

# AirlineDelayAnalysis

2023-05-25

## Loading necessary libraries

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.2.3
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##      filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

## Loading data and Summarization for Airlines Delay

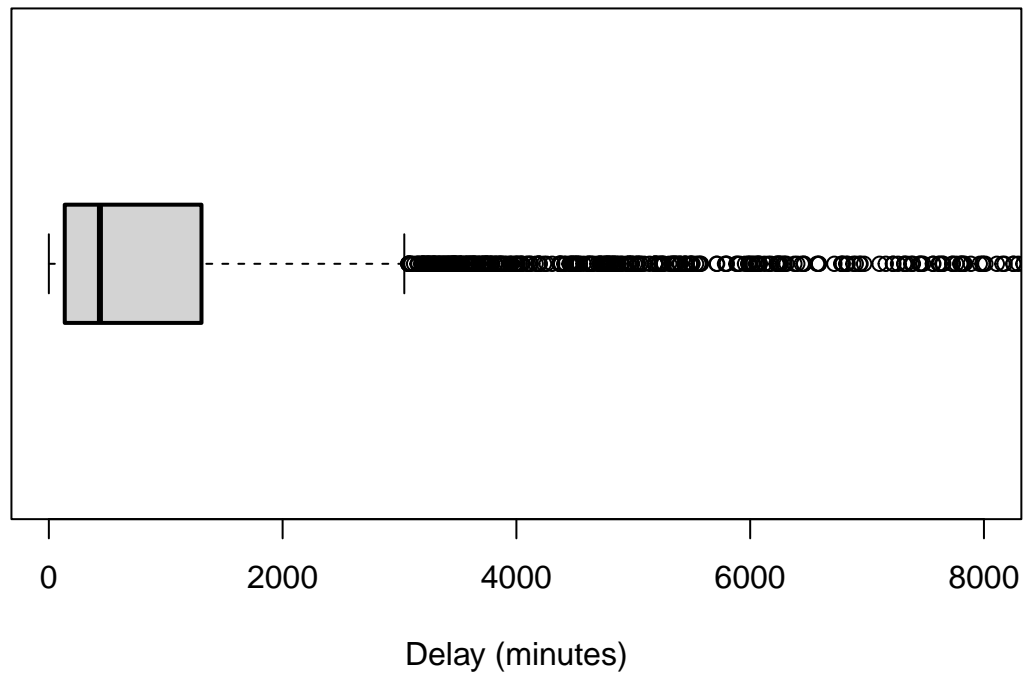
```
airlines_data <- read.csv("Airline_Delay_Cause.csv")
```

## Box Plots

```
# Box Plot - carrier_delay
```

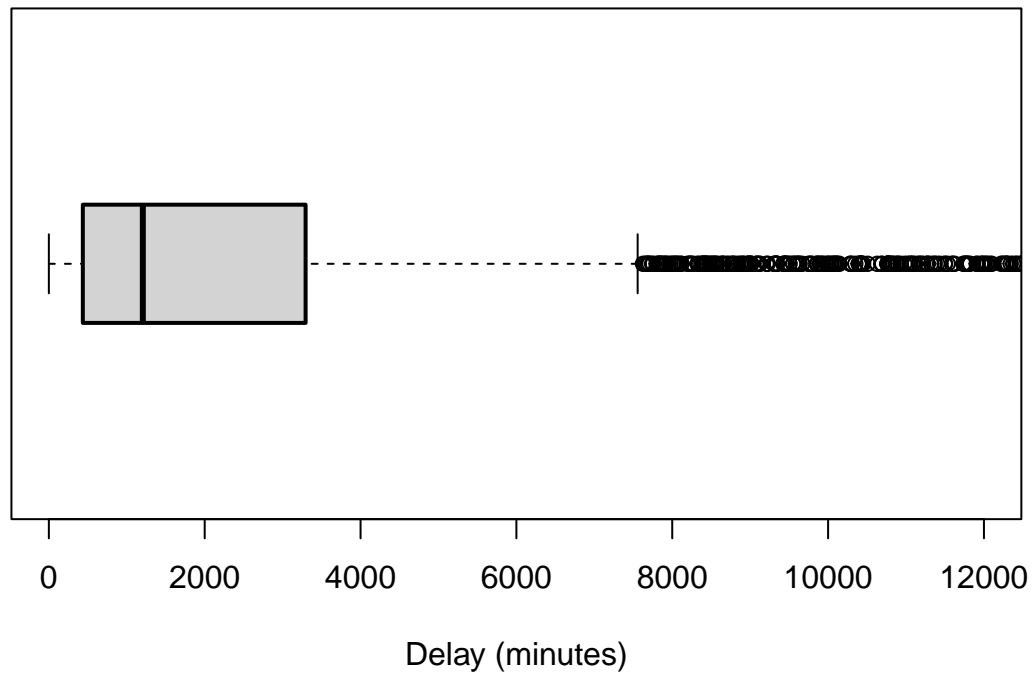
```
boxplot(airlines_data$carrier_delay, main = "Distribution of Carrier Delays", xlab = "Delay (minutes)"  
        boxlwd = 2, outwex = 0.5, boxwex = 0.5 , outline = TRUE , horizontal = TRUE ,  
        ylim=c(0,8000))
```

## Distribution of Carrier Delays



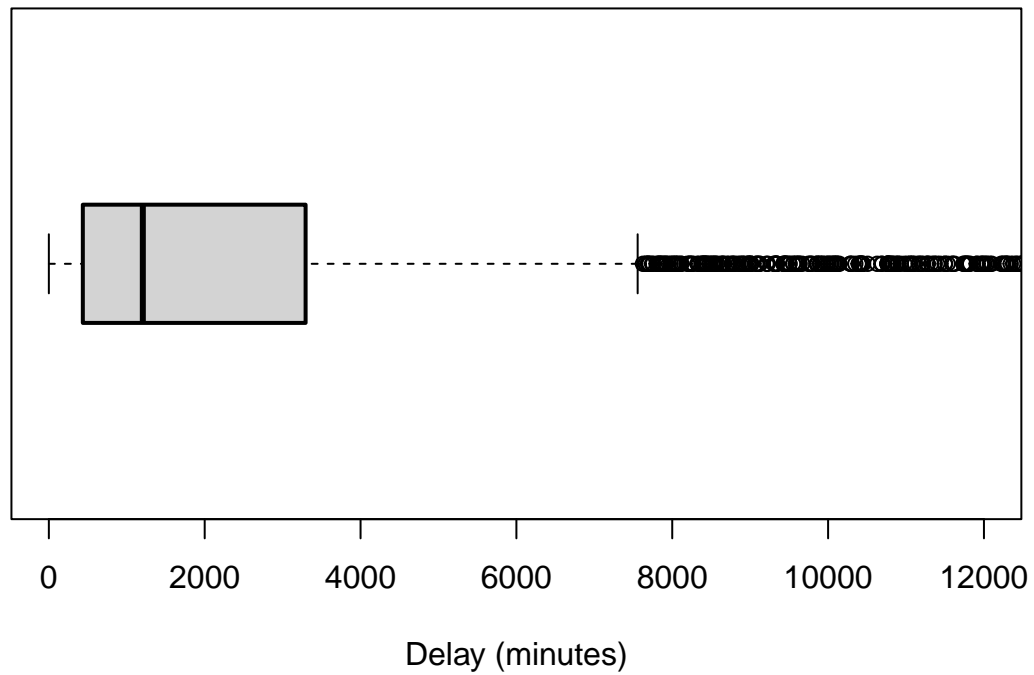
```
# Box Plot - arr_delay
boxplot(airlines_data$arr_delay, main = "Distribution of Arrival Delays", xlab = "Delay (minutes)" ,
        boxlwd = 2, outwex = 0.5, boxwex = 0.5 , outline = TRUE , horizontal = TRUE ,
        ylim=c(0,12000))
```

## Distribution of Arrival Delays



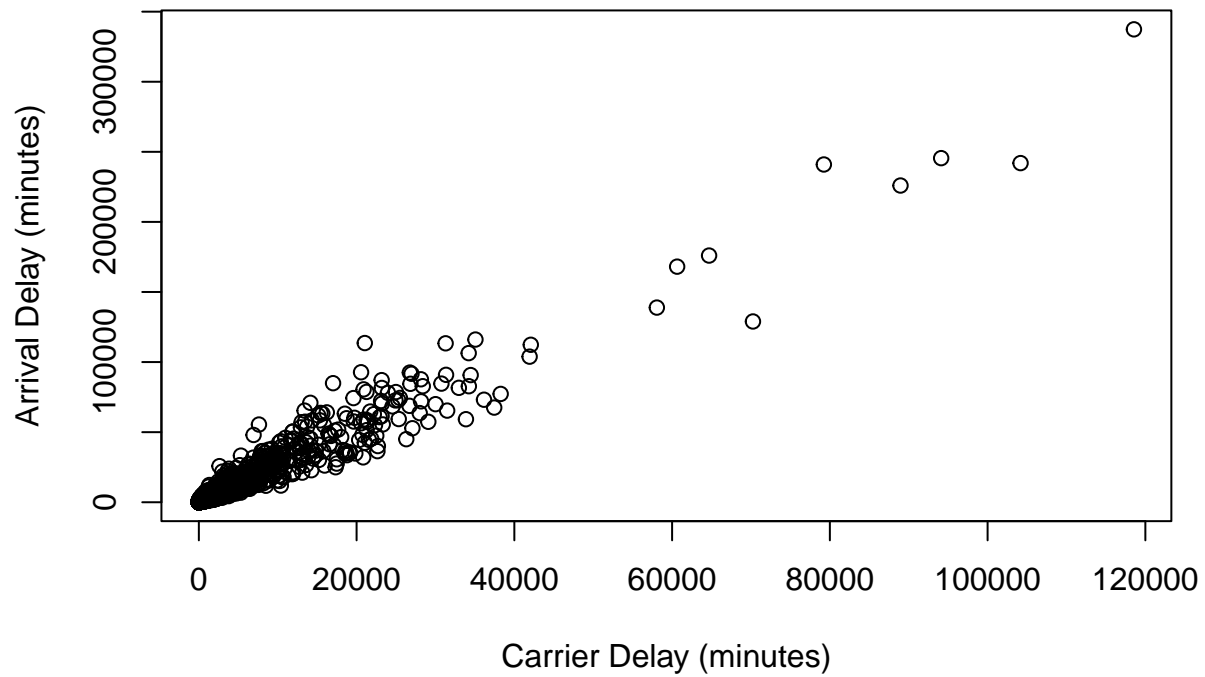
```
# Box Plot - arr_delay
boxplot(airlines_data$arr_delay, main = "Distribution of Weather Delays", xlab = "Delay (minutes)" ,
        boxlwd = 2, outwex = 0.5, boxwex = 0.5 , outline = TRUE , horizontal = TRUE ,
        ylim=c(0,12000))
```

## Distribution of Weather Delays



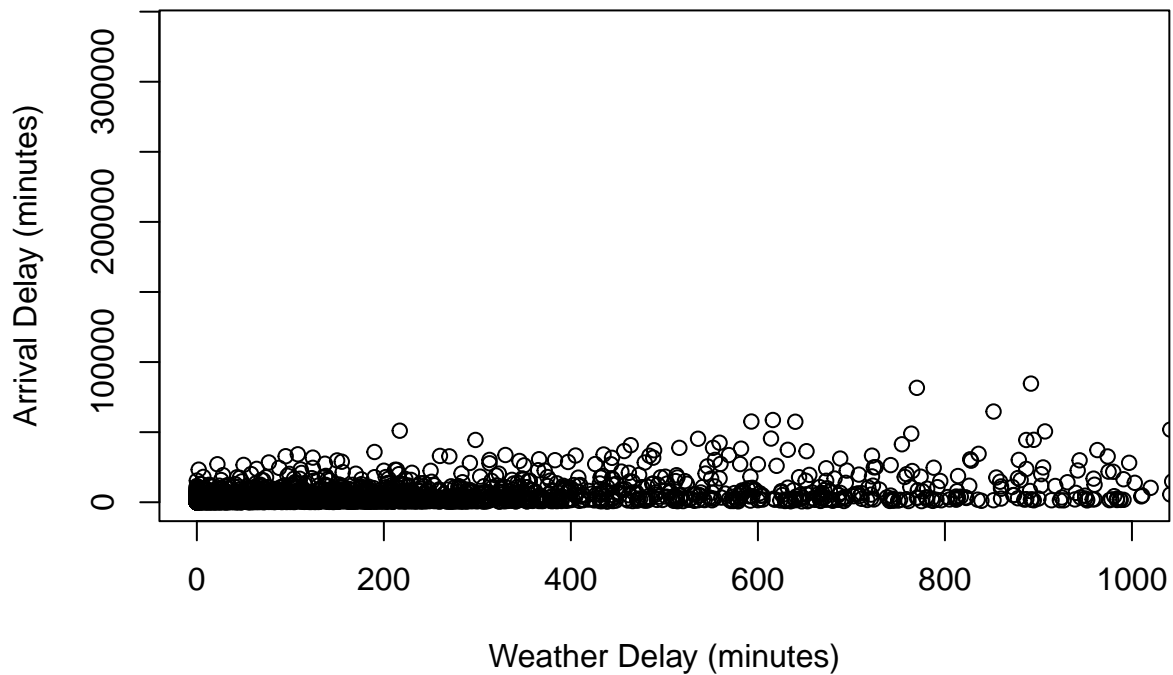
```
# Scatter Plot
plot(airlines_data$carrier_delay, airlines_data$arr_delay,
     main = "Arrival Delay vs Carrier Delay",
     xlab = "Carrier Delay (minutes)", ylab = "Arrival Delay (minutes)")
```

## Arrival Delay vs Carrier Delay



```
# Scatter Plot
plot(airlines_data$weather_delay, airlines_data$arr_delay,
     main = "Arrival Delay vs Weather Delay",
     xlab = "Weather Delay (minutes)", ylab = "Arrival Delay (minutes)",
     xlim = c(0,1000))
```

## Arrival Delay vs Weather Delay



## Linear Regression

```
# Select the relevant columns for the linear regression analysis
airlines_data_1m <- airlines_data %>%
select (arr_delay, carrier_delay, weather_delay, nas_delay, security_delay, late_aircraft_delay)

# Fit a linear regression model
arr_carrier_model <- lm(arr_delay ~ carrier_delay , data = airlines_data_1m)
arr_weather_model <- lm(arr_delay ~ weather_delay , data = airlines_data_1m)
arr_nas_model <- lm(arr_delay ~ nas_delay , data = airlines_data_1m)
arr_security_model <- lm(arr_delay ~ security_delay , data = airlines_data_1m)
arr_aircraft_model <- lm(arr_delay ~ late_aircraft_delay , data = airlines_data_1m)

#Print the model summary
print(summary(arr_carrier_model))
```

```
##
## Call:
## lm(formula = arr_delay ~ carrier_delay, data = airlines_data_1m)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -57616   -548    -191     223   57544
##
```

```
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  215.55778   54.86562   3.929 8.66e-05 ***
## carrier_delay  2.65278    0.00974 272.365 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3528 on 4605 degrees of freedom
## (5 observations deleted due to missingness)
## Multiple R-squared:  0.9416, Adjusted R-squared:  0.9415
## F-statistic: 7.418e+04 on 1 and 4605 DF, p-value: < 2.2e-16
```

```
print(summary (arr_weather_model))
```

```
##
## Call:
## lm(formula = arr_delay ~ weather_delay, data = airlines_data_1m)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -110151   -2062   -1577    -206   172750
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2095.2375   160.0484   13.09 <2e-16 ***
## weather_delay  10.3403    0.1562   66.18 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10450 on 4605 degrees of freedom
## (5 observations deleted due to missingness)
## Multiple R-squared:  0.4875, Adjusted R-squared:  0.4874
## F-statistic: 4380 on 1 and 4605 DF, p-value: < 2.2e-16
```

```
print (summary (arr_nas_model))
```

```
##
## Call:
## lm(formula = arr_delay ~ nas_delay, data = airlines_data_1m)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -70239   -1025    -718    176  145597
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  967.63751  107.92052   8.966 <2e-16 ***
## nas_delay     4.02231   0.03233 124.428 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6988 on 4605 degrees of freedom
## (5 observations deleted due to missingness)
```

```
## Multiple R-squared:  0.7707, Adjusted R-squared:  0.7707
## F-statistic: 1.548e+04 on 1 and 4605 DF,  p-value: < 2.2e-16
```

```
print(summary(arr_security_model))
```

```
##
## Call:
## lm(formula = arr_delay ~ security_delay, data = airlines_data_1m)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
	-188983	-3170	-2502	-806	276897

```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	3471.691	195.605	17.75	<2e-16 ***
security_delay	141.454	4.009	35.28	<2e-16 ***

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12950 on 4605 degrees of freedom
## (5 observations deleted due to missingness)
## Multiple R-squared:  0.2128, Adjusted R-squared:  0.2126
## F-statistic: 1245 on 1 and 4605 DF,  p-value: < 2.2e-16
```

```
print(summary(arr_aircraft_model))
```

```
##
## Call:
## lm(formula = arr_delay ~ late_aircraft_delay, data = airlines_data_1m)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
	-26529	-672	-416	113	50317

```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	547.24086	51.70081	10.59	<2e-16 ***
late_aircraft_delay	2.34028	0.00813	287.85	<2e-16 ***

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3349 on 4605 degrees of freedom
## (5 observations deleted due to missingness)
## Multiple R-squared:  0.9473, Adjusted R-squared:  0.9473
## F-statistic: 8.286e+04 on 1 and 4605 DF,  p-value: < 2.2e-16
```

## Chi-Square Test



```

delta <- subset(airlines_data, carrier_name == "Delta Air Lines Inc.")
american <- subset(airlines_data, carrier_name == "American Airlines Inc.")

# carrier_ct vs weather_ct
delta_table<- table (delta$carrier_ct, delta$weather_ct)
american_table <- table(american$carrier_ct, american$weather_ct)
chi_sq <- chisq.test(delta_table, american_table)

## Warning in chisq.test(delta_table, american_table): Chi-squared approximation
## may be incorrect

print(chi_sq)

##
## Pearson's Chi-squared test
##
## data:  delta_table
## X-squared = 90954, df = 85360, p-value < 2.2e-16

#carrier_ct vs nas_ct
delta_table<- table (delta$carrier_ct, delta$nas_ct)
american_table <- table(american$carrier_ct, american$nas_ct)
chi_sq <- chisq.test(delta_table, american_table)

## Warning in chisq.test(delta_table, american_table): Chi-squared approximation
## may be incorrect

print(chi_sq)

##
## Pearson's Chi-squared test
##
## data:  delta_table
## X-squared = 146650, df = 138904, p-value < 2.2e-16

#carrier_ct vs security_ct
delta_table <- table(delta$carrier_ct, delta$security_ct)
american_table <- table (american$carrier_ct, american$security_ct)
chi_sq <- chisq.test(delta_table, american_table)

## Warning in chisq.test(delta_table, american_table): Chi-squared approximation
## may be incorrect

print(chi_sq)

##
## Pearson's Chi-squared test
##
## data:  delta_table
## X-squared = 7303.2, df = 6596, p-value = 1.293e-09

```

```
#carrier_ct vs late_aircraft_ct
delta_table <- table (delta$carrier_ct, delta$late_aircraft_ct)
american_table <- table (american$carrier_ct, american$late_aircraft_ct)
chi_sq <- chisq.test(delta_table, american_table)

## Warning in chisq.test(delta_table, american_table): Chi-squared approximation
## may be incorrect

print(chi_sq)
```

```
##
## Pearson's Chi-squared test
##
## data: delta_table
## X-squared = 147352, df = 137740, p-value < 2.2e-16
```

## Correlation Matrix

```
data <- airlines_data %>%
  select(carrier_delay, weather_delay, nas_delay, security_delay, late_aircraft_delay)

correlation_matrix <- cor(data, use = "complete.obs")
print(correlation_matrix)
```

```
##
## carrier_delay weather_delay nas_delay security_delay
## carrier_delay 1.0000000 0.7406985 0.7743065 0.4501237
## weather_delay 0.7406985 1.0000000 0.5202790 0.4013624
## nas_delay 0.7743065 0.5202790 1.0000000 0.3883256
## security_delay 0.4501237 0.4013624 0.3883256 1.0000000
## late_aircraft_delay 0.9234710 0.5888050 0.8175952 0.4364896
##
## late_aircraft_delay
## carrier_delay 0.9234710
## weather_delay 0.5888050
## nas_delay 0.8175952
## security_delay 0.4364896
## late_aircraft_delay 1.0000000
```