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, 9^{th} Decile, 10^{th} 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?

Note: Comment on all your findings, charts, and diagrams.

```
#Cleaning and reading the csv file

# Read the CSV file

df <- read.csv("United Airlines Aircraft Operating Statistics- Cost Per Block Hour (Unadjusted).csv", header = TRUE)

df <- data.frame(lapply(df, function(x) {

 if (is.character(x)) {

  x <- gsub(",", "", x)  #comma removing

  x <- gsub("\\$", "", x)  #dollar symbol removing

 }

 return(x)

}))

# Save the cleaned data to a new CSV file

write.csv(df, "Cleaned_United_Airlines_Aircraft_Operating_Statistics.csv", row.names = FALSE

cleaned_df <- read.csv("Cleaned_United_Airlines_Aircraft_Operating_Statistics.csv", header = TRUE)
```

<u>Q1 and Q3:</u>

```
# Load the necessary library

library(dplyr)

# Define the datasets containing salaries and wages for each aircraft type

salary_data_small <- c(388, 410, 444, 439, 437, 484, 553, 568, 371, 327, 292, 293, 296, 325, 354, 358, 367, 414, 528, 615, 649)

salary_data_large <- c(495, 517, 583, 539, 546, 585, 629, 654, 440, 393, 350, 353, 368, 382, 401, 409, 429, 413, 542, 621, 648)

salary_data_wide <- c(759, 858, 990, 980, 949, 994, 1082, 1131, 739, 653, 584, 581, 595, 647, 656, 654, 684, 837, 922, 1087, 1096)

# Set the size of the sample

n <- 21
```

```r
# Function to perform frequency analysis
analyze_frequency <- function(salary_data, aircraft_type) {
 # Randomly sample 'n' elements from the salary data
 sampled_data <- sample(salary_data, n)
 # Use Sturges' formula to determine the number of class intervals
 k <- ceiling(log2(n) + 1)
 # Create intervals (breakpoints) for the frequency distribution
 intervals <- seq(min(sampled_data), max(sampled_data), length.out = k + 1)
 # Group the data into the defined intervals and generate a frequency table
 freq_table <- cut(sampled_data, breaks = intervals, include.lowest = TRUE, right = FALSE) %>%
  table()
 # Display the frequency table
 cat(sprintf("\nFrequency Table for %s Aircraft:\n", aircraft_type))
 print(freq_table)
 # Plot a histogram for the frequency distribution
 hist(sampled_data,
    breaks = intervals,
    col = switch(aircraft_type,
           "Small Narrow Body" = "lavender",
           "Large Narrow Body" = "pink",
           "Wide Body" = "skyblue"),
    border = "black",
    main = paste("Distribution of Salaries and Wages -", aircraft_type),
    xlab = "Salary Range",
    ylab = "Frequency")
}

# Analyze each aircraft type
par(mfrow = c(3,1))  # Arrange plots in 3 rows, 1 column
# Analyze Small Narrow Body Aircraft
analyze_frequency(salary_data_small, "Small Narrow Body")
# Analyze Large Narrow Body Aircraft
analyze_frequency(salary_data_large, "Large Narrow Body")
# Analyze Wide Body Aircraft
analyze_frequency(salary_data_wide, "Wide Body")
# Reset the plotting layout
```

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

OUTPUT:

Q2:

```
# Define salary datasets for all aircraft types

salary_data_small <- c(388, 410, 444, 439, 437, 484, 553, 568, 371, 327, 292, 293, 296, 325, 354, 358, 367, 414, 528, 615, 649)

salary_data_large <- c(495, 517, 583, 539, 546, 585, 629, 654, 440, 393, 350, 353, 368, 382, 401, 409, 429, 413, 542, 621, 648)

salary_data_wide <- c(759, 858, 990, 980, 949, 994, 1082, 1131, 739, 653, 584, 581, 595, 647, 656, 654, 684, 837, 922, 1087, 1096)


# Function to find mode (most frequent value)

calculate_mode <- function(values) {

 unique_vals <- unique(values)

 frequency_count <- tabulate(match(values, unique_vals))

 max_frequency <- max(frequency_count)

 if (max_frequency == 1) {

  return("No mode, all values occur only once")

 } else {

  return(unique_vals[which(frequency_count == max_frequency)])

 }

}

# Function to calculate all statistics for a given dataset

calculate_statistics <- function(salary_data, aircraft_type) {

 # Calculate basic statistics
```

```r
  avg_salary <- mean(salary_data)
  median_salary <- median(salary_data)
  mode_salary <- calculate_mode(salary_data)
  standard_deviation <- sd(salary_data)
  variance_salary <- var(salary_data)
  quartile_values <- quantile(salary_data)
  decile_90th <- quantile(salary_data, 0.9)
  percentile_10th <- quantile(salary_data, 0.1)
  range_salary <- diff(range(salary_data))
  # Compile results
  statistics_summary <- list(
    Aircraft_Type = aircraft_type,
    Average = avg_salary,
    Median = median_salary,
    Mode = mode_salary,
    Standard_Deviation = standard_deviation,
    Variance = variance_salary,
    Quartiles = quartile_values,
    `90th_Decile` = decile_90th,
    `10th_Percentile` = percentile_10th,
    Range = range_salary
  )
  return(statistics_summary)
}
# Calculate statistics for each aircraft type
small_narrow_stats <- calculate_statistics(salary_data_small, "Small Narrow Body")
large_narrow_stats <- calculate_statistics(salary_data_large, "Large Narrow Body")
wide_body_stats <- calculate_statistics(salary_data_wide, "Wide Body")
# Display results for each aircraft type
cat("\nStatistics for Small Narrow Body Aircraft:\n")
print(small_narrow_stats)
cat("\nStatistics for Large Narrow Body Aircraft:\n")
print(large_narrow_stats)
cat("\nStatistics for Wide Body Aircraft:\n")
print(wide_body_stats)
# Create a comparative summary of key metrics
```

```
comparative_summary <- data.frame(

  Aircraft_Type = c("Small Narrow Body", "Large Narrow Body", "Wide Body"),

  Average = c(small_narrow_stats$Average, large_narrow_stats$Average, wide_body_stats$Average),

  Median = c(small_narrow_stats$Median, large_narrow_stats$Median, wide_body_stats$Median),

  StdDev = c(small_narrow_stats$Standard_Deviation, large_narrow_stats$Standard_Deviation,
wide_body_stats$Standard_Deviation),

  Range = c(small_narrow_stats$Range, large_narrow_stats$Range, wide_body_stats$Range)

)
```

OUTPUT:

```
> print(small_narrow_stats)
$Aircraft_Type
[1] "Small Narrow Body"

$Average
[1] 424.381

$Median
[1] 410

$Mode
[1] "No mode, all values occur only once"

$Standard_Deviation
[1] 106.4836

$Variance
[1] 11338.75

$Quartiles
  0%  25%  50%  75% 100%
 292  354  410  484  649

$`90th_Decile`
90%
568

$`10th_Percentile`
10%
296

$Range
[1] 357
```

```
> print(large_narrow_stats)
$Aircraft_Type
[1] "Large Narrow Body"

$Average
[1] 490.3333

$Median
[1] 495

$Mode
[1] "No mode, all values occur only once"

$Standard_Deviation
[1] 103.915

$Variance
[1] 10798.33

$Quartiles
  0%  25%  50%  75% 100%
 350  401  495  583  654

$`90th_Decile`
90%
629

$`10th_Percentile`
10%
368

$Range
[1] 304
```

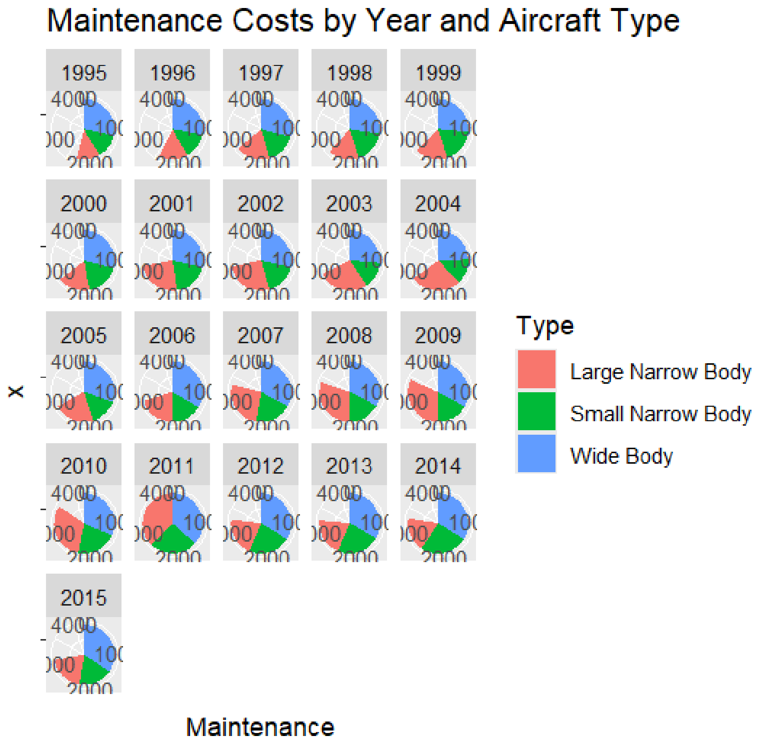
```
> print(wide_body_stats)
$Aircraft_Type
[1] "Wide Body"

$Average
[1] 832.2857

$Median
[1] 837

$Mode
[1] "No mode, all values occur only once"

$Standard_Deviation
[1] 190.7491

$Variance
[1] 36385.21

$Quartiles
  0%  25%  50%  75% 100%
 581  654  837  990 1131

$`90th_Decile`
 90%
1087

$`10th_Percentile`
10%
595

$Range
[1] 550
```

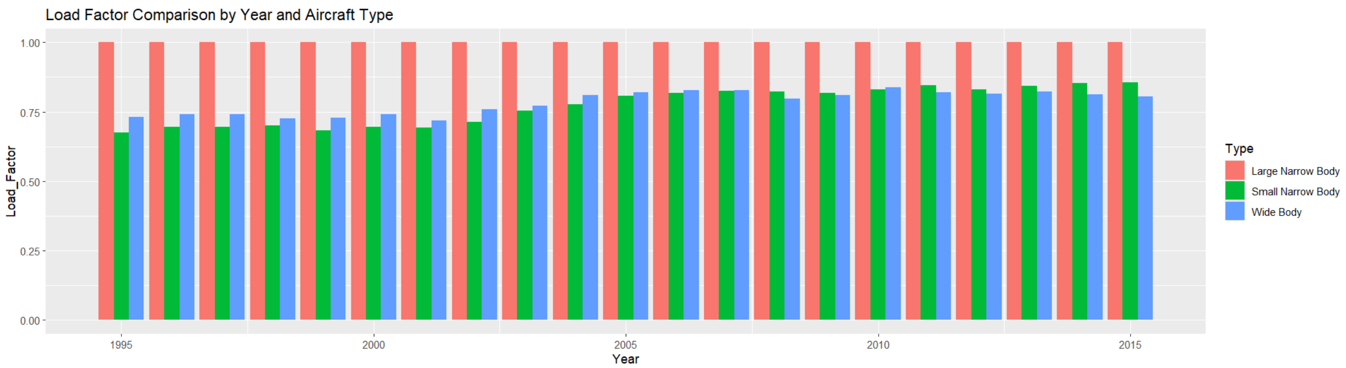
Q4:

```
# Import necessary libraries

library(ggplot2)

library(reshape2)

library(gridExtra)

library(dplyr)

library(tidyr)

# Seed for reproducibility

set.seed(123)

# Data setup for different aircraft types

# Data for Small Narrow Body Aircraft

small_narrow_data <- data.frame(

  Year = 1995:2015,

  Maintenance = c(552, 588, 696, 737, 788, 757, 808, 710, 594, 553, 614, 721, 833, 750, 712, 913, 1103, 961,
993, 1076, 774),

  Load_Factor = c(0.676, 0.696, 0.696, 0.701, 0.683, 0.696, 0.694, 0.712, 0.754, 0.777, 0.808, 0.818, 0.826,
0.822, 0.818, 0.830, 0.845, 0.831, 0.844, 0.854, 0.855),
```

```
  Type = "Small Narrow Body"
)
# Data for Large Narrow Body Aircraft
large_narrow_data <- data.frame(
  Year = 1995:2015,
  Maintenance = c(532, 652, 775, 711, 739, 734, 1032, 1117, 1148, 1155, 939, 834, 1108, 1262, 1334, 1323, 1578, 867, 842, 770, 815),
  Load_Factor = rep(1, 21),
  Type = "Large Narrow Body"
)
# Data for Wide Body Aircraft
wide_body_data <- data.frame(
  Year = 1995:2015,
  Maintenance = c(1196, 1176, 1225, 1159, 1116, 1225, 1206, 1203, 1112, 1038, 1281, 1410, 1386, 1371, 1395, 1331, 1538, 1421, 1391, 1426, 1446),
  Load_Factor = c(0.732, 0.741, 0.740, 0.726, 0.728, 0.740, 0.717, 0.758, 0.773, 0.811, 0.820, 0.829, 0.829, 0.798, 0.809, 0.839, 0.819, 0.816, 0.822, 0.813, 0.805),
  Type = "Wide Body"
)
# Merge data for all aircraft types
combined_data <- bind_rows(small_narrow_data, large_narrow_data, wide_body_data)


# Pie chart for Maintenance, showing distribution by aircraft type for each year
p1 <- ggplot(combined_data, aes(x = "", y = Maintenance, fill = Type)) +
  geom_bar(stat = "identity") +
  coord_polar(theta = "y") +
  facet_wrap(~Year) +
  labs(title = "Maintenance Costs by Year and Aircraft Type")
# Bar chart for Load Factor, grouped by Year and Aircraft Type
p2 <- ggplot(combined_data, aes(x = Year, y = Load_Factor, fill = Type)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Load Factor Comparison by Year and Aircraft Type")
# Display the charts in a grid
grid.arrange(p1, p2, ncol = 1)
```

OUTPUT:





Q5:

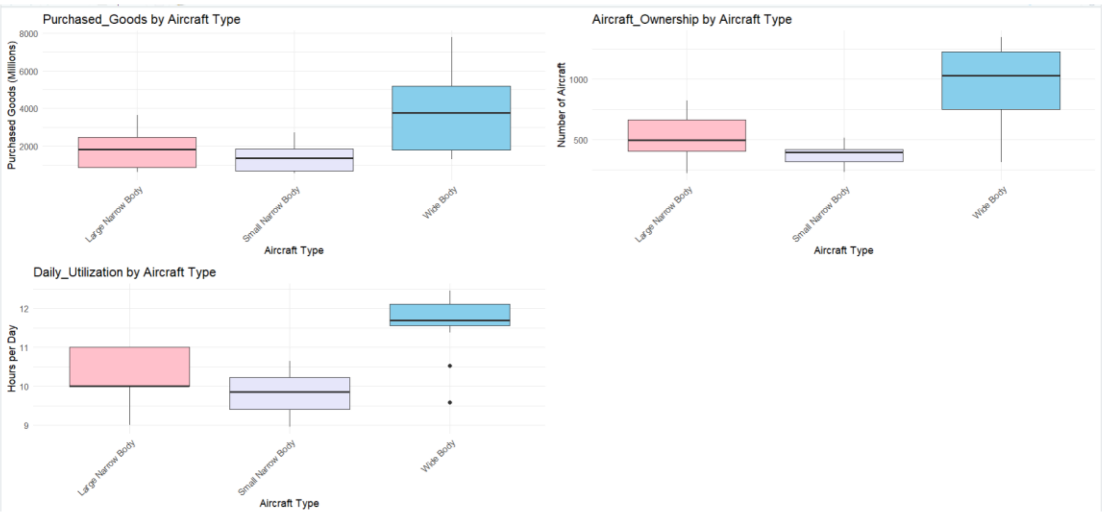
```
# Load required libraries

library(ggplot2)

library(reshape2)

library(gridExtra)  # For arranging multiple plots

# Set seed for reproducibility

set.seed(123)

# Create data frames for each aircraft type

small_narrow_data <- data.frame(

  Year = 1995:2015,

  Purchased_Goods = c(563, 693, 665, 572, 544, 769, 796, 652, 770, 981, 1384, 1633, 1681, 2726, 1357, 1858, 2335, 2583, 2410, 2297, 1538),

  Aircraft_Ownership = c(445,411,387,396,408,444,485,512,399,358,315,366,391,414,435,379,229,295,280,293,291),
```

```
  Daily_Utilization =
c(9.28,9.40,9.59,9.78,9.95,9.85,9.22,8.96,9.00,9.81,10.36,10.57,10.65,9.96,10.23,10.46,10.39,10.01,10.07
,9.65,9.18),

  Aircraft_Type = "Small Narrow Body"

)

large_narrow_data <- data.frame(

  Year = 1995:2015,

  Purchased_Goods =
c(644,784,756,650,631,864,961,873,1021,1354,1897,2200,2236,3644,1844,2478,3124,3166,2951,2775,18
36),

  Aircraft_Ownership =
c(821,769,737,661,575,581,769,734,590,465,403,385,418,467,516,492,424,329,295,274,217),

  Daily_Utilization = c(11,11,11,11,11,11,10,10,9,10,11,11,11,10,10,10,10,10,10,10,10),

  Aircraft_Type = "Large Narrow Body"

)


wide_body_data <- data.frame(

  Year = 1995:2015,

  Purchased_Goods =
c(1513,1788,1683,1356,1309,1815,1935,1701,1927,2660,3953,4634,4884,7761,3762,5188,6521,7017,673
7,6406,4216),

  Aircraft_Ownership =
c(1340,1344,1215,1140,1121,1235,1232,1225,1266,1165,1029,815,793,787,837,750,313,369,386,348,312
),

  Daily_Utilization =
c(11.54,11.38,11.48,11.69,11.94,11.86,11.56,10.53,9.59,11.61,12.07,12.14,12.41,12.45,12.09,12.28,12.10
,12.25,11.65,11.68,11.62),

  Aircraft_Type = "Wide Body"

)


# Combine all data frames

all_data <- rbind(small_narrow_data, large_narrow_data, wide_body_data)


# Reshape data for ggplot

data_long <- melt(all_data, id.vars = c("Year", "Aircraft_Type"))


# Load required libraries

library(ggplot2)

library(reshape2)

library(gridExtra)
```

Create function to generate box plot

```
create_boxplot <- function(data, var_name, y_label) {
 ggplot(subset(data, variable == var_name),
     aes(x = Aircraft_Type, y = value, fill = Aircraft_Type)) +
  geom_boxplot() +
  scale_fill_manual(values = c("Large Narrow Body" = "pink",
                "Small Narrow Body" = "lavender",
                "Wide Body" = "skyblue")) +
  labs(title = paste(var_name, "by Aircraft Type"),
     x = "Aircraft Type",
     y = y_label) +
  theme_minimal() +
  theme(legend.position = "none",
     axis.text.x = element_text(angle = 45, hjust = 1))
}

# Create the three plots using the function
plots <- list(
 create_boxplot(data_long, "Purchased_Goods", "Purchased Goods (Millions)"),
 create_boxplot(data_long, "Aircraft_Ownership", "Number of Aircraft"),
 create_boxplot(data_long, "Daily_Utilization", "Hours per Day")
)

# Arrange plots in a grid
grid.arrange(grobs = plots, ncol = 2, nrow = 2)
```

OUTPUT:



Q6:

Here's an analysis of the information presented in the plots :

1. Salary Distributions

- **Small Narrow Body**: Salaries are concentrated around $300 to $650, with most employees earning in the $350–500 range.
- **Large Narrow Body**: Salaries range from $350 to $650, and more employees earn around $500–600.
- **Wide Body**: Salaries are significantly higher, between $600 and $1100, with a large concentration in the $900–1000 range.

Key Point: Wide-body aircraft employees earn higher salaries compared to narrow-body ones.

2. Maintenance Costs and Load Factors

- **Maintenance Costs**: The small scatter plots show how maintenance costs vary over years for different aircraft types. Wide-body planes generally have higher maintenance costs, as expected from their size.
- **Load Factor Comparison**:
    - Large narrow-body planes consistently achieve higher load factors (~1.0 or 100% utilization), indicating better passenger occupancy.
    - Wide-body planes have lower load factors but still perform well.
    - Small narrow-body planes show the lowest load factors.

Key Point: Large narrow-body planes are more efficient in passenger occupancy, while wide-body planes require higher maintenance.

3. Purchased Goods, Aircraft Ownership, and Daily Utilization

- **Purchased Goods**: Wide-body planes have the highest spending on purchased goods, while small narrow-body planes have the lowest.
- **Aircraft Ownership**: Wide-body aircraft are owned in larger numbers, indicating higher utilization or demand.
- **Daily Utilization**:
    - Wide-body planes operate for the most hours per day (up to 12 hours).
    - Large narrow-body planes follow closely.
    - Small narrow-body planes have the least daily utilization.

Key Point: Wide-body planes dominate in ownership, daily usage, and associated costs.