

```

library(MASS)
library(help=MASS)          # functions and datasets in MASS - also go online

d1=Cars93
str(d1)
dim(d1)
n = nrow(d1)

# SUBSETTING
#=====
d2 = d1[c(1:10),c(2,4,9)]
d2
#      Model Min.Price      AirBags
#1   Integra    12.9          None
#2   Legend    29.2 Driver & Passenger
#3     90     25.9   Driver only
#4    100    30.8 Driver & Passenger
#5   535i     23.7   Driver only
#6  Century    14.2   Driver only
#7  LeSabre    19.9   Driver only
#8 Roadmaster    22.6   Driver only
#9  Riviera    26.3   Driver only
#10 DeVille    33.0   Driver only

d2 = d1[c(1:10),]
d2 = d1[,c(2,4,9)]

# Manufacturers and Prices
d2 = data.frame(d1$Manufacturer,d1$Price)
head(d2)
  d1.Manufacturer d1.Price
1         Acura    15.9
2         Acura    33.9
3          Audi    29.1
4          Audi    37.7
5          BMW    30.0
6         Buick    15.7

d2 = subset(d1,select=c(Manufacturer,Price))    # must use select explicitly

# Ford cars
d2 = subset(d1,subset = Manufacturer=="Ford")
d2 = subset(d1,Price,subset = Manufacturer=="Ford")    # only prices
d2 = subset(d1,c(Manufacturer,Price),subset = Manufacturer=="Ford")

# Ford and Nissan cars
d2 = subset(d1,subset=Manufacturer=="Ford"|Manufacturer=="Nissan")
d2 = subset(d1,c(Manufacturer,Price),subset=Manufacturer=="Ford"|Manufacturer=="Nissan")

```

```

# cars weighting > 3500
d2 = d1[d1$Weight>4000,]    # there are 4

# COUNTING (by factor levels)
#=====
# how many exceeding 3000 lbs?
aux = d1$Weight
cars1=aux[aux>3000]
length(cars1)    # [1] 48

# number of cars by DriveTrain?
table(d1$DriveTrain)
# 4WD Front  Rear
# 10      67    16

# relative freq
prop.table(table(d1$DriveTrain))
#      4WD      Front      Rear
# 0.1075269 0.7204301 0.1720430

# by two factors
table(d1$AirBags,d1$DriveTrain)
#
#      4WD Front Rear
# Driver & Passenger  0   11   5
# Driver only        5   28  10
# None               5   28   1

# how many cars by AirBags & DriveTrain & Passengers?
ftable(d1$AirBags,d1$DriveTrain,d1$Passengers)
#
#      2  4  5  6  7  8
#Driver & Passenger 4WD    0  0  0  0  0  0
#
#      Front    0  2  3  6  0  0
#
#      Rear    0  3  1  1  0  0
#Driver only      4WD    0  1  1  0  3  0
#
#      Front    0  5 16  7  0  0
#
#      Rear    2  2  3  3  0  0
#None            4WD    0  2  1  0  1  1
#
#      Front    0  8 15  1  4  0
#
#      Rear    0  0  1  0  0  0

# MEASURING
#=====
# median weight per DriveTrain
aux1=tapply(d1$Weight,d1$DriveTrain,median)
# 4WD Front  Rear
# 3720 2910 3520
sort1=aux1[order(aux1)]    # in ascending order

```

```

# relative freq
rel1=prop.table(aux1)
#      4WD      Front      Rear
# 0.3665025 0.2866995 0.3467980

# median weight per Airbags & DriveTrain
aux = list(d1$AirBags,d1$DriveTrain)
tapply(d1$Weight,aux,median) # factors in a list()
#           4WD Front Rear
#Driver & Passenger   NA 3490.0 3515
#Driver only         3735 2970.0 3510
#None                2640 2552.5 3610

# change NA to 0
m1=tapply(d1$Weight,aux,median)
m1[is.na(m1)]=0
m1
#           4WD Front Rear
# Driver & Passenger    0 3490.0 3515
# Driver only         3735 2970.0 3510
# None                2640 2552.5 3610

# SORTING
#=====

d2 = subset(d1,select=c(Manufacturer,Price,Weight,Width))
head(d2)
# Manufacturer Price Weight Width
#1      Acura  15.9   2705    68
#2      Acura  33.9   3560    71
#3      Audi   29.1   3375    67
#4      Audi   37.7   3405    70
#5      BMW    30.0   3640    69
#6      Buick  15.7   2880    69

# sort Width
d3 = d2[order(d2$Width),]
head(d3)
# Manufacturer Price Weight Width
#80     Subaru   8.4   2045    60
#31      Ford   7.4   1845    63
#39      Geo    8.4   1695    63
#44     Hyundai  8.0   2345    63
#83      Suzuki  8.6   1965    63
#88     Volkswagen  9.1   2240    63
tail(d3)
# Manufacturer Price Weight Width
#75     Pontiac  17.7   3240    75
#18     Chevrolet 18.8   3910    77

```

```
#52      Lincoln 36.1 4055 77
#8       Buick  23.7 4105 78
#17     Chevrolet 16.6 4025 78
#38       Ford  20.9 3950 78
```

```
# sort by Width and break ties by Weight
d3 = d2[order(d2$Width,d2$Weight),]
head(d3)
#   Manufacturer Price Weight Width
#80      Subaru   8.4   2045    60
#39        Geo   8.4   1695    63
#31        Ford   7.4   1845    63
#83      Suzuki   8.6   1965    63
#88 Volkswagen   9.1   2240    63
#44      Hyundai   8.0   2345    63
```

```
# NAs
#=====
# rows with NAs
totals = rowSums(is.na(d1))
totals[totals>0]
#16 17 19 26 36 56 57 66 70 87 89
# 1  1  2  1  1  1  2  1  1  1  1
```

```
d2 = d1[totals > 0,]
rownames(d2)
# "16" "17" "19" "26" "36" "56" "57" "66" "70" "87" "89"
```

```
# rows excluding NAs
index = as.integer(rownames(d2)) # 16 17 19 26 36 56 57 66 70 87 89
d3 = d1[-index,] # wo NAs
totals2 = rowSums(is.na(d3))
totals2[totals2>0]
# named numeric(0)
```

```
# Choose 4 cars at random
#=====
set.seed(1928)
x = nrow(d1)
idx = sample(x,4) # [1] 63 51 36 41
d1[idx,1:9]
#   Manufacturer      Model      Type Min.Price Price Max.Price MPG.city MPG.highway      AirBag
#63 Mitsubishi    Diamante Midsize    22.4  26.1    29.9    18        24      Driver onl
#51 Lincoln Continental Midsize    33.3  34.3    35.3    17        26 Driver & Passenge
#36 Ford      Aerostar   Van    14.5  19.9    25.3    15        20      Driver onl
#41 Honda      Prelude   Sporty 17.0  19.8    22.7    24        31 Driver & Passenge
```