

# Pueblo ZooMonitor Debriefing #4

4/14/2020

## Goals:

**Goal:** Create an improved version of the prototype R Shiny program (compatible with more datasets)

→ Measured by the provision of a link for people to test out the refined prototype R Shiny Program

**Subgoal 1:** Update the generalized cleaning script as necessary

**Subgoal 2:** Enhance functionalities

**Subgoal 3:** Improve aesthetics

→ All subgoals are measured by the provision of bullet point comments enumerating the various additions and modifications

## Meeting with the Client:

We had our fourth meeting with our client on April 1st, 2020. The primary objective of the meeting was to receive input regarding the program. We had sent our client the link to the published prototype R Shiny program a day prior to the meeting so that she could test out the program in advance.

## Notes:

- She seemed to be pleased with the program that we showed her and was also open to changes that would enhance the program's usability.
- We proceeded to explain some issues with the program that could potentially come up, as well as changes that we had been considering.
- The client appreciated the content of this discussion and agreed on the changes.
- The suggestions that we made to our client regarding modifications to the program are addressed in subgoals 1 and 2.
- The client mentioned that she would contact us if she happened to think of any new ideas.

## Goal:

**Link:** <https://qishuyi.shinyapps.io/ZooMonitor/>

## Notes:

- Please use the link above to access the updated prototype R Shiny program.
  - It may be more beneficial to access the program after reading through the various updates we made in the three subgoals.
- The updated prototype R Shiny program is fully tested and compatible with 23 animal data sets. (The 3 datasets we already had on Painted African Dogs, Cattles, and Squirrel Monkeys, alongside 20 new datasets)
- The revised program incorporates all updates made in the three subgoals.
- We provided anonymous five datasets as a zip file for Professor Miller and STA-395 classmates to use when testing out the program; this addresses data privacy concerns that arose during our last debriefing.
  - These five datasets originate from the real datasets but use fake animal names, and each fake dataset is a randomly selected subset of the full data.

## Subgoal 1:

The generalized cleaning script is a single R script designed to be able to clean all animal data sets using the same procedure. This script is embedded within the R Shiny program, and cleans the raw data set that is initially uploaded by the user. We provide a very brief overview of the cleaning procedure below.

1. Read in a CSV file
2. Remove spaces in column names and replace them with underscores
3. Rename specific column names
4. Unite certain columns into a single column called Category
5. Unite certain columns into a single column called Behavior
6. Create, rename, or unite Social Modifier column
7. Filter out rows without behavioral observations
8. Fix NA issues for certain column caused by uniting in (4), (5), and (6)
9. Split rows containing multiple behavioral observations
10. Remove redundant and unnecessary columns
11. Standardize the format of Category and Behavior values
12. Add a Day of Week column

As of March 31st, the generalized cleaning script was only fully tested with three data sets. For this debriefing, we tested 20 new data sets and updated the script to resolve issues that came up.

### Updates:

- The cleaning script now standardizes the format of Category and Behavior values by using title case.
  - This update was implemented primarily to fix issues where a single category or behavior was recorded under multiple names due to a typing error made by the zoo keepers.
  - A limitation is that this update can only fix capitalization errors, and not spelling errors.
  - Standardizing Category and Behavior values also improves the aesthetics of visuals and tables in the program.
  - This update is reflected in **(11)** of the overview above.
- The cleaning script now artificially creates an empty Social Modifier column for any data set that lack this column
  - Since the previous version of the script was written under the assumption that all animal data sets contain at least one Social Modifier column, we ran into issues with any newly tested dataset that lacked this column.
  - Specifically, the code that fixes NA issues for the Social Modifier column in **(8)** crashes when there is no Social Modifier column in the data set.
  - This update fixes this issue, and also ensures that all cleaned data sets are identical in structure (All cleaned data sets will have a Social Modifier column, empty or not).
  - This update is reflected in the “Create” portion of **(6)**.

### Subgoal 2:

The refined version of the prototype R Shiny program includes an assortment of new functionalities. The new features make the program easier to use, provide more information to the user, and allows for a more detailed analysis of the animals.

### Updates:

- As you may recall, the first version of the prototype R Shiny program allowed users to choose between the three datasets that we had (Painted African Dogs, Cattle, and Squirrel Monkey). The “Upload Data” tab now allows users to upload their own CSV file.

- The “Categories” and “Behaviors” tabs are now merged together in a single tab called “Activities”. Instead of having two separate tabs, users can use radio buttons to choose whether to filter by Category or Behavior under the “Activities” tab.
- The “Activities” tab now includes a “Select All” and “Deselect All” button. The user can click this button to select all and deselect all Category/Behavior checkboxes.
- The “Activities” tab now includes a fourth subtab called “Information Table”. When users are filtering by Category, a table will appear listing the behaviors associated with the categories that the user selected.
- Information regarding the pie charts in the "Events" tab:
  - To review, this tab was created in order to explore potential changes in the animals' behavior after a certain event, such as some animal's birth or death.
  - In the last version of our program, the tab only had a date input where the user could select a specific date from the data set that they uploaded. Based on the selected date, a pie chart/pie charts would appear in the main panel, showing the proportion of various behaviors in each period (if there are two). The pie charts were labeled as "Before" and "After," representing the periods from the first day of the data set to the selected date and from the selected date to the last day of the data set respectively (When the selected date is the first or last date of the data set, it will only show the “After” or “Before” pie chart, respectively).
  - Now, in our updated version of the program, users can choose between using all observations from the uploaded data and using data that excludes the subject animal.
    - Here, the subject animal refers to the individual that caused the event (birth, death, etc). If they chose the latter, radio buttons with all animal names will appear, letting the user choose one animal to exclude from the data set.
    - By doing so, the user will not have to be concerned about the difference in behavioral proportions that is caused by the absence of the subject animal in one of the pie charts. For example, if animal  $A$  died on the selected date, then without the exclusion of this animal, the pie chart shows a behavioral proportion of all animals in the "Before" period, while all animals except for  $A$  in the "After" period, which creates an inconsistency between the two periods.
    - In other words, the user can now evaluate the meaningful impact of an event on the behaviors of the other animals
  - However, there could still be an inconsistency between the animals included in the two pie charts even after the user excluded the subject animal. That is, when another animal/other animals passed away or joined the zoo during either period, i.e. when we have another subject animal on a different date. To handle this issue, we also

manipulated the date range, ensuring that both periods (before and after) include the same set of animals (Note: The set of animals may not include all remaining animals)

- Finally, if there's no such period in the data set where all animals except for the subject animal coexist, then notes instructing the user to manually choose an animal/animals to instead include in the plot will appear together with checkbox options. These checkboxes only include the names of the remaining animals. When at least one of the checkboxes is selected, a plot that includes only the animals chosen by the user will appear.

### **Subgoal 3:**

Aesthetic improvements are also reflected in the latest version of the prototype R Shiny program. Updates include the modification of the program interface, the adjustment of colors and labels of visualizations, and the fixing of minor miscellaneous issues. These changes serve to both polish and strengthen the usability of the program.

#### **Updates:**

- The program now uses the Yeti theme instead of the default theme.
- The size of all input descriptions were increased to h4.
- The typo in the "Observations" tab has been fixed (dash line → dashed line).
- All the help texts and notes in each main panel now appear only after a data set is uploaded to the program.
- All Categories and Behaviors are now in title case.
- The pie charts in the "Events" tab are now constructed using the (continuous) "wesanderson" palette, which is more eye soothing than the "rainbow" palette employed in our last version.
- The facet labels in the “Barplots of Behavior by Hour of Day” visual under the “Daily/Weekly” tab have been changed to time format (e.g. 9 → 9:00).
- Previously, in the stacked bar plots under the “Activities” tab, the color corresponding to each Category/Behavior would change each time users selected/deselected checkboxes. We assigned and fixed a color to each Category/Behavior so that the colors remain unchanged even when users select/deselect checkboxes.
- In addition to the change described above, we also changed the colors to be distinct from each other for accessibility purposes.
- There was a scalability issue in our previous version of the program: in the Hour of Day plot under the “Observations” tab, the x-axis used to have a hard-coded range of values between 9 to 16. The hard-coded range of values caused minor issues with the plot, specifically when there were hours that had no observations. We modified the code so that the range of values on the x-axis change dynamically with each dataset, fixing the issues that we previously observed.

# Progress Reflection

## Things that went well

- Equally distributed workload using Trello
- Improvements in the aesthetic aspect of the program
- Most new datasets worked perfectly on the program. The couple datasets that did not work well all had the issue of not having a Social Modifier column. All datasets worked well after a modification in the data cleaning script.
- The current version of the pie charts in the "Events" tab exhausts almost all possible cases.

## Things that were difficult

- Figuring out how to remove the inconsistencies in the pie charts was troublesome. The code required many divisions into cases, although it still does not capture all possibilities
- Overall, tasks regarding the pie chart and the creation of the information table were the most difficult and took the longest amount of time.

# Goal Reflection

**Goal:** Create an improved version of the prototype R Shiny program → **Complete**

- The updated program incorporates all the changes made in the three subgoals
- The program is also fully compatible with 23 animal datasets
- We published and posted a link so that Professor Miller and STA-395 classmates can test out the program.

**Subgoal 1:** Update the generalized cleaning script as necessary → **Complete**

- We made two changes to the generalized cleaning script
- The first change solved capitalization issues with Category and Behavior names, while the second change ensured that the program is compatible with all data sets

**Subgoal 2:** Enhance functionalities → **Complete**

- We made numerous modifications to enhance the functionalities of the program and listed all of the updates

### **Subgoal 3: Improve aesthetics → Complete**

- We made a variety of adjustments to improve the visual aspects of the program and listed all of the updates
- We were able to successfully complete our main goal and all three of our subgoals
- The main goal and all three subgoals were feasible, but as always, some parts required more extensive efforts than others
- The creation of the information table and the modification of the pie chart were the two most complicated tasks

## **Next Steps**

This time, our main focus was on the enhancement of the program in all aspects. In addition to the general improvements and modifications, we fixed any problem that came up with the newly obtained data sets.

### **Considerations for our next goal:**

1. We will fix the date range setting in the "Events" tab to have them capture all possibilities
2. We will continue to reflect on opinions from our client, STA-395 classmates, and Professor Miller and work to further improve the program.
3. We will continue to test out the program with the rest of the animal datasets, which will be around 20 more, and fix issues alongside, if any.
4. We will create a short tutorial video for this program targeted towards potential users at the Pueblo Zoo. We will base our user guide, which is one of the components of the final deliverable, on this video, taking into consideration our client's feedback.