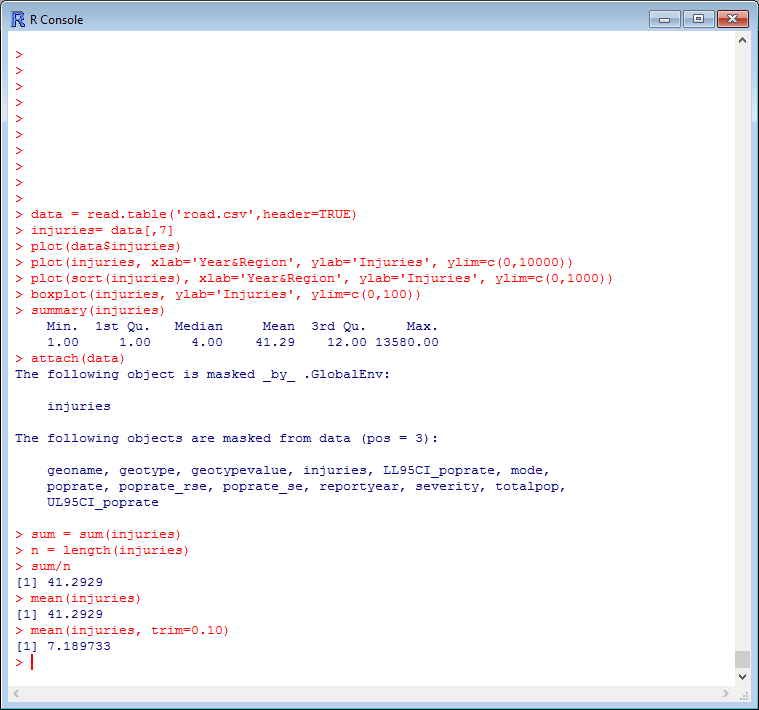
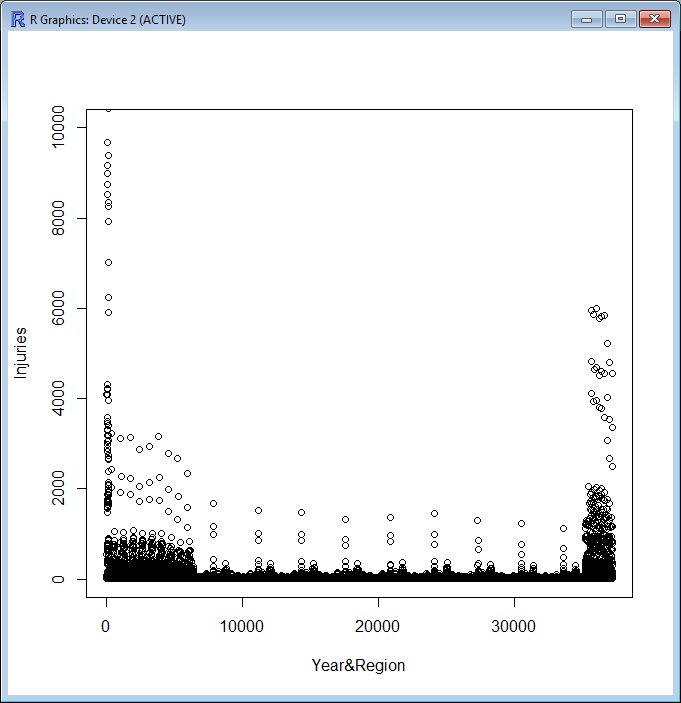
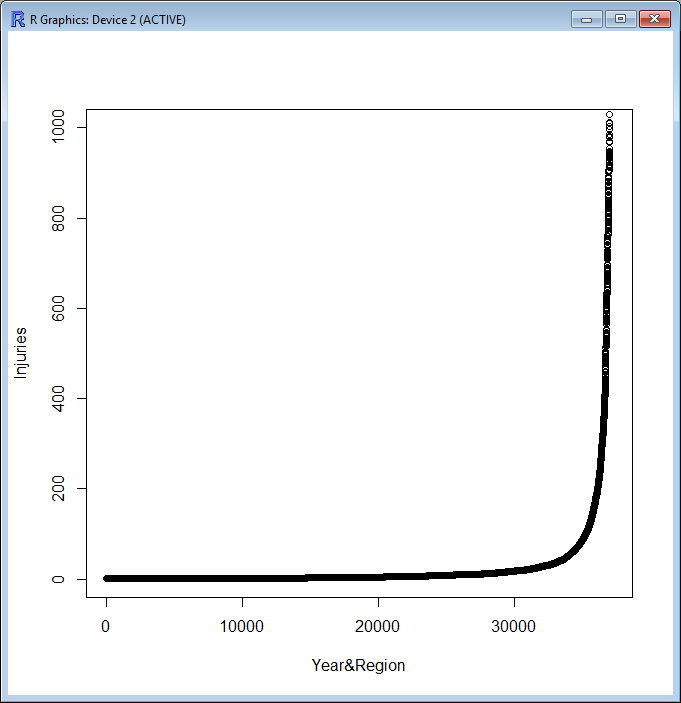
**Big Data Report**

İstifadə etdiyim data 2002-2010-cu illər üçün Kalifornia ştatında qəzalar zamanı yaralanmalar və itkiləri əks etdirən təqribən 32000 sətirdən ibarətdir.Əsas burada illərə görə itkilərin necə dəyişməsi və qəza növləri arasında münasibət araşdırlır.

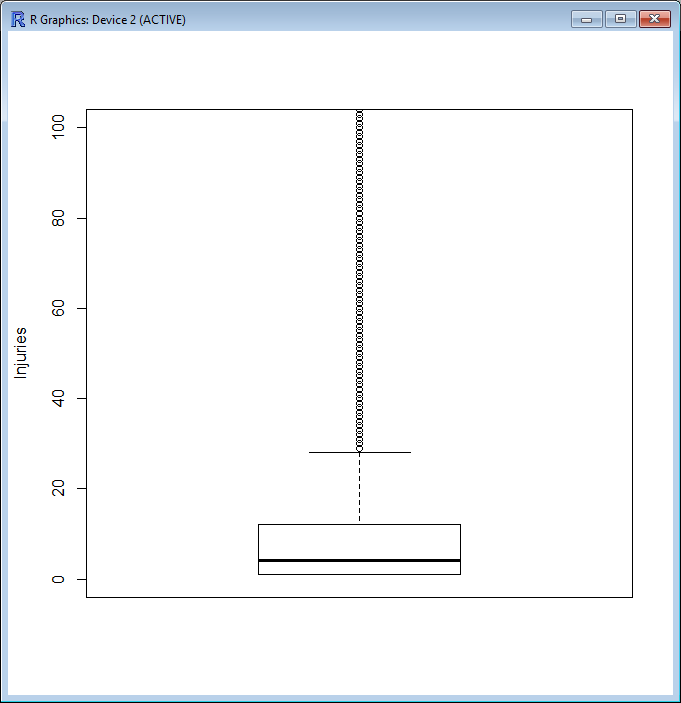
Əvvəlcə datanı import edək və başlayaq.  
  
Data vizual olaraq aşağıdakı kimi paylanmışdır.Burada biz **injuries** sütunu üçün baxırıq.



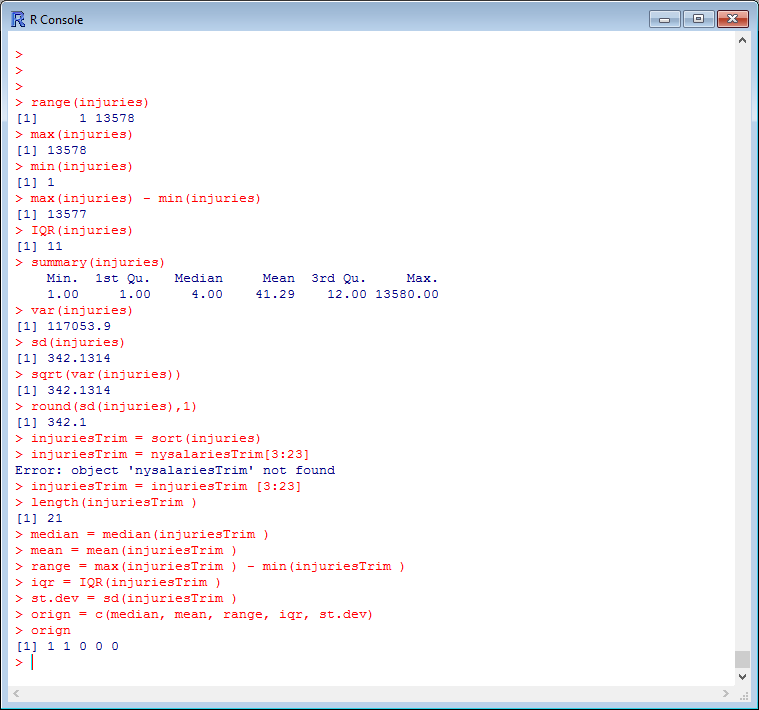
Datanı artan sırada düzəndən sonra vizual paylanma aşağıdakı kimi olur.

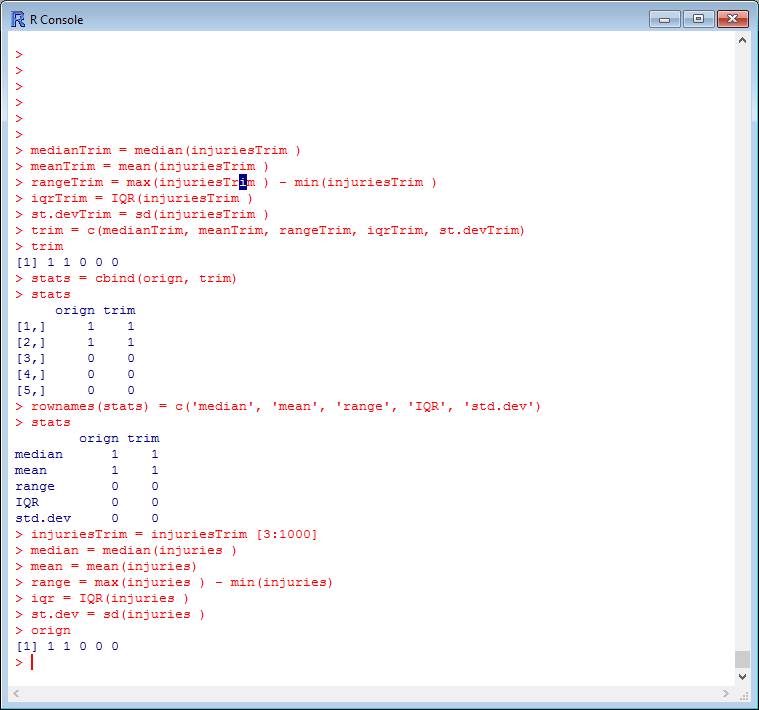


Bu data üçün boxplot quraq.Göründüyü kimidatanın əsas hissəsi 0-20 arasında dəyişir.Amma buna baxmayaraq outlayer-lər çoxdur.

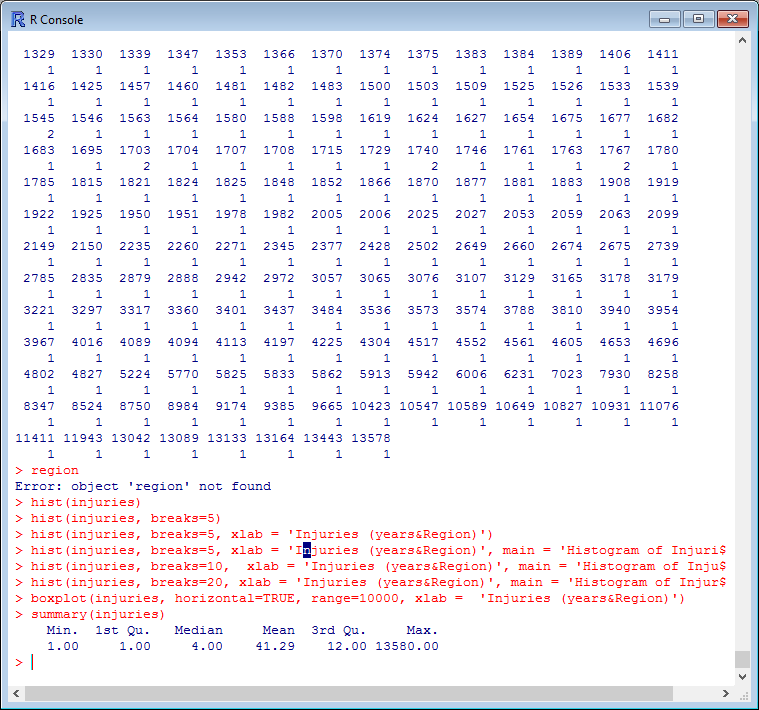


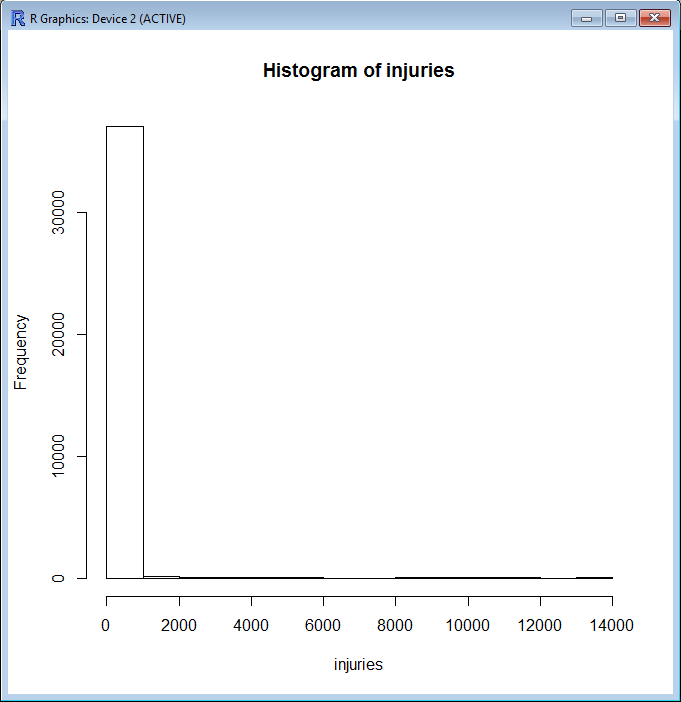
Data üçün Var,SD ni hesablayaq.Datanı trim edərək bu qiymətlərin necə dəyişməsinə baxaq.



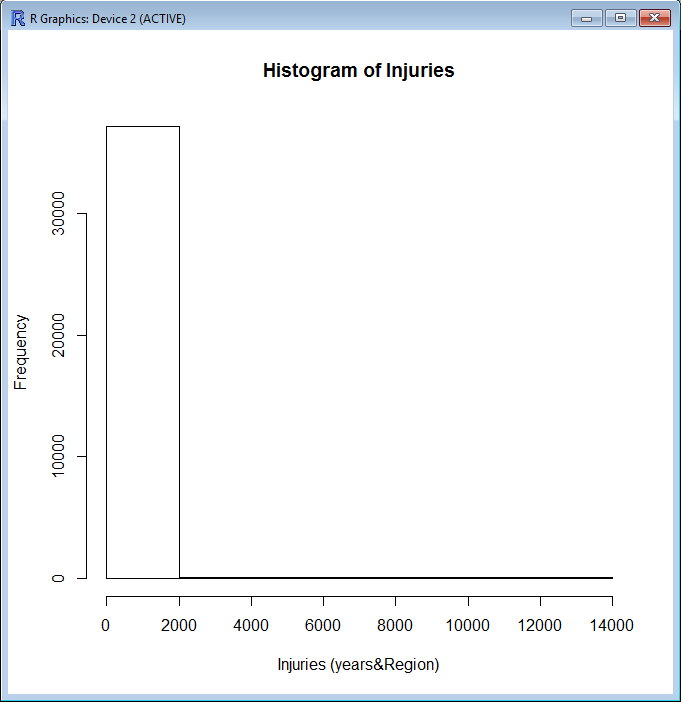


Datanın təsviri üçün üçün histogram quraq.

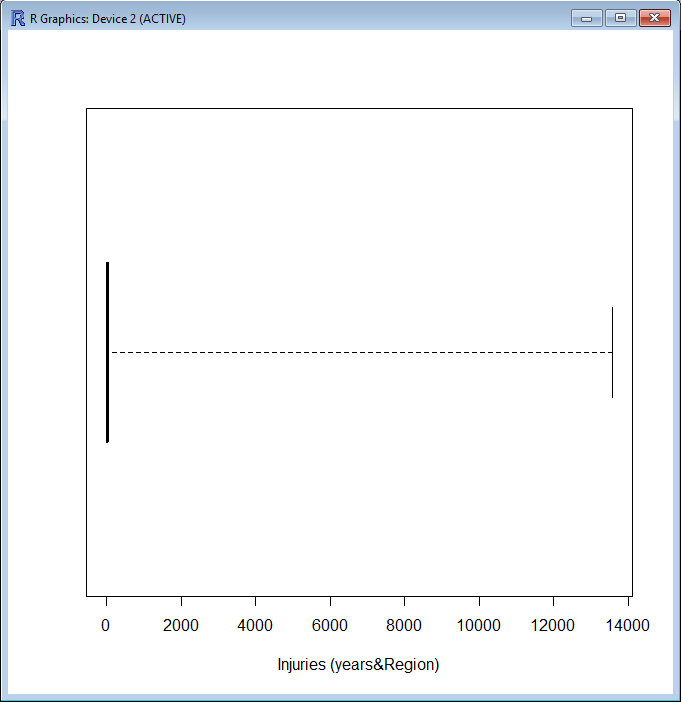




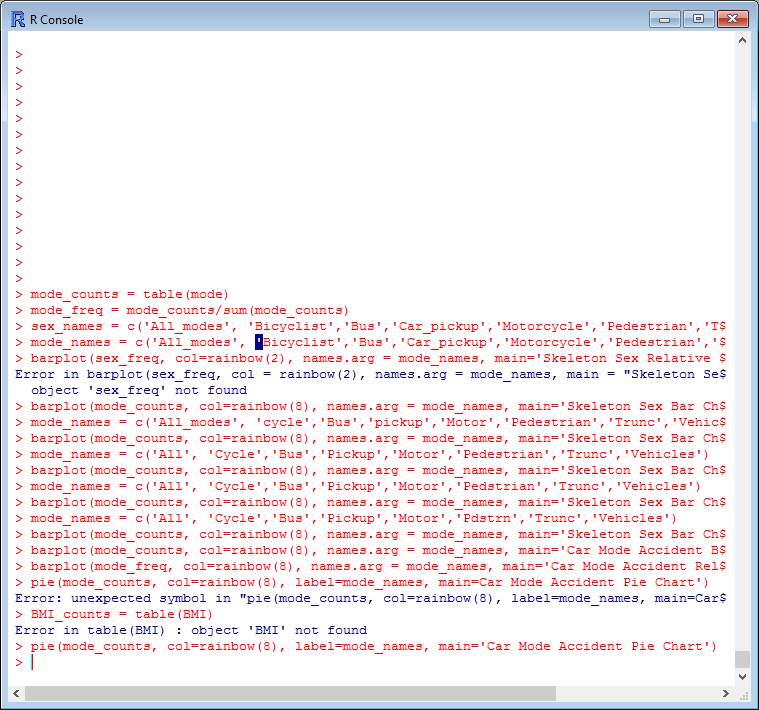
Bölünmələrin sayı 5 olduqda histoqramı yenidən quraq.

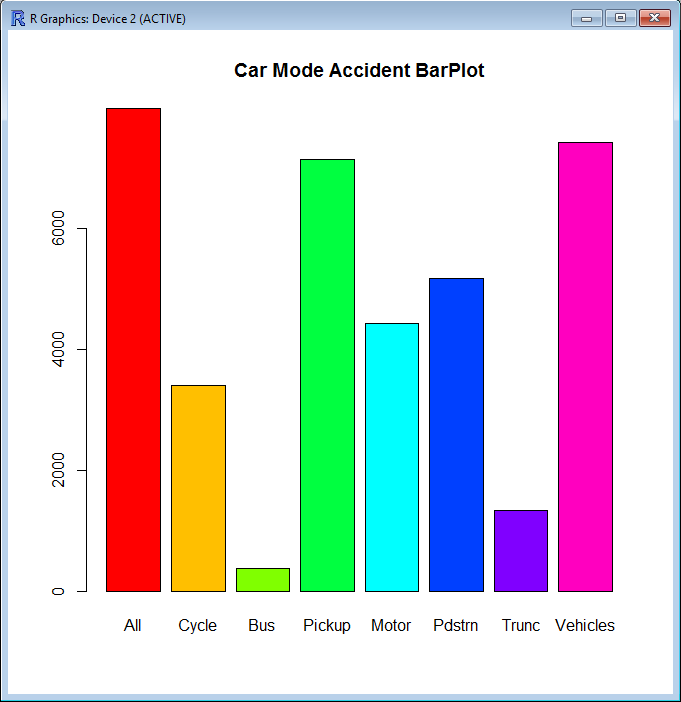


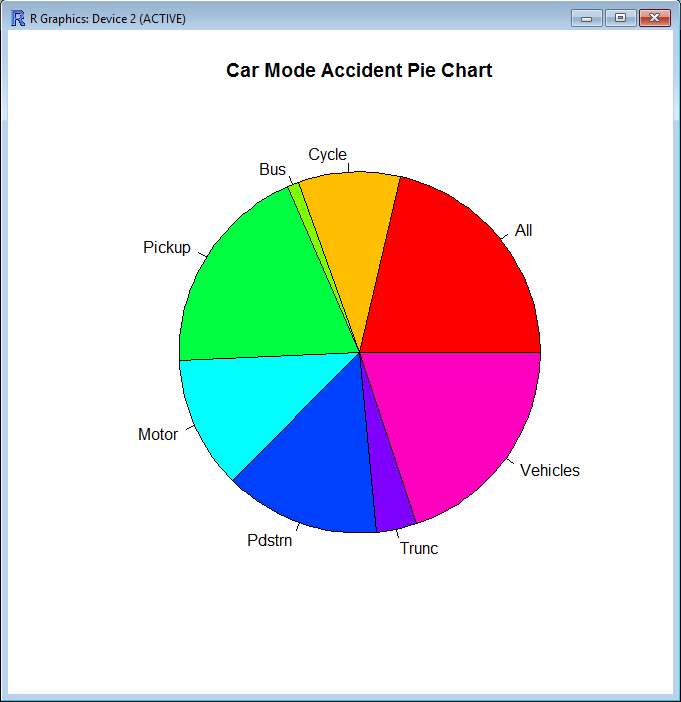
Data üçün horizontal boxplot quraq.



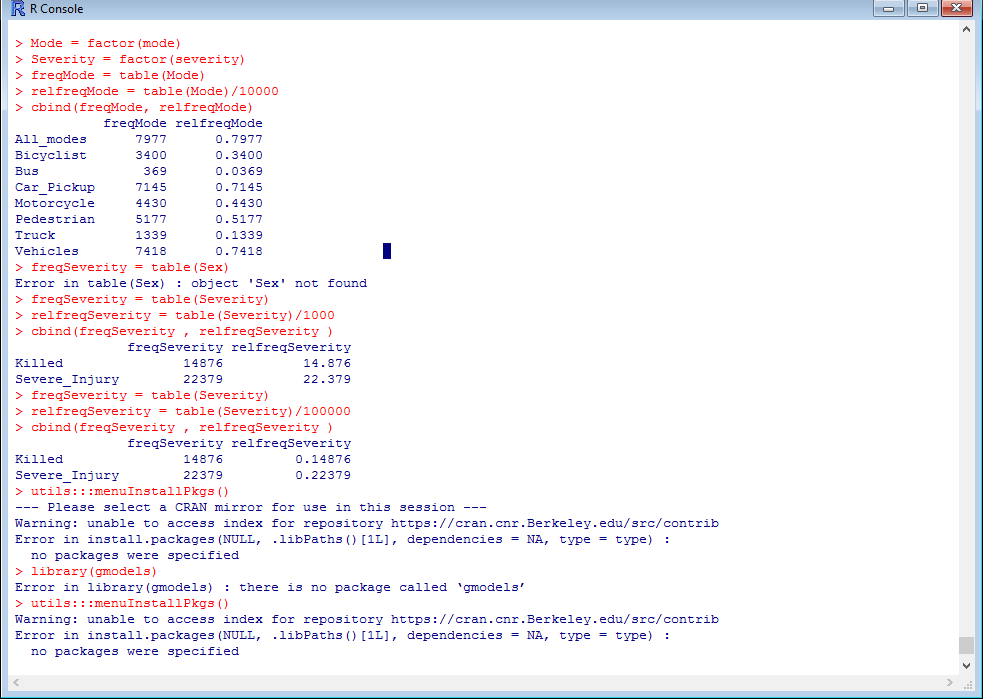
Datanı vizuallaşdırmaq üçün barplot və pie-chart quraq.

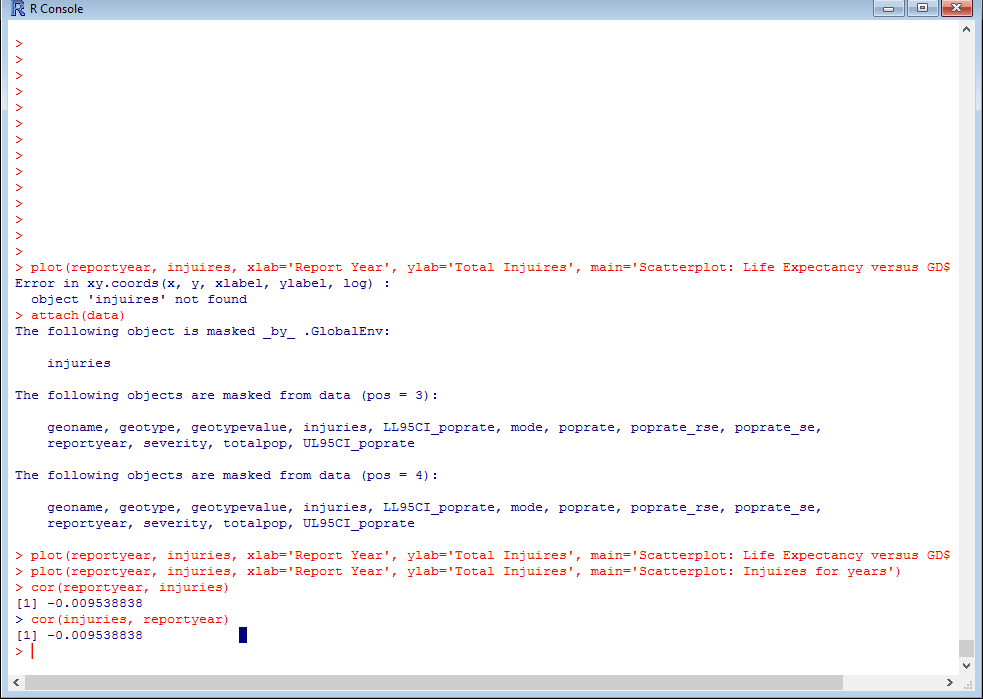


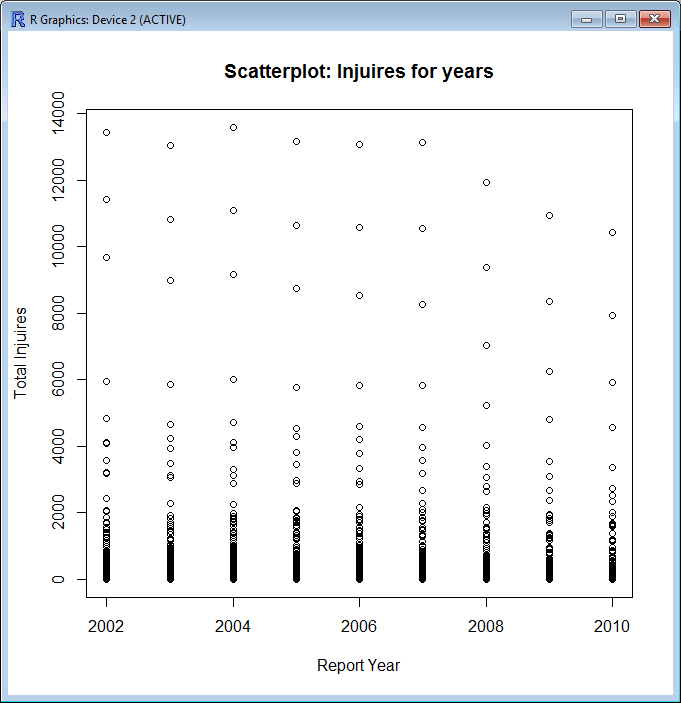




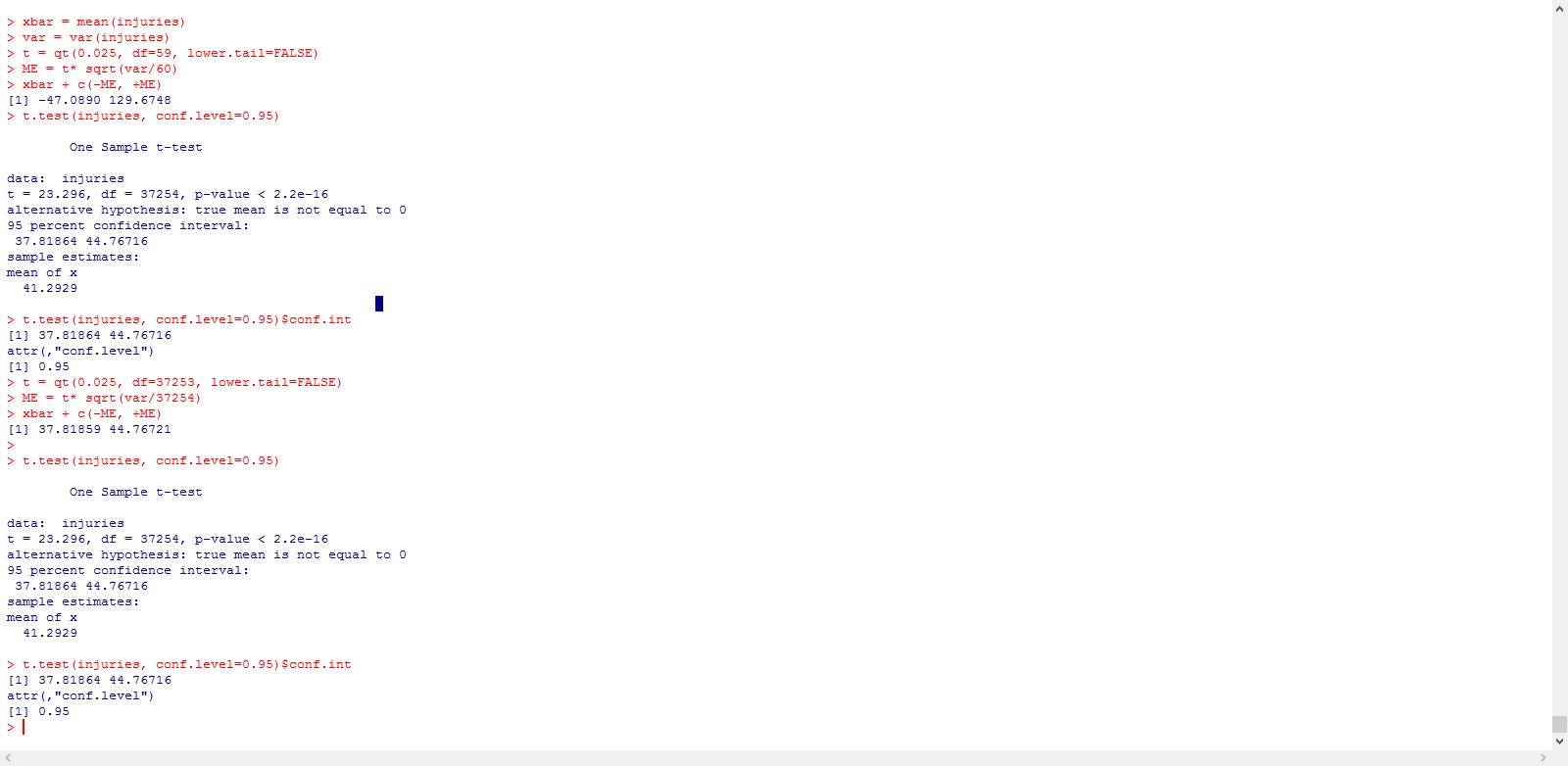
Mode və Severity kateqorik dəyişənləri üçün münasibətlərə baxaq.Relative Frequency-ləri hesablayaq.



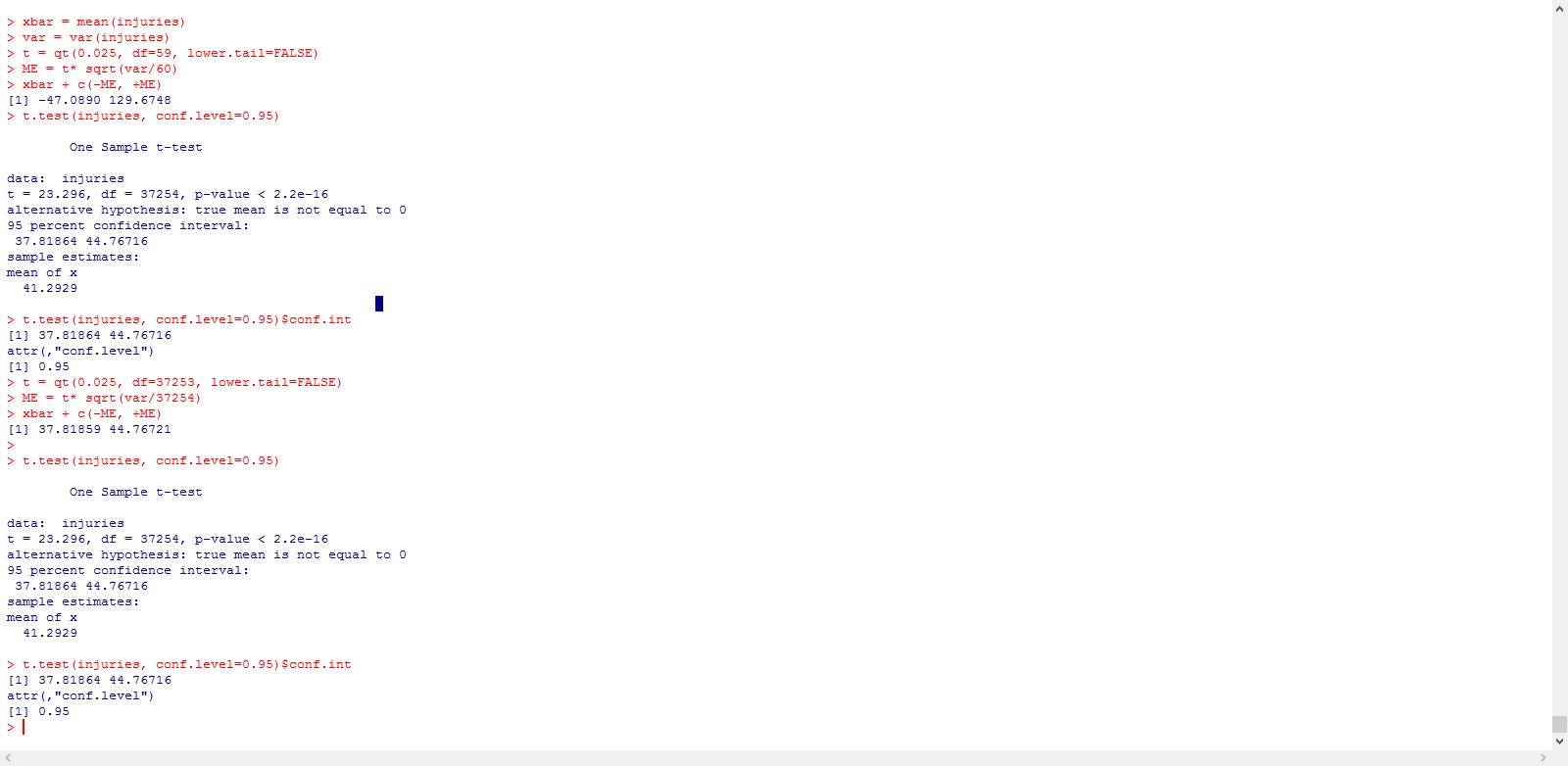


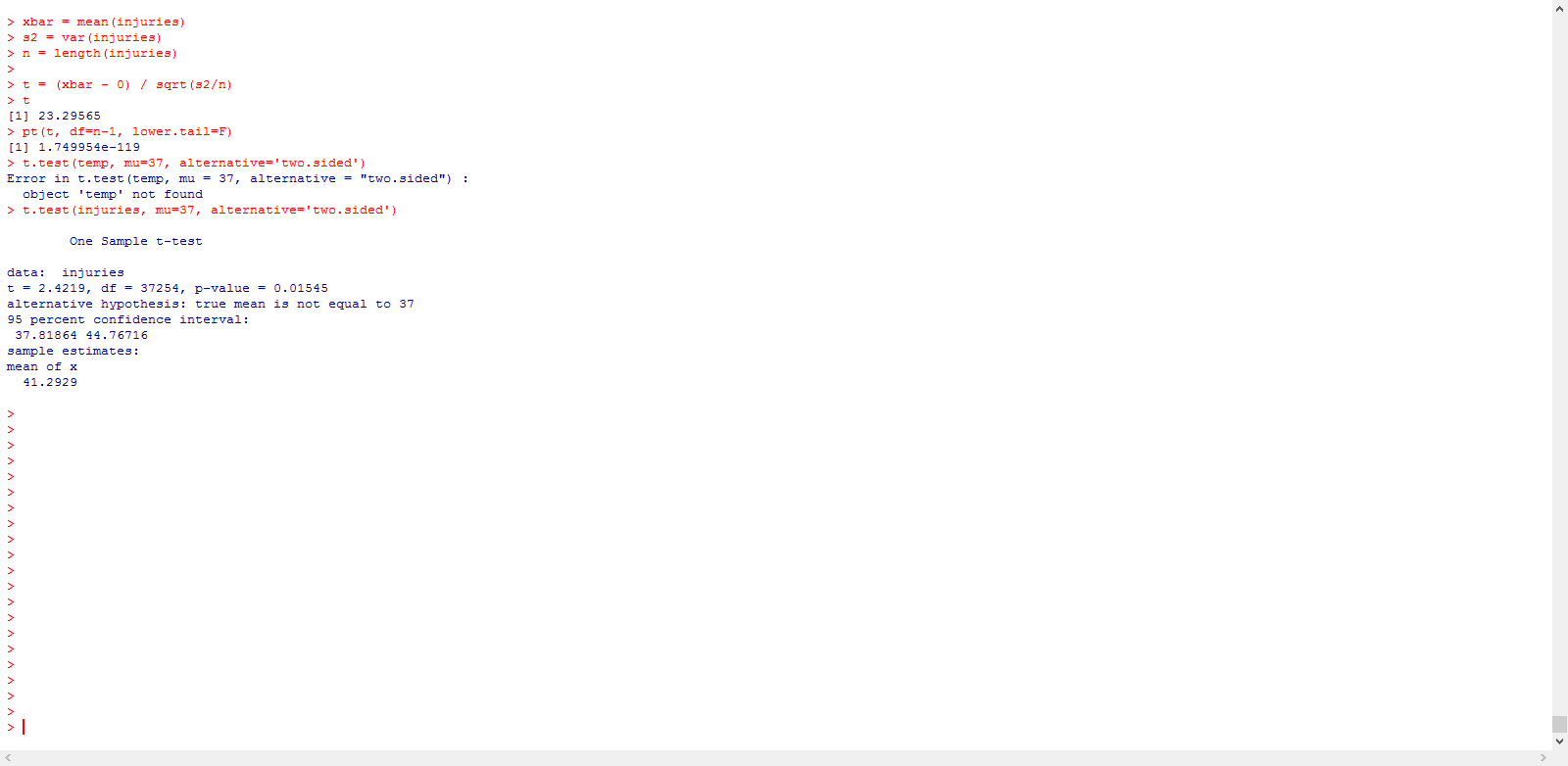


T paylanma və ME(Margin of Error) –ni bizim data üçün hesablayaq.

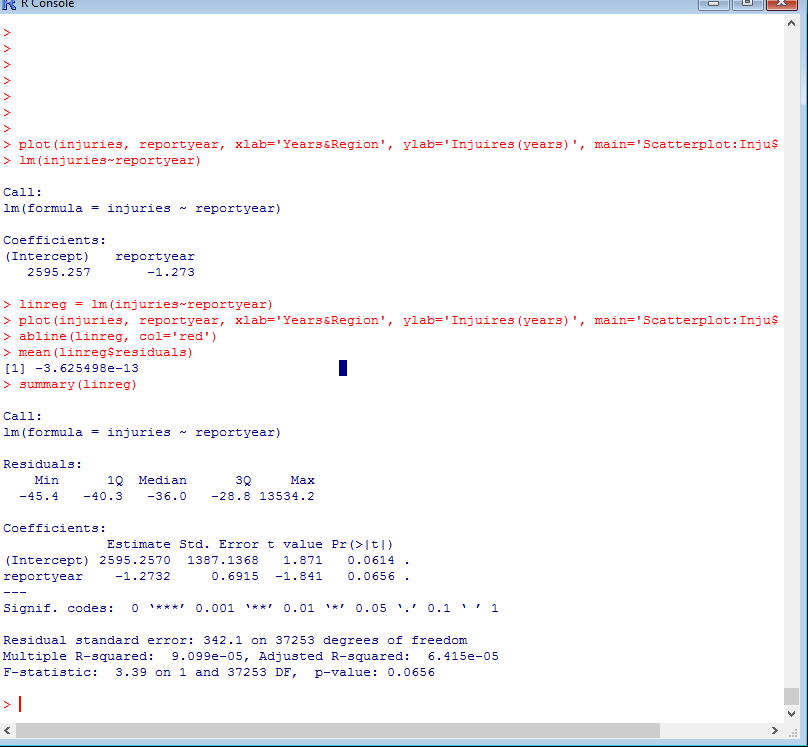


Data üçün null hipotezis və p-value-ni hesablayaq.

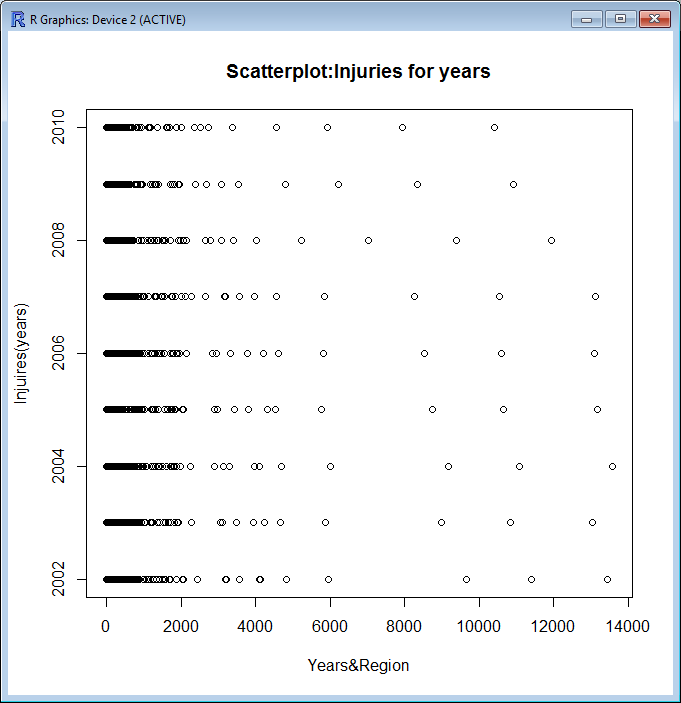




Data üçün Xətti reqresiya məsələsinə baxaq.



İlk öncə data-nın illər üzrə necə paylanmsaına baxaq.



İllər və İtki və Yaralanmaların asılığından ibarət xətti reqresiya məsələsinə baxaq.Kəmiyyətləri loqarifmləyib yenidən münasibətə baxaq.

