Sabneet Bains

Data Visualization – EN.605.662

Project #2: Data Exploration and Design

Data Description

Near-Earth objects or NEOs are neither a planet, a dwarf planet, nor natural satellite, and the International Astronomical Union states, "All other objects, except satellites, orbiting the Sun shall be referred to collectively as 'Small Solar System Bodies'." [1] Since 1998, NASA and other world space agencies have collected data on NEOs under project Spaceguard. Thus, one area of interest that piqued my attention from this massive amount of data is studying NEOs that have close approaches to Earth, i.e., Earth-grazing meteoroids.

NASA's Jet Propulsion Laboratory at Caltech keeps a public record of NEOs' close approaches to the Earth from 1900 CE to 2200 CE. [2] The data is available to download as a CSV, Excel file, or a tabular print of the website table.

I utilized the Excel option and looked at the data from the past year (2021-Jun-27 to 2022-Jun-27). With trivial reformatting, I corrected the faulty units from "Km to m" or vice versa and fragmented the "Diameter" ranges into two discrete attributes defining the estimated minimum and maximum diameters. Likewise, the "Date" and "Time" were split for more accessible statistical analyses.

[1] https://solarsystem.nasa.gov/planets/in-depth/
[2] https://cneos.jpl.nasa.gov/ca/

Thus, the data stands at 11 variables x 1560 records. With individual variable descriptions as follows,

Variable	***Description***
Object	Primary designation
Close-Approach (CA) Date	Date of closest Earth approach
Close-Approach (CA) Time	Time of closest Earth approach
CA Distance Nominal	The most likely close-approach distance
CA Distance Minimum	The minimum possible close-approach distance
V relative	Object velocity relative to Earth
V infinity	Object velocity relative to a massless Earth
H	Asteroid absolute magnitude
Diameter Minimum	Estimated minimum diameter
Diameter Maximum	Estimated maximum diameter
Rarity	0 means an average frequency of 100 per year, 2 is roughly once a year, 3 is roughly once a decade

Table 1*:* Table column variable descriptions

Data Exploration

Variable	*Data Type*	*Mean*	*Standard Error*	*Median*	*Mode*	*Standard Deviation*	*Skewness*	*Minimum*	*Maximum*
Object	Nominal	-	-	-	-	-	-	-	-
Close-Approach (CA) Date	Interval	2021-Dec-24	-	2021-Dec-10	2021-Nov-03	-	-	2021-Jun-27	2022-Jun-26
Close-Approach (CA) Time	Interval	12:12	-	12:19	17:55	-	-	0:00	23:59
CA Distance Nominal	Nominal	3070159.91	54534.75	2801164.00	-	2153950.22	0.39	9427.00	7471513.00
CA Distance Minimum	Ratio	3017477.06	53784.61	2761558.00	-	2124322.07	0.40	1037.00	7439811.00
V relative	Ratio	10.62	0.12	9.68	8.47	4.82	1.02	0.65	35.28
V infinity	Ratio	10.57	0.12	9.62	5.65	4.83	1.02	0.27	35.28
H	Ratio	26.27	0.04	26.30	26.20	1.76	-0.51	16.60	31.80
Diameter Minimum	Ratio	0.02	0.00	0.02	0.01	0.06	18.41	0.00	1.80
Diameter Maximum	Ratio	0.05	0.00	0.03	0.02	0.10	12.10	0.00	1.80
Rarity	Ordinal /Discrete	0.12	0.01	0.00	0.00	0.37	3.31	0.00	3.00

Table 2*:* A subset of the statistical analysis that properly fits the paper aspect ratio is given here; for a more holistic analysis, please refer to the included Excel file.

Based on Table 2, it is evident that different statistical descriptors are possible/practical for specific data types; for example, it does not make much sense to find the mean of the ”Object” names as they are purely nominal. However, for something like “V relative,” finding the minimum and maximum provides the much-needed range for further calculations or visualizations.

Analytical Questions

Initially, I had the following five analytical questions that I wanted to be answered visually:

1. *Which near-Earth objects were the closest, and which were the farthest?*
2. *Which near-Earth object had the closest approach to Earth in the past year, and at what date and time?*
3. *How many rare close approaches were recorded in the past year?*
4. *Which near-Earth objects have the fastest relative velocities?*
5. *Which near-Earth objects are the biggest and hence most hazardous?*

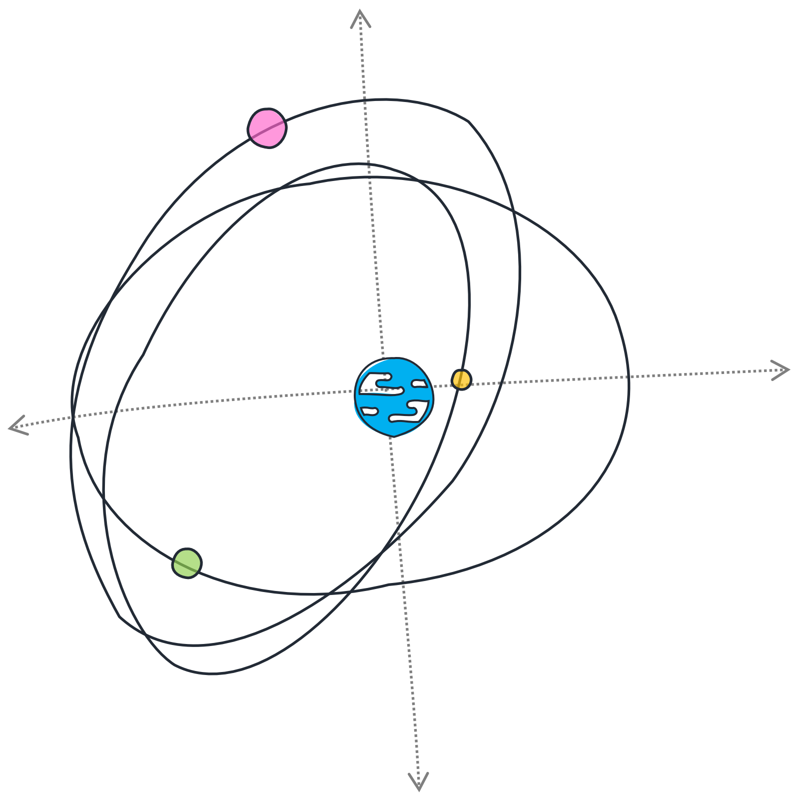
Visualization Techniques

After further analysis, I wanted to visualize the following three analytical questions:

- *Which near-Earth objects were the closest, and which were the farthest?*
- *How many rare close approaches were recorded in the past year?*
- *Which near-Earth objects are the biggest and hence deemed hazardous?*

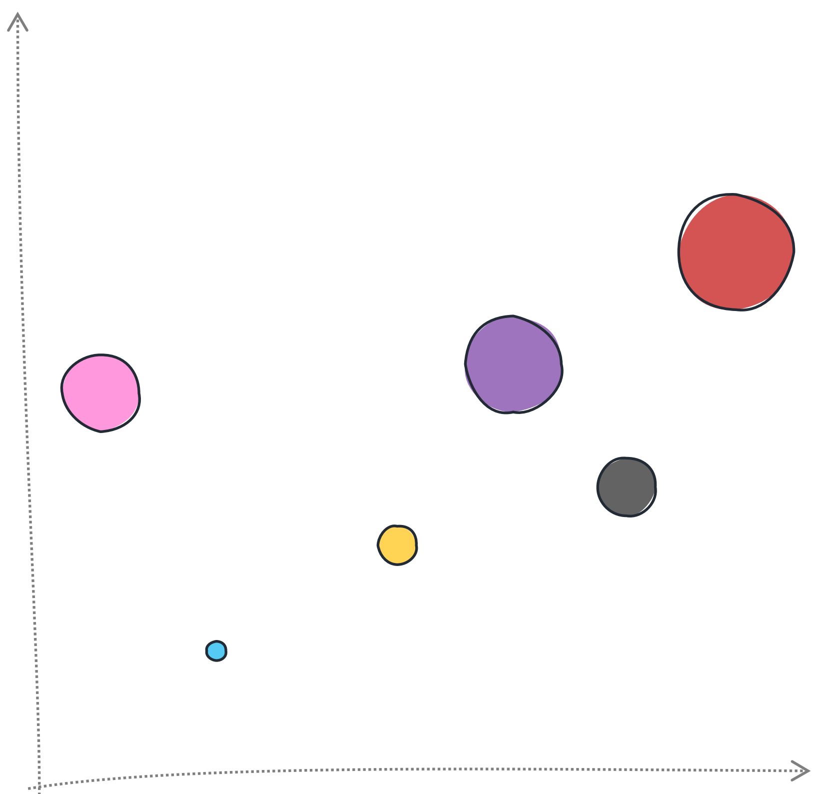
- To answer: *Which Near-Earth objects were the closest, and which were the farthest?*

First visualization



A modified radar chart could be ideal for depicting the closest and farthest approaches by each NEO. If we were interested in the relative position at a particular date, an animated version with a simple date slider could present the orbital positions of the respective NEOs observed. Since this specific dataset only provides the magnitude from the Earth's center to NEO's center, we would have to use trigonometry and spherical coordinates to recreate the pseudo-orbits. – This might be misleading, even though the visualization will be pretty!

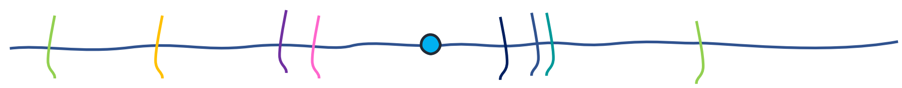
Second visualization



A bubble chart also depicts the position of each NEO compared to the origin (Earth), and if the scope of the question included relative sizes, then that could be answered immediately here as well.

Bonus: the overall visualization is more straightforward and two-dimensional.

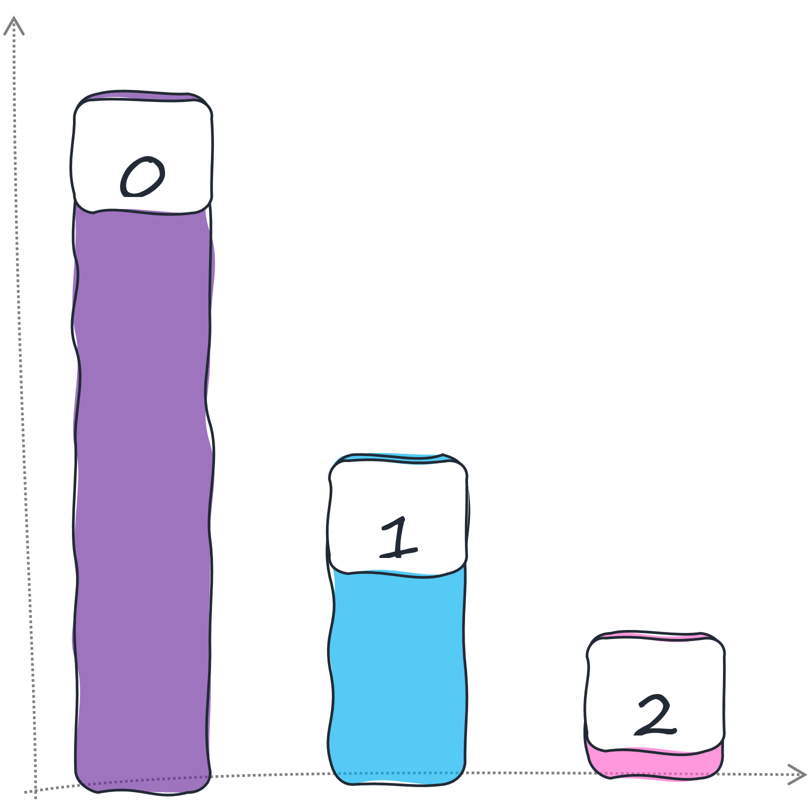
Third Visualization



Let's take simplicity even further, all the way to one-dimensional graphics - A mere number line can also display the relative distances given color-coordinated objects.

- *To answer: How many rare close approaches were recorded in the past year?*

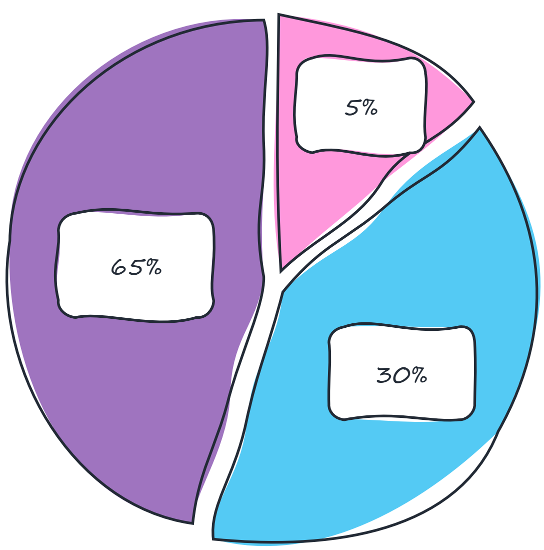
First visualization



A simple histogram best depicts each category and their respective counts. One can quickly realize if there were any ultrarare close approaches in the past year or not. This is indeed extremely powerful and simple to construct at the same time.

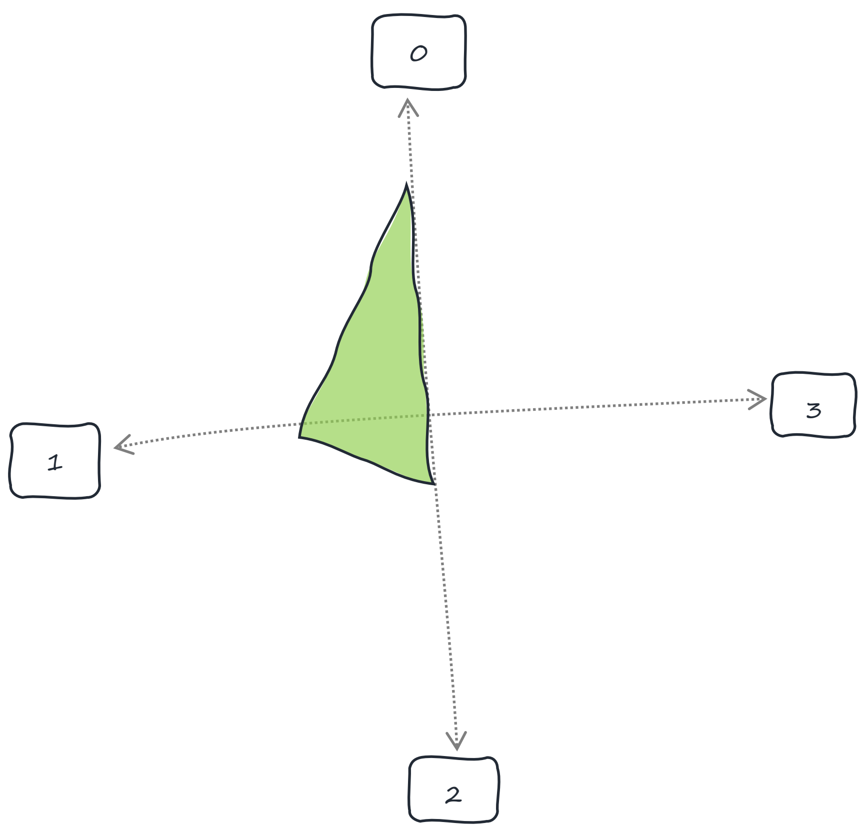
Second visualization

Although not commonly used in conjunction with ordinal data, given the scope of the question, a pie chart could tell us the same information in a percentage format. Nevertheless, this could be somewhat misleading as small percentage changes in ultrarare close approaches may be missed as "insignificant." I.e., needs some initial documentation and clarifications to understand correctly.



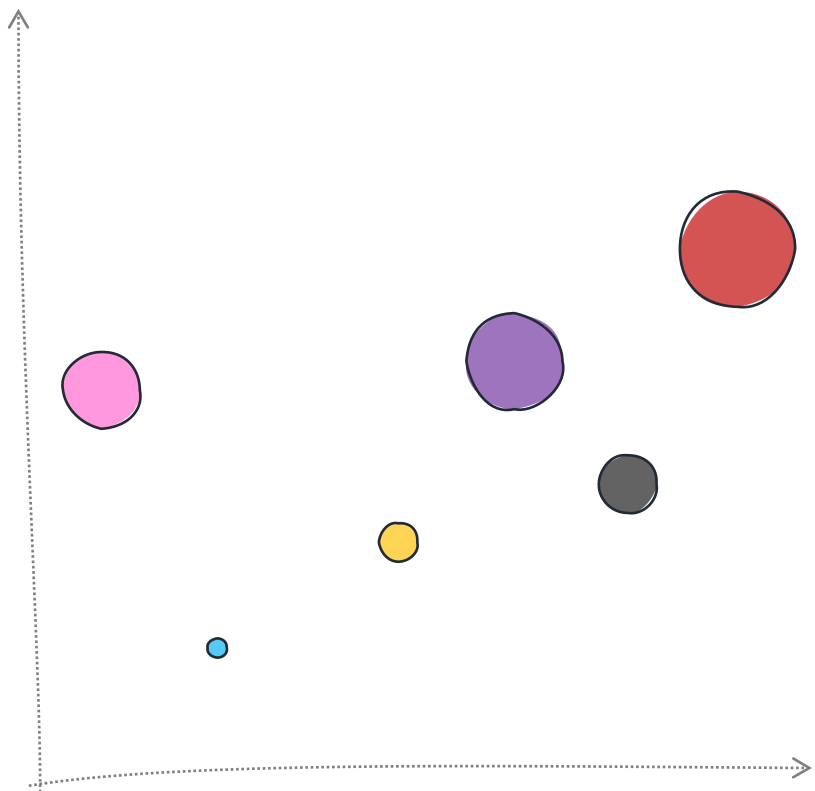
Third Visualization

Additionally, even a radar chart could work! Likewise, given similar caveats as the pie chart, a user may be rapidly able to recognize the existence of a skew towards rare or ultrarare events but with limited knowledge. i.e., missing the count unless augmented with interactivity like mouse hover function, etc.



- *Which near-Earth objects are the biggest and hence deemed hazardous?*

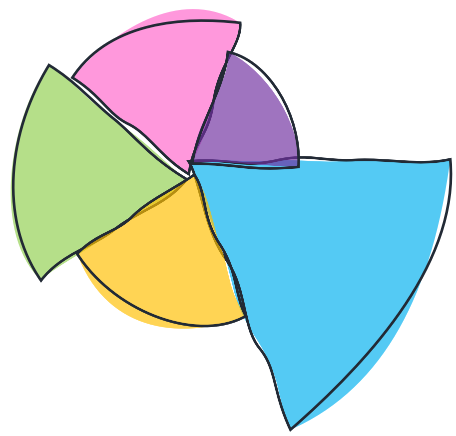
First visualization



A bubble chart works the best here to depict the size of each NEO and let the user quickly realize which specific NEO can be hazardous given their size alone.

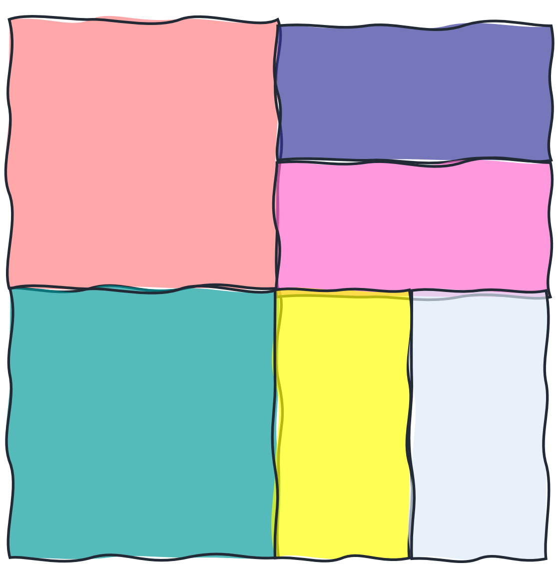
Second visualization

A polar area chart may be able to depict the relative sizes of the biggest NEOs given some truncation to the data, i.e., limit the visualization to the top 10, etc.



Third Visualization

Another useful visualization would be an area graph to quickly scan the sizes of the objects, again limited to a lower count, like the top 10.



Conclusion

Hence, as always, there is no perfect visualization, it all depends on the scope of the question, practicality and what the designer intends to actually convey. However, ultimately to a larger extent, what the user comprehends from the visualization makes or breaks the said visualization. Here, I tried brainstorming a total of nine visualization for the three analytical question proposed.