ADSP31006_FinalProject

Marian Xu

2024-05-15

Prepare Data

crime data

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

```
# Convert the entire dataset into a time series object
crime_ts <- ts(crime_clean[, 3:35], 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))
```

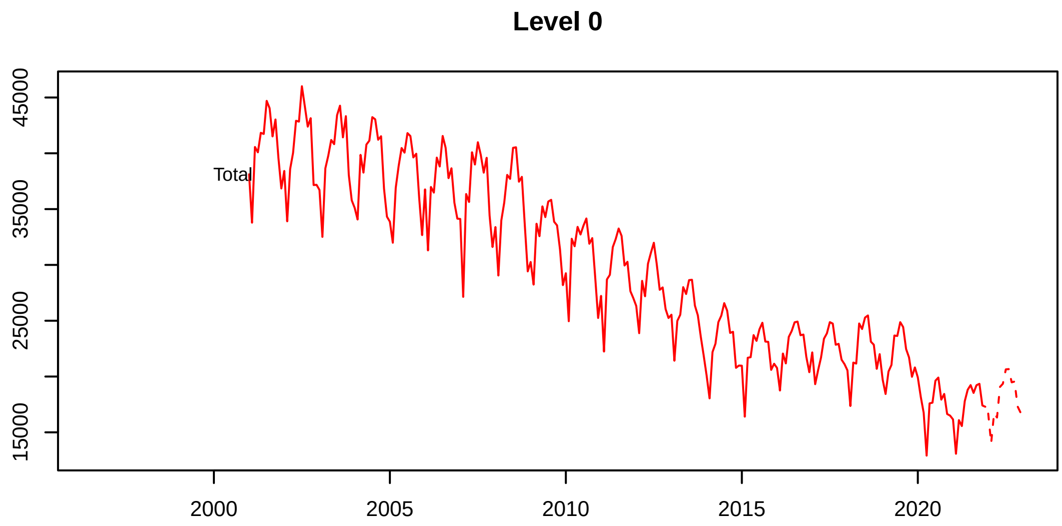
HTS

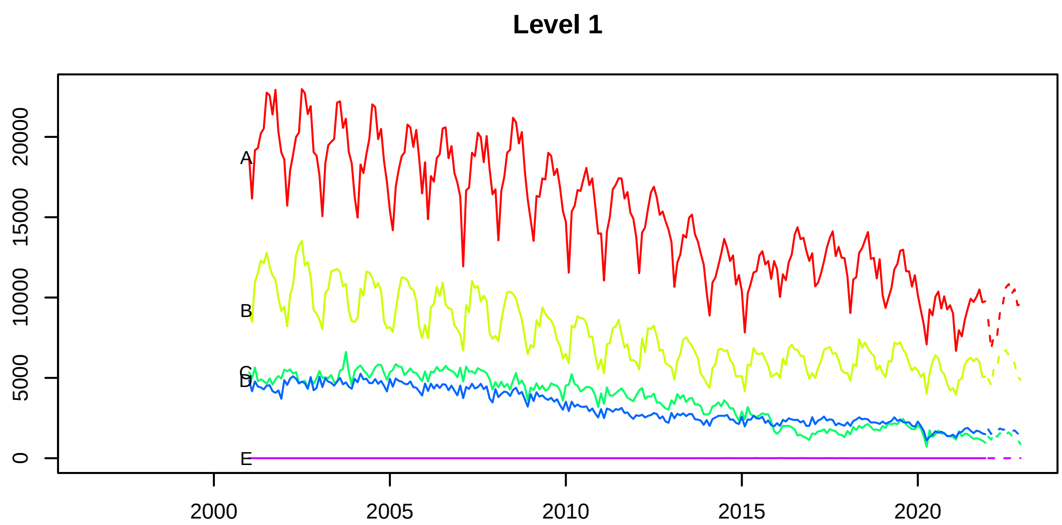
```
nodes <- list(5, c(7, 9, 7, 9, 1))
hts_data <- hts(train_ts, nodes = nodes)
```

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

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

```
plot(hts_pred)
```


Level 0
Total
45000
35000
25000
15000
2000
2005
2010
2015
2020


Level 1
A
B
C
D
E
20000
15000
10000
5000
0
2000
2005
2010
2015
2020


Level 2
theft_cases
battery_cases
narcotics_cases
riminal_damage_cases
other_offense_cases
burglary_cases
prostitution_cases
10000
8000
6000
4000
2000
0
2000
2005
2010
2015
2020

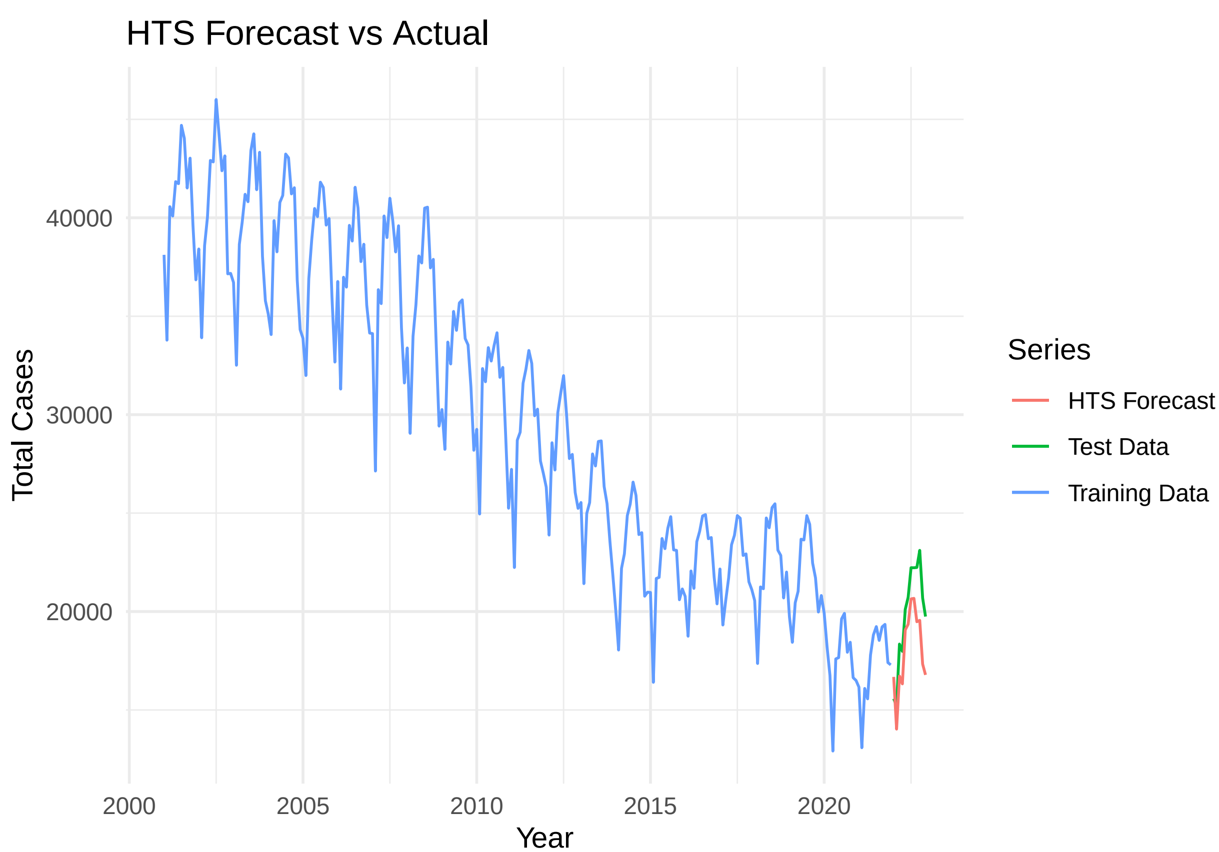
```r
hts_pred_ts <- ts(rowSums(hts_pred$bts, na.rm = TRUE), start = start(test_ts), frequency = 12)
total_train_ts <- ts(rowSums(train_ts, na.rm = TRUE), start = start(train_ts), frequency = 12)
total_test_ts <- ts(rowSums(test_ts, na.rm = TRUE), start = start(test_ts), frequency = 12)

forecast::accuracy(hts_pred_ts, total_test_ts)
```

```
##                 ME    RMSE      MAE      MPE     MAPE      ACF1 Theil's U
## Test set 1794.54 2163.34 1979.106 8.636766 9.822011 0.4850786  1.271955
```

```r
smape(hts_pred_ts, total_test_ts)
```

```
## [1] 0.103682
```

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

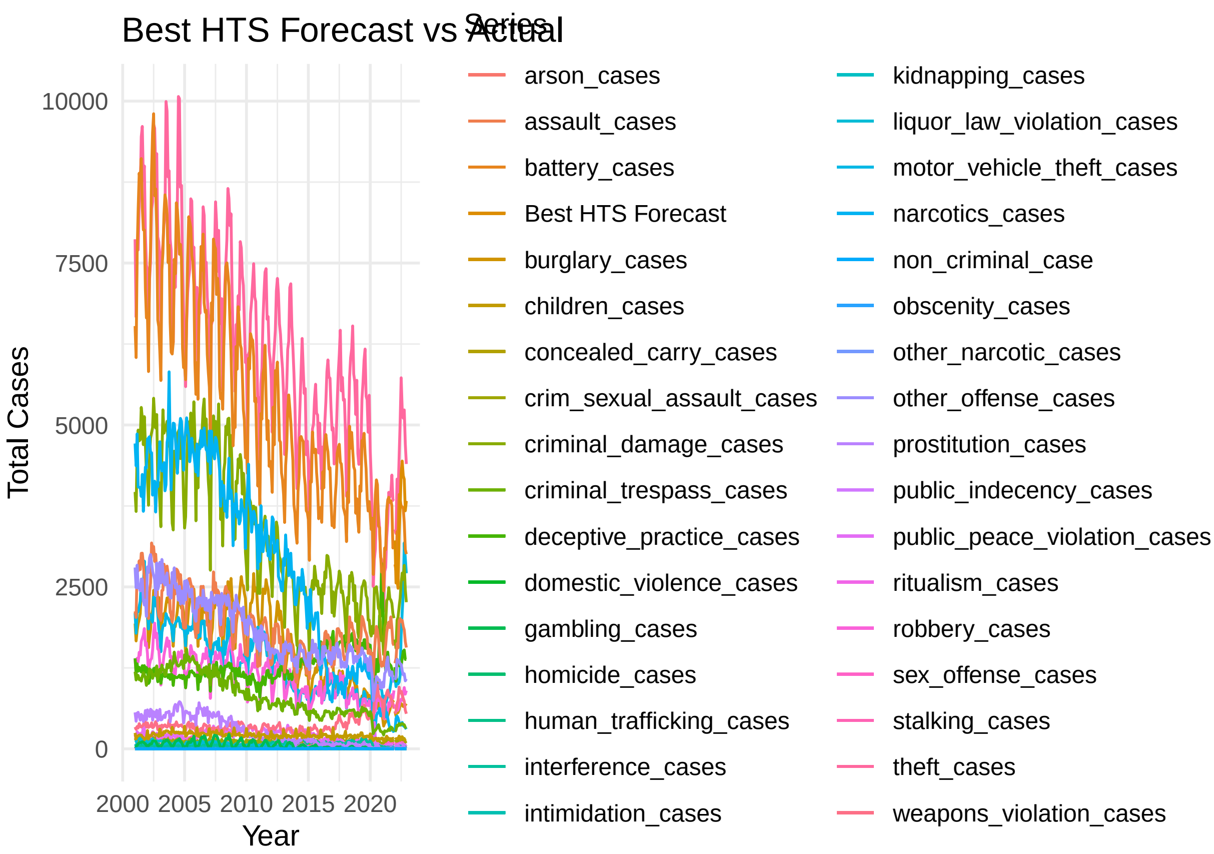



```r
hts_pred_ts <- ts(hts_pred$bts[,1], start = c(2022, 1),end = c(2022, 12), frequency = 12)
```

```r
# 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_clean$Date <- as.yearmon(crime_clean$Date, "%Y-%m")
```

```r
# Convert the entire dataset into a time series object
crime_ts <- ts(crime_clean[, 2], 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))
```

Holt Winters

```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 = 0.3615
##     beta  = 0.0042
##     gamma = 0.1903
##
##   Initial states:
##     l = 41096.9516
##     b = 3.8995
##     s = 0.8983 0.952 1.0653 1.0395 1.0992 1.1126
##            1.046 1.0537 0.995 0.9861 0.8302 0.9222
##
##   sigma:  0.044
##
##      AIC     AICc      BIC
## 5006.629 5009.245 5066.630
##
## Error measures:
##                     ME    RMSE      MAE        MPE     MAPE      MASE      ACF1
## Training set -76.05566 1038.39 740.0221 -0.3783071 2.863598 0.4768056 0.1777524
##
## Forecasts:
##          Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## Jan 2022       16424.26 15499.04 17349.48 15009.26 17839.26
## Feb 2022       14220.61 13367.10 15074.12 12915.28 15525.94
## Mar 2022       16315.35 15277.65 17353.05 14728.32 17902.37
## Apr 2022       15975.96 14903.93 17047.98 14336.44 17615.48
## May 2022       18605.19 17292.91 19917.47 16598.23 20612.15
## Jun 2022       18857.25 17463.56 20250.94 16725.79 20988.71
## Jul 2022       19718.44 18195.48 21241.40 17389.28 22047.60
## Aug 2022       19448.14 17882.08 21014.20 17053.05 21843.22
## Sep 2022       18261.86 16731.81 19791.92 15921.85 20601.88
## Oct 2022       18116.29 16539.84 19692.73 15705.32 20527.25
## Nov 2022       16333.94 14860.09 17807.79 14079.88 18588.00
## Dec 2022       16145.87 14637.33 17654.41 13838.76 18452.98
## Jan 2023       15520.53 13970.12 17070.94 13149.39 17891.67
## Feb 2023       13434.51 12051.09 14817.93 11318.75 15550.26
## Mar 2023       15409.26 13774.99 17043.52 12909.86 17908.65
```

```
## Apr 2023       15084.57 13438.17 16730.96 12566.63 17602.51
## May 2023       17562.22 15591.10 19533.35 14547.64 20576.81
## Jun 2023       17795.16 15742.72 19847.61 14656.22 20934.11
## Jul 2023       18602.59 16399.15 20806.02 15232.73 21972.44
## Aug 2023       18342.34 16112.47 20572.20 14932.06 21752.62
## Sep 2023       17218.54 15071.33 19365.76 13934.66 20502.43
## Oct 2023       17076.31 14893.05 19259.56 13737.31 20415.30
## Nov 2023       15391.75 13375.16 17408.34 12307.64 18475.85
## Dec 2023       15210.00 13168.83 17251.18 12088.29 18331.71
```

```
# Forecast using the Holt-Winters model
hw_forecast <- forecast(hw_model, h = 12)
print(hw_forecast)
```

```
##          Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## Jan 2022       16424.26 15499.04 17349.48 15009.26 17839.26
## Feb 2022       14220.61 13367.10 15074.12 12915.28 15525.94
## Mar 2022       16315.35 15277.65 17353.05 14728.32 17902.37
## Apr 2022       15975.96 14903.93 17047.98 14336.44 17615.48
## May 2022       18605.19 17292.91 19917.47 16598.23 20612.15
## Jun 2022       18857.25 17463.56 20250.94 16725.79 20988.71
## Jul 2022       19718.44 18195.48 21241.40 17389.28 22047.60
## Aug 2022       19448.14 17882.08 21014.20 17053.05 21843.22
## Sep 2022       18261.86 16731.81 19791.92 15921.85 20601.88
## Oct 2022       18116.29 16539.84 19692.73 15705.32 20527.25
## Nov 2022       16333.94 14860.09 17807.79 14079.88 18588.00
## Dec 2022       16145.87 14637.33 17654.41 13838.76 18452.98
```

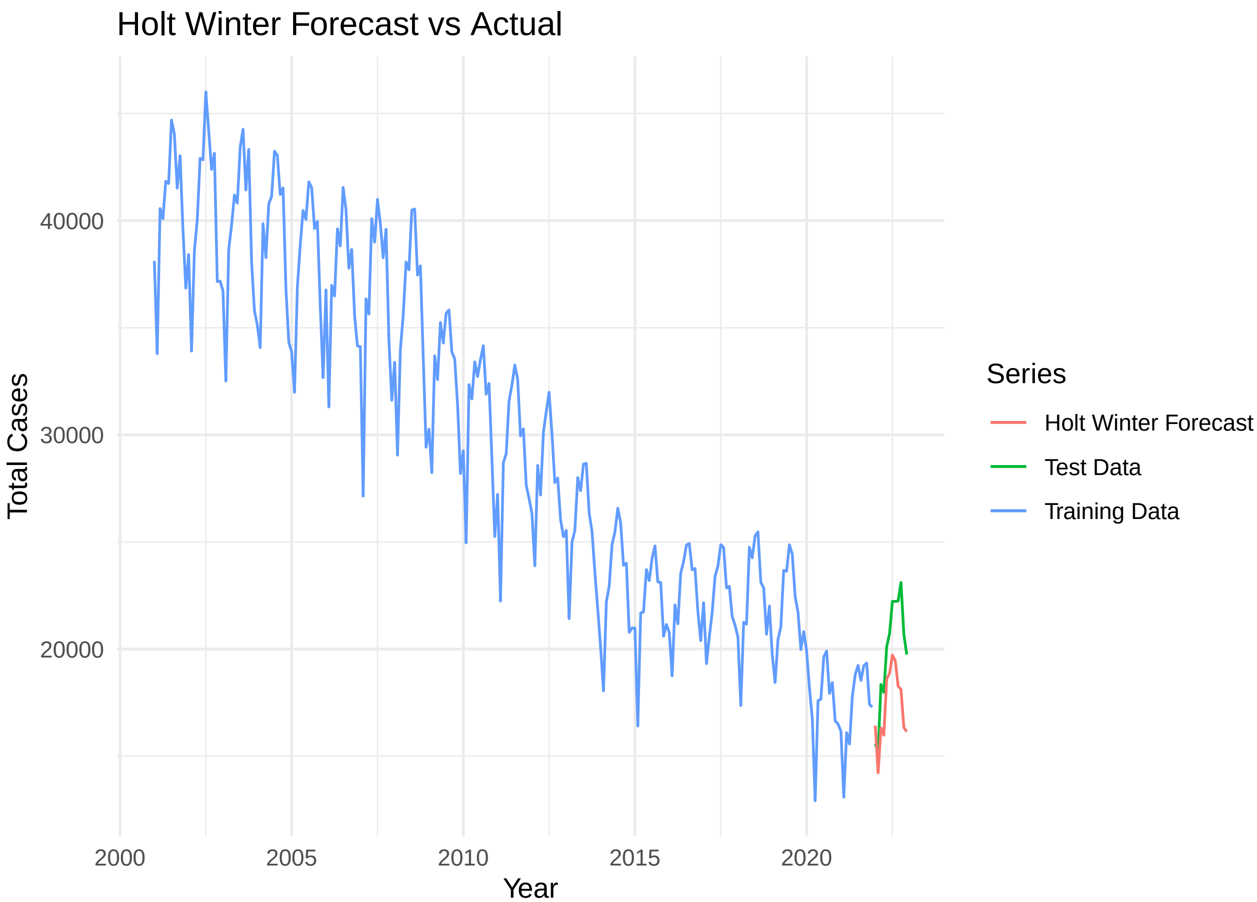
```
forecast::accuracy(hw_forecast$mean, test_ts)
```

```
##                ME     RMSE      MAE      MPE     MAPE     ACF1 Theil's U
## Test set 2478.654 2918.202 2620.697 11.85441 12.76659 0.628668  1.706287
```

```
smape(hw_forecast$mean, test_ts)
```

```
## [1] 0.1378925
```

```
hw_pred_ts <- ts(hw_forecast$mean, 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(hw_pred_ts, series="Holt Winter Forecast") +
  ggtitle("Holt Winter Forecast vs Actual") +
  xlab("Year") + ylab("Total Cases") +
  theme_minimal() +
  guides(colour=guide_legend(title="Series"))
```


Holt Winter Forecast vs Actual
40000
30000
20000
Total Cases
2000
2005
2010
2015
2020
Year
Series
Holt Winter Forecast
Test Data
Training Data