Data visualization practical work Report

Authors

FINOS SIMON

IAVARONE SIMON

LE GALL ERWAN

Problem characterization in the application domain

Dataset selection:

We decided to conduct this data visualization practical work using the 2018 Central Park Squirrel Census available in NYC Open Data here. This dataset consists of squirrel data for each of the 3,023 sightings, including location coordinates, age, primary and secondary fur color, elevation, activities, communications, and interactions between squirrels and with humans. More information is available on the Central Park Squirrel Census website here, expanding on the census process, objectives and making of their 2019 report.

Objectives definition:

Our goal is to provide a tool to an end-user able to guide him in his squirrel observations in Central Park. More precisely, we identify the end-user as a nature enthusiast, an animal photographer, or a scientist willing to conduct studies on the Central Park squirrels social behaviors. Here are the questions this end-user may ask and that should be addressed by our visualization tool:

- When and where is best to observe squirrels in Central Park?
- What is the organization of squirrel behaviors and interactions with humans and with squirrels?

Data and task abstraction

Data abstraction:

The basic abstract type of dataset that corresponds best to the dataset we chose is a table where each item is a squirrel observation with its position when sighted and the following described attributes (considered attributes only):

Attribute name	Date	Approaches	Run from	Indifferent	Fur color
Semantic	Time the squirrel was seen	The squirrel approached humans	The squirrel ran away from humans	The squirrel did not react to human presence	Primary fur color of the squirrel
Attribute type	Ordered, Ordinal, considered as value	Categorical (binary)	Categorical (binary)	Categorical (binary)	Categorical

Task abstraction:

The goal of the task abstraction step is to define more precisely the need of the user with no domain-specific language. We will first describe the actions and then the targets.

Actions:

- Use: The tool will aim at offering information consumption, allowing the user to discover new knowledge by helping with hypothesis generation and validation. (eg: “This area should be fitted for squirrel observation.”)
- Search: The user should mainly use the tool to perform a “locate” kind of search if he wants to verify a hypothesis but it could also be used in a more “explore” fashion if the target of the user isn’t as well defined.
- Query: Our tool and the dataset we chose seem more effective to perform comparisons and overviews, drawing conclusions and discovering information more from the trends and the global repartition of the data rather than single entries.

Targets:

- To identify patterns in the squirrel social organization, the user might be interested in features of the dataset like discovering spatial structures or trends linked to their social behaviors (eg. Correlating spatial distribution and social behavior characteristics).

Interaction and visual encoding

Encoding:

As space arrangement is the main interest of the defined user and the dataset, we chose to prioritize this form of visual encoding to reflect the geographical nature of the data. More precisely we chose to plot points on maps to help visualization. We also considered deriving more info from the data and presenting it using grids, expanding on the spatial nature of the data. However due to the relative small size of the dataset we did not manage to achieve any efficient visual using this lead.

Since the regional position was dedicated to the positions in the dataset, we decided to use other visual channels to deal with the behavioral attributes. In fact we used the color and size channels to separate the different behaviors. An attempt was made to use a diverging colormap representing some sort of “sociability amount” of the squirrels as for example we could define a sociability score to a region based on the observed behaviors but no visually efficient results were achieved. More precisely we used size to emphasize more noticeable behaviors, in opposition to less informative attributes.

Manipulation:

Since we want the user to be able to confront his observations and conclusions to reality we made sure to include a map allowing for a free navigation and giving reference points. Moreover we added the possibility to display more info about a squirrel sighting by clicking on the point representing it, giving for the exemple the latitude and longitude thus allowing the user to find the exact same spot.

Filtering:

We made sure to allow the user to focus on more specific behavior by hiding other features and only analyze a specific time window within the dataset. Moreover this filtering process is applied simultaneously to the different visuals so as to not hamper comparisons between the different views.

Algorithmic implementation

The algorithmic implementation was conducted using R and Shiny as instructed. To launch the application, the “SquirrelApp.R” application should be launched and might need the download of the necessary packages before running correctly. The data is already provided within the data folder and no further download is necessary to run the application. Moreover the final application was published at this address: SquirrelApp.