week2 R md practice

Shibo

1/13/2020

Big Section

subsection

sub-subection

Load packages

```r
library(dplyr)
library(ggplot2)
```

Practice dataset

work with the `mtcars` dataset.
New line of text, after ending the line with 2 spaecs

```r
mtcars %>% glimpse
```

```
## Observations: 32
## Variables: 11
## $ mpg  <dbl> 21.0, 21.0, 22.8, 21.4, 18.7, 18.1, 14.3, 24.4, 22.8, 19.2, 17...
## $ cyl  <dbl> 6, 6, 4, 6, 8, 6, 8, 4, 4, 6, 6, 8, 8, 8, 8, 8, 8, 4, 4, 4, 4,...
## $ disp <dbl> 160.0, 160.0, 108.0, 258.0, 360.0, 225.0, 360.0, 146.7, 140.8,...
## $ hp   <dbl> 110, 110, 93, 110, 175, 105, 245, 62, 95, 123, 123, 180, 180, ...
## $ drat <dbl> 3.90, 3.90, 3.85, 3.08, 3.15, 2.76, 3.21, 3.69, 3.92, 3.92, 3....
## $ wt   <dbl> 2.620, 2.875, 2.320, 3.215, 3.440, 3.460, 3.570, 3.190, 3.150,...
## $ qsec <dbl> 16.46, 17.02, 18.61, 19.44, 17.02, 20.22, 15.84, 20.00, 22.90,...
## $ vs   <dbl> 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1,...
## $ am   <dbl> 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0,...
## $ gear <dbl> 4, 4, 4, 3, 3, 3, 3, 4, 4, 4, 4, 3, 3, 3, 3, 3, 3, 4, 4, 4, 3,...
## $ carb <dbl> 4, 4, 1, 1, 2, 1, 4, 2, 2, 4, 4, 3, 3, 3, 4, 4, 4, 1, 2, 1, 1,...
```

```r
class(mtcars$mpg)
```

```
## [1] "numeric"
```

```r
mtcars$mpg
```

```
##  [1] 21.0 21.0 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 17.8 16.4 17.3 15.2 10.4
## [16] 10.4 14.7 32.4 30.4 33.9 21.5 15.5 15.2 13.3 19.2 27.3 26.0 30.4 15.8 19.7
## [31] 15.0 21.4
```

```r
'car and make are the row names for mtcars'
```

```
## [1] "car and make are the row names for mtcars"
```

```
mtcars
```

```
##                      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
## Merc 280C           17.8   6 167.6 123 3.92 3.440 18.90  1  0    4    4
## Merc 450SE          16.4   8 275.8 180 3.07 4.070 17.40  0  0    3    3
## Merc 450SL          17.3   8 275.8 180 3.07 3.730 17.60  0  0    3    3
## Merc 450SLC         15.2   8 275.8 180 3.07 3.780 18.00  0  0    3    3
## Cadillac Fleetwood  10.4   8 472.0 205 2.93 5.250 17.98  0  0    3    4
## Lincoln Continental 10.4   8 460.0 215 3.00 5.424 17.82  0  0    3    4
## Chrysler Imperial   14.7   8 440.0 230 3.23 5.345 17.42  0  0    3    4
## Fiat 128            32.4   4  78.7  66 4.08 2.200 19.47  1  1    4    1
## Honda Civic         30.4   4  75.7  52 4.93 1.615 18.52  1  1    4    2
## Toyota Corolla      33.9   4  71.1  65 4.22 1.835 19.90  1  1    4    1
## Toyota Corona       21.5   4 120.1  97 3.70 2.465 20.01  1  0    3    1
## Dodge Challenger    15.5   8 318.0 150 2.76 3.520 16.87  0  0    3    2
## AMC Javelin         15.2   8 304.0 150 3.15 3.435 17.30  0  0    3    2
## Camaro Z28          13.3   8 350.0 245 3.73 3.840 15.41  0  0    3    4
## Pontiac Firebird    19.2   8 400.0 175 3.08 3.845 17.05  0  0    3    2
## Fiat X1-9           27.3   4  79.0  66 4.08 1.935 18.90  1  1    4    1
## Porsche 914-2       26.0   4 120.3  91 4.43 2.140 16.70  0  1    5    2
## Lotus Europa        30.4   4  95.1 113 3.77 1.513 16.90  1  1    5    2
## Ford Pantera L      15.8   8 351.0 264 4.22 3.170 14.50  0  1    5    4
## Ferrari Dino        19.7   6 145.0 175 3.62 2.770 15.50  0  1    5    6
## Maserati Bora       15.0   8 301.0 335 3.54 3.570 14.60  0  1    5    8
## Volvo 142E          21.4   4 121.0 109 4.11 2.780 18.60  1  1    4    2
```

```
class(mtcars)
```

```
## [1] "data.frame"
```

```
#pipe mtcars to tible
mtcars %>%
  tbl_df %>% class
```

```
## [1] "tbl_df"     "tbl"        "data.frame"
```

```
#colon-colon operator access a specific function from a class
mtcars %>%
  tbl_df() %>%
  tibble::rownames_to_column('my_rowname') %>% #generate a chr column
  mutate(car_name = rownames(mtcars)) #mutate is a verb that creates a new column
```

```
## # A tibble: 32 x 13
##    my_rowname   mpg   cyl  disp    hp  drat    wt  qsec    vs    am  gear  carb
##    <chr>      <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 1             21     6   160   110   3.9  2.62  16.5     0     1     4     4
```

```
##  2 2            21       6 160     110 3.9   2.88  17.0       0    1     4     4
##  3 3            22.8     4 108      93 3.85  2.32  18.6       1    1     4     1
##  4 4            21.4     6 258     110 3.08  3.22  19.4       1    0     3     1
##  5 5            18.7     8 360     175 3.15  3.44  17.0       0    0     3     2
##  6 6            18.1     6 225     105 2.76  3.46  20.2       1    0     3     1
##  7 7            14.3     8 360     245 3.21  3.57  15.8       0    0     3     4
##  8 8            24.4     4 147.     62 3.69  3.19  20         1    0     4     2
##  9 9            22.8     4 141.     95 3.92  3.15  22.9       1    0     4     2
## 10 10           19.2     6 168.    123 3.92  3.44  18.3       1    0     4     4
## # ... with 22 more rows, and 1 more variable: car_name <chr>
```

Assign out actions to a new variable or object

```r
mtcars_b <- mtcars %>%
  tbl_df() %>%
  tibble::rowid_to_column() %>%
  tibble::rownames_to_column('my_rowname') %>%
  mutate(car_name = rownames(mtcars))
```

select() verb, filter(), desc, arrange, distinct, count and summarise verbs

```r
mtcars_b %>%
  dplyr::select(car_name, my_rowname, rowid) %>%
  select(c(1, 3)) %>%
  select(ends_with('e'))
```

```
## # A tibble: 32 x 1
##    car_name
##    <chr>
##  1 Mazda RX4
##  2 Mazda RX4 Wag
##  3 Datsun 710
##  4 Hornet 4 Drive
##  5 Hornet Sportabout
##  6 Valiant
##  7 Duster 360
##  8 Merc 240D
##  9 Merc 230
## 10 Merc 280
## # ... with 22 more rows
```

```r
mtcars_b %>%
  filter(mpg > 11 & disp > 000170.5)
```

```
## # A tibble: 14 x 14
##    my_rowname rowid   mpg   cyl  disp    hp  drat    wt  qsec    vs    am  gear
##    <chr>      <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
##  1 4              4  21.4     6   258   110  3.08  3.22  19.4     1     0     3
##  2 5              5  18.7     8   360   175  3.15  3.44  17.0     0     0     3
##  3 6              6  18.1     6   225   105  2.76  3.46  20.2     1     0     3
##  4 7              7  14.3     8   360   245  3.21  3.57  15.8     0     0     3
##  5 12            12  16.4     8   276.  180  3.07  4.07  17.4     0     0     3
##  6 13            13  17.3     8   276.  180  3.07  3.73  17.6     0     0     3
##  7 14            14  15.2     8   276.  180  3.07  3.78  18       0     0     3
##  8 17            17  14.7     8   440   230  3.23  5.34  17.4     0     0     3
##  9 22            22  15.5     8   318   150  2.76  3.52  16.9     0     0     3
## 10 23            23  15.2     8   304   150  3.15  3.44  17.3     0     0     3
```

```
## 11 24               24  13.3    8  350    245  3.73  3.84  15.4      0    0    3
## 12 25               25  19.2    8  400    175  3.08  3.84  17.0      0    0    3
## 13 29               29  15.8    8  351    264  4.22  3.17  14.5      0    1    5
## 14 31               31  15      8  301    335  3.54  3.57  14.6      0    1    5
## # ... with 2 more variables: carb <dbl>, car_name <chr>
```

```
mtcars_b %>%
  arrange(desc(mpg))
```

```
## # A tibble: 32 x 14
##    my_rowname rowid   mpg   cyl  disp    hp  drat    wt  qsec    vs    am  gear
##    <chr>      <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
##  1 20            20  33.9     4  71.1    65  4.22  1.84  19.9     1     1     4
##  2 18            18  32.4     4  78.7    66  4.08  2.2   19.5     1     1     4
##  3 19            19  30.4     4  75.7    52  4.93  1.62  18.5     1     1     4
##  4 28            28  30.4     4  95.1   113  3.77  1.51  16.9     1     1     5
##  5 26            26  27.3     4  79      66  4.08  1.94  18.9     1     1     4
##  6 27            27  26       4 120.     91  4.43  2.14  16.7     0     1     5
##  7 8              8  24.4     4 147.     62  3.69  3.19  20       1     0     4
##  8 3              3  22.8     4 108      93  3.85  2.32  18.6     1     1     4
##  9 9              9  22.8     4 141.     95  3.92  3.15  22.9     1     0     4
## 10 21            21  21.5     4 120.     97  3.7   2.46  20.0     1     0     3
## # ... with 22 more rows, and 2 more variables: carb <dbl>, car_name <chr>
```

```
mtcars_b %>%
  distinct(cyl) %>%
  arrange(desc(cyl))
```

```
## # A tibble: 3 x 1
##     cyl
##   <dbl>
## 1     8
## 2     6
## 3     4
```

```
mtcars_b %>%
  group_by(cyl) %>%
  summarise(num_rows = n(), avg_mpg = mean(mpg)) #summarise the dataset based on the group_by performed
```

```
## # A tibble: 3 x 3
##     cyl num_rows avg_mpg
##   <dbl>    <int>   <dbl>
## 1     4       11    26.7
## 2     6        7    19.7
## 3     8       14    15.1
```