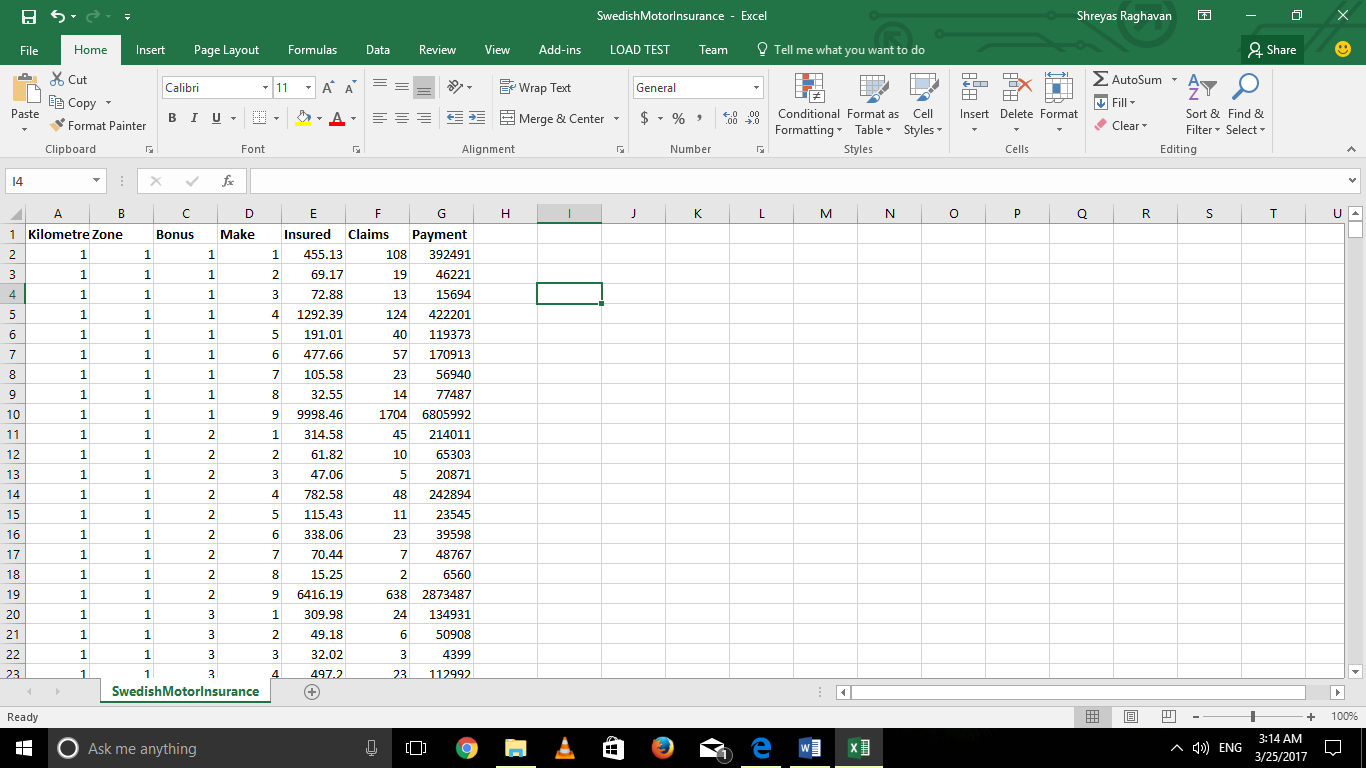
# Analyse the report of Swedish Motor Insurance

## Problem Statement:

The data gives the details of third party motor insurance claims in Sweden for the year 1977. In Sweden, all motor insurance companies apply identical risk arguments to classify customers, and thus their portfolios and their claims statistics can be combined. The data were compiled by a Swedish Committee on the Analysis of Risk Premium in Motor Insurance. The Committee was asked to look into the problem of analysing the real influence on the claims of the risk arguments and to compare this structure with the actual tariff.

## Dataset:



2182 x 7 is the dimension of the dataset.

As you could see from the dataset above, there are 7 variables.

Variable Description

Kilometre Kilometre travelled per year   
 1: <1000  
 2: 1000-15000  
 3: 15000-20000  
 4: 20000-25000  
 5: >25000

Zone Geographical zone

1: Stockholm, Göteborg, and Malmö with surroundings

2: Other large cities with surroundings

3: Smaller cities with surroundings in southern Sweden

4: Rural areas in southern Sweden

5: Smaller cities with surroundings in northern Sweden

6: Rural areas in northern Sweden 7: Gotland

Bonus No claims bonus; equal to the number of years, plus one, since the last claim

Make 1-8 represents eight different common car models. All other models are c combined in class 9.

Insured Number of insured in policy-years

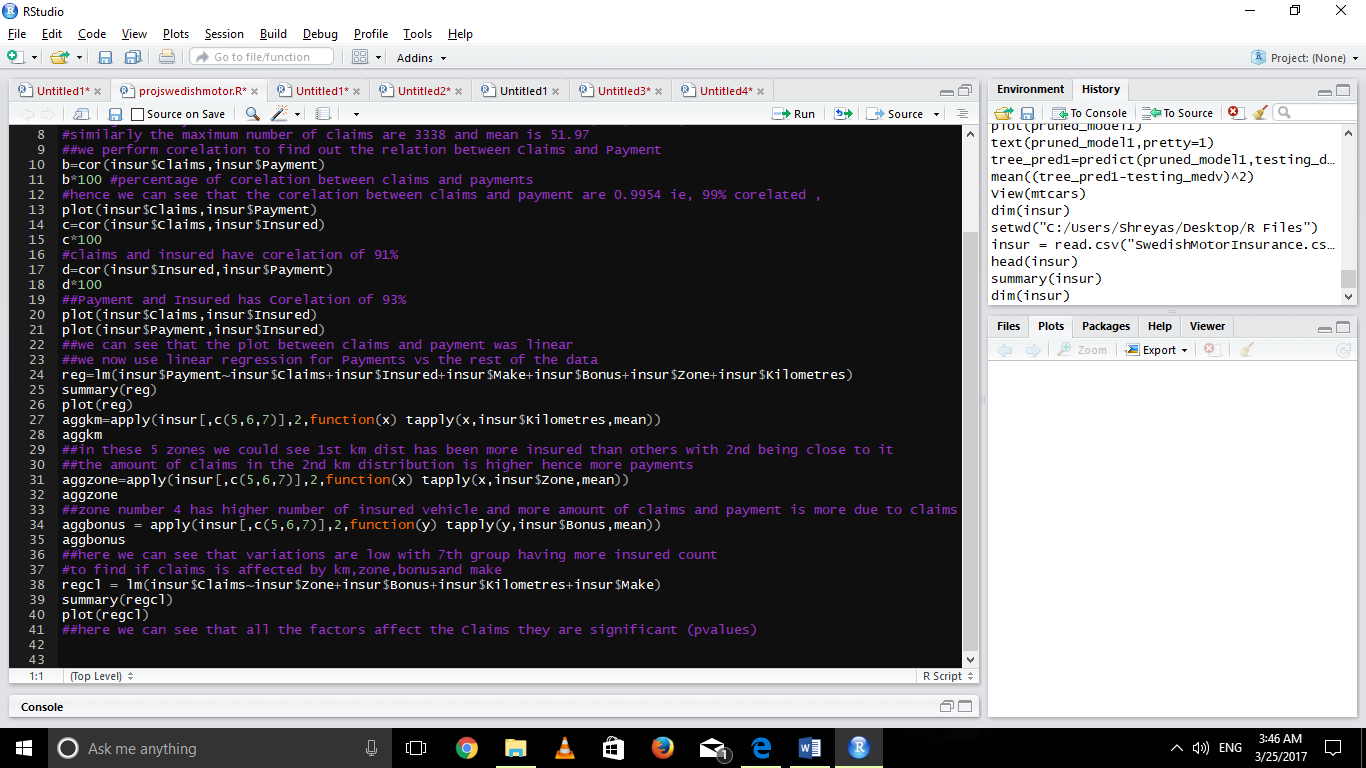
Claims Number of claims

Payment Total value of payments in Skr (Swedish Krona)

# Code:

# 

# 

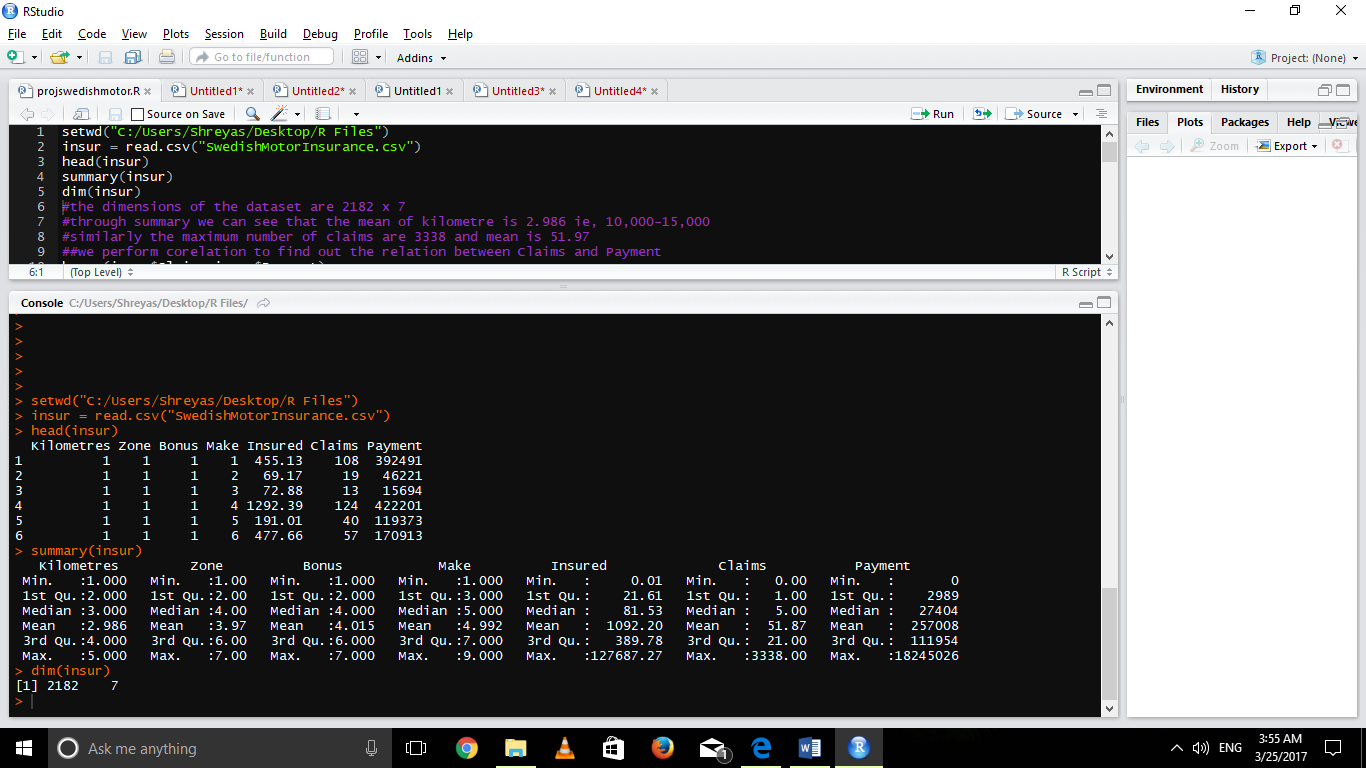


# Solution:

We start by importing the dataset using read.csv function and find out the head, summary and dimension of the data we’ll be working on.

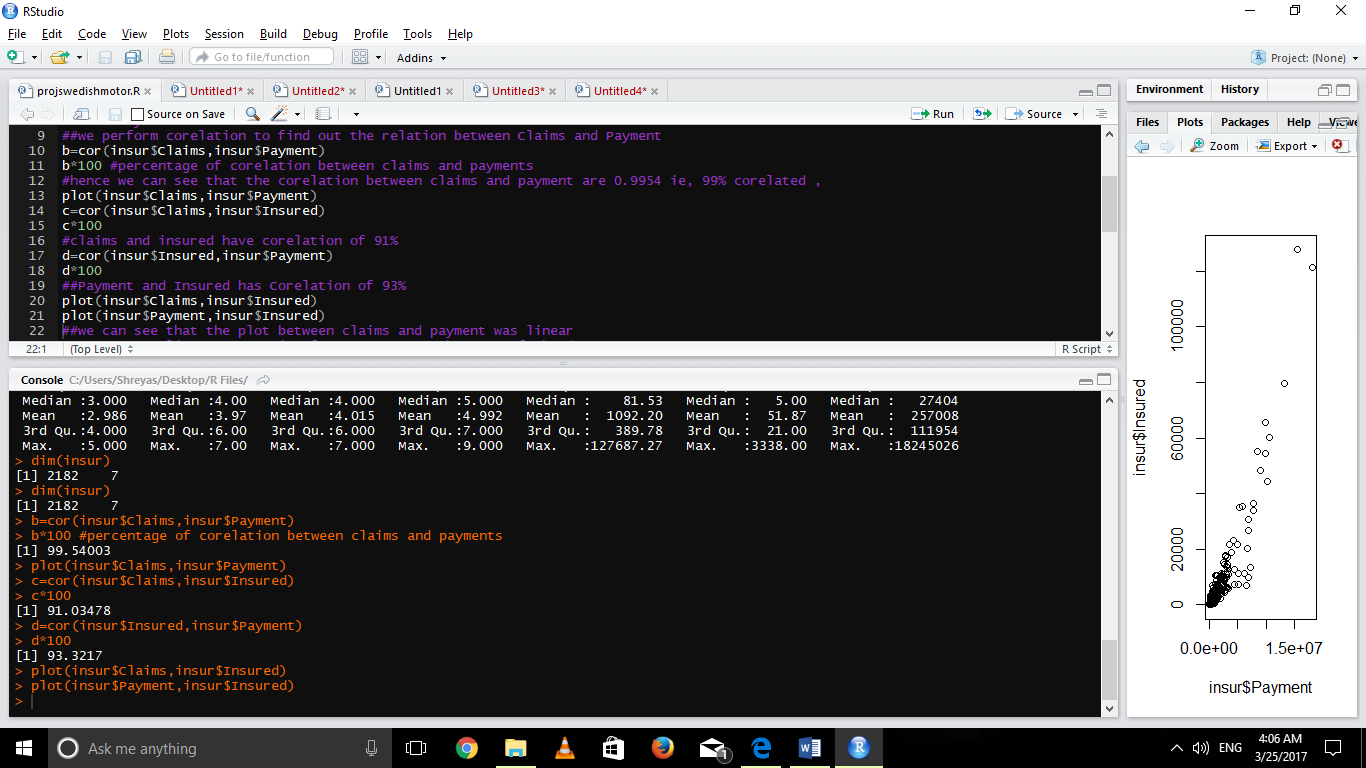
The problems are:

1. **The committee is interested to know each field of the data collected through descriptive analysis to gain basic insights into the data set and to prepare for further analysis.**

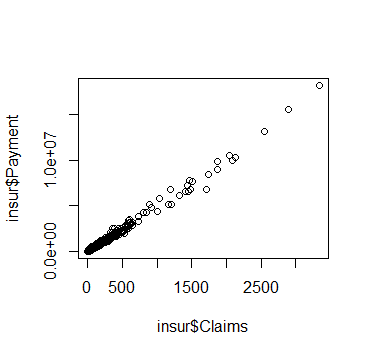


The result gives us the median, mean, quartile, max and min values. We can infer that insured minimum claims and payment is 0 from which we can deduce that **there are cars which are insured but there has been no insurance claimed.**

1. **The total value of payment by an insurance company is an important factor to be monitored. So, the committee has decided to find whether this payment is related to number of claims and the number of insured policy years. They also want to visualize the results for better understanding.**

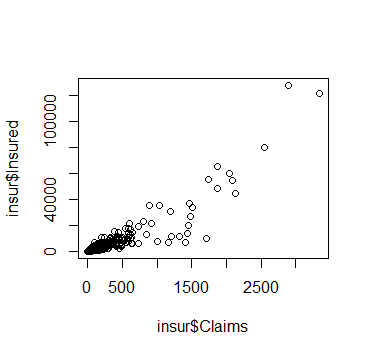


**The Claims vs Payment Graph:**



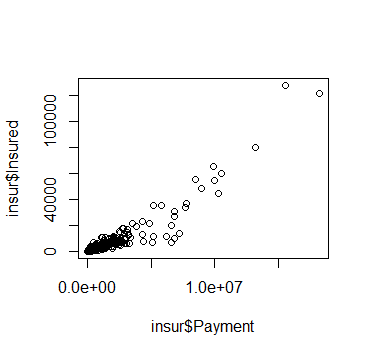
As we could see the graph is linear. The correlation between payment and claim is **99.54%.**

**The Claims vs Insured Graph:**



It follows the same linear trend. It has a correlation of **91.034%.**

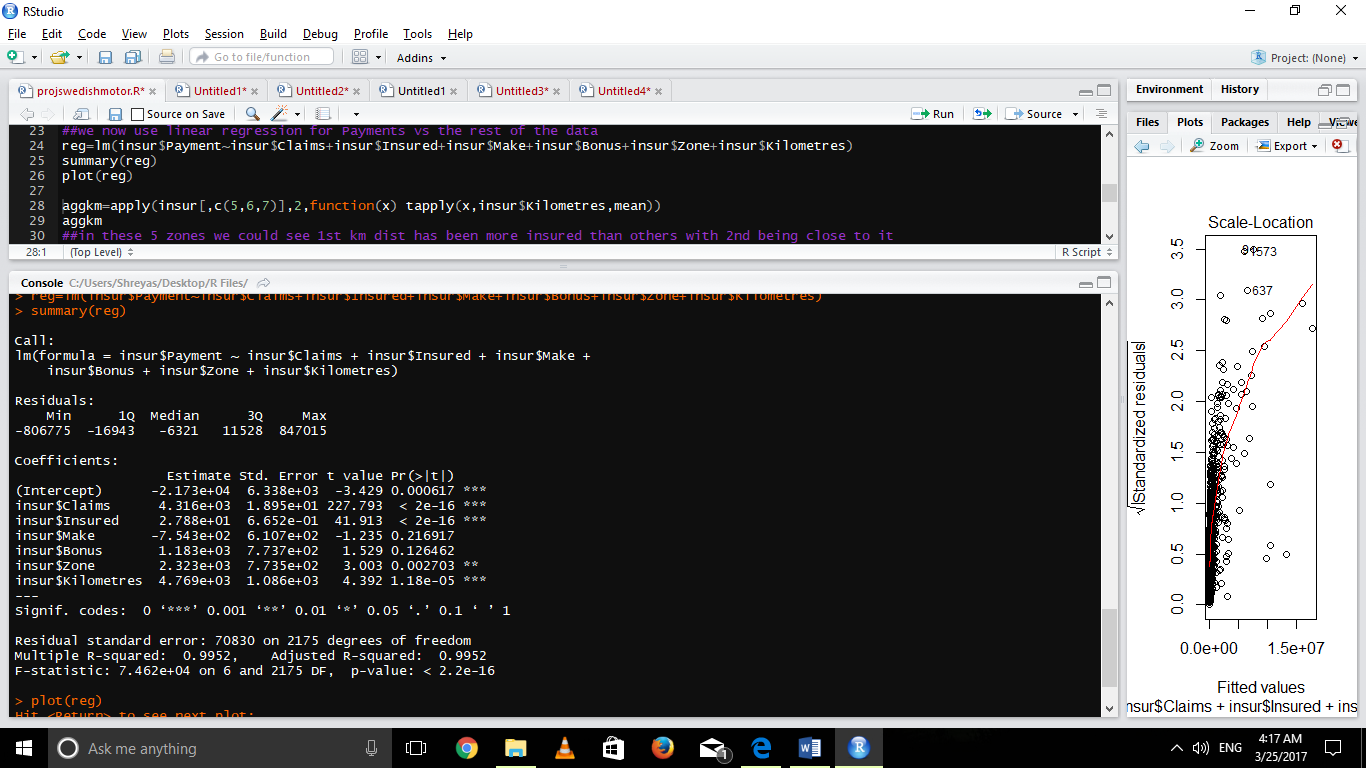
**The Payment vs Insured Graph:**



This graph is very similar to that of Claims vs Insured and the Correlation is **93.321%.**

**Hence the relation between Claims and Payments are correlated and as the value of claim increases the payment value will also increase.**

1. **The committee wants to figure out the reasons for insurance payment increase and decrease. So, they have decided to find whether distance, location, bonus, make, and insured amount or claims are affecting the payment or all or some of these are affecting it.**

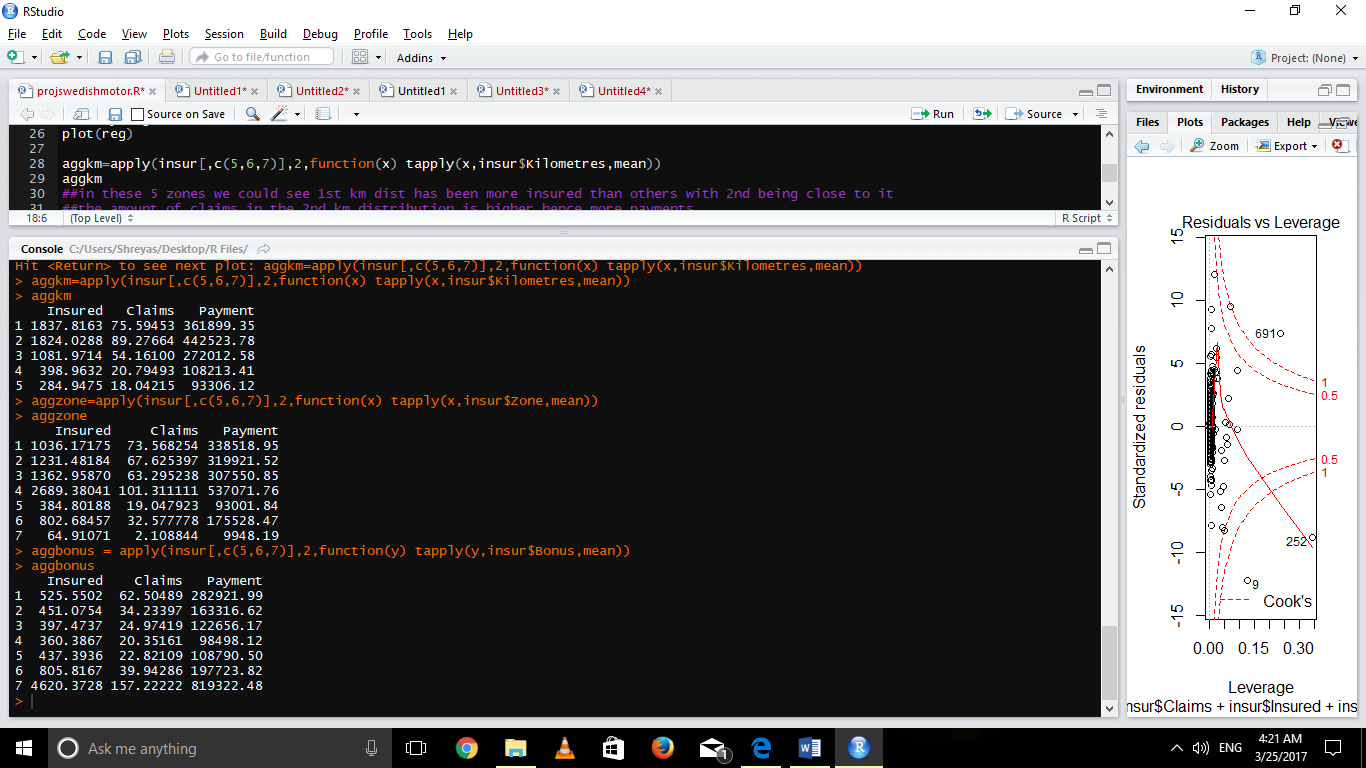
****

**Independent variable: insured, claims, make, bonus, zone, and kilometres.**

**Dependent variable: payment.**

From this we can see the regression lines (as seen on the left side) and this would help us predict the payment values. **The high p value of make and bonus show they do not make much influence on the payment.**

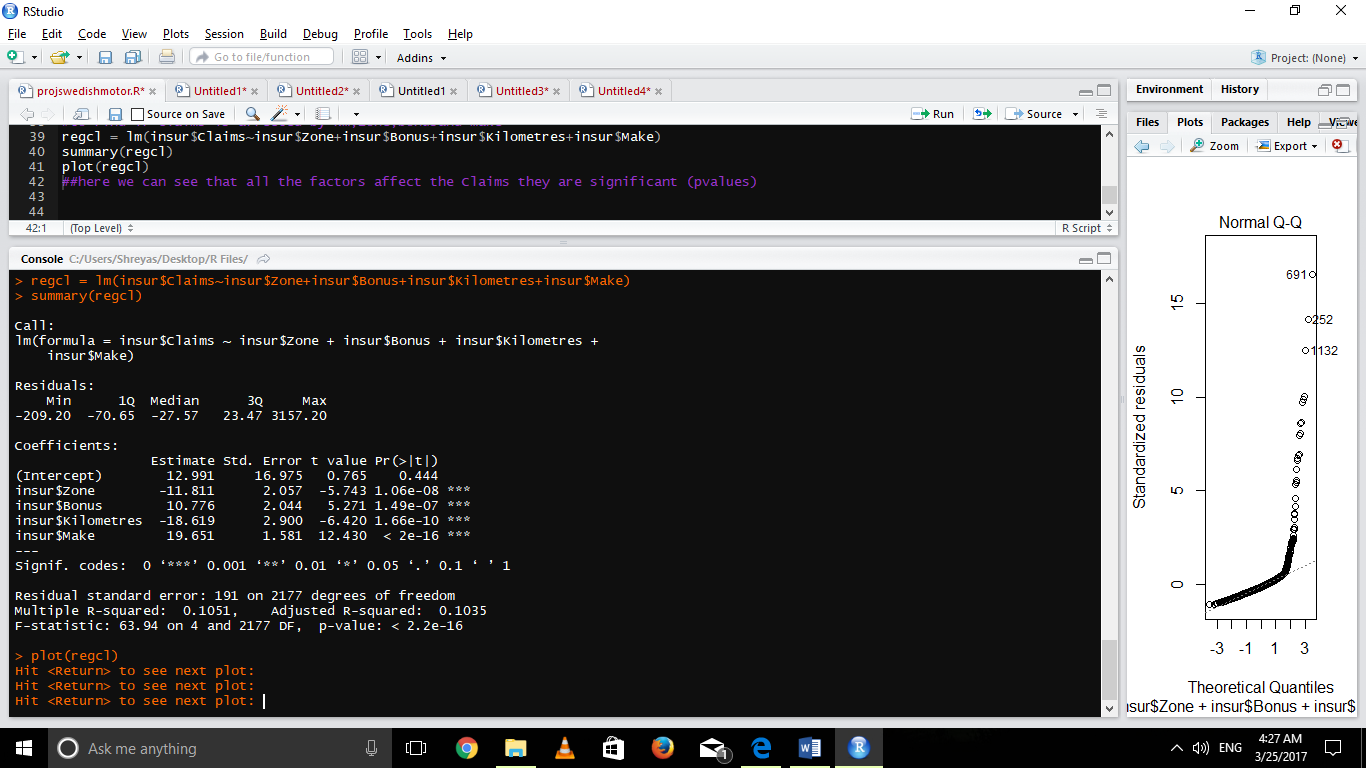
1. **The insurance company is planning to establish a new branch office, so they are interested to find at what location, kilometer, and bonus level their insured amount, claims, and payment get increased. (Hint: Aggregate Dataset)**

****

We find the mean value of value of insured, payment, and claims based on zone, kilometre, and bonus variables, group all the result variables based on individual categorical variables.

We can observe that

1. **Zone 1-4 have more insured years.**
2. **KM group2 has maximum payments though are insured lesser number of years than Group1 and even have higher claims than group1.**
3. **Zone 4 has highest claims hence higher payment.**
4. **Group 7 has extremely high amount of insured years, Payments, claims in groups of bonus.**
5. **The committee wants to understand what affects their claim rates so as to decide the right premiums for a certain set of situations. Hence, they need to find whether the insured amount, zone, kilometer, bonus, or make affects the claim rates and to what extent.**

****

**Dependent variable: claims**

**Independent variable: kilometres, zone, bonus, make, and insured**

In the last case we see all the **P values are significantly high hence we can assume that kilometer, zone , insured, bonus , make are making impacts on claims.**