`mtcars` data exploration

Rachel Han

15/09/2019

mtcars as tibble

mtcars data set is of type data.frame, not a tibble. A tibble lets us see the data without printing the entire data set to the console. An example is shown below:

```
mtcars_tibble <-as_tibble(mtcars)
head(mtcars_tibble)
```

```
## # A tibble: 6 x 11
##     mpg   cyl  disp    hp  drat    wt  qsec    vs    am
##   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1  21       6   160   110  3.9   2.62  16.5     0     1
## 2  21       6   160   110  3.9   2.88  17.0     0     1
## 3  22.8     4   108    93  3.85  2.32  18.6     1     1
## 4  21.4     6   258   110  3.08  3.22  19.4     1     0
## 5  18.7     8   360   175  3.15  3.44  17.0     0     0
## 6  18.1     6   225   105  2.76  3.46  20.2     1     0
```

Dimensions of data

Let's see how many rows and columns are there:

```
rownum <- nrow(mtcars_tibble)
colnum <- ncol(mtcars_tibble)
cat("row:",rownum)
```

```
## row: 32
```

```
cat("column:",colnum)
```

```
## column: 11
```

Glimpse

We can use the `glimpse` function to see the glimpse of the data. Notice that the data is transposed upon applying this function: the column names are now in rows.

```
glimpse(mtcars_tibble)
```

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

Row names

Notice that by changing the data type to tibble, the row names of `mtcars` are not the strings of car names, but numbers representing the row.

Originally, the row names are:

```
rownames(mtcars)
```

```
##  [1] "Mazda RX4"           "Mazda RX4 Wag"       "Datsun
##  [4] "Hornet 4 Drive"      "Hornet Sportabout"   "Valian
##  [7] "Duster 360"          "Merc 240D"           "Merc 2
## [10] "Merc 280"            "Merc 280C"           "Merc 4
## [13] "Merc 450SL"          "Merc 450SLC"         "Cadill
## [16] "Lincoln Continental" "Chrysler Imperial"   "Fiat 1
## [19] "Honda Civic"         "Toyota Corolla"      "Toyota
## [22] "Dodge Challenger"    "AMC Javelin"         "Camaro
## [25] "Pontiac Firebird"    "Fiat X1-9"           "Porsch
## [28] "Lotus Europa"        "Ford Pantera L"      "Ferrar
## [31] "Maserati Bora"       "Volvo 142E"
```

After conversion:

```
rownames(mtcars_tibble)
```

```
##  [1] "1"  "2"  "3"  "4"  "5"  "6"  "7"  "8"  "9"  "10" "
## [15] "15" "16" "17" "18" "19" "20" "21" "22" "23" "24" "
## [29] "29" "30" "31" "32"
```

Summary

We can use the function summary to get some basic statistical information about the data. Here, we just select the first five columns.

```
summary(mtcars_tibble[,1:5])
```

```
##       mpg             cyl             disp              h
##  Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.
##  1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.
##  Median :19.20   Median :6.000   Median :196.3   Median
##  Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean
##  3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.
##  Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.
##       drat
##  Min.   :2.760
##  1st Qu.:3.080
##  Median :3.695
##  Mean   :3.597
##  3rd Qu.:3.920
```