**Daugiamačių duomenų vizualizacija**

**3 laboratorinio darbo ataskaita**

**Ignas Jatulis 1815628**

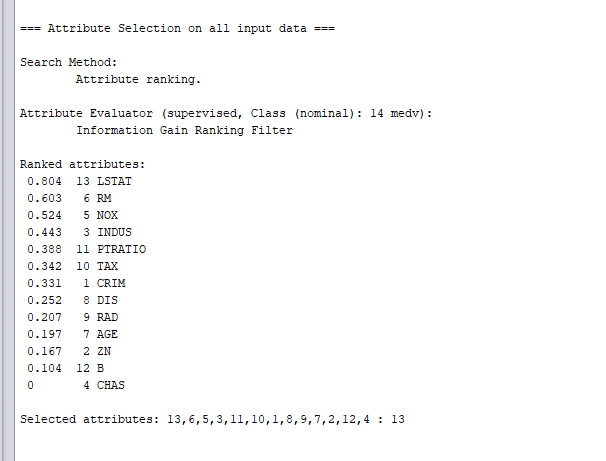
**Trečioji užduotis:**

*Atlikite pagrindinių komponenčių analizę daugiamačiams duomenims. Nustatykite, kokia procentinė dalis nuo visos dispersijos tenka kiekvienai komponentei, kiek tenka dviem pagrindinėm. Vizualizuokite dvi pagrindines komponentes. Galima naudoti* [*Orange*](http://orange.biolab.si/)*, Matlab, octave ar kitas programas.*

**Sprendimas:**

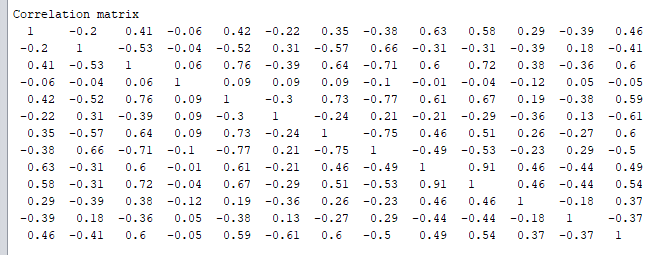
Prieš nustantant kokia proncentinė dalis tenka kiekvienai komponentai, klasės atributo duomenys buvo diskretizuoti pasitelkiant *WEKA* programinę įrangą. Klasės atributas *medv* buvo sutrauktas į penkis intervalus.

Toliau programinio paketo *WEKA* pagalba, buvo apskaičiuota kiekvieno atributų dispersija:

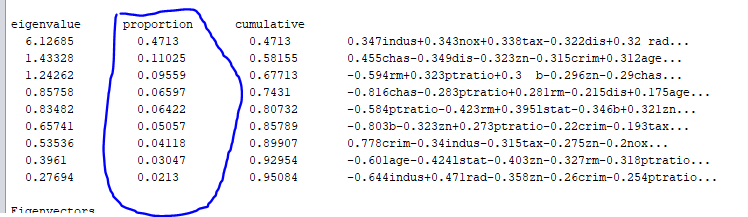


Toliau, pritaikius *PrincipalComponents* įvertinimo metodą, buvo gauta:

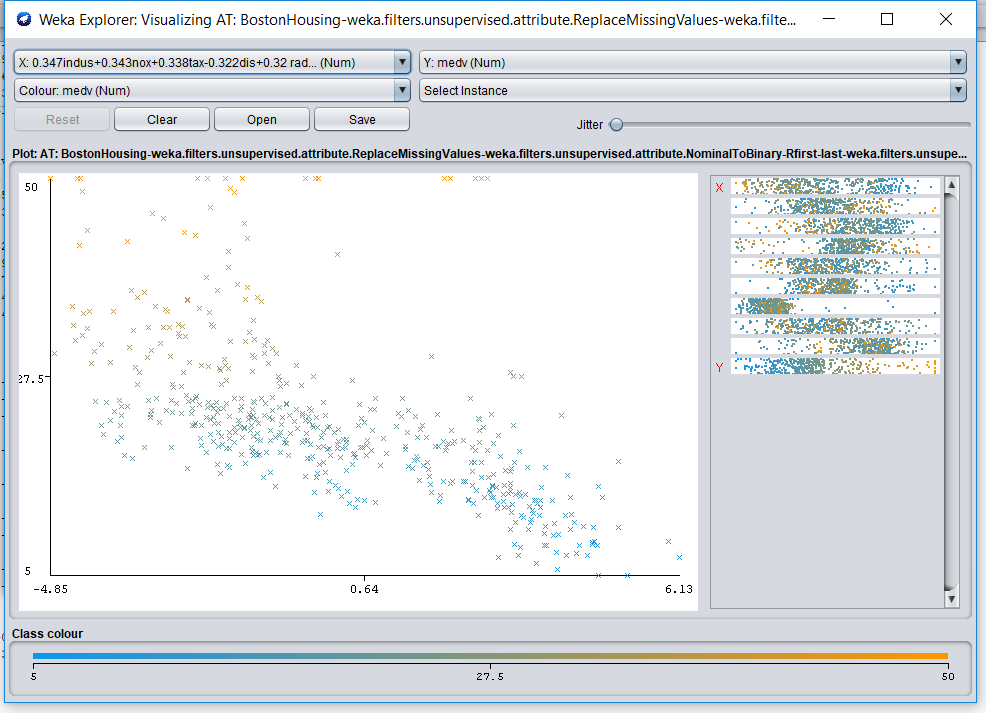
Koreliacijos matrica:



Buvo nustatyta, kokia procentinė dalis nuo visos dispersijos tenka kiekvienai komponentei



Pagrindinių komponenčių vizualizacija:





**Išvada**:

Matome, kad labiausiai yra atsiskyrę intervalai, kur būsto kainos yra > 27.5 (tūkst.)