IT497 Assignment 2

Kaveh Arvand
School of Information Technology
Illinois State University
karvand@ilstu.edu

October 22, 2014

1 Introduction

Bitcoin is the first decentralized virtual currency that can be transferred from person to person online without the existence of a central repository and control of an authority or institute. Bitcoins are created digitally using mathematical formula by a community of people called miners who use computing power in a distributed network. Everybody can join this community. There are free open source software programs that follow the mathematical formula . Bit coins can be used to obtain fiat money, products and services. There are exchanges that allow people to buy and sell coins with each other. A general ledger, called the block chain, stores the details of every single transaction that ever happened in the network. Furthermore, once the bicoins are sent to another person, there is no way to get them back, unless the recipient returns them. The value of the currency fluctuates a lot.However, since only about 20 million bitcoins can ever be mined, it is predicted that the value of bitcoin will rise as more and more people start using them. Many believe that bitcoin is about to disrupt the monetary system in the similar fashion to how email changed the way we communicate.

This cryptocurrency was created in 2009 by a person under the alias of *"Satoshi Nakamoto"*.However, No one knows about his exact identity. This shows that bitcoin is all about being anonymous. The real names of buyers and sellers are never revealed. Due to its anonymous nature, bitcoin can be used for illegal activities as well, such as selling drugs and paying hitmen. Based on an article published in Wall Street Journal on 5 November 2013, *Bi tcoin Comes Under Senate Scrutiny*, the FBI shut down the *Silk Road* online black market and seized 144,000 bitcoins worth US$28.5 million at the time. Wikipedia states that *"the United States is considered more bitcoin-friendly than other governments"* while buying bitcoin in yuan is restricted in China. In this assignment, I will use R statistical data analysis tool to obtain , scrub, explore, model and interpret the data about the volume of bitcoin exchange as well as

its rate within the last 30 days in Bitstamp exchange market, using Quandl website(www.quandl.com) as the source of the data.

2 Recent Bitcoin Exchange Rate on Bitstamp

2.1 Obtaining the Data

Bitstamp is the world's second bitcoin exchange market by volume. It is located in UK. In order to gather the data about bitcoin exchange rate in USD on Bitstamp during the last 30 days(from 21 Sep 2014 to 20 Oct 2014) from Quandl webpage *https:www.quandl.com/BITCOIN/BITSTAMPUS D* using R, the Quandl package should be installed and loaded. Furthermore re, an authentication code is required to have an unlimited access to the data in Quandl website. It can be obtained after signing in to Quandl website from the user's profile page. In order to download the data related to a specific chart, it is needed to have its Quandl code as well . The Quandl code of the chart in our case is "BITCOIN/BITSTAMPUSD":

```
#load Quandl package to extract data
library(Quandl)

#Use the authentication code to obtain access to the data
Quandl.auth("6p2sEsU5szSygtTrZ2t-")

#Download the data using the chart's Quandl code and the required range
BitCoinData <- Quandl("BITCOIN/BITSTAMPUSD"
                      ,trim_start="2014-09-21",trim_end="2014-10-20")

#Show the first rows of the data frame
head(BitCoinData)
```

```
##         Date  Open  High   Low Close Volume (BTC) Volume (Currency)
## 1 2014-10-20 389.1 390.6 376.2 381.2        12082           4614454
## 2 2014-10-19 390.6 394.2 385.0 387.5         3242           1262683
## 3 2014-10-18 383.2 397.3 377.0 390.6         7075           2746148
## 4 2014-10-17 384.2 386.0 371.0 383.6        10508           3989422
## 5 2014-10-16 394.5 399.0 370.1 383.9        22777           8706162
## 6 2014-10-15 402.0 404.3 385.9 394.5        19148           7537827
##   Weighted Price
## 1          381.9
## 2          389.4
## 3          388.2
## 4          379.7
## 5          382.2
## 6          393.7
```

```
#Show the size of the data frame
dim(BitCoinData)

## [1] 30  8
```

2.2 Data Scrubbing

As it is observed, the data frame includes some unwanted data that should be eliminated. Only the "Date", "Volume BTC", "Volume Currency" and "Weighted Price" variable records of the last 30 days are required. Therefore, a new data frame will be created having only the required variables:

```
#Select the required columns and store in BitCoinData.new
BitCoinData.new<-BitCoinData[,c(1,6:8)]

#Show the column names of BitCoinData.new
names(BitCoinData.new)

## [1] "Date"              "Volume (BTC)"      "Volume (Currency)"
## [4] "Weighted Price"
```

Now BitCoinData data frame can be stored into a .csv file:

```
write.csv(BitCoinData.new,"BitCoinData.csv")
```

2.3 Exploring the Data

The following code shows the type of BitCoinData.new :

```
class(BitCoinData.new)

## [1] "data.frame"
```

As expected it is a data frame. The structure of the BitCoinData.new that includes the name, data type and number of variables(columns) as well as the number of observations(rows) is shown below:

```
str(BitCoinData.new)

## 'data.frame': 30 obs. of  4 variables:
##  $ Date             : Date, format: "2014-10-20" "2014-10-19" ...
##  $ Volume (BTC)     : num  12082 3242 7075 10508 22777 ...
##  $ Volume (Currency): num  4614454 1262683 2746148 3989422 8706162 ...
##  $ Weighted Price   : num  382 389 388 380 382 ...
```

In addition, information about the basic descriptive statistics(Minimum, Median, Mean, Maximum, etc.) of the variables in BitCoinData.new data frame can be obtained by executing the summary command:

```
summary(BitCoinData.new)
```

```
##       Date              Volume (BTC)   Volume (Currency)  Weighted Price
##  Min.   :2014-09-21   Min.   : 3242   Min.   : 1262683   Min.   :306
##  1st Qu.:2014-09-28   1st Qu.:11123   1st Qu.: 4410998   1st Qu.:365
##  Median :2014-10-05   Median :15578   Median : 6189163   Median :383
##  Mean   :2014-10-05   Mean   :20700   Mean   : 7550623   Mean   :378
##  3rd Qu.:2014-10-12   3rd Qu.:24357   3rd Qu.: 9541542   3rd Qu.:400
##  Max.   :2014-10-20   Max.   :69538   Max.   :21937271   Max.   :429
```

As it is observed, BitCoinData.new data frame has 4 columns:

- Date: shows the date of each observation. It has date as its data type with minimum value of 2014-09-21 and 2014-10-20 as the maximum value.
- **Volume (BTC):** includes the volume of bitcoins transacted per day. Its data type is number and no negative value is allowed for this variable.
- **Volume (Currency):** includes the volume of bitcoin transactions in USD per day. The Data type for this variable is number. It cannot have a negative value.
- **Weighted Price:** is the is the daily unit price of bitcoin in USD. Number is its data type and the value of this variable cannot be negative as well.

2.4 Results

Table 1. shows the Bitstamp daily exchange rates(number of bitcoins transacted, the value of transactions in USD and bitcoin's unit price in USD) during the last 30 days.

```
BitCoinData.new
```

```
##          Date Volume (BTC) Volume (Currency) Weighted Price
## 1  2014-10-20        12082           4614454          381.9
## 2  2014-10-19         3242           1262683          389.4
## 3  2014-10-18         7075           2746148          388.2
## 4  2014-10-17        10508           3989422          379.7
## 5  2014-10-16        22777           8706162          382.2
## 6  2014-10-15        19148           7537827          393.7
## 7  2014-10-14        24822          10060958          405.3
## 8  2014-10-13        26084          10034403          384.7
## 9  2014-10-12        14250           5236991          367.5
```

```
## 10 2014-10-11          9973           3594482          360.4
## 11 2014-10-10         20543           7440722          362.2
## 12 2014-10-09         47817          17652426          369.2
## 13 2014-10-08         28615           9820002          343.2
## 14 2014-10-07         22962           7504310          326.8
## 15 2014-10-06         69538          21937271          315.5
## 16 2014-10-05         61726          18876496          305.8
## 17 2014-10-04         28982           9853710          340.0
## 18 2014-10-03         20409           7433824          364.2
## 19 2014-10-02         10414           3931403          377.5
## 20 2014-10-01         13038           5010052          384.3
## 21 2014-09-30         14290           5491849          384.3
## 22 2014-09-29         20027           7549530          377.0
## 23 2014-09-28         16140           6202488          384.3
## 24 2014-09-27          5473           2200884          402.2
## 25 2014-09-26          9921           4017441          404.9
## 26 2014-09-25         15017           6175837          411.3
## 27 2014-09-24         14141           6065026          428.9
## 28 2014-09-23         26999          11528952          427.0
## 29 2014-09-22         10803           4343179          402.0
## 30 2014-09-21         14183           5699759          401.9
```

Table 1: **Bitcoin Exchange Rate on Bitstamp from 21 Sep 2014 to 20 Oct 2014**

As it is observed in Table 1., the value of the Volume (Currency) variable is the product of values of the Volume (BTC) and Weighted Price variables.
The table shows that on 6 Oct the number of transacted bitcoins reached its maximum number(69,537) which was worth about $22,000,000 while on 19 Oct the minimum number of bitcoins were transacted(3,242) with the total price of $1,262 ,683. Also 24 Sep and 5 Oct are the dates when the unit price of bitcoin in USD had its highest($428) and lowest($305) values respectively.

In order to create line graphs to illustrate the changes in the number and value of transacted bitcoins as well as the total volume of the transactions during this 30 -day period (Figure 1.), the following steps are required:

```
#load reshape2 package
library(reshape2)

#Melt BitCoinData from wide to long format, without melting Date variable
BitCoinData.m <- melt(BitCoinData.new,id.vars="Date")
```

The long format data frame has three columns; Date, variable and value. variable column consists of 3 levels Volume (BTC), Volume (Currency) and

Weighted Price. Value column includes the correspondent values of the variable column:

```
#Show the first lines of BitCoinData.m
head(BitCoinData.m)

##         Date     variable value
## 1 2014-10-20 Volume (BTC) 12082
## 2 2014-10-19 Volume (BTC)  3242
## 3 2014-10-18 Volume (BTC)  7075
## 4 2014-10-17 Volume (BTC) 10508
## 5 2014-10-16 Volume (BTC) 22777
## 6 2014-10-15 Volume (BTC) 19148
```

Three line graphs are generated in Figure 1. Each shows the value of one of the levels in the variable column per day. Therefore, x axis shows the date while y axis shows the value of Volume (BTC), Volume (Currency) and Weighted Price. In order to scale x and y axes, *scales* package is used. X axis tick marks are set on a daily basis using *scale_x_date*. Since the values of variables on y axis cover a wide range of numbers, they should be transformed to base 10 logarithmic scale. Therefore, the data is transformed before properties such as breaks(tick mark locations) and range of the axis are decided. This provides virtually equal spacing between the tick marks of the y axis that is desired in this case. In addition, in order to to set the y axis tick marks to show exponents, trans_breaks and trans_format are used.

```
#load ggplot2 and scales packages
library(ggplot2)
library(scales)

#Define the data frame that is used to generate the plot
ggplot(data=BitCoinData.m,

#Define the variables that are shown by x and y axes
       aes(x=Date, y=value ,

#Define how each line graph is categorized
           group=variable))+

#Define a separate color for each line based on its category
       geom_line(aes(color = variable))+

#Define labels for x and y axes
       xlab("Date") + ylab("Volume")+

#Set the x axis tick mark breaks on a daily basis
```

```
       scale_x_date(breaks = date_breaks("day"))+

#Transform the values of y axis to base 10 logarithmic scale and
#setezset the y axis tick marks to show exponents
       scale_y_log10(breaks = trans_breaks("log10", function(x) 10^x),
       labels = trans_format("log10",math_format(10^.x)))+

#Set the theme of the graph and change the orientation of labels of
#tickmarks of x-axis
       theme_bw()+theme(axis.text.x= element_text(angle=90, vjust=0.5),
       legend.title=element_blank())
```

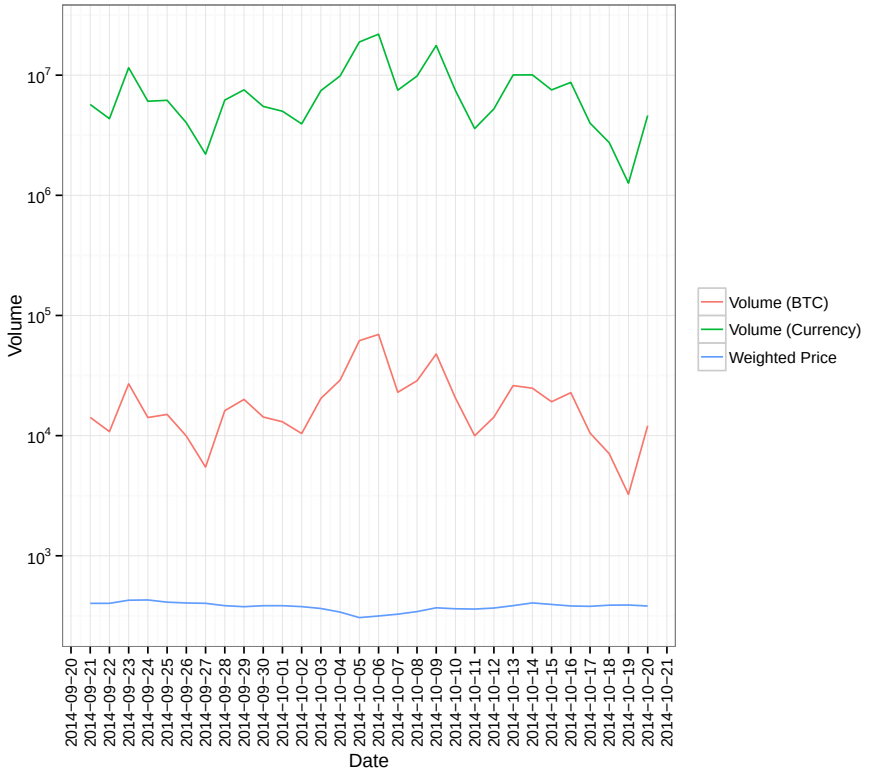

Figure 1: **Bitcoin Exchange Rate (BTC vs. USD) on Bitstamp**

Figure 1. shows that the volume of transacted bitcoins drops for more than 61% from 14,182 to 5,472 on the first 7 days(21 to 27 Sep), after an increase on 23rd. Then it starts rising drastically till it reaches its maximum value (69,537) on 6 October.Afterwards, we can observe a gradual decrease in the number of transacted bicoins till 20 Oct when the recorded volume is 9,385. The total price

of transacted bitcoins in USD follows the same trend as well. By comparing the fluctuations in the number of transacted bitcoins with the changes in bicoin's unit price, we can figure out that there is an inverse relationship between them. In other words, the lower the price is, the more transactions have taken place. This explains the high rate of transactions on 6 October(about 70,000) when the price had a very low value ($305).