

Neuronal Regeneration Project Proposal

Augustus Lee

Ajay Rao

Harris Lussenhop

ABSTRACT

In collaboration with Northeastern University's NeuroLab, headed by Professor Samuel Chung, this project aims to create effective visualizations displaying neuronal regeneration data in *Caenorhabditis elegans* (*C. elegans*). NeuroLab observed a novel form of neuronal regeneration in *C. elegans* that is independent of common known neuronal regeneration signaling factors. Through the use of an interactive web-based visualization, the goal of this project is to effectively communicate the novel neuronal regeneration data to other researchers and to elucidate key data trends found across the various experimental conditions/groups. The visualization consists of stacked bar charts and dot plots to display the neuronal regeneration data. The linking of the charts enables users to extract key information when examining specific data. Together, the dot plots and the stacked bar charts enable the user to compare novel neuronal regeneration data across the different experimental conditions/groups and to interpret any newly-evident significant data findings.

Github Pages: [here](#)

1 INTRODUCTION

When studying biological phenomena, effective means of communicating research findings between other researchers and to the general public is essential toward making progress in that particular field of study. Data figures and charts are the primary method of how researchers share and communicate their experimental data. Finding methods to create effective visualizations of experimental data is crucial for researchers to collaboratively make progression in any field of science. Northeastern University's NeuroLab, headed by Professor Samuel Chung, is looking for ways to effectively convey its experimental data to progress the field of neuronal regeneration. Data collected on neuronal regeneration varies but may contain information pertaining to axonal length measurements or axonal percent regeneration. NeuroLab is looking to visualize some of its complex data in an interactive web-based visualization to reduce the visual cluttering associated with complex static 2D figures. More importantly, NeuroLab is looking for ways to visualize its data in a dynamic manner that is able to convey more information with greater clarity compared to how biological research information is traditionally represented in 2D static figures. The goal of this project is to approach the neuronal regeneration data from a non-biological perspective and instead approach it from a data scientist/visualization perspective. The project aims to find powerful ways to display the neuronal regeneration data in a way that enables and improves the transmission of the data findings to the other scholars in the field and to ultimately help progress the field of neuroscience.

Interactive and linked dot plots and stacked bar charts were chosen as the most effective charts to visualize the neuronal regeneration data. Two-way linking between the dot plots and the stacked bar charts allows for users to extract more information from each data point compared to a industry-standard static visualization. In addition, the implemented details-on-demand interactive feature reduces general visual clutter by allowing the user to dynamically

and selectively extract pertinent information by choice.

Github Repo: [here](#)

Final Presentation Slides: [here](#)

Project Video: [here](#)

2 RELATED WORK

Figures from [8] show how neuronal activity from different electrodes can be visualized at once using unique color encoding.

[6] provides a good example of how to create a high-level overview visualization of a project. We can make such a visualization interactive via D3.

[9] shows how neuronal area can be effectively mapped using a line chart against some other variable.

[5] provides an example for how neuron complexity over time can be displayed in three plots under three different conditions. For our project, an interesting idea would be to think about ways to display similar data from various experimental conditions on one graph while avoiding unnecessary visual clutter.

Figures from [3] provide an example of effective 3-d visualizations to show the direction of growth of neuron cells.

[11] shows effective uses of diagrams and line charts to show how different nerves recover from surgery over time.

[2] effectively uses colored bar charts to display the density of different neuron cells.

[7] uses groups of neuron histograms to display different responses to a stimulus.

[1] shows how iron-oxide labeled neural stem cells are detected using magnetic resonance imaging. Image histograms combined with diagrams provide an effective visualization.

[10] models the transmembrane potential of Type-1 Spiral Ganglion Neurons. It effectively uses colored maps to show spatial heterogeneity of tissues

As the final visualization was developed, our plan had shifted from focusing on task 3 to tasks 1 and 2. As a result, our visualization had significantly changed from our initial sketches to our final sketches. For the final visualization, elements of some visualization design choices from the related works were used. Particularly, [2] uses colored bar charts to represent quantitative neuronal data and we utilized stacked colored bar charts in our visualization for a similar purpose.

Compared to previous works, our project is related because it serves to function as a method of conveying complex biological data to its users. Using interactivity, our visualization is able to reduce the visual clutter of complex data and makes data details easily accessible to the user. Most related works on this topic utilize static visualizations for data reporting but our visualization uses interactivity to allow users to explore and extract pertinent data based on the user's data collection goals.

3 PARTNER

Northeastern University's NeuroLab seeks to investigate possible avenues for treating central nervous system (CNS) injuries and neurodegenerative diseases. Particularly, the NeuroLab focuses on CNS

axon regeneration, which is typically the fundamental barrier to the poor recovery observed with CNS or neurodegenerative pathologies. Professor Samuel Chung is an Assistant Professor of Bioengineering at Northeastern University who leads the NeuroLab's neuronal regeneration research efforts. Professor Chung specializes in research involving neuronal regeneration observed in genetic studies involving the nematode *C. elegans* (popular genetic model organism). Professor Chung and NeuroLab are looking for ways to transform some of its data into interactive web-based visualizations to improve the clarity of conveying his experimental findings.

The interview with the partner was informative and eye-opening. The partner explained that neuronal regeneration studies rely on visualizations that allow for visual comparisons between different experimental conditions/groups. Thus, visualizations that allow for easy data comparisons are preferred by the partner. The partner also explained that a common issue with biological research figures is that the complex data can often visually clutter the visualization. The partner expressed a desire to have a visualization with details-on-demand functionality to hide data details unless they are interactively requested by the user. Apart from visualization expectations, the partner helped explain the context behind the neuronal regeneration data. By explaining the significance of the data, the partner enabled us to plan and think of how to make an appropriate visualization with any necessary interactivity features. The partner also explained that the target audience for the visualization is other researchers, which helped influence how we would convey the complex data to our visualization users.

4 DATA

The data used in this project comes directly from neuronal regeneration [findings](#) outlined in [4]. Note The data is primarily quantitative and pertains specifically to neuronal regeneration, such as axon length measurements or percentages of neuronal growth over time. Categorical data pertaining to regeneration type was also provided. Data is readily available in Excel formats.

5 TASK ANALYSIS

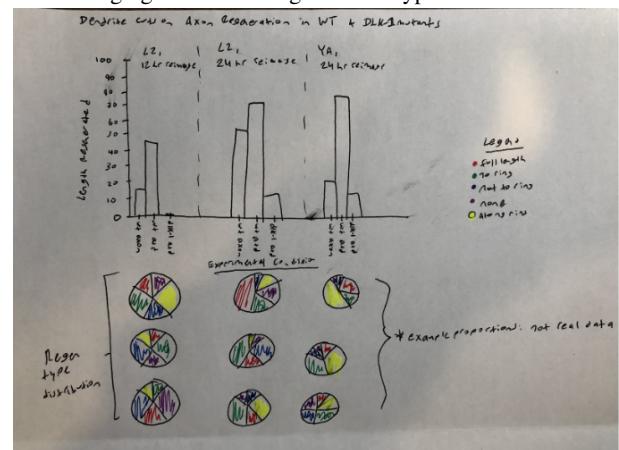
After careful review of the project tasks, it is evident that the primary type of consumption for this visualization project is an exploratory visualization. When exploring novel biological phenomena fueled by an existing hypothesis, the primary goal is to uncover knowledge that is not previously known. Patterns among quantitative neuronal regeneration data are not evident until visualized, and the resulting data patterns/trends observed from the visualizations help researchers verify or disconfirm their hypotheses. The partner is unclear of what data trends to expect from visualizing the quantitative neuronal regeneration data, so the primary goal of this visualization is to explore and discover new insights into the data.

The primary consumer for this visualization will be other scientific researchers, especially those studying the field of neurobiology. The ultimate goal of the visualization is to convey the dense neuronal regeneration data in a clear and effective manner to help educate and inform the research efforts of other scientists in similar fields. See Table 1 for the high level goals of the user.

6 EXECUTION & DESIGN PROCESS

Preliminary Sketch 1 displays the initial attempt to visualize the neuronal regeneration data. The visualization depicts a bar chart for the neuronal regeneration measurements and corresponding pie charts to represent the distribution of regeneration type for each group's experimental condition. A bar chart was chosen for its ability to compare quantitative data and a pie chart was chosen for its strength in comparing relative percentage values. Since the quantitative neuronal regeneration data was organized by each worm, the bar chart would have visualized an aggregation of the regeneration data (i.e. sum, mean, etc.). The legend displays a qualitative color map

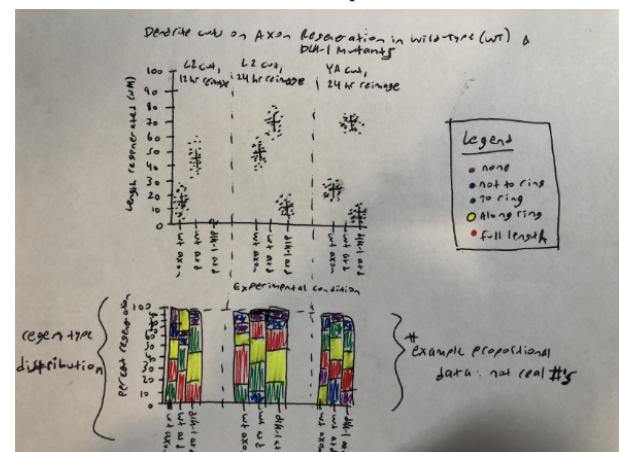
with five distinct colors to represent the five different regeneration types. The colors only differed in hue to establish clear differences in data belonging to different regeneration types.



Preliminary Sketch 1

After consultation with other research papers and scientists, we realized that pie charts were not the best option to represent the distribution of categorical regeneration types. While pie charts are strong at showing the relative contribution of data compared to the whole, they are not as strong as stacked bar charts in performing the same function. Pie charts require significant screen space compared to stacked bar charts and the angle channel in a pie chart is not as precise as the length channel in a stacked bar chart. Given the weaknesses of pie charts for our visualization purposes, stacked bar charts were used for encoding the distribution of regeneration types.

The use of a bar chart to represent the aggregate neuronal regeneration data was also determined to be a weak visualization option. After consultation with the project partner and with the research paper the data is derived from, the use of displaying an aggregation of the neuronal regeneration data was determined to not be a powerful tool for researchers. The individual neuronal regeneration data from each worm was important to be independently plotted because the heterogeneity of the neuronal regeneration data was a key feature that was essential for comparative data analysis. Dot plots were chosen as the ideal plots to represent the heterogeneity of the neuronal regeneration data because of their ability to represent individual data points in a way that enables easy visual comparisons between adjacent plots. Additionally, aggregation data, such as the mean, could also be overlaid on a dot plot if desired.



Preliminary Sketch 2

Task ID #	Domain Task	Analytic Task (low-level, "query")	Search Task (mid-level)	Analyze Task (high-level)
1	Examining regeneration effects of dendrite cuts on wild-type and DLK-1 mutant worms over time. How do dendritic lesions affect axon regeneration in DLK-1 mutant worms over time?	Compare	Explore	Discover
2	Analyzing distribution of regeneration patterns over time. How does the distribution of regeneration types change over time?	Compare	Explore	Discover
3 *	Examining quantitative neuronal regeneration data across gene mutants. Which gene mutant condition allows for the best (most) DLK-1-independent regeneration?	Compare	Explore	Discover

Table 1: Task Analysis

* chose to omit due to the shift in visualization focus towards Tasks 1 and 2. The goals of Task 3 did not align with the current visualization model

Preliminary Sketch 2 contains many of the design changes discussed after Preliminary Sketch 1 and is most similar to the current final visualization. The sketch depicts dot plots for representing the neuronal regeneration data and stacked bar charts representing the distribution of regeneration type for each experimental condition. Thin bars representing the mean and standard deviation (2 std's) of the neuronal regeneration data were overlaid on the dot plots to display useful aggregate data while also allowing users to observe the heterogeneous data points beneath them. The same color encoding from Preliminary Sketch 1 was used given its strength at differentiating between the five distinct regeneration types.

Preliminary Sketch 2 was heavily influential in the production of the final visualization. The final visualization improved data understandability from Preliminary Sketch 2 by employing interactive features between the dot plots and stacked bar charts, such as brushing, linking, and details-on-demand.

When the virtual visualization was initially created with all its interactive features, the visualization was subjected to usability testing from users with little biological backgrounds. Feedback from usability testing revealed that users had little complaints about the visualization types used but had difficulty with contextualizing the data being presented. Despite these users not being our target audience, we still realized that our visualization could include more functionality to help the user contextualize the data. A details-on-demand feature was implemented as a tooltip for the three experimental conditions used in the visualization: wt axon, wt a+d, dlk-1 a+d. Upon mousing over these three experimental conditions on the X axes of the dot plots, more information about each experimental condition was provided. The goal of this tooltip was to aid in data understandability by elaborating on data encoded by confusing/complex abbreviations (sometimes even researchers may be confused about abbreviations too!).

7 VISUALIZATION DESIGN

Significant differences between the Digital sketches and the current visualization were a result of design changes that occurred naturally during the creation of the visualization. The Digital sketches included an “Overview/Compare” functionality that would allow for direct comparison between selected data groups. The purpose of the “Overview/Compare” functionality was to assist with data comparisons between experimental groups but the current visualization accomplishes this goal through its brushing and linking functionality. As a result, the “Overview/Compare” feature was not essential for a viewer’s ability to compare experimental group data. Another significant change between the Digital sketch and the current visualization was the omission of a second larger stacked bar chart that would be visible after interaction with the main stacked bar charts. Upon selecting a bar of interest from the main stacked bar charts, a large version of the selected barchart would have appeared to allow for an

easier visual breakdown of the data differences within the selected bar chart. This feature was not included in the current visualization since the underlying purpose of the larger stacked bar charts was to illuminate data that may have been obscured by the small size of the original stacked bar charts. However, the current visualization largely scales the main stacked bar charts, clearly displaying all the data and rendering the inclusion of a second larger stacked bar chart as unnecessary. The omission of the second larger stacked bar chart also freed up screen space that allowed for larger dot plots and stacked barcharts, which aids in viewer data interpretation.

Any other minor interactivity features specified in the digital sketches that were not included in the final visualization was intentional because those features did not aid in user data interpretation/analysis.

The current visualization contains linked stacked bar charts and dot plots. The dot plots display the length of axon regeneration given three experimental conditions (wt axon, wt a+d, DLK-1 a+d) for three different cohort groups (L2 cut 12 hr reimage, L2 cut 24 hr reimage, Young adult cut, 24 hr reimage) for a total of nine plots. L2 and Young Adult refer to the larval stage of the C. Elegans worms. The reimage time describes how long after surgery the particular worm was evaluated for data collection. The three experimental conditions describe the genetics and type of regeneration stimulation used on each worm. “Wt” refers to normal, wild-type worms, while “dlk-1” refers to worms that do not have the DLK-1 gene (typically associated with traditional neuronal regeneration). “Axon” refers to the neuronal axon being cut to stimulate regeneration while “a+d” refers to the neuronal axon and the dendrite being cut to stimulate regeneration.

Stacked bar charts are organized below each of the nine dot plots with data displaying the distribution of regeneration types that correspond to the dots in the respective dot plot. When a dot in a dot plot is brushed over, the bar corresponding to that dot’s regeneration type is highlighted. If multiple dots in a dot plot are brushed over, all the respective bars corresponding to the selected dots are highlighted. The linking between the dots in the dot plots and their corresponding regeneration type in the stacked bar charts persists even when selecting dots across different dot plots (i.e. selecting dots in L2 cut, 12 hr reimage and dots in L2 cut, 24 hr reimage). In addition, when dots are selected via brushing, the dots are colored corresponding to their respective regeneration type in accordance with the color scheme in the legend and in the stacked bar charts. The dots are normally filled black if not brushed to allow the user to see the overall data points without the confusion that may be experienced if the user sees many colorful dots in a small viewing area.

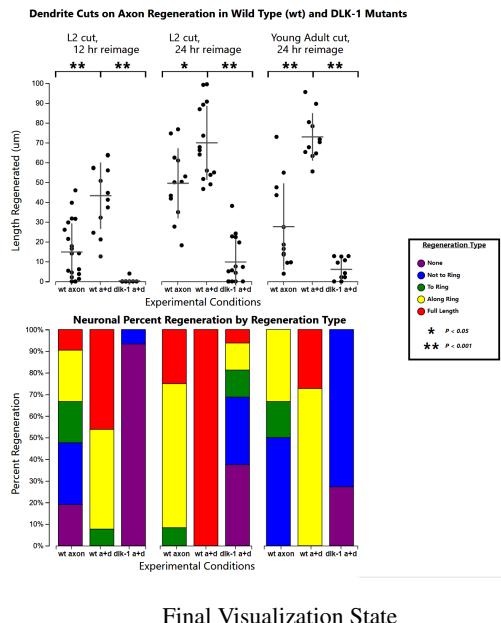
When a bar within the stacked bar chart is moused over, the bar becomes highlighted and utilizes a detail-on-demand feature: displaying the relative percentage of the selected bar’s regeneration

type in comparison to the other regeneration types found within the dots that correspond with the selected bar. In addition, the bar that is moused over also colors in the dots that correspond to the selected bar's regeneration type. For example, if a red bar within a stacked bar chart is moused over, the percentage of that regeneration type in relation to all the regeneration types in that group is visible and all the dots that correspond to the red bar are filled in red.

The linking between the stacked bar charts and the dot plots is intended to provide the user with dynamic feedback when exploring the data and should aid the user in analyzing and comparing the data with useful contextual information (i.e. the dot plots data along with the stacked bar charts data should allow for contextualized data results that aid in data interpretation/analysis).

The dot plots contain a tool-tip functionality for the three types of experimental conditions (wt axon, wt a+d, DLK-1 a+d). When the experimental condition is moused over, the tooltip displays a text box that contains more details about the selected experimental condition to aid the user in interpreting the data. The purpose of the tooltip is to reduce any confusion about the contextual significance of the dot plots data. The tooltip is not included for the stacked bar charts since they contain the same experimental groups as the dot plots (repeating the same information is unnecessary).

The legend is positioned to the right of both the stacked bar charts and the dot plots to provide the user an easy reference point when interpreting the color encodings found throughout the visualization. Throughout the visualization, five different colors were used: purple, blue, green, yellow, red. All colors had the same saturation and brightness, only differing in hue. These colors were selected to differentiate the types of regeneration (categorical data) because their stark contrast with each other reduces any ambiguity when a user attempts to visually gauge the regeneration type of a dot or bar. All the colors were given the same saturation and brightness to eliminate any potential data importance/magnitude interpretations of the categorical data. The legend also includes information about the statistical tests performed (denoted by * or **) that are present in the dot plots.



8 DISCUSSION

After completing this project, key information about the neuronal regeneration data became apparent. From the dot plots and stacked bar charts, it was observed that the wild-type axon + dendrite cut experimental condition within the L2 cut, 24 hr reimage group possessed the highest mean regeneration, and all the worms in that

group exhibited full-length regeneration. Notably, the worms in the DLK-1 axon + dendrite cut experimental condition within the same L2 cut, 24hr reimage group exhibited lower mean regeneration than that of the wild-type axon + dendrite cut but the distribution of regeneration type within the experimental condition consisted of all five regeneration types. Such data observations and trends became clearly evident when the data was viewed using the created visualization.

From a visualization creation standpoint, working on this project enforced the importance of contextualized data. Without a clear understanding of the meaning and significance of the data of interest, the appropriate visualization would be impossible to create. Since the quantitative regeneration data needed to be represented in a way to preserve the heterogeneity of the data points, dot plots were found to be an effective and appropriate visualization type. Since relative percentages of categorical variables (i.e. regeneration type) needed to be represented in a way that enables comparative analysis, stacked bar charts were identified as a strong visualization choice.

Another key takeaway from completing the project includes an understanding about the importance of interactivity features in visualizations. When used improperly, interactivity features can be distracting to the user and can reduce the effectiveness of the visualization. However, when used properly, interactivity features can reduce visual clutter and improve user understanding of the data (compared to a static visualization).

An important lesson from this visualization project is to establish clear communication between all members and with the partner. Creating an effective visualization is difficult if there is miscommunication among the different groups of involved individuals.

9 CONCLUSION

Using quantitative and qualitative neuronal regeneration data, a visualization consisting of linked, interactive dot plots and stacked bar charts was created. The dot plots convey quantitative neuronal regeneration length data and the corresponding stacked bar charts display the distribution of regeneration type (categorical data) for the data points in the respective dot plots. Interactive features, such as brushing, linking, and details-on-demand, were utilized across the dot plots and stacked bar charts. A qualitative color map was used to encode the different types of neuronal regeneration in both the dot plots and the stacked bar charts. The purpose of the visualization was to convey the complex neuronal regeneration data to other scientists (target audience) and to elucidate any key data observations/trends.

An area for improvement could be to include more details-on-demand when brushing. In addition to highlighting the stacked bars when the dots are brushed, the bar percentage could also be shown instead of it only being shown when mousing over the stacked bar chart.

Future directions could be to modify the code to allow for easily visualizing new regeneration data. Some data from the csv files were manipulated to create the visualization, which slightly hinders the ability of the visualization to be compatible with new data that is not similarly formatted. Further optimizing the code to allow for easily visualizing differently-formatted neuronal regeneration data would allow for this visualization to be a useful and robust tool for many neuroscientists looking to visualize their data.

10 ACKNOWLEDGEMENTS

Special thanks to Professor Samuel Chung and Noa Grooms for supplying the data used in the visualization. Professor Chung and Noa Grooms also assisted with data interpretation and analysis.

Thanks to Professor Cody Dunne and David Saffo for their help with visualization theory advice and with the practical implementation of the visualization.

REFERENCES

- [1] M. Baghchechi, A. Obenaus, S. Ashwal, M. Hamer, A. Plaia, and N. Ghosh. Hierarchical region splitting detects iron-oxide labeled neural stem cells in susceptibility weighted magnetic resonance imaging. In *2016 International Conference on Systems in Medicine and Biology (ICSMB)*, pp. 74–77, 2016. doi: 10.1109/ICSMB.2016.7915091.
- [2] Z. Cao and W. He. Neuronal responses to simply prepared chitosan composite gels. In *2010 Biomedical Sciences and Engineering Conference*, pp. 1–4, 2010. doi: 10.1109/BSEC.2010.5510825.
- [3] X. Che, J. Boldrey, X. Zhong, I. Schneider, D. Jiles, and L. Que. Studies of single neuron cell growth on nanoporous surface and under transcranial magnetic stimulation. In *2017 IEEE 17th International Conference on Nanotechnology (IEEE-NANO)*, pp. 1056–1059, 2017. doi: 10.1109/NANO.2017.8117495.
- [4] S. H. Chung, M. R. Awal, J. Shay, M. M. McLoed, E. Mazur, and C. V. Gabel. Novel dlk-independent neuronal regeneration in *caenorhabditis elegans* shares links with activity-dependent ectopic outgrowth. *Proceedings of the National Academy of Sciences*, 113(20):E2852–E2860, 2016. doi: 10.1073/pnas.1600564113.
- [5] K. P. Lam, K. P. Dempsy, W. A. Smith, K. T. Wright, W. E. Masri, and J. B. Richardson. A computational approach to quantifying axon regeneration in the presence of mesenchymal stem cells (mscs). In *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, pp. 1541–1544, 2013. doi: 10.1109/NER.2013.6696240.
- [6] N. Mokarram and R. V. Bellamkonda. Overcoming endogenous constraints on neuronal regeneration. *IEEE Transactions on Biomedical Engineering*, 58(7):1900–1906, 2011. doi: 10.1109/TBME.2010.2103075.
- [7] F. Panetsos, V. Bonacasa, and V. A. Makarov. Neural activity in the lower pathway of the somatosensory system in the presence of silicon interfaces. In *First International IEEE EMBS Conference on Neural Engineering, 2003. Conference Proceedings.*, pp. 211–213, 2003. doi: 10.1109/CNE.2003.1196795.
- [8] K. Shimba, K. Sakai, K. Kotani, T. Yagi, and Y. Jimbo. Microtunnel-electrode device for elucidating axon features: Toward pharmacological manipulation of individual axons. In *2017 8th International IEEE/EMBS Conference on Neural Engineering (NER)*, pp. 267–270, 2017. doi: 10.1109/NER.2017.8008342.
- [9] B. F. Sisken. Nerve limb regeneration. *IEEE Engineering in Medicine and Biology Magazine*, 2(4):32–39, 1983. doi: 10.1109/EMMB.1983.5005989.
- [10] K. K. Sriperumbudur, H. W. Pau, and U. van Rienen. Effect of tissue heterogeneity on the transmembrane potential of type-1 spiral ganglion neurons: A simulation study. *IEEE Transactions on Biomedical Engineering*, 65(3):658–668, 2018. doi: 10.1109/TBME.2017.2700361.
- [11] N. Zhang, S. Wang, X. Chen, Q. Shi, J. Li, J. Zhu, S. Wang, B. Yang, Y. Guo, and J. Zhu. Study on neural regeneration effect of rat by using pulsed functional magnetic stimulation. *IEEE Transactions on Magnetics*, 51(11):1–4, 2015. doi: 10.1109/TMAG.2015.2437997.

11 GROUP CHARTER

11.1 Group Purpose

This group was formed to assist with effectively visualizing a novel dataset collected from Northeastern University's NeuroLab (led by Professor Samuel Chung) involving neuronal regeneration in *C. elegans* models. The group's purpose is to convey critical biological information in a way that maximizes interpretation and readability for scholars with both science and possibly those with non-science backgrounds. The intended users of our visualizations would be other researchers hoping to understand the experimental findings from Professor Samuel Chung's research. In addition, we hope that the other researchers will be inspired by the visualization and will

look to utilize an interactive web-based display to convey their biological research findings (not just limited to neurological data) in a digestible format to allow other scientific and possibly non-scientific academics to effectively interpret their data. The expectations of our partner for this project are to improve upon existing published figures that exist for the data and/or find new, creative ways to visualize the same data.

11.2 Group Goals

The group's project goal is to produce an effective visualization that can aid our partner's research endeavors and help convey the novel biological findings observed by the NeuroLab research team. The group hopes to contribute to the neuroscience field by finding novel or other effective means of conveying neurological data. The group's process goal will be to meet weekly and consistently complete assigned action items. Steady, consistent work will be enforced to ensure constant project progress throughout the semester. The group aspires to complete work at the highest quality possible. All group members are able to commit enough time and energy to producing a visualization of high quality. All group members are collectively aiming for a course grade of "A".

11.3 Group Roles and Responsibilities

Given the collaborative nature of the group project, specific group member roles must be established in order to make efficient project progress throughout the semester. Ajay will take charge of working with the partner and interpreting the biological data in a way to effectively communicate the data significance to the rest of the team (given his academic background in biology). He will convey any group confusions to the partner about data meaning/interpretation (via the communications director). Harris will be the group's designated communications director and will handle communications between the group and the professor and the partner. He will also focus on technical aspects of the project such as working with JavaScript and D3 for the web page as specified by the requirements of the project. Augustus will be the meeting facilitator. He will take charge of ensuring that meetings follow the general structure specified in the "Ground Rules". During the meeting, questions from each group member will be answered, and by the end of the meeting, each member must have clear action items.

11.4 Ground Rules

To ensure that project work proceeds smoothly throughout the semester, certain ground rules must be set in place. The group will have weekly meetings on Wednesdays from 12:00 - 1:00pm (timing is flexible week-to-week depending on group member schedules). Meetings will follow a general structure: 1) Brief review of prior weeks discussion, 2) Individual member report on progress made from the prior meeting, 3) Open QA among group members pertaining to any relevant topics about the project, 4) Set of action items assigned to each member for completion before the next meeting.

Each member will be expected to complete their tasks in a timely fashion. If they are busy, they are expected to plan ahead accordingly and communicate issues with the group at least 2 days before the upcoming meeting. If group members consistently fail to complete their weekly action items, the professor and/or TAs will be notified of the member's apparent lack of project contribution. Group members are all expected to devote considerable time towards executing this project. The time commitment of the project may fluctuate weekly depending on the immediate required work, but group members should be willing to adapt to each different weekly situation.

During meetings, discussions will be held to share information between group members. Discussions will be an open space where any group member can freely share their opinions. If there are any dissenting views, the remaining group members will respectfully listen to those views and all members will work together to find a compromise. In order for the group to make a decision, all three members must be in agreement or agree to a compromise. If no agreements or compromises can be made, the professor and/or TAs will be notified for external assistance.

11.5 Potential Barriers and Coping Strategies

Potential barriers to group work may arise due to confusions with the biological logic of the data. Not all group members have taken biology-related courses so their understanding of key biological concepts may be limited. If group members cannot understand the biological logic behind visualization ideas proposed by the project partner, then expanding upon the partner's ideas in an effective manner may be cumbersome and unproductive. To avoid group work barriers centered around understanding the provided data, all group members will be expected to read all pertinent literature about the particular topic. Each group member will read neurological publications centered around neuronal regeneration in *C.elegans* models in an effort to better contextualize the neuronal regeneration data that will be provided to the group. In addition, frequent meetings with the project partner will be scheduled to address any key data comprehension questions that any group member may have when reading relevant literature. Following the aforementioned guidelines should avoid potential barriers to group work, but issues may still arise if group members do not complete their literature readings. If group members fail to complete any of the required readings to improve that individual's understanding of the data, then guidelines found under the "Ground Rules" section (above) will be followed (i.e. may contact professor and/or TA's about group member's lack-luster efforts). Problems that have arisen in the past from poor group work have normally been addressed and resolved with an intra-group confrontation of the offending member. Clear and effective communication will be utilized to avoid any barriers to group work as much as possible.

11.6 Mid-Way Reflection

All group members have been abiding by the agreed-upon rules. Members are adhering well to group roles and project progress has improved steadily. All members are comfortable with their group roles and no issues have arisen. There are no significant issues that need troubleshooting at this time.

12 APPENDIX A: INTERVIEW

NOTES:

- Our approach to the project: what would the professor benefit from us
 - We could potentially use imaging data to improve 1B
 - Figure out how to improve figure 2:
 - a and b
 - Make them look more intuitive, less complicated
 - look at s5 as well from the supplementary figures.
 - ex: hover over and show data
- Improve Figure 3:
 - Find a better way of showing a lot of data.
- Improve Figure 1c,d:
 - Timelapse data (unclear if data is suitable)
 - ex: color code 1d points with colors bars below
 - Hover over colors, show related data on hover, popup of an image or line diagram.
 - Animation of 1b, neural regeneration.
- Data file sent over was bulk of regeneration data
- Ectopic outgrowth:
 - There is a corresponding dataset as well.
 - Corresponding figures included in supplement.
- Fig 2b data included in summary of datafile.
- Some fields in the excel file are calculated.
- 'for n > 100' excel tab used for calculating p values.
- Ectopic outgrowth raw-data-mut corresponds to fig. S5.
- Figure 2b: black bars are a non-mutated DLK gene.
 - white bar: DLK knocked out/DLK Mutant
- Worm websites exist with a lot of related data
- End Users:
 - Involved researchers.
 - Familiar with genes, regeneration, not familiar with Prof. Chungs model
- Web based visualizations are ok.
- Make a tool others can use to visualize worm/neural data?
 - Wormbase.org -> worm gene and protein information
- link gene information to wormbase.com
- Ju476 and km12
- wormatlas.org
 - Used to visualize neurons.
 - Link to whenever a neuron is mentioned.
 - ex: ASJL neuron
- Professor and Noa typically available on thursday

13 APPENDIX B: DATA EXPLORATION

13.1 Data types

The data contains multiple quantitative data columns organized by multiple categorical column labels.

The quantitative data contains numerical/percentage data about neuron ectopic outgrowth and neuron percent regeneration in DLK-1 mutant cells. The categorical data includes each gene mutant and the general experimental condition group each mutant is in. Quantitative data is available for each gene mutant and each gene mutant is grouped by a parent categorical label based on the experimental condition it belongs to.

The data was collected by laboratory research efforts by NeuroLab at Northeastern University led by Professor Samuel Chung. All gene mutants and experimental conditions were tested with NeuroLab personnel.

13.2 Potential issues

The data contains some non-specific column labels, which currently have unknown importance in our analysis. In addition, many columns contain abbreviated labels that non-biologists have difficulty understanding.

13.3 Insights

The data exploration was a difficult and arduous process. The data was formatted within excel files that prioritized human readability rather than machine readability. Processing the data into Jupyter Notebooks or into Tableau required some data wrangling and modification to isolate our data columns of interest.

Before beginning our analysis, we were expecting to see similar regeneration and ectopic outgrowth trends for each mutant that have already been reported in the corresponding research paper. Upon exploration using Tableau and Jupyter Notebook (Plotly library), the expected data trends were observed. The Tax-2 mutant was expected to have the highest ectopic outgrowth percentage for the DLK-1(+) condition and our data results were consistent with that. The control categories were expected to have low percent ectopic outgrowth and neuronal regeneration and our findings were consistent with those expectations.

Some messy data was identified during our data exploration. Regeneration data for several gene mutants were not aligned with their corresponding row, so manual manipulation of row data was necessary for Tableau and Jupyter Notebook (Plotly library) compatibility. Without this manual data wrangling, visualization methods using Tableau and Jupyter Notebook (Plotly library) were not displaying the correct data relationships.

13.4 Screenshots

Ectopic outgrowth of DLK-mediated and DLK-independent Neurons

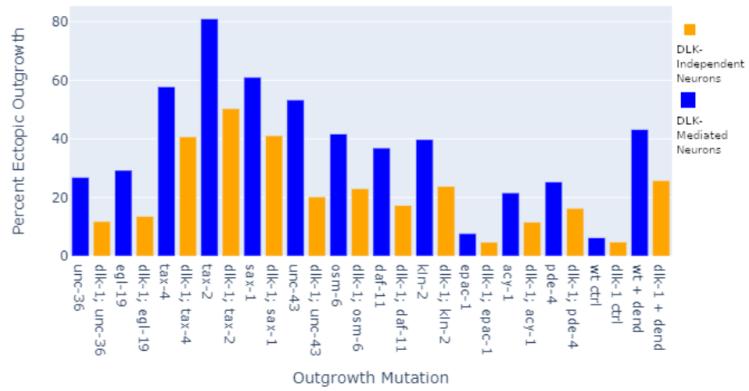


Figure 1: This figure explores the ectopic outgrowth of DLK (DLK-1)-mediated and DLK (DLK-1)-independent neurons, omitting the regeneration column from the data. Bars were used to represent the percent of ectopic outgrowth (quantitative data). The blue color represents DLK-1(+) (DLK-mediated) neurons and the orange bars represent DLK-1(-) (DLK-independent) neurons. Each type of respective neuron is listed on the x-axis. This visualization shows a trend of more ectopic outgrowth taking place in DLK-1-mediated neurons than DLK-1-independent neurons.

DLK-independent Regeneration

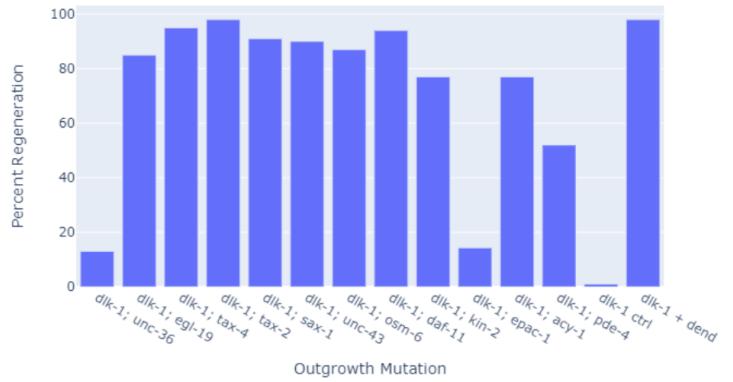


Figure 2: This figure explores percent regeneration in DLK (DLK-1)-independent neurons. This is a subset of the full dataset, omitting the ectopic outgrowth data. Bars were used to represent the percent of neuronal regeneration (quantitative data). Color channels were unnecessary since only one type of quantitative data is visualized here. Each individual neuron is listed on the x-axis as categorical data. This visualization shows that for most DLK-independent neurons, the percent regeneration is near 100%.

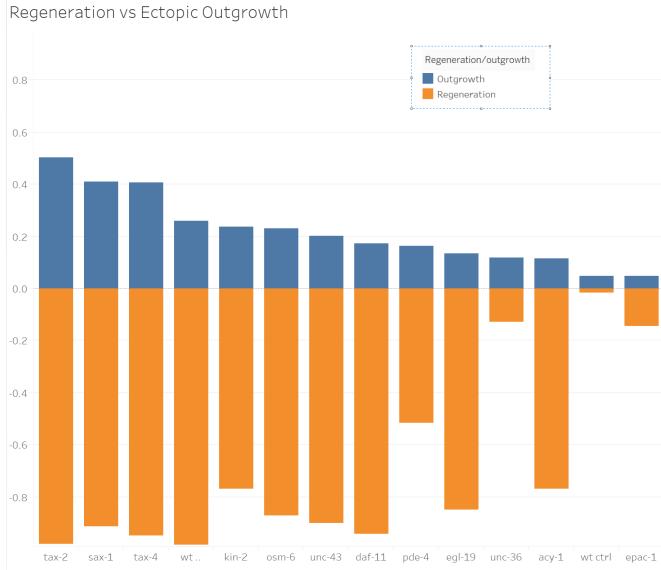


Figure 3: This figure explores the quantitative data on ectopic outgrowth and neuronal regeneration for different outgrowth mutations of *C. elegans*. A two-sided bar chart was used to compare differences in percent ectopic outgrowth and neuronal regeneration magnitudes for each mutant. The unique color encoding differentiates between the ectopic outgrowth and neuronal regeneration data for each mutant. In addition, the colors of the bars are different from the color of the background as well as different from each other to show distinctions between the data. We see that the Tax-2 mutation shows the greatest percent magnitude of ectopic outgrowth, while the wild type group and Tax-2 group show the greatest percent neuronal regeneration. From these observations, we also notice that the Tax-2 mutant has the most percent magnitude of ectopic outgrowth and neuronal regeneration. We also see that the control group exhibits the lowest percent ectopic outgrowth and neuronal regeneration compared to any other mutant type.

14 APPENDIX C:DESIGN SKETCHES

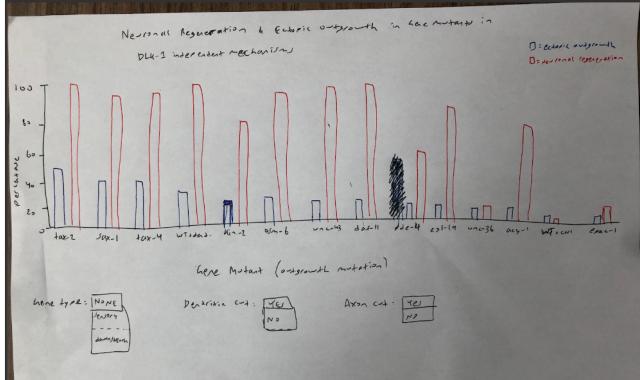


Figure 4: (FAVORITE) SKETCH 1: Illustrated by Ajay Rao. A bar chart was used in this design, with bars being used as the marks. The different types of regeneration (ectopic outgrowth versus neuronal regeneration) were represented in bar marks and were differentiated from each other by distinct color encodings. A bar chart was used for this design because the goal of the visualization was to compare data values and explore the dataset for high and low values. Different color encodings for the bars were used to allow easy visual data magnitude comparisons between bars of the same type (i.e. can compare neuronal regeneration for each mutant group since all the respective bars are the same red color). This design addresses Task 3: Examining quantitative neuronal regeneration data across gene mutants. Which gene mutant condition allows for the best (most) DLK-1-independent regeneration?

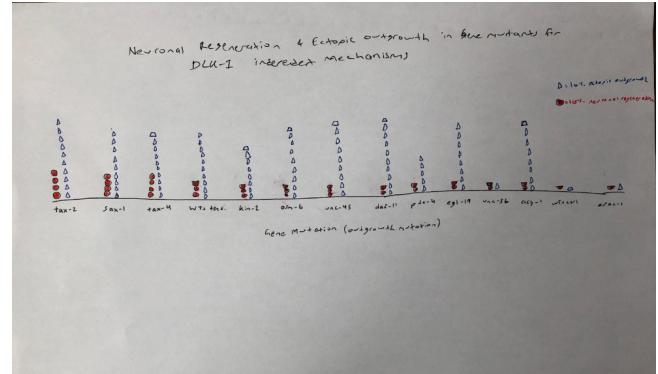


Figure 5: SKETCH 2: Illustrated by Ajay Rao. The design above is a Pictogram graph, with different symbols used as marks. The different types of regeneration (ectopic outgrowth versus neuronal regeneration) were represented with symbols and distinct color encodings. A pictogram graph was used for this design to allow comparison of relative data magnitudes between ectopic outgrowth and neuronal regeneration data. The goal of the visualization was to compare data values and explore the dataset for extreme values. Different symbols for the symbol mark were used to differentiate between the two types of regeneration seen in the graph. In addition, the different symbols have a distinct color encoding to provide easy visual differentiation between the distinct types of regeneration. The use of different symbols and color encodings was intended to allow easy relative visual magnitude comparisons between data for different mutants. This design addresses Task 3: Examining quantitative neuronal regeneration data across gene mutants. Which gene mutant condition allows for the best (most) DLK-1-independent regeneration?

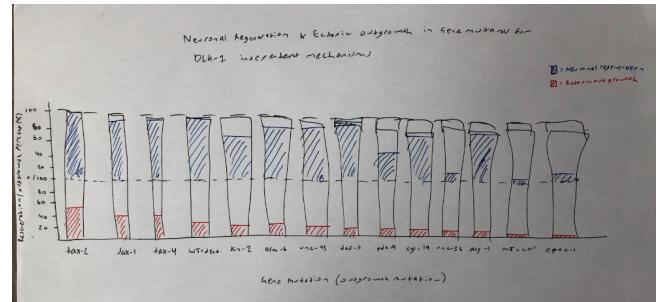


Figure 6: (FAVORITE) SKETCH 3: Illustrated by Ajay Rao. The design above is a modified stacked bar chart, with separate baseline axes for the neuronal regeneration and ectopic outgrowth data. Bars were used as marks to represent the different regeneration data for each mutant. In addition, distinct color encodings were used to specify which bars represent neuronal regeneration or ectopic outgrowth. A stacked bar chart with separate baseline axes for the two datasets was used to allow for easy visual comparison of data magnitudes of neuronal regeneration and ectopic outgrowth across different mutant groups. The different baseline axes for the two datasets were used to enable the viewer to compare the different data magnitudes without allowing the two datasets for a single mutant influence the perception of the data trends (a problem typically associated with stacked bar charts is that stacked data is hard to compare between various groups since the stacked bars have different vertical/horizontal origin positions). Distinct color encodings were used to enable easy visual differentiation and comparisons for the two datasets for each mutant. This design addresses Task 3: Examining quantitative neuronal regeneration data across gene mutants. Which gene mutant condition allows for the best (most) DLK-1-independent regeneration?

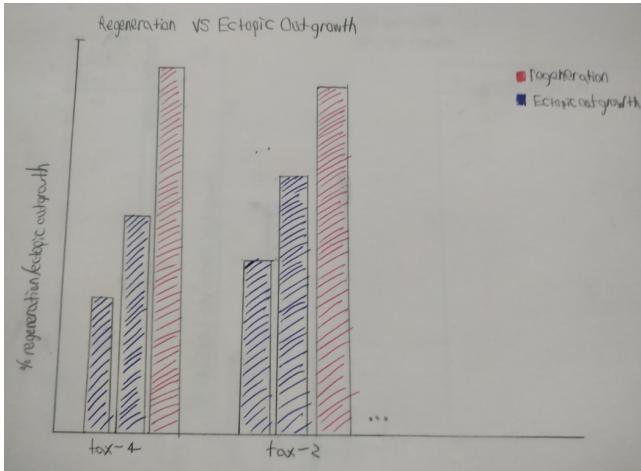


Figure 7: SKETCH 4: Illustrated by Harris Lussenhop: This figure showcases bars (marks) for ectopic outgrowth and regeneration data. A bar chart was used to allow for easy visual comparison of data magnitudes of neuronal regeneration and ectopic outgrowth across different mutant groups. The color encoding indicates if the bar represents ectopic outgrowth or neuronal regeneration and there are three bars for every mutant group. The bars are placed side-by-side to allow the reader to easily compare ectopic outgrowth and neuronal regeneration within one view. This design addresses Task 3: Examining quantitative neuronal regeneration data across gene mutants. Which gene mutant condition allows for the best (most) DLK-1-independent regeneration?

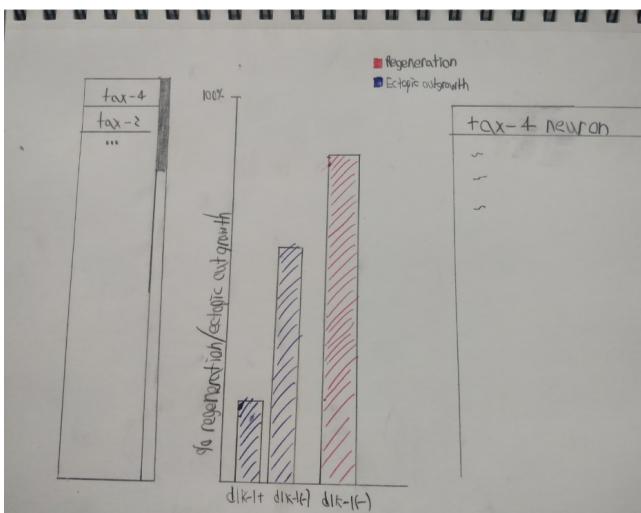


Figure 8: (FAVORITE) SKETCH 5: Illustrated by Harris Lussenhop: This figure represents bars (marks) for ectopic outgrowth and regeneration data. The color encoding indicates if the bar represents ectopic outgrowth or neuronal regeneration and there are three bars for every mutant group. On the left of this figure is a menu where the user can choose specific neurons to examine. Once a specific neuron is chosen by the user, the middle figure displays the percent ectopic outgrowth and neuronal regeneration for DLK (DLK-1)-independent and DLK (DLK-1)-mediated neurons. On the right is a space where more data about the specific neuron can be displayed, whether it be textual descriptions or other relevant data. This design addresses Task 3: Examining quantitative neuronal regeneration data across gene mutants. Which gene mutant condition allows for the best (most) DLK-1-independent regeneration?

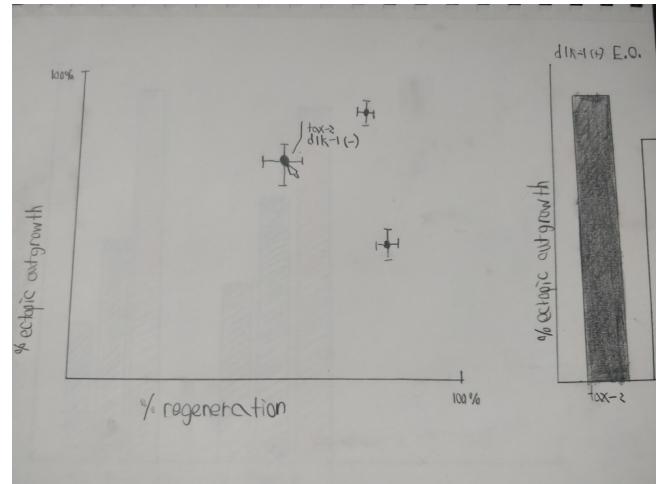


Figure 9: SKETCH 6: Illustrated by Harris Lussenhop: This sketch showcases a dot plot on the left showcasing percent ectopic outgrowth and percent regeneration for DLK (DLK-1)-independent neurons. The dot plot uses dots (marks) that allow for quantitative data to be displayed on both the x and y axis instead of having two channels in the y axis. In addition, this figure allows for error bars to be clearly represented in relation to the dots. The dot plot does not display the data for DLK (DLK-1)-mediated neurons so that data must be put on a separate chart to the left. When a user hovers over a dot on the dot plot, additional data will be displayed and relevant bars on the bar chart will also displayed. This allows for trends in ectopic outgrowth and regeneration to be more easily seen. This design addresses Task 3: Examining quantitative neuronal regeneration data across gene mutants. Which gene mutant condition allows for the best (most) DLK-1-independent regeneration?

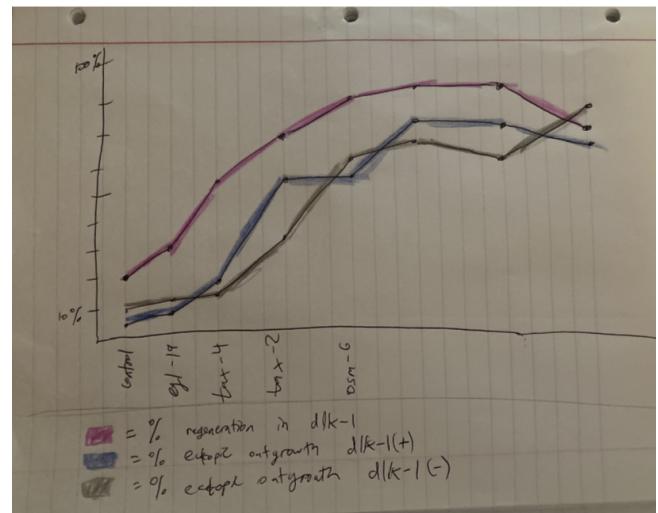


Figure 10: SKETCH 7: Illustrated by Gus Lee: This design uses a line chart to display the neuronal regeneration and ectopic outgrowth data for each mutant. Colored lines were used to encode the type of regeneration (neuronal regeneration versus ectopic outgrowth). The use of lines (marks) can resolve the issue of visualizing closely overlapping data. In addition, contrasting colors can enable the viewer to easily distinguish between the different data that are present in the figure. The x-axis contains the categorical variable of genetic mutant, and the y axis encodes percentage of neuronal regeneration/ectopic outgrowth. This design addresses Task 3: Examining quantitative neuronal regeneration data across gene mutants. Which gene mutant condition allows for the best (most) DLK-1-independent regeneration?

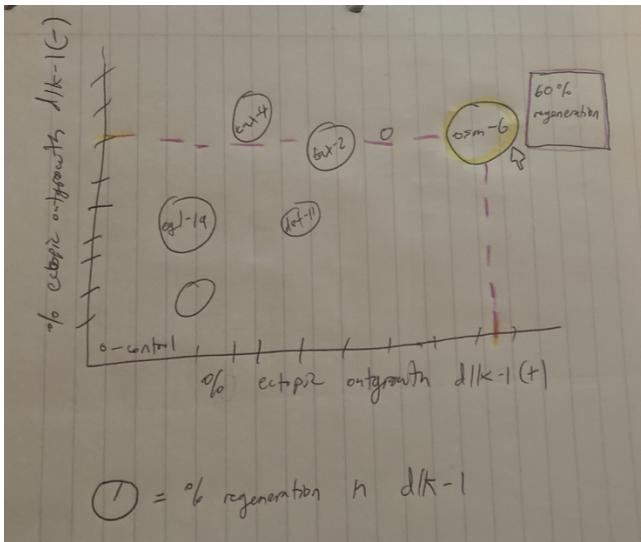


Figure 11: SKETCH 8: Illustrated by Gus Lee: This design uses a bubble plot to compare neuronal regeneration and ectopic outgrowth data in DLK (DLK-1)-mediated and DLK (DLK-1)-independent conditions. The x-axis represents the percentage of ectopic outgrowth in DLK-1-+ (DLK-mediated) worms and the y-axis represents the percentage of ectopic outgrowth in DLK-1- (-) (DLK-independent) worms. Bubbles (marks) are used to represent each genetic mutant. The radius of the bubble is proportional to the percentage of neuronal regeneration, and the x and y coordinates represent the respective ectopic outgrowth values along each axis. A bubble plot was chosen since it would be an effective way to encode data that contains potentially three quantitative variables and one categorical variable. Additionally, hovering over each bubble with the mouse pointer shows more precise data values for each mutant. This design addresses Task 3: Examining quantitative neuronal regeneration data across gene mutants. Which gene mutant condition allows for the best (most) DLK-1-independent regeneration?

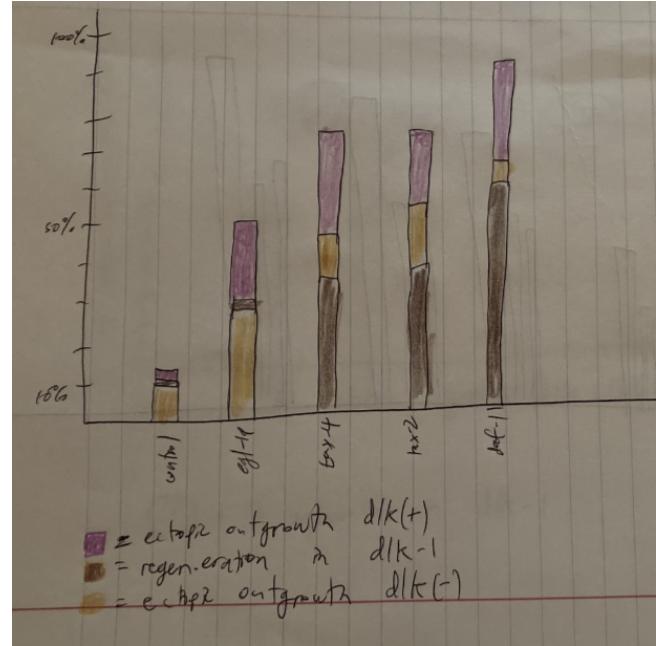


Figure 12: SKETCH 9: Illustrated by Gus Lee: This design uses a stacked bar chart to compare neuronal regeneration and ectopic outgrowth data for each genetic mutant. Colored stacked bars encode the type of regeneration (neuronal regeneration versus ectopic outgrowth). In addition, contrasting colors make it easier for the viewer to distinguish between the different data. The x axis represents the categorical variable of genetic mutant, and the y axis represents the percentage of neuronal regeneration/ectopic outgrowth. A stacked bar chart was chosen because of its ability to display multiple quantitative variables against a categorical variable. This design addresses Task 3: Examining quantitative neuronal regeneration data across gene mutants. Which gene mutant condition allows for the best (most) DLK-1-independent regeneration?

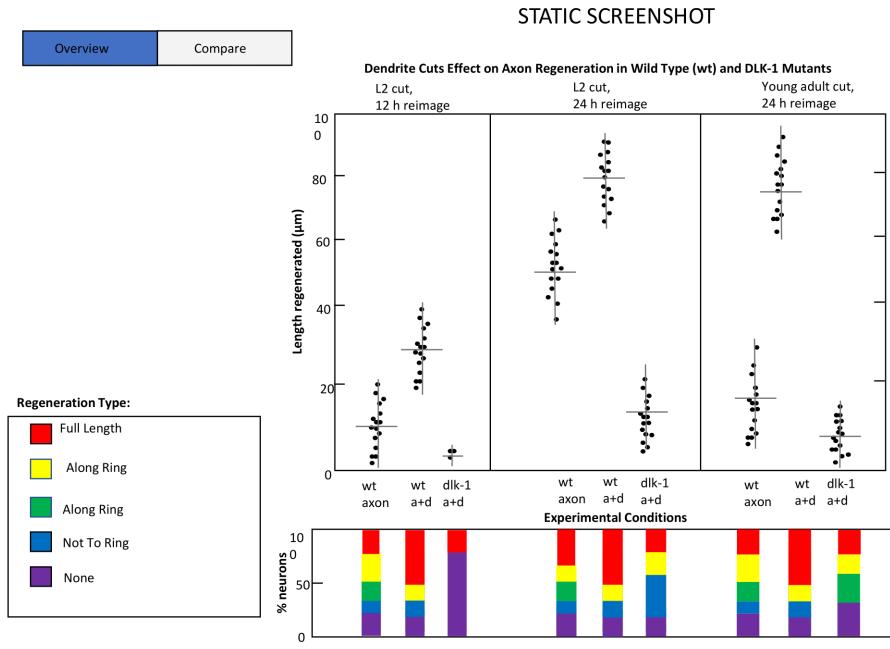
The favorites were chosen for their strong ability to represent the neuronal regeneration and ectopic outgrowth data in a manner that allows for easy visual magnitude comparisons. Favorite Sketch 1 uses a bar chart to compare neuronal regeneration and ectopic outgrowth data. The differing color encodings for the two datasets allow for easy visual comparisons of each dataset across all the mutants. In addition, the chart contains clickable filtering features to differentiate the viewable data based on selected criteria. The “type of gene mutant” or “if the axon/dendrite is cut” can be manipulated to selectively show a specific subset of data. This type of interactive chart is useful since it provides an overview of the data that allows for exploration of extreme values and also allows for filtering of the data to facilitate specific scientific analyses.

Favorite Sketch 3 uses a modified stacked bar chart with different origin axes for neuronal regeneration and ectopic outgrowth data. The use of a stacked bar chart without adjacent bars allows for easy visual data magnitude comparisons between ectopic outgrowth and neuronal regeneration across the different mutants. The design choice of not using adjacent bars enables comparison of the respective bar magnitudes on a common respective axis, which aids in quick visual magnitude comparisons across mutants. If bars for ectopic outgrowth and neuronal regeneration data were adjacent in the stacked bar chart, then the bars would have varying origins and would be difficult to quickly and accurately compare their results across mutants. A modified stacked bar chart is useful here because it allows for comparison between ectopic outgrowth and neuronal regeneration data across mutants in a way that does not distort the appearance of the other bar. In addition, the distinct color encodings for the neuronal regeneration and ectopic outgrowth data prevents confusion when comparing the respective data across mutants.

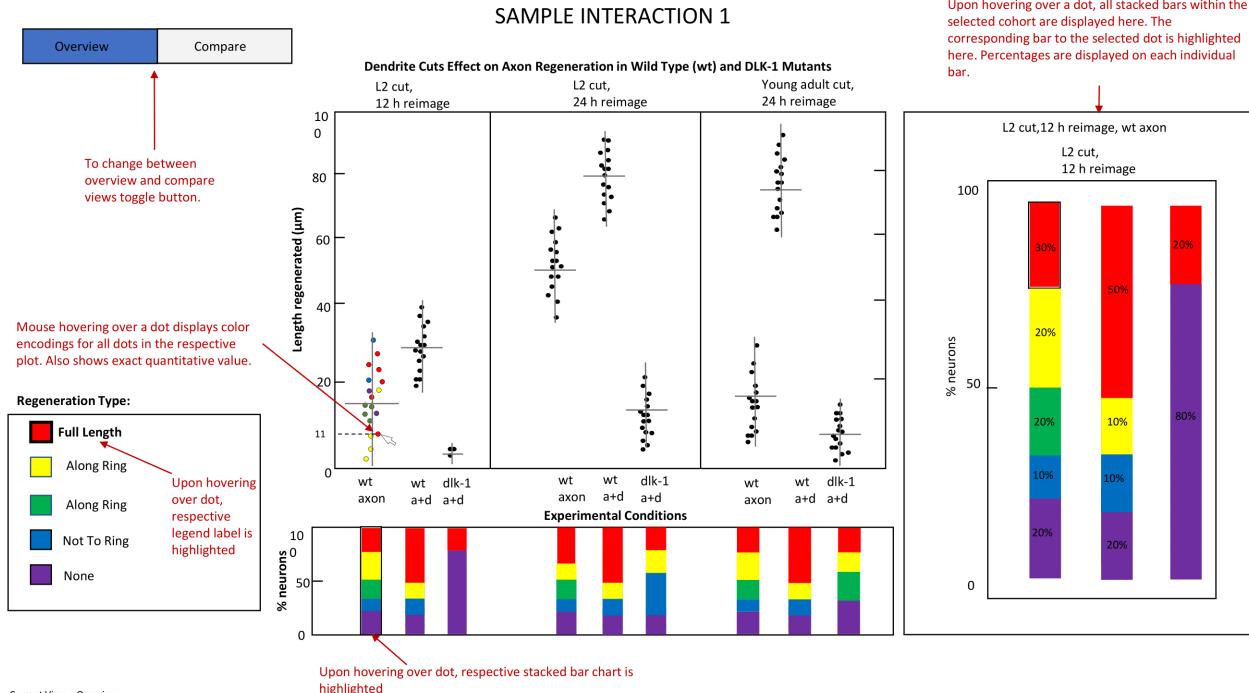
Favorite Sketch 5 uses a bar chart to compare the neuronal regeneration and ectopic outgrowth data. Similar to Sketch 1, the use of a bar chart enables the easy visual data magnitude comparisons between ectopic outgrowth and neuronal regeneration across the different mutants. The use of different color encodings for the neuronal regeneration and the ectopic outgrowth data instructs the viewer to be sure of which data to visually compare across mutants (ideally, should compare bars of the same color across mutants). The bar chart contains an interactive filtering system that allows filtering based on the specific gene of interest. When the gene of interest is specified, the overview bar chart will be filtered and the specified bars will be shown alongside any scientific information about the selected gene (will be displayed on the right side of the page). This type of interactive chart is useful because it allows for users to focus on specific genes from the overview bar chart and to learn more about the gene of interest. The extra information from the gene of interest can enable users to better contextualize the bar chart results within the scope of relevant biological processes.

PLEASE NOTE: sketches illustrated above focus on Task 3 and may not resemble/be representative of visualizations that focus on Task 1 and/or Task 2.

15 APPENDIX D: DIGITAL SKETCHES



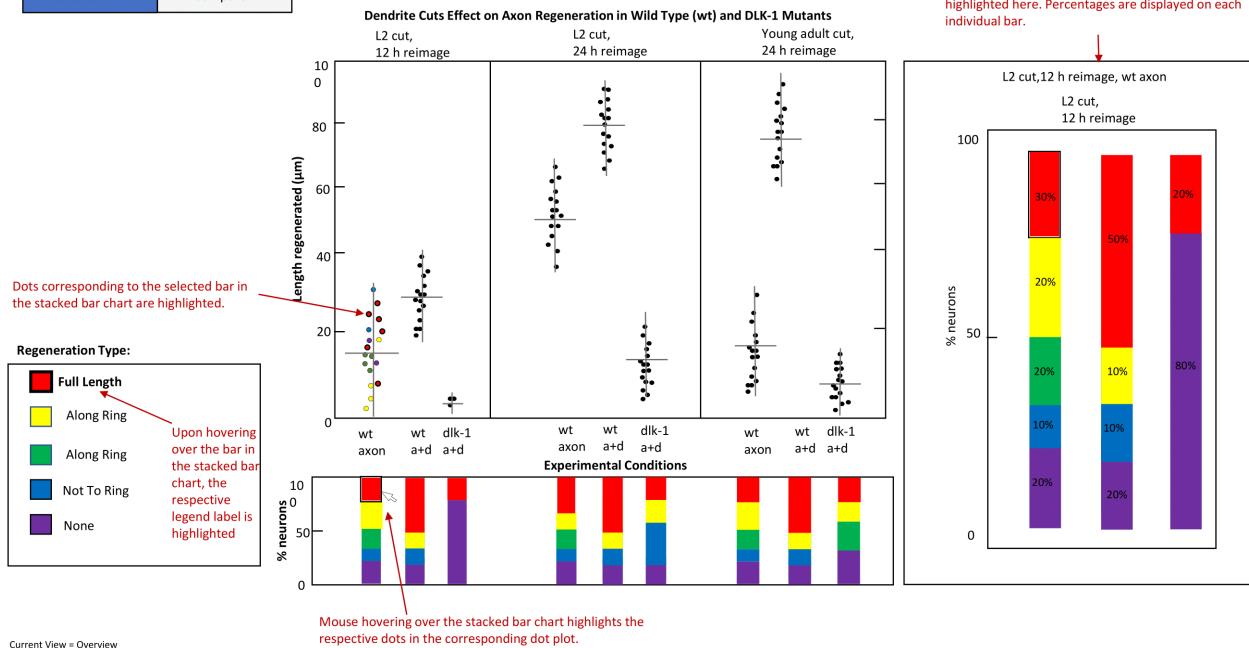
Current View = Overview



Current View = Overview

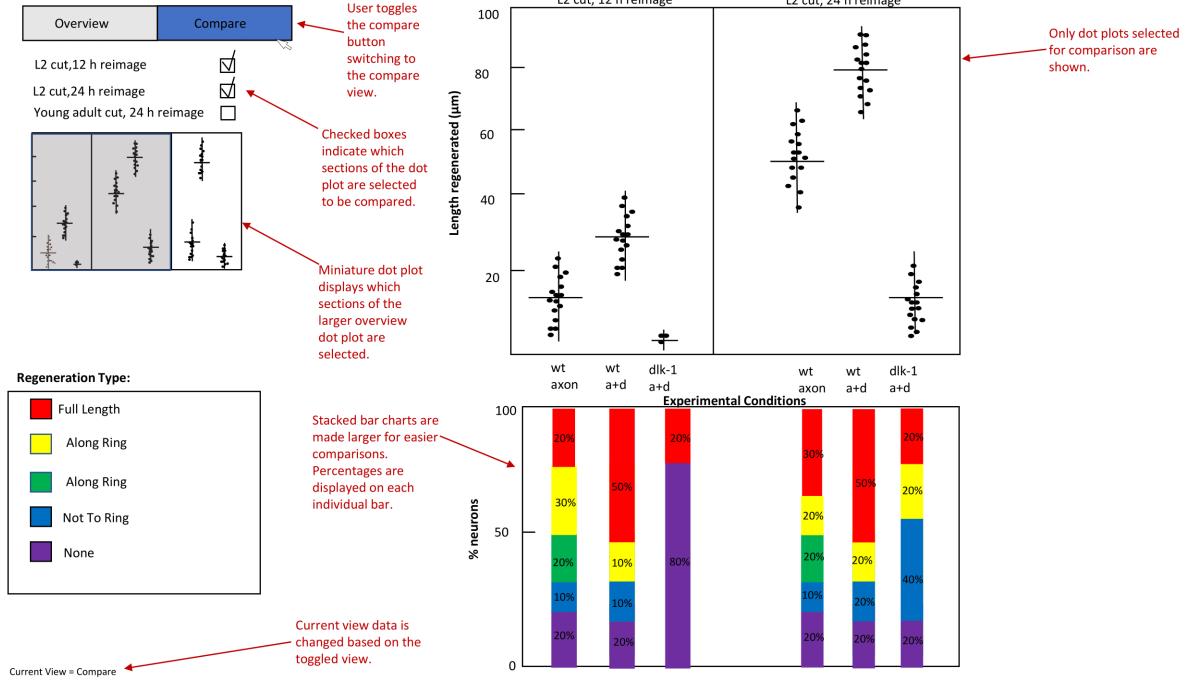
Overview Compare

SAMPLE INTERACTION 2



Current View = Overview

SAMPLE INTERACTION 3



The sketches address Tasks 1 and 2. In short, the sketches look to see how dendritic lesions affect axon regeneration in Wild Type and DLK-1 mutant worms and how the distribution of regeneration type (i.e. to ring, along ring, etc.) changes over time. There are no new tasks to add, since Task 1 and Task 2 can be addressed in one visualization using interactive and multiple view systems.

Task 1 and Task 2 will be prioritized for the design. Using multiple views and interaction, data from axon regeneration in Wild Type and DLK-1 mutant worms with dendritic lesions at a certain time point can be linked to another view that reflects the distribution of regeneration type at that time point. Our design helps accomplish these tasks by using dot plots and stacked bar charts to convey regeneration changes and regeneration type changes over time. Using interactive functionality, powerful data comparisons are possible to further explore the specific experimental conditions that are linked with regeneration and regeneration type changes over time.

Our previous plans were to focus on a visualization that addresses Task 3. However, upon group reflection and external advice, we decided that a visualization that addresses Tasks 1 and 2 would be a more interesting and useful endeavor. A visualization that can address multiple tasks is a more powerful tool to researchers than a visualization that only addresses a single task.

16 APPENDIX E: REFLECTIONS

Ajay: Throughout the project, I relied on messaging services and email to communicate with my group members and the partner, respectively. Group member communication over messaging services was reliable due to all group members being attentive to the group chat. Similarly, the partner was prompt with email responses when emailed by us. In general, communication with the partner has been effective but relatively infrequent. The partner was receptive to any questions we had and was willing to offer data interpretation advice. However, partner communication could be improved if we sent more frequent update emails to notify them about the status of the visualization. Towards the end of the semester, the partner did not inquire about the project and our group was so hyperfocused on creating the visualization that we rarely sent the partner any major updates. However, despite the infrequent communication towards the end of the semester, the partner was generally kept updated about any major progress with our project and was well-informed about our project goals/direction.

Gus: As a group, we communicated using a texting group chat. This was effective seeing as we all checked it and responded regularly, making sure to stay up to date with assignments and project work. We commonly reminded each other about meeting times, individual work, pushing to the github, and many other course related matters. With our partner, Professor Chung, we regularly emailed him questions about the dataset, which he responded to promptly. We never had long term issues with the datasets, and he was very useful when it came to answering our questions. One thing that could be improved upon was the frequency that we talked to each other. Other than general questions about the data, we rarely spoke with each other, mostly because our primary focus was on the code. With regards to partner communication, our means of communication was generally effective.

Harris: Our group primarily communicated with each other using a texting group chat, this is what we used to set up group meetings, and also talk about minor problems and adjustments that did not require a group zoom call. Every group member was attentive in messaging the group chat and there were little issues regarding it. We emailed with our partner and we were reliable with on-time responses as was he when we had questions about his project and the data used in it. Although our communication was effective, we could have communicated more with our partner towards the end of the project.

17 APPENDIX F: SLIDES

[Slides](#)