Problem Set 3

- Use the Iris built-in data.

    1. Put all rows of Species 'versicolor' in a new data frame. Call this data frame: 'iris.vers'

    2. Put a vector called 'sepal.dif' with the difference between 'Sepal.Length' and 'Sepal.Width' of 'versicolor' plants.

    3. Update (add) 'iris.vers' with the new column 'sepal.dif'.

- Use the mtcars built-in data.

    4. Using the apply family of functionals, check the class of each variable in 'mtcars'.

    5. Change 'am', 'cyl' and 'vs' to integer and store the new dataset as 'newmtc'.

    6. Round the 'newmtc' data frame to one digit.

- Use the Iris built-in data.

    7. Use dplyr to filter the Iris data frame for all data of Species 'virginica' with a 'Sepal.Width' of greater than 3.5.

    8. How would you use R Base to get a data frame of all data of Species 'virginica' with a 'Sepal.Width' of greater than 3.5, but without the last column Species in the data frame?

    9. Get the row IDs of the rows matching the two filtering criteria provided above.

       Expected result: "110" "118" "132"

- Use the Diamonds built-in data.

    10. How many observations of diamonds have a cut of 'ideal' and have less than 0.21 carat?

    11. How many observations of diamonds have a combined 'x' + 'y' + 'z' dimension greater than 40?

12. How many observations of diamonds have either a price above 10.000 USD or a depth of at least 70?

13. Get a data frame with observations '67' and '982' of variables color and y.

14. Get a data frame with the full info on observations '453', '792' and '10489'.

15. Get the first 10 rows of the dataset 'diamonds' with the variables 'x', 'y', 'z'.

16. Create the object 'newdiam' which is a subset of the first 1000 rows of 'diamonds'.

17. Order 'newdiam' according to price, starting with the lowest.

18. Use 'dplyr', 'sample_n' to get the object 'diam750' which contains 750 randomly sampled observations of 'diamonds'. I use seed nr. 56 for reproduction.

19. Get a summary of the new data frame.

20. Using 'diam750' plot a scatter plot of price vs depth, using R Base plot and ggplot2.