Ketten-Questionnaire

Christian 'Becko' Beck

21 Oktober 2016

Import

Data had to be imported. The original xlsx was improved: two items (Person2 - Pompfe; 103 - best number of stones) were seperated into item groups, since multiple selections were effectively possible and encuraged by the questions. The file was exported into a csv for import into R (note: i know of the package 'xlsx', but it often does not what i want to do, while the other way through a csv has been profen to be rock solid)

```
Testergebnis.Fragebogen <- read.csv2("~/ownCloud/Programming/jugger_data/chain2016/Testergebnis Fragebog

summary(Testergebnis.Fragebogen)
```

```
##     Person.1       Person.2           P2.Stab       P2.LP
##  Min.   : 0.000   Length:52          Min.   :1    Min.   :1
##  1st Qu.: 2.000   Class :character   1st Qu.:1    1st Qu.:1
##  Median : 3.000   Mode  :character   Median :1    Median :1
##  Mean   : 3.755                      Mean   :1    Mean   :1
##  3rd Qu.: 6.000                      3rd Qu.:1    3rd Qu.:1
##  Max.   :11.000                      Max.   :1    Max.   :1
##  NA's   :3                           NA's   :26   NA's   :45
##    P2.Schild       P2.Q       P2.Kette     P2.Quick       X10
##  Min.   :1    Min.   :1    Min.   :1    Min.   :1    Min.   :1.000
##  1st Qu.:1    1st Qu.:1    1st Qu.:1    1st Qu.:1    1st Qu.:4.000
##  Median :1    Median :1    Median :1    Median :1    Median :5.000
##  Mean   :1    Mean   :1    Mean   :1    Mean   :1    Mean   :4.373
##  3rd Qu.:1    3rd Qu.:1    3rd Qu.:1    3rd Qu.:1    3rd Qu.:5.000
##  Max.   :1    Max.   :1    Max.   :1    Max.   :1    Max.   :5.000
##  NA's   :43   NA's   :42   NA's   :37   NA's   :47   NA's   :1
##       X20             X30             X41             X42
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:1.250   1st Qu.:1.000   1st Qu.:1.000   1st Qu.:1.000
##  Median :3.000   Median :2.000   Median :1.000   Median :1.000
##  Mean   :2.778   Mean   :2.364   Mean   :1.918   Mean   :1.885
##  3rd Qu.:3.750   3rd Qu.:3.000   3rd Qu.:3.000   3rd Qu.:3.000
##  Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.000
##  NA's   :34      NA's   :8       NA's   :3
##       X43             X44             X45             X51
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :0.0000
##  1st Qu.:1.000   1st Qu.:1.000   1st Qu.:3.000   1st Qu.:0.0000
##  Median :2.000   Median :2.000   Median :4.000   Median :0.0000
##  Mean   :1.939   Mean   :1.837   Mean   :3.808   Mean   :0.2941
##  3rd Qu.:3.000   3rd Qu.:2.000   3rd Qu.:5.000   3rd Qu.:1.0000
##  Max.   :5.000   Max.   :4.000   Max.   :5.000   Max.   :1.0000
##  NA's   :3       NA's   :3                       NA's   :1
##       X52            X53             X54             X55
##  Min.   :0.00   Min.   :1.000   Min.   :1.000   Min.   :1.00
##  1st Qu.:0.00   1st Qu.:3.000   1st Qu.:1.000   1st Qu.:2.00
```

```
##  Median :0.00   Median :4.000   Median :2.000   Median :3.00  
##  Mean   :0.69   Mean   :3.974   Mean   :1.927   Mean   :3.13  
##  3rd Qu.:1.00   3rd Qu.:5.000   3rd Qu.:3.000   3rd Qu.:4.00  
##  Max.   :6.00   Max.   :5.000   Max.   :4.000   Max.   :5.00  
##  NA's   :2      NA's   :14      NA's   :11      NA's   :6     
##       X56            X61             X62             X63       
##  Min.   :1.00   Min.   :1.000   Min.   :1.000   Min.   :1.000  
##  1st Qu.:2.00   1st Qu.:2.000   1st Qu.:3.000   1st Qu.:2.000  
##  Median :3.00   Median :4.000   Median :4.000   Median :3.000  
##  Mean   :2.87   Mean   :3.423   Mean   :3.667   Mean   :3.192  
##  3rd Qu.:4.00   3rd Qu.:5.000   3rd Qu.:5.000   3rd Qu.:4.000  
##  Max.   :5.00   Max.   :5.000   Max.   :5.000   Max.   :5.000  
##  NA's   :6                      NA's   :1                      
##       X64             X65             X66             X67       
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.00  
##  1st Qu.:2.000   1st Qu.:2.000   1st Qu.:3.000   1st Qu.:2.00  
##  Median :3.000   Median :3.000   Median :4.000   Median :3.00  
##  Mean   :2.725   Mean   :2.706   Mean   :3.608   Mean   :2.94  
##  3rd Qu.:3.000   3rd Qu.:3.000   3rd Qu.:4.000   3rd Qu.:4.00  
##  Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.00  
##  NA's   :1       NA's   :1       NA's   :1       NA's   :2     
##       X71             X72             X81             X91       
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000  
##  1st Qu.:3.000   1st Qu.:2.000   1st Qu.:2.000   1st Qu.:2.000  
##  Median :4.000   Median :3.000   Median :2.500   Median :3.000  
##  Mean   :3.308   Mean   :3.157   Mean   :2.654   Mean   :2.923  
##  3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:4.000  
##  Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.000  
##                  NA's   :1                                      
##       X92             X93            X94             X95       
##  Min.   :1.000   Min.   : 5.0   Min.   :1.000   Min.   :1.000  
##  1st Qu.:2.000   1st Qu.:60.0   1st Qu.:2.000   1st Qu.:2.000  
##  Median :3.000   Median :67.5   Median :3.000   Median :3.000  
##  Mean   :2.694   Mean   :63.2   Mean   :2.923   Mean   :3.231  
##  3rd Qu.:4.000   3rd Qu.:70.0   3rd Qu.:4.000   3rd Qu.:4.000  
##  Max.   :5.000   Max.   :95.0   Max.   :5.000   Max.   :5.000  
##  NA's   :3       NA's   :2                                     
##       X96            X97            X98             X99       
##  Min.   :1.00   Min.   :1.00   Min.   :1.000   Min.   :1.000  
##  1st Qu.:2.00   1st Qu.:1.00   1st Qu.:3.000   1st Qu.:2.750  
##  Median :3.00   Median :2.00   Median :4.000   Median :4.000  
##  Mean   :3.14   Mean   :2.18   Mean   :3.412   Mean   :3.327  
##  3rd Qu.:4.00   3rd Qu.:3.00   3rd Qu.:4.000   3rd Qu.:4.000  
##  Max.   :5.00   Max.   :5.00   Max.   :5.000   Max.   :5.000  
##  NA's   :2      NA's   :2      NA's   :1                      
##       X101            X102           X103               X103.5   
##  Min.   :1.000   Min.   :1.000   Length:52          Min.   :1   
##  1st Qu.:1.000   1st Qu.:3.000   Class :character   1st Qu.:1   
##  Median :1.000   Median :4.000   Mode  :character   Median :1   
##  Mean   :1.788   Mean   :3.519                      Mean   :1   
##  3rd Qu.:2.250   3rd Qu.:5.000                      3rd Qu.:1   
##  Max.   :5.000   Max.   :5.000                      Max.   :1   
##                                                     NA's   :32  
##     X103.6       X103.7       X103.8   
```

```
##  Min.   :1    Min.   :1    Min.   :1
##  1st Qu.:1    1st Qu.:1    1st Qu.:1
##  Median :1    Median :1    Median :1
##  Mean   :1    Mean   :1    Mean   :1
##  3rd Qu.:1    3rd Qu.:1    3rd Qu.:1
##  Max.   :1    Max.   :1    Max.   :1
##  NA's   :38   NA's   :44   NA's   :35
```

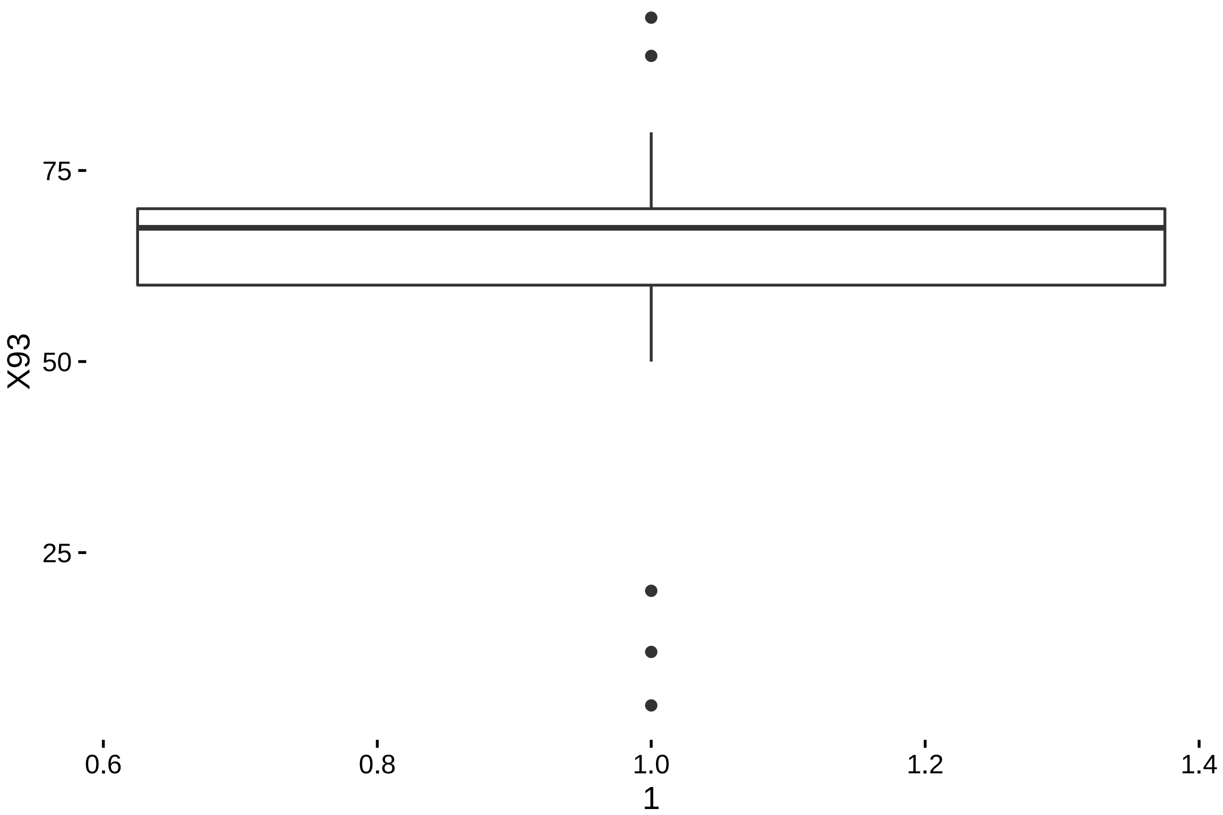
Cleaning

```
library(magrittr)
library(dplyr)
##
## Attaching package: 'dplyr'
## Die folgenden Objekte sind maskiert von 'package:stats':
##
##     filter, lag
## Die folgenden Objekte sind maskiert von 'package:base':
##
##     intersect, setdiff, setequal, union

Testergebnis.Fragebogen %<>% select(-Person.2, -X103)
Testergebnis.Fragebogen$P2.Stab[is.na(Testergebnis.Fragebogen$P2.Stab)] <- 0
Testergebnis.Fragebogen$P2.LP[is.na(Testergebnis.Fragebogen$P2.LP)] <- 0
Testergebnis.Fragebogen$P2.Schild[is.na(Testergebnis.Fragebogen$P2.Schild)] <- 0
Testergebnis.Fragebogen$P2.Q[is.na(Testergebnis.Fragebogen$P2.Q)] <- 0
Testergebnis.Fragebogen$P2.Kette[is.na(Testergebnis.Fragebogen$P2.Kette)] <- 0
Testergebnis.Fragebogen$P2.Quick[is.na(Testergebnis.Fragebogen$P2.Quick)] <- 0

Testergebnis.Fragebogen$X103.5[is.na(Testergebnis.Fragebogen$X103.5)] <- 0
Testergebnis.Fragebogen$X103.6[is.na(Testergebnis.Fragebogen$X103.6)] <- 0
Testergebnis.Fragebogen$X103.7[is.na(Testergebnis.Fragebogen$X103.7)] <- 0
Testergebnis.Fragebogen$X103.8[is.na(Testergebnis.Fragebogen$X103.8)] <- 0

Testergebnis.Fragebogen %<>% mutate(P2.Stab = as.factor(P2.Stab), P2.LP = as.factor(P2.LP), P2.Schild =
```

Note: There are a lot of questions that are asked wrong - the meaning of some of the answers is in many cases not clear. This is due to relative statements with no given baseline in the questions. Given the context and that the hypothesis to each statement is known (or guessable) beforehand makes it still possible - but quit dirty - to analyse all questions. However i issue a warning to not give these results to much credibility.

Note on Q93: Gerd already found that two outliers exist. Some more do exist and the downward outliers might be caused by reading/understanding the question wrong. I do not try to correct this (since i do not know the truth) and keep them in the data set. At a later point i might decide to drop them and will note this accordingly

```
library(dplyr)
library(ggplot2)

ggplot(Testergebnis.Fragebogen %>% select(X93) %>% filter(!is.na(X93))) + geom_boxplot(aes(x = 1, y = X9
```



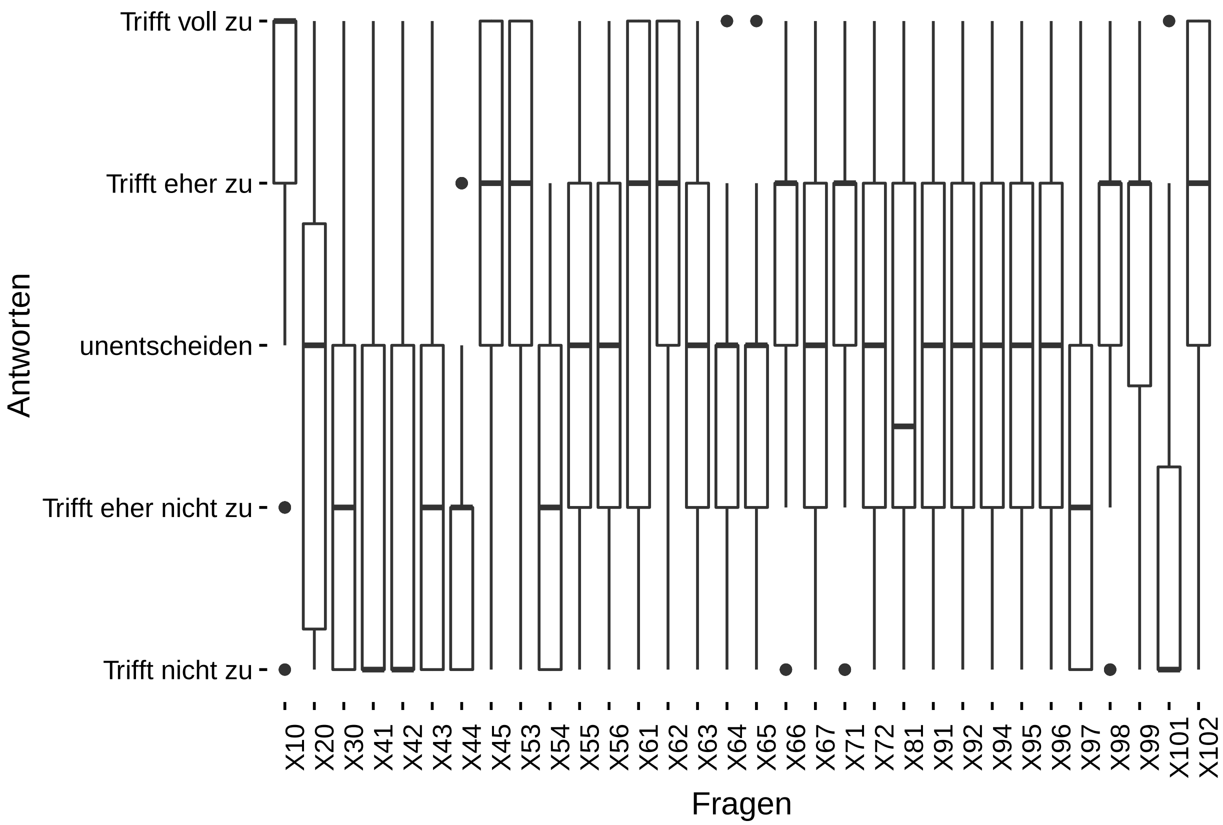

```
# tX93 <- scale(Testergebnis.Fragebogen$X93)
# Testergebnis.Fragebogen$X93[abs(tX93) > 1.5] <- NA
```

Basic Visualisation and exploration

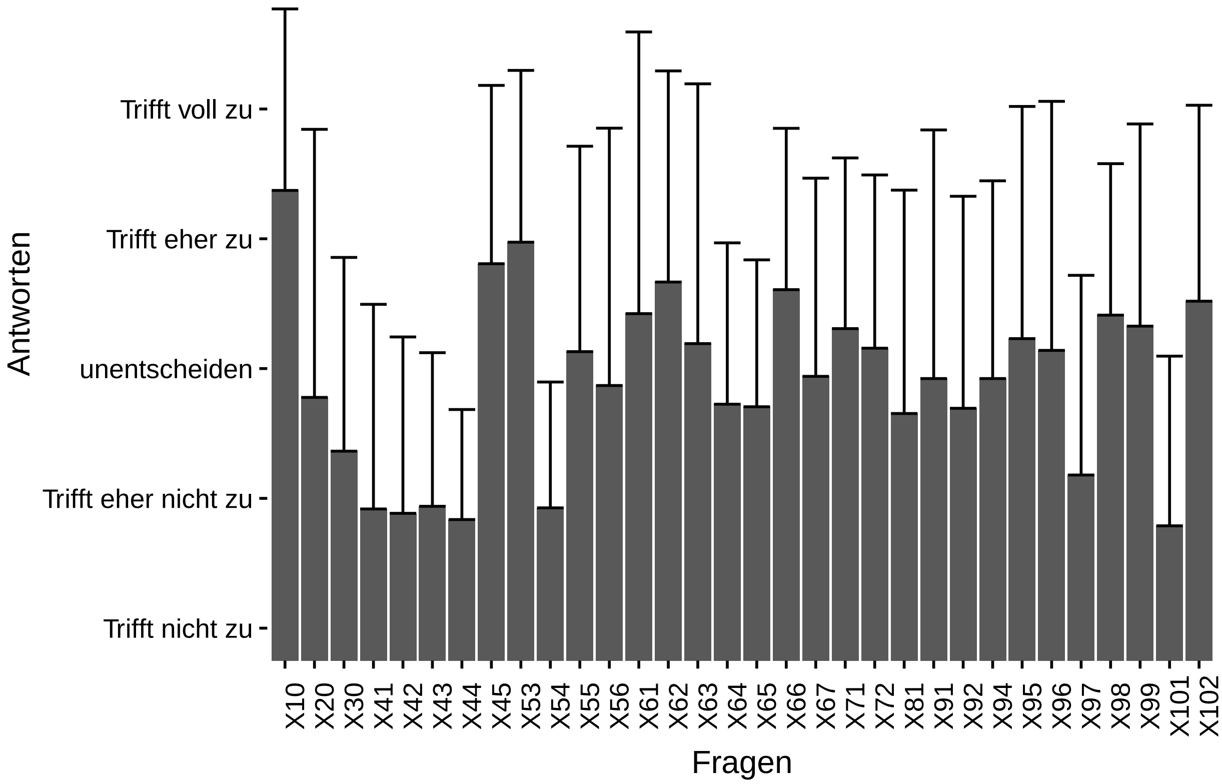
Note: Question Person.2 (which is your MAIN Pompfe) is somewhat ill posed. However important in the current context is just the number of chain players, all others can be treated the same.

- Count of questionnaires: 52
- Count of chains: 15
- Ratio of chains: 0.2884615

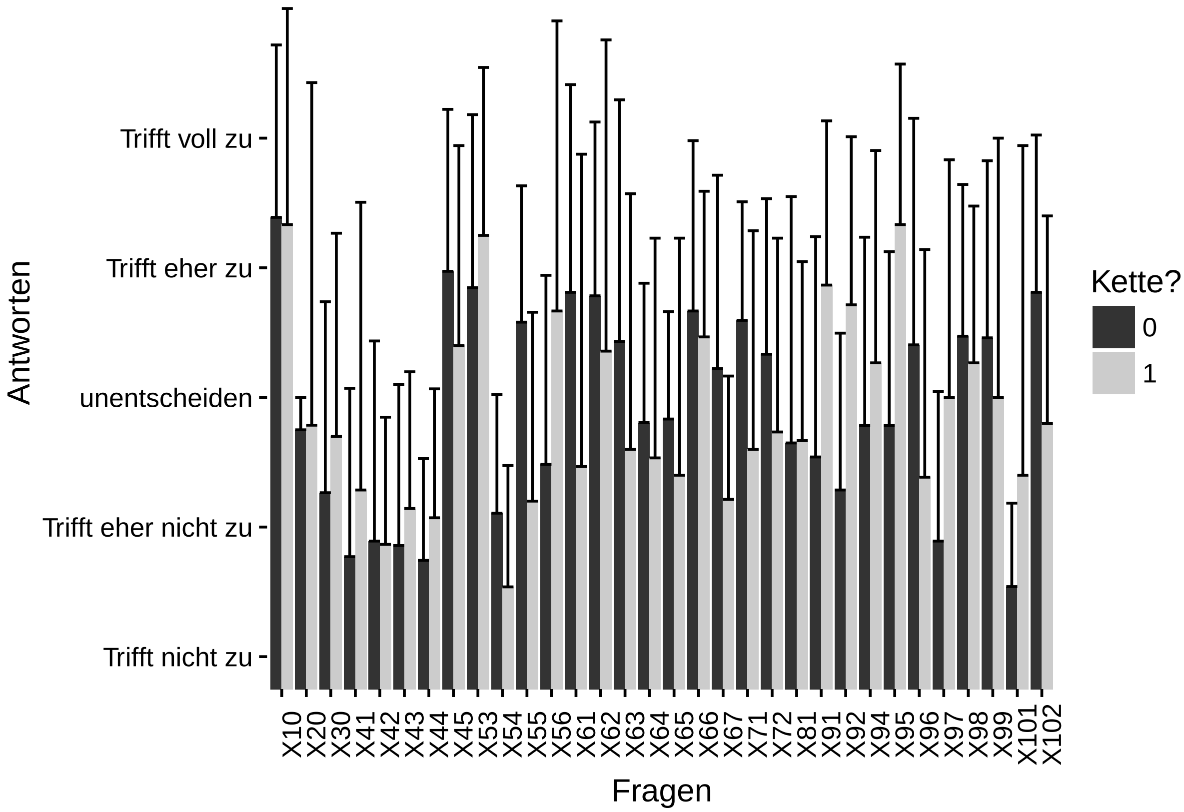
```
library(ggplot2)
library(reshape2)
library(knitr)

# boxplot for each score
ggplot(melt(Testergebnis.Fragebogen %>% select(-Person.1, -P2.Q,-P2.Stab,-P2.LP,-P2.Schild,-P2.Quick, -I
  geom_boxplot(aes(x = variable, y = value)) +
  theme_classic() + theme(axis.text.x = element_text(angle = 90)) +
  scale_y_continuous(name = "Antworten", breaks = 1:5, minor_breaks = NULL, labels = c("Trifft nicht zu
  scale_x_discrete(name = "Fragen")
```




```r
# Aggregate of all scores
Testergebnis.stats <- Testergebnis.Fragebogen %>% select(-Person.1, -P2.Q,-P2.Stab,-P2.LP,-P2.Schild,-P2
ggplot(Testergebnis.stats) +
  geom_bar(aes(x = variable, y = avg),stat = "identity") +
  geom_errorbar(aes(x = variable, ymin = avg, ymax = avg+var)) +
  coord_cartesian(ylim = c(1,6)) +
  theme_classic() + theme(axis.text.x = element_text(angle = 90)) +
  scale_y_continuous(name = "Antworten", breaks = 1:5, minor_breaks = NULL, labels = c("Trifft nicht zu
  scale_x_discrete(name = "Fragen")
```




```
# Aggregate of all scores nach Kette/nicht-Kette
Testergebnis.stats2 <- Testergebnis.Fragebogen %>% select(-Person.1, -P2.Q,-P2.Stab,-P2.LP,-P2.Schild,-P
ggplot(Testergebnis.stats2) +
  geom_bar(aes(x = variable, y = avg, fill = P2.Kette),stat = "identity", position = "dodge") +
  geom_errorbar(aes(x = variable, ymin = avg, ymax = avg+var, fill = P2.Kette), position = "dodge", wid
  coord_cartesian(ylim = c(1,6)) +
  theme_classic() + theme(axis.text.x = element_text(angle = 90)) +
  scale_y_continuous(name = "Antworten", breaks = 1:5, minor_breaks = NULL, labels = c("Trifft nicht zu
  scale_x_discrete(name = "Fragen") +
#  scale_fill_discrete(name = "Kette") +
  scale_fill_grey(name = "Kette?")
```



First statistics

```
chain.stats <- Testergebnis.Fragebogen %>% select(-Person.1, -P2.Q,-P2.Stab,-P2.LP,-P2.Schild,-P2.Quick

##### normality
ntest <- data.frame(Question = colnames(chain.stats), W = NA, p = NA, "H0 rejected?" = NA, stringsAsFac
for (ci in 1:ncol(chain.stats)) {
  st <- shapiro.test(x = chain.stats[,ci])
  ntest[ci,2] <- st$statistic
  ntest[ci,3] <- st$p.value
  ntest[ci,4] <- st$p.value < 0.05
}

kable(x = ntest, digits = 2,
      caption = "Results of the Shapiro-test for normality. TRUE means H0 was rejected at confidence le
```

Table 1: Results of the Shapiro-test for normality. TRUE means H0 was rejected at confidence level 0.05.

Question	W	p	H0.rejected.
X10	0.59	0.00	TRUE
X20	0.88	0.03	TRUE
X30	0.87	0.00	TRUE
X41	0.72	0.00	TRUE

Question	W	p	H0.rejected.
X42	0.76	0.00	TRUE
X43	0.80	0.00	TRUE
X44	0.79	0.00	TRUE
X45	0.85	0.00	TRUE
X53	0.82	0.00	TRUE
X54	0.82	0.00	TRUE
X55	0.91	0.00	TRUE
X56	0.89	0.00	TRUE
X61	0.83	0.00	TRUE
X62	0.84	0.00	TRUE
X63	0.88	0.00	TRUE
X64	0.90	0.00	TRUE
X65	0.86	0.00	TRUE
X66	0.86	0.00	TRUE
X67	0.90	0.00	TRUE
X71	0.89	0.00	TRUE
X72	0.90	0.00	TRUE
X81	0.89	0.00	TRUE
X91	0.87	0.00	TRUE
X92	0.90	0.00	TRUE
X94	0.90	0.00	TRUE
X95	0.89	0.00	TRUE
X96	0.89	0.00	TRUE
X97	0.83	0.00	TRUE
X98	0.85	0.00	TRUE
X99	0.89	0.00	TRUE
X101	0.72	0.00	TRUE
X102	0.89	0.00	TRUE

```
###### H0 test
wtest <- data.frame(Question = colnames(chain.stats), V = NA, p = NA, "H0 rejected?" = NA, "H0 rejected
for (ci in 1:ncol(chain.stats)) {
  wt <- wilcox.test(x = chain.stats[,ci], alternative = "two.sided", mu = 3)
  wtest[ci,2] <- wt$statistic
  wtest[ci,3] <- wt$p.value
  wtest[ci,4] <- wt$p.value < 0.05
  wtest[ci,5] <- wt$p.value < 0.05 / ncol(chain.stats)
}

kable(x = wtest, digits = 2,
      caption = "Results of the Wilcoxon signed rank test with continuity correction. TRUE means H0 was
```

Table 2: Results of the Wilcoxon signed rank test with continuity correction. TRUE means H0 was rejected at confidence level 0.05 - either without or with Bonferroni correction for multiple comparison.

Question	V	p	H0.rejected.	H0.rejected..Boferroni
X10	1137.5	0.00	TRUE	TRUE
X20	30.5	0.51	FALSE	FALSE

Question	V	p	H0.rejected.	H0.rejected..Boferroni
X30	99.0	0.00	TRUE	FALSE
X41	55.0	0.00	TRUE	TRUE
X42	89.0	0.00	TRUE	TRUE
X43	44.5	0.00	TRUE	TRUE
X44	48.0	0.00	TRUE	TRUE
X45	643.5	0.00	TRUE	TRUE
X53	437.0	0.00	TRUE	TRUE
X54	22.5	0.00	TRUE	TRUE
X55	264.5	0.50	FALSE	FALSE
X56	296.0	0.55	FALSE	FALSE
X61	760.0	0.07	FALSE	FALSE
X62	755.0	0.00	TRUE	FALSE
X63	488.0	0.45	FALSE	FALSE
X64	149.5	0.08	FALSE	FALSE
X65	88.0	0.04	TRUE	FALSE
X66	705.0	0.00	TRUE	TRUE
X67	252.5	0.61	FALSE	FALSE
X71	508.5	0.08	FALSE	FALSE
X72	428.5	0.38	FALSE	FALSE
X81	281.0	0.07	FALSE	FALSE
X91	442.5	0.53	FALSE	FALSE
X92	256.0	0.09	FALSE	FALSE
X94	409.5	0.59	FALSE	FALSE
X95	487.5	0.29	FALSE	FALSE
X96	472.5	0.58	FALSE	FALSE
X97	93.0	0.00	TRUE	TRUE
X98	580.0	0.02	TRUE	FALSE
X99	531.0	0.09	FALSE	FALSE
X101	74.0	0.00	TRUE	TRUE
X102	616.0	0.00	TRUE	FALSE

Note: the above test confirm Gerds intuition on most of the questions and answers.

Another note: I hoped for the data to be such that i can highlight some possible p-Hacking in there to support an argument about statistics - how one can (intended or not) shift the results of an anlysis based on his or her prejudices. But "sadly" the data did not follow and so even with two-sided testing and quit conservative Bonferroni correction the pattern of significant answers emerges. (ps: I'm not t-testing here, because the data is not normal enough)

Next: group differences between chain and non-chain players!

```
chain.stats <- Testergebnis.Fragebogen %>% select(-Person.1, -P2.Q,-P2.Stab,-P2.LP,-P2.Schild,-P2.Quick

###### two-sample H0 test
twtest <- data.frame(Question = colnames(chain.stats)[-1], V = NA, p = NA, "H0 rejected?" = NA, "H0 rej
for (ci in 2:ncol(chain.stats)) {
  twt <- wilcox.test(x = chain.stats[chain.stats[,1] == 1,ci], y = chain.stats[chain.stats[,1] != 1,ci]
  twtest[ci-1,2] <- twt$statistic
  twtest[ci-1,3] <- twt$p.value
  twtest[ci-1,4] <- twt$p.value < 0.05
  twtest[ci-1,5] <- twt$p.value < 0.05 / ncol(chain.stats)
}
```

```r
kable(x = twtest, digits = 2,
      caption = "Results of the two-sample Wilcoxon signed rank test with continuity correction - diffe
```

Table 3: Results of the two-sample Wilcoxon signed rank test with continuity correction - differences in ratings between chain players and non-chain players. TRUE means H0 was rejected at confidence level 0.05 - either without or with Bonferroni correction for multiple comparison.

Question	V	p	H0.rejected.	H0.rejected..Boferroni
X10	270.0	1.00	FALSE	FALSE
X20	28.0	1.00	FALSE	FALSE
X30	205.5	0.31	FALSE	FALSE
X41	294.5	0.21	FALSE	FALSE
X42	291.5	0.77	FALSE	FALSE
X43	289.0	0.30	FALSE	FALSE
X44	294.0	0.25	FALSE	FALSE
X45	205.5	0.13	FALSE	FALSE
X53	191.5	0.25	FALSE	FALSE
X54	116.5	0.05	FALSE	FALSE
X55	88.5	0.00	TRUE	TRUE
X56	340.5	0.01	TRUE	FALSE
X61	152.5	0.01	TRUE	FALSE
X62	224.0	0.45	FALSE	FALSE
X63	185.5	0.06	FALSE	FALSE
X64	216.5	0.25	FALSE	FALSE
X65	216.5	0.24	FALSE	FALSE
X66	239.5	0.51	FALSE	FALSE
X67	130.0	0.01	TRUE	FALSE
X71	152.5	0.01	TRUE	FALSE
X72	192.0	0.10	FALSE	FALSE
X81	284.5	0.89	FALSE	FALSE
X91	431.0	0.00	TRUE	TRUE
X92	399.0	0.00	TRUE	TRUE
X94	339.5	0.20	FALSE	FALSE
X95	473.5	0.00	TRUE	TRUE
X96	141.0	0.02	TRUE	FALSE
X97	349.5	0.01	TRUE	FALSE
X98	240.0	0.51	FALSE	FALSE
X99	225.5	0.28	FALSE	FALSE
X101	356.5	0.07	FALSE	FALSE
X102	150.0	0.01	TRUE	FALSE

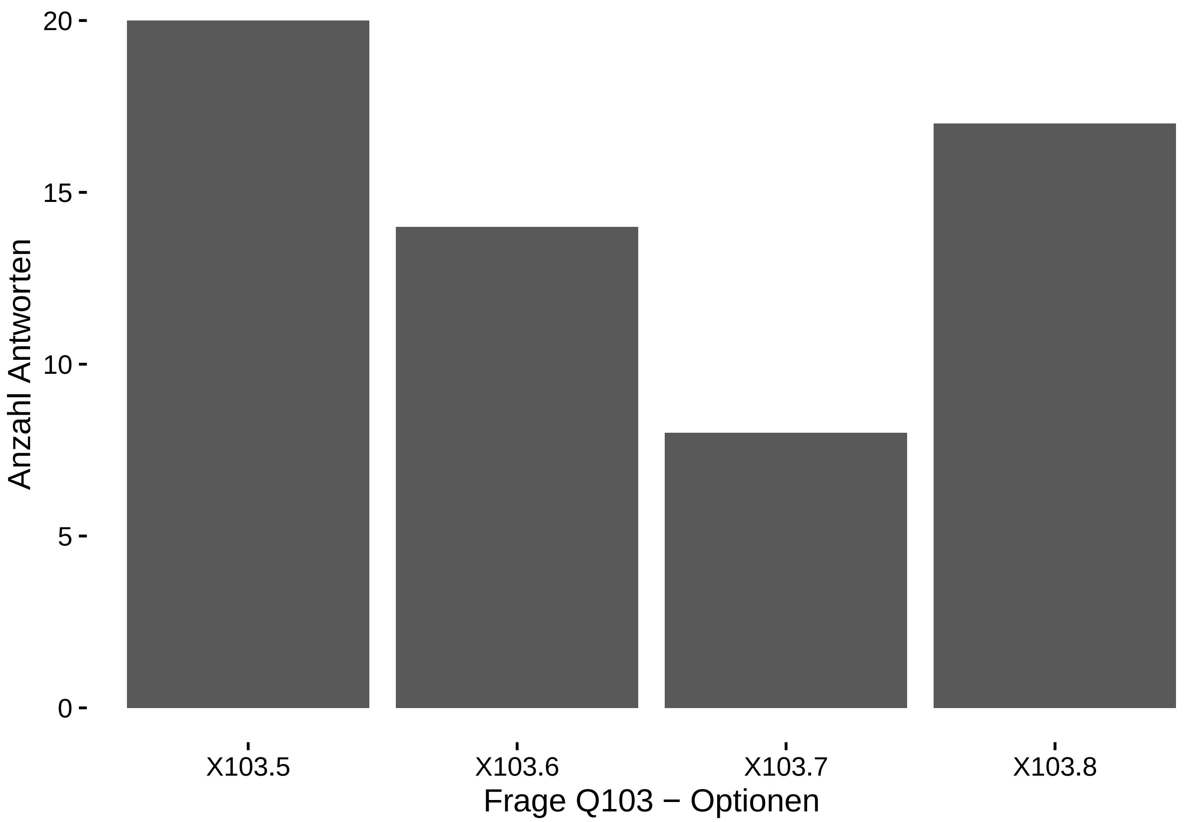
Now it's getting interesting (i report only Bonferroni corrected significance, since no hypothesis exists):

- On doubles chain players and non-chain players seem to disagree (Q55)
- The three questions about the value of chain specialities - such as hit-rate, reach and pin (Q91, Q92, Q95) - are seen differently by chains than by non-chains.

And now onto Gerds big news: **Question 103!** What number of stones is preferred? 5,6,7 or *more than 8*? (note that *more than 8* was recorded by some subjects as *8 or above*, nevertheless the answer options were erronouse and we can't dismiss any effect of this on the results with absoulte certainty - even though it seems plausible)

```
library(reshape2)
X103.stats <- Testergebnis.Fragebogen %>% select(X103.5,X103.6,X103.7,X103.8) %>% summarise_all(funs(su

ggplot(melt(X103.stats)) +
  geom_bar(aes(x = variable, y = value), stat = "identity") +
  theme_classic() +
  scale_y_continuous(name = "Anzahl Antworten") +
  scale_x_discrete(name = "Frage Q103 - Optionen")
```

```
## No id variables; using all as measure variables
```




```
# only consider answers for 5 and 8 (6 and 7 are noise to me, since no hypothesis nor test for those wa
binom.test(x = c(X103.stats$X103.5[1], X103.stats$X103.8[1]), p = 0.5, alternative = "greater")
```

```
##
##  Exact binomial test
##
## data:  c(X103.stats$X103.5[1], X103.stats$X103.8[1])
## number of successes = 20, number of trials = 37, p-value = 0.3714
## alternative hypothesis: true probability of success is greater than 0.5
## 95 percent confidence interval:
##  0.3938002 1.0000000
## sample estimates:
```

```
## probability of success
##               0.5405405
```

```
binom.test(x = X103.stats$X103.5[1], n = nrow(Testergebnis.Fragebogen), p = 0.5, alternative = "greater
```

```
##
##  Exact binomial test
##
## data:  X103.stats$X103.5[1] and nrow(Testergebnis.Fragebogen)
## number of successes = 20, number of trials = 52, p-value = 0.9648
## alternative hypothesis: true probability of success is greater than 0.5
## 95 percent confidence interval:
##  0.2714462 1.0000000
## sample estimates:
## probability of success
##               0.3846154
```

As we can see, there is no significant trend towards one side or the other. Given that we have a very likely recruiters bias (among other methodlogical problems) i might add that the 8-to-5 ratio might be totally diffferent in the general population, as to 8 being the significant more prominent answer - this of course is speculative, but on the other side many of Gerds conclusions are speculative too.

Relations

Years of experience against ratings and ratings against ratings:

```
chain.stats <- Testergebnis.Fragebogen %>% select(-P2.Q,-P2.Stab,-P2.LP,-P2.Schild,-P2.Quick,-P2.Kette,

chain.stats.cor <- chain.stats %>% psych::corr.test(use = "pairwise", method = "pearson", adjust = "holm
# note: holm multiple comparison correction is already applied here

kable(x = data.frame(Question = colnames(chain.stats.cor$r)[-1],
                     ChainR = chain.stats.cor$r[-1,"Person.1"],
                     ChainP = chain.stats.cor$p[-1,"Person.1"],
                     ChainPs = chain.stats.cor$p[-1,"Person.1"] < 0.05),
      digits = 2, caption = "Correlation results of years of experience (Person.1) with all other rating
```

Table 4: Correlation results of years of experience (Person.1) with all other ratings

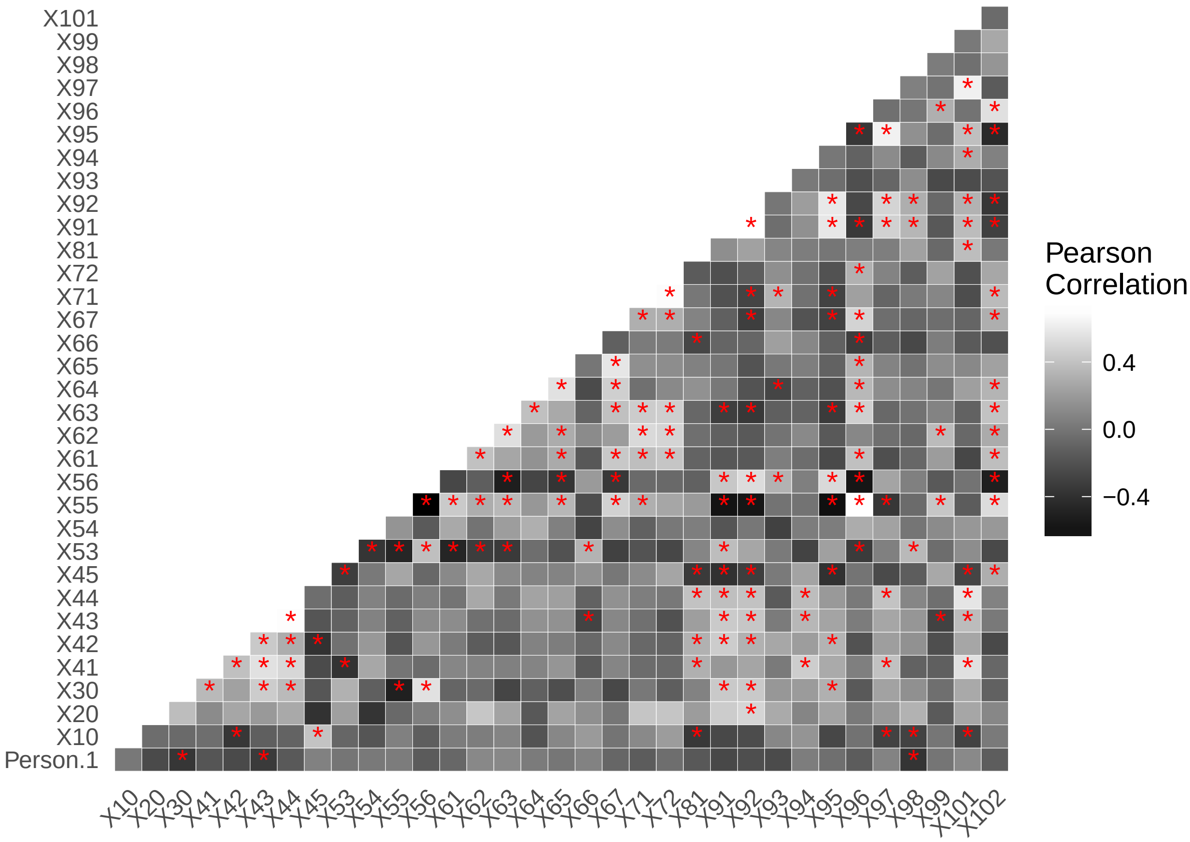
	Question	ChainR	ChainP	ChainPs
X10	X10	0.02	0.91	FALSE
X20	X20	-0.25	0.33	FALSE
X30	X30	-0.38	0.01	TRUE
X41	X41	-0.19	0.21	FALSE
X42	X42	-0.25	0.08	FALSE
X43	X43	-0.35	0.02	TRUE
X44	X44	-0.16	0.28	FALSE
X45	X45	0.06	0.67	FALSE
X53	X53	0.00	0.99	FALSE
X54	X54	0.02	0.91	FALSE

	Question	ChainR	ChainP	ChainPs
X55	X55	0.04	0.80	FALSE
X56	X56	-0.15	0.34	FALSE
X61	X61	-0.08	0.59	FALSE
X62	X62	0.10	0.50	FALSE
X63	X63	0.10	0.49	FALSE
X64	X64	0.03	0.83	FALSE
X65	X65	0.00	0.99	FALSE
X66	X66	0.07	0.64	FALSE
X67	X67	-0.09	0.55	FALSE
X71	X71	-0.14	0.32	FALSE
X72	X72	-0.04	0.76	FALSE
X81	X81	-0.17	0.25	FALSE
X91	X91	-0.27	0.06	FALSE
X92	X92	-0.23	0.13	FALSE
X93	X93	-0.25	0.09	FALSE
X94	X94	0.04	0.77	FALSE
X95	X95	-0.04	0.79	FALSE
X96	X96	-0.14	0.34	FALSE
X97	X97	0.07	0.62	FALSE
X98	X98	-0.39	0.01	TRUE
X99	X99	0.00	1.00	FALSE
X101	X101	0.11	0.46	FALSE
X102	X102	-0.14	0.33	FALSE

Note on years: interesting relations of experience

- more experience, less important was the missing of tactical interactions with chains (Q30)
- more experience, less was the game perceived influence of the rule change (Q43)
- more experience, less faster/wilder/stressful (Q98)

```
library(reshape2)
library(ggplot2)

chain.stats.cor$r[upper.tri(chain.stats.cor$r, diag = TRUE)] <- NA
chain.stats.cor <- chain.stats.cor %$% inner_join(melt(r), melt(p), by = c("Var1" = "Var1", "Var2" = "V

ggplot(chain.stats.cor, aes(Var1, Var2, fill = r)) + geom_tile(colour = "white") +
  geom_text(aes(Var1, Var2, label = ifelse(p < 0.05, "*", "")), color = "red", size = 4) +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1), axis.title.x = element_blank(),
        axis.title.y = element_blank(), panel.grid.major = element_blank(), panel.border = element_blan
  scale_fill_gradient(low = "black", high = "white", name = "Pearson\nCorrelation")
```




```
# red dots mark significant correlations
```

It is a bit confusing - many variables (34 variables resulting in 561 unique correlations) and many significant correlations (147!). Lets try something else more visually convenient: Let's treat the correlation matrix as a adjacency matrix (thresholded by $p < 0.05$)

Also note: The high number of significant correlations is a hint that the items are not meassuring independent components, but rather rely on a few latent dimensions. Lets test this first!

```
chain.stats <- Testergebnis.Fragebogen %>% select(-Person.1, -P2.Q,-P2.Stab,-P2.LP,-P2.Schild,-P2.Quick

factors <- prcomp(formula = ~ ., data = chain.stats, center = TRUE, scale = TRUE, retx = TRUE, na.actio

kable(x = data.frame(PC = colnames(factors$rotation), Standard.deviations = round(factors$sdev,2)), dig
      caption = "Standard deviations of all primary components of the questionnaire data set.")
```

Table 5: Standard deviations of all primary components of the questionnaire data set.

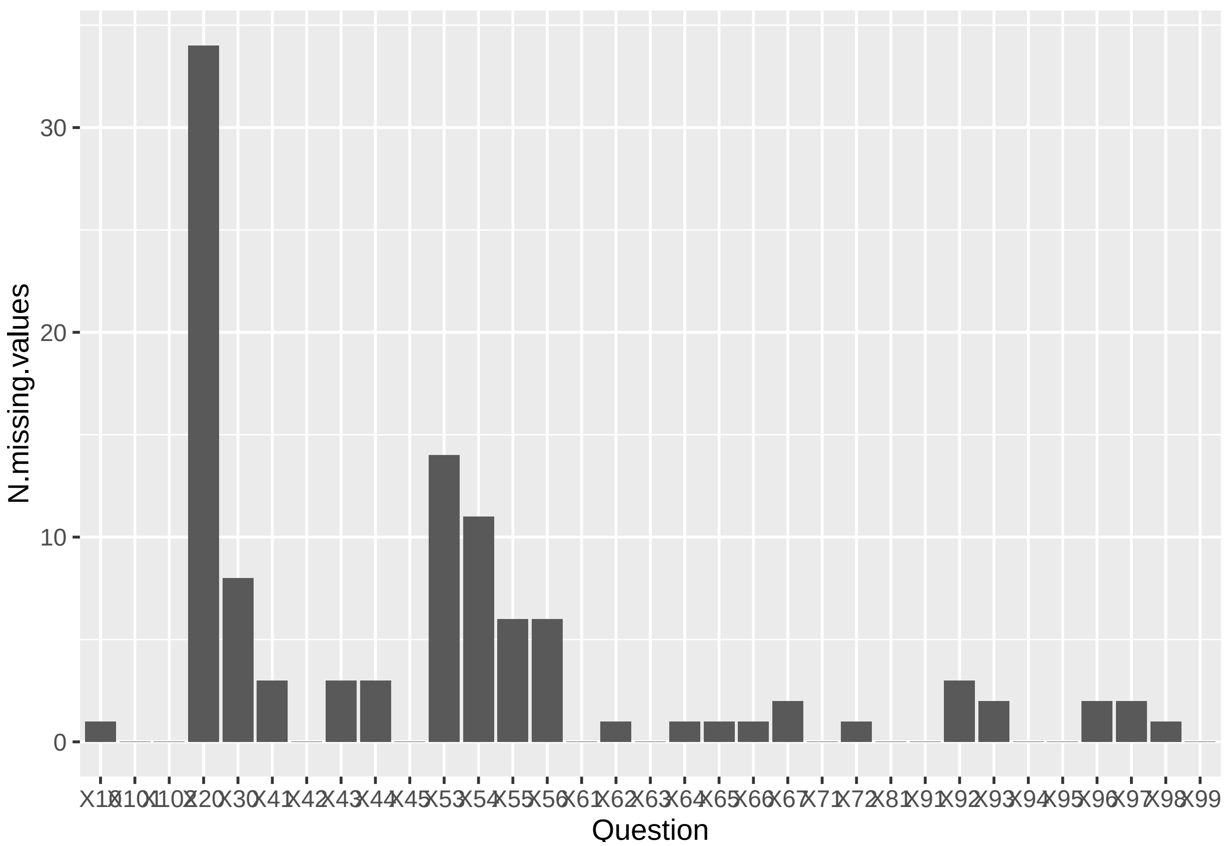
PC	Standard.deviations
PC1	2.90
PC2	2.76
PC3	2.59
PC4	2.06
PC5	1.73
PC6	1.56

PC	Standard.deviations
PC7	0.74
PC8	0.00

```
#kable(x = as.data.frame(factors$rotation), digits = 2,
#      caption = "Rotations matrix of the primary factor analysis. Shows decomposition of the factors i
```

Only eight components are within this one data set and the last one is effectively zero, which reduces the number of componentes to 7. At first this seems like an interesting result, but hold on! How many rows of the result table are without any missing values? Exactly 8! Well that explains a lot ... But lets have a look at the questions without answers: It turns out that question 20 askes only chain players (and thus 34 missing values), 30 asks only Pompfers (8 mising values - discrepancies with the numbers of the question Person2 are due to the fact that Person2 asks only for the main Pompfe, while Q20 and Q30 apply also for secondary and tertiary Pompfe), 53 and 54 are only applicable if a chain-double occured to that player (14 and 11 missing values). This is good news, since subjects read and followed the instructions :) So next take on primary components:

```
library(ggplot2)
ggplot(data = data.frame(Question = colnames(chain.stats), N.missing.values = colSums(is.na(chain.stats)
```



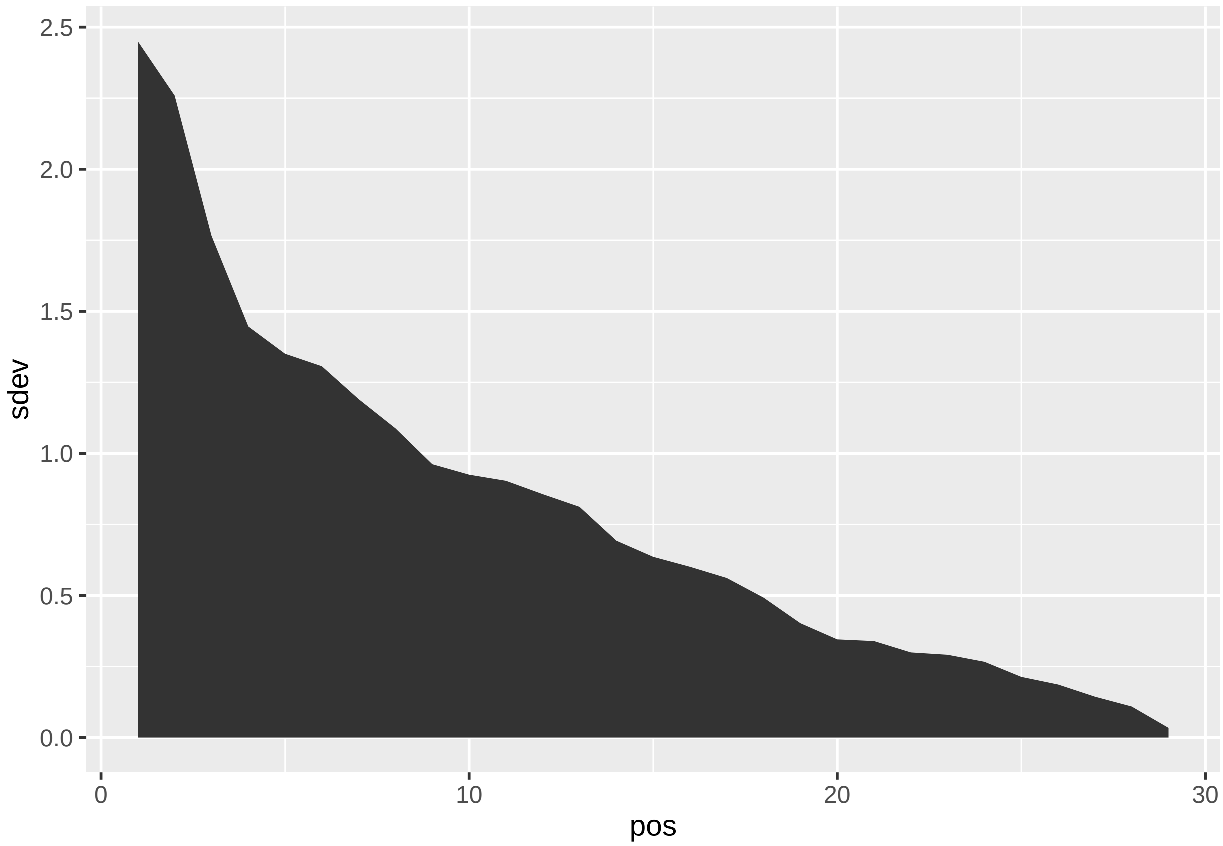

```
chain.stats <- Testergebnis.Fragebogen %>% select(-Person.1, -P2.Q,-P2.Stab,-P2.LP,-P2.Schild,-P2.Quick

factors <- prcomp(formula = ~ ., data = chain.stats, center = TRUE, scale = TRUE, retx = TRUE, na.actio
```

```
kable(x = data.frame(PC = colnames(factors$rotation), Standard.deviations = round(factors$sdev,2)), dig
      caption = "Standard deviations of all primary components of the questionnaire data set.")
```

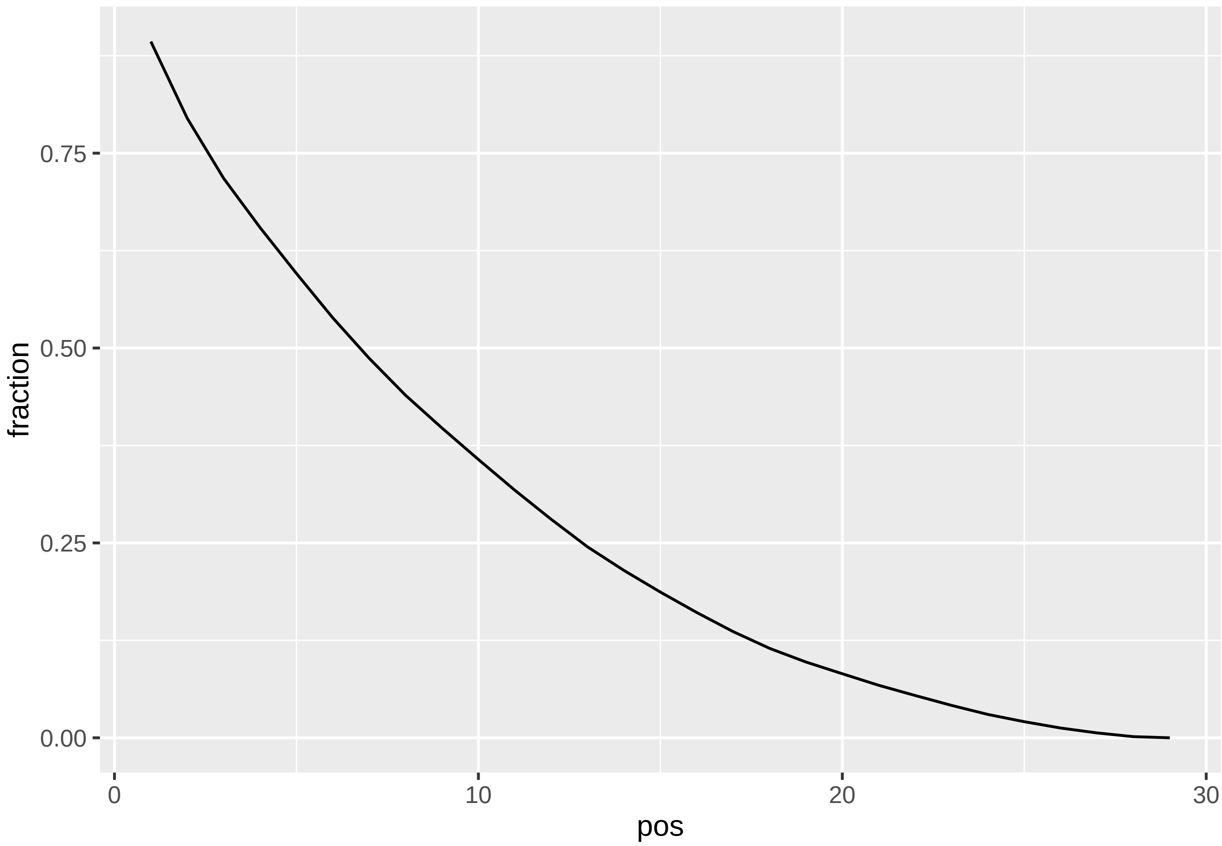
Table 6: Standard deviations of all primary components of the questionnaire data set.

PC	Standard.deviations
PC1	2.45
PC2	2.26
PC3	1.77
PC4	1.45
PC5	1.35
PC6	1.31
PC7	1.19
PC8	1.09
PC9	0.96
PC10	0.93
PC11	0.90
PC12	0.86
PC13	0.81
PC14	0.69
PC15	0.64
PC16	0.60
PC17	0.56
PC18	0.49
PC19	0.40
PC20	0.35
PC21	0.34
PC22	0.30
PC23	0.29
PC24	0.27
PC25	0.21
PC26	0.19
PC27	0.14
PC28	0.11
PC29	0.03

```
ggplot(data = data.frame(pos = 1:length(colnames(factors$rotation)),PC = colnames(factors$rotation), sd
```




```
ggplot(data = data.frame(pos = 1:length(colnames(factors$rotation)),PC = colnames(factors$rotation), fra
```




```
kable(x = as.data.frame(factors$rotation), digits = 2,
      caption = "Rotations matrix of the primary factor analysis. Shows decomposition of the factors int
```

Table 7: Rotations matrix of the prim
decomposition of the factors into the s

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15
X10	0.06	-0.25	0.34	0.00	0.06	-0.22	0.14	-0.26	0.07	0.13	-0.13	-0.09	0.10	0.03	0.00
X41	-0.05	0.25	0.23	-0.20	0.25	-0.09	-0.24	-0.13	-0.04	-0.08	-0.19	-0.16	0.02	-0.24	0.12
X42	-0.10	0.23	0.13	0.23	0.05	-0.13	-0.15	0.01	-0.05	0.40	-0.03	0.14	0.59	-0.11	0.13
X43	-0.09	0.31	0.16	0.09	0.19	-0.19	0.15	-0.02	-0.04	-0.29	-0.06	0.20	-0.20	-0.10	0.24
X44	-0.11	0.30	0.23	-0.09	0.14	0.03	0.17	0.12	-0.04	-0.12	0.18	0.06	-0.15	0.29	0.14
X45	0.16	-0.23	0.14	-0.14	0.07	-0.11	0.29	0.30	-0.33	0.08	-0.04	-0.25	0.03	0.27	-0.04
X55	0.34	0.10	-0.10	-0.07	0.21	-0.04	0.05	-0.10	0.18	0.08	0.05	-0.11	-0.01	0.23	0.09
X56	-0.27	-0.12	0.14	0.07	-0.08	0.05	0.16	0.13	-0.28	0.15	0.13	0.01	-0.39	-0.45	0.15
X61	0.24	0.13	0.14	0.01	-0.01	-0.05	0.24	-0.45	-0.17	-0.16	-0.04	0.06	-0.02	-0.02	-0.39
X62	0.17	0.10	0.26	-0.18	-0.16	0.29	0.12	0.11	0.04	-0.30	0.12	-0.33	0.12	-0.12	0.05
X63	0.25	0.14	0.07	0.03	-0.23	0.16	-0.25	0.06	-0.18	-0.24	-0.10	-0.18	0.16	0.01	0.33
X64	0.10	0.23	-0.19	-0.07	-0.28	-0.24	0.05	0.27	-0.38	0.14	0.04	0.08	-0.04	-0.13	-0.19
X65	0.13	0.16	0.12	-0.07	-0.27	-0.45	0.04	0.06	0.34	-0.08	0.00	-0.13	-0.02	-0.21	-0.33
X66	0.00	-0.24	0.26	-0.15	-0.19	0.01	-0.08	0.19	0.48	0.00	0.07	0.31	-0.08	0.06	0.19
X67	0.19	0.23	0.06	0.13	-0.20	-0.30	-0.05	0.08	0.01	0.08	0.39	-0.09	-0.03	0.10	0.29
X71	0.15	0.11	0.31	0.28	-0.04	0.35	-0.10	0.07	0.12	-0.01	0.01	0.18	-0.12	-0.20	-0.23
X72	0.16	-0.01	0.29	0.09	-0.23	0.34	-0.05	-0.20	-0.25	0.28	0.11	0.09	0.03	0.16	-0.05
X81	-0.02	0.22	-0.13	0.20	0.37	0.23	0.02	0.12	0.17	0.16	0.33	-0.32	-0.02	-0.12	-0.19
X91	-0.27	0.17	0.09	0.07	-0.26	-0.01	0.24	0.05	0.12	-0.02	0.17	0.17	0.21	0.17	-0.18

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15
X92	-0.30	0.10	0.11	0.04	0.05	0.05	0.33	-0.27	-0.01	-0.03	0.10	0.01	0.08	0.12	0.09
X93	0.02	-0.08	0.27	0.42	-0.01	-0.14	-0.03	0.05	0.12	0.28	-0.18	-0.37	-0.32	0.10	0.01
X94	0.02	0.03	0.36	-0.25	0.32	-0.08	0.00	0.34	-0.11	0.14	-0.19	0.17	0.11	0.03	-0.09
X95	-0.32	0.08	0.04	-0.12	-0.21	0.02	-0.07	-0.18	0.02	0.04	-0.27	-0.28	0.05	-0.15	0.04
X96	0.26	0.21	-0.13	-0.04	0.00	-0.06	0.02	-0.27	0.01	0.27	-0.07	0.20	-0.31	0.04	0.25
X97	-0.23	0.19	0.03	-0.18	-0.26	0.08	-0.24	-0.05	-0.05	0.09	-0.19	-0.05	-0.25	0.36	-0.09
X98	-0.08	0.15	-0.17	0.25	-0.15	0.12	0.43	0.17	0.15	-0.01	-0.40	-0.15	0.11	0.12	0.17
X99	0.13	0.00	-0.05	-0.44	-0.12	0.16	0.35	-0.10	0.15	0.36	0.08	-0.06	0.05	-0.30	0.20
X101	-0.15	0.28	-0.01	-0.33	0.06	0.12	-0.13	0.08	0.12	0.25	0.01	-0.11	-0.11	0.15	-0.21
X102	0.25	0.18	-0.08	0.06	0.03	0.19	0.16	0.20	0.09	0.05	-0.44	0.24	-0.06	-0.09	-0.06

The resulting 29 questions produce a typical decrease in primary component power. It shows that i can drop 15 factors (of 29) and still only loos less than 20% of the variance within the data. This serves as a minor hint that there might be some hidden structure in the data, although it is weak since the steepness of the decrease is not that radical. But as a remninder keep in mind the highest values of the first component: Q55, Q95, Q92, Q56, Q91, Q96, Q102, Q63, Q61, Q97, . . .

So much for that. Now back to the graph analysis. As a reminder: I wanted to highlight some structures of the correelation matrix.

```
# library(visNetwork)

chain.stats.cor.nw <- chain.stats.cor %>% filter(Var1 != "Person.1", Var2 != "Person.1", p < 0.05)

# cv <- chain.stats.cor.nw %$% visNetwork(nodes = data.frame(id = unique(chain.stats.cor$Var1),
#                                                          label = unique(chain.stats.cor$Var1),
#                                                          title = paste("<b>Question", unique(chain.
#                                                                        "Fill in question texts late
#                   edges = data.frame(to = Var1, from = Var2,
#                                      title = paste("<b>",Var1,"~",Var2,"</b></br>", "r =",  r),
#                                      value = abs(r))) %>%
#   visOptions(highlightNearest = TRUE, nodesIdSelection = TRUE)
#
# cv
```

Still a lot to digest within one picture - a dense strongly interconnected graph. But look: It's interactive!!! This neat feature allows you to explore the structure by hand. Just hover, click, drag and drop.

However i still have the feeling that this visualisation doesn't tell me everything - still it gives me a very good intuition of whats going on. So lets drill in with some graph methods.

```
library(igraph) # just for the convenient implementation of largest.clique
cvig <- chain.stats.cor.nw %$% graph.data.frame(data.frame(to = Var1, from = Var2), directed = FALSE)
largest.cliques(cvig)
```

```
## [[1]]
## + 7/33 vertices, named:
## [1] X96  X55  X56  X63  X67  X95  X102
##
## [[2]]
## + 7/33 vertices, named:
## [1] X96  X55  X56  X63  X91  X95  X102
```

```
##
## [[3]]
## + 7/33 vertices, named:
## [1] X92  X71  X55  X95  X102 X63  X67
##
## [[4]]
## + 7/33 vertices, named:
## [1] X92  X67  X55  X56  X95  X102 X63
##
## [[5]]
## + 7/33 vertices, named:
## [1] X92  X91  X95  X55  X56  X102 X63
```

So it seems that there are five cliques of size seven. But those cliques are comprised of only only 10 nodes (Q95, Q96, Q56, Q63, Q55, Q102, Q91, Q92, Q71, Q67) and 4 of them contained within all cliques (Q55, Q63, Q95, Q102). An surprisingly there is a huge overlap between those nodes and the questions forming the first principal component of the above factor analysis. Let's do one last statistic to finish this topic.

```r
# note: im not importing psych here, since the select of psych overwrites the select of dplyr
psych::alpha(chain.stats %>% select(X55,X63,X95,X102))
```

```
## Warning in psych::alpha(chain.stats %>% select(X55, X63, X95, X102)): Some items were negatively cor
## should be reversed.
## To do this, run the function again with the 'check.keys=TRUE' option
```

```
## Some items ( X95 ) were negatively correlated with the total scale and
## probably should be reversed.
## To do this, run the function again with the 'check.keys=TRUE' option
```

```
##
## Reliability analysis
## Call: psych::alpha(x = chain.stats %>% select(X55, X63, X95, X102))
##
##   raw_alpha std.alpha G6(smc) average_r    S/N  ase mean   sd
##      -0.11    -0.077    0.32    -0.018 -0.071 0.22  3.3 0.62
##
##  lower alpha upper     95% confidence boundaries
## -0.55 -0.11 0.33
##
##  Reliability if an item is dropped:
##      raw_alpha std.alpha G6(smc) average_r   S/N alpha se
## X55      -0.54     -0.54  -0.038     -0.13 -0.35    0.361
## X63      -0.92     -0.79   0.077     -0.17 -0.44    0.454
## X95       0.69      0.70   0.619      0.43  2.31    0.075
## X102     -0.98     -1.02  -0.152     -0.20 -0.50    0.451
##
##  Item statistics
##       n raw.r std.r r.cor r.drop mean  sd
## X55  46  0.62  0.66  0.69   0.18  3.1 1.3
## X63  52  0.76  0.72  0.55   0.27  3.2 1.4
## X95  52 -0.16 -0.21 -0.91  -0.59  3.2 1.3
## X102 52  0.72  0.77  0.78   0.36  3.5 1.2
##
```

```
## Non missing response frequency for each item
##         1    2    3    4    5 miss
## X55  0.13 0.15 0.35 0.20 0.17 0.12
## X63  0.19 0.12 0.21 0.27 0.21 0.00
## X95  0.15 0.13 0.23 0.29 0.19 0.00
## X102 0.06 0.17 0.23 0.27 0.27 0.00
```

```
psych::alpha(chain.stats %>% select(X55,X63,X95,X102,X96,X56,X91,X92,X71,X67))
```

```
## Warning in psych::alpha(chain.stats %>% select(X55, X63, X95, X102, X96, : Some items were negatively
## should be reversed.
## To do this, run the function again with the 'check.keys=TRUE' option

## Some items ( X95 X56 X91 X92 ) were negatively correlated with the total scale and
## probably should be reversed.
## To do this, run the function again with the 'check.keys=TRUE' option

##
## Reliability analysis
## Call: psych::alpha(x = chain.stats %>% select(X55, X63, X95, X102,
##     X96, X56, X91, X92, X71, X67))
##
##   raw_alpha std.alpha G6(smc) average_r  S/N  ase mean   sd
##      0.064       0.1    0.64     0.011 0.11 0.21  3.1 0.43
##
##  lower alpha upper     95% confidence boundaries
## -0.34 0.06 0.47
##
##  Reliability if an item is dropped:
##      raw_alpha std.alpha G6(smc) average_r      S/N alpha se
## X55      0.112   0.13104    0.59   1.6e-02  0.15080     0.20
## X63     -0.027  -0.00553    0.56  -6.1e-04 -0.00550     0.23
## X95      0.186   0.23585    0.69   3.3e-02  0.30865     0.17
## X102    -0.024  -0.00043    0.61  -4.7e-05 -0.00043     0.23
## X96     -0.145  -0.11342    0.52  -1.1e-02 -0.10187     0.26
## X56      0.310   0.32468    0.68   5.1e-02  0.48078     0.15
## X91      0.101   0.14878    0.62   1.9e-02  0.17478     0.18
## X92      0.109   0.17578    0.64   2.3e-02  0.21326     0.18
## X71     -0.103  -0.08194    0.55  -8.5e-03 -0.07574     0.25
## X67     -0.125  -0.09492    0.58  -9.7e-03 -0.08669     0.25
##
##  Item statistics
##       n raw.r  std.r  r.cor r.drop mean  sd
## X55  46  0.23  0.275  0.269 -0.064  3.1 1.3
## X63  52  0.45  0.460  0.420  0.129  3.2 1.4
## X95  52  0.15  0.094 -0.063 -0.175  3.2 1.3
## X102 52  0.41  0.454  0.373  0.143  3.5 1.2
## X96  50  0.58  0.578  0.606  0.275  3.1 1.4
## X56  46 -0.05 -0.097 -0.228 -0.367  2.9 1.4
## X91  52  0.28  0.247  0.184 -0.040  2.9 1.4
## X92  49  0.26  0.202  0.130 -0.057  2.7 1.3
## X71  52  0.50  0.546  0.486  0.272  3.3 1.1
## X67  50  0.53  0.559  0.484  0.282  2.9 1.2
```

```
##
## Non missing response frequency for each item
##         1    2    3    4    5 miss
## X55  0.13 0.15 0.35 0.20 0.17 0.12
## X63  0.19 0.12 0.21 0.27 0.21 0.00
## X95  0.15 0.13 0.23 0.29 0.19 0.00
## X102 0.06 0.17 0.23 0.27 0.27 0.00
## X96  0.18 0.16 0.18 0.30 0.18 0.04
## X56  0.22 0.22 0.22 0.17 0.17 0.12
## X91  0.23 0.17 0.15 0.33 0.12 0.00
## X92  0.22 0.24 0.22 0.22 0.08 0.06
## X71  0.10 0.13 0.25 0.40 0.12 0.00
## X67  0.18 0.14 0.34 0.24 0.10 0.04
```

So it seems that the 4 core questions of all maximal cliques are asking information from the same source-construct, while all 10 involved in the cliques do not constitut a single source. Some more exploration might be possible, but lets keep it at that for now.