I Want to Believe:
A Visualization of Linguistic Features in UFO Sighting Reports

Hayley Guillou & Theodore Smith
National UFO Reporting Centre

● 24-hour hotline since 1974
● 110,000 UFO reportings
● Each report has:
  ○ date and time of sighting
  ○ date reported
  ○ shape
  ○ duration
  ○ location
  ○ text comment
Previous Visualizations

● Chloropleth maps
  ○ Sightings per capita
  ○ Population density vs number of sightings

● Trends of shape over time
  ○ Line charts, bar charts, unit charts, text tables
  ○ Tableau
What is our goal?

● Tool for exploring patterns in text
  ○ High variability in open response comments
  ○ Inconsistent grammatical structure

● Breadth-wise visualization

● Possible use cases
  ○ Geographic dispersement of sightings based on filtering
  ○ Textual comparison of different geographic areas
How did we build it?

- Data cleaning
  - Geolocation and standardized date-time information drawn from pre-existing NUFORC Kaggle data set
  - Missing / corrupted data removed using Python/pandas
- Geospatial data
  - Leaflet
  - Records clustered using MarkerCluster
- Linguistic analysis
  - NLP using SentenTree
Features

- Filtering
  - Shape (options specified by NUFORC)
  - Date (year, month, day)
  - Time (hours and minutes)
  - Duration (seconds)
  - Keywords (comma-separated exact matches)
- Detail on demand
- SentenTree comment aggregation
  - Load from current view
  - Clone to preserve results
SentenTree

- Text visualization technique for summarizing a collection of social media text
- Improvement over word cloud as it provides directionality and indication of common sentence structures
Design Choices

- What: Tabular data
  - Items: UFO sighting reports
  - Attributes: date/time of sighting, shape, duration, location, comment, date posted
- Why: Discovery/Trends
- Why: Enjoy

- How: Geographic view
  - filter items
  - hue
  - zooming
- How: Node-link view
  - directed graph
  - size encoding
- How: Facet
  - Multi-form, overview/detail
Limitations/Critiques

- SentenTree clones do not have filter information
- Lacks 2-way intractability with the map
  - Must modify keywords rather than simply clicking on words from SentenTree output
- No option to add multiple specific dates
- SentenTree can be ambiguous
Expanding the project

- Weighted comments
  - {id: 0, count: 1, text: “demo text”}
  - Count commonly used for retweets/likes
- Different marks based on shape
- Using different datasets