

ClusterMouseAutism

Introduction

This project is an interactive web application coded in [R](#) for clustering mouse models of autism using neuroanatomical data. It makes use of the [Shiny](#) web application framework (a package in R). The app presents data from a research paper published in Nature Neuroscience (reference) entitled (reference).

Workflow

The app can be accessed in many ways. Options include:

1. Launching the app locally, assuming you have R installed
 - download the app folder from GitHub
 - launch R in the terminal or launch RStudio
 - from within R, load the Shiny package and set your working directory to the folder *containing* the app folder you just downloaded:
 - `setwd(insert folder containing app folder you just downloaded)`
 - launch the app with:
 - `runApp(appDir = 'ClusterMouseAutism', launch.browser = TRUE, display.mode = 'normal')`, which will launch the app in your default browser
2. Launch the app directly from GitHub
 - use this command:
 - `runGitHub(repo='ClusterMouseAutism', username='jeffbruce')`
 - this option distorts the style of the app for whatever reason
3. Access the app through the web
 - the app is currently hosted on a private Shiny server which resides on a Linux virtual machine
 - the app can be accessed by visiting the following site:
 -

Dependencies

- R packages:
 - shiny
 - gplots

Development Notes and Rationales

There's probably no point in giving the option to cluster by just 1 dimension, despite the option of doing so in the heatmap call, because the dimensions are clustered independently. The dendrograms don't make the plot significantly more difficult to read or anything.

Inspirations

- Devium Web (reference)
- Radiant (reference)

TO DO

Fix Bugs

- make the row/column dendrogram labels actually line up with the rows/columns
 - WAIT
 - maybe this bug doesn't exist on all browsers / displays
- verify that effect size calculations / data are correct
 - ask Jacob for original alleffects code
- heatmap doesn't immediately show up when you launch the app
 - WAIT
 - only resizing the window or clicking Recalculate makes the heatmap show – the heatmap on the second page always shows
 - maybe wait on this bug until I've launched it to see if it exists in the launch version

Lower Priority Issues

- ensure that the dendrogram lengths make sense
- improve y-axis scaling for bar plot
- “Warning: Stacking not well defined when ymin != 0”
- “Error in mousedata[isolate(input*strains*), inputselectBoxStrainRegion]”
- “ymax not defined: adjusting position using y instead”
 - only occurs for boxplots – tried setting the ymax but it didn't work
 - the number of errors output multiplies by the number of subplots

Add Hosting

- bug Fernando about configuring VNC to allow me to launch graphical applications from the vm
- figure out how to host the app on a Linux server here
- 3 (4?) options:
 - can host on Github so R users can run the app directly from the R command line
 - can host on Shiny's public servers (shinyapps.io)
 - * only professional version (\$3000 a year) offers custom domains, authentication of users, and free version only allows 10 applications, 50 active hours per month
 - can use a private ShinyServer hosted on a Linux server (there is a free version and a professional version)
 - * main differences between free and (\$10000) paid version is that paid version allows multiple R processes per app, provides admin dashboard with performance data, and supports SSL/authentication
 - apparently the Shiny library comes with a web server, and is designed to only host one app at a time – maybe if we're only hosting one app then this would be easiest?
 - * can this be done on a virtual machine? how to change the domain name?

Add New Features

- upgraded plots
 - upgraded violin plot to also include IQR and median which is standard in many violin plots?
 - bootstrapped violin plot (in general, page with bootstrapped means / effect sizes?)
 - change dot plot / bar plot to have standard error instead of standard deviation?
 - I need to understand bootstrapping better (more specifically, the conditions under which it fails)
- speed up bootstrapping somehow?
- show statistical tests?
- demonstrate bootstrapping on a separate Shiny page
- plot dashed line on effect size plot showing associated .05 p value?
- allow clustering by different methods by changing hclustfun? (e.g. Ward's, single-linkage, complete-linkage, top-down K-means?)
 - was this investigated in the paper to see if you could cluster the strains/regions into more meaningful groups with different clustering methods?
- allow other users to view/download the summary data?
- sort by a specific column or row and list correlations on side instead of dendrograms?
 - might be possible by supplying an argument to reorderfun
 - could be more interpretable by others
 - use pvclust to plot the actual appropriateness of the clusters
 - this doesn't seem very useful because you can infer the most similar strain directly from the length of the dendrogram branch
- provide user with a control to select groups of regions according to common attributes (gene models that affect the synapse, white matter / social perception & autonomic regulation, etc.)
- replace collaborator pdf summaries of scanned brains with generic Shiny app
- relate Shiny app to genetic information of the mouse models?
- add confidence of analysis as you change the number of strains / regions to cluster on?
- instead of weighting each brain region equally, allow user to add in their own weighting system, for example, where each region gets a weight corresponding to its size, or proportion of different cell types, etc.
 - but, total brain size isn't predictive on its own of autism so size isn't necessarily an interesting thing to cluster on
- group mouse models by differences in connectivity?
 - any papers that do this or groups working on this?
 - are we working on this with DTI data?
- any methods for combining multimodal data (neuroanatomical, genetic, behavioural) to subtype these models?
 - use something like PCA to predetermine how many clusters there might be?
- upgrade to heatmap.3 over heatmap.2?
 - what are the differences between these functions?
- in general, would be useful if more stats and code were on the micewiki
 - seems like people might be wasting time trying to find code templates in their own libraries, leading to sub-optimal ways of sharing code
 - i wonder if you can associate github repos to higher level repos for given users, so a lab can then search a collection of repos for the specific plot or function that they need

- given a dendrogram, can one break it out into an optimal number of clusters?
 - the length of a branch represents how similar the two rows/columns are
- best papers on the reliability of methods at MICE (there's that one MICE paper claiming that you need x number of subjects to maintain a given confidence level, which I need to read)

Theoretical Quandaries Effect size calculation

- why always normalize by the wildtype group and not vice versa? basically the variance of one of the groups is ignored, which seems inappropriate to me.

Bootstrapping in the paper

- used to determine the consistency of the clustering in the paper
- bootstrapped and calculated how often regions / groups were connected with each other
 - what does 'connected' mean here?
- what does this sentence mean: "Different group assumptions (+/- 1) and connection thresholds (+/-5%) were tested and the results were consistent"
- pvclust assesses clustering uncertainty via multiscale bootstrap resampling
- the hierarchical clustering method used was 'complete', and the distance metric was 'correlation' which used the Pearson method
- an $\alpha < 0.10$ was used to determine the clusters

Verification of clusters section

- not exactly sure what's going on this paragraph