St Petersburg University
www.spbu.ru

Data Consolidation and Analysis System for Brain Research

Vladislav Volosnikov, Vladimir Korkhov, Andrey Vorontsov, Kirill Gribkov, Alexander Degtyarev, and Alexander Bogdanov
Saint Petersburg State University

Natalia Zalutskaya, Nikolay Neznanov, Natalia Ananyeva
V.M.Bekhterev National Research Medical Center for Psychiatry and Neurology

Volosnikov.apmath@gmail.com

GRID'18, Dubna, Russia, September 14, 2018

St Petersburg University
www.spbu.ru

There were around **47 million** people worldwide living with dementia in 2015 and this number will reach 132 million in 2050.

World Alzheimer Report 2016
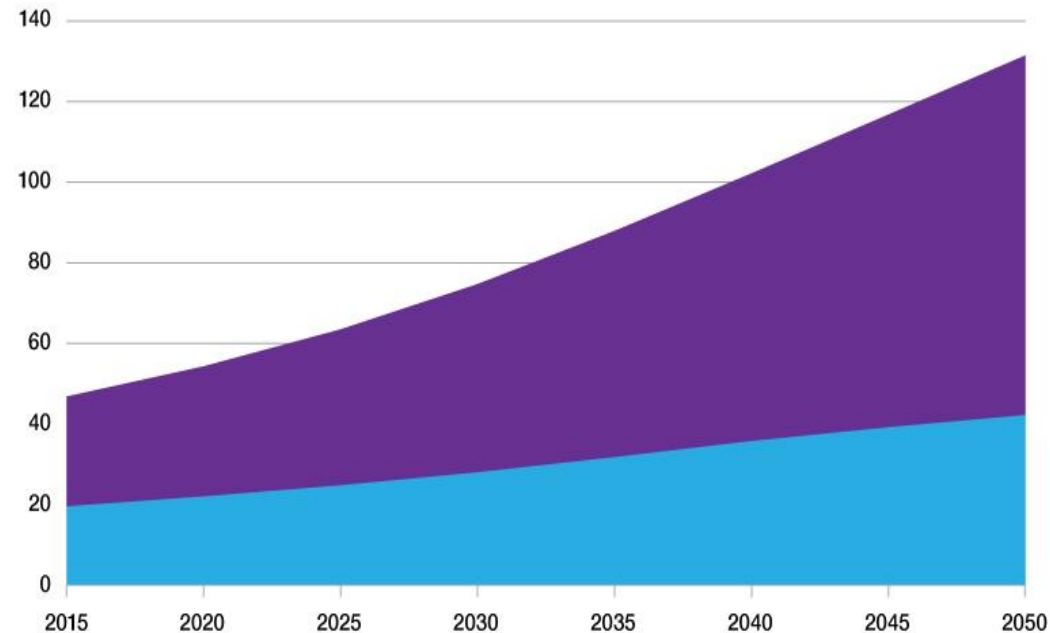
We need to detect and diagnose Alzheimer's earlier.
A more reliable, affordable, and accessible diagnostic — such as a blood test — would make it easier to see how Alzheimer's progresses and track how effective new drugs are.

Why I'm digging deep into Alzheimer's by Bill Gates

A Growing Health Crisis
The projected number of people with dementia, *millions*

■ High-income countries   ■ Low- & middle-income countries

Source: Alzheimer's Disease International – World Alzheimer Report 2015

It is chronic neurodegenerative disease leading to **irreversible** brain damage.

Nowadays medicine can **only slow** its progression. There is no consensus on the reasons for its occurrence.

Methods of neuroimaging, which allow to detect changes in the brain *in vivo* have great prospects for practical application.

Brain Atrophy in Advanced Alzheimer's Disease

Normal / AD

St Petersburg
University
www.spbu.ru

- Create **information system** for doctors to collect heterogeneous patient data.

- Perform **consolidation of heterogeneous patient data** to enable complex data analysis.

- Create **data analysis** system for clusterization and classification patient data to estimate Alzheimer's development probability.

- Provide a convenient **interface** for working with data and analysis results.

- Heterogeneous data from various data sources: **MRI**; **fMRI**; **EEG**; **clinical, biochemical**, **genetic analysis**; **psycho tests**…
- E.g., volumes of subcortical structures and thickness of the cortex for parts of both hemispheres from MRI.
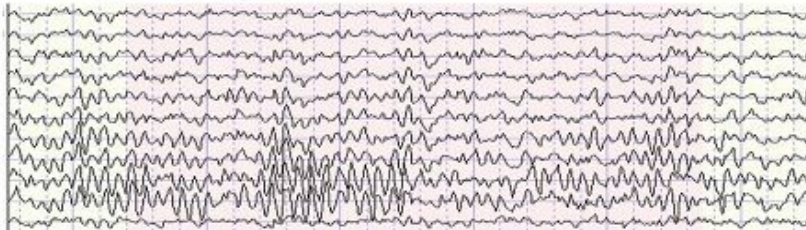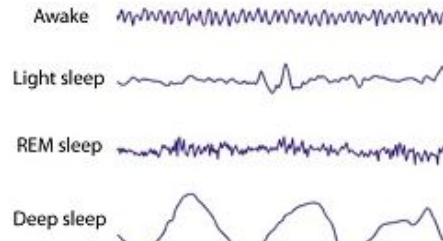
# ColHeaders	Index	SegId	NVoxels	Volume_mm3	StructName	normMean	normStdDev	normMin	normMax	normRange
1	4	11594	11594.0	Left-Lateral-Ventricle	36.6615	11.3303	7.0000	86.0000	79.0000
2	5	1091	1091.0	Left-Inf-Lat-Vent	45.6393	11.7438	16.0000	82.0000	66.0000
3	7	14363	14363.0	Left-Cerebellum-White-Matter	89.8912	6.3888	39.0000	111.0000	72.0000
4	8	42465	42465.0	Left-Cerebellum-Cortex	70.0583	10.1552	22.0000	115.0000	93.0000
5	10	5326	5326.0	Left-Thalamus-Proper	86.2262	10.3070	44.0000	108.0000	64.0000
6	11	3669	3669.0	Left-Caudate	80.2315	9.0591	43.0000	104.0000	61.0000
7	12	4567	4567.0	Left-Putamen	88.6310	6.5598	62.0000	108.0000	46.0000
8	13	1648	1648.0	Left-Pallidum	96.9097	5.6036	71.0000	113.0000	42.0000
9	14	2474	2474.0	3rd-Ventricle	33.3027	10.2338	13.0000	77.0000	64.0000
10	15	2468	2468.0	4th-Ventricle	50.8719	9.0996	26.0000	84.0000	58.0000
11	16	18653	18653.0	Brain-Stem	86.3133	10.9907	0.0000	117.0000	117.0000
12	17	1842	1842.0	Left-Hippocampus	72.0411	9.5104	29.0000	99.0000	70.0000
13	18	535	535.0	Left-Amygdala	71.9239	7.0689	52.0000	91.0000	39.0000
14	24	2124	2124.0	CSF	49.4938	11.9826	23.0000	89.0000	66.0000
15	26	530	530.0	Left-Accumbens-area	77.8528	7.2036	52.0000	93.0000	41.0000
16	28	3656	3656.0	Left-VentralDC	93.2075	9.7105	50.0000	118.0000	68.0000
17	30	12	12.0	Left-vessel	68.3333	4.0332	62.0000	72.0000	10.0000
18	31	2241	2241.0	Left-choroid-plexus	50.6993	11.0664	23.0000	97.0000	74.0000
19	43	9872	9872.0	Right-Lateral-Ventricle	37.9887	12.1891	0.0000	91.0000	91.0000
20	44	1065	1065.0	Right-Inf-Lat-Vent	47.4778	11.6087	20.0000	85.0000	65.0000
21	46	14474	14474.0	Right-Cerebellum-White-Matter	91.7947	5.9579	42.0000	114.0000	72.0000

Awake
Light sleep
REM sleep
Deep sleep
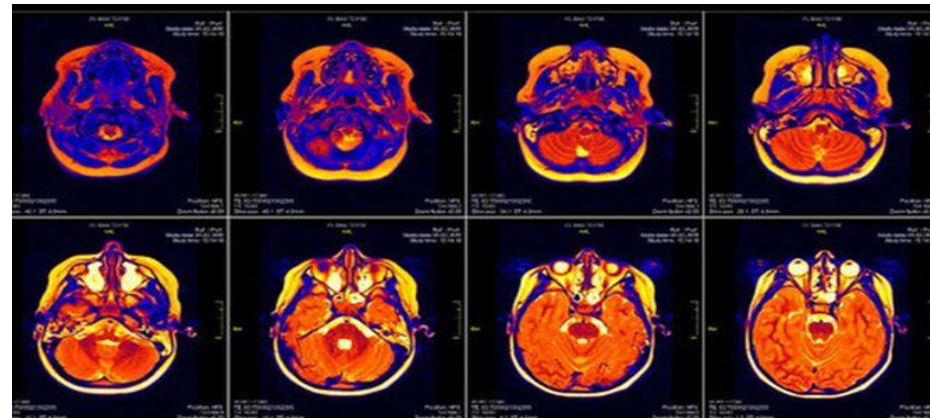
- **EEG (electroencephalography)** measures the electrical activity of our brain via electrodes that are placed on the scalp. It tells us, from the surface measurements, how active the brain is. This can be useful for quickly determining how brain activity can change in response to stimuli, and can also be useful for measuring abnormal activity, such as with epilepsy.

- **MRI (magnetic resonance imaging)** provides a map of the brain – how it looks at a set moment in time. This structural information can be useful for determining how the sizes of certain brain areas compare across people, or if there is something abnormal about a particular brain (a tumor for example).

- **Functional magnetic resonance imaging or functional MRI (fMRI)** measures brain activity by detecting changes associated with blood flow. This technique relies on the fact that cerebral blood flow and neuronal activation are coupled. When an area of the brain is in use, blood flow to that region also increases.

- **Distributed** data storage due to the **large** data size.

- Collection and integration of data from multiple sources.

- Various data formats: reason for **NoSQL** database.

Within the confines of cooperation, the system **unites** the resources of **Bekhterev' Institute** and **Saint Petersburg State University.**
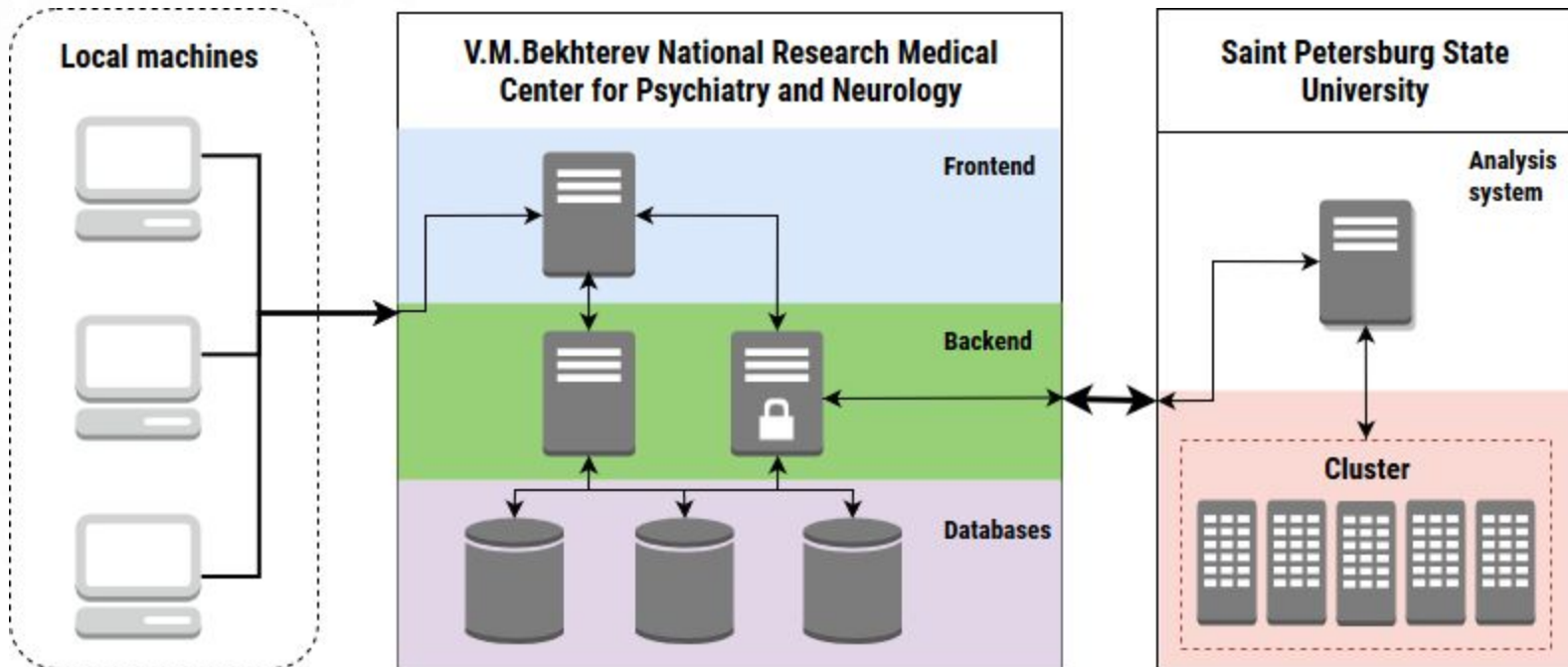
System scheme

It was decided to use **MEAN** stack.
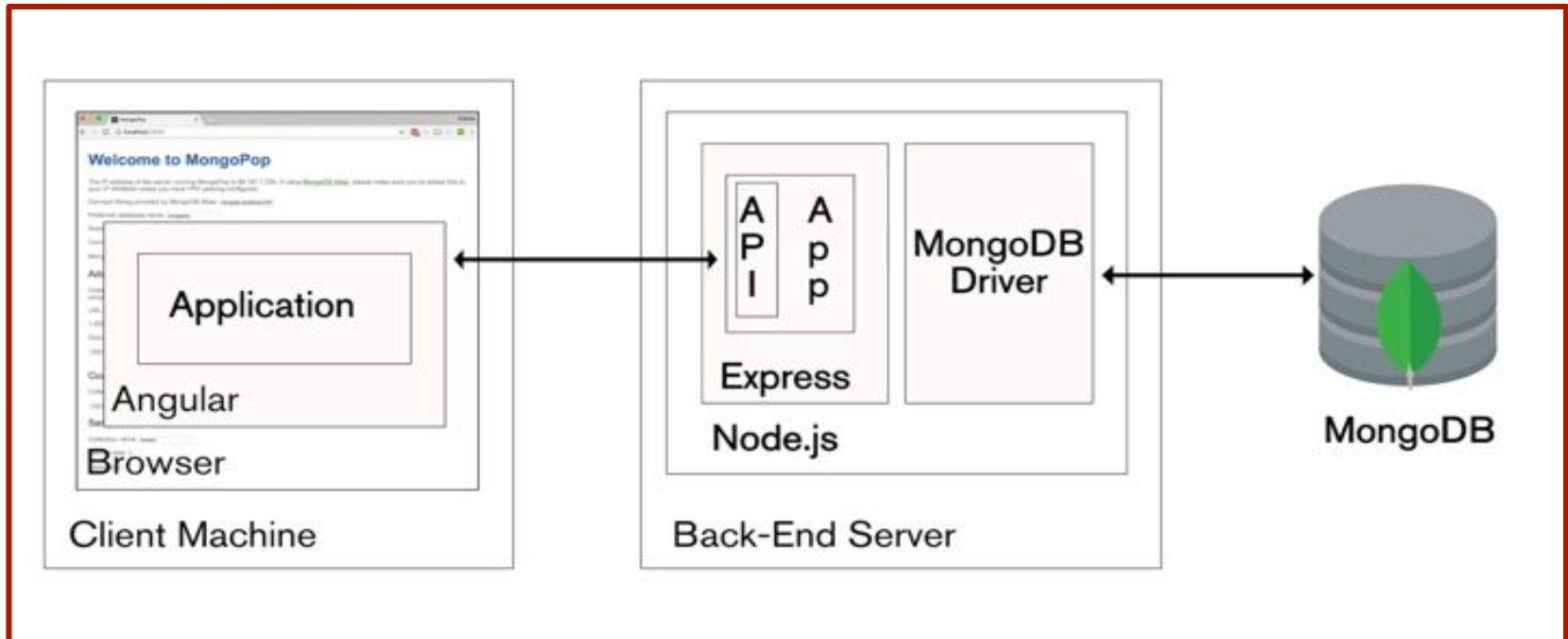
MEAN stack: MongoDB, Express, Angular, Node

St Petersburg
University
www.spbu.ru

- Flexible support for various data sources and types.

- Import from Excel, CSV tables.

- Role-based access.

- Patient data obfuscation.

- API for integrated data analysis.

- **Anonymous** MRI examinations and other tests.

- Group of **145** patients not less then 55 years of age.

- Control group of **healthy** volunteers and 3 groups of patients with various types of **dementia**.

- Perform data clustering separately by parts of the brain: temporal lobe, parietal lobe, insula, limbic system, entorhinal cortex, etc.

MRI data was post-processed by open source **Freesurfer** package.

Using anatomical atlases, the package does automatic segmentation and parcellation of brain.

As result we receive a **numerical** data from MRI.
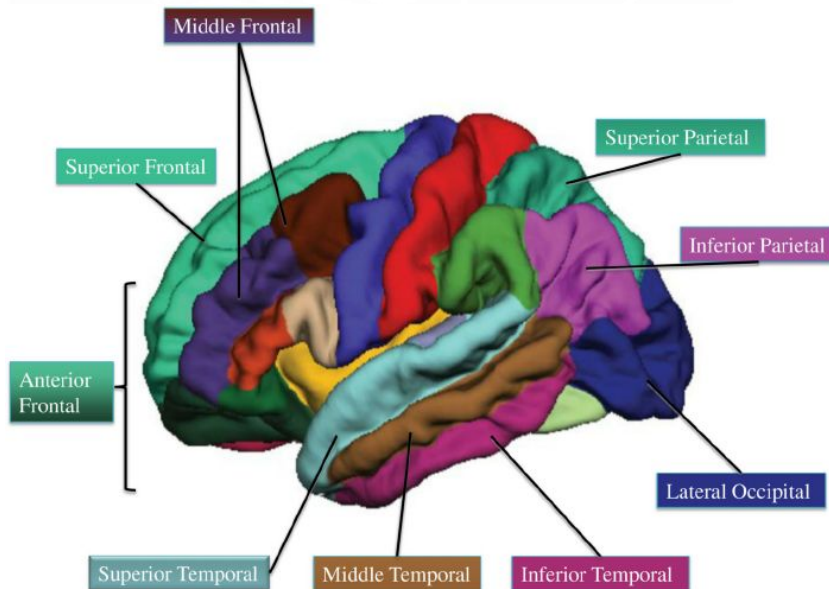
What FreeSurfer Does...

FreeSurfer creates computerized models of the brain from MRI data.
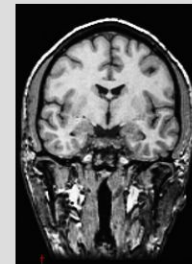
Input:
T1-weighted (MPRAGE)
$1mm^3$ resolution
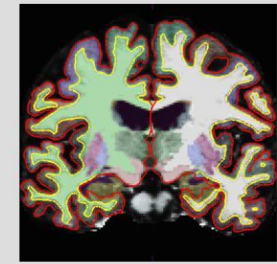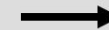
Output:
Segmented & parcellated conformed volume
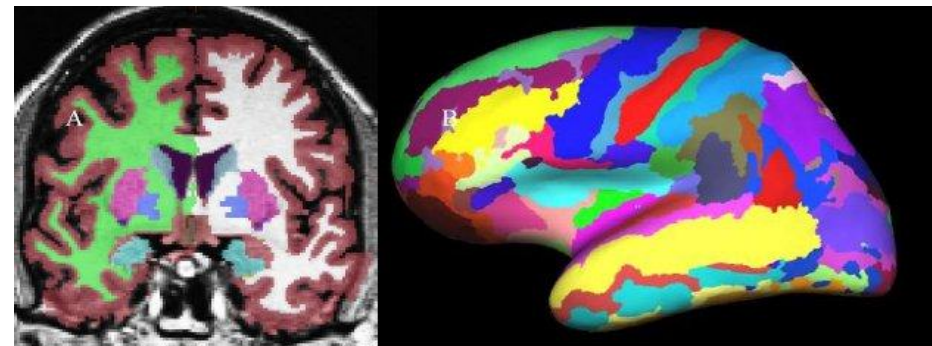
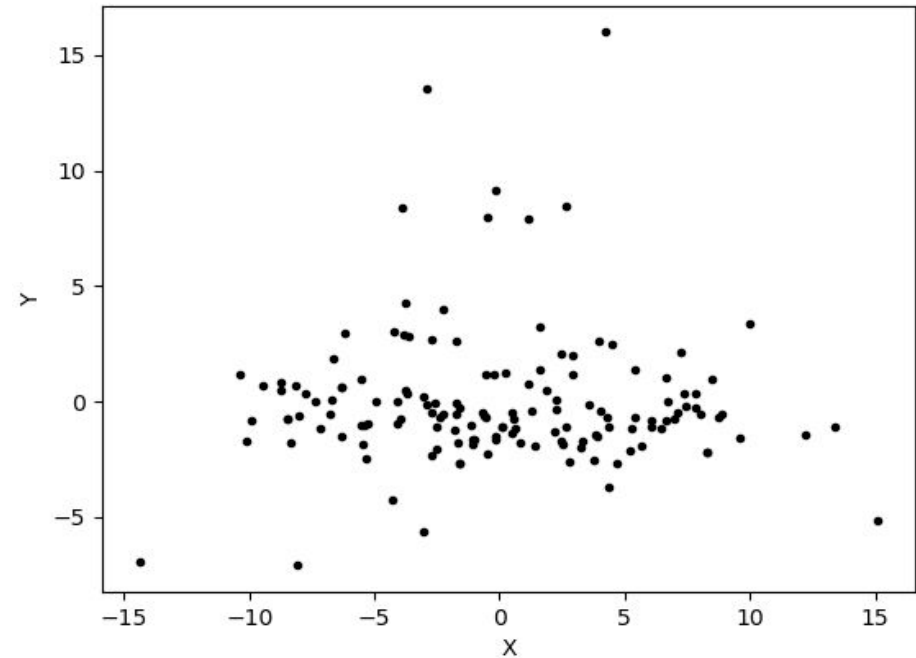Each patient was represented by a point in a feature space.

The total dimension is about **550**.
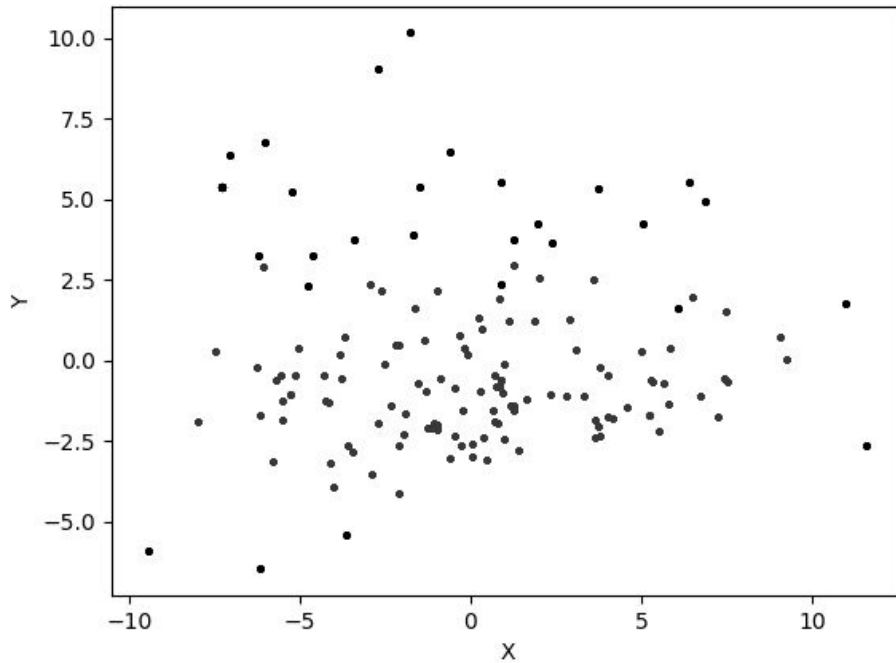
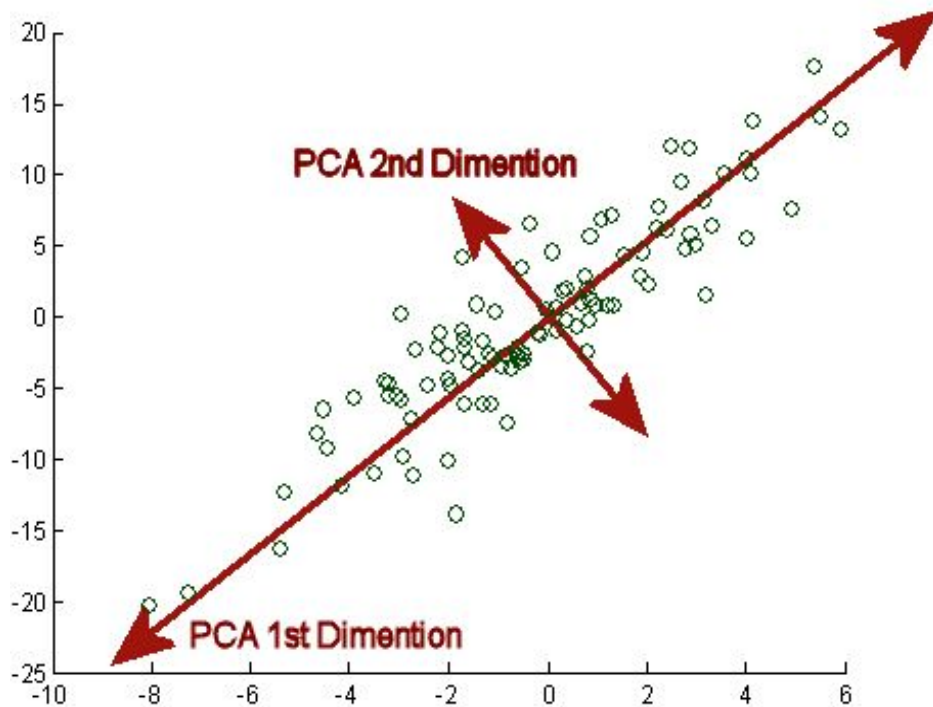

The distribution of parietal lobe characteristics

The distribution of frontal lobe characteristics

Due to "*Curse of dimensions*", dimension was reduced by **PCA method**.

In all cases, the final dimension was constrained: the residual variance of the data should be **at least 90%** of the original.
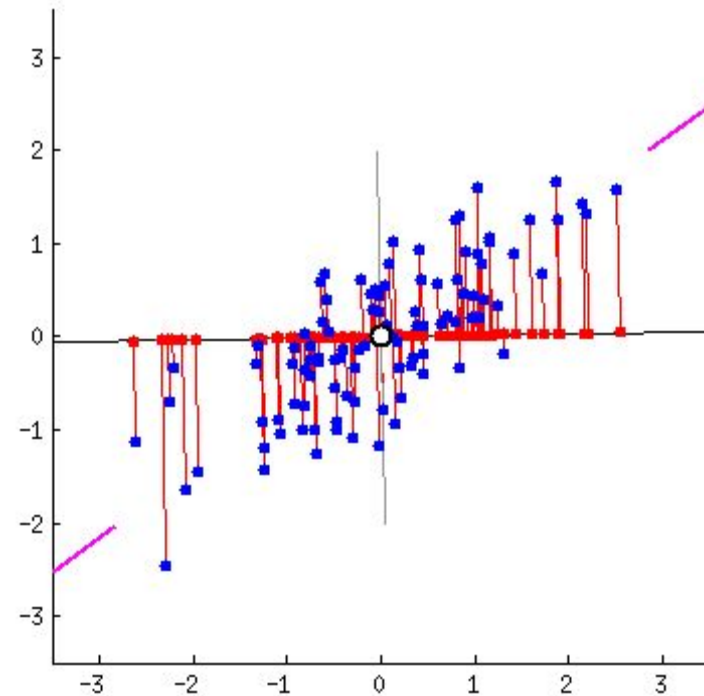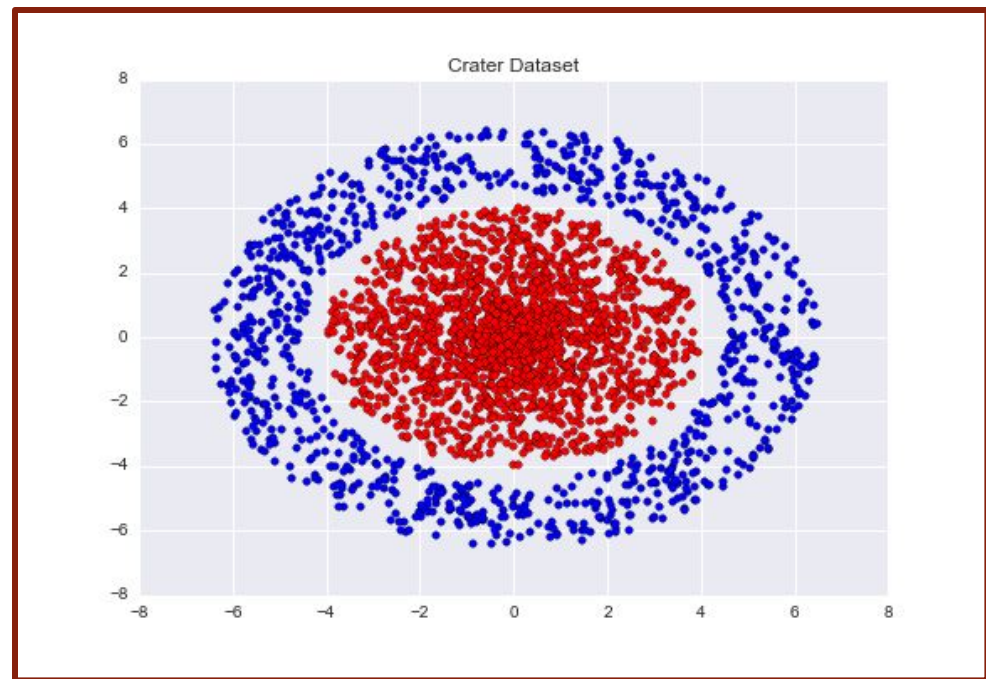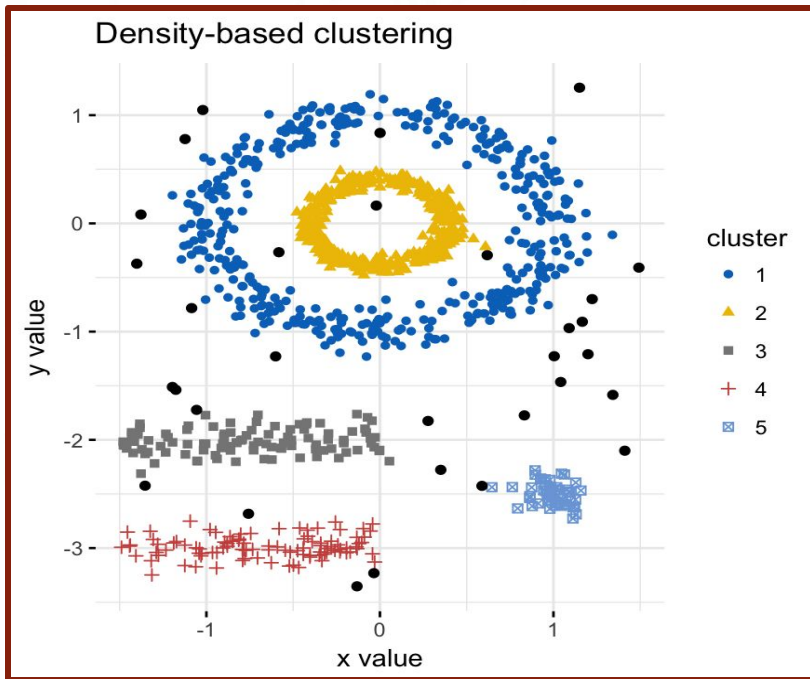
PCA method idea illustrations

Using **density-based** clustering algorithms.

Some advantages of this class are: ability to detect clusters of **arbitrary shape** and size, the resistance to noise, **no need to know the number of groups**.

Density-based clusterization illustrations

For clustering was used instance of "OPTICS" **(ordered points to identify the clustering structure)** algorithm from PyClustering package.

- Basic idea is similar to DBSCAN, but it addresses one of DBSCAN's major weaknesses: the problem of **detecting** meaningful **clusters** in data of **varying density**.

- In order to do so, the points of the database are (linearly) ordered such that points which are spatially closest become neighbors in the ordering.

- Additionally, a special distance is stored for each point that represents the density that needs to be accepted for a cluster in order to have both points belong to the same cluster.
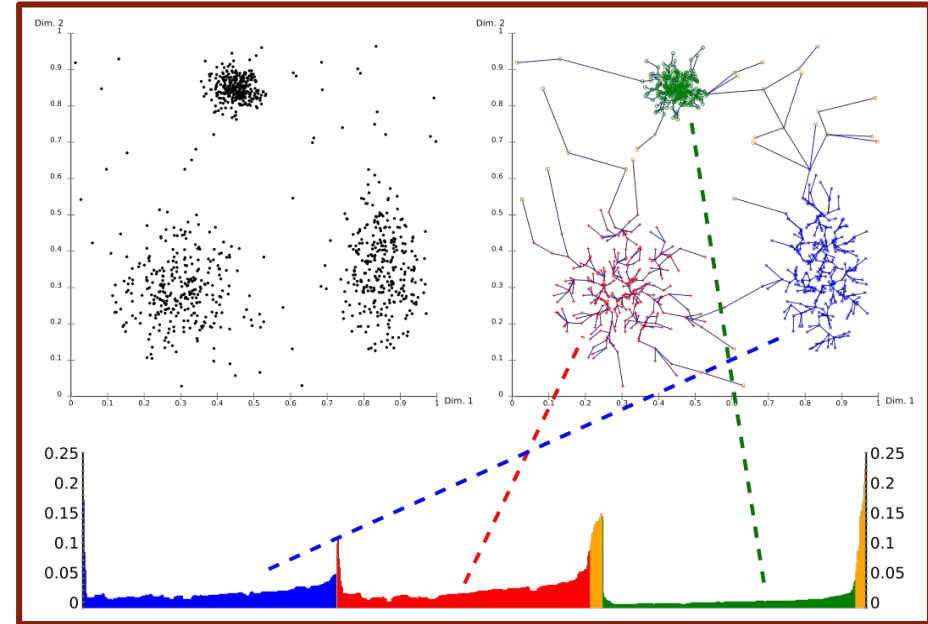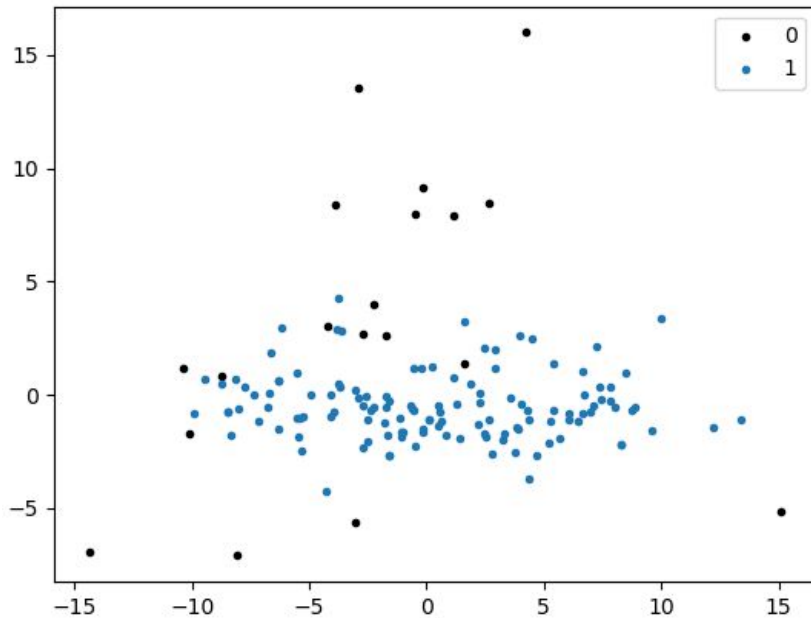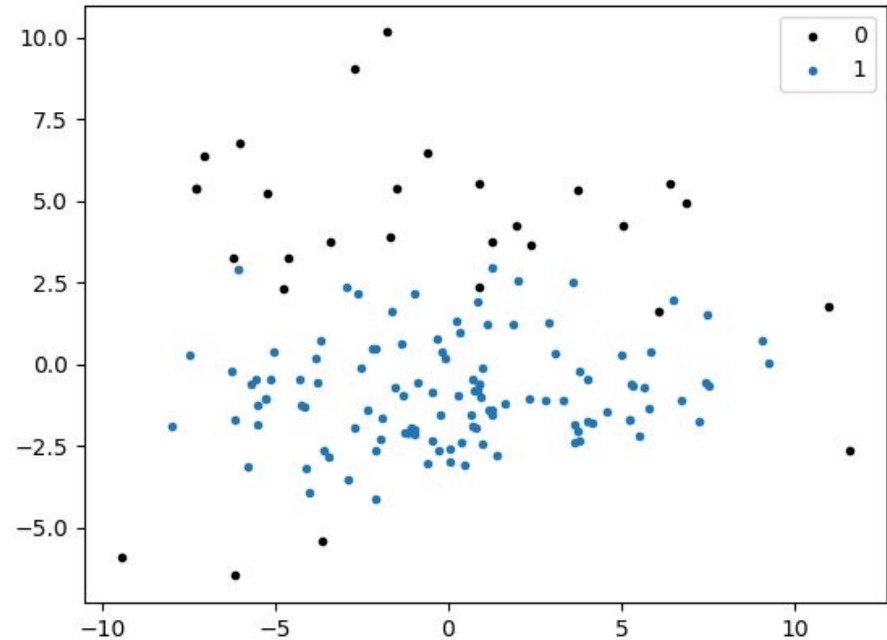

OPTICS idea illustration

The study of most brain parts showed a large cluster, surrounded by emissions.

Increasing number of patients can lead to the discovery of new groups.
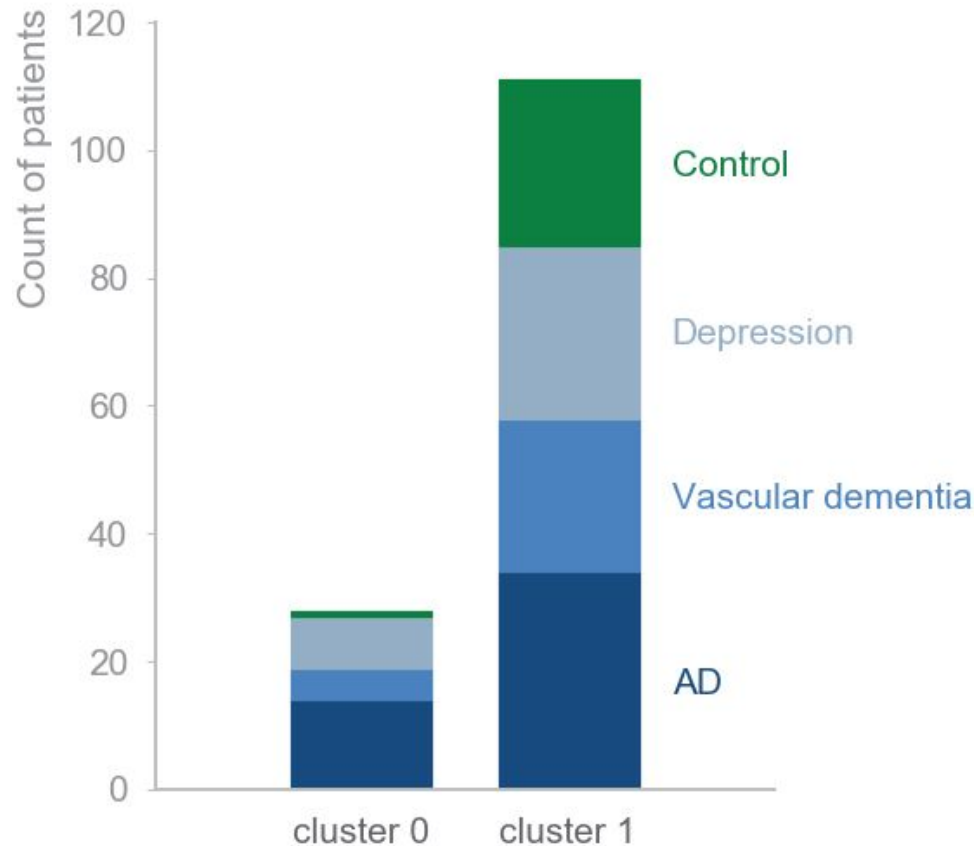

Clustering by occipital lobe characteristics

Clustering by frontal lobe characteristics

0 - anomaly
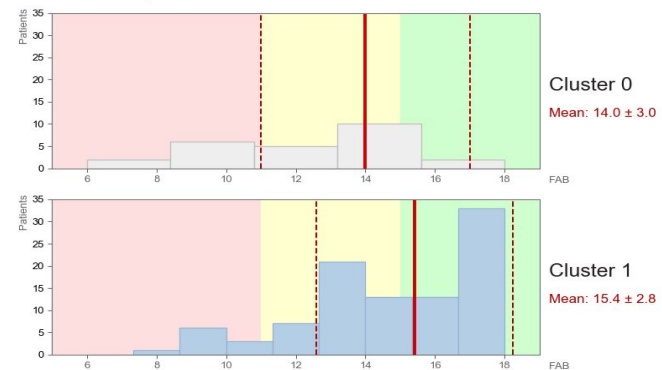## 1 - "normal"

Distribution by Temporal lobe

Temporal lobe

According to p-value =

0.016,

a statistically significant relationship of **medium** strength have been found.
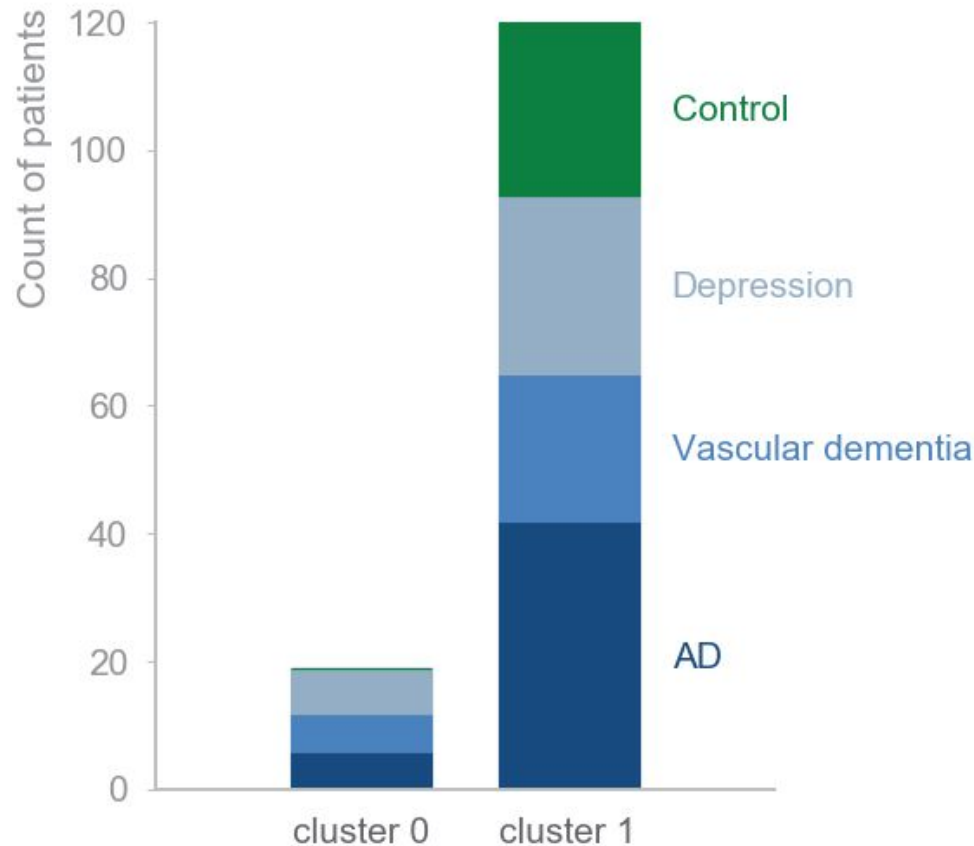
OR: 0.12, 95% CI: 0.02-0.93. Using Fisher exact test.

FAB, Temporal

Cluster 0
Mean: 14.0 ± 3.0

Cluster 1
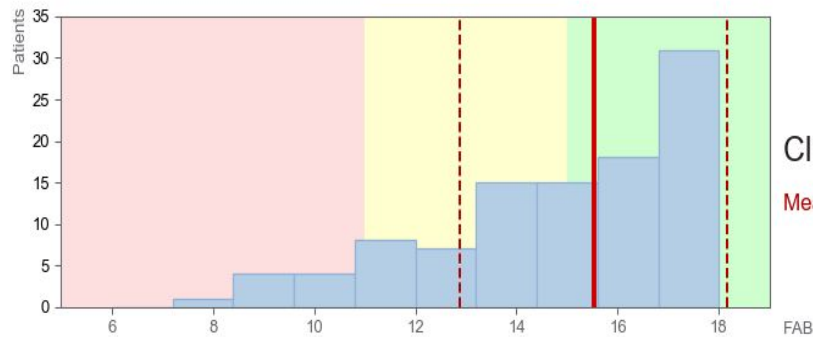Mean: 15.4 ± 2.8

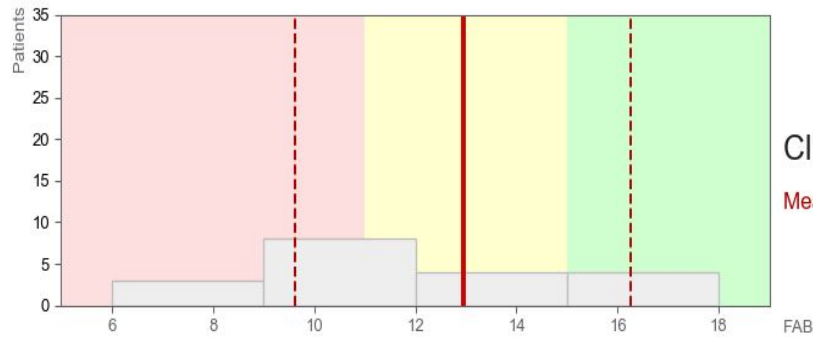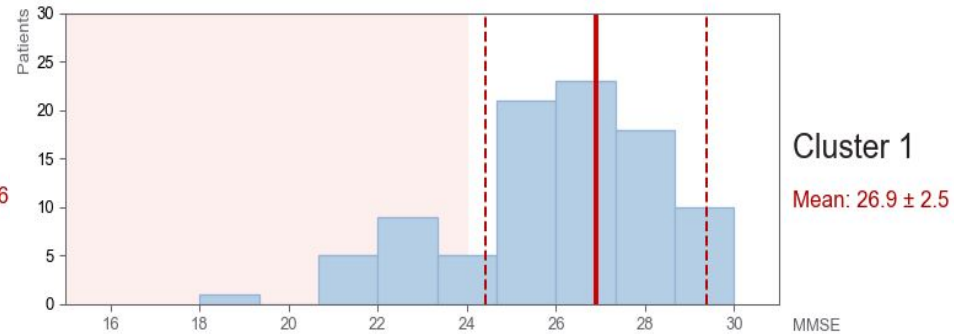*Result of cluster analysis, generated by system*
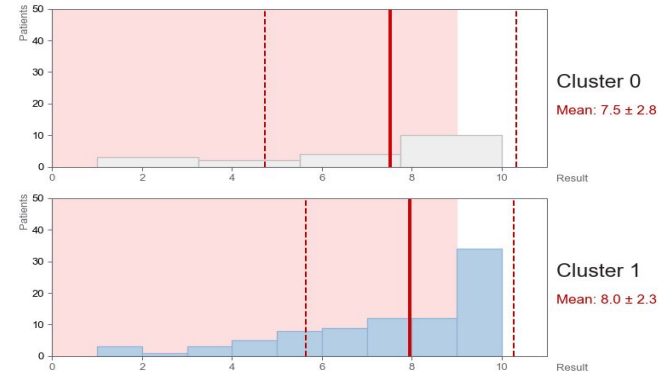
Result of cluster analysis, generated by system

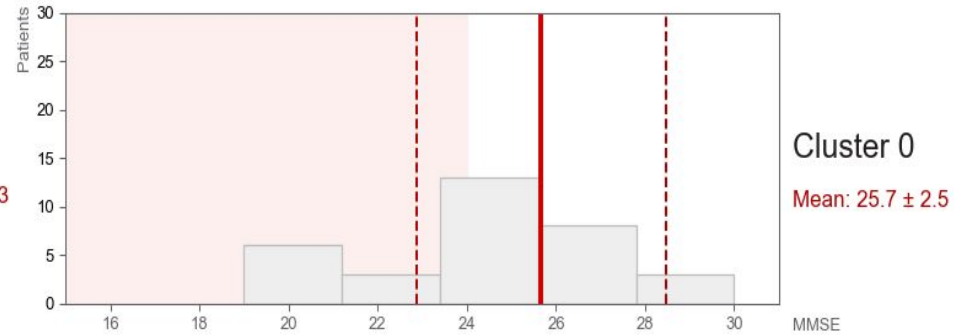System produces various histograms and graphics for **clear** representation of analysis' results.

FAB, Frontal

Cluster 0 — Mean: 12.9 ± 3.3
Cluster 1 — Mean: 15.5 ± 2.6

MMSE, Parietal

Cluster 0 — Mean: 25.7 ± 2.5
Cluster 1 — Mean: 26.9 ± 2.5

We created a prototype of information system for patient data consolidation and analysis.

Flexible API to load, store and access patient data of any type.

Results of data analysis with OPTICS algorithm on patient grouping **correspond** to expectations.

Specialists of St. Petersburg V.M.Bekhterev National Research Medical Center for Psychiatry and Neurology **confirmed** the importance and validity of the research and its results.

Thank you for attention! Questions?

St Petersburg University
spbu.ru