SPACE: A Visualization Tool for Principal Components Analysis Data

Table of Contents

[Installation 2](#_Toc516574593)

[Local Installation Steps 2](#_Toc516574594)

[Web Installation Steps 2](#_Toc516574595)

[Getting Started 3](#_Toc516574596)

[Select a file to load 3](#_Toc516574597)

[Choose which PCs to plot 4](#_Toc516574598)

[Choose a label to color the plot 5](#_Toc516574599)

[Select points to inspect the data 6](#_Toc516574600)

[Double click to zoom 6](#_Toc516574601)

[Hide and show subsets by label 8](#_Toc516574602)

[Tune the plot’s look 9](#_Toc516574603)

[Advanced 11](#_Toc516574604)

[Palette options 11](#_Toc516574605)

[Toggle 12](#_Toc516574606)

[Thresholding 12](#_Toc516574607)

[Source 13](#_Toc516574608)

[Got an idea for a new feature? 13](#_Toc516574609)

# Installation

SPACE can be launched from an R session. Alternately it can run on server running Shiny server. In either case it opens in a web browser such as Chrome, Firefox or Internet Explorer.

## Local Installation Steps

1. Download the repo from github: <https://github.com/Ancestry/SPACE>
2. Put server.R and ui.R into a directory together, e.g. C:\Users\user\_name\R\_projects\SPACE-master
3. Install R (if it’s not already installed)
4. Open an R session
5. Install and load Shiny  
   > install.packages("shiny") # first time only  
   > library(shiny)
6. Navigate to the parent directory, e.g.   
   > setwd("C:\Users\user\_name\R\_projects\")  
   > runApp("SPACE-master")
7. SPACE will open in your default web browser.
8. Optionally, load the example file from the repo

## Web Installation Steps

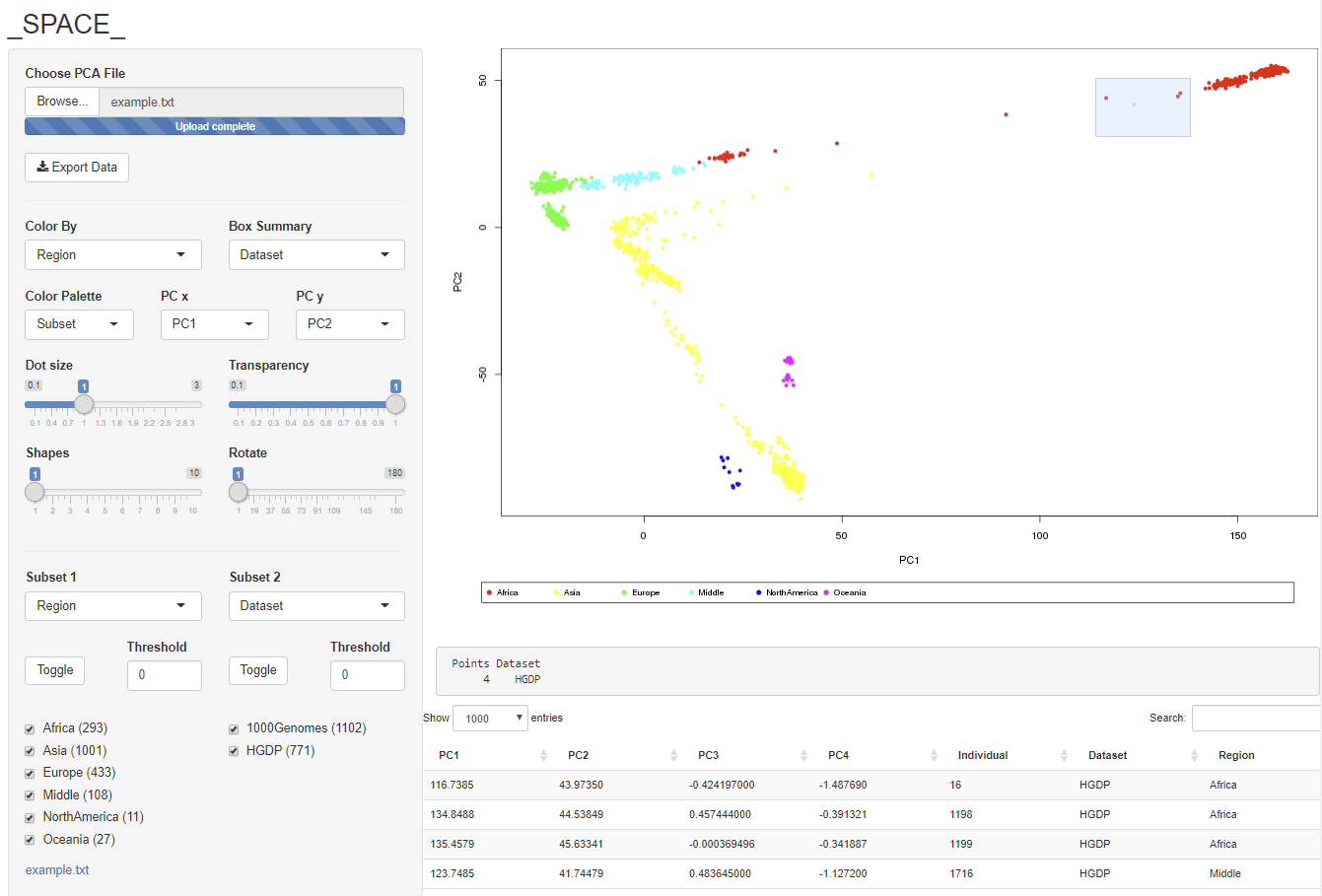
1. Download the repo from github: <https://github.com/Ancestry/SPACE>
2. Install R (if it's not already installed)   
   > sudo apt-get install r-base
3. Install shiny  
   sudo su - -c "R -e \"install.packages('shiny')\"“
4. Install gdebi-core and shiny-server  
   > sudo apt-get install gdebi-core  
   > sudo gdebi shiny-server-1.5.4.deb
5. Start the shiny server  
   > sudo systemctl start shiny-server
6. create a new directory in /srv/shiny-server e.g.  
   > mkdir /srv/shiny-server/SPACE
7. put ui.R and server.R in the new directory
8. Navigate to your server in a browser. By default, shiny server runs on port 3838 e.g.   
   <http://127.0.0.1:3838/SPACE/>

# Getting Started

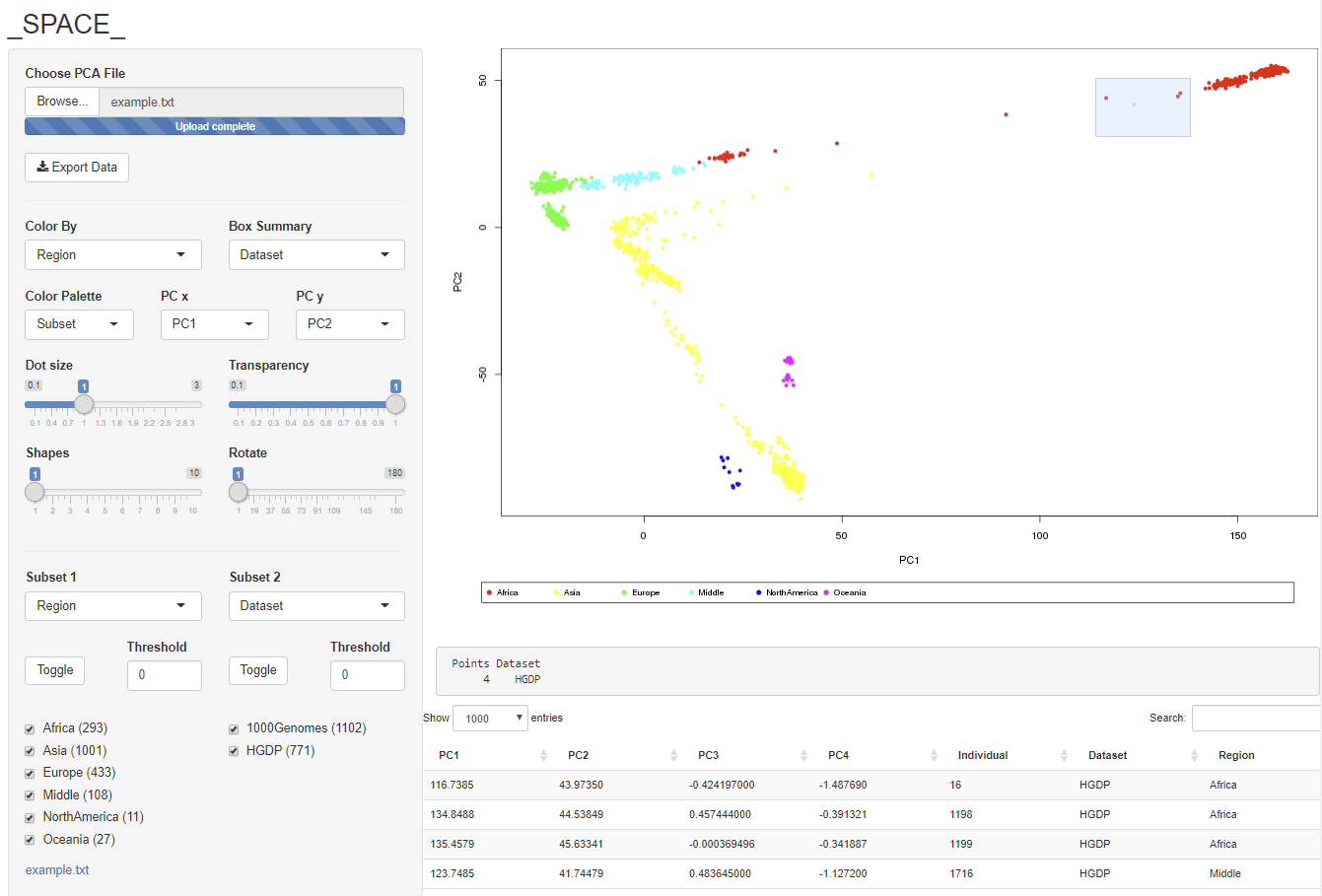
## Select a file to load

SPACE takes tab delimited text files. Any column that starts with “PC” is considered a data column. Other columns are considered labels. Columns can be in any order. An example input file might look like this:

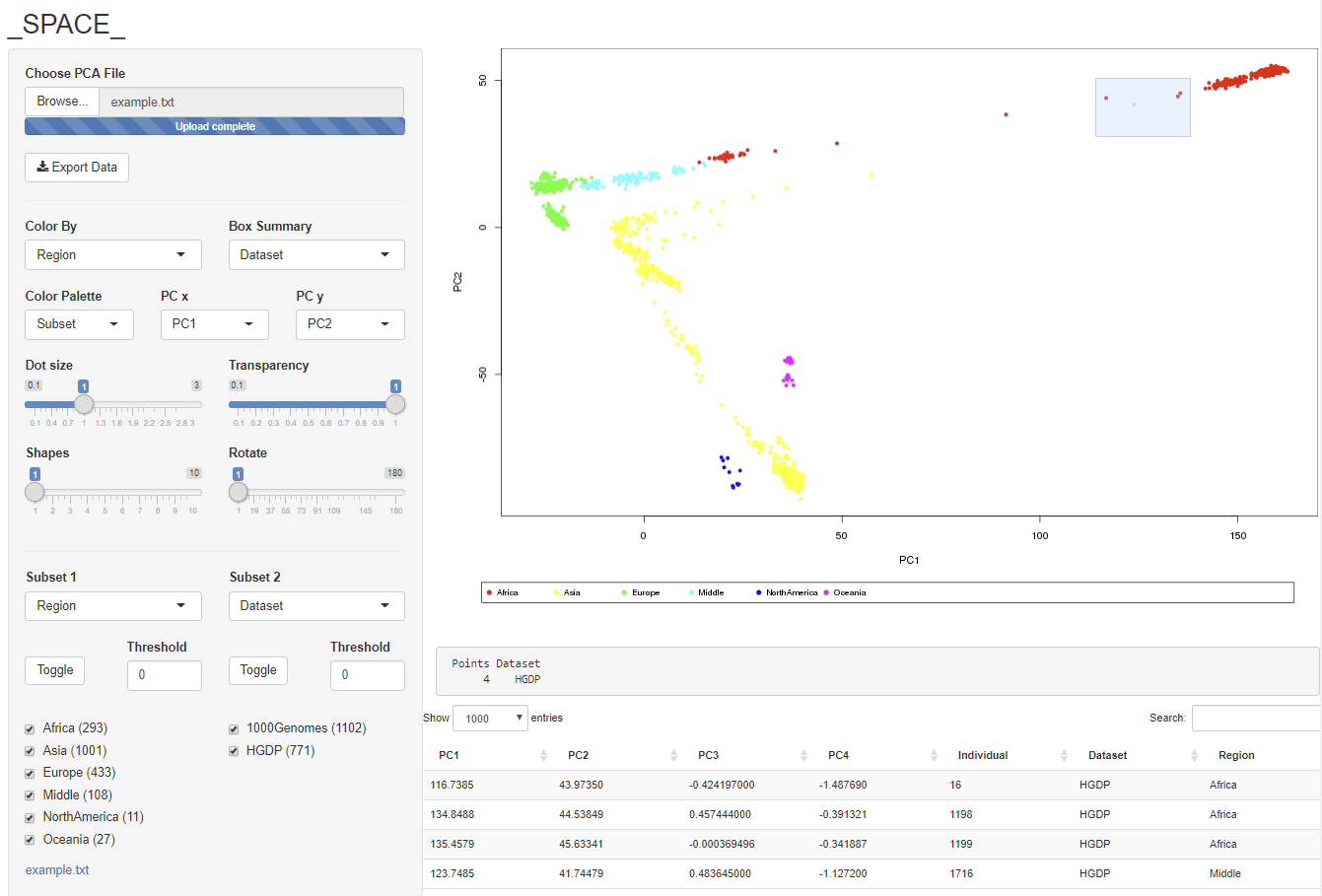
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ID** | **PC1** | **PC2** | **PC3** | **PC4** | **Region** | **Data Set** |
| 1 | .5 | .5 | .2 | .1 | Africa | HGDP |
| 2 | .6 | .9 | .2 | .4 | Asia | HGDP |
| 3 | .8 | .1 | .5 | .7 | Europe | HGDP |

In this case PC1, PC2, PC3 and PC4 are data. Continent and Data Set are labels. Clicking the Browse button opens a file browser.

## Choose which PCs to plot

By default, the first two PC columns in the file are used to make the plot. Changing the PC dropdowns will control which PCs are used

## Choose a label to color the plot

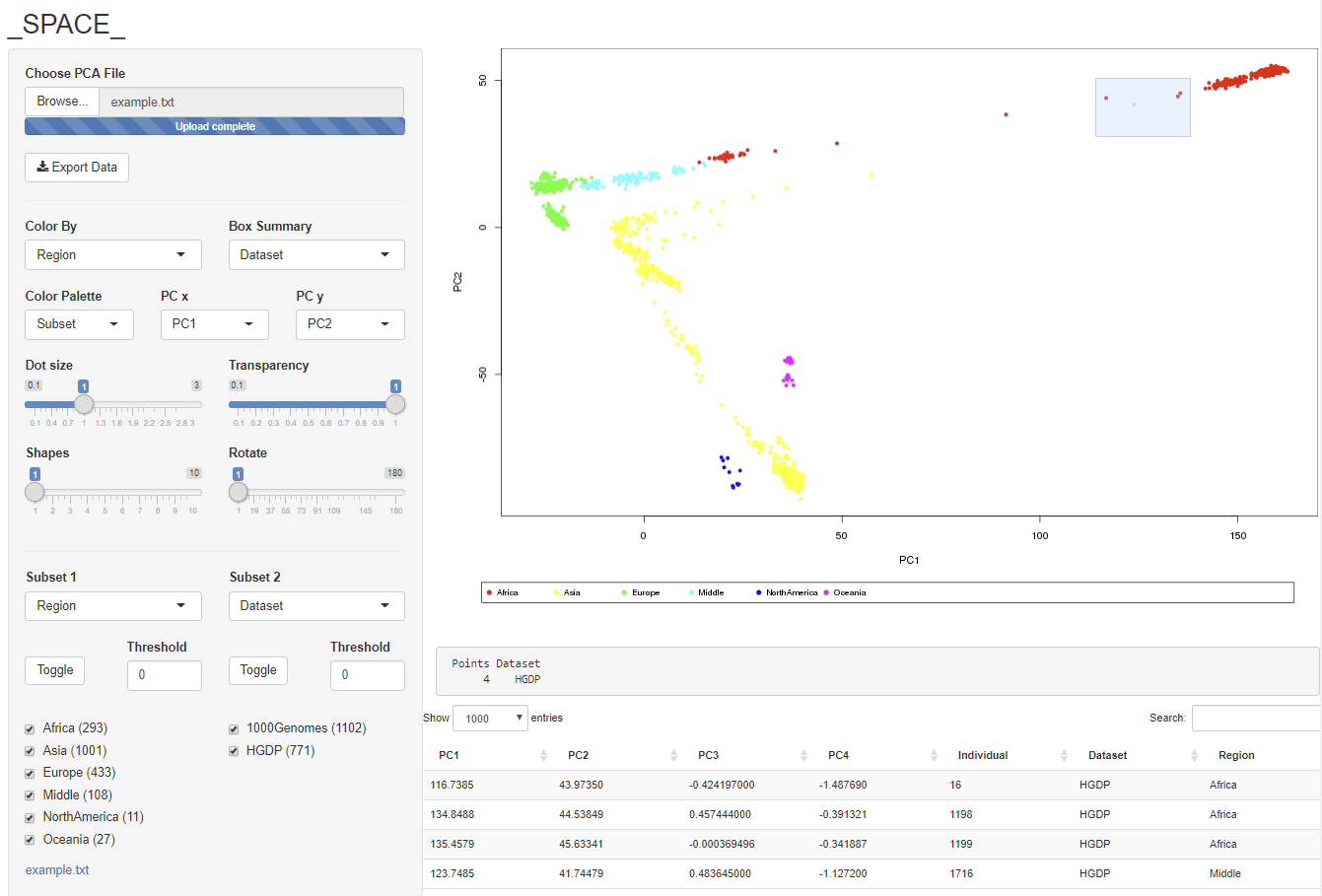
The Color By dropdown is automatically populated with all the label columns from the file, in this example there are just two: Region and Data Set, however any number of label columns is supported.

## Select points to inspect the data

Clicking and dragging on the plot will create a selection box. The data from the input file for the selected points is shown below. This is a useful method for inspecting outliers and other points of interest.

Selecting points also creates a summary based on the label columns. Here 4 points were selected, all from the HGDP data set.

The summary can be set to display any label or combination of labels using the Box Summary drop down.

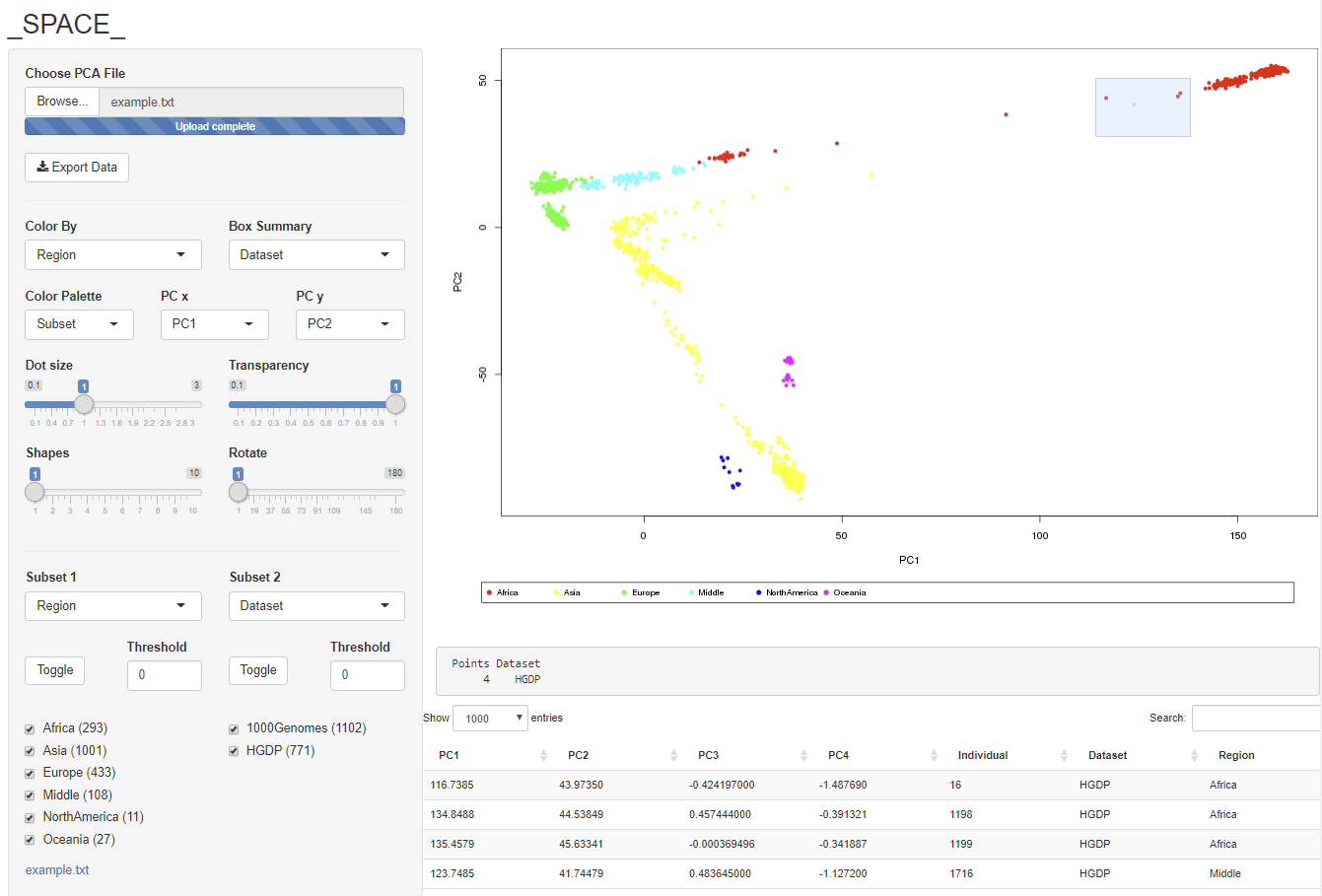


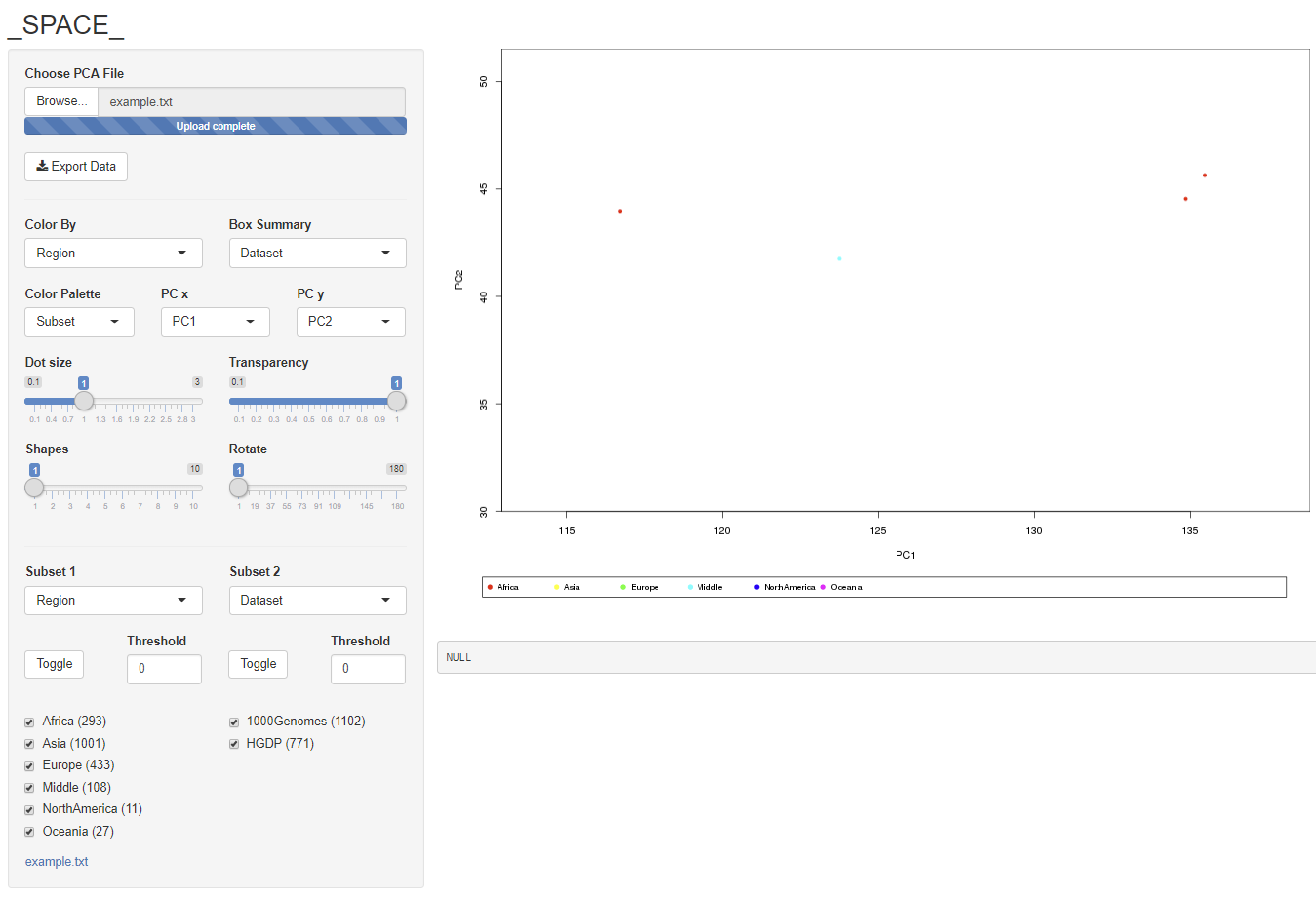
## Double click to zoom

Summary of selected points

Data for selected points

Control the summary

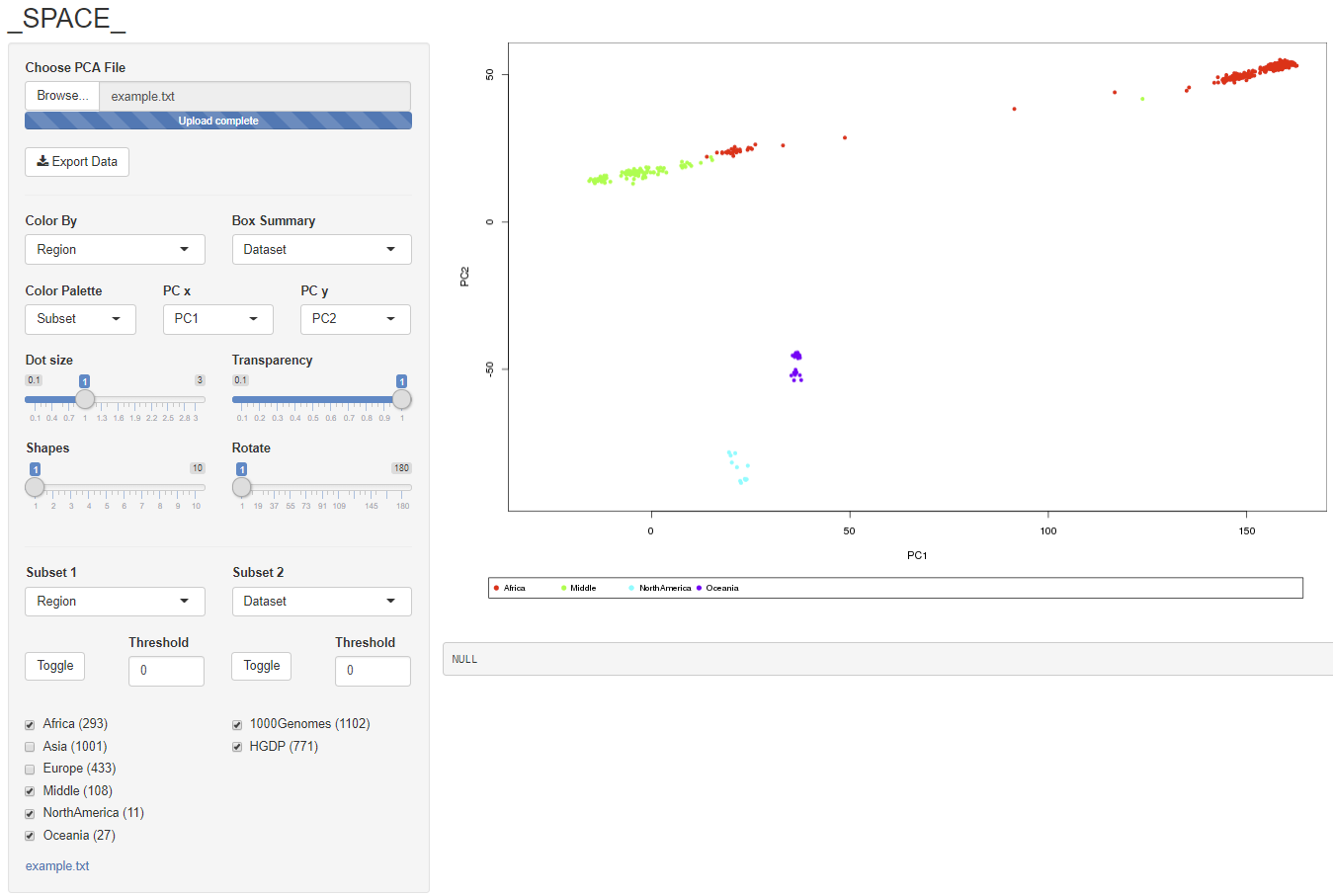
Double clicking a selection zooms in on it. Double clicking again zooms back out.



## Hide and show subsets by label

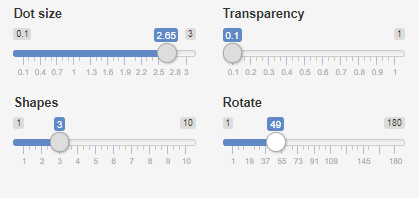
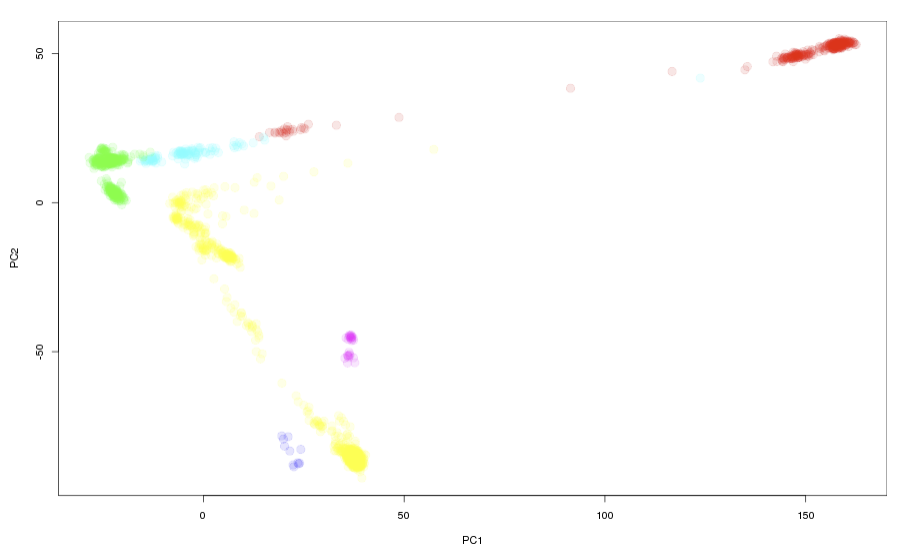
SPACE makes a panel of check boxes based on the labels in the file. By default, all boxes are checked. Unchecking a box will hide all data points the that label. The label column used to make the check boxes is set by a drop down. Selecting a different label will re-create the check boxes.

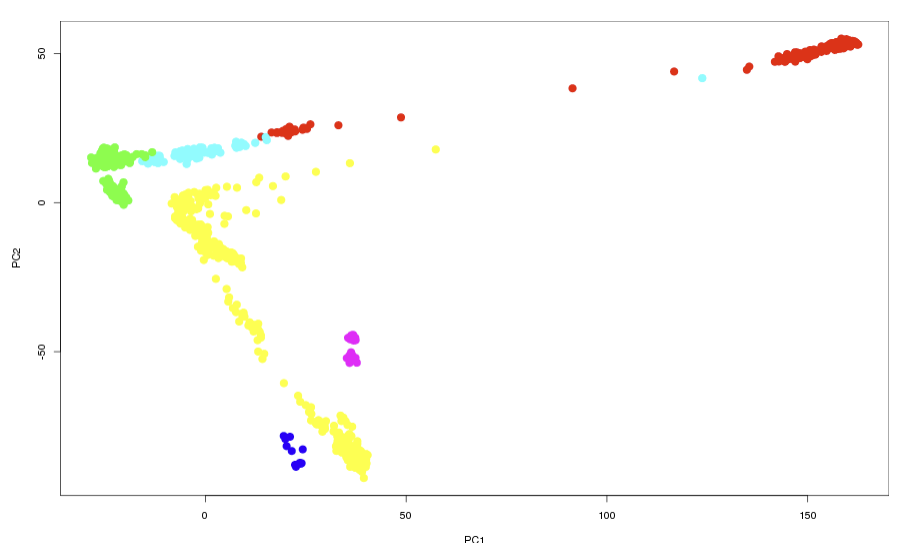
A second set of boxes can be used simultaneously. In this example, the user has chosen to show data from all data sets but has hidden samples from Asia and Europe.



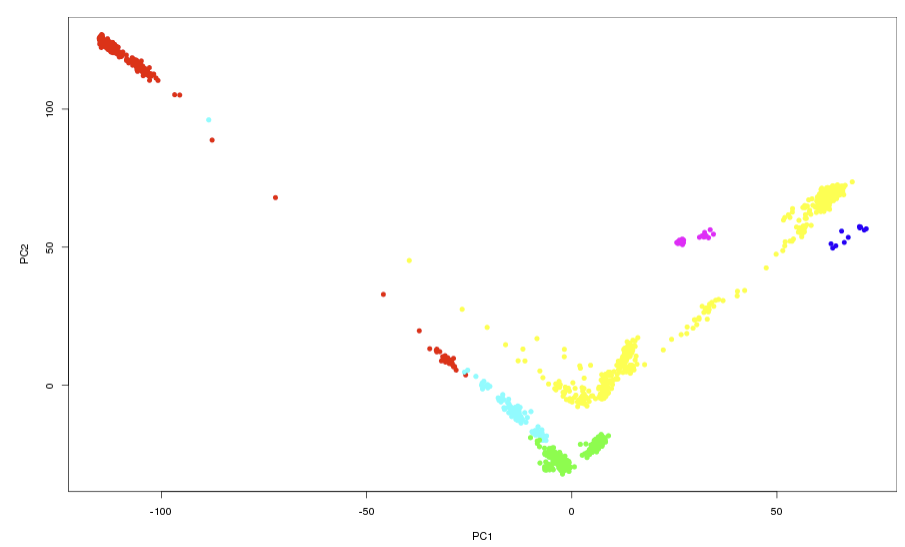
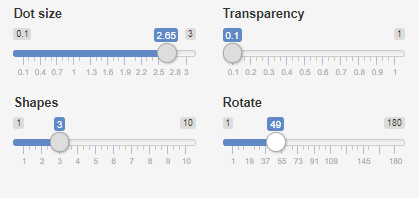
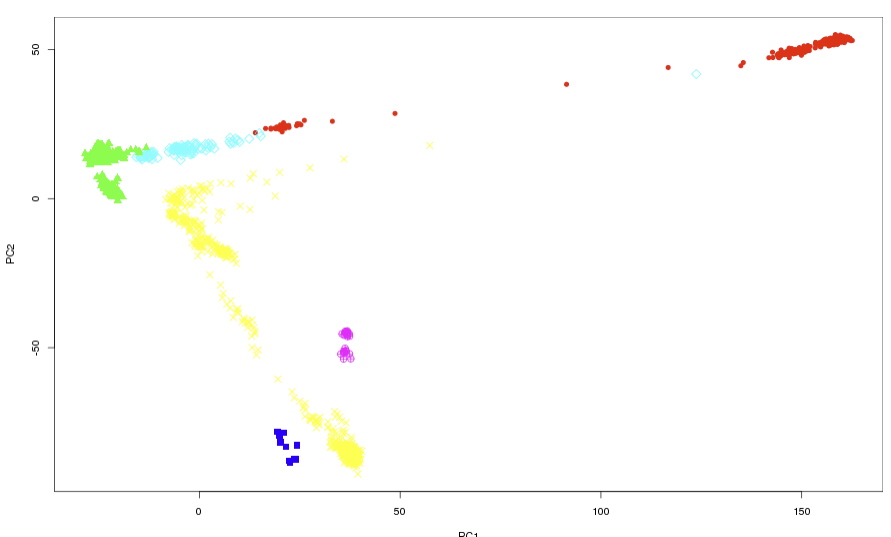
## Tune the plot’s look

A set of sliders control the look of the plot. Make the points larger or smaller. Visualize cluster density by changing the opacity of the plot





When the number of labels is large, it can be difficult to visually differentiate similar colors. In this case, SPACE can use more symbols for the points. SPACE can also rotate the plot for better clarity.



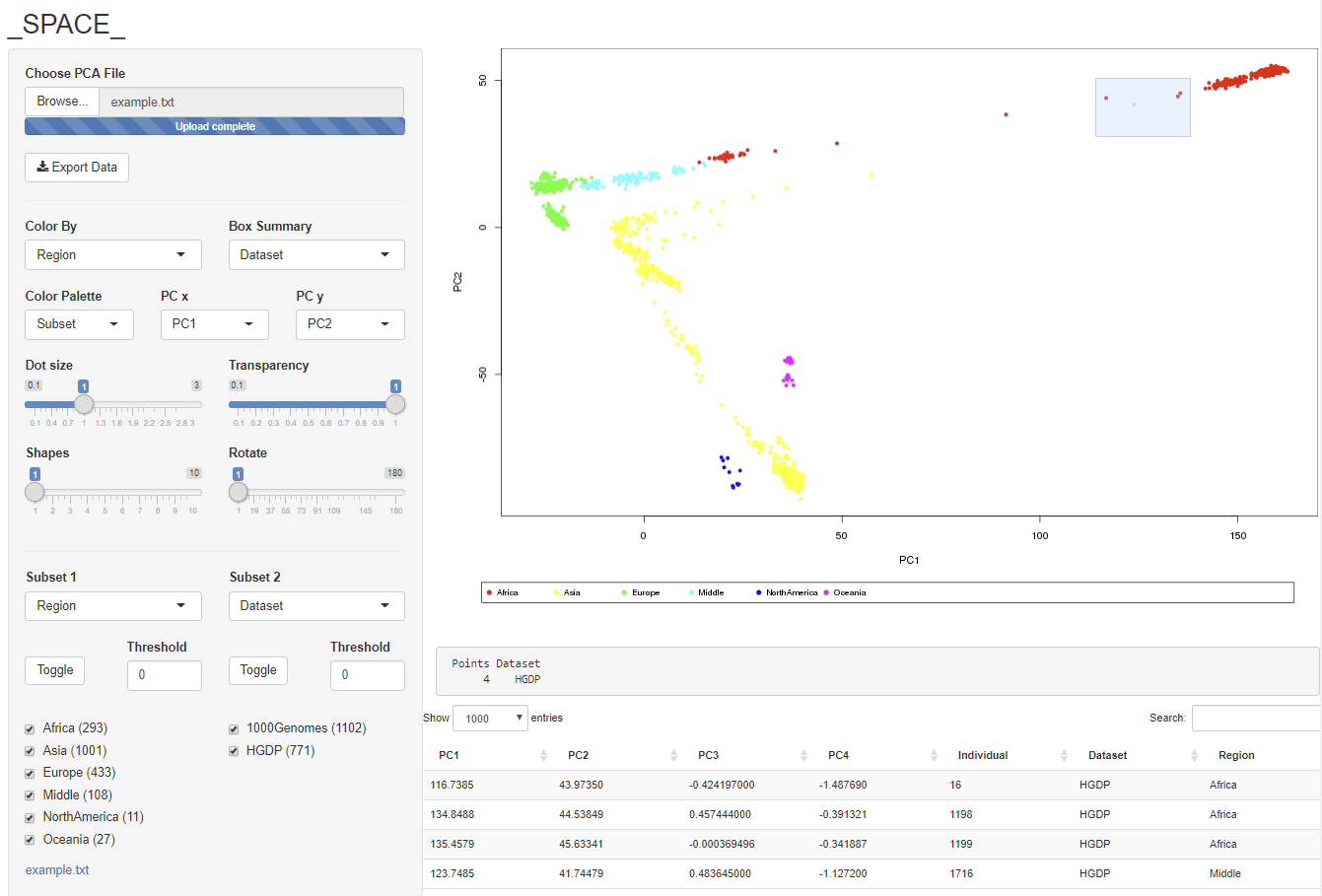
# Advanced

## Palette options

By default, SPACE colors using a rainbow palette based on the selected label. Hiding points will cause the plot to be recolored to spread out the tones.

Some users may wish to avoid the recoloring step to keep colors visually consistent during analysis. Choosing “All” from the palette drop down will force SPACE to keep the original colors.

As users hide and show points they may wish to orient themselves within the full PCA space. Choosing “Greyed Out” will show hidden points in grey. Clicking and dragging will not select greyed out points.

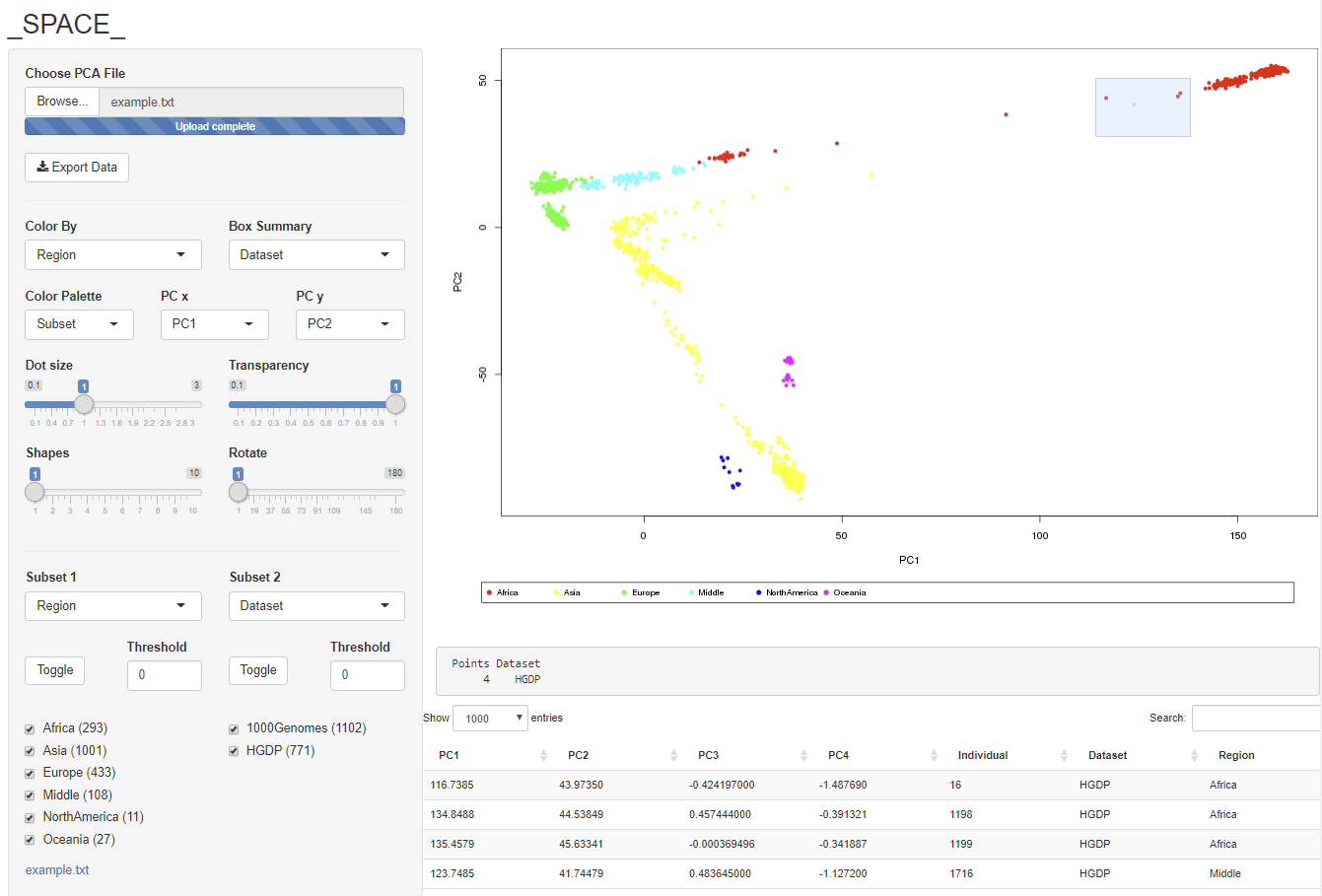


## Toggle

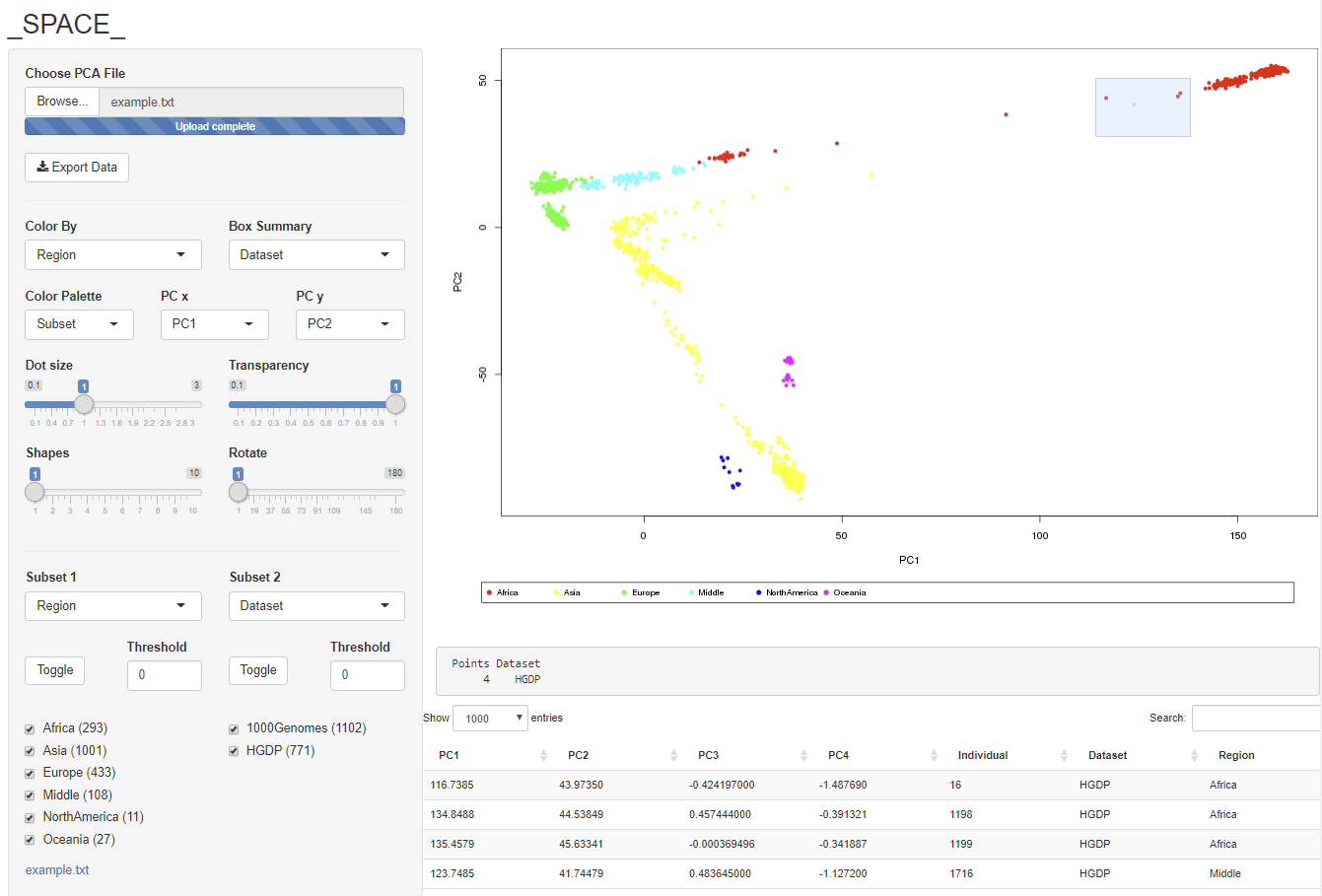
The toggle button unchecks all of the boxes. This is useful for users who want to add populations one by one rather than start with the full set and subtract them. Pressing the toggle a second time checks all the boxes.

## Thresholding

Some data sets have a large number of labels, and certain labels may only apply to a small number of individuals. Depending on the experiment, these populations may be of interest or may simply contribute to visual noise. Setting a threshold automatically hides populations with a small number of individuals. For example, setting the threshold to 10 will filter out individuals from populations smaller than 10



Export

To make a new file containing a subset of individuals, select the group of interest and click the export button.

# Source

## Got an idea for a new feature?

SPACE is open source and free to use for academic and non-profit entities. Please review the license in the GitHub repository.