

# G Luku 1 Yksinkertainen korrespondenssianalyysi

Jussi Hirvonen

versio 1.5.5 dev , tulostettu 2020-02-02

## Sisältö

<b>1</b>	<b>Data</b>	<b>4</b>
1.1	Luvun 1 tavoitteet . . . . .	4
1.2	Perhe ja muuttuvat sukupuoliroolit - ISSP:n kyselytutkimuksen data 2012 . . . . .	5
1.3	Substanssimuuttajat, taustamuuttajat, muut . . . . .	6
1.4	Aineiston rajaaminen . . . . .	7
1.5	Datan valinnan vaiheet ja puuttuvat tiedot . . . . .	21
1.6	Perusmuunnoksesta ISSP2012 - datalle . . . . .	23
<b>2</b>	<b>Yksinkertainen korrespondenssianalyysi - kahden luokittelu- muuttujan taulukko</b>	<b>57</b>
2.1	Äiti työssä . . . . .	59
2.2	Korrespondenssianalyysin käsitteet . . . . .	79
<b>3</b>	<b>Tulkinnan perusteita</b>	<b>79</b>

**Versiot - vanha Galku - 5.6.2019 versio 1.5.1 Uusi Galku - 2.2.2020 versio 1.5.5**

Siivotaan datan käsittelyn koodilohkot, kopioidaan mahdollisesti hyödylliset koodipätkät tiedostoon siivous1.R (30.1.2020).

Uudet datan luku- ja muunnosskriptit (treeni2-projektista), korjaillaan virheitä ja editoidaan koodia.(31.1.2020)

(2.2.20) Toimii johdattelevaan esimerkkiin asti, myös PDF-tulostus. Kuvien otsikot vähän mitä sattuu, ja ´profiilikuviin asti maa-muuttujan järjestys ´väärä”, ts. eri kuin vanhemmissa versioissa. Korjattava (**TODO 2.2.20**)

## HISTORIAA

**6.8.2018 versio 1.0**

**Siistitään -> 12.8.2018 versio 1.05**

**Kommentit ja korjaukset -> 4.9.2018 versio 1.1**

puuttuva riviprofilikuva, siistimmät interaktiomuuttujien koodaukset, ensimmäinen “pinottu taulu” - analyysi -> 19.9.2018 versio 1.2

**25.9.2018 siistitään datan käsittelyä; ei huomioida puuttuvan tiedon tarkempaa koodausta (read\_spss - funktion user\_na = TRUE asetus)**

**1.10.2018** Versio 1.3

Muutokset tarkemmin Readme.md - tiedostossa.

Uusi jakso yksinkertaisen CA:n laajennuksille, joissa otetaan analyysiin useampia muuttujia “pinoamalla” ja/tai yhdistämällä taulkoita. Tässä jaksossa otetaan myös käyttöön isompi aineisto (enemmän maita ja muuttujia). Siisti koodipätkä täydentävien muuttujien lisäämiseen.

**3.10.2018** Versio 1.4

Siistitään pois turhat datan listaukset. Aineiston rajaaminen selkeäksi. Ensin kuusi maata, sitten 27 (Espanja pois). Valitaan myös muuttujat, jotta käsiteltävän datan listaukset ovat järkevämpiä. Aineistossa esim. Espanjan ja muutaman Unkarin poikkeavien vastausvaihtoehtojen vastaukset ovat omina muuttujina, ja niiden arvo muille havainnoille on NAP (Not applicable). Samoin paljon maa-kohtaisia muuttujia, esim. koulutustaso. Mukaan otetaan vain kv-vertailuihin kelpaavat muuttujat, muutama sellainen on myös aineistoon rakennettu. Jätetään pois kaikki perhesuhteisiin liittyvät kysymykset (esim. kotitöiden jakaminen) ja taustatiedot (esim. rahankäyttö, puolison eri tiedot jne.), koska muuten jouduttaisiin miettimään miten näiden osalta käsitellään perheettömiä. Muutamia muuttujia otetaan mukaan (lasten lkm jne.).

**8.10.2018**

Datan valinta. Data-jaksossa aluksi, voi miettiä siirtääkö esimerkki-lukuun ja “pinotut taululut” - luvun alkuun kuvailut. Tavallaan siistiä, jos alussa lyhyesti.

**10.10.2018**

Maiden ja muuttujien valinta. TOPBOT halutaan mukaan, joten USA ja GB on jätettävä pois. Muuttuja on kuitenkin hankala, usealla maalla puuttuva tieto yli 10 prosentissa, ja muutamalla nolla tai ihan muutamia. Pohditaan aikanaan. **5.11.2018** Puuttuvat tiedot ovat puuttuvia, ei voi mitään. Jos vähän ja selviää virheitä (ikä, sukupuoli), voidaan pudottaa havainnot. Muuten mukaan, periaatteessa.

Data-jaksosta siirretään aineiston laajentamisen yhteyteen laajemman muuttujajoukon deskriptiiviset tarkastelu. Taulukko muuttujakuvauksesta jää data-lukuun.

**5.12.18** Puuttuneisuuden taulukointia on, mutta siisti NA-tila taulukko puuttu.

**11.10.2018** Versio 1.4

- paperitilosteessa v1.3 kommentteja karttoihin ja ca:n numeerisiin tuloksiin, samoin muuttujalistauksiin.
- paperitilosteessa v1.4 samoin, ja puuttuneisuuden taulukointeja

**11.10.2018 aloitetaan versio 1.5** - pieniä muutoksia ja kommentteja, aloitetaan uusi versio 1.51 5.12.2018

**6.12.2018 1.5.1** - as\_factor - funktio käyttöön; testaillaan miten toimii kun (a) user\_na - arvoja ei lueta ja (b) puuttuvat ovat mukana.

#### **Muistilista:**

1. Taulukot ja kuvat luvusta 2. alkaen eivät ole "bookdown-muodossa". CA-tulokset on tulostettu siiteinä taulukoina Bookdown-demo - dokumentissa. Voi tulostaa myös ca-outputin. Ominaisarvojen taulukko keskeneräinen, samoin "scree plot" kuvana puuttuu.
2. Osa kuvista (esim. profilikuva) pitää varmaan tulostaa pdf-muodossa ja ottaa capaper-dokkariin include\_graphics - funktiolla.
3. Puuttuvia tai mahdollisesti lisättäviä taulukoita (nämä saa ca-funktion tuloksista suoraan)
  - khii2 - etäisyydet riveille ja sarakkeille - on tulostettu ilman muotoiluja (11.10.18)
  - massoilla painotetut khii2-etäisyyden keskiarvorivistä/sarakkeesta?
4. Kuvissa vielä hiottavaa, pdf-kuvia lisäilty img-hakemistoon.
5. Data-tiedostojen nimeäminen (27.12.18)

**\*\*ISSP2012\*.data\*\*** - täysi aineisto

**\*\*ISSP2012\*jh1.data\*\*** - valikoitu aineisto (maat, muuttujat)

**\*\*ISSP2012\*esim1.dat\*\*** - muuttujien muunnoksia ja uusia muuttujia; analyysissä käytettävä data, tarkenne dat.

6. kasitteet1.rmd - taulukko käsitteistä ja tärkeimmistä ISSP-dokumenteista

#### **Historiaa** (11.10.18)

Vanhoja kommentteja

- kirjastot/paketit ladataan jokaisessa Rmd-dokumentissa
- bib-formaatin viitetietokantaa tullaan kokeilemaan
- kuvasuhde (aspect ratio) edelleen epäselvä juttu! Mutta näyttää PDF-tulosteessa olevan ok.
- Datien käsittely ja hallinta +SPSS:n sallima kolme puuttuvan tiedon koodia saadaan mukaan read\_spss-funktion (haven) parametrilla USER\_NA = TRUE (mutta tarkistettava!) (25.4.18)
  - faktoreita ei ainakaan toistaiseksi muuteta ordinaaliasteikolle, CA ei tästä välitä
  - pidetään muuttujien ja tiedostojen nimeäminen selkeänä, tarkistetaan aika ajoin
- Taulukot: lisättiin riviprosentti- ja sarakeprosenttitaulut (25.4.18), kuva riviprofileista puuttu vielä (15.5.2018)

- Datan esittelyssä on turhaa välitulostusta, ja samoin vähän muuallakin. Html on helpompi lukea, kun koodi on oletuksena piilossa
- PDF-tulosteessa koodi pääsääntöisesti näkyy toistaiseksi
- kokeiluja CA-karttojen tulostamiseen (a) suoraan koodilla ja (b) r-grafiikkaikkunasta tallennetun pdf-kuvan avulla. Paras toistaiseksi (a), jätin kokeilu näkyviin. Analyysit R:n grafiikkaikkunassa, jotta asp=1, ja tulkintaa varten voi tallentaa PDF-muodossa.
- rakenteeseen muutoksia (näkyvät sisällysluettelossa), ei erillistä teorialiitettä vaan sopivina annoksina. Lukuun 3 perusasiat, kaavat, määritelmät
- tehdään käsitetaulukko (kirjoittamista varten)
- 20.5.2018 (a) tulkita-osuuteen karttakuvia ja ca-tulokset (b) siistimpi taulukoiden tulostus löytyi (c) kaavaliite laajeni (dispo-haarassa)
- 23.5.2018 lisätään dataan toinen maa-muuttuja maa2, ikäluokkamuuttuja age\_cat ja iän ja sukupuolen vuorovaikutusmuuttuja ga.
- 24.5.2018 lisättiin ca-kartta, jossa Saksan ja Belgian ositteet ja summarivit täydentävinä (passiivisina)

---

## 1 Data

**edit** tässä luvussa on paljon siistittävää, mutta data on ok. (13.5.2018). **edit** capaper - dokumentissa parempi uusi jäsentely (4.9.2018) **edit** ISSP-datan perustietoa dokumentissa ISSP\_data1.docx (4.9.2018) **edit** koodilohkoja ei vielä siistitä, eikä nimetä capaper-vaatimusten mukaan. **edit** Poistetaan aineistosta havainnot, joissa puuttuva tieto muuttujassa SEX tai AGE

**edit 24.9.18 Poistettiin turhaa, uusi versio tiedostosta (G1\_1\_data1.Rmd -> G1\_1\_data2.Rmd).**

**edit 30.1.20** Siivotaan, luodaan faktori-muuttujat heti alussa koko datalle. Uusi G1\_1\_data\_fct1.Rmd tekee muunnokset.

### 1.1 Luvun 1 tavoitteet

**Datan esittely ja kuvailut - tämä luku täysin uusiksi (24.9.18)**

**10.10.2018 maat ja muuttujat valittu.**

1. Eksploratiivinen ja graafinen menetelmä tarvitseen aineiston, hankalaa esitellä jollain synteettisellä esimerkkiaineistolla. **edit** Eksp&graaf menetelmät määriteltävä johdantoluvussa. Esimerkkiaineistoja (synteettisiä kuten smoke, myös muita) on mm. ca - paketissa.
2. CA (ja MCA) sopivat isojen moniulotteisten ja mutkikkaiden aineistojen analyysiin, siksi iso aineisto. Samalla analyysiä voi laajentaa moneen suuntaan. **V** Benzecri: "kun data menee miljoonaan suuntaan".
3. Aineiston esittely, laajan kyselytutkimusaineiston tyypilliset ominaisuudet

4. Laadukkaan ja hyvin dokumentoidun aineiston edut
5. Huom! CA sopii ja sitä on käytetty myös hyvin toisen tyyppisiin aineistoihin (ekologia ja biologia, arkeologia, kielen tutkimus)

## 1.2 Perhe ja muuttuvat sukupuoliroolit - ISSP:n kyselytutkimuksen data 2012

Hieman historiaa datasta, sosiaalisesti määräytyneen sukupuoliroolit (gender) tutkimusaiheena neljässä kansainvälisessä kyselytutkimuksessa. **luvun pitäisi olla mahdollisimman lyhyt (5.12.18)**

### Tärkeät linkit

Toimivat html-tulosteessa, PDFtiedostoissa saa toimimaan (vaati tarkat formoinnit Rmd-koodissa).

[www.issp.org](http://www.issp.org), tutkimushankkeen historiaa. Löytyy myös bibliografia tutkimuksista, joissa aineistoja on käytetty.

[www.gesis.org](http://www.gesis.org) - tutkimuksen “sihteeristö”, dokumentaatio ja datat.

data ja dokumentaatio (selattavissa): [zacat.gesis.org](http://zacat.gesis.org)

**edit** tässä järkevä viite ISSP - dataan ISSP Research Group (2016): International Social Survey Programme: Family and Changing Gender Roles IV - ISSP 2012. GESIS Data Archive, Cologne. ZA5900 Data file Version 4.0.0, doi:10.4232/1.12661 **tämä doi-linkki ei toimi**

### Linkitys on hankalaa

- monta portaalia, joista pääsee monien organisaationimien taakse
- tästä lyhyt selostus
- tärkeimmät linkit ISSP-tutkimuksen “kotisivu” ja selkeä **muuttujakuvaukset ja muut tiedot**
- käytännössä linkittäminen “syvälle” johonkin sivustoon tai www-palveluun ei ole järkevää, parempi antaa selkeät viitetiedot ja tiedot organisaatioista. Ne kyllä säilyvät, tai jäljille pääsee.

**Edit** Refworksiin on kerätty viitteitä, tässä pärjätään kolmen saitin osoitelilla. Voi laittaa taulukon tärkeimmistä dokumenteista, tarvittaessa liitteeksi (tiedostonimet ja kuvaus). Alla linkkejä jotka eivät näy PDF-tulosteessa, lisätty tekstinä.

Aineistot <https://dbk.gesis.org/dbksearch/sdesc2.asp?no=5900&db=e2012> **toimii**

[Muuttujakuvaukset ja muut tiedot] (<http://zacat.gesis.org/webview/index.jsp?object=http://zacat.gesis.org/obj/fStudy/ZA5900>) <http://zacat.gesis.org/webview/index.jsp?object=http://zacat.gesis.org/obj/fStudy/ZA5900> **OK - täältä löytyy oikeastaan kaikki!** Suomenkielinen lomake (ZA5900\_q-fi-fi.pdf)

löytyy helpoiten täältä, samoin muu dokumntaati tiedostoina. Veppisivuilla kerrotaan, mitä ne dokumentit ovat.

Data ja dokumentit **vie vain aineiston dokumentoinnin etusivulle** <https://dbk.gesis.org/dbksearch/sdesc2.asp?no=5900&db=e>

Käyttöehdot: **GESIS-palvelun datan yleiset käyttöehdot, viittauskäytännöt**

Havaintojen lukumäärät voi tarkistaa täältä <http://zacat.gesis.org/webview/index.jsp?object=http://zacat.gesis.org/obj/fStudy/ZA5900>. **Dokumentointisivusto/katalogi, jossa helppo navigoida** [zacat.gesis.org](http://zacat.gesis.org).

Dokumentointi on hyvin tarkka, tiedot löytyvät haastattelumenetelmistä (parerilomake, tietokoneavusteinen haastattelu, jne), maakohtaisten taustamuuttujien harmonisoinnista maittain, otantamenetelmistä jne. Esittelen vain aineiston tärkeimmät rajaukset. MOnitorointiraportti kertoo puuttuneisuuden määrän, otantamenetlmät jne maittain. “Code book” kertoo muuttujien määritelmät sekä yhteisille että maakohtaisille muuttujille. Kaikista muuttujista on taulukko maittain.

```
issp_docname <- c("Variable Report", "Study Monitoring Report", "Basic Questionnaire",
                  "Contents of ISSP 2012 module")
issp_docdesc <- c("Perusdokumentti, muuttujien kuvaukset ja taulukot",
                  "tiedokeruun toteutus eri maissa",
                  "Maittain sovellettava kyselylomake", "substanssikysymykset taulukkona")
issp_docfile <- c("ZA5900_cdb.pdf", "ZA5900_mr.pdf", "ZA5900_bq.pdf", "ZA5900_overview.pdf")

ISSPdocsT.df <- data_frame(issp_docname, issp_docdesc, issp_docfile)

## Warning: `data_frame()` is deprecated, use `tibble()`.
## This warning is displayed once per session.

col_isspdocs <- c("dokumentti", "sisältö", "tiedosto")
colnames(ISSPdocsT.df) <- col_isspdocs

knitr::kable(ISSPdocsT.df, booktab=TRUE)
```

dokumentti	sisältö	tiedosto
Variable Report	Perusdokumentti, muuttujien kuvaukset ja taulukot	ZA5900_cdb.pdf
Study Monitoring Report	tiedokeruun toteutus eri maissa	ZA5900_mr.pdf
Basic Questionnaire	Maittain sovellettava kyselylomake	ZA5900_bq.pdf
Contents of ISSP 2012 module	substanssikysymykset taulukkona	ZA5900_overview.pdf

### 1.3 Substanssimuuttujat, taustamuuttujat, muut

**zxy** capaper - dokumentissa uusi jäsentely (4.9.2018)

**zxy** Aineiston luonne: maakohtaisesti eri tavoin kerätty data, jossa pyritään yhtenäisiin käytäntöihin ja tietosisältöihin. Silti myös substanssikysymyksissä eroja, isoja ja pienempiä. Näin vain on, en pohdi miksi. Ei ole mitenkään ainutlaatuista. Aineiston editoinnissa ja tiedonkeruun suunnittelussa on nähty paljon vaivaa vertailukelpoisuuden vuoksi. Tästä esimerkkejä, esim. “mitä puoluetta äänestit”.

**zxy** yksi kappale: Aineitoa on harmonisoitu, kysymyksiä hiottu, vertailukelpoisuuteen on pontevasti pyritty. Silti eroja löytyy, osa ymmärrettäviä (lisäkysymykset jne) ja osa ei (Espanja!). Tällaista on kansainvälisen kyselytutkimuksen data.

Parempi muotoilu: Varsinaiset substanssimuuttujat eli kyselylomakkeet on koitettu hioa mahdollisimman yhdenmukaisiksi. Silti pieniä eroja löytyy, ja isojaakin (Espanja on pudottanut neutraalin “en samaa enkä eri mieltä” - vaihtoehdon pois, ja Unkarissakin on muuttamat vastausvaihtoehdot valittu omalla tyylillä). Taustamuuttujissa on pyritty samaan, ja aineistoon on myös rakennettu kansainvälisesti vertailukelpoisia muuttujia kansallisesti kerätyistä tiedoista. Näitä ovat erityisesti tuloihin liittyvät tiedot, ja mone muutkin. Muuttujat jakautuvat substanssi- ja taustamuuttujiin, ja taustamuuttujista monet tiedot on kerätty kansallisiin aineistossa maan kirjantunnisteella alkaviin muuttujiin.

**zxy** HUOM! Dataa ei ole kerätty vain kansainvälisiin vertailuihin! Sitä voi ja ehkä pitäisikin analysoida maa kerrallaan, ja vertailla näitä tuloksia. (#V Blasiuksen artikkeli, jossa arvioidaan yhden ISSP-tutkimuksen vertailukelpoisuutta. Kysymykset eivät kovin hyvin näytä toimivan samalla tavalla eri maissa.)

## 1.4 Aineiston rajaaminen

### 1. Eurooppa ja samankaltaiset maat (28 aluksi -> 25)

(**Pois 13:** Argentiina, Turkki, Venezuela, Etelä-Afrikka, Korea, Intia, Kiina, Taiwan, Filippiinit, Meksiko, Israel, Japani, Chile.)

Bulgaria, Czech Republic, Denmark, Finland, France, Germany, Great Britain, Ireland, Latvia, Lithuania, Norway, Poland, Sweden, Slovakia, Slovenia, Spain, Switzerland, Australia, Austria, Canada, Croatia, Iceland, Russia, United States, Belgium, Hungary, Netherlands, Portugal (**28**) - **Espanja, Great Britain, USA pois -> 25 (11.10.18)**

### 2. Maat joissa varsinaisissa tutkimuskysymyksissä on käytetty poikkeavia luokitituksia tms.

Espanja pois, USA ja GB pois koska TOPBOT-muuttuja puuttuu (11.10.18)

### 3. kaikki havainnot, joissa on puuttuvia tietoja.

Johdattelevassa esimerkissä on kolme muuttujaa, ei ongelma, aika vähän puuttuvia.

Isomman 25 aineiston osalta tarkistetaan, mitä “listwise deletion” saa aikaan. Aineisto pienenee nopeasti, ja vaikeasti hahmotettavalla tavalla. Tämä erävas-  
tauskato ei ole tutkielman ydinaihe, mutta laajemman aineiston käytössä sitä ei  
voi ohittaa. Yksikkövastaukskatoa ei käsitellä.

CA:n eräs etu on se, että muuttujien oletetaan olevan luokitteluasteikon (nomi-  
naaliasteikon) muuttujia, ja puuttuva havainto on yksi luokka lisää. Puuttuvat  
havainnot otetaan mukaan, niitä ei poisteta aineistosta.

Yksittäisten vastausten puuttuminen eli erävasauskato ohitetaan aluksi, mut-  
ta siihen palataan. Korrespondenssianalyysiin on helppo ottaa mukaan myös  
puuttuvat tiedot, sillä data on luokitteluasteikon dataa.

Yksikkövastaukskato eli otokseen poimitut joita ei ole tavoitettu ollenkaan on  
kansallisen tason ongelma, joka on ratkaistu vaihtelevin tavoin. Tiedot löytyvät  
aineiston dokumentaatiosta. Aineistossa on myös mukana maakohtaiset paino-  
muuttujat, mutta ei painoja maiden vertailuun. Vastausprosentit (response rate)  
vaihtelevat maittain, kts. monitoring report. Aika tyypillisiä, ei kauhean huonoja  
paitsi muutama.

**edit** toistoa (30.1.20)

#### 4. Datan hallinta **liittyy reproducible research- periaatteeseen**

Aineistoa käsitellään ja muokataan niin, että jokaisen analyysin voi mahdollisman  
yksinkertaisesti toistaa suoraan alkuperäisestä datasta.

1. Valitaan maat ja muuttujat
2. Poistetaan havainnot, joissa puuttuva tieto sukupuolesta tai iästä (tai  
molemmista)

Aineiston muokkauksen (muuttujien ja havaintojen valikointi, muunnokset ja  
uusien muuttujien luonti jne.) dokumentoidaan r-koodiin.

**zxy 3.10.18** R-spesifiä: R-koodissa tarkemmin, kaikki yksityiskohdat.

Kun SPSS-tiedosto luetaan R:n data frame - tiedostoksi, mukana tulee myös  
metadata. Uusien muuttujien luonnissa tai data-formaatin vaihtuessa (esim.  
matriisiksi, taulukoksi jne) metadata katoaa. Siksi muuttujien tyyppimuunnokset  
(yleensä faktorointi) tallennetaan uusiksi muuttujiksi, metatieto säilyy vanhassa  
muuttujassa.

Helposti toistettava tutkimus: polku alkuperäisestä datasta analyysien dataan  
selkeä (ja lyhyt jos mahdollista).

Puuttuva tieto voidaan koodata monella tavalla (ei halua vastata jne), ja SPSS  
(datan jakelutiedosto) sallii kolme koodia puuttuville tiedoille. Ne voi lukea R-  
dataan, mutta puuttuneisuutta ei tässä työssä tutkita sen tarkemmin. Detaljit  
R-koodissa (haven-paketin read\_spss-funktion user\_na -optio).

**Tiedostonimistä** (10.10.18, 30.1.20)



ISSP2012.data - täysi aineisto ISSP2012jh1.data - valittu osa aineistosta (yleensä: maita) ja muuttujista ISSP2012\*.jh1.dat - valittu osa aineistosta, luotu uusia muuttujia ja muunnettu muuttujia ISSP2012esim1, 2 jne, tarkenne .dat muunnettuja muuttujia, rajattuja aineistoja jne

ISSP2012.data (df) jossa alkuperäinen SPSS-data ISSP2012jh1.data osajoukko edellisestä ISSP2012jh1a.data - valitaan maat jne. Kerrottu alempana.

ISSP2012esim1.dat edellisen osajoukkoja, joissa uusia muuttujia ja tyyppimuunnoksia. Nämä vaihtuvat analyysin vaiheen mukaan, jotta polku olisi lyhyt. Jaksot erillisiä Rmd-tiedostoja, jokaisen alussa ladataan r-paketit ja data. Tallennetaan datan lukukoodi omaksi tiedostoksi, näin on jo tehty paketeille (paketit.R)

Muuta: alkuperäinen muuttuja säilytetään, voi tarkistaa

**zxy** R-koodiin jätetään myös tarkistuksia yms. joita ei raportoida tässä, samoin niiden tuloksia. Voiko R-koodi olla fingelskaa? Olkoon toistaiseksi.

#### DATA RAJAAMISTA - maat(5.10.2018)

```
# Aineiston rajaamisen kolme vaihetta (10.2018)
#
# TIEDOSTOJEN NIMEÄMINEN
#
# R-datatiedostot .data - tarkenteella ovat osajoukkoja koko ISSP-datasta ISSP2012.data
# R-datatiedostot .dat - tarkenteella: mukana alkuperäisten muuttujien muunnoksia
# (yleensä as_factor), alkuperäisissä muuttujissa mukana SPSS-tiedoston metadata.

# Muutetaan R-datatiedossa alunperin ordinaali- tai nominaaliasteikon muuttuja haven-paketin
# as_factor - funktiolla faktoreiksi. R:n faktortyyppin muuttujille voidaan tarvittaessa
# määritellä järjestys, toistaiseksi niin ei tehdä (25.9.2018).
# Luokittelumuuttujan tyyppi on datan lukemisen jälkeen yleensä merkkijono (char), vaikka se
# kokonaisluvuksi. R:n ns. "implisiittinen konversio" muuntaa arvot merkkijoiksi, jos joukossa
# havaintoja (NA)
#
# Muunnetun muuttujan rinnalla säilytetään SPSS-tiedostosta luettu muuttuja, metatiedot säilytetään
# alkuperäisessä.
#
# R-datatiedostot joiden nimen loppuosa on muotoa *esim1.dat: käytetään analyyseissä
#
# 1. VALITAAN MAAT (25) -> ISSP2012jh1a.data. Muuttujat koodilohkossa dataset_vars1
#
# kolme maa-muuttujaa datassa. V3 erotelee joidenkin maiden alueita, V4 on koko maan
# ja C_ALPHAN on maan kaksimerkkinen tunnus.
#
# V3 - Country/ Sample ISO 3166 Code (see V4 for codes for whole nation states)
# V3 erot valituissa maissa
# 5601 BE-FLA-Belgium/ Flanders
```

```

# 5602 BE-WAL-Belgium/ Wallonia
# 5603 BE-BRU-Belgium/ Brussels
# 27601 DE-W-Germany-West
# 27602 DE-E-Germany-East
# 62001 PT-Portugal 2012: first fieldwork round (main sample)
# 62002 PT-Portugal 2012: second fieldwork round (complementary sample)
# Myös tämä on erikoinen, näyttää olevan vakio kun V4 = 826:
# 82601 GB-GBN-Great Britain
# Portugalissa ainestoa täydennettiin, koska siinä oli puutteita. Jako ei siis ole oleellinen
# mutta muut ovat. Tähdellä merkityt maat valitaan johdattelevaan esimerkkiin.
#
# Maat (25, ei Espanjaa).Myös USA ja Iso-Britannia pois, koska muuttuja TOPBOT puuttuu.
#
# 36 AU-Australia
# 40 AT-Austria
# 56 BE-Belgium*
# 100 BG-Bulgaria*
# 124 CA-Canada
# 191 HR-Croatia
# 203 CZ-Czech Republic
# 208 DK-Denmark*
# 246 FI-Finland*
# 250 FR-France
# 276 DE-Germany*
# 348 HU-Hungary*
# 352 IS-Iceland
# 372 IE-Ireland
# 428 LV-Latvia
# 440 LT-Lithuania
# 528 NL-Netherlands
# 578 NO-Norway
# 616 PL-Poland
# 620 PT-Portugal
# 643 RU-Russia
# 703 SK-Slovakia
# 705 SI-Slovenia
# 752 SE-Sweden
# 756 CH-Switzerland
# 826 GB-Great Britain and/or United Kingdom - jätetään pois jotta saadaan TOPBOT
# -muuttuja mukaan (top-bottom self-placement) .(9.10.18)
# 840 US-United States - jätetään pois, jotta saadaan TOPBOT-muuttuja mukaan.(10.10.18)
#
# Belgian ja Saksan alueet:
# V3
# 5601 BE-FLA-Belgium/ Flanders

```

```

# 5602      BE-WAL-Belgium/ Wallonia
# 5603      BE-BRU-Belgium/ Brussels
# 27601     DE-W-Germany-West
# 27602     DE-E-Germany-East
#
# Unkari (348) toistaiseksi mukana, mutta joissain kysymyksissä myös Unkarilla on
# poikkeavia vastausvaihtoehtoja(HU_V18, HU_V19,HU_V20). Jos näitä muuttujia käytetään,
# Unkari on parempi jättää pois.
#
#
# (25.4.2018) user_na
# haven-paketin read_spss - funktiolla voi r-tiedostoon lukea myös SPSS:n sallimat kolme
# (yleensä 7, 8, 9) tarkempaa koodia puuttuvalle tiedolle.
# "If TRUE variables with user defined missing will be read into labelled_spss objects.
# If FALSE, the default, user-defined missings will be converted to NA"
# https://www.rdocumentation.org/packages/haven/versions/1.1.0/topics/read\_spss
#

ISSP2012jh.data <- read_spss("data/ZA5900_v4-0-0.sav") #luetaan alkuperäinen data R- dataks

#str(ISSP2012jh.data)

incl_countries25 <- c(36, 40, 56,100, 124, 191, 203, 208, 246, 250, 276, 348, 352,
                     372, 428, 440, 528, 578, 616, 620, 643, 703, 705, 752, 756)

#str(ISSP2012jh.data)
#str(ISSP2012jh.data) #61754 obs. of 420 variables - kaikki

ISSP2012jh1a.data <- filter(ISSP2012jh.data, V4 %in% incl_countries25)

#head(ISSP2012jh1a.data)
#str(ISSP2012jh1a.data) #34271 obs. of 420 variables, Espanja ja Iso-Britannia
#                               pois (9.10.2018)
# str(ISSP2012jh1a.data) # 32969 obs. of 420 variable, Espanja Iso-Britannia,
#                               USA pois (10.10.2018)
#
# names() # muuttujen nimet
# Maakohtaiset muuttujat (kun on poikettu ISSP2012 - vastausvaihtoehtoista tms.)
# on aineistossa eroteltu maatunnus-etuliitteellä (esimerkiksi ES_V7).
# Demografisissa ja muissa taustamuuttujissa suuri osa tiedoista on kerätty maa-
# kohtaisilla lomakkeilla. Vertailukelpoiset muuttujat on konstruoitu niistä.
# Muuttujia on 420, vain osa yhteisiä kaikille maille.

```

**DATAN RAJAAMISTA - MUUTTUJAT** (5.10.2018) Kolme ensimmäistä muuttujaa ovat datan metatietoja.

```

# 2. VALITAAN MUUTTUJAT -> ISSP2012jh1b.data. Maat valittu koodilohkossa dataset_country1
#
#
# Muuttujat on luokiteltu dokumentissa ZA5900_overview.pdf
# https://zacat.gesis.org/webview/index.jsp?object=http://zacat.gesis.org/obj/fStudy/ZA5900
# Study Description -> Other Study Description -> Related Materials
#
#

# METADATA

metavars1 <- c("V1", "V2", "DOI")

#MAA - maakoodit ja maan kahden merkin tunnus

countryvars1 <- c("V3", "V4", "C_ALPHAN")

# SUBSTANSSIMUUTTUJAT - Attitudes towards family and gender roles (9)
#
# Yhdeksän kysymystä (lyhennetyt versiot, englanniksi), vastausvaihtoehdot Q1-Q2
#
# 1 = täysin samaa mieltä, 2 = samaa mieltä, 3 = ei samaa eikä eri mieltä,
# 4 = eri mieltä, 5 = täysin eri mieltä
#
# Q1a Working mother can have warm relation with child
# Q1b Pre-school child suffers through working mother
# Q1c Family life suffers through working mother
# Q1d Women's preference: home and children
# Q1e Being housewife is satisfying
#
# Q2a Both should contribute to household income
# Q2b Men's job is earn money, women's job household
#
# Q3a Should women work: Child under school age
# Q3b Should women work: Youngest kid at school
# 1= kokopäivätyö, 2 = osa-aikatyö, 3 = pysyä kotona, 8 = en osaa sanoa (can't choose), 9 =
#
# Kysymysten Q3a ja Q3b eos-vastaus ei ole sama kuin "en samaa enkä eri mieltä" (ns. neutraali
# vaihtoehto), mutta kieltäytymisiä jne. (koodi 9) on aika vähän. Kolmessa
# maassa ne on yhdistetty:
# (8 Can't choose, CA:can't choose+no answer, KR:don't know+refused, NL:don't know).
# Kun SPSS-tiedostosta ei ole tuotu puuttuvan tiedon tarkempaa luokittelua,
# erottelua ei voi tehdä.
#
#

```

```

#

substvars1 <- c("V5","V6","V7","V8","V9","V10","V11","V12","V13") # 9 muuttujaa

# Nämä yhteiset muuttujat pois (maaspesifien muuttujien lisäksi) :
#
# "V14","V15","V16", "V17","V18","HU_V18","V19","HU_V19","V20","HU_V20","V21","V28","V29",
# "V34", "V35", "V36", "V37", "V38", "V39", "V40", "V41", "V42", "V43", "V44", "V45",
# "V46", "V47", "V48", "V49", "V50", "V51", "V52", "V53", "V54", "V55", "V56", "V57", "V58",
# "V60", "V61", "V62", "V63", "V64", "V65", "V65a","V66", "V67"
#
#
# DEMOGRAFISET JA MUUT TAUSTAMUUTTUAJAT (8)
#
# AGE, SEX
#
# DEGREE - Highest completed degree of education: Categories for international comparison.
# Slightly re-arranged subset of ISCED-97
#
# 0 No formal education
# 1 Primary school (elementary school)
# 2 Lower secondary (secondary completed does not allow entry to university: obligatory school)
# 3 Upper secondary (programs that allow entry to university or programs that allow to enter university)
# other ISCED level 3 programs - designed to prepare students for direct entry into the labour market
# 4 Post secondary, non-tertiary (other upper secondary programs toward labour market or toward tertiary education)
# 5 Lower level tertiary, first stage (also technical schools at a tertiary level)
# 6 Upper level tertiary (Master, Dr.)
# 9 No answer, CH: don't know
# Yhdistelyt?
#
# MAINSTAT - main status: Which of the following best describes your current situation?
#
# 1 In paid work
# 2 Unemployed and looking for a job, HR: incl never had a job
# 3 In education
# 4 Apprentice or trainee
# 5 Permanently sick or disabled
# 6 Retired
# 7 Domestic work
# 8 In compulsory military service or community service
# 9 Other
# 99 No answer
# Armeijassa tai yhdyskuntapalvelussa muutamia, muutamissa maissa. Kategoriassa 9
# on hieman väkeä. Yhdistetään 8 ja 9. Huom! Esim Puolassa ei yhtään eläkeläistä
# eikä kategoriassa 9, Saksassa ei ketään kategoriassa 9.

```

```

#
# TOPBOT - Top-Bottom self-placement (10 pt scale)
#
# "In our society, there are groups which tend to be towards the top and groups
# which tend to be towards the bottom. Below is a scale that runs
# from the top to the bottom. Where would you put yourself on this scale?"
# Eri maissa hieman erilaisia kysymyksiä.
#
# HHCHILDR - How many children in household: children between [school age] and
# 17 years of age
#
# 0 No children
# 1 One child
# 2 2 children
# 21 21 children
# 96 NAP (Code 0 in HOMPOP)
# 97 Refused
# 99 No answer
#
# Voisi koodata dummymuuttujaksi lapsia (1) - ei lapsia (0).
# Ranskan datassa on erittäin iso osa puuttuvia tietoja ( "99", n. 20 %), myös Austerliall
# aika paljon. Sama tilanne myös muissa perheen kokoon liittyvissä kysymyksissä.
#
#
# MARITAL - Legal partnership status
#
# What is your current legal marital status?
# The aim of this variable is to measure the current 'legal' marital status '.
# PARTLIV - muuttujassa on 'de facto' - tilanteen tieto parisuhteesta
#
# 1 Married
# 2 Civil partnership
# 3 Separated from spouse/ civil partner (still legally married/ still legally
#   in a civil partnership)
# 4 Divorced from spouse/ legally separated from civil partner
# 5 Widowed/ civil partner died
# 6 Never married/ never in a civil partnership, single
# 7 Refused
# 8 Don't know
# 9 No answer
#
# URBRURAL - Place of living: urban - rural
#
# 1 A big city
# 2 The suburbs or outskirts of a big city

```

```

# 3 A town or a small city
# 4 A country village
# 5 A farm or home in the country
# 7 Other answer
# 9 No answer
# 1 ja 2 vaihtelevat aika paljon maittain, parempi laskea yhteen. Unkarista puuttuu
# jostain syystä kokonaan vaihtoehto 5. Vaihtoehtoon 7 on valinnut vain 4 vastaajaa Ranskassa
# Yhdistetään 1 ja 2 = city, 3 = town, rural= 4, 5, 7
#

bgvars1 <- c( "SEX","AGE","DEGREE", "MAINSTAT", "TOPBOT", "HHCHILDR", "MARITAL", "URBRURAL")

#Valitaan muuttujat

jhvars1 <- c(metavars1,countryvars1, substvars1,bgvars1)

#jhvars1
ISSP2012jh1b.data <- select(ISSP2012jh1a.data, jhvars1)

# laaja aineisto - mukana havainnot joissa puuttuvia tietoja
# hauska detalji URBRURAL - muuttujan metatiedoissa viite jonkun työaseman hakemistoon
# str(ISSP2012jh1b.data) #32969 obs. of 23 variables
#
# SUBSTANSSIMUUTTUJAT
#
# $ V5      : 'haven_labelled' num  5 1 2 2 1 NA 2 4 2 2 ...
# ..- attr(*, "label")= chr "Q1a Working mom: warm relationship with children as a not wor
# ..- attr(*, "labels")= Named num  0 1 2 3 4 5 8 9
#
# ISSP2012jh1b.data$V5 näyttää tarkemmin rakenteen
#
# glimpse(ISSP2012jh1b.data)
# str(ISSP2012jh1b.data) # 32969 obs. of 23 variables

# Poistetaan havainnot, joissa ikä (AGE) tai sukupuoli tieto puuttuu (5.7.2019)
ISSP2012jh1c.data <- filter(ISSP2012jh1b.data, (!is.na(SEX) & !is.na(AGE)))
str(ISSP2012jh1c.data) # 32823 obs. of 23 variables, 32969-32823 = 146

## Classes 'tbl_df', 'tbl' and 'data.frame': 32823 obs. of 23 variables:
## $ V1      : 'haven_labelled' num  5900 5900 5900 5900 5900 5900 5900 5900 5900 5900 ...
## ..- attr(*, "label")= chr "GESIS Data Archive Study Number"
## ..- attr(*, "labels")= Named num  5900
## .. ..- attr(*, "names")= chr "GESIS Data Archive Study Number ZA5900"
## $ V2      : chr  "4.0.0 (2016-11-23)" "4.0.0 (2016-11-23)" "4.0.0 (2016-11-23)" "4.0.0 (
## ..- attr(*, "label")= chr "GESIS Archive Version"

```

```

##   .- attr(*, "format.spss")= chr "A25"
##   .- attr(*, "display_width")= int 26
## $ DOI      : chr "doi:10.4232/1.12661" "doi:10.4232/1.12661" "doi:10.4232/1.12661" "doi:10.4232/1.12661"
##   .- attr(*, "label")= chr "Digital Object Identifier"
##   .- attr(*, "format.spss")= chr "A50"
##   .- attr(*, "display_width")= int 26
## $ V3       : 'haven_labelled' num  36 36 36 36 36 36 36 36 36 36 ...
##   .- attr(*, "label")= chr "Country/ Sample ISO 3166 Code (see V4 for codes for whole na
##   .- attr(*, "labels")= Named num  32 36 40 100 124 152 156 158 191 203 ...
##   .. .- attr(*, "names")= chr  "AR-Argentina" "AU-Australia" "AT-Austria" "BG-Bulgaria"
## $ V4       : 'haven_labelled' num  36 36 36 36 36 36 36 36 36 36 ...
##   .- attr(*, "label")= chr "Country ISO 3166 Code (see V3 for codes for the sample)"
##   .- attr(*, "labels")= Named num  32 36 40 56 100 124 152 156 158 191 ...
##   .. .- attr(*, "names")= chr  "AR-Argentina" "AU-Australia" "AT-Austria" "BE-Belgium"
## $ C_ALPHAN: chr  "AU" "AU" "AU" "AU" ...
##   .- attr(*, "label")= chr "Country Prefix ISO 3166 Code - alphanumeric"
##   .- attr(*, "format.spss")= chr "A20"
##   .- attr(*, "display_width")= int 22
## $ V5       : 'haven_labelled' num   5 1 2 2 1 NA 2 4 2 2 ...
##   .- attr(*, "label")= chr "Q1a Working mom: warm relationship with children as a not wo
##   .- attr(*, "labels")= Named num   0 1 2 3 4 5 8 9
##   .. .- attr(*, "names")= chr  "NAP: ES" "Strongly agree" "Agree" "Neither agree nor dis
## $ V6       : 'haven_labelled' num   1 5 4 4 4 NA 4 3 4 3 ...
##   .- attr(*, "label")= chr "Q1b Working mom: Preschool child is likely to suffer"
##   .- attr(*, "labels")= Named num   0 1 2 3 4 5 8 9
##   .. .- attr(*, "names")= chr  "NAP: ES" "Strongly agree" "Agree" "Neither agree nor dis
## $ V7       : 'haven_labelled' num   3 5 2 4 4 NA 4 2 4 2 ...
##   .- attr(*, "label")= chr "Q1c Working woman: Family life suffers when woman has full-t
##   .- attr(*, "labels")= Named num   0 1 2 3 4 5 8 9
##   .. .- attr(*, "names")= chr  "NAP: ES" "Strongly agree" "Agree" "Neither agree nor dis
## $ V8       : 'haven_labelled' num   3 5 5 2 4 NA 4 5 4 5 ...
##   .- attr(*, "label")= chr "Q1d Working woman: What women really want is home and kids"
##   .- attr(*, "labels")= Named num   0 1 2 3 4 5 8 9
##   .. .- attr(*, "names")= chr  "NAP: ES" "Strongly agree" "Agree" "Neither agree nor dis
## $ V9       : 'haven_labelled' num   3 1 2 3 4 NA 2 4 4 1 ...
##   .- attr(*, "label")= chr "Q1e Working woman: Being housewife is as fulfilling as worki
##   .- attr(*, "labels")= Named num   0 1 2 3 4 5 8 9
##   .. .- attr(*, "names")= chr  "NAP: ES" "Strongly agree" "Agree" "Neither agree nor dis
## $ V10      : 'haven_labelled' num   1 3 4 2 2 NA 2 5 2 1 ...
##   .- attr(*, "label")= chr "Q2a Both should contribute to household income"
##   .- attr(*, "labels")= Named num   0 1 2 3 4 5 8 9
##   .. .- attr(*, "names")= chr  "NAP: ES" "Strongly agree" "Agree" "Neither agree nor dis
## $ V11      : 'haven_labelled' num   3 5 4 4 4 NA 2 5 4 1 ...
##   .- attr(*, "label")= chr "Q2b Men's job earn money, women's job look after home"
##   .- attr(*, "labels")= Named num   0 1 2 3 4 5 8 9
##   .. .- attr(*, "names")= chr  "NAP: ES" "Strongly agree" "Agree" "Neither agree nor dis

```



```

## $ V12      : 'haven_labelled' num  3 NA NA 2 2 NA 2 NA 2 2 ...
##   ..- attr(*, "label")= chr "Q3a Should women work: Child under school age"
##   ..- attr(*, "labels")= Named num  1 2 3 6 8 9
##   .. ..- attr(*, "names")= chr  "Work full-time" "Work part-time" "Stay at home" "TW: work part-time"
## $ V13      : 'haven_labelled' num  2 NA 2 1 2 NA 2 NA 2 2 ...
##   ..- attr(*, "label")= chr "Q3b Should women work: Youngest kid at school"
##   ..- attr(*, "labels")= Named num  1 2 3 6 8 9
##   .. ..- attr(*, "names")= chr  "Work full-time" "Work part-time" "Stay at home" "TW: work part-time"
## $ SEX      : 'haven_labelled' num  1 2 2 2 2 1 2 1 2 2 ...
##   ..- attr(*, "label")= chr "Sex of Respondent"
##   ..- attr(*, "labels")= Named num  1 2 9
##   .. ..- attr(*, "names")= chr  "Male" "Female" "No answer"
## $ AGE      : 'haven_labelled' num  58 59 40 20 72 68 64 57 45 71 ...
##   ..- attr(*, "label")= chr "Age of respondent"
##   ..- attr(*, "labels")= Named num  15 16 17 18 102 999
##   .. ..- attr(*, "names")= chr  "15 years" "16 years" "17 years" "18 years" ...
## $ DEGREE   : 'haven_labelled' num  2 5 5 3 2 NA NA 6 5 6 ...
##   ..- attr(*, "label")= chr "Highest completed degree of education: Categories for international comparison"
##   ..- attr(*, "labels")= Named num  0 1 2 3 4 5 6 9
##   .. ..- attr(*, "names")= chr  "No formal education" "Primary school (elementary school)" "Secondary school" "Tertiary school"
## $ MAINSTAT : 'haven_labelled' num  6 6 3 1 6 5 6 2 1 5 ...
##   ..- attr(*, "label")= chr "Main status"
##   ..- attr(*, "labels")= Named num  1 2 3 4 5 6 7 8 9 99
##   .. ..- attr(*, "names")= chr  "In paid work" "Unemployed and looking for a job, HR: in the last 12 months" "Retired" "Other"
## $ TOPBOT   : 'haven_labelled' num  3 7 8 NA 7 2 7 NA 10 6 ...
##   ..- attr(*, "label")= chr "Top-Bottom self-placement"
##   ..- attr(*, "labels")= Named num  0 1 2 3 4 5 6 7 8 9 ...
##   .. ..- attr(*, "names")= chr  "Not available: GB,US" "Lowest, Bottom, 01" "02" "03" ...
## $ HHCHILDR : 'haven_labelled' num  NA NA 3 1 0 NA 0 0 1 NA ...
##   ..- attr(*, "label")= chr "How many children in household: children between [school age and 18 years]"
##   ..- attr(*, "labels")= Named num  0 1 2 21 96 97 99
##   .. ..- attr(*, "names")= chr  "No children" "One child" "2 children" "21 children" ...
## $ MARITAL  : 'haven_labelled' num  6 1 1 6 1 6 1 1 1 NA ...
##   ..- attr(*, "label")= chr "Legal partnership status"
##   ..- attr(*, "labels")= Named num  1 2 3 4 5 6 7 8 9
##   .. ..- attr(*, "names")= chr  "Married" "Civil partnership" "Separated from spouse/ civil partner" "Widowed"
## $ URBURURAL : 'haven_labelled' num  1 1 1 NA 1 2 NA 2 2 NA ...
##   ..- attr(*, "label")= chr "Place of living: urban - rural"
##   ..- attr(*, "labels")= Named num  1 2 3 4 5 7 9
##   .. ..- attr(*, "names")= chr  "A big city" "The suburbs or outskirts of a big city" "A small town or village"
## - attr(*, "notes")= chr  "document Plan File: /Users/marcic/Desktop/old/GPS2011 sampling"

```

ISSP2012jh1c.data %>% summary()

V1	V2	DOI	V3	V4	C_ALPHAN
Min. :5900	Length:32823	Length:32823	Min. : 36	Min. : 36.0	Length:32823

V1	V2	DOI	V3	V4	C_ALPHAN	
1st Qu.:5900	Class :character	Class :character	1st Qu.: 208	1st Qu.:203.0	Class :character	1
Median :5900	Mode :character	Mode :character	Median : 428	Median :276.0	Mode :character	M
Mean :5900	NA	NA	Mean : 4063	Mean :362.1	NA	1
3rd Qu.:5900	NA	NA	3rd Qu.: 705	3rd Qu.:578.0	NA	3
Max. :5900	NA	NA	Max. :62002	Max. :756.0	NA	1
NA	NA	NA	NA	NA	NA	

Metatietojen (3) ja maa-muuttujien (3) lisäksi aineistossa on seitsemäntoista muuttujaa. Yhdeksän muuttujaa ovat ns. substanssikysymysten vastauksia, joilla luodataan asenteita sukupuolirooleihin ja perhearvoihin. Kahdeksan taustamuuttujaa.

**Yhdeksän kysymystä (lyhennetyt versiot, englanniksi), vastausvaihtoehdot edit** tämä pätkä tuottaa virheilmoituksia (31-1-20), esim Error: unexpected symbol in “Q1a Working”

Vastausvaihtoehdot:

1 = täysin samaa mieltä, 2 = samaa mieltä, 3 = ei samaa eikä eri mieltä, 4 = eri mieltä, 5 = täysin eri mieltä

Q1a Working mother can have warm relation with child Q1b Pre-school child suffers through working mother Q1c Family life suffers through working mother Q1d Women’s preference: home and children Q1e Being housewife is satisfying Q2a Both should contribute to household income Q2b Men’s job is earn money, women’s job household

Q3a Should women work: Child under school age Q3b Should women work: Youngest kid at school

Vastausvaihtoehdot: “Work full-time” “Work part-time” “Stay at home”, “Can’t choose” 1 = W, 2 = w, 3 = H, NA = 6,8,9 ei tässä eriteltyinä. 6 on Taiwanin oma vastausvaihtoehdot, 8 = en osaa sanoa ja 9 = no answer.

*# Muuttuja taulukkona - karkea tapa*

```
tabVarNames <- c(substvars1,bgvars1) # muuttujanimet muuttujille

# Kysymysten lyhyet versiot englanniksi
tabVarDesc <- c("Q1a Working mother can have warm relation with child ",
  "Q1b Pre-school child suffers through working mother",
  "Q1c Family life suffers through working mother",
  "Q1d Women’s preference: home and children",
  "Q1e Being housewife is satisfying",
  "Q2a Both should contribute to household income",
  "Q2b Men’s job is earn money, women’s job household",
  "Q3a Should women work: Child under school age",
```

```

        "Q3b Should women work: Youngest kid at school",
        "Respondents age ",
        "Respondents gender",
        "Highest completed degree of education: Categories for international comparison",
        "Main status: work, unemployed, in education...",
        "Top-Bottom self-placement (10 pt scale)",
        "How many children in household: children between [school age] and 17 years",
        "Legal partnership status: married, civil partnership...",
        "Place of living: urban - rural"
    )
#tabVarDesc

# Taulukko
# luodaan df
jhVarTable1.df <- data_frame(tabVarnames,tabVarDesc)
cols_jhVarTable1 <- c("muuttuja","Kysymyksen tunnus, lyhennetty kysymys")
colnames(jhVarTable1.df) <- cols_jhVarTable1
#jhVarTable1.df

# Suomalaiset pitkät kysymykset
vastf1 <- c("Q1a Työssäkäyvä äiti pystyy luomaan lapsiinsa aivan yhtä lämpimän
           ja turvallisen suhteen kuin äiti, joka ei käy työssä")

vastf2 <- c("Q1b Alle kouluikäinen lapsi todennäköisesti kärsii, jos hänen äitinsä käy työssä")
vastf3 <- c("Q1c Kaiken kaikkiaan perhe-elämä kärsii, kun naisella on kokopäivätyö.")
vastf4 <- c("Q1d On hyvä käydä töissä mutta tosiasiaissa useimmat naiset haluavat
           ensisijaisesti kodin ja lapsia.")
vastf5 <- c("Q1e Kotirouvana oleminen on aivan yhtä antoisaa kuin ansiotyön tekeminen.")
vastf6 <- c("Q2a Sekä miehen että naisen tulee osallistua perheen toimeentulon hankkimiseen.")
vastf7 <- c("Q2b Miehen tehtävä on ansaita rahaa; naisen tehtävä on huolehtia kodista ja perheestä.")
vastf8 <- c("Q3a Millä tavoin naisten pitäisi mielestäsi käydä työssä seuraavissa tilanteissa:
           Kun perheessä on alle kouluikäinen lapsi")
vastf9 <- c("Q3b Millä tavoin naisten pitäisi mielestäsi käydä työssä seuraavissa tilanteissa:
           Kun nuorin lapsi on aloittanut koulunkäynnin")

tabVarDesc_fi <- c(vastf1,vastf2,vastf3,vastf4,vastf5,vastf6,vastf7, vastf8,vastf9)
#tabVarDesc_fi
tabVarnames_subst <- c(substvars1)
jhVarTable1_fi.df <- data_frame(tabVarnames_subst,tabVarDesc_fi)
cols_jhVarTable1 <- c("muuttuja","Kysymyksen tunnus, suomenkielisen lomakkeen kysymys")
colnames(jhVarTable1_fi.df) <- cols_jhVarTable1

# TAULUKODEN TULOSTUS

# kable(booktab = T) # booktab = T gives us a pretty APA-ish table

```

```
# Lyhyet kysymykset englanniksi
```

```
knitr::kable(jhVarTable1.df, booktab=TRUE,
  fig.cap="ISSP2012:Työelämä ja perhearvot - valitut muuttajat")
```

muuttuja	Kysymyksen tunnus, lyhennetty kysymys
V5	Q1a Working mother can have warm relation with child
V6	Q1b Pre-school child suffers through working mother
V7	Q1c Family life suffers through working mother
V8	Q1d Women's preference: home and children
V9	Q1e Being housewife is satisfying
V10	Q2a Both should contribute to household income
V11	Q2b Men's job is earn money, women's job household
V12	Q3a Should women work: Child under school age
V13	Q3b Should women work: Youngest kid at school
SEX	Respondents age
AGE	Respondents gender
DEGREE	Highest completed degree of education: Categories for international comparison
MAINSTAT	Main status: work, unemployed, in education...
TOPBOT	Top-Bottom self-placement (10 pt scale)
HHCHILDR	How many children in household: children between [school age] and 17 years of age
MARITAL	Legal partnership status: married, civil partnership...
URBRURAL	Place of living: urban - rural

```
# Suomen lomakkeen kysymykset (löytyy myös kuva lomakkeen sivusta)
```

```
knitr::kable(jhVarTable1_fi.df, booktab=TRUE,
  fig.cap="ISSP2012: suomenkielisen lomakkeen kysymykset")
```

muuttuja	Kysymyksen tunnus, suomenkielisen lomakkeen kysymys
V5	Q1a Työssäkäyvä äiti pystyy luomaan lapsiinsa aivan yhtä lämpimän ja turvallisen suhteen kuin äiti, joka ei käy työssä
V6	Q1b Alle kouluikäinen lapsi todennäköisesti kärsii, jos hänen äitinsä käy työssä.
V7	Q1c Kaiken kaikkiaan perhe-elämä kärsii, kun naisella on kokopäivätyö.
V8	Q1d On hyvä käydä töissä mutta tosiasiassa useimmat naiset haluavat ensisijaisesti kodin ja lapsia.
V9	Q1e Kotirouvana oleminen on aivan yhtä antoisaa kuin ansiotyön tekeminen.
V10	Q2a Sekä miehen että naisen tulee osallistua perheen toimeentulon hankkimiseen.
V11	Q2b Miehen tehtävä on ansaita rahaa; naisen tehtävä on huolehtia kodista ja perheestä.
V12	Q3a Millä tavoin naisten pitäisi mielestäsi käydä työssä seuraavissa tilanteissa? Kun perheessä on alle kouluikäinen lapsi
V13	Q3b Millä tavoin naisten pitäisi mielestäsi käydä työssä seuraavissa tilanteissa? Kun nuorin lapsi on aloittanut koulunkäynnin

# Taulukot voivat olla hankalia eristyisesti PDF-tulostuksessa, jos ne ovat  
# monimutkaisia tai solujen "koot" (merkkiä/solu) vaihtelevat paljon.

# Kokeillaan taulukoiden yhdistämistä, jos aikaa jää. Ei luultavasti kannata, kun halutaan j  
# html-tulostus samalla koodilla (26.12.18).

Tarkemmat kysymysten muotoilut poikkeavat tietysti hieman eri maiden välillä. Suomen lomakkeet täydelliset kysymykset voi tarkista tiedostosta ZA5900\_q\_fi-fi.pdf, löytyy zcat-sivustolta. Tarkemmat kuvaukset lähes tuhatsivuisessa koodikirjassa ZA5900\_cdb.pdf (**refworks-viite pitäisi löytyä**, ja ISSP dokumentit kerrotaan luvun alussa).

Bookdown-versiossa taulukot omiksi koodilohkoiksi, ja fig.caption - optiolla taulukon otsikko.

Kysymyslomakkeen kuva, vai kuva liitteisiin? **Liitteisiin.**

```
knitr::include_graphics('img/substvar_fi_Q1Q2.png')
```

Seuraavaksi perheeseen, työhön ja kotiin liittyviä kysymyksiä.							
<b>23. Mitä mieltä olet seuraavista väittämistä?</b>							
Rengasta jokaisell... lulla vain yksi vaihtoehto							
	Täysin samaa mieltä	Samaa mieltä	En samaa mieltä	Eri mieltä	Täysin eri mieltä	En osaa sanoa	
a)	Työssäkäyvä äiti pystyy luomaan lapsinsa aivan yhtä lämpimän ja turvallisen suhteen kun äiti, joka ei käy työssä.....						
b)	Alle kouluikäinen lapsi todennäköisesti kärsii, jos hänen äitinsä käy työssä.....						
c)	Kaiken kaikkiaan perhe-elämä kärsii, kun naisella on kokopäivätyö.....						
d)	On hyvä käydä toisessa mutta tosiasiasissa useimmat naiset haluavat ensisijaisesti kodin ja lapsia.....						
e)	Kotona olo on aivan yhtä antoisaa kuin ansaitseminen.....						
<hr/>							
<b>24. Mitä mieltä olet seuraavista väittämistä?</b>							
Rengasta kummaltakin riviltä vain yksi vaihtoehto.							
	Täysin samaa mieltä	Samaa mieltä	En samaa mieltä	Eri mieltä	Täysin eri mieltä	En osaa sanoa	
a)	Sekä miehen että naisen tulee osallistua perheen toimeentulon hankkimiseen.....						
b)	Miehen tehtävä on ansaita rahaa, naisen tehtävä on huolehtia kodista ja perheestä.....						
<hr/>							
<b>25. Millä tavoin naisten pitäisi mielestäsi käydä työssä seuraavissa tilanteissa?</b>							
Rengasta kummaltakin riviltä vain yksi vaihtoehto.							
Naisen tulisi...		käydä koko-päivätyössä	käydä osa-aikatyössä	pyydyä kotona	En osaa sanoa		
a)	Kun perheessä on alle kouluikäinen lapsi.....	1	2	3	8		
b)	Kun nuorin lapsi on aloittanut koulunkäynnin.....	1	2	3	8		

Kuva 1: Suomen lomake

## 1.5 Datan valinnan vaiheet ja puuttuvat tiedot

**ks** Perusasiat havaintojen puuttellisuudesta kyselytutkimuksissa. Yksikkövas-tauskato (unit non-response), eräsvastauskato (item non-response). Mitä on raportoitava, kun käytetään valmista aineistoa? Eräsvastauskatoa analysoidaan, kun käytetään kaikkia valittuja muuttujia.

Yksikkövastauskato on otettu vaihtelevasti huomioon, kun kyselyn toteuttaja on editoinut ja tarkastanut datan. Eri maiden datassa on (mutta ei aina!) mukana

painot mm. vastauskadon oikaiksemiseen **Viittet - tekninen raportti**. Myös selaimella voi zcat-sivustolla tutkailla kysymyksittäin.

Datakatalogi-dokumentista näkee vastausten jakauman jokaisen kysymyksen osalta, myös puuttuvien tietojen tarkemman koodauksen.

1. Valitaan 25 maata ja muuttujat
2. Johdattalevissa esimerkeissä valitaan kuusi maata ja kolme muuttujaa. Jätetään pois kaikki havainnot (vastaukset) joissa on puuttuvia tietoja ("listwise deletion")
3. Kun laajempi aineisto otetaan käyttöön, joudutaan pohtimaan miten puuttuvia havaintoja käsitellään. Jos kyse on selvistä virheistä (esim. haastateltavan ikä puuttu) havainnot jätetään pois, muuten mietitään.

**Miten puuttuvia tietoja (erävastuskato, havainnossa puuttu joku tieto) käsitellään? edit** Tämä on vähän hämää, ehkä pois? (30.1.20)

1. Miksi tieto puuttuu, mitä "puuttuva tieto" tarkoittaa?

Joissain kysymyksissä (V12, V13) puuttuvaksi tiedoksi kirjautuu vastaus ("en osaa sanoa") "ei vastausta" - vaihtoehdon lisäksi. Nämä mukaan.

Ikä ja sukupuoli: ilmeinen virhe, joten jätetään havainnot pois (näitä ei ole paljon)

2. Puuttuvien tietojen jakauma?

Onko puuttuvia tietoja tasaisesti eri maissa, vai vaihtelee niiden suhteellinen osuus?

Onko joissain tai jossain maassa huomattava määrä puuttuvia tietoja?

Onko puuttuvia tietoja paljon vai vähän?

Puuttuvat tiedot otetaan mukaan analyysiin, mutta tulkinnassa on pidettävä mielessä niiden jakauma. **edit** Mitäköhän tarkoittaa?

Tarkemmin puuttuneisuutta ei analysoida. Esimerkkejä löytyy (MG, CAiP ja "vihreä kirja"). **edit** Viite!

Koko aineistossa (valitut 25 maata) kysymyksen Q1b (muuttuja V6) vastauksista puuttuvia tietoja on 3,5 prosenttia (1219/34271). **Huom:** kun pudotetaan havainnot joilta SEX tai AGE puuttuu, N = 32823! On oikea määrä (5.7.2019, kts. treeni2- projekti, Data\_iso1.R

Lisätietoa puuttuvien havaintojen päähkäilyyn, apumateriaalia (6.12.2018). Ei tarvita, kun luodaan muuttuja jossa puuttuva tieto on mukana uutena luokittelumuuttujan arvona (30.1.20)

Puuttuvien tietojen tarkempi koodaus ISSP-datassa:

0: Not applicable (NAP), Not available (NAV) 7: (97,997, 9997,...): Refused 8: (98, 998, 9998,...): Don't know 9: (99, 999, 9999,...): No answer

NAP ja NAV määritellään

"GESIS adds 'Not applicable'(NAP) codes for questions that have filters. NAP indicates that only a subsample and not all of respondents were asked. Also in the case of country specific variables, all the other countries are coded NAP.

GESIS adds 'Not available' for variables, which in single countries may not have been conducted for whatever reason."

---

## 1.6 Perusmuunnoksesta ISSP2012 - datalle

Datatiedosto on ISSP2012jh1.data, ja luokittelumuuttujat muunnetaan R:n factor- muuttujaksi.

Jokaisesta muuttujasta on kaksi versiota, toisessa puuttuvat tiedot ovat R:n "NA"- arvoja ja toisessa "NA"-arvo on eksplisiittinen muuttuja ("missing").

Substanssimuuttujien luokkien tunnuksien (faktorilabelit, levels?) muutetaan graafisiin analyysiin sopivan lyhyiksi. Taustamuuttujien luokittelua ja luokkien tunnuksia pohditaan, kun ne otetaan käyttöön.

**TODO 30.1.20** Tarkistukset, varmistukset jne.

**TODO2.2.20** Muunnetaanko muuttujan maa (C\_ALPHAN as\_factor) järjestys heti samaksi kuin C\_ALPHAN? Nyt tehdään G1\_2\_johdesim.Rmd:ssä.

```
# VAIHE 1 - muuttujat joissa ei ole puuttuvia tietoja

# vaihe 1.1 haven_labelled ja chr -> as_factor

ISSP2012jh1d.dat <- ISSP2012jh1c.data %>%
  mutate(maa = as_factor(C_ALPHAN), # ei puuttuvia, ei tyhjiä tasoja
         sp1 = as_factor(SEX), # ei puuttuvia, tyhjä taso "no answer" 999
        )

# str(ISSP2012jh1d.dat$maa) #Country Prefix ISO 3166 Code - alphanumeric
# attributes(ISSP2012jh1d.dat$maa) # ei tyhjiä tasoja, 25 tasoja

#C_ALPHAN - maa - tarkistuksia

ISSP2012jh1d.dat$maa %>% fct_unique()

## [1] AU AT BG CA HR CZ DK FI FR HU IS IE LV LT NL NO PL RU SK SI SE CH BE DE PT
## 25 Levels: AU AT BG CA HR CZ DK FI FR HU IS IE LV LT NL NO PL RU SK SI ... PT

ISSP2012jh1d.dat$maa %>% fct_count() %>% kable()
```

f	n
AU	1557
AT	1182
BG	1003
CA	953
HR	997
CZ	1804
DK	1403
FI	1171
FR	2409
HU	1012
IS	1172
IE	1166
LV	1000
LT	1187
NL	1315
NO	1444
PL	1115
RU	1525
SK	1128
SI	1034
SE	1059
CH	1237
BE	2192
DE	1761
PT	997

```
ISSP2012jh1d.dat$maa %>% summary()
```

```
##   AU   AT   BG   CA   HR   CZ   DK   FI   FR   HU   IS   IE   LV   LT   NL   NO
## 1557 1182 1003  953  997 1804 1403 1171 2409 1012 1172 1166 1000 1187 1315 1444
##   PL   RU   SK   SI   SE   CH   BE   DE   PT
## 1115 1525 1128 1034 1059 1237 2192 1761  997
```

```
sum(is.na(ISSP2012jh1d.dat$maa)) # ei puuttuvia tietoja
```

```
## [1] 0
```

```
# C_ALPHAN alkuperäinen järjestys, maa aakkosjärjestyssä (21.2.20)
```

```
#
```

```
# Huom1: Myös merkkijonomuuttujaa C_ALPHAN tarvitaan jatkossa
```

```
#
```

```
# Huom2: kun dataa rajataan, on tarkistettava ja tarvittaessa poistettava
```

```
# "tyhjät" R-factor - muuttujan "maa" luokat
```

```
ISSP2012jh1d.dat %>% tableX(C_ALPHAN, maa)
```



C_ALPHAN/maa	AU	AT	BG	CA	HR	CZ	DK	FI	FR	HU	IS	IE	LV
AT	0	1182	0	0	0	0	0	0	0	0	0	0	0
AU	1557	0	0	0	0	0	0	0	0	0	0	0	0
BE	0	0	0	0	0	0	0	0	0	0	0	0	0
BG	0	0	1003	0	0	0	0	0	0	0	0	0	0
CA	0	0	0	953	0	0	0	0	0	0	0	0	0
CH	0	0	0	0	0	0	0	0	0	0	0	0	0
CZ	0	0	0	0	0	1804	0	0	0	0	0	0	0
DE	0	0	0	0	0	0	0	0	0	0	0	0	0
DK	0	0	0	0	0	0	1403	0	0	0	0	0	0
FI	0	0	0	0	0	0	0	1171	0	0	0	0	0
FR	0	0	0	0	0	0	0	0	2409	0	0	0	0
HR	0	0	0	0	997	0	0	0	0	0	0	0	0
HU	0	0	0	0	0	0	0	0	0	1012	0	0	0
IE	0	0	0	0	0	0	0	0	0	0	0	1166	0
IS	0	0	0	0	0	0	0	0	0	0	1172	0	0
LT	0	0	0	0	0	0	0	0	0	0	0	0	0
LV	0	0	0	0	0	0	0	0	0	0	0	0	100
NL	0	0	0	0	0	0	0	0	0	0	0	0	0
NO	0	0	0	0	0	0	0	0	0	0	0	0	0
PL	0	0	0	0	0	0	0	0	0	0	0	0	0
PT	0	0	0	0	0	0	0	0	0	0	0	0	0
RU	0	0	0	0	0	0	0	0	0	0	0	0	0
SE	0	0	0	0	0	0	0	0	0	0	0	0	0
SI	0	0	0	0	0	0	0	0	0	0	0	0	0
SK	0	0	0	0	0	0	0	0	0	0	0	0	0
Total	1557	1182	1003	953	997	1804	1403	1171	2409	1012	1172	1166	100

```
# vaihe 1.2 tyhjät luokat (levels) pois faktoreista

ISSP2012jh1d.dat <- ISSP2012jh1d.dat %>%
  mutate(sp = fct_drop(sp1))
# sp, sp1, SEX - tarkistuksia
ISSP2012jh1d.dat$sp %>% fct_count()
```

f	n
Male	14789
Female	18034

```
ISSP2012jh1d.dat$sp %>% fct_count()
```

	f	n
Male		14789
Female		18034

```
ISSP2012jh1d.dat %>% tableX(SEX,sp1)
```

SEX/sp1	Male	Female	No answer	Total
1	14789	0	0	14789
2	0	18034	0	18034
Total	14789	18034	0	32823

```
ISSP2012jh1d.dat %>% tableX(SEX,sp)
```

SEX/sp	Male	Female	Total
1	14789	0	14789
2	0	18034	18034
Total	14789	18034	32823

```
ISSP2012jh1d.dat %>% tableX(sp1,sp)
```

sp1/sp	Male	Female	Total
Male	14789	0	14789
Female	0	18034	18034
No answer	0	0	0
Total	14789	18034	32823

```
# vaihe 1.3 uudet "faktorilabelit"
ISSP2012jh1d.dat <- ISSP2012jh1d.dat %>%
  mutate(sp =
    fct_recode(sp,
      "m" = "Male",
      "f" = "Female")
  )

# Tarkistuksia
```

```
ISSP2012jh1d.dat$sp %>% fct_unique()
```

```
## [1] m f
## Levels: m f
```

```
ISSP2012jh1d.dat$sp %>% fct_count()
```

	n
m	14789
f	18034

```
ISSP2012jh1d.dat$sp %>% summary()
```

```
##      m      f
## 14789 18034
```

```
# AGE -> ika
# AGE----
```

```
ISSP2012jh1d.dat$ika <- ISSP2012jh1d.dat$AGE
```

```
# Tarkistuksia
```

```
attributes(ISSP2012jh1d.dat$ika) # tyhjä level "No answer"
```

```
## $label
## [1] "Age of respondent"
##
## $labels
## 15 years 16 years 17 years 18 years 102 years No answer
##      15      16      17      18      102      999
##
## $class
## [1] "haven_labelled"
```

```
ISSP2012jh1d.dat$ika %>% summary()
```

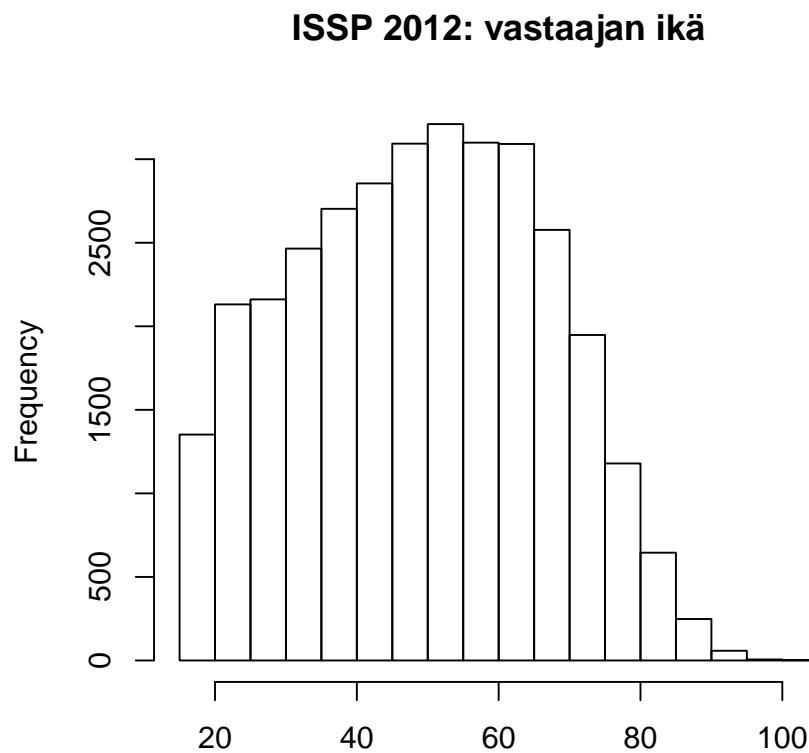
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
15	36	50	49.51607	63	102

```
ISSP2012jh1d.dat %>%
tableC(AGE, ika, cor_type = "pearson", na.rm = FALSE, rounding = 5,
output = "text", booktabs = TRUE, caption = NULL, align = NULL,
float = "htb") %>% kable()
```

```
## N = 32823
## Note: pearson correlation (p-value).
```

		[1]	[2]
[1]AGE	[1]AGE	1.00	
[2]ika	[2]ika	1.00	1.00

```
ISSP2012jh1d.dat$ika %>% hist(main = "ISSP 2012: vastaajan ikä")
```



```
str(ISSP2012jh1d.dat)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame': 32823 obs. of 27 variables:
## $ V1 : 'haven_labelled' num 5900 5900 5900 5900 5900 5900 5900 5900 5900 5900 ...
## ..- attr(*, "label")= chr "GESIS Data Archive Study Number"
## ..- attr(*, "labels")= Named num 5900
## .. ..- attr(*, "names")= chr "GESIS Data Archive Study Number ZA5900"
## $ V2 : chr "4.0.0 (2016-11-23)" "4.0.0 (2016-11-23)" "4.0.0 (2016-11-23)" "4.0.0 (2016-11-23)" ...
## ..- attr(*, "label")= chr "GESIS Archive Version"
```

```

##   .- attr(*, "format.spss")= chr "A25"
##   .- attr(*, "display_width")= int 26
## $ DOI      : chr "doi:10.4232/1.12661" "doi:10.4232/1.12661" "doi:10.4232/1.12661" "doi:10.4232/1.12661"
##   .- attr(*, "label")= chr "Digital Object Identifier"
##   .- attr(*, "format.spss")= chr "A50"
##   .- attr(*, "display_width")= int 26
## $ V3       : 'haven_labelled' num  36 36 36 36 36 36 36 36 36 36 ...
##   .- attr(*, "label")= chr "Country/ Sample ISO 3166 Code (see V4 for codes for whole na
##   .- attr(*, "labels")= Named num  32 36 40 100 124 152 156 158 191 203 ...
##   .. .- attr(*, "names")= chr  "AR-Argentina" "AU-Australia" "AT-Austria" "BG-Bulgaria"
## $ V4       : 'haven_labelled' num  36 36 36 36 36 36 36 36 36 36 ...
##   .- attr(*, "label")= chr "Country ISO 3166 Code (see V3 for codes for the sample)"
##   .- attr(*, "labels")= Named num  32 36 40 56 100 124 152 156 158 191 ...
##   .. .- attr(*, "names")= chr  "AR-Argentina" "AU-Australia" "AT-Austria" "BE-Belgium"
## $ C_ALPHAN: chr  "AU" "AU" "AU" "AU" ...
##   .- attr(*, "label")= chr "Country Prefix ISO 3166 Code - alphanumeric"
##   .- attr(*, "format.spss")= chr "A20"
##   .- attr(*, "display_width")= int 22
## $ V5       : 'haven_labelled' num   5 1 2 2 1 NA 2 4 2 2 ...
##   .- attr(*, "label")= chr "Q1a Working mom: warm relationship with children as a not wo
##   .- attr(*, "labels")= Named num   0 1 2 3 4 5 8 9
##   .. .- attr(*, "names")= chr  "NAP: ES" "Strongly agree" "Agree" "Neither agree nor dis
## $ V6       : 'haven_labelled' num   1 5 4 4 4 NA 4 3 4 3 ...
##   .- attr(*, "label")= chr "Q1b Working mom: Preschool child is likely to suffer"
##   .- attr(*, "labels")= Named num   0 1 2 3 4 5 8 9
##   .. .- attr(*, "names")= chr  "NAP: ES" "Strongly agree" "Agree" "Neither agree nor dis
## $ V7       : 'haven_labelled' num   3 5 2 4 4 NA 4 2 4 2 ...
##   .- attr(*, "label")= chr "Q1c Working woman: Family life suffers when woman has full-t
##   .- attr(*, "labels")= Named num   0 1 2 3 4 5 8 9
##   .. .- attr(*, "names")= chr  "NAP: ES" "Strongly agree" "Agree" "Neither agree nor dis
## $ V8       : 'haven_labelled' num   3 5 5 2 4 NA 4 5 4 5 ...
##   .- attr(*, "label")= chr "Q1d Working woman: What women really want is home and kids"
##   .- attr(*, "labels")= Named num   0 1 2 3 4 5 8 9
##   .. .- attr(*, "names")= chr  "NAP: ES" "Strongly agree" "Agree" "Neither agree nor dis
## $ V9       : 'haven_labelled' num   3 1 2 3 4 NA 2 4 4 1 ...
##   .- attr(*, "label")= chr "Q1e Working woman: Being housewife is as fulfilling as worki
##   .- attr(*, "labels")= Named num   0 1 2 3 4 5 8 9
##   .. .- attr(*, "names")= chr  "NAP: ES" "Strongly agree" "Agree" "Neither agree nor dis
## $ V10      : 'haven_labelled' num   1 3 4 2 2 NA 2 5 2 1 ...
##   .- attr(*, "label")= chr "Q2a Both should contribute to household income"
##   .- attr(*, "labels")= Named num   0 1 2 3 4 5 8 9
##   .. .- attr(*, "names")= chr  "NAP: ES" "Strongly agree" "Agree" "Neither agree nor dis
## $ V11      : 'haven_labelled' num   3 5 4 4 4 NA 2 5 4 1 ...
##   .- attr(*, "label")= chr "Q2b Men's job earn money, women's job look after home"
##   .- attr(*, "labels")= Named num   0 1 2 3 4 5 8 9
##   .. .- attr(*, "names")= chr  "NAP: ES" "Strongly agree" "Agree" "Neither agree nor dis

```

```

## $ V12      : 'haven_labelled' num  3 NA NA 2 2 NA 2 NA 2 2 ...
##   .- attr(*, "label")= chr "Q3a Should women work: Child under school age"
##   .- attr(*, "labels")= Named num  1 2 3 6 8 9
##   .. .- attr(*, "names")= chr  "Work full-time" "Work part-time" "Stay at home" "TW: wor
## $ V13      : 'haven_labelled' num  2 NA 2 1 2 NA 2 NA 2 2 ...
##   .- attr(*, "label")= chr "Q3b Should women work: Youngest kid at school"
##   .- attr(*, "labels")= Named num  1 2 3 6 8 9
##   .. .- attr(*, "names")= chr  "Work full-time" "Work part-time" "Stay at home" "TW: wor
## $ SEX      : 'haven_labelled' num  1 2 2 2 2 1 2 1 2 2 ...
##   .- attr(*, "label")= chr "Sex of Respondent"
##   .- attr(*, "labels")= Named num  1 2 9
##   .. .- attr(*, "names")= chr  "Male" "Female" "No answer"
## $ AGE      : 'haven_labelled' num  58 59 40 20 72 68 64 57 45 71 ...
##   .- attr(*, "label")= chr "Age of respondent"
##   .- attr(*, "labels")= Named num  15 16 17 18 102 999
##   .. .- attr(*, "names")= chr  "15 years" "16 years" "17 years" "18 years" ...
## $ DEGREE   : 'haven_labelled' num  2 5 5 3 2 NA NA 6 5 6 ...
##   .- attr(*, "label")= chr "Highest completed degree of education: Categories for intern
##   .- attr(*, "labels")= Named num  0 1 2 3 4 5 6 9
##   .. .- attr(*, "names")= chr  "No formal education" "Primary school (elementary school)
## $ MAINSTAT: 'haven_labelled' num  6 6 3 1 6 5 6 2 1 5 ...
##   .- attr(*, "label")= chr "Main status"
##   .- attr(*, "labels")= Named num  1 2 3 4 5 6 7 8 9 99
##   .. .- attr(*, "names")= chr  "In paid work" "Unemployed and looking for a job, HR: inc
## $ TOPBOT   : 'haven_labelled' num  3 7 8 NA 7 2 7 NA 10 6 ...
##   .- attr(*, "label")= chr "Top-Bottom self-placement"
##   .- attr(*, "labels")= Named num  0 1 2 3 4 5 6 7 8 9 ...
##   .. .- attr(*, "names")= chr  "Not available: GB,US" "Lowest, Bottom, 01" "02" "03" ...
## $ HHCHILDR: 'haven_labelled' num  NA NA 3 1 0 NA 0 0 1 NA ...
##   .- attr(*, "label")= chr "How many children in household: children between [school age
##   .- attr(*, "labels")= Named num  0 1 2 21 96 97 99
##   .. .- attr(*, "names")= chr  "No children" "One child" "2 children" "21 children" ...
## $ MARITAL  : 'haven_labelled' num  6 1 1 6 1 6 1 1 1 NA ...
##   .- attr(*, "label")= chr "Legal partnership status"
##   .- attr(*, "labels")= Named num  1 2 3 4 5 6 7 8 9
##   .. .- attr(*, "names")= chr  "Married" "Civil partnership" "Separated from spouse/ ci
## $ URBURAL  : 'haven_labelled' num  1 1 1 NA 1 2 NA 2 2 NA ...
##   .- attr(*, "label")= chr "Place of living: urban - rural"
##   .- attr(*, "labels")= Named num  1 2 3 4 5 7 9
##   .. .- attr(*, "names")= chr  "A big city" "The suburbs or outskirts of a big city" "A
## $ maa      : Factor w/ 25 levels "AU","AT","BG",...: 1 1 1 1 1 1 1 1 1 1 ...
##   .- attr(*, "label")= chr "Country Prefix ISO 3166 Code - alphanumeric"
## $ sp1      : Factor w/ 3 levels "Male","Female",...: 1 2 2 2 2 1 2 1 2 2 ...
##   .- attr(*, "label")= chr "Sex of Respondent"
## $ sp       : Factor w/ 2 levels "m","f": 1 2 2 2 2 1 2 1 2 2 ...
##   .- attr(*, "label")= chr "Sex of Respondent"

```

```
## $ ika      : 'haven_labelled' num  58 59 40 20 72 68 64 57 45 71 ...
##   ..- attr(*, "label")= chr "Age of respondent"
##   ..- attr(*, "labels")= Named num  15 16 17 18 102 999
##   .. ..- attr(*, "names")= chr  "15 years" "16 years" "17 years" "18 years" ...
```

Vaiheessa 2 luodaan samalla samalla periaatteella substanssi- ja taustamuuttujille kaksi R-factor- tyyppin muuttujaa. Toisessa (esim. Q1a) puuttuva tieto on R-ohjelmiston sisäinen NA-arvo. Toisessa (Q1am) puuttuva tieto on yksi luokittelumuuttujan arvo(“missing”).

```
# Substanssi- ja taustamuuttujat R-faktoreiksi
ISSP2012jh1d.dat <- ISSP2012jh1d.dat %>%
  mutate(Q1a1 = as_factor(V5), #labels
         Q1b1 = as_factor(V6),
         Q1c1 = as_factor(V7),
         Q1d1 = as_factor(V8),
         Q1e1 = as_factor(V9),
         Q2a1 = as_factor(V10),
         Q2b1 = as_factor(V11),
         Q3a1 = as_factor(V12), #labels = vastQ3_labels (W,w,H)
         Q3b1 = as_factor(V13), #labels = vastQ3_labels
         edu1 = as_factor(DEGREE),
         msta1 = as_factor(MAINSTAT),
         sosta1 = as_factor(TOPBOT),
         nchild1 = as_factor(HHCHILDR),
         lifsta1 = as_factor(MARITAL),
         urbru1 = as_factor(URBRURAL)

  )

# Muuttujat Q1a1...urbru1 ovat apumuuttujia, joissa on periaatteessa kaikki SPSS-
# tiedostosta siirtyvä metatieto. Poikkeus on SPSS:n kolme tarkentavaa koodia
# puuttuvalle tiedolle, ne saisi mukaan read_spss - parametrin avulla (user_na=TRUE)
#

# Tarkistuksia
ISSP2012jh1d.dat %>% summary()
```

V1	V2	DOI	V3	V4	C_ALPHAN	
Min. :5900	Length:32823	Length:32823	Min. : 36	Min. : 36.0	Length:32823	
1st Qu.:5900	Class :character	Class :character	1st Qu.: 208	1st Qu.:203.0	Class :character	1
Median :5900	Mode :character	Mode :character	Median : 428	Median :276.0	Mode :character	M
Mean :5900	NA	NA	Mean : 4063	Mean :362.1	NA	M
3rd Qu.:5900	NA	NA	3rd Qu.: 705	3rd Qu.:578.0	NA	3
Max. :5900	NA	NA	Max. :62002	Max. :756.0	NA	1
NA	NA	NA	NA	NA	NA	

```
# Substanssimuttajat
```

```
ISSP2012jh1d.dat$Q1a1 %>% fct_count()
```

f	n
NAP: ES	0
Strongly agree	11116
Agree	12352
Neither agree nor disagree	3382
Disagree	4074
Strongly disagree	1051
Can't choose, CA:+NA, KR:DK,ref., NL:DK	0
No answer	0
NA	848

```
ISSP2012jh1d.dat$Q1b1 %>% fct_count()
```

f	n
NAP: ES	0
Strongly agree	2747
Agree	8389
Neither agree nor disagree	5949
Disagree	9003
Strongly disagree	5547
Can't choose, CA:+NA, KR:DK,ref., NL:DK	0
No answer	0
NA	1188

```
ISSP2012jh1d.dat$Q1c1 %>% fct_count()
```

f	n
NAP: ES	0
Strongly agree	2838
Agree	8263
Neither agree nor disagree	6000
Disagree	8706
Strongly disagree	5960
Can't choose, CA:+NA, KR:DK,ref., NL:DK	0
No answer	0
NA	1056



```
ISSP2012jh1d.dat$Q1d1 %>% fct_count()
```

f	n
NAP: ES	0
Strongly agree	2818
Agree	7672
Neither agree nor disagree	7403
Disagree	7863
Strongly disagree	5016
Can't choose, CA:+NA, KR:DK,ref., NL:DK	0
No answer	0
NA	2051

```
ISSP2012jh1d.dat$Q1e1 %>% fct_count()
```

f	n
NAP: ES	0
Strongly agree	3357
Agree	8342
Neither agree nor disagree	7841
Disagree	7267
Strongly disagree	3462
Can't choose, CA:+NA, KR:DK,ref., NL:DK	0
No answer	0
NA	2554

```
ISSP2012jh1d.dat$Q2a1 %>% fct_count()
```

f	n
NAP: ES	0
Strongly agree	11305
Agree	13464
Neither agree nor disagree	5039
Disagree	1929
Strongly disagree	403
Can't choose, CA:+NA, KR:DK,ref., NL:DK	0
No answer	0
NA	683

```
ISSP2012jh1d.dat$Q2b1 %>% fct_count()
```

f	n
NAP: ES	0
Strongly agree	2704
Agree	5164
Neither agree nor disagree	6109
Disagree	9210
Strongly disagree	8917
Can't choose, CA:+NA, KR:DK,ref., NL:DK	0
No answer	0
NA	719

```
ISSP2012jh1d.dat$Q3a1 %>% fct_count()
```

f	n
Work full-time	5373
Work part-time	15655
Stay at home	8367
TW: women should decide	0
Can't choose, CA:+NA, KR:DK,ref., NL:DK	0
No answer	0
NA	3428

```
ISSP2012jh1d.dat$Q3b1 %>% fct_count()
```

f	n
Work full-time	13722
Work part-time	13817
Stay at home	1762
TW: women should decide	0
Can't choose, CA:+NA, KR:DK,ref., NL:DK	0
No answer	0
NA	3522

```
# Taustamuuttujat
```

```
ISSP2012jh1d.dat$edu1 %>% fct_count()
```

---

f

---

No formal education

Primary school (elementary school)

Lower secondary (secondary completed does not allow entry to university: obligatory school)

Upper secondary (programs that allows entry to university)

Post secondary, non-tertiary (other upper secondary programs toward labour market or technical formation)

Lower level tertiary, first stage (also technical schools at a tertiary level)

Upper level tertiary (Master, Dr.)

No answer, CH: don't know

NA

---

```
ISSP2012jh1d.dat$msta1 %>% fct_count()
```

---

f	n
In paid work	17967
Unemployed and looking for a job, HR: incl never had a job	1769
In education	1763
Apprentice or trainee	189
Permanently sick or disabled	1093
Retired	7999
Domestic work	1180
In compulsory military service or community service	9
Other	484
No answer	0
NA	370

---

```
ISSP2012jh1d.dat$sosta1 %>% fct_count()
```

---

f	n
Not available: GB,US	0
Lowest, Bottom, 01	562
02	866
03	2221
04	3346
05	6798
06	6889
07	5778
08	3477
09	667
Highest, Top, 10	442
Refused	0
Don't know	0

---

f	n
No answer	0
NA	1777

```
ISSP2012jh1d.dat$nchild1 %>% fct_count()
```

f	n
No children	24102
One child	4378
2 children	2643
3	598
4	117
5	20
6	13
7	7
8	3
18	1
21 children	1
NAP (Code 0 in HOMPOP); not available: TR	0
Refused	0
No answer	0
NA	940

```
ISSP2012jh1d.dat$lifsta1 %>% fct_count()
```

f	n
Married	17573
Civil partnership	1035
Separated from spouse/ civil partner (still legally married/ still legally in a civil partnership)	486
Divorced from spouse/ legally separated from civil partner	2997
Widowed/ civil partner died	2763
Never married/ never in a civil partnership, single	7535
Refused	0
Don't know	0
No answer	0
NA	434

```
ISSP2012jh1d.dat$urbru1 %>% fct_count()
```

f	n
A big city	8442

f	n
The suburbs or outskirts of a big city	4386
A town or a small city	9203
A country village	8646
A farm or home in the country	1902
Other answer	0
No answer	0
NA	244

Taustamuuttujien luokitteluja (esim. luokkien yhdistäminen) pohditaan tarkemmin, kun muuttujat otetaan käyttöön.

Poistetaan muuuttujista luokittelumuuttujien arvot, joissa ei ole havaintoja. Näitä tyhjiä luokkia siirtyy SPSS-tiedostosta haven\_labelled -luokan tietohin.

*# Poistetaan tyhjät luokat muuttujista*

```
ISSP2012jh1d.dat <- ISSP2012jh1d.dat %>%
  mutate(Q1a = fct_drop(Q1a1),
         Q1b = fct_drop(Q1b1),
         Q1c = fct_drop(Q1c1),
         Q1d = fct_drop(Q1d1),
         Q1e = fct_drop(Q1e1),
         Q2a = fct_drop(Q2a1),
         Q2b = fct_drop(Q2b1),
         Q3a = fct_drop(Q3a1),
         Q3b = fct_drop(Q3b1),
         edu = fct_drop(edu1),
         msta = fct_drop(msta1),
         sosta = fct_drop(sosta1),
         nchild = fct_drop(nchild1),
         lifsta = fct_drop(lifsta1),
         urbru = fct_drop(urbru1)
  )
# Tarkistuksia 1

ISSP2012jh1d.dat %>% summary()
```

V1	V2	DOI	V3	V4	C_ALPHAN	
Min. :5900	Length:32823	Length:32823	Min. : 36	Min. : 36.0	Length:32823	
1st Qu.:5900	Class :character	Class :character	1st Qu.: 208	1st Qu.:203.0	Class :character	1
Median :5900	Mode :character	Mode :character	Median : 428	Median :276.0	Mode :character	M
Mean :5900	NA	NA	Mean : 4063	Mean :362.1	NA	M
3rd Qu.:5900	NA	NA	3rd Qu.: 705	3rd Qu.:578.0	NA	3

V1	V2	DOI	V3	V4	C_ALPHAN
Max. :5900	NA	NA	Max. :62002	Max. :756.0	NA
NA	NA	NA	NA	NA	NA

```
ISSP2012jh1d.dat %>%
```

```
  select(Q1a, Q1b, Q1c, Q1d, Q1e,Q2a,Q2b,Q3a, Q3b) %>%
  str()
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':   32823 obs. of  9 variables:
## $ Q1a: Factor w/ 5 levels "Strongly agree",...: 5 1 2 2 1 NA 2 4 2 2 ...
##   ..- attr(*, "label")= chr "Q1a Working mom: warm relationship with children as a not w
## $ Q1b: Factor w/ 5 levels "Strongly agree",...: 1 5 4 4 4 NA 4 3 4 3 ...
##   ..- attr(*, "label")= chr "Q1b Working mom: Preschool child is likely to suffer"
## $ Q1c: Factor w/ 5 levels "Strongly agree",...: 3 5 2 4 4 NA 4 2 4 2 ...
##   ..- attr(*, "label")= chr "Q1c Working woman: Family life suffers when woman has full-t
## $ Q1d: Factor w/ 5 levels "Strongly agree",...: 3 5 5 2 4 NA 4 5 4 5 ...
##   ..- attr(*, "label")= chr "Q1d Working woman: What women really want is home and kids"
## $ Q1e: Factor w/ 5 levels "Strongly agree",...: 3 1 2 3 4 NA 2 4 4 1 ...
##   ..- attr(*, "label")= chr "Q1e Working woman: Being housewife is as fulfilling as worki
## $ Q2a: Factor w/ 5 levels "Strongly agree",...: 1 3 4 2 2 NA 2 5 2 1 ...
##   ..- attr(*, "label")= chr "Q2a Both should contribute to household income"
## $ Q2b: Factor w/ 5 levels "Strongly agree",...: 3 5 4 4 4 NA 2 5 4 1 ...
##   ..- attr(*, "label")= chr "Q2b Men's job earn money, women's job look after home"
## $ Q3a: Factor w/ 3 levels "Work full-time",...: 3 NA NA 2 2 NA 2 NA 2 2 ...
##   ..- attr(*, "label")= chr "Q3a Should women work: Child under school age"
## $ Q3b: Factor w/ 3 levels "Work full-time",...: 2 NA 2 1 2 NA 2 NA 2 2 ...
##   ..- attr(*, "label")= chr "Q3b Should women work: Youngest kid at school"
```

```
ISSP2012jh1d.dat %>%
```

```
  select(Q1a1, Q1b1, Q1c1, Q1d1, Q1e1,Q2a1,Q2b1,Q3a1, Q3b1) %>%
  str()
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':   32823 obs. of  9 variables:
## $ Q1a1: Factor w/ 8 levels "NAP: ES","Strongly agree",...: 6 2 3 3 2 NA 3 5 3 3 ...
##   ..- attr(*, "label")= chr "Q1a Working mom: warm relationship with children as a not w
## $ Q1b1: Factor w/ 8 levels "NAP: ES","Strongly agree",...: 2 6 5 5 5 NA 5 4 5 4 ...
##   ..- attr(*, "label")= chr "Q1b Working mom: Preschool child is likely to suffer"
## $ Q1c1: Factor w/ 8 levels "NAP: ES","Strongly agree",...: 4 6 3 5 5 NA 5 3 5 3 ...
##   ..- attr(*, "label")= chr "Q1c Working woman: Family life suffers when woman has full-t
## $ Q1d1: Factor w/ 8 levels "NAP: ES","Strongly agree",...: 4 6 6 3 5 NA 5 6 5 6 ...
##   ..- attr(*, "label")= chr "Q1d Working woman: What women really want is home and kids"
## $ Q1e1: Factor w/ 8 levels "NAP: ES","Strongly agree",...: 4 2 3 4 5 NA 3 5 5 2 ...
##   ..- attr(*, "label")= chr "Q1e Working woman: Being housewife is as fulfilling as worki
## $ Q2a1: Factor w/ 8 levels "NAP: ES","Strongly agree",...: 2 4 5 3 3 NA 3 6 3 2 ...
##   ..- attr(*, "label")= chr "Q2a Both should contribute to household income"
```

```
## $ Q2b1: Factor w/ 8 levels "NAP: ES","Strongly agree",...: 4 6 5 5 5 NA 3 6 5 2 ...
##   ..- attr(*, "label")= chr "Q2b Men's job earn money, women's job look after home"
## $ Q3a1: Factor w/ 6 levels "Work full-time",...: 3 NA NA 2 2 NA 2 NA 2 2 ...
##   ..- attr(*, "label")= chr "Q3a Should women work: Child under school age"
## $ Q3b1: Factor w/ 6 levels "Work full-time",...: 2 NA 2 1 2 NA 2 NA 2 2 ...
##   ..- attr(*, "label")= chr "Q3b Should women work: Youngest kid at school"
```

```
ISSP2012jh1d.dat %>%
```

```
  select(edu, msta, sosta, nchild, lifsta, urbru) %>%
  str()
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame': 32823 obs. of 6 variables:
## $ edu : Factor w/ 7 levels "No formal education",...: 3 6 6 4 3 NA NA 7 6 7 ...
##   ..- attr(*, "label")= chr "Highest completed degree of education: Categories for intern
## $ msta : Factor w/ 9 levels "In paid work",...: 6 6 3 1 6 5 6 2 1 5 ...
##   ..- attr(*, "label")= chr "Main status"
## $ sosta : Factor w/ 10 levels "Lowest, Bottom, 01",...: 3 7 8 NA 7 2 7 NA 10 6 ...
##   ..- attr(*, "label")= chr "Top-Bottom self-placement"
## $ nchild: Factor w/ 11 levels "No children",...: NA NA 4 2 1 NA 1 1 2 NA ...
##   ..- attr(*, "label")= chr "How many children in household: children between [school age
## $ lifsta: Factor w/ 6 levels "Married","Civil partnership",...: 6 1 1 6 1 6 1 1 1 NA ...
##   ..- attr(*, "label")= chr "Legal partnership status"
## $ urbru : Factor w/ 5 levels "A big city","The suburbs or outskirts of a big city",...: 1
##   ..- attr(*, "label")= chr "Place of living: urban - rural"
```

```
ISSP2012jh1d.dat %>%
```

```
  select(edu1, msta1, sosta1, nchild1, lifsta1, urbru1) %>%
  str()
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame': 32823 obs. of 6 variables:
## $ edu1 : Factor w/ 8 levels "No formal education",...: 3 6 6 4 3 NA NA 7 6 7 ...
##   ..- attr(*, "label")= chr "Highest completed degree of education: Categories for intern
## $ msta1 : Factor w/ 10 levels "In paid work",...: 6 6 3 1 6 5 6 2 1 5 ...
##   ..- attr(*, "label")= chr "Main status"
## $ sosta1 : Factor w/ 14 levels "Not available: GB,US",...: 4 8 9 NA 8 3 8 NA 11 7 ...
##   ..- attr(*, "label")= chr "Top-Bottom self-placement"
## $ nchild1: Factor w/ 14 levels "No children",...: NA NA 4 2 1 NA 1 1 2 NA ...
##   ..- attr(*, "label")= chr "How many children in household: children between [school age
## $ lifsta1: Factor w/ 9 levels "Married","Civil partnership",...: 6 1 1 6 1 6 1 1 1 NA ...
##   ..- attr(*, "label")= chr "Legal partnership status"
## $ urbru1 : Factor w/ 7 levels "A big city","The suburbs or outskirts of a big city",...: 1
##   ..- attr(*, "label")= chr "Place of living: urban - rural"
```

```
# Tarkistuksia 2
```

```
ISSP2012jh1d.dat %>% tableX(Q1a,Q1a1)
```

Q1a/Q1a1	2	3	4	5	6	Missing	Total
1	11116	0	0	0	0	0	11116
2	0	12352	0	0	0	0	12352
3	0	0	3382	0	0	0	3382
4	0	0	0	4074	0	0	4074
5	0	0	0	0	1051	0	1051
Missing	0	0	0	0	0	848	848
Total	11116	12352	3382	4074	1051	848	32823

ISSP2012jh1d.dat %>% tableX(Q1b,Q1b1)

Q1b/Q1b1	2	3	4	5	6	Missing	Total
1	2747	0	0	0	0	0	2747
2	0	8389	0	0	0	0	8389
3	0	0	5949	0	0	0	5949
4	0	0	0	9003	0	0	9003
5	0	0	0	0	5547	0	5547
Missing	0	0	0	0	0	1188	1188
Total	2747	8389	5949	9003	5547	1188	32823

ISSP2012jh1d.dat %>% tableX(Q1c,Q1c1)

Q1c/Q1c1	2	3	4	5	6	Missing	Total
1	2838	0	0	0	0	0	2838
2	0	8263	0	0	0	0	8263
3	0	0	6000	0	0	0	6000
4	0	0	0	8706	0	0	8706
5	0	0	0	0	5960	0	5960
Missing	0	0	0	0	0	1056	1056
Total	2838	8263	6000	8706	5960	1056	32823

ISSP2012jh1d.dat %>% tableX(Q1d,Q1d1)

Q1d/Q1d1	2	3	4	5	6	Missing	Total
1	2818	0	0	0	0	0	2818
2	0	7672	0	0	0	0	7672
3	0	0	7403	0	0	0	7403
4	0	0	0	7863	0	0	7863
5	0	0	0	0	5016	0	5016
Missing	0	0	0	0	0	2051	2051



Q1d/Q1d1	2	3	4	5	6	Missing	Total
Total	2818	7672	7403	7863	5016	2051	32823

```
ISSP2012jh1d.dat %>% tableX(Q1e,Q1e1)
```

Q1e/Q1e1	2	3	4	5	6	Missing	Total
1	3357	0	0	0	0	0	3357
2	0	8342	0	0	0	0	8342
3	0	0	7841	0	0	0	7841
4	0	0	0	7267	0	0	7267
5	0	0	0	0	3462	0	3462
Missing	0	0	0	0	0	2554	2554
Total	3357	8342	7841	7267	3462	2554	32823

```
ISSP2012jh1d.dat %>% tableX(Q2a,Q2a1)
```

Q2a/Q2a1	2	3	4	5	6	Missing	Total
1	11305	0	0	0	0	0	11305
2	0	13464	0	0	0	0	13464
3	0	0	5039	0	0	0	5039
4	0	0	0	1929	0	0	1929
5	0	0	0	0	403	0	403
Missing	0	0	0	0	0	683	683
Total	11305	13464	5039	1929	403	683	32823

```
ISSP2012jh1d.dat %>% tableX(Q2b,Q2b1)
```

Q2b/Q2b1	2	3	4	5	6	Missing	Total
1	2704	0	0	0	0	0	2704
2	0	5164	0	0	0	0	5164
3	0	0	6109	0	0	0	6109
4	0	0	0	9210	0	0	9210
5	0	0	0	0	8917	0	8917
Missing	0	0	0	0	0	719	719
Total	2704	5164	6109	9210	8917	719	32823

```
ISSP2012jh1d.dat %>% tableX(Q3a,Q3a1)
```

Q3a/Q3a1	1	2	3	Missing	Total
1	5373	0	0	0	5373
2	0	15655	0	0	15655
3	0	0	8367	0	8367
Missing	0	0	0	3428	3428
Total	5373	15655	8367	3428	32823

```
ISSP2012jh1d.dat %>% tableX(Q3b,Q3b1)
```

Q3b/Q3b1	1	2	3	Missing	Total
1	13722	0	0	0	13722
2	0	13817	0	0	13817
3	0	0	1762	0	1762
Missing	0	0	0	3522	3522
Total	13722	13817	1762	3522	32823

```
ISSP2012jh1d.dat %>% tableX(edu,edu1)
```

edu/edu1	1	2	3	4	5	6	7	Missing	Total
1	491	0	0	0	0	0	0	0	491
2	0	1531	0	0	0	0	0	0	1531
3	0	0	7811	0	0	0	0	0	7811
4	0	0	0	7115	0	0	0	0	7115
5	0	0	0	0	5658	0	0	0	5658
6	0	0	0	0	0	5147	0	0	5147
7	0	0	0	0	0	0	4762	0	4762
Missing	0	0	0	0	0	0	0	308	308
Total	491	1531	7811	7115	5658	5147	4762	308	32823

```
ISSP2012jh1d.dat %>% tableX(msta,msta1)
```

msta/msta1	1	2	3	4	5	6	7	8	9	Missing	Total
1	17967	0	0	0	0	0	0	0	0	0	17967
2	0	1769	0	0	0	0	0	0	0	0	1769
3	0	0	1763	0	0	0	0	0	0	0	1763
4	0	0	0	189	0	0	0	0	0	0	189
5	0	0	0	0	1093	0	0	0	0	0	1093
6	0	0	0	0	0	7999	0	0	0	0	7999
7	0	0	0	0	0	0	1180	0	0	0	1180
8	0	0	0	0	0	0	0	9	0	0	9

msta/msta1	1	2	3	4	5	6	7	8	9	Missing	Total
9	0	0	0	0	0	0	0	0	484	0	484
Missing	0	0	0	0	0	0	0	0	0	370	370
Total	17967	1769	1763	189	1093	7999	1180	9	484	370	32823

```
ISSP2012jh1d.dat %>% tableX(sosta,sosta1)
```

sosta/sosta1	10	11	2	3	4	5	6	7	8	9	Missing	Total
1	0	0	562	0	0	0	0	0	0	0	0	562
10	0	442	0	0	0	0	0	0	0	0	0	442
2	0	0	0	866	0	0	0	0	0	0	0	866
3	0	0	0	0	2221	0	0	0	0	0	0	2221
4	0	0	0	0	0	3346	0	0	0	0	0	3346
5	0	0	0	0	0	0	6798	0	0	0	0	6798
6	0	0	0	0	0	0	0	6889	0	0	0	6889
7	0	0	0	0	0	0	0	0	5778	0	0	5778
8	0	0	0	0	0	0	0	0	0	3477	0	3477
9	667	0	0	0	0	0	0	0	0	0	0	667
Missing	0	0	0	0	0	0	0	0	0	0	1777	1777
Total	667	442	562	866	2221	3346	6798	6889	5778	3477	1777	32823

```
ISSP2012jh1d.dat %>% tableX(nchild,nchild1)
```

nchild/nchild1	1	10	11	2	3	4	5	6	7	8	9	Missing	Total
1	24102	0	0	0	0	0	0	0	0	0	0	0	24102
10	0	1	0	0	0	0	0	0	0	0	0	0	1
11	0	0	1	0	0	0	0	0	0	0	0	0	1
2	0	0	0	4378	0	0	0	0	0	0	0	0	4378
3	0	0	0	0	2643	0	0	0	0	0	0	0	2643
4	0	0	0	0	0	598	0	0	0	0	0	0	598
5	0	0	0	0	0	0	117	0	0	0	0	0	117
6	0	0	0	0	0	0	0	20	0	0	0	0	20
7	0	0	0	0	0	0	0	0	13	0	0	0	13
8	0	0	0	0	0	0	0	0	0	7	0	0	7
9	0	0	0	0	0	0	0	0	0	0	3	0	3
Missing	0	0	0	0	0	0	0	0	0	0	0	940	940
Total	24102	1	1	4378	2643	598	117	20	13	7	3	940	32823

```
ISSP2012jh1d.dat %>% tableX(lifsta,lifsta1)
```

lifsta/lifsta1	1	2	3	4	5	6	Missing	Total
1	17573	0	0	0	0	0	0	17573
2	0	1035	0	0	0	0	0	1035
3	0	0	486	0	0	0	0	486
4	0	0	0	2997	0	0	0	2997
5	0	0	0	0	2763	0	0	2763
6	0	0	0	0	0	7535	0	7535
Missing	0	0	0	0	0	0	434	434
Total	17573	1035	486	2997	2763	7535	434	32823

```
ISSP2012jh1d.dat %>% tableX(urbru,urbru1)
```

urbru/urbru1	1	2	3	4	5	Missing	Total
1	8442	0	0	0	0	0	8442
2	0	4386	0	0	0	0	4386
3	0	0	9203	0	0	0	9203
4	0	0	0	8646	0	0	8646
5	0	0	0	0	1902	0	1902
Missing	0	0	0	0	0	244	244
Total	8442	4386	9203	8646	1902	244	32823

Luodaan uusi muuttuja, jossa puuttuva tieto (NA) on mukana luokittelumuuttujan uutena arvona ("missing").

*# Uusi muuttuja, jossa NA-arvot ovat mukana muuttujan uutena luokkana. Muuttujat # nimetään Q1a -> Q1am.*

```
ISSP2012jh1d.dat <- ISSP2012jh1d.dat %>%
  mutate(Q1am = fct_explicit_na(Q1a, na_level = "missing"),
         Q1bm = fct_explicit_na(Q1b, na_level = "missing"),
         Q1cm = fct_explicit_na(Q1c, na_level = "missing"),
         Q1dm = fct_explicit_na(Q1d, na_level = "missing"),
         Q1em = fct_explicit_na(Q1e, na_level = "missing"),
         Q2am = fct_explicit_na(Q2a, na_level = "missing"),
         Q2bm = fct_explicit_na(Q2b, na_level = "missing"),
         Q3am = fct_explicit_na(Q3a, na_level = "missing"),
         Q3bm = fct_explicit_na(Q3b, na_level = "missing"),
         edum = fct_explicit_na(edu, na_level = "missing"),
         mstam = fct_explicit_na(msta, na_level = "missing"),
         sostam = fct_explicit_na(sosta, na_level = "missing"),
         nchildm = fct_explicit_na(nchild, na_level = "missing"),
         lifstam = fct_explicit_na(lifsta, na_level = "missing"),
         urbrum = fct_explicit_na(urbru, na_level = "missing"),
```

```

)
# Tarkistuksia

ISSP2012jh1d.dat %>%
  select(Q1am, Q1bm, Q1cm, Q1dm, Q1em, Q2am, Q2bm, Q3am, Q3bm) %>%
  summary()

```

Q1am	Q1bm	Q1cm	
Strongly agree :11116	Strongly agree :2747	Strongly agree :2838	
Agree :12352	Agree :8389	Agree :8263	
Neither agree nor disagree: 3382	Neither agree nor disagree:5949	Neither agree nor disagree:6000	Nei
Disagree : 4074	Disagree :9003	Disagree :8706	
Strongly disagree : 1051	Strongly disagree :5547	Strongly disagree :5960	
missing : 848	missing :1188	missing :1056	

```

ISSP2012jh1d.dat %>%
  select(edum,mstam, sostam,nchildm,lifstam, urbrum) %>%
  summary()

```

edum
Lower secondary (secondary completed does not allow entry to university: obligatory school) :7811
Upper secondary (programs that allows entry to university :7115
Post secondary, non-tertiary (other upper secondary programs toward labour market or technical formation
Lower level tertiary, first stage (also technical schools at a tertiary level) :5147
Upper level tertiary (Master, Dr.) :4762
Primary school (elementary school) :1531
(Other) : 799

```

ISSP2012jh1d.dat %>%
  select(Q1am, Q1bm, Q1cm, Q1dm, Q1em, Q2am, Q2bm, Q3am, Q3bm) %>%
  str()

```

```

## Classes 'tbl_df', 'tbl' and 'data.frame': 32823 obs. of 9 variables:
## $ Q1am: Factor w/ 6 levels "Strongly agree",...: 5 1 2 2 1 6 2 4 2 2 ...
## ..- attr(*, "label")= chr "Q1a Working mom: warm relationship with children as a not w
## $ Q1bm: Factor w/ 6 levels "Strongly agree",...: 1 5 4 4 4 6 4 3 4 3 ...
## ..- attr(*, "label")= chr "Q1b Working mom: Preschool child is likely to suffer"
## $ Q1cm: Factor w/ 6 levels "Strongly agree",...: 3 5 2 4 4 6 4 2 4 2 ...
## ..- attr(*, "label")= chr "Q1c Working woman: Family life suffers when woman has full-t
## $ Q1dm: Factor w/ 6 levels "Strongly agree",...: 3 5 5 2 4 6 4 5 4 5 ...
## ..- attr(*, "label")= chr "Q1d Working woman: What women really want is home and kids"
## $ Q1em: Factor w/ 6 levels "Strongly agree",...: 3 1 2 3 4 6 2 4 4 1 ...

```

```
##   ..- attr(*, "label")= chr "Q1e Working woman: Being housewife is as fulfilling as worki
##   $ Q2am: Factor w/ 6 levels "Strongly agree",...: 1 3 4 2 2 6 2 5 2 1 ...
##   ..- attr(*, "label")= chr "Q2a Both should contribute to household income"
##   $ Q2bm: Factor w/ 6 levels "Strongly agree",...: 3 5 4 4 4 6 2 5 4 1 ...
##   ..- attr(*, "label")= chr "Q2b Men's job earn money, women's job look after home"
##   $ Q3am: Factor w/ 4 levels "Work full-time",...: 3 4 4 2 2 4 2 4 2 2 ...
##   ..- attr(*, "label")= chr "Q3a Should women work: Child under school age"
##   $ Q3bm: Factor w/ 4 levels "Work full-time",...: 2 4 2 1 2 4 2 4 2 2 ...
##   ..- attr(*, "label")= chr "Q3b Should women work: Youngest kid at school"
```

```
ISSP2012jh1d.dat %>%
  select(edum,mstam, sostam,nchildm,lifstam, urbrum) %>%
  str()
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':   32823 obs. of  6 variables:
##  $ edum   : Factor w/ 8 levels "No formal education",...: 3 6 6 4 3 8 8 7 6 7 ...
##  ..- attr(*, "label")= chr "Highest completed degree of education: Categories for intern
##  $ mstam   : Factor w/ 10 levels "In paid work",...: 6 6 3 1 6 5 6 2 1 5 ...
##  ..- attr(*, "label")= chr "Main status"
##  $ sostam  : Factor w/ 11 levels "Lowest, Bottom, 01",...: 3 7 8 11 7 2 7 11 10 6 ...
##  ..- attr(*, "label")= chr "Top-Bottom self-placement"
##  $ nchildm: Factor w/ 12 levels "No children",...: 12 12 4 2 1 12 1 1 2 12 ...
##  ..- attr(*, "label")= chr "How many children in household: children between [school age
##  $ lifstam : Factor w/ 7 levels "Married","Civil partnership",...: 6 1 1 6 1 6 1 1 1 7 ...
##  ..- attr(*, "label")= chr "Legal partnership status"
##  $ urbrum  : Factor w/ 6 levels "A big city","The suburbs or outskirts of a big city",...
##  ..- attr(*, "label")= chr "Place of living: urban - rural"
```

```
# Taustamuuttuja, puuttuva tieto mukana
```

```
ISSP2012jh1d.dat$edum %>% fct_count()
```

---

```
f
```

```
No formal education
```

```
Primary school (elementary school)
```

```
Lower secondary (secondary completed does not allow entry to university: obligatory school)
```

```
Upper secondary (programs that allows entry to university)
```

```
Post secondary, non-tertiary (other upper secondary programs toward labour market or technical formation)
```

```
Lower level tertiary, first stage (also technical schools at a tertiary level)
```

```
Upper level tertiary (Master, Dr.)
```

```
missing
```

---

```
ISSP2012jh1d.dat$mstam %>% fct_count()
```

f	n
In paid work	17967
Unemployed and looking for a job, HR: incl never had a job	1769
In education	1763
Apprentice or trainee	189
Permanently sick or disabled	1093
Retired	7999
Domestic work	1180
In compulsory military service or community service	9
Other	484
missing	370

```
ISSP2012jh1d.dat$sostam %>% fct_count()
```

f	n
Lowest, Bottom, 01	562
02	866
03	2221
04	3346
05	6798
06	6889
07	5778
08	3477
09	667
Highest, Top, 10	442
missing	1777

```
ISSP2012jh1d.dat$nchildm %>% fct_count()
```

f	n
No children	24102
One child	4378
2 children	2643
3	598
4	117
5	20
6	13
7	7
8	3
18	1
21 children	1
missing	940

```
ISSP2012jh1d.dat$lifstam %>% fct_count()
```

f	n
Married	17573
Civil partnership	1035
Separated from spouse/ civil partner (still legally married/ still legally in a civil partnership)	486
Divorced from spouse/ legally separated from civil partner	2997
Widowed/ civil partner died	2763
Never married/ never in a civil partnership, single	7535
missing	434

```
ISSP2012jh1d.dat$surbrum %>% fct_count()
```

f	n
A big city	8442
The suburbs or outskirts of a big city	4386
A town or a small city	9203
A country village	8646
A farm or home in the country	1902
missing	244

*# Substanssimuuttujat, puuttuva tieto mukana*

```
ISSP2012jh1d.dat$Q1am %>% fct_count()
```

f	n
Strongly agree	11116
Agree	12352
Neither agree nor disagree	3382
Disagree	4074
Strongly disagree	1051
missing	848

```
ISSP2012jh1d.dat$Q1bm %>% fct_count()
```

f	n
Strongly agree	2747
Agree	8389
Neither agree nor disagree	5949
Disagree	9003



f	n
Strongly disagree	5547
missing	1188

```
ISSP2012jh1d.dat$Q1cm %>% fct_count()
```

f	n
Strongly agree	2838
Agree	8263
Neither agree nor disagree	6000
Disagree	8706
Strongly disagree	5960
missing	1056

```
ISSP2012jh1d.dat$Q1dm %>% fct_count()
```

f	n
Strongly agree	2818
Agree	7672
Neither agree nor disagree	7403
Disagree	7863
Strongly disagree	5016
missing	2051

```
ISSP2012jh1d.dat$Q1em %>% fct_count()
```

f	n
Strongly agree	3357
Agree	8342
Neither agree nor disagree	7841
Disagree	7267
Strongly disagree	3462
missing	2554

```
ISSP2012jh1d.dat$Q2am %>% fct_count()
```

f	n
Strongly agree	11305
Agree	13464

f	n
Neither agree nor disagree	5039
Disagree	1929
Strongly disagree	403
missing	683

```
ISSP2012jh1d.dat$Q2bm %>% fct_count()
```

f	n
Strongly agree	2704
Agree	5164
Neither agree nor disagree	6109
Disagree	9210
Strongly disagree	8917
missing	719

```
ISSP2012jh1d.dat$Q3am %>% fct_count()
```

f	n
Work full-time	5373
Work part-time	15655
Stay at home	8367
missing	3428

```
ISSP2012jh1d.dat$Q3bm %>% fct_count()
```

f	n
Work full-time	13722
Work part-time	13817
Stay at home	1762
missing	3522

Lopuksi luodaan uuden “faktorilabelit” substanssimuuttujille. Graafisessa analyysissä kuviin on saatava mukaan kaikki oleellinen, mutta ei mitään sen lisäksi. Näitä muuttujan arvojen tunnuksia muokataan tarvittaessa.

```
# Vaihe 2.4.1
```

```
# Viisi vastausvaihtoehtoa - ei eksplisiittistä NA-tietoa("missing")
# Q3a - Q3b kolme vastausvaihtoehtoa
```

```

ISSP2012jh1d.dat <- ISSP2012jh1d.dat %>%
  mutate(Q1a = fct_recode(Q1a,
    "S" = "Strongly agree",
    "s" = "Agree",
    "?" = "Neither agree nor disagree",
    "e" = "Disagree",
    "E" = "Strongly disagree"),
    Q1b = fct_recode(Q1b,
    "S" = "Strongly agree",
    "s" = "Agree",
    "?" = "Neither agree nor disagree",
    "e" = "Disagree",
    "E" = "Strongly disagree"),
    Q1c = fct_recode(Q1c,
    "S" = "Strongly agree",
    "s" = "Agree",
    "?" = "Neither agree nor disagree",
    "e" = "Disagree",
    "E" = "Strongly disagree"),
    Q1d = fct_recode(Q1d,
    "S" = "Strongly agree",
    "s" = "Agree",
    "?" = "Neither agree nor disagree",
    "e" = "Disagree",
    "E" = "Strongly disagree"),
    Q1e = fct_recode(Q1e,
    "S" = "Strongly agree",
    "s" = "Agree",
    "?" = "Neither agree nor disagree",
    "e" = "Disagree",
    "E" = "Strongly disagree"),
    Q2a = fct_recode(Q2a,
    "S" = "Strongly agree",
    "s" = "Agree",
    "?" = "Neither agree nor disagree",
    "e" = "Disagree",
    "E" = "Strongly disagree" ),
    Q2b = fct_recode(Q2b,
    "S" = "Strongly agree",
    "s" = "Agree",
    "?" = "Neither agree nor disagree",
    "e" = "Disagree",
    "E" = "Strongly disagree"),
    Q3a = fct_recode(Q3a,
    "W" = "Work full-time",

```

```

        "w" = "Work part-time",
        "H" = "Stay at home" ),
  Q3b = fct_recode(Q3b,
    "W" = "Work full-time",
    "w" = "Work part-time",
    "H" = "Stay at home" )
)

# Tarkistuksia 1
ISSP2012jh1d.dat %>%
  select(Q1a, Q1b, Q1c, Q1d, Q1e, Q2a, Q2b, Q3a, Q3b) %>%
  summary()

```

Q1a	Q1b	Q1c	Q1d	Q1e	Q2a	Q2b	Q3a	Q3b
S :11116	S :2747	S :2838	S :2818	S :3357	S :11305	S :2704	W : 5373	W : 5373
s :12352	s :8389	s :8263	s :7672	s :8342	s :13464	s :5164	w :15655	w :15655
? : 3382	? :5949	? :6000	? :7403	? :7841	? : 5039	? :6109	H : 8367	H : 8367
e : 4074	e :9003	e :8706	e :7863	e :7267	e : 1929	e :9210	NA's: 3428	NA's: 3428
E : 1051	E :5547	E :5960	E :5016	E :3462	E : 403	E :8917	NA	NA
NA's: 848	NA's:1188	NA's:1056	NA's:2051	NA's:2554	NA's: 683	NA's: 719	NA	NA

```

# Vaihe 2.4.2 - muuttujassa eksplisiittinen NA-tieto
ISSP2012jh1d.dat <- ISSP2012jh1d.dat %>%
  mutate(Q1am = fct_recode(Q1am,
    "S" = "Strongly agree",
    "s" = "Agree",
    "?" = "Neither agree nor disagree",
    "e" = "Disagree",
    "E" = "Strongly disagree",
    "P" = "missing"),
    Q1bm = fct_recode(Q1bm,
    "S" = "Strongly agree",
    "s" = "Agree",
    "?" = "Neither agree nor disagree",
    "e" = "Disagree",
    "E" = "Strongly disagree",
    "P" = "missing"),
    Q1cm = fct_recode(Q1cm,
    "S" = "Strongly agree",
    "s" = "Agree",
    "?" = "Neither agree nor disagree",
    "e" = "Disagree",
    "E" = "Strongly disagree",
    "P" = "missing"))

```

```

        "P" = "missing"),
Q1dm =fct_recode(Q1dm,
        "S" ="Strongly agree",
        "s" = "Agree",
        "?" = "Neither agree nor disagree",
        "e" = "Disagree",
        "E"= "Strongly disagree",
        "P" = "missing"),
Q1em =fct_recode(Q1em,
        "S" ="Strongly agree",
        "s" = "Agree",
        "?" = "Neither agree nor disagree",
        "e" = "Disagree",
        "E"= "Strongly disagree",
        "P" = "missing"),
Q2am =fct_recode(Q2am,
        "S" ="Strongly agree",
        "s" = "Agree",
        "?" = "Neither agree nor disagree",
        "e" = "Disagree",
        "E"= "Strongly disagree",
        "P" = "missing"),
Q2bm =fct_recode(Q2bm,
        "S" ="Strongly agree",
        "s" = "Agree",
        "?" = "Neither agree nor disagree",
        "e" = "Disagree",
        "E"= "Strongly disagree",
        "P" = "missing"),
Q3am = fct_recode(Q3am,
        "W" = "Work full-time",
        "w" = "Work part-time",
        "H" = "Stay at home",
        "P" = "missing"),
Q3bm = fct_recode(Q3bm,
        "W" = "Work full-time",
        "w" = "Work part-time",
        "H" = "Stay at home",
        "P" = "missing")
)

# Tarkistuksia 2

ISSP2012jh1d.dat %>%
  select(Q1am, Q1bm, Q1cm, Q1dm, Q1em, Q2am, Q2bm, Q3am, Q3bm) %>%

```

```
summary()
```

Q1am	Q1bm	Q1cm	Q1dm	Q1em	Q2am	Q2bm	Q3am	Q3bm
S:11116	S:2747	S:2838	S:2818	S:3357	S:11305	S:2704	W: 5373	W:13722
s:12352	s:8389	s:8263	s:7672	s:8342	s:13464	s:5164	w:15655	w:13817
?: 3382	?:5949	?:6000	?:7403	?:7841	?: 5039	?:6109	H: 8367	H: 1762
e: 4074	e:9003	e:8706	e:7863	e:7267	e: 1929	e:9210	P: 3428	P: 3522
E: 1051	E:5547	E:5960	E:5016	E:3462	E: 403	E:8917	NA	NA
P: 848	P:1188	P:1056	P:2051	P:2554	P: 683	P: 719	NA	NA

```
# Tarkistuksia 3
```

```
ISSP2012jh1d.dat %>%
```

```
tableX(Q1a,Q1am)
```

Q1a/Q1am	S	s	?	e	E	P	Total
1	11116	0	0	0	0	0	11116
2	0	12352	0	0	0	0	12352
3	0	0	3382	0	0	0	3382
4	0	0	0	4074	0	0	4074
5	0	0	0	0	1051	0	1051
Missing	0	0	0	0	0	848	848
Total	11116	12352	3382	4074	1051	848	32823

```
ISSP2012jh1d.dat %>%
```

```
tableX(Q1b,Q1bm)
```

Q1b/Q1bm	S	s	?	e	E	P	Total
1	2747	0	0	0	0	0	2747
2	0	8389	0	0	0	0	8389
3	0	0	5949	0	0	0	5949
4	0	0	0	9003	0	0	9003
5	0	0	0	0	5547	0	5547
Missing	0	0	0	0	0	1188	1188
Total	2747	8389	5949	9003	5547	1188	32823

```
ISSP2012jh1d.dat %>%
```

```
tableX(Q1c,Q1cm)
```

Q1c/Q1cm	S	s	?	e	E	P	Total
1	2838	0	0	0	0	0	2838
2	0	8263	0	0	0	0	8263
3	0	0	6000	0	0	0	6000
4	0	0	0	8706	0	0	8706
5	0	0	0	0	5960	0	5960
Missing	0	0	0	0	0	1056	1056
Total	2838	8263	6000	8706	5960	1056	32823

ISSP2012jh1d.dat %>%  
tableX(Q1d,Q1dm)

Q1d/Q1dm	S	s	?	e	E	P	Total
1	2818	0	0	0	0	0	2818
2	0	7672	0	0	0	0	7672
3	0	0	7403	0	0	0	7403
4	0	0	0	7863	0	0	7863
5	0	0	0	0	5016	0	5016
Missing	0	0	0	0	0	2051	2051
Total	2818	7672	7403	7863	5016	2051	32823

ISSP2012jh1d.dat %>%  
tableX(Q1e,Q1em)

Q1e/Q1em	S	s	?	e	E	P	Total
1	3357	0	0	0	0	0	3357
2	0	8342	0	0	0	0	8342
3	0	0	7841	0	0	0	7841
4	0	0	0	7267	0	0	7267
5	0	0	0	0	3462	0	3462
Missing	0	0	0	0	0	2554	2554
Total	3357	8342	7841	7267	3462	2554	32823

ISSP2012jh1d.dat %>%  
tableX(Q2a,Q2am)

Q2a/Q2am	S	s	?	e	E	P	Total
1	11305	0	0	0	0	0	11305
2	0	13464	0	0	0	0	13464
3	0	0	5039	0	0	0	5039
4	0	0	0	1929	0	0	1929

Q2a/Q2am	S	s	?	e	E	P	Total
5	0	0	0	0	403	0	403
Missing	0	0	0	0	0	683	683
Total	11305	13464	5039	1929	403	683	32823

```
ISSP2012jh1d.dat %>%
  tableX(Q2b,Q2bm)
```

Q2b/Q2bm	S	s	?	e	E	P	Total
1	2704	0	0	0	0	0	2704
2	0	5164	0	0	0	0	5164
3	0	0	6109	0	0	0	6109
4	0	0	0	9210	0	0	9210
5	0	0	0	0	8917	0	8917
Missing	0	0	0	0	0	719	719
Total	2704	5164	6109	9210	8917	719	32823

```
ISSP2012jh1d.dat %>%
  tableX(Q3a,Q3am)
```

Q3a/Q3am	W	w	H	P	Total
1	5373	0	0	0	5373
2	0	15655	0	0	15655
3	0	0	8367	0	8367
Missing	0	0	0	3428	3428
Total	5373	15655	8367	3428	32823

```
ISSP2012jh1d.dat %>%
  tableX(Q3b,Q3bm)
```

Q3b/Q3bm	W	w	H	P	Total
1	13722	0	0	0	13722
2	0	13817	0	0	13817
3	0	0	1762	0	1762
Missing	0	0	0	3522	3522
Total	13722	13817	1762	3522	32823

```
ISSP2012jh1d.dat %>% # tableX muotoilee taulukkoa!
  tableX(Q3am,Q3a)
```



Q3am/Q3a	1	2	3	Missing	Total
W	5373	0	0	0	5373
w	0	15655	0	0	15655
H	0	0	8367	0	8367
P	0	0	0	3428	3428
Total	5373	15655	8367	3428	32823

```
ISSP2012jh1d.dat$Q3a %>% levels()
```

```
## [1] "W" "w" "H"
```

```
ISSP2012jh1d.dat$Q3am %>% levels()
```

```
## [1] "W" "w" "H" "P"
```

## 2 Yksinkertainen korrespondenssianalyysi - kahden luokittelumuuttujan taulukko

### Vanhaa jäsennystä

Yksinkertainen esimerkki, yksi kysymys (V6) ja kuusi maata ristiintaulukoituna. Johdatteluna aiheeseen esitellään ca-käsitteet profiili, massa ja reunajakauma. Havainnollistetaan rivi- ja sarakeprofiilien vertailua vastaaviin keskiarvoprofiileihin.

Taulukoita tarkastella ensin rivien ja sitten sarakkeiden suhteen. Miten ne poikkeavat keskiarvostaan, miten toisistaan saman kategorian profiilista. Usein taulukoissa muuttujilla on selvästi eri rooli, kuten tässä. Koitan hahmottaa maiden (=aggregoituja yksilöitä) eroja ja yhtäläisyyksiä. Sarakkeiden vertailussa taas näemme, miten muuttujien profiilit poikkeavat keskiarvostaan. Monia riippuvuusia ja poikkeamia näyttäisi olevan. Klassinen ongelma, Pearson ja Fisher (ehkä turhaa tässä?).

Riippumattomuushypoteesi ja  $\chi^2$  - riippumattomuustesti (pieni huomautus - on monta tapaa testata taulukon riippuvuuksia). Riippumattomuushypoteesi ehdollisena todennäköisyytenä reunajakauman suhteen. **zxy** Tämä puuttuu kaavoista!

**zxy**

### Tarvitaanko käsitteellistä täsmentämistä, tai selkiinnyttämistä?

#### 1. Taulukon käsite

Erityisesti CA, jossa "ranskalaisella terminologialla" käsitellään yksilöiden tai havaintoyksiköiden pilveä ja muuttujien pilvelä (nominaaliasteikko). Taulukot

saadaan yksinkertaisen CA:n tapauksessa aggregoimalla “cloud of individuals”.  
#V MOOC, LeReoux

2. Kontingenssitaulu (kts. viite, jossa ohje “yhteys aina riviä pitkin”), frekvenssitaulu, ristiintaulukointi
  - dataa valitaan, aggregoidaan, ryhmitellään. Aktiivisia valintoja. Blasius emt. “data ei löydy kadulta”, ja vaikka siitä ei ole epäilystäkään ISSP-datan tapauksessa, niin siitäkin jatketaan eteenpäin.
3. Peruskäsitteiden yksinkertaisessa esityksessä tärkein lähde MG:n CAiP  
#V Siellä tästäkin on sananen: substanssiero usein on.
4. CA:ssa hämäävä juttu (Blasius, “vizualisation - verkkokirja”) rivien ja sarakkeiden **tekniinen** symmetria.

$\chi^2$  - etäisyys, yhteys hajontaan eli inertiaan ca-terminologiassa.

Dimensioiden vähentämisen idea (“the essence”), joka ei pienessä taulossa ole ihan ilmeinen. Toinen tavoite on visualisointi, yleensä kaksiulotteisena kuvana (karttana).

Yksinkertainen korrespondenssianalyysi on kahden luokitteluasteikon muuttujan riippuvuuksien geometrista analyysiä. Lähtökohta on kahden muuttujan ristiintaulukointi, alkuperäinen data voi olla muillakin asteikoilla mitattua. Menetelmän ydin on tarkastella molempien muuttujien – taulukon rivien ja sarakkeiden – riippuvuuksia kaksiulotteisena kuvana. Kuvaa kutsutaan myös kartaksi, ja tulkinnan ensimmäinen askel on kartan “koordinaatiston” tulkinta. Kaikki etäisyydet kuvassa ovat suhteellisia, vain rivi- ja sarakepisteiden etäisyydet kuvan origosta voidaan tulkita tarkasti. Koordinaatiston tulkinta aloitetaan “katsomalla mitä on oikealla ja vasemmalla, ja mitä on ylhäällä ja alhaalla” (viite LeRoux et.al, Bezecri-sitaatti). Vaikka pisteiden etäisyyksiä edes rivi- ja sarakepisteiden välillä ei voi tarkkaan tulkita (approksimaatioita), projektiossa kaukana toisistaan olevat pisteet ovat kaukana toisistaan myös alkuperäisessä “pistepilvessä”.

Akseleiden tulkinta “ääripäiden” kautta (“kontrasti”?). Huom “ääripää” ei välttämättä Likert-asteikolla tarkoita “äärimielipidettä”, vaan se voi tarkoittaa myös selvää tai varmaa mielipidettä.(3.10.18).

#### Vanha lista:

1. Ensimmäinen taulukko: profiilit, massat, keskiarvoprofiilit, khii2 - riippumattomuustesti ja etäisyysmitta
2. Hyvin tiivis esitys CA:n perusideasta, mutta ilman aivan simppeleitä kolmiulotteisia kuvia (niitä on jo)
3. Ensimmäinen symmetrinen kartta, perustulkinta (mitä kuvasta voidaan sanoa, mitä ei)
4. Lyhyt viittaus graafisen esityksen tulkintapulmiin, jotka eivät ole kovin

pahoja. CA-kartta kaksoiskuvana (ts. informaatio voidaan palauttaa, skaalaritulo)?

#### 5. Tulkinnan syventäminen - CA-käsitteiden tarkempi esittely

Haaste: käsitteet ja niiden suhteet ovat abstraktien matemaattisten rakenteiden tuloksia (barycentric, sentroidi), ja ne pitää jotenkin johdonmukaisesti palakerrallaan tuoda esimerkkien kautta tekstiin. Käsitteistä oma Rmd (ja Excel jos osoittautuu kätevämmäksi), kaavaliite Dispo-repossa ja myös Rmd-muodossa. **edit** Kaavaliitteessä pieniä eroja, ja tekstiä on LaTeX-versiossa enemmän.

#### Ensimmäinen symmetrinen kartta

Tulkinnat ja yksinkertaisimmat perussäännöt. Dimensiot ja kuinka paljon alkuuperäisen taulukon inertiaa saadaan esitettyä kartalla. Sitten asian ydin, akseleiden tulkinta (“mitä on oikealla ja vasemmalla”). Jos pisteet ovat alkuperäisessä “pilvessä” kaukana toisistaan, ne ovat sitä myös projektiossa. Kartta, mutta etäisyyksillä ei suoraa tulkintaa paitsi eteisyyksinllä origoon. Rivipisteiden suhteelliset etäisyydet, samoin sarakepisteidet. Mitä tarkoittavat prosentit akseleilla?

Varoitus virhetulkinnasta: ryhmien tunnistaminen rivi, jopa rivi- ja sarakepisteistä koostuvien ryhmien. **zxy** Ja silti tavallaan voi. Sarake- ja rivipisteiden etäisyyksille ei ole suoraa tulkintaa, mutta on “vetovoima” (attraktio) ja “työntövoima” (repulsio). Jos profilissa sarakemuuttujan osuus on suuri (siis suurempi kuin keskiarvopisteessä, suhteellinen ero), se “ajautuu” lähelle sarekepistettä. MG: “loose ends” - paperi, symmetrinen kuva eräs suurin sekaannuksen lähde. Tätä koitetaan selvittää myös MG:n JASA-artikkelissa.

**zxy** termi korrespondenssi: “neglected multivariate method” - paperissa käännetty näin englanniksi ransk. termi (Benzecri) rivien ja sarakkeiden “correspondence” eli yhteys/“riippuvuus”/vastaavuus tms.

**zxy** . Tarina: valitaan edellisessä luvussa esitetyn pohjalta osa muuttujista, perustellaan miksi työmarkkia-asenteen ovat kiinnostavia, valitaan esimerkkianalyysiin **yksi** muuttuja ja kuusi maata.

## 2.1 Äiti työssä

**zxy** Perustellaan aineiston valinnan vaiheet. Esimerkiksi otetaan yksi kysymys.

**zxy** Suhde data-lukuun, siellä pitäisi esitellä aineisto sisällöllisesti. Tässä vain valitan esimerkkiä varten yksi kysymys ja kuusi maata.

**zxy** Muuttujien nimeäminen vaikuttaa (a) muuttujien faktorointiin ja (b) kuviin ja taulukoihin.

Aineisto muuttujat V5-V9 ovat vastauksia ensimmäiseen kysymyspatteriin (Q1a-Q1e, arvot 1-5, täysin samaa mieltä - täysin eri mieltä).

(V6/Q1b) Alle kouluikäinen lapsi todennäköisesti kärsii, jos hänen äitinsä käy työssä. V6 muunnetaan uudeksi luokittelumuuttujaksi (R:ssä factor) Q1b. Tämä

ei vielä tee kuvista ahtaita kun sarakkeita ja rivejä on vähän. Pudotetaan tarvittaessa turha Q-kirjain pois. Alkuperäisessä muuttujassa metatieto säilyy varmemmin, ja tarkistuksia on helpompi tehdä.

Valitaan esimerkin data edellisessä luvussa luodusta R-datasta ISSP2012jh1c.data (df). Ihan yhtä hyvin voisi aina lukea suoraan alkuperäisestä spss-tiedostosta, mutta pidemmässä raportissa tämä on siistimpi tapa (23.3.2019).

```
# UUSI DATA 30.1.20
#
# LUETAAN DATA G1_1_data2.Rmd - tiedostossa, luodaan faktorimuuttujat
# G1_1_data_fct1.Rmd-tiedostossa -> ISSP2012jh1d.dat (df)
# 23 muuttujaa (9 substanssimuuttujaa, 8 taustamuuttujaa, 3 maa-muuttujaa, 3 metadatamuuttujaa)
# 25 maata.
# Poistettu 146 havaintoa, joilla SEX tai AGE puuttuu
# Johdattelevassa esimerkissä kuusi maata, kaksi taustamuuttujaa ja yksi kysymys (V6/Q1b)

# Kuusi maata

countries_esim1 <- c(56, 100, 208, 246, 276, 348) #BE,BG,DK,FI,DE,HU
ISSP2012esim3.dat <- filter(ISSP2012jh1d.dat, V4 %in% countries_esim1)
str(ISSP2012esim3.dat)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':   8542 obs. of  72 variables:
## $ V1      : 'haven_labelled' num  5900 5900 5900 5900 5900 5900 5900 5900 5900 5900 ...
## ..- attr(*, "label")= chr "GESIS Data Archive Study Number"
## ..- attr(*, "labels")= Named num 5900
## .. ..- attr(*, "names")= chr "GESIS Data Archive Study Number ZA5900"
## $ V2      : chr  "4.0.0 (2016-11-23)" "4.0.0 (2016-11-23)" "4.0.0 (2016-11-23)" "4.0.0 ..."
## ..- attr(*, "label")= chr "GESIS Archive Version"
## ..- attr(*, "format.spss")= chr "A25"
## ..- attr(*, "display_width")= int 26
## $ DOI      : chr  "doi:10.4232/1.12661" "doi:10.4232/1.12661" "doi:10.4232/1.12661" "doi:10.4232/1.12661" ...
## ..- attr(*, "label")= chr "Digital Object Identifier"
## ..- attr(*, "format.spss")= chr "A50"
## ..- attr(*, "display_width")= int 26
## $ V3      : 'haven_labelled' num  100 100 100 100 100 100 100 100 100 100 ...
## ..- attr(*, "label")= chr "Country/ Sample ISO 3166 Code (see V4 for codes for whole nation)"
## ..- attr(*, "labels")= Named num  32 36 40 100 124 152 156 158 191 203 ...
## .. ..- attr(*, "names")= chr  "AR-Argentina" "AU-Australia" "AT-Austria" "BG-Bulgaria" "BE-Belgium" "DE-Germany" "DK-Denmark" "FI-Finland" "FR-France" "GB-Great Britain" "HU-Hungary" "IS-Iceland" "IT-Italy" "JP-Japan" "KR-Korea" "LT-Lithuania" "LU-Luxembourg" "NL-Netherlands" "NO-Norway" "PL-Poland" "PT-Portugal" "SE-Sweden" "SI-Slovenia" "SK-Slovakia" "US-United States" "ZA-South Africa"
## $ V4      : 'haven_labelled' num  100 100 100 100 100 100 100 100 100 100 ...
## ..- attr(*, "label")= chr "Country ISO 3166 Code (see V3 for codes for the sample)"
## ..- attr(*, "labels")= Named num  32 36 40 56 100 124 152 156 158 191 ...
## .. ..- attr(*, "names")= chr  "AR-Argentina" "AU-Australia" "AT-Austria" "BE-Belgium" "BG-Bulgaria" "DE-Germany" "DK-Denmark" "FI-Finland" "FR-France" "GB-Great Britain" "HU-Hungary" "IS-Iceland" "IT-Italy" "JP-Japan" "KR-Korea" "LT-Lithuania" "LU-Luxembourg" "NL-Netherlands" "NO-Norway" "PL-Poland" "PT-Portugal" "SE-Sweden" "SI-Slovenia" "SK-Slovakia" "US-United States" "ZA-South Africa"
## $ C_ALPHAN: chr  "BG" "BG" "BG" "BG" ...
## ..- attr(*, "label")= chr "Country Prefix ISO 3166 Code - alphanumeric"
## ..- attr(*, "format.spss")= chr "A20"
```

```

##   ..- attr(*, "display_width")= int 22
## $ V5       : 'haven_labelled' num  3 3 2 1 3 2 2 4 4 5 ...
##   ..- attr(*, "label")= chr "Q1a Working mom: warm relationship with children as a not wo
##   ..- attr(*, "labels")= Named num  0 1 2 3 4 5 8 9
##   .. ..- attr(*, "names")= chr  "NAP: ES" "Strongly agree" "Agree" "Neither agree nor dis
## $ V6       : 'haven_labelled' num  3 2 3 4 3 3 4 3 2 3 ...
##   ..- attr(*, "label")= chr "Q1b Working mom: Preschool child is likely to suffer"
##   ..- attr(*, "labels")= Named num  0 1 2 3 4 5 8 9
##   .. ..- attr(*, "names")= chr  "NAP: ES" "Strongly agree" "Agree" "Neither agree nor dis
## $ V7       : 'haven_labelled' num  3 2 2 3 3 3 2 2 2 3 ...
##   ..- attr(*, "label")= chr "Q1c Working woman: Family life suffers when woman has full-t
##   ..- attr(*, "labels")= Named num  0 1 2 3 4 5 8 9
##   .. ..- attr(*, "names")= chr  "NAP: ES" "Strongly agree" "Agree" "Neither agree nor dis
## $ V8       : 'haven_labelled' num  2 1 2 4 3 3 2 3 2 3 ...
##   ..- attr(*, "label")= chr "Q1d Working woman: What women really want is home and kids"
##   ..- attr(*, "labels")= Named num  0 1 2 3 4 5 8 9
##   .. ..- attr(*, "names")= chr  "NAP: ES" "Strongly agree" "Agree" "Neither agree nor dis
## $ V9       : 'haven_labelled' num  3 3 2 4 4 4 2 3 2 1 ...
##   ..- attr(*, "label")= chr "Q1e Working woman: Being housewife is as fulfilling as worki
##   ..- attr(*, "labels")= Named num  0 1 2 3 4 5 8 9
##   .. ..- attr(*, "names")= chr  "NAP: ES" "Strongly agree" "Agree" "Neither agree nor dis
## $ V10      : 'haven_labelled' num  2 1 1 1 1 1 1 3 2 1 ...
##   ..- attr(*, "label")= chr "Q2a Both should contribute to household income"
##   ..- attr(*, "labels")= Named num  0 1 2 3 4 5 8 9
##   .. ..- attr(*, "names")= chr  "NAP: ES" "Strongly agree" "Agree" "Neither agree nor dis
## $ V11      : 'haven_labelled' num  4 2 1 4 3 4 1 5 1 5 ...
##   ..- attr(*, "label")= chr "Q2b Men's job earn money, women's job look after home"
##   ..- attr(*, "labels")= Named num  0 1 2 3 4 5 8 9
##   .. ..- attr(*, "names")= chr  "NAP: ES" "Strongly agree" "Agree" "Neither agree nor dis
## $ V12      : 'haven_labelled' num  2 2 3 2 3 3 3 2 3 3 ...
##   ..- attr(*, "label")= chr "Q3a Should women work: Child under school age"
##   ..- attr(*, "labels")= Named num  1 2 3 6 8 9
##   .. ..- attr(*, "names")= chr  "Work full-time" "Work part-time" "Stay at home" "TW: won
## $ V13      : 'haven_labelled' num  2 2 1 1 2 2 2 2 2 2 ...
##   ..- attr(*, "label")= chr "Q3b Should women work: Youngest kid at school"
##   ..- attr(*, "labels")= Named num  1 2 3 6 8 9
##   .. ..- attr(*, "names")= chr  "Work full-time" "Work part-time" "Stay at home" "TW: won
## $ SEX      : 'haven_labelled' num  2 2 1 2 2 2 1 1 2 1 ...
##   ..- attr(*, "label")= chr "Sex of Respondent"
##   ..- attr(*, "labels")= Named num  1 2 9
##   .. ..- attr(*, "names")= chr  "Male" "Female" "No answer"
## $ AGE      : 'haven_labelled' num  64 43 63 31 52 46 51 40 57 64 ...
##   ..- attr(*, "label")= chr "Age of respondent"
##   ..- attr(*, "labels")= Named num  15 16 17 18 102 999
##   .. ..- attr(*, "names")= chr  "15 years" "16 years" "17 years" "18 years" ...
## $ DEGREE   : 'haven_labelled' num  2 3 2 4 3 3 3 4 4 6 ...

```

```

##   .- attr(*, "label")= chr "Highest completed degree of education: Categories for intern
##   .- attr(*, "labels")= Named num  0 1 2 3 4 5 6 9
##   .. .- attr(*, "names")= chr  "No formal education" "Primary school (elementary school)
## $ MAINSTAT: 'haven_labelled' num  6 1 6 7 1 2 2 5 5 6 ...
##   .- attr(*, "label")= chr "Main status"
##   .- attr(*, "labels")= Named num  1 2 3 4 5 6 7 8 9 99
##   .. .- attr(*, "names")= chr  "In paid work" "Unemployed and looking for a job, HR: in
## $ TOPBOT  : 'haven_labelled' num  4 5 3 5 5 5 5 5 7 ...
##   .- attr(*, "label")= chr "Top-Bottom self-placement"
##   .- attr(*, "labels")= Named num  0 1 2 3 4 5 6 7 8 9 ...
##   .. .- attr(*, "names")= chr  "Not available: GB,US" "Lowest, Bottom, 01" "02" "03" ...
## $ HHCHILDR: 'haven_labelled' num  0 1 0 1 0 0 0 1 0 0 ...
##   .- attr(*, "label")= chr "How many children in household: children between [school age
##   .- attr(*, "labels")= Named num  0 1 2 21 96 97 99
##   .. .- attr(*, "names")= chr  "No children" "One child" "2 children" "21 children" ...
## $ MARITAL : 'haven_labelled' num  4 1 1 1 1 2 1 1 1 1 ...
##   .- attr(*, "label")= chr "Legal partnership status"
##   .- attr(*, "labels")= Named num  1 2 3 4 5 6 7 8 9
##   .. .- attr(*, "names")= chr  "Married" "Civil partnership" "Separated from spouse/ ci
## $ URBURURAL: 'haven_labelled' num  3 3 3 3 3 3 3 3 3 3 ...
##   .- attr(*, "label")= chr "Place of living: urban - rural"
##   .- attr(*, "labels")= Named num  1 2 3 4 5 7 9
##   .. .- attr(*, "names")= chr  "A big city" "The suburbs or outskirts of a big city" "A
## $ maa      : Factor w/ 25 levels "AU","AT","BG",...: 3 3 3 3 3 3 3 3 3 3 ...
## $ sp1      : Factor w/ 3 levels "Male","Female",...: 2 2 1 2 2 2 1 1 2 1 ...
## $ sp       : Factor w/ 2 levels "m","f": 2 2 1 2 2 2 1 1 2 1 ...
## $ ika      : 'haven_labelled' num  64 43 63 31 52 46 51 40 57 64 ...
##   .- attr(*, "label")= chr "Age of respondent"
##   .- attr(*, "labels")= Named num  15 16 17 18 102 999
##   .. .- attr(*, "names")= chr  "15 years" "16 years" "17 years" "18 years" ...
## $ Q1a1     : Factor w/ 8 levels "NAP: ES","Strongly agree",...: 4 4 3 2 4 3 3 5 5 6 ...
## $ Q1b1     : Factor w/ 8 levels "NAP: ES","Strongly agree",...: 4 3 4 5 4 4 5 4 3 4 ...
## $ Q1c1     : Factor w/ 8 levels "NAP: ES","Strongly agree",...: 4 3 3 4 4 4 3 3 3 4 ...
## $ Q1d1     : Factor w/ 8 levels "NAP: ES","Strongly agree",...: 3 2 3 5 4 4 3 4 3 4 ...
## $ Q1e1     : Factor w/ 8 levels "NAP: ES","Strongly agree",...: 4 4 3 5 5 5 3 4 3 2 ...
## $ Q2a1     : Factor w/ 8 levels "NAP: ES","Strongly agree",...: 3 2 2 2 2 2 2 4 3 2 ...
## $ Q2b1     : Factor w/ 8 levels "NAP: ES","Strongly agree",...: 5 3 2 5 4 5 2 6 2 6 ...
## $ Q3a1     : Factor w/ 6 levels "Work full-time",...: 2 2 3 2 3 3 2 3 3 ...
## $ Q3b1     : Factor w/ 6 levels "Work full-time",...: 2 2 1 1 2 2 2 2 2 2 ...
## $ edu1     : Factor w/ 8 levels "No formal education",...: 3 4 3 5 4 4 4 5 5 7 ...
## $ msta1    : Factor w/ 10 levels "In paid work",...: 6 1 6 7 1 2 2 5 5 6 ...
## $ sosta1   : Factor w/ 14 levels "Not available: GB,US",...: 5 6 4 6 6 6 6 6 6 8 ...
## $ nchild1  : Factor w/ 14 levels "No children",...: 1 2 1 2 1 1 1 2 1 1 ...
## $ lifsta1  : Factor w/ 9 levels "Married","Civil partnership",...: 4 1 1 1 1 2 1 1 1 1 ...
## $ urbru1   : Factor w/ 7 levels "A big city","The suburbs or outskirts of a big city",...
## $ Q1a      : Factor w/ 5 levels "S","s","?","e",...: 3 3 2 1 3 2 2 4 4 5 ...

```

```
## $ Q1b      : Factor w/ 5 levels "S","s","?", "e",...: 3 2 3 4 3 3 4 3 2 3 ...
## $ Q1c      : Factor w/ 5 levels "S","s","?", "e",...: 3 2 2 3 3 3 2 2 2 3 ...
## $ Q1d      : Factor w/ 5 levels "S","s","?", "e",...: 2 1 2 4 3 3 2 3 2 3 ...
## $ Q1e      : Factor w/ 5 levels "S","s","?", "e",...: 3 3 2 4 4 4 2 3 2 1 ...
## $ Q2a      : Factor w/ 5 levels "S","s","?", "e",...: 2 1 1 1 1 1 1 3 2 1 ...
## $ Q2b      : Factor w/ 5 levels "S","s","?", "e",...: 4 2 1 4 3 4 1 5 1 5 ...
## $ Q3a      : Factor w/ 3 levels "W","w","H": 2 2 3 2 3 3 3 2 3 3 ...
## $ Q3b      : Factor w/ 3 levels "W","w","H": 2 2 1 1 2 2 2 2 2 2 ...
## $ edu      : Factor w/ 7 levels "No formal education",...: 3 4 3 5 4 4 4 5 5 7 ...
## $ msta     : Factor w/ 9 levels "In paid work",...: 6 1 6 7 1 2 2 5 5 6 ...
## $ sosta    : Factor w/ 10 levels "Lowest, Bottom, 01",...: 4 5 3 5 5 5 5 5 5 7 ...
## $ nchild   : Factor w/ 11 levels "No children",...: 1 2 1 2 1 1 1 2 1 1 ...
## $ lifsta   : Factor w/ 6 levels "Married","Civil partnership",...: 4 1 1 1 1 2 1 1 1 1 ...
## $ urbru    : Factor w/ 5 levels "A big city","The suburbs or outskirts of a big city",...: 3 3 2 1 3 2 2 4 4 5 ...
## $ Q1am     : Factor w/ 6 levels "S","s","?", "e",...: 3 3 2 1 3 2 2 4 4 5 ...
## $ Q1bm     : Factor w/ 6 levels "S","s","?", "e",...: 3 2 3 4 3 3 4 3 2 3 ...
## $ Q1cm     : Factor w/ 6 levels "S","s","?", "e",...: 3 2 2 3 3 3 2 2 2 3 ...
## $ Q1dm     : Factor w/ 6 levels "S","s","?", "e",...: 2 1 2 4 3 3 2 3 2 3 ...
## $ Q1em     : Factor w/ 6 levels "S","s","?", "e",...: 3 3 2 4 4 4 2 3 2 1 ...
## $ Q2am     : Factor w/ 6 levels "S","s","?", "e",...: 2 1 1 1 1 1 1 3 2 1 ...
## $ Q2bm     : Factor w/ 6 levels "S","s","?", "e",...: 4 2 1 4 3 4 1 5 1 5 ...
## $ Q3am     : Factor w/ 4 levels "W","w","H","P": 2 2 3 2 3 3 3 2 3 3 ...
## $ Q3bm     : Factor w/ 4 levels "W","w","H","P": 2 2 1 1 2 2 2 2 2 2 ...
## $ edum     : Factor w/ 8 levels "No formal education",...: 3 4 3 5 4 4 4 5 5 7 ...
## $ mstam    : Factor w/ 10 levels "In paid work",...: 6 1 6 7 1 2 2 5 5 6 ...
## $ sostam   : Factor w/ 11 levels "Lowest, Bottom, 01",...: 4 5 3 5 5 5 5 5 5 7 ...
## $ nchildm  : Factor w/ 12 levels "No children",...: 1 2 1 2 1 1 1 2 1 1 ...
## $ lifstam  : Factor w/ 7 levels "Married","Civil partnership",...: 4 1 1 1 1 2 1 1 1 1 ...
## $ urbrum   : Factor w/ 6 levels "A big city","The suburbs or outskirts of a big city",...: 3 3 2 1 3 2 2 4 4 5 ...

#neljä maamuuttujaa, kysymys Q1b, ikä ja sukupuoli

vars_esim1 <- c("C_ALPHAN", "maa", "V3", "V4", "Q1b", "sp", "ika")
ISSP2012esim2.dat <- select(ISSP2012esim3.dat, vars_esim1)

str(ISSP2012esim2.dat) # 8542 obs. of 7 variables

## Classes 'tbl_df', 'tbl' and 'data.frame':    8542 obs. of  7 variables:
## $ C_ALPHAN: chr  "BG" "BG" "BG" "BG" ...
##   ..- attr(*, "label")= chr "Country Prefix ISO 3166 Code - alphanumeric"
##   ..- attr(*, "format.spss")= chr "A20"
##   ..- attr(*, "display_width")= int 22
## $ maa      : Factor w/ 25 levels "AU","AT","BG",...: 3 3 3 3 3 3 3 3 3 3 ...
## $ V3       : 'haven_labelled' num  100 100 100 100 100 100 100 100 100 100 ...
##   ..- attr(*, "label")= chr "Country/ Sample ISO 3166 Code (see V4 for codes for whole na
##   ..- attr(*, "labels")= Named num  32 36 40 100 124 152 156 158 191 203 ...
```

```

## .. ..- attr(*, "names")= chr "AR-Argentina" "AU-Australia" "AT-Austria" "BG-Bulgaria"
## $ V4 : 'haven_labelled' num 100 100 100 100 100 100 100 100 100 100 ...
## ..- attr(*, "label")= chr "Country ISO 3166 Code (see V3 for codes for the sample)"
## ..- attr(*, "labels")= Named num 32 36 40 56 100 124 152 156 158 191 ...
## .. ..- attr(*, "names")= chr "AR-Argentina" "AU-Australia" "AT-Austria" "BE-Belgium"
## $ Q1b : Factor w/ 5 levels "S","s","?", "e",...: 3 2 3 4 3 3 4 3 2 3 ...
## $ sp : Factor w/ 2 levels "m","f": 2 2 1 2 2 2 1 1 2 1 ...
## $ ika : 'haven_labelled' num 64 43 63 31 52 46 51 40 57 64 ...
## ..- attr(*, "label")= chr "Age of respondent"
## ..- attr(*, "labels")= Named num 15 16 17 18 102 999
## .. ..- attr(*, "names")= chr "15 years" "16 years" "17 years" "18 years" ...

# C_ALPHAN: chr, maa: Factor w/ 25

# Poistetaan havainnot, joilla Q1b - muuttujassa puuttuva tieto 'NA'

ISSP2012esim1.dat <- filter(ISSP2012esim2.dat, !is.na(Q1b))

str(ISSP2012esim1.dat) # 8143 obs. of 6 variable

## Classes 'tbl_df', 'tbl' and 'data.frame': 8143 obs. of 7 variables:
## $ C_ALPHAN: chr "BG" "BG" "BG" "BG" ...
## ..- attr(*, "label")= chr "Country Prefix ISO 3166 Code - alphanumeric"
## ..- attr(*, "format.spss")= chr "A20"
## ..- attr(*, "display_width")= int 22
## $ maa : Factor w/ 25 levels "AU","AT","BG",...: 3 3 3 3 3 3 3 3 3 3 ...
## $ V3 : 'haven_labelled' num 100 100 100 100 100 100 100 100 100 100 ...
## ..- attr(*, "label")= chr "Country/ Sample ISO 3166 Code (see V4 for codes for whole na"
## ..- attr(*, "labels")= Named num 32 36 40 100 124 152 156 158 191 203 ...
## .. ..- attr(*, "names")= chr "AR-Argentina" "AU-Australia" "AT-Austria" "BG-Bulgaria"
## $ V4 : 'haven_labelled' num 100 100 100 100 100 100 100 100 100 100 ...
## ..- attr(*, "label")= chr "Country ISO 3166 Code (see V3 for codes for the sample)"
## ..- attr(*, "labels")= Named num 32 36 40 56 100 124 152 156 158 191 ...
## .. ..- attr(*, "names")= chr "AR-Argentina" "AU-Australia" "AT-Austria" "BE-Belgium"
## $ Q1b : Factor w/ 5 levels "S","s","?", "e",...: 3 2 3 4 3 3 4 3 2 3 ...
## $ sp : Factor w/ 2 levels "m","f": 2 2 1 2 2 2 1 1 2 1 ...
## $ ika : 'haven_labelled' num 64 43 63 31 52 46 51 40 57 64 ...
## ..- attr(*, "label")= chr "Age of respondent"
## ..- attr(*, "labels")= Named num 15 16 17 18 102 999
## .. ..- attr(*, "names")= chr "15 years" "16 years" "17 years" "18 years" ...

# EDELLINEN DATA - havaintojen määrät samat kuin uudella datalla (31.1.20)
#
# 8557 obs. ennen kuin sexagemissing poistettiin, nyt 8542, 8557-8542 = 15
#
# Poistetaan havainnot joissa puuttuva tieto muuttujassa V6 (Q1b) n = 399
# 8542-399 = 8143

```



```
# Tyhjät "faktorilabelit" on poistettava
```

```
ISSP2012esim1.dat <- ISSP2012esim1.dat %>%  
  mutate(maa = fct_drop(maa))
```

```
ISSP2012esim1.dat %>% tableX(maa, Q1b, type = "count")
```

maa/Q1b	S	s	?	e	E	Total
BG	118	395	205	190	13	921
DK	70	238	152	232	696	1388
FI	47	188	149	423	303	1110
HU	219	288	225	190	75	997
BE	191	451	438	552	381	2013
DE	165	375	198	538	438	1714
Total	810	1935	1367	2125	1906	8143

```
fct_count(ISSP2012esim1.dat$Q1b)
```

f	n
S	810
s	1935
?	1367
e	2125
E	1906

```
# fct_count(ISSP2012esim1.dat$sp)
```

```
# fct_unique(ISSP2012esim1.dat$maa)
```

```
# fct_count(ISSP2012esim1.dat$maa)
```

```
ISSP2012esim1.dat %>% tableX(maa, C_ALPHAN, type = "count")
```

maa/C_ALPHAN	BE	BG	DE	DK	FI	HU	Total
BG	0	921	0	0	0	0	921
DK	0	0	0	1388	0	0	1388
FI	0	0	0	0	1110	0	1110
HU	0	0	0	0	0	997	997
BE	2013	0	0	0	0	0	2013
DE	0	0	1714	0	0	0	1714
Total	2013	921	1714	1388	1110	997	8143

```

# Muutetaan muuttujan "maa" arvojen (levels) järjestys samaksi kuin alkuperäisen
# muuttujan C_ALPHAN. Helpomi verrata aikaisempiin tuloksiin.

# "väärässä järjestyksessä" oleva maa talteen, olisiko järjestys jossa ovat
# SPSS-datassa?

ISSP2012esim1.dat <- ISSP2012esim1.dat %>%
  mutate(maa2 = maa)
# maa samaan järjestykseen kuin C_ALPHAN - olisiko aakkosjärjestys?

ISSP2012esim1.dat <- ISSP2012esim1.dat %>%
  mutate(maa =
    fct_relevel(maa,
      "BE",
      "BG",
      "DE",
      "DK",
      "FI",
      "HU"))

# Tarkistus
ISSP2012esim1.dat %>% tableX(maa2,maa, type = "count")

```

maa2/maa	BE	BG	DE	DK	FI	HU	Total
BG	0	921	0	0	0	0	921
DK	0	0	0	1388	0	0	1388
FI	0	0	0	0	1110	0	1110
HU	0	0	0	0	0	997	997
BE	2013	0	0	0	0	0	2013
DE	0	0	1714	0	0	0	1714
Total	2013	921	1714	1388	1110	997	8143

```
ISSP2012esim1.dat %>% tableX(maa,C_ALPHAN, type = "count")
```

maa/C_ALPHAN	BE	BG	DE	DK	FI	HU	Total
BE	2013	0	0	0	0	0	2013
BG	0	921	0	0	0	0	921
DE	0	0	1714	0	0	0	1714
DK	0	0	0	1388	0	0	1388
FI	0	0	0	0	1110	0	1110
HU	0	0	0	0	0	997	997
Total	2013	921	1714	1388	1110	997	8143

**zxy** taulukot erotettava omiksi koodilohkoiksi bookdowniin.

```
# Taulukoita (31.1.2020)
```

```
# toinen maa-muuttuja, jossa Saksan ja Belgian jako
```

```
# V3
```

```
# 5601 BE-FLA-Belgium/ Flanders
```

```
# 5602 BE-WAL-Belgium/ Wallonia
```

```
# 5603 BE-BRU-Belgium/ Brussels
```

```
# 27601 DE-W-Germany-West
```

```
# 27602 DE-E-Germany-East
```

```
#
```

```
# Tarkastuksia
```

```
#
```

```
ISSP2012esim1.dat %>% tableX(maa,Q1b,type = "count")
```

maa/Q1b	S	s	?	e	E	Total
BE	191	451	438	552	381	2013
BG	118	395	205	190	13	921
DE	165	375	198	538	438	1714
DK	70	238	152	232	696	1388
FI	47	188	149	423	303	1110
HU	219	288	225	190	75	997
Total	810	1935	1367	2125	1906	8143

```
ISSP2012esim1.dat %>% tableX(maa,ika,type = "count")
```

maa/ika	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
BE	0	0	0	16	26	25	24	31	30	27	29	24	34	27	28	33	32	4
BG	0	0	0	13	10	9	9	7	11	7	11	13	9	9	10	14	9	1
DE	0	0	0	19	29	21	27	24	33	25	27	20	24	21	25	21	26	2
DK	0	0	0	24	18	18	31	35	38	27	16	15	24	19	20	14	29	2
FI	8	20	13	18	12	15	13	7	14	17	15	16	14	13	29	13	18	1
HU	0	0	0	9	12	10	8	19	19	11	15	10	14	13	19	16	18	1
Total	8	20	13	99	107	98	112	123	145	114	113	98	119	102	131	111	132	1

```
ISSP2012esim1.dat %>% tableX(maa,sp ,type= "count")
```

maa/sp	m	f	Total
BE	955	1058	2013
BG	375	546	921
DE	834	880	1714

maa/sp	m	f	Total
DK	686	702	1388
FI	476	634	1110
HU	473	524	997
Total	3799	4344	8143

```
# Riviprofiilit
```

```
# ISSP2012esim1.dat %>% tableX(maa,ika,type = "row_perc")
ISSP2012esim1.dat %>% tableX(maa,sp ,type= "row_perc")
```

maa/sp	m	f	Total
BE	47.44	52.56	100.00
BG	40.72	59.28	100.00
DE	48.66	51.34	100.00
DK	49.42	50.58	100.00
FI	42.88	57.12	100.00
HU	47.44	52.56	100.00
All	46.65	53.35	100.00

```
# Kysymyksen Q1b vastaukset
```

```
# ISSP2012esim1.dat %>% tableX(maa,Q1b,type = "row_perc")
# ISSP2012esim1.dat %>% tableX(maa,Q1b,type = "cell_perc")
```

```
# str(ISSP2012esim1.dat) # 8143 obs. of 7 variable,
# sama kuin vanhassa Galku-koodissa.
```

**Taulukot ja kuvat omina koodilohkoina**

**Frekvenssitaulukko**

```
taulu2 <- ISSP2012esim1.dat %>% tableX(maa, Q1b, type = "count")
knitr::kable(taulu2,digits = 2, booktabs = TRUE,
             caption = "Kysymyksen Q1b vastaukset maittain")
```

Taulukko 84: Kysymyksen Q1b vastaukset maittain

	S	s	?	e	E	Total
BE	191	451	438	552	381	2013
BG	118	395	205	190	13	921
DE	165	375	198	538	438	1714
DK	70	238	152	232	696	1388
FI	47	188	149	423	303	1110

	S	s	?	e	E	Total
HU	219	288	225	190	75	997
Total	810	1935	1367	2125	1906	8143

### Riviprosentit

```
taulu3 <- ISSP2012esim1.dat %>% tableX(maa,Q1b,type = "row_perc")
knitr::kable(taulu3,digits = 2, booktabs = TRUE,
              caption = "Kysymyksen Q1b vastaukset, riviprosentit")
```

Taulukko 85: Kysymyksen Q1b vastaukset, riviprosentit

	S	s	?	e	E	Total
BE	9.49	22.40	21.76	27.42	18.93	100.00
BG	12.81	42.89	22.26	20.63	1.41	100.00
DE	9.63	21.88	11.55	31.39	25.55	100.00
DK	5.04	17.15	10.95	16.71	50.14	100.00
FI	4.23	16.94	13.42	38.11	27.30	100.00
HU	21.97	28.89	22.57	19.06	7.52	100.00
All	9.95	23.76	16.79	26.10	23.41	100.00

### Sarakeprosentit

```
taulu4 <- ISSP2012esim1.dat %>% tableX(maa,Q1b,type = "col_perc")
knitr::kable(taulu4,digits = 2, booktabs = TRUE,
              caption = "Kysymyksen Q1b vastaukset, sarakeprosentit")
```

Taulukko 86: Kysymyksen Q1b vastaukset, sarakeprosentit

	S	s	?	e	E	All
BE	23.58	23.31	32.04	25.98	19.99	24.72
BG	14.57	20.41	15.00	8.94	0.68	11.31
DE	20.37	19.38	14.48	25.32	22.98	21.05
DK	8.64	12.30	11.12	10.92	36.52	17.05
FI	5.80	9.72	10.90	19.91	15.90	13.63
HU	27.04	14.88	16.46	8.94	3.93	12.24
Total	100.00	100.00	100.00	100.00	100.00	100.00

Taulukoissa on kuuden maan vastausten jakauma kysymykseen “Alle kouluikäinen lapsi todennäköisesti kärsii, jos hänen äitinsä käy työssä”. Taulukko on pieni,

mutta havaintoja 8143. Alemman suhteellisten frekvenssien taulukon rivejä voi verrata toisiinsa ja alimpaan ("Total") keskimääräiseen riviin, sarakemuuttujien eli vastausvaihtoehtojen reunajakaumaan. Vastavasti sarakkeita voi verrata rivi-muuttujien reunajakaumasarakkeeseen ("Total2"). Eniten vastajia on Belgiasta (25 %) ja Saksasta (21 %), vähiten Unkarista (12 %).

**EDIT:** Pienenkin taulukon pyörittely johdattelee hyvin, mihin korrespondenssianalyysiä tarvitaan. Näistähän riippuvuuden rakenteet näkee ilmankin, jos on tarpeeksi nokkela. Muiden pitää käyttää CA:ta.

*# CA tässä, jotta saadaan rivi- ja sarakeprofiilikuvat*

```
simpleCA1 <- ca(~maa + Q1b,ISSP2012esim1.dat)
```

*# Maiden järjestys kääntää kuvan (1.2.20)*

```
simpleCA2 <- ca(~maa2 + Q1b,ISSP2012esim1.dat)
```

**TODO 2.2.20** Onko tämä kuva tallennettava kuvatiedostoksi, vai onnistuuko sen tuottaminen Bookdownissa. Ei taida onnistua? (4.9.18)

Sarakeprofiilit, oikea järjestys maa-muuttujan tasoilla.

*#mutkikas kuvan piirto - sarakeprofiilit vertailussa*

*#ggplot vaatii df-rakenteen ja 'long data' - muotoon*

*##https://stackoverflow.com/questions/9563368/create-stacked-barplot-where-each-stack-is-sc*  
*#*

*# käytetään ca - tuloksia*

```
apu1 <- (simpleCA1$N)
```

```
colnames(apu1) <- c("S", "s", "?", "e", "E")
```

```
rownames(apu1) <- c("BE", "BG", "DE", "DK", "FI", "HU")
```

```
apu1_df <- as.data.frame(apu1)
```

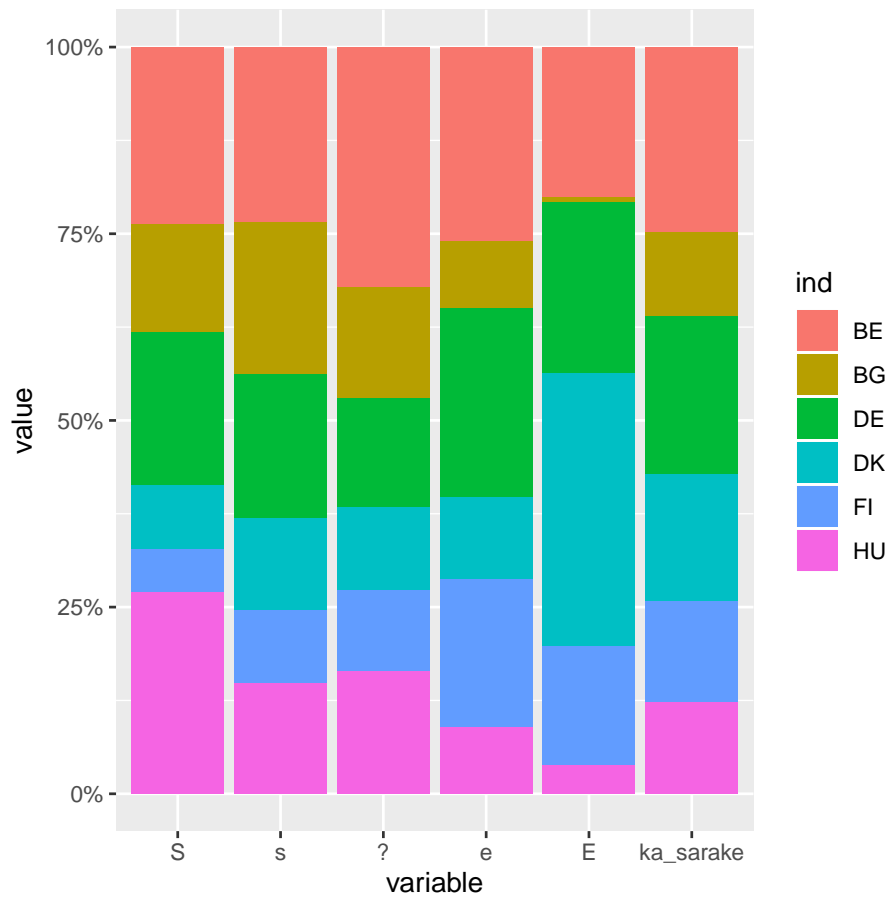
*#lasketan rivien reunajakauma*

```
apu1_df$ka_sarake <- rowSums(apu1_df)
```

*#muokataan 'long data' - muotoon*

```
apu1b_df <- melt(cbind(apu1_df, ind = rownames(apu1_df)), id.vars = c('ind'))
```

```
ggplot(apu1b_df, aes(x = variable, y = value, fill = ind)) +  
  geom_bar(position = "fill", stat = "identity") +  
  scale_y_continuous(labels = percent_format())
```



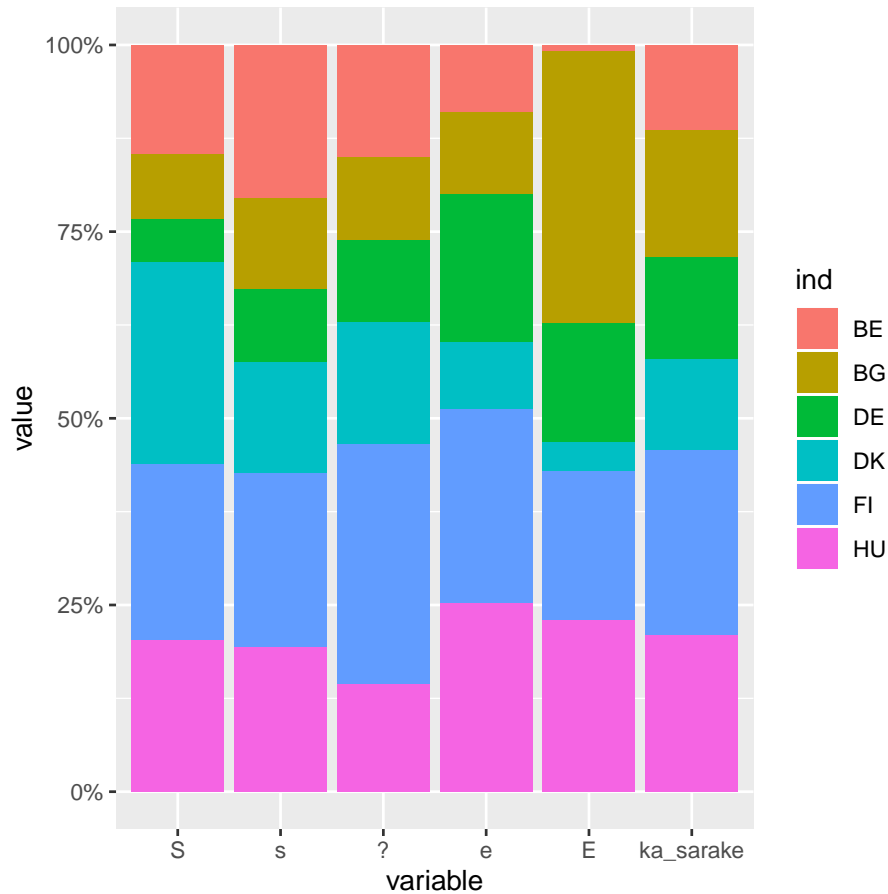
```
#apu1b_df
```

Testaus: maa2, eri järjestys kuin C\_ALPHAN (joka oli käytössä vanhemmissa Galku-versioissa)

```
#mutkikas kuvan piirto - sarakeprofiilit vertailussa
#ggplot vaatii df-rakenteen ja 'long data' - muotoon
##https://stackoverflow.com/questions/9563368/create-stacked-barplot-where-each-stack-is-sc
#
# käytetään ca - tuloksia
apu1test <- (simpleCA2$N)
colnames(apu1test) <- c("S", "s", "?", "e", "E")
rownames(apu1test) <- c("BE", "BG", "DE", "DK", "FI", "HU")
apu1_dftest <- as.data.frame(apu1test)
#lasketan rivien reunajakauma
apu1_dftest$ka_sarake <- rowSums(apu1_dftest)
#muokataan 'long data' - muotoon
```

```
apu1b_dftest <- melt(cbind(apu1_dftest, ind = rownames(apu1_dftest)), id.vars = c('ind'))

ggplot(apu1b_dftest, aes(x = variable, y = value, fill = ind)) +
  geom_bar(position = "fill", stat = "identity") +
  scale_y_continuous(labels = percent_format())
```



**TODO 2.2.20** Massat saa mukaan vaikka viittaamalla frekvenssitauluun (4.9.2018)

Riviprofiilikuva toimii, mutta vaatii vielä viilausta (18.9.2018).

```
# riviprofiilit ja keskiarvorivi - 18.9.2018
apu2_df <- as.data.frame(apu1)
apu2_df <- rbind(apu2_df, ka_rivi = colSums(apu2_df))

#apu2_df
#str(apu2_df)
```

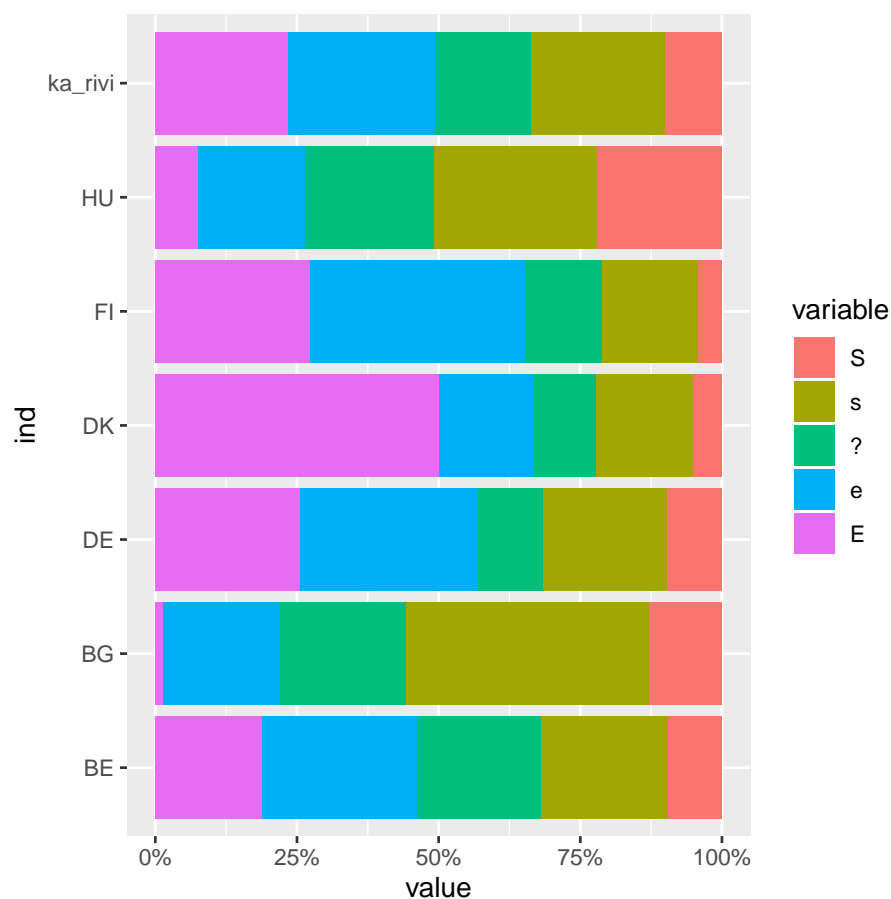


```

## typeof(apu2_df) # what is it?
## class(apu2_df) # what is it? (sorry)
## storage.mode(apu2_df) # what is it? (very sorry)
## length(apu2_df) # how long is it? What about two dimensional
## objects?
# attributes(apu2_df)

# temp1 <- cbind(apu2_df, ind = rownames(apu2_df))
# temp1
##muokataan 'long data' - muotoon
apu2b_df <- melt(cbind(apu2_df, ind = rownames(apu2_df)), id.vars = c('ind'))
#apu2b_df
#
#
#ggplot(apu2b_df, aes(x = value, y = ind, fill = variable)) +
#  geom_bar(position = "fill", stat = "identity") +
#  #coord_flip() +
#  scale_x_continuous(labels = percent_format())
#versio2 # perkele, tämä toimii! 18.9.2018
ggplot(apu2b_df, aes(x = ind, y = value, fill = variable)) +
  geom_bar(position = "fill", stat = "identity") +
  coord_flip() +
  scale_y_continuous(labels = percent_format())

```



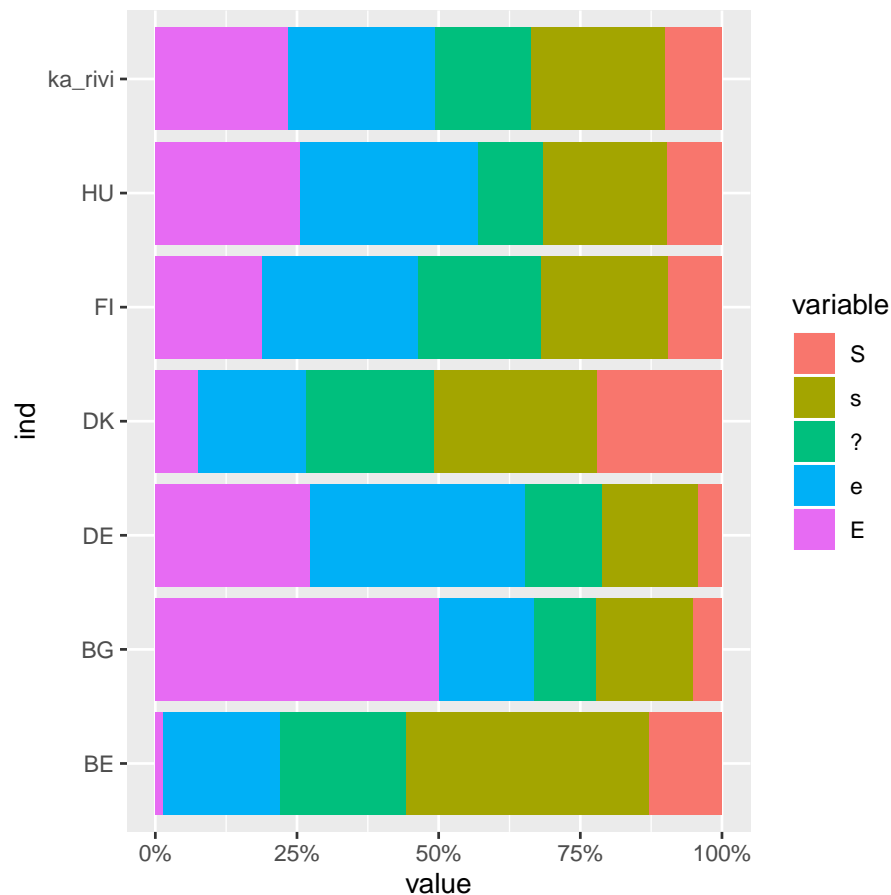
Ja sama testaus kuin sarakeprofilikuvilla

```
# riviprofiilit ja keskiarvorivi - 18.9.2018
apu2_dftest <- as.data.frame(apu1test)
apu2_dftest <- rbind(apu2_dftest, ka_rivi = colSums(apu2_dftest))

#apu2_df
#str(apu2_df)
## typeof(apu2_df) # what is it?
## class(apu2_df) # what is it? (sorry)
## storage.mode(apu2_df) # what is it? (very sorry)
## length(apu2_df) # how long is it? What about two dimensional
## objects?
# attributes(apu2_df)

# temp1 <- cbind(apu2_df, ind = rownames(apu2_df))
# temp1
```

```
##muokataan 'long data' - muotoon
apu2b_dftest <- melt(cbind(apu2_df, ind = rownames(apu2_df)), id.vars = c('ind'))
#apu2b_df
#
#
#ggplot(apu2b_df, aes(x = value, y = ind, fill = variable)) +
#  geom_bar(position = "fill", stat = "identity") +
#  #coord_flip() +
#  scale_x_continuous(labels = percent_format())
#versio2 # perkele, tämä toimii! 18.9.2018
ggplot(apu2b_df, aes(x = ind, y = value, fill = variable)) +
  geom_bar(position = "fill", stat = "identity") +
  coord_flip() +
  scale_y_continuous(labels = percent_format())
```



Graafinen analyysi ja R

Käytännön neuvoja data-analyysiin, kuulunee tekstiin, vai meneekö “ohjelmistoympäristö” -liitteeseen? Tärkeä juttu!

Kuvasuhteen saa oikeaksi, kun avaa g-ikkunan (X11()) ja sitten plot. Voi tallentaa pdf-muodossa grafiikkaikkunasta, ja ladata outputiin knitr-vaiheessa. Parempi tulostaa kuvatdsto pdf-ajurilla, jos lopulliseen versioon joutuu näin tekemään (13.5.2018 ). Tämä voi olla järkevä tapa analyysivaiheessa? Teksti kopsattu alla olevasta koodilohkosta.

Ensimmäinen korrespondenssianalyysi - kokeiluja kuvasuhteen säätämiseksi output-dokumentissa. RStudiassa voi avata komentokehoitteessa grafiikkaikkunan. Siitä käsin tallennettu pdf-kuva on ladattu alla Rmarkdowenin omalla komennolla, kohdistus keskelle. Parhaiten näyttäisi toimivan knitrin funktio, mutta oletuskuvakolla saa ca-kuvasta näköjään aika lähelle oikeanlaisen ilman mitään temppuja.

**zxy** Selventäisikö vielä khii2-etäisyyksien taulukko, tai ehkä seuraavassa luvussa? **#V** MG&Blasius, “vihreän kirja”, johdanto.

**Rivien (1) ja sarakkeiden (2) khii2-etäisyydet keskiarvosta.**

```
# khii2 - etäisyyksien taulukko
#str(simpleCA1)
#simpleCA1$rowdist
#str(simpleCA1$rowdist)
#tblRowDist <- simpleCA1$rowdist
#rownames(tblRowDist) <- simpleCA1$rownames
simpleCA1$rowdist

## [1] 0.1579735 0.6309909 0.1750128 0.6340627 0.3477331 0.5504040
simpleCA1$coldist

## [1] 0.5246525 0.3248840 0.3078230 0.2721699 0.6271108

CA-ratkaisun lähtötieto: suhteelliset frekvenssit (korrespondenssimatriisi P)
taulu5 <- ISSP2012esim1.dat %>% tableX(maa,Q1b,type = "cell_perc")
knitr::kable(taulu5,digits = 2, booktabs = TRUE,
             caption = "Kysymyksen V6 vastaukset maittain (%)")
```

Taulukko 87: Kysymyksen V6 vastaukset maittain (%)

	S	s	?	e	E	Total
BE	2.35	5.54	5.38	6.78	4.68	24.72
BG	1.45	4.85	2.52	2.33	0.16	11.31
DE	2.03	4.61	2.43	6.61	5.38	21.05
DK	0.86	2.92	1.87	2.85	8.55	17.05
FI	0.58	2.31	1.83	5.19	3.72	13.63
HU	2.69	3.54	2.76	2.33	0.92	12.24

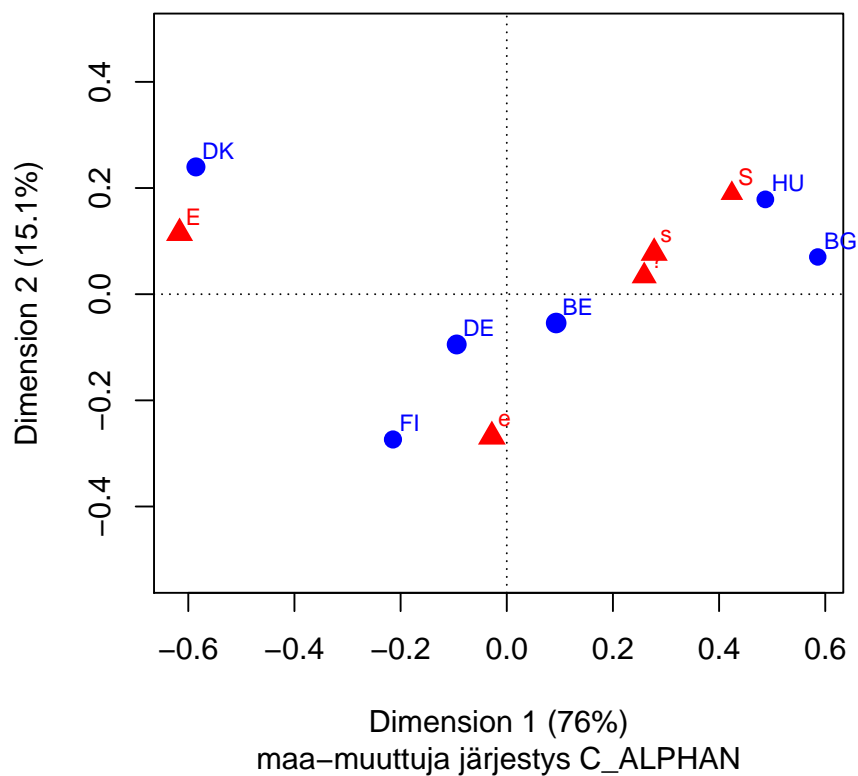
	S	s	?	e	E	Total
Total	9.95	23.76	16.79	26.10	23.41	100.00

**zxy** Tätä ensimmäistä kuvaa on muistiinpanoissa kommentoitu (löytyy printatuna)

```
#simpleCA1 <- ca(~maa + V6,ISSP2012esim1.dat) suoritetaan ennen värikuvaa, tuloksia tarvitaan
#symmetrinen kartta
```

```
plot(simpleCA1, map = "symmetric", mass = c(TRUE,TRUE),
     main = "Lapsi kärsii jos äiti on töissä -symmetrinen kartta",
     sub = "maa-muuttuja järjestys C_ALPHAN")
```

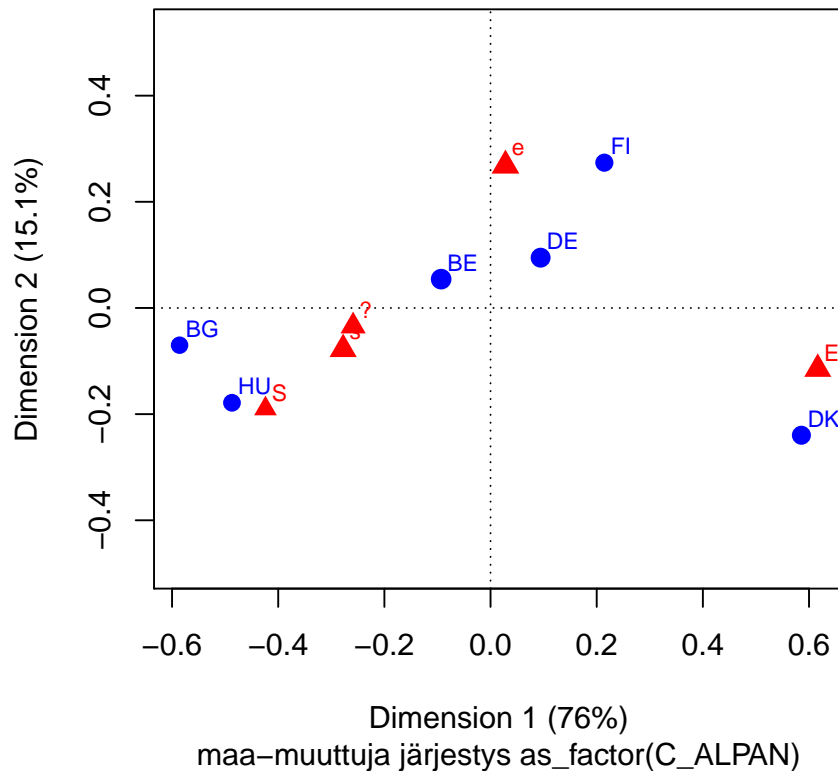
## Lapsi kärsii jos äiti on töissä –symmetrinen kartta



Kuva 2: V6: lapsi kärsii jos äiti on töissä

```
plot(simpleCA2, map = "symmetric", mass = c(TRUE,TRUE),
     main = "Lapsi kärsii jos äiti on töissä -symmetrinen kartta ",
     sub = "maa-muuttuja järjestys as_factor(C_ALPAN)")
```

## Lapsi kärsii jos äiti on töissä –symmetrinen kartta



Kuva 3: V6: lapsi kärsii jos äiti on töissä

```
#str(simpleCA1)
# 13.5.2018
# kuvasuhteen saa oikeaksi, kun avaa g-ikkunan (X11()) ja sitten plot. Voi tallentaa pdf-muotoon
# grafiikkaikkunasta, ja ladata outputiin knitr-vaiheessa. Parempi tulostaa kuvatdsto pdf-aikana
# näin tekemään.
# näitä kokeiln chunk-optioissa mutta ei toimineet (out.width = "6", out.height = "6")
# (13.5.2018), vaan pdf-konversiossa pandoc failed with error 43
```

Ja toinen tapa - kuvatiedoston lataaminen include\_graphics - funktiolla. Ei esitetä tässä. Nämä toiminevat vain pdf-tulostuksessa?

## 2.2 Korrespondenssianalyysin käsitteet

1. Profiilit
2. Massat
3. Profiilien etäisyydet (khii2)

**zxy** Ja tätä “triplettiä” täydentää neljä siitä johdettua käsitettä, viite muistiinpanoissa. **#V** Tässäkin CAiP ja MG2017HY-luentokalvot.

---

## 3 Tulkinnan perusteita

Luvussa syvennetään esimerkin tulkinnan perusteita. Miksi symmetrinen kartta on yleensä paras vaihtoehto, siksi se oletusarvoisesti esitetäänkin. Milloin voi käyttää vaihtoehtoisia esitystapoja? **Ydinluku.**

Esimerkkiaineistossa tulee jo pohdittavaa, Guttman (arc, horseshoe) - efekti, ratkaisun dimensiot jne.

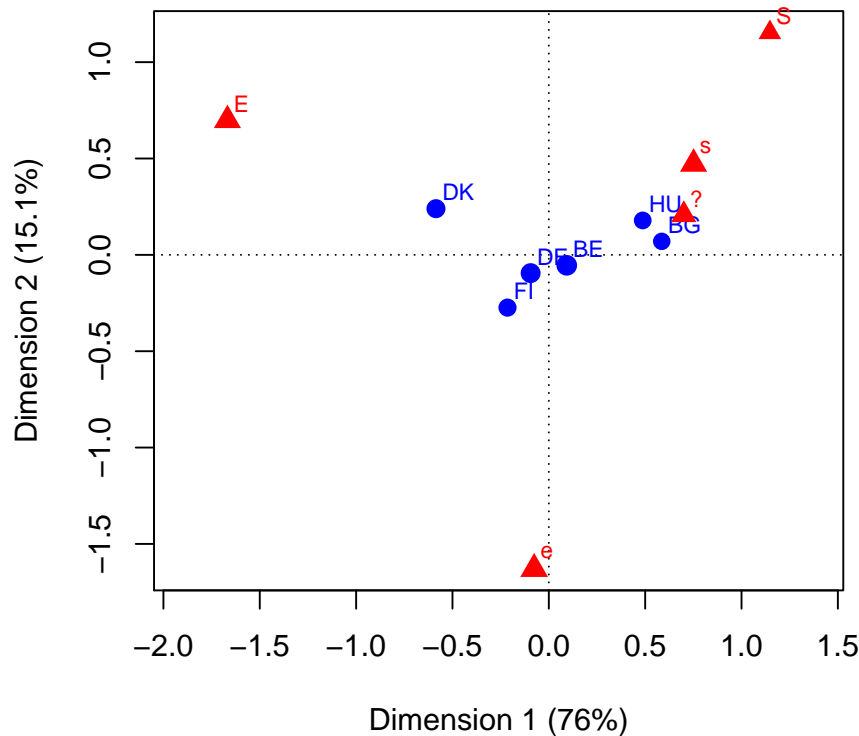
Asymmetrinen kartta, jossa riviprofiilit ovat pääkomponentti-koordinaateissa ja sarakeprofiilit standardikoordinaateissa.

- (1) Sarakkeet ideaalipisteinä, edustavat kuvittellisia maita joissa kaikki ovat vastanneet vain yhdellä tavalla.
- (2) Sarakepisteet kaukana origosta, koska skaalattu
- (3) Rivipisteet kasautuneet keskiarvopisteen ympärille
- (4) Rivi- ja sarakepisteiden suhteelliset sijannit samat kuin symmetrisessä kuvassa
- (5) Tässäkin kuvassa pisteen koko kuvaa sen massaa. Sarakkeista “täysin samaa mieltä” (ts) ja “ei samaa eikä eri mieltä” ovat massoiltaan pienimmät.
- (6) Pisteiden koko kuvaa rivin tai sarakkeen massaa.

```
# asymmetrinen kartta - rivit pc ja sarakkeet sc
# HUOM! simpleCA1 luodaan G1_2_johdesim.Rmd - tiedostossa

plot(simpleCA1, map = "rowprincipal",
     mass = c(TRUE,TRUE),
     main = "Lapsi kärsii jos äiti on töissä -asymmetrinen kartta" )
```

## Lapsi kärsii jos äiti on töissä –asymmetrinen kartta



**HUOM (1.2.20)** Kuva on kääntynyt x-akselin ympäri? (1.2.20) Maat ovat eri järjestyksessä, johtuisiko siitä? Tarkista myös edellisen luvun rivi- ja sarakeprofiilikuvat, niissä on muokkailtu taulukkoa...

Tarinaa voi tarvittaessa jatkaa, tämä on CA:n hankalin asia. Kaksi koordinaattis-  
toa, ja niiden yhteys.

(7) Asymmetrinen kuva ja akseleiden / dimensioiden tulkinta

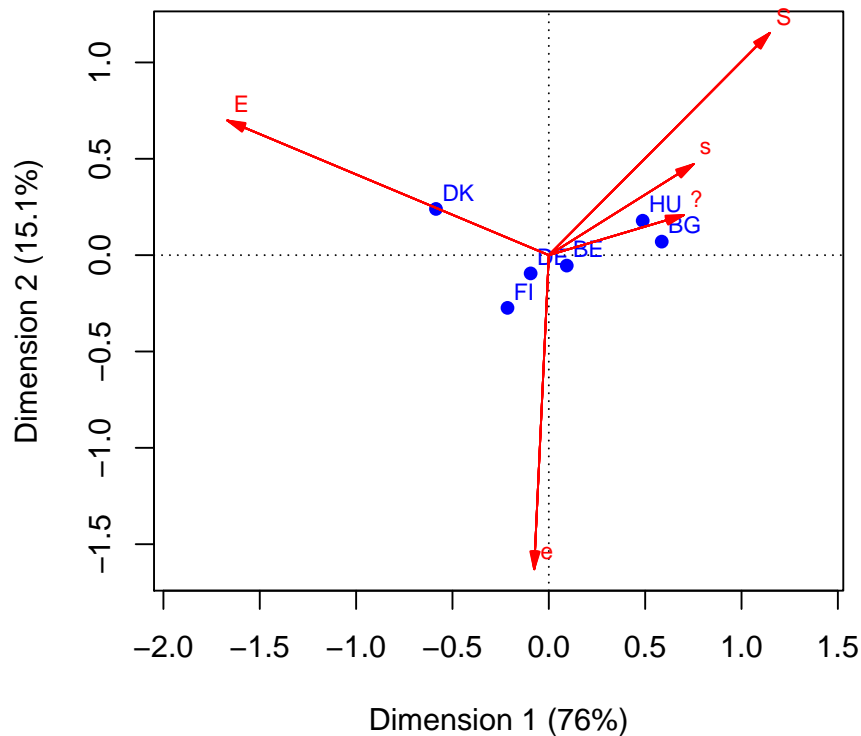
Piirretään sama asymmetrinen kartta uudelleen, mutta yhdistetään sarakepis-  
teet keskiarvopisteeseen (sentroidiin) suorilla. Mitä terävämpi on sarakesuoran  
(vektorin?) ja akselin kulma, sitä enemmän sarake määrittää tätä ulottuvuutta.  
Jos vektori on lähettä 45 asteen kulmaa, sarake määrittää yhtä paljon molempia  
ulottuvuuksia.

```
# asymmetrinen kartta - rivit pc ja sarakkeet sc
# sarakkeet vektorikuvina
# HUOM! simpleCA1 luodaan G1_2_johdesim.Rmd - tiedostossa
```



```
plot(simpleCA1, map = "rowprincipal",
     arrows = c(FALSE,TRUE),
     main = "Lapsi kärsii jos äiti on töissä -asymmetrinen kartta 1" )
```

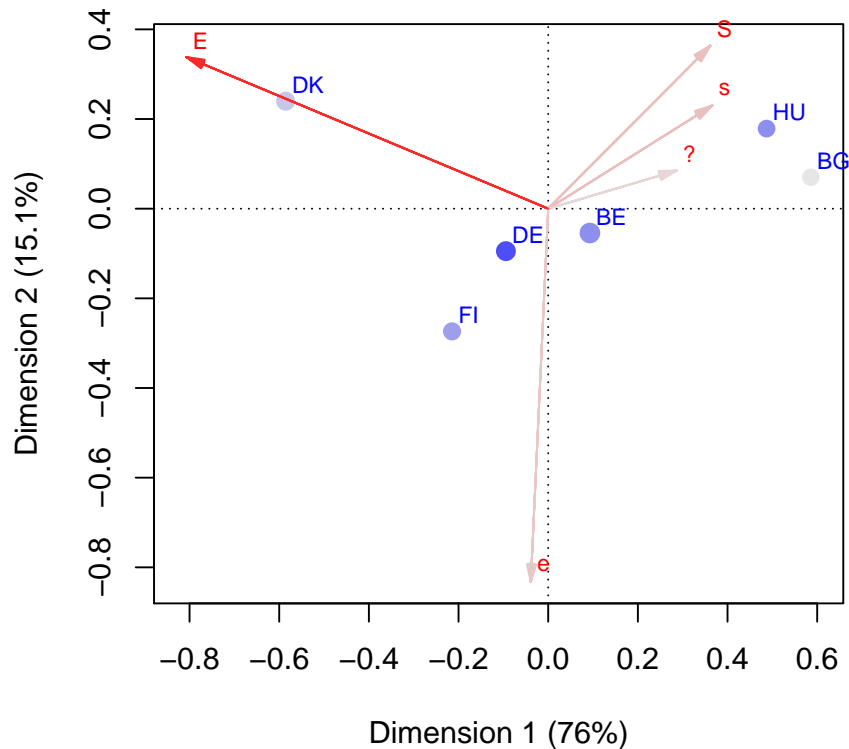
## Lapsi kärsii jos äiti on töissä –asymmetrinen kartta



Tärkein havainto on sarakkeen “Eri mieltä” (e) ja toisen ulottuvuuden yhteys. Myös sarake “täysin samaa mieltä” (ts) määrittää toista ulottuvuutta lähes yhtä paljon kuin ensimmäistä.

```
#X11() komentoriville ja plot-komento
plot(simpleCA1, map = "rowgreen",
     contrib= c("absolute", "absolute"),
     mass = c(TRUE,TRUE),
     arrows = c(FALSE, TRUE),
     main = "Lapsi kärsii jos äiti on töissä - asymmetrinen kartta 2" )
```

## Lapsi kärsii jos äiti on töissä – asymmetrinen kartta



Greenacre (2006, “loose ends -artikkeli”) ehdotti asymmetrisessä kuvassa standardikoordinaattien skaalaamista niin, että ne kerrotaan massan neliöjuurella. Tämä skaalaus toimii hyvin pienen ja suuren inertian tapauksessa. Kartoissa pätee sama sääntö kuin muussakin graafisessa data-analyysissä, kuvien on esitettävä oleelliset yhteydet, mutta mielellään vain ne.

### Tulkinta: rivipisteiden ortogonaalinen projektio “sarakevektorille”

Asymmetrisessä kartassa 2 pisteiden koko on suhteessa niiden massaan, ja värisävy absoluuttiseen kontribuutioon (voi olla myös suhteellinen kontribuutio).

```
# CA:n numeeriset tulokset
```

```
summary(simpleCA1)
```

```
##
## Principal inertias (eigenvalues):
##
## dim      value      %   cum%   scree plot
```

```

## 1      0.136619  76.0  76.0  *****
## 2      0.027089  15.1  91.1  ****
## 3      0.010054   5.6  96.7  *
## 4      0.005988   3.3 100.0  *
##      -----
## Total: 0.179751 100.0
##
##
## Rows:
##      name  mass  qlt  inr      k=1 cor ctr      k=2 cor ctr
## 1 |  BE |  247  465  34 |  93 347  16 | -54 118  27 |
## 2 |  BG |  113  874  251 | 586 862 284 |  70  12  21 |
## 3 |  DE |  210  584  36 | -94 291  14 | -95 293  70 |
## 4 |  DK |  170  996  381 | -586 853 428 | 240 143 362 |
## 5 |  FI |  136 1000  92 | -214 380  46 | -274 620 377 |
## 6 |  HU |  122  889  206 |  487 783 213 |  179 105 144 |
##
## Columns:
##      name  mass  qlt  inr      k=1 cor ctr      k=2 cor ctr
## 1 |  S |  99  784  152 | 424 653 131 | 190 131 132 |
## 2 |  s | 238  788  140 | 278 731 134 |  78  57  53 |
## 3 |   | 168  720  88 | 259 707  82 |  34  12  7 |
## 4 |  e | 261  982  108 | -28  11  2 | -268 971 693 |
## 5 |  E | 234 1000  512 | -616 966 651 |  115  34 114 |

```

**zxy** Taulukon käsitteiden läpikäynti ja pureskelu kuulunee seuraavaan lukuun.

MG & Blasius, “vihreä kirja”: kontribuutiot inertiaan

---



---