CS591 Data Mining Final Project
Siddharth Padture

Description:
For my project I am drawing from multiple datasets. Primarily I will be analyzing a table of ~55,000 self-reported UFO sightings in the United States scraped from the National UFO Reporting Center website. This dataset contains date, location, description, and other comments for UFO sightings for the past ~65 years. The NUFORC claims to have 'cleaned' the data for hoaxes but of course given the subject matter the ability of anyone to do this is suspect. The data is organized by date of sighting and also contains information on how long the unidentified object was observed, its general shape, and location of the sighting. Additionally each observation is accompanied by a paragraph long report with a description of the event.

Plan for Analysis:
In my analysis I will be taking the coordinates of the each report and referencing them against the known locations airfield, flight paths and military installations which have been downloaded from the Google Maps api. Additionally I will also try to correlate UFO reports with historical weather and astronomical data provided by the WeatherSource.com API. The report descriptions provide a large text corpus that I will examine to see if I can determine any trends in observations decade over decade and by location.

Hypotheses:
My prediction is that UFO reports will be clustered around airbases and airfields, as well as more prevalent in rural areas (which I will define using county population density statistics). Furthermore I predict that UFO reports will have peaked (when standardized for population growth) during the cold war, and that reports from that era will contain descriptions of

Data Description:

N	Location	Duration	Year	Description
55,401	MAX = California	AVG = 2.47 minutes	MAX = 2014	MODE = 'Light'