Time Series Covid Project

Reagan Meagher and Megan Riley

#Modeling Covid19 Time Series

Covid19 is a worldwide pandemic that will likely define 2020. In the United States, currently over four million people have been infected and over 150,000 have died as a result of Covid19. As the pandemic continues, limiting infections, serious harm, and death is a primary concern for all involved.

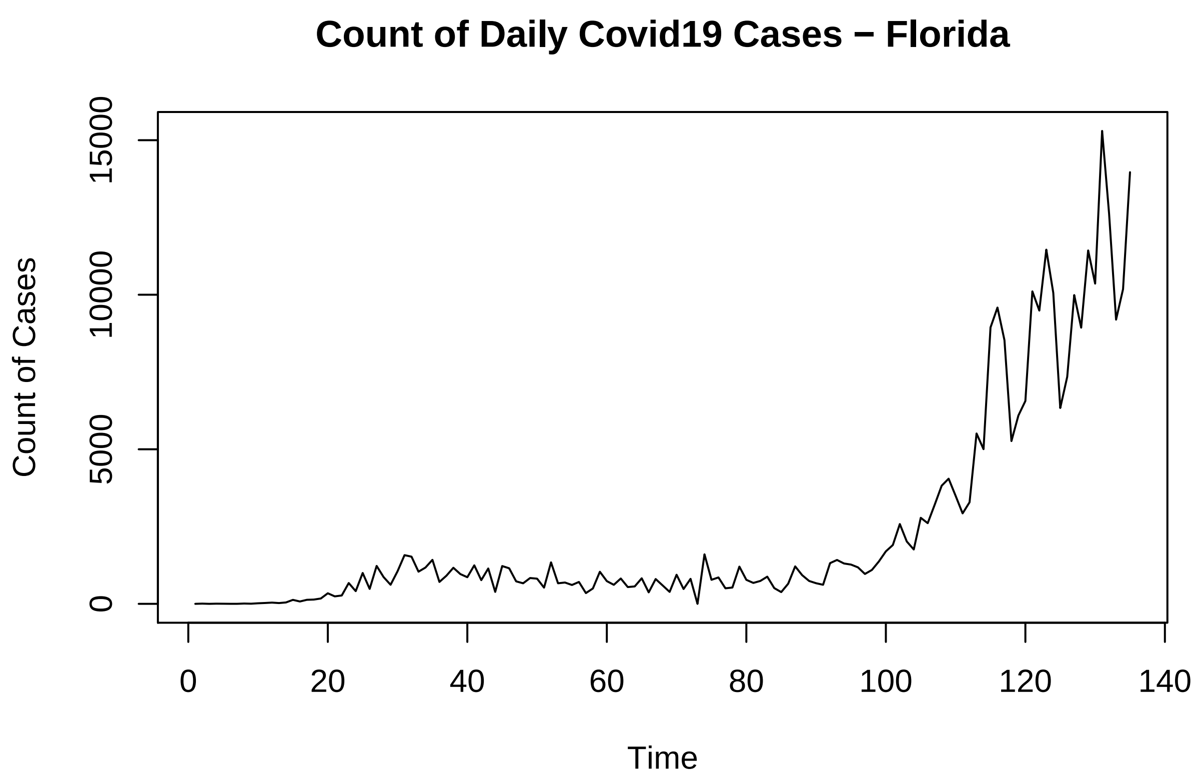
As this is a novel illness, we know relatively little, but understanding how Covid19 is spreading and judging the severity of an outbreak can be approximated with the data we have available. In this report we aim to build effective time series models to forecast future Covid19 cases using the techniques we have learned from this Time Series course. ##*Goal One: Data Collection*

The data source we are using is sourced from The Covid Tracking Project.

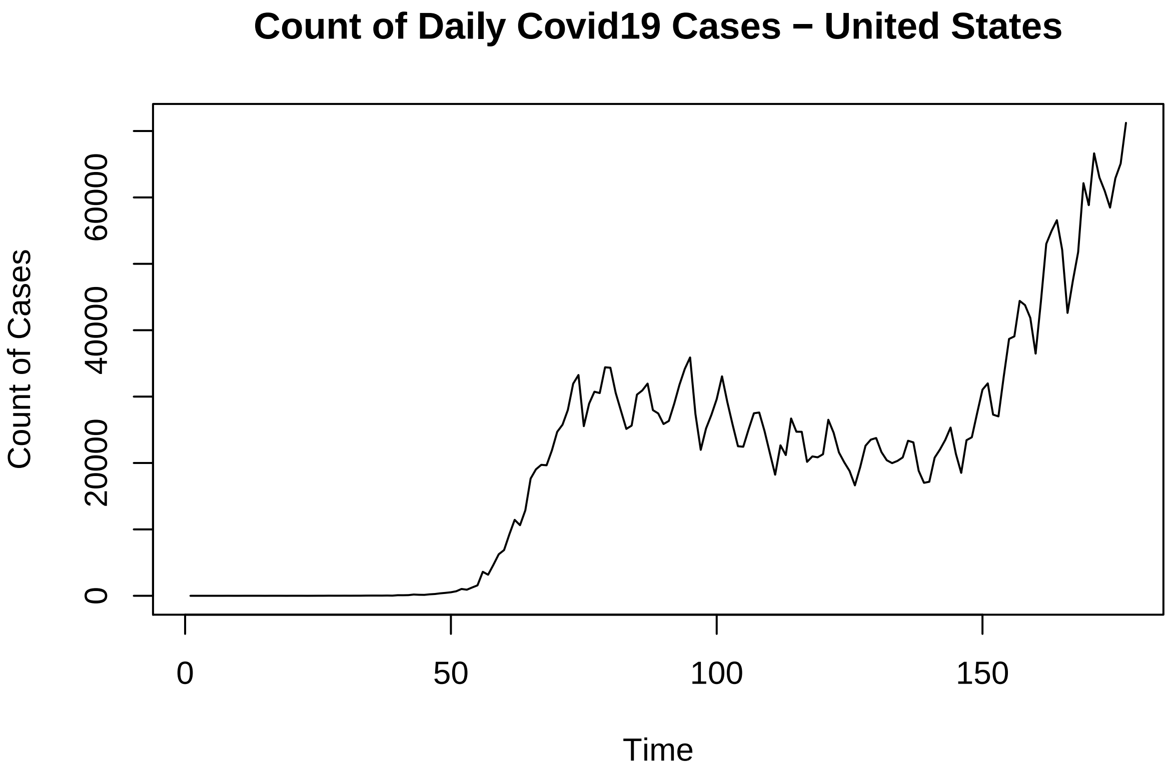
```
initial_data_fl <- read.csv(file="https://raw.githubusercontent.com/megnn/TimeSeries_Covid/master/covid
initial_data_us <- read.csv(file="https://raw.githubusercontent.com/megnn/TimeSeries_Covid/master/covid

initial_data_fl = initial_data_fl[order(nrow(initial_data_fl):1),]
initial_data_us = initial_data_us[order(nrow(initial_data_us):1),]

len_fl = dim(initial_data_fl)[1]
len_us = dim(initial_data_us)[1]
```

Below we plot the realizations of daily new cases from both Florida and the United States as a whole.

```
plot(x = seq(1,len_fl), y = initial_data_fl$positiveIncrease, type = "l", ylab = "Count of Cases", xlab
```



```r
plot(x = seq(1,len_us), y = initial_data_us$positiveIncrease, type = "l", ylab = "Count of Cases", xlab
```

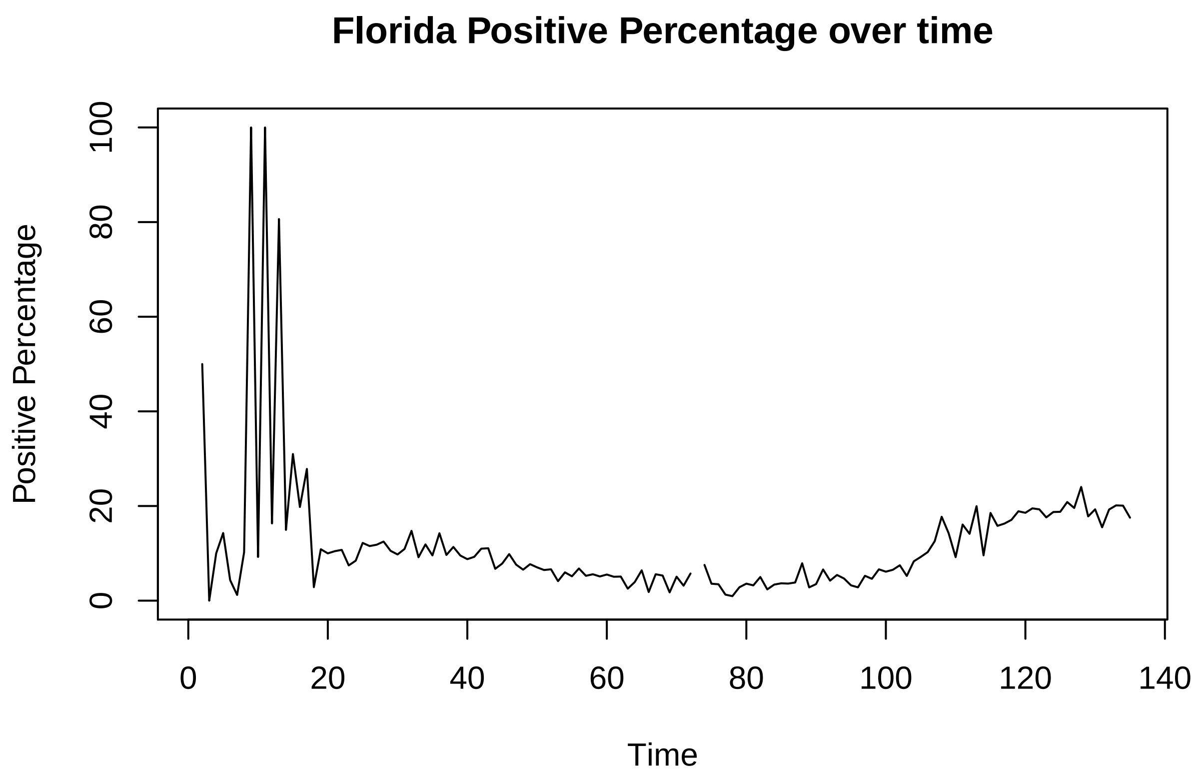


##*Positive Percentage*

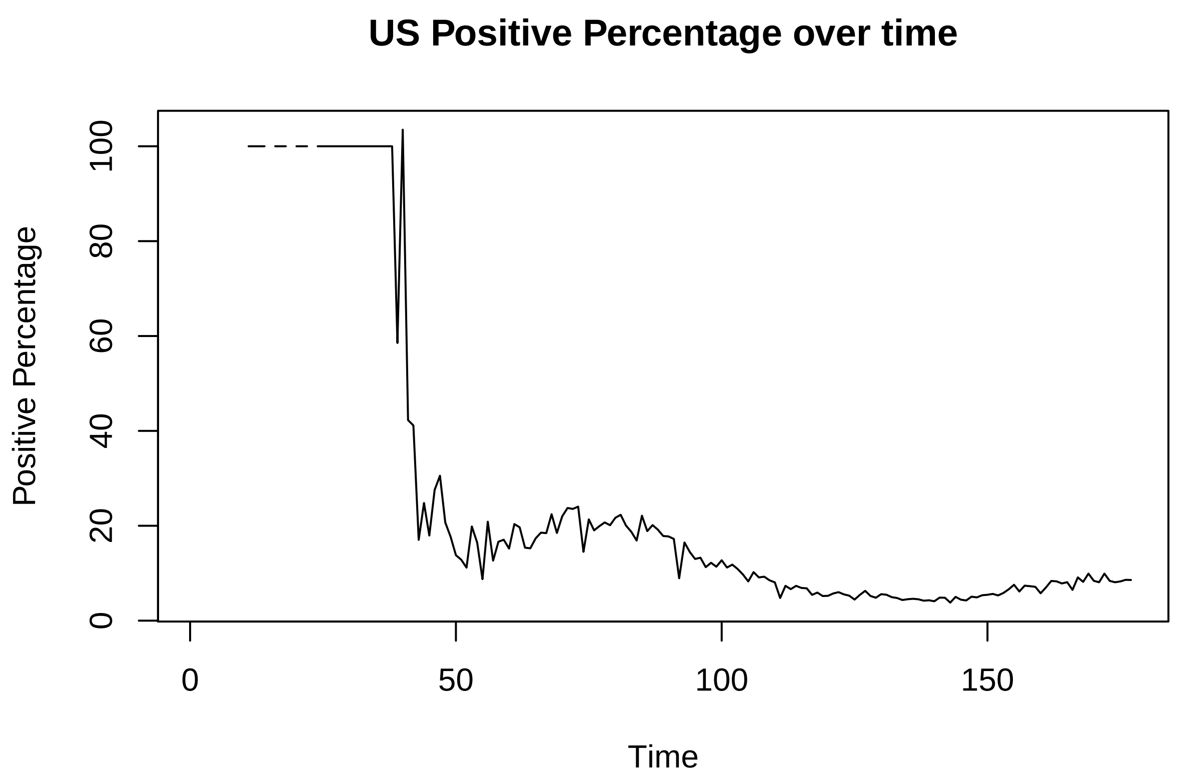
Positive Percentage is a statistic that calculates daily positive tests as a percentage of daily overall tests returned. We calculated this column and added it to our data below followed by some visual exploration of the statistic itself.

Overall we see a clear instance of high and often 100% positive test rates early on in the first days and weeks of the pandemic spread. We understand this as a result of the fact that Covid19 spread fast and we had more community spread than anticipated early on without the testing available. It is abundantly clear that when we have extremely high percent positive rates near 100% we can expect true positive case numbers at the time to be under represented. But without better epidemlogical understanding we can't make judgement calls on true case numbers when percent positives rise from 5% to 10% as we see begin to happen somewhat in recent days in Florida.

```r
for (i in 1:nrow(initial_data_fl)) {
  n <- round((initial_data_fl$positiveIncrease / initial_data_fl$totalTestResultsIncrease) * 100, digit
  initial_data_fl$positive_percentage <- n
}

for (i in 1:nrow(initial_data_us)) {
  n <- round((initial_data_us$positiveIncrease / initial_data_us$totalTestResultsIncrease) * 100, digit
  initial_data_us$positive_percentage <- n
}


#Percent Positive Exploration
plot(x = seq(1:len_fl), y = initial_data_fl$positive_percentage, type = "l", main = "Florida Positive P
```




```
plot(x = seq(1:len_us), y = initial_data_us$positive_percentage, type = "l", main = "US Positive Percen
```

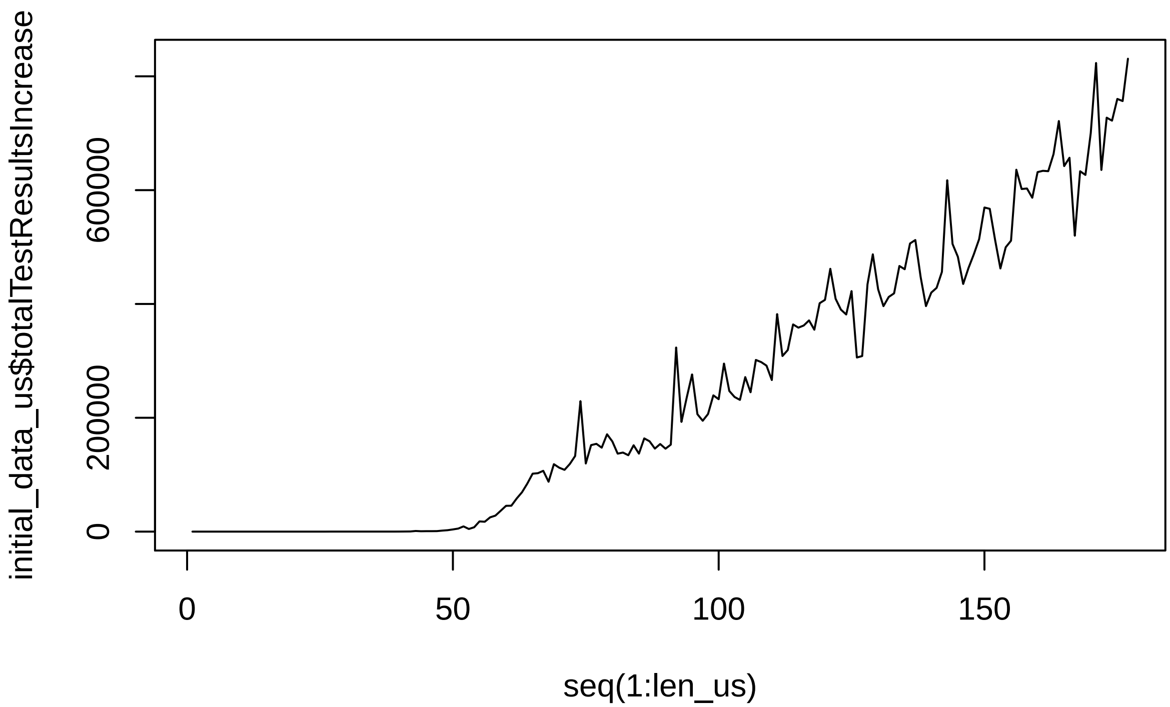


Positive Percentage as a metric is a measure of two main things, how many tests are we administering and how many positives are we receiving. If tests are skyrocketing while positive cases are increasing, we would see a stable or even diminishing line which could indicate not a pandemic under control but simply better testing resources but could be interpreted as a pandemic managed.

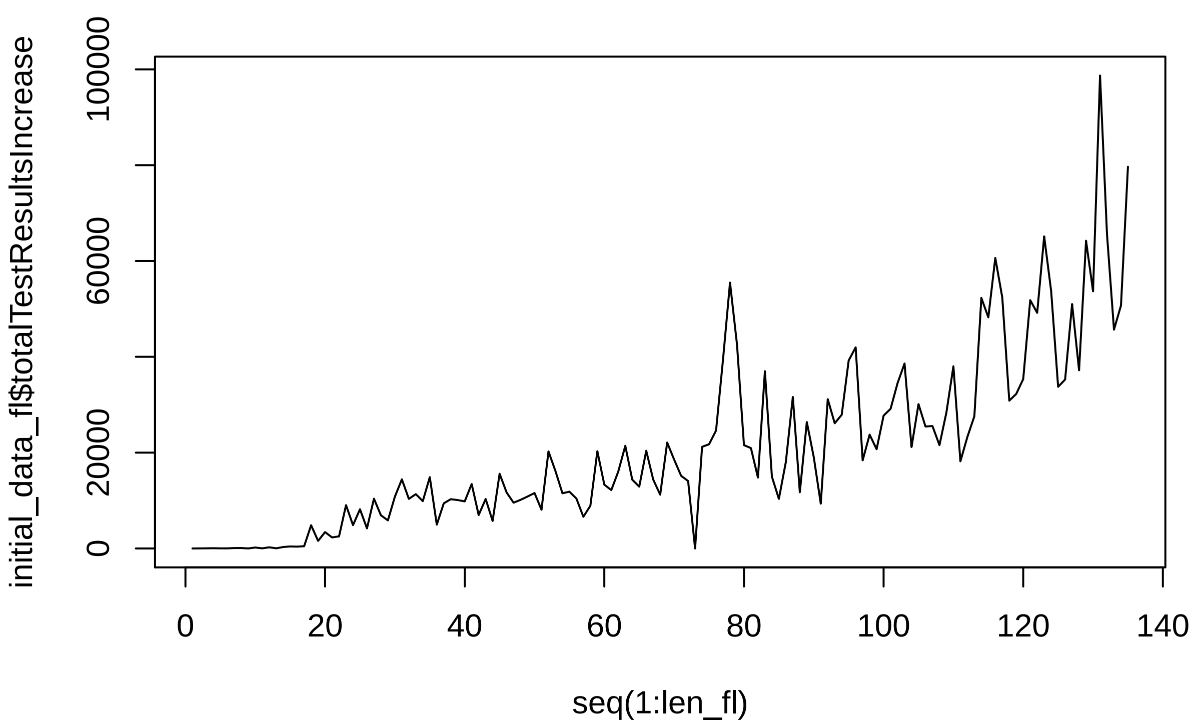
Keeping tests increasing to continue to keep percent positives level is a good indication we have leveled up our resources to continue to diagnose the pandemic at the same level, but if we need to scale up our testing to keep the same positive percentage, there is more covid spread.

However, an increasing positive percentage is a good indicator that our testing resources may not be up to actually up to tracking the current stage of the pandemic.

```
plot(x = seq(1:len_us), y = initial_data_us$totalTestResultsIncrease, type = "l")
```




```
plot(x = seq(1:len_fl), y = initial_data_fl$totalTestResultsIncrease, type = "l")
```



####Data Preperation

In order to model new case numbers by day we set up dataframes with only our date and positive increase amount per day.

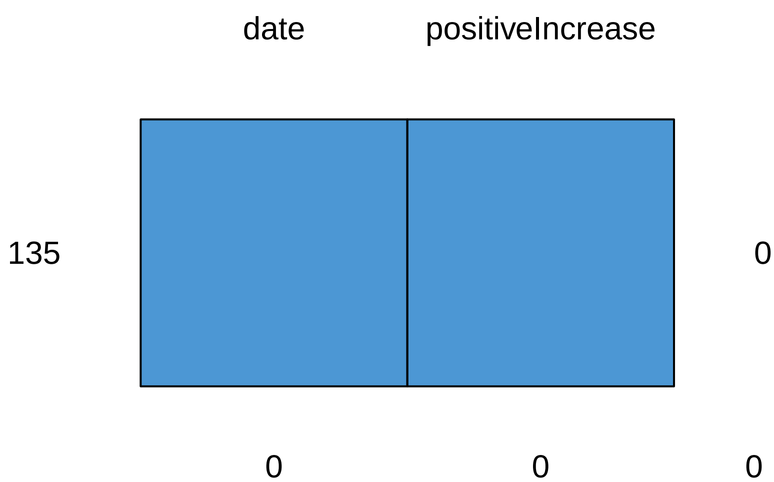
```
newcases_fl <- dplyr::select(initial_data_fl, c("date", "positiveIncrease"))
newcases_us <- dplyr::select(initial_data_us, c("date", "positiveIncrease"))
```

####Checking for NAs

We can see with the missing value analysis below that we have no NAs present in our new case data.

```
#Checking for NAs
md.pattern(newcases_fl)
```

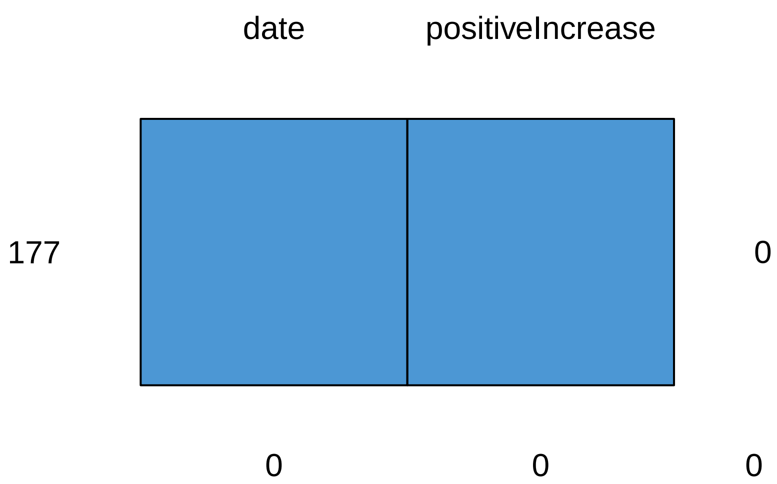
```
##  /\     /\
## {  `---'  }
## {  O   O  }
## ==>  V <==  No need for mice. This data set is completely observed.
##  \  \|/  /
##   `-----'
```




```
##     date positiveIncrease
## 135    1                1 0
##        0                0 0
```

```
# No NAs present


#Checking for NAs
md.pattern(newcases_us)
```

```
##  /\     /\
## {  `---'  }
## {  O   O  }
## ==>  V <==  No need for mice. This data set is completely observed.
##  \  \|/  /
##   `-----'
```

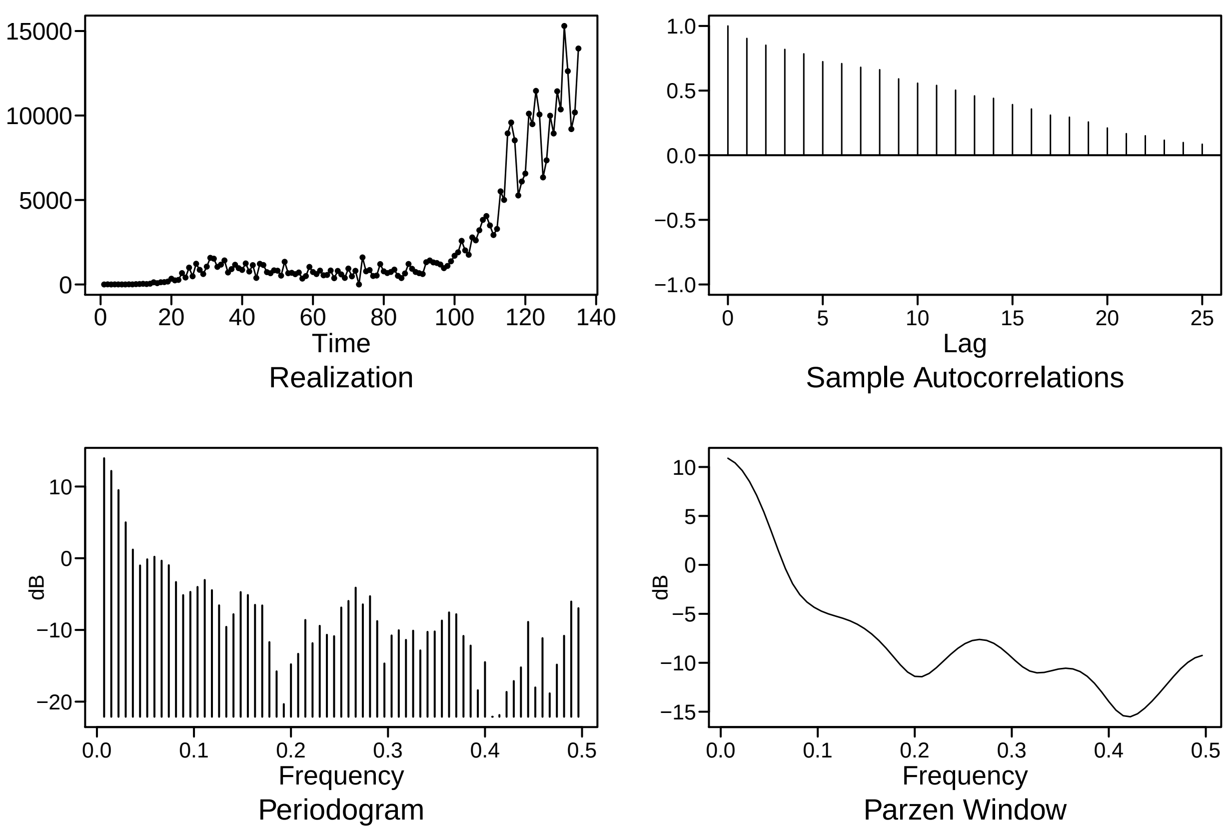



```
##      date positiveIncrease
## 177     1                1 0
##         0                0 0
```

```
#No NAs present
```

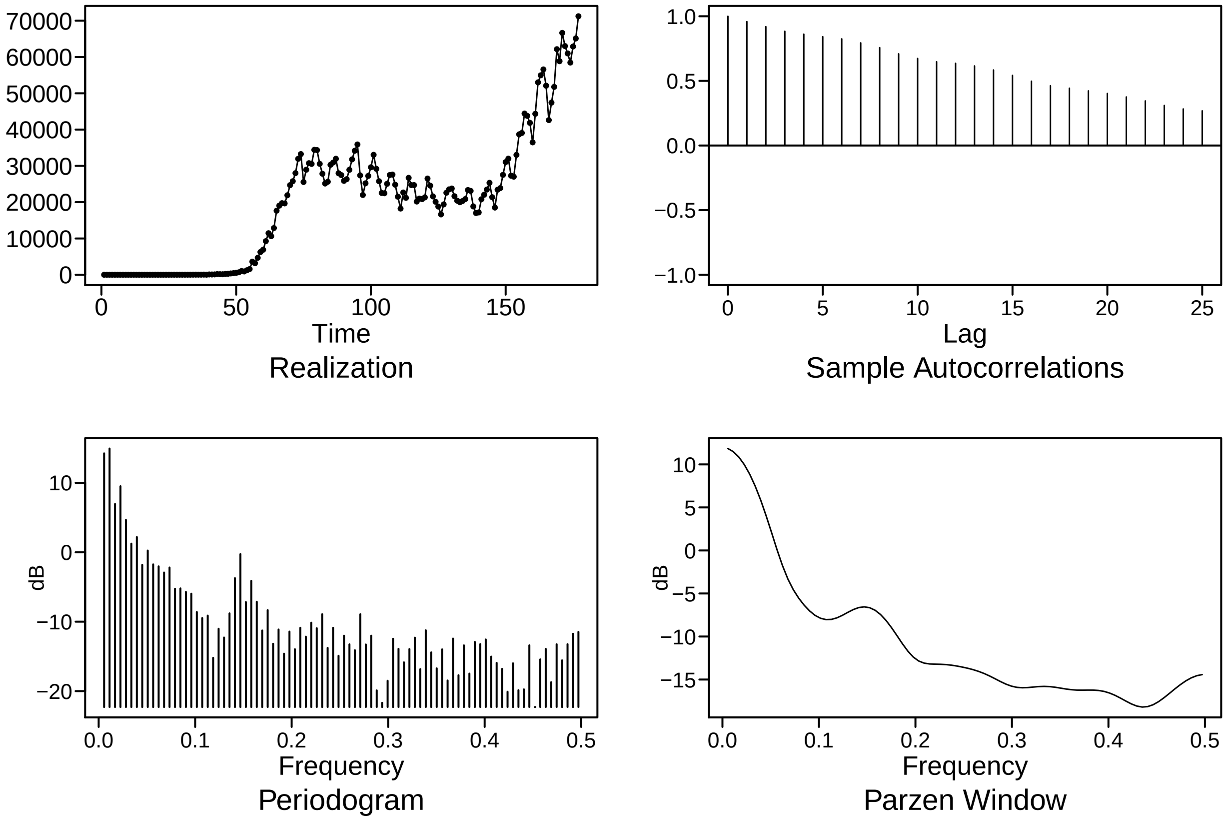
Florida Daily Cases:

```
#no text output, just the plots
x = plotts.sample.wge(newcases_fl$positiveIncrease)
```



###**US Daily Cases:**

```r
#no text output
x = plotts.sample.wge(newcases_us$positiveIncrease)
```




```r
dev.off()
plot(x = seq(1,len_us), y = newcases_us$positiveIncrease, type = "l", main = 'United States Covid-19 Da
```

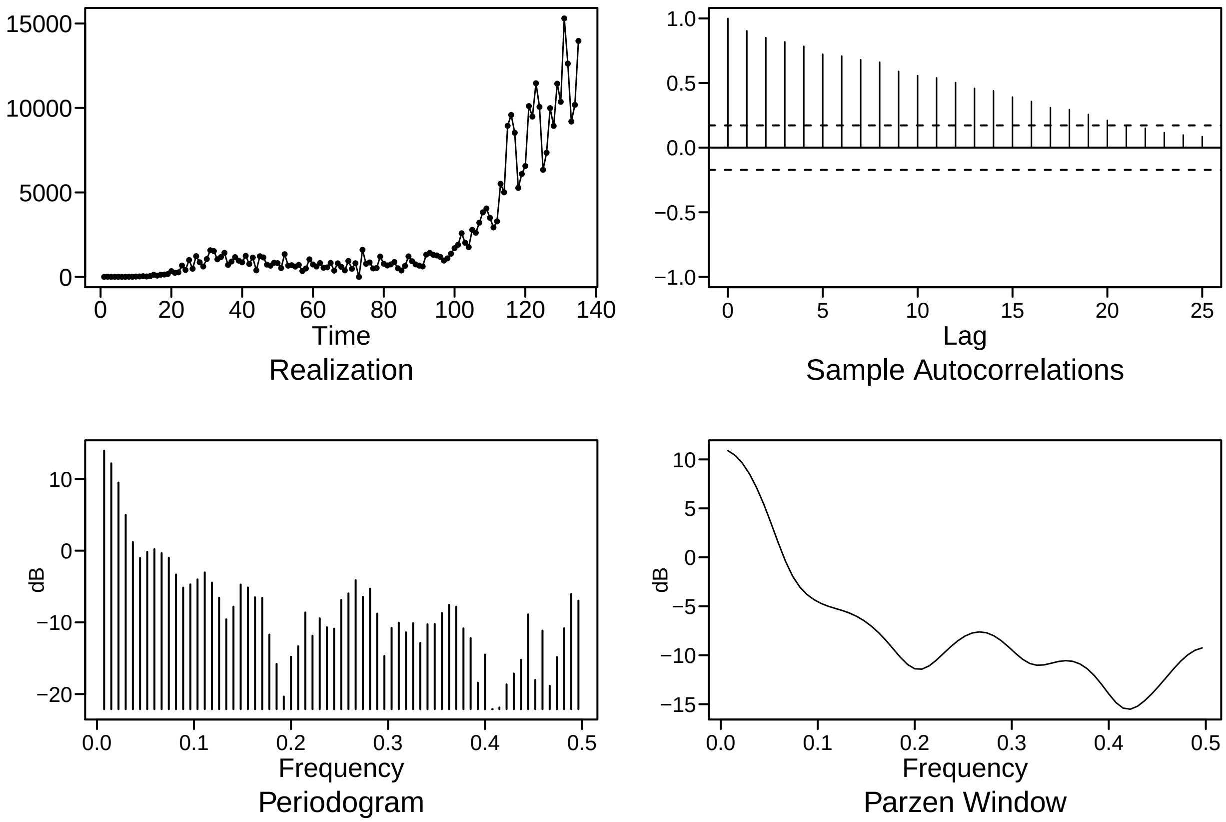
##*Goal Two: Univariate Analysis*

###Model Building for Cases in Florida

A. Stationarity vs Non-Stationarity

Overall we see slowly dampening ACFs, combined with a strong aperiodic frequency at zero in our spectral density. These measures alone with a recently quickly rising case count in recent days gives us strong evidence that our data is non-stationary. Given Covid19 spread, it is likely we see continued rising behavior in the short term, some return to lower numbers in the coming months but more uncertainty as new spikes could arise, and in the longest term of years on, we expect new cases to diminish to zero once the pandemic has ended spread.

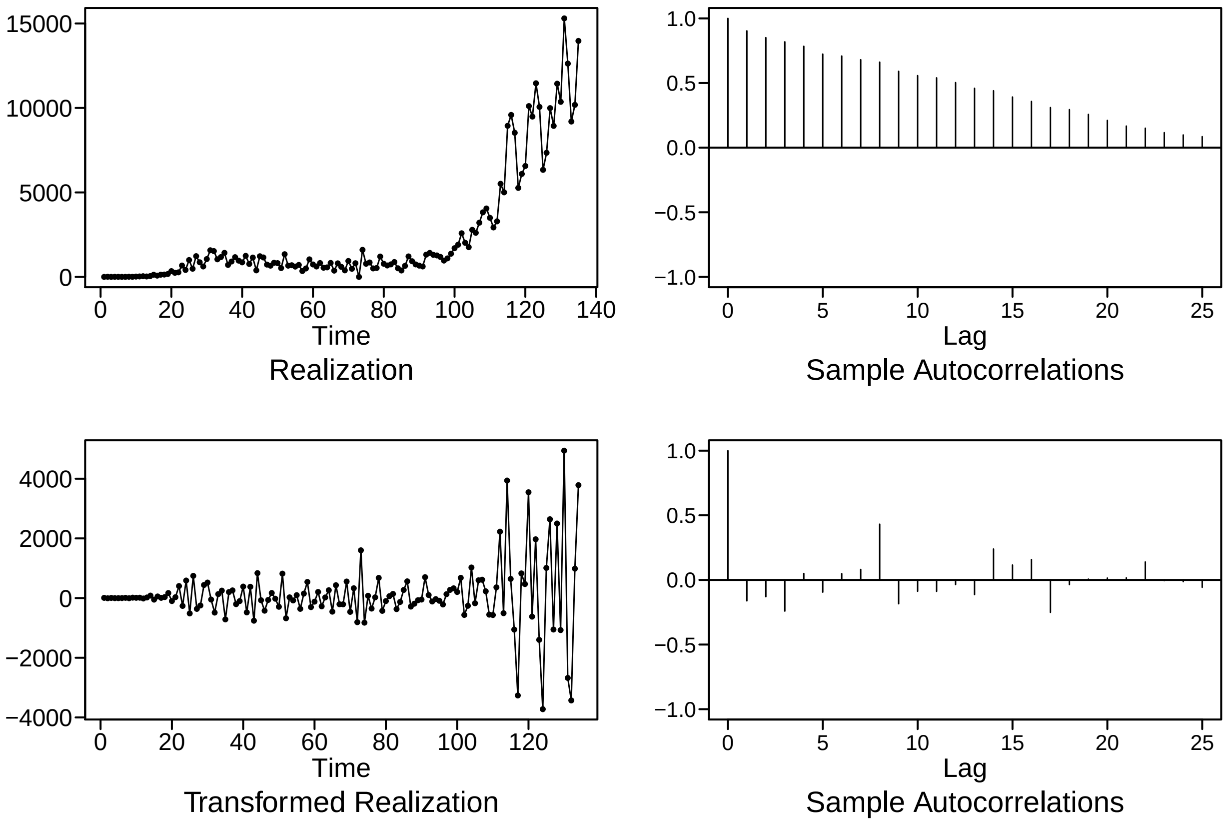
```r
x = plotts.sample.wge(newcases_fl$positiveIncrease, arlimits = TRUE)
```



B. Non Stationary Modeling

We did not do any differencing of our data set to account for this non-stationarity. Going into this project we knew that because of the failure to contain the Covid-19 outbreak we would see large spikes of cases in recent time periods compared to distant time periods. We feel that this is an important aspect of our data that we want to portray in our models because we can see empirically in Florida and the United States as a whole that both individual behavior and political policy continue to trend towards further outbreak and rapid, almost exponential daily case growth. While some states with compliant individual behavior and strong political Covid-19 policies have shown "completed" Covid-19 curves, where daily case count begins to trend downwards towards zero, Florida is the opposite. Therefore, since we empirically expect the trend of non-stationarity to continue, we want that represented in our models. This is a fundamental assumption that our models are built on.

```
diff_fl = artrans.wge(newcases_fl$positiveIncrease, 1)
```



C. Model IDing of stationary models

For ARIMA models we identify the stationary components below.

```
#Model Differenced Data
aic5.wge(diff_fl)
```

```
## ---------WORKING... PLEASE WAIT...
##
##
## Error in aic calculation at 3 2
## Error in aic calculation at 4 2
## Error in aic calculation at 5 2
## Five Smallest Values of  aic

##       p    q        aic
## 9     2    2   13.71539
## 16    5    0   13.82831
## 17    5    1   13.83000
## 11    3    1   13.84291
## 14    4    1   13.85759
```

```
aic5.wge(diff_fl, type = "bic") #2,2 produced
```

```
## ---------WORKING... PLEASE WAIT...
##
##
## Error in aic calculation at 3 2
## Error in aic calculation at 4 2
## Error in aic calculation at 5 2
```

```
## Five Smallest Values of  bic
##       p    q        bic
## 9     2    2   13.82352
## 3     0    2   13.94850
## 11    3    1   13.95104
## 10    3    0   13.95289
## 5     1    1   13.95515
```

```r
#Modeling original data if stationary
aic5.wge(newcases_fl$positiveIncrease)
```

```
## ---------WORKING... PLEASE WAIT...
##
##
## Error in aic calculation at 1 1
## Error in aic calculation at 1 2
## Error in aic calculation at 2 0
## Error in aic calculation at 2 2
## Error in aic calculation at 3 0
## Error in aic calculation at 3 1
## Error in aic calculation at 3 2
## Error in aic calculation at 4 0
## Error in aic calculation at 4 1
## Error in aic calculation at 4 2
## Error in aic calculation at 5 0
## Error in aic calculation at 5 1
## Error in aic calculation at 5 2
## Five Smallest Values of  aic
```

```
##      p    q        aic
## 8    2    1   13.97724
## 4    1    0   14.01245
## 3    0    2   14.80561
## 2    0    1   15.43958
## 1    0    0   16.27428
```

```r
aic5.wge(newcases_fl$positiveIncrease, type = 'bic')
```

```
## ---------WORKING... PLEASE WAIT...
##
##
## Error in aic calculation at 1 1
## Error in aic calculation at 1 2
## Error in aic calculation at 2 0
## Error in aic calculation at 2 2
## Error in aic calculation at 3 0
## Error in aic calculation at 3 1
## Error in aic calculation at 3 2
## Error in aic calculation at 4 0
## Error in aic calculation at 4 1
## Error in aic calculation at 4 2
## Error in aic calculation at 5 0
## Error in aic calculation at 5 1
## Error in aic calculation at 5 2
## Five Smallest Values of  bic
```

```
##      p    q        bic
## 4    1    0   14.05549
## 8    2    1   14.06332
## 3    0    2   14.87017
## 2    0    1   15.48262
## 1    0    0   16.29580
```

D. Model Building

Florida Cases - ARIMA Model

```r
fl_arima = est.arma.wge(diff_fl, p = 2, q = 2)
```

```
##
## Coefficients of Original polynomial:
## 1.1966 -0.7536
##
## Factor                 Roots                Abs Recip    System Freq
## 1-1.1966B+0.7536B^2    0.7939+-0.8346i      0.8681       0.1290
##
##
```

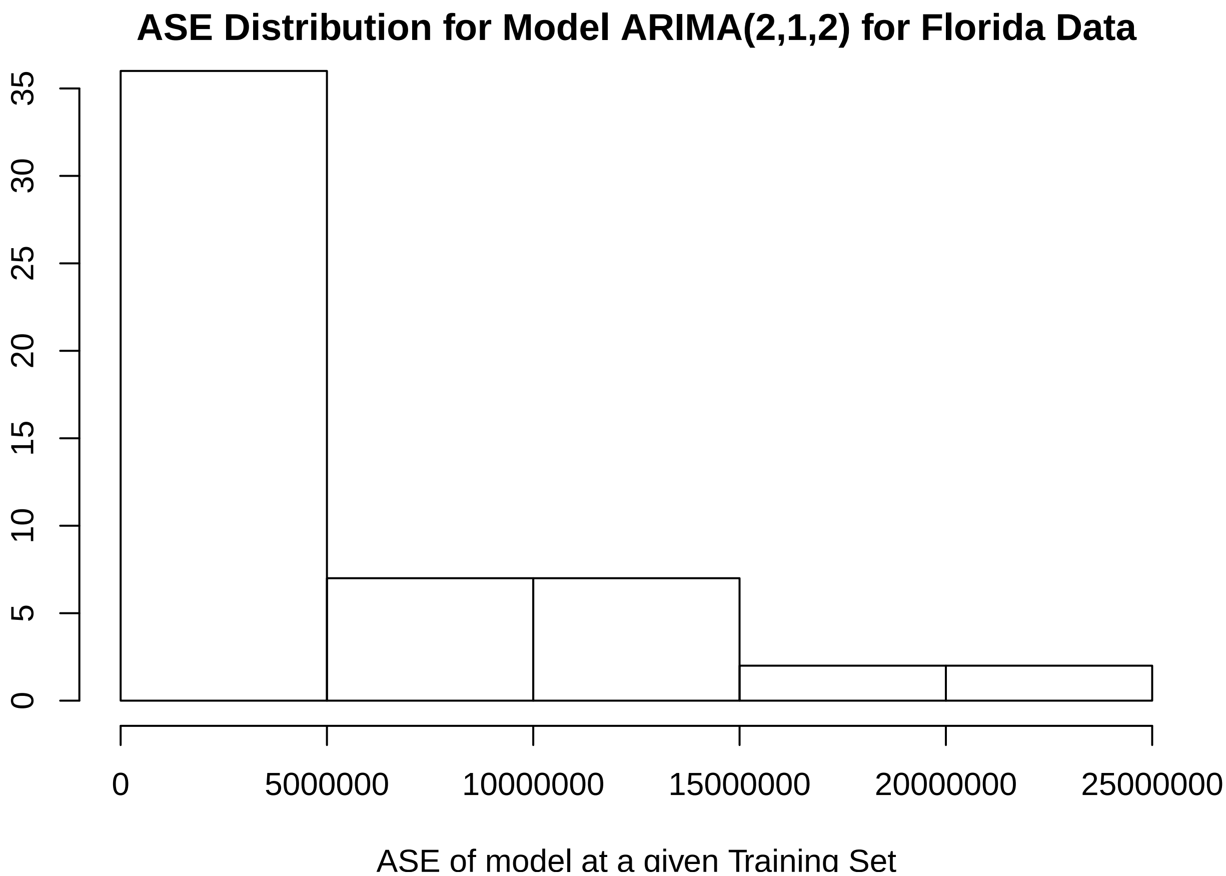
```r
fl_arima$aic
```

```
## [1] 13.71539
```

```r
trainingSize = 70
horizon = 12
ASEHolder = numeric()

for( i in 1:(135-(trainingSize + horizon) + 1))
{

  forecasts = fore.aruma.wge(newcases_fl$positiveIncrease[i:(i+(trainingSize-1))],phi = fl_arima$phi, th

  ASE = mean((newcases_fl$positiveIncrease[(trainingSize+i):(trainingSize+ i + (horizon) - 1)] - foreca
  ASEHolder[i] = ASE

}

ASEHolder
```

```
##  [1]   173447.49   229133.67   168384.33   341904.83   191371.33
##  [6]    66746.55   106055.45    55793.09    51760.38    98347.57
## [11]    69300.69   139565.30   172736.75   171867.75   247073.12
## [16]   318483.37   249008.34    79288.60   153567.66   232083.96
## [21]   465674.63   679351.45   377844.85   576114.33   618238.92
## [26]   859790.13  1264646.77  2049131.44  2573862.78  2774562.47
## [31]  2726131.29  3450072.02  2727073.18  6406627.51 10231802.96
## [36] 10857622.78 11399600.50 10532859.25  9384763.58 11294574.46
## [41] 14138091.79 20793805.72 23521406.71 17866340.71 18388442.48
## [46]  5099762.18  1515023.47  1884767.08  1726135.90  5054543.44
## [51]  9125960.97  6894983.65 10460145.72  7559902.47
```

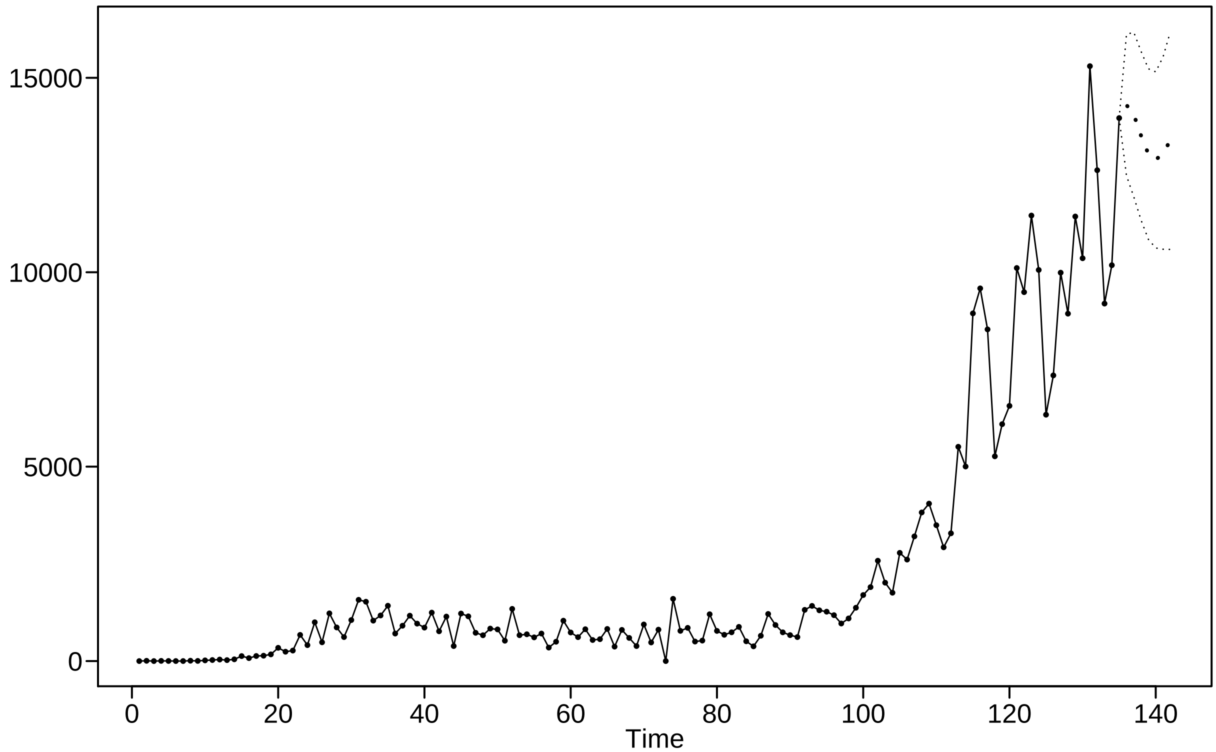
```r
#Distribution of ASEs on Two Week Periods
hist(ASEHolder, xlab = "ASE of model at a given Training Set",  main = "ASE Distribution for Model ARIM
```



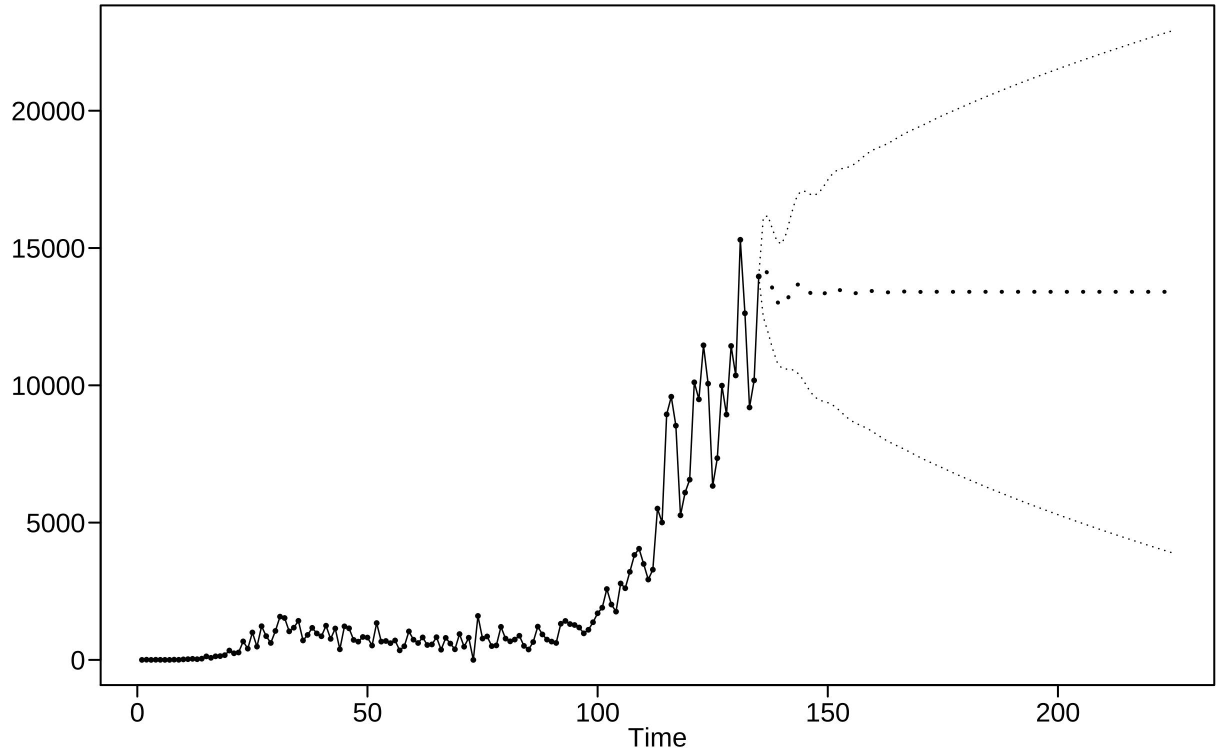

```r
#Mean ASE
WindowedASE = mean(ASEHolder)
WindowedASE
```

```
## [1] 4418437
```

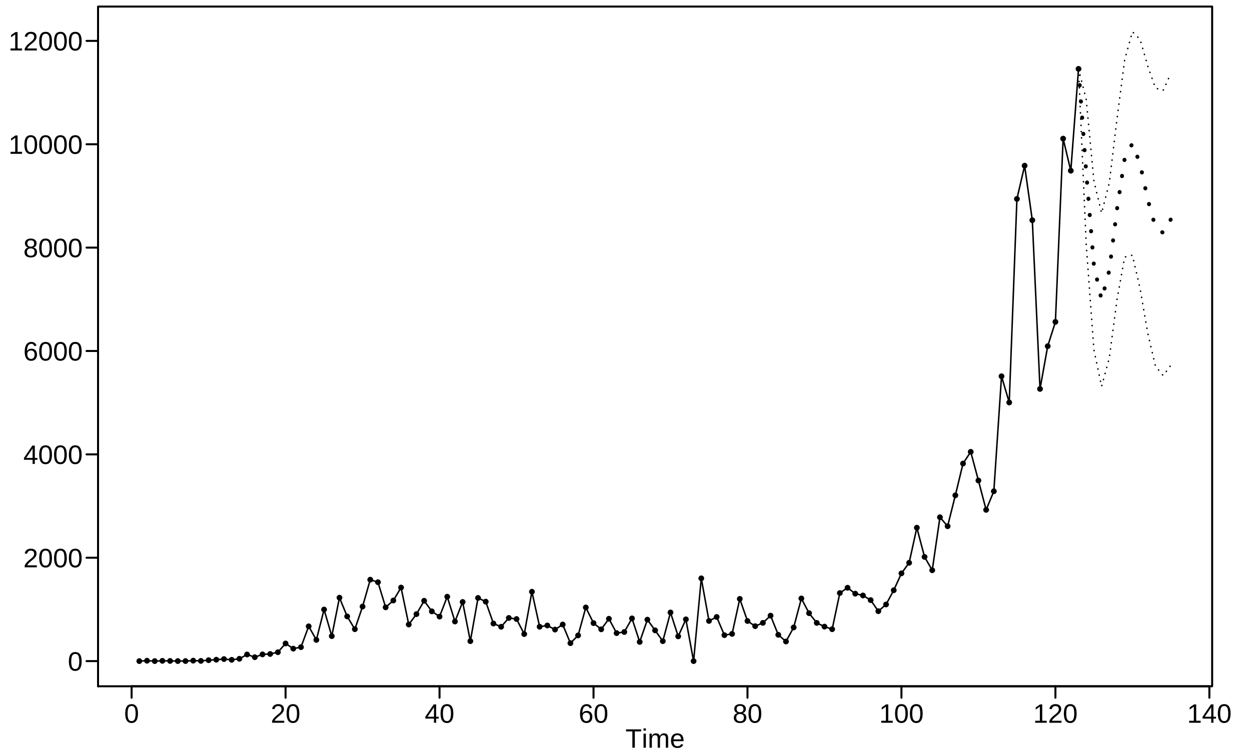
```r
##



short_fl_arima = fore.aruma.wge(newcases_fl$positiveIncrease,phi = fl_arima$phi,theta = fl_arima$theta,
```



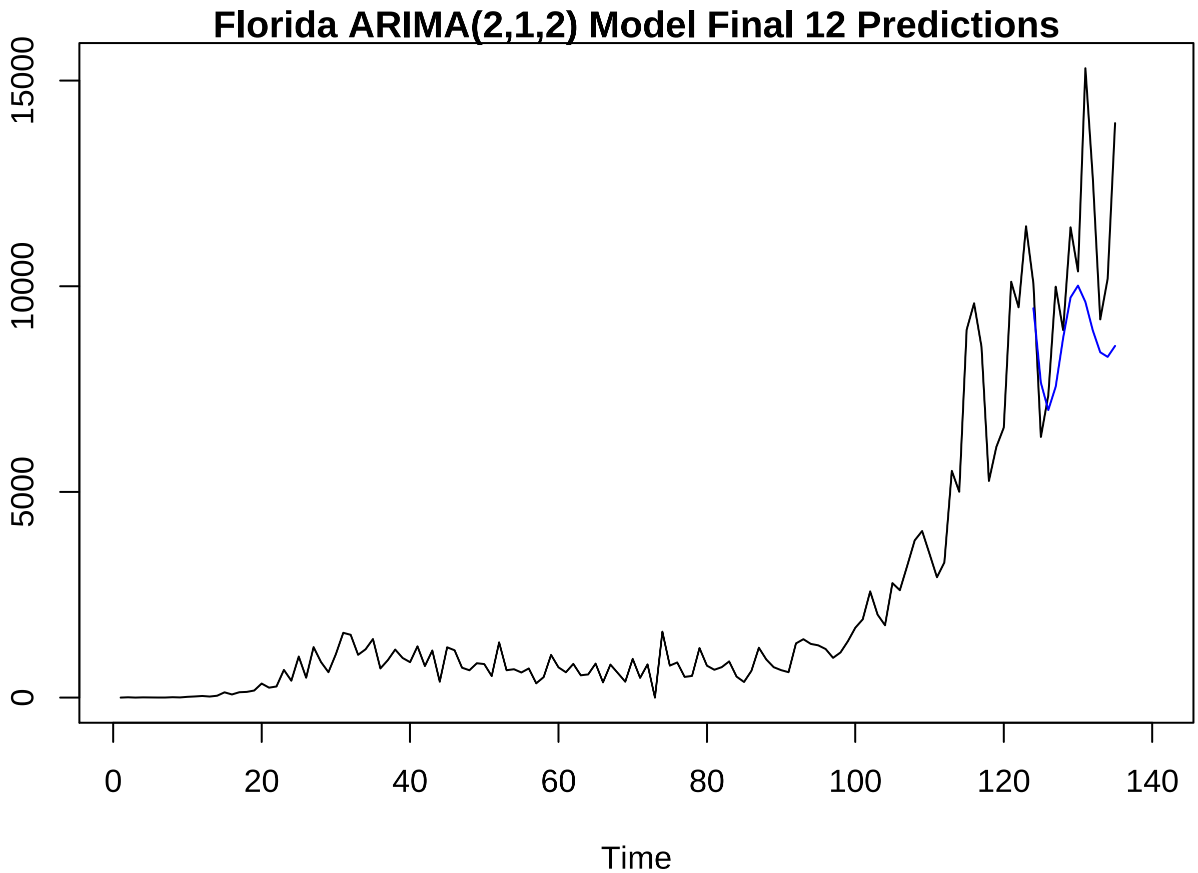

```
long_fl_arima = fore.aruma.wge(newcases_fl$positiveIncrease,phi = fl_arima$phi,theta = fl_arima$theta, c
```




```
final_pred = fore.aruma.wge(newcases_fl$positiveIncrease[1:123],phi = fl_arima$phi,theta = fl_arima$thet
```




```
final_pred_df = data.frame(t = seq(124:135), final_pred$f)

plot(newcases_fl$positiveIncrease, type = "l", ylab = "Count of New Cases", xlab = "Time", main = "Flor
lines(ts(final_pred$f, start = 124, end = 135),  col = "blue")
```




```
final_ASE = mean((newcases_fl$positiveIncrease[124:135] - final_pred$f)^2)
final_ASE
```

```
## [1] 7562545
```

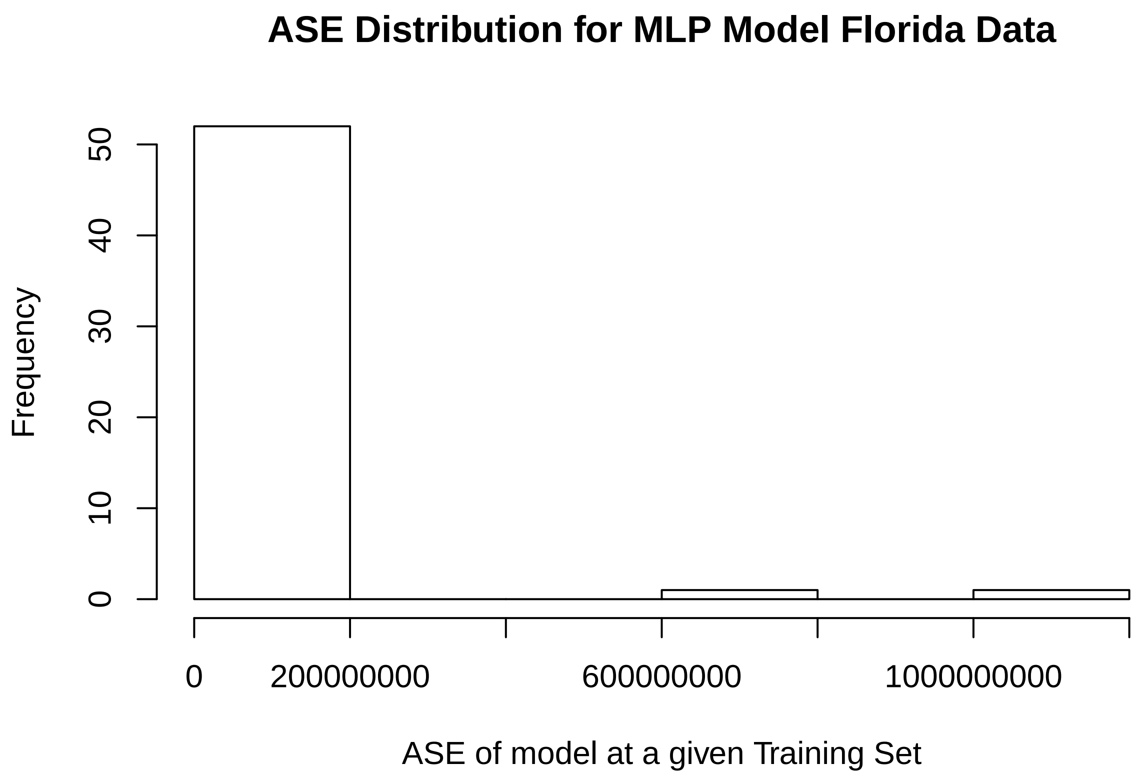
####MLP Model for Florida Cases

```
trainingSize = 70
horizon = 12
ASEHolder = numeric()


for( i in 1:(135-(trainingSize + horizon) + 1))
{
  mlp.fit = mlp(ts(newcases_fl$positiveIncrease[1:trainingSize+i]), hd = 5, comb = "median")
  forecasts = forecast(mlp.fit,h = horizon)

  ASE = mean((newcases_fl$positiveIncrease[(trainingSize+i):(trainingSize+ i + (horizon) - 1)] -forecas

  ASEHolder[i] = ASE

}

ASEHolder
```

```
##  [1]   143763.04   124054.34   278614.85   160612.19    65987.22
##  [6]    90854.71    70154.93    85186.05    59639.47    63197.74
## [11]   117068.83   101686.93   110588.72   162454.92   225414.41
```

```
## [16]      295904.65      184371.61      135652.12      143280.42      283467.27
## [21]     1121047.25      620038.17      649608.42      895662.35      896700.96
## [26]      786773.42     1870488.66     2630511.25     3060536.54     2350241.29
## [31]     2310265.12     1579835.32     2739523.67     7000339.51    10742667.16
## [36]    11694236.48    11091932.56     7334210.01     6642625.92    12055258.26
## [41]    17425464.88    21775454.59    16359541.14    11186050.41   636419916.19
## [46] 1175130786.81     5427954.55    33135286.07    15345235.31    25459893.02
## [51]    15272787.02    15055502.44     8893742.66     5273119.84
```

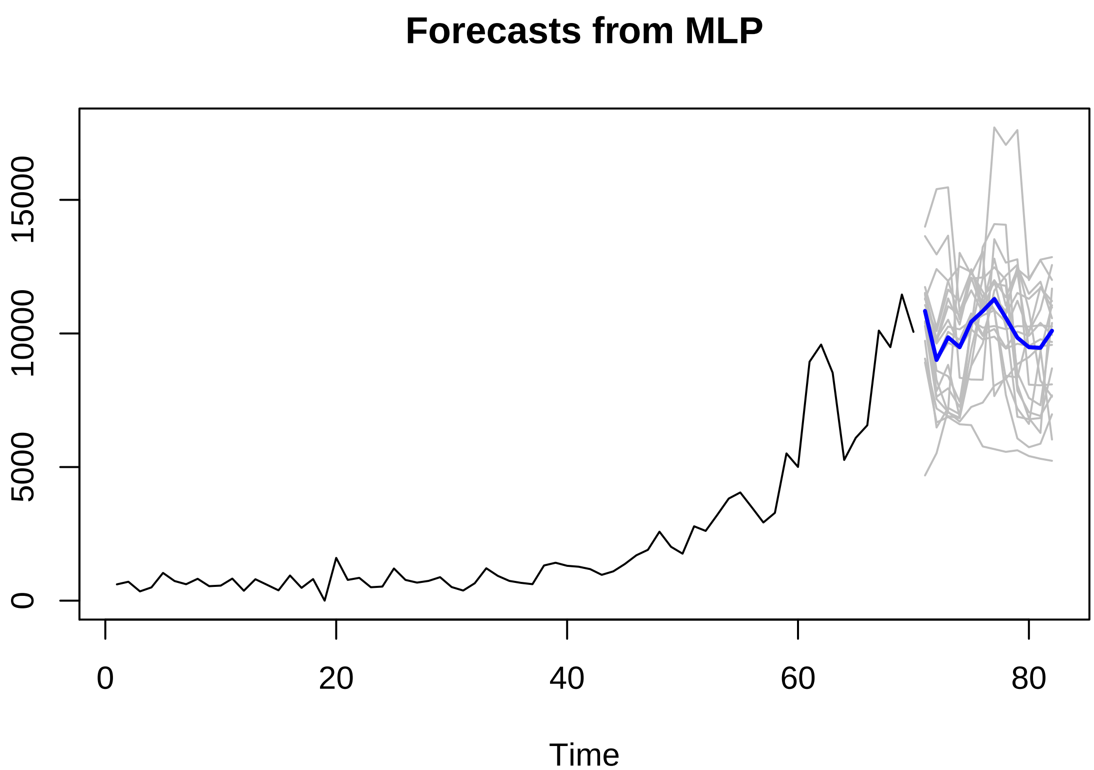
```r
#Distribution of ASEs on Two Week Periods
hist(ASEHolder, xlab = "ASE of model at a given Training Set",  main = "ASE Distribution for MLP Model F
```




```r
#Mean ASE
WindowedASE = mean(ASEHolder)
WindowedASE
```

```
## [1] 38761763
```

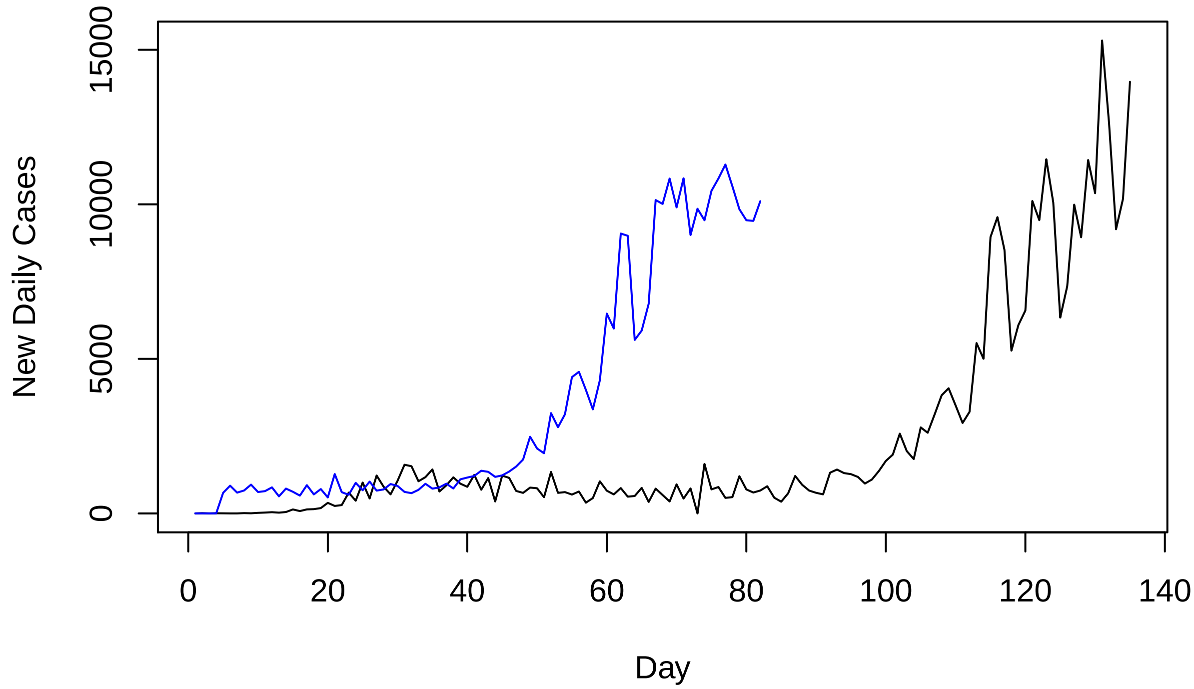
```r
plot(forecasts)
```



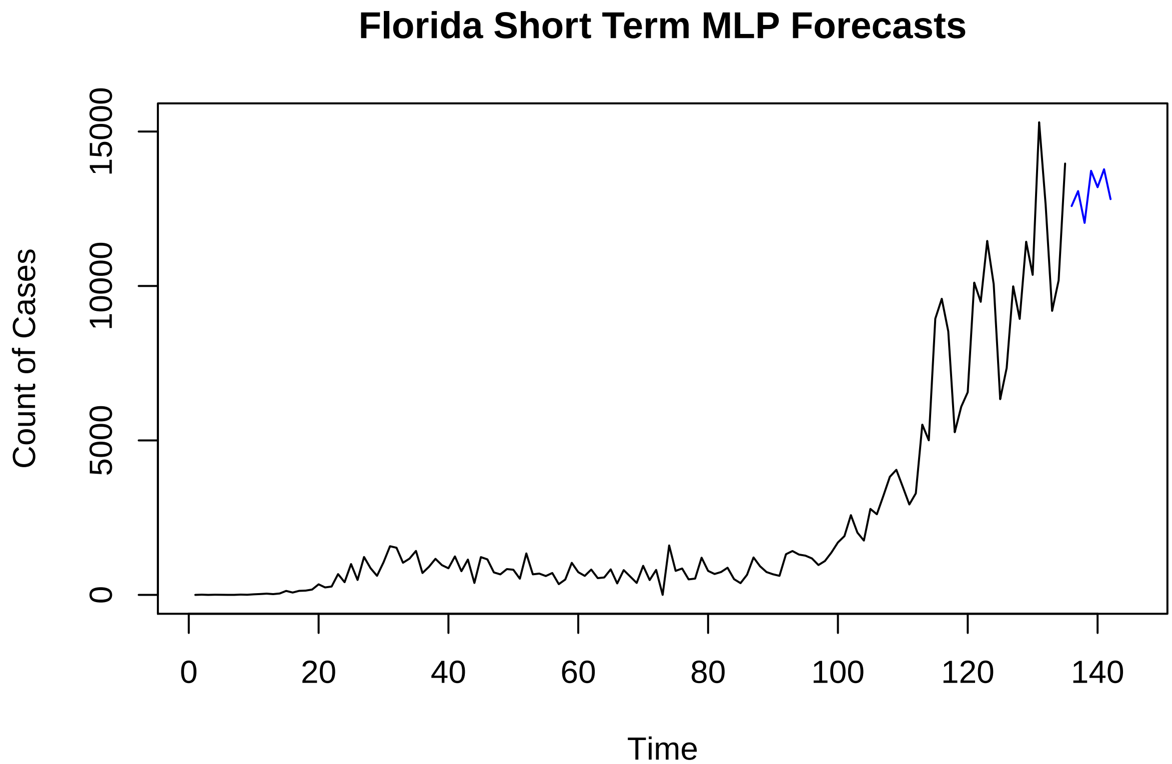

```r
#Final Forecasts with data known
mlp.fit_fl_final = mlp(ts(newcases_fl$positiveIncrease[1:123]), hd = 5, comb = "median")
forecasts_fl_mlp = forecast(mlp.fit,h = 12)

final12_ase = mean((newcases_fl$positiveIncrease[124:135] -forecasts_fl_mlp$mean)^2)
final12_ase
```

```
## [1] 5273120
```

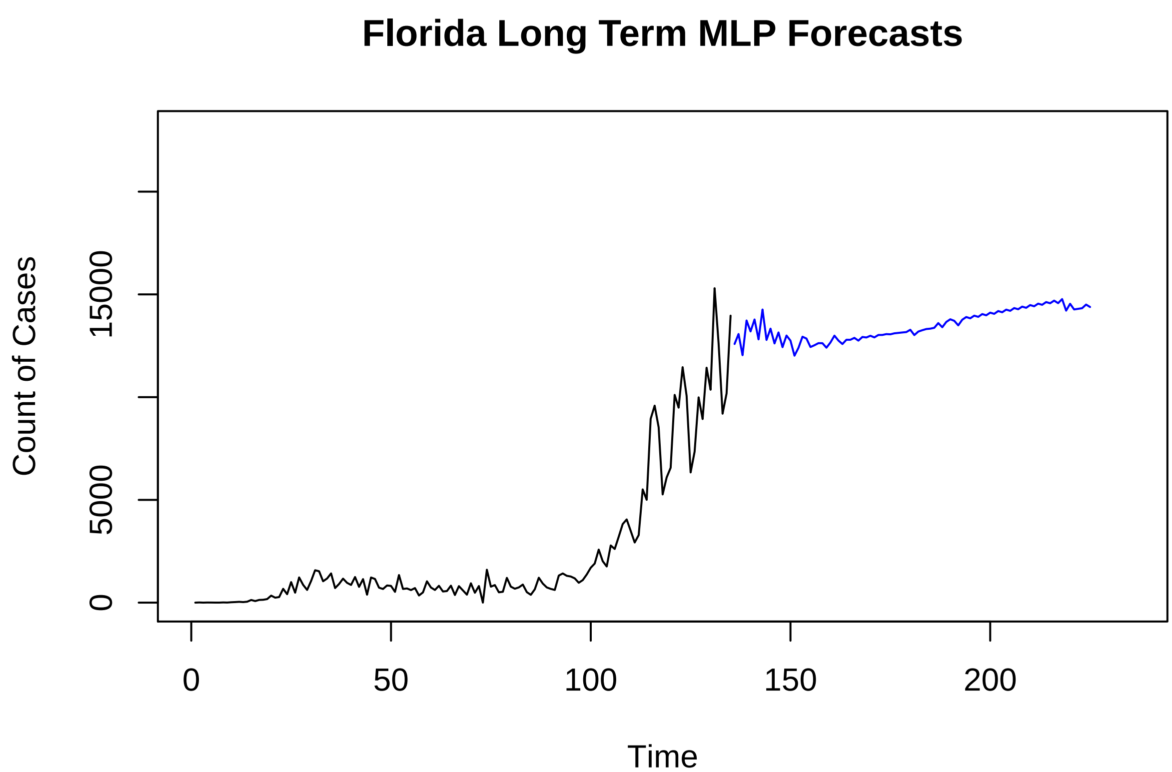
```r
all_f = c(rep(1,4), forecasts$fitted, forecasts$mean)
plot(newcases_fl$positiveIncrease, type = "l", ylab = "New Daily Cases", xlab = "Day", main = "")
lines(all_f, col = "blue")
```




```
#Future Forecasts

mlp.fit_fl_future =mlp(ts(newcases_fl$positiveIncrease), hd = 5, comb = "median")

short_fl_mlp = forecast(mlp.fit_fl_future,h = 7)
long_fl_mlp = forecast(mlp.fit_fl_future,h = 90)

plot(newcases_fl$positiveIncrease, type = "l", xlim = c(1,145), main = "Florida Short Term MLP Forecasts
lines(short_fl_mlp$mean, col = "blue")
```




```
plot(newcases_fl$positiveIncrease, type = "l", xlim = c(1,235), ylim = c(0,23000), main = "Florida Long
lines(long_fl_mlp$mean, col = "blue")
```

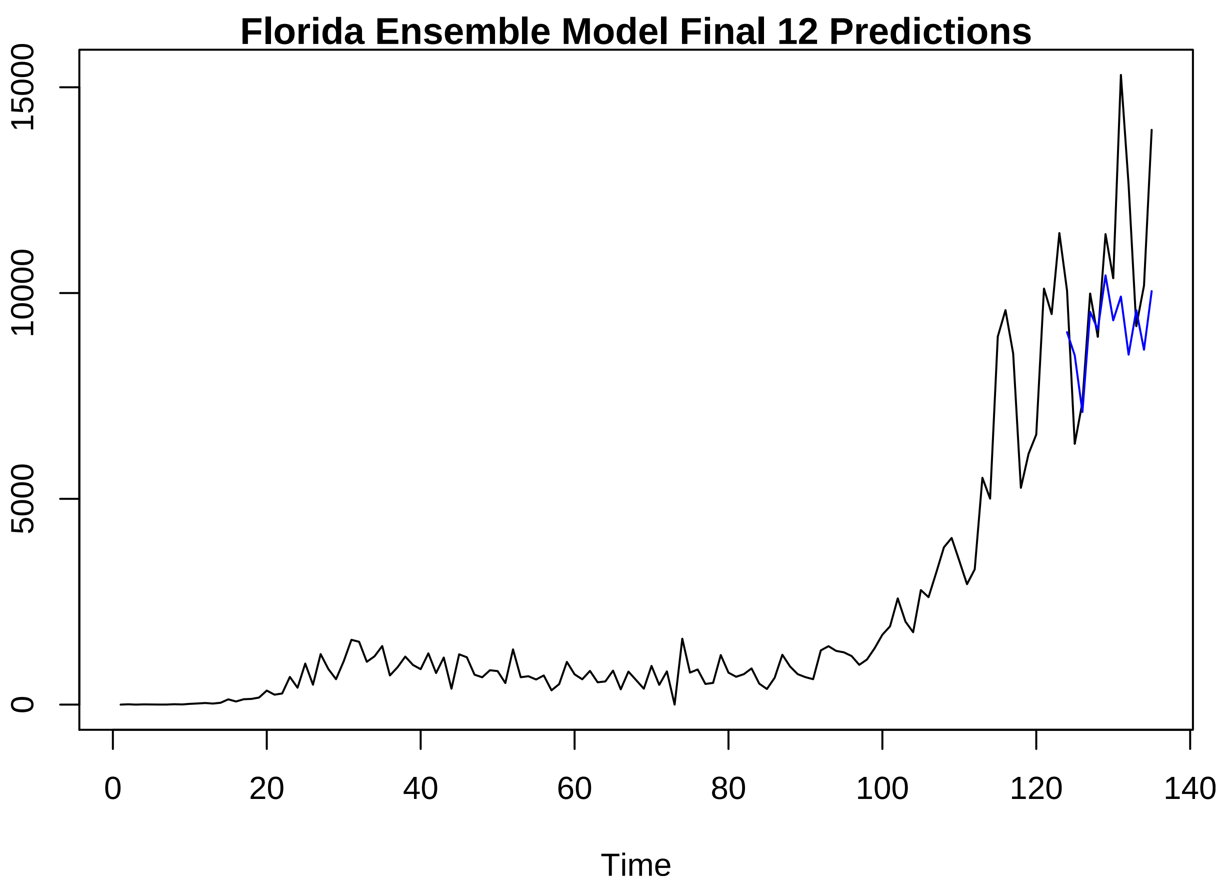


####Florida Cases Ensemble Model

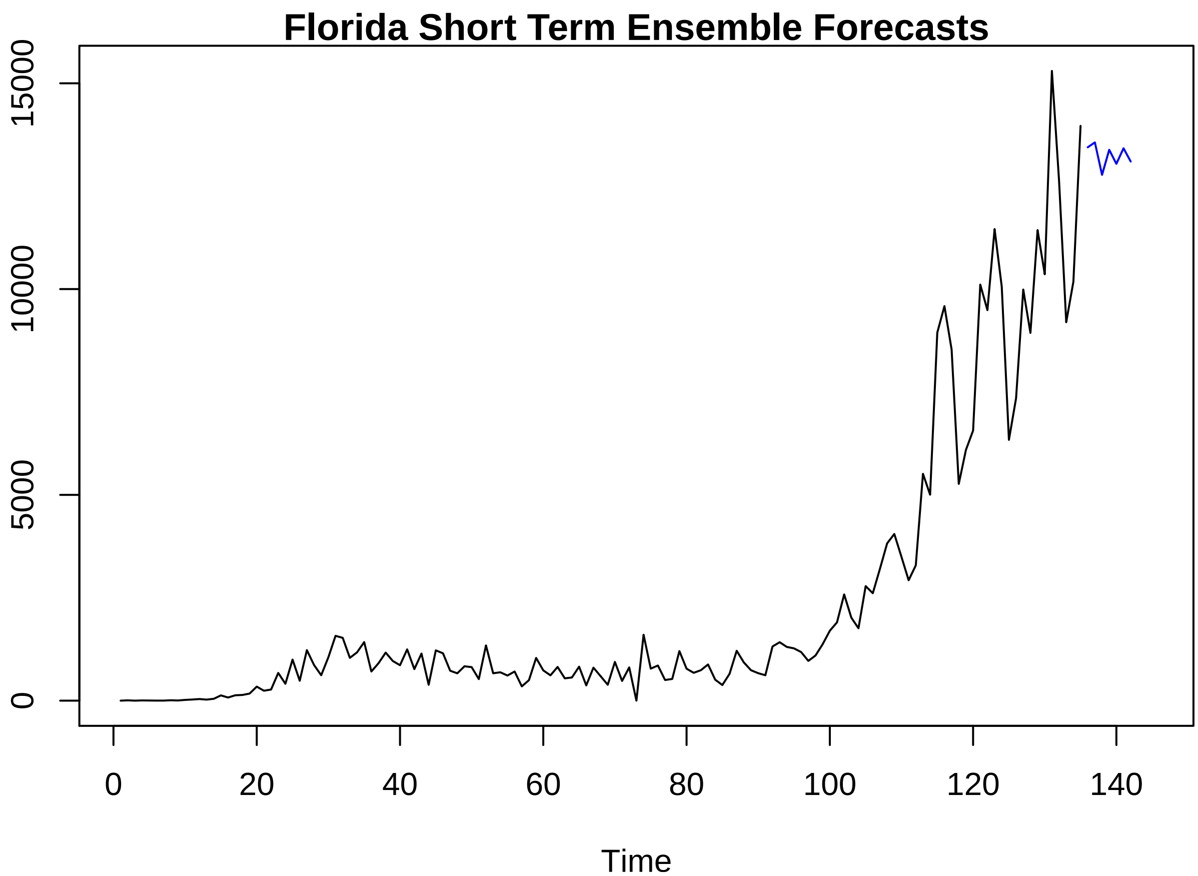
```r
#ASE fits for ensemble
mlp.fit_fl_final = mlp(ts(newcases_fl$positiveIncrease[1:123]), hd = 5, comb = "median")
forecasts_fl_mlp = forecast(mlp.fit_fl_final, h = 12)

forecasts_fl_arima = fore.aruma.wge(newcases_fl$positiveIncrease[i:(i+(trainingSize-1))],phi = fl_arima$

ensemble_fl_fore = (forecasts_fl_mlp$mean + final_pred_df$f) / 2

ensemble_ASE = mean((newcases_fl$positiveIncrease[124:135] - ensemble_fl_fore)^2)
ensemble_ASE
```

```
## [1] 5990983
```

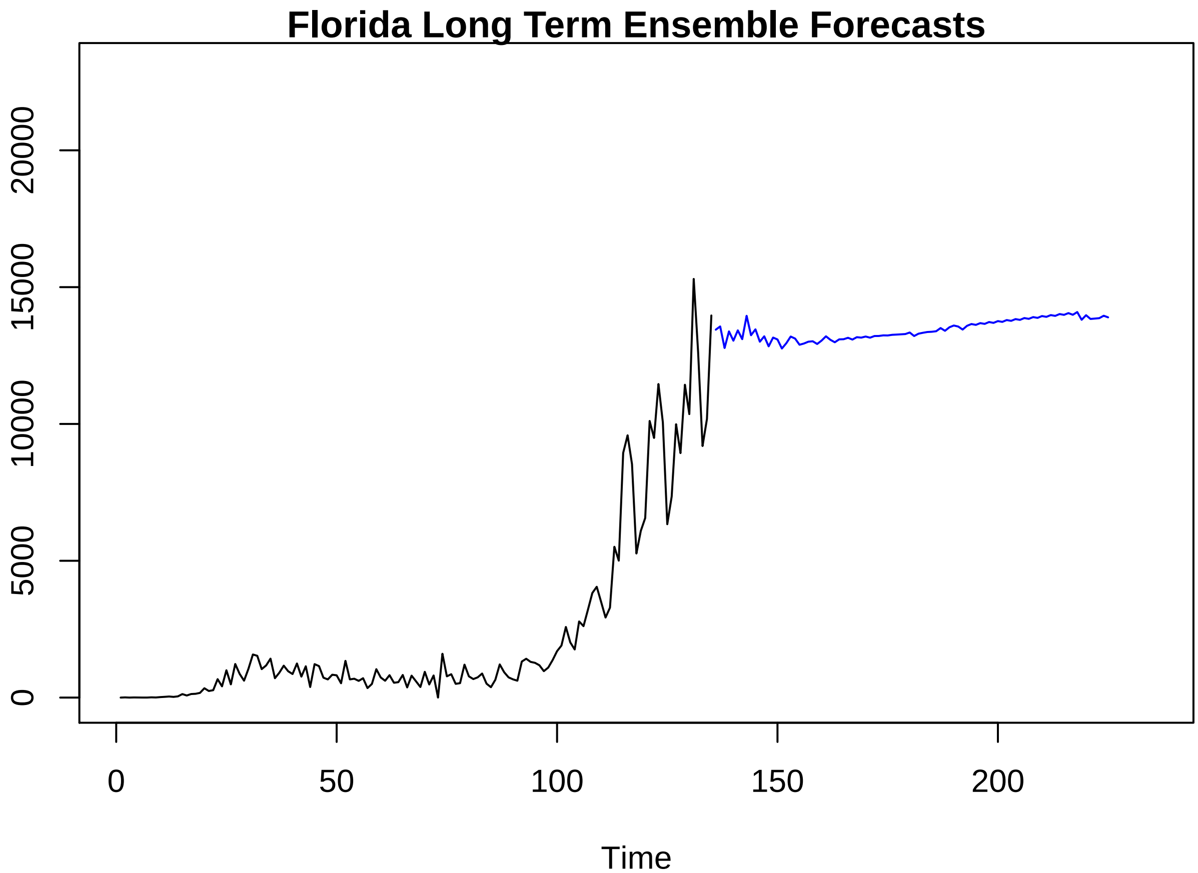
```r
#8.4 Mill

plot(newcases_fl$positiveIncrease, type = "l", ylab = "Count of New Cases", xlab = "Time", main = "Flori
lines(ensemble_fl_fore, col = "blue")
```




```r
#future forecasting

short_fl_ensemble = (short_fl_mlp$mean + short_fl_arima$f)/2
long_fl_ensemble = (long_fl_mlp$mean + long_fl_arima$f)/2



plot(newcases_fl$positiveIncrease, type = "l", xlim = c(1,145), main = "Florida Short Term Ensemble Fore
lines(short_fl_ensemble, col = "blue")
```




```
plot(newcases_fl$positiveIncrease, type = "l", xlim = c(1,235), ylim = c(0,23000), main = "Florida Long
lines(long_fl_ensemble, col = "blue")
```



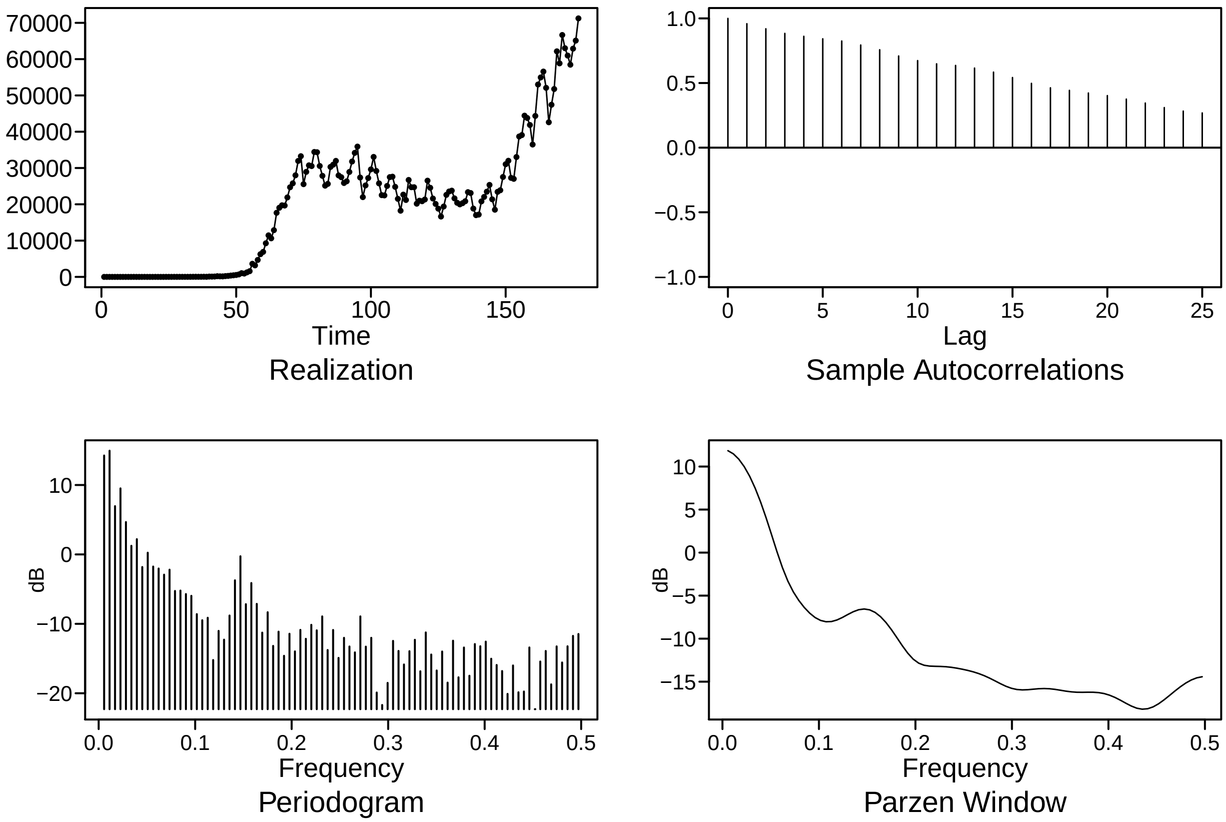
Comparing and Assessing Models

###Model Building for Cases United States

A. Stationarity vs Non-Stationarity

Overall we see slowly dampening ACFs, combined with a strong aperiodic frequency at zero in our spectral density. These measures alone with a recently quickly rising case count in recent days gives us strong evidence that our data is non-stationary. Given Covid19 spread, it is likely we see continued rising behavior in the short term, some return to lower numbers in the coming months but more uncertainty as new spikes could arise, and in the longest term of years on, we expect new cases to diminish to zero once the pandemic has ended spread.

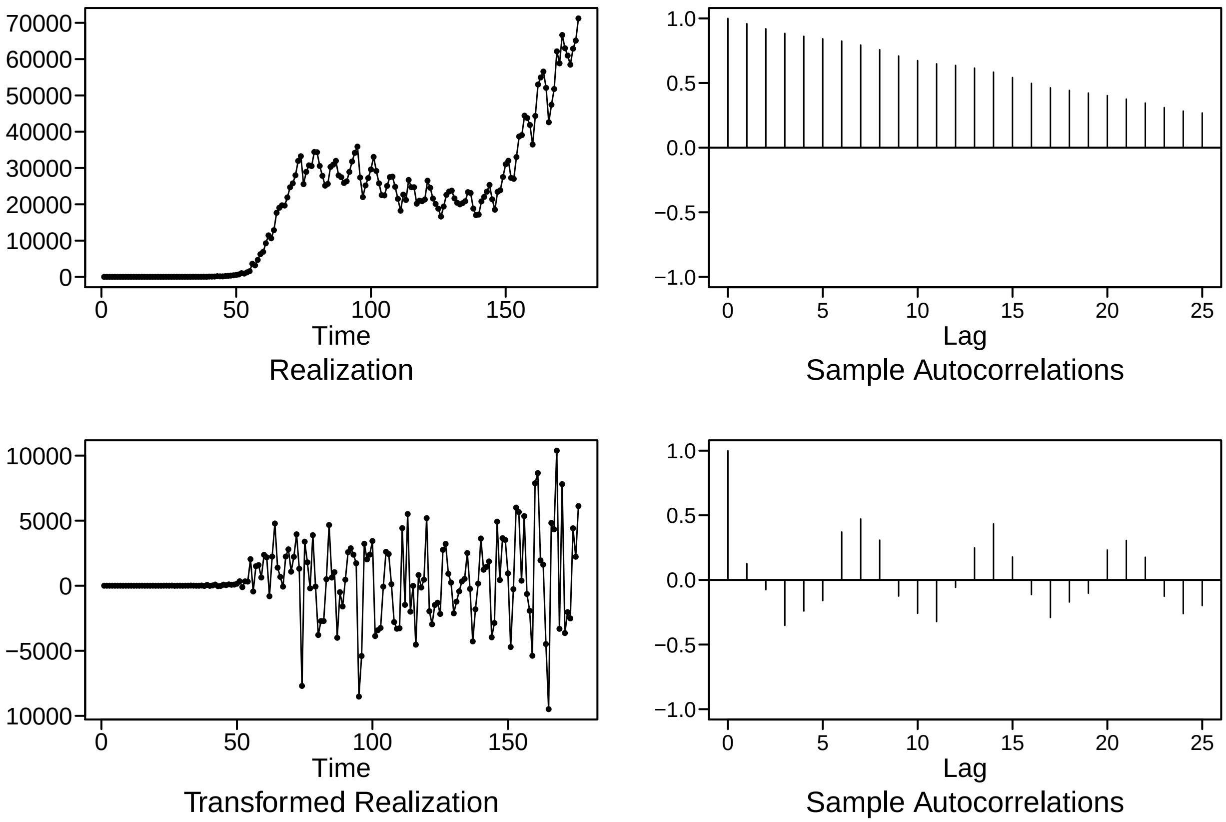
```
#no text output
x = plotts.sample.wge(newcases_us$positiveIncrease)
```



B. Non-Stationary Modeling

We did not do any differencing of our data set to account for this non-stationarity. Going into this project we knew that because of the failure to contain the Covid-19 outbreak we would see large spikes of cases in recent time periods compared to distant time periods. We feel that this is an important aspect of our data that we want to portray in our models because we can see empirically in Florida and the United States as a whole that both individual behavior and political policy continue to trend towards further outbreak and rapid, almost exponential daily case growth. While some states with compliant individual behavior and strong political Covid-19 policies have shown "completed" Covid-19 curves, where daily case count begins to trend downwards towards zero, Florida is the opposite. Therefore, since we empirically expect the trend of non-stationarity to continue, we want that represented in our models. This is a fundamental assumption that our models are built on.

```
diff_us = artrans.wge(newcases_us$positiveIncrease, 1)
```



C. Model IDing of stationary models

For ARIMA models we identify the stationary components below.

```r
#modeling as non-stationary
#0,5 maxes out
aic5.wge(diff_us, p = 3:10)
```

```
## ---------WORKING... PLEASE WAIT...
##
##
## Error in aic calculation at 3 2
## Error in aic calculation at 4 2
## Five Smallest Values of  aic

##        p    q        aic
## 12     6    2   15.35301
## 15     7    2   15.36326
## 18     8    2   15.37452
## 24    10    2   15.37494
## 17     8    1   15.37902
```

```r
aic5.wge(diff_us, type = 'bic',p = 3:10)
```

```
## ---------WORKING... PLEASE WAIT...
##
##
## Error in aic calculation at 3 2
## Error in aic calculation at 4 2
## Five Smallest Values of  bic
```

```
##       p    q        bic
## 12    6    2   15.51514
## 15    7    2   15.54340
## 16    8    0   15.54364
## 14    7    1   15.55598
## 17    8    1   15.55916
```

```
#Modeling as stationary
aic5.wge(newcases_us$positiveIncrease)
```

```
## ---------WORKING... PLEASE WAIT...
##
##
## Error in aic calculation at 1 1
## Error in aic calculation at 2 0
## Error in aic calculation at 2 1
## Error in aic calculation at 2 2
## Error in aic calculation at 3 0
## Error in aic calculation at 3 1
## Error in aic calculation at 3 2
## Error in aic calculation at 4 0
## Error in aic calculation at 4 1
## Error in aic calculation at 4 2
## Error in aic calculation at 5 0
## Error in aic calculation at 5 1
## Error in aic calculation at 5 2
## Five Smallest Values of  aic
```

```
##      p    q        aic
## 6    1    2   15.95786
## 4    1    0   15.96008
## 3    0    2   17.57501
## 2    0    1   18.44652
## 1    0    0   19.55796
```

```
aic5.wge(newcases_us$positiveIncrease, type = 'bic')
```

```
## ---------WORKING... PLEASE WAIT...
##
##
## Error in aic calculation at 1 1
## Error in aic calculation at 2 0
## Error in aic calculation at 2 1
## Error in aic calculation at 2 2
## Error in aic calculation at 3 0
## Error in aic calculation at 3 1
## Error in aic calculation at 3 2
## Error in aic calculation at 4 0
## Error in aic calculation at 4 1
## Error in aic calculation at 4 2
## Error in aic calculation at 5 0
## Error in aic calculation at 5 1
## Error in aic calculation at 5 2
## Five Smallest Values of  bic
```

```
##      p    q        bic
```

```
## 4    1    0   15.99597
## 6    1    2   16.02964
## 3    0    2   17.62884
## 2    0    1   18.48240
## 1    0    0   19.57590
```

D. Model Building

####US Cases ARIMA

```r
us_arima = est.arma.wge(diff_us, p = 6, q = 2)
```

```
##
## Coefficients of Original polynomial:
## 0.9297 -0.4193 -0.1050 0.1192 -0.1654 0.4349
##
## Factor                  Roots                Abs Recip    System Freq
## 1-1.2345B+0.9676B^2    0.6379+-0.7916i      0.9837       0.1420
## 1-0.9236B              1.0827               0.9236       0.0000
## 1+0.7898B             -1.2661               0.7898       0.5000
## 1+0.4386B+0.6161B^2   -0.3560+-1.2233i      0.7849       0.2951
##
##
```

```r
us_arima$aic
```

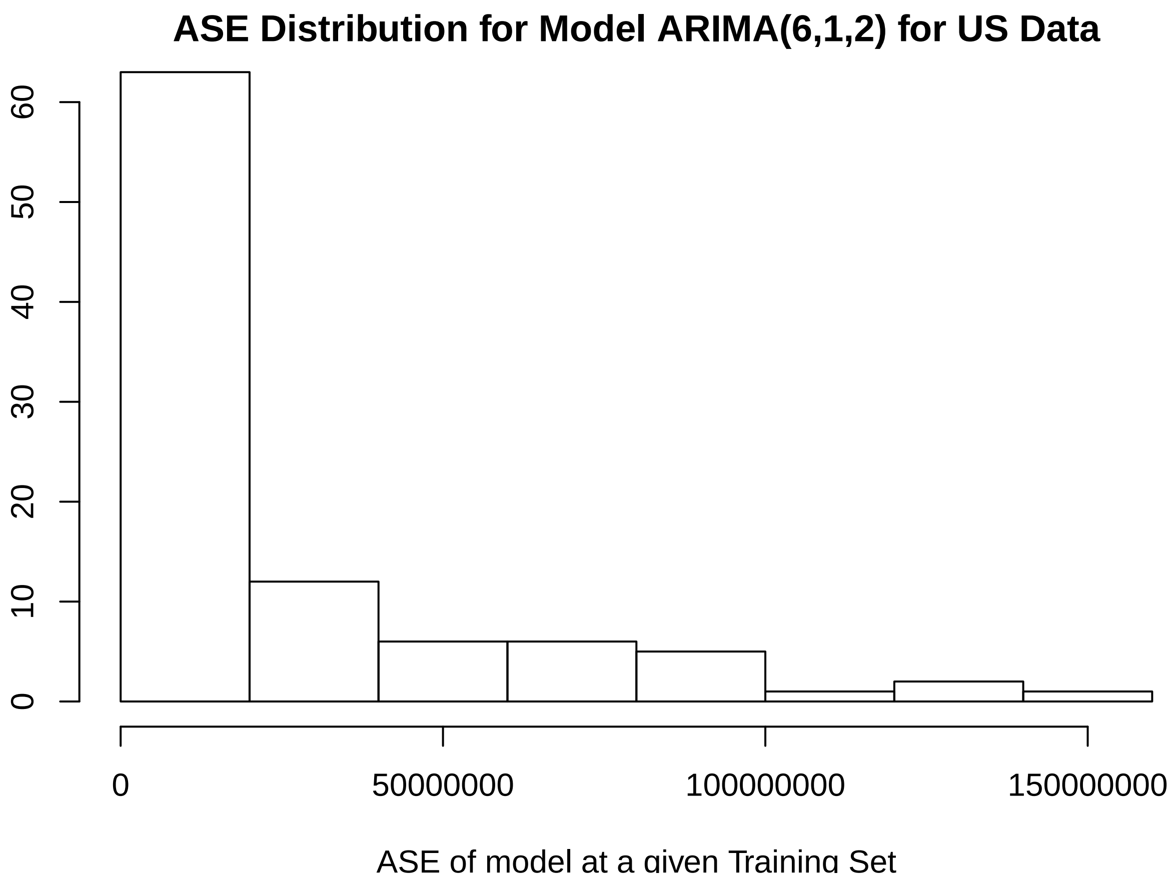
```
## [1] 15.35301
```

```r
trainingSize = 70
horizon = 12
ASEHolder = numeric()

for( i in 1:(177-(trainingSize + horizon) + 1))
{

  forecasts = fore.aruma.wge(newcases_us$positiveIncrease[i:(i+(trainingSize-1))],phi = us_arima$phi, t

  ASE = mean((newcases_us$positiveIncrease[(trainingSize+i):(trainingSize+ i + (horizon) - 1)] - foreca
  ASEHolder[i] = ASE

}

ASEHolder
```

```
##   [1]  10675056  18600503  30808202  65947936 107923495  17482463  30122239
##   [8]  34296758  18901087  41985555  22196078   8038259   4001099  16220456
##  [15]  25623600  10988291  16015628   7577133  11764724   6676069   7463415
##  [22]   7011389   8632723  11233687  29607353  92783091  28689956   2678473
##  [29]   6273351   3213464   2655246   5598353   2367409   5317189   7163735
##  [36]   7454121   9185823   3972393   2978218   9144903  12066578  17419336
##  [43]   1643801   5211231   3376700   2165611   4101283   3272886   6092837
##  [50]   7387112   2477927   9583839   2591774   3062112   2918252   4802053
##  [57]  14084685  16012246   9870509   4635991   3884471   6951811  14030208
##  [64]  16685027   7880299   4905266   6986689  12231151  42430184  68416051
##  [71]  72678321  36882715  43820966  61733016  41676411  78582278 122236456
##  [78]  73735402  97503595  89508118  88460562  90058778 135316356 149071511
##  [85]  57389370  16692038  24213370  17023010  16284869  21377733  59860784
```

```
## [92]  35431300  12398822  12621652  14419232  20118462
```

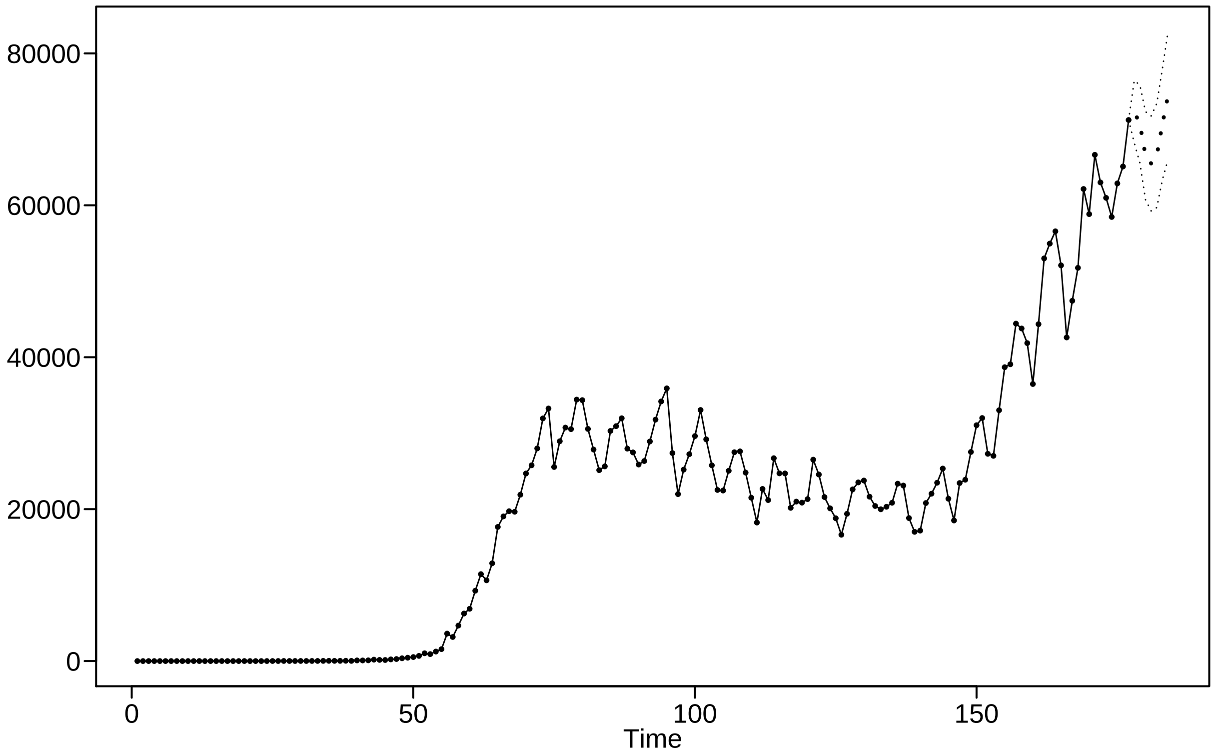
```r
#Distribution of ASEs on Two Week Periods
hist(ASEHolder, xlab = "ASE of model at a given Training Set",  main = "ASE Distribution for Model ARIM
```



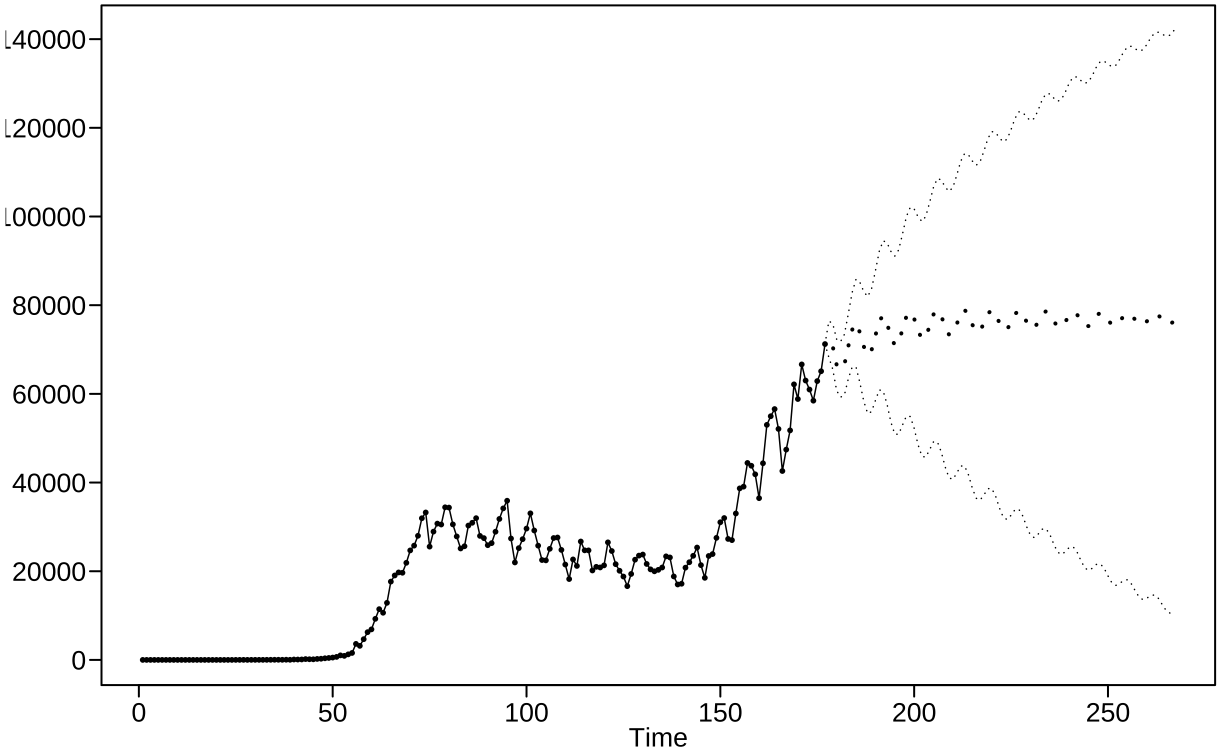

```r
#Mean ASE
WindowedASE = mean(ASEHolder)
WindowedASE
```

```
## [1] 26724396
```

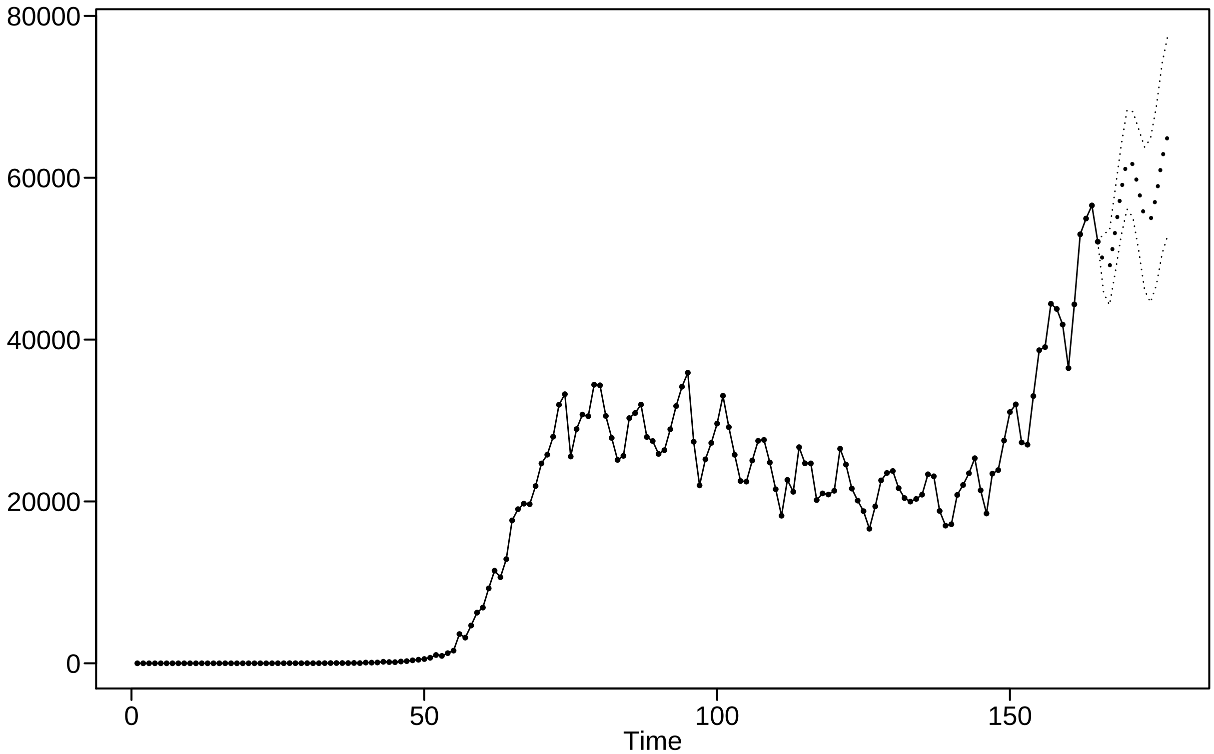
```r
#26724396


short_us_arima = fore.aruma.wge(newcases_us$positiveIncrease,phi = us_arima$phi,theta = us_arima$theta,
```




```
long_us_arima = fore.aruma.wge(newcases_us$positiveIncrease,phi = us_arima$phi,theta = us_arima$theta, d
```



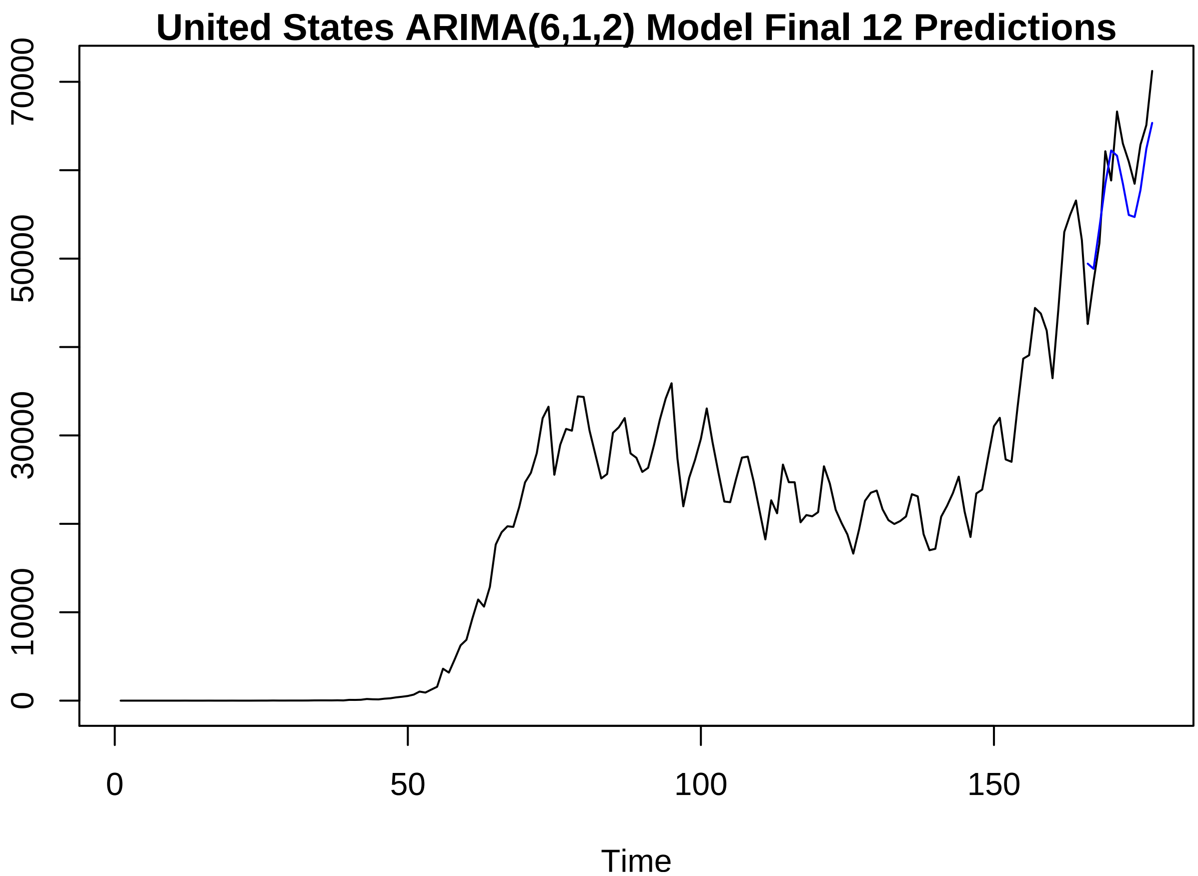

```
final_pred = fore.aruma.wge(newcases_us$positiveIncrease[1:165],phi = us_arima$phi,theta = us_arima$thet
```




```
final_12_ase = mean((newcases_us$positiveIncrease[166:177] - final_pred$f)^2)
final_12_ase
```

```
## [1] 20118462
```

```
plot(newcases_us$positiveIncrease, type = "l", ylab = "Count of New Cases", xlab = "Time", main = "Unit
lines(ts(final_pred$f, start = 166, end = 177),  col = "blue")
```



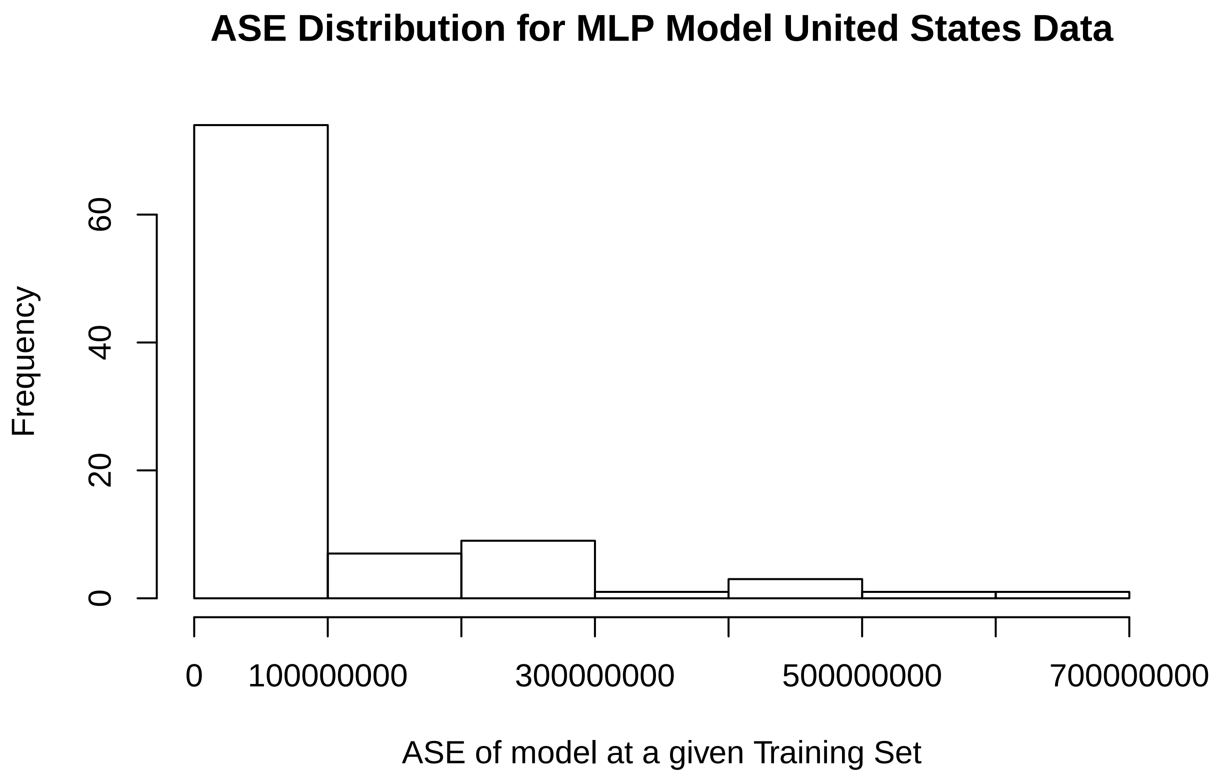
####US Cases MLP

```
trainingSize = 70
horizon = 12
ASEHolder = numeric()


for( i in 1:(177-(trainingSize + horizon) + 1))
{
  mlp.fit = mlp(ts(newcases_us$positiveIncrease[1:trainingSize+i]), hd = 5, reps = 20, lags = c(1,3,4),
  forecasts = forecast(mlp.fit,h = horizon)

  ASE = mean((newcases_us$positiveIncrease[(trainingSize+i):(trainingSize+ i + (horizon) - 1)] -forecas
  ASEHolder[i] = ASE

}

ASEHolder
```

```
##  [1] 109816116 297816634 287105287 432385367  11332866  17563339  17526596
##  [8]  15298739 164076580 119381783  42150192  10085044   6520820   9779106
## [15]   9096543  14807288  17848664  15161895  14593212  14799941  14789418
## [22]  15167655  18841428  25090063  32408782  19777547  12740601  14610824
## [29]  19662723  31381929  38723313  40435470  34219264  27423156  25621217
## [36]  33618702  40459391  46570671  43760932  34577803  24702177  31271953
## [43]  28995686  38447903  44179583  48224540  36948855  36211173  32297912
## [50]  31926019  42504757  39219494  28823327  32879421  22899968  10935029
```

```
## [57]   4570911   7501687  14534334   8232079   6423252   5877739   6206097
## [64]   4511452   8180231   8006157   8052028   8836821  13545447  19223523
## [71]  21617327  22991811  35571592  60587075  88054709 126461837 167499379
## [78] 214932487 232113012 129809804 208280841 298939040 291724934 395013220
## [85] 403252714 403509101  94337827 218411715 522585766 621589282 183795297
## [92]  80600017  61277318  41403128  79444543 239295383
```

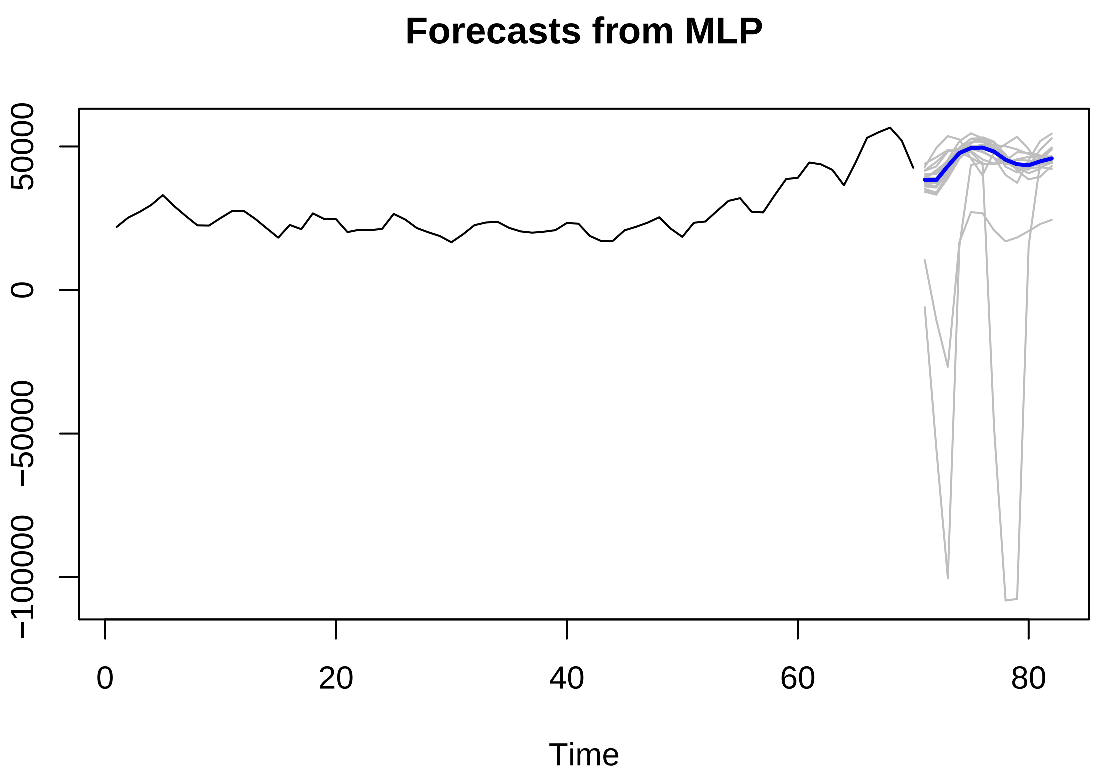
```r
#Distribution of ASEs on Two Week Periods
hist(ASEHolder, xlab = "ASE of model at a given Training Set",  main = "ASE Distribution for MLP Model U
```




```r
#Mean ASE
WindowedASE = mean(ASEHolder)
WindowedASE
```

```
## [1] 84315579
```

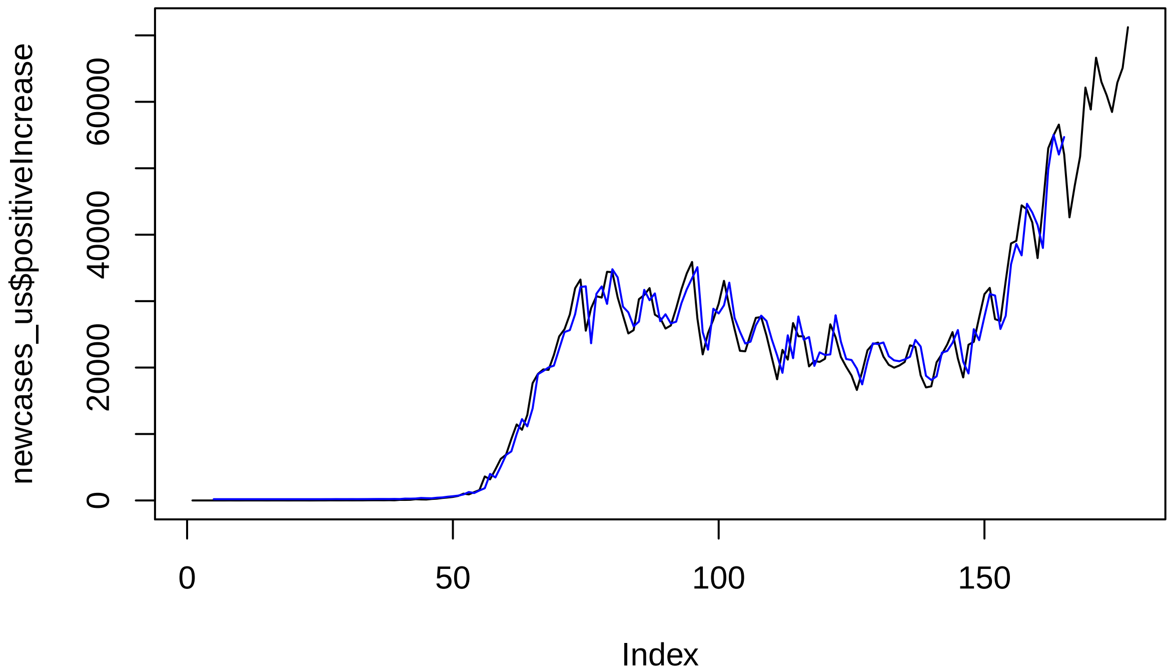
```r
#228 mill

plot(forecasts)
```



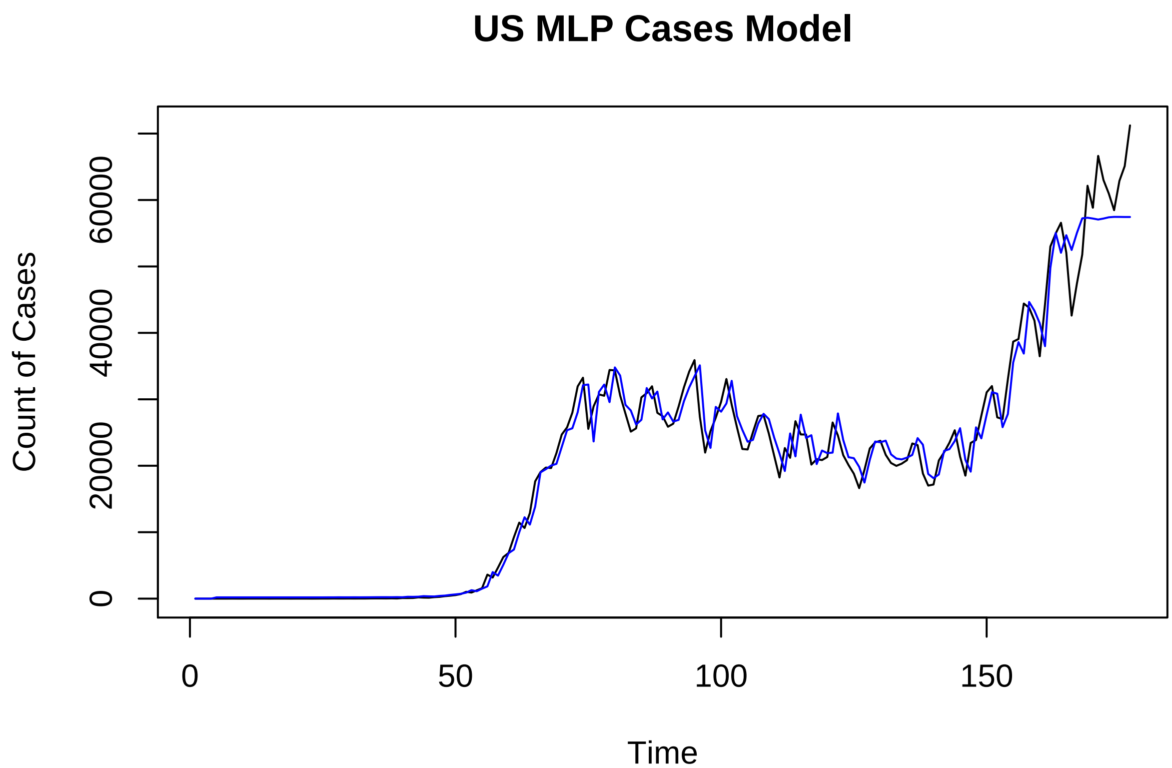

```r
#Actual Forecasting on last segment of data

mlp.fit = mlp(ts(newcases_us$positiveIncrease[1:165]), hd = 5, comb = "median")
forecasts_us_mlp = forecast(mlp.fit,h = 12)

ASE = mean((newcases_us$positiveIncrease[166:177] -forecasts_us_mlp$mean)^2)
ASE
```

```
## [1] 52414812
```

```r
#53,843,551
plot(newcases_us$positiveIncrease, type = "l")
lines(forecasts_us_mlp$fitted, col = "blue")
```

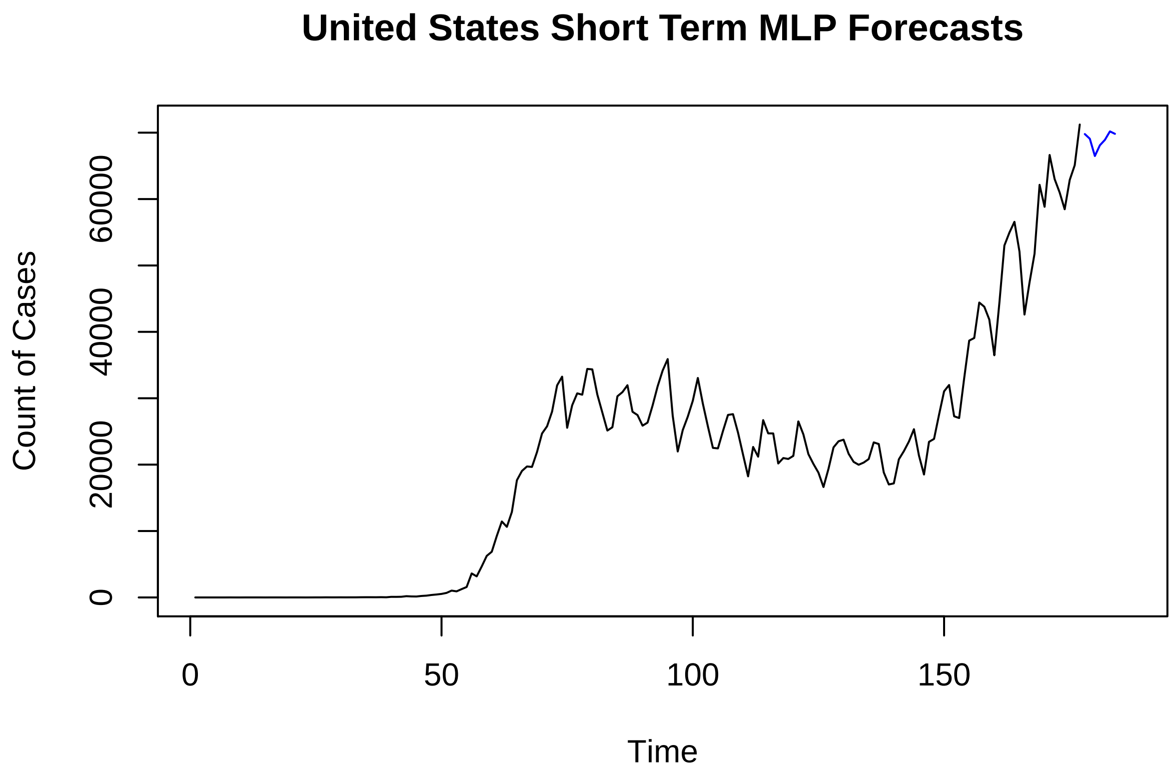



```
all_f = c(rep(1,4),forecasts_us_mlp$fitted, forecasts_us_mlp$mean)
plot(newcases_us$positiveIncrease, type = "l", ylab = "Count of Cases", xlab = "Time", main = "US MLP Ca
lines(all_f, col = "blue")
```

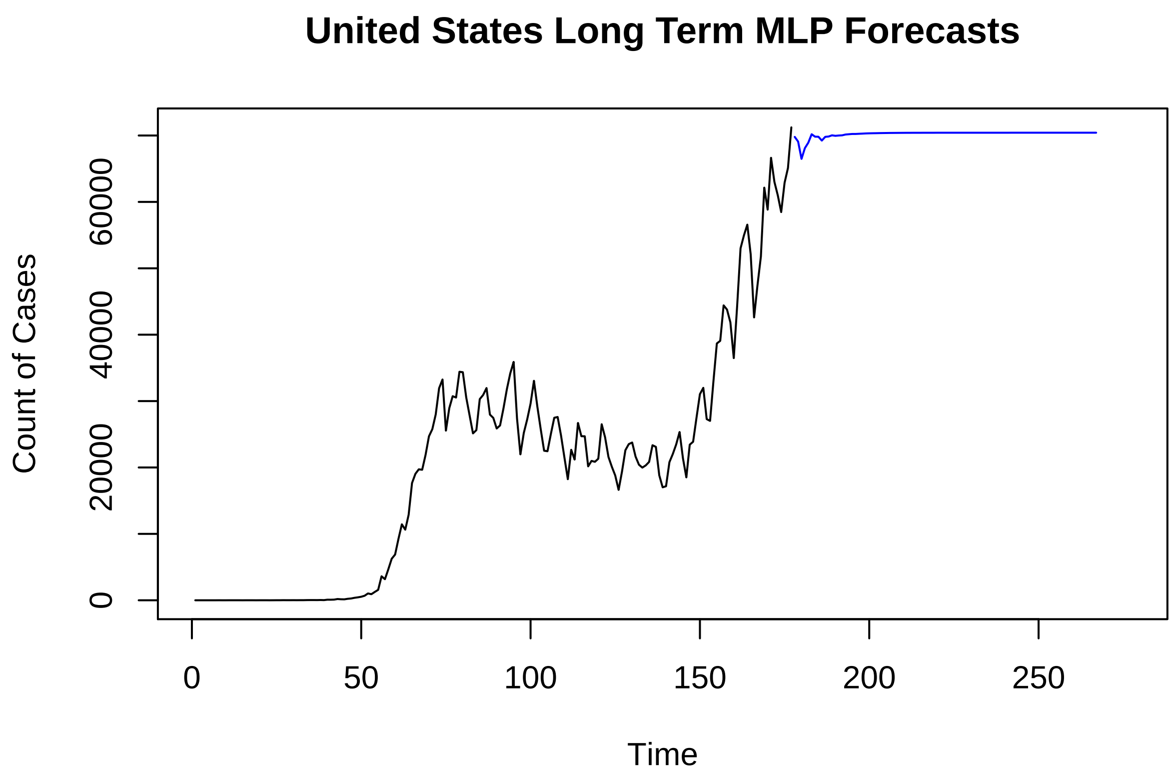



```
#Future Predictions

mlp.fit_us_future =mlp(ts(newcases_us$positiveIncrease), hd = 5, comb = "median")
short_us_mlp = forecast(mlp.fit_us_future,h = 7)
```

```
long_us_mlp = forecast(mlp.fit_us_future,h = 90)

plot(newcases_us$positiveIncrease, type = "l", xlim = c(1,187), main = "United States Short Term MLP Fo
lines(short_us_mlp$mean, col = "blue")
```

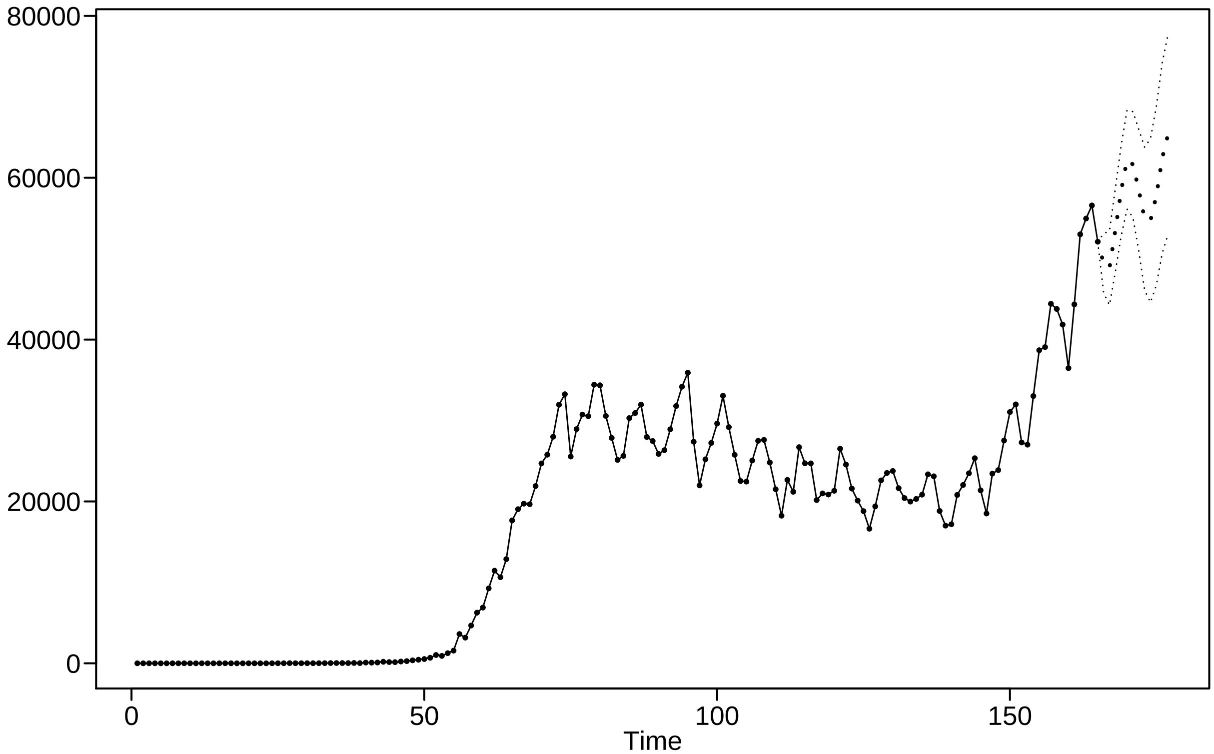



```
plot(newcases_us$positiveIncrease, type = "l", xlim = c(1,277), main = "United States Long Term MLP For
lines(long_us_mlp$mean, col = "blue")
```



####US Ensemble

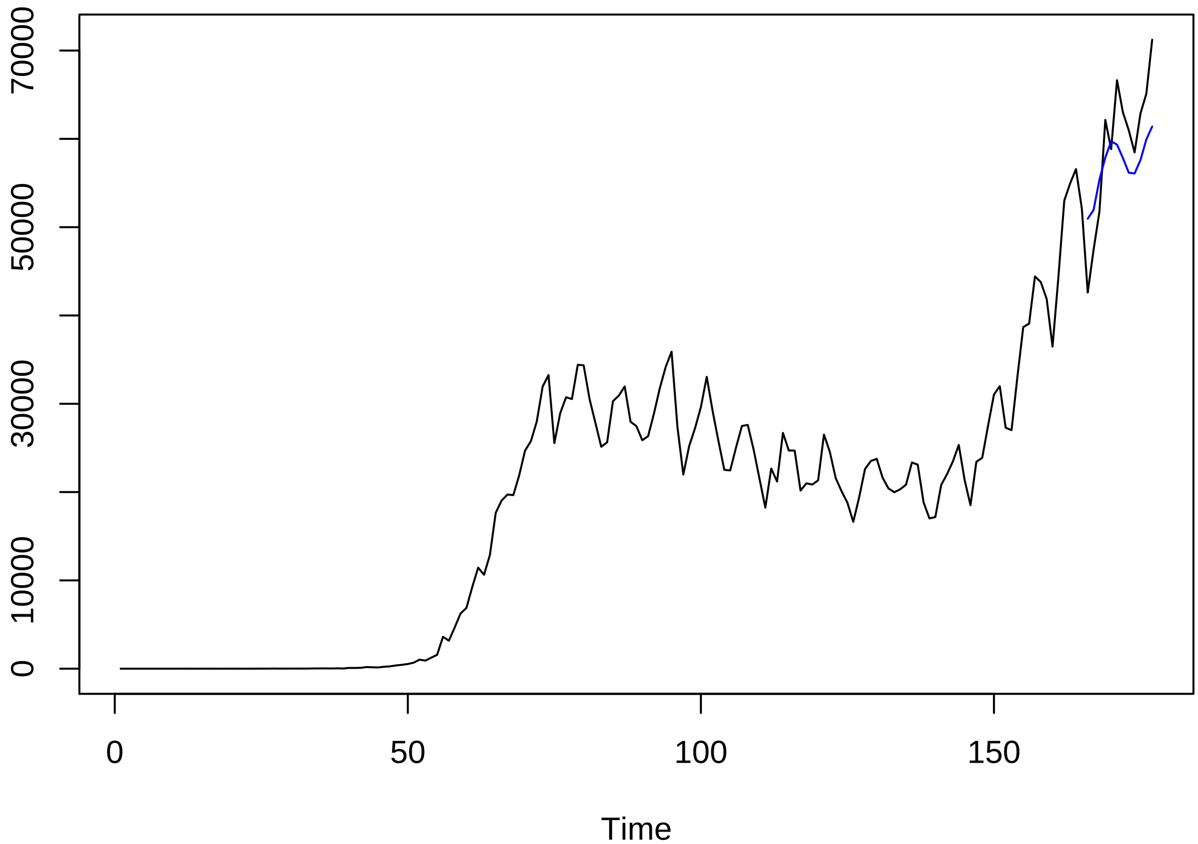
```
#ASE fits for ensemble
#mlp.fit_us_final = mlp(ts(newcases_us$positiveIncrease[1:165]), hd = 5, comb = "median")
#forecasts_us_mlp = forecast(mlp.fit_us_final,h = 12)


forecasts_arima_us = fore.aruma.wge(newcases_us$positiveIncrease[1:165],phi = us_arima$phi,theta = us_ar
```



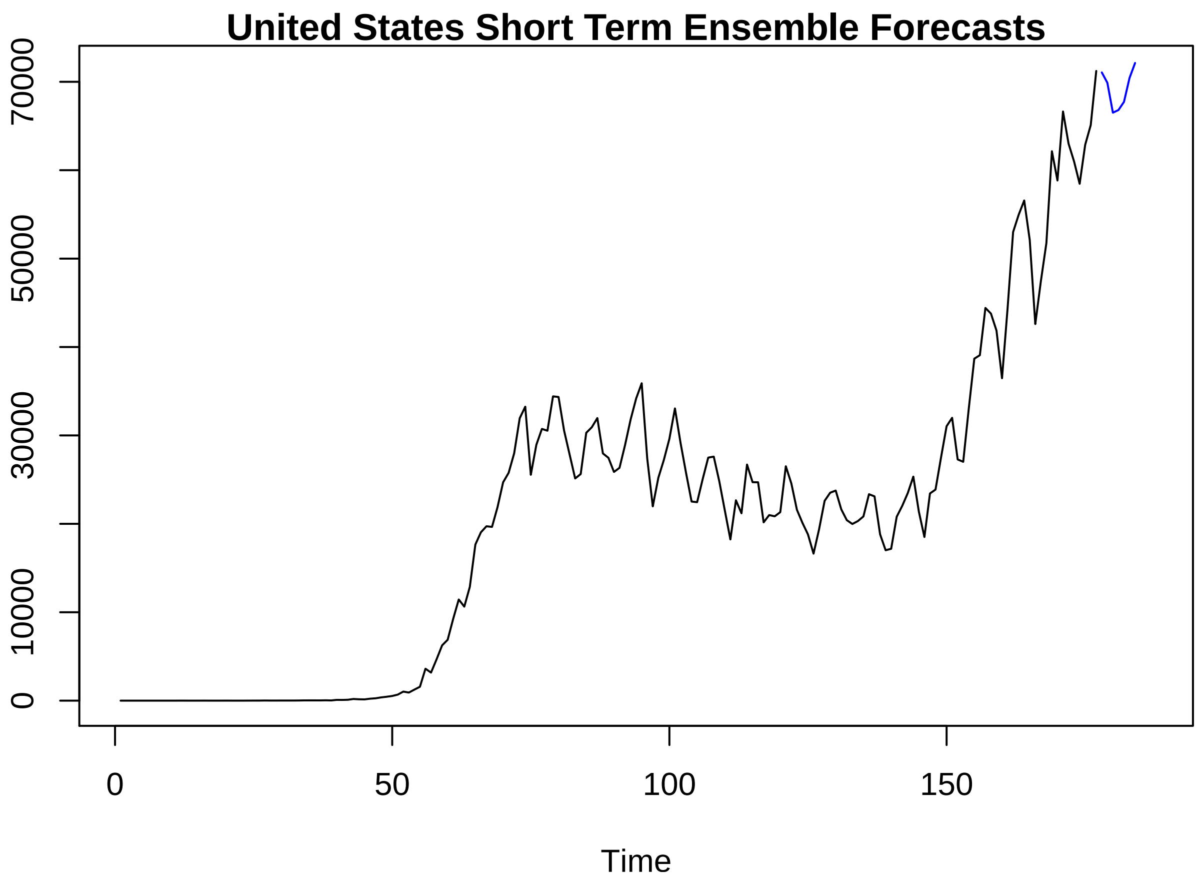

```r
ensemble_fore = (forecasts_us_mlp$mean + forecasts_arima_us$f) / 2

ensemble_ASE = mean((newcases_us$positiveIncrease[166:177] -ensemble_fore)^2)
ensemble_ASE
```

```
## [1] 31855669
```

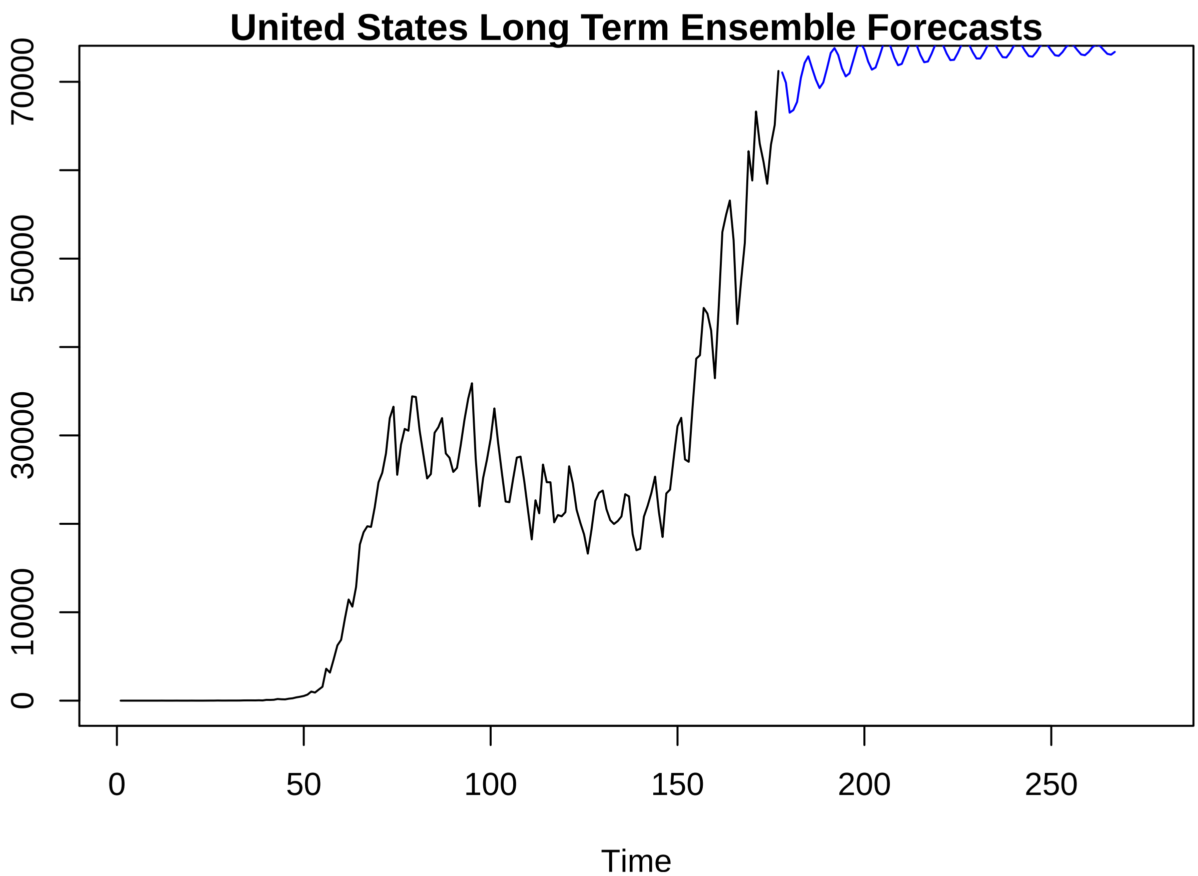
```r
#

plot(newcases_us$positiveIncrease, type = "l", ylab = "Count of New Cases", xlab = "Time")
lines(ensemble_fore, col = "blue")
```




```
#Forecasting ahead
short_ensemble_us = (short_us_mlp$mean +short_us_arima$f)/2

long_ensemble_us = (long_us_mlp$mean +long_us_arima$f)/2


plot(newcases_us$positiveIncrease, type = "l", xlim = c(1,187), main = "United States Short Term Ensembl
lines(short_ensemble_us, col = "blue")
```



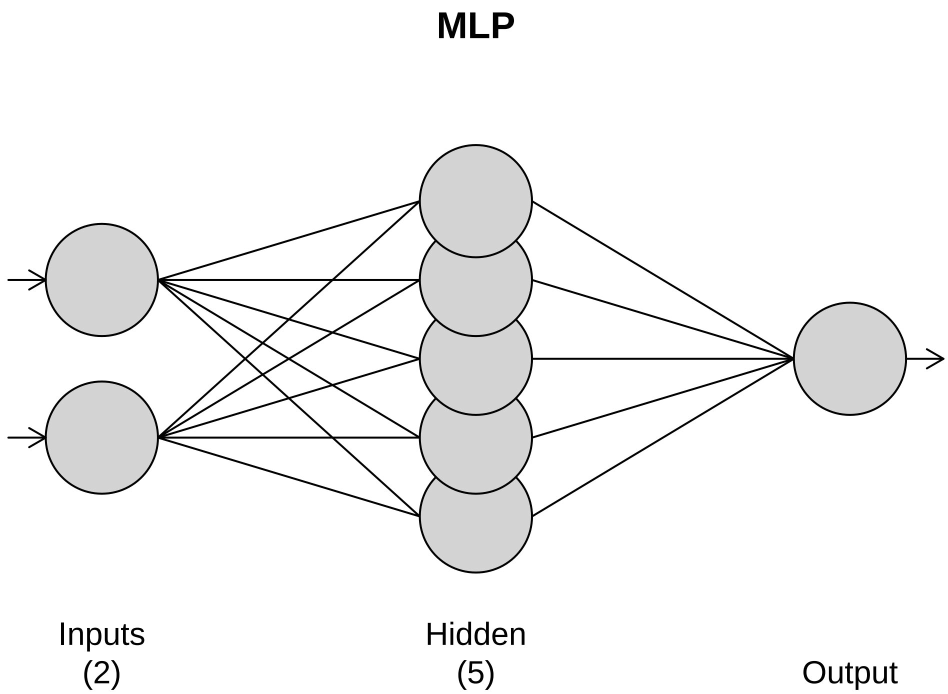

```
plot(newcases_us$positiveIncrease, type = "l", xlim = c(1,277), main = "United States Long Term Ensembl
lines(long_ensemble_us, col = "blue")
```



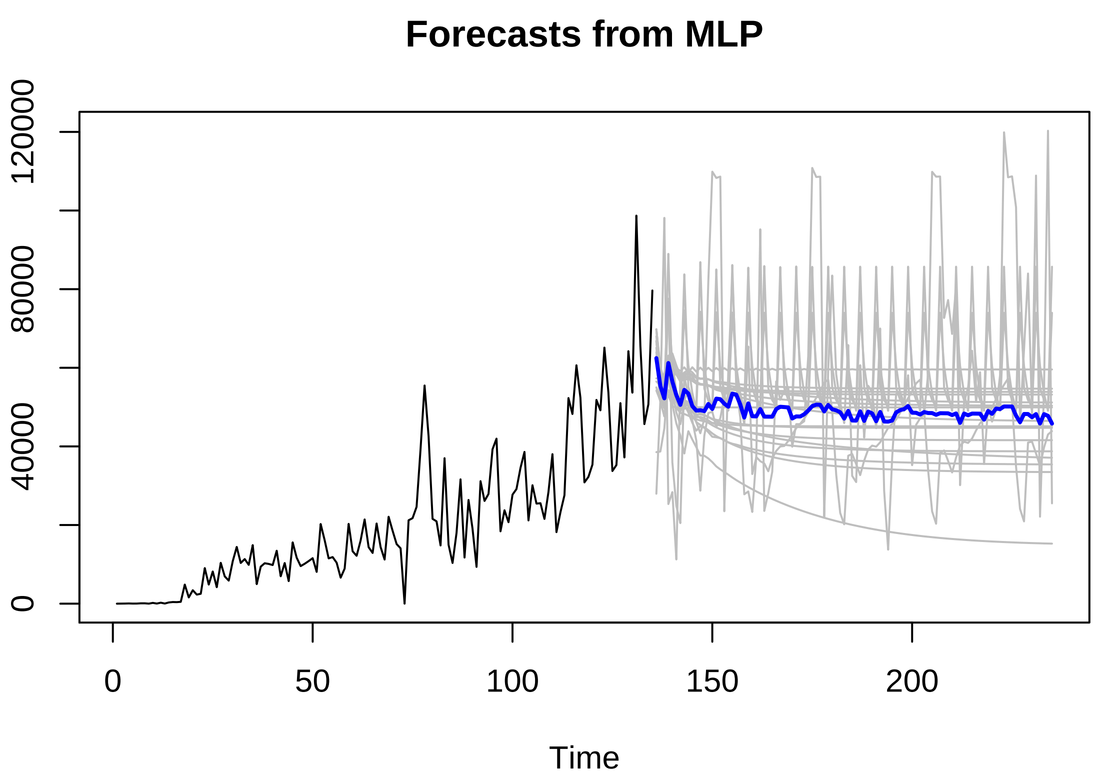
Model Building

##Goal Three: Multivariate Analysis

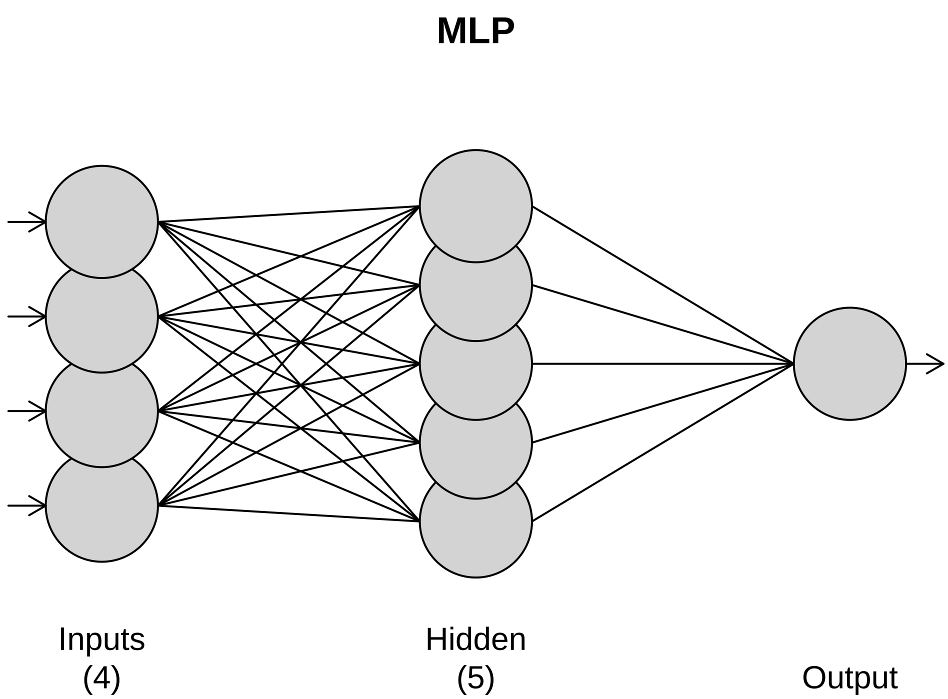
```
newcases_fl_multi =  initial_data_fl %>% dplyr::select(positiveIncrease, totalTestResultsIncrease, hospi

#Forecast beyond data for Florida

#Forecast future variables
fit.mlp.1 = mlp(ts(newcases_fl_multi$totalTestResultsIncrease),reps = 20, comb = "median")
plot(fit.mlp.1)
```



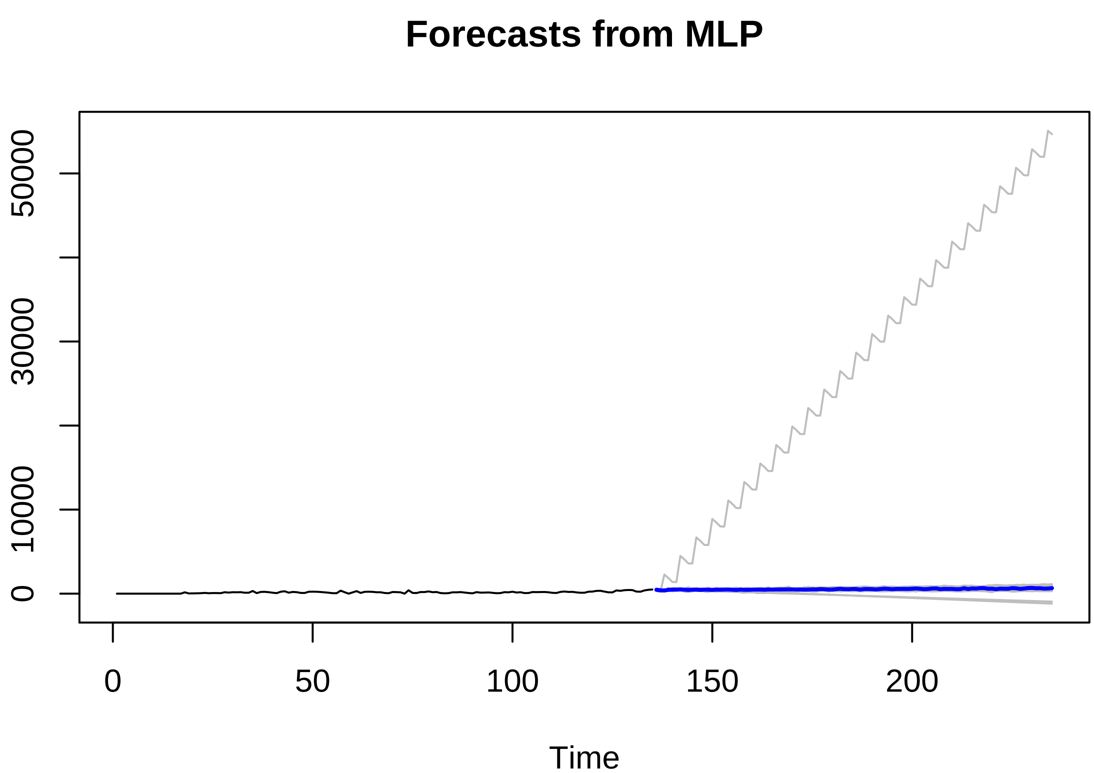

```
fore.mlp.1 = forecast(fit.mlp.1, h = 100)
plot(fore.mlp.1)
```



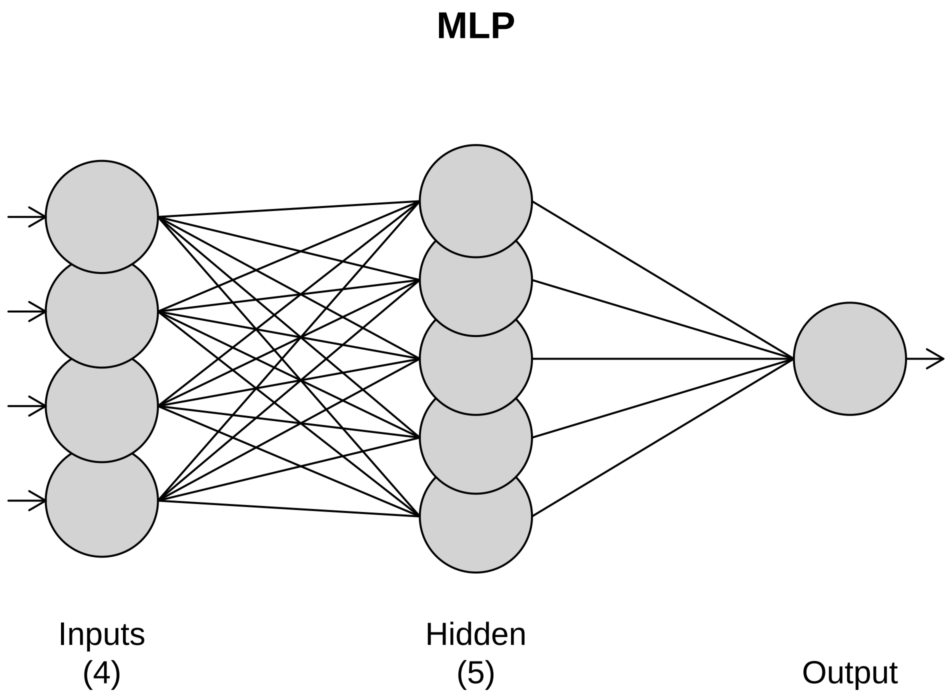

```
fit.mlp.2 = mlp(ts(newcases_fl_multi$hospitalizedIncrease),reps = 20, comb = "median")
plot(fit.mlp.2)
```



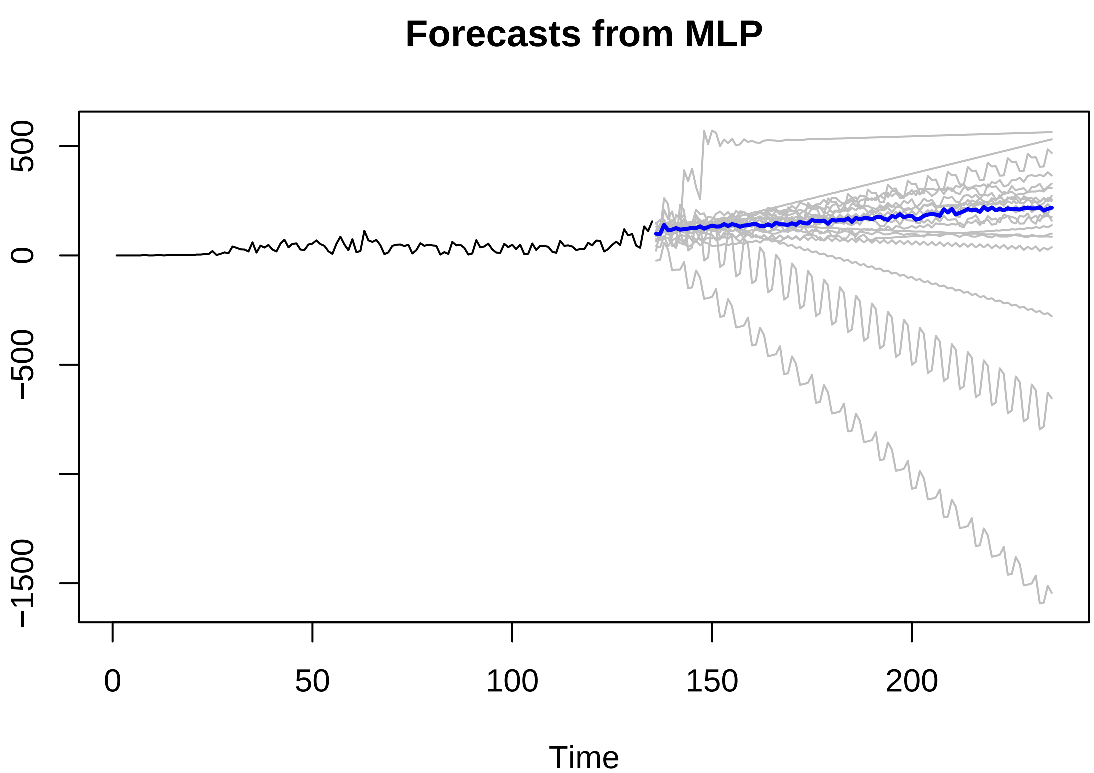

```
fore.mlp.2 = forecast(fit.mlp.2, h = 100)
plot(fore.mlp.2)
```




```
fit.mlp.3 = mlp(ts(newcases_fl_multi$deathIncrease),reps = 20, comb = "median")
plot(fit.mlp.3)
```




```
fore.mlp.3 = forecast(fit.mlp.3, h = 100)
plot(fore.mlp.3)
```




```
#package them up in data frame.
newvar_fore_fl = data.frame(totalTestResultsIncrease = ts(c(newcases_fl_multi$totalTestResultsIncrease,

#Data has 100 instances beyond current data
dim(newvar_fore_fl)
```

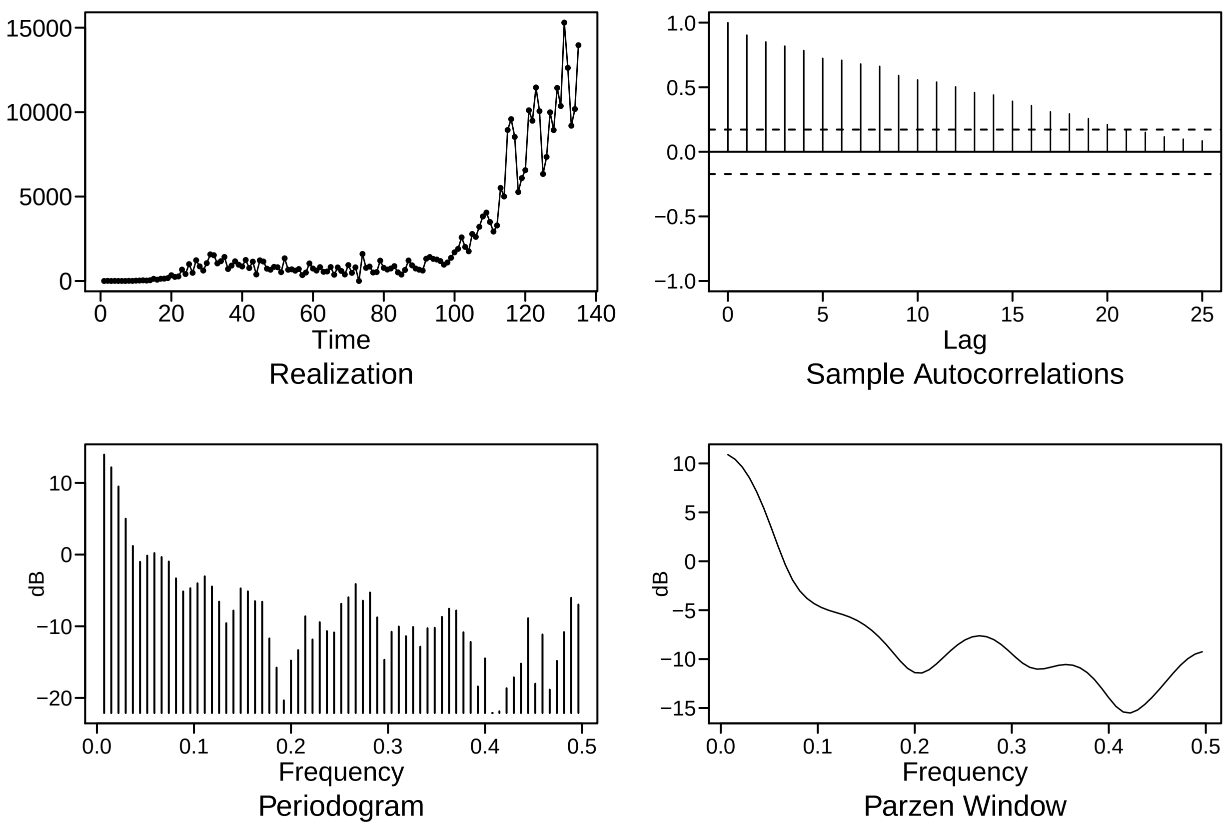
```
## [1] 235   3
```

###Multivariate Model Building for Florida Cases

A. Stationarity vs Non-Stationarity

Overall we see slowly dampening ACFs, combined with a strong aperiodic frequency at zero in our spectral density. These measures alone with a recently quickly rising case count in recent days gives us strong evidence that our data is non-stationary. Given Covid19 spread, it is likely we see continued rising behavior in the short term, some return to lower numbers in the coming months but more uncertainty as new spikes could arise, and in the longest term of years on, we expect new cases to diminish to zero once the pandemic has ended spread.

```r
x = plotts.sample.wge(newcases_fl$positiveIncrease, arlimits = TRUE)
```



B. Non-Stationary Modeling

We did not do any differencing of our data set to account for this non-stationarity. Going into this project we knew that because of the failure to contain the Covid-19 outbreak we would see large spikes of cases in recent time periods compared to distant time periods. We feel that this is an important aspect of our data that we want to portray in our models because we can see empirically in Florida and the United States as a whole that both individual behavior and political policy continue to trend towards further outbreak and rapid, almost exponential daily case growth. While some states with compliant individual behavior and strong political Covid-19 policies have shown "completed" Covid-19 curves, where daily case count begins to trend downwards towards zero, Florida is the opposite. Therefore, since we empirically expect the trend of non-stationarity to continue, we want that represented in our models. This is a fundamental assumption that our models are built on.

C. Model ID

In multivariate modeling, our identification of models occured specifically for each model and can be found at the beginning of those sections in particular.

D. Model Building

####Florida MLR Model

```
fit = lm(positiveIncrease~totalTestResultsIncrease + hospitalizedIncrease, data = newcases_fl_multi)
summary(fit)
```

```
##
## Call:
## lm(formula = positiveIncrease ~ totalTestResultsIncrease + hospitalizedIncrease,
##     data = newcases_fl_multi)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6794.8 -1047.8     1.2  1357.6  3712.4
##
## Coefficients:
##                             Estimate   Std. Error t value
## (Intercept)              -1358.579010   250.514502  -5.423
## totalTestResultsIncrease     0.137289     0.009719  14.126
## hospitalizedIncrease         5.588560     1.626045   3.437
##                                    Pr(>|t|)
## (Intercept)                      0.00000027 ***
## totalTestResultsIncrease < 0.0000000000000002 ***
## hospitalizedIncrease               0.000787 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1687 on 132 degrees of freedom
## Multiple R-squared:  0.7585, Adjusted R-squared:  0.7548
## F-statistic: 207.3 on 2 and 132 DF,  p-value: < 0.00000000000000022
```

```
aic5.wge(fit$residuals)#picks 1,1
```

```
## ---------WORKING... PLEASE WAIT...
##
##
## Five Smallest Values of  aic
```

```
##       p    q        aic
## 5     1    1   13.87980
## 8     2    1   13.89429
## 6     1    2   13.89497
## 11    3    1   13.90310
## 9     2    2   13.90630
```

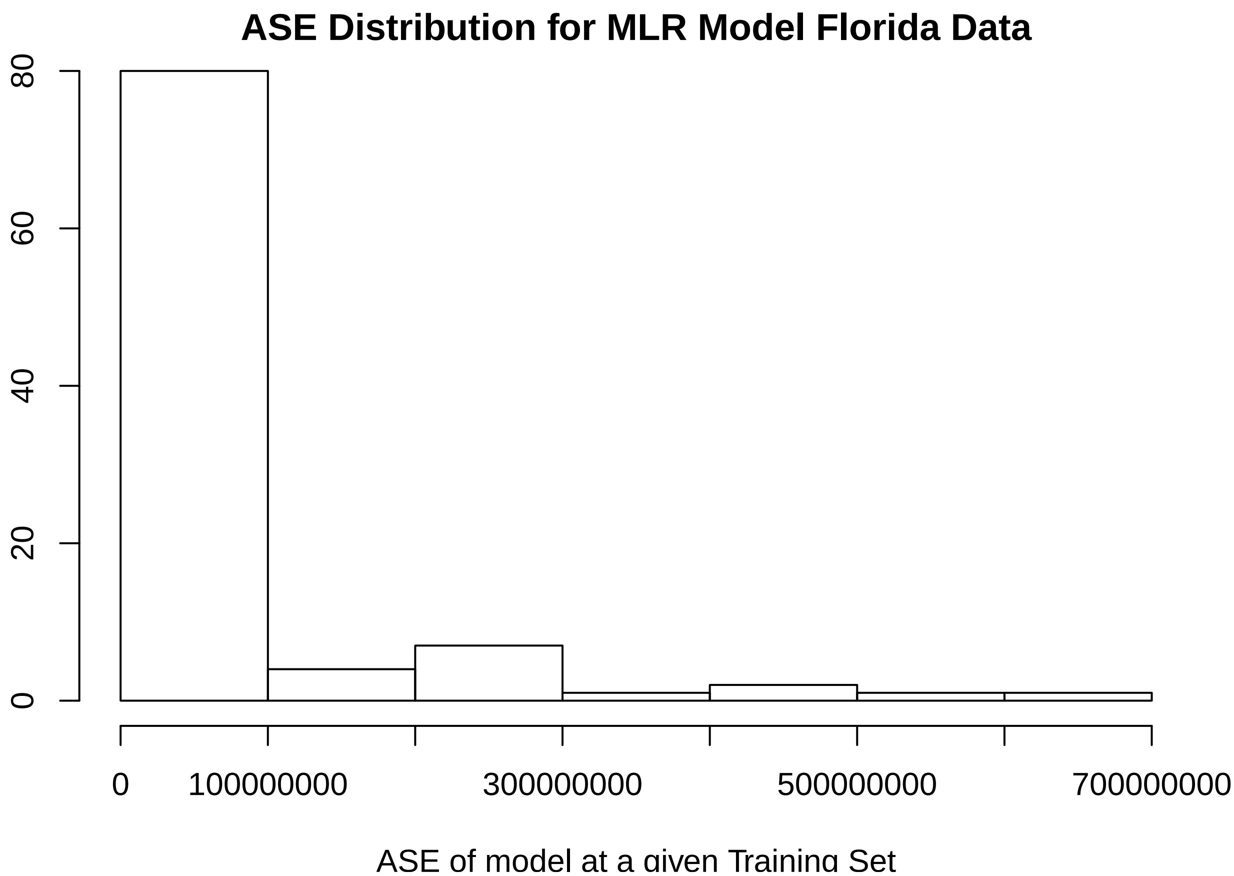
```
est1 = est.arma.wge(fit$residuals, p = 1, q = 1)
```

```
##
## Coefficients of Original polynomial:
## 0.9626
##
## Factor                 Roots                Abs Recip    System Freq
## 1-0.9626B              1.0388               0.9626       0.0000
##
##
```

```r
est_tests = mean(tail(newcases_fl_multi$totalTestResultsIncrease))
est_hospital= mean(tail(newcases_fl_multi$hospitalizedIncrease))

for( i in 1:(135-(trainingSize + horizon) ))
{
  fit = lm(positiveIncrease~totalTestResultsIncrease + hospitalizedIncrease, data = newcases_fl_multi[1
  newdata = data.frame(totalTestResultsIncrease = rep(est_tests,horizon), hospitalizedIncrease = rep(es

  preds = predict(fit, newdata = newdata)
  forecasts = fore.arma.wge(fit$residuals,phi = est1$phi,theta = est1$theta, lastn = FALSE,n.ahead = ho

  final_pred  = preds + forecasts$f


  ASE = mean((newcases_fl_multi$positiveIncrease[(trainingSize+i):(trainingSize+ i + (horizon) - 1)] -f
  ASEHolder[i] = ASE

}

ASEHolder
```

```
##   [1]   4376563.6   4490588.9   4713585.0   4996534.0   4552540.4
##   [6]   3889871.3   1517553.2    461262.7    539850.0    616442.4
##  [11]    496137.0    523596.0    524626.6    403695.8    221782.6
##  [16]    136819.1    182184.4    193169.1    143050.5    131257.9
##  [21]    264319.3    225747.0    213421.5    346797.5    413300.1
##  [26]    628089.3   1013180.8   1478979.6   1599746.7   1372768.6
##  [31]   1172744.6   1472886.3   1908251.3   5303180.7   8244606.3
##  [36]  10064419.8   8986978.8   7931125.0   7225911.1   9791042.3
##  [41]  12183663.8  15548906.9  12226294.3  10461543.8   4229821.9
##  [46]   4189764.2   4857911.5   5308658.7   4130739.7   6332777.4
##  [51]   5791097.6   6141545.0   6741713.6  32879420.5  22899968.2
##  [56]  10935029.1   4570911.4   7501687.3  14534334.4   8232078.7
##  [61]   6423252.2   5877738.6   6206097.2   4511451.6   8180230.5
##  [66]   8006156.6   8052028.0   8836821.1  13545447.0  19223522.8
##  [71]  21617326.7  22991810.6  35571592.3  60587074.8  88054709.5
##  [76] 126461836.7 167499378.8 214932487.0 232113012.4 129809804.3
##  [81] 208280841.0 298939040.5 291724934.3 395013219.6 403252713.6
##  [86] 403509101.2  94337827.4 218411714.6 522585765.8 621589282.2
##  [91] 183795297.5  80600016.7  61277317.6  41403127.9  79444543.1
##  [96] 239295383.3
```

```r
#Distribution of ASEs on Two Week Periods
hist(ASEHolder, xlab = "ASE of model at a given Training Set",  main = "ASE Distribution for MLR Model I
```




```r
#Mean ASE
WindowedASE = mean(ASEHolder)
WindowedASE
```

```
## [1] 58691962
```

```r
fit = lm(positiveIncrease~totalTestResultsIncrease + hospitalizedIncrease, data = newcases_fl_multi)
summary(fit)
```

```
##
## Call:
## lm(formula = positiveIncrease ~ totalTestResultsIncrease + hospitalizedIncrease,
##     data = newcases_fl_multi)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6794.8 -1047.8     1.2  1357.6  3712.4
##
## Coefficients:
##                              Estimate   Std. Error t value
## (Intercept)              -1358.579010   250.514502  -5.423
## totalTestResultsIncrease     0.137289     0.009719  14.126
## hospitalizedIncrease         5.588560     1.626045   3.437
##                                      Pr(>|t|)
## (Intercept)                        0.00000027 ***
## totalTestResultsIncrease < 0.0000000000000002 ***
## hospitalizedIncrease                 0.000787 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1687 on 132 degrees of freedom
## Multiple R-squared:  0.7585, Adjusted R-squared:  0.7548
## F-statistic: 207.3 on 2 and 132 DF,  p-value: < 0.000000000000000022
```

```
newdata = data.frame(totalTestResultsIncrease = rep(est_tests,12), hospitalizedIncrease = rep(est_hospi

preds = predict(fit, newdata = newdata)
aic5.wge(fit$residuals)#picks 1,1
```

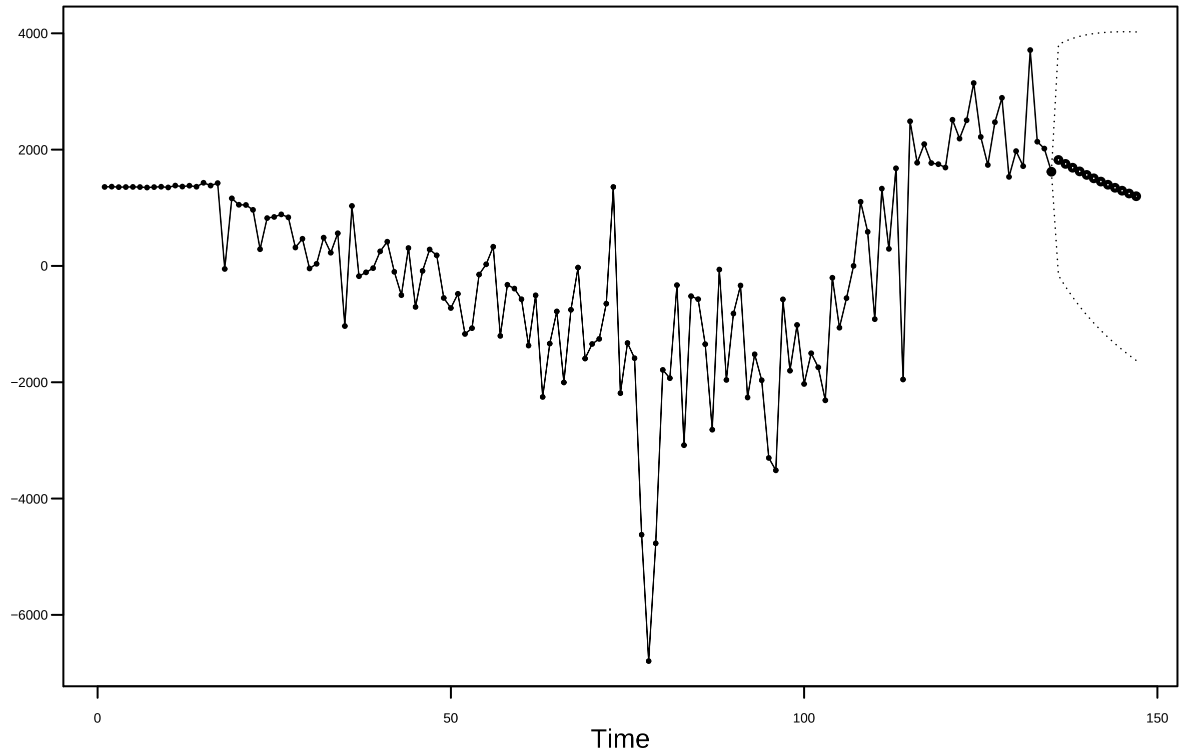
```
## ---------WORKING... PLEASE WAIT...
##
##
## Five Smallest Values of  aic
```

```
##       p    q        aic
## 5     1    1   13.87980
## 8     2    1   13.89429
## 6     1    2   13.89497
## 11    3    1   13.90310
## 9     2    2   13.90630
```

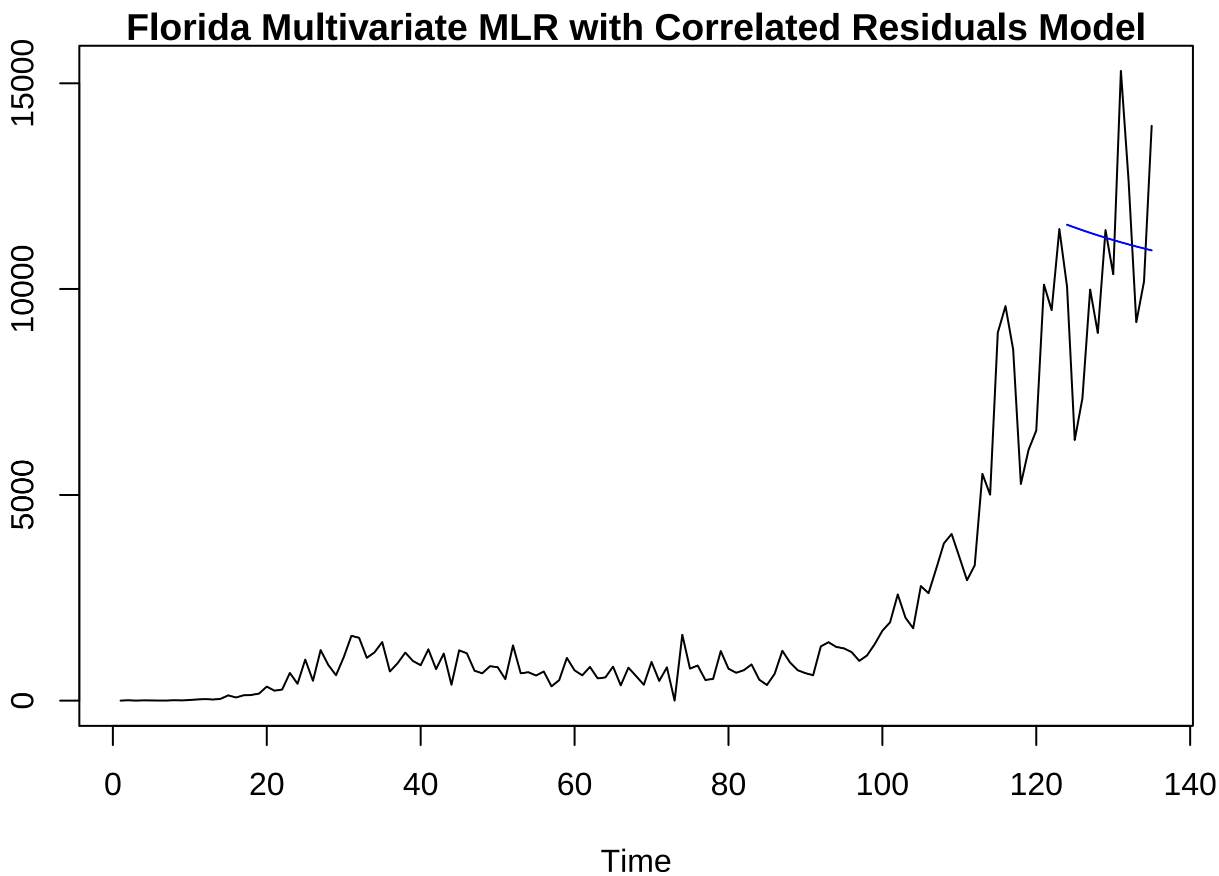
```
est1 = est.arma.wge(fit$residuals, p = 1, q = 1)
```

```
##
## Coefficients of Original polynomial:
## 0.9626
##
## Factor                 Roots                Abs Recip    System Freq
## 1-0.9626B              1.0388               0.9626       0.0000
##
##
```

```
forecasts = fore.arma.wge(fit$residuals,phi = est1$phi,theta = est1$theta, lastn = FALSE,n.ahead = 12)
```



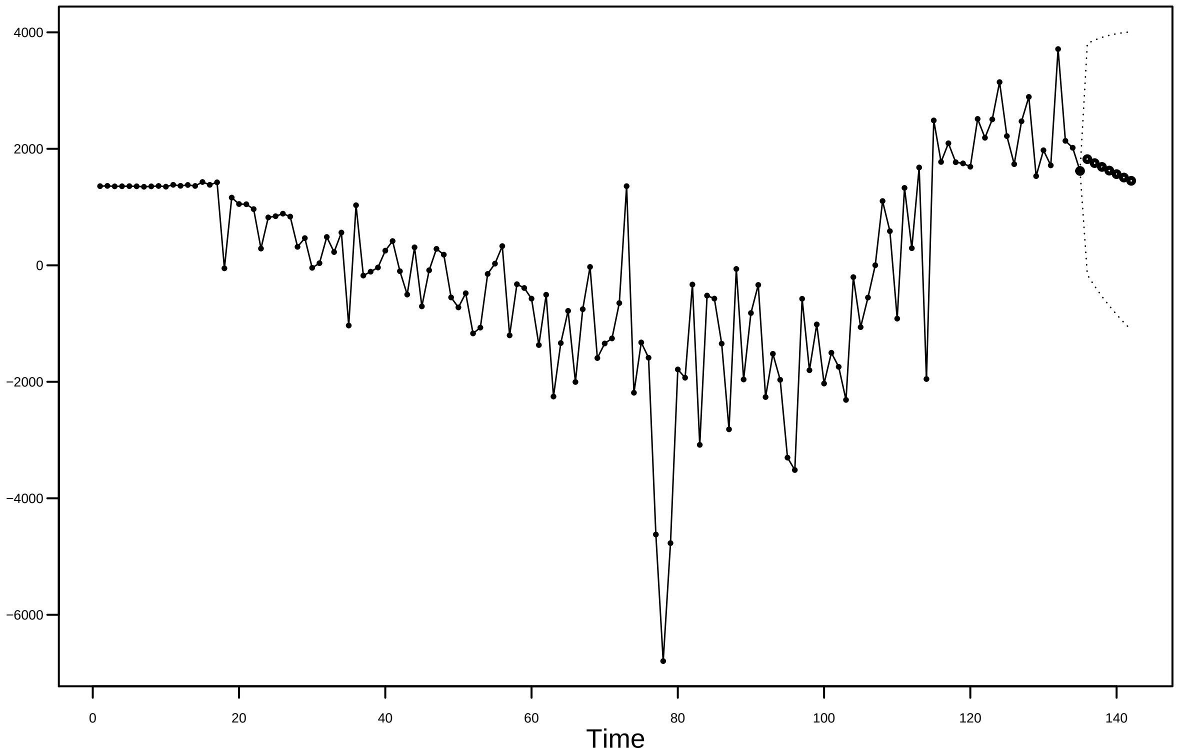

```
FinalPredictions_fl_MLR  = preds + forecasts$f


plot(newcases_fl$positiveIncrease, type = "l", main ="Florida Multivariate MLR with Correlated Residuals
lines(ts(FinalPredictions_fl_MLR, start = 124), col = "blue")
```



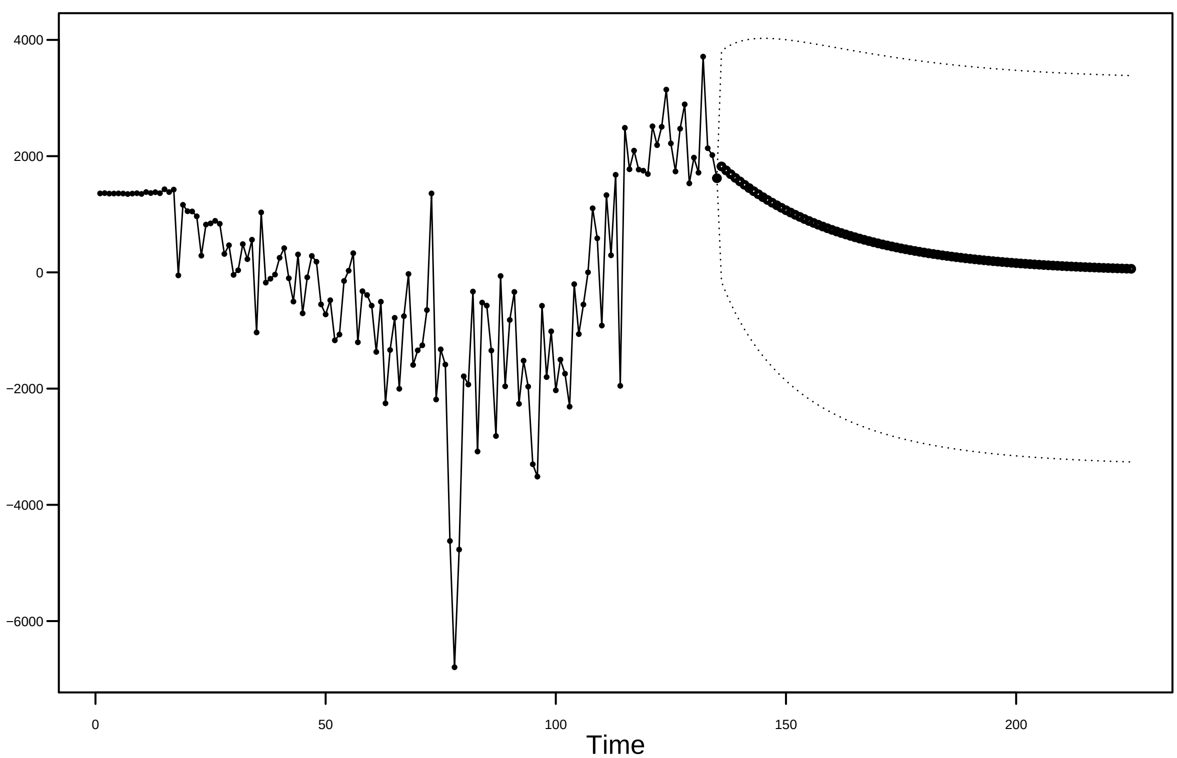

```
ASE = mean((newcases_fl_multi$positiveIncrease[124:135] - FinalPredictions_fl_MLR)^2)
ASE
```

```
## [1] 7223923
```

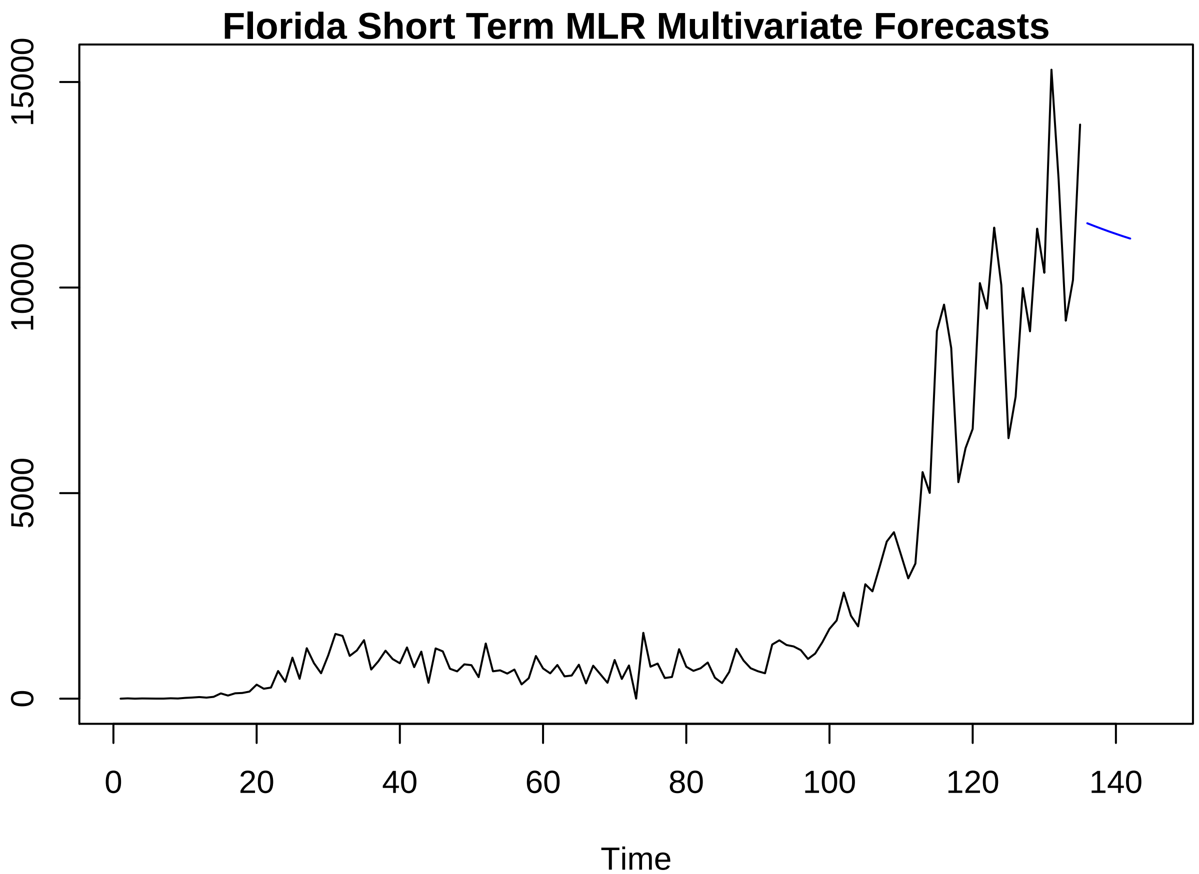
```
#Forecasting Ahead

shortdata = data.frame(totalTestResultsIncrease = rep(est_tests,7), hospitalizedIncrease = rep(est_hosp
longdata =data.frame(totalTestResultsIncrease = rep(est_tests,90), hospitalizedIncrease = rep(est_hospi




fit = lm(positiveIncrease~totalTestResultsIncrease + hospitalizedIncrease, data = newcases_fl_multi)
#short
preds = predict(fit, newdata = shortdata)
forecasts = fore.arma.wge(fit$residuals,phi = est1$phi,theta = est1$theta, lastn = FALSE,n.ahead = 7)
```



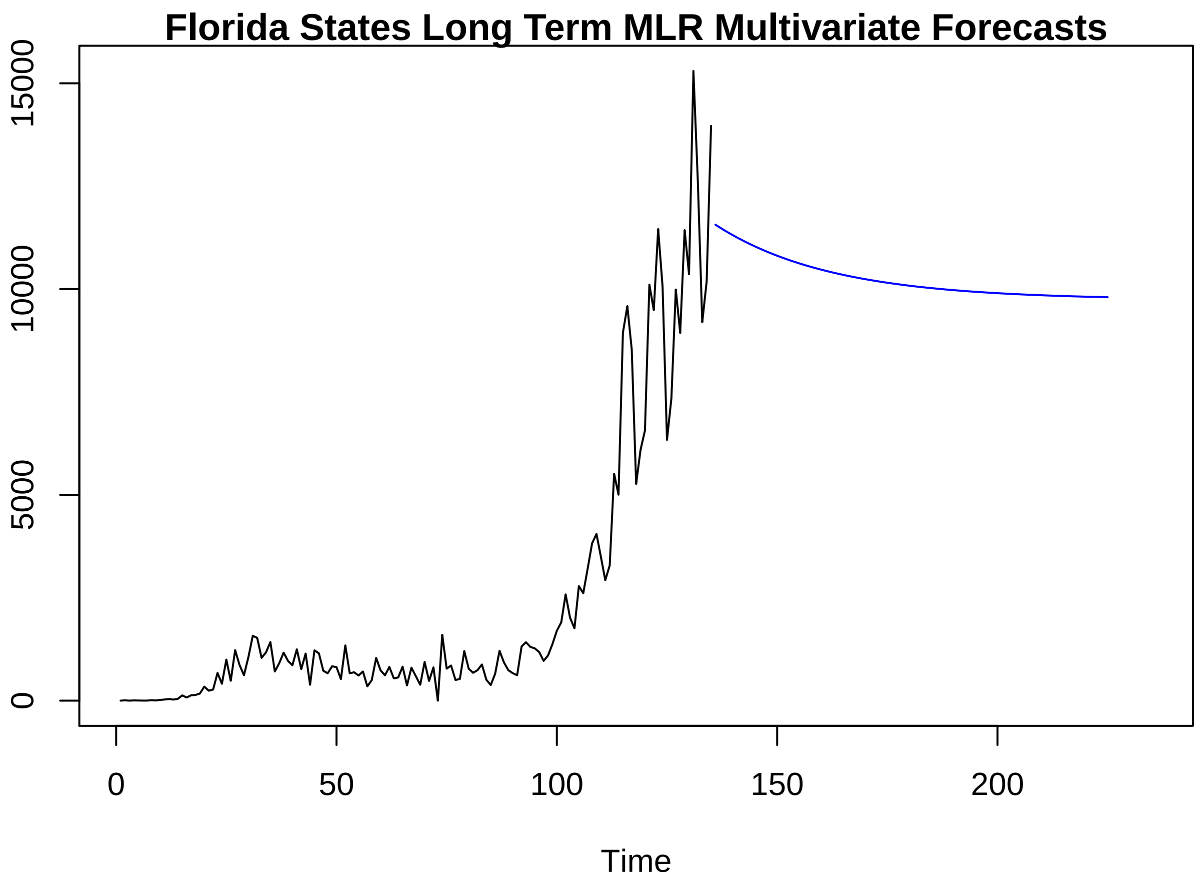

```
short_fl_mlr_m  = preds + forecasts$f

#long
preds = predict(fit, newdata = longdata)
forecasts = fore.arma.wge(fit$residuals,phi = est1$phi,theta = est1$theta, lastn = FALSE,n.ahead = 90)
```



```
long_fl_mlr_m  = preds + forecasts$f


plot(newcases_fl$positiveIncrease, type = "l", xlim = c(1,145), main = "Florida Short Term MLR Multivari
lines(ts(short_fl_mlr_m, start = 136), col = "blue")
```




```
plot(newcases_fl$positiveIncrease, type = "l", xlim = c(1,235), main = "Florida States Long Term MLR Mul
lines(ts(long_fl_mlr_m,start = 136), col = "blue")
```



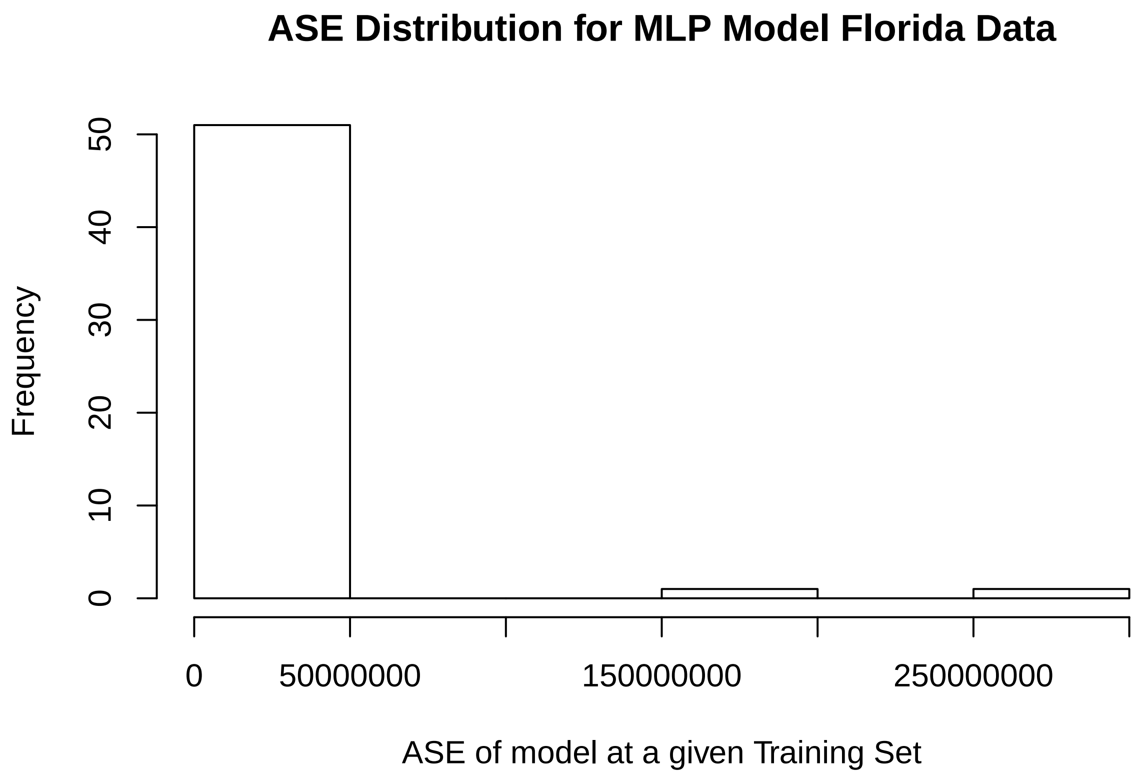
####Florida Multivariate MLP Cases Model

```
newcases_fl_multi =  initial_data_fl %>% dplyr::select(positiveIncrease, totalTestResultsIncrease, hosp
newcases_fl_var = cbind(ts(newcases_fl_multi$totalTestResultsIncrease),ts(newcases_fl_multi$hospitalize

trainingSize = 70
horizon = 12
ASEHolder = numeric()

#Out of bounds if it goes for 54 runs, this ASE will be slightly less wide than the others. But the win
for( i in 1:(135-(trainingSize + horizon) ))
{
  mlp.fit = mlp(ts(newcases_fl_multi$positiveIncrease[1:trainingSize+i]), hd = 5, comb = "median", xreg
  forecasts = forecast(mlp.fit,h = horizon, xreg = newcases_fl_var[1:(trainingSize + i + 12),])

  ASE = mean((newcases_fl_multi$positiveIncrease[(trainingSize+i):(trainingSize+ i + (horizon) - 1)] -f
  ASEHolder[i] = ASE

}

ASEHolder
```

```
##  [1]    444153.44    194060.03    712958.65    903230.74     93626.46
##  [6]    262553.62    683313.71   1311789.25    525867.47     77420.21
## [11]    151935.46    111753.21    277586.38    221647.60    294752.61
## [16]    401838.94    142857.79    263538.47    162420.22    298807.46
## [21]    612585.87    707084.56    618467.53   1092515.95   1078234.70
```

```
## [26]    860323.96   2116833.27   2896532.04   2513659.93   2965045.82
## [31]   2223309.84   1449933.05   2619105.27   6259345.20  11106408.50
## [36]  11151344.74  10602421.29   8788979.76   8739962.12   8360816.99
## [41]  13457932.79  21281998.77  12992823.99  10721605.74 152094225.32
## [46] 250526636.74  44318528.65  40428553.15  15077537.49  18401073.75
## [51]  16003705.56  16142526.33   8093564.28
```

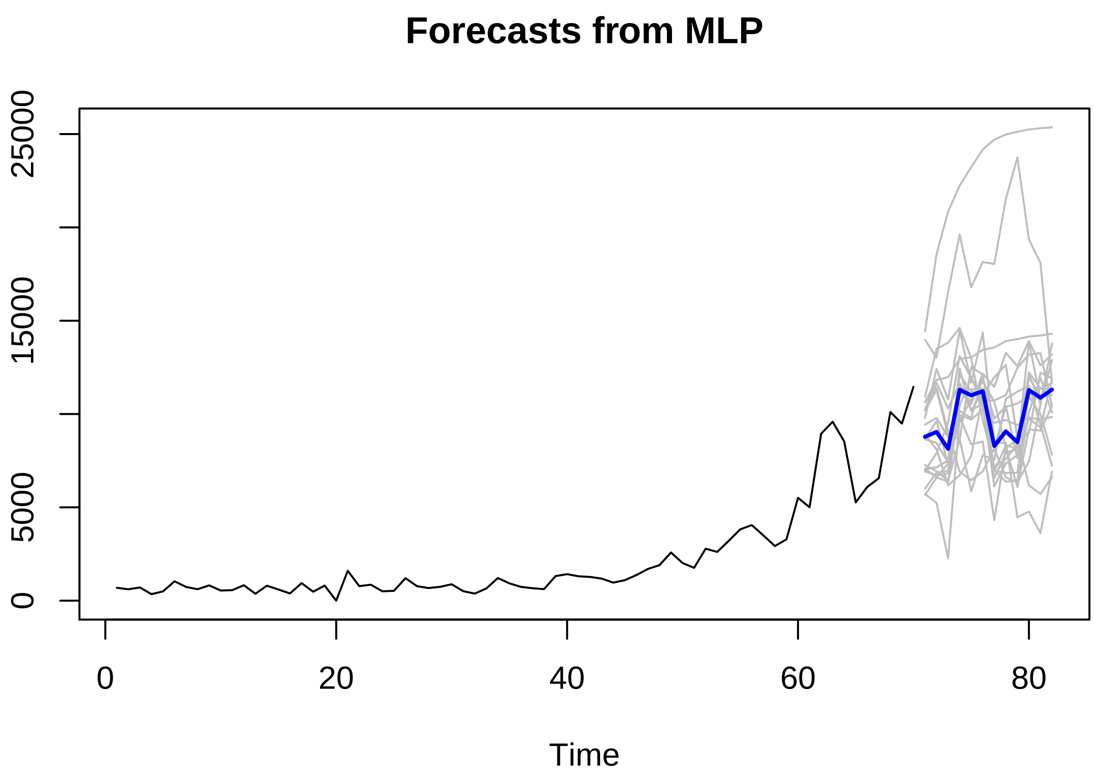
```
#Distribution of ASEs on Two Week Periods
hist(ASEHolder, xlab = "ASE of model at a given Training Set",  main = "ASE Distribution for MLP Model F
```



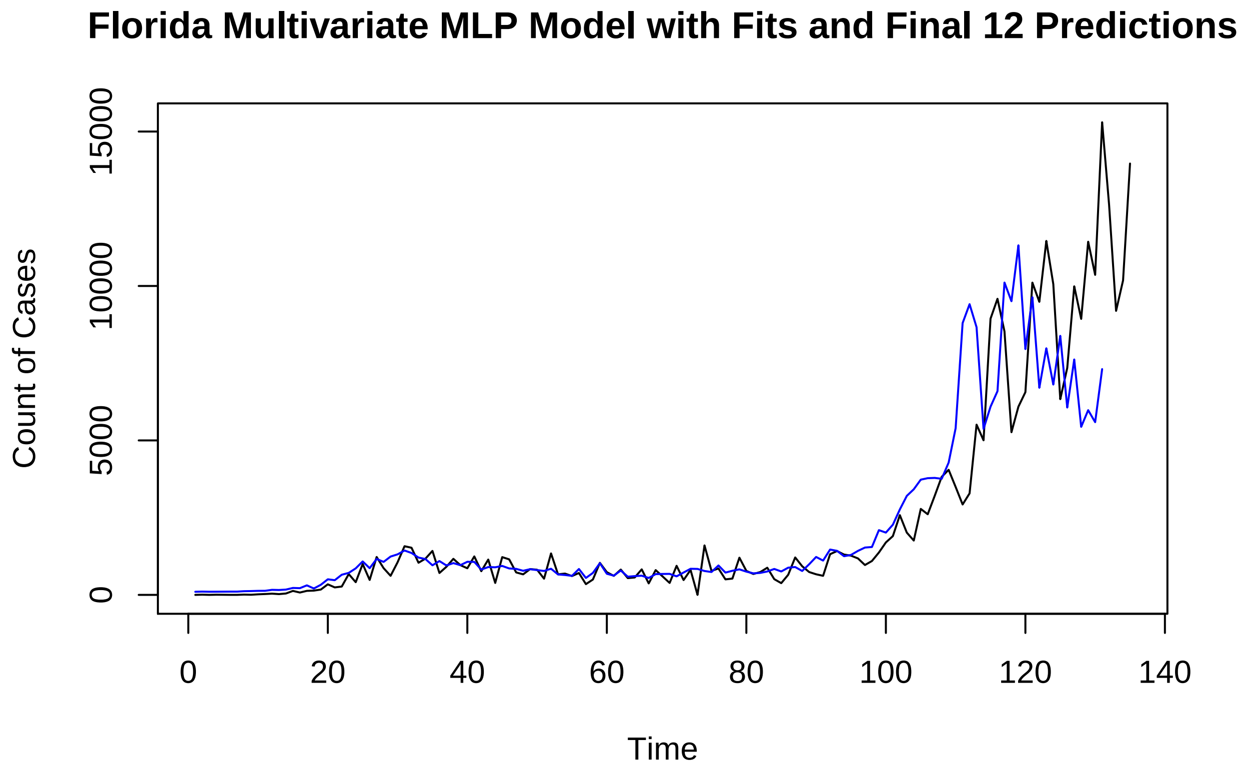

```
#Mean ASE
WindowedASE = mean(ASEHolder)
WindowedASE
```

```
## [1] 13468674
```

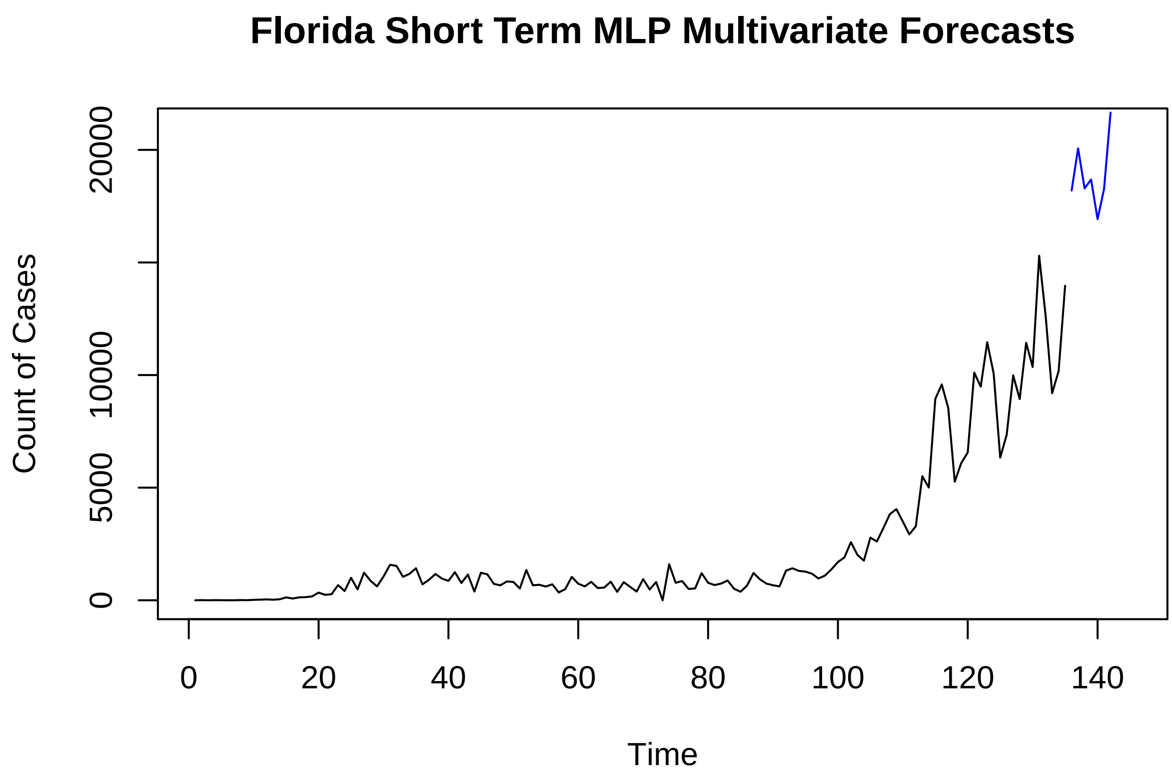
```
#18757436 - 18 mill

plot(forecasts)
```



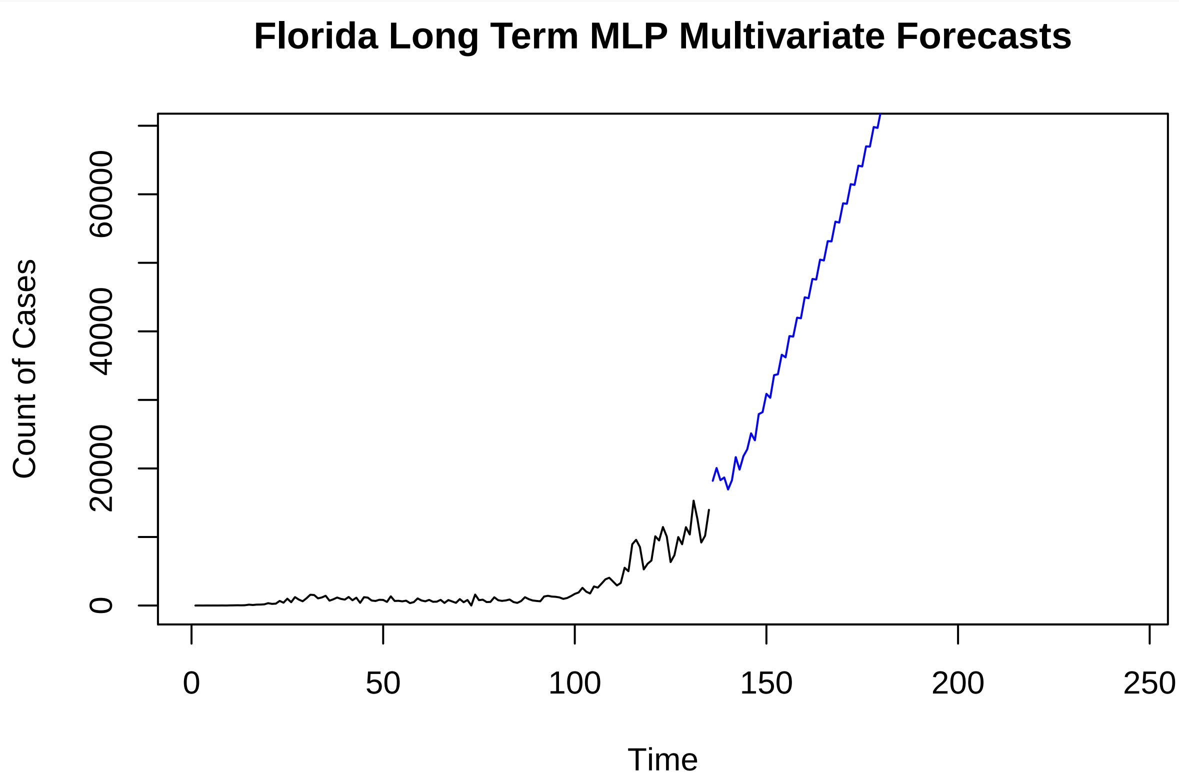

```
#Final Forecasts with data known
mlp.fit = mlp(ts(newcases_fl_multi$positiveIncrease[1:123]), hd = 5, comb = "median", xreg =newcases_fl_
forecasts = forecast(mlp.fit,h = 12, xreg = newcases_fl_var[1:135,])
fl_multi_mlp_fore = forecasts$mean

all_f = c(forecasts$fitted, forecasts$mean)
plot(newcases_fl_multi$positiveIncrease, type = "l", main = "Florida Multivariate MLP Model with Fits an
lines(all_f, col = "blue")
```




```r
#Forecast beyond data
mlp.fit = mlp(ts(newcases_fl_multi$positiveIncrease), hd = 5, comb = "median", xreg = newvar_fore_fl[1:
short_fl_mlp_m = forecast(mlp.fit,h = 7, xreg = newvar_fore_fl[1:145,])
long_fl_mlp_m = forecast(mlp.fit,h = 90, xreg = newvar_fore_fl[1:225,])

plot(newcases_fl$positiveIncrease, type = "l", xlim = c(1,145),ylim = c(0,21000), main = "Florida Short
lines(short_fl_mlp_m$mean, col = "blue")
```




```
plot(newcases_fl$positiveIncrease, type = "l", xlim = c(1,245),ylim = c(0,69000), main = "Florida Long T
lines(long_fl_mlp_m$mean, col = "blue")
```

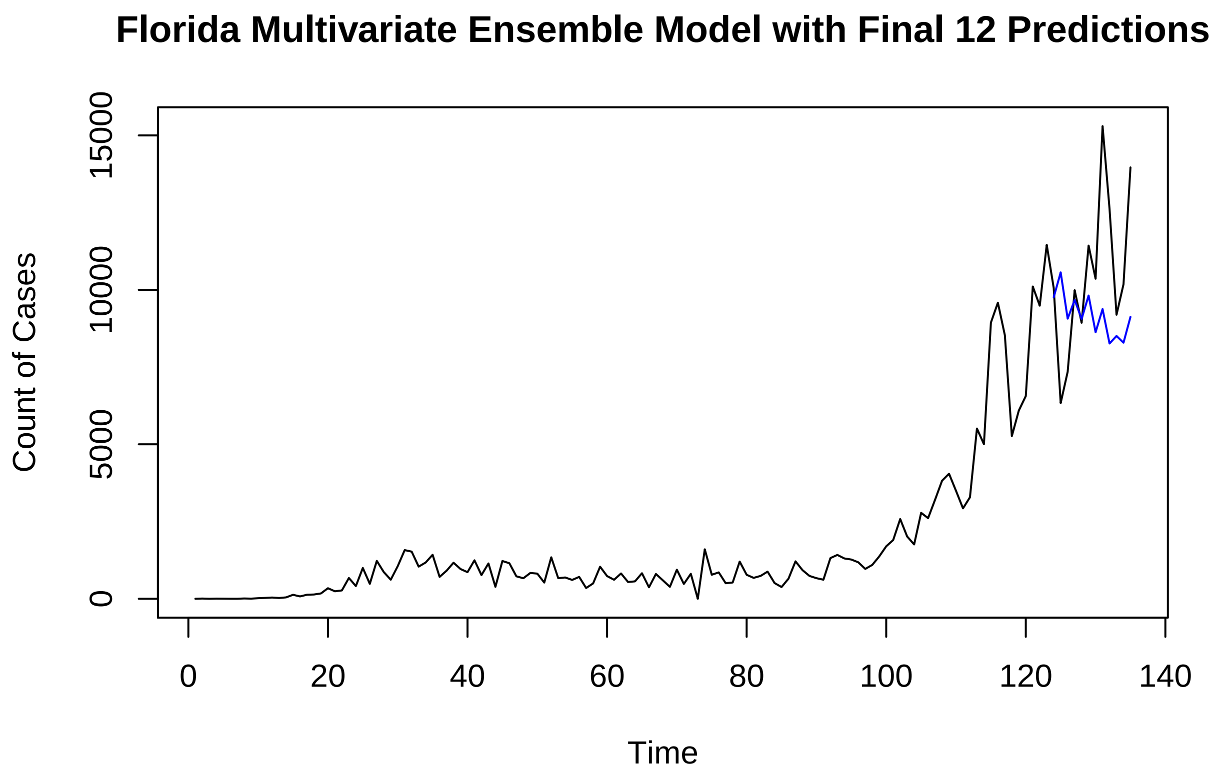



```
final_12_ase = mean((newcases_fl_multi$positiveIncrease[124:135] -fl_multi_mlp_fore)^2)
final_12_ase
```

```
## [1] 19855091
```

####Florida Multivariate Ensemble Model

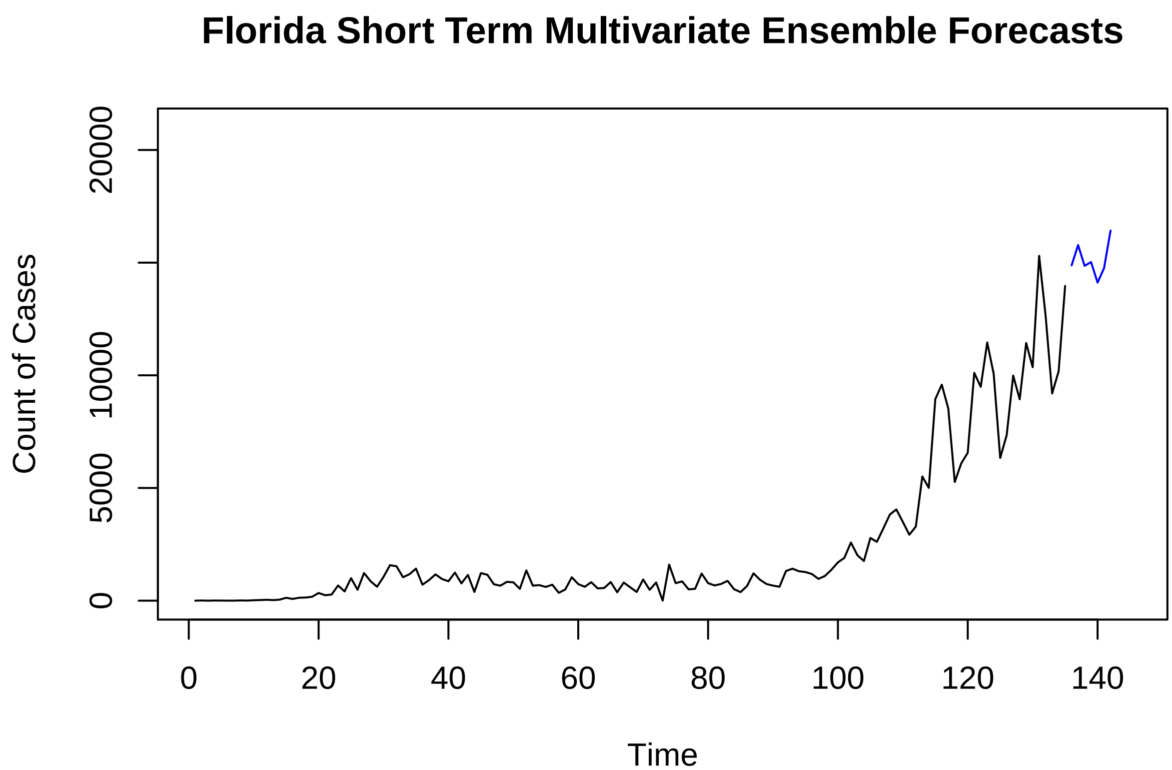
```r
ensemble_fore = (fl_multi_mlp_fore + FinalPredictions_fl_MLR)/2

plot(newcases_fl_multi$positiveIncrease, type = "l", main = "Florida Multivariate Ensemble Model with Fi
lines(ensemble_fore, col = "blue")
```



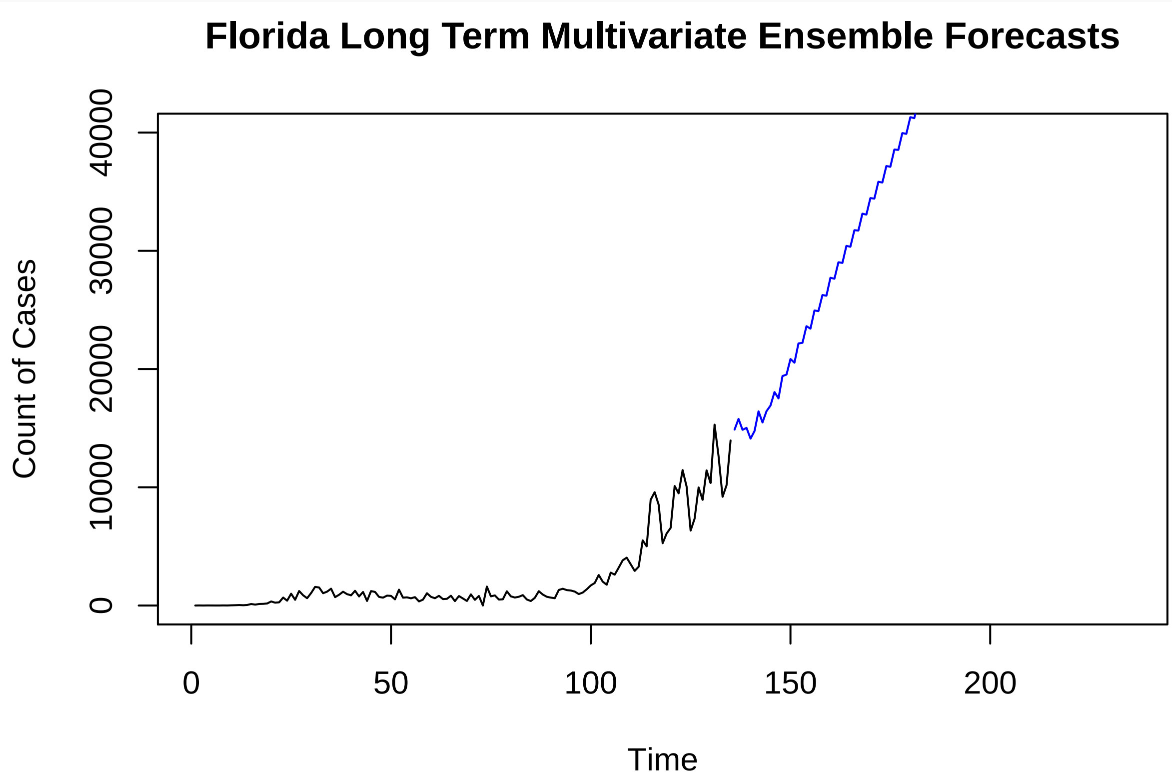

```r
ASE_fl_multi = mean((newcases_fl_multi$positiveIncrease[124:135] -ensemble_fore)^2)
ASE_fl_multi
```

```
## [1] 9018759
```

```r
#ASE of 8,427,522

#future

#long_fl_mlp_m
#short_fl_mlp_m


ensemble_fl_fore_short = ( short_fl_mlp_m$mean+ short_fl_mlr_m)/2
ensemble_fl_fore_long = (long_fl_mlp_m$mean + long_fl_mlr_m)/2

plot(newcases_fl$positiveIncrease, type = "l", xlim = c(1,145),ylim = c(0,21000), main = "Florida Short
lines(ensemble_fl_fore_short, col = "blue")
```




```
plot(newcases_fl$positiveIncrease, type = "l", xlim = c(1,235), ylim = c(0,40000),main = "Florida Long T
lines(ensemble_fl_fore_long, col = "blue")
```

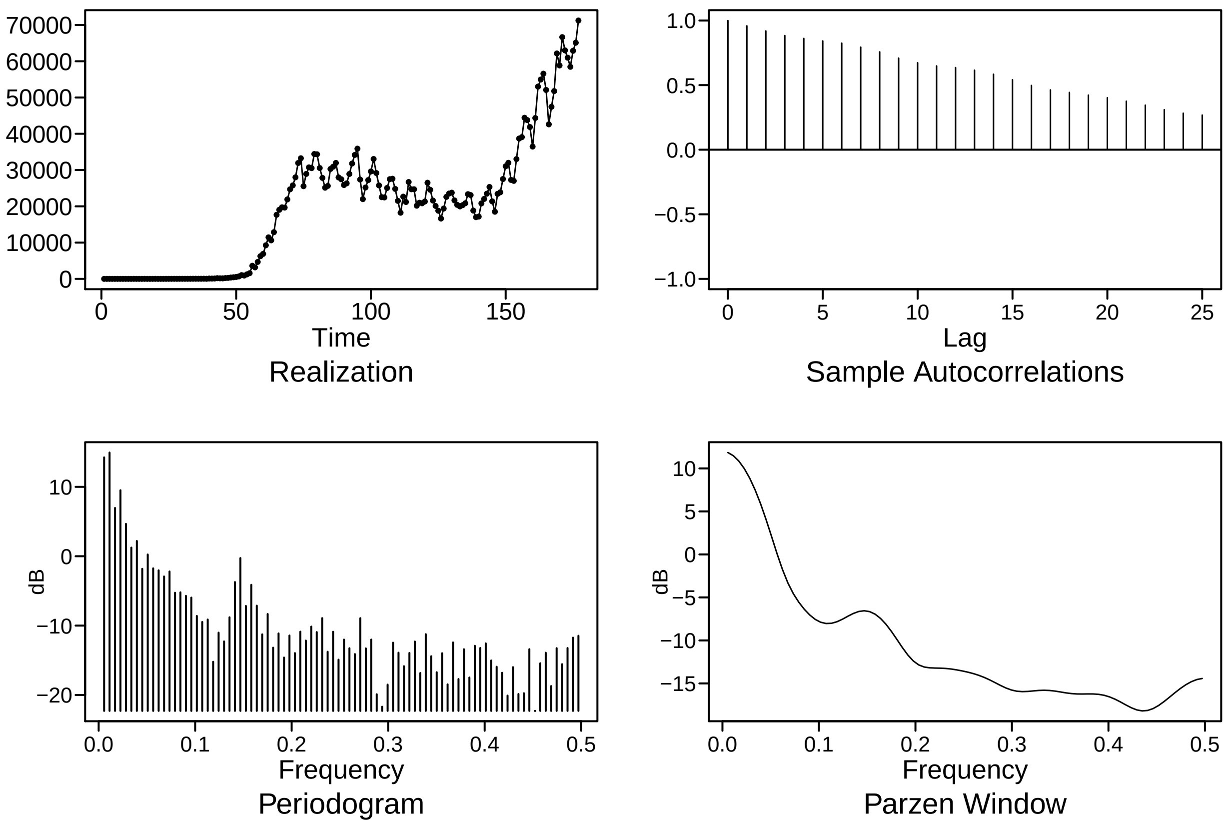


###Multivariate US Models

A. Stationarity vs Non-Stationarity

Overall we see slowly dampening ACFs, combined with a strong aperiodic frequency at zero in our spectral density. These measures alone with a recently quickly rising case count in recent days gives us strong evidence that our data is non-stationary. Given Covid19 spread, it is likely we see continued rising behavior in the short term, some return to lower numbers in the coming months but more uncertainty as new spikes could arise, and in the longest term of years on, we expect new cases to diminish to zero once the pandemic has ended spread.

```
#no text output
x = plotts.sample.wge(newcases_us$positiveIncrease)
```



B. Non-Stationary Modeling

We did not do any differencing of our data set to account for this non-stationarity. Going into this project we knew that because of the failure to contain the Covid-19 outbreak we would see large spikes of cases in recent time periods compared to distant time periods. We feel that this is an important aspect of our data that we want to portray in our models because we can see empirically in Florida and the United States as a whole that both individual behavior and political policy continue to trend towards further outbreak and rapid, almost exponential daily case growth. While some states with compliant individual behavior and strong political Covid-19 policies have shown "completed" Covid-19 curves, where daily case count begins to trend downwards towards zero, Florida is the opposite. Therefore, since we empirically expect the trend of non-stationarity to continue, we want that represented in our models. This is a fundamental assumption that our models are built on.

C. Model IDing of stationary models

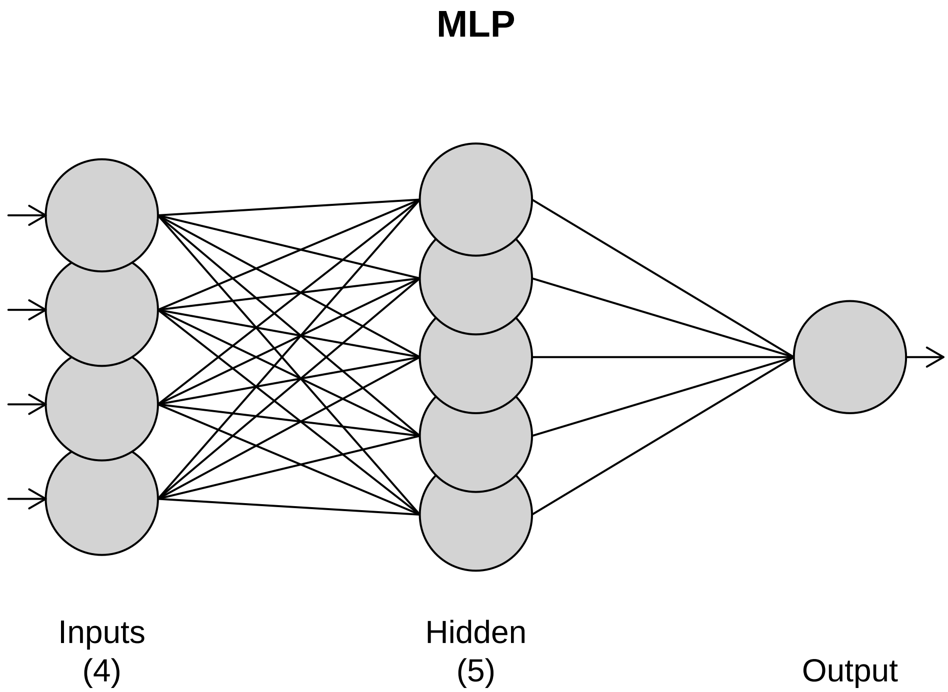
In multivariate modeling, our identification of models occured specifically for each model and can be found at the beginning of those sections in particular.

D. Model Building

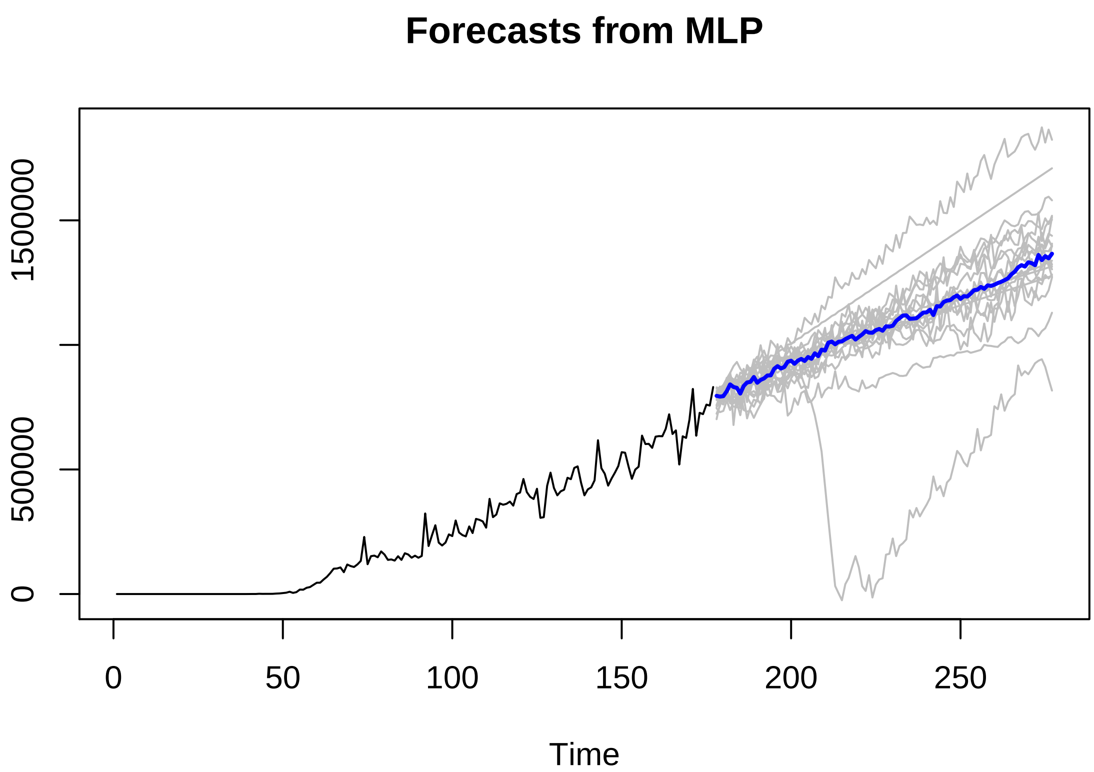
In order to forecast multivariate models of new case numbers, we will first fit some new variables for future

MLP type models.

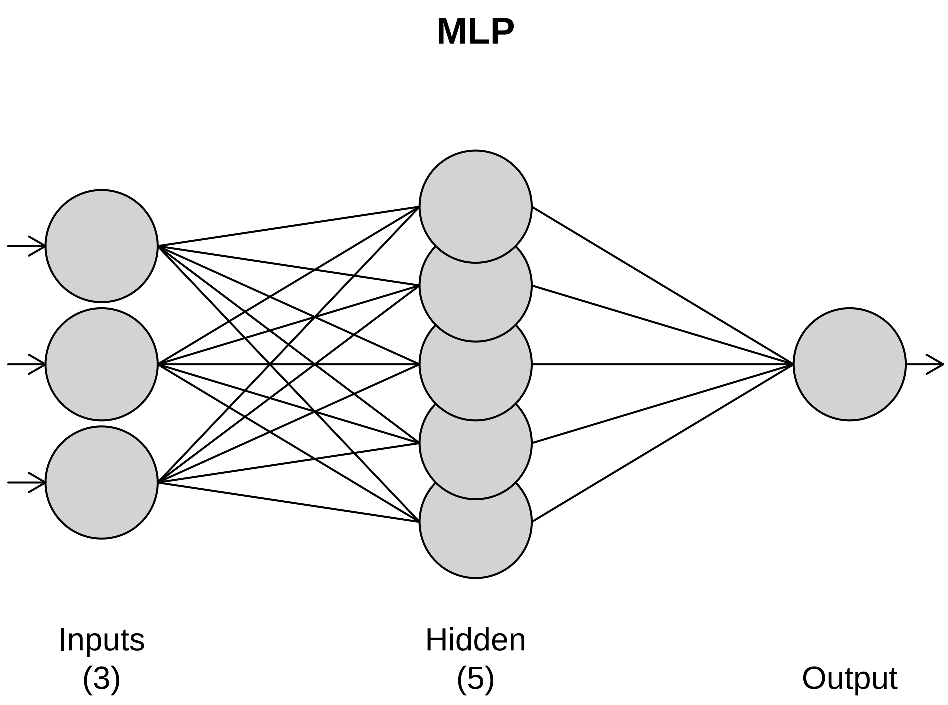
```
newcases_us_multi =  initial_data_us %>% dplyr::select(positiveIncrease, totalTestResultsIncrease, hospi

#Forecast Future


#Forecast future variables
fit.mlp.1 = mlp(ts(newcases_us_multi$totalTestResultsIncrease),reps = 20, comb = "median")
plot(fit.mlp.1)
```



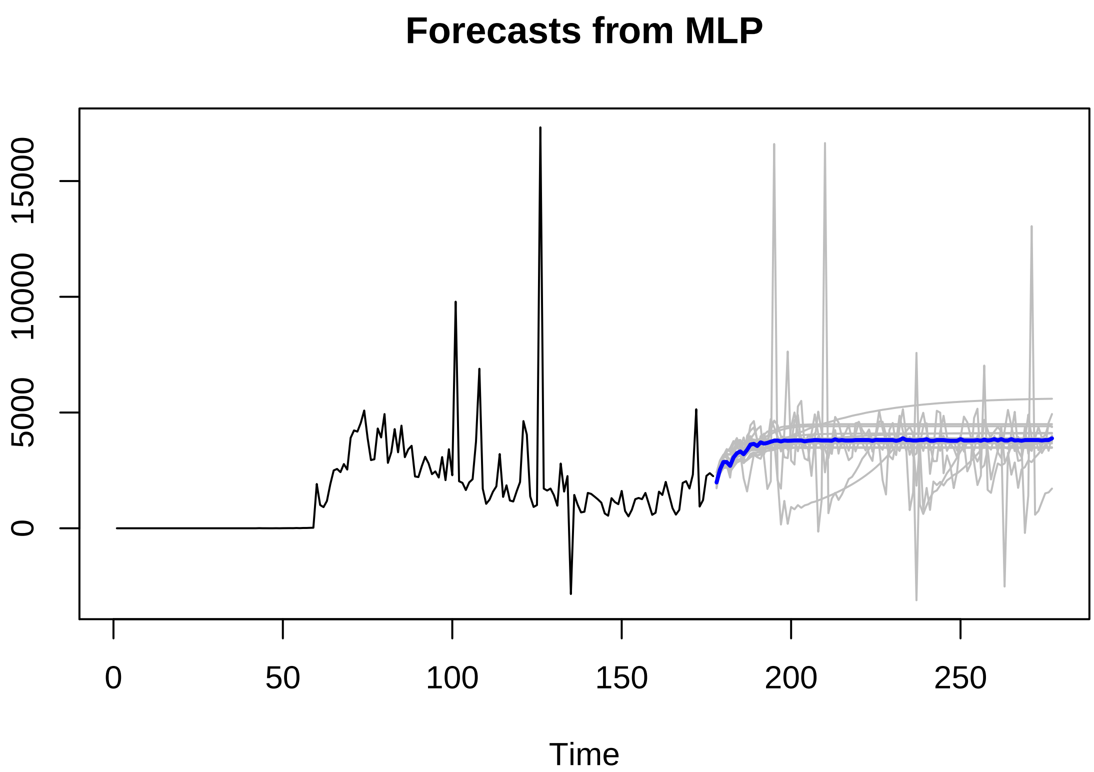

```
fore.mlp.1 = forecast(fit.mlp.1, h = 100)
plot(fore.mlp.1)
```



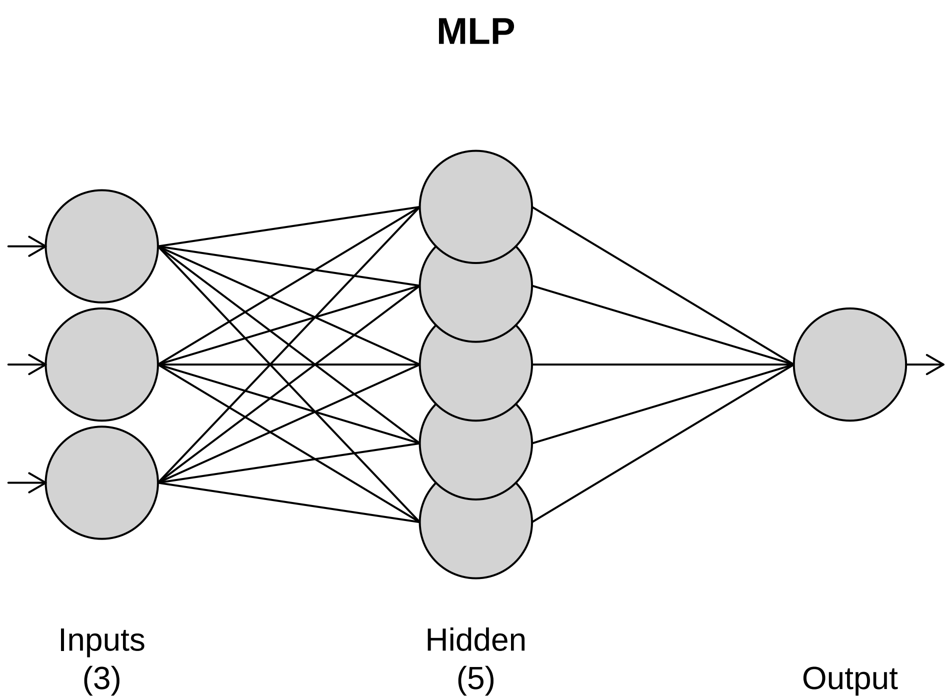

```
fit.mlp.2 = mlp(ts(newcases_us_multi$hospitalizedIncrease),reps = 20, comb = "median")
plot(fit.mlp.2)
```



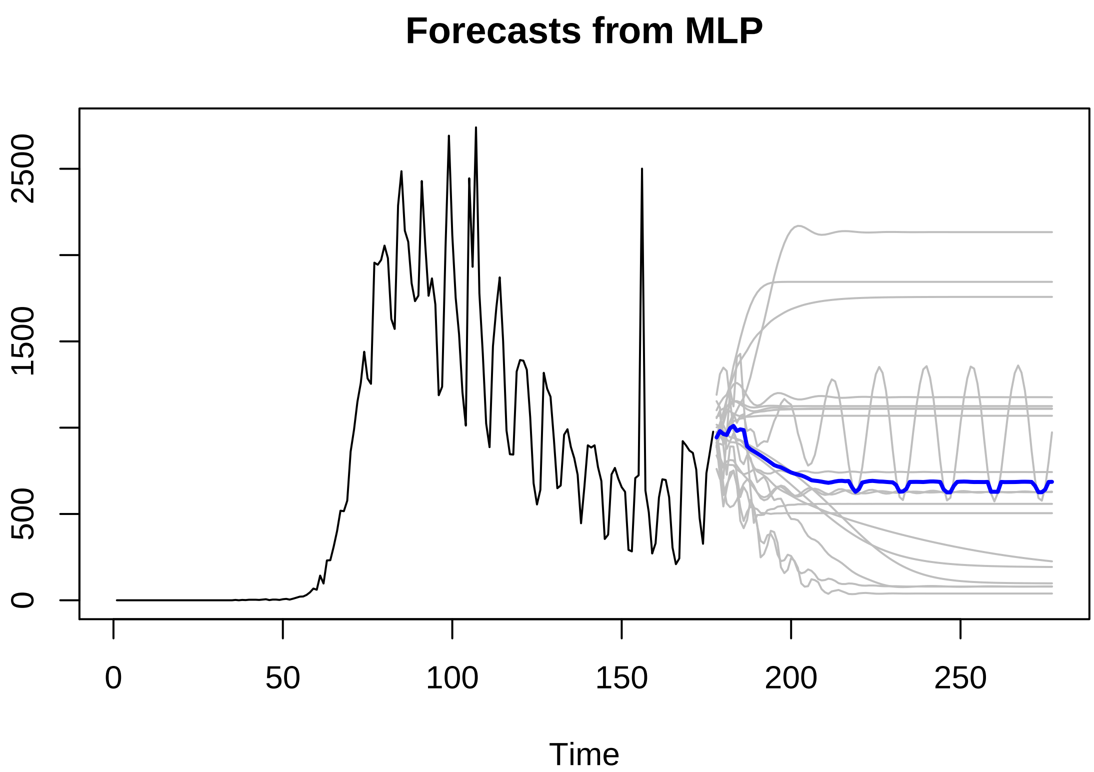

```
fore.mlp.2 = forecast(fit.mlp.2, h = 100)
plot(fore.mlp.2)
```




```
fit.mlp.3 = mlp(ts(newcases_us_multi$deathIncrease),reps = 20, comb = "median")
plot(fit.mlp.3)
```




```
fore.mlp.3 = forecast(fit.mlp.3, h = 100)
plot(fore.mlp.3)
```




```
#package them up in data frame.
newvar_fore_us = data.frame(totalTestResultsIncrease = ts(c(newcases_us_multi$totalTestResultsIncrease,i

dim(newvar_fore_us)
```

```
## [1] 277   3
```

####US MLR with Correlated Errors Model

```
fit = lm(positiveIncrease~totalTestResultsIncrease + hospitalizedIncrease, data = newcases_us_multi[1:16
summary(fit)
```

```
##
## Call:
## lm(formula = positiveIncrease ~ totalTestResultsIncrease + hospitalizedIncrease,
##     data = newcases_us_multi[1:165, ])
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -42222  -3101  -2878   5316  14897
##
## Coefficients:
##                             Estimate  Std. Error t value
## (Intercept)              3101.227341  869.662497   3.566
## totalTestResultsIncrease    0.049977    0.002809  17.793
## hospitalizedIncrease        2.336046    0.290613   8.038
##                                      Pr(>|t|)
## (Intercept)                          0.000477 ***
## totalTestResultsIncrease < 0.0000000000000002 ***
## hospitalizedIncrease         0.00000000000018 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7349 on 162 degrees of freedom
## Multiple R-squared:  0.743,  Adjusted R-squared:  0.7399
## F-statistic: 234.2 on 2 and 162 DF,  p-value: < 0.00000000000000022
```

```
est_tests = mean(tail(newcases_us_multi$totalTestResultsIncrease))
est_hospital= mean(tail(newcases_us_multi$hospitalizedIncrease))
newdata = data.frame(totalTestResultsIncrease = rep(est_tests,12), hospitalizedIncrease = rep(est_hospi

preds = predict(fit, newdata = newdata)



aic5.wge(fit$residuals)#picks 3,2 with full data
```

```
## ---------WORKING... PLEASE WAIT...
##
##
## Five Smallest Values of  aic
##       p    q        aic
## 5     1    1   16.83224
## 6     1    2   16.83708
## 11    3    1   16.83910
## 8     2    1   16.83932
## 13    4    0   16.83996
```

```
est1 = est.arma.wge(fit$residuals, p = 3, q = 2)
```

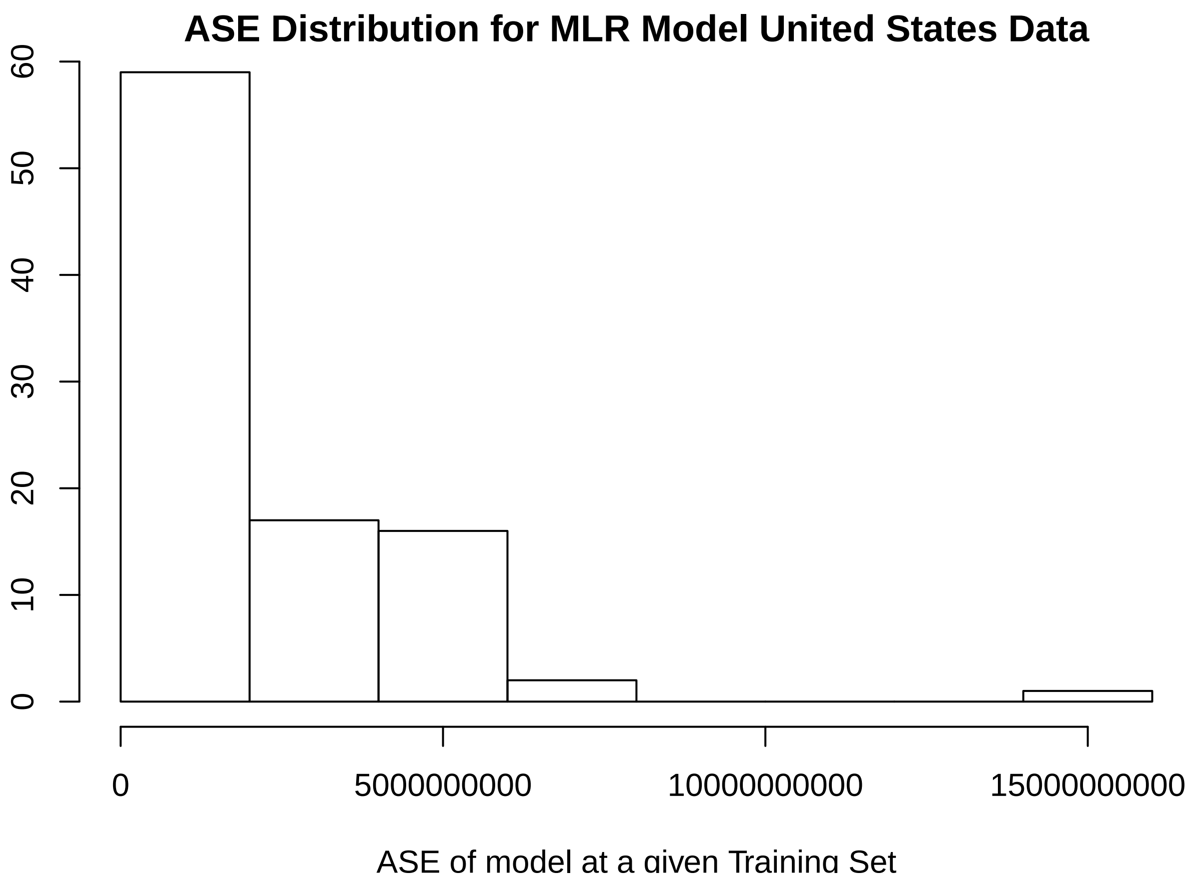
```
##
## Coefficients of Original polynomial:
## 1.9444 -0.9016 -0.0473
##
## Factor                 Roots                Abs Recip    System Freq
## 1-1.9919B+0.9961B^2    0.9998+-0.0652i      0.9981       0.0104
## 1+0.0474B             -21.0791               0.0474       0.5000
##
##
```

```
for( i in 1:(177-(trainingSize + horizon) ))
{
  fit = lm(positiveIncrease~totalTestResultsIncrease + hospitalizedIncrease, data = newcases_us_multi[1
  newdata = data.frame(totalTestResultsIncrease = rep(est_tests,horizon), hospitalizedIncrease = rep(es

  preds = predict(fit, newdata = newdata)
  forecasts = fore.arma.wge(fit$residuals,phi = est1$phi,theta = est1$theta, lastn = FALSE,n.ahead = ho

  final_pred  = preds + forecasts$f


  ASE = mean((newcases_us_multi$positiveIncrease[(trainingSize+i):(trainingSize+ i + (horizon) - 1)] -f
  ASEHolder[i] = ASE

}
```

```r
ASEHolder
```

```
##   [1]   5380274223  4914564759  4555455224  1708071709  1694436967
##   [6]   2545354790  3587609643  3594983226  4067557383  4050506483
##  [11]   5077832323  5176752257  5348294416  4838019350  4708622612
##  [16]   5227841612  5654222458  5676101641  6000663028  6078253699
##  [21]   5900331399  1554258063  1809395334  2252786028  2416734053
##  [26]   2461093918  2165006104  2061993158  1812244338  1917445023
##  [31]   2481645385  2516051039  2398456166  2214123680  1852137279
##  [36]   1749588012  1559140675  1356556650  1285173533  1199287506
##  [41]    771158389   673913491   642870645   532293973   548308227
##  [46]    520596381   438583785   414419021   374067423   251862631
##  [51]    205905745   121400107   974304484    23442192   504413661
##  [56]    323729831   112461824  1042354738   612690301    97034966
##  [61]    339480392  2226182776   152121062    15325519   116703533
##  [66]    566689479   474690881    21056191   291275630  1192411920
##  [71]    841358069 14317882811   812372140    99566604   332052455
##  [76]    974270102   417823732  3699683219  2362942726  2400494736
##  [81]    126143855   816363425    83089298   124121842  3827308783
##  [86]    512370767   971484081   147200875   119892934  1526454702
##  [91]     75284485    49826470  5221093458  4388516429   628353183
```

```r
#Distribution of ASEs on Two Week Periods
hist(ASEHolder, xlab = "ASE of model at a given Training Set",  main = "ASE Distribution for MLR Model U
```

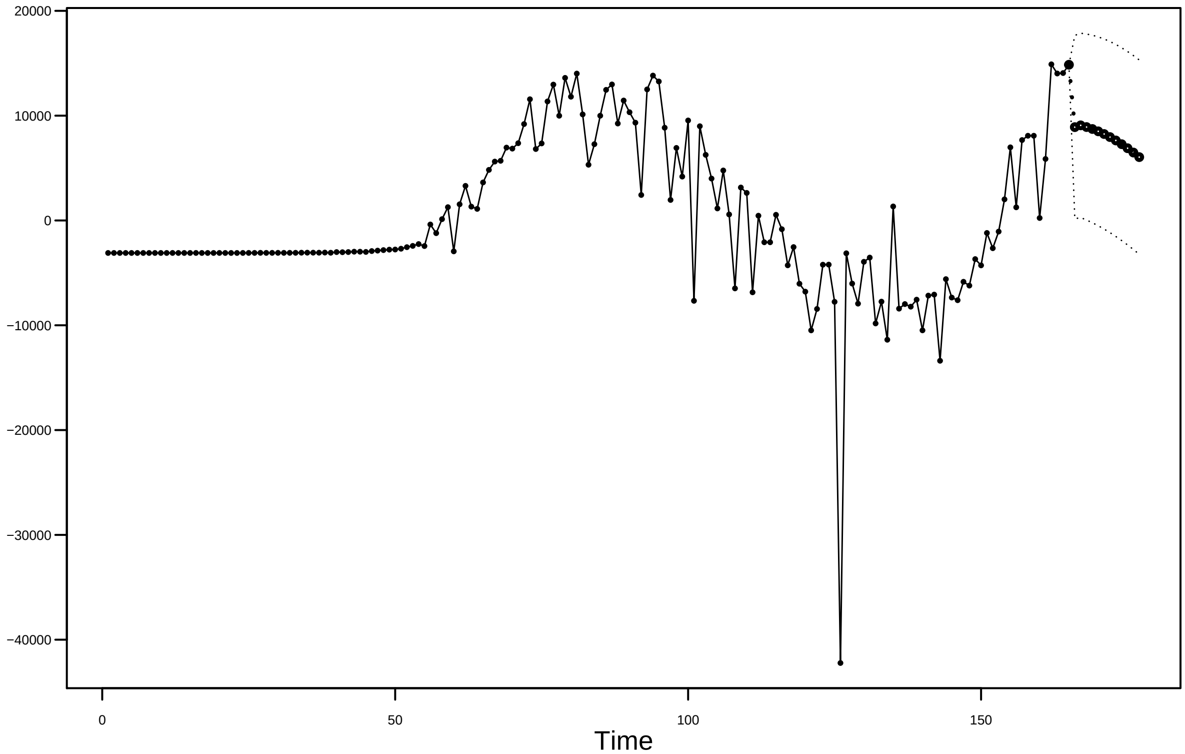



```r
#Mean ASE
WindowedASE = mean(ASEHolder)
```

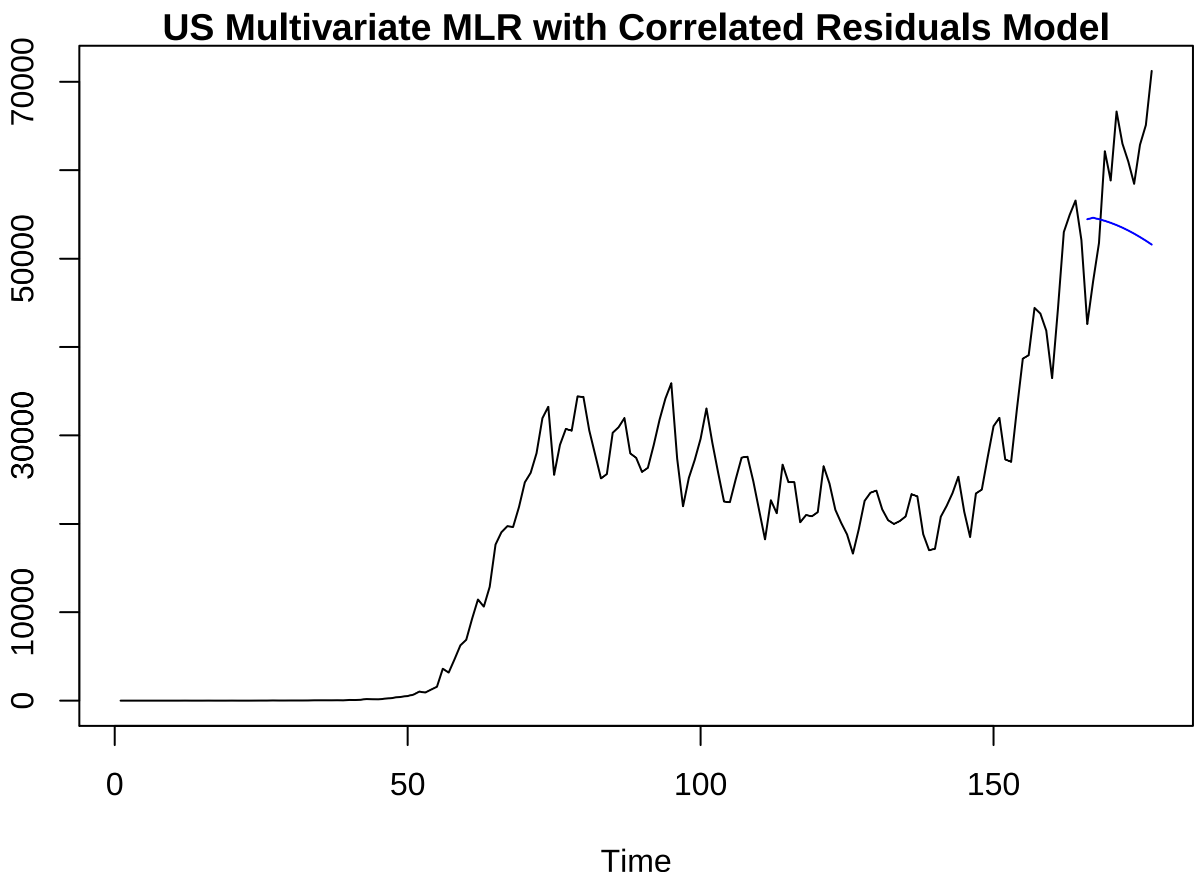
```r
WindowedASE
```

```
## [1] 2024279637
```

```r
fit = lm(positiveIncrease~totalTestResultsIncrease + hospitalizedIncrease, data = newcases_us_multi[1:16
preds = predict(fit, newdata = newdata)

forecasts = fore.arma.wge(fit$residuals,phi = est1$phi,theta = est1$theta, lastn = FALSE,n.ahead = 12)
```



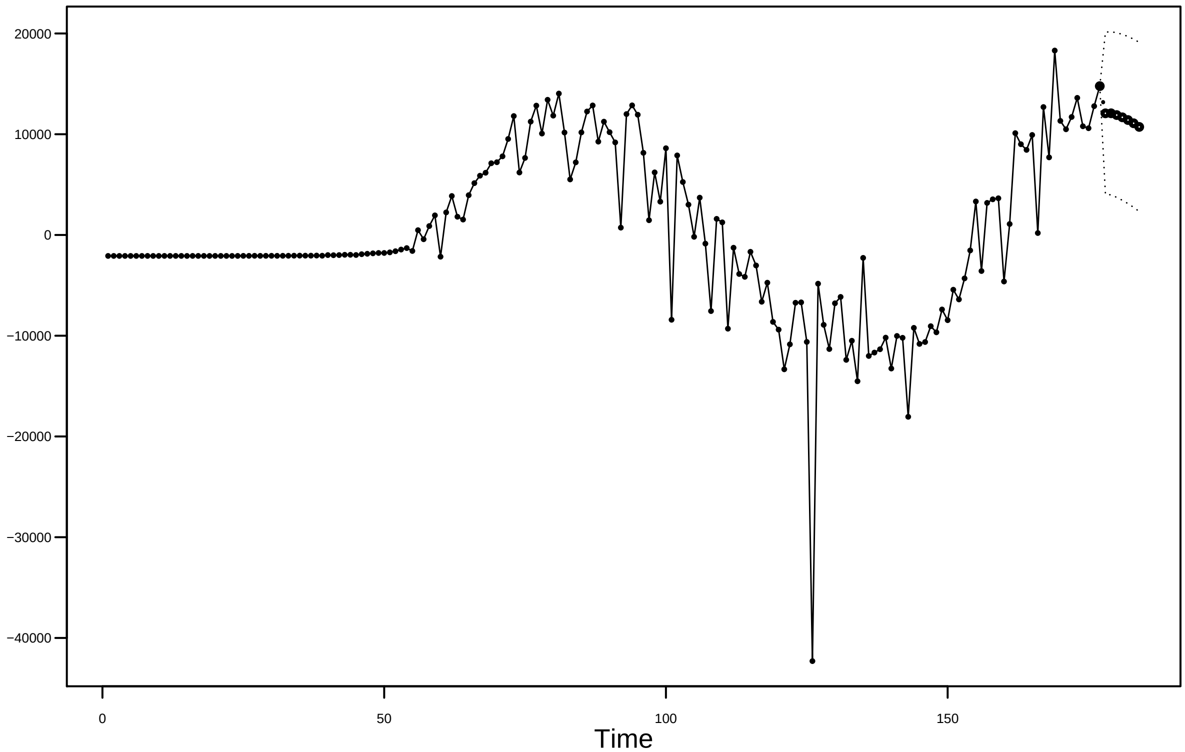

```r
FinalPredictions_us_MLR  = preds + forecasts$f

plot(newcases_us$positiveIncrease, type = "l", main ="US Multivariate MLR with Correlated Residuals Mode
lines(ts(FinalPredictions_us_MLR, start = 166), col = "blue")
```



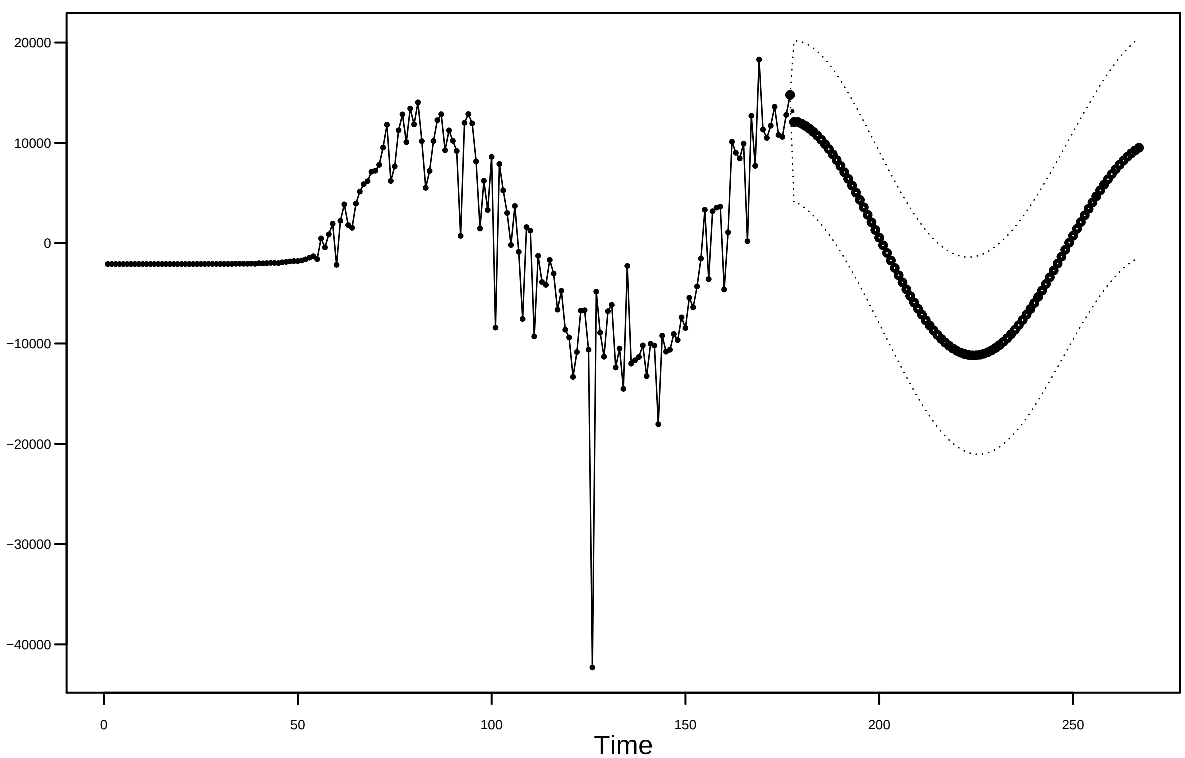

```
ASE = mean((newcases_us_multi$positiveIncrease[166:177] - FinalPredictions_us_MLR)^2)
ASE
```

```
## [1] 108181241
```

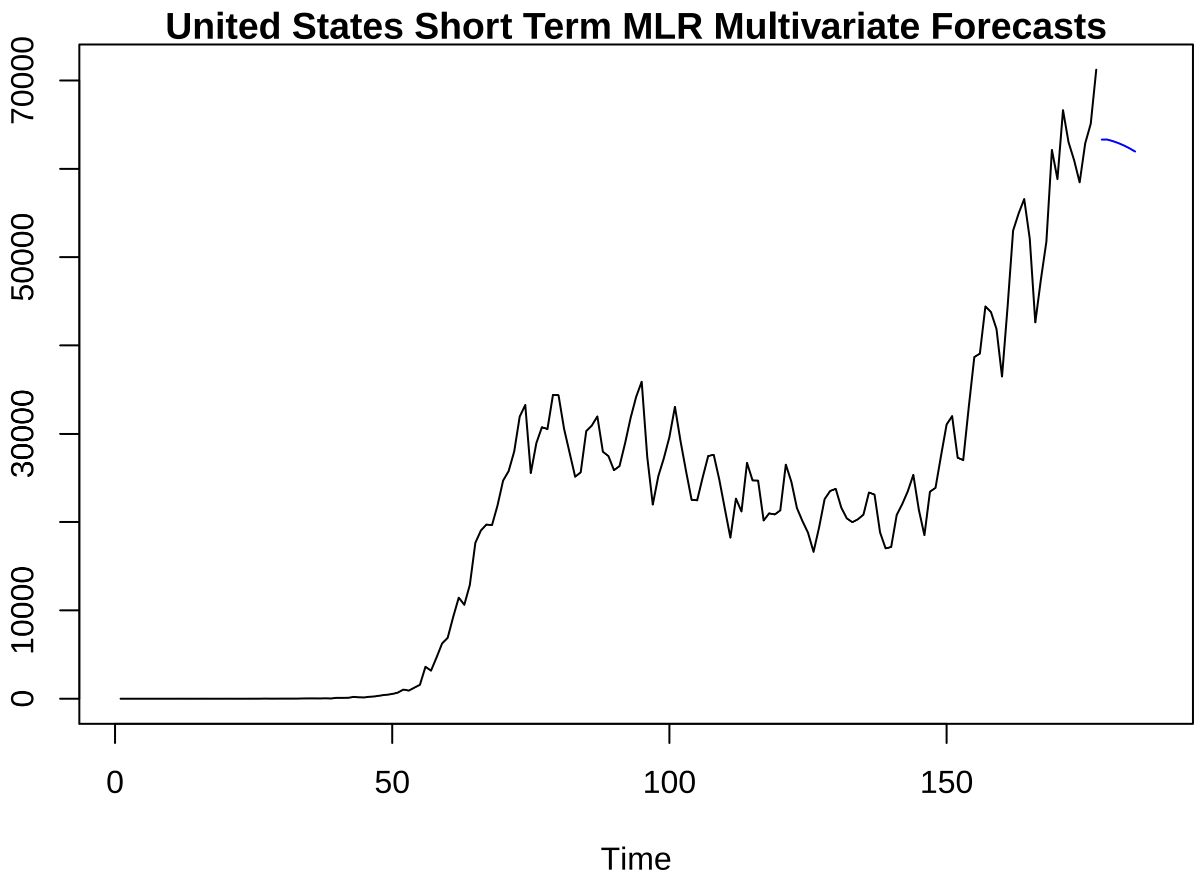
```
#

#Forecasting Ahead

shortdata = data.frame(totalTestResultsIncrease = rep(est_tests,7), hospitalizedIncrease = rep(est_hospi
longdata =data.frame(totalTestResultsIncrease = rep(est_tests,90), hospitalizedIncrease = rep(est_hospit


fit = lm(positiveIncrease~totalTestResultsIncrease + hospitalizedIncrease, data = newcases_us_multi)
#short
preds = predict(fit, newdata = shortdata)
forecasts = fore.arma.wge(fit$residuals,phi = est1$phi,theta = est1$theta, lastn = FALSE,n.ahead = 7)
```



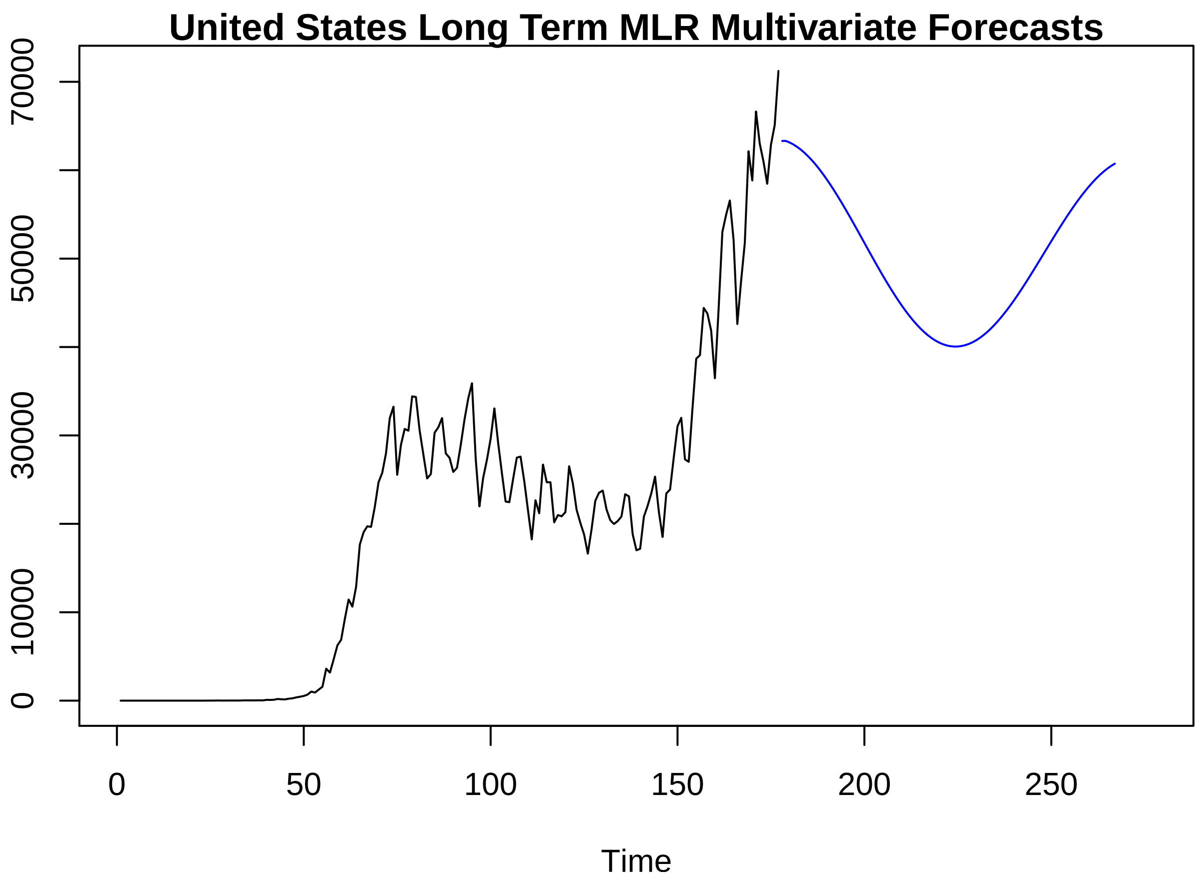

```r
short_us_mlr_m  = preds + forecasts$f

#long
preds = predict(fit, newdata = longdata)
forecasts = fore.arma.wge(fit$residuals,phi = est1$phi,theta = est1$theta, lastn = FALSE,n.ahead = 90)
```



```
long_us_mlr_m  = preds + forecasts$f


plot(newcases_us$positiveIncrease, type = "l", xlim = c(1,187), main = "United States Short Term MLR Mul
lines(ts(short_us_mlr_m, start = 178), col = "blue")
```




```
plot(newcases_us$positiveIncrease, type = "l", xlim = c(1,277), main = "United States Long Term MLR Mult
lines(ts(long_us_mlr_m,start = 178), col = "blue")
```



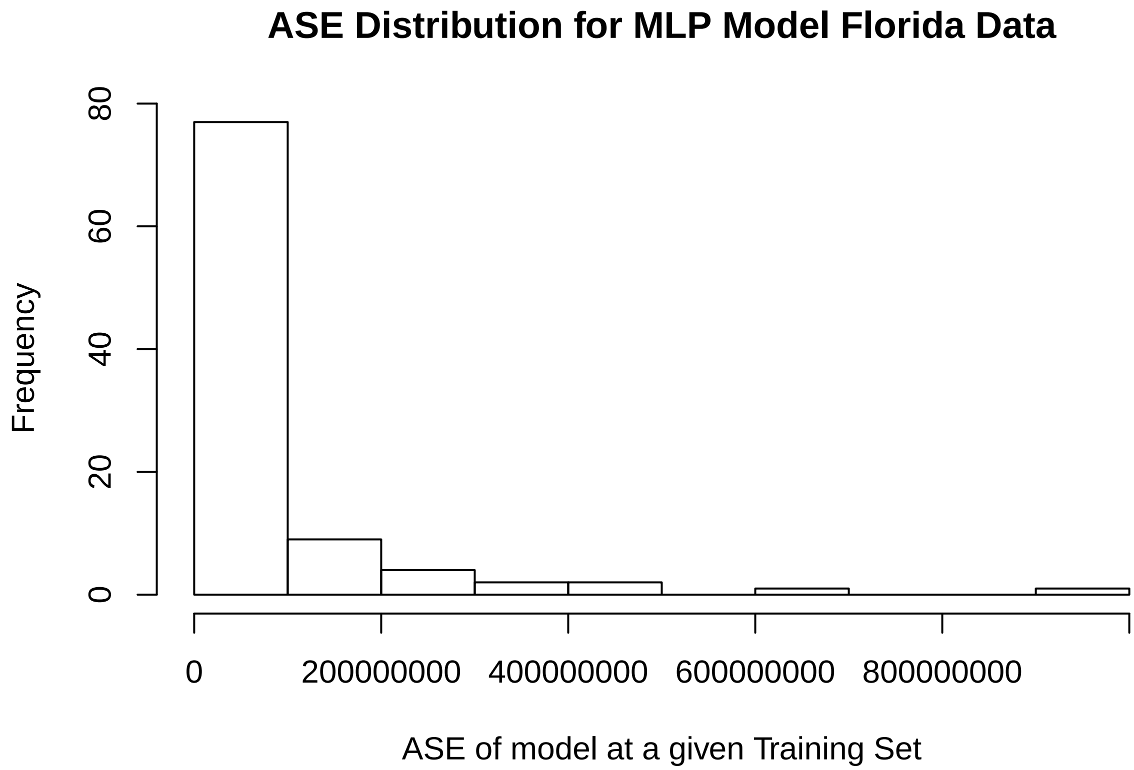
####US MLP/RNN Model

```
trainingSize = 70
horizon = 12
ASEHolder = numeric()

for( i in 1:(177-(trainingSize + horizon) + 1))
{
  mlp.fit = mlp(ts(newcases_us_multi$positiveIncrease[1:trainingSize+i]), hd = 5, comb = "median", xreg
  forecasts = forecast(mlp.fit,h = horizon, xreg = newvar_fore_us[1:(trainingSize + i + 13),])

  ASE = mean((newcases_us_multi$positiveIncrease[(trainingSize+i):(trainingSize+ i + (horizon) - 1)] -f
  ASEHolder[i] = ASE

}

ASEHolder
```

```
##  [1]  31611678 295608440 967166724 405102072 695004108  20181861  15107576
##  [8]  83099979 303918712 212524108 137064433  67148067  32795839  24937462
## [15]  18441362  33736934  26224626  34670369  38855630  33774080  25651710
## [22]  12259544  16708986  27452912  30499574  43702326  26145257  24846826
## [29]  27941411  43867937  52731830  70910206  61789627  52126325  46466672
## [36]  53222073  55769011  64291897  68278449  68270659  48428787  51984400
## [43]  54120095  70892537  83546948  94144578  88682358  88395125  79435008
## [50]  75032208  82763190  90642265  98382648  91049845  79403607  73247047
## [57]  25748267  40988375  93550417 100912375 107185941  93530400  85980007
```

```
## [64]   64603005  87205883  98720247  98064459  97248332  70454087  42253671
## [71]   77843813  68973543  72465252  66399147  49534046  59976850  60123877
## [78]   71602880  71176770  56921589  83978818  87467857 113859219 247857197
## [85] 157432693 160004389 156337986 110268093 329716206 246679351 425528281
## [92]   63281279 105103364  65869811  77484459  69528459
```

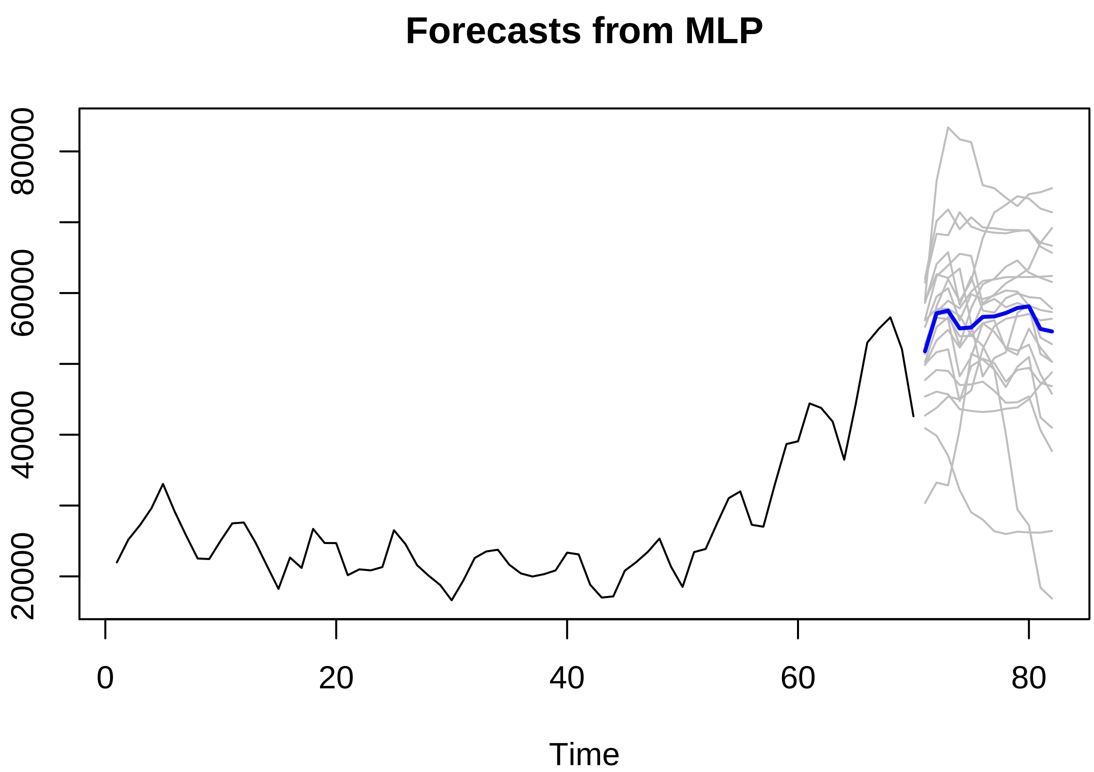
```
#Distribution of ASEs on Two Week Periods
hist(ASEHolder, xlab = "ASE of model at a given Training Set",  main = "ASE Distribution for MLP Model I
```



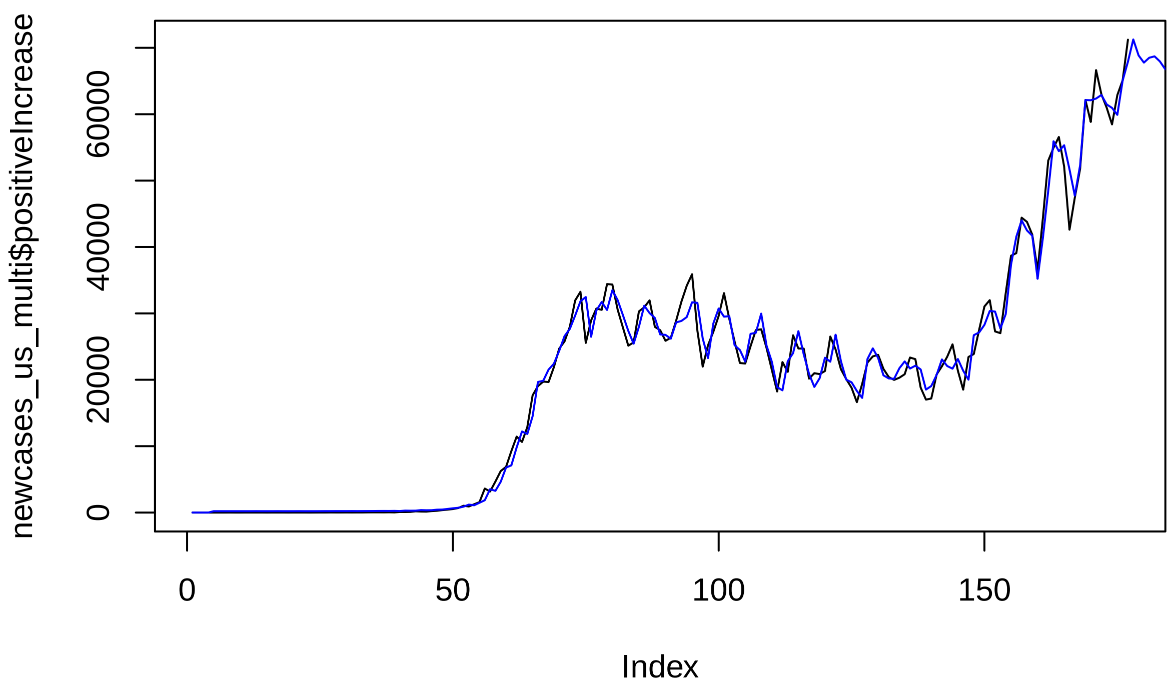

```
#Mean ASE
WindowedASE = mean(ASEHolder)
WindowedASE
```

```
## [1] 102728319
```

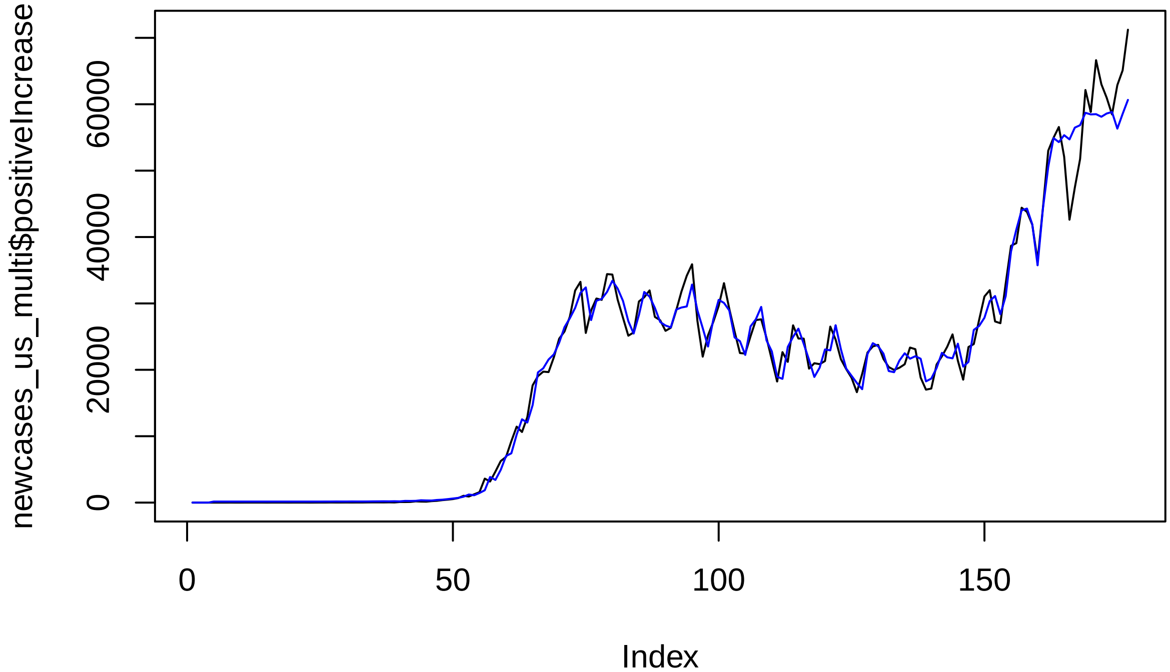
```
#97494363

plot(forecasts)
```




```
#Final Forecasts with data known
mlp.fit = mlp(ts(newcases_us_multi$positiveIncrease[1:177]), hd = 5, comb = "median", xreg = newvar_fore
forecasts_us_mlp = forecast(mlp.fit,h = 12, xreg = newvar_fore_us[1:190,])

all_f = c(rep(1,4),forecasts_us_mlp$fitted, forecasts_us_mlp$mean)
plot(newcases_us_multi$positiveIncrease, type = "l")
lines(all_f, col = "blue")
```



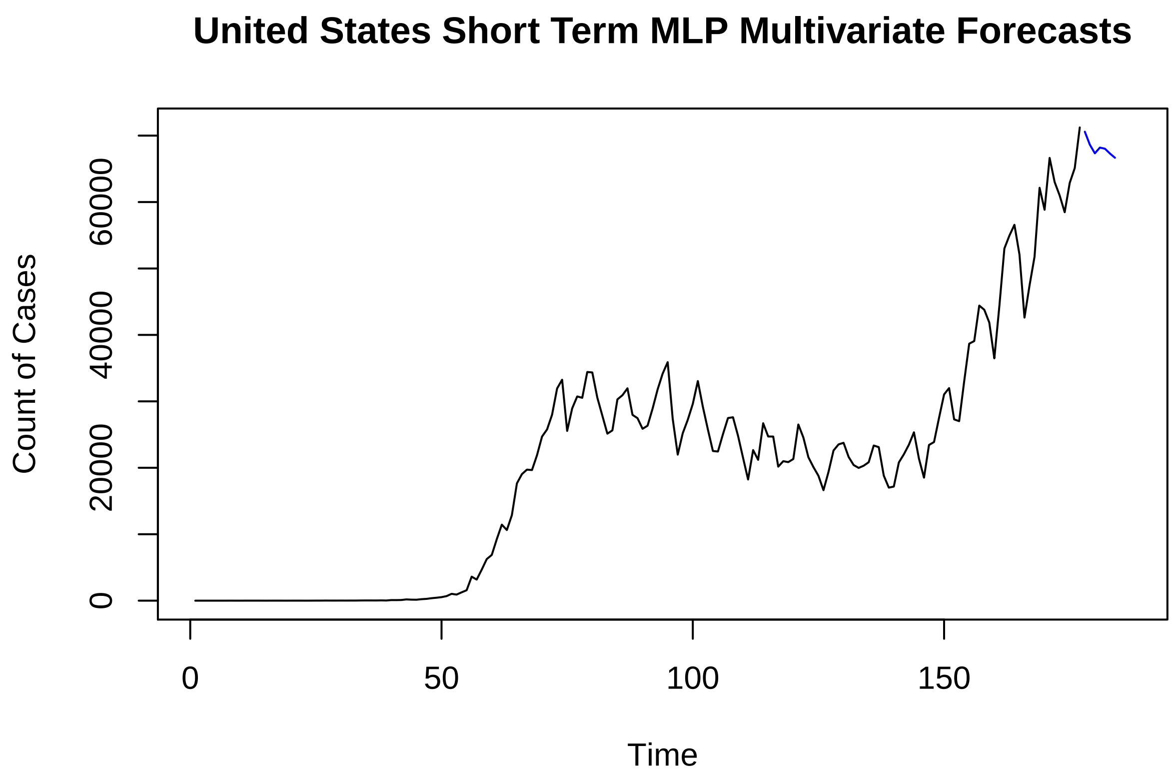
```r
#final 12 forecasts

mlp.fit = mlp(ts(newcases_us_multi$positiveIncrease[1:165]), hd = 5, comb = "median", xreg = newvar_fore
forecasts_us_mlp = forecast(mlp.fit,h = 12, xreg = newvar_fore_us[1:177,])

all_f = c(rep(1,4),forecasts_us_mlp$fitted, forecasts_us_mlp$mean)
plot(newcases_us_multi$positiveIncrease, type = "l")
lines(all_f, col = "blue")
```



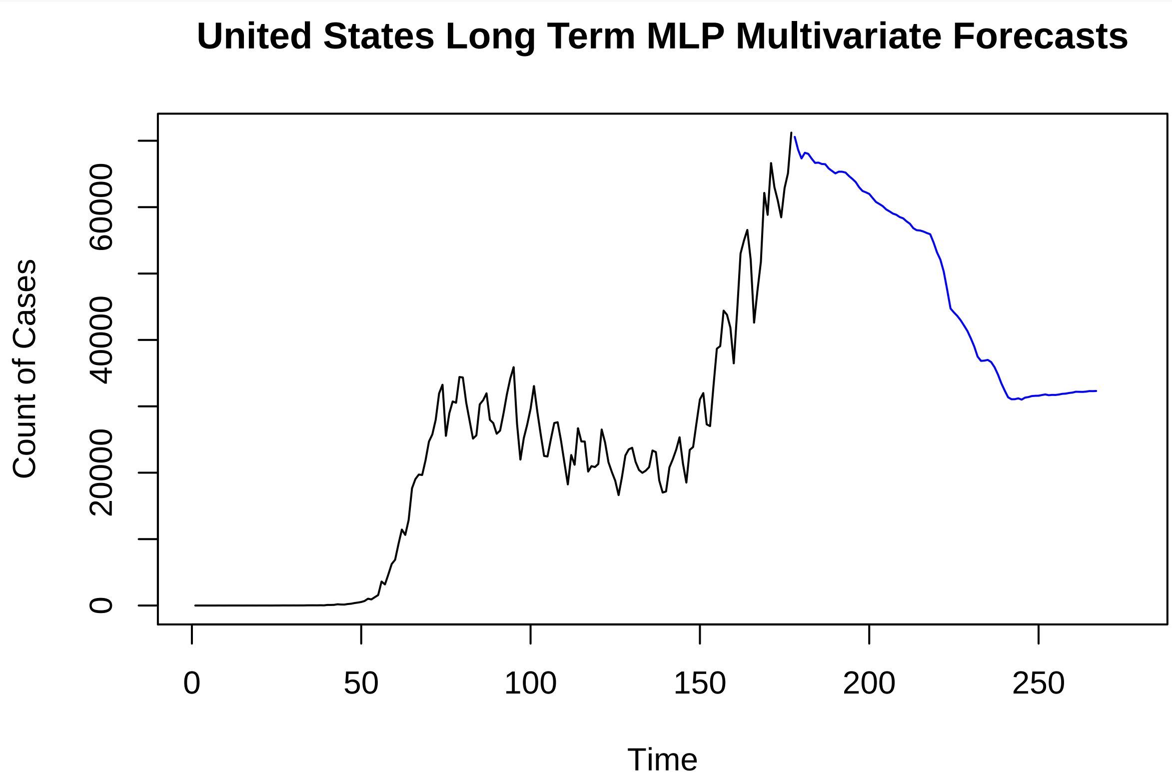

```r
ASE_final12 = mean((newcases_us_multi$positiveIncrease[166:177] -forecasts_us_mlp$mean)^2)
ASE_final12
```

```
## [1] 46652500
```

```r
#45799110


#Future Forecasts
mlp.fit = mlp(ts(newcases_us_multi$positiveIncrease), hd = 5, comb = "median", xreg = newvar_fore_us[1:
short_us_mlp_m = forecast(mlp.fit,h = 7, xreg = newvar_fore_us[1:187,])
long_us_mlp_m = forecast(mlp.fit,h = 90, xreg = newvar_fore_us[1:267,])

plot(newcases_us$positiveIncrease, type = "l", xlim = c(1,187), main = "United States Short Term MLP Mul
lines(short_us_mlp_m$mean, col = "blue")
```

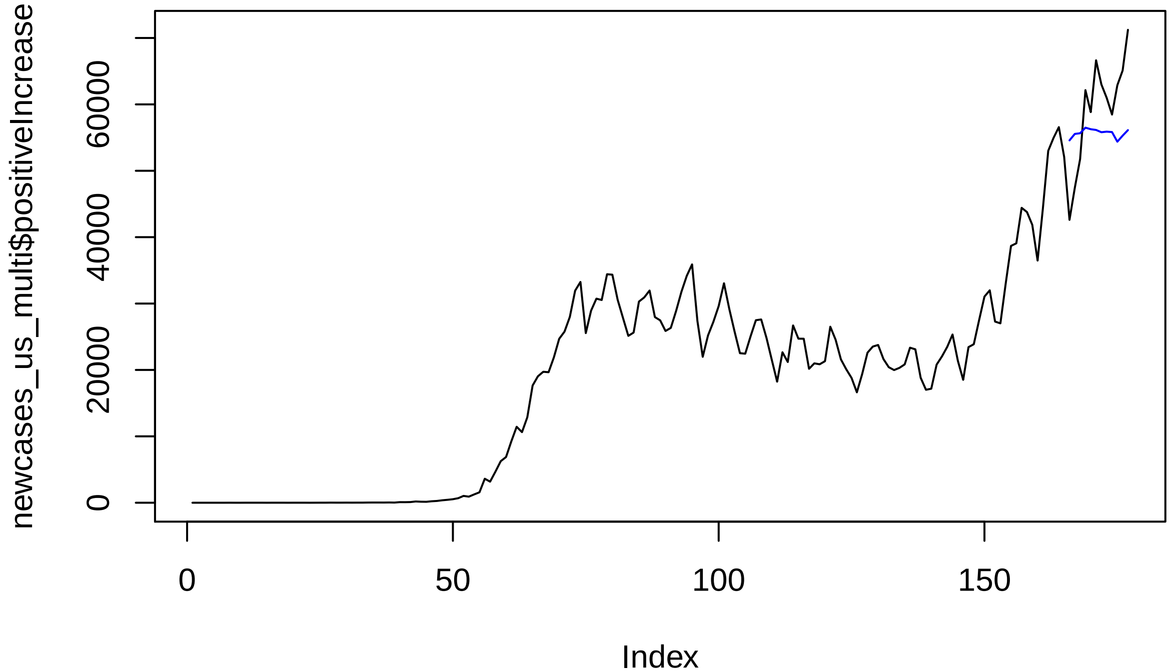



```
plot(newcases_us$positiveIncrease, type = "l", xlim = c(1,277), main = "United States Long Term MLP Mul
lines(long_us_mlp_m$mean, col = "blue")
```



####US Ensemble

```
ensemble_us_fore = (forecasts_us_mlp$mean + FinalPredictions_us_MLR)/2
plot(newcases_us_multi$positiveIncrease, type = "l")
lines(ensemble_us_fore, col = "blue")
```

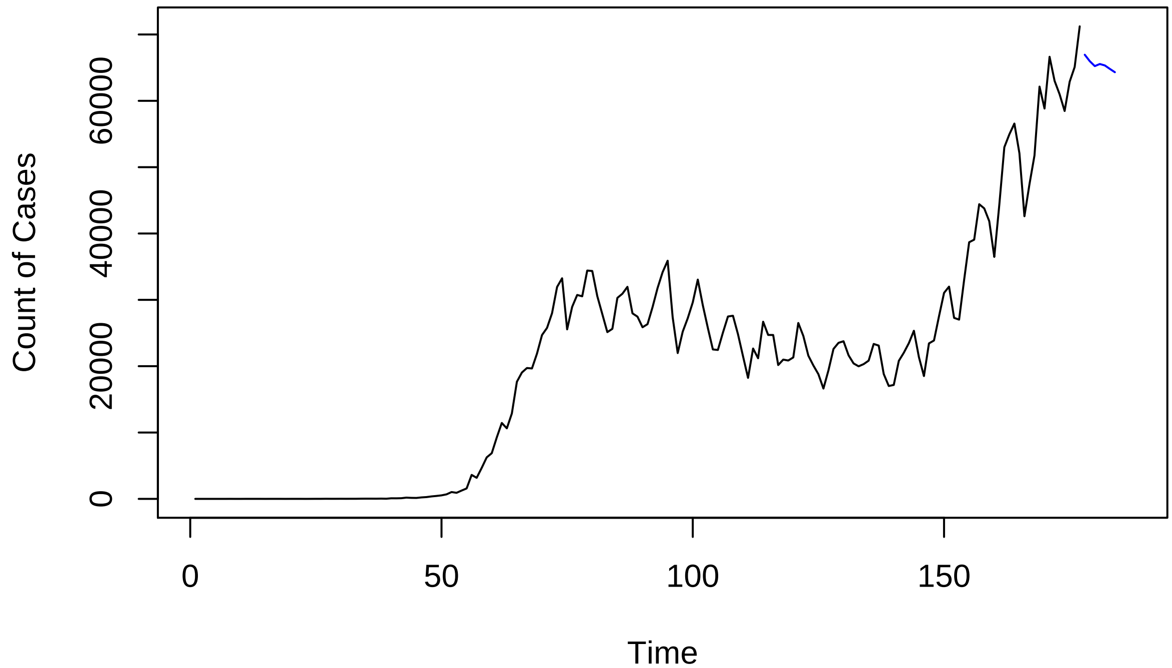



```
#Final 12 ASE

ASE_final12 = mean((newcases_us_multi$positiveIncrease[166:177] -ensemble_us_fore)^2)
ASE_final12
```

```
## [1] 71221939
```

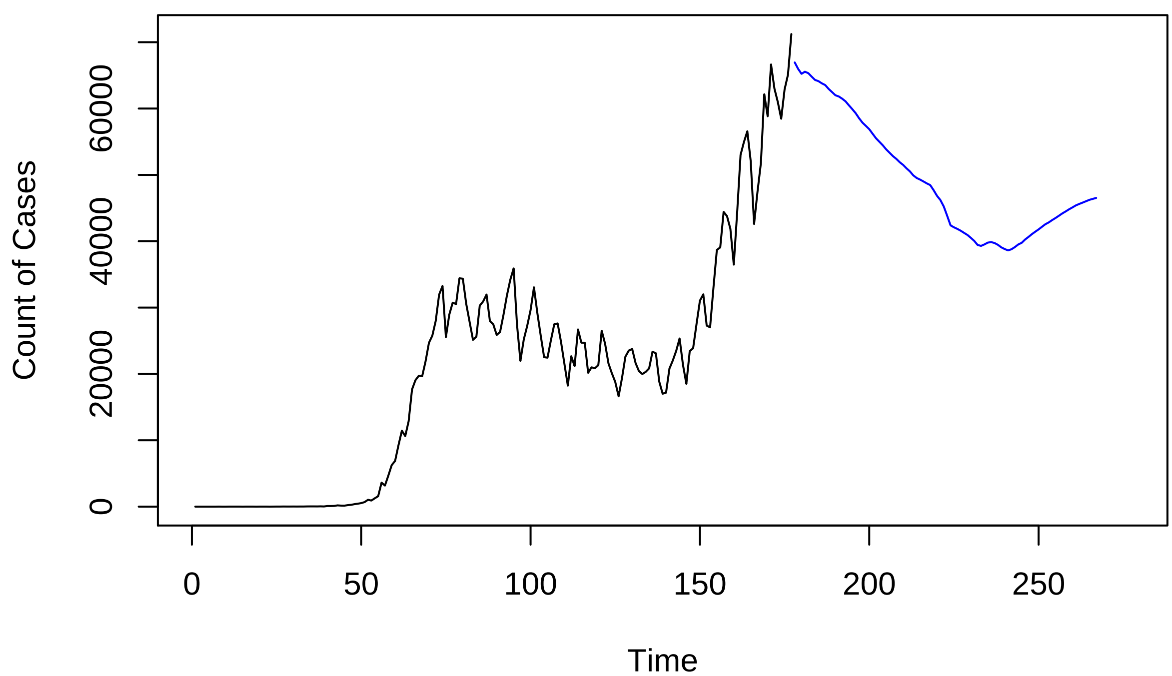
```
#70596024

#Forecasting
#short_us_mlr_m
#short_us_mlp_m

#long_us_mlr_m
#long_us_mlp_m

ensemble_us_fore_short = ( short_us_mlp_m$mean+ short_us_mlr_m)/2
ensemble_us_fore_long = (long_us_mlp_m$mean + long_us_mlr_m)/2

plot(newcases_us$positiveIncrease, type = "l", xlim = c(1,187), main = "United States Short Term Multiva
lines(ensemble_us_fore_short, col = "blue")
```

United States Short Term Multivariate Ensemble Forecasts




```
plot(newcases_us$positiveIncrease, type = "l", xlim = c(1,277), main = "United States Long Term Multiva
lines(ensemble_us_fore_long, col = "blue")
```

United States Long Term Multivariate Ensemble Forecasts



##Conclusion

Given our data at hand we found our ARIMA(6,1,2) model to offer the best predictions for United States overall data. We found our Multiple Linear Regression model with correlated errors best predicted our Florida data. Given our current state of spread and recent trends of Covid19 spread, we expect increasing not optimistic forecasts of new case counts to show as most accurate in the coming days and months.