

IDA Introduktion til R

Tue Hellstern

2020-04-20

Contents

1	Webinar	5
1.1	Filer	5
2	R og RStudio	7
2.1	R	7
2.2	RStudio	7
3	Grundlæggende R	9
3.1	Navne	9
3.2	Variable / Objekter	9
3.3	Funktioner	10
3.4	Logiske operatører	10
3.5	Hjælp	11
3.6	Working Directory	11
4	Pakker	13
4.1	Installation af pakker	13
5	Data	15
5.1	Grafisk (GUI)	15
5.2	Kode	15
6	Plots	17
6.1	Standard plot	17
6.2	ggplot2	18

7 Demo Shiny	19
7.1 Struktur af en Shiny app	19
8 Resourcer	21
8.1 Cheatsheets	21
8.2 Links	21

Chapter 1

Webinar

- Webinar den 22-04-2020
- Kl. 19:00
- Tue Hellstern
- Jelgren Consult

Dette webinar er udviklet af Tue Hellstern i samarbejde med Jelgren Consult - <http://www.jelgrenconsult.dk>

Dette materiale må ikke kopieres eller mangfoldiggøres uden forudgående aftale med Jelgren Consult.



1.1 Filer

Det er følgende filer der bruges på dette webinar:

R kode filer

- webinar.R

Data filer

- webinar_data.xlsx - *Indeholder tre ark; height_weight, issalg og salgs_data*
- bmi.csv
- issalg.csv
- salgs_data.csv

Introduktion {#intro}

Chapter 2

R og RStudio

Du skal have installeret R og det IDE udviklingsmiljø, der hedder RStudio, begge er open source og begge findes til Mac, Linux og Windows.

2.1 R

Du skal downloade R fra <https://cran.r-project.org> Du skal her vælge den version, der passer til din computer – Linux, OS X (Mac) eller Windows. Der er vejledninger for de forskellige installationer på www siden.

2.2 RStudio

Når R er installeret, kan du bruge det, men for at gøre det nemmere bruger vi et udviklingsmiljø, der hedder RStudio. Rstudio findes i flere forskellige versioner, den du skal bruge er *RStudio Desktop Open Source License*

Som du finder her: <https://www.rstudio.com/products/rstudio/download/> Du skal her vælge den version, der passer til din computer.

The screenshot displays the RStudio environment with four main panels:

- Code Editor:** Contains R code for linear regression and plotting. The code includes comments, data loading, model fitting, prediction, and ggplot2 commands for two plots.
- Workspace:** Lists objects in the global environment, including 'a', 'folk2', 'folk4', 'folk5', 'folk6', 'folk7', 'folk8', 'folk9', 'folk10', 'folk11', 'folk12', 'folk13', 'folk14', 'folk15', 'folk16', 'folk17', 'folk18', 'folk19', 'folk20', 'folk21', 'folk22', 'folk23', 'folk24', 'folk25', 'folk26', 'folk27', 'folk28', 'folk29', 'folk30', 'folk31', 'folk32', 'folk33', 'folk34', 'folk35', 'folk36', 'folk37', 'folk38', 'folk39', 'folk40', 'folk41', 'folk42', 'folk43', 'folk44', 'folk45', 'folk46', 'folk47', 'folk48', 'folk49', 'folk50', 'folk51', 'folk52', 'folk53', 'folk54', 'folk55', 'folk56', 'folk57', 'folk58', 'folk59', 'folk60', 'folk61', 'folk62', 'folk63', 'folk64', 'folk65', 'folk66', 'folk67', 'folk68', 'folk69', 'folk70', 'folk71', 'folk72', 'folk73', 'folk74', 'folk75', 'folk76', 'folk77', 'folk78', 'folk79', 'folk80', 'folk81', 'folk82', 'folk83', 'folk84', 'folk85', 'folk86', 'folk87', 'folk88', 'folk89', 'folk90', 'folk91', 'folk92', 'folk93', 'folk94', 'folk95', 'folk96', 'folk97', 'folk98', 'folk99', 'folk100'.
- R Console:** Shows the execution of the code, including the output of the linear regression model and the ggplot2 commands.
- Plot & Files:** Displays a scatter plot titled "Vægt og højde i forhold til hinanden" with a subtitle "Data er fiktive". The plot shows a positive correlation between weight (Vægt i Kg) and height (Højde i cm) using red points and a blue regression line.

Chapter 3

Grundlæggende R

Som alle andre programmeringssprog har R nogle forskellige regler for, hvordan koden skal bygges op – altså syntaksen for R.

3.1 Navne

”Navne” i R skal overholde nogle regler:

- Bogstaver – a til z – store og små. Du kan ikke bruge de danske tegn æøå
- Tal
- Tegnene ”.” og ”_”. Altså ikke mellemrum eller bindestreg
- Et navn kan ikke starte med et Tal, men skal starte med et bogstav eller et gyldigt tegn
- R er case-sensitiv. Der er forskel på store og små bogstaver

3.2 Variable / Objekter

Alt hvad du kommer til at bruge i R vil blive gemt i Objekter, det kan være et enkelt tal eller et helt datasæt. I R skal du bruge ”<-” for at gemme en værdi i et objekt. Mange andre programmeringssprog bruger ”=”

```
2 + 2                # Beregn
1:100                # printer 1 til 100
print("Hej med dig")

x <- 2                # Tildel x værdien 2
x * 2
```

```

x * x
x + x
x                                # Viser værdien af x

y <- c(6, 7, 8, 9, 10) # Gemmer 6 til 10 i variabelen y
# c() står for concatenate eller combine - Giver en vektor

x + y    # x lagt sammen med y
y
x * 2    # Ganger alle elementer i x med 2 - LOOP uden at være et LOOP ?

ls()     # Vis / List alle objekter

```

3.3 Funktioner

R har de samme funktioner som du måske kender fra Excel eller andre programmeringssprog

```

length(y)    # Antallet af numre i vektoren
min(y)       # Minimumsværdien
max(y)       # Maksimumsværdien
median(y)    # Medianen
sum(y)       # Summen
mean(y)      # Gennemsnittet
var(y)       # Variansen
sd(y)        # Standardafvigelsen

```

3.4 Logiske operatorer

Meget af det du kommer til at arbejde med i R, vil kræve en eller anden form for logisk operator. Med en logisk operator finder du ud af om et udsagn er Sandt eller Falsk.

```

x == 2      # "==" betyder "lig med"
x != 2      # "!=" betyder "ikke lig med"
x < 1       # "<" betyder "mindre end"
x > 1       # ">" betyder "større end"
x <= 2      # "<=" betyder "mindre eller lig med"

```

Husk, at dette også gælder for de mapper, du arbejder med.

3.5 Hjælp

Du kan få hjælp direkte i R ved at bruge nedstående kommandoer:

```
help("mean") # Hjælp til en specifik funktion - mean
help("plot")
```

```
?mean
?plot
```

```
# Hjælp til en specifik pakke - base og tidyverse
help(package = "base")
help(package = "tidyverse")
```

Hjælpen vises i vinduet nederst til højre

3.6 Working Directory

R bruger en mappe som "Working folder". Det er nemmere at tilgå filer, der ligger i denne mappe.

```
getwd()
setwd("C:/Users/Tue Hellstern/Documents/GitHub/r_webinar/r_kode/webinar_dag1")
```


Chapter 4

Pakker

Du kan ”tilføje” funknationalitet til R ved at installere forskellige pakker. Du kan nederst til højre se, hvilke pakker du har installeret. Det er også her muligt at opdatere de pakker, du har installeret. Du kan finde en oversigt over de forskellige pakker her: cran.rstudio.com

4.1 Installation af pakker

Du kan installere pakker på to måder:

- Via RStudio - GUI
- Via R kode

Grafisk I vinduet nederst til højre kan du klikke på fanen **Packages** og derefter på knappen **Install**, du skriver navnet på den eller de pakker du vil installere.

Det er også muligt at opdatere dine pakker vi knappen **Update**

I sammme vindue kan du også se hvilke pakker der er i *brug*.

Kode Det er nemt at installere og bruge pakker via kode:

```
library()                # Pakker der er på din computer
install.packages("ggplot2") # install
library(dplyr)           # load

update.packages("tidyverse") # Update
remove.packages("tidyverse") # Fjern
```

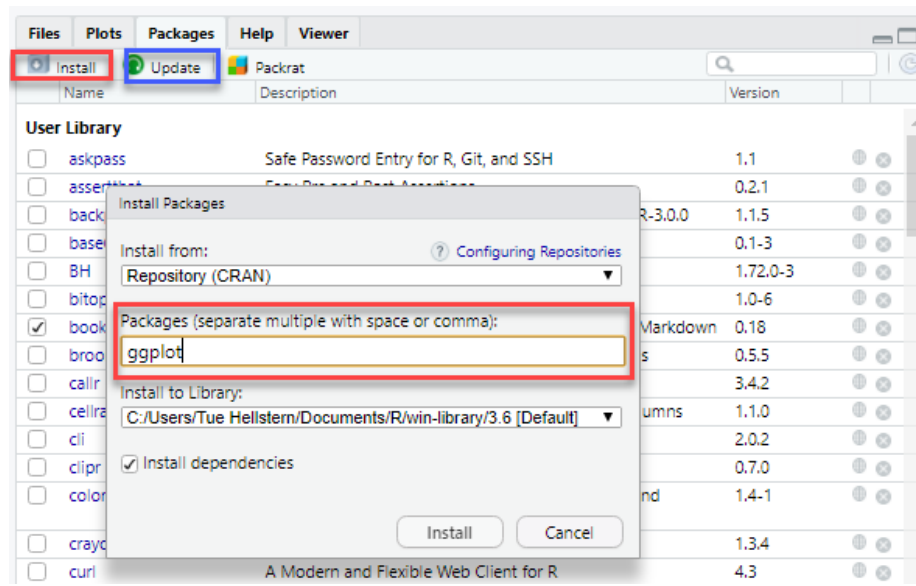


Figure 4.1: Pakker

Chapter 5

Data

For at arbejde med data skal de importeres ind i R. Det er simpelt, hvis de filer der skal importeres ligger i R's aktive bibliotek. Det er en god ide at organisere dine data inden de importeres, hvilket der findes mange gode værktøjer til.

R kan importere fra de fleste formater, her vil jeg bruge; **CSV** og **Excel**

Du kan i RStudio vælge mellem at bruge den grafiske brugerflade (GUI) eller skrive den kode, du vil bruge til importen.

5.1 Grafisk (GUI)

Fordelen ved at bruge den grafiske brugerflade er, at du ikke skal huske de forskellige importfunktioner. Det ligner lidt import i f.eks. Excel.

Import af fra Excel

Når du vil importere data fra en Excel fil, kan du bruge: File > Import Dataset > From Excel Derefter vælger du den Excel fil, du vil importere

5.2 Kode

Excel import

```
salg <- read_excel("webinar_data.xlsx")           # Indlæser første ark
salg <- read_excel("webinar_data.xlsx", sheet = "salgs_data") # Indlæser arket salgs_data
salg <- read_excel("webinar_data.xlsx", sheet = 3)   # Indlæser ark 3
```

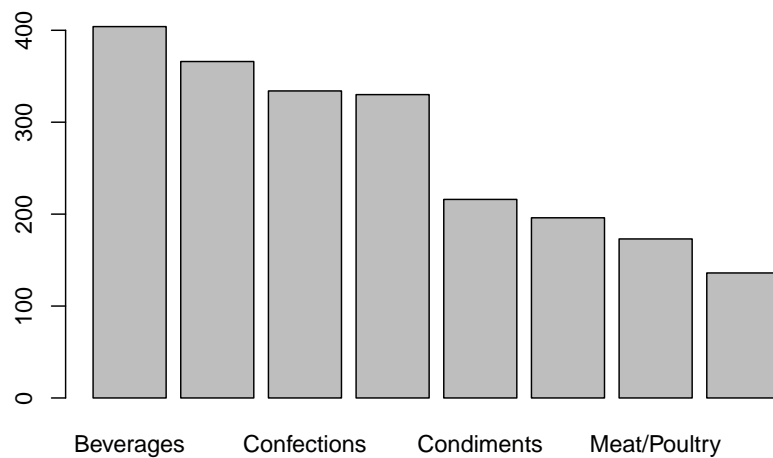

Chapter 6

Plots

6.1 Standard plot

```
category <- table(salg$CategoryName)
str(category)
barplot(category[order(category, decreasing = T)])
```

Giver dette plot



6.2 ggplot2

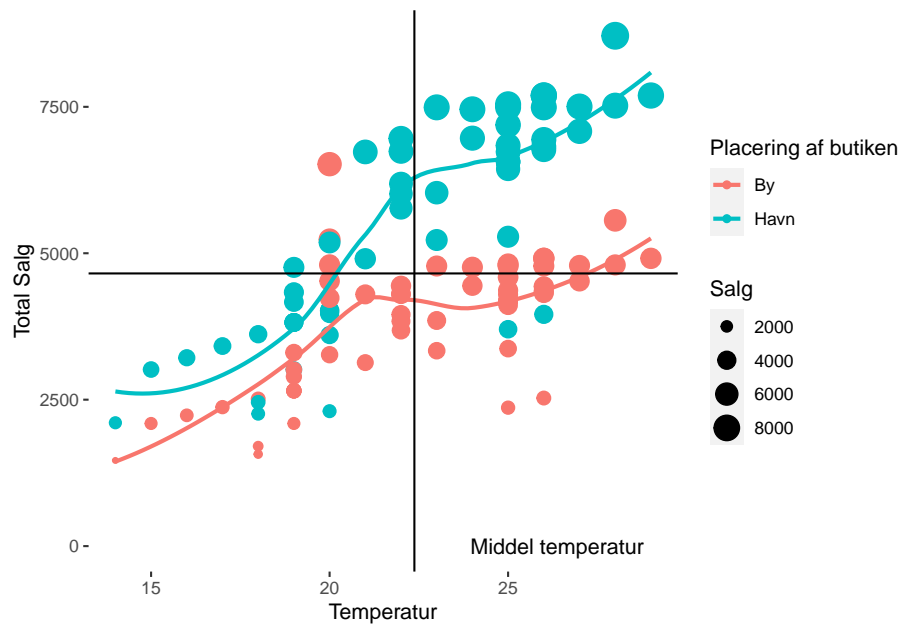
Kode

```
ggplot(data=issalg) +
  geom_point(mapping = aes(x=Temperatur, y=Salg, color=Butik_Placering, size=Salg)) +
  geom_smooth(mapping = aes(x=Temperatur, y=Salg, color=Butik_Placering), se=FALSE) +
  geom_hline(yintercept = mean(issalg$Salg)) +
  geom_vline(xintercept = mean(issalg$Temperatur)) +
  annotate("text", label="Middel temperatur", y=1, x=mean(issalg$Temperatur+4)) +
  theme(panel.background = element_blank()) +
  theme(plot.background = element_blank()) +
  scale_x_continuous(name="Temperatur") +
  scale_y_continuous(name="Total Salg") +
  scale_color_discrete(name="Placering af butikken") +
  scale_size_continuous(name="Salg")
```

Giver dette plot

```
## Warning: package 'ggplot2' was built under R version 3.6.3
```

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



Chapter 7

Demo Shiny

Når du har lavet en "analyse" i R vil et tit være en fordel at kunne dele denne med andre, det kan du gøre ved at sende den dit R script og data, men det er nok ikke den mest brugervenlige måde. En anden ulighed er vise det på en webside så det kan tilgås via en browser. Det kan du opnå via en R pakke der hedder shiny. Du kan finde en del eksempler på løsninger oprettet i shiny her: shiny.rstudio.com/gallery/#demos

Det er også muligt at afvikle de 11 eksempler der følger med shiny på følgende måde:

```
# Shiny eksempler
runExample("01_hello")
runExample("02_text")
runExample("03_reactivity")
runExample("04_mpg")
runExample("05_sliders")
runExample("06_tabsets")
runExample("07_widgets")
runExample("08_html")
runExample("09_upload")
runExample("10_download")
runExample("11_timer")
```

7.1 Struktur af en Shiny app

En Shiny App er "samlet" i et enkelt script app.R som du kan afvikle. app.R består af tre dele:

- Userinterface (ui.R) - *Her styrer du layout af din applikation*
- Server (server.R) - *Her definere du det der skal til for at opbygge din løsning*
- Kald til shinyApp function - *Her bygges selev løsningen*

Du kan oprette en ny shiny App under: **File – New File – Shiny Web App**. I dette vindue skal du indtaste navnet på din App og placeringen. Du har også mulighed for at vælge mellem om din App skal opdeles i to file, **ui.R** og **server.R**, eller om du vil have alt i en fil.

Opdelingen i to filer giver dig de fleste muligheder og en bedre kontrol.

Du får to filer der indeholder en demo løsning som du kan bruge som udgangspunkt for din egen løsning.

Chapter 8

Resourcer

8.1 Cheatsheets

- RStudio Cheatsheets - rstudio.com/resources/cheatsheets/
- RStudio IDE
- readr
- dplyr
- Shiny
- ggplot2

8.2 Links

Her er en lille samling af nyttige links, det er på ingen måde en fuldstændig liste
– Brug Google.

- www.rproject.org
- mran.microsoft.com/open
- www.rstudio.com
- support.rstudio.com/hc/en-us/categories/200035113-Documentation
- community.rstudio.com
- www.statmethods.net