Table of Contents

Introduction 3

Question 01: Frequency Distribution of “Salaries and Wages” 3

Question 02: Central Tendency of “Salaries and Wages” 4

Question 03: Histograms for “Grouped Salaries” 6

Question 04: Pie Chart and Bar Diagram for “Maintenance” and “Load factor” 8

Question 05: Box Plot for “Purchased Goods”, “Aircraft Ownerships” and “Daily Utilization per Aircraft” 12

Question 06: Plot summary 14

Introduction:

This report provides a comprehensive statistical analysis of fleet performance, focusing on salaries, maintenance costs, and aircraft utilization. It employs frequency distributions and graphical representations to examine operational data from 1995 to 2004 across fleet categories, including Small Narrowbodies, Large Narrowbodies, Widebodies, and the Total Fleet. The findings aim to offer insights into United Airlines' operational strategies, supporting improved resource utilization and more effective decision-making.

Question 01: Frequency Distribution of "Salaries and Wages"

The dataset comprised information on the Total Fleet, Large Narrowbodies, Small Narrowbodies, and Widebodies, with relevant data points ranging from B2 to W158. The data was loaded and analyzed using the following code.

Table I: Code for the Q1

```
install.packages("readxl")
install.packages("fdth")

library(readxl)
library(fdth)

# Load the dataset
data <- read_excel(
  path =
"C:/Users/shoae/Desktop/Assignment_22024003/United_Airlines_Aircraft_Operating_S
tatistics-_Cost_Per_Block_Hour_(Unadjusted).xls",
  range = "b2:w158"  # Specify the range of cells to read
)
data <- as.data.frame(data)  # Convert to a data frame for easier manipulation

# Define fleet names and corresponding row indices for "Salaries and Wages"
fleets <- c("Small Narrowbodies", "Large Narrowbodies", "Widebodies", "Total
Fleet")
wages_rows <- c(6, 45, 84, 123)  # Adjusted row indices (8, 47, 86, 125) - 2

# Extracts numeric data from a specific row and removes NAs
get_row <- function(row, data) {
  # Convert row to numeric and exclude NA values, selecting the first 19 entries
  na.omit(as.numeric(data[row, -1]))[1:13]
}

# Creates and prints a frequency distribution table
create_table <- function(data, fleet_name) {
  # Calculate class interval using Sturges' formula
  interval <- as.integer((max(data) - min(data)) / (log2(length(data)) + 1))

  # Generate the frequency distribution table
  freq_table <- fdt(data, start = min(data), end = max(data), h = interval)

  # Print the results
  cat("\nFrequency Distribution Table for", fleet_name, ":\n")
  print(as.data.frame(freq_table$table))  # Convert table to data frame for
readability
}
```

```
#main code

# Iterate through each fleet type to generate and display frequency tables
for (i in seq_along(wages_rows)) {
  # Extract "Salaries and Wages" data for the current fleet
  pilots_wages <- get_row(wages_rows[i], data)

  # Create and print the frequency distribution table
  create_table(pilots_wages, fleets[i])
}
```

The output of this code provides following frequency distribution tables for small narrowbodies, large narrowbodies, wide bodies and total fleet.

Table I.I: Small Narrowbodies

Class Intervals	Frequency
291.555,349.555	4
349.555,407.555	2
407.555,465.555	4
465.555,523.555	1

Table I.II: Large Narrowbodies

Class Intervals	Frequency
349.52,413.52	4
413.52,477.52	1
477.52,541.52	3
541.52,605.52	3

Table I.III: Widebodies

Class Intervals	Frequency
581.0153,698.0153	4
698.0153,815.0153	2
932.0153,1049.015	4
815.0153,932.0153	1

Table I.IV: Total Fleet

Class Intervals	Frequency
385.51,464.51	4
464.51,543.51	2
543.51,622.51	3
622.51,701.51	2

Question 02: Central Tendency of “Salaries and Wages”

Mean, Median, Mode, Standard Deviation, Variance, Quartiles, 9th Decile, 10th Percentile and Range of “Salaries and Wages”

```
library(readxl)

# Load data
data <-
read_excel("C:/Users/shoae/Desktop/Assignment_22024003/United_Airlines_Ai
rcraft_Operating_Statistics-_Cost_Per_Block_Hour_(Unadjusted).xls", range
= "b2:w158")
data <- as.data.frame(data)

# Define rows to analyze
wages_rows <- c(8, 47, 86, 125) - 2
fleets <- c("Small Narrowbodies", "Large Narrowbodies", "Widebodies",
"Total Fleet")

# Function to get the first 19 non-NA values from a specified row
get_row <- function(row, data) {
  na.omit(as.numeric(data[row, -1]))[1:13]
}

# Function to calculate mode
get_mode <- function(v) {
  uniqv <- unique(v)
  uniqv[which.max(tabulate(match(v, uniqv)))]
}

# Function to format numbers to 3 decimal places
format_number <- function(x) {
  format(round(x, 3), nsmall = 3)
}

# Calculate statistics for specified rows
results_list <- lapply(wages_rows, function(i) {
  wages <- get_row(i, data)
  list(
    mean = format_number(mean(wages, na.rm = TRUE)),
    mode = format_number(get_mode(wages)),
    median = format_number(median(wages, na.rm = TRUE)),
    range = format_number(max(wages, na.rm = TRUE) - min(wages, na.rm =
TRUE)),
    sd = format_number(sd(wages, na.rm = TRUE)),
    variance = format_number(var(wages, na.rm = TRUE)),
```

```
    quartiles = paste(format_number(quantile(wages, probs = c(0.25, 0.5, 
0.75), na.rm = TRUE)), collapse = ", "),
    deciles = paste(format_number(quantile(wages, probs = c(0.9, 1), 
na.rm = TRUE)), collapse = ", ")
  )
})

# Convert results into a matrix format
results_matrix <- do.call(cbind, lapply(results_list, function(x) {
  c(x$mean, x$mode, x$median, x$range, x$sd, x$variance, x$quartiles, 
x$deciles)
}))

# Assign row names and column names
rownames(results_matrix) <- c("mean", "mode", "median", "range", "sd", 
"variance", "quartiles", "deciles")
colnames(results_matrix) <- fleets

# Convert the matrix to data frame for better printing
results_df <- as.data.frame(results_matrix)

# Print the results in the desired format
print(results_df, row.names = TRUE)
```

Table II: Code for the Q2

The output of the given code gives the following output for the data of 4 catagories.

Table III: Central tendency for the data

	Small Narrowbodies	Large Narrowbodies
mean	407.720	496.256
mean	838.118	547.802
mode	758.697	510.407
median	858.441	552.980
range	550.117	372.286
sd	197.047	126.609
variance	38827.643	16029.790
quartiles	652.931, 858.441, 990.396	429.842, 552.980, 623.455
deciles	1064.090, 1131.132	705.841, 757.800

Question 03: Histograms for "Grouped Salaries"

There can be four histograms for the frequency distribution tables from the question

```
library(readxl)
library(fdth)

# Load data
data <-
read_excel("C:/Users/shoae/Desktop/Assignment_22024003/United_Airlines
_Aircraft_Operating_Statistics-_Cost_Per_Block_Hour_(Unadjusted).xls",
range="b2:w158")
data <- as.data.frame(data)

fleets <- c("Small Narrowbodies", "Large Narrowbodies", "Widebodies",
"Total Fleet")
wages_rows <- c(8, 47, 86, 125) - 2

# Function to get the first 19 non-NA values from a specified row
get_row <- function(row, data) {
  na.omit(as.numeric(data[row, -1]))[1:13]
}

# Function to create a frequency distribution table and save the plot
create_and_save_table <- function(data, fleet_name) {
  interval <- as.integer((max(data) - min(data)) / (log2(length(data))
+ 1))
  table <- fdt(data, start = min(data), end = max(data), h = interval)

  # Save plot as PNG
  png_filename <- paste0("C:\\Users\\shoae\\Desktop\\MAT-4509",
fleet_name, "_wages_distribution.png")
  png(filename = png_filename, width = 800, height = 600)
  plot(table, main = paste("Wages Distribution for", fleet_name))
  dev.off()  # Close the PNG device
}

# Main code to create tables and save plots
for (i in seq_along(wages_rows)) {
  pilots_wages <- get_row(wages_rows[i], data)
  create_and_save_table(pilots_wages, fleets[i])
}
```

Table IV: Code for Q3

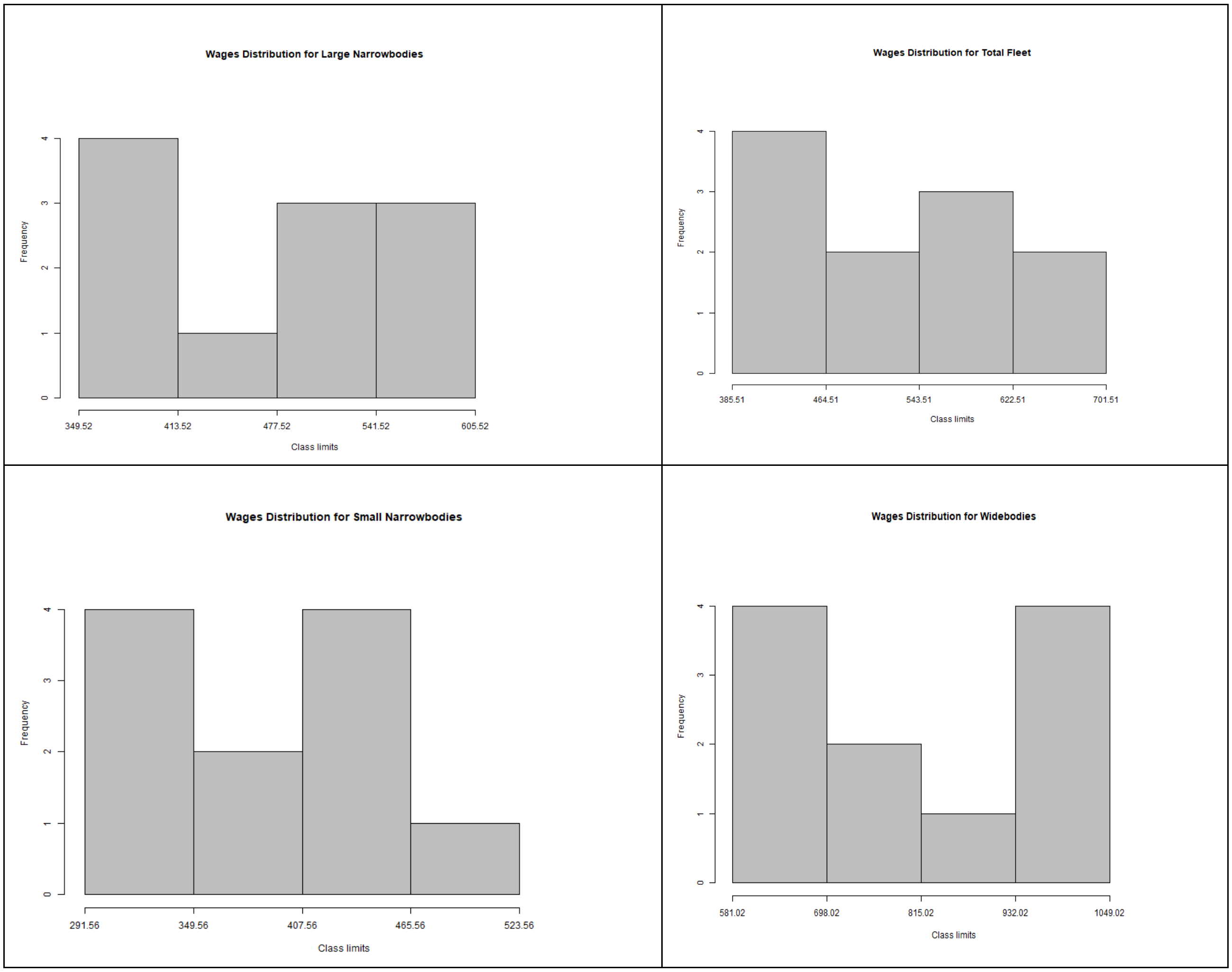


Figure 1: Frequency distribution histograms

Question 04: Pie Chart and Bar Diagram for "Maintenance" and "Load factor"

The following code plots pie charts for the "Maintenance" variables and bar plots for "Load factors".

Table V: Code for Q4

```
library(readxl)

#Load Data
data <-
read_excel("C:/Users/shoae/Desktop/Assignment_22024003/United_Airlines_Aircraft_Opera
ting_Statistics-_Cost_Per_Block_Hour_(Unadjusted).xls", range="b2:w158")
data <- as.data.frame(data)
fleets <- c("Small Narrowbodies", "Large Narrowbodies", "Widebodies", "Total Fleet")
years_row <- 3 - 2
load_factor_rows <- c(36, 75, 114, 153) - 2
maintanance_rows <- matrix(c(19,20,21,23,58,59,60,62,97,98,99,101,136,137,138,140) -
2, nrow = 4, byrow = FALSE)

get_row <- function(row, data) {
  na.omit(as.numeric(data[row, -1]))[1:13]
}

get_row_sum <- function(row, data) {
  sum(na.omit(as.numeric(data[row, -1]))[1:13])
}

get_totals <- function(row_matrix, data) {
  sapply(1:4, function(i) sapply(1:4, function(j) get_row_sum(row_matrix[i, j],
data)))
}

plot_and_save_pies <- function(labour, materials, third_party, burden, labels,
save_path) {
  for (i in 1:4) {
    values <- c(labour[i], materials[i], third_party[i], burden[i])
    percentages <- round(values / sum(values) * 100)
    pie_labels <- paste(c("Labour", "Materials", "Third party", "Burden"),
percentages, "%")
    file_name <- paste0(save_path, "/", labels[i], "_maintenance_pie.png")
    # Open a PNG device and save the plot
    png(filename = file_name, width = 600, height = 600)
    pie(values, labels = pie_labels, main = labels[i], col = rainbow(length(values)))
    dev.off()  # Close the device to save the file
  }
}

create_barplot <- function(years_row, load_factor_rows, fleets, save_path) {
  years <- get_row(years_row, data)
  for (i in seq_along(load_factor_rows)) {
    load_factors <- get_row(load_factor_rows[i], data)
    fleet_name <- fleets[i]
```

```
    file_name <- paste0(save_path, "/barplot_", fleet_name, ".png")
    png(file_name, width = 800, height = 600)
    barplot(load_factors, names.arg = years, col = "skyblue",
            main = paste("Load Factors for", fleet_name),
            xlab = "Years", ylab = "Load Factors")
    dev.off()
  }
  message("Bar plots saved successfully in ", save_path)
}

# Define save path
save_path <- "C:\\Users\\shoae\\Desktop\\MAT-4509"
# Create directory if it doesn't exist
if (!dir.exists(save_path)) {
  dir.create(save_path, recursive = TRUE)
}
# Calculate totals for each category
totals <- get_totals(maintanance_rows, data)
total_labour <- totals[1, ]
total_materials <- totals[2, ]
total_third_party <- totals[3, ]
total_burden <- totals[4, ]
# Generate and save pie charts for each fleet
plot_and_save_pies(total_labour, total_materials, total_third_party, total_burden,
fleets, save_path)

if (!dir.exists(save_path)) {
  dir.create(save_path, recursive = TRUE)
}

create_barplot(years_row, load_factor_rows, fleets, save_path)
```

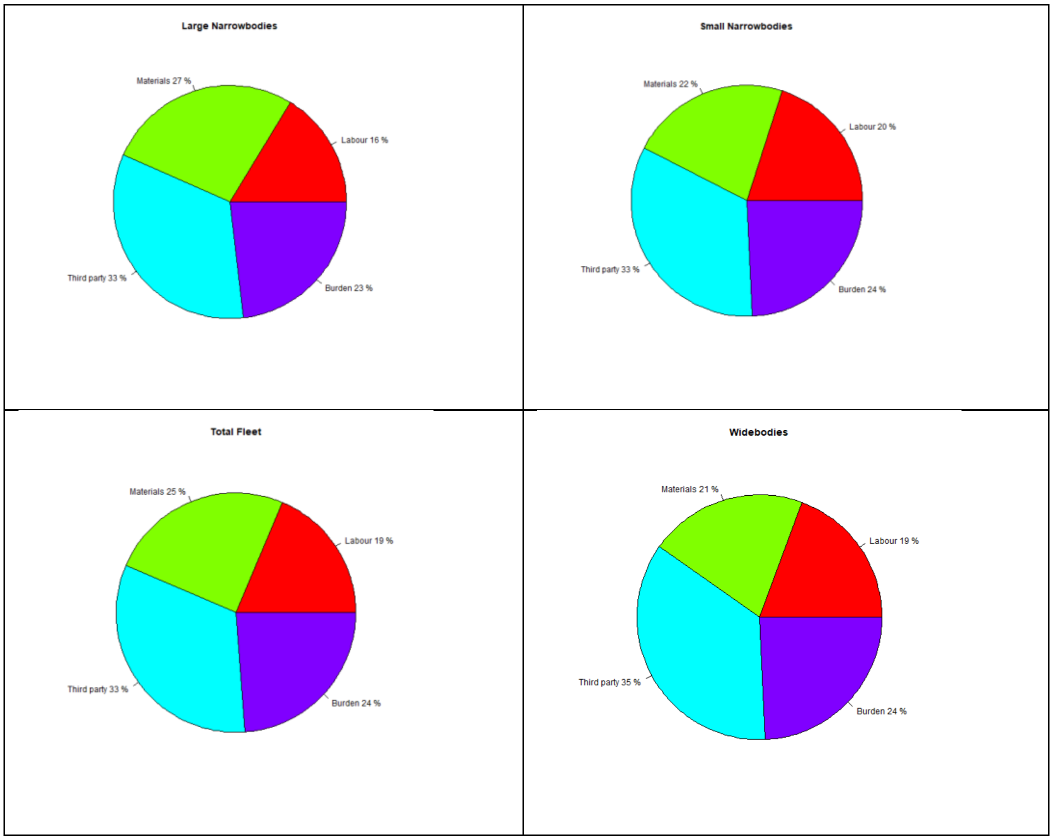


Figure 2: Pie Charts for the "Maintenance" variable

Here the "Maintenance" variable had the following "Labour", "Materials", "Third party", "Burden" sections. The pie chart was plotted by talking the average of 10 years for each type of fleets. We get 4 pie charts out of this.

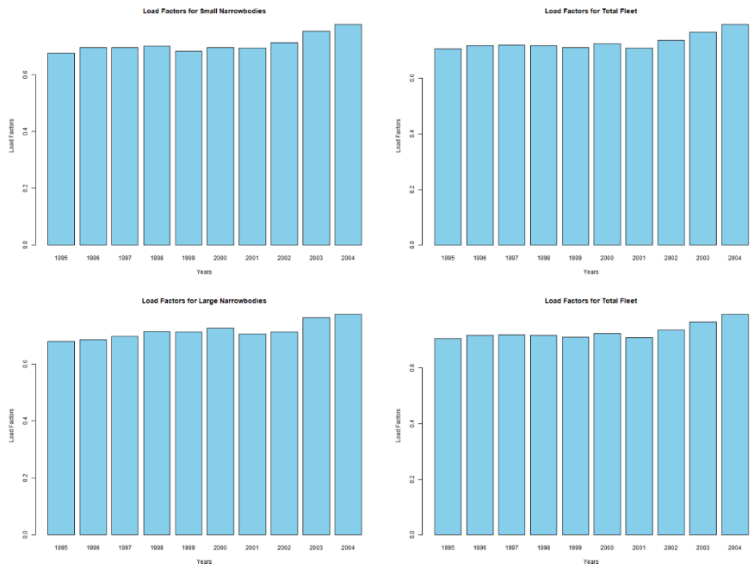


Figure 3: Barplot for "Load factor" variable

At the 2nd part of the code the bar plot for the "Load factor" variable was plotted, for each type of fleets the load factor was plotted for different years, in this case for 10 years

Question 05: Box Plot for "Purchased Goods", "Aircraft Ownerships" and "Daily Utilization per Aircraft"

The following code generates box plot for the following variables,

- Purchased Goods:
    - Fuel/Oil
    - Insurance
    - Other (inc. Tax)
- Aircraft Ownership:
    - Rentals
    - Depreciation and Amortization
- Daily Utilization per Aircraft:
    - Block Hours
    - Airborne Hours
    - Departures

Each of these sections contains multiple variables. The box plots were plotted for those variables for each section and each type of aircrafts.

Purchased goods:

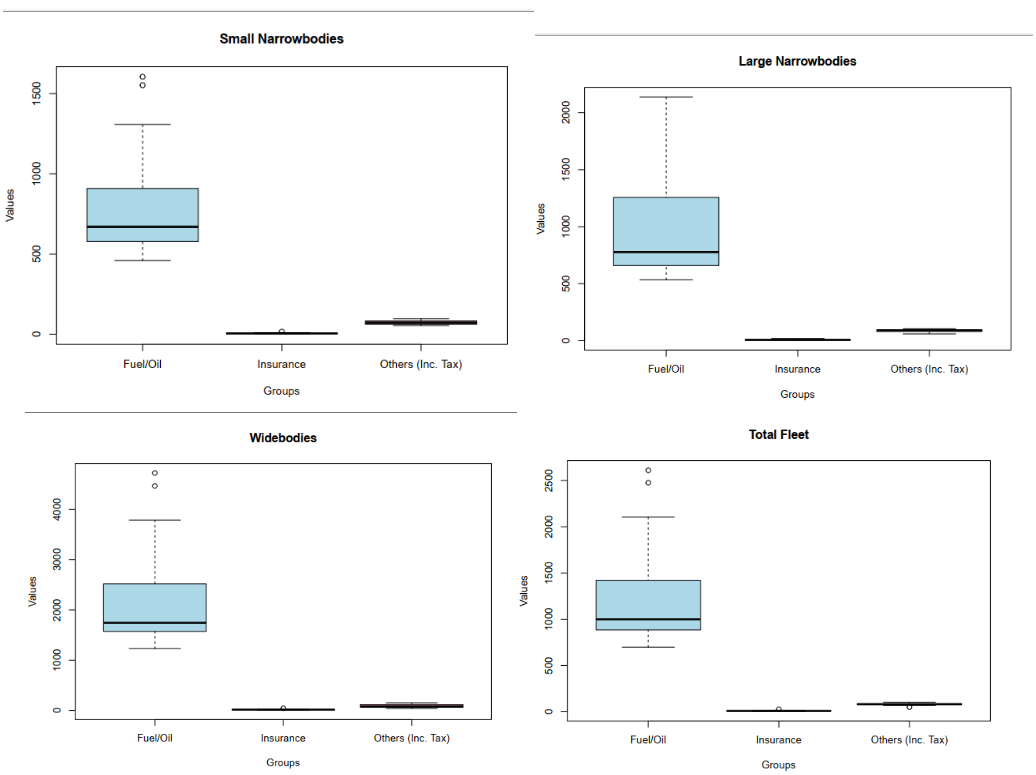


Figure 4: Box plots for "Purchased goods"

Aircraft Ownership:

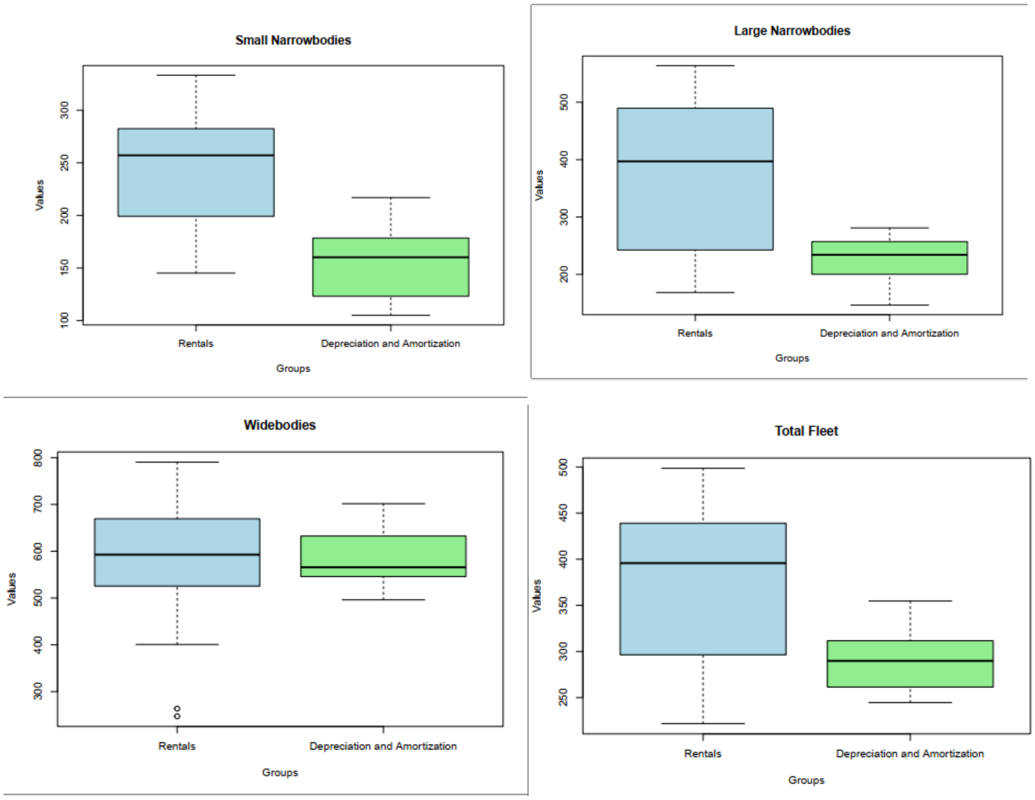
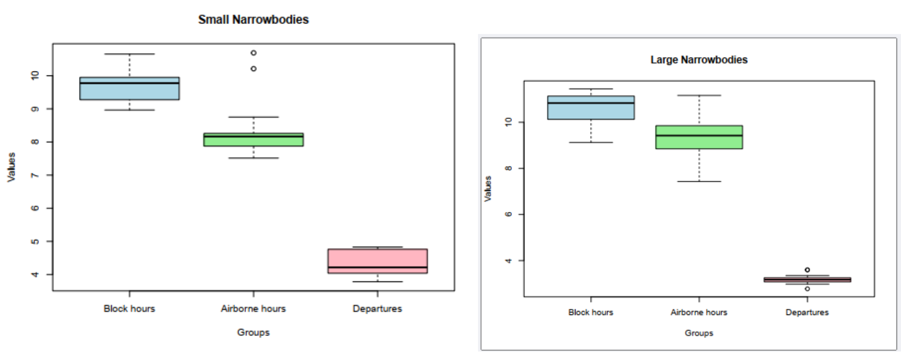


Figure 5: Box plots for "Aircraft ownership

Daily Utilization per Hour:



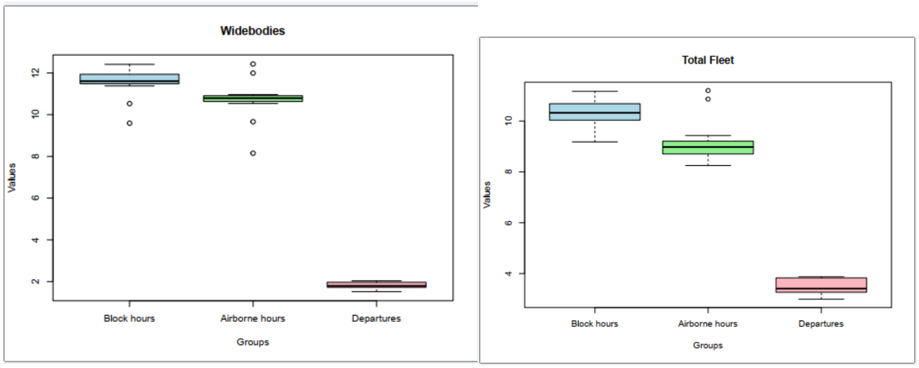


Figure 6: Box plots for "Daily Utilization per Aircraft"

Question 06: Plot summary

United Airlines' data offers a detailed comparison of various fleet types across operational and financial metrics.

Salaries Distribution:

- *Small Narrowbodies* exhibit a bimodal distribution, with most salaries concentrated in the middle class, and no representation in the highest class.
- *Large Narrowbodies* have a nearly uniform distribution, with higher salary classes more represented.
- *Widebodies* display a highly skewed distribution, with most salaries at the extremes and minimal representation in the middle.
- *Total Fleet* shows a consistent distribution, peaking in the middle to upper salary range.

Maintenance Costs:

- *Small and Large Narrowbodies* are balanced, with third-party services constituting the largest share, and labor and burden costs being similar.
- *Widebodies* have higher labor costs but still rely heavily on third-party services.
- *Total Fleet* provides a comprehensive view of maintenance costs, aligning with United's overall strategy while accommodating specific aircraft needs.

Load Factors (1995-2004):

- *Small Narrowbodies* show a rising trend, with higher values towards 2004 compared to lower values in 1995.
- *Large Narrowbodies* improve significantly, peaking higher than Small Narrowbodies by 2004.
- *Widebodies* maintain consistently high load factors throughout the decade.
- *Total Fleet* demonstrates steady improvement, reflecting an upward trend over the period.

Fuel/Oil Costs:

- Highest variability is observed in *Widebodies*.
- *Large Narrowbodies* show moderate variability.
- *Small Narrowbodies* have more compact distributions.
- *Insurance and Other Costs*, including taxes, are lower with compact distributions and fewer outliers.

Rentals vs. Depreciation & Amortization:

- *Small Narrowbodies* have higher median rental costs than depreciation, with moderate variability.
- *Large Narrowbodies* follow a similar pattern but with a wider spread in rental costs and an outlier in depreciation.
- *Widebodies* are the most diverse, with rentals dominating costs.

- *Total Fleet* sees higher median rental costs with a wider spread than depreciation, including an outlier in rentals.

Utilization Metrics (Block Hours, Airborne Hours, Departures):

- *Small Narrowbodies* lead in departures but have middling block and airborne hours, reflecting consistent short-haul use.
- *Large Narrowbodies* have higher block and airborne hours but fewer departures, indicating balanced usage.
- *Widebodies* achieve the highest block and airborne hours with the fewest departures, showcasing a variable usage pattern.
- *Total Fleet* maintains a balanced profile, with reasonable block and airborne hours and well-distributed departures.

Summary:
Smaller aircraft prioritize frequent, shorter flights, while larger aircraft, like Widebodies, emphasize longer flight hours with fewer departures. These patterns align with operational strategies tailored to each aircraft type's intended use and route demands.