

# VoQs: A Web Application for Visualization of Questionnaire Surveys

Xiaowei Zhang<sup>1</sup>, Frank Klawonn<sup>1,2</sup>, Lorenz Grigull<sup>3</sup>, and Werner Lechner<sup>4</sup>

<sup>1</sup> Helmholtz Center for Infection Research  
Inhoffenstrasse 7, 38124 Braunschweig, Germany

{Frank.Klawonn, Xiaowei.Kortum}@helmholtz-hzi.de

<sup>2</sup> Department of Computer Science, Ostfalia University of Applied Sciences  
Salzdahlumer Str. 46/48, 38302 Wolfenbuettel, Germany

f.klawonn@ostfalia.de

<sup>3</sup> Department of Paediatric Haematology and Oncology,  
Medical University Hanover

Carl-Neuberg Str.1, 30625 Hannover, Germany

grigull.lorenz@mh-hannover.de

<sup>4</sup> Improved Medical Diagnostics IMD GmbH,  
Ostfeldstr. 25, Hannover 30559, Germany

werner@improvedmedicaldiagnostics.com

**Abstract.** This paper is motivated by analyzing questionnaire data that collected from patients who suffer from an orphan disease. In order to decrease misdiagnoses and shorten the diagnosis time for people who have not been diagnosed yet but already have a long history of health problems, a research project about using questionnaire mode and data analysis methods to predetermine orphan disease has been set up and questionnaires were designed based on experiences from patients who already have a diagnosis.

The main focus of this work is to visualize answering patterns that characterize patients with a specific orphan disease, which questions are most useful to distinguish between certain orphan diseases and how well an answering pattern of a specific patient fits to the general pattern of those patients who share the same disease.

We borrow from the concept of sequence logos, commonly used in genetics to visualize the conservation of nucleotides in a strand of DNA or RNA. Instead of nucleotides, we have possible answers from a question. Our proposed visualization techniques are not limited to questionnaires on orphan diseases but also can be applied to any questionnaire survey with closed-ended questions for which we are interested in answering characteristics of different groups.

## 1 Introduction

Questionnaire surveys, which enable researchers to obtain and gather information directly from participants, can be used for measuring participants' knowledge, preferences, attitudes or personal beliefs, also for exploring facts, experiences and present situations. [7] Many questionnaires are using "close-ended

questions”. Possible answers of such questions are limited to a categorical or ordinal scale, e.g. the Likert scale (yes/no; satisfied/unsatisfied). Instead, “open-ended questions” offer respondents to express opinions, feelings or suggestions without being influenced by the researcher, which can help to obtain responses that individuals give spontaneously. “Close-ended questions” are often researchers’ first choice because they limit the respondent choices to a fixed set of finite alternatives that are offered. The ordered options can not only let respondents easily choose one appropriate position in a series of approval degrees for a questions but also make the statistical analysis of such questionnaires much easier [9], [3] .

Although a sophisticated statistical analysis of such questionnaires can provide a deeper insight, the domain expert for the questionnaires might not be familiar with the statistical methods and may have difficulties to interpret the results correctly. Therefore, we have developed visualization techniques, called VoQs, which helps domain experts to identify interesting, characteristic and also deviating answering patterns as well as questions of high interest.

Before we describe VoQs in more detail, Section 2 reviews the original motivation and background to develop this application. Section 3 introduces the core technology of VoQs followed by preliminary results, including the principal statistics and computational methods used in VoQs. Section 4 presents all functionalities and important characteristics of this application and the visualization aspects with graphical results. Finally, we briefly outline further extensions of VoQs.

## 2 The Origin of VoQs

The motivation of developing VoQs application originated from questionnaires we had developed for patients who might suffer from an orphan disease. Because rare diseases are often overlooked or misdiagnosed, a clear statement of a final diagnosis and medical treatment is often linked to reviewing the patient’s history and the trend of the patient’s constitution [6]. For our questionnaires it was of interest to visualize how patients with different orphan diseases differ in their answer patterns, which questions are most useful to distinguish between certain orphan diseases and how well an answering pattern of a specific patient fits to the general pattern of those patients who share the same disease. Although we had already introduced visualization techniques based on multidimensional scaling that helps to discover outliers and could explain how well different diseases could be distinguished [2], these visualizations were still too abstract and provided only an overall view on the data without focus on specific questions.

Individual questionnaires were set up for different groups of orphan diseases, such as bleeding disorders, pulmonological and neuromuscular diseases. Patients data were collected through these anonymous questionnaires, which contain, for instance, information about noticed symptoms, physical constitution in general and specific sports activities. All questions are close-ended with four ordered options from “No” to “Absolutely”.

As an example, we focus on the questionnaire for neuromuscular diseases. So far, it contains data from 366 persons with 13 different diagnoses (e.g. amyotrophic lateral sclerosis (ALS), Morbus Pompe). To see how much variation the answering pattern of a person from a specific diagnosis shows, one could simply use a bar chart for each question to visualize the absolute or relative frequencies for each possible answer. Figure 1 shows the result for diagnosis 2 (Morbus Pompe). Questions showing a very uniform answer pattern cannot be considered to be characteristic for the corresponding disease, whereas questions for which (almost) all patients choose the same answer might be of more interest – unless people from all other diagnoses also choose the same answer. Such more interesting questions for Morbus Pompe are questions *Q3* and *Q4* where patients tend to choose option 4 (Absolutely) or questions *Q29* and *Q33* where option 1 (No) is favored.

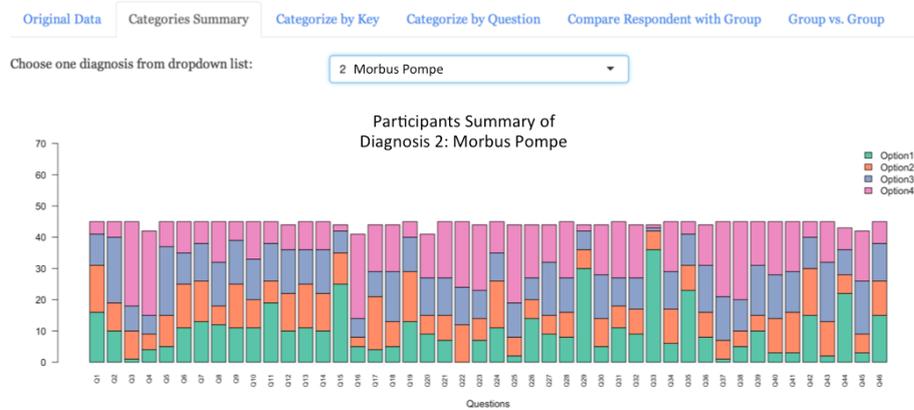


Fig. 1: Data summary of diagnosis 2: Morbus Pompe

However, it is quite tedious, to identify these more interesting questions in this simple visualization because one can only see the distribution of answers for each question without emphasizing those that might be of more interest. In order to highlight those questions with a stronger agreement among the patients under the considered group, we borrow ideas from sequence logo and other statistical tests.

### 3 Involved Statistical Methods

The improved visualization compared to Figure 1 is shown in Figure 2. The length of the bars is no longer equal to the number of participants who answered questions. Higher bars indicate a stronger consensus among the answers to the

specific question, and colored points at the bottom indicate a significant deviation from a uniform distribution for the corresponding question in the diagnosis. In order to explain how the height of the bar is calculated, and the settings for calculating p-values, we briefly review the concepts of entropy, sequence logos, Chi-square test and Fisher’s exact test.

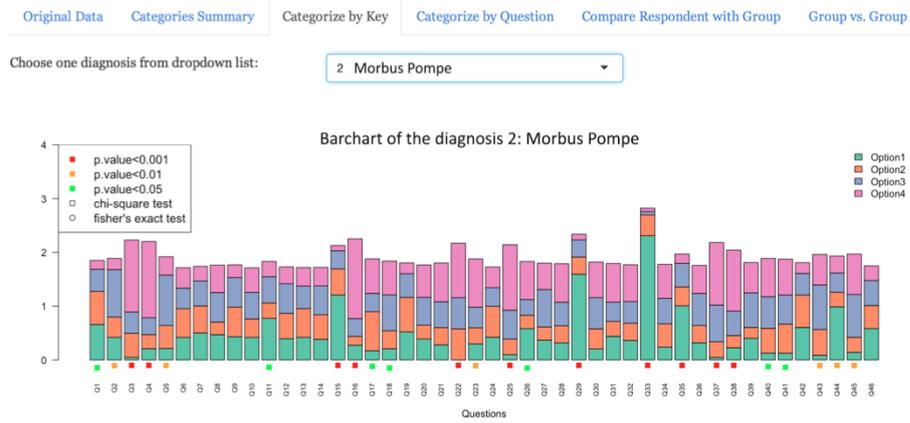


Fig. 2: Diagnosis 2 – Morbus Pompe Disease

### 3.1 Entropy

For each question  $j$ , we have  $i$  possible answers – in our case  $i = 4$ .

$$Q_j = \{Option_{1,j}, \dots, Option_{i,j}\}.$$

The relative frequency of option  $i$  in question  $j$  is denoted  $p_{i,j}$ . The entropy (see for instance [5]) for the answers to question  $Q_j$  is then

$$H(Q_j) = - \sum_i p_{i,j} \cdot \log_2 p_{i,j}.$$

Entropy is a measure of unpredictability of an outcome. The greater the entropy value, the higher the uncertainty of the outcome. If a question has an entropy of zero, this would mean that all patients of the corresponding disease group have chosen the same answer to this question.

### 3.2 Sequence Logos

Sequence logos, invented by Tom D. Schneider and R. Michael Stephens [8], are usually applied in genetics or proteomics to represent aligned sets of DNA, RNA

or protein sequences. It is used to discover conserved patterns in such sequences. In case of DNA sequences, every position would have one of the four nucleotides: A,C,T,G. At each position of a sequence, from which one has multiple samples, it is of interest whether always the same nucleotide occurs – indicating conservation – or whether there is a high variance of nucleotides at this position. Sequence logos display the distribution of the nucleotides by using the letters A,C,T,G as stacks in a bar. The height of the bar chart depends on the entropy w.r.t. A,C,T,G. The lower the entropy, the higher the bar.

In our case, we do not deal with nucleotides but with  $i$  possible answers question to instead. We use colored stacked bar charts where each color represents one possible answer to a question. Combined with the case used in this paper, a derivative work of sequence logo consists of stacks of colors. The questions correspond to the sequence. The overall height of each stack in a sequence logo depicts the information content or conservation of the sequence at that position [1], whereas in our case the relative height of each color within the stack reflects the frequency of the corresponding option in the sequence of questions.  $R_{sequence}(Q_j)$  defines the height of the bar for question  $Q_j$ . We introduce a small correction factor  $e_n$  (like a Laplace correction) for the case when no one has chosen an answer.

$$\begin{aligned} R_{sequence}(Q_j) &= \log_2 N - (H(Q_j) + e_n) \\ &= \log_2 N - \left( - \sum_i p_{i,j} \cdot \log_2 p_{i,j} + e_n \right) \end{aligned}$$

The height (denoted by  $h_{i,j}$ ) of the color stack of each option at the corresponding position (bar) is determined by multiplying the frequency  $p_{i,j}$  by the total information at that question.

$$\begin{aligned} h_{i,j} &= p_{i,j} \cdot R_{sequence}(Q_j) \\ &= p_{i,j} \cdot [\log_2 N - (H(Q_i) + e_n)] \\ &= p_{i,j} \cdot \left[ \log_2 N - \left( - \sum_i p_{i,j} \cdot \log_2 p_{i,j} + e_n \right) \right] \end{aligned}$$

### 3.3 Chi-square Test and Fisher's Exact Test

Before more rigorous and deeper statistical analysis starts, it is helpful to perform some basic inferential statistical tests like chi-square test and Fisher's exact test to determine whether the answering pattern of one question in one diagnosis deviates significantly from a uniform distribution or the answering patterns of two diagnoses for a specific question differ significantly. Since we are considering frequencies of possible answers of each question with usually more than two options, the chi-square test is an obvious choice to check whether the answering patterns deviate from the uniform distribution or whether two answering patterns differ from each other. However, the chi-square test is an asymptotic test

and should only be used for sufficiently large sample sizes. Therefore, we only apply the chi-square test when the sample size – total number of answers to a question – exceeds 40 and when the theoretical absolute frequencies exceed 5 in each considered category. Otherwise, we apply Fisher’s exact test by reducing the contingency table to  $2 \times 2$ -table.

We indicate low p-values by colors. Red stands for p-values lower than 0.001, yellow for p-values between 0.001 and 0.01 and green for p-values between 0.01 and 0.05.

For observed data that meet the conditions that mentioned above, the chi-square test is perfect for ascertaining the association between categorical variables [4]. In our case, participants’ answer patterns for a specific question in two different diagnoses ( $x$  and  $y$ ) generate a new matrix:

$$mat = \begin{vmatrix} a_x & a_y \\ b_x & b_y \\ c_x & c_y \\ d_x & d_y \end{vmatrix}$$

where entries correspond to frequencies of answers. Then the p-value can be directly calculated by applying the chi-square test.

When the sample size is less than 40 or the expected number of answer in a category is less than 5, Fisher’s exact test is used instead. Since Fisher’s exact test is designed for  $2 \times 2$  contingency tables, the participants’ answer summary for a specific question needs some further processing before applying Fisher’s exact test.

In our case, we have four possible answers in each question. The matrix  $mat$  can be organized into four  $2 \times 2$  contingency tables:  $mat_1$ ,  $mat_2$ ,  $mat_3$ , and  $mat_4$ , each one corresponding to one possible answer against all other possible answers.

$$mat_1 = \begin{vmatrix} a_x & a_y \\ b_x + c_x + d_x & b_y + c_y + d_y \end{vmatrix}, mat_2 = \begin{vmatrix} b_x & b_y \\ a_x + c_x + d_x & a_y + c_y + d_y \end{vmatrix}$$

$$mat_3 = \begin{vmatrix} c_x & c_y \\ a_x + b_x + d_x & a_y + b_y + d_y \end{vmatrix}, mat_4 = \begin{vmatrix} d_x & d_y \\ a_x + b_x + c_x & a_y + b_y + c_y \end{vmatrix}$$

In this way, we obtain four p-values and we finally carry out a correction for multiple testing, i.e. we multiply each p-value with the number of tests – corresponding to the number of possible answers to the considered question.

Different test methods are represented by different shapes of dots. In the bar plot, we use “□” to mark the chi-square test and “○” to mark Fisher’s exact test.

## 4 VoQs - a Functional Data Visualization Application

With the above mentioned visualization principle, various views can be offered to analysis questionnaire data.

By analysis the data set transmitted by users, VoQs supports users check out options probability distribution of each question, observe salient features of identified respondent groups, meanwhile figure out possible internal relationships and differences among data.

In addition to the graphic of summarizing participates number of each diagnosis (Figure 1), VoQs mainly provides four functions. The main function is to display bar charts of different grouping elements based on statistical calculation results. Using the neuromuscular disease questionnaire as an example, one can display the answering pattern for a specific disease as already shown in Figure 2.

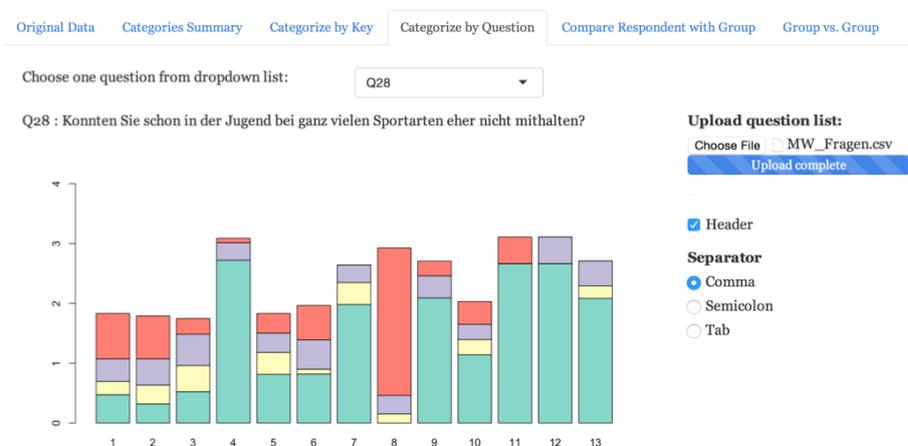


Fig. 3: Compare different diagnoses answer patterns for question 28

The second function that VoQs provides is to compare all diagnoses answer patterns for any designated question. Figure 3 shows the answer pattern of patients from different diagnoses on question 28. It can be seen that most patients in diagnosis 8 (Morbus McArdle) tend to choose option 4 (Absolutely) while in most other groups option 1 (No) is favored. This means that this question is one relevant indicator to distinguish diagnosis 8 from others.

The third function of VoQs is mapping a particular patient's questionnaire data to any diagnosis in order to check whether the answers of the patient fit to the answer pattern of a specific diagnosis. 46 points on the graph represent the answers the patient selected for each question. Points are positioned in the middle of the color corresponding to the selected answer. Figures 4 and 5 show the answer pattern of patient with diagnosis 2 to the correct diagnosis (Figure 4) and to another diagnoses (Figure 5). It is obvious that the patient's answer are more in line with diagnosis 2 than diagnosis 8.

Choose respondent ID:

MW\_P160

Choose one Group to Compare:

2 Morbus Pompe

The System record of the respondent 'MW\_P160' is Group Type: '2'

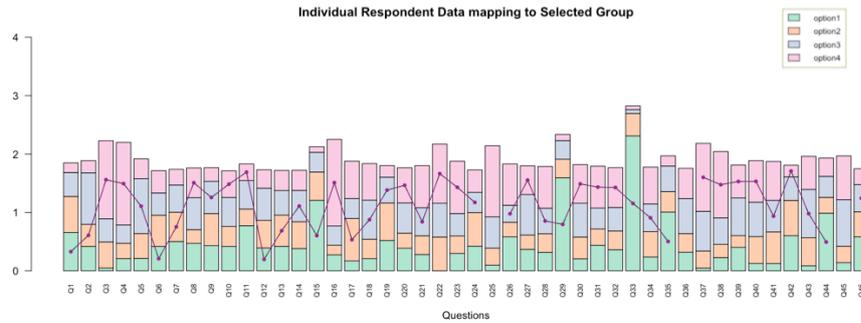


Fig. 4: Patient data mapped to diagnosis 2: Morbus Pompe

Choose respondent ID:

MW\_P160

Choose one Group to Compare:

8 Morbus McArdle

The System record of the respondent 'MW\_P160' is Group Type: '2'

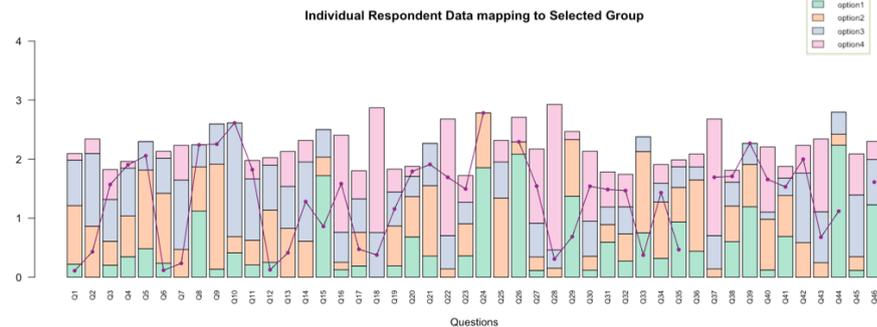


Fig. 5: Patient data mapped to diagnosis 8: Morbus McArdle

Another function is the comparison of answering patterns for two different diagnoses. Figure 6 shows a comparison between diagnosis 1 (muscular dystrophies, myotonies) and diagnosis 11 (multiple sclerosis). The focus of comparison two different groups will fall on the high peaks of the bar chart. Most interesting are two high peaks for one question with different main colors.

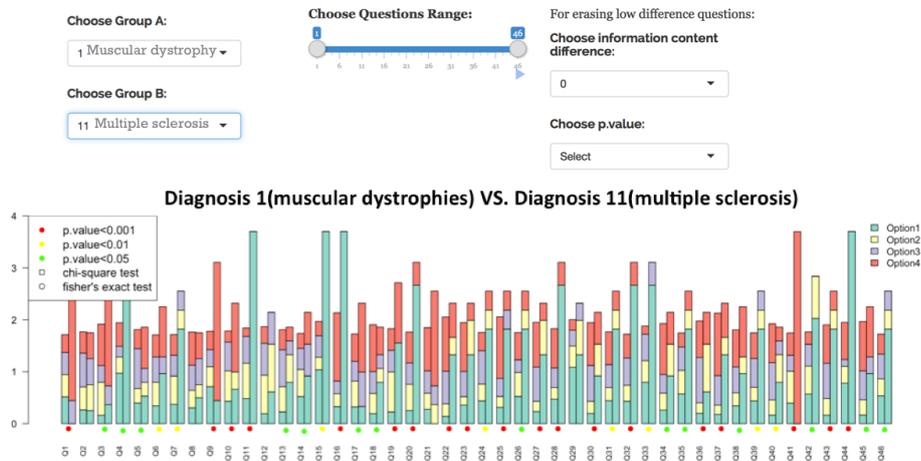


Fig. 6: Compare two diagnoses: Diagnosis 1 vs. Diagnosis 11

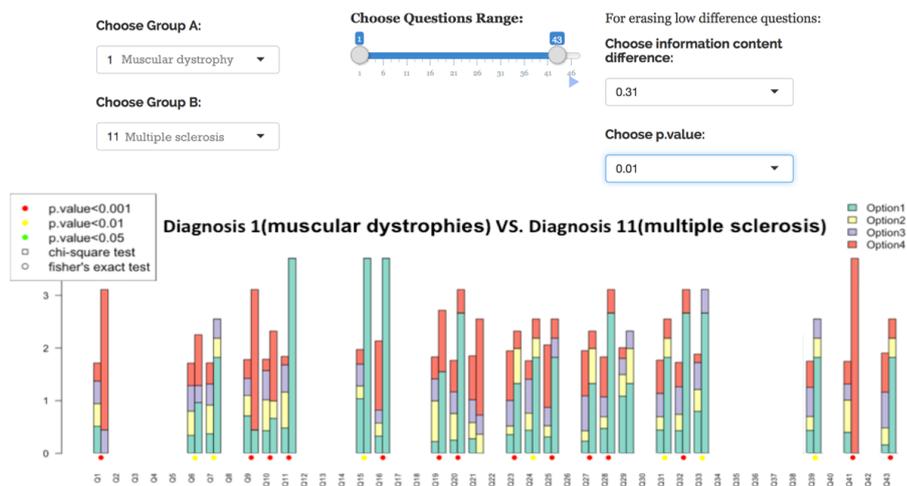


Fig. 7: Set the intervals of the graphic

In this case, the answers to the corresponding question are homogeneous within each diagnosis but heterogeneous between two diagnoses. It is also interesting if there is one tall and one short bar in some specific question, which means for one diagnosis patients tend to choose the same answer whereas for the other diagnosis patients show a very heterogeneous answering scheme. In order to obtain a more focused visualization, one can select specific questions restrict to those questions with significant differences as shown in Figure 7. This web based application is already published in ([shiny.improvedmedicaldiagnostics.com](http://shiny.improvedmedicaldiagnostics.com)).

## 5 Conclusions and Future work

VoQs is not limited to this particular area, it can be generally used for every kind of questionnaire with close-ended questions to better understand the answering patterns of certain groups, to identify questions where people from different groups typically choose different answers and to discover unusual cases that deviate strongly from usual answering patterns. With the help of VoQs application, researchers can collect important attributes and grouping questions in a more effective way.

In the future, this application will be enhanced with additional features that focus on further complex statistical testing for significant differences and the relevance of combinations of questions.

## 6 Acknowledgements

This work is partly funded by the Robert Bosch Stiftung (Germany) and Genzyme GmbH, Germany.

## References

1. Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E.: Weblogo: a sequence logo generator. *Genome research* 14(6), 1188–1190 (2004)
2. F. Klawonn, W. Lechner, L.G.: Case-centred multidimensional scaling for classification visualisation in medical diagnosis. In: *Health Information Science. Lecture Notes in Computer Science*, vol. 7798, pp. 137–148. Springer Berlin Heidelberg (March 2013)
3. Heiberger, R.M., Robbins, N.B.: Design of diverging stacked bar charts for likert scales and other applications. *Journal of Statistical Software* 57(5), 1–32 (2014)
4. Moore, D.S.: Chi-square tests. Tech. rep., DTIC Document (1976)
5. Paninski, L.: Estimation of entropy and mutual information. *Neural Computation* 15 (2003)
6. Shire: Rare disease impact report: Insights from patients and the medical community (April 2013), <http://www.geneticalliance.org.uk/docs/e-update/rare-disease-impact-report.pdf>
7. Sincero, S.M.: Types of survey questions (April 2012), <https://explorable.com/types-of-survey-questions>
8. Thomas D. Schneider, R.M.S.: Sequence logos: A new way to display consensus sequences. *Nucleic Acids Res.* 18, 6097–6100 (1990)
9. U. Reja, Katja Lozar Manfreda, V.H., Vehovar, V.: Open-ended vs. close-ended questions in web questionnaires. *Developments in Applied Statistics* 19, 159–177 (2003)