null

null

null

Data Visualization with `ggplot2`

Load `ggplot2` package, supposing it is already installed.

```
require(ggplot2)
```

Data

iris

Almost all the following exercises are based on the `iris` data, taken from the `datasets` package. It is a base package so it is already installed and loaded.

```
data("iris")
```

This dataset gives the measurements in centimeters of length and width of sepal and petal, respectively, for 50 flowers from each of 3 species of iris. The species are Iris setosa, versicolor, and virginica.

`iris` dataset contains the following variables:

- `Sepal.Length`: length of iris sepal
- `Sepal.Width`: width of iris sepal
- `Petal.Length`: length of iris petal
- `Petal.Width`: width of iris petal
- `Species`: species of iris

```
dim(iris)
```

```
## [1] 150   5
```

```
head(iris)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

```
str(iris)
```

```
## 'data.frame':    150 obs. of  5 variables:
##  $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
##  $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
##  $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
##  $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
##  $ Species     : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...
```

mpg

Some of the exercises are based on `mpg` dataset, taken from the `datasets` package.
It is a base package so it is already installed and loaded.

```
data("mpg")
```

This dataset contains the fuel economy data from 1999 and 2008 for 38 popular models of car.

```
dim(mpg)
```

```
## [1] 234  11
```

```
head(mpg)
```

```
##   manufacturer model displ year cyl      trans drv cty hwy fl   class
## 1         audi    a4   1.8 1999   4   auto(l5)   f  18  29  p compact
## 2         audi    a4   1.8 1999   4 manual(m5)   f  21  29  p compact
## 3         audi    a4   2.0 2008   4 manual(m6)   f  20  31  p compact
## 4         audi    a4   2.0 2008   4   auto(av)   f  21  30  p compact
## 5         audi    a4   2.8 1999   6   auto(l5)   f  16  26  p compact
## 6         audi    a4   2.8 1999   6 manual(m5)   f  18  26  p compact
```

```
str(mpg)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    234 obs. of  11 variables:
##  $ manufacturer: chr  "audi" "audi" "audi" "audi" ...
##  $ model       : chr  "a4" "a4" "a4" "a4" ...
##  $ displ       : num  1.8 1.8 2 2 2.8 2.8 3.1 1.8 1.8 2 ...
##  $ year        : int  1999 1999 2008 2008 1999 1999 2008 1999 1999 2008 ...
##  $ cyl         : int  4 4 4 4 6 6 6 4 4 4 ...
##  $ trans       : chr  "auto(l5)" "manual(m5)" "manual(m6)" "auto(av)" ...
##  $ drv         : chr  "f" "f" "f" "f" ...
##  $ cty         : int  18 21 20 21 16 18 18 18 16 20 ...
##  $ hwy         : int  29 29 31 30 26 26 27 26 25 28 ...
##  $ fl          : chr  "p" "p" "p" "p" ...
##  $ class       : chr  "compact" "compact" "compact" "compact" ...
```

Scatterplot

Exercise 1

    a. Generate a scatterplot to analyze the relationship between `Sepal.Width` and `Sepal.Length` variables.

    b. Set the size of the point as 3 and their colour (`colour` and `fill` arguments as "green").
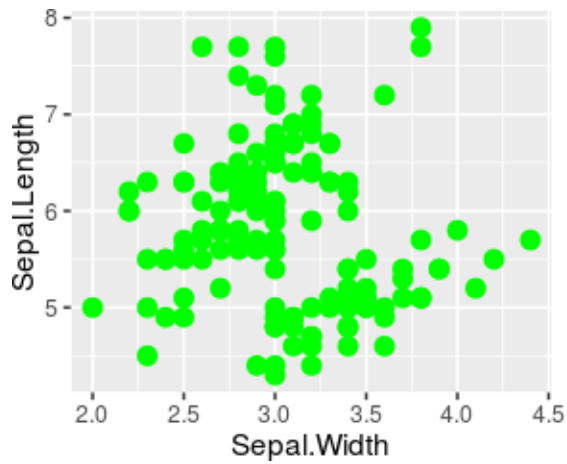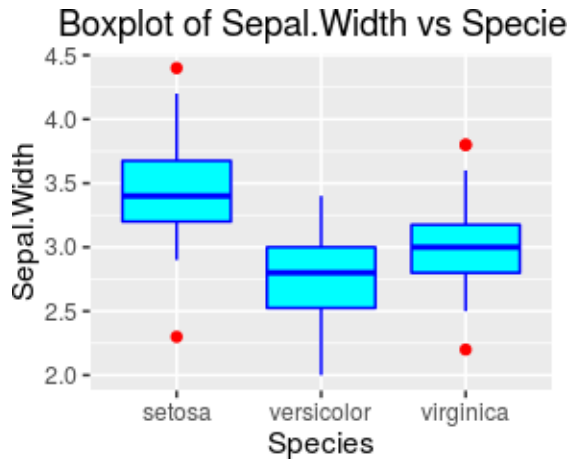

Exercise 2

    a. Generate a scatterplot to analyze the relationship between `Petal.Width` and `Petal.Length` variables according to iris species, mapped as `colour` aes.
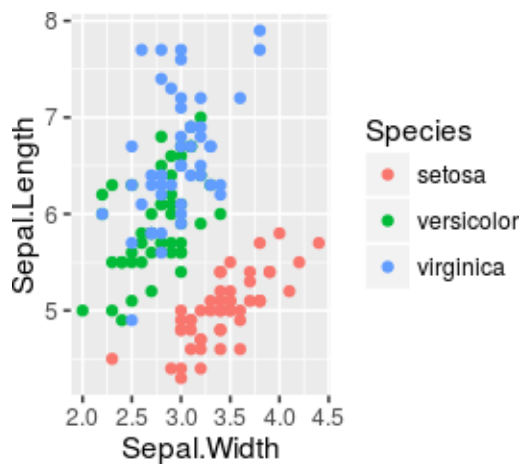
Box Plot

Exercise 1

    a. Build a box plot to compare the differences of sepal width accordingly to the type of iris species.
    b. Set the fill of boxes as "#00FFFF", the colour as "#0000FF" and the outlier colours as "red".
    c. Add the plot title: "Boxplot of Sepal.Width vs Species"

Histogram

Exercise 1

    a. Represent the distribution of sepal length with an histogram.
    b. Set bins fill as "hotpink" and colour as "deeppink".
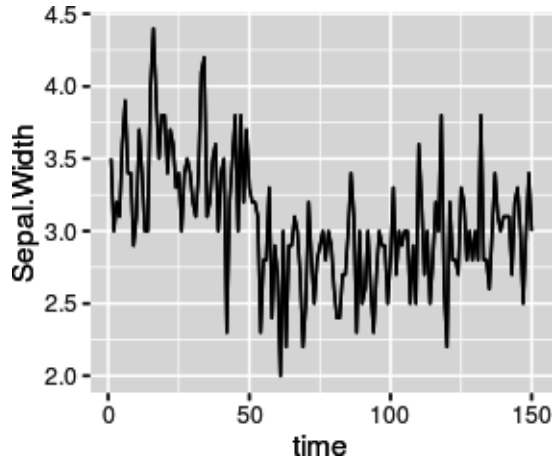   c. Set the number of bins as 15.
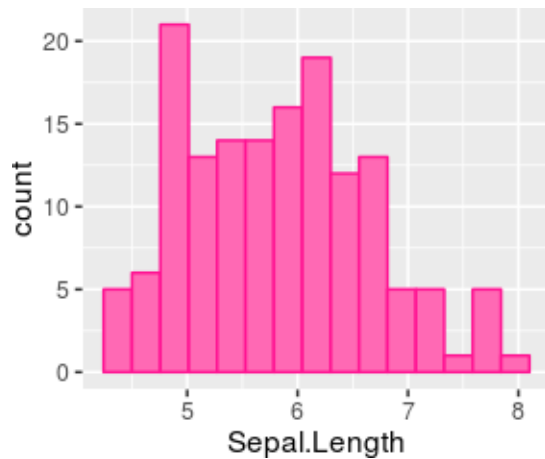
Lineplot

Exercise 1

Let us suppose that the observations on flowers are taken along time, so let us consider the following dataset:

```
require(dplyr)
iris2 <- iris %>% mutate(time=1:150)
```

    a. Build a line plot to visualize the `Sepal.Length` along time.
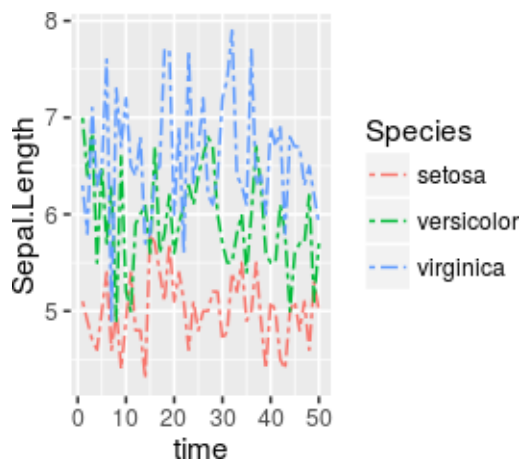

Exercise 2

Let us suppose that the observations on flowers are taken along time, so let us consider the following dataset:

```
iris3 <- iris %>% mutate(time=rep(1:50, times=3))
```

    a. Build a line plot to visualize the `Sepal.Length` along time, according to the `Species`.
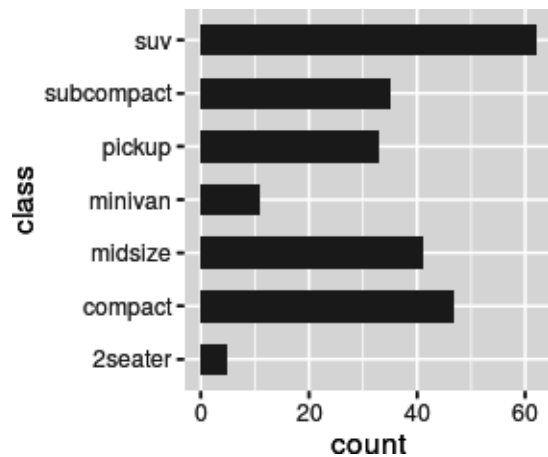   b. Set linetype as "twodash".

Bar graph

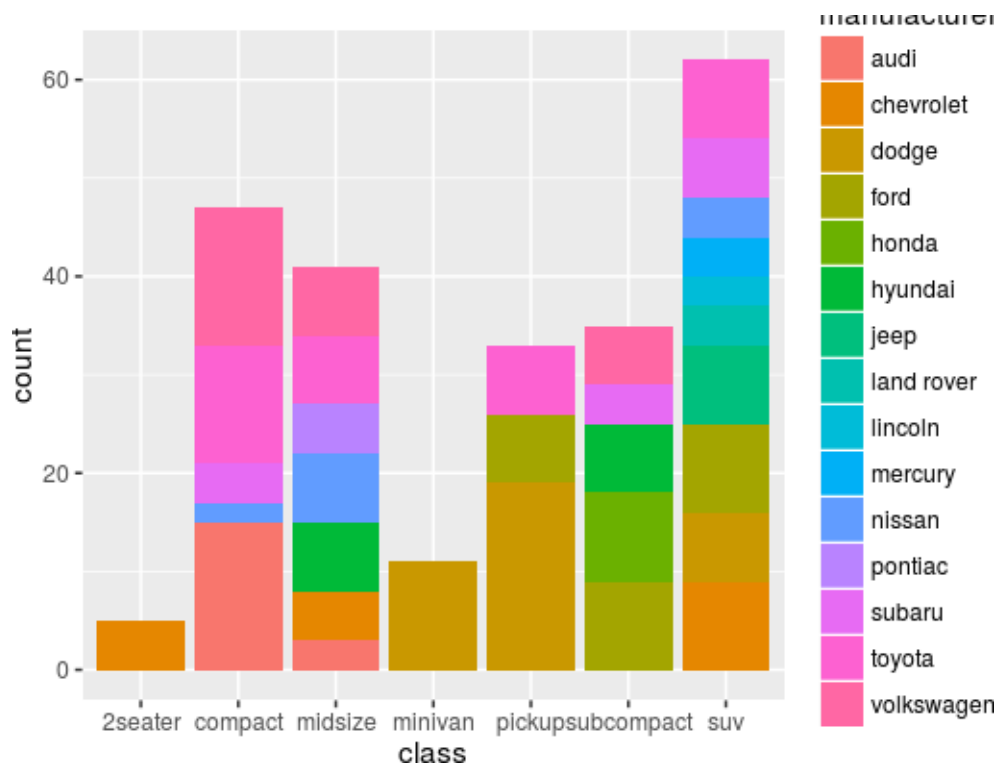Let us consider `mpg` dataset.

Exercise 1

   a. Represent graphically with a bar graph, how many cars there are for each class.
   b. Represent horizontal bar and set bar width as 0.6

Exercise 2

   a. Represent graphically with a bar graph, how many cars there are for each class according to manifacturer.

b. Represent graphically with a bar graph, the distribution of manifacturerfor each class.