**Software Development Lab Documentation & Notes**

**Date : 19th July 2016**

**Senthil R 2013103602**

**( Teamed up with Sanjana Jain S 2013013601 )**

**Project Title : Movie Data Analysis and Recommender**

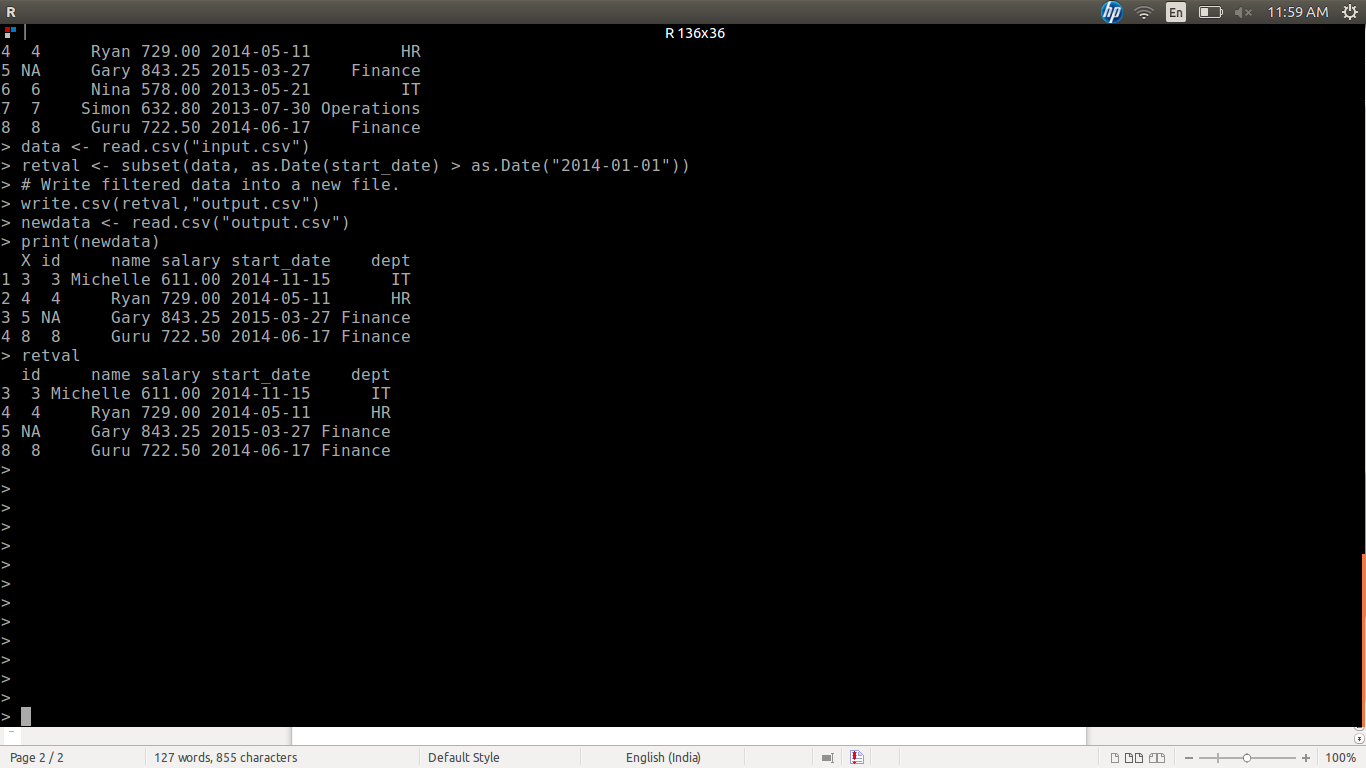
**1. Reading files as csv xls and writing it back to the file**

data <- read.csv("input.csv")

print(is.data.frame(data))

print(ncol(data))

print(nrow(data))



**-> and basic ops tried out on this data**

**2. Summary of data**

**-> summary concept**

print(is.data.frame(data))

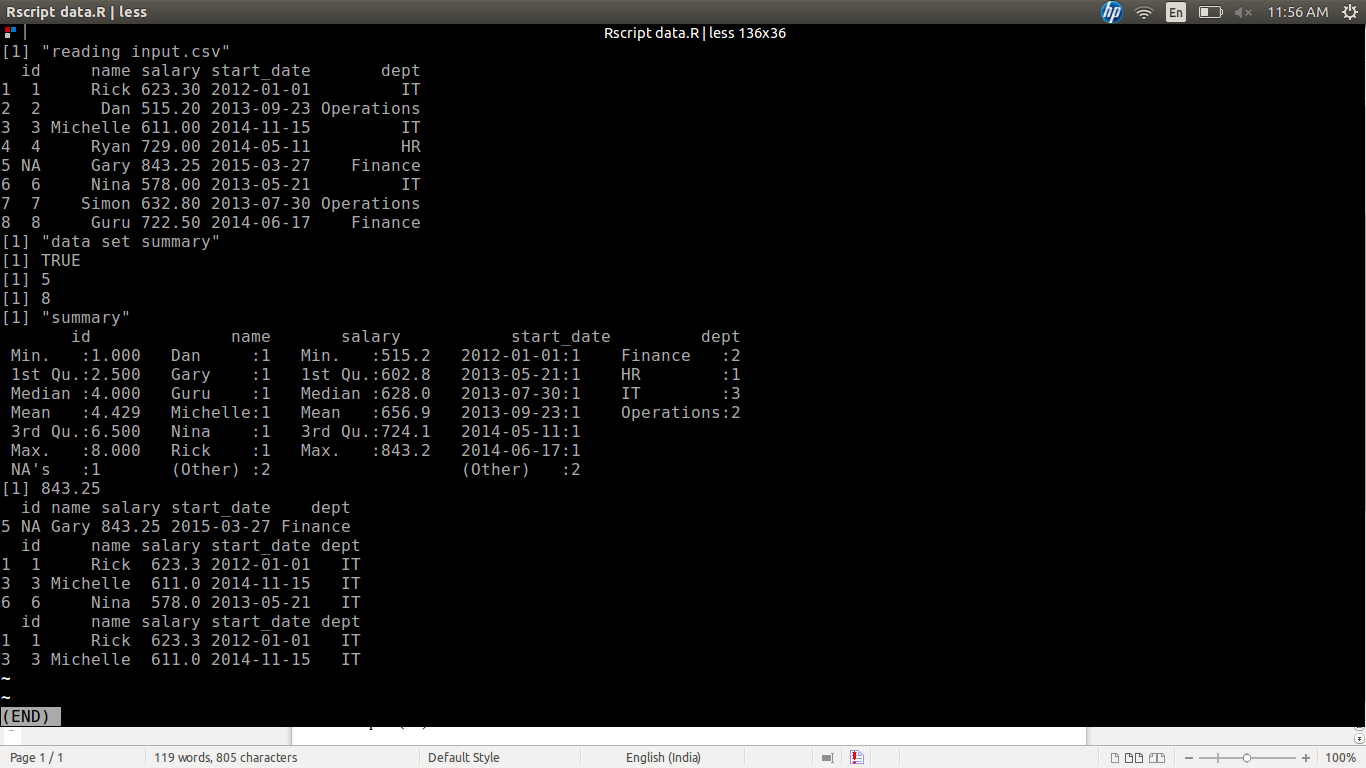
print(ncol(data))

print(nrow(data))

print ("summary")

print(summary(data))

**3. Analysing the data**

**-> to get the maximim of it**

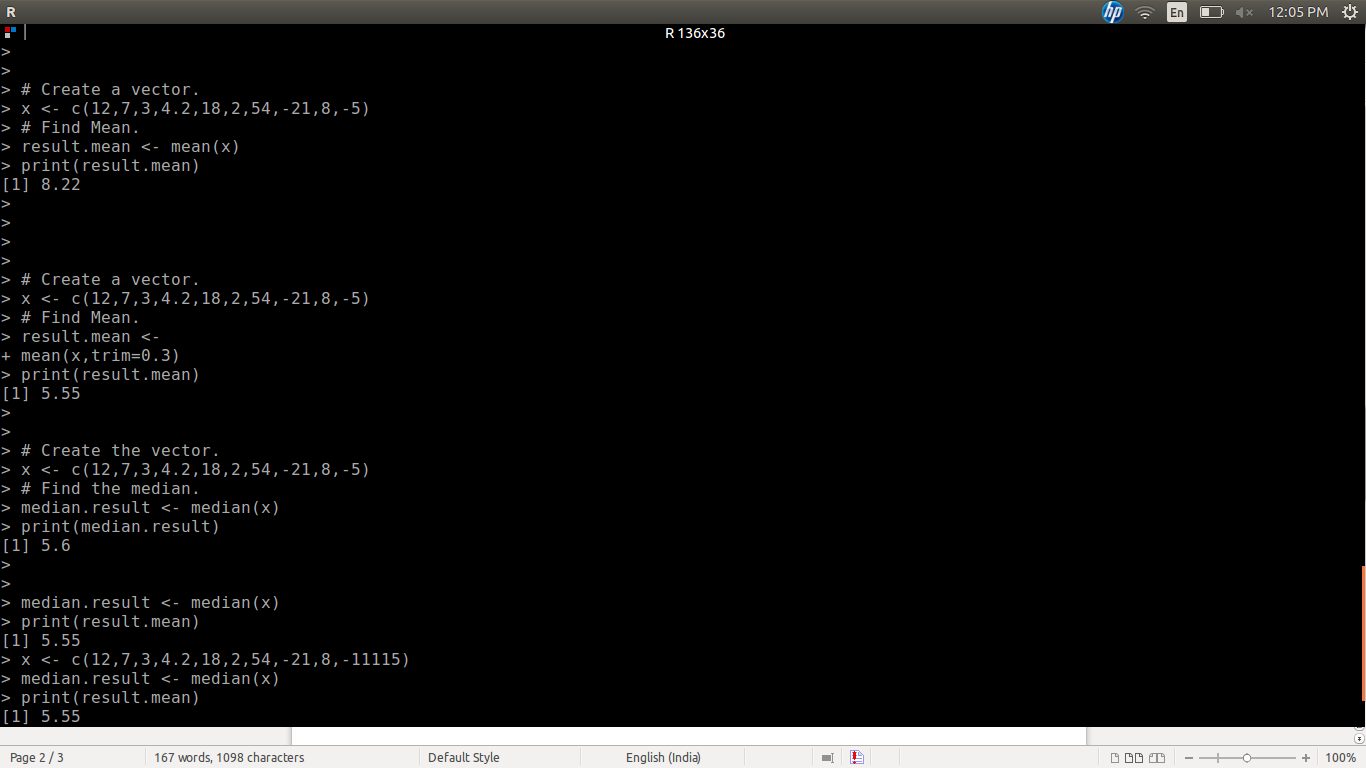
data <- read.csv("input.csv")

# Get the max salary from data frame.

sal <- max(data$salary)

print(sal)

**-> mean median mode**

****

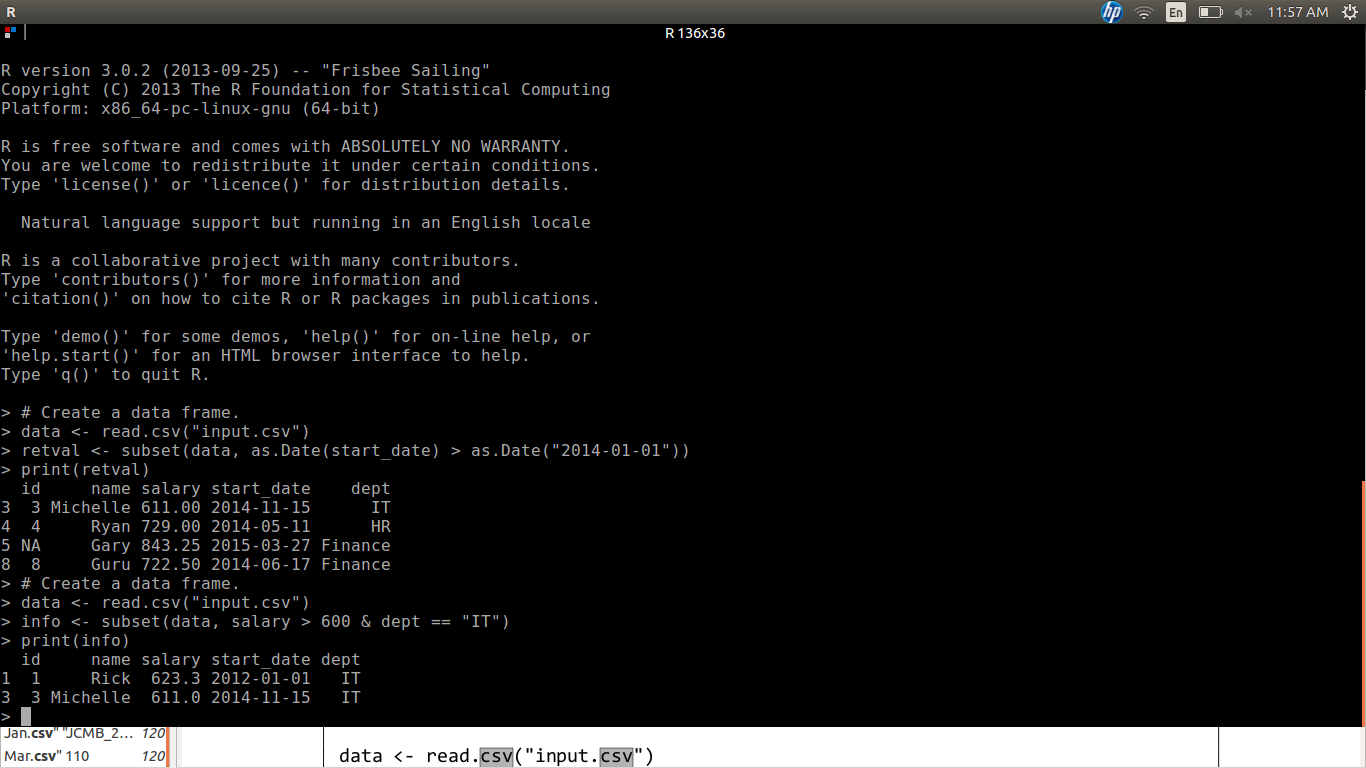
**-> individual max extraction**

sal <- max(data$salary)

# Get the person detail having max salary.

retval <- subset(data, salary == max(salary))

print(retval)



**-> Subset analysis ( individual constraint basis )**

**4. Took the data, tried to understand and find details**

**-> read about scrapping web data using ( couldn't try it coz of internet issues )**

**RCurl**

**XML**

**stringr**

**pylr**

**Dataset Link :** [**www.grouplens.org/datasets/movielens**](http://www.grouplens.org/datasets/movielens)

**Conclusion :** Lots of null values found, so we are trying to get a cleaned version of the same by processing.