# NeuroStream User Guide

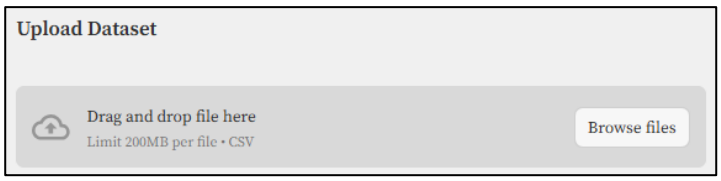
## Overview – What is NeuroStream?

NeuroStream is an interactive visualization tool designed to support exploratory data analysis (EDA) of regional brain features derived from neuroimaging pipelines (e.g., MRI-based processing).

The tool enables users to explore, without writing code**:** Distributional differences across cohorts, Changes in data characteristics under different preprocessing methods, Relationships between brain measures and clinical variables (e.g., age, years of education).

## 1. Data Upload

This screen allows users to upload the input dataset for analysis.




| cohort | age_group | gender | edu_level | htn | dm | icv_300 | whole_brai | ctx_cerebr | cerebral_v |
|---|---|---|---|---|---|---|---|---|---|
| Cohort1 | 3 | 2 | 0 | 1 | 2 | 1528.97 | 1087.181 | 475.854 | 456.194 |
| Cohort1 | 3 | 1 | 2 | 1 | 1 | 1672.834 | 1155.588 | 510.907 | 468.693 |
| Cohort1 | 1 | 2 | 1 | 2 | 2 | 1270.509 | 916.385 | 417.169 | 386.732 |
| Cohort1 | 3 | 2 | 1 | 2 | 2 | 1606.304 | 969.974 | 438.748 | 412.385 |
| Cohort1 | 3 | 2 | 1 | 1 | 2 | 1489.17 | 957.11 | 429.918 | 418.315 |
| Cohort1 | 4 | 2 | 1 | 2 | 2 | 1247.569 | 857.186 | 387.895 | 359.236 |
| Cohort1 | 3 | 1 | 6 | 1 | 2 | 1529.645 | 1083.463 | 466.072 | 464.204 |
| Cohort1 | 3 | 2 | 1 | 2 | 2 | 1518.339 | 956.883 | 411.802 | 448.83 |
| Cohort1 | 3 | 1 | 1 | 1 | 2 | 1480.643 | 1072.522 | 428.421 | 436.001 |
| Cohort1 | 2 | 2 | 3 | 2 | 2 | 1422.969 | 999.328 | 456.327 | 425.394 |
| Cohort1 | 3 | 2 | 3 | 1 | 1 | 1447.983 | 1042.655 | 487.187 | 476.371 |

Supported format: CSV file

Data structure: One row per subject

Required contents:

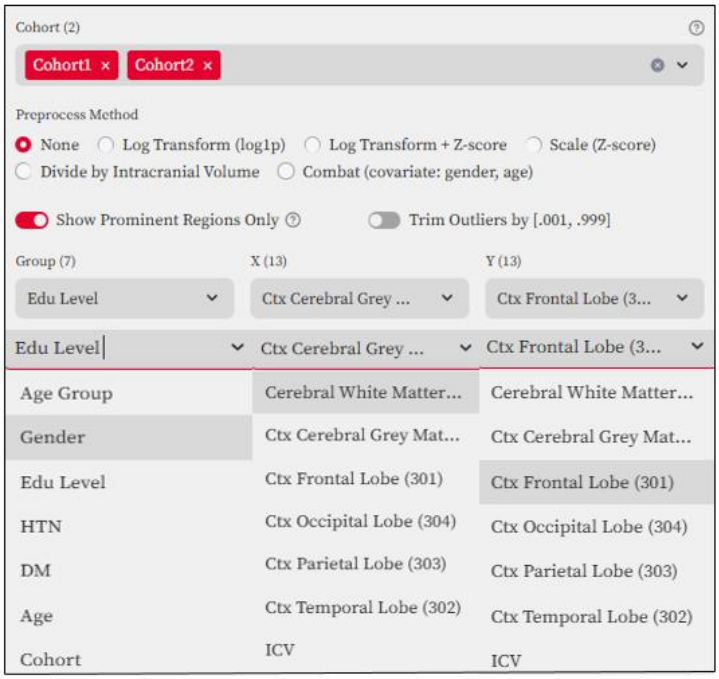
· Cohort information

· Clinical variables (e.g., age, education level)

· Regional brain volumes

The uploaded dataset serves as the foundation for all subsequent visualizations and statistical summaries within NeuroStream.

## 2. Cohort and Preprocessing Selection




At this stage, users select the cohorts and preprocessing methods to include in the analysis.

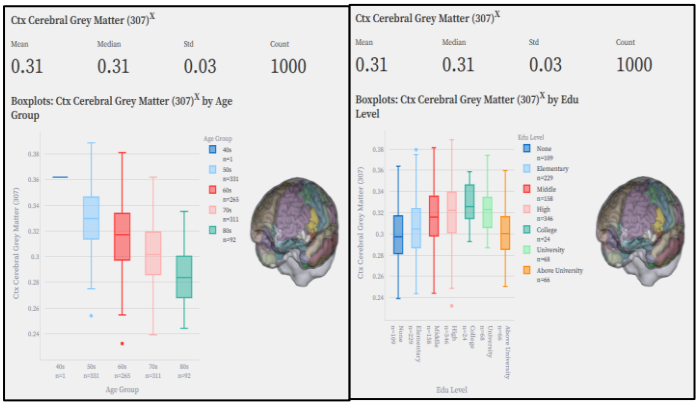
· Multiple cohorts can be selected for group-wise comparison

· Different preprocessing methods can be applied to adjust data scale and distributional properties

Changes in preprocessing are immediately reflected in downstream visualizations, including boxplots, scatter plots, and PCA results.

## 3. Exploratory Data Analysis (EDA)

### 3.1. Boxplot, descriptive statistics




This view presents descriptive statistics and boxplot visualizations for the selected variables.

Reported statistics:

· Mean, Median, Standard deviation, Sample size

Boxplots summarize distributional differences across categorical variables

(e.g., age groups or education groups) and are useful for comparing central tendency and variability between cohorts.

### 3.2. Groupwise T-test

**Groupwise T-test: Ctx Cerebral Grey Matter (307)X by Age Group**

* p<0.05, ** <0.01, *** <0.001, or Not Significant

| | 40s | 50s | 60s | 70s | 80s |
|---|---|---|---|---|---|
| 40s | - | NS | NS | * | *** |
| 50s | 1.403 | - | *** | *** | *** |
| 60s | 1.872 | 7.028 | - | *** | *** |
| 70s | 2.53 | 15.055 | 6.96 | - | *** |
| 80s | 3.481 | 16.519 | 10.577 | 5.989 | - |

**Groupwise T-test: Ctx Cerebral Grey Matter (307)X by Edu Level**

* p<0.05, ** <0.01, *** <0.001, or Not Significant

| | None | Elementary | Middle | High | College | University | Above University |
|---|---|---|---|---|---|---|---|
| None | - | NS | *** | *** | *** | *** | *** |
| Elementary | -1.913 | - | *** | *** | *** | *** | *** |
| Middle | -6.382 | -5.417 | - | ** | *** | *** | NS |
| High | -9.538 | -9.859 | -3.135 | - | * | *** | NS |
| College | -7.391 | -6.354 | -3.609 | -2.107 | - | NS | ** |
| University | -11.351 | -10.871 | -6.447 | -4.509 | -0.668 | - | *** |
| Above University | -5.574 | -4.538 | -0.542 | 1.634 | 2.846 | 4.738 | - |

This section provides results from pairwise group-wise t-tests between selected groups.

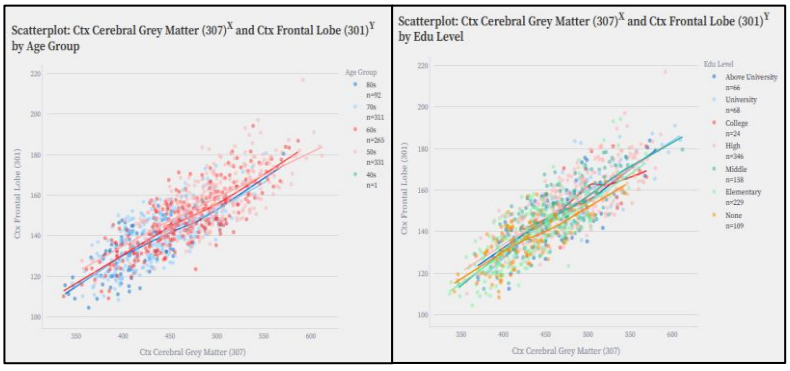
· Pairwise comparison of mean differences between age or education categories

· Comparisons can be performed across cohorts or preprocessing conditions

This feature is intended for rapid exploratory assessment of potential group differences during early-stage analysis.

Note: NeuroStream is intended for exploratory data analysis and visualization. Statistical results are provided for descriptive purposes only and should not be interpreted as confirmatory inference.

### 3.3. Scatter plot



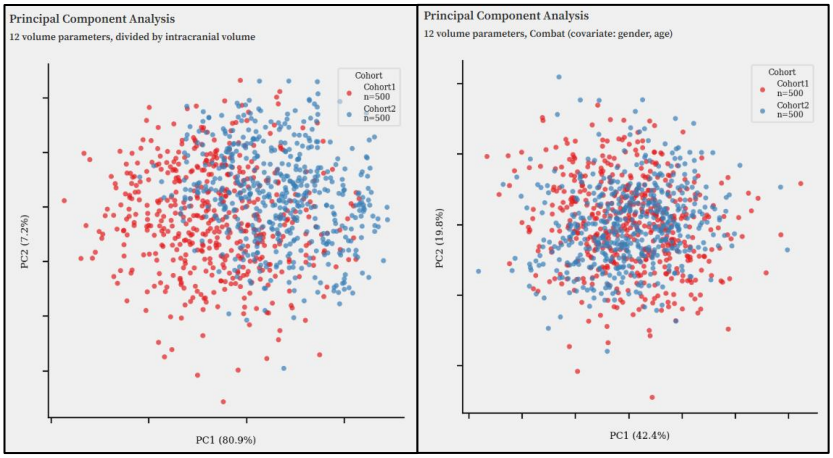

The scatter plot visualizes relationships between two continuous variables.

· Example: age vs. regional brain volume

· Points can be color-coded by cohort or group

This view enables intuitive inspection of trends such as age-related changes and cohort-specific patterns.

## 4. PCA (Principal Component Analysis)



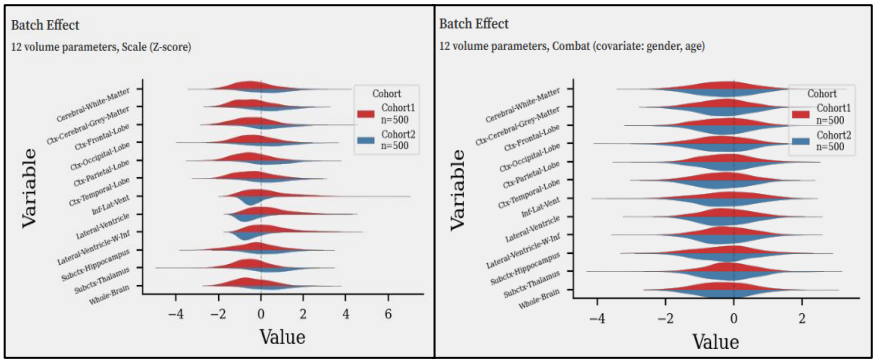

PCA projects high-dimensional regional brain features into a lower-dimensional space for visualization.

· Overview of the global data structure

· Assessment of cohort-level separation or overlap

· Comparison of data structure before and after ComBat harmonization to visually assess reduction of batch effects.

### 4.1. Violin plot




Violin plots visualize the full distribution of regional brain measures using kernel density estimation.

· Provide more detailed distributional information than boxplots

· Useful for inspecting asymmetry and distributional spread

By comparing violin plots before and after ComBat harmonization, users can intuitively assess how harmonization affects the overall distribution of brain features across cohorts.