Desciptive Statistics and Graphs

Jie Wang

2025-05-07

This file contains sample R code for the lesson "Descriptive Statistics and Graphs".

Example 1

Let's use a toy example. For more details, please read the lesson notes.

```
Time = c(44, 37, 40, 46, 44, 30, 40, 34, 34, 37,
         36, 42, 39, 36, 37, 37, 42, 34, 37, 31)
Time
```

```
##  [1] 44 37 40 46 44 30 40 34 34 37 36 42 39 36 37 37 42 34 37 31
```

Let's find the mean and standard deviation of Time.

```
time.bar = mean(Time)
time.sd <- sd(Time)
```

The (sample) mean time is 37.85 with a (sample) standard deviation 4.295.

If you would like to obtain a set of summary statistics, you may try the following:

```
# The five number summary and the mean
summary(Time)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   30.00   35.50   37.00   37.85   40.50   46.00
```

```
# Range
range(Time)
```

```
## [1] 30 46
```

```
# IQR
IQR(Time)
```

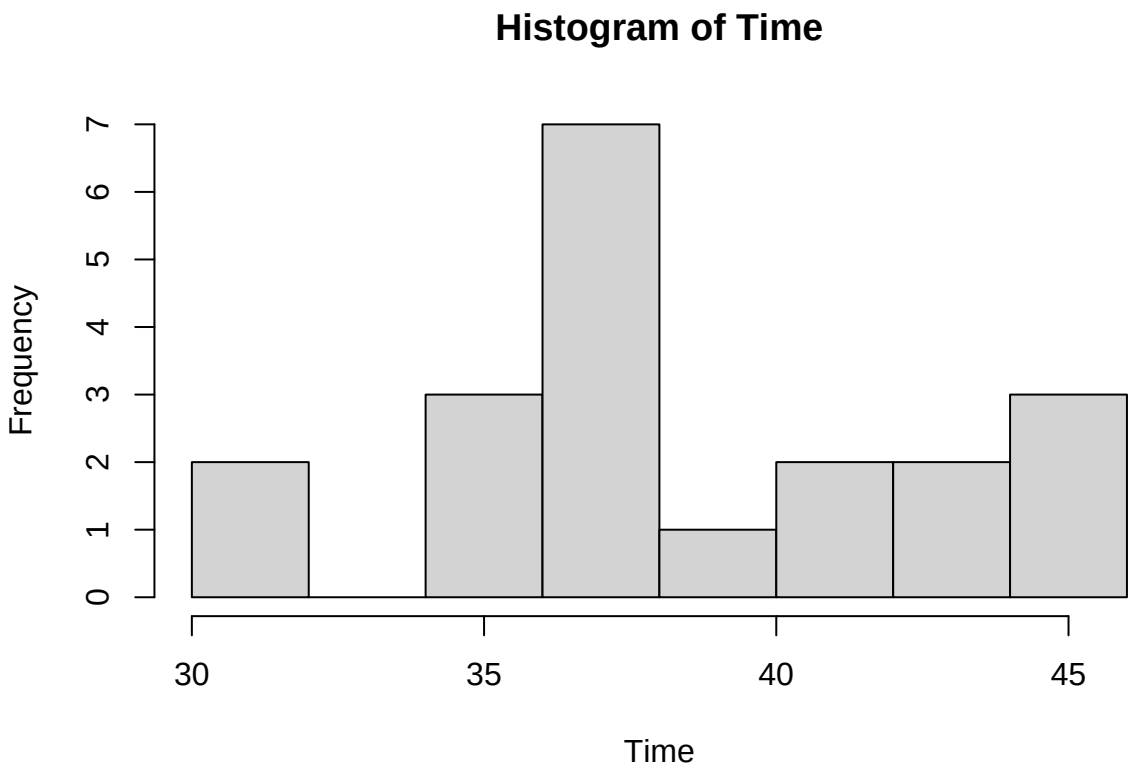
```
## [1] 5
```

```
# Variance
var(Time)
```

```
## [1] 18.45
```

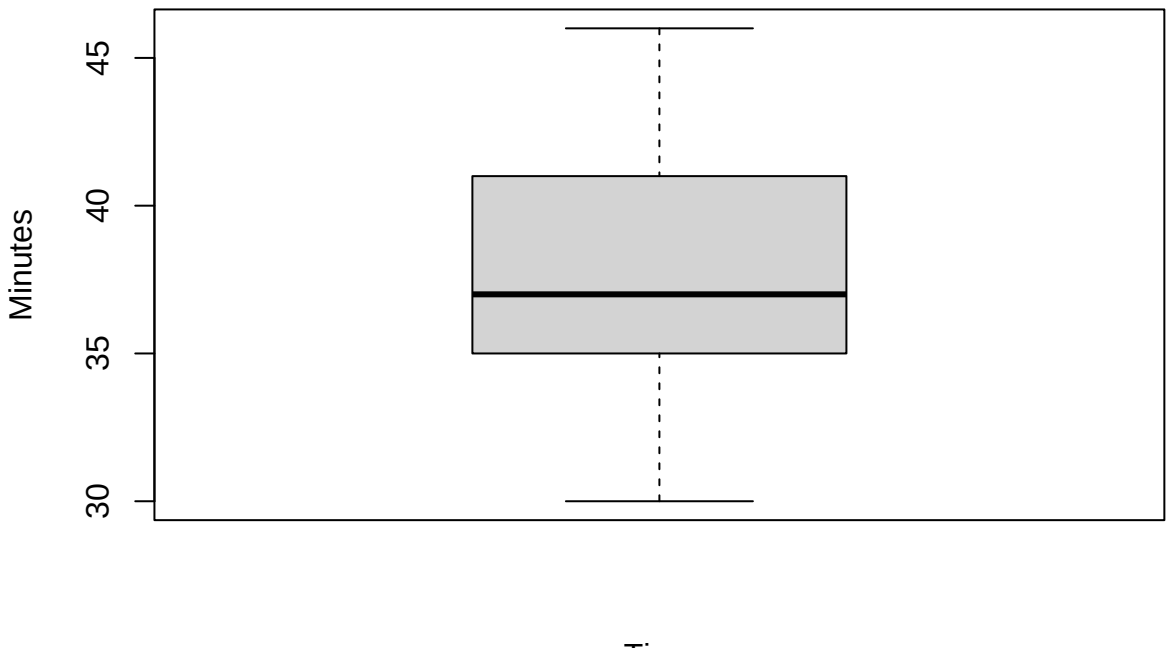
Let's get the graphs.

```
# Histogram
hist(Time, breaks=seq(30,46,2), right = FALSE)
```




```
# right= is logical; if TRUE, the histogram cells are right-closed (left open) intervals.

# Boxplot
boxplot(Time, xlab="Time", ylab="Minutes")
```



Example 2

Let's see another type of variable.

```
Food = c('Pinapple', 'Pinapple', 'Pasta', 'Pizza', 'Pasta', 'Pasta', 'Pinapple', 'Pizza', 'Pasta', 'Pas
print(Food)
```

```
##  [1] "Pinapple" "Pinapple" "Pasta"    "Pizza"    "Pasta"    "Pasta"
##  [7] "Pinapple" "Pizza"    "Pasta"    "Pasta"    "Pinapple" "Pasta"
## [13] "Pizza"    "Pinapple" "Pasta"    "Pizza"    "Pinapple" "Pinapple"
## [19] "Pinapple" "Pinapple"
```

Obtain the frequency (count) or relative frequency of each type of food.

```
table(Food)
```

```
## Food
##    Pasta Pinapple    Pizza
##        7        9        4
```

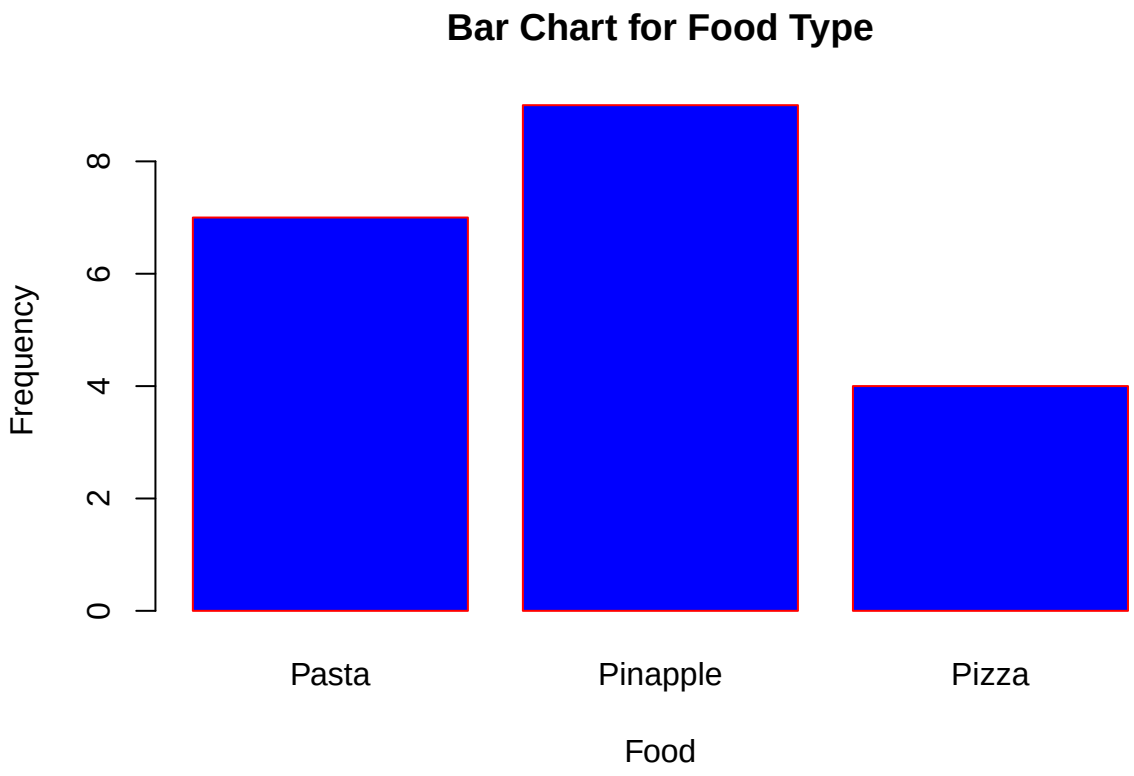
```
table(Food)/length(Food)
```

```
## Food
##    Pasta Pinapple    Pizza
##     0.35     0.45     0.20
```

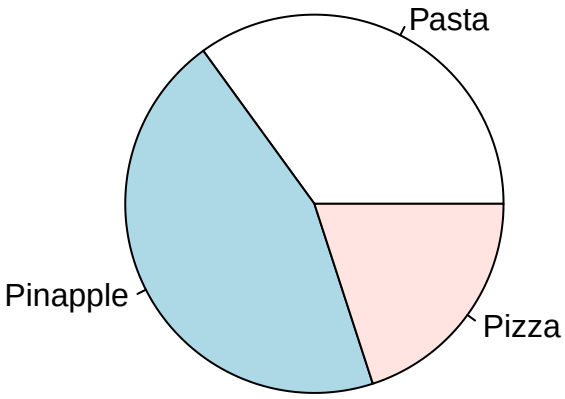
Making graphs.

```
Freq <- c(7,9,4)
FoodType <- c("Pasta","Pinapple","Pizza")
```

```r
# Bar chart. You can add a lot of attributes.
barplot(Freq,names.arg=FoodType,xlab="Food",ylab="Frequency",
        col="blue", main="Bar Chart for Food Type",border="red")
```




```r
# Pie chart. You can make it very simple.
pie(Freq, FoodType)
```



Example 3

```
data(mtcars)
mtcars
```

```
##                      mpg cyl  disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4           21.0   6 160.0 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag       21.0   6 160.0 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710          22.8   4 108.0  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive      21.4   6 258.0 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout   18.7   8 360.0 175 3.15 3.440 17.02  0  0    3    2
## Valiant             18.1   6 225.0 105 2.76 3.460 20.22  1  0    3    1
## Duster 360          14.3   8 360.0 245 3.21 3.570 15.84  0  0    3    4
## Merc 240D           24.4   4 146.7  62 3.69 3.190 20.00  1  0    4    2
## Merc 230            22.8   4 140.8  95 3.92 3.150 22.90  1  0    4    2
## Merc 280            19.2   6 167.6 123 3.92 3.440 18.30  1  0    4    4
## Merc 280C           17.8   6 167.6 123 3.92 3.440 18.90  1  0    4    4
## Merc 450SE          16.4   8 275.8 180 3.07 4.070 17.40  0  0    3    3
## Merc 450SL          17.3   8 275.8 180 3.07 3.730 17.60  0  0    3    3
## Merc 450SLC         15.2   8 275.8 180 3.07 3.780 18.00  0  0    3    3
## Cadillac Fleetwood  10.4   8 472.0 205 2.93 5.250 17.98  0  0    3    4
## Lincoln Continental 10.4   8 460.0 215 3.00 5.424 17.82  0  0    3    4
## Chrysler Imperial   14.7   8 440.0 230 3.23 5.345 17.42  0  0    3    4
## Fiat 128            32.4   4  78.7  66 4.08 2.200 19.47  1  1    4    1
## Honda Civic         30.4   4  75.7  52 4.93 1.615 18.52  1  1    4    2
## Toyota Corolla      33.9   4  71.1  65 4.22 1.835 19.90  1  1    4    1
```

```
## Toyota Corona       21.5   4 120.1  97 3.70 2.465 20.01  1  0    3    1
## Dodge Challenger    15.5   8 318.0 150 2.76 3.520 16.87  0  0    3    2
## AMC Javelin         15.2   8 304.0 150 3.15 3.435 17.30  0  0    3    2
## Camaro Z28          13.3   8 350.0 245 3.73 3.840 15.41  0  0    3    4
## Pontiac Firebird    19.2   8 400.0 175 3.08 3.845 17.05  0  0    3    2
## Fiat X1-9           27.3   4  79.0  66 4.08 1.935 18.90  1  1    4    1
## Porsche 914-2       26.0   4 120.3  91 4.43 2.140 16.70  0  1    5    2
## Lotus Europa        30.4   4  95.1 113 3.77 1.513 16.90  1  1    5    2
## Ford Pantera L      15.8   8 351.0 264 4.22 3.170 14.50  0  1    5    4
## Ferrari Dino        19.7   6 145.0 175 3.62 2.770 15.50  0  1    5    6
## Maserati Bora       15.0   8 301.0 335 3.54 3.570 14.60  0  1    5    8
## Volvo 142E          21.4   4 121.0 109 4.11 2.780 18.60  1  1    4    2
```

```
# Or you can print just a few rows.
head(mtcars)
```

```
##                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

Now, let's play the game of finding statistics and making graphs using this set of data.

Can you find the descriptive statistics of the variable mpg?

Can you find the descriptive statistics of cyl, treating cyl as a categorical variable?

Can you make a histogram for mpg? a bar chart for cyl?

Let me show you what I do to answer these questions.

```
# The (sample) mean of mpg
mpg.bar = mean(mtcars$mpg) # Use the $ sign to reference a variable in a data frame.
mpg.bar
```

```
## [1] 20.09062
```

```
# If you don't want to use the $ sign, you can try with() as shown below.
# The sample standard deviation of mpg
mpg.sd = with(mtcars, sd(mpg))
mpg.sd
```

```
## [1] 6.026948
```

```
# The sample variance can be found by
var(mtcars$mpg)
```

```
## [1] 36.3241
```

```
# or
mpg.sd^2
```

```
## [1] 36.3241
```

```
# The quartiles and the extremes can be found individually or together as shown below.
quantile(mtcars$mpg)
```

```
##     0%    25%    50%    75%   100%
```

```
## 10.400 15.425 19.200 22.800 33.900
```

Thus the mean and standard deviation of mpg are 20.091 and 6.027.

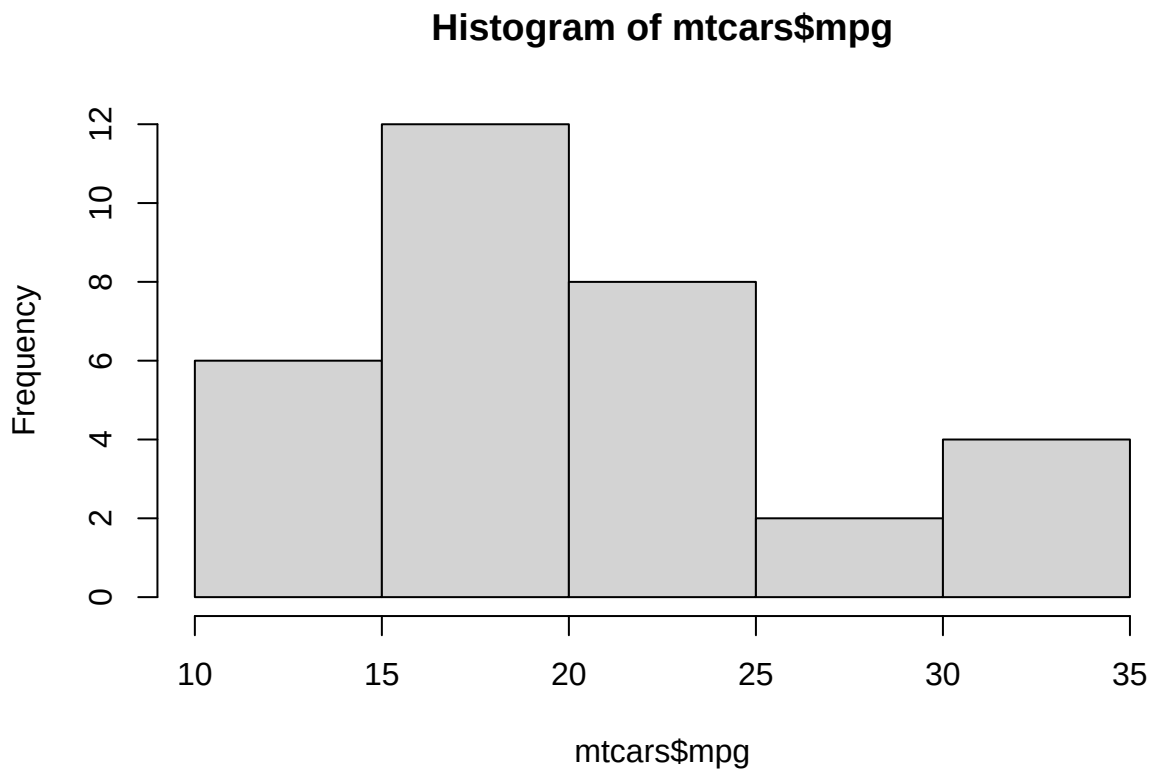
```
# The frequencies of cyl values
table(mtcars$cyl)
```

```
##
##  4  6  8
## 11  7 14
```

```
# Relative frequencies
table(mtcars$cyl)/length(mtcars$cyl)
```

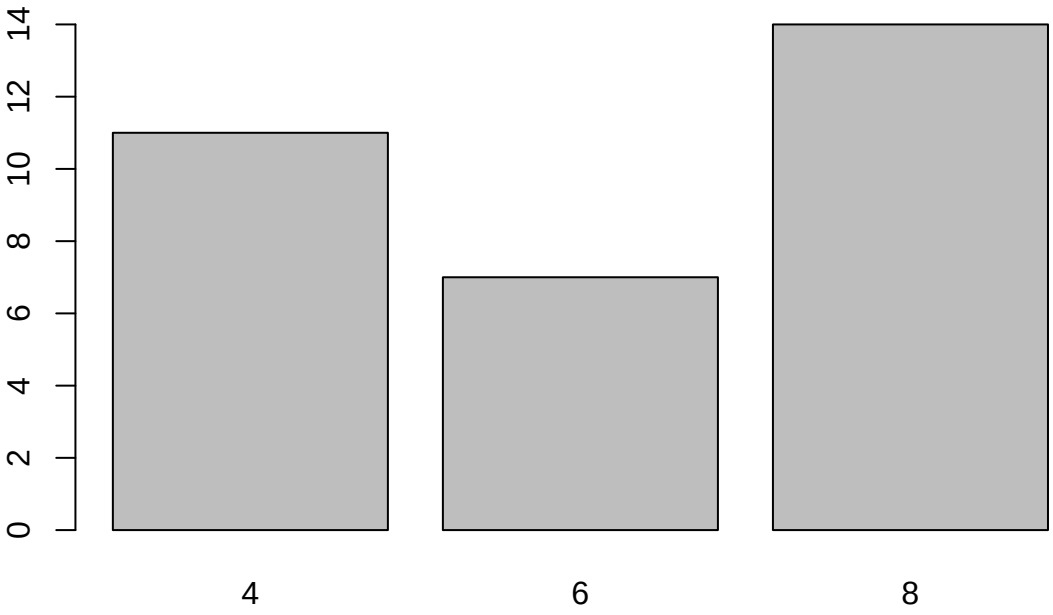
```
##
##       4       6       8
## 0.34375 0.21875 0.43750
```

```
# The histogram
hist(mtcars$mpg)
```




```
# The bar chart
Freq=c(11,7,14)
cyl_cate=c("4","6","8")

barplot(Freq,names.arg=cyl_cate)
```


0
2
4
6
8
10
12
14
4
6
8