Continuous x Continuous data (2 of 2)

Sep 17, 2023.

Exploring bivariate Continuous x Continuous data, using `ggplot2`

This chapter demonstrates the use of the popular **ggplot2** and **ggpubr** packages to further explore the interaction between *bivariate continuous data.*

Data: Suppose we run the following code to prepare the `mtcars` data for subsequent analysis and save it in a tibble called `tb`.

```
# Load the required libraries, suppressing annoying startup messages
library(dplyr, quietly = TRUE, warn.conflicts = FALSE)
library(tibble, quietly = TRUE, warn.conflicts = FALSE)
library(ggplot2, quietly = TRUE, warn.conflicts = FALSE) # For data visualization
library(ggpubr, quietly = TRUE, warn.conflicts = FALSE) # For data visualization
# Read the mtcars dataset into a tibble called tb
data(mtcars)
tb <- as_tibble(mtcars)
# Convert relevant columns into factor variables
tb$cyl <- as.factor(tb$cyl) # cyl = {4,6,8}, number of cylinders
tb$am <- as.factor(tb$am) # am = {0,1}, 0:automatic, 1: manual transmission
tb$vs <- as.factor(tb$vs) # vs = {0,1}, v-shaped engine, 0:no, 1:yes
tb$gear <- as.factor(tb$gear) # gear = {3,4,5}, number of gears
# Directly access the data columns of tb, without tb$mpg
attach(tb)
```

```
The following object is masked from package:ggplot2:

    mpg
```

Scatterplots using ggplot2

```
# Create a scatter plot
ggplot(tb,
       aes(x = hp, y = mpg)) +
  geom_point() +
  xlab("Horsepower of Engline") +
  ylab("Miles per gallon") +
  ggtitle("Scatter Plot of mpg vs. hp")
```
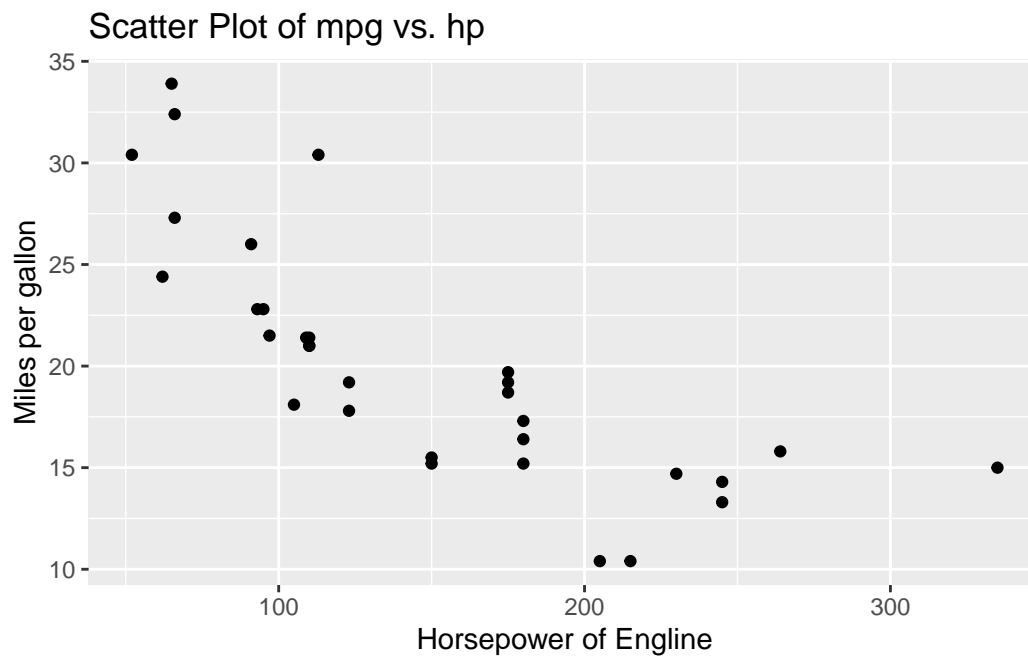

This code creates a scatter plot of the wt variable (weight in 1000 lbs) on the x-axis and the mpg variable (miles per gallon) on the y-axis. The geom_point() function is used to add the points to the plot, and xlab(), ylab(), and ggtitle() are used to add axis labels and a plot title, respectively. You can adjust the aesthetics of the plot, such as the color and size of the points, by adding additional arguments to the geom_point() function.
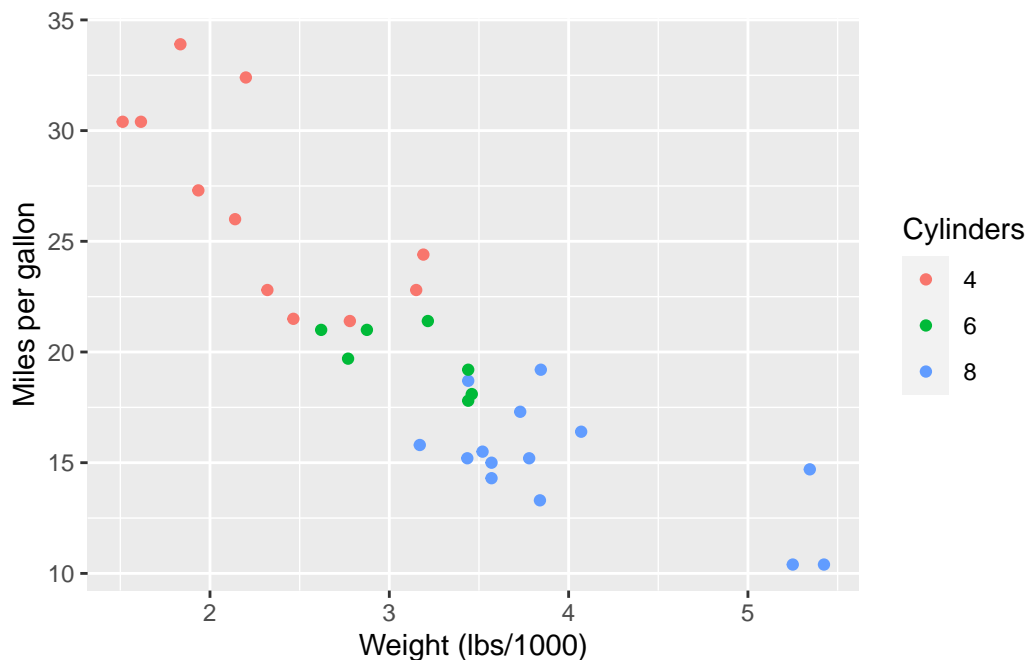
Scatterplots broken down by Categorical Variables

Scatterplot with colored by Categorical Variable Using ggplot()

This will create a scatterplot of miles per gallon (mpg) against weight, with each point colored according to the number of cylinders in the engine (cyl).

```r
# Load the ggplot2 package
library(ggplot2)

# Create a scatterplot of mpg vs. wt, colored by cyl
ggplot(tb, aes(x = wt, y = mpg, color = factor(cyl))) +
  geom_point() +
  labs(x = "Weight (lbs/1000)", y = "Miles per gallon") +
  scale_color_discrete(name = "Cylinders")
```


Scatterplot with broken down by Categorical Variable Using ggplot()

This will create a scatterplot of miles per gallon (mpg) against weight, with each plot faceted by the number of cylinders in the engine (cyl).

```
# Load the ggplot2 package
library(ggplot2)

# Create a scatterplot matrix using ggplot()
ggplot(tb, aes(x = mpg, y = disp)) +
  geom_point() +
  facet_grid(. ~ cyl)
```