

SKP

Risto Hinn

Friday, June 26, 2015

Sissejuhatus

Kati “Makro” Ökonoomika on tavaline Eesti naine, kes saab kõigega hakkama: ta kasvatab üksi last, maksab kogupereleenu ning töötab Euroopa Komisjoni struktuuriüksuses ametnikuna. 2013. aasta alguses andis Olli Rehn, üks toonastest 28-st Euroopa Komisjoni volinikust, talle ülesandeks uurida võlakoorma mõju majanduskasvule.

Kati teadis, et hea uurimus peab põhinema andmetel mitte ekspertide kõhutundel. Peagi leidis ta artikli *Growth in a Time of Debt*, mille põhitulemuseks oli, et kõrgem riigivõlg on seotud madalama majanduskasvuga ning väga probemaatiline on riigivõlg, mis on üle 90% SKP-st. Artikkel tundus usaldusväärne, sest artikli autoriteks on tunnustatud majandusteadlased Harvardist, artiklit oli tsiteeritud sel hetkel juba üle 500 korra ning see põhines 50-aastaselt perioodil.

Juba 9. aprillil 2013 kasutas Olli Rehn oma kõnes Rahvusvahelisele Tööorganisatsioonile Kati leitud:

public debt in Europe is expected to stabilise only by 2014 and to do so at above 90% of GDP. Serious empirical research has shown that at such high levels, public debt acts as a permanent drag on growth.

Kati tundis, et ta oma tööd hasti teinud ja unistas aastalõpu boonusest. Tema sisemine rahulolu jäi aga üürikeseks. Majandusteadlased (Herndon, Ash, Pollin, edasises HAP) ülikoolist UMass Amherst avaldasid teadusartikli, kus väitsid, et Reinhart-Rogoffi (edasises RR) meetodikas on tõsiseid vigu sees, alustades Exceli arvutusveast, andmete väljajätmisest ja valest kaalumismeetodist. Majandusteemalised ajalehed olid täis kriitilisi kommentaare, sest Exceli viga on ju hoomatav kõigile.

Moodustati kriisikomisjon, mida pandi juhtima Kati. Vaja on teha suurele hulgale poliitikutele võimalikult lihtsasti selgeks, mille vastu eksisid Reinhart-Rogoff ja kui palju esialgsed tulemused muutusid.

Tutvumine andmestikuga

20 arenenud riigi kohta on teada SKP kasvunumber ja võlakoorma-SKP suhe aastatel 1946-2009. Andmestikus on järgnevad tunnused:

- riik: mis riigi kohta näitajad on toodud
- aasta: mis aasta kohta käivad SKP kasv ja võlakoorma-SKP suhe
- vola_skp_suhe: võlakoorma ja SKP suhe
- skp_kasv: SKP kasvunumber
- exceli_viga: binaarne tunnus, mis näitab, kas andmepunkt jäi RR analüüsist välja Exceli arvutusvea tõttu (kui on 1, siis jäi välja)
- valikuline: binaarne tunnus, mis näitab, kas andmepunkt jäi RR analüüsist välja (HAP väitel selekteerimise tõttu, RR väitel, et neid andmeid analüüsi tegemise ajal polnud. Kui on 1, siis jäi RR analüüsist välja.)

Ülesanne 1 (2 punkti) - RR ja HAP tulemuste reprodutseerimine

Kontrolli, kas suudad reprodutseerida RR ja HAP tulemused.

Juhised:

- Lisa uus tunnus, mis näitab vola_skp_suhe kategooriat (< 30%, 30-60%, 60-90%, > 90%). Näpunäide: kasuks tuleb käsk cut.
- kaalumiskiisid. HAP kaalub igas kategoorias andmepunkte võrdselt (ehk võtab tavalise aritmeetilise keskmise), RR arvutab igas grupis riikide keskmise ja võtab neist aritmeetilise keskmise.
- RR-i tulemuste reprodutseerimiseks jäta arvutustest välja andmepunktid, mis jäid välja Exceli vea tõttu (vt tunnus exceli_viga) ja andmete puudumise tõttu (vt tunnus valikuline).
- RR tegid ka kopeerimisvea: kopeerides riikide keskmisi ühest Exceli tabelist teise, muutus Uus-Meremaa keskmine SKP tõus grupis "> 90%" väärtuselt -7.6 väärtuseks -7.9. (Näpunäide: kasuks tuleb käsk ifelse.)

```
library(dplyr)
andmed=read.csv("./data/skp_ja_volg.csv")
#lisame tunnuse, mis näitab vola_skp_suhe kategooriat
miinimum=min(andmed$vola_skp_suhe, na.rm=T)
maksimum=max(andmed$vola_skp_suhe, na.rm=T)
andmed$vola_skp_suhe_klass=cut(andmed$vola_skp_suhe,
                               breaks=c(miinimum,30,60,90, maksimum), include.lowest=T,
                               labels=c("< 30%", "30-60%", "60-90%", "> 90%"))
andmed_grupp= group_by(andmed, vola_skp_suhe_klass)
#HAPi kaalud
HAP_tulem=andmed %>%
  group_by(vola_skp_suhe_klass)%>%
  summarise(HAP_mean = round(mean(skp_kasv, na.rm=T),1),
            HAP_median=round(median(skp_kasv, na.rm=T),1))
#RR-i reprodutseerimine
RR_tulem=andmed_grupp%>%
  subset( exceli_viga==0)%>%
  subset(valikuline==0)%>%
  group_by(vola_skp_suhe_klass,riik)%>%
  summarise(SKp_keskmise_kasv= round(mean(skp_kasv, na.rm=T),1),
            SKp_mediaan_kasv=round(median(skp_kasv, na.rm=T),1))%>%
  group_by(vola_skp_suhe_klass)%>%
  summarise(RR_mean=round(mean(ifelse(
    riik=="New Zealand" & SKp_keskmise_kasv== -7.6, -7.9, SKp_keskmise_kasv), na.rm=T),1),
            RR_median=round(median(SKp_mediaan_kasv),1))
#tulemid kokku
tulem=cbind(RR_tulem[1:4,], HAP_tulem[1:4,])
head(tulem)
```

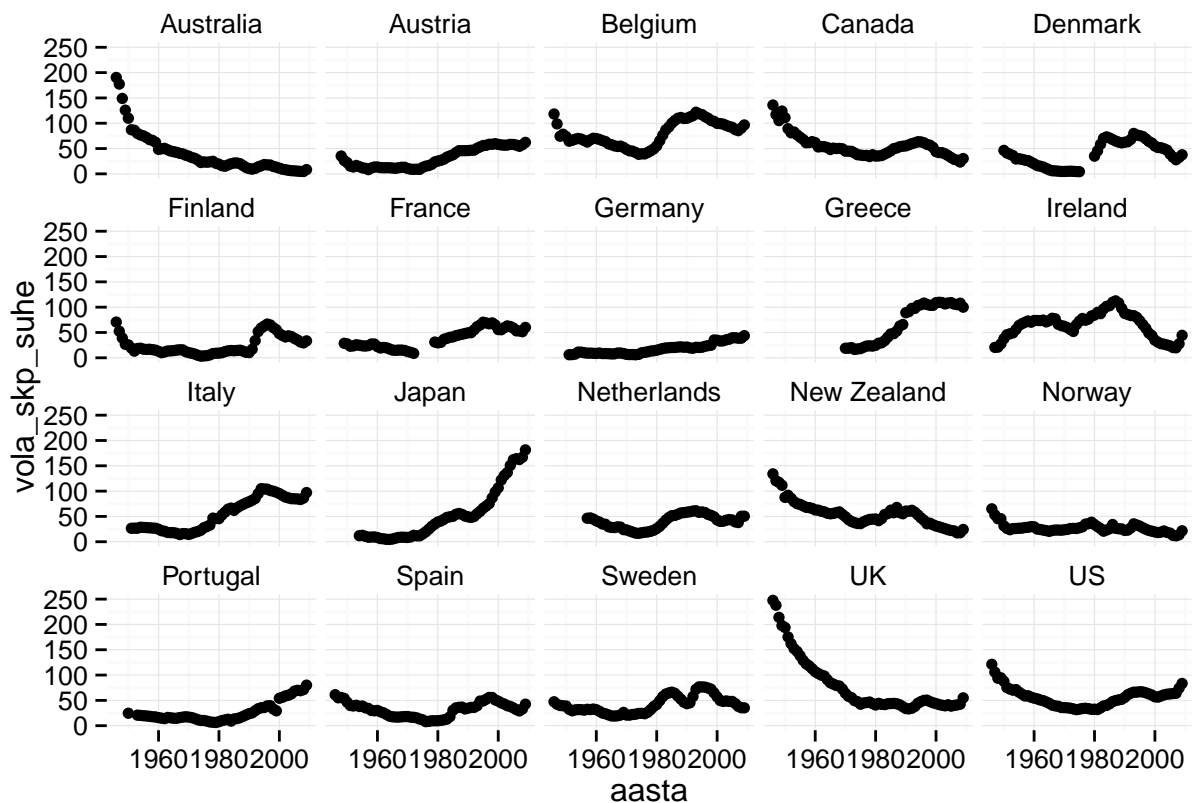
```
##   vola_skp_suhe_klass RR_mean RR_median vola_skp_suhe_klass HAP_mean
## 1                < 30%    4.1      4.2                < 30%    4.2
## 2               30-60%    2.9      3.0               30-60%    3.1
## 3               60-90%    3.4      3.1               60-90%    3.2
## 4                > 90%   -0.1      1.6                > 90%    2.2
##   HAP_median
## 1          4.1
```

```
## 2      3.1
## 3      2.9
## 4      2.3
```

Ülesanne 2 (1 punkt) - visualiseeri võlakooormuse muutumist ajas

Visualiseeri, kuidas võlakooormus on aastate jooksul muutunud riikide lõikes

```
library(ggplot2)
ggplot(andmed, aes(x=aasta, y=vola_skp_suhe))+
  geom_point()+
  facet_wrap(~riik)+
  theme_minimal()
```



Kuidas efektiivselt visualiseerida ...

Järgnevalt püüame leida parima viisi, kuidas efektiivselt visualiseerida

- millised vaatlused jäid RR analüüsist välja
- kuivõrd erinesid RR ja HAP analüüside tulemused
- kas võlakooormus suurem kui 90% on maagilise tähtsusega (st kas piir on just täpselt 90%)
- milline on seos SKP ja võlakooormuse vahel

Ülesanne 3 (1 punkt) - millised vaatlused jäid RR analüüsist välja

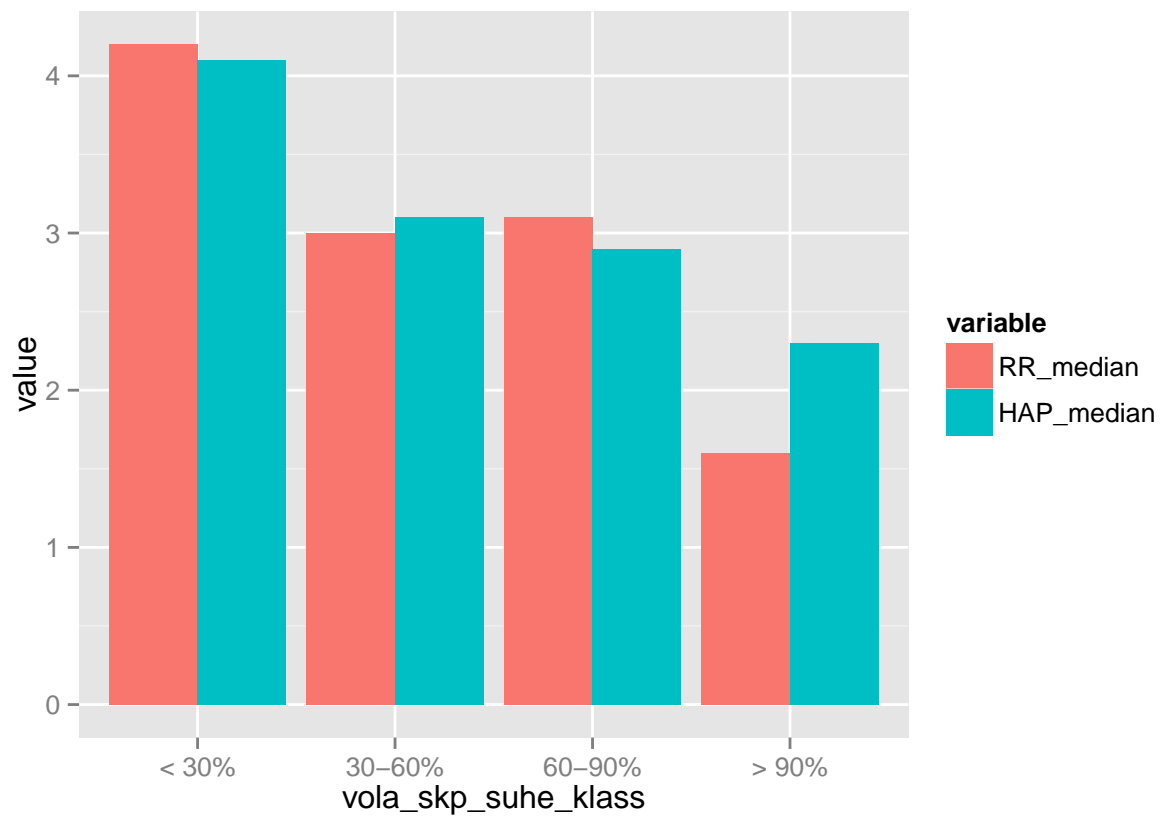
```
pole=subset(andmed, exceli_viga==1 | valikuline==1)
head(pole)
```

```
##      riik aasta vola_skp_suhe  skp_kasv valikuline exceli_viga
## 1 Australia 1946    190.41908 -3.557951         1         1
## 2 Australia 1947    177.32137  2.459475         1         1
## 3 Australia 1948    148.92981  6.437534         1         1
## 4 Australia 1949    125.82870  6.611994         1         1
## 5 Australia 1950    109.80940  6.920201         1         1
## 6 Australia 1951     87.09448  4.272612         0         1
##  vola_skp_suhe_klass
## 1                > 90%
## 2                > 90%
## 3                > 90%
## 4                > 90%
## 5                > 90%
## 6            60-90%
```

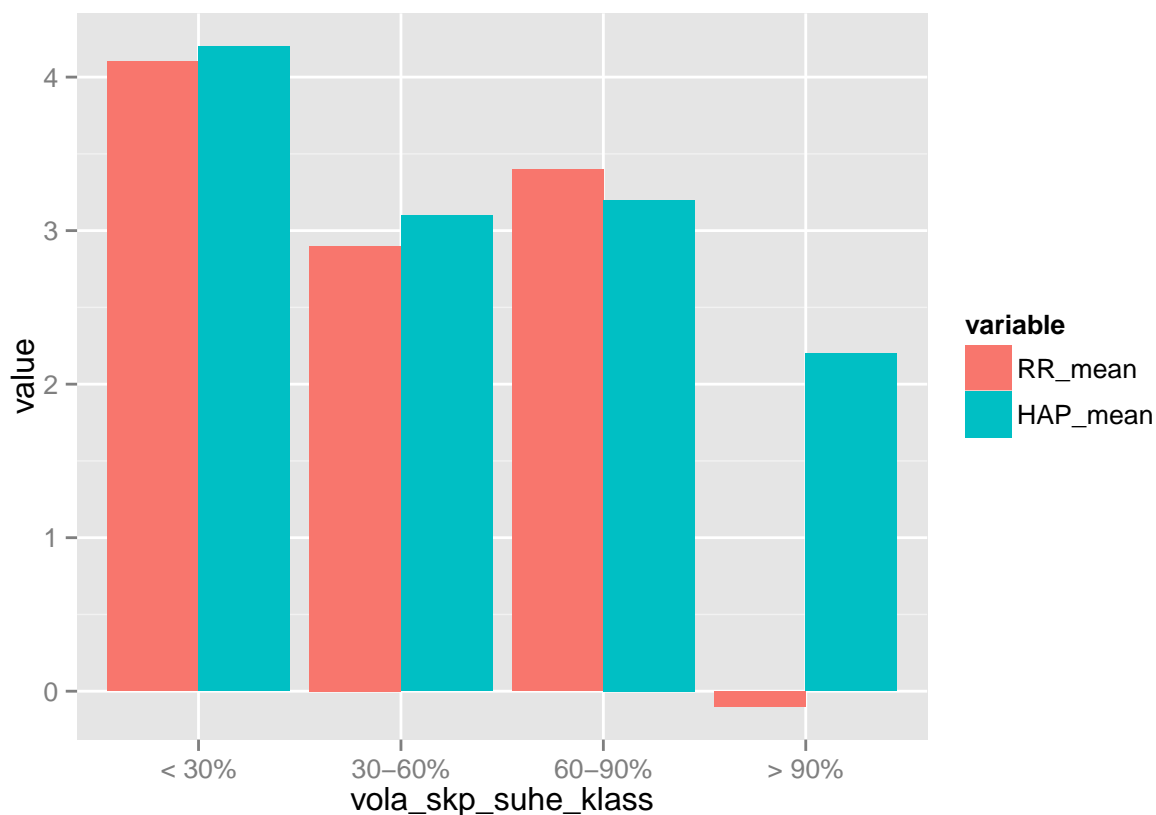
Boonusülesanne (2 punkti) - kuidas erinesid RR ja HAP analüüside tulemused

```
library(reshape2)
tulem_melt=melt(tulem, id=c("vola_skp_suhe_klass"))
tulem_melt_median=tulem_melt[grepl("median", tulem_melt$variable),]
tulem_melt_mean=tulem_melt[grepl("mean", tulem_melt$variable),]

ggplot(tulem_melt_median, aes(x=vola_skp_suhe_klass, y=value, fill=variable)) +
  geom_bar(position="dodge", stat="identity")
```



```
ggplot(tulem_melt_mean, aes(x=vola_skp_suhe_klass, y=value, fill=variable)) +  
  geom_bar(position="dodge", stat="identity")
```



Ülesanne 4 (2 punkti) - kas võlakoormus suurem kui 90% on maagilise tähtsusega

```
andmed2=andmed
andmed2$vola_skp_suhe_klass=cut(andmed2$vola_skp_suhe,
                               breaks=c(miinimum,30,60,90,120, maksimum), include.lowest=T,
                               labels=c("< 30%", "30-60%", "60-90%", "90-120%", "> 120%"))

andmed_grupp2= group_by(andmed2, vola_skp_suhe_klass)

RR_tulem2=andmed_grupp2%>%
  subset( exceli_viga==0)%>%
  subset(valikuline==0)%>%
  group_by(vola_skp_suhe_klass,riik)%>%
  summarise(SKP_keskmine_kasv= round(mean(skp_kasv, na.rm=T),1),
            SKP_mediaan_kasv=round(median(skp_kasv, na.rm=T),1))%>%
  group_by(vola_skp_suhe_klass)%>%
  summarise(RR_mean=round(mean(ifelse(
    riik=="New Zealand" & SKP_keskmine_kasv== -7.6,-7.9,SKP_keskmine_kasv),na.rm=T),1),
            RR_median=round(median(SKP_mediaan_kasv),1))

RR_tulem2

## Source: local data frame [6 x 3]
##
##   vola_skp_suhe_klass RR_mean RR_median
```

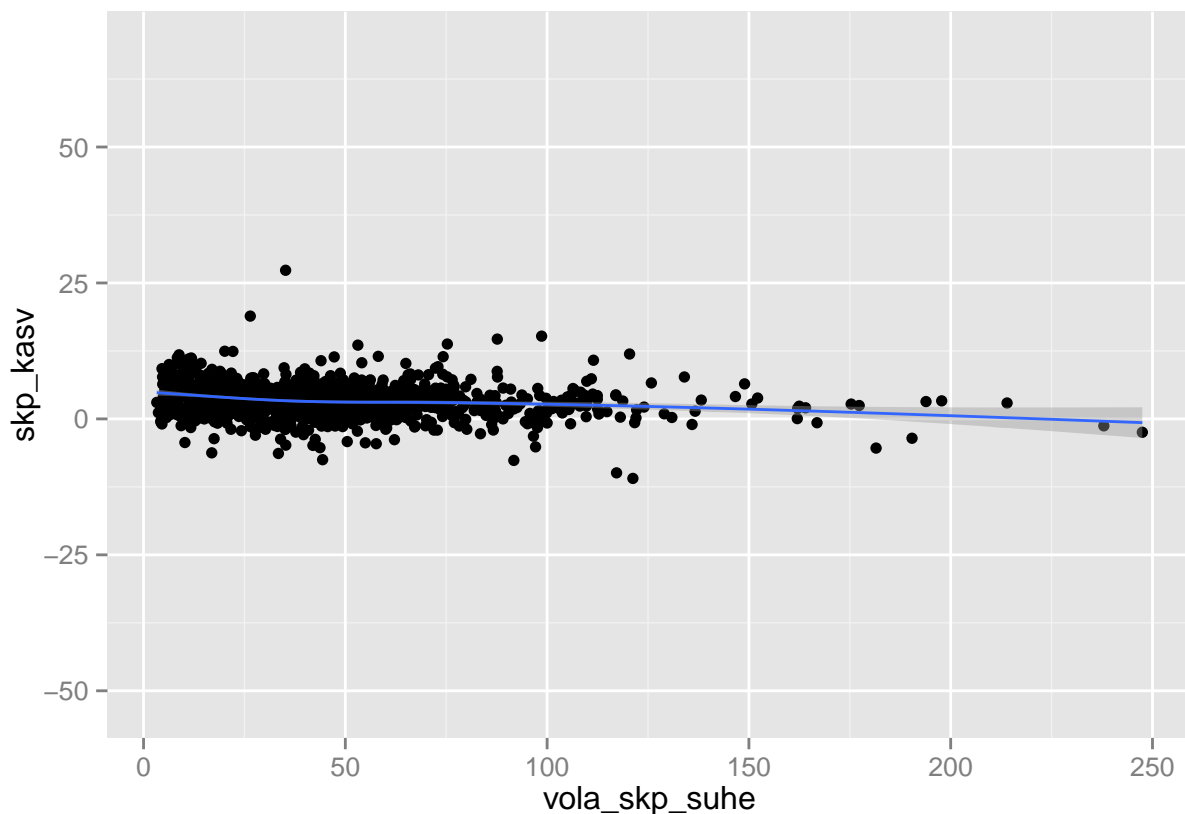
| | | | |
|------|---------|------|-----|
| ## 1 | < 30% | 4.1 | 4.2 |
| ## 2 | 30-60% | 2.9 | 3.0 |
| ## 3 | 60-90% | 3.4 | 3.1 |
| ## 4 | 90-120% | 0.6 | 1.6 |
| ## 5 | > 120% | -2.8 | 1.4 |
| ## 6 | NA | 5.5 | NA |

Ülesanne 5 (4 punkti) - Kuidas visuaalselt uurida, milline on seos SKP ja võlakoormuse vahel?

Kõigepealt, tehke joonis, kus oleks näha seos SKP ja võlakoormuse vahel. Seose iseloomustamiseks võite kasutada `stat_smooth()` abil leitavat joont.

Näinud seost andmestikus, tekib küsimus, ega see seos ei ole lihtsalt juhuslik. Ehk kas vaadeldud seos erineb oluliselt seostest sellistes andmestikes, kus tegelikult SKP ja võlakoormuse vahel mingisugust seost ei eksisteeri.

```
ggplot(andmed, aes(x=vola_skp_suhe, y=skp_kasv))+
  geom_point()+
  stat_smooth()
```



Selle visuaalseks kontrollimiseks võime kasutada järgmist skeemi. See põhineb permutatsioonitestil, mille kohta vaadake esmalt kodutöö osa III [videot](#)

Skeem:

- Nullhüpotees on, et SKP ja riigivõla vahel seos puudub.
- Genereerime meie andmetest permuteerimise teel sellise andmestiku, mis vastab nullhüpoteesile. Näiteks võib fikseerida SKP väärtused ning neile vastavusse seatavad riigivõla väärtused permuteerida. (Näpunäide: permuteerimisel on kasuks funktsioon sample.)

```
#jätame alles read, kus pole NAd SKP kasvus ja võla SKP suhtes
andmed_permut=andmed[!is.na(andmed$vola_skp_suhe)&!is.na(andmed$skp_kasv),]

#permuteerime vola_skp_suhe väärtuseid
andmed_permut$iter1=sample(andmed_permut$vola_skp_suhe, replace=T)

plot_permut=ggplot(andmed_permut, aes(x=vola_skp_suhe, y=skp_kasv))+
  geom_point()+
  stat_smooth(aes(x=iter1, y=skp_kasv), colour="grey")

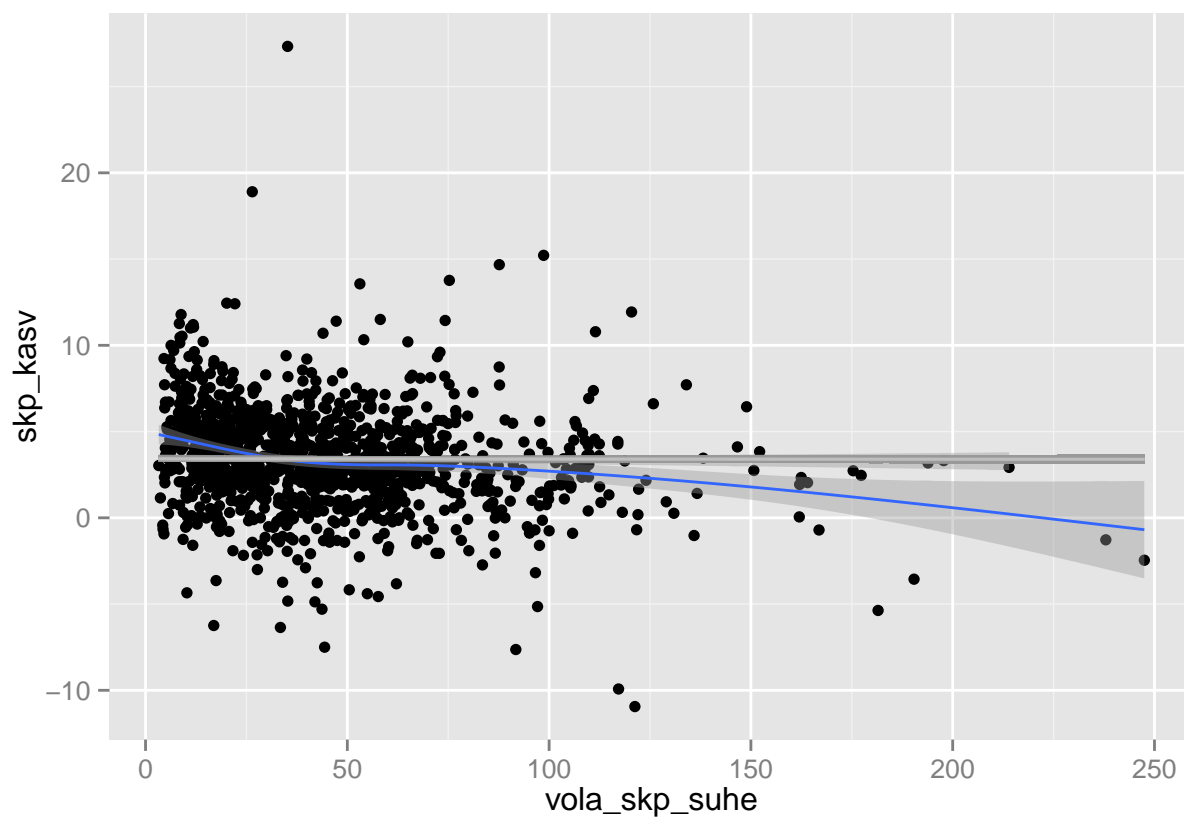
#jätame ainult vajaliku alles
andmed_permut=andmed_permut[, c("vola_skp_suhe", "skp_kasv", "iter1" )]
andmed_permut_iter=andmed_permut

#graafik itereerimiseks
plot_permut_iter=ggplot(andmed_permut_iter, aes(x=vola_skp_suhe, y=skp_kasv))+
  geom_point()+
  stat_smooth()

plot_permut_iter=plot_permut_iter+stat_smooth(aes(x=iter1, y=skp_kasv), colour="grey")

#nüüd tuleb seda 100 korda teha
#library(ggplot2) lisaks tuleb esimene graafik valmis teha
j=4
for (i in 1:100) {
  andmed_permut_iter[,j]=sample(andmed_permut_iter$vola_skp_suhe, replace=T)
  plot_permut_iter=plot_permut_iter+stat_smooth(aes(x=andmed_permut_iter[,j],
                                                    y=skp_kasv), colour="grey")
  j+1
}

plot_permut_iter
```

muidu opereerib hästi, kuid siin teeb jama