

UCLA

UCLA Electronic Theses and Dissertations

Title

Cluster Analysis of Questionnaire Responses to MyTherapistMatch.com

Permalink

<https://escholarship.org/uc/item/3jh0h35j>

Author

Zhou, Yang

Publication Date

2012

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

**Cluster Analysis of Questionnaire Responses to
MyTherapistMatch.com**

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Science in Statistics

by

Yang Zhou

2012

© Copyright by
Yang Zhou
2012

ABSTRACT OF THE THESIS

**Cluster Analysis of Questionnaire Responses to
MyTherapistMatch.com**

by

Yang Zhou

Master of Science in Statistics

University of California, Los Angeles, 2012

Professor Frederic Paik Schoenberg, Chair

MyTherapistMatch.com seeks to match visiting patients with suitable therapists after patients fill out the online questionnaire which consists of many psychological questions. However, a problem with this website is that many patients in fact do not end up scheduling a session with a therapist. The website founder believes that one of the major reasons is the length of the questionnaire. Therefore, to reduce annoyance for users, the task becomes selecting a subset of necessary questions from the questionnaire. The website provides patient selection data and patient action data which records how a patient interacts with a matched therapist. This thesis tries to implement hierarchical clustering method on both the question responses and the questions themselves, in order to find a reasonable way to pick the necessary questions. Correlation coefficients and Pearson's chi-squared test are used to define the metrics in hierarchical clustering. Satisfiable results are obtained. A linear model is also used to find the relationship between question responses and patient actions.

The thesis of Yang Zhou is approved.

Hongquan Xu

Qing Zhou

Frederic Paik Schoenberg, Committee Chair

University of California, Los Angeles

2012

To my beloved parents . . .

TABLE OF CONTENTS

1	Introduction	1
2	Data Description	3
3	Correlation, Hierarchical Clustering and Question Selection . .	5
3.1	Response Coding, Correlation Coefficient and Hierarchical Clustering	5
3.2	Contingency Table and Pearson's χ^2 Test	12
3.3	Hierarchical Clustering for Questions With P-values of χ^2 Test . .	13
3.4	Question Selection With Cluster Center	14
4	Question Selection With Customer Behavior Data	18
5	Psychological Background and Result Analysis	22
5.1	Some Psychological Concepts and Question Domains	22
5.2	Result Analysis	25
6	Discussion	32
A	Linear Regression Model	34
B	Survey Questions and Their Domains	39
	References	46

LIST OF FIGURES

2.1	Procedures to Find a Therapist	3
2.2	Question Example	4
2.3	Patients Distribution	4
3.1	Hierarchical Clustering Dendrogram Using d_1	9
3.2	Hierarchical Clustering Dendrogram Using d_2	10
3.3	Multidimensional Scaling Plot of Question Responses	11
3.4	Hierarchical Clustering Dendrogram for Questions Using P-value .	14
3.5	P-value Matrix Ordered by Clusters	15
4.1	Patient Action Points Distribution	20

LIST OF TABLES

2.1	Typical User Action Record	4
3.1	Response Coding	6
3.2	Contingency Table Example	12
3.3	Clusters Produced by d_2 (Left) and P-value (Right)	16
3.4	Cluster Centers Produced by d_2 (Left) and P-value (Right)	17
4.1	Action Points	18
4.2	Dummy Variable Coding	19
5.1	Question Selection Results According to Question Domains	27
5.2	Question Domain and Cluster Center Produced by d_2 (Left) and P-value (Right)	27
5.3	Summarization of Clusters by P-value	30

ACKNOWLEDGMENTS

I would like to show my deepest gratitude to my supervisor, Professor Frederic Paik Schoenberg, for his guidance in every stage of this research project. I would also extend my thanks to Professor Qing Zhou and Professor Hongquan Xu, who patiently help me formulate some important ideas in this thesis.

CHAPTER 1

Introduction

MyThetapistMatch.com is a website which finds suitable therapists for visiting patients. Therapists in this website include psychologists, psychiatrists, professional counselors, psychoanalysts, marriage and family therapists, rehabilitation counselors and clinical social workers. The therapists provide help on depression, anxiety, stress management, anger management, teen issues, divorce recovery, marriage issues, emotional issues, life transitions and self improvement.

Traditionally, websites with similar functions, for example GoodTherapy.org and PsychologyToday.com, try to match patients with random referrals within a certain area. This recommendation procedure in fact only considers location and fails to consider other factors while connecting a patient with a therapist, which sometimes results in user dissatisfaction. MyTherapistMatch.com seeks to generate personalized lists of therapist matches for visiting patients after they fill out their online questionnaires. The website hopes to improve user satisfaction and help produce successful therapeutic results in this way.

Its matching strategy is based on elements selected from the theories of Isabel Briggs Myers, Telos Programs, the Destination Method, Gestalt, Representational Systems, Cognitive-Behavioral Perspectives, Meta-Modeling, Epistemology and Axiology. Their matching decisions are based on 23 weighted criteria, including Preferred Learning Styles, Motivation Strategies, Perceptual Positions, Decision

Strategies, Introversion/Extroversion and Preferred Orientation in Time¹.

However, a major problem that has been identified is that a high proportion of visitors to the website do not end up scheduling a session with a therapist. The website founder believes that one of the major reasons for this is the length of the questionnaire. Therefore, to reduce the annoyance for users, the task becomes selecting questions in the questionnaire using statistical methods.

MyTherapistMatch.com provides the user selection data of different survey questions and user behavior records like viewing a therapist's profile, clicking on the referral link, emailing or calling a therapist, etc. The website also emails follow-up surveys to collect user satisfaction feedbacks, which is important for question selection. Unfortunately, few users fill out the survey. As a result, the following statistical analysis will focus on the user selection and behavior data.

The statistical methods involved in the following analysis include hierarchical clustering and linear regression. Hierarchical clustering will be first applied to question responses and then to questions themselves by using different metrics. It is a method which divides objects (here question responses or questions) into several different groups after studying the relations among them. Question selection can be performed within clusters then. Linear regression is used here to predict user behaviors using the question responses. We can also find important questions for to user behaviors with this method.

¹<http://www.mytherapistmatch.com/howitworks.aspx>

CHAPTER 2

Data Description

As mentioned in the introduction, a patient follows three steps to find therapists who are compatible using this website:

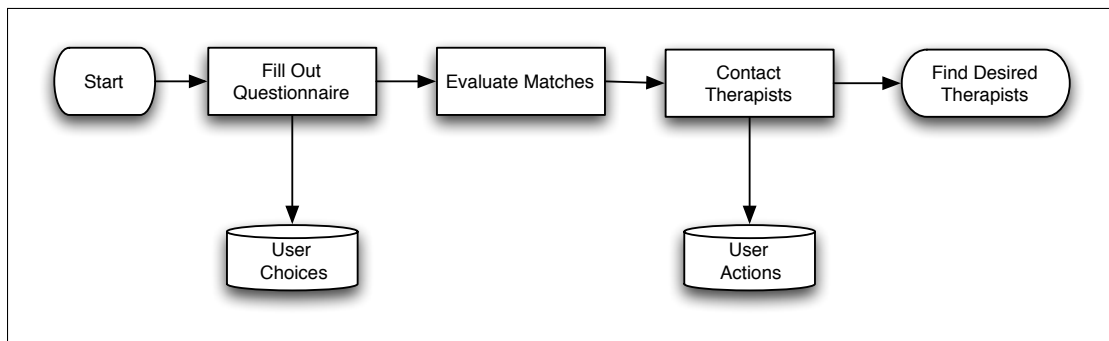


Figure 2.1: Procedures to Find a Therapist

1. Complete the questionnaire of the website which helps produce ideal therapists according to the answers.
2. Receive a list of therapists and evaluate them.
3. Pick satisfiable therapists and contact them via email, phone or url links provided.

User choices and actions are recorded by the website at the same time. These procedures are shown in *Figure 2.1*. Questions from the website look like *Figure 2.2*. A typical user action record looks like *Table 2.1*.

1. When solving a problem, I tend to:
- consult with someone about it.
 - get in touch with my deeper self.
 - look at the big picture.
 - talk it over with myself or another person.

Figure 2.2: Question Example

Patient Id	Therapist Id	Insert Date	Action
{xxxxxx}	{xxxxxx}	xxxx-xx-xx xx:xx:xx.xx	ProfileView

Table 2.1: Typical User Action Record

The distribution of patients across different states is shown in *Figure 2.3*. From this figure, we can see that patients are mainly from California and New York State, especially California, where there are 2206 patients, a large portion of the 3763 patients nationwide.

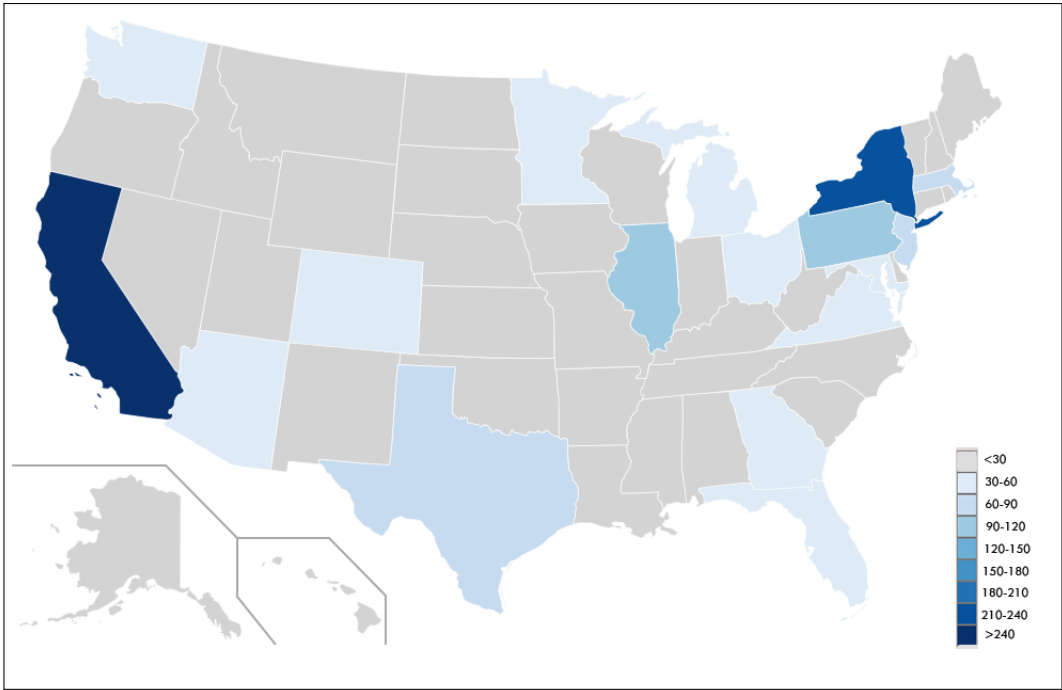


Figure 2.3: Patients Distribution

CHAPTER 3

Correlation, Hierarchical Clustering and Question Selection

3.1 Response Coding, Correlation Coefficient and Hierarchical Clustering

In order to pick questions from the large pool of given questions, a natural way is to study the relationship of these questions. The point is, if users' responses to two different questions present similar patterns, we can conclude that these two questions are similar and then pick only one of them.

Pearson's correlation coefficient r is often used to study the relationship between two samples. It ranges from -1 to 1 . It indicates strong positive relationship when r is close to 1 , strong negative relationship when r is close to -1 , and weak relationship when $|r|$ is close to 0 .

Since the survey questions do not have numerical response which is needed by r , we can try to study their responses instead, after proper coding. The coding strategy is shown in *Table 3.1*. 1 means selection and 0 means absence of selection. For example, patient 1 picks response 1 for question 1 and response 1 for question 2, patient 2 picks response 2 for question 1 and response 2 for question 2, etc. If we always observe that different patients pick response 1 for question 1 and response

2 for question 2 at the same time, then we can say these two responses are highly correlated. When two responses from two questions are highly correlated, we can say that these two questions are in fact highly correlated. As a result, we only need to pick one question from two.

	Q1R1	Q1R2	Q2R1	Q2R2	Q2R3	...
patient1	1	0	1	0	0	...
patient2	0	1	0	1	0	...
patient3	1	0	0	0	1	...
patient4	1	0	0	1	0	...
			...			

Table 3.1: Response Coding

If we code responses like *Table 3.1*, then the Pearson’s correlation coefficient $cor(r_1, r_2)$ of two responses r_1 and r_2 goes up towards 1 when the high correlation described above exists. Also, according to *Table 3.1*, if there are not too many responses for the same question, any two of them should present strong negative correlation, which means their correlation coefficient would be close to -1 . If the coefficient is near 0, these two responses do not present strong correlation.

Cluster analysis, also called data segmentation, has a variety of goals. All relate to grouping or segmenting a collection of objects into subsets or clusters, such that those within each cluster are more closely related to one another than objects assigned to different clusters [TF01].

Hierarchical clustering methods divide the data into a collection of clusters of various sizes and numbers of clusters, typically with a branching structure. One

outstanding feature of hierarchical clustering is that it does more than produce a flat list of clusters; it also shows their relationships in an explicit way [Jan10]. For agglomerative hierarchical clustering, at the first stage of the algorithm, each object is assigned to its own cluster and then the algorithm proceeds iteratively, at each stage joining the two most similar clusters, continuing until there is just a single cluster. There are many ways to measure the dissimilarity between two objects. All these metrics are small when two objects are similar and large when they are quite different.

We only need to pick one from those similar objects. Therefore, if we build clusters of question responses with their dissimilarity, we find a way to pick responses according to their clusters. We hope to put the responses from the same question in the same cluster, since we need to decide which questions are important at last, rather than only responses. A natural approach would be to put strongly correlated responses, whether negative or positive correlated responses, into the same cluster. In this way, responses from the same question would have a good chance to be in the same cluster. We can define the dissimilarity measure between two responses as:

$$d_1(r_1, r_2) = 1 - |cor(r_1, r_2)| \tag{3.1}$$

The dissimilarity between strongly correlated responses will be close to 0. Euclidean distance $\|r_1 - r_2\|_2$, Manhattan distance $\|r_1 - r_2\|_1$, maximum distance $\|r_1 - r_2\|_\infty$ and Mahalanobis distance $\sqrt{(r_1 - r_2)^T S^{-1} (r_1 - r_2)}$ do not work here, since all of them tend to put responses from the same question into different clusters.

The definition of dissimilarity between two clusters is also important in the agglomerative hierarchical clustering algorithm in order to merge the clusters step

by step. In fact we only need to find representative responses from each cluster and define dissimilarity with them as group dissimilarity. Three methods are often discussed: single linkage (SL), complete linkage (CL) and group average (GA). Single linkage method takes the intergroup dissimilarity to be that of the closest pair, complete linkage takes the intergroup dissimilarity to be that of the furthest pair and group average clustering uses the average dissimilarity between groups. If the data similarities exhibit a strong clustering tendency, these three methods produce similar results. However, when the data are not very neatly organized into distinct clusters, single linkage can violate the "compactness" property that all observations within each cluster tend to be similar to one another. As a result, it will present a chaining phenomenon. On the contrary, complete linkage produces compact clusters[TF01]. Group average stands in the middle of the other two methods. In this problem, we pay attention to "compactness". We need all the responses in the cluster to be similar to each other, since we need a representative of the cluster. "Closeness" property, which is emphasized by single linkage, is not as important as "compactness". Therefore, we implement complete linkage method to this dataset using d_1 in *Equation 3.1* as the metric. The dendrogram is shown in *Figure 3.1* to visualize this clustering.

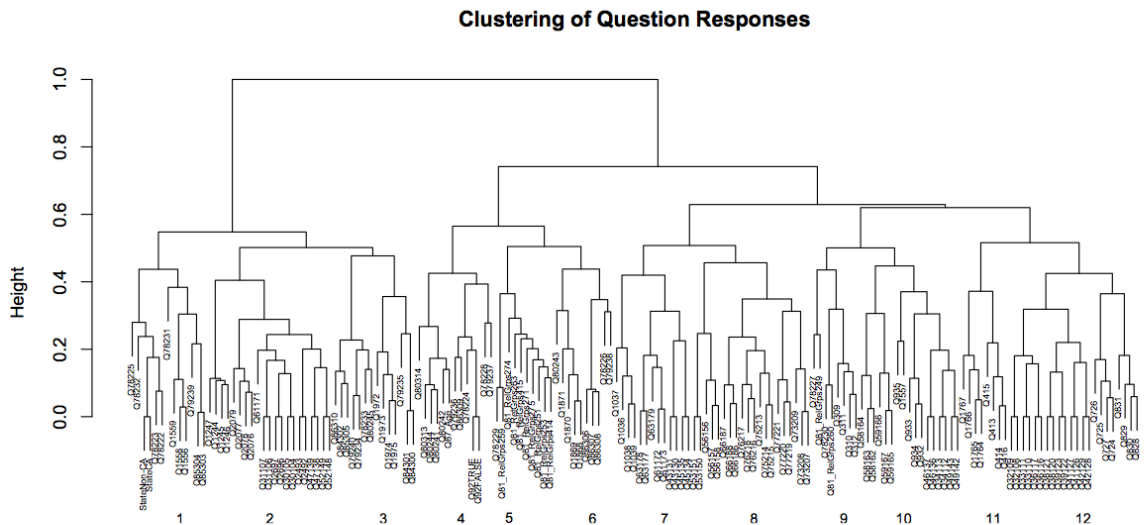


Figure 3.2: Hierarchical Clustering Dendrogram Using d_2

Multidimensional scaling (MDS) is a method in information visualization for exploring dissimilarities in data. It is often used to find lower dimensional representation of the original dataset while retains the dissimilarity structure [Seg11]. For example, two-dimensional representation can be plotted out on a page, therefore makes patterns and structures in the data intuitive.

This method assigns each object a random or predetermined location and starts with the dissimilarity matrix of objects. For each pair of objects, the current distance is compared to the target dissimilarity. Then all the objects are pushed or pulled towards the similarity structure according to the ratio of current distance and target dissimilarity.

There are many algorithms to calculate this lower dimensional representation. Here we implement one of them, metric MDS, only to show the structure of question responses. *Figure 3.3* plots the result using this method. We can observe the

3.2 Contingency Table and Pearson's χ^2 Test

Contingency table is often used to record and analyze the relation between categorical variables in statistics. Take $Q3$ and $Q4$ in this thesis as an example, the contingency table is shown in *Table 3.2*.

	Q4R1	Q4R2	Q4R3	Q4R4
Q3R1	53	323	23	120
Q3R2	80	314	35	395
Q3R3	67	271	38	253
Q3R4	197	723	117	742

Table 3.2: Contingency Table Example

This table shows that there are 53 people who choose question 3's response 1 and question 4's response 1, 323 people who choose question 3's response 1 and question 4's response 2, etc. Each observation is put into the particular cell according to its responses of these two questions.

If we assume question 3 and question 4 are independent, then we can try Pearson's χ^2 test to test this assumption statistically. If the assumption holds, then these two questions are independent, otherwise there should be some relationship between them.

When the assumption of independence holds (null hypothesis), the theoretical frequency for a cell is [Rez09]:

$$e_{i,j} = \frac{(\sum_{n_a=1}^a f_{i,n_a}) \cdot (\sum_{n_b=1}^b f_{n_b,j})}{N}, \quad (3.3)$$

when there are a columns and b rows. f_{i,n_a} is the observed frequency in cell (i, n_a) and N is the total observation number. Then the χ^2 test statistics is set to be:

$$\chi^2 = \sum_{i,j} \frac{(f_{i,j} - e_{i,j})^2}{e_{i,j}} \quad (3.4)$$

If the χ^2 probability for this value (p-value) is less than a given significance level α , then the null hypothesis of independence is rejected at level α . This means the two questions are in fact correlated in this example. We can produce p-values for each two questions and summarize these values in a data matrix.

3.3 Hierarchical Clustering for Questions With P-values of χ^2 Test

P-values produced by statistical comparison tests can in fact be used as dissimilarity measure in hierarchical clustering [Jor05]. In this problem, small p-values means correlation of two questions and in this way these two questions should be put into the same cluster. Therefore, we use p-values directly as dissimilarity measure. *Q87_Age* is not considered in this situation, since it is not a category variable. The hierarchical clustering result is shown in *Figure 3.4*.

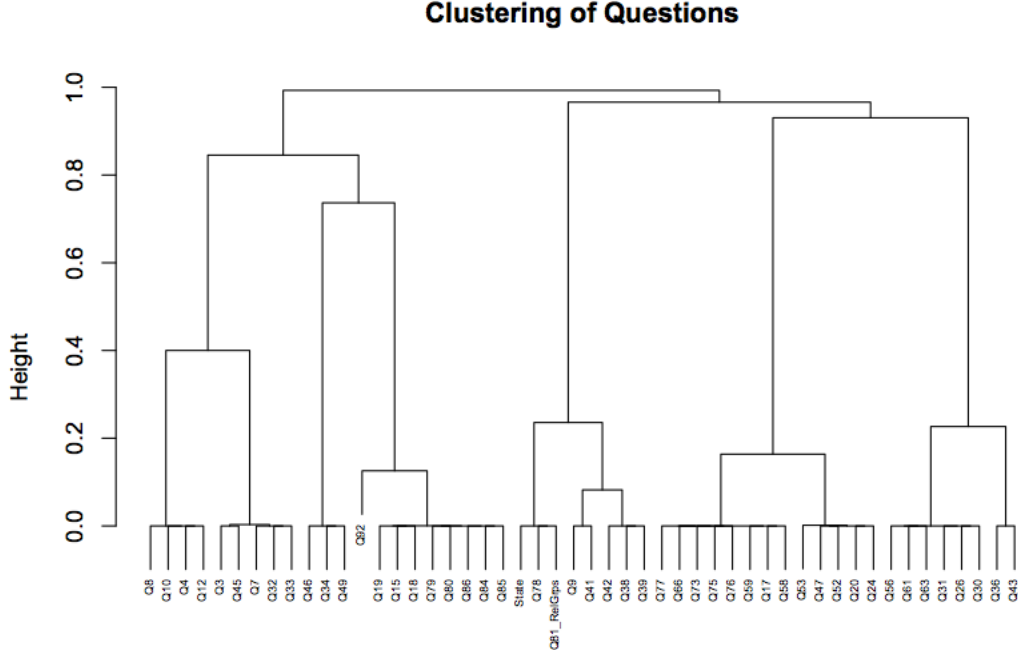


Figure 3.4: Hierarchical Clustering Dendrogram for Questions Using P-value

In fact, we can observe the clusters if we plot the p-value matrix for survey questions according to the order of clusters. This is shown in *Figure 3.5*. White is for two correlated questions and blue is for two questions that are not that correlated. From the diagonal of the matrix, we can see the sequence of 10 clusters. In a cluster, every question is closely related to each other. In this way, the white squares of different sizes along the matrix diagonal are indicators of clusters.

3.4 Question Selection With Cluster Center

The clusters produced by the above two methods are summarized in *Table 3.3*. Blue and deep blue are used to label clusters with their "parent" clusters, whose hierarchies are higher than them in hierarchical clustering. In other words, clusters may have common "parent" cluster, and therefore belong to the same bigger

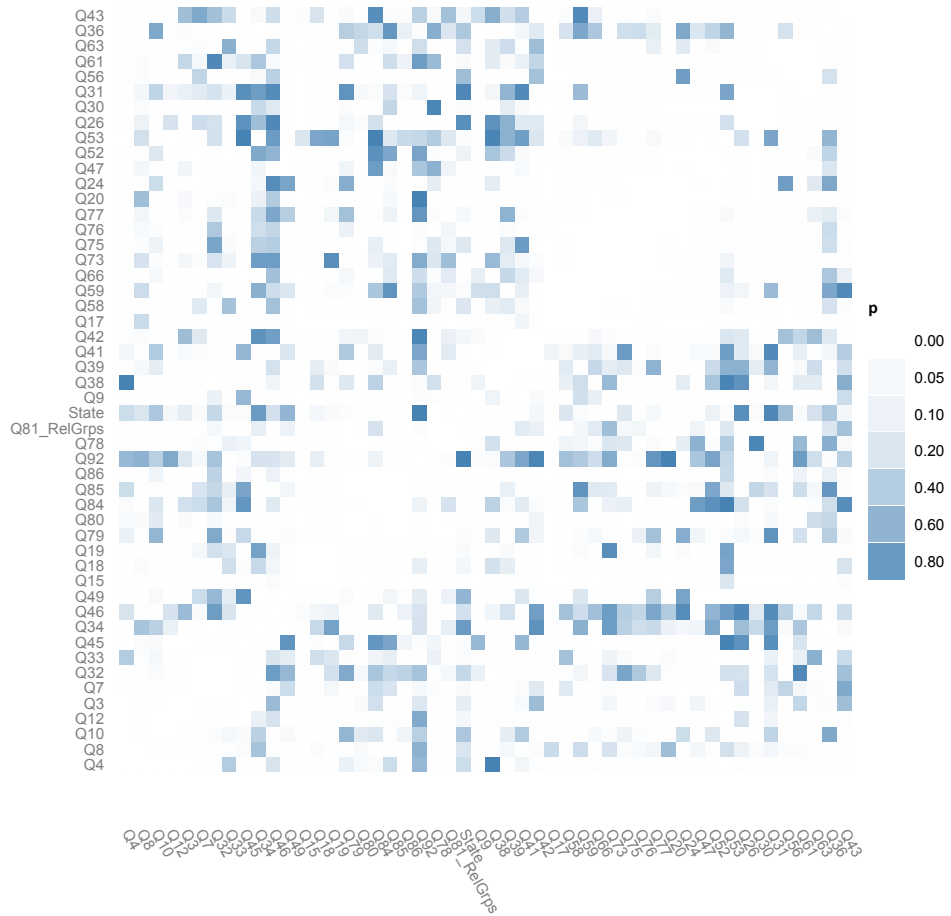


Figure 3.5: P-value Matrix Ordered by Clusters

Cluster	Questions	Cluster	Questions
1	Q15, Q78, Q79, Q85, State	1	Q4, Q8, Q10, Q12
2	Q12, Q20, Q24, Q26, Q30, Q31, Q47, Q52, Q61	2	Q3, Q7, Q32, Q33, Q45
3	Q19, Q78, Q79, Q80, Q84, Q85, Q86	3	Q34, Q46, Q49
4	Q78, Q79, Q80, Q86, Q87_Age, Q92	4	Q15, Q18, Q19, Q79, Q80, Q84, Q85, Q86, Q92
5	Q78, Q81_RelGrps	5	Q78, Q81_RelGrps, State
6	Q18, Q78, Q79, Q80, Q86	6	Q9, Q38, Q39, Q41, Q42
7	Q10, Q43, Q45, Q61, Q63	7	Q17, Q58, Q59, Q66, Q73, Q75, Q76, Q77
8	Q53, Q56, Q66, Q73, Q75, Q76, Q77	8	Q20, Q24, Q47, Q52, Q53
9	Q3, Q78, Q81_RelGrps	9	Q26, Q30, Q31, Q56, Q61, Q63
10	Q9, Q15, Q34, Q46, Q49, Q58, Q59	10	Q36, Q43
11	Q4, Q17		
12	Q7, Q8, Q32, Q33, Q36, Q38, Q39, Q41, Q42		

Table 3.3: Clusters Produced by d_2 (Left) and P-value (Right)

cluster. We observe that sometimes the same question falls into different clusters in the left table. The reason is that these questions have too many responses, which weakens the negative correlation between responses of the same question. For the right table, there won't be such problems, since we perform clustering on questions themselves, rather than question responses.

If we pick one question from each cluster, we can get 12 questions (although perhaps not distinct ones) from clusters shown on the left of *Table 3.3*, and 10 questions from clusters on the right. Then we satisfy the website's need. We therefore define a cluster center that is the representative of its cluster. This is the object which is closest to other objects in the same cluster. The mathematical definition is:

$$c = \arg \min_{m_a} \sum_{m_b \neq m_a} d(m_a, m_b), \quad (3.5)$$

where c is the cluster center, m_a is any member of the cluster, m_b is any member of the same cluster except m_a , $d(m_a, m_b)$ is the dissimilarity measure of m_a and m_b . In our problem, the dissimilarity can be defined as $1 - |\text{cor}(m_a, m_b)|$ in the response clustering, and as p-value in the question clustering. After calculation, the centers are selected in *Table 3.4*.

Cluster	Questions	Cluster	Questions
1	Q15, Q78, Q79, Q85, State	1	Q4, Q8, Q10, Q12
2	Q12, Q20, Q24, Q26, Q30, Q31, Q47, Q52, Q61	2	Q3, Q7, Q32, Q33, Q45
3	Q19, Q78, Q79, Q80, Q84, Q85, Q86	3	Q34, Q46, Q49
4	Q78, Q79, Q80, Q86, Q87_Age, Q92	4	Q15, Q18, Q19, Q79, Q80, Q84, Q85, Q86, Q92
5	Q78, Q81_RelGrps	5	Q78, Q81_RelGrps, State
6	Q18, Q78, Q79, Q80, Q86	6	Q9, Q38, Q39, Q41, Q42
7	Q10, Q43, Q45, Q61, Q63	7	Q17, Q58, Q59, Q66, Q73, Q75, Q76, Q77
8	Q53, Q56, Q66, Q73, Q75, Q76, Q77	8	Q20, Q24, Q47, Q52, Q53
9	Q3, Q78, Q81_RelGrps	9	Q26, Q30, Q31, Q56, Q61, Q63
10	Q9, Q15, Q34, Q46, Q49, Q58, Q59	10	Q36, Q43
11	Q4, Q17		
12	Q7, Q8, Q32, Q33, Q36, Q38, Q39, Q41, Q42		

Table 3.4: Cluster Centers Produced by d_2 (Left) and P-value (Right)

CHAPTER 4

Question Selection With Customer Behavior Data

Other information might help us with question selection. If we take the recorded user action into account, we can build a response for each user and then use linear regression to select significant questions for this response.

First, we build an action point system for user actions. we attempt to assign a high point value for a high level action, for example "PhoneClicked", which reflects the satisfaction of the user for the match. A user gets his/her total points as the response value by adding all his/her action points together. The point system is shown in *Table 4.1*.

Action Recorded	Action	Points
PhoneClicked	Patient clicks on the 'contact this therapist via phone' link on the therapist's profile	10
ContactClicked	Patient clicks on the 'contact this therapist via email' link on the therapist's profile	10
WebsiteReferral	Patient clicks on the 'therapist's website' link on the therapist's profile	5
MyVirtualShrink	Patient doesn't get any matches and they click on a link referring to an external website: myvirtualshrink.com	3
ProfileView	Patient clicks on the 'contact this therapist via email' link on the therapist's profile	3
Match	Patient gets matched with one therapist	1

Table 4.1: Action Points

All the total points for each user are calculated in this way. The distribution of action points across states is shown in *Figure 4.1*. From this figure, we can see that Hawaii has the largest average point value, Wyoming takes the second place and West Virginia takes the third. California is in the next level. Montana and

Idaho has no user behavior data and therefore looks grey. When we pay attention to the states whose average point value is higher than that of California, we can observe from *Figure 2.2* that their patient numbers are too small for consideration. Also, for most of the "blue" states where there are lots of prospective patients using the MyTherapistMatch.com website, the points average is lower than that of California.

A linear regression model is built with all the questions mentioned above in the clustering methods. For categorical variables, however, we need to build some dummy variables in order to code them for regression. These variables takes the values 0 or 1 to indicate the absence or presence of some categorical effect that might affect the outcome. In this way, one categorical variable is in fact separated into several variables with values of 0 and 1. In our case, when we have k responses to one question, we set $k - 1$ variables with 0 and 1 values, with each variable corresponding to a different possible response to the question. *Table 4.2* is the coding example.

Response 2	Response 3	Response 4
1	0	0
0	1	0
0	0	1
0	0	0

Table 4.2: Dummy Variable Coding

The first row in *Table 4.2* shows the user picks Response 2, the second shows Response 3, the third shows Response 4 and the fourth shows Response 1.

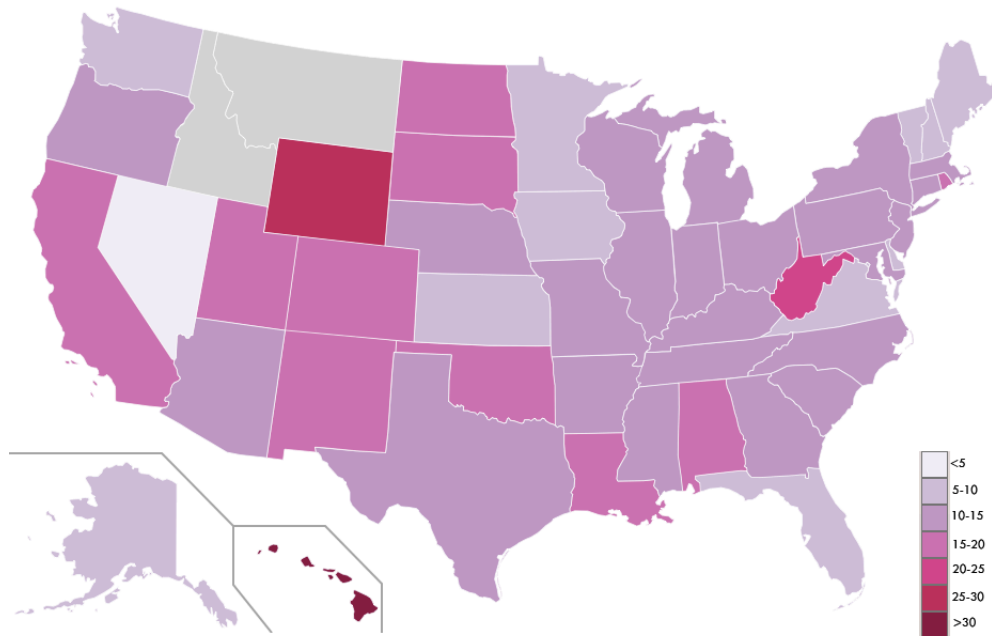


Figure 4.1: Patient Action Points Distribution

Therefore, the problem of telling whether a question is statistically significant becomes the problem of testing whether other responses are statistically "different", which means they have different impact on the outcome when compared with this specific response to the same question.

The result of linear regression is appended in *Appendix A* and the details of the questions are shown in *Appendix B*. The significant responses are $Q1765$, $Q41127$, $Q78223$, $Q79239$, $Q80245$, $Q81_RelGrps253$, $Q81_RelGrps274$, $Q81_RelGrps275$, $Q81_RelGrps414$, $Q81_RelGrps415$, $Q85304$, $StateNon-CA$ and $Q92TRUE$. So the questions selected are $Q17$, $Q41$, $Q78$, $Q79$, $Q80$, $Q81$, $Q85$, $State$ and $Q92$. Most of them are in fact demographics questions, except $Q17$ and $Q41$. Therefore, these questions are perhaps important for user action. However, we notice that the multiple R^2 value for this model is as low as 0.1847, the adjusted R^2 value is as low as 0.1575. Therefore, the linear model is in fact not a very good model

to predict user action. More efforts are needed to explore other factors which are related to user action.

CHAPTER 5

Psychological Background and Result Analysis

5.1 Some Psychological Concepts and Question Domains

The questions in the questionnaire are from several question domains set by the website. These question domains are selected according to psychological theories listed in the introduction part. All the questions and their corresponding domains are shown in *Appendix B*. We will briefly discuss some necessary psychological concepts here in this section.

Preferred representational system is a neuro-linguistic programming model (NLP) in psychology, which examines how the human mind process information. NLP calls each individual's perception of the world their 'map'. NLP teaches that our mind-body (neuro) and what we say (language) all interact together to form our perceptions of the world, or maps (programming). Each person's map of the world determines feelings and behavior. Therefore, impoverished - and unrealistic - maps can restrict choices and result in problems. As an approach to personal development or therapy it involves understanding that people create their own internal 'map' or world, recognizing unhelpful or destructive patterns of thinking based on impoverished maps of the world, then modifying or replacing these patterns with more useful or helpful ones. There is also an emphasis on ways to change internal representations or maps of the world in order to increase behavioral flexibility [Gri79, BG75, BG83].

When people engage in activities, they are in fact using some internal representation of the materials they are involved with, such as a conversation, a rifle shot, a spelling task. These representations can be visual, auditory, kinesthetic, or involve the other senses [Dru88]. NLP states that people would have a preference on one representational system over other representational systems. Theoretically speaking, the psychotherapist can achieve better communication with the client and hence more effective results by matching and working within the preferred representational system. Therefore, the website includes this important type of questions in their online questionnaire in order to provide better therapist matching.

For example, the first question in this questionnaire, $Q3$ is a representational system question:

Q: When solving a problem, I tend to:

R1: consult with someone about it.

R2: get in touch with my deeper self.

R3: look at the big picture.

R4: talk it over with myself or another person.

$R2$ is kinesthetic, $R3$ is visual, $R4$ is auditory and $R1$ is hence a neutral response. All the representational system questions in this questionnaire are designed like this example. All of them have one kinesthetic, one visual, one auditory and one neutral response, while sometimes change the order of these response types.

Internal/External questions are from the famous theory behind the MBTI personality test, which is created by Isabel Briggs Myers. It seeks to help people find their best fit personality type in order to help them succeed most in life. Internal/External is like the Extraversion/Introversion type in MBTI personality test. "Internal" means someone likes to get energy from dealing with the ideas in one's inner world. "External" means someone likes to get energy from involvement in events and communication with others. Therefore, questions of this type should have two responses, one is "internal" and the other is "external". An example is *Q24*:

Q: I know I've done a good job when:

R1: I notice it myself.

R2: someone lets me know.

All other questions are designed in a similar way as questions from the above two question domains. We do not go into details and only briefly introduce the main idea of them here.

Towards/Away questions judge whether people act in order to avoid something or towards something.

Options/Procedures questions decide whether people like to create their own options or follow existent rules.

Sameness/Difference questions decide whether people like to change their way of doing things.

Specific/General questions decide whether people like details or the big picture of events.

Proactive/Reactive questions decide whether people like to act themselves or seek help from others.

Perceptual Positions questions decide whether people think from their own perspective or from others' perspective.

Experience of Time questions decide whether people like to relate to past, current or future events.

Demographics questions record people's ethnicity, sexual orientation, relationship status, religions, attitude towards smoking and alcohol, exercise frequency, location, age and attitude towards online therapy.

All the question domains and their explanations have been listed above. We hope to explain our result using these explanations.

5.2 Result Analysis

After the above introduction of some psychological concepts and explanation of question domains, we can start to analyze the results produced in *Chapter 3*. The significant factors in the linear regression model are mostly demographics questions and the model is in fact not a very good prediction for user action. Therefore,

we will not discuss the linear model result in this section and focus only on the clustering results.

The clusters produced are already presented in *Table 3.4*. However, when we take question domains into account, we might hope to see how these methods select across different domains. This is shown in *Table 5.1*. Also, we hope to examine the inner constitution of these clusters from the perspective of question domain. This is shown in *Table 5.2*.

We begin our analysis from *Table 3.4* and *Table 5.2*. The right table can show us the relationship of different questions, while the left table can tell us how question responses relate to each other. Analysis combining these two sets of tables can give us a comprehensive view of the result. The right table is the first to be analyzed.

In the right table, the first cluster consists of four preferred representational system questions. The survey questions shown in *Appendix B* are worth examining. We can see that although there are 7 preferred representational system questions, the four questions in the first cluster ($Q4$, $Q8$, $Q10$, $Q12$) are all about communication and talking except $Q12$, which looks like what all other preferred representational system questions are in fact asking. Therefore it seems reasonable for me to pick one of them as the cluster representative. Here, the cluster center turns out to be $Q4$.

The second cluster consists of two preferred representational system questions $Q3$ and $Q7$, two options/procedures questions $Q32$ and $Q33$, and one specific/general question $Q45$. Although $Q3$, $Q7$, $Q32$ and $Q33$ are from different question domains, but their questions are in fact similar. For example, although responses

Domain #	Question Domain	Method 1	Method 2
1	Preferred Representational System	Q3, Q4, Q9	Q4, Q7
2	Towards/Away	Q18	/
3	Internal/External	Q30	/
4	Options/Procedures	Q33	Q34, Q36
5	Sameness/Difference	/	Q41
6	Specific/General	/	/
7	Proactive/Reactive	/	Q47
8	Perceptual Positions	Q63	Q61
9	Experience of Time	Q76	Q66
10	Demographics	Q78, Q80, Q81_RelGrps, Q84	Q78, Q80

Table 5.1: Question Selection Results According to Question Domains

Cluster	Questions	Cluster	Questions
1	2, 10, 10, 10, 10	1	1, 1, 1, 1
2	1, 2, 3, 3, 3, 3, 7, 7, 8	2	1, 1, 4, 4, 6
3	2, 10, 10, 10, 10, 10, 10	3	4, 6, 7
4	10, 10, 10, 10, 10, 10	4	2, 2, 2, 10, 10, 10, 10, 10, 10
5	10, 10	5	10, 10, 10
6	2, 10, 10, 10, 10	6	1, 5, 5, 5, 5
7	1, 6, 6, 8, 8	7	2, 8, 8, 9, 9, 9, 9, 9
8	7, 8, 9, 9, 9, 9	8	2, 3, 7, 7, 7
9	1, 10, 10	9	3, 3, 3, 8, 8, 8
10	1, 2, 4, 6, 7, 8, 8	10	4, 6
11	1, 2		
12	1, 1, 4, 4, 4, 5, 5, 5, 5		

Table 5.2: Question Domain and Cluster Center Produced by d_2 (Left) and P-value (Right)

are very different, Q_3 and Q_{32} are both asking: "when solving a problem, I tend to prefer". These four questions are all asking one's attitude when one is dealing with things. Q_{45} seems irrelevant at first glance, but its response "I generally prefer thinking about the big picture in life" is like the response "look at the big picture" of Q_3 . What's more, in fact options/procedures questions are similar as specific/general questions if we take a look at their responses. In fact options/procedures questions are always correlated with specific/general questions since we can also see Q_{34} and Q_{46} , Q_{36} and Q_{43} in the same cluster. Perhaps

we can compare the nature of these two type of questions more in the future.

As mentioned above, options/procedures question *Q34* and specific/general question *Q46* reside in the third cluster. Proactive/reactive question *Q49* also reside in this cluster. In fact, *Q34* and *Q49* are asking the exact same thing: whether you will do the planning for traveling. It is reasonable for me to retain only one of them.

The majority of demographics questions (*Q78*, *Q79*, *Q80*, *Q84*, *Q85*, *Q86* and *Q92*) are placed in cluster 4, along with three towards/away questions (*Q15*, *Q18*, *Q19*). It is not a surprise that so many demographics questions are in the same cluster. The towards/away questions here are all related to health and life. The cluster center is *Q80*, about relationship status. In fact relationship status is indeed an important factor with strong influence on other things, like sexual orientation (*Q79*), smoking and drinking habits (*Q84* and *Q85*).

The fifth cluster consists of three demographics questions (*Q78*, *Q81_RelGrps* and *State*). Ethnicity, religions and location surely have tight connection.

All the sameness/difference questions (*Q38*, *Q39*, *Q41* and *Q42*) are placed in cluster 6, along with preferred representational system question *Q9*. An interpretation of why *Q9* is placed in this cluster is that *Q9* is about buying cars, like what *Q41* asks here. Anyway, the center *Q41* is from sameness/difference questions.

The seventh cluster includes all the experience of time questions (*Q66*, *Q73*, *Q75*, *Q76* and *Q77*), two perceptual positions questions (*Q58*, *Q59*) and one towards/away question (*Q17*). The experience of time questions are almost the

same, therefore I strongly recommend to retain only one of them. For example, we can use the cluster center *Q66*, a typical experience of time question as the representative. *Q58* and *Q59* are perceptual position questions which ask about one's attitude towards one's own misfortune. We notice that other perceptual position questions, which are placed in cluster 9, all ask about one's attitude towards others' misfortune. *Q59* asks about the past experience, therefore it is correlated with the experience of time questions. It seems *Q17* also relates to lonely feeling, and perhaps this can be considered related to one's own misfortune. Anyway, this cluster is mainly about experience of time and the attitude towards unpleasant experience.

The eighth cluster consists of three proactive/reactive questions (*Q47*, *Q52* and *Q53*), one towards/away question (*Q20*) and one internal/external question (*Q24*). The cluster center is *Q47*, a proactive/reactive question. *Q24* and *Q20* are somewhat related to the proactive/reactive idea.

The ninth cluster includes three internal/external questions (*Q26*, *Q30*, *Q31*) and three perceptual positions questions (*Q56*, *Q61*, *Q63*). As mentioned above, the perceptual positions questions are all asking people's attitude towards their friends' misfortune. Perhaps *Q26*, *Q30* and *Q31* are correlated with these questions. Anyway, the cluster center *Q61* is a perceptual positions question.

The last cluster consists of one options/procedures question *Q36* and one specific/general question *Q43*. In fact I'm not quite sure about the relationship between these two questions here. However, users' selection shows these two questions are in fact strongly correlated. Either question can be a representative of this cluster.

Cluster	Focus
1	Preferred Representational System Questions about Communication
2	Ways to Solve a Problem
3	Do Planning for Travelling or Not
4	Demographics Except for Ethnicity, Religions and Location
5	Ethnicity, Religions and Location
6	Sameness / Difference Questions
7	Experience of Time, Unpleasant Experience
8	Proactive / Reactive Questions
9	Internal / External Questions and Perceptual Positions Questions
10	Q36 and Q43

Table 5.3: Summarization of Clusters by P-value

The left table can sometimes help us understand the relationship of questions through the relationship of their responses. For example, the internal/external questions ($Q26$, $Q30$, $Q31$) in the above ninth cluster are in fact in the second cluster of the left table of *Table 3.4* and *Table 5.2*. We try to understand their relationship with $Q56$, $Q61$ and $Q63$ in the above analysis, but the relationship seems vague. We can see that $Q61$ is also placed in the second cluster of the left table. So the responses of these internal/external questions are correlated with the responses of $Q61$.

We can often observe that the correlated questions in the right table have correlated responses in the left table. But the left table cannot explain everything. For example, we still do not know how to explain the relationship between $Q36$ and $Q43$ using the relationship between their responses, since their responses are not in the same cluster in the left table.

The advantage of clustering responses is that it helps us understand the user re-

sponse patterns. But one problem is that when the number of responses to the same question is large, the correlations of these responses will become weak and as a result separate the responses into different clusters. This can sometimes make interpretation hard, since we are picking questions rather than question responses.

The results of *Table 5.1* are worth examining closely. The Table shows us that perhaps specific/general questions are not important at all, since both of these two methods ignore them. We already observe specific/general questions appear with options/procedures questions together in the same cluster in the right table of *Table 5.2* again and again. So when we ask options/procedures questions, it is probably not necessary to ask the specific/general ones.

All in all, clusters in the right table can be summarized as *Table 5.3*. After our analysis, we find that we can generally trust the cluster center method in picking questions. More domain knowledge is needed to decide a better way of selection and whether to pick more than one question from the same cluster. As a result, we only need to provide the website with these clusters, make a question selection suggestion according to the cluster centers and let them decide which questions are essential from their point of view.

CHAPTER 6

Discussion

Our goal is to pick questions from the question list provided by MyTherapist-Match.com.

We first implement hierarchical clustering on both question responses and questions themselves. After the clusters are generated, cluster centers are regarded as a suggestion for question selection. We find that question clustering performs well, since different clusters have their own topics and we are in fact selecting representative questions from all the topics.

There are two advantages of hierarchical clustering used here. First, it is an intuitive way to illustrate the relationship of questions. In this way, we only need to study the relationships between questions within the same cluster, rather than study all the relationships. Second, if the website implement a weighting strategy for questions, this method suggests a possible way to update the weights by combining weights of questions in the same cluster as the weight of the cluster representative. This can perhaps maintain the original matching results while reducing the number of survey questions at the same time.

We can improve the selection results of the above hierarchical clustering method in the future. After further research in the psychological concepts, we can then decide whether to pick one or more questions from the same cluster. Anyway,

the statistical approach above only suggests one way to pick the questions. There are many other issues that we need to consider in practice. For example, if the website need to record all the demographics information of users, then we do not need to pick a representative from the demographics questions.

For the linear regression part, in fact when user satisfaction data is not available, user action might serve as a good approximation for satisfaction. When a user is active, it means the user is somewhat interested in the matching results returned. The significant questions produced in this way might be important for user satisfaction. Most of the significant questions are demographics questions. However, the linear model here is not a very good model for prediction. There are still many factors left for study to improve the model.

APPENDIX A

Linear Regression Model

Residuals:

Min	1Q	Median	3Q	Max
-24.427	-7.074	-2.651	4.283	157.450

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	18.634786	2.743604	6.792	1.29e-11	***
Q310	0.752377	0.787980	0.955	0.33973	
Q311	0.232458	0.803176	0.289	0.77227	
Q312	0.673038	0.686063	0.981	0.32665	
Q414	-0.152581	0.762404	-0.200	0.84139	
Q415	0.483599	1.127774	0.429	0.66809	
Q416	-1.188135	0.760387	-1.563	0.11825	
Q725	0.386329	0.625374	0.618	0.53677	
Q726	-0.767083	0.771894	-0.994	0.32040	
Q727	-0.416098	0.530967	-0.784	0.43329	
Q829	-0.413658	0.560676	-0.738	0.46069	
Q830	0.234070	0.542689	0.431	0.66627	
Q831	-1.173431	0.996715	-1.177	0.23915	
Q933	-1.098200	0.629665	-1.744	0.08123	.
Q934	0.228979	0.522031	0.439	0.66096	

Q935	-1.007006	1.894440	-0.532	0.59506	
Q1037	-0.938787	1.032766	-0.909	0.36341	
Q1038	-1.044891	0.747241	-1.398	0.16210	
Q1039	-0.642860	0.744679	-0.863	0.38805	
Q1245	-0.872944	0.649731	-1.344	0.17918	
Q1246	-0.149822	0.625574	-0.239	0.81074	
Q1247	-0.433746	0.700125	-0.620	0.53561	
Q1557	-2.549336	1.839541	-1.386	0.16588	
Q1558	0.488200	0.490185	0.996	0.31934	
Q1559	-0.090477	0.771686	-0.117	0.90667	
Q1765	-1.501791	0.578492	-2.596	0.00947	**
Q1766	0.273298	0.709282	0.385	0.70003	
Q1767	-0.895110	0.759923	-1.178	0.23892	
Q1869	-0.774329	0.541343	-1.430	0.15269	
Q1870	-0.934983	0.642486	-1.455	0.14569	
Q1871	-0.207720	0.998789	-0.208	0.83526	
Q1973	1.153589	0.997133	1.157	0.24739	
Q1974	0.722305	0.882604	0.818	0.41320	
Q1975	0.486844	0.925507	0.526	0.59890	
Q2077	-0.382727	0.645864	-0.593	0.55350	
Q2078	-0.172930	0.552785	-0.313	0.75443	
Q2079	1.423161	0.735141	1.936	0.05296	.
Q2493	-0.399034	0.476256	-0.838	0.40217	
Q2697	-0.027259	0.971818	-0.028	0.97762	
Q30105	-0.137263	0.622620	-0.220	0.82552	
Q31107	0.813260	0.831409	0.978	0.32806	
Q32109	-0.101040	0.458146	-0.221	0.82546	
Q33111	-0.171044	0.464082	-0.369	0.71247	

Q34113	0.029374	0.474300	0.062	0.95062
Q36117	-0.788324	0.581522	-1.356	0.17531
Q38121	0.538194	0.477355	1.127	0.25963
Q39123	0.775305	0.466650	1.661	0.09672 .
Q41127	-1.005697	0.459831	-2.187	0.02880 *
Q42129	0.064765	0.462287	0.140	0.88859
Q43131	0.006272	0.478278	0.013	0.98954
Q45135	-0.083250	0.491717	-0.169	0.86557
Q46137	-0.268679	0.489931	-0.548	0.58345
Q47139	-0.402451	0.505878	-0.796	0.42635
Q49143	0.118830	0.473710	0.251	0.80194
Q52149	0.547553	0.524373	1.044	0.29646
Q53151	0.142547	0.461931	0.309	0.75765
Q56157	-0.064219	0.662501	-0.097	0.92278
Q56158	-0.644476	0.694520	-0.928	0.35350
Q58163	-0.427474	0.510358	-0.838	0.40231
Q58164	-0.326135	0.770741	-0.423	0.67222
Q59166	-0.154734	0.797742	-0.194	0.84621
Q59167	-0.636445	0.493679	-1.289	0.19742
Q61172	-0.176955	0.881077	-0.201	0.84084
Q61173	0.020251	0.997200	0.020	0.98380
Q63178	0.239092	0.539686	0.443	0.65778
Q63179	-0.724307	0.956961	-0.757	0.44917
Q66187	-0.692960	0.642516	-1.079	0.28088
Q66188	-0.083583	0.534465	-0.156	0.87574
Q73208	0.367427	0.502068	0.732	0.46432
Q73209	0.752405	0.910449	0.826	0.40863
Q75214	-0.332256	0.787968	-0.422	0.67330

Q75215	-0.758078	0.693058	-1.094	0.27411
Q76217	-0.288817	0.682068	-0.423	0.67200
Q76218	0.616021	0.580440	1.061	0.28863
Q77220	-0.549260	0.550333	-0.998	0.31832
Q77221	-0.186664	0.657038	-0.284	0.77635
Q78223	1.637438	0.811616	2.018	0.04372 *
Q78224	-1.359142	0.967834	-1.404	0.16031
Q78225	0.446751	1.262025	0.354	0.72336
Q78226	3.984122	2.588057	1.539	0.12379
Q78227	-3.838957	3.295495	-1.165	0.24413
Q78228	0.241211	1.981142	0.122	0.90310
Q78229	-2.557781	2.236770	-1.144	0.25290
Q78230	-2.400200	2.439248	-0.984	0.32519
Q78231	-3.991858	3.384226	-1.180	0.23826
Q78232	0.468272	1.104101	0.424	0.67150
Q78233	1.706941	1.009588	1.691	0.09098 .
Q79235	0.395096	0.907767	0.435	0.66341
Q79236	-1.134529	0.949926	-1.194	0.23243
Q79237	3.158435	3.957698	0.798	0.42490
Q79238	1.921845	2.195199	0.875	0.38137
Q79239	5.968988	1.657047	3.602	0.00032 ***
Q79240	-1.586604	0.873753	-1.816	0.06948 .
Q80242	-1.493458	1.040789	-1.435	0.15140
Q80243	3.358131	2.992132	1.122	0.26180
Q80244	0.924313	0.635771	1.454	0.14608
Q80245	2.815875	1.306830	2.155	0.03125 *
Q80313	0.411140	0.632720	0.650	0.51586
Q80314	2.058294	1.310497	1.571	0.11636

Q81_RelGrps251	-2.082102	1.158532	-1.797	0.07239	.
Q81_RelGrps253	-2.450058	1.082316	-2.264	0.02365	*
Q81_RelGrps259	-4.308572	3.141952	-1.371	0.17037	
Q81_RelGrps260	1.312376	2.921869	0.449	0.65335	
Q81_RelGrps263	-0.493908	1.375890	-0.359	0.71964	
Q81_RelGrps271	-0.916390	1.269470	-0.722	0.47042	
Q81_RelGrps274	-4.088028	1.887694	-2.166	0.03041	*
Q81_RelGrps275	-2.888982	1.184056	-2.440	0.01474	*
Q81_RelGrps414	-2.543865	1.117116	-2.277	0.02283	*
Q81_RelGrps415	-2.746331	1.326809	-2.070	0.03854	*
Q84301	-0.017102	0.609274	-0.028	0.97761	
Q84302	-0.554697	1.133925	-0.489	0.62474	
Q85304	-1.005726	0.505952	-1.988	0.04691	*
Q85305	-0.511892	1.040757	-0.492	0.62286	
Q86307	0.304692	0.572402	0.532	0.59455	
Q86308	-0.372601	0.584491	-0.637	0.52385	
Q86309	-2.341110	1.278664	-1.831	0.06720	.
Q86310	-0.256993	1.245727	-0.206	0.83657	
Q87_Age	0.024418	0.024691	0.989	0.32275	
StateNon-CA	-9.697659	0.463473	-20.924	< 2e-16	***
Q92TRUE	5.048887	0.497154	10.156	< 2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.9 on 3566 degrees of freedom

(75 observations deleted due to missingness)

Multiple R-squared: 0.1847, Adjusted R-squared: 0.1575

F-statistic: 6.789 on 119 and 3566 DF, p-value: < 2.2e-16

APPENDIX B

Survey Questions and Their Domains

Preferred Representational System	Q3
	Q4
	Q7
	Q8
	Q9
	Q10
	Q12

Specific/General	Q43
	Q45
	Q46

Proactive/ Reactive	Q47
	Q49
	Q52
	Q53

Towards/Away	Q15
	Q17
	Q18
	Q19
	Q20

Perceptual Positions	Q56
	Q58
	Q59
	Q61
	Q63

Internal/External	Q24
	Q26
	Q30
	Q31

Experience of Time	Q66
	Q73
	Q75
	Q76
	Q77

Options/ Procedures	Q32
	Q33
	Q34
	Q36

Demographics	Q78
	Q79
	Q80
	Q81_RelGrps
	Q84
	Q85
	Q86
	Q87_Age
	State
	Q92

Sameness/ Difference	Q38
	Q39
	Q41
	Q42

Survey Questions Ordered by QuestionId						
Cluster # 1	Cluster # 2	QuestionId	Question	AnswerId	Answer	AnswerType
9	2	Q3	When solving a problem, I tend to:	9	look at the big picture.	VISU
9	2	Q3	When solving a problem, I tend to:	10	consult with someone about it.	NEUT
9	2	Q3	When solving a problem, I tend to:	11	get in touch with my deeper self.	KINE
9	2	Q3	When solving a problem, I tend to:	12	talk it over with myself or another person.	AUDI
11	1	Q4	I tend to communicate best with:	13	the volume and tone of my voice.	AUDI
11	1	Q4	I tend to communicate best with:	14	logic.	NEUT
11	1	Q4	I tend to communicate best with:	15	the way I look.	VISU
11	1	Q4	I tend to communicate best with:	16	my emotions.	KINE
12	2	Q7	I accomplish my work more easily if I:	24	clearly see what is wanted.	VISU
12	2	Q7	I accomplish my work more easily if I:	25	have a feeling for what is required.	KINE
12	2	Q7	I accomplish my work more easily if I:	26	talk with myself about what is needed.	AUDI
12	2	Q7	I accomplish my work more easily if I:	27	get instructions about what is wanted.	NEUT
12	1	Q8	It is easy to understand a presentation if:	28	I have hands-on experience.	KINE
12	1	Q8	It is easy to understand a presentation if:	29	visual aids are used.	VISU
12	1	Q8	It is easy to understand a presentation if:	30	it is based on logically presented ideas.	NEUT
12	1	Q8	It is easy to understand a presentation if:	31	the speaker emphasizes with tone and volume.	AUDI
10	6	Q9	I buy a car based on:	32	my thoughts about the price, mpg, and safety features.	NEUT
10	6	Q9	I buy a car based on:	33	how it feels.	KINE
10	6	Q9	I buy a car based on:	34	its color, shape and look.	VISU
10	6	Q9	I buy a car based on:	35	the sound of the engine, the stereo system or how quiet it is.	AUDI
7	1	Q10	When talking with someone, I mostly notice:	36	whether or not the person is logical.	NEUT
7	1	Q10	When talking with someone, I mostly notice:	37	the person's tone of voice.	AUDI
7	1	Q10	When talking with someone, I mostly notice:	38	how I feel about the person.	KINE
7	1	Q10	When talking with someone, I mostly notice:	39	the person's body language and their point of view.	VISU
2	1	Q12	I am good at:	44	seeing the big picture.	VISU
2	1	Q12	I am good at:	45	understanding new facts and data.	NEUT
2	1	Q12	I am good at:	46	listening for what is right.	AUDI
2	1	Q12	I am good at:	47	embracing my feelings.	KINE
1	4	Q15	If I were to exercise, I would do so in order to:	56	improve my health.	TOWA
1	4	Q15	If I were to exercise, I would do so in order to:	58	get fit.	TOWA
1	4	Q15	If I were to exercise, I would do so in order to:	59	avoid criticism from others.	AWAY
10	4	Q15	If I were to exercise, I would do so in order to:	57	avoid injury.	AWAY
11	7	Q17	I seek personal relationships, in order to:	64	enjoy another's company.	TOWA
11	7	Q17	I seek personal relationships, in order to:	65	not be lonely.	AWAY
11	7	Q17	I seek personal relationships, in order to:	66	have my needs met.	TOWA
11	7	Q17	I seek personal relationships, in order to:	67	avoid isolation.	AWAY
6	4	Q18	I brush my teeth to:	68	keep them healthy.	TOWA
6	4	Q18	I brush my teeth to:	69	avoid getting cavities.	AWAY
6	4	Q18	I brush my teeth to:	70	have a bright smile.	TOWA
6	4	Q18	I brush my teeth to:	71	reduce the chance of illness.	AWAY
3	4	Q19	When I wear my seatbelt, I do so to:	72	conform to the law.	TOWA
3	4	Q19	When I wear my seatbelt, I do so to:	73	avoid a ticket.	AWAY
3	4	Q19	When I wear my seatbelt, I do so to:	74	be safe.	TOWA
3	4	Q19	When I wear my seatbelt, I do so to:	75	protect myself from injury.	AWAY
2	8	Q20	What is likely to motivate you more?	76	working toward a goal	TOWA
2	8	Q20	What is likely to motivate you more?	77	avoiding failure	AWAY
2	8	Q20	What is likely to motivate you more?	78	achievement	TOWA
2	8	Q20	What is likely to motivate you more?	79	fear of loss	AWAY
2	8	Q24	I know I've done a good job when:	92	someone lets me know.	EXTE
2	8	Q24	I know I've done a good job when:	93	I notice it myself.	INTE
2	9	Q26	When buying new clothes, I tend to buy whatever:	96	looks and/or feels right to me.	INTE
2	9	Q26	When buying new clothes, I tend to buy whatever:	97	my friends will probably like.	EXTE
2	9	Q30	I know I am right when:	104	I feel it in my gut.	INTE
2	9	Q30	I know I am right when:	105	others tell me so.	EXTE
2	9	Q31	If I were to dance, I would do so:	106	to be seen.	EXTE
2	9	Q31	If I were to dance, I would do so:	107	because it feels good.	INTE
12	2	Q32	When solving a problem, I prefer:	108	many alternatives.	OPTI
12	2	Q32	When solving a problem, I prefer:	109	a step-by-step method.	PROC
12	2	Q33	When cooking a meal, I tend to:	110	deviate from the recipe.	OPTI

Survey Questions Ordered by QuestionId						
Cluster # 1	Cluster # 2	QuestionId	Question	AnswerId	Answer	AnswerType
12	2	Q33	When cooking a meal, I tend to:	111	follow the recipe.	PROC
10	3	Q34	When planning a vacation, I prefer to:	112	create a detailed itinerary.	PROC
10	3	Q34	When planning a vacation, I prefer to:	113	figure out what to do when I arrive.	OPTI
12	10	Q36	If I were to buy a bird house that required assembly, I would	116	follow the instructions.	PROC
12	10	Q36	If I were to buy a bird house that required assembly, I would	117	wing it.	OPTI
12	6	Q38	Regarding employment, I prefer to:	120	be with the same employer for life.	SAME
12	6	Q38	Regarding employment, I prefer to:	121	change employers or significantly changes roles within the same company every two to three years.	DIFF
12	6	Q39	I prefer to live:	122	where I have roots.	SAME
12	6	Q39	I prefer to live:	123	in various places, as it suits me.	DIFF
12	6	Q41	When buying a car, I tend to prefer purchasing:	126	the same brand to stay with what works.	SAME
12	6	Q41	When buying a car, I tend to prefer purchasing:	127	a different brand to try something new.	DIFF
12	6	Q42	When going out to eat, I prefer eating at:	128	the same restaurant.	SAME
12	6	Q42	When going out to eat, I prefer eating at:	129	new restaurants.	DIFF
7	10	Q43	I agree with the following statement:	130	After attending a movie, I can tell a friend how the story unfolded.	SPEC
7	10	Q43	I agree with the following statement:	131	After attending a movie, I know if I liked it or not, but can't completely recall how the story unfolded.	GENE
7	2	Q45	I agree with the following statement:	134	I generally prefer thinking about the big picture in life.	GENE
7	2	Q45	I agree with the following statement:	135	I generally prefer thinking about particular details (people, places, things, etc.).	SPEC
10	3	Q46	At a restaurant, when paying the bill, I tend to:	136	review the bill closely, looking at all the details.	SPEC
10	3	Q46	At a restaurant, when paying the bill, I tend to:	137	just pay it.	GENE
2	8	Q47	When involved in a misunderstanding, I	138	take initiative to solve the problem.	PROA
2	8	Q47	When involved in a misunderstanding, I	139	wait for the other person(s) to approach me.	REAC
10	3	Q49	When traveling with someone, I:	142	usually do the planning/organizing.	PROA
10	3	Q49	When traveling with someone, I:	143	let others do the planning/organizing.	REAC
2	8	Q52	When at work, I tend to:	148	be a self starter.	PROA
2	8	Q52	When at work, I tend to:	149	wait for direction from others.	REAC
8	8	Q53	When in an intimate relationship, I tend to:	150	be the first to express my feelings.	PROA
8	8	Q53	When in an intimate relationship, I tend to:	151	let the other person express him/her feelings first.	REAC
8	9	Q56	When expressing sympathy to someone who has lost a loved one, I feel:	156	my own sorrow.	1ST
8	9	Q56	When expressing sympathy to someone who has lost a loved one, I feel:	157	the other person's sorrow.	2ND
8	9	Q56	When expressing sympathy to someone who has lost a loved one, I feel:	158	that the other's loss is unfortunate.	3RD
10	7	Q58	When I watch a sad movie, I:	162	feel sad about my life.	1ST
10	7	Q58	When I watch a sad movie, I:	163	feel sad for the characters in the movie.	2ND
10	7	Q58	When I watch a sad movie, I:	164	remind myself that it is just a movie.	3RD
10	7	Q59	When I think of a painful event from my past, I:	165	relive my feelings as though it were happening now.	1ST
10	7	Q59	When I think of a painful event from my past, I:	166	think of the suffering the other person(s) went through.	2ND
10	7	Q59	When I think of a painful event from my past, I:	167	observe that event from a distance.	3RD
2	9	Q61	When a friend gets injured, I:	171	think of my own pain.	1ST
7	9	Q61	When a friend gets injured, I:	172	imagine his/her pain.	2ND
7	9	Q61	When a friend gets injured, I:	173	mentally remove myself.	3RD
7	9	Q63	When someone complains about a pain I've never experienced, I:	177	think I'm lucky that it didn't happen to me.	1ST
7	9	Q63	When someone complains about a pain I've never experienced, I:	178	try to imagine what he/she must be going through	2ND
7	9	Q63	When someone complains about a pain I've never experienced, I:	179	think its time for him/her to get over it.	3RD
8	7	Q66	I often think about what:	186	I did in the past	PAST
8	7	Q66	I often think about what:	187	I'm doing right now	PRES
8	7	Q66	I often think about what:	188	I'll be doing in the future	FUTU
8	7	Q73	I often think about people I:	207	used to know.	PAST
8	7	Q73	I often think about people I:	208	currently know.	PRES
8	7	Q73	I often think about people I:	209	want to know in the future.	FUTU
8	7	Q75	I often think about things I:	213	used to have.	PAST
8	7	Q75	I often think about things I:	214	have now.	PRES
8	7	Q75	I often think about things I:	215	want to have in the future.	FUTU
8	7	Q76	I often think about activities I:	216	used to engage in.	PAST

Survey Questions Ordered by QuestionId						
Cluster # 1	Cluster # 2	QuestionId	Question	AnswerId	Answer	AnswerType
8	7	Q76	I often think about activities I:	217	do now.	PRES
8	7	Q76	I often think about activities I:	218	want to do in the future.	FUTU
8	7	Q77	I often think about what I:	219	learned in the past.	PAST
8	7	Q77	I often think about what I:	220	am learning now.	PRES
8	7	Q77	I often think about what I:	221	will learn in the future.	FUTU
1	5	Q78	My ethnicity (select one)	222	White, non-Hispanic	
1	5	Q78	My ethnicity (select one)	223	Hispanic or Latino	
1	5	Q78	My ethnicity (select one)	225	Asian/Pacific Islander	
1	5	Q78	My ethnicity (select one)	231	Native American	
1	5	Q78	My ethnicity (select one)	232	Other	
3	5	Q78	My ethnicity (select one)	233	No comment	
4	5	Q78	My ethnicity (select one)	224	African-American	
4	5	Q78	My ethnicity (select one)	228	Chinese	
5	5	Q78	My ethnicity (select one)	229	Indian	
6	5	Q78	My ethnicity (select one)	226	Korean	
9	5	Q78	My ethnicity (select one)	227	Japanese	
9	5	Q78	My ethnicity (select one)	230	Arab	
1	4	Q79	I am (select one)	239	Celibate	
3	4	Q79	I am (select one)	234	Heterosexual	
3	4	Q79	I am (select one)	235	Homosexual	
3	4	Q79	I am (select one)	240	No comment	
4	4	Q79	I am (select one)	236	Bi-Sexual	
4	4	Q79	I am (select one)	237	Transgendered	
6	4	Q79	I am (select one)	238	Nonsexual	
3	4	Q80	I am (select one)	245	No comment	
4	4	Q80	I am (select one)	241	Married	
4	4	Q80	I am (select one)	242	Divorced	
4	4	Q80	I am (select one)	244	Single	
4	4	Q80	I am (select one)	313	In a relationship	
4	4	Q80	I am (select one)	314	Separated	
6	4	Q80	I am (select one)	243	Widowed	
5	5	Q81_RelGr	I identify with the following religion(s) / spirituality	251	Catholic	
5	5	Q81_RelGr	I identify with the following religion(s) / spirituality	253	Christian	
5	5	Q81_RelGr	I identify with the following religion(s) / spirituality	259	Hindu	
5	5	Q81_RelGr	I identify with the following religion(s) / spirituality	263	Jewish	
5	5	Q81_RelGr	I identify with the following religion(s) / spirituality	271	Spiritual	
5	5	Q81_RelGr	I identify with the following religion(s) / spirituality	274	Not listed	
5	5	Q81_RelGr	I identify with the following religion(s) / spirituality	275	No comment	
5	5	Q81_RelGr	I identify with the following religion(s) / spirituality	414	Agnostic	
5	5	Q81_RelGr	I identify with the following religion(s) / spirituality	415	Atheist	
9	5	Q81_RelGr	I identify with the following religion(s) / spirituality	249	Buddhist	
9	5	Q81_RelGr	I identify with the following religion(s) / spirituality	260	Islamic	
3	4	Q84	I smoke cigarettes? (select one)	300	Yes	
3	4	Q84	I smoke cigarettes? (select one)	301	No	
3	4	Q84	I smoke cigarettes? (select one)	302	No Comment	
1	4	Q85	I drink alcohol? (select one)	303	Yes	
1	4	Q85	I drink alcohol? (select one)	304	No	
3	4	Q85	I drink alcohol? (select one)	305	No Comment	
3	4	Q86	I exercise (select one):	310	No comment	
4	4	Q86	I exercise (select one):	309	7 days a week	
6	4	Q86	I exercise (select one):	306	Rarely	
6	4	Q86	I exercise (select one):	307	Sometime	
6	4	Q86	I exercise (select one):	308	Frequently	
4		Q87_Age	Age			
4	4	Q92FALSE	Include therapists who offer online and/or tele-sessions		FALSE	
4	4	Q92TRUE	Include therapists who offer online and/or tele-sessions		TRUE	
1	5		State		Non-CA	
1	5		State		CA	

Survey Questions Ordered by Cluster # 2						
Cluster # 1	Cluster # 2	QuestionId	Question	AnswerId	Answer	AnswerType
11	1	Q4	I tend to communicate best with:	13	the volume and tone of my voice.	AUDI
11	1	Q4	I tend to communicate best with:	14	logic.	NEUT
11	1	Q4	I tend to communicate best with:	15	the way I look.	VISU
11	1	Q4	I tend to communicate best with:	16	my emotions.	KINE
12	1	Q8	It is easy to understand a presentation if:	28	I have hands-on experience.	KINE
12	1	Q8	It is easy to understand a presentation if:	29	visual aids are used.	VISU
12	1	Q8	It is easy to understand a presentation if:	30	it is based on logically presented ideas.	NEUT
12	1	Q8	It is easy to understand a presentation if:	31	the speaker emphasizes with tone and volume.	AUDI
7	1	Q10	When talking with someone, I mostly notice:	36	whether or not the person is logical.	NEUT
7	1	Q10	When talking with someone, I mostly notice:	37	the person's tone of voice.	AUDI
7	1	Q10	When talking with someone, I mostly notice:	38	how I feel about the person.	KINE
7	1	Q10	When talking with someone, I mostly notice:	39	the person's body language and their point of view.	VISU
2	1	Q12	I am good at:	44	seeing the big picture.	VISU
2	1	Q12	I am good at:	45	understanding new facts and data.	NEUT
2	1	Q12	I am good at:	46	listening for what is right.	AUDI
2	1	Q12	I am good at:	47	embracing my feelings.	KINE
9	2	Q3	When solving a problem, I tend to:	9	look at the big picture.	VISU
9	2	Q3	When solving a problem, I tend to:	10	consult with someone about it.	NEUT
9	2	Q3	When solving a problem, I tend to:	11	get in touch with my deeper self.	KINE
9	2	Q3	When solving a problem, I tend to:	12	talk it over with myself or another person.	AUDI
12	2	Q7	I accomplish my work more easily if I:	24	clearly see what is wanted.	VISU
12	2	Q7	I accomplish my work more easily if I:	25	have a feeling for what is required.	KINE
12	2	Q7	I accomplish my work more easily if I:	26	talk with myself about what is needed.	AUDI
12	2	Q7	I accomplish my work more easily if I:	27	get instructions about what is wanted.	NEUT
12	2	Q32	When solving a problem, I prefer:	108	many alternatives.	OPTI
12	2	Q32	When solving a problem, I prefer:	109	a step-by-step method.	PROC
12	2	Q33	When cooking a meal, I tend to:	110	deviate from the recipe.	OPTI
12	2	Q33	When cooking a meal, I tend to:	111	follow the recipe.	PROC
7	2	Q45	I agree with the following statement:	134	I generally prefer thinking about the big picture in life.	GENE
7	2	Q45	I agree with the following statement:	135	I generally prefer thinking about particular details (people, places, things, etc.).	SPEC
10	3	Q34	When planning a vacation, I prefer to:	112	create a detailed itinerary.	PROC
10	3	Q34	When planning a vacation, I prefer to:	113	figure out what to do when I arrive.	OPTI
10	3	Q46	At a restaurant, when paying the bill, I tend to:	136	review the bill closely, looking at all the details.	SPEC
10	3	Q46	At a restaurant, when paying the bill, I tend to:	137	just pay it.	GENE
10	3	Q49	When traveling with someone, I:	142	usually do the planning/organizing.	PROA
10	3	Q49	When traveling with someone, I:	143	let others do the planning/organizing.	REAC
1	4	Q15	If I were to exercise, I would do so in order to:	56	improve my health.	TOWA
1	4	Q15	If I were to exercise, I would do so in order to:	58	get fit.	TOWA
1	4	Q15	If I were to exercise, I would do so in order to:	59	avoid criticism from others.	AWAY
10	4	Q15	If I were to exercise, I would do so in order to:	57	avoid injury.	AWAY
6	4	Q18	I brush my teeth to:	68	keep them healthy.	TOWA
6	4	Q18	I brush my teeth to:	69	avoid getting cavities.	AWAY
6	4	Q18	I brush my teeth to:	70	have a bright smile.	TOWA
6	4	Q18	I brush my teeth to:	71	reduce the chance of illness.	AWAY
3	4	Q19	When I wear my seatbelt, I do so to:	72	conform to the law.	TOWA
3	4	Q19	When I wear my seatbelt, I do so to:	73	avoid a ticket.	AWAY
3	4	Q19	When I wear my seatbelt, I do so to:	74	be safe.	TOWA
3	4	Q19	When I wear my seatbelt, I do so to:	75	protect myself from injury.	AWAY
1	4	Q79	I am (select one)	239	Celibate	
3	4	Q79	I am (select one)	234	Heterosexual	
3	4	Q79	I am (select one)	235	Homosexual	
3	4	Q79	I am (select one)	240	No comment	
4	4	Q79	I am (select one)	236	Bi-Sexual	
4	4	Q79	I am (select one)	237	Transgendered	
6	4	Q79	I am (select one)	238	Nonsexual	
3	4	Q80	I am (select one)	245	No comment	
4	4	Q80	I am (select one)	241	Married	
4	4	Q80	I am (select one)	242	Divorced	
4	4	Q80	I am (select one)	244	Single	
4	4	Q80	I am (select one)	313	In a relationship	
4	4	Q80	I am (select one)	314	Separated	
6	4	Q80	I am (select one)	243	Widowed	

Survey Questions Ordered by Cluster # 2						
Cluster # 1	Cluster # 2	QuestionId	Question	AnswerId	Answer	AnswerType
3	4	Q84	I smoke cigarettes? (select one)	300	Yes	
3	4	Q84	I smoke cigarettes? (select one)	301	No	
3	4	Q84	I smoke cigarettes? (select one)	302	No Comment	
1	4	Q85	I drink alcohol? (select one)	303	Yes	
1	4	Q85	I drink alcohol? (select one)	304	No	
3	4	Q85	I drink alcohol? (select one)	305	No Comment	
3	4	Q86	I exercise (select one):	310	No comment	
4	4	Q86	I exercise (select one):	309	7 days a week	
6	4	Q86	I exercise (select one):	306	Rarely	
6	4	Q86	I exercise (select one):	307	Sometime	
6	4	Q86	I exercise (select one):	308	Frequently	
4	4	Q92FALSE	Include therapists who offer online and/or tele-sessions		FALSE	
4	4	Q92TRUE	Include therapists who offer online and/or tele-sessions		TRUE	
1	5	Q78	My ethnicity (select one)	222	White, non-Hispanic	
1	5	Q78	My ethnicity (select one)	223	Hispanic or Latino	
1	5	Q78	My ethnicity (select one)	225	Asian/Pacific Islander	
1	5	Q78	My ethnicity (select one)	231	Native American	
1	5	Q78	My ethnicity (select one)	232	Other	
3	5	Q78	My ethnicity (select one)	233	No comment	
4	5	Q78	My ethnicity (select one)	224	African-American	
4	5	Q78	My ethnicity (select one)	228	Chinese	
5	5	Q78	My ethnicity (select one)	229	Indian	
6	5	Q78	My ethnicity (select one)	226	Korean	
9	5	Q78	My ethnicity (select one)	227	Japanese	
9	5	Q78	My ethnicity (select one)	230	Arab	
5	5	Q81_RelGr	I identify with the following religion(s) / spirituality	251	Catholic	
5	5	Q81_RelGr	I identify with the following religion(s) / spirituality	253	Christian	
5	5	Q81_RelGr	I identify with the following religion(s) / spirituality	259	Hindu	
5	5	Q81_RelGr	I identify with the following religion(s) / spirituality	263	Jewish	
5	5	Q81_RelGr	I identify with the following religion(s) / spirituality	271	Spiritual	
5	5	Q81_RelGr	I identify with the following religion(s) / spirituality	274	Not listed	
5	5	Q81_RelGr	I identify with the following religion(s) / spirituality	275	No comment	
5	5	Q81_RelGr	I identify with the following religion(s) / spirituality	414	Agnostic	
5	5	Q81_RelGr	I identify with the following religion(s) / spirituality	415	Atheist	
9	5	Q81_RelGr	I identify with the following religion(s) / spirituality	249	Buddhist	
9	5	Q81_RelGr	I identify with the following religion(s) / spirituality	260	Islamic	
1	5		State		Non-CA	
1	5		State		CA	
10	6	Q9	I buy a car based on:	32	my thoughts about the price, mpg, and safety features.	NEUT
10	6	Q9	I buy a car based on:	33	how it feels.	KINE
10	6	Q9	I buy a car based on:	34	its color, shape and look.	VISU
10	6	Q9	I buy a car based on:	35	the sound of the engine, the stereo system or how quiet it is.	AUDI
12	6	Q38	Regarding employment, I prefer to:	120	be with the same employer for life.	SAME
12	6	Q38	Regarding employment, I prefer to:	121	change employers or significantly changes roles within the same company every two to three years.	DIFF
12	6	Q39	I prefer to live:	122	where I have roots.	SAME
12	6	Q39	I prefer to live:	123	in various places, as it suits me.	DIFF
12	6	Q41	When buying a car, I tend to prefer purchasing:	126	the same brand to stay with what works.	SAME
12	6	Q41	When buying a car, I tend to prefer purchasing:	127	a different brand to try something new.	DIFF
12	6	Q42	When going out to eat, I prefer eating at:	128	the same restaurant.	SAME
12	6	Q42	When going out to eat, I prefer eating at:	129	new restaurants.	DIFF
11	7	Q17	I seek personal relationships, in order to:	64	enjoy another's company.	TOWA
11	7	Q17	I seek personal relationships, in order to:	65	not be lonely.	AWAY
11	7	Q17	I seek personal relationships, in order to:	66	have my needs met.	TOWA
11	7	Q17	I seek personal relationships, in order to:	67	avoid isolation.	AWAY

Survey Questions Ordered by Cluster # 2						
Cluster # 1	Cluster # 2	QuestionId	Question	AnswerId	Answer	AnswerType
10	7	Q58	When I watch a sad movie, I:	162	feel sad about my life.	1ST
10	7	Q58	When I watch a sad movie, I:	163	feel sad for the characters in the movie.	2ND
10	7	Q58	When I watch a sad movie, I:	164	remind myself that it is just a movie.	3RD
10	7	Q59	When I think of a painful event from my past, I:	165	relive my feelings as though it were happening now.	1ST
10	7	Q59	When I think of a painful event from my past, I:	166	think of the suffering the other person(s) went through.	2ND
10	7	Q59	When I think of a painful event from my past, I:	167	observe that event from a distance.	3RD
8	7	Q66	I often think about what:	186	I did in the past	PAST
8	7	Q66	I often think about what:	187	I'm doing right now	PRES
8	7	Q66	I often think about what:	188	I'll be doing in the future	FUTU
8	7	Q73	I often think about people I:	207	used to know.	PAST
8	7	Q73	I often think about people I:	208	currently know.	PRES
8	7	Q73	I often think about people I:	209	want to know in the future.	FUTU
8	7	Q75	I often think about things I:	213	used to have.	PAST
8	7	Q75	I often think about things I:	214	have now.	PRES
8	7	Q75	I often think about things I:	215	want to have in the future.	FUTU
8	7	Q76	I often think about activities I:	216	used to engage in.	PAST
8	7	Q76	I often think about activities I:	217	do now.	PRES
8	7	Q76	I often think about activities I:	218	want to do in the future.	FUTU
8	7	Q77	I often think about what I:	219	learned in the past.	PAST
8	7	Q77	I often think about what I:	220	am learning now.	PRES
8	7	Q77	I often think about what I:	221	will learn in the future.	FUTU
2	8	Q20	What is likely to motivate you more?	76	working toward a goal	TOWA
2	8	Q20	What is likely to motivate you more?	77	avoiding failure	AWAY
2	8	Q20	What is likely to motivate you more?	78	achievement	TOWA
2	8	Q20	What is likely to motivate you more?	79	fear of loss	AWAY
2	8	Q24	I know I've done a good job when:	92	someone lets me know.	EXTE
2	8	Q24	I know I've done a good job when:	93	I notice it myself.	INTE
2	8	Q47	When involved in a misunderstanding, I	138	take initiative to solve the problem.	PROA
2	8	Q47	When involved in a misunderstanding, I	139	wait for the other person(s) to approach me.	REAC
2	8	Q52	When at work, I tend to:	148	be a self starter.	PROA
2	8	Q52	When at work, I tend to:	149	wait for direction from others.	REAC
8	8	Q53	When in an intimate relationship, I tend to:	150	be the first to express my feelings.	PROA
8	8	Q53	When in an intimate relationship, I tend to:	151	let the other person express him/her feelings first.	REAC
2	9	Q26	When buying new clothes, I tend to buy whatever:	96	looks and/or feels right to me.	INTE
2	9	Q26	When buying new clothes, I tend to buy whatever:	97	my friends will probably like.	EXTE
2	9	Q30	I know I am right when:	104	I feel it in my gut.	INTE
2	9	Q30	I know I am right when:	105	others tell me so.	EXTE
2	9	Q31	If I were to dance, I would do so:	106	to be seen.	EXTE
2	9	Q31	If I were to dance, I would do so:	107	because it feels good.	INTE
8	9	Q56	When expressing sympathy to someone who has lost a loved one, I feel:	156	my own sorrow.	1ST
8	9	Q56	When expressing sympathy to someone who has lost a loved one, I feel:	157	the other person's sorrow.	2ND
8	9	Q56	When expressing sympathy to someone who has lost a loved one, I feel:	158	that the other's loss is unfortunate.	3RD
2	9	Q61	When a friend gets injured, I:	171	think of my own pain.	1ST
7	9	Q61	When a friend gets injured, I:	172	imagine his/her pain.	2ND
7	9	Q61	When a friend gets injured, I:	173	mentally remove myself.	3RD
7	9	Q63	When someone complains about a pain I've never experienced, I:	177	think I'm lucky that it didn't happen to me.	1ST
7	9	Q63	When someone complains about a pain I've never experienced, I:	178	try to imagine what he/she must be going through	2ND
7	9	Q63	When someone complains about a pain I've never experienced, I:	179	think its time for him/her to get over it.	3RD
12	10	Q36	If I were to buy a bird house that required assembly, I would	116	follow the instructions.	PROC
12	10	Q36	If I were to buy a bird house that required assembly, I would	117	wing it.	OPTI
7	10	Q43	I agree with the following statement:	130	After attending a movie, I can tell a friend how the story unfolded.	SPEC
7	10	Q43	I agree with the following statement:	131	After attending a movie, I know if I liked it or not, but can't completely recall how the story unfolded.	GENE
4		Q87_Age	Age			

REFERENCES

- [BG75] Richard Bandler and John Grinder. *The Structure of Magic I: A Book About Language and Therapy*. Science and Behavior Books, Palo Alto, CA, 1975.
- [BG83] Richard Bandler and John Grinder. *Reframing: Neurolinguistic programming and the transformation of meaning*. Real People Press, Moab, UT, 1983.
- [Def77] Defays, D. “An efficient algorithm for a complete link method.” *The Computer Journal (British Computer Society)*, **20**(4):364–366, 1977.
- [Dru88] Daniel Druckman. *Enhancing Human Performance: Issues, Theories, and Techniques*. 1988.
- [Gri79] Bandler, R., Grinder, J. *Frogs into Princes: Neuro Linguistic Programming*. Real People Press, Moab, UT, 1979.
- [Jan10] Philipp K. Janert. *Data Analysis with Open Source Tools*. O’Reilly, Sebastopol, CA, 2010.
- [Jor05] Jornsten, R. “PCLust Statistics seminar.” University Lecture, 2005.
- [Rez09] Rezankova, H. “Cluster analysis and categorical data.” *Statistika*, pp. 223–234, 2009.
- [Seg11] Toby Segaran. *Programming Collective Intelligence*. O’Reilly, Sebastopol, CA, 2011.
- [Sib73] R. Sibson. “SLINK: an optimally efficient algorithm for the single-link cluster method.” *The Computer Journal (British Computer Society)*, **16**(1):30–34, 1973.
- [Sta89] R. M. Star. *Foo Bar Baz*. MIT Press, Cambridge, MA, 1989.
- [TF01] Hastie, T., Tibshirani, R. and Friedman, J. *Elements of Statistical Learning*. Springer, New York, 2001.