Data Exploration and Visualisation

Scott Lee

Visualising Near Earth Object Data

Project Proposal

This project will look to create data visualisations of Near Earth Objects (NEO) according to NASA and other research databases. A NEO is defined as an asteroid or comit which passes within 1.3 Astronomical Units (AU) of the Earth, or about 45 million kilometres (Near-Earth Objects, n.d.)

Space is full of an enormous number (as the data will show) of asteroids and comets. Thanks to Science Fiction movies, along with a healthy dose of self-preservation instincts, the general public and the scientific community alike have an interest in understanding and tracking different objects that may be on a collision course with the Earth. Like many areas in science, Hollywood can often portray things to be worse then they are. This project is intended to show that humans, as a species, are relatively safe from an extinction level asteroid collision for the next 200 years or so, which is as far out as the data goes.

Specifically, this project will show the audience:

- the number of NEOs that will pass by the Earth
- NEO velocities
- NEO size
- NEO minimum distance from Earth.

Data Set

The data set used will be from the Centre for Near Earth Object Studies.

https://cneos.jpl.nasa.gov/ca/

A sample of the data is below. I am using data for NEOs in the future and that are projected to pass within 0.05 astronomical units (the distance between the Earth and the Sun) of the Earth resulting in **12,408** NEOs to analyse.

Object	Close-Approach (CA) Date	CA Distance Nominal (LD | au)	CA Distance Minimum (LD | au)	V relative (km/s)	V infinity (km/s)	H (mag)	Diameter
(2022 DX2)	2022-Mar-13 04:32 ± < 00:01	15.69 | 0.04031	15.64 | 0.04019	12.69	12.68	25.0	26 m - 59 m
(2022 DR3)	2022-Mar-13 08:08 ± < 00:01	15.60 | 0.04008	15.51 | 0.03985	12.42	12.42	25.3	24 m - 53 m
(2018 GY)	2022-Mar-13 08:33 ± < 00:01	11.86 | 0.03048	11.86 | 0.03048	10.68	10.67	24.6	32 m - 71 m
(2022 ES3)	2022-Mar-13 19:17 ± 00:04	0.87 | 0.00223	0.86 | 0.00222	18.37	18.30	27.1	10 m - 23 m
(2022 EO4)	2022-Mar-13 19:36 ± 00:09	2.75 | 0.00708	2.74 | 0.00705	15.64	15.61	25.4	22 m - 49 m
(2022 EA1)	2022-Mar-14 04:24 ± < 00:01	8.06 | 0.02072	8.03 | 0.02064	5.46	5.44	27.1	9.9 m - 22 m
(2022 EK4)	2022-Mar-14 15:54 ± 00:08	4.74 | 0.01218	4.71 | 0.01209	8.33	8.31	26.3	15 m - 33 m
(2022 ED4)	2022-Mar-15 14:10 ± 00:11	2.06 | 0.00528	2.04 | 0.00525	17.01	16.98	26.6	13 m - 28 m
(2022 DP3)	2022-Mar-15 16:34 ± 00:02	2.99 | 0.00768	2.97 | 0.00763	10.79	10.76	24.7	31 m - 69 m
(2022 EG3)	2022-Mar-15 18:21 ± 00:26	13.71 | 0.03523	13.47 | 0.03461	6.53	6.52	28.2	6.1 m - 14 m

Preparation

Data preparations for this project are done by passing the csv file with 12408 records in Python with Jupyter Notebooks.

An initial data audit has shown that there were missing values as well as formatting issues which needed to be addressed.

Some particular examples included:

1. Close-Approach (CA) Date needed to be converted to date format with the error interval removed. The error interval was not deemed, in my view, significant enough on the analysis I was trying to conduct need to keep it.
2. The CA Distance Nominal and CA Distance Maximum columns needed to be split into LD (lunar distance) and AU (astronomical units) columns and converted to floats
3. The Diameter column needed to be split into a Diameter Minimum and a Diameter Max column
4. NA values needed to be replaced (they appeared in the Diameter column data)
5. Velocity columns needed to be changed to floats instead of strings.

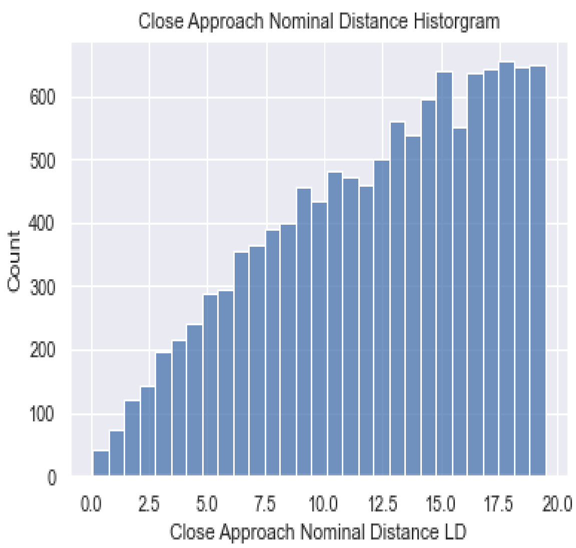
My final data frame looked as the below:

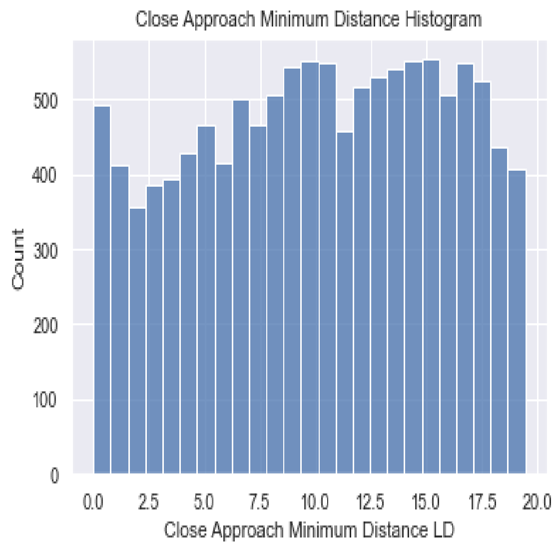
	Object	Close Approach Date and Time	Close Approach Time Error	Close Approach Nominal Distance LD	Close Approach Nominal Distance AU	Close Approach Minimum Distance LD	Close Approach Minimum Distance AU	Diameter Min (m)	Diameter Max (m)	Relative Velocity (km/s)	V Infinity (km/s)	Absolute Magnitude (mag)
0	(2022 DX2)	2022-03-13 04:32:00	00:01	15.69	0.04031	15.64	0.04019	26.0	59.0	12.69	12.68	25.0
1	(2022 DR3)	2022-03-13 08:08:00	00:01	15.60	0.04008	15.51	0.03985	24.0	53.0	12.42	12.42	25.3
2	(2018 GY)	2022-03-13 08:33:00	00:01	11.86	0.03048	11.86	0.03048	32.0	71.0	10.68	10.67	24.6
3	(2022 ES3)	2022-03-13 19:17:00	00:04	0.87	0.00223	0.86	0.00222	10.0	23.0	18.37	18.30	27.1
4	(2022 EO4)	2022-03-13 19:36:00	00:09	2.75	0.00708	2.74	0.00705	22.0	49.0	15.64	15.61	25.4

Exploratory Data Analysis

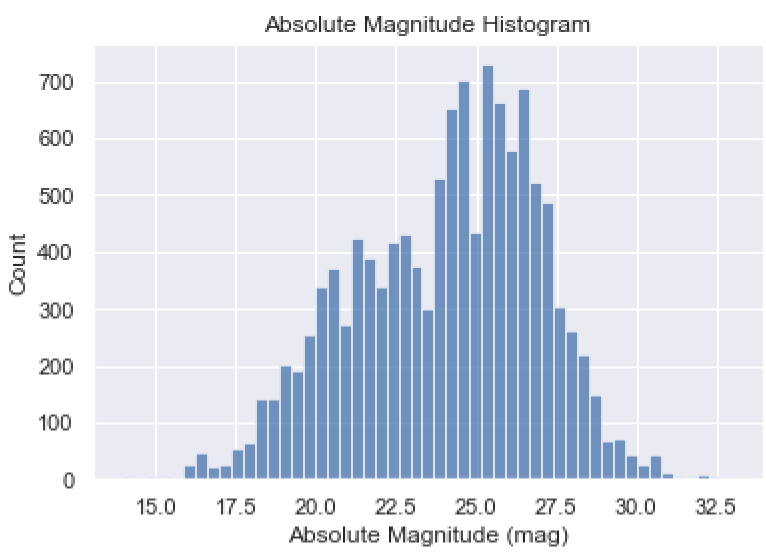
After the data was wrangled, insight was provided to help shape the design (five design sheet) methodology to produce the visualisation.

Close Approach Distance

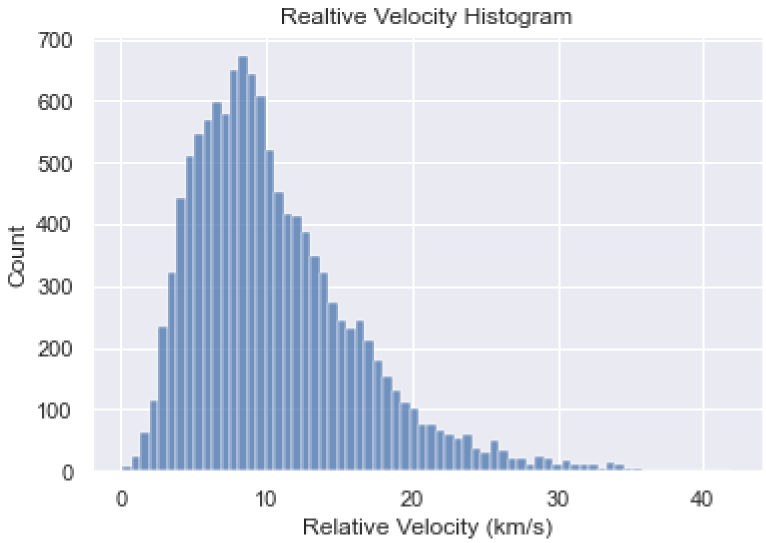




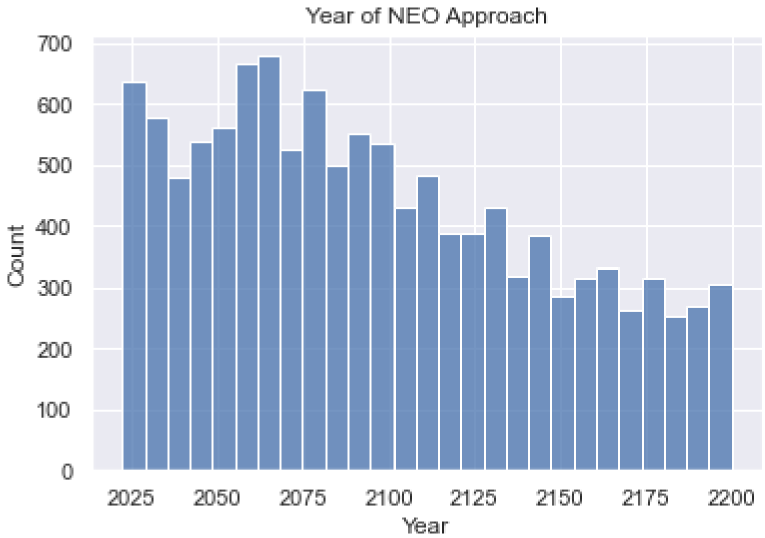
Absolute Magnitude of NEOs



Relative Velocity of NEOs



Year of NEOs



Comments On The Above Graphics

From initial exploration using histograms some interesting questions can be now be asked.

Firstly there seems to be a relatively uniform distribution of minimum distances that NEOs pass to the Earth.

Secondly, the initial data tends to suggest that most of the NEOs are at the smaller (absolute magnitude is an inverse relationship, so a higher number is a smaller NEO) slower ends of the size and velocity scales respectively.

Finally, it looks like there are a lot more NEOs expected to orbit close to the Earth in this century than the next century.

Story To Be Told

The aim of this project is to give the audience a better understanding of NEOs, and how they exist in space relative to the Earth. Specifically, the visualisation will address:

- The number of NEOs each year
- Some physical characteristics e.g. diameter and velocity
- How close they actually get to the Earth

The graphics will be interactive to allow the audience to explore the data in an intuitive fashion, being able to focus on elements of the data that grab their attention. The aim is that after exploring the visualisation, the audience better understand NEOs.

Five Design Sheet Methodology

The five design sheet methodology was adopted for the development of the visualisation. See the end of the document for the five design sheet handwritten notes.

Final Design Considerations

During the exploration of the data and the five design sheet process the following challenges were identified:

1. How to capture so much data in relatively few charts
2. How to make relatively complex physics (orbital mechanics) understood by a non-technical/non-specialist audience.

In doing so the final design was arrived on which contains the three main figures:

1. Histogram showing frequency of NEOs over time
2. Interactive scatter plot showing NEO velocity and closeness to Earth
3. Interactive scatter plot showing NEO maximum diameter and closeness to Earth

NEO Histogram

This histogram simply shows the audience the number of NEOs that pass close to the Earth. A drop down menu was created so that they could view either “Close”, “Mid-Distance” or “Far” NEOs. This was designed to give the audience an appreciation/reference to the number of NEOs being considered. Distance defined as:

- Close < 1 LD
- 1 LD <= Mid-Distance <= 3 LD
- Far >= 3 LD

Note: LD refers to lunar distance i.e. the average distance between the Earth and the Moon.

Relative Velocity Vs Distance To Earth

The chart was chosen to show how fast different NEOs are travelling and how close they get to the Earth. The point was to give the audience an understanding of how much variation there is with speed and also to give them an appreciation of how different the minimum distances can be.

A reduced data set of NEOs within 1 LD was used to avoid oversaturating the plot with too much data.

NEO Diameter and Minimum Distance to Earth

The purpose of this graphic is to show the audience the variation in size of the different NEOs. Inferring the mass of the NEOs can be difficult, so the maximum diameter is a useful gauge the size of the body. This is plotted against minimum distance to Earth to highlight how close they get to Earth.

Again, a reduced data set of NEOs within 1 LD was used to avoid oversaturating the plot with too much data.

Code

I extensively used the course content as well as the Shiny tutorials on shiny.rstudio.com to develop my code. These resources allowed me to find and adapt visualisations that were a similar style to what I wanted to create in my use of the five design sheet methodology.

Running the Visualisation Using Shiny.

To run this visualisation, the user will need:

- The NEO data file, which I have called “neo_data.csv”
- RStudio
- An internet connection.
- The R script which consist of two files;
    - ui.R
    - server.R

When the shiny script is open the “neo_data.csv” **must** be in the same directory and the name not changed.

The following packages need to be installed prior to running the script:

- shiny
- ggplot2
- tidyverse

If they are not currently installed then they need to be by using:

- install.packages(“shiny”)
- install.packages(“ggplot2”)

- install.packages(“tidyverse”)

Once the above is set up, the shiny script can simply be run in the RStudio console and the interactive visualisation will appear.

The visualisation video demonstration can be found at: https://youtu.be/E1ZEFjP0xPA. Please give it a moment for the image to render so the text can be read.

Conclusion

This project was a very interesting exploration of data which taught me a lot about setting up a web visualisation. The main challenge was being able to distil a broad range of information into easily understandable graphics for the user. This is very typical of the challenges of being a Data Scientist, in that you are trying to get easily expressible insights from complex data.

I think it does broadly achieve its goals in terms of enabling the audience to gain a better appreciation for the number of NEOs, as well as understanding some of their physical characteristics. Importantly, it will hopefully put any anxious minds at ease that the scenarios displayed in Hollywood movies (large Earth ending asteroids on a collision course for Earth) are just science fiction, and not something that the average person needs to be concerned about.

If I was to do this project again, I would want to find a good approximation for NEO mass so I could show the potential impact energy of NEOs. I think that would be a good visualisation. It would also be good to have been able to create some sort of animation, so that the audience could visualise the later two charts over time.

References

1) *NEO Earth Close Approaches*. (2022). Centre for Near Earth Object Studies. https://cneos.jpl.nasa.gov/ca/
2) *Near-Earth Objects*. (n.d.). Www.unoosa.org. https://www.unoosa.org/oosa/en/ourwork/topics/neos/index.html#:~:text=A%20near%2DEarth%20object%20is
3) *Shiny - Gallery*. (n.d.). Shiny.rstudio.com. https://shiny.rstudio.com/gallery/