

Comp. Biology Final Project

An Evolutionary Programming Approach to Principal Component Analysis

Grant Belsterling

Spring 2017

Comp Bio Final Project

An Evolutionary Programming Approach to Principal Component Analysis

Introduction

Evolutionary programming is a method used to solve optimization problems. It is unique in that unlike other common methods of solving optimization problems (such as constraint programming), evolutionary programming is stochastic. This gives it the advantage of requiring no knowledge of the system being optimized other than being able to recognize a “good” solution from a bad one.

Evolutionary programming is based on the metaphor of natural selection. Evolutionary algorithms work by randomly generating a population of valid solutions, culling the weakest solutions (the least fit), and then randomly mutating the surviving, most fit solutions. After each generation, the most fit survive, and we hope that after many generations, an acceptable solution will be selected for.

Since mutations happen randomly, evolutionary programming is particularly applicable when we have no prior knowledge about the shape or structure of an optimized solution. For this reason, I chose to apply an evolutionary algorithm approach to principal component analysis.

Principal component analysis is a type of dimensionality reduction, a way to “simplify” high dimensional data. For example, if we have data with many columns, but we suspect that some columns are dependent on each other, we can eliminate these “redundant” columns, and simplify our model to only those columns that are dependent of each other. Thus, dimensionality reduction uncovers latent, underlying variables that are responsible for variations in multiple columns. We apply evolutionary programming, to randomly generate weights for a component, and then selecting for the fittest components.

The data set used in this dimensionality reduction was the results of the British government’s Department of Education’s “Longitudinal Study of Young People in England.” In the study, participants aged 15-30 were interviewed about

music preferences, movie preferences, hobbies/interests, phobias, health habits, personality traits, views on life, spending habits, and demographic factors. The goal of this is to hopefully uncover some interesting underlying traits that are responsible for variations in a wide spectrum of apparently unrelated behaviors.

Method

As a starting point, the Watchmaker Framework written in Java was utilized in developing the evolutionary algorithm. It was developed by Daniel W. Dyer and has an open license.

Using this framework, I engineered a solution for evolving principal components for the data set (described in the appendix). A principal component is a vector with length equal to the number of variables in the data set, and all entries being between -1 and 1. We can multiply the principal component vector with an observation, which results in a single scalar result. This scalar is a single “reduced dimension,” or a way to describe an entire observation, which has a set of variables, with only one variable.

A key element in evolutionary programming is the fitness function, which is used to assign a numerical value to the fitness of a candidate. In my solution, I defined the fitness of a principal component as the variance of the set of dot products of the principal component with all observations in the data set. Since these dot products are the observations “boiled down” to one scalar, finding a principal component that produces the most variance of these scalars will also produce a principal component that selects for the most “interesting” variables that cause the most variation among observations—which allows us to identify which variables make an observation unique.

A principal component is a new feature of an observation—it expresses an underlying latent variable that causes other variables contained in the data to vary together. Thus, finding a principal component allows us to get a grasp on our data. In this particular example, I was interesting in learning whether neuroticism and mood swings were a symptom of an underlying personality trait, and what other personality traits/interests/factors occur commonly along with neuroticism.

Results

The results of four different runs of the evolutionary algorithm are reproduced below. The key mapping the names of variables below to the full questions asked in the survey are available in the appendix.

| Trait | Trial 1 | Trial 2 | Trial 3 | Trial 4 |
|--------------------------|---------|---------|---------|---------|
| Music | -0.91 | 0.52 | -0.53 | 0.42 |
| Slow.songs.or.fast.songs | 0.21 | -0.28 | 0.52 | -0.21 |
| Dance | 0.01 | 0.71 | -0.66 | 0.78 |
| Folk | -1.00 | 0.85 | -0.82 | 0.87 |
| Country | 0.00 | -0.77 | -0.11 | -0.70 |
| Classical.music | -0.53 | 0.09 | -0.38 | 0.40 |
| Musical | -0.74 | 0.74 | -0.90 | 0.95 |
| Pop | -0.67 | 0.95 | -0.12 | 0.83 |
| Rock | -0.64 | -0.22 | -0.89 | -0.16 |
| Metal.or.Hardrock | 0.50 | -0.55 | 0.91 | -0.97 |
| Punk | 0.61 | -0.95 | 0.44 | -0.94 |
| Hiphop..Rap | 0.31 | -0.88 | 0.05 | -0.50 |
| Reggae..Ska | 0.82 | -0.82 | -0.69 | -0.60 |
| Swing..Jazz | -0.67 | 1.00 | -0.57 | 0.56 |
| Rock.n.roll | -0.90 | -0.76 | -0.61 | 0.95 |
| Alternative | -0.47 | -0.46 | -0.16 | -0.33 |
| Latino | -0.94 | 0.82 | -0.81 | 0.54 |
| Techno..Trance | 0.77 | -0.61 | 0.53 | -0.17 |
| Opera | -0.38 | 0.51 | -0.80 | -0.15 |
| Movies | -0.15 | -0.86 | 0.24 | 0.32 |
| Horror | 0.37 | -0.83 | 0.21 | -0.93 |
| Thriller | 0.15 | -0.73 | 0.00 | -0.67 |

| | | | | |
|-----------------------|-------|-------|-------|-------|
| Comedy | 0.65 | 0.56 | -0.21 | -0.30 |
| Romantic | -0.82 | 0.94 | -0.45 | 0.93 |
| Sci.fi | 0.84 | -0.67 | 0.79 | -0.32 |
| War | 0.96 | -0.84 | 0.93 | -0.96 |
| Fantasy.Fairy.tales | -0.57 | 0.93 | 0.03 | 0.84 |
| Animated | -0.33 | 0.37 | -0.68 | -0.24 |
| Documentary | 0.52 | -0.10 | -0.54 | -0.59 |
| Western | 0.42 | -0.86 | 0.70 | 0.83 |
| Action | 0.87 | -0.51 | 0.81 | -0.81 |
| History | 0.87 | -0.26 | 0.80 | -0.45 |
| Psychology | -0.96 | 0.73 | -0.57 | 0.61 |
| Politics | 0.81 | -0.60 | 0.27 | -0.96 |
| Mathematics | 0.88 | -0.92 | -0.12 | -0.74 |
| Physics | 0.96 | -0.96 | -0.18 | -0.91 |
| Internet | 0.68 | -0.38 | 0.81 | -0.43 |
| PC | 0.90 | -0.89 | 0.91 | -0.96 |
| Economy.Management | 0.49 | 0.57 | 0.49 | -0.26 |
| Biology | 0.17 | 0.26 | -0.92 | 0.22 |
| Chemistry | -0.51 | 0.39 | -0.03 | -0.87 |
| Reading | -0.97 | 0.94 | -0.91 | 0.42 |
| Geography | 0.23 | -0.53 | 0.32 | -0.84 |
| Foreign.languages | -0.84 | 0.82 | -0.87 | 0.88 |
| Medicine | -0.71 | 0.75 | -0.48 | 0.54 |
| Law | -0.08 | -0.09 | -0.74 | -0.14 |
| Cars | 0.61 | 0.03 | 0.23 | -0.75 |
| Art.exhibitions | -0.70 | 0.52 | -0.87 | 0.01 |
| Religion | 0.92 | 0.65 | 0.11 | 0.12 |
| Countryside..outdoors | 0.15 | 0.08 | -0.58 | 0.78 |

| | | | | |
|-----------------------------|-------|-------|-------|-------|
| Dancing | -0.61 | 0.67 | -0.98 | 0.98 |
| Musical.instrum ents | -0.60 | -0.75 | -0.93 | 0.38 |
| Writing | -0.38 | 0.47 | -0.39 | 0.66 |
| Passive.sport | 0.45 | 0.10 | 0.53 | -0.61 |
| Active.sport | 0.96 | -0.99 | 0.65 | -0.70 |
| Gardening | -0.10 | 0.18 | -0.63 | 0.77 |
| Celebrities | -0.94 | 0.27 | 0.25 | 0.08 |
| Shopping | -0.88 | 0.76 | -0.94 | 1.00 |
| Science.and.tech nology | 0.74 | -0.77 | 0.90 | -0.89 |
| Theatre | -0.66 | 0.28 | -0.74 | 0.46 |
| Fun.with.friends | -0.99 | 0.19 | -0.63 | -0.73 |
| Adrenaline.sport s | 0.93 | -0.91 | 0.77 | -0.74 |
| Pets | -0.65 | 0.35 | -0.99 | 0.77 |
| Flying | -0.79 | 0.30 | -0.64 | 0.88 |
| Storm | -0.52 | -0.19 | -0.89 | 0.96 |
| Darkness | -0.93 | 0.83 | -0.96 | 0.75 |
| Heights | 0.04 | 0.85 | -0.79 | -0.02 |
| Spiders | -0.23 | 0.49 | -0.75 | 0.73 |
| Snakes | -0.96 | 0.85 | -0.44 | 0.94 |
| Rats | -0.31 | 0.58 | -0.56 | -0.01 |
| Ageing | -0.89 | 0.94 | -0.68 | 0.62 |
| Dangerous.dogs | -0.41 | 0.35 | -0.59 | 0.95 |
| Fear.of.public.sp eaking | -0.49 | 0.47 | -0.37 | 0.43 |
| Healthy.eating | -0.96 | 0.98 | -0.31 | 0.50 |
| Daily.events | 0.42 | 0.63 | -0.12 | 0.54 |
| Prioritising.workl oad | -0.57 | 0.27 | -0.44 | -0.73 |
| Writing.notes | -0.52 | 0.78 | -0.74 | 0.91 |
| Workaholism | -0.39 | 0.04 | -0.62 | -0.15 |

| | | | | |
|---------------------------|-------|-------|-------|-------|
| Thinking.ahead | -0.05 | 0.49 | -0.44 | -0.85 |
| Final.judgement | -0.64 | 0.78 | -0.86 | -0.03 |
| Reliability | 0.04 | 0.65 | -0.69 | 0.26 |
| Keeping.promise s | -0.45 | 0.05 | -0.99 | -0.25 |
| Loss.of.interest | 0.91 | 0.65 | -0.31 | 0.08 |
| Friends.versus.m oney | -0.12 | 0.42 | -0.49 | 0.84 |
| Funniness | 0.91 | -0.68 | 0.15 | -0.69 |
| Fake | 0.48 | -0.67 | 0.92 | -0.99 |
| Criminal.damage | 0.61 | -0.26 | 0.74 | -0.97 |
| Decision.making | -0.98 | 0.98 | -0.51 | 0.53 |
| Elections | 0.38 | 0.74 | 0.00 | -0.66 |
| Self.criticism | -0.61 | 0.93 | -0.87 | 0.23 |
| Judgment.calls | -0.94 | 0.79 | -0.71 | -0.37 |
| Hypochondria | -0.88 | -0.64 | 0.05 | -0.34 |
| Empathy | -0.68 | 0.83 | -0.65 | 0.76 |
| Eating.to.survive | 1.00 | -0.72 | 0.99 | -0.72 |
| Giving | -0.77 | 0.25 | 0.08 | 0.94 |
| Compassion.to.a nimals | -0.95 | 0.93 | -0.93 | 0.57 |
| Borrowed.stuff | -0.89 | 0.46 | -0.62 | 0.77 |
| Loneliness | -0.65 | 0.75 | 0.22 | 0.90 |
| Cheating.in.scho ol | -0.09 | -0.28 | -0.32 | -0.44 |
| Health | 0.43 | 0.84 | -0.75 | 0.29 |
| Changing.the.pa st | 0.78 | -0.20 | 0.39 | -0.61 |
| God | -0.72 | 0.87 | -0.99 | 0.90 |
| Dreams | 0.41 | -0.34 | 0.48 | -0.31 |
| Charity | 0.06 | 0.74 | -0.53 | -0.14 |
| Number.of.friend s | -0.97 | -0.93 | -0.55 | -0.92 |

| | | | | |
|--------------------------------|-------|-------|-------|-------|
| New.environment | 0.58 | -0.32 | 0.29 | -0.54 |
| Mood.swings | -0.83 | 0.60 | -0.56 | 0.97 |
| Appearance.and.gestures | -0.66 | 0.41 | -0.41 | 0.99 |
| Socializing | 0.97 | -0.46 | -0.93 | -0.87 |
| Achievements | 0.87 | 0.66 | 0.98 | -0.08 |
| Responding.to.a.serious.letter | -0.51 | -0.82 | 0.61 | 0.19 |
| Children | -0.63 | 0.66 | -0.85 | 0.07 |
| Assertiveness | 0.12 | -0.33 | 0.67 | -0.93 |
| Getting.angry | -0.49 | 0.86 | -0.71 | -0.35 |
| Knowing.the.right.people | 0.62 | -0.59 | 0.33 | -0.79 |
| Public.speaking | -0.07 | 0.91 | -0.84 | 0.89 |
| Unpopularity | 0.21 | 0.89 | -0.03 | -0.27 |
| Life.struggles | -0.99 | 0.76 | -0.73 | 0.87 |
| Happiness.in.life | 0.46 | 0.24 | 0.44 | 0.35 |
| Energy.levels | 0.64 | 0.04 | -0.49 | -0.73 |
| Small...big.dogs | 0.17 | -0.61 | 0.56 | -0.82 |
| Personality | -0.06 | -0.99 | 0.56 | -0.66 |
| Finding.lost.valuables | -0.63 | 0.25 | 0.37 | 1.00 |
| Getting.up | -0.05 | 0.99 | -0.59 | -0.15 |
| Interests.or.hobbies | -0.30 | -0.16 | -0.41 | -0.57 |
| Parents..advice | -0.53 | -0.06 | -1.00 | 0.19 |
| Questionnaires.or.polls | -0.98 | 0.82 | 0.58 | -0.12 |
| Finances | -0.77 | 0.88 | -0.55 | 0.40 |
| Shopping.centres | -0.52 | 0.04 | -0.10 | 0.87 |
| Branded.clothing | -0.26 | -0.67 | -0.70 | -0.74 |
| Entertainment.spend | 0.82 | -0.67 | 0.10 | -0.50 |

| | | | | |
|--------------------------------|-------|-------|-------|-------|
| Spending.on.loo ks | -0.72 | 0.55 | -0.33 | 0.95 |
| Spending.on.gad gets | 0.72 | -0.17 | 0.86 | -0.86 |
| Spending.on.hea lthy.eating | -0.12 | -0.01 | 0.36 | 0.29 |
| Age | -0.21 | 0.47 | -0.90 | -0.77 |
| Height | 0.90 | -0.97 | 0.98 | -0.97 |
| Weight | 0.99 | -0.98 | 0.94 | -1.00 |

These results were obtained using 200 individuals per trial, and running 1000 generations per trial. We can see that even with this high number of generations, we don't see too much stability or convergence over the trials: the numbers in each trial vary widely. This suggests that evolutionary programming is particularly vulnerable to the randomness in generating the initial population and the randomness in the genetic crossover in each generation. It also brings in to question whether a method which produces such widely varying answers to the "best" solution when asked the same question multiple times is really producing *any* solutions that are the "best."

However, these differences are somewhat explainable. We first notice that the first and third trial are relatively similar: differences are usually in magnitude, and never in direction (ie the signs of the coefficients for variables are usually the same in the two trials). We could say something similar for the second and fourth trial, those these differences are usually larger in magnitude.

We can also note that the first/third trial seem to be *inverses* of results from the second/fourth trial; where a coefficient is positive in the first and third, the coefficients tend to be negative and of similar magnitude in the second/fourth trial. This is understandable when we consider that the variance of a data set is identical to the variance of that data set if we changed the sign of every value. Since the evolutionary algorithm operates on variances, we see that some indeterminacy in the algorithm's convergence may not be due to poor performance of the algorithm itself, but due to properties of the variance. However, this does not explain away all of the differences in each trial: the trials are not identical, merely with flipped sign, which implies that there is still significant randomness in each trial not attributable to the properties of variance.

Now we will discuss the significance of our findings— the motivation for doing component analysis on this data set in the first place. We find variables that are all either very high or very low in the same principal component, as this suggests that all of these variables are caused by another underlying factor. We examine the variables in the first trial with large coefficients, to discover a factor that describes someone who: does not cry (-.99), likes war movies (.96), likes action movies (.87), history (.87), physics (.96), PCs (.90), likes active sports (.96) and adrenaline sports (.93), dislikes psychology (-.96), dislikes reading (-.97), dislikes celebrities (-.94), dislikes shopping (-.88), has fewer mood swings (-.83), has fewer friends (-.97), eats to survive (1), shows little compassion towards animals (-.95), likes socializing (.97), “can fall for someone very quickly and then completely lose interest” (.91), makes quick or impulsive decisions (-.98), cannot “tell if people listen to me or not when I talk to them” (-.94). We appear to have discovered gender. This assumption is verified once we consider that height and weight are also correlated (.90 and .99 respectively), which of course vary with gender as well.

It is interesting that gender was selected for as producing the most variance among people. Although, perhaps gender doesn’t separate the human species as much as these results might lead us to believe. Again, this might be an artifact from using variance as the fitness function in principal component analysis. An evenly distributed data set will have a lesser variance than a data set with the same mean, but bimodally distributed. This applies to gender, in that responses are expected to be bimodal, rather than uniformly distributed, and thus will be selected as more fit. This leads to gender being selected over other factors that produce more variance between people, but not among the entire population. These factors perhaps might have been more interesting— perhaps further investigation is required in future projects.

Now we turn our attention to what originally motivated me to carry out this project— correlations with neuroticism as measured by the question "My moods change quickly." We examine the principal component that had the highest value for this trait— the fourth trial, with a value of .97. In this principal component, we notice people who: like folk music (.87), musical theatre (.95), dislike metal (-.97), like rock (.95), dislike war movies (-.96) and like romance movies (.93), enjoy dancing (.98), the outdoors/the countryside (.78), writing (.66), gardening (.77),

shopping (1), are generous (.94), lonely (.9), have few friends (-.92), and look after their appearance (.99). Some of these are indeed associated with the female gender (in this component, height and weight are negatively correlated with -.97, and -1). But this component seems also to have identified a certain type of neurotic feminine characteristic— some traits are not as obviously stereotypically feminine. Some traits are more interesting, such as enjoying the countryside, enjoying writing, being generous, lonely, and aware of their appearance. Perhaps we have identified the “Emily Dickinson and Emily Bronte” archetype— the neurotic, writerly type who enjoys the solitude of the countryside and transmutes their neuroticism and loneliness into art.

Analysis

We compare these results to the results of principal component analysis carried out analytically, using an R package:

| | RC1 | RC2 | RC3 | RC4 | RC5 |
|--------------------------|-------|-------|-------|-------|-------|
| Music | 0.17 | 0.24 | 0.01 | 0.07 | -0.1 |
| Slow.songs.or.fast.songs | -0.05 | -0.12 | 0.11 | 0.19 | -0.08 |
| Dance | 0.41 | -0.11 | 0.18 | 0.22 | 0.17 |
| Folk | -0.09 | 0.36 | -0.1 | 0.04 | 0.3 |
| Country | -0.15 | 0.24 | 0.12 | -0.02 | 0.26 |
| Classical.music | -0.16 | 0.59 | 0.06 | -0.16 | 0.24 |
| Musical | 0.22 | 0.39 | -0.1 | 0 | 0.25 |
| Pop | 0.49 | -0.13 | 0.07 | 0.06 | 0.14 |
| Rock | -0.21 | 0.45 | 0.05 | -0.13 | -0.17 |
| Metal.or.Hardrock | -0.45 | 0.3 | 0.12 | -0.11 | -0.14 |
| Punk | -0.23 | 0.32 | 0.13 | -0.1 | -0.15 |
| Hiphop..Rap | 0.34 | -0.17 | 0.32 | 0.25 | -0.02 |
| Reggae..Ska | 0.01 | 0.25 | 0.15 | 0.16 | -0.02 |
| Swing..Jazz | -0.02 | 0.59 | 0.06 | 0.01 | 0.11 |
| Rock.n.roll | -0.09 | 0.55 | 0.07 | 0 | -0.05 |
| Alternative | -0.2 | 0.57 | -0.01 | -0.07 | -0.16 |

| | | | | | |
|---------------------|-------|-------|-------|-------|-------|
| Latino | 0.3 | 0.22 | -0.14 | 0.18 | 0.27 |
| Techno..Trance | 0.15 | -0.07 | 0.31 | 0.09 | 0.03 |
| Opera | -0.12 | 0.5 | 0.05 | -0.07 | 0.3 |
| Movies | 0.19 | 0.18 | 0.18 | 0.02 | -0.1 |
| Horror | 0 | -0.03 | 0.32 | -0.02 | -0.18 |
| Thriller | -0.04 | -0.03 | 0.4 | -0.11 | -0.1 |
| Comedy | 0.28 | -0.08 | 0.16 | 0.12 | 0.07 |
| Romantic | 0.54 | 0.11 | -0.22 | 0.01 | 0.18 |
| Sci.fi | -0.26 | 0.09 | 0.41 | -0.07 | 0.01 |
| War | -0.22 | 0.11 | 0.43 | 0.03 | -0.07 |
| Fantasy.Fairy.tales | 0.3 | 0.29 | -0.21 | 0.03 | 0.14 |
| Animated | 0.19 | 0.29 | -0.11 | 0 | 0.06 |
| Documentary | -0.23 | 0.28 | 0.22 | -0.02 | 0.3 |
| Western | -0.3 | 0.13 | 0.38 | 0.06 | 0.08 |
| Action | -0.12 | -0.1 | 0.53 | 0.07 | 0 |
| History | -0.15 | 0.42 | 0.21 | 0.02 | 0.09 |
| Psychology | 0.07 | 0.42 | -0.02 | 0.01 | 0.09 |
| Politics | -0.07 | 0.22 | 0.38 | 0.12 | 0.08 |
| Mathematics | -0.26 | -0.02 | 0.29 | 0.03 | 0.36 |
| Physics | -0.34 | 0.06 | 0.36 | 0.04 | 0.37 |
| Internet | 0.09 | -0.11 | 0.39 | -0.06 | -0.01 |
| PC | -0.24 | -0.09 | 0.57 | -0.02 | 0.11 |
| Economy.Management | 0.17 | -0.05 | 0.29 | 0.15 | 0.12 |
| Biology | 0.08 | 0.23 | -0.08 | 0 | 0.37 |
| Chemistry | -0.05 | 0.1 | -0.05 | 0.01 | 0.37 |
| Reading | 0.06 | 0.56 | -0.27 | -0.12 | 0.07 |
| Geography | -0.06 | 0.24 | 0.21 | 0.12 | 0.13 |
| Foreign.languages | 0.19 | 0.43 | 0 | 0.05 | 0.05 |
| Medicine | 0.05 | 0.24 | -0.05 | 0.03 | 0.34 |

| | | | | | |
|-------------------------|-------|-------|-------|-------|-------|
| Law | 0.15 | 0.18 | 0.28 | 0.13 | 0.08 |
| Cars | -0.03 | -0.17 | 0.58 | 0.21 | 0.04 |
| Art.exhibitions | 0.11 | 0.63 | -0.11 | -0.01 | 0.11 |
| Religion | -0.07 | 0.32 | -0.09 | 0.03 | 0.3 |
| Countryside..outdoors | 0.02 | 0.27 | -0.11 | 0.08 | 0.32 |
| Dancing | 0.4 | 0.27 | -0.15 | 0.24 | 0.17 |
| Musical.instruments | -0.04 | 0.46 | -0.03 | 0.01 | 0.13 |
| Writing | 0.02 | 0.44 | -0.04 | -0.14 | 0.05 |
| Passive.sport | 0.06 | -0.02 | 0.24 | 0.18 | -0.03 |
| Active.sport | 0.05 | 0.07 | 0.28 | 0.39 | 0.05 |
| Gardening | 0.1 | 0.11 | -0.15 | 0.04 | 0.32 |
| Celebrities | 0.59 | -0.09 | 0.09 | -0.07 | -0.02 |
| Shopping | 0.73 | 0.01 | -0.07 | 0.08 | -0.02 |
| Science.and.technology | -0.22 | 0.17 | 0.55 | 0.03 | 0.24 |
| Theatre | 0.18 | 0.59 | -0.2 | 0.01 | 0.13 |
| Fun.with.friends | 0.26 | 0.2 | 0.09 | 0.37 | -0.16 |
| Adrenaline.sports | 0.03 | 0.13 | 0.38 | 0.4 | -0.09 |
| Pets | 0.24 | 0.07 | -0.03 | 0.01 | 0.03 |
| Flying | 0.18 | -0.05 | -0.15 | -0.21 | 0.15 |
| Storm | 0.39 | 0.04 | -0.24 | -0.12 | 0.05 |
| Darkness | 0.41 | 0.13 | -0.22 | -0.19 | -0.09 |
| Heights | 0.15 | -0.04 | -0.07 | -0.22 | -0.01 |
| Spiders | 0.43 | 0.01 | -0.1 | -0.2 | -0.05 |
| Snakes | 0.43 | -0.08 | -0.1 | -0.14 | 0.04 |
| Rats | 0.48 | -0.07 | -0.03 | -0.16 | 0 |
| Ageing | 0.34 | 0.06 | 0.04 | -0.2 | -0.1 |
| Dangerous.dogs | 0.36 | -0.03 | -0.15 | -0.27 | 0.08 |
| Fear.of.public.speaking | 0.07 | -0.08 | -0.08 | -0.48 | 0.03 |
| Healthy.eating | 0.19 | 0.1 | 0.06 | 0.19 | 0.31 |

| | | | | | |
|-----------------------|-------|-------|-------|-------|-------|
| Daily.events | 0.12 | 0.18 | 0.33 | 0.05 | 0.1 |
| Prioritising.workload | 0.1 | -0.03 | -0.01 | 0.06 | 0.51 |
| Writing.notes | 0.25 | 0.17 | -0.11 | -0.05 | 0.41 |
| Workaholism | 0 | 0.24 | 0.05 | -0.02 | 0.46 |
| Thinking.ahead | -0.02 | 0.02 | 0.18 | -0.19 | 0.43 |
| Final.judgement | 0.15 | 0.02 | -0.04 | -0.03 | 0.28 |
| Reliability | 0.11 | 0.01 | 0.09 | 0.07 | 0.44 |
| Keeping.promises | 0.04 | 0.06 | 0.1 | 0.03 | 0.36 |
| Loss.of.interest | 0.14 | 0.08 | 0.11 | -0.08 | -0.15 |
| Friends.versus.money | 0.04 | 0.19 | -0.18 | 0.14 | 0.16 |
| Funniness | -0.03 | 0.11 | 0.32 | 0 | -0.11 |
| Fake | -0.03 | 0.07 | 0.15 | -0.36 | -0.14 |
| Criminal.damage | 0.05 | -0.04 | 0.32 | -0.16 | -0.25 |
| Decision.making | 0.1 | 0.01 | -0.08 | -0.32 | 0.3 |
| Elections | -0.05 | 0.19 | 0.03 | 0.1 | 0.17 |
| Self.criticism | 0.02 | 0.24 | 0.04 | -0.36 | 0.13 |
| Judgment.calls | 0.14 | 0.21 | 0.1 | 0.09 | -0.02 |
| Hypochondria | 0.16 | 0.1 | 0.04 | -0.32 | -0.06 |
| Empathy | 0.19 | 0.2 | -0.17 | 0.03 | 0.13 |
| Eating.to.survive | -0.02 | -0.06 | 0.12 | -0.19 | 0.06 |
| Giving | 0.34 | 0.08 | -0.03 | 0.14 | 0.22 |
| Compassion.to.animals | 0.25 | 0.13 | -0.1 | -0.09 | 0.03 |
| Borrowed.stuff | 0.13 | 0.01 | 0.06 | -0.08 | 0.34 |
| Loneliness | 0.05 | 0.13 | 0.03 | -0.55 | -0.02 |
| Cheating.in.school | 0.15 | -0.01 | 0.13 | 0.15 | -0.46 |
| Health | 0.42 | -0.01 | 0.11 | -0.2 | 0.19 |
| Changing.the.past | 0.15 | 0.06 | 0.15 | -0.4 | -0.16 |
| God | 0.17 | 0.08 | -0.19 | 0.05 | 0.26 |
| Dreams | -0.03 | -0.07 | 0.01 | 0.3 | 0.07 |

| | | | | | |
|--------------------------------|-------|-------|-------|-------|-------|
| Charity | 0.1 | 0.32 | 0.1 | 0.03 | 0.24 |
| Number.of.friends | 0.22 | 0.15 | 0.07 | 0.55 | -0.11 |
| Waiting | -0.2 | 0.06 | -0.05 | 0.21 | 0.04 |
| New.environment | -0.05 | 0.11 | 0.07 | 0.47 | -0.13 |
| Mood.swings | 0.18 | 0.1 | -0.11 | -0.41 | -0.06 |
| Appearance.and.gestures | 0.45 | 0.07 | 0.08 | 0.05 | 0 |
| Socializing | 0.18 | 0.18 | 0.1 | 0.49 | -0.1 |
| Achievements | 0.15 | 0.01 | 0.17 | 0.05 | -0.1 |
| Responding.to.a.serious.letter | -0.05 | 0.04 | -0.02 | -0.19 | -0.01 |
| Children | 0.25 | 0.09 | -0.17 | 0.17 | 0.23 |
| Assertiveness | 0.04 | 0.09 | 0.19 | 0.25 | -0.05 |
| Getting.angry | 0.27 | -0.03 | 0.04 | -0.29 | -0.15 |
| Knowing.the.right.people | 0.3 | 0.09 | 0.35 | 0.17 | -0.04 |
| Public.speaking | 0.16 | -0.22 | -0.11 | -0.43 | 0.11 |
| Unpopularity | 0.03 | 0.12 | -0.04 | -0.2 | 0.12 |
| Life.struggles | 0.41 | 0.14 | -0.45 | -0.22 | 0.1 |
| Happiness.in.life | 0.04 | -0.07 | 0.02 | 0.59 | 0.19 |
| Energy.levels | 0.15 | 0.02 | 0.06 | 0.67 | 0.04 |
| Small...big.dogs | -0.24 | 0.07 | 0.2 | 0.15 | -0.28 |
| Personality | 0.11 | -0.13 | 0.1 | 0.48 | 0.07 |
| Finding.lost.valuables | 0.05 | 0.16 | -0.14 | 0 | 0.27 |
| Getting.up | 0.12 | 0.11 | 0 | -0.14 | -0.37 |
| Interests.or.hobbies | 0.03 | 0.26 | 0.15 | 0.48 | 0.06 |
| Parents..advice | 0.2 | 0.06 | -0.03 | -0.04 | 0.33 |
| Questionnaires.or.polls | 0.07 | 0.1 | -0.14 | -0.1 | 0.24 |
| Finances | -0.06 | -0.08 | -0.09 | -0.09 | 0.48 |
| Shopping.centres | 0.6 | -0.04 | 0.07 | 0.09 | -0.02 |
| Branded.clothing | 0.27 | -0.09 | 0.39 | 0.16 | -0.13 |
| Entertainment.spending | 0.14 | 0.12 | 0.3 | 0.26 | -0.37 |

| | | | | | |
|----------------------------|-------|-------|------|------|-------|
| Spending.on.looks | 0.6 | 0.01 | 0.2 | 0.16 | -0.16 |
| Spending.on.gadgets | 0.09 | -0.02 | 0.54 | 0.09 | -0.13 |
| Spending.on.healthy.eating | 0.14 | 0.15 | 0.16 | 0.17 | 0.13 |
| Age | -0.08 | 0 | 0.12 | 0.02 | 0.27 |
| Height | -0.38 | -0.17 | 0.52 | 0.05 | -0.06 |
| Weight | -0.36 | -0.12 | 0.51 | 0.03 | -0.02 |

First we notice that gender appears to be less heavily selected for; height and weight have much less extreme values. We can summarize the first principal component as interest in shopping/celebrities/spending on looks. In the evolutionary algorithm, these traits were indeed included, but they were combined with a component that combined all stereotypically gendered traits. The second principal component seems to measure an appreciation of art, with the highest values being enjoying classical music, jazz, opera, history, psychology, reading, art exhibitions, musical instruments, writing, and theatre. This is a trait that our evolutionary algorithm failed to pick up on. Overall, we notice that values in the analytic solution are less extreme, with very few values being near 1 or -1, suggesting that our evolutionary algorithm has a tendency to gravitate towards extreme values, and is a “greedier” algorithm than the deterministic one.

Conclusion

We analyzed the effectiveness of an evolutionary algorithm in exploring a data set using principal component analysis. We concluded that evolutionary algorithms are particularly prone to initial randomness in generating the population, and also prone to gravitating towards extreme, greedy values. Future possibilities for research include running with more computational power (more generations, larger samples, as my computer was being taxed with these trials), controlling for gender by optimizing the variance calculations for bimodality, and carrying out multiple principal component analysis instead of just a single principal component, which would make each component orthogonal.

Works Cited

Koza, John. "Welcome to." *Genetic-programming.org-Home-Page*. N.p., n.d. Web. 03 May 2017.

LSYPE : Longitudinal Study of Young People in England. N.p., n.d. Web. 03 May 2017.

Poli, Riccardo, W. B. Langdon, Nicholas F. McPhee, and John R. Koza. *A Field Guide to Genetic Programming*. S.l.: Lulu, 2008. Web.

"Q1.2: What's Evolutionary Programming (EP)?" *FAQ: Comp.ai.genetic Part 2/6 (A Guide to Frequently Asked Questions) - Q1.2: What's Evolutionary Programming (EP)?* N.p., n.d. Web. 03 May 2017.

When Are Evolutionary Algorithms Useful? N.p., n.d. Web. 03 May 2017.

Programs

PCA.java

```
import java.util.ArrayList;
import java.util.List;
import org.uncommons.maths.random.MersenneTwisterRNG;
import org.uncommons.maths.random.Probability;
import org.uncommons.watchmaker.framework.EvolutionaryOperator;
import org.uncommons.watchmaker.framework.GenerationalEvolutionEngine;
import org.uncommons.watchmaker.framework.operators.DoubleArrayCrossover;
import org.uncommons.watchmaker.framework.operators.EvolutionPipeline;
import org.uncommons.watchmaker.framework.selection.RouletteWheelSelection;
import org.uncommons.watchmaker.framework.termination.GenerationCount;
import java.io.FileNotFoundException;

//main class to run program
public class PCA
{

    //prints an array, as a way to read output of selection
    public static void printArr(double[] list) {
        String result= "";
        for(double d : list) {
            result = result+" "+Double.toString(d);
        }
        System.out.println(result);
    }

    //main method, runs four trials and prints them
    public static void main(String[] args) throws FileNotFoundException
    {
        printArr(evolveVect());
        printArr(evolveVect());
        printArr(evolveVect());
        printArr(evolveVect());
    }

    //evolves a single principal component vector
    public static double[] evolveVect() throws FileNotFoundException
    {
        List<EvolutionaryOperator<double[]>> operators = new
ArrayList<EvolutionaryOperator<double[]>>(2);

        operators.add(new DoubleArrayCrossover(3, new Probability(0.7d)));

        EvolutionaryOperator<double[]> pipeline = new
EvolutionPipeline<double[]>(operators);

        GenerationalEvolutionEngine<double[]> engine = new
GenerationalEvolutionEngine<double[]>(new VectorFactory(),
pipeline,
```

```

new VectorEvaluator(),

new RouletteWheelSelection(),

new MersenneTwisterRNG());
    engine.setSingleThreaded(true);
    return engine.evolve(200, // 200 individuals in each generation.
        5, // use elitism.
        new GenerationCount(1000));
}
}

```

VectorFactory.java

```

import java.util.Random;

import org.uncommons.watchmaker.framework.factories.AbstractCandidateFactory;

public class VectorFactory extends AbstractCandidateFactory<double[]>
{
    public double[] generateRandomCandidate(Random r)
    {
        double[] vector = new double[139];
        Random randgen = new Random();
        for (int i = 0; i < 138; i++)
        {
            vector[i] = (randgen.nextDouble()*2)-1;
        }
        return vector;
    }
}

```

VectorEvaluator.java

```

import java.util.List;
import org.uncommons.watchmaker.framework.*;
import java.io.File;
import java.io.FileNotFoundException;
import java.util.*;

public class VectorEvaluator implements FitnessEvaluator<double[]>
{
    ArrayList<double[]> obsList = new ArrayList<double[]>();

    //returns the mean of a list
    public double getMean(ArrayList<Double> list) {
        double sum = 0.0;
        for(double a : list)
            sum += a;
        return sum/list.size();
    }
}

```

```

}

//calculates the variances of a list
public double calcVariance(ArrayList<Double> list) {
    double mean = getMean(list);
    double temp = 0;
    for(double a :list)
        temp += (a-mean)*(a-mean);
    return temp/list.size();
}

//the fitness function for a principal component
public VectorEvaluator() throws FileNotFoundException{
    Scanner scanner = new Scanner(new File("numerdata.csv"));
    scanner.useDelimiter("\n");
    scanner.nextLine();

    while(scanner.hasNextLine()){
        String[] curLine = scanner.nextLine().split(",");
        double[] cur = new double[139];
        for(int i =0; i<curLine.length; i++){
            cur[i] = Double.parseDouble(curLine[i]);
        }
        obsList.add(cur);
    }
    scanner.close();
}

//calculates the dot product of two vectors
public double dotProduct(double[] vect1, double[] vect2) {
    double result = 0;
    for (int i =0; i<vect1.length; i++) {
        result = result + vect1[i]*vect2[i];
    }
    return result;
}

//the fitness function
public double getFitness(double[] candidate,
                        List<? extends double[]> population)
{
    ArrayList<Double> dotProducts = new ArrayList<Double>();
    for(double[] o : obsList){
        dotProducts.add(dotProduct(o,candidate));
    }

    return calcVariance(dotProducts);
}

/**
 * Always returns true. A higher score indicates a fitter individual.
 * @return True.

```

```

    */
    public boolean isNatural()
    {
        return true;
    }
}

```

Questionnaire

MUSIC PREFERENCES

- 1 I enjoy listening to music.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 2 I prefer.: Slow paced music 1-2-3-4-5 Fast paced music (integer)
- 3 Dance, Disco, Funk: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
- 4 Folk music: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
- 5 Country: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
- 6 Classical: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
- 7 Musicals: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
- 8 Pop: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
- 9 Rock: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
- 10 Metal, Hard rock: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
- 11 Punk: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
- 12 Hip hop, Rap: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
- 13 Reggae, Ska: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
- 14 Swing, Jazz: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
- 15 Rock n Roll: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
- 16 Alternative music: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
- 17 Latin: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
- 18 Techno, Trance: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
- 19 Opera: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)

MOVIE PREFERENCES

- 1 I really enjoy watching movies.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 2 Horror movies: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
- 3 Thriller movies: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
- 4 Comedies: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
- 5 Romantic movies: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
- 6 Sci-fi movies: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
- 7 War movies: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
- 8 Tales: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
- 9 Cartoons: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
- 10 Documentaries: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
- 11 Western movies: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
- 12 Action movies: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)

HOBBIES & INTERESTS

- 1 History: Not interested 1-2-3-4-5 Very interested (integer)
- 2 Psychology: Not interested 1-2-3-4-5 Very interested (integer)
- 3 Politics: Not interested 1-2-3-4-5 Very interested (integer)
- 4 Mathematics: Not interested 1-2-3-4-5 Very interested (integer)
- 5 Physics: Not interested 1-2-3-4-5 Very interested (integer)

- 6 Internet: Not interested 1-2-3-4-5 Very interested (integer)
- 7 PC Software, Hardware: Not interested 1-2-3-4-5 Very interested (integer)
- 8 Economy, Management: Not interested 1-2-3-4-5 Very interested (integer)
- 9 Biology: Not interested 1-2-3-4-5 Very interested (integer)
- 10 Chemistry: Not interested 1-2-3-4-5 Very interested (integer)
- 11 Poetry reading: Not interested 1-2-3-4-5 Very interested (integer)
- 12 Geography: Not interested 1-2-3-4-5 Very interested (integer)
- 13 Foreign languages: Not interested 1-2-3-4-5 Very interested (integer)
- 14 Medicine: Not interested 1-2-3-4-5 Very interested (integer)
- 15 Law: Not interested 1-2-3-4-5 Very interested (integer)
- 16 Cars: Not interested 1-2-3-4-5 Very interested (integer)
- 17 Art: Not interested 1-2-3-4-5 Very interested (integer)
- 18 Religion: Not interested 1-2-3-4-5 Very interested (integer)
- 19 Outdoor activities: Not interested 1-2-3-4-5 Very interested (integer)
- 20 Dancing: Not interested 1-2-3-4-5 Very interested (integer)
- 21 Playing musical instruments: Not interested 1-2-3-4-5 Very interested (integer)
- 22 Poetry writing: Not interested 1-2-3-4-5 Very interested (integer)
- 23 Sport and leisure activities: Not interested 1-2-3-4-5 Very interested (integer)
- 24 Sport at competitive level: Not interested 1-2-3-4-5 Very interested (integer)
- 25 Gardening: Not interested 1-2-3-4-5 Very interested (integer)
- 26 Celebrity lifestyle: Not interested 1-2-3-4-5 Very interested (integer)
- 27 Shopping: Not interested 1-2-3-4-5 Very interested (integer)
- 28 Science and technology: Not interested 1-2-3-4-5 Very interested (integer)
- 29 Theatre: Not interested 1-2-3-4-5 Very interested (integer)
- 30 Socializing: Not interested 1-2-3-4-5 Very interested (integer)
- 31 Adrenaline sports: Not interested 1-2-3-4-5 Very interested (integer)
- 32 Pets: Not interested 1-2-3-4-5 Very interested (integer)

PHOBIAS

- 1 Flying: Not afraid at all 1-2-3-4-5 Very afraid of (integer)
- 2 Thunder, lightning: Not afraid at all 1-2-3-4-5 Very afraid of (integer)
- 3 Darkness: Not afraid at all 1-2-3-4-5 Very afraid of (integer)
- 4 Heights: Not afraid at all 1-2-3-4-5 Very afraid of (integer)
- 5 Spiders: Not afraid at all 1-2-3-4-5 Very afraid of (integer)
- 6 Snakes: Not afraid at all 1-2-3-4-5 Very afraid of (integer)
- 7 Rats, mice: Not afraid at all 1-2-3-4-5 Very afraid of (integer)
- 8 Ageing: Not afraid at all 1-2-3-4-5 Very afraid of (integer)
- 9 Dangerous dogs: Not afraid at all 1-2-3-4-5 Very afraid of (integer)
- 10 Public speaking: Not afraid at all 1-2-3-4-5 Very afraid of (integer)

HEALTH HABITS

- 1 Smoking habits: Never smoked - Tried smoking - Former smoker - Current smoker (categorical)
- 2 Drinking: Never - Social drinker - Drink a lot (categorical)
- 3 I live a very healthy lifestyle.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)

PERSONALITY TRAITS, VIEWS ON LIFE & OPINIONS

- 1 I take notice of what goes on around me.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)

- 2 I try to do tasks as soon as possible and not leave them until last minute.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 3 I always make a list so I don't forget anything.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 4 I often study or work even in my spare time.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 5 I look at things from all different angles before I go ahead.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 6 I believe that bad people will suffer one day and good people will be rewarded.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 7 I am reliable at work and always complete all tasks given to me.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 8 I always keep my promises.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 9 I can fall for someone very quickly and then completely lose interest.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 10 I would rather have lots of friends than lots of money.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 11 I always try to be the funniest one.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 12 I can be two faced sometimes.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 13 I damaged things in the past when angry.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 14 I take my time to make decisions.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 15 I always try to vote in elections.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 16 I often think about and regret the decisions I make.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 17 I can tell if people listen to me or not when I talk to them.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 18 I am a hypochondriac.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 19 I am emphatic person.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 20 I eat because I have to. I don't enjoy food and eat as fast as I can.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 21 I try to give as much as I can to other people at Christmas.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 22 I don't like seeing animals suffering.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 23 I look after things I have borrowed from others.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 24 I feel lonely in life.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 25 I used to cheat at school.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 26 I worry about my health.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 27 I wish I could change the past because of the things I have done.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 28 I believe in God.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 29 I always have good dreams.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 30 I always give to charity.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 31 I have lots of friends.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)

- 32 Timekeeping.: I am often early. - I am always on time. - I am often running late.
(categorical)
- 33 Do you lie to others?: Never. - Only to avoid hurting someone. - Sometimes. - Everytime
it suits me. (categorical)
- 34 I am very patient.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 35 I can quickly adapt to a new environment.: Strongly disagree 1-2-3-4-5 Strongly agree
(integer)
- 36 My moods change quickly.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 37 I am well mannered and I look after my appearance.: Strongly disagree 1-2-3-4-5
Strongly agree (integer)
- 38 I enjoy meeting new people.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 39 I always let other people know about my achievements.: Strongly disagree 1-2-3-4-5
Strongly agree (integer)
- 40 I think carefully before answering any important letters.: Strongly disagree 1-2-3-4-5
Strongly agree (integer)
- 41 I enjoy childrens' company.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 42 I am not afraid to give my opinion if I feel strongly about something.: Strongly disagree
1-2-3-4-5 Strongly agree (integer)
- 43 I can get angry very easily.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 44 I always make sure I connect with the right people.: Strongly disagree 1-2-3-4-5 Strongly
agree (integer)
- 45 I have to be well prepared before public speaking.: Strongly disagree 1-2-3-4-5 Strongly
agree (integer)
- 46 I will find a fault in myself if people don't like me.: Strongly disagree 1-2-3-4-5 Strongly
agree (integer)
- 47 I cry when I feel down or things don't go the right way.: Strongly disagree 1-2-3-4-5
Strongly agree (integer)
- 48 I am 100% happy with my life.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 49 I am always full of life and energy.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 50 I prefer big dangerous dogs to smaller, calmer dogs.: Strongly disagree 1-2-3-4-5 Strongly
agree (integer)
- 51 I believe all my personality traits are positive.: Strongly disagree 1-2-3-4-5 Strongly agree
(integer)
- 52 If I find something the doesn't belong to me I will hand it in.: Strongly disagree 1-2-3-4-5
Strongly agree (integer)
- 53 I find it very difficult to get up in the morning.: Strongly disagree 1-2-3-4-5 Strongly agree
(integer)
- 54 I have many different hobbies and interests.: Strongly disagree 1-2-3-4-5 Strongly agree
(integer)
- 55 I always listen to my parents' advice.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 56 I enjoy taking part in surveys.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 57 How much time do you spend online?: No time at all - Less than an hour a day - Few
hours a day - Most of the day (categorical)

SPENDING HABITS

- 1 I save all the money I can.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)

- 2 I enjoy going to large shopping centres.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 3 I prefer branded clothing to non branded.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 4 I spend a lot of money on partying and socializing.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 5 I spend a lot of money on my appearance.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 6 I spend a lot of money on gadgets.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
- 7 I will happily pay more money for good, quality or healthy food.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)

DEMOGRAPHICS

- 1 Age: (integer)
- 2 Height: (integer)
- 3 Weight: (integer)
- 4 How many siblings do you have?: (integer)

Key

"original", "short"
 "I enjoy listening to music.", "Music"
 "I prefer.", "Slow songs or fast songs"
 "Dance, Disco, Funk", "Dance"
 "Folk music", "Folk"
 "Country", "Country"
 "Classical", "Classical music"
 "Musicals", "Musical"
 "Pop", "Pop"
 "Rock", "Rock"
 "Metal, Hard rock", "Metal or Hardrock"
 "Punk", "Punk"
 "Hip hop, Rap", "Hiphop, Rap"
 "Reggae, Ska", "Reggae, Ska"
 "Swing, Jazz", "Swing, Jazz"
 "Rock n Roll", "Rock n roll"
 "Alternative music", "Alternative"
 "Latin", "Latino"
 "Techno, Trance", "Techno, Trance"
 "Opera", "Opera"
 "I really enjoy watching movies.", "Movies"
 "Horror movies", "Horror"
 "Thriller movies", "Thriller"
 "Comedies", "Comedy"
 "Romantic movies", "Romantic"
 "Sci-fi movies", "Sci-fi"
 "War movies", "War"

"Tales", "Fantasy/Fairy tales"
 "Cartoons", "Animated"
 "Documentaries", "Documentary"
 "Western movies", "Western"
 "Action movies", "Action"
 "History", "History"
 "Psychology", "Psychology"
 "Politics", "Politics"
 "Mathematics", "Mathematics"
 "Physics", "Physics"
 "Internet", "Internet"
 "PC Software, Hardware", "PC"
 "Economy, Management", "Economy Management"
 "Biology", "Biology"
 "Chemistry", "Chemistry"
 "Poetry reading", "Reading"
 "Geography", "Geography"
 "Foreign languages", "Foreign languages"
 "Medicine", "Medicine"
 "Law", "Law"
 "Cars", "Cars"
 "Art", "Art exhibitions"
 "Religion", "Religion"
 "Outdoor activities", "Countryside, outdoors"
 "Dancing", "Dancing"
 "Playing musical instruments", "Musical instruments"
 "Poetry writing", "Writing"
 "Sport and leisure activities", "Passive sport"
 "Sport at competitive level", "Active sport"
 "Gardening", "Gardening"
 "Celebrity lifestyle", "Celebrities"
 "Shopping", "Shopping"
 "Science and technology", "Science and technology"
 "Theatre", "Theatre"
 "Socializing", "Fun with friends"
 "Adrenaline sports", "Adrenaline sports"
 "Pets", "Pets"
 "Flying", "Flying"
 "Thunder, lightning", "Storm"
 "Darkness", "Darkness"
 "Heights", "Heights"
 "Spiders", "Spiders"
 "Snakes", "Snakes"
 "Rats, mice", "Rats"
 "Ageing", "Ageing"
 "Dangerous dogs", "Dangerous dogs"

"Public speaking", "Fear of public speaking"
 "Smoking habits", "Smoking"
 "Drinking", "Alcohol"
 "I live a very healthy lifestyle.", "Healthy eating"
 "I take notice of what goes on around me.", "Daily events"
 "I try to do tasks as soon as possible and not leave them until last minute.", "Prioritising workload"
 "I always make a list so I don't forget anything.", "Writing notes"
 "I often study or work even in my spare time.", "Workaholism"
 "I look at things from all different angles before I go ahead.", "Thinking ahead"
 "I believe that bad people will suffer one day and good people will be rewarded.", "Final judgement"
 "I am reliable at work and always complete all tasks given to me.", "Reliability"
 "I always keep my promises.", "Keeping promises"
 "I can fall for someone very quickly and then completely lose interest.", "Loss of interest"
 "I would rather have lots of friends than lots of money.", "Friends versus money"
 "I always try to be the funniest one.", "Funniness"
 "I can be two faced sometimes.", "Fake"
 "I damaged things in the past when angry.", "Criminal damage"
 "I take my time to make decisions.", "Decision making"
 "I always try to vote in elections.", "Elections"
 "I often think about and regret the decisions I make.", "Self-criticism"
 "I can tell if people listen to me or not when I talk to them.", "Judgment calls"
 "I am a hypochondriac.", "Hypochondria"
 "I am empathetic person.", "Empathy"
 "I eat because I have to. I don't enjoy food and eat as fast as I can.", "Eating to survive"
 "I try to give as much as I can to other people at Christmas.", "Giving"
 "I don't like seeing animals suffering.", "Compassion to animals"
 "I look after things I have borrowed from others.", "Borrowed stuff"
 "I feel lonely in life.", "Loneliness"
 "I used to cheat at school.", "Cheating in school"
 "I worry about my health.", "Health"
 "I wish I could change the past because of the things I have done.", "Changing the past"
 "I believe in God.", "God"
 "I always have good dreams.", "Dreams"
 "I always give to charity.", "Charity"
 "I have lots of friends.", "Number of friends"
 "Timekeeping.", "Punctuality"
 "Do you lie to others?", "Lying"
 "I am very patient.", "Waiting"
 "I can quickly adapt to a new environment.", "New environment"
 "My moods change quickly.", "Mood swings"
 "I am well mannered and I look after my appearance.", "Appearance and gestures"
 "I enjoy meeting new people.", "Socializing"
 "I always let other people know about my achievements.", "Achievements"
 "I think carefully before answering any important letters.", "Responding to a serious letter"
 "I enjoy childrens' company.", "Children"

"I am not afraid to give my opinion if I feel strongly about something.", "Assertiveness"
 "I can get angry very easily.", "Getting angry"
 "I always make sure I connect with the right people.", "Knowing the right people"
 "I have to be well prepared before public speaking.", "Public speaking"
 "I will find a fault in myself if people don't like me.", "Unpopularity"
 "I cry when I feel down or things don't go the right way.", "Life struggles"
 "I am 100% happy with my life.", "Happiness in life"
 "I am always full of life and energy.", "Energy levels"
 "I prefer big dangerous dogs to smaller, calmer dogs.", "Small - big dogs"
 "I believe all my personality traits are positive.", "Personality"
 "If I find something the doesn't belong to me I will hand it in.", "Finding lost valuables"
 "I find it very difficult to get up in the morning.", "Getting up"
 "I have many different hobbies and interests.", "Interests or hobbies"
 "I always listen to my parents' advice.", "Parents' advice"
 "I enjoy taking part in surveys.", "Questionnaires or polls"
 "How much time do you spend online?", "Internet usage"
 "I save all the money I can.", "Finances"
 "I enjoy going to large shopping centres.", "Shopping centres"
 "I prefer branded clothing to non branded.", "Branded clothing"
 "I spend a lot of money on partying and socializing.", "Entertainment spending"
 "I spend a lot of money on my appearance.", "Spending on looks"
 "I spend a lot of money on gadgets.", "Spending on gadgets"
 "I will happily pay more money for good, quality or healthy food.", "Spending on healthy eating"
 "Age", "Age"
 "Height", "Height"
 "Weight", "Weight"
 "How many siblings do you have?", "Number of siblings"
 "Gender", "Gender"
 "I am", "Left - right handed"
 "Highest education achieved", "Education"
 "I am the only child", "Only child"
 "I spent most of my childhood in a", "Village - town"
 "I lived most of my childhood in a", "House - block of flats"