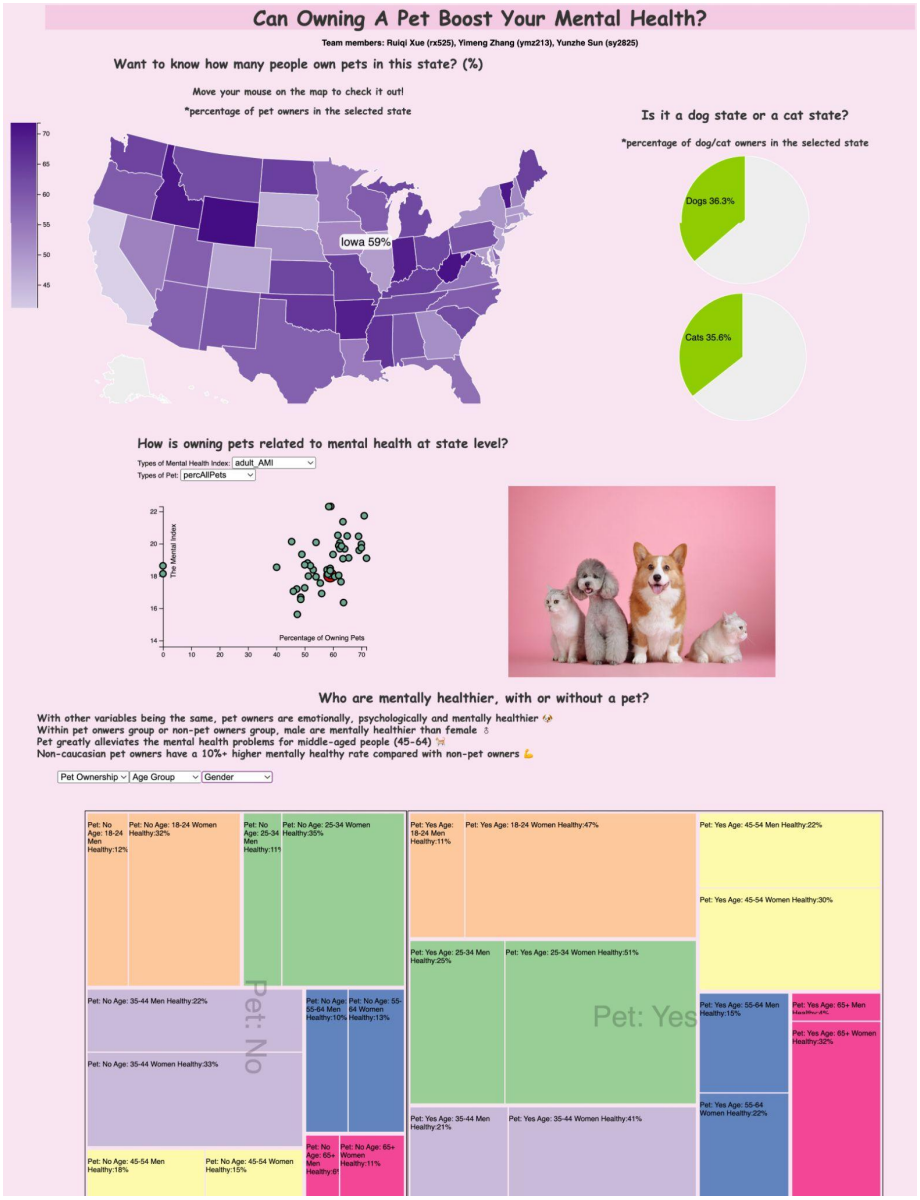
Final Project Report

Title: Can Owning A Pet Boost Your Mental Health?
Team members: Ruiqi Xue (rx525), Yimeng Zhang (ymz213), Yunzhe Sun (sy2825)

1. Overview



Our visualization project seeks to study a contentious scientific myth surrounding pet ownership and its potential impact on mental health. Using a linked design approach that incorporates a geo-map, pie chart, and scatter plot, we provide an in-depth analysis of pet ownership statistics in the United States in 2016, and their correlation with mental health conditions at the state level. Our treemap delves even further into individual-level data, exploring how various socioeconomic factors impact pet ownership and mental health outcomes. With public health officers and therapists among our target audience, our visualization offers valuable insights into the complex relationship between pet ownership and mental health, providing a valuable tool for anyone seeking to explore this important topic in greater detail.

2. Data

Three datasets related to pets and people's mental health conditions in the US are used in this project. In detail, the first dataset contains the records of household pet ownership for 50 states. Each record has 3 attributes that entail the popularity of a specific breed (*Any pet(s), Dog(s), Cat(s)*), i.e., the percentage of owning a pet(s)/dog(s)/cat(s) for each state. Notably, a few data points are unavailable so we replace them with a N/A and use the color white in the view design.

In the same "state" level, the second dataset contains the mental health indexes for each state from the same year. Each record has 6 attributes that reflect the specific mental health condition of the adult and youth population, such as the *total mental health score* and *the percentages of the adult population with serious thoughts of suicide*. For convenience, the first two datasets are joined to have a coherent structure of data.

After examining the relationship between pet ownership and mental health, we would like to dive into the question of how pet ownership affects mental health differently for different groups of people. Therefore, we utilize the third dataset used in a research paper titled "Association between Pet Ownership and Mental Health and Wellbeing in a Cross-sectional Study" from Harvard Dataverse. This dataset contains 23 variables and 1501 records. Given that we are going to design a tree map using the dataset, for conciseness, we would like to select a portion of variables relevant to our topic. The variables we want to use can be mainly divided into three categories: 1) *pet ownership*; 2) demographic variables such as *age, gender, education, and ethnicity*; 3) *mental health condition* as target variable (binary: either in good or bad conditions). Moreover, we transformed quantitative variables such as age into categorical variables (groups) for better illustration.

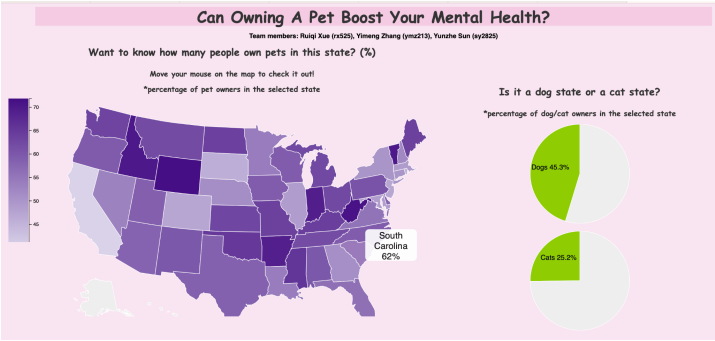
3. Goals and tasks

We aim to make a visualization that obtains simplicity and aesthetics at the same time to show the potential pattern and relationships between pet ownership and mental health. We hope that users can first get a general view of the situation of pet ownership in the United States on a state level along with being aware of the geographics. S/he would be able to get the state name, percentage of pet owners in a state's population, and the situation of dog and cat ownership via interactions, and further through comparison, s/he can get a taste of if a certain state is a "dog state" or a "cat state" in a relative term. Then the users can be walked to a scatterplot that contains the information on the mental health index and pet owning percentage to discover the possible relationships between owning a pet and mental health with detectable patterns. In the meantime, these two views are linked, so the users can have a geo sense while exploring the patterns. There are several drop-down boxes for different mental health issues by which users can select a specific mental health issue and filter the information they are typically interested in. Then there is a treemap for users to dig deeper into an individual level after having a sense of how owning pets affect mental health at the state level. By selecting multiple dropdown boxes that contain nested attributes (e.g., gender, age group, income group), they can know which socioeconomic factors can significantly have a relationship with pet ownership and have an impact on mental health. Either for a public health officer who wants to determine which socioeconomic factors moderate improvements in mental health, a therapist who plans to suggest your client consider animal therapy, or even for someone like a kid who wants to statistically convince his/her parents to buy a pet, we hope this website could satisfy their needs with an integrative view on how owning a pet can affect one's mental health.

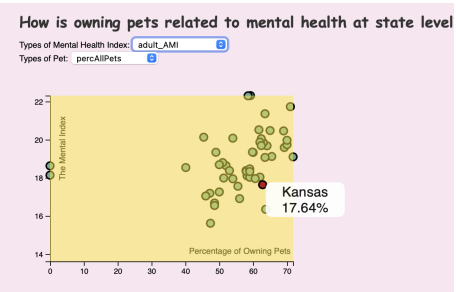
4. Visualization

To achieve the goals mentioned above, we first have a geo-map of the United States by states. We use channel saturation to indicate the number of percent of pet ownership of each state, the darker the color, the higher the percentage. Users can hover their mouse on each state and whenever the mouse is on, the state block (the area mark) on the map is highlighted and there is a tooltip that shows the state name and the percentage of the selected state.

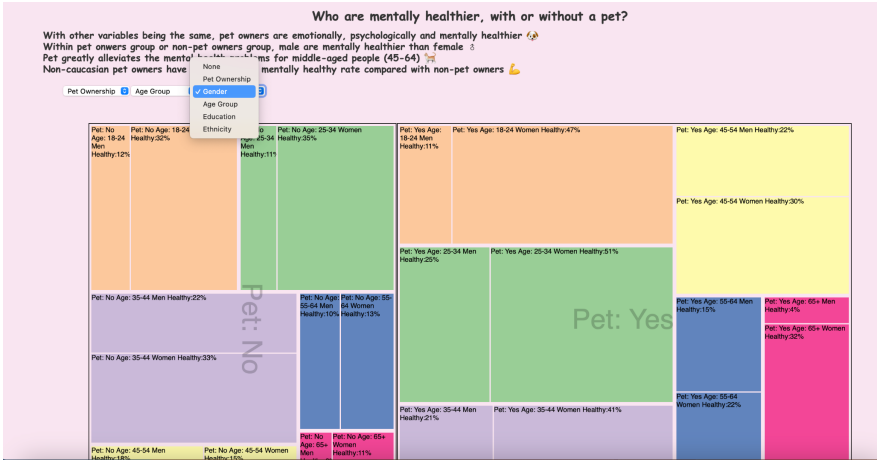
One innovative point is that we also create a pie chart containing two pies for dogs and cats right next to the geo-map to show the numbers of the percentage of dog and cat ownership in each selected state. The two pies are placed together and aligned so the comparison can be easily made in a clear and straightforward way for if the selected state has more dog owners or cat owners in relative terms.



The second main view is a scatte rplot. The x-axis represents the pet ownership percentage and the y-axis represents the mental health index, and each dot represents a state. Users can choose a specific mental health issue by selecting different options in the dropdown boxes which can better address the types of mental health and pet ownership. When users hover their mouse on the scatterplot, the selected dot will turn red and the background will turn yellow with some opacity to highlight the dot. Alongside, there is a tooltip showing the state name and percentage of ownership of selected types of pets. In addition, the two views of the geo-map and the scatterplot are linked using d3 which gives users a clearer view of each state both geographically and mental health conditions.



Finally, we have a tree map illustrating how pet ownership affects individuals' mental health differently for different groups of people. Given the individual-level data, we first divide the dataset using the primary attribute – pet ownership. Then, we further divide pet owners and non-pet owners based on secondary attributes such as gender, age group, education, ethnicity, etc. The data division is represented by three dropdown boxes for users to select. There are texts inside each block showing the demographic information of the particular group, with their mentally healthy rate (calculated by the number of people in good mental conditions divided by the overall number of that group). In this tree map, we use channel hue to represent the categorical variable for separating each demographic group visually. By having this tree map, users are able to tell whether pet ownership will have a different mental health effect on females versus males, older people versus the young, etc.



5. Reflection

Our final visualization is almost the same as what we stated in the proposal: we have four views including a geo-map, pie charts, a scatterplot, and a tree map. Each of them serves its own purpose in illustrating different aspects of our goal.

We started our project with data collection, cleaning, and pre-processing. The first challenge we encountered is that there is no available complete dataset containing both pet ownership and mental health indices, so we tried our best to find both data from the same location and year to make sure our analysis is accurate. Moreover, given the fact that the pandemic might have a great impact on people's mental health, we choose a dataset prior to the pandemic to control for this effect. Given that these two variables are from different datasets, we manually match them together to create our own dataset.

Next, each of us worked on an individual view. Regarding pie charts, our original design was to demonstrate the percentage of dog owners and cat owners within one pie. However, we realized there can be an overlap between dog and cat owners. Therefore, a single pie chart is not appropriate to visualize this information. Therefore, we change our design to two adjacent pie charts and comparisons can be made in an accurate manner. Looking back, if given more time, we would probably use a double-donut chart which makes it further easy to make comparisons using two rings within one chart.

After completing individual views, we connected the pie chart – geo map – scatterplot together to provide dynamic interaction with the audience. We encountered several technical difficulties, for example: since we are using d3 for this part, we defined several different functions to make the connection between geomap and scatterplot happen, taking into account with and without the mouse hovering on the geomap using an if-else statement. In addition, since we integrate two versions of d3 (version 3 and version 6), it takes extra effort to figure out their differences in syntax to make codes work.

When all the views were finished, we embellished our visualizations with different styles including changing the background color, distinguishing font size, and adding a cute picture of pets to make it more interesting and engaging. Most importantly, we think about how we can make our project more meaningful by enhancing storytelling. We add questions and prompts that encourage the audience to engage with the views and present key summaries to make the conclusions clearer.

We have been solving challenges and aiming for better design throughout this visualization project. Our takeaway is that we need to have a clear goal, a deep understanding of the available data, and present the data in its most appropriate shape and form precisely and concisely to convey information that makes sense to the audience through storytelling.