Jose Arreluce
Dominic DiCarlo
Hasmik Grigoryan

Potential Asteroid Impacts: Visualization Goals and Rationale

Context and purpose:

The data consists of asteroids that Sentry, NASA's Near Earth Object monitoring system has examined and determined to be potential impactors. The purpose of the visualization is to examine the potential impact danger these asteroids pose. Our variables are potential damage (calculated by estimating kinetic energy) and impact probability. Potential damage and impact probability can be ordinally compared by looking at circle size and x-axis position, and color and y-axis position, respectively. The viewer can explore the visualization to identify points of interest, such as meteors with high impact probability, high potential damage, or a combination of the two. Previous impactors are included in the data set in order to give the viewer an understanding of the relative threat level of each asteroid. These are encoded with the color red and are annotated to draw attention to them, as well as give the viewer an understanding of the hazards a meteor impact can present.

Pre-processing:

Each meteor observation came with a number of descriptor variables; out of those we used *Cumulative Impact Probability*, *Asteroid Diameter*, and *Asteroid Velocity*. From *Diameter* and *Velocity*, we calculated the variable *Energy***,** an estimate of kinetic energy (used to express potential damage). *Energy* was estimated by the following formula:

$$Energy = Diameter^3 * Velocity^2$$

Which is based on the classical mechanics formula for kinetic energy:

$$E = mv^2 / 2$$

Diameter was cubed to estimate mass, approximating an asteroid's shape as a cube or sphere. Since we are only interested in the *comparison* of energy between meteors, we chose not to apply any constants. The damage of an asteroid is to be understood by the viewer in relation to the reference impactors, such as the Chelyabinsk meteor and the Chicxulub impactor. The variable distributions were first visualized using the python library 'matplotlib' in Jupyter.

Different scaling functions were tested so that no circle was obnoxiously large or microscopically small, while still maintaining a perceptible range of circle sizes. We decided on using $log(x)$ to scale *Cumulative Impact Probability* and $x^{1/3}$ to scale *Energy*.

Variable encoding:

Damage is encoded with circle area and probability of impact is encoded with color. In addition to this, the location of a circle along the x-axis is determined by damage, with damage increasing from left to right. The location along the y-axis is *approximately* determined by probability of impact, with the probability decreasing the further away the circle is from the center y-axis. Encoding damage by circle size will assist the viewer in exploration tasks, since points of interest physically take up more space in the visualization. We have also doubly encoded damage using the x-axis. As a result, tasks such as identifying the minimum-damage asteroid or locating an asteroid by relative damage are easy to undertake. This allows the viewer to see a clear distribution of asteroids' potential damage with respect to the probability of impact.

The cubehelix color scale gives a luminance-ordered color gradient that also varies in hue. The variation in hue makes it easier to identify discrete probability ranges in the legend, while the logarithmically varying luminance establishes a general trend from low probability to high probability. By dual encoding with both color and y-axis positioning, location, identification, and exploration tasks are easily undertaken by the viewer.

Past impact events were singled out and filled with the color red. This identifies them distinctively while not interfering with the cubehelix color scale.

Alternatives and arbitrary choices:

A simpler visualization such as a scatterplot could use spatial layout to effectively present distribution of impact probability and damage. However, this sort of visualization would give equal space to all data points, regardless of their significance to the viewer. Our visualization keeps the benefits of using spatial orientation *and* makes data points that could be of interest larger and/or gives them higher contrast against the background. We want our viewers to pay more attention to the more dangerous asteroids in addition to getting a sense of the overall distribution of asteroids.

Some arbitrariness was introduced to the layout due to the use of d3 collision force when placing the circles on the visualization. Each circle is pulled to a position on the x-axis based on the damage variable for the asteroid it is encoding. The circles start out at a certain position on the y-axis based on impact probability and are pulled toward $y = 0$ with a strength proportional to impact probability. While these forces act together, the collision force keeps the circles from overlapping each other. The unfortunate side effect of this is that the collision force prevents some circles from achieving their desired final position, so in some cases lower probability circles are closer to the x-axis than some of their y counterparts. While the distance of a circle from the x-axis is only partially arbitrary, the side of the x-axis that a circle lays on is decided arbitrarily. Also worth mentioning is the partial arbitrariness in our color encoding. While the cubehelix scale varies logarithmically in luminance, the choice of the particular pattern in hue variance is arbitrary. It is the change in hue itself that is significant, as it allows the probability band to be broken into discrete sections.

Possible confusers

It should be noted that *Cumulative* Impact Probability was used to estimate likelihood of an asteroid impact. Some of these asteroids have only a single occurrence or fly-by where they have a chance of impacting with Earth, while others pass Earth hundreds of times with a small probability of collision each time. The visualization loses this information, and combines the probability of all occurrences into one variable.

Possible hallucinators

The size of circles had to be scaled so that the most dangerous asteroids weren't obnoxiously large and the least dangerous asteroids weren't so small as to disappear on the poster. Because of this, viewers might initially misconceive the difference in danger between two asteroids, such as the largest potential impactor and the Chicxulub Dinosaur Extinction Event Asteroid, where the latter is over 100 times as damaging as the former, but only takes up about 5 times as much area. A damage scale bar has been added to the visualization to allow the viewer to better understand this scaling. It features a cube root transformation of tick marks to illustrate the difference in area between circles compared to the difference in damage.

Distribution of labor:

All group members of the project worked diligently to complete various tasks for the project. Jose worked on annotations, the force layout used to place the circles, the general layout of the poster, code style, and on the write-up. Dominic worked on the cubehelix color encoding, the legend and scales, historical asteroid impacts research, and on the writeup. Hasmik worked on processing the data from pre-processing to discovering the appropriate scaling, the force layout used to place the circles, and on the write-up.

Sources:

The bulk asteroid potential impact data was taken from Kaggle, which was sourced from NASA:
https://www.kaggle.com/nasa/asteroid-impacts

The data on historical asteroid impacts was taken from multiple sources:
http://www.fallingstar.com/historical.php
https://science.nasa.gov/science-news/science-at-nasa/2008/30jun_tunguska

Wikipedia articles on each of the historical impactors were also consulted.