

# R Notebook template DS at HVL

```
---
title: "R Notebook template DS at HVL"
bibliography: ds_course_tmpl.bib
csl: springer-basic-author-date.csl
header-includes:
- \usepackage{amsmath}
- \usepackage{mathtools}
output:
  html_notebook: default
  pdf_document:
    latex_engine: xelatex
    citation_package: natbib
author: Arnstein Gjestland
fig_caption: true
---
```

Det en ser ovenfor her er den såkalte YAML headeren til documentet. Den bestemmer bl.a. hvilket format vi vil ha RMarkdown dokumentet konvertert til. Her settes også navnet på Bibliographyfilen vi vil benytte. Den inneholder vår samling av referanser. Bare de referansene vi siterer i teksten vil bli lagt til referanselisten. RStudio vil selv endre litt i YAML, spesielt gjelder dette **output**: verdien som blir endret hvis vi velger et annet format i Previewmenyen.  $a^2 + b^2 = c^2$

Her tester vi matte  $\bar{x} = \frac{\sum_{i=1}^N x_i}{N}$

$$\bar{x} = \frac{\sum_{i=1}^N x_i}{N}$$

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

```
my_mod <- lm(dist ~ speed, data = cars)
summary(my_mod)
```

```
##
## Call:
## lm(formula = dist ~ speed, data = cars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -29.069  -9.525  -2.272   9.215  43.201
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -17.5791     6.7584  -2.601   0.0123 *
## speed        3.9324     0.4155   9.464 1.49e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 15.38 on 48 degrees of freedom
## Multiple R-squared:  0.6511, Adjusted R-squared:  0.6438
## F-statistic: 89.57 on 1 and 48 DF,  p-value: 1.49e-12
```

```
pander(my_mod)
```

Table 1: Fitting linear model: dist ~ speed

	Estimate	Std. Error	t value	Pr(> t )
<b>(Intercept)</b>	-17.58	6.758	-2.601	0.01232
<b>speed</b>	3.932	0.4155	9.464	1.49e-12

```
pander(summary(my_mod))
```

	Estimate	Std. Error	t value	Pr(> t )
<b>(Intercept)</b>	-17.58	6.758	-2.601	0.01232
<b>speed</b>	3.932	0.4155	9.464	1.49e-12

Table 3: Fitting linear model: dist ~ speed

Observations	Residual Std. Error	$R^2$	Adjusted $R^2$
50	15.38	0.6511	0.6438

```
cars[1:10,]
```

```
##      speed dist
## 1      4     2
## 2      4    10
## 3      7     4
## 4      7    22
## 5      8    16
## 6      9    10
## 7     10    18
## 8     10    26
## 9     10    34
## 10     11    17
```

```
pander(cars[1:10,])
```

speed	dist
4	2
4	10
7	4
7	22
8	16
9	10
10	18
10	26

speed	dist
10	34
11	17

```
pander(density(runif(10)))
```

Table 5: Kernel density of *runif(10)* (bandwidth: 0.1817241)

	Coordinates	Density values
<b>Min.</b>	-0.531	0.00253
<b>1st Qu.</b>	-0.01751	0.08491
<b>Median</b>	0.496	0.493
<b>Mean</b>	0.496	0.4863
<b>3rd Qu.</b>	1.009	0.864
<b>Max.</b>	1.523	1.007

```
chisq.test(table(mtcars$am, mtcars$gear))
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(mtcars$am, mtcars$gear)
## X-squared = 20.945, df = 2, p-value = 2.831e-05
pander(chisq.test(table(mtcars$am, mtcars$gear)))
```

Table 6: Pearson's Chi-squared test: `table(mtcars$am, mtcars$gear)`

Test statistic	df	P value
20.94	2	2.831e-05 * * *

Vi starter med en RMarkdown notebook og vil til slutt konvertere den til andre formater.

## Vivamus sagittis lacus vel augue laoreet rutrum faucibus. (Second level header)

Plura mihi bona sunt, inclinet, amari petere vellent. Idque Caesaris facere voluntate liceret: sese habere. Etiam habebis sem dicantur magna mollis euismod. Quid securi etiam tamquam eu fugiat nulla pariat.

### Formatering av tekst

Overskriften ovenfor er satt vha. `### Formatering av tekst` på en egen linje (blank linje før og etter) og er tredje nivå.

*Kursiv* settes vha. `*Kursiv*` og **uthevet** skrift vha. `**uthevet**`. Hevet tekst får vi vha. `^` (f.eks.  $x^2$  er `x^2`), mens senket tekst får vi ved `~` gjerne kalt «tilde» (f.eks.  $y_2$  er `y~2~`). Ønsker vi å skrive **Markdown** kode som *ikke* skal evelures setter vi den mellom to `'`, f.eks. `'x^2'`. Det siste bruker jeg mye av i dette dokumentet. Ofte er `'` en såkalt «dead letter», den er brukt til å sette merke på bokstaver, så vi trenger en `space` etter den siste.

**Liste** Overskriften ovenfor er satt vha. `#### Liste` på en egen linje (blank linje før og etter) og er fjerde nivå.

[illegible]

## Eksempel på lister

- Listen ovenfor er satt vha.

- 4

- 4) Sed haec quis possit intrepidus aestimare tellus.
- 7) Tityre, tu patulae recubans sub tegmine fagi dolor.
- 8) Idque Caesaris facere voluntate liceret: sese habere.

Listen over er satt vha.

- 1) Cras mattis iudicium purus sit amet fermentum.
- 1) Nihil hic munitissimus habendi senatus locus, nihil horum?
  - 1) At nos hinc posthac,
  - 2) sitientis piros Afros.
- 1) Integer legentibus erat a ante historiarum dapibus.
- 1) Sed haec quis possit intrepidus aestimare tellus.
  - 7) Tityre, tu patulae recubans sub tegmine fagi dolor.
  - 7) Idque Caesaris facere voluntate liceret: sese habere.

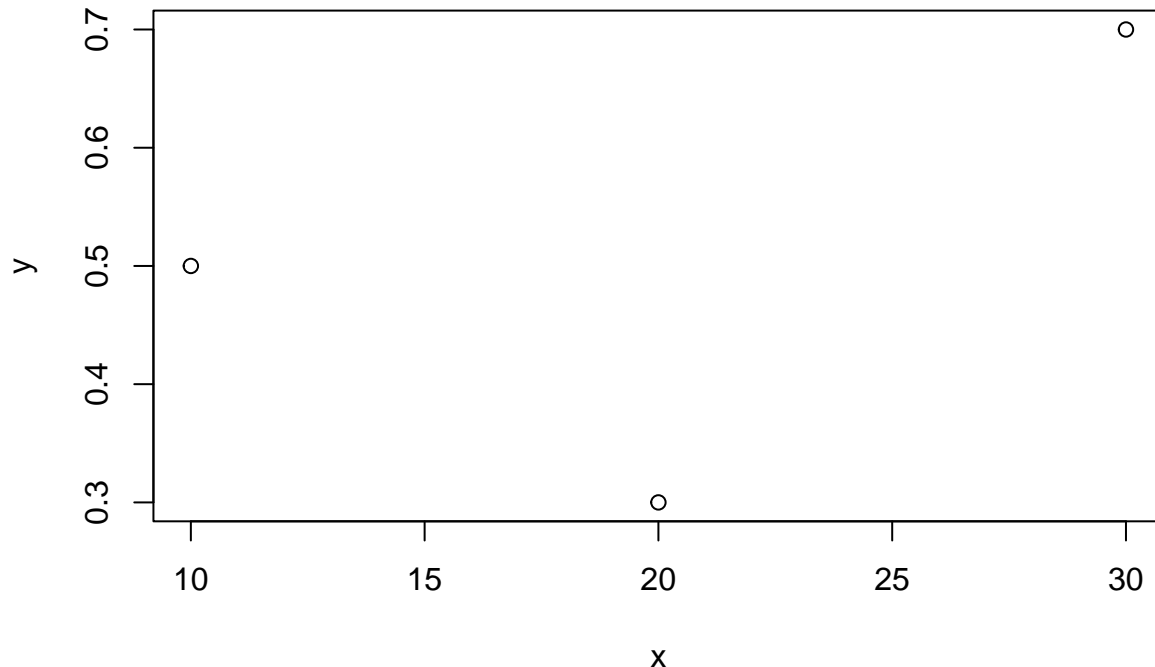
Når vi vil angi Markdown kode over flere linjer benytter vi “” på egen linje før og etter koden. Se i kildekode filen (.Rmd filen) for eksempler på dette.

**R code** Vi kan også sette R kode i dokumentet vårt. Da starter vi med “{r}” på egen linje og avslutter med “” på egen linje. Nedenfor er et eksempel på litt R kode som er riktig, men med dårlig stil

```
```{r} x=c(10,20,30) y=c(0.5,0.3,0.7) plot(x,y) ```
```

Benytter vi «Styler» (marker koden og velg **Style selection** fra **Addins** menyen) kan vi få den tilpasset til **tidyverse** stil. Siden vi er i en såkalt «Notebook» må vi trykke på trekanten til høyre for å se resultatet.

```
x <- c(10, 20, 30)
y <- c(0.5, 0.3, 0.7)
plot(x, y)
```



**Matematikk** For å skrive matematikk i RMarkdown benytter man LaTeX (Lamport 1986) som er en makropakke bygget fra TeX (Knuth 1986) for å skrive strukturerte dokumenter. Når vi genererer pdf filer fra RMarkdown går vi også via LaTeX. Vi skriver matematikk i teksten vha.  $\$ \$$ , f.eks.  $r_m = \sum_{i=1}^N w_i \cdot \bar{r}_i$

vha. `\$r_m = \sum_{i=1}^N w_i \cdot \bar{r}_i\$`. Ønsker vi å sette dette som et uttrykk på egen linje benytter vi `$$$` med blank linje før og etter

$$r_m = \sum_{i=1}^N w_i \cdot \bar{r}_i, \quad \beta_m = \sum_{i=1}^N w_i \cdot \beta_i$$

Dette er satt vha. markup koden

```
$$$
r_m = \sum_{i=1}^N w_i \cdot \bar{r}_i \quad, \quad \beta_m = \sum_{i=1}^N w_i \cdot \beta_i
$$$
```

Ønsker vi likninger over flere linjer kan vi bruke

```
\begin{math}
\begin{aligned}
a^2 + b^2 &= c^2 \quad \&= 5
\end{aligned}
\end{math}
```

som gir oss

$$a^2 + b^2 = c^2 \\ = 5$$

`\begin{aligned}` kommer fra en ekstra-pakke til  $\text{\LaTeX}$  kalt **amsmath** som gir oss muligheter utover standard  $\text{\LaTeX}$  matte. Vi lastet **amsmath** i YAML-headeren i starten av dokumentet.

Vi kan også få til matte over flere linjer vha. standard  $\text{\LaTeX}$ . Nummererte får vi vha. **eqnarray**. F.eks vil

```
\begin{eqnarray}
r_m &= \sum_{i=1}^N w_i \cdot \bar{r}_i, \quad \\
\beta_m &= \sum_{i=1}^N w_i \cdot \beta_i
\end{eqnarray}
```

gi oss

$$r_m = \sum_{i=1}^N w_i \cdot \bar{r}_i, \tag{1}$$

$$\beta_m = \sum_{i=1}^N w_i \cdot \beta_i \tag{2}$$

Ønsker vi ikke nummerering kan vi benyttet **eqnarray\*** f.eks

```
\begin{eqnarray*}
r_m &= \sum_{i=1}^N w_i \cdot \bar{r}_i, \quad \\
\beta_m &= \sum_{i=1}^N w_i \cdot \beta_i
\end{eqnarray*}
```

som gir

$$r_m = \sum_{i=1}^N w_i \cdot \bar{r}_i,$$

$$\beta_m = \sum_{i=1}^N w_i \cdot \beta_i$$

Et tredje alternativ er å benytte `eqnarray` og så slå av nummerering på enkelte linjer vha. `\nonumber`

```
\begin{eqnarray}
r_m &=& \sum_{i=1}^N w_i \cdot \bar{r}_i, \\
\beta_m &=& \sum_{i=1}^N w_i \cdot \beta_i \nonumber
\end{eqnarray}
```

som gir

$$r_m = \sum_{i=1}^N w_i \cdot \bar{r}_i, \tag{3}$$

$$\beta_m = \sum_{i=1}^N w_i \cdot \beta_i$$

Som vi ser ovenfor benytter vi `\\` for å angi ny linje (eller riktigere: slutten på en linje). Vær veldig nøye med `&` tegnene. Ellers blir det mye krøll.

Her er et eksempel sjølet lånt fra Wikibooks LaTeX/Mathematics som er et greit sted for å finne mer om å skrive matematikk vha. `$\LaTeX$`. Det siste eksemplet benytter også `amsmath` (`\cfrac` er definert der) som gir oss flere muligheter.

$$x = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \frac{1}{a_4}}}}$$

**Sitering/referanser** Siteringer skjer i markdown vha. `[@citekey]`, f.eks. `[@knuth1986texbook]` som jeg brukte for å sitere TeX boken ovenfor. Citekey bestemmer vi i utgangspunktet selv, event. mha. Zotero, men det kan være greit å ha et system. Pakken `citr` vil installere en Addin i RStudio som gjør det lettere å legge inn siteringer. Opplysningene som skal i referanselisten hentes inn fra filen som vi anga oppe i YAML headeren, `bibliography: ds_course_tmpl.bib`. Det er veldig viktig at denne ligger på øverste nivå, dvs det må ikke være noe «whitespace» før `bibliography`. Ellers får vi en «Error»

Informasjon til referanselisten hentes fra bib-filen, her `ds_cource_tmpl.bib`. For TeX boken er dette

```
@book{knuth1986texbook,
  title = {The {{TeXbook}}},
  author = {Knuth, D.E.},
  year = {1986},
  publisher = {{Addison-Wesley}},
  isbn = {978-0-201-13447-6},
  lccn = {85030845},
  series = {Computers \& Typesetting}
}
```

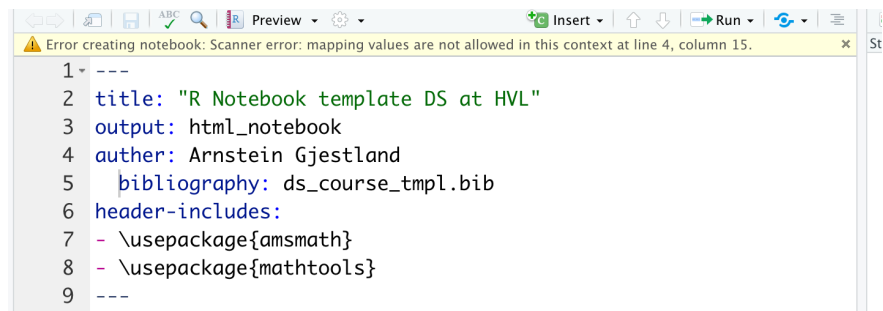


Figure 1: Error

Dette er ingen fornøyelse å skrive inn selv (tro meg, jeg har gjort det) så heldigvis har vi nå verktøy som gjør at vi lett kan stjele hente opplysningene fra andre f.eks. Oria (Bibsys) eller andre nettarkiv (se eget dok. om Zotero).

Systemet fungerer slik at bare det vi referer i teksten blir tatt inn i referanselisten. Bib-filen kan altså inneholde mange flere kilder en de vi bruker. Hvordan siteringene ser ut og hvordan referanselisten er formatert bestemmes av en såkalt cs1 fil (kan hentes fra Zotero Style Repository). Denne legges helst i samme mappe som vårt dokument og må angis i YAML headeren.

Vi har også tre ulike varianter av cite kommandoen som vi skifter mellom alt etter hvordan vi siterer. Disse er (Knuth 1986) vha. kommandoen `[@knuth1986texbook]`, (1986) vha. kommandoen `[-@knuth1986texbook]` og Knuth (1986) vha. kommandoen `@knuth1986texbook`.

En siste ting er av vi kanskje bør referere til R (R Core Team 2019) og de fantastiske pakkene vi bruker slik som `citr` (Aust 2019) etc. Dette kan vi enkelt gjøre ved å benytte R funksjonen `citation()` i Console.

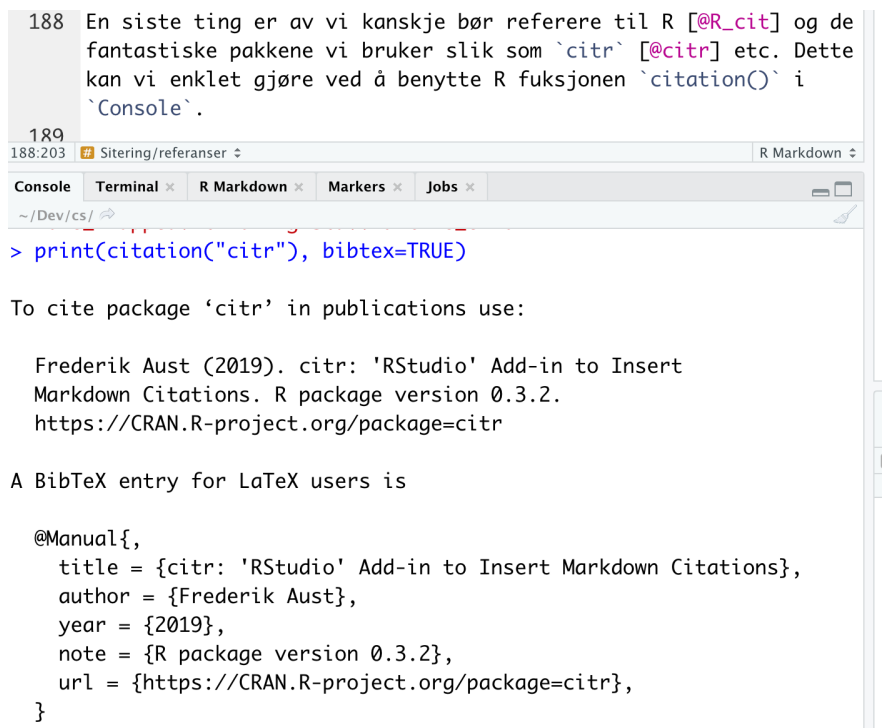


Figure 2: "citr cite"

Fig. 1: Bruk av `citation()` funksjonen in Console. `@Manual{}` etc. må kopieres inn i bib-filen. I `citr` må



*en så klikke på det tomme området under tabellen for å få mulighet til å lese in den oppdaterte bib-filen.*

Merk at RMarkdown ikke har direkte støtte for å sette en forklaring («caption») på figurene. Se gjerne .Rmd filen for å se hvordan jeg har trikset det til her.

Etter å ha gjort tilsvarende for **knitr** (Xie 2014) (husk å legge inn cite-key før første komma i bib-filen. Jeg valgte å bruke `cit_knitr`).

```
@InCollection{cit_knitr,  
  booktitle = {Implementing Reproducible Computational Research},  
  editor = {Victoria Stodden and Friedrich Leisch and Roger D. Peng},  
  title = {kn
```

Jeg har også benyttet pakkene **RMarkdown** (Xie et al. 2018) og **styler** (Müller and Walthert 2020).

Foreløpig har vi bare generert html-dokumenter, men mye av poenget med **RMarkdown** er at vi lett skal kunne generere dokumentet i ulike format. Bruk Knit/Preview menyen og generer pdf og MS Word versjoner.

## Referanser

Aust F (2019) Citr: 'RStudio' add-in to insert markdown citations

Knuth DE (1986) The TeXbook. Addison-Wesley

Lamport L (1986) LATEX: A document preparation system. Addison-Wesley, Reading, Mass

Müller K, Walthert L (2020) Styler: Non-invasive pretty printing of r code

R Core Team (2019) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Xie Y (2014) Knitr: A comprehensive tool for reproducible research in R. In: Stodden V, Leisch F, Peng RD (eds) Implementing reproducible computational research. Chapman; Hall/CRC

Xie Y, Allaire JJ, Golemund G (2018) R markdown: The definitive guide. Chapman; Hall/CRC, Boca Raton, Florida