

Data Visualization_Bikeshare

Mingi Um

2023-10-12

```
library("tidyverse")
library("skimr")
library("dplyr")
library("ggplot2")
library("scales")
```

1.1 Loading the data and preparation

```
# Load the data set
# read the CSV file into a data frame 'bikesharingdata'
bikesharingdata <- read_csv("bikesharedailydata.csv")

## Rows: 731 Columns: 16
## -- Column specification -----
## Delimiter: ","
## chr  (1): dteday
## dbl (15): instant, season, yr, mnth, holiday, weekday, workingday, weathersi...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
# look at all the variables by using the skim function
skim(bikesharingdata)
```

Table 1: Data summary

Name	bikesharingdata
Number of rows	731
Number of columns	16
Column type frequency:	
character	1
numeric	15
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
dteday	0	1	6	8	0	731	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
instant	0	1	366.00	211.17	1.00	183.50	366.00	548.50	731.00	
season	1	1	2.50	1.11	1.00	2.00	3.00	3.00	4.00	
yr	0	1	0.50	0.50	0.00	0.00	1.00	1.00	1.00	
mnth	1	1	6.53	3.45	1.00	4.00	7.00	10.00	12.00	
holiday	0	1	0.03	0.17	0.00	0.00	0.00	0.00	1.00	
weekday	0	1	3.00	2.00	0.00	1.00	3.00	5.00	6.00	
workingday	0	1	0.68	0.47	0.00	0.00	1.00	1.00	1.00	
weathersit	0	1	1.40	0.54	1.00	1.00	1.00	2.00	3.00	
temp	0	1	0.50	0.18	0.06	0.34	0.50	0.66	0.86	
atemp	0	1	0.47	0.16	0.08	0.34	0.49	0.61	0.84	
hum	0	1	0.63	0.14	0.00	0.52	0.63	0.73	0.97	
windspeed	0	1	0.19	0.08	0.02	0.13	0.18	0.23	0.51	
casual	0	1	848.18	686.62	2.00	315.50	713.00	1096.00	3410.00	
registered	0	1	3656.17	1560.26	20.00	2497.00	3662.00	4776.50	6946.00	
cnt	0	1	4504.35	1937.21	22.00	3152.00	4548.00	5956.00	8714.00	

```
# We observe that the 'season' and 'mnth' columns each have one missing value.
# We can omit these missing values as they are minimal and unlikely to impact the analysis.
bikesharingdata_no_na <- na.omit(bikesharingdata)
```

```
# Verify that missing values have been successfully omitted
skim(bikesharingdata_no_na)
```

Table 4: Data summary

Name	bikesharingdata_no_na
Number of rows	729
Number of columns	16
Column type frequency:	
character	1
numeric	15
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
dteday	0	1	6	8	0	729	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
instant	0	1	366.98	210.62	1.00	185.00	367.00	549.00	731.00	
season	0	1	2.50	1.11	1.00	2.00	3.00	3.00	4.00	
yr	0	1	0.50	0.50	0.00	0.00	1.00	1.00	1.00	

skim_variabl	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
mnth	0	1	6.53	3.44	1.00	4.00	7.00	10.00	12.00	
holiday	0	1	0.03	0.17	0.00	0.00	0.00	0.00	1.00	
weekday	0	1	3.00	2.00	0.00	1.00	3.00	5.00	6.00	
workingday	0	1	0.68	0.47	0.00	0.00	1.00	1.00	1.00	
weathersit	0	1	1.40	0.54	1.00	1.00	1.00	2.00	3.00	
temp	0	1	0.50	0.18	0.06	0.34	0.50	0.66	0.86	
atemp	0	1	0.48	0.16	0.08	0.34	0.49	0.61	0.84	
hum	0	1	0.63	0.14	0.00	0.52	0.63	0.73	0.97	
windspeed	0	1	0.19	0.08	0.02	0.13	0.18	0.23	0.51	
casual	0	1	850.24	686.42	2.00	317.00	721.00	1097.00	3410.00	
registered	0	1	3662.58	1557.58	20.00	2506.00	3667.00	4790.00	6946.00	
cnt	0	1	4512.82	1933.08	22.00	3190.00	4549.00	5976.00	8714.00	

```
# Define common month labels
month_labels <- c('1' = 'Jan', '2' = 'Feb', '3' = 'Mar', '4' = 'Apr', '5' = 'May', '6' = 'Jun',
                  '7' = 'Jul', '8' = 'Aug', '9' = 'Sep', '10' = 'Oct', '11' = 'Nov', '12' = 'Dec')
```

1.2 Define the color and characteristics of charts

```
# We define the color to use
chartcolor <- "#00356b"

# And then we create 'mychartattributes' to make the code more simple
mychartattributes <- theme_bw() + theme(text=element_text(family="serif")) + theme(panel.border = element_rect(fill="white", stroke="black", strokewidth=1))

# Filtering 2012 data
bikesharingdata_2012 <- bikesharingdata_no_na %>% filter(yr == 1)
```

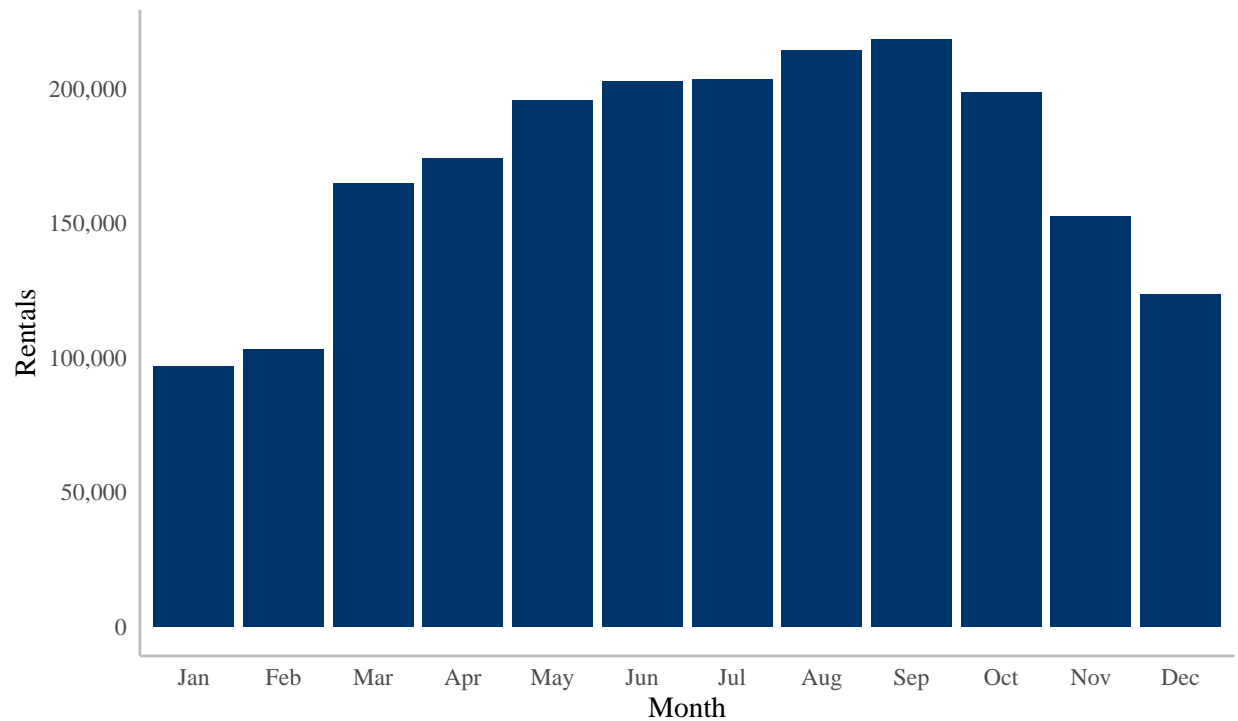
2.1 Bar chart

```
# Bar chart with 2012 data

ggplot(bikesharingdata_2012, aes(x=as.factor(mnth), y=cnt)) + geom_bar(stat="identity", fill = chartcolor)
```

Monthly Trend of Bike Rentals

Highlighting monthly variations in bike rentals for 2012



resents the total number of bike rentals per month, indicating noticeable fluctuations and potentially identifying peak usage periods

2.2 Line chart

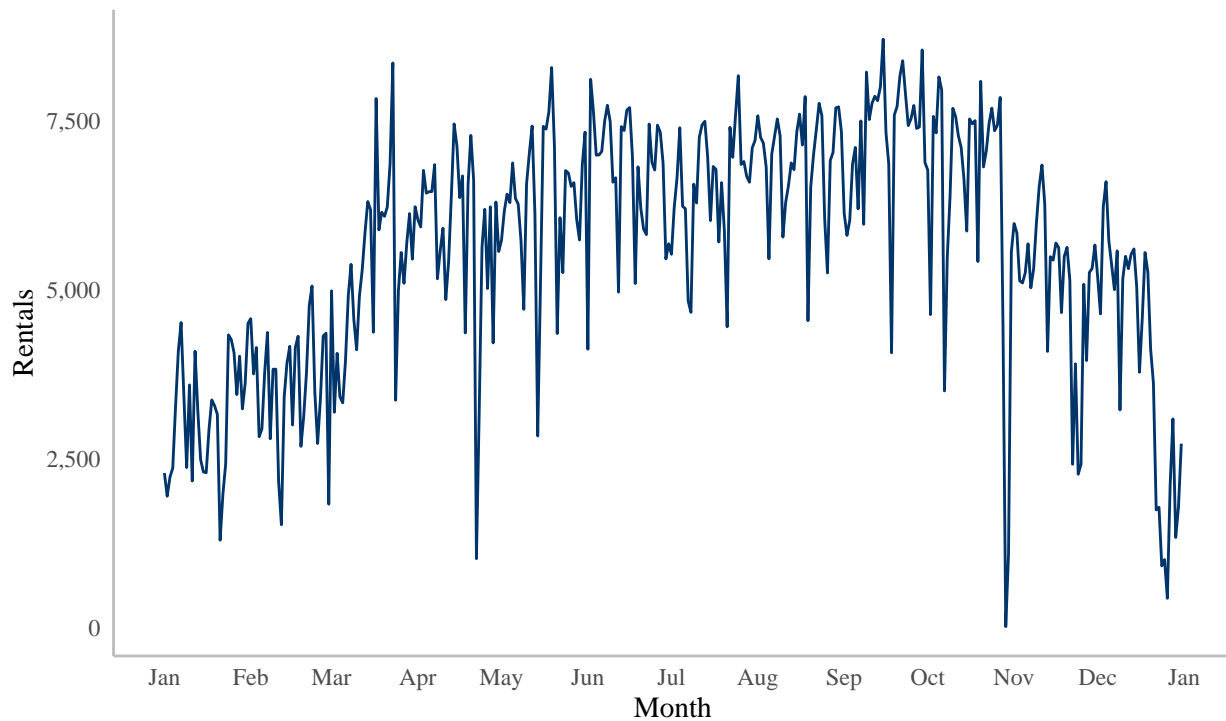
```
# Line chart with 2012 data
```

```
linechart_bikeshare <- ggplot(bikesharingdata_2012, aes(x=as.Date(dteday, format="%m/%d/%y"), y=cnt)) +  
  geom_line(color = chartcolor) +  
  scale_x_date(date_labels = "%b", date_breaks = "1 month") +  
  scale_y_continuous(labels = comma) +  
  labs(title = "Daily Trend of Bike Rentals in 2012",  
        subtitle = "Observing fluctuations and patterns in daily rentals",  
        caption = "Line chart represents daily rental counts, providing a detailed view of its variability",  
        x = "Month",  
        y = "Rentals") + mychartattributes
```

```
linechart_bikeshare
```

Daily Trend of Bike Rentals in 2012

Observing fluctuations and patterns in daily rentals



Line chart represents daily rental counts, providing a detailed view of its variability throughout the year.

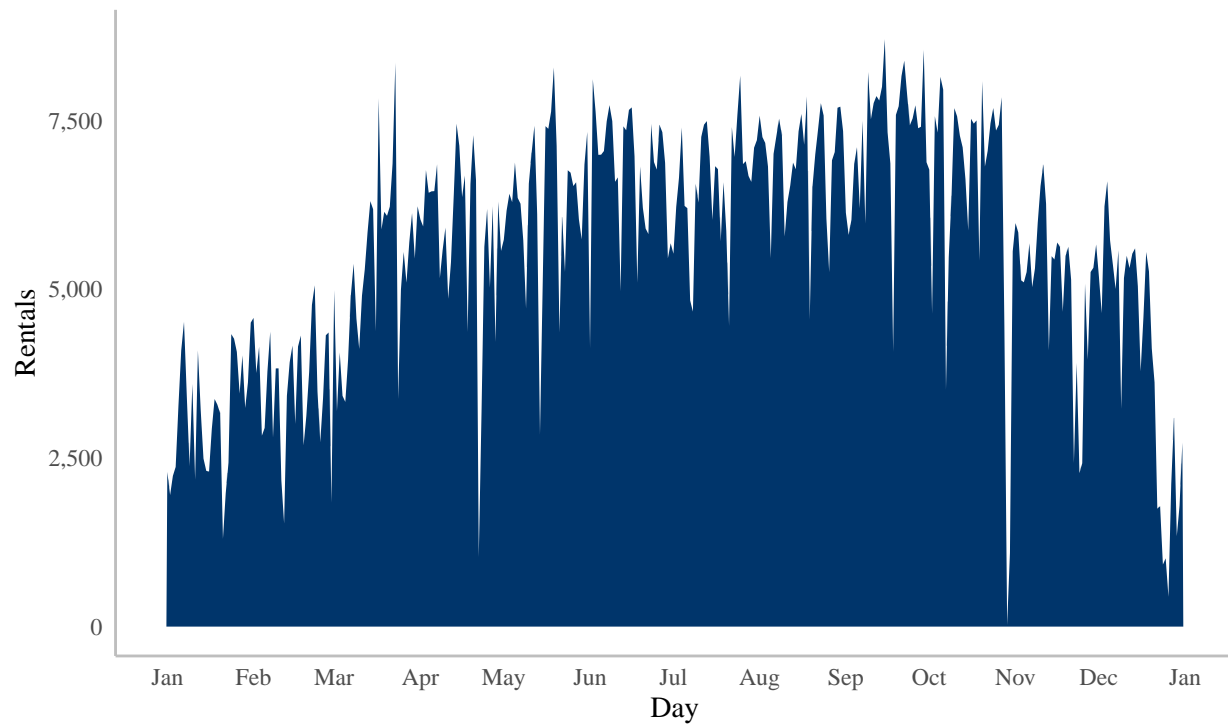
2.3 Stacked area chart

```
# Area chart with 2012 data
```

```
areachart_bikeshare <- ggplot(bikesharingdata_2012, aes(x=as.Date(dteday, format="%m/%d/%y"), y=cnt)) +  
  geom_area(fill = chartcolor) +  
  scale_x_date(date_labels = "%b", date_breaks = "1 month") +  
  scale_y_continuous(labels = comma) +  
  labs(title = "Daily Trend of Bike Rentals in 2012",  
        subtitle = "Visualizing the magnitude and flow of daily rentals",  
        caption = "Area chart explains the volume of daily bike rentals, unfolding the progression and p",  
        x = "Day",  
        y = "Rentals") +  
  mychartattributes  
areachart_bikeshare
```

Daily Trend of Bike Rentals in 2012

Visualizing the magnitude and flow of daily rentals



Area chart explains the volume of daily bike rentals, unfolding the progression and potential peaks throughout the year.

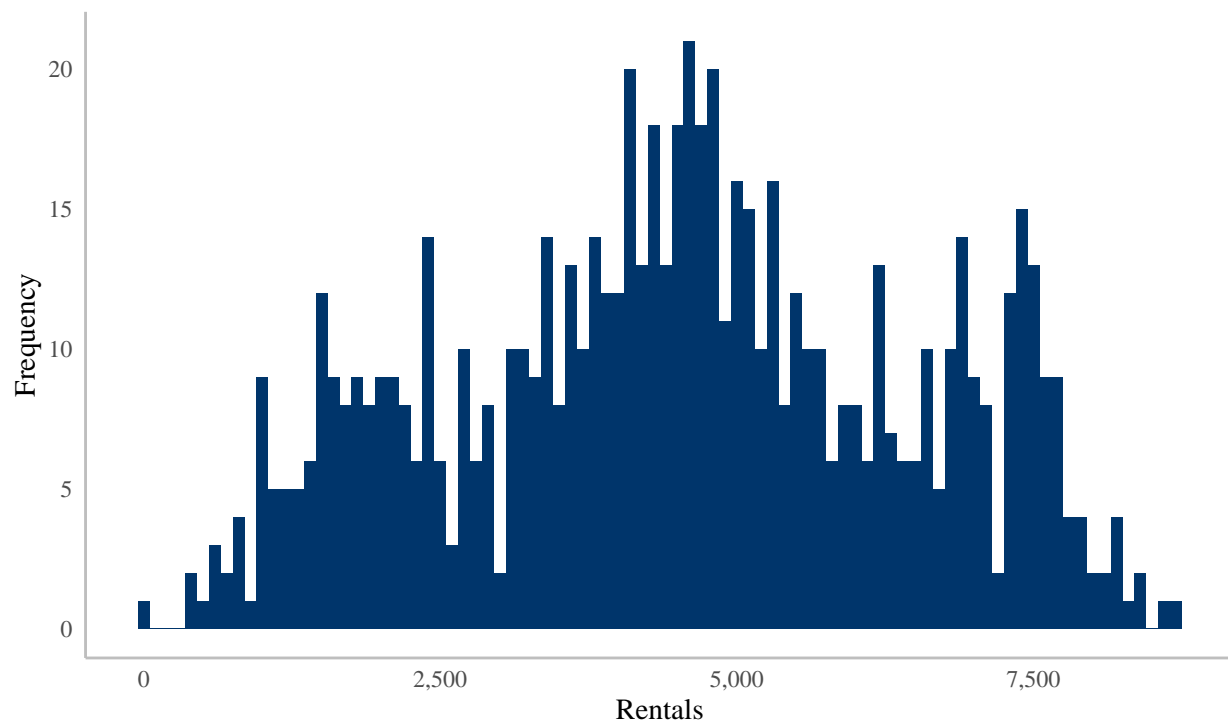
2.4 Histogram

```
# Histogram
```

```
hist_bikeshare <- ggplot(bikesharingdata_no_na, aes(x=cnt)) + geom_histogram(binwidth = 100, fill = cha  
hist_bikeshare
```

Distribution of Daily Bike Rentals

Examining the frequency distribution of rental counts



Histogram displays the distribution of rental counts

2.5 Density plot

```
# Density plot
```

```
densityplot_bikeshare <- ggplot(bikesharingdata_no_na, aes(x=cnt)) + geom_density(fill = chartcolor, col = chartcolor)
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
```

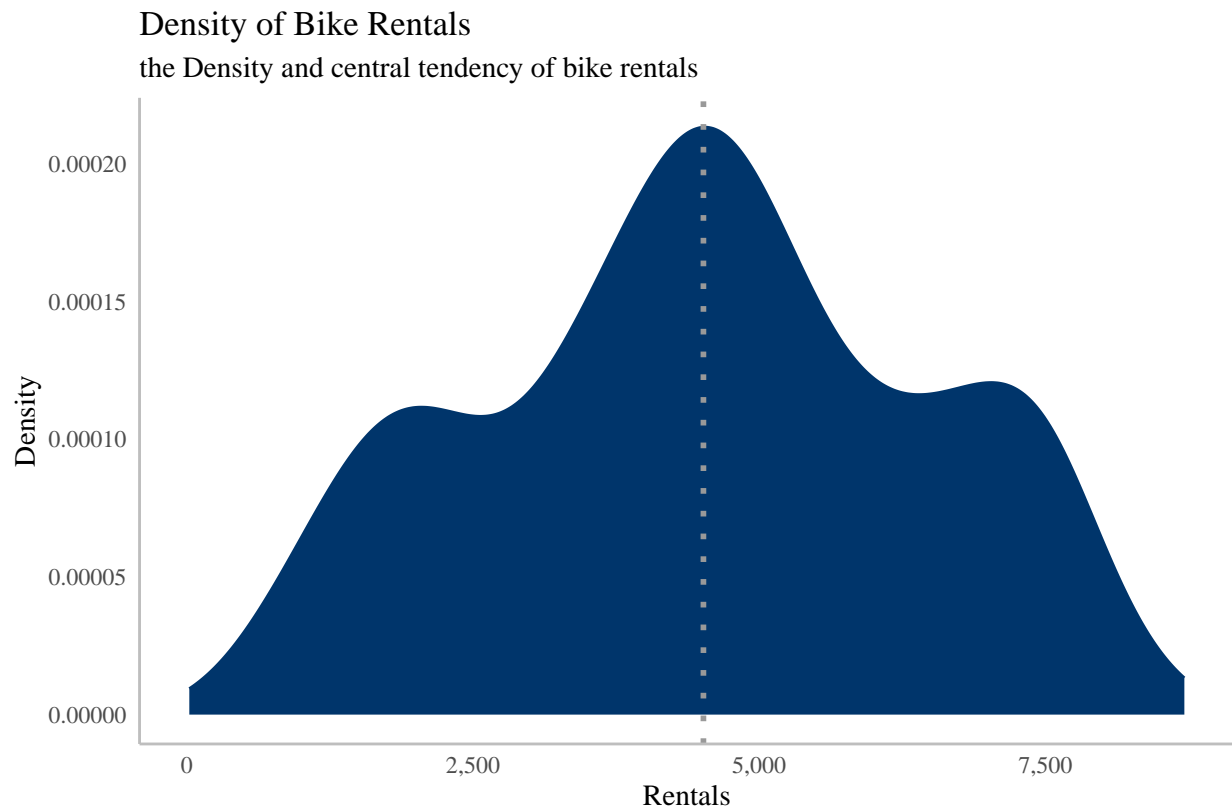
```
## i Please use `linewidth` instead.
```

```
## This warning is displayed once every 8 hours.
```

```
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
```

```
## generated.
```

```
densityplot_bikeshare
```



The chart provides insight into the pattern and concentration of the overall rental distributions

2.6 Boxplot

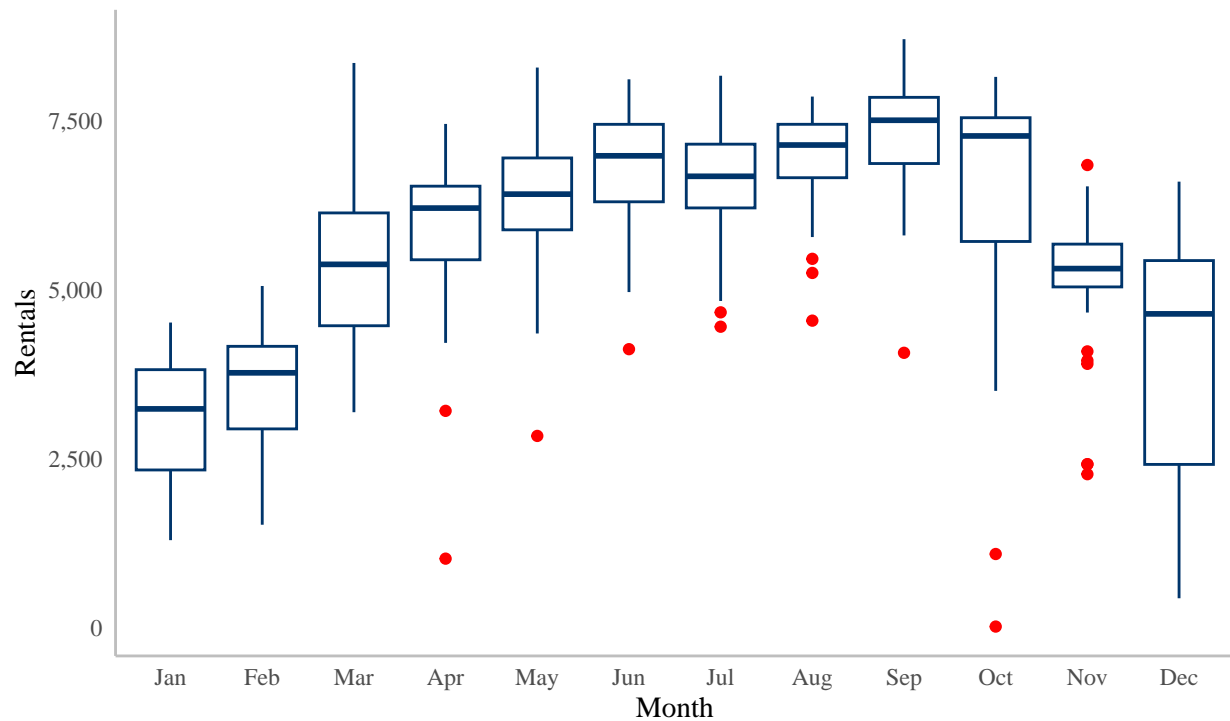
```
# Boxplot with 2012 data
```

```
boxplot_bikeshare <- ggplot(bikesharingdata_2012, aes(x=as.factor(mnth), y=cnt)) + geom_boxplot(color =
```

```
boxplot_bikeshare
```


Monthly Trend of Bike Rentals

Boxplots representing monthly distributions of rentals in 2012



Each boxplot visualizes the central tendency and variability of rentals each month, offering insights into seasonal rental patterns

2.7 Scatterplot chart

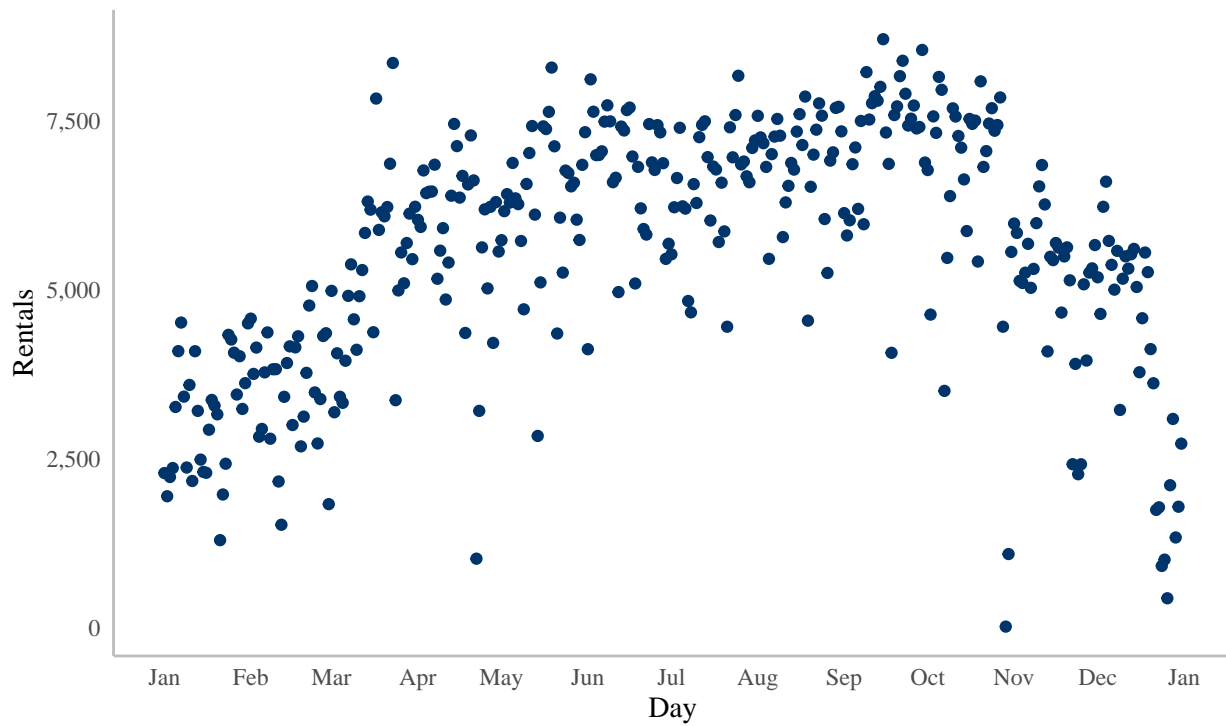
```
# Scatterplot with 2012 data
```

```
scatter_bikeshare <- ggplot(bikesharingdata_2012, aes(x=as.Date(dteday, format="%m/%d/%y"), y=cnt)) +  
  geom_point(color = chartcolor) +  
  labs(title = "Scatterplot of Daily Bike Rentals in 2012",  
        subtitle = "Exploring the Day-to-Day variability in rental counts",  
        caption = "Scatterplot illuminates the daily fluctuations in bike rentals, pinpointing specific c",  
        x = "Day",  
        y = "Rentals") +  
  scale_x_date(date_labels = "%b", date_breaks = "1 month") +  
  scale_y_continuous(labels = comma) +  
  mychartattributes
```

```
scatter_bikeshare
```

Scatterplot of Daily Bike Rentals in 2012

Exploring the Day-to-Day variability in rental counts



Scatterplot illuminates the daily fluctuations in bike rentals, pinpointing specific days with spikes or drops.