lubricate-package-usage

Léo Dange

5/7/2020

Introduction

Here is a quick usage of the lubricate package include in tidyverse. This package help extract value from date or times, convert timestamps, ect...

We are going to use the following libraries:

```
library(dslabs)
library(lubridate)
library(dplyr)
library(ggplot2)
library(gridExtra)
library(tinytex)
```

and load the data include in dslabs:

```
data(movielens)
```

Commandes

Here is the head of our file containing movie rating

```
head(movielens,1)
```

```
##   movieId           title year genres userId rating  timestamp
## 1      31 Dangerous Minds 1995  Drama      1    2.5 1260759144
```

First we are going to translate the timestamps in date in a new column "date" :

```
movielens <-  movielens %>% mutate(date = as_datetime(timestamp))
head(movielens,1)
```

```
##   movieId           title year genres userId rating  timestamp
## 1      31 Dangerous Minds 1995  Drama      1    2.5 1260759144
##                  date
## 1 2009-12-14 02:52:24
```

then we will extract for each review, which year and which hour a review has been posted. The "year" from the timestamp will replace the year of the movie as we do not need it for this analysis.

```r
movielens <-  movielens %>% mutate(year = year(date), hour = hour(date))
head(movielens,1)
```

```
##   movieId           title year genres userId rating  timestamp
## 1      31 Dangerous Minds 2009  Drama      1    2.5 1260759144
##                   date hour
## 1 2009-12-14 02:52:24    2
```

Last we'll creat two objects "y" & "h" to stores the count of year and hour and easily see what year has the most review

```r
y <- movielens %>% group_by(year) %>% count()
head(y,3)
```

```
## # A tibble: 3 x 2
## # Groups:   year [3]
##    year     n
##   <dbl> <int>
## 1  1995     3
## 2  1996  6239
## 3  1997  3294
```

```r
h <- movielens %>% group_by(hour) %>% count()
head(h,3)
```

```
## # A tibble: 3 x 2
## # Groups:   hour [3]
##    hour     n
##   <int> <int>
## 1     0  3960
## 2     1  5296
## 3     2  4056
```

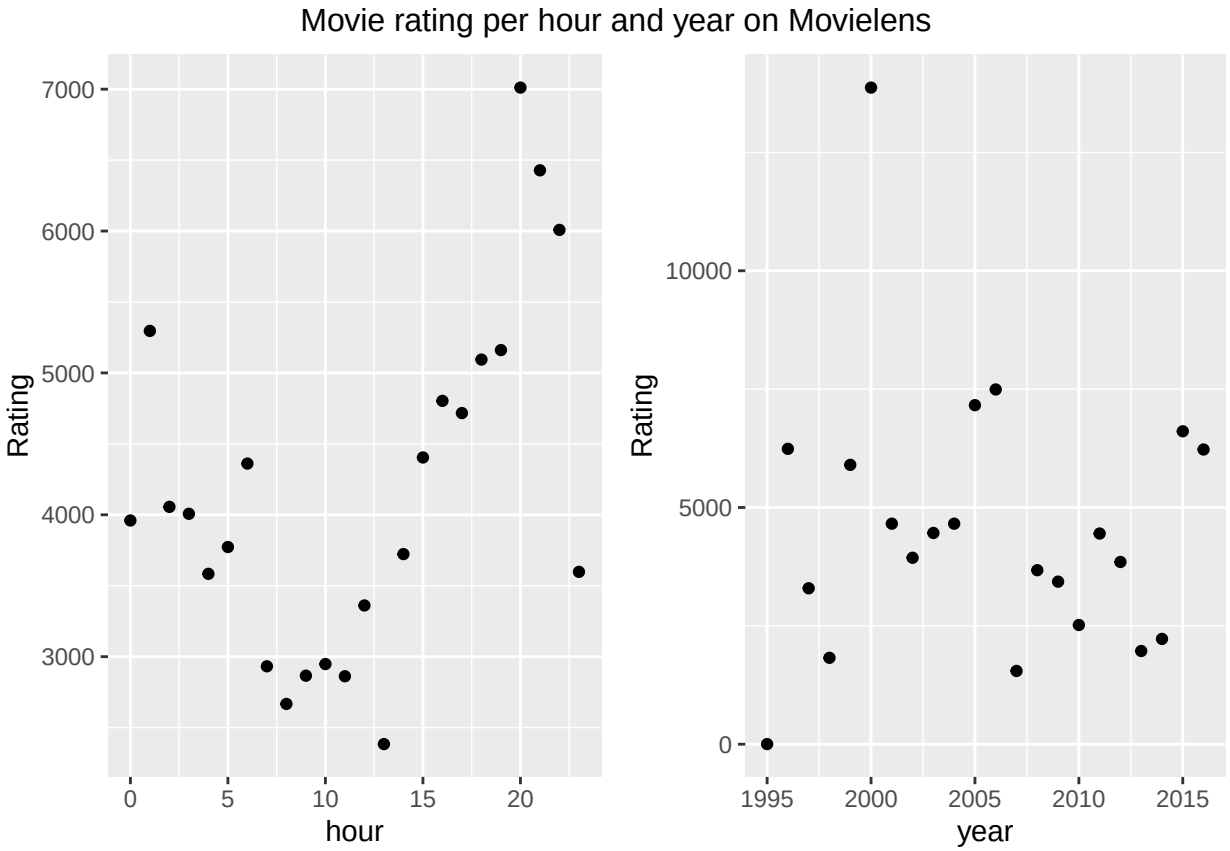
Graph

If we plot them we can see the evolution of rating during the day (per hour) and time (per year)

```r
y <- data.frame(y, row.names = NULL) %>%
    transform(y, year = as.numeric(year),
    n = as.numeric(n))
h <- data.frame(h, row.names = NULL) %>%
    transform(h, hour = as.numeric(hour),
    n = as.numeric(n))

h_plot <- h %>% ggplot() +
    geom_point(aes(x = hour, y = n)) +
    ylab("Rating")
y_plot <- y %>% ggplot() +
    geom_point(aes(x = year, y = n)) +
    ylab("Rating")

grid.arrange(h_plot,y_plot, ncol = 2, top = "Movie rating per hour and year on Movielens")
```


Movie rating per hour and year on Movielens
Rating
7000
6000
5000
4000
3000
0
5
10
15
20
hour
Rating
10000
5000
0
1995
2000
2005
2010
2015
year