

Laura Botzet
9/15/2016

Data import

Library

```
library(devtools)
library(haven)
devtools::install_github("hadley/haven")
```

```
## Skipping install of 'haven' from a github remote, the SHA1 (68e0c2d5) has not changed since last ins
##   Use `force = TRUE` to force installation
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##   filter, lag
```

```
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(plyr)
```

```
## -----
```

```
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)
```

```
## -----
```

```
##
## Attaching package: 'plyr'
```

```
## The following objects are masked from 'package:dplyr':
##
##   arrange, count, desc, failwith, id, mutate, rename, summarise,
##   summarize
```

```
library(tidyr)
```

Load data

```
### Individual and household Information
## Informations about individuals living in the household in 2014/2015
# Individuals living in the household
bk_ar1 = read_dta("data/hh14_all_dta/bk_ar1.dta")

### Informations about earlier waves
ptrack = read_dta("data/hh14_all_dta/ptrack.dta")

### IQ Information
ek_ek1 = read_dta("data/hh14_all_dta/ek_ek1.dta")
ek_ek2 = read_dta("data/hh14_all_dta/ek_ek2.dta")
ek_time1 = read_dta("data/hh14_all_dta/ek_time1.dta")
ek_time2 = read_dta("data/hh14_all_dta/ek_time2.dta")
# additional information (counting backwards, adaptive testing) for adults
b3b_cob = read_dta("data/hh14_all_dta/b3b_cob.dta")
b3b_co1 = read_dta("data/hh14_all_dta/b3b_co1.dta")
### Personality Information (only for adults)
b3b_psn = read_dta("data/hh14_all_dta/b3b_psn.dta")

### Risk taking
b3a_si = read_dta("data/hh14_all_dta/b3a_si.dta")
```

Select data

```
### Individuals
individuals = bk_ar1 %>% select(hhid14_9, pidlink, ar02b, ar10, ar11, ar07, ar08day, ar08mth, ar08yr, a

##create number that should be identical for all siblings in one family
individuals <- unite(individuals,
  col = "hhid14_9_ar10_ar11",
  hhid14_9, ar10, ar11,
  sep = "_",
  remove = FALSE)

#calculate sibling size
individuals <- dplyr::ddply(individuals,.(hhid14_9_ar10_ar11),transform,siblingcount=length(hhid14_9_ar10_ar11)
##calculate birth order??

### IQ Informations
##ek1 (7-14yrs old)
iq1 = ek_ek1 %>% select(hhid14_9, pidlink, age, sex, ektype, resptype, result, reason, ek1_ans, ek2_ans
# Participants who took the ek1 in 2007 were given ek2 again - to make sure that the dont appear in bot
```

```

iq1 = subset(iq1, age<15)

##ek2 (>14yrs)
iq2.1 = ek_ek2 %>% select(hhid14_9, pidlink, age, sex, ektype, resptype, result, reason, ek1_ans, ek2_a
#additional informations for adults: counting backwards
iq2.2 = b3b_co1 %>% select(hhid14_9, pidlink, co04a, co04b, co04c, co04d, co04e, co07count, co10count)
#additional informations for adults: adaptive number test
iq2.3 = b3b_cob %>% select(hhid14_9, pidlink, cob01_b, cob02_b, cob03_b, cob04_b, cob05_b, cob06_b, cob
# put all the informations for participants >= 15 together
iq2 = full_join(iq2.1, iq2.2, by = "pidlink")
iq2 = full_join(iq2, iq2.3, by = "pidlink")

### Put both iq files together
iq = bind_rows(iq1, iq2)

### Personality
##Rearrange personality data so that every individual has only one row
pers = spread(b3b_psn, psntype, psn01)
##name columns
colnames(pers) <- c("hhid14_9", "pid14", "hhid14", "pidlink", "version", "module", "e1", "c1", "o1", "e
pers = pers %>% select(hhid14_9, pidlink, e1, c1, o1, e2r, n1r, a1, n2r, o2, c1r, o3, a2, c2, e3, a3r, r

###Risktaking
risk = b3a_si %>% select(hhid14_9, pidlink, random_si, si01, si02, si03, si04, si05, si11, si12, si13, s

```

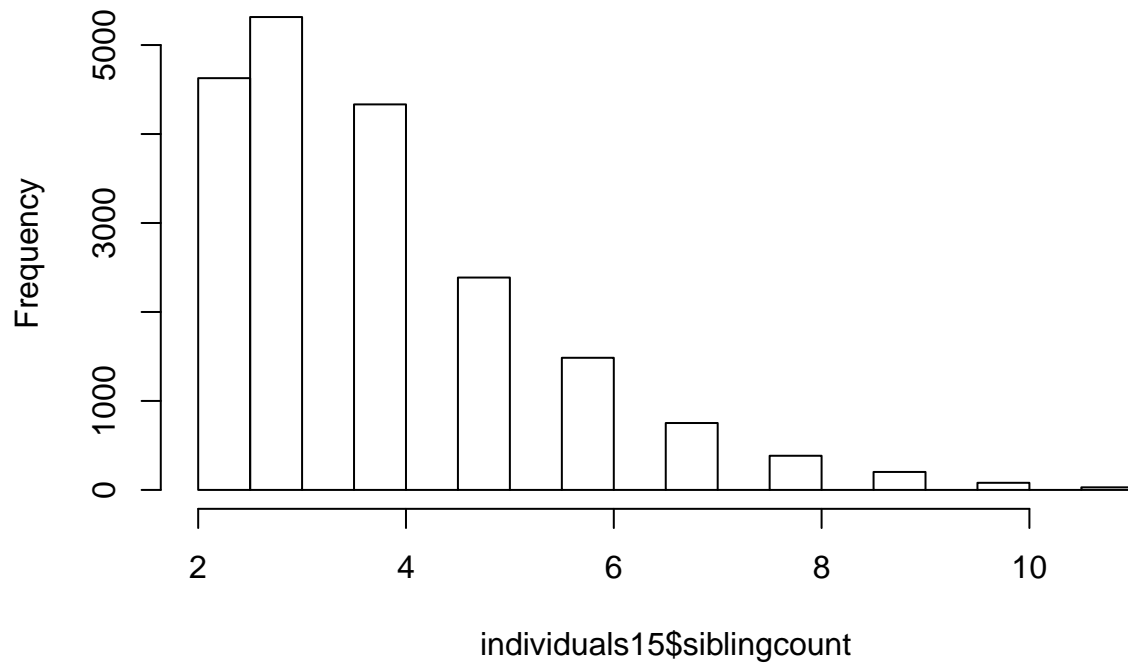
Find out if its possible to analyse data for personality (siblings older than 14)

```

### how many siblings older than 14 are there?
##older than 14
individuals15 = subset(individuals, ar09>14)
##exclude people whose parents are not known
individuals15 = subset(individuals15, ar10 <51)
individuals15 = subset(individuals15, ar11 <51)
##exclude only childs (watch out: sibling size was conduted WITH all younger siblings!)
individuals15=subset(individuals15, siblingcount != 1)
hist(individuals15$siblingcount)

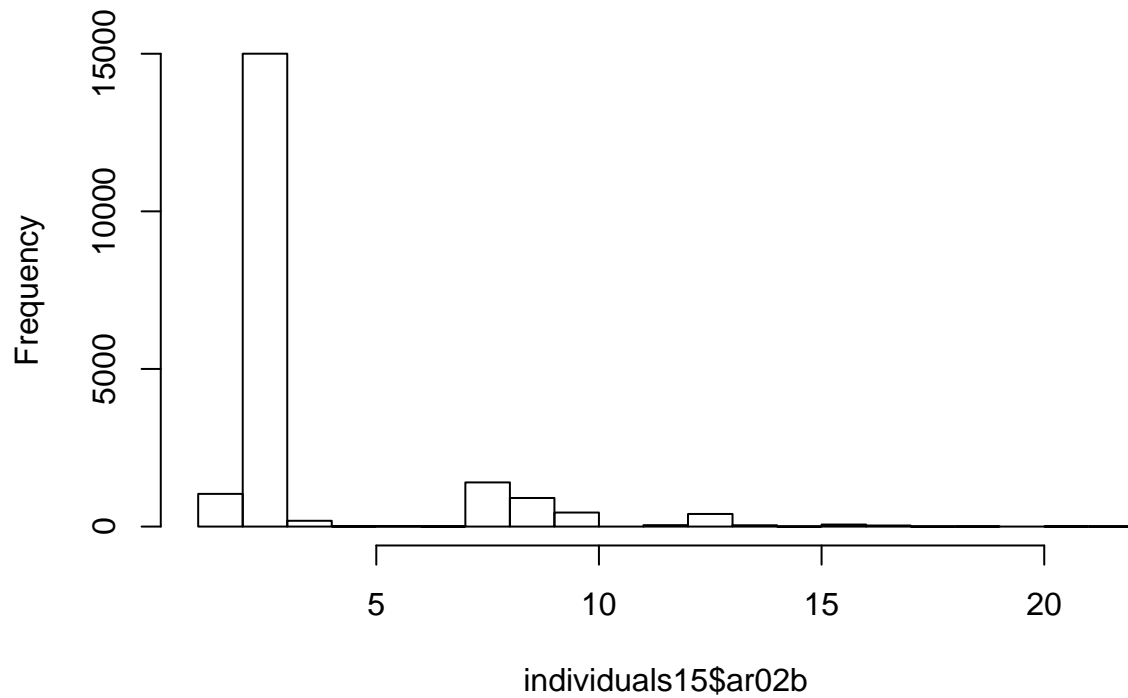
```

Histogram of individuals15\$siblingcount



```
hist(individuals15$ar02b)
```

Histogram of individuals15\$ar02b



Merge data

```
### Merge ALL data (including all participants)
alldata = full_join(individuals, iq, by="pidlink")
alldata = full_join(alldata, pers, by="pidlink")
alldata = full_join(alldata, risk, by="pidlink")
```

Save data

for future analyses

```
saveRDS(alldata, file = "data/hh14_all_dta/alldata.dta")
```