ADSP31006_FinalProject

Marian Xu

2024-05-15

Prepare Data

crime data

```r
crime_clean <- read.csv("~/Documents/ADSP31006/crime_clean (1).csv")
crime_clean$Date <- as.yearmon(crime_clean$Date, "%Y-%m")
```

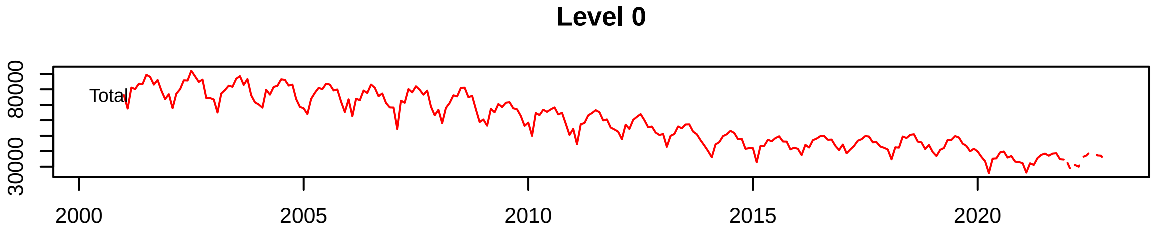
```r
# Convert the entire dataset into a time series object
crime_ts <- ts(crime_clean[, -1], start = c(2001, 1), frequency = 12)

train_ts <- window(crime_ts, start = c(2001, 1), end = c(2021, 12))
test_ts <- window(crime_ts, start = c(2022, 1), end = c(2022, 12))
```

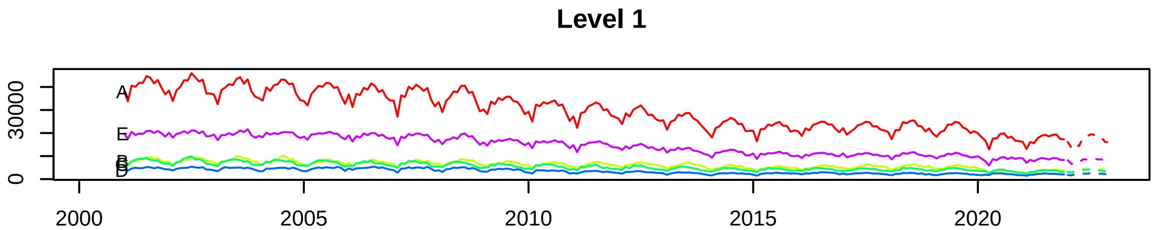
HTS

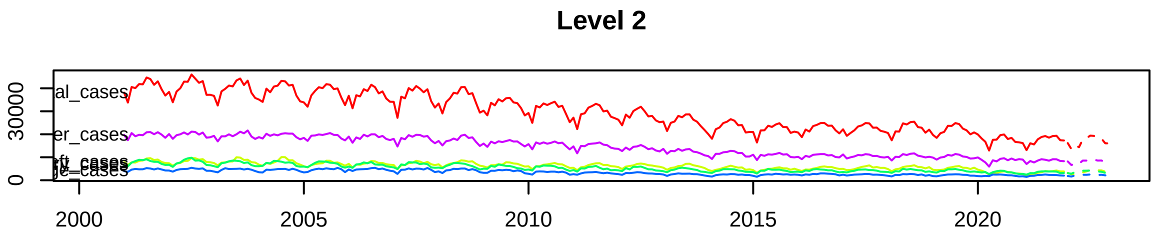
```r
nodes <- list(5, c(1, 1, 1, 1, 1))
hts_data <- hts(train_ts, nodes = nodes)
```

```
## Since argument characters are not specified, the default labelling system is used.
```

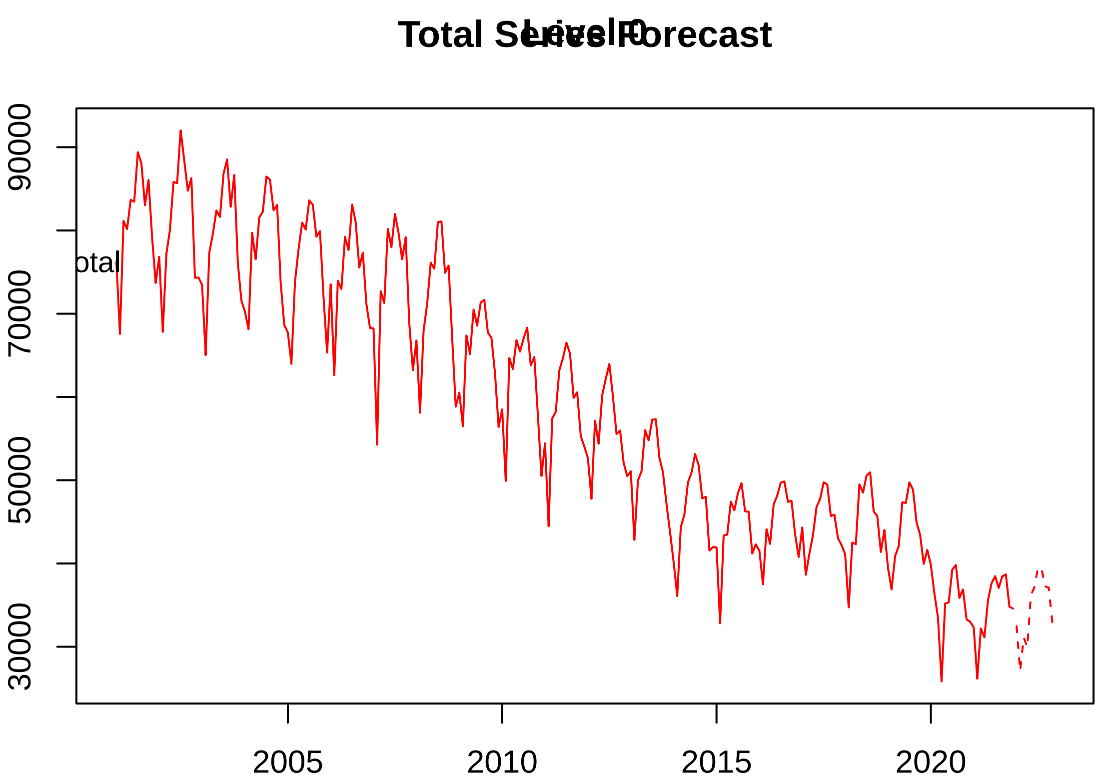
```r
hts_pred <- forecast(hts_data, h=12)
plot(hts_pred)
```



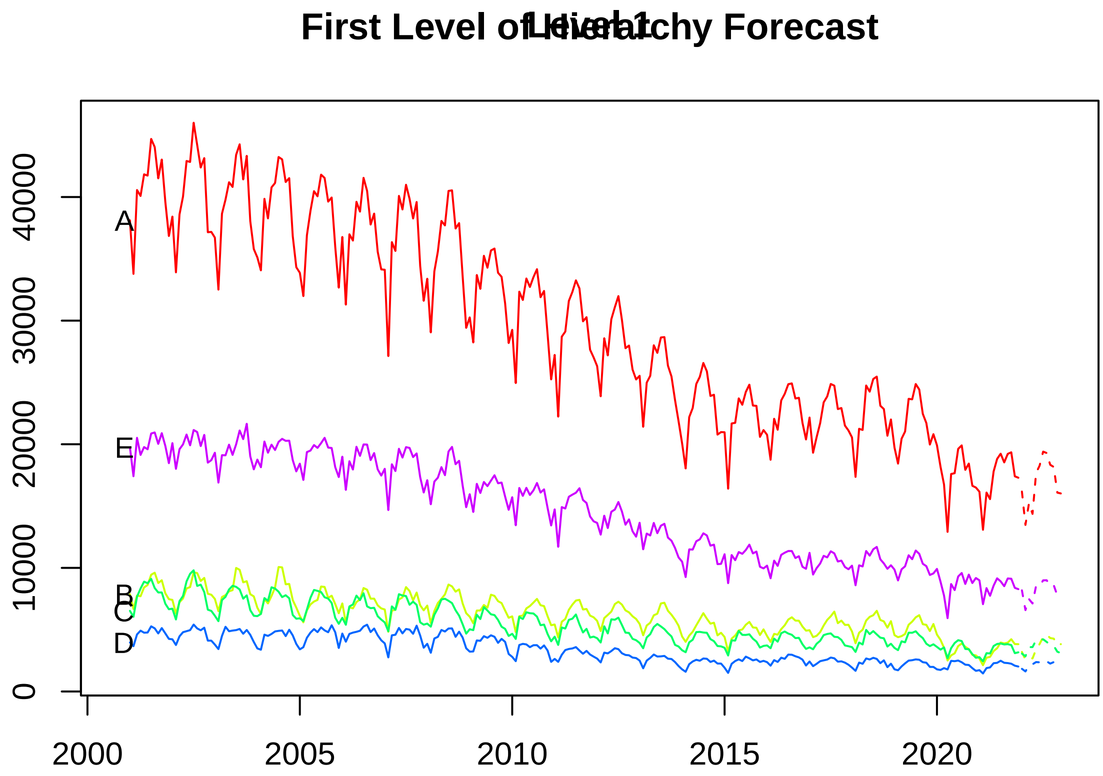




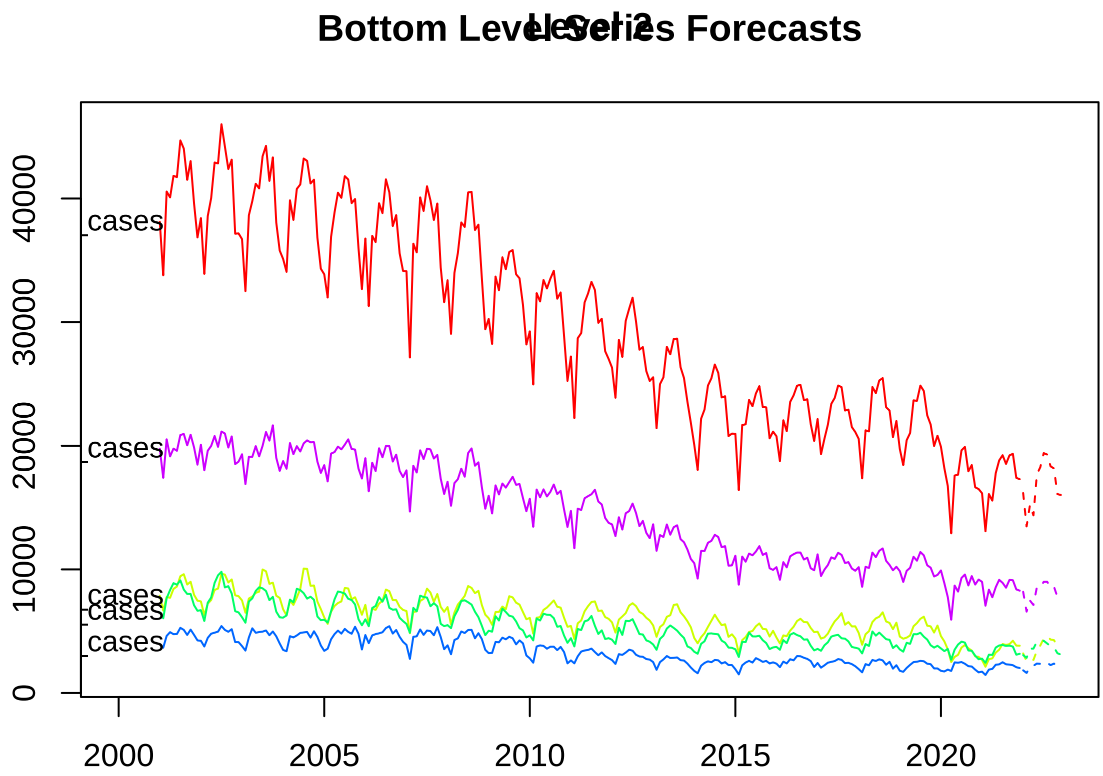

```r
# Plot the total series forecast
plot(hts_pred, levels = 0)
title(main = "Total Series Forecast")
```



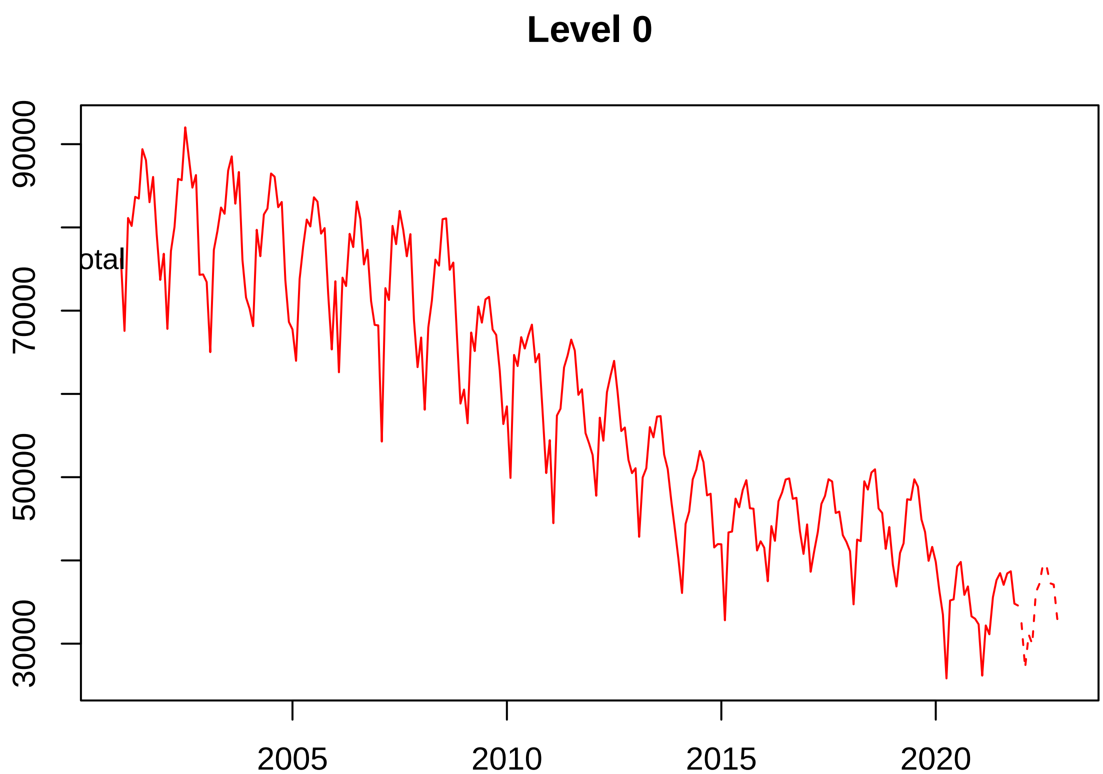
```r
# Plot the first level of the hierarchy forecast
plot(hts_pred, levels = 1)
title(main = "First Level of Hierarchy Forecast")
```



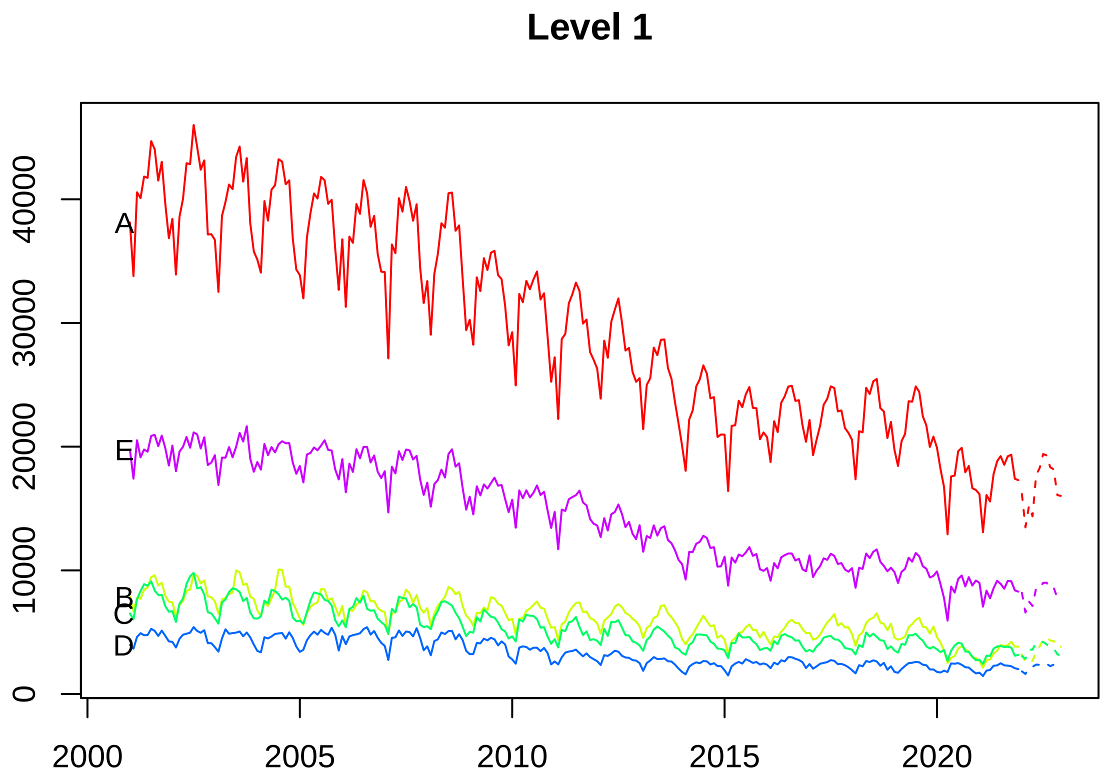

```r
# Plot the bottom level series forecasts
plot(hts_pred, levels = 2)
title(main = "Bottom Level Series Forecasts")
```



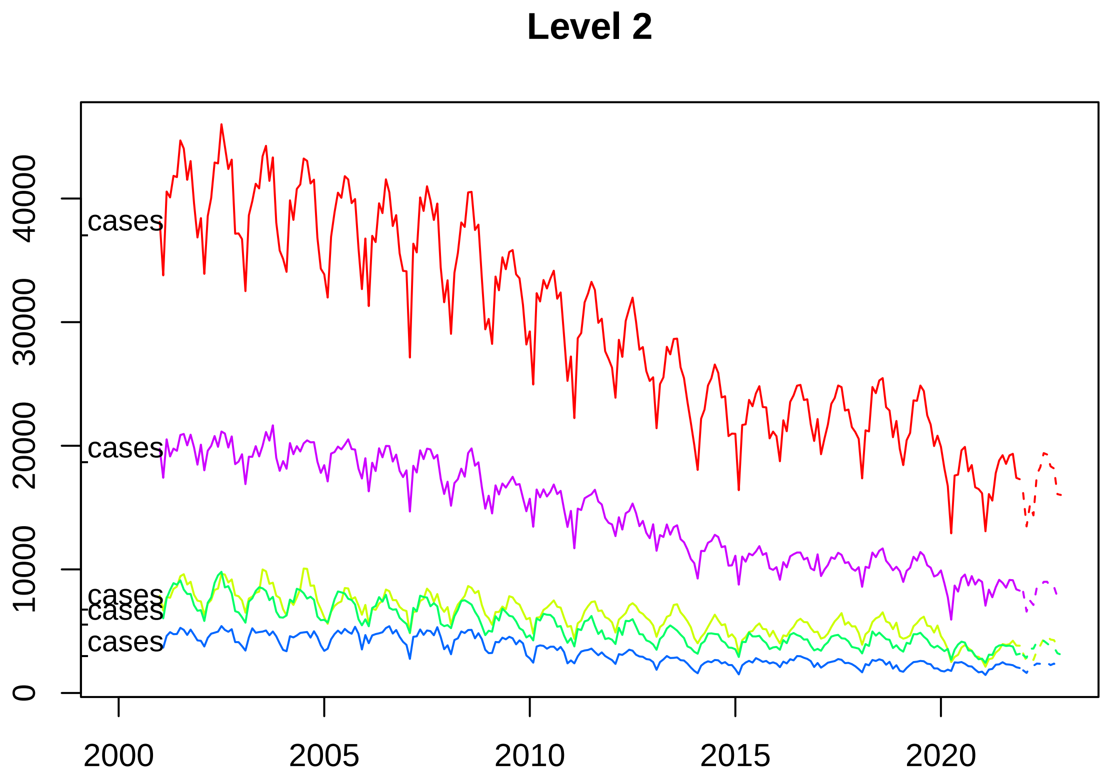

```r
# Plot the total series forecast
plot(hts_pred, levels = 0)
```



```r
# Plot the first level of the hierarchy forecast
plot(hts_pred, levels = 1)
```




```r
# Plot the bottom level series forecasts
plot(hts_pred, levels = 2)
```




```
forecast::accuracy(hts_pred$bts[,1], test_ts[,1])
```

```
##                  ME     RMSE      MAE     MPE     MAPE      ACF1 Theil's U
## Test set 2973.609 3283.665 3074.044 14.5455 15.19047 0.4538166  2.006162
```

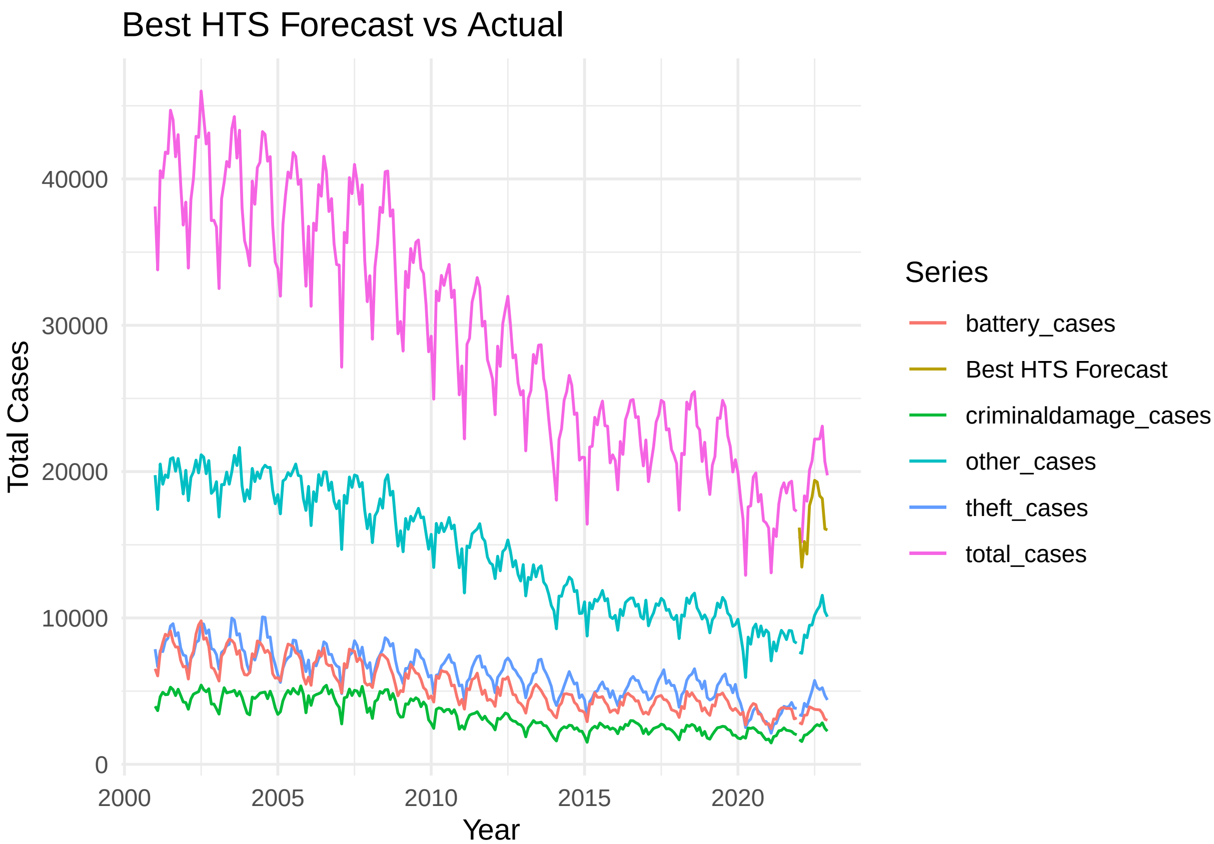
```
smape(hts_pred$bts[,1], test_ts[,1])
```

```
## [1] 0.1658701
```

```
hts_pred_ts <- ts(hts_pred$bts[,1], start = c(2022, 1),end = c(2022, 12), frequency = 12)

# Plot the forecasts and include training data
autoplot(train_ts, series="Training Data") +
  autolayer(test_ts, series="Test Data") +
  autolayer(hts_pred_ts, series="Best HTS Forecast") +
  ggtitle("Best HTS Forecast vs Actual") +
  xlab("Year") + ylab("Total Cases") +
  theme_minimal() +
  guides(colour=guide_legend(title="Series"))
```

```
## Warning in ggplot2::geom_line(na.rm = TRUE, ...): Ignoring unknown parameters:
## `series`
```

```
## For a multivariate time series, specify a seriesname for each time series. Defaulting to column name
```

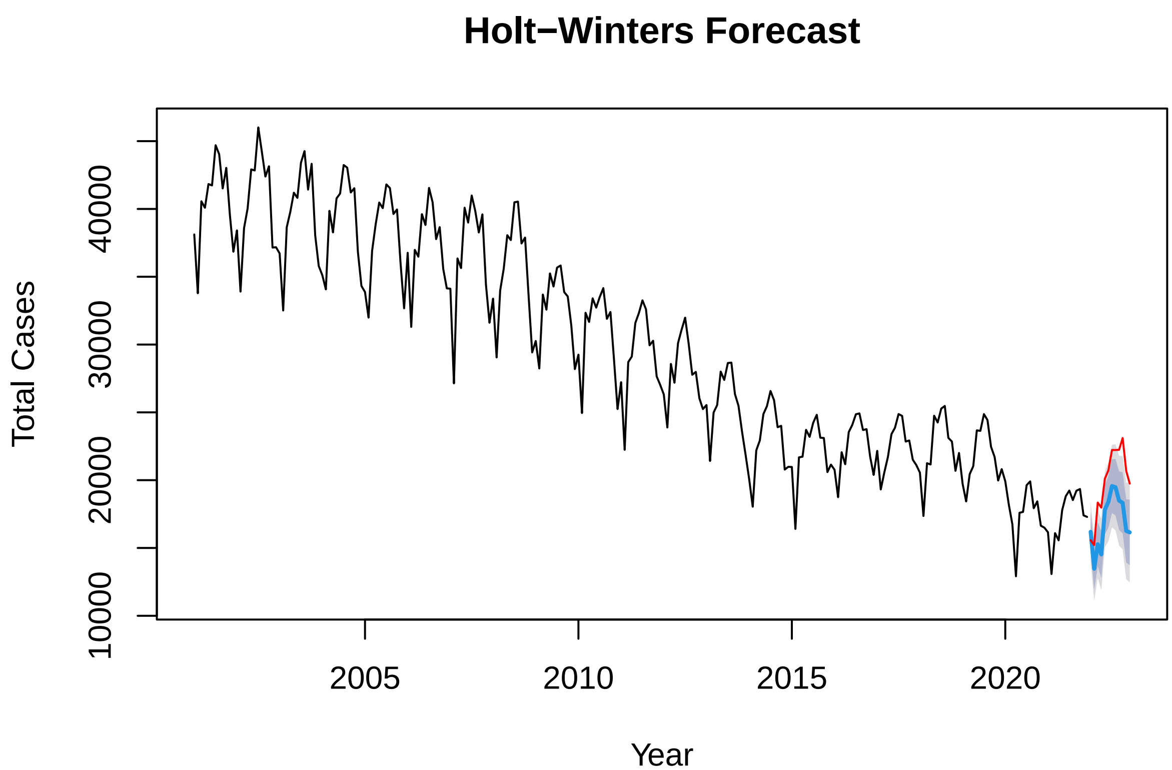



```r
crime_data<- read.csv("~/Documents/ADSP31006/crime_data_type.csv")
```

```r
# Convert Date column to yearmon format
crime_data$Date <- as.yearmon(crime_data$Date, "%Y-%m")

# Convert total_cases to a time series object
crime_ts <- ts(crime_data$total_cases, start = c(2001, 1), frequency = 12)

# Split into training and test datasets
train_ts <- window(crime_ts, start = c(2001, 1), end = c(2021, 12))
test_ts <- window(crime_ts, start = c(2022, 1), end = c(2022, 12))

# Perform Holt-Winters exponential smoothing
hw_model <- HoltWinters(train_ts)
```

```
## Warning in HoltWinters(train_ts): optimization difficulties: ERROR:
## ABNORMAL_TERMINATION_IN_LNSRCH
```

```r
# Forecast the next 12 months
hw_forecast <- forecast(hw_model, h=12)

# Plot the forecast
plot(hw_forecast, main="Holt-Winters Forecast", xlab="Year", ylab="Total Cases")
lines(test_ts, col='red')
```




```r
# Print the forecasted values
print(hw_forecast)
```

```
##          Point Forecast     Lo 80    Hi 80    Lo 95    Hi 95
## Jan 2022       16192.04 14755.08 17629.00 13994.40 18389.68
## Feb 2022       13465.92 11930.39 15001.44 11117.54 15814.29
## Mar 2022       15278.32 13647.32 16909.31 12783.93 17772.71
## Apr 2022       14519.30 12795.37 16243.23 11882.78 17155.82
## May 2022       17821.89 16007.13 19636.64 15046.46 20597.32
## Jun 2022       18404.76 16500.94 20308.57 15493.12 21316.39
## Jul 2022       19568.40 17577.02 21559.78 16522.84 22613.96
## Aug 2022       19477.42 17399.74 21555.10 16299.89 22654.96
## Sep 2022       18490.22 16327.33 20653.12 15182.36 21798.09
## Oct 2022       18327.52 16080.34 20574.70 14890.75 21764.28
## Nov 2022       16256.83 13926.16 18587.50 12692.37 19821.28
## Dec 2022       16149.67 13736.19 18563.15 12458.57 19840.77
```

```r
# Fit the Holt-Winters model
hw_model <- hw(train_ts, seasonal = "multiplicative")
summary(hw_model)
```

```
##
## Forecast method: Holt-Winters' multiplicative method
##
## Model Information:
## Holt-Winters' multiplicative method
##
## Call:
```

```
##  hw(y = train_ts, seasonal = "multiplicative")
##
##   Smoothing parameters:
##     alpha = 0.3634
##     beta  = 0.0041
##     gamma = 0.1808
##
##   Initial states:
##     l = 41096.8363
##     b = 3.9761
##     s = 0.898 0.9512 1.0646 1.0389 1.0979 1.1121
##            1.0477 1.0567 0.9951 0.986 0.8291 0.9226
##
##   sigma:  0.044
##
##      AIC     AICc      BIC
## 5006.622 5009.238 5066.623
##
## Error measures:
##                     ME    RMSE      MAE        MPE     MAPE      MASE      ACF1
## Training set -76.15209 1038.28 737.9151 -0.3798232 2.857123 0.4754531 0.1782807
##
## Forecasts:
##          Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## Jan 2022       16428.56 15503.13 17354.00 15013.24 17843.89
## Feb 2022       14231.21 13376.56 15085.86 12924.13 15538.29
## Mar 2022       16361.77 15319.99 17403.55 14768.51 17955.03
## Apr 2022       16049.71 14971.16 17128.26 14400.21 17699.21
## May 2022       18648.17 17330.52 19965.81 16633.00 20663.33
## Jun 2022       18880.15 17481.94 20278.35 16741.78 21018.52
## Jul 2022       19733.86 18206.32 21261.40 17397.69 22070.03
## Aug 2022       19460.55 17889.74 21031.35 17058.21 21862.89
## Sep 2022       18257.44 16723.90 19790.98 15912.10 20602.79
## Oct 2022       18120.81 16539.82 19701.79 15702.90 20538.71
## Nov 2022       16344.81 14865.98 17823.65 14083.13 18606.49
## Dec 2022       16137.38 14625.45 17649.32 13825.08 18449.69
## Jan 2023       15528.37 13975.84 17080.90 13153.98 17902.76
## Feb 2023       13447.83 12061.63 14834.02 11327.83 15567.82
## Mar 2023       15456.93 13815.73 17098.13 12946.93 17966.93
## Apr 2023       15157.99 13501.51 16814.48 12624.62 17691.37
## May 2023       17607.24 15628.39 19586.09 14580.85 20633.63
## Jun 2023       17821.32 15762.91 19879.73 14673.26 20969.39
## Jul 2023       18621.94 16412.90 20830.97 15243.50 22000.37
## Aug 2023       18358.82 16123.47 20594.17 14940.15 21777.49
## Sep 2023       17218.90 15068.19 19369.62 13929.67 20508.14
## Oct 2023       17085.11 14897.13 19273.08 13738.89 20431.33
## Nov 2023       15406.13 13384.27 17427.98 12313.96 18498.29
## Dec 2023       15206.13 13162.01 17250.26 12079.91 18332.35
```

```r
# Forecast using the Holt-Winters model
hw_forecast <- forecast(hw_model, h = length(test_ts))
print(hw_forecast)
```

```
##          Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
```

```
## Jan 2022       16428.56 15503.13 17354.00 15013.24 17843.89
## Feb 2022       14231.21 13376.56 15085.86 12924.13 15538.29
## Mar 2022       16361.77 15319.99 17403.55 14768.51 17955.03
## Apr 2022       16049.71 14971.16 17128.26 14400.21 17699.21
## May 2022       18648.17 17330.52 19965.81 16633.00 20663.33
## Jun 2022       18880.15 17481.94 20278.35 16741.78 21018.52
## Jul 2022       19733.86 18206.32 21261.40 17397.69 22070.03
## Aug 2022       19460.55 17889.74 21031.35 17058.21 21862.89
## Sep 2022       18257.44 16723.90 19790.98 15912.10 20602.79
## Oct 2022       18120.81 16539.82 19701.79 15702.90 20538.71
## Nov 2022       16344.81 14865.98 17823.65 14083.13 18606.49
## Dec 2022       16137.38 14625.45 17649.32 13825.08 18449.69
```

```r
# Extract the forecasted mean values
forecasted_values <- hw_forecast$mean
print(forecasted_values)
```

```
##           Jan      Feb      Mar      Apr      May      Jun      Jul      Aug
## 2022 16428.56 14231.21 16361.77 16049.71 18648.17 18880.15 19733.86 19460.55
##           Sep      Oct      Nov      Dec
## 2022 18257.44 18120.81 16344.81 16137.38
```