Playing with Messy Data in Tableau

Entity collection data: The National UFO Reporting Center http://www.nuforc.org/

Data set: Complete UFO sightings, including time standardization and geocoding
Download **complete** dataset[1] from Blackboard Session 3

1) Identify at least 10 different instances of messy data in this dataset. Explore each variable.

2) Review the event reporting form: http://www.ufocenter.com/reportformsubmit.html What are your observations? What do you see in the form that will lead to poor data quality?

Data Exploration & Quality Review

1) How many total sightings are captured in this dataset? How many have a complete information for location?

2) How complete is the dataset? (for each variable/total)

3) How accurate/reliable is the data?

4) What year was the first recording sighting? Where was it? How long was it?

5) Where in the world are the most frequently reported sightings?

6) Are there any trends or patterns in the sightings?

7) What is the most common shape reported?

[1] This dataset was scraped, geolocated, and time standardized from NUFORC data by Sigmond Axel. https://www.kaggle.com/NUFORC/ufo-sightings

Correcting Messy Data in Tableau

There are several options for correcting messy data in Tableau.

- Check the nulls
- Correct location data
- Split the variable (based on string calculations)
- Edit the alias of values
- Use filters

Using Tableau Prep to clean your data → upcoming class
https://www.tableau.com/learn/training#prep

https://interworks.com/blog/sparker/2018/05/09/tableau-prep-how-to-cleanse-your-data-and-prepare-it-for-world-domination-analysis/