

# Eduskuntavaalien listojen jakaumia

Tässä dokumentissa analysoidaan eduskuntavaalien ehdokkaiden saamia äänimääriä. Tarkoituksena on arvioida, noudattavatko äänimäärät suunnilleen eksponenttijakaumaa.

Analysoidaan vuoden 2007 ja 2015 äänestystuloksia (Ainoat täydet setit jotka saatavilla).

```
library(dplyr)

##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##   filter, lag
##
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(ggplot2)
library(stringr)
```

Analyze data

```
minListSize <- 10

fullData <- as_tibble(read.csv("CleanData.csv"))
```

## Suhteelliset osuudet

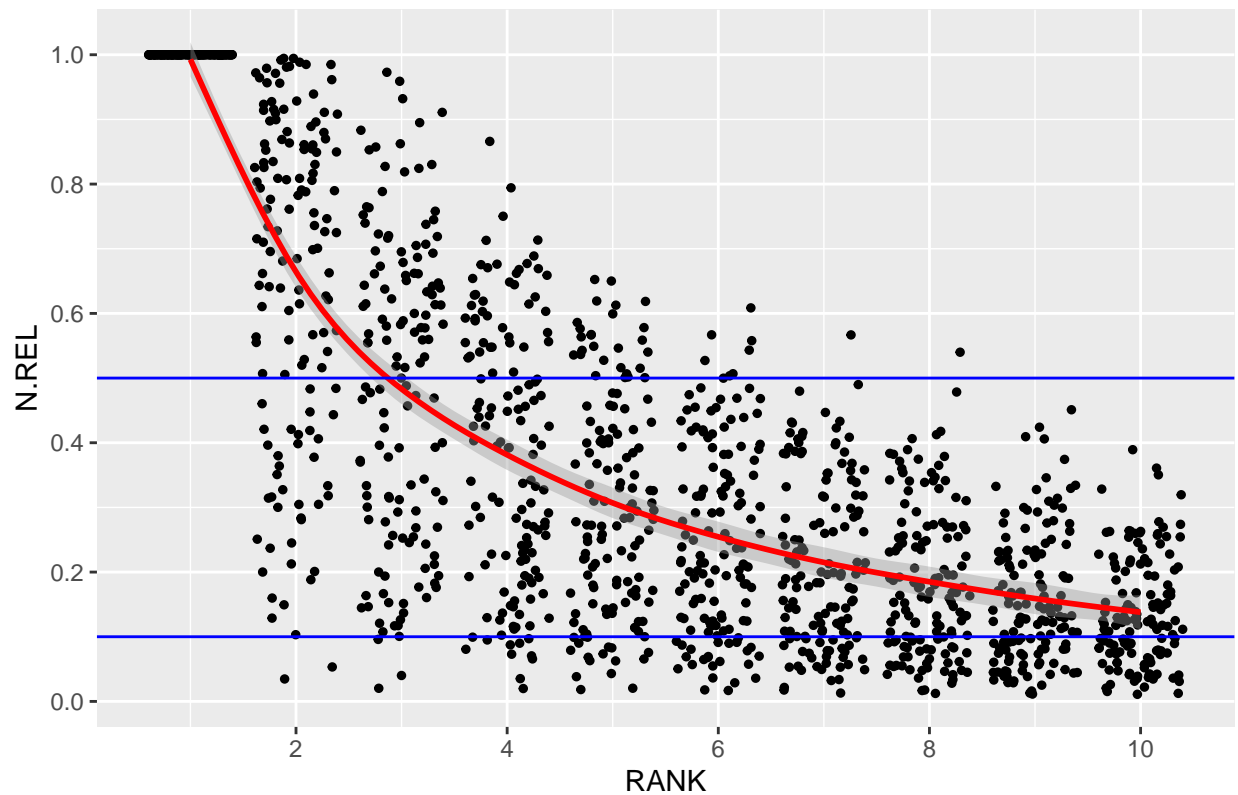
Plotataan kaikkien vaalipiirien kaikkien listojen edustajien äänimäärät niin, että eniten ääniä saaneen ehdokkaan suhteelliseksi äänimääräksi normitetaan 1. Nähdään, että eksponenttijakauma ei ole ainakaan huono arvio.

```
g1 <- ggplot(fullData %>% filter(RANK<=minListSize,LISTPASSED>0))+
  geom_jitter(aes(x=RANK,y=N.REL,group=interaction(YEAR,PARTY,PROVINCE)),col="black",size=1)+
  scale_x_continuous(breaks=seq(0,minListSize,2))+
  scale_y_continuous(breaks=seq(0,1,0.2))+
  geom_smooth(aes(x=RANK,y=N.REL),col="red")+
  geom_hline(yintercept = c(0.1,0.5),col="blue")+
  ggtitle("Listalta ainakin yksi kansanedustaja")

plot(g1)

## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

## Listalta ainakin yksi kansanedustaja



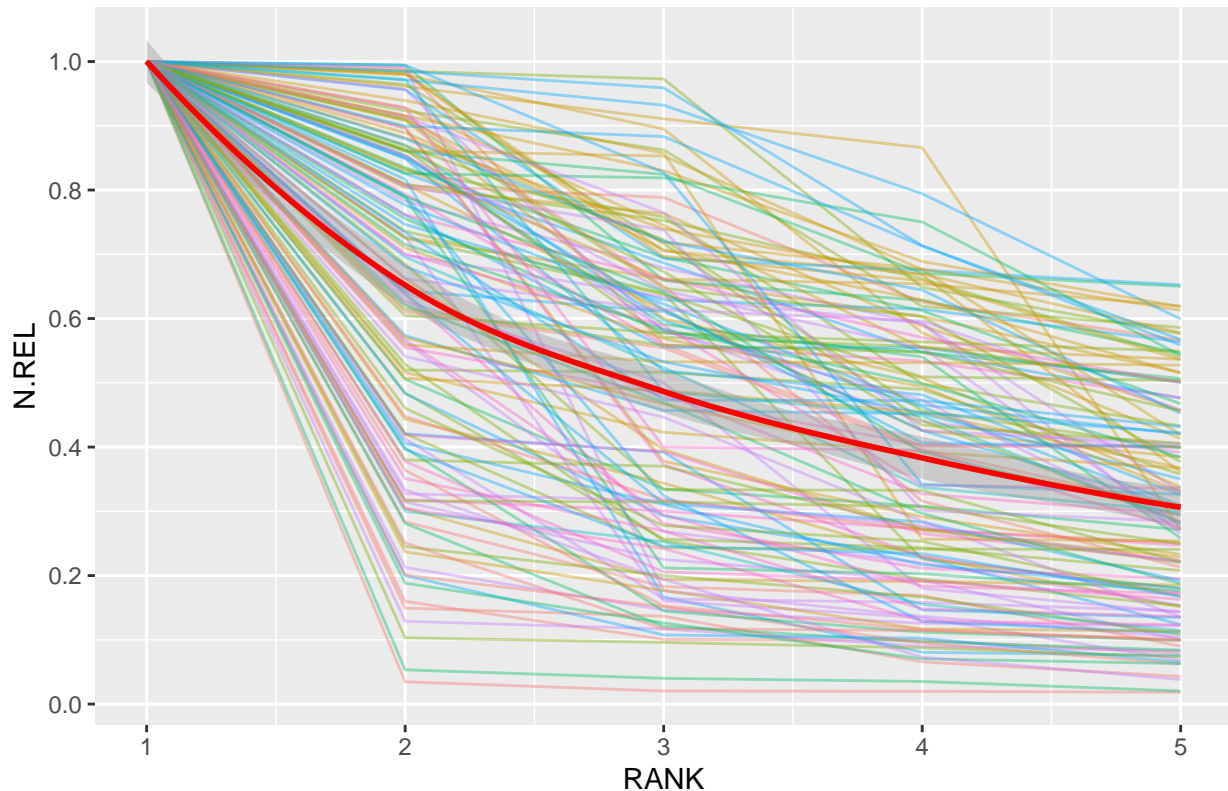
## Puolueittain

Koitetaan arvioida, onko eri puolueilla systemaattisesti erilaisia jakaumia (ts onko jokin puolue systemaattisesti liikkeellä yhdellä kärkiehdokkaalla). Systemaattista eroa ei ainakaan silmin pysty näkemään.

```
g2 <- ggplot(fullData %>% filter(RANK<=5,LISTPASSED>0))+  
  geom_line(aes(x=RANK,y=N.REL,group=interaction(YEAR,PARTY,PROVINCE),col=PARTY),size=0.5,alpha=0.4)+  
  scale_x_continuous(breaks=seq(0,12,1))+  
  scale_y_continuous(breaks=seq(0,1,0.2))+  
  geom_smooth(aes(x=RANK,y=N.REL),col="red")+  
  theme(legend.position="none")+  
  ggtitle("Ainakin yksi kansanedustaja, puolueet värikoodattu")  
  
plot(g2)
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

## Ainakin yksi kansanedustaja, puolueet värikoodattu



## Normitettu data: Yksi vs useampi edustaja

Arvioidaan, onko usean läpipäässeeseen ehdokkaan listoilla myös vähemmän saaneiden ehdokkaiden jakauma tasaisempi. Listojen koot ovat eri puolilla maata eri pituisia, joten ehdokkaan sijainti listalla kannattaa normittaa listan pituuteen.

Tällöin nähdään, että ei ole juurikaan eroa, onko listalla ollut yksi läpimenijä vai useampia.

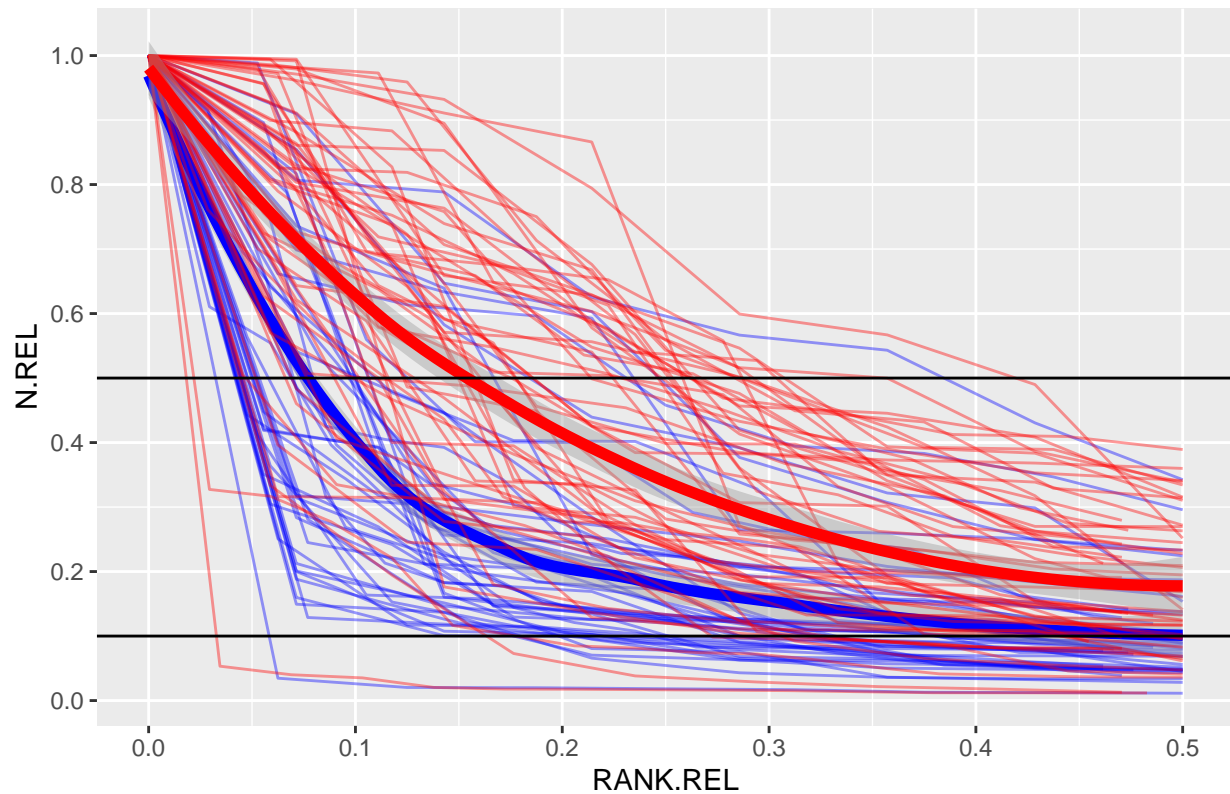
```
closeup <- fullData %>% filter(RANK.REL<=0.5, LISTPASSED > 0)

g3 <- ggplot(closeup)+
  geom_line(data=closeup %>% filter(LISTPASSED == 1), aes(x=RANK.REL,y=N.REL,group=interaction(YEAR,PARTY))) +
  geom_smooth(data=closeup %>% filter(LISTPASSED == 1), aes(x=RANK.REL,y=N.REL),col="blue",size=2)+
  geom_line(data=closeup %>% filter(LISTPASSED > 1 & LISTPASSED <4), aes(x=RANK.REL,y=N.REL,group=interaction(YEAR,PARTY))) +
  geom_smooth(data=closeup %>% filter(LISTPASSED > 1 & LISTPASSED <4), aes(x=RANK.REL,y=N.REL),col="red",size=2)+
  scale_x_continuous(breaks=seq(0,1,0.1))+
  scale_y_continuous(breaks=seq(0,1,0.2))+
  # scale_color_continuous(low="blue",high="red")+
  # geom_smooth(aes(x=RANK,y=N.REL),col="red")+
  theme(legend.position="none")+
  geom_hline(yintercept = c(0.1,0.5),col="black")+
  ggtitle("Listalta 1 kansanedustaja (sin) tai 2-3 (pun)")

plot(g3)
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

### Listalta 1 kansanedustaja (sin) tai 2–3 (pun)



## Normitettu data

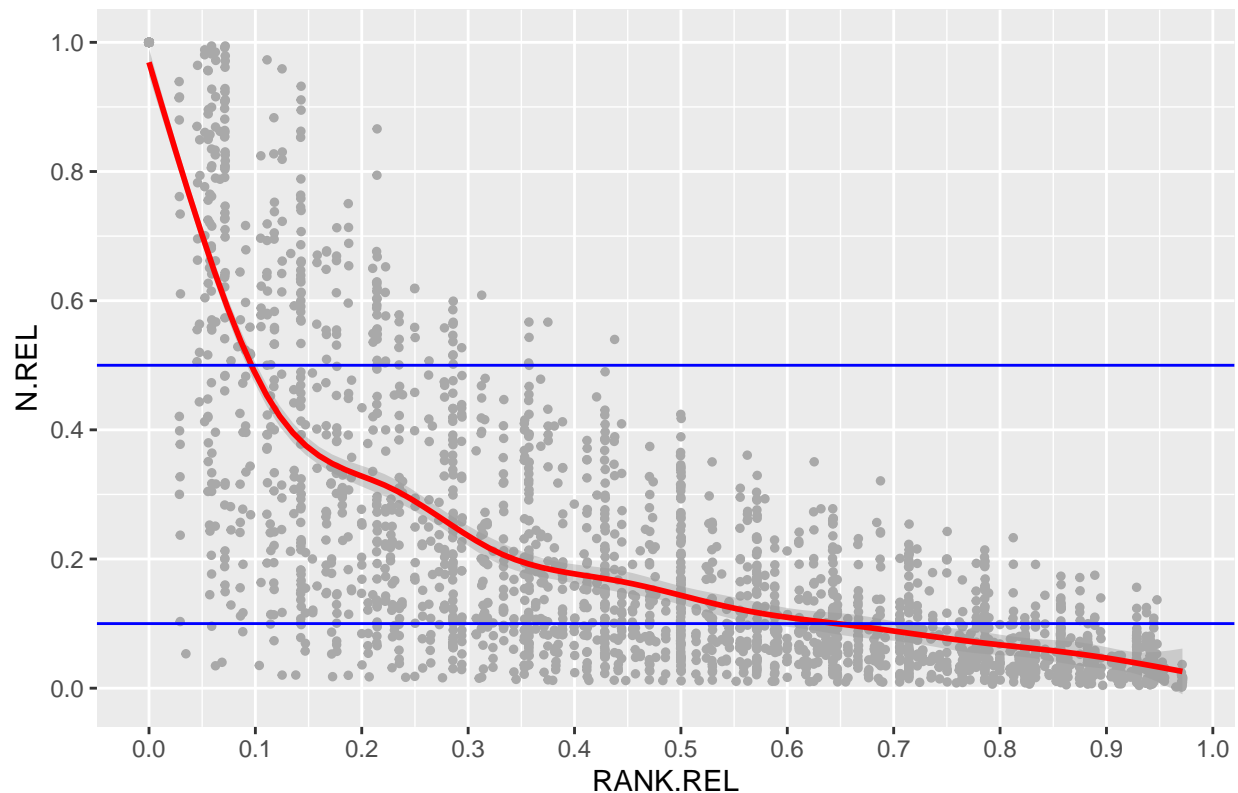
Kun normitetaan sekä sijainti listalla että suhteellinen äänimäärä, nähdään erittäin systemaattinen tulos. Listan suosituin 10% saa lähes aina noin puolet kaikista listan äänistä. Listan viimeisessä kolmanneksessa äänimäärät ovat käytännössä aina alle 10% kärkiehdokkaan äänistä.

```
g4 <- ggplot(fullData %>% filter(RANK.REL<=1,LISTPASSED>0))+
  geom_jitter(aes(x=RANK.REL,y=N.REL,group=interaction(YEAR,PARTY,PROVINCE)),col="darkgray",size=1)+
  scale_x_continuous(breaks=seq(0,minListSize,0.1))+
  scale_y_continuous(breaks=seq(0,1,0.2))+
  geom_smooth(aes(x=RANK.REL,y=N.REL),col="red")+
  geom_hline(yintercept = c(0.1,0.5),col="blue")+
  ggtitle("Listalta ainakin yksi kansanedustaja")

plot(g4)
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

## Listalta ainakin yksi kansanedustaja



## Normitettu tulos: useita ehdokkaita vai yksi ehdokas?

Jos plotataan erikseen tapaukset, joissa listalta pääsi läpi alle ja yli 10% listalla olijoista, ei nähdä käytännössä merkittävää eroa.

```
closeup <- fullData %>% filter(RANK.REL<=0.33, LISTPASSED > 0)

lowerLimit <- 0.1
midLimit <- 0.10
upperLimit <- 0.20

g5 <- ggplot(closeup)+
  geom_line(data=closeup %>% filter(PASSED.REL < lowerLimit), aes(x=RANK.REL,y=N.REL,group=interaction(
    geom_smooth(data=closeup %>% filter(PASSED.REL < lowerLimit), aes(x=RANK.REL,y=N.REL),col="blue",si
  geom_line(data=closeup %>% filter(PASSED.REL > midLimit & PASSED.REL <= upperLimit), aes(x=RANK.REL,y
    geom_smooth(data=closeup %>% filter(PASSED.REL > midLimit & PASSED.REL <= upperLimit),aes(x=RANK.R

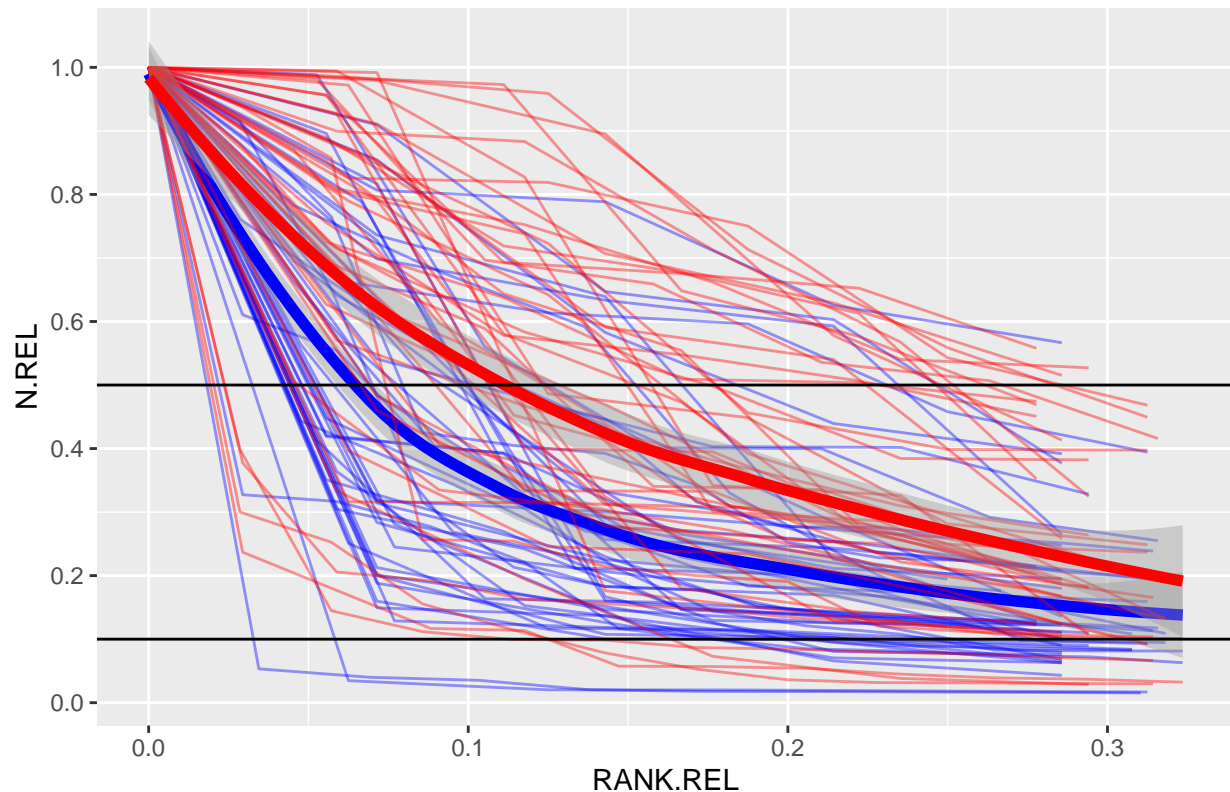
  scale_x_continuous(breaks=seq(0,1,0.1))+
  scale_y_continuous(breaks=seq(0,1,0.2))+
  # scale_color_continuous(low="blue",high="red")+
  # geom_smooth(aes(x=RANK,y=N.REL),col="red")+
  theme(legend.position="none")+
```

```
geom_hline(yintercept = c(0.1,0.5),col="black")+
ggtitle("Listasta l?pi < 10% (sin) tai 10-20% (pun)")
```

```
plot(g5)
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

Listasta l?pi < 10% (sin) tai 10-20% (pun)



## Varsinais-Suomen jakaumat

Piirretään Varsinais-Suomen ehdokkaiden äänimäärät kaksissa vaaleissa, ja verrataan sen vuoden äänikynnykseen (pienimpään listan kokonaisäänimäärään jolla ehdokas saatiin läpi).

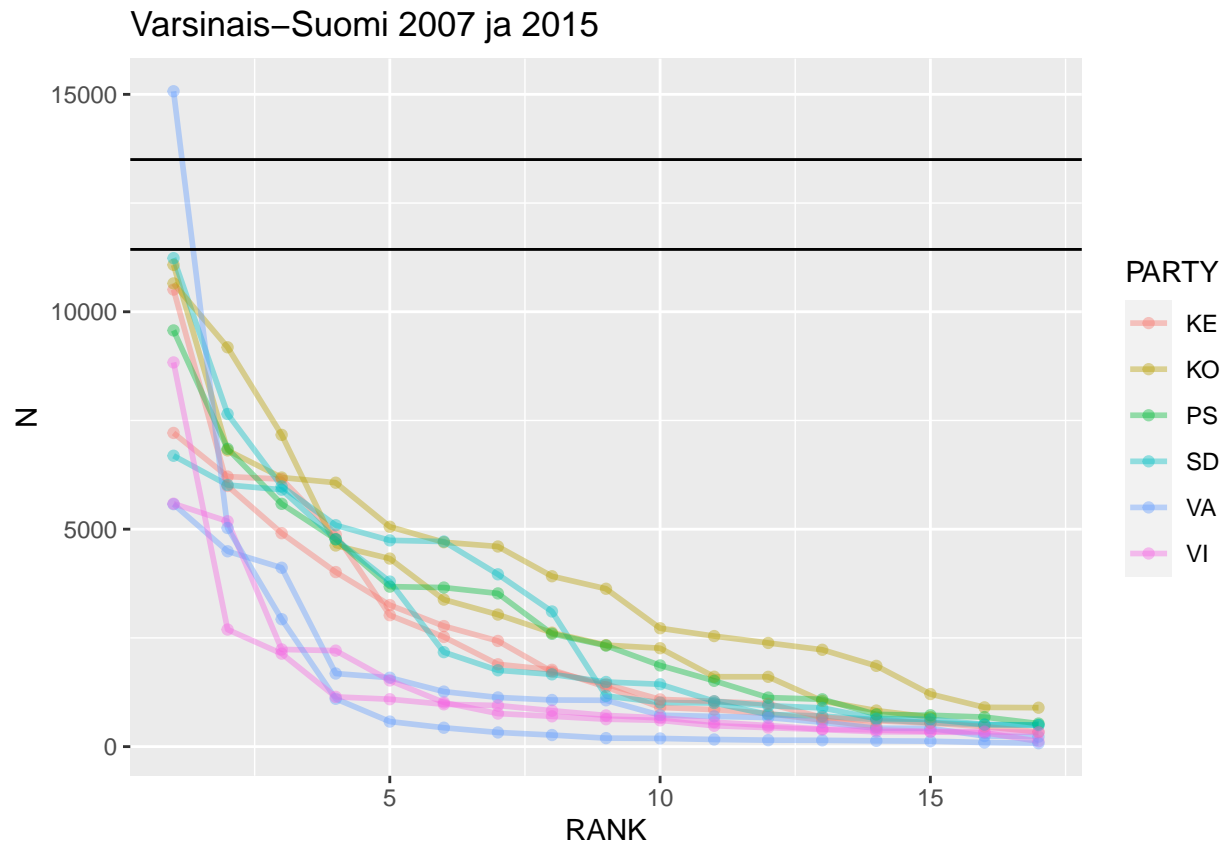
Huomataan, että vain yhdessä tapauksessa (VAS 2015) kärkiehdokas on yksinään päässyt yli äänikynnyksen. Erityisesti KOK ja PS ovat kuitenkin saaneet kärkiehdokkaita, jotka ovat päässeet yksinään hyvin lähelle äänikynnystä. Toisessa ääripäässä Erityisesti SDP ja VIHR ovat puolueita, joissa kärkiehdokas on tyypillisesti jäänyt kauas kynnyksestä, ja läpimeno on vaadittu paljon muitakin listan ehdokkaita.

```
Vsuomi <- filter(fullData, PROVINCE=="VAR" & LISTPASSED > 0)
```

```
g10 <- ggplot(Vsuomi)+
  geom_line(aes(x=RANK,y=N,group=interaction(PARTY, YEAR),col=PARTY),size=1,alpha=0.4)+
  geom_point(aes(x=RANK,y=N,group=interaction(PARTY, YEAR),col=PARTY),size=1.5,alpha=0.4)+
  # scale_x_continuous(breaks=seq(0,12,1))+
  # scale_y_continuous(breaks=seq(0,1,0.2))+
```

```
# geom_smooth(aes(x=RANK,y=N),col="red")+
# theme(legend.position="none")+
geom_hline(yintercept=unique(Vsuomi$THRESHOLD))+
ggtitle("Varsinais-Suomi 2007 ja 2015")

plot(g10)
```



```
#Usually not needed
if (F) { sessionInfo() }
```