**Daugiamačių duomenų vizualizacija**

**2 laboratorinio darbo ataskaita**

**Ignas Jatulis 1815628**

**Antroji užduotis:**

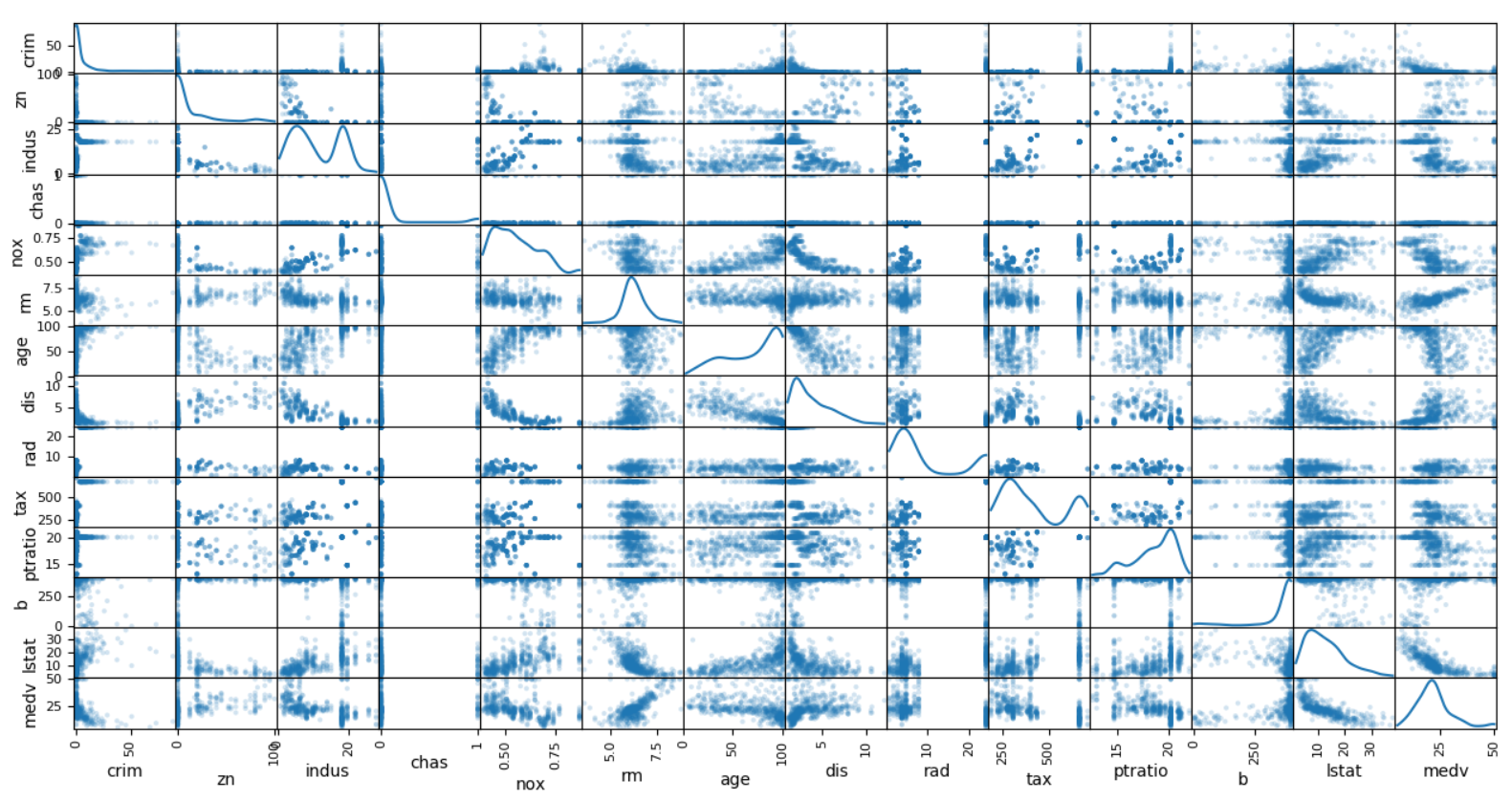
*Vizualizuokite skaitinius daugiamačių duomenų parametrus bent trimis tiesioginio vizualizavimo metodais. Galima naudoti Matlab (*[*Statistical Toolbox*](http://www.mathworks.se/help/stats/multivariate-visualization-1.html)*),* [*Orange*](http://orange.biolab.si/)*,* [*Visulab*](http://www.inf.ethz.ch/personal/hinterbe/Visulab/)*,* [*Xmdv*](http://davis.wpi.edu/xmdv/)*,* [*Pandas*](http://pandas.pydata.org/) *ar kitas programas.*

**Sprendimas:**

Duomenų vizualizavimui buvo panaudota Python programavimo kalba ir jos biblioteka Pandas. Su ja buvo realizuotas tiek pasrinkto duomenų failo, pavertus į CSV formatą, nuskaitymas, tiek grafikų generavimas.

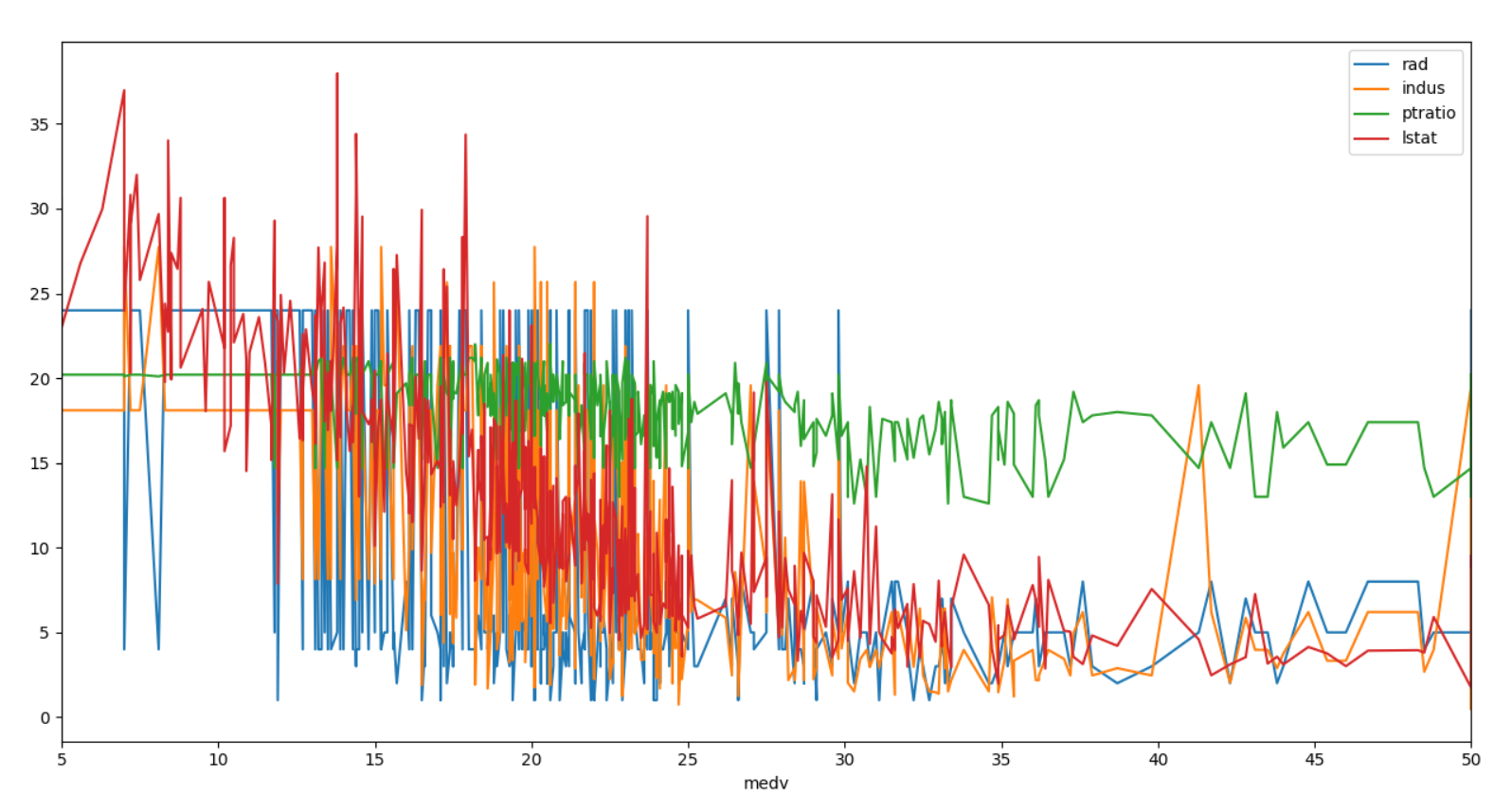
**Taškinių grafikų matrica**

Taškinių grafikų matricoje galima pamatyt, kaip kos duomuo yra susijęs tiek su kitu, tiek kaip pačio atributo duomenys yra pasiskirstę. Šiame grafike galime pamatyti, kad pavyzdžiui jei būstai yra seni (*age* atributas), tai tada jų vertė yra maža, ko ir galima buvo tikėtis.



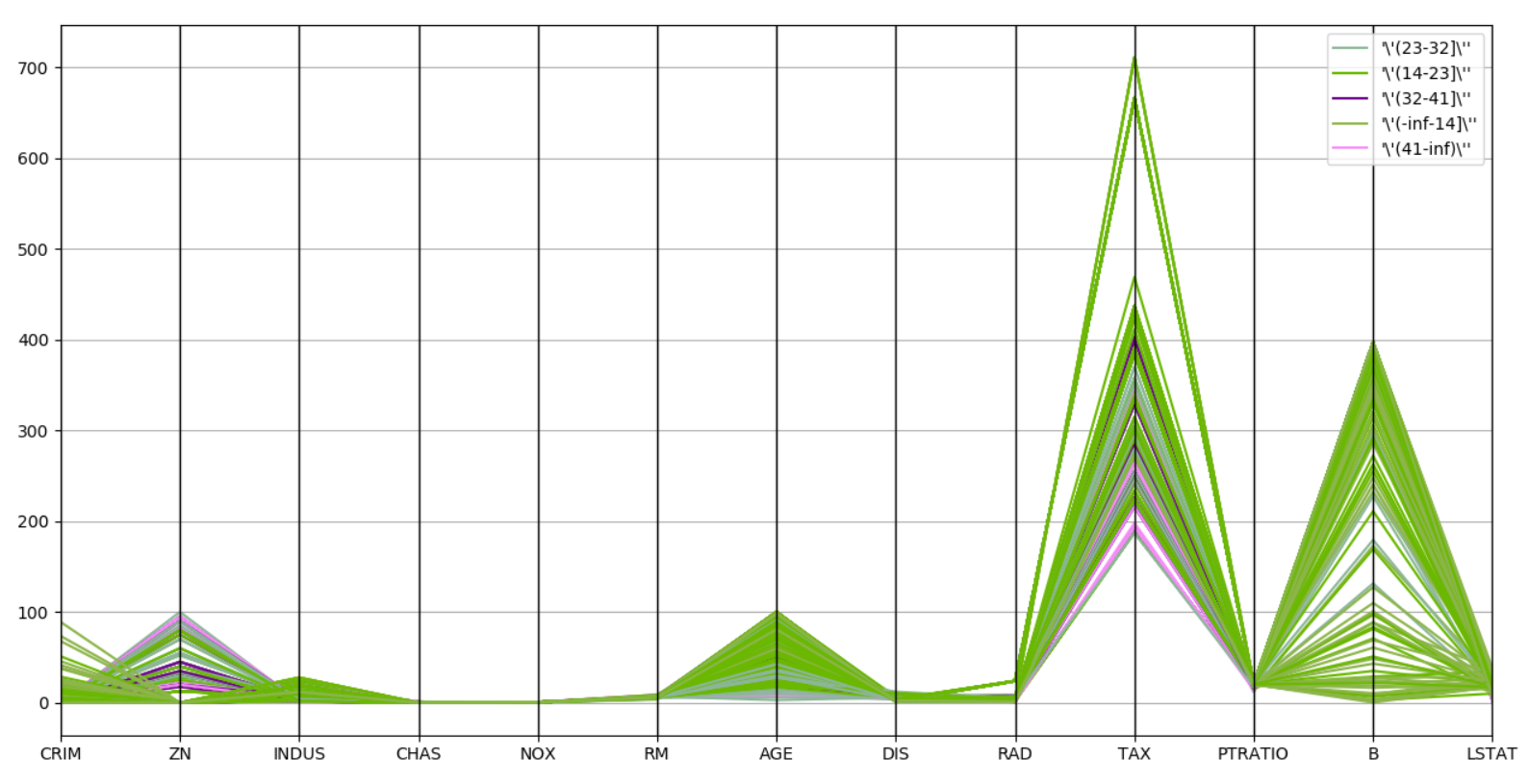
**Daugialinijiniai grafikai**

Šiame grafike galima pamatyt, kad duomenų rinkinyje yra daug įrašų, kurie apibūdina vidutinę būsto kainą, kuri yra mažiau vidutinio. Taip pat, galima pamatyti, kad mokinių ir mokytojų santykis (*ptratio* atributas) yra ganėtinai tolygiai pasiskirstęs skirtinguose Bostono rajonuose ir jis nedaro didelės įtakos būsto kainai.



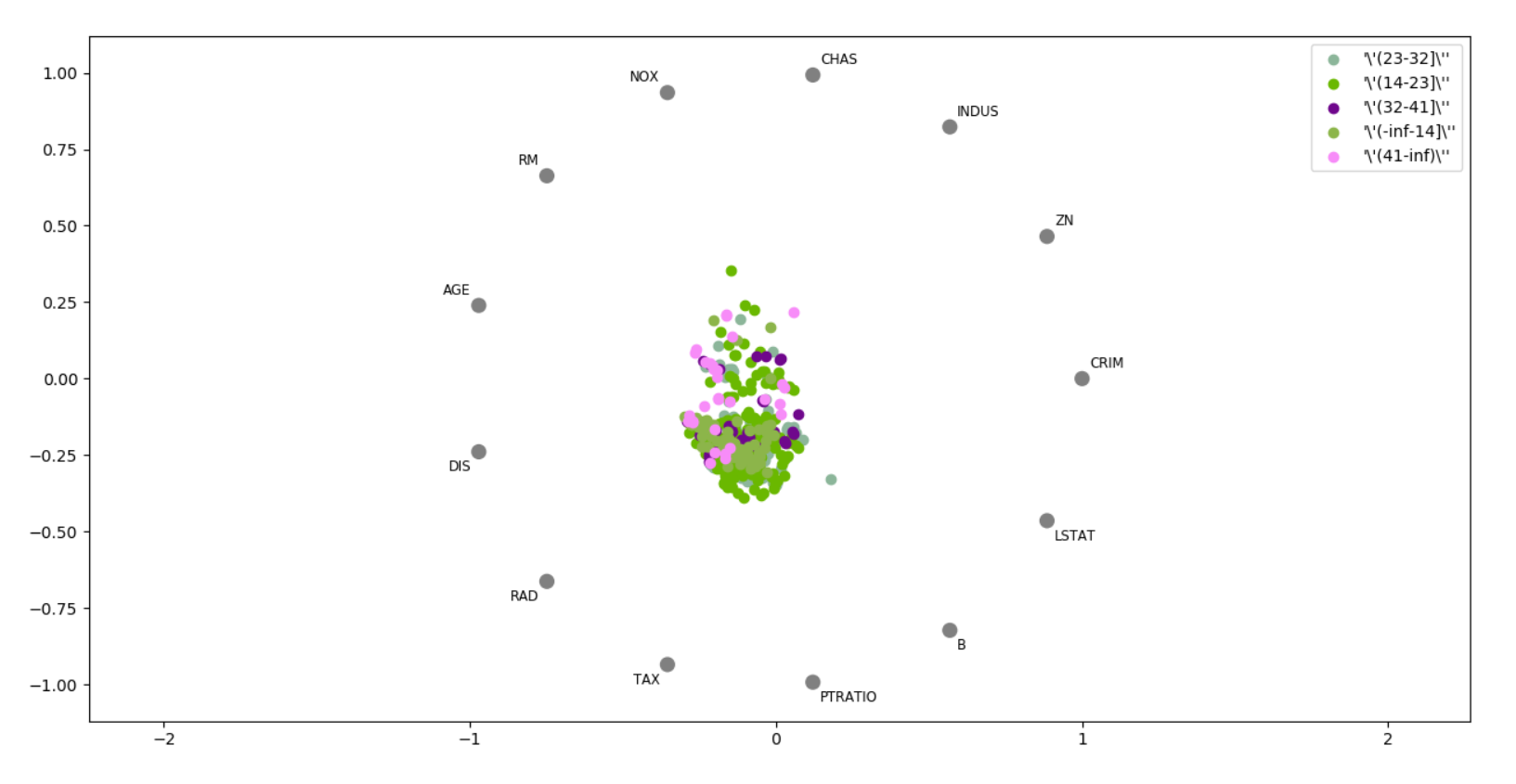
**Lygiagrečios koordinatės**

Duomenys buvo diskretizuoti pasitelkiant *WEKA* programinę įrangą. Klasės atributas *medv* buvo sutrauktas į penkis intervalus, bandant pastebėti, gal yra grupė būstų kainų, kurias nulėmė tam tikras atributas. Tačiau iš šio grafiko to negalime pamatyti, duomenys yra persipinę.



**Spindulinio vizualizavimo metodas**

Spinduliniame vizualizavimo metode galima grafiškai pastebėti, prie kokių atributų telkiasi klasės atributas, t.y. koks atributas turi didžiausią reikšmę. Iš šio grafiko matome, kad didelę įtaką būsto kainai turi *TAX* atributas, kuris reiškia mokesčio tarifą.



**Išvada:**

Grafiškai galima pamatyti kaip duomenys yra pasiskirste tarpusavy, kokių reikšmių yra daugiausiai, kurie atributai turi didžiausią įtaką klasės atributui.