Prctice from14-24

Muhammad Huzaifah Zia

10/12/2021

Subsetting And Sorting - Using Dates

Subsetting and Sorting

Setting seed for reproducible result from the random int generated by sample

```
set.seed(13435)
```

making a data frame with 3 columns

shuffling the rows usinng sample

and making the 1st and 3rd elements of var2 column NA

displaying x

```
x <- data.frame("var1" = sample(1:5), "var2" = sample(6:10), "var3" = sample(11:15))
x <- x[sample(1:5),]; x$var2[c(1,3)] =NA
x
```

```
##   var1 var2 var3
## 5    2   NA   11
## 4    4   10   12
## 1    3   NA   14
## 2    1    7   15
## 3    5    6   13
```

Displaying column 1 of the dataframe x

subsetting based on column

```
x[,1]
```

```
## [1] 2 4 3 1 5
```

Displaying column where column = "var1" of dataframe x

subsetting data based on column

```
x [,"var1"]
```

```
## [1] 2 4 3 1 5
```

Displaying rows 1 to 2 of where column = "var2" of dataframe x

```
x [1:2,"var2"]
```

```
## [1] NA 10
```

subsetting x by logical comparisons

```
x[(x$var1 <= 3 & x$var3 > 11),]
```

```
##   var1 var2 var3
## 1    3   NA   14
## 2    1    7   15
```

```
x[(x$var1 <= 3 | x$var3 > 15),]
```

```
##   var1 var2 var3
## 5    2   NA   11
## 1    3   NA   14
## 2    1    7   15
```

subsetting by NAs

```
x[which(x$var2 > 8),]
```

```
##   var1 var2 var3
## 4    4   10   12
```

sorting data

```
sort(x$var1)
```

```
## [1] 1 2 3 4 5
```

```
sort(x$var2, na.last = TRUE)
```

```
## [1]  6  7 10 NA NA
```

sorting by order

```
x[order(x$var1),]
```

```
##   var1 var2 var3
## 2    1    7   15
## 5    2   NA   11
## 1    3   NA   14
## 4    4   10   12
## 3    5    6   13
```

```
x[order(x$var1, x$var3),]
```

```
##   var1 var2 var3
## 2    1    7   15
## 5    2   NA   11
## 1    3   NA   14
## 4    4   10   12
## 3    5    6   13
```

ordering with plyr

```
library(plyr)
arrange(x,var1)
```

```
##   var1 var2 var3
## 1    1    7   15
## 2    2   NA   11
## 3    3   NA   14
## 4    4   10   12
## 5    5    6   13
```

```
arrange(x,desc(var1))
```

```
##   var1 var2 var3
## 1    5    6   13
## 2    4   10   12
## 3    3   NA   14
## 4    2   NA   11
## 5    1    7   15
```

Adding rows and columns

```
x$var4 <- rnorm(5)
x
```

```
##   var1 var2 var3       var4
## 5    2   NA   11 -0.4150458
## 4    4   10   12  2.5437602
## 1    3   NA   14  1.5545298
## 2    1    7   15 -0.6192328
## 3    5    6   13 -0.9261035
```

```
Y <- cbind(x,rnorm(5))
Y
```

```
##   var1 var2 var3       var4    rnorm(5)
## 5    2   NA   11 -0.4150458 -0.66549949
## 4    4   10   12  2.5437602 -0.02166735
## 1    3   NA   14  1.5545298 -0.17411953
## 2    1    7   15 -0.6192328  0.23900438
## 3    5    6   13 -0.9261035 -1.83245959
```

_*_

Summarizing data

Fetching data from web

```
if (!file.exists("./data")){
  dir.create("./data")
}
fileUrl <- "https://data.baltimorecity.gov/api/views/k5ry-ef3g/rows.csv?accessType=DOWNLOAD"
download.file(fileUrl,destfile="./data/restaurants.csv", method = "curl" )
restData <- read.csv("./data/restaurants.csv")
```

lookat a bit of data

```
head(restData,n=3)
```

```
##       X..DOCTYPE.html.
## 1       <html lang=en>
## 2               <head>
## 3 <meta charset=utf-8>
```