Questionnaire results

Eduard Szöcs

November 2, 2015

Results from the questionnaire

This is just a Quick and Dirty analysis...

Load the data

```
prj <- '/home/edisz/Documents/Uni/Projects/PHD/CONFERENCES/AufLand'
df <- read.csv(file.path(prj, 'questionnaire/questionnaire.csv'),
                 sep = ',', header = TRUE, quote = "\"",
               stringsAsFactors = FALSE)
```

```
## 'data.frame':    11 obs. of  18 variables:
##  $ Timestamp                                                                    : chr  "
##  $ How.would.you.categorize.your.theoretical.statistical.knowledge.             : chr  "
##  $ Do.you.have.any..irrespective.of.language..programming.experience.           : chr  "
##  $ How.would.you.categorize.your.R.programing.knowledge.                        : chr  "
##  $ How.do.you.currently.do.your.statistical.analyses.                           : chr  "
##  $ Which.of.these.topics.is.the.most.interesting.for.you.                       : chr  "
##  $ Rank.the.five.most.interesting.topics....1..most.interesting..               : chr  "
##  $ Rank.the.five.most.interesting.topics....2.                                  : chr  "
##  $ Rank.the.five.most.interesting.topics....3.                                  : chr  "
##  $ Rank.the.five.most.interesting.topics....4.                                  : chr  "
##  $ Rank.the.five.most.interesting.topics....5..least.interesting..              : chr  "
##  $ Do.you.have.any.other.topic.suggestions.                                     : chr  "
##  $ Do.you.have.the.necessary.R.and.theoritcal.background.for.the.topic.you.re.interested.in.: chr  "
##  $ Name                                                                         : chr  "
##  $ Email                                                                        : chr  "
##  $ Affiliation                                                                  : chr  "
##  $ Position                                                                     : chr  "
##  $ I.will.bring.my.own.laptop                                                   : chr  "
```

Looks OK, but quite long column names....

Clean the data

```
# shorten the colnames
colnames(df) <- c('date', 'theory', 'prog', 'rknow', 'curr', 'inter', 'one',
                  'two', 'three', 'four', 'five', 'sugg', 'backg', 'name',
                  'email', 'affl', 'pos', 'laptop')
str(df)
```

```
## 'data.frame':    11 obs. of  18 variables:
##  $ date  : chr  "12/10/2015 12:23:49" "12/10/2015 12:23:53" "12/10/2015 13:37:45" "12/10/2015 14:02:4
##  $ theory: chr  "intermediate (e.g. I'm familiar with GLMs, model comparisons, PCA)" "intermediate (
##  $ prog  : chr  "no" "no" "yes" "no" ...
##  $ rknow : chr  "intermediate (e.g. I know data structures and data types of R and can aggregate & pl
##  $ curr  : chr  "I use solely R for my analyses (clean, plot, model)" "I use solely R for my analyses
##  $ inter : chr  "Generlized linear models (GLMs) with R (from basic regression, to count/binomial da
##  $ one   : chr  "Generlized linear models (GLMs) with R (from basic regression, to count/binomial da
##  $ two   : chr  "Creating publication ready plots with ggplot2 (An introduction to the ggplot2 packag
##  $ three : chr  "Multivariate Statistics using the vegan package" "Multivariate Statistics using the
##  $ four  : chr  "Introduction to R programming (data types & structures, writing functions, repetitiv
##  $ five  : chr  "Introduction to R for ecologists (Basic R, How to read data, clean and aggregate, pl
##  $ sugg  : chr  "" "" "" "" ...
##  $ backg : chr  "Maybe (I am unsure about my theory knowledge, but R is good)" "Maybe (I am unsure ab
##  $ name  : chr  "Steffi" "Anna" "Friedrich" "Bonny" ...
##  $ email : chr  "@uni-landau.de" "@uni-landau.de" "@uni-landau.de" "@uni-landau.de" ...
##  $ affl  : chr  "Uni Landau" "Uni Landau" "Universität Koblenz-Landau " "University Koblenz Landau"
##  $ pos   : chr  "PhD" "PhD student" "" "PhD" ...
##  $ laptop: chr  "yes" "yes" "no" "yes" ...
```

From the `affl` column, we see that the Uni-Landau was entered in different ways. Lets create a new column coding if the attendee comes from Landau.

We can use columns `email` and `affl` to check if an attendee come form landau. We use regular expressions to match the word 'landau' in both columns. Note, that I compare only lowercase strings, as this saves 2 comparisons.

Character vectors can be translated to lowercase using `tolower()`:

```
tolower(df$affl)
```

```
##  [1] "uni landau"                      "uni landau"
##  [3] "universität koblenz-landau "     "university koblenz landau"
##  [5] ""                                "wassercluster lunz"
##  [7] "csic"                            "uni landau"
##  [9] "university of basel"             "uni landau, molecular ecology "
## [11] "uni landau"
```

We can check if a string contains `'landau'` using the function `grepl()` (= global regular expression, l stands for logical which is the output):

```
grepl(pattern = 'landau', df$email)
```

```
##  [1]  TRUE  TRUE  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE  TRUE  TRUE
```

I a last step we combine both using the logical `OR` and save as new variable:

```
df$fromlandau <- grepl(pattern = 'landau', df$email) | grepl(pattern = 'landau', tolower(df$affl))
```

Plot the data

How many attendies replied and where are they from?

We have only 11 repsonses:

```
nrow(df)
```

```
## [1] 11
```

from which the majority

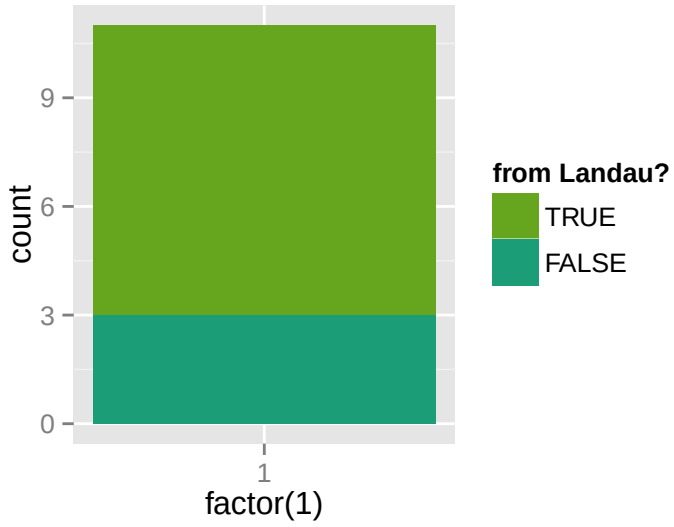
```
sum(df$fromlandau)
```

```
## [1] 8
```

are from Landau.

```
require(ggplot2)
require(RColorBrewer)
classPalette = colorRampPalette(brewer.pal(5, "Dark2"))
```

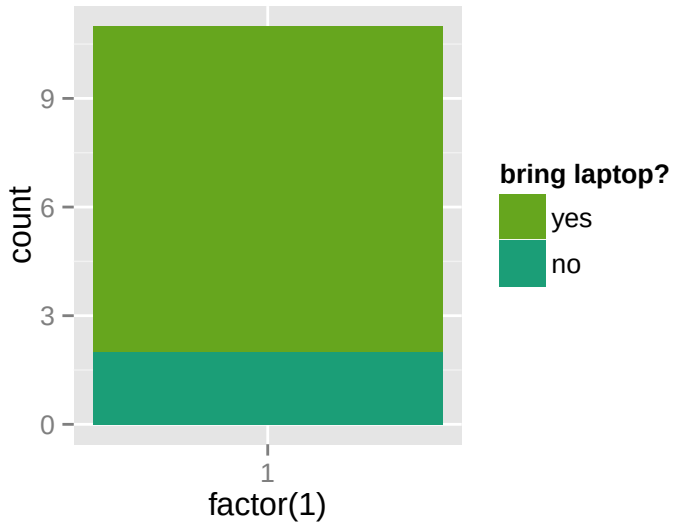
```
nlev <- length(unique(df$fromlandau))
ggplot(df, aes(x = factor(1), fill = df$fromlandau)) +
  geom_bar() +
  scale_fill_manual(values = classPalette(nlev)) +
  guides(fill= guide_legend(reverse=TRUE, title = 'from Landau?'))
```



What room do we need?

Note that we need a computer room, as two attendees won't bring their laptop. Or we provide laptops?!

```
nlev <- length(unique(df$laptop))
ggplot(df, aes(x = factor(1), fill = df$laptop)) +
  geom_bar() +
  scale_fill_manual(values = classPalette(nlev)) +
  guides(fill= guide_legend(reverse=TRUE, title = 'bring laptop?'))
```

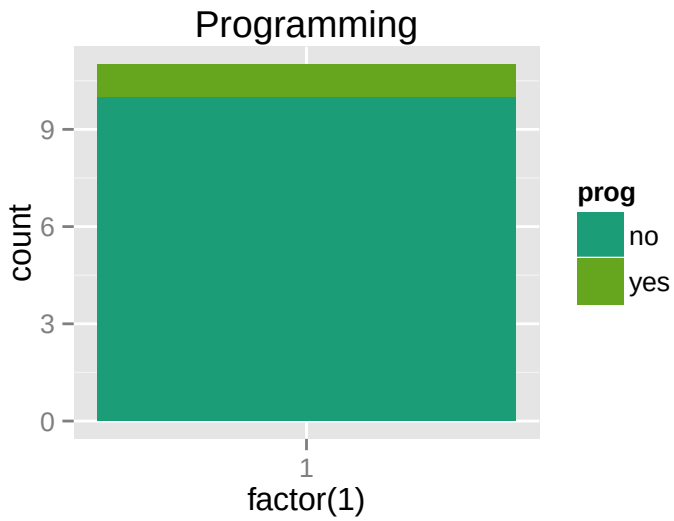



```
sum(df$laptop == 'no') / nrow(df) * 100
```

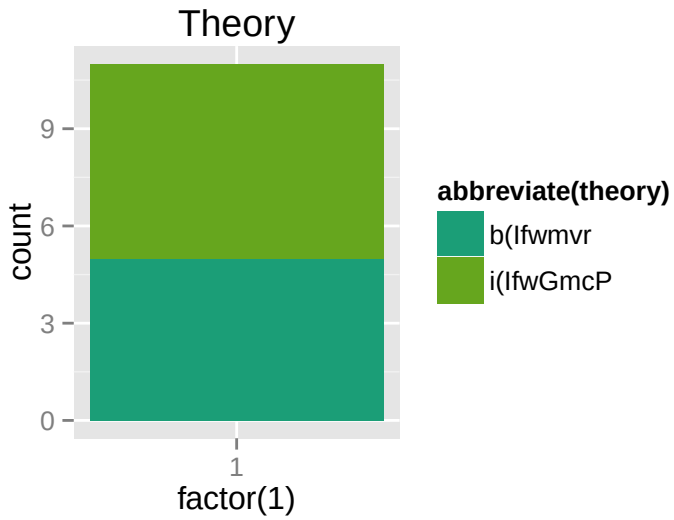
```
## [1] 18.18182
```

What's the backgound knowledge?

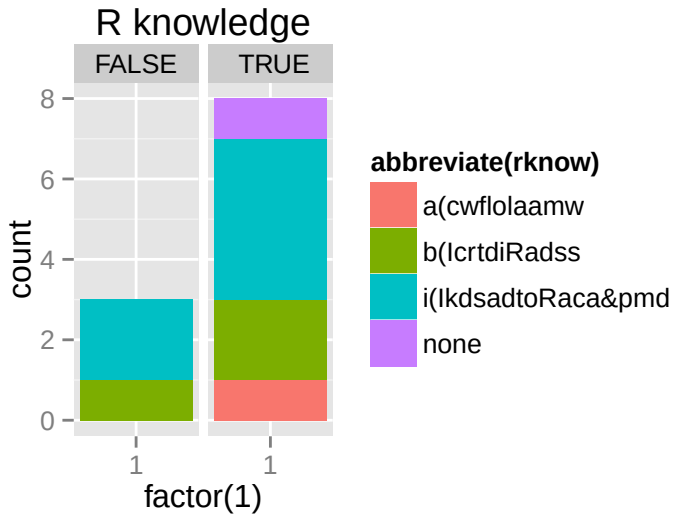
```
nlev <- length(unique(df$prog))
ggplot(df, aes(x = factor(1), fill = prog)) +
  geom_bar() +
  scale_fill_manual(values = classPalette(nlev)) +
  ggtitle('Programming')
```



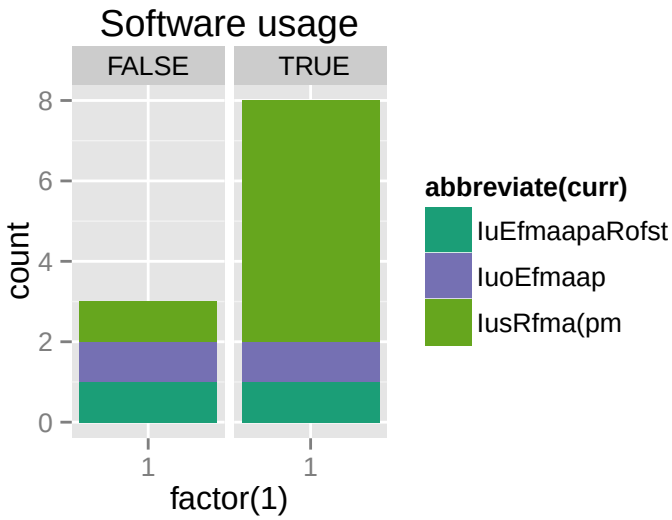

```
nlev <- length(unique(df$theory))
ggplot(df, aes(x = factor(1), fill = abbreviate(theory))) +
  geom_bar() +
  scale_fill_manual(values = classPalette(nlev)) +
  ggtitle('Theory')
```




```
nlev <- length(unique(df$rknow))
ggplot(df, aes(x = factor(1), fill = abbreviate(rknow))) +
  geom_bar() +
  ggtitle('R knowledge') +
  facet_wrap(~fromlandau)
```




```
nlev <- length(unique(df$curr))
ggplot(df, aes(x = factor(1), fill = abbreviate(curr))) +
  geom_bar() +
  scale_fill_manual(values = classPalette(nlev)) +
  ggtitle('Software usage') +
  facet_wrap(~fromlandau)
```


Software usage
FALSE
TRUE
8
6
4
2
0
count
1
1
factor(1)
abbreviate(curr)
luEfmaapaRofst
luoEfmaap
lusRfma(pm