Contents lists available at ScienceDirect

Applied Computing and Geosciences

journal homepage: www.sciencedirect.com/journal/applied-computing-and-geosciences

Using a 3D heat map to explore the diverse correlations among elements and mineral species

Jiyin Zhang[a], Xiang Que[a], Bhuwan Madhikarmi[b], Robert M. Hazen[c], Jolyon Ralph[d], Anirudh Prabhu[c], Shaunna M. Morrison[c], Xiaogang Ma[a,c,*]

[a] *Department of Computer Science, University of Idaho, Moscow, ID, 83844, USA*
[b] *12 N Taylor Ave, Norwalk, CT, 06854, USA*
[c] *Earth and Planets Laboratory, Carnegie Institution for Science, Washington, DC, 20015, USA*
[d] *Hudson Institute of Mineralogy, Keswick, VA, 22947, USA*

ARTICLE INFO

Keywords:
Exploratory data analysis
Data science
Data visualization
Correlation analysis
Mineral informatics

ABSTRACT

This paper presents an enhanced 3D heat map for exploratory data analysis (EDA) of open mineral data, addressing the challenges caused by rapidly evolving datasets and ensuring scientifically meaningful data exploration. The Mindat website, a crowd-sourced database of mineral species, provides a constantly updated open data source via its newly established application programming interface (API). To illustrate the potential usage of the API, we constructed an automatic workflow to retrieve and cleanse mineral data from it, thus feeding the 3D heat map with up-to-date records of mineral species. In the 3D heat map, we developed scientifically sound operations for data selection and visualization by incorporating knowledge from existing mineral classification systems and recent studies in mineralogy. The resulting 3D heat map has been shared as an online demo system, with the source code made open on GitHub. We hope this updated 3D heat map system will serve as a valuable resource for researchers, educators, and students in geosciences, demonstrating the potential for data-intensive research in mineralogy and broader geoscience disciplines.

1. Introduction

Mineralogy, like many other geoscience disciplines, faces the opportunities enabled by the fast-growing open data facilities and data science methods (Hazen et al., 2019). For example, in recent years, significant data-driven scientific discoveries have been achieved in the field of mineral ecology (Hystad et al., 2015, 2019), mineral evolution (Hazen and Ferry, 2010; Morrison et al., 2017), as well as the co-evolution between geosphere and biosphere (Hazen and Papineau, 2012; Dong et al., 2022). As a result of these converging efforts, mineral informatics (Prabhu et al., 2023a, 2023b) has been proposed as a field that leverages cyberinfrastructure, data science, and informatics to discover patterns in various datasets relevant to the study of mineralogy. A unique topic in mineral informatics is exploratory data analysis (EDA), highlighted as an effective way to tackle the challenges caused by big open mineral data (Ma et al., 2017; Ma, 2023). EDA aims to gain a quick view of target datasets and build plausible insights (Tukey, 1977). In a data science workflow, EDA is treated as a functional step for generating hypotheses about the dataset under study (Schutt and O'Neil, 2013; Ma et al., 2017). EDA employs various data visualization methods, including histograms, pie charts, scatter plots, and box plots, alongside statistical techniques like linear regression, to facilitate comprehensive data analysis.

The heat map is a standard visualization method for illustrating complex relationships among large groups of factors (Wilkinson and Friendly, 2009). In recent years it has been increasingly used as an EDA method in the research of chemical elements and mineral species (Feltrin and Bertelli, 2020; Emami and Emami, 2020; Carvalho et al., 2022; Hazen et al., 2023a). In our previous work, an initial three-dimensional (3D) heat map (i.e., 3D Klee diagram) has been built by Ma et al. (2017) to visualize the correlations in mineralogy, such as the co-existence of elements among mineral species and co-occurrence of mineral species among localities. Three case studies were then analyzed using the mineral species list approved by the International Mineralogical Association (IMA; rruff.info/ima). Overall, the study of Ma et al. (2017) demonstrated the potential of using visualization techniques in the EDA step of a data science process.

Nevertheless, that work was based on static datasets of correlations.

[*] Corresponding author. Department of Computer Science, University of Idaho, Moscow, ID, 83844, USA.
E-mail address: max@uidaho.edu (X. Ma).

https://doi.org/10.1016/j.acags.2024.100154
Received 30 August 2023; Received in revised form 22 December 2023; Accepted 3 January 2024
Available online 5 January 2024

It could not easily be updated along with the fast-growing IMA mineral species list (IMA approves about 100 new mineral species annually). The outdated datasets reduce the utility of the demo system built by Ma et al. (2017). Moreover, the user interface of the original demo system needs to add more flexibility: it merely listed the elements for data selection while needing more scientific guidance and explanation from the perspective of mineralogy. The 3D heat maps demonstrated the co-existence of element triplets among all mineral species, yet more scientifically meaningful operations for element selection and sub-setting can further facilitate the interpretation of EDA results and add more scientific value to the demo system.

The OpenMindat project (Ma et al., 2023) provides an opportunity to solve the outdated mineral data mentioned earlier. Mindat (mindat.org) is a crowd-sourced data website that collects and shares records of mineral species and their corresponding attributes, such as chemical formulas, classification systems, localities, and more. As of February 2023, Mindat has over 5884 IMA-approved mineral species, over 389,000 localities, over 1,463,000 mineral occurrences, and much other relevant geology, petrology, and paleontology information. Those abundant records make Mindat one of the best resources for retrieving mineralogical datasets that match the latest IMA mineral species list. The OpenMindat project has established an application programming interface (API) that allows users to query and download data. This API allows us to construct an automatic workflow to retrieve and cleanse up-to-date open mineral data from Mindat and feed them into the 3D heat map.

On the other hand, studies in mineralogy provide many clues and resources to address the need for more scientifically-sound operations in the demo system. For example, we can obtain structured records of mineral classification and mineral-element correlation from the Dana Classification (Gaines et al., 1997), Hey's Mineral Index (Clark, 1993), and the recently proposed Evolutionary System of Mineralogy (Hazen, 2019; Hazen et al., 2023a). Reviewing those geoscience literature resources will help us retrieve mineral species' structured classification and formational contexts. We can also quickly obtain each mineral species' chemical formula and element list. Using those structured, scientifically meaningful records, we can establish a list of new operations at both the demo system's data selection and visualization stages.

This paper presents our work on an updated 3D heat map for the EDA of open mineral data, highlighting how we solved the earlier issues of out-of-date data and scientifically-meaningful data operations. With those updates, we hope the demo system will be a long-lasting and helpful resource for researchers, educators, and students in geosciences. We have also shared the code and data on GitHub for interested researchers to adapt and extend. In the remainder of the paper, Section 2 describes the workflow and structure of the updated 3D heat map demo system, the live data resource from the OpenMindat API, and the new data selection and visualization operations. Section 3 presents a detailed illustration of the demo system on how EDA of mineral data can be performed through the functions provided. This section will utilize two use cases, one on igneous minerals and the other on significant elements in mineral species. Section 4 discusses the scientific value of this work from the perspective of data science and mineral informatics and presents some thoughts on future extensions. Finally, Section 5 concludes the paper.

2. Methods and datasets

2.1. Overview of the workflow

With the live data service from the OpenMindat Data API (Ma et al., 2023), we designed the workflow for the 3D heat map (Fig. 1). The workflow consists of four significant steps: data retrieving, data cleansing, data selection, and the visual exploration based on 3D heat map results. The last step adapts the code from Ma et al. (2017) and extends the new interactive operations on the 3D heat map result, such as rendering and filtering. The first three steps are all new developments centered on the new data source enabled by the OpenMindat Data API. Python was used for data retrieving and cleansing functions, and JavaScript was used for data selection and heat map generation functions.

The derived datasets in the workflow and the critical packages used are also shown in Fig. 1. The OpenMindat API provides a list of parameters that can be used as filters to retrieve names and attributes of mineral species approved by the International Mineralogical Association (IMA). For example, in this study, we have retrieved records of the co-existence of elements in mineral species to illustrate the workflow and use cases. The obtained raw data from the API are stored in JSON (JavaScript Object Notation) format. Then, the raw data are parsed and restructured into a specific CSV (Comma-Separated Values) format to represent the co-existence of three elements in mineral species. In our workflow, we convert JSON to CSV format to align with the technical needs of our visualization tools, initially developed for CSV input. This transformation also provides a user-friendly format for researchers, allowing easier inspection and comparison of element distributions within the datasets.

The structure of the CSV is designed in this way to make it easy for 3D heat map visualization. To build interactive data selections from the cleansed CSV file, two JavaScript packages, Papaparse.js and Alasql.js, are used. Papaparse.js is an effective and convenient CSV parser that can load and convert the CSV file into a more interactable data object. The structured CSV files enable data selection on a user interface, such as using a visualized periodic table or a dropdown list of predefined element compositions. After selecting, the picked elements will be applied as querying parameters in the Alasql.js package to obtain corresponding element subsets from the data object.

Once the data selection is successfully run, the extracted element subsets will be rendered as a 3D heat map using the Three.js package (Fig. 2). Each cube in the heat map refers to a triplet of elements (on X, Y,

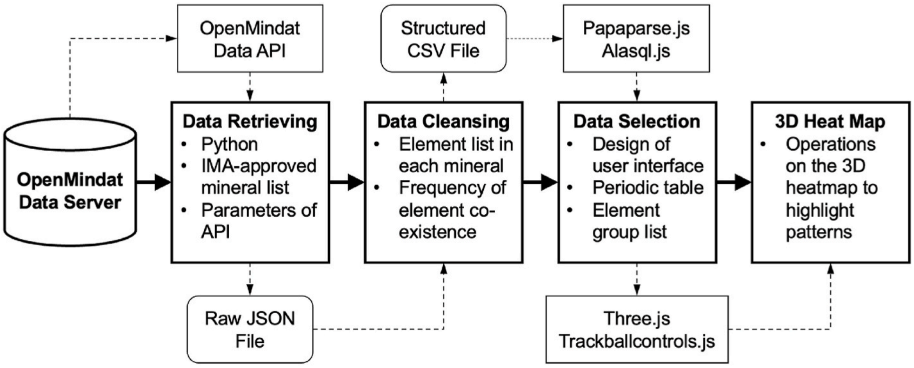


Fig. 1. Workflow of the developed 3D heat map for EDA of open mineral data.

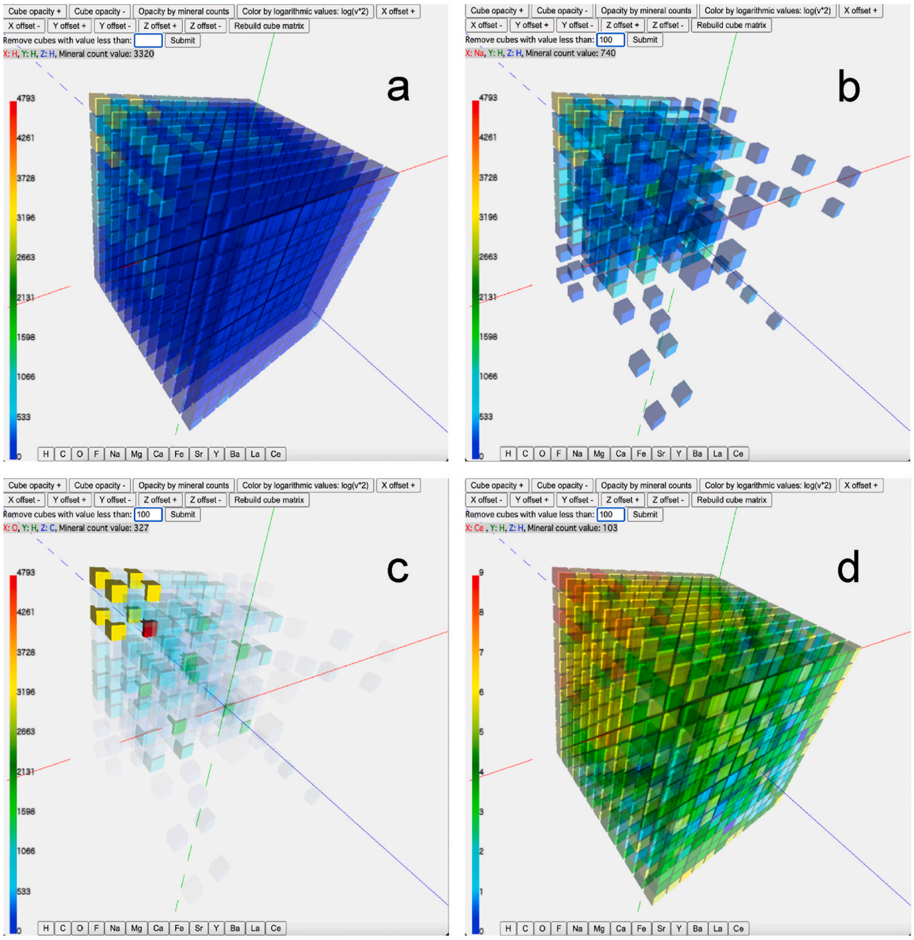

a
b
c
d
Cube opacity +
Cube opacity -
Opacity by mineral counts
Color by logarithmic values: log(v*2)
X offset +
X offset -
Y offset +
Y offset -
Z offset +
Z offset -
Rebuild cube matrix
Remove cubes with value less than:
100
Submit
X: H, Y: H, Z: H, Mineral count value: 3320
X: Na, Y: H, Z: H, Mineral count value: 740
X: O, Y: H, Z: C, Mineral count value: 327
X: Ce , Y: H, Z: H, Mineral count value: 103
4793
4261
3728
3196
2663
2131
1598
1066
533
0
9
8
7
6
5
4
3
2
1
H
C
O
F
Na
Mg
Ca
Fe
Sr
Y
Ba
La
Ce

Fig. 2. Interactive operations on the 3D heat map. (a) Initial color rendering of selected element subset; (b) Set a cut-off value to remove unwanted cubes; (c) Set opacity of cubes based on values; and (d) Re-rendering the color using a logarithmic transformation. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

and Z axes). In Fig. 2a, the initial color rendering corresponds to the value in each cube, which is the number of mineral species in which the three elements co-exist, with blue for the lowest value and red for the highest. Accompanied by Three.js, the Trackballcontrols.js package controls the interactive operations on the 3D heat map result. The basic operations include changing camera or object positions, tilting camera directions, and zooming in and out (Ma et al., 2017). In this study, we enhanced the visualization availabilities by extending new interactions, such as removing unexpected cubes and featuring the primary ones, adjusting the cube opacities, re-rendering the cubes with different algorithms to highlight value variance, and more (Fig. 2b, c, and d).

2.2. *Stable and live mineral data source*

The OpenMindat API data used in this study is sourced from the Mindat website, one of the world's largest crowd-sourced mineral databases. In recent years, the management team of Mindat has faced fast-increasing open data requests from various researchers and organizations. Although the Mindat website is open and accessible for browsing, before the Spring of 2023, there was no machine-readable interface for data query and access. The OpenMindat data server addresses those needs. OpenMindat is a refactoring version of the Mindat database for effective and convenient machine accessibility (Ralph et al., 2022; Ma et al., 2023). The work on OpenMindat includes a list of data enriching and integration and server construction activities towards a stable and live data API that enables machine query and access. From a broad perspective, the OpenMindat efforts follow the FAIR principles (Findable, Accessible, Interoperable, and Reusable) (Wilkinson et al., 2016) to facilitate a smooth data science workflow from the machine interoperability and accessibility of data sources to the data reusability at the users' end. For this study, the OpenMindat data server provides a stable, live, and up-to-date mineral data source that can refresh the generated 3D heat map and extend the development of new functions. The OpenMindat API, characterized as 'stable' in this manuscript, reflects our commitment to providing a reliable data service. While instances of downtime are inevitable in any online system, we have established a rigorous maintenance protocol to ensure these are quickly resolved. Our dedication to swift action in the face of such challenges underpins our promise of stability to our users.

The accessible data from OpenMindat API cover a series of subjects, including but not limited to mineral species, mineral classification, petrological classification, localities, mineral occurrences, and more. The corresponding data records can be obtained through a list of querying parameters on the API, such as “name (of the mineral species)”, “mindat_formula”, “ima_status”, “groupid”, “elements”, “occurrence”, and the classification codes from Nickel-Strunz and Dana classification systems. The OpenMindat API can fulfill different data queries and export data in the open-source JSON format (Ma et al., 2023). Users can conduct thematic data query tasks by combining different querying parameters and filtering conditions on the OpenMindat API, such as retrieving all the IMA-approved mineral species that incorporate copper, sulfur, and oxygen into a specific part of the petrological classification system.

A few use cases of data query on OpenMindat API and how the retrieved data are used to create the 3D heat map will be demonstrated. The focus of these use cases is on the element co-existence among all the IMA-approved mineral species. So far, a workflow to automatically perform the data retrieval and cleansing has been established (see Fig. 1). While the IMA-approved mineral species list keeps growing and the records on the OpenMindat data server are continuously updating, the 3D heat map can always have up-to-date data input through our automated data retrieving and cleansing workflow. Moreover, the OpenMindat API will also gradually expand the scope and accessibility of data subjects from the original Mindat database, enabling more exciting scientific topics to be explored through the 3D heat map and many other EDA techniques.

2.3. *Scientifically meaningful data operations on the user interface*

The OpenMindat API can output many datasets to illustrate mineralogy correlations. So far in this study, we have focused on the co-existence of elements in mineral species, and we have generated a list of datasets for the updated 3D heat map demo. The datasets consist of all 73 mineral-forming elements and the 30 primary mineral-forming elements/element groups. The detailed list of datasets and descriptions are made open online, and the links can be found in this paper's Code and Dataset Availability section. Since the publication of Ma et al. (2017), we have received many comments suggesting that more interactive, visualized, and scientifically-sound operations should be provided for users to select a subset of data to visualize in the heat map rather than visualize the whole 73ˆ3 or 30ˆ3 matrix. In this study, we have developed corresponding operations on the user interface to address those comments (Fig. 3).

As depicted in Fig. 3a, the user interface is structured around four primary operations:

a) Dataset Loading: Upon uploading the dataset labeled '73_elements' (indicative of its 3D matrix structure of 73ˆ3), elements within it are prominently highlighted in black on the periodic table, while non-included elements appear grey.
b) Element Selection via the Periodic Table: Users select elements by first clicking an empty box under each axis, then choosing elements from the periodic table to populate the axis box. Once all axes have their chosen elements, users can either generate a 3D heat map by clicking "Visualize Selected Elements" or download the subset data as a CSV via "Download Subset Data."
c) Element Selection by Mineral Classes: Once the 73ˆ3 dataset is selected, a dropdown menu activates, titled "Select Element Combinations by Mineral Classes." This menu primarily lists igneous mineral classes. By selecting a mineral class, the relevant element combinations will automatically populate an axis. For instance, in Fig. 3a, selecting the "oxides" mineral class for all three axes yields the heat map shown in Fig. 3b.
d) Visualization: The 3D heat map, as presented in Fig. 3b, visualizes correlations between elements across the X, Y, and Z axes, where each colored cell signifies the number of mineral species in which the triad of elements co-exist. This heat map offers a potent way to discern patterns, especially when axes showcase different mineral class combinations.

For a deeper dive into specific use cases, datasets, and interactive operations, readers can refer to Section 3. Here, we will further elucidate the system's capabilities and the scientific insights gleaned from its exploratory data analysis (EDA).

3. Use cases, results and analysis

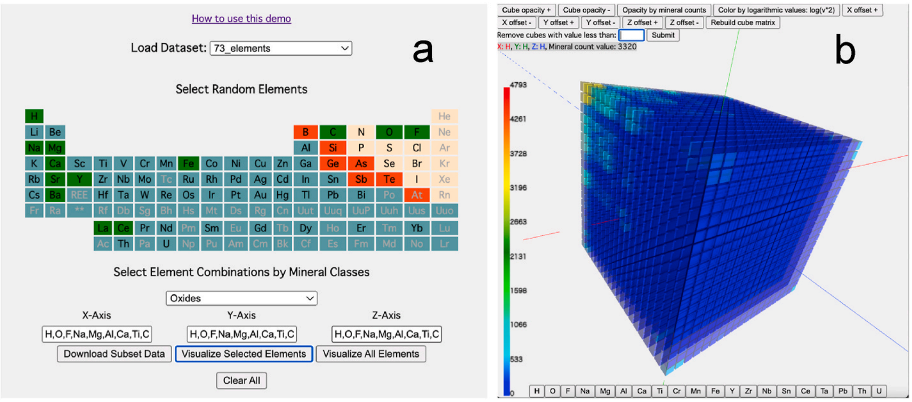
Using datasets retrieved from the OpenMindat API, two typical use cases were built to demonstrate the utility of the 3D heatmap demo system to support EDA in mineralogy. The datasets were collected in early February 2023, based on the 5899 mineral species recorded by Mindat at that time. A list of eight datasets was cleansed for the use cases. As automated workflows were established for the data retrieving and cleansing steps, all the use cases can be easily updated with the latest OpenMindat datasets.

3.1. *Element co-existence in primary igneous minerals*

The first use case centered on the co-existence of elements among igneous minerals. Igneous minerals are a topic of long-term interest in mineralogy.

Bowen (1928) proposed the "reaction series" to study igneous mineral co-occurrence, which can be applied to explain and predict the fractional crystallization sequence as silicate magma's temperature decreases. For example, the continuous reaction series of olivines → pyroxenes → amphiboles → biotites indicate well-attested igneous mineral serial co-occurrences in many rock varieties. In addition to the fractional crystallization caused by temperature, igneous mineral species demonstrate association patterns based on chemical factors, such as quartz-alkali feldspar pair, hornblende-intermediate plagioclase pair, and more. We prepared two data objects to further study element co-existence among igneous minerals: the igneous mineral list and the element-mineral correlation dataset.

We referred to Hazen et al. (2023a) for the igneous mineral list. When determining igneous minerals, their work adopted the Dana Classification System for categorizing minerals in this research, a method also used by Mindat.org based on mineral chemical composition. They conducted a co-occurrence analysis of igneous minerals using


How to use this demo
a
Select Random Elements
H He
Li Be B C N O F Ne
Na Mg Al Si P S Cl Ar
K Ca Sc Ti V Cr Mn Fe Co Ni Cu Zn Ga Ge As Se Br Kr
Rb Sr Y Zr Nb Mo Tc Ru Rh Pd Ag Cd In Sn Sb Te I Xe
Cs Ba REE Hf Ta W Re Os Ir Pt Au Hg Tl Pb Bi Po At Rn
Fr Ra ** Rf Db Sg Bh Hs Mt Ds Rg Cn Uut Uuq UuP Uuh Uus Uuo
La Ce Pr Nd Pm Sm Eu Gd Tb Dy Ho Er Tm Yb Lu
Ac Th Pa U Np Pu Am Cm Bk Cf Es Fm Md No Lr
Select Element Combinations by Mineral Classes
Download Subset Data
Visualize Selected Elements
Visualize All Elements
Clear All
Cube opacity +
Cube opacity -
Opacity by mineral counts
Color by logarithmic values: log(v*2)
X offset +
X offset -
Y offset +
Y offset -
Z offset +
Z offset -
Rebuild cube matrix
Submit
b
4793
4261
3728
3196
2663
2131
1598
1066
533
0
H O F Na Mg Al Ca Ti Cr Mn Fe Y Zr Nb Sn Ce Ta Pb Th U

Load Dataset: 73_elements

Oxides

X-Axis: H,O,F,Na,Mg,Al,Ca,Ti,C

Y-Axis: H,O,F,Na,Mg,Al,Ca,Ti,C

Z-Axis: H,O,F,Na,Mg,Al,Ca,Ti,C

Remove cubes with value less than:

X: H, Y: H, Z: H, Mineral count value: 3320

Fig. 3. Scientifically meaningful data operations. (a) User interface for data loading and selection, and (b) The 3D heat map result when X, Y, Z axes are the same list of elements from the mineral class "oxides".

a hierarchical clustering heat map. Their results delineated a clear association pattern among chemical composition-based mineral classes. Noticing that the recognized igneous minerals in the work are related to chemistry-based mineral classes, we felt that the element combinations of mineral species could be used to explore the correlations between mineral species further. Accordingly, we built a table to show the element combination of each igneous mineral class (Table 1), and we used them in the user interface of the demo system (Fig. 3a).

In our analysis, the selected element combinations associated with igneous minerals aim to highlight the distribution of elements within IMA minerals. This method does not strictly categorize minerals based on their geological origins but rather investigates potential patterns in elemental makeup across various mineral classes, acknowledging the diverse geological processes that contribute to the formation of these minerals.

All the 73 mineral-forming elements were sourced from OpenMindat, then the information was processed into various 3D matrices. OpenMindat's API offers attributes like "elements", "sigelements", and "imaformula", which define the chemical composition of mineral species. For example, the "elements" attribute for the mineral species "aeschynite-(Ce)" lists as "-Ca-Ce-Fe-Nb-Th-Ti-O-H-", while the "sigelements", or major structural site elements, are "-Ce-Ti-O-". The latter is derived from the main dominant elements in lattice positions in the chemical formula of Mindat. Documentation for defining API attributes can be accessed in the Code and Dataset Availability section.

We then structured two primary 3D matrices: "73_elements" and "73_sigelements". Both matrices use the same 73 elements on the X, Y, and Z axes. Each cell within these matrices indicates the number of mineral species with the corresponding X, Y, and Z elements. These matrices were generated using the "elements" and "sigelements" attributes from the 5899 IMA-approved mineral species.

Further, we developed two derived 3D matrices: "normalized _73_elements" and "normalized_73_sigelements". Here, each cell value denotes the fraction of mineral species containing the Z element, which also simultaneously includes X and Y elements. In mathematical terms, it's represented as (number of species with X, Y, and Z elements) ÷ (number of species with Z element).

These four datasets can be accessed on our demo system. Users can load them to create 3D heat maps as demonstrated in Fig. 4. Specifically, Fig. 4a showcases the "73_elements" dataset, visualized as a comprehensive 73ˆ3 matrix heat map. The color intensity in each cell corresponds to its value, with a reference color bar situated to the screen's left. However, considering the sheer size of such a matrix, it is more practical for users to generate a smaller 3D heat map from a dataset subset. Conveniently, we have integrated an igneous mineral list (seen in Table 1) into the data selection phase, aiding users in this process.

Table 1
Primary igneous mineral classes and corresponding element combinations.

Classes of Igneous Minerals (From Hazen et al., 2023a)	Element Combinations
Native Elements	C
Sulfides	S, Fe, Cu, Zn, Mo, Pb
Oxides	H, O, F, Na, Mg, Al, Ca, Ti, Cr, Mn, Fe, Y, Zr, Nb, Sn, Ce, Ta, Pb, Th, U
Halides	F, Ca
Carbonates	H, C, O, F, Na, Mg, Ca, Fe, Sr, Y, Ba, La, Ce
Sulphates	O, S, Ba
Phosphates	H, Li, O, F, Al, P, Cl, Ca, Fe, As, Y
Nesosilicates or Orthosilicates	H, Li, Be, O, F, Mg, Al, Si, S, Ca, Ti, Mn, Fe, Y, Zr, I, Th
Sorosilicates or Disilicates	H, Be, O, F, Na, Mg, Al, Si, K, Ca, Ti, Cr, Mn, Fe, Sr, Y, Zr, Nb, Ba, W
Cyclosilicates	H, Li, Be, B, O, F, Na, Mg, Al, Si, Cl, Ca, Fe, Zr, Cs
Inosilicates	H, Li, O, F, Na, Mg, Al, Si, Cl, K, Ca, Ti, Mn, Fe, Zr, Nb, Cs
Phyllosilicates	H, Li, O, F, Mg, Al, Si, Cl, K, Ti, V, Cr, Mn, Fe, Rb, Cs
Tectosilicates	H, C, O, Na, Mg, Al, Si, S, Cl, K, Ca, Fe, Cs

Fig. 4a illustrates a 3D heat map comprising the entire list of elements. Leveraging the dropdown menu featuring igneous mineral classes, we can swiftly choose element combinations for the X, Y, and Z axes. For instance, extending upon Hazen et al. (2023)'s research on primary igneous mineral clustering associations, we executed an Exploratory Data Analysis (EDA) by selecting elements from carbonates, sulfides, and oxides mineral classes for the X, Y, and Z axes, respectively, resulting in Fig. 4b.

The interactive demo allows users to manipulate the 3D heat map, offering rotation, zoom, and even a feature to produce a 2D projection by selecting an element (the chosen element forms the Z axis). A prime example is the right segment of Fig. 4b, focusing on the Z-axis element "Fe". However, Fig. 4a and b reveal that many cells, colored in blue, represent minimal or zero values due to sparse element co-existence records. To enhance visibility, we can apply a logarithmic transformation (log(value*2)) to each cell, ensuring more apparent distinctions between cells, as depicted in Fig. 4c.

One challenge is the obscured cells within the 3D heat map's core. To address this, beyond the slicing function demonstrated in Fig. 4b, we can increase spacing between cell layers along each axis, as shown in Fig. 4d. This adjustment allows for a more precise comparison of patterns, highlighting, for example, the pronounced value (red) of the cell where the element 'Fe' is consistent on all axes. Moreover, by implementing a cut-off value, we can prune the 3D heat map, eliminating less significant cells. This technique yields Fig. 4e (from the "73_elements" dataset) and f (from the "73_sigelements" dataset), both using carbonates, sulfides, and oxides for the X, Y, and Z axes. Any cell below the value of 45 is omitted, with opacity enhanced for the remaining cells to accentuate differences. Notably, Fig. 4e retains more cells than Fig. 4f. The "73_sigelements" dataset, focusing on significant elements, reduces the count of mineral species with X, Y, and Z element co-existence. Applying a consistent cut-off value of 45 to both heat maps results in Fig. 4f having more cells removed.

In Fig. 4e and f, the axes are defined by the elements from the carbonates, sulfides, and oxides mineral classes. Notably, the first two axes encapsulate the components of the Ca–Mg carbonatite mineral community. This community integrates 11 of the 13 primary igneous carbonates and 5 of the 6 sulfides, as cataloged in the definitive list of 115 prevalent primary igneous mineral species by Hazen et al. (2023). The significance of this arrangement becomes particularly evident when scrutinizing the high-frequency elements, where the deep-colored cubes act as beacons of chemical association among this mineral community and oxides. The pronounced clustering along the sulfides axis in Fig. 4e and f provide unequivocal evidence of intrinsic chemical associations between the carbonates and oxides classes.

Contrastingly, the expansive void observed in the anterior segment of the oxides axis tells its own compelling story. This noticeable absence of prominent triplets signifies chemical element antipathies. A meticulous examination of these visualizations elucidates certain elements in the oxides class—specifically, Cr, Y, Zr, Sn, Ta, and Th—that appear to be geochemically incompatible with the carbonatite mineral community.

Furthermore, the filtering capability of the model illuminates rare triplet combinations within the extensive dataset of 5800 IMA-approved mineral species. To illustrate, the unique triplet "Ce, Mo, Fe" has been documented only once, associated with the mineral "tancaite-(Ce)", a secondary mineral recently identified in cavities within quartz veins (Bonaccorsi and Orlandi, 2020). Such infrequent or entirely absent triplets offer an invaluable, data-driven avenue for recognizing geochemically improbable element combinations while alerting mineralogists to the prospect of discovering novel mineral species. This analytical approach is thus promising, not only in showcasing how visual analytics can contribute to understanding known occurrences, but also to the expansion of our mineralogical knowledge base.

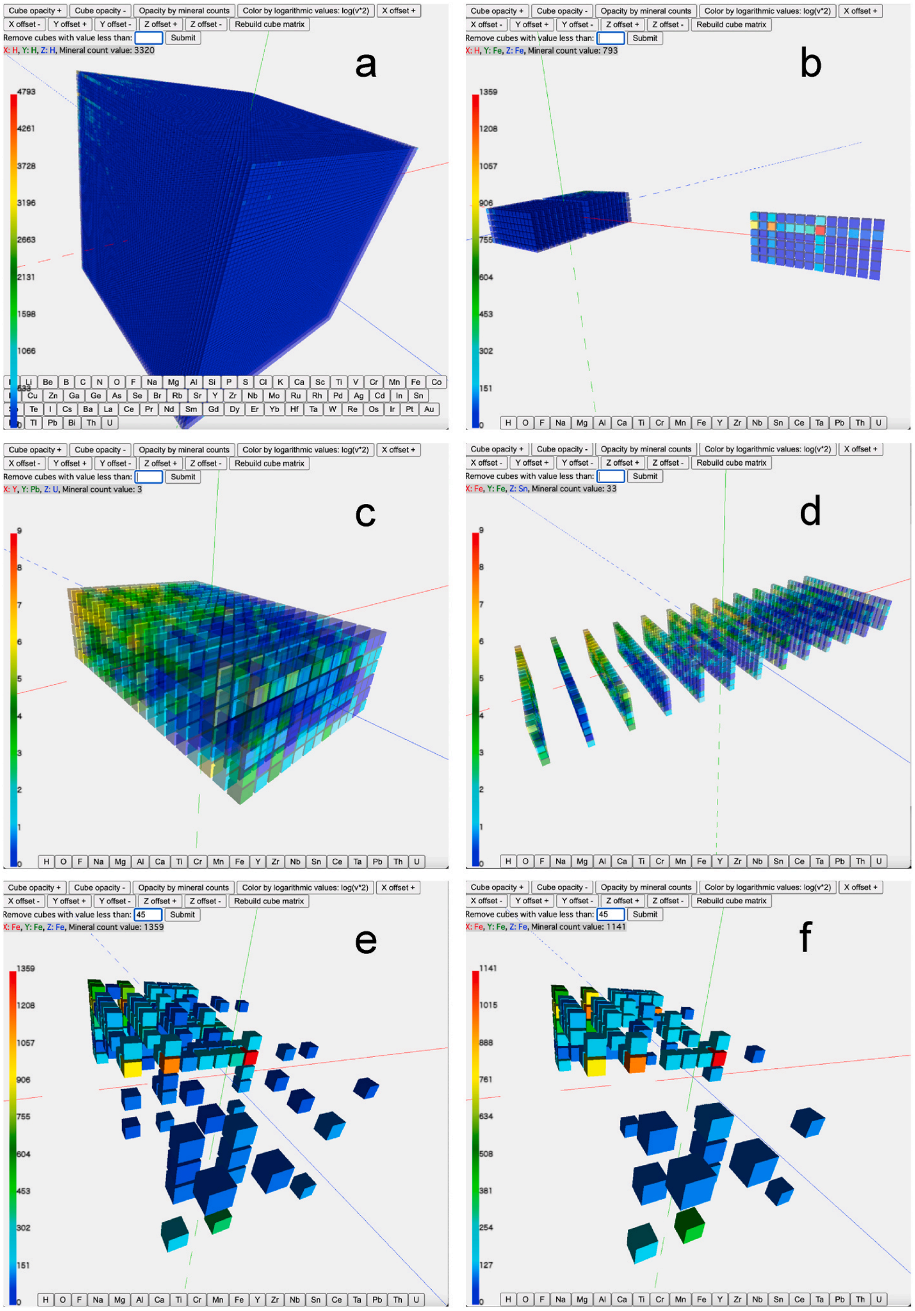

a
b
c
d
e
f
Cube opacity +
Cube opacity -
Opacity by mineral counts
Color by logarithmic values: log(v^2)
X offset +
X offset -
Y offset +
Y offset -
Z offset +
Z offset -
Rebuild cube matrix
Remove cubes with value less than:
Submit
X: H, Y: H, Z: H, Mineral count value: 3320
X: H, Y: Fe, Z: Fe, Mineral count value: 793
X: Y, Y: Pb, Z: U, Mineral count value: 3
X: Fe, Y: Fe, Z: Sn, Mineral count value: 33
X: Fe, Y: Fe, Z: Fe, Mineral count value: 1359
X: Fe, Y: Fe, Z: Fe, Mineral count value: 1141
45

Fig. 4. Different ways to explore the patterns of element co-existence among igneous minerals. (a) The overview of 73 elements 3D heat map cubes; (b) The 3D heat map of carbonates, sulfides, and oxides in the corresponding X (red), Y (green), and Z (blue) axis, with the slicing out of Fe on the Z axis; (c) Re-coloring the cubes with the logarithmic values; (d) Expanding the 3D heat map along X axis; (e) Applying the filter function and removing cubes with fewer than 45 values in the 73_elements dataset; (f) Applying the filter function and removing cubes with fewer than 45 values in the 73_sigelements dataset. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

3.2. Significant elements in mineral species

Upon initial analysis, we discerned considerable sparsity in the 73-element co-existence dataset, where many matrix cells displayed negligible values. A notable observation was the diminished values exhibited by the Rare Earth Elements (REE) cells, in contrast to those representing more prevalent mineral-forming elements. This disparity raised concerns over the potential exclusion of REE cells during the filtration operations, despite their intrinsic significance.

As highlighted by Hazen and Morrison (2022), while the rare chemical elements such as REE, platinum-group elements, As, Mo, and Sn constitute a trivial fraction of the Earth's crust (approximately 0.01%), they are manifest in nearly 40% of identified mineral species. This observation underscores the need for a nuanced approach to our dataset construction.

In our second use-case analysis, focusing on the 30 most predominant elements in IMA-approved mineral species, we gave special attention to REEs. While scandium (Sc) is often grouped with REEs, we acknowledge its distinct geochemical behavior and rare substitution in REE minerals. For the purposes of this study, Sc was included in the REE category to demonstrate our methodology's adaptability. This inclusion, particularly regarding scandium, should be seen as an example of flexibility rather than a definitive classification. For instance, in 'shakhdaraite-(Y)' denoted as (-Nb-Sc-O-Y-), a format specific to the Mindat

API, we treated '-Sc-Y-' collectively as REEs for simplicity, transforming its chemical composition to "-Nb-O-REE-". Such groupings are adjustable in our workflow, and in section 4, we will discuss the contentious topic of whether to consider Sc (scandium) and Y (yttrium) as REE.

Echoing the methodology of our preliminary analysis, we curated four datasets, leveraging the attributes "elements" and "sigelements" from OpenMindat and implementing appropriate fraction calculations to generate normalized datasets. Visualization of these datasets culminated in Fig. 5, where Fig. 5a presents a 3D heatmap of the "30_elements" dataset, spotlighting an REE layer (with REE designated to the Z axis). Enhanced clarity was achieved through a logarithmic transformation, showcased in Fig. 5b. The resultant heatmap vividly illuminates elements exhibiting heightened co-existence propensities with REE. The dominance of O (observed in 386 mineral species), H (219), Si (180), Ca (158), and Fe (96) emerges palpably, offering invaluable insights into elemental associations, particularly those embedded within the REE framework.

As elucidated in the preliminary use case, the value encapsulated within a specific cell is emblematic of the fraction of mineral species harboring Z elements that concurrently contain X and Y elements in the normalized datasets. For instance, within the specific context of the cell with "X: Si, Y: Ca, Z: REE," a value of 100 is juxtaposed against a value of 393 for the cell corresponding to "X: REE, Y: REE, Z: REE." This convention leads to a normalized value of 100/393 for the former cell, and a unitary value for the latter.

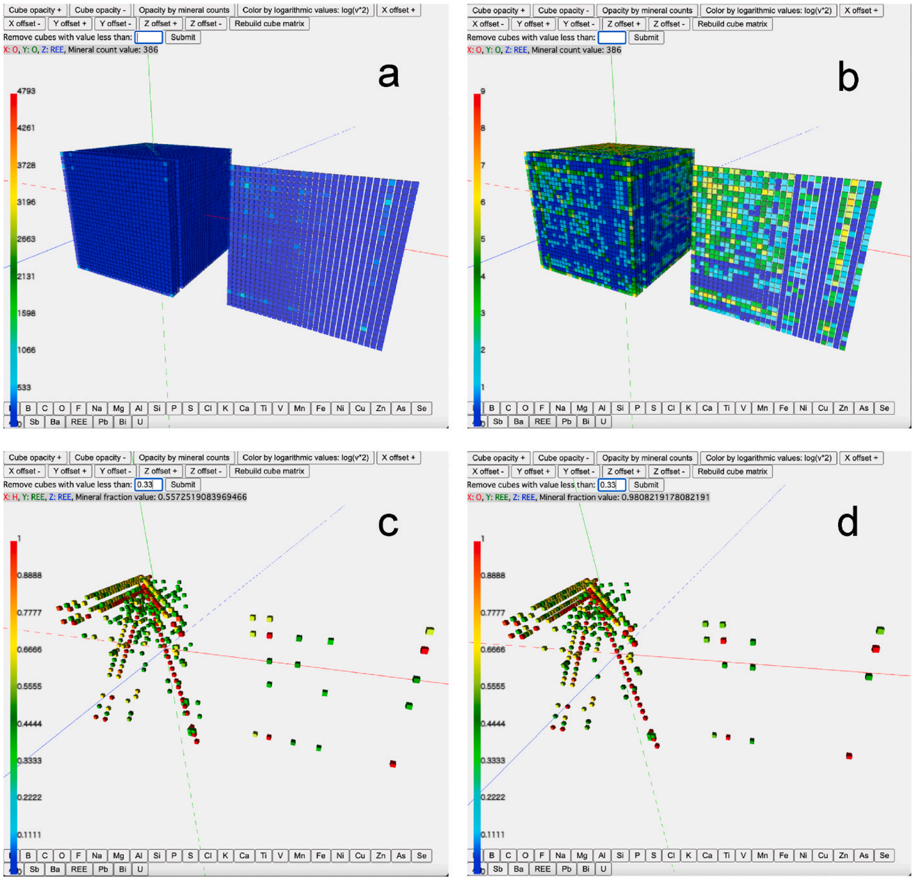
Fig. 5c and d, representing the datasets "normalized_30_elements" and "normalized_30_sigelements," are rendered with the REE layers selectively extracted and a cutoff value of 0.33 applied, thereby removing cells with values falling beneath this threshold. Upon careful examination, several intriguing patterns emerge.

First, the asymmetry on the left side of Fig. 5c and d is conspicuous, deviating from the symmetrical pattern observed in the initial use case. This distinction arises from the inherent asymmetry of the normalized values, as opposed to the original co-existence counts.

Second, the manifestation of high-value cell clusters indicates the prominent elements depicted in Fig. 5b. An illustrative example is the demonstration of parallel cell chains in the upper left quadrant of both Fig. 5c and d, an indicator of the dominance of mineral species integrating H and O. Furthermore, two inclined cell chains exemplify the pair-wise coexistence between H or O and other elements.

Third, the presence of green cells outside the aforementioned chains yields valuable insights into the coexistence patterns of elements independent of the H and O influence. Within Fig. 5c and d, where the REE layer is sliced, distinct variations between the REE slices become apparent. Specifically, in Fig. 5c, the residual cells correspond to Ca, O, and REE, whereas in Fig. 5d, all Ca-associated cells are absent. This disparity underscores a potential tendency within the Mindat database to recognize Ca as significant in the presence of REE within mineral species.

In conclusion, the use cases elucidated herein represent a mere fraction of the possibilities latent within the current 3D heat map demo system. We invite interested scholars and researchers to explore the demo website (refer to the section "Code and Dataset Availability" for the link), engaging with diverse data selections and uncovering hitherto concealed patterns. The dynamic interplay of these elements, as captured by our 3D visualization model, can significantly augment our understanding of mineral species and their complex interrelationships.



(a) Remove cubes with value less than:
X: O, Y: O, Z: REE, Mineral count value: 386

(b) Remove cubes with value less than:
X: O, Y: O, Z: REE, Mineral count value: 386

(c) Remove cubes with value less than: 0.33
X: H, Y: REE, Z: REE, Mineral fraction value: 0.5572519083969466

(d) Remove cubes with value less than: 0.33
X: O, Y: REE, Z: REE, Mineral fraction value: 0.9808219178082191

Fig. 5. Interactions with the 3D heat map of the 30 common elements among IMA-approved mineral species. (a) Sliced out layer with REE on Z axis; (b) Re-coloring cubes based on the logarithmically transformed cell values; (c) The REE slice in normalized_30_elements dataset, with all the cubes having values greater than 0.33; (d) The REE slice in normalized_30_sigelements dataset, with all the cubes having values greater than 0.33.

4. Discussion

In previous sections, we presented our work on designing, developing, and using case analyses of a 3D heat map demo system for EDA of mineral data. The results show that EDA methods such as heat maps effectively expose patterns of various correlations in mineral data. Data visualization techniques such as the 3D matrices used in the demo system add more features of interactions in the result interpretation. Those techniques and scientific-meaning operations in data cleansing and selection make the resulting 3D heat map demo system a helpful tool for data-intensive research in mineralogy.

Although the work has a scientific focus on mineralogy, the designed workflow is an excellent example of addressing the challenges caused by the rapid evolution of open big data in geosciences. The open data and open science activities (Stall et al., 2019; Gentemann, 2023) have resulted in fast-growing records in open data portals. For any data visualization or analysis results, if they cannot elaborate on the updated datasets, their findings will soon be out of date. IMA approves about 100 new mineral species annually for the mineral species data alone. Accordingly, the Mindat data contributor community (about 7000 people) (Ma et al., 2023) is doing its best to upload and curate the records of the new species, integrate them with the existing mineral species records, and then make the data machine accessible through the Mindat API. With the ever-growing OpenMindat database, new scientific questions or ideas can always be explored. The utility of the 3D heat map demo system built in this study has been tested through a list of use cases to illustrate how EDA can facilitate research on mineralogy. Moreover, the workflow built on the live OpenMindat data API (Fig. 1) is effective for tackling the velocity of open mineral data. Beyond mineralogy, the EDA method and the techniques for correlation visualization can also be adapted in other geoscience disciplines, as well as those outside geosciences.

As a reflection, this work's most noteworthy technical contributions are in two parts: the established gateway to the live data source and the scientifically-sound data operations on the user interface. Our previous work (Ma et al., 2017) built the initial 3D heat map visualization environment and constructed a few use cases with separate datasets. Those datasets were collected from multiple sources and were cleansed manually. Although the 3D heat map results were sound, it is hard to extend to other datasets to build new use cases quickly. By elaborating on the existing software building blocks, the new developments in this study established a framework (Fig. 1) to interconnect live mineral data service, meaningful data operations, and interactive heat map results. Those new features significantly improve the reusability and extensibility of the resulting demo system regarding both new datasets and data operation functions. The OpenMindat API, once fully established, will provide accessibility to massive valuable mineral data (Ma et al., 2023). The 3D heat map can be used to do an EDA of various correlations in mineral data. All the established use cases can be regularly refreshed using the live data service from OpenMindat, and the workflow for data retrieving and cleansing developed in this study.

On the other hand, mineralogy has many established and evolving knowledge systems. The igneous mineral classes used in this study for data selection are part of them, and more such structured knowledge can be elaborated in EDA. We aim to continuously maintain and update this 3D heat map demo system and make it a long-lasting tool and reference for research, education, and outreach in mineralogy, mineral informatics, and data science.

We have planned a list of extensions to the demo website. As the technical workflow was established from the live data source to the visualization output, we will work on several new datasets and functions for the immediate next stage to extend the scientific coverage and the utility of the demo system among various users. For example, in this study, we showed the use case of igneous mineral classes. We can also build use cases for metamorphic and sedimentary minerals and many other classes in the Dana Classification (Gaines et al., 1997) and Hey's Mineral Index (Clark, 1993).

The proposed data retrieval and cleansing workflow boasts remarkable adaptability to diverse user requirements. The work is particularly evident when addressing historical ambiguities, such as the classification of scandium (Sc) and yttrium (Y) as REE. While normally, REE referred to the lanthanide series, comprising 15 elements from atomic numbers 57 (lanthanum) to 71 (lutetium), both Sc and Y, due to their association with lanthanides in mineral deposits and similar properties, have often been studied together with the lanthanide series and occasionally been regarded as “rare earth elements” by mineralogists (Balaram, 2019). Such debates underscore the need for flexible data preparation and pre-processing. The proposed workflow ensures that, at the early stages of EDA, scientists can handle these variances and generate customized datasets. This adaptability establishes a strong foundation for varied research aims and requirements.

The output can be new datasets and drop-down lists at the data selection step of the demo website. Using the OpenMindat API, we can obtain mineral species records of certain areas of interest, such as by countries or states. A 3D heat map can be built to illustrate the correlations between minerals and elements in each area. We can also update the user interface of the demo system to display the heat maps of several areas. The inter-comparison of those heat maps has the potential to expose other interesting mineralogy patterns across those areas. For all the current 3D heat map results, elements are listed along the three axes (i.e., the element-wise heat maps). We can also try new data structures for mineral-wise heat maps. For a simple example, we can list subsets of mineral species on the X, Y, and Z axes, and then in each cell, we can fill in the number of localities where the three mineral species co-occur. Also, one could examine coexisting minerals associated with the 57 mineral paragenetic modes (Hazen and Morrison, 2022), or track mineral/chemical co-occurrences versus age.

We can also think about other potential ideas for future work from a broad perspective. One idea is to demonstrate the pattern of mineral forming temperature and pressure conditions in the 3D heat map. For that direction, we need to extend the data structure by considering what can be treated as proxy properties for temperature and pressure, for example, by considering the attributes of mineral formation modes (Hazen et al., 2023b). Many such properties are hidden in the big geoscience literature data. Text mining may help retrieve them to configure an appropriate data structure (Wang et al., 2018). Another idea for future updates is to elaborate a 3D virtual globe to show the paleogeographic distribution of mineral species on the user interface of the 3D heat map, such as by using the GPlates API (Müller et al., 2018). In the study of mineral evolution (Hazen et al., 2008), an open Mineral Evolution Database was built to record mineral species' temporal and spatial properties (Golden et al., 2019). We can calculate the paleo-coordinates of a mineral species by using its age attribute (e.g., its first appearance on Earth) and visualize the records with a paleogeographic map (e.g., plate tectonics) as background. Once fully established, this will be a valuable tool to demonstrate the scientific topics in mineral ecology, evolution, and informatics (Hazen et al., 2008, 2015; Hystad et al., 2019; Prabhu et al., 2023a, 2023b).

5. Conclusions

This study demonstrates how exploratory data analysis can effectively discover initial patterns in open and big mineral data with appropriate technical developments. As the IMA-approved mineral list continuously grows, the crowd-sourced Mindat database rapidly updates, and the OpenMindat API will thus provide a stable, current service of mineral data. The automated workflow for data retrieval and cleansing built in this study can be adapted to collect various types of correlation data from OpenMindat and then conduct exploratory data analysis in the 3D heat map. Existing knowledge frameworks in mineralogy, such as different parts of the mineral classification systems and the recent studies on mineral ecology and mineral evolution, can also be elaborated on the user interface to add scientific meanings to the data operations. We have already planned a list of new use cases and functions for the extension of the demo system. However, we firmly believe there will be much more innovative scientific topics and data exploration ideas from the geoscience community. The 3D heat map demo system, the data retrieval and cleansing workflow, and all the source code, datasets, and documentation developed in this study are open online. We welcome other researchers to try the existing use cases, adapt the code to build their studies, or send requests to us to collaborate on new use cases and functions for exploratory data analysis.

Code and dataset availability

Name of the code/library: 3D Heat Map Data Preprocessing.

Contact: Jiyin Zhang: jiyinz@uidaho.edu or Xiaogang Ma: max@uidaho.edu; +1 208 885 1547.

Hardware requirements: No specific requirement. A general laptop will work well.

Program language: Python.

Software required: No specific requirement. Any programming environment support Python will work.

Program size: About 30 MB (including the sample datasets).

The source codes are available for download at the link: https://github.com/ChuBL/3DHeatmapDataPreprosses.

The demo system of the 3D heat map is accessible at http://tickmap.nkn.uidaho.edu/D3Cube.

The definition for API attributes: https://github.com/smrgeoinfo/How-to-Use-Mindat-API/blob/main/geomaterialfields.csv.

The Mindat API data used in this study is in alignment with the open access policy stated by Mindat.org, which is transitioning to a Creative Commons share-alike license. We have adhered to Mindat's current data use guidelines, which allow for this application. Detailed licensing information is available at Mindat.org's copyright policy page (https://www.mindat.org/copyrights.php).

CRediT authorship contribution statement

Jiyin Zhang: Writing - review & editing, Writing - original draft, Software, Methodology, Conceptualization. **Xiang Que:** Writing - review & editing, Software, Methodology. **Bhuwan Madhikarmi:** Writing - review & editing, Software, Methodology. **Robert M. Hazen:** Writing - review & editing, Validation. **Jolyon Ralph:** Writing - review & editing, Data curation. **Anirudh Prabhu:** Writing - review & editing, Validation. **Shaunna M. Morrison:** Writing - review & editing, Validation. **Xiaogang Ma:** Writing - review & editing, Writing - original draft, Validation, Software, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The source codes are open at the link: https://github.com/ChuBL/3DHeatmapDataPreprosses. The demo system of the 3D heat map is accessible at: http://tickmap.nkn.uidaho.edu/D3Cube.

Acknowledgments

This work was supported by the U.S. National Science Foundation (Grant No. 2126315).

References

Balaram, V., 2019. Rare earth elements: a review of applications, occurrence, exploration, analysis, recycling, and environmental impact. Geosci. Front. 10 (4), 1285–1303.

Bonaccorsi, E., Orlandi, P., 2020. Tancaite-(Ce), ideally FeCe $(MoO_4)_3$•$3H_2O$: description and average crystal structure. Eur. J. Mineral 32 (3), 347–354.

Bowen, N.L., 1928. The Evolution of the Igneous Rocks. Princeton University Press, Princeton, NJ, p. 334.

Carvalho, P.C.S., Antunes, I.M.H.R., Albuquerque, M.T.D., Santos, A.C.S., Cunha, P.P., 2022. Stream sediments as a repository of U, Th and as around abandoned uranium mines in central Portugal: implications for water quality management. Environ. Earth Sci. 81 (6), 175.

Clark, A.M., 1993. Hey's Mineral Index: Mineral Species, Varieties and Synonyms (Third Index). Chapman & Hall, London, p. 852pp.

Dong, H., Huang, L., Zhao, L., Zeng, Q., Liu, X., Sheng, Y., Shi, L., Wu, G., Jiang, H., Li, F., Zhang, L., 2022. A critical review of mineral–microbe interaction and co-evolution: mechanisms and applications. Natl. Sci. Rev. 9 (10), nwac128.

Emami, M., Emami, S.N., 2020. A time-dependent statistical evaluation of the ceramic manufacturing process based on the mineralogical chemical analysis. ArchéoSciences 44 (2), 145–159.

Feltrin, L., Bertelli, M., 2020. Using clustered heat maps in mineral exploration to visualize volcanic-hosted massive sulfide alteration and mineralization. Nat. Resour. Res. 29 (1), 311–344.

Gaines, R.V., Skinner, H.C.W., Foord, E.E., Mason, B., Rosenzweig, A., 1997. Dana's New Mineralogy, eighth ed. John Wiley & Sons, Inc., New York, p. 1819.

Gentemann, C., 2023. Why NASA and federal agencies are declaring this the Year of Open Science. Nature 613 (7943), 217.

Golden, J.J., Downs, R.T., Hazen, R.M., Pires, A.J., Ralph, J., 2019. Mineral evolution database: data-driven age Assignment, how does a mineral Get an age? GSA. https://doi.org/10.1130/abs/2019AM-334056. Annual Meeting 2019, Phoenix, Arizona, USA.

Hazen, R.M., 2019. An evolutionary system of mineralogy: Proposal for a classification of planetary materials based on natural kind clustering. Am. Mineral. 104 (6), 810–816.

Hazen, R.M., Downs, R.T., Eleish, A., Fox, P., Gagne, O., Golden, J.J., Grew, E.S., Hummer, D.R., Hystad, G., Krivovichev, S.V., Li, C., Liu, C., Ma, X., Morrison, S.M., Pan, F., Pires, A.J., Prabhu, A., Ralph, J., Rumyon, S.E., Zhong, H., 2019. Data-driven discovery in mineralogy: recent advances in data resources, analysis, and visualization. Engineering 5 (3), 397–405.

Hazen, R.M., Ferry, J.M., 2010. Mineral evolution: mineralogy in the fourth dimension. Elements 6 (1), 9–12.

Hazen, R.M., Grew, E.S., Downs, R.T., Golden, J., Hystad, G., 2015. Mineral ecology: chance and necessity in the mineral diversity of terrestrial planets. Can. Mineral. 53 (2), 295–324.

Hazen, R.M., Morrison, S.M., 2022. On the paragenetic modes of minerals: a mineral evolution perspective. Am. Mineral. 107 (7), 1262–1287.

Hazen, R.M., Morrison, S.M., Prabhu, A., Walter, M.J., Williams, J.R., 2023a. An evolutionary system of mineralogy, Part VII: the evolution of the igneous minerals (> 2500 Ma). Am. Mineral. https://doi.org/10.2138/am-2022-8539 (in press).

Hazen, R.M., Morrison, S.M., Prabhu, A., Williams, J.R., Wong, M.L., Krivovichev, S.V., Bermanec, M., 2023b. On the attributes of mineral poaragenetic modes. The Canadian Journal of Mineralogy and Petrology 61. https://doi.org/10.3749/2200022 (in press).

Hazen, R.M., Papineau, D., 2012. Mineralogical co-evolution of the geosphere and biosphere. In: Knoll, A.H., Canfield, D.E., Konhauser, K.O. (Eds.), Fundamentals of Geobiology. Wiley-Blackwell, Oxford, UK, pp. 333–350.

Hazen, R.M., Papineau, D., Bleeker, W., Downs, R.T., Ferry, J.M., McCoy, T.J., Sverjensky, D.A., Yang, H., 2008. Mineral evolution. Am. Mineral. 93 (11–12), 1693–1720.

Hystad, G., Downs, R.T., Grew, E.S., Hazen, R.M., 2015. Statistical analysis of mineral diversity and distribution: earth's mineralogy is unique. Earth Planet Sci. Lett. 426, 154–157.

Hystad, G., Morrison, S.M., Hazen, R.M., 2019. Statistical analysis of mineral evolution and mineral ecology: the current state and a vision for the future. Applied Computing and Geosciences 1, 100005.

Ma, X., 2023. Data science for geoscience: recent progress and future Trends from the perspective of a data Life Cycle. In: Ma, X., Mookerjee, M., Hsu, L., Hills, D. (Eds.), Recent Advancement in Geoinformatics and Data Science. Geological Society of America Special Paper V. 558, Boulder, CO, pp. 57–69, 2023.

Ma, X., Hummer, D., Golden, J.J., Fox, P.A., Hazen, R.M., Morrison, S.M., Downs, R.T., Madhikarmi, B.L., Wang, C., Meyer, M.B., 2017. Using visual exploratory data analysis to facilitate collaboration and hypothesis generation in cross-disciplinary research. ISPRS Int. J. Geo-Inf. 6 (11), 368.

Ma, X., Ralph, J., Zhang, J., Que, X., Prabhu, A., Morrison, S.M., Hazen, R.M., Wyborn, L., Lehnert, K., 2023. OpenMindat: open and FAIR mineralogy data from the Mindat database. Geoscience Data Journal.

Morrison, S.M., Liu, C., Eleish, A., Prabhu, A., Li, C., Ralph, J., Downs, R.T., Golden, J.J., Fox, P., Hummer, D.R., Meyer, M.B., 2017. Network analysis of mineralogical systems. Am. Mineral. 102 (8), 1588–1596.

Müller, R.D., Cannon, J., Qin, X., Watson, R.J., Gurnis, M., Williams, S., Pfaffelmoser, T., Seton, M., Russell, S.H., Zahirovic, S., 2018. GPlates: building a virtual Earth through deep time. G-cubed 19 (7), 2243–2261.

Prabhu, A., Morrison, S.M., Fox, P., Ma, X., Wong, M.L., Williams, J., McGuinness, K.N., Krivovichev, S., Lehnert, K.A., Ralph, J.P., Lafuente, B., 2023a. What is mineral informatics? Am. Mineral. 108 (7), 1242–1257.

Prabhu, A., Morrison, S.M., Hazen, R.M., 2023b. Mineral informatics: origins. In: Bindi, L., Cruciani, G. (Eds.), Celebrating the International Year of Mineralogy: Progress and Landmark Discoveries of the Last Decades. Springer Nature, Cham, Switzerland, pp. 39–68.

Ralph, J., Ma, X., Prabhu, A., Martynov, P., 2022. Building OpenMindat for FAIR mineralogical data access. In: EarthCube 2022 Annual Meeting (San Diego, CA. Poster).

Schutt, R., O'Neil, C., 2013. Doing Data Science: Straight Talk from the Frontline. O'Reilly, Sebastopol, CA, p. 375pp.

Stall, S., Yarmey, L., Cutcher-Gershenfeld, J., Hanson, B., Lehnert, K., Nosek, B., Parsons, M., Robinson, E., Wyborn, L., 2019. Make scientific data FAIR. Nature 570 (7759), 27–29.

Tukey, J.W., 1977. Exploratory Data Analysis. Addison-Wesley, Reading, PA, USA, p. 688p.

Wang, C., Ma, X., Chen, J., Chen, J., 2018. Information extraction and knowledge graph construction from geoscience literature. Comput. Geosci. 112, 112–120.

Wilkinson, L., Friendly, M., 2009. The history of the cluster heat map. Am. Statistician 63 (2), 179–184.

Wilkinson, M.D., Dumontier, M., Aalbersberg, IjJ., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3 (1), 160018.