**Exploratory Data Analysis – Assignment 1**

Last Date: 3rd March 2018

**Notes:**

* Dataset: Please download the dataset (parliament.zip) from this link <https://github.com/skathirmani/datasets>. In this data set each row represents votes gained by a contestant in his/her constituency for a specific year. Along with this we also have his party, gender and winning position details (Candidate with position 1 has won that constituency). Likewise data set contains information about all the parliament elections till 2014
* Please use R Markdown. Submit the .Rmd & .html file only
* Have proper titles, explanation wherever necessary. Suppress unnecessary print statements and warning messages. Submission which doesn’t follow these criteria will not be consider for evaluation.

**Questions**

1. Create a function which will take a data frame as input and returns two vector as output. The first vector should have names of those columns which are of type numeric. And the second vector should have names of those columns which are of type factor. Remember that the types float and integer should be consider as numeric
2. Create a function which will take a data frame as input and returns the following data frame. The function should automatically filter the numerical columns. The number of columns in the output data frame should be equal to number of numerical columns

|  |  |  |  |
| --- | --- | --- | --- |
| Statistic | Column 1 | Column 2 | Column 3 |
| Min |  |  |  |
| 25th percentile |  |  |  |
| 50th percentile |  |  |  |
| Mean |  |  |  |
| 75th percentile |  |  |  |
| Max |  |  |  |
| Range |  |  |  |
| Variance |  |  |  |
| Standard Deviation |  |  |  |
| % of missing values |  |  |  |
| % of outliers |  |  |  |

1. Create a function which will take a data frame as input and returns the following data frame. The function should automatically filter the categorical columns. The number of columns in the output data frame should be equal to number of categorical columns

|  |  |  |  |
| --- | --- | --- | --- |
| Statistic | Column 1 | Column 2 | Column 3 |
| Unique no. of levels |  |  |  |
| Most repeating level (Mode) |  |  |  |
| Freq. of most repeating level |  |  |  |
| % of levels that together contribute to 80% of the frequency |  |  |  |

1. Using parliament data, identify year wise percentage of male and female candidates?
2. Using parliament data, Compute party wise total number of candidates won so far? Display top ten parties alone