R Code For Reading and Cleaning Data:

```
#Cleaning and reading the csv file
getwd()
setwd("E:/all files for r studio")
getwd()

# 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)
```

R Code: Q1, Q3

```
library(dplyr)  # Load necessary library
##for small narrowbodies
# Sample of size n=11 from Salaries and Wages data
salaries_wages <- c(388, 410, 444, 439, 437, 484, 553, 568, 371, 327, 292, 293, 296, 325, 354, 358, 367, 414, 528,
615, 649)
# Set the sample size
set.seed(123) # To ensure reproducibility
sample_size <- 11
sample_data <- sample(salaries_wages, sample_size)
# Determine appropriate class intervals using Sturges' formula
num_classes <- ceiling(log2(sample_size) + 1)
# Create frequency distribution
breaks <- seq(min(sample_data), max(sample_data), length.out = num_classes + 1)
frequency_distribution <- cut(sample_data, breaks = breaks, include.lowest = TRUE, right = FALSE)
frequency_table <- table(frequency_distribution)
# Display the frequency distribution
print(frequency_table)
hist(sample_data, breaks = breaks, main = "Frequency Distribution of Salaries and Wages (Small Narrowbodies)",
     xlab = "Salaries and Wages", ylab = "Frequency", col = "Pink", border = "Black")

##For large narrow bodies
# Sample of size n=11 from Salaries and Wages data
salaries_wages <- c(495, 517, 583, 539, 546, 585, 629, 654, 440, 393, 350, 353, 368, 382, 401, 409, 429, 413, 542,
621, 648)
# Set the sample size
set.seed(123) # To ensure reproducibility
sample_size <- 11
sample_data <- sample(salaries_wages, sample_size)
# Determine appropriate class intervals using Sturges' formula
num_classes <- ceiling(log2(sample_size) + 1)
# Create frequency distribution
```

```
breaks <- seq(min(sample_data), max(sample_data), length.out = num_classes + 1)
frequency_distribution <- cut(sample_data, breaks = breaks, include.lowest = TRUE, right = FALSE)
frequency_table <- table(frequency_distribution)
# Display the frequency distribution
print(frequency_table)
hist(sample_data, breaks = breaks, main = "Frequency Distribution of Salaries and Wages (Large Narrowbodies)",
   xlab = "Salaries and Wages", ylab = "Frequency", col = "Pink", border = "Black")

##Wide bodies
# Sample of size n=11 from Salaries and Wages data
salaries_wages <- c(5040, 5424, 5376, 4948, 4804, 5631, 5882, 5885, 5530, 5890, 7200, 7747, 7986, 10946, 7050,
8452, 9990, 10667, 10604, 10567, 8373)
# Set the sample size
set.seed(123) # To ensure reproducibility
sample_size <- 11
sample_data <- sample(salaries_wages, sample_size)
# Determine appropriate class intervals using Sturges' formula
num_classes <- ceiling(log2(sample_size) + 1)
# Create frequency distribution
breaks <- seq(min(sample_data), max(sample_data), length.out = num_classes + 1)
frequency_distribution <- cut(sample_data, breaks = breaks, include.lowest = TRUE, right = FALSE)
frequency_table <- table(frequency_distribution)
# Display the frequency distribution
print(frequency_table)
hist(sample_data, breaks = breaks, main = "Frequency Distribution of Salaries and Wages (Widebodies)",
   xlab = "Salaries and Wages", ylab = "Frequency", col = "Pink", border = "Black")
```

R Code Output:

frequency_distribution

[292,363)	[363,435)	[435,506)	[506,578)	[578,649]
4	1	4	1	1

Fig01: Frequency Distribution Table for Small narrow bodies.

frequency_distribution

[350,410)	[410,469)	[469,529)	[529,588)	[588,648]
4	0	1	5	1

Fig02: Frequency Distribution Table for large narrow bodies.

frequency_distribution

[4.8e+03,6.03e+03)	[6.03e+03,7.26e+03)	[7.26e+03,8.49e+03)	[8.49e+03,9.72e+03)	[9.72e+03,1.09e+04]
6	2	1	0	2

Fig03: Frequency Distribution Table for wide bodies.

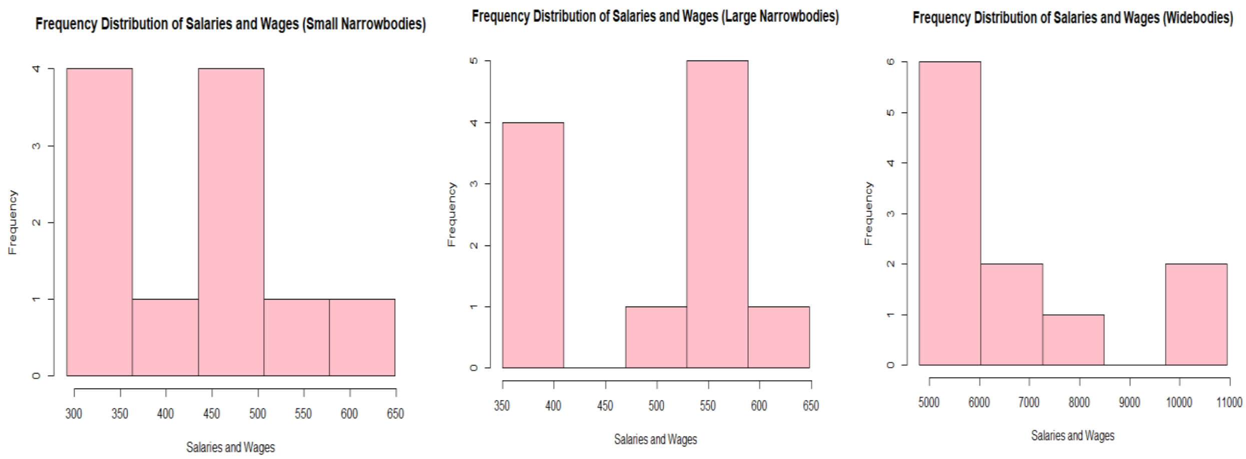

Fig04: Histograms for small , large narrow bodies and wide bodies

<u>R Code</u>: Q2

```
# Define get_mode function to be used for all datasets
  get_mode <- function(v) {
  uniq_vals <- unique(v)
  tabulated_vals <- tabulate(match(v, uniq_vals))
  max_count <- max(tabulated_vals)
  if (max_count == 1) {
    return
  } else {
    return(uniq_vals[which(tabulated_vals == max_count)])
  }
}
# Function to compute all statistical measures
  compute_stats <- function(data, aircraft_type) {
  mean_value <- mean(data)
  median_value <- median(data)
  mode_value <- get_mode(data)
  std_dev <- sd(data)
  variance_value <- var(data)
  quartiles <- quantile(data)
  decile_9 <- quantile(data, 0.9)
  percentile_10 <- quantile(data, 0.1)
  range_value <- range(data)

  result_list <- list(
    Aircraft_Type = aircraft_type,
    Mean = mean_value,
    Median = median_value,
    Mode = mode_value,
    Standard_Deviation = std_dev,
    Variance = variance_value,
    Quartiles = quartiles,
    `9th_Decile` = decile_9,
    `10th_Percentile` = percentile_10,
    Range = range_value
  )
  return(result_list)
}
```

```
# Small Narrowbodies Data
small_narrowbodies <- c(388, 410, 444, 439, 437, 484, 553, 568, 371, 327, 292, 293, 296, 325, 354, 358, 367, 414, 528, 615, 649)
cat("\nSmall Narrowbodies Statistics:\n")
print(compute_stats(small_narrowbodies, "Small Narrowbodies"))

# Large Narrowbodies Data
large_narrowbodies <- c(495, 517, 583, 539, 546, 585, 629, 654, 440, 393, 350, 353, 368, 382, 401, 409, 429, 413, 542, 621, 648)
cat("\nLarge Narrowbodies Statistics:\n")
print(compute_stats(large_narrowbodies, "Large Narrowbodies"))

# Widebodies Data
widebodies <- c(5040, 5424, 5376, 4948, 4804, 5631, 5882, 5885, 5530, 5890, 7200, 7747, 7986, 10946, 7050, 8452, 9990, 10667, 10604, 10567, 8373)
cat("\nWidebodies Statistics:\n")
print(compute_stats(widebodies, "Widebodies"))
```

R Code Output:

```
$Aircraft_Type
[1] "Small Narrowbodies"

$Mean
[1] 424.381

$Median
[1] 410

$Mode
.Primitive("return")

$Standard_Deviation
[1] 106.4836

$Variance
[1] 11338.75

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

$`9th_Decile`
90% 
568 

$`10th_Percentile`
10% 
296 

$Range
[1] 292 649
```

```
$Aircraft_Type
[1] "Large Narrowbodies"

$Mean
[1] 490.3333

$Median
[1] 495

$Mode
.Primitive("return")

$Standard_Deviation
[1] 103.915

$Variance
[1] 10798.33

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

$`9th_Decile`
90% 
629 

$`10th_Percentile`
10% 
368 

$Range
[1] 350 654
```

```
$Aircraft_Type
[1] "Widebodies"

$Mean
[1] 7332.952

$Median
[1] 7050

$Mode
.Primitive("return")

$Standard_Deviation
[1] 2152.493

$Variance
[1] 4633226

$Quartiles
   0%   25%   50%   75%  100% 
 4804  5530  7050  8452 10946 

$`9th_Decile`
  90% 
10604 

$`10th_Percentile`
 10% 
5040 

$Range
[1]  4804 10946
```

Fig05: Statistics for different aircraft types.

<u>R Code:</u> Q4

```
# Load required libraries
library(ggplot2)
library(reshape2)
library(gridExtra)
library(dplyr)
library(tidyr)

set.seed(123)  # Set seed for reproducibility

# Create data frames for each aircraft type
# Small Narrow Body
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"
)
```

```
# Large Narrow Body
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 = c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1),
  Type = "Large Narrow Body"
)

# Wide Body
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"
)

# Randomly sample 11 years
sampled_years <- sample(1995:2015, 11)

# Filter all datasets using the same sampled years
small_narrow_sample <- small_narrow_data[small_narrow_data$Year %in% sampled_years, ]
large_narrow_sample <- large_narrow_data[large_narrow_data$Year %in% sampled_years, ]
wide_body_sample <- wide_body_data[wide_body_data$Year %in% sampled_years, ]

# Combine all samples into one data frame
combined_data <- rbind(
  small_narrow_sample,
  large_narrow_sample,
  wide_body_sample
)

# Simple pie chart for Maintenance
p1 <- ggplot(combined_data, aes(x = "", y = Maintenance, fill = Type)) +
  geom_col() +
  coord_polar("y") +
  facet_wrap(~Year) +
  ggtitle("Maintenance Costs")

# Simple bar chart for Load Factor
p2 <- ggplot(combined_data, aes(x = Year, y = Load_Factor, fill = Type)) +
  geom_col(position = "dodge") +
  ggtitle("Load Factor")

# Show both plots
grid.arrange(p1, p2, nrow = 2)
```

R Code Output:

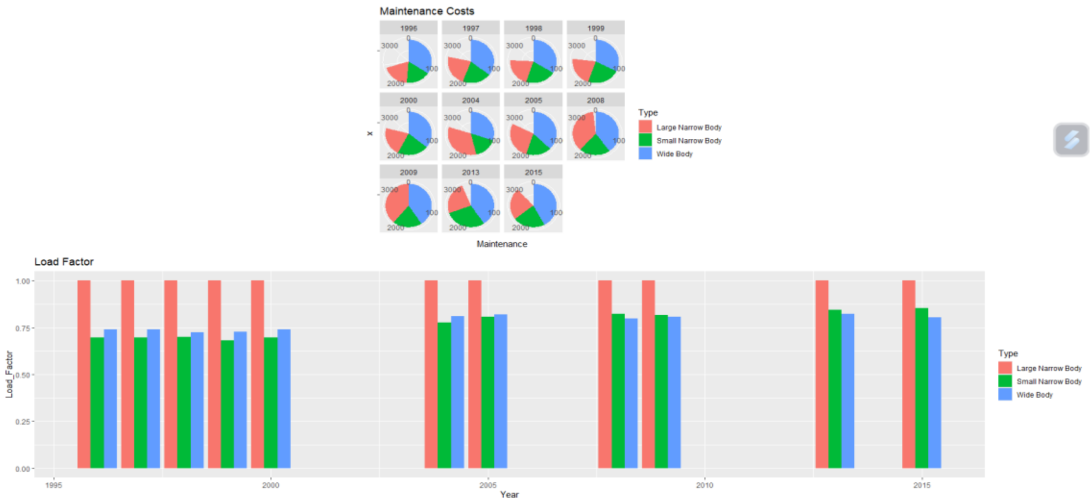

Fig06: Maintenance Pie chart & Load Factor Bar Chart

<u>R Code:</u> 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(644,784,756,650,631,864,961,873,1021,1354,1897,2200,2236,3644,1844,2478,3124,3166,2951,2775,1836),
  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(1513,1788,1683,1356,1309,1815,1935,1701,1927,2660,3953,4634,4884,7761,3762,5188,6521,7017,6737,6406,42
16),
  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(11.54,11.38,11.48,11.69,11.94,11.86,11.56,10.53,9.59,11.61,12.07,12.14,12.41,12.45,12.09,12.28,12.10,12.25,11.6
5,11.68,11.62),
  Aircraft_Type = "Wide Body"
)
# Combine all data frames
all_data <- rbind(small_narrow_data, large_narrow_data, wide_body_data)

# Randomly sample 11 years
sampled_years <- sample(unique(all_data$Year), 11)
# Filter data for sampled years
sampled_data <- all_data[all_data$Year %in% sampled_years, ]

# Reshape data for ggplot
data_long <- melt(sampled_data, id.vars = c("Year", "Aircraft_Type"))
# Create the box plot
ggplot(data_long, aes(x = variable, y = value, fill = Aircraft_Type)) +
  # Add box plots with colors
  geom_boxplot() +
  # Set colors
  scale_fill_manual(values = c(
    "Large Narrow Body" = "pink",
    "Small Narrow Body" = "lavender",
    "Wide Body" = "skyblue"
  )) +
  labs(
    title = "United Airlines Variables by Aircraft Type",
    x = "Variable",
    y = "Value",
    fill = "Type"
  ) +
  # clean look
  theme_minimal() +
  # legend at bottom
  theme(legend.position = "bottom")
  )
```

R Code Output:

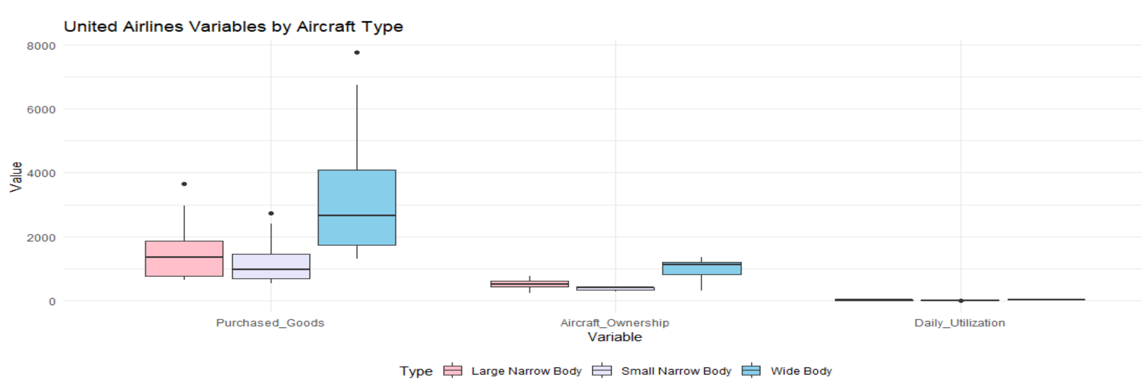

Fig07: Box plots for different aircraft types

<u>R Code:</u> Q6

The box plots for different aircraft categories within United Airlines provide information on

1. Central Tendency: The thick line within each box represents the median of the data, which gives an idea of each variable's "middle" value.
2. Dispersion or Spread:
   - The box itself spans the interquartile range (IQR), which contains the middle 50% of the data. A longer box indicates greater variability.
   - The lines (whiskers) extending from the box show the overall range of the data, excluding outliers.
   - Outliers are often plotted as individual points, suggesting unusual or extreme values.
3. Small Narrowbodies: "Purchased Goods" has a higher median and greater variability than "Aircraft Ownership" and "Daily Utilization."
4. Large Narrowbodies: "Daily Utilization" shows a very compact box plot, indicating little variation in daily usage for this aircraft type.
5. Widebodies: "Aircraft Ownership" has a skewed distribution with a few low outliers, suggesting a few years with shallow ownership values.

Overall, these box plots compare the distributions of different variables across various aircraft categories, highlighting key differences in their central tendency, spread, and potential outliers