Lab 1: Exercise - Answer Key

Load the dataset "mpg" and work through the exercises below. Note, "mpg" is included in the tidyverse package, so you will need to load the package first.

1. Calculate the **mean**, **range**, **minimum**, and **maximum** of the variable "hwy" across all models. Then, combine these statistics into one vector. (Tip: look up the RDocumentation for the functions `mean`, `range`, `min`, and `max`).

```
library(tidyverse)
data(mpg)
mean_hwy <- mean(mpg$hwy)
range_hwy <- range(mpg$hwy)
min_hwy <- min(mpg$hwy)
max_hwy <- max(mpg$hwy)
sum_stats <- c(mean_hwy, range_hwy, min_hwy, max_hwy)
sum_stats
```

```
[1] 23.44017 12.00000 44.00000 12.00000 44.00000
```

2. Since "hwy" is measured in miles per gallon, create a new variable in mpg that expresses "hwy" in litres per 100 km.

```
temp <- 235.215 / mpg$hwy
mpg$hwy_Lp100km_1 <- temp


# Or, in one step
mpg$hwy_Lp100km_2 <- 235.215 / mpg$hwy


unique(mpg$hwy_Lp100km_1 == mpg$hwy_Lp100km_2) # sanity check
```

```
[1] TRUE
```

3. Identify the models of cars that are most fuel-efficient. Which classes of cars are least fuel-efficient? (Tip: to remove duplicates, use the function `unique`).

```
mpg$model[mpg$hwy == max(mpg$hwy)]
```

```
[1] "jetta"      "new beetle"
```

```
mpg$class[mpg$hwy_Lp100km_1 == max(mpg$hwy_Lp100km_1)]
```

```
[1] "pickup" "suv"    "pickup" "pickup" "suv"
```

```
unique(mpg$class[mpg$hwy_Lp100km_1 == max(mpg$hwy_Lp100km_1)])
```

```
[1] "pickup" "suv"
```

4. Compute the quantiles of "hwy". Can you also calculate the tertiles instead? (Tip: look up for the RDocumentation for the function `quantile`).

```
quantile(mpg$hwy)
```

```
  0%  25%  50%  75% 100%
  12   18   24   27   44
```

```
quantile(mpg$hwy, probs = seq(0, 1, 0.333))
```

```
    0%  33.3%  66.6%  99.9%
12.000 19.589 26.000 44.000
```

```
quantile(mpg$hwy, c(0, 0.333, 0.666, 1)) # you can achieve the same using
```

```
    0%  33.3%  66.6%   100%
12.000 19.589 26.000 44.000
```

5. Now, based on the tertiles you calculated, assign "least efficient", "medium", and "most efficient" labels to all models. Try using both base R indexing functions and the function `ifelse`.

```r
tertile <- quantile(mpg$hwy, c(0, 0.333, 0.666, 1))

mpg$efficient_class_1 <- NA # optional, this eliminates the warning message of "Unknown or uninitialised column"
mpg$efficient_class_1[mpg$hwy < tertile[2]] <- "least efficient"
mpg$efficient_class_1[mpg$hwy >= tertile[2] & mpg$hwy < tertile[3]] <- "medium"
mpg$efficient_class_1[mpg$hwy >= tertile[3]] <- "most efficient"

# OR
mpg$efficient_class_2 <- NA
mpg$efficient_class_2 <- ifelse(mpg$hwy < tertile[2], "least efficient", mpg$efficient_class_2)
mpg$efficient_class_2 <- ifelse(mpg$hwy >= tertile[2] & mpg$hwy < tertile[3], "medium", mpg$efficient_class_2)
mpg$efficient_class_2 <- ifelse(mpg$hwy >= tertile[3], "most efficient", mpg$efficient_class_2)

unique(mpg$efficient_class_1 == mpg$efficient_class_2) # sanity check
```

```
[1] TRUE
```