

Psych 253

Advanced statistical modeling

Correlation Analysis

Daniel Yamins

Wu Tsai Neurosciences Institute
Departments of Psychology and Computer Science
Stanford University

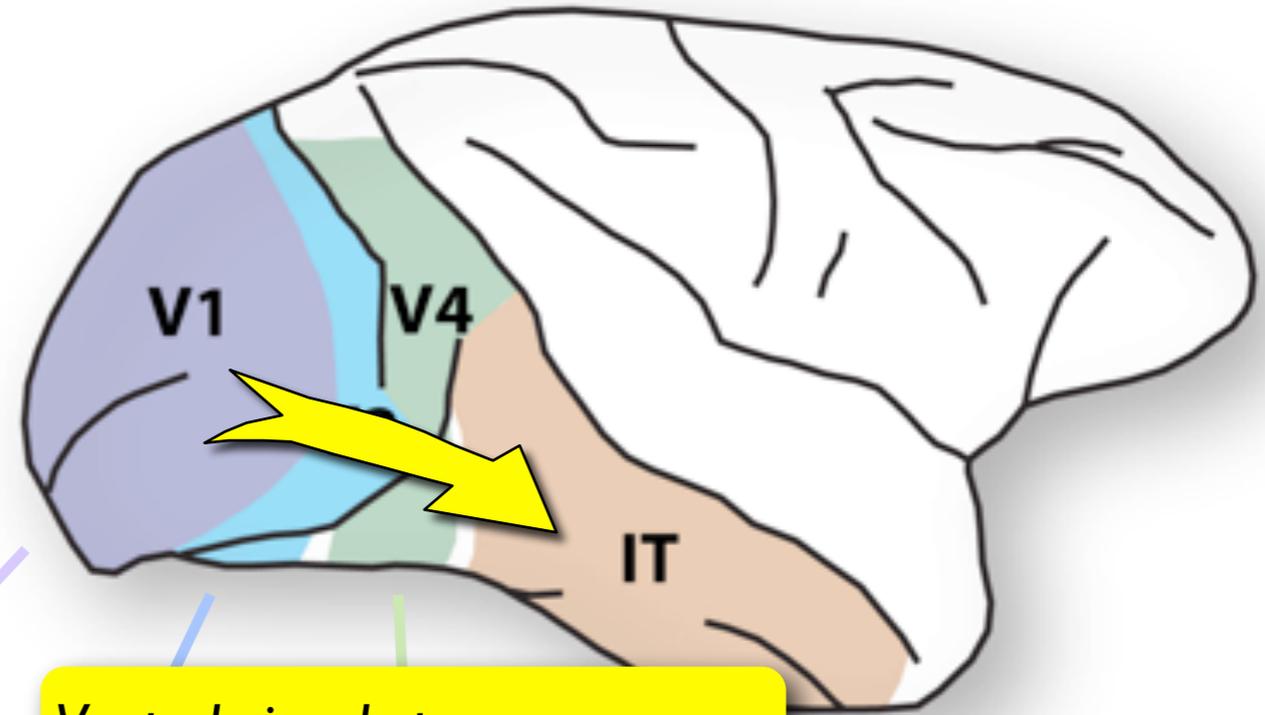
Russ Poldrack

Department of Psychology
Stanford University

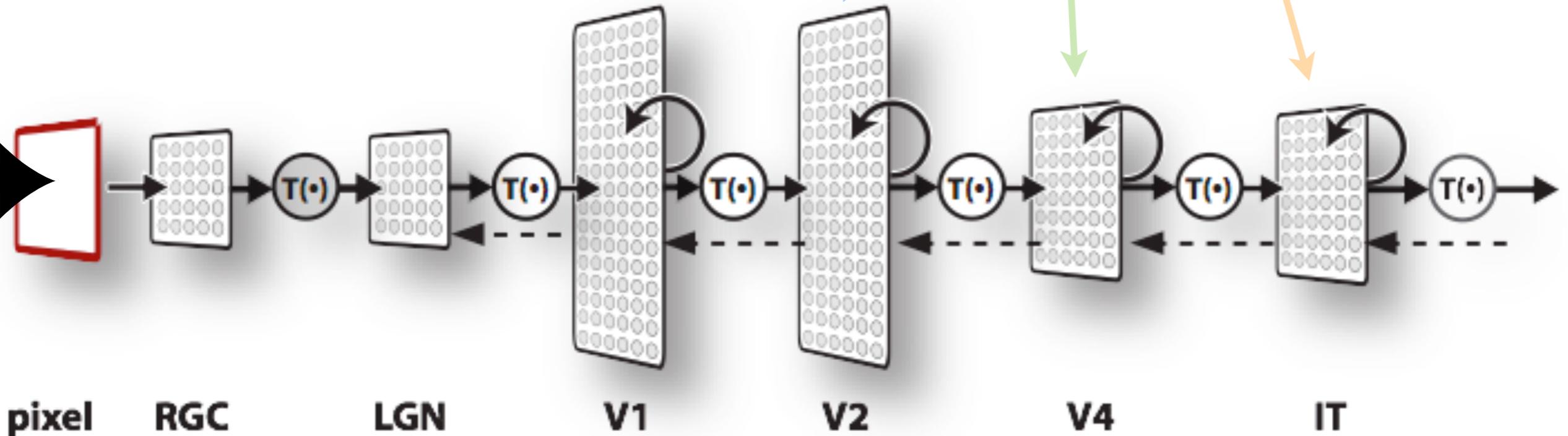
Example I: The Ventral Visual Pathway



rhesus macaque (macaca mulatta)



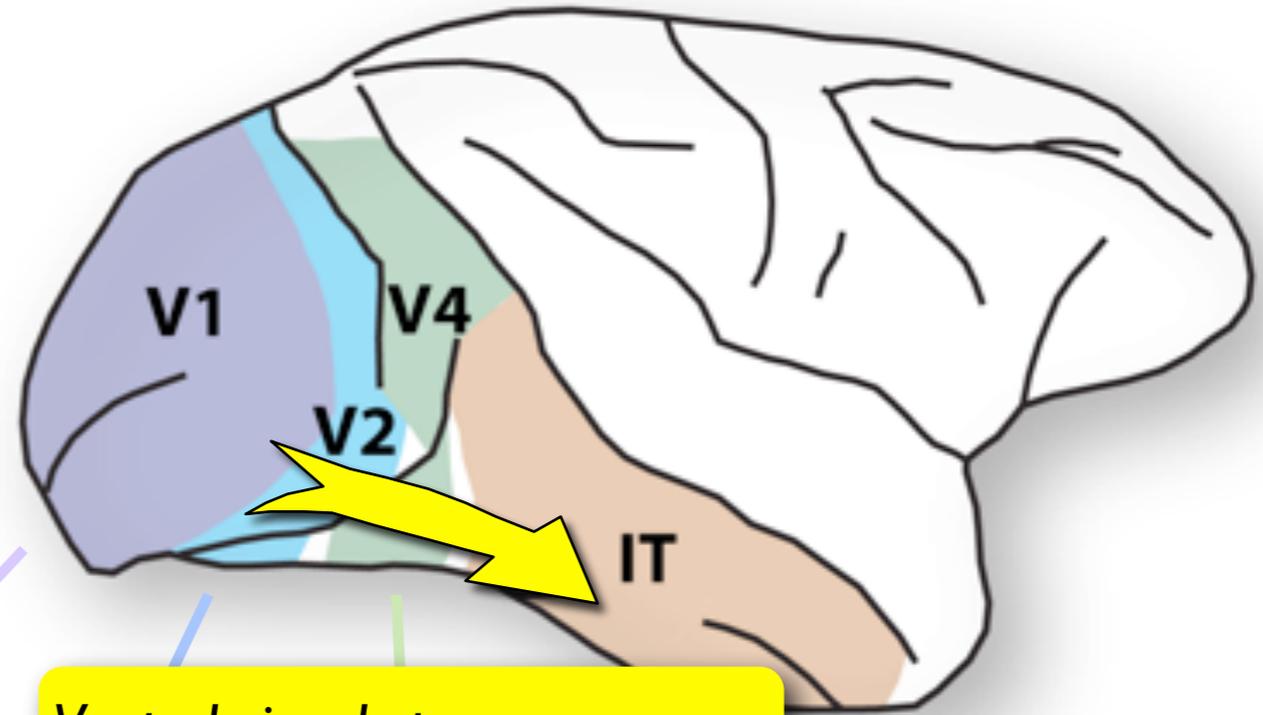
Ventral visual stream



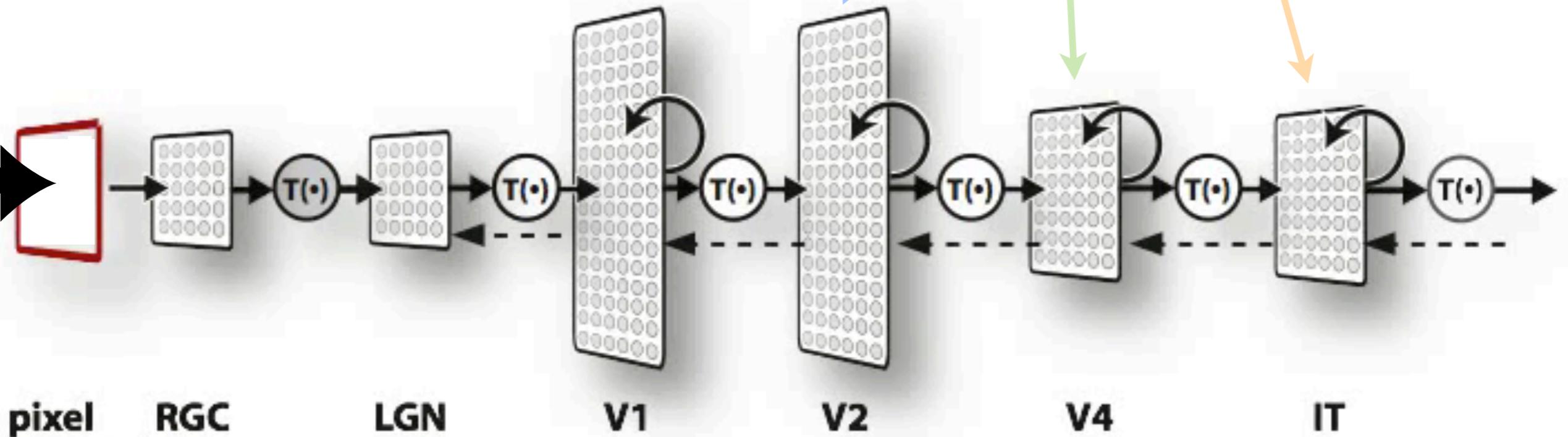
Example I: The Ventral Visual Pathway



rhesus macaque (macaca mulatta)

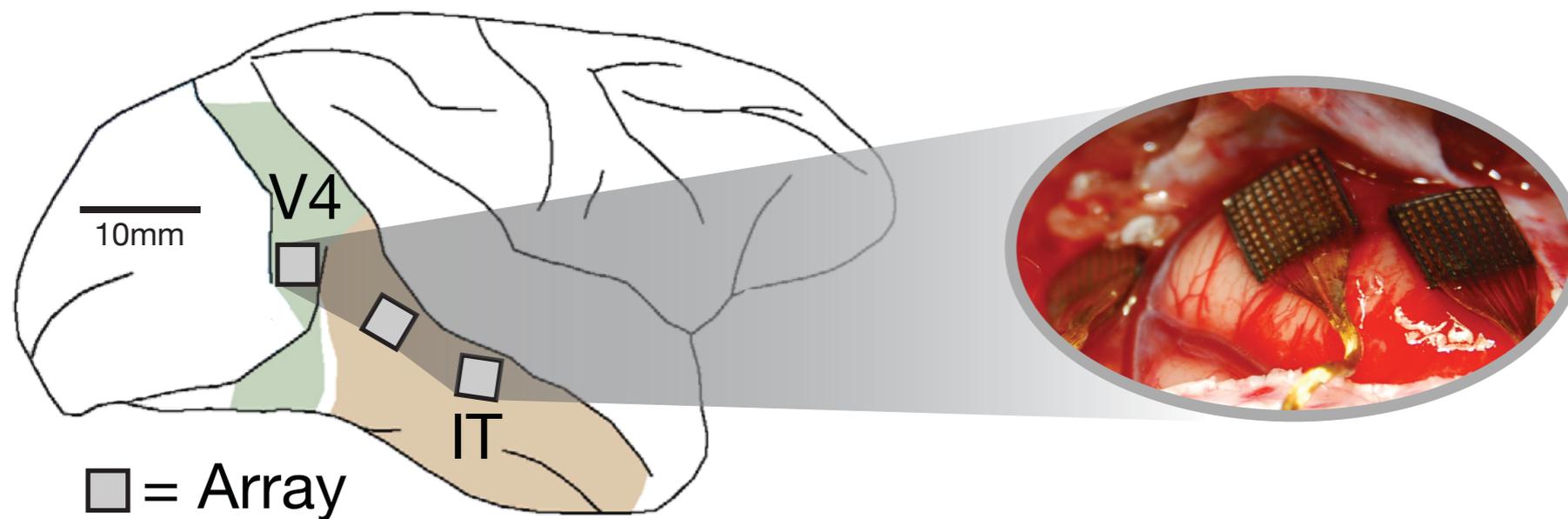


Ventral visual stream

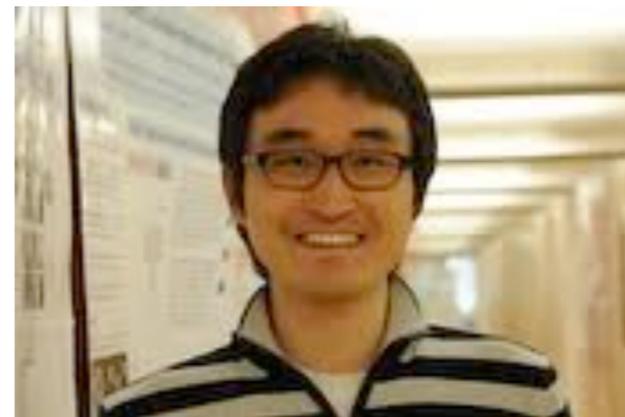


Example I: The Ventral Visual Pathway

Multi-array electrophysiology in macaque V4 and IT.



About 300 total sites



Ha Hong



Jim DiCarlo

Example I: The Ventral Visual Pathway

5760 images

64 objects

8 categories

uncorrelated photo backgrounds

Low variation



... 640 images

Medium variation



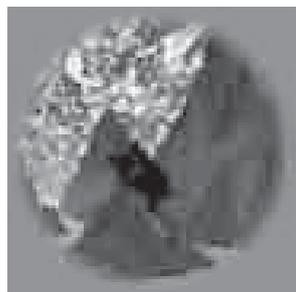
... 2560 images

High variation



... 2560 images

Animals



Boats



Cars



Chairs



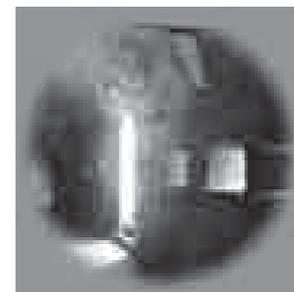
Faces



Fruits



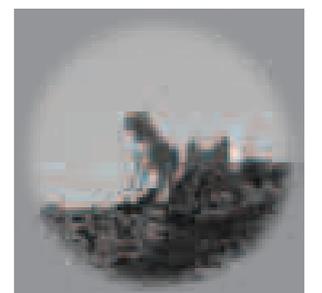
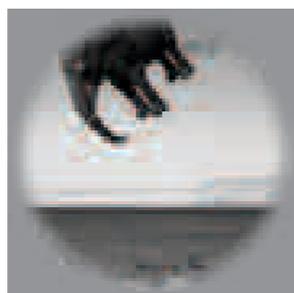
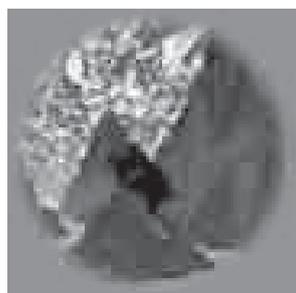
Planes



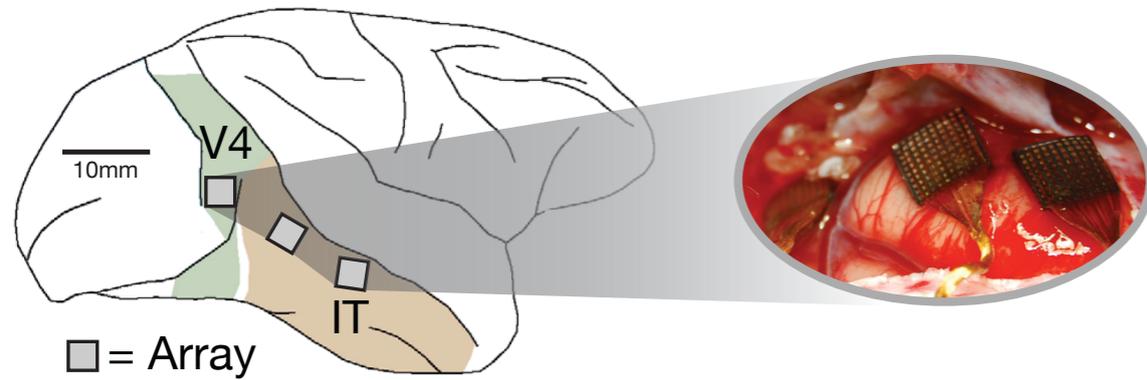
Tables



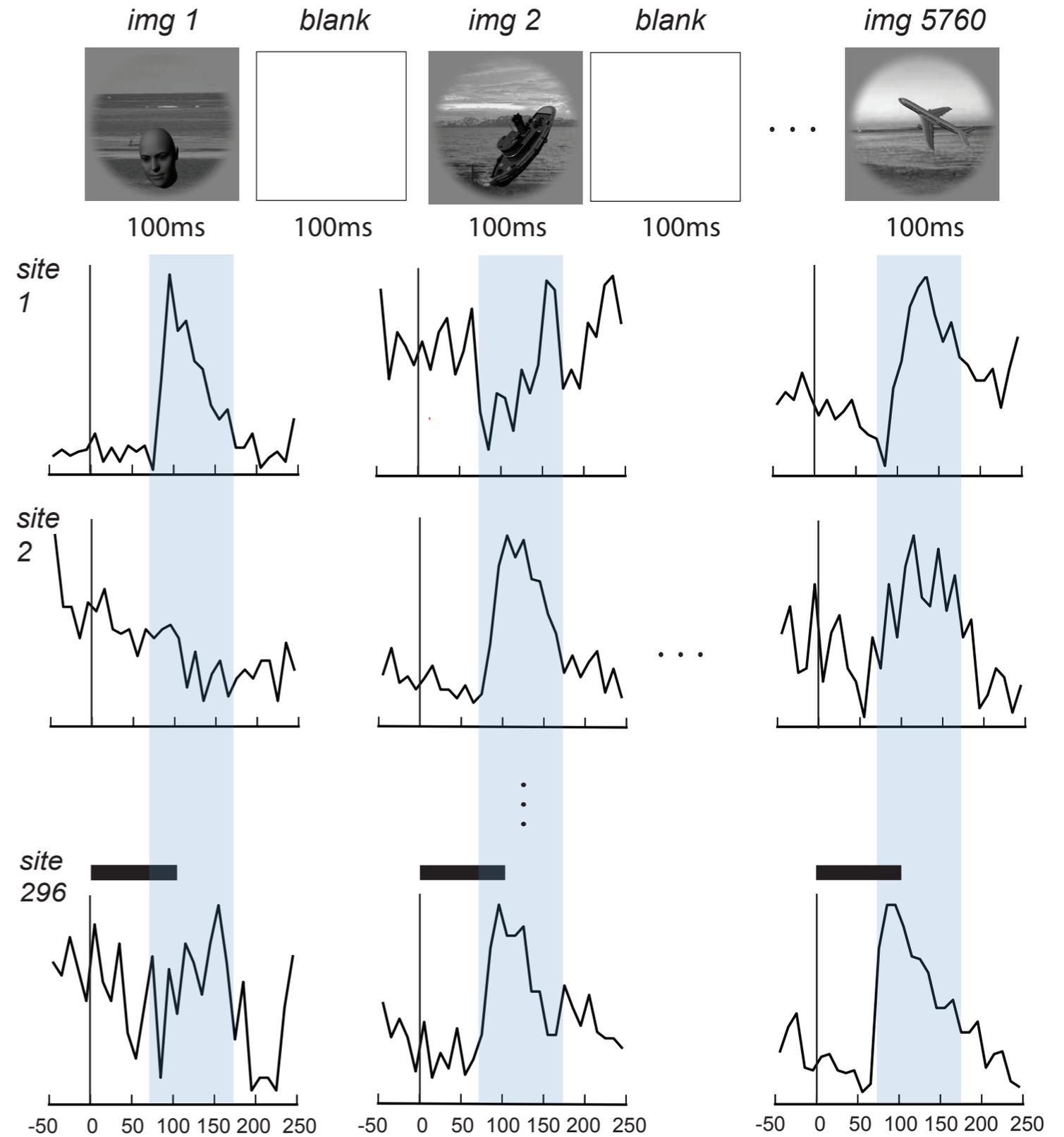
Pose, position, scale, and background variation



Example I: The Ventral Visual Pathway



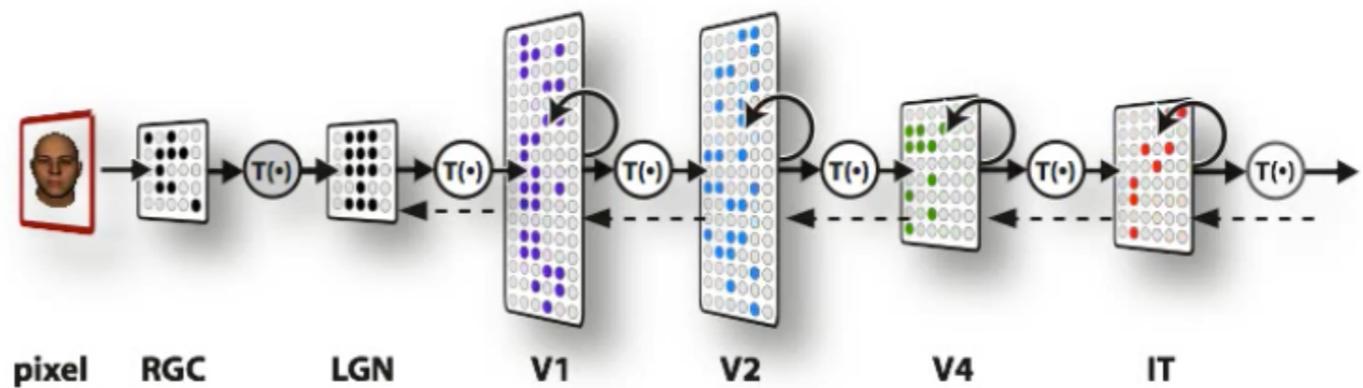
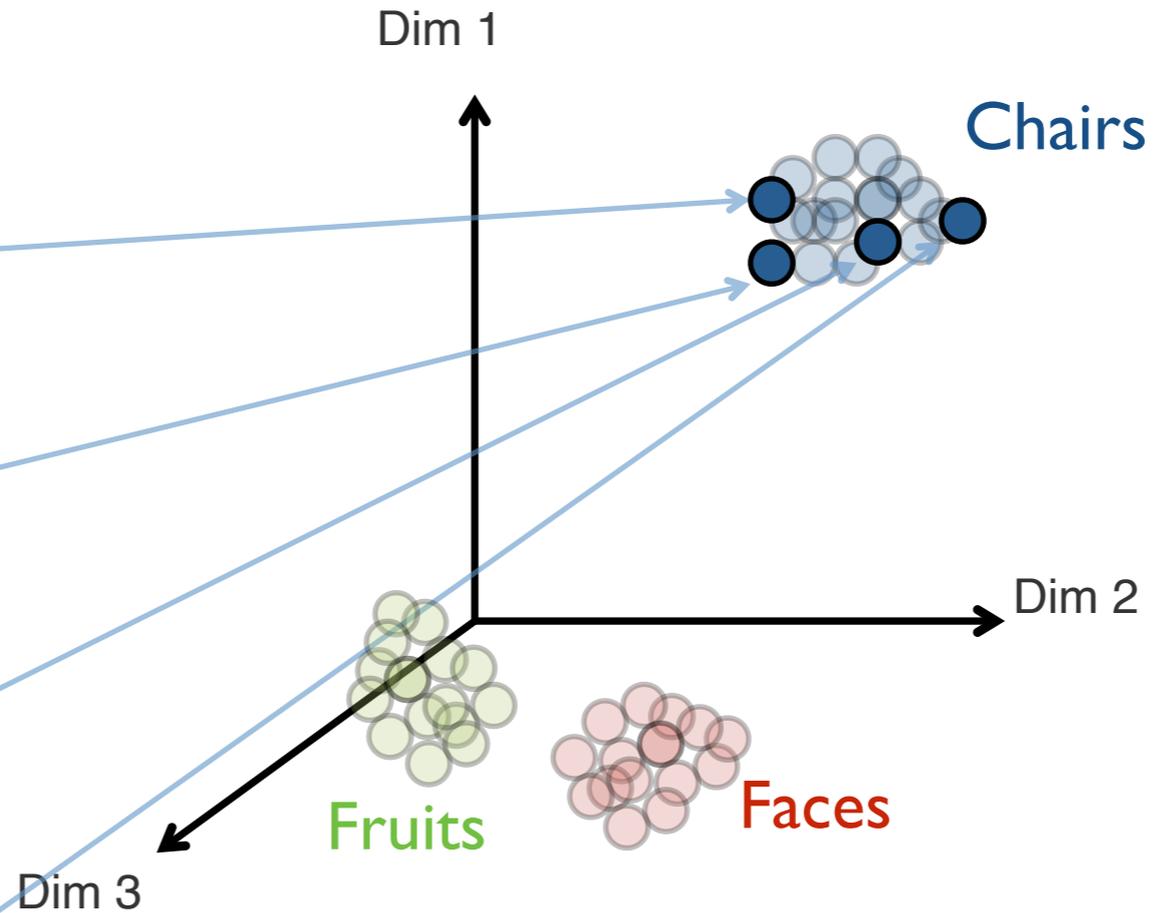
About 300 total sites



Data as Feature Representation

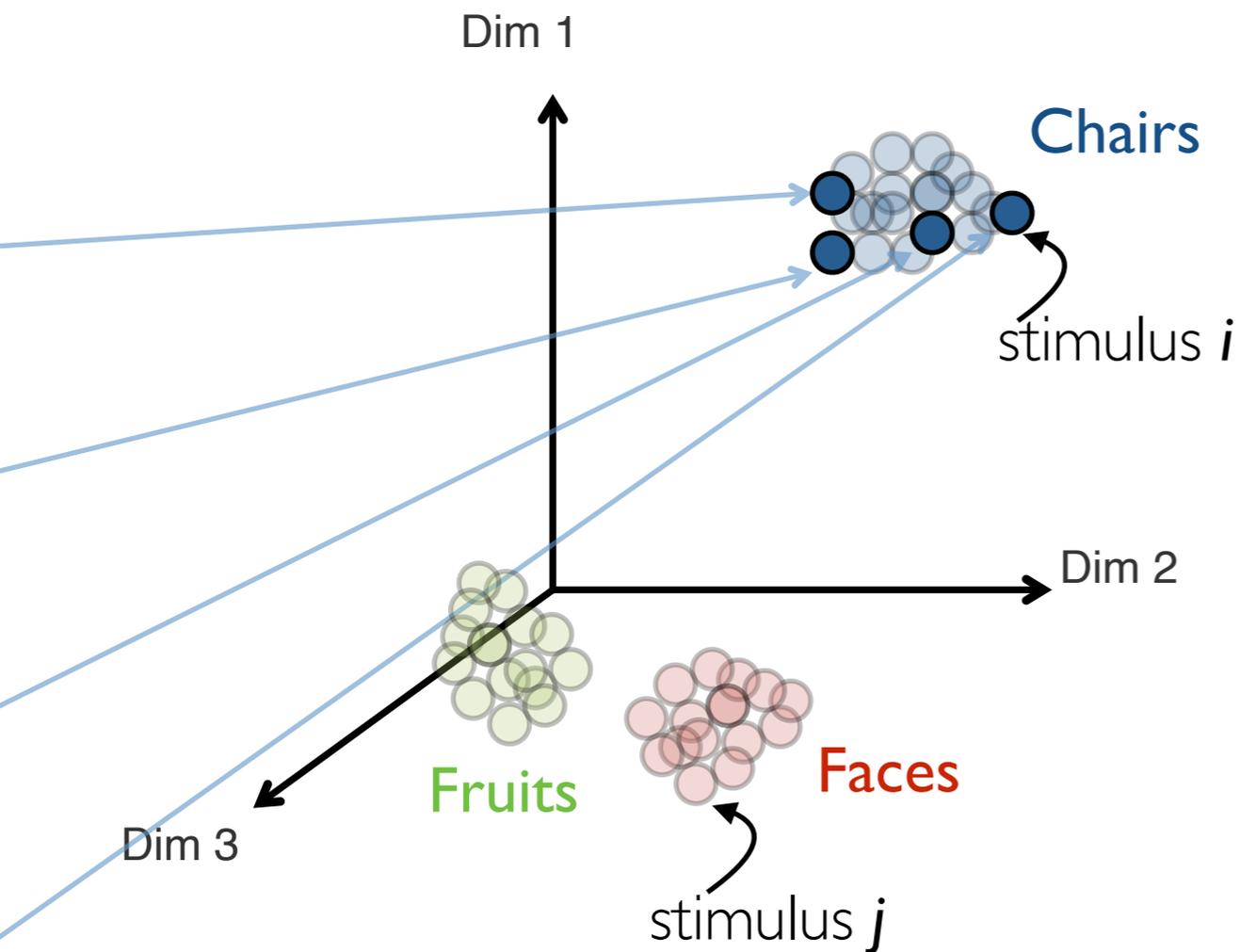
Pixel space: $R \sim 10000000$

IT feature space: $R^{4000(?)}$



Data as Feature Representation

IT feature space

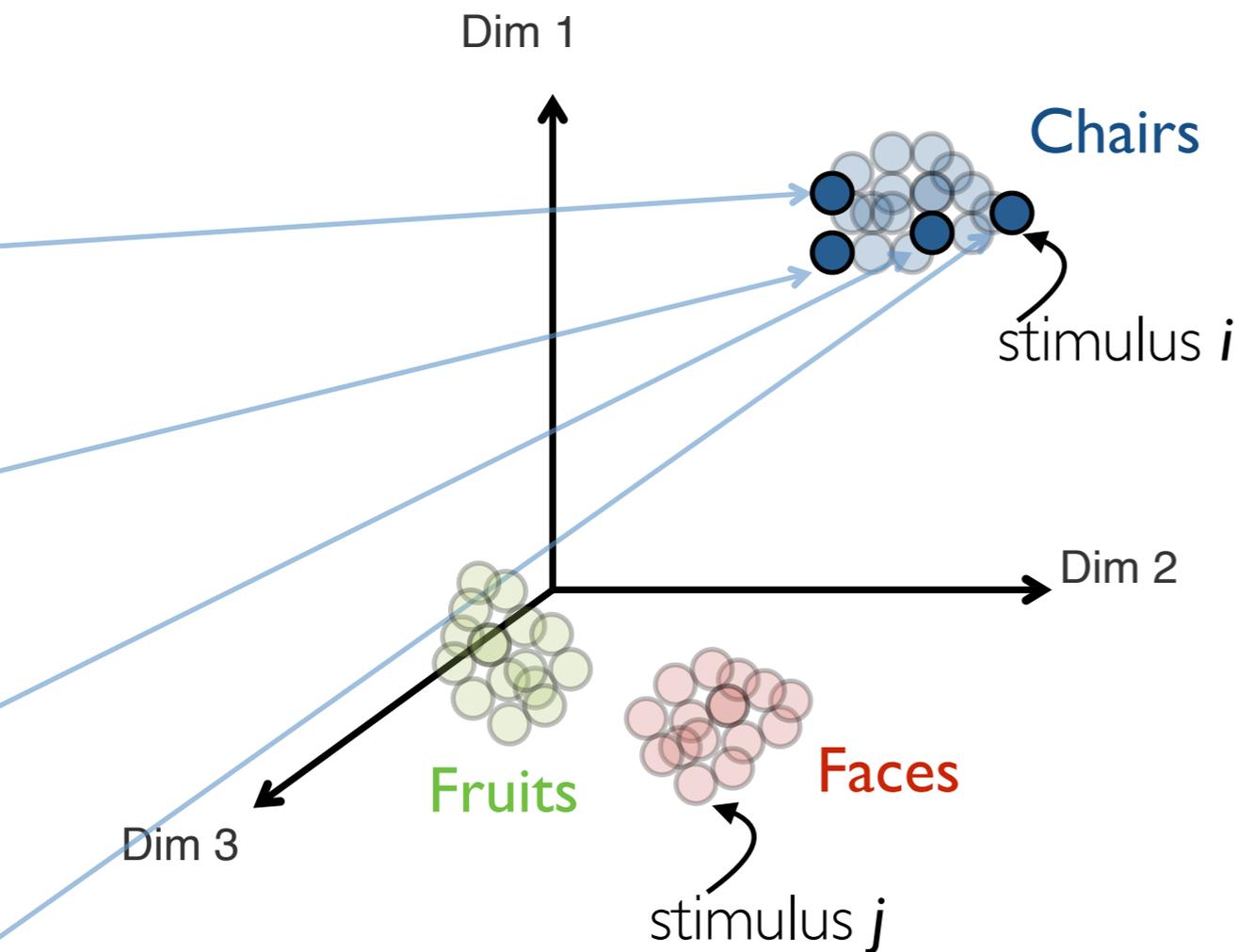


r_i = vector of neural responses to stimulus i

r_j = vector of neural responses to stimulus j

Data as Feature Representation

IT feature space



r_i = vector of neural responses to stimulus i

r_j = vector of neural responses to stimulus j

$$dist(i, j) = euclidean(r_i, r_j)$$

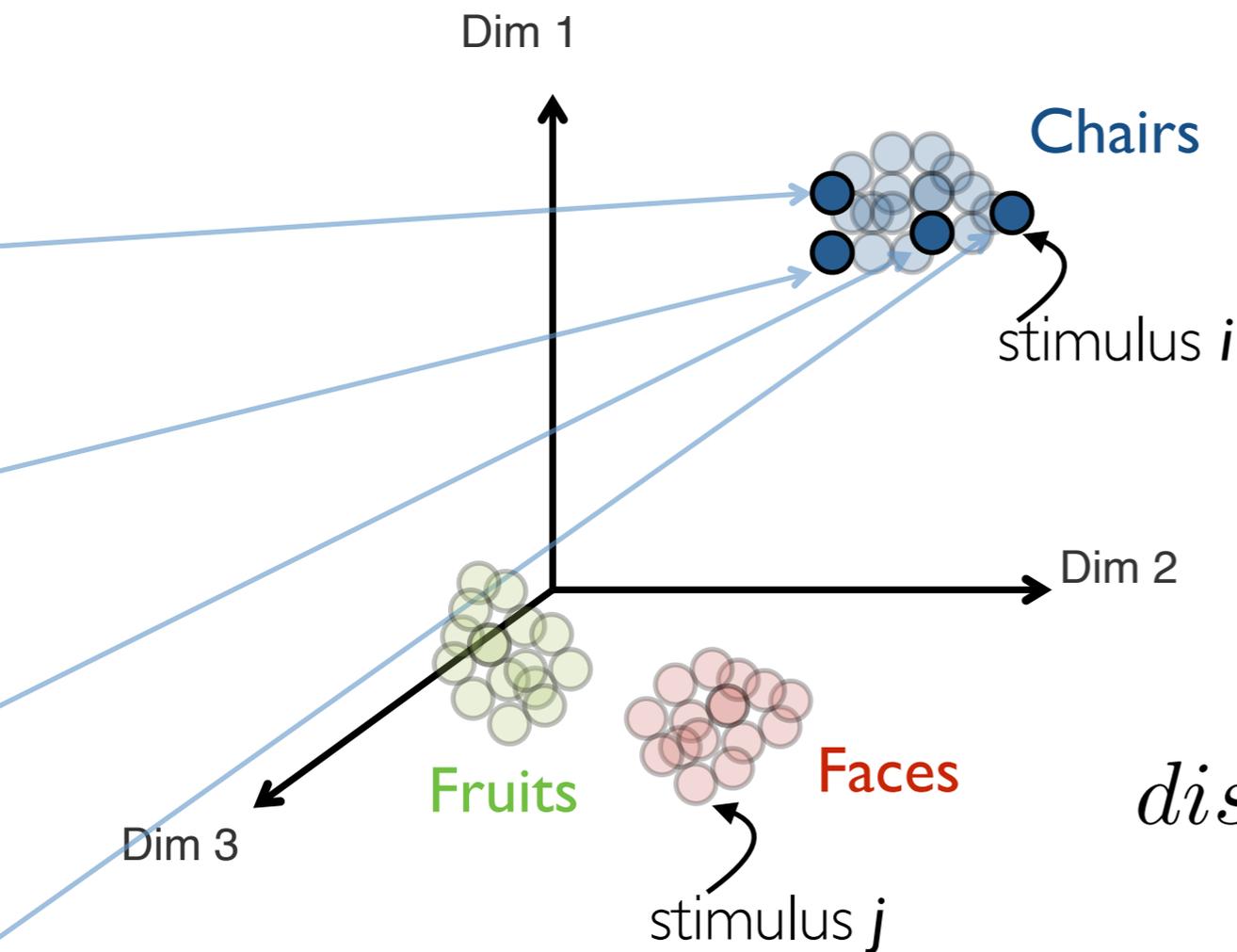
$$= \sqrt{\sum_{k=0}^{N-1} (r_{ik} - r_{jk})^2}$$

N = # of neurons (# of dimensions)

k = variable ranging over neurons

Data as Feature Representation

IT feature space



r_i = vector of neural responses to stimulus i

r_j = vector of neural responses to stimulus j

instead of euclidean distance:

$$dist(i, j) = 1 - correlation(r_i, r_j)$$

$$= 1 - \frac{cov(r_i, r_j)}{\sqrt{var(r_i) \cdot var(r_j)}}$$

$$\mathbf{E}_k[r_i r_j] - \mathbf{E}_k[r_i] \mathbf{E}_k[r_j]$$

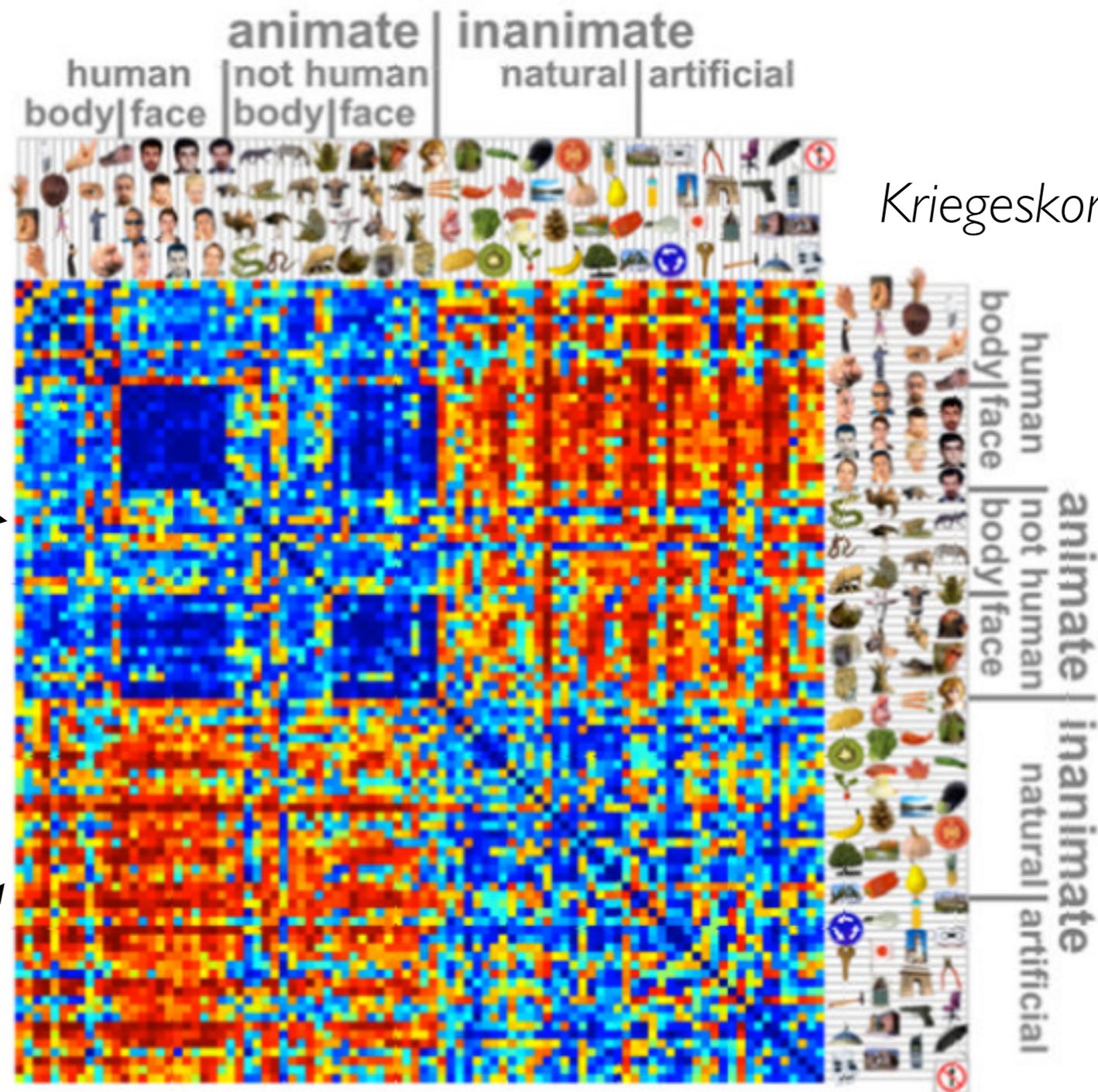
$$\mathbf{E}_k[r_i^2] - \mathbf{E}_k[r_i]^2$$

expectations over neurons

Representational Similarity Analysis

$$M_{ij} = 1 - \text{correlation}(r_i, r_j)$$

Low M_{ij} (blue) means neurons think the stimuli are similar



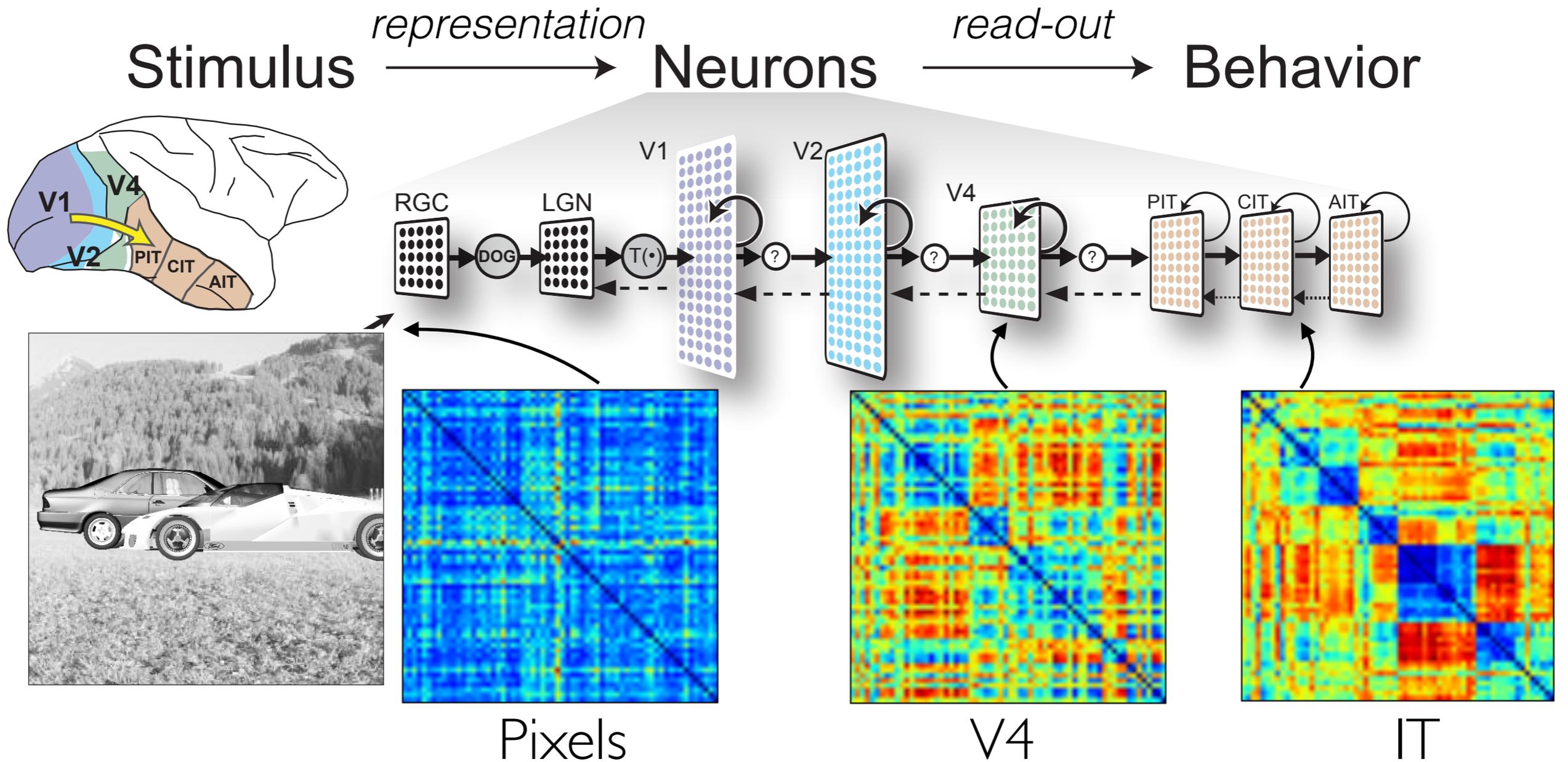
Kriegeskorte, 2008

High M_{ij} (red) means neurons think the two stimuli are different.

Correlation Analysis

[IPYNB: *Correlation Analysis*]

Representational Similarity Analysis

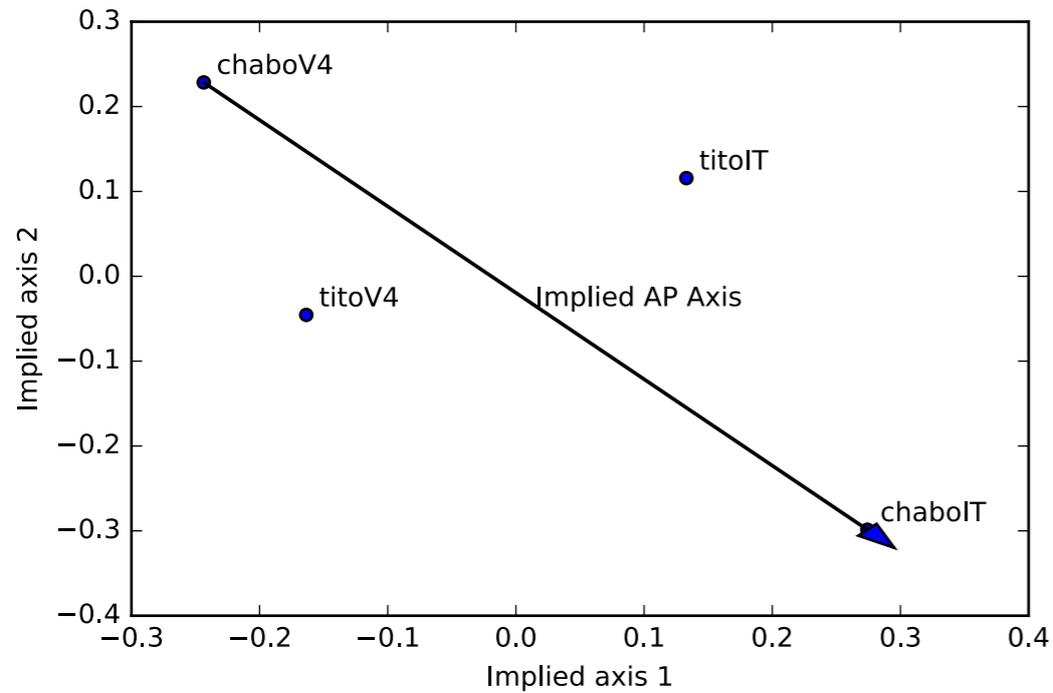


How do we measure similarity or differentness of these correlation matrices?

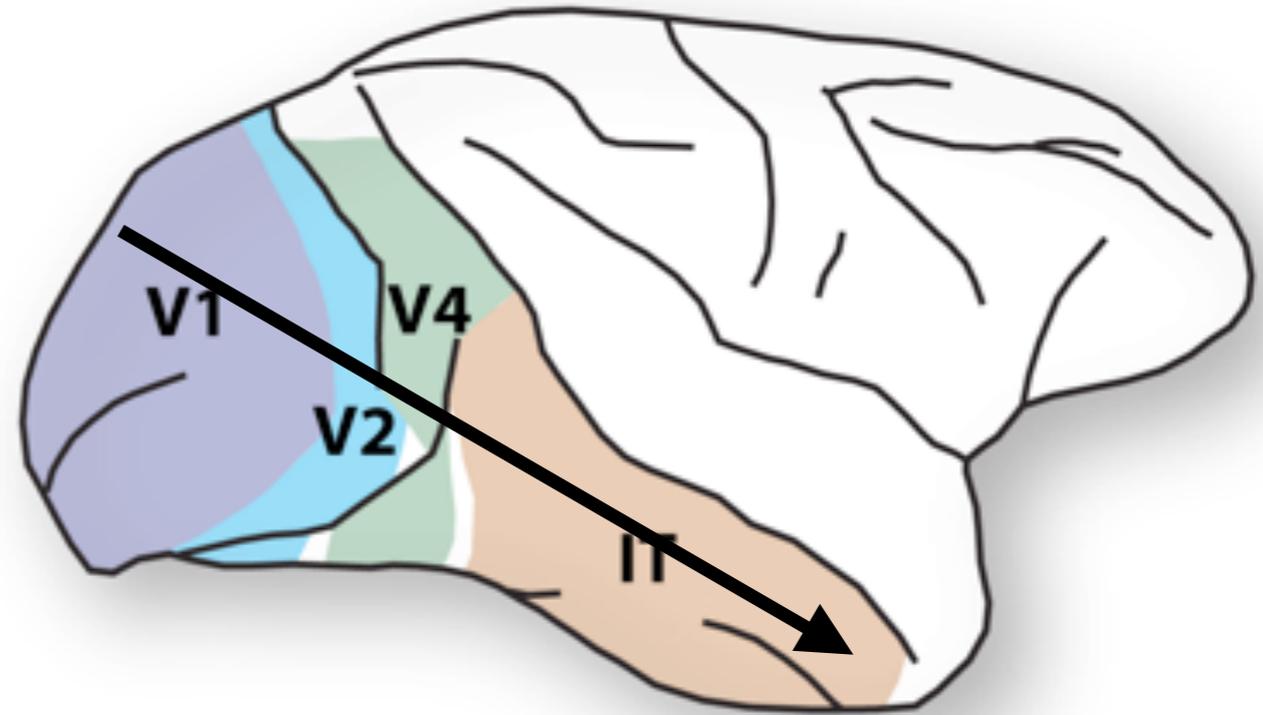
Correlation Analysis

[IPYNB: RSA]

Representational Similarity Analysis



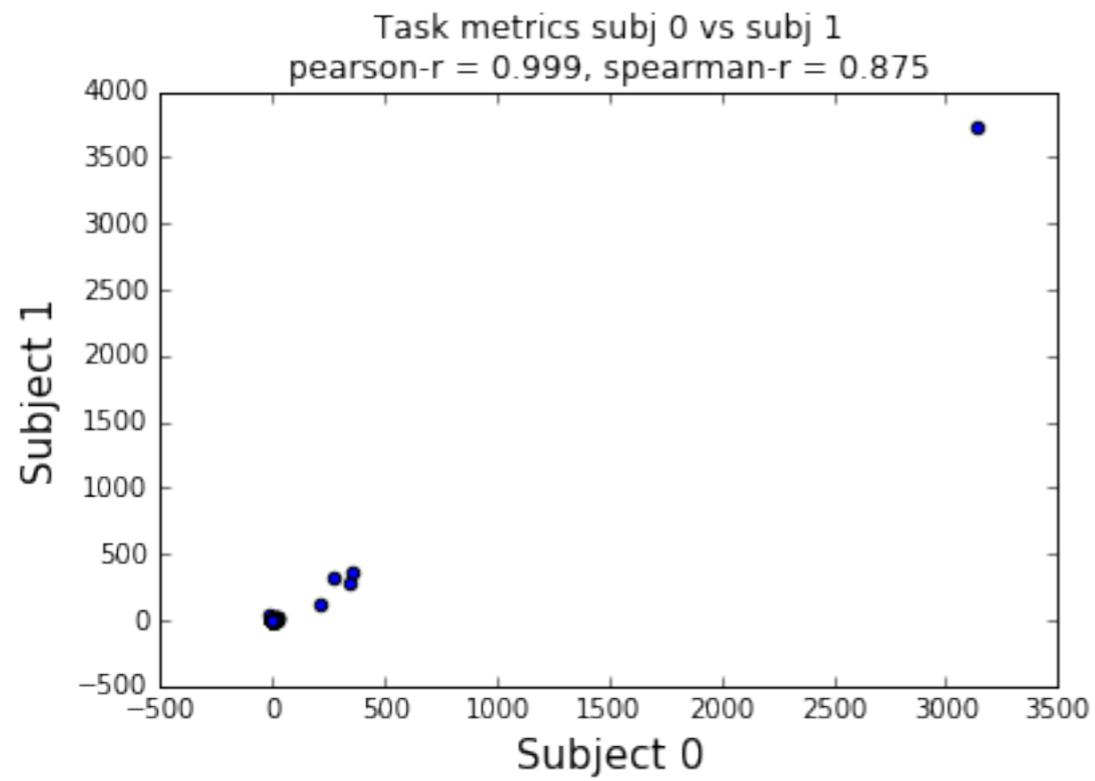
from RSA analysis
+ 2D MDS embedding



Anterior-Posterior Axis
("getting more abstract")

Correlation Analysis

Example from Self-Regulation Ontology Data



Correlation Analysis

Spearman's rank correlation coefficient

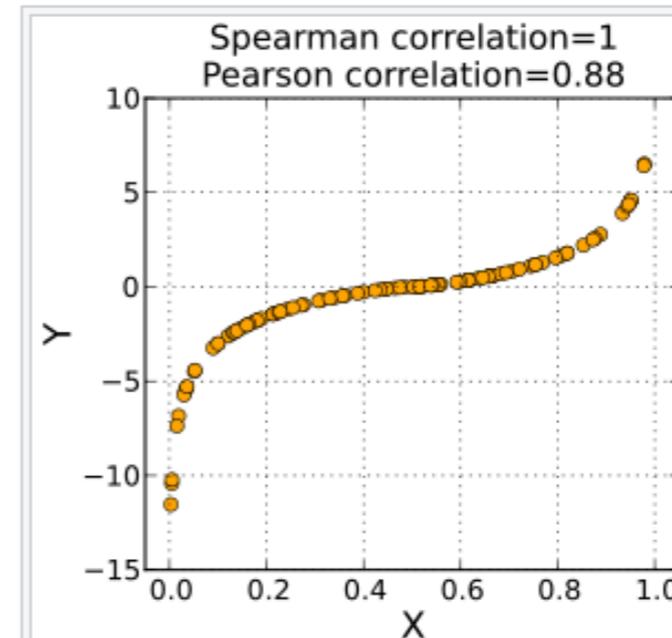
From Wikipedia, the free encyclopedia

In [statistics](#), **Spearman's rank correlation coefficient** or **Spearman's ρ** , named after [Charles Spearman](#) and often denoted by the Greek letter ρ (rho) or as r_s , is a [nonparametric](#) measure of [rank correlation](#) ([statistical dependence](#) between the [rankings](#) of two [variables](#)). It assesses how well the relationship between two variables can be described using a [monotonic](#) function.

The Spearman correlation between two variables is equal to the [Pearson correlation](#) between the rank values of those two variables; while Pearson's correlation assesses linear relationships, Spearman's correlation assesses monotonic relationships (whether linear or not). If there are no repeated data values, a perfect Spearman correlation of +1 or -1 occurs when each of the variables is a perfect monotone function of the other.

Intuitively, the Spearman correlation between two variables will be high when observations have a similar (or identical for a correlation of 1) [rank](#) (i.e. relative position label of the observations within the variable: 1st, 2nd, 3rd, etc.) between the two variables, and low when observations have a dissimilar (or fully opposed for a correlation of -1) rank between the two variables.

Spearman's coefficient is appropriate for both [continuous](#) and discrete [ordinal variables](#).^{[1][2]} Both Spearman's ρ and [Kendall's \$\tau\$](#) can be formulated as special cases of a more [general correlation coefficient](#).



A Spearman correlation of 1 results when the two variables being compared are monotonically related, even if their relationship is not linear. This means that all data points with greater x values than that of a given data point will have greater y values as well. In contrast, this does not give a perfect Pearson correlation.

For a sample of size n , the n [raw scores](#) X_i, Y_i are converted to ranks $rg X_i, rg Y_i$, and r_s is computed as

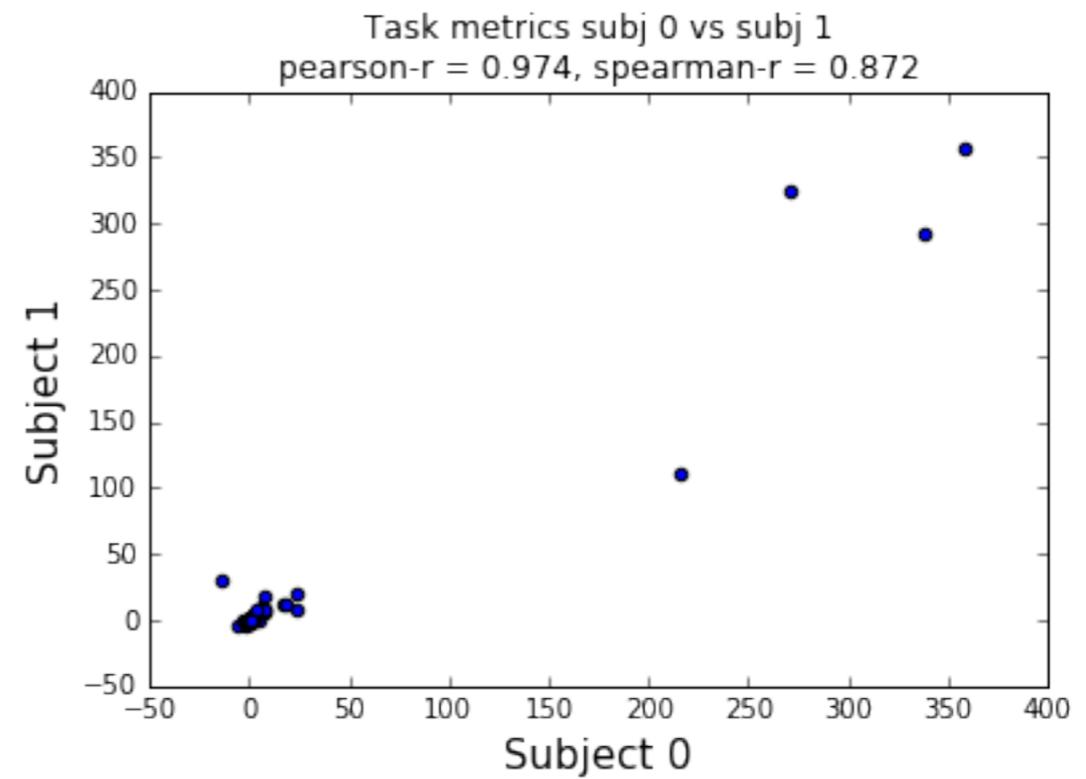
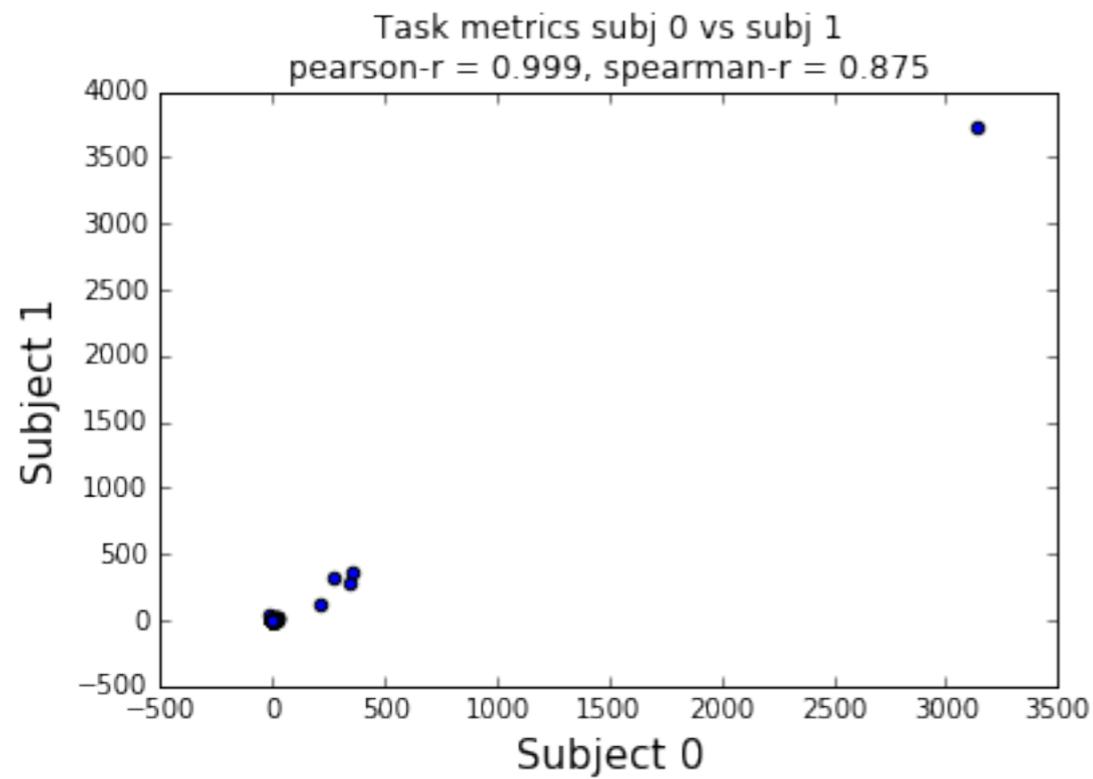
$$r_s = \rho_{rgX, rgY} = \frac{\text{cov}(rgX, rgY)}{\sigma_{rgX} \sigma_{rgY}},$$

where

ρ denotes the usual [Pearson correlation coefficient](#), but applied to the rank variables, $\text{cov}(rgX, rgY)$ is the [covariance](#) of the rank variables, σ_{rgX} and σ_{rgY} are the [standard deviations](#) of the rank variables.

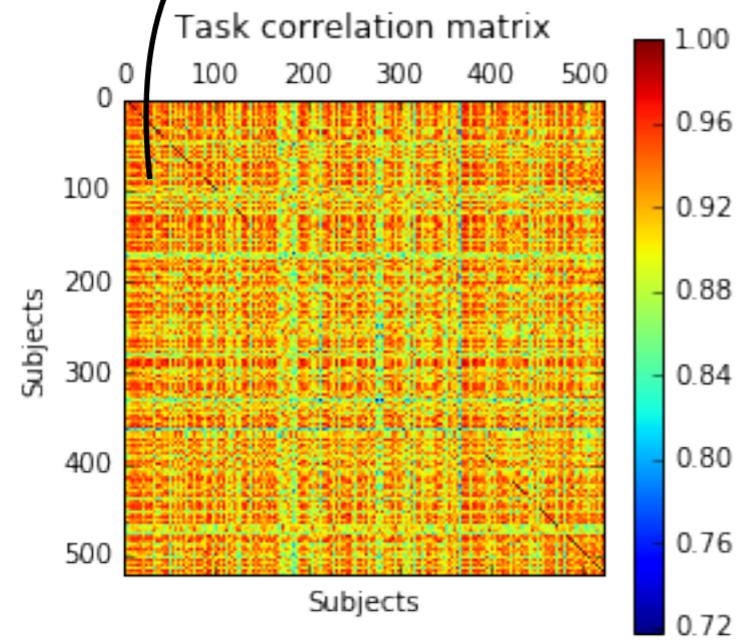
Correlation Analysis

Example from Self-Regulation Ontology Data



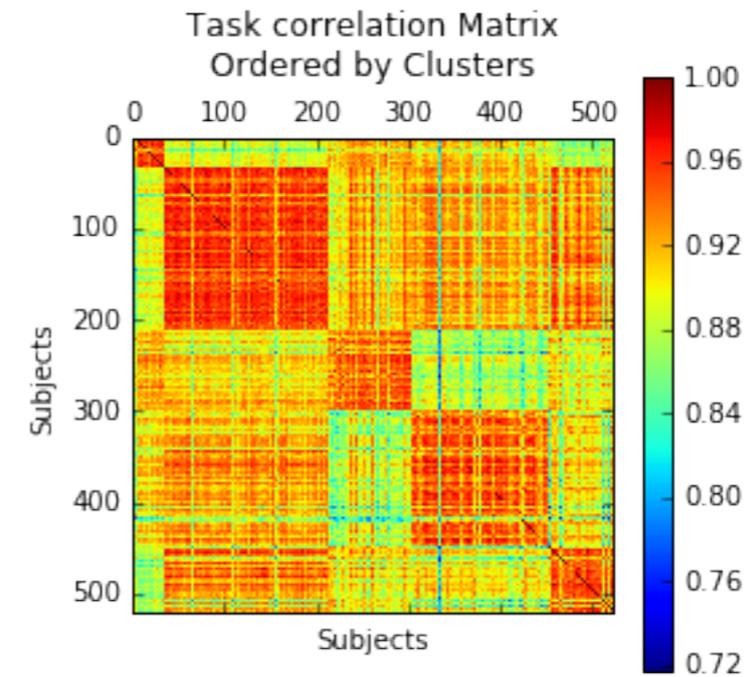
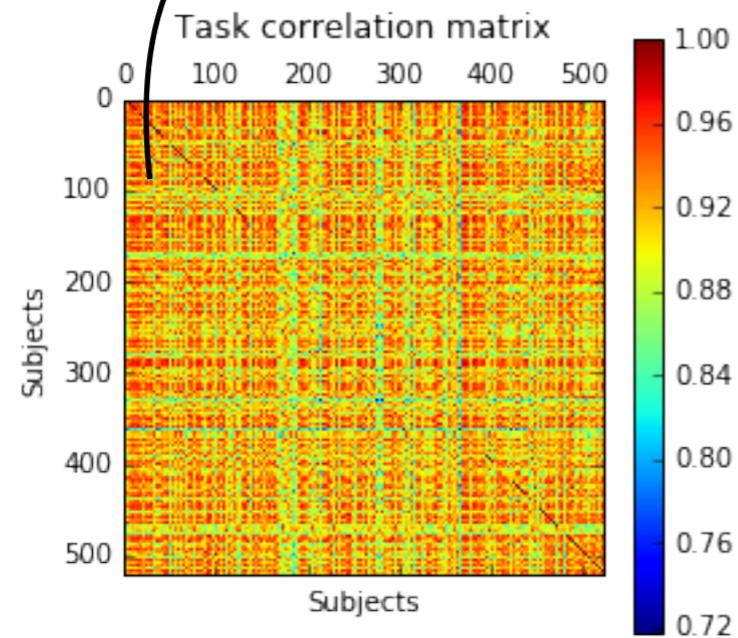
Correlation Analysis

Higher (redder) value means subjects more similar



Correlation Analysis

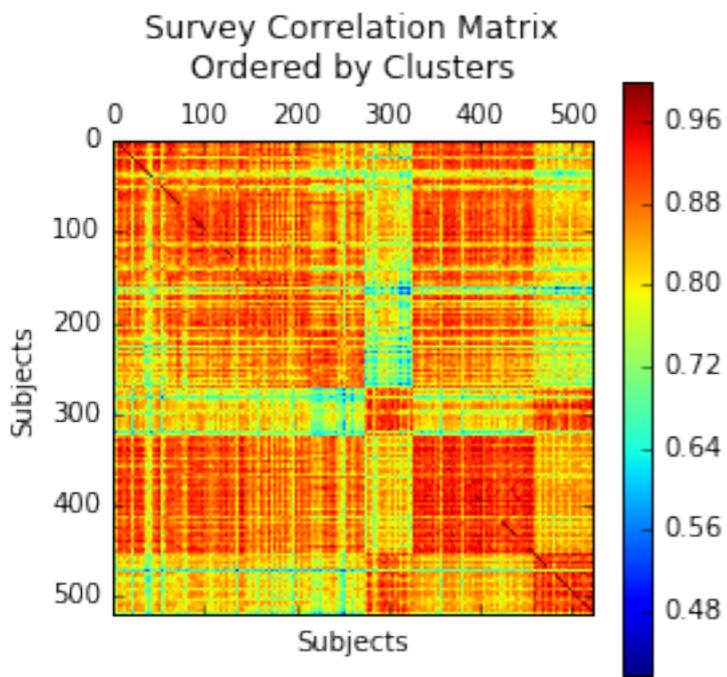
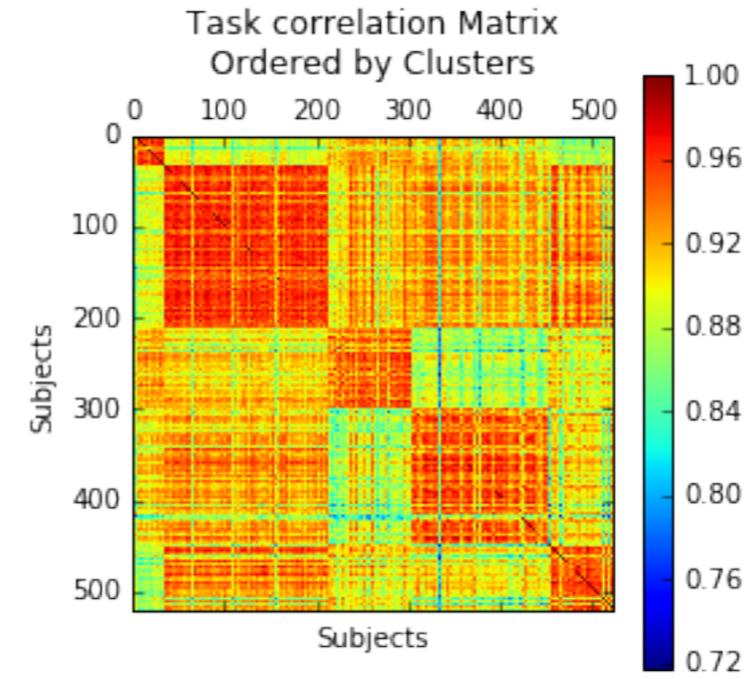
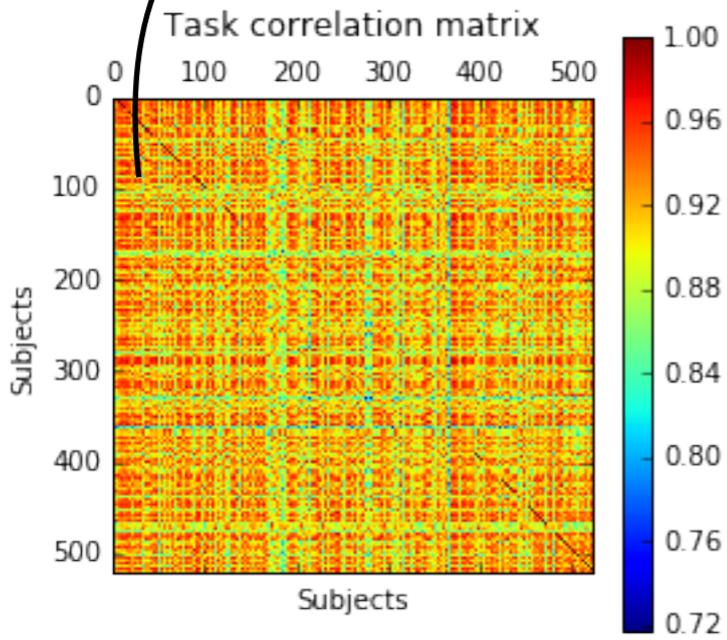
Higher (redder) value means subjects more similar



Ordering based on clustering of subjects, using
Correlation matrix as affinities
(will return to this later lectures)

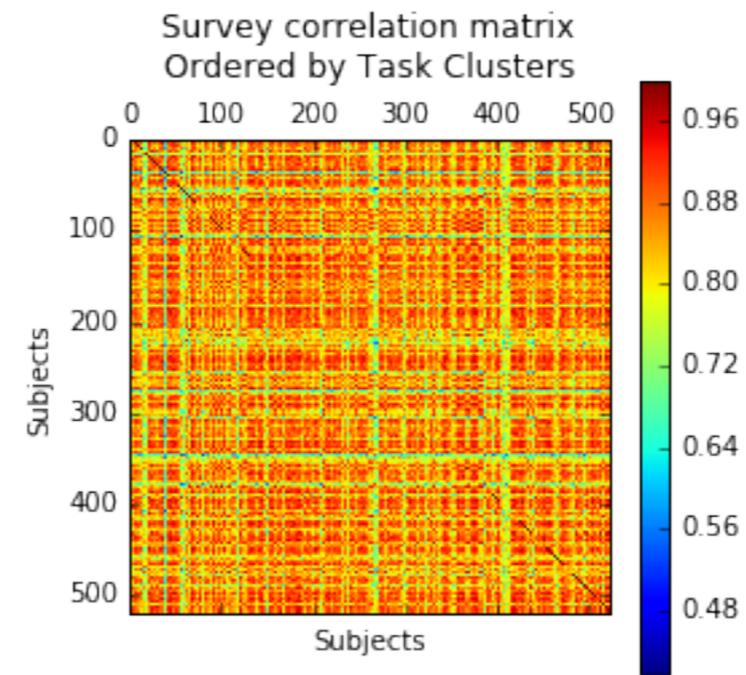
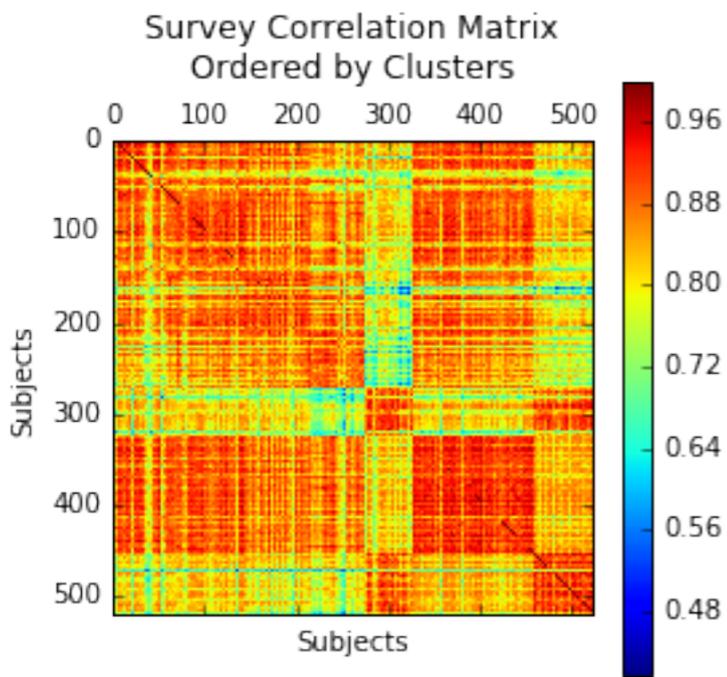
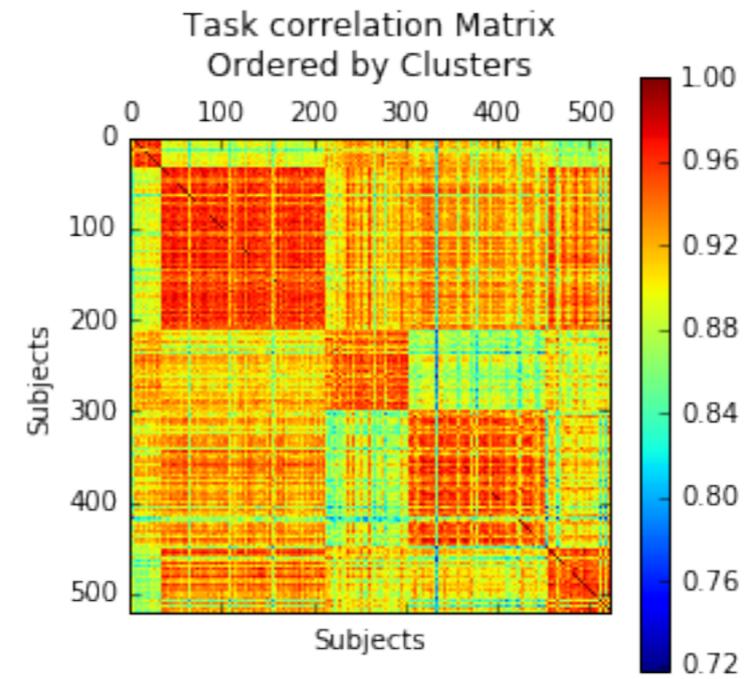
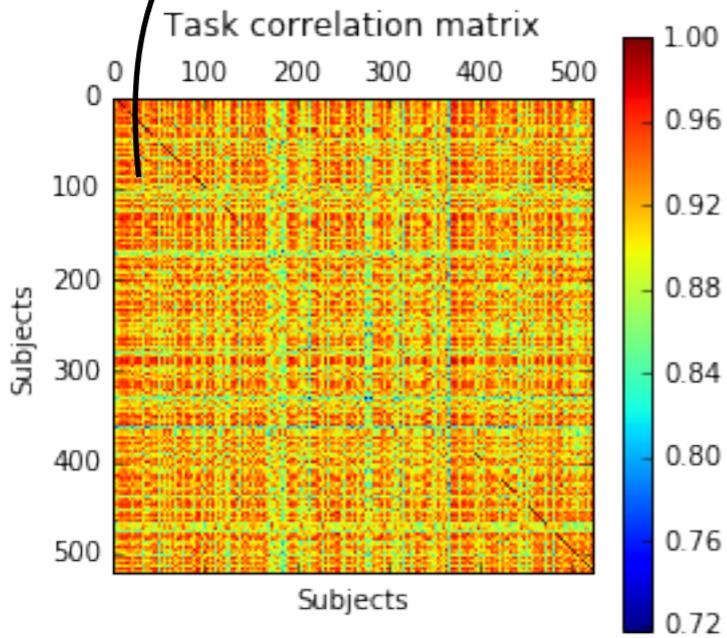
Correlation Analysis

Higher (redder) value means subjects more similar



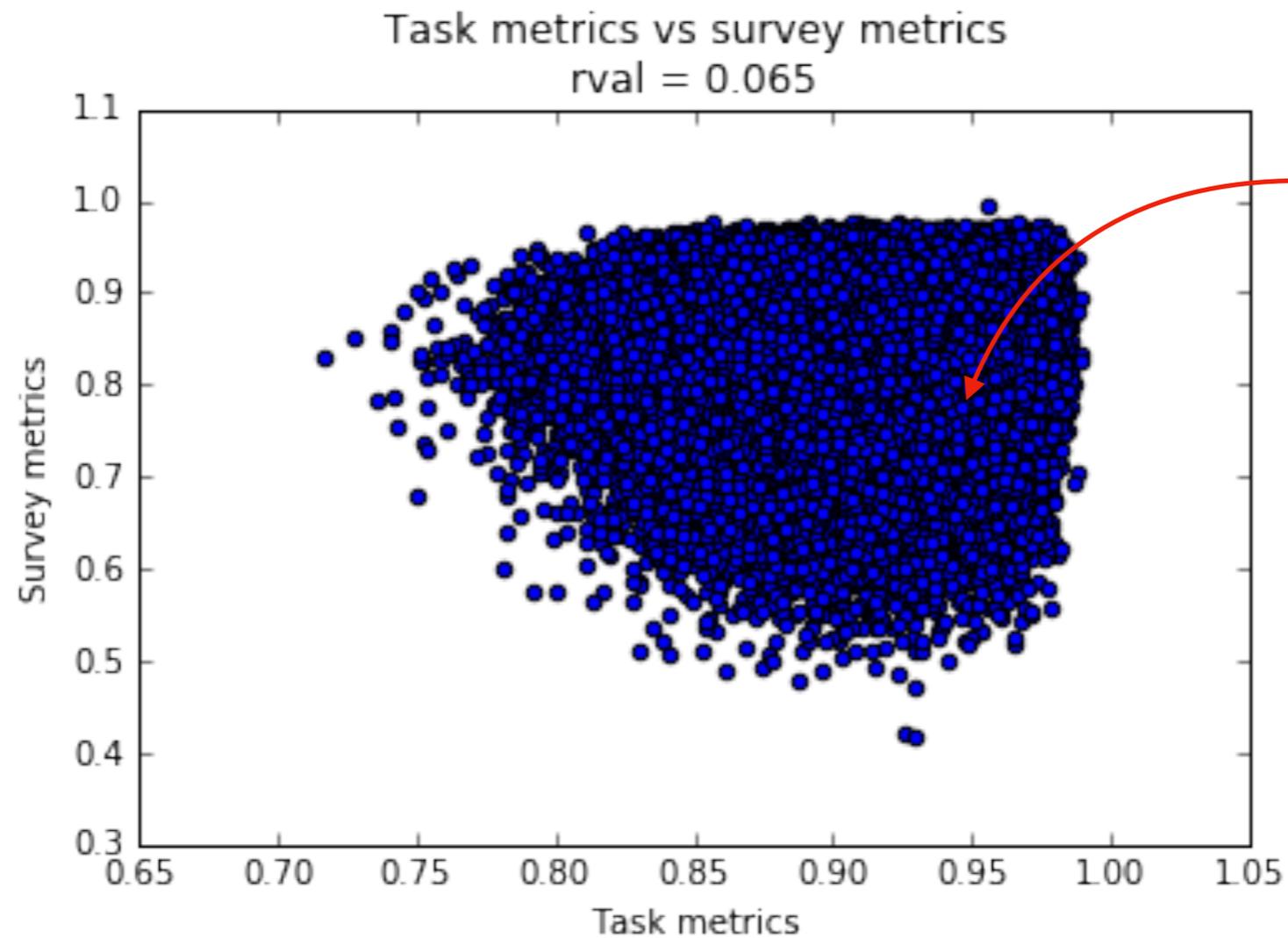
Correlation Analysis

Higher (redder) value means subjects more similar



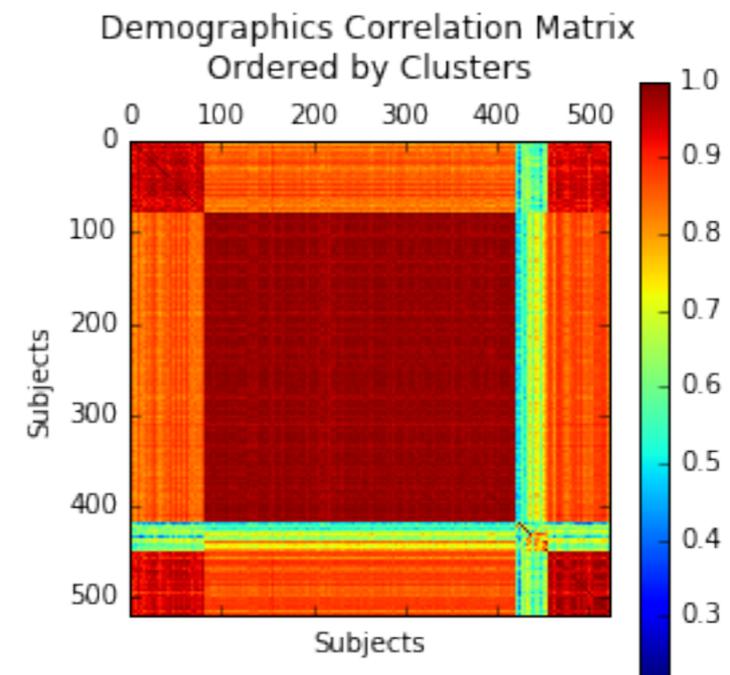
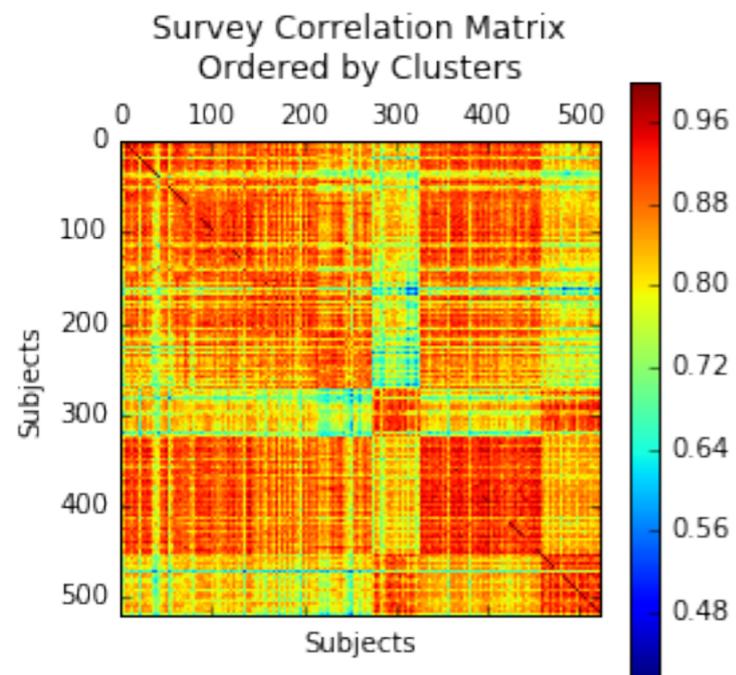
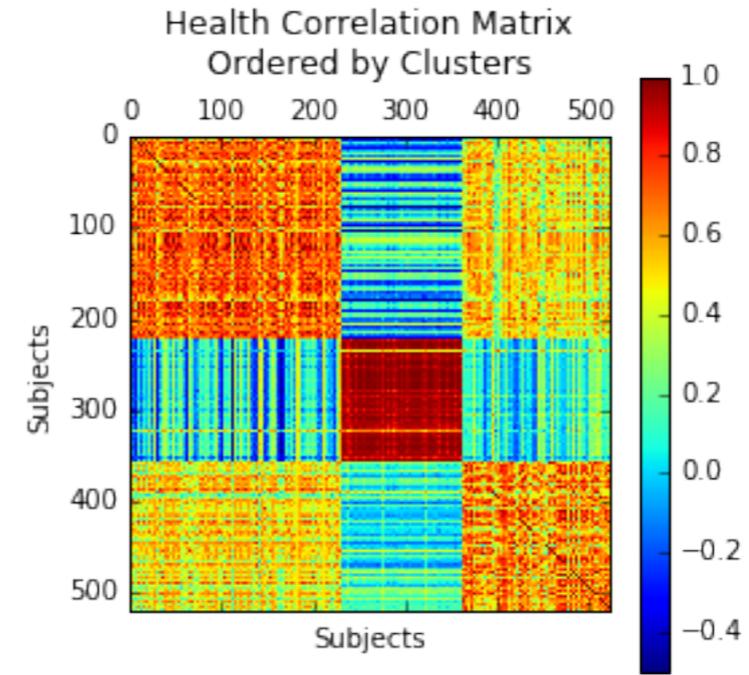
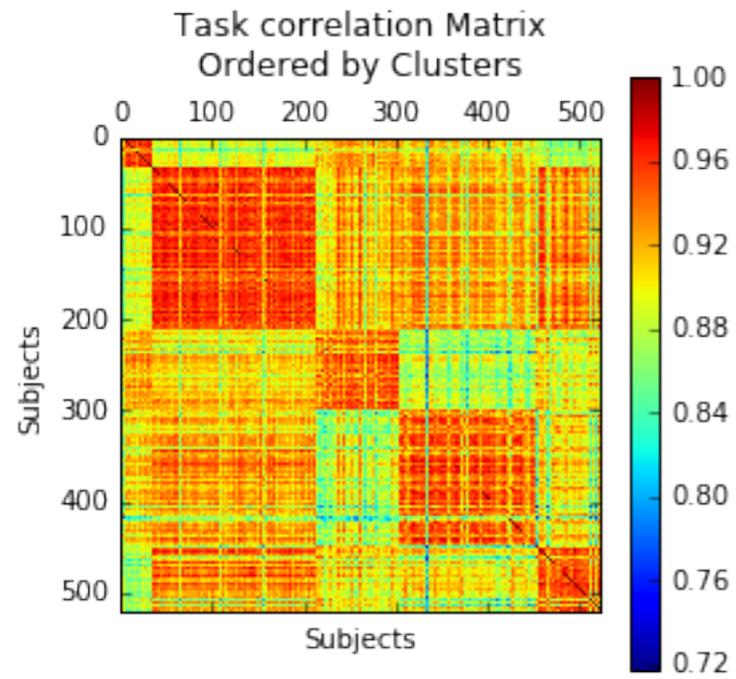
Correlation Analysis

The survey and task metrics aren't too related



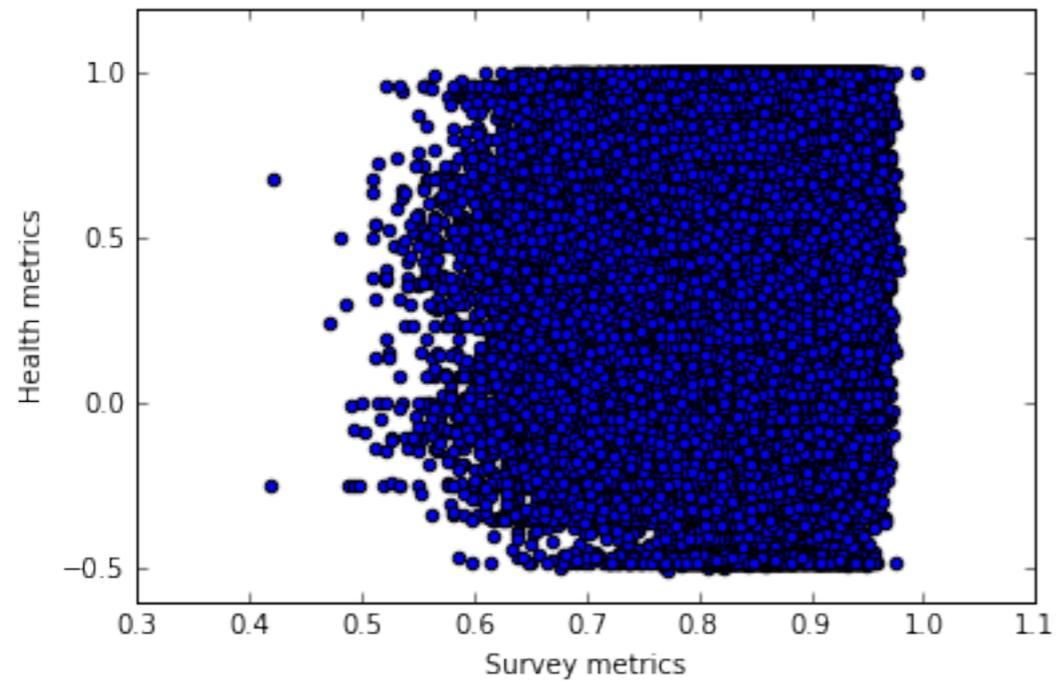
Each dot is a point
In the correlation matrix,
e.g. a **pair of subjects**

Correlation Analysis



Correlation Analysis

Survey metrics vs health metrics
rval = 0.078



Demographics metrics vs health metrics
rval = 0.008

