

## 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
```

Ricerca solo determinate 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 database 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

```
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
```

```
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

```
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
```

```
# 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
## 2
## 3
```

```
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
## 1326
## 1327
```

```
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 8
## $ 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 NA ...
## $ X2010.Census.Wards.Precincts: logi  NA NA NA NA NA NA NA ...
## $ Zip.Codes      : logi  NA 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))
```

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

```
#guardo le probabilità che mi interessano
```

## Table

```
table(restData$zipCode , useNA="ifany")
```

```
##
## -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
```

*#quante volte appaiono i numeri nella tabella, aggiungo una colonna se ci sono NA per contarli*

```
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)) # restituisce 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

```
table(restData$zipCode %in% c("21212", "21213"))
```

```
##
## FALSE TRUE
## 1268 59
```

*#cerca quante volte appaiono i due valori nella colonna*

*se voglio avere solo i ristoranti con quelle caratteristiche faccio un subset*

```
restData[restData$zipCode %in% c("21212", "21213"),]
```

##		name	zipCode	neighborhood
## 29		BAY ATLANTIC CLUB	21212	Downtown
## 39		BERMUDA BAR	21213	Broadway East
## 92		ATWATER'S	21212	Chinquapin Park-Belvedere
## 111		BALTIMORE ESTONIAN SOCIETY	21213	South Clifton Park
## 187		CAFE ZEN	21212	Rosebank
## 220		CERIELLO FINE FOODS	21212	Chinquapin Park-Belvedere
## 266		CLIFTON PARK GOLF COURSE SNACK BAR	21213	Darley Park
## 276		CLUB HOUSE BAR & GRILL	21213	Orangeville Industrial Area
## 289		CLUBHOUSE BAR & GRILL	21213	Orangeville Industrial Area
## 291		COCKY LOU'S	21213	Broadway East
## 362		DREAM TAVERN, CARRIBEAN U.S.A.	21213	Broadway East
## 373		DUNKIN DONUTS	21212	Homeland
## 383		EASTSIDE SPORTS SOCIAL CLUB	21213	Broadway East
## 417		FIELDS OLD TRAIL	21212	Mid-Govans
## 475		GRAND CRU	21212	Chinquapin Park-Belvedere
## 545		RANDY'S BAR	21213	Broadway East
## 604		MURPHY'S NEIGHBORHOOD BAR & GRILL	21212	Mid-Govans
## 616		NEOPOL	21212	Chinquapin Park-Belvedere
## 620		NEW CLUB THUNDERBIRD INC.	21213	Middle East
## 626		NEW MAYFIELD, INC.	21213	Belair-Edison
## 678		IKAN SEAFOOD	21212	Chinquapin Park-Belvedere
## 711		KAY-CEE CLUB	21212	Homeland
## 763		LA'RAE	21213	Oliver
## 777		LEMONGRASS BALTIMORE	21213	Little Italy
## 779		LEN'S SANDWICH SHOP	21213	Broadway East
## 845		MCDONALD'S	21213	South Clifton Park
## 852		MCDONALD'S	21212	Radnor-Winston
## 873		NEW REX LIQUORS, INC.	21212	Wilson Park
## 895		OK TAVERN	21213	Biddle Street
## 919		PANERA BREAD	21212	Lake Walker
## 940		PEIWEI ASIAN DINER	21212	Cedarcroft
## 949		PERGUSA ENTERPRISES	21212	Rosebank
## 957		PHANTOM'S BAR AND GRILL	21213	Belair-Edison
## 976		POPEYES FAMOUS FRIED CHICKEN	21212	Winston-Govans
## 994		ROBBIE'S NEST	21213	Broadway East
## 1017		RUTLAND BAR	21213	Broadway East
## 1018		RYAN'S DAUGHTER	21212	Chinquapin Park-Belvedere
## 1022		saigon remembered restaurant	21212	Mid-Govans



## 1053	SHIRLEY'S HONEY HOLE	21213	Broadway East
## 1120	STEEPLE CHASE II	21213	Biddle Street
## 1122	SUBWAY	21213	Oliver
## 1153	TAM-TAM	21212	Mid-Govans
## 1155	TASTE	21212	Mid-Govans
## 1159	TAYLORS EAST	21213	Berea
## 1186	THE EDGE BAR & LOUNGE	21213	Broadway East
## 1187	THE EDGE BAR & LOUNGE - KITCHEN AREA	21213	Broadway East
## 1198	THE HOLLOW BAR & GRILL	21212	Rosebank
## 1209	THE NEW BUCKETT'S LOUNGE	21213	Broadway East
## 1232	THREE ACE'S	21213	Belair-Edison
## 1246	TORAIN'S HIDE-A-WAY	21213	Broadway East
## 1259	TSUNAMI BALTIMORE	21213	Little Italy
## 1287	VITO'S PIZZA	21212	Cedarcroft
## 1298	WENDY'S OLD FASHIONED HAMBURGERS #96	21212	Homeland
## 1304	WHITTEN'S (4502-04)	21213	Claremont-Freedom
## 1312	wozi lounge	21212	Guilford
## 1319	YETI RESTAURANT & CARRYOUT	21212	Rosebank
## 1320	YORK CLUB TAVERN	21212	Homeland
## 1323	ZEN WEST ROADSIDE CANTINA	21212	Rosebank
## 1325	ZINK'S CAF\u0090	21213	Belair-Edison
##	councilDistrict	policeDistrict	Location.1
## 29	11	CENTRAL	206 REDWOOD ST\nBaltimore, MD
## 39	12	EASTERN	1801 NORTH AVE\nBaltimore, MD
## 92	4	NORTHERN	529 BELVEDERE AVE\nBaltimore, MD
## 111	12	EASTERN	1932 BELAIR RD\nBaltimore, MD
## 187	4	NORTHERN	438 BELVEDERE AVE\nBaltimore, MD
## 220	4	NORTHERN	529 BELVEDERE AVE\nBaltimore, MD
## 266	14	NORTHEASTERN	2701 ST LO DR\nBaltimore, MD
## 276	13	EASTERN	4217 ERDMAN AVE\nBaltimore, MD
## 289	13	EASTERN	4217 ERDMAN AVE\nBaltimore, MD
## 291	12	EASTERN	2101 NORTH AVE\nBaltimore, MD
## 362	13	EASTERN	2300 LAFAYETTE AVE\nBaltimore, MD
## 373	4	NORTHERN	5422 YORK RD\nBaltimore, MD
## 383	13	EASTERN	1203 COLLINGTON AVE\nBaltimore, MD
## 417	4	NORTHERN	5723 YORK RD\nBaltimore, MD
## 475	4	NORTHERN	527 BELVEDERE AVE\nBaltimore, MD
## 545	12	EASTERN	2135 NORTH AVE\nBaltimore, MD
## 604	4	NORTHERN	5847 YORK RD\nBaltimore, MD
## 616	4	NORTHERN	529 BELVEDERE AVE\nBaltimore, MD
## 620	13	EASTERN	2201 CHASE ST\nBaltimore, MD
## 626	13	NORTHEASTERN	3349 BELAIR RD\nBaltimore, MD
## 678	4	NORTHERN	529 BELVEDERE AVE\nBaltimore, MD
## 711	4	NORTHERN	201 HOMELAND AVE\nBaltimore, MD
## 763	12	EASTERN	1000 HOFFMAN ST\nBaltimore, MD
## 777	1	SOUTHEASTERN	1300 BANK STREET\nBaltimore, MD
## 779	12	EASTERN	1500 WASHINGTON ST\nBaltimore, MD
## 845	12	EASTERN	2001 BROADWAY\nBaltimore, MD
## 852	4	NORTHERN	5100 YORK RD\nBaltimore, MD
## 873	4	NORTHERN	4637 YORK RD\nBaltimore, MD
## 895	13	EASTERN	2301 BIDDLE ST\nBaltimore, MD
## 919	4	NORTHERN	6307 1 2 YORK RD\nBaltimore, MD
## 940	4	NORTHERN	6302 YORK RD\nBaltimore, MD
## 949	4	NORTHERN	5928 YORK RD\nBaltimore, MD

## 957	3	NORTHEASTERN	3539 BELAIR RD\nBaltimore, MD
## 976	4	NORTHERN	5002 YORK RD\nBaltimore, MD
## 994	12	EASTERN	2250 NORTH AVE\nBaltimore, MD
## 1017	12	EASTERN	1508 RUTLAND AVE\nBaltimore, MD
## 1018	4	NORTHERN	600 BELVEDERE AVE\nBaltimore, MD
## 1022	4	NORTHERN	5857 york rd\nBaltimore, MD
## 1053	13	EASTERN	2300 OLIVER ST\nBaltimore, MD
## 1120	13	EASTERN	2401 CHASE ST\nBaltimore, MD
## 1122	12	EASTERN	1400 NORTH AVE\nBaltimore, MD
## 1153	4	NORTHERN	5722 YORK RD\nBaltimore, MD
## 1155	4	NORTHERN	510 BELVEDERE AVE\nBaltimore, MD
## 1159	13	EASTERN	1201 POTOMAC ST\nBaltimore, MD
## 1186	12	EASTERN	2015 FEDERAL ST\nBaltimore, MD
## 1187	12	EASTERN	2015 FEDERAL ST\nBaltimore, MD
## 1198	4	NORTHERN	5921 YORK RD\nBaltimore, MD
## 1209	13	EASTERN	1432 CHESTER ST\nBaltimore, MD
## 1232	3	NORTHEASTERN	3534 belair RD\nBaltimore, MD
## 1246	12	EASTERN	1701 ELLSWORTH ST\nBaltimore, MD
## 1259	1	SOUTHEASTERN	1300 BANK ST\nBaltimore, MD
## 1287	4	NORTHERN	6304 YORK RD\nBaltimore, MD
## 1298	4	NORTHERN	5615 YORK RD\nBaltimore, MD
## 1304	13	NORTHEASTERN	4502 ERDMAN AVE\nBaltimore, MD
## 1312	4	NORTHERN	4515 YORK RD\nBaltimore, MD
## 1319	4	NORTHERN	5926 YORK RD\nBaltimore, MD
## 1320	4	NORTHERN	5407 YORK RD\nBaltimore, MD
## 1323	4	NORTHERN	5916 YORK RD\nBaltimore, MD
## 1325	13	NORTHEASTERN	3300 LAWNVIEW AVE\nBaltimore, MD
##	X2010.Census.Neighborhoods X2010.Census.Wards.Precincts Zip.Codes		
## 29		NA	NA NA
## 39		NA	NA NA
## 92		NA	NA NA
## 111		NA	NA NA
## 187		NA	NA NA
## 220		NA	NA NA
## 266		NA	NA NA
## 276		NA	NA NA
## 289		NA	NA NA
## 291		NA	NA NA
## 362		NA	NA NA
## 373		NA	NA NA
## 383		NA	NA NA
## 417		NA	NA NA
## 475		NA	NA NA
## 545		NA	NA NA
## 604		NA	NA NA
## 616		NA	NA NA
## 620		NA	NA NA
## 626		NA	NA NA
## 678		NA	NA NA
## 711		NA	NA NA
## 763		NA	NA NA
## 777		NA	NA NA
## 779		NA	NA NA
## 845		NA	NA NA

## 852	NA	NA	NA
## 873	NA	NA	NA
## 895	NA	NA	NA
## 919	NA	NA	NA
## 940	NA	NA	NA
## 949	NA	NA	NA
## 957	NA	NA	NA
## 976	NA	NA	NA
## 994	NA	NA	NA
## 1017	NA	NA	NA
## 1018	NA	NA	NA
## 1022	NA	NA	NA
## 1053	NA	NA	NA
## 1120	NA	NA	NA
## 1122	NA	NA	NA
## 1153	NA	NA	NA
## 1155	NA	NA	NA
## 1159	NA	NA	NA
## 1186	NA	NA	NA
## 1187	NA	NA	NA
## 1198	NA	NA	NA
## 1209	NA	NA	NA
## 1232	NA	NA	NA
## 1246	NA	NA	NA
## 1259	NA	NA	NA
## 1287	NA	NA	NA
## 1298	NA	NA	NA
## 1304	NA	NA	NA
## 1312	NA	NA	NA
## 1319	NA	NA	NA
## 1320	NA	NA	NA
## 1323	NA	NA	NA
## 1325	NA	NA	NA

## Cross tabs

```
data("UCBAdmissions")
df = as.data.frame(UCBAdmissions) # creo il database
summary(df)
```

##	Admit	Gender	Dept	Freq
##	Admitted:12	Male :12	A:4	Min. : 8.0
##	Rejected:12	Female:12	B:4	1st Qu.: 80.0
##			C:4	Median :170.0
##			D:4	Mean :188.6
##			E:4	3rd Qu.:302.5
##			F:4	Max. :512.0

*cerco delle relazioni nei data*

```
xt <- xtabs(Freq ~ Gender + Admit, data= df )
#freq è la variabile che voglio far vedere nella tabella
# gli altre sono le categorie
xt
```

```
##           Admit
## Gender   Admitted Rejected
##   Male         1198       1493
##   Female        557       1278
```

## Creating sequences

```
s1 <- seq(1,10,by=2); s1
```

```
## [1] 1 3 5 7 9
```

```
s2 <- seq(1,10,length=3); s2
```

```
## [1] 1.0 5.5 10.0
```

## Subsetting variables

```
restData$nearme = restData$neighborhood %in% c("Roland Park", "Homeland")
table(restData$nearme)
```

```
##
## FALSE  TRUE
##  1314    13
```

```
#cerco tutti i ristoranti che ci trovano vicino a Roland e Homeland
```

## Creating binary variables

```
restData$zipWrong = ifelse(restData$zipCode < 0, TRUE, FALSE)
table(restData$zipWrong, restData$zipCode < 0)
```

```
##
##           FALSE TRUE
##   FALSE  1326    0
##   TRUE    0     1
```

## Creating factor variables

```
restData$zcf <- factor(restData$zipCode)
#crea dei livelli
restData$zcf[1:3]
```

```
## [1] 21206 21231 21224
## 32 Levels: -21226 21201 21202 21205 21206 21207 21208 21209 21210 ... 21287
```

```
class(restData$zcf)# guardo di che classe è
```

```
## [1] "factor"
```

## Levels of factor variables

```
yesno <- sample(c("yes","no"), size = 10, replace=TRUE)
yesnofac = factor(yesno, levels=c("yes","no"))
relevel(yesnofac, ref="yes")
```

```
## [1] no yes yes no yes yes no yes no no
## Levels: yes no
```

## Reshape

```
library(reshape2)
head(mtcars)
```

```
##           mpg  cyl  disp  hp  drat    wt  qsec vs  am  gear  carb
## Mazda RX4      21.0   6  160 110 3.90 2.620 16.46 0   1    4    4
## Mazda RX4 Wag  21.0   6  160 110 3.90 2.875 17.02 0   1    4    4
## Datsun 710     22.8   4  108  93 3.85 2.320 18.61 1   1    4    1
## Hornet 4 Drive  21.4   6  258 110 3.08 3.215 19.44 1   0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02 0   0    3    2
## Valiant        18.1   6  225 105 2.76 3.460 20.22 1   0    3    1
```

## melting

```
#renderli più leggibili
mtcars$carname <- rownames(mtcars)
carMelt <- melt(mtcars, id=c("carname","gear","cyl"), measure.vars = c("mpg","hp")) # divido in varibili
head(carMelt)
```

```
##           carname gear cyl variable value
## 1      Mazda RX4    4   6      mpg    21.0
```

```
## 2      Mazda RX4 Wag      4  6      mpg  21.0
## 3          Datsun 710      4  4      mpg  22.8
## 4      Hornet 4 Drive      3  6      mpg  21.4
## 5 Hornet Sportabout      3  8      mpg  18.7
## 6          Valiant        3  6      mpg  18.1
```

```
tail(carMelt)
```

```
##           carname gear cyl variable value
## 59   Porsche 914-2    5  4      hp      91
## 60    Lotus Europa    5  4      hp     113
## 61 Ford Pantera L     5  8      hp     264
## 62   Ferrari Dino     5  6      hp     175
## 63 Maserati Bora      5  8      hp     335
## 64    Volvo 142E      4  4      hp     109
```

## Casting dataframes

```
cylData <- dcast(carMelt, cyl ~ variable )
```

```
## Aggregation function missing: defaulting to length
```

```
cylData
```

```
##   cyl mpg hp
## 1   4  11 11
## 2   6   7  7
## 3   8  14 14
```

```
cylData <- dcast(carMelt, cyl ~ variable, mean)
cylData
```

```
##   cyl      mpg      hp
## 1   4 26.66364 82.63636
## 2   6 19.74286 122.28571
## 3   8 15.10000 209.21429
```

## Managing Dataframes with dplyr

```
#chicago <- readRDS("chicago.rds") erroe
```

## Merging dat

```
df1=data.frame(id=sample(1:10), x=rnorm(10))
df2=data.frame(id=sample(1:10), y=rnorm(10))
dflist = list(df1, df2)
library(plyr)
join_all(dflist)
```

```
## Joining by: id
```

```
##      id      x      y
## 1    7 0.4906615 0.4868619
## 2    3 1.4912935 -0.3144474
## 3    5 -0.1842727 -0.4993640
## 4    4 0.5127249 0.5745255
## 5   10 -0.9409096 0.3654714
## 6    9 -0.3808710 0.3121766
## 7    2 0.1015107 1.3186876
## 8    8 0.1731997 -0.8792612
## 9    6 -0.4339347 0.2910936
## 10   1 -0.1505799 1.3030966
```