Case study: personality assessment

Sanjoy Dasgupta

University of California, San Diego

Quantifying personality

What are the dimensions along which personalities differ?

- *Lexical hypothesis*: most important personality characteristics have become encoded in natural language.
- Allport and Odbert (1936): identified 4500 words describing personality traits.
- Group these words into (approximate) synonyms, by manual clustering. E.g. Norman (1967):

Spirit	Jolly, merry, witty, lively, peppy
Talkativeness	Talkative, articulate, verbose, gossipy
Sociability	Companionable, social, outgoing
Spontaneity	Impulsive, carefree, playful, zany
Boisterousness	Mischievous, rowdy, loud, prankish
Adventure	Brave, venturous, fearless, reckless
Energy	Active, assertive, dominant, energetic
Conceit	Boastful, conceited, egotistical
Vanity	Affected, vain, chic, dapper, jaunty
Indiscretion	Nosey, snoopy, indiscreet, meddlesome
Sensuality	Sexy, passionate, sensual, flirtatious

- Data collection: subjects whether these words describe them.

Personality assessment: the data
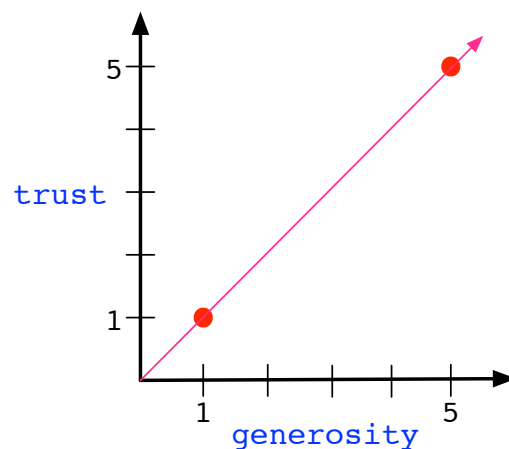
Matrix of data (1 = strongly disagree, 5 = strongly agree)

	shy	merry	tense	boastful	forgiving	quiet
Person 1	4	1	1	2	5	5
Person 2	1	4	4	5	2	1
Person 3	2	4	5	4	2	2
$\vdots$						

How to extract important directions?

- Treat each column as a data point, find tight clusters
- Treat each row as a data point, apply PCA
- Or factor analysis, independent component analysis, etc.

What would PCA accomplish?

E.g.: Suppose two traits (generosity, trust) are so highly correlated that each person either answers "1" to both or "5" to both.

A single PCA dimension would entirely account for both traits.

Personality assessment: the data

Matrix of data (1 = strongly disagree, 5 = strongly agree)

	shy	merry	tense	boastful	forgiving	quiet
Person 1	4	1	1	2	5	5
Person 2	1	4	4	5	2	1
Person 3	2	4	5	4	2	2

⋮

Methodology: apply PCA to the rows of this matrix.

The "Big Five" taxonomy

Extraversion
−: quiet (-.83), reserved (-.80), shy (-.75), silent (-.71)
+: talkative (.85), assertive (.83), active (.82), energetic (.82)

Agreeableness
−: fault-finding (-.52), cold (-.48), unfriendly (-.45), quarrelsome (-.45)
+: sympathetic (.87), kind (.85), appreciative (.85), affectionate (.84)

Conscientousness
−: careless (-.58), disorderly (-.53), frivolous (-.50), irresponsible (-.49)
+: organized (.80), thorough (.80), efficient (.78), responsible (.73)

Neuroticism
−: stable (-.39), calm (-.35), contented (-.21)
+: tense (.73), anxious (.72), nervous (.72), moody (.71)

Openness
−: commonplace (-.74), narrow (-.73), simple (-.67), shallow (-.55)
+: imaginative (.76), intelligent (.72), original (.73), insightful (.68)