VISUALIZATION PROJECT

An Attempt to Understand the Leading Indicator through Cryptocurrency using
Interactive Visualization

CREATOR NAME: WEI XIN TAN

LAST MODIFIED DATE: 04/06/2018

PROGRAM USED: R

Table of Contents

1. Introduction

To recap, my exploration project is about market anomaly in cryptocurrency. In previous project I ended with below graph showing correlation between the close price and search trend.
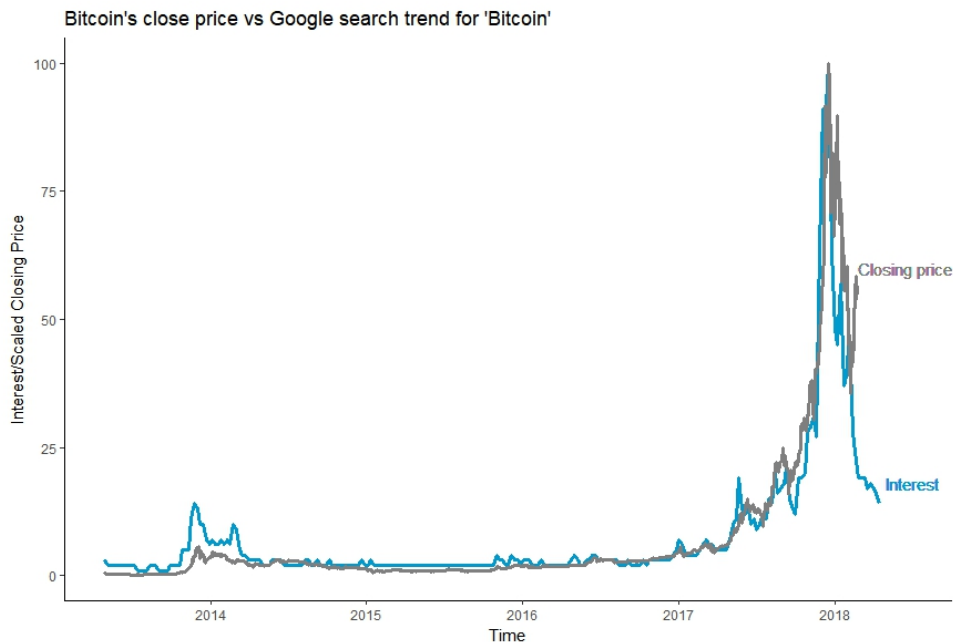

Figure 1: Graph from previous exploration project

However, in the world finance, we are always trying to predict the future, and knowing the fact that two items are correlated to each other is not useful per se. If they are happening at the same time you would not be able use one to predict another. Therefore, there is a concept called leading indicator. When I was taking a finance subject in my undergraduate, I stumble upon this concept, and I remember learning it in a textbook, it tries to explain the relationship between two markets by showing us a static graph. It took me sometime to understand it because I struggled to imagine how two variables are correlated when you shift the curve forward. In this project, I will attempt to communicate the concept through interactive visualization. The target audience would be any student who are studying finance, or someone who has little to none finance knowledge but is interested in this concept.

2. Design with Five Design Sheet Methodology

2. 1 Sheet 1 – Brainstorm

2.1.1. Generate Ideas

Text

1. Time series
4. Indexed charts
7. Center Aligned
10. Dropdown list
13. Drag and drop
16. Scaling data

2. Sliders to shift the curves
5. Labelling
8. Live data for search trend
11. Step by step guide
14. Search bar
17. Use example data to illustrate the concept

3. Dual axes
6. Brief introduction
9. Less colour
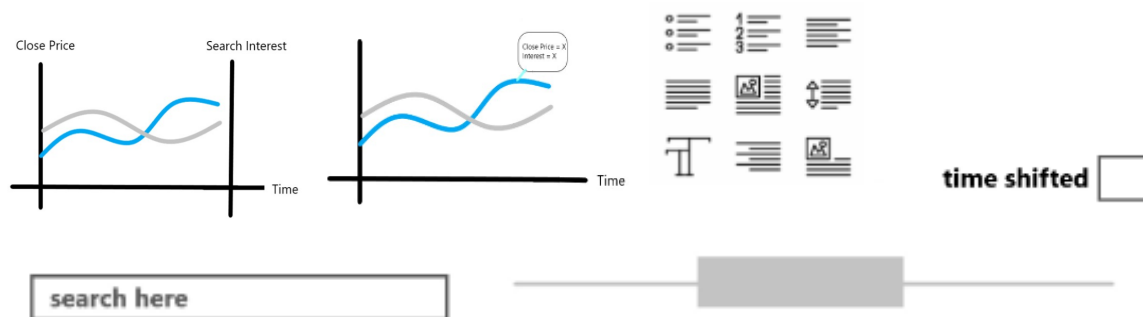12. Date range
15. More info section

Sketches

Figure 2: Generating Ideas for sheet one of five design sheet

We have generated a couple of ideas, and the main idea that can help student to understand the concept is moving the sliders to interact with the time series. All the other ideas are just trying to facilitate the main idea. Some ideas might not make sense at this stage, but together they can create a better narrative visualization. For example, more information section will provide some links to other website for more explanation on certain terminology, this might aid people that are not familiar with some of the financial terms used in this visualization.

2.1.2. Filtering the Ideas

After Categorizing

Main Plot	Sidebar	Design Element	Layout
1. Time series	1. Sliders to shift the curves	1. Drag and drop	1. Brief introduction
2. Labelling	2. Search bar	2. Center Aligned	2. Step by step guide
3. Scaling data	3. Live data for search trend	3. Less colour	3. More info section
	4. Dropdown list	4. Use example data to illustrate the concept	
	5. Date Range		

Figure 3: Filtering Ideas for sheet one of five design sheet

There are a few ideas that aim to deal with two data series with different measure and value. So the first idea is dual axes and it is very common in economic magazine. However, according to an article[1], recently shared by our tutor, it says that "The scales of dual axis charts are arbitrary and can therefore (deliberately) mislead readers about the relationship between the two data series." For our case, although the measure and value of x axis is not really important, but to avoid unnecessary confusion, dual axes will not be used. For indexed chart, as recommended by the same article[1], and shown in below, is a viable alternative. However, imagine if we indexed both our data, and start with a same data point, we will lose our original trend and no matter how we shift the curve it will never be able to align to each other. Therefore, for the main plot, a time series in combination with labeling and scaled data of close price will be used.
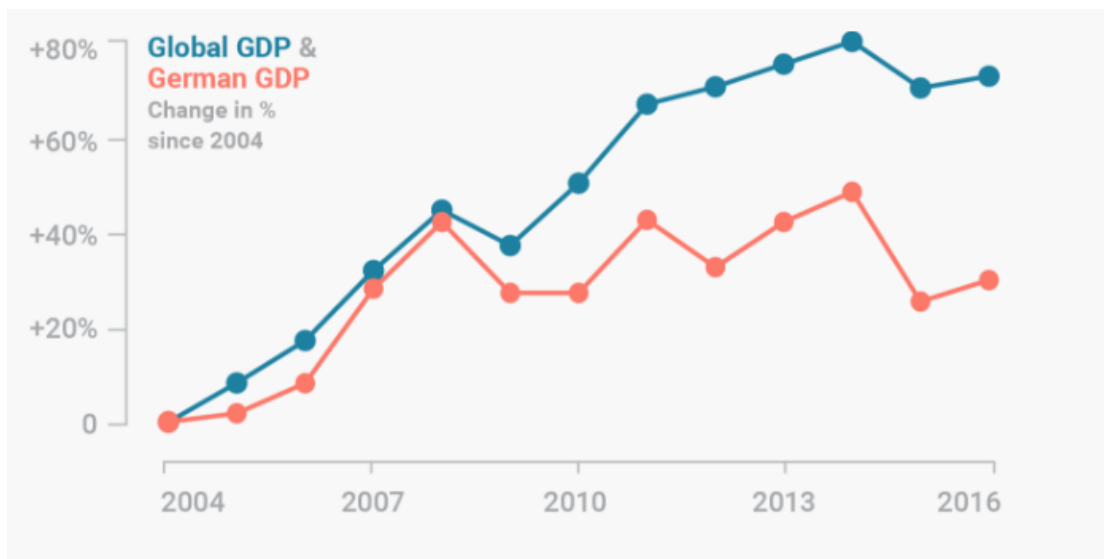

Figure 4: Indexed chart from the article[1]

2.1.3. Categorize the sketches

After filtering the ideas, it is categorized into four types: main plot, sidebar, design element and layout. So now we get a fuller idea of how the visualization would be.

2.1.4 Combine & Refine

2.1.4.1 Main Plot

X's Close Price vs Y's Search Trend

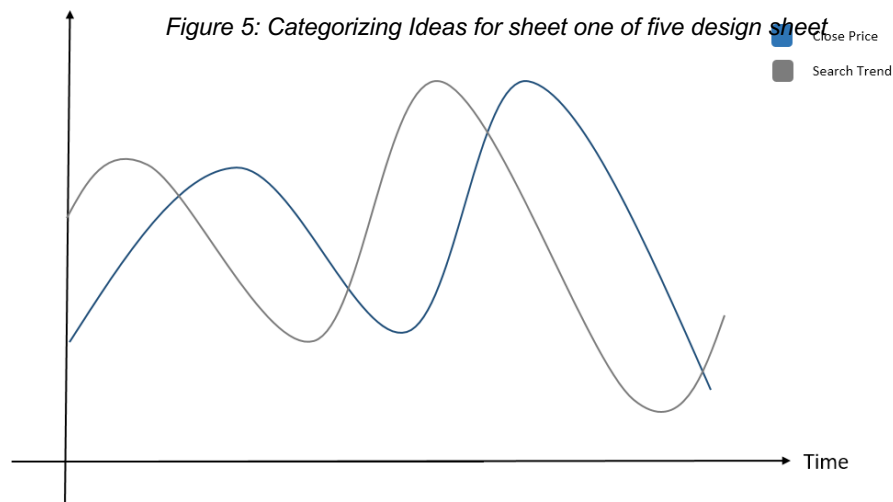*Figure 5: Categorizing Ideas for sheet one of five design sheet*

Figure 6: Main plot of the categorized ideas

Figure 6 is the result of combining the time series, labelling and scaled data. Labelling is used to indicate which curve is close price and which is search trend. The only variable that will be scaled is the close price, scaling from 0 to 100. The google search trend data is already in the scale of 0 to 100.
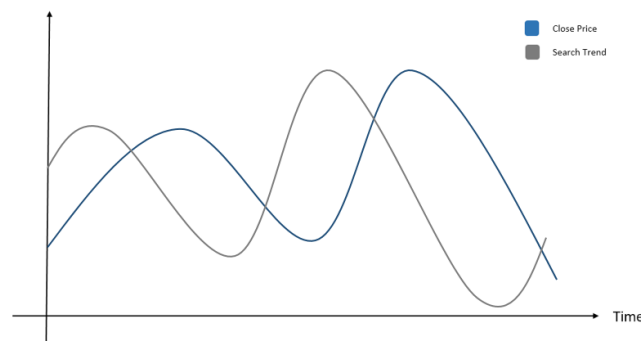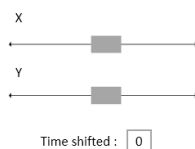
2.1.4.2 Sidebar & Design Elements

Figure 7: Sidebar of the categorized ideas

After adding in the design elements and the sidebars, the graph will look like figure 7. On the left hand side we have the sliders, and on the right hand side we have a dropdown list for selecting the cryptocurrency, and a search bar for google search trend. Lesser colour will be used to prevent distraction, and also to cater for colour blindness. It is tested using Coblis(Color Blindness Simulator)[2]. Since it has only two colours , changing the colour hue will increase the contrast, and ultimately still will be visible in extreme colour blindness condition. In short, a dark colour and a light colour will be chosen for our visualization, refer to appendix to see the test result on the final visualization.

My design choices are often leaning towards simplicity. We can see this theme throughout the presentations, product announcements and talks from Apple, Tesla, Ted, Google and many more. However, we must bear in mind that these presentations are trying to convey very specific ideas or facts, so simplicity might be appealing. But simplicity comes with a price which is missing information. So it might not be useful when we are in the exploration phase, when we are trying to generate new ideas. Also, missing information can sometimes misled us toward certain misconception. Hence, we must be careful with what we are not showing in our graph.
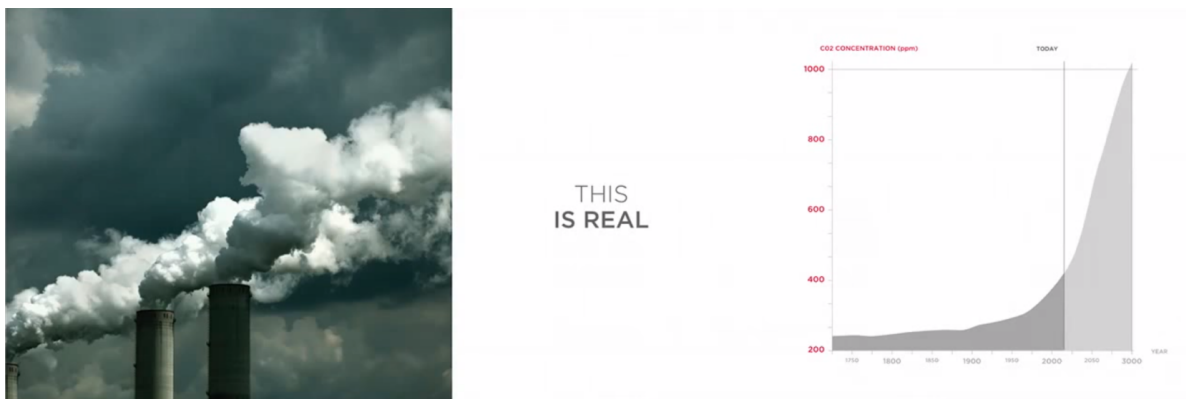

Figure 8: One of the slides when Elon Musk was announcing the Tesla energy

2.1.4.3 Layout

Finally, the layout will be as per figure 9, it will start with introduction and step by step guide, then it goes to plot, and lastly it will end with conclusion and more information section.
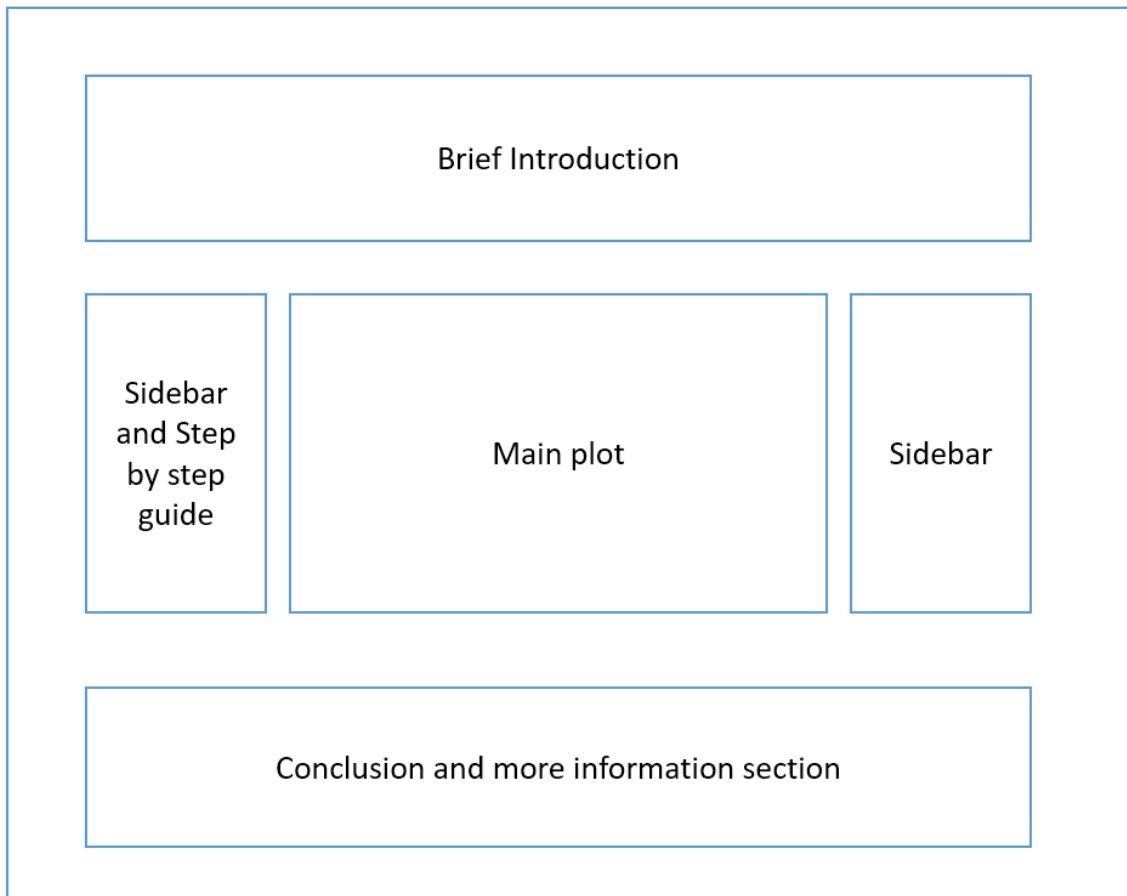
Figure 9: Layout for categorized ideas

2.1.5 Question

So the question would be does this interactive visualization design help student better understand the concept of "leading indicator". Is time series graph the correct and the best way to communicate the idea? I personally think this would be a better learning experience as compared to learning it from a static graph or pure text description. When I design this layout I kept reminding myself to imagine as if I was learning it for the first time.
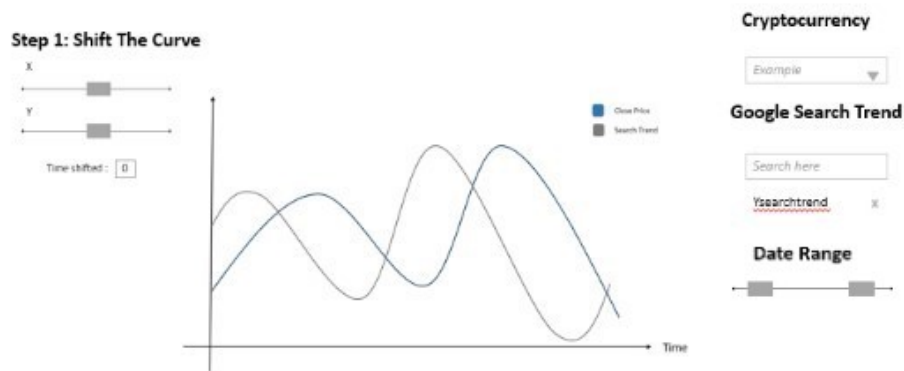
2.2 Sheet 2,3,4 - Initial designs

2.2.1 Layout

After combining all the categories in sheet one, we will get a final layout that looks like below illustration.

1. Brief introduction

2. Main Plot and step by step guide

Step 1: Shift The Curve

X

Y

Time shifted : 0

Close Price
Search Trend

Time

Cryptocurrency

Example

Google Search Trend

Search here

Ysearchtrend   X

Date Range

3. Conclusion and more infomation section

Figure 10: Overall Layout for Sheet 2 of Five Sheet Design

2.2.2 Focus / Zoom

2.2.2.1 Brief introduction and step by step guide

The introduction will be brief. It will explain what 'leading indicator' is by using simple example. The language and tone that will used are more friendly and less formal, since it

is communicating with the student and not professional. This part will also include title and some meta-information.

2.2.2.2 Main Plot

This part is when the user engage with the graph and really try to understand the concept. So it is important to be intuitive and clear. Hopefully, our implementation and design will be able achieve that.

2.2.2.3. Conclusion and more information section

The final part will be a brief conclusion and some additional sources to check out on the topic and the terminologies used.

2.2.3 Operates

There will be step by step guide in the main plot to guide users on how to operate the visualization. Firstly, the user needs to move the sliders to shift one of the search trend curve to match the close price of the cryptocurrency. It will start off with example data, then the users can change it to real data to understand to the concept further. The sliders and the dropdown list are fairly easy to operate, and the output will be showed instantaneously. More on this will be discussed in later section.

2.2.4 Discussion

2.2.4.1 Advantages

I think the strength of this interactive visualization is having a step by step guide. I noticed there are some beautiful and complex visualizations did not include step by step guide, which makes it harder to navigate around.

2.2.4.2. Disadvantage

The shortcoming which mentioned earlier is there is no way to know the actual close price of the cryptocurrency, since it is scaled variable. The plot is also very inflexible, there are only a few things that can be changed. These are the things that being sacrificed for simplicity. The main goal of this visualization is just to understand one simple concept. However, it can be further develop to become a certain tools for predicting the future price of cryptocurrency. For starter, we need to add real-time data for both variables, and the variable must be allowed to assign certain weightage. This will also lead to higher complexity in implementation.

2.2.5 Meta-information

Title: An Attempt to Understand the Leading Indicator through Cryptocurrency using Interactive Visualization

Author: Wei Xin, Tan

Sheet Number: 2

Some of these data will be shown in the final visualization

2.3 Sheet 5 - Realization Design

Other segments remain the same as sheet 2, except discussion is exchanged for some detail.

2.3.1 Details

2.3.1.1 Algorithms

For the title, introduction, conclusion and more information section, the algorithms will be based on Shiny HTML Tags Glossary[3] and some html guide from w3school.com[4]. The Shiny HTML Tags are incredibly useful for creating header, paragraphs, line breaks and hyperlinks.

For the sidebar, the algorithms will be from Shiny Widgets Gallery by Rstudio[5], which includes slider, slider range, and select box. Additionally, the interactive text in the sidebar will be using method from Shiny Tutorial: Text Ouput[6].

Lastly, the main plot will be constructed using 'ggplot', All the algorithms for plotting, theming, scaling data, labelling and colouring are created based on the 'ggplot' reference page by 'tidyverse'[7].

2.3.1.2 Dependencies

The dependency of this visualization will be heavily on 'shiny' and 'ggplot' libraries. Meanwhile, 'dplyr' will also be used for data manipulation. More on this will be discussed in implementation section later. I try to limit the number of dependency as not all machine have the same configuration. This might be trivial but user will appreciate the convenience of it.

2.3.1.3 Hardware requirement

Any machine with browsing (Google Chrome would be preferable) capability will be sufficient to run this visualization. However, the focus is on screens that have wider aspect ratio, instead of taller aspect ratio like mobile devices. These days most media consumption is having 16:9, 16:10 (Macbooks) and similar ratios, for this reason, when designing the sidebars, the vertical space was utilized, instead of placing them top or bottom.

3. Implementation

3.1 Software

This visualization was implemented using R with Shiny. The libraries used are 'dplyr' and 'ggplot'. 'dplyr' is used for grouping data and 'ggplot' is used to create the time series.

3.2 Data

The cryptocurrency data is from Kaggle Dataset[8], It contains 1581 cryptocurrencies, and it has 758,835 rows. Even loading the data takes some time, so I decided to shrink the data by selecting only 10 cryptocurrencies based on their ranking and existence time. For google search trend, initially 'gtrendsR' package was used to extract the live data, but during the development time it has some problems[9]. Therefore, I have to resort to download the data manually from Google Trend website[10]. I randomly picked 10 of them, some are closely related to cryptocurrency. However, the silver lining is now it is faster compared to extracting it from the R package which takes a few second. I also remained the code for future development. Both "Example" data are created using cosine function for demonstration. The purpose of data processing here is mainly to increase the speed.

3.3 Program design

3.3.1 Static Plot

The first stage is to come out with the main plot with the intended design. In this case, ''ggplot' is used to construct the title, subtitle, labels and the curves for both variable. Through several trials and errors, the static main plot will be built as per design requirement. After that, new variables will be redefined for those variables that need to be interactive.

3.3.2 Shiny Server

The second stage would be building the Shiny server around the main plot. The pipeline feature from 'dplyr' will used for filtering and selecting data as per below code.

```
interest_data <- interest_csv %>% filter(search_term==searchinput) %>% select('date','hits')
```

The third stage would be building the Shiny UI. At this stage, the narration, sliders, dropdown list, step by step guide and more information section will be added.

Lastly, all the design elements will be added in, details on design choices will be discussed in the next section. It is done by looking at code using browser's inspection tools, and changing it in the Shiny UI.

3.4. Aesthetic Design

Although the visualization looks plain and simple, but it takes some efforts to make it this clean. From the font to the line-height, everything is handpicked to specifically make it look this way. Roboto is my favorite font which is widely used by Google. Pair it with slightly off black (#444444), 1.4 line spacing and justify text alignment, I think it makes the text easier to read. If the font is not installed in your machine, it will use Helvetica instead, Which is a good font too, used by Apple website, and it used to be on iOS too, unsurprisingly it is also the default font for Shiny.

As for the sidebar, we remove all the default box style, background color, shadow et cetera to make it looks flatter as it does not provide any utility in this context. The sliders button also has been changed from round to rectangular rounded edges to make it feel

more like a slider. If you notice the recent development in operating systems, especially android and IOS, you will see they have changed their design from 3D looking to a flatter looking design. This is where I got the inspiration.

Figure 11: Images downloaded from Googe, from left to right: Android icon evolution, IOS icon evolution and Google icon evolution.
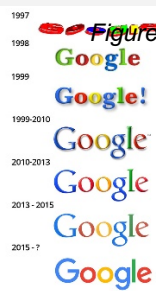
For the main plot, the classic theme from 'ggplot' is used, as it provided a more minimalistic look. With that, we can easily shift the focus to the curves (our most important element) by adding colours to them. Furthermore, to ease visibility, the tick labels for x axis are rotated to 45 degrees, and only displaying month and year for every three months.

3.5 Failed to implement

Unfortunately due to my incompetence, the drag and drop feature was not able to implement.

3.6 Final Product

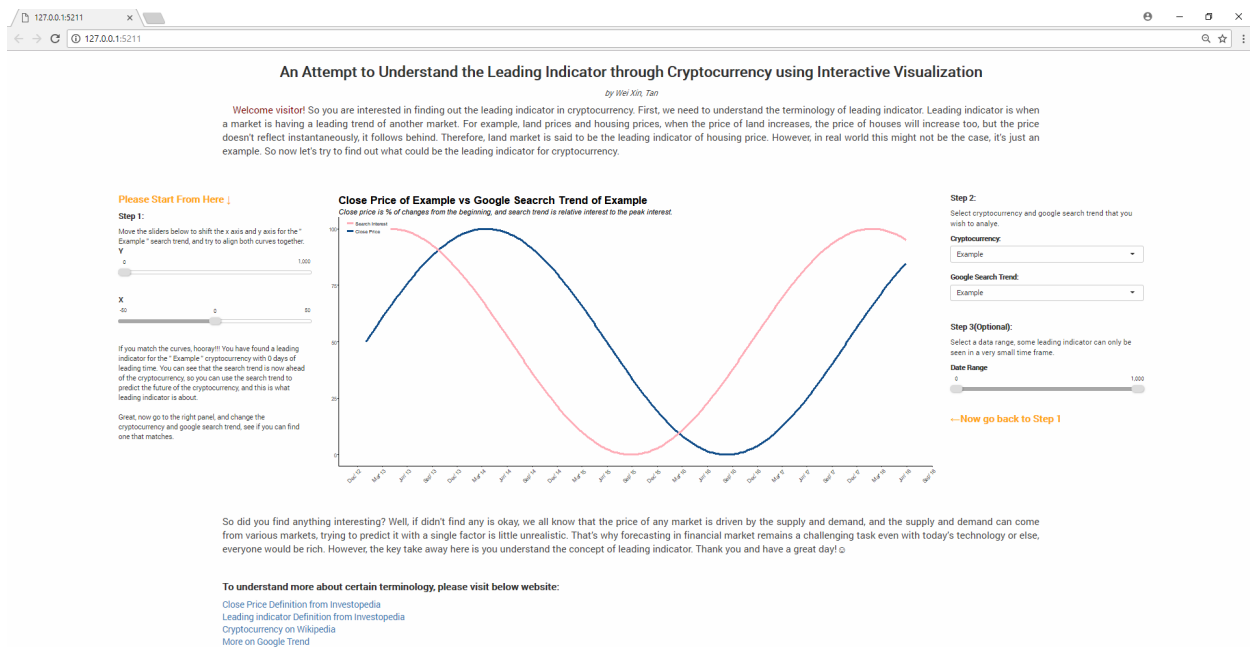Below is the final product, and to get a better viewing experience, please open it in full screen mode.

Figure 12: Final product viewed in Google Chrome

4. User Guide

4.1 Step 1

Read the title and the introduction.

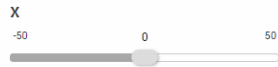Figure 13: Introduction part from the final visualization

4.2 Step 2

This part you would need to move the search trend forward to align with the close price as show below to understand what is leading indicator.

Step 1:

Move the sliders below to shift the x axis and y axis for the " Example " search trend, and try to align both curves together.

Y

0   352   1,000

X

-50   0   50

If you match the curves, hooray!!! You have found a leading indicator for the " Example " cryptocurrency with 352 days of leading time. You can see that the search trend is now ahead of the cryptocurrency, so you can use the search trend to predict the future of the cryptocurrency, and this is what leading indicator is about.

Great, now go to the right panel, and change the cryptocurrency and google search trend, see if you can find one that matches.

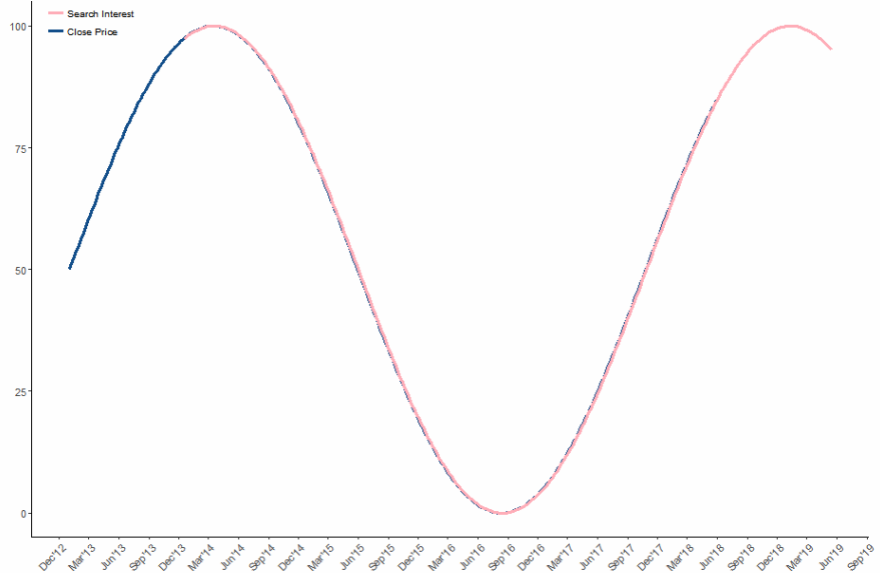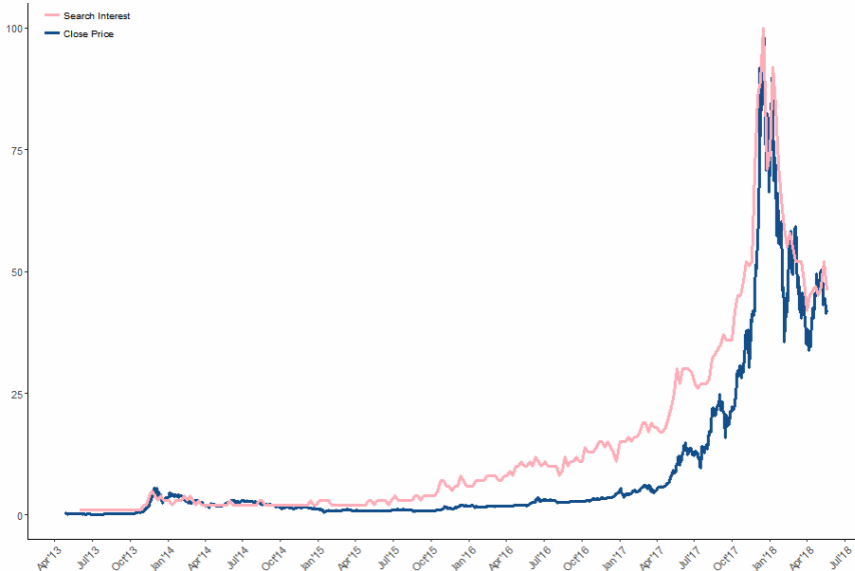*Figure 14: Step 1 of the interactive visualization*

4.3 Step 3

Now go to the right panel, and change the cryptocurrency and google search trend.

Step 2:

Select cryptocurrency and google search trend that you wish to analye.

Cryptocurrency:

Bitcoin

Google Search Trend:

Blockchain

Step 3(Optional):

Select a data range, some leading indicator can only be seen in a very small time frame.

Date Range

0   1,000

←Now go back to Step 1

Figure 15: Step 2 of the interactive visualization

4.4 Step 4

Go back to step 2, and repeat the process, see if you can find one that matches like the example data.

4.5 Step 5

Read the conclusion to understand why it is very difficult to find leading indicator in the real world. If there is some terminology that you want to understand more, click the links provided in the more information section.

So did you find anything interesting? Well, if didn't find any is okay, we all know that the price of any market is driven by the supply and demand, and the supply and demand can come from various markets, trying to predict it with a single factor is little unrealistic. That's why forecasting in financial market remains a challenging task even with today's technology or else, everyone would be rich. However, the key take away here is you understand the concept of leading indicator. Thank you and have a great day! ☺

To understand more about certain terminology, please visit below website:

Close Price Definition from Investopedia
Leading indicator Definition from Investopedia
Cryptocurrency on Wikipedia
More on Google Trend

Figure 16: Conclusion and more information section of the final visualization

5. Conclusion

By looking at the code and final presentation, it would not be deemed as difficult. But there are considerable efforts have been made to come up with the design and the implementation code. For design, the difficulty would be choosing the right colour, the right graph, the number of optimal sliders before the interactive becomes too complicated and the right design element to make the visualization more pleasant. As for implementation, creating a website-like experience using Shiny required some learning from Shiny documentation and HTML syntax. By no means am I implying this would be difficult for other people. But doing it as a first hand and for the first time, it was a challenge.

Hopefully I have achieved my goal of explaining 'leading indicator' in a better way using interactive visualization. I learnt that creating a good visualization require tremendous efforts in design and implementation, so next time when I see a good visualization, I would not take that for granted, I will instead try to learn from it and improve upon my own visualization. The thing that I would have done differently is adding a tooltip so that there is a way to find out the close price, and a reset to default button would be a nice touch too.

6. Reference

1. https://blog.datawrapper.de/dualaxis/

2. http://www.color-blindness.com/coblis-color-blindness-simulator/

3. https://shiny.rstudio.com/articles/tag-glossary.html

4. https://www.w3schools.com/Html/

5. https://shiny.rstudio.com/gallery/widget-gallery.html

6. https://shiny.rstudio.com/reference/shiny/1.0.2/renderText.html

7. http://ggplot2.tidyverse.org/reference/index.html

8. https://www.kaggle.com/jessevent/all-crypto-currencies/data

9. https://github.com/PMassicotte/gtrendsR/issues/166

10. https://trends.google.com/trends/

7. Appendix

Result from COBLIS, even with Monochromacy/Achromatopsia, the contrast still remain visible.