Data Frame Slicing

HDS, Mike Kozlowski

2/2/2022

updated 1/31/2023

Slicing or indexing a data frame

There are a set of useful examples of working with data frames, the ideas are covered in book and in the swirl exercises, but let's look at these using the built-in data set mtcars, data from a 1974 motor trend puplication

```
data(mtcars)

summary(mtcars)
```

```
##       mpg             cyl             disp             hp
##  Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0
##  1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5
##  Median :19.20   Median :6.000   Median :196.3   Median :123.0
##  Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7
##  3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0
##  Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0
##       drat             wt             qsec             vs
##  Min.   :2.760   Min.   :1.513   Min.   :14.50   Min.   :0.0000
##  1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000
##  Median :3.695   Median :3.325   Median :17.71   Median :0.0000
##  Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375
##  3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000
##  Max.   :4.930   Max.   :5.424   Max.   :22.90   Max.   :1.0000
##       am             gear             carb
##  Min.   :0.0000   Min.   :3.000   Min.   :1.000
##  1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000
##  Median :0.0000   Median :4.000   Median :2.000
##  Mean   :0.4062   Mean   :3.688   Mean   :2.812
##  3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000
##  Max.   :1.0000   Max.   :5.000   Max.   :8.000
```

For each example bit of code, run it, and then above it, state what the code is doing and how you might make use of it.

Suggesting a change in the code, and run it, making sure the change did what you expected

indexing rows and columns Code is displaying the first 4 rows with all columns associated with that data. Changing it to 1:10 will result in the first 10 rows being displayed.

```
mtcars[1:10,]
```

```
##                     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
```

#Question: What does the code below do? This code displays the first 4 rows with only the first 3 columns associated with that data.By changing the 1:3 to 4:6, we will see columns 4-6 instead of 1-3.

```
mtcars[1:4,4:6]
```

```
##                    hp drat    wt
## Mazda RX4         110 3.90 2.620
## Mazda RX4 Wag     110 3.90 2.875
## Datsun 710         93 3.85 2.320
## Hornet 4 Drive    110 3.08 3.215
```

```
min_mpg=min(mtcars$mpg)

car_min_mpg=which(mtcars$mpg==min_mpg)

mtcars[car_min_mpg,]
```

```
##                      mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Cadillac Fleetwood  10.4   8  472 205 2.93 5.250 17.98  0  0    3    4
## Lincoln Continental 10.4   8  460 215 3.00 5.424 17.82  0  0    3    4
```

#Question/Action

Change the code above to find the car with the worst mileage

```
unique(mtcars$cyl)
```

```
## [1] 6 4 8
```

#Question/Action

How many different numbers of cylinders are found in this set of cars? 3 different numbers of cylinders

Selecting based on a condition Changed it to 6 to select the 6 cylinder vehicles

```
print(mtcars[mtcars$cyl==6,])
```

```
##                      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
## Hornet 4 Drive     21.4   6 258.0 110 3.08 3.215 19.44  1  0    3    1
## Valiant            18.1   6 225.0 105 2.76 3.460 20.22  1  0    3    1
## Merc 280           19.2   6 167.6 123 3.92 3.440 18.30  1  0    4    4
## Merc 280C          17.8   6 167.6 123 3.92 3.440 18.90  1  0    4    4
## Ferrari Dino       19.7   6 145.0 175 3.62 2.770 15.50  0  1    5    6
```

#Question

In the cell below, what is happening Apply will apply the function in the third field onto the data set in the first field, with 2 in the second field indicating that it is applying to the columns, 1 would indicate rows.

```
apply(mtcars,2,mean)
```

```
##        mpg        cyl       disp         hp       drat         wt       qsec
##  20.090625   6.187500 230.721875 146.687500   3.596563   3.217250  17.848750
##         vs         am       gear       carb
##   0.437500   0.406250   3.687500   2.812500
```

#This is a table calculation, what is it telling you?

What do the entries in this table show you?

This is also called a "pivot table" or "cross table"

This table calculation shows how many entries in the table include both values from both entries, so how many vehicles with 3 gears and 4 cylinders, 6 cylinders, and 8 cylinders, vehicles with 4 gears and 4,6,8 cylinders, etc.

```
table(mtcars$gear,mtcars$cyl)
```

```
##
##      4  6  8
##   3  1  2 12
##   4  8  4  0
##   5  2  1  2
```

Question: How do these tables differ from the one above

The first one shows the average horsepower for the different cylinder options, and the second one shows the average horsepower for the different cylinder options given the number of gears that cylinder has.

```
tapply(mtcars$hp,c(as.factor(mtcars$cyl)),mean)
```

```
##         4         6         8
##  82.63636 122.28571 209.21429
```

```
tapply(mtcars$hp,list(as.factor(mtcars$cyl),as.factor(mtcars$gear)),mean)
```

```
##         3     4     5
## 4  97.0000  76.0 102.0
## 6 107.5000 116.5 175.0
## 8 194.1667    NA 299.5
```

#Questions:

Which of the 4 cylinder cars has the highest horsepower? Lotus Europa

```
max(mtcars$hp[mtcars$cyl==4])
```

```
## [1] 113
```

```
mtcars[mtcars$cyl==4&mtcars$hp==113,]
```

```
##                 mpg cyl disp  hp drat    wt qsec vs am gear carb
## Lotus Europa 30.4   4 95.1 113 3.77 1.513 16.9  1  1    5    2
```

Compute the mean horse power per number of gears

```
tapply(mtcars$hp,c(as.factor(mtcars$gear)),mean)
```

```
##         3        4        5
## 176.1333  89.5000 195.6000
```