R-kielen perusteet

Asiantuntijaluennot

Bragge Janne TTM23SAI

2025

Table of Contents

[R-kielen Asentaminen 2](#_Toc184227265)

[1. R-kieli 2](#_Toc184227266)

[1.1. Windows 2](#_Toc184227267)

[1.2. MacOS 2](#_Toc184227268)

[1.3. Linux 2](#_Toc184227269)

[2. RStudio-työkalun asentaminen 3](#_Toc184227270)

[3. Testaa asennus 3](#_Toc184227271)

[4. Lisäpakettien asentaminen 3](#_Toc184227272)

[5. Vinkkejä ja huomioitavaa 3](#_Toc184227273)

[Käytön aloittaminen 4](#_Toc184227274)

[Pseudo-datasetin luominen 4](#_Toc184227275)

[1. Avaa RStudio 4](#_Toc184227276)

[2. Luo uusi skriptitiedosto 4](#_Toc184227277)

[3. Tallenna tiedosto 4](#_Toc184227278)

[4. Aloita koodin kirjoittaminen 4](#_Toc184227279)

[5. Suorita koodi 4](#_Toc184227280)

[Pseudodatan visualisointi 5](#_Toc184227281)

[1. Luo pseudodata 5](#_Toc184227282)

[2. Perusvisualisointi: pistekaavio 5](#_Toc184227283)

[3. Histogrammi 6](#_Toc184227284)

[4. Laatikkokaavio (Boxplot) 7](#_Toc184227285)

[5. Viivakaavio (jos aikasarjadatan tyyppistä) 7](#_Toc184227286)

[Vinkkejä 8](#_Toc184227287)

[Muuttujat (Variables) 9](#_Toc184227288)

[Kappale 2: R-kielen muuttujat (Variables) 9](#_Toc184227289)

[Muuttujien luonti 9](#_Toc184227290)

[Muuttujien nimeäminen 10](#_Toc184227291)

[Muuttujatyypit 10](#_Toc184227292)

[Muuttujien muuttaminen 11](#_Toc184227293)

[Muuttujien käyttäminen laskutoimituksissa 11](#_Toc184227294)

[Erikoismuuttujat 11](#_Toc184227295)

[Muuttujien hallinta 12](#_Toc184227296)

[Näytä kaikki nykyiset muuttujat 12](#_Toc184227297)

[Poista muuttuja 12](#_Toc184227298)

# R-kielen Asentaminen

Tässä raportissa kuvataan vaiheittain, kuinka asentaa R-kieli ja siihen liittyvät työkalut (RStudio) Windows-, MacOS- ja Linux-käyttöjärjestelmissä.

## 1. R-kieli

R on avoimen lähdekoodin ohjelmointikieli, joka on erityisen suosittu data-analyysissä ja tilastollisessa laskennassa.

### 1.1. Windows

1. Siirry R:n viralliselle verkkosivustolle: [CRAN R Project](https://cran.r-project.org/).
2. Valitse “Download R for Windows”.
3. Klikkaa "base" ja lataa viimeisin versio ("Download R X.X.X for Windows").
4. Suorita ladattu asennustiedosto.
   * Valitse asennuksen aikana oletusasetukset, ellet tiedä tarvitsevasi erityisiä määrittelyjä.
5. Kun asennus on valmis, R voidaan avata joko komentoriviltä tai RStudion kautta.

### 1.2. MacOS

1. Siirry CRAN-sivustolle: [CRAN R Project](https://cran.r-project.org/).
2. Valitse “Download R for macOS”.
3. Lataa sopiva versio (yleensä "arm64.pkg" tai "x86\_64.pkg" riippuen prosessoristasi).
4. Avaa ladattu tiedosto ja noudata ohjeita asentaaksesi R:n.
5. R on nyt käytettävissä. Asenna RStudio lisämukavuuden saavuttamiseksi.

### 1.3. Linux

1. Avaa komentorivi.
2. Lisää CRAN:n pakettivarasto järjestelmääsi. Esimerkiksi Ubuntu/Debian-järjestelmissä:
3. sudo apt update
4. sudo apt install r-base
5. Asenna tarvittaessa ylimääräiset paketit komennolla install.packages() R-konsolissa.
6. Käynnistä R kirjoittamalla komentoriville R.

## 2. RStudio-työkalun asentaminen

RStudio on R:n kehitysympäristö, joka tarjoaa helppokäyttöisen käyttöliittymän ja monipuolisia ominaisuuksia.

1. Siirry RStudion verkkosivustolle: [RStudio Download](https://posit.co/download/rstudio-desktop/).
2. Lataa ilmainen versio, joka on saatavilla kaikille käyttöjärjestelmille.
3. Asenna ohjelma ja avaa se. Varmista, että R on asennettu ennen RStudion käyttämistä.

## 3. Testaa asennus

1. Avaa R tai RStudio.
2. Kirjoita konsoliin:
3. print("R toimii oikein!")

Jos saat tulosteena tekstin "R toimii oikein!", asennus on onnistunut.

## 4. Lisäpakettien asentaminen

Voit lisätä R:ään laajennuksia (paketteja), jotka tarjoavat uusia toimintoja:

install.packages("dplyr")

library(dplyr)

Tämä komento asentaa ja lataa dplyr-paketin käyttöön.

## 5. Vinkkejä ja huomioitavaa

* **Päivitykset:** Pidä R ja RStudio ajan tasalla uusien ominaisuuksien ja parannusten vuoksi.
* **RMarkdown:** Asenna rmarkdown-paketti raporttien luomiseen:
* install.packages("rmarkdown")
* **Dokumentaatio:** Hyödynnä R:n laajaa verkkodokumentaatiota ja yhteisön resursseja.

Raportin avulla voit suorittaa R:n ja siihen liittyvien työkalujen asennuksen onnistuneesti eri käyttöjärjestelmissä.

# Käytön aloittaminen

## Pseudo-datasetin luominen

### 1. Avaa RStudio

* Käynnistä RStudio tietokoneellasi.

### 2. Luo uusi skriptitiedosto

1. Valitse ylävalikosta **File > New File > R Script**.
2. Uusi tyhjä R-skriptitiedosto avautuu editoriin.

### 3. Tallenna tiedosto

1. Klikkaa ylävalikosta **File > Save** tai paina **Ctrl + S** (Windows) tai **Cmd + S** (Mac).
2. Anna tiedostolle nimi, esimerkiksi:  
   **pseudo\_dataset.R**
3. Valitse tallennuspaikka ja paina **Save**.

### 4. Aloita koodin kirjoittaminen

Esimerkki pseudodatasetin luomisesta:

# Tämä on uusi R-tiedosto

set.seed(123)

pseudo\_data <- data.frame(

ID = 1:50,

Value = rnorm(50, mean = 100, sd = 15)

)

print(head(pseudo\_data))

View(pseudo\_data)

### 5. Suorita koodi

1. Valitse haluamasi koodirivi tai lohko.
2. Paina **Ctrl + Enter** (Windows) tai **Cmd + Enter** (Mac) suorittaaksesi sen.

## Pseudodatan visualisointi

RStudiossa pseudodatan visualisointi onnistuu helposti esimerkiksi **ggplot2**-kirjastolla, joka on yksi suosituimmista R:n visualisointikirjastoista. Tässä on vaiheittainen ohje, kuinka voit visualisoida pseudodatan.

**Esimerkki pseudodatan luomisesta ja visualisoinnista**

### 1. Luo pseudodata

Aloitetaan yksinkertaisella pseudodatasetillä:

# Lataa ggplot2-kirjasto (asennus: install.packages("ggplot2"))

library(ggplot2)

# Luo pseudodata

set.seed(123)

pseudo\_data <- data.frame(

ID = 1:100,

Age = sample(18:65, 100, replace = TRUE),

Income = round(rnorm(100, mean = 50000, sd = 10000), 0),

Gender = sample(c("Male", "Female"), 100, replace = TRUE)

)

# Tarkastele dataa

head(pseudo\_data)

### 2. Perusvisualisointi: pistekaavio

Visualisoidaan esimerkiksi **ikä ja tulot sukupuolen mukaan**:

ggplot(pseudo\_data, aes(x = Age, y = Income, color = Gender)) +

geom\_point(size = 3, alpha = 0.7) + # Lisää pisteet

labs(

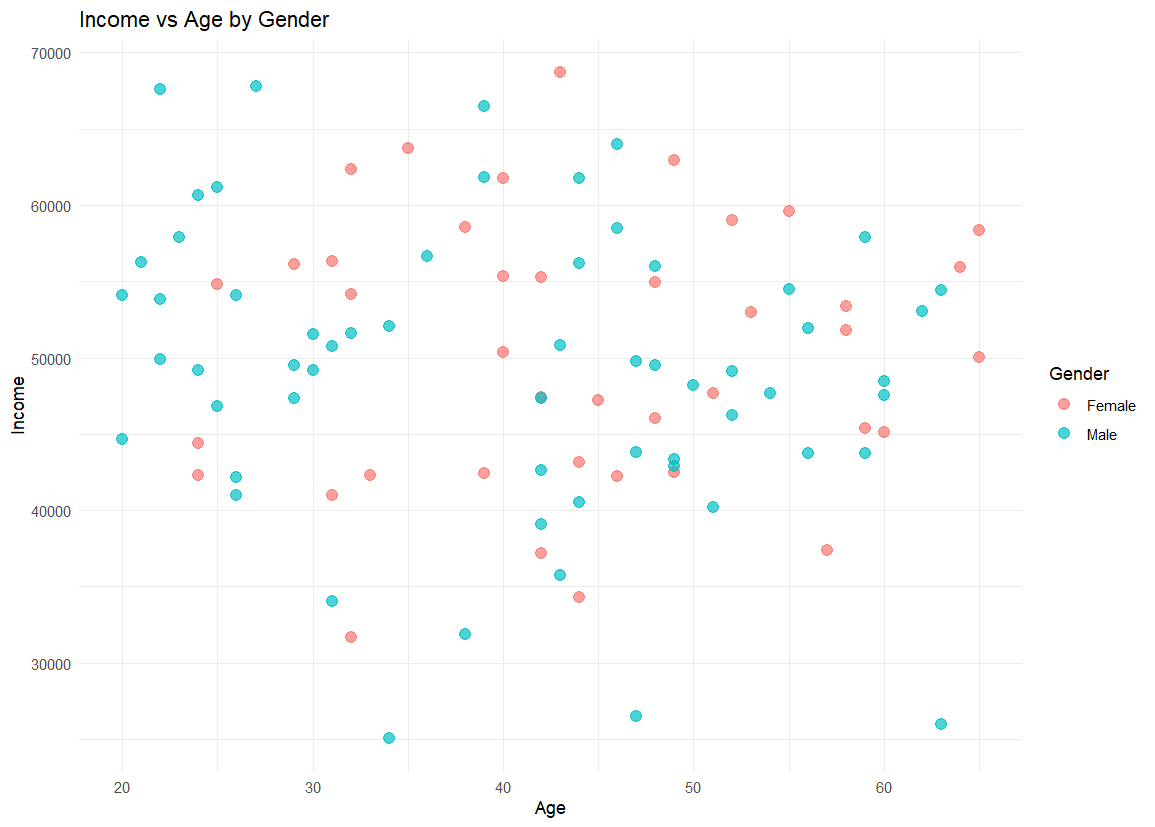
title = "Income vs Age by Gender",

x = "Age",

y = "Income"

) +

theme\_minimal() # Käytä minimalistista teemaa



### 3. Histogrammi

Histogrammin avulla voit tarkastella esimerkiksi ikäjakaumaa:

ggplot(pseudo\_data, aes(x = Age, fill = Gender)) +

geom\_histogram(binwidth = 5, alpha = 0.7, position = "dodge") + # Histogrammi sukupuolen mukaan

labs(

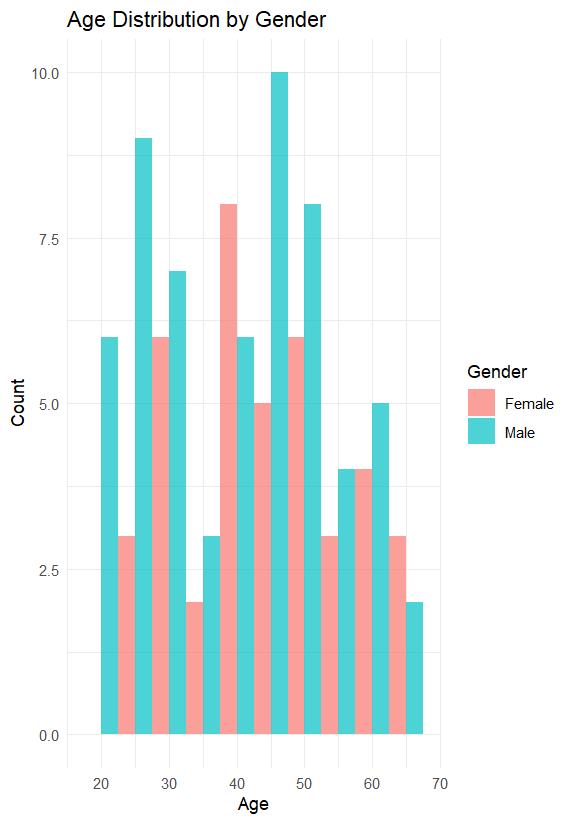
title = "Age Distribution by Gender",

x = "Age",

y = "Count"

) +

theme\_minimal()



### 4. Laatikkokaavio (Boxplot)

Laatikkokaavio sopii hyvin tulojen jakauman tarkasteluun sukupuolen mukaan:

ggplot(pseudo\_data, aes(x = Gender, y = Income, fill = Gender)) +

geom\_boxplot() +

labs(

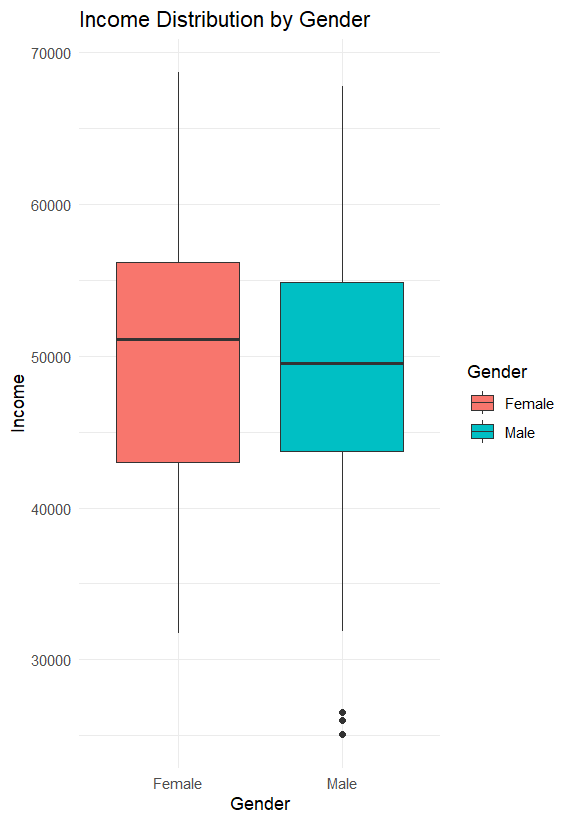
title = "Income Distribution by Gender",

x = "Gender",

y = "Income"

) +

theme\_minimal()



### 5. Viivakaavio (jos aikasarjadatan tyyppistä)

Jos sinulla olisi aikasarjaa, voit visualisoida sen näin:

pseudo\_data$Time <- seq.Date(from = as.Date("2023-01-01"), by = "day", length.out = 100)

ggplot(pseudo\_data, aes(x = Time, y = Income, color = Gender)) +

geom\_line() +

labs(

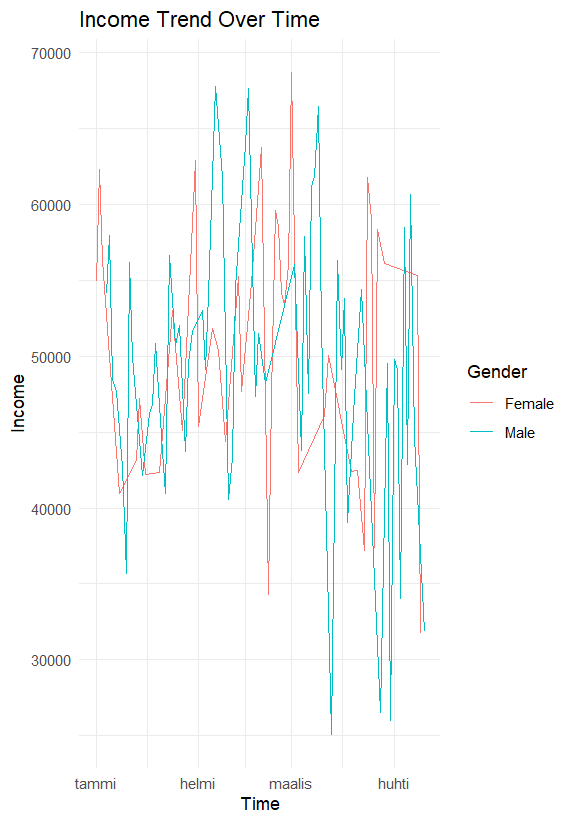
title = "Income Trend Over Time",

x = "Time",

y = "Income"

) +

theme\_minimal()



### Vinkkejä

1. **Mukauta visualisointeja**:  
   Lisää teemaa, värejä tai fontteja oman tarpeesi mukaan käyttämällä theme()-funktiota.
2. **Tallentaminen**:  
   Tallenna visualisointi tiedostoksi:
3. g <- ggplot(pseudo\_data, aes(x = Age, y = Income, color = Gender)) +
4. geom\_point(size = 3, alpha = 0.7) +
5. labs(title = "Income vs Age by Gender", x = "Age", y = "Income") +
6. theme\_minimal()
7. ggsave("income\_vs\_age.png", plot = g, width = 8, height = 6)

# Muuttujat (Variables)

## Kappale 2: R-kielen muuttujat (Variables)

**Mikä on muuttuja?**

Muuttuja on nimetty säilö, joka tallentaa arvoja, kuten numeroita, tekstiä, loogisia arvoja tai muita tietotyyppejä. R-kielessä muuttujien avulla voimme tallentaa, muokata ja käyttää tietoja ohjelmoinnissa.

## Muuttujien luonti

Muuttuja luodaan käyttämällä **<-**-operaattoria tai yhtälailla **=**-operaattoria.

**Esimerkkejä:**

# Muuttujan luonti

x <- 10 # Luvun tallentaminen muuttujaan x

y = "Hello" # Merkkijonon tallentaminen muuttujaan y

z <- TRUE # Loogisen arvon tallentaminen muuttujaan z

Voit myös tarkistaa muuttujan arvon kirjoittamalla sen nimen:

x # Tulostaa: 10

y # Tulostaa: "Hello"

z # Tulostaa: TRUE

A screenshot of a computer

Description automatically generated

## Muuttujien nimeäminen

R-kielessä muuttujien nimet voivat sisältää kirjaimia, numeroita, alaviivoja (\_) ja pisteitä (.), mutta niiden on aloitettava kirjaimella tai pisteellä, joka ei seuraa numeroa.

**Hyviä käytäntöjä:**

* Käytä kuvaavia nimiä, kuten age tai total\_sales.
* Noudatetaan **snake\_case** tai **camelCase**-tyylejä:
  + **snake\_case**: total\_sales
  + **camelCase**: totalSales

**Esimerkkejä:**

valid\_variable <- 5 # Sallittu

variable.name <- "text" # Sallittu

total\_sales <- 200 # Sallittu

**Virheelliset nimet:**

2variable <- 5 # Ei sallittu (ei voi alkaa numerolla)

variable-name <- 5 # Ei sallittu (viiva ei ole sallittu merkki)

## Muuttujatyypit

Muuttuja voi sisältää monentyyppisiä tietoja. Tässä yleisimmät tietotyypit:

1. **Numerot (numeric)**  
   Esimerkiksi desimaaliluvut tai kokonaisluvut.
2. a <- 42 # Kokonaisluku
3. b <- 3.14 # Desimaaliluku
4. **Merkkijonot (character)**  
   Tekstiä, jonka ympärillä käytetään lainausmerkkejä.
5. text <- "Hello, R!" # Merkkijono
6. **Loogiset arvot (logical)**  
   Arvot **TRUE** tai **FALSE**.
7. is\_active <- TRUE
8. **Vektorit (vector)**  
   Useiden arvojen lista, kuten [1, 2, 3].
9. numbers <- c(1, 2, 3, 4, 5)

**huom!**

A computer code with black text

Description automatically generated with medium confidence

## Muuttujien muuttaminen

Muuttuja voidaan päivittää helposti antamalla sille uusi arvo:

x <- 10 # Alkuperäinen arvo

x <- x + 5 # Päivitetään arvo (x nyt 15)

Muuttujan tyyppi voi myös vaihtua:

x <- "text" # x oli aiemmin numero, nyt se on merkkijono

## Muuttujien käyttäminen laskutoimituksissa

Muuttujia voidaan käyttää laskutoimituksissa ja muissa operaatioissa:

a <- 20

b <- 5

# Yhteenlasku

sum <- a + b # 25

# Kertolasku

product <- a \* b # 100

# Jakolasku

division <- a / b # 4

## Erikoismuuttujat

R sisältää kolme erikoisarvoa:

1. **NA**: Tarkoittaa "ei saatavilla" (missing value).
2. value <- NA
3. is.na(value) # Tarkistaa, onko arvo NA
4. **NULL**: Tarkoittaa "tyhjää arvoa" (null value).
5. empty <- NULL
6. **Inf ja -Inf**: Käytetään äärettömyyden ilmaisemiseen.
7. infinity <- 1 / 0 # Inf
8. **NaN**: Tarkoittaa "ei numero" (not a number).
9. not\_number <- 0 / 0 # NaN

## Muuttujien hallinta

### Näytä kaikki nykyiset muuttujat

ls() # Tulostaa kaikki ympäristössä olevat muuttujat

### Poista muuttuja

rm(x) # Poistaa muuttujan x

#### Käytännön esimerkki

# Myynnin seuranta

products\_sold <- 120 # Kuinka monta tuotetta myytiin

price\_per\_unit <- 25 # Tuotteen hinta

# Laske kokonaistulot

total\_revenue <- products\_sold \* price\_per\_unit

print(total\_revenue) # Tulostaa: 3000