**1. Introduction**

This report presents an analysis of baggage complaints for three airlines: United, American Eagle, and Hawaiian. It highlights insights gained through data visualization, evaluation of model performance, and various forecasting techniques, aimed at informing strategies for operational improvements.

**What is the objective?**  
To forecast monthly baggage complaints for United Airlines, American Eagle, and Hawaiian Airlines.

**Why is it important?**  
Baggage complaints significantly impact customer satisfaction and an airline’s reputation. Understanding and predicting complaint trends can help airlines optimize baggage handling processes, especially during peak travel periods.

**2. Data Description**

**Dataset:**

* Monthly baggage complaints dataset from January 2004 to December 2010.
* **Key Variables**:
  + **Airline**: Name of the airline.
  + **Baggage**: Number of complaints (dependent variable).
  + **Scheduled**: Total number of scheduled flights (independent variable).
  + **Cancelled**: Total number of canceled flights (independent variable).
  + **Enplaned**: Total number of passengers enplaned (independent variable).

**Airlines:**

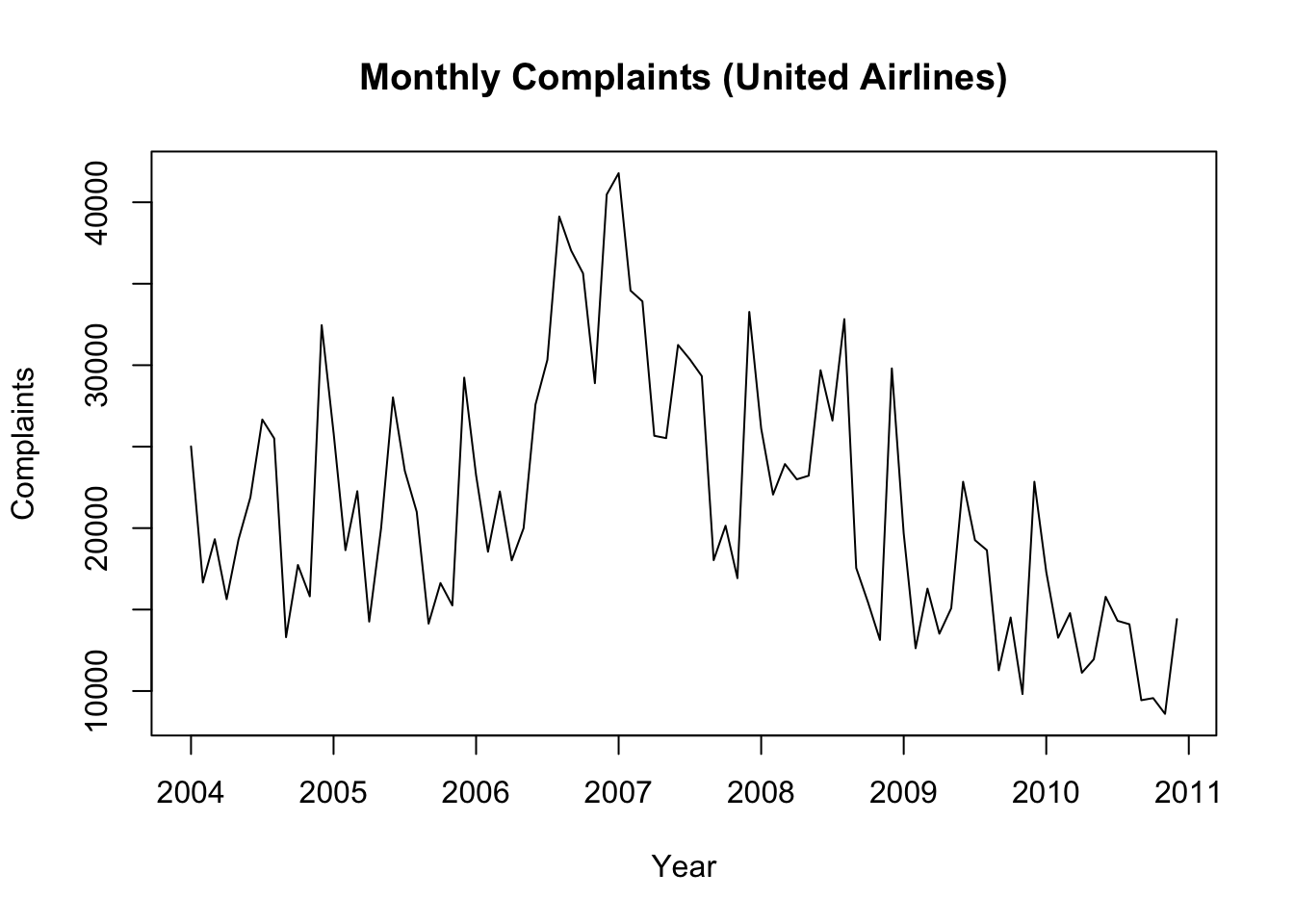
* **United Airlines**: High volume of complaints, fluctuating significantly.
* **American Eagle**: Moderate complaints with visible seasonality.
* **Hawaiian Airlines**: Low, stable complaints over time.

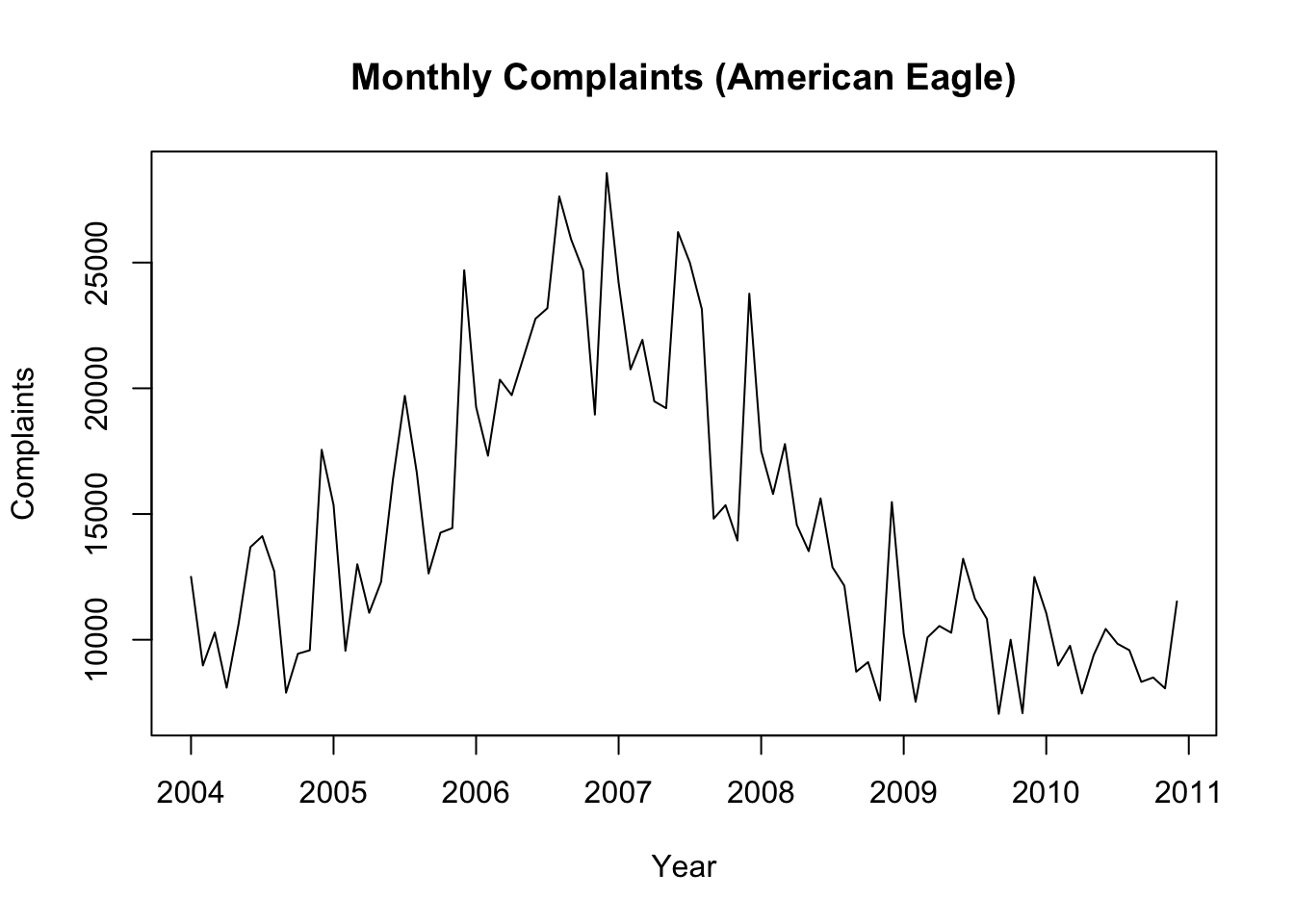
**3. Insights from Exploratory Data Analysis (EDA)**

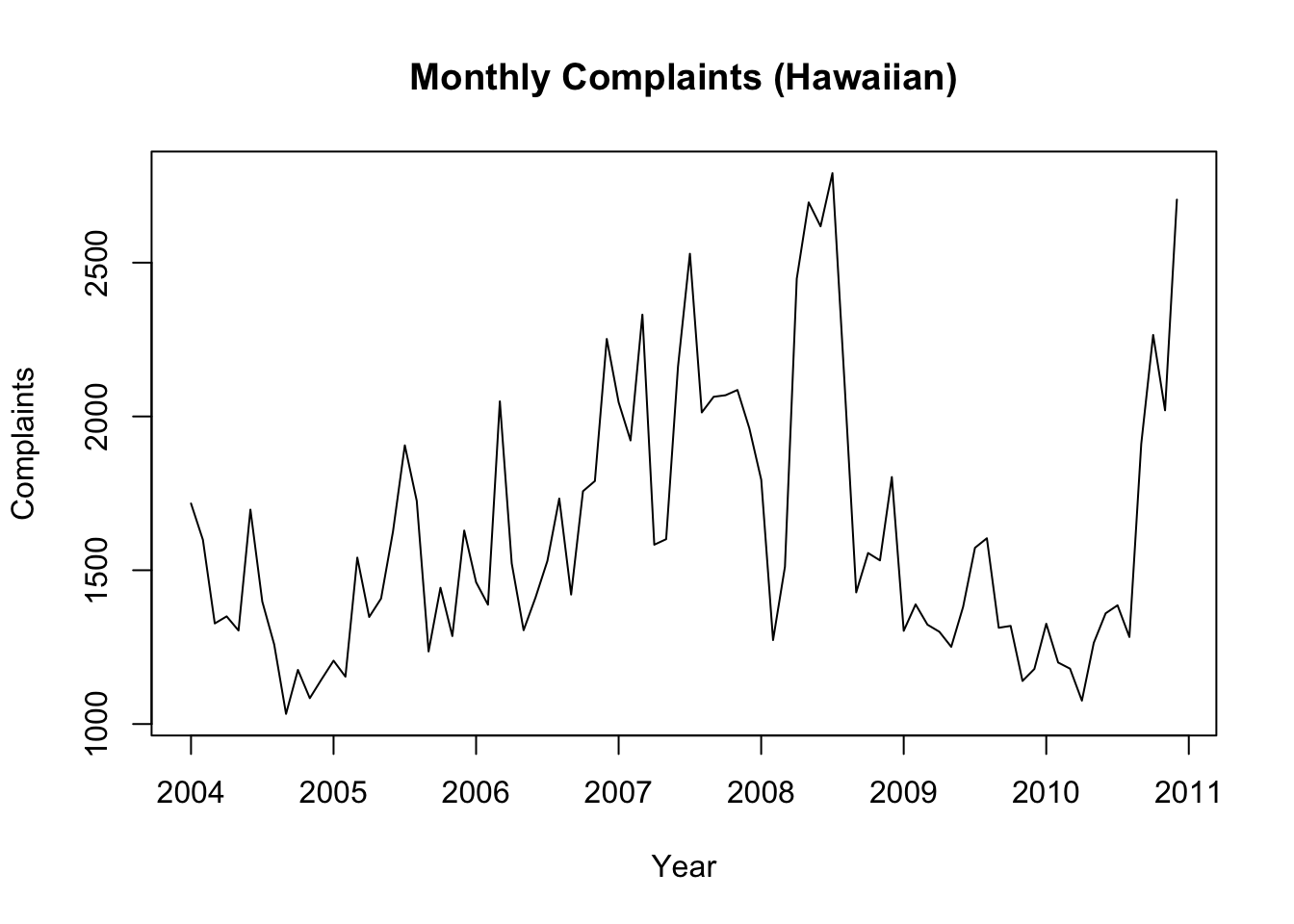
* **Purpose:** To identify trends, seasonality, and fluctuations over time for each airline.
* **Techniques Used**:
  + Time series visualization.
  + Faceted and aggregated plots
  + Histogram
  + Boxplots and ACF.
  + Decomposition.

**Time Series Plot**

* Purpose: To compare complaint trends across airlines







baggageUnited <- baggageUnited %>% mutate(Airline = "United")

baggageAmerican <- baggageAmerican %>% mutate(Airline = "American Eagle")

baggageHawaiian <- baggageHawaiian %>% mutate(Airline = "Hawaiian")

combined\_baggage <- bind\_rows(baggageUnited, baggageAmerican, baggageHawaiian)

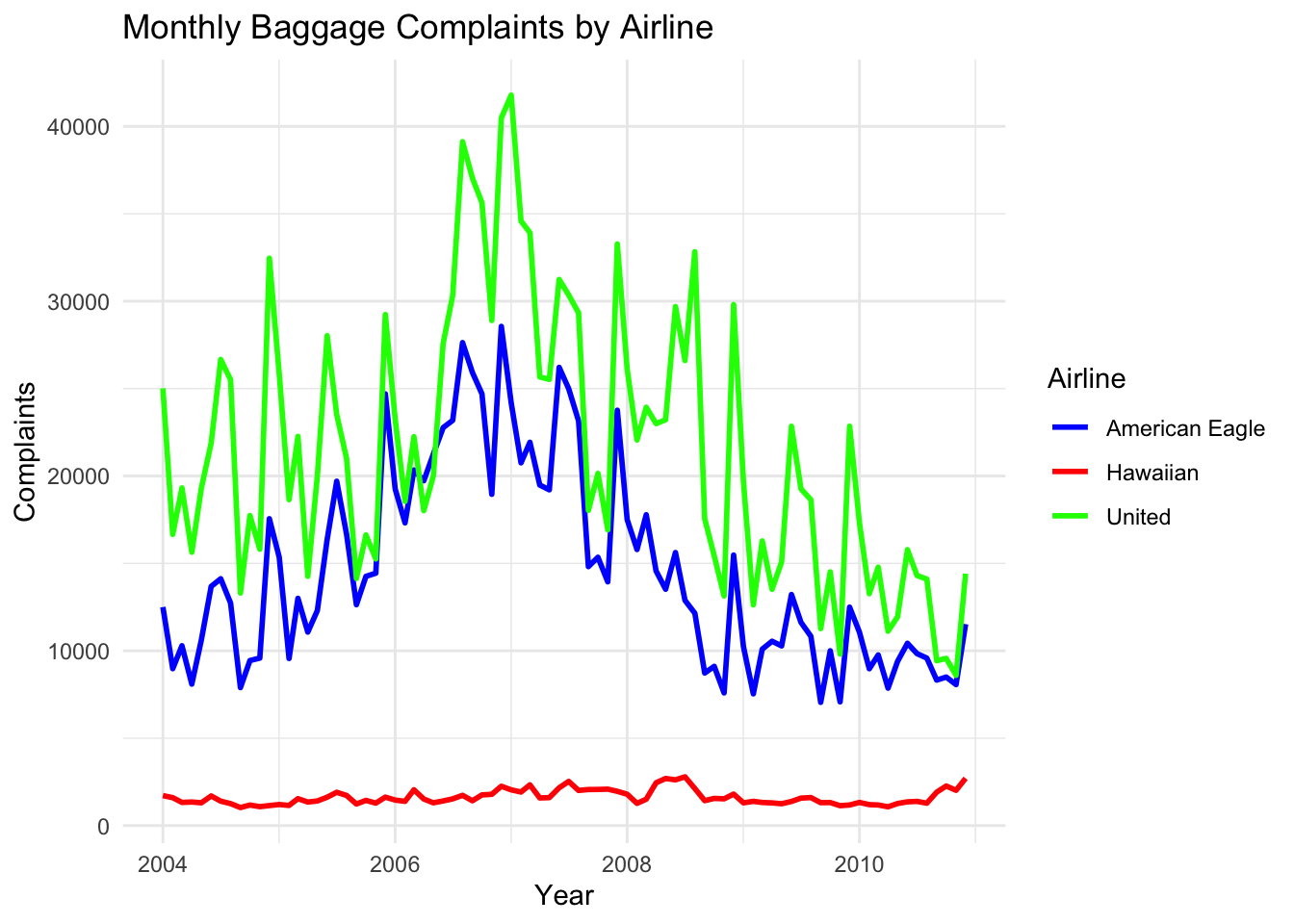
ggplot(combined\_baggage, aes(x = as.Date(paste(Year, Month, "1", sep = "-")), y = Baggage, color = Airline)) +

geom\_line(linewidth = 1) +

labs(title = "Monthly Baggage Complaints by Airline", x = "Year", y = "Complaints") +

theme\_minimal() +

scale\_color\_manual(values = c("blue", "red", "green"))



**Insights**: United Airlines has the highest complaints, followed by American Eagle, and Hawaiian Airlines has consistently low complaints.

ggplot(combined\_baggage, aes(x = as.Date(paste(Year, Month, "1", sep = "-")), y = Baggage)) +

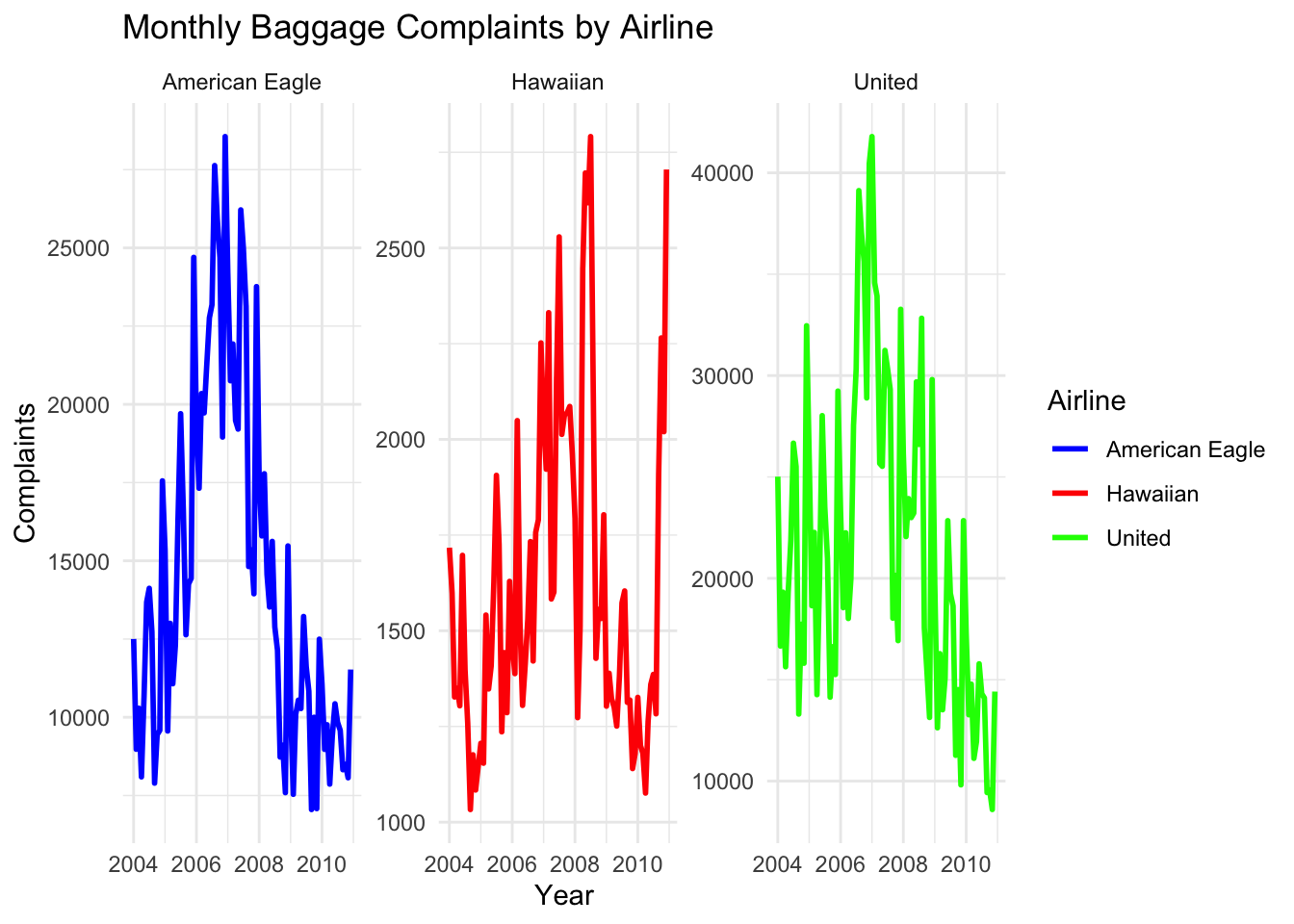
geom\_line(aes(color = Airline), linewidth = 1) +

facet\_wrap(~Airline, scales = "free\_y") + # Facet by airline

labs(title = "Monthly Baggage Complaints by Airline", x = "Year", y = "Complaints") +

theme\_minimal() +

scale\_color\_manual(values = c("blue", "red", "green"))



* United Airlines: High seasonal peaks with a declining trend post-2007.
* American Eagle: Moderate variability, especially during busy travel times.
* Hawaiian Airlines: Low and steady complaints.

**Aggregated Data by Year**

* Purpose: To understand yearly complaint totals for each airline.

# Aggregate total baggage complaints by Year and Airline

baggage\_by\_year <- baggagecomplaints %>%

group\_by(Year, Airline) %>%

summarise(Total\_Baggage\_Complaints = sum(Baggage, na.rm = TRUE))

ggplot(baggage\_by\_year, aes(x = factor(Year), y = Total\_Baggage\_Complaints, fill = Airline)) +

geom\_bar(stat = "identity", position = "dodge") + # Dodge separates bars by airline

theme\_minimal() +

labs(

title = "Total Baggage Complaints by Airline and Year",

x = "Year",

y = "Total Baggage Complaints",

fill = "Airline"

) +

scale\_fill\_manual(

values = c("American Eagle" = "#FF9999", "Hawaiian" = "#99CCFF", "United" = "#FFCC99")

) +

theme(

plot.title = element\_text(hjust = 0.5, face = "bold", size = 14),

axis.title = element\_text(size = 12),

axis.text = element\_text(size = 10)

)

A graph of baggage complaints

Description automatically generated

* United Airlines shows a peak around 2006-2007.
* American Eagle exhibits moderate fluctuation.
* Hawaiian Airlines maintains consistently low complaints.

**Boxplot Analysis**

* Purpose: To explore the distribution of complaints for each airline.

par(mfrow = c(1, 3))

boxplot(united\_window, main = "Monthly Complaints United Airlines")

boxplot(american\_window, main = "Monthly Complaints American Eagle")

boxplot(hawaiian\_window, main = "Monthly Complaints Hawaiian Airlines")

A row of gray rectangular boxes with numbers

Description automatically generated with medium confidence

* Boxplots for each airline highlight:
  + United Airlines: Wide range and higher median complaints.
  + American Eagle: Moderate range with some variability.
  + Hawaiian Airlines: Narrow range with consistently low complaints.

**Autocorrelation Function (ACF) Analysis**

* Purpose: To check for autocorrelation in the complaint data for each airline.

acf(united\_window, main = "ACF: United Airlines", col = "green", lwd = 2)

acf(american\_window, main = "ACF: American Eagle", col = "blue", lwd = 2)

acf(hawaiian\_window, main = "ACF: Hawaiian Airlines", col = "red", lwd = 2)

A group of graphs showing different types of flight

Description automatically generated with medium confidence

* United Airlines: Significant autocorrelation showing strong seasonal patterns.
* American Eagle: Moderate autocorrelation.
* Hawaiian Airlines: Minimal autocorrelation due to low variability.

**Decomposition**

* Purpose: To separate complaint data into trend, seasonal, and residual components.

united\_decomp <- decompose(united\_window)

american\_decomp <- decompose(american\_window)

hawaiian\_decomp <- decompose(hawaiian\_window)

plot(united\_decomp)

plot(american\_decomp)

plot(hawaiian\_decomp)

A graph of different types of time

Description automatically generated with medium confidence A graph of different types of time

Description automatically generated with medium confidence

A graph of different types of lines

Description automatically generated with medium confidence

* United Airlines: Strong seasonal pattern with a declining trend.
* American Eagle: Moderate seasonality with a weaker downward trend.
* Hawaiian Airlines: Minimal seasonality and a stable trend.

summary(united\_window)

## Min. 1st Qu. Median Mean 3rd Qu. Max.

## 8597 14256 18340 20315 25778 41787

summary(american\_window)

## Min. 1st Qu. Median Mean 3rd Qu. Max.

## 7052 9537 11576 13393 15664 26213

summary(hawaiian\_window)

## Min. 1st Qu. Median Mean 3rd Qu. Max.

## 1076 1318 1578 1729 2065 2791

**Historical Data included in time series**

Since trend is decline after 2007, took the data from 2007 for all the airlines.

united\_window <- window(united\_ts, start = c(2007, 1), end = c(2010,12), frequency = 12)

american\_window <- window(american\_ts, start = c(2007, 1), end = c(2010,12), frequency = 12)

hawaiian\_window <- window(hawaiian\_ts, start = c(2007, 1), end = c(2010,12), frequency = 12)

par(mfrow = c(1, 3))

plot(united\_window, main = "Monthly Complaints United Airlines",

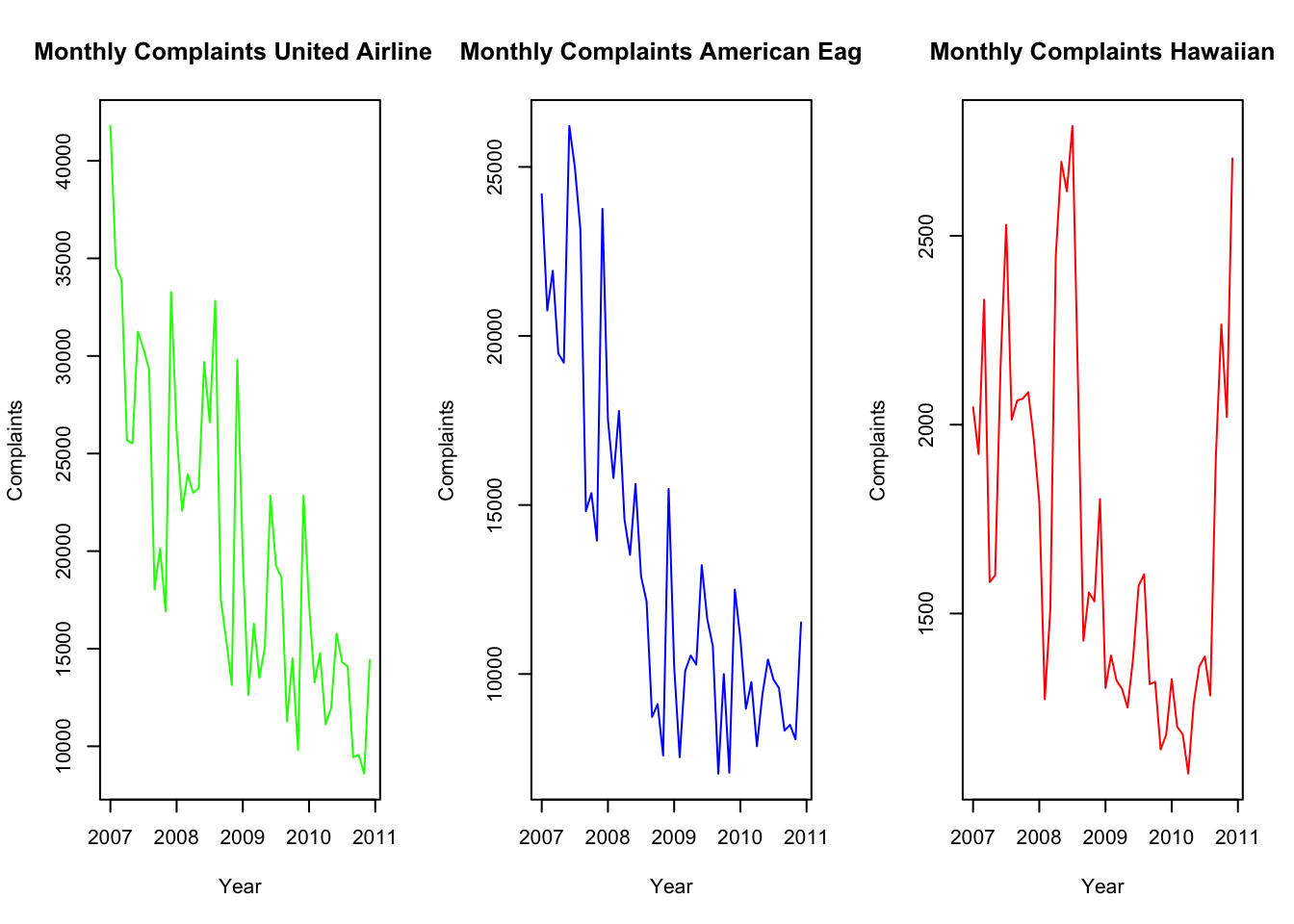
xlab = "Year", ylab = "Complaints", col = "green")

plot(american\_window, main = "Monthly Complaints American Eagle",

xlab = "Year", ylab = "Complaints", col = "blue")

plot(hawaiian\_window, main = "Monthly Complaints Hawaiian",

xlab = "Year", ylab = "Complaints", col = "red")



**United Airline**

* Most Complaints: United has the highest number of baggage complaints almost every month and year compared to the other two airlines.
* Complaints shot up a lot during 2006-2007, meaning something might have gone wrong—maybe too many passengers, poor baggage handling, or other issues.
* After 2007, complaints started going down steadily, showing that United probably made changes to fix their baggage issues.
* Has the highest complaints across all metrics, indicating significant challenges in baggage handling.
* Their average (20315) and maximum (41787) complaints highlight operational in efficiencies, especially during peak months.
* Both seasonality and trend are significant factors. Complaints tend to increase during specific seasons and persist for multiple months.

**American Eagle Airline**

* Moderate Complaints: American Eagle has fewer complaints than United but still more than Hawaiian.
* Complaints go up and down a lot, especially during busy travel times like summer and the holiday season.
* Like United, American Eagle’s complaints also started decreasing after 2007.
* Performs better than United but still has more complaints than Hawaiian Airlines.
* Their average (13393) and maximum (26213) complaints suggest room for improvement, particularly during busy periods.
* Shows weaker seasonality and trend compared to United. Complaints might still rise slightly during peak travel periods but don’t persist long.

**Hawaiian Airline**

* Lowest Complaints: Hawaiian Airlines has the least baggage complaints out of all three airlines, and the numbers stay consistently low throughout the years.
* Unlike the other two airlines, Hawaiian’s complaints don’t spike much, showing they handle baggage well no matter the season.
* Consistently outperforms both United and American Eagle in baggage handling.
* Their average (1729) and maximum (2791) complaints are significantly lower, showcasing their reliable operations.
* No seasonal influence or trend is visible. Complaints remain stable and consistent across all months.

**4. Accuracy Measure**

**Mean Absolute Percentage Error (MAPE)**:  
MAPE was selected for its ease of interpretation as a percentage. It provides a clear indication of how far off predictions are from actual values, making it ideal for comparing models across different scales of data.

**5. Insights from Forecasting Methods**

**1. Naive Forecasting**

* A simple forecasting method that assumes future values are equal to the most recent observed value.
* Best suited for time series data without trends or seasonality.

par(mfrow = c(1, 3))

united\_naive <- naive(united\_window, h = 12)

plot(united\_naive, main = "Naive Forecast: United Airlines",

xlab = "Year", ylab = "Complaints", col = "green")

american\_naive <- naive(american\_window, h = 12)

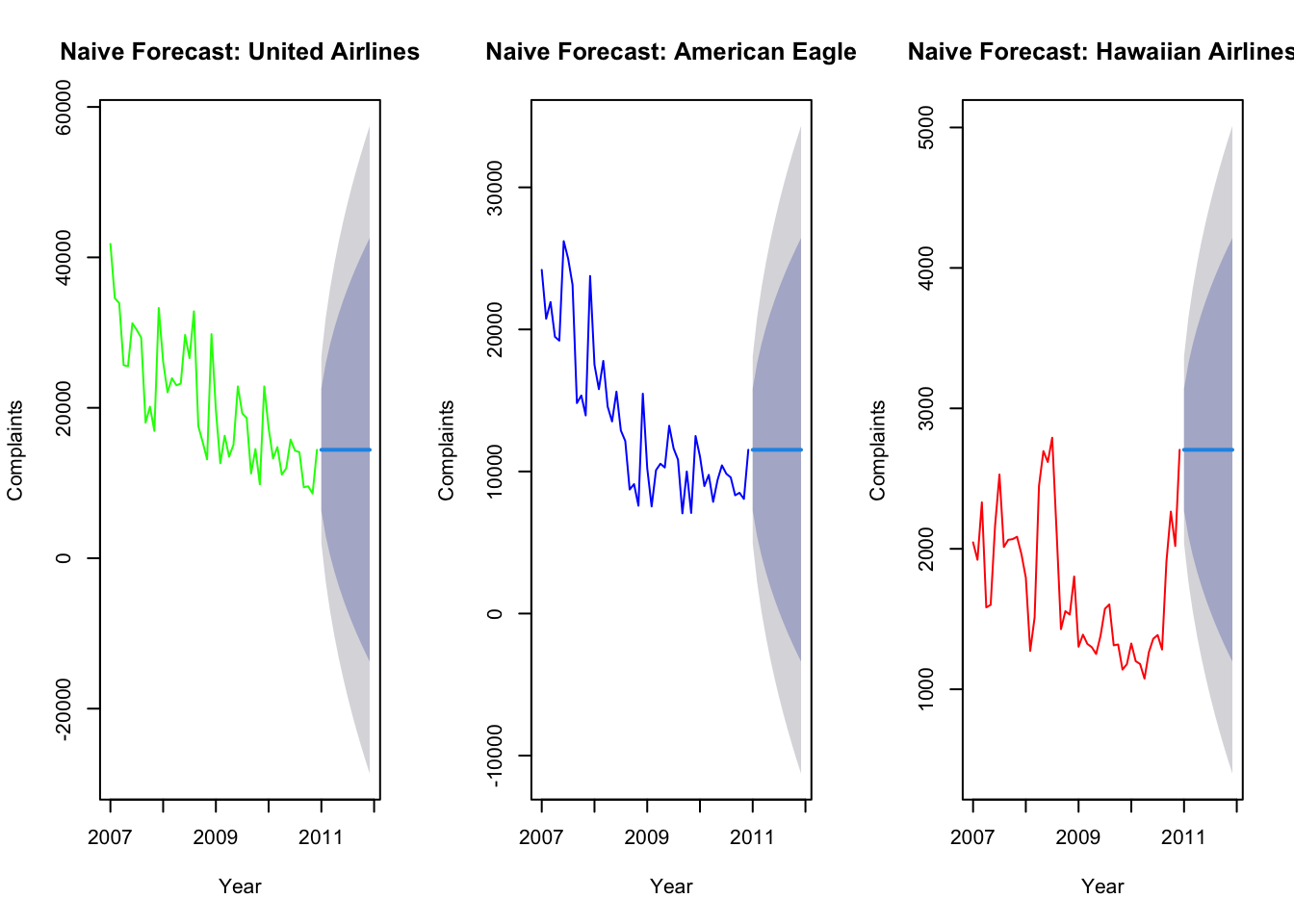
plot(american\_naive, main = "Naive Forecast: American Eagle",

xlab = "Year", ylab = "Complaints", col = "blue")

hawaiian\_naive <- naive(hawaiian\_window, h = 12)

plot(hawaiian\_naive, main = "Naive Forecast: Hawaiian Airlines",

xlab = "Year", ylab = "Complaints", col = "red")



#### **Residual Analysis**

residuals\_naive\_united <- residuals(united\_naive)

fitted\_naive\_united <- fitted(united\_naive)

residuals\_naive\_american <- residuals(american\_naive)

fitted\_naive\_american <- fitted(american\_naive)

residuals\_naive\_hawaiian <- residuals(hawaiian\_naive)

fitted\_naive\_hawaiian <- fitted(hawaiian\_naive)

par(mfrow = c(1, 3))

plot(residuals\_naive\_united, main = "United Airlines (Naive Model)",

ylab = "Residuals", xlab = "Time",col ='green')

plot(residuals\_naive\_american, main = "American Eagle Airlines (Naive Model)",

ylab = "Residuals", xlab = "Time", col ='blue')

plot(residuals\_naive\_hawaiian, main = "Hawaiian Airlines (Naive Model)",

ylab = "Residuals", xlab = "Time", col= 'red')



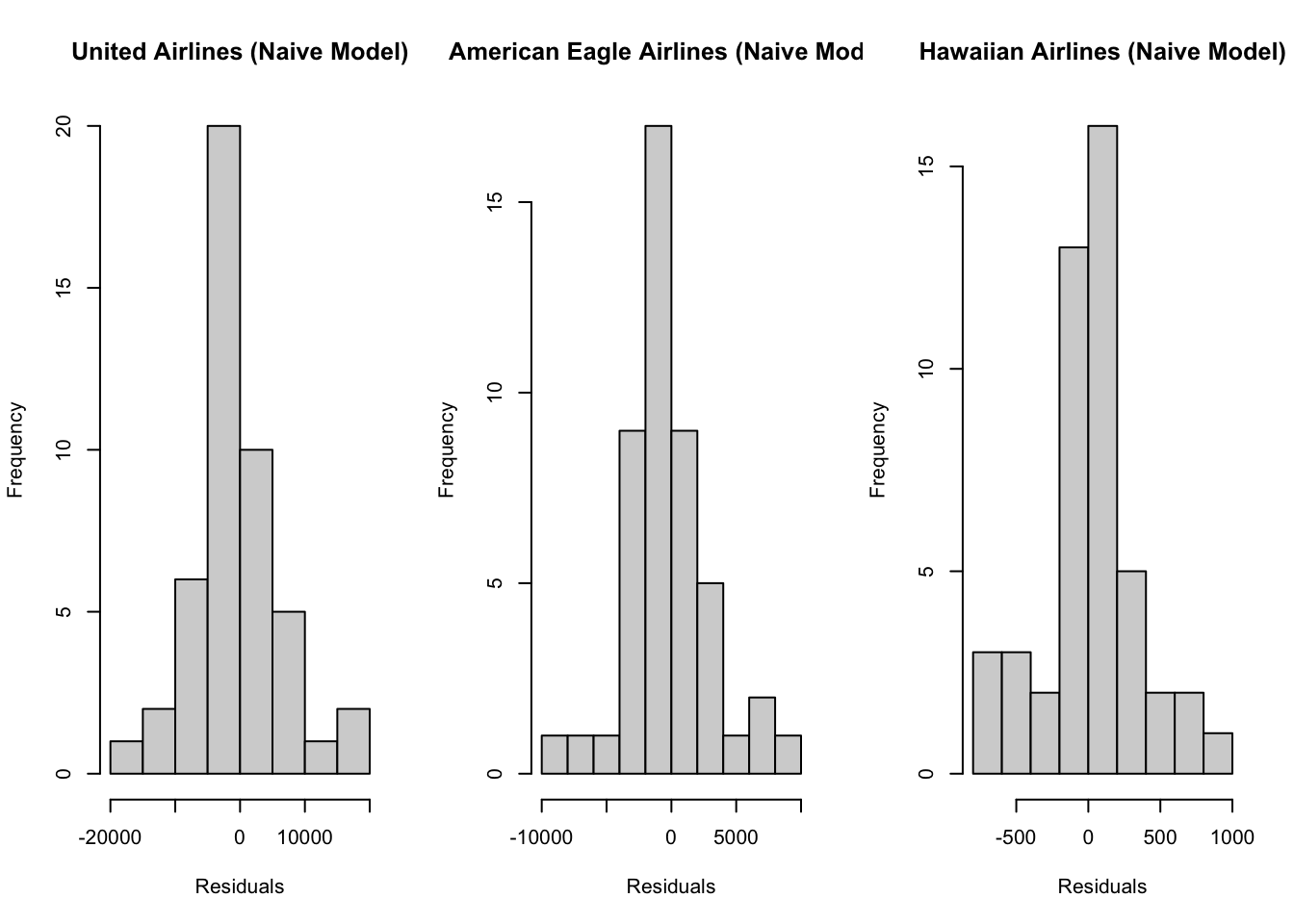
**Insights:**

* United Airlines: Produces a flat-line forecast, unsuitable for capturing trends or seasonality.
* American Eagle: Over-simplifies predictions, ignoring fluctuations.
* Hawaiian Airlines: Reasonable for stable data with no significant seasonality.

hist(residuals\_naive\_united, main = "United Airlines (Naive Model)", xlab = "Residuals")

hist(residuals\_naive\_american, main = "American Eagle Airlines (Naive Model)", xlab = "Residuals")

hist(residuals\_naive\_hawaiian, main = "Hawaiian Airlines (Naive Model)", xlab = "Residuals")



* **United Airlines**: Residuals are approximately centered around zero, with slight asymmetry.
* **American Eagle**: Residuals are more symmetrical, indicating a good fit.
* **Hawaiian Airlines**: Residuals are tightly centered, consistent with the stable data.

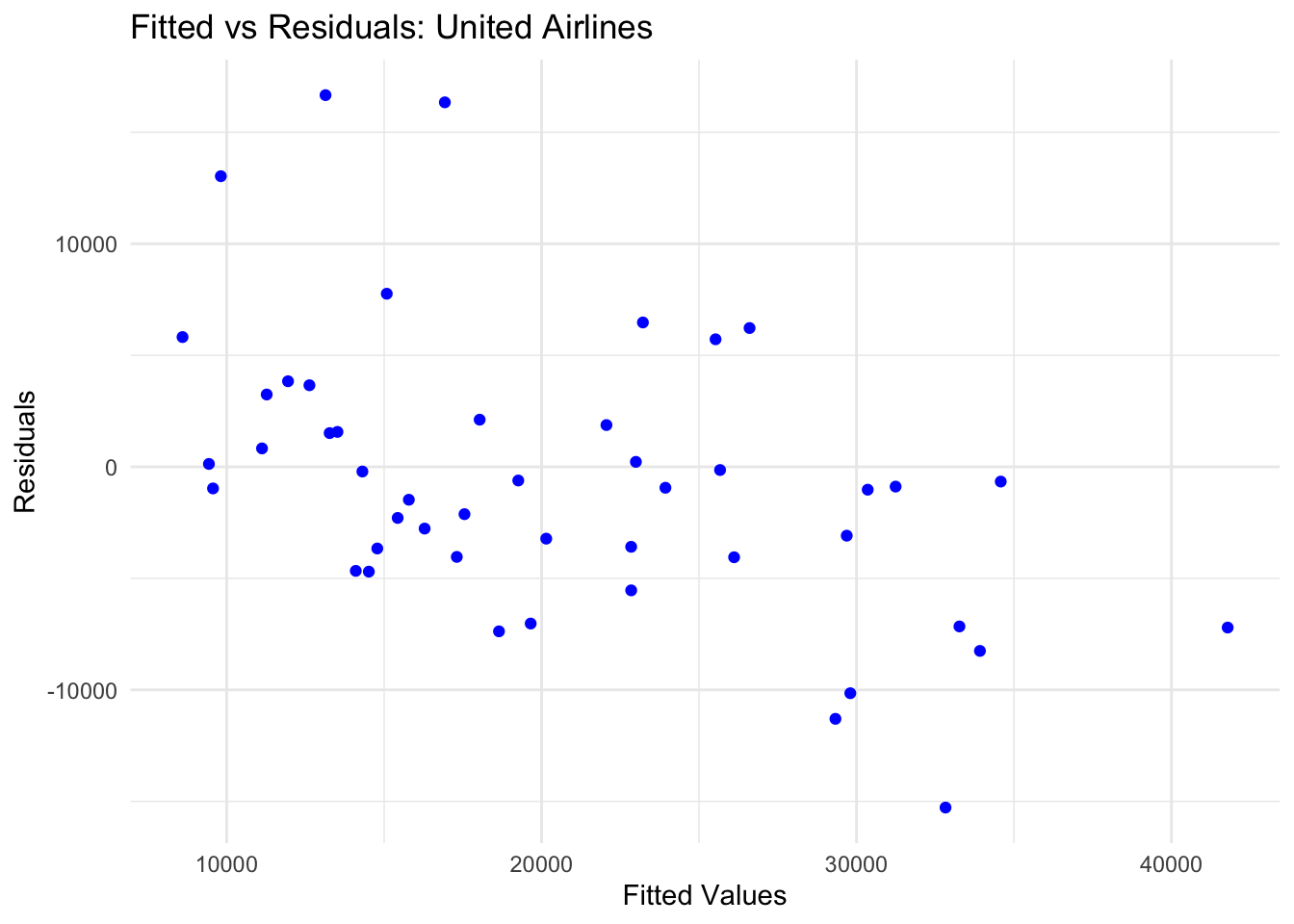
ggplot(data = NULL, aes(x = fitted\_naive\_united, y = residuals\_naive\_united)) +

geom\_point(color = "blue") +

labs(title = "Fitted vs Residuals: United Airlines",

x = "Fitted Values", y = "Residuals") +

theme\_minimal()



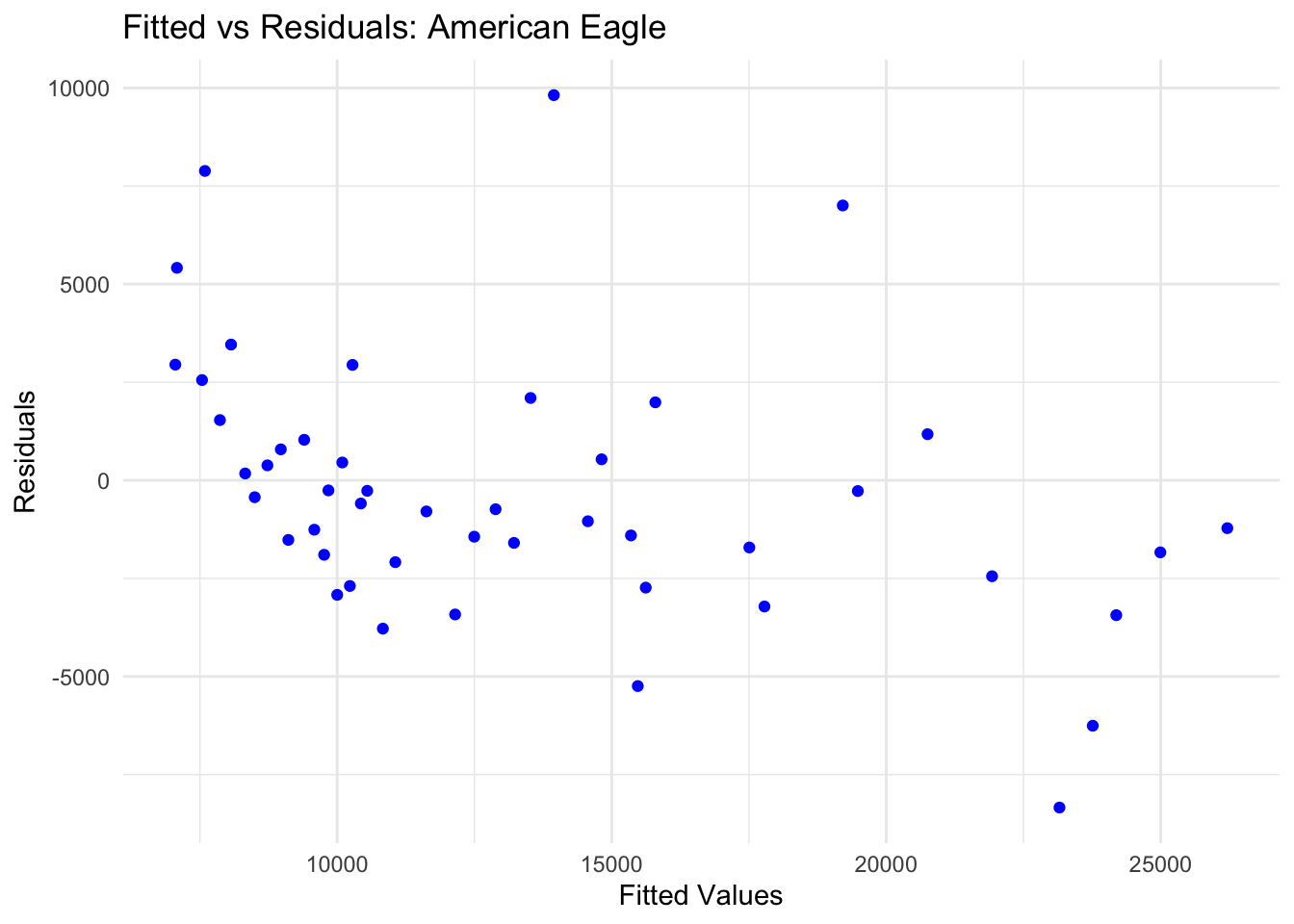
ggplot(data = NULL, aes(x = fitted\_naive\_american, y = residuals\_naive\_american)) +

geom\_point(color = "blue") +

labs(title = "Fitted vs Residuals: American Eagle ",

x = "Fitted Values", y = "Residuals") +

theme\_minimal()



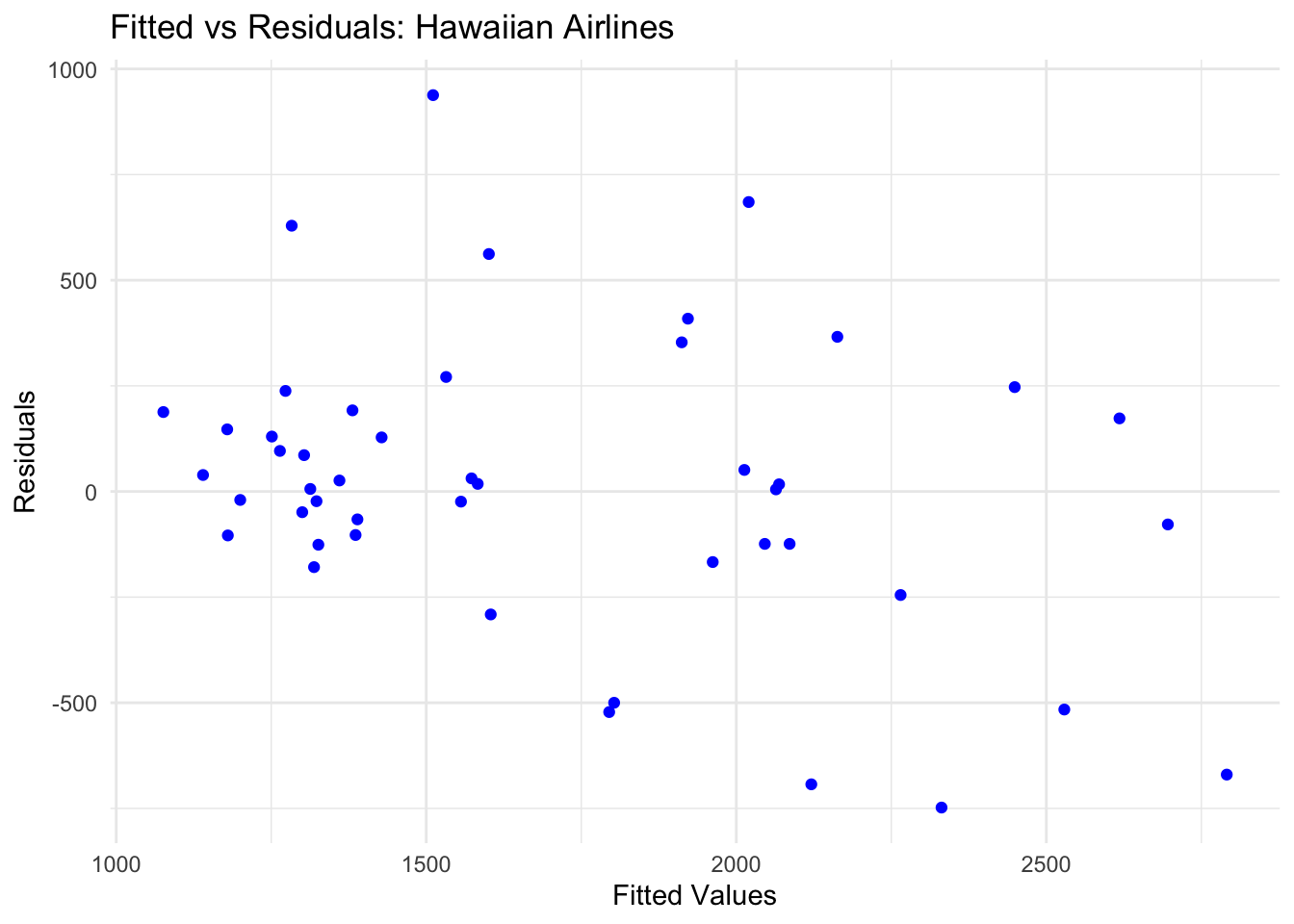
ggplot(data = NULL, aes(x = fitted\_naive\_hawaiian, y = residuals\_naive\_hawaiian)) +

geom\_point(color = "blue") +

labs(title = "Fitted vs Residuals: Hawaiian Airlines",

x = "Fitted Values", y = "Residuals") +

theme\_minimal()



## Actual Vs Residual

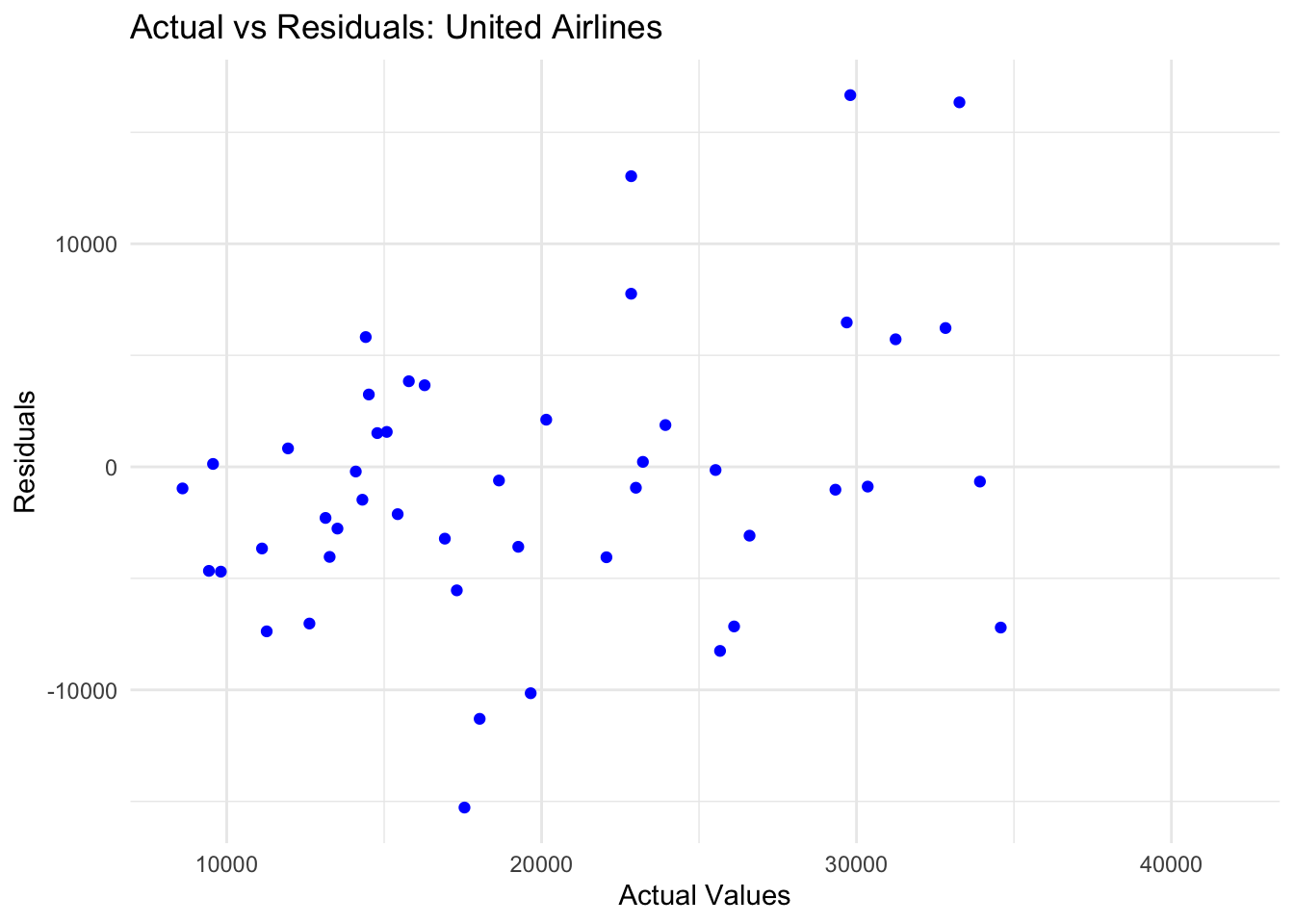
ggplot(data = NULL, aes(x = united\_window, y = residuals\_naive\_united)) +

geom\_point(color = "blue") +

labs(title = "Actual vs Residuals: United Airlines",

x = "Actual Values", y = "Residuals") +

theme\_minimal()



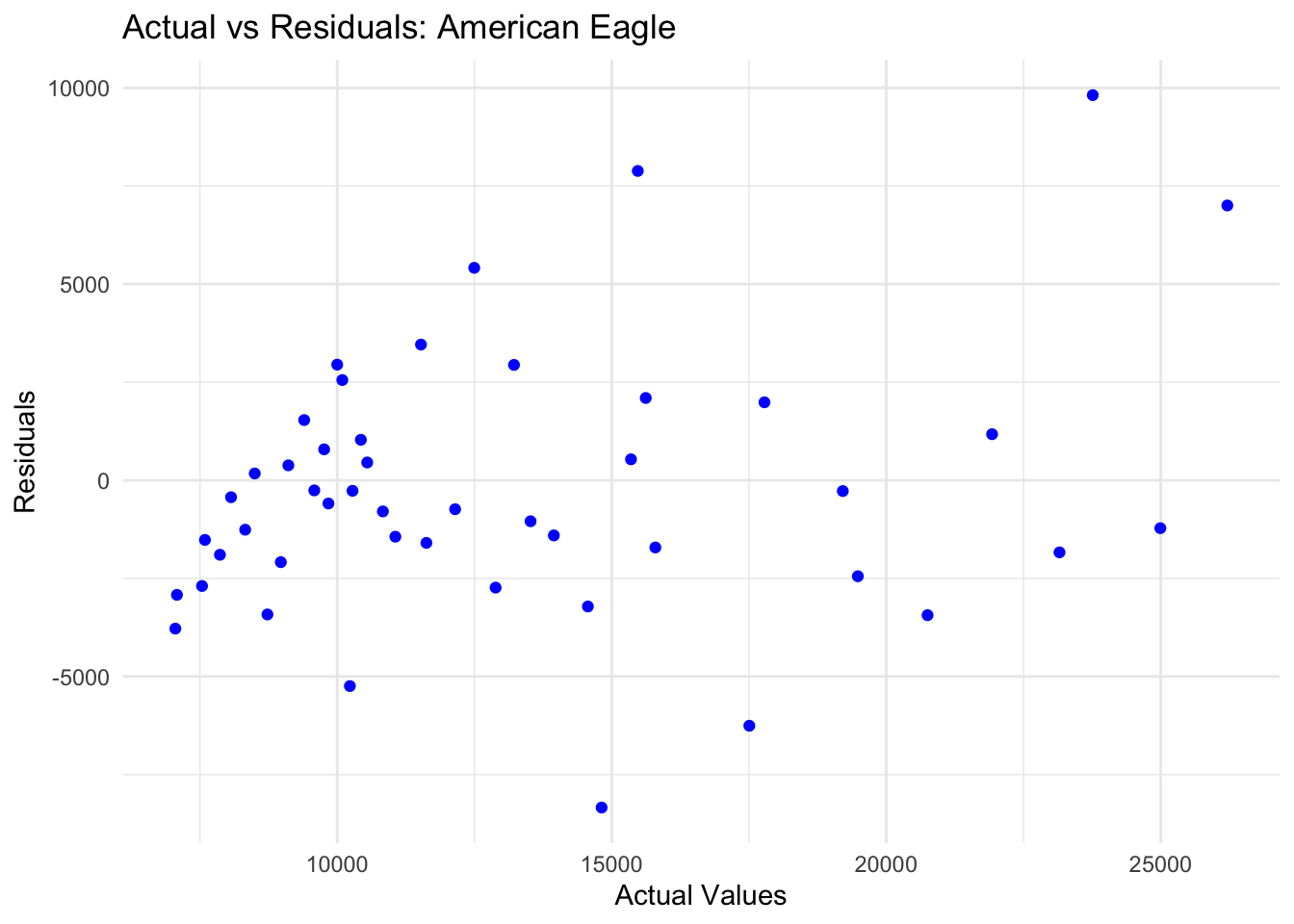
ggplot(data = NULL, aes(x = american\_window, y = residuals\_naive\_american)) +

geom\_point(color = "blue") +

labs(title = "Actual vs Residuals: American Eagle ",

x = "Actual Values", y = "Residuals") +

theme\_minimal()



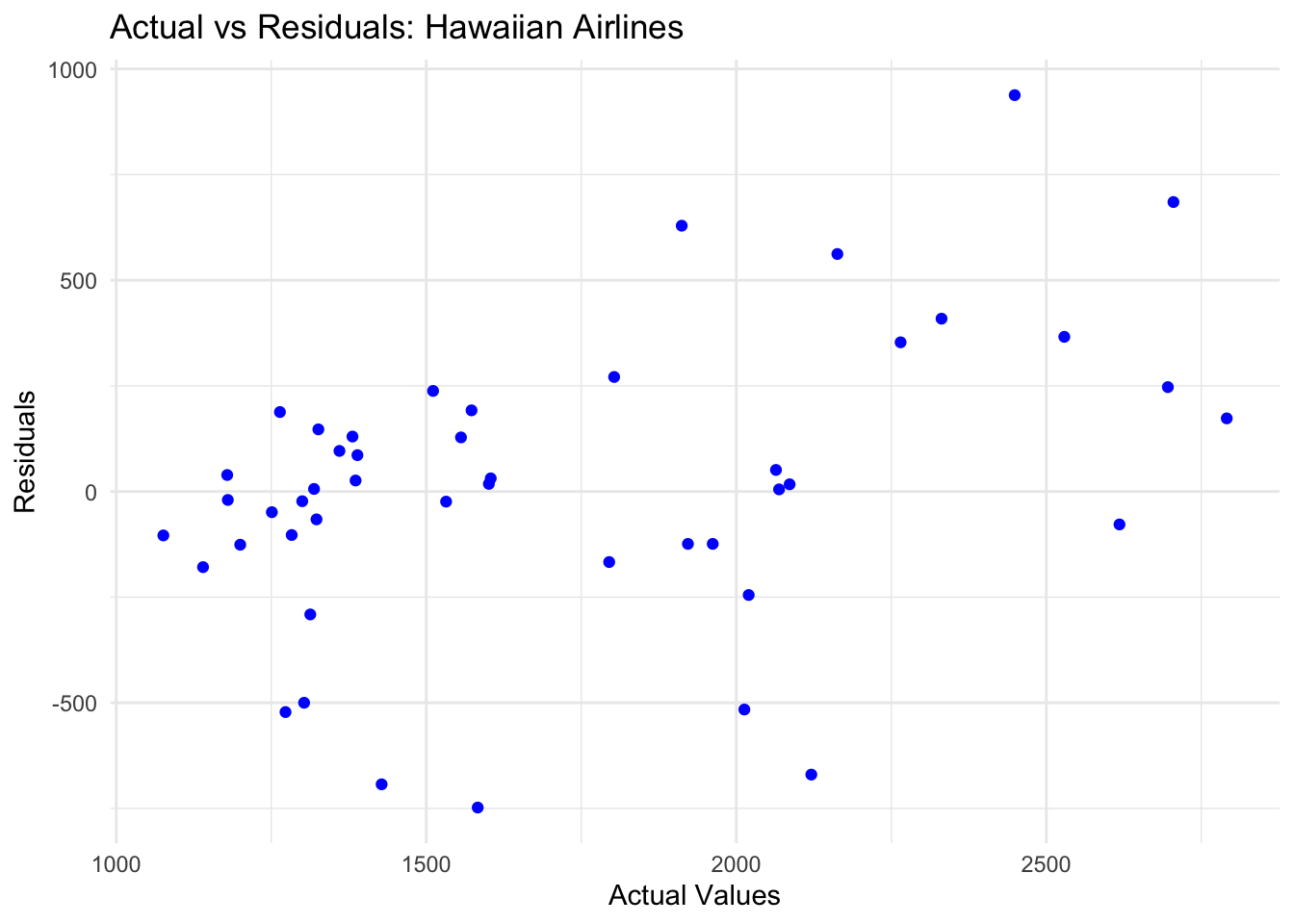
ggplot(data = NULL, aes(x = hawaiian\_window, y = residuals\_naive\_hawaiian)) +

geom\_point(color = "blue") +

labs(title = "Actual vs Residuals: Hawaiian Airlines",

x = "Actual Values", y = "Residuals") +

theme\_minimal()



* **United Airlines**: Residuals show slight patterns, possibly due to seasonality not captured by simpler models.
* **American Eagle**: Minimal patterns observed, validating the model.
* **Hawaiian Airlines**: Random distribution, reflecting good model fit.

**2. Simple Average (Mean Forecasting)**

* Predicts future values as the mean of all historical observations.
* Ignores trends and seasonality, making it suitable for stable series.

united\_sa <- meanf(united\_window, h=12)

american\_sa <- meanf(american\_window, h=12)

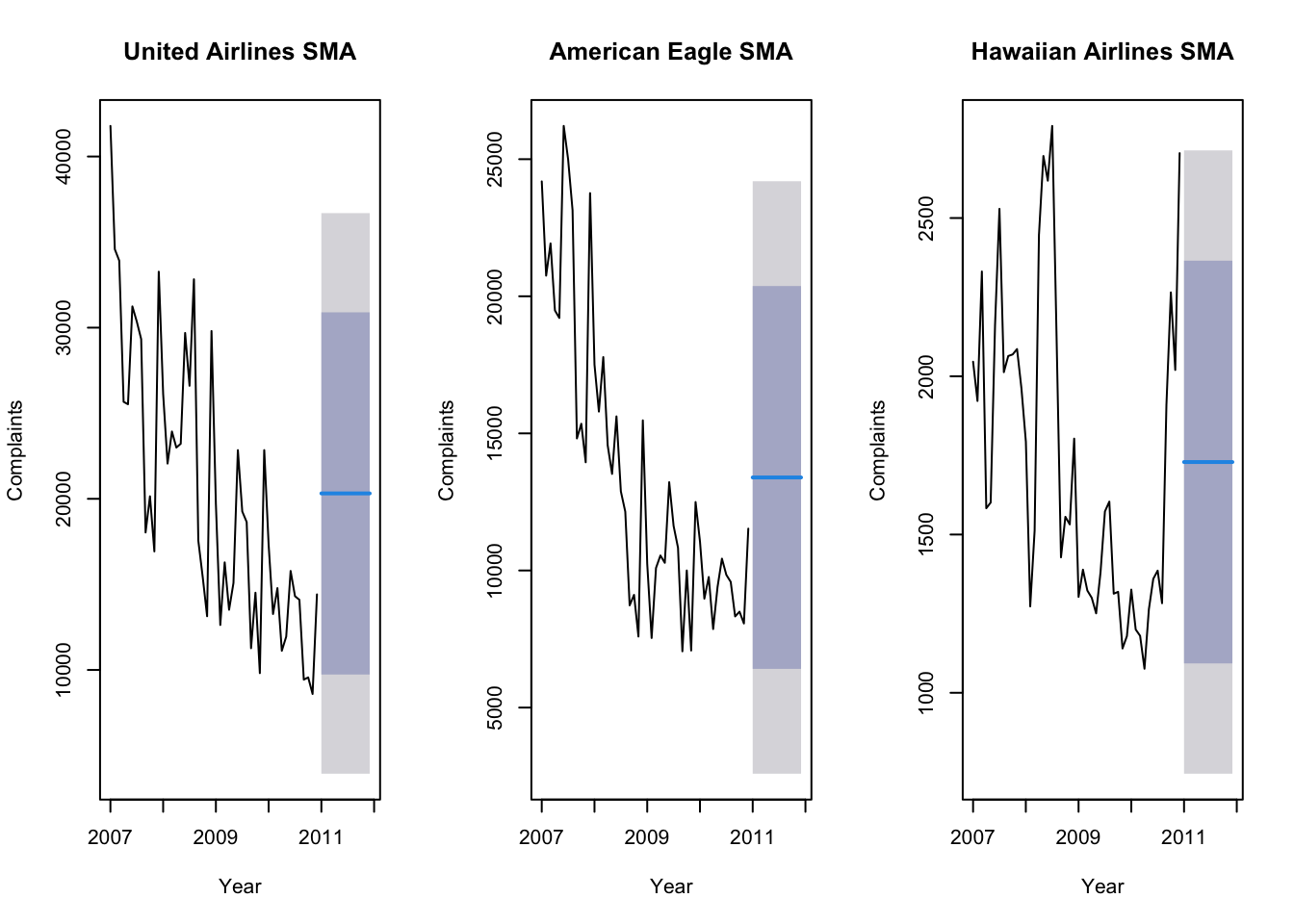
hawaiian\_sa <- meanf(hawaiian\_window, h=12)

par(mfrow = c(1, 3))

plot(united\_sa, main = "United Airlines SMA", xlab = "Year", ylab = "Complaints")

plot(american\_sa, main = "American Eagle SMA", xlab = "Year", ylab = "Complaints")

plot(hawaiian\_sa, main = "Hawaiian Airlines SMA", xlab = "Year", ylab = "Complaints")



* **United Airlines**: The flat forecast fails to capture downward trends or seasonal patterns.
* **American Eagle**: Similar limitations due to fluctuating data.
* **Hawaiian Airlines**: Works well for stable, non-trending complaints.

**3. Moving Average (MA)**

* Smooths the time series by averaging observations over a sliding window.
* Helps identify underlying trends by reducing noise.

united\_ma <- ma(united\_window, order = 3)

american\_ma <- ma(american\_window, order = 3)

hawaiian\_ma <- ma(hawaiian\_window, order = 3)

par(mfrow = c(1, 3))

plot(united\_window, main = "United Airlines MA", xlab = "Year", ylab = "Complaints")

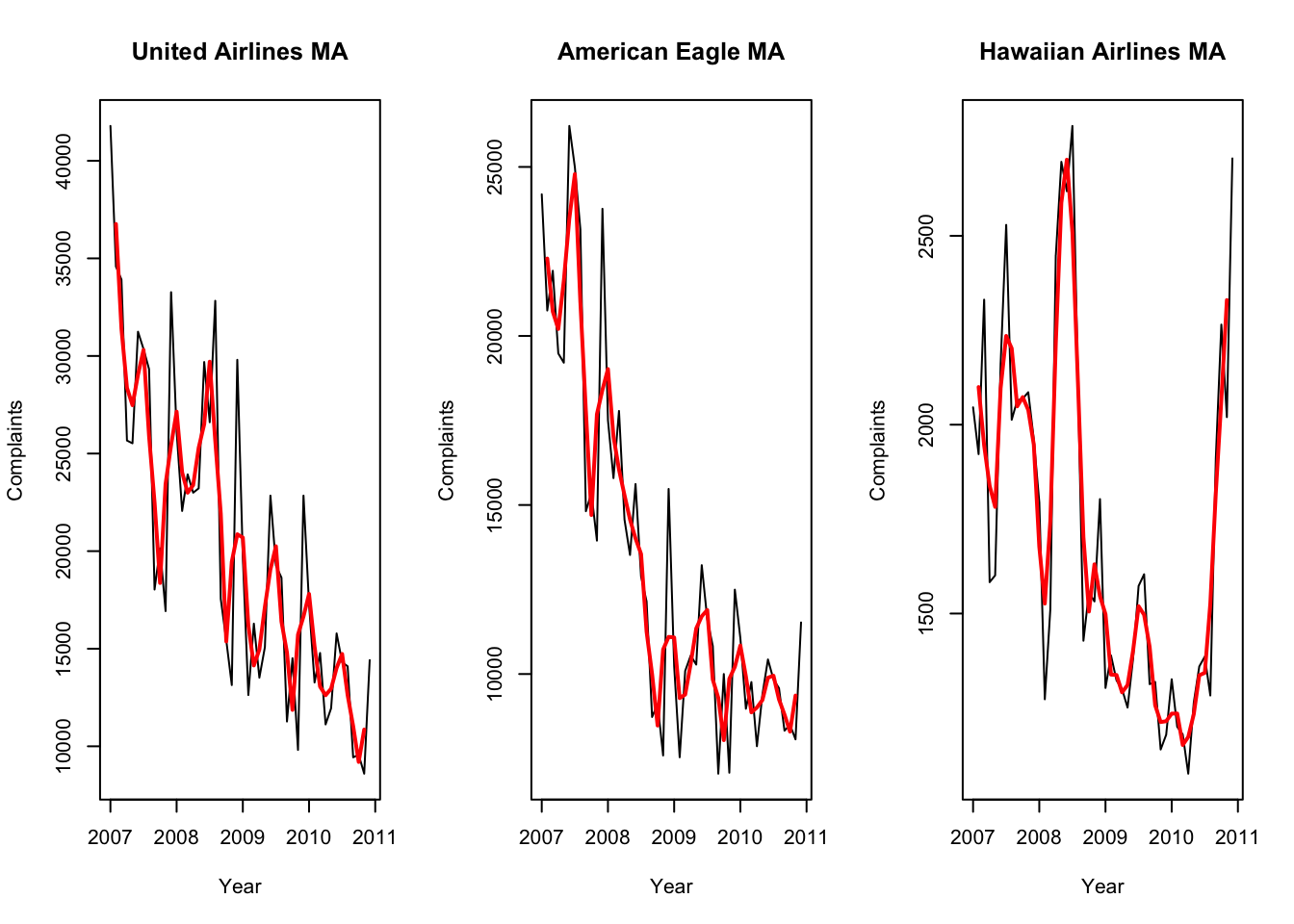
lines(united\_ma, col = "red", lwd = 2)

plot(american\_window, main = "American Eagle MA", xlab = "Year", ylab = "Complaints")

lines(american\_ma, col = "red", lwd = 2)

plot(hawaiian\_window, main = "Hawaiian Airlines MA", xlab = "Year", ylab = "Complaints")

lines(hawaiian\_ma, col = "red", lwd = 2)



**United Airlines** The moving average smooths out the fluctuations, highlighting the downward trend in complaints over time.

**American Eagle** The moving average closely follows the data, reflecting smoother declines with smaller fluctuation.

**Hawaiian Airlines** The moving average shows how stable Hawaiian Airlines is compared to the other airlines, with minimal fluctuations.

**4. Exponential Smoothing (ETS)**

* A weighted average of past observations, assigning exponentially decreasing weights to older data.
* Captures trend and seasonality for better accuracy.

ets\_united <- ets(united\_window)

ets\_american <- ets(american\_window)

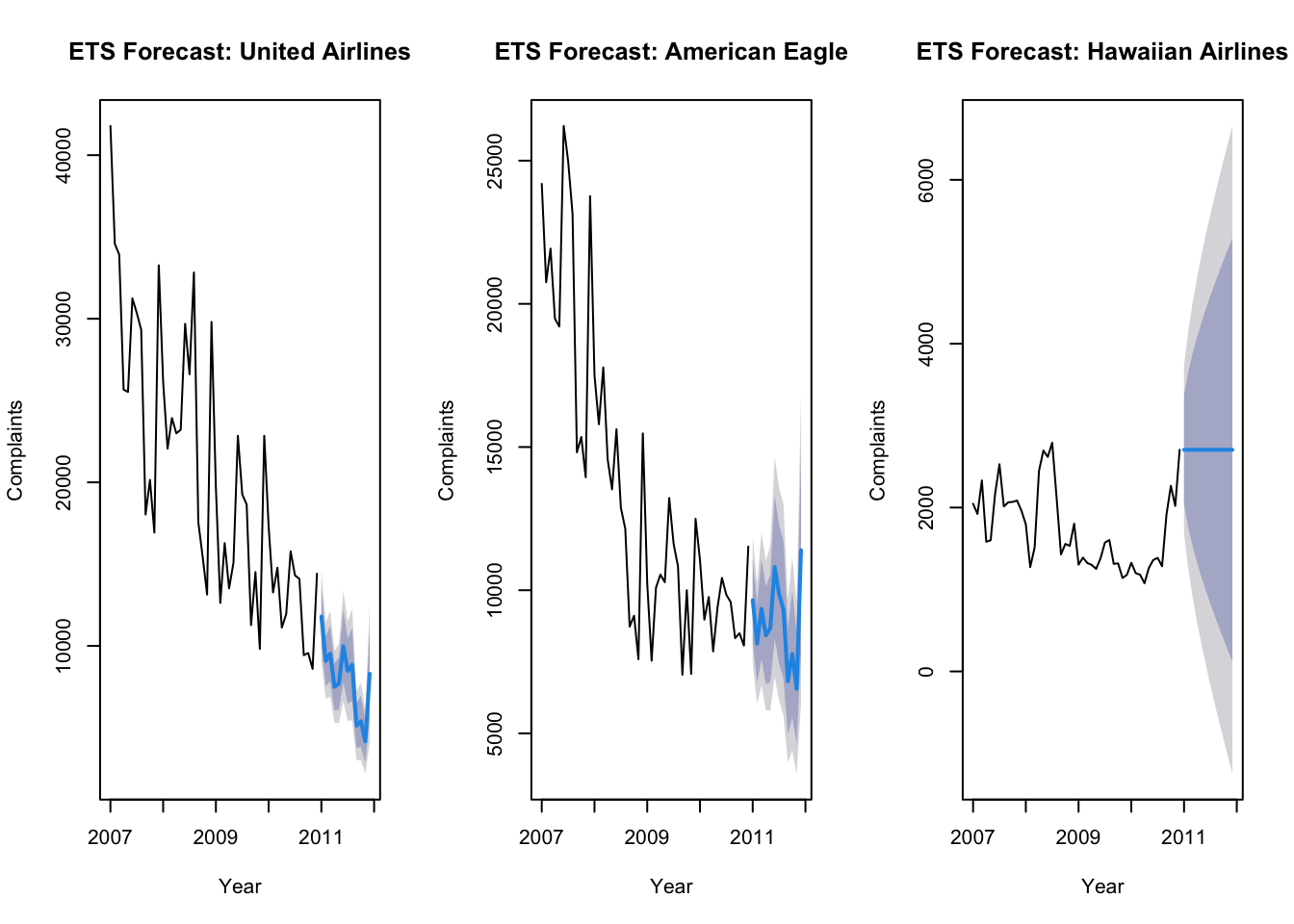
ets\_hawaiian <- ets(hawaiian\_window)

par(mfrow = c(1, 3))

plot(forecast(ets\_united, h = 12), main = "ETS Forecast: United Airlines", xlab = "Year", ylab = "Complaints")

plot(forecast(ets\_american, h = 12), main = "ETS Forecast: American Eagle", xlab = "Year", ylab = "Complaints")

plot(forecast(ets\_hawaiian, h = 12), main = "ETS Forecast: Hawaiian Airlines", xlab = "Year", ylab = "Complaints")



**United Airlines**

* Complaints are going down steadily over time.
* The forecast shows that complaints will stay low and stable in the future
* Accurately captures the declining trend and seasonal patterns.

**American Eagle**

* Complaints are also decreasing, but not as fast as United Airlines.
* The forecast predicts complaints will drop per month.
* Effective in reflecting seasonal fluctuations.

**Hawaiian Airlines**

* Complaints are very steady with only small ups and downs over time.
* The forecast predicts complaints will stay low.

summary(ets\_united)

## ETS(M,A,M)

##

## Training set error measures:

## ME RMSE MAE MPE MAPE MASE ACF1

## Training set -111.5725 2197.537 1593.736 -1.010138 7.813287 0.2950907 0.3705239

summary(ets\_american)

## ETS(M,Ad,M)

##

## Call:

## ets(y = american\_window)

##

## Smoothing parameters:

## alpha = 0.5059

## beta = 1e-04

## gamma = 1e-04

## phi = 0.962

##

## Initial states:

## l = 24352.5504

## b = -488.4151

## s = 1.3272 0.7592 0.8949 0.7806 1.0595 1.1138

## 1.2125 0.9661 0.9313 1.0259 0.8863 1.0428

##

## sigma: 0.1176

##

## AIC AICc BIC

## 901.9431 925.5293 935.6247

##

## Training set error measures:

## ME RMSE MAE MPE MAPE MASE ACF1

## Training set -173.7653 1179.871 965.1317 -1.613666 7.767323 0.2391307 0.246096

summary(ets\_hawaiian)

## ETS(M,N,N)

##

## Call:

## ets(y = hawaiian\_window)

##

## Smoothing parameters:

## alpha = 0.9999

##

## Initial states:

## l = 1975.8368

##

## sigma: 0.1937

##

## AIC AICc BIC

## 743.7838 744.3293 749.3974

##

## Training set error measures:

## ME RMSE MAE MPE MAPE MASE ACF1

## Training set 15.191 336.2452 239.0193 -1.012701 13.55208 0.4488625 -0.001138732

**United Airlines**

* Alpha = 0.3861, The model gives some weight to recent complaints but also considers historical data.
* Beta = 0.0001, very low, slow change in trend
* Gamma = 0.0002 very low, seasonal patterns are stable.

**American Eagle**

* Aplha = 0.5059, the model reacts more quickly to changes in the level.
* Beta = 0.0001, very low, slow change in trend
* Gamma = 0.0001, Very low. Seasonal patterns remain stable over time.

**Hawaiian Airlines**

* Alpha = 0.9999, Strong reliance on the most recent data, there is no trend or seasonality.

#### **Residual Analysis**

residuals\_ets\_united <- residuals(ets\_united)

fitted\_ets\_united <- fitted(ets\_united)

residuals\_ets\_american <- residuals(ets\_american)

fitted\_ets\_american <- fitted(ets\_american)

residuals\_ets\_hawaiian <- residuals(ets\_hawaiian)

fitted\_ets\_hawaiian <- fitted(ets\_hawaiian)

par(mfrow = c(1, 3))

plot(residuals\_ets\_united, main = "United Airlines (ETS)",

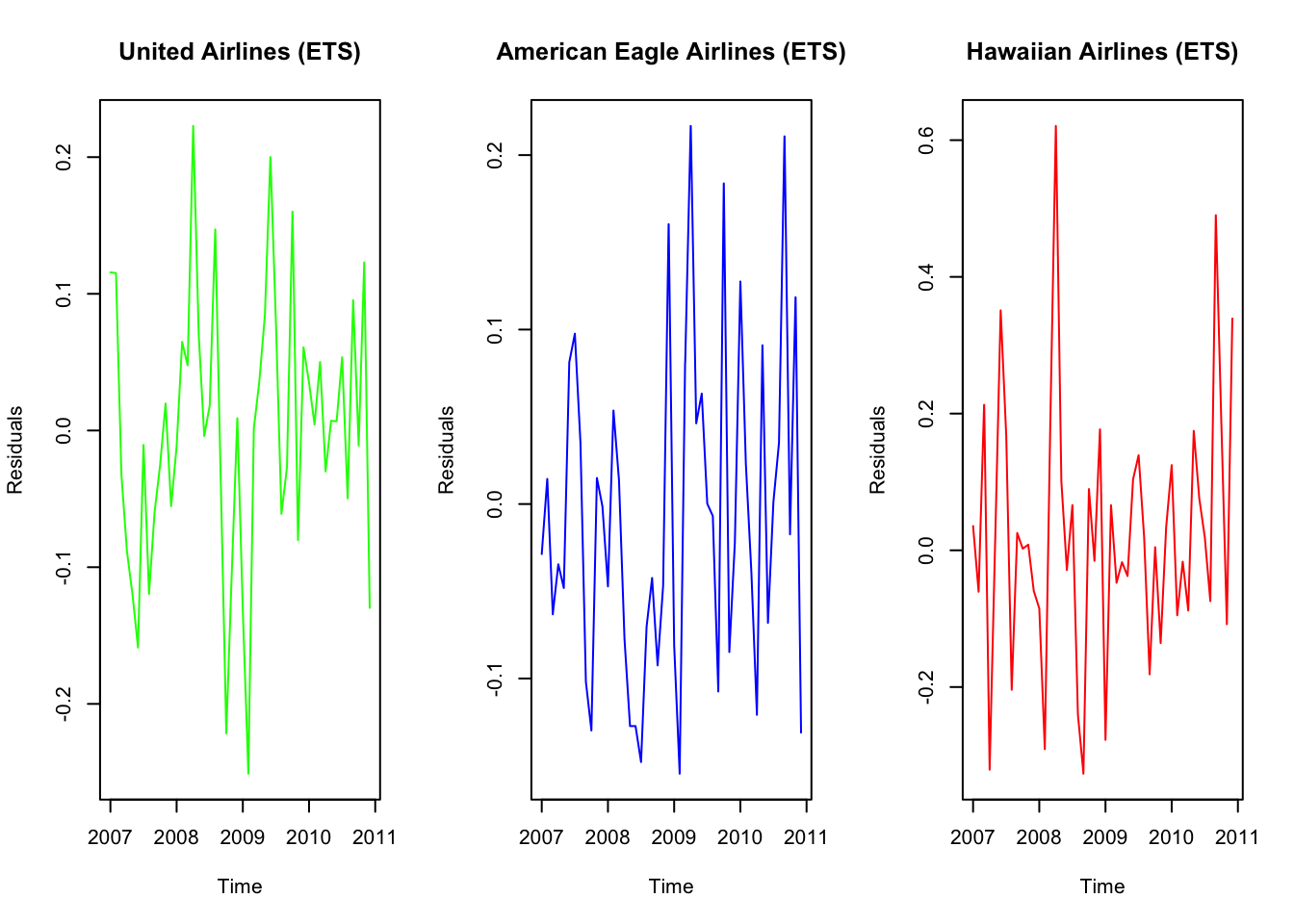
ylab = "Residuals", xlab = "Time",col ='green')

plot(residuals\_ets\_american, main = "American Eagle Airlines (ETS)",

ylab = "Residuals", xlab = "Time", col ='blue')

plot(residuals\_ets\_hawaiian, main = "Hawaiian Airlines (ETS)",

ylab = "Residuals", xlab = "Time", col= 'red')

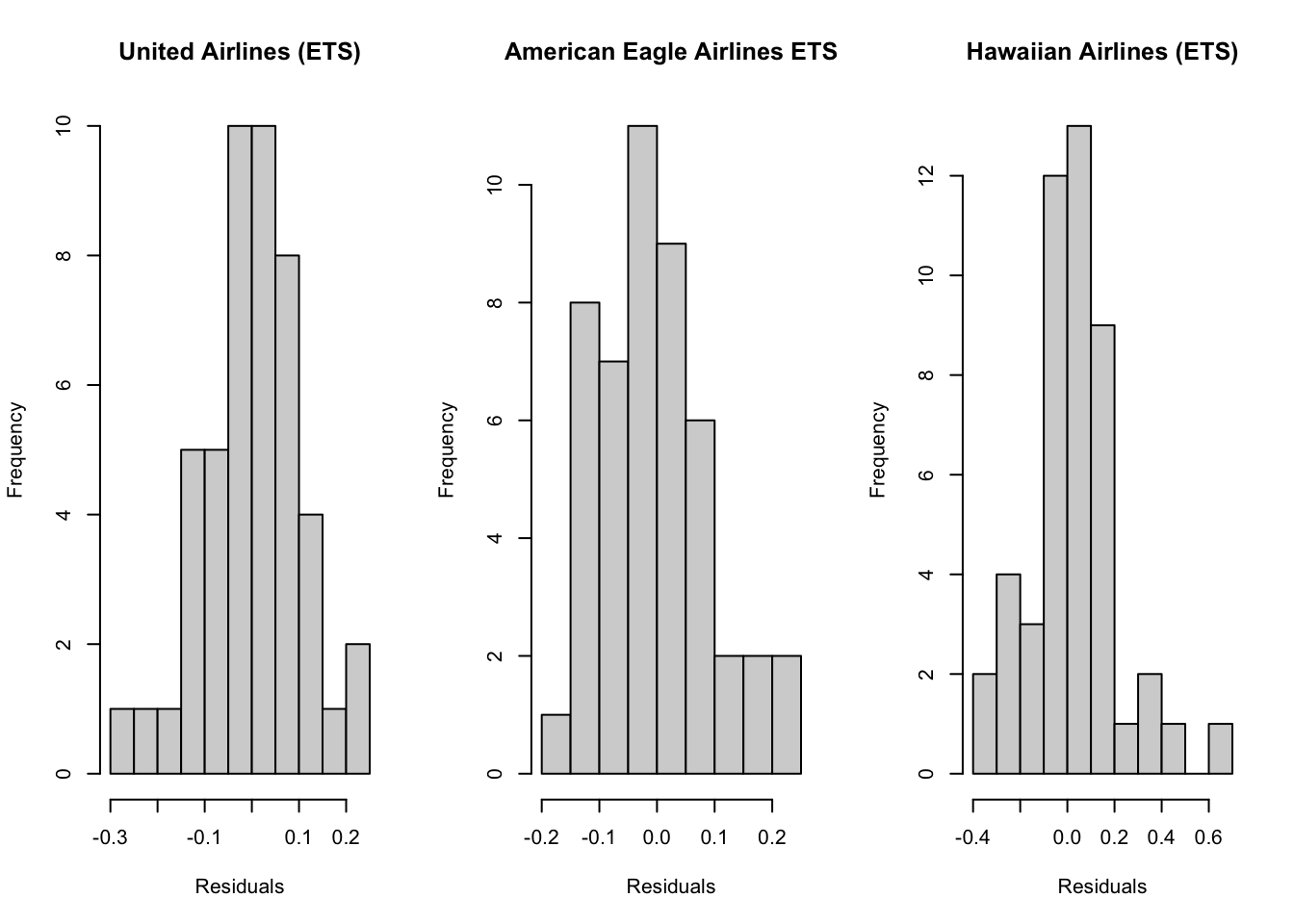


* **United Airlines**: Residuals fluctuate around zero, indicating the model captures the trend well but may have some unexplained variability.
* **American Eagle**: Residuals are random and show less variance, validating a good model fit.
* **Hawaiian Airlines**: Residuals are consistently centered around zero with minimal variability.

hist(residuals\_ets\_united, main = "United Airlines (ETS)", xlab = "Residuals")

hist(residuals\_ets\_american, main = "American Eagle Airlines ETS", xlab = "Residuals")

hist(residuals\_ets\_hawaiian, main = "Hawaiian Airlines (ETS)", xlab = "Residuals")



* + **United Airlines**: Residuals are slightly skewed but mostly symmetric.
  + **American Eagle**: Symmetrical distribution, supporting model assumptions.
  + **Hawaiian Airlines**: Residuals are tightly clustered, indicating a strong fit for stable data.

ggplot(data = NULL, aes(x = fitted\_ets\_united, y = residuals\_ets\_united)) +

geom\_point(color = "blue") +

labs(title = "Fitted vs Residuals: United Airlines",

x = "Fitted Values", y = "Residuals") +

theme\_minimal()



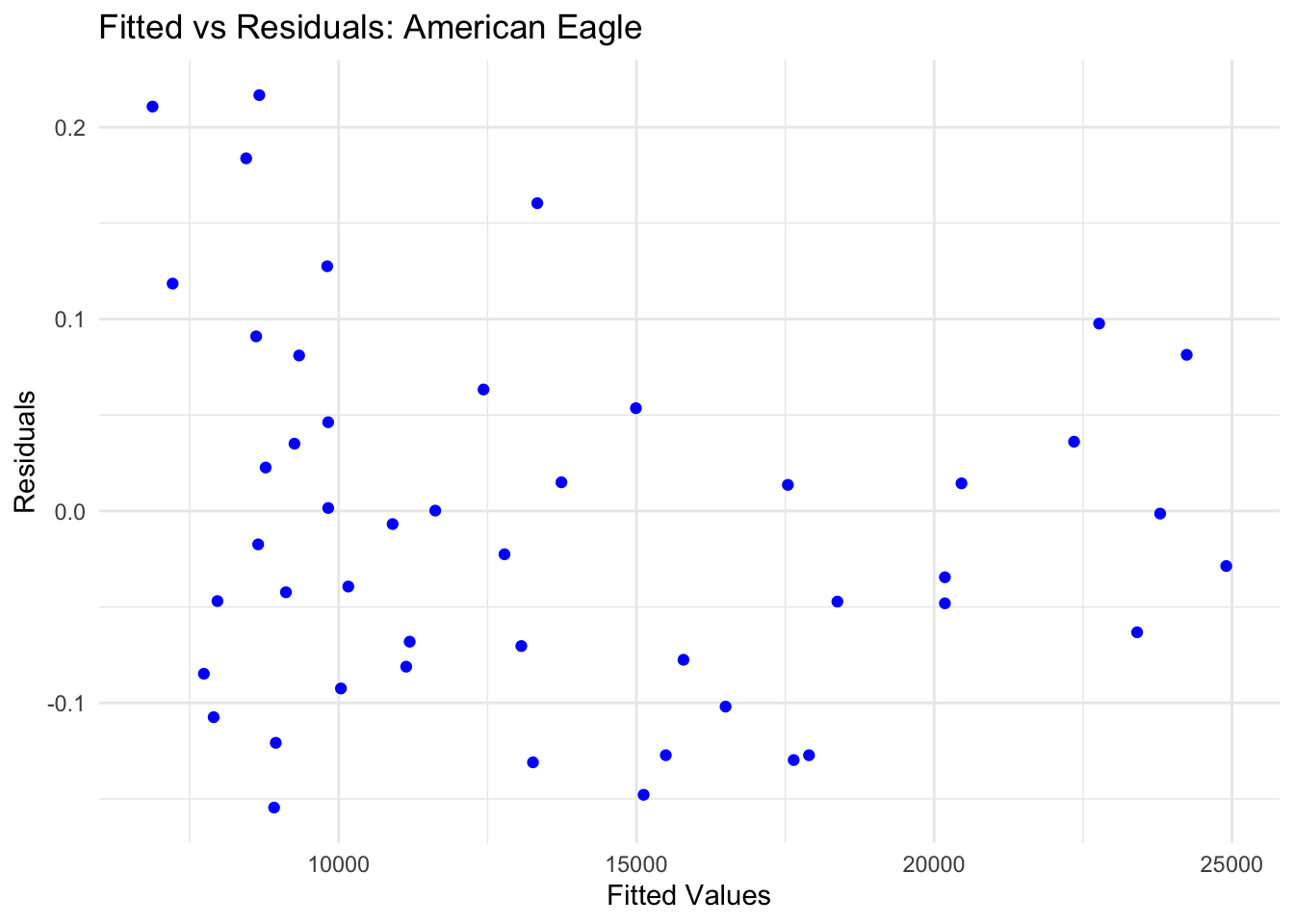
ggplot(data = NULL, aes(x = fitted\_ets\_american, y = residuals\_ets\_american)) +

geom\_point(color = "blue") +

labs(title = "Fitted vs Residuals: American Eagle ",

x = "Fitted Values", y = "Residuals") +

theme\_minimal()



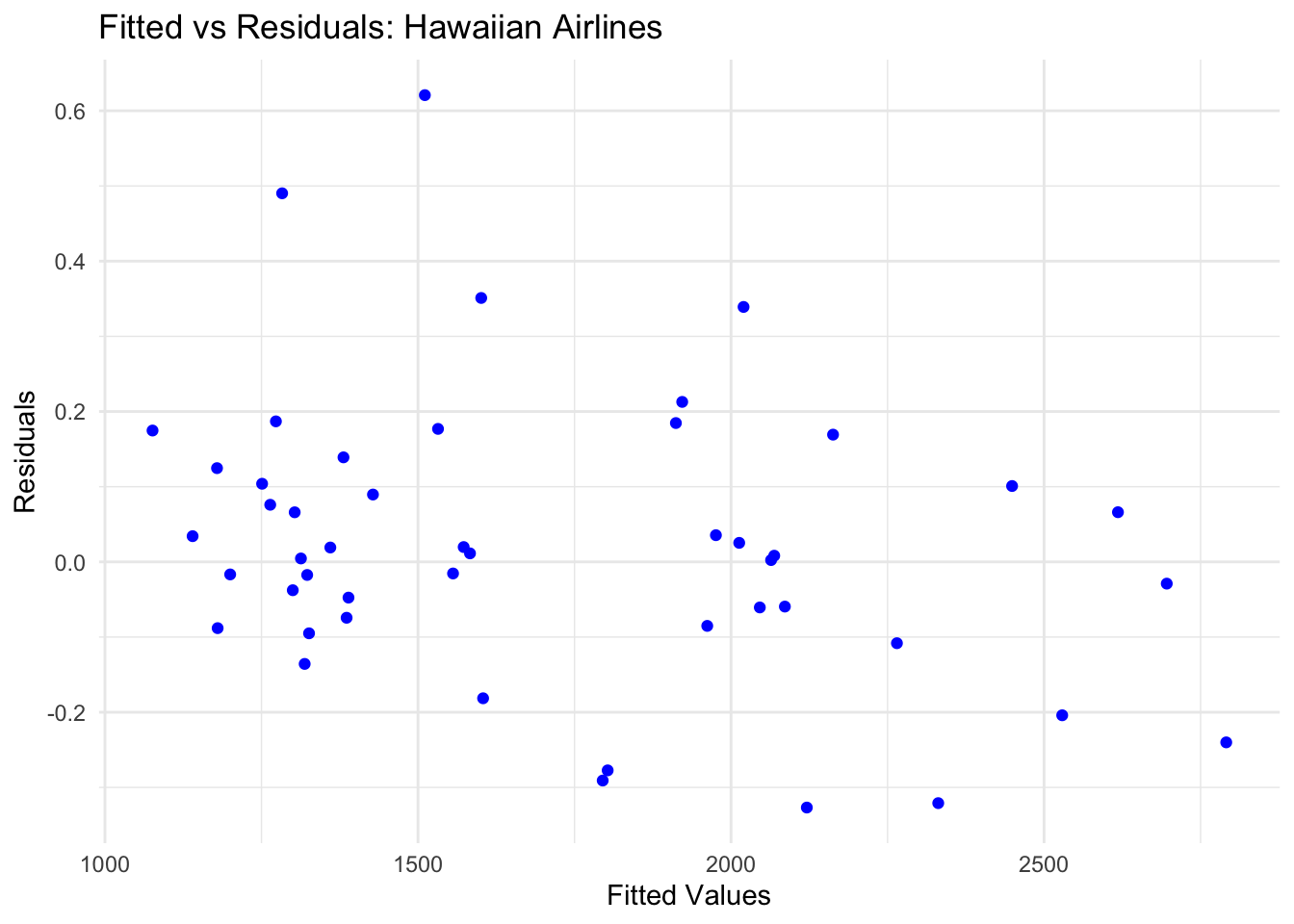
ggplot(data = NULL, aes(x = fitted\_ets\_hawaiian, y = residuals\_ets\_hawaiian)) +

geom\_point(color = "blue") +

labs(title = "Fitted vs Residuals: Hawaiian Airlines",

x = "Fitted Values", y = "Residuals") +

theme\_minimal()



## Actual Vs Residual

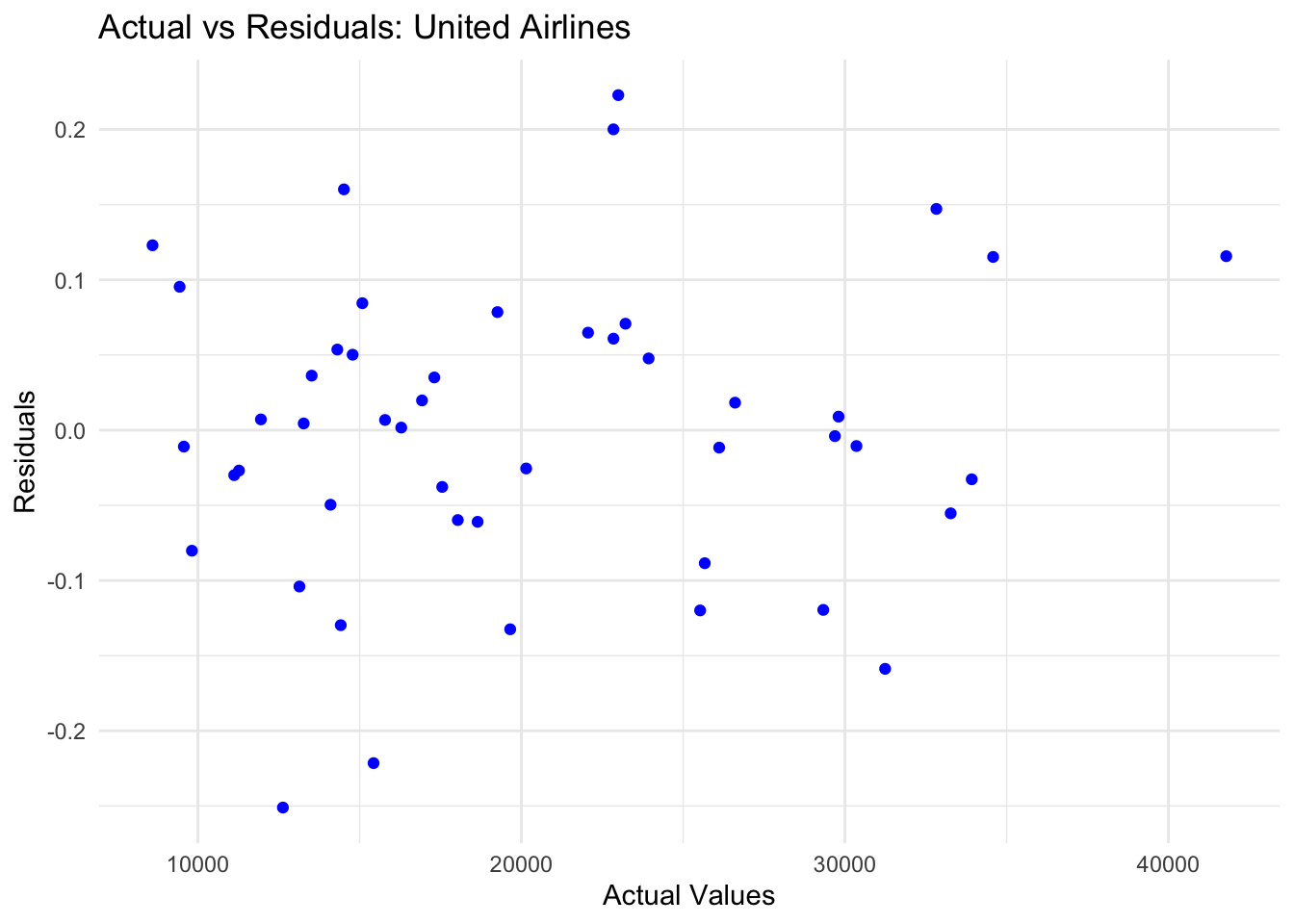
ggplot(data = NULL, aes(x = united\_window, y = residuals\_ets\_united)) +

geom\_point(color = "blue") +

labs(title = "Actual vs Residuals: United Airlines",

x = "Actual Values", y = "Residuals") +

theme\_minimal()



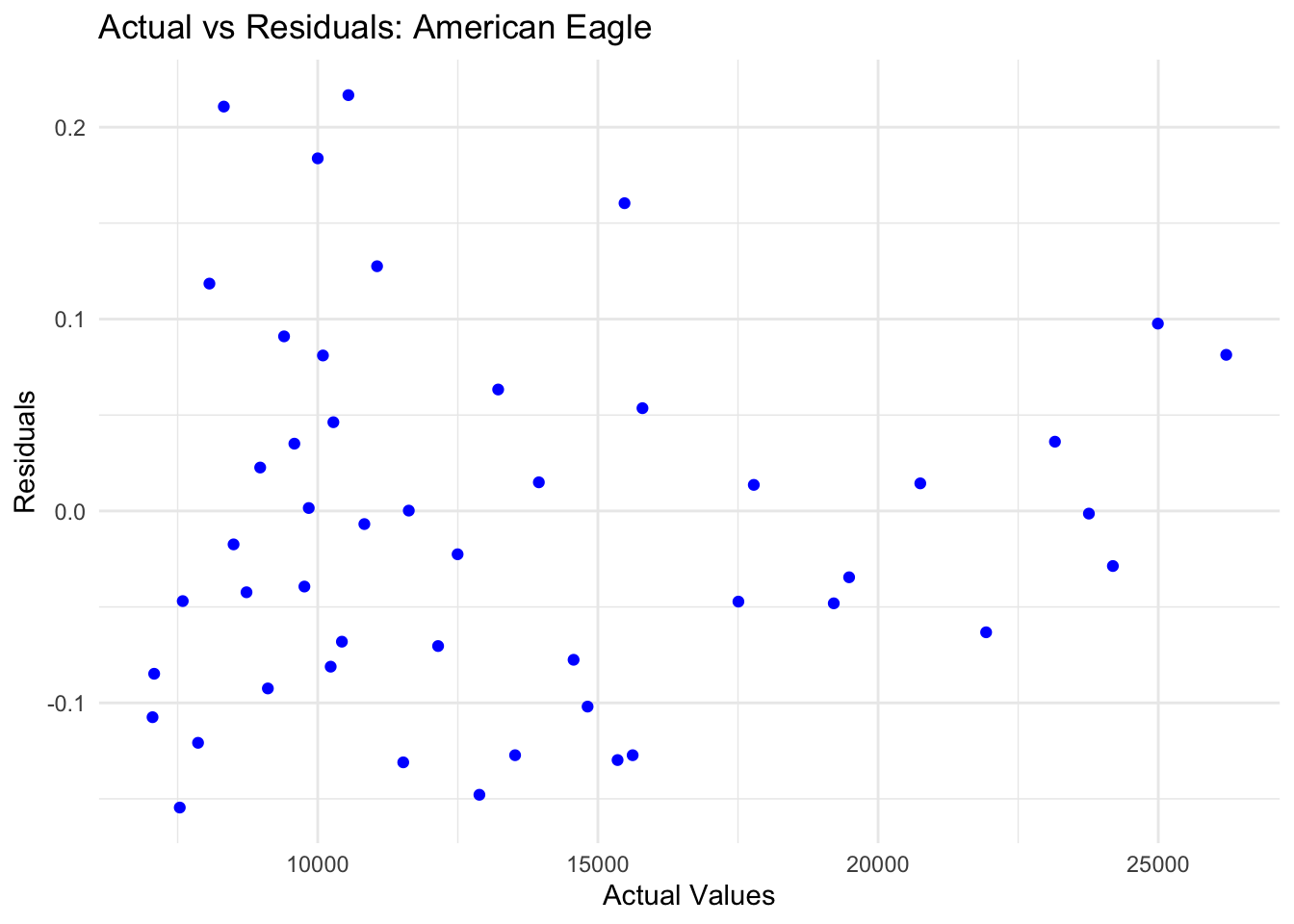
ggplot(data = NULL, aes(x = american\_window, y = residuals\_ets\_american)) +

geom\_point(color = "blue") +

labs(title = "Actual vs Residuals: American Eagle ",

x = "Actual Values", y = "Residuals") +

theme\_minimal()



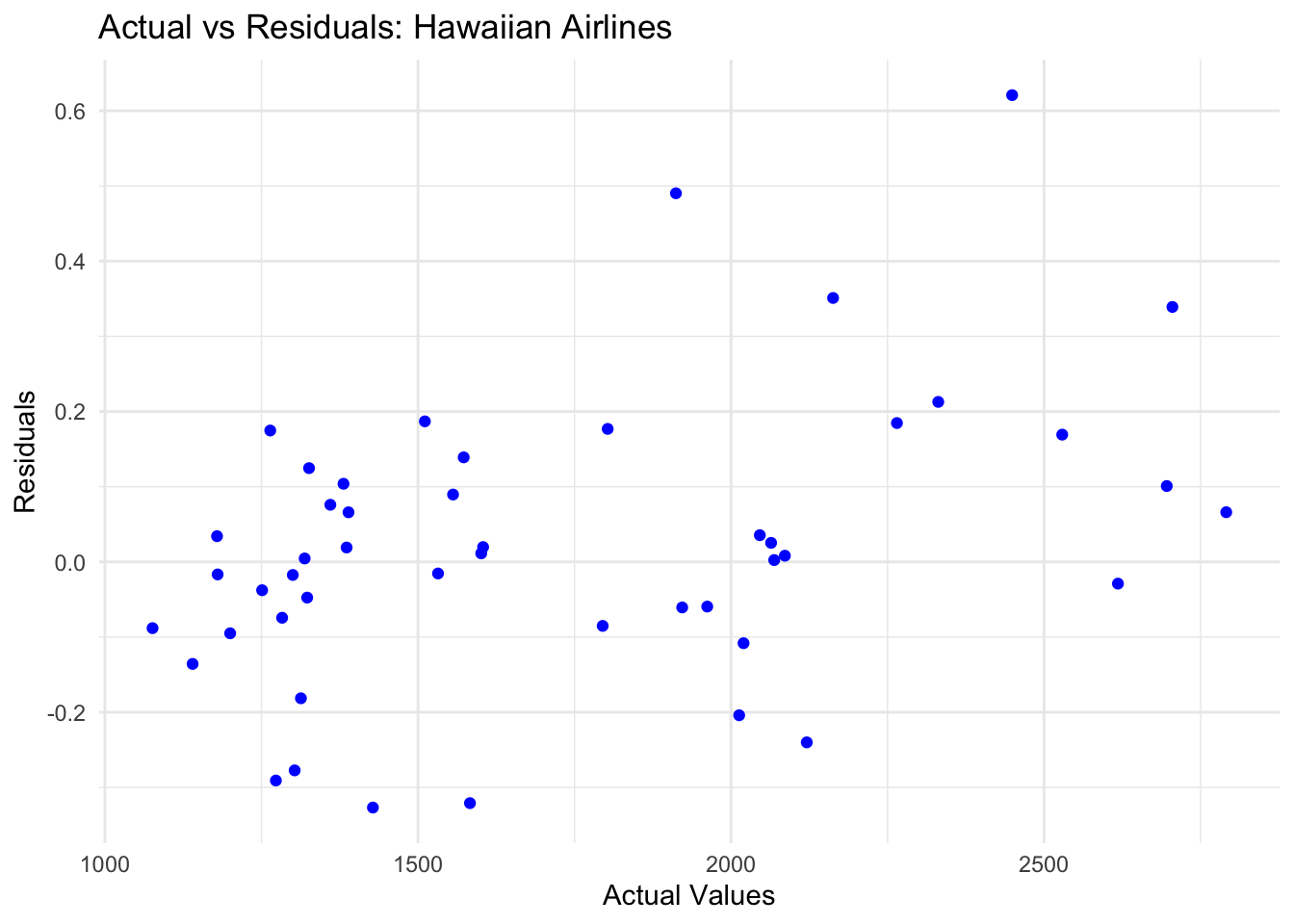
ggplot(data = NULL, aes(x = hawaiian\_window, y = residuals\_ets\_hawaiian)) +

geom\_point(color = "blue") +

labs(title = "Actual vs Residuals: Hawaiian Airlines",

x = "Actual Values", y = "Residuals") +

theme\_minimal()



* **United Airlines**: Residuals show slight patterns due to remaining seasonality.
* **American Eagle**: Residuals are randomly scattered, confirming good model performance.
* **Hawaiian Airlines**: Tight clustering near zero indicate no systematic error.

**5. Holt-Winters Method**

* Extends exponential smoothing by incorporating level, trend, and seasonality components.
* Suitable for time series with strong seasonal patterns.

hw\_united <- hw(united\_window, h = 12)

hw\_american <- hw(american\_window, h = 12)

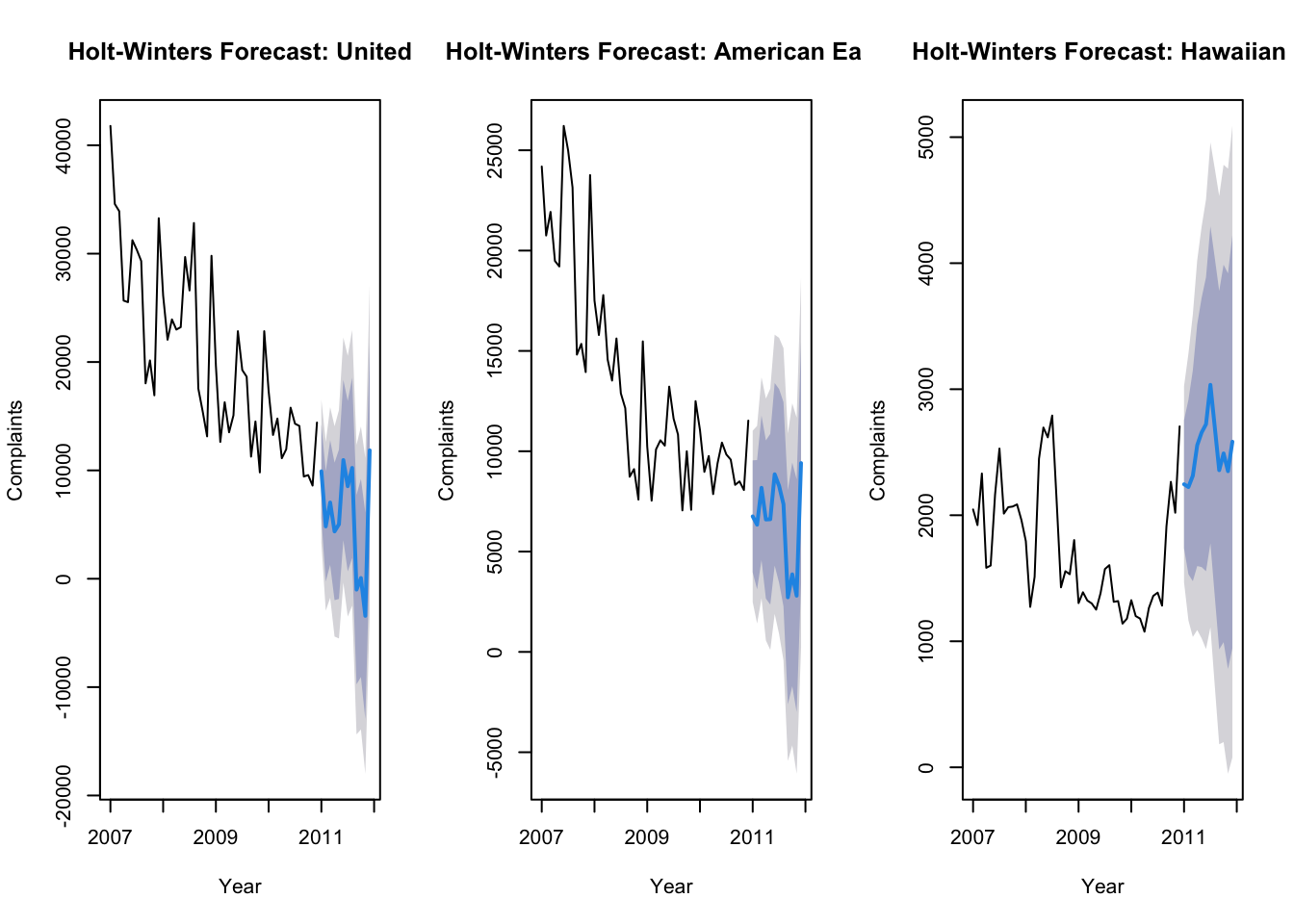
hw\_hawaiian <- hw(hawaiian\_window, h = 12)

par(mfrow = c(1, 3))

plot(hw\_united, main = "Holt-Winters Forecast: United", xlab = "Year", ylab = "Complaints")

plot(hw\_american, main = "Holt-Winters Forecast: American Eagle", xlab = "Year", ylab = "Complaints")

plot(hw\_hawaiian, main = "Holt-Winters Forecast: Hawaiian ", xlab = "Year", ylab = "Complaints")



**United Airlines**

* Complaints are expected to continue declining steadily.
* United Airlines is making progress in reducing complaints, but attention should still be given to peak complaint months to further improve customer satisfaction.

**American Eagle**

* Complaints are predicted to drop.
* Improvements are visible but focus should shift toward maintaining current levels of performance.

**Hawaiian Airlines**

* Complaints are expected to remain stable
* Hawaiian Airlines demonstrates a stable and reliable performance in handling complaints. Efforts should focus on maintaining this level of consistency.

summary(hw\_united)

##

## Forecast method: Holt-Winters' additive method

##

## Model Information:

## Holt-Winters' additive method

##

## Call:

## hw(y = united\_window, h = 12)

##

## Smoothing parameters:

## alpha = 0.621

## beta = 1e-04

## gamma = 1e-04

##

## Initial states:

## l = 32571.691

## b = -476.2188

## s = 8771.368 -6977.65 -3964.608 -5511.03 5243.736 3078.918

## 5014.77 -1386.432 -2520.716 -331.1463 -3010.162 1592.952

##

## sigma: 3370.045

##

## AIC AICc BIC

## 980.1327 1000.5327 1011.9431

##

## Error measures:

## ME RMSE MAE MPE MAPE MASE ACF1

## Training set -30.65225 2751.63 1960.97 0.253576 11.1306 0.3630865 0.08283619

##

## Forecasts:

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

## Jan 2011 9912.59367 5593.7077 14231.480 3307.4276 16517.76

## Feb 2011 4832.39596 -251.7909 9916.583 -2943.1968 12607.99

## Mar 2011 7034.73097 1286.0396 12783.422 -1757.1339 15826.60

## Apr 2011 4368.42332 -1975.7289 10712.576 -5334.1203 14070.97

## May 2011 5026.68049 -1861.8172 11915.178 -5508.3676 15561.73

## Jun 2011 10951.66895 3558.6417 18344.696 -354.9906 22258.33

## Jul 2011 8539.56203 674.1553 16404.969 -3489.5395 20568.66

## Aug 2011 10227.75699 1916.6372 18538.877 -2483.0039 22938.52

## Sep 2011 -1002.60384 -9736.8537 7731.646 -14360.4863 12355.28

## Oct 2011 67.02262 -9070.9121 9204.957 -13908.2425 14042.29

## Nov 2011 -3422.17875 -12946.8257 6102.468 -17988.8692 11144.51

## Dec 2011 11849.34916 1952.9726 21745.726 -3285.8528 26984.55

summary(hw\_american)

##

## Forecast method: Holt-Winters' additive method

##

## Model Information:

## Holt-Winters' additive method

##

## Call:

## hw(y = american\_window, h = 12)

##

## Smoothing parameters:

## alpha = 0.5743

## beta = 3e-04

## gamma = 0.0321

##

## Initial states:

## l = 24321.1432

## b = -270.2226

## s = 4446.341 -2396.847 -1678.507 -3156.506 1301.747 1997.55

## 2099.493 -264.0999 -509.9658 835.7922 -1184.309 -1490.689

##

## sigma: 2178.696

##

## AIC AICc BIC

## 938.2576 958.6576 970.0680

##

## Error measures:

## ME RMSE MAE MPE MAPE MASE

## Training set -108.7539 1778.898 1433.417 -0.5826443 11.43462 0.3551579

## ACF1

## Training set 0.1264468

##

## Forecasts:

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

## Jan 2011 6754.584 3962.473 9546.695 2484.4187 11024.75

## Feb 2011 6340.152 3119.911 9560.393 1415.2180 11265.09

## Mar 2011 8176.599 4578.487 11774.711 2673.7618 13679.44

## Apr 2011 6599.022 2658.811 10539.232 572.9900 12625.05

## May 2011 6611.690 2356.517 10866.863 103.9644 13119.42

## Jun 2011 8851.045 4302.405 13399.685 1894.5003 15807.59

## Jul 2011 8274.570 3450.033 13099.106 896.0781 15653.06

## Aug 2011 7362.390 2276.668 12448.112 -415.5504 15140.33

## Sep 2011 2721.340 -2613.019 8055.698 -5436.8581 10879.54

## Oct 2011 3863.880 -1708.247 9436.006 -4657.9525 12385.71

## Nov 2011 2801.717 -2998.645 8602.079 -6069.1715 11672.61

## Dec 2011 9409.915 3389.765 15430.065 202.8896 18616.94

summary(hw\_hawaiian)

##

## Forecast method: Holt-Winters' additive method

##

## Model Information:

## Holt-Winters' additive method

##

## Call:

## hw(y = hawaiian\_window, h = 12)

##

## Smoothing parameters:

## alpha = 0.9226

## beta = 1e-04

## gamma = 1e-04

##

## Initial states:

## l = 2119.2169

## b = -7.342

## s = 104.2992 -136.8299 -3.4078 -143.5896 188.3205 519.1114

## 199.258 126.5693 18.7822 -231.2163 -328.0321 -313.2649

##

## sigma: 396.884

##

## AIC AICc BIC

## 774.7851 795.1851 806.5956

##

## Error measures:

## ME RMSE MAE MPE MAPE MASE

## Training set 18.08109 324.0544 242.455 -0.6423617 14.38428 0.4553146

## ACF1

## Training set 0.008432555

##

## Forecasts:

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

## Jan 2011 2245.877 1737.2494 2754.504 1467.99837 3023.755

## Feb 2011 2223.701 1531.6485 2915.753 1165.29825 3282.103

## Mar 2011 2313.332 1477.1393 3149.525 1034.48562 3592.179

## Apr 2011 2555.966 1597.0352 3514.897 1089.40792 4022.524

## May 2011 2656.584 1588.9103 3724.257 1023.71826 4289.449

## Jun 2011 2722.070 1555.7287 3888.411 938.30502 4505.835

## Jul 2011 3034.575 1777.2662 4291.883 1111.68719 4957.462

## Aug 2011 2696.582 1354.4399 4038.725 643.95265 4749.212

## Sep 2011 2357.520 935.5796 3779.461 182.84972 4532.191

## Oct 2011 2490.342 992.8333 3987.850 200.10017 4780.584

## Nov 2011 2349.674 780.2168 3919.132 -50.60377 4749.952

## Dec 2011 2583.529 945.2646 4221.793 78.02005 5089.037

**United Airlines**

* Alpha = 0.621, The model reacts moderately to changes in the level, giving a fair balance to recent and historical data.
* Beta = 0.0001, The trend changes very slowly.
* Gamma = 0.0001, Seasonal changes are minimal and evolve very slowly.
* Focus on high-complaint months to further improve customer satisfaction.

**American Eagle**

* Alpha = 0.5743, the model reacts moderately to level changes.
* Beta = 0.0003, The trend changes slightly faster than United
* Gamma = 0.0321, Seasonal adjustments are slightly more compared to United Airlines.
* Monitor months with seasonal peaks and optimize operations during these periods.

**Hawaiian Airlines**

* Alpha = 0.9226, the model heavily relies on recent data,
* Beta = 0.0001, Negligible trend, as complaints remain stable.
* Gamma = 0.0001, Seasonal adjustments are minimal
* Maintain consistency in operations to keep complaints stable.
* Address occasional spikes in complaints.

#### **Residual Analysis**

residuals\_hw\_united <- residuals(hw\_united)

fitted\_hw\_united <- fitted(hw\_united)

residuals\_hw\_american <- residuals(hw\_american)

fitted\_hw\_american <- fitted(hw\_american)

residuals\_hw\_hawaiian <- residuals(hw\_hawaiian)

fitted\_hw\_hawaiian <- fitted(hw\_hawaiian)

par(mfrow = c(1, 3))

plot(residuals\_hw\_united, main = "United Airlines (/hw)",

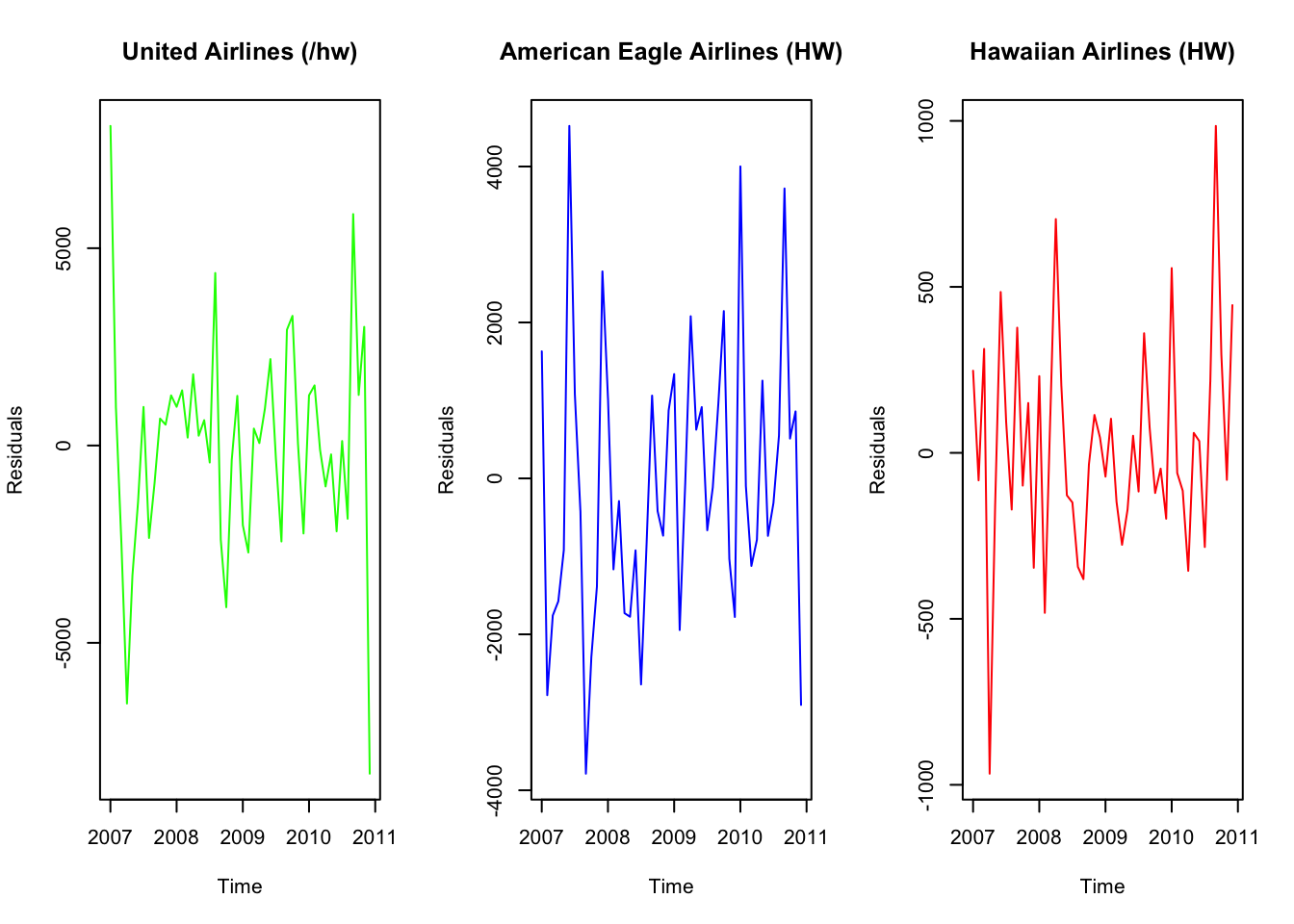
ylab = "Residuals", xlab = "Time",col ='green')

plot(residuals\_hw\_american, main = "American Eagle Airlines (HW)",

ylab = "Residuals", xlab = "Time", col ='blue')

plot(residuals\_hw\_hawaiian, main = "Hawaiian Airlines (HW)",

ylab = "Residuals", xlab = "Time", col= 'red')

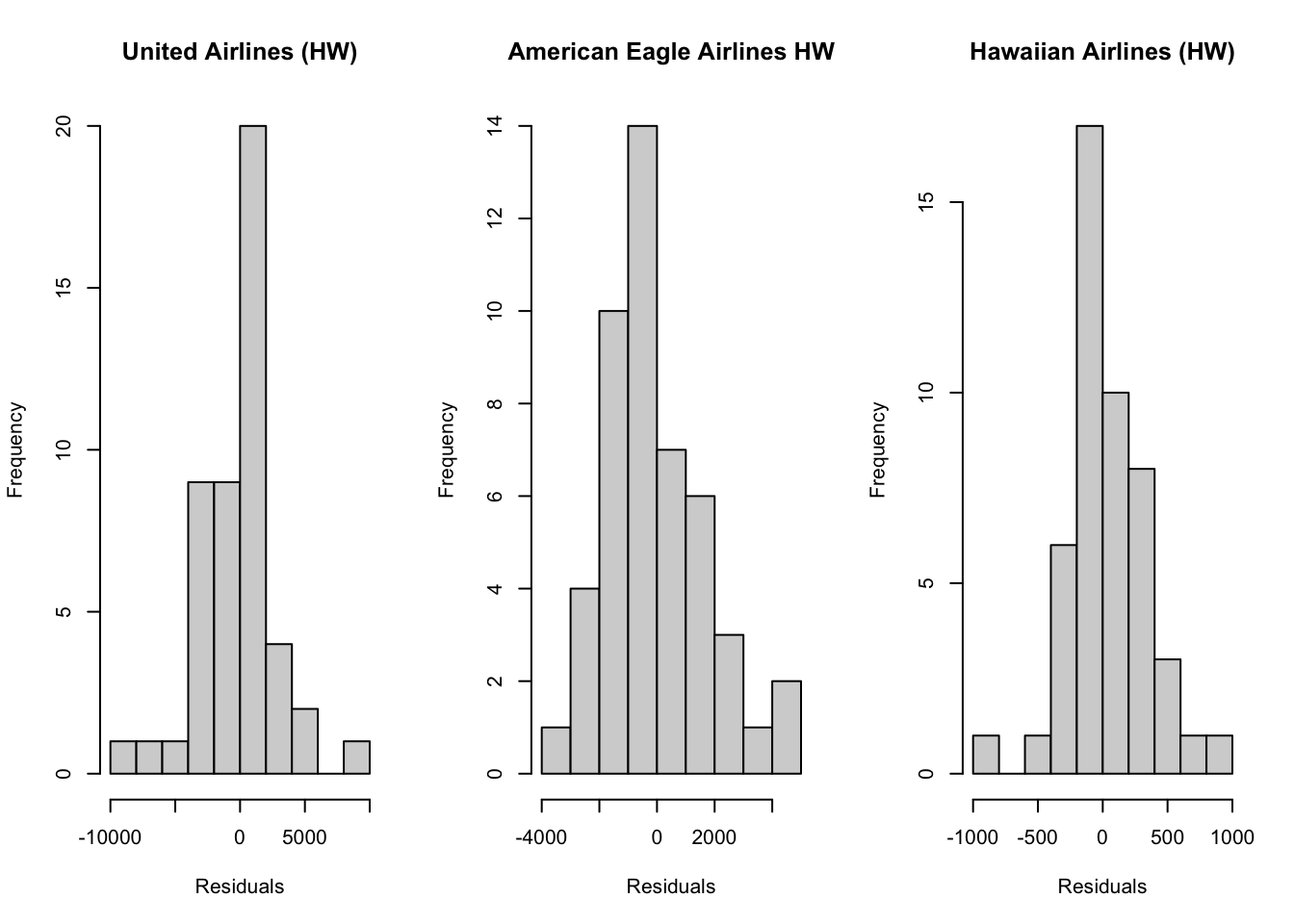


* **United Airlines**: Residuals fluctuate around zero, indicating the model captures the main trend but may leave some seasonality unexplained.
* **American Eagle**: Residuals are mostly random with small deviations, indicating a good fit.
* **Hawaiian Airlines**: Residuals are tightly clustered around zero, reflecting the stability of the data and a good model fit.

hist(residuals\_hw\_united, main = "United Airlines (HW)", xlab = "Residuals")

hist(residuals\_hw\_american, main = "American Eagle Airlines HW", xlab = "Residuals")

hist(residuals\_hw\_hawaiian, main = "Hawaiian Airlines (HW)", xlab = "Residuals")



* + **United Airlines**: Residuals show slight skewness, possibly indicating remaining patterns.
  + **American Eagle**: Residuals are more symmetrical, supporting the model's validity.
  + **Hawaiian Airlines**: Residuals are normally distributed, aligning with the stable data.

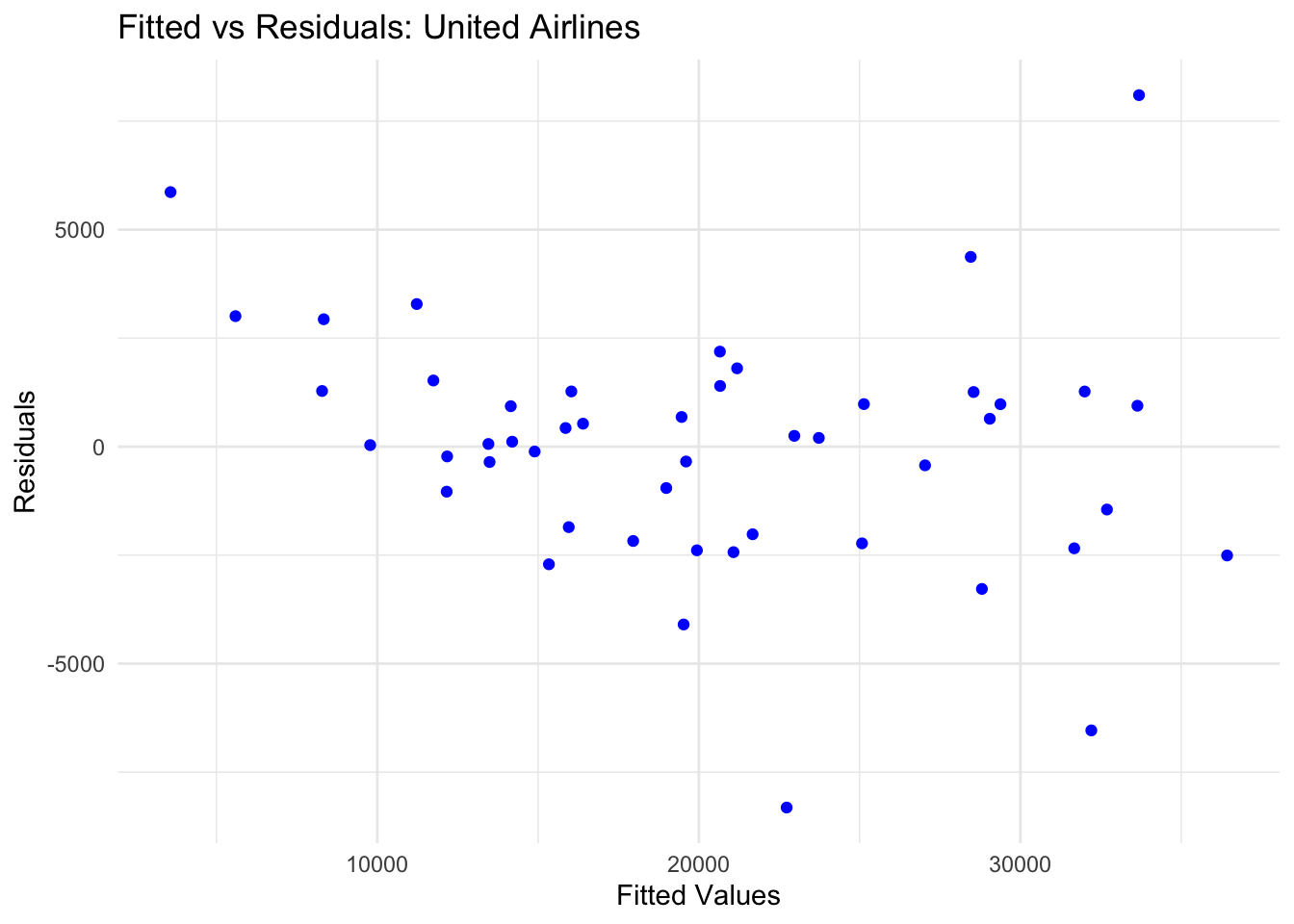
ggplot(data = NULL, aes(x = fitted\_hw\_united, y = residuals\_hw\_united)) +

geom\_point(color = "blue") +

labs(title = "Fitted vs Residuals: United Airlines",

x = "Fitted Values", y = "Residuals") +

theme\_minimal()



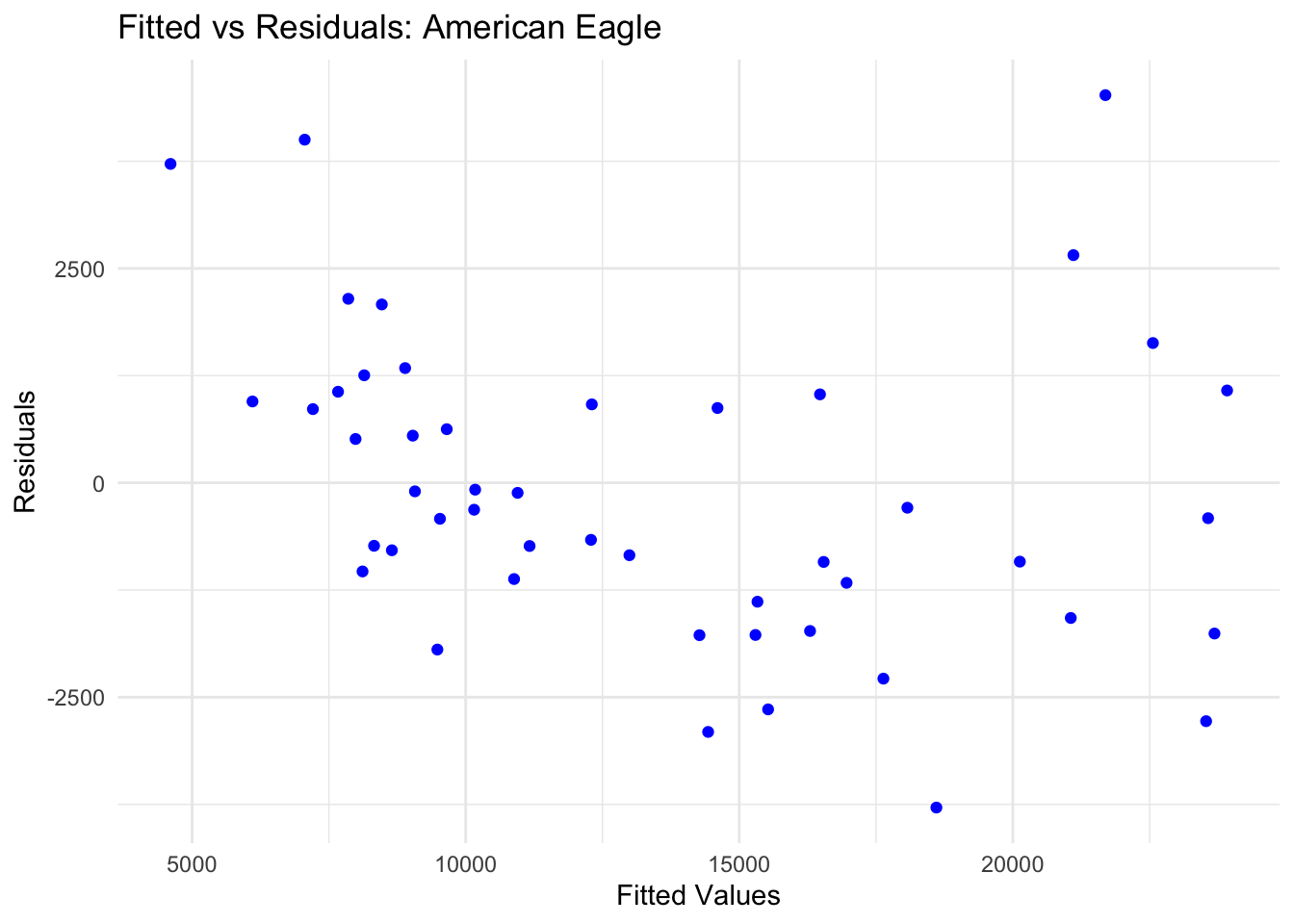
ggplot(data = NULL, aes(x = fitted\_hw\_american, y = residuals\_hw\_american)) +

geom\_point(color = "blue") +

labs(title = "Fitted vs Residuals: American Eagle ",

x = "Fitted Values", y = "Residuals") +

theme\_minimal()



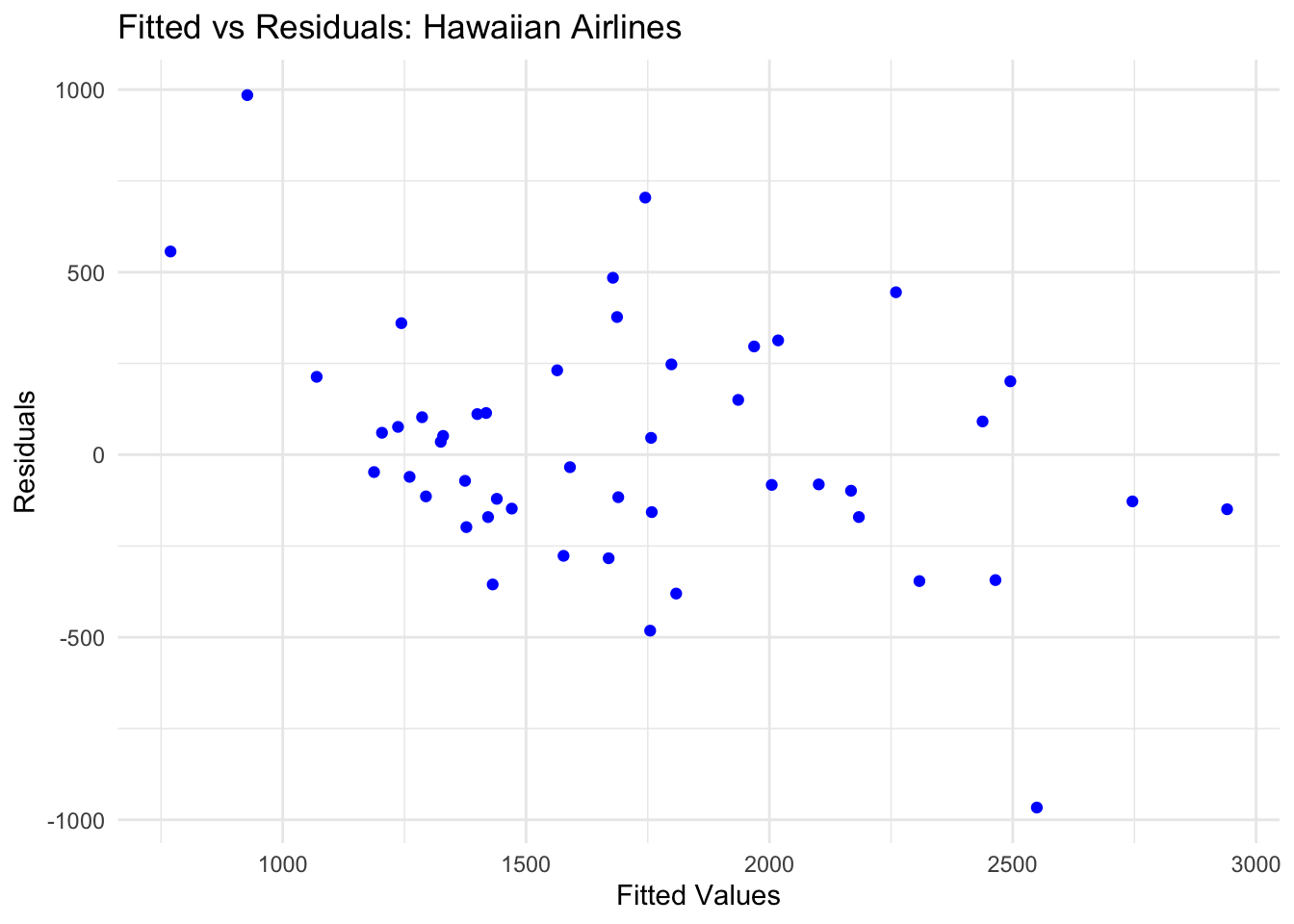
ggplot(data = NULL, aes(x = fitted\_hw\_hawaiian, y = residuals\_hw\_hawaiian)) +

geom\_point(color = "blue") +

labs(title = "Fitted vs Residuals: Hawaiian Airlines",

x = "Fitted Values", y = "Residuals") +

theme\_minimal()



## Actual Vs Residual

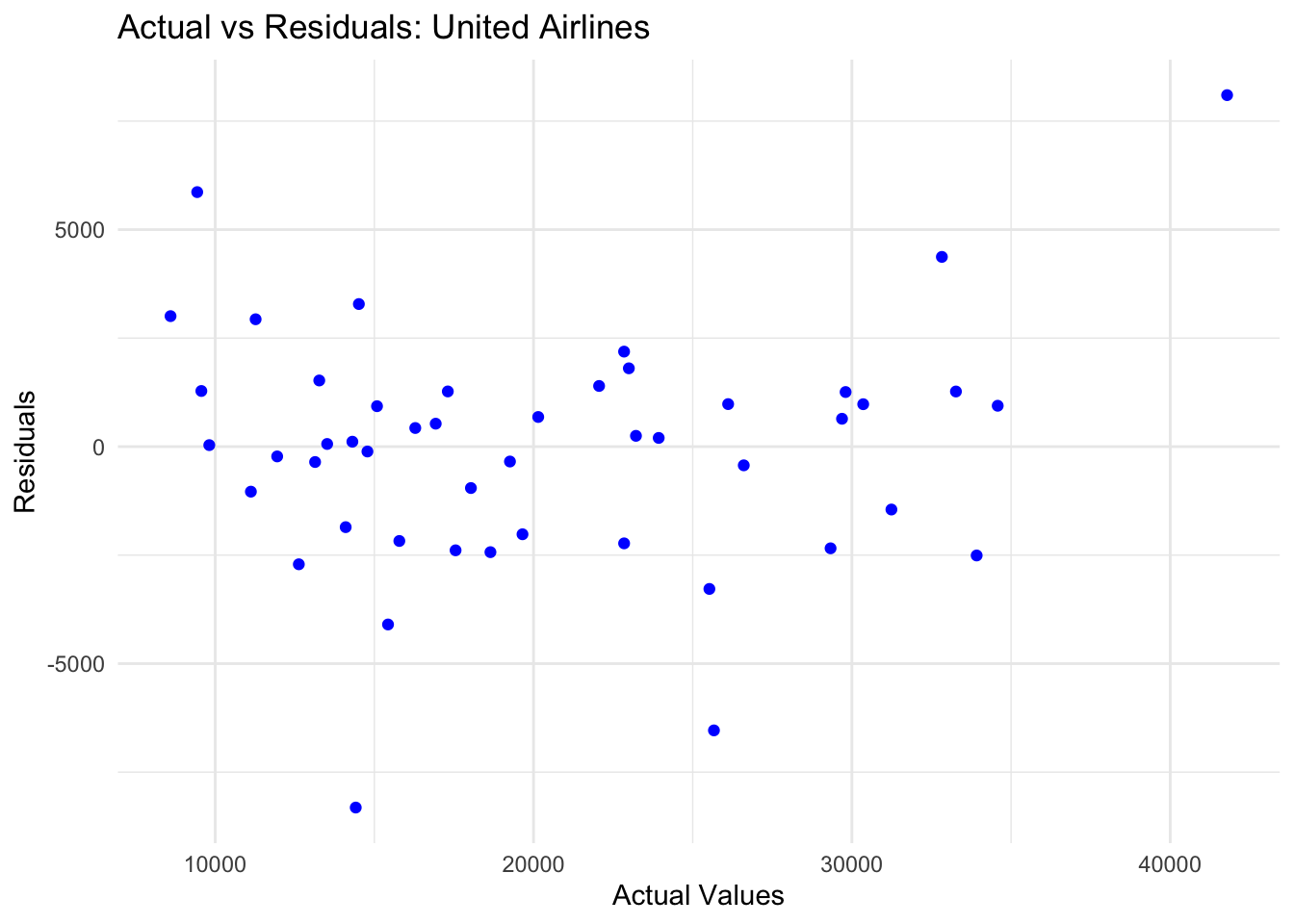
ggplot(data = NULL, aes(x = united\_window, y = residuals\_hw\_united)) +

geom\_point(color = "blue") +

labs(title = "Actual vs Residuals: United Airlines",

x = "Actual Values", y = "Residuals") +

theme\_minimal()



ggplot(data = NULL, aes(x = american\_window, y = residuals\_hw\_american)) +

geom\_point(color = "blue") +

labs(title = "Actual vs Residuals: American Eagle ",

x = "Actual Values", y = "Residuals") +

theme\_minimal()



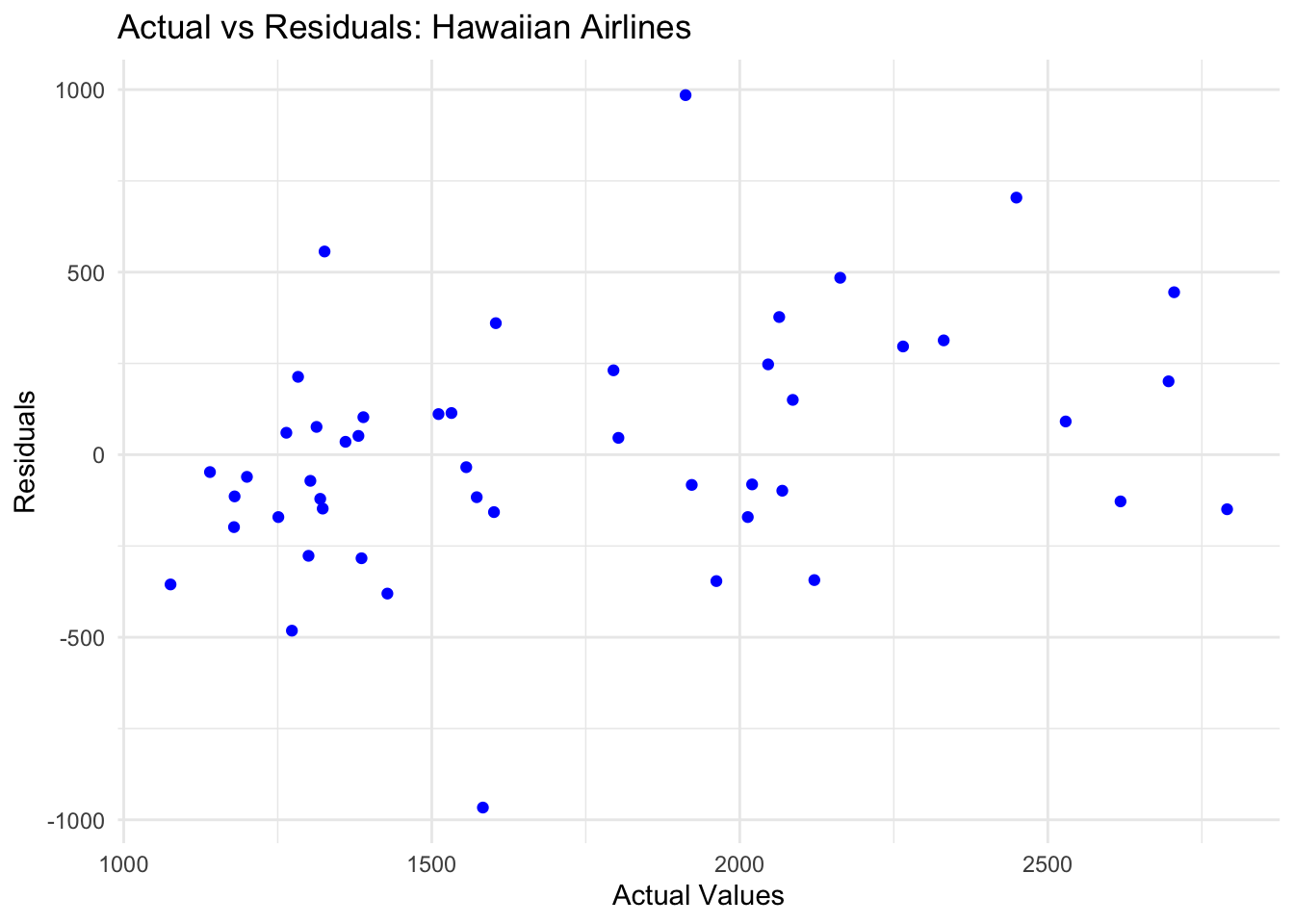
ggplot(data = NULL, aes(x = hawaiian\_window, y = residuals\_hw\_hawaiian)) +

geom\_point(color = "blue") +

labs(title = "Actual vs Residuals: Hawaiian Airlines",

x = "Actual Values", y = "Residuals") +

theme\_minimal()



* **United Airlines**: Residuals show a slight pattern, suggesting potential underfitting in capturing all seasonality.
* **American Eagle**: Residuals are randomly distributed, indicating a strong model.
* **Hawaiian Airlines**: Tight clustering around zero, reflecting the model's ability to fit the stable data.

#### **Accuracy**

accuracy\_naive\_us <- accuracy(united\_naive)

accuracy\_naive\_am <- accuracy(american\_naive)

accuracy\_naive\_ha <- accuracy(hawaiian\_naive)

accuracy\_avg\_us <- accuracy(united\_sa)

accuracy\_avg\_am <- accuracy(american\_sa)

accuracy\_avg\_ha <- accuracy(hawaiian\_sa)

accuracy\_ets\_us <- accuracy(ets\_united)

accuracy\_ets\_am <- accuracy(ets\_american)

accuracy\_ets\_ha <- accuracy(ets\_hawaiian)

accuracy\_hw\_us <- accuracy(hw\_united)

accuracy\_hw\_am <- accuracy(hw\_american)

accuracy\_hw\_ha <- accuracy(hw\_hawaiian)

# Create accuracy data frame for all measures (ME, RMSE, MAE, MPE, MAPE, MASE)

accuracy\_data <- data.frame(

Airline = rep(c("United", "American Eagle", "Hawaiian"), each = 4), # Repeat for each airline

Model = rep(c("Naive", "Average", "ETS", "Holt-Winters"), 3), # Repeat for each model

ME = c(

accuracy\_naive\_us["Training set", "ME"], accuracy\_avg\_us["Training set", "ME"], accuracy\_ets\_us["Training set", "ME"], accuracy\_hw\_us["Training set", "ME"],

accuracy\_naive\_am["Training set", "ME"], accuracy\_avg\_am["Training set", "ME"], accuracy\_ets\_am["Training set", "ME"], accuracy\_hw\_am["Training set", "ME"],

accuracy\_naive\_ha["Training set", "ME"], accuracy\_avg\_ha["Training set", "ME"], accuracy\_ets\_ha["Training set", "ME"], accuracy\_hw\_ha["Training set", "ME"]

),

RMSE = c(

accuracy\_naive\_us["Training set", "RMSE"], accuracy\_avg\_us["Training set", "RMSE"], accuracy\_ets\_us["Training set", "RMSE"], accuracy\_hw\_us["Training set", "RMSE"],

accuracy\_naive\_am["Training set", "RMSE"], accuracy\_avg\_am["Training set", "RMSE"], accuracy\_ets\_am["Training set", "RMSE"], accuracy\_hw\_am["Training set", "RMSE"],

accuracy\_naive\_ha["Training set", "RMSE"], accuracy\_avg\_ha["Training set", "RMSE"], accuracy\_ets\_ha["Training set", "RMSE"], accuracy\_hw\_ha["Training set", "RMSE"]

),

MAE = c(

accuracy\_naive\_us["Training set", "MAE"], accuracy\_avg\_us["Training set", "MAE"], accuracy\_ets\_us["Training set", "MAE"], accuracy\_hw\_us["Training set", "MAE"],

accuracy\_naive\_am["Training set", "MAE"], accuracy\_avg\_am["Training set", "MAE"], accuracy\_ets\_am["Training set", "MAE"], accuracy\_hw\_am["Training set", "MAE"],

accuracy\_naive\_ha["Training set", "MAE"], accuracy\_avg\_ha["Training set", "MAE"], accuracy\_ets\_ha["Training set", "MAE"], accuracy\_hw\_ha["Training set", "MAE"]

),

MPE = c(

accuracy\_naive\_us["Training set", "MPE"], accuracy\_avg\_us["Training set", "MPE"], accuracy\_ets\_us["Training set", "MPE"], accuracy\_hw\_us["Training set", "MPE"],

accuracy\_naive\_am["Training set", "MPE"], accuracy\_avg\_am["Training set", "MPE"], accuracy\_ets\_am["Training set", "MPE"], accuracy\_hw\_am["Training set", "MPE"],

accuracy\_naive\_ha["Training set", "MPE"], accuracy\_avg\_ha["Training set", "MPE"], accuracy\_ets\_ha["Training set", "MPE"], accuracy\_hw\_ha["Training set", "MPE"]

),

MAPE = c(

accuracy\_naive\_us["Training set", "MAPE"], accuracy\_avg\_us["Training set", "MAPE"], accuracy\_ets\_us["Training set", "MAPE"], accuracy\_hw\_us["Training set", "MAPE"],

accuracy\_naive\_am["Training set", "MAPE"], accuracy\_avg\_am["Training set", "MAPE"], accuracy\_ets\_am["Training set", "MAPE"], accuracy\_hw\_am["Training set", "MAPE"],

accuracy\_naive\_ha["Training set", "MAPE"], accuracy\_avg\_ha["Training set", "MAPE"], accuracy\_ets\_ha["Training set", "MAPE"], accuracy\_hw\_ha["Training set", "MAPE"]

),

MASE = c(

accuracy\_naive\_us["Training set", "MASE"], accuracy\_avg\_us["Training set", "MASE"], accuracy\_ets\_us["Training set", "MASE"], accuracy\_hw\_us["Training set", "MASE"],

accuracy\_naive\_am["Training set", "MASE"], accuracy\_avg\_am["Training set", "MASE"], accuracy\_ets\_am["Training set", "MASE"], accuracy\_hw\_am["Training set", "MASE"],

accuracy\_naive\_ha["Training set", "MASE"], accuracy\_avg\_ha["Training set", "MASE"], accuracy\_ets\_ha["Training set", "MASE"], accuracy\_hw\_ha["Training set", "MASE"]

)

)

# Display the accuracy data frame

print(accuracy\_data)

# Optional: Rank models based on each metric

accuracy\_table <- accuracy\_data %>%

group\_by(Airline) %>%

mutate(

ME\_Rank = rank(ME),

RMSE\_Rank = rank(RMSE),

MAE\_Rank = rank(MAE),

MPE\_Rank = rank(MPE),

MAPE\_Rank = rank(MAPE),

MASE\_Rank = rank(MASE)

)

# Display the ranked table

print(accuracy\_table)

## # A tibble: 12 × 4

## # Groups: Airline [3]

## Airline Model MAPE Model\_Rank

## <chr> <chr> <dbl> <dbl>

## 1 United Naive 24.8 3

## 2 United Average 38.7 4

## 3 United ETS 7.81 1

## 4 United Holt-Winters 11.1 2

## 5 American Eagle Naive 19.5 3

## 6 American Eagle Average 34.7 4

## 7 American Eagle ETS 7.77 1

## 8 American Eagle Holt-Winters 11.4 2

## 9 Hawaiian Naive 13.8 2

## 10 Hawaiian Average 25.1 4

## 11 Hawaiian ETS 13.6 1

## 12 Hawaiian Holt-Winters 14.4 3

**ETS** is consistently the best-performing model across all three airlines.

**Naïve Model** is not suitable for this time series data

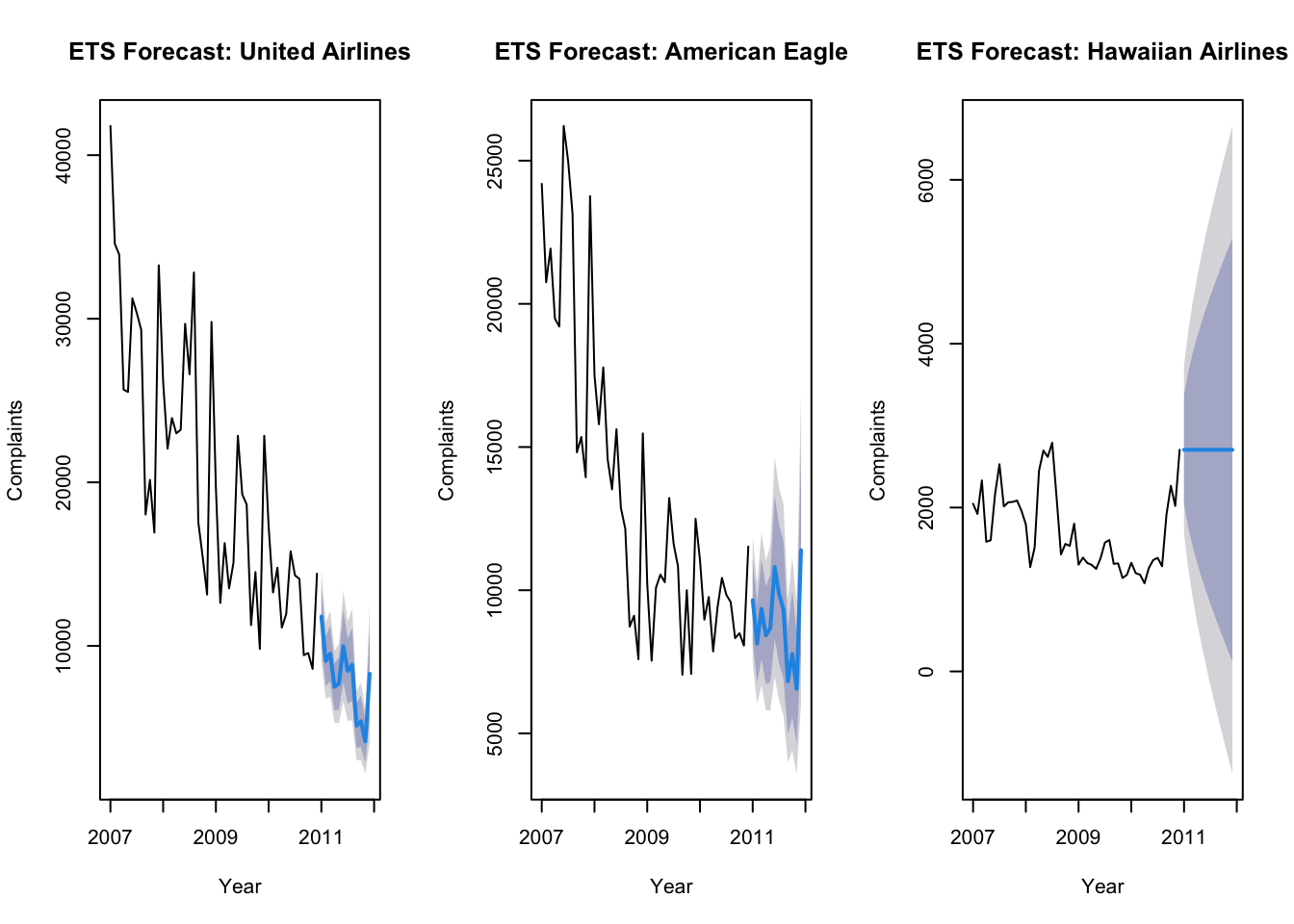
#### **Forecast of baggage complaints using ETS model**

par(mfrow = c(1, 3))

plot(forecast(ets\_united, h = 12), main = "ETS Forecast: United Airlines", xlab = "Year", ylab = "Complaints")

plot(forecast(ets\_american, h = 12), main = "ETS Forecast: American Eagle", xlab = "Year", ylab = "Complaints")x

plot(forecast(ets\_hawaiian, h = 12), main = "ETS Forecast: Hawaiian Airlines", xlab = "Year", ylab = "Complaints")



#### **Conclusion**

**United Airlines**

* The forecast suggests that baggage complaints will generally go down over time, but there will still be some increases during busy months, like holidays or peak travel seasons.
* Complaints are expected to be highest in December and July, when many people are traveling, and lowest in February and September, which are quieter months.
* The predictions are reliable, with only a small chance of error, so United can use them confidently for planning.

**American Eagle**

* Complaints are expected to level out over time, with a slight decrease, showing that American Eagle is making improvements. However, there may still be occasional spikes in complaints during certain months.
* The seasonal pattern isn’t as strong as United’s, but there could still be a small increase in complaints during summer months.
* The forecast is moderately reliable, so while the predictions are helpful, there’s still a chance of unexpected changes.

**Hawaiian Airlines**

* Complaints are predicted to stay stable over the next year, around 2,000 per month, with minimal seasonal changes.
* Hawaiian has fewer fluctuations compared to United and American Eagle, so there’s less risk of unexpected complaint spikes.
* The forecast is less precise because the data is very stable and doesn’t have a clear trend or strong seasonality, but overall, Hawaiian Airlines seems consistent.

#### **Decision Based on Analysis**

1. **United Airlines**: Focus on Peak Travel Seasons: United should prioritize improving baggage handling during high-complaint months like December and July. The seasonal pattern is strong, so targeted measures during peak periods will have the most impact on reducing complaints.
2. **American Eagle**: Sustain Gradual Improvements: With complaints stabilizing and a slight downward trend, American Eagle should continue its improvement strategies while preparing for occasional spikes, especially in summer months.
3. **Hawaiian Airlines**: Maintain Current Operations: Hawaiian Airlines demonstrates consistent performance with minimal seasonal changes. The focus should remain on sustaining this stability while looking for opportunities to further optimize operations.

**9. Suggestions for Improvement**

1. **Data Enrichment**:
   * Include external factors like weather disruptions, economic conditions, and staffing levels.
2. **Validation**:
   * Validate forecasts with data from different airlines or time periods.
3. **Operational Insights**:
   * Use predictions to adjust baggage handling protocols and staffing strategies proactively.