Method: PCA

Principal component analysis (PCA) is used to analyze one table of quantitative data. PCA mixes the input variables to give new variables, called principal components. The first principal component is the line of best fit. It is the line that maximizes the inertia (similar to variance) of the cloud of data points. Subsequent components are defined as orthogonal to previous components and maximize the remaining inertia.

PCA gives one map for the rows (called factor scores), and one map for the columns (called loadings). These 2 maps are related, because they both are described by the same components. However, these 2 maps project different kinds of information onto the components, and so they are interpreted differently. Loadings describe the column variables. Loadings are interpreted by the angle between them, and their distance from the origin.

The distance from the origin is important in both maps, because squared distance from the mean is inertia (variance, information; see sum of squares as in ANOVA/regression). Because of the Pythagorean Theorem, the total information contributed by a data point (its squared distance to the origin) is also equal to the sum of its squared factor scores.

Dataset : oasis\_cross-sectional

The Open Access Series of Imaging Studies (OASIS) is a project aimed at making MRI data sets of the brain freely available to the scientific community. It is a cross-sectional MRI Data in Young, Middle Aged, Nondemented and Demented Older Adults. This set consists of a cross-sectional collection of 216 subjects (rows) aged 33 to 96. The subjects include both men and women. It measures 216 subjects (rows) on 4 quantitative variables (columns). The 5 qualitative variables have been used to interpret the data.

Measures included in the study:

Age: Age (years)

Gender: Male or Female

Education: Years of Education - Recoded into categories from 1 (highest) to 5 (lowest)

SES: Socioeconomic status classified into categories from 1 (highest) to 5 (lowest)

MMSE: Mini-Mental State Examination score – range from 0 (worst) to 30 (best)

CDR: Clinical Dementia Rating - (0 = no dementia, 0.5 = very mild AD, 1 = mild AD, 2 = moderate AD)

ASF : Atlas scaling factor (unitless). Computed scaling factor that transforms native-space brain and skull to the atlas target (i.e. the determinant of the transform matrix)

eTIV: Estimated total intracranial volume (cm3)

nWBV: Normalized whole brain volume, expressed as a percent of all voxels in the atlas-masked image that are labeled as gray or white matter by the automated tissue segmentation process