

Data Visualizations

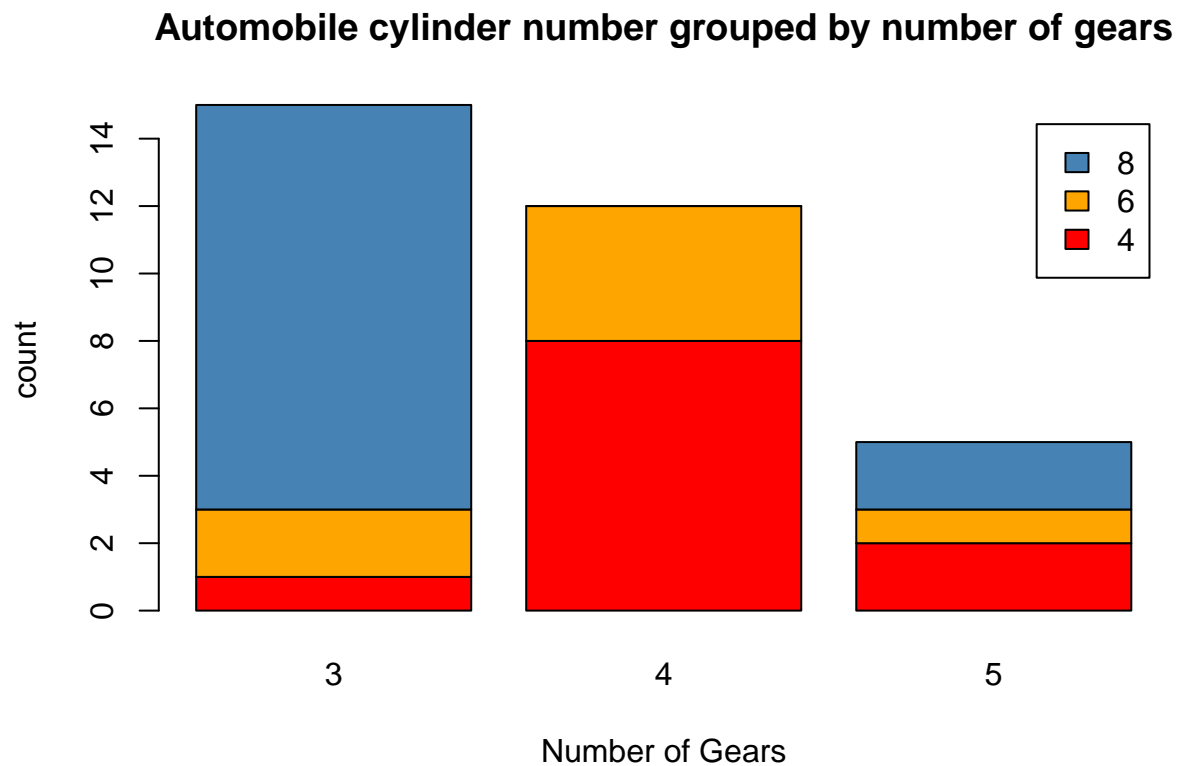
Chapter 1 : Bar Chart

You're probably already familiar with the basic bar chart from elementary school, high school and college. The concept of the bar chart in R is the same as it was in the past scenarios — to show a categorical comparison between two or more variables. However, there are several different types of bar charts to know and understand.

Horizontal and vertical bar charts are already common and familiar — they are standard formats in most academic or professional presentations. But R provides a stacked bar chart that lets you introduce different variables to each category.

```
library(datasets)
```

```
Numbers<-table(mtcars$cyl,mtcars$gear)
barplot(Numbers,main='Automobile cylinder number grouped by number of gears',
        col=c('red','orange', 'steelblue'),legend=rownames(Numbers),xlab='Number of Gears',
        ylab='count')
```

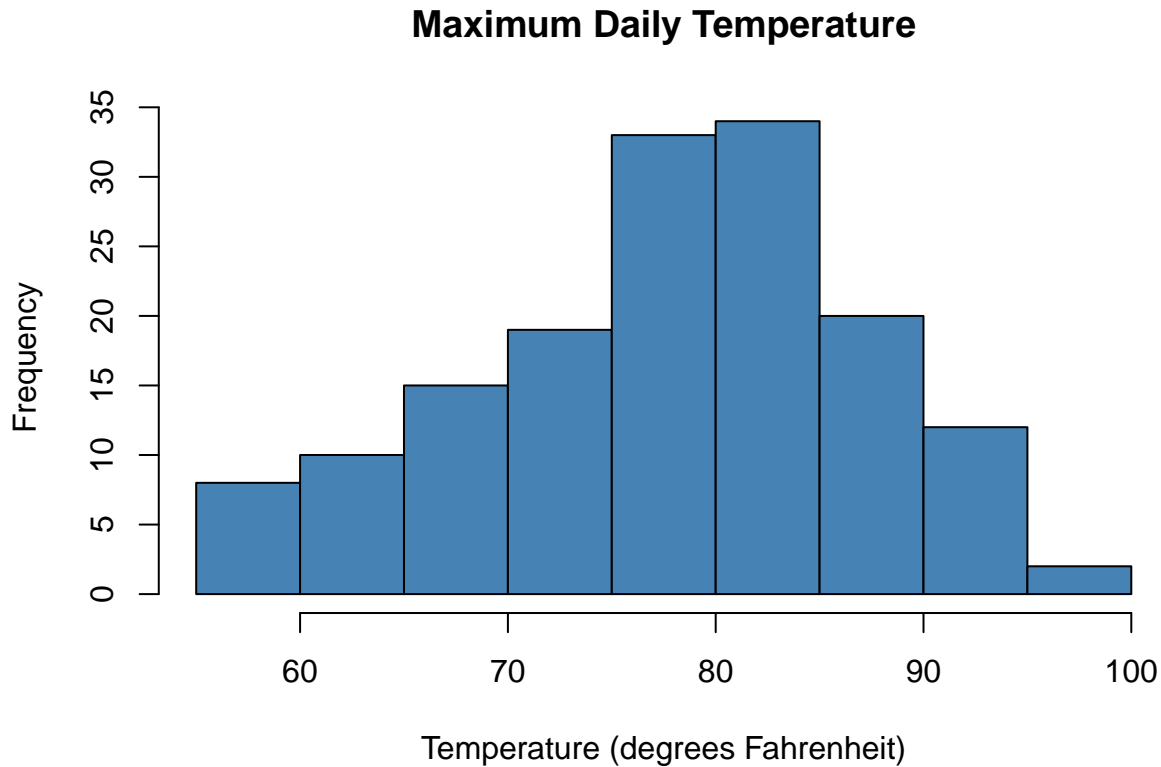


Chapter 2 : Histogram

Histograms are standard in some academic fields, but they're usually reserved for the senior-most levels. These charts are best with highly precise or accurate numbers in R.

It ultimately provides a probability estimate of a variable — the period of time before a project's completion, for example. R provides a simple function for this as well.

```
# histogram of frequency of ozone values in 'airquality' dataset  
hist(airquality$Temp,col='steelblue',main='Maximum Daily Temperature',  
      xlab='Temperature (degrees Fahrenheit)')
```

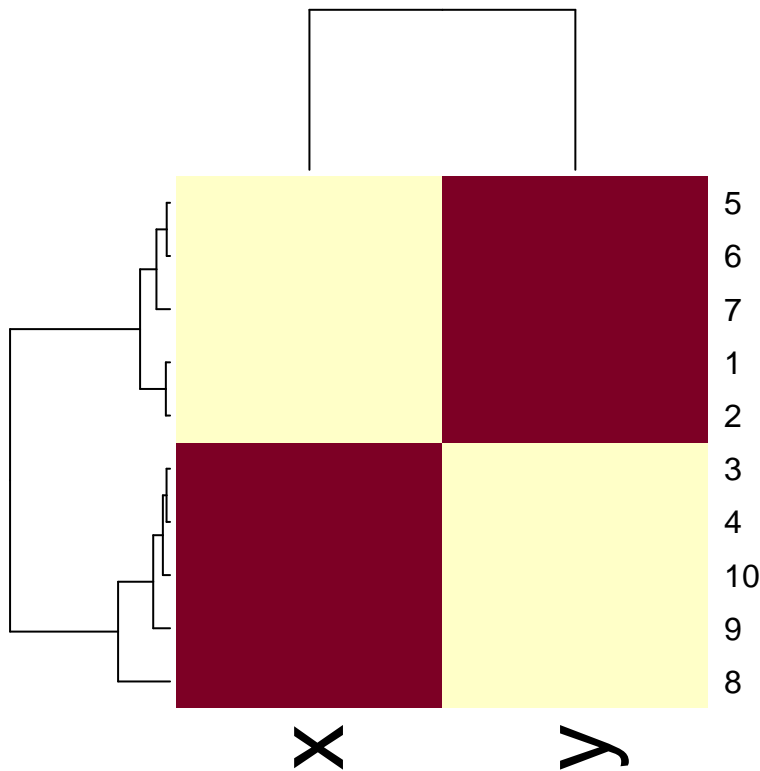


Chapter 3 : Heat Map

One of the most innovative data visualizations in R, the heat map emphasizes color intensity to visualize relationships between multiple variables.

The result is an attractive 2D image that is easy to interpret. As a basic example, a heat map highlights the popularity of competing items by ranking them according to their original market launch date. It breaks it down further by providing sales statistics and figures over the course of time.

```
# simulate a dataset of 10 points  
x<-rnorm(10,mean=rep(1:5,each=2),sd=0.7)  
y<-rnorm(10,mean=rep(c(1,9),each=5),sd=0.1)  
dataFrame<-data.frame(x=x,y=y)  
set.seed(143)  
dataMatrix<-as.matrix(dataFrame)[sample(1:10),] # convert to class 'matrix', then shuffle the rows of t  
heatmap(dataMatrix) # visualize hierarchical clustering via a heatmap
```

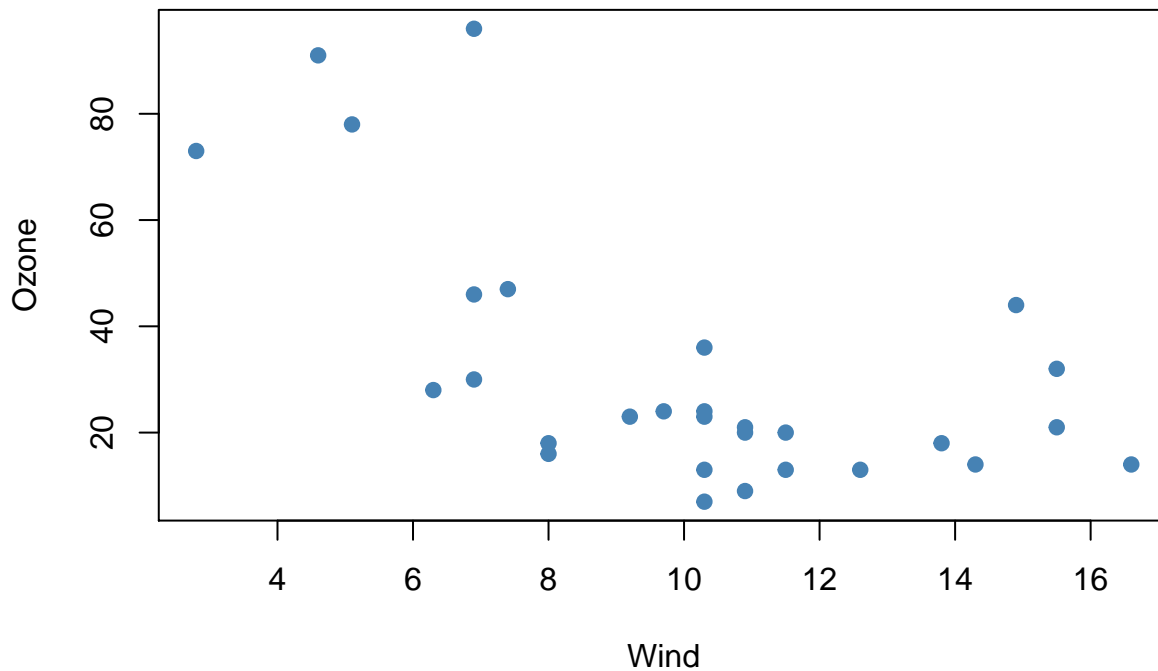


Chapter 4 : Scatter Plot

Plotting is a popular alternative to charting or graphing. It provides a unique visualization involving various dots. The most standard iteration — the scatter plot — tracks two continuous variables over the course of time. A basic application of the scatter plot involves tracking the height and weight of children throughout the years.

Scatter plots are useful when trying to avoid misinformation in the visualization. Only use a plot if you're sure the audience is familiar with that type of chart, and always use it sparingly. When in doubt, go with one of your other options.

```
# Plot Ozone and Temperature measurements for only the month of September  
with(subset(airquality,Month==9),plot(Wind,Ozone,col='steelblue',pch=20,cex=1.5))
```



```
#title('Wind and Temperature in NYC in September of 1973')
```

Chapter 5 : Box Plot

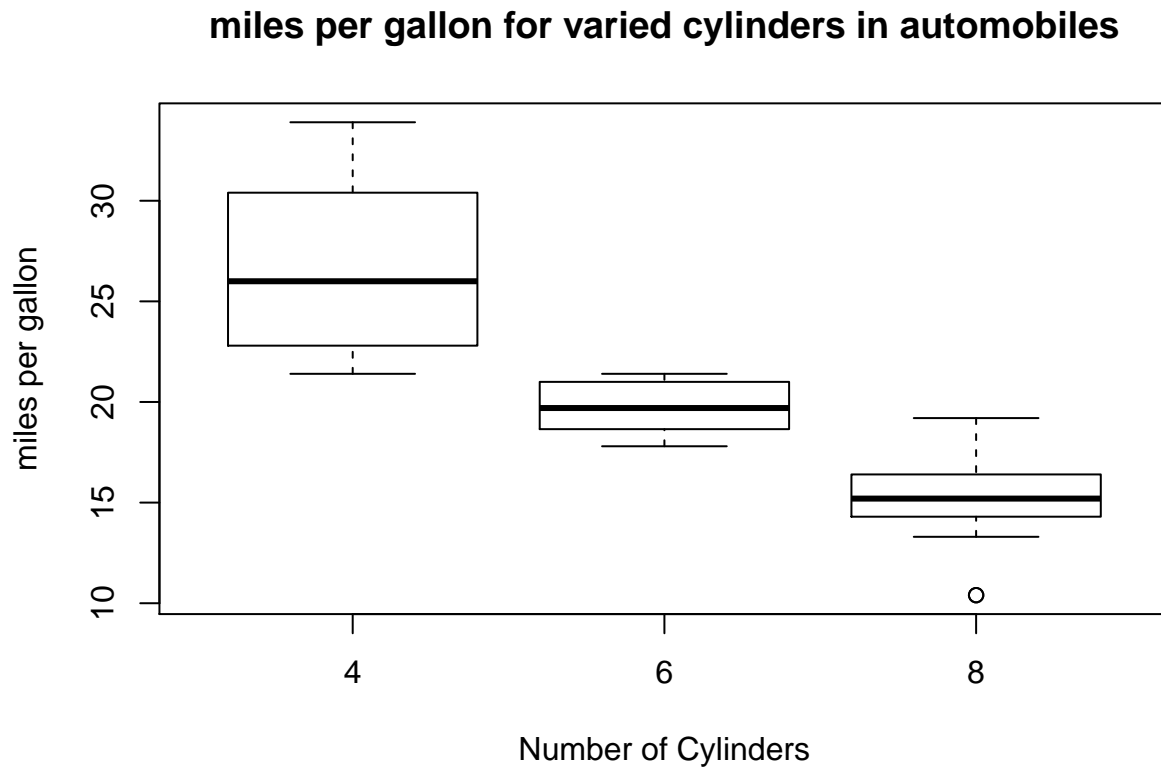
The box plot resembles a bar chart in many respects. Instead of focusing on categorical data, box plots provide visualization for both categorical and continuous variable data.

In the real world, box plots give detailed information on weather patterns and how they change over the course of time.

```
mtcars<-transform(mtcars,cyl=factor(cyl)) # convert 'cyl' column from class 'numeric' to class 'factor'  
class(mtcars$cyl) # 'cyl' is now a categorical variable
```

```
## [1] "factor"
```

```
boxplot(mpg~cyl,mtcars,xlab='Number of Cylinders',ylab='miles per gallon',  
        main='miles per gallon for varied cylinders in automobiles',cex.main=1.2)
```



Chapter 6 : Correlogram

Correlated data is best visualized through corplot. The 2D format is similar to a heat map, but it highlights statistics that are directly related.

Most correlograms highlight the amount of correlation between datasets at various points in time. Comparing sales data between different months or years is a basic example.

```
library(dplyr) #data manipulation

##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(ggplot2) #data visualization

## Registered S3 methods overwritten by 'ggplot2':
##   method      from
##   [.quosures   rlang
##   c.quosures   rlang
##   print.quosures rlang

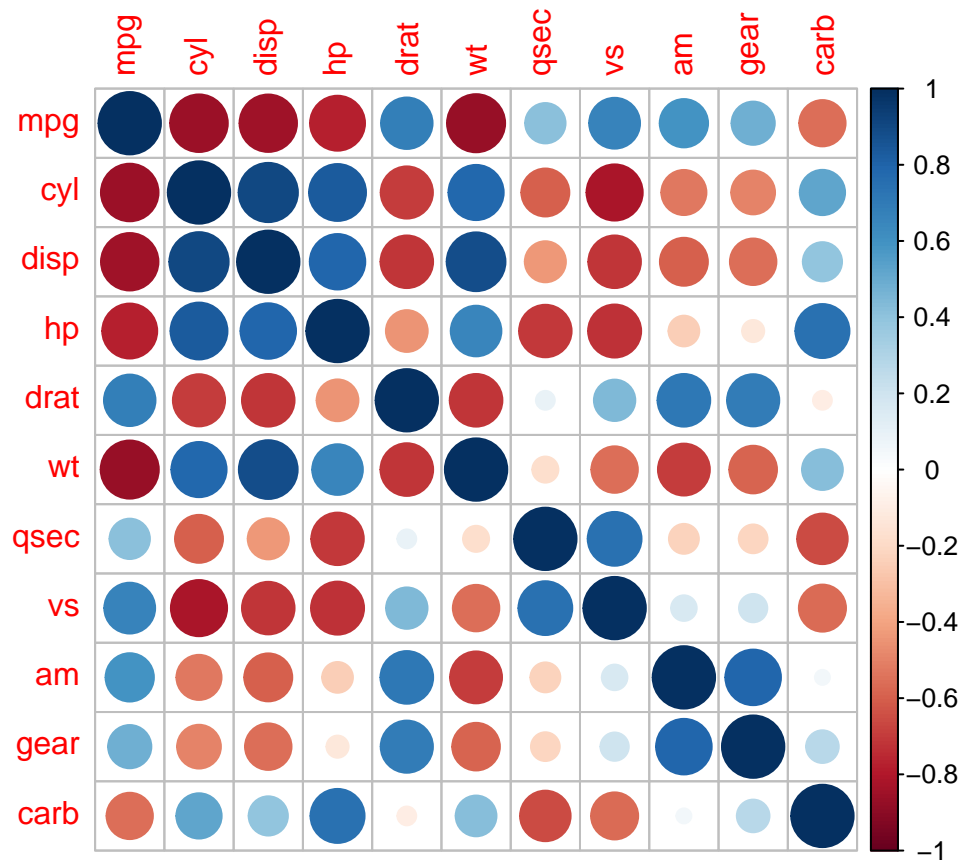
library(corrplot) #correlogram

## corrplot 0.84 loaded

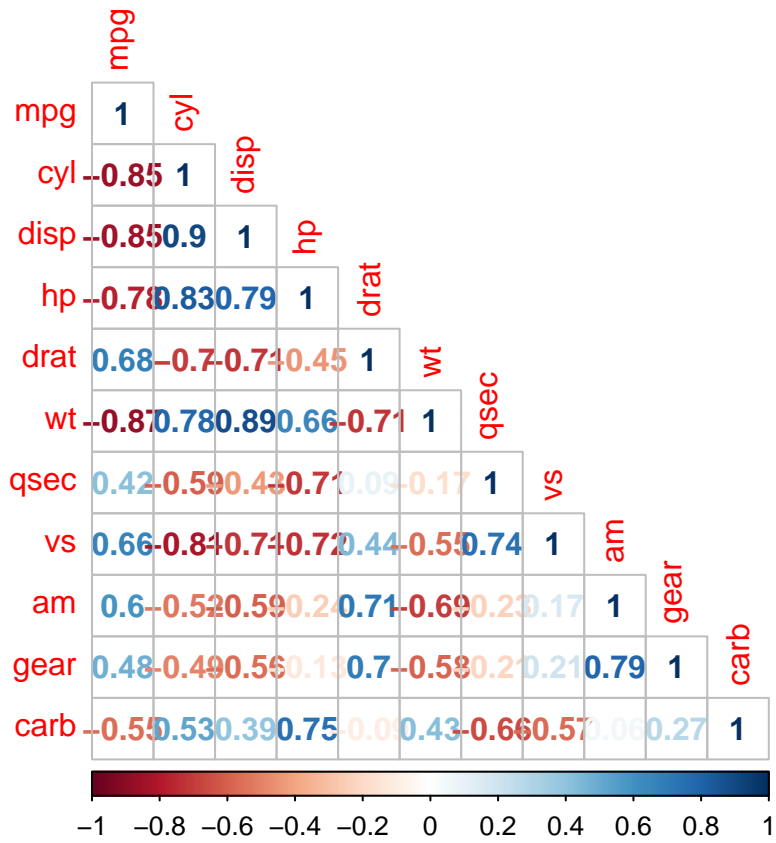
data("mtcars")

corr_matrix <- cor(mtcars)

# with circles
corrplot(corr_matrix)
```



```
# with numbers and lower
corrplot(corr_matrix,
         method = 'number',
         type = "lower")
```



Chapter 7 : Area Chart

Area charts express continuity between different variables or data sets. It's akin to the traditional line chart you know from grade school and is used in a similar fashion.

Most area charts highlight trends and their evolution over the course of time, making them highly effective when trying to expose underlying trends — whether they're positive or negative.

```
data("airquality") #dataset used

airquality %>%
  group_by(Day) %>%
  summarise(mean_wind = mean(Wind)) %>%
  ggplot() +
  geom_area(aes(x = Day, y = mean_wind)) +
  labs(title = "Area Chart of Average Wind per Day",
        subtitle = "using airquality data",
        y = "Mean Wind")
```

Area Chart of Average Wind per Day
using airquality data

