

Riigikogu

Risto Hinn

Friday, June 26, 2015

Tegeleme sama andmestikuga, mis jäi viienda kodutöö viiendas ülesandes koostada. Sellele oleme lisanud ka fraktsiooni tunnuse ning iga hääletuse nimetuses on lisaks kuupäev. Andmestikku on jäetud saadikud, kes hääletasid vähemalt 10 korda. Täpsemalt on andmestikus tunnused:

- nimi - saadiku nimi
- fraktsioon - millisesse fraktsiooni kuulus saadik (kui saadik vahetas fraktsiooni, siis on näidatud kõige sagedasem fraktsioon, kuhu saadik Riigikogu XII vältel kuulus)
- 2011-04-05 2821 - hääletuse tulemus, mis toimus 2011-04-05 (2821 on kontrollindeks ja ei oma sisulist tähendust)
- 2011-04-11 2825 - hääletuse tulemus, mis toimus 2011-04-05 (2825 on kontrollindeks ja ei oma sisulist tähendust)
- 2014-12-18 5363 - hääletuse tulemus, mis toimus 2014-12-18 (5363 on kontrollindeks ja ei oma sisulist tähendust)

```
andmed=read.csv("./data/riigikogu_xii.csv", as.is = TRUE)
```

Ülesanne 1 (1 punkt) - andmestiku mudimine 1

Muuda tunnused numbriliseks. Selleks kodeeri hääletused järgmiselt:

- vastu tähistab arvuga -1
- poolt tähistab arvuga 1
- puudub, erapooletu, ei hääletanud, NA tähistab arvuga 0

Pane tähele, et tunnused oleksid numbrilised mitte sõnad.

```
library(car)
data <- apply(andmed, 2, function(x) {x <- recode(x, "'vastu'=-1; 'poolt'=1; NA=0;
                                                'puudub'=0; 'erapooletu'=0;
                                                'ei hääletanud'=0"); x})

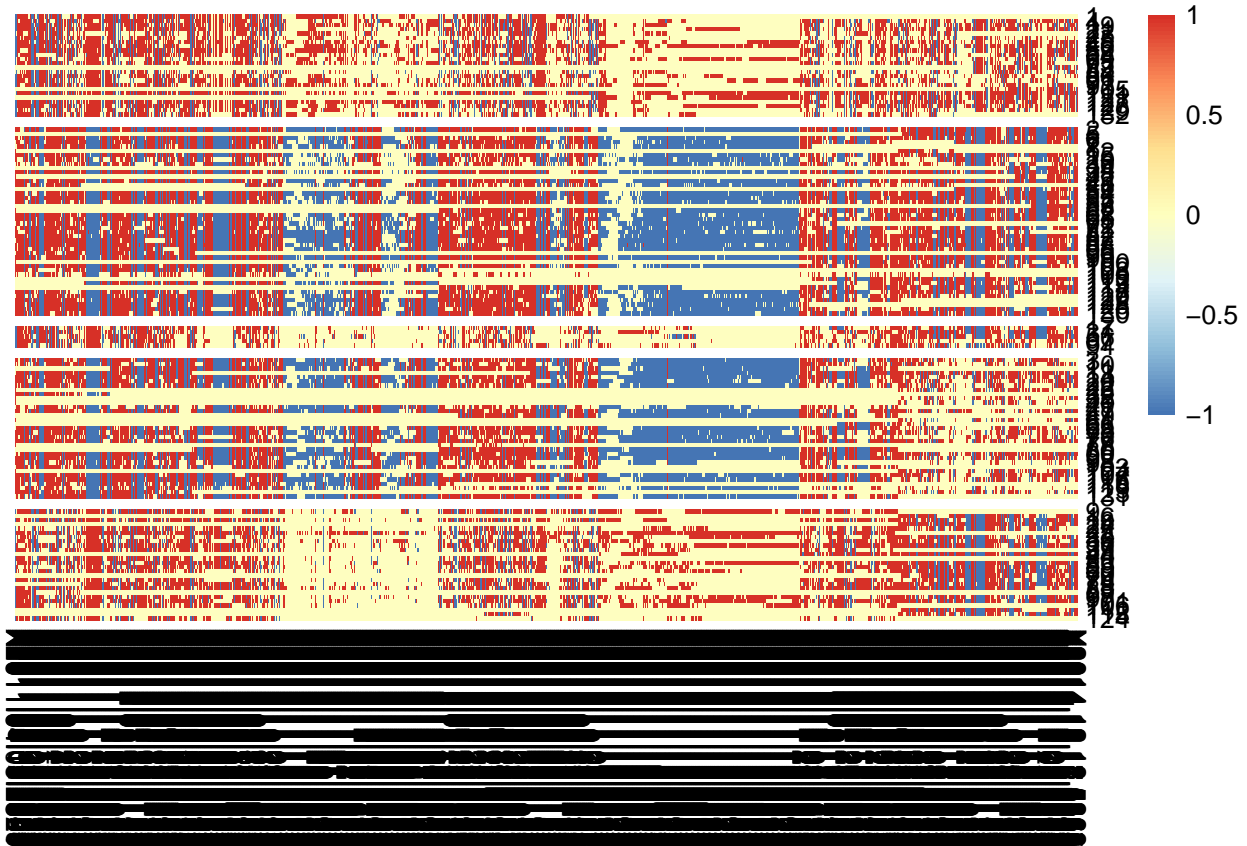
#teeme numericuks
data_numeric <- apply(data[, 3:1847], 2, function(x) {x <- as.numeric(x); x})
data=as.data.frame(data)
data[,3:1847]=data_numeric
```

Ülesanne 2 (2 punkti) - visualiseeri andmeid heatmapi abil

Sorteeri andmestiku read fraktsiooni põhjal ja visualiseeri neid heatmapi abil. Tõlgenda tulemusi.

```
data_order=data[order(data$fraktsioon),]

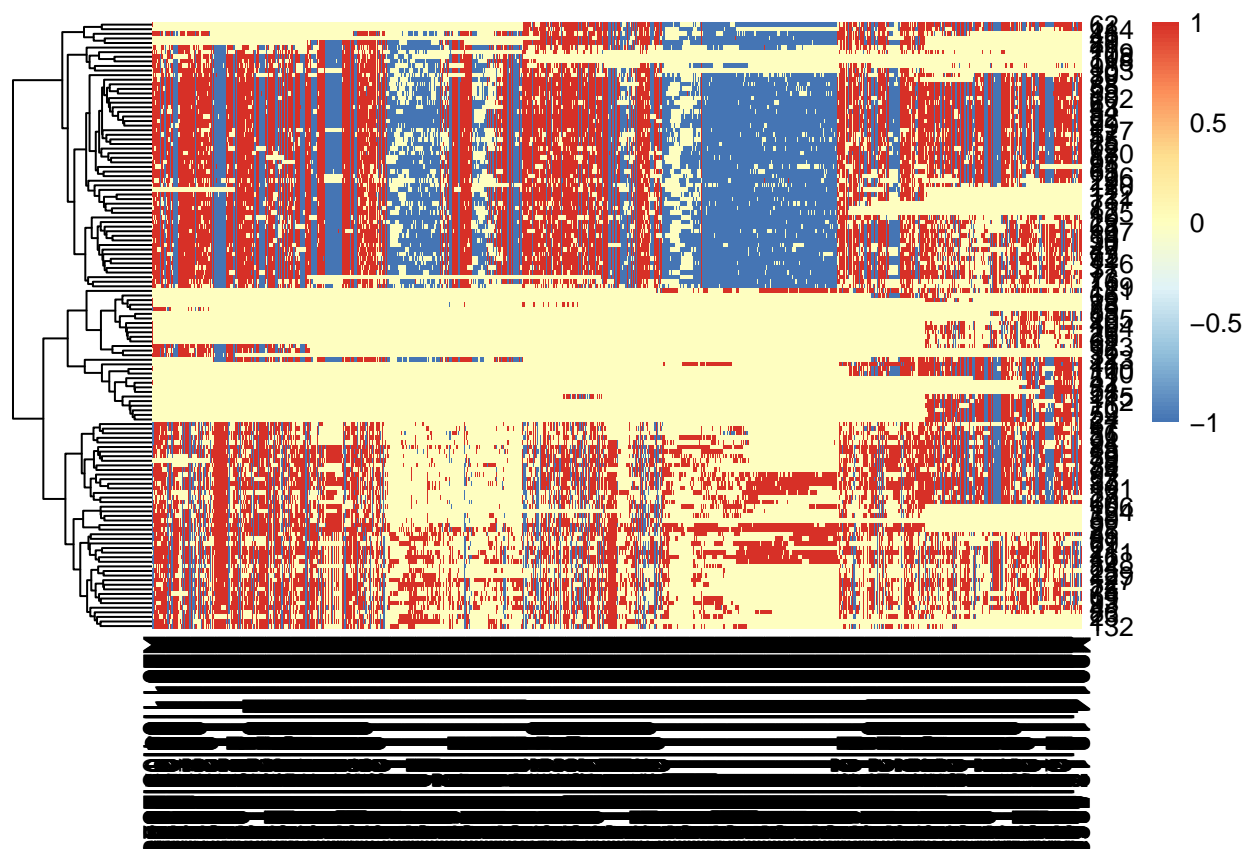
library(pheatmap)
pheatmap(data_order[, 3:1847], cluster_rows=FALSE, cluster_cols=FALSE,
         gaps_row = c(24, 68, 73,106)) #panin ka vahed iga fraktsiooni vahele
```



- Tõlgendus: Erinevate fraktsioonide hääletustulmustes on näha erinevusi, samuti on näha erisusi ajalises plaanis, kus mõnel perioodil on sarnasused suuremad.

Visualiseeri andmestikku heatmapi abil, aga seekord klasterda ka read. Tõlgenda tulemusi.

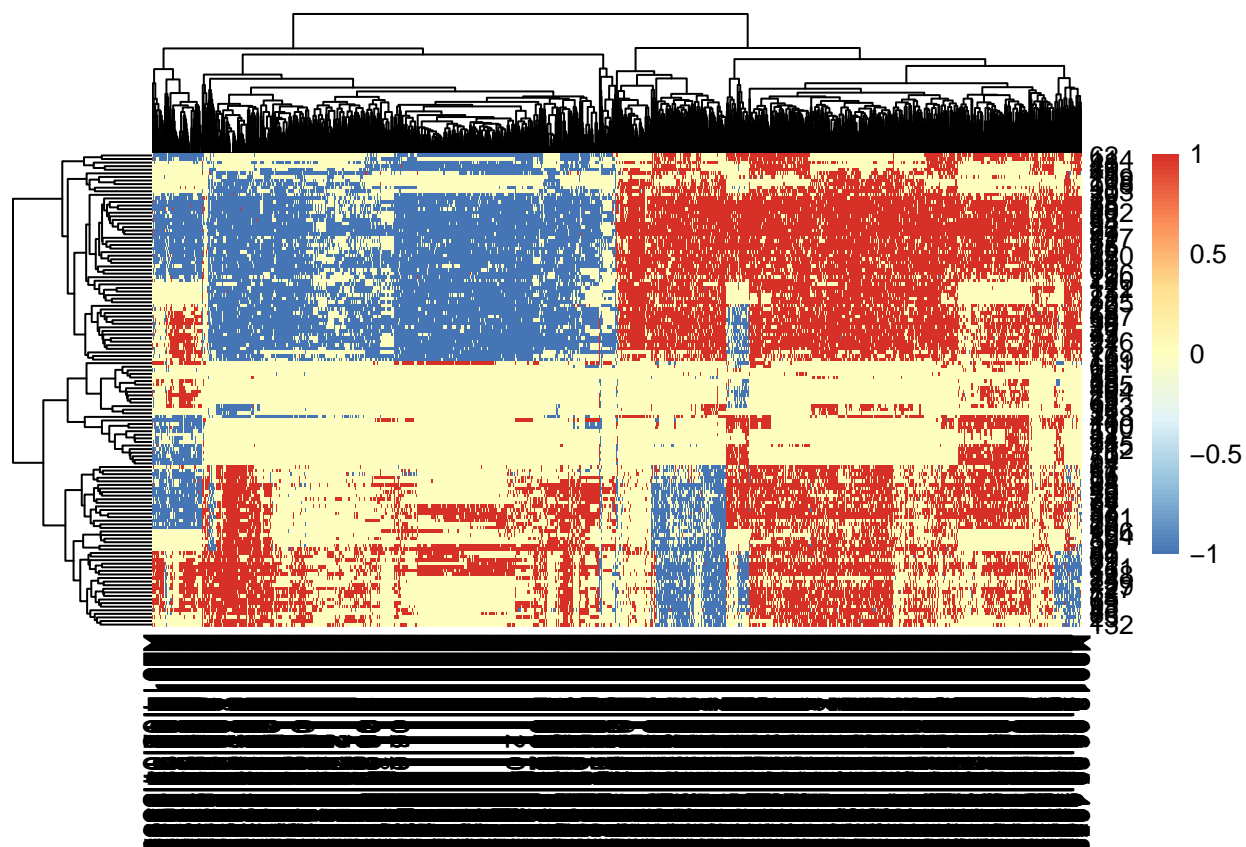
```
pheatmap(data_order[, 3:1847], cluster_rows=T, cluster_cols=FALSE)
```



- Tõlgendus: kõige laiemal tasandil jaguneb fraktsioonide hääletusmustrid kaheks.

Visualiseeri andmestikku heatmapi abil, aga seekord klasterda nii read kui veerud. Tõlgenda tulemusi.

```
pheatmap(data_order[, 3:1847], cluster_rows=T, cluster_cols=T)
```



- Tõlgendus: ajalisel plaanis on näha, et eristuvad selgelt 2 perioodi, kus suure osa saadikute hääletusmuster on muutunud vastupidiseks alperioodil olevaga

Ülesanne 3 (1 punkt)

Riigikogu 2011. aasta valimised võitis Reformierakond ning koalitsioonilepe sõlmiti REF ja IRL vahel. See koalitsioon tegi tööd 3 aastat enne kui purunes. 26. märtsil 2014 algas Taavi Rõivase valitsus, kus koalitsioonis olid REF ja SDE.

Jaga sisseloetud andmestik data kaheks alamandmestikuks data1 ja data2. Esimene neist sisaldagu hääletusi kuni 2014-03-26. Andmestik data2 sisaldagu hääletusi alates kuupäevast 2014-03-26.

```
#grepl("2014.03.26",names(data_order))
data1=data_order[, 1:1526]
data2=data_order[, c(1,2, 1527:1847)]
```

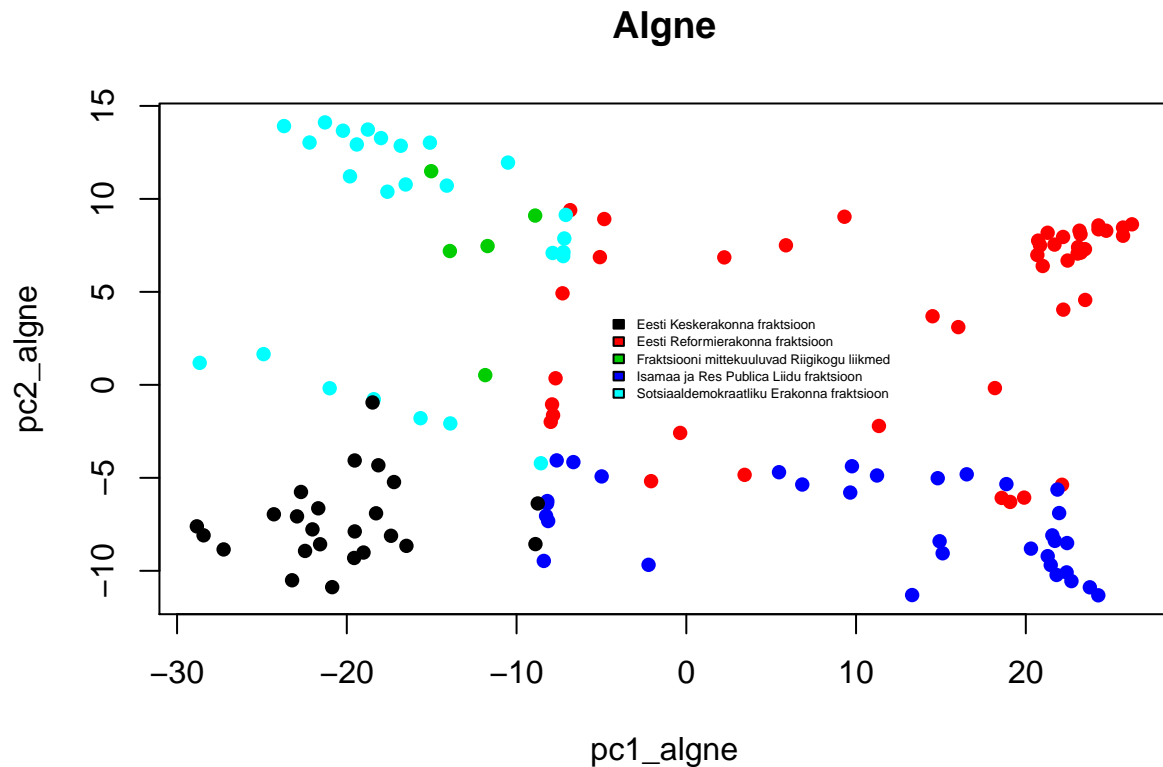
Ülesanne 4 (3 punkti)

Tee PCA esialgsele andmestikule. Visualiseeri kahte esimest peakomponenti hajuvusdiagrammi abil. Tõlgenda tulemusi.

```
pca_algne=prcomp(data[, -(1:2)])
pc1_algne = pca_algne$x[, 1]
```

```
pc2_algne = pca_algne$x[, 2]

plot(pc1_algne, pc2_algne, pch=16, col=as.factor(data$fraktsioon))
title("Algne")
legend("center", "groups", legend = unique(data2$fraktsioon),
      text.width = 5, cex=0.45, ncol=1,
      fill = 1:5, bty = "n")
```



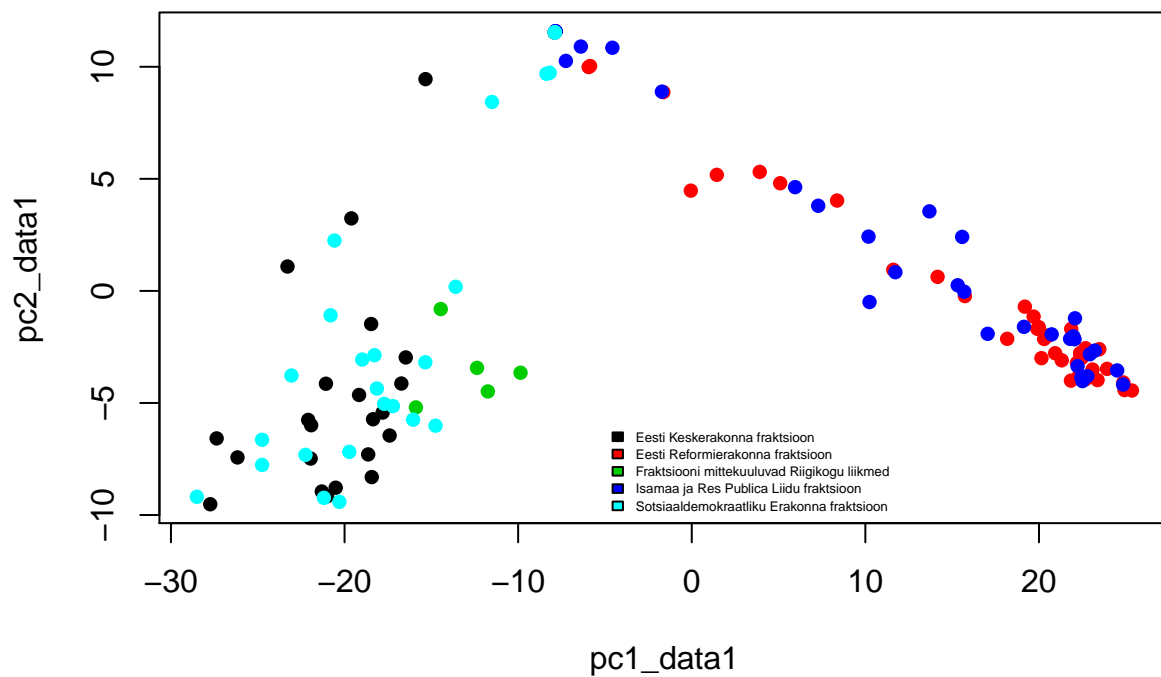
- Tõlgendus: PC1 jagab erakonnad enam-vähem koalitsiooniks ja opositsiooniks. PC2 tundub näitavat, erakondade võimalikku valitsusse kuulumist (koostööd).

Tee PCA nii andmestikule data1 kui ka data2. Mõlemal juhul visualiseeri kahte esimest peakomponenti hajuvusdiagrammi abil. Tõlgenda tulemusi.

```
#data1
pca_data1=prcomp(data1[, -(1:2)])

pc1_data1 = pca_data1$x[, 1]
pc2_data1 = pca_data1$x[, 2]
plot(pc1_data1, pc2_data1, pch=16, col=as.factor(data1$fraktsioon))

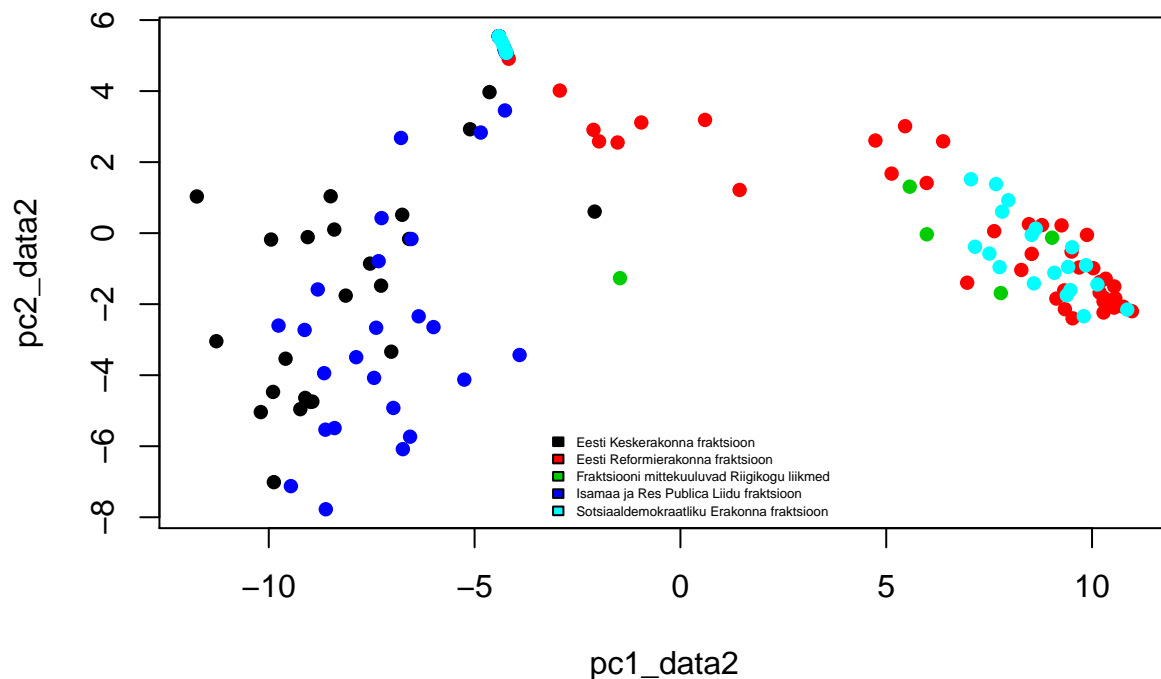
legend("bottom", "groups", legend = unique(data2$fraktsioon),
      text.width = 5, cex=0.45, ncol=1,
      fill = 1:5, bty = "n")
```



```
#data2
pca_data2=prcomp(data2[, -(1:2)])

pc1_data2 = pca_data2$x[, 1]
pc2_data2 = pca_data2$x[, 2]
plot(pc1_data2, pc2_data2, pch=16, col=as.factor(data2$fraktsioon))

legend("bottom","groups", legend = unique(data2$fraktsioon),
      text.width = 5,cex=0.45, ncol=1,
      fill = 1:5, bty = "n")
```



- Tõlgendus: jooniselt on hästi näha, kuidas toimus koalitsioonivahetus, REFist sai keskerakonnaga sarnasem partei ning SDE lähenes reformile.

Ülesanne 5 (3 punkt)

Eelmises ülesandes rakendasid PCA-d andmestikule data1 ja visualiseerisid seda kasutades kahte esimest komponenti. Kas piisab ka ühest komponendist?

```
#praegu ei jookusta läbi, liiga pikk tabel
summary(pca_data1)
```

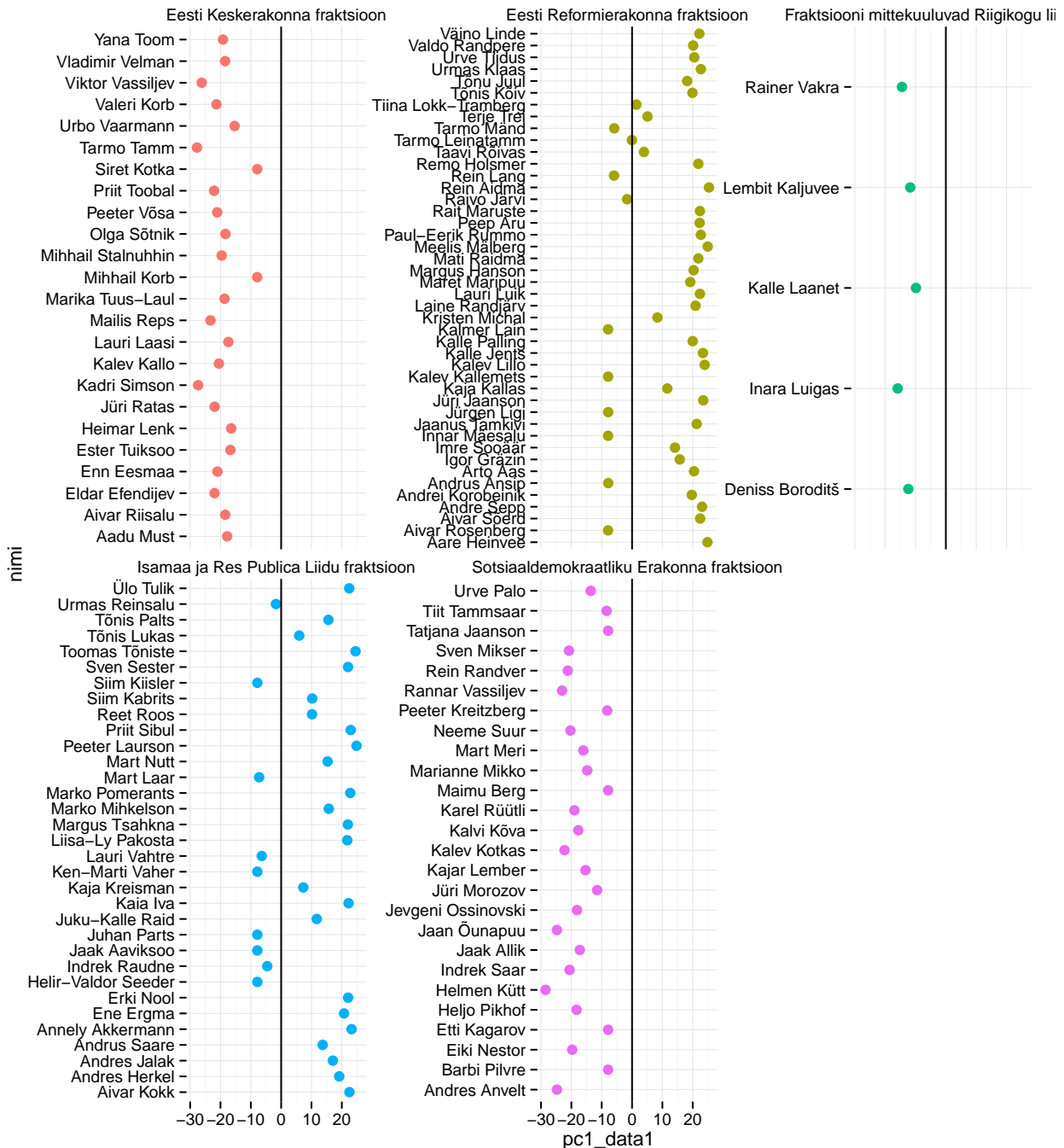
- Vastus: pigem mitte, kuna PC1 selgitab umbes 55% variatsioonist ning pole seega väga täpne

Visualiseeri data1 esimest peakomponenti selliselt, et näha oleksid ka saadikute nimed. Kas esimest peakomponenti võib tõlgendada kui vasak- ja parempoolsust?

```
vahetabel=as.data.frame(pc1_data1)
vahetabel$nimi=data_order$nimi
vahetabel$fraktsioon=data_order$fraktsioon

library(ggplot2)
ggplot(vahetabel, aes(x=pc1_data1, colour=fraktsioon))+
  geom_point(aes(y=nimi),size=3)+
```

```
geom_vline(xintercept = 0)+
facet_wrap(~fraktsioon,scales="free_y")+
guides(color=FALSE)+
theme_minimal()
```



- Vastus: Tundub, et jagab vasak-parem poolseks küll parteisid.