Histograms

Dr. Niladri Chakraborty

1. Histograms in ggplot2 Packages

Histograms are a useful way to display the distribution of a single continuous variable. In this tutorial, we will use the ggplot2 package in R to create histograms and customize them for effective data visualization.**You can also use the tidyverse package that includes all the tools from ggplot2 package.**

Note: ggplot2 is the package name. ggplot() is the function.

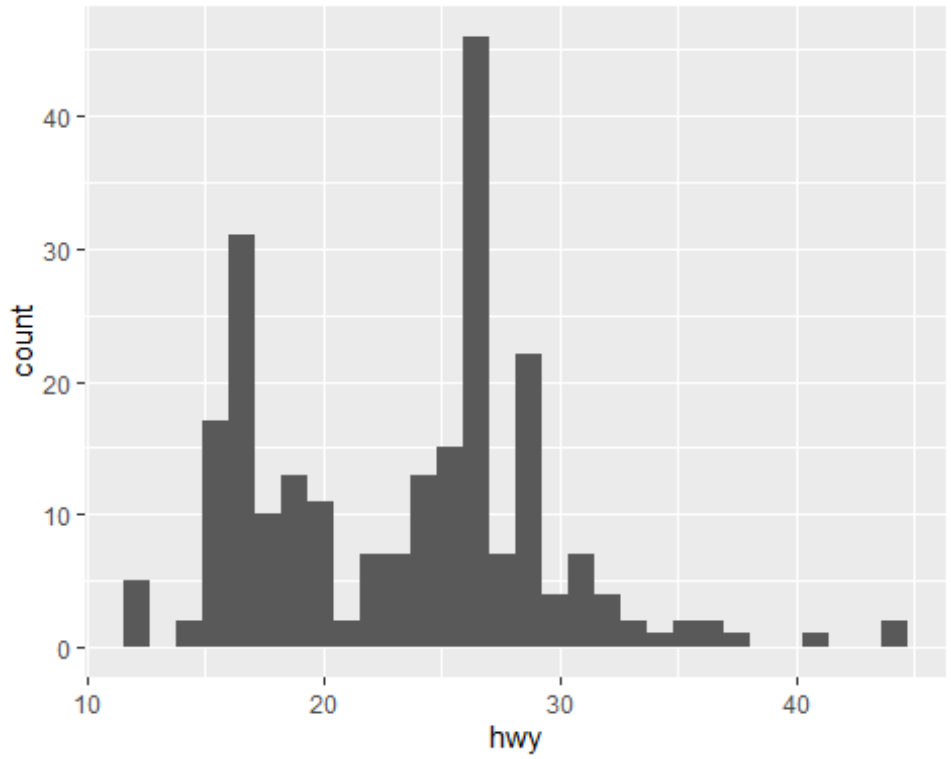
Before we can create histograms in ggplot2, we need to load the package and create a sample dataset. For this tutorial, we will use the **mpg** dataset, which contains information on the fuel economy of cars:

```
# Load ggplot2 package
library(ggplot2)

# Load mpg dataset
data(mpg)
```

2. Creating a Basic Histogram

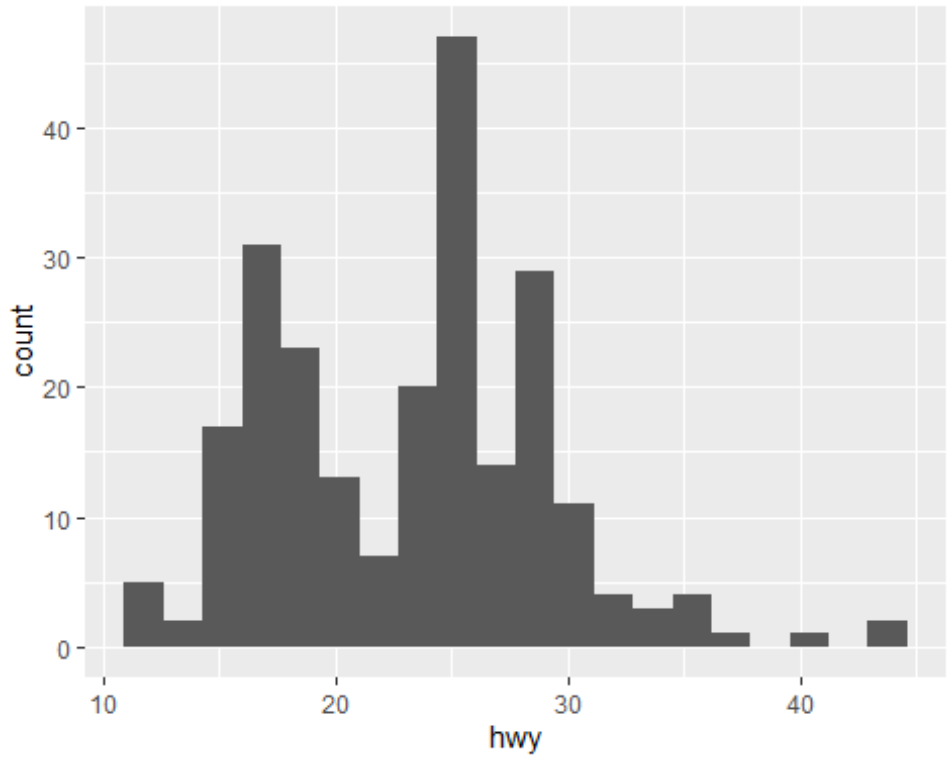
To create a basic histogram in ggplot2, we first need to specify the data and the variable we want to plot. We can then use the geom_histogram() function to create the histogram. For example, to create a histogram of the hwy variable in the mpg dataset, we can use the following code:

```
# Create basic histogram
ggplot(mpg, aes(x = hwy)) +
  geom_histogram()

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



This will create a histogram with default bin widths and colors. We can adjust the appearance of the histogram by adding various arguments to the geom_histogram() function. For example, we can specify the number of bins using the bins argument:

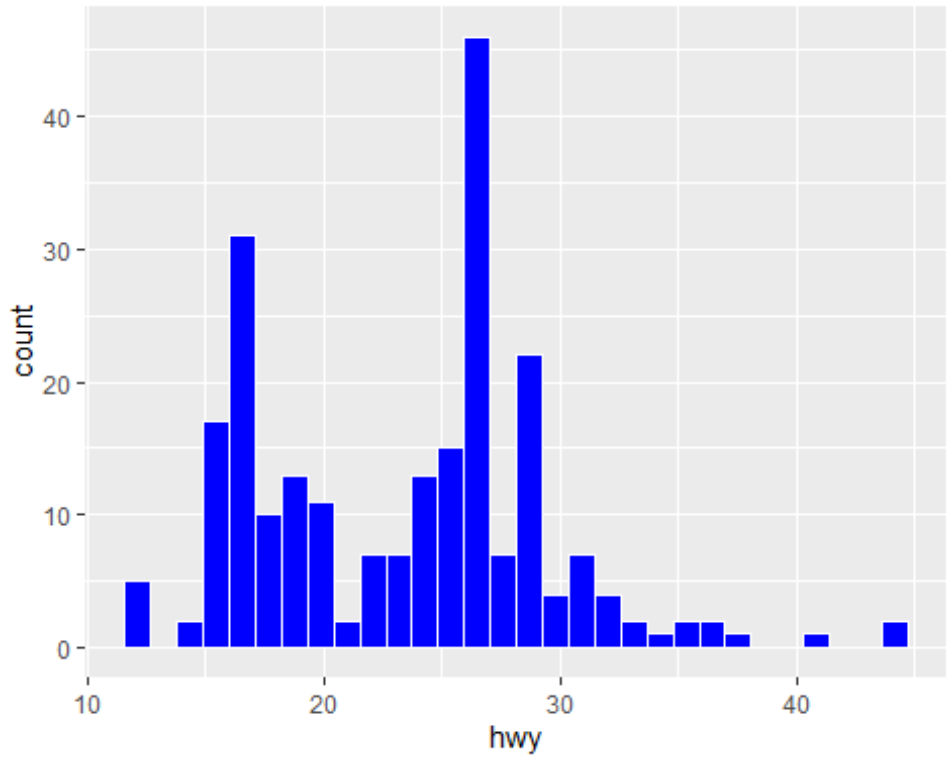
```
# Create histogram with 20 bins
ggplot(mpg, aes(x = hwy)) +
  geom_histogram(bins = 20)
```



3. Customizing Histogram Appearance

To customize the appearance of the histogram, we can add various arguments to the ggplot() function and the geom_histogram() function. For example, we can change the color and fill of the bars using the fill and color arguments:

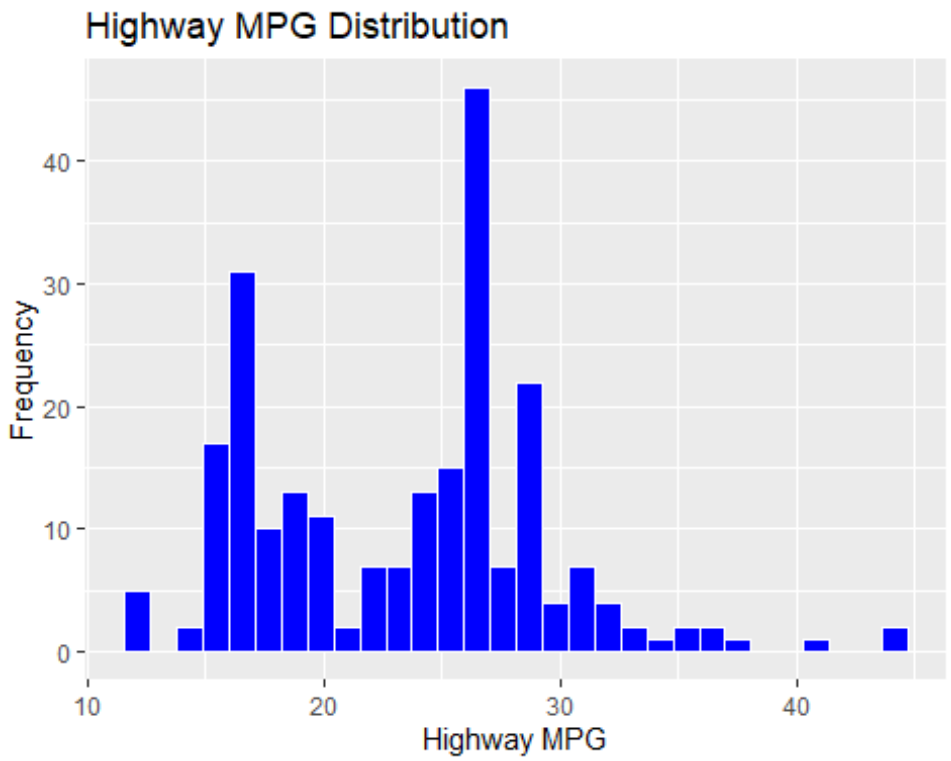
```
# Create histogram with customized colors
ggplot(mpg, aes(x = hwy)) +
  geom_histogram(fill = "blue", color = "white")

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



This will create a histogram with blue bars and white borders.

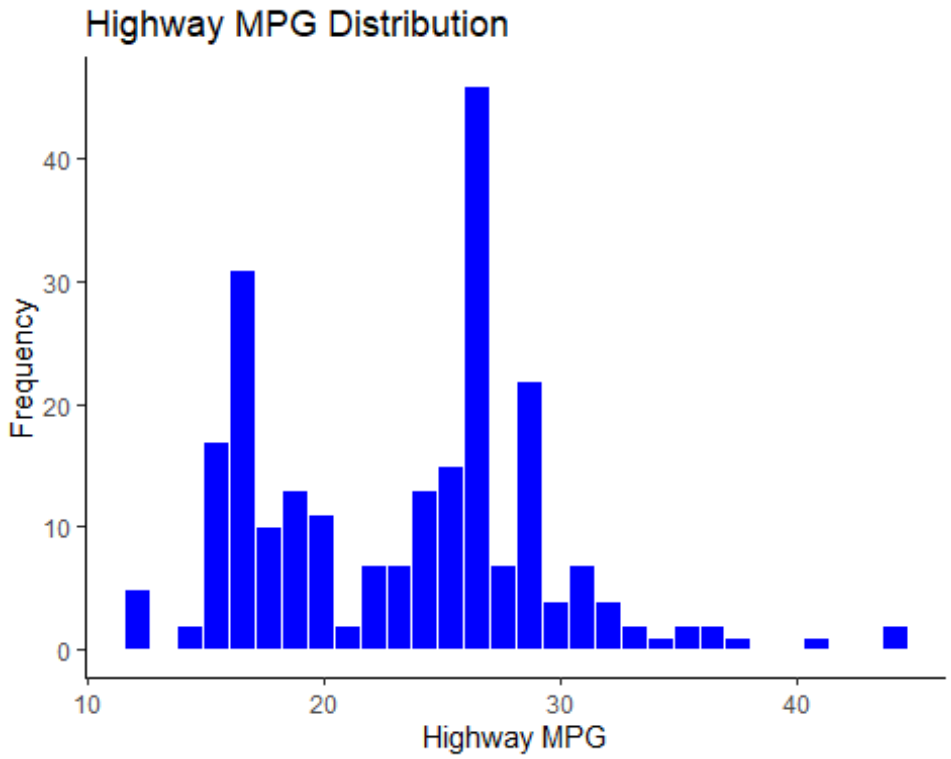
We can also add a title and axis labels using the ggtitle() and labs() functions, respectively:

```
# Create histogram with title and axis labels
ggplot(mpg, aes(x = hwy)) +
  geom_histogram(fill = "blue", color = "white") +
  ggtitle("Highway MPG Distribution") +
  labs(x = "Highway MPG", y = "Frequency")

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



You can change the background by using the 'hrbrthemes' package.

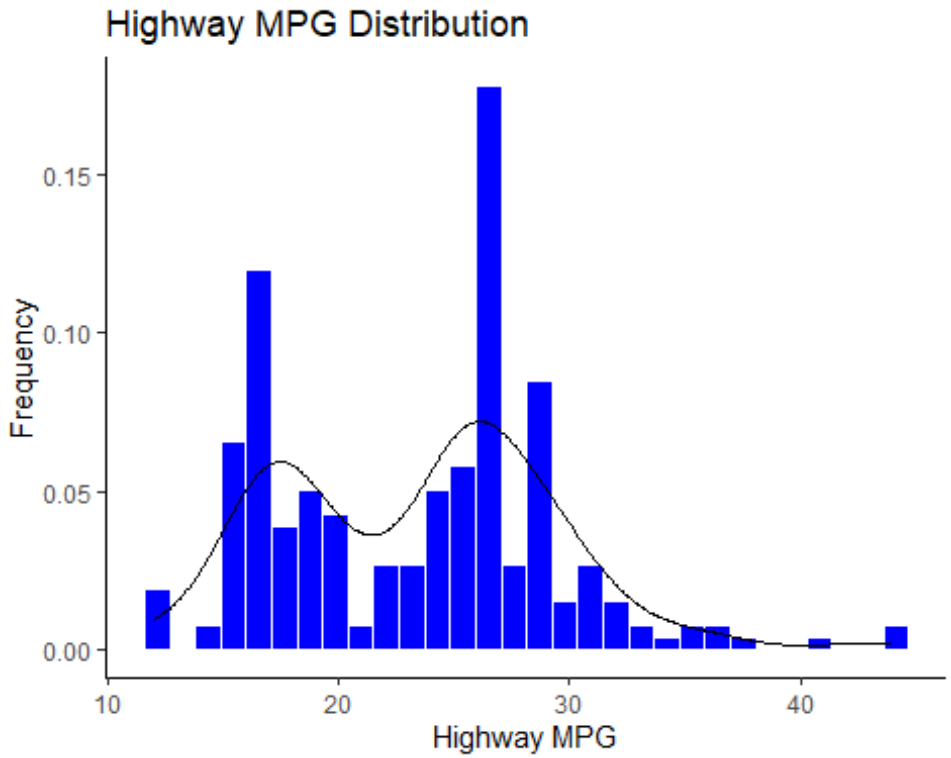
```
# Create histogram with title and axis labels
ggplot(mpg, aes(x = hwy)) +
  geom_histogram(fill = "blue", color = "white") +
  ggtitle("Highway MPG Distribution") +
  labs(x = "Highway MPG", y = "Frequency")+
  theme_classic()

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Adding a density curve to the histogram

```
ggplot(mpg, aes(x = hwy)) +
  geom_histogram(aes(y = ..density..), fill = "blue", color = "white") +
  geom_density()+
  ggtitle("Highway MPG Distribution") +
  labs(x = "Highway MPG", y = "Frequency")+
  theme_classic()
```

```
## Warning: The dot-dot notation (`..density..`) was deprecated in ggplot2
3.4.0.
## ℹ Please use `after_stat(density)` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```


Highway MPG Distribution
Frequency
0.15
0.10
0.05
0.00
10
20
30
40
Highway MPG