tests given were positive?'

U.S.

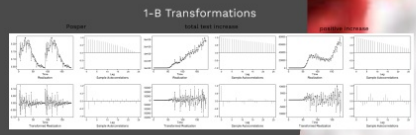
Arizona

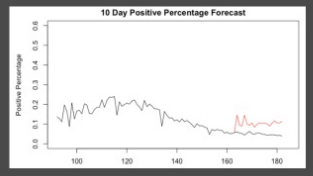
Neural Network Models

U.S. Multivariate Forecasting

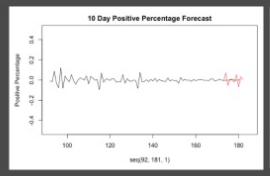
Our Var Model included a multivariate analysis of Positive Percentage using:
* Total Test Results Increase
* Positive Test Results Increase

Three Var Models
* Stationary
* Trend
* Airplane Model

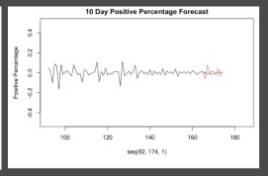




ASE = .0028



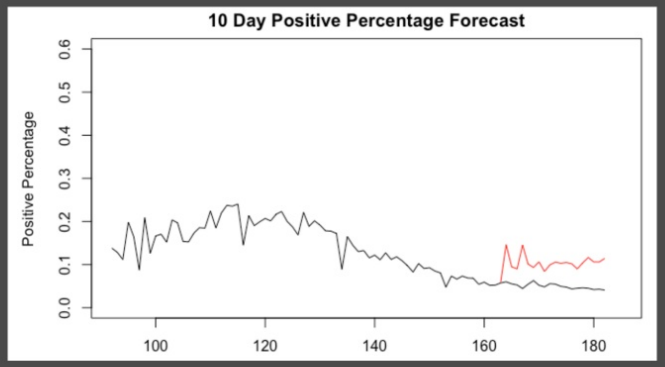
ASE = .0017

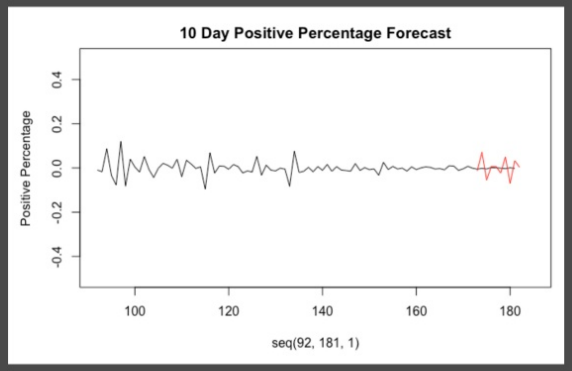


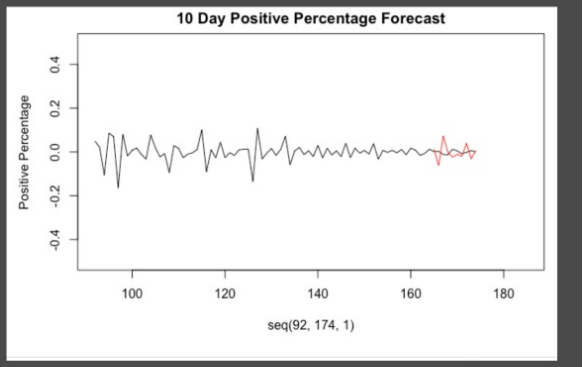
ASE = .0016

Our Var Model included a multivariate analysis of Positive Percentage using:
* Total Test Results Increase
* Positive Test Results Increase

Three Var Models
* Stationary
* Trend
* Airplane Model







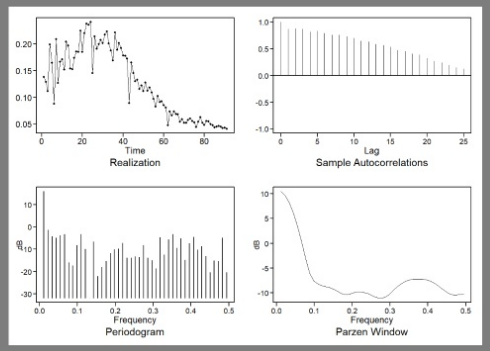
ASE = .0028

ASE = .0017

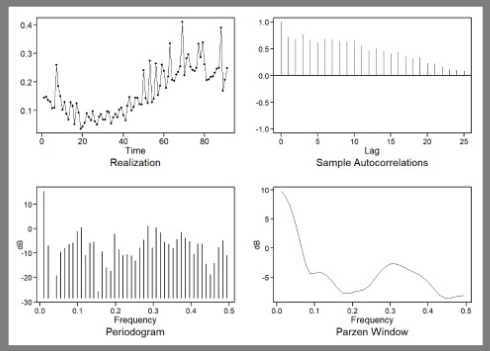
ASE = .0016

Multivariate Analysis

U.S.



AZ



'Percent Positive'
'What percent of total tests given were positive?'

U.S.

Arizona

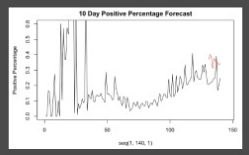
Neural Network Models

Arizona Multivariate Models

Our Var Model included a multivariate analysis of Positive Percentage using:
* Total Test Results Increase
* Positive Test Results Increase

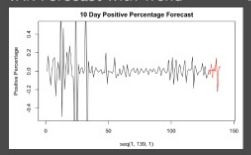
Three Var Models
* Stationary
* Trend
* Airplane Model

VAR Model



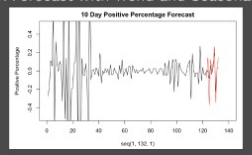
ASE = 0.064

VAR Forecast with Trend



ASE = 0.0017

VAR Forecast with Trend and Seasonality



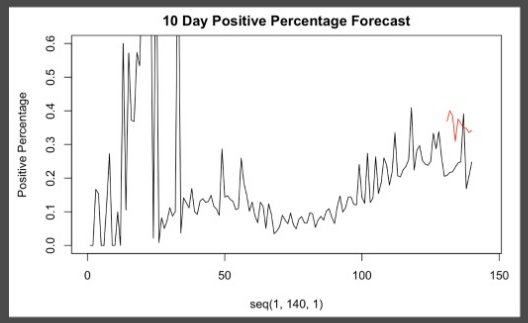
ASE = 0.0148

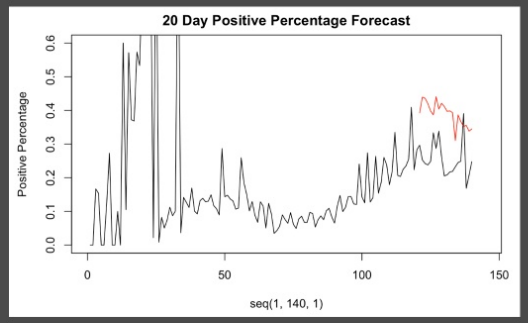
Arizona Multivariate Models

Our Var Model included a multivariate analysis of Positive Percentage using:
* Total Test Results Increase
* Positive Test Results Increase

Three Var Models
* Stationary
* Trend
* Airplane Model

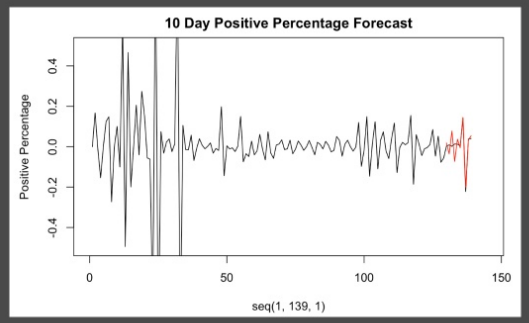
VAR Model

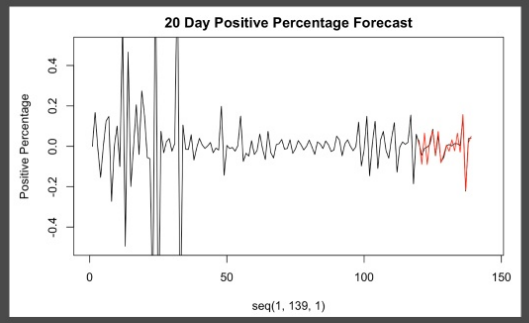




ASE = 0.064

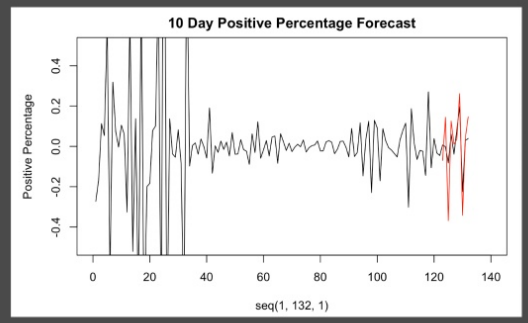
VAR Forecast with Trend

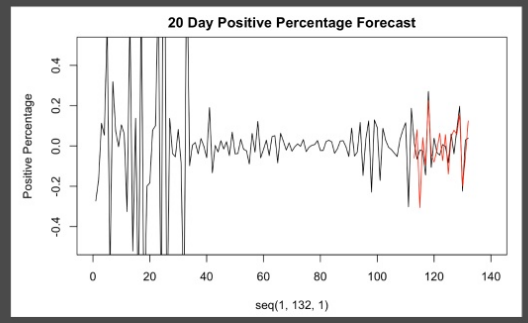




ASE = 0.0017

VAR Forecast with Trend and Seasonality

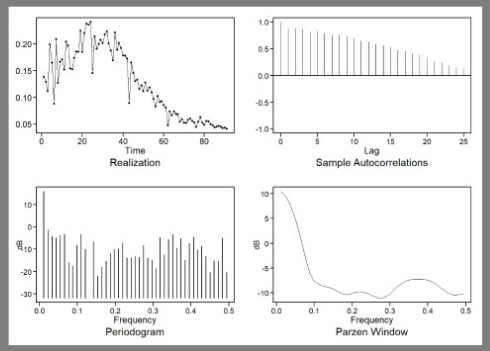




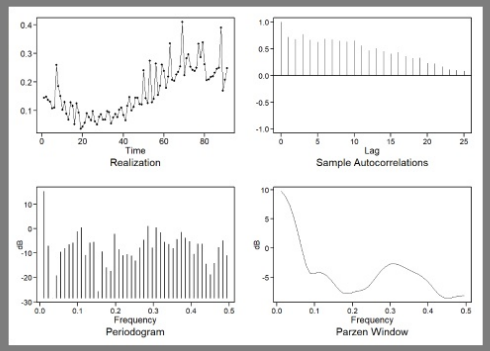
ASE = 0.0148

Multivariate Analysis

U.S.



AZ



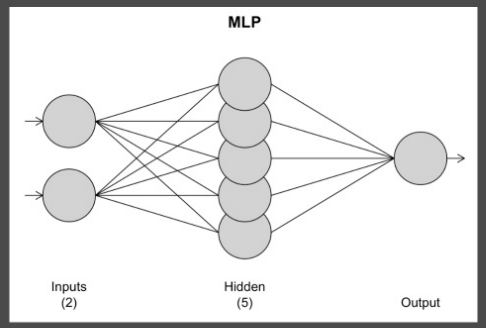
'Percent Positive'
'What percent of total tests given were positive?'

U.S.

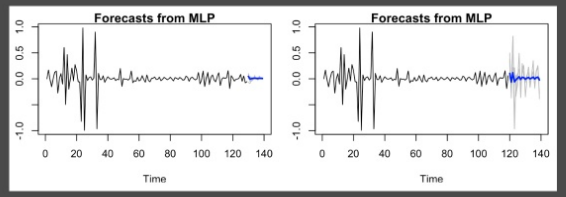
Arizona

Neural Network Models

AZ MLP Model



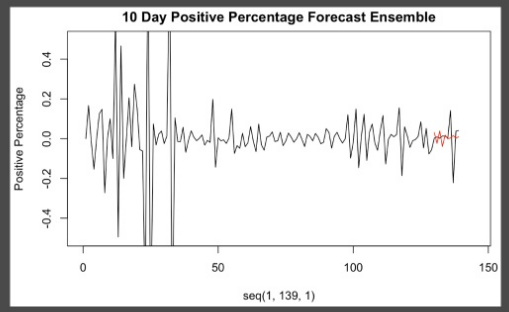
The MLP output gave us 5 nodes and 20 repititions, with two regressors included.



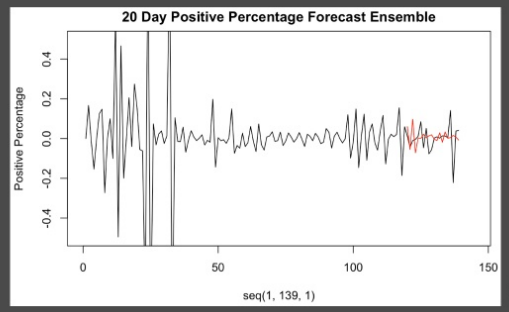
ASE MLP10 = 0.0077

ASE MLP20 = 0.0076

Ensemble Model

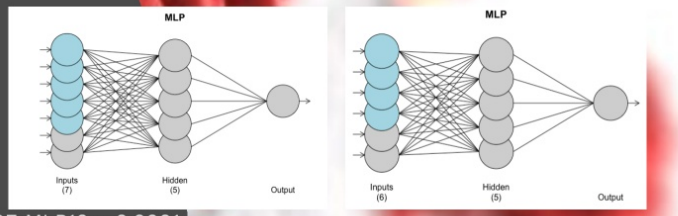


ASE = 0.0084



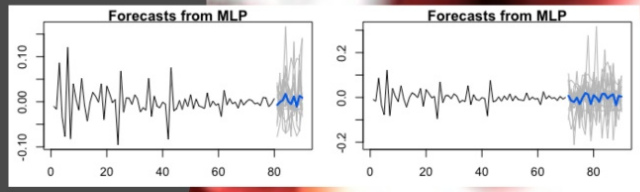
ASE = 0.0063

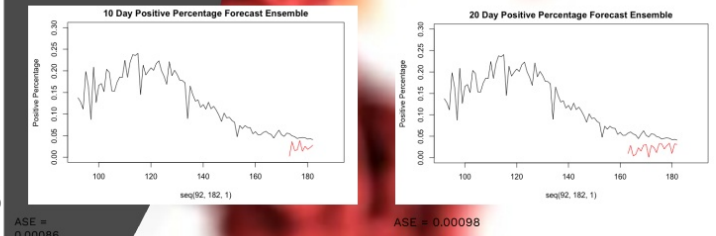
US MLP Model



ASE MLP10 = 0.0001

ASE MLP20 = 0.0005





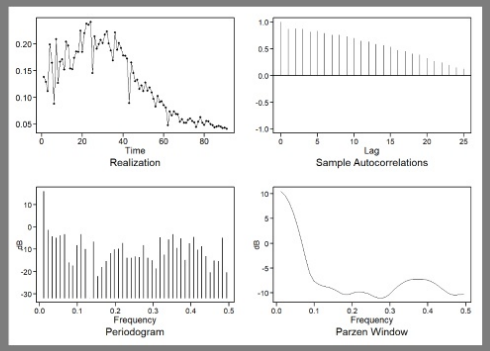
ASE = 0.00086

ASE = 0.00098

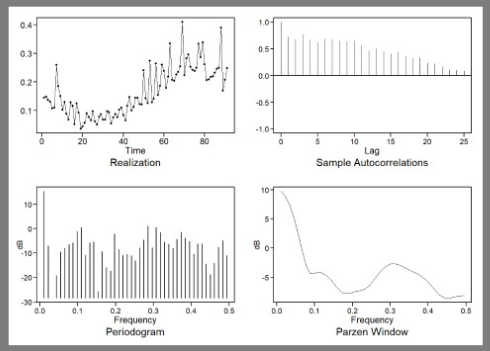
ASE = 0.00073

Multivariate Analysis

U.S.



AZ



'Percent Positive'
'What percent of total tests given were positive?'

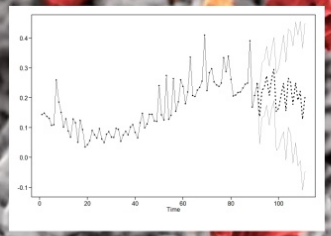
U.S.

Arizona

Neural Network Models

US AND ARIZONA COVID-19 TIME SERIES ANALYSIS AND FORECASTS

Daniel Clark and Gavin Hudgeons
Southern Methodist University



Data Collection and EDA

Univariate Analysis

Multivariate Analysis

Summary

SUMMARY

US Models (10 Day Forecast)

"Arma(2,2): 0.000259881145874531"
"Arima(2,1,0): 3.54076321872339e-05"
"Aruma(9,1,0)s=7: 0.000763959941806105"
"Var Stationary: 0.00282287523520302"
"Var Trend: 0.00174724210122012"
"Var Airplane: 0.00165197254640644"
"MLP: 0.000122143914746698"
"Ensemble: 0.000726832950246414"

AZ Models (10 Day Forecast)

"Arma(2,2): 0.00258659893879501"
"Arima(2,1,0): 0.0053589563896785"
"Aruma(9,1,0)s=7: 0.0118374559290337"
"Var Stationary: 0.0590967906499526"
"Var Trend: 0.00168893770664753"
"Var Airplane: 0.0147814492897754"
"MLP: 0.00766458411127551"
"Ensemble: 0.00834489705311712"

The best performing model with the US data is the ARIMA(2,1,0) while the best AZ model was VAR Trend. Both models were ARIMA models that differenced for Trend. Interestingly, while we see what visually appears to be a seasonal trend, accounting for it does not produce the best model

Thank you.

SUMMARY

US Models (10 Day Forecast)

"Arma(2,2): 0.0002598811458745 31"
"Arima(2,1,0): 3.54076321872339e-05"
"Aruma(9,1,0)s=7: 0.0007639599418061 05"
"Var Stationary: 0.00282287523520302"
"Var Trend: 0.00174724210122012"
"Var Airplane: 0.00165197254640644"
"MLP: 0.000122143914746698"
"Ensemble: 0.0007268329502464 14"

AZ Models (10 Day Forecast)

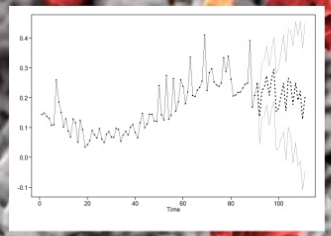
"Arma(2,2): 0.00258659893879501"
"Arima(2,1,0): 0.0053589563896785"
"Aruma(9,1,0)s=7: 0.0118374559290337"
"Var Stationary: 0.0590967906499526"
"Var Trend: 0.00168893770664753"
"Var Airplane: 0.0147814492897754"
"MLP: 0.0076645841112755 1"
"Ensemble: 0.0083448970531171 2"

The best performing model with the US data is the ARIMA(2,1,0) while the best AZ model was VAR Trend. Both models were ARIMA models that differenced for Trend. Interestingly, while we see what visually appears to be a seasonal trend, accounting for it does not produce the best model

Thank you.

US AND ARIZONA COVID-19 TIME SERIES ANALYSIS AND FORECASTS

Daniel Clark and Gavin Hudgeons
Southern Methodist University



Data Collection and EDA

Univariate Analysis

Multivariate Analysis

Summary