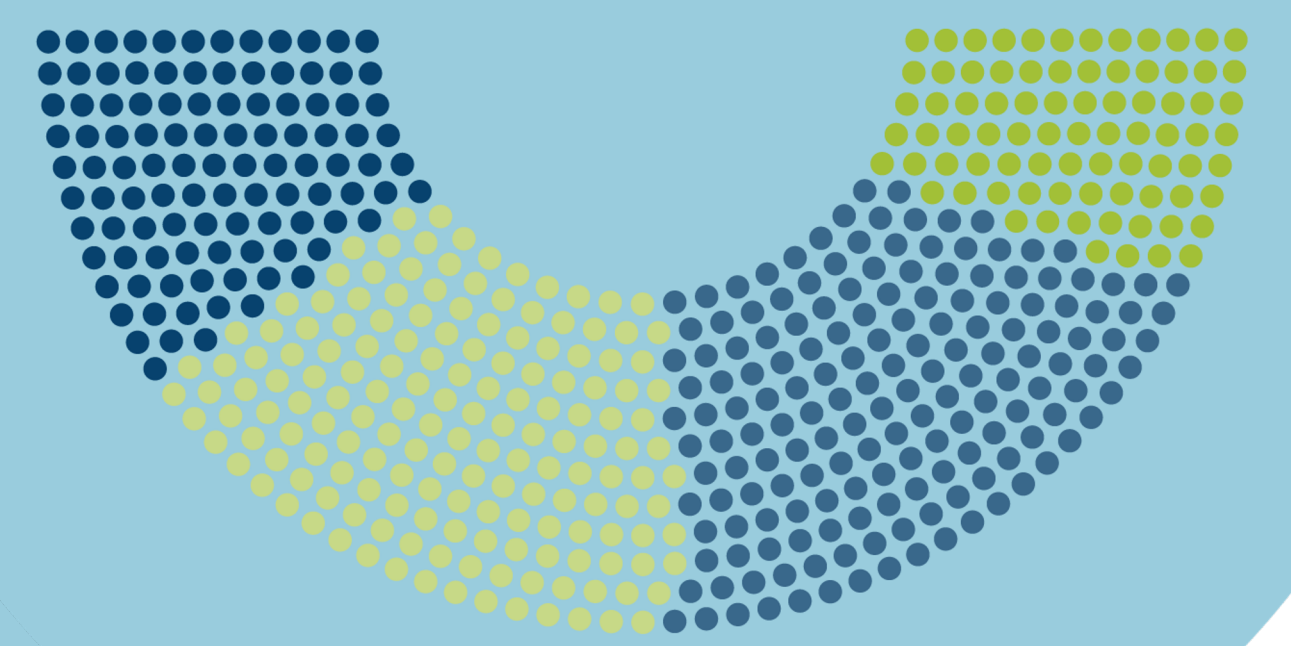


Parlamendikorpused ja mis neist leida võib

ParlaMint



Kadri Vider, Neeme Kahusk, Martin Mölder

Eesti Kirjandusmuuseum, Tartu Ülikool (Johan Skytte Instituut)

Kadri.vider@eesti.ee, Neeme.kahusk@gmail.com, Martin.molder@ut.ee

Eesti ParlaMinti korpus

- Riigikogu istungite stenogrammid
4. aprill 2011 -- 17. juuni 2022
12., 13. ja 14. Riigikogu koosseis
- 1318 stenogrammi**
Iga tekst sisaldab ka linki Riigikogu kodulehel olevale originaalstenogrammile.
- sõnu: 22 882 661**
koos kirjavahemärkidega: 26 967 450
- lauseid: 1 810 808

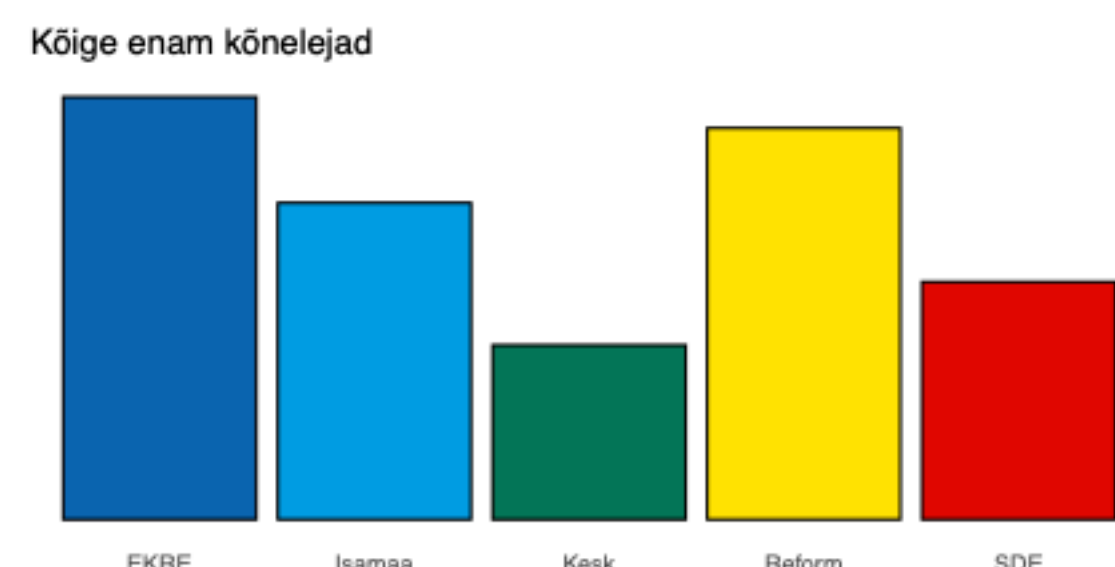
Kõnelejad

- ca **490 kõnelejat**, kõige rohkem sõnavõtte ehk lauseid ja sõnu on spiikeritel
- Kõnelejate **rollid**:
spiiker, asespiiker
RK liige
külaline (s.h president ja valitsuse liikmed)
- Iga kõneleja kohta **meta-andmed**: sugu, sünnikuupäev, erakond, fraktsioon, valimisinfo (nimekiri, mandaat, valimisringkond, häälte arv)

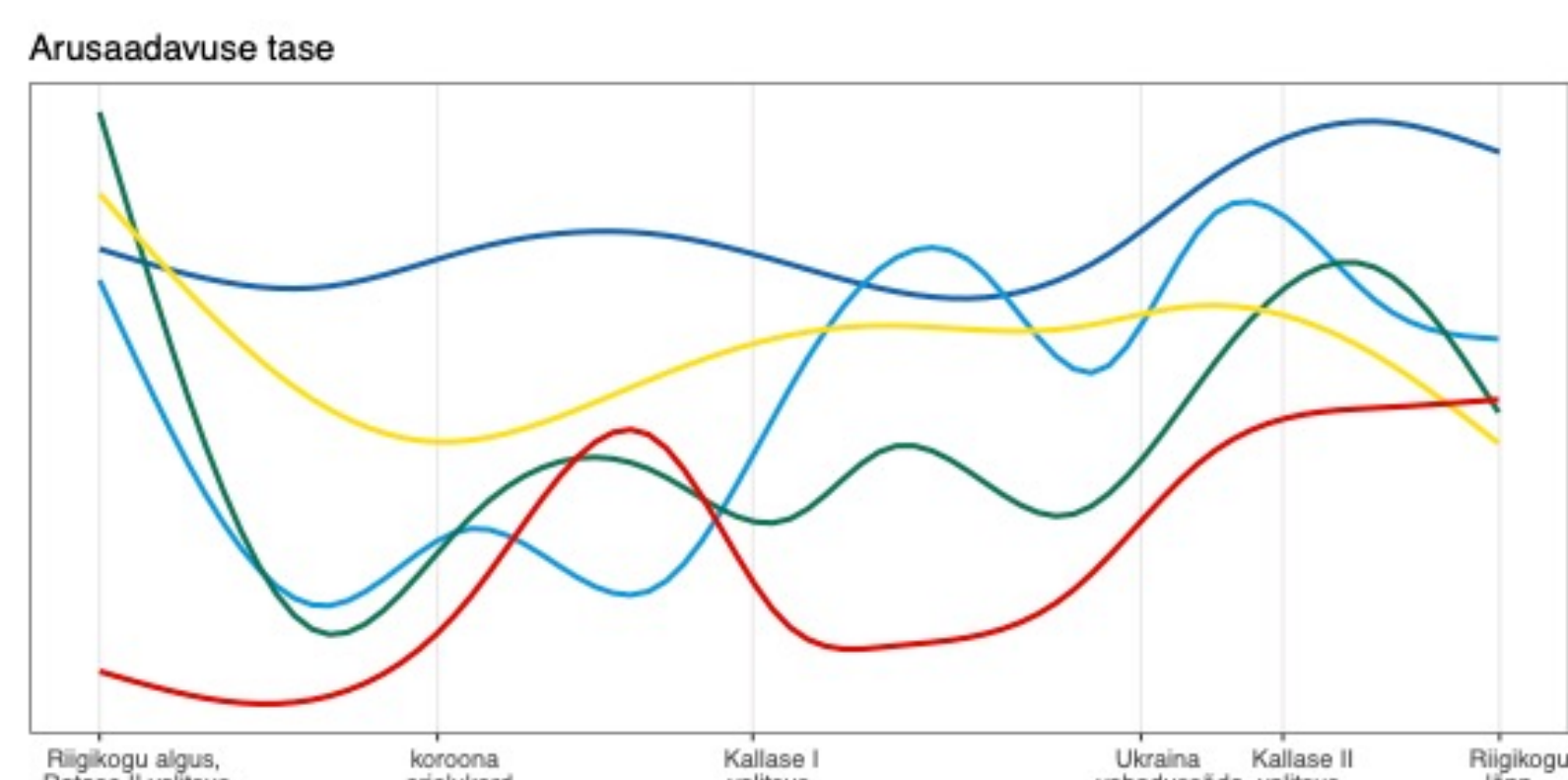
Märgendus

- Märgenduskihid**:
lemma,
sõnaliik,
vormianalüüs,
lauseanalüüs,
nimeüksused (isik, koht, organisatsioon)
- TEI vormingus XML-skeem, Parla-CLARINi arendus

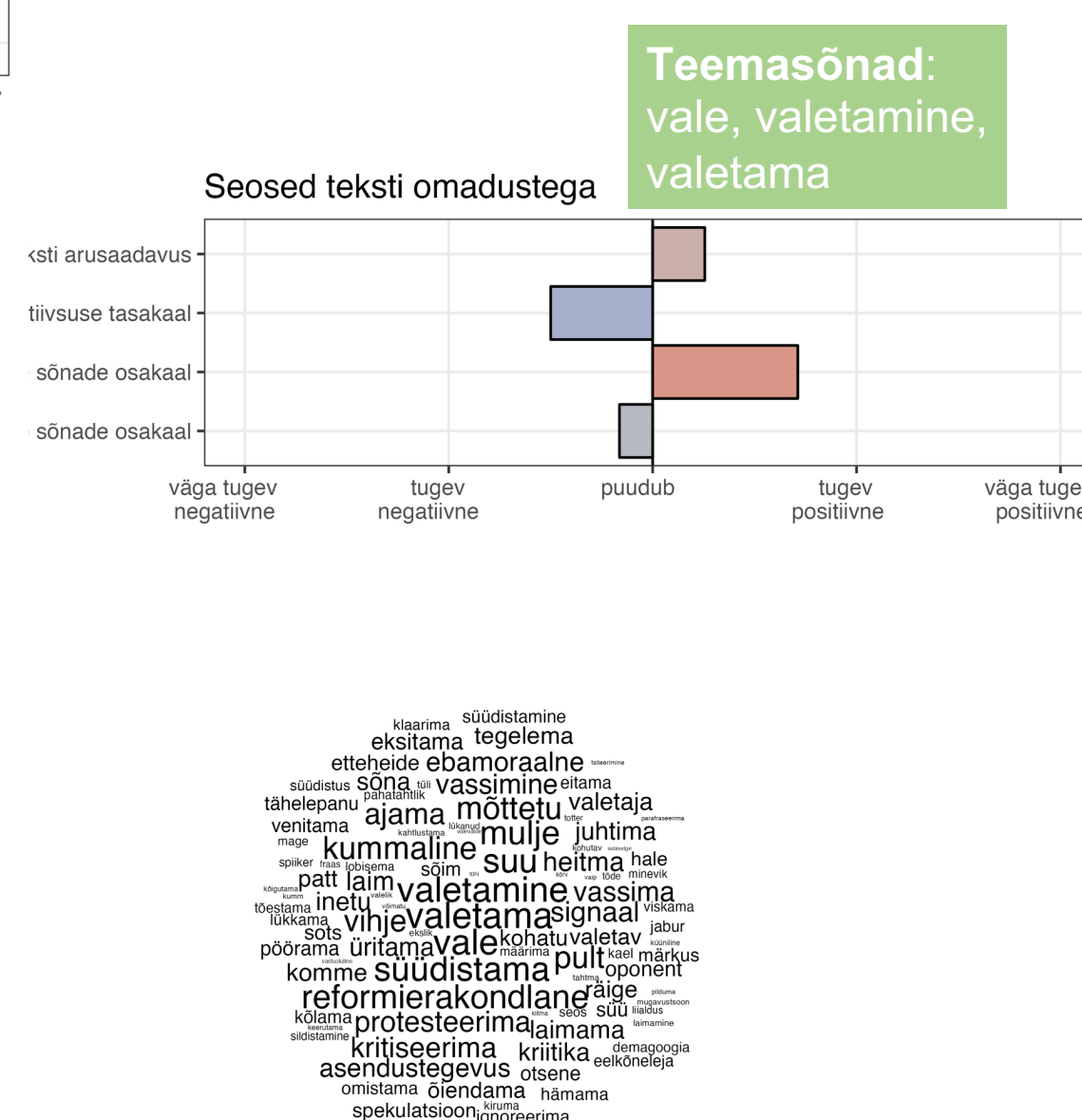
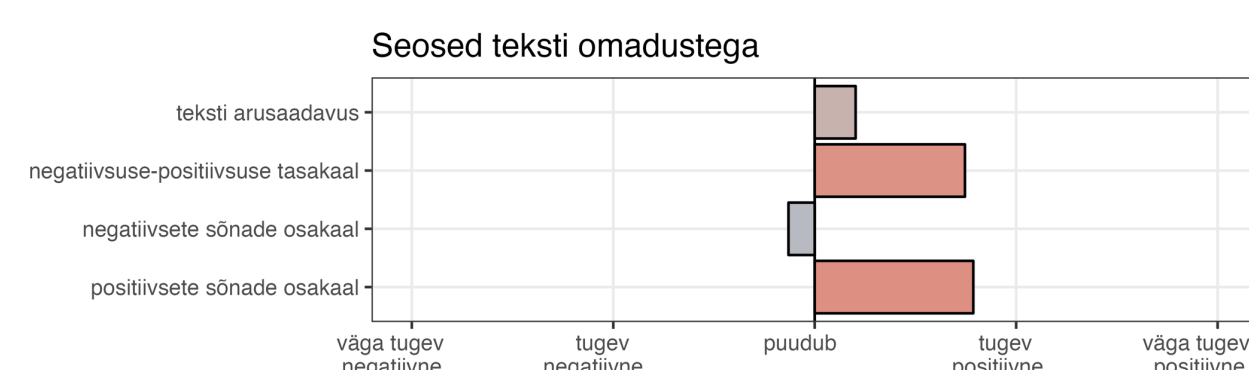
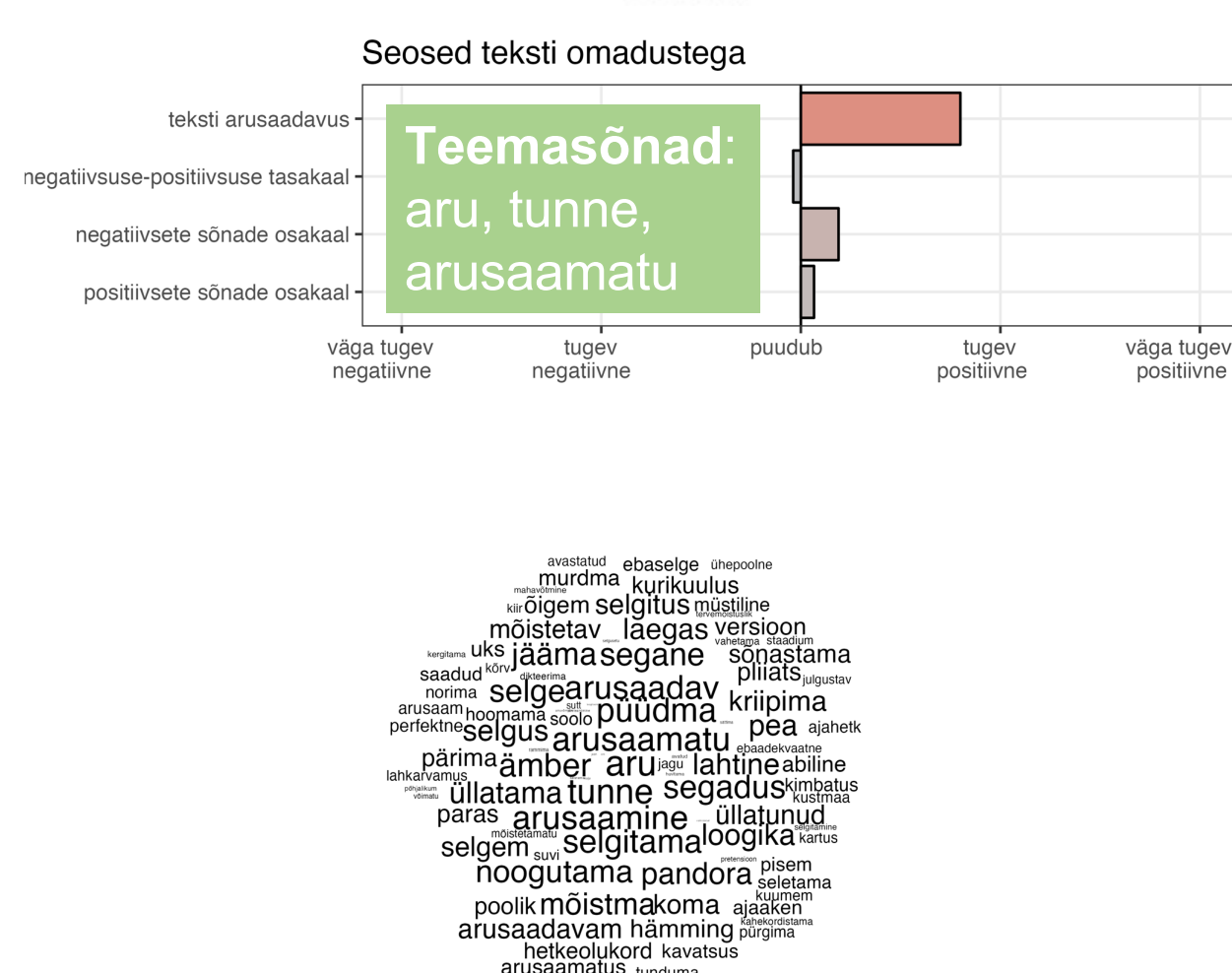
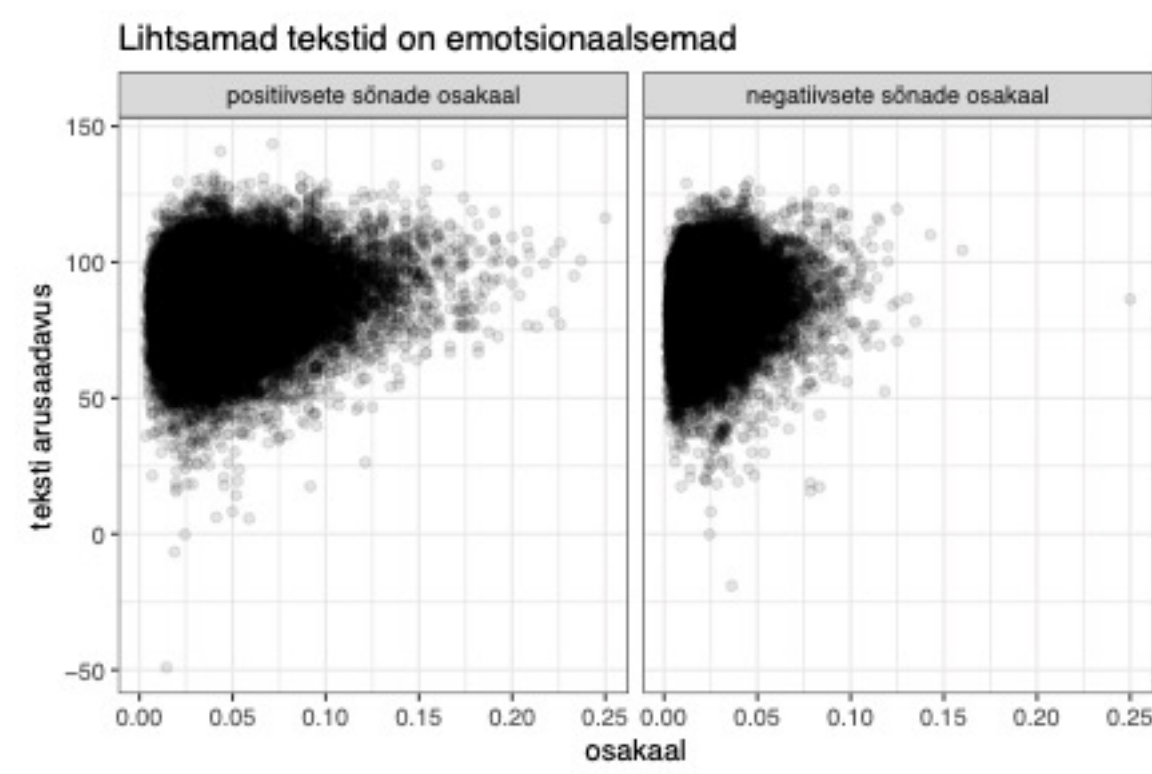
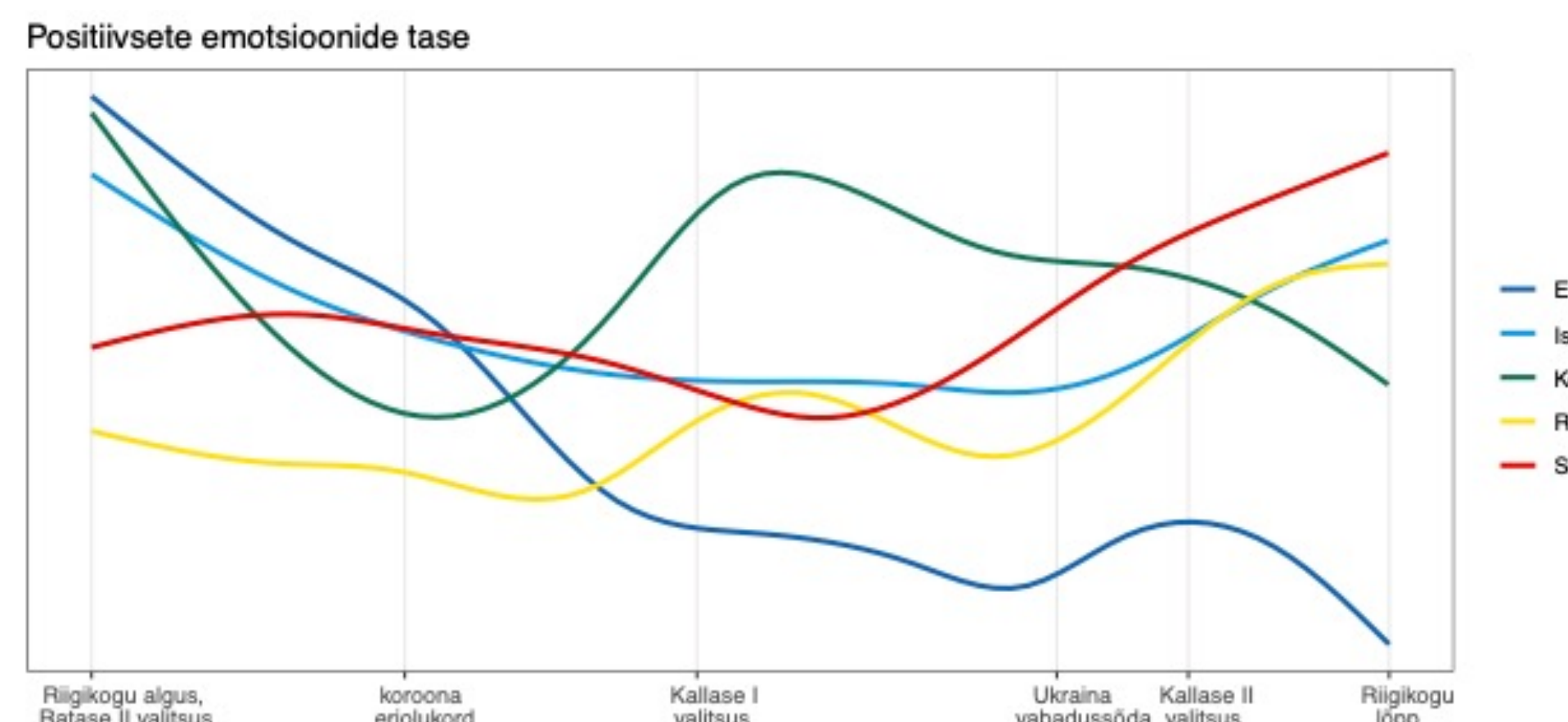
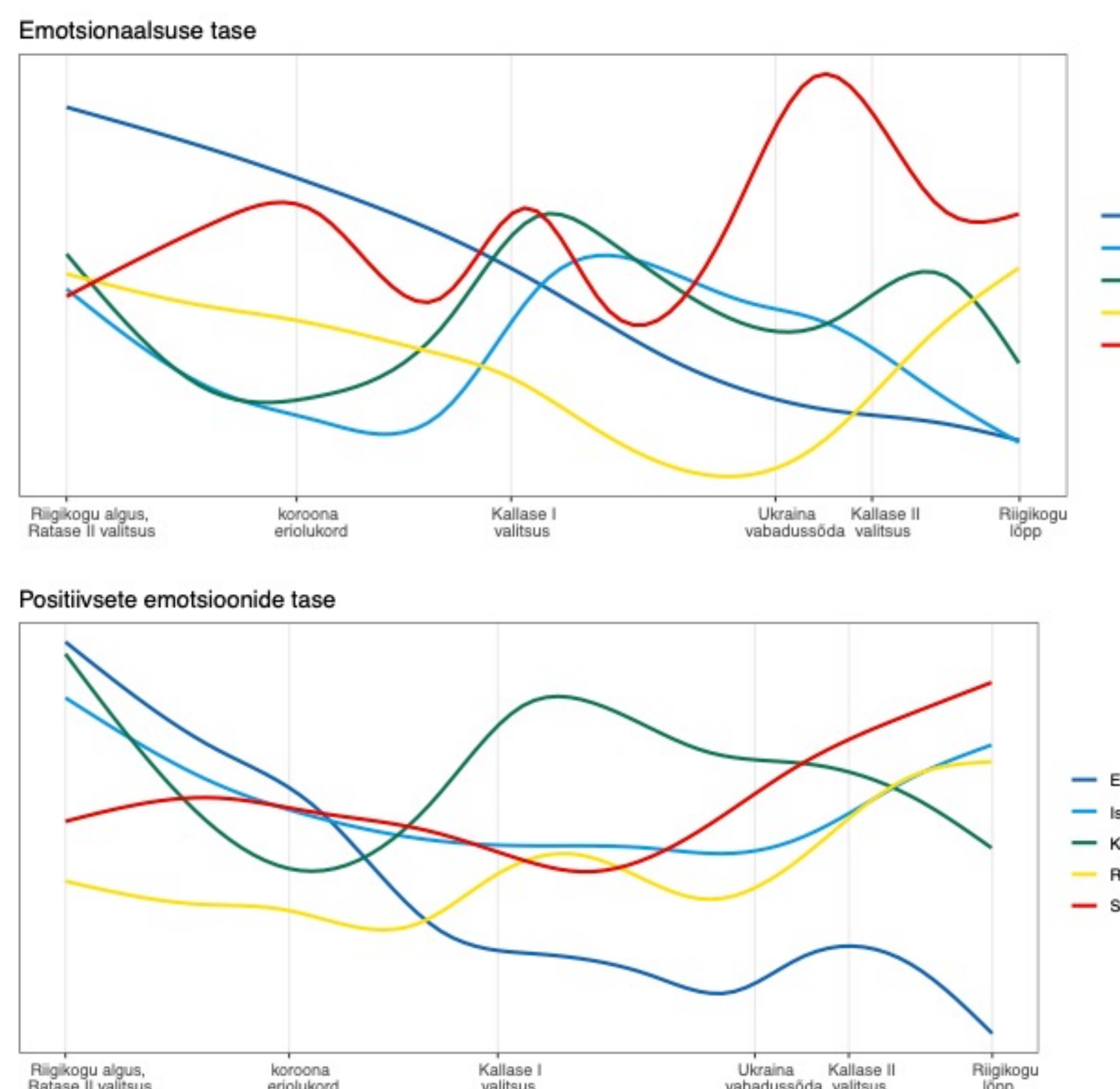
XIV Riigikogu kõnede lihtsus, meelsus ja teemad



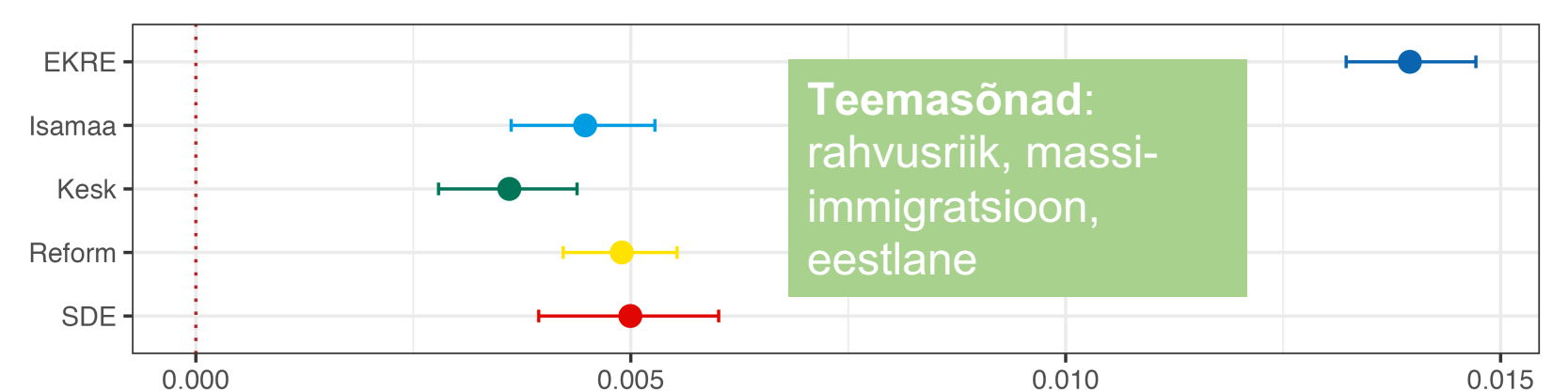
Arusaadavus ehk lihtsus



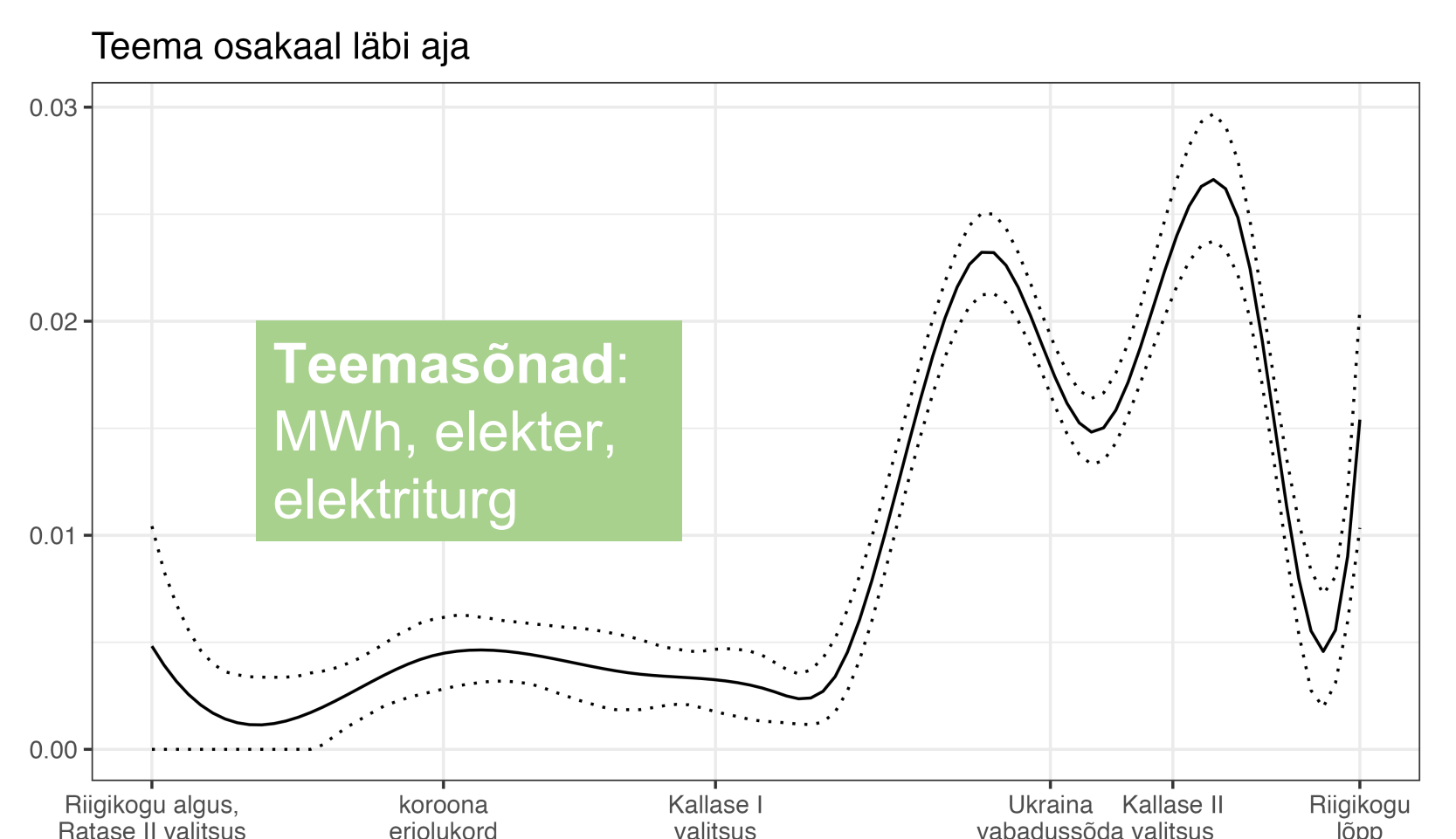
Meelsus, nii positiivne kui negatiivne



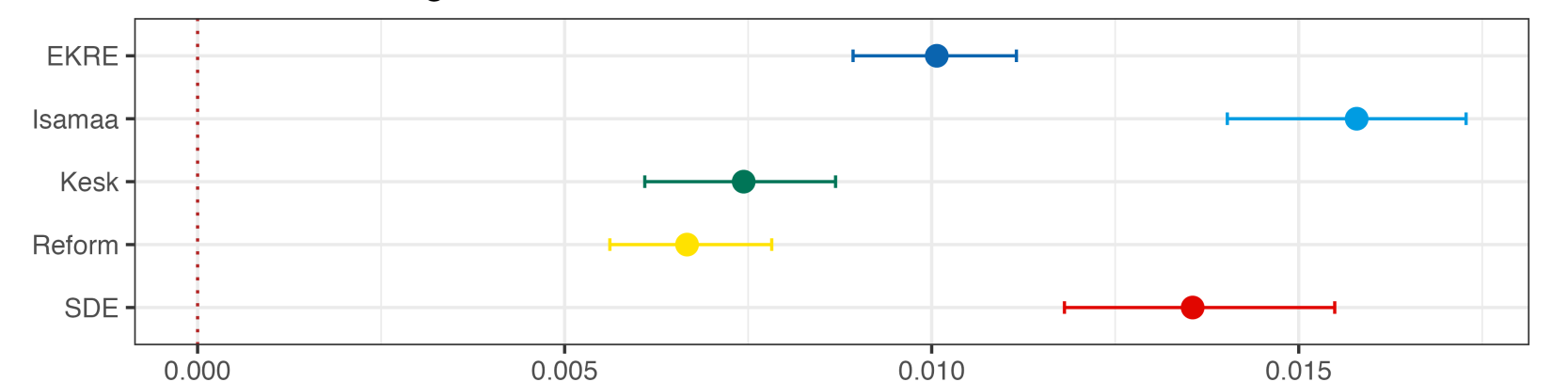
Millise erakonnaga on teema seotud?



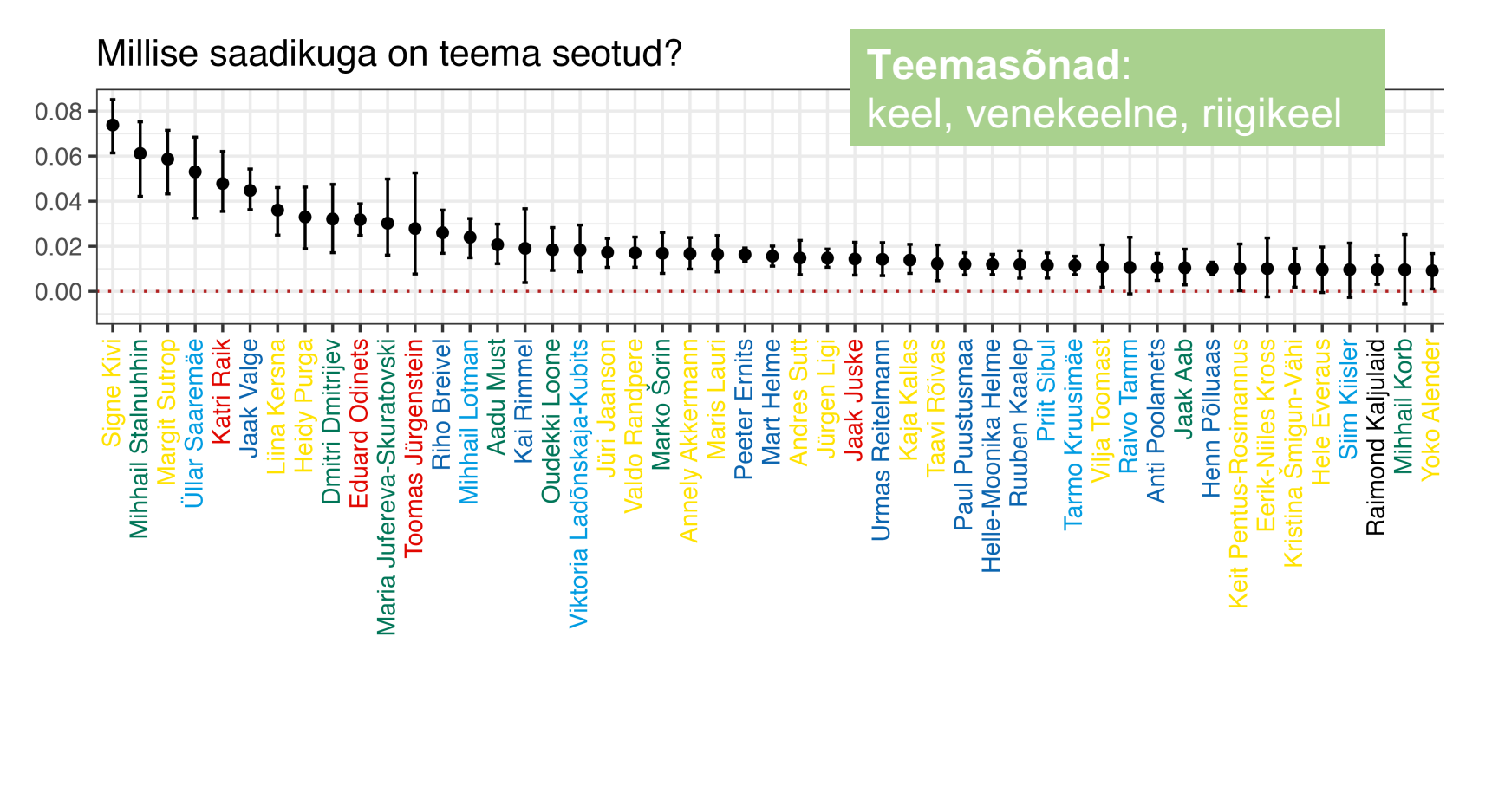
Millal teemad aktuaalsed olid?



Millise erakonnaga on teema seotud?



Millise saadikuga on teema seotud?



Viited

ParlaMint projekt: <https://www.clarin.eu/parlamint>

ParlaMint GitHub: <https://github.com/clarin-eric/ParlaMint>

Riigikogu stenogrammid 1990-2021 ja nende erinevad analüüsid (Martin Mölder veebileht):

<https://eestipoliitika.ee/>

Eesti parlamendikorpused korpuste päringusüsteemis KORP: <https://korp.keeleressursid.ee/>

Fleschi teksti lihtsuse skoor (sõltub sõnade arvust lauses ja silpide arvust sõnas)

Laur, Sven, Siim Orasmaa, Dage Särg, and Paul Tammo (2020). "EstNLTk 1.6: Remastered Estonian NLP Pipeline". In: Proceedings of The 12th Language Resources and Evaluation Conference. Marseille, France: European Language Resources Association, pp. 7154–7162. url: <https://www.aclweb.org/anthology/2020.lrec-1.884>.

Emotsionaalsus eesti valentsisõnastiku põhjal (sõltub pos ja neg sõnade osakaalust teksti sõnade koguhulka)

Pajupuu, Hille, Rene Altrov, and Jaan Pajupuu (2016). "Identifying polarity in different text types". In: Folklore: Electronic Journal of Folklore 64, pp. 126–138. <http://dx.doi.org/10.7592/FEJF2016.64.polarity>

Teemade klassifitseerimiseks kasutatud strukturealne teemamudel (200 teemat)

Roberts ME, Stewart BM, Tingley D (2019). "stm: An R Package for Structural Topic Models." Journal of Statistical Software_ *91*(2), 1-40 <https://doi.org/10.18637/jss.v091.i02>



CLARIN
ERIC

Common Language Resources and
Technology Infrastructure