Writeup

**Introduction**

My final project supplements two separate research projects of Neeti Jain, who is Master of Environmental Science Candidate at Yale School of the Environment (and my brother’s fiancé). The first project comes from Neeti’s current partnership with Yale Peabody Museum of Natural History. Neeti is working with Peabody to retool the museum’s education curricula to incorporate narratives from an environmental justice framework. Some of her work makes use of surveys that Peabody sent out to different target groups including museum members, Yale students, and grade-level students. These surveys asked some questions with multiple choice responses, but also questions that requested written responses. I take these written responses and apply sentiment analysis to tease out broad trends from two surveys as well as trends from individual questions. The second project comes stems from research Neeti conducted as an undergraduate in the Department of Ecology and Evolution at the University of Chicago. She assisted in the data collection and analysis of physiological and behavioral responses of two zoo-housed female giraffes to construction at Lincoln Park Zoo from 2014-2020. I utilize data collected on cortisol metabolite levels (CC (ng/g)) – which can be used to determine stress levels – from this project in 2014-2015 during a period of demolition and active construction (with a brief period of no construction in-between those two phases). I plot these data by phases of construction and fit a model to see if there are statistically significant changes in CC (ng/g) levels during construction (demolition and active construction) relative to the period of no construction.

**Peabody**

The data Neeti shared with me from the Peabody came from two separate surveys: Evaluation A was administered to paid Yale students and Peabody members who responded to an e-newsletter, and Evaluation B was administered to Yale students that work for the interpretation team at Peabody. The surveys asked participants to read labels that appear alongside museum exhibits and respond to several questions about these labels. Neeti asked me to find general trends from these responses and words or sentiments that were most frequently used.

The survey responses were download onto an Excel file in separate sheets from the survey software Qualtrics. I imported the two sheets with the two evaluations of interests and cleaned and merged the two data sets organized by question and evaluation. I cleaned the data so as to only include the text responses from the individual respondents. After talking over the research goals and what my analysis could offer, Neeti indicated that classifying words used by respondents as positive or negative would be most helpful to her. I also included a word cloud of the most frequently used words used in both evaluations combined as well as an option to generate word clouds specific to individual survey questions. She can use this information to conduct analysis on how survey participants reacted to the exhibit labels and use this data to inform decisions on how the labels could be re-written. The work I contributed to this project could taken farther to probe more into the types of words that were assigned positive and negative and if those designations were appropriate given the context of the survey.

**Lincoln Park Zoo**

The giraffe data Neeti shared with me was from the beginning stages of her research over the period of late 2014 through November 2015 when there were three separate phases of the giraffe exhibit reconstruction: demolition, a pause on construction activities, and then active construction. The research team collected fecal samples from two female giraffes, Etana and Sabrena, consistently throughout this period and collected data on several hormonal levels. I plot these data by phases of construction and fit a model to see if there are statistically significant changes in CC (ng/g) levels during construction (demolition and active construction) relative to the period of no construction. My contribution to this research could be expanded by thinking more critically about the models I fitted. I could also make use of additional data that the team collected on these two giraffes from 2016-2020 that included other behavioral observations that help better inform how construction impacted the stress levels of these two giraffes.

**Code Organization**

My code for this project is contained in two locations: final.R and app.R. The final.R script contains data wrangling, static plots, initial sentiment analysis, and regression analysis. The app.R is an R Shiny app that houses the findings from these two projects in three tabs. The “Home” tab offers brief descriptions the Neeti’s work and the Peabody and Lincoln Park Zoo projects. The “Peabody” tab contains two interactive bar graphs with sentiment analysis and two word clouds. The “Lincoln Park Zoo” tab contains interactive versions of the static plots from the giraffe analysis.

**GitHub Repository Contents**

R code and commit history

* final.R
* app.R (https://balaseric.shinyapps.io/data\_skills\_2\_r\_project/)
* lpz.R (rough draft of Lincoln Park Zoo code)
* text.R (rough draft of Peabody code)
* wordcloud.R (rough draft of Shiny app)

The initial, \*unmodified\* dataframes

* \_Evaluation A and B Text Responses ONLY.xlsx
* Etana EIA Results – Updated.xlsx
* Sabrena EIA Results – Updated.xlsx

Saved .png versions of your static plots

* etana.png
* sabrena.png
* giraffe.png
* evaluations.png

The final versions of the dataframe(s) you built (with the same caveat as in 2)

* giraffes\_clean.csv
* evaluations\_clean.csv

Writeup (Word or markdown are both fine)

* Writeup.docx