|  |
| --- |
| Process Book |

### Claire Danaher

### Yibo Wu

Cryptic Tweets

## CS 582- Data Visualization

# Table of Contents

IN THIS SECTION:

* List your contents here
* List your contents here
* List your contents here
* List your contents here

Process Notebook

Tab 2

# Overview

As part of a project completed in the Fall of 2017, data was collected over the course of a little over a week from Twitter on the topic of cryptocurrencies. The primary purpose of this previously completed project was use data collection techniques to access and store twitter data. Some preliminary analysis was completed. This data visualization project is a continuation of the previously completed work.

The purpose of this project was to tell a story about how data collected from twitter can be used to explore a topic of interest. We defined this goal more specifically as exploring the relationship between the sentiment of tweets and cryptocurrency valuation. We hypothesized that changes in Twitter sentiment could be used as an indicator for monitoring the cryptocurrency markets.

Unlike other financial markets, cryptocurrencies are highly unregulated. The lack of regulation makes information-gathering difficult when compared to conventional markets. Twitter, similar to the cryptocurrency environment, is also highly unregulated and attracts a segment of the population actively engaged with technology. Therefore, it seemed like an interesting information source to explore for the purposes of collection information about cryptocurrencies.

When we started this project, our idea was to build a console that could be used as prototype for live stream data. Over the course of the development of the project, it became clear that from a visualization standpoint the data was more interested if not explored purely for market evaluation.

* Overview and Motivation: Provide an overview of the project goals and the motivation for it. Consider that this will be read by people who did not see your project proposal.
* Related Work: Anything that inspired you, such as a paper, a web site, visualizations we discussed in class, etc.
* Questions: What questions are you trying to answer? How did these questions evolve over the course of the project? What new questions did you consider in the course of your analysis?

# Data Collection and Analysis

This project was completed using previously collected twitter data from the fall of 2017. The data was collected via the twitter streaming API. Using this API, users can query the twitter stream for tweets containing keywords. A summary of the data collection times is provided below:

|  |  |  |  |
| --- | --- | --- | --- |
| Date | Number of Tweets Collected | | |
| Morning | Midday | Evening |
| 9/13/2017 | - | 1,968 | 1,654 |
| 9/14/2017 | 3,162 | 2,040 | 1,688 |
| 9/15/2017 | 1,874 | 1,917 | 1,560 |
| 9/16/2017 | 1,615 | - | 848 |
| 9/17/2017 | 1,727 | - | - |

The data was imported into Python to be aggregated and analyzed. A Python library was used to analyze the sentiments of the tweets. The library used was Textblob. Textblob is a natural language processing program which used the content of the tweets to calculate a polarity score. The polarity score ranged form very negative(value=-1) to very positive(value=+1).

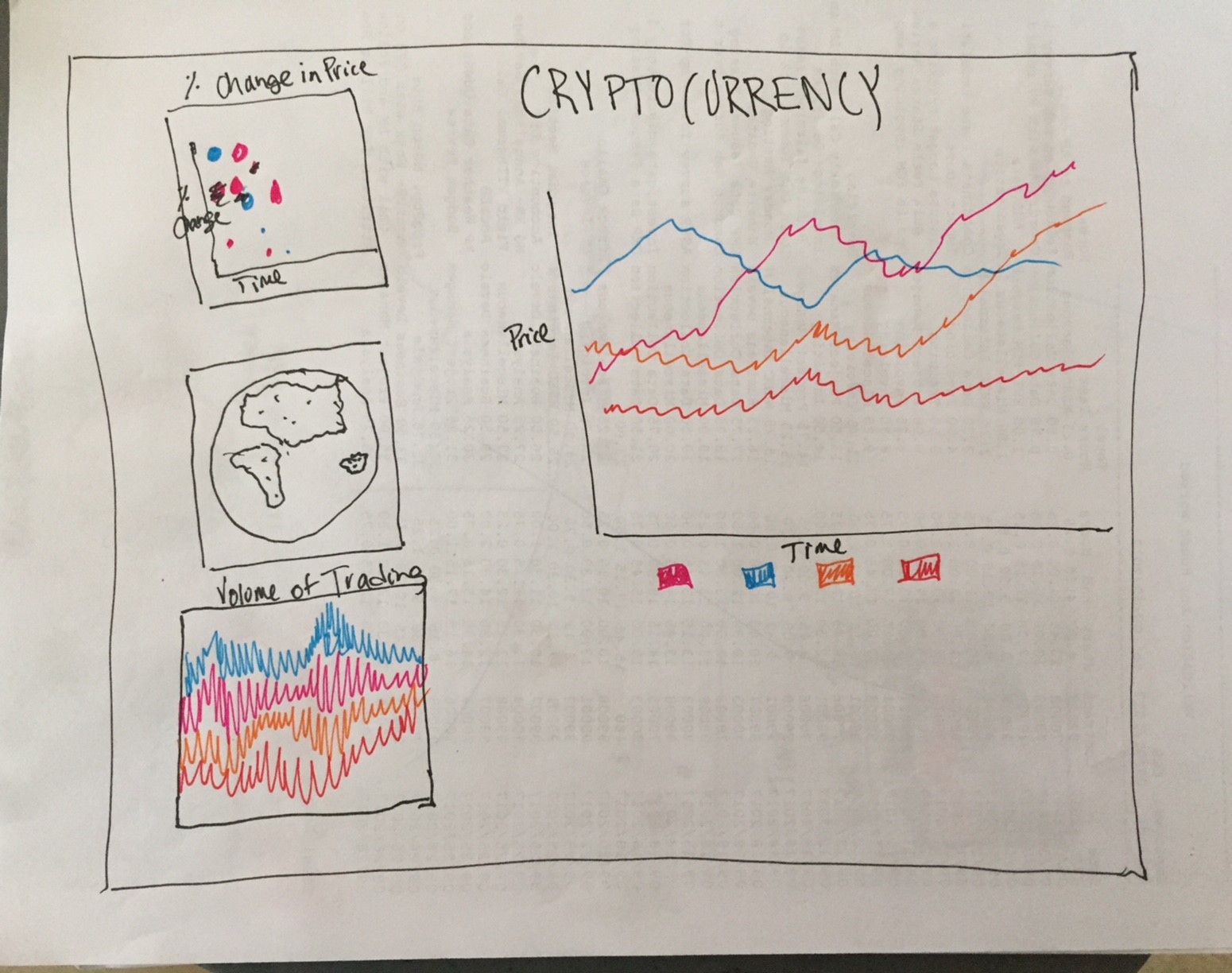
While preliminary analyses were completed as part of past work, additional analysis and exploration was conducted for the purposes of this project. The first step in reviewing the data involved exploring the twitter json objects to identify variables of interest. Initial fields of interest included: datetime information, user location information, tweet content.

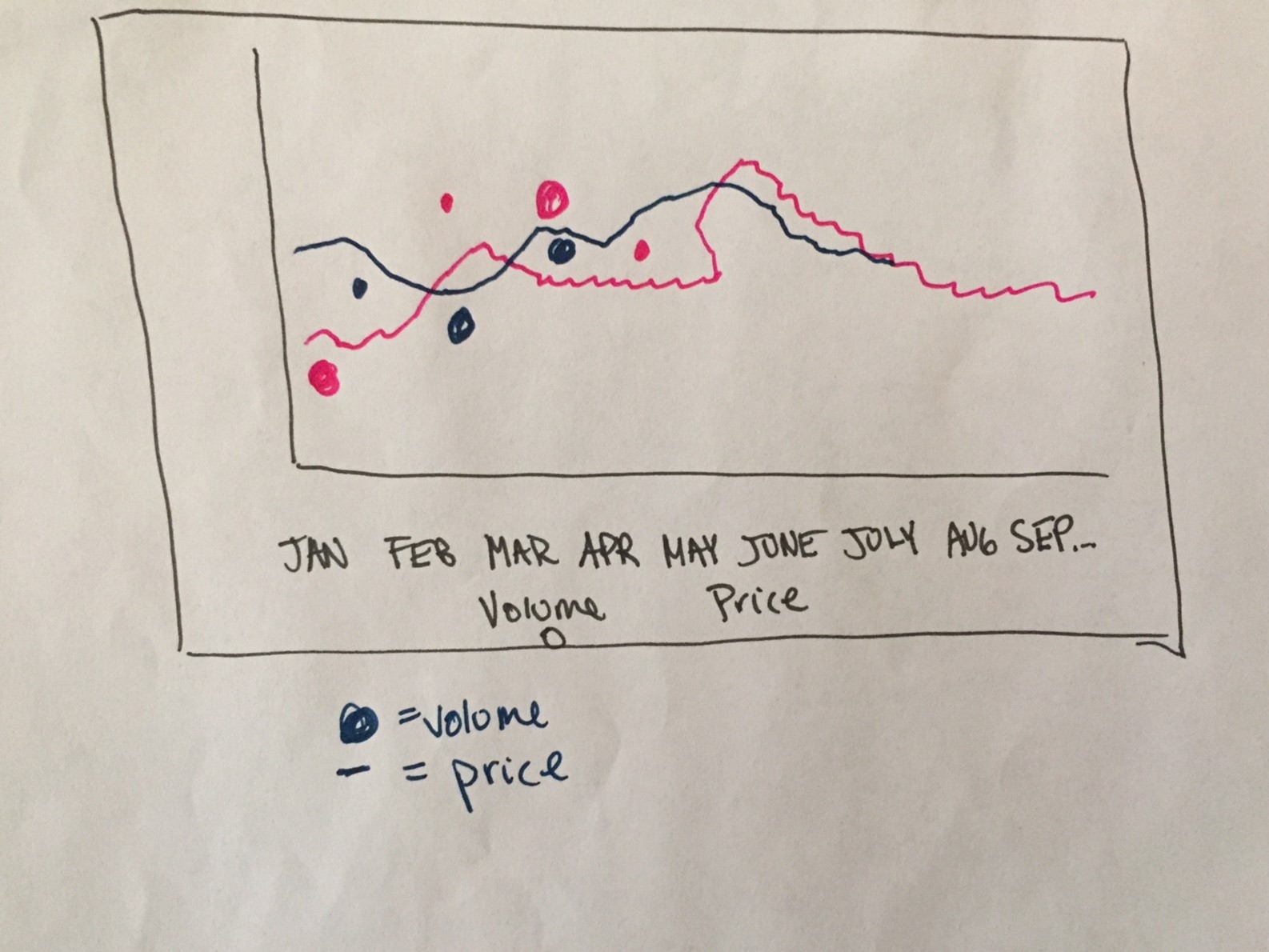
For each of the cryptocurrencies analyzed as part of our analysis, we need the associated financial data. Our research showed that public raw data sets used a daily interval. This interval was not sufficiently granular for our analysis and therefore the data had to be collected manually. The data was collected by manually identifying and collecting the data for each the data collection periods.

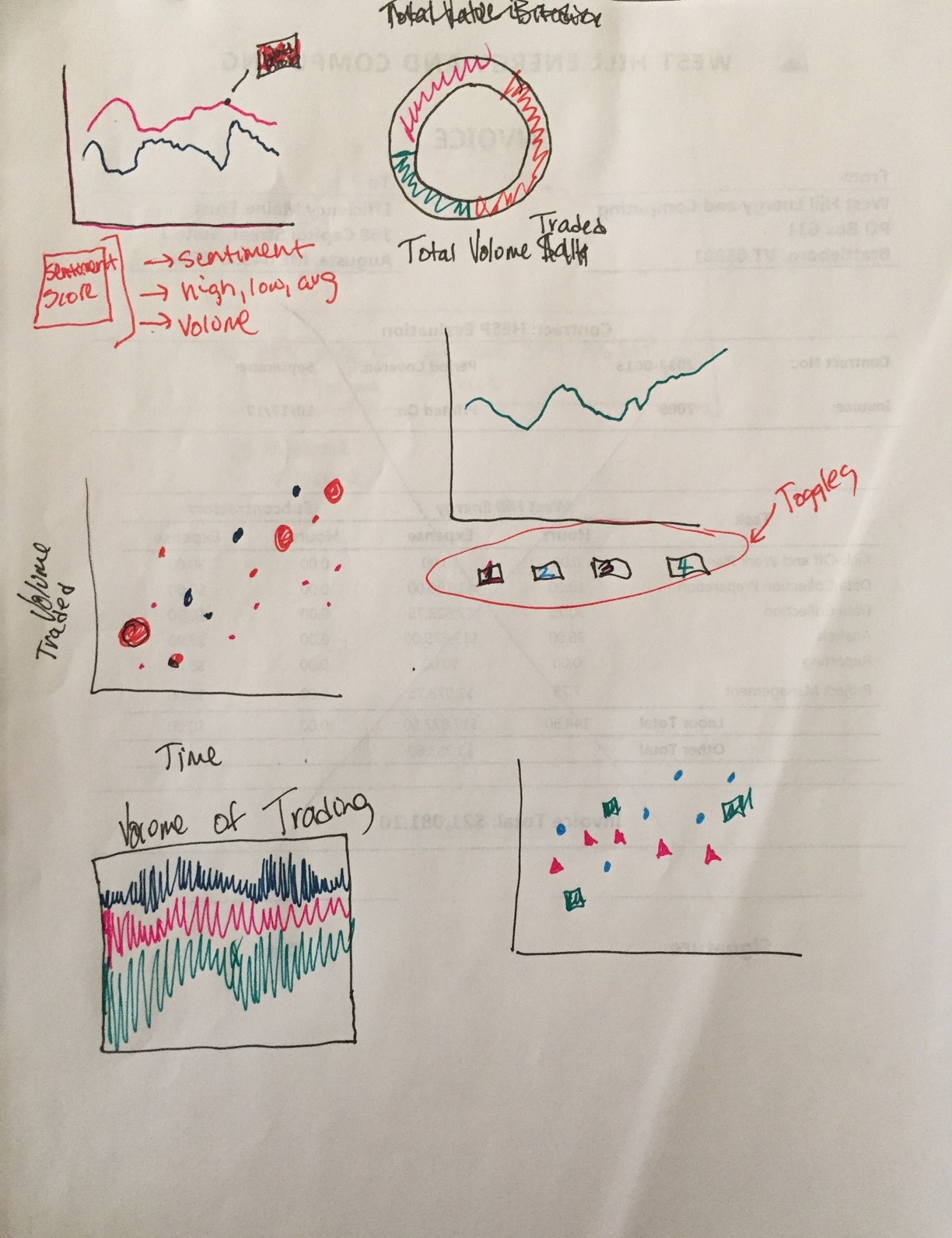
A first glance review of the financial data showed that USD would not be a viable scale of comparison as the range of unitary values is very wide. Therefore, consideration needed to be given to determine an appropriate scale. This and other design issues are discussed in greater detail in the following section.

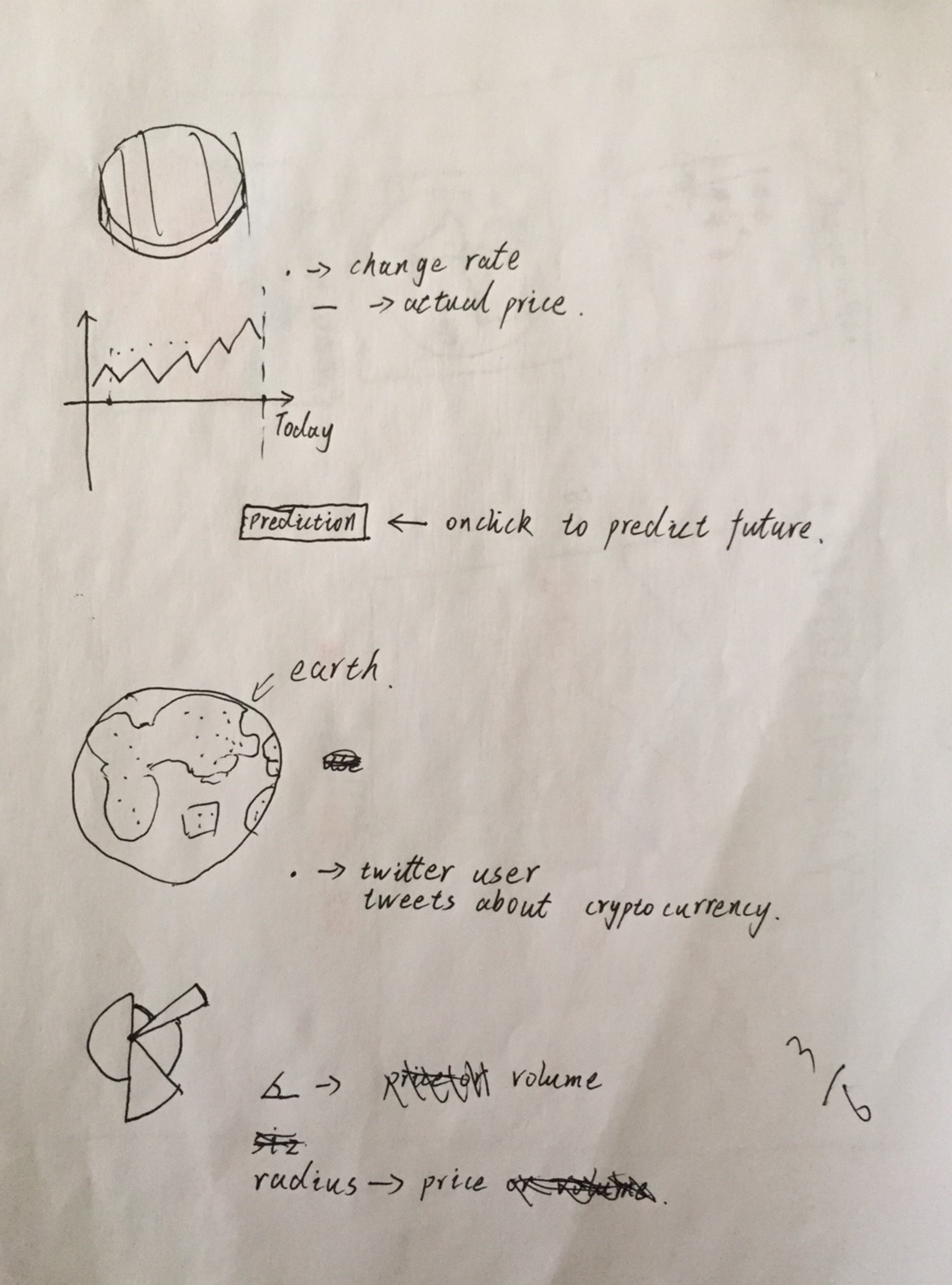
# DeSign Evolution

At the time the project proposal was completed for this project, our idea was to develop a prototype for an interface that could be used to stream data from other sources for the purposes of monitoring the cryptocurrency markets. Our group used a 5 sheet methodology to brainstorm visualization techniques that could be used to express interesting information contained in our data. The results of this brainstorming are provided below.

**Design Sheet 1**

**Design Sheet 2**

**Design Sheet 2**

**Design Sheet 5**

Design Evolution: What are the different visualizations you considered? Justify the design decisions you made using the perceptual and design principles you learned in the course. Did you deviate from your proposal?

# IMPLEMENTATION

* Implementation: Describe the intent and functionality of the interactive visualizations you implemented. Provide clear and well-referenced images showing the key design and interaction elements.

# Evaluation

* Evaluation: What did you learn about the data by using your visualizations? How did you answer your questions? How well does your visualization work, and how could you further improve it?