R_2_select_subset_booleans

Mike Montes

2024-01-13

1. Load and call data

```
library(readxl)
SP500 <- read_excel("C:/Users/mmsax/School_Portfolio/Coding_Skills/SP500.xls")
View(SP500)
print(SP500)
```

```
## # A tibble: 1,773 x 6
##      Date SP500 Dividend Earnings   CPI  Rate
##     <dbl> <dbl>    <dbl>    <dbl> <dbl> <dbl>
##  1 1871.   4.44     0.26      0.4  12.5  5.32
##  2 1871.   4.5      0.26      0.4  12.8  5.32
##  3 1871.   4.61     0.26      0.4  13.0  5.33
##  4 1871.   4.74     0.26      0.4  12.6  5.33
##  5 1871.   4.86     0.26      0.4  12.3  5.33
##  6 1871.   4.82     0.26      0.4  12.1  5.34
##  7 1871.   4.73     0.26      0.4  12.1  5.34
##  8 1871.   4.79     0.26      0.4  11.9  5.34
##  9 1871.   4.84     0.26      0.4  12.2  5.35
## 10 1871.   4.59     0.26      0.4  12.4  5.35
## # i 1,763 more rows
```

2. Number of rows and columns in dataset using 'dim', 'nrow', & 'ncol'

nrow = 1773

ncol = 6

```
dim(SP500)
```

```
## [1] 1773    6
```

```
nrow(SP500)
```

```
## [1] 1773
```

```
ncol(SP500)
```

```
## [1] 6
```

3. Three methods of selecting ' SP500', 'CPI', & 'Rate'

```
SP500[ , c('SP500', 'CPI', 'Rate')]
```

```
## # A tibble: 1,773 x 3
##      SP500   CPI  Rate
##      <dbl> <dbl> <dbl>
##  1   4.44  12.5  5.32
##  2   4.5   12.8  5.32
##  3   4.61  13.0  5.33
##  4   4.74  12.6  5.33
##  5   4.86  12.3  5.33
##  6   4.82  12.1  5.34
##  7   4.73  12.1  5.34
##  8   4.79  11.9  5.34
##  9   4.84  12.2  5.35
## 10   4.59  12.4  5.35
## # i 1,763 more rows
```

```
SP500[ , -which(names(SP500) %in% c('Date', 'Dividend', 'Earnings'))]
```

```
## # A tibble: 1,773 x 3
##      SP500   CPI  Rate
##      <dbl> <dbl> <dbl>
##  1   4.44  12.5  5.32
##  2   4.5   12.8  5.32
##  3   4.61  13.0  5.33
##  4   4.74  12.6  5.33
##  5   4.86  12.3  5.33
##  6   4.82  12.1  5.34
##  7   4.73  12.1  5.34
##  8   4.79  11.9  5.34
##  9   4.84  12.2  5.35
## 10   4.59  12.4  5.35
## # i 1,763 more rows
```

```
head(subset(SP500, select = -c(Date, Dividend, Earnings)))
```

```
## # A tibble: 6 x 3
##     SP500   CPI  Rate
##     <dbl> <dbl> <dbl>
## 1   4.44  12.5  5.32
## 2   4.5   12.8  5.32
## 3   4.61  13.0  5.33
## 4   4.74  12.6  5.33
## 5   4.86  12.3  5.33
## 6   4.82  12.1  5.34
```

4. Select rows 10, 100, 500 & 1500

```
SP500[c(10, 100, 500, 1500), ]
```

```
## # A tibble: 4 x 6
##    Date  SP500 Dividend Earnings   CPI  Rate
##   <dbl>  <dbl>    <dbl>    <dbl> <dbl> <dbl>
## 1 1871.   4.59     0.26      0.4  12.4  5.35
## 2 1879.   3.77    0.187    0.333  8.18  4.17
## 3 1912.   9.81    0.477    0.663  9.70  4.27
## 4 1995. 615.      13.8     34.0  154.   5.71
```

5. Two methods tp select SP500 > 500 or CPI < 100

```
SP500[c((SP500$SP500 > 500) | (SP500$CPI < 100)), ]
```

```
## # A tibble: 1,633 x 6
##     Date SP500 Dividend Earnings   CPI  Rate
##    <dbl> <dbl>    <dbl>    <dbl> <dbl> <dbl>
##  1 1871.  4.44     0.26      0.4  12.5  5.32
##  2 1871.  4.5      0.26      0.4  12.8  5.32
##  3 1871.  4.61     0.26      0.4  13.0  5.33
##  4 1871.  4.74     0.26      0.4  12.6  5.33
##  5 1871.  4.86     0.26      0.4  12.3  5.33
##  6 1871.  4.82     0.26      0.4  12.1  5.34
##  7 1871.  4.73     0.26      0.4  12.1  5.34
##  8 1871.  4.79     0.26      0.4  11.9  5.34
##  9 1871.  4.84     0.26      0.4  12.2  5.35
## 10 1871.  4.59     0.26      0.4  12.4  5.35
## # i 1,623 more rows
```

```
SP_CPI <- subset(SP500, (SP500 > 500) | (CPI <100))
print(SP_CPI)
```

```
## # A tibble: 1,633 x 6
##     Date SP500 Dividend Earnings   CPI  Rate
##    <dbl> <dbl>    <dbl>    <dbl> <dbl> <dbl>
##  1 1871.  4.44     0.26      0.4  12.5  5.32
##  2 1871.  4.5      0.26      0.4  12.8  5.32
##  3 1871.  4.61     0.26      0.4  13.0  5.33
##  4 1871.  4.74     0.26      0.4  12.6  5.33
##  5 1871.  4.86     0.26      0.4  12.3  5.33
##  6 1871.  4.82     0.26      0.4  12.1  5.34
##  7 1871.  4.73     0.26      0.4  12.1  5.34
##  8 1871.  4.79     0.26      0.4  11.9  5.34
##  9 1871.  4.84     0.26      0.4  12.2  5.35
## 10 1871.  4.59     0.26      0.4  12.4  5.35
## # i 1,623 more rows
```

6. Two methods to select Earnings > 50 & Rate < 3 only showing columns 'SP500' & "Dividend'

() used to print

```
(Earnings_Rate <-  subset(SP500, (Earnings > 50) & (Rate < 3), select = c(SP500, Dividend)))
```

```
## # A tibble: 89 x 2
##     SP500 Dividend
##     <dbl>    <dbl>
##  1 1087.      22.2
##  2 1122.      22.4
##  3 1172.      22.5
##  4 1199.      22.6
##  5 1185.      24.9
##  6 1174.      25.2
##  7 1207.      25.6
##  8 1226.      26.0
##  9 1243.      26.4
## 10 1301.      26.7
## # i 79 more rows
```

7. Remove 'Rate' column

```
SP500$Rate <- NULL
colnames(SP500)
```

```
## [1] "Date"     "SP500"     "Dividend" "Earnings" "CPI"
```

8. Add column 'RealPrice' according to given formula

There are several ways to do this. The dataset was received sorted by ascending date, so I'm choosing to use the last entry as the latest CPI. Alternatively, this value could be chosen by sorting by date and choosing the latest CPI or by choosing the CPI row using a conditional maximum date.

```
SP500$RealPrice <- SP500$SP500 * SP500$CPI / tail(SP500$CPI, 1)
head(SP500)
```

```
## # A tibble: 6 x 6
##     Date SP500 Dividend Earnings   CPI RealPrice
##    <dbl> <dbl>    <dbl>    <dbl> <dbl>     <dbl>
## 1 1871.  4.44     0.26      0.4  12.5     0.219
## 2 1871.  4.5      0.26      0.4  12.8     0.229
## 3 1871.  4.61     0.26      0.4  13.0     0.238
## 4 1871.  4.74     0.26      0.4  12.6     0.236
## 5 1871.  4.86     0.26      0.4  12.3     0.236
## 6 1871.  4.82     0.26      0.4  12.1     0.231
```

9. Add column 'RealEarnings' according to given formula

```
SP500$RealEarnings <- SP500$Earnings * SP500$CPI / tail(SP500$CPI, 1)
head(SP500)
```

```
## # A tibble: 6 x 7
##    Date SP500 Dividend Earnings   CPI RealPrice RealEarnings
##   <dbl> <dbl>    <dbl>    <dbl> <dbl>     <dbl>        <dbl>
## 1 1871.  4.44     0.26      0.4  12.5     0.219       0.0197
## 2 1871.  4.5      0.26      0.4  12.8     0.229       0.0204
## 3 1871.  4.61     0.26      0.4  13.0     0.238       0.0207
## 4 1871.  4.74     0.26      0.4  12.6     0.236       0.0199
## 5 1871.  4.86     0.26      0.4  12.3     0.236       0.0194
## 6 1871.  4.82     0.26      0.4  12.1     0.231       0.0191
```

10. Add column 'PERatio' according to given formula

```
SP500$PERatio <- SP500$RealPrice / SP500$RealEarnings
head(SP500)
```

```
## # A tibble: 6 x 8
##    Date SP500 Dividend Earnings   CPI RealPrice RealEarnings PERatio
##   <dbl> <dbl>    <dbl>    <dbl> <dbl>     <dbl>        <dbl>   <dbl>
## 1 1871.  4.44     0.26      0.4  12.5     0.219       0.0197    11.1
## 2 1871.  4.5      0.26      0.4  12.8     0.229       0.0204    11.2
## 3 1871.  4.61     0.26      0.4  13.0     0.238       0.0207    11.5
## 4 1871.  4.74     0.26      0.4  12.6     0.236       0.0199    11.9
## 5 1871.  4.86     0.26      0.4  12.3     0.236       0.0194    12.1
## 6 1871.  4.82     0.26      0.4  12.1     0.231       0.0191    12.0
```