



Euroopa Liit
Euroopa Sotsiaalfond



Eesti
tuleviku heaks



RAHANDUSMINISTEERIUM



RIIGI TUGITEENUSTE
KESKUS

praxis
mõttekoda



EESTI RAKENDUSUURINGUTE
KESKUS CENTAR

Sissejuhatus andmeanalüüsi tarkvarasse R

I päev

Esimesel päeval antakse esmalt ülevaade sellest, miks R on täna tõusnud andmeanalüüsi keelte tippu. Mis on tema eelised, aga ka puudused võrreldes alternatiividega. Jätkame seejärel RStudio – Ri enimlevinud kasutajaliidese tutvustamisega ning Ri kasutamise põhitõdedega. Vaatame üle, kuidas andmed Ris saada (nii failidest, kui otse nt Eurostatist või mujalt internetiandmebaasidest), ja tulemusi sealt kätte. Käime läbi Ri käskude põhitõed ja abiinfo loogika, peamised andmestruktuurid Ris.

09:00 – 09:30 Kogunemine

09:30 – 10:45 Sissejuhatus Ri ja Rstudiosse.

Mis on R, kus teda kasutatakse ja mis on tema viimase aja suure populaarsuse taga. Millised on peamised alternatiivid, Ri head ja vead nende kõrval. Vaatame üle Ri enimlevinud kasutajaliidese Rstudio.

10:45 – 11:00 Kohvipaus

11:00 - 12:45 Baasokused Ris

Ri kasutamine – skriptifailide kirjutamine ning käskude käivitamine. Peamised andmestruktuurid.

12:45 - 13.30 Lõunapaus

13.30 - 15.00 Funktsioonid Ris

15.00 - 15.15 Kohvipaus

15.15 - 16.30 Lisapaketid R-is.

II koolituspäev – andmegraafika Ris

Ri populaarsuse kasvu üks esimesi põhjuseid on olnud eriti vinge andmegraafika. Ris realiseeriti esimesena inimsõbralikult graafika grammatika (grammar of graphics), mis tänaseks jõudnud paljudesse teistesse andmeanalüüsivahenditesse (nii et siinomandatavad oskused on kasutatavad laiemalt). Vaatame üle graafika grammatika põhitõed. Näeme, kuidas mõnda baaskontseptsiooni tundes võimalik kiirelt sotti saada, mis seosed andmetes leiduvad, kas nad mingite gruppide kaupa erinevad. Vaatame ka kuidas luua professionaalset andmegraafikat tulemuste esitamiseks. Muuhulgas käime läbi mõned vajalikud kõrvalfunktsioonid - nagu kuidas järjestada riike mõne tunnuse alusel, et nad siis graafikutele õiges järjestuses ilmuksid.

09:30 - 10:00 Kordamine – kinnistame möödunud korral õpitut.

10:30 – 10:45 Graafika grammatika alused

Andme graafika alused Ris. Ggplot pakett, eksploratiivne visuaalne andmete analüüs ning publitseerimiskvaliteediga graafikute loomine

10:45 - 11:00 Kohvipaus

11:15 - 12:45 Graafika grammatika alused

12:45 - 13.30 Lõunapaus

13.30 - 15.00 Graafika grammatika alused

15.00 - 15.15 Kohvipaus

15.15 - 16.30 Andmete grammatika alused

Esmane ülevaade andmete grammatikast (dplyr), millest põhjalikumalt kolmandal päeval.

III päev - moodne R ja raportite kirjutamine Ris

R on andmetöötlemise innovatsiooni avangardis, kuhu tihti ilmuvad esimesena uued ideed, mis seejärel levivad mujale. Astume *tidyversumisse* - moodsasse Ri, mis võimaldab päris keerulisi andmetöötlemise ja -analüüsi ülesandeid lahendada lihtsalt ja elegantselt. Seda on nimetatud andmetöötlemise grammatikaks, kus kümnekonna "tehusõnaga" on võimalik teha ära suur osa kogu andmetöötlemisest.

Kui teie töö sisaldab vähegi rutiinset andmete gruppideks loomist, nende gruppide võrdlust jms, siis võivad need oskused aidata teil oma tööd olulises osas automatiseerida.

09:30 - 10:00 Kordamine – kinnistame möödunud korral õpitut.

10:30 - 11:00 Sissejuhatus andmete grammatikasse

dplyr ja tidyr andmete viimiseks erinevatele kujudele, andmete analüüsimiseks gruppide kaupa

10:45 – 11:00 Kohvipaus

11:00 - 12:45 Sissejuhatus andmete grammatikasse

12:45 - 13.30 Lõunapaus

13.30 - 15.00 Sissejuhatus andmete grammatikasse

15.00 - 15.15 Kohvipaus

15.15 - 16.30 Rmarkdown ja raportite automaatne genereerimine Ris.