`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[,1:5])
```

```
## # A tibble: 6 x 5
##     mpg   cyl  disp    hp  drat
##   <dbl> <dbl> <dbl> <dbl> <dbl>
## 1  21       6   160   110  3.9
## 2  21       6   160   110  3.9
## 3  22.8     4   108    93  3.85
## 4  21.4     6   258   110  3.08
## 5  18.7     8   360   175  3.15
## 6  18.1     6   225   105  2.76
```

For the purpose of the presentation, we will use a smaller subset of the data.

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[1:4,1:4])
```

```
## Observations: 4
## Variables: 4
## $ mpg  <dbl> 21.0, 21.0, 22.8, 21.4
## $ cyl  <dbl> 6, 6, 4, 6
## $ disp <dbl> 160, 160, 108, 258
## $ hp   <dbl> 110, 110, 93, 110
```

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[1:3,])
```

```
## [1] "Mazda RX4"      "Mazda RX4 Wag" "Datsun 710"
```

After conversion:

```
rownames(mtcars_tibble[1:3,])
```

```
## [1] "1" "2" "3"
```

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[,1:3])
```

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