

# R-ohjaus

Yliopistonlehtori Juho Kopra

Itä-Suomen yliopisto, Tietojenkäsittelytieteen laitos

11.9.2023

# R:n ja RStudio asentaminen

- 1 Mene sivulle  
<https://posit.co/download/rstudio-desktop/#download> ja klikkaa Download and install R.
- 2 Paina “Download R for Windows” tai “Download R for macOS” riippuen käyttöjärjestelmästäsi.
- 3 Paina “base”.
- 4 “Download R-4.3.1 for Windows” tai vastaava. Tiedosto latautuu. Avaa exe ja aloita asentaminen, kuten asentaisit minkä tahansa ohjelman.
- 5 Jatka asennusta käyttäen oletusasetuksia.
- 6 Tarkasta Käynnistä-valikosta tai vastaavasta että R:n asennus onnistui.

- 1 Mene sivulle  
<https://posit.co/download/rstudio-desktop/#download> ja rullaa alaspäin taulukkoon jossa on lueteltu asennuspaketit eri käyttöjärjestelmille.
- 2 Valitse omalle koneellesi sopiva asennuspaketti (yleisin Windows 10/11 tai macille macOS 11+).
- 3 Jatka asennusta käyttäen oletusasetuksia.
- 4 Tarkasta Käynnistä-valikosta tai vastaavasta että RStudion asennus onnistui.

# Aineiston ottaminen käyttöön (datas4uef)

# Aineiston ottaminen käyttöön (datas4uef)

- 1 Käydään seuraavaksi läpi datas4uef -paketin asentaminen kuten luennolla
- 2 Otetaan paketti ja siinä olevat aineistot käyttöön

# Pakettien asentaminen ja lataaminen

- R koostuu perusosasta (base) sekä lisäosista eli paketeista, joilla R:n toiminnallisuutta voidaan laajentaa. Asennetaan nyt remotes -niminen paketti, jota tarvitaan toisen paketin asentamista varten.

```
install.packages("remotes")
```

- Tyypillisesti saat asennettua tarvitsemasi paketit samalla tavalla kuin yllä asennettiin remotes.
- Jos haluamaasi pakettia ei ole lähetetty R:n yleiseen pakettiarkistoon (Comprehensive R Archive Network, CRAN) vaan se on githubissa, niin sen voi asentaa seuraavasti. Asennetaan siis kurssilla tarvittava paketti datas4uef.

```
remotes::install_github("jukop/datas4uef")
```

# datas4uef -paketin käyttöönotto

```
library(datas4uef)
data("aineistoA_dat")
data()
```

- Kirjoittamalla `data()` RStudio täydentää oikeanpuoleisen sulun ja näyttöön tulee saatavilla olevien aineistojen luettelo kursorin kohdalle. Voit valita oikean aineiston hiirellä tai nuolinäppäimen ja Enter-näppäimen avulla. Mikäli luettelo poistuu näkyvistä, voit poistaa komennosta pari merkkiä ja täyttää komennon uudelleen.
- Objektin nimi, jossa aineisto on tietokoneen muistissa tulee näkyviin Environment-välilehteen RStudios oikeanpuoleiseen yläneljännekseen. Jos haluat aineiston `aineistoA_dat` uudella nimellä, esim. nimelle `dat` voit komentaa

```
dat <- aineistoA_dat
rm(aineistoA_dat) # poistaa objektin aineistoA_dat
```



# Aineiston ottaminen käyttöön (tiedostot)

- 1 Päätä kansio, johon haluat laittaa datatiedostot. Esim.  
C:/Users/jukop/Documents/TJK
- 2 Lataa datatiedostot elearnista kyseiseen kansioon.
  - Jos olet R-kieli -kurssilla niin muista purkaa pakattu zip-tiedosto ja siirtää siinä olevat tiedostot haluamaasi kansioon!
- 3 Aseta työskentelykansio komennolla

```
setwd("C:/Users/jukop/Documents/TJK")
```

- 4 Voit nyt kutsua kansiossa C:/Users/jukop/Documents/TJK olevia tiedostoja ilman että sinun tarvitsee kirjoittaa koko hakemistopolkua R-koodiin.

# Aineiston lukeminen sisään (csv)

- Useimmiten aineistot ovat csv-tiedostomuodossa. csv-tiedosto on tekstimuotoinen tiedosto, joka usein oletusarvoisesti avautuu Excelissä (jos asennettu), mutta nähdäksesi miltä aineisto näyttää tekstinä, avaa se esim. Notepadissa tai vastaavassa (ei Wordissa). **Tiedostoa ei saa muokata Excelissä!**
- Luetaan tiedosto sisään funktiolla `read.csv2`. Jos ei toimi, niin kokeile `read.csv`-funktiolla.

```
aineistoA_dat <- read.csv2("AineistoA.csv")
```

- csv-tiedostosta ladatut aineistot eivät ole valmiiksi faktoreina (laatuero ja järjestysasteikko), joten ne täytyy tarvittaessa muuttaa faktoreiksi luentojen ohjeen mukaan (seuraava kalvo on kopioitu luentokalvoista).
  - Myös muuttujien nimet ovat vanhastaan eri. Ne voi muuttaa

```
names(aineistoA_dat) <- c("eka uusi nimi", "toinen uusi nimi")  
# anna uudet nimet tässä muodossa s.e. nimien määrä on sama  
# kuin aineiston sarakkeiden määrä
```

- R-kieli olettaa, että kaikki laskutoimitukset ovat sallittuja
  - Ts. jos muuttuja on numeerinen, niin R olettaa suhdeasteikon
- Toisaalta käytännön kannalta ei yleensä ole väliä onko muuttuja suhdeasteikollinen vai välimatka-asteikollinen (jakolasku on harvinainen vaatimus perusmenetelmissä)
  - Siispä keskitytään siihen onko muuttuja numeerinen (vähintään välimatka-asteikollinen) vai järjestysasteikollinen tai laatueroasteikollinen
- Esimerkiksi olkoon muuttuja, joka on koodattu lukuarvoin 1, 2 ja 3 siten että 1 = "Eri mieltä", 2="Samapa tuo" ja 3="Samaa mieltä".

```
x <- c(1,2,2,3)
```

- laatueroasteikolliseksi

```
x_nom <- factor(x,levels=c(1,2,3),  
               labels=c("Eri mieltä", "Samapa tuo","Samaa mieltä"))  
x_nom
```

```
## [1] Eri mieltä   Samapa tuo   Samapa tuo   Samaa mieltä  
## Levels: Eri mieltä Samapa tuo Samaa mieltä
```

- järjestysasteikolliseksi

```
x_ord <- factor(x,levels=c(1,2,3),  
               labels=c("Eri mieltä", "Samapa tuo","Samaa mieltä"),  
               ordered=TRUE)  
x_ord
```

```
## [1] Eri mieltä   Samapa tuo   Samapa tuo   Samaa mieltä  
## Levels: Eri mieltä < Samapa tuo < Samaa mieltä
```