Explorative DataViz Short Paper

Ji, Ye

10/15/2017

Assignment Requirement

This assignment builds off the last assignment, resulting in a single document either in PDF or HTML form exported from an R markdown document. You must use the data exclusively from your portfolio proposal! Further, this document should have:

- Five data visualizations of most distinct graph types, made with ggplot2; This can include the three visualizations from last week, though you can replace them if you want. Note that graph requirements always include:

    - Correct usage of all visual encodings;

    - Appropriate data sourcing;

    - Proper labeling of ALL visual encodings, as well as an appropriate title and subtitle;

    - The code you used to generate each graphic, right above the graph (R Markdown should make this easy).

- A written narrative for each graph, from three to five sentences, describing the context of the graph and how it informs about your policy topic. You can write this as a narrative.

Note that, over time, you are expected to tackle progressively more varied and ambitious graph types. At this stage in the course, the standard is still not very high. That said, you should be able to clearly articulate the value of the graphs you created so far - meaning they need to make sense and have some value-add.

Please submit the URL of the file on a Git repository (it can be one git repo you keep adding to, or a stand alone repository).

Introduction

There are several advantages to dedicate a data visualization portfolio and 10-week effort to digital currency:

- Massive and Robust data source: Blockchain is a public transaction log that exists as a distributed database, validated by its powerful algorithm.

- Fitting data types: Blockchain data includes time and IP addresses, and can be updated frequently, which means it is useful for both the static and dynamic data visualizations.

- Well-supported: I can save the entire blockchain file locally,use community-maintained tools to parse it, and keep the data up-to-date.

- Policy-setting: Digital currencies are being used as store of value to fight inflation in turbulent economies such as Zimbabwe and Latin American countries. Central Banks are issuing experimental policies in response to the dessemination of digital currency and blockchain technology.
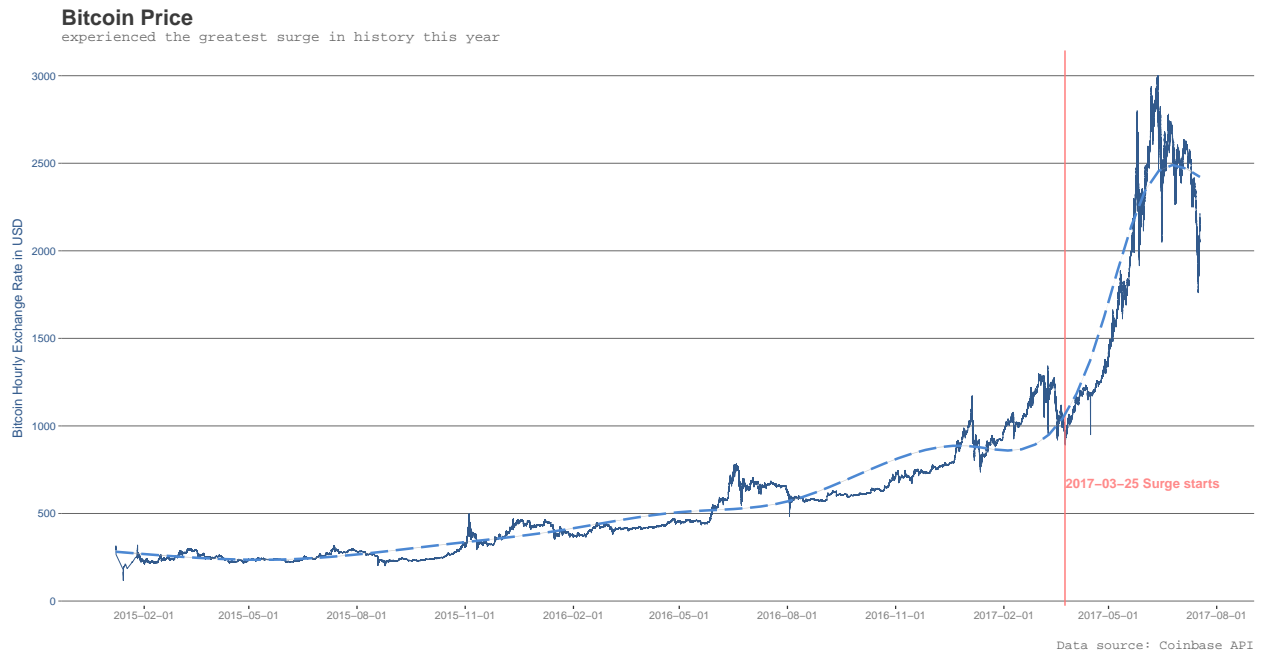
This paper will explore some data sources published by major Bitcoin exchanges and platforms as an market overview.

Graph one - line chart

```r
line <- ggplot(data=coinbase_lim) +
  geom_line(size=.25,
            aes(unix_timestamp, price),
            color="#325a8c") +
  geom_vline(xintercept = as.POSIXct("2017-03-25"),
             colour="#ff7575",
             size=.75,
             alpha=.75) +
  geom_smooth(aes(unix_timestamp,price),
              span=2,
              linetype="longdash",
              color="#4c88d3",
              alpha=.5) +
  scale_x_datetime(name = "",
                   date_breaks = "3 month") +
  scale_y_continuous(name ="Bitcoin Hourly Exchange Rate in USD",
                     breaks = c(0, 500, 1000, 1500, 2000, 2500, 3000)) +
  ggtitle("Bitcoin Price",
          subtitle = "experienced the greatest surge in history this year") +
  labs(caption = "Data source: Coinbase API") +
  annotate(geom="text",
           x=as.POSIXct("2017-05-30"),
           y=675,
           label="2017-03-25 Surge starts",
           colour="#ff7575",
           fontface="bold",
           alpha=.85) +
  theme(panel.grid.major.y = element_line( size=.1, color="#666666"),
        panel.grid.major.x = element_blank(),
        panel.background = element_blank(),
        plot.title = element_text(size=18,
                                  family = "Helvetica",
                                  colour = "#3a3a3a",
                                  face = "bold"),
        plot.subtitle = element_text(size=12, family = "mono", colour = "#666666"),
        axis.title.y.right = element_text(color="#85bb65"),
        axis.text.y.right = element_text(color="#85bb65"),
        axis.text.y = element_text(color="#325a8c"),
        axis.title.y = element_text(color="#325a8c"),
        axis.ticks.y = element_line( size=.25, color="#666666"),
        axis.text.x = element_text(color="#7f7f7f"),
        legend.position = "none",
        plot.caption = element_text(size=11, family = "mono", colour = "#666666"),
        plot.margin = unit(c(2,2,2,2), "cm"))

line

## `geom_smooth()` using method = 'gam'
```

Bitcoin Price

experienced the greatest surge in history this year

Data source: Coinbase API

Bitcoin price is volatile

The price of bitcoin can increase or decrease drastically over a short period of time. This graph mainly serves as a priliminary clue to pinpoint certain time points where a new policy may have impacted the price of Bitcoin, or vice versa. And potentially, this graph could be used to find global price trend correlations.
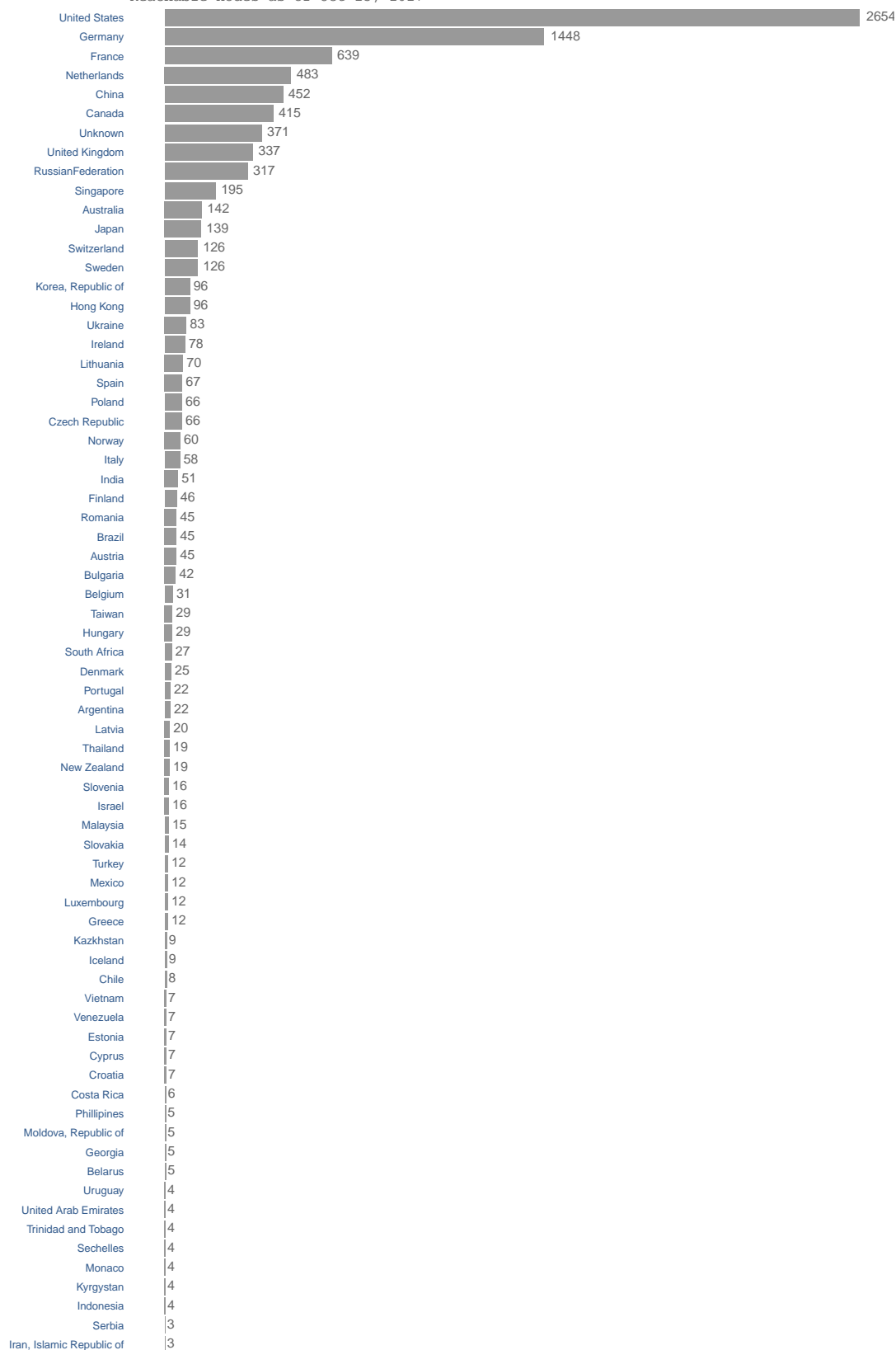
Graph two - histogram

```r
hist <- ggplot(data=nodes, aes(reorder(factor(country), node), node)) +
  geom_col(width=0.9, position=position_dodge(width=5), fill = "#999999") +
  geom_text(aes(label=node),
            color="#666666",
            position=position_dodge(width=0.9),
            vjust=0.35,
            hjust=-.25) +
  ggtitle("Bitcoin Nodes Distribution by Countries",
          subtitle = "Reachable nodes as of Oct 15, 2017") +
  ylab("") +
  xlab("") +
  labs(caption = "Data source: BitNodes.21.co") +
  theme(panel.grid.major.y = element_blank(),
        panel.grid.major.x = element_blank(),
        panel.background = element_blank(),
        plot.title = element_text(size=18,
                                  family = "Helvetica",
                                  colour = "#3a3a3a",
                                  face = "bold"),
        plot.subtitle = element_text(size=12, family = "mono", colour = "#666666"),
        axis.text.y = element_text(color="#325a8c"),
        axis.title.y = element_text(color="#325a8c"),
        axis.ticks.y = element_blank(),
        axis.ticks.x = element_blank(),
        axis.text.x = element_blank(),
        legend.position = "none",
        plot.caption = element_text(size=11, family = "mono", colour = "#666666"),
        plot.margin = unit(c(2,2,2,2), "cm")) +
  scale_x_discrete(expand=c(0,0)) +
  coord_flip()

hist
```

Bitcoin Nodes Distribution by Countries

Reachable nodes as of Oct 15, 2017

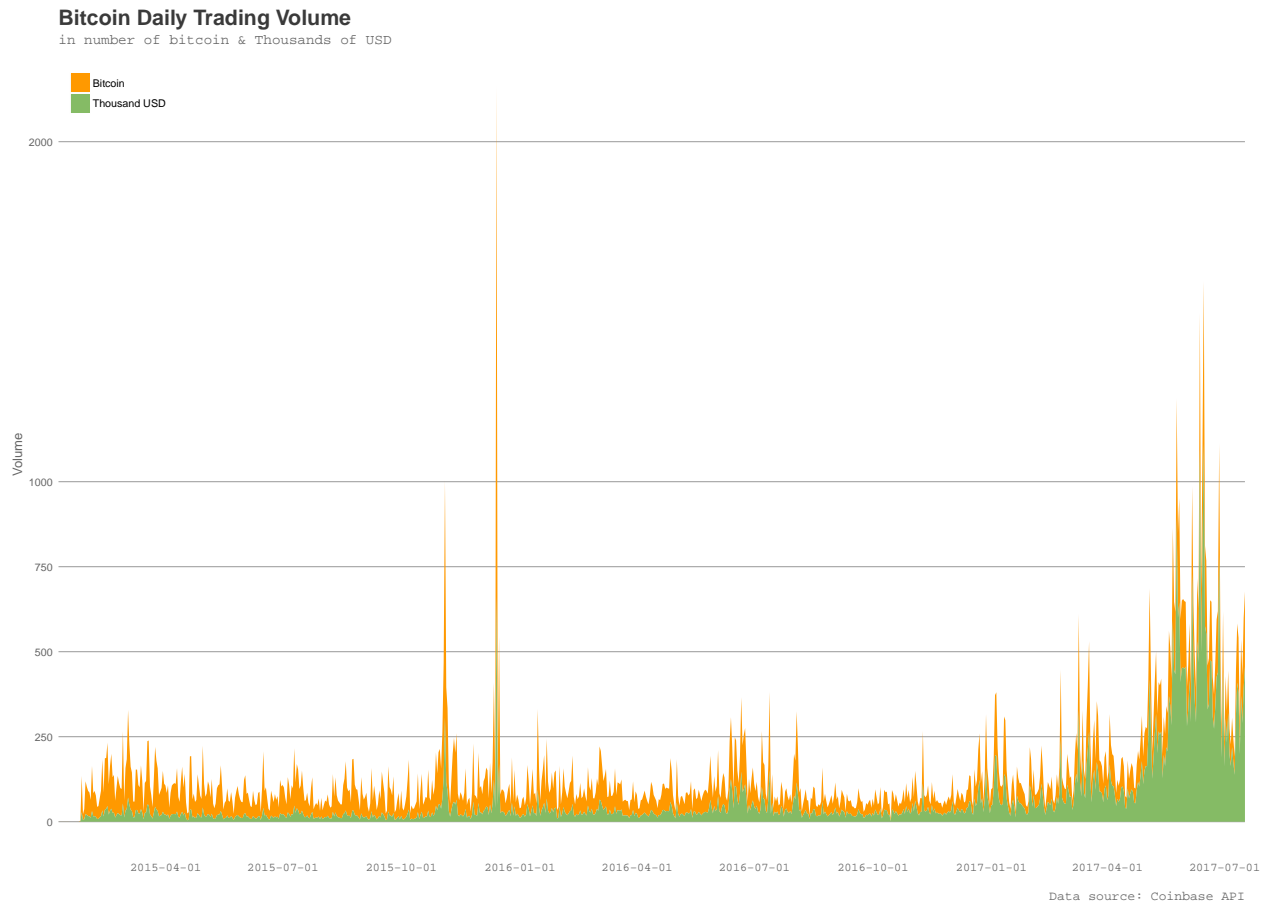| Country	Nodes
United States	2654
Germany	1448
France	639
Netherlands	483
China	452
Canada	415
Unknown	371
United Kingdom	337
Russian Federation	317
Singapore	195
Australia	142
Japan	139
Switzerland	126
Sweden	126
Korea, Republic of	96
Hong Kong	96
Ukraine	83
Ireland	78
Lithuania	70
Spain	67
Poland	66
Czech Republic	66
Norway	60
Italy	58
India	51
Finland	46
Romania	45
Brazil	45
Austria	45
Bulgaria	42
Belgium	31
Taiwan	29
Hungary	29
South Africa	27
Denmark	25
Portugal	22
Argentina	22
Latvia	20
Thailand	19
New Zealand	19
Slovenia	16
Israel	16
Malaysia	15
Slovakia	14
Turkey	12
Mexico	12
Luxembourg	12
Greece	12
Kazkhstan	9
Iceland	9
Chile	8
Vietnam	7
Venezuela	7
Estonia	7
Cyprus	7
Croatia	7
Costa Rica	6
Phillipines	5
Moldova, Republic of	5
Georgia	5
Belarus	5
Uruguay	4
United Arab Emirates	4
Trinidad and Tobago	4
Sechelles	4
Monaco	4
Kyrgystan	4
Indonesia	4
Serbia	3
Iran, Islamic Republic of	3

Data source: BitNodes.21.co

Which countries are more crytocurrency-friendly?

Countries that have more reachable nodes means more people in the countries are involved in Bitcoin mining, which indicates that these countries are more likely to have policies, or the lack of, to support the development of cryptocurrency. For an explorative analysis, this graph is helpful for narrowing down the origin of some major policies that have shaped the development of Bitcoin. If we consider this graph as a ranking of countries' involvement in digital currency. It is interesting to note that, this ranking is not necessarily ordered in exact accordance with the ranking of GDP, population, technological advancement, or even perceived internet presence.

Graph Three - stacked area

```r
# note: color #85bb65 is nicknamed as "dollar bill green".
#       color #FF9900 is the orange from bitcoin logo.
area <- ggplot(data=agg, aes(x = as.POSIXct(date), y = volume, fill = as.factor(unit))) +
  geom_area(position = "stack") +
  scale_x_datetime(date_breaks = "3 months", expand=c(0,0)) +
  scale_y_continuous(breaks = c(0, 250, 500, 750, 1000, 2000, 3000)) +
  scale_fill_manual(values = alpha(c("#FF9900", "#85bb65")), 0.2) +
  ggtitle("Bitcoin Daily Trading Volume",
          subtitle = "in number of bitcoin & Thousands of USD") +
  labs(caption = "Data source: Coinbase API",
       x = "",
       y = "Volume") +
  theme(panel.grid.major.y =element_line( size=.1, color="#999999"),
        panel.grid.major.x = element_blank(),
        panel.background = element_blank(),
        plot.title = element_text(size=18,
                                  family = "Helvetica",
                                  colour = "#3a3a3a",
                                  face = "bold"),
        plot.subtitle = element_text(size=12, family = "mono", colour = "#666666"),
        axis.text.y = element_text(color="#666666"),
        axis.title.y = element_text(color="#666666"),
        axis.ticks.y = element_blank(),
        axis.ticks.x = element_blank(),
        axis.text.x = element_text(size=10, family = "mono", colour = "#666666"),
        legend.title = element_blank(),
        legend.position = c(0.05, 0.95),
        plot.caption = element_text(size=11, family = "mono", colour = "#666666"),
        plot.margin = unit(c(2,2,2,2), "cm"))

area
```

Bitcoin Daily Trading Volume
in number of bitcoin & Thousands of USD

■ Bitcoin
■ Thousand USD

Volume

2000

1000

750

500

250

0

2015-04-01 2015-07-01 2015-10-01 2016-01-01 2016-04-01 2016-07-01 2016-10-01 2017-01-01 2017-04-01 2017-07-01

Data source: Coinbase API

High Trading Volume Volatility:

In this graph, we can see that bitcoin's daily trading volume is extremely volatile, perhaps even more so than its price. This level of volatility can be a sign for market manipulation, especially given Bitcoin's young economy, novel nature, and lack of regulation. Together with the price trend, this graph can help identify major policy changes that either encourage (spike) or discourage (trough) speculation.

Graph Four - network of trust

```r
# View in full screen for clarity.
heatmap <- ggplot(soc, aes(x=rater, y=ratee, fill=score)) + geom_tile() +
  scale_fill_distiller(type="div", palette = "RdYlBu") +
  ggtitle("The trust nework of Bitcoin traders on a platform called Bitcoin OTC",
          subtitle = "is sparse and unregulated") +
  labs(caption = "Data source: Bitcoin OTC trust network by Stanford University") +
  theme(panel.grid.major.y = element_blank(),
        panel.grid.major.x = element_blank(),
        panel.background = element_blank(),
        plot.title = element_text(size=18,
                                  family = "Helvetica",
                                  colour = "#3a3a3a",
                                  face = "bold"),
        plot.subtitle = element_text(size=12, family = "mono", colour = "#666666"),
        axis.text.y = element_text(color="#666666"),
        axis.title.y = element_text(color="#666666"),
        axis.ticks.y = element_blank(),
        axis.ticks.x = element_blank(),
        axis.text.x = element_text(size=10, family = "mono", colour = "#666666"),
        legend.title = element_text(),
        legend.position = c(0.1, 0.9),
        plot.caption = element_text(size=11, family = "mono", colour = "#666666"),
        plot.margin = unit(c(2,2,2,2), "cm"))
heatmap
```

The trust nework of Bitcoin traders on a platform called Bitcoin OTC

is sparse and unregulated

Data source: Bitcoin OTC trust network by Stanford University
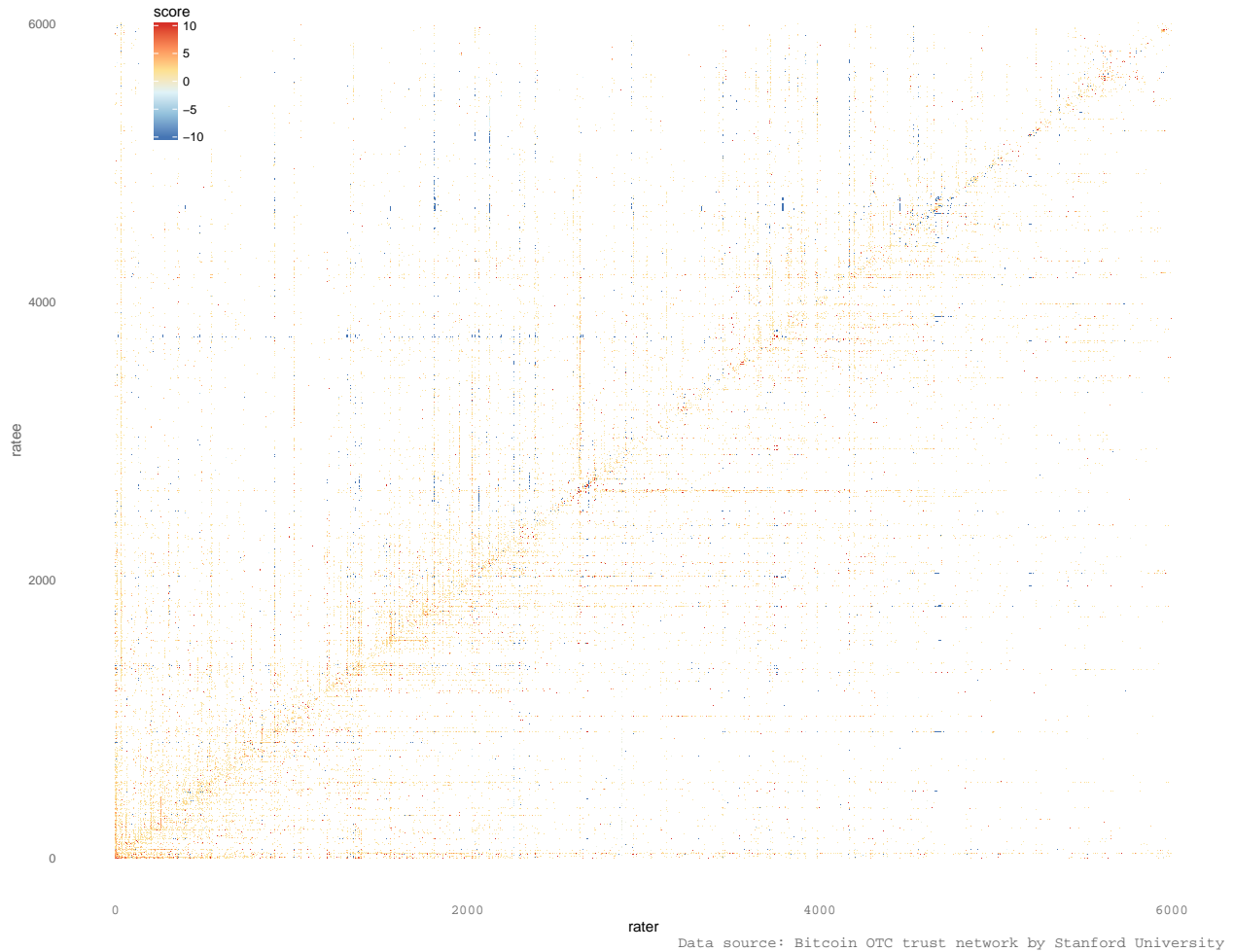
The Who-trust-whom Network:

While sparse and barely visible, this graph does provide some insights about the Bitcoin market. First, it allows us to identify

some super raters and ratees, who are more likely to be the major player capable of manipulating Bitcoin price. Second, although orange (trustworthy) is more prevalent, there are quite a few blue spots on this heatmap, marking the users that are rated as fraud by others. At a glimpse, fraud is frequent in digital currency market, which underscores its unregulated nature.

Graph Five: long-term user network

Low Liquidity:

This is a graph that visualizes the transaction activities of several thousand long-term Bitcoin users during 2009 - 2013. We can see that the majority is inactive, with one user radiating assets to many others, and a few others repeatedly send money to themselves. Not only can this network graph serves as a piece of evidence to show that most people purcahse Bitcoin for the purpose of speculation, it also shows that, while the technological philosophy of Bitcoin is decentralization, its ownership and commercial activity is very much the opposite.

Graph Five: network
```{r, echo=FALSE, include=FALSE, result="hide"}
links <- read_tsv("./data/lt_graph_ELTE.txt", col_names = c("sender", "recipient", "transaction_timestamp"))
net <- graph_from_data_frame(d=links, directed = T)
```

```{r}
# set color
col.1 <- adjustcolor("#666666", alpha=0.4)
col.2 <- adjustcolor("#000000", alpha=0.4)
edge.pal <- colorRampPalette(c(col.1, col.2), alpha = TRUE)
edge.col <- edge.pal(100)

colrs_transparent <- adjustcolor("#FFFFFF", alpha=0)

#net_simplified <- simplify(net,remove.multiple = F,edge.attr.comb=list(Weight="sum"), remove.loops = T)
l <- layout_on_sphere(net)
plot(net,
    layout=l,
    edge.arrow.size=0.1,
    vertex.size=2,
    vertex.color="#FF9900",
    vertex.frame.color=colrs_transparent,
    vertex.label=NA,
    edge.color=edge.col,
    main="The majority of long-term Bitcoin users does not engage in transaction",
    sub="Data source: ELTE Bitcoin Project website and resources (2009 - 2013)",
    output.mode='inline'
    )
legend(x=-0.3, y=-1.1, c("long-term bitcoin user"), col="#FF9900", pch=21, pt.bg="#FF9900")
```


The majority of long-term Bitcoin users does not engage in transaction

long-term bitcoin user

Data source: ELTE Bitcoin Project website and resources (2009 - 2013)