Week 3 Tasks — Data Preparation, EDA, and Intro Modeling (R)

Cao Pham Minh Dang

Instructions

- Use this template to complete Week 3 tasks. Replace placeholders with your work.
- Ensure the document can knit top-to-bottom without errors.
- Add short captions/annotations below each plot and metric output.

Task 1 — Load Data and Inspect

```
data <- read.csv("./ncr_ride_bookings.csv", na = c("", "NA", "null"))
print(dim(data))
```

```
## [1] 150000     21
```

```
summary(data)
```

```
##      Date               Time            Booking.ID        Booking.Status
##  Length:150000      Length:150000      Length:150000      Length:150000
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  Customer.ID        Vehicle.Type       Pickup.Location    Drop.Location
##  Length:150000      Length:150000      Length:150000      Length:150000
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##     Avg.VTAT         Avg.CTAT     Cancelled.Rides.by.Customer
##  Min.   : 2.000   Min.   :10.00   Min.   :1
##  1st Qu.: 5.300   1st Qu.:21.60   1st Qu.:1
##  Median : 8.300   Median :28.80   Median :1
##  Mean   : 8.456   Mean   :29.15   Mean   :1
##  3rd Qu.:11.300   3rd Qu.:36.80   3rd Qu.:1
##  Max.   :20.000   Max.   :45.00   Max.   :1
##  NA's   :10500    NA's   :48000   NA's   :139500
```

```
##  Reason.for.cancelling.by.Customer Cancelled.Rides.by.Driver
##  Length:150000                     Min.   :1
##  Class :character                  1st Qu.:1
##  Mode  :character                  Median :1
##                                    Mean   :1
##                                    3rd Qu.:1
##                                    Max.   :1
##                                    NA's   :123000
##  Driver.Cancellation.Reason Incomplete.Rides Incomplete.Rides.Reason
##  Length:150000              Min.   :1        Length:150000
##  Class :character           1st Qu.:1        Class :character
##  Mode  :character           Median :1        Mode  :character
##                             Mean   :1
##                             3rd Qu.:1
##                             Max.   :1
##                             NA's   :141000
##  Booking.Value    Ride.Distance   Driver.Ratings  Customer.Rating
##  Min.   :  50.0   Min.   : 1.00   Min.   :3.000   Min.   :3.000
##  1st Qu.: 234.0   1st Qu.:12.46   1st Qu.:4.100   1st Qu.:4.200
##  Median : 414.0   Median :23.72   Median :4.300   Median :4.500
##  Mean   : 508.3   Mean   :24.64   Mean   :4.231   Mean   :4.405
##  3rd Qu.: 689.0   3rd Qu.:36.82   3rd Qu.:4.600   3rd Qu.:4.800
##  Max.   :4277.0   Max.   :50.00   Max.   :5.000   Max.   :5.000
##  NA's   :48000    NA's   :48000   NA's   :57000   NA's   :57000
##  Payment.Method
##  Length:150000
##  Class :character
##  Mode  :character
##
##
##
##
```

```
head(data, 10)
```

```
##          Date     Time   Booking.ID  Booking.Status  Customer.ID  Vehicle.Type
## 1  2024-03-23 12:29:38 "CNR5884300" No Driver Found "CID1982111"         eBike
## 2  2024-11-29 18:01:39 "CNR1326809"      Incomplete "CID4604802"      Go Sedan
## 3  2024-08-23 08:56:10 "CNR8494506"       Completed "CID9202816"          Auto
## 4  2024-10-21 17:17:25 "CNR8906825"       Completed "CID2610914" Premier Sedan
## 5  2024-09-16 22:08:00 "CNR1950162"       Completed "CID9933542"          Bike
## 6  2024-02-06 09:44:56 "CNR4096693"       Completed "CID4670564"          Auto
## 7  2024-06-17 15:45:58 "CNR2002539"       Completed "CID6800553"       Go Mini
## 8  2024-03-19 17:37:37 "CNR6568000"       Completed "CID8610436"          Auto
## 9  2024-09-14 12:49:09 "CNR4510807" No Driver Found "CID7873618"      Go Sedan
## 10 2024-12-16 19:06:48 "CNR7721892"      Incomplete "CID5214275"          Auto
##        Pickup.Location     Drop.Location Avg.VTAT Avg.CTAT
## 1          Palam Vihar           Jhilmil       NA       NA
## 2        Shastri Nagar Gurgaon Sector 56      4.9     14.0
## 3              Khandsa     Malviya Nagar     13.4     25.8
## 4  Central Secretariat          Inderlok     13.1     28.5
## 5     Ghitorni Village       Khan Market      5.3     19.6
## 6                AIIMS       Narsinghpur      5.1     18.1
## 7            Vaishali      Punjabi Bagh      7.1     20.4
```

```
## 8          Mayur Vihar         Cyber Hub     12.1     16.5
## 9      Noida Sector 62   Noida Sector 18       NA       NA
## 10              Rohini      Adarsh Nagar      6.1     26.0
##    Cancelled.Rides.by.Customer Reason.for.cancelling.by.Customer
## 1                           NA                              <NA>
## 2                           NA                              <NA>
## 3                           NA                              <NA>
## 4                           NA                              <NA>
## 5                           NA                              <NA>
## 6                           NA                              <NA>
## 7                           NA                              <NA>
## 8                           NA                              <NA>
## 9                           NA                              <NA>
## 10                          NA                              <NA>
##    Cancelled.Rides.by.Driver Driver.Cancellation.Reason Incomplete.Rides
## 1                         NA                       <NA>               NA
## 2                         NA                       <NA>                1
## 3                         NA                       <NA>               NA
## 4                         NA                       <NA>               NA
## 5                         NA                       <NA>               NA
## 6                         NA                       <NA>               NA
## 7                         NA                       <NA>               NA
## 8                         NA                       <NA>               NA
## 9                         NA                       <NA>               NA
## 10                        NA                       <NA>                1
##    Incomplete.Rides.Reason Booking.Value Ride.Distance Driver.Ratings
## 1                     <NA>            NA            NA             NA
## 2          Vehicle Breakdown          237          5.73             NA
## 3                     <NA>          627         13.58            4.9
## 4                     <NA>          416         34.02            4.6
## 5                     <NA>          737         48.21            4.1
## 6                     <NA>          316          4.85            4.1
## 7                     <NA>          640         41.24            4.0
## 8                     <NA>          136          6.56            4.4
## 9                     <NA>           NA            NA             NA
## 10              Other Issue          135         10.36             NA
##    Customer.Rating Payment.Method
## 1               NA           <NA>
## 2               NA            UPI
## 3              4.9     Debit Card
## 4              5.0            UPI
## 5              4.3            UPI
## 6              4.6            UPI
## 7              4.1            UPI
## 8              4.2            UPI
## 9               NA           <NA>
## 10              NA           Cash
```

Briefly describe the dataset, its purpose, and key variables.

Overview This comprehensive dataset contains detailed ride-sharing data from Uber operations for the year 2024.

The dataset contains 150,000 Uber ride bookings in the NCR region, with each row representing a single

ride booking event. Its purpose is to provide detailed analytics on ride bookings, cancellations, completions, and customer/driver interactions.

Key variables: - `Date`, `Time`: When the booking was made. - `Booking ID`, `Customer ID`: Unique identifiers for each booking and customer. - `Booking Status`: Status of the ride (e.g., Completed, Incomplete, No Driver Found). - `Vehicle Type`: Type of vehicle booked. - `Pickup Location`, `Drop Location`: Start and end points of the ride. - `Avg VTAT` (Vehicle Turnaround Time), `Avg CTAT` (Customer Turnaround Time): Operational metrics. - `Cancelled Rides by Customer/Driver`, `Reason for cancelling by Customer/Driver`: Cancellations and their reasons. - `Incomplete Rides`, `Incomplete Rides Reason`: Rides that were not completed and why. - `Booking Value`: Fare amount for the ride. - `Ride Distance`: Distance covered in the ride. - `Driver Ratings`, `Customer Rating`: Feedback scores. - `Payment Method`: How the ride was paid for (e.g., UPI, Debit Card).

Purpose The primary purpose of this dataset is to enable in-depth analysis of Uber's ride-hailing operations within the NCR region during 2024. By capturing every stage of the ride lifecycle, from booking to completion or cancellation, it provides a foundation for answering both operational and strategic questions. Therefore, this data can offer rich insights into booking patterns, vehicle performance, revenue streams, cancellation behaviors, and customer satisfaction metrics.

Task 2 — Data Types, Summary Stats, and Missingness

```r
# Glimpse types
glimpse(data)
```

```
## Rows: 150,000
## Columns: 21
## $ Date                                <chr> "2024-03-23", "2024-11-29", "2024-08~
## $ Time                                <chr> "12:29:38", "18:01:39", "08:56:10", ~
## $ Booking.ID                          <chr> "\"CNR5884300\"", "\"CNR1326809\"", ~
## $ Booking.Status                      <chr> "No Driver Found", "Incomplete", "Co~
## $ Customer.ID                         <chr> "\"CID1982111\"", "\"CID4604802\"", ~
## $ Vehicle.Type                        <chr> "eBike", "Go Sedan", "Auto", "Premie~
## $ Pickup.Location                     <chr> "Palam Vihar", "Shastri Nagar", "Kha~
## $ Drop.Location                       <chr> "Jhilmil", "Gurgaon Sector 56", "Mal~
## $ Avg.VTAT                            <dbl> NA, 4.9, 13.4, 13.1, 5.3, 5.1, 7.1, ~
## $ Avg.CTAT                            <dbl> NA, 14.0, 25.8, 28.5, 19.6, 18.1, 20~
## $ Cancelled.Rides.by.Customer         <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ Reason.for.cancelling.by.Customer   <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ Cancelled.Rides.by.Driver           <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ Driver.Cancellation.Reason          <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ Incomplete.Rides                    <int> NA, 1, NA, NA, NA, NA, NA, NA, NA, 1~
## $ Incomplete.Rides.Reason             <chr> NA, "Vehicle Breakdown", NA, NA, NA,~
## $ Booking.Value                       <int> NA, 237, 627, 416, 737, 316, 640, 13~
## $ Ride.Distance                       <dbl> NA, 5.73, 13.58, 34.02, 48.21, 4.85,~
## $ Driver.Ratings                      <dbl> NA, NA, 4.9, 4.6, 4.1, 4.1, 4.0, 4.4~
## $ Customer.Rating                     <dbl> NA, NA, 4.9, 5.0, 4.3, 4.6, 4.1, 4.2~
## $ Payment.Method                      <chr> NA, "UPI", "Debit Card", "UPI", "UPI~
```

```r
# Summary statistics (numeric and categorical)
summary(data)
```

```
##      Date               Time            Booking.ID        Booking.Status
```

```
##  Length:150000      Length:150000      Length:150000      Length:150000
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  Customer.ID        Vehicle.Type       Pickup.Location    Drop.Location
##  Length:150000      Length:150000      Length:150000      Length:150000
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##     Avg.VTAT         Avg.CTAT     Cancelled.Rides.by.Customer
##  Min.   : 2.000   Min.   :10.00   Min.   :1
##  1st Qu.: 5.300   1st Qu.:21.60   1st Qu.:1
##  Median : 8.300   Median :28.80   Median :1
##  Mean   : 8.456   Mean   :29.15   Mean   :1
##  3rd Qu.:11.300   3rd Qu.:36.80   3rd Qu.:1
##  Max.   :20.000   Max.   :45.00   Max.   :1
##  NA's   :10500    NA's   :48000   NA's   :139500
##  Reason.for.cancelling.by.Customer Cancelled.Rides.by.Driver
##  Length:150000                      Min.   :1
##  Class :character                   1st Qu.:1
##  Mode  :character                   Median :1
##                                     Mean   :1
##                                     3rd Qu.:1
##                                     Max.   :1
##                                     NA's   :123000
##  Driver.Cancellation.Reason Incomplete.Rides Incomplete.Rides.Reason
##  Length:150000              Min.   :1        Length:150000
##  Class :character           1st Qu.:1        Class :character
##  Mode  :character           Median :1        Mode  :character
##                             Mean   :1
##                             3rd Qu.:1
##                             Max.   :1
##                             NA's   :141000
##  Booking.Value    Ride.Distance   Driver.Ratings  Customer.Rating
##  Min.   :  50.0   Min.   : 1.00   Min.   :3.000   Min.   :3.000
##  1st Qu.: 234.0   1st Qu.:12.46   1st Qu.:4.100   1st Qu.:4.200
##  Median : 414.0   Median :23.72   Median :4.300   Median :4.500
##  Mean   : 508.3   Mean   :24.64   Mean   :4.231   Mean   :4.405
##  3rd Qu.: 689.0   3rd Qu.:36.82   3rd Qu.:4.600   3rd Qu.:4.800
##  Max.   :4277.0   Max.   :50.00   Max.   :5.000   Max.   :5.000
##  NA's   :48000    NA's   :48000   NA's   :57000   NA's   :57000
##  Payment.Method
##  Length:150000
##  Class :character
##  Mode  :character
##
##
##
```

##

```r
# Missingness per column
tibble(col = names(data), n_missing = colSums(is.na(data))) %>%
  mutate(p_missing = n_missing / nrow(data)) %>%
  arrange(desc(p_missing))
```

```
## # A tibble: 21 x 3
##    col                                n_missing p_missing
##    <chr>                                  <dbl>     <dbl>
##  1 Incomplete.Rides                      141000      0.94
##  2 Incomplete.Rides.Reason               141000      0.94
##  3 Cancelled.Rides.by.Customer           139500      0.93
##  4 Reason.for.cancelling.by.Customer     139500      0.93
##  5 Cancelled.Rides.by.Driver             123000      0.82
##  6 Driver.Cancellation.Reason            123000      0.82
##  7 Driver.Ratings                         57000      0.38
##  8 Customer.Rating                        57000      0.38
##  9 Avg.CTAT                               48000      0.32
## 10 Booking.Value                          48000      0.32
## # i 11 more rows
```

Task 3 — Data Cleaning

```r
# Standardize column names → snake_case
names(data) <- names(data) %>%
  tolower() %>%
  str_trim() %>%
  str_replace_all("[^a-z0-9]+", "_") %>%   # replace any sequence of non-alphanumeric chars with "_"
  str_replace_all("^_|_$", "")             # remove leading/trailing underscores


# Remove duplicates
data <- data %>%
  distinct()

# Handle missing values
data <- data %>%
  mutate(across(where(is.character), ~replace_na(., "Unknown"))) %>%
  mutate(across(where(is.numeric), ~replace_na(., median(., na.rm = TRUE))))
```

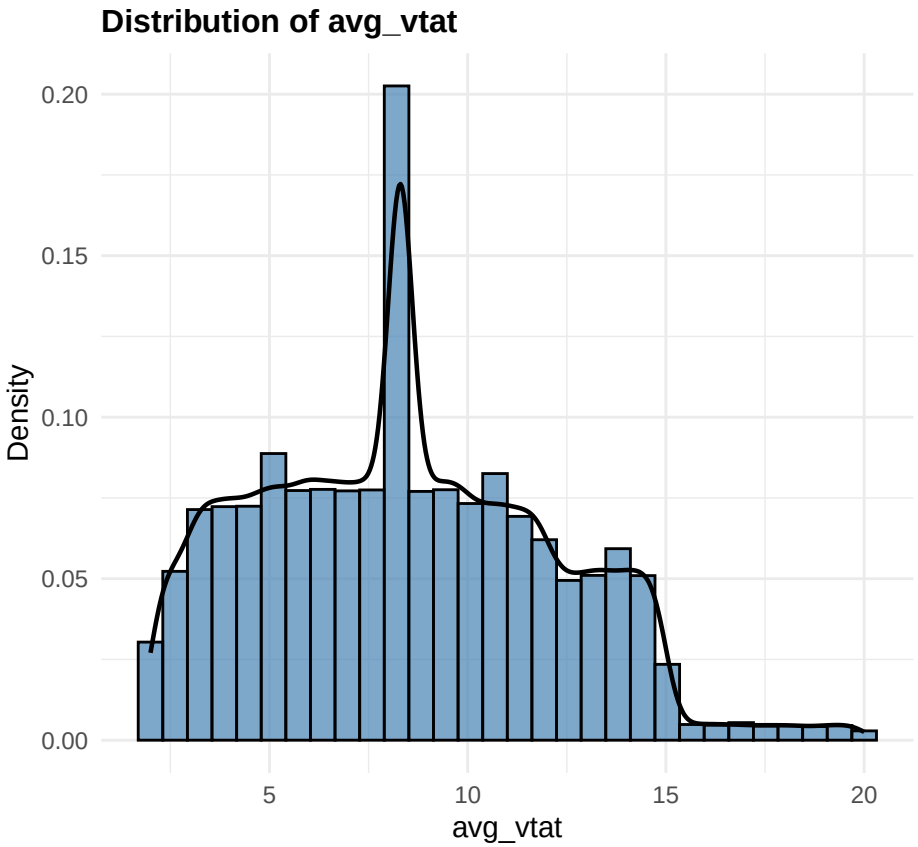
Data Cleaning Decisions

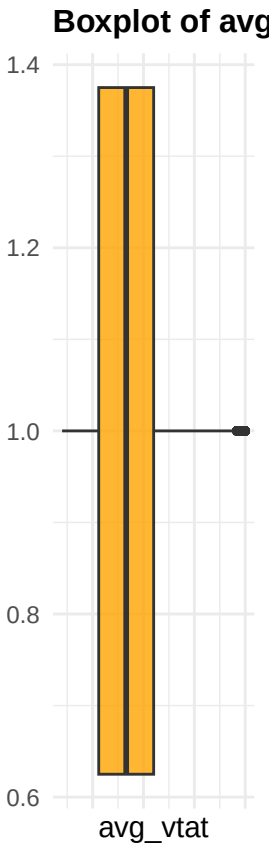
1. **Missing Values**
   - Treated `""`, `"NA"`, and `"null"` as missing.
   - For categorical variables → replaced missing with `"Unknown"` to preserve information.
   - For numeric variables → imputed missing values with the **median** to avoid skewness compared to mean imputation.

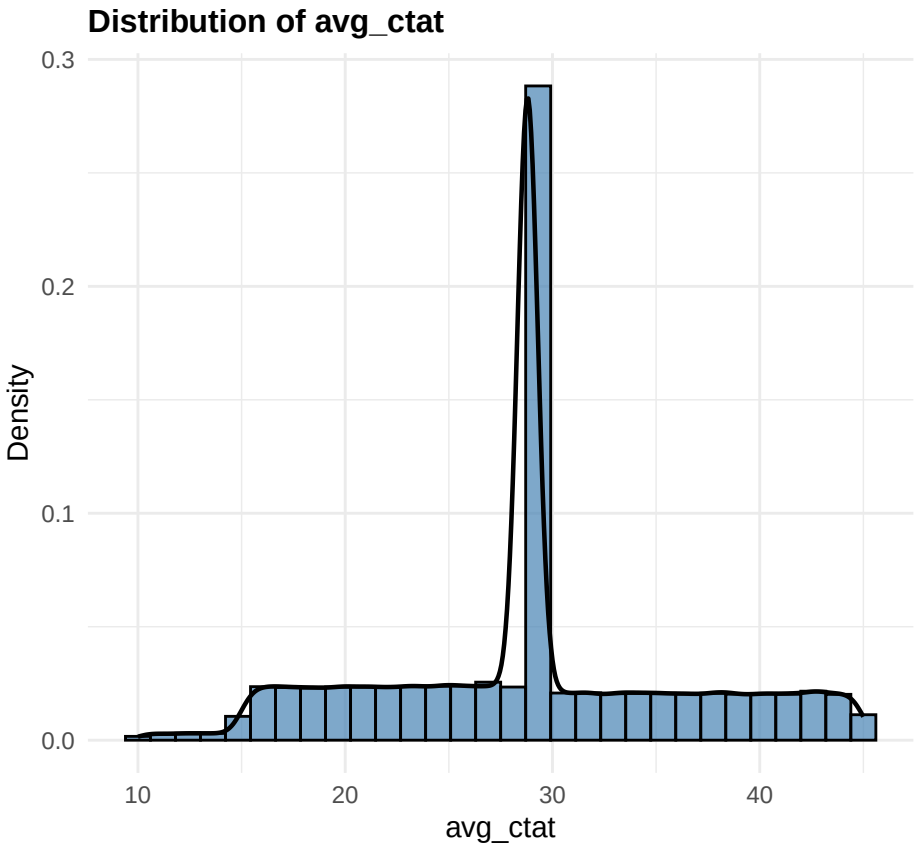
2. **Duplicates**
   - Removed duplicate rows using `distinct()` to prevent overcounting rides.
3. **Column Names**
   - Standardized column names to **snake_case** (e.g., `Booking.ID` → `booking_id`) for consistency in coding and readability.

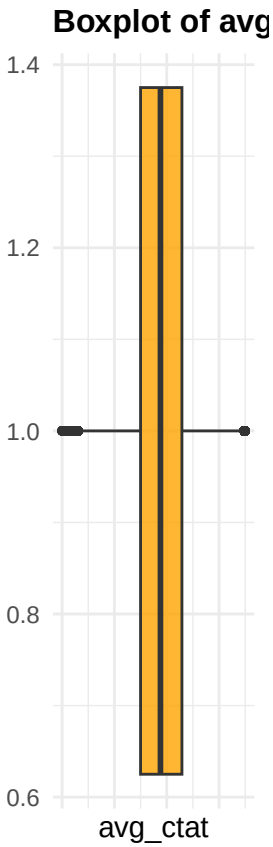
Task 4 — Exploratory Data Analysis (EDA)

```r
# Full EDA plotting script (no conditionals) ----
library(ggplot2)
library(cowplot)    # plot_grid()
library(forcats)    # fct_infreq()
library(dplyr)

# Dataframe: `data` is expected to exist in the environment.

# 1) Univariate distributions + boxplots for numeric columns
numeric_cols <- c("avg_vtat", "avg_ctat", "booking_value",
                  "ride_distance", "driver_ratings", "customer_rating")

for (col in numeric_cols) {
  # histogram with density overlay (density scale)
  p_hist <- ggplot(data, aes_string(x = col)) +
    geom_histogram(aes(y = ..density..),
                   bins = 30, fill = "steelblue", color = "black", alpha = 0.7) +
    geom_density(alpha = 0.35, size = 0.8) +
    labs(title = paste("Distribution of", col),
         x = col, y = "Density") +
    theme_minimal() +
    theme(plot.title = element_text(size = 12, face = "bold"))

  # horizontal boxplot
  p_box <- ggplot(data, aes_string(x = "1", y = col)) +
    geom_boxplot(fill = "orange", alpha = 0.8, outlier.size = 1) +
    coord_flip() +
    labs(title = paste("Boxplot of", col),
         x = "", y = col) +
    theme_minimal() +
    theme(axis.text.x = element_blank(),
          axis.ticks.x = element_blank(),
          plot.title = element_text(size = 12, face = "bold"))

  # side-by-side with relative widths like the original (3:1)
  print(plot_grid(p_hist, p_box, nrow = 1, rel_widths = c(3, 1)))
}
```

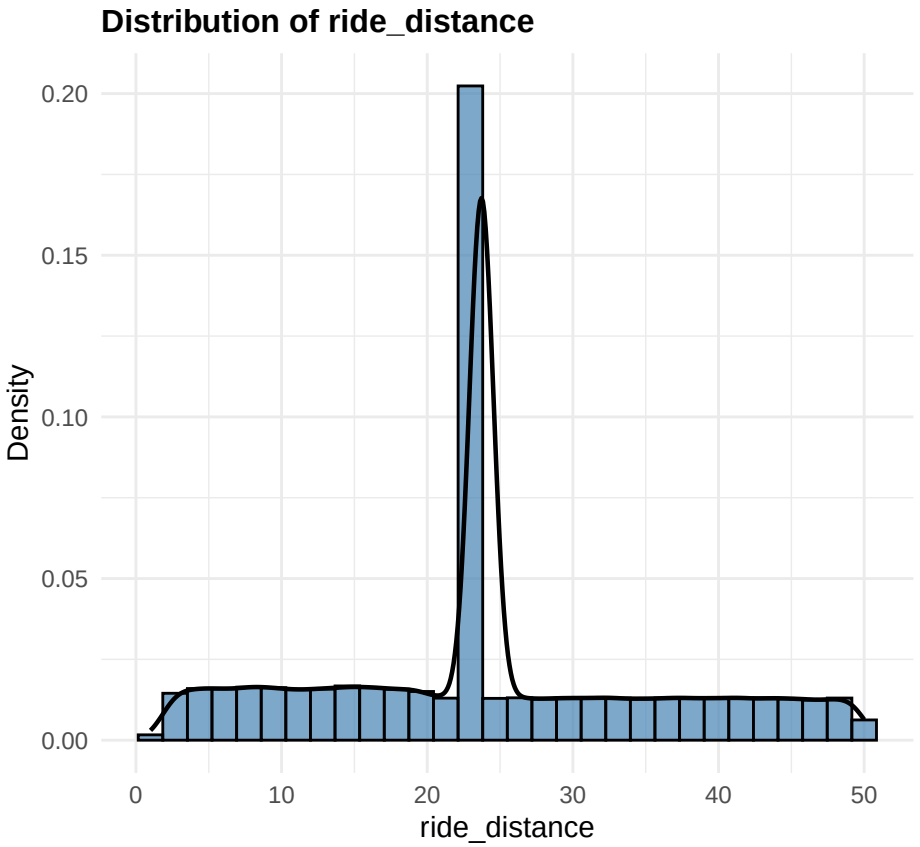

Distribution of avg_vtat
Density
0.20
0.15
0.10
0.05
0.00
5
10
15
20
avg_vtat

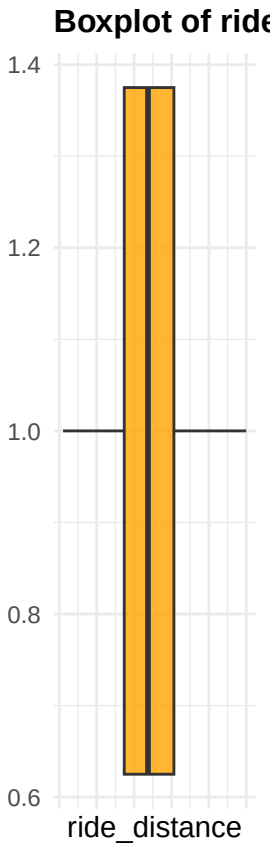

Boxplot of avg
1.4
1.2
1.0
0.8
0.6
avg_vtat

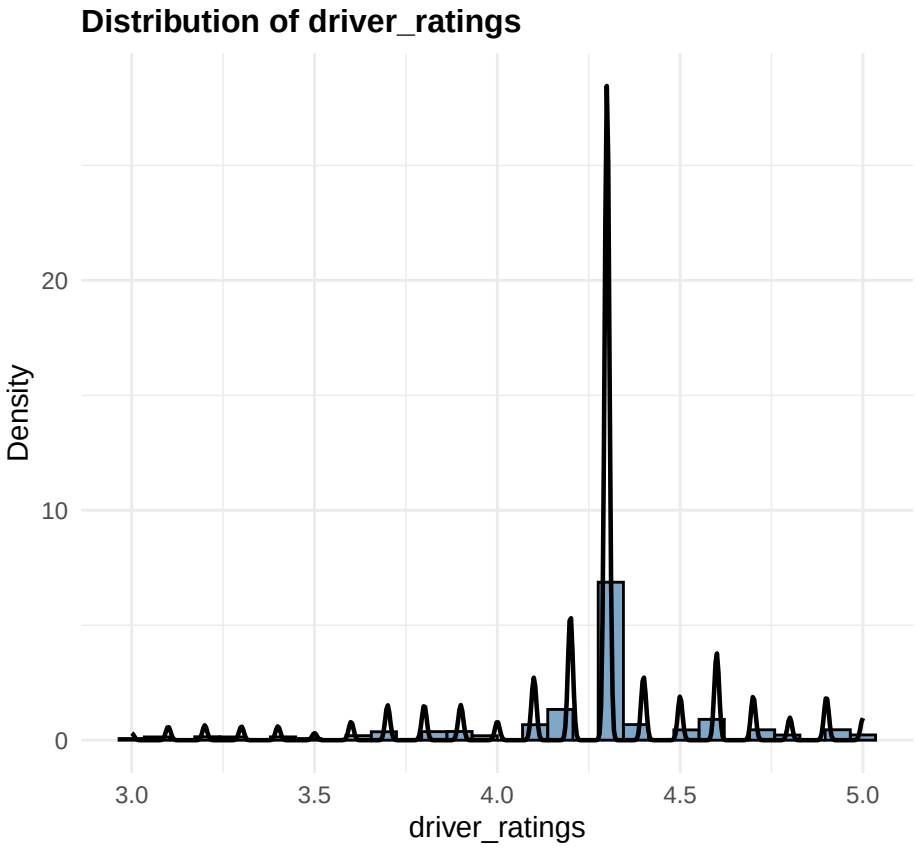

Distribution of avg_ctat
Density
0.3
0.2
0.1
0.0
10
20
30
40
avg_ctat

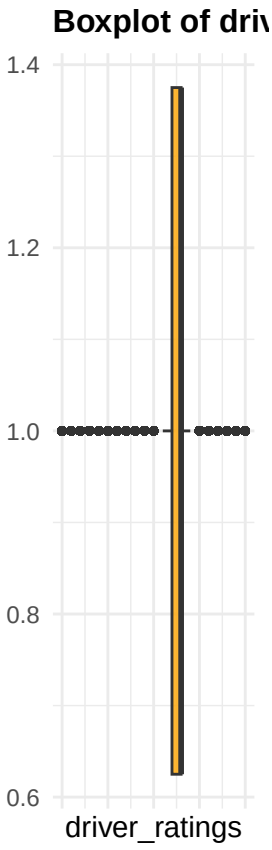

Boxplot of avg
1.4
1.2
1.0
0.8
0.6
avg_ctat

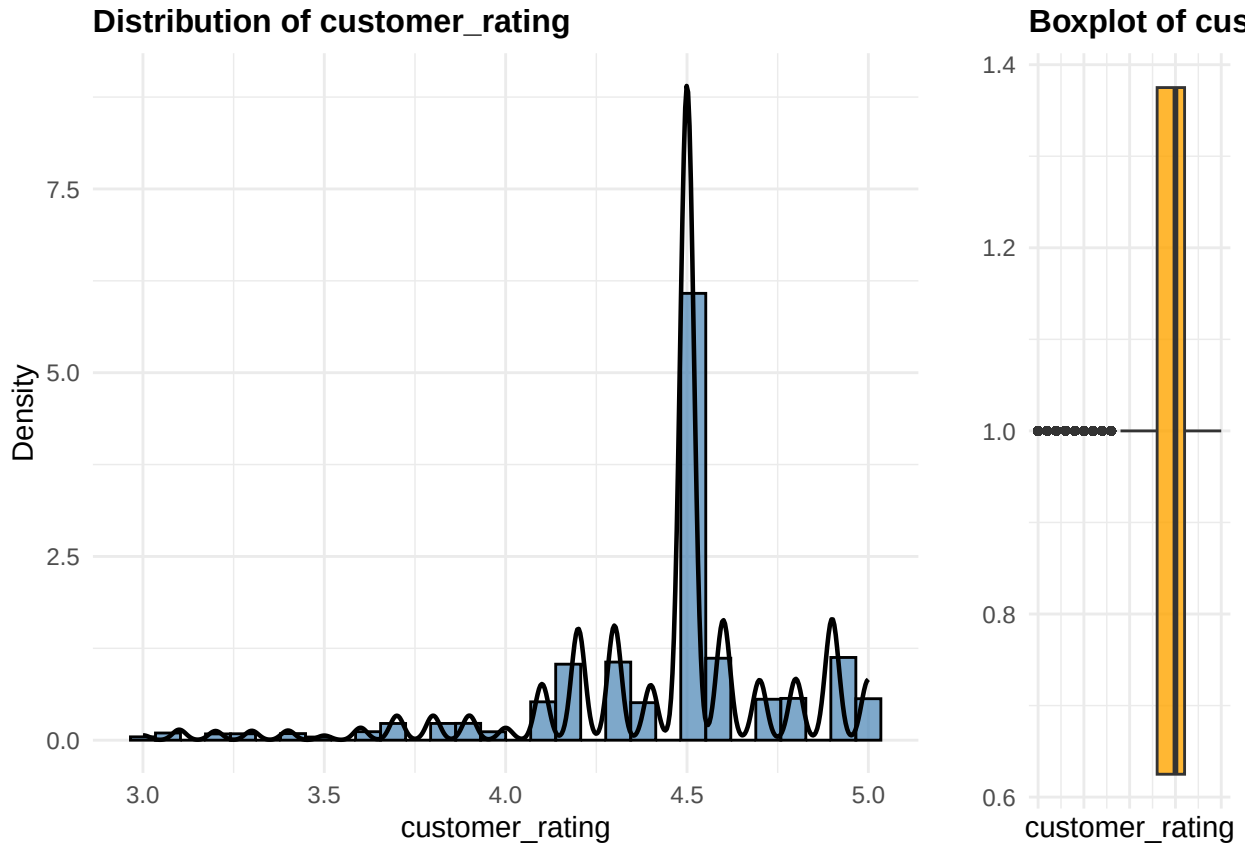

Distribution of booking_value
Density
0.009
0.006
0.003
0.000
0
1000
2000
3000
4000
booking_value
Boxplot of boo
1.4
1.2
1.0
0.8
0.6
booking_value


Distribution of ride_distance
Density
0.20
0.15
0.10
0.05
0.00
0
10
20
30
40
50
ride_distance

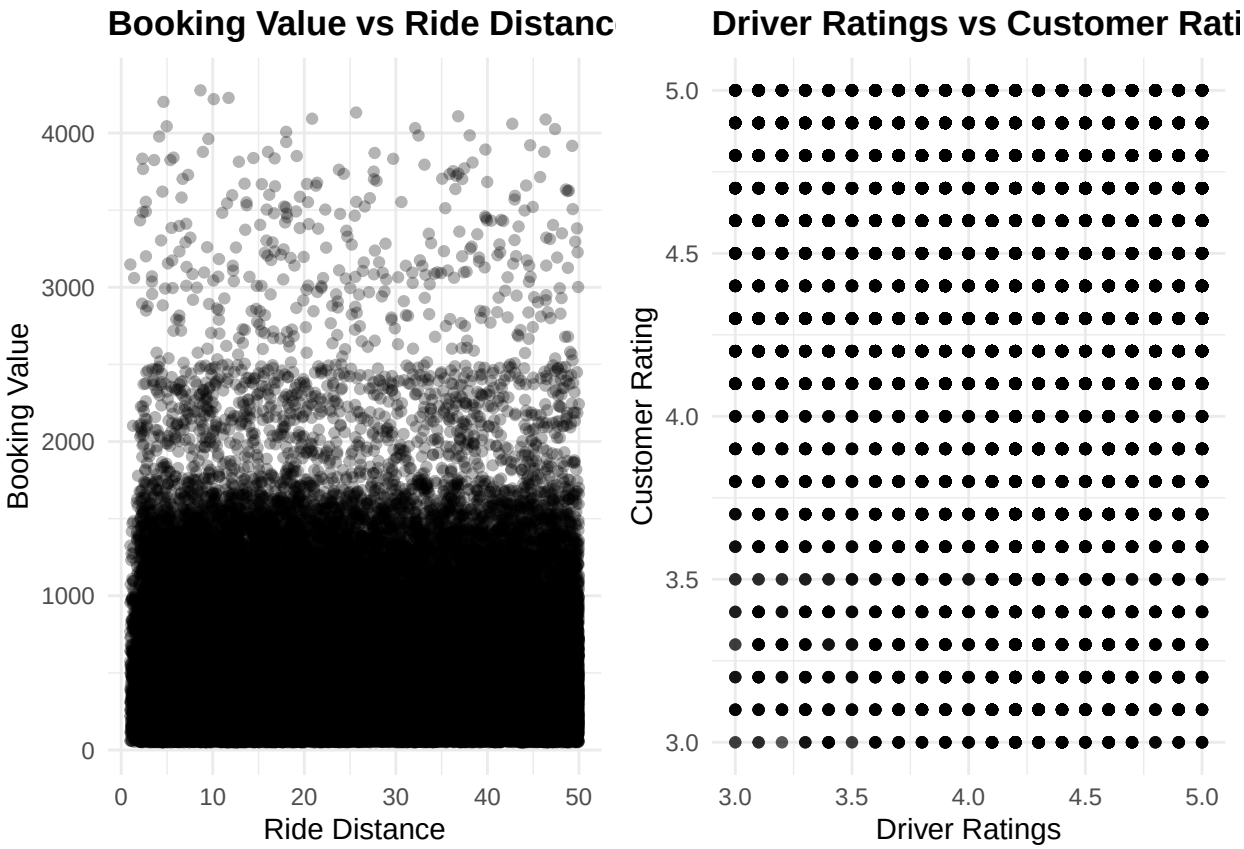

Boxplot of ride
1.4
1.2
1.0
0.8
0.6
ride_distance


Distribution of driver_ratings
Density
20
10
0
3.0
3.5
4.0
4.5
5.0
driver_ratings

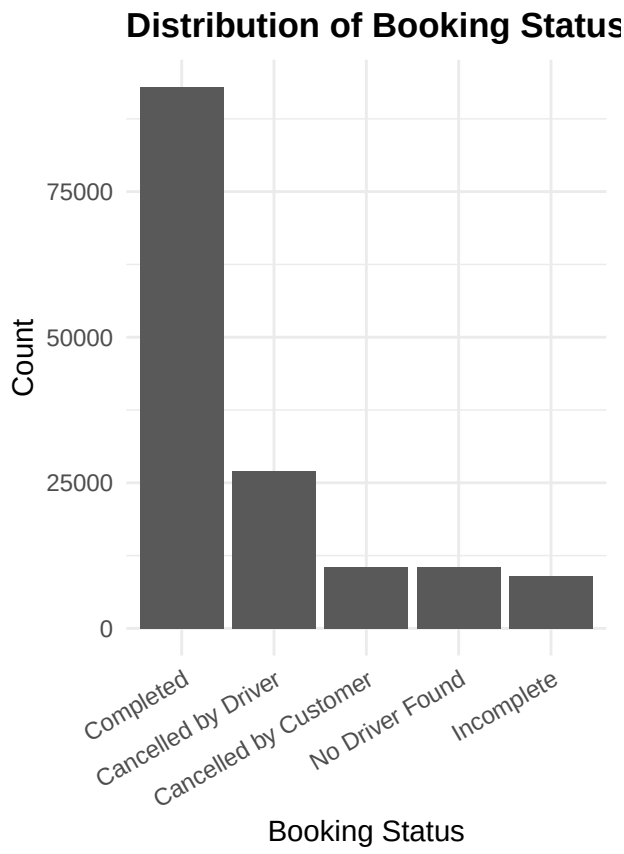

Boxplot of driv
1.4
1.2
1.0
0.8
0.6
driver_ratings

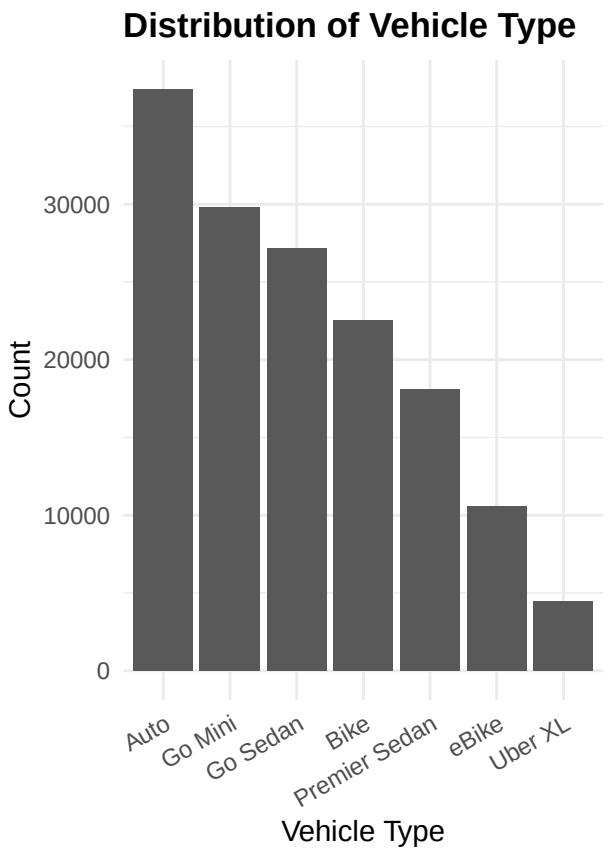



```r
# 2) Pair scatterplots: Booking Value vs Ride Distance & Driver Ratings vs Customer Rating
p_scatter_ride_booking <- ggplot(data, aes(x = ride_distance, y = booking_value)) +
  geom_point(alpha = 0.3) +
  labs(title = "Booking Value vs Ride Distance",
       x = "Ride Distance", y = "Booking Value") +
  theme_minimal() +
  theme(plot.title = element_text(face = "bold"))

p_scatter_driver_customer <- ggplot(data, aes(x = driver_ratings, y = customer_rating)) +
  geom_point(alpha = 0.3) +
  labs(title = "Driver Ratings vs Customer Rating",
       x = "Driver Ratings", y = "Customer Rating") +
  theme_minimal() +
  theme(plot.title = element_text(face = "bold"))

print(plot_grid(p_scatter_ride_booking, p_scatter_driver_customer, nrow = 1))
```

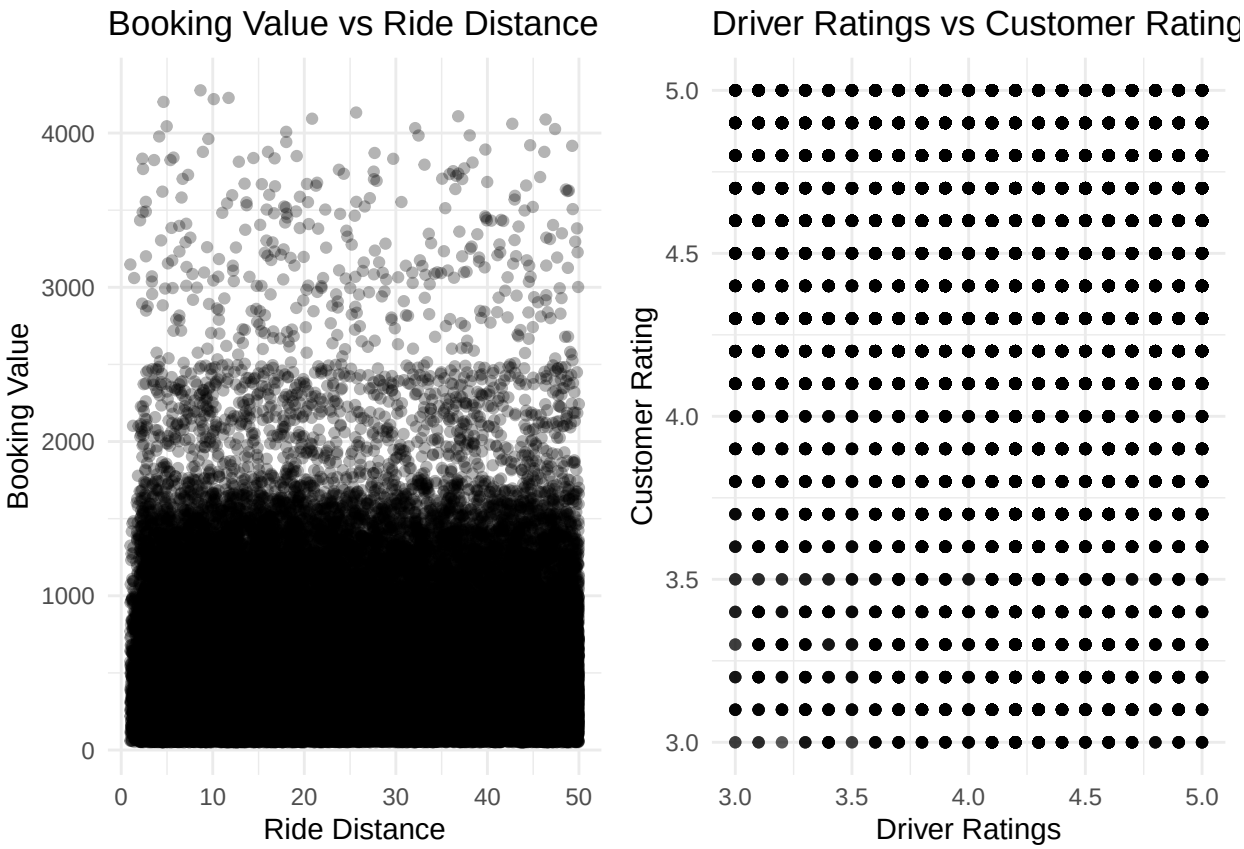



```r
# 3) Categorical distributions: Booking Status and Vehicle Type (ordered by frequency)
p_booking_status <- ggplot(data %>% mutate(booking_status = fct_infreq(booking_status)),
                           aes(x = booking_status)) +
  geom_bar() +
  labs(title = "Distribution of Booking Status", x = "Booking Status", y = "Count") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 30, hjust = 1),
        plot.title = element_text(face = "bold"))

p_vehicle_type <- ggplot(data %>% mutate(vehicle_type = fct_infreq(vehicle_type)),
                         aes(x = vehicle_type)) +
  geom_bar() +
  labs(title = "Distribution of Vehicle Type", x = "Vehicle Type", y = "Count") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 30, hjust = 1),
        plot.title = element_text(face = "bold"))

print(plot_grid(p_booking_status, p_vehicle_type, nrow = 1))
```

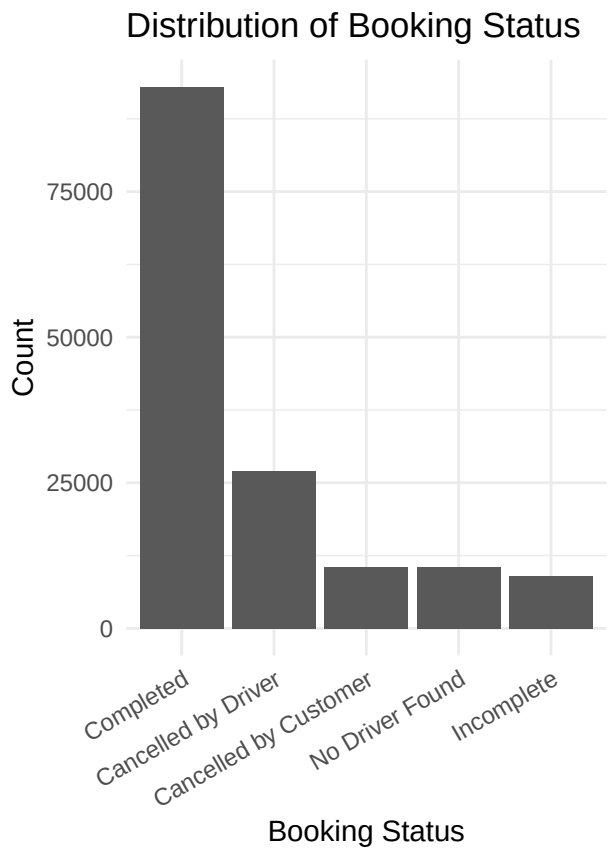


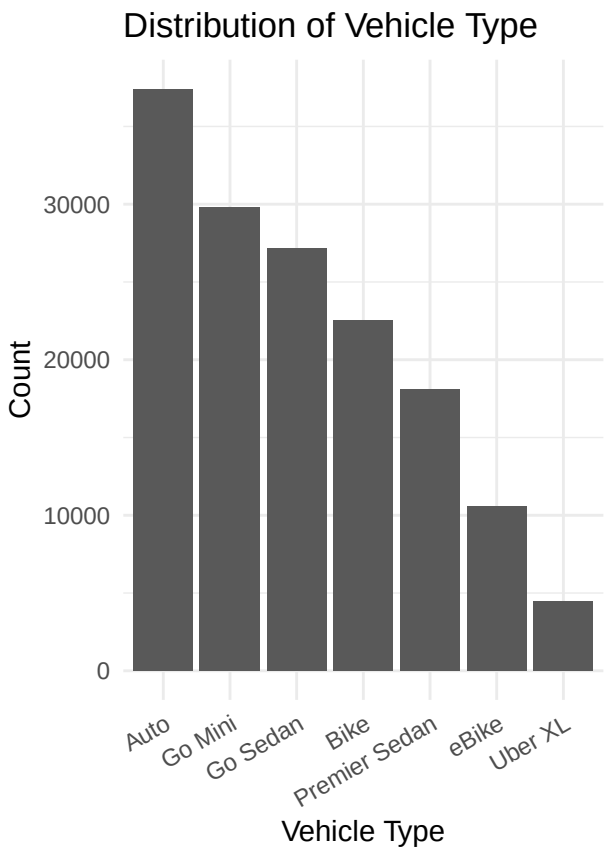



```r
# 4) Repeat the scatterplots (explicitly matching the Python block that creates separate figure)
# (keeps parity with your original script which shows these twice)
p_scatter_ride_booking2 <- ggplot(data, aes(x = ride_distance, y = booking_value)) +
  geom_point(alpha = 0.3) +
  labs(title = "Booking Value vs Ride Distance", x = "Ride Distance", y = "Booking Value") +
  theme_minimal()

p_scatter_driver_customer2 <- ggplot(data, aes(x = driver_ratings, y = customer_rating)) +
  geom_point(alpha = 0.3) +
  labs(title = "Driver Ratings vs Customer Rating", x = "Driver Ratings", y = "Customer Rating") +
  theme_minimal()

print(plot_grid(p_scatter_ride_booking2, p_scatter_driver_customer2, nrow = 1))
```

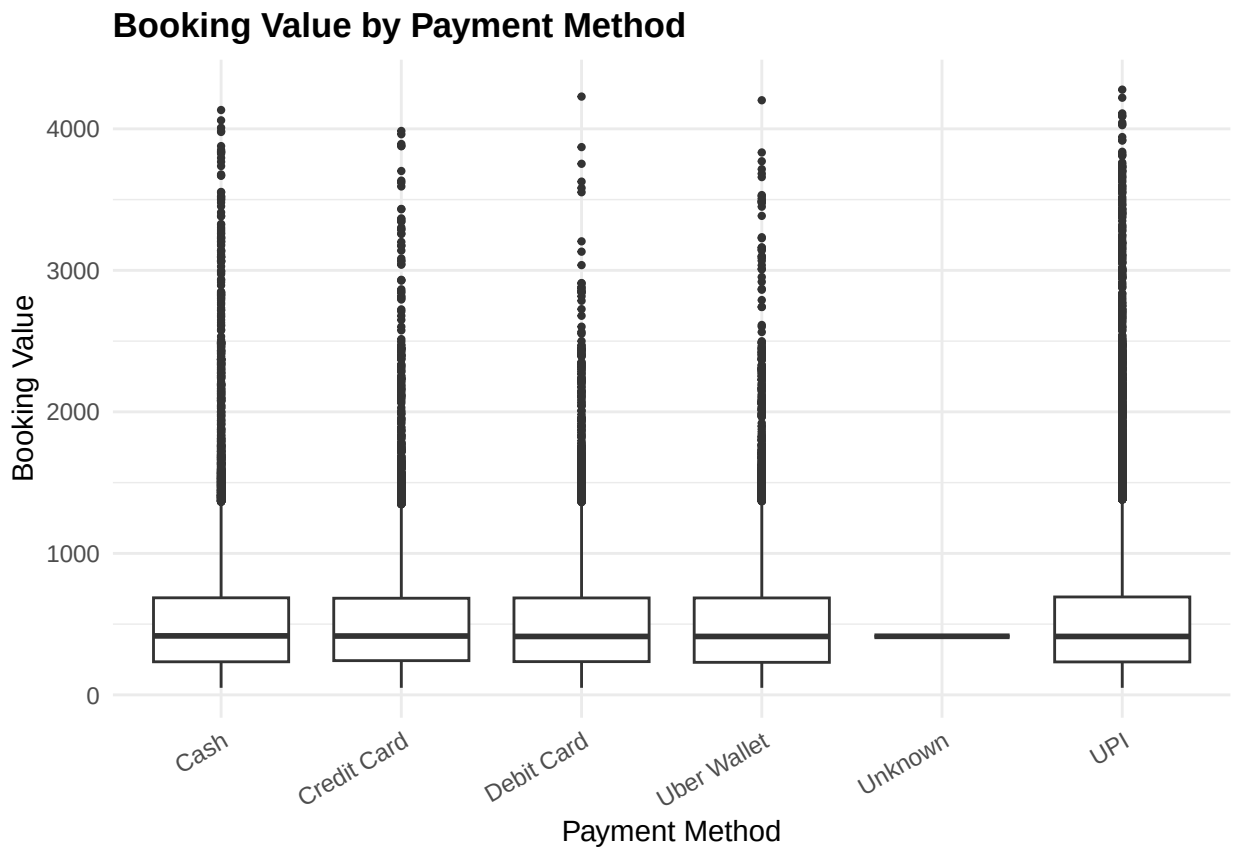



```r
# 5) Repeat the categorical count plots (explicit parity with Python)
p_booking_status2 <- ggplot(data %>% mutate(booking_status = fct_infreq(booking_status)),
                            aes(x = booking_status)) +
  geom_bar() +
  labs(title = "Distribution of Booking Status", x = "Booking Status", y = "Count") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 30, hjust = 1))

p_vehicle_type2 <- ggplot(data %>% mutate(vehicle_type = fct_infreq(vehicle_type)),
                          aes(x = vehicle_type)) +
  geom_bar() +
  labs(title = "Distribution of Vehicle Type", x = "Vehicle Type", y = "Count") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 30, hjust = 1))

print(plot_grid(p_booking_status2, p_vehicle_type2, nrow = 1))
```






```r
# 6) Boxplot: Booking Value by Payment Method
p_payment_box <- ggplot(data, aes(x = payment_method, y = booking_value)) +
  geom_boxplot(outlier.size = 0.8) +
  labs(title = "Booking Value by Payment Method",
       x = "Payment Method", y = "Booking Value") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 30, hjust = 1),
        plot.title = element_text(face = "bold"))

print(p_payment_box)
```



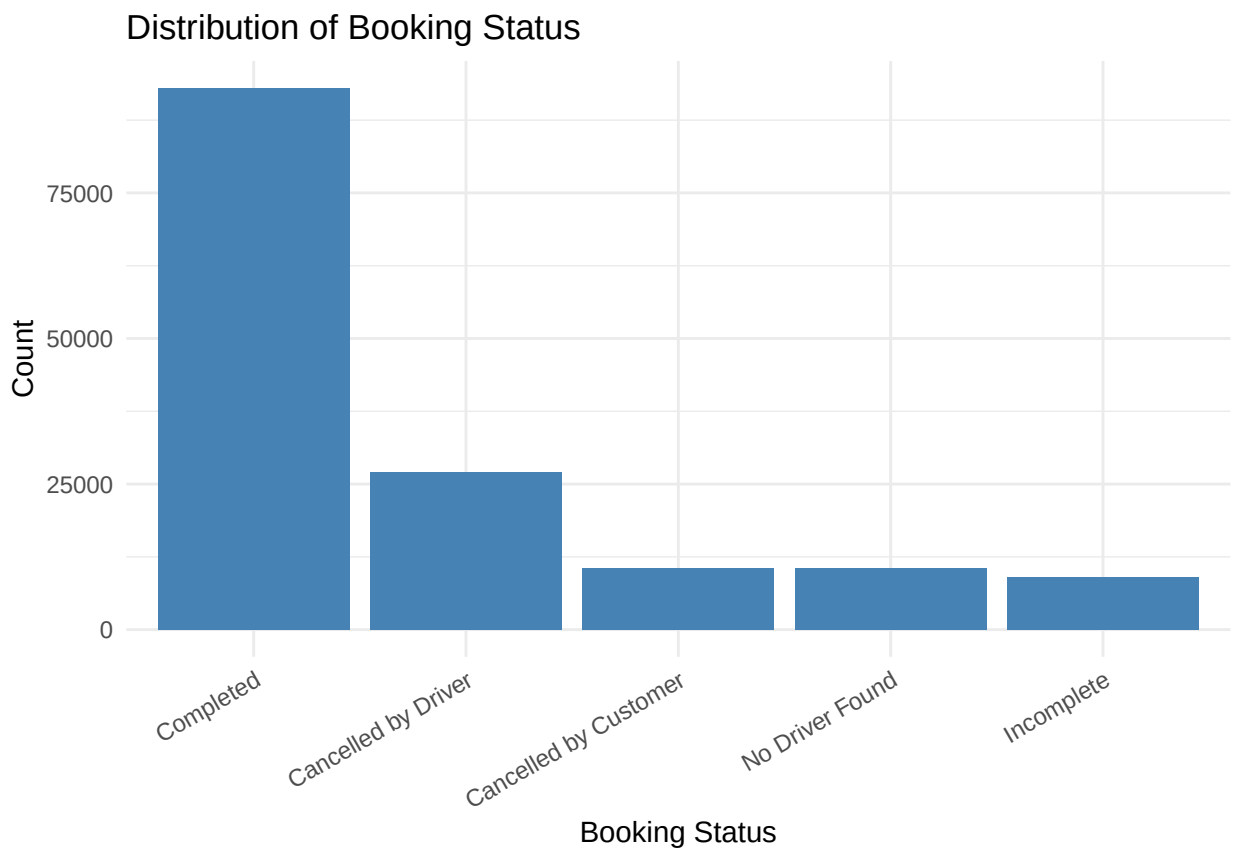
Write 3–5 short observations from EDA here.

- **Customer Satisfaction:** The majority of customers provide an average driver rating of 4.5 stars, reflecting a consistently high level of satisfaction with ride experiences.
- **Cancellations:** Driver-initiated cancellations (~25,000) significantly exceed customer-initiated cancellations (~10,000). This points to supply-side challenges such as inadequate incentives, route mismatches, or operational constraints (e.g., traffic conditions).
- **Pricing Dynamics:** Booking values exhibit high variability even at comparable ride distances, suggesting that fare determination is influenced by multiple factors beyond distance alone, including vehicle type, surge pricing, and demand-supply dynamics.

Task 5 — Data Visualization

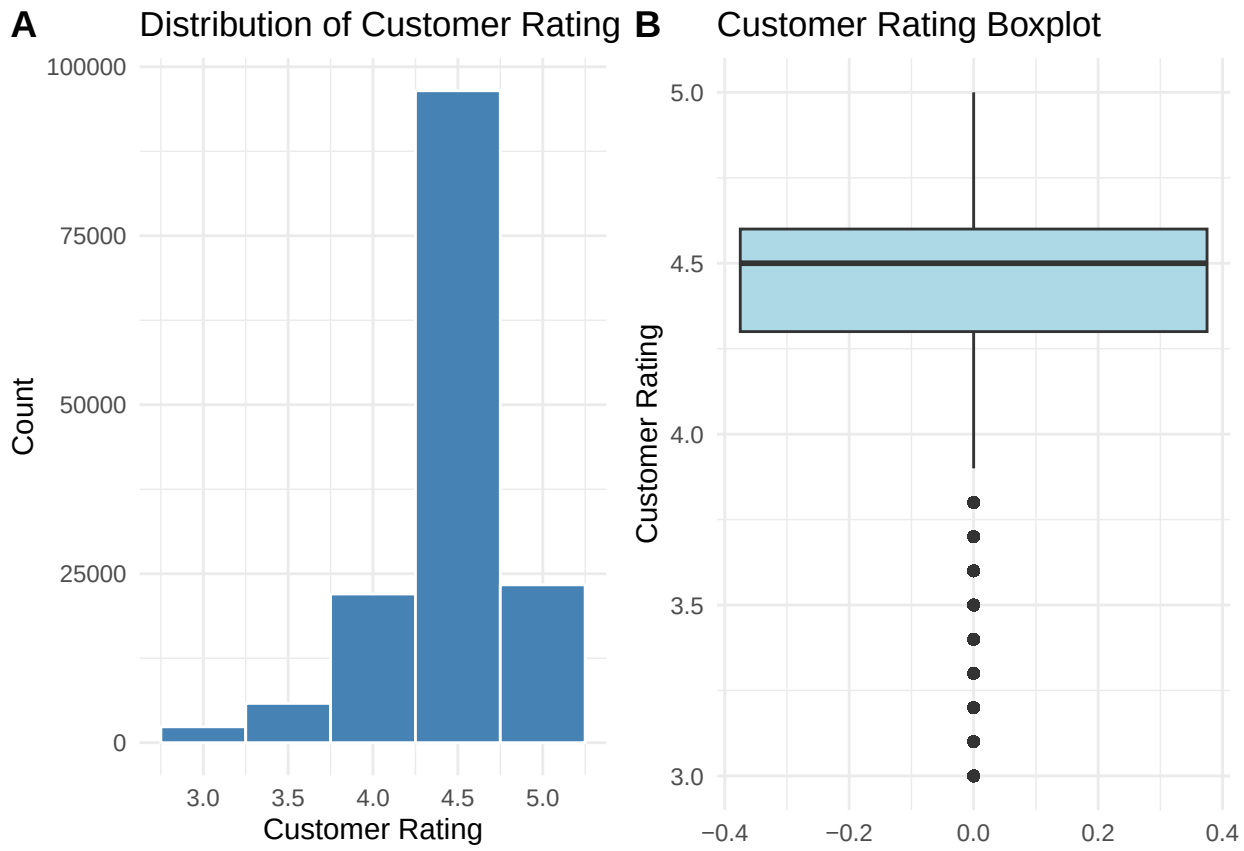
```
# --- Plot 1: Booking Status distribution ---
p1 <- ggplot(data, aes(x = booking_status)) +
  geom_bar(fill = "steelblue") +
  scale_x_discrete(limits = names(sort(table(data$booking_status), decreasing = TRUE))) +
  labs(
    title = "Distribution of Booking Status",
    x = "Booking Status",
    y = "Count"
```

```
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 30, hjust = 1))

print(p1)
```




```
# --- Plot 2: Booking Value by Payment Method ---

# Histogram of Customer Rating
p2 <- ggplot(data, aes(x = customer_rating)) +
  geom_histogram(binwidth = 0.5, fill = "steelblue", color = "white") +
  labs(
    title = "Distribution of Customer Rating",
    x = "Customer Rating",
    y = "Count"
  ) +
  theme_minimal()

# Boxplot of Customer Rating
p3 <- ggplot(data, aes(y = customer_rating)) +
  geom_boxplot(fill = "lightblue") +
  labs(
    title = "Customer Rating Boxplot",
    y = "Customer Rating"
  ) +
```

```r
  theme_minimal()

# Combine plots side by side
plot_grid(p2, p3, labels = c("A", "B"))
```



Key takeaway 1: Driver cancelations significantly exceeds customer cancelations. This points to supply-side challenges such as inadequate incentives, route mismatches, or operational constraints (e.g., traffic conditions).

Key takeaway 2: Most customers assign drivers an average rating of 4.5 stars, indicating a consistently high level of satisfaction and a strong overall service experience.