WEEK 3

```
set.seed(13435)
X <- data.frame("var1"=sample(1:5), "var2"=sample(6:10), "var3"=sample(11:15))
X
```

```
##   var1 var2 var3
## 1    3    8   14
## 2    1    7   15
## 3    5    6   13
## 4    4   10   12
## 5    2    9   11
```

le mischio e inserisco dei NA

```
X <- X[sample(1:5),]
X
```

```
##   var1 var2 var3
## 5    2    9   11
## 4    4   10   12
## 1    3    8   14
## 2    1    7   15
## 3    5    6   13
```

```
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
```

guardare una colonna

```
X[,1]
```

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

```
X[,"var1"]
```

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

Ricercare solo determine colonne o righe in base a logiche

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

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

Which, ritorna gli indici che approvano determinate condizioni

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

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

per ordinare i valori

```
sort(X$var1) #crescente
```

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

```
sort(X$var1 , decreasing=TRUE)
```

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

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

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

ordinare il databse in base ai valori di una colonna

```
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
```

Libreria plyr

```r
library(plyr)
arrange(X, var1) #ordina il database rispetto alla colonna
```

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

```r
arrange(X, desc(var1)) #ordina decrescente
```

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

aggiungere una nuova colonna

```r
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
```

```r
# posso fare lo stess con
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
```

manipolare i dati

```
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")
```

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

```
##    name zipCode neighborhood councilDistrict policeDistrict
## 1   410   21206    Frankford               2   NORTHEASTERN
## 2  1919   21231  Fells Point               1   SOUTHEASTERN
## 3 SAUTE   21224       Canton               1   SOUTHEASTERN
##                        Location.1 X2010.Census.Neighborhoods
## 1 4509 BELAIR ROAD\nBaltimore, MD                         NA
## 2    1919 FLEET ST\nBaltimore, MD                         NA
## 3   2844 HUDSON ST\nBaltimore, MD                         NA
##   X2010.Census.Wards.Precincts Zip.Codes
## 1                           NA        NA
## 2                           NA        NA
## 3                           NA        NA
```

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

```
##                  name zipCode  neighborhood councilDistrict policeDistrict
## 1325 ZINK'S CAF\u0090   21213 Belair-Edison              13   NORTHEASTERN
## 1326     ZISSIMOS BAR   21211       Hampden               7       NORTHERN
## 1327           ZORBAS   21224     Greektown               2   SOUTHEASTERN
##                            Location.1 X2010.Census.Neighborhoods
## 1325 3300 LAWNVIEW AVE\nBaltimore, MD                         NA
## 1326      1023 36TH ST\nBaltimore, MD                         NA
## 1327  4710 EASTERN Ave\nBaltimore, MD                         NA
##      X2010.Census.Wards.Precincts Zip.Codes
## 1325                           NA        NA
## 1326                           NA        NA
## 1327                           NA        NA
```

```
summary(restData)
```

```
##                            name         zipCode            neighborhood
##  MCDONALD'S                  :   8   Min.   :-21226   Downtown    :128
##  POPEYES FAMOUS FRIED CHICKEN:   7   1st Qu.: 21202   Fells Point : 91
##  SUBWAY                      :   6   Median : 21218   Inner Harbor: 89
##  KENTUCKY FRIED CHICKEN      :   5   Mean   : 21185   Canton      : 81
##  BURGER KING                 :   4   3rd Qu.: 21226   Federal Hill: 42
##  DUNKIN DONUTS               :   4   Max.   : 21287   Mount Vernon: 33
##  (Other)                     :1293                    (Other)     :863
##  councilDistrict       policeDistrict                            Location.1
##  Min.   : 1.000   SOUTHEASTERN:385    1101 RUSSELL ST\nBaltimore, MD:   9
##  1st Qu.: 2.000   CENTRAL     :288    201 PRATT ST\nBaltimore, MD   :   8
##  Median : 9.000   SOUTHERN    :213    2400 BOSTON ST\nBaltimore, MD :   8
##  Mean   : 7.191   NORTHERN    :157    300 LIGHT ST\nBaltimore, MD   :   5
##  3rd Qu.:11.000   NORTHEASTERN: 72    300 CHARLES ST\nBaltimore, MD :   4
##  Max.   :14.000   EASTERN     : 67    301 LIGHT ST\nBaltimore, MD   :   4
##                    (Other)     :145    (Other)                       :1289
##  X2010.Census.Neighborhoods X2010.Census.Wards.Precincts Zip.Codes
##  Mode:logical               Mode:logical                 Mode:logical
##  NA's:1327                  NA's:1327                    NA's:1327
##
##
##
##
##
```

```
str(restData)
```

```
## 'data.frame':    1327 obs. of  9 variables:
##  $ name                      : Factor w/ 1277 levels "#1 CHINESE KITCHEN",..: 9 3 992 1 2 4 5 6 7
##  $ zipCode                   : int  21206 21231 21224 21211 21223 21218 21205 21211 21205 21231 ..
##  $ neighborhood              : Factor w/ 173 levels "Abell","Arlington",..: 53 52 18 66 104 33 98
##  $ councilDistrict           : int  2 1 1 14 9 14 13 7 13 1 ...
##  $ policeDistrict            : Factor w/ 9 levels "CENTRAL","EASTERN",..: 3 6 6 4 8 3 6 4 6 6 ...
##  $ Location.1                : Factor w/ 1210 levels "1 BIDDLE ST\nBaltimore, MD",..: 835 334 554
##  $ X2010.Census.Neighborhoods  : logi  NA NA NA NA NA NA ...
##  $ X2010.Census.Wards.Precincts: logi  NA NA NA NA NA NA ...
##  $ Zip.Codes                 : logi  NA NA NA NA NA NA ...
```

Quantiles of quantitative variables

The generic function quantile produces sample quantiles corresponding to the given probabilities. The smallest observation corresponds to a probability of 0 and the largest to a probability of 1.

```
quantile(restData$councilDistrict , na.rm=TRUE)
```

```
##   0%  25%  50%  75% 100%
##    1    2    9   11   14
```

```
quantile(restData$zipCode , na.rm=TRUE)
```

```
##       0%      25%      50%      75%     100%
## -21226.0  21202.0  21218.0  21225.5  21287.0
```

```
quantile(restData$zipCode, probs=c(0.5,0.75,0.9)) #guardo le probabilità che mi interessano
```

```
##     50%     75%     90%
## 21218.0 21225.5 21231.0
```

Table

```
table(restData$zipCode , useNA="ifany") #quante volte appaiono i numeri nella tabella, aggiungo una col
```

```
##
## -21226  21201  21202  21205  21206  21207  21208  21209  21210  21211  21212
##      1    136    201     27     30      4      1      8     23     41     28
##  21213  21214  21215  21216  21217  21218  21220  21222  21223  21224  21225
##     31     17     54     10     32     69      1      7     56    199     19
##  21226  21227  21229  21230  21231  21234  21237  21239  21251  21287
##     18      4     13    156    127      7      1      3      2      1
```

```
table (restData$policeDistrict, restData$zipCode ) #conto quanti ristoranti ci sono in un distretto di
```

```
##
##                 -21226 21201 21202 21205 21206 21207 21208 21209 21210 21211
##   CENTRAL            0   129   143     0     0     0     1     0     0     0
##   EASTERN            0     1    12    20     0     0     0     0     0     0
##   NORTHEASTERN       0     0     0     0    30     0     0     0     0     0
##   NORTHERN           0     0     0     0     0     0     0     8    23    41
##   NORTHWESTERN       0     0     0     0     0     3     0     0     0     0
##   SOUTHEASTERN       0     0    42     7     0     0     0     0     0     0
##   SOUTHERN           1     6     4     0     0     0     0     0     0     0
##   SOUTHWESTERN       0     0     0     0     0     1     0     0     0     0
##   WESTERN            0     0     0     0     0     0     0     0     0     0
##
##                 21212 21213 21214 21215 21216 21217 21218 21220 21222 21223
##   CENTRAL           1     0     0     0     0    10     1     0     0     0
##   EASTERN           0    23     0     0     0     0     7     0     0     0
##   NORTHEASTERN      0     6    17     1     0     0     6     0     0     0
##   NORTHERN         27     0     0     2     0     0    55     0     0     0
##   NORTHWESTERN      0     0     0    48     1     0     0     0     0     0
##   SOUTHEASTERN      0     2     0     0     0     0     0     0     7     0
##   SOUTHERN          0     0     0     0     0     0     0     1     0    24
##   SOUTHWESTERN      0     0     0     0     6     1     0     0     0    21
##   WESTERN           0     0     0     3     3    21     0     0     0    11
##
##                 21224 21225 21226 21227 21229 21230 21231 21234 21237 21239
##   CENTRAL           0     0     0     0     0     3     0     0     0     0
##   EASTERN           1     0     0     1     0     0     1     0     0     0
##   NORTHEASTERN      0     0     0     1     0     0     0     7     0     2
##   NORTHERN          0     0     0     0     0     0     0     0     0     1
##   NORTHWESTERN      0     0     0     0     0     0     0     0     0     0
##   SOUTHEASTERN    198     1     0     0     0     1   126     0     1     0
##   SOUTHERN          0    18    18     0     0   141     0     0     0     0
##   SOUTHWESTERN      0     0     0     2    13    11     0     0     0     0
##   WESTERN           0     0     0     0     0     0     0     0     0     0
##
##                 21251 21287
##   CENTRAL           0     0
##   EASTERN           0     1
##   NORTHEASTERN      2     0
##   NORTHERN          0     0
##   NORTHWESTERN      0     0
##   SOUTHEASTERN     0     0
##   SOUTHERN         0     0
##   SOUTHWESTERN     0     0
##   WESTERN          0     0
```

Check for missing values

```
sum(is.na(restData$councilDistrict)) # resituisce zero se non mancano dati
```

```
## [1] 0
```

```
# oppure
any(is.na(restData$councilDistrict))
```

```
## [1] FALSE
```

```
# oppure se ogni valore soddisfa una condizione
all(restData$zipCode > 0)
```

```
## [1] FALSE
```

Row and column sums

```
colSums(is.na(restData)) #conta gli NA
```

```
##                         name                      zipCode
##                            0                            0
##                 neighborhood              councilDistrict
##                            0                            0
##               policeDistrict                   Location.1
##                            0                            0
##   X2010.Census.Neighborhoods X2010.Census.Wards.Precincts
##                         1327                         1327
##                    Zip.Codes
##                         1327
```

```
all(colSums(is.na(restData))==0) #verifica che sono zero i Na de quella colonna
```

```
## [1] FALSE
```

Visualizza con carateristiche precise