

# 01 Basic R for Finance

Boni

9/1/2020

## R Basic

### Working with Matrix

We can create matrix from one atomic vector.

```
x_vec <- c(9, 4, 6, 20, 19, 29)
x_mat <- matrix(data = x_vec, nrow = 2, ncol = 3, byrow = TRUE)
x_mat
```

```
##      [,1] [,2] [,3]
## [1,]    9    4    6
## [2,]   20   19   29
```

```
cor(x_mat)
```

```
##      [,1] [,2] [,3]
## [1,]    1    1    1
## [2,]    1    1    1
## [3,]    1    1    1
```

We can also create matrix from 2 vectors.

```
y_vec = rnorm(6, 23, 98)

matrix_yx <- cbind(x_vec, y_vec)
matrix_yx
```

```
##      x_vec      y_vec
## [1,]    9 120.88997
## [2,]    4 -48.33703
## [3,]    6 123.68690
## [4,]   20 -190.29827
## [5,]   19 146.33734
## [6,]   29 -95.21845
```

```
cor(matrix_yx)
```

```
##           x_vec      y_vec
## x_vec  1.0000000 -0.4362036
## y_vec -0.4362036  1.0000000
```

## Working with DataFrame

```
cash_flow <- rnorm(10, 4, 9)
year <- runif(10, 5, 45)
company <- c("Google", "Google", "Microsoft", "Microsoft", "Apple", "Apple", "Google", "Google", "Microsoft", "Microsoft")
company_data <- data.frame(company, cash_flow, year)
company_data
```

```
##      company cash_flow      year
## 1    Google  7.132735 43.442155
## 2    Google 17.672833 42.371643
## 3 Microsoft 20.320641  9.639075
## 4 Microsoft 13.147890 18.921111
## 5     Apple  6.961772  7.902813
## 6     Apple 18.059448 31.331998
## 7    Google  2.064183 38.361934
## 8    Google  5.842083 29.747339
## 9 Microsoft -1.175229  6.904006
## 10 Microsoft 11.159130 26.739124
```

```
# sub-setting the DF
company_data[1:3,1, drop=FALSE]
```

```
##      company
## 1    Google
## 2    Google
## 3 Microsoft
```

```
company_data$cash_flow
```

```
## [1]  7.132735 17.672833 20.320641 13.147890  6.961772 18.059448  2.064183
## [8]  5.842083 -1.175229 11.159130
```

```
subset(company_data, cash_flow < 6.00)
```

```
##      company cash_flow      year
## 7    Google  2.064183 38.361934
## 8    Google  5.842083 29.747339
## 9 Microsoft -1.175229  6.904006
```

```
# Delete a column
company_data$year = NULL
company_data
```

```
##      company cash_flow
## 1      Google  7.132735
## 2      Google 17.672833
## 3  Microsoft 20.320641
## 4  Microsoft 13.147890
## 5       Apple  6.961772
## 6       Apple 18.059448
## 7      Google  2.064183
## 8      Google  5.842083
## 9  Microsoft -1.175229
## 10 Microsoft 11.159130
```

## Working with Factor

```
# Factor will make R treat string as integer.
investment <- c("stock", "bonds", "stock", "bonds", "stock")
investment_factor <- factor(investment)
# Summary of factor is more informative.
summary(investment)
```

```
##      Length      Class      Mode
##           5 character character
```

```
summary(investment_factor)
```

```
## bonds stock
##      2      3
```

```
# Change the levels.
investment
```

```
## [1] "stock" "bonds" "stock" "bonds" "stock"
```

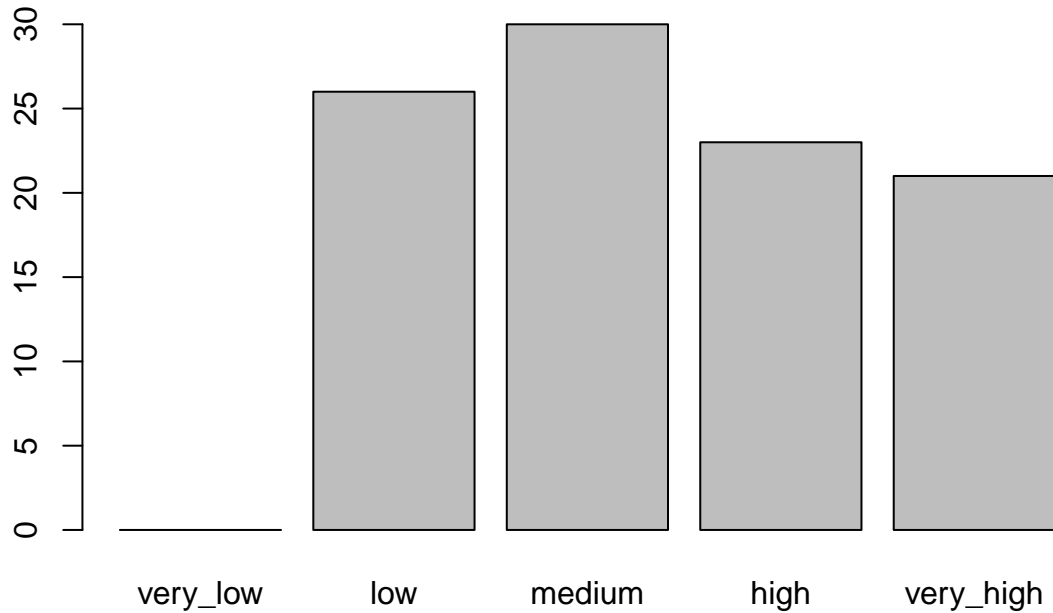
```
levels(investment) <- c("B", "S")
investment
```

```
## [1] "stock" "bonds" "stock" "bonds" "stock"
## attr(,"levels")
## [1] "B" "S"
```

```
# cut() will factoring the numbers based on interval.
price <- runif(100, 20, 100)
breaks <- c(0, 20, 40, 60, 80, 100)
grouped_price <- cut(price, breaks = breaks)
head(grouped_price)
```

```
## [1] (80,100] (40,60] (20,40] (60,80] (40,60] (20,40]
## Levels: (0,20] (20,40] (40,60] (60,80] (80,100]
```

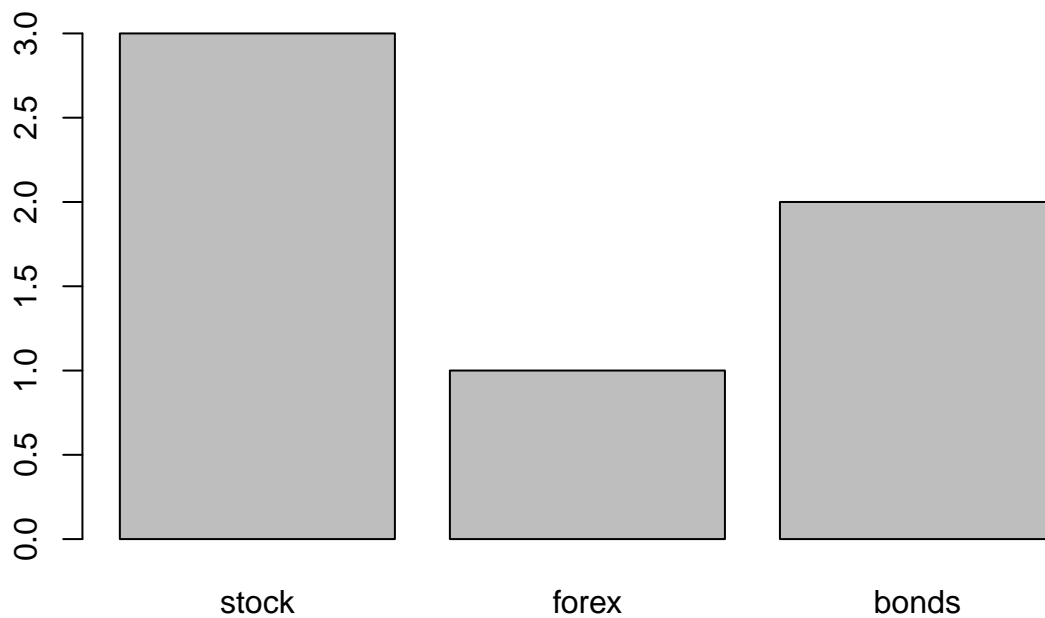
```
levels(grouped_price) <- c("very_low", "low", "medium", "high", "very_high")
# Plot the factor
plot(grouped_price)
```



```
# Order factor
investment <- c("stock", "bonds", "stock", "bonds", "stock", "forex")
ranked_investment <- factor(investment, ordered = TRUE, levels = c("stock", "forex", "bonds"))
ranked_investment
```

```
## [1] stock bonds stock bonds stock forex
## Levels: stock < forex < bonds
```

```
plot(ranked_investment)
```



```
# To remove unavailable level after subset, use drop=TRUE
summary(ranked_investment[1:3, drop = TRUE])
```

```
## stock bonds
##      2      1
```

*notes* R automatically treat string as factor, when using `data.frame()`, to remove this behavior we can add argument `stringsAsFactors = FALSE`.

## Working with List

```
# List can contain multiple data type.
list1 <- list("Boni", investment)
list1
```

```
## [[1]]
## [1] "Boni"
##
## [[2]]
## [1] "stock" "bonds" "stock" "bonds" "stock" "forex"
```

```
# To subset list we use [[]] instead of []
list1[[2]]
```

```
## [1] "stock" "bonds" "stock" "bonds" "stock" "forex"
```

```
# Add names to list.
names(list1) <- c("name", "investment")
list1$name
```

```
## [1] "Boni"
```

```
# Split DataFrame to list
cash_flow <- rnorm(10, 4, 9)
year <- runif(10, 5, 45)
company <- c("Google", "Google", "Microsoft", "Microsoft", "Apple", "Apple", "Google", "Google", "Microso
company_data <- data.frame(company, cash_flow, year)
company_data
```

```
##      company  cash_flow    year
## 1    Google  9.6228782 41.458526
## 2    Google  0.9987962 10.967773
## 3 Microsoft 17.2591199 29.933456
## 4 Microsoft  9.4634295  8.116101
## 5     Apple  1.4394185 15.257383
## 6     Apple -3.0913255 26.840584
## 7     Google -1.4772187  9.728086
## 8     Google -26.4477562 14.735821
## 9 Microsoft 11.3534048 35.183784
## 10 Microsoft -7.6808229  8.870288
```

```
new_list <- split(company_data, company_data$company)
new_list
```

```
## $Apple
##   company cash_flow    year
## 5   Apple  1.439418 15.25738
## 6   Apple -3.091326 26.84058
##
## $Google
##   company  cash_flow    year
## 1  Google  9.6228782 41.458526
## 2  Google  0.9987962 10.967773
## 7  Google -1.4772187  9.728086
## 8  Google -26.4477562 14.735821
##
## $Microsoft
##   company cash_flow    year
## 3 Microsoft 17.259120 29.933456
## 4 Microsoft  9.463430  8.116101
## 9 Microsoft 11.353405 35.183784
## 10 Microsoft -7.680823  8.870288
```

```
# Unsplit list  
unsplit(new_list, company)
```

```
##      company  cash_flow    year  
## 1    Google  9.6228782 41.458526  
## 2    Google  0.9987962 10.967773  
## 3 Microsoft 17.2591199 29.933456  
## 4 Microsoft  9.4634295  8.116101  
## 5     Apple  1.4394185 15.257383  
## 6     Apple -3.0913255 26.840584  
## 7    Google -1.4772187  9.728086  
## 8    Google -26.4477562 14.735821  
## 9 Microsoft 11.3534048 35.183784  
## 10 Microsoft -7.6808229  8.870288
```

```
# Getting attributes  
attributes(new_list)
```

```
## $names  
## [1] "Apple"      "Google"     "Microsoft"
```

```
attr(new_list, "names")
```

```
## [1] "Apple"      "Google"     "Microsoft"
```