



Grafična predstavitev multivariatnih podatkov

Multivariatna analiza



Načrt

- Načela dobre grafične predstavitve
- Različni prikazi multivariatnih podatkov

10 osnovnih načel dobre grafične predstavitve

(Cleroux, Lepage, Ranger, 1984)

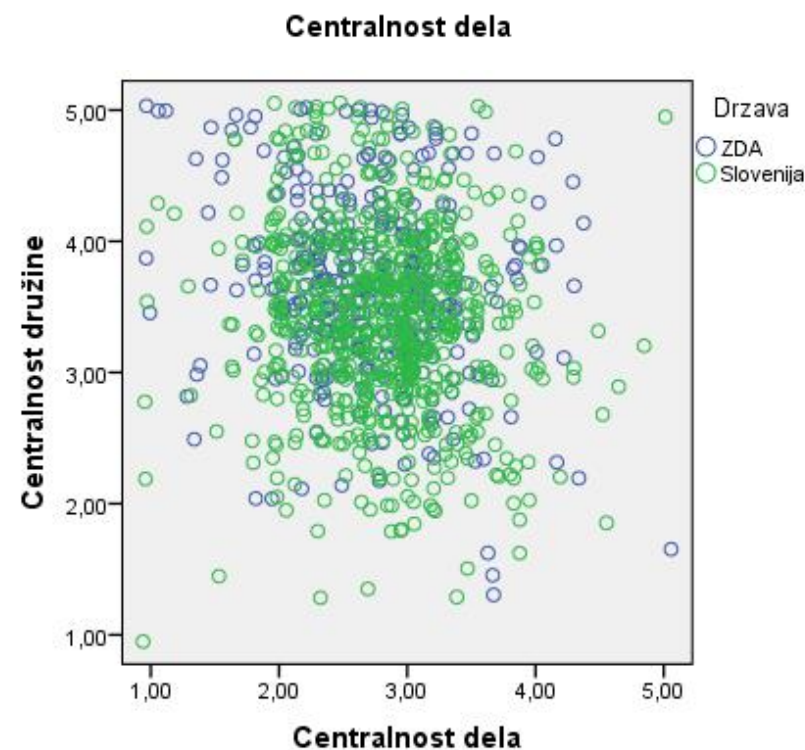
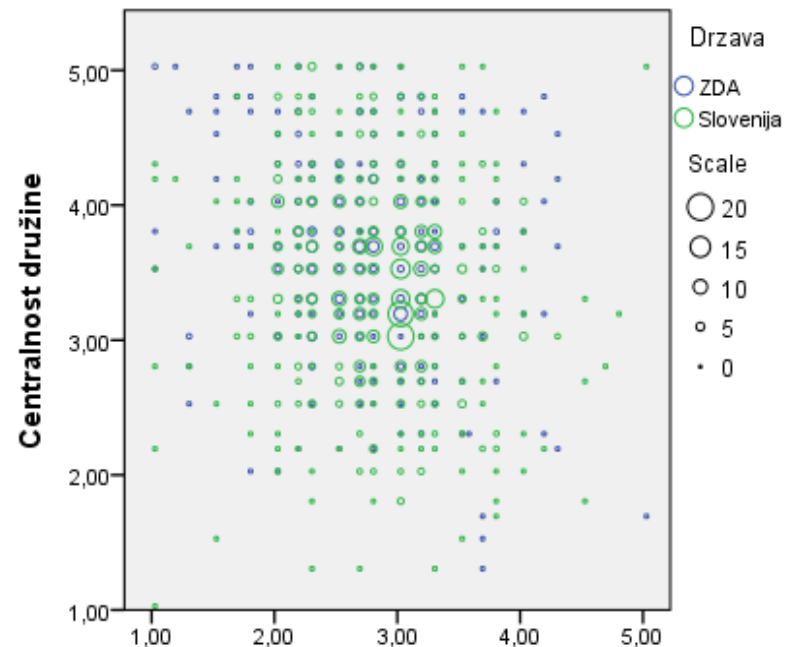
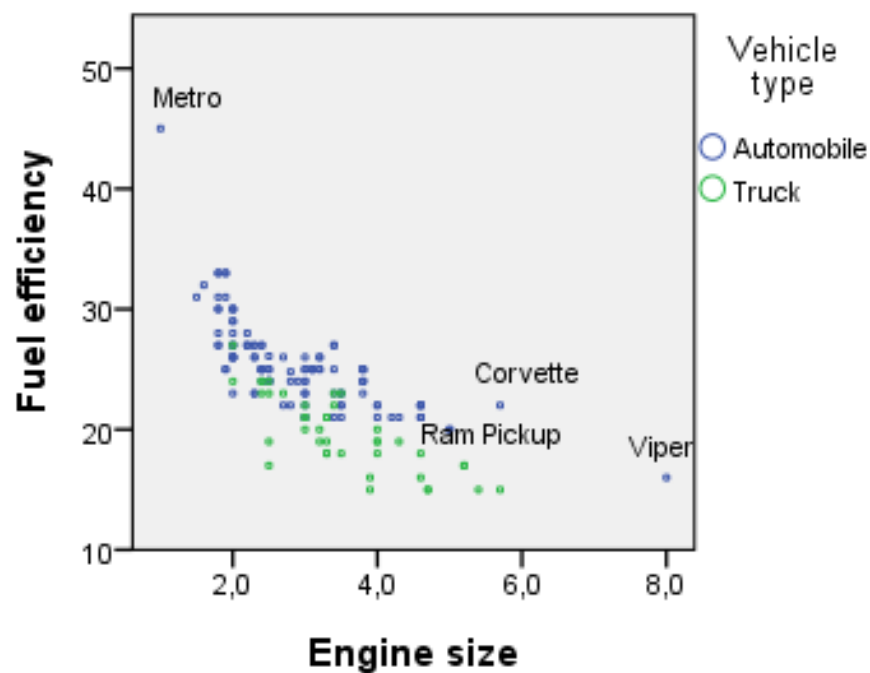
- Preprosto in hitro posredovanje informacije
- Pomoč pri razumevanju informacije
- Poudarek na pomembnih informacijah
- Mnemoničnost (zapomljivost)
- Preprosta, zgoščena in privlačna
- Jasna, točna, nezmaličena
- Lahko in hitro razumljiva
- Uporaba običajnih oblik
- Možnost hkratne predstavitve več razsežnosti
- Možnost primerjave in razvrščanja med predstavitvami.

Grafični prikazi multivariatnih podatkov

- Razsevni grafikon 2D in 3D
- Večkratni (matrični) razsevni grafikoni
- Linijski grafikoni
- Zvezde
- Obrazi (Chernoff)
- Panelni grafikoni
- Redukcija prostora
- “biplot”

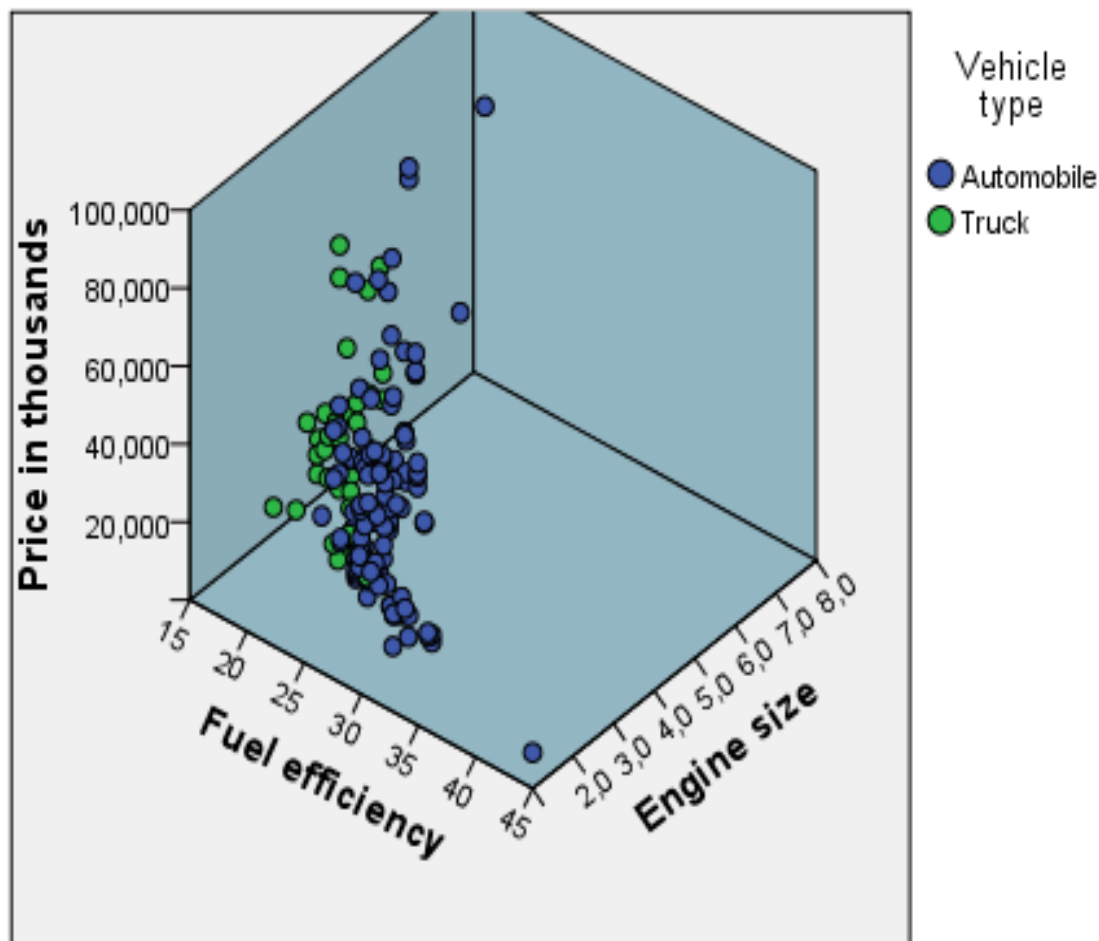
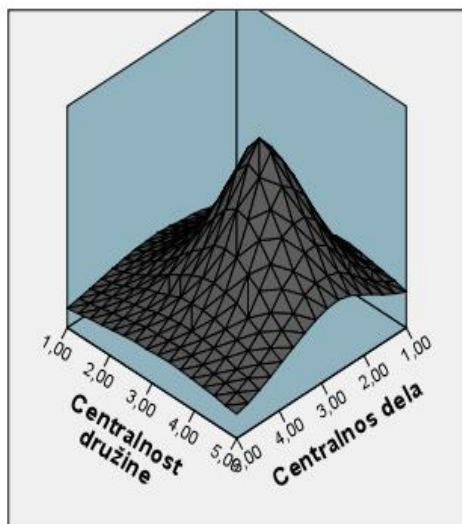
Razsevni grafikoni (2D in 3D)

2D Primeri:



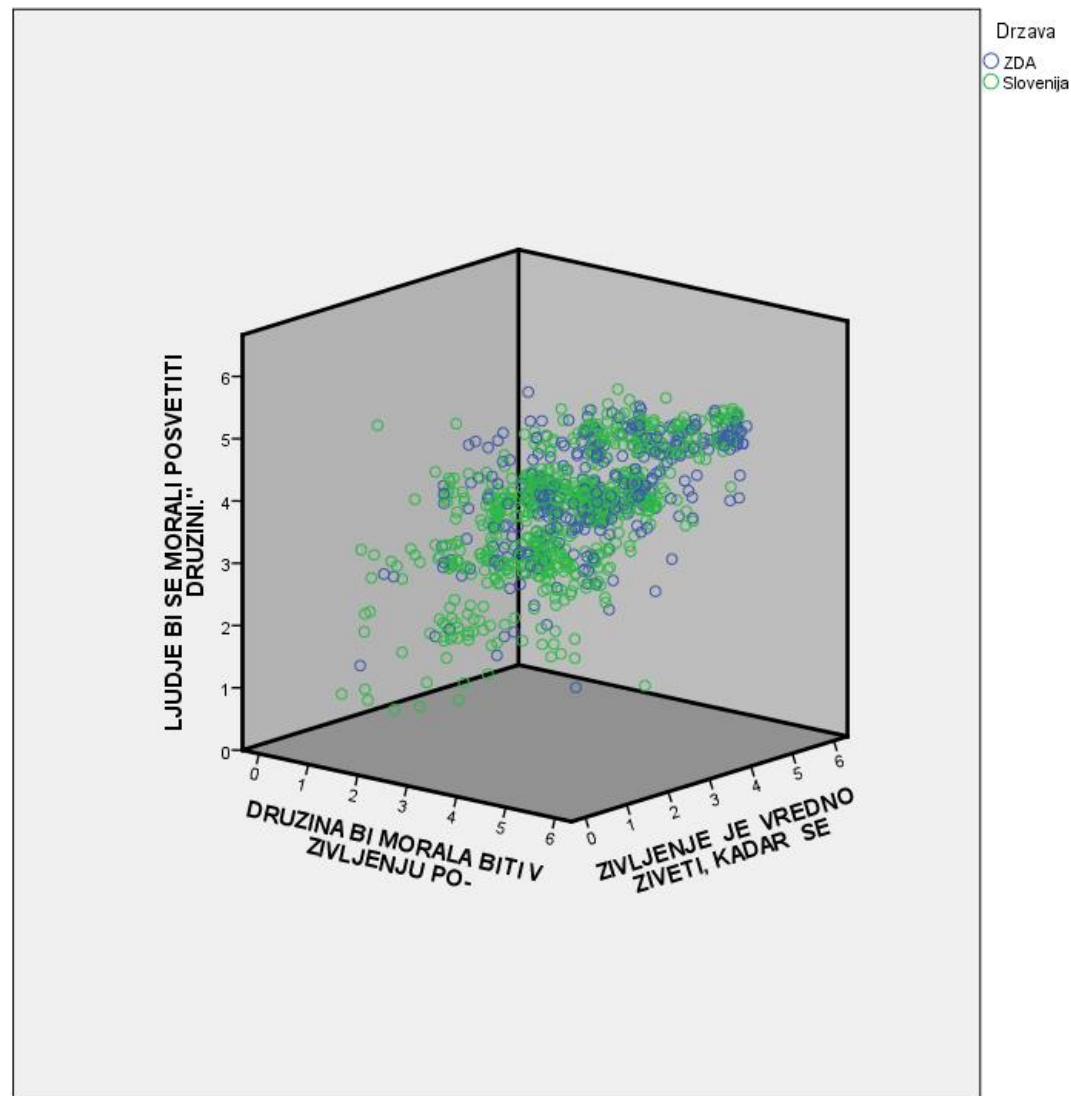
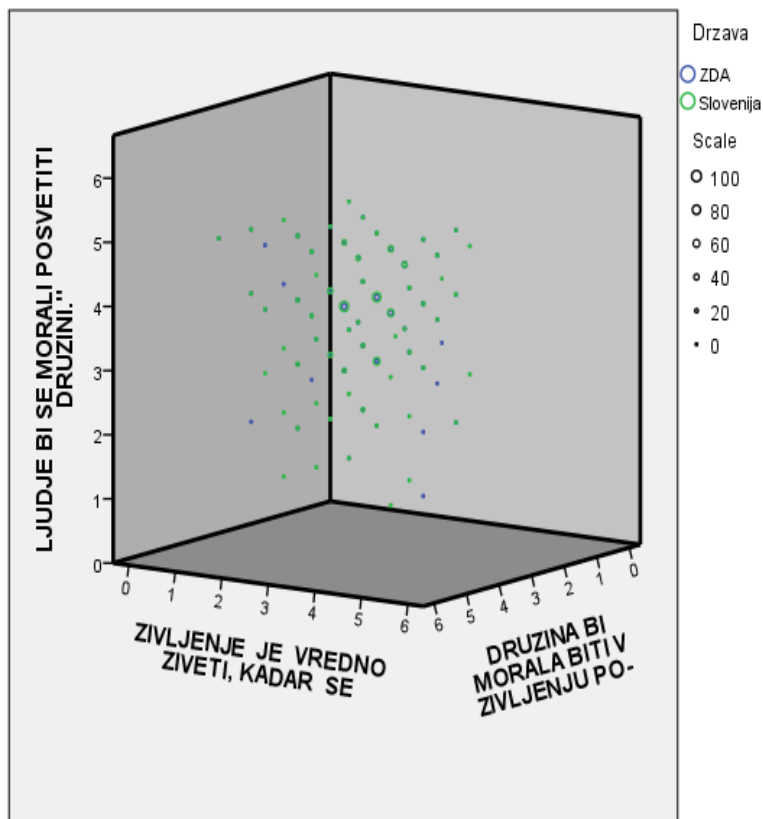
Razsevni grafikoni (2D in 3D)

3D Primeri:



Razsevni grafikoni (2D in 3D)

3D Primeri:



Razsevni grafikoni (2D in 3D)

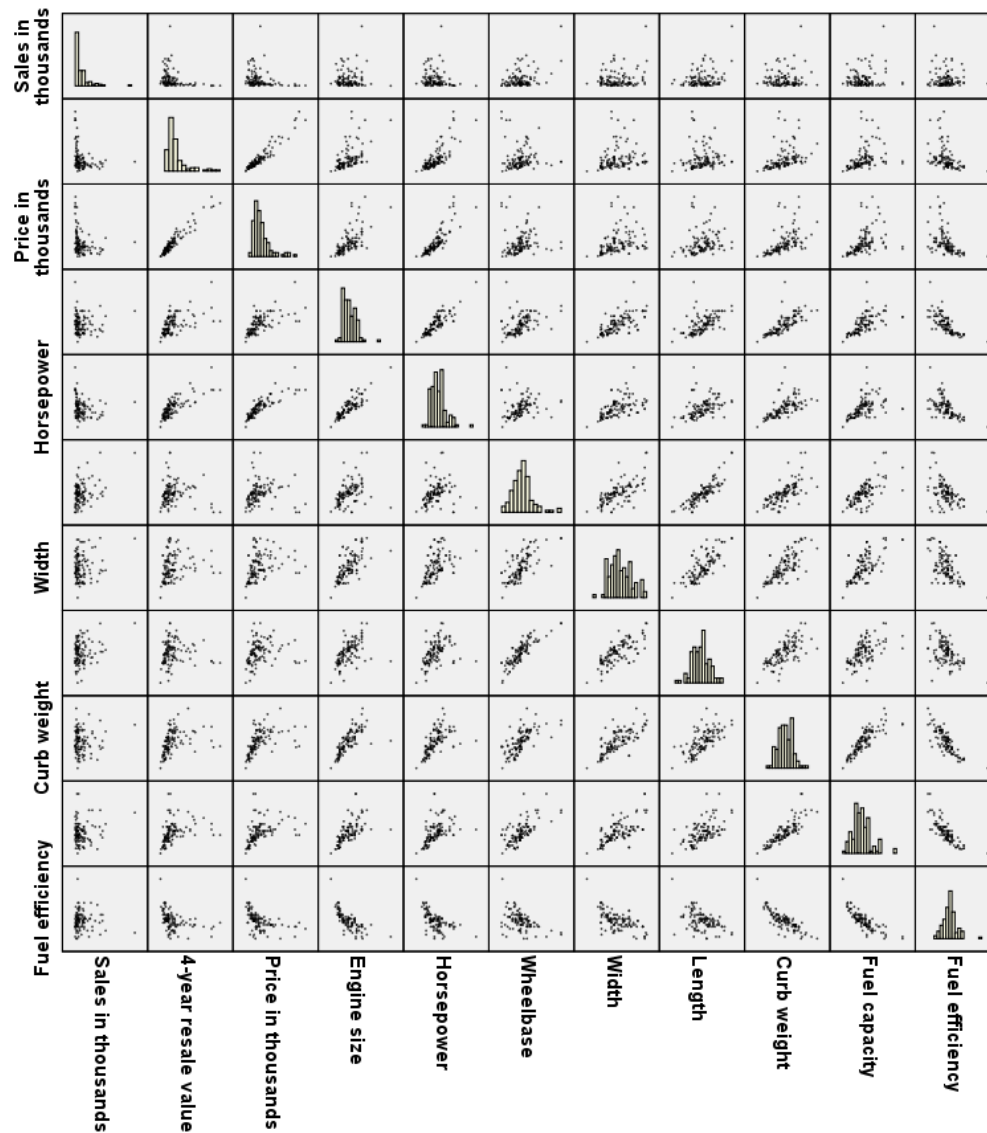
Prednosti:

- Omogoča predstavitev velikega števila enot
- Enostavno ocenimo podobnost med enotami
- Z barvami in oblikami možnost dodatnih nominalnih spremenljivk
- Z velikostjo točk, “dodatki”, ..., tudi možnost dodatnih intervalnih spremenljivk
- Lahko razumljiv prikaz (pri 3D lažje z interakcijo)
- Jasno vidni odnosi med spremenljivkami

Slabosti:

- V osnovi se lahko prikažejo samo 2 oz. 3 spremenljivke, z dodatki kakšno več
- Brez “popravkov” niso primerni za diskretne spremenljivke

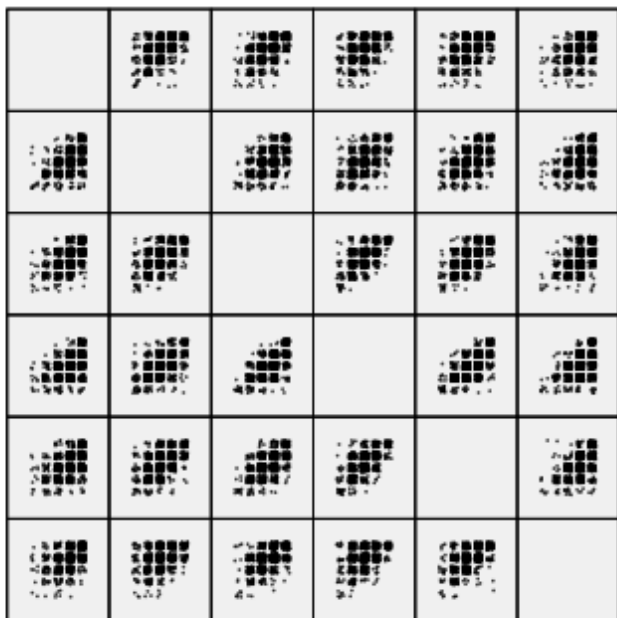
Večkratni/matrični razsevni grafikon



Večkratni/matrični razsevni grafikon

POSAMEZNIKOVI
ZIVLJENJSKI CILJI
BI MORAL

LJUDJE BI
MORALI POSV
DRUZINI.



LJUDJE BI SE MORALI POSVETITI DRUZINI."

ZIVLJENJE JE VREDNO ZIVETI, KADAR SE

DRUZINA BI MORALA BITI V ZIVLJENJU PO.

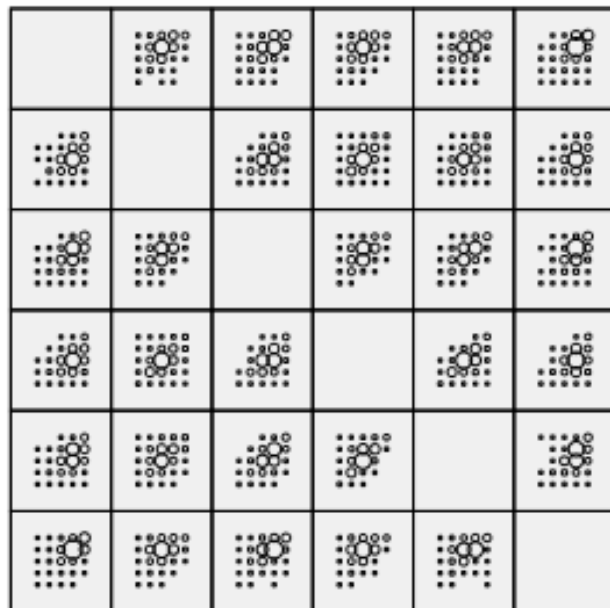
POSAMEZNIKOVI ZIVLJENJSKI CILJI BI MORAL

NAJPOMEMBNEJSE STVARI V ZIVLJENJU SE

DRUZINA BI MORALA BITI POMEMBEN DEL PO.

POSAMEZNIKOVI
ZIVLJENJSKI
CILJI BI MORAL

LJUDJE BI
MORALI
POSVETI
DRUZINI.



LJUDJE BI SE MORALI POSVETITI DRUZINI."

ZIVLJENJE JE VREDNO ZIVETI, KADAR SE

DRUZINA BI MORALA BITI V ZIVLJENJU PO.

POSAMEZNIKOVI ZIVLJENJSKI CILJI BI MORAL

NAJPOMEMBNEJSE STVARI V ZIVLJENJU SE

DRUZINA BI MORALA BITI POMEMBEN DEL PO.

Scale
○ 600
○ 500
○ 400
○ 300
○ 200
○ 100
• 0

Večkratni/matrični razsewni grafikon

Več 2D razsewnih grafikonov, zloženih v matriko

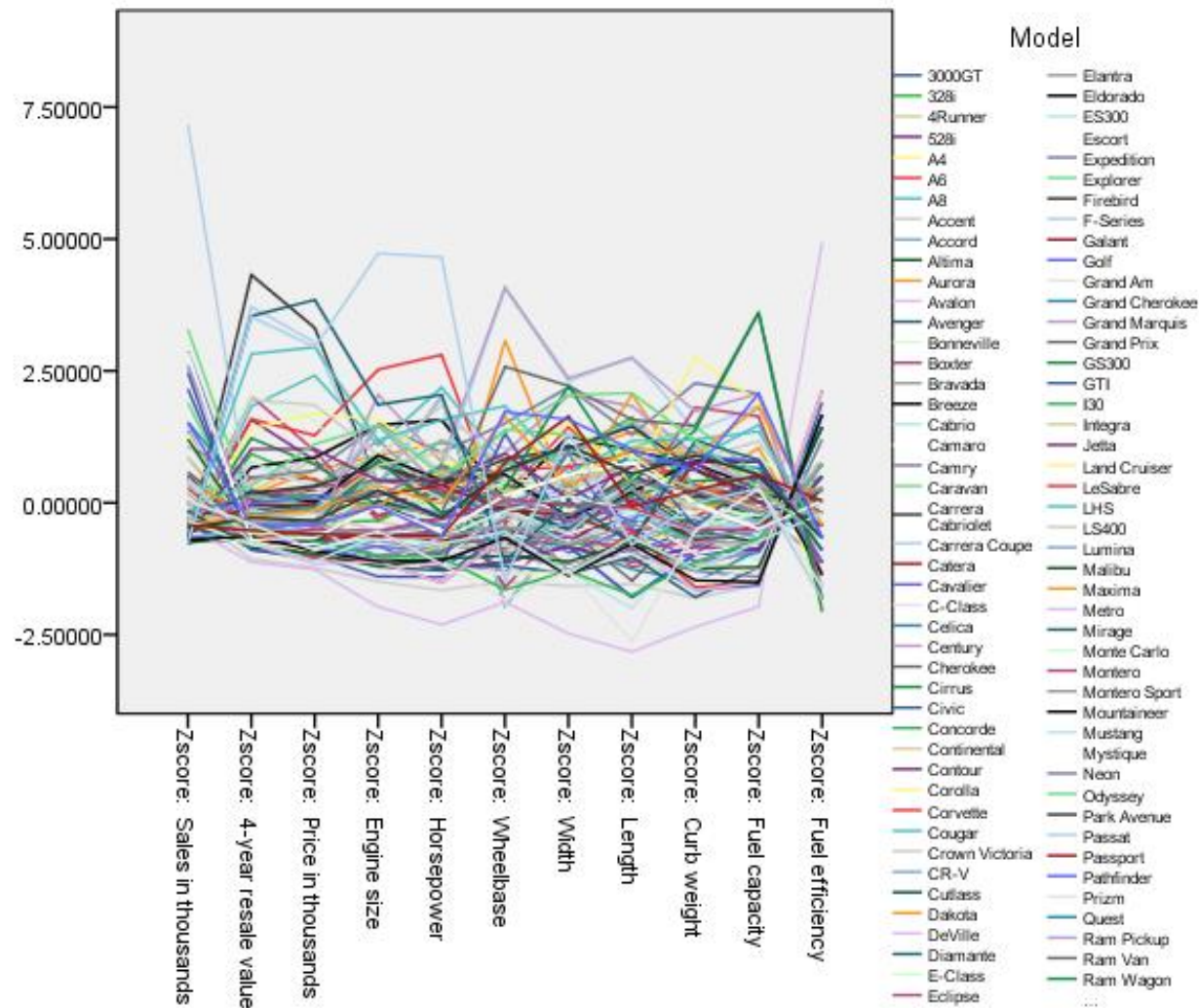
Prednosti:

- Možnost prikaza velikega števila enot in večjega števila spremenljivk (3-10+)
- Načeloma možnost prikaza dodatnih nominalnih spremenljivk (preko barv, oblik)

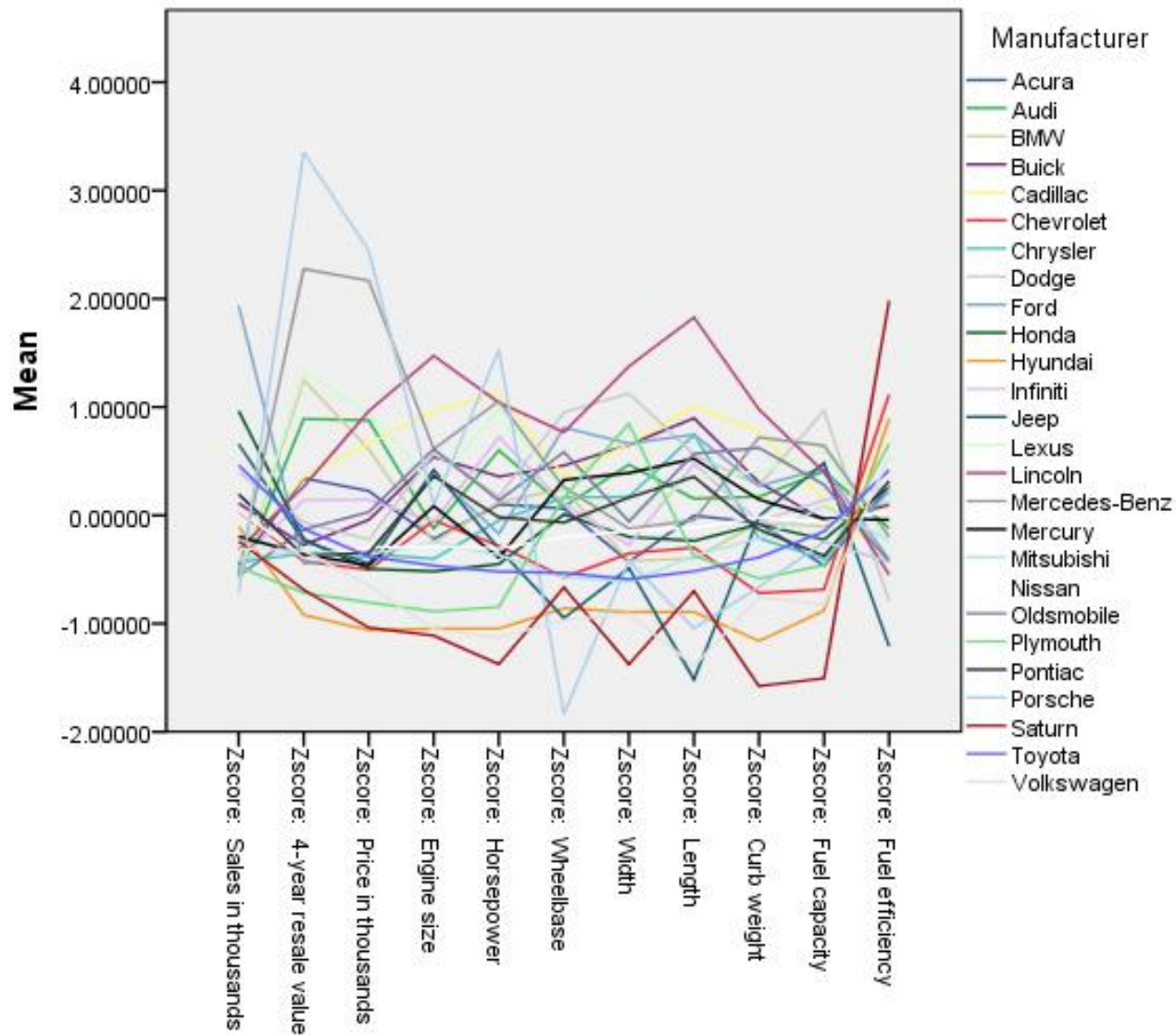
Slabosti:

- Samo bivariata povezanost
- Nepregledno pri velikem številu spremenljivk

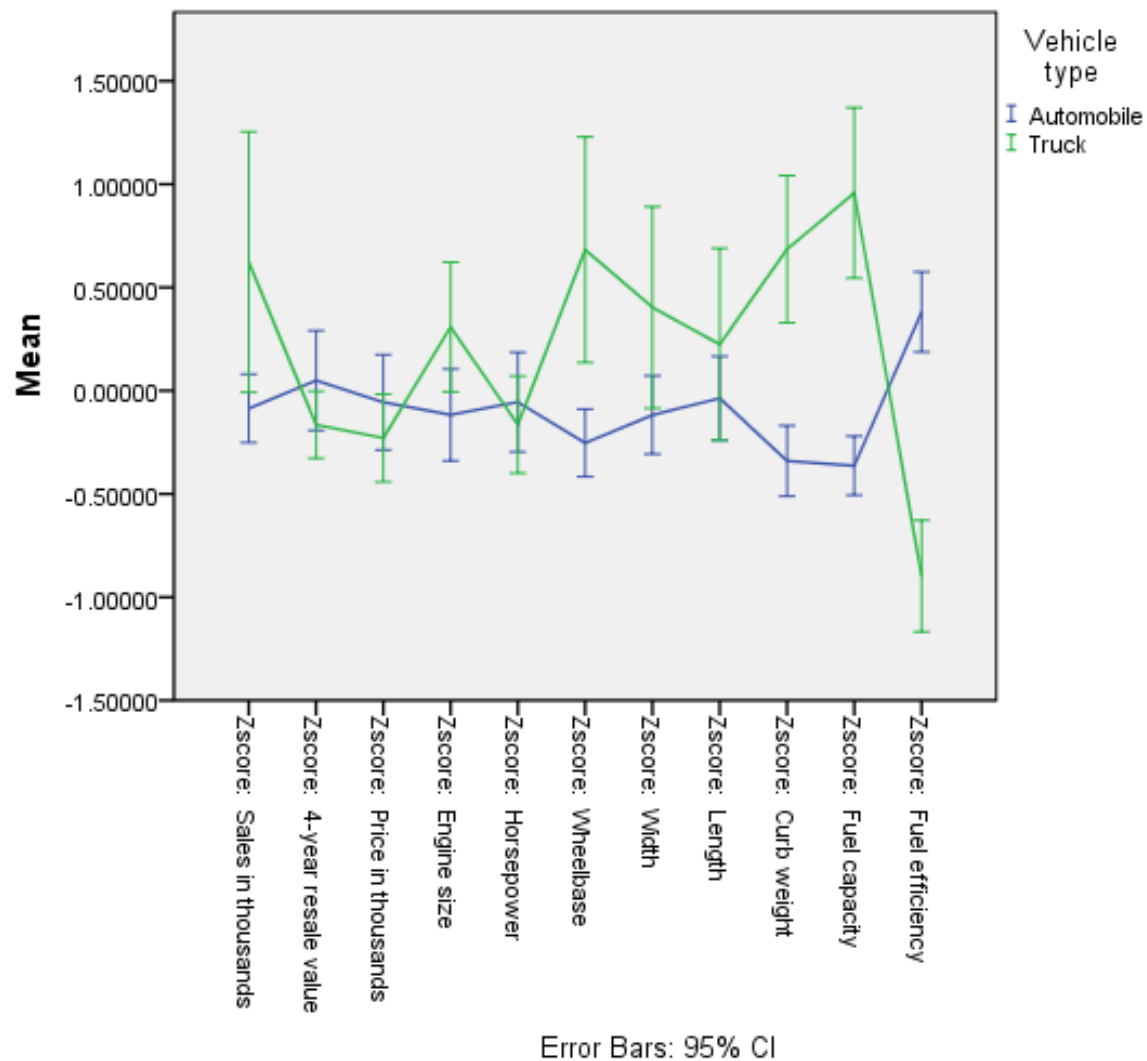
Linijski grafikoni



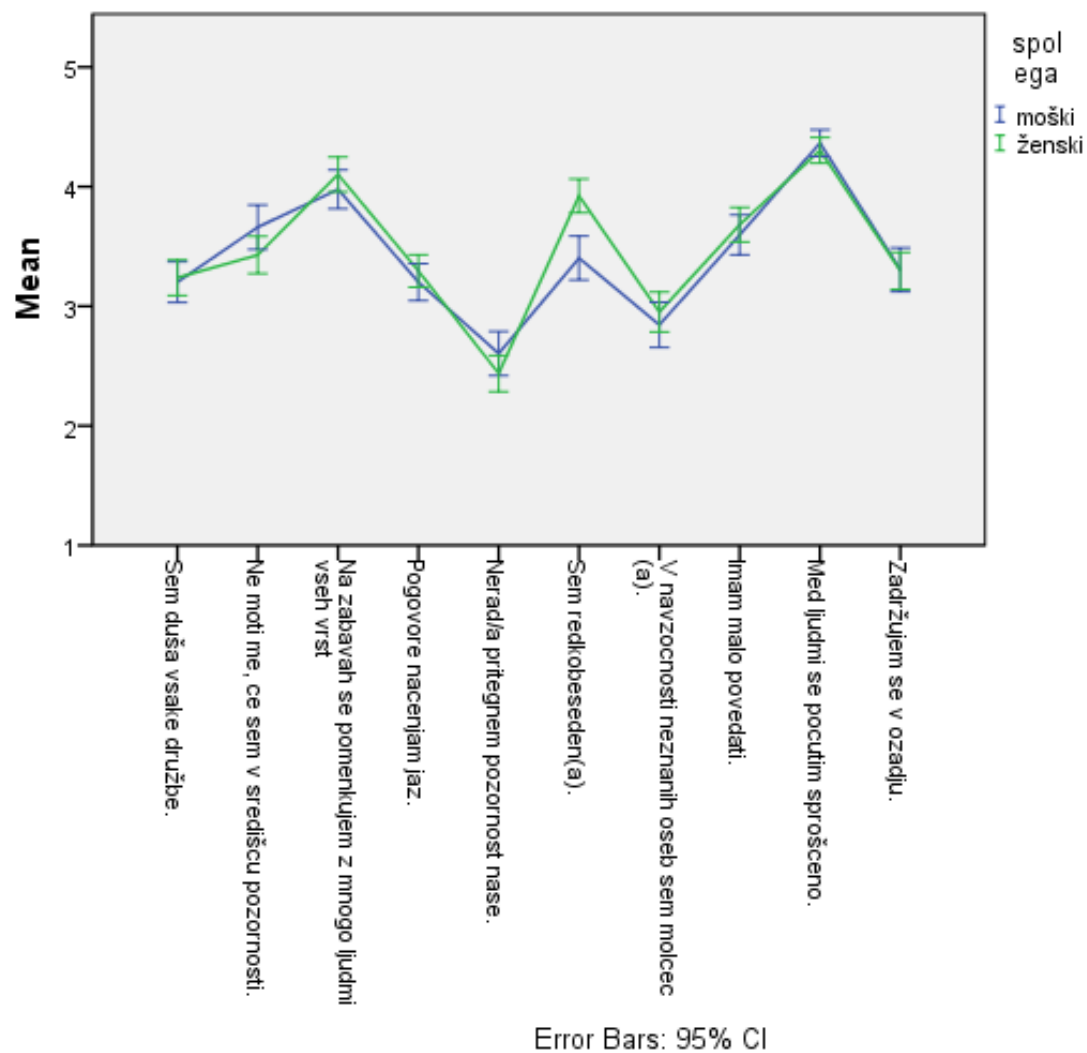
Linijski grafikoni



Linijski grafikoni



Linijski grafikoni



Linijski grafikoni

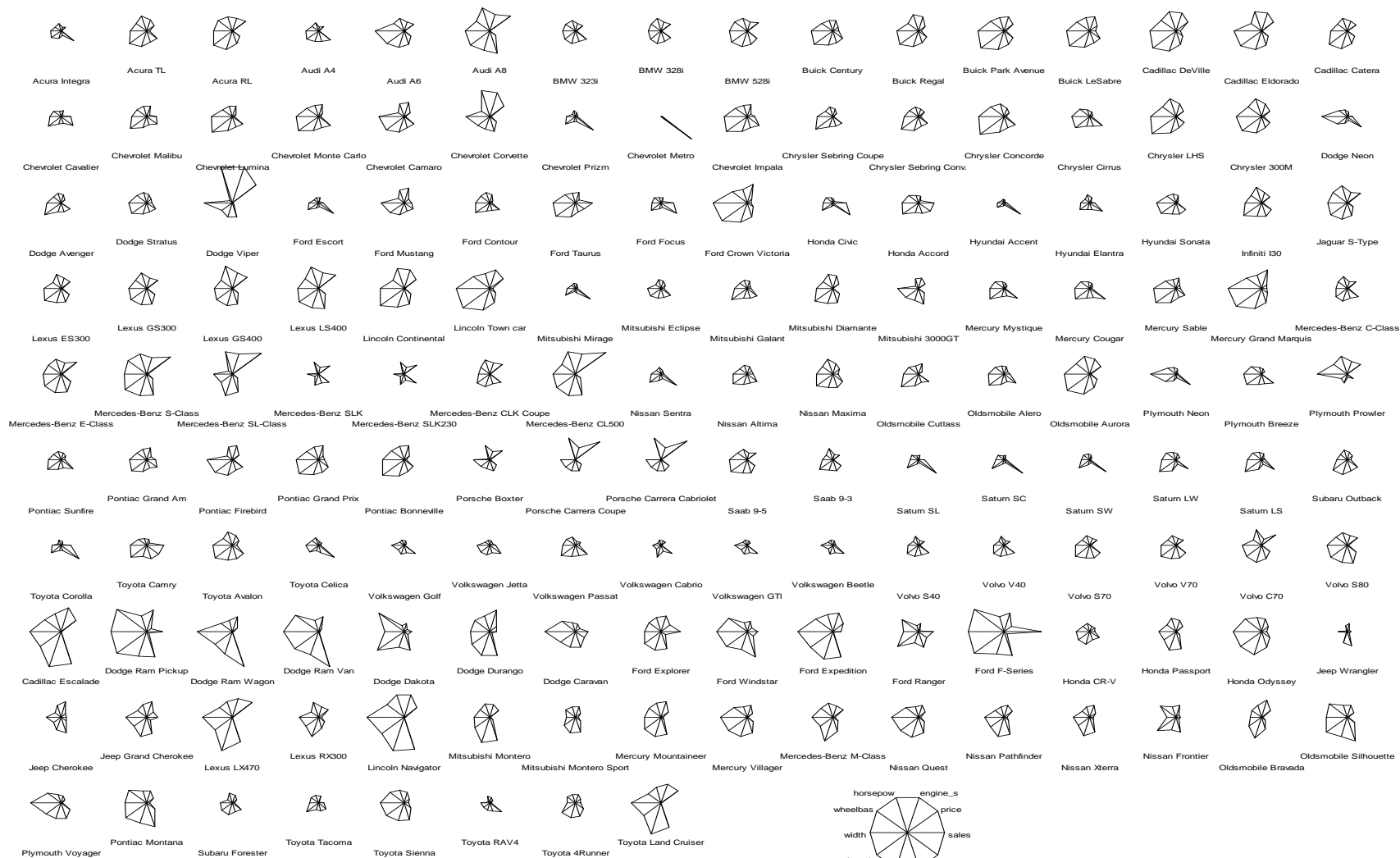
Prednosti:

- Omogoča prikaz velikega števila spremenljivk
- Enostavna primerjava vrednosti spremenljivk
- Če prikazujemo srednje vrednosti skupin lahko dodamo tudi mere variabilnosti

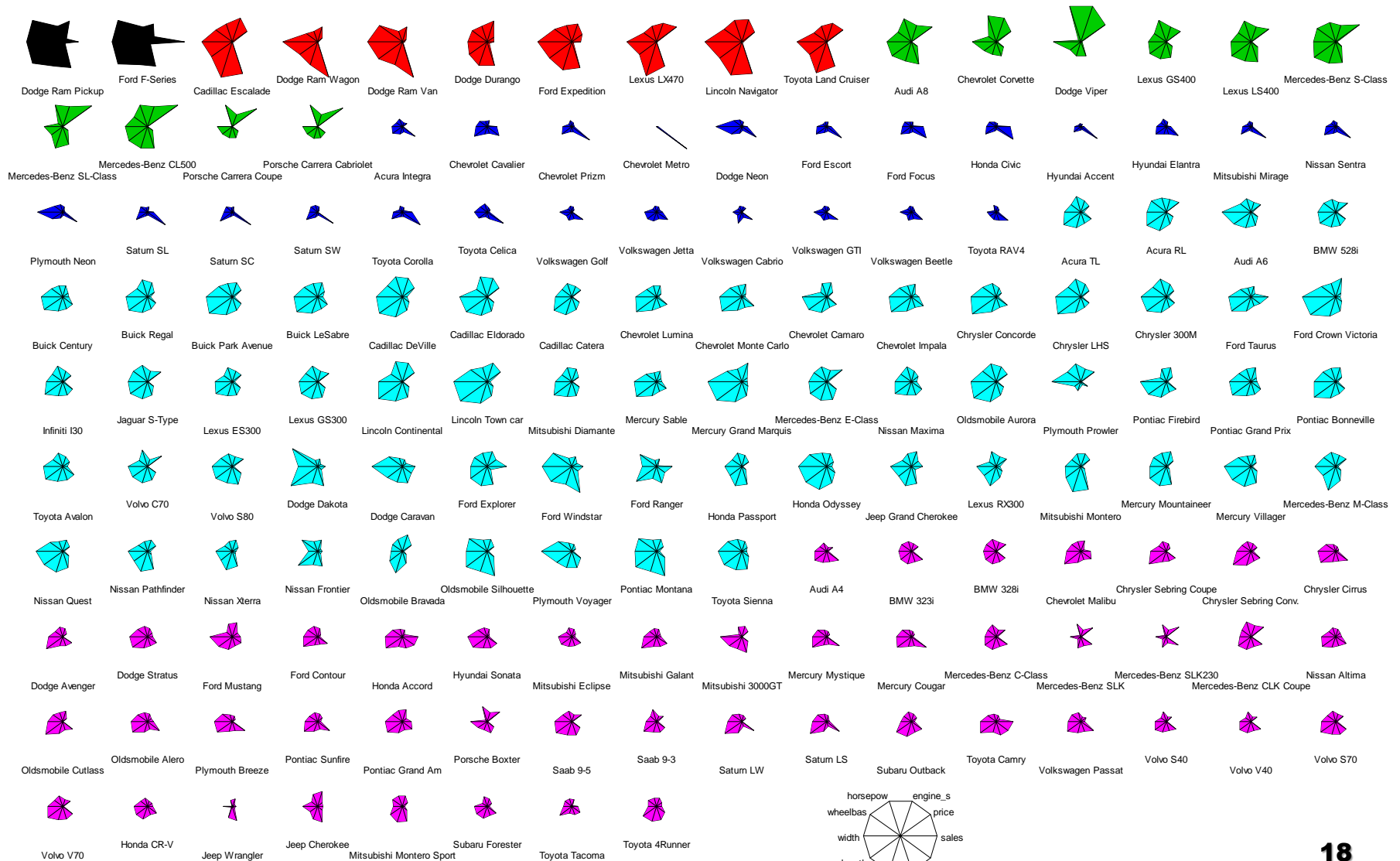
Slabosti:

- Prikaže se lahko le omejeno število enot oz. skupin enot
- Spremenljivke morajo imeti podobne merske lestvice (kar lahko dosežemo tudi s trasformacijo)

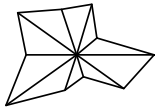
Zvezde



Zvezde



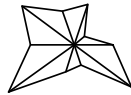
Zvezde



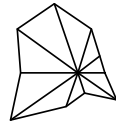
Buick LeSabre



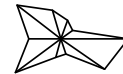
Cadillac Catera



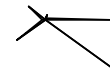
Chevrolet Monte Carlo



Chrysler 300M



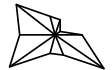
Dodge Stratus



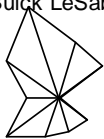
Ford Escort



Hyundai Elantra



Hyundai Sonata



Lexus GS400



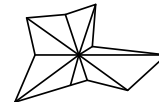
Mercedes-Benz SLK



Mercedes-Benz SLK230



Nissan Altima



Pontiac Grand Prix



Porsche Boxster



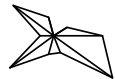
Porsche Carrera Coupe



Subaru Outback



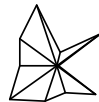
Toyota Celica



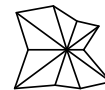
Volkswagen Passat



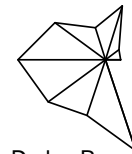
Volkswagen GTI



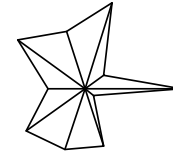
Volvo C70



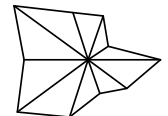
Volvo S80



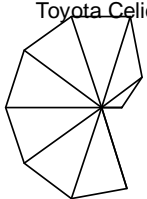
Dodge Ram Wagon



Dodge Durango



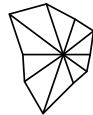
Honda Odyssey



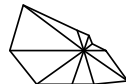
Lincoln Navigator



Mercury Villager



Mercedes-Benz M-Class



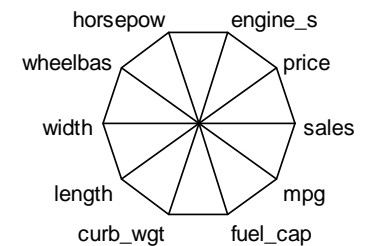
Plymouth Voyager



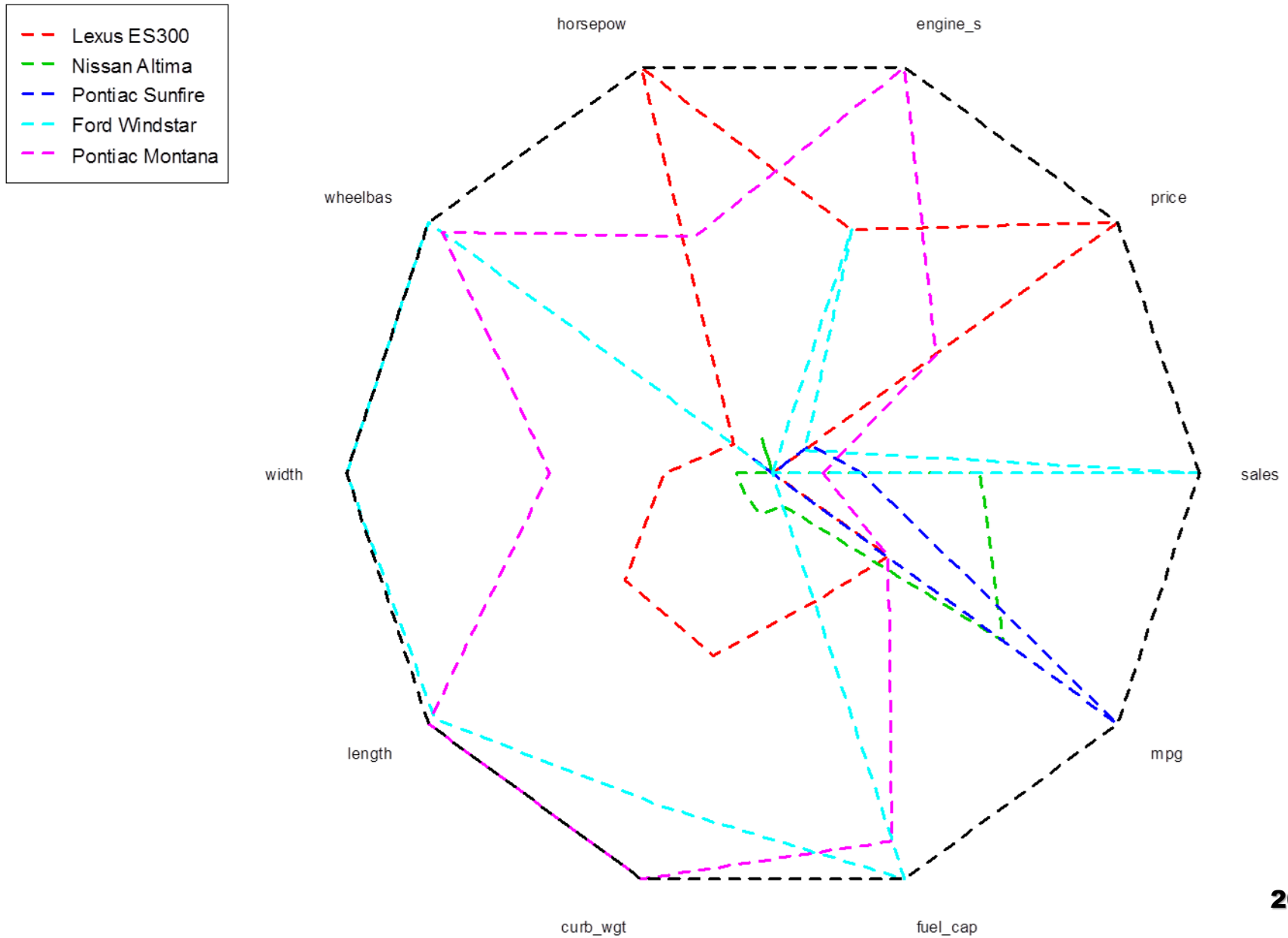
Toyota Tacoma



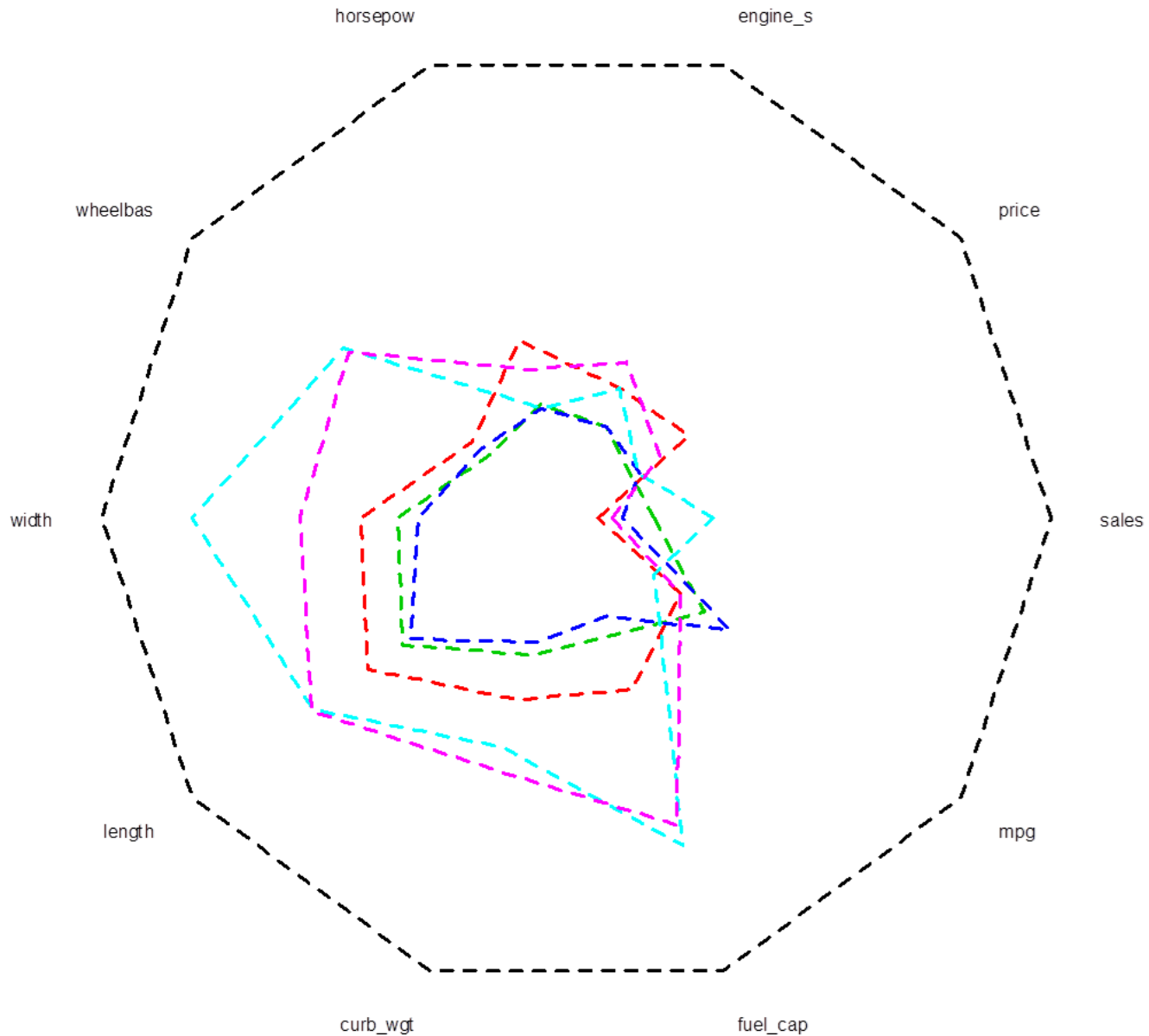
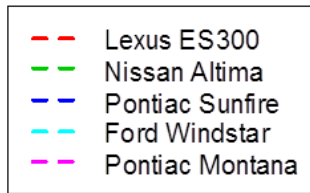
Toyota RAV4



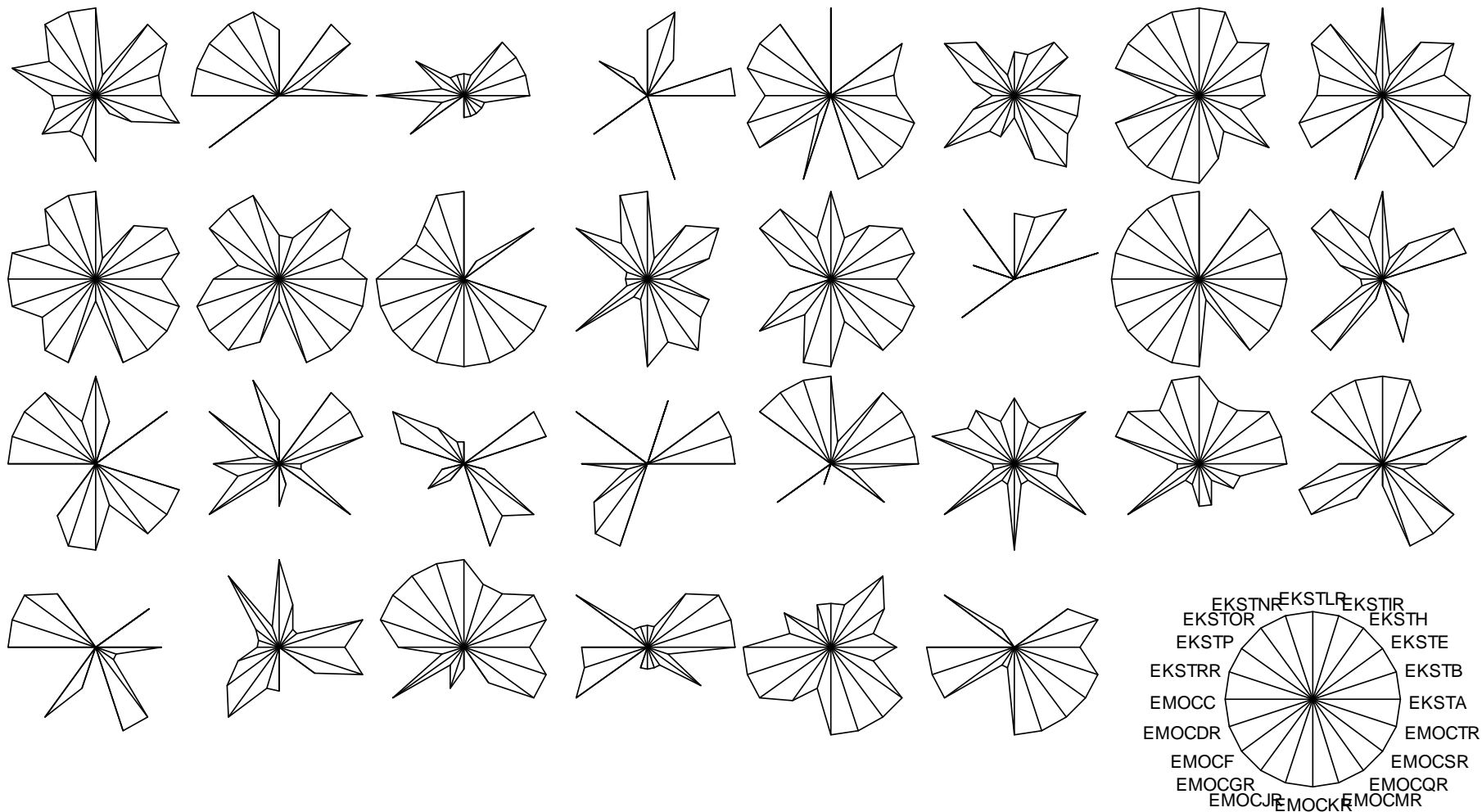
Zvezde



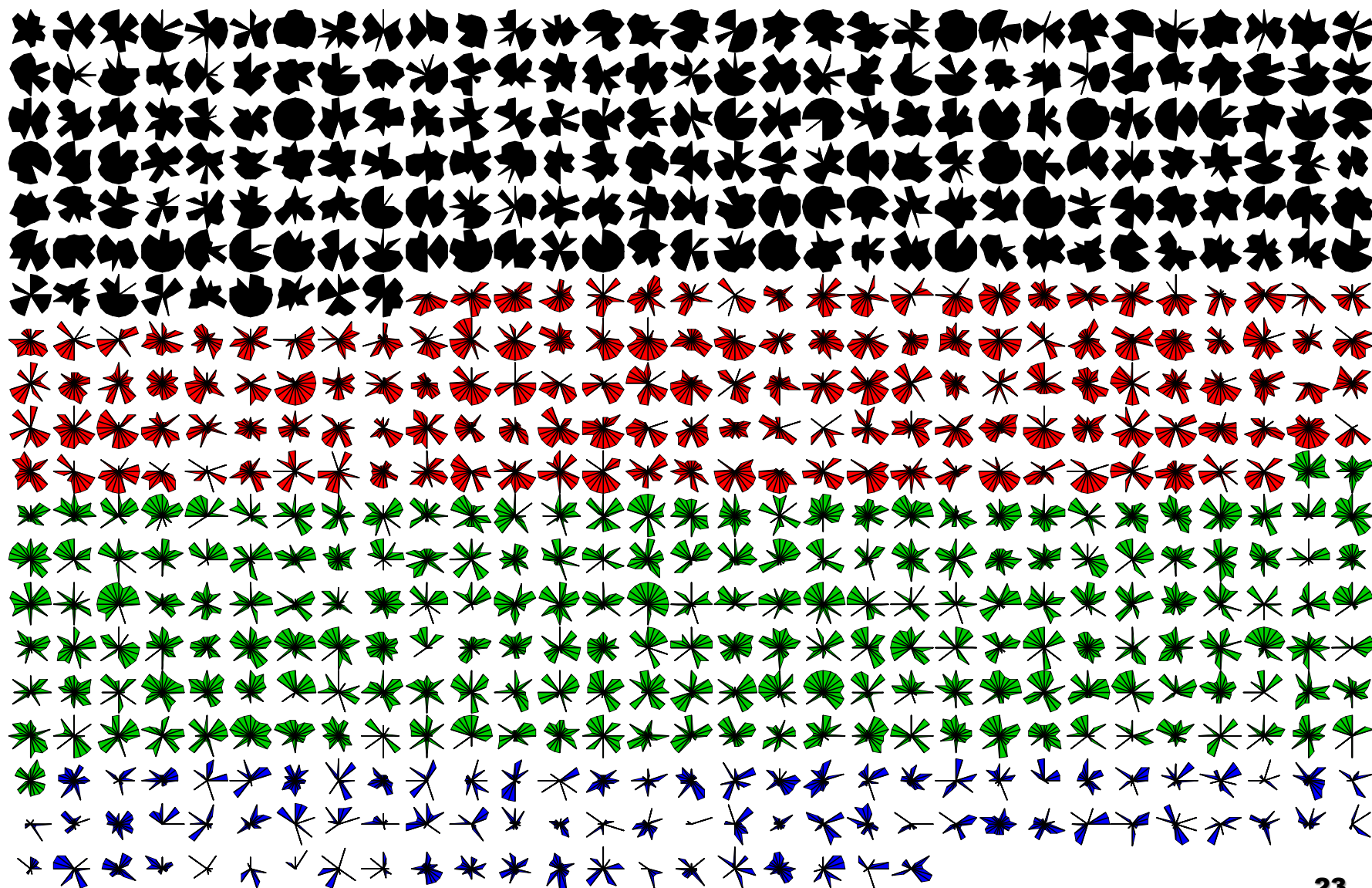
Zvezde



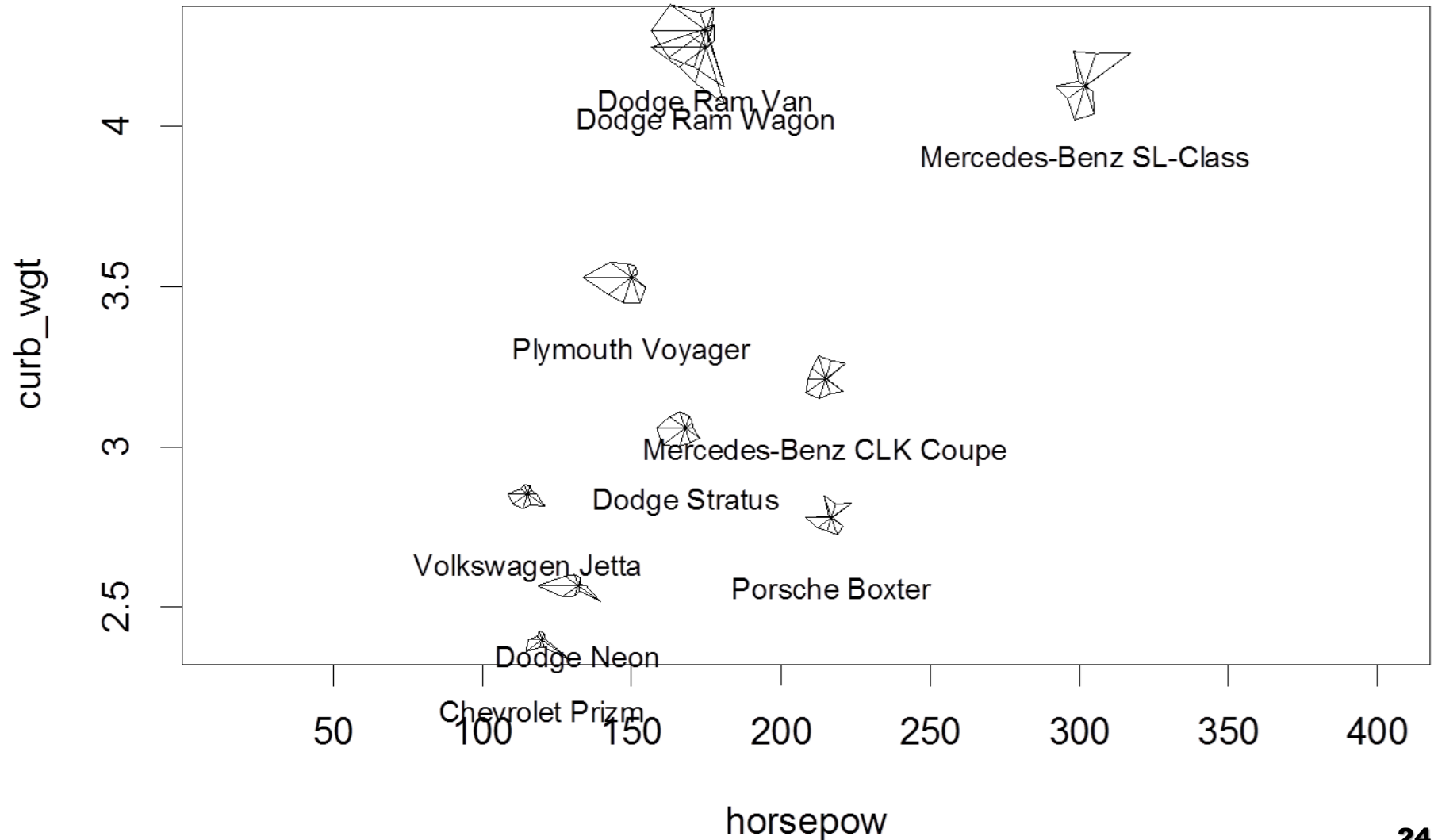
Zvezde



Zvezde



Zvezde



Zvezde

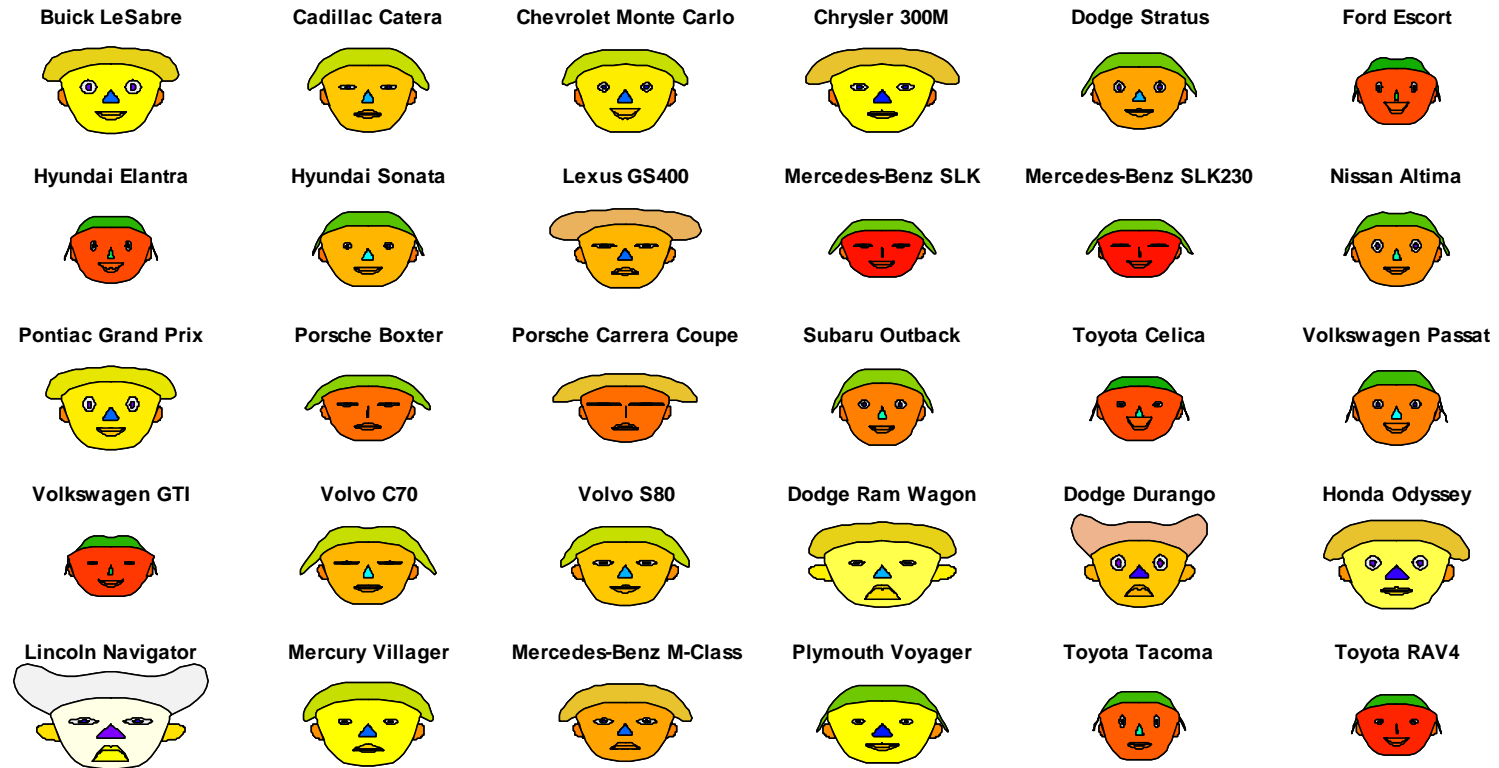
Prednosti:

- Omogoča prikaz več spremenljivk (cca. 10)
- Enostavna primerjava posameznih enot

Slabosti:

- Težko oceniti vrednosti spremenljivk
- Spremenljivke morajo imeti podobne merske lestvice (kar običajno dosežemo s transformacijo)
- Težko dobiti “sliko” celotnega podatkovja
- Niso primerne za zelo veliko enot ali spremenljivk

Obrazi (Chernoff)



"height of face
"width of face
"smiling
"height of eyes
"width of eyes

" "length"
" "width"
" "mpg"
" "sales"
" "price"

"height of hair
"width of hair
"style of hair
"width of nose
"width of ear

" "curb_wgt"
" "horsepow"
" "engine_s"
" "wheelbas"
" "fuel_cap"



"height of face

"width of face

"smiling

"height of eyes

"width of eyes

" "length"

" "width"

" "mpg"

" "sales"

" "price"

"height of hair

"width of hair

"style of hair

"width of nose

"width of ear

" "curb_wgt"

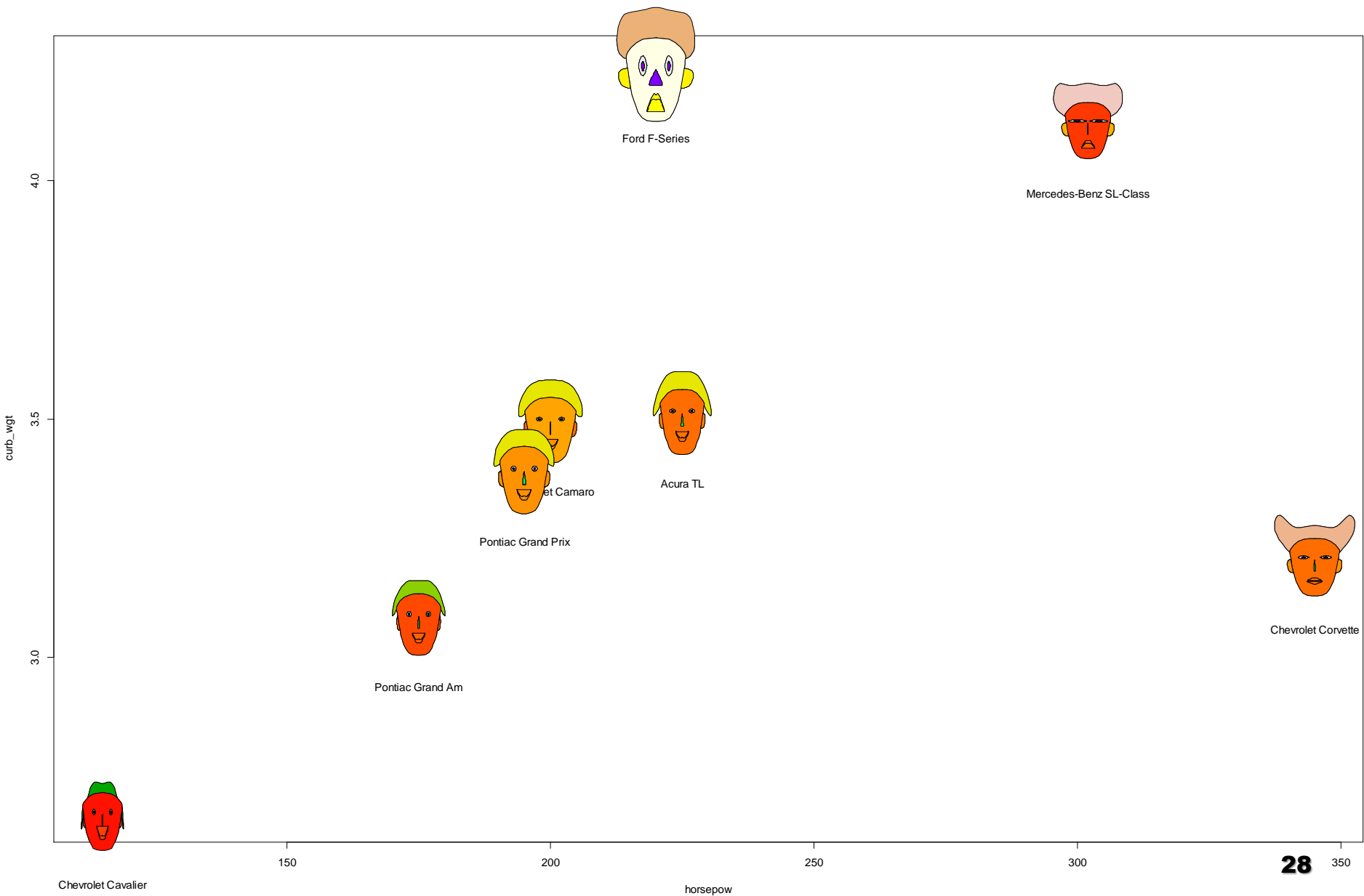
" "horsepow"

" "engine_s"

" "wheelbas"

" "fuel_cap"

Obrazi



Obrazi (Chernoff)

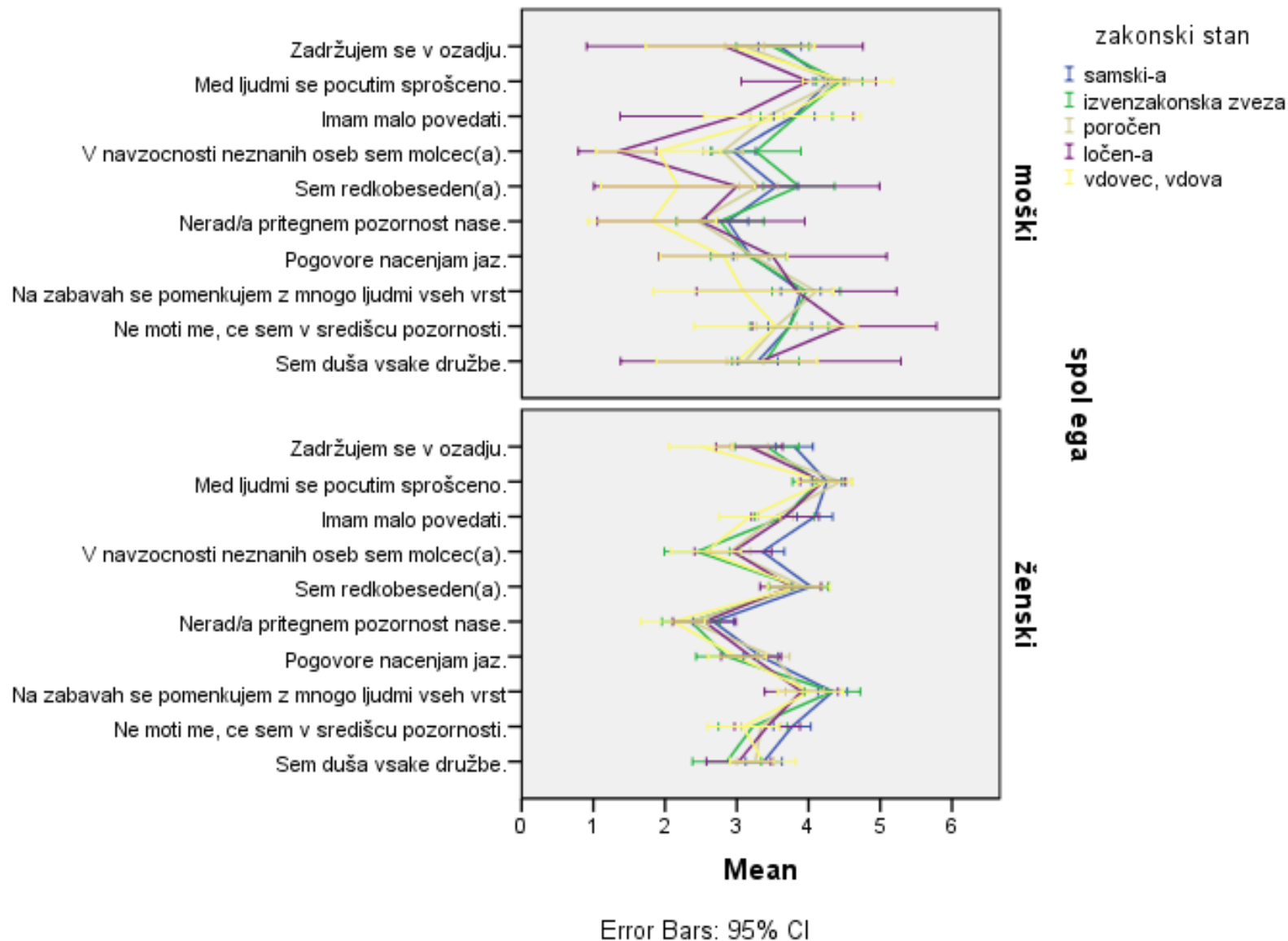
Prednosti:

- Omogoča prikaz velikega števila spremenljivk (20+, nesimterični obrazi x2)
- Enostavna primerjava posameznih enot

Slabosti:

- Subjektivna določitev pomembnosti posamezne spremenljivke
- Težko oceniti vrednosti spremenljivk
- Težko dobiti “sliko” celotnega podatkovja
- Niso primerne za zelo veliko enot

Panelni grafikoni

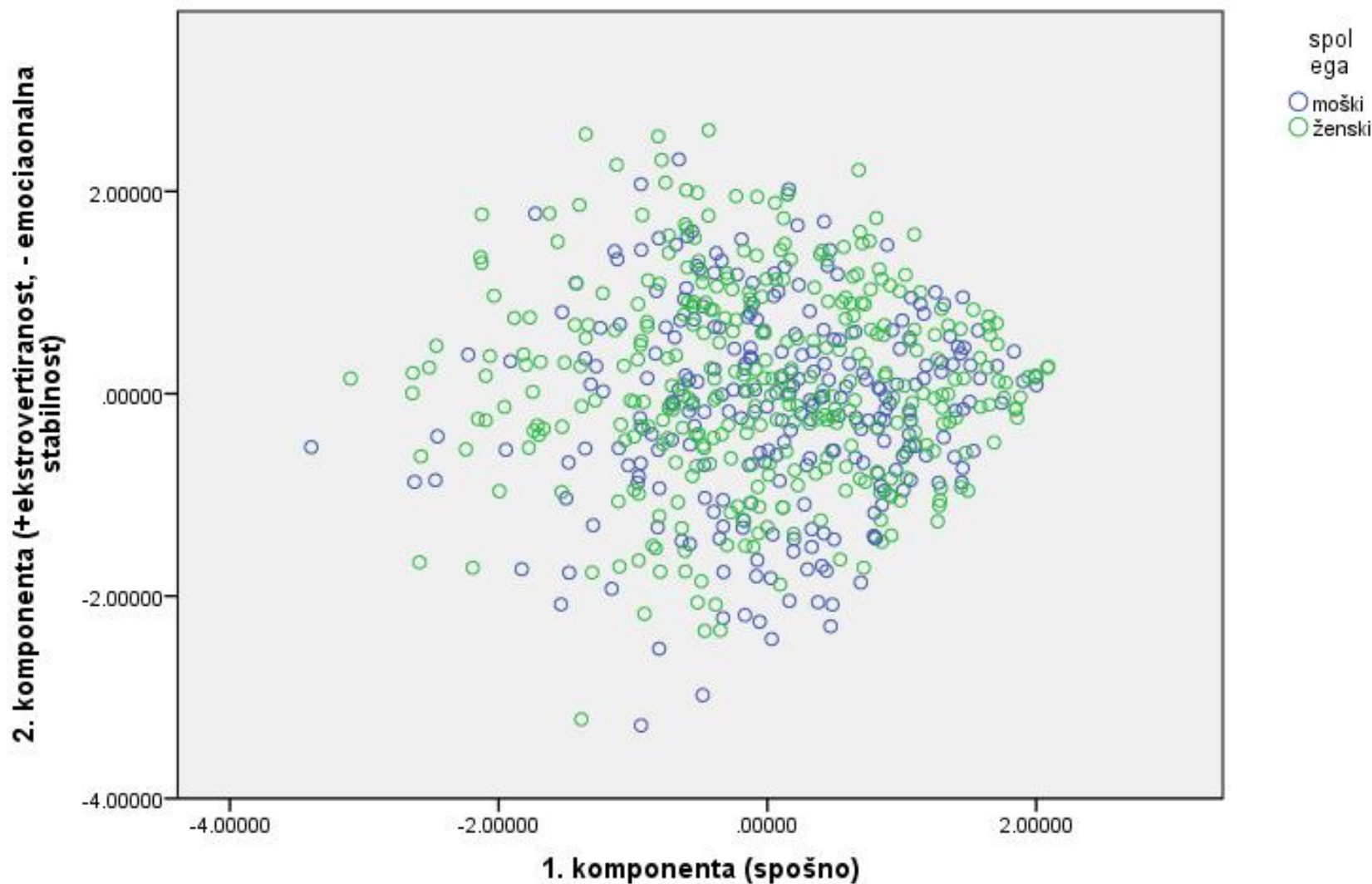


Panelni grafikoni

- Ideja je, da vzamemo enega izmed „navadnih“ grafikonov in jih narišemo ločeno glede na vrednost neke druge spremenljivke.
- To omogoča prikaz dodatne nove spremenljivke, vendar pa tudi naredi graf manj pregleden.

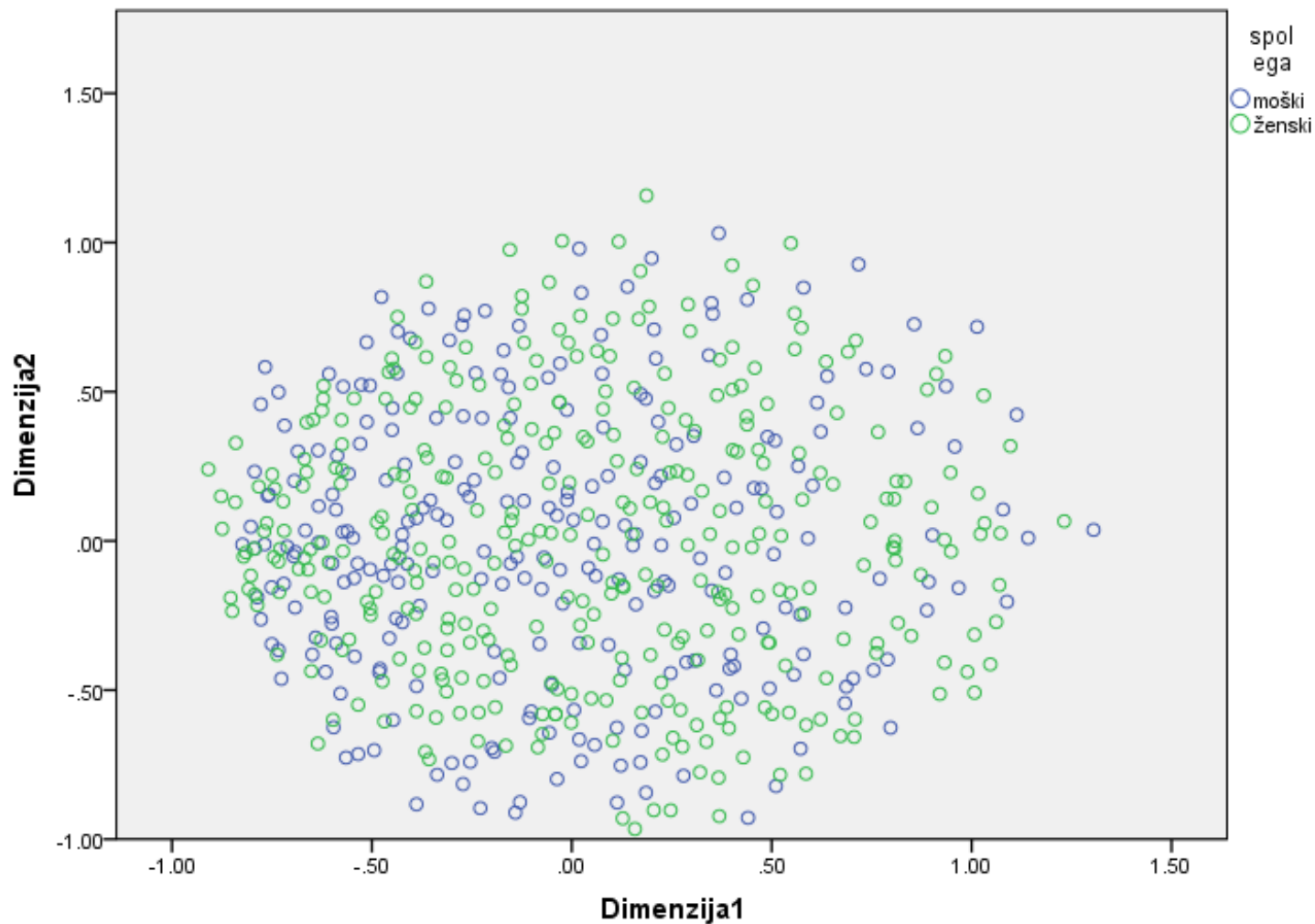
Redukcija prostora

Metoda glavnih komponent



Redukcija prostora

Večrazmernostno lestvičenje



Redukcija prostora

- Pogosto problem velikega števila spremenljivk poskušamo rešiti tako, da večje število spremenljivk reduciramo na 2 ali 3 spremenljivke.
- To poskušamo storiti tako, da izgubimo čim manj informacije.
- Nekaj metod:
 - Sestavljene spremenljivke (likartove lestvice)
 - Metoda glavnih komponent
 - Faktorska analiza
 - Večrazmernostno lestvičenje

Redukcija prostora

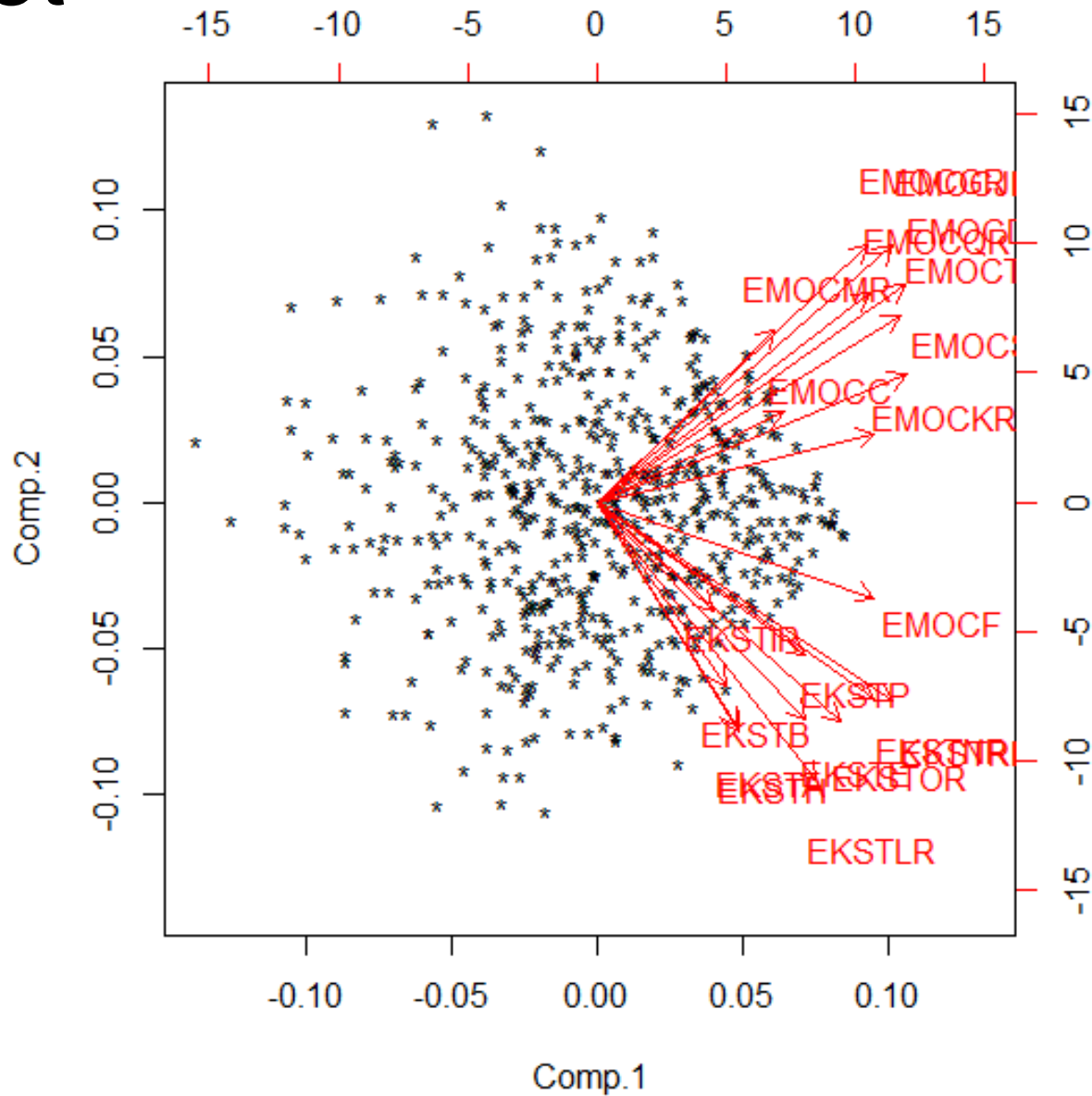
Prednosti:

- Omogoča prikaz velikega števila enot ob upoštevanju velikega števila spremenljivk
- Omogoča pridobiti „globalno“ sliko multivariatnih podatkov

Slabosti:

- Vrednosti posameznih spremenljivk niso vidne
- Dimenzije prostora (osi) včasih ni mogoče enostavno interpretirati
- Nekaj informacije (odvisno od povezanosti spremenljivk) se izgubi → če spremenljivke niso močno povezane (in jih je veliko), je izguba informacije velika.

„Biplot“



„Biplot“

Pravzaprav „razširitev“ klasičnega razsevnega grafikona na podlagi redukcije prostora.

Dodatna prednost:

- Prikaže „položaj“ oz. „smer“ spremenljivk v „generiranem prostoru“

Interaktivni grafi

- Interaktivni grafi:

- <http://webales.fdv.uni-lj.si/jsGraf.html>

- Prilagodljivi grafi (shiny):

- <http://webales.fdv.uni-lj.si/shiny/ShinyGraf/>

- <https://ales.shinyapps.io/ShinyGraf/>