SEAHORS: Spatial Exploration of ArcHaeological Objects in R Shiny

Aurélien Royer1,\*, Emmanuel Discamps2, Sébastien Plutniak3, Marc Thomas2

1 Biogéosciences, UMR 6282 CNRS, Université de Bourgogne, 6 Boulevard Gabriel, 21000 Dijon, France.

2 UMR 5608 TRACES, CNRS & Université Toulouse Jean Jaurès, Toulouse 31058, France.

3 CNRS/Université de Tours, UMR 7324 CITERES, Laboratoire Archéologie et Territoires, 40 Rue James Watt, ActiCampus 1, 37200, Tours, France.

\* corresponding author: aurelien.royer@u-bourgogne.fr

# **Abstract**

This paper presents SEAHORS, a R shiny application dedicated to the intra-site spatial analysis of piece-plotted archaeological remains. This open-source script generates 2D and 3D scatter and density plots for archaeological objects located with cartesian coordinates. Many different GIS software already exist for this, but they mostly require specific skills and training to be used and are rarely designed for specific archaeological applications. The goal of SEAHORS is to make the two and three-dimensional spatial exploration of archaeological data as user-friendly as possible, in order to give opportunities to researchers not familiar with GIS and R software to undertake such approaches. SEAHORS has an easily accessible interface and can import data from text and Excel files (.csv and .xls/xlsx respectively) without preformatting. The application includes functions to concatenate columns and to merge databases, for example when spatial data (XYZ coordinates) and analytical data (e.g. taxonomical attribution of faunal remains, typo-technological attributes of artifacts, etc.) are stored in separate files. SEAHORS can generate five types of plots: 3D, 2D and density plots, as well as 2D plot by slices (i.e. subdivisions according to a third dimension) and 2D plot allowing to modify the angle of projection (i.e. to explore spatial organization free from the constraints of the field grid orientation). SEAHORS has visualization tools with several sorting and formatting options (colors) applicable to coordinates and all possible analytical variables (i.e. levels, spits, identified species, taphonomical alterations, etc.). Orthophotos can also be imported and directly used in the program. The application also allows the grouping of objects into new variables by selecting items on the interactive 2D plots. We demonstrate the utility of the application, using the case study of the Cassenade Paleolithic site (France).

**Keywords:** archaeological methods, data visualization, spatial analysis, R shiny application, post-excavation stratigraphies

# **Introduction**

Archaeological excavation is a practice that, by its very nature, destroys the context of the studied material, preventing any replication of field observations. To minimize such losses, many archaeologists since the 19th century promoted the recording of 3D coordinates (X, Y and Z) of archaeological objects during excavation (cf. Plutniak 2021a for a review). Using these field 3D spatial data to virtually reconstruct the excavated deposits offers the opportunity to investigate spatial (horizontal and vertical) organizations of plotted archaeological objects. Such exploration allows to attest, confirm or infer stratigraphic and spatial observations made in the field, and to generate post-excavation stratigraphies and assemblages (see Discamps et al. submitted for a recent review of such use).

Following early calls for the development of spatial analyses of archaeological objects (e.g. Laplace & Méroc 1954), this practice was renewed when computer power became increasingly accessible to archaeologists from the end of the 1980s. Software dedicated to spatial analysis were developed in the next decades: Esri *ARC/INFO* in 1982 (Morehoust 1985); *Newplot* developed since 1986 (Dibble & McPherron 1988); *Paleo III* (by Pierre-Alain Gillioz, see application in Chadelle et al. 1996); *FrAcTool* (Houllier & Arnoux 2001); the adaptation of DataDesk to archaeological applications (Lacrampe-Cuyaubère 1997); QGIS in 2002 (QGIS.org 2023); the web application and R package *archeoViz* (Plutniak 2023; Plutniak submitted). More recently, web applications dedicated to specific archaeological sites were developed, such as *Plotit* for Combe-Capelle and others sites (R shiny application; [https://oldstoneage.com](https://oldstoneage.com/)); *PyCoCu* for Combe-Cullier (Python script; Sécher et al. 2020), *Virtual Poeymaü* for the Poeymaü Cave (R shiny application; Plutniak 2021b) and a R shiny application for the La Roche à Pierrot site (Couillet et al. 2022). However, using these software for analyzing spatial distributions of archaeological remains can be challenging due to several factors, such as complex interfaces, difficulties in adapting software to specific needs or databases to the required format, lack of user training, cost, or the absence of a supporting community.

This paper presents SEAHORS (Spatial Exploration of ArcHaeological Objects in R Shiny), an R Shiny free open-source application, that allows easy and quick exploration of spatial distribution of archaeological objects. SEAHORS is available on <https://github.com/AurelienRoyer/SEAHORS>. A web application is deployed on the shinyapps.io platform and is available here: <https://aurelienroyer.shinyapps.io/Seahors/>

**SEAHORS overview**

SEAHORS is a R shiny application. R is a free and open-source statistical environment, with a large supportive community (R core team 2022). However, using a programming language can be challenging for untrained archaeologists. The *Shiny* package offers the possibility to build a point-and-click interactive interface, allowing for a more user-friendly experience (Chang et al. 2022). SEAHORS also relies on the *Plotly* package (Sievert et al. 2017), offering interactive data visualization by displaying information associated with each point directly on the plots. This overview presents SEAHORS’ main features. In addition, a tutorial video using data of the Paleolithic site of Cassenade (Discamps et al. 2019) as an example is available at <https://nakala.fr/10.34847/nkl.3fdd6h8j>.

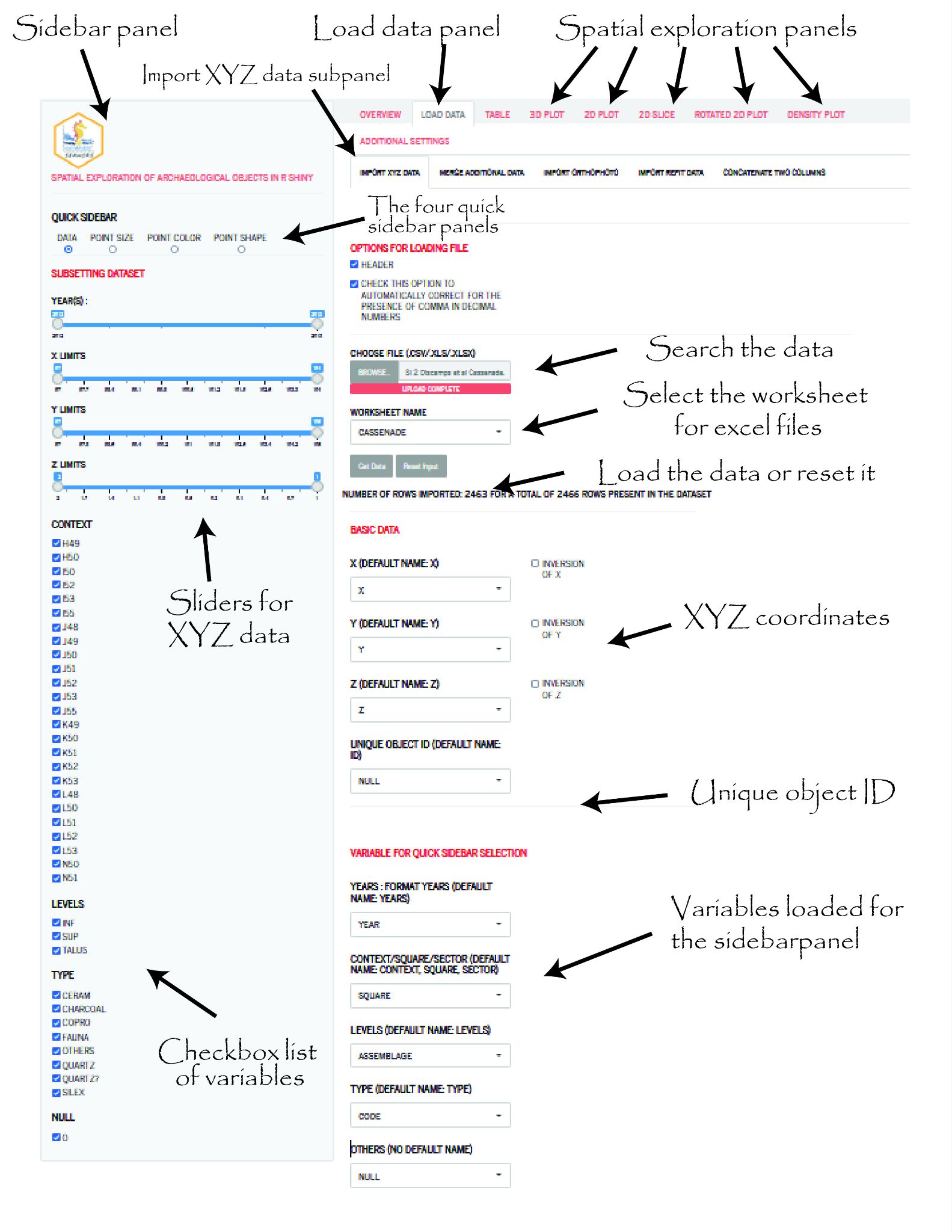
***Data import: supported files***

SEAHORS requires tabular data and supports Excel spreadsheet (.xls/xlsx) and comma-separated values (.csv) formats. We define a “variable” as the information stored in one column of the table, an “object” as information recorded in one row of the table, and “modalities” as the variables’ values (i.e. all the forms that a variable can take in the column). Variables can thus be numeric (e.g. XYZ coordinates) or categorical (e.g. stratigraphic unit or type of finds, with variable modalities such as “flint”, “fauna” and “other”). SEAHORS requires a table with at least two columns with X, Y, or Z coordinates (for bi-dimensional exploration) or three columns (X, Y, and Z).

***Data import:* the *Load data* panel**

The *Load data* panel allows importing different types of data for spatial exploration. This panel is divided into five subpanels:

1. The first subpanel, *Import XYZ data*, is the most important and is a necessary step before spatial exploration. It allows to set up the four basic data to import, namely the XYZ coordinates and the ID of objects. If the three columns of the XYZ coordinates are named “X”, “Y” and “Z”, respectively, in the imported data frame, the script will automatically load them. Otherwise the column names can be manually set up. Note that only rows with numeric values for the three X, Y and Z coordinates will be taken into account by SEAHORS. The values of the X, Y and Z coordinates can be inverted (i.e. multiplied by -1) by checking the boxes in front of these data: this is useful for instance when Z values were recorded in the field as depth and not altitude. The ID of objects must be unique (i.e. the variable dedicated to identifying the objects has a distinct modality for each object), and a warning message is displayed if that is not the case. Although the absence of unique IDs does not prevent analysis, it is required for some data manipulations, notably to avoid errors while merging the XYZ coordinates with additional data. In the bottom part of this subpanel, the user can select the names of the variables corresponding to the excavation year (numeric format), the sectors, the levels, the type of objects, and any other categorical variables to be displayed in the quick selection sidebar on the left (Figure 1, see below). None of these variables is mandatory, and the names listed as “default names” are only indicative, just allowing these variables to be loaded automatically when the data frame is imported; whatever name is used in the imported table will not affect the program. The box in front of each variable allows the user to select the corresponding column of the imported sheet.



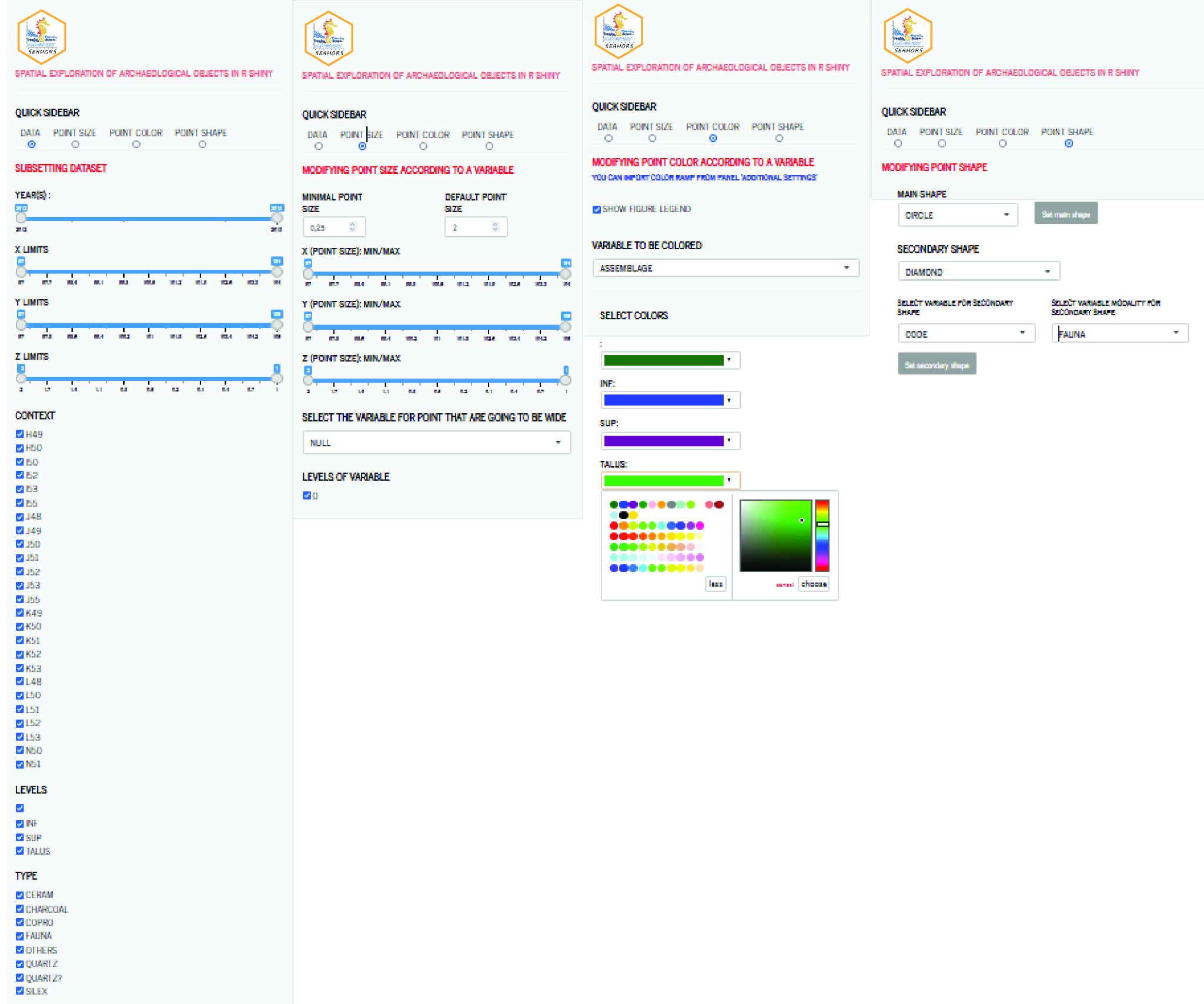
**Figure 1:** The*Load data panel* of the application *SEAHORS* (see text and supporting video for more details).

1. The second subpanel, entitled *Merge additional data*, offers the possibility to import supplementary data from a second dataset (i.e. a second Excel or text file). To do this, determining unique object IDs in the *Import XYZ data* subpanel is required. These unique IDs are used to merge the XYZ dataset and the supplementary dataset; it must then be present in the two files. After merging, a report displayed in the bottom part of the subpanel describes how the result of the merging by listing: the IDs of objects that are not unique 1) in the XYZ dataset, and 2) in the supplementary dataset, 3) the objects present in the XYZ dataset only (i.e. no object ID correspondence in the supplementary dataset), and 4) objects from the supplementary dataset not merged because of a lack of correspondence in the XYZ dataset.
2. The third subpanel, *Import orthophoto*, allows loading an orthophotographic image that could be then used in the *2D plot* and *Density plot*. The orthophotographic image should be obtained from a geo-referenced model using the same excavation grid as the XYZ dataset. The orthophotographic image should be in .TIFF and smaller than 150 MB.
3. The four subpanel is used to *Import refit data*. Users can either import a separate file that lists refit data or select the appropriate column from the main dataset (the one already imported using the *Import XYZ data* subpanel). The refit data should be organized as a spreadsheet listing refitted objects (one row per object) with one column the object’s unique ID and one column for the ID of the refit group to which the object belongs. Groups of refitted objects can link two objects or more. For instance, if we consider that three objects (unique IDs 1300, 1502 and 1603) refit together, then they belong to one refit group ID, which we name “refit group 1”. The refit database should then consist of 2 columns (unique IDs, refit group ID) with three rows with “1300”, “1502” and “1603” in the first column, and in the second column “refit\_group\_1”, “refit\_group\_1”, “refit\_group\_1”.
4. The last subpanel (*Concatenate two columns*) gives the possibility to concatenate two columns. This can be very handy, for example, to merge the square values (e.g. C50) and the object number (e.g. 3) to generate unique object IDs (e.g. C50-3).

***The “Quick sidebar” panel***

The “Quick sidebar” allows the user to easily subset the dataset to facilitate visualization, as well as to modify point size, color or shape in the different plots. Modifying the sliders and check-boxes directly modify the objects that are included in projections. This panel is subdivided into four subpanels easily accessible with buttons (Figure 2):

1. The first one is dedicated to subset the dataset shown in the plots. Numeric variables will be displayed with sliders and categorical ones as check-box lists;
2. The second one can be used to adjust the point size according to their XYZ coordinates or to categories (i.e. variable modalities);
3. The third subpanel is used to color points according to the modalities of a variable;
4. The last one is used to change the point shapes, modifying the point shape (circle, triangle, square, diamond) either for all points or for specific variable modalities.

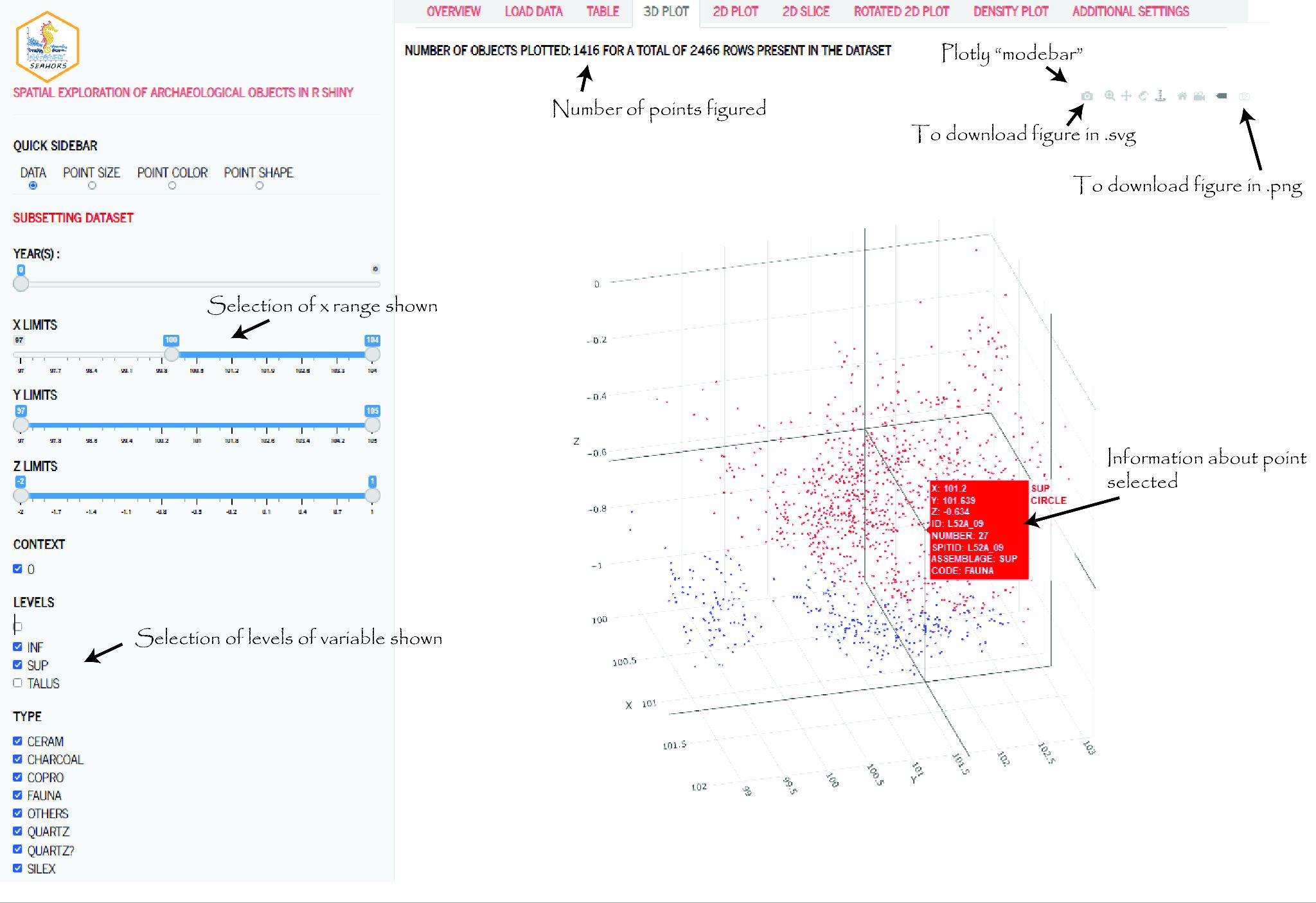
**Figure 2:** The *Quick sidebar* panel and its four subpanels, that allows to subset the dataset, modify the point size, color and shape.

***The Table panel***

The *Table* panel is divided in two subpanels: *Raw table* and *Pivot table*. The first offers the opportunity to download the table of data of the selected objects to make the projections, including the additional data that might have been merged, as well as the new column(s) that might have been created (cf. *2D plot* in “spatial exploration of data” paragraph). The second subpanel allows the user to create pivot tables to count objects according to a variable and its modalities.

***Spatial exploration of data***

Spatial exploration of data can be carried out using five different types of plots, all directly accessible with one click on the tabs at the top: *3D plot*, *2D plot*, *2D slice*, *Rotated 2D plot* and *Density plot*. The first two allow users to plot the objects in 3D and 2D (Figures 3, 4 and 5).

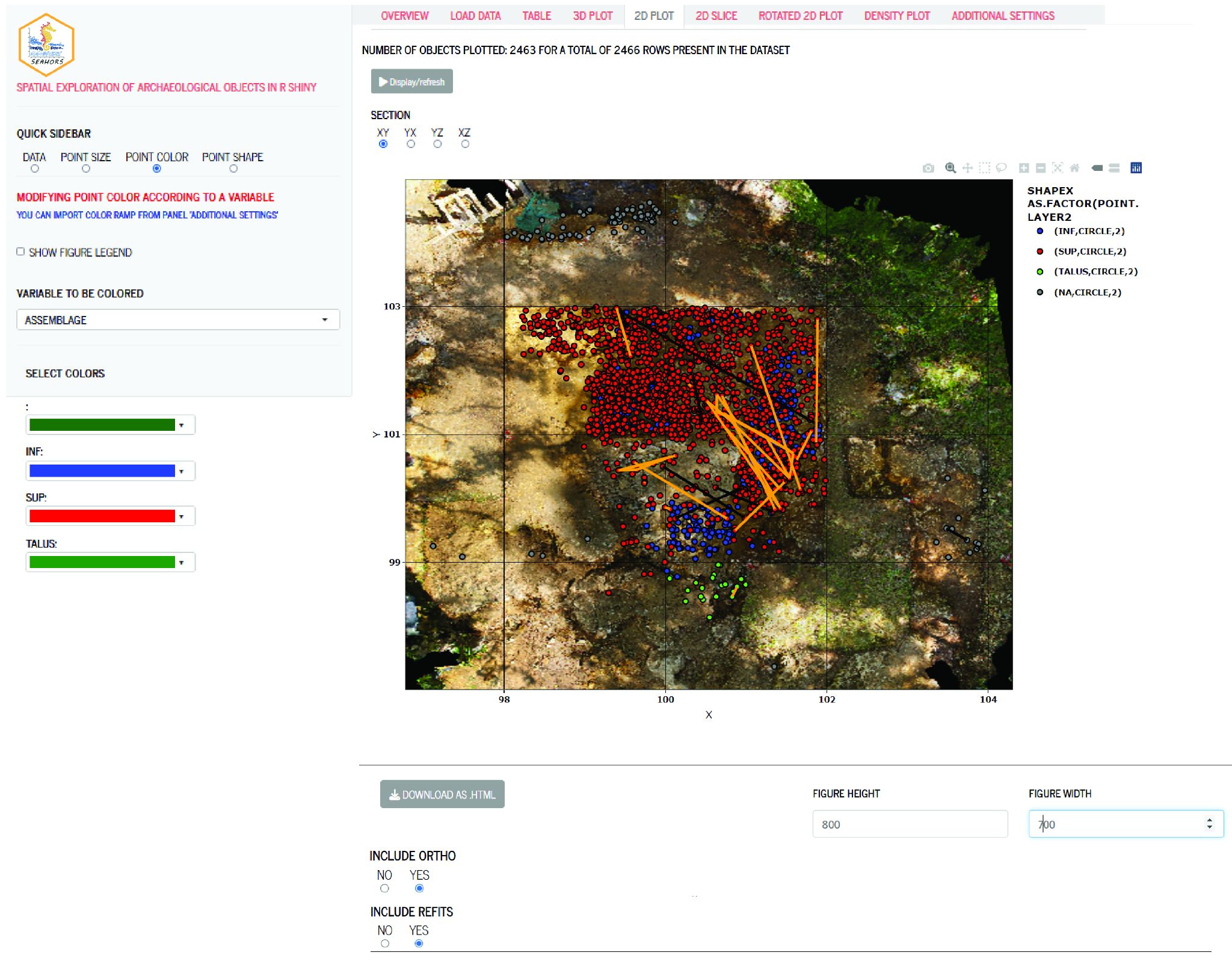


**Figure 3:** Example of a 3D plot of the Cassenade site. Hovering on a point makes it possible to obtain information on it. Here, a faunal fragment from the upper assemblage of the site. The plotly “modebar” allows to download the 3D projection in a graphic format (.png or .svg).

The *2D slice* panel offers the option of automatically creating sequential 2D plots after virtually “slicing” the site according to the axis of a third coordinate (X or Y). The user can select the range of the third axis to be used, as well as the thickness desired for slices. This slicing is automatic and produces the corresponding plots in the two other coordinates (e.g. YZ) for each slice. Users should first define the total range and thickness of the slices and the script will then calculate and display the 2D plot slices.



**Figure 4:** Example of Cassenade with 2D plots. Information of the point selected in the *2D plot panel* corresponds to a fauna element from the upper levels. The plotly “modebar” allows to download the 2D projection in a graphic format (.png or .svg).



**Figure 5:** Example of 2D plots of the Cassenade site. Here, plotted objects are colored according to the variable “Assemblage”, lines connect refits, and an orthophotographic image of the site is used as a background. Lithic and faunal refits are displayed in orange and black, respectively.

The *rotated 2D plot* panel is useful when users suspect that the orientation provided by the original XY grid of the excavation is not optimal for visually exploring the spatial organization of archaeological objects. In this panel, users can modify the angle of the projection (from -180 to 180°) to explore the spatial organization of objects without being constrained by the grid pattern. These new values of X and Y can be loaded in the sidebar panel for further investigations in the *2D plot* or *3D plot* slides.

Finally, the *Density plot* panel offers the possibility to plot objects with a gradient color depending on the density of the nearest points. The density of points and the density lines are calculated using a two-dimensional Kernel density estimation from the functions ‘Ke2D’ and ‘geom\_density\_2d’ of the package *Mass* (Ripley et al. 2013). The density curves are obtained using the ‘geom\_density’ function from the *ggplot2* package (Wickham et al. 2016).

**Additional settings**

In the last panel, a few more additional settings are available to modify the way plots are displayed:

* Color ramps can be saved (or loaded) if the user is frequently using the same set of colors according to variable modalities.
* Color of refits can also be changed according to the modalities of a variable (and this parameter can similarly be saved or loaded). In the specific case of a refit between two objects that have two different modalities of the chosen variable, the color of the refit will be black in the *2D plot*, and gradient in the *3D plot*.
* The “steps” used for sliders can be modified.
* The information displayed while hovering a point on projections can be customized.
* New variables can also be created and coded directly in SEAHORS. This is for example useful when the user needs to define new groups of objects (e.g. assemblages), and assign objects to each group on the projections. These new groups can later be used in the quick sidebar panel like any other variable, loading them in the *Load data* panel. The first step is to go to the *record new group* subpanel in the *Additional settings* panel. Then, the user chooses a name for the new group variable and creates it. At this step, data from another variable can be copied, a useful option if the new group consists of a slight modification of an already present variable. The user can create as many new group variables as necessary. When the new group is created, you can select points in *2D plots* by using the “box” or “lasso” tools and change their attribution by using the button “Change Group Assignment”. The list of points selected is detailed below the latter. This new data can be downloaded using the *raw table* panel.
* Finally, the last subpanel of *Additional settings* can extract a report in the format of a .html file (using the *rmarkdown* package, Allaire et al. 2023). This report details the parameters used and includes the data tables and the figures obtained.

**Exporting data and figures**

SEAHORS allows for the export of table data and figures. The tables (*raw table*, *pivot table*, *refit table* and the rotated coordinate data) can be downloaded in .csv format, through buttons placed in each corresponding panel. All modifications of the data made using “merge additional data”, “merge two columns”, or “change group assignment” functions can be downloaded through the *raw table* panel. The figures can be downloaded in .html by using the buttons below projections or in .png and in .svg formats by using the buttons in the bar above the plot (the plotly “modebar”, cf. Figures 3 and 4). The latter format (.svg) is the most convenient, making possible to easily edit the figures in a vector-based graphics software (e.g. Inkscape), or convert them into other formats, such as .tiff or .pdf.

**Conclusion**

Since the middle of the 20th century, archaeologists have been recording an increasing amount of spatial data during fieldwork, with the intent of preserving contextual information as much as possible. Such tedious work is now indispensable to propose reliable spatial (i.e. planimetrical and stratigraphical) interpretations. As a result, it leads to the accumulation of thousands or more datasets and object coordinates, the analysis of which requires adequate software and extensive training in GIS software. The aim of the SEAHORS application is to offer tools for the spatial exploration of archaeological objects to a broader audience, through a point-and-click interactive interface, in a free and open-source environment. We hope that this software (in its current form and through future additions) will foster the spatial exploration of archaeological sites and, in particular, thus the practice of post-excavation stratigraphic analysis.

**Acknowledgements**

The authors are also grateful to Rémi Laffont for his help.

**Data, scripts, code, and supplementary information availability**

R shiny script: <https://github.com/AurelienRoyer/SEAHORS>

Zenodo. DOI:<https://doi.org/10.5281/zenodo.7664192>

A web application: <https://aurelienroyer.shinyapps.io/Seahors/>

The web application does not allow to import orthophotographic images nor to export a rmarkdown report.

The tutorial video of SEAHORS in english: <https://nakala.fr/10.34847/nkl.3fdd6h8j>

The tutorial video of SEAHORS in french: [https://nakala.fr/10.34847/nkl.65bf1h72](https://nakala.fr/10.34847/nkl.3fdd6h8j)

Cassenade spatial data: <https://nakala.fr/10.34847/nkl.e30aie2u>

Cassenade refit:<https://nakala.fr/10.34847/nkl.fe27j10z>

Cassenade orthophoto:<https://nakala.fr/10.34847/nkl.7ea78e6u>

**Installation**

Install R (<https://www.r-project.org/>) and eventually Rstudio Desktop (<https://posit.co/download/rstudio-desktop/>) to have a more comfortable R environment.

Open the script in R or Rstudio.

To run it in R, the user must highlight the code (“Select all” or CTRL + A) and click on “Run all” or press CTRL + R.

To run it in Rstudio, the user can either highlight the code (“Select all” or CTRL + A) and click on “Run selected lines” or CTRL + Enter, or simply click on the "Run App" button (at the top of the editor window).

The SEAHORS application will be launched and the necessary R package will be installed automatically the first time. The first launch may therefore take some time.

**Conflict of interest disclosure**

The authors declare that they comply with the PCI rule of having no financial conflicts of interest in relation to the content of the article.

**Funding**

This research was funded by the DeerPal project (French National Research Agency, ANR-18-CE03-0007).

**References**

Allaire, J.J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W., & Iannone, R. (2023). *rmarkdown: Dynamic Documents for R*. R package version 2.20,<https://github.com/rstudio/rmarkdown>.

Chadelle, J.-P., Geneste, J.-M., & Texier, J.-P. (1996). Les gisements paléolithiques récemment découverts à Pombonne (Lembras, Dordogne). *PALEO. Revue d'archéologie préhistorique,* *8*(1), 279-291. <https://doi.org/10.3406/pal.1996.926>

Chang, W., Cheng, J., Allaire, J.J., Xie, Y., McPherson, J., Dipert, A., & Borges, B. (2022). *shiny: Web Application Framework for R.* R package version 1.7.4, <https://cran.r-project.org/web/packages/shiny/index.html>

Couillet, A., Rougier, H., Todisco, D., Marot, J., Gillet, O., & Crevecoeur, I. (2022). New Visual Analytics Tool and Spatial Statistics to Explore Archeological Data: The Case of the Paleolithic Sequence of La Roche-à-Pierrot, Saint-Césaire, France. J*ournal of Computer Applications in Archaeology, 5*(1), 19-34. <http://doi.org/10.5334/jcaa.81>

Dibble, H.L., & McPherron, S.P. (1988). On the computerization of archaeological projects. *Journal of Field Archaeology*, *15*(4), 431-440. <https://doi.org/10.1179/jfa.1988.15.4.431>

Discamps, E., Bachellerie, F., Baillet, M., & Sitzia, L. (2019). The use of spatial taphonomy for interpreting Pleistocene palimpsests: an interdisciplinary approach to the Châtelperronian and carnivore occupations at Cassenade (Dordogne, France). *Paleoanthropology*, *2019*, 362-388. <https://doi.org/10.4207/PA.2019.ART136>

Discamps, E., Thomas, M., Dancette, C., Gravina, B., Plutniak, S., Royer, A., Angelin, A., Bachellerie, F., Beauval, C., Bordes, J.-G., Deschamps, M., Ducasse, S., Langlais, M., Laroulandie, V., Mallye, J.-B., Michel, A., Perrin, T., & Rendu, W. (submitted). Breaking free from field layers: the interest of Post-Excavation Stratigraphies (PES) for producing reliable archaeological interpretations and increasing chronological resolution. *Research Square Preprint*, <https://doi.org/10.21203/rs.3.rs-2405909/v1>

Houllier, J.R., & Arnoux, T. (2001). Informatisation d’une fouille: réalisation et déploiement du logiciel d’acquisition de données «FrAcTool» (Fressignes Acquisition Tool). *PALEO. Revue d'archéologie préhistorique*, 13, 251-264. <https://doi.org/10.4000/paleo.1096>

Lacrampe-Cuyaubere, F. (1997). Utilisation du logiciel Data Desk™ pour la construction de projections de vestiges archéologiques: une application automatisée à l'usage des préhistoriens. *PALEO. Revue d'archéologie préhistorique*, *9*(1), 335-349. <https://doi.org/10.3406/pal.1997.1240>

Laplace-Jauretche, G., & Méroc, L. (1954). Application des coordonnées cartésiennes à la fouille d'un gisement. *Bulletin de la Société préhistorique de France*, *51*(Fasc. 1/2), 58-66. [https://doi.org/](https://doi.org/10.3406/pal.1997.1240)[10.3406/bspf.1954.12427](https://doi.org/10.3406/bspf.1954.12427)

Morehouse, S. (1985). ARC/INFO: A geo-relational model for spatial information. In *Auto-Carto 7*.

Plutniak, S. (2021a). Abstraction in Archaeological Stratigraphy: a Pyrenean Lineage of Innovation (late 19th-early 21th century). In S. de Beaune, A. Guidi, O. Moro Abadía, M. Tarantini (eds). New Advances in the History of Archaeology, Oxford: Archaeopress, 78-92. Available at: <https://hal.archives-ouvertes.fr/hal-03338343v2>

Plutniak, S. (2021b). *Virtual Poeymaü: a web application to explore the archaeological data from the excavation archives of the Poeymaü cave (France) (v0.2)*. Zenodo. DOI: [10.5281/zenodo.4765693](https://doi.org/10.5281/zenodo.4765693)

Plutniak, S. (2023). *archeoViz. Visualisation, Exploration, and Web Communication of Archaeological Excavation Data*. DOI: [10.5281/zenodo.7512526](https://doi.org/10.5281/zenodo.7512526); R package v0.2.1,<https://cran.r-project.org/web/packages/archeoViz/index.html>.

Plutniak, S. (submitted), “Visualiser et explorer la distribution spatiale du mobilier archéologique : l’application archeoViz et son portail web”.

QGIS Development Team (2002). QGIS Geographic Information System. QGIS Association. <https://www.qgis.org>

Ripley, B., Venables, B., Bates, D.M., Hornik, K., Gebhardt, A., Firth, D., & Ripley, M.B. (2013). Package ‘MASS’. *Cran r*, *538*, 113-120. <https://cran.r-project.org/web/packages/MASS/index.html>

Sécher, A., Langlais, M., Berlioz, É., Delvigne, V., Pétillon, J.-M, & Renou, S. (2020). Combe-Cullier (Lacave, Lot) : une séquence oubliée du Magdalénien. *Rapport de deuxième année d’aide à la préparation de publication, DRAC Occitanie*, 142 p.

Sievert, C., Parmer, C., Hocking, T., Chamberlain, S., Ram, K., Corvellec, M., & Despouy, P. (2017). plotly: Create Interactive Web Graphics via ‘plotly. js’. *R package version*, *4*(1), 110. <https://cran.r-project.org/web/packages/plotly/>

Team, R. C. (2022). R: A language and environment for statistical computing. <https://www.r-project.org/>

Wickham, H., Chang, W., Henry, L., Lin Pedersen, T., Takahashi, K., Wilke, C., Woo, K., Yutani, H., & Dunnington, D. (2016). *Package ‘ggplot2’.* *Create elegant data visualisations using the grammar of graphics. Version*, *2*(1), 1-189. <https://cran.r-project.org/web/packages/ggplot2/index.html>