**EDA using R**

**(Total marks: 30) Duration: 3.5 hours**

1. Consider the US weather History Dataset (KSEA.csv) attached. This dataset contains observations on Temperature in US and changes in it. Following are the fields in the dataset:

| Column | Description |
| --- | --- |
| Date | The date of the weather record, formatted YYYY-M-D |
| actual\_mean\_temp | The measured average temperature for that day |
| actual\_min\_temp | The measured minimum temperature for that day |
| actual\_max\_temp | The measured maximum temperature for that day |
| average\_min\_temp | The average minimum temperature on that day since 1880 |
| average\_max\_temp | The average maximum temperature on that day since 1880 |
| record\_min\_temp | The lowest ever temperature on that day since 1880 |
| record\_max\_temp | The highest ever temperature on that day since 1880 |
| record\_min\_temp\_year | The year that the lowest ever temperature occurred |
| record\_max\_temp\_year | The year that the highest ever temperature occurred |
| actual\_precipitation | The measured amount of rain or snow for that day |
| average\_precipitation | The average amount of rain or snow on that day since 1880 |
| record\_precipitation | The highest amount of rain or snow on that day since 1880 |

Using the dataset, perform an exploratory data analysis (EDA). Remember that this is an open ended question, and while there are no certain yardsticks for EDA you should at a minimum try and check data summaries (five point summary and describe), check out missing values, outliers in dataset, distributions, and basic plots so as to familiarize yourself with the dataset. While you are running the code, also note down your observations about the dataset and outputs. Finally, address the following questions:

1. A) Which day was the maximum temperature recorded?
2. B) Highest rainfall or snow recorded in which year?
3. C) What is generally the distribution of temperature that a particular year have? You do not have to be very specific here but generally try to find out what proportion of high, low, medium temperature in a year? You can try and put temperature into high, low, and medium bucket based on data given.
4. D) How much of the rainfall and the hottest temperature recorded over last six months?
5. E) Is temperature consistent during any period of time? Examine?