



A comparison of user interactions in lab and crowdsourced studies

Gruber Julian
a1207718

March 16, 2017

Abstract

This thesis describes a Lab and Crowded sourced Study analysis tool capable for a various amount of interfaces. Realizing the tool with a visualization based on D3js, algorithms in JavaScript and PHP and HTML as frontend language with support from JQuery and Bootstrap.

Contents

1	Introduction	2
2	Preparations	2
3	Study design	2
3.1	Demographics	2
3.2	Understanding	2
3.3	Interfaces	2
3.3.1	No Sa	3
3.3.2	Global sa sl both	3
3.4	Short Summary Test	4
3.5	Investment Knowledge Test	4
3.6	Personality survey	4
4	Pilot Study	4
5	Study	4
6	Analysis Tool	4
6.1	Overview Graph	5
6.2	Detail Graph	5
6.3	Comparison Graph	5
6.4	Risk-Return Graph	6
6.5	Filters and Details	7
6.6	User Interactions	7
7	Results	7
8	Conclusion	9

1 Introduction

This bachelor thesis is about an investment study in a lab and crowd way. The lab study as well as the crowd sources study are based on a HTML template with JavaScript functionality. The visualizations are coded with the D3js library [Zhu13]. The purpose of the survey is to find a correlation between taken investments and the interface which was used and correlating those interface correlations against each other. In order to analyze the correlations in a visual and numerical way, a evaluation tool for those two classifications was designed. For designing the evaluation tool the 5 design sheet methodology Was used in a not that large way [RHR16].

2 Preparations

The original study was designed to run on Amazon AWS, as Amazon AWS runs on Linux the study was built in node.js. For a lab study the original study has to be transformed into a **node.js** application because of the node.js, with all the dependencies the study has, has to run on every machine the study is used to run on. Therefore node has an inbuilt option to compile all JavaScript to a main.js file with all the functionality. For the reason that participants are not able to do the survey twice, which is automatically provided via the original study with an generated key but it doesn't work within the compiled version as the same key is still in use. The key would only be changed as if the script starts. For that reason a login has been created. The login window is a simple form that triggers within every refresh of the page as well as the initial state form the page. The participants get a username and a key for login. The login will check those credentials and mark the user as used in the database. For the database a MySQL database engine is used. The compiled study as well as the MySQL engine run on my private webserver.

The data transmission to Amazon AWS also had to be changed. A single JavaScript object had to be created with all the data collected within the study. A Ajax command triggers a PHP script for saving the data as a .JSON file on the webserver. The .JSON files named as pilot or mainstudy as well as the current timestamp.

3 Study design

The study was build up by several steps which once be finished it cant be undone. All different steps has to be done at once. The first one, the user is confronted with, is the demographics page. A short form for age, gender, computer experience and investment experience. Next is the basic understanding for the following step. The next section is the important one. The users had to set up an optimal portfolio by a given interface, 5 times. After each step the total amount of money will rise as the portfolio is worth. on the next page the users recognition is checked with three simple questions regarding the step before. After that a investment quiz will be displayed. Last but not least a personality survey and after that you may want to give some feedback.

3.1 Demographics

This section of the study is used to categorize the participants in age, informatics expertise and investment expertise. The data is used to get a sense of comparison between investment and computer beginners/ intermediates/ experts and a high return value and setup age groups for checking the different ages to return values.

3.2 Understanding

In this understanding step, the participant has to pick the best portfolio out of four. Each portfolio has a different risk return setting. The interface has disabled sliders for each stock and a risk return graph is just like in figure 1 which can be seen in subsection "no sa". To choose on of the four stocks, a radio button is used.

3.3 Interfaces

The interface is divided into two parts. One for displaying the current stock settings of each of the five stocks, the total amount of money left to spent on and the other for displaying the current

risk-return ratio. Overall there are 2 Interfaces but for **one test** only one interface is used. For every step, that's five, the participant has to spend all the money. there are no restriction for setting the stock in every way possible. A bookmark button, displayed beneath the interface marks the current stocks settings. this is used to compare different settings. The bookmark will not affect the next portfolio. This means it the current portfolio cant be transfered to the next step.

3.3.1 No Sa

In this interface the stock selection is represented with five sliders for easy investing in each stock. The total amount spent on each stock is displayed on the right hand side of the sliders. The risk return graph shows the current risk return settings and will be displayed as a dot. The very simple design can be seen in [figure 1](#).

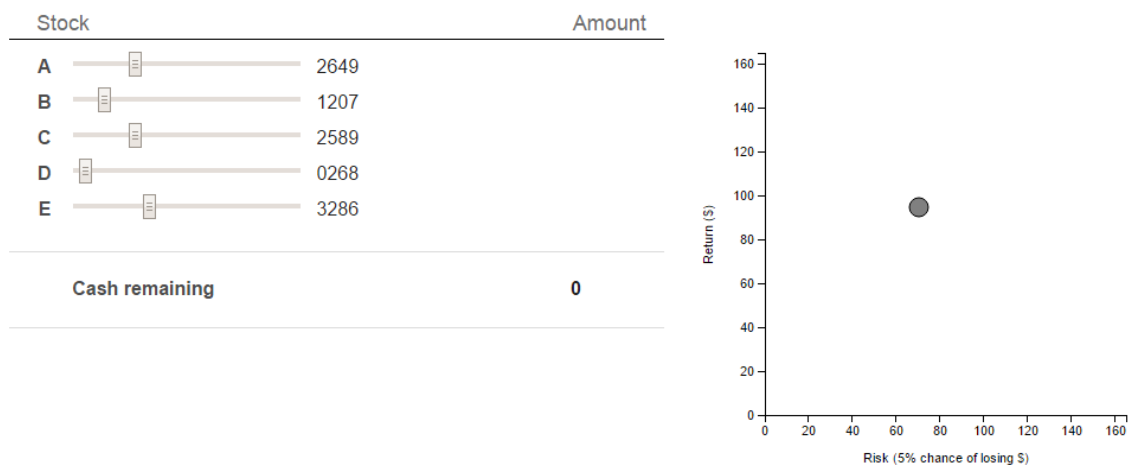


Figure 1: No Sa interface

3.3.2 Global sa sl both

In this particular interface the sliders for adjusting the stock settings were divided into two sliders per stock. The first one is used for displaying the particular return of the stock as a line chart and the other one is used for displaying the particular risk of the stock as line chart. The risk-return layout on the right is the same as the layout from the "no sa" interface. For adjusting the amount per stock either the dot on the Risk graph or the dot on the return graph can be used. The advanced design from global sa sl both can be seen in [figure 2](#).

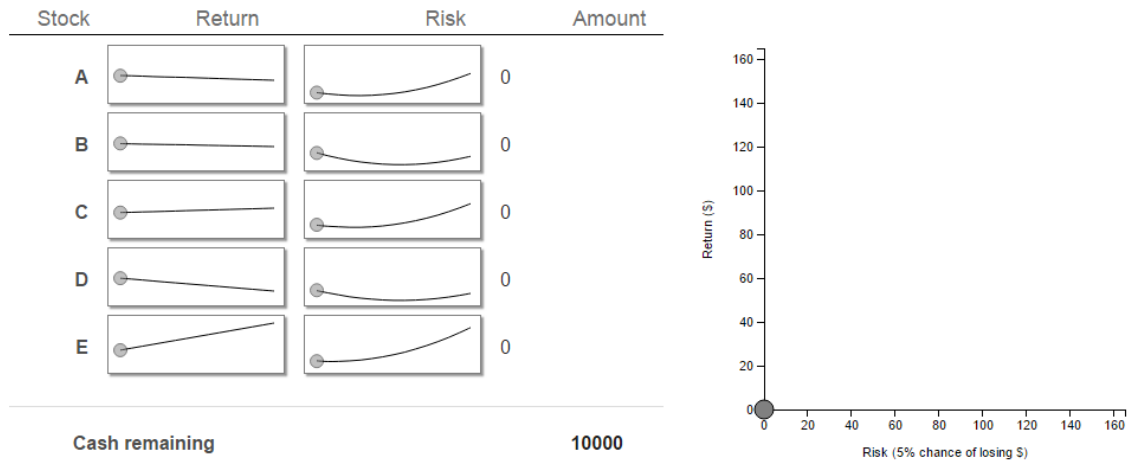


Figure 2: Global Sa Sl Both interface

3.4 Short Summary Test

The short summary test is designed to check for comprehension from the Study done before. Three simple questions will be asked which can be answered if the participant has read the instructions clearly. For not giving the right answers, the participant will fail the study and the survey will be rejected. The first two questions had to be right, the last one fails the test if the answer will be "No strategy" or "other location". The others don't fail the test but it would skip the investment test if not upper left or middle top is selected.

3.5 Investment Knowledge Test

In this section the participant is confronted with ten questions which must be answered to proceed to the next stage. Within every question there is a not sure / don't know possibility. The questions are in some general questions which can be answered without being an investment expert and some questions which can only be answered if the participant had at least a medium knowledge of investment. Every question must be answered in this section to proceed the study.

3.6 Personality survey

18 Questions with an option array from strong disagree to strong agree about the behavior and personal settings of the participants. More than half of the questions were about the mental perceptions of the participants. The others are used to get information about their thinking about work and solving problems and tasks. Again all questions had to be answered to finish the study.

4 Pilot Study

The pilot study was designed to simulate the original study, but only with a few participants, for checking the working methods and clearness of the study. Ten people were asked to join this pilot study by randomness between an age of 20 to 30 and divided into two groups. Each group will perform on a different interface. The rest of the survey **are** the same for both groups. Within the pilot study several errors occurred and had been fixed for the main study. The most of them were typing errors as well as some programmatic changes due to portability.

The tests were held at Währingerstraße 29, Faculty of Informatics, second floor open study place. The overall conditions and the all-around climate were at **comparable conditions. Mid full study place, not quiet, just say normal noise level, nothing special for distraction.** These comparable conditions were necessary for getting a clean lab study.



Out of the ten people eight people had satisfied results on the study. The others dropped out because of not reading the instructions clearly and failed the short summary test, **the other one got rejected, because he just used one slider in under one minute overall.** To get serious results the one described before got canceled for the evaluation.

5 Study

Most of the study where held at the dancing school Zehender in Baden. This location was chosen because there are comparable conditions throughout all of the participants. A room with a good noise reduction behind the wardrobe with a desk for working seems to be a perfect place for the study. All the other surveys were held at private places with comparable conditions to the surveys held in the dancing school. The age group chosen reaches from 16 to mid 40 to get a random sample throughout all possible ages who are willing to invest. The timeslot for each participant were between 16:00 and 20:00.

6 Analysis Tool

The analysis tool is a web-based application with JavaScript, PHP and HTML as base languages. For saving the results a JSON File per user is created. The tool is divided in several parts. A

overview Graph with the totals of the five investment steps displayed as line chart, a detail graph for displaying stock distributions with options as line or bar chart. A compare chart for comparing certain datasets and a Risk-Return chart as Scatter plot. A filter section and detail section are displayed as separation purpose.

To load the Data into the tool a Ajax command triggers a PHP script which itself reads through all the .JSON files in a specified folder on the webserver. Only those .JSON files were accepted which data isn't null. This mechanism is used to filtered data out which can cause the application to stop working.

6.1 Overview Graph

The purpose of the overview graph is to get a sense of the data. It can show the total amount of money after each investment month. The x-axis is an ordinal axis with the 5 months step. on the y-axis the total amount of money is show between 10000 and 11000. The first thing shown is the steepness of each portfolio. see the best or the worst or even an equable portfolios. On mouse over the current portfolio information will be displayed in den Details section. This will be discussed in chapter 4.5. on mouse click the stock per investment step are shown in the details graph. more in chapter 4.2 .the Current detail information is saved in the Details for direct comparison.

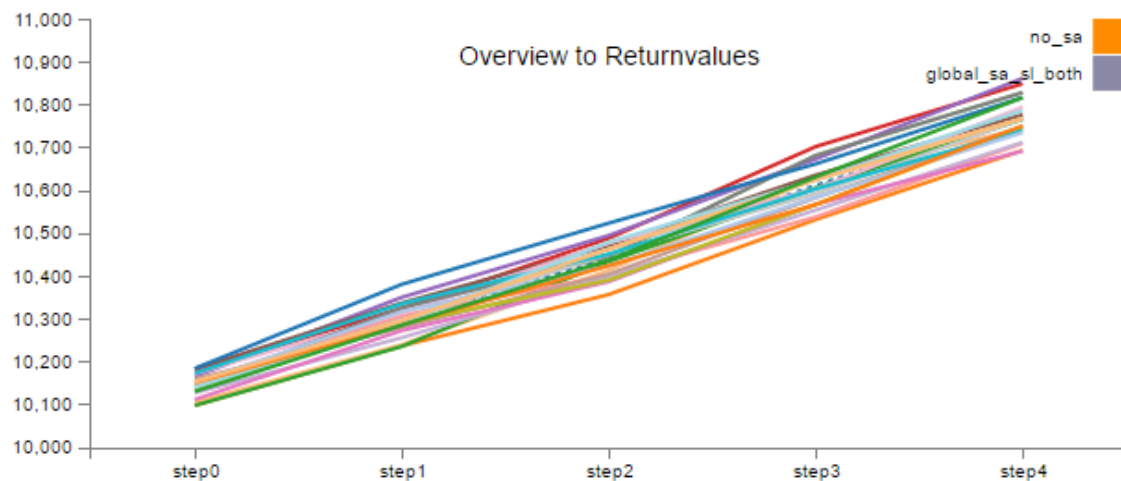


Figure 3: Overview Graph

6.2 Detail Graph

the Detail graph displayed either as line chart or as bar chart shows the distribution of each stock per investment step. The design as line chart is used for getting a sense of how the users set the amount per stock. Either in a random way, meaning he's not following the investment distribution from the step before or slightly adjusting the sliders in order to get the best return value. This is used to compare results of each type of users and check if randomness leads to more return. Within the bar chart it is easier to see die distribution of each step. Both line as well as bar chart can be seen in figure 4 .

6.3 Comparison Graph

The purpose of the comparison graph is to compare a variety of datasets with each other. A checkbox is used to set a dataset for comparison. A possible way is to compare the best male with the best female according to their maximum returns of each investment step. Another thing reasonable to compare is the best returns form each interface. The interface looks the same as in overview graph - figure 3.

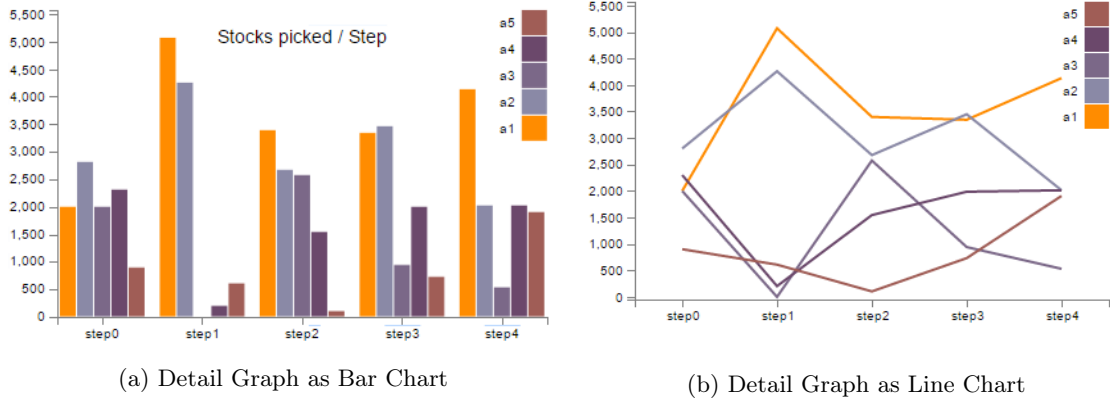


Figure 4: Detail Graph

6.4 Risk-Return Graph

The risk return graph is used to find the portfolio with the best returns. A scatter plot is used for visualizing the positions from the risk/return value and a curve segment for representing the optimal investment line. The distance is measured for each Dataset and showed as mean value per interface. The other thing to see is the correlation of the dots. Is there a tending that one interface has a higher correlation as one other.

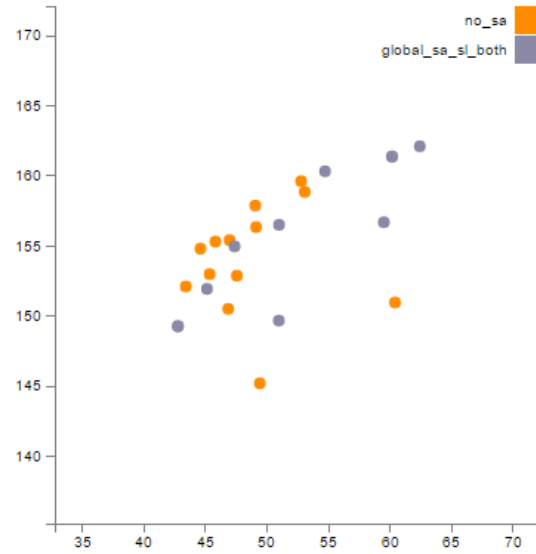


Figure 5: Risk Return Graph

6.5 Filters and Details

The filter section has two purposes. On the one hand there are, of course, filters which are clustered in gender group, an investment experience, a computer experience, an age group and a selection of each interface used in the study and on the other hand it divides the tool into a graph analysis on maximum outputs, and the Risk-Return-Curve analysis. All filters trigger on a checkbox change. As initial setup all filters are checked and therefore all data will be showed. As a checkbox is unchecked, the filter for this particular checkbox will be applied to the filtered data array. On a recheck of a checkbox the filtered data array will be set to the data which saved the whole data as a big array and all the checkboxes that are unchecked will be applied.

The detail section gives a short information on the current dataset – which is triggered by mouseover – on Computer Experience, Investment Experience and the time for completing the five investment steps. On a mouseclick the values will be permanently shown next to the mouseovered data.

After clicking on a Dataset on the overview line chart the Checkbox for comparing this particular dataset shows up. With this checkbox checked the current dataset will be sent to the compare graph. The other thing that will show up is a "Delete This Dataset" button. Which, when triggered, deletes the current dataset from the dataarray and triggers the apply filter function.

6.6 User Interactions

On all line charts a mouseover and an mouseout function is deployed. On mouseover the actual line will be highlighted and the detail information will be displayed as mentioned above. Another thing showed on mouseover is the maximum return value and the id of the dataset as a tooltip. On mouseout the initial layout will be set. A legend mouseover triggered a highlighting of all datasets according to the label on all charts except the detail graphs.



On a mouseclick, either on overview graph, compare graph or the risk return graph enabled the "to compare" and the "Delete This Dataset" functionality. Furthermore the mouseclick saves the detail information from the actual graph in the detail section next to the hovered informations as compare tool. The purpose of delete a dataset is, to get rid of outliers which distort the calculation of the correlation coefficient. This action can be undone without a reload of the page. Within the filter section the filters can be applied to the dataset. underneath the filters the switching between line and bar chart is located. this action applies to the detail graph.

7 Results

In terms of gender there are more female participants than male ones. The ratio between those are 60 to 40. Through all different interfaces the **mean** with only female view is 10779. the mean from all male participants is located at 10769. On selection of every Computer experience and low investment experience in comparison to low and middle investment experience there the mean return goes slightly back.

Filters

(click here to see Full)

Filter Gender

☒ Female ☒ Male

Filter Investment experience L/M/H

☒ Low ☒ Medium ☒ High

Filter Computer experience L/M/H

☒ Low ☒ Medium ☒ High

Filter Age

☒ 0 < 20 ☒ 21 < 30 ☒ 31+

Filter Dataset

☒ Mainstudy no_sa

☒ Prestudy no_sa

☒ Prestudy global_sa_sl_both

☒ Mainstudy global_sa_sl_both

Select between Line / Barchart

☐ Line ☒ Bar

Figure 6: Filter Section

The following graphic shows the distribution of female and male, the mean values and the important correlation which is computed without an outliers detection

Interface	correl	mean Age	mean Return	mean time	M/F[%]
no_sa	0.076	21.62	10766.75	7.68	46.2 / 53.8
global_sa_sl_both	0.911	28.5	10780.03	8.48	33.3 / 66.7
Global		24.92	10773.12	8.07	40 / 60

Figure 7: statistic overview

In terms of data correlation there is a lightly increase after getting rid of outliers. Two datasets from the "no sa" study were marked as outliers. In figure 5 the orange dot in the bottom mid and the dot slightly right above are outliers. With the outliers still in the calculation the correlation coefficient is only 7.6 percent as shown in the figure 7. Without these outliers the correlation coefficient reached 79.6 percent. The dataset of "global sa sl both" had no influence of outliers and therefor the correlation coefficient reached 88 percent. **The difference between those two interfaces are 9 percent.**

The Correlation Coefficient is calculated as follows:

$$\frac{n \sum xy - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2] * [n \sum y^2 - (\sum y)^2]}}$$

For checking the Correlation coefficient for if it is significant the t test will be applied. The T-test is calculated as follows:

$$r * \sqrt{\frac{n-2}{1-r^2}}$$

r is the correlation coefficient and n is the number of datasets.

The test size for 0.88 is 5.570 and for 0.791 is 4.324. the corresponding value in t-table for n-2 degrees of freedom with a significant **niveau** of 5 percent is 1.895. With both values greater than t-value we could speak from significant correlations.

The analysis from age groups in comparison to time for completing all five stage showed that the age group including participants from an age of 21 to 30 took them 5 minutes less the the to other age groups. The time in comparison to the maximal return the age group 21 to 30 has on average the lowest returns. we could analyze that amount of the return is in correlation with time they spent.

The time spent with an low investment experience is slightly higher than the time spent with medium investment experience. low: 8m 30s med: 6m 36s

As we analyze the different age groups with an eye on the mean return value per interface, there is a trend at the interface "global sa sl both". the mean values rise from 10748 from participants younger than 21 to 10780 from between 21 and 30 year olds to 10799 from the last age group, 31 plus. Within the interface "no sa" the returns starts at 10781 decrease to 10733 and increase again to 10766. If we compare those values to the mean values for each group, we can see that the "global sa sl both" interface is not that easy to handle for young people but with a experience a slightly higher investment time the "global sa sl both" interface is the one to prefer.

interface	0 - 20	21 - 30	31 +
no sa	10783	10733	10766
global sa sl both	10748	10768	10799
mean	10773	10753	10791

In the risk return graph, seen in figure 5, the interface "no sa", represented by the orange dots, shows that the most of the participants used a portfolio with not high risk setting. in comparison to that, the interface "global sa sl both", represented by the gray dots, are more spread within the ranges of the portfolio. A reason for that could be the interface which is settings are divided into a risk curve and a return curve and the participants are only controlling the return slider.

8 Conclusion

As we have a look at the mean values per age group as well as the mean values per experience the "global sa sl both" interface has a slightly higher returns as the other interface. in terms of the correlation we can say that "global sa sl both" had the higher correlation but as both correlation are significant we cant say that the "global sa sl both" interface is completely better than the "no sa" interface. According to the amount of participants the study has, to make a definite decision there are way to less participants. Crowd sourcing the hole study may have helped to get more participants but condition cant be controlled.

List of Figures

1	No Sa interface	3
2	Global Sa Sl Both interface	3
3	Overview Graph	5
4	Detail Graph	6
5	Risk Return Graph	6
6	Filter Section	7
7	statistic overview	8

References

- [RHR16] Jonathan C Roberts, Chris Headleand, and Panagiotis D Ritsos. Sketching designs using the five design-sheet methodology. *IEEE transactions on visualization and computer graphics*, 22(1):419–428, 2016.
- [Zhu13] Nick Qi Zhu. *Data visualization with D3.js cookbook*. Packt Publishing Ltd, 2013.

