PCA: an example from Data Science 101
http://mypersonality.org/

Thanks to M. Kosinski we have data on 10000 individuals who took the *IPIP proxy for Costa and McCrae’s NEO-PI-R domains (Five Factor Model)*

Personality theory talks about 5 traits (OCEAN)

**Openness** to experience describes a dimension of personality that distinguishes imaginative, creative people from down-to-earth, conventional people.

**Conscientiousness** concerns the way in which we control, regulate, and direct our impulses.

**Extraversion** is marked by pronounced engagement with the external world, versus being comfortable with your own company.

**Agreeableness** reflects individual differences in concern with cooperation and social harmony.

**Neuroticism** refers to the tendency to experience negative emotions.
The questions

The data contains answers to the first 20 questions.

The questions presumably probe the different 5 dimensions as follows:

1. Openness
2. Agreeableness
3. Extraversion
4. Conscientiousness
5. Neuroticism

You can take the test.
A quick look at the demographics

- **Age**
  - Frequency distribution for ages 20 to 80 years

- **Age of males**
  - Frequency distribution for ages 20 to 80 years

- **Gender**
  - Frequency distribution for males and females

- **Age of females**
  - Frequency distribution for ages 30 to 70 years
Principal components of the Personality questions

Variances of the principal components

![Graph of Variances of the principal components](image)
Coefficients of questions and their association to OCEAN

PCA 1

PCA 2

PCA 3

PCA 4

PCA 5

PCA 6
Projection of subjects onto the first 6 PCA