

Bike Sharing Washington

EB04 - Group 2

Zhanica Arrindell (596427) Vera Gak Anagrova (772713)
Aleksandra Tatkó (648925) Ly Le (644801)

November 30, 2025

Contents

| | | |
|----------|--|----------|
| 1 | Introduction | 1 |
| 2 | Distribution of Rentals by Weather | 2 |
| 3 | Weekly Bike Rentals in 2012 | 3 |
| 4 | Proportion of Casual Rentals by Weekday | 4 |
| 5 | Rentals vs Temperature | 5 |
| 6 | Appendix | 6 |

1 Introduction

This report presents the analysis of the Bike Sharing Washington dataset, following tasks A–D from the assignment instructions. All plots are included in the main body, while the full R code can be found in the appendix.

2 Distribution of Rentals by Weather

We investigate how the total number of bike rentals varies across different weather conditions. Three density plots are created, one for each level of `weathersit` (Clear, Cloudy/Not Clear, Rain/Snow). The densities are transparent to allow for overlap and easy comparison.

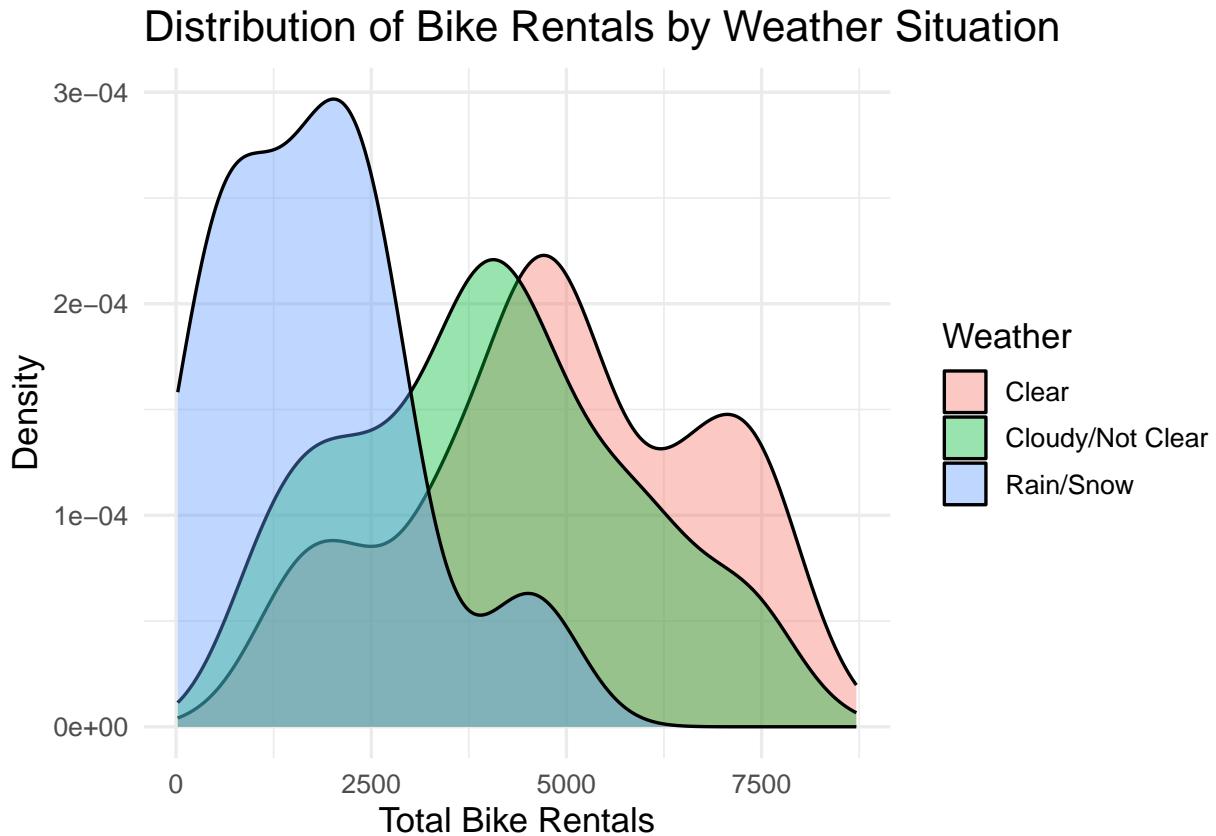


Figure 1. Distribution of Rentals by Weather

3 Weekly Bike Rentals in 2012

The dates are converted to proper Date objects and filtered to include only observations from 2012. We then compute weekly totals for overall rentals, registered users, and casual users.

A line plot shows the weekly patterns of all three groups.

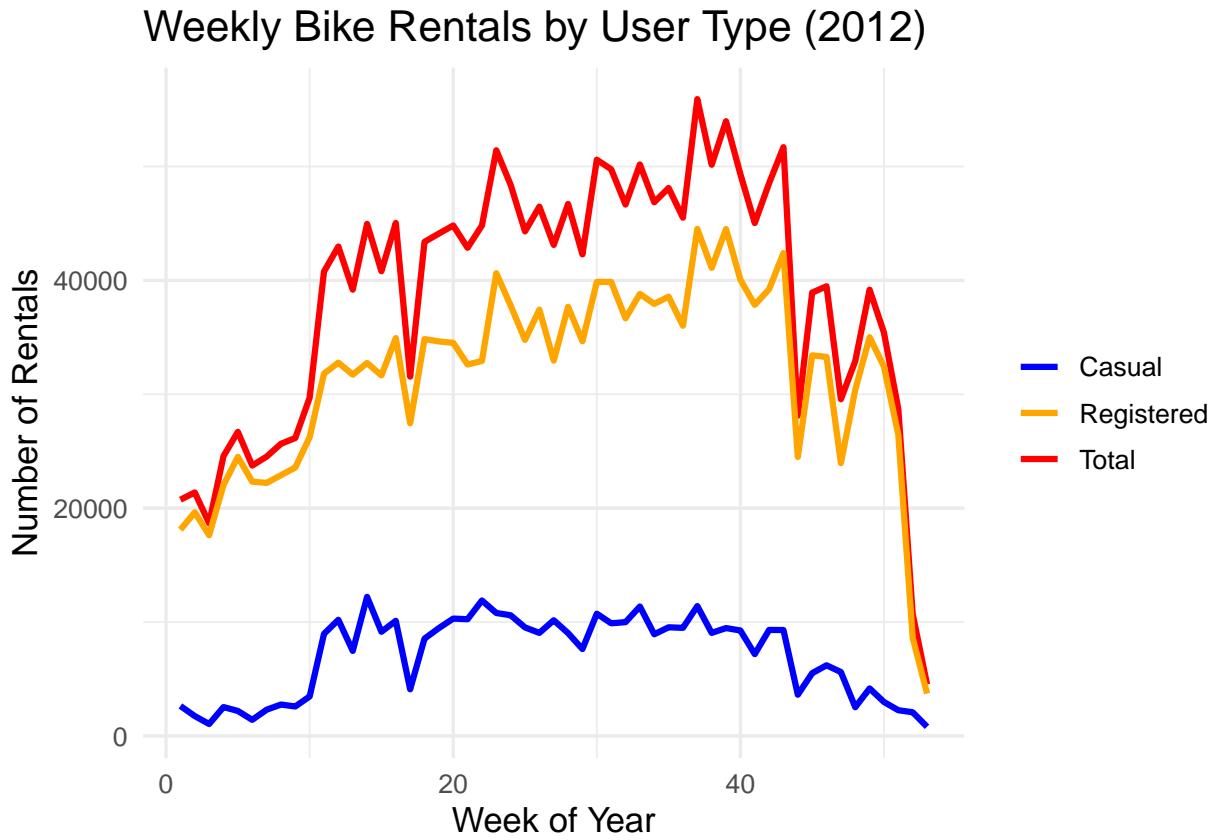


Figure 2. Weekly Bike Rentals by User Type (2012)

4 Proportion of Casual Rentals by Weekday

To study usage behavior by day of the week, we compute the share of rentals by casual users relative to total rentals.

We then plot a boxplot of this proportion across weekdays, adding a red point to indicate the mean for each day.

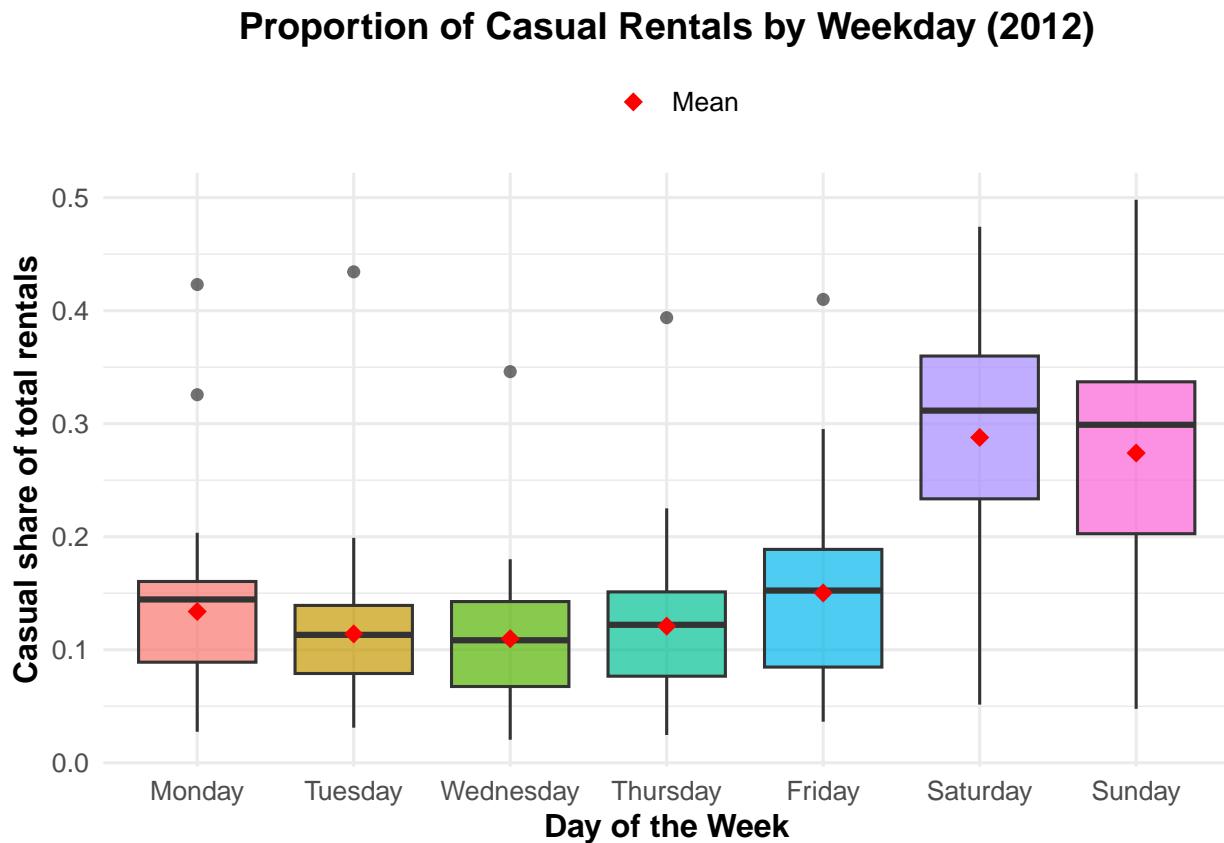


Figure 3. Proportion of Casual Rentals by Weekday (2012)

5 Rentals vs Temperature

Finally, we investigate how temperature affects total rentals. The temp variable is denormalized to Celsius ($^{\circ}\text{C}$).

Using base R graphics, we create a scatter plot of rentals against temperature, distinguishing holidays (orange triangles) from non-holidays (blue circles).

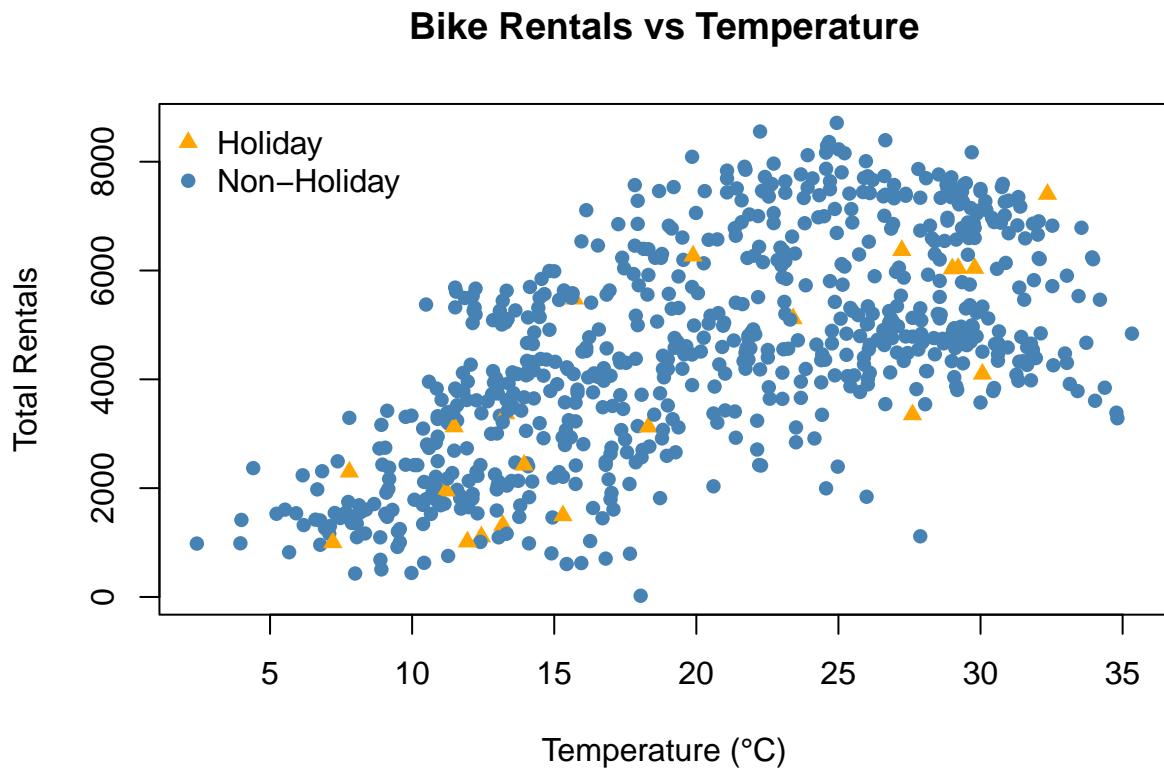


Figure 4. Bike Rentals vs. Temperature

6 Appendix

```
# Load all the libraries
knitr::opts_chunk$set(echo = FALSE, message=FALSE, warning=FALSE)
library(ggplot2)
library(dplyr)
library(lubridate)
library(tidyr)
library(kableExtra)
library(stargazer)

# Load the dataset
library(readr)
Bikes <- read_csv("Bikes_Washington.csv")
str(Bikes)

Bikes <- read.csv("Bikes_Washington.csv")

Bikes$weathersit <- factor(
  Bikes$weathersit,
  levels = c(1,2,3),
  labels = c("Clear", "Cloudy/Not Clear", "Rain/Snow")
)

ggplot(Bikes, aes(x = cnt, fill = weathersit)) +
  geom_density(alpha = 0.4) +
  labs(
    title = "Distribution of Bike Rentals by Weather Situation",
    x = "Total Bike Rentals",
    y = "Density",
    fill = "Weather"
  ) +
  theme_minimal(base_size = 13)

# Convert the dates in dteday to Date objects. Next, select the dates in 2012.
Bikes$dteday <- as.Date(Bikes$dteday)
bikes2012 <- filter(Bikes, format(dteday, "%Y") == "2012")
bikes2012$week <- as.integer(strftime(bikes2012$dteday, format = "%U"))

weekly <- aggregate(cbind(cnt, registered, casual) ~ week, data = bikes2012, sum)

ggplot(weekly, aes(x = week)) +
  geom_line(aes(y = cnt, color = "Total"), size = 1.1) +
  geom_line(aes(y = registered, color = "Registered"), size = 1.1) +
  geom_line(aes(y = casual, color = "Casual"), size = 1.1) +
  scale_color_manual(
    values = c("Total" = "red", "Registered" = "orange", "Casual" = "blue"),
    name = NULL
  ) +
  labs(
    title = "Weekly Bike Rentals by User Type (2012)",
    x = "Week of Year",
    y = "Number of Rentals"
```

```

) +
theme_minimal(base_size = 13)

bikes2012$weekday <- factor(
  weekdays(bikes2012$dteday),
  levels = c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday", "Sunday")
)

bikes2012$prop_casual <- bikes2012$casual / bikes2012$cnt

ggplot(bikes2012, aes(x = weekday, y = prop_casual, fill = weekday)) +
  geom_boxplot(alpha = 0.7) +
  # Add mean with legend
  stat_summary(
    aes(color = "Mean"),
    fun = mean,
    geom = "point",
    shape = 18,
    size = 3
  ) +
  scale_color_manual(
    name = "",
    values = c("Mean" = "red")
) +
  labs(
    title = "Proportion of Casual Rentals by Weekday (2012)",
    x = "Day of the Week",
    y = "Casual share of total rentals"
) +
  theme_minimal(base_size = 12) +
  theme(
    plot.title = element_text(hjust = 0.5, face = "bold"),
    axis.title = element_text(face = "bold"),
    legend.position = "top"
  ) +
  guides(fill = "none")

Bikes$temp_celsius <- Bikes$temp * 41
cols <- ifelse(Bikes$holiday == 1, "orange", "steelblue")
shaps <- ifelse(Bikes$holiday == 1, 17, 16)

plot(Bikes$temp_celsius, Bikes$cnt,
      col = cols, pch = shaps,
      xlab = "Temperature (°C)",
      ylab = "Total Rentals",
      main = "Bike Rentals vs Temperature")
legend("topleft",
       legend = c("Holiday", "Non-Holiday"),
       col = c("orange", "steelblue"),
       pch = c(17, 16),
       bty = "n")

```