# Bookdown-dokumetti - testi 1

*Jussi Hirvonen*

*2018-12-04 (Versio 3.05)*

# Sisältö

# Luku 1

# Bookdown-paketin testidokumentti

Esimerkki Rmarkdownin ja bookdown-paketin käytöstä. Kuvat, taulukot ja kaavat numeroidaan ja niihin voi viitata tekstissä. Lähdeviitteet toimivat, myös ne joissa on ns.

A sample document using RMarkdown with bookdown-package to do statistical analysis and publish a report in html and pdf formats.

# Luku 2

# Johdanto

## 2.1 Alkutoimia

Ero YAML-headerissa lang-parametri (lang: fi). (verrattuna bookdown-demoon). Bookdown-demossa lisäksi output: pdf_document, mutta lienee tarpeeton kun kaksi outputformaattia annettu output.yaml-tiedostossa

Bookdown - formaatissa "juuritiedoston" indexBD.Rmd tekstit eivät tulostu jos siellä ei ole luvun (chapter) aloittavaa ensimmäisen tason otsikkoa. Siellä on YAML-headeri (metadata).

Lisää YAML-parametreja voi antaa tiedostoissa _bookdown.yml ja _output.yml. Nämä lienee välittyvät Pandocille?

Bookdown - demon esimerkkitiedostot ovat nämä:

ouput.yml (huomaa, että _ - merkki jätetty pois!) (tässä oli bookdown-demo-paketin yml-tiedostot, poistin 3.7.2018)

## 2.2 Tärkeimmät ohjelmistot

```
system("pdflatex --version")
#getwd()

rmarkdown::pandoc_version()
```

```
## [1] '2.2.1'
```

Viimeinen rivi kertoo pandoc-version.

## 2.3 Muutoksia, tilannetietoja ja puutteita

Nyt toimii gitbook [1] ja pdf_book tulostusformaatteina. Molemmat ovat html-paketteja, ja tarvitsevat ehkä r-datahakemistosta (omalta koneelta) libs-hakemiston jQuery- ja Gitbook-paketit (javaskriptiä ja css-tyylitiedostoja).

test1_preamble.tex -tiedostoa kokeiltiin, mutta sitä ei saatu heinäkuussa toimimaan.

Lähdeluettelossa Å tulee heti A - kirjaimen jälkeen gitbook-versiossa. PDF-tiedostossa taas Å-alkuinen sukunimi sijoittuu vähän toisin virheellisesti. Ikävä juttu!

[1] Virheilmoitukset ovat aika hyödyttömiä. Niiden sijaan tähän alkuun sopisi kuvaus perusideoista ja tekstin muotoiluista. Alaviitteistä esimerkiksi.

# Luku 3

# Kaavat ja matemattiset merkinnät

Kaavat on esitettävä bookdown-paketin määrityksillä. Viittausnimien on oltava yksikäsitteisiä koko dokumentissa, jos käytetään “merge and knit” menetelmää. Jos taas jokainen lapsidokumentti on “itsenäinen” (“knit and merge”), tämä koskee vain kyseistä dokumenttia (kts. Bookdown - webkirja).

Kaavoissa iso ongelma heinäkuussa oli tämä:

**equation-tägien välissä ei saa olla tyhjiä rivejä!**

## 3.1 Kahden luokittelumuuttuja taulukko

Kahden luokittelumuuttujan riippuvuutta voidaan testata $\chi^2$ - testillä. Testisuure saadaan laskemalla yhteen jokaisen solun havaittujen ja odotetettujen (riippumattomuushypoteesi) frekvenssien erotukset muodossa

$$\chi^2 = \frac{(havaittu - odotettu)^2}{odotettu} \tag{3.1}$$

Tämä voidaan esittää ca:han sopivammalla tavalla parilla muunnoksella, jolloin saamme riveittäin vastaavat termit rivisummalla painotettuna:

$$rivisumma \times \frac{(havaittu\, riviprofiili - odotettu\, riviprofiili)^2}{odotettu\, riviprofiili} \tag{3.2}$$

Kun jaamme nämä tekijät havaintojen kokonaismäärällä $n$, rivisumma muuntuu rivin massaksi, ja niiden summa muotoon $\frac{\chi^2}{n}$.

$$\frac{\chi^2}{n} = \phi^2 \tag{3.3}$$

Tunnusluku $\phi^2$ on korrespondenssianalyysissä kokonaisinertia (total inertia). Se kuvaa, kuinka paljon varianssia taulukossa on ja on riippumaton havaintojen lukumäärästä. Tilastotieteessä tunnusluvulla on useita vaihtoehtoisia nimiä (esim. mean square contingency coefficient), ja sen neliöjuurta kutsutaan $\phi$ - kertoimeksi.

Tässä siirrytään kahden luokittelumuuttujan taulukosta suhteellisten frekvenssien taulukkoon, ja pieni pohdinta taulukoista yleensä olisi paikallaan. Yhtälöihin voi viitata (3.1) . Kokeillaan vielä, toimivatko kirjallisuusviittet, kuten tärkeä lähde(Greenacre, 2017).

# Luku 4

# Taulukot ja kuvat

Tähän taulukoita ja kuvia, esimerkkiaineistoilla.

Kirjallisuutta on myös (Roux and Rouanet, 2004), ja (Greenacre and Hastie, 1987) esittelee geometrisen tulkinnan peruskäsitteet yksinkertaisen kahden luokittelumuuttujan korrespondenssianalyysin avaulla. Mitenköhän skandit toimivat lähteissä, bib-tiedostossa on niitä myös escape-muodossa (katso esim. (Älli Åhlgren, 1994), kritiikkiä on esittänyt (Ahlgren, 1994))

Viitteet saa tulostusasetuksilla yhdelle sivulle, oletuksena on viitteiden esittäminen jokaisen sivun alareunassa.

## 4.1 Taulukoita

Tästä poistettu koodilohko data_1, ei tarvita jos ca-paketti on ladattu. Ja alaviiva on aikanakin ref-labeleissa kielletty. Koodilohkojen nimissä taitaa olla sallittu?

Taulukot tulostetaan funktiolla knitr::kable(). Taulukko numeroidaan ja se saa automaattisesti labelin etutunnisteella 'tab', ja siihen liitetään chunk-label (esim alla tab:smoketable1).

Tämä koodipätkä ei antaa yhden kappaleen esikatselussa virheilmoituksen, "smoke"-dataa ei löydy.

```
knitr::kable(smoke[,1:4], booktabs = TRUE,
  caption = 'CA-paketin smoke-data (keinotekoinen)'
)
```

```
# Taulukkoon viittaaminen tekstissä \@ref(label)
```

Taulukossa 4.1 on kahden luokittelumuuttujan keinotekoinen esimerkkiaineisto tupakonnin määrästä henkilöstöryhmittäin (SM = senior managemet, JM = junior management, SE ja JE vastaavasti ryhmälle employee, SC = secretary).

Taulukko 4.1: CA-paketin smoke-data (keinotekoinen)

| | none | light | medium | heavy |
|---|---|---|---|---|
| SM | 4 | 2 | 3 | 2 |
| JM | 4 | 3 | 7 | 4 |
| SE | 25 | 10 | 12 | 4 |
| JE | 18 | 24 | 33 | 13 |
| SC | 10 | 6 | 7 | 2 |

Taulukko 4.2: Riviprofiilit ja keskiarvoprofiili

| | none | light | medium | heavy |
|---|---|---|---|---|
| SM | 0.364 | 0.182 | 0.273 | 0.182 |
| JM | 0.222 | 0.167 | 0.389 | 0.222 |
| SE | 0.490 | 0.196 | 0.235 | 0.078 |
| JE | 0.205 | 0.273 | 0.375 | 0.148 |
| SC | 0.400 | 0.240 | 0.280 | 0.080 |

| none | light | medium | heavy |
|---|---|---|---|
| 0.316 | 0.233 | 0.321 | 0.13 |

Useampi taulukko saadaan taulukkoympäristöön (table environment) yhdistämällä data-objektit listaksi.

```
# riviprofiilit
smoke.rpro <- smoke / rowSums (smoke)
# keskiarvoprofiili
smoke.avrpro <- colSums(smoke) / sum(smoke)

knitr::kable(
  list(smoke.rpro, t(smoke.avrpro)   ), digits = 3,
  caption = 'Riviprofiilit ja keskiarvoprofiili', booktabs = TRUE
)
```

Taulukossa 4.2 on laskettu jokaisen rivin riviprofiilit. Ne saadan kun rivin luvut jaetaan rivin summalla. Yhden rivin taulukossa on esitetty riviprofiilien keskikarvo, sarakesummat jaettuna koko taulukon havaintojen lukumäärällä. Sen prosenttiluvut kertovat tupakoititapojen jakauman koko henkilöstössä.

Jos PDF-tulostuksessa ei haluta ns. kelluvaa taulukkoa (float), voi kable-funktiossa käyttää LaTeXin pakettia longtable. Silloin on myös muistettava ottaa paketti käyttöön (usepackage{}) LaTeX - pohjatiedostossa (preamble).

Pandoc tukee monia Markdownin taulukkotyyppejä. Viittaaminen vaaati labeloidun otsikon, ja sen on oltava otsikkotestin alussa määrämuotoisena (esim. ab:hienotaulu). Tämä vaatii tarkkuutta, jos taulukon pitää toimia html- ja LaTex-outputissa. kable-funktiota kannattaa käyttää!

## 4.2 Korrespondenssianalyysin numeeriset tulokset taulukoina

Korrespondenssianalyysin idea on vähentää aineiston dimensioita, ja esittää taulukon rivien ja sarakkeiden riippuvuudet yleensä kaksiulotteisena karttana.

Numeeriset tulokset ovat tärkeitä diagnostiikassa ja kartan laadun varmistuksessa. Niistä näkee myös täsmällisesti, mitkä rivit ja sarakkeet määrittävät koordinaatiston.

```
smokeCA <- ca(smoke)
#temp1 <- smokeCA tämä kai tarpeeton ? (4.12.2018)
numres1CA1 <- summary(smokeCA)
#str(smokeCA)
#knitr::kable( smokeCA,
#  digits = 3,
#  caption = 'Riviprofiilit ja keskiarvoprofiili', booktabs = TRUE
#)
#str(temp1)
#stargazer(temp2$rows, type = "text", title = "CA-tuloksia")
# LateX-tulostuksessa float vaatii jotain tällaista:Table: (\#tab:cataul1)
#str(temp2)
#str(temp2$scree)
```

Taulukko 4.3: Korrespondenssianalyysin diagnostiikkaa - rivit

| name | mass | qlt | inr | k=1 | cor | ctr | k=2 | cor | ctr |
|---|---|---|---|---|---|---|---|---|---|
| SM | 57 | 893 | 31 | -66 | 92 | 3 | -194 | 800 | 214 |
| JM | 93 | 991 | 139 | 259 | 526 | 84 | -243 | 465 | 551 |
| SE | 264 | 1000 | 450 | -381 | 999 | 512 | -11 | 1 | 3 |
| JE | 456 | 1000 | 308 | 233 | 942 | 331 | 58 | 58 | 152 |
| SC | 130 | 999 | 71 | -201 | 865 | 70 | 79 | 133 | 81 |

```
#temp2$scree
numres1CA1
```

```
##
## Principal inertias (eigenvalues):
##
##  dim    value      %   cum%   scree plot
##  1      0.074759  87.8  87.8  **********************
##  2      0.010017  11.8  99.5  ***
##  3      0.000414   0.5 100.0
##         -------- -----
##  Total: 0.085190 100.0
##
##
## Rows:
##     name   mass  qlt  inr    k=1 cor ctr    k=2 cor ctr
## 1 |   SM |   57  893   31 |  -66  92   3 | -194 800 214 |
## 2 |   JM |   93  991  139 |  259 526  84 | -243 465 551 |
## 3 |   SE |  264 1000  450 | -381 999 512 |  -11   1   3 |
## 4 |   JE |  456 1000  308 |  233 942 331 |   58  58 152 |
## 5 |   SC |  130  999   71 | -201 865  70 |   79 133  81 |
##
## Columns:
##     name   mass  qlt  inr    k=1 cor ctr    k=2 cor ctr
## 1 | none |  316 1000  577 | -393 994 654 |  -30   6  29 |
## 2 | lght |  233  984   83 |   99 327  31 |  141 657 463 |
## 3 | medm |  321  983  148 |  196 982 166 |    7   1   2 |
## 4 | hevy |  130  995  192 |  294 684 150 | -198 310 506 |
```

Taulukot ovatkin aika vaikeita, tulostiedoista! Stargazer toki tekee monenlaista, mutta kun kyse on hyvin yksinkertaisista tulostaulukoista kablen pitäisi toimia.

Kokeillaan summary(smokeCA) - listan dataframe-olioden tulostusta kablella. Voisi harkita funktiota, joka poimii CA:n tuloslistasta sopivat objektit kable-funktiolle? Stargazer taas vaatisi (luultavasti) jonkun ehdollisen tulostuksen (PDF ja html)?

```
knitr::kable( numres1CA1$rows,
    digits = 3,
    caption = 'Korrespondenssianalyysin diagnostiikkaa - rivit', booktabs = TRUE
)
```

Rivien ja sarakkeiden diagnotiikkataulukot eivät mahdu rinnakkain, siksi ne tulostetaan erikseen.

```
knitr::kable( numres1CA1$columns,
    digits = 3,
```

Taulukko 4.4: Korrespondenssianalyysin diagnostiikkaa - sarakkeet

| name | mass | qlt | inr | k=1 | cor | ctr | k=2 | cor | ctr |
|---|---|---|---|---|---|---|---|---|---|
| none | 316 | 1000 | 577 | -393 | 994 | 654 | -30 | 6 | 29 |
| lght | 233 | 984 | 83 | 99 | 327 | 31 | 141 | 657 | 463 |
| medm | 321 | 983 | 148 | 196 | 982 | 166 | 7 | 1 | 2 |
| hevy | 130 | 995 | 192 | 294 | 684 | 150 | -198 | 310 | 506 |

Taulukko 4.5: Korrespondenssianalyysin diagnostiikkaa - ominaisarvot

| | values | values2 | values3 |
|---|---|---|---|
| 1 | 0.075 | 87.756 | 87.756 |
| 2 | 0.010 | 11.759 | 99.515 |
| 3 | 0.000 | 0.485 | 100.000 |

```
    caption = 'Korrespondenssianalyysin diagnostiikkaa - sarakkeet', booktabs = TRUE
)
```

Taulukoiden 4.3 ja 4.4 luvut on kerrottu tuhannella (“per milles”).

Dimensioiden ominaisarvot (eli niiden osuus kokonaisinertiasta) saadaan ca-funktion tuloslistasta taulukoksi. Se esitetään joskus myös ns. scree - kuvana, jos dimensoita on paljon ja joudutaan pohtimaan kuinka monta valitaan (vaikea kysymys!).

```
knitr::kable( numres1CA1$scree,
    digits = 3,
    caption = 'Korrespondenssianalyysin diagnostiikkaa - ominaisarvot', booktabs = TRUE
)
```

Taulukko 4.5 vaatii selityksen, mutta kuvaa ei tässä tapauksessa tarvita.

## 4.3 Kuvat

**chunk-optiot**

fig.cap: R plot - kuvat figure-ympäristöön, automaattiset labelit (fig: + koodipätkän label) ja niihin voi viitata.

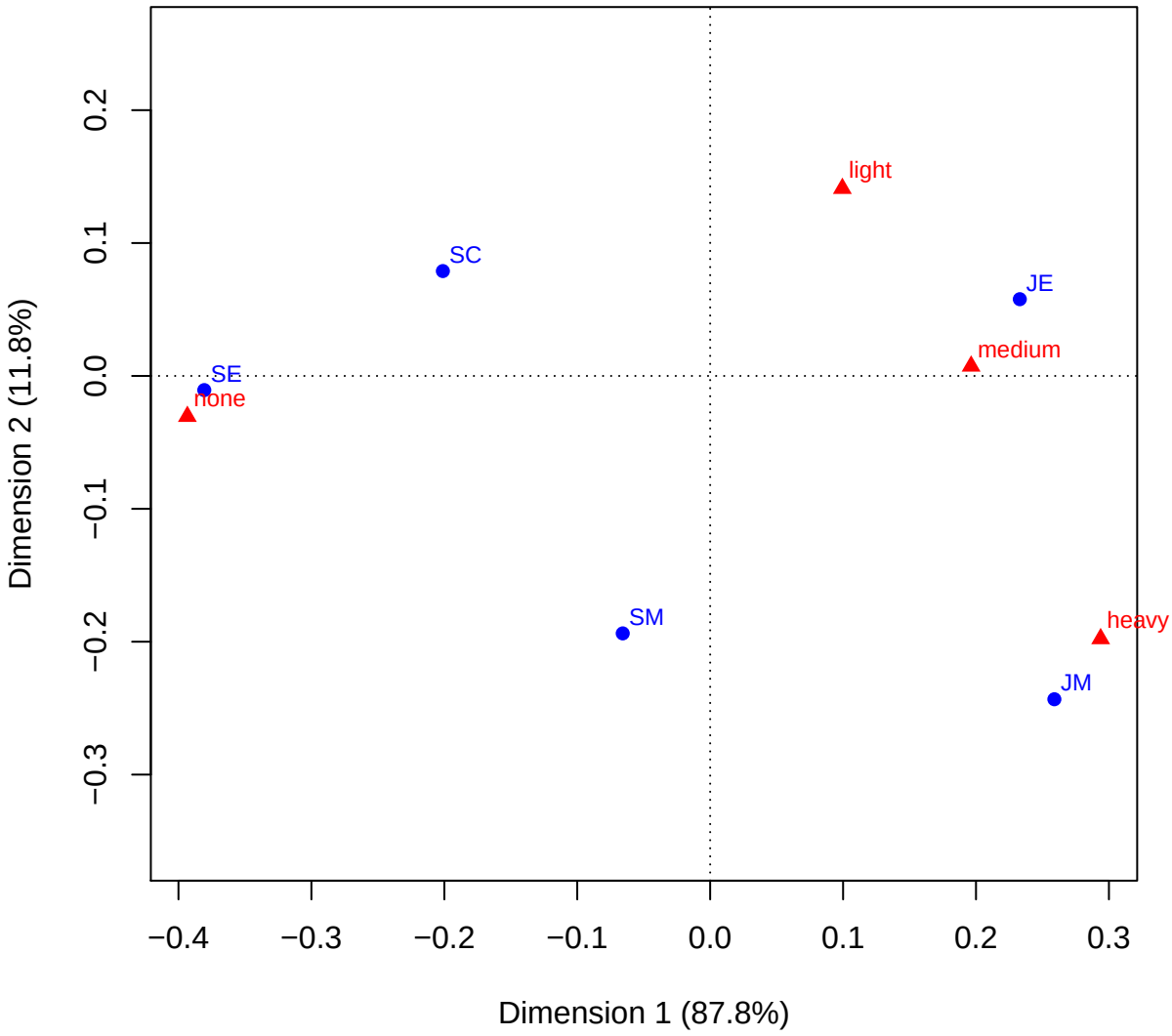
fig.asp oikeaan arvoon 1.

```
plot(smokeCA)
```

Kuviin (kuten 4.1) ja taulukoihin voi viitata tekstissä. Kuvan otsikko tulostuu kuvan alapuolelle, ehkä vähän huono idea?

Näköjään stargazer-kokeilu tulostusoptiolla “html” loi R-projektihakemistoon kansion ja sinne png-kuvan. finnish.ldf tiedoston muokkaus MikTeX:ssä tehty, mutta se ei vaikuta html-viiteotsikkoon. Korjattu “ehdollisessa viitesivussa” viitteet.Rmd jossa html-viiteluettelon otsikko annetaan.

Saisiko numeeristen tulosten scree-kuvan samalla tavalla kuvaksi?




Kuva 4.1: CA-kartta

# Luku 5

# Bookdown ja Rmarkdown

Bookdown- R-paketti “paketoi” RMarkdownin tulostutoiminnot (output) ja sen monet säädettävät optiot. Samat Rmd-dokumentit saadaan koottua moneen eri formaattiin: html- sivuiksi, PDF-dokumentiksi tai Ebook-kirjaksi. Kaikissa tulostusvaihtoehdoissa on monia eri vaihtoehtoja. Html-tulostuksessa voi valita yhden tai useamman html-sivun lisäksi gitbook- tai Tufte- vaihtoehdon. Ne on toteutettu css-tyylitiedostoilla ja JavaScript-kirjastoilla. Tässä on käytetty gitbook-formaattia.

LaTeX-formaatti renderöidään jollain LaTeX-vaihtoehdolla PDF-tiedostoksi. **ToDo** PDF-formaattejakin on useita variantteja, mikä niistä. Tässä vaihtoehdossa konfigurointimahdollisuudet ovat käytännössä rajattomat, sillä välitulosteena syntyvää TeX-tiedostoa voi muokata ja muuntaa sen sitten PDF-muotoon.

Prosessissa on monta vaihtetta, ja eri parametrien yhteisvaikutusta on vaikea hahmottaa.

```
knitr::include_graphics('BookdownProc.png')
```

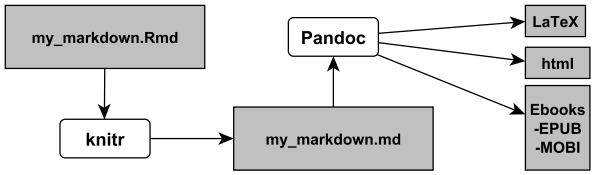
Perusopas bookdown paketin käyttöön on Yihui Xien “bookdown: Authoring Books and Technical Documents with R Markdown”". Siinä pääidea on tuottaa yhdellä Rmd-koodilla kuvan **??** kolme vaihtoehtoista tulostiedostoa mahdollisimman yksinkertaisesti. Knitr- ohjelma “kutoo” Rmd-tiedoston r-koodilohkojen tulokset ja tekstin markdown-tiedostoksi (md). Rmd-tiedostojen YAML-asetukset siirtyvät Pandocille, joka täydentää niillä omia mallitiedostojaan (template).

Laajempi ja tarkempi opas ilmestyi 15.7.2018, kolmen kirjoittajan “R Markdown: The Definitive Guide”. Siinä eri asetusten hierarkia on kuvattu tarkemmin ja selkeämmin. Tulostusvaihtoehtoja esitellään laajemmin, bookdown on vain yksi luku.

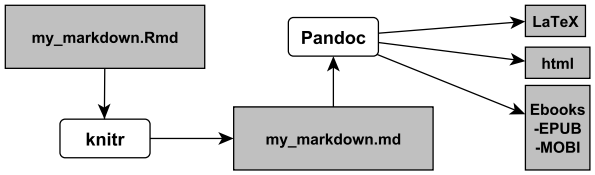
R Studiolla alkuun pääse helposti, kun lataa bookdown-paketin, ja luo uuden bookdown-projektin. Xien ensimmäisen kirjan alku-luvut ja uudemman teoksen johdattelut auttavat jatkoon.

**Käytännön vinkkejä**

1. Kuvasuhde pitää olla 1:1 . Ehkä hankalin juttu Rmarkdownin kanssa työskennellessä, mutta aina voi avata oman grafiikkaikkunan. Dataa analysoidessa voi tallentaa kuvat pdf-muodossa, lisäillä kommentteja




Kuva 5.1: Tulostiedoston prosessointi - png




Kuva 5.2: Tulostiedoston prosessointi - pdf

yms. Lopullisessa dokumentissa kuvasuhden pitää erikseen tarkista, säätämiseen on monta vipua.

2. Bookdown-työskentelyssä pdf-tuloste ei ole kätevä, yleensä analyysiä hiotaan Rmd-tiedosto kerrallaan. R Studio voi yllättää aina joskus! Knit-napin takaa löytyy kuitenkin eri renderöintifuktiot kuin oikean laidan yläikkunan "Build Book" - valikosta. Knitr-funktiota kannattaa käyttää, jos haluaa katsoa yhden Rmd-tiedoston tulostetta. Tarkista kuitenkin, että (a) Rmd-tiedostoon ei automaattisesti lisäillä YAML-headereita ja (b) projektin hakemistoon ei ilmesty ylimääräisiä Rmd-tiedostoja. Joskus bookdown R Studion kanssa kasaa yhden Rmd-tiedoston tulostuksessa "väliaikaiseksi" tiedostoksi koko dokumentin yhteen .Rmd -tiedostoon. Jos bookdown-projektiin kuuluvia Rmd-tiedostoja ei eksplitsiittisesti luetella (suositeltavaa, laita bookdown.yml - tiedostoon lista) syntyy hassua sotkua.
3. Koko raportin tulostus html-muodossa käy kätevimmin "Build book" - valikon html-book- funktiolla/formaatilla. Tämä pitää tsekata! (4.12.18)
4. Suositus: koko dokumentit tulostukset aina "puhtaalta pöydältä", käynnistä R uudelleen. Myös silloin, kun tulostat ensin vaikka gitbookin ja sitten pdf-tiedoston.

Windows-ympäristössä (Windows 10) MikTeXin kanssa voi tulla ongelmia, jos käytät konetta tavallisen käyttäjän oikeuksilla. Bookdown-paketin kanssa on kätevää käyttää tinytex - r-pakettia, ja konfiguroida oman koneen MikTeX - asennus asentamaan tarvittavat paketit "lennossa". Peruskäyttäjän omat paketit voivat vaatia päivitystä, mutta oikeudet eivät riitä. Pulman voi ratkaista, kun käynnistää MikTeXin paketinhallintasovelluksen (jolla on monta nimeä, admin console jne) peruskäyttäjänä, ja katsoo mitä päivityksiä on tarjolla. Nämä paketit voi sitten asentaa admin-oikeuksilla.

Kokeillaan vielä PDF-kuvan liittämistä dokumenttiin. Ei näy html-tulosteessa.

```
knitr::include_graphics('BookdownProc.pdf')
```

Testataan koodilohkojen listausta, näyttää toimivan mutta vaatii vielä säätämistä. Ohje löytyi Yihui Xienin blogista (luettu 26.10.2018).

```
#pitääkö kirjastot ladata tässä, vai jokaisen rmd-tiedoston alussa?
library(rgl)
library(ca)
library(haven)
library(dplyr)
library(knitr)
library(tidyverse)
library(lubridate)
library(rmarkdown)
library(ggplot2)
library(furniture)
library(likert)
library(scales) # G_1_2 - kuva
library(reshape2)  # G_1_2 - kuva
library(printr) #19.5.18 taulukoiden ja matriisien tulostukseen
library(stargazer) # 28.5.2018 taulukoiden yms. tulostietojen siistiin tulostukseen
library(bookdown)
```

```r
library(tinytex)
system("pdflatex --version")
#getwd()

rmarkdown::pandoc_version()

knitr::kable(smoke[,1:4], booktabs = TRUE,
  caption = 'CA-paketin smoke-data (keinotekoinen)'
)
# Taulukkoon viittaaminen tekstissä \@ref(label)
# riviprofiilit
smoke.rpro <- smoke / rowSums (smoke)
# keskiarvoprofiili
smoke.avrpro <- colSums(smoke) / sum(smoke)

knitr::kable(
  list(smoke.rpro, t(smoke.avrpro)   ), digits = 3,
  caption = 'Riviprofiilit ja keskiarvoprofiili', booktabs = TRUE
)

smokeCA <- ca(smoke)
#temp1 <- smokeCA tämä kai tarpeeton ? (4.12.2018)
numres1CA1 <- summary(smokeCA)
#str(smokeCA)
#knitr::kable( smokeCA,
#  digits = 3,
#  caption = 'Riviprofiilit ja keskiarvoprofiili', booktabs = TRUE
#)
#str(temp1)
#stargazer(temp2$rows, type = "text", title = "CA-tuloksia")
# LateX-tulostuksessa float vaatii jotain tällaista:Table: (\#tab:cataul1)
#str(temp2)
#str(temp2$scree)
#temp2$scree
numres1CA1

knitr::kable( numres1CA1$rows,
    digits = 3,
    caption = 'Korrespondenssianalyysin diagnostiikkaa - rivit', booktabs = TRUE
)

knitr::kable( numres1CA1$columns,
    digits = 3,
    caption = 'Korrespondenssianalyysin diagnostiikkaa - sarakkeet', booktabs = TRUE
)

knitr::kable( numres1CA1$scree,
    digits = 3,
    caption = 'Korrespondenssianalyysin diagnostiikkaa - ominaisarvot', booktabs = TRUE
)

plot(smokeCA)
```

```
str(numres1CA1$scree)
test2 <- as.table(numres1CA1$scree)
#str(test1$V1)
str(test2)
test2[[dimnames]]
# Vielä kokeilua!


knitr::include_graphics('BookdownProc.png')
knitr::include_graphics('BookdownProc.pdf')
```

“New line” vaaditaan koodilohkon jälkeen.

# Lähteet


Ahlgren, A. (1994). *Öljyntuotanto Hämeessä - outo idea*. Kluwer Academic Publishers, Dordrecht.

Greenacre, M. and Hastie, T. (1987). The geometric interpretation of correspondence analysis. *Journal of the American Statistical Association*, 82(398):437–447. doi: 10.1080/01621459.1987.10478446.

Greenacre, M. J. (2017). *Correspondence analysis in practice*. CRC Press, Boca Raton, Florida, third edition edition.

Älli Åhlgren (1994). *Öljyntuotanto Hämeessä*. Kluwer Academic Publishers, Dordrecht.

Roux, B. L. and Rouanet, H. (2004). *Geometric data analysis: from correspondence analysis to structured data analysis*. Kluwer Academic Publishers, Dordrecht.