Bangabandhu Sheikh Mujibur Rahman
Aviation and Aerospace University
(BSMRAAU)

ASSIGNMENT

Assignment No: 1

Assignment Info: Solving Statistical findings from a data sheet and building Necessary plots for analysis through R programming.

Date of Submission: 26-11-2024

Submitted by: Snigdha Das
Student ID: 22024012

Department: Aeronautical Engineering (Avionics) [Batch-2]

Course Name: Probability and Statistics

Course Code: MAT 4509

Course Instructor: Md. Siddikur Rahman, PhD
Associate Professor
Dept. of Statistics, BRUR

Assignment Questions.
####_1. First enter the United Airlines Aircraft Operating Statistics and then select a sample of size, n=last two digits of your ID and answer the exercises.

i) Select an appropriate class interval and organize the "Salaries and Wages" into a frequency distribution.

ii) Compute the Mean, Median, Mode, Standard Deviation, Variance, Quartiles, 9th Decile, 10th Percentile and Range of "Salaries and Wages" from the raw data of your sample and interpret.'

iii) Develop a histogram (Using the question "i") for the variable "Grouped Salaries".

iv) Develop a Pie chart and a Bar diagram for the variables "Maintenance" and "Load factor".

v) Develop a Box plot for the variables "Purchased Goods", "Aircraft Ownerships" and "Daily Utilization per Aircraft".

vi) What information can you give from these plots?

Assignment Report:
Answer code 1(i,ii,iii):

```
# Snigdha Das(ID_22024012)
# Changes should be made to the working directory for using it in other PC.
rm(list = ls())# to clear the environment.

library(readxl)

setwd("F:/Rstudio codes/MAT_4509")
#variable with my ID
id<-12

# load dataset
xl_data <- read_excel("United Airlines Aircraft Operating Statistics- Cost Per Block Hour (Unadjusted).xls",
range = "b2:w158")

# creating function to salarys and wages data by row
#took the same number of data input as my id's last two digits.
get_salary_wages <- function(row_num, data = xl_data){
  return((as.numeric(data[row_num, 1:id+1])))
}
# Extract salary and wages data for different fleets
salary_wages_snbodies <- get_salary_wages(6)
salary_wages_lnbodies <- get_salary_wages(45)
salary_wages_wbodies <- get_salary_wages(84)
salary_wages_tfleet <- get_salary_wages(123)

get_modes <- function(data) {
  freq_table <- table(data)
  max_freq <- max(freq_table)
```

```
  modes <- as.numeric(freq_table[freq_table == max_freq])
  if (length(modes) == length(data)) {
    return(NULL)
  }
  return(modes)
}
make_Freq_Dis <- function(
    wage_data) {
  # number of observations
  n <- length(wage_data)

  # calculating the value of k (smallest k such that 2^k > n)
  k <- 0
  for (i in 1:(n / 2)) {
    if (2^i > n) {
      k <- i
      break
    }
  }

  # calculating class interval ( interval >= (max - min)/k)
  min_salary <- min(wage_data)
  max_salary <- max(wage_data)
  class_interval <- (max_salary - min_salary) / k
  class_interval <- ceiling(class_interval)

  # Creating breakpoints
  break_points <- seq(
    min_salary - (class_interval / 2),
    max_salary + (class_interval / 2),
    by = class_interval
  )

  # Creating frequency distribution
  salary_bins <- cut(wage_data, breaks = break_points, right = TRUE)
  frequency_distribution <- table(salary_bins)
  #cut() function is used to divide a numeric vector into different ranges.
  return(frequency_distribution)
}
print_analysis <- function(wage_data, title) {
  mean <- mean(wage_data)
  median <- median(wage_data)
  modes <- get_modes(wage_data)
  sample_sd <- sd(wage_data) # sample
  sample_var <- var(wage_data) # sample
  quartiles <- quantile(wage_data, probs = c(0.25, 0.5, 0.75))
  tenth_percentile <- quantile(wage_data, probs = 0.10)
  ninth_decile <- quantile(wage_data, probs = 0.90)
  range <- max(wage_data) - min(wage_data)
```

```
  # print results
    # Here,I used cat() to concatenate and print strings and values from the variable together.
  cat("Analysis of ", title, ":\n")
  cat("Mean:", mean, "\n")
  cat("Median:", median, "\n")
  if (is.null(modes) || length(modes) == 0) {
    cat("Modes: None\n")
  } else {
    cat("Modes:", paste(modes, collapse = ", "), "\n")
  }

  cat("Sample Standard Deviation:", sample_sd, "\n")
  cat("Sample Variance:", sample_var, "\n")
  cat("Quartiles (Q1, Q2, Q3):", quartiles, "\n")
  cat("10th Percentile:", tenth_percentile, "\n")
  cat("9th Decile:", ninth_decile, "\n")
  cat("Range:", range, "\n")
  cat("\n\n")
}
set_window_size <- function() {
  windows(width = 1800 / 100, height = 1080 / 100)
}

plot_histogram <- function(frequency_distribution, title,xlim,ylim) {
  set_window_size()
  barplot(frequency_distribution,
          main = title,
          xlab = "Salary Ranges",
          ylab = "Frequency",
          col = "steelblue",
          border = "black",
          space = 0.5, # No space between bars
          width = 1 # Adjust width to fill the space better
  )

}

# get frequency distribution (i)
SN_freq_dis <- make_Freq_Dis(salary_wages_snbodies)
LN_freq_dis <- make_Freq_Dis(salary_wages_lnbodies)
WB_freq_dis <- make_Freq_Dis(salary_wages_wbodies)
TO_freq_dis <- make_Freq_Dis(salary_wages_tfleet)

# print Frequency Distribution
cat("Frequency Distribution for Small Narrowbodies:\n")
print(SN_freq_dis)
cat("\nFrequency Distribution for Large Narrowbodies:\n")
print(LN_freq_dis)
```

```
cat("\nFrequency Distribution for Widebodies:\n")
print(WB_freq_dis)
cat("\nFrequency Distribution for Total Fleet:\n")
print(TO_freq_dis)

# print analysis (ii)
print_analysis(salary_wages_snbodies, "salary wages of small narrowbodies")
print_analysis(salary_wages_lnbodies, "salary wages of large narrowbodies")
print_analysis(salary_wages_wbodies, "salary wages of widebodies")
print_analysis(salary_wages_tfleet, "salary wages of total fleet")

# histogram using i. (iii)
plot_histogram(SN_freq_dis, "salary wages of small narrowbodies")
plot_histogram(LN_freq_dis, "salary wages of large narrowbodies")
plot_histogram(WB_freq_dis, "salary wages of widebodies")
plot_histogram(TO_freq_dis, "salary wages of total fleet")
```

Output 1(i):

```
> # print Frequency Distribution
> cat("Frequency Distribution for Small Narrowbodies:\n")
Frequency Distribution for Small Narrowbodies:
> print(SN_freq_dis)
salary_bins
(257,327] (327,397] (397,467] (467,537]
        2         3         4         1
> cat("\nFrequency Distribution for Large Narrowbodies:\n")

Frequency Distribution for Large Narrowbodies:
> print(LN_freq_dis)
salary_bins
(311,388] (388,465] (465,542] (542,619]
        2         2         3         3
> cat("\nFrequency Distribution for Widebodies:\n")

Frequency Distribution for Widebodies:
> print(WB_freq_dis)
salary_bins
     (512,650]      (650,788]      (788,926] (926,1.06e+03]
             2              3              1              4
> cat("\nFrequency Distribution for Total Fleet:\n")

Frequency Distribution for Total Fleet:
> print(TO_freq_dis)
salary_bins
(339,433] (433,527] (527,621] (621,715]
        3         2         3         2
```

Output 1(ii):

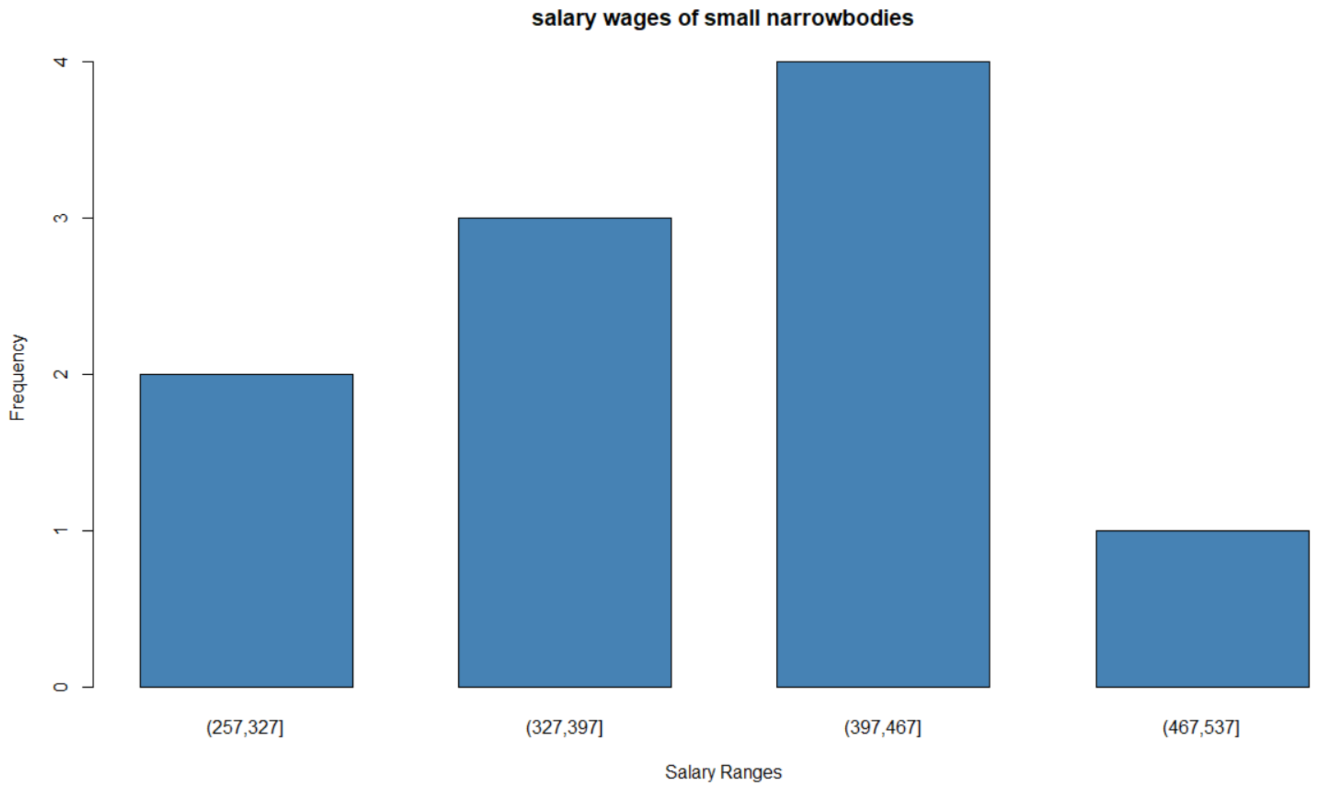
> # print analysis (ii)
> print_analysis(salary_wages_snbodies, "salary wages of small narrowbodies")
Analysis of salary wages of small narrowbodies :
Mean: 417.0428
Median: 423.6756
Modes: None
Sample Standard Deviation: 90.3098
Sample Variance: 8155.86
Quartiles (Q1, Q2, Q3): 359.6435 423.6756 453.95
10th Percentile: 296.1847
9th Decile: 545.78
Range: 276.8887

> print_analysis(salary_wages_lnbodies, "salary wages of large narrowbodies")
Analysis of salary wages of large narrowbodies :
Mean: 506.9169
Median: 528.1182
Modes: None
Sample Standard Deviation: 103.3259
Sample Variance: 10676.24
Quartiles (Q1, Q2, Q3): 427.8891 528.1182 583.0776
10th Percentile: 356.6653
9th Decile: 624.8474
Range: 304.8817

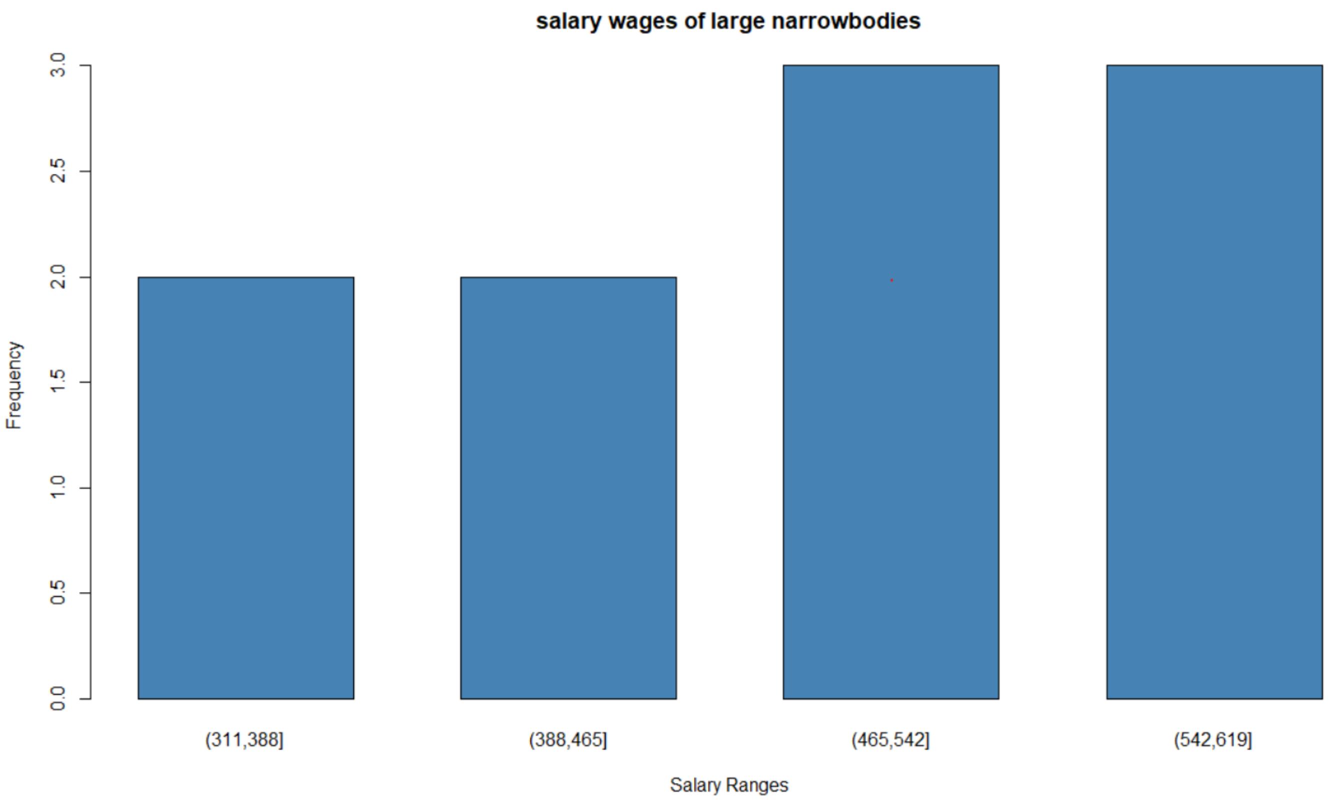
> print_analysis(salary_wages_wbodies, "salary wages of widebodies")
Analysis of salary wages of widebodies :
Mean: 858.3783
Median: 903.9113
Modes: None
Sample Standard Deviation: 191.1442
Sample Variance: 36536.1
Quartiles (Q1, Q2, Q3): 717.6633 903.9113 991.408
10th Percentile: 590.5458
9th Decile: 1072.796
Range: 550.1166

> print_analysis(salary_wages_tfleet, "salary wages of total fleet")
Analysis of salary wages of total fleet :
Mean: 560.3813
Median: 579.3378
Modes: None
Sample Standard Deviation: 123.4623
Sample Variance: 15242.94
Quartiles (Q1, Q2, Q3): 472.8665 579.3378 630.0795
10th Percentile: 389.9788
9th Decile: 712.827
Range: 372.2857

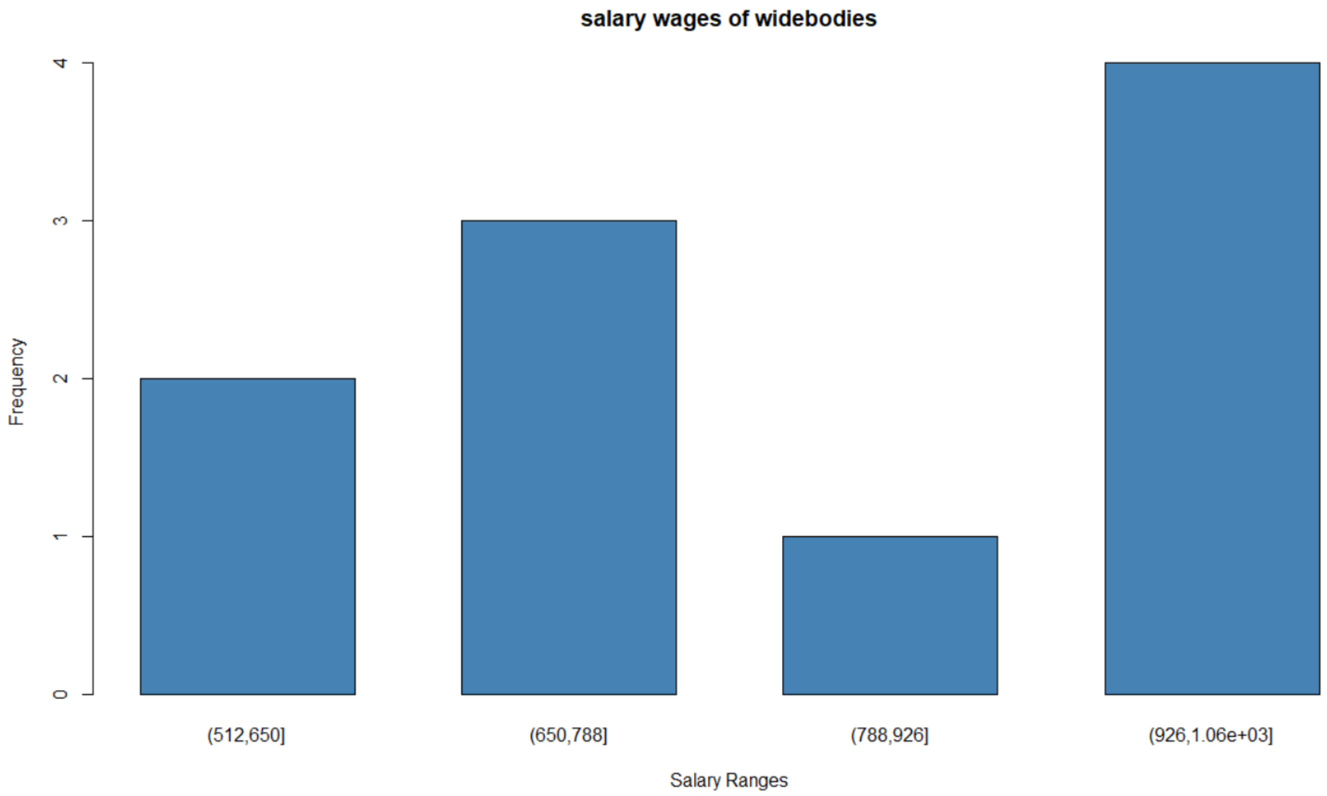
Output 1(iii):

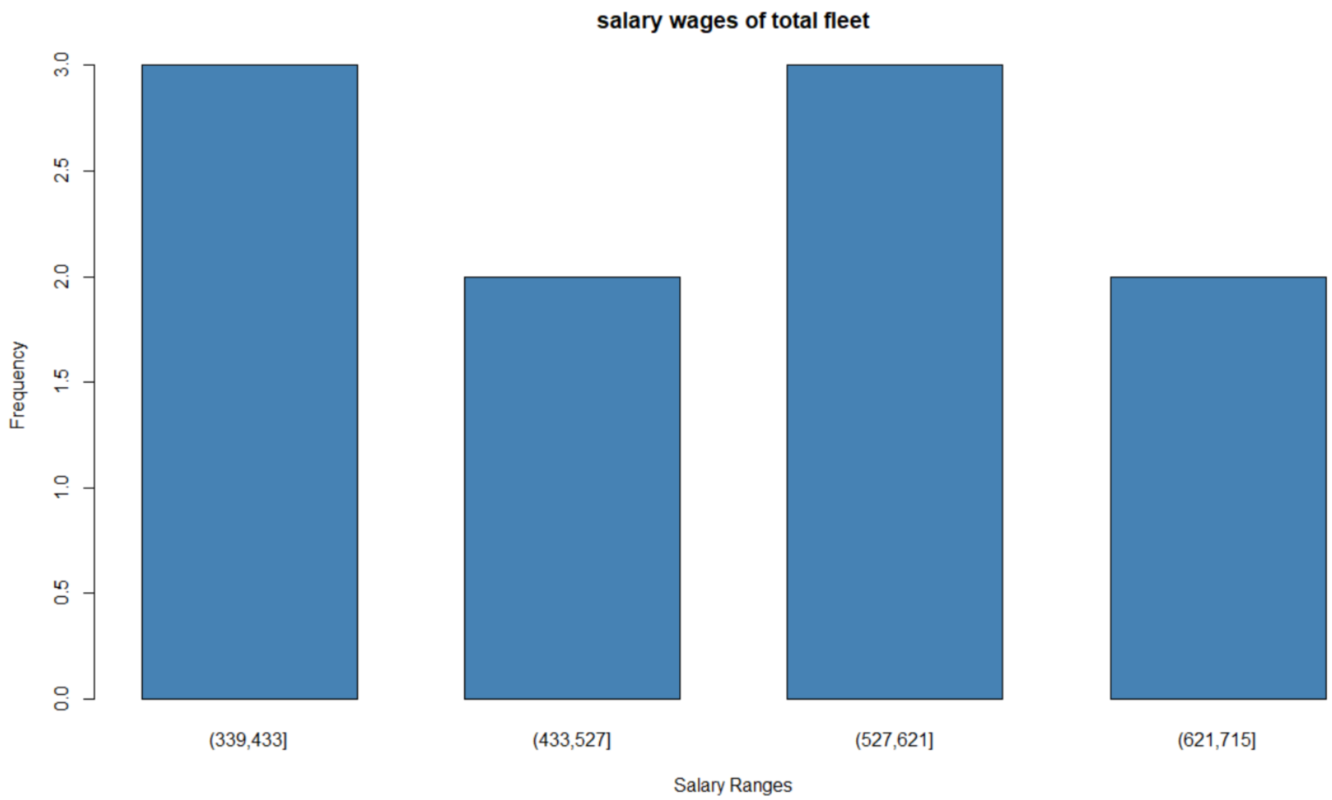


salary wages of small narrowbodies
Frequency
0
1
2
3
4
(257,327]
(327,397]
(397,467]
(467,537]
Salary Ranges



salary wages of large narrowbodies
Frequency
0.0
0.5
1.0
1.5
2.0
2.5
3.0
(311,388]
(388,465]
(465,542]
(542,619]
Salary Ranges


salary wages of widebodies
Frequency
0
1
2
3
4
(512,650]
(650,788]
(788,926]
(926,1.06e+03]
Salary Ranges

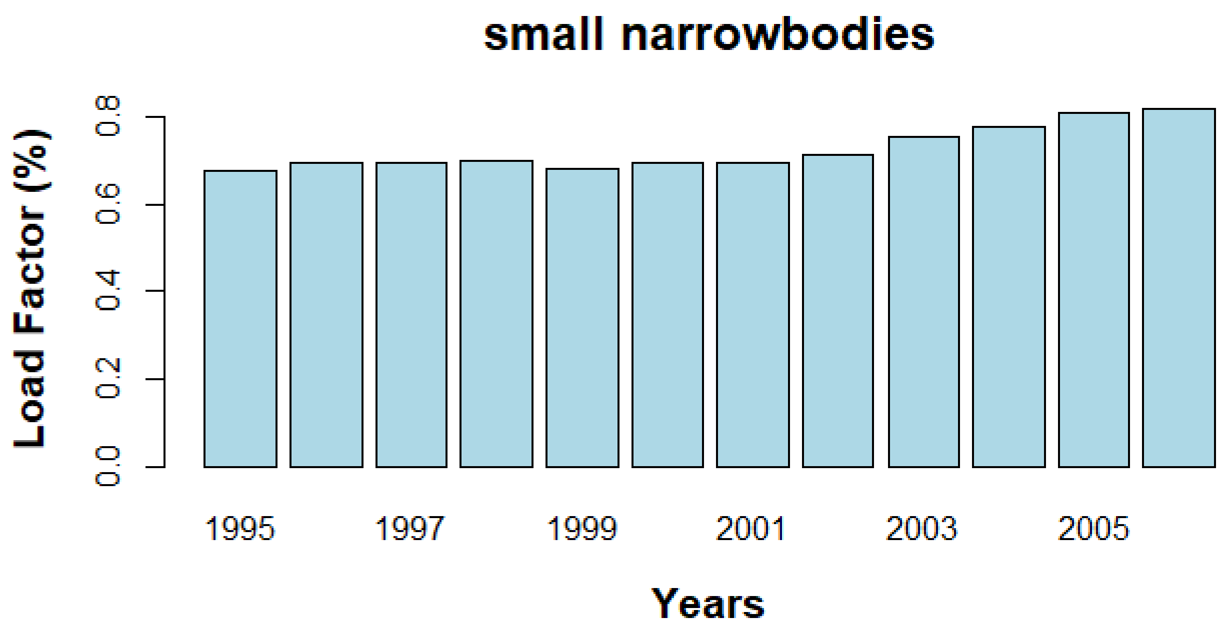

salary wages of total fleet
Frequency
0.0
0.5
1.0
1.5
2.0
2.5
3.0
(339,433]
(433,527]
(527,621]
(621,715]
Salary Ranges

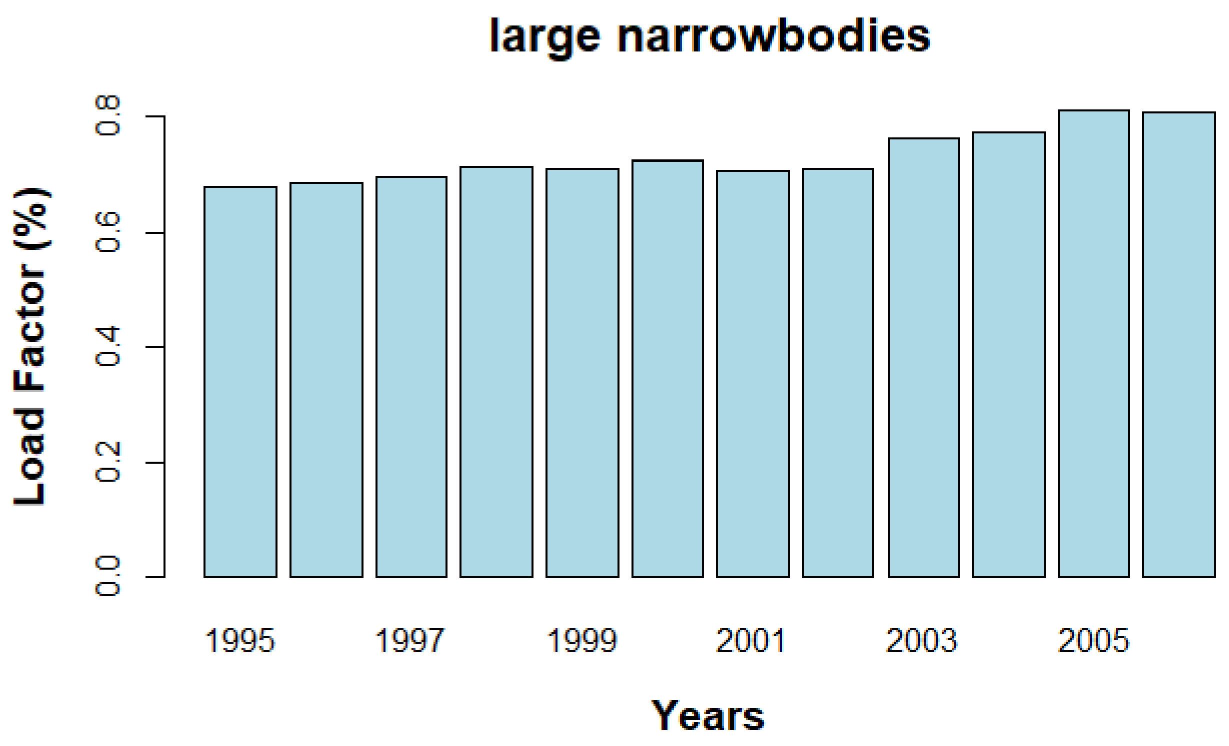
Answer Code 1(iv):

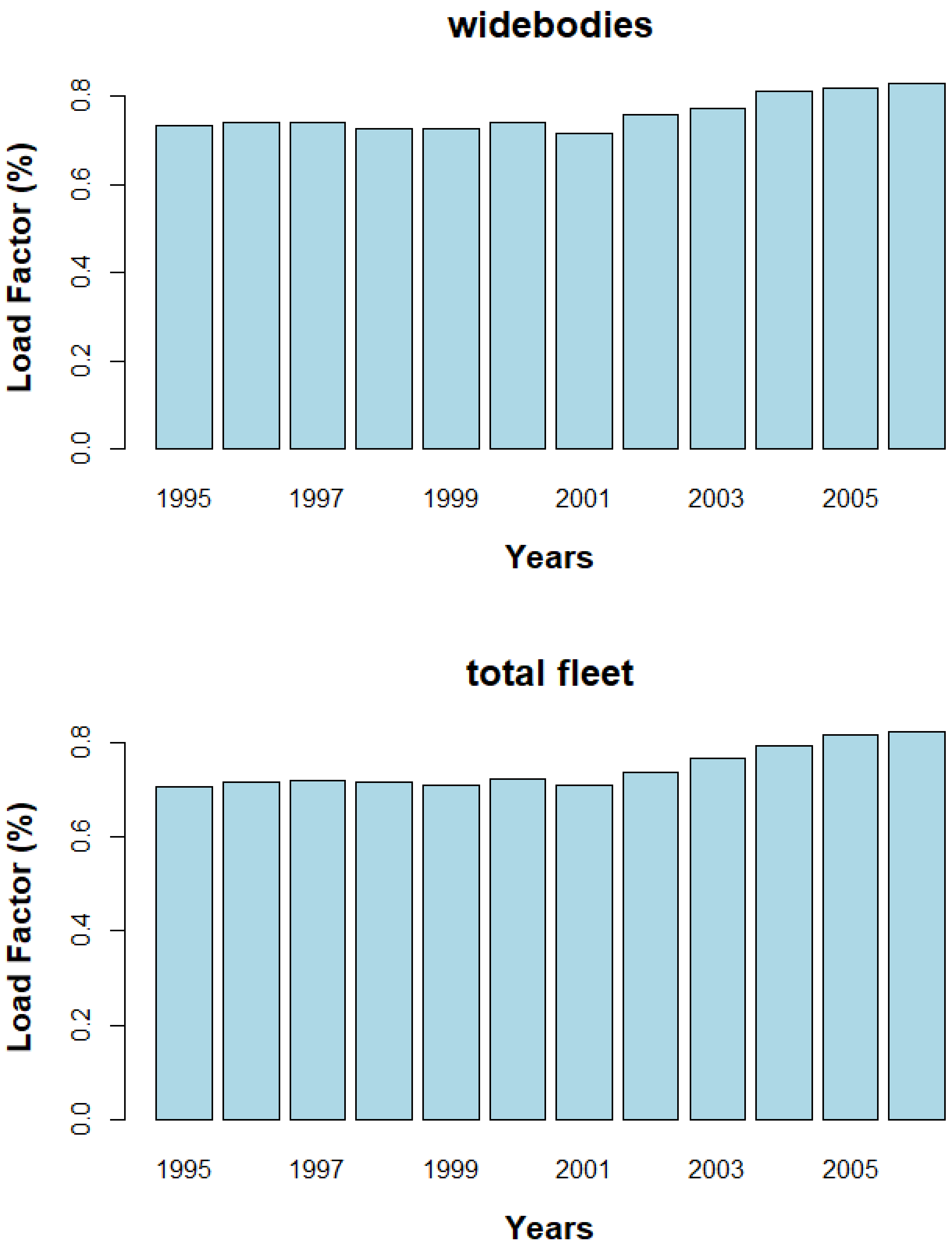
```
# Snigdha Das(ID_22024012)
# Changes should be made to the working directory for using it in other PC.
Rm(list = ls()) # Clearing  the environment
graphics.off()  # Closeing all open graphics windows

library(readxl)
library(graphics)

setwd("F:/Rstudio codes/MAT_4509")  # setting the directory name
# Variable with my ID
id <- 12

# Load dataset
xl_data <- read_excel("United Airlines Aircraft Operating Statistics- Cost Per Block Hour (Unadjusted).xls",
range = "b2:w158")

# Took the same number of data input as my ID's last two digits.
Maintenance_categories <- c("labor", "materials", "third party", "burden")
maintenance_rows <- c(16, 55, 94, 133)

years <- 1995 :(1995 + (id – 1))
load_factor_rows <- maintenance_rows + 18

# Function to extract data for a given row number (Maintenance/Load Factor)
get_data_by_row <- function(row_num) {
  return((as.numeric(xl_data[row_num, 1:(id) + 1])))
}

get_maintenance_category <- function(row_num) {
  labor <- get_data_by_row(row_num + 1)
  materials <- get_data_by_row(row_num + 2)
  third_party <- get_data_by_row(row_num + 3)
  burden <- get_data_by_row(row_num + 5)
  return(setNames(
    c(sum(labor), sum(materials), sum(third_party), sum(burden)),
    maintenance_categories
  ))
}

# For plotting Load Factor bar plot
plot_bar <- function(data, title) {
  barplot(data,
        xlab = "Years",
        ylab = "Load Factor (%)",
        col = "lightblue",
        border = "black",
        main = title,
        cex.main = 1.5,  # Increase main title size
```

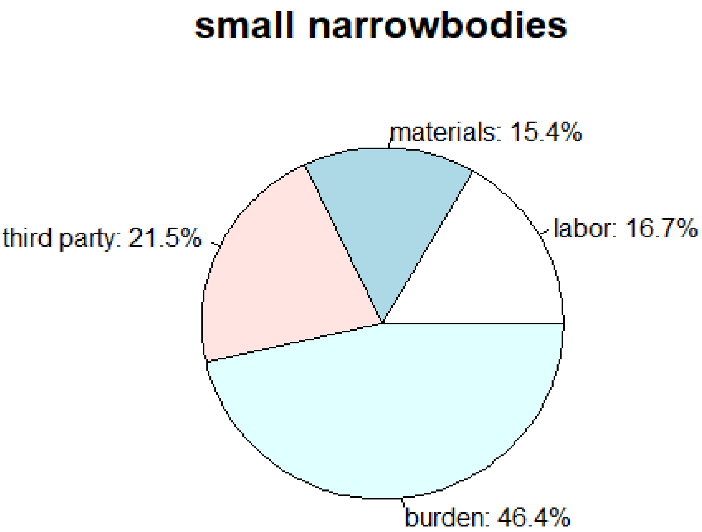
```
        cex.lab = 1.3,   # Increase axis label size
        font.main = 2,   # Bold main title
        font.lab = 2     # Bold axis labels
  )
}

fleet_category <- c("small narrowbodies", "large narrowbodies",
                    "widebodies", "total fleet")

# Function to create pie chart with percentages
plot_pie_with_percentages <- function(data, main_title) {
  percentages <- round(data / sum(data) * 100, 1)  # Calculate percentages
  labels <- paste0(names(data), ": ", percentages, "%")  # Create labels with percentages
  pie(data, labels = labels, main = main_title,
      cex.main = 1.5,       # Increase main title size
      cex.lab = 1.3,        # Increase label size
      font.main = 2,        # Bold main title
      font.axis = 2         # Bold axis font (though not typically used in pie charts)
  )
}

# Plot pie charts for maintenance in separate figures
for (I in 1:4) {
  data_1 <- get_maintenance_category(maintenance_rows[i])
  windows()  # Open a new window for Windows OS
  # quartz()  # Uncomment this line if you're on macOS
  plot_pie_with_percentages(data_1, fleet_category[i])  # Call the new function
  Sys.sleep(2)  # Pause to view the pie chart
}

# Plot bar charts for load factor in separate figures
for (I in 1:4) {
  data_2 <- setNames(get_data_by_row(maintenance_rows[i] + 18), years)
  windows()  # Open a new window for Windows OS
  # quartz()  # Uncomment this line if you're on macOS
  plot_bar(data_2, fleet_category[i])
  Sys.sleep(2)  # Pause to view the bar chart
}
```

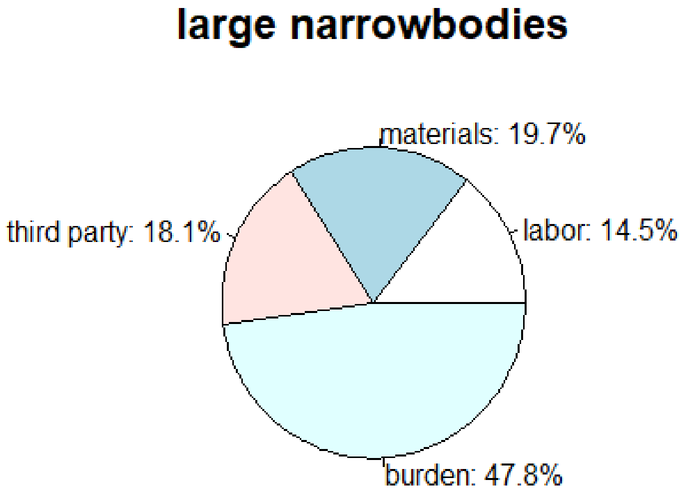
Code Output 1(iv):

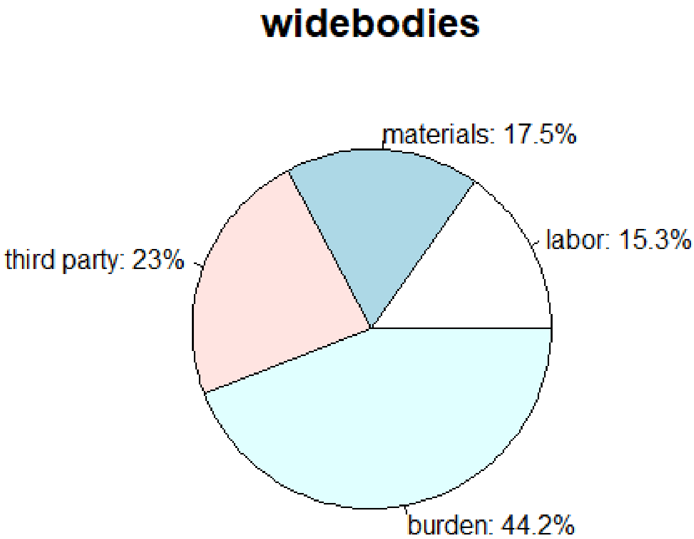


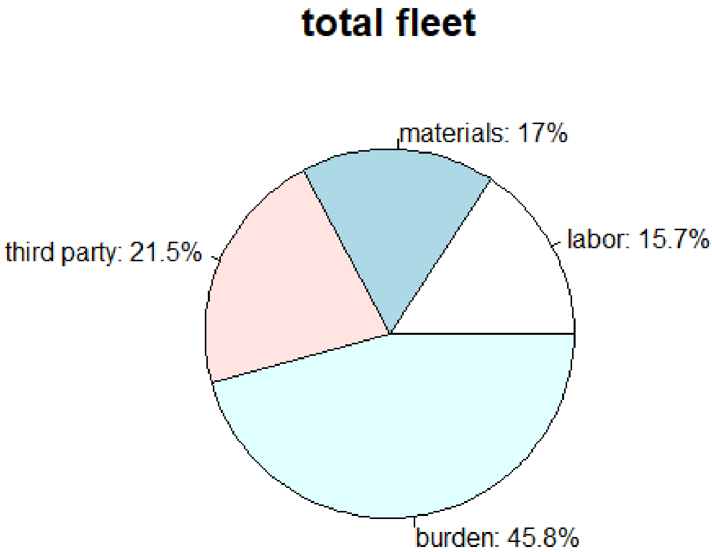



widebodies
Load Factor (%)
0.0
0.2
0.4
0.6
0.8
1995
1997
1999
2001
2003
2005
Years
total fleet
Load Factor (%)
0.0
0.2
0.4
0.6
0.8
1995
1997
1999
2001
2003
2005
Years









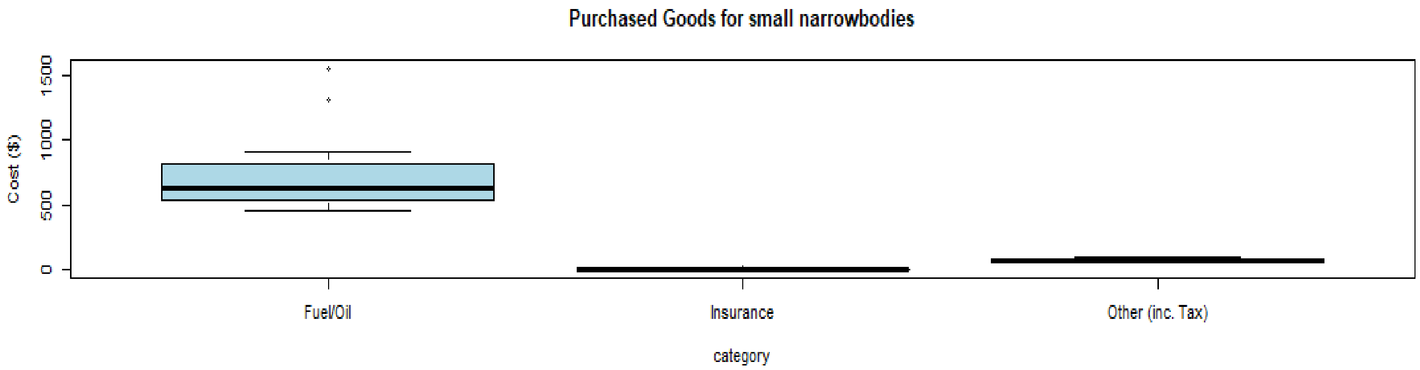
<u>Answer code 1(v):</u>

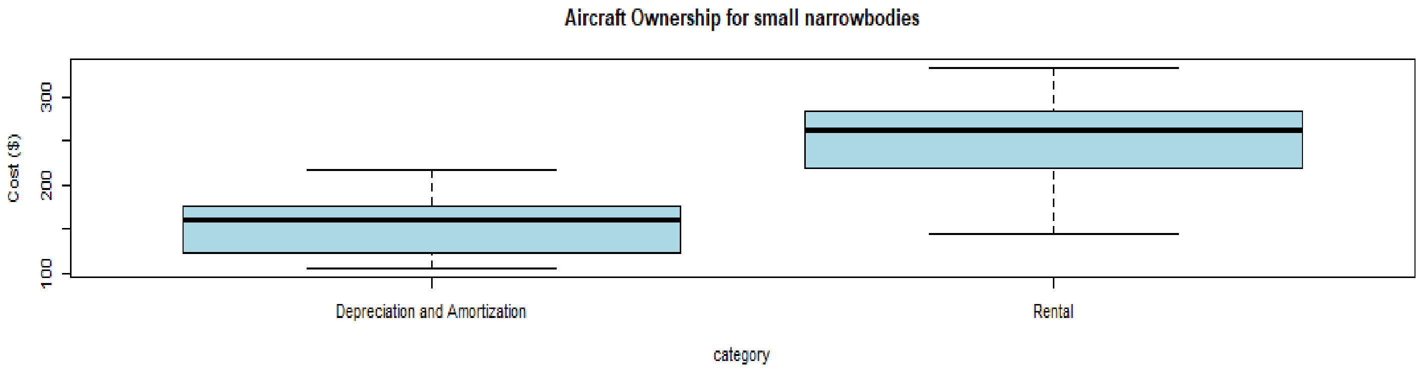
```
# Snigdha Das(ID_22024012)
# Changes should be made to the working directory for using it in other PC.
rm(list = ls()) # Clear the environment
graphics.off()  # Close all open graphics windows

library(readxl)
library(graphics)

# Load dataset
xl_data <- read_excel("United Airlines Aircraft Operating Statistics- Cost Per Block Hour (Unadjusted).xls",
range = "b2:w158")

# Variable with my ID
id <- 12  # This should be set based on your specific requirements

# Define categories
daily_utilization_categories <- c("Block hours", "Airborne hours", "Departures")
ownership_categories <- c("Rental", "Depreciation and Amortization")
```

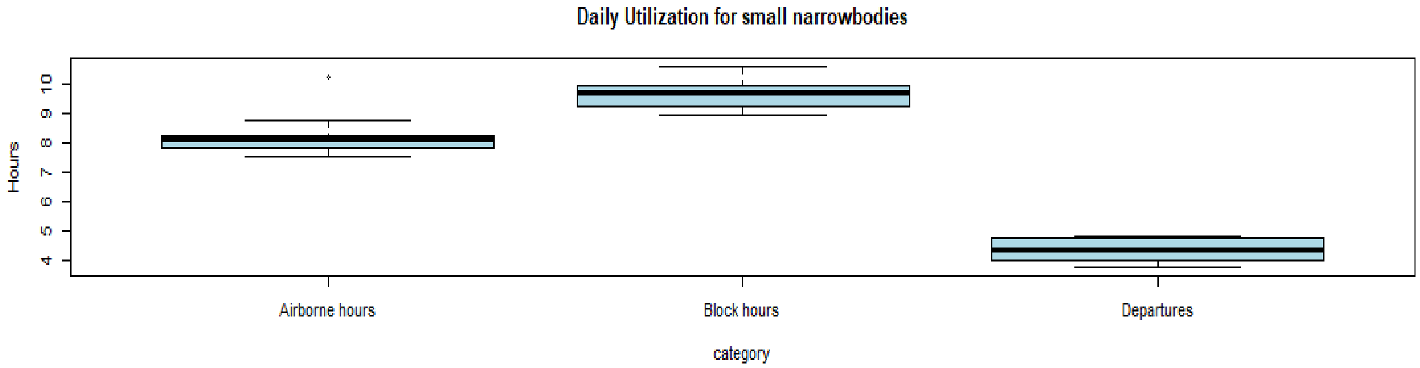
```
purchased_goods_categories <- c("Fuel/Oil", "Insurance", "Other (inc. Tax)")
fleet_category <- c(
  "small narrowbodies",
  "large narrowbodies",
  "widebodies",
  "total fleet"
)

# Row numbers for each category
purchased_goods_rows <- c(16, 55, 94, 133) - 5
ownership_rows <- purchased_goods_rows + 12
daily_utilization_rows <- ownership_rows + 13

get_data_by_row <- function(row_num) {
  return((as.numeric(xl_data[row_num, 1:(id) + 1])))  # Using 'id' to select number of values
}

get_category_data <- function(row_num, categories) {
  rows_data <- lapply(
    seq_along(categories),
    function(i) get_data_by_row(row_num + i)
  )
  costs <- unlist(rows_data)
  category <- factor(rep(categories, sapply(rows_data, length)))
  return(data.frame(costs = costs, category = category))
}

# Function to create box plots
box_plot <- function(data, title, ylab) {
  boxplot(costs ~ category,
          data = data,
          col = "lightblue",
          ylab = ylab,
          main = title,
          border = "black"
  )
}

# Plot all box plots for each fleet category in separate windows
for (i in 1:4) {
  windows(width = 15, height = 10)  # Open a new window with increased dimensions

  # Set up layout for 3 rows (1 for each category)
  par(mfrow = c(3, 1))  # 3 rows, 1 column

  # Plot Purchased Goods
  box_plot(
    get_category_data(purchased_goods_rows[i], purchased_goods_categories),
    paste("Purchased Goods for", fleet_category[i]),
```

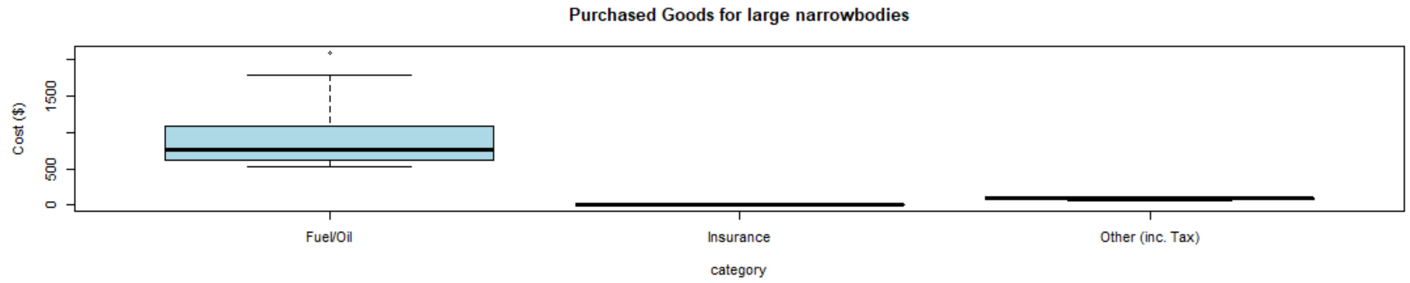
```
    "Cost ($)"
  )

  # Plot Aircraft Ownership
  box_plot(
    get_category_data(ownership_rows[i], ownership_categories),
    paste("Aircraft Ownership for", fleet_category[i]),
    "Cost ($)"
  )

  # Plot Daily Utilization
  box_plot(
    get_category_data(daily_utilization_rows[i], daily_utilization_categories),
    paste("Daily Utilization for", fleet_category[i]),
    "Hours"
  )
}
```

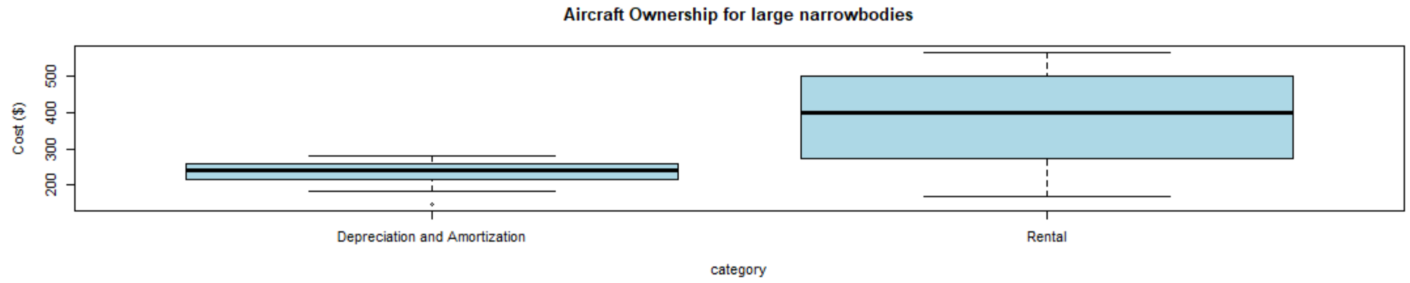
Code Output 1(v):

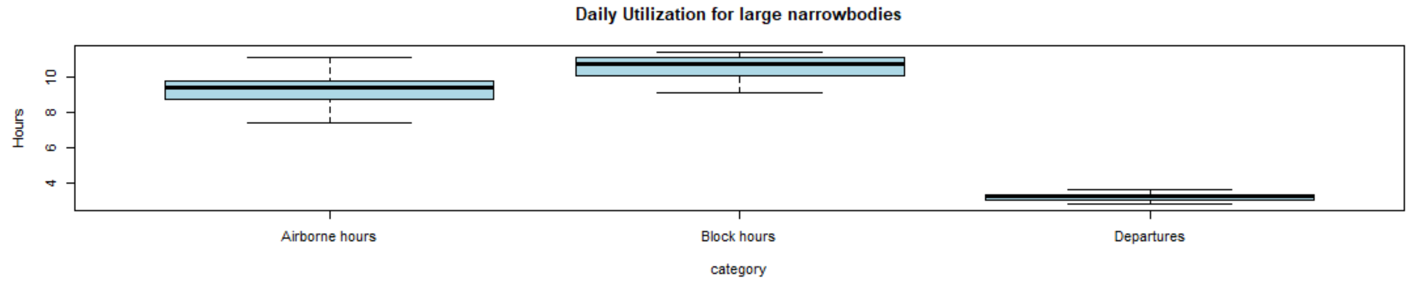


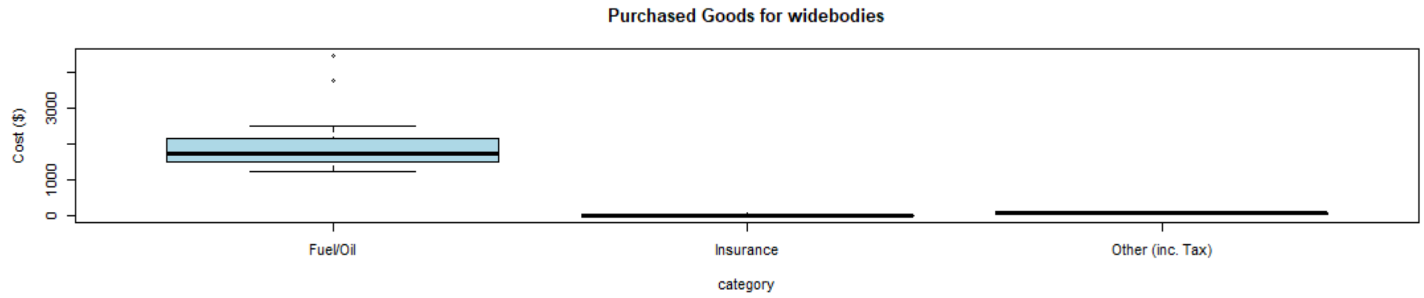


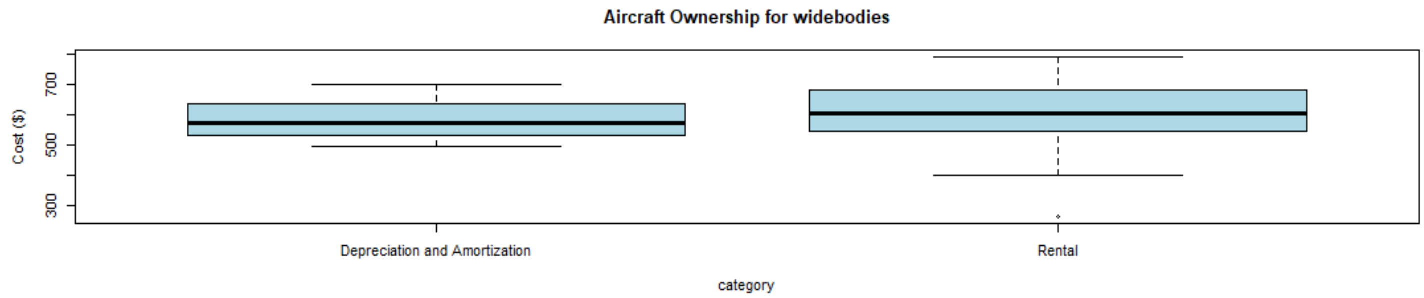


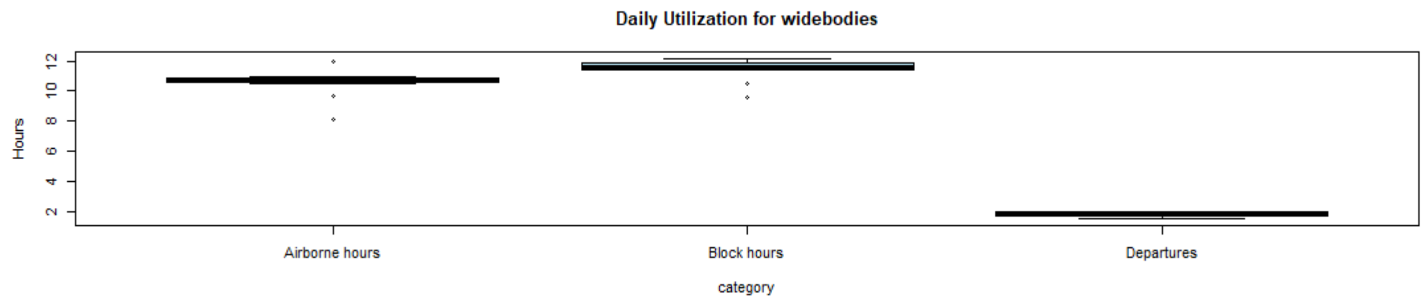

Purchased Goods for large narrowbodies
Cost ($)
0
500
1500
Fuel/Oil
Insurance
Other (inc. Tax)
category

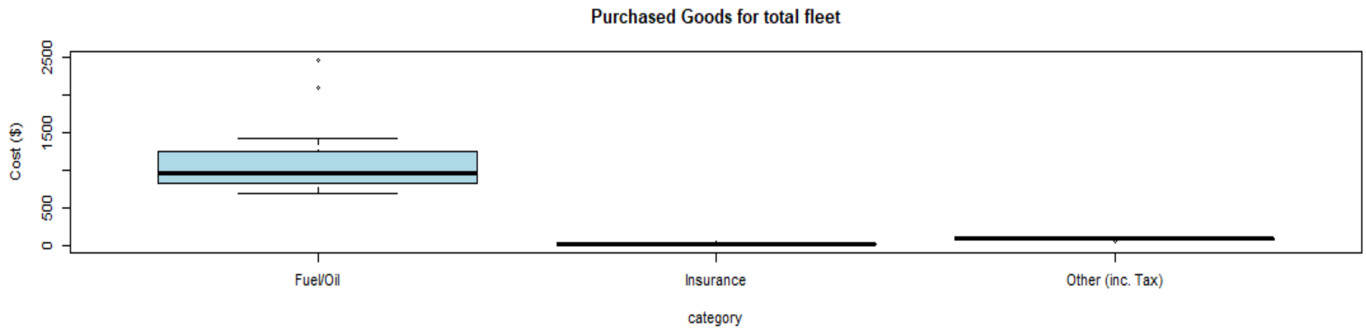

Aircraft Ownership for large narrowbodies
Cost ($)
200
300
400
500
Depreciation and Amortization
Rental
category

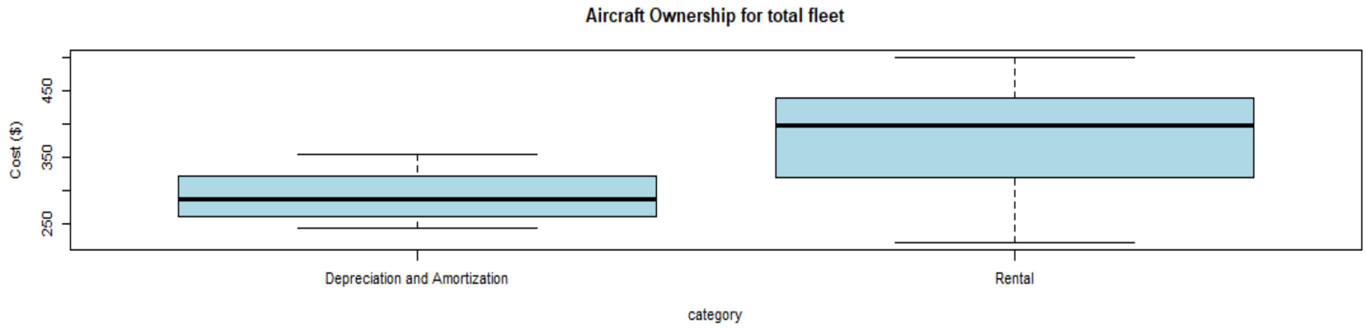

Daily Utilization for large narrowbodies
Hours
4
6
8
10
Airborne hours
Block hours
Departures
category

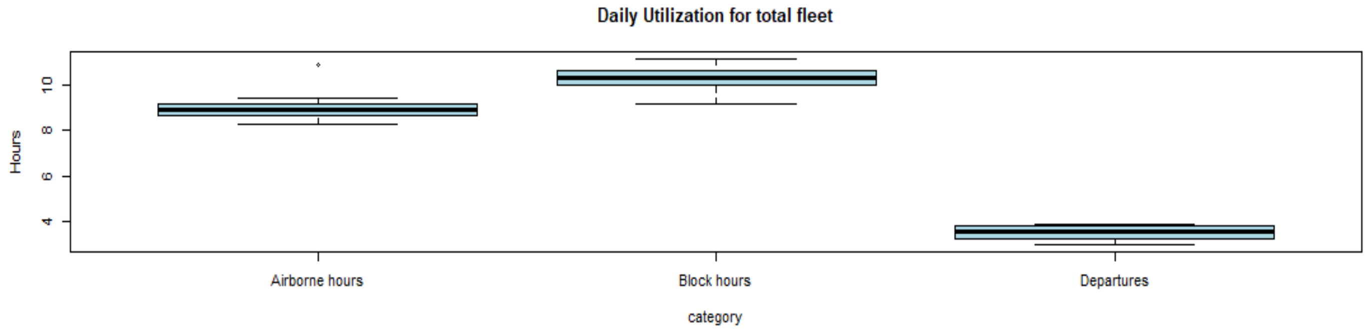

Purchased Goods for widebodies
Cost ($)
0
1000
3000
Fuel/Oil
Insurance
Other (inc. Tax)
category


Aircraft Ownership for widebodies
Cost ($)
300
500
700
Depreciation and Amortization
Rental
category


Daily Utilization for widebodies
Hours
2
4
6
8
10
12
Airborne hours
Block hours
Departures
category







__Analysis of Question 1(vi):__

From the box plots , information on costs and daily metrics of utilization of aircrafts within the categorized type (small narrowbodies, large narrowbodies, wide bodies, total fleets is displayed. Here's a concise overview:

1. Purchased Goods

Small Narrowbodies:

Fuel/Oil: These costs fluctuate with their extreme values but are not important in this case while Insurance and Other (inc. Tax) while they are important cost indices are steady with slight effects.

Large Narrowbodies:

Higher for Fuel/Oil compared with small narrow bodies due, perhaps, to higher operation requirements, while its Insurance fluctuation is comparably less volatile as in small narrow bodies.

Widebodies:

Much higher relative to size and operation requirements the Fuel/Oil costs in comparing of the two.

Total Fleet:

Comparable trends are observable with high fuel prices, indicating that greater planes affect total costs.

2. Aircraft Ownership

Small Narrowbodies:

In this area, costs are slightly lower compared to Depreciation, which means more preference at the ownership level.

Large Narrowbodies:

Planes are in general more expensive than small narrow bodies in year one and also in year five.

Widebodies:

Both categories experienced cost increments since the development of these aircraft required high expenditures.

Total Fleet:

In line with the hypothesis and similar to previous studies revealing that larger fleets have higher ownership costs.

3. Daily Utilization

Small Narrowbodies:

Total Flight Hours by subcategories of Airborne hours and Block hours are fairly constant, but the Departures one is lower – it means fewer flights.

Large Narrowbodies:

Like other applications, such usage profiles, have marginally better usage density relative to smaller aircraft.

Widebodies:

This means higher utilization for longer flights and fewer frequencies.

Total Fleet:

This aspect involves the adoption characteristics of wide different bodies that depict efficient use of larger aircraft.

Conclusion

The box plots show the financial effects of sustaining different sorts of aircraft that show variability in the fuel cost and the steady experience of the ownership cost. They give a reliable insight into how operating parameters vary with the kind of aircraft, a situation that explains the general approaches to fleet planning and cost management.