Visualization of Texas Austin Animal Shelter

Vitid Nakareseisoon, Wenjia Wang

Viterbi School of Engineering, University of Southern California

Abstract. This paper presents a visualization of Texas Austin Animal Shelter's incoming and outgoing data. The main motivation of the project is to provide informed decision and promote animal wellness. This paper explains associated visualization components and consideration factors. Various visualization technologies and principles are employed to increase the effectiveness of the presentation.

1. Introduction

This project aims to investigate and visualize Texas Austin Animal Shelter dataset. The data consisted of animal's incoming and outgoing records that span from October 1st, 2013 to November 30th, 2016. We developed an interactive web-based presentation that allows shelter officers, researchers, and interested audiences to easily explore the shelter's operation. Along with the development process, we followed the principles described by Cairo [1] and Ware [2] in order to derive an effective visualization.

2. Dataset

The incoming dataset consisted of 57,938 records, and the outgoing dataset consisted of 58,013 records. We wrote R scripts to perform necessary data wrangling/ data munging processes and transform the raw data into a number of appropriated formats. Additionally, we wrote a Python script to execute GoogleMap's API and translate location descriptions to geographical coordinates (latitude & longitude).

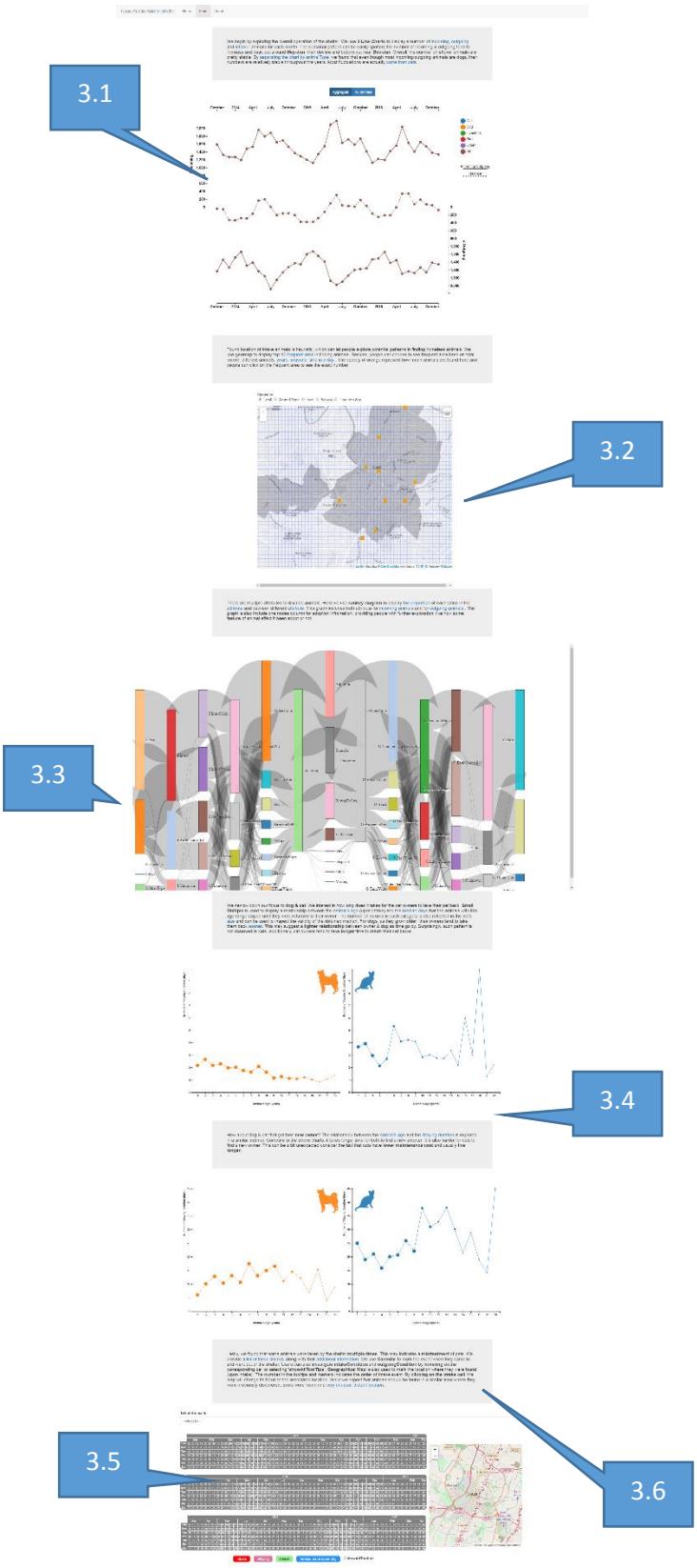
The animal dataset consisted of the following columns:

Column	Description
Animal ID	Id of the animal
Name	Animal's name
DateTime	Date upon Incoming/ Outgoing
MonthYear	MonthYear upon Incoming/ Outgoing
**Date of Birth	Birth date
*Found Location	Location descript when the animal was found
*IntakeType	e.g.: Stray, Wildlife
*Intake Condition	e.g.: Normal, Injured
**Outcome Type	e.g.: Adoption, Return to Owner
**Outcome Subtype	e.g.: Suffering, Medical
Animal Type	Cat/ Dog/ Livestock/ Bird/ Other
*Sex upon Intake	Male/ Female & Intact/ Neutered
**Sex upon Outcome	
*Age upon Intake	Age upon Incoming/ Outgoing
**Age upon Outcome	
Breed	Breed of the animal
Color	Color of the animal

(* - pertained to Incoming data, ** - pertained to Outgoing data)

3. Visualization

We structure our presentation in a vertical manner where we align the contents to go from general to fine-grained details. The overall structure of our main visualization is shown in the following picture:



3.1. Incoming, Outgoing, and Balance

In this graph, we intend to display overall number of animals that came to and went out of the shelter across the whole period. Readers can inspect any seasonal patterns and the shelter's resources can then be adjusted accordingly. We use 3 Time Series charts to display the number of incoming, outgoing and leftover animals. The data is aggregated into a monthly basis for easier digestion. The upper line is used to display incoming figures, the lower line is used to display outgoing figures, and the middle line is used to display leftover figures. Readers can select between "Aggregate" mode, which sums up the record of all animal types, or drill down to "All Animals" mode, which separates the figures by animal type. We use a different color to represent each animal type. Colors are carefully selected to easily distinguish between each animal type. For easier navigation we use smooth lines to represent incoming/ outgoing figures and the dotted line to represents leftover figures. For further clarification, hovering on any dots on a line chart will display the actual number figures and blur out irrelevant information.

3.2. Found Location

We use geo-map to show animal's found location. this is heuristic, which can let people to explore potential patterns of found location for homeless animals. our map focus on Austin neighborhood. which included in dark grey area. Top-10 frequent area for finding animals are highlighted by orange, which opacity represent frequency. The orange is outstanding compared to back ground grey. when mouse clicks on frequent area, the count number will be showed up. Above the map. we add buttons to let people to choose by which way they want the data to be shown, like

displayed by animal type, by year, by season, or time in a day, from morning to midnight. And these result can be combine together.

3.3. Animal's features

Multiple attributes for describing animals is important. They are animal' type, age, sex, color etc. People maybe want to see how these attribute distributed, what is the relationship between these attributes, if certain attribute will affect the animal been adopted or not. So, we use Sankey diagram to display the relationship between attributes.

Sankey diagrams are a specific type of flow diagram, in which the width of flow is shown proportionally to the flow quantity. We also add Parallel Coordinates, which is similar to time-series for non-time-dependent data, in order to display multivariate information. In the graph, each node represents one attribute, the flow between nodes shows how many animals from one attribute to another. Same kind of attributes aligned in one column. Besides, there is one node for incoming and one node for outgoing to divide attributes into two part. The column between them is for animal outgoing situation, like been adopted or transferred or etc.

For example, once the mouse hover over the flow between node "BirthControlled" and node "NeuteredMale" in the left side of incoming attributes, animal number will popped up. We can conclude from the graph that there are more than thirteen thousand animals have surgery to become neutered Male.

People can also drag nodes to see attribute they want to focus on.

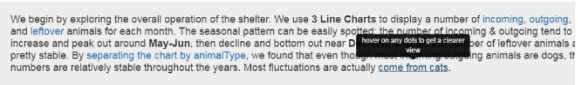
3.4. Animal's Age vs. Stay Duration

The main motivation for this visualization is to explore the relationship between animal's age and the number of days that the animals stayed in the shelter. We narrow down the focus to only dog and cat since they form a large part of the data. We use Small Multiple to distinguish between dog/ cat and animals that were returned to their previous owner/ adopted by new owner. For each age range, we calculate the median days the animals within that range stay and plot the number using Line Chart. Additionally, we realized that the validity of the obtained median depended on the amount of sample size, so we decided to encode the number of records (in another word, confidence in the population inference) into the size of the corresponding dot. The radius of the dot is maxed out at 100 records because we think that this sampling size is sufficient to reflect a true population's median. To get a precise number of days and records, readers can hover on any dots on the chart. Moreover, to attract reader's attention and make the chart memorable, we also put SVG images of a dog and cat into the graph. The images are carefully placed so that they do not turn the visualization into "Junk Chart" [1].

3.5. Multiple Adoptions

We interest in animals that were taken by the shelter multiple times. This may indicate owners' mistreatment of their pet, or a particular problem pertained with the animal. We compiled a list of animals that were taken more than once (all are dog) and display it in the dropdown list. By selecting an animal from the list, Calendar will show temporal information by highlighting each cell according to its associated event. We use 4 different colors to indicate Intake/Staying/Leave,and Intake/Leave_on_the_same_day

event. The colors chosen are based on the global perception of colors (red indicates undesired event, vice versa for green) [2]. IntakeCondition and outgoingCondition can be investigated by hovering on the corresponding cell or selecting 'showAllToolTips' option. Furthermore, spatial information of the found locations will be displayed in the Geographical Map. The number in the tooltips and map markers indicates the order of Intake event. By clicking on the cell that contains Intake event, the map will change its focus to the found location associated with that event.

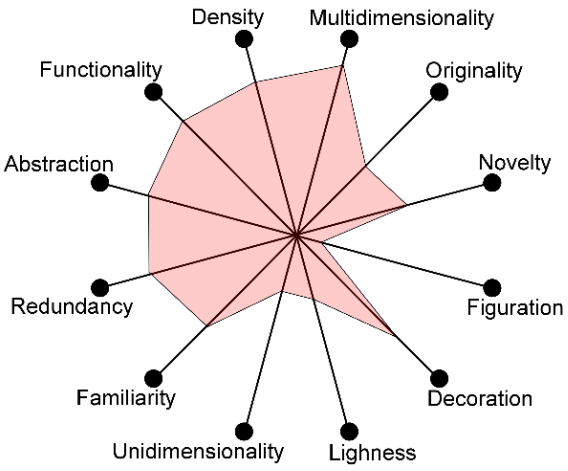


(3.6. Reader's Aid)

3.6. Reader's Aid

Lastly, we realized that some of visualizations can be quite unfamiliar to the readers. In order to minimize the confusion, we decided to increase Redundancy in our presentation [1]. We use a number of tooltips to enhance reader's comprehensibility and emphasize important texts to guide reader's attention.

4. Design Decisions



Abstraction-Figuration

Our data consisted mostly of number figures and graphical representations, and thus are highly abstract. Geographical representation and animal images are used in a small extent to compensate for this.

Functionality-Decoration

We designed various functionalities that can be performed by a reader; such as comparing pattern, switching condition, and focusing on a particular figure. These functionalities drive up the Functionality score. However, we still use well-established web-based styling techniques (Bootstrap), animal images, and animations to enhance reader's memorization and capture reader's attention.

Density-Lightness

Our presentation is dense with figures, graphical representations, and textual information. We decided to compensate for this by putting them in a vertical layout so that our readers will not be overwhelmed by all of the contents at once.

Multidimensionality-Unidimensionality

We decided that readers can explore the data from various dimensions, such as time, geographical region, and many aggregated figures. Thus, our visualization is designed to lean toward Multidimensionality axis.

Originality-Familiarity

To effectively explore various relationships in the data, we utilized various novel visualizations; such as the visualization in section 3.1 and 3.3. Such visualizations are, however, composed of basic visualization elements,

such as Map and Histogram, that readers are likely to be already familiar with.

Novelty-Redundancy

In order to avoid reader's confusion because of high Originality, we increase Redundancy in our presentation, as described in section 3.6. Tooltip is used to avoid over-emphasizing because it can be shown/ hidden as needed by the reader.

5. Tools Used

Tool	Description
D3.js	Translating data to graphical representations/ performing animations
Leaflet	Displaying geographical-related data
Bootstrap	Styling web pages
jQuery	Facilitating scripting
GoogleMaps API	Translating location data to latitude/ longitude coordinates
Moment.js	Parsing date/time data
R	Performing data wrangling/ munging
Python	Calling bulk operations on GoogleMaps API

6. Related Work

The similar visualization is created by Alexandru [3] to explore the relationship between Breed and Age. The main different is that the data used in his work are based solely on the outgoing dataset. The author intends to perform exploratory analysis by focusing on Breed and Animal's Age as well as publish his code (written in R) in the source website.

7. Conclusion

We explore Texas Austin Animal Shelter dataset by creating the web-based visualization product. It is designed with effective visualization principles ([1] and [2]) in mind to maximize reader's comprehensibility. Various graphical representations are utilized to investigate the data from various angles. We hope that our project can be harnessed by the relevant readers to provide invaluable insight and informed decisions.

References

1. Cairo, A.: The Functional Art: An Introduction to Information Graphics and Visualization. New Riders (2012)
2. Ware, C.: Visual Thinking for Design. Morgan Kaufman (2008)
3. Alexandru, P.: Visualizing Breed and Age by Outcome. https://www.kaggle.com/apapiu/shelter-animal-outcomes/visualizing-breeds-and-ages-by-outcome (2016)