**https://www.kaggle.com/miroslavsabo/young-people-survey**

**Introduction**

In 2013, students of the Statistics class at [FSEV UK](https://fses.uniba.sk/en/) were asked to invite their friends to participate in this survey.

* The data file (responses.csv) consists of 1010 rows and 150 columns (139 integer and 11 categorical).
* For convenience, the original variable names were shortened in the data file. See the columns.csv file if you want to match the data with the original names.
* The data contain missing values.
* The survey was presented to participants in both electronic and written form.
* The original questionnaire was in Slovak language and was later translated into English.
* All participants were of Slovakian nationality, aged between 15-30.

The variables can be split into the following groups:

* **Music preferences** (19 items)
* **Movie preferences** (12 items)
* **Hobbies & interests** (32 items)
* **Phobias** (10 items)
* **Health habits** (3 items)
* **Personality traits, views on life, & opinions** (57 items)
* **Spending habits** (7 items)
* **Demographics** (10 items)

**Research questions**

Many different techniques can be used to answer many questions, e.g.

* **Clustering:** Given the music preferences, do people make up any clusters of similar behavior?
* **Hypothesis testing:** Do women fear certain phenomena significantly more than men? Do the left handed people have different interests than right handed?
* **Predictive modeling:** Can we predict spending habits of a person from his/her interests and movie or music preferences?
* **Dimension reduction:** Can we describe a large number of human interests by a smaller number of latent concepts?
* **Correlation analysis:** Are there any connections between music and movie preferences?
* **Visualization:** How to effectively visualize a lot of variables in order to gain some meaningful insights from the data?
* **(Multivariate) Outlier detection:** Small number of participants often cheats and randomly answers the questions. Can you identify them? Hint: [Local outlier factor](https://en.wikipedia.org/wiki/Local_outlier_factor) may help.
* **Missing values analysis:** Are there any patterns in missing responses? What is the optimal way of imputing the values in surveys?
* **Recommendations:** If some of user's interests are known, can we predict the other? Or, if we know what a person listen, can we predict which kind of movies he/she might like?

**Past research**

* (in slovak) Sleziak, P. - Sabo, M.: Gender differences in the prevalence of specific phobias. Forum Statisticum Slovacum. 2014, Vol. 10, No. 6. [Differences (gender + whether people lived in village/town) in the prevalence of phobias.]
* Sabo, Miroslav. Multivariate Statistical Methods with Applications. Diss. Slovak University of Technology in Bratislava, 2014. [Clustering of variables (music preferences, movie preferences, phobias) + Clustering of people w.r.t. their interests.]

**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 emphatetic 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 hapilly 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)
5. **Gender**: Female - Male (categorical)
6. **I am**: Left handed - Right handed (categorical)
7. **Highest education achieved**: Currently a Primary school pupil - Primary school - Secondary school - College/Bachelor degree (categorical)
8. **I am the only child**: No - Yes (categorical)
9. **I spent most of my childhood in a**: City - village (categorical)
10. **I lived most of my childhood in a**: house/bungalow - block of flats (categorical)