

Creating a Graphical Tool for Non-Programmers to Use to Make Heatmaps

Nicholas Alicea

Centre College

Danville, KY

nicholas.alicea@centre.edu

Akenpaul Chani

Centre College

Danville, KY

aken.chani@centre.edu

Lam Le

Centre College

Danville, KY

lam.le@centre.edu

Hayata Suenaga

Centre College

Danville, KY

hayata.suenaga@centre.edu

David Toth

Centre College

Danville, KY

david.toth@centre.edu

Selam Van Voorhis

Centre College

Danville, KY

selam.vanvoorhis@centre.edu

Jessica Wooten

Piedmont University

Demorest, GA

jwooten@piedmont.edu

ABSTRACT

Heatmaps are used to visualize data to enable people to quickly understand them. While there are libraries that enable programmers to create heatmaps with their data, scientists who do not typically write programs need a way to quickly create heatmaps to understand their data and use those figures in their publications. One of the authors is not a programmer but needed a way to generate heatmaps for their research. For a summer undergraduate research experience, we created a program with a graphical user interface to allow non-programmers, including that author, to create heatmaps to visualize their data with just a few mouse clicks. The program allows the user to easily customize their heatmaps and export them as PNG or PDF files to use in their publications.

Categories and Subject Descriptors

J3 [Life and Medical Sciences] — Biology and genetics

General Terms

Design, human factors

Keywords

Heatmap, Bioinformatics, Genetics, Software development, Student experience, Student research

1. INTRODUCTION

One of the authors is a biologist who needed a tool to enable the creation of a visual representation of bioinformatics research data

for research projects on which they were working. That author has minimal experience in computer programming. They reached out to one of the other authors, a computer scientist, to see if some of the computer scientist's students could do a project that could produce the necessary tool, a program with a simple graphical user interface (GUI) that would allow the creation of heatmaps from data sets stored in comma-separated value (CSV) files. The GUI would allow the user to create heatmaps without needing to write computer programs. The project would allow the students to gain experience in software development and user interface design while building a tool to advance science.

2. BACKGROUND

A heatmap is a representation of data in the form of a map or diagram in which data values are represented as colors depending on the magnitude of the value [1]. "They are often used to visualize high-frequency data or when seeing general patterns is more important than exact values" [2]. A general heatmap takes a two-dimensional set of values as its input and maps each of the values to a color. Heatmaps can be used to visualize or represent raw data in an easily digestible format that can enable users to understand their data. Figure 1 shows a sample heatmap with made-up data for levels of contamination in a small region of a city. The heatmap makes it easy to spot the extremely high levels of contamination in the regions represented by Row 5 Column 4 and Row 8 Column 10, in contrast to areas with high, moderate, and low levels of contamination. Our project was to create a GUI to allow scientists who are inexperienced with coding to easily create a heatmap to display their data.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Copyright ©JOCSE, a supported publication of the Shodor Education Foundation Inc.

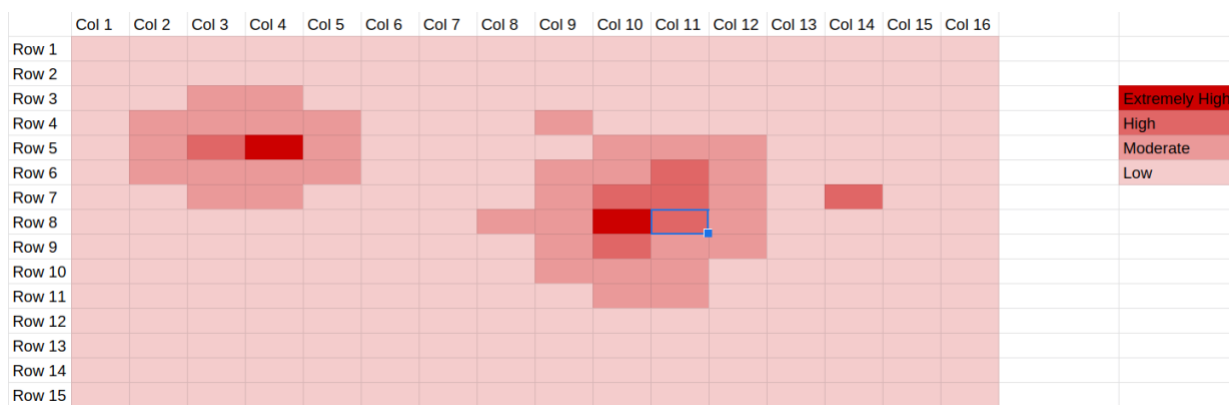


Figure 1. A sample heatmap showing contamination levels in a city.

The biologist author’s primary focus was identifying taxonomies of bacteria that are particularly prevalent among similar microbiome samples. A microbiome is the total collection of microorganisms present in a particular environment. A bioinformatics lab took the author’s data and analyzed them for occurrence of operational taxonomic units (OTUs), resulting in a CSV file composed of raw data representing the occurrence of each OTU in each sample. An OTU can be defined as a categorical classification based on genetic similarity, providing a method of understanding what microbes are present in a sample through genetic analysis. A heatmap is the accepted way of displaying the data for such a project.

3. METHODS

Several different factors impacted the development of the program. The data input file format, the importance of ease of use of the program, and the existence of any libraries that would aid in the development were all important.

3.1 Data

The data sets that would be used as input to the program were stored in CSV files. They contained the information about bacteria found on two different populations studied by one of the authors. The first set consisted of ninety samples taken from either salamanders or environmental controls, and the second set consisted of forty samples taken from either dead bodies or environmental controls.

3.2 Interface Design

When designing the user interface, we felt it was critical that a person with no programming experience could easily use it without investing more than a couple minutes to learn to use the tool. We also wanted to make the tool prevent the user from being able to make mistakes, to make it even easier to use. To ensure we accomplished these goals, we spent a considerable amount of time discussing the ways the user interface could be implemented and making sure it prevented the user from making errors. We made mockups of different possible interface designs and discussed the benefits and drawbacks of them. Ultimately, we chose to implement the graphical user interface as the common user interface pattern known as “task wizard.” In a task wizard interface, the user answers a small number of questions on each screen, clicking a “next” button or “back” button to navigate between screens. On the final screen, the user clicks a “finish” button to complete the task. Since most computer users have

installed software using one of these task wizard interfaces, and since we could prevent errors using this kind of interface, we designed the program’s user interface as a task wizard. Figures 2, 3, and 4 show the screens of our user interface.

3.3 Implementation Details

To create the GUI, we used Tkinter, a module in the Python standard library which serves as a GUI toolkit. Tkinter is centered around event driven programming, which means the program responds to actions the user takes such as clicking on buttons or selecting an option from a dropdown menu [3]. We used a Python plotting library, Matplotlib, and its extension NumPy [4,5]. Both provide an object-oriented API for embedding plots into applications with Tkinter. To create heatmaps, we used seaborn, a Python data visualization library based on Matplotlib [6]. We also used another software library, pandas, which is used for data manipulation and analysis and provides data structures and operations for manipulating numerical tables of data such as the OTU occurrence data sets [7].

4. RESULTS

The GUI has two screens, a setup screen shown in Figure 2 and a heatmap screen shown in Figures 3 and 4.

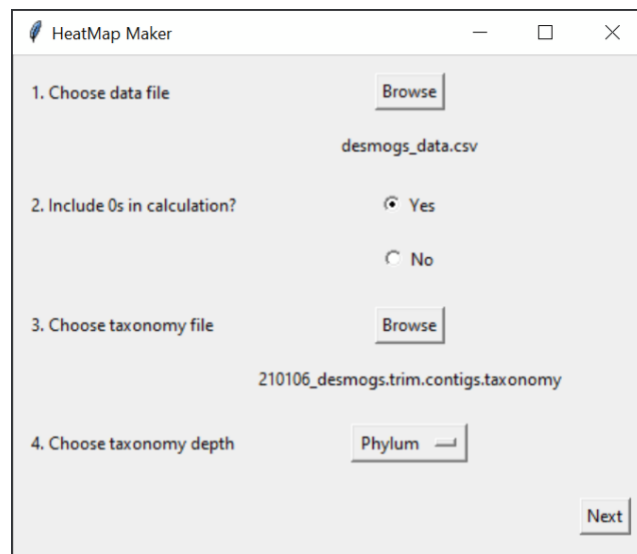


Figure 2. The setup screen.

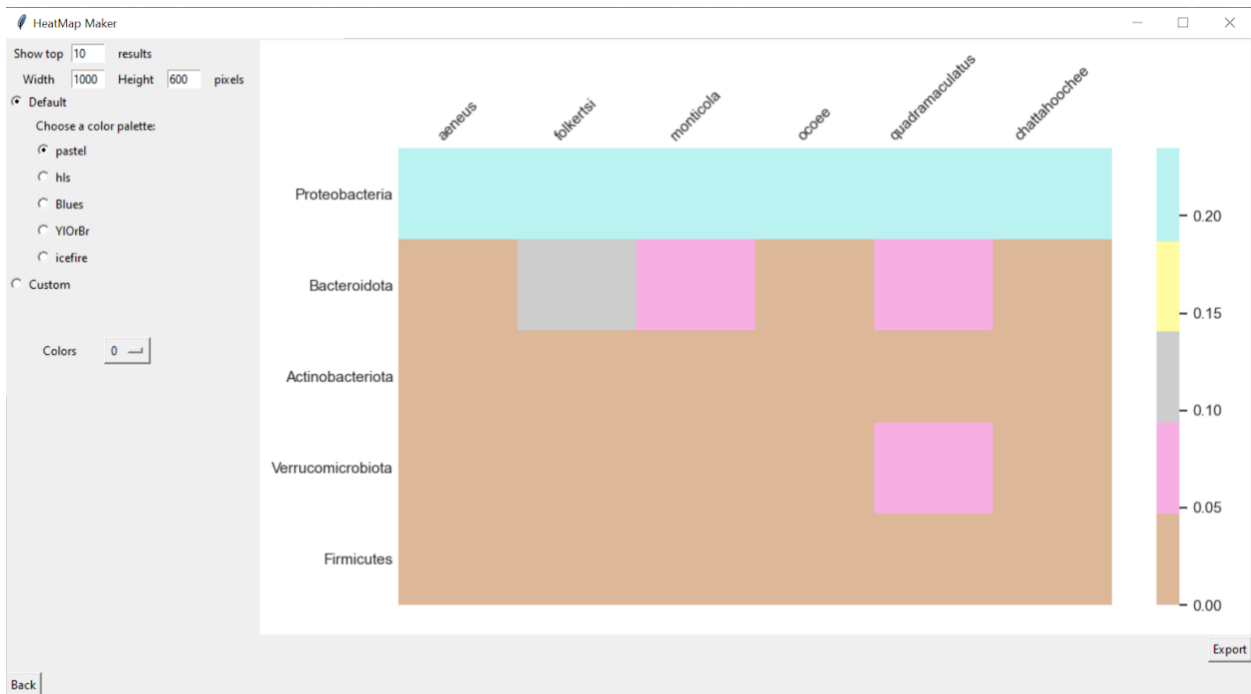


Figure 3. The heatmap screen with the default color scheme selected.

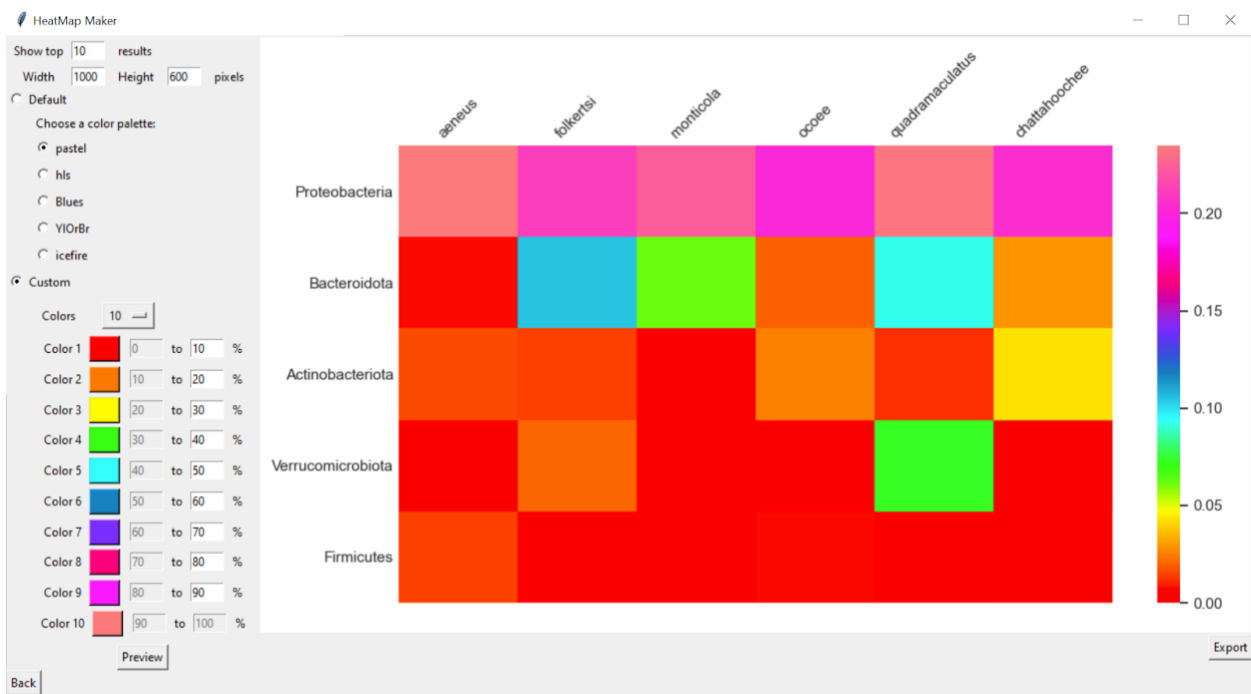


Figure 4. The heatmap screen with the custom color scheme selected.

4.1 Setup Screen

On the setup screen, the user selects two input files, a data file, and a taxonomy file. The program forces the user to select the two required files before allowing them to continue to the Heatmap screen. The user may choose whether to include zeros in the category averages as well as choose the taxonomy depth represented in the heatmap. If an unavailable taxonomy depth is chosen, an error message will appear, and the user must select a new depth before continuing.

4.2 Heatmap Screen

The heatmap screen is displayed after the user clicks the Next button on the setup screen and the calculations have finished. The user may return to the setup screen by hitting the Back button at any point. On the heatmap screen, the user can choose how many results (rows) they want to be displayed on the heatmap. If the maximum number of results is less than what the user chooses, no extra rows will be added. The rows are sorted in descending order by an average of the row's values. The user can also determine the width and height of the heatmap in pixels. There are two color options, default and custom. With the default option, the user can choose one of the default color schemes provided by seaborn [6]. With the custom option, the user can choose from four to ten colors to create their own color scheme as well as designate each color a percentage range. The user must choose all colors and enter continuously ascending range values before clicking the Preview button, or an error message will appear. Once all input fields are filled, the percentage of each color will be scaled correspondingly to the data, and the custom heatmap will be drawn. Clicking the Preview button will automatically change the plot option from Default to Custom, and selecting any of the default color schemes will change the plot option from Custom to Default. When the user is satisfied with the colors they have chosen, they can click the Export button to export the heatmap in either PNG or PDF format. To quit the program, the user can click the X button in the window's title bar and a prompt will ask them to confirm if they want to quit the program.

5. FUTURE WORK

There are several possible opportunities for further development of this GUI. Currently, the program requires the input files to be in a specific format. The first thing we could work on is to generalize the program to accept more input file formats, enabling the GUI to a wider scope of use. We could also allow for greater customization of what in the data file is being analyzed, such as letting the user choose the column in the input file by which the heatmap will be categorized. Since we made this GUI to make biological heatmaps, we could also include the option to italicize labels that are species names.

6. REFLECTIONS

This is the first time we got to do a real-world project that is both enriching and helpful. We had to learn how to facilitate cross-disciplinary communication and collaboration between biologists and computer scientists. We are fortunate that one of the authors on our team had prior experience working with bioinformatics to help bridge the knowledge gap. We had former experience with Python in class, but this project provided us a chance to apply what we learned, further familiarizing us with the language and opening up our programming possibilities. We also had to learn tools for collaboration and version control such as Git and GitHub. To create a GUI, we had to plan ahead to really get a handle on our code. We got to experience the software development process like real software developers. We first had a mock-up for the GUI then proceeded to prototyping and programming it. We had several meetings for feedback with our "client" along the way. Since we carefully planned our design before getting into coding, adjusting the software was not too challenging. As we made the GUI, it was important to think from a user perspective, because something that may make sense to implement as a programmer may not make sense to a user. All in-class assignments have rubrics and well-defined goals, so this is also our first time working with something where the goals for this project were fluid and open-ended. The team experience we had was invaluable since we could have diverse approaches to solve the problem and work together to achieve a product.

7. REFERENCES

- [1] HEAT MAP English Definition and Meaning | Lexico.com. Retrieved from https://www.lexico.com/en/definition/heat_map
- [2] Schwabish, J. 2021. *Better Data Visualizations*. Columbia University Press, New York.
- [3] Introduction to GUI Programming with tkinter — Object-Oriented Programming in Python 1 Documentation. Retrieved from https://python-textbok.readthedocs.io/en/1.0/Introduction_to_GUI_Programming.html
- [4] Matplotlib - Wikipedia. Retrieved from <https://en.wikipedia.org/wiki/Matplotlib>
- [5] NumPy. Retrieved from <https://numpy.org/>
- [6] seaborn.heatmap — seaborn 0.11.1 documentation. Retrieved from <https://seaborn.pydata.org/generated/seaborn.heatmap.htm>
- [7] Pandas (software) - Wikipedia. Retrieved from [https://en.wikipedia.org/wiki/Pandas_\(software\)](https://en.wikipedia.org/wiki/Pandas_(software))