Milestone #2

Jason Orender

October 4, 2017

1 Data

1.1 Cryptocurrency pricing and volume data

Cryptocurrency selection The most comprehensive pricing and exchange volume data with the largest time frame available is for Bitcoin. While other cryptocurrencies may provide interesting supporting data, the primary focus throughout the project is expected to be on Bitcoin.

Data Source The data used in this project was originally harvested from the "blockchain.info" website and then formatted and uploaded to Kaggle. The blockchain.info website collects real time data regarding cryptocurrency transactions by reading the blockchain for Bitcoin and it also retains complete pricing data back to its inception.

Time Range Bitcoin pricing data is available starting January 3rd, 2009, though the market price is zero at that point since there was as yet no widely agreed upon pricing mechanism or third party exchange. A reasonable starting point for this analysis is **January 1st, 2011**. This is shortly after the first exchange was opened (MtGox) and after "Bitcoin" was established in the public lexicon, achieving a certain level of general awareness. As the project evolves, I expect to continue to harvest real-time data from the blockchain.info website and incorporate it into the final product.

1.2 Data on global financial instability

Proxy data Since there is no specific data regarding global instability per se, a reasonable proxy will be used instead. When cataclysms of a financial nature strike a country, their currency value versus the US Dollar declines in value. This exchange rate is a good indicator of financial instability. Since Bitcoin is also valued globally with respect the US Dollar, this will also allow an apples-to-apples comparison.

Selected event specifics In order to provide a more interesting backdrop to the instability proxy data as represented by the US Dollar exchange rate, areas of interesting cross-correlation may require additional research (why a specific spike in Bitcoin volume occurred on a specific date, for instance).

1.3 Other supporting data

Google Trends Ancillary data, such as that provided by Google Trends may also provide some interesting cross-correlations. This data is available on both a region specific and time-range specific basis, making the search term "bitcoin", for example, a valuable indicator for when cryptocurrency interest peaks.

LocalBitcoins Trading Volume LocalBitcoins is a peer to peer trading platform that is very popular among those without access to a financial institution or a digital currency exchange. These types of transactions are more representative of the types of transactions that a member of a third world nation might participate in. Country specific data is available for a large number of countries.

1.4 Data fields

- cryptocurrency price (quantitative)
- date (ordinal)
- fiat currency exchange rate (quantitative)
- trading volume (quantitative)
- incident explanatory categories (categorical)
- country of trade (categorical)

1.5 Derived Data

- Total trading volume per country per month - *quantitative*
- Lowest exchange value for a specific country in a given month - *quantitative*

2 Abstract Tasks

Analyze. Consume. Present.

Search. Locate. Query. Identify. I expect to use fiat currency exchange rates to locate areas of interest and then to use that location to identify instability and then present any correlations that might exist.

Search. Explore. Query. Summarize Cross-Correlations. Using spikes in exchange rate, cross-correlated with spikes in Bitcoin exchange volume, I expect to be able to show the relationship between a succession of these events and jumps in the worldwide price of bitcoin.

Search. Locate. Query. Compare Local to Worldwide trading volume. The local growth in trading volume should show a departure from worldwide growth in trading volume during instability. This vis should be able to show if, in fact, that is true.