Data Wrangling in R: Additional Examples with Answers (Part 2)

Use the `mtcars` dataset and answer the following questions:

Exercise 1

Print out the *hp* column using the `select` function.

```
mtcars %>% select(hp)
```

```
##                     hp
## Mazda RX4          110
## Mazda RX4 Wag      110
## Datsun 710          93
## Hornet 4 Drive     110
## Hornet Sportabout  175
## Valiant            105
## Duster 360         245
## Merc 240D           62
## Merc 230            95
## Merc 280           123
## Merc 280C          123
## Merc 450SE         180
## Merc 450SL         180
## Merc 450SLC        180
## Cadillac Fleetwood 205
## Lincoln Continental 215
## Chrysler Imperial  230
## Fiat 128            66
## Honda Civic         52
## Toyota Corolla      65
## Toyota Corona       97
## Dodge Challenger   150
## AMC Javelin        150
## Camaro Z28         245
## Pontiac Firebird   175
## Fiat X1-9           66
## Porsche 914-2       91
## Lotus Europa       113
## Ford Pantera L     264
## Ferrari Dino       175
## Maserati Bora      335
## Volvo 142E         109
```

Exercise 2

Print out the all **but** *hp* column using the `select` function.

```
mtcars %>% select(-hp)# just to practice the concept of piping %>%
```

or

```
select(mtcars ,-hp)
```

```
##                    mpg cyl  disp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6 160.0 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6 160.0 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4 108.0 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6 258.0 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8 360.0 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6 225.0 2.76 3.460 20.22  1  0    3    1
## Duster 360        14.3   8 360.0 3.21 3.570 15.84  0  0    3    4
## Merc 240D         24.4   4 146.7 3.69 3.190 20.00  1  0    4    2
## Merc 230          22.8   4 140.8 3.92 3.150 22.90  1  0    4    2
## Merc 280          19.2   6 167.6 3.92 3.440 18.30  1  0    4    4
## Merc 280C         17.8   6 167.6 3.92 3.440 18.90  1  0    4    4
## Merc 450SE        16.4   8 275.8 3.07 4.070 17.40  0  0    3    3
## Merc 450SL        17.3   8 275.8 3.07 3.730 17.60  0  0    3    3
## Merc 450SLC       15.2   8 275.8 3.07 3.780 18.00  0  0    3    3
## Cadillac Fleetwood  10.4   8 472.0 2.93 5.250 17.98  0  0    3    4
## Lincoln Continental 10.4   8 460.0 3.00 5.424 17.82  0  0    3    4
## Chrysler Imperial   14.7   8 440.0 3.23 5.345 17.42  0  0    3    4
## Fiat 128          32.4   4  78.7 4.08 2.200 19.47  1  1    4    1
## Honda Civic       30.4   4  75.7 4.93 1.615 18.52  1  1    4    2
## Toyota Corolla    33.9   4  71.1 4.22 1.835 19.90  1  1    4    1
## Toyota Corona     21.5   4 120.1 3.70 2.465 20.01  1  0    3    1
## Dodge Challenger  15.5   8 318.0 2.76 3.520 16.87  0  0    3    2
## AMC Javelin       15.2   8 304.0 3.15 3.435 17.30  0  0    3    2
## Camaro Z28        13.3   8 350.0 3.73 3.840 15.41  0  0    3    4
## Pontiac Firebird  19.2   8 400.0 3.08 3.845 17.05  0  0    3    2
## Fiat X1-9         27.3   4  79.0 4.08 1.935 18.90  1  1    4    1
## Porsche 914-2     26.0   4 120.3 4.43 2.140 16.70  0  1    5    2
## Lotus Europa      30.4   4  95.1 3.77 1.513 16.90  1  1    5    2
## Ford Pantera L    15.8   8 351.0 4.22 3.170 14.50  0  1    5    4
## Ferrari Dino      19.7   6 145.0 3.62 2.770 15.50  0  1    5    6
## Maserati Bora     15.0   8 301.0 3.54 3.570 14.60  0  1    5    8
## Volvo 142E        21.4   4 121.0 4.11 2.780 18.60  1  1    4    2
```

Exercise 3

Print out the *mpg, hp, vs, am, gear* columns. Consider using the colon (:) symbol.

```
select(mtcars, mpg, hp, vs:gear)
```

or

```
mtcars %>% select(mpg, hp, vs:gear)
```

```
##                    mpg  hp vs am gear
## Mazda RX4         21.0 110  0  1    4
## Mazda RX4 Wag     21.0 110  0  1    4
## Datsun 710        22.8  93  1  1    4
## Hornet 4 Drive    21.4 110  1  0    3
## Hornet Sportabout 18.7 175  0  0    3
## Valiant           18.1 105  1  0    3
## Duster 360        14.3 245  0  0    3
## Merc 240D         24.4  62  1  0    4
## Merc 230          22.8  95  1  0    4
## Merc 280          19.2 123  1  0    4
## Merc 280C         17.8 123  1  0    4
## Merc 450SE        16.4 180  0  0    3
## Merc 450SL        17.3 180  0  0    3
## Merc 450SLC       15.2 180  0  0    3
## Cadillac Fleetwood  10.4 205  0  0    3
## Lincoln Continental 10.4 215  0  0    3
## Chrysler Imperial   14.7 230  0  0    3
## Fiat 128          32.4  66  1  1    4
## Honda Civic       30.4  52  1  1    4
## Toyota Corolla    33.9  65  1  1    4
## Toyota Corona     21.5  97  1  0    3
## Dodge Challenger  15.5 150  0  0    3
## AMC Javelin       15.2 150  0  0    3
## Camaro Z28        13.3 245  0  0    3
## Pontiac Firebird  19.2 175  0  0    3
## Fiat X1-9         27.3  66  1  1    4
## Porsche 914-2     26.0  91  0  1    5
## Lotus Europa      30.4 113  1  1    5
## Ford Pantera L    15.8 264  0  1    5
## Ferrari Dino      19.7 175  0  1    5
## Maserati Bora     15.0 335  0  1    5
## Volvo 142E        21.4 109  1  1    4
```

Exercise 4

Create the object **cars_m_h** containing the columns *mpg, hp* columns but let the column names be ‘miles_per_gallon’, and ‘horse_power’ respectively.

```
cars_m_h <- mtcars %>% select(miles_per_gallon = mpg, horse_power = hp)
```

or

```
cars_m_h <- select(mtcars ,miles_per_gallon = mpg, horse_power = hp)
```

Exercise 5

Change the column names of **cars_m_h** from ‘miles_per_gallon’, and ‘horse_power’ to ‘mpg’ and ‘hp’ respectively.

```
cars_m_h <- rename(cars_m_h, miles_per_gallon = mpg , horse_power=hp)
```

Exercise 6

Print out a randomly half the observations of **cars_m_h**.
Hint : consider using the `sample_frac` function

```
cars_m_h %>% sample_frac(size = 0.5, replace = FALSE)
##                     mpg  hp
## Dodge Challenger    15.5 150
## Toyota Corona       21.5  97
## Cadillac Fleetwood  10.4 205
## Ford Pantera L      15.8 264
## Honda Civic         30.4  52
## Chrysler Imperial   14.7 230
## Camaro Z28          13.3 245
## Duster 360          14.3 245
## Volvo 142E          21.4 109
## Merc 450SL          17.3 180
## Fiat 128            32.4  66
## Merc 280C           17.8 123
## Merc 230            22.8  95
## Lincoln Continental 10.4 215
## Pontiac Firebird    19.2 175
## AMC Javelin         15.2 150
```

Exercise 7

Print out from `mtcars` object all the observations which have mpg>20 and hp>100.

```
filter(mtcars , mpg>20, hp >100)
```

Exercise 8

Create a new object named *cars_cyl* and assign to it the *mtcars*data frame grouped by the variable *cyl*
Hint: be careful about the data type of the variable, in order to be used for grouping it has to be a factor.

```
mtcars$cyl <- as.factor(mtcars$cyl)
cars_cyl <- mtcars %>% group_by(cyl)
```

Exercise 10

Print out the average and standard division of *hp* for every cyl category

```
mtcars %>% group_by(cyl) %>% summarize(Mean = mean(hp),
                                       Std = sd(hp))
## # A tibble: 3 x 3
##      cyl      Mean      Std
##   <fctr>     <dbl>    <dbl>
## 1      4  82.63636 20.93453
## 2      6 122.28571 24.26049
## 3      8 209.21429 50.97689
```