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?

Answer:

Source code-1(i, ii, iii):

```
library(readxl)


# URL of the Excel file in the GitHub repository

github_url <- "https://github.com/jamee47/MAT-4509/raw/refs/heads/main/United%20Airlines%20Aircraft%20Operating%20Statistics-%20Cost%20Per%20Block%20Hour%20(Unadjusted).xls"


# Download the Excel file temporarily

temp_file <- tempfile(fileext = ".xls")

download.file(github_url, temp_file, mode = "wb")


# Verify if the file exists

if (file.exists(temp_file)) {

  cat("File downloaded successfully.\n")

} else {

  stop("Failed to download the file from GitHub.")

}

all_data <- read_excel(data_file, range = "B2:W158")

all_data

# Helper function to extract salary data by row

get_salary_wages <- function(row_num, data = all_data) {

  return(na.omit(as.numeric(data[row_num, -1])))

}

get_salary_wages

# Ensure that salary_wages data has 28 points

# Extract salary data from the dataset using get_salary_wages()

salary_wages_snbodies <- get_salary_wages(6)    # For small narrowbodies

salary_wages_lnbodies <- get_salary_wages(45)   # For large narrowbodies

salary_wages_wbodies <- get_salary_wages(84)    # For widebodies

salary_wages_tfleet <- get_salary_wages(123)    # For total fleet

# Now, combine the extracted salary data into one sample

# Assuming you want a combined sample of all these salary data sets

salary_wages_sample <- c(salary_wages_snbodies, salary_wages_lnbodies, salary_wages_wbodies, salary_wages_tfleet)

salary_wages_sample

# Check the number of observations in the combined sample

length(salary_wages_sample)
```

```
# If you need exactly 20 observations, you can either take the first 20, sample 20 randomly, or apply some selection method.

# Randomly select 20 observations from the combined data

set.seed(123)  # For reproducibility

salary_wages_sample_20 <- sample(salary_wages_sample, 20, replace = FALSE)

salary_wages_sample_20

# View the sample of 20 observations

print(salary_wages_sample_20)

## Assuming these functions and salary data extraction methods have already been defined:

get_modes <- function(data) {

  freq_table <- table(data)

  max_freq <- max(freq_table)

  modes <- as.numeric(names(freq_table[freq_table == max_freq]))

  if (length(modes) == length(data)) {

    return(NULL)

  }

  return(modes)

}

get_frequency_distribution <- function(wage_data) {

  # Number of observations

  n <- length(wage_data)

  # Calculate k directly as log2(n), and round it up

  k <- ceiling(log2(n))

  # Calculate 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)  # Ensure class interval is a whole number

  # Create breakpoints

  break_points <- seq(

    min_salary - (class_interval / 2),  # Start the first break point slightly before the min value

    max_salary + (class_interval / 2),  # End the last break point slightly after the max value

    by = class_interval

  )

  # Create frequency distribution

  salary_bins <- cut(wage_data, breaks = break_points, right = TRUE)

  frequency_distribution <- table(salary_bins)

  return(frequency_distribution)

}

# Example: Extract salary data from each category

salary_wages_snbodies <- get_salary_wages(6)    # For small narrowbodies

salary_wages_lnbodies <- get_salary_wages(45)   # For large narrowbodies

salary_wages_wbodies <- get_salary_wages(84)    # For widebodies
```

```
salary_wages_tfleet <-
get_salary_wages(123)    # For total fleet

# Combine the extracted salary data

combined_salary_wages <-
c(salary_wages_snbodies,
salary_wages_lnbodies,
salary_wages_wbodies,
salary_wages_tfleet)

# If you want to take exactly 20 samples,
you can sample from combined data

set.seed(123)  # For reproducibility

salary_wages_sample_20 <-
sample(combined_salary_wages, 20,
replace = FALSE)

# Get frequency distribution for the sample
data

frequency_distribution_sample <-
get_frequency_distribution(salary_wages_s
ample_20)

# Print the frequency distribution for the
sample

cat("Frequency Distribution for Sample of
20 Observations:\n")

print(frequency_distribution_sample)

# Perform analysis on the sample

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

  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")

}

# Perform analysis on the 15 sample

print_analysis(salary_wages_sample_15,
"Salary Wages Sample of 15 Observations")

# Plot the histogram for the sample data

plot_histogram <-
function(frequency_distribution,
window_title) {

  barplot(frequency_distribution,
```

```
        xlab = "Salary Ranges",
        ylab = "Frequency",
        col = "lightblue",
        border = "black",
        space = 0,  # No space between bars
        width = 1,  # Adjust width to fill the space better
        main = window_title
  )
}
```

Plot histogram for the sample frequency distribution

plot_histogram(frequency_distribution_sample, "Histogram of Salary Wages Sample of 20 Observations")

output 1(i):

Print the frequency distribution for the sample

> cat("Frequency Distribution for Sample of 15 Observations:\n")

Frequency Distribution for Sample of 15 Observations:

> print(frequency_distribution_sample)

```
salary_bins
    (244,406]     (406,568]     (568,730]
            5             7             4
    (730,892] (892,1.05e+03]
            3             0
```

Output 1(ii):

print_analysis(salary_wages_sample_15, "Salary Wages Sample of 15 Observations")

Analysis of Salary Wages Sample of 15 Observations ::

Mean: 566.7791

Median: 542.6755

Modes: None

Sample Standard Deviation: 188.4018

Sample Variance: 35495.23

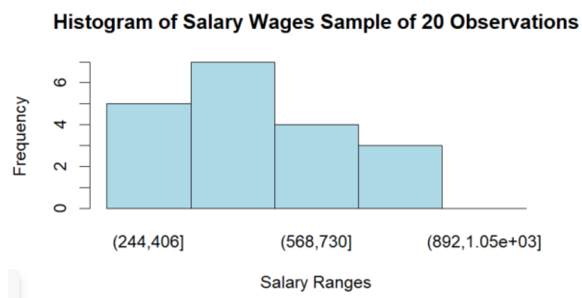
Quartiles (Q1, Q2, Q3): 426.2836 542.6755 651.3924

10th Percentile: 368.8777

9th Decile: 741.1863

Range: 806.1665

Output 1(iii):



Source code 1(iv):

library(readxl)

library(RColorBrewer)

URL of the Excel file in the GitHub repository

github_url <- "https://github.com/jamee47/MAT-4509/raw/refs/heads/main/United%20Airlines%20Aircraft%20Operating%20Statistics-%20Cost%20Per%20Block%20Hour%20(Unadjusted).xls"

Download the Excel file temporarily

temp_file <- tempfile(fileext = ".xls")

```r
download.file(github_url, temp_file, mode =
"wb")

# Verify if the file exists
if (file.exists(temp_file)) {
  cat("File downloaded successfully.\n")
} else {
  stop("Failed to download the file from
GitHub.")
}
all_data <- read_excel(data_file, range =
"B2:W158")
all_data
maintenance_categories <- c("labor",
"materials", "third party", "burden")
years <- 1995:2015
# Function to extract data for a given row
number (Maintenance/Load Factor)
get_data_by_row <- function(row_num) {

return(na.omit(as.numeric(all_data[row_nu
m, -1])))
}
get_maintenace_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 ploting Load Factor bar plot
plot_bar <- function(data, title) {
  barplot(data,
          main = title,
          xlab = "Years",
          ylab = "Load Factor (%)",
          col = "lightgreen",
          border = "pink"
  )
}
# Maintenance and Load Factor row
numbers
maintenance_rows <- c(16, 55, 94, 133)
load_factor_rows <- c(34, 73, 112, 151)
fleet_category <- c(
  "small narrowbodies",
  "large narrowbodies",
  "widebodies",
  "total fleet"
)
# Load necessary library
library(RColorBrewer)  # For color palettes
# Pie chart for maintenance
windows(width = 1920 / 100, height = 1080 /
100) # Set window size
par(mfrow = c(2, 2), oma = c(0, 0, 3, 0))
# Define a color palette
```

```r
colors <- brewer.pal(4, "Set3")  # Using RColorBrewer for a set of 4 distinct colors

# Create pie charts for each maintenance category with enhancements

lapply(1:4, function(i) {

  data <- get_maintenace_category(maintenance_rows[i])

  # Calculate percentages

  percentages <- round(100 * data / sum(data), 1)

  # Create labels with category names and percentages

  labels <- paste0(names(data), ": ", percentages, "%")


  # Create pie chart

  pie(data,

      labels = labels,        # Use labels with percentages

      main = paste("Maintenance Costs for", fleet_category[i]),  # Descriptive title

      col = colors,           # Set colors for slices

      border = "black")       # Add border to slices

})

# Add an outer title for all pie charts

mtext("Maintenance Cost Distribution", outer = TRUE, cex = 1.5)

# Reset plotting parameters to default

par(mfrow = c(1, 1))

# bar chart for load factor

windows(width = 1920 / 100, height = 1080 / 100) # Set window size

par(mfrow = c(2, 2), oma = c(0, 0, 3, 0))

lapply(1:4, function(i) {

  data <- setNames(get_data_by_row(load_factor_rows[i]), years)

  plot_bar(data, fleet_category[i])

})

mtext("Load Factor", outer = TRUE, cex = 1.5)

par(mfrow = c(1, 1))
```

output 1(iv):

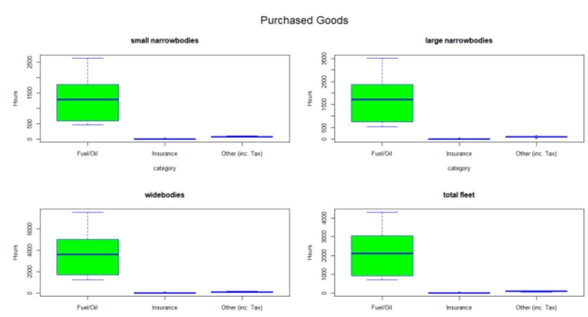
Source code 1(v):

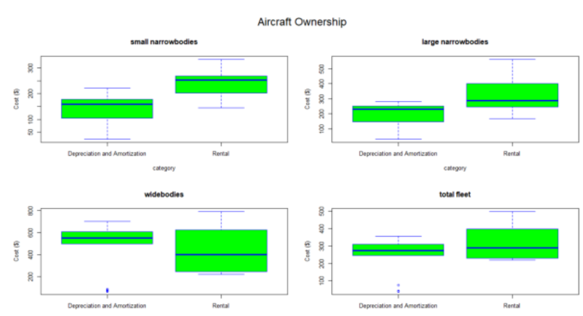
```r
library(readxl)

library(RColorBrewer)
```

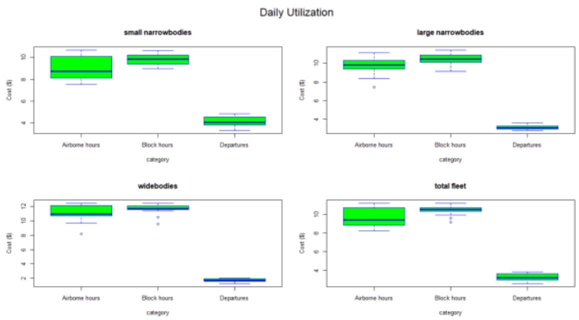
```
# URL of the Excel file in the GitHub repository

github_url <- "https://github.com/jamee47/MAT-4509/raw/refs/heads/main/United%20Airlines%20Aircraft%20Operating%20Statistics-%20Cost%20Per%20Block%20Hour%20(Unadjusted).xls"


# Download the Excel file temporarily

temp_file <- tempfile(fileext = ".xls")

download.file(github_url, temp_file, mode = "wb")


# Verify if the file exists

if (file.exists(temp_file)) {

  cat("File downloaded successfully.\n")

} else {

  stop("Failed to download the file from GitHub.")

}all_data <- read_excel(data_file, range = "B2:W158")

all_data

# 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

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) {

  if (row_num > nrow(all_data)) {

    stop("Row number exceeds data range.")

  }


return(na.omit(as.numeric(all_data[row_num, -1])))

}

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))

}

box_plot <- function(data, title, ylab) {

  boxplot(costs ~ category,

        data = data,
```

```
        main = title,
        col = "green",
        ylab = ylab,
        border = "blue"
  )
}
plot_category <- function(rows, categories, title, ylab) {
  windows(width = 1920 / 100, height = 1080 / 100) # Set window size
  par(mfrow = c(2, 2), oma = c(0, 0, 3, 0))
  # Create the box plot using the formula interface
  lapply(
    seq_along(rows),
    function(i) {
      box_plot(
        get_category_data(
          rows[i], categories
        ), fleet_category[i], ylab
      )
    }
  )
  mtext(title, outer = TRUE, cex = 1.5)
  par(mfrow = c(1, 1))
}
plot_category(
  purchased_goods_rows,
  purchased_goods_categories,
  "Purchased Goods",
  "Hours"
)
plot_category(
  ownership_rows,
  ownership_categories,
  "Aircraft Ownership",
  "Cost ($)"
)
plot_category(
  daily_utilization_rows,
  daily_utilization_categories,
  "Daily Utilization",
  "Cost ($)"
)
```

Output 1(v):







Vi. Findings from i to v

1. purchased goods:

Small narrowbodies:

"Fuel/Oil" has a smaller range (up to ~2,500 hours) compared to other types.

Widebodies:

"Fuel/Oil" exhibits the highest range (up to ~6,000 hours), reflecting the greater fuel usage of these larger aircraft.

Large narrowbodies and Total fleet:

Both show similar patterns, with "Fuel/Oil" distribution peaking around ~3,500 hours.

2. Aircraft Ownership:

Small narrowbodies:

Both *Depreciation* and *Rental* costs are relatively low compared to other fleets.

Median costs for *Depreciation* are slightly below $200, while *Rental* costs have a wider range (~$100–$300).

Large narrowbodies:

Rental costs exhibit higher variability, with a range extending up to ~$500.

Depreciation is concentrated around $200–$300, showing less variability.

Widebodies:

Depreciation is significantly higher compared to other fleet types, with medians around ~$600.

Rental costs for widebodies also show a wide range, from ~$200 to ~$800, with some outliers.

Total fleet:

The combined trends align with individual fleet types, with *Rental* consistently higher and more variable than *Depreciation*.

3. Daily Utilization:

Small Narrowbodies:

Airborne Hours and Block Hours have similar cost distributions, with median costs near $8–9.

Departures show a much lower cost range (around $4–6), indicating a significant difference in cost-per-departure compared to hours flown. **Large Narrowbodies**:

Similar trends to small narrowbodies, but with a slightly higher overall cost distribution for Airborne and Block Hours.

Departures remain in the lower range of cost.

Widebodies:

These aircraft incur higher costs across all metrics.

Airborne Hours and Block Hours exhibit a median cost close to $12, while Departures costs are again significantly lower (around $4–5).

Total Fleet:

The combined view of all aircraft types shows a trend consistent with individual categories:

Airborne Hours and Block Hours have relatively high median costs.

Departures have the lowest costs among the metrics.