# TEAM\_02\_CSCI599\_HW\_BIGDATA\_Report

## Input File: ufo\_awesome\_FINAL\_OUTPUT\_v2.tsv

## Output file:

## Data Preparation and Cleaning

We started with manually cleaning the date field in the v2 TSV file in order to make the British UFO sightings dates correspond to the same standard format as that of the UFO\_awessme sightings date format.

Once we were done with that, we have written a script which converts the entire V2 TSV into json and also indexes the same into ElasticSearch. We then wrote 10 scripts to aggregate and extract the data we required for the 10 D3 Visualizations in TSV format. Once we had the 10 TSV’s, we wrote programs to put them into json and ingest it to their corresponding D3 visualizations.

Observations made about the extracted dataset:

### Our D3 Visualizations:

**The count of the number of UFO sightings based on year**

This is the first aggregation that comes to mind when deciding to collect useful information whether the number of sightings increased or decreased every year.

From the visualization it is evident that after remaining constant till almost 1995, the number of sightings reported grew almost exponentially by the year.

**The different shapes of the UFO sightings and its count**

With the data consisting of the different types of shapes of UFO sighted, we thought it would be a good thing if we could analyze of which shape did most people encounter their UFO sightings.  
  
From the visualization we can see that majority of the people found the shape of the UFO to be …. followed by ….

**The number of sightings per year within a 25 mile radius vs outside the 25 mile radius**

**The total duration of the sightings in each year**

**A**

**The count of UFO sightings in densely populated areas vs sparsely populated areas by year**

**The count of UFO sightings in a rural area vs an urban area by year**

**How many Sci-Fi movies vs sightings in each year**

**The count of meteoroid in rural vs urban areas by year**

**A Bubble graph based on the number of sightings in the states of USA**

**A graph representing who the sighting was reported by**