**KDM HACKATON REPORT**

**DEBT PREDICTOR**



Submitted by,

Tej Kiran Meka

**Data Model (Features):**

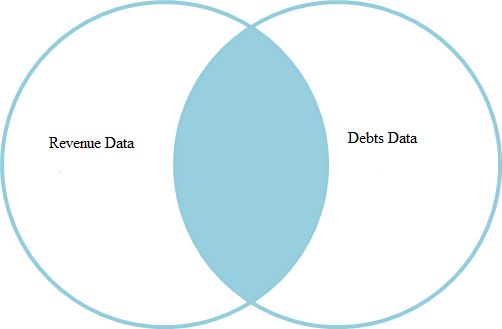
In this project I have used prediction mechanism for the project. I have collected data sets and integrated those databases and I have implemented the Regression analysis and made prediction on the debt.

Features of this model:

Predicts the value of the Debt when the population and revenue data are entered.

**Integration Model Algorithm:**

For this exam, I have taken two datasets debts of the cities and Revenues of the cities. Since I have taken the datasets in the form of Microsoft Access Database, I have developed a python code to integrate both datasets .

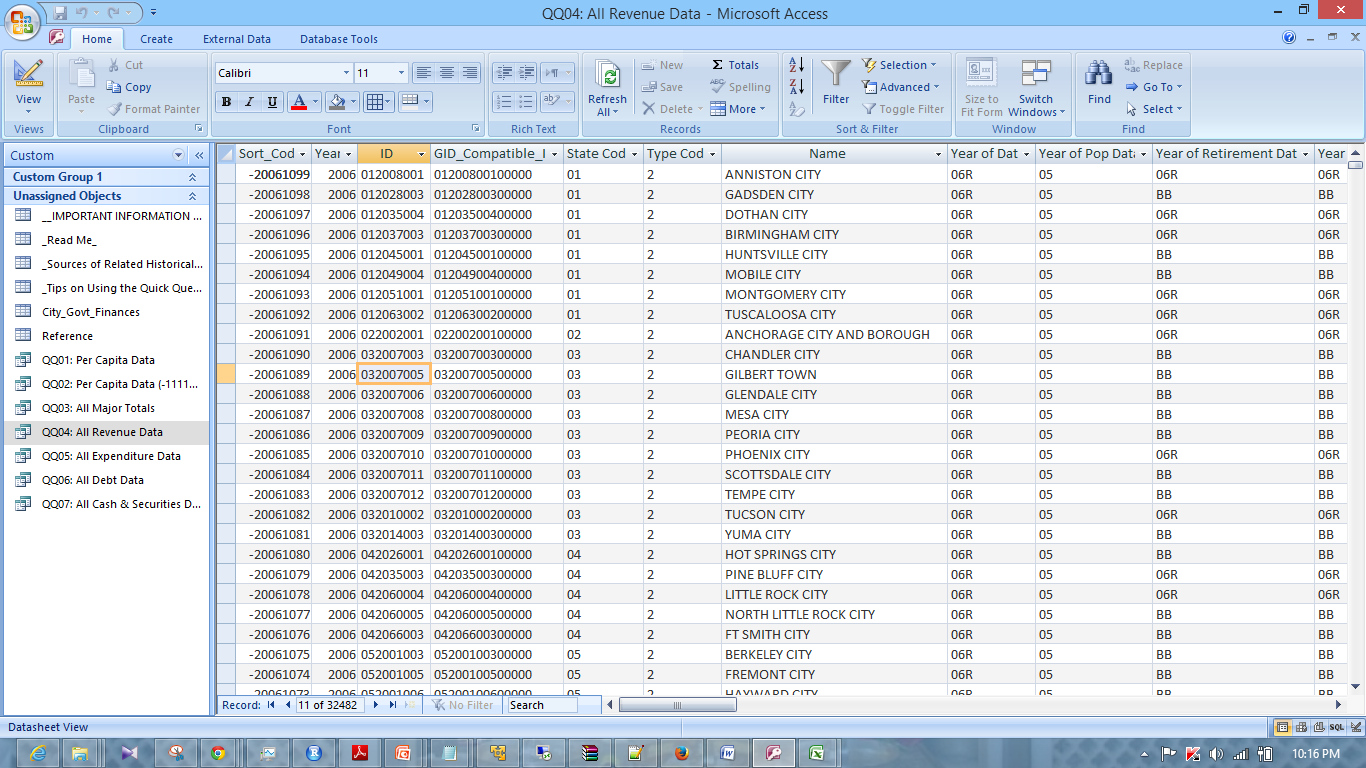


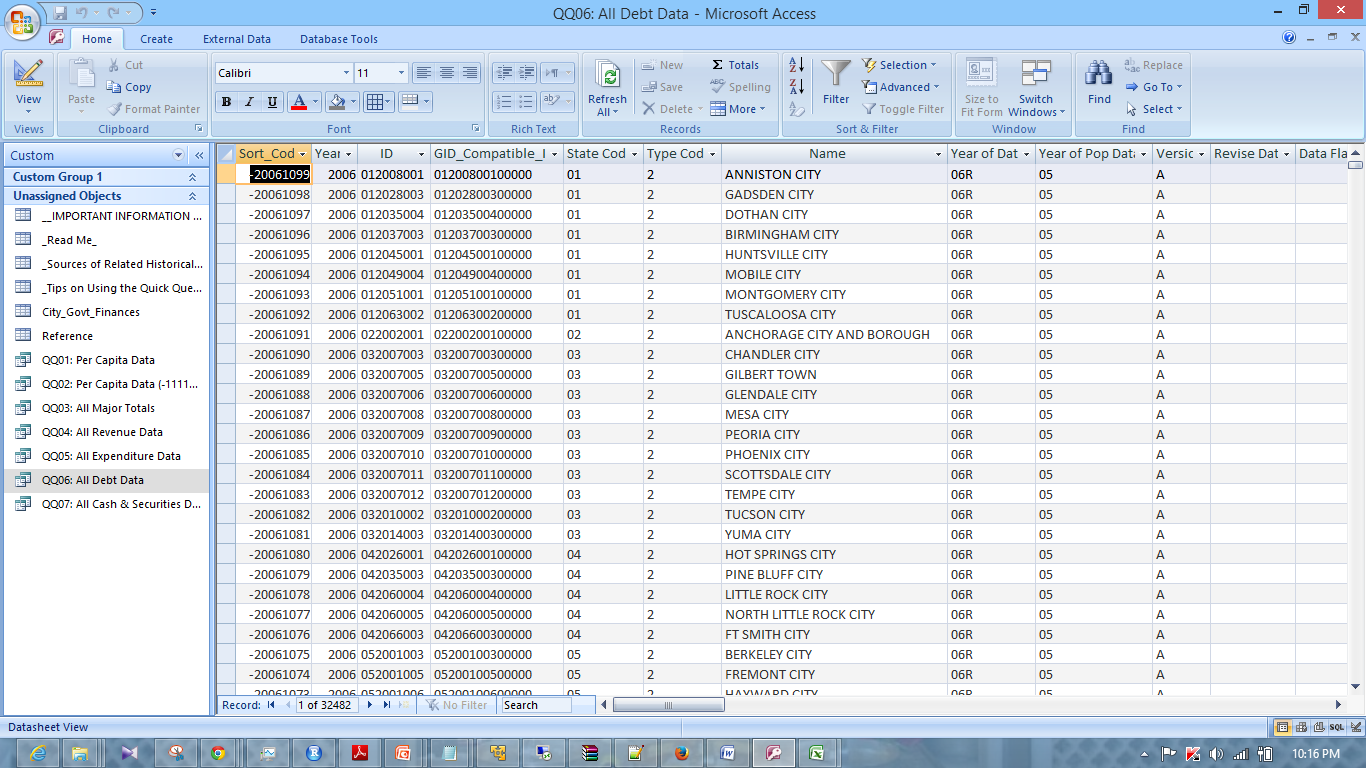
**Prediction Model and Algorithm:**

For making predictions in this model I have used the Multi linear Regression Model and Algorithms. For implementing this algorithm I have used R-Project tool developed by the R-Project which has the inbuilt Multi linear Regression Algorithm. Through this multi linear regression algorithm we can make predictions on one dependent Y variable using the two X variables X1, X2. Formula for making these predictions using the multiple linear regression algorithm is Y=a+ b1\*X1 + b2\*X2+b3\*X3+b4\*X4+.... bn\*Xn. For implementing this prediction model I have taken the Total Revenue and Population as the independent X variables and I have considered the Debt as the dependent variables and will be predicted through this algorithms.

**Selection of Datasets:**

For this Hackaton I have collected data from www.data.gov. Through this I have collected the cities and their revenue data collected from the year 1951-2006 for different cities in United States.





By using these Datasets I have joined the two tables All Revenue Data and All Debt Data using the native database joins by writing join condition on Name of the city. I have developed a python script to convert the joined tables into CSV files. I got the required CSV files and I run the Regression analysis on those tables to make predictions.

**Mobile App/Web Design**:

For this part of the project I have used twitter bootstrap to design the mobile web page.

**Features Implemented:**

**Integration Algorithms:**

For integration of the tables in msaccess I have developed a code in python to join the tables from the two datasets and to process the datasets.

**Predictive Algorithms:**

For this part I have used Project-R and the R-studio to run the multiple linear regression on the data set that I have imported the dataset into the R-studio and then I run the multiple linear regression analysis on the data set to predict the models.

**Solr Indexing:**

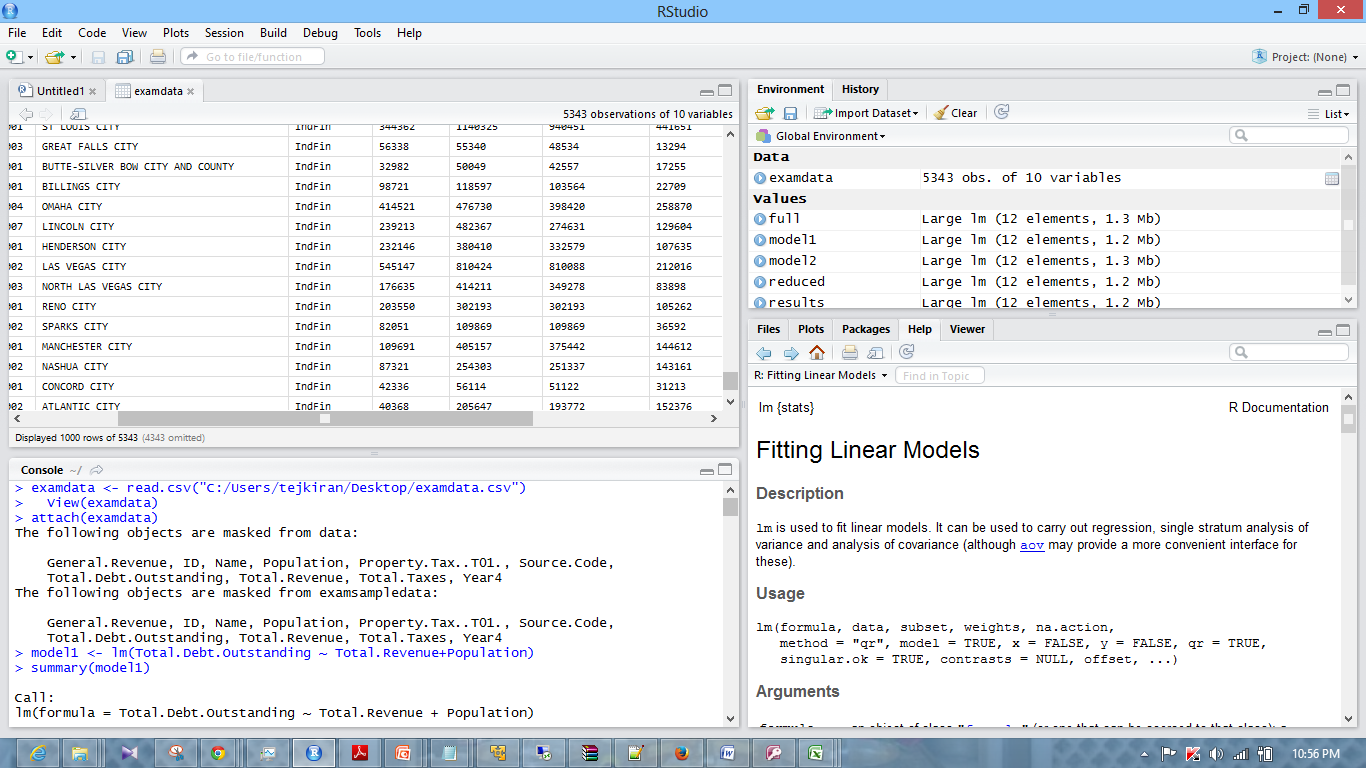
For solr indexing I have followed the steps mentioned in the tutorials to upload the json file into the solr using the putty and File Zilla. I collected the prediction output data from the R and I have used the python script to convert the text file into the json file. Through this I got the json data that can be used with the mobile web service.

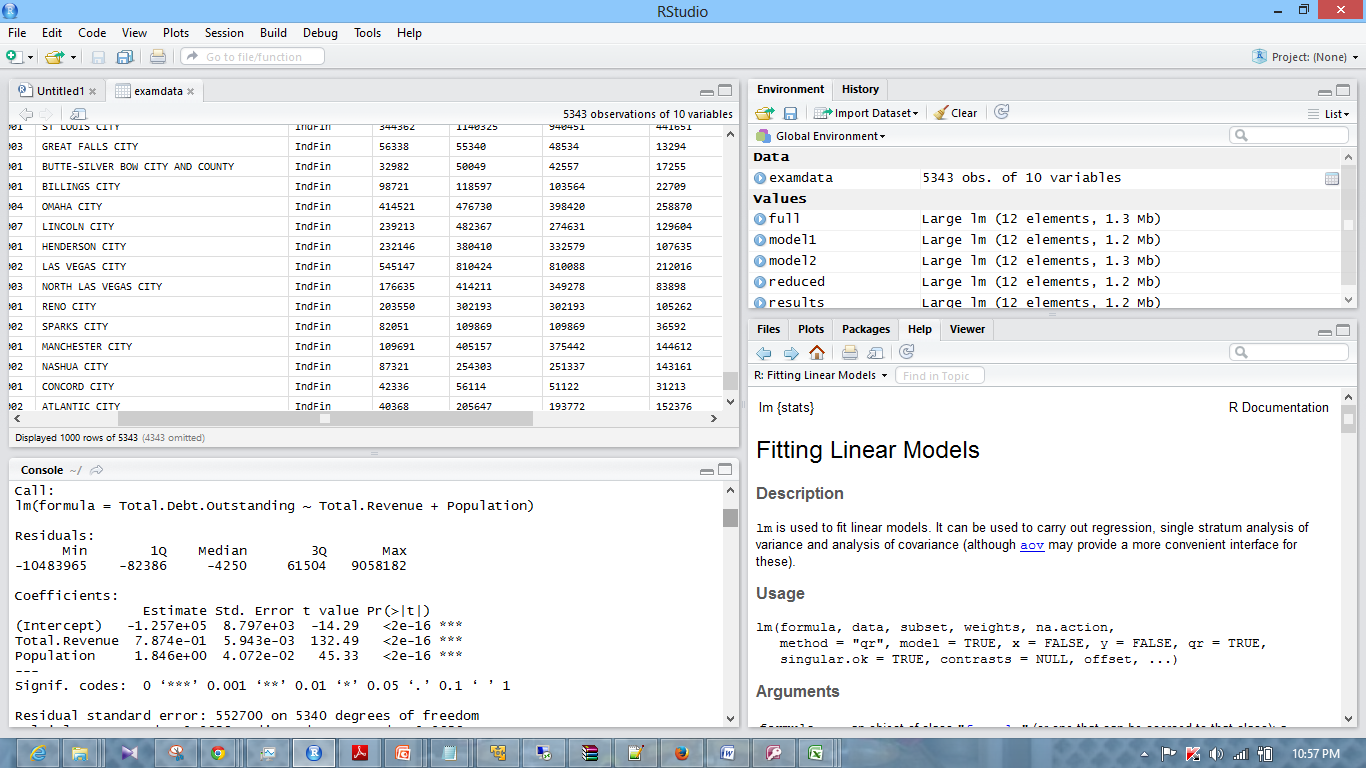
**Mobile interface Design:**

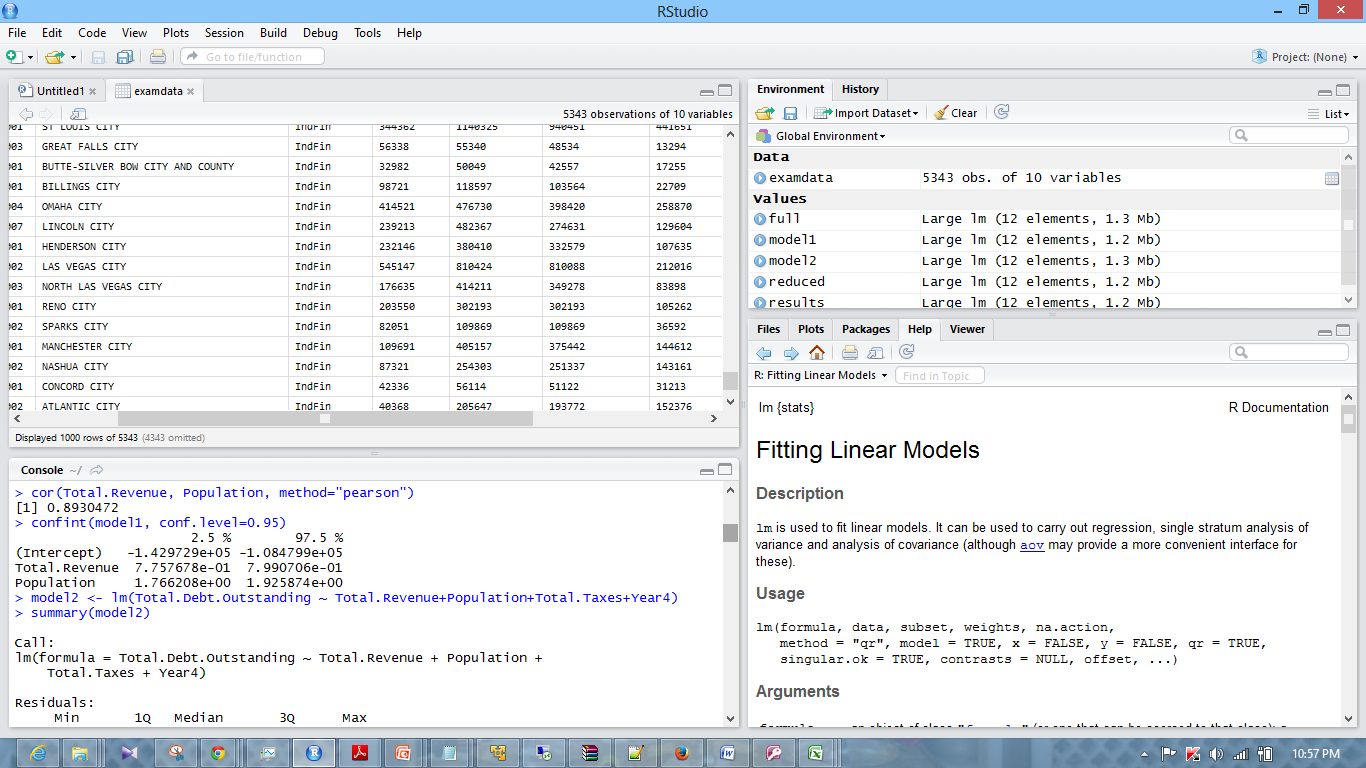
For designing the mobile interface I used twitter bootstrap for the website.

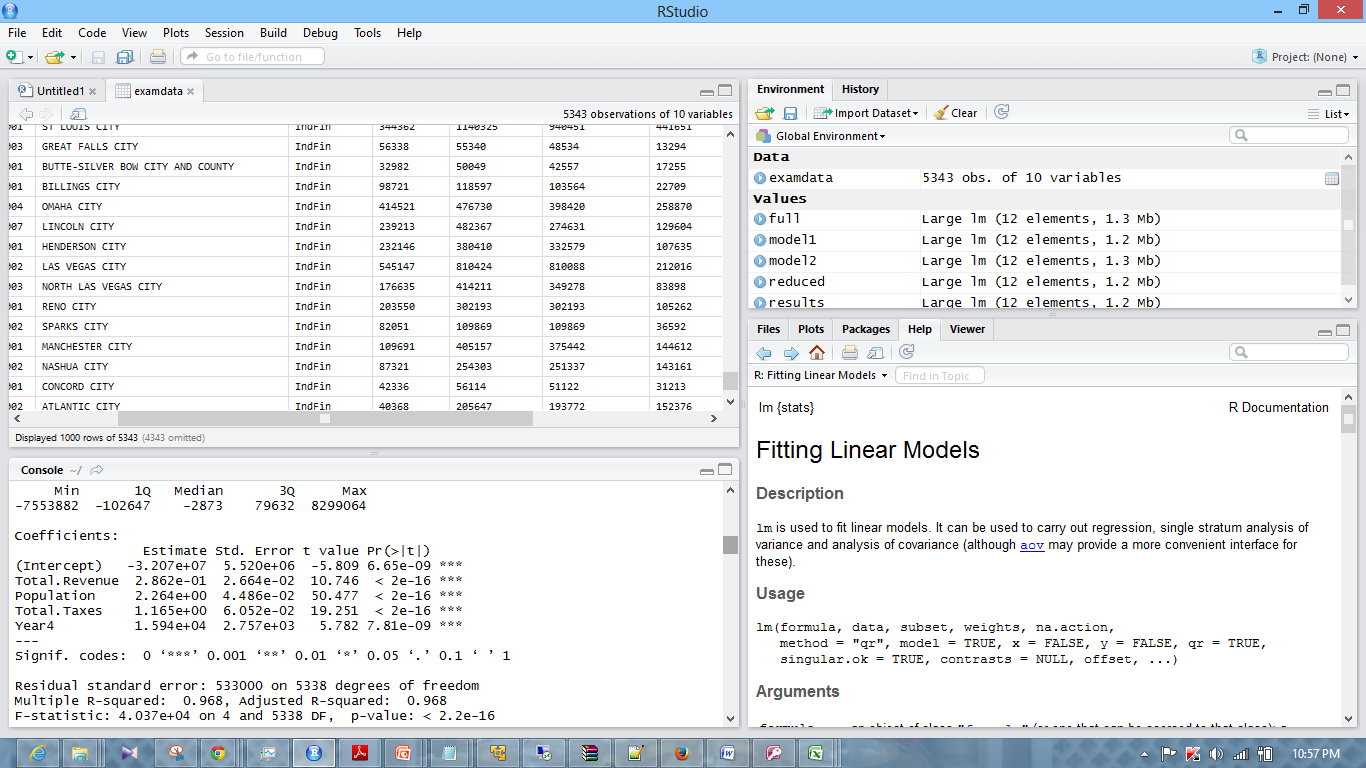
**Outputs:**

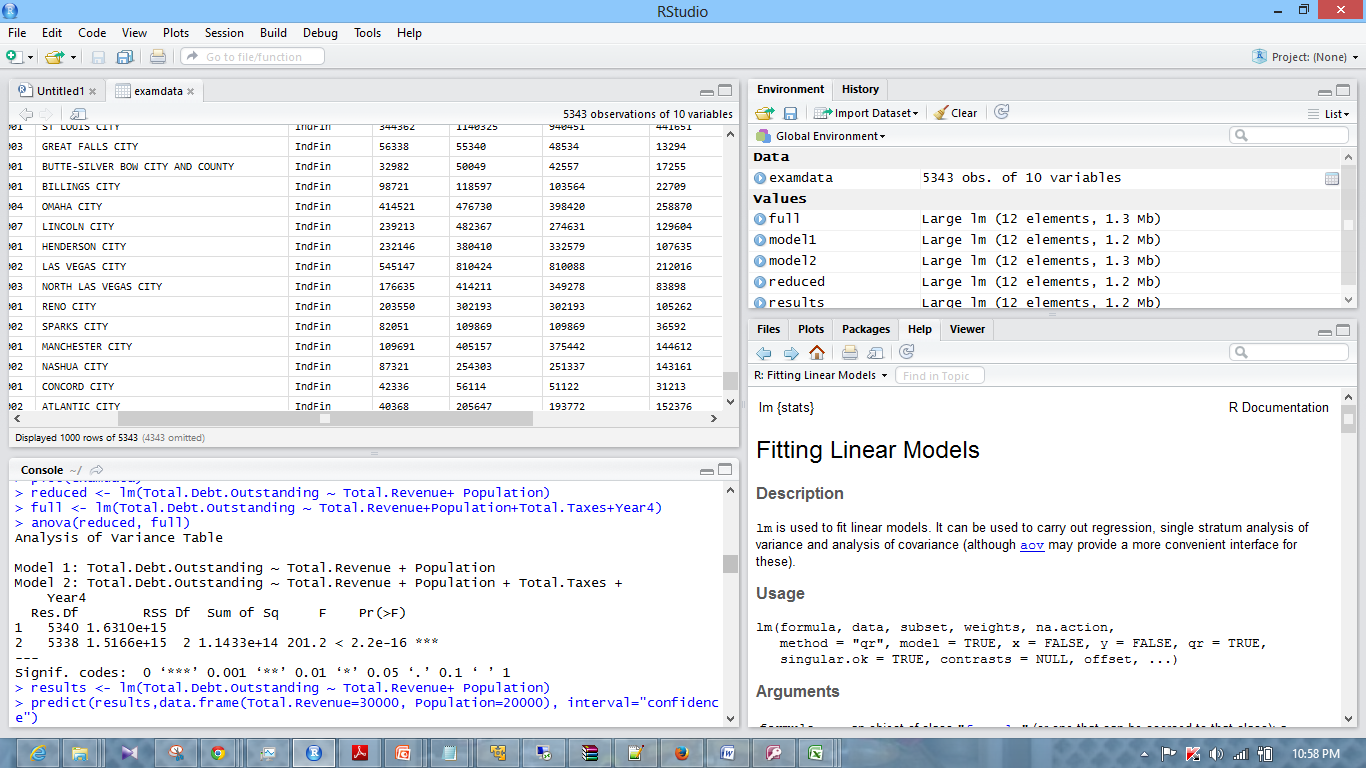
**R-Screenshots:**



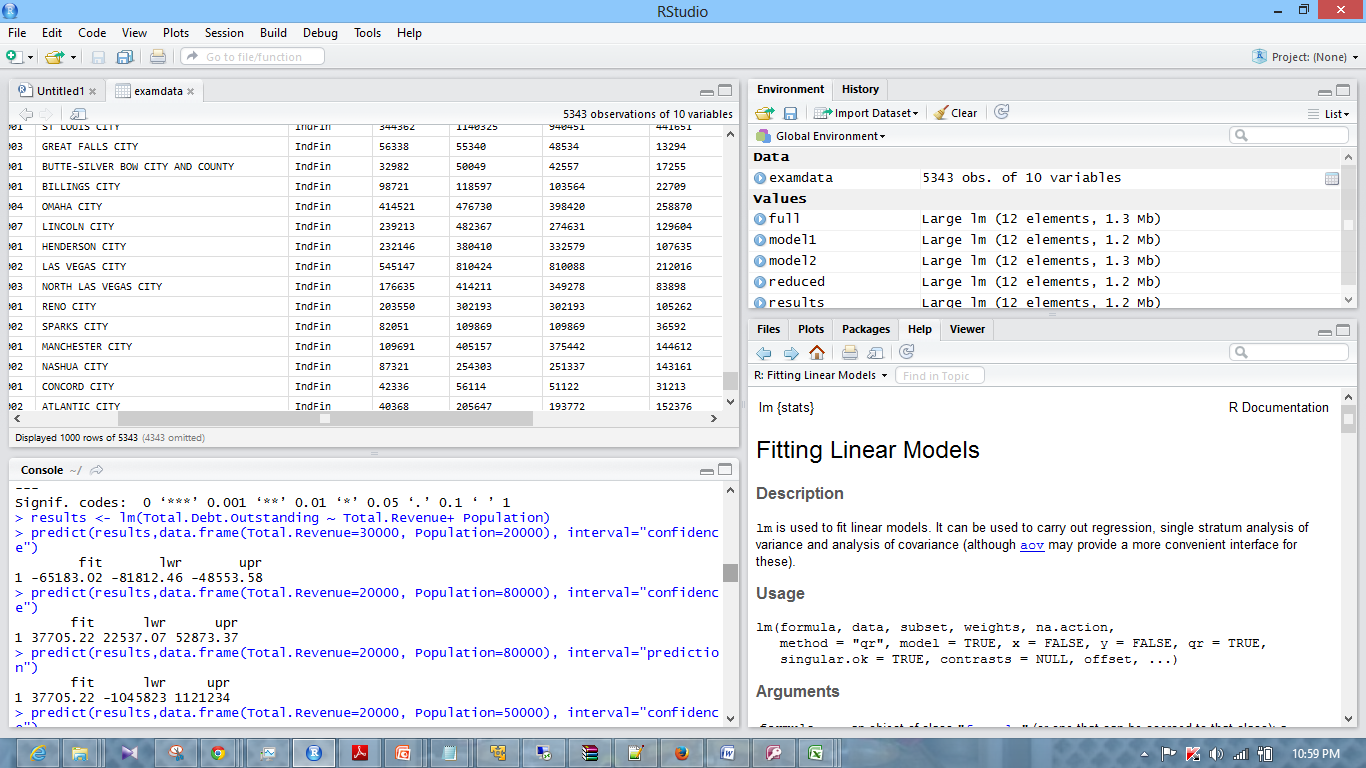


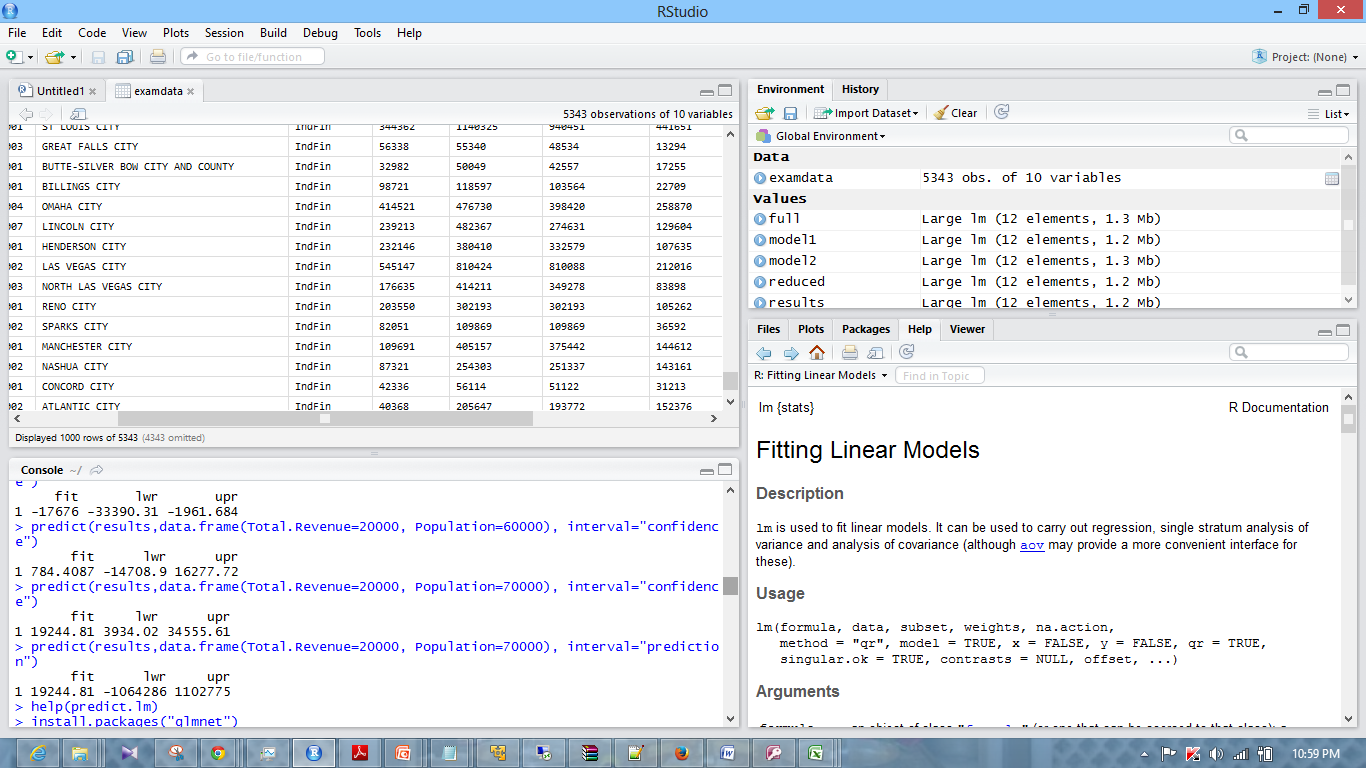


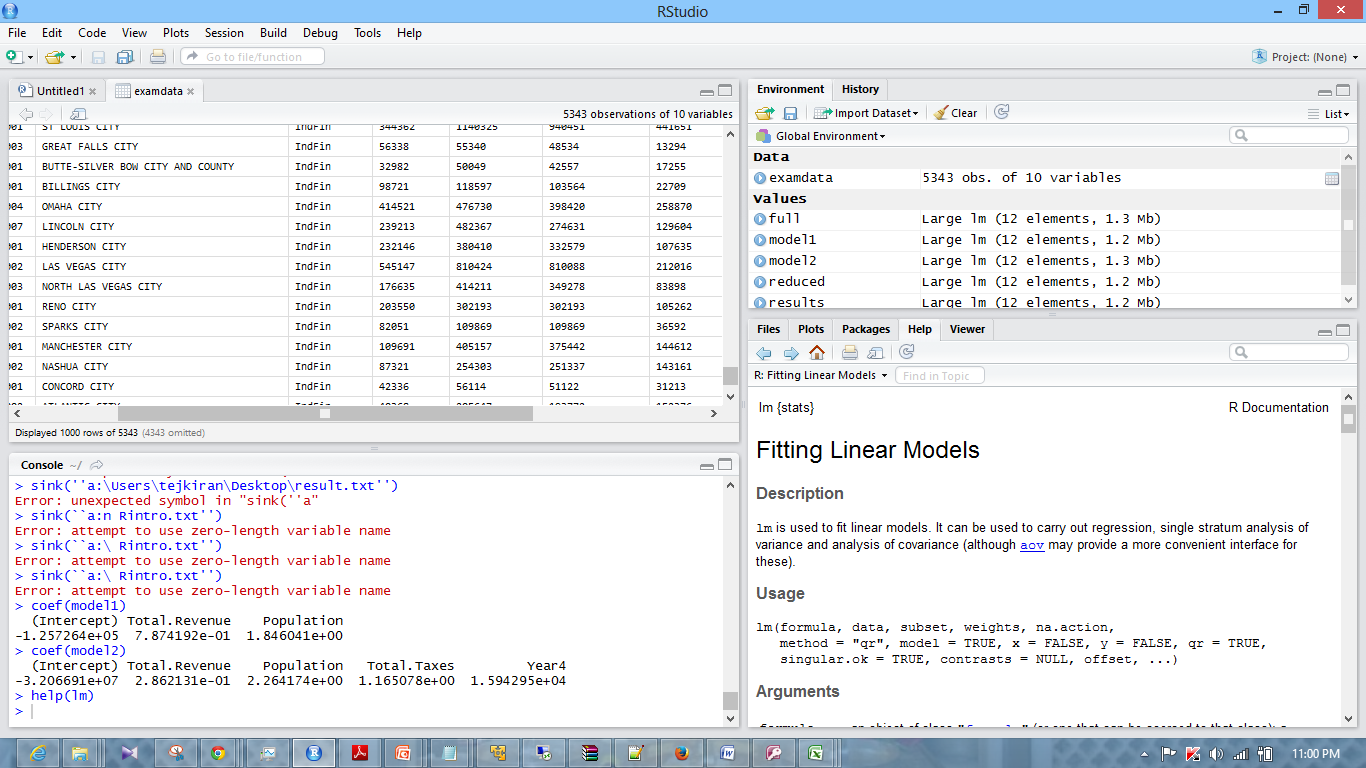


****

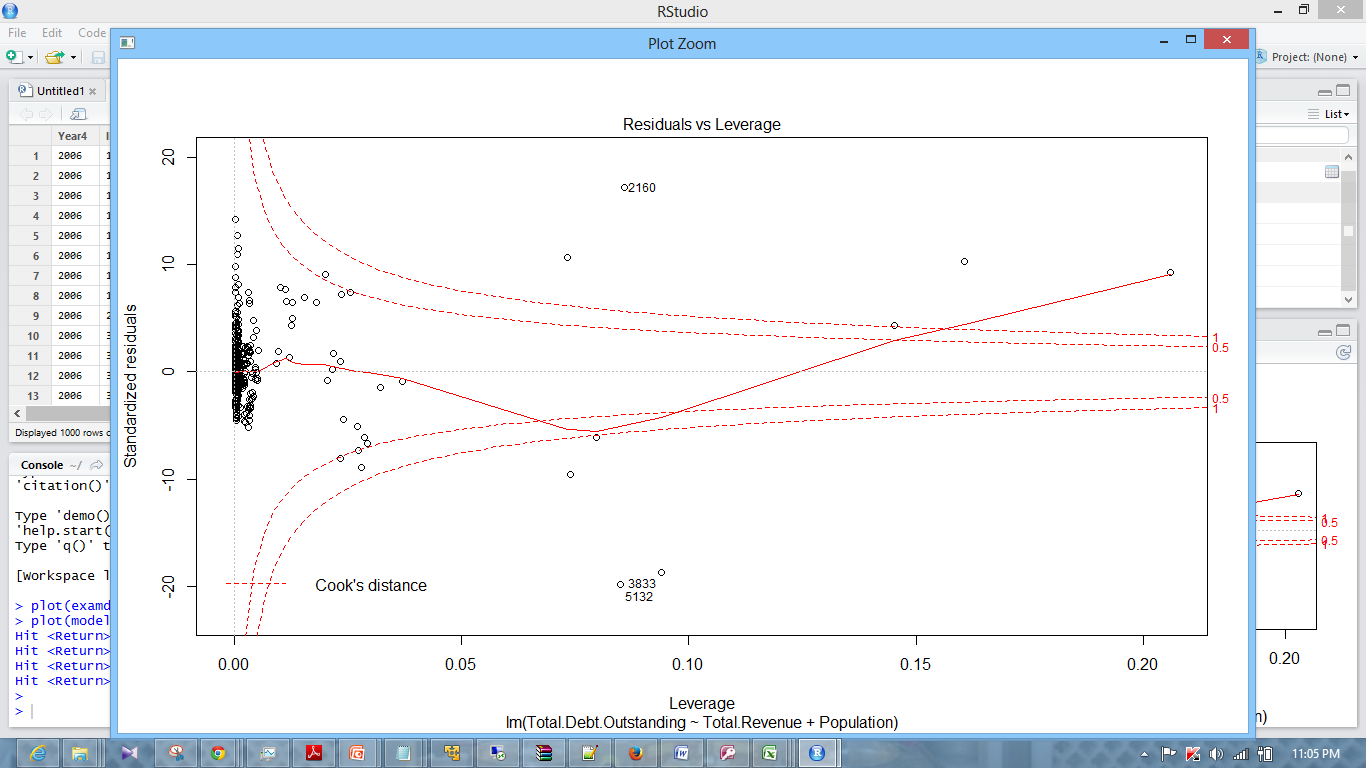
Below Screenshots represents the prediction results of the Multiple Linear Regression using R.

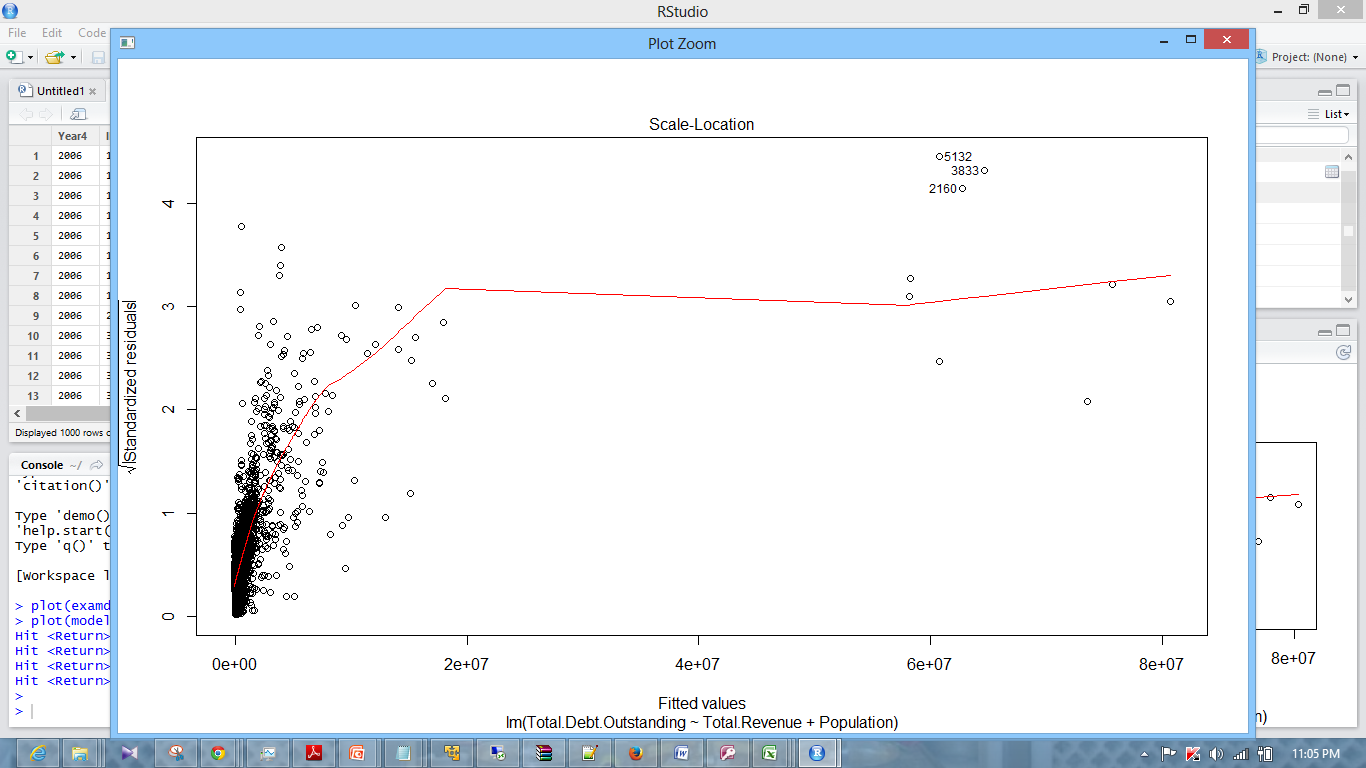


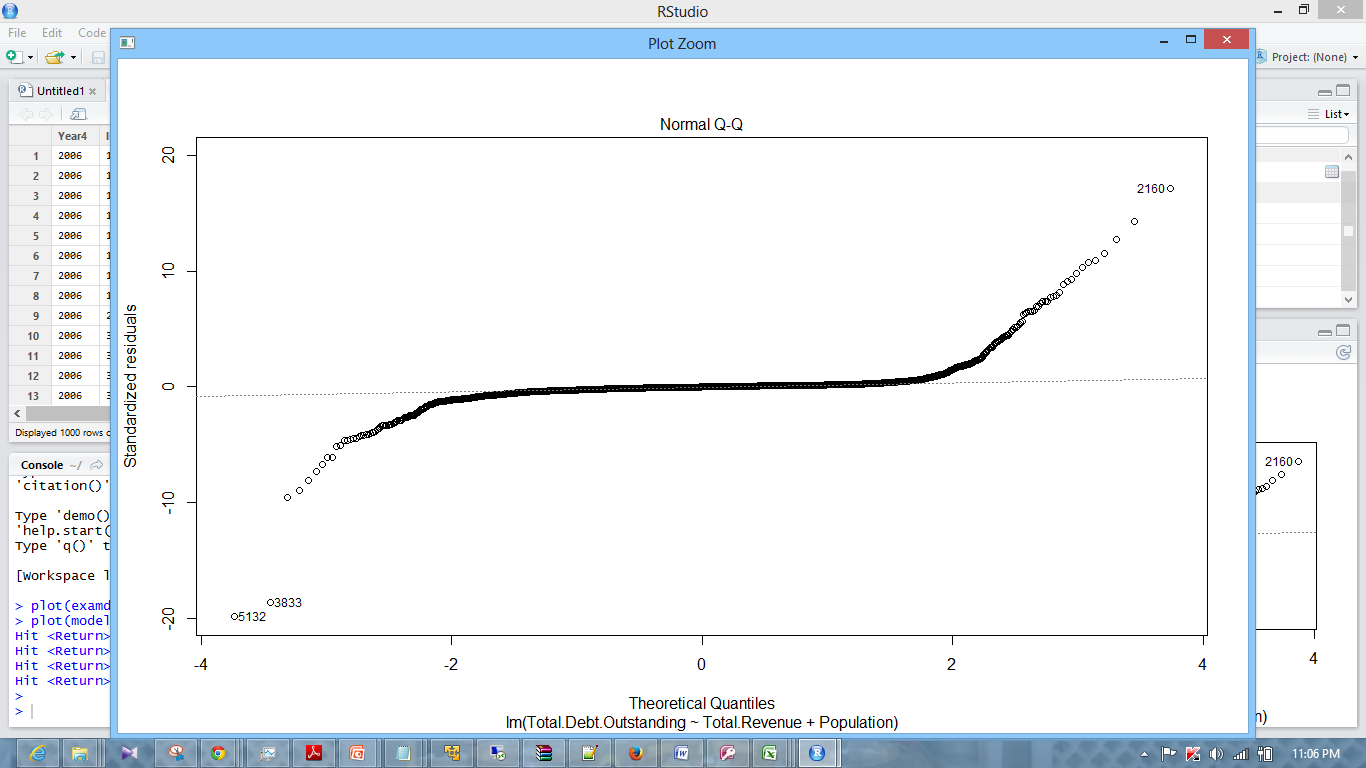


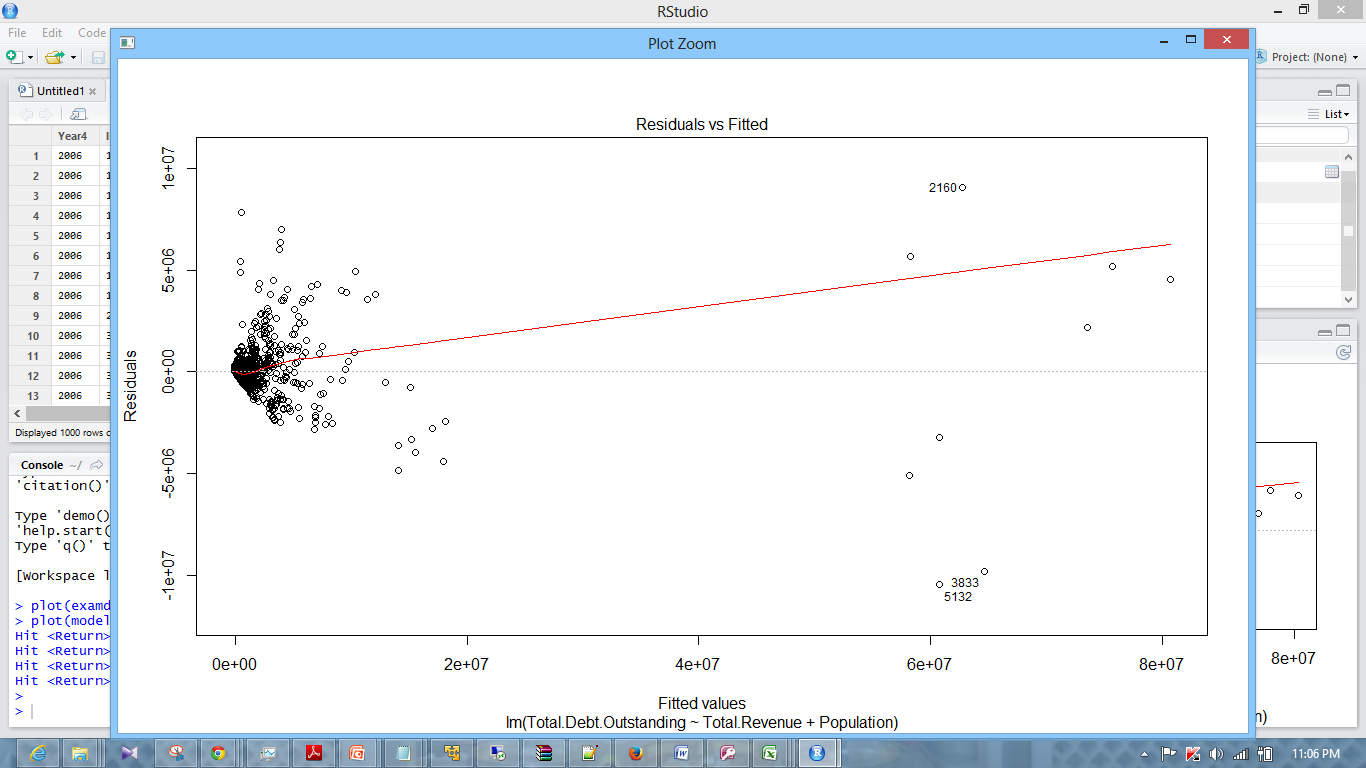


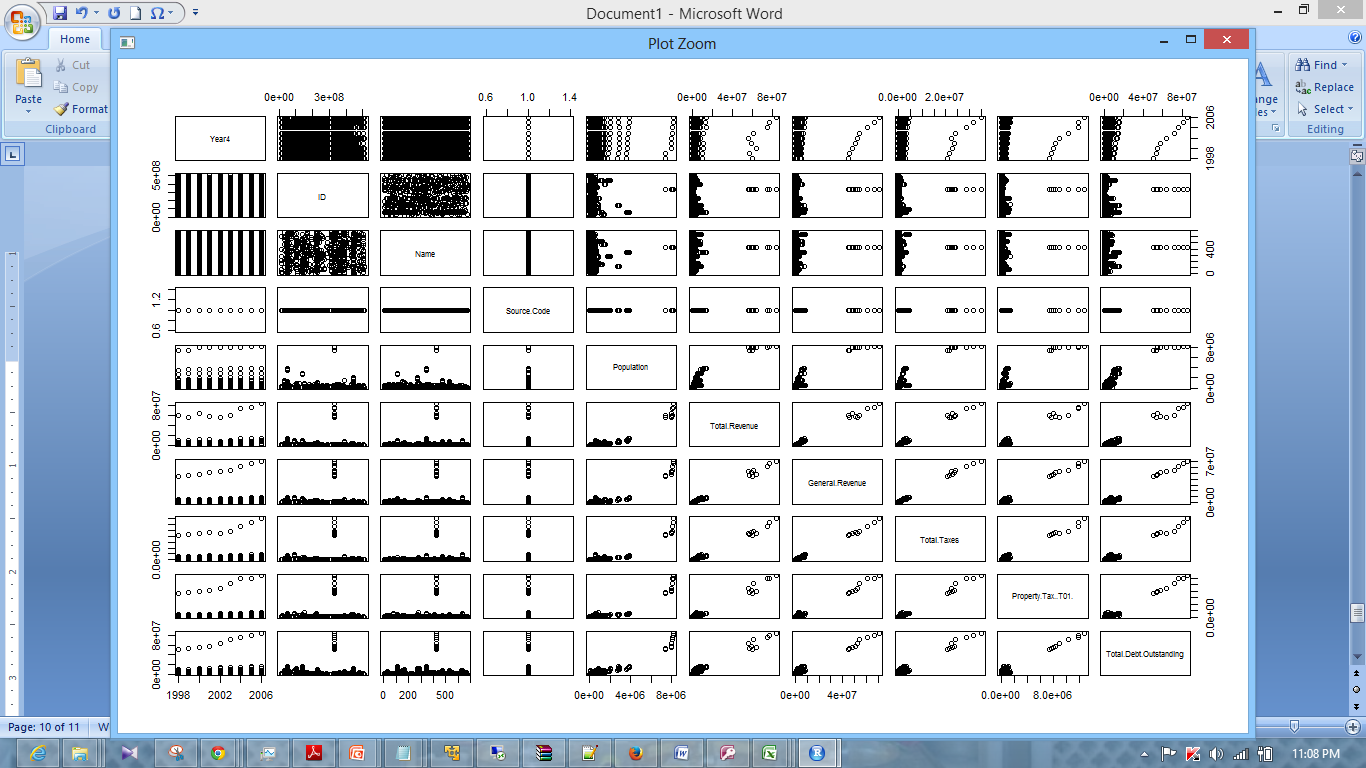
Following screenshots represents the plots of the Multiple Linear Regressions;

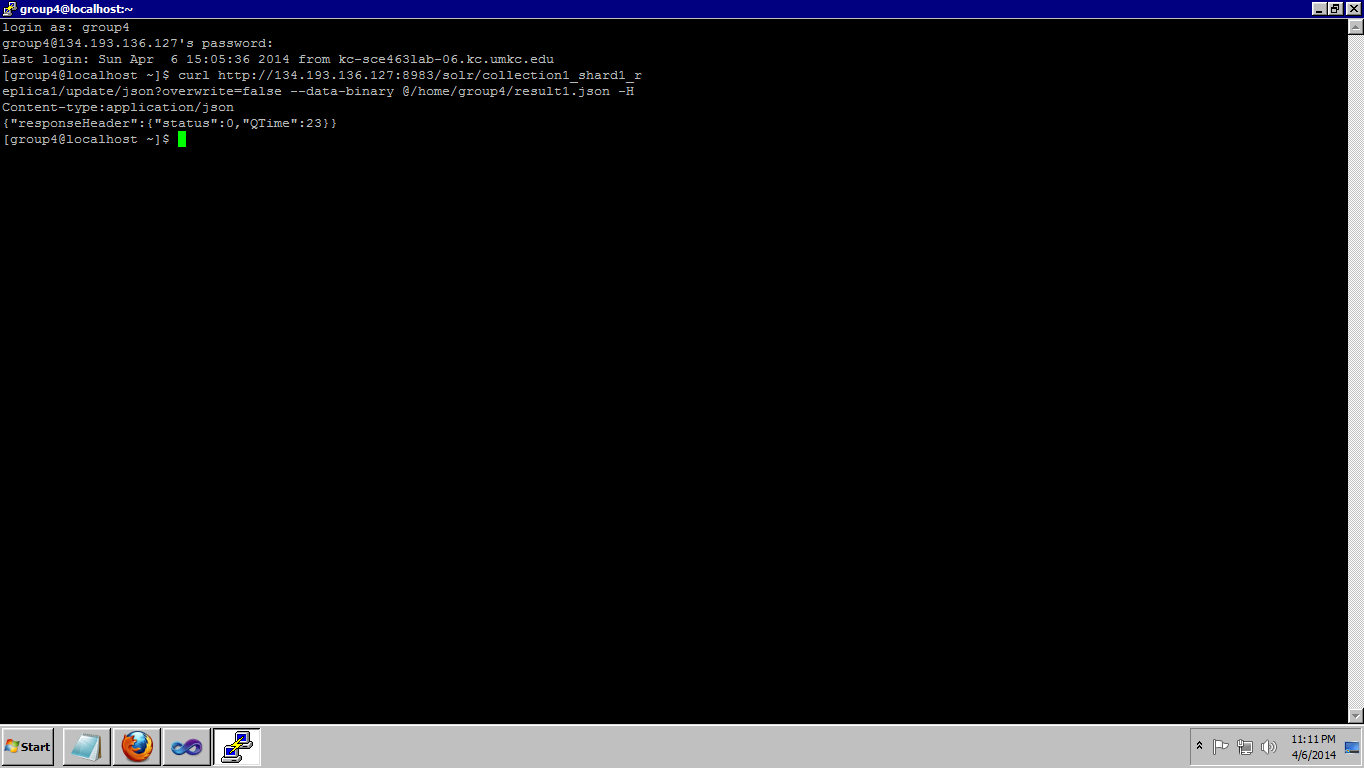




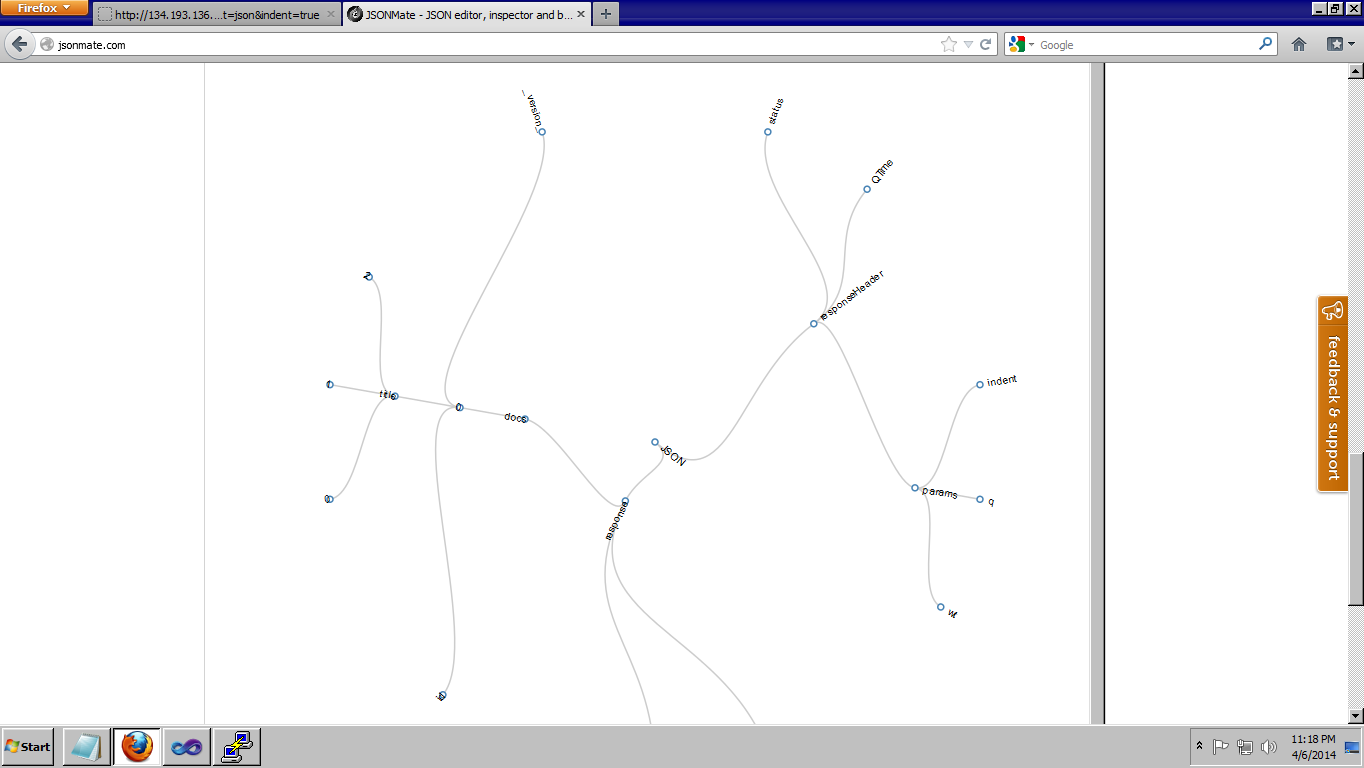












**Github URL:**

https://github.com/tmrhc/Knowledge-Discovery-Management/tree/master/Hackaton

http://134.193.136.127:8983/solr/collection1\_shard1\_replica1/select?q=id%3ATej&wt=json&indent=true

**Limitations:**

Limitations of the project that I have implemented are I was unable to run the Regression analysis on the variables strings. Since regression analysis on R doesn't allow to run on the strings I was limited to run the Regression analysis on the integer valued variables. I have designed a web service. But I was unable to parse the values of the json data into the webservice.

**References:**

**www.wikipedia.com**

**www.stackoverflow.com**

**www.getbootstrap.com**

**www.w3schools.com**