CS 732: Data Visualization Assignment 3 Report

Rachna S Kedigehalli
IMT2019069
IIIT Bangalore
rachna.kedigehalli@iiitb.ac.in

Nandakishore S Menon
IMT2019057
IIIT Bangalore
nandakishore.menon@iiitb.ac.in

Chinthan Chandra
IMT2020109
IIIT Bangalore
chinthan.chandra@iiitb.ac.in

I. Dataset

In this assignment, we work with the Cryptocurrency Historical Prices [1]. The given dataset has values of the following variables on a daily basis from the time when the particular cryptocurrencies were first introduced till July 06, 2021 for 23 cryptocurrencies:

- Date: date of observation.
- Open: Opening price on the given day.
- High: Highest price on the given day.
- Low: Lowest price on the given day.
- Close: Closing price on the given day.
- Volume: Volume of transactions on the given day.
- Market Cap: Market capitalization in USD, which refers to the total dollar market value of a company's outstanding shares of stock [2].

The cryptocurrencies are namely, Chainlink, Cardano, Solana, Dogecoin, Polkadot, NEM, XRP, Ethereum, Aave, Bitcoin, Cosmos, Litecoin, Uniswap, EOS, Binance Coin, Crypto.com Coin, USD Coin, Monero, TRON, Wrapped Bitcoin, Tether, IOTA, and Stellar.

II. Libraries

- *matplotlib.pyplot*: Used to plot the desired visualizations.
- *plotly*: Used to plot sunburst and treemaps.
- *networkx*: Used to plot node-link diagram and matrix visualization of the adjacency matrix.
- *d3.js*: Used to plot force-directed node-link diagram [3].

III. Data Remodelling

A. Remodelling for Network Visualization

We were given the task to remodel this multivariate data for network and tree visualization. Networks give an impression of some relationship between the nodes of the network. Since we have the prices of cryptocurrencies for different dates, we can compute the percentage change in different prices (we have considered close price since that is the most commonly used market indicator). The similarity in the behavior of prices of the two currencies can be seen using the similarity of percentage change in their closing prices.

A constant `max_diff` is set and a link between two nodes is added if the difference of percentage change of price (close price) is less than the `max_diff`. Similarity in the percentage change of price determines the distance between the nodes in the graph. So, edge weight is set as inversely proportional to the difference in the percentage change of prices of the crypto-currencies. The graph is considered to be an un-directed graph. Thus only a single link is added between any two nodes.

B. Remodelling for Hierarchy Visualization

The tree data structure has a relationship corresponding to parent and child in two adjacent levels. The data given to us is time-series data, so we used time as a metric to group data together and form a hierarchy. To get a tree from the dataset, we have used the `Date` column to get the `Year, Month, Date1` columns for the dataset. The hierarchy levels are `Year, Month, Date1`. The weight of each node is the `Volume` traded in that period. The dataset used here is the complete dataset corresponding to that coin. Another hierarchy we made was the share of volume traded on a given day for all coins. This hierarchy has only one level. So the dataset used was a subset of the dataset given, taking one row from each coin based on the date. This would give insights into the market composition on that particular day.

IV. Visualizations

A. Node-link diagram

Data is remodelled for a particular date as described in Section III. *networkx* has been used to generate the node-link diagrams. Nodes are colored based on the trend of change they show on that particular day, i.e. if the change is positive, then the node color is set to green, otherwise red. Figure. 1 shows the nodes in a circular layout. Figure. 2 shows a force-directed node-link diagram. Edge weights are used to set the repulsive forces. Instead of setting two colors for the nodes, we also tried giving a sequential colormap for the nodes as shown in Figure. 3. Green was chosen for the positive end and red for the negative end as this is intuitive for market prices. Since this is a diverging color map, percentage change of 0 is set to the mid of colormap.

In Figure. 3, it can be seen that *DOGE* crypto-curreny has a very high positive change compared to all other coins. *BTC*, *ETC*, etc. are in the same community in the node-link diagrams, which shows that the popular coins have similar behaviour.

To get aesthetically pleasing visualization and interactivity of force-directed node-link diagram, we plot the same data using *d3.js*. The result is shown in Figure 4.

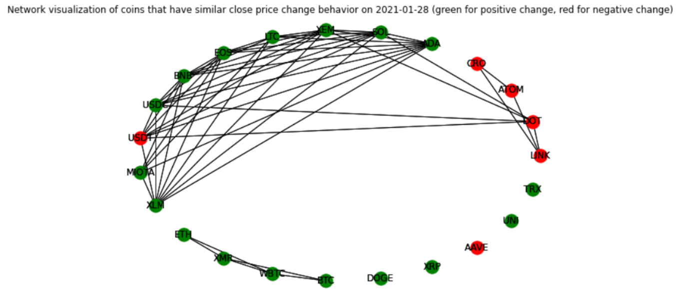

Fig. 1: Node-link diagram on a circular layout

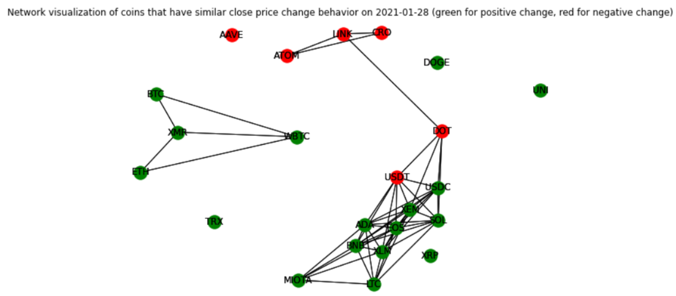

Fig. 2: Force-directed node-link diagram

B. Matrix Visualization

Adjacency matrix is drawn with crypto-currencies as nodes and edge weights as the difference in % change of close prices of the currencies. Figure. 5 shows the adjacency matrix with *magma* colormap on linear scale. Since *DOGE* crypto-currency had a very high % change value, the visualization with linear scale was skewed. So, we used log scale as shown in Figure. 6. Rows and columns corresponding to *DOGE*, *XRP*, *UNI*, *TRX* have lighter shades overall, indicating that the difference of these currencies with respect to other currencies is relatively high. This is also indicated by the visualizations in assignment 4.

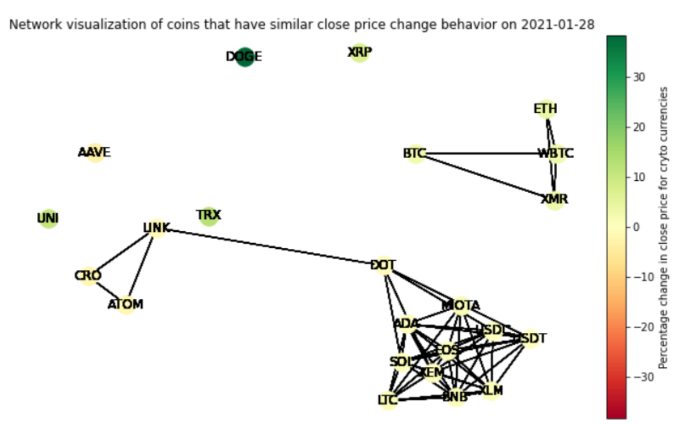

Fig. 3: Force-directed node-link diagram with sequential colormap

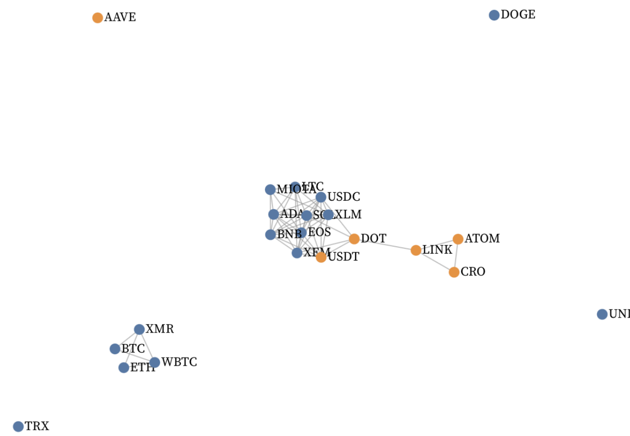

Fig. 4: Interactive Force-directed node-link diagram in D3.js

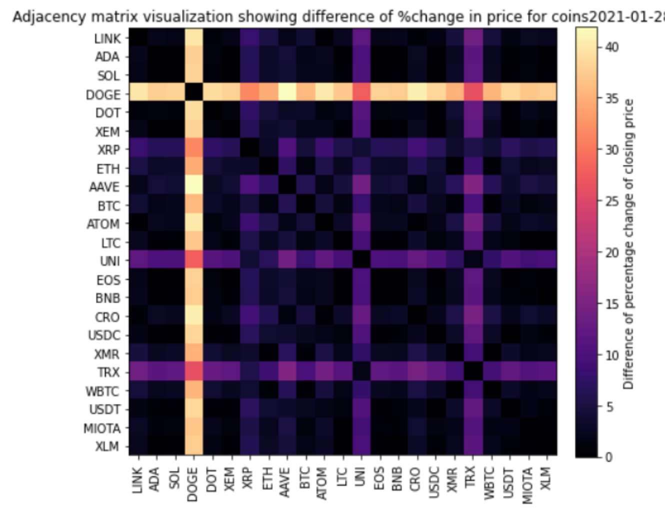

Fig. 5: Adjacency matrix with linear scale

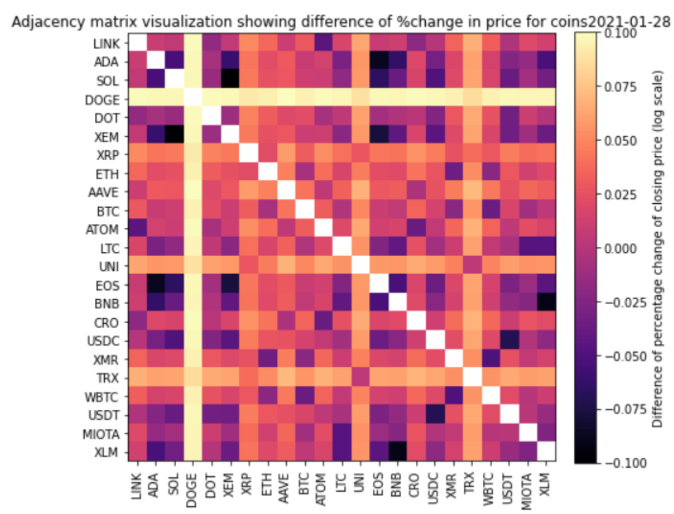

Fig. 6: Adjacency matrix with log scale

C. Treemap Visualization

The remodeling done to get a hierarchy was used for treemap visualization. Both hierarchies were visualized. The first hierarchy was done for the volume of the coin traded with time as hierarchy, the second was the volume traded for each coin on a given day. The colormap used for the first hierarchy is `magma_r` which is the original `magma` colormap reversed, the reason for reversing is to give a darker color to higher volume. Figure 7 shows an example for the first hierarchy. For the second hierarchy, the colormap used is `viridis_r`, the reason being the same as before. Figure 8 shows an example for the second hierarchy.

From figure 7 we can say that for Ethereum the volume traded has increased gradually over the years as its popularity grew. This is the same for any other coin. Few discrepancies that can be inferred are that the volume traded does not increase gradually over the months in a year, this shows the variation in the market sentiment of any coin.

Similarly for the treemap, in Fig 8, for all coin at a given date tells that Bitcoin and Ethereum are the cryptocurrencies which take the majority of the volume traded on a given day. But this is not true all dates, as the market is majorly affected by the world events.

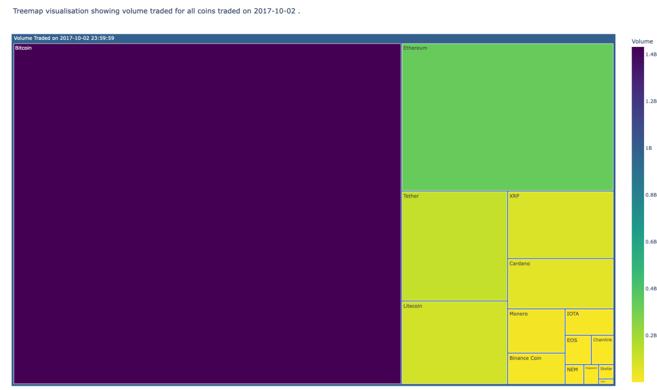

Fig. 8: Treemap visualiation showing the volume traded for all coins on 2017-10-02

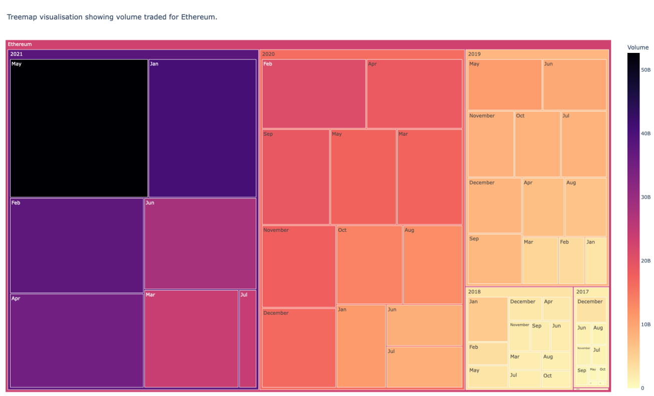

Fig. 7: Treemap visualization showing the volume traded for ETH with hierarchy of Year, Month

D. Sunburst Visualisation

The same hierarchies are visualized in this method too. The colormap used for the first hierarchy is `thermal_r`, with the reason for using the reversed colormap being the same as before and for the second hierarchy `viridis_r` is used. The logarithm of the values could be taken for volume column but this would not show the actual market composition, rather only give a representation for all the coins in the visualization of the second hierarchy. Figure 9 shows the first hierarchy and Fig 10 shows the second hierarchy.

The inference from sunburst visualization in Figure 9, is that the volume traded has gradually increased over the years, with surges in recent years. This is same as the ineferences drawn from the treemap visualization for this hierarchy. Some more observation is that the volume traded in a month is more than the volume traded in a year few years back.

From the figure 10 also says the same inferences as the treemap visualization for the same hierarchy.

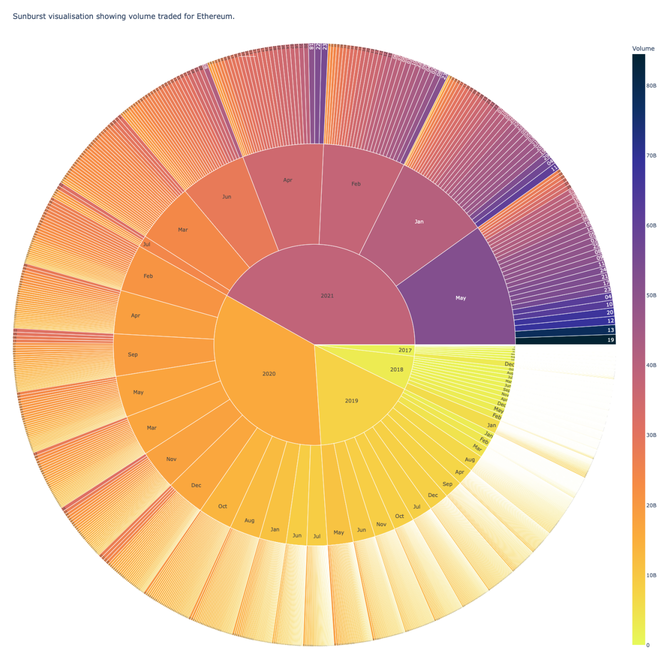

Fig. 9: Sunburst visualization showing the volume traded for ETH with hierarchy of Year, Month and Day

E. Scatterplot Matrix

Scatterplot matrix visualization was done using data pertaining to every cryptocurrency. Each attribute is plotted against every other attribute and is arranged in the form of a matrix. The data points corresponding to every cryptocurrency are added as points on individual graphs and are mapped to a color based on the currency name.

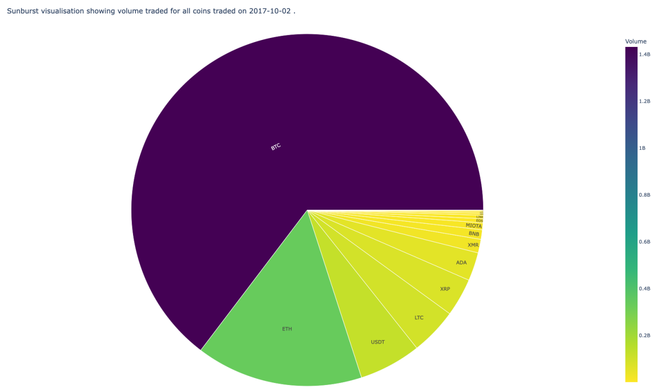

Fig. 10: Sunburst visualization showing the volume traded for all coins on 2017-10-02

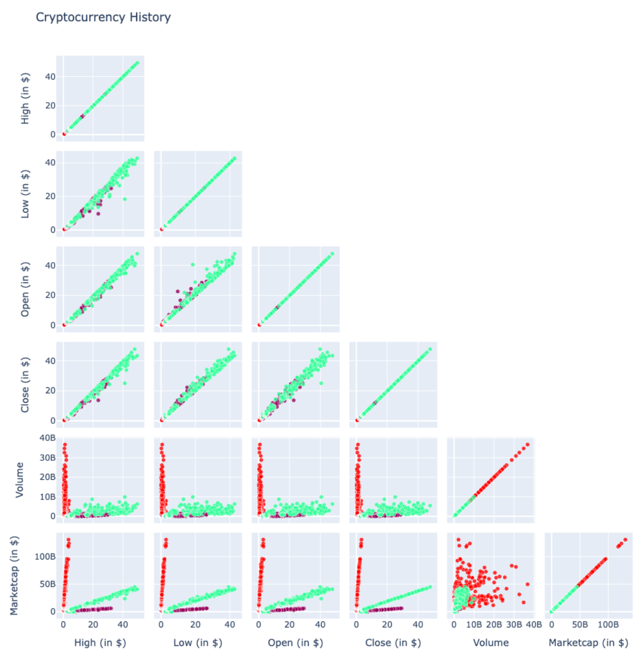

Fig. 11: Scatterplot Matrix visualization for Polkadot, XRP and Cosmos

Scatterplot matrix visualization is particularly useful to make inferences about the relationship between attributes. As an example, Figure 11 shows the Scatterplot Matrix visualization for three coins: Polkadot (in green), XRP (in purple) and Cosmos (in blue). In the Marketcap vs Closing price plot, we can see that XRP has a higher Marketcap, despite having a low Closing price. This means that the number of coins in circulation of XRP must be significantly higher than the other coins represented here.

Additionally, the generated visualization is interactive, and allows the user to highlight data points by means of brushing (as seen in Figure 12). It is also a linked visualization, as brushing on one component in the visualization would change

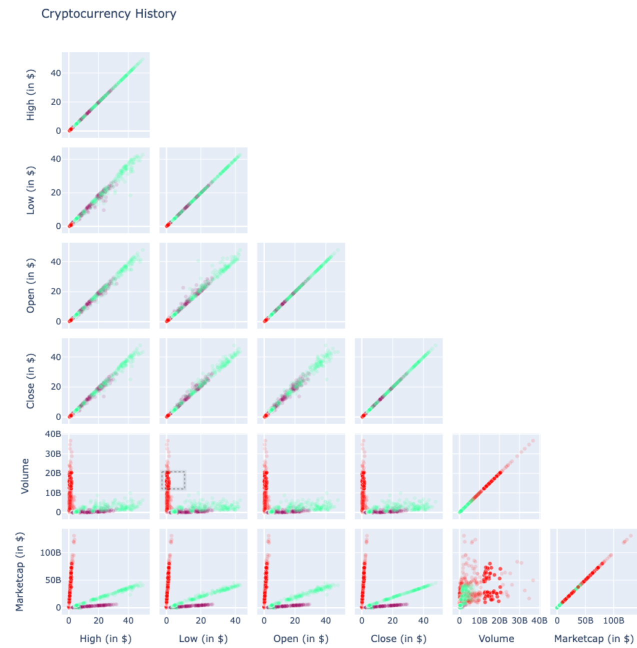

Fig. 12: Scatterplot Matrix visualization for Polkadot, XRP and Cosmos after brushing

the entire visualization.

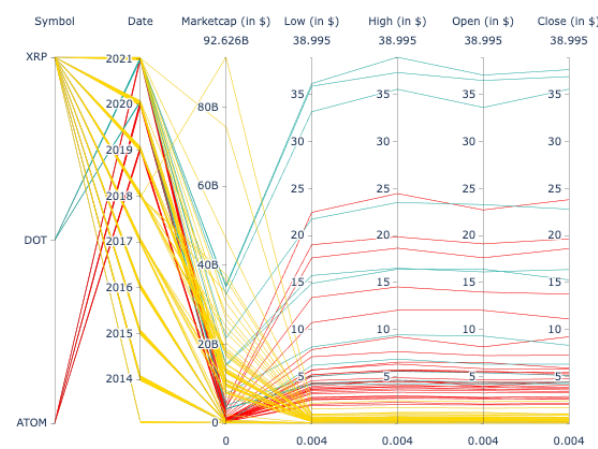

Fig. 13: Parallel Co-ordinate plot created on sampled data for DOT, ATOM and XRP

F. Parallel Co-ordinates Plot

Parallel coordinates are a common way of visualizing and analyzing high-dimensional datasets. We've considered three coins (the same coins considered in the scatterplot matrix) and only the data of the first day from every month is considered. This sampling was done in order to reduce clutter. Axes can

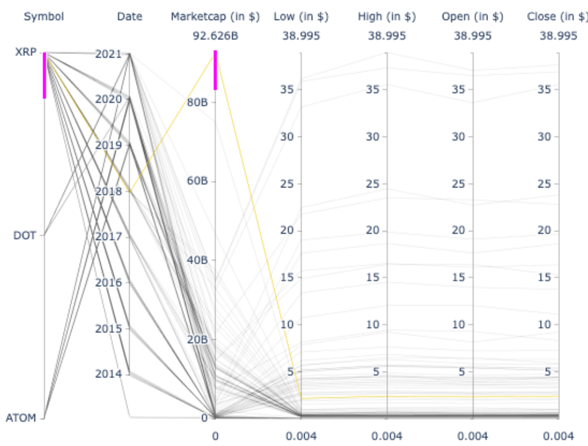

Fig. 14: Parallel Co-ordinates plot with brushing highlighting an outlier for Marketcap attribute

be moved around to make observations using that attribute and its neighbors. This visualization has also been made to be interactive and allows users to highlight data points through brushing. The lines corresponding to a data point will be highlighted by clicking and dragging on the axis. Figure 13 shows the Parallel Co-ordinates plot for ATOM (Cosmos), DOT (Polkadot) and XRP. We can see that XRP is the oldest coin among the lot, but has relatively lower price compared to the other coins.

It can be observed that there is an outlier in the Marketcap attribute for a XRP datapoint. Brushing has been used to highlight it in 14. This spike in Marketcap is due to the spike in price, which can be observed in the candlestick plot from the 4th Assignment.

V. Collective Inferences

The collective inferences we can draw is that Bitcoin and Ethereum are the ones which are traded more and whose prices are higher than the rest of the crypto currencies. Other inferences which we can draw is by spotting anomalies in any of the visualizations and then focusing on that particular date/anomaly attribute in the rest of the visualizations to draw inferences. One such inference we could find was the volume traded for Tether(USDT) was exorbitantly high, higher than the usual cryptocoins in the market, such as Bitcoin and Ethereum. This can be explained by the fact that USDT is generally used for payments and trading, so its traded volume is generally high. Volume traded for DOGE was very low before 28 Jan 2021. But volume traded for DOGE suddenly increased, as seen in Figure. 15. This coincides with Elon Musk tweeting about Dogecoin. Parallel plot in Figure. 16 also shows that high marketcap for DOGE corresponds to 2021. The inferences mentioned in Section. IV-A and IV-B also indicate the same.

Parallel plot for Tether (USDT) crypto-currency shows all the lines cluttered at a stable value of $ 1 for all 4 metrics: high, low, open, close. This shows that USDT is a stable coin.

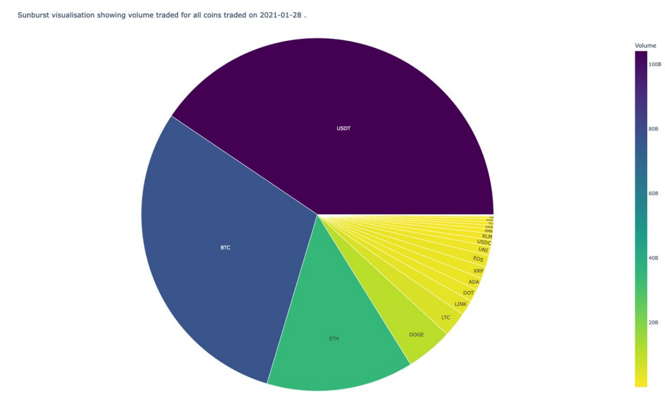

Fig. 15: Volume traded for multiple currencies on 28 Jan 2021

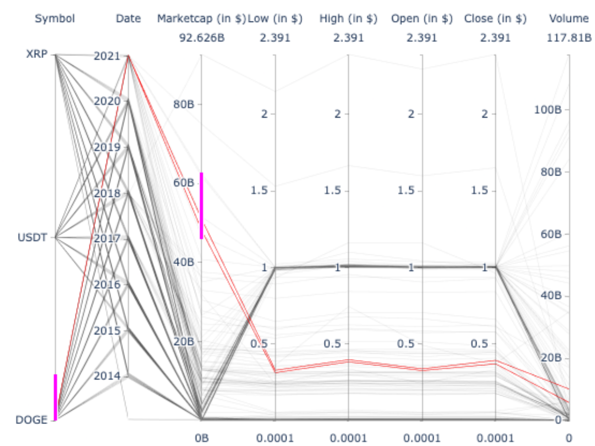

Fig. 16: Brushing on a Parallel Co-ordinates plot highlighting the data from 28 Jan 2021 for DOGE

References

[1] SRK. Cryptocurrency historical prices. [Online]. Available: https://www.kaggle.com/datasets/sudalairajkumar/cryptocurrencypricehistory

[2] J. Fernando. Market capitalization. [Online]. Available: https://www.investopedia.com/terms/m/marketcapitalization.asp

[3] M. Bostock. Force-directed graph d3.js example. [Online]. Available: https://observablehq.com/@d3/force-directed-graph

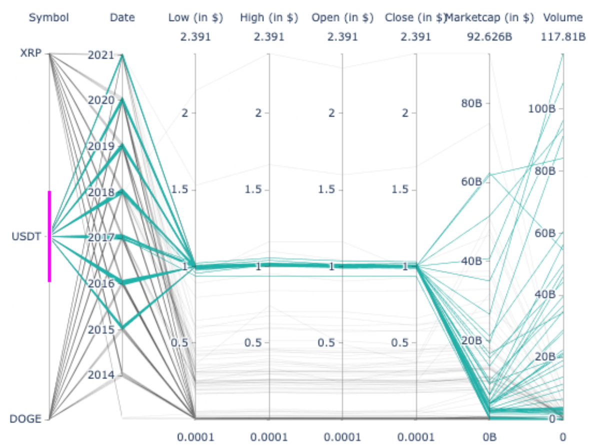

Fig. 17: Parallel co-ordinates plot, highlighting USDT by means of brushing