Cars_Data_Visualisation_R

Dara Abb

2025-04-10

```
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✔ dplyr     1.1.4     ✔ readr     2.1.5
## ✔ forcats   1.0.0     ✔ stringr   1.5.1
## ✔ ggplot2   3.5.1     ✔ tibble    3.2.1
## ✔ lubridate 1.9.3     ✔ tidyr     1.3.1
## ✔ purrr     1.0.2
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(ggplot2)
```

Import datasets

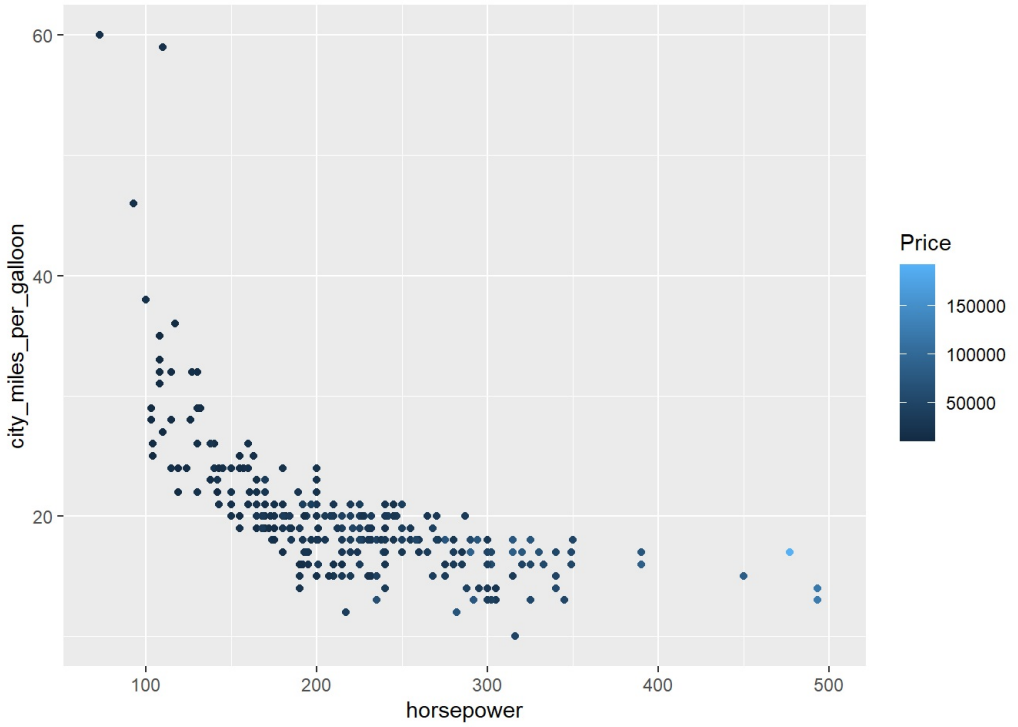
```
cars <- read.csv("cars.csv")

retail_clean <- read.csv("retail_clean.csv")
```

Scatterplots:are used to examine the relationship between two continuous variables

1st Scatterplot: examines the relation between horsepower an city miler per gallon

```
cars %>% ggplot(aes(x= horsepower, y=city_miles_per_galloon, color=Price))+geom_point()
```

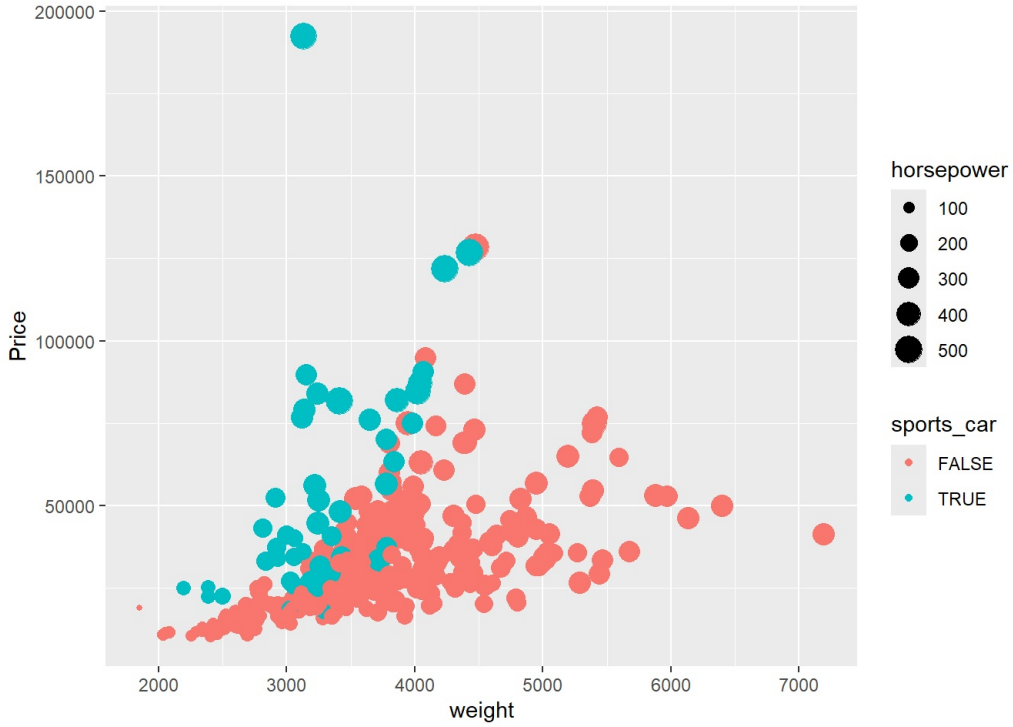
```
## Warning: Removed 14 rows containing missing values or values outside the scale range
## (`geom_point()`).
```



2nd Scatterplot: Weight vs. Price with Size and Color Modifications

```
cars %>% ggplot(aes(x=weight, y= Price, color=sports_car, size=horsepower))+geom_point()
```

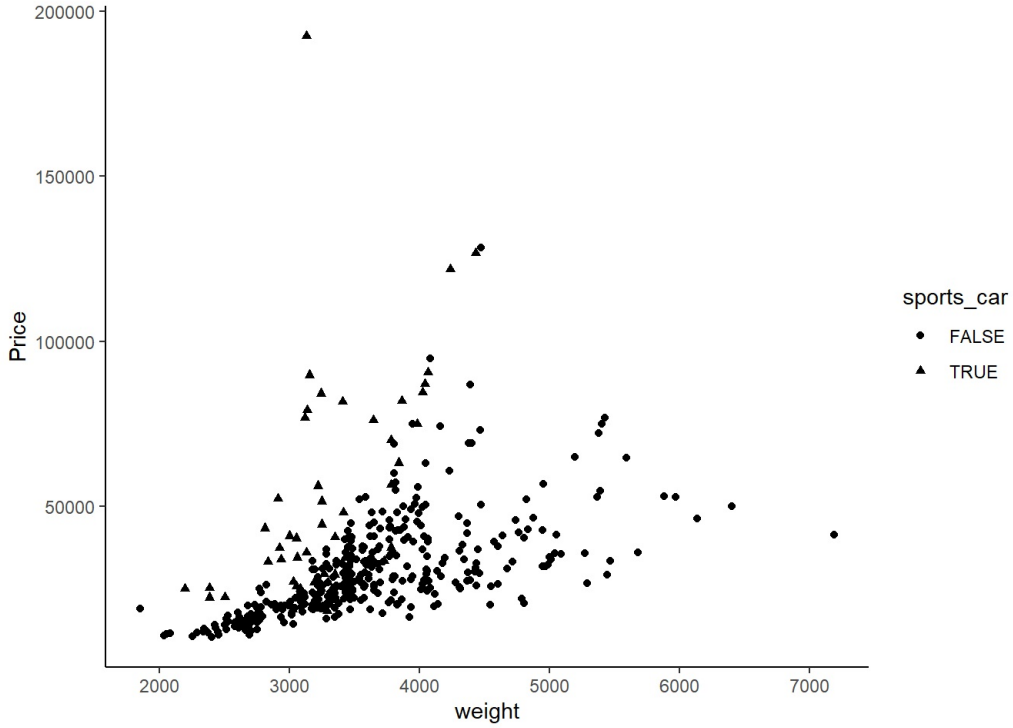
```
## Warning: Removed 2 rows containing missing values or values outside the scale range
## (`geom_point()`).
```



3rd Scatterplot: Scatterplot with Different Shape

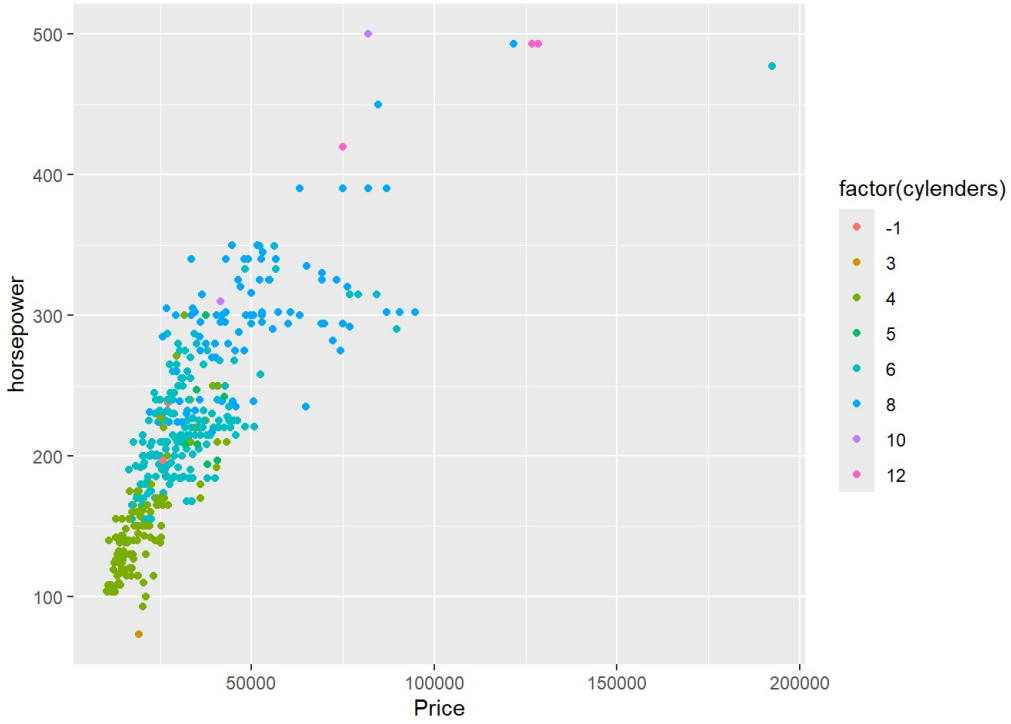
```
cars %>% ggplot(aes(x=weight, y= Price, shape=sports_car))+geom_point()+theme_classic()
```

```
## Warning: Removed 2 rows containing missing values or values outside the scale range
## (`geom_point()`).
```



4th Scatterplot: Price vs. Horsepower with Cylinders Factor

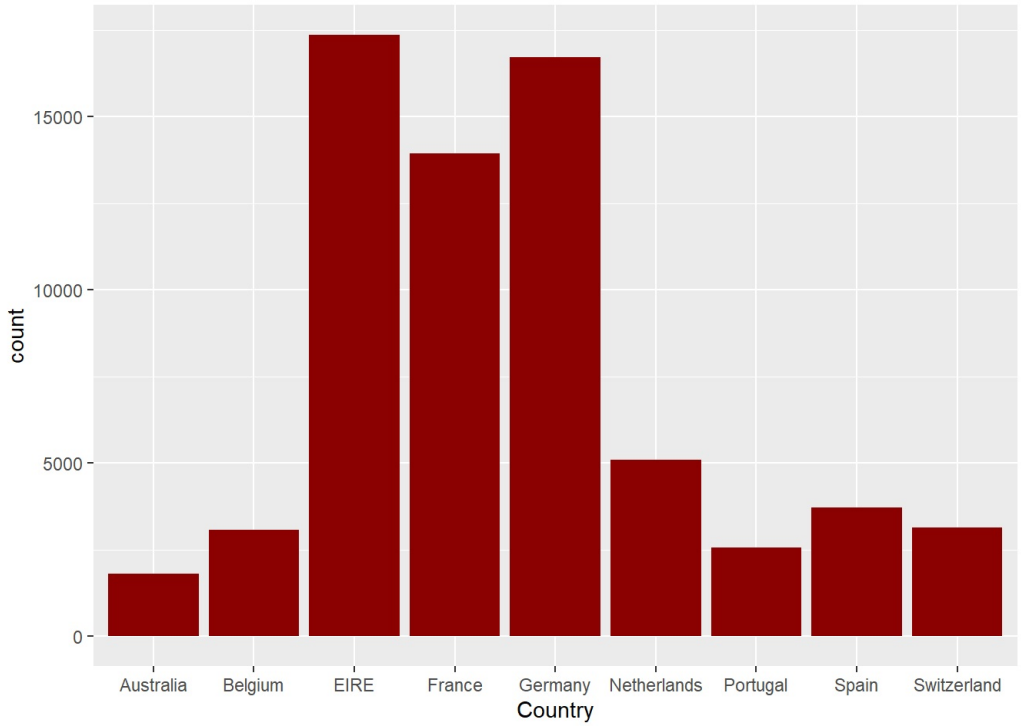
```
cars %>% ggplot(aes(x=Price, y= horsepower, color=factor(cylenders)))+geom_point()
```



Bar plots help to visualize categorical data by representing the frequency of each category.

Top 9 Countries with Highest Appearances

```
top9_international <- retail_clean %>% group_by(Country) %>% summarise(appearnece = n()) %>% arrange(desc(appearn
ece))%>%
  slice(2:10)

top9 <- as.vector(top9_international[["Country"]])

retail_clean %>% filter(Country %in% top9) %>% ggplot(aes(x= Country))+geom_bar(fill= "darkred")
```



Weekday Distribution Plot

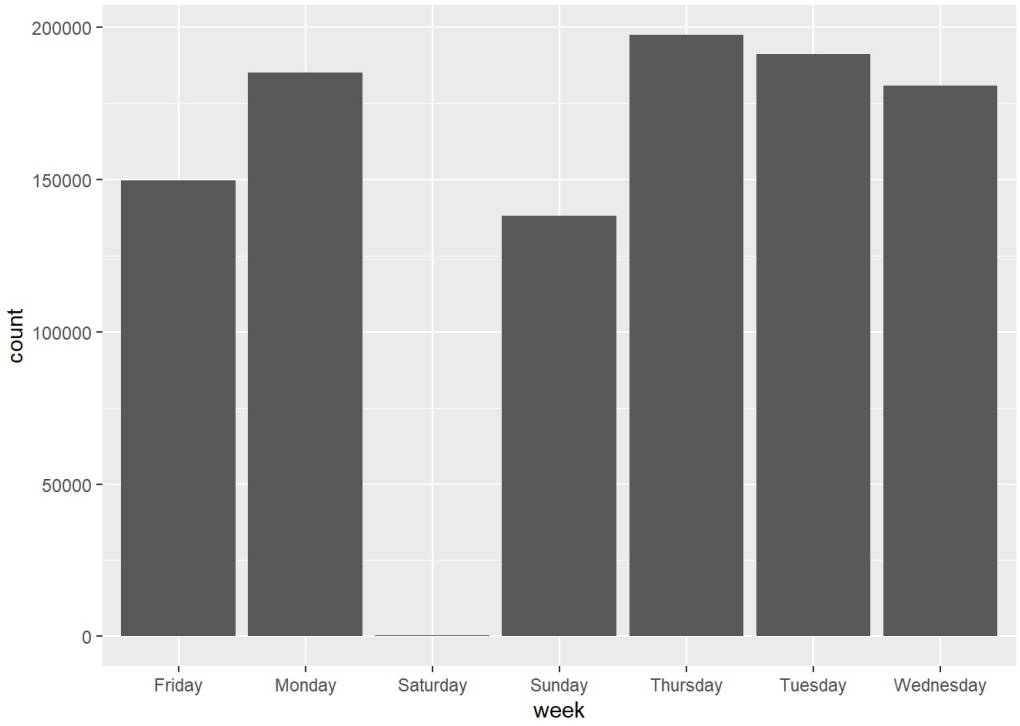
```
retail_clean<- retail_clean %>% separate(col = InvoiceDate,
                                         into = c("date","time"),
                                         sep = " ")
```

```
## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 1042721 rows [1, 2, 3, 4,
## 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].
```

```
head(retail_clean)
```

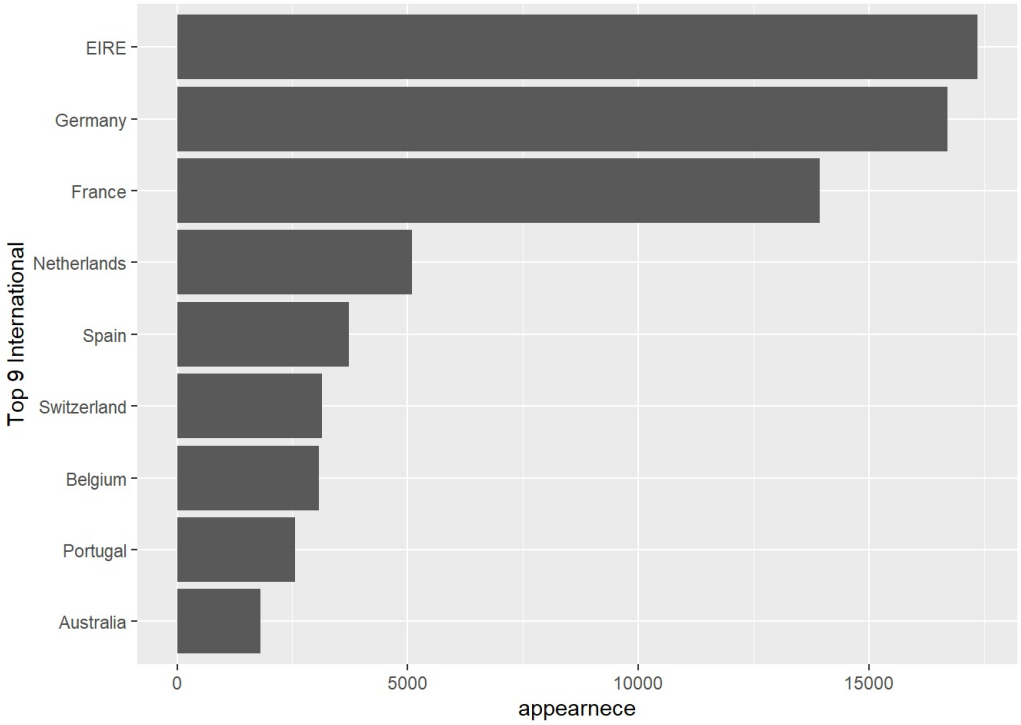
```
##   Invoice StockCode                         Description Quantity
## 1  489434     85048 15CM CHRISTMAS GLASS BALL 20 LIGHTS       12
## 2  489434    79323P                  PINK CHERRY LIGHTS       12
## 3  489434    79323W                 WHITE CHERRY LIGHTS       12
## 4  489434     22041         RECORD FRAME 7" SINGLE SIZE       48
## 5  489434     21232       STRAWBERRY CERAMIC TRINKET BOX       24
## 6  489434     22064            PINK DOUGHNUT TRINKET POT       24
##                   date time Price Customer.ID        Country
## 1 2009-12-01T07:45:00Z <NA>  6.95       13085 United Kingdom
## 2 2009-12-01T07:45:00Z <NA>  6.75       13085 United Kingdom
## 3 2009-12-01T07:45:00Z <NA>  6.75       13085 United Kingdom
## 4 2009-12-01T07:45:00Z <NA>  2.10       13085 United Kingdom
## 5 2009-12-01T07:45:00Z <NA>  1.25       13085 United Kingdom
## 6 2009-12-01T07:45:00Z <NA>  1.65       13085 United Kingdom
```

```
#####weekday analysis
retail_clean$date <- as.Date(retail_clean$date)

retail_clean$week <- weekdays(retail_clean$date)

retail_clean %>% ggplot(aes(x=week))+geom_bar()
```



Bar Plot for Top 9 Countries Reordered

```
top9_international %>%
  ggplot(aes(x=reorder(Country, appearnece), y=appearnece))+geom_col()+coord_flip()+xlab("Top 9 International")
```

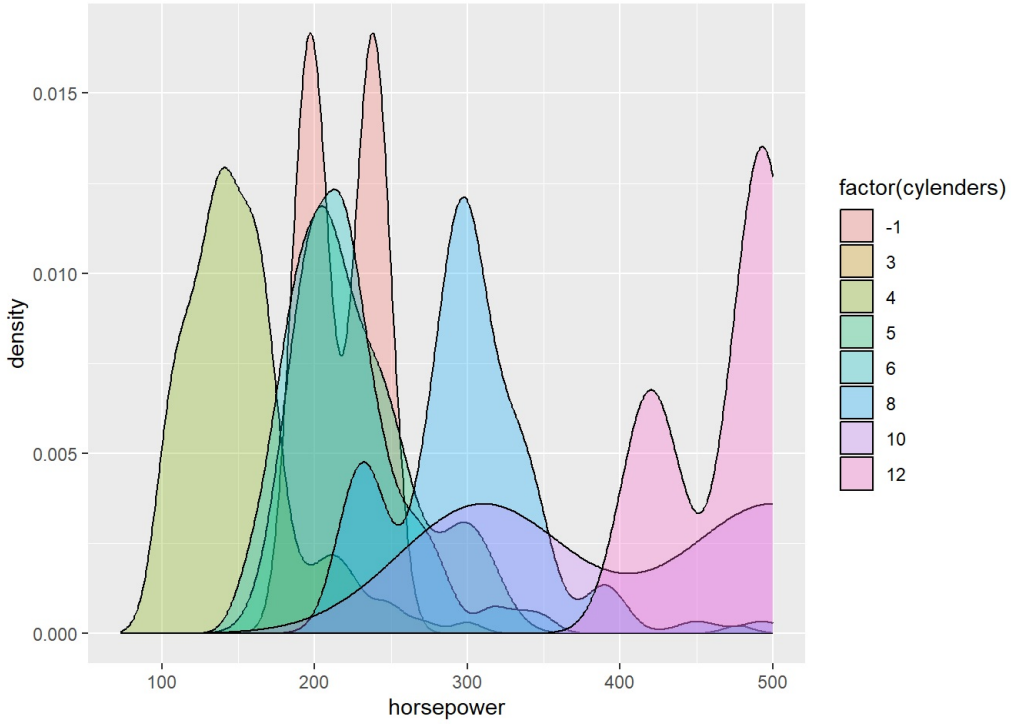


Density Plot: Horsepower by Cylinders

```
cars %>% ggplot(aes(x=horsepower, fill = factor(cylenders)))+ geom_density(alpha=0.3)
```

```
## Warning: Groups with fewer than two data points have been dropped.
```

```
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf
```

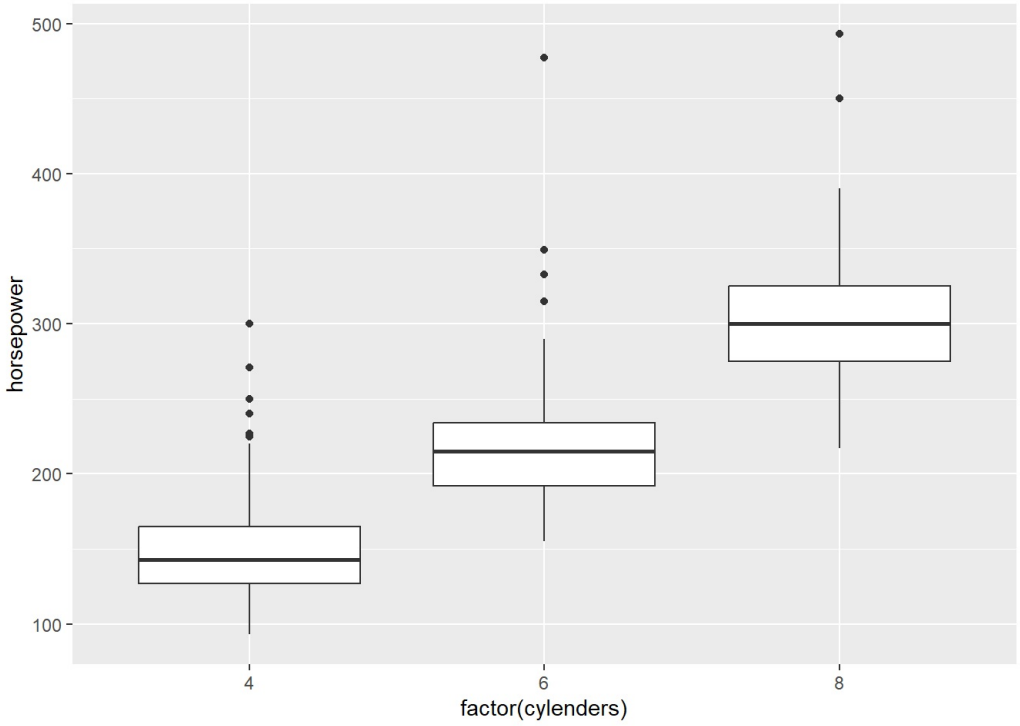


Box plots: Box plots are used to visualize the distribution of a continuous variable and identify outliers.

```
##Box Plot: Horsepower by Cylinders
table(cars$cylenders)
```

```
##
##  -1   3   4   5   6   8  10  12
##   2   1 136   7 190  87   2   3
```

```
cars %>% filter(cylenders %in% c(4,6,8)) %>% ggplot(aes(x=factor(cylenders), y=horsepower))+geom_boxplot()
```


500
400
300
200
100
horsepower
4
6
8
factor(cylenders)