

TSM Előadás

Balázs Frigyes, Gedeon Anna, Nagy Balázs

Adatok

- Autós adatbázis

- Forrás: **Használtautó.hu**
- Letöltés ideje: 2023.04.03.
- Sok változó (22), ebből 13-at használtunk fel
- Eredetileg majdnem 90.000 adat, a végleges adatbázis 50 ezer megfigyelést tartalmazott



Tömeg



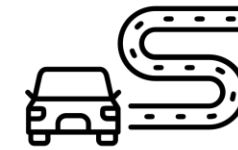
Üzemanyag
típusa



Klímaberendezés



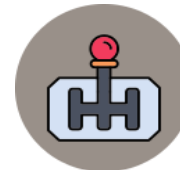
Teljesítmény



Km óra állása



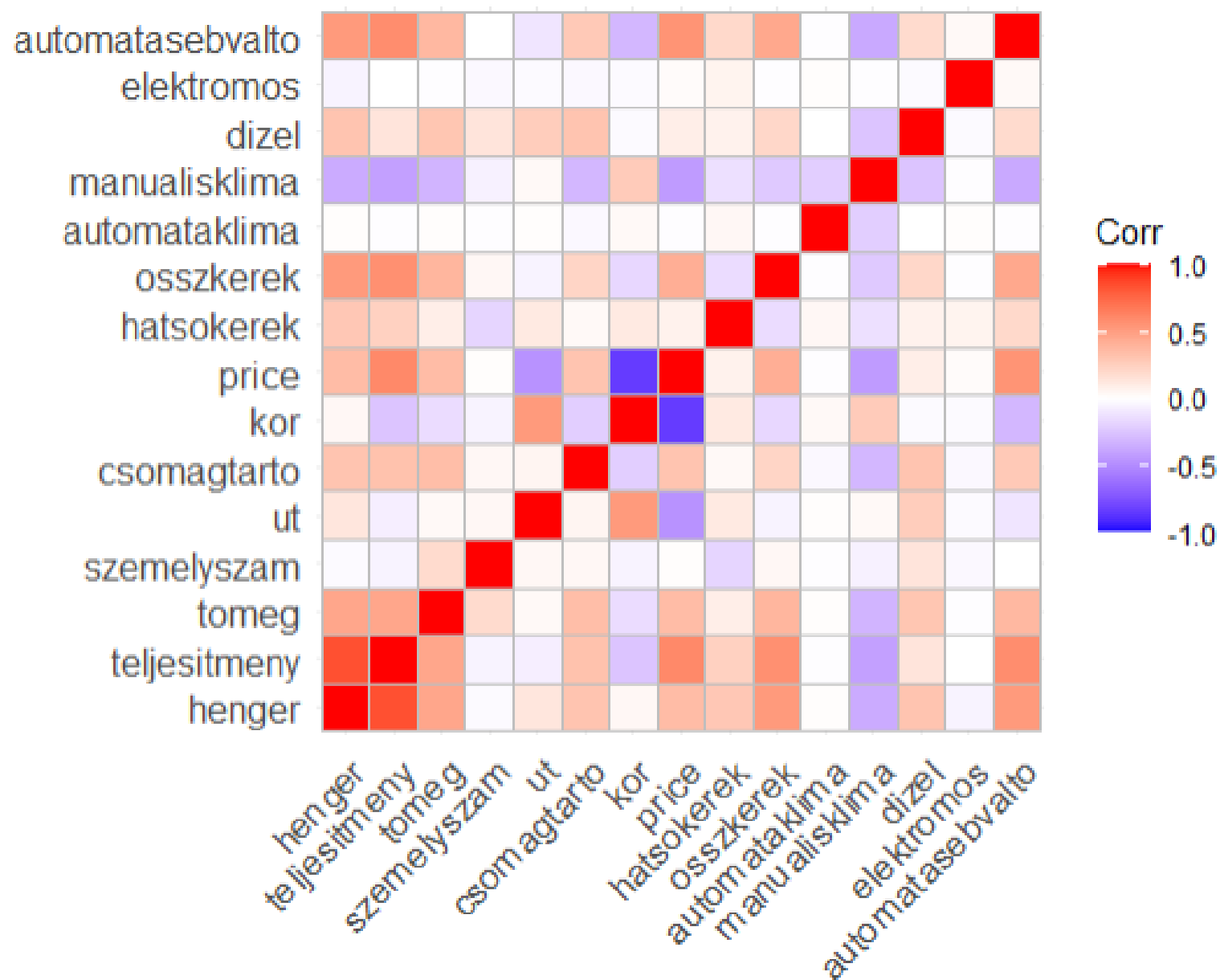
Kor



Sebességváltó
típusa



Hengerűrtartalom



A modell felépítése

- ▲ Használt megfigyelt változók száma (p): 8
- ▲ Becsült paraméterek száma (m): 16
- ▲ Szabadságfok: $df = \frac{p(p-1)}{2} - m = 12$
- ▲ Becslés módszere: robosztus Maximum Likelihood (MLF)

SEM

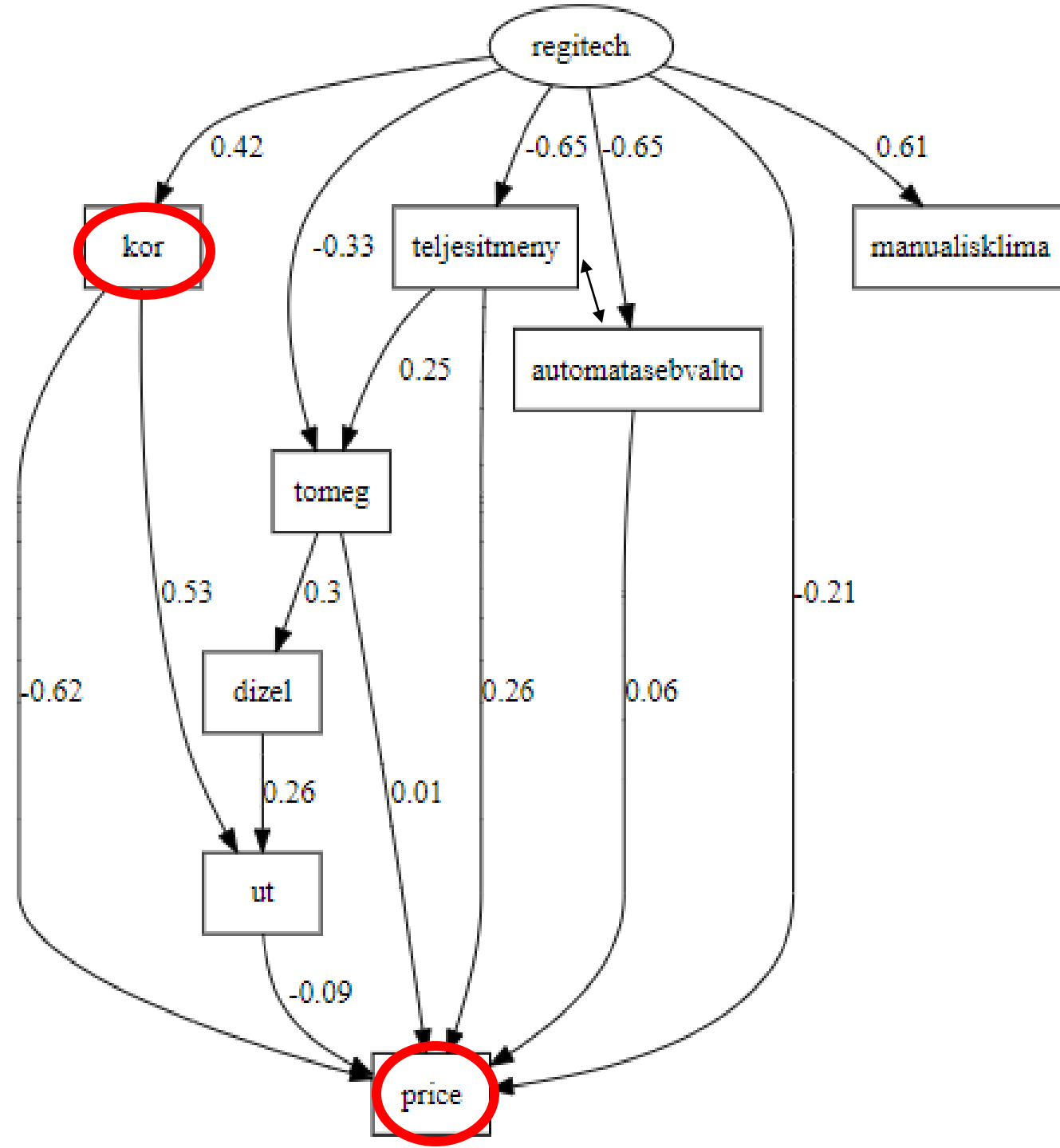
▲ A modell az ár varianciájár **89,6%**-át magyarázza

▲ A CFI és TLI alapján kiváló;
RMSEA alapján elfogadható

▲ Látenis változó

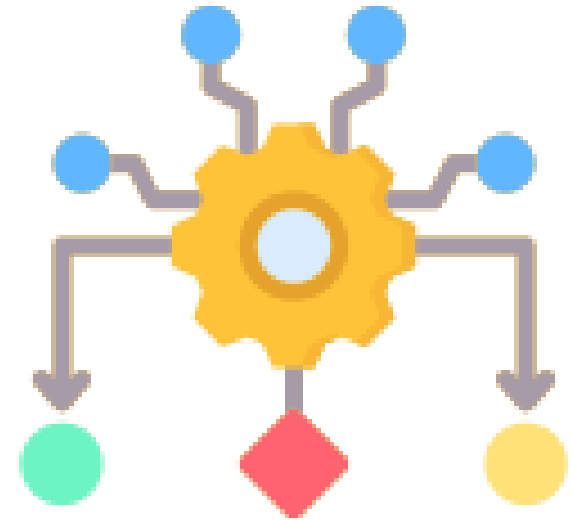
▲ Útelemzés -példa

Ceteris paribus plusz egy év az autó
koránál közvetlenül **46%**-kal, a megtett
úton keresztül még **4,5%**-kal, összesen
49%-kal növeli az autó árának várható
értékét.



Endogenitás

- ▲ Confounder változók
 - Gazdasági körülmények
 - Marketinghatások
 - Szabályozásváltozás
- ▲ Fordított okság nem merül fel

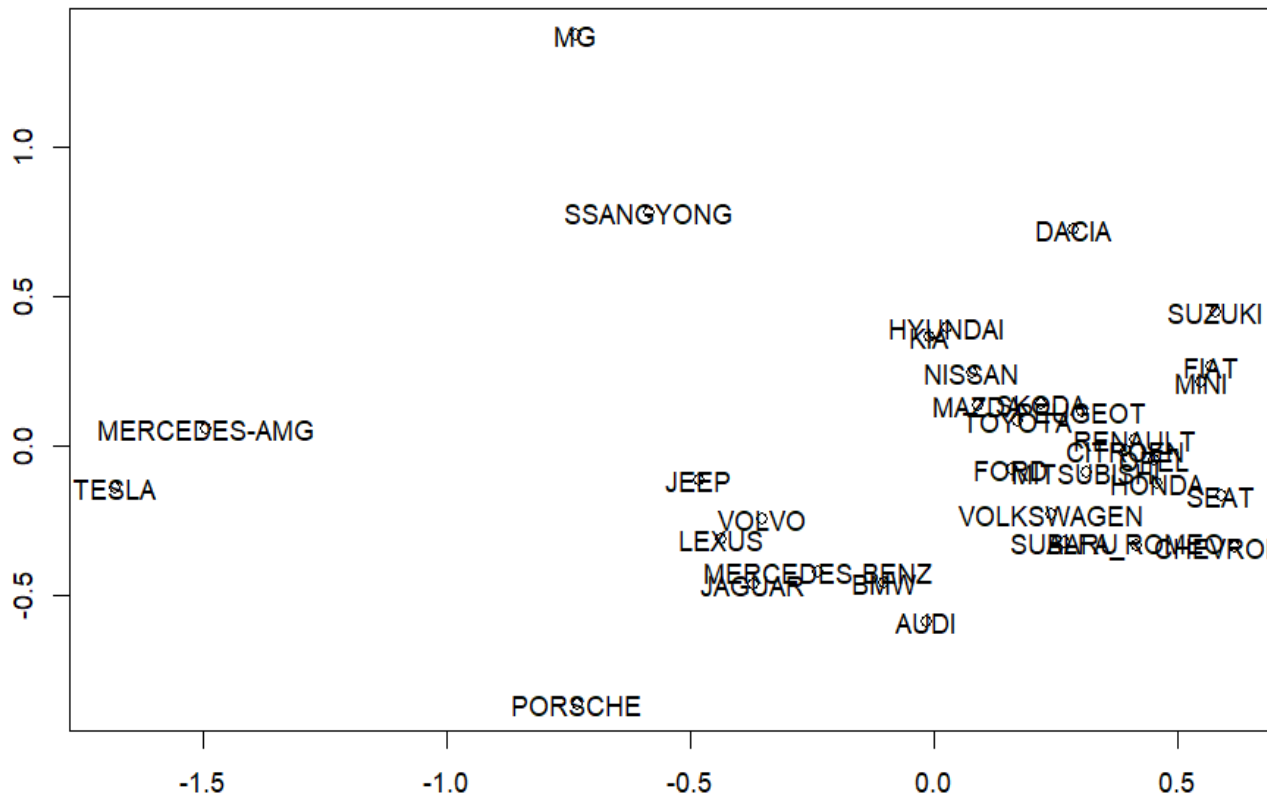


MDS

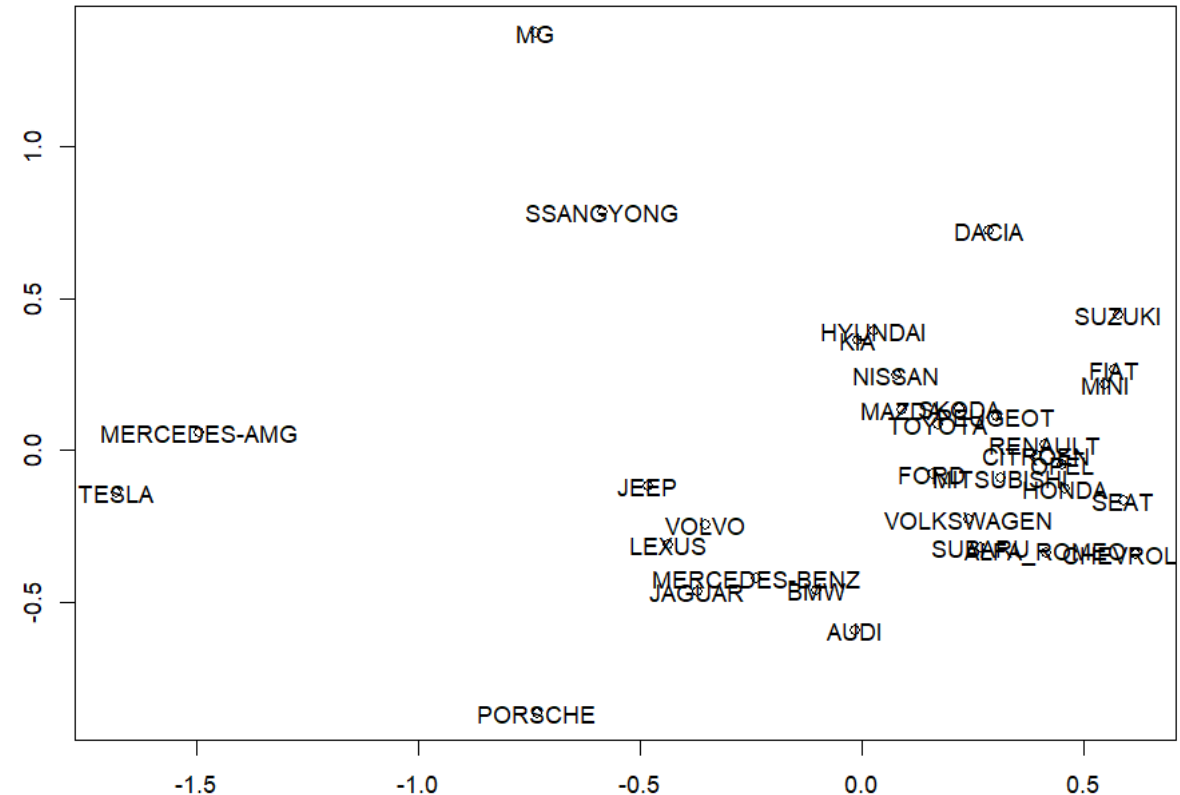
- Autós adatbázis
- Márkák „közelségének” mérése (márkánkénti átlag
- Használt numerikus (standardizált) változók: Vételár, teljesítmény, kor, Km-óraállás
- Kevésszer előforduló márkák kiszűrése
- Sok kipróbált modell, általában jó stressz értékekkel
 - Márkák száma szerint
 - Metrikusság szerint
 - Dimenzió szerint
 - Használt változók szerint

1. Modell 34 márkával

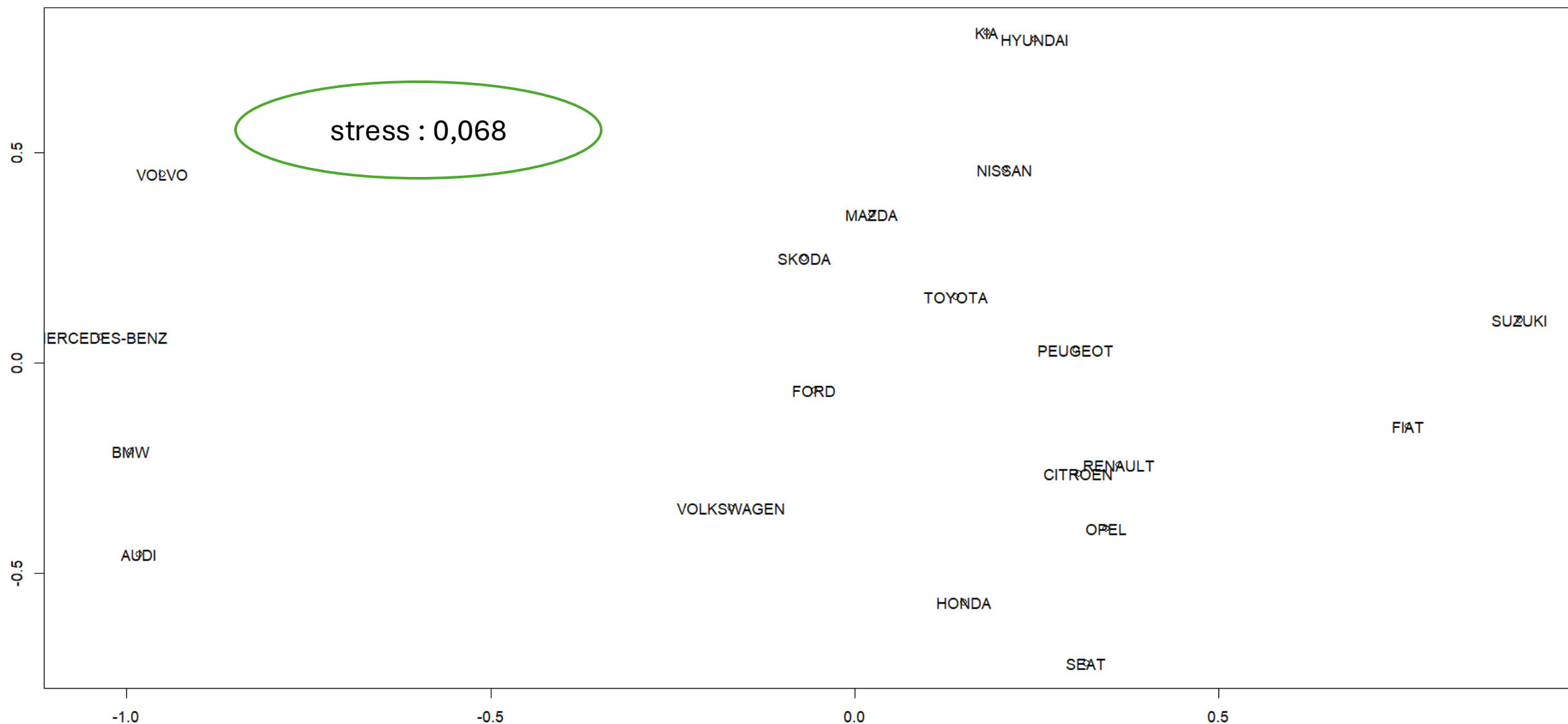
stress metrikus:
0,049



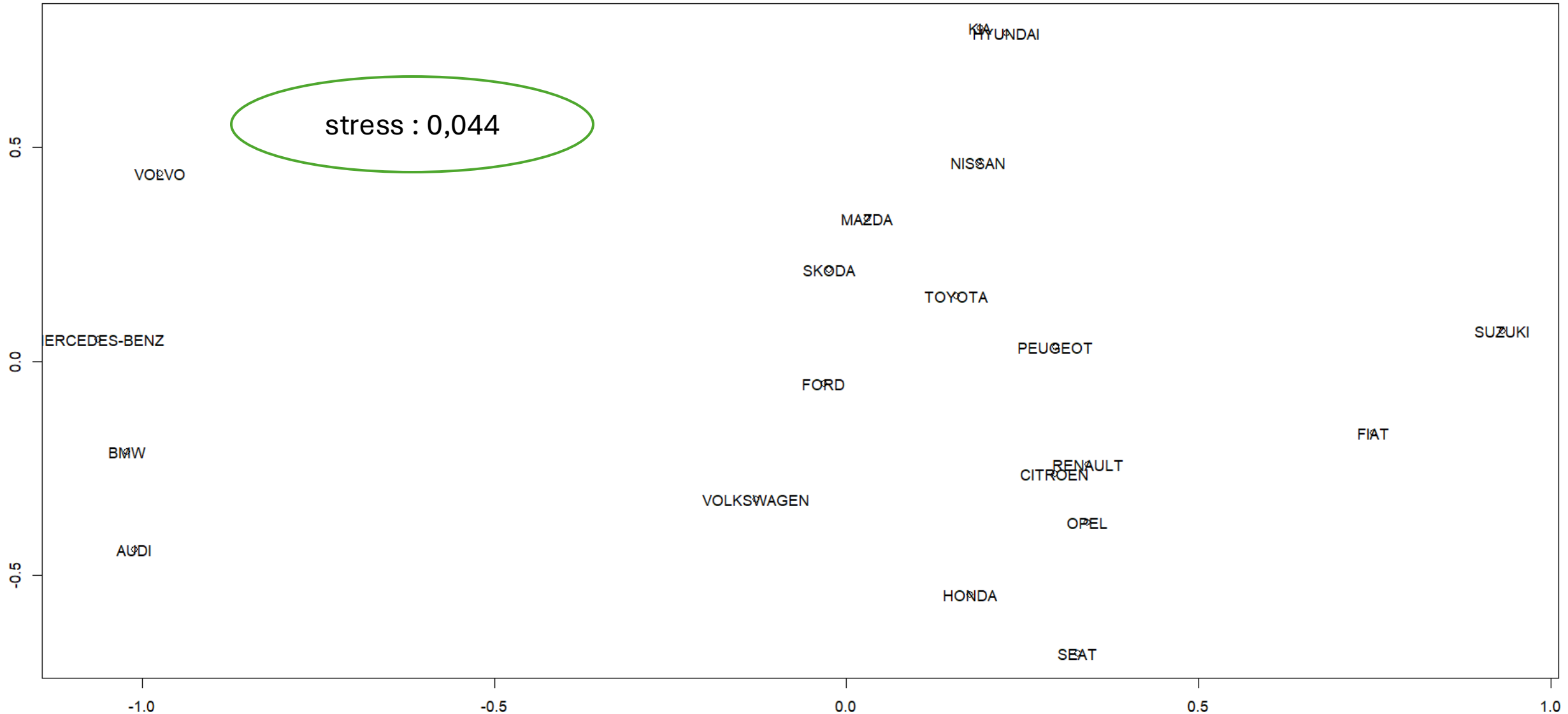
stress nem
metrikus: 0,03



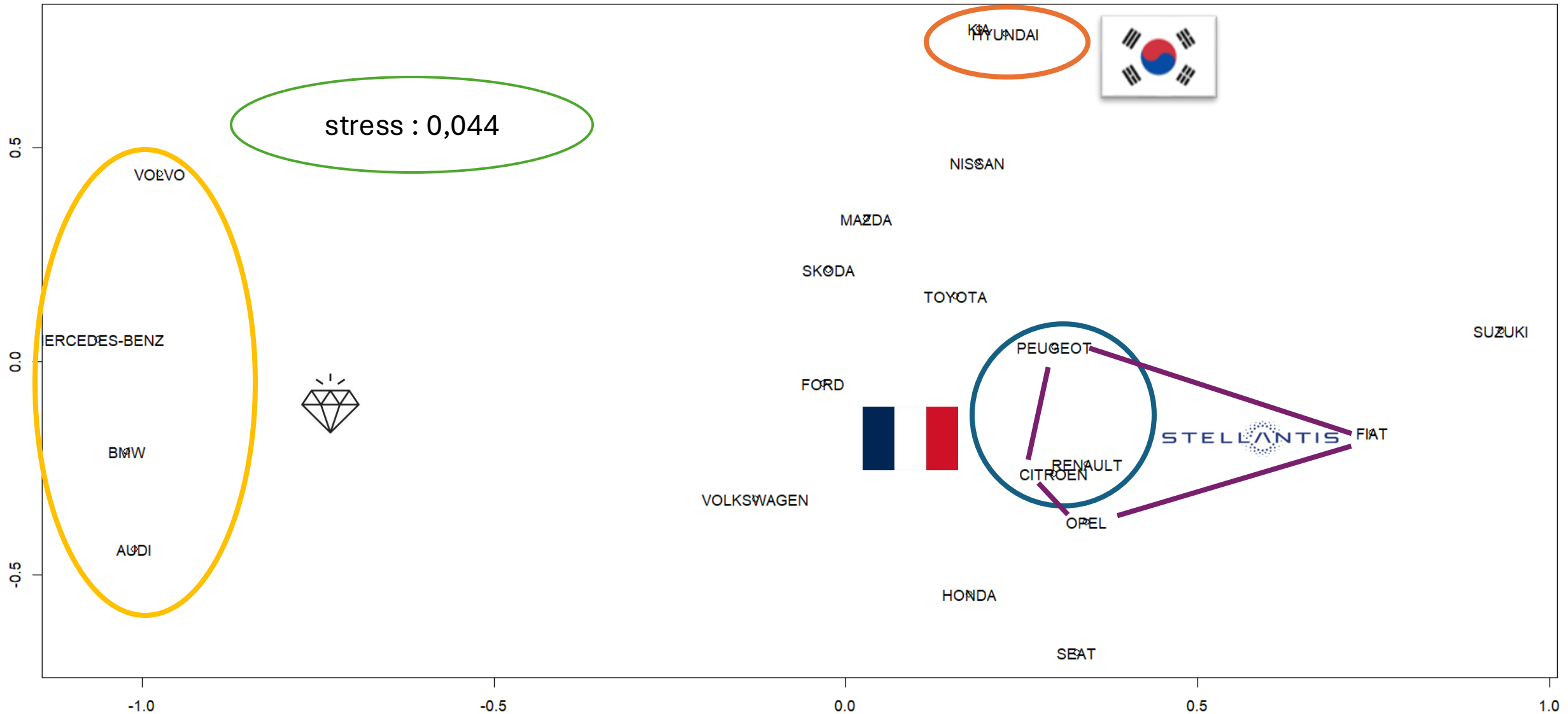
2. Modell 20 márkával



2. Modell 20 márkával, nem metrikus

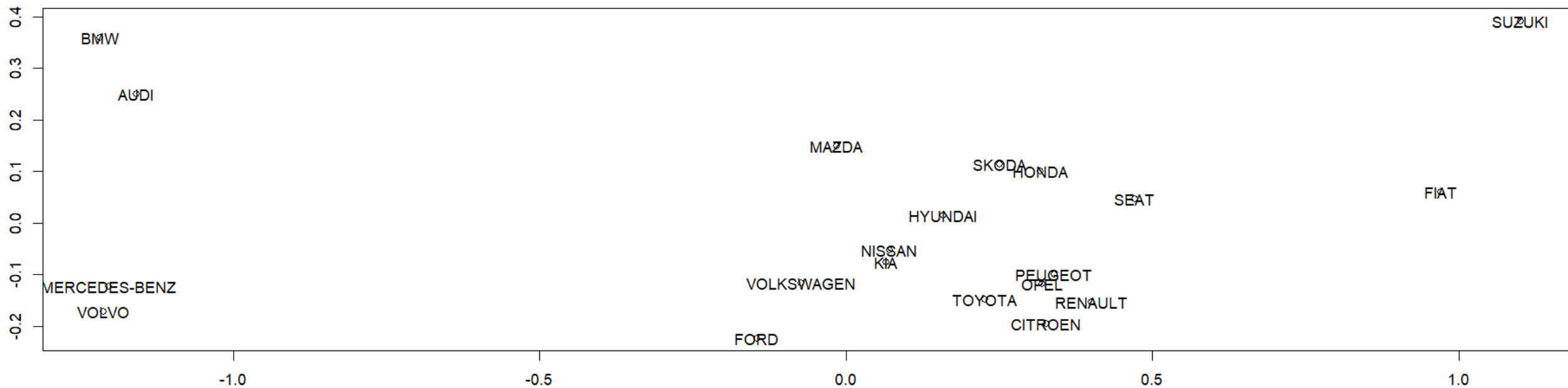


2. Modell 20 márkával, nem metrikus



Egyéb modellek

- Egy dimenzióban nem használható
- Két változó esetén nagyon pontos térképet ad vissza: $7.921811e-16$ stress



SVM

- Használt adatbázis: Kaggle-Loan Approval Classification Dataset
 - Valódi adatokon alapul, de szintetikus
 - Eredeti adatok kiegészítése plusz változókkal
 - Új adatpontok generálása SMOTNEC technikával
 - Adatból véletlen minta vétel (3000 megfigyelés)
- Numerikus (használt) változók: kor, jövedelem, munkatapasztalat, hitel nagysága, kamat, jövedelemhez viszonyított hitelnagyság, hiteltörténet hossza, hitelminősítés

Változók átlagos értékei

hiteligény eredménye	Sikeres (682)	Sikertelen (2318)
kor	27,6	28,1
jövedelem	62047	84638
munkatapasztalat	5,2	5,7
hitel nagysága	11091	9381
kamat	13	10
jövedelemhez viszonyított hitelnagyság	0,203	0,124
hiteltörténet hossza	5,8	6,1
hitelminősítés	631,5	632,2

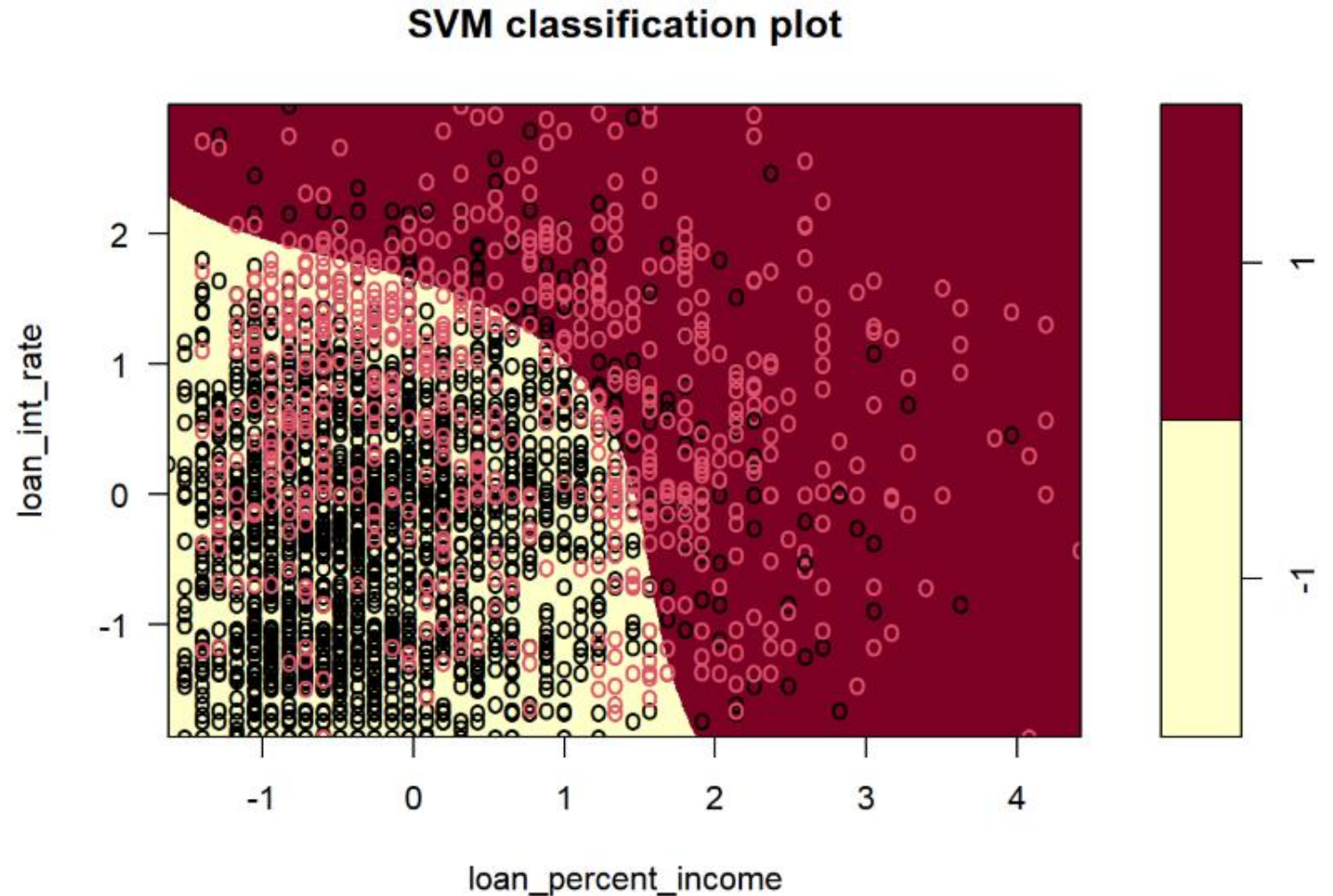
1. Modell két változóval

- Magyarázóváltozók a kamat, és a hitel nagysága a jövedelem arányában
- radiális kernel függvény
- 83,8% pontosság

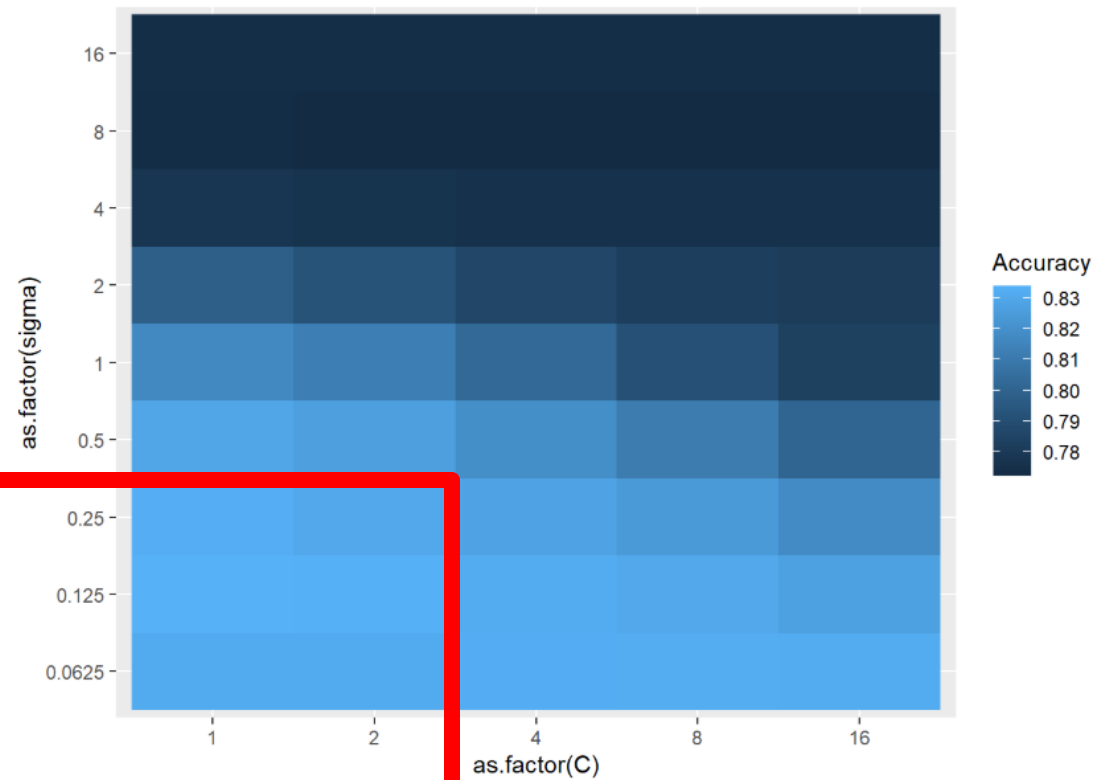


2. Modell összes változóval

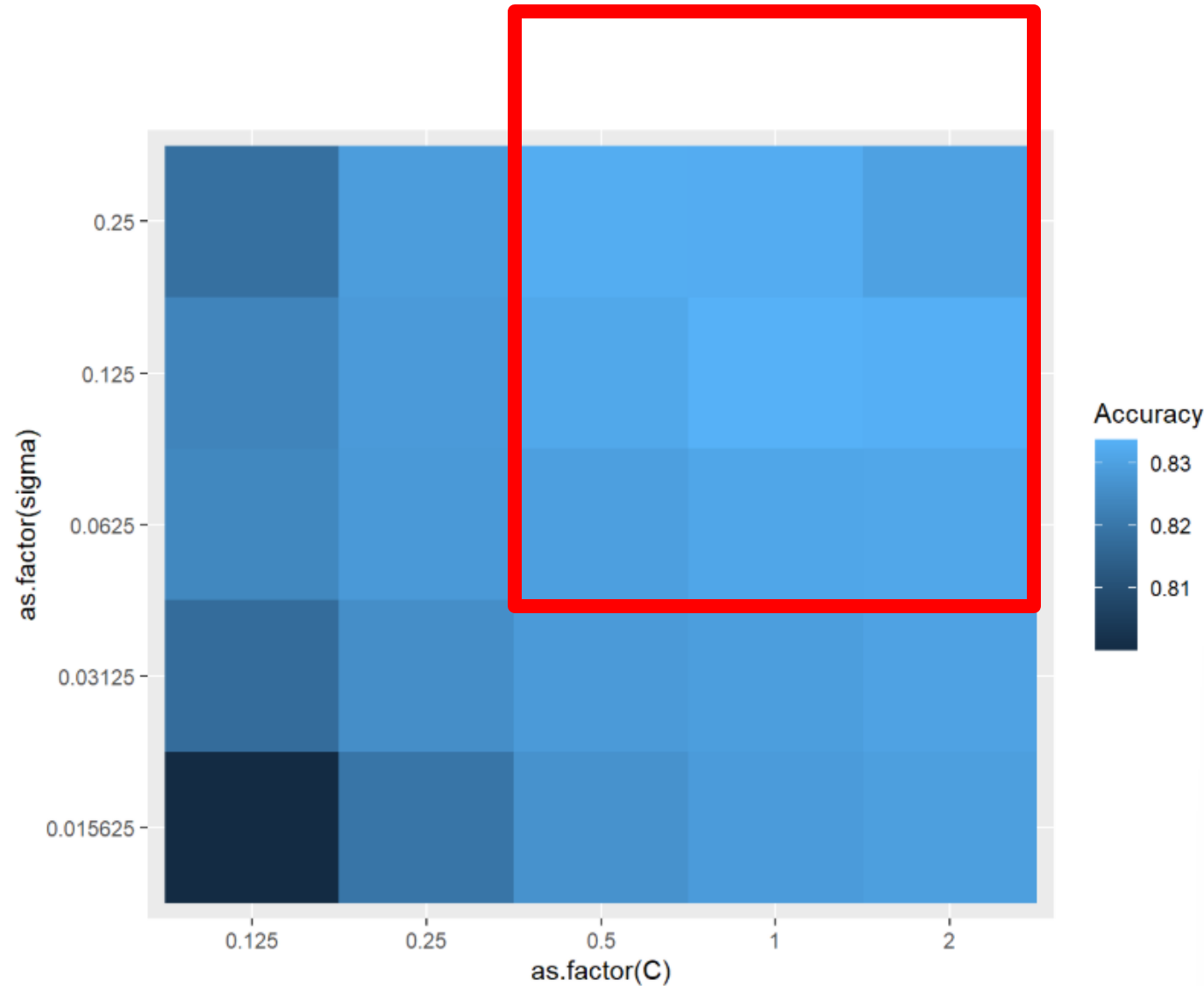
- Radiális kernel
- 84,5% pontosság



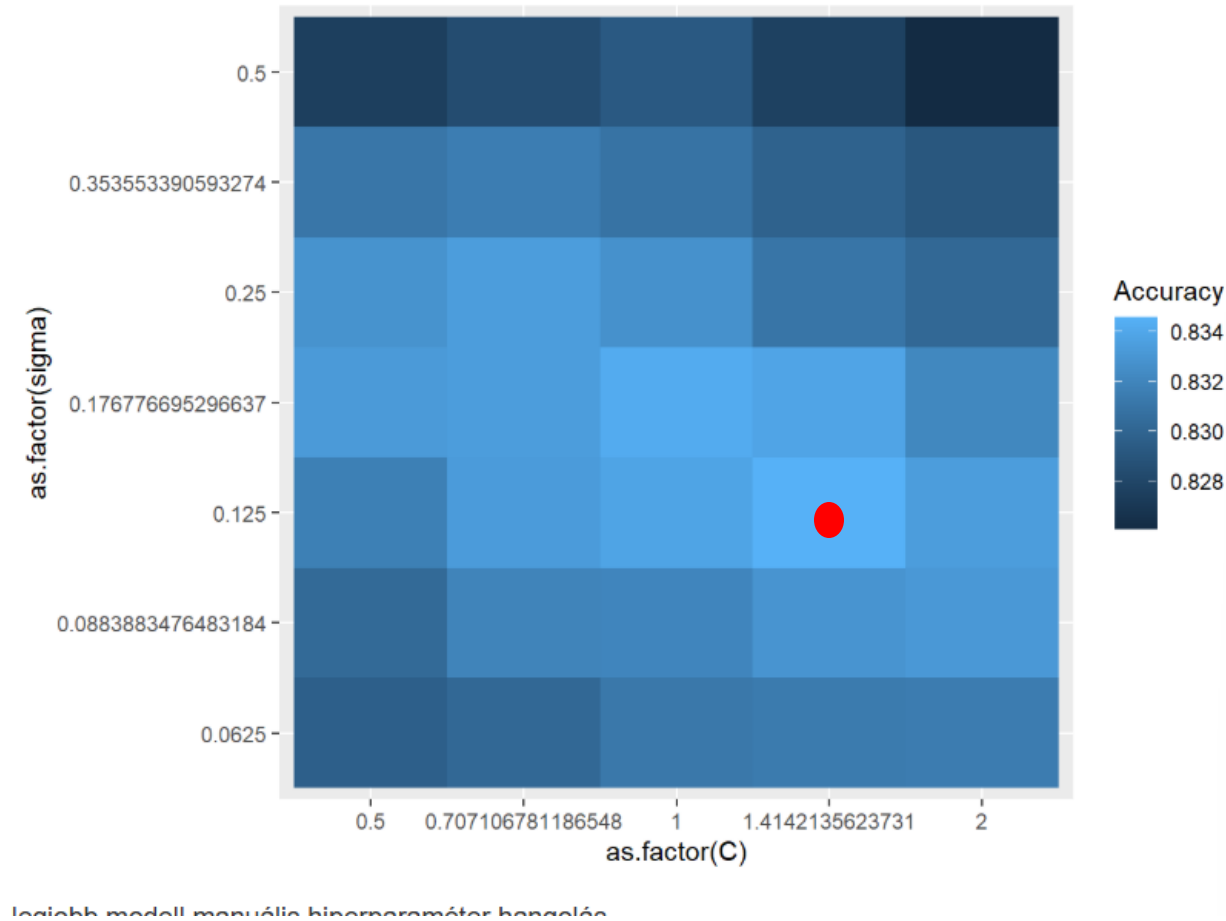
Hiperparaméter hangolás 1



Hiperparaméter hangolás 2



Hiperparaméter hangolás 3



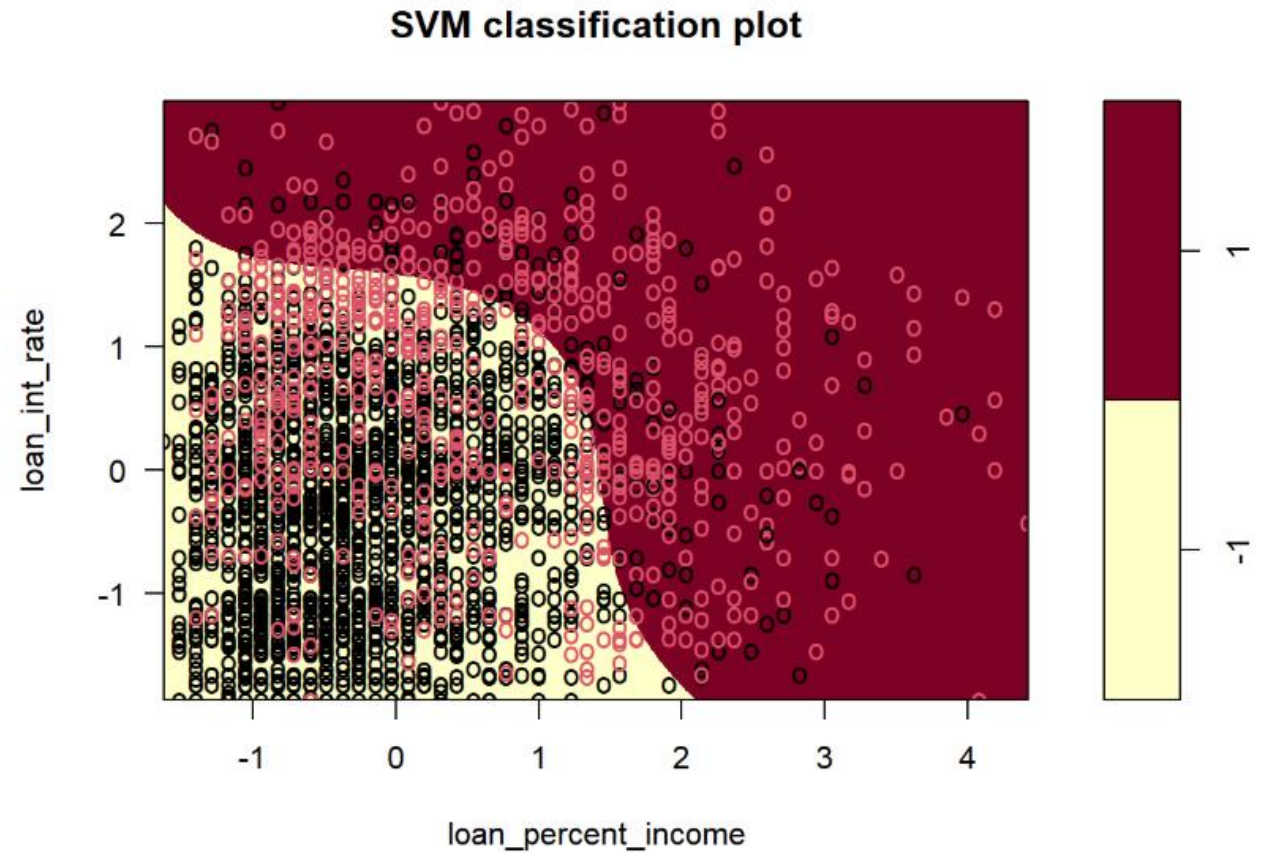
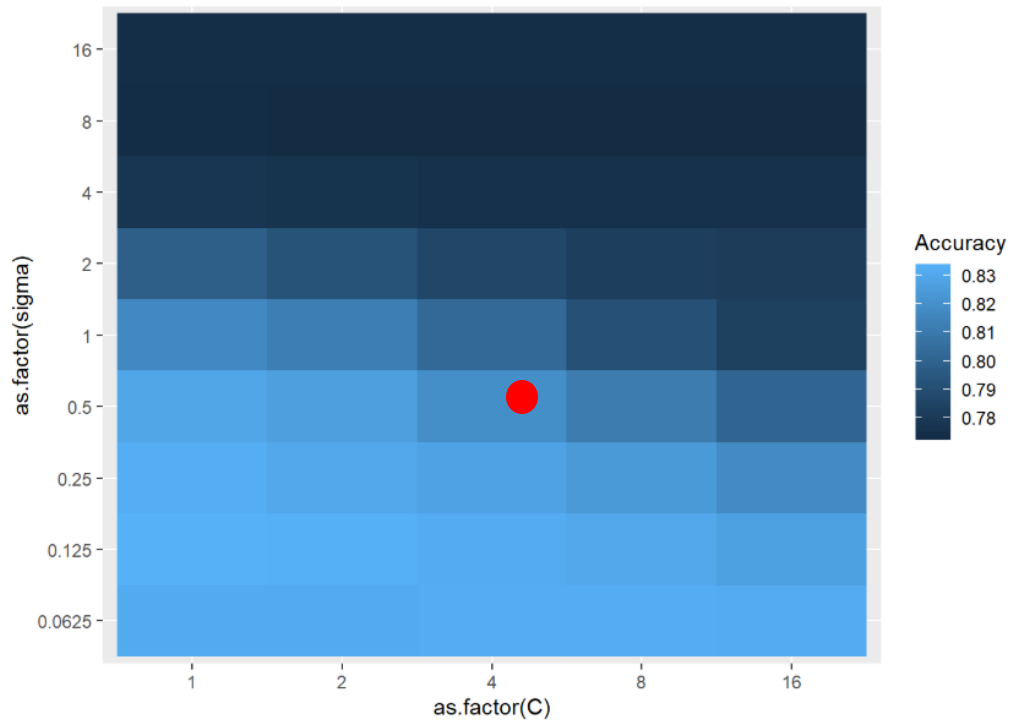
3. modell

- Manuális
Hiperparaméter
Hangolás
- 84,9% pontosság



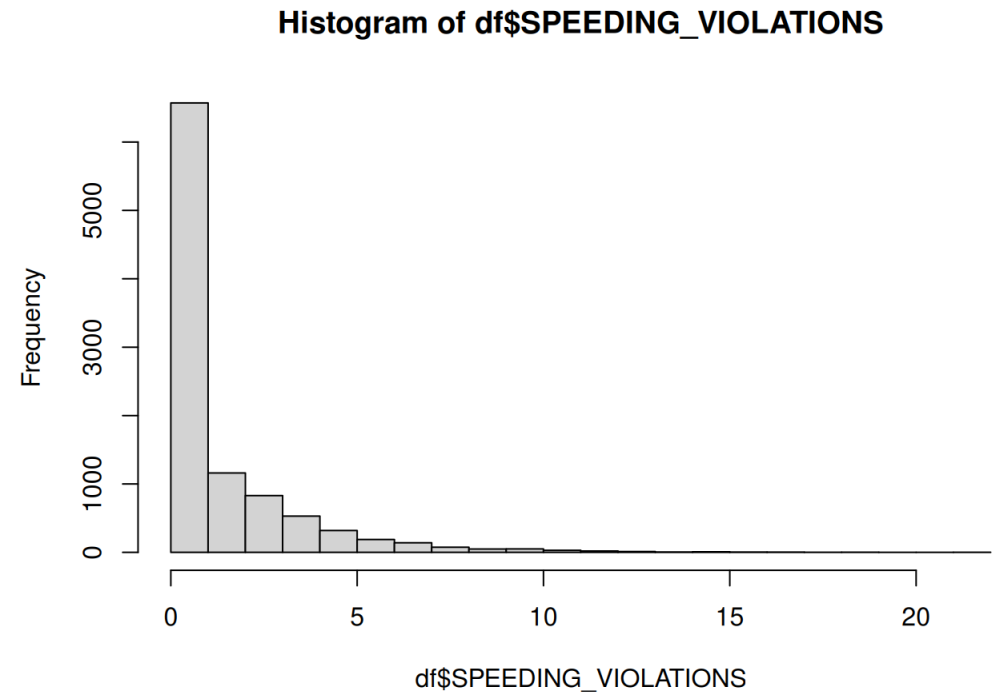
Automatikus hiperparaméterhangolás

- 85,9% pontosság



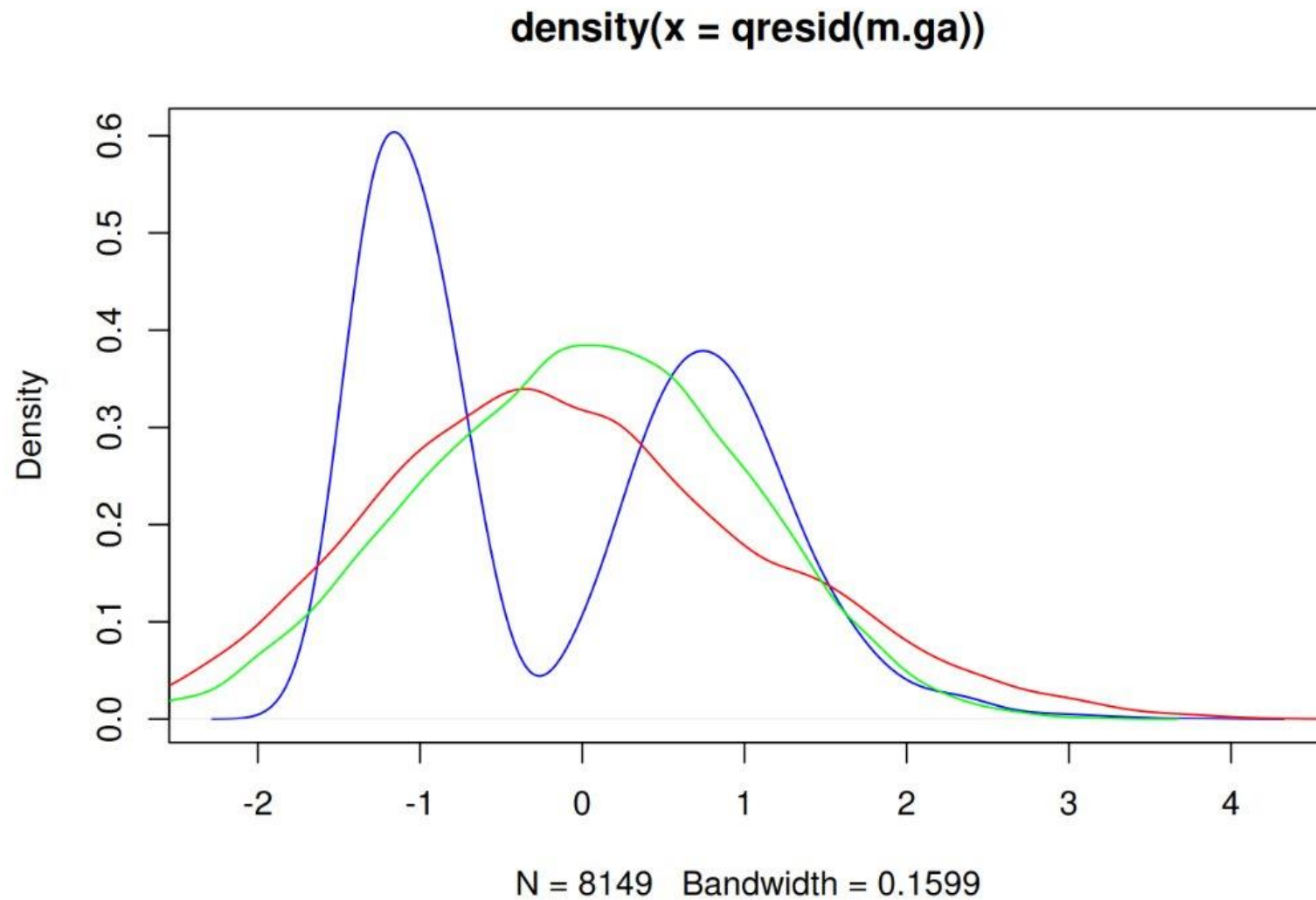
GLM

- Kaggle adatbázis, autósok gyorsajtásairól
- Nagyonbrészt kategorikus változók
 - Korcsoport, nem, rassz, oktatottság
 - Credit score, éves megtett kilométerek, gyorsajtások száma
- ~8000 adatpont



Kipróbált modellek

- Poisson, Gamma és negatív binomiális
- Modellszelekció AIC alapján
- Korcsoportok középpontjai alapján offsetelve
- Ez alapján Gamma a legjobb
- Viszont hibatagok eloszlását ellenőrizni kell
- Dunn és Gordon (2018) féle kvantilis maradéktagok



Kék - Gamma
Piros - Poisson
Zöld - Negatív binomiális

Modell értelmezése (negatív binom)

Szignifikánsak

VÁLTOZÓ	Hatás iránya
Férfi	+
Egyetemen tanult	-
Szegény és munkásosztálybéli	-
Felsőosztálybéli	+
Házass	+
Gyermekek	+
Log(Éves megtett Km-ek)	-

Nem szignifikánsak

- Rasz
- Maximum általános iskolában tanult
- Credit Score
- Autó tulajdonos
- 5 évnél idősebb autó
- Sportautó
- Irányítószám, kivéve egyet

Köszönjük a figyelmet!