

Bikshare Datset Visualization

Willy Feid

2023-10-25

1. Data Exploration and Cleaning

Data Summary & Cleaning

- **instant**: This is an index
 - Rename to ID
- **dteday**: date column
 - Rename to Date
- **season**: Values 1-4. By looking at the data, we can determine that 1: Winter, 2: Spring, 3: Summer, 4: Winter
 - Rename seasons to Winter, Spring, Summer, Fall
 - Impute NA value with 1 (season NA value has data 1/7/11)
- **'yr'**: Has values 0 and 1. 0 = 2011, 1 = 2012
 - Rename to Year, and values to 2011 and 2012
- **mnth**: Values 1-12 for month
 - Rename to Month, and values to month names.
 - Impute NA value with 1 for January (date for this row is 1/10/11)
- **holiday**: 1= Holiday, 0 = Not Holiday
- **weekday**: 0=Sun, 1=Mon, ..., 6=Sat
 - Re label values to names of day of the week.
- **workingday**: 1=workday, 0=weekend or holiday.
- **weathersit**: values of 1, 2, 3. Seem to indicate general weather (ie, clear, cloudy, etc)
- **temp**: Temperature that has been normalized.
- **atemp**: Could be average temp or feeling temp. Also normalized.
- **hum**: Humidity, normalized
- **windspeed**: Wind speed normalized.
- **casual, registered, cnt**: casual + registered = cnt. We can infer that cnt is the total number of bike rentals for that day, and casual and registered are two types of renters.

Cleaning

```

# Fill NA values:
bikeshare$season[is.na(bikeshare$season)] <- 1
bikeshare$mnth[is.na(bikeshare$mnth)] <- 1

# Rename columns:
bikeshare <- bikeshare %>%
  rename(
    index = instant,
    date = dteday,
    year = yr,
    month = mnth,
    humidity = hum,
    rentals = cnt
  )

# Rename Values for season, year, and weekday
bikeshare <- bikeshare %>%
  mutate(
    season = case_when(
      season == 1 ~ "Winter",
      season == 2 ~ "Spring",
      season == 3 ~ "Summer",
      season == 4 ~ "Fall"),
    year = case_when(
      year == 0 ~ "2011",
      year == 1 ~ "2012"),
    weekday = case_when(
      weekday == 0 ~ "Sunday",
      weekday == 1 ~ "Monday",
      weekday == 2 ~ "Tuesday",
      weekday == 3 ~ "Wednesday",
      weekday == 4 ~ "Thursday",
      weekday == 5 ~ "Friday",
      weekday == 6 ~ "Saturday"),
    month = month(month, label = TRUE)
  )

# Reorder the factor levels of the date column
bikeshare$date <- factor(bikeshare$date, levels = unique(bikeshare$date))

# Subset for 2011
bikeshare_2011 <- bikeshare %>%
  filter(year == "2011")

# Subset for 2012
bikeshare_2012 <- bikeshare %>%
  filter(year == "2012")

head(bikeshare)

```

```

## # A tibble: 6 x 16
##   index date   season year  month holiday weekday  workingday weathersit temp
##   <dbl> <fct>   <chr>  <chr> <ord>   <dbl> <chr>          <dbl>      <dbl> <dbl>

```

```
## 1      1 1/1/11 Winter 2011 Jan          0 Saturday          0          2 0.344
## 2      2 1/2/11 Winter 2011 Jan          0 Sunday            0          2 0.363
## 3      3 1/3/11 Winter 2011 Jan          0 Monday            1          1 0.196
## 4      4 1/4/11 Winter 2011 Jan          0 Tuesday            1          1 0.2
## 5      5 1/5/11 Winter 2011 Jan          0 Wednesday          1          1 0.227
## 6      6 1/6/11 Winter 2011 Jan          0 Thursday           1          1 0.204
## # i 6 more variables: atemp <dbl>, humidity <dbl>, windspeed <dbl>,
## #   casual <dbl>, registered <dbl>, rentals <dbl>
```

2. Visualizations

```
# Defining chart attributes

chartcolor <- "#4E79A7"
chartcolor2 <- "#053661"
chartcolor_hunter <- "#589d62"
chartcolor_hunter2 <- "#467c4e"
chartcolor_gray <- "#5a5a5a"
caption <- "Source: Bikeshare Dataset (2011 - 2012) | Willy Feid"
caption_2011 <- "Source: Bikeshare Dataset (2011) | Willy Feid"
caption_2012 <- "Source: Bikeshare Dataset (2012) | Willy Feid"
mychartattributes <- theme_minimal() +
  theme(text=element_text(family = ".New York")) +
  theme(panel.border = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.line = element_line(color = "gray"),
        axis.line.y = element_blank(),
        axis.line.x = element_blank(),
        axis.ticks.x = element_blank(),
        axis.ticks.y = element_blank(),
        plot.title.position = "plot",
        plot.caption = element_text(size = 8, color=chartcolor_gray, ),
        plot.subtitle = element_text(size = 10),
        axis.text.x = element_text(size = 10),
        axis.text.y = element_text(size = 10),
        axis.title.x = element_text(size = 10, color = chartcolor_gray),
        axis.title.y = element_text(size = 10, color = chartcolor_gray))
```

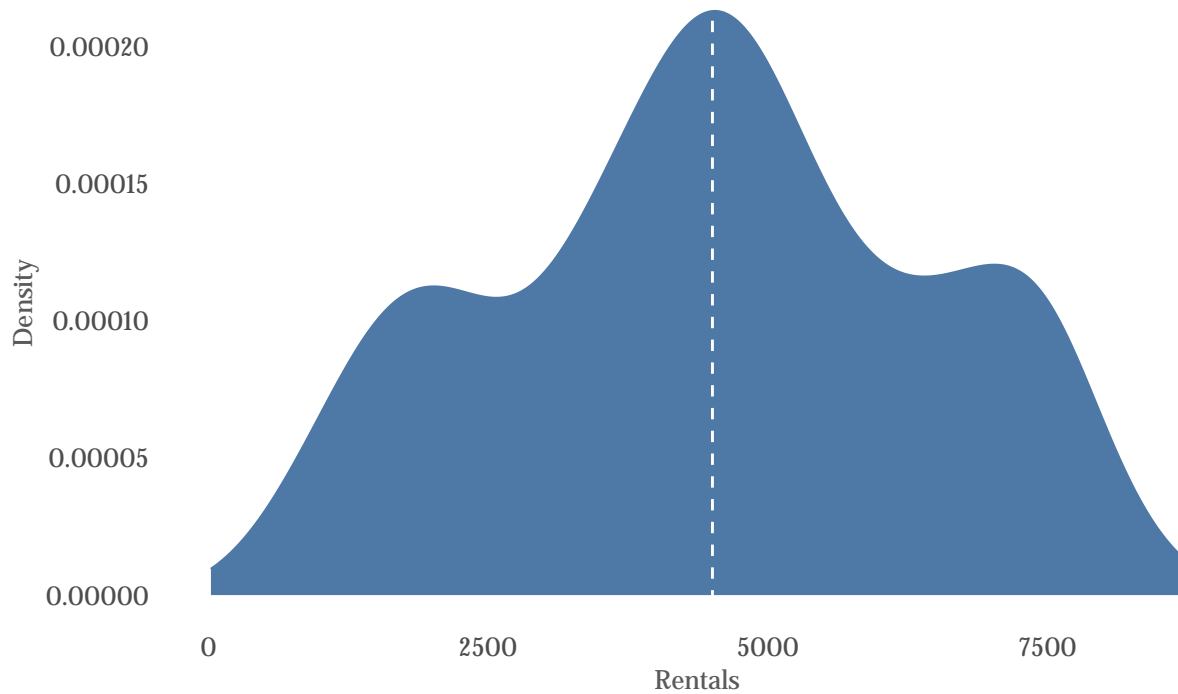
Density Plot

```
caption = "Source: Bikeshare Dataset (2011 - 2012) | Willy Feid"
ggplot(bikeshare, aes(x=rentals)) +
  geom_density(
    fill = chartcolor, color = chartcolor) + #add fill color
  labs(title = "Density Plot of Daily Bike Rentals",
        subtitle = "Daily bike rental fall around the mean (4,504) with notable smaller peaks above and below")
```

```
caption = caption,
x = "Rentals",
y = "Density") +
mychartattributes +
geom_vline(xintercept = round(mean(bikeshare$rentals),2), linewidth=0.5, color = "white", linetype =
```

Density Plot of Daily Bike Rentals

Daily bike rental fall around the mean (4,504) with notable smaller peaks above and below the mean.



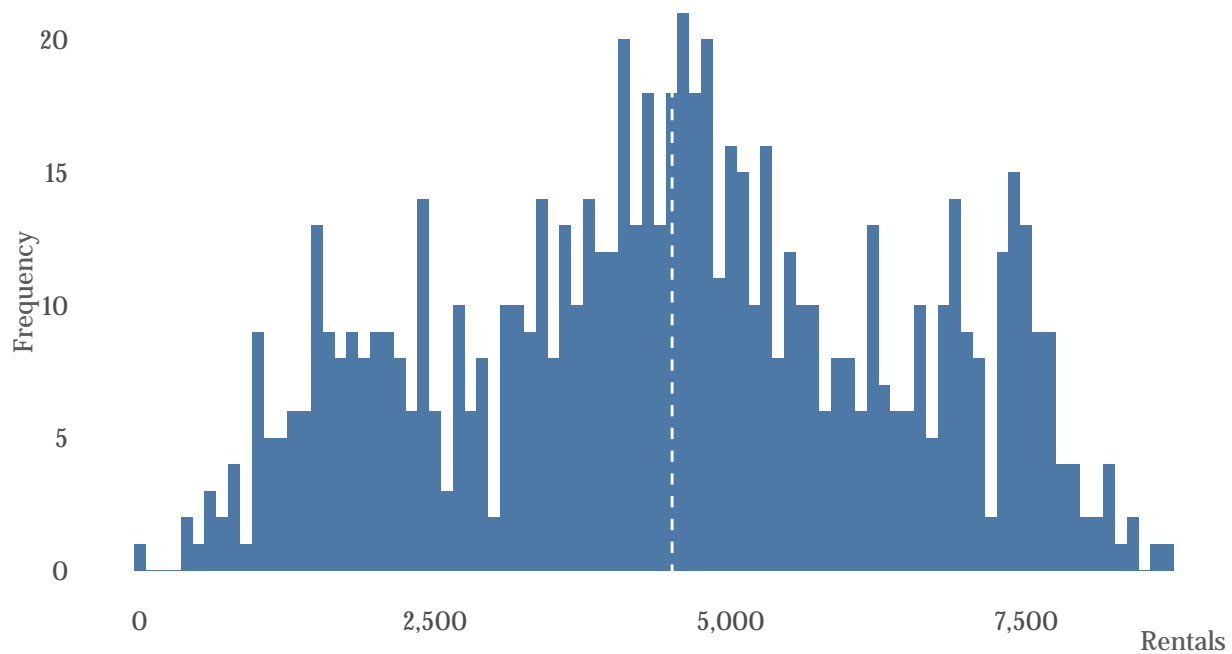
Source: Bikeshare Dataset (2011 – 2012) | Willy Feid

Histogram

```
ggplot(bikeshare, aes(x=rentals)) +
  geom_histogram(binwidth = 100, # set thickness of bins
                 bins = 100,
                 fill = chartcolor) +
  labs(title = "Histogram of Daily Bike Rentals from 2011 to 2012",
        subtitle = "Daily bike rental fall around the mean but there are peaks of very busy days, and \n",
        caption = caption,
        x = "Rentals",
        y = "Frequency") +
  mychartattributes +
  geom_vline(xintercept = round(mean(bikeshare$rentals),2), linewidth=0.5, color = "white", linetype =
  scale_x_continuous(labels = comma) +
  theme(axis.title.x = element_text(vjust = 3, hjust = 1),
        axis.title.y = element_text(angle = 90, vjust = 0, hjust = 0.5))
```

Histogram of Daily Bike Rentals from 2011 to 2012

Daily bike rental fall around the mean but there are peaks of very busy days, and very slow days (could be weather or seasonal conditions).



Source: Bikeshare Dataset (2011 – 2012) | Willy Feid

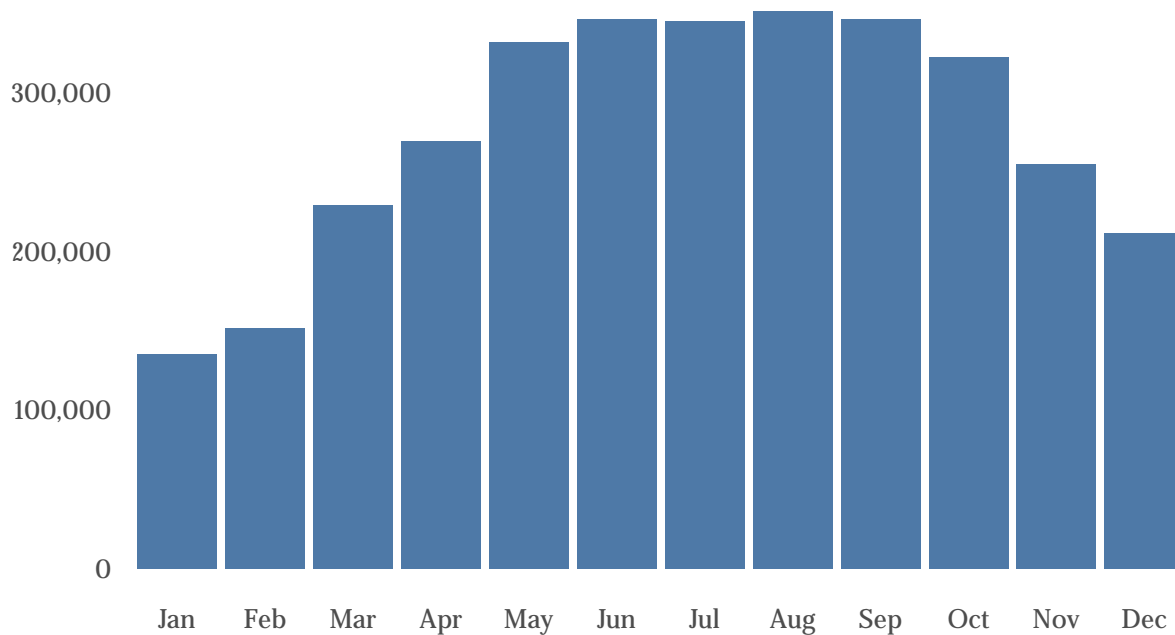
Bar Chart

Bike Rentals by Month between 2011 & 2012

```
# Getting the sum of rentals for each month (2011 + 2012)
monthly_rentals <- bikeshare %>%
  group_by(month) %>%
  summarise(total_rentals = sum(rentals))
# Plot
ggplot(monthly_rentals, aes(x=month, y=total_rentals)) +
  geom_bar(stat="identity", fill = chartcolor) +
  labs(title = "Bike Rentals by Month between 2011 & 2012",
       subtitle = "Bike rentals show a seasonality trends. Rentals are most popular in late summer and early fall.",
       caption = caption,
       x = " ",
       y = " ") +
  mychartattributes +
  scale_y_continuous(labels = comma)
```

Bike Rentals by Month between 2011 & 2012

Bike rentals show a seasonality trends. Rentals are most popular in late summer and early fall, with clear drop offs in colder winter months.



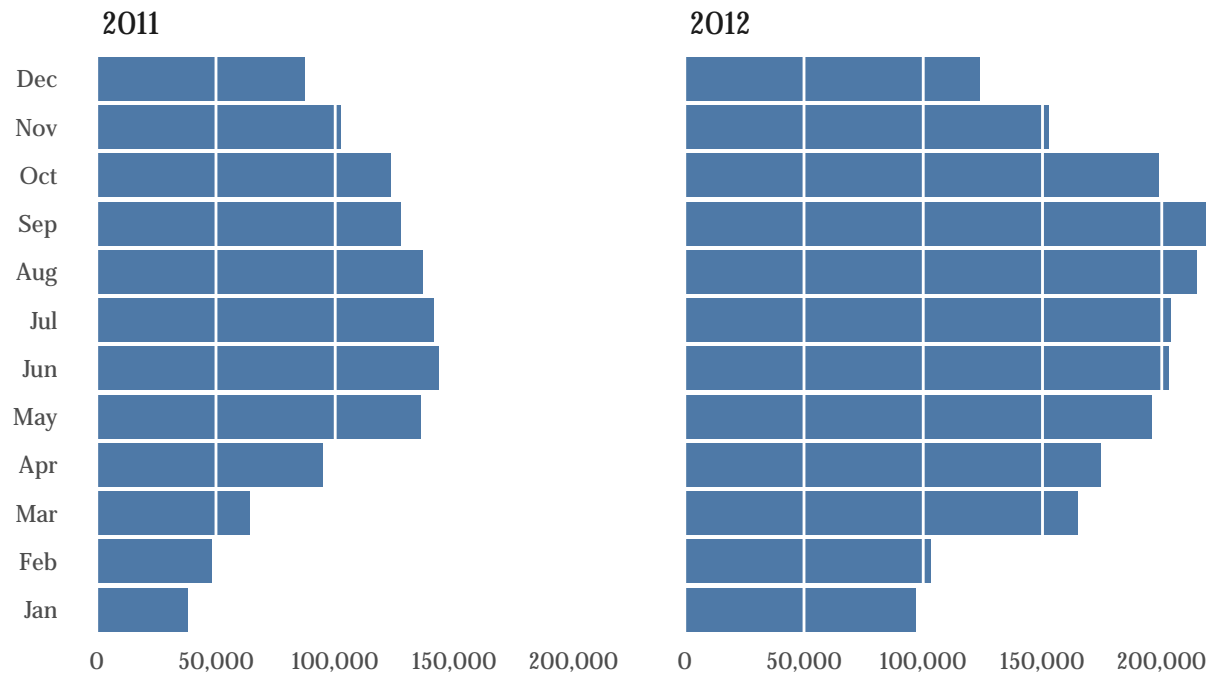
Source: Bikeshare Dataset (2011 – 2012) | Willy Feid

Rentals per Month by Year

```
ggplot(bikeshare) +  
  geom_bar(aes(x=rentals, y=month), stat="identity", fill = chartcolor) +  
  facet_grid(.~year) +  
  labs(title = "Bike Rentals by Month in 2011 and 2012",  
        subtitle = "Seasonality factors were consistent in both years, but overall rentals increased in 2012",  
        caption = caption,  
        x = " ",  
        y = NULL) +  
  theme_minimal() +  
  theme(  
    text=element_text(family = ".New York"),  
    plot.title.position = "plot",  
    plot.subtitle = element_text(size=10),  
    strip.text = element_text(size=12, hjust = 0.04),  
    plot.caption = element_text(hjust = 1, vjust = 1, size = 8, color="#5a5a5a"),  
    panel.grid.major = element_line('white', linewidth = 0.5),  
    panel.grid.minor = element_blank(),  
    panel.grid.major.y = element_blank(),  
    panel.ontop = TRUE  
  ) +  
  scale_x_continuous(labels = comma)
```

Bike Rentals by Month in 2011 and 2012

Seasonality factors were consistent in both years, but overall rentals increased in 2012.



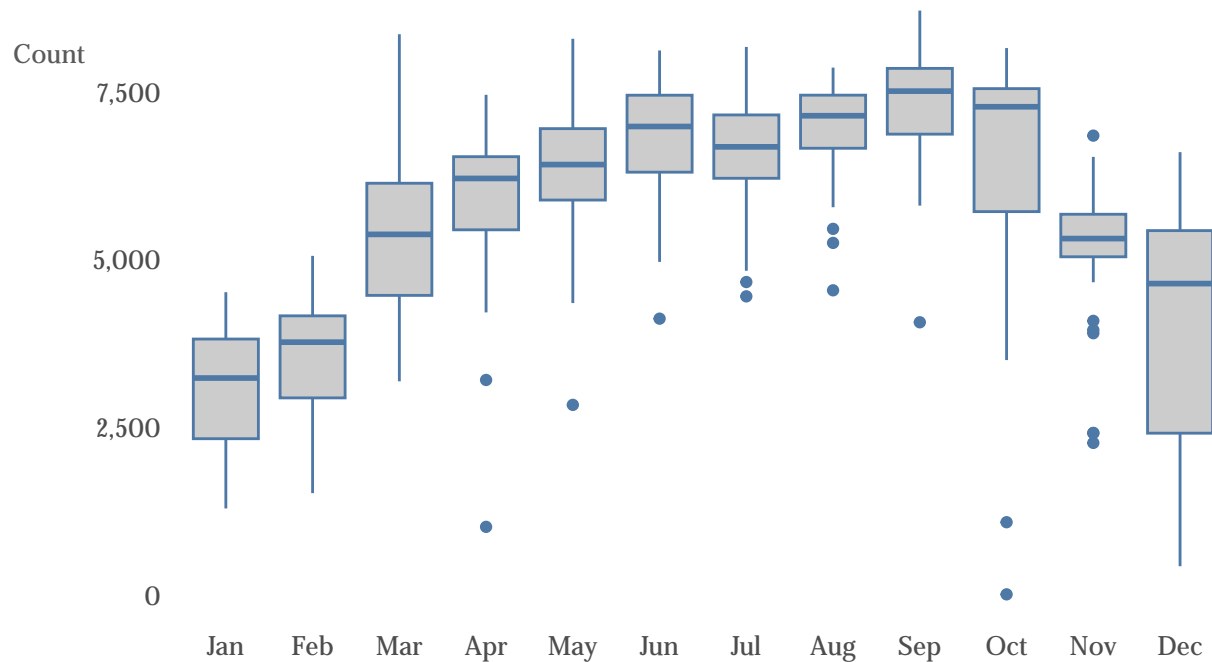
Source: Bikeshare Dataset (2011 – 2012) | Willy Feid

Box Plot

```
ggplot(bikeshare_2012, aes(x=month, y=rentals)) +  
  geom_boxplot(color = chartcolor, fill = "#cccccc") +  
  labs(title = "Monthly Distributions of Bike Rentals in 2012",  
        subtitle = "Higher variability in less popular monthly and more consistency in popular months.",  
        caption = caption_2012,  
        x = " ",  
        y = "Count") +  
  mychartattributes +  
  theme(axis.title.y = element_text(angle = 0, vjust = 0.9, hjust = 1)) +  
  scale_y_continuous(labels = comma)
```

Monthly Distributions of Bike Rentals in 2012

Higher variability in less popular monthly and more consistency in popular months.



Source: Bikeshare Dataset (2012) | Willy Feid

Line Chart

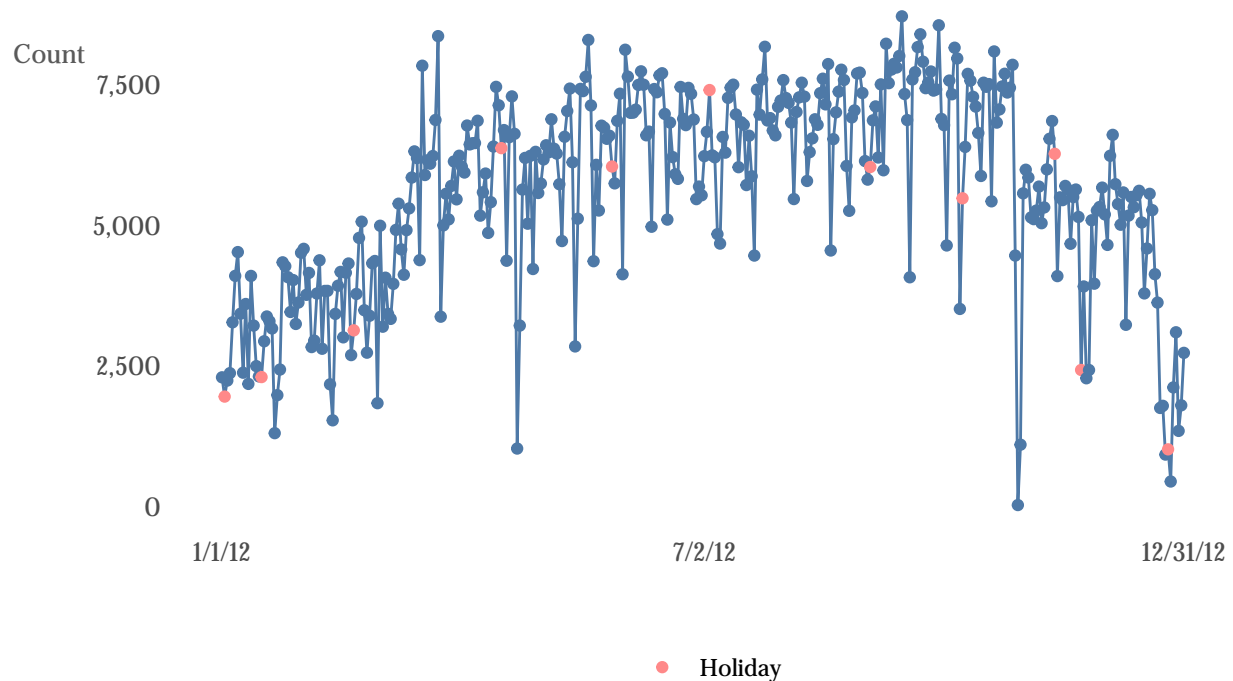
```
# for x-axis ticks
first_date <- "1/1/12"
mid_date <- "7/2/12"
last_date <- "12/31/12"
bikeshare_2012$holiday <- factor(bikeshare_2012$holiday, levels = c(0, 1))
#plot
ggplot(bikeshare_2012, aes(x=date, y=rentals, group=1)) +
  geom_line(color = chartcolor) +
  geom_point(aes(color=factor(holiday))) +
  labs(title = "Daily Bike Rentals in 2012",
       subtitle = "Holidays are not outliers but instead stay along the general trend.",
       caption = caption_2012,
       x = " ",
       y = "Count",
       color = NULL) +
  mychartattributes +
  theme(axis.title.y = element_text(angle = 0, vjust = 0.9, hjust = 1)) +
  scale_x_discrete(breaks = c(first_date, mid_date, last_date), expand = c(0.05, 0.05)) +
  scale_color_manual(values = c("0" = chartcolor, "1" = "#ff8989"),
                    breaks = "1",
                    labels = "Holiday") +
```



```
theme(legend.position = "bottom") +
scale_y_continuous(labels = comma)
```

Daily Bike Rentals in 2012

Holidays are not outliers but instead stay along the general trend.



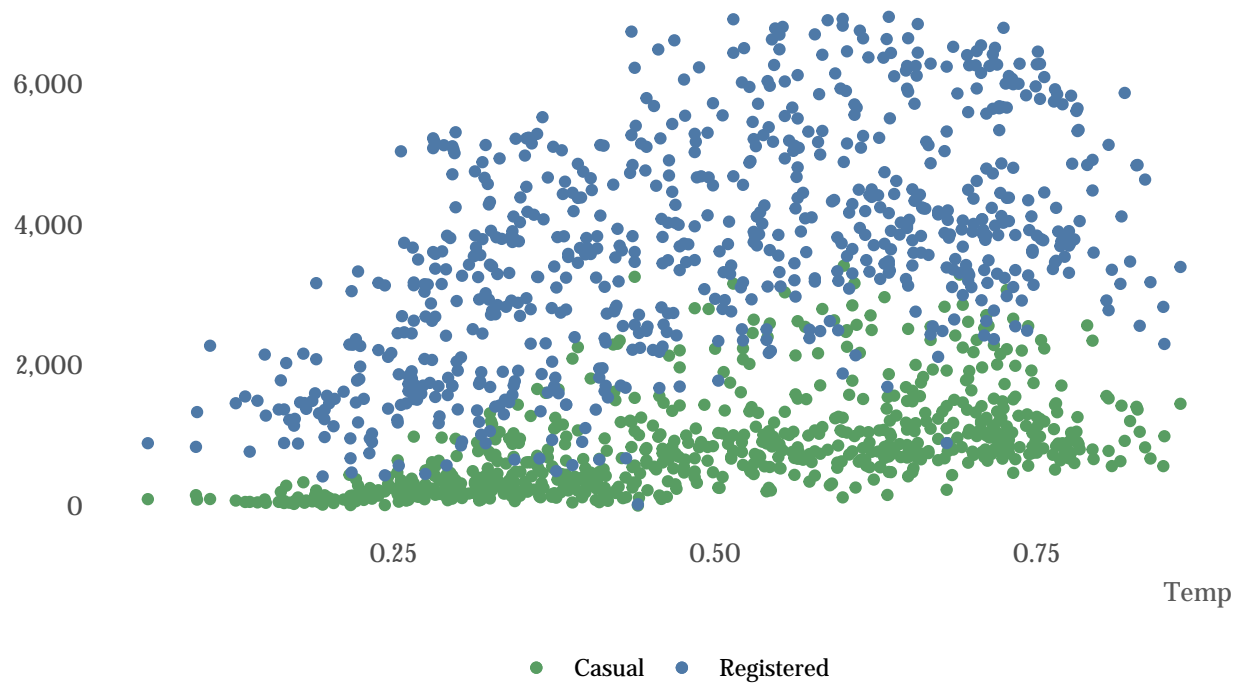
Source: Bikeshare Dataset (2012) | Willy Feid

Scatter Plot

```
ggplot(bikeshare) +
  geom_point(aes(x = temp, y = casual, color = "Casual")) +
  geom_point(aes(x = temp, y = registered, color = "Registered")) +
  labs(title = "Correlation of Temperature and Number of Rentals",
       subtitle = "Temperature has a greater effect on casual users than registered users. Temperature :
       caption = caption,
       x = "Temp",
       y = NULL,
       color = NULL) + # Optional: Remove the legend title
  mychartattributes +
  theme(
    axis.title.x = element_text(vjust = 0, hjust = 1),
    legend.position = "bottom") +
  scale_y_continuous(labels = comma) +
  scale_color_manual(values = c("Casual" = chartcolor_hunter, "Registered" = chartcolor))
```

Correlation of Temperature and Number of Rentals

Temperature has a greater effect on casual users than registered users. Temperature is normalized.



Source: Bikeshare Dataset (2011 – 2012) | Willy Feid

Stacked Area

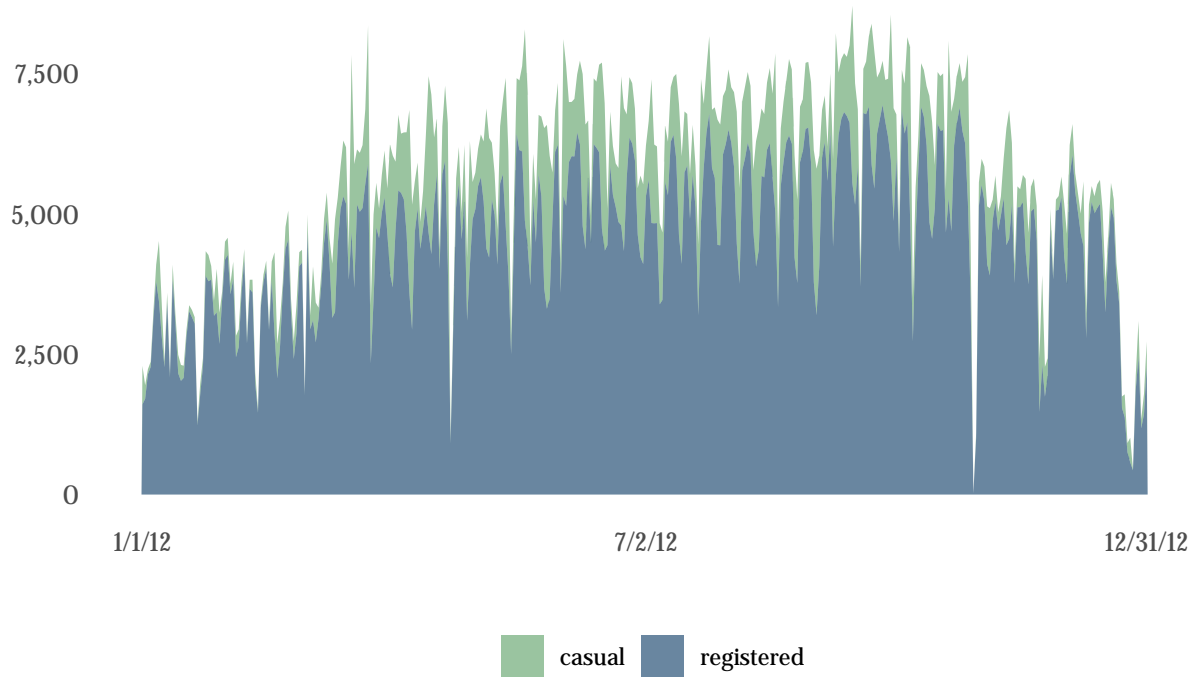
Area graph of daily rentals in 2012 by user type

```
# Reshape the data to long format
bikeshare_2012_long <- bikeshare_2012 %>%
  dplyr::select(date, casual, registered) %>%
  pivot_longer(cols = c(casual, registered), names_to = "type", values_to = "count")

# Plot the stacked area chart
ggplot(bikeshare_2012_long, aes(x = date, y = count, fill = type, group = type)) +
  geom_area(alpha = 0.6, color = NA) +
  labs(title = "Daily Bike Rentals in 2012",
       subtitle = "Registered users make up the majority of rentals, and are responsible for almost all",
       caption = caption_2012,
       x = " ",
       y = " ") +
  mychartattributes +
  scale_x_discrete(breaks = c(first_date, mid_date, last_date), expand = c(0.05, 0.05)) +
  scale_fill_manual(values = c(casual = chartcolor_hunter, registered = chartcolor2)) +
  guides(fill = guide_legend(title = NULL)) +
  theme(legend.position = "bottom") +
  scale_y_continuous(labels = comma)
```

Daily Bike Rentals in 2012

Registered users make up the majority of rentals, and are responsible for almost all bike rentals in the colder



Source: Bikeshare Dataset (2012) | Willy Feid