

R Notebook

Title: "IST687 – Writing functions and doing some initial data analysis"
Name: Sathish Kumar Rajendiran
Week: 2
Date: 04/18/2020

Exercise: Explore the mtcars dataset (which is already included in R). Copy the mtcars dataset into a new variable (called it myCars), so that if you mess up, you can start again very easily (by copying mtcars into myCars again)

```
{ }
```

Step 0: Clone the dataset into new Dataframe.

```
# clone mtcars into mycarsdf dataframe
mycarsdf <- data.frame(mtcars)
head(mycarsdf,10)
```

```
##           mpg cyl  disp  hp drat   wt  qsec vs am gear carb
## Mazda RX4      21.0   6 160.0 110 3.90 2.620 16.46 0  1   4    4
## Mazda RX4 Wag  21.0   6 160.0 110 3.90 2.875 17.02 0  1   4    4
## Datsun 710      22.8   4 108.0  93 3.85 2.320 18.61 1  1   4    1
## Hornet 4 Drive  21.4   6 258.0 110 3.08 3.215 19.44 1  0   3    1
## Hornet Sportabout 18.7   8 360.0 175 3.15 3.440 17.02 0  0   3    2
## Valiant         18.1   6 225.0 105 2.76 3.460 20.22 1  0   3    1
## Duster 360      14.3   8 360.0 245 3.21 3.570 15.84 0  0   3    4
## Merc 240D       24.4   4 146.7  62 3.69 3.190 20.00 1  0   4    2
## Merc 230        22.8   4 140.8  95 3.92 3.150 22.90 1  0   4    2
## Merc 280        19.2   6 167.6 123 3.92 3.440 18.30 1  0   4    4
```

```
mycarsdf <- cbind(name = rownames(mycarsdf), mycarsdf)
rownames(mycarsdf) <- 1:nrow(mycarsdf)
head(mycarsdf,10)
```

```
##           name  mpg cyl  disp  hp drat   wt  qsec vs am gear carb
## 1      Mazda RX4 21.0   6 160.0 110 3.90 2.620 16.46 0  1   4    4
## 2    Mazda RX4 Wag 21.0   6 160.0 110 3.90 2.875 17.02 0  1   4    4
## 3      Datsun 710 22.8   4 108.0  93 3.85 2.320 18.61 1  1   4    1
## 4    Hornet 4 Drive 21.4   6 258.0 110 3.08 3.215 19.44 1  0   3    1
## 5  Hornet Sportabout 18.7   8 360.0 175 3.15 3.440 17.02 0  0   3    2
## 6         Valiant 18.1   6 225.0 105 2.76 3.460 20.22 1  0   3    1
## 7      Duster 360 14.3   8 360.0 245 3.21 3.570 15.84 0  0   3    4
## 8      Merc 240D 24.4   4 146.7  62 3.69 3.190 20.00 1  0   4    2
## 9      Merc 230 22.8   4 140.8  95 3.92 3.150 22.90 1  0   4    2
## 10     Merc 280 19.2   6 167.6 123 3.92 3.440 18.30 1  0   4    4
```

Step 1: What is the hp (hp stands for “horse power”)

```
# 1) What is the highest hp?
maxHp <- max(mycarsdf$hp)
sprintf("%s is the highest horse power!", maxHp)
```

```
## [1] "335 is the highest horse power!"
```

```
# 2) Which car has the highest hp?
maxHp_index <- which.max(mycarsdf$hp)
print(mycarsdf[maxHp_index,1],max.levels = 0)
```

```
## [1] Maserati Bora
```

```
#Maserati Bora
```

Step 2: Explore mpg (mpg stands for “miles per gallon”)

```
# 3) What is the highest mpg?

maxMPG <- max(mycarsdf$mpg)
sprintf("%s is the highest mpg!", maxMPG)
```

```
## [1] "33.9 is the highest mpg!"
```

```
# 4) Which car has the highest mpg?
maxMPG_index <- which.max(mycarsdf$mpg)
print(mycarsdf[maxMPG_index,1],max.levels = 0)
```

```
## [1] Toyota Corolla
```

```
#Toyota Corolla
```

```
# 5) Create a sorted dataframe, based on mpg

carsDFbyMPG <- mycarsdf[order(-mycarsdf$mpg),]
head(carsDFbyMPG,5)
```

```
##           name  mpg cyl disp  hp drat   wt  qsec vs am gear carb
## 20 Toyota Corolla 33.9   4  71.1  65 4.22 1.835 19.90  1  1    4    1
## 18      Fiat 128  32.4   4  78.7  66 4.08 2.200 19.47  1  1    4    1
## 19   Honda Civic  30.4   4  75.7  52 4.93 1.615 18.52  1  1    4    2
## 28   Lotus Europa  30.4   4  95.1 113 3.77 1.513 16.90  1  1    5    2
## 26      Fiat X1-9  27.3   4  79.0  66 4.08 1.935 18.90  1  1    4    1
```

```
tail(carsDFbyMPG,5)
```

```
##           name  mpg cyl disp  hp drat   wt  qsec vs am gear carb
## 17 Chrysler Imperial 14.7   8  440 230 3.23 5.345 17.42  0  0    3    4
## 7         Duster 360 14.3   8  360 245 3.21 3.570 15.84  0  0    3    4
## 24         Camaro Z28 13.3   8  350 245 3.73 3.840 15.41  0  0    3    4
## 15 Cadillac Fleetwood 10.4   8  472 205 2.93 5.250 17.98  0  0    3    4
## 16 Lincoln Continental 10.4   8  460 215 3.00 5.424 17.82  0  0    3    4
```

Step 3: Which car has the “best” combination of mpg and hp?

```
# 6) What logic did you use?

# lets divide hp by mpg to find the best value car
mycarsdf$hpBympg <- mycarsdf$hp/mycarsdf$mpg

# re-arrange columns
bestcarsDF <- mycarsdf[,c("name", "mpg", "hp", "hpBympg", "cyl", "disp", "drat", "wt", "qsec", "vs", "am", "gear")]
bestcarsDF
```

	name	mpg	hp	hpBympg	cyl	disp	drat	wt	qsec	vs	am	gear
## 1	Mazda RX4	21.0	110	5.238095	6	160.0	3.90	2.620	16.46	0	1	4
## 2	Mazda RX4 Wag	21.0	110	5.238095	6	160.0	3.90	2.875	17.02	0	1	4
## 3	Datsun 710	22.8	93	4.078947	4	108.0	3.85	2.320	18.61	1	1	4
## 4	Hornet 4 Drive	21.4	110	5.140187	6	258.0	3.08	3.215	19.44	1	0	3
## 5	Hornet Sportabout	18.7	175	9.358289	8	360.0	3.15	3.440	17.02	0	0	3
## 6	Valiant	18.1	105	5.801105	6	225.0	2.76	3.460	20.22	1	0	3
## 7	Duster 360	14.3	245	17.132867	8	360.0	3.21	3.570	15.84	0	0	3
## 8	Merc 240D	24.4	62	2.540984	4	146.7	3.69	3.190	20.00	1	0	4
## 9	Merc 230	22.8	95	4.166667	4	140.8	3.92	3.150	22.90	1	0	4
## 10	Merc 280	19.2	123	6.406250	6	167.6	3.92	3.440	18.30	1	0	4
## 11	Merc 280C	17.8	123	6.910112	6	167.6	3.92	3.440	18.90	1	0	4
## 12	Merc 450SE	16.4	180	10.975610	8	275.8	3.07	4.070	17.40	0	0	3
## 13	Merc 450SL	17.3	180	10.404624	8	275.8	3.07	3.730	17.60	0	0	3
## 14	Merc 450SLC	15.2	180	11.842105	8	275.8	3.07	3.780	18.00	0	0	3
## 15	Cadillac Fleetwood	10.4	205	19.711538	8	472.0	2.93	5.250	17.98	0	0	3
## 16	Lincoln Continental	10.4	215	20.673077	8	460.0	3.00	5.424	17.82	0	0	3
## 17	Chrysler Imperial	14.7	230	15.646259	8	440.0	3.23	5.345	17.42	0	0	3
## 18	Fiat 128	32.4	66	2.037037	4	78.7	4.08	2.200	19.47	1	1	4
## 19	Honda Civic	30.4	52	1.710526	4	75.7	4.93	1.615	18.52	1	1	4
## 20	Toyota Corolla	33.9	65	1.917404	4	71.1	4.22	1.835	19.90	1	1	4
## 21	Toyota Corona	21.5	97	4.511628	4	120.1	3.70	2.465	20.01	1	0	3
## 22	Dodge Challenger	15.5	150	9.677419	8	318.0	2.76	3.520	16.87	0	0	3
## 23	AMC Javelin	15.2	150	9.868421	8	304.0	3.15	3.435	17.30	0	0	3
## 24	Camaro Z28	13.3	245	18.421053	8	350.0	3.73	3.840	15.41	0	0	3
## 25	Pontiac Firebird	19.2	175	9.114583	8	400.0	3.08	3.845	17.05	0	0	3
## 26	Fiat X1-9	27.3	66	2.417582	4	79.0	4.08	1.935	18.90	1	1	4
## 27	Porsche 914-2	26.0	91	3.500000	4	120.3	4.43	2.140	16.70	0	1	5
## 28	Lotus Europa	30.4	113	3.717105	4	95.1	3.77	1.513	16.90	1	1	5
## 29	Ford Pantera L	15.8	264	16.708861	8	351.0	4.22	3.170	14.50	0	1	5
## 30	Ferrari Dino	19.7	175	8.883249	6	145.0	3.62	2.770	15.50	0	1	5
## 31	Maserati Bora	15.0	335	22.333333	8	301.0	3.54	3.570	14.60	0	1	5
## 32	Volvo 142E	21.4	109	5.093458	4	121.0	4.11	2.780	18.60	1	1	4
##	carb											
## 1	4											
## 2	4											
## 3	1											
## 4	1											
## 5	2											
## 6	1											
## 7	4											
## 8	2											
## 9	2											

```
## 10    4
## 11    4
## 12    3
## 13    3
## 14    3
## 15    4
## 16    4
## 17    4
## 18    1
## 19    2
## 20    1
## 21    1
## 22    2
## 23    2
## 24    4
## 25    2
## 26    1
## 27    2
## 28    2
## 29    4
## 30    6
## 31    8
## 32    2
```

```
# 7) Which car?
# find the car that has best of value
bestcar_index <- which.max(bestcarsDF$hpBypmg)
print(bestcarsDF[bestcar_index,1],max.levels = 0)
```

```
## [1] Maserati Bora
```

Step 4: Which car has “best” car combination of mpg and hp, where mpg and hp must be given equal weight?

```
bestcarsDF$scalempg <- scale(bestcarsDF$mpg)
bestcarsDF$scalehp <- scale(bestcarsDF$hp)
cols <- c('scalempg','scalehp')
bestcarsDF$mostvalued <- apply(bestcarsDF[,cols],1,sum)

finaldf <- bestcarsDF[c("name","mpg","hp","hpBypmg","scalempg","scalehp","mostvalued")]
finaldf <- finaldf[order(-finaldf$mostvalued),]

mostvalued_index <- which.max(finaldf$mostvalued)
print(finaldf[mostvalued_index,1],max.levels = 0)
```

```
## [1] Maserati Bora
```

```
#Maserati Bora
```