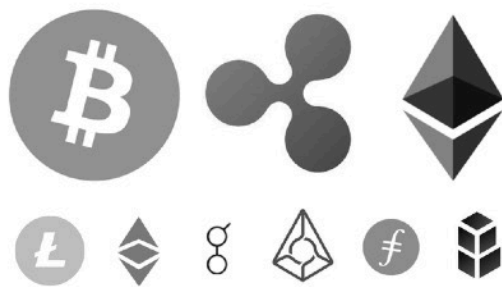


Process Book

Data Visualization - Project

Visualize cryptocurrencies
in a new way



Olivier Couque, Emma Lejal Glaude, Matthieu Sauvé

Fall 2017

Overview, Motivation and Target Audience

The initial idea came from our impression of seeing always similar visualizations when talking about the value of cryptocurrencies among time. This usual representation as lines made us dazzled because people without any financial knowledge would not understand much. Here is an example from the website coinmarketcap.com/currencies/bitcoin.



When we look at this representation with a fresh eye, the difference between the blue and green curve is unclear. The grey bar graph below is also unclear. Also when looking at the duplicated curve in the brush area it seems that the evolution is different than the green curve. Moreover the logarithmic scale can be very difficult to understand for people without background in mathematics.

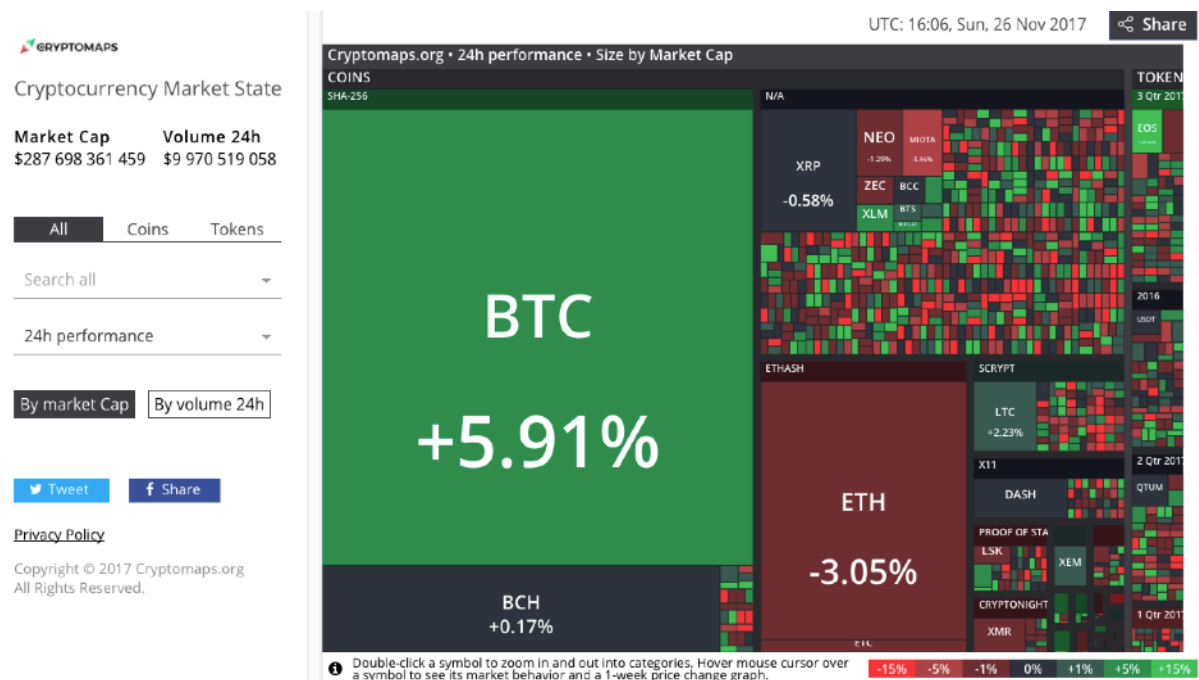
This aside, with this many information it is impossible to have multiple cryptocurrencies displayed at the same time.

For this project, we decided to tackle the problem of visualize the evolution of cryptocurrencies for a public without strong financial or mathematical backgrounds. The idea is to express quickly the intuition of win or loss on a given period of time for the selected cryptocurrencies.

We would then target newcomers to cryptocurrency world and help them appreciate the evolution by simulating a past investment.

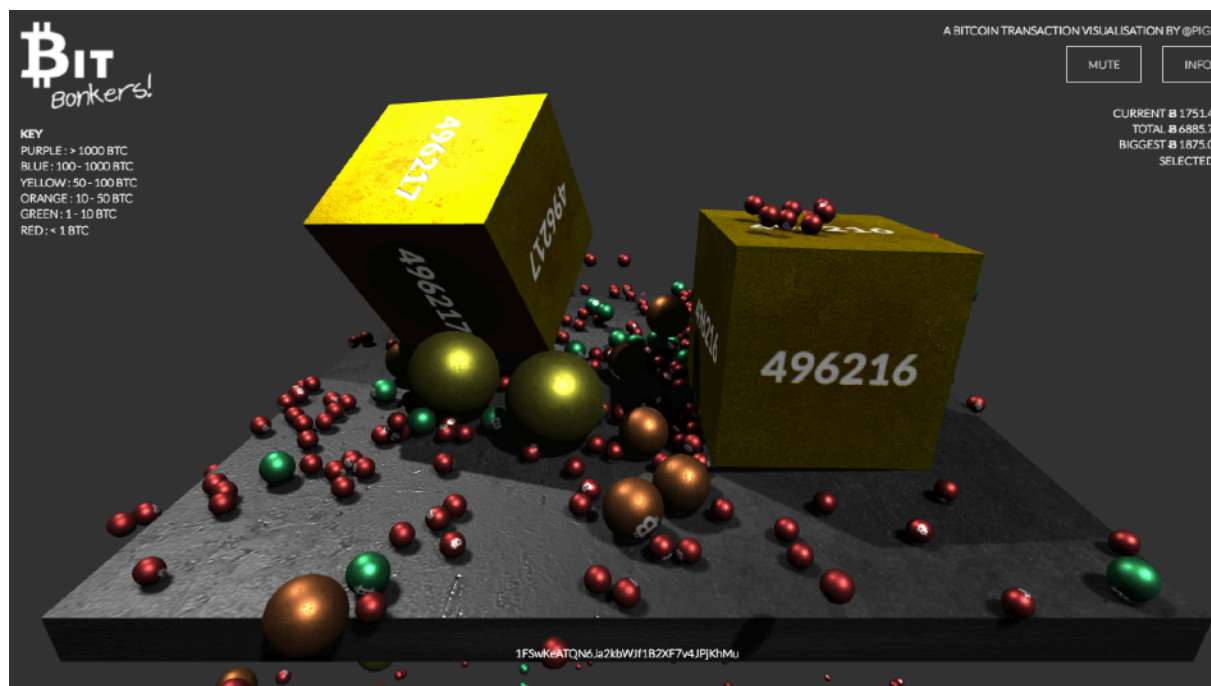
Related Work and Inspiration

During our research to get ideas of visualization, here is what we encountered :



Source : <http://cryptomaps.org/>

What we liked about this visualization is that area provides instant impression and comparison (when the differences are as big as in this view). Although the smaller areas are very difficult to understand. We thought that the colors did not mean much either.



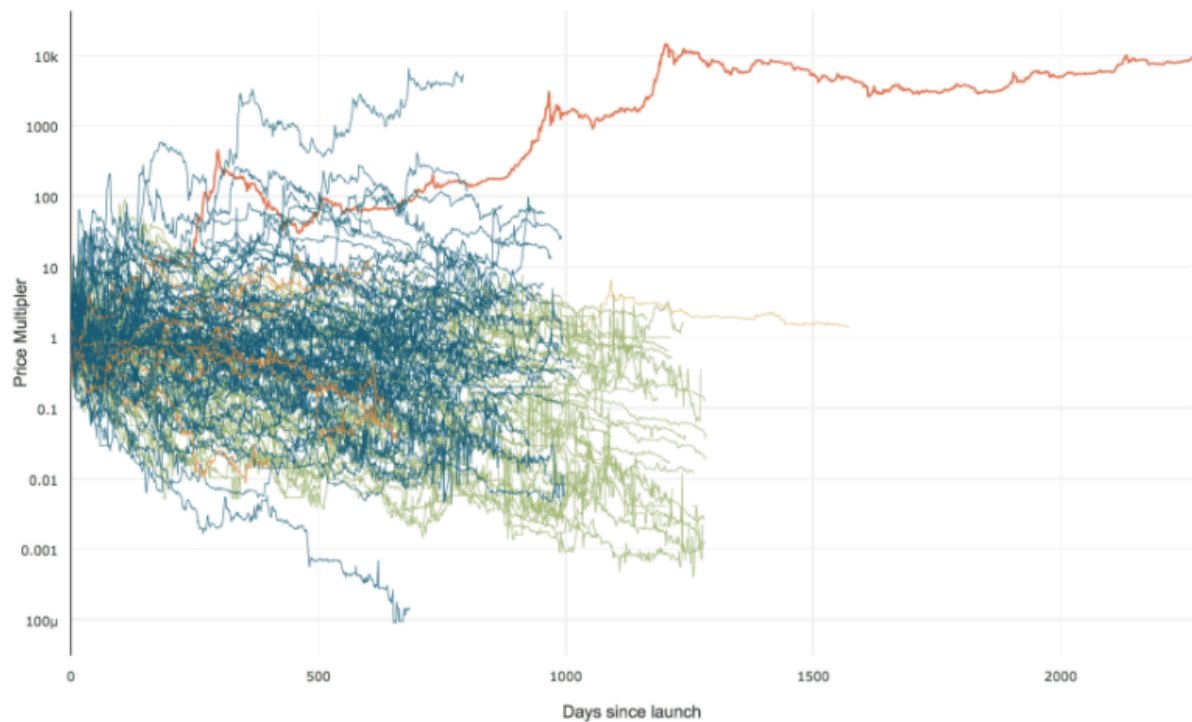
Source : <https://bitbonkers.com/>

What we liked about this visualization is that you realize easily that the cryptocurrency market is always moving. It took us some time to understand what the colors meant and what the cubes are. The colors are actually interval of values of transactions which is also encoded with the size of the spheres. The cubes are the last mined blocks. The gravity to make the sphere is not fully respected, which is acceptable, but the only purpose of having the spheres drop is not to surcharge the screen. In the end, we did not really get the use of 3D scene.



Source : <https://blocks.wizb.it/>

This visualization helped us understand what we did not want in our visualization. The animation is moving very fast, making it hard for the eye to understand that the globe actually rotated from the USA to Europe and back to USA in a very short time. Also the exact position and amount is, in our opinion, too much detail for the goal of our visualization.



Source : <http://woobull.com/data-visualisation-118-coins-plotted-over-time-this-is-why-hodl-alt-coin-indexes-dont-work/>

This is more data art than a visualization but we enjoyed that it showed the fact that we hear a lot of the very successful cryptocurrencies but the many loose value. It is a rare point of view. We also see clearly that this phenomenon is recent and has a lot of activity.

Question

What am I trying to show with my visualization ?

The aim of our visualization is to show quickly the gain or loss of different cryptocurrencies and easily compare them.

We wish to provide a tool that beginners in cryptocurrencies can use to evaluate which cryptocurrency would have provided the best investment on the time period of their choice. This would be in percentages or with an absolute amount that the user can input.

Indeed when start wondering about investments in cryptocurrencies, you could wish to play with different setups (which cryptocurrency to invest in, one or several, what amount to invest, at what time, what would have been your output money if you had sold on this specific date, etc.). Our goal is to provide this tool with a clear distinction between win and loss. We would not want people to only think that they can win with this kind of investments. This is also a way to demonstrate the variations over time and how unpredictable your win or loss is.

Dataset

Where does it come from, what are your processing steps ?

Our original dataset comes from the platform Kaggle. We used the dataset originally aimed to be used in a Machine Learning challenge. Here is the link : <https://www.kaggle.com/sudalairajkumar/cryptocurrencypricehistory> .

We chose this dataset as we had various cryptocurrencies in the same format. Moreover with the preview available on the Kaggle platform we knew that the data was clean enough to be able to work quickly. We also had the possibility to work on specific visualization of either Bitcoin or Ethereum as those two cryptocurrencies had some additional files.

This dataset is interesting as we have the data for a significant time period. Indeed we have the evolution from April 28, 2013 to September 14, 2017. Although the Bitcoin existed before this period of time we did not try to gather the former data as the cryptocurrency world was at the time only composed of experts and scientists. The comparison would also have been poor as Litecoin was the only existing alternative to Bitcoin.

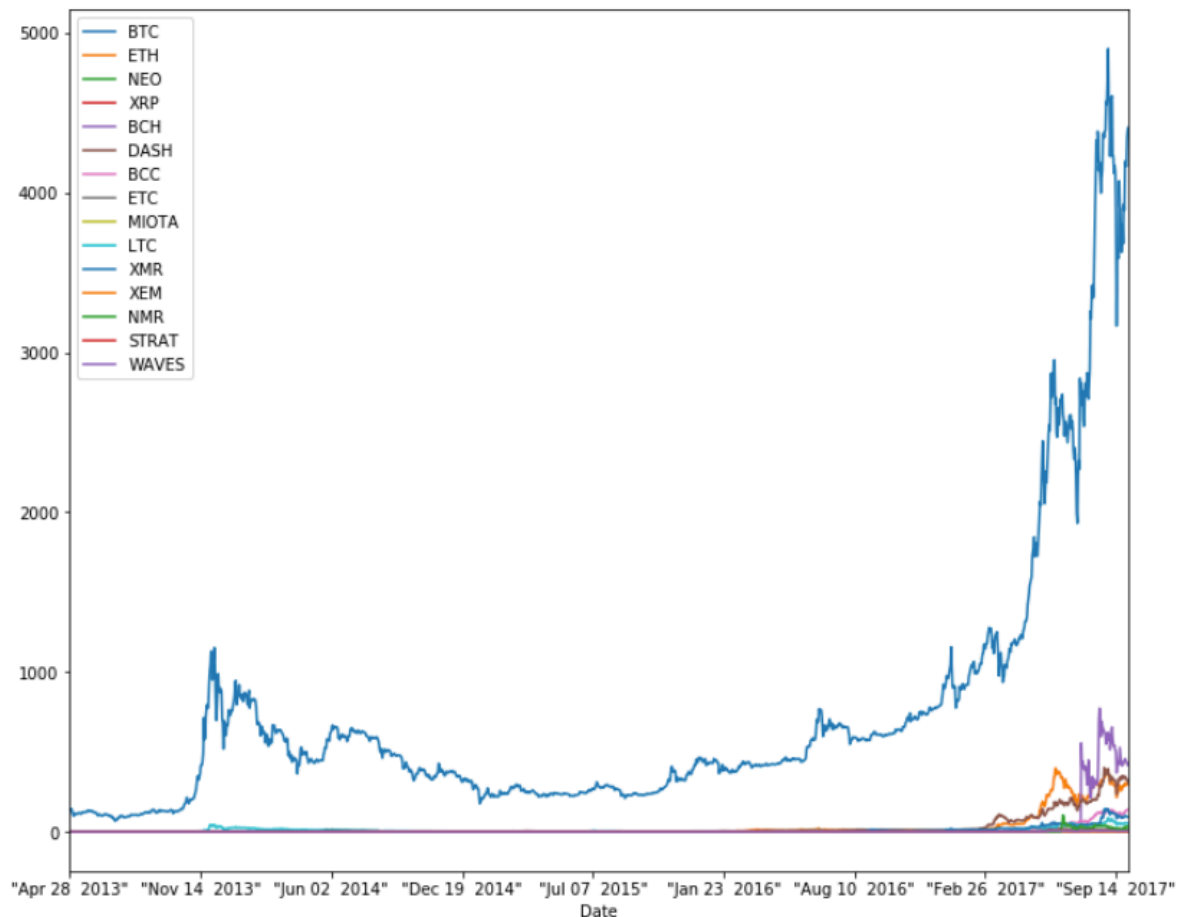
About the preprocessing, what we did is basically merge all the separate price files. In order to do this we used a Jupyter notebook and the Pandas library. The header of every file, giving the column names, is the following : Date, Open price, High price, Low price, Close price, Volume, Market Cap. In the purpose of this visualization we are only interested in the Open price. This can easily be changed later.

We implement a function, named `csv_into_df`, that given the path to a file and the name to attach to this cryptocurrency returns a Pandas Dataframe with each Date associated to the Open price. Once we have this we create the final Pandas Dataframe by merging the data frames given by the function `csv_into_df`. The result will be a column with the date and a column per cryptocurrency with the open price as values. We make sure that there are no missing values by setting the unknown to zero. Finally we export the final Dataframe as a csv file, which is a format easily readable with the d3 library.

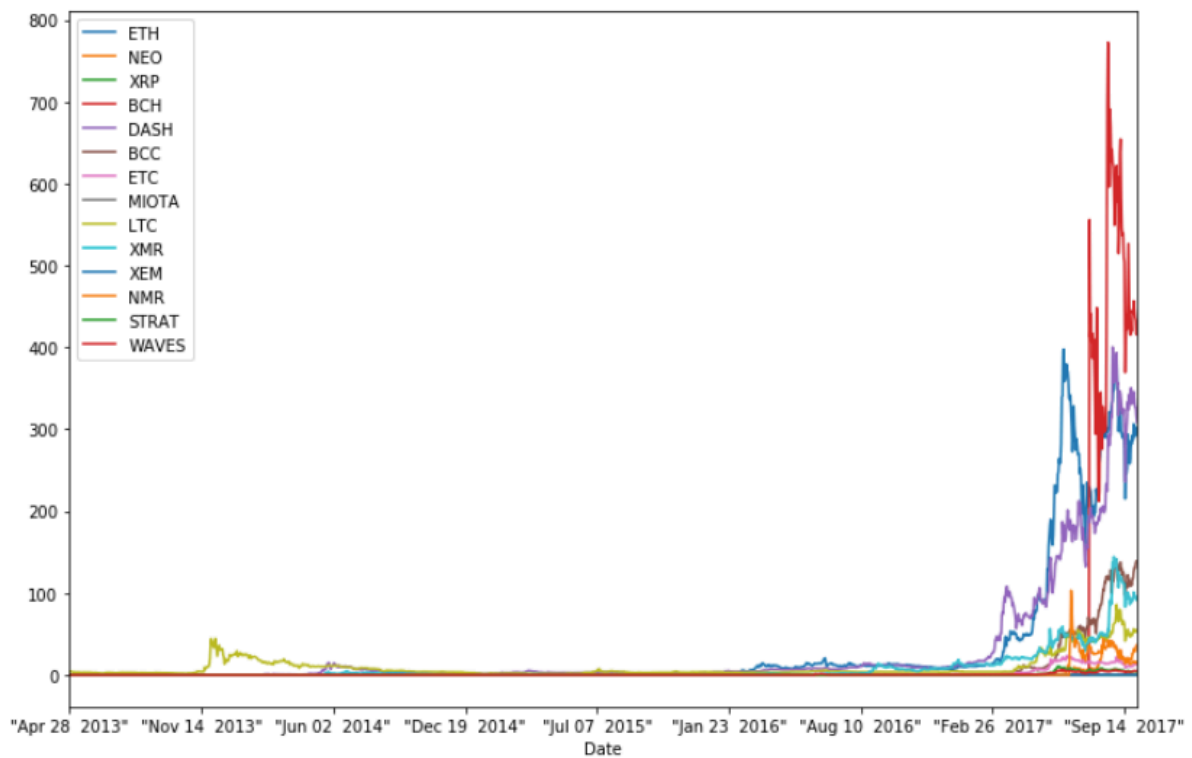
Exploratory Analysis

What visualization have you used to gain insights on the data ?

As our data is based on a suite of numerical values that corresponds to a time evolution, we used the basic representation that we want to escape from to know what our data looks like. That is to say we plotted a line plot with time as x-axis and a curve per cryptocurrency.



In this first plot we show the evolution of the opening price of all the cryptocurrencies for which we have some data. As we see that the graph is very dominated by the Bitcoin, we provide another view without this cryptocurrency.

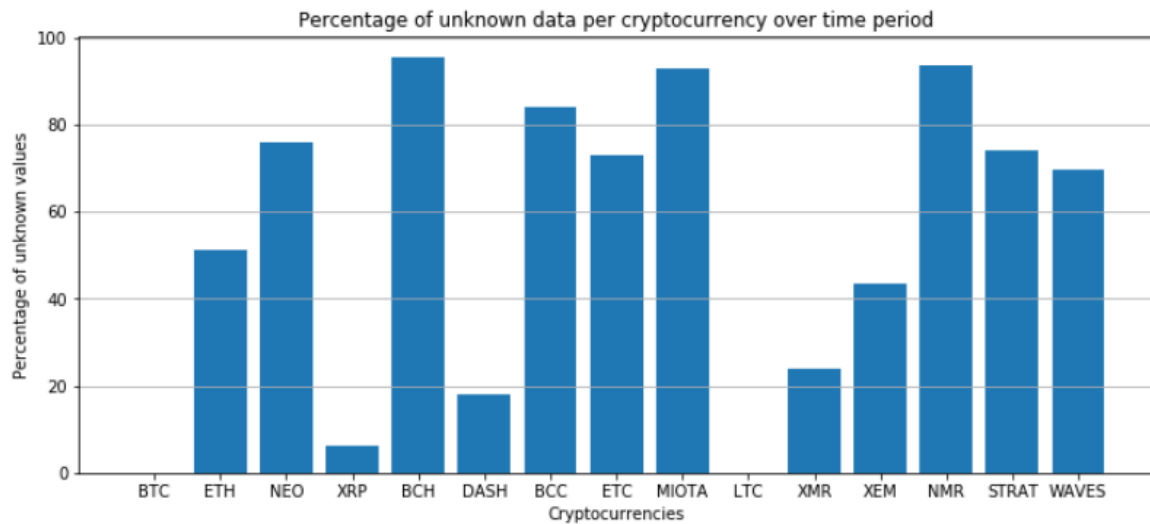


We can see in this visualization that the data is concentrated after February 2017. Most cryptocurrencies saw their value rise considerably after this date. This is something to keep in mind as we go and implement our visualization. Indeed, for most of the time intervals (approximately before January 2016), we have a limited number of cryptocurrencies to compare.

To investigate this last point we illustrate in the following figure the start date of each cryptocurrency, according to our dataset. We can see that we have the value of Bitcoin (BTC) from the moment that Litecoin had a significant value (LTC). Indeed some cryptocurrencies with values below 10 cents appear as 0 in the dataset.

BTC	from	"Apr 28 2013"
ETH	from	"Aug 07 2015"
NEO	from	"Sep 09 2016"
XRP	from	"Aug 04 2013"
BCH	from	"Jul 23 2017"
DASH	from	"Feb 14 2014"
BCC	from	"Jan 20 2017"
ETC	from	"Jul 24 2016"
MIOTA	from	"Jun 13 2017"
LTC	from	"Apr 28 2013"
XMR	from	"May 21 2014"
XEM	from	"Apr 01 2015"
NMR	from	"Jun 23 2017"
STRAT	from	"Aug 12 2016"
WAVES	from	"Jun 02 2016"

To quantify in a better way what proportion of the data was unknown, or equivalently how recent a cryptocurrency is, we did a plot of the percentage of unknown data per cryptocurrency.



We verify here that Bitcoin and LiteCoin have no unknown data (percentage is 0) and that the most recent cryptocurrency, namely Bitcoin Cash (BCH), has the higher percentage of unknown data (95.49 %).

Design

What are the different visualizations you considered ? Justify the design decisions you made using the perceptual and design principles

Our design process followed the design sheets that were provided in the course. You will find them in the appendix in full version. We will provide some insights here.

Understand

Brainstorm design requirements :

- Differentiate cryptocurrencies (while being consistent with other website representations)
- Select different periods of time (start & duration)
- Easily measure win or loss values
- Show timeline

Rank requirements :

1. Show timeline
2. Easily measure win or loss values
3. Differentiate cryptocurrencies
4. Select different period of time

Ideate

How might we questions :

- How might we show the win/loss value ?
- How might we select the cryptocurrency ?
- How might we select the time period ?

Our first sketch had drop down select menus for start date, end date and cryptocurrencies.

We later changed this to a list of button to select or not each cryptocurrencies and a time line under the graph to select an interval of dates.

The second sketch displayed the cryptocurrencies as lines which is what we don't want to keep in the usual graphs.

In the final sketch we made a bar chart centered at 0 bars would be going up to show a positive value (meaning a win) and down for a negative value (meaning a loss).

We still regret that we can only have one period of time selected at once on a single window.

Make

Achievable goal : communicate without financial knowledge the potential past win or loss of investment in cryptocurrencies.

Encoding & Layouts :

- Cryptocurrency : name, abbreviation, logo, usually associated color
- Time : horizontal axis, selection with an interval on the axis
- Win/Loss : bar chart centered around 0, different colors for positive and negative bars

Interactions :

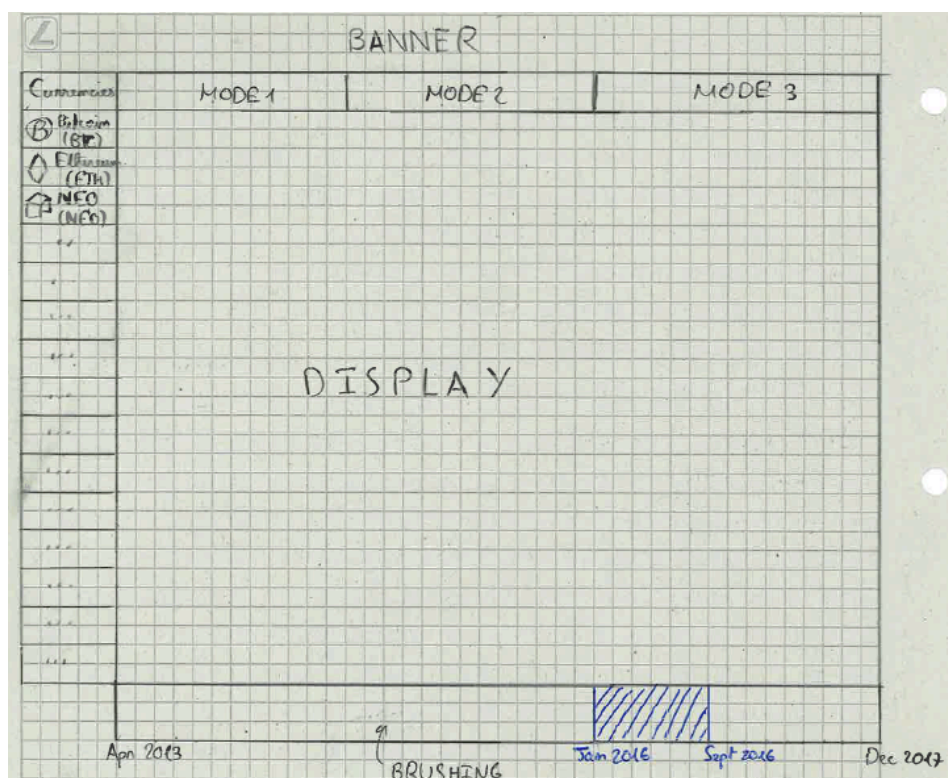
- Selection of time period
- Shift selected period in time
- Selection of desired cryptocurrencies
- Hover a cryptocurrency to make it more visible

Additional views :

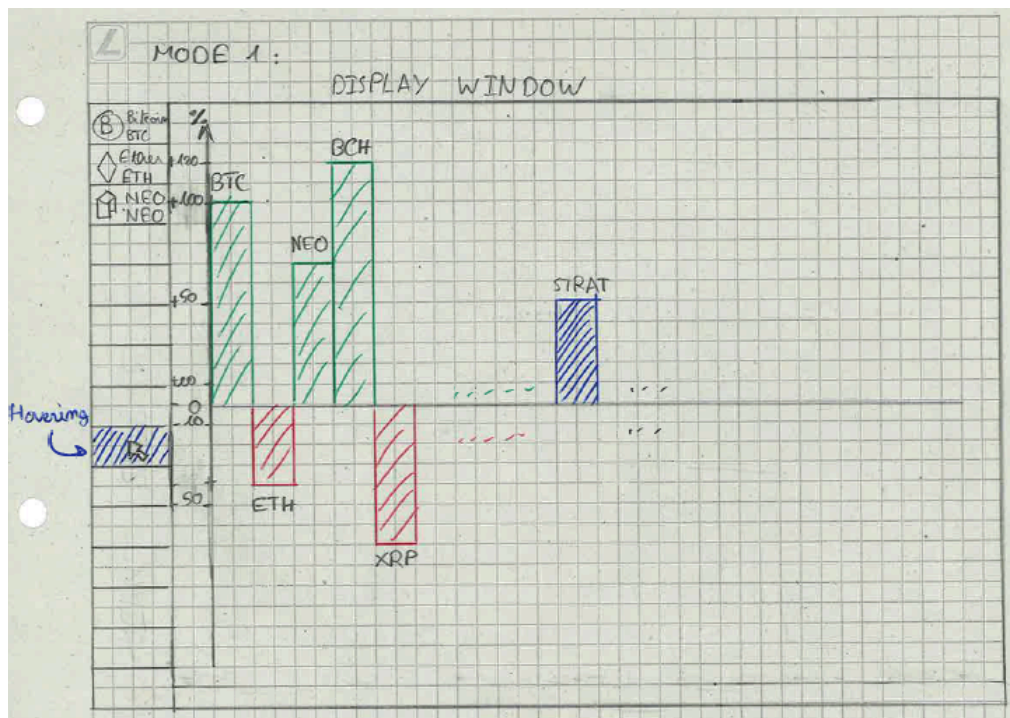
1. Isolated view : Given a duration (a week, a month, two months,...), give for every interval the best cryptocurrency to invest in
2. Consecutive : Given a duration and an amount, give the final amount you would have earned if you had invested this amount every duration in the best cryptocurrency during this duration.

Limitation : Data is limited to a few number of cryptocurrencies (15 out of +1000) and we cannot predict the future evolution, all computations are past-based.

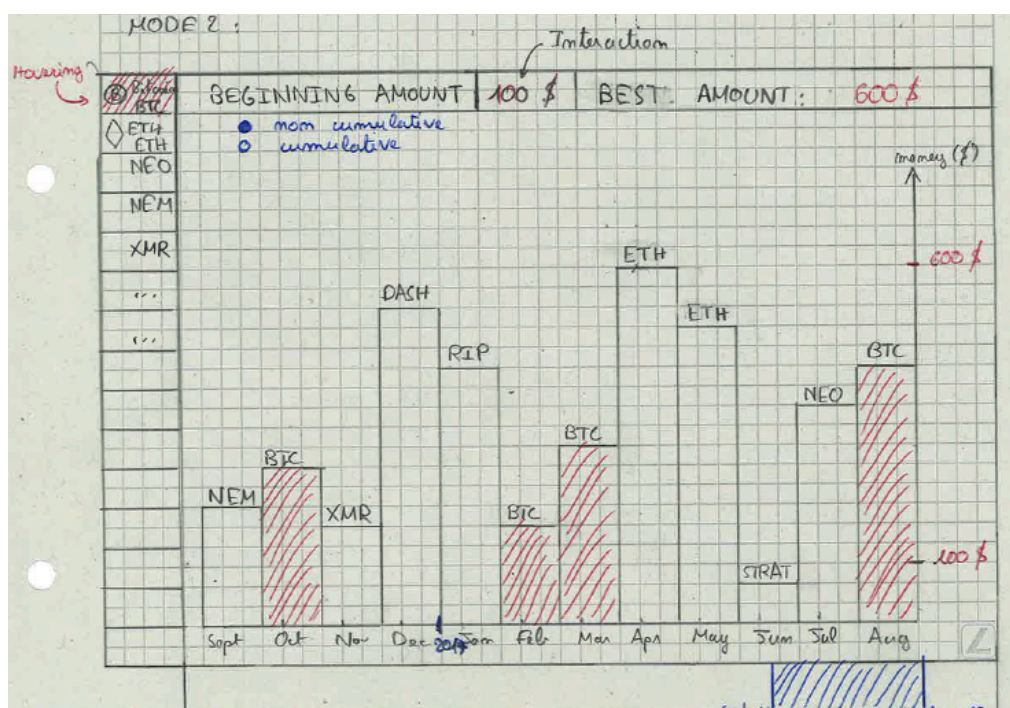
The following scheme shows the general structure of our visualization. The « Display » area will welcome the graphs, the buttons on the left will present the different cryptocurrencies, the button on the top will provide the different modes and we find in the bottom the time axis.



With the following figure we illustrate the first mode, we see in this mode the cryptocurrencies win or loss in percentages on the selected period of time according to the brush cursor on the time axis. By hovering one button we can highlight the desired cryptocurrency.



In this last view we present the second mode, here we find the input to enter the amount of investment. We can also see on the time period given (here one month) the best cryptocurrency to invest for every time period and the amount you get at the end.



Deploy

Usability concerns :

- Find a way to nicely represent the time interval (the area on the time axis might not be precise enough)
- Find a way to indicate which cryptocurrencies present on the left side were not existing in an early period.

Aesthetics :

We want to keep the accent toward the graph. The bars will be the most colorful part of the visualization. This means that the background and the side list of all cryptocurrencies must remain a light element regarding the design. That means not eye-catching colors or bold characteristics.

The time axis is the second element, we should see very distinctly which period of time we have selected.

Although plain text is not the best element of aesthetics, we feel like it is necessary to start the page with an explanation of how one should use the visualisation. It must remain explanatory and concise.

Method to evaluate :

We will present the visualisation to someone matching our target audience and ask him to go over it. The strategy will be « Show and don't tell », this means that we provide the product but no further explanations. We will observe the behavior of our test user, listen to his comments and examine whether the goals of the project are met. This should easily reveal if we made a visualisation in a logical way.

Deviation from initial project

First of all, we changed our idea of having one screen and a switch button to go from one mode to the other. We preferred using as much space as possible on one screen for the visualisation and then have a main screen that lets you choose between the two modes. This provides us with the possibility to integrate easily the link toward the code, this document and the screencast on our website.

About mode 1, named « Classic Mode », the implementation is very close to the schemes present in this document. Minor changes would be that the y-axis with the percentages of win or loss is not fixed for every time interval. Indeed the percentages have a very diverse range of possible values, it would be unfeasible to keep the same axis. For example, for a given interval you could have a range of percentages between $[-300\%, 600\%]$ and for some other time interval a range between $[0, 60000\%]$. This is why we change the y-axis once we change the time interval. This also implies moving the x-axis to fit the 0 value on the y-axis, hence it is not fixed in the middle of the screen. We also added some additional information such as the exact start and end date of the selected time interval and the list of cryptocurrencies created since the beginning year of the interval. This is important so that the user can set the date to a more precise moment and understand why a cryptocurrency would not appear even if it is selected.

About mode 2, named « Invest Mode », we did quite deviate from our original idea. We fixed the duration to split the time interval to one month to prevent having too many data to display (e.g. : if we select a duration of one day and the interval over 2 years). A month seemed appropriate since when selecting the entire axis we can still interpret the data. We also decided to delete the y-axis and display the amount along with the name of the best cryptocurrency. This is also for readability when there are many bars. In the same sense we display only the abbreviation of the cryptocurrency, it was then necessary to display the list of cryptocurrencies with their full name next to the abbreviation. The major change that we made is that instead of making the cumulative feature, we simply display the amount one would own if one made all the investments and transactions as shown on the graph.

Implementation

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

On the main screen, you will find a description of the different elements available concerning the project. The main functionality is the selection of between the two different mode provided by our visualisation, the access to the code, to the process book or the screencast. To access any of these, the user must simply click on the button of his choice.

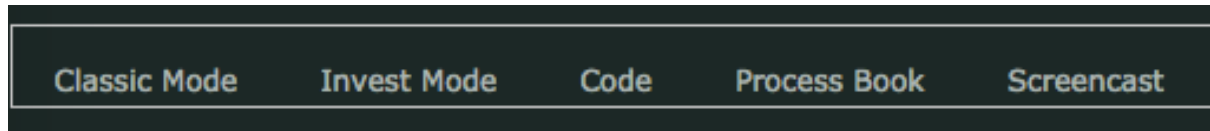


Fig 1 : Selection tool present on main screen

Concerning the Classic Mode, the user will find a description to guide him through the possibilities offered to him. The first move is to select the cryptocurrencies we want to compare. This is done with the buttons on the right.



Figure 2 : Extract of the list of available cryptocurrencies, with Bitcoin selected

The next move is to select the time interval we want to base the comparison on. This is done with the time axis present below the graph area. We can also have the precise interval below the axis.



Figure 3 : Extract of the time axis with the cursor selecting the time interval.

Time interval Apr 11 2017 to Oct 21 2017

Figure 4 : Example of precise interval display

This will let the user have the customer view of the graph. To help him make the comparison of his choice, we implemented a hover property on the list of cryptocurrencies, simply by putting the mouse on a cryptocurrency name only this once will appear on the graph.

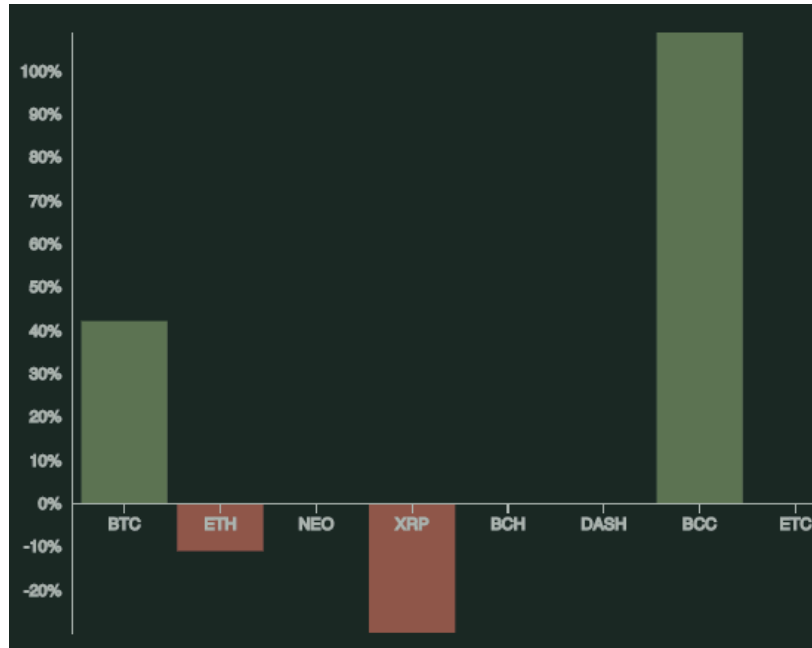


Figure 5 : Example of graph

Finally, to explicit the fact that all the cryptocurrencies were not available at all time, we added a text giving the list of cryptocurrencies created since the beginning year of the selected time interval.

From the year 2017, the following currencies appeared BCH,NMR,MIOTA,START,WAVES

Figure 6 : List of unavailable cryptocurrencies for time interval Feb 23 2017 to Jun 12 2017

Concerning Invest Mode, in this configuration, we find again the time-axis, the precise time interval, the list of unavailable cryptocurrencies and the list of all cryptocurrencies with their abbreviation (although this time there is no interaction with the latter one).

The first interaction available for the user is the investment amount we will simulate for every month. The default value is 100\$ and the user is free to change it.

Investment amount

Figure 7 : Choice of monthly investment amount

Below the former item, the user will find the cumulative amount he would have if he investment the given amount every month in the best cryptocurrency. This is in fact the ideal case, if the user chose the best match every month in the selected time interval.

You would have won a total amount
of 1462\$

Figure 8 : Display of optimal sequence of investments

Finally, we chose to delete the y-axis of the graph and to help the user understand the graph we made the decision to display both the abbreviation of the best cryptocurrency for each month along with the amount that represent the virtual win.

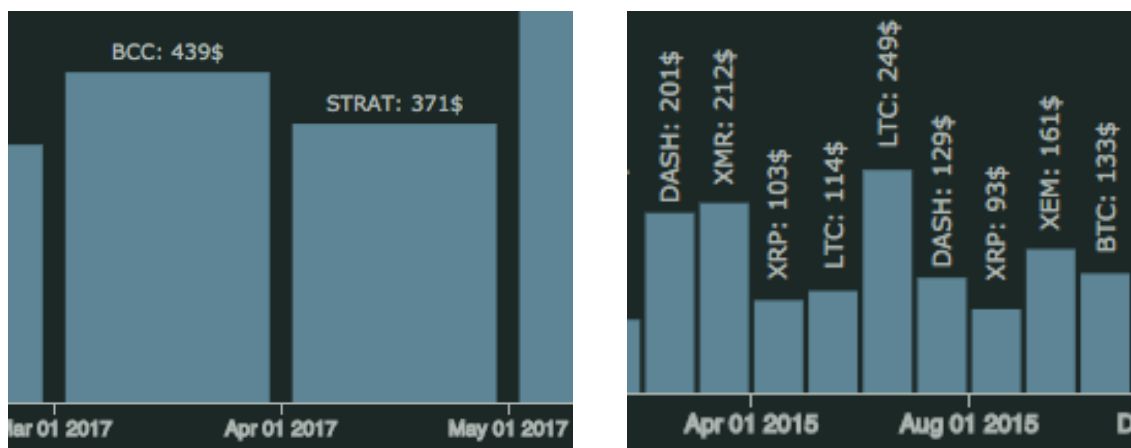


Figure 9 : Extracts of a graph showing the name and value of best transaction for a given month, with a few months or many months

Evaluation

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

While doing this visualisation, we discovered that the percentages of win or loss change extremely quickly, have no simple pattern and can have very diverse values.

We got intuition about this huge scale of rises with the coinmarketcap.com visualisation as we saw the evolution increase quickly even in logarithmic scale although the percentages going up to , for example, +14000% for New (XEM) between Jun 12 2016 and Aug 12 2017 was more striking to us.

We also see very well in the invest mode that this cryptocurrency world is very volatile, from one month to the other the best shots are rarely identical.

In a general sense, after playing around a little bit with our visualisation, we can really appreciate how emergent this area is, we see many cryptocurrencies appearing over time, how wins in Classic Mode are way more frequent than losses and how big the amounts of cumulative earning in Invest Mode can get. In this sense we would say that we achieved our goal to give an impression of cryptocurrency world to beginners.

The more precise goal to provide a way to easily know the outcome of a past investment is explicit with the use of bar graph paired with colors and percentages.

We are pretty happy about how well our visualization works. Indeed both modes have transitions as we intended, the brush area is functional (we even implemented some changes of axis in Invest mode to have an optimal reading). In Classic mode, the color code, selection of cryptocurrencies and hovering are efficient. In Invest mode the user input is well considered by the code.

We could improve it further by many points. We list here some of our ideas : giving the creation dates of every cryptocurrency when we hover the related button, display the start and end date of the time interval below the selection area instead of in text, provide a nicer way to switch from one mode to another, elaborate Invest Mode by including a profile parameter representing the risk-level the user desires for his investments and compute the best moves according to this constraint, try predicting in some way the future evolution so that people can base their investments on this work instead of only exploring past data.

Appendix

1. Design Sheets
 - 1.1. Understand Sheet
 - 1.2. Ideate Sheet
 - 1.3. Make Sheet
 - 1.4. Deploy Sheet
2. Testing

Understand

○○○○

goal: gather, observe, and research available information to find the needs of the user

artifacts: design requirements

1) identify the challenge & users

think big! what is the **problem**? who is affected by it? what is known/unknown? orient yourself with all of the project's who, what, why, when, & how.

Cryptocurrencies charts are hard to understand w/o financial background.
Difficult for people who are beginners to finance.

3) check with users or explore data

users: what did you find out? what sparked curiosity?
data: characterize aspects of the data. what is it like?

I knew what I would have earned

5) compare and rank design requirements

choose a method for comparison: **pros/cons table**, **rank** based on your findings/user needs/tasks, **cross out** the list based on listed justifications, or **pick top 3** to keep and why. explain and review with a group or partner.

- 1 - Show timeline (make evident that it is in the past)
- 2 - Easily measure +/- values (represent win or loss of money)
- 3 - Differentiate Cryptocurrencies (people rarely invest in everything)
- 4 - Select different periods of time

!! is this the right challenge to tackle? is there enough detail? or too much? too many or not enough requirements? complete this worksheet again to refocus the project.

Ideate

○○○○

goal: generate good concepts and ideas for supporting some of the project's design requirements

artifacts: ideas & sketches

1) select a design requirement

how might we address the challenge using the requirement? which questions would a user ask? revisit this worksheet for each important design requirement.
How show the win/loss value?
How select the cryptocurrency?
How select time period?

!! revisit this worksheet for all important design requirements for your project

3) sketch another idea

try another sketch, think of a new perspective, be different, do not build off of your previous sketch.



5) compare and relate your ideas

for each sketch, break apart what works well (+) and what doesn't (-) in the table below. make connections, reflect on best parts. can you **combine ideas**? review the table with a partner or group.

sketch #1	sketch #2	sketch #3
+ win/loss is obvious + length of bars gives the magnitude - manually input period of time - manually input currency	+ comparison b/w the currencies + selection of currency is easy - No choice of time. - classical view	+ Selection of time easy + Selection of currency easy + Comparison easy + win/loss is obvious - only one period of time selected on one screen

!! combining ideas and sketches is not easy. sometimes it may open up new possibilities and ideas - guess what, ideate again!

Make

○○○○

goal: concretize ideas into tangible prototypes which are approximations of a product in some aspects

artifacts: prototypes

1) set an achievable goal

what should the prototype **achieve**? what are the specific criteria for success? break a larger goal into parts with clearer features.

Communicate without financial knowledge the potential profit/loss of investment in cryptocurrencies.

!! break a goal apart into multiple and create a worksheet for each sub-goal

3) plan support for interactions

what can the user **do**? what is required given the chosen encodings? justify your design decisions.

User can select his time period and make it shift on axis.
He can select the desired currencies.
He can hover on currency to make it more visible.

5) build the prototype and check-in

are your **goals met** by the prototype? test with users if possible, are design decisions properly justified? do any need to be revisited? were any new constraints or limitations discovered? write down your progress and additional justifications below; review this progress and the prototype with a partner or your group.

Goals are met.
Limitations: retrieved data of limited number of cryptocurrencies and cannot predict in future evolution

!! did the prototype meet its goals? measure its success, make sure you have addressed the design requirement, does the prototype try to do too much?

Deploy

○○○○

goal: bring a prototype into effective action in order to support real world users' work & goals

artifacts: visualization system

1) pinpoint a target audience

who are you deploying to? what are their **goals**? what will qualify this deployment as a success?

Beginners interested in investments in cryptocurrency.

!! does this audience match your users back on the Understand sheet? if not, revisit previous sheets!

3) improve points of integration

integrate data/tools, maximize algorithmic or storage efficiency, how does this fit in a user's workflow?

- Speed is ok (only 3 years of data)
- Hard to input new raw files, code suited for one dataset but quite classic format

5) consider a method to evaluate your system

take a look at the provided supplement of possible methods, how would you test your system? what would be a successful test of this system? write an evaluation plan here, talk through this plan with a partner or your group. if you have time: test with one or more users, summarize your findings, insights, and recommendations below.

Test by "Show and don't tell".
Maybe ask user their thoughts & few questions
- What do you remember the most?
- What do you think you would use the most?
- What was a surprise?
- What didn't you understand?

!! did any of the usability, integration, or aesthetic changes result in new ideas or requirements? revisit earlier worksheets as needed!