

Vizuelizacija i analiza podataka

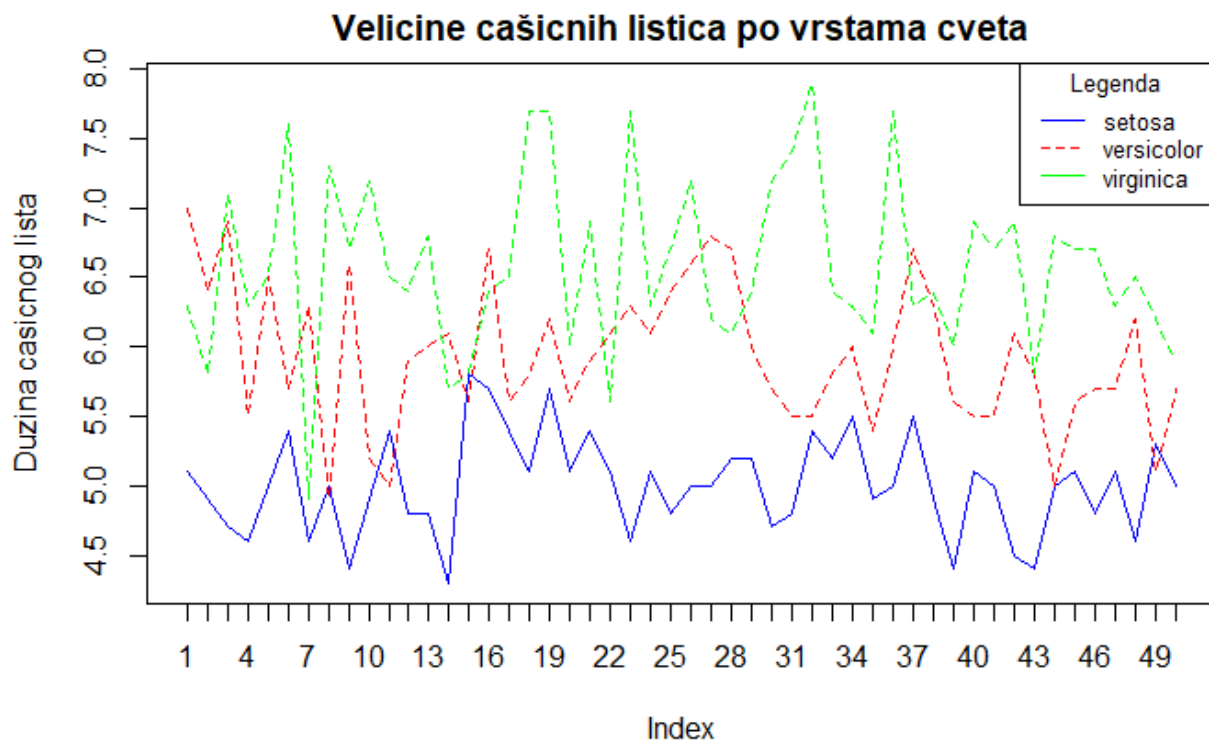
Drugi kolokvijum
24. decembar 2019.

1. Napisati *R* skriptu koja ućivata listu od najviše 20 brojeva i za nju prikazuje sledeće:
 - a. Pronaći i ispisati sve elemente iz liste koji su veći od vrednosti za treći kvartil.
 - b. Za vrednosti koje su ispod vrednosti trećeg kvartila ispisati: prosečnu vrednost i varijansu.
2. Date su dve matrice *A* i *B*:

$$A = \begin{bmatrix} 1 & 7 & 11 & 6 \\ 2 & 8 & 12 & 8 \\ 3 & 9 & 9 & 3 \\ 4 & 10 & 4 & 12 \end{bmatrix}$$

$$B = \begin{bmatrix} 2 & 6 & 6 & 5 \\ 3 & 4 & 2 & 11 \\ 4 & 8 & 3 & 12 \\ 5 & 9 & 4 & 7 \end{bmatrix}$$

- a. Spojiti dve matrice po kolonama (jedna pored druge), za zatim za svaku vrstu odrediti njen prosek.
 - b. Pronaći sve elemente spojene matrice koji su veći od 10, a zatim ih zamenti brojem 100.
 - c. Tako dobijenu matricu transponovati, a zatim uraditi sumu po kolonama.
3. Ugrađena baza podataka **iris** sadrži podatke o veličinama i oblicima 3 vrste cveta. Nacrtati sledeći grafik sa slike.



4. U fajlu apartmani.csv dati su podaci o udaljenosti apartmana u primorju od plaže, kao i o cenama tih apartmana na dnevnom nivou. Na osnovu datih podataka odrediti:
 - a. Pearson-ov koeficijent korelacije za te dve varijable.
 - b. Koeficijente linearne regresije koji pokazuju vezu izmedju udaljenosti i cene apartmana. Ukloniti netipične vrednosti ukoliko ih ima, pa nakon toga ponoviti postupak odredjivanja koeficijenata regresione prave. Na grafiku predstaviti podatke i obe regresione prave (različitim bojama). Dodati legendu.
 - c. Obrazložiti koji od ova dva linearna modela je bolji. Pomoću tog modela odrediti predviđene vrednosti za cenu svakog apartmana iz baze.