

# Math23C Spring 2018 Final Project

*Patrick Watts, David Wihl*

*May 6, 2018*

## Abstract

Cryptocurrencies such as Bitcoin, Ethereum, etc. generated significant attention in 2017. Cryptocurrencies have significant volatility as there is rampant speculation. Given the high variance in prices, can data science methods explored in this class be used to model the market dynamics?

## Data Source

We obtained a history cryptocurrency price data from <https://www.kaggle.com/sudalairajkumar/cryptocurrencypricehistory/data> for four different cryptocurrencies (BTC, ETH, XMR, XRP)

(REQ: a dataframe, at least two numeric columns, at least 20 rows)

## Data Ingestion, clean up and normalization

```
# Read one price history file per currency
BTCdf = read.csv("data/bitcoin_price.csv", stringsAsFactors = F)
ETHdf = read.csv("data/ethereum_price.csv", stringsAsFactors = F)
XMRdf = read.csv("data/monero_price.csv", stringsAsFactors = F)
XRPdf = read.csv("data/ripple_price.csv", stringsAsFactors = F)

# Make dates native format
BTCdf$Date = as.Date(BTCdf$Date,"%b %d, %Y")
ETHdf$Date = as.Date(ETHdf$Date,"%b %d, %Y")
XMRdf$Date = as.Date(XMRdf$Date,"%b %d, %Y")
XRPdf$Date = as.Date(XRPdf$Date,"%b %d, %Y")
# Ensure that all data start from the same date
earliestCommonDate = max(min(BTCdf$Date),
                           min(ETHdf$Date),
                           min(XMRdf$Date),
                           min(XRPdf$Date))
nrow(BTCdf);BTCdf = BTCdf[BTCdf$Date>=earliestCommonDate,];nrow(BTCdf)

## [1] 1760
## [1] 929
nrow(ETHdf);ETHdf = ETHdf[ETHdf$Date>=earliestCommonDate,];nrow(ETHdf)

## [1] 929
## [1] 929
nrow(XMRdf);XMRdf = XMRdf[XMRdf$Date>=earliestCommonDate,];nrow(XMRdf)

## [1] 1371
## [1] 929
```

```
nrow(XRPdf);XRPdf = XRPdf[XRPdf$Date>=earliestCommonDate,];nrow(XRPdf)
```

```
## [1] 1662
```

```
## [1] 929
```

```
earliestCommonDate
```

```
## [1] "2015-08-07"
```

## Exploratory Data Analysis

### Topic 1 - Volatility Exploration

### Topic 2 -

## Summary of Project Requirements

### Required dataset standards

- ☒ A dataframe
- ☐ At least two categorical or logical columns
- ☒ At least two numeric columns
- ☒ At least 20 rows, preferably more, but real-world data may be limited

### Required graphical displays (all graphs must be colored and nicely labeled)

- ☐ A barplot
- ☐ A histogram
- ☐ A probability density graph overlaid on a histogram
- ☐ A contingency table

### Required Analysis

- ☐ A permutation test
- ☐ A p-value or other statistic based on a distribution function
- ☐ Analysis of a contingency table
- ☐ Comparison of analysis by classical methods (chi-square, CLT) and simulation methods

### Required submission uploads

- ☐ A .csv with the dataset
- ☐ A long, well-commented script that loads the dataset, explores it, and does all the analysis.
- ☐ A shorter .Rmd with compiled .pdf or .html file that presents highlights in ten minutes.
- ☐ A one-page handout that explains the dataset and summarizes the analysis.

### Additional points for creativity or complexity (up to 10 points)

## References