Assignment 1 solutions

You can access datasets from the R datasets package by using

```
data(NAME_OF_DATASET)
```

For this question, we will use the dimaonds data from the `ggplot2` library.

```
library(tidyverse) # Note the tidyverse package loads the ggplot2 library
data(diamonds)
```

Note you can learn about this dataset by using

```
help(diamonds)
```

a. Determine the (i) mode and (ii) class of the `diamonds` data object.

```
mode(diamonds)
```

```
[1] "list"
```

```
class(diamonds)
```

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

b. How would you find how many rows and columns the object has by using R functions `nrow` and `ncol` ? Give the code and the result.

```
nrow(diamonds)
```

```
[1] 53940
```

```
ncol(diamonds)
```

```
[1] 10
```

c. What is the value contained in row 12345 and the `depth` column (which contains the depth percentage)?

```
diamonds[12345,"depth"]
```

```
# A tibble: 1 x 1
  depth
  <dbl>
1  64.5
```

d. Write a line of code that creates a new data object called `diamonds_imp` which is of the same mode and class as the original `diamonds` data object and contains the same columns as the original, but also contains three new columns: `x_imp` , `y_imp` , `z_imp` where each of these measurements are Imperial measurements in inches, i.e. `x_imp` is equal to `x` divided by 25.4, as there are 25.4 mm in 1 inch. Show the first 6 rows of the resulting data object.

```
diamonds_imp <- diamonds %>% mutate(x_imp=x/25.4, y_imp=y/25.4, z_imp=z/25.4)
head(diamonds_imp)
```

```
# A tibble: 6 x 13
  carat cut       color clarity depth table price     x     y     z x_imp y_imp
  <dbl> <ord>     <ord> <ord>   <dbl> <dbl> <int> <dbl> <dbl> <dbl> <dbl> <dbl>
1  0.23 Ideal     E     SI2      61.5    55   326  3.95  3.98  2.43 0.156 0.157
2  0.21 Premium   E     SI1      59.8    61   326  3.89  3.84  2.31 0.153 0.151
3  0.23 Good      E     VS1      56.9    65   327  4.05  4.07  2.31 0.159 0.160
4  0.29 Premium   I     VS2      62.4    58   334  4.2   4.23  2.63 0.165 0.167
5  0.31 Good      J     SI2      63.3    58   335  4.34  4.35  2.75 0.171 0.171
6  0.24 Very Good J     VVS2     62.8    57   336  3.94  3.96  2.48 0.155 0.156
# … with 1 more variable: z_imp <dbl>
```

e. Write a line of code that adds a column named `over_under` to the `diamonds_imp` data object that contains the difference between the price of the diamond in that row and the `median` of the prices of other diamonds with the same `color` .

```
diamonds_imp <- diamonds_imp %>% group_by(color) %>% mutate(over_under = price-median
(price))
diamonds_imp
```

```
# A tibble: 53,940 x 14
# Groups:   color [7]
   carat cut       color clarity depth table price     x     y     z x_imp y_imp
   <dbl> <ord>     <ord> <ord>   <dbl> <dbl> <int> <dbl> <dbl> <dbl> <dbl> <dbl>
 1  0.23 Ideal     E     SI2      61.5    55   326  3.95  3.98  2.43 0.156 0.157
 2  0.21 Premium   E     SI1      59.8    61   326  3.89  3.84  2.31 0.153 0.151
 3  0.23 Good      E     VS1      56.9    65   327  4.05  4.07  2.31 0.159 0.160
 4  0.29 Premium   I     VS2      62.4    58   334  4.2   4.23  2.63 0.165 0.167
 5  0.31 Good      J     SI2      63.3    58   335  4.34  4.35  2.75 0.171 0.171
 6  0.24 Very Good J     VVS2     62.8    57   336  3.94  3.96  2.48 0.155 0.156
 7  0.24 Very Good I     VVS1     62.3    57   336  3.95  3.98  2.47 0.156 0.157
 8  0.26 Very Good H     SI1      61.9    55   337  4.07  4.11  2.53 0.160 0.162
 9  0.22 Fair      E     VS2      65.1    61   337  3.87  3.78  2.49 0.152 0.149
10  0.23 Very Good H     VS1      59.4    61   338  4     4.05  2.39 0.157 0.159
# … with 53,930 more rows, and 2 more variables: z_imp <dbl>, over_under <dbl>
```

f. Write a line of code that creates a new data object from the original `diamonds` data object named `Expensive` that contains only the diamonds whose price is *strictly* greater than $18800 and show the contents of that data object.

```
Expensive <- diamonds %>% filter(price > 18800)
Expensive
```

```
# A tibble: 5 x 10
  carat cut       color clarity depth table price     x     y     z
  <dbl> <ord>     <ord> <ord>   <dbl> <dbl> <int> <dbl> <dbl> <dbl>
1  2    Very Good H     SI1      62.8    57 18803  7.95  8     5.01
2  2.07 Ideal     G     SI2      62.5    55 18804  8.2   8.13  5.11
3  1.51 Ideal     G     IF       61.7    55 18806  7.37  7.41  4.56
4  2    Very Good G     SI1      63.5    56 18818  7.9   7.97  5.04
5  2.29 Premium   I     VS2      60.8    60 18823  8.5   8.47  5.16
```

Question 2

The Statistical Society of Canada (the SSC) is the professional society for statisticians in academics and industry in Canada. The Board of Directors for the Society is made up of an elected executive committee and elected regional representatives. Below is a *partial* list of the members of the Board of Directors of the SSC, along with their roles and the dates of the ends of their elected terms.

```
board_of_directors <- list(
  Exec = tibble(
    Name=c("Grace Yi","Bruno Rémillard"),
    Position=c("President","President-Elect"),
    Term_End=c("2022-06-30","2022-06-30")
  ),
  Regional_Reps = list(
    Atlantic_Region = tibble(
    Name=c("Michael McIsaac","Wilson Lu"),
    Term_End=c("2022-06-30","2023-06-30")
    ),
    Quebec = tibble(
      Name=c("Paramita Saha Chaudhuri","Cody Hyndman",
             "Johanna Neslehova","Denis Talbot"),
      Term_End=c("2022-06-30","2022-06-30","2023-06-30","2023-06-30")
    )
  )
)
```

a. What objects (or values) are returned by the following lines of R code?

```
board_of_directors$Exec[1,2]
```

```
# A tibble: 1 x 1
  Position
  <chr>
1 President
```

```
board_of_directors[[2]][1,]
```

```
Error in board_of_directors[[2]][1, ]: incorrect number of dimensions
```

```
board_of_directors[[2]][1]
```

```
$Atlantic_Region
# A tibble: 2 x 2
  Name            Term_End
  <chr>           <chr>
1 Michael McIsaac 2022-06-30
2 Wilson Lu       2023-06-30
```

```
board_of_directors[2][[1]][[1]]
```

```
# A tibble: 2 x 2
  Name            Term_End
  <chr>           <chr>
1 Michael McIsaac 2022-06-30
2 Wilson Lu       2023-06-30
```

b. Using R code, write statements which yield the following three results:

```
cat("Result 1:\n")
```

```
Result 1:
```

```
board_of_directors$Regional_Reps$Quebec[3,1]
```

```
# A tibble: 1 x 1
  Name
  <chr>
1 Johanna Neslehova
```

```
cat("\n")
```

```
cat("Result 2: \n")
```

```
Result 2:
```

```
board_of_directors[[1]][,2]
```

```
# A tibble: 2 x 1
  Position
  <chr>
1 President
2 President-Elect
```

```
cat("Result 3: \n")
```

```
Result 3:
```

```
board_of_directors[[2]][[2]][1,]
```

```
# A tibble: 1 x 2
  Name                     Term_End
  <chr>                    <chr>
1 Paramita Saha Chaudhuri 2022-06-30
```

```
cat("Result 4: \n")
```

```
Result 4:
```

```
board_of_directors[[2]][[2]][c(1,3),]
```

```
# A tibble: 2 x 2
  Name                     Term_End
  <chr>                    <chr>
1 Paramita Saha Chaudhuri 2022-06-30
2 Johanna Neslehova       2023-06-30
```