

Vizualizacija podatkov

<http://goo.gl/forms/hUiEEANMNh>



Zakaj vizualizacija?

- V delovnem pomnilniku je prostora za tri kose informacije naenkrat: številka, krivulja,
- Vid je najpomembnejši čut.

Primer – vrednosti valut (v EUR)

	Jan	Feb	Mart	Apr	Maj	Jun	Jul	Aug	Sept	Okc	Nov	Dic
USD	1,2147	1,2048	1,1816	1,1978	1,1848	1,2182	1,2119	1,2092	1,2012	1,1812	1,1811	1,2018
GBP	0,6621	0,6601	0,6603	0,6603	0,6603	0,6601	0,6611	0,6611	0,6611	0,6611	0,6611	0,6611
YEN	100,7200	100,5704	118,7200	124,5000	120,4000	118,4000	100,0000	118,0000	120,0000	112,5470	120,0000	121,0000

Kaj se je dogajalo z vrednostjo dolarja?
 Ali dolar in jen padata in rasteta hkrati?
 Kateri par valut se obnaša podobno?

Za odgovor na prvo vprašanje naš delovni
 pomnilnik zadošča, za drugi dve pa moramo
 primerjati (pre)več števil naenkrat.

Primer – vrednosti valut (v EUR)

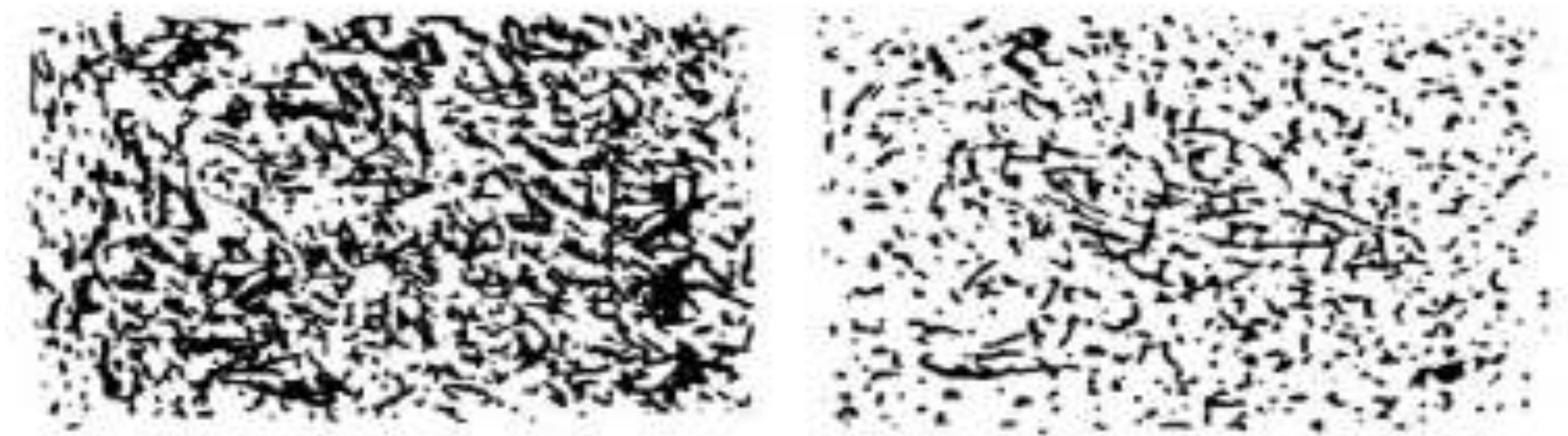
	Jan	Feb	Mar	Apr	Maj	Jun	Jul	Avg	Sep	Oct	Nov	Dek
USD	1,4147	1,3748	1,3458	1,3127	1,2787	1,2364	1,2117	1,2137	1,2187	1,2407	1,2117	1,2804
GBP	0,6477	0,6307	0,6140	0,6100	0,600	0,5897	0,5817	0,5877	0,5818	0,5807	0,5847	0,5877
YEN	120,7203	122,5724	119,7227	124,5590	123,8990	119,833	120,2847	119,5544	120,7000	112,5470	109,9254	107,0260



Kaj se je dogajalo z vrednostjo dolarja?
Ali dolar in jen padata in rasteta hkrati?
Kateri par valut se obnaša podobno?

Za odgovor na prvo vprašanje naš delovni pomnilnik zadošča, za drugi dve pa moramo primerjati (pre)več številok naenkrat.

Kaj je na sliki?



- Pri iskanju vzorcev v slikah smo ljudje izredno dobri.
- Manj pa pri iskanju vzorcev v številkah (le tri naenkrat).

Človekove omejitve

Težko primerjamo dve sliki.

Kadar je namen vizualizacije primerjava, postavi eno sliko tesno ob (ali čez) drugo.



Kakšna je razlika med slikama?

Ne verjamete?



Ne verjamete?



Ne verjamete?



Verjamete?



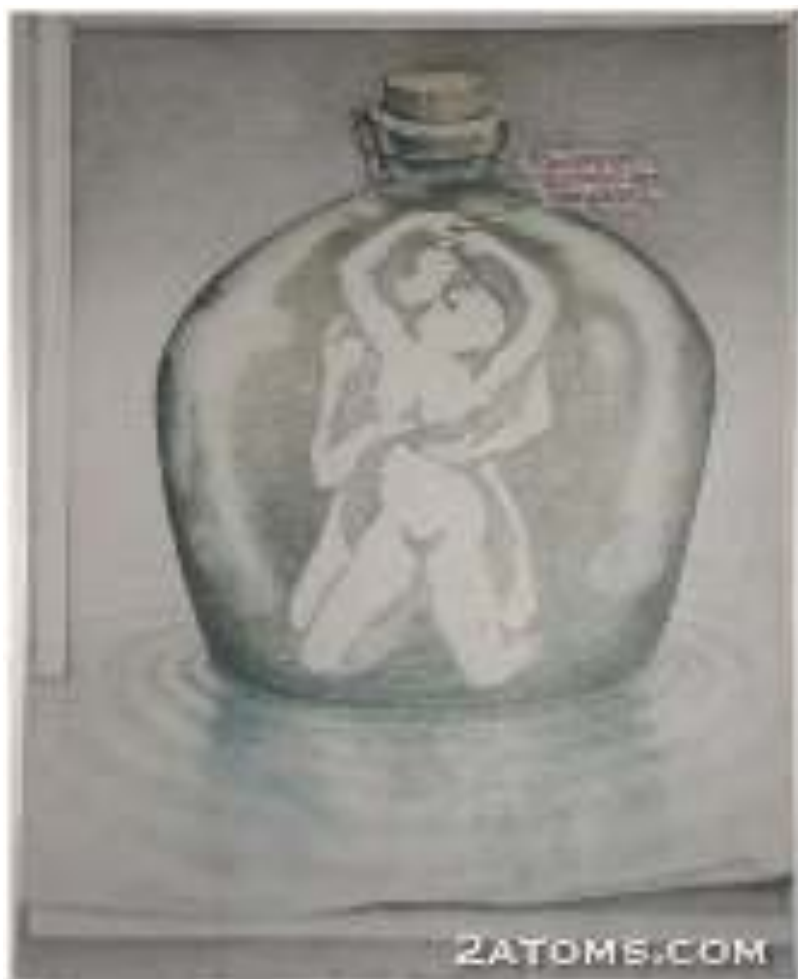
Verjamete?



Še nekaj...



Človekove omejitve - II



Naša pričakovanja.

- Vidimo, kar želimo videti in tisto, kar pogosto videvamo.
- Nepričakovane vzorce težje opazimo.
- Potrebno se je uriti v zaznavanju neznanega in uporabljati vizualizacije, ki izpostavijo “nepričakovano.”

Pri branju slik imamo določene konvencije:

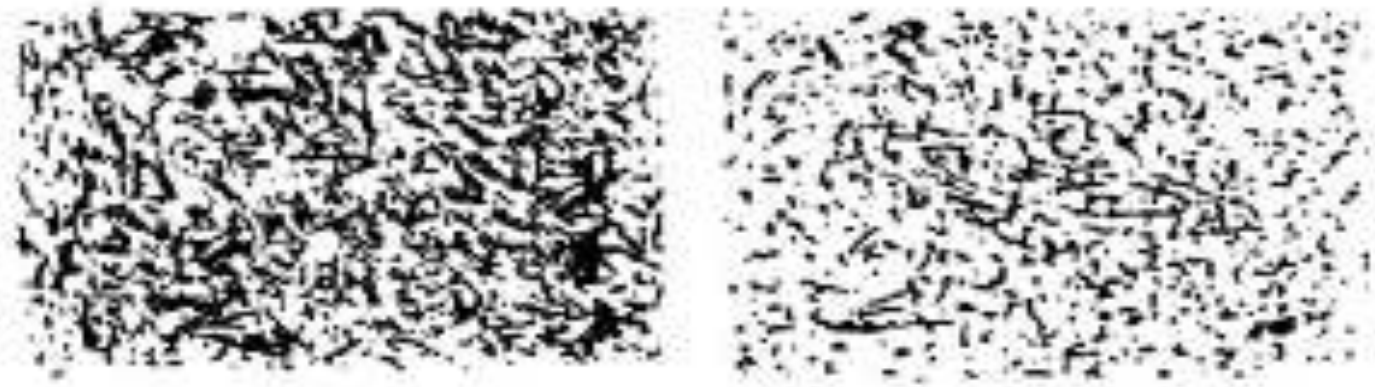
- čas teče od leve proti desni
- rdeča barva slabo, zelena dobro,
- ...

Človekove omejitve - III

Vidimo vzorce, ki jih ni.

Vidimo, saj HOČEMO videti vzorce, brez njih se počutimo izgubljene.

Ljudje prepoznajo vzorce v slikah, ki so naključni šum.



Eksperti prepoznajo vzorce v zmedi na Wall Streetu med gospodarsko recesijo, čeprav verjetno večinoma naključni.

"The researchers found that when people were primed to feel out of control, they were more likely to see patterns where none exist." (See a Pattern on Wall Street?, [John Jansz](#), on Science)

Kdaj ne risati?



Ko malo števil, jih raje napiši.

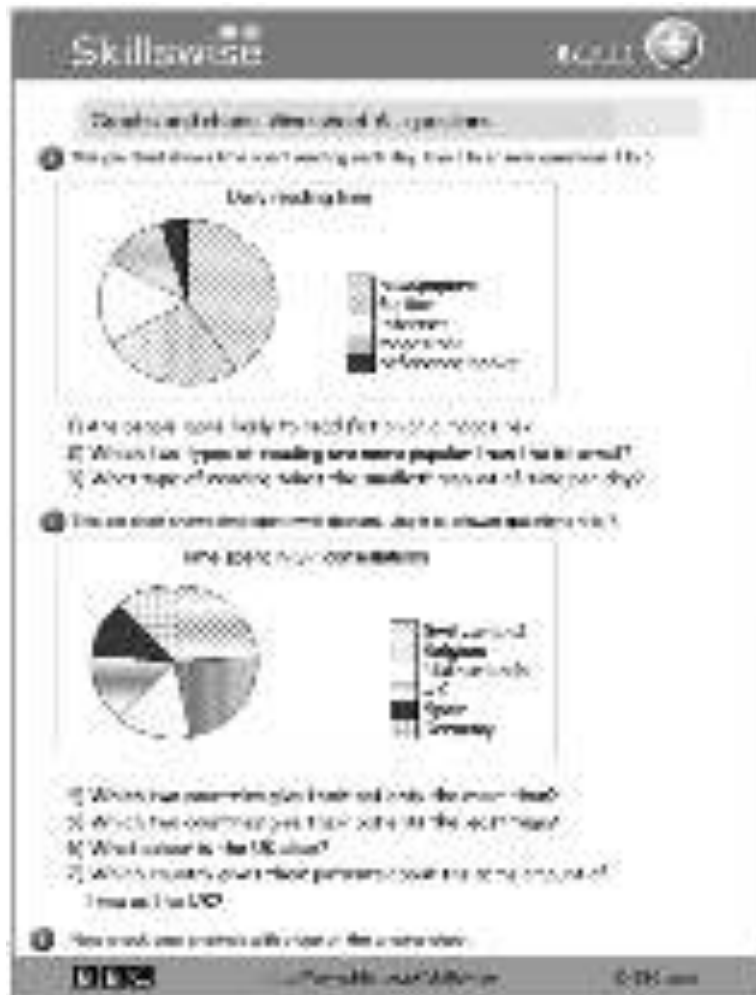
Primer:

- Sedem med seboj bolj ali manj nepovezanih števil.
- Je število patentnih uradnikov zakonsko določeno med 6000 in 8000?
- Kazalci ne kažejo pravega razpona in kje se trenutno nahajamo.

Osnovne vizualizacije

- tortni diagram (pie chart)
- box plot
- stolpični diagrami (bar chart)
- črte oz. krivulje

Tortni diagrami (pie chart)



Pretežno **neuporabni**.

Naj bi prikazovali deleže, ki pa jih težko razberemo.

Primer:

Na prva tri vprašanja bi lažje in točneje odgovorili, če deleži podani v urejenem seznamu.

Drugi graf je nejasen. Deleže česa prikazuje?

Če so to povprečni časi preživeti pri zdravniku, boljši prikaz stolpični diagram.

Alternativa tortnemu diagramu

Slabo:

- nagnjena torta,
- preveč podobne barve,
- legenda predaleč od diagrama,
- (ne)urejenost kategorij



Sector Allocation of Holding

Financials	21.45%
Non-Cyclical Consumer Goods	18.00%
Cyclical Services	14.17%
Information Technology	12.81%
Resources	9.81%
General Industries	8.99%
Utilities	3.83%
Basic Industries	3.70%
Non-Cyclical Services	3.87%
Cyclical Consumer Goods	1.87%
Total	100%



Bistveno boljša
alternativa:

+ stolpični diagram

Še slabši primer tortnega diagrama

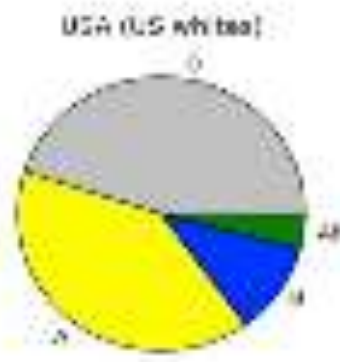
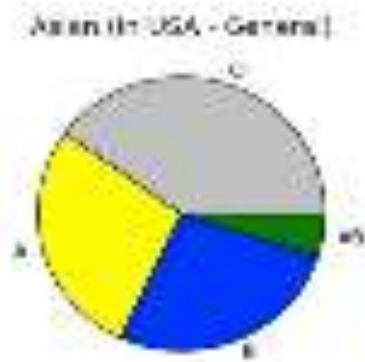


Če že uporabljamo, naj se deleži seštejejo v 100%!

Tortne diagrame težko primerjamo

Katera etična skupina ZDA ima največji delež krvne skupine B?

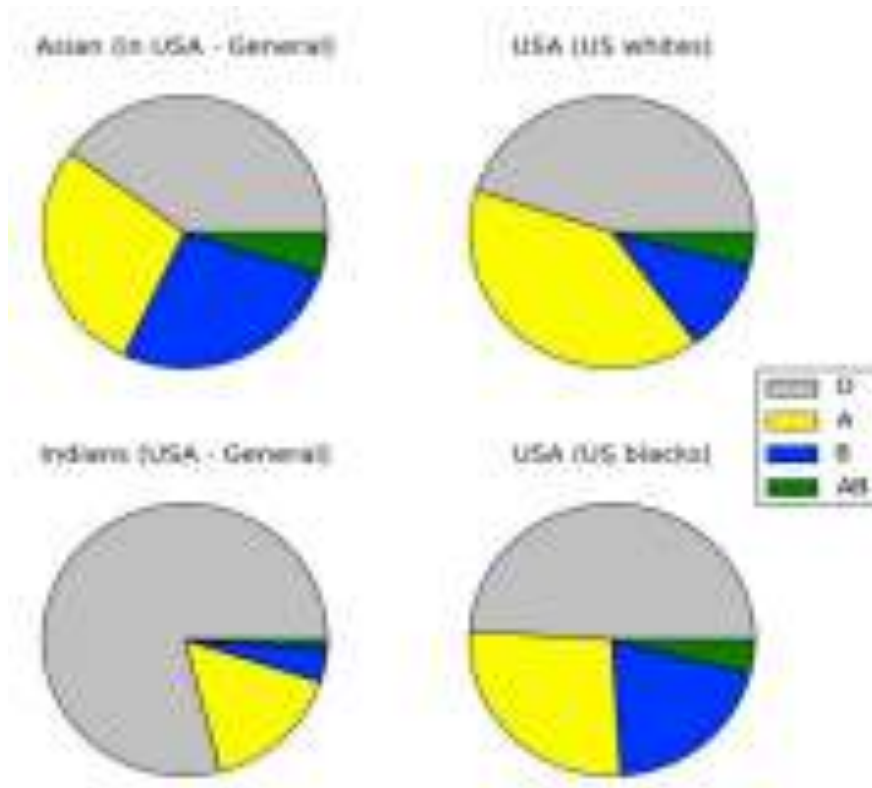
Potrebno je kar nekaj časa in skakanja s pogledom od enega do drugega grafa.



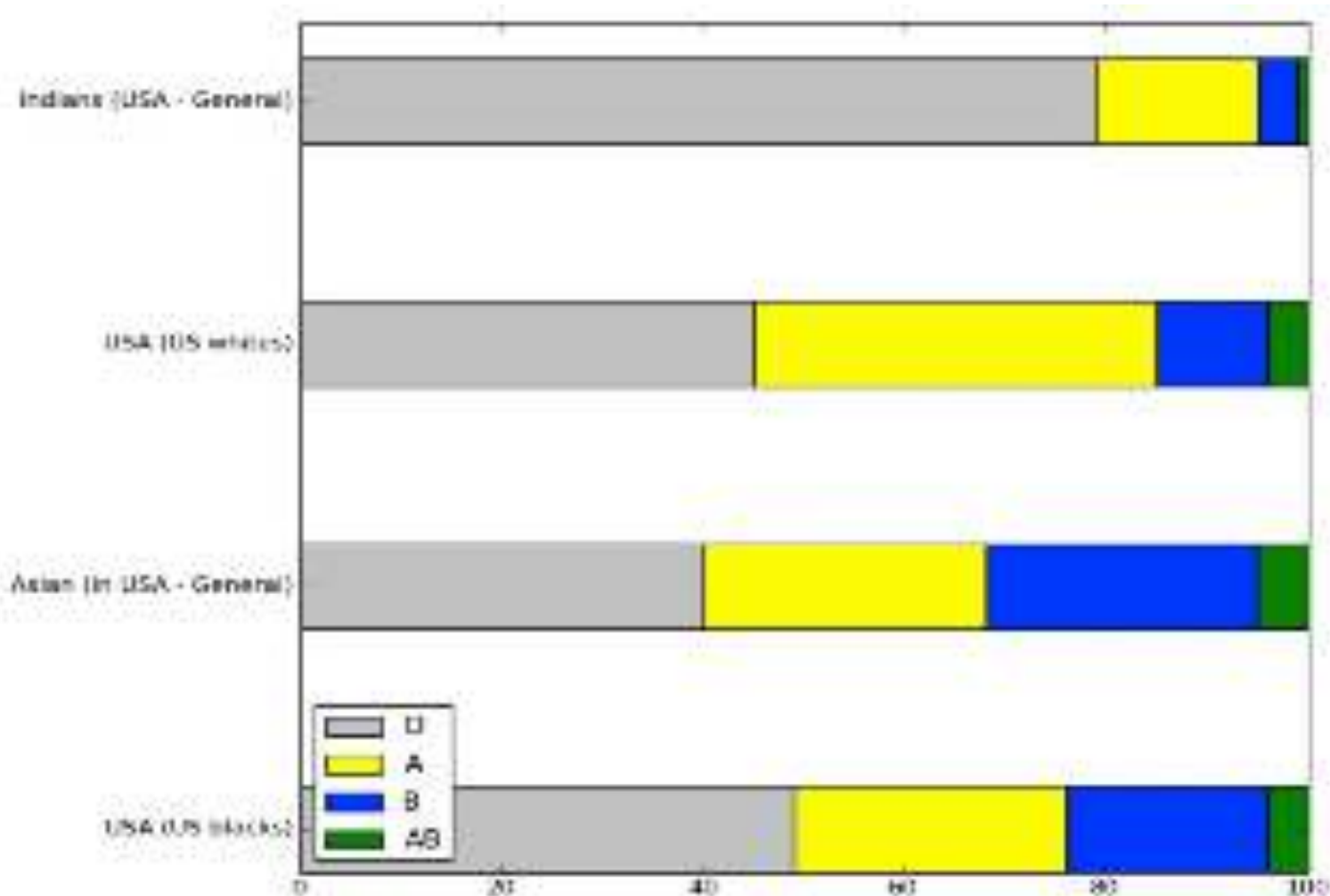
Tortne diagrame težko primerjamo

Katera etična skupina ZDA ima največji delež krvne skupine B?

Če umaknemo legendo, je še prijetneje.



Boljša alternativa – štrudelj



Ker je tortne diagrame težko primerjati...



Kaj se dogaja z deležem japonske avtomobilske industrije?

Slabo izbrane barve: Japonska ni modra (kvečjemu rumena ali oranžna), Nemčija tudi ni rumena in Francija zelena.

Boljši način primerjave deležev v času

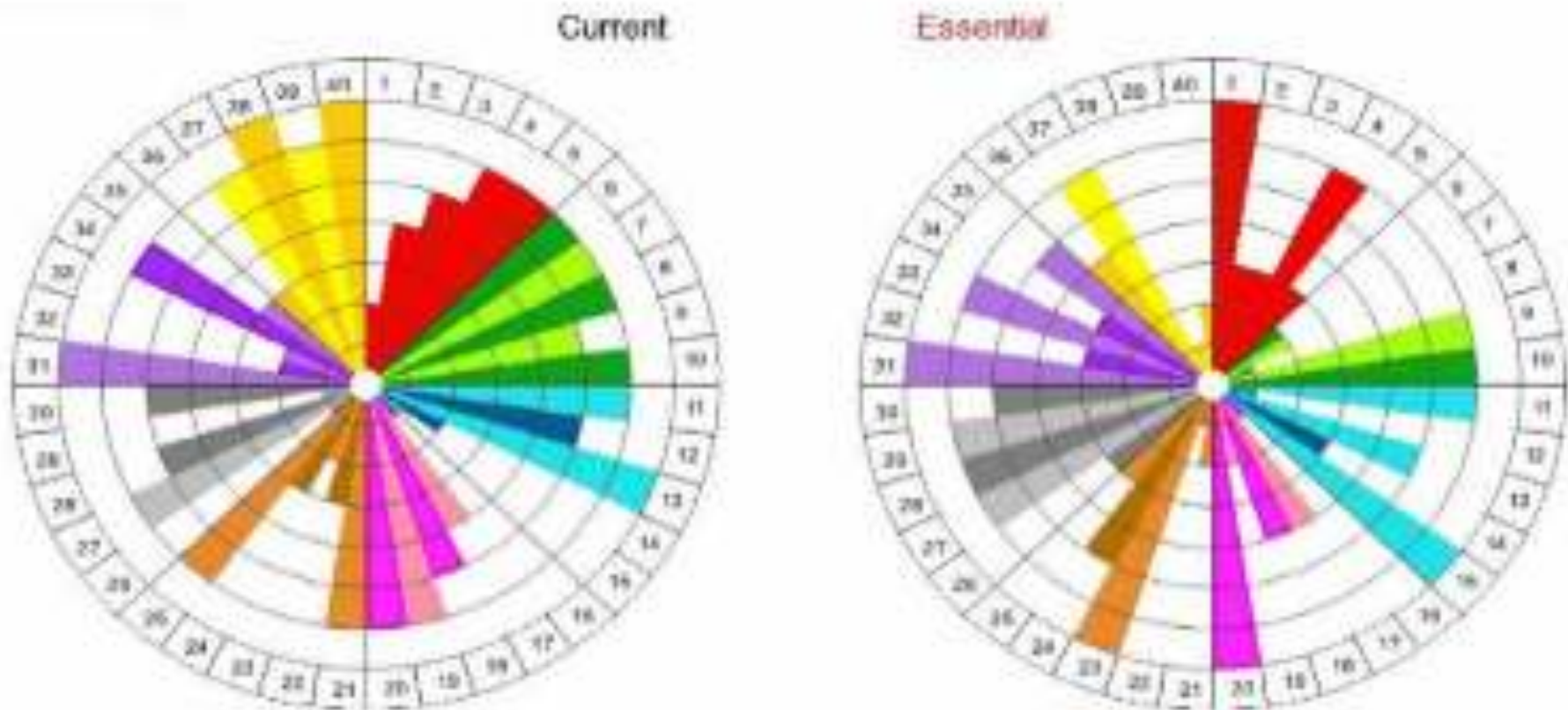


Kaj se dogaja z deležem japonske avtomobilske industrije?

Opazujemo dogajanje skozi čas.

Oznake x-osi so tudi glava tabele.

Kako primerjati tole?



Namen te vizualizacije je narediti primerjavo med tem, kaj imamo in kaj bi morali imeti.

Kaj kaže ta slika?



Da čas teče.

Če pobrišemo številke, pa sploh ne vidimo ničesar več.
Diagonalna postavitev oteži sicer nemogoče primerjanje.



Javnomnenjska anketa o podpori angleškim strankam

Popularity by generation



Naslov se nanaša na desni del. Levi kaže, kako se je podpora spremenila v zadnjem tednu.

Nemogoče barve.

Kdo opazi, da se oranžni "krivulji" križata?

Kateri teden kaže obroč na levi? Pravzaprav vseeno, 31% ali 32% ne ločimo.

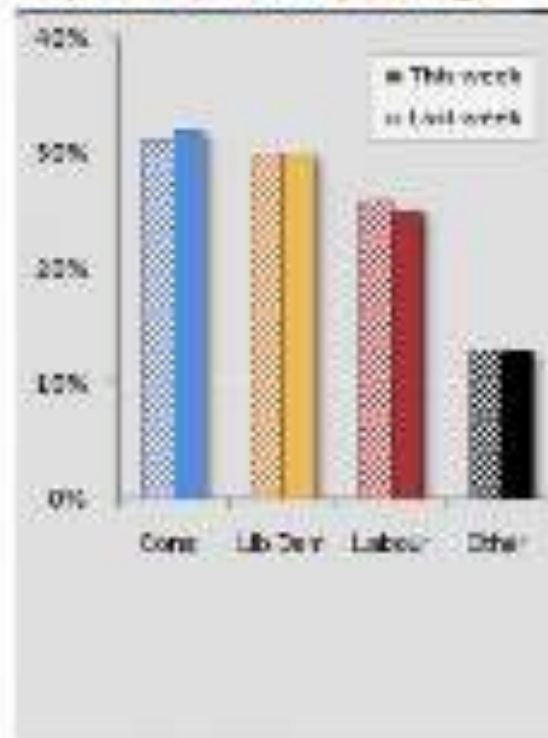
Zakaj razlika posebej podana?

Popularity by generation

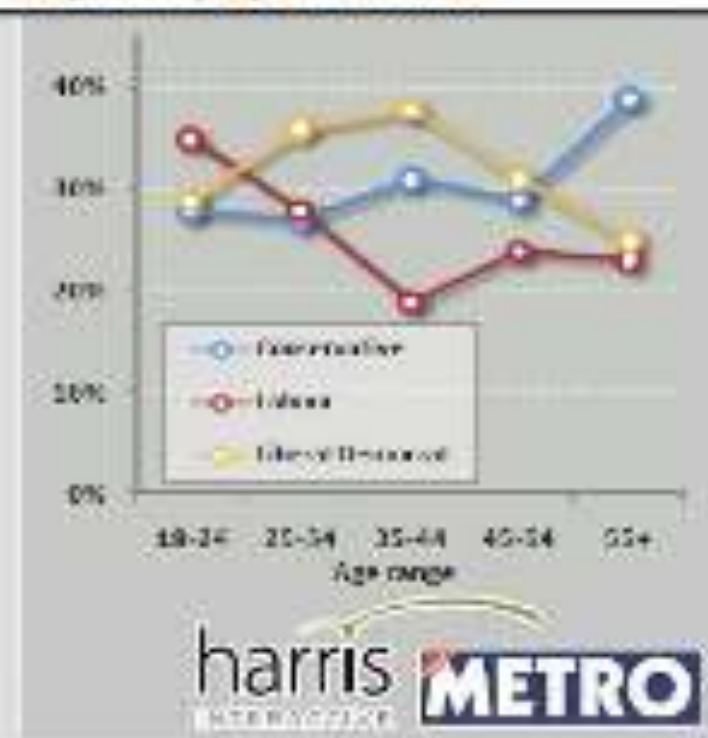


Lepši prikaz sprememb.
Jasne barve.
Prejšnji teden je prikazan
"šibkejšje."
Imamo krivulje.

Popularity Weekly Change



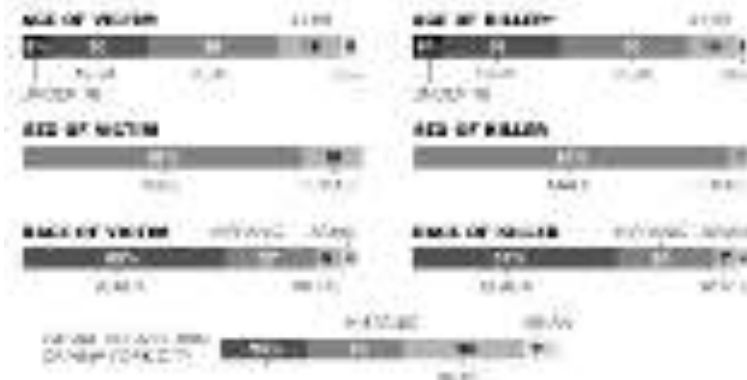
Popularity By Generation



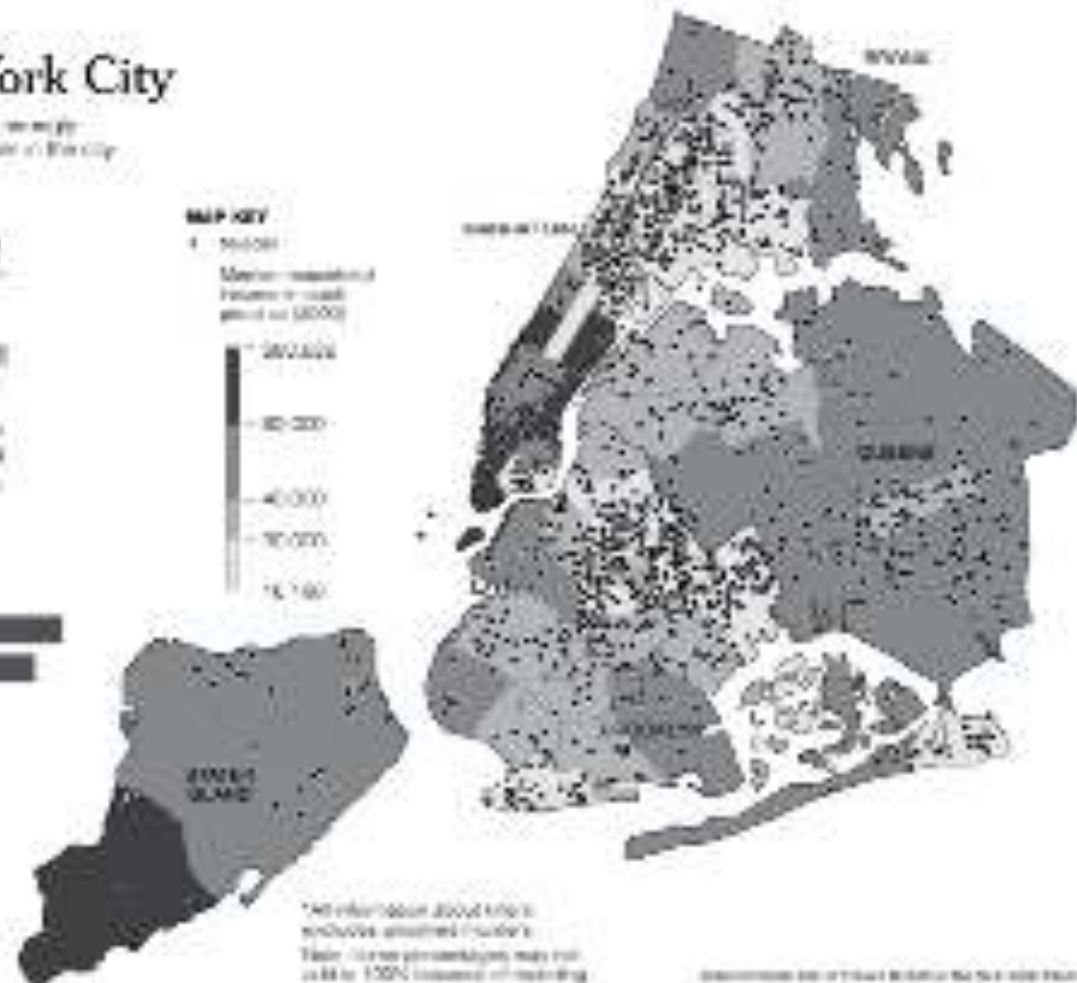
Štrudelj – nadomestek za torto

Three Years of Murder in New York City

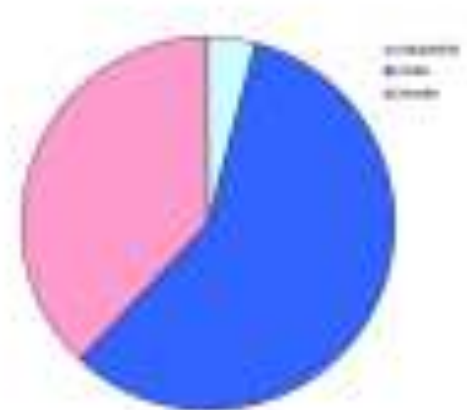
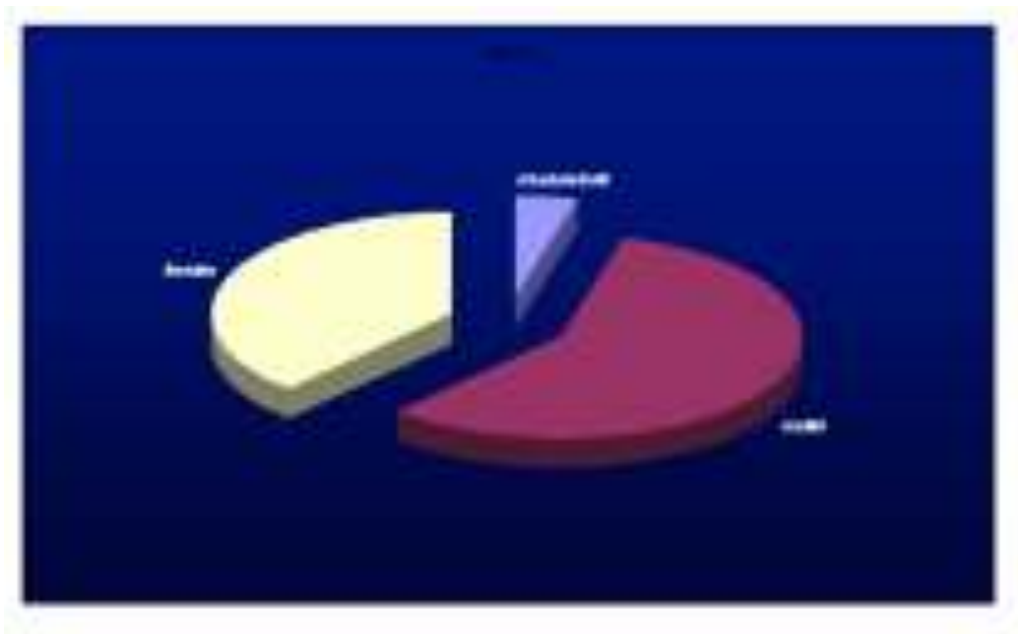
1990-1991 and 1992-1993. New York City, NY through 2010. Sample
 collected in place on 2nd floor. Here we have all the data of the 2010-2011



Please note that the following information is for informational purposes only and is not intended to be used for any other purpose.



~~3D torte~~



Za to ni niti enega dobrega razloga.

3D torte

Perception of 3D pie charts



Dobre le za norčevanje iz tridimenzionalnih tort.

Najboljša torta

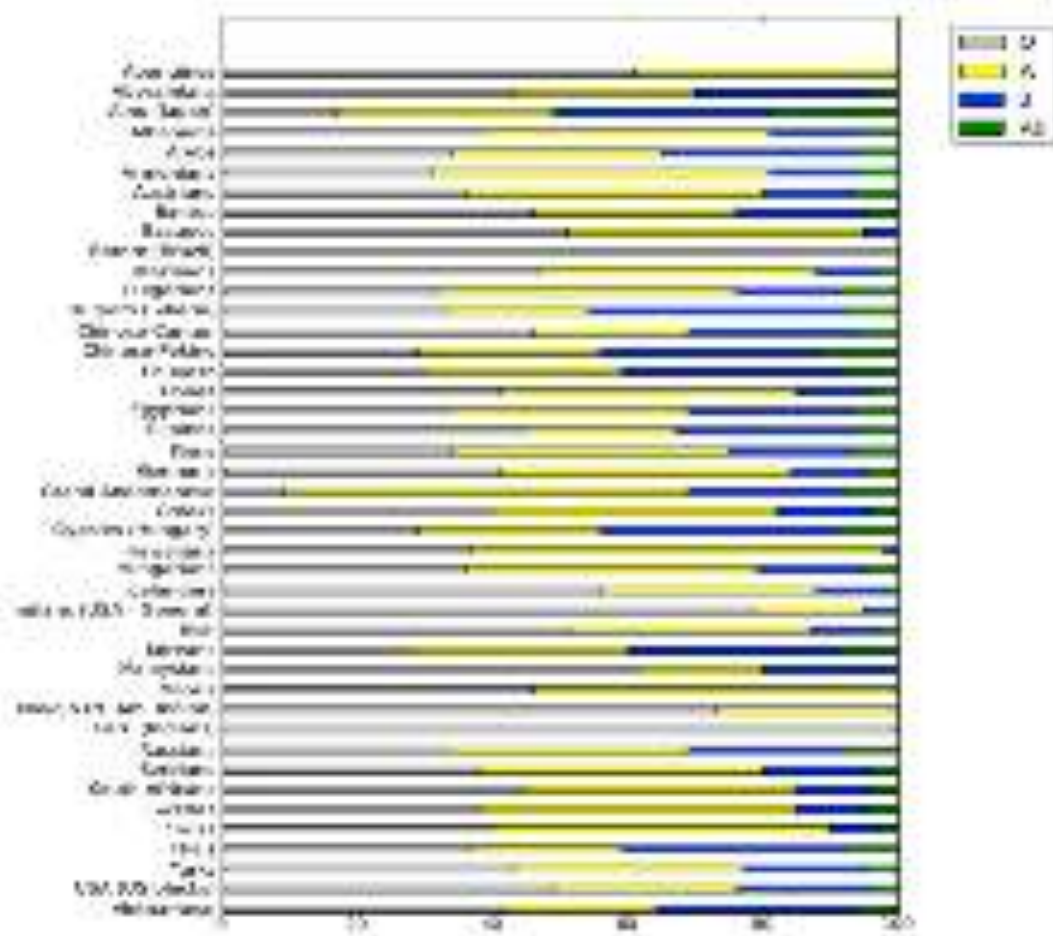


Box plot

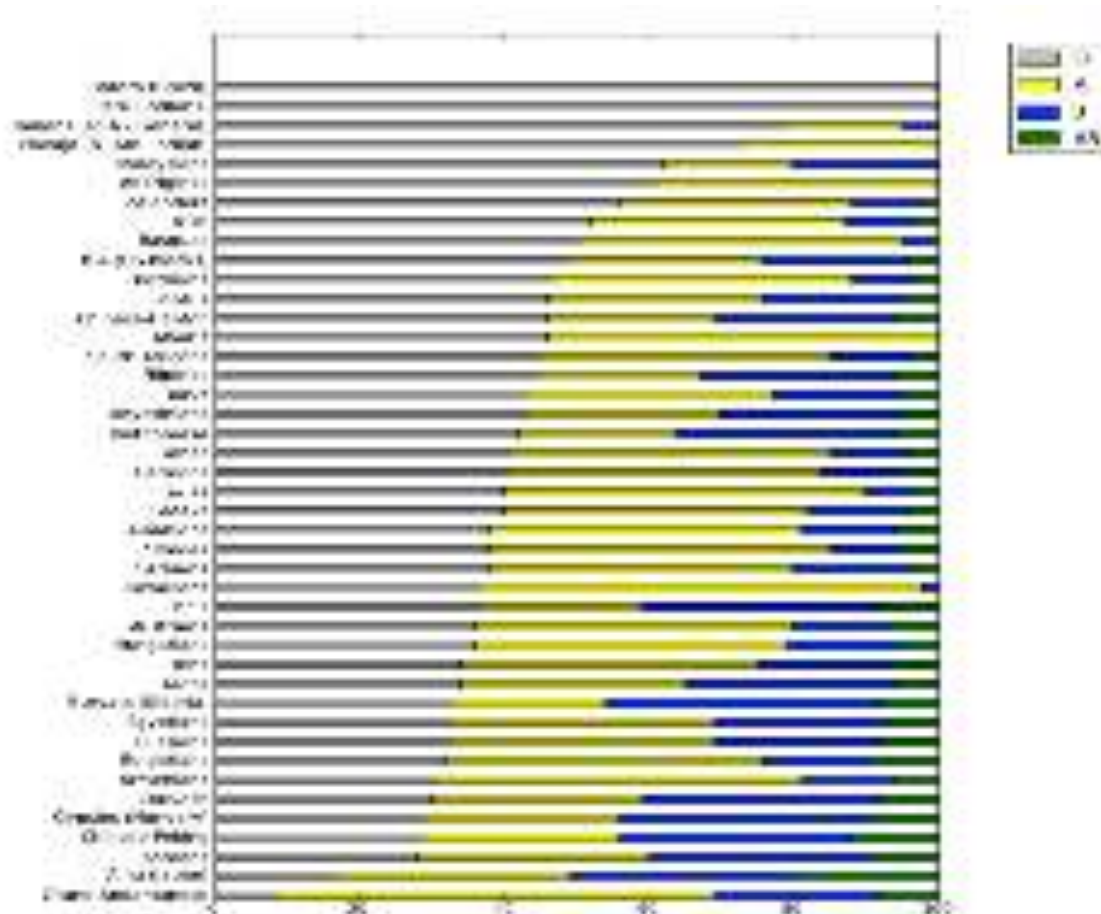


- Kažejo osnovne statistike:
 - povprečje,
 - mediano, ...
- Poleg histogramov primerne za prikaz zveznih distribucij.
- Lahko postavimo enega zraven drugega in primerjamo populacije, dele vzorca (moški, ženske), ...
- Hitro poiščemo čudne primere, napake v podatkih, ...

Stolpični diagrami (bar chart)



- Primerni za prikaz diskretnih atributov (ordinalnih ali nominalnih).
- Uporabni tudi za zvezne, vendar jih prej diskretiziramo.
 - Ordinalne uredimo naravno (mladi, srednji, stari), od leve proti desni, od zgoraj navzdol.
 - Nominalne uredimo na nek smislen način (abeceda ni nujno najboljši način).
- Vrstni red stolpcev pomemben, odvisen od namena primerjave posameznih primerov. S kom želimo primerjati Egipčane? Z geografskimi (Libicij in Izraelci) ali abecednimi sosedi?



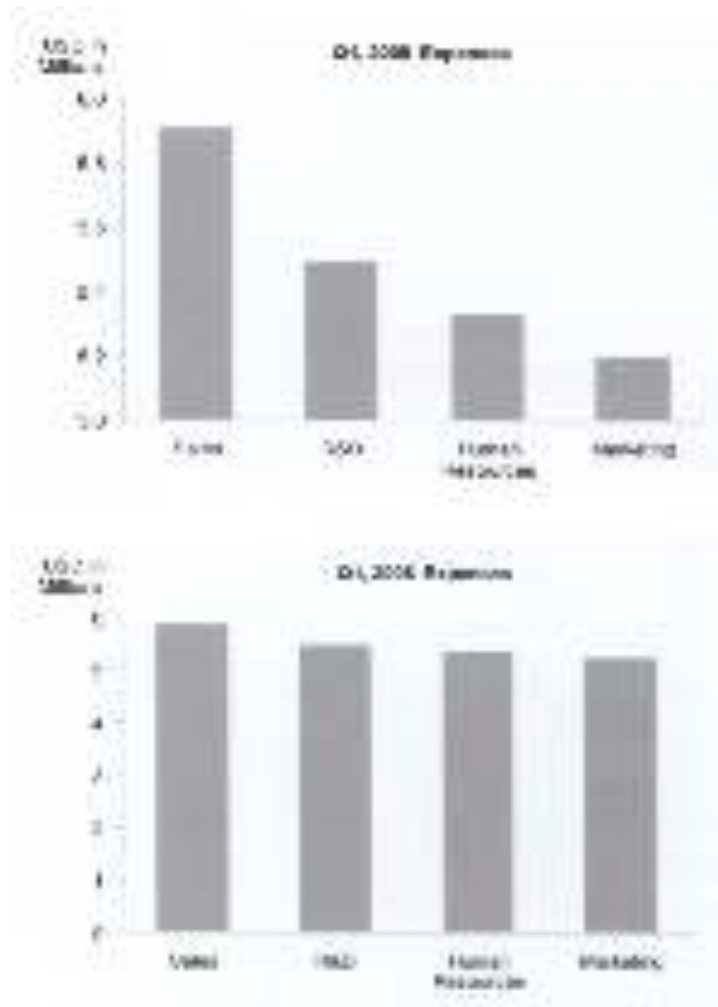
Uredimo po deležu
krvne skupine 0, in hitro
vidimo kam Egipčani
sodijo po deležu krvne
skupine 0.

Diagrami kažejo dve
spremenljivki:

- diskretna: etnično skupino
- zvezna: distribucija deležev krvnih skupin (štrudelj)

Postavimo **vodoravno**
(več teksta obnje) ali
navpično (včasih
“naravnejši,” recimo
višina Soče po
mesecih).

Skala lahko zavaja



Skala naj začne pri 0, sicer popačimo vizualizacijo.

Gre za R&D res dvakrat manj denarja kot za prodajo?

Histogrami



Kadar količine (deleža ali števila nečesa) kažemo v vsakem stolpcu posebej (ne v obliki štrudlov).

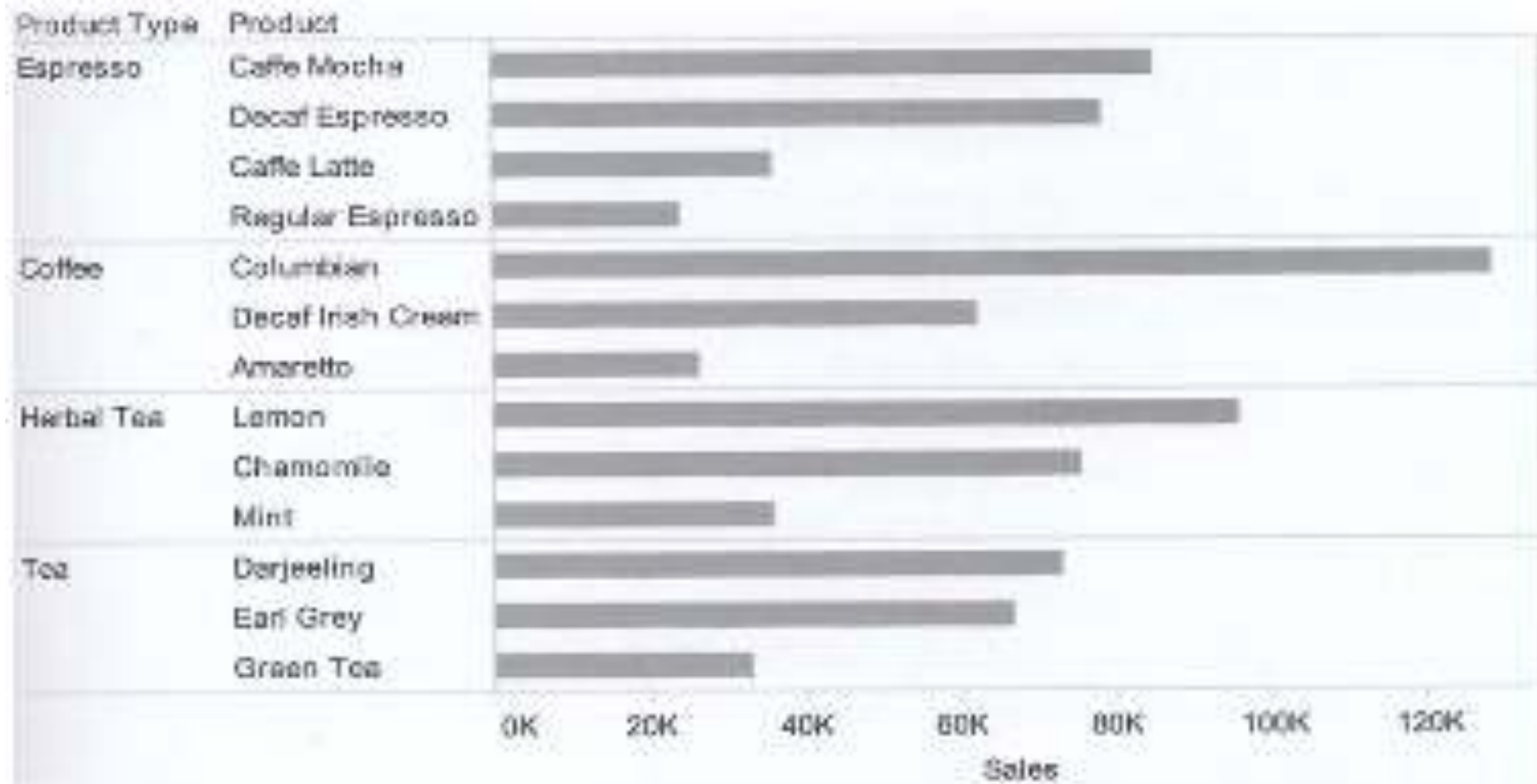
Kadar kažemo distribucije, jim rečemo histogrami.

Vizualizacija na levi je zgrešena, ker dodana slika na desni navidezno poviša stolpec, čeprav pol manjši od srednjega.

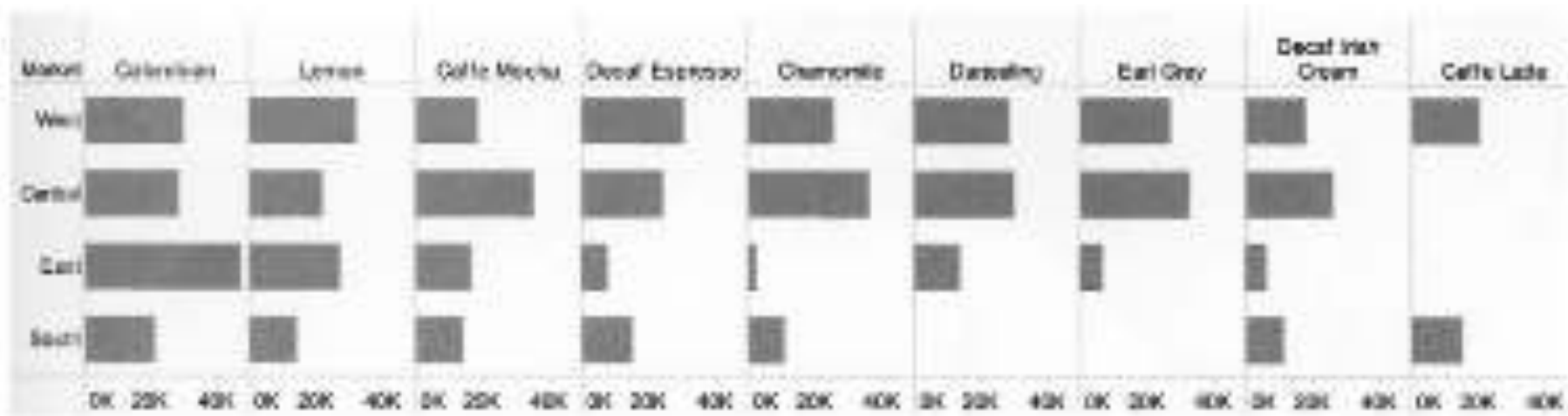
Kaj je tu narobe?



Po kategorijah



Po dveh kategorijah hkrati



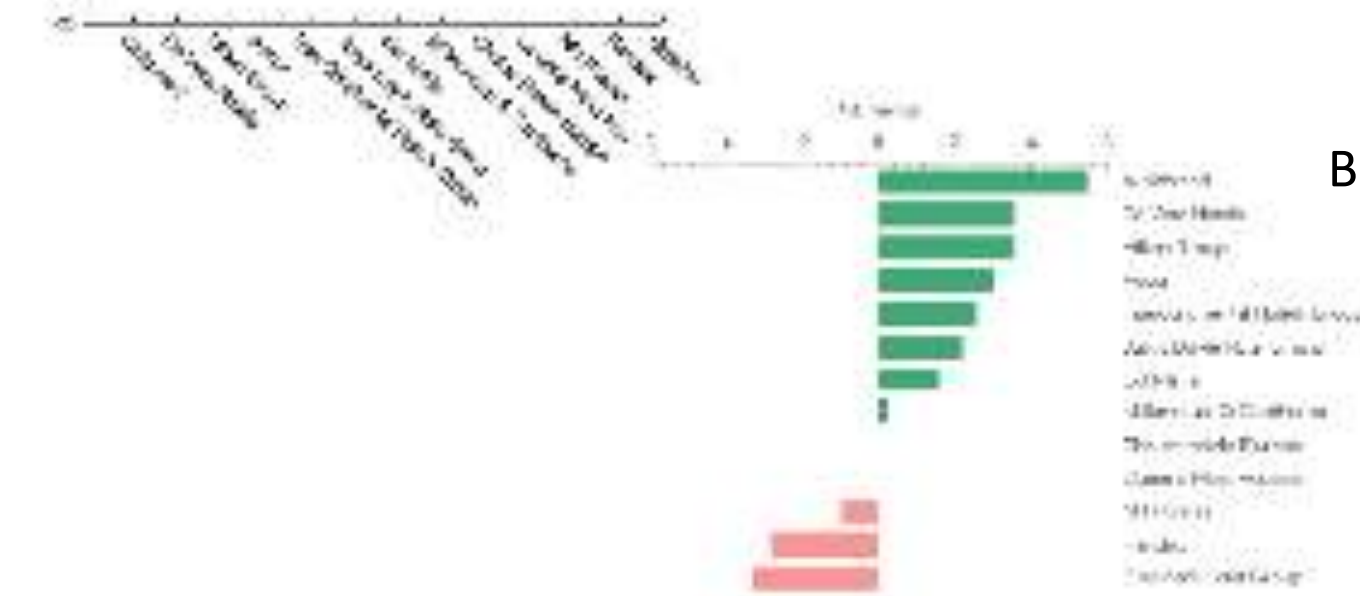
Po vrsti proizvoda in tržišču.

Pozitivni in negativni odkloni



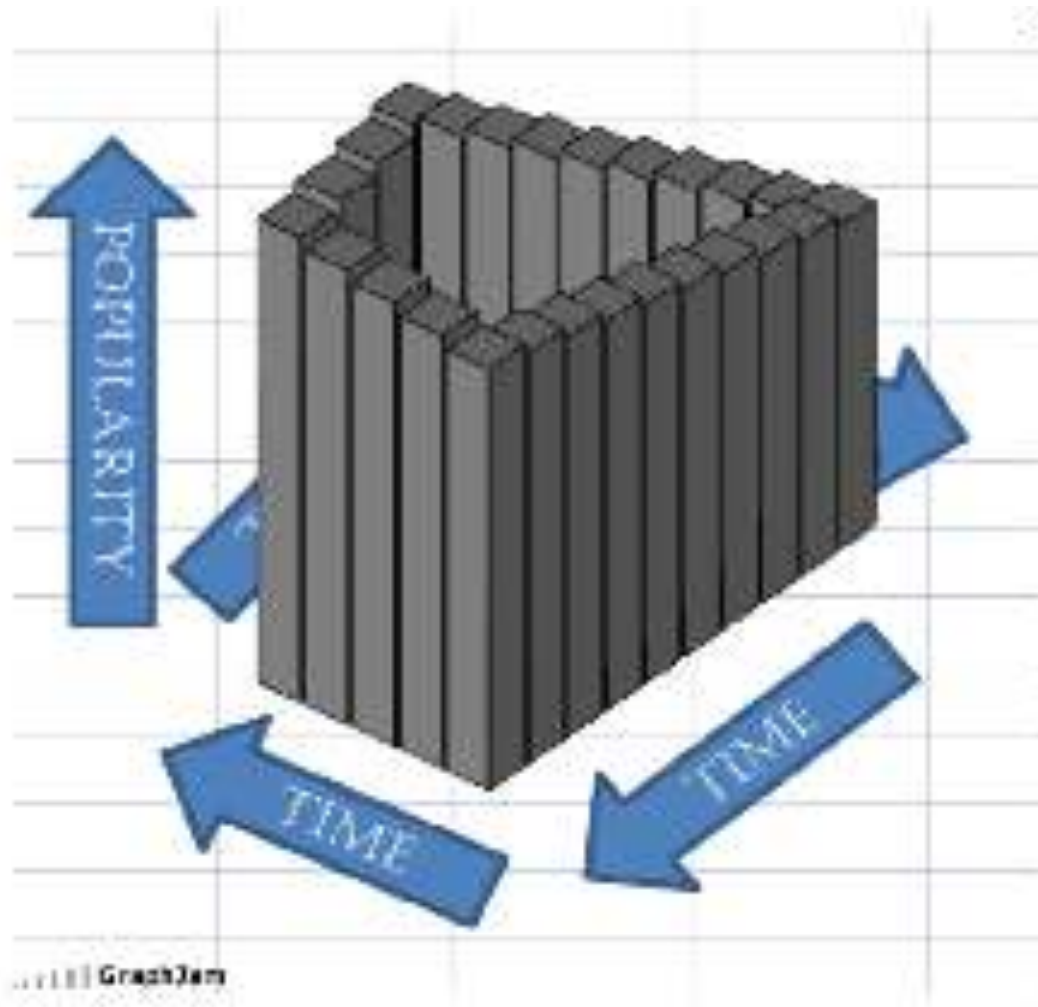
3D nepregledno.

Preširoka skala (skrči prikazani del).



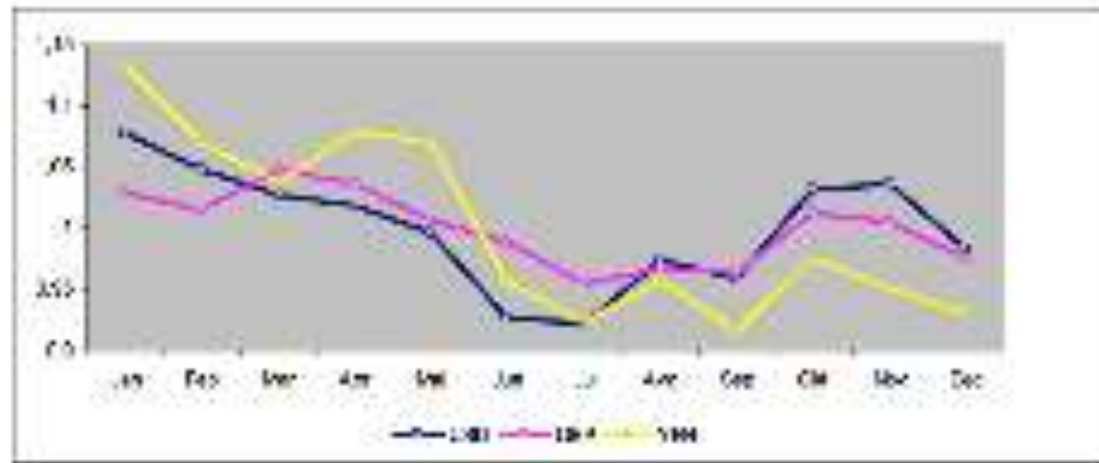
Bistveno boljši.

~~3D~~ grafi



Krivulje

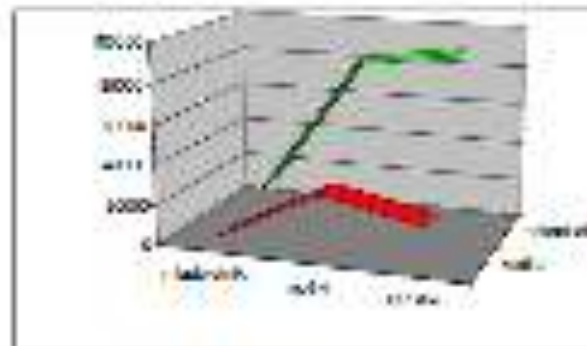
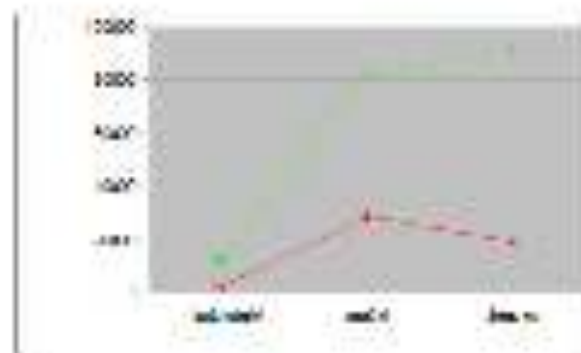
	Jan	Feb	Mar	Apr	Mai	Jun	Jul	Aug	Sep	Oct	Nov	Dec
IMR	1.448	1.395	1.436	1.500	1.505	1.294	1.213	1.379	1.291	1.262	1.158	1.205
GNP	7.4234	7.4871	7.4775	7.4887	7.4945	7.5071	7.5111	7.5213	7.5341	7.5476	7.5611	7.5734
YEN	10.1286	10.5724	10.7557	10.5583	10.7862	11.0465	10.6547	10.5044	10.5785	10.5478	10.5064	10.4305



Dobri predvsem za prikaz sprememb v času in opazovanju trendov.
 Atribut na osi x: zvezen ali urejen (nikoli neurejen).
 Atribut na osi y: (praviloma) zvezen

Krivulje - II

	mladoletniki	moški	ženske
izkušnje	2761	2070	1807
neizkušnje	1553	8294	1733

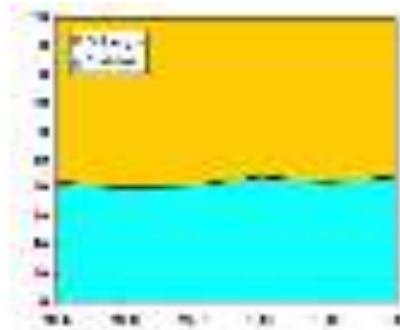
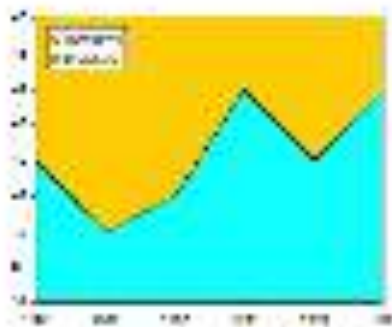


Nikoli ne
kažemo
neurejenih
atributov.

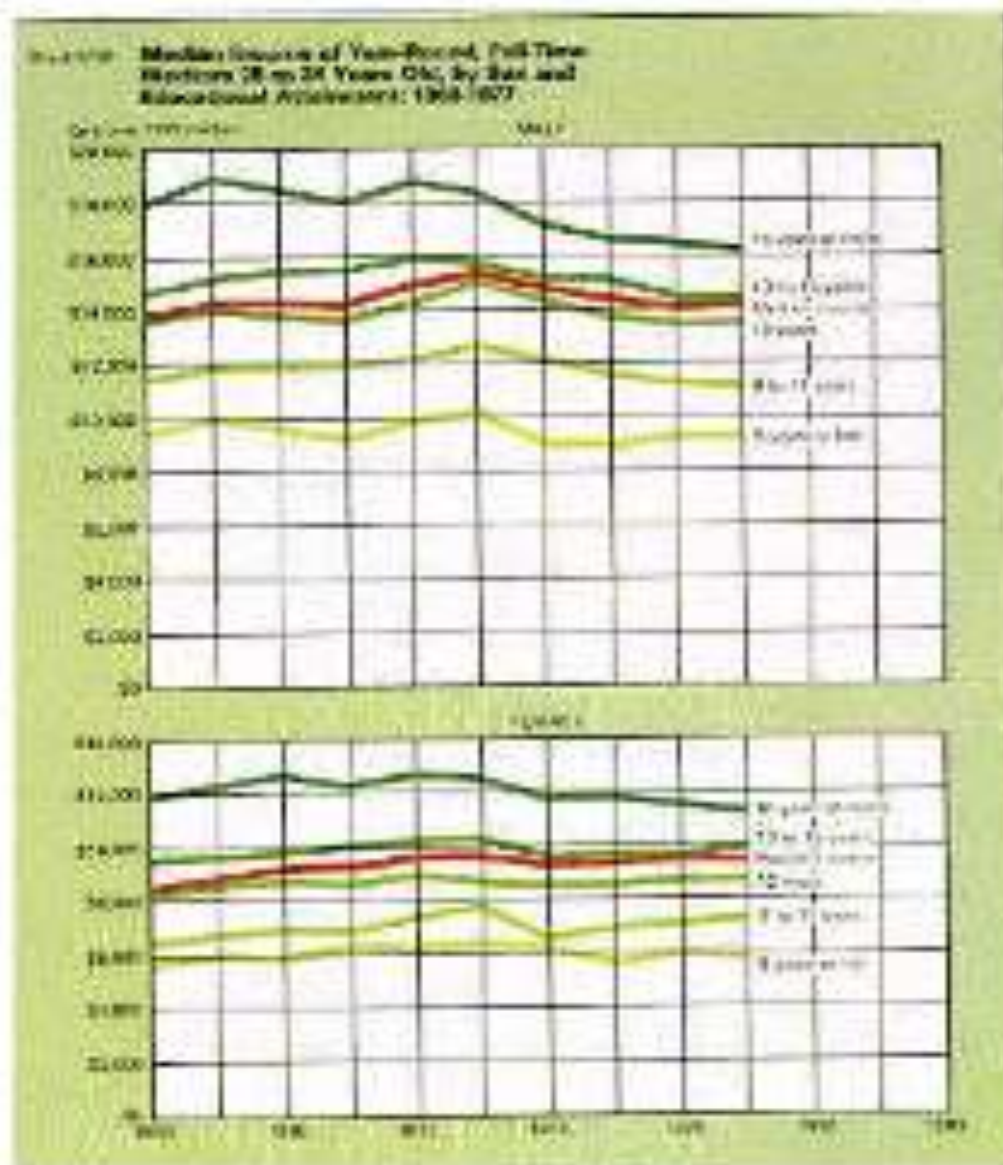
Ko mladoletnik
postane moški,
začne kaditi?

Ko se kasneje
spremeni v
žensko, bo
morda nehal.

Prikaz deležev z obarvanimi površinami pod krivuljami



- Čas zdaj teče na osi x.
- Problem skale:
 - prva kaže med 0 in 100, majhne razlike.
 - druga kaže med 38 in 46, razlike videti ogromne.Kaj je prav?
- Spodnja desna slika manj pristranska do (manj favorizira) Christie's.



- Višina povprečne letne plače glede na število let izobrazbe.
- Daljše izobraževanje, višja plača. Razlike se z leti manjšajo.
- Primerjava moški in ženske težka, ker grafa na različnih grafih, z dvema različnima skalama!
- Ženska s 16 leti izobraževanja (najvišja spodaj) plačana enako kot moški z 8-11 leti (druga od spodaj).

Manipuliranje – pritoževanje nad šolninami

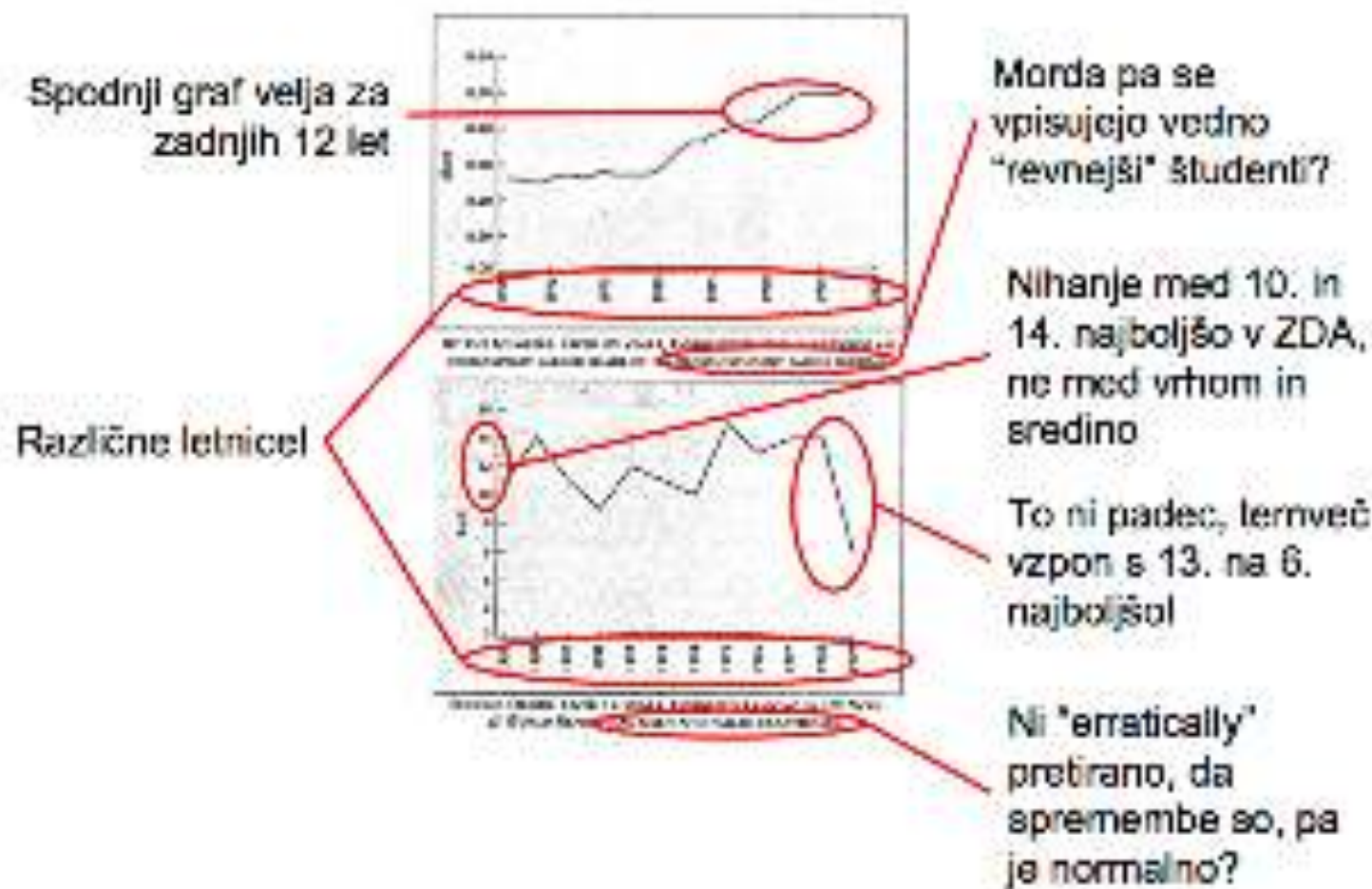


Kvaliteta univerze pod ceno.

Kvaliteta pada, vendar še vedno nad ceno šolnine.

Nič od tega res, saj teh krivulj ne moremo postavljati na isti graf, ker itak nimata iste skale!

Podrobnosti manipulacije

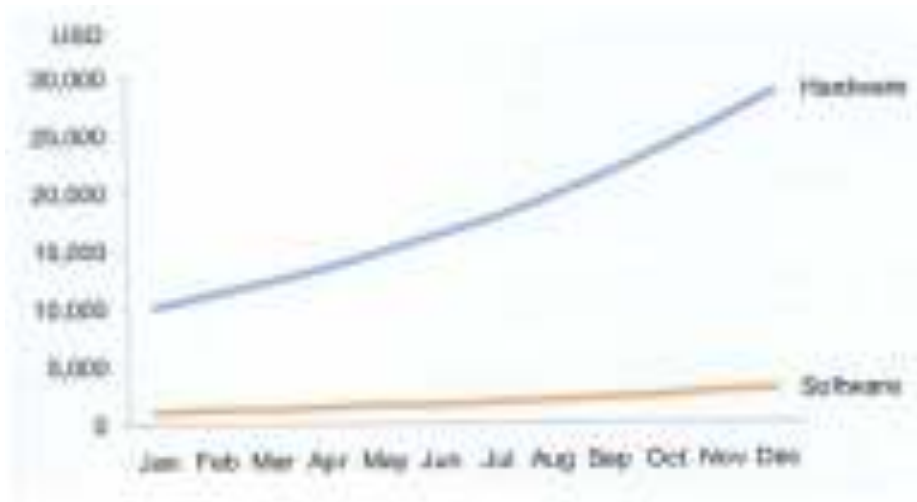


Še ena manipulativna vizualizacija



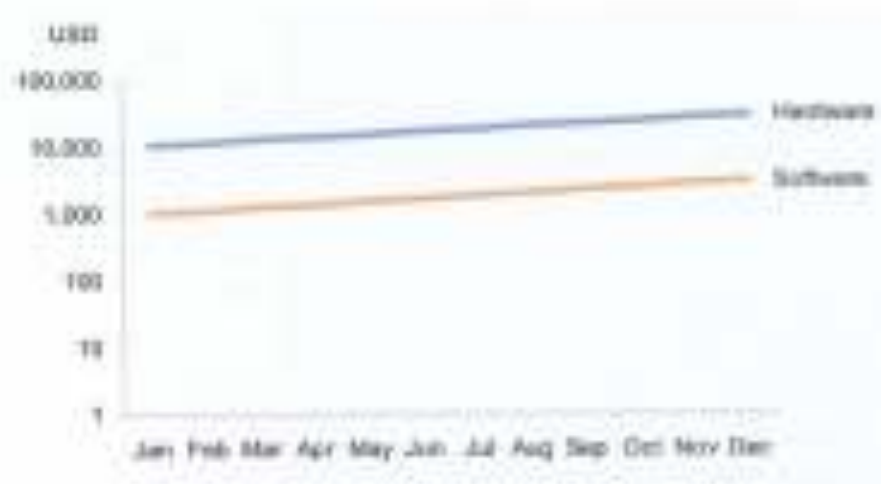
Različne skale.

Linearna ali logaritemska skala?

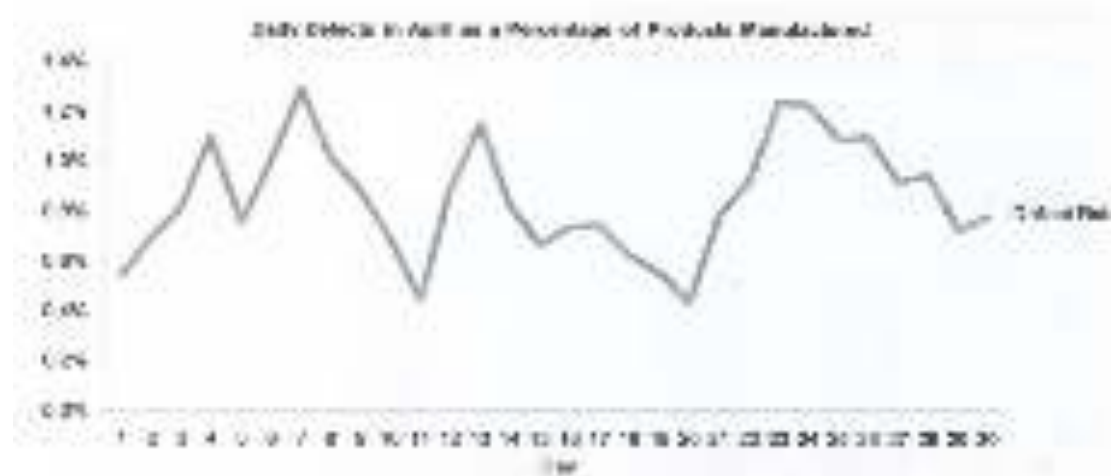


Recimo, da se stroški za strojno in programsko opremo vsako leto potrojijo.

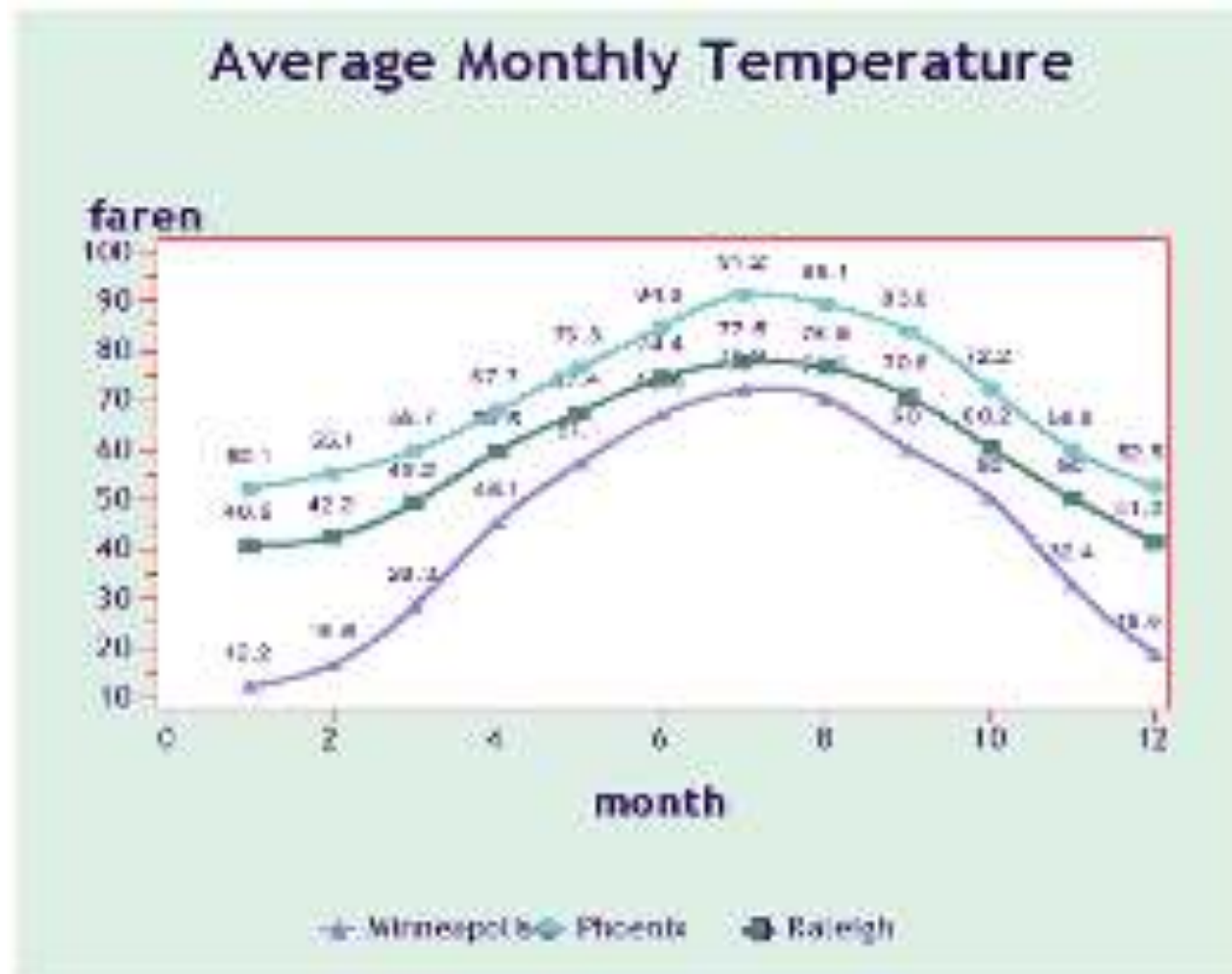
Logaritemski graf kaže relativne spremembe: eksponentno povečanje bo prikazano linearno.

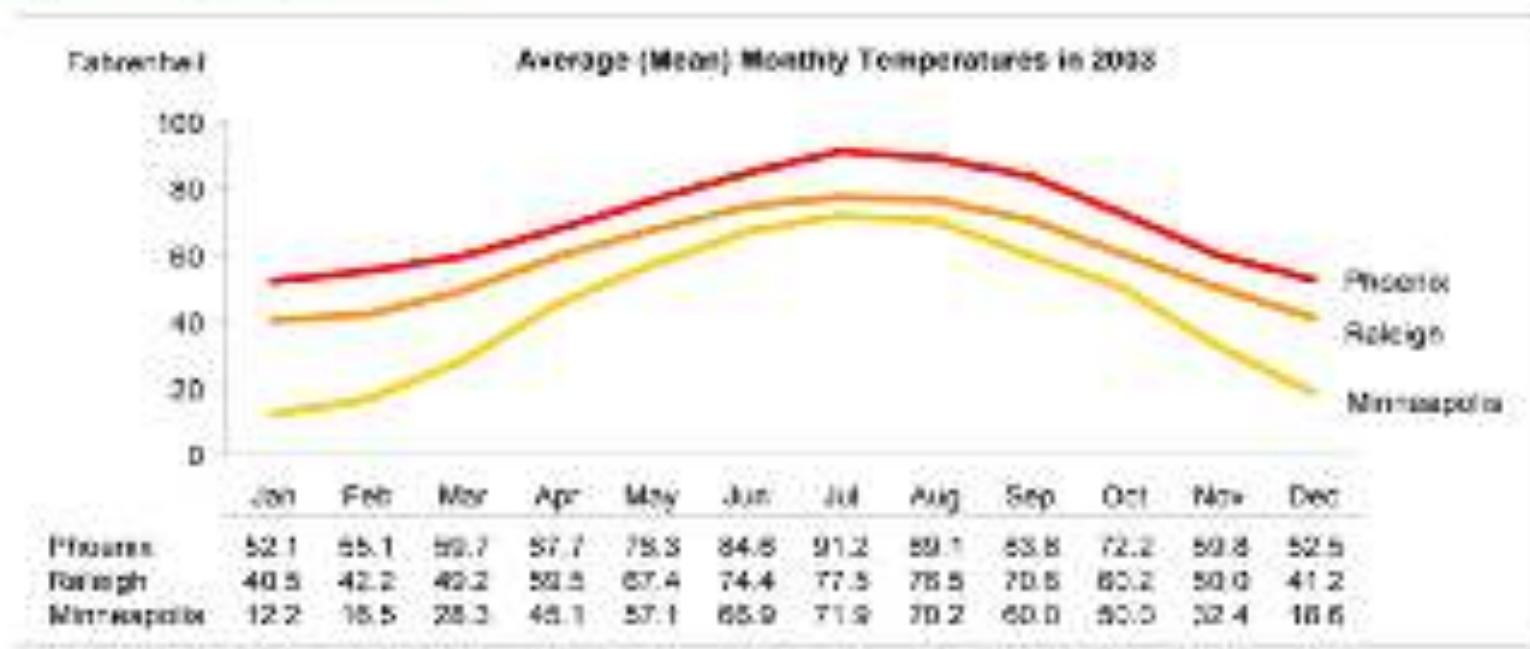
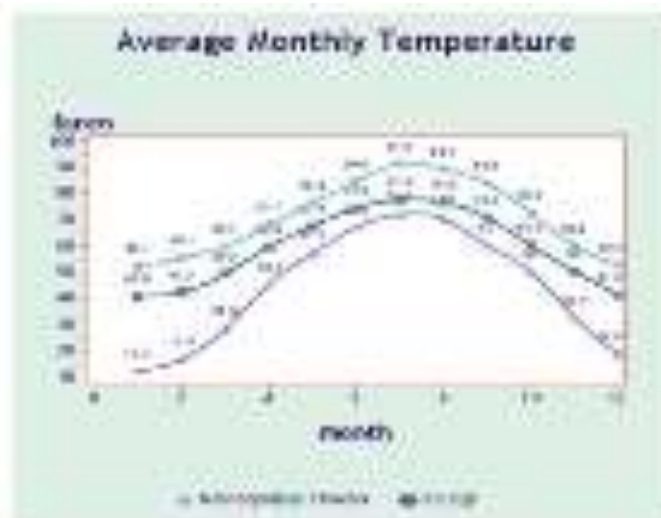


Krivulje - referenčna vrednost

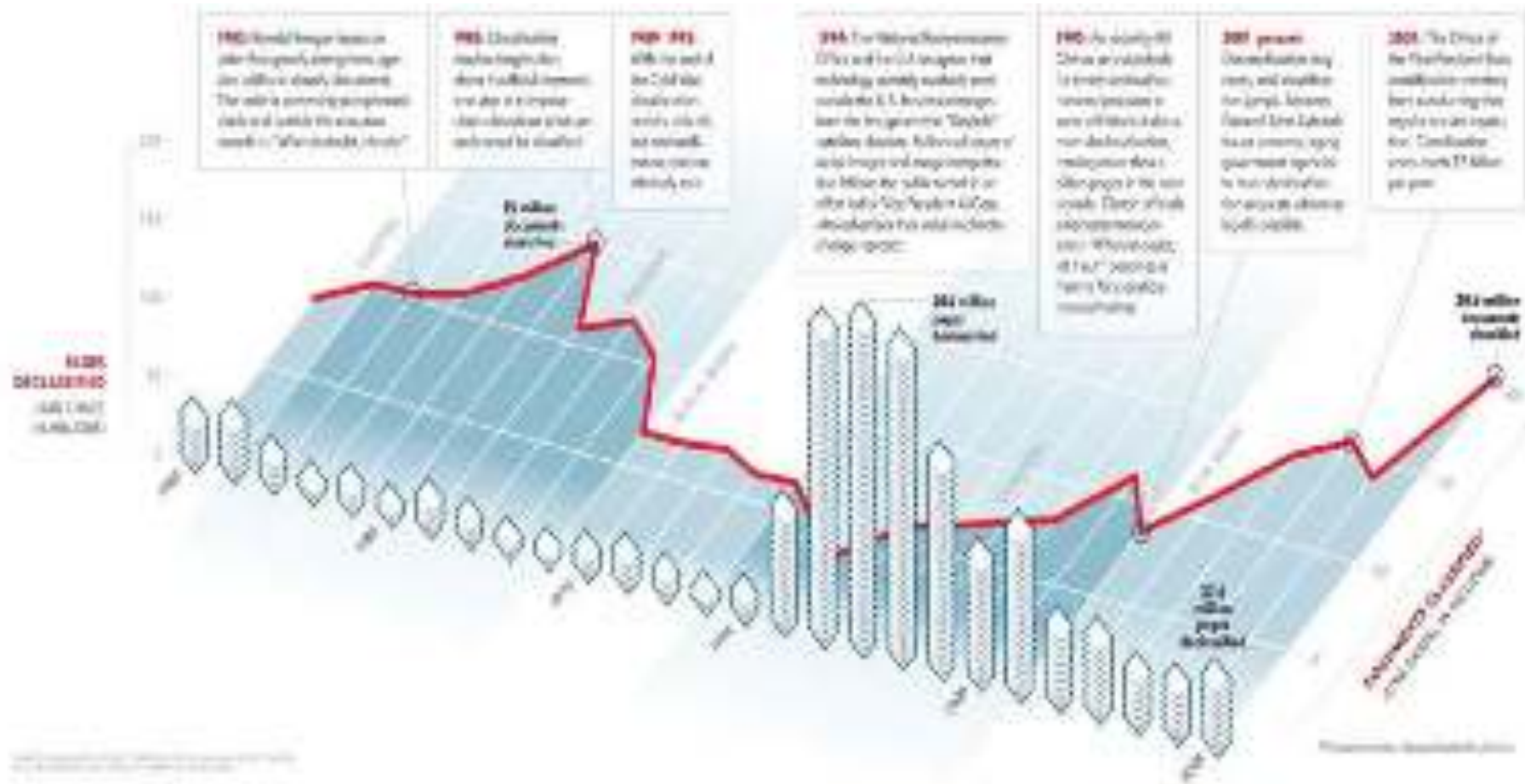


Krivulje – izogibanje pisanja števil v grafe

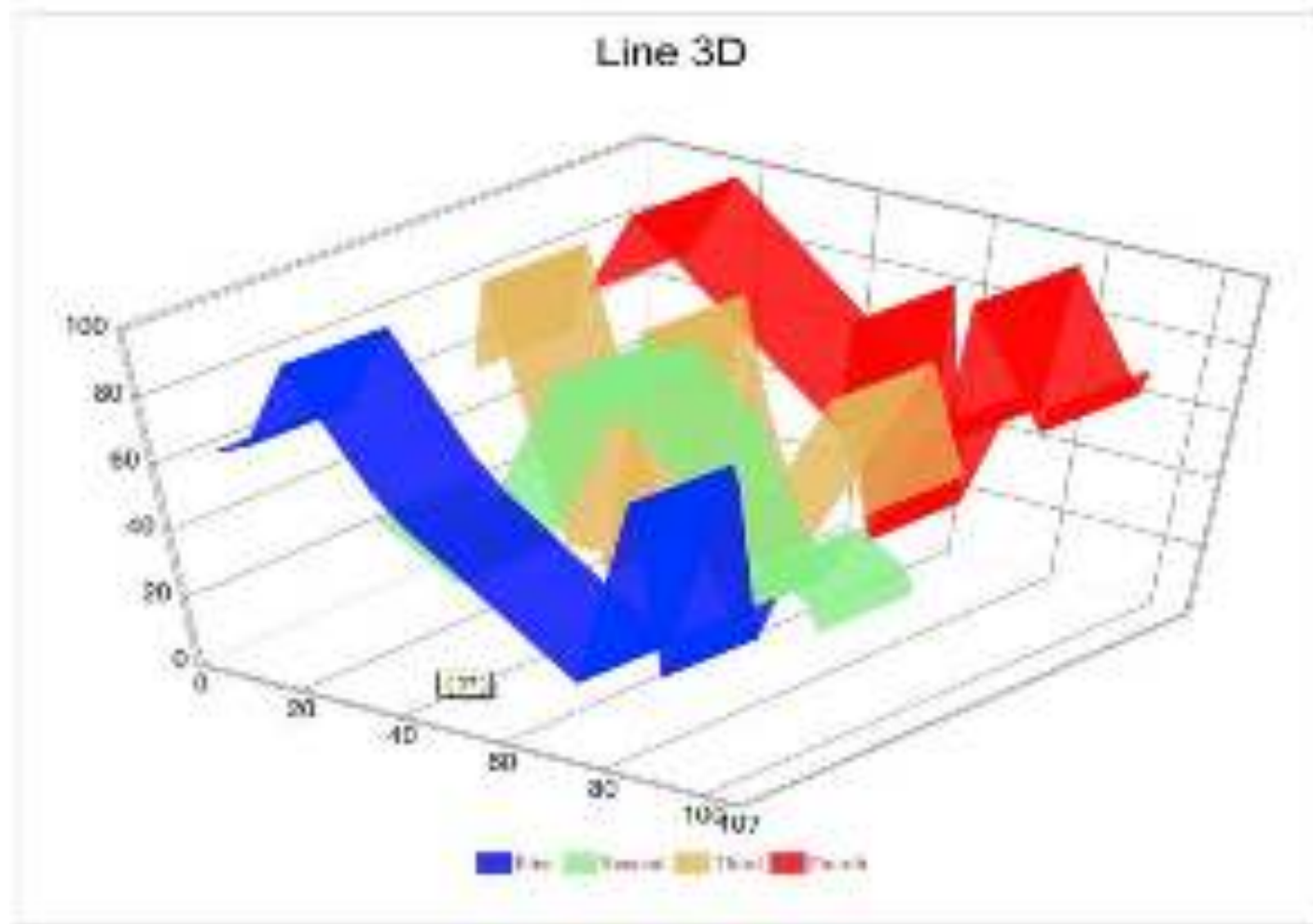




Nepotreben 3D pokvari preglednost

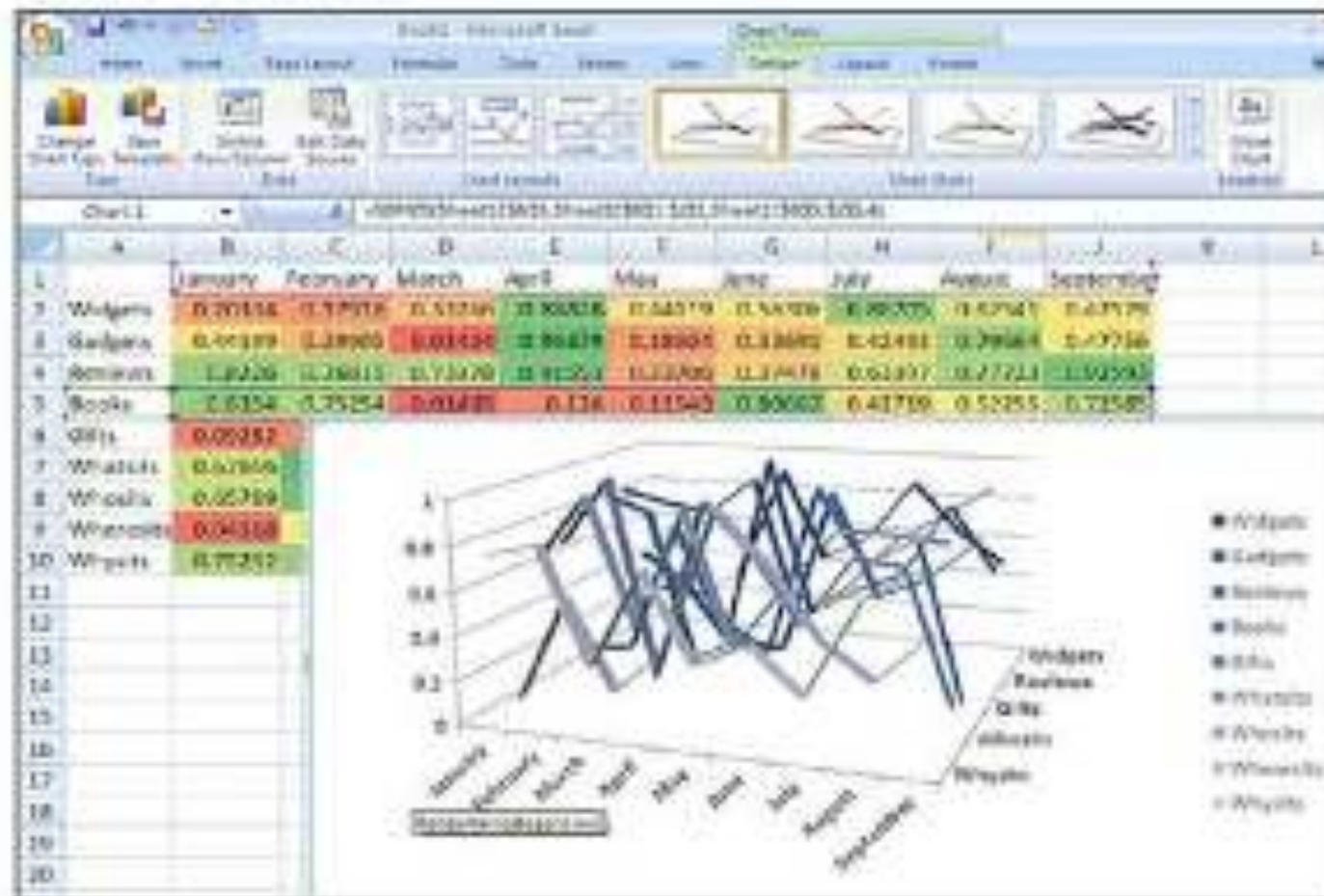


Bowl-of-noodles graph



Spider plot

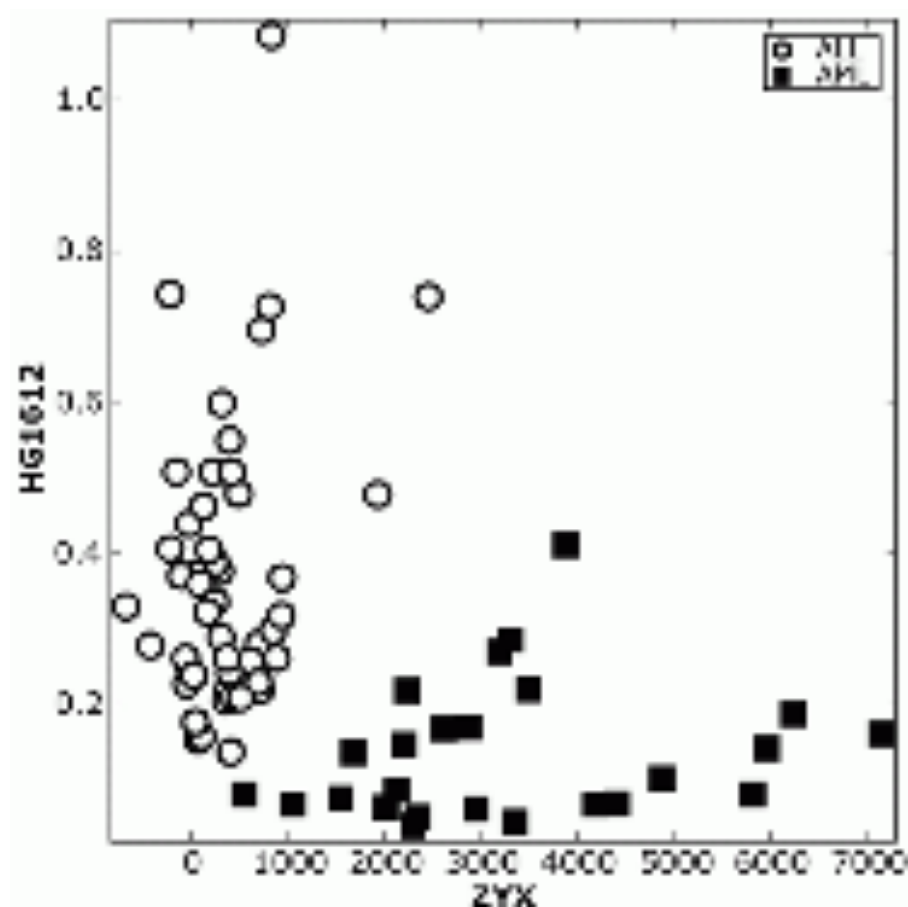
Microsoft Office 2007 Beta 2



Slike več spremenljivk hkrati

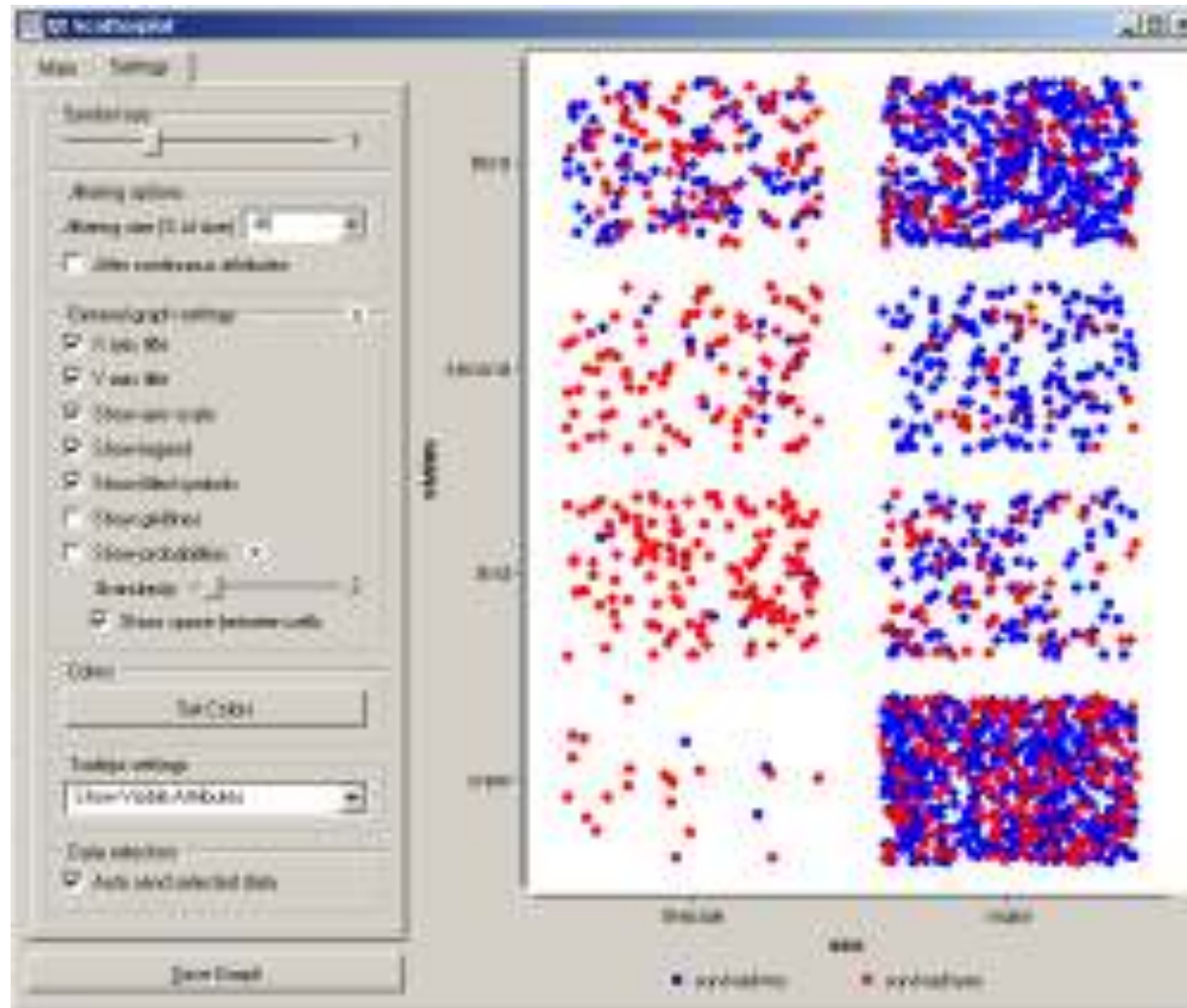
- razsevni diagram (scatter plot)
- radviz
- polyviz
- vzporedne koordinate (parallel plot)
- parketni diagram (mosaic)
- sieve

Razsevni diagram (scatter plot)



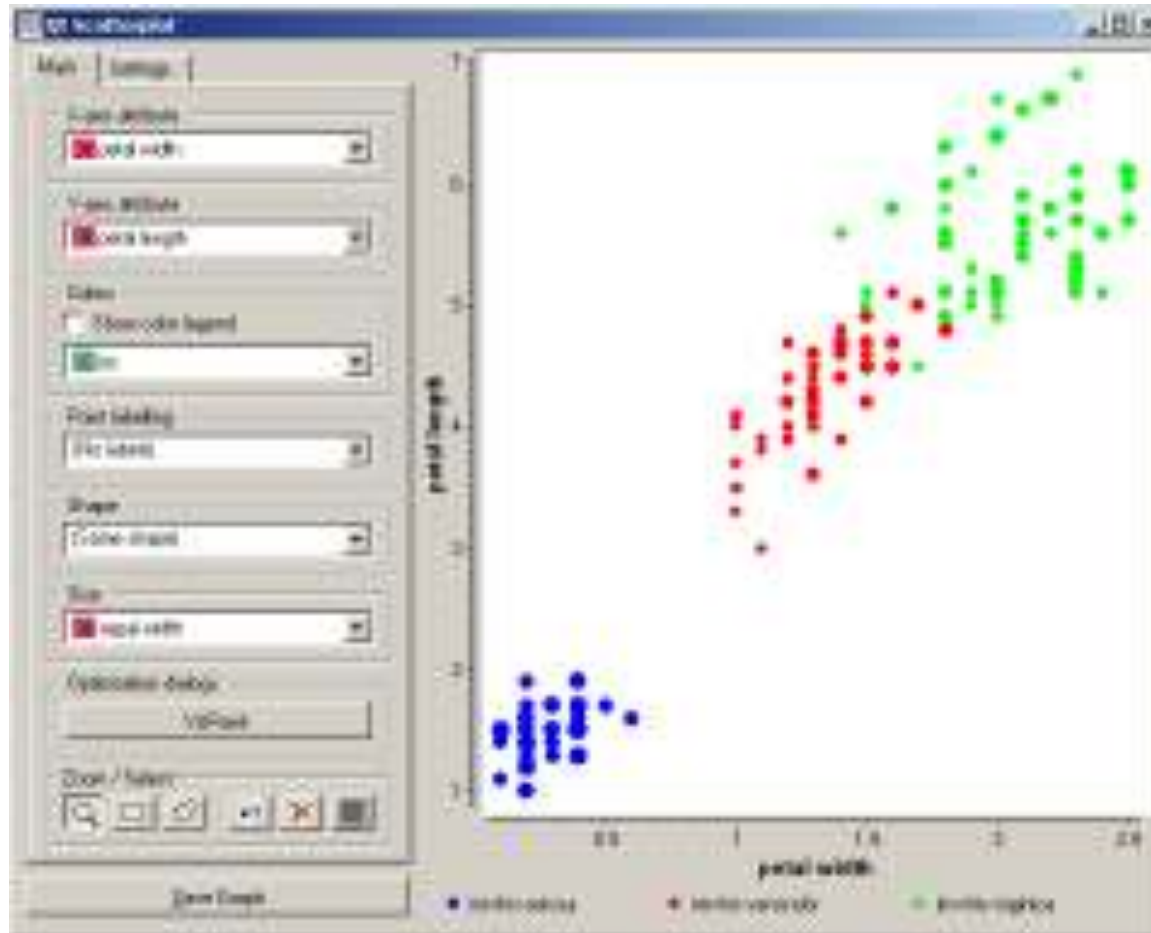
- Kažejo posamezne primere, ne statistik.
- Kažejo vrednost dveh spremenljivk (zveznih) + še kaj z obliko, barvo, velikostjo.
- Prikažejo korelacije in druge zveze.

Razsevni diagram (scatter plot)



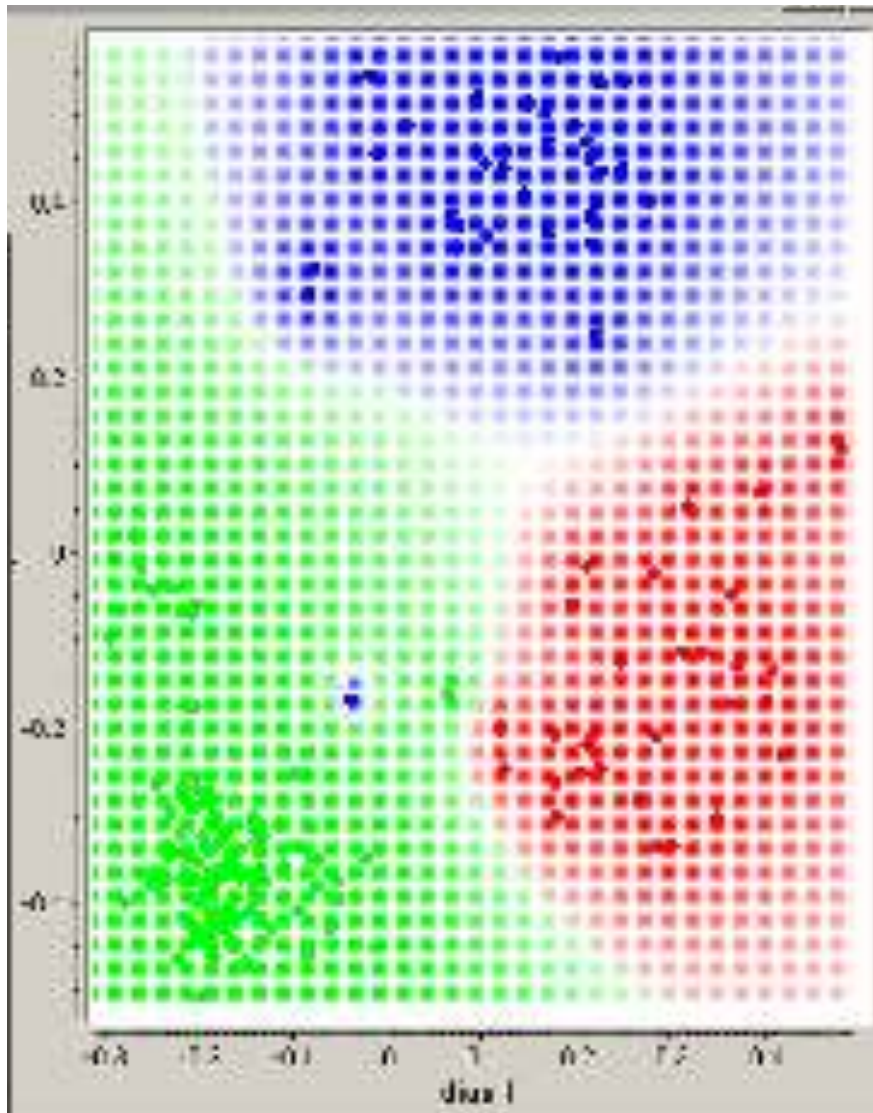
- Prekrivanje rešimo z jitteringom in transparentnostjo.

Razsevni diagram (scatter plot)



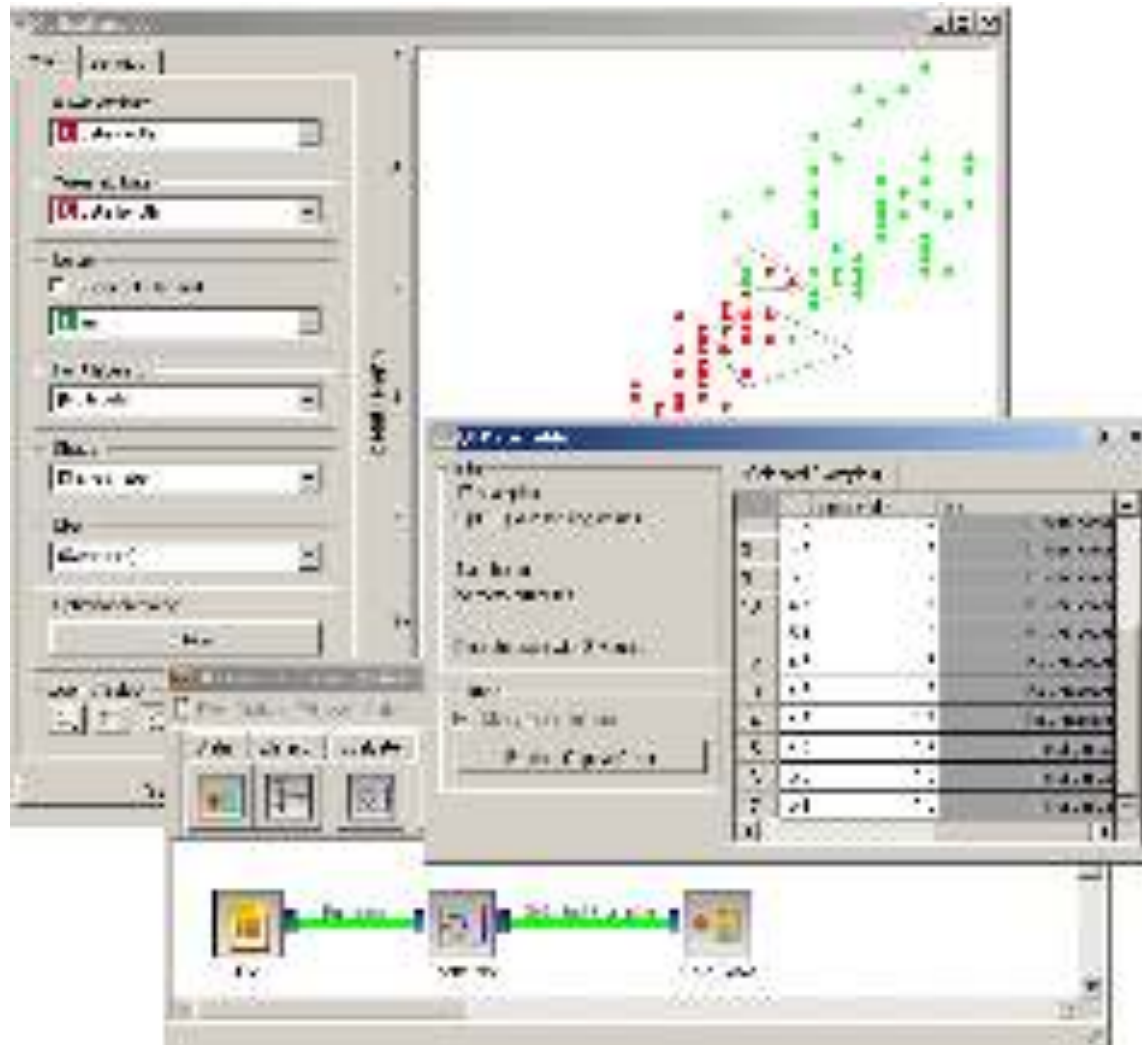
- Dodatni atributi:
 - barva: dobra za diskretne sicer težko razberemo skalo
 - velikost: težko razbrati
 - oblika: slabo vidno
 - labele: prekrivanje

Razsevni diagram (scatter plot)



- Ozadje prikaže verjetnosti.
- Iskanje osamelcev (outliers).

Razsevni diagram (scatter plot)



- Izbiranje posameznih primerov.

Public Data

Immunization, DPT (% of child...
 Immunization, measles (% of c...
 Improved sanitation facilities (%...
 Improved water source (% with...
 Incidence of tuberculosis (per 1...
Life expectancy
 Lifetime risk of maternal death (...
 Lifetime risk of maternal death (...
 Low-birthweight babies (% of b...
 Malnutrition prevalence, height...
 Malnutrition prevalence, weight...

Clear

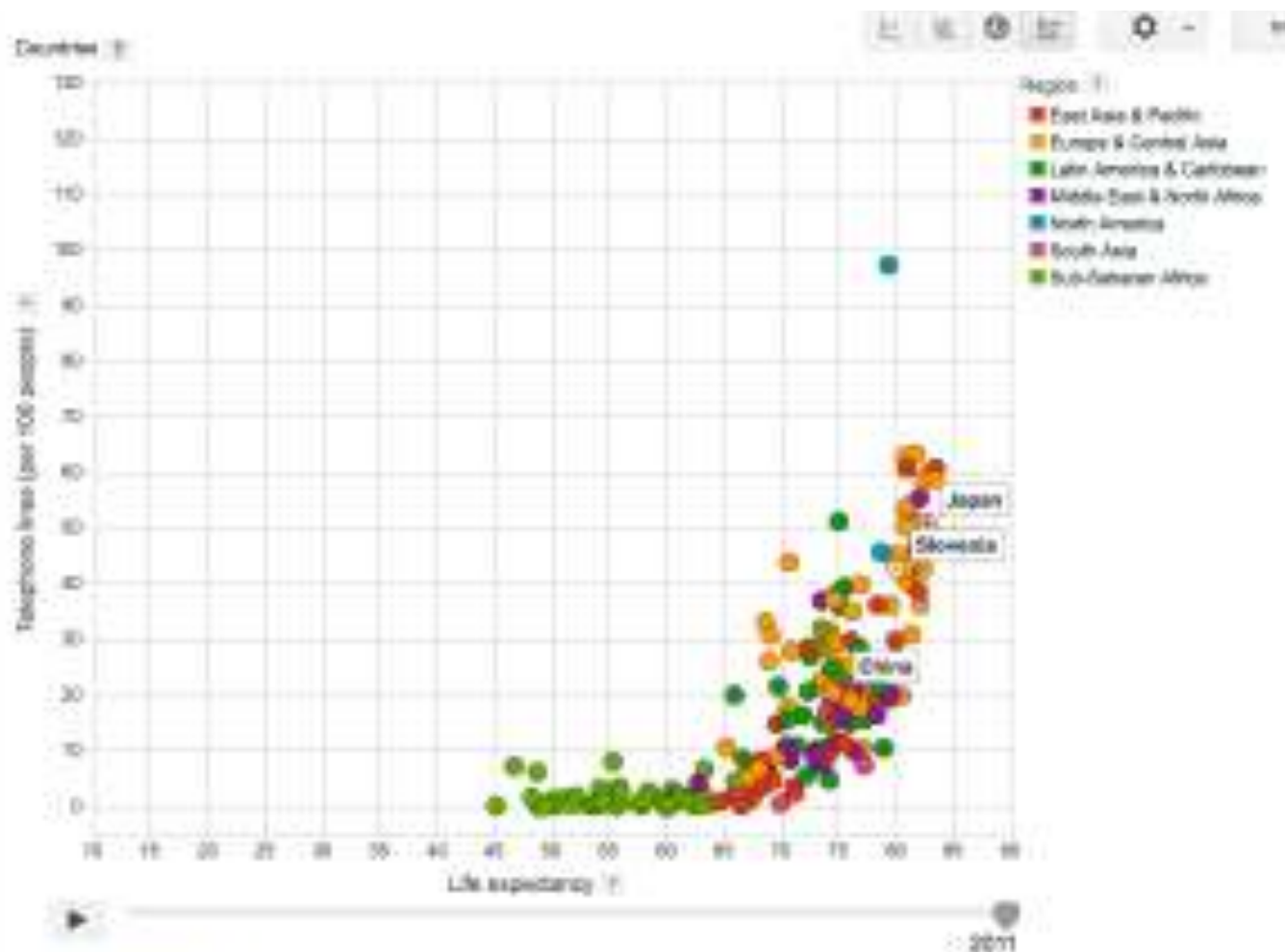
Compare by: Country

Region: Color

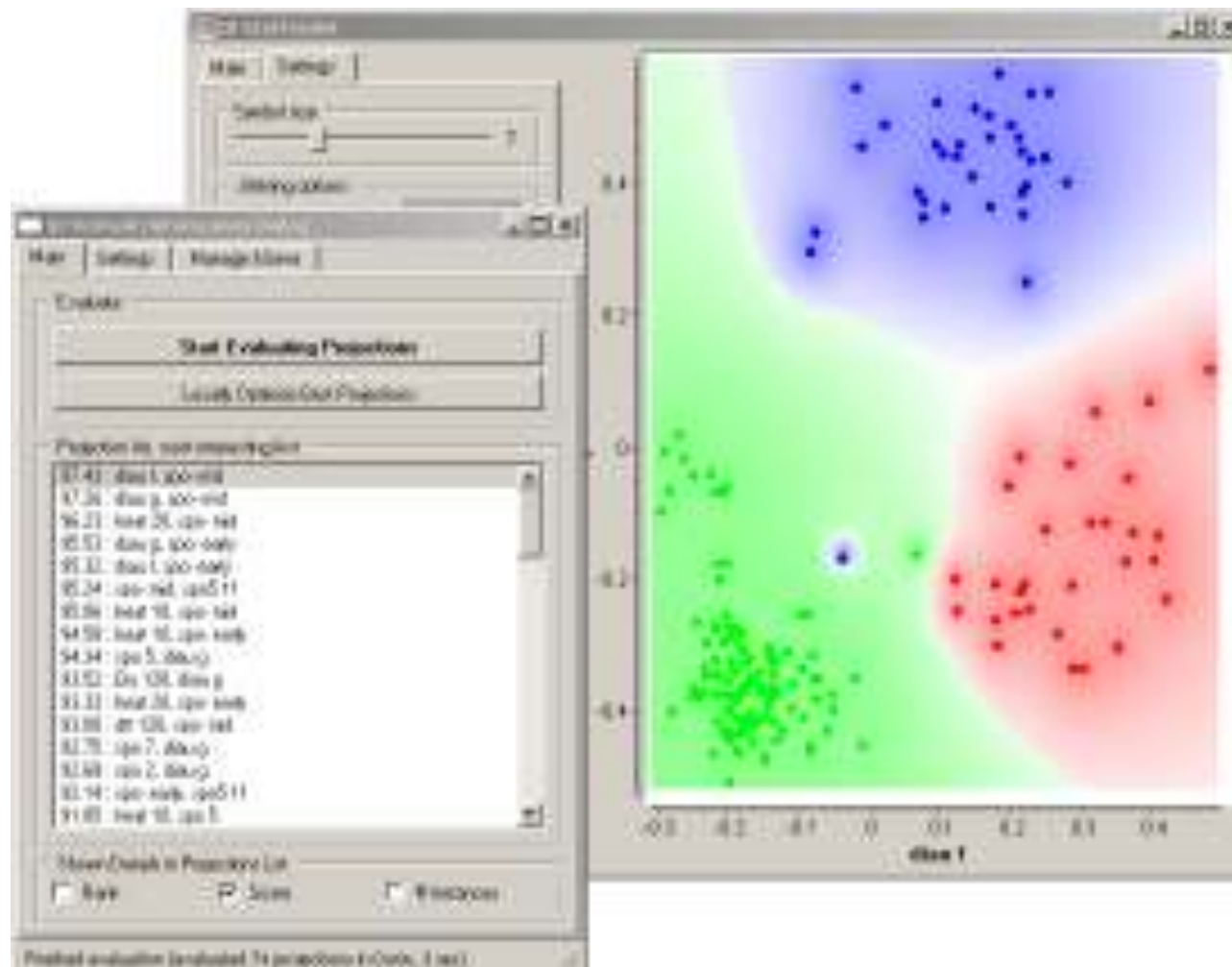
Loading Type: All

Income Level: All

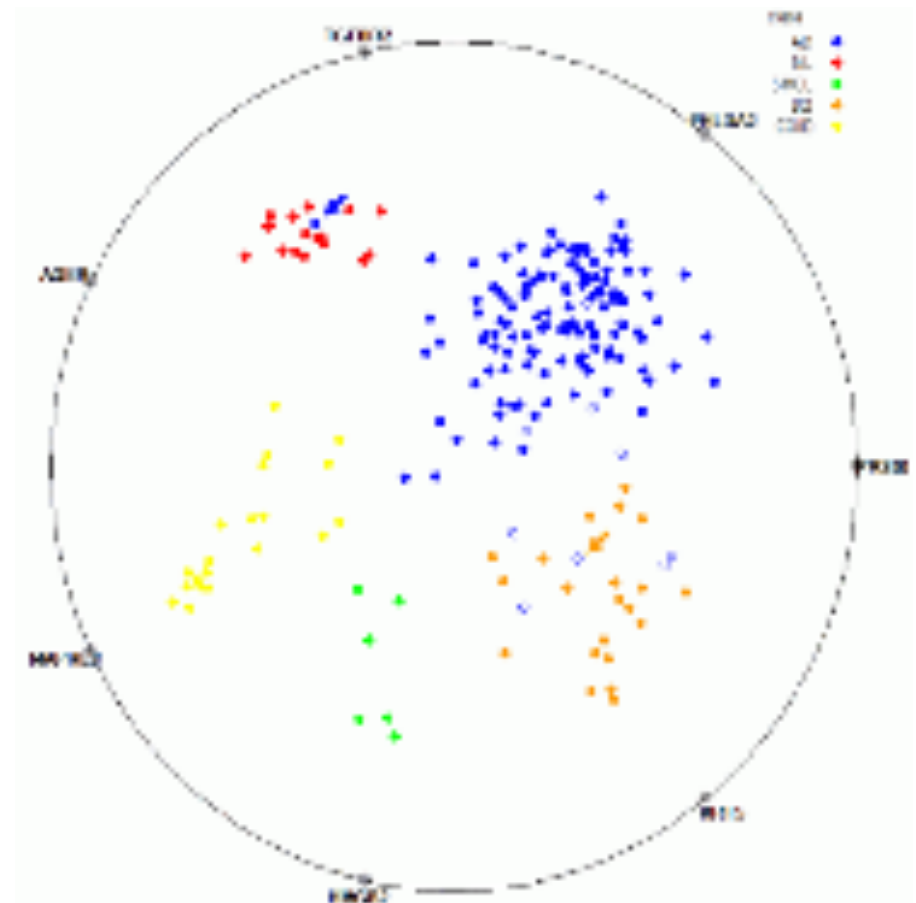
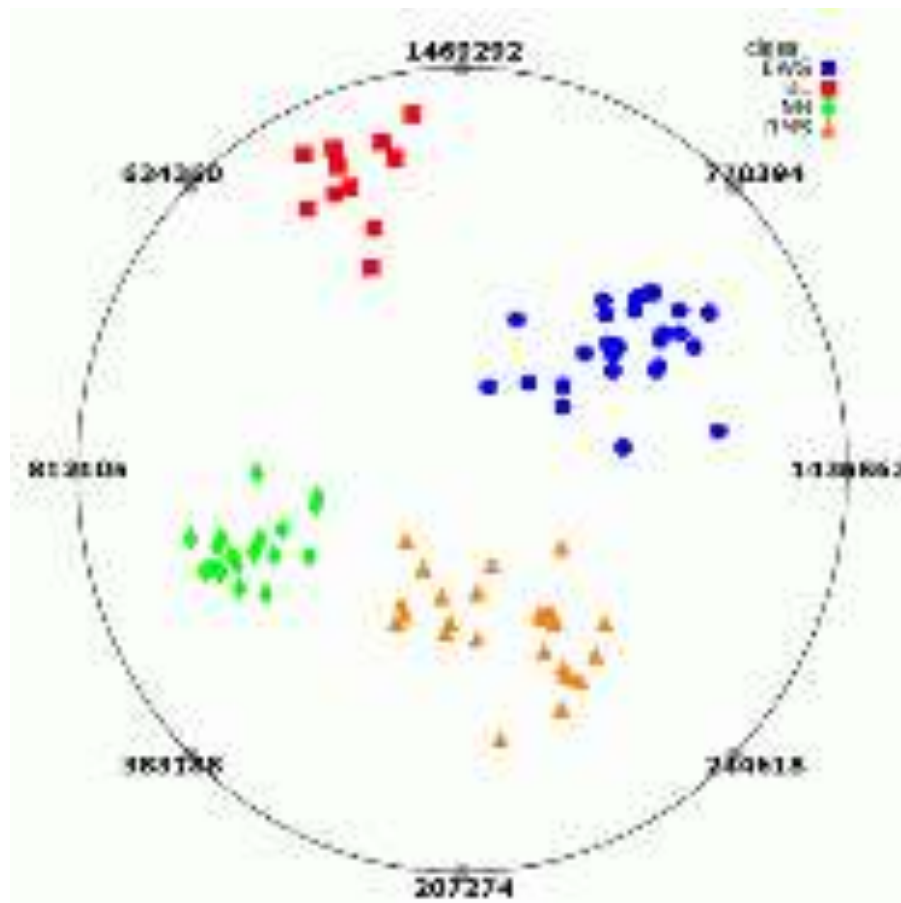
São Tomé and Príncipe
 Saudi Arabia
 Senegal
 Serbia
 Seychelles
 Sierra Leone
 Singapore
 Sint Maarten (Dutch part)
 Slovak Republic
 ✓ Slovakia
 Solomon Islands
 Clear selections



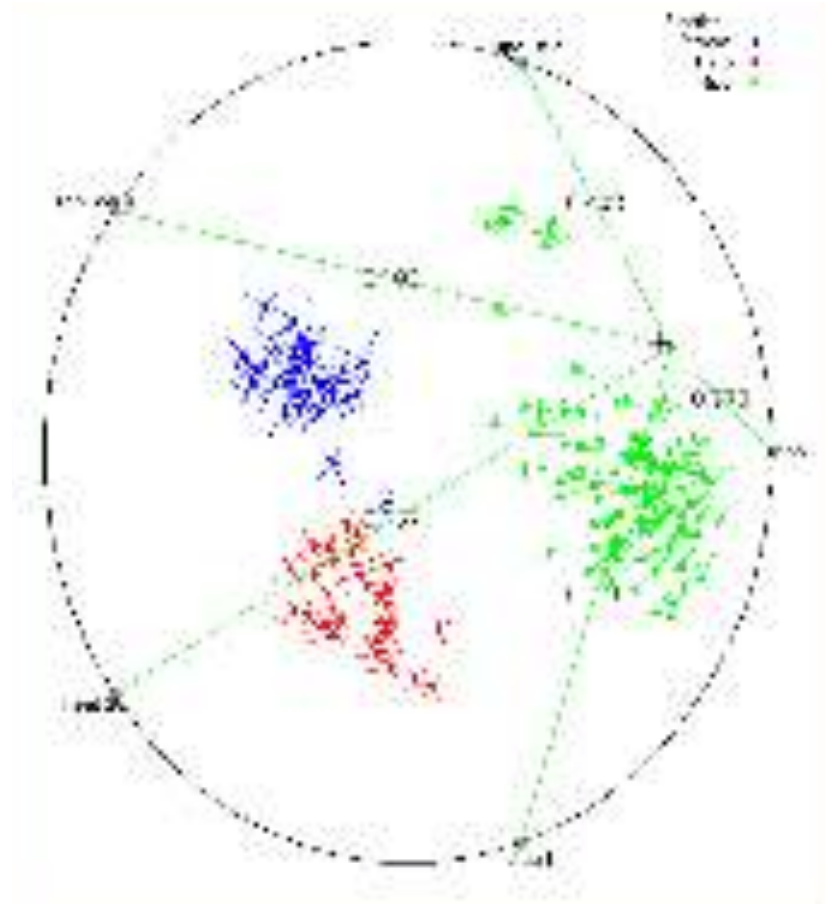
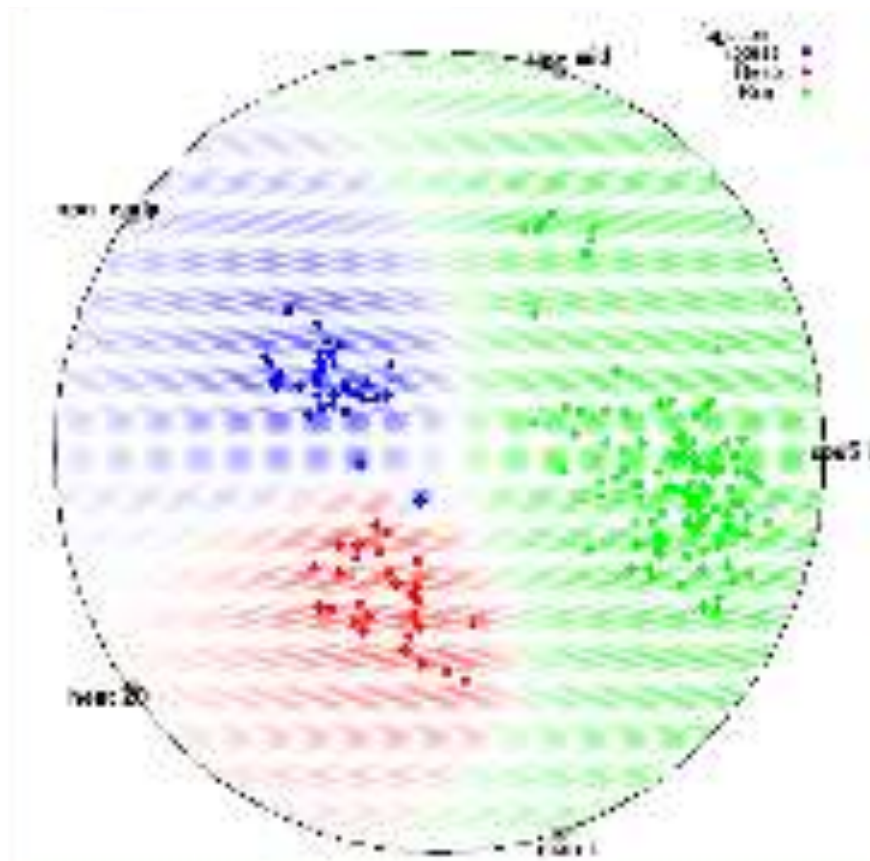
VizRank – iskanje informativnih diagramov



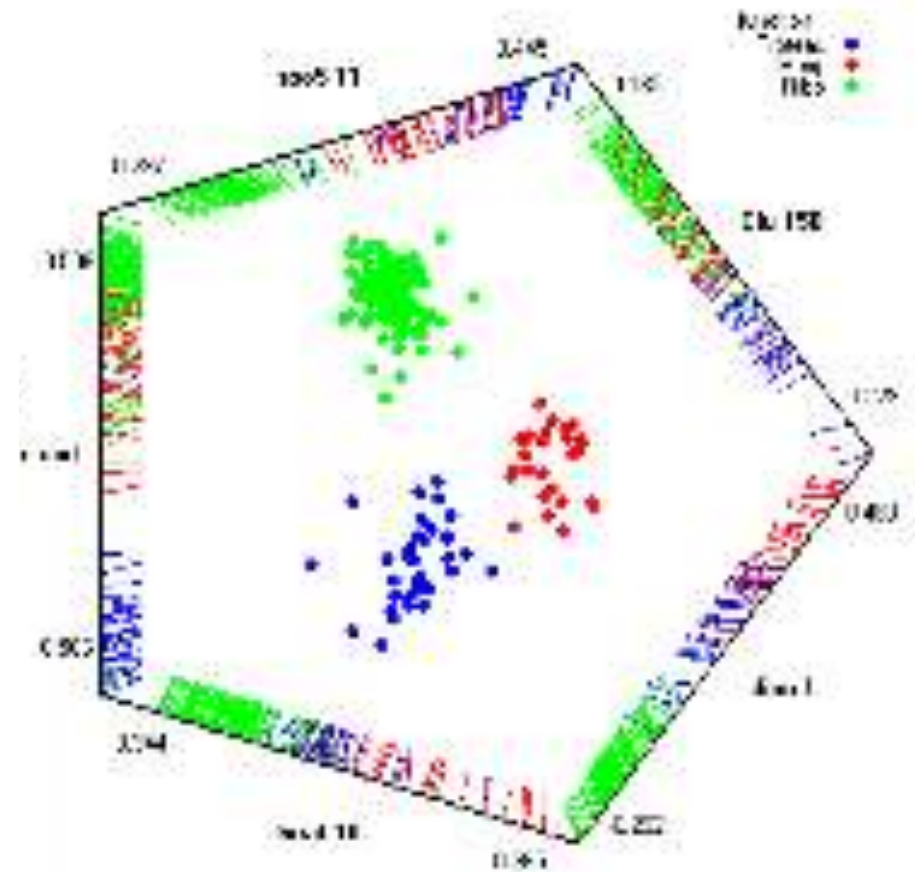
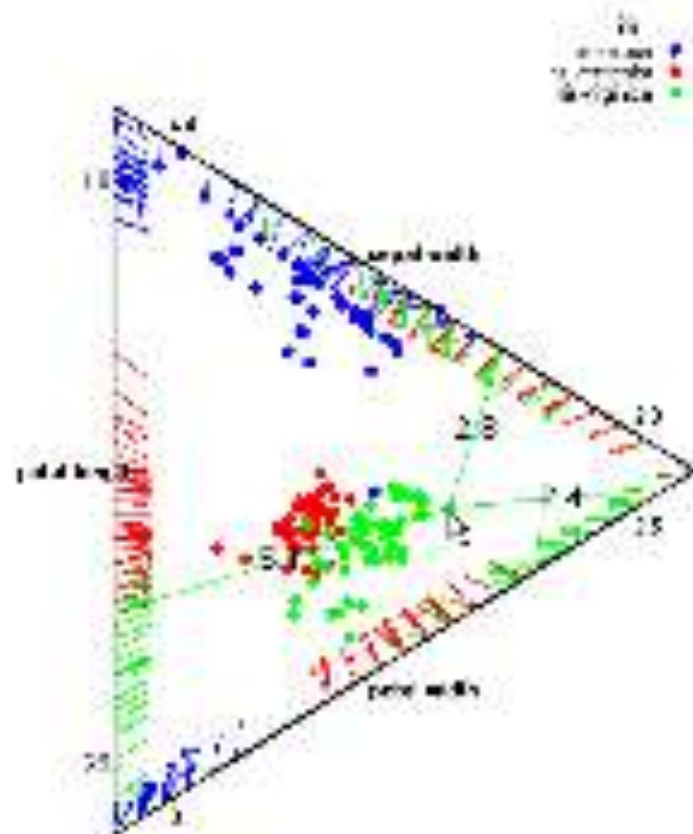
RadViz



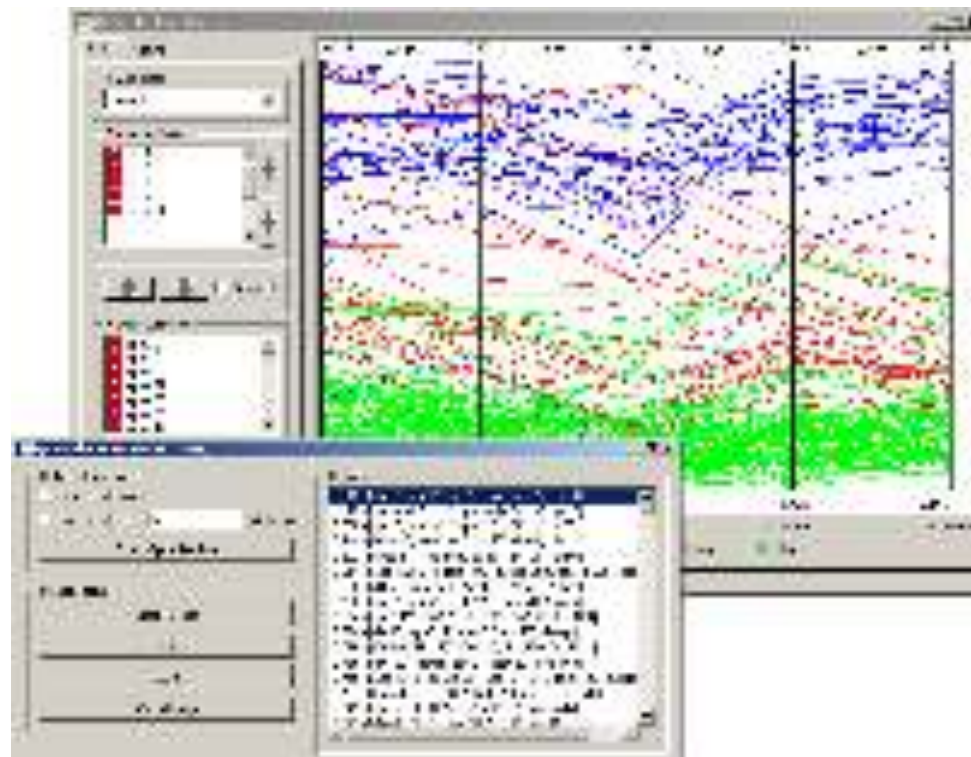
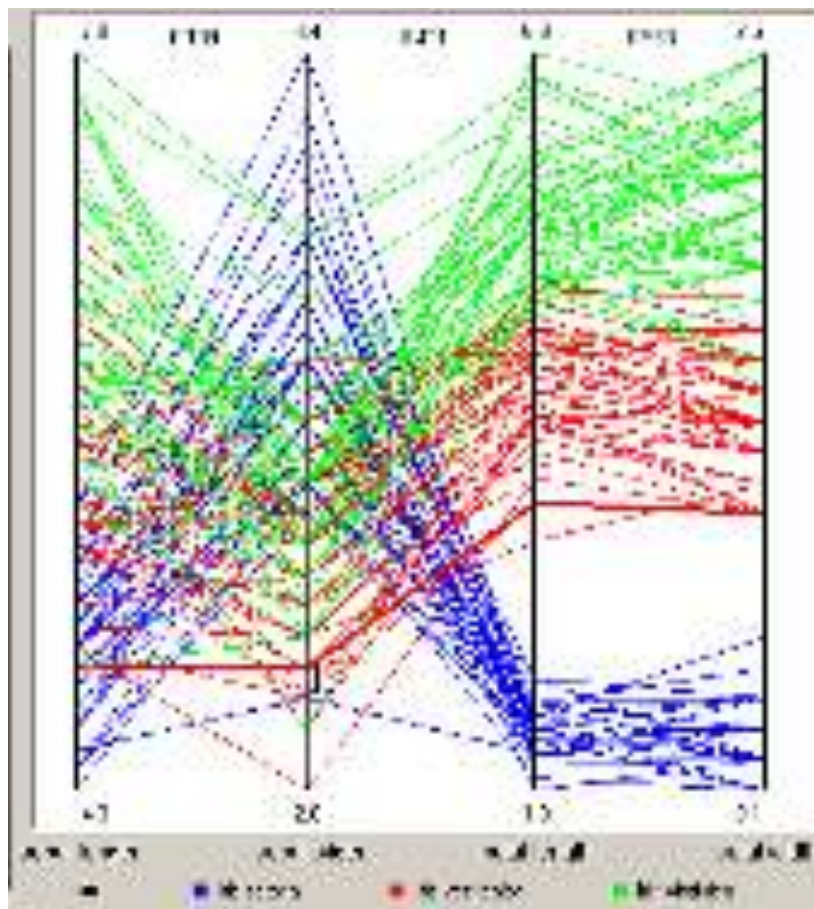
RadViz



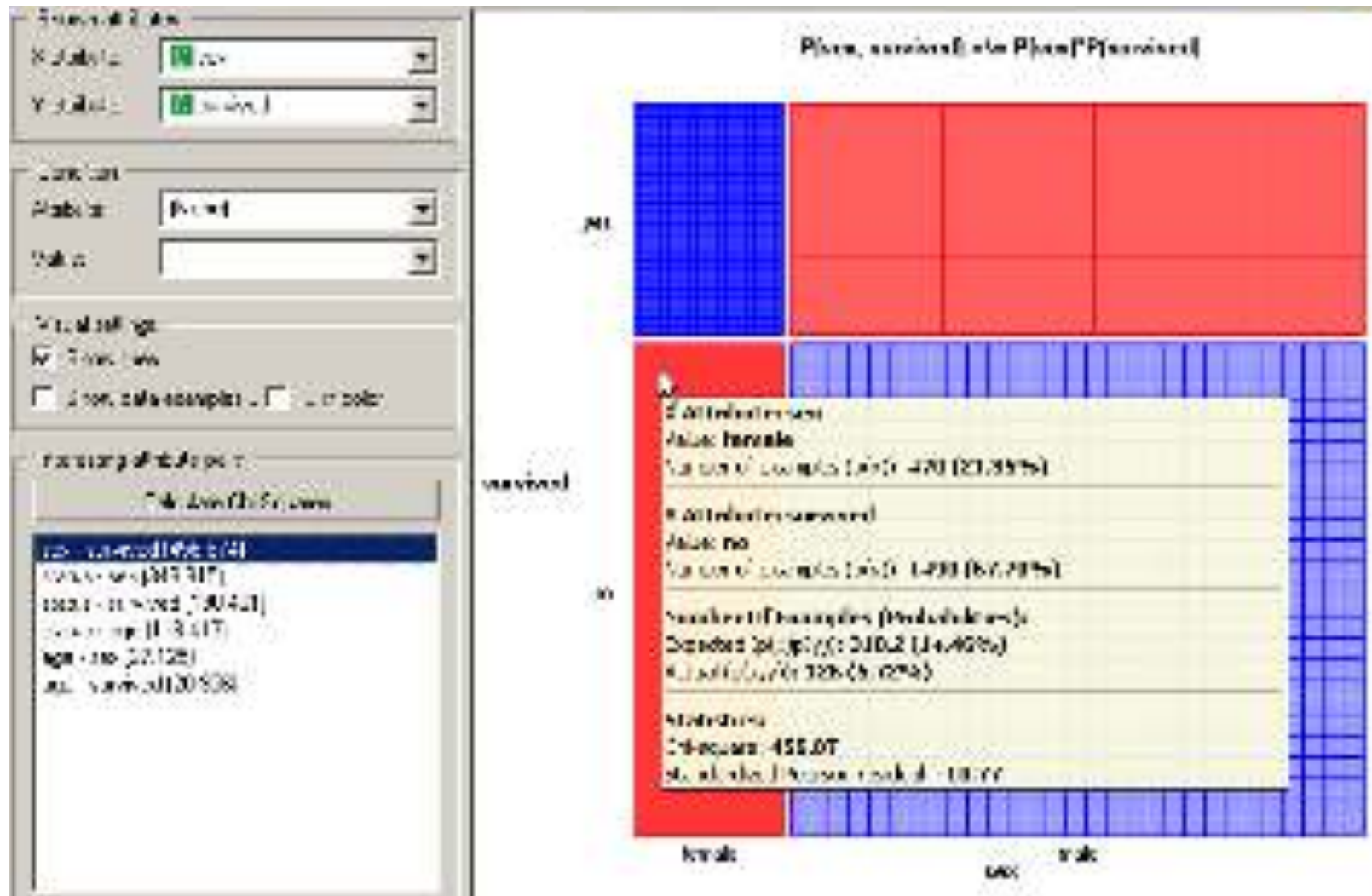
PolyViz



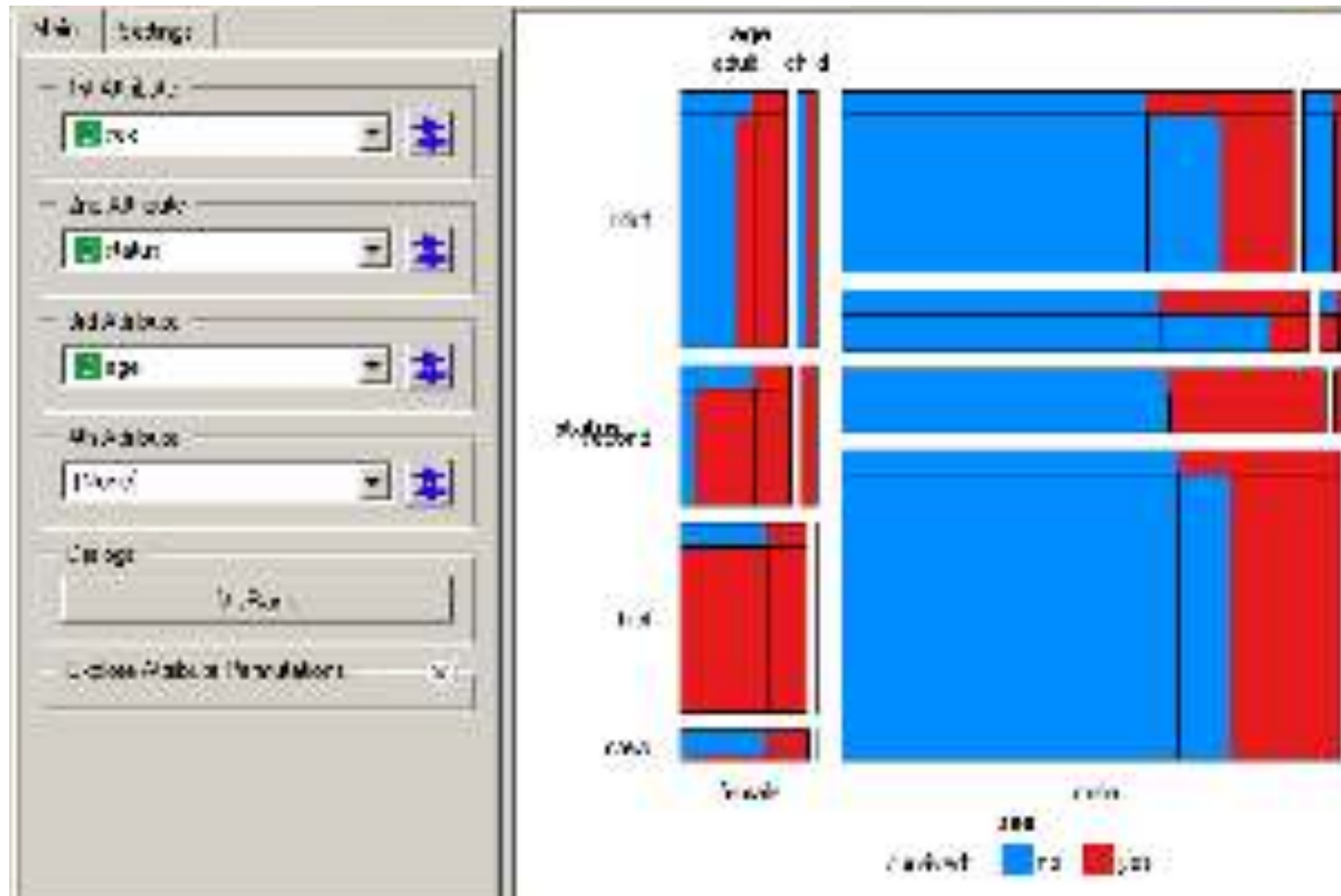
Vzporedne koordinate (parallel coordinates)



Sieve diagram



Parquetni diagram (mosaic)



Feedback

<http://goo.gl/forms/1Cy4PA3MRn>

