

DATA VISUALIZATION - KIRELL BENZI

PROCESS BOOK

ROBIN CLERC, KRISTINA SATARA, LARA WIETSCHORKE

Contents

Finding the Perfect Idea 9

Creating a Concept 15

Final Concept 19

Implementation 21

Evaluation 25

Further Work 27

Conclusion 29

Peer reviews 31

List of Figures

- 1 Mockup of the map, when entering the visualization. 10
- 2 Mockup of the map, when a player was selected. 10
- 3 Mockup of the visualization, when the information section of the player was extended. 11
- 4 The first idea of possible visualization. The elements are noted around the circle, connected by the lines if they appear together. 13
- 5 Final Concept: Paintings 19
- 6 Final Concept: Elements 20
- 7 Final Concept: Colors 20
- 8 Inspiration for concept map visualization - taken from website <http://www.findtheconversation.com/concept-map/> 21
- 9 First implementation of the concept map 22
- 10 Painting details - When clicked on one of the images in the center of the circle, right from the painting the details are shown. 22
- 11 Element gallery - When one of the elements is selected, all images where this elements appears are shown on the right side of the window in the Gallery section 23

List of Tables

Finding the Perfect Idea

THE FIRST STEP we took was brainstorming. After the announcement of the project we collected several ideas. Finally we choose the idea of football player transfers, which we will explain in detail in the following section.

Football Player Transfers

Inspired by famous games like the FIFA series and Football Manager, we want to visualize the transfers of professional football players all over the world, but also visualize data about the players and clubs themselves. For these games there was already a lot of data collected, but it was only used for the gameplay. Unfortunately it's not possible to explore the data within these games.

Our motivation is to make this data available to the many people all over the world which are not only interested in football game statistics, but also in an overview of the transfer flow of players. The teams are always in change and there are a lot of transfers made in this business. It's hard to keep every transfer in mind to have an overview. And the people who were observing this business for some years usually develop the feeling, that young players come to Europe to play in the famous leagues while they leave to Asia and Northern America when the players become older. But is this really true? If so, what are the possible reasons? Are the performances of players in leagues from different continents really that different as it seems to us? And do clubs maybe have a typical transfer strategy? If we match the transfer data with additional information of the player, can we maybe discover more about the transfer? Does the money used to buy a new player always depend on his last performances? Or are there also other factors? There are a lot of questions and hopefully our visualization will bring us a bit closer to an explanation. The target audience is on the one hand young people interested in football players and clubs. For them we will create a search function to make exactly the data accessible they are looking for. On the other hand we want to give people who are new to that topic the possibility to discover it with predefined filters.

In fact we wanted to visualize the transfers with curves on a world map, using different colors to encode the age. By using sev-

eral filters we could e.g. also use the color to encode the transfer sum. We wanted to give the user the possibility to search for a player or a club and by this only visualize the transfers of this player (the color encodes the chronology of his transfers) or the club.

By clicking on the name and picture/logo, we would extend an overlay with additional information like a brief curriculum vitae of the player and a brief information part about his current club. Also we could show a list of players, with whom the selected player has played in another club before.

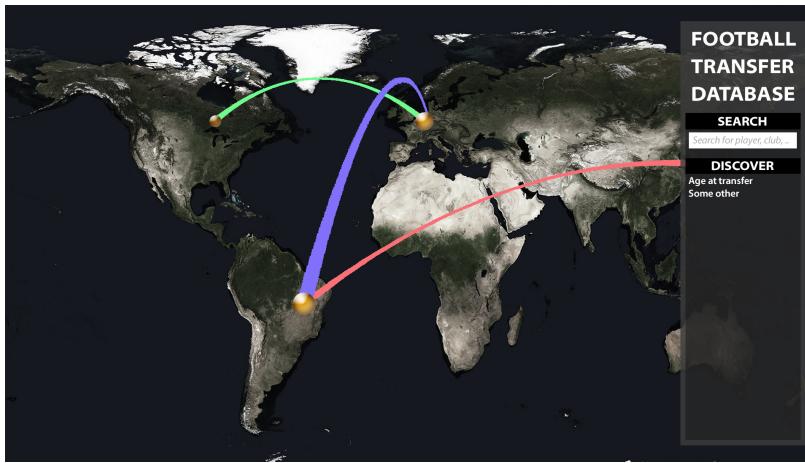


Figure 1: Mockup of the map, when entering the visualization.

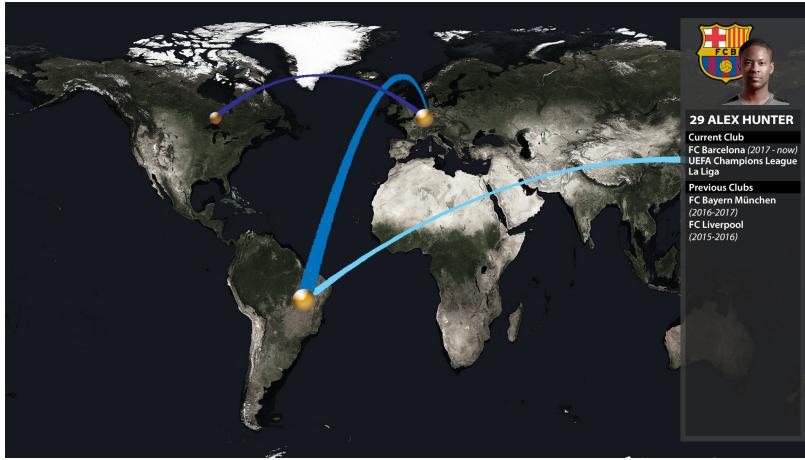


Figure 2: Mockup of the map, when a player was selected.

Unfortunately we had to cancel this idea due to several problems:

- We could not find a database containing all the transfer data. We would have had to gather this information by crawling through the code of several websites which contain the information.
- Moreover the website we could crawl does not contain the transfer sum.
- For the players we wanted to use the FIFA database. It also contains links to pictures of the players and the logos of the clubs.

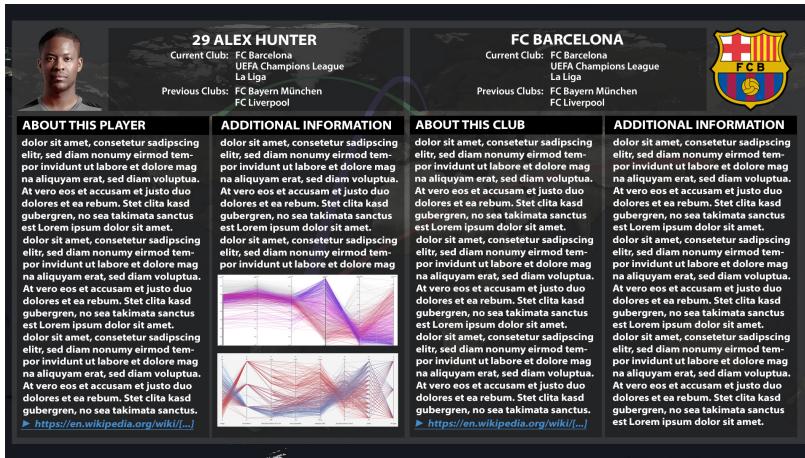


Figure 3: Mockup of the visualization, when the information section of the player was extended.

But the resolution of these pictures was too low, so we would have had to replace them by higher resolution pictures.

- Also the encoding of the FIFA database threw up a problem with special characters in the player's names. In addition the names of the players are not normalized, so there were entries like «Cristiano Ronaldo» and «L. Messi». This makes it harder to join this database with other databases.
- We need a database with the GPS coordinates of the stadiums of the football clubs. The database we found did only contain some countries of Europe, but there were still many important clubs missing.



«We don't make mistakes. We just have happy accidents.» – *Bob Ross*

Bob Ross

Statistics is everywhere, even in the art! What we want to show is that the paintings can be observed in statistical manner. Inspired by instructional television program *The Joy of the Painting* created and hosted by Bob Ross, we want to visualize the data about his paintings created during this 11-year television show. During 403 episodes Bob Ross painted 381 works. Each of this painting consists of distinct set of elements - «happy trees», «almighty mountains», «fluffy clouds» how he used to call them. The data set that we plan to use comes from fivethirtyeight's website¹ and is created by exploring each episode and painting in particular. In total there are 67 keywords that describe the content of the image (trees, water, mountains, etc.). We want to use these keywords (tags) to make conclusions not just about this particular television program, but also about instructional painting. In order to explain some of the basic art concepts, what did he draw? When observing his paintings on the first sight, we can conclude that he painted mountains and water really often. But how often is it in numbers? Did the nature he was presenting change during the years (seasons), or maybe during the months (episodes)? Can we make some assumptions – that if he draws mountains, how likely is he going to draw a house or trees? We think that the results could give us some interesting insights and conclusions about these paintings.

The target audience are his fans all over the world as well as people which are interested in art in general. Also people who might want to decorate their apartment with some drawings might be interested in exploring Bob Ross' paintings by selecting the elements they would like (or don't want) to see in the painting.

In Figure 4² you can see our first idea for a visualization with the data set described above.

MOREOVER we discovered the website [TwoInchBrush](#) which contains a database of all links to pictures of Bob's paintings, the link to each episode on YouTube as well as a matrix of the used colours per painting and paintings made by guests. We contacted the owner of the website, Felix Auer, and he shared his database³ with us. Thank you, Felix!

¹

²

³

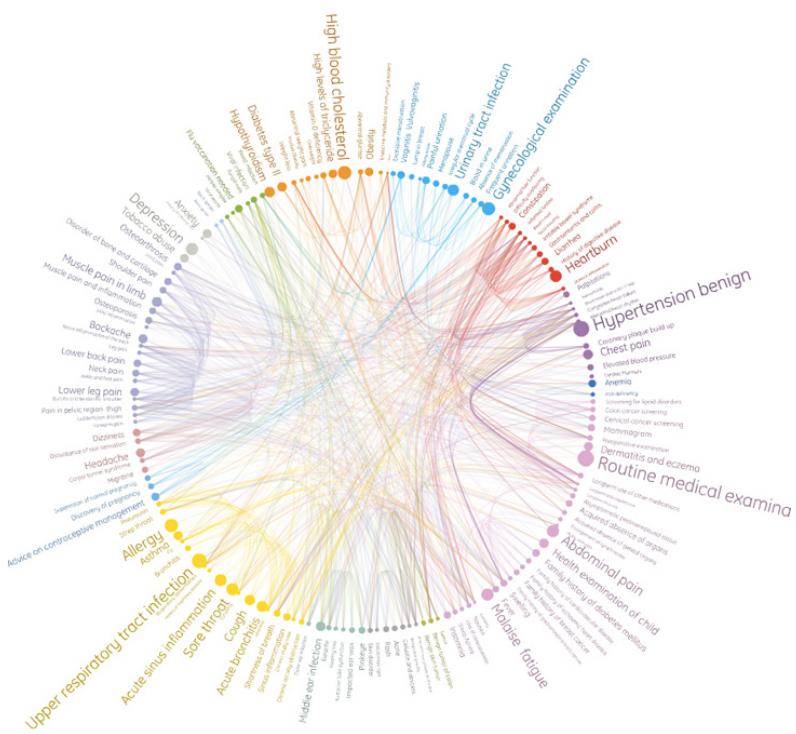
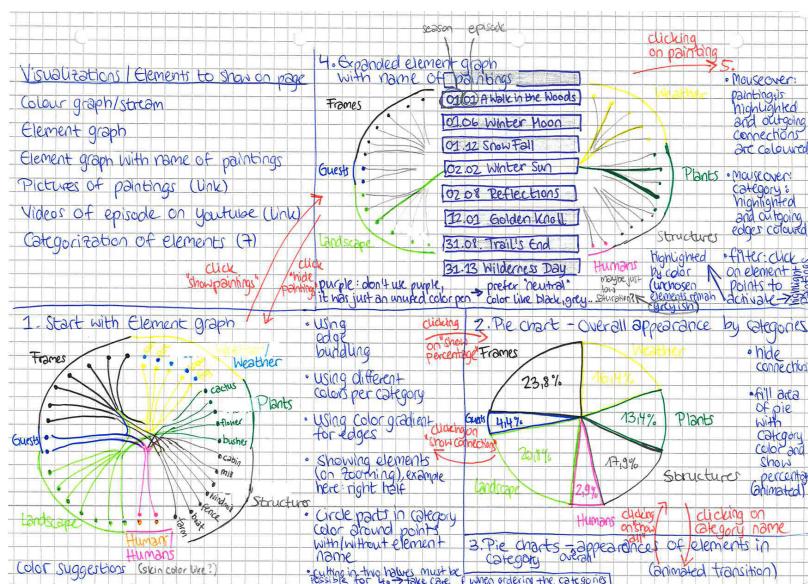


Figure 4: The first idea of possible visualization. The elements are noted around the circle, connected by the lines if they appear together.

Creating a Concept

First Concept

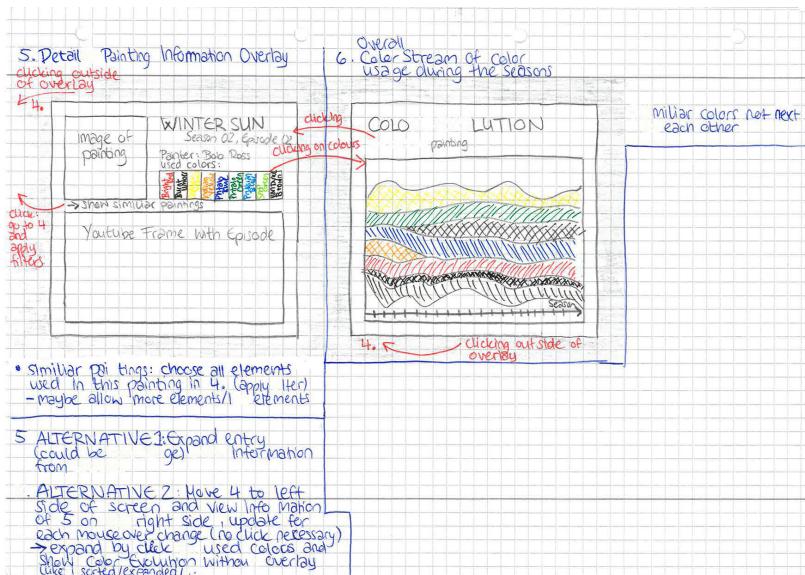
We created a draft of the concept of our visualization. Starting with collecting all the ideas we had up to this point we created an overview of the designs of the several graphs and transitions between them.



THE BASIC ELEMENT GRAPH is shown in 1 in 8. We plan this to be our start visualization, so the first visualization that is shown to the user when he enters the website. In this graph we show all the elements and their connections to each other. That means, if there appear «fog» and «boat» together in a painting, there will be a connection. If these two elements appear several times together in the paintings, the connection will be stronger. We want to use a color gradient on the edges.

In order to make edge bundling possible, we categorized the different elements into seven categories. For each category we chose a color (and sometimes also an alternative color) that is mostly associated with the category.

We want to draw the elements as dots in a circle and show the name of the element next to it's dot. Next to the element's names



there is a circle line part in the category's color and the name of the category.

PIE CHARTS are also part of our concept. They will show the overall appearance. In 8 they are labeled with 2. The first pie chart is shown by hiding the edges from the element graph and filling the pies with the category color and the percentage. By clicking on a category the category's pie extend to the whole circle and adds the element pies. The element pies should have a similar color to the category color, but they should all have different colors. This is referred to in 8 as 3.

THE EXPANDED ELEMENT GRAPH is shown in 4 in 8. Here we basically use the element graph of 1 and add the paintings in boxed in the middle of the splitted circle. The edges are no longer between the elements, but refer now to the appearance in the paintings. For increasing the readability of the graph, we transfer all colors to greyscale and show only the colors of the highlighted elements. Highlighted are elements by mouseover. If the mouse is over a painting, it will be highlighted for example by setting the background of its box to white (instead of grey for the non-highlighted elements) and its edges will be shown in the colors of the corresponding category. The same highlighting principle will be used for a mouseover on the categories respectively on the elements.

We can apply filters by clicking on elements (one click: add element to filter, second click: remove element from filter) and multiple elements can be chosen. By this, they will be highlighted as well as the paintings to which the filters apply.

DETAILED PAINTING INFORMATIONS can be visualized as shown in 5 in 9. This painting information section can be reached by click-

ing on a painting in the expanded element graph. For this visualization we drafted several alternatives which we have to try out in our implementation. The first idea is with an overlay (which is drawn in 9). But the disadvantage of the overlay it that it does not fit in the other concepts we used. So we looked for alternatives. The second idea is to expand the painting's box in the expanded element graph and show the information in this box. But this could easily be too big for a nice visualization and destroy the expanded element graph, since it would grow very high. The third alternative is to move the expanded element graph to the left side and show the detailed painting information on the right side. This section would be updated with every change of the mouseover on a painting and by this, the user does not need to click on the painting anymore.

AN OVERALL COLOR STREAM of the color usage during the seasons is also a nice way to visualize how Bob's preferences for colors might change over 31 seasons. We show how often every color was used in a season and create a graph out of this information. This section can be either shown by clicking on the used colors section in the detailed painting information (for the overlay version, as drawn in 9) or in case of the third idea for the detailed painting information also on the right side. In order to represent the colors as exact as possible we created another database with the hexadecimal values of the used colours in the paintings. We took all hexcodes from ⁴ with the exception of «Black Gesso»⁵, «Burnt Umber»⁶, «Indian Red»⁷, «Liquid Black»⁸ and «Liquid Clear»⁹.

Critic and Problems

THE BASIC ELEMENT GRAPH had to much data in the first implementation. The time it took to react was not acceptable. So we decided that it is better to show the connections between the categories instead of the connections between the elements.

PIE CHARTS are hard to read as we learned in the lecture. So we exchanged the pie chart idea by two bar diagrams. The first one shows as well the overall values of the categories we created. The second shows the elements of the category that was chosen in the first bar chart.

WE DECIDED THAT TELLING A STORY will be better than trying to link all diagrams together. So created a whole new overall concept in which we will embed the graphs. Hereby we decided to show the detailed painting information on the right part of the website, next to the expanded element graph, which we call now concept map (since it is an own visualization now).

⁴

⁵ colour value was chosen by us

⁶

⁷

⁸ value was set by us

⁹ value was set by us

MOREOVER, THE DATABASE needed to be cleaned in the way of how the elements, paintings and colors were saved (as capslock string). We wanted to be able to take directly this data, so we changed it's style according to the rules for titles. Also we realized that there are some elements that can we delete or merge with other elements.

Final Concept

THE FINAL CONCEPT divides the visualizations we have in their three main topics: paintings, elements and colors. As seen in figure 5 the connections of the elements with the painting will have the color of their category. By this it will be easier to understand the visualization. We also added a season brush. Without the season brush we would have to many elements to show and the visualization would clutter. The season brush will have a fixed maximum size. On the right side we see the detailed information for the painting that is selected on the graph.

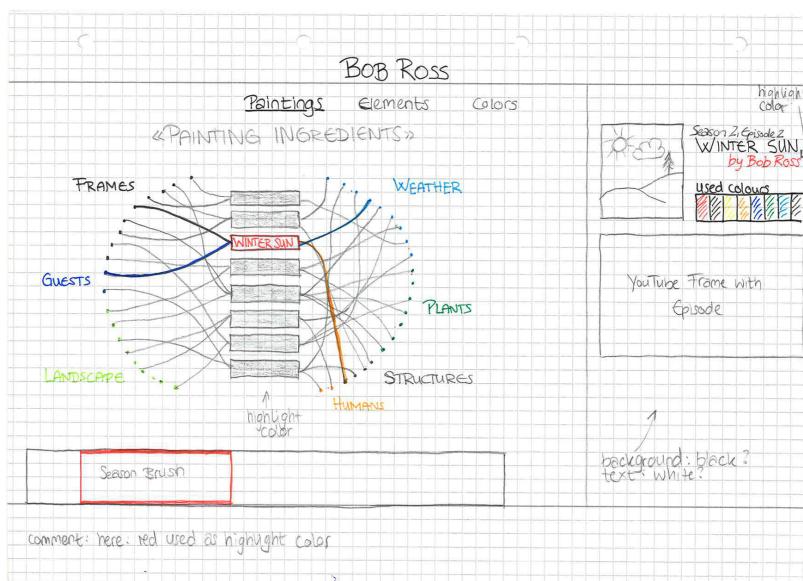


Figure 5: Final Concept: Paintings

In figure 6 we want to show the connections between the categories of the elements. Since it is too hard to draw we represented it only by a circle in this draft. Here we also have the bar charts that show the total appearances of categories and their elements. We also have the season brush as in the draft before. This is an element that connects all the different parts of the websites. The season brush in this part applies as well to the circle as to the bar charts.

In figure 7 we see the color stream as described before. We added an axis to show the seasons and highlight the color that was selected. The season brush will apply also to this visualization.

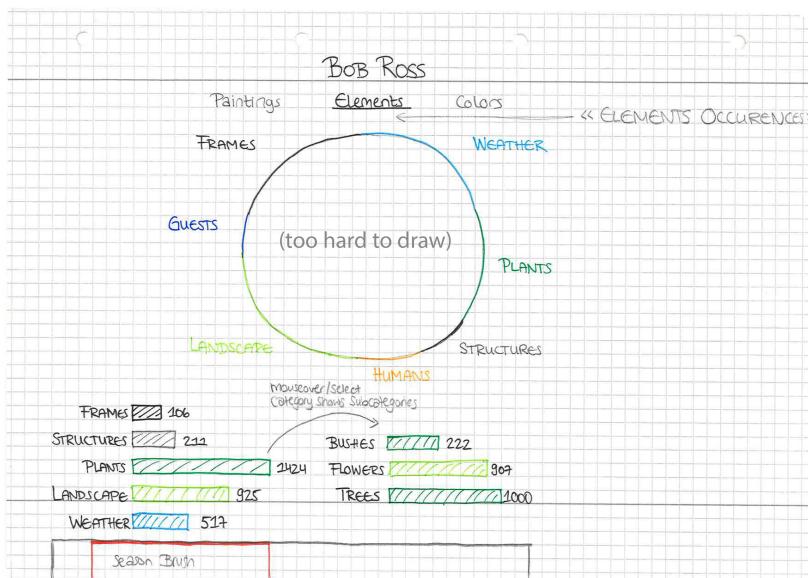


Figure 6: Final Concept: Elements

THE FINAL CONCEPT

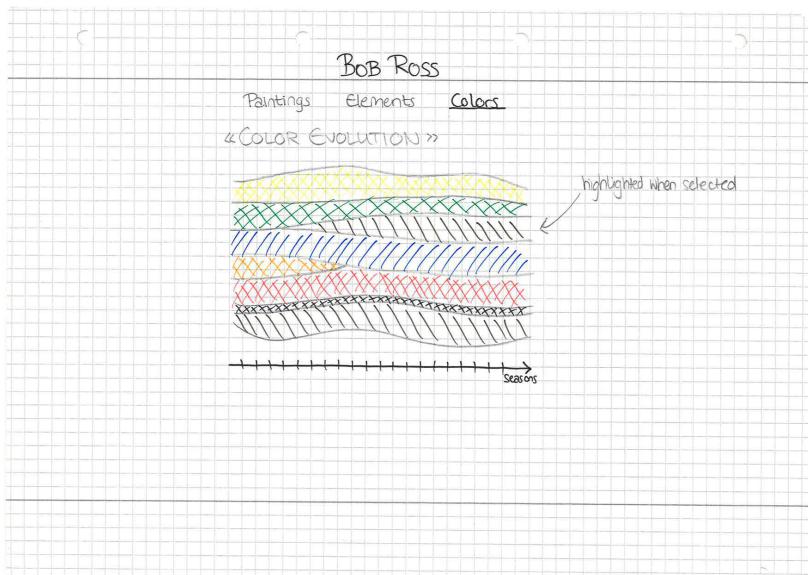


Figure 7: Final Concept: Colors

Implementation

The Concept Map

WHILE THINKING ABOUT THE RIGHT WAY to present the elements in a graph, on the website - "The Conversation" website - we found the inspiration - concept map shown in Figure 8. We decided to use this idea to present elements drawn in each episode. First implementation is shown in Figure 9 and it presents elements drawn in each episode during the first season of the TV show. Further we want to show the elements for all seasons - in total there are 31 seasons. As this probably will not fit into single graph, we will enable filtering by seasons, initially showing first few seasons. We also plan to provide filtering by different elements (for example trees, cottages, etc) or similar groups of elements (mountain landscape, winter landscape, etc).

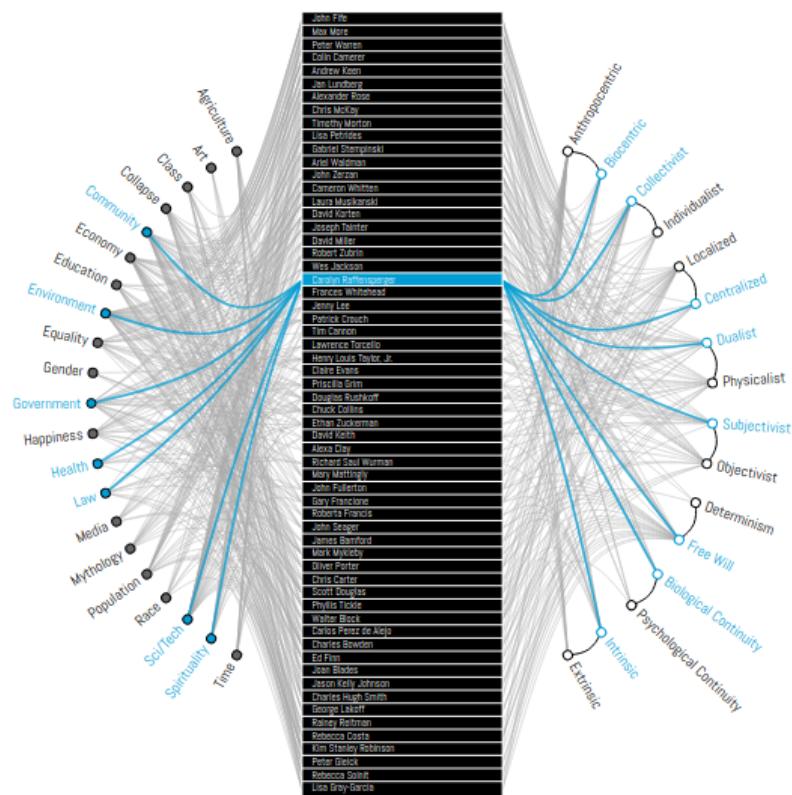


Figure 8: Inspiration for concept map visualization - taken from website
<http://www.findtheconversation.com/concept-map/>

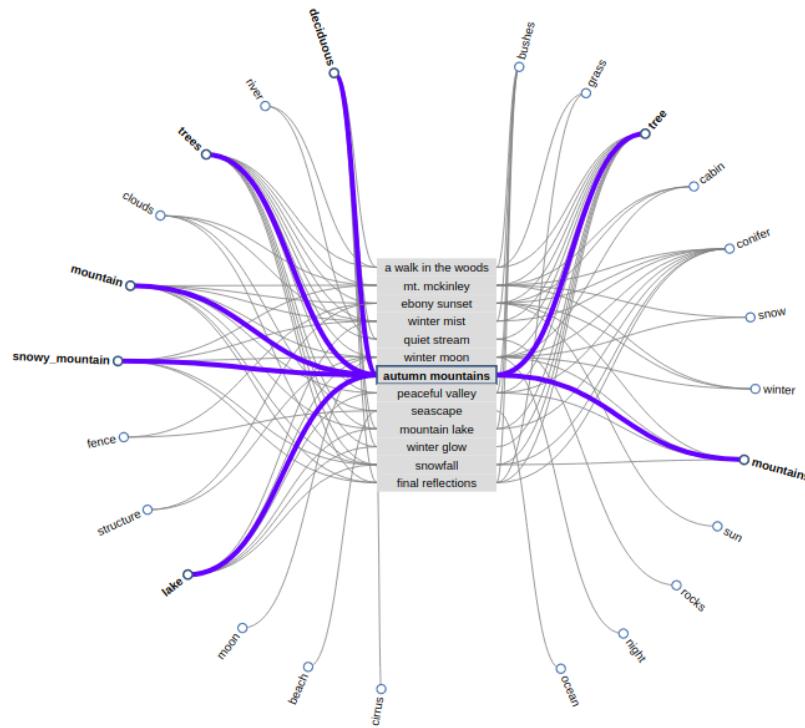


Figure 9: First implementation of the concept map

The first implementation showed in the Figure 9 is using the mocked data, the next step was to add the real dataset. This resulted with graph that we were not able to understand, as there were too many links between the nodes. Because of this we decided to use the brush and to bound the maximum number of the seasons to two. In this way we achieve that the concept map is always useful. In Figure 10 this can be observed.

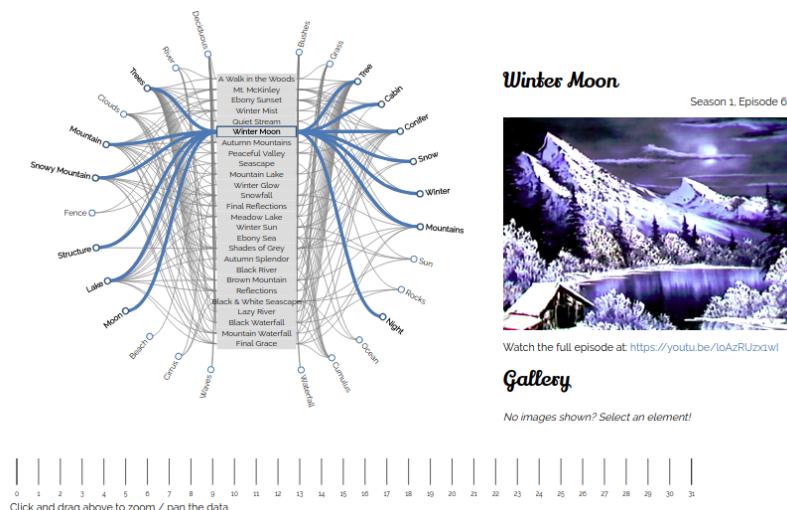


Figure 10: Painting details - When clicked on one of the images in the center of the circle, right from the painting the details are shown.

BESIDES OF THE BRUSH, in order to make our chart easier to use and also to give more information to the users, we decided to add

details about each painting. When selecting one of the paintings places in the center of the circle, right from the chart details are appearing, as shown in Figure 10. First of all, the links to the elements, which are appearing on this painting, are bolded in steel blue. Painting details, such as season and episode numbers are shown on the right side of the window.

ELEMENT GALLERY is shown also on the right side of the window. One example can be observed on Figure 11, where element Cabin is selected. This element appears on four images that are shown on the right side of the window after selection.

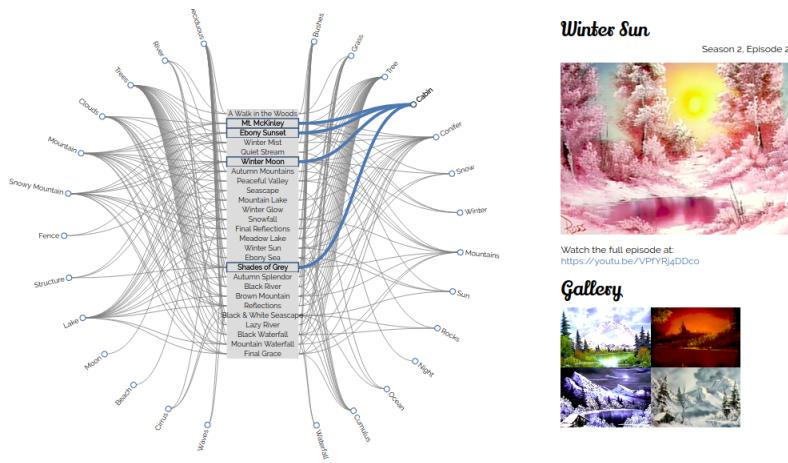
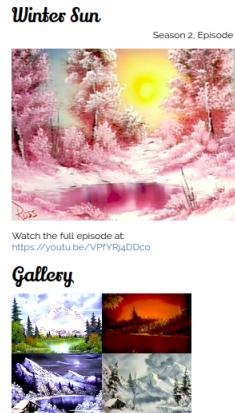


Figure 11: Element gallery - When one of the elements is selected, all images where this elements appears are shown on the right side of the window in the Gallery section



Bar Charts

FOR CREATING A DATASTRUCTURE, we first read the data about the appearances of elements in paintings and the data about categories and their corresponding elements from the csv files. In fact, the data has the form of a matrix with a 1, if the element appears in a painting, else the entry is 0. Since some of the columns were redundant, we had to clean the data first. In fact, we deleted the columns «frames» (since we can compute this column from all the other elements that belong to the category «frames») and «lakes» since it was zero. After reading the file, we use a filter to extract only the data of the seasons that were chosen with the brush. We construct an array with the length of the total amount of all possible elements. Each entry has another array of length 3. The first entry is for category, the second for the element and the third for the total amount of the element. By iterating through the data from the csv file, we increase the total amount for every element, if it appears in a painting. In the next we use the data from the other csv file, that tells us to which category the elements belong. The corresponding category is added to the data structure. Thereafter we construct two more arrays from this datastructure. The first array

contains the summed up values for each category. The second array contains an array for each category. This array contains all elements of this category and their total amount for the selected part of the dataset.

THE OVERALL (CATEGORY) BAR CHART shows the summed up appearances of elements in the seasons that are chosen with the brush. These are grouped by category. Moreover, we determine an array with the categories and the colors the categories will use as well a color for not highlighted content. Generally speaking, we create the overall (category) bar chart as we did it in our exercise lesson, but adding a brush a some additional behavior for mouseover, click and mouseout actions. When doing a mouseover over a category, this category will keep it's color while all other categories will get the color for not highlighted content. The detailed (element) bar chart for this category will be shown. If a category is clicked the mouseover is deactivated and the detailed (element) bar chart is kept when moving the mouse. When another category is clicked, the detailed (element) is updated to the clicked category. When the already chosen category is clicked again, the mouseover is activated again and the detailed (element) graph updates to the category on which the mouse points. When a mouseout is performed the chart is transformed to default again (with category colors for each category). Important is, that the mouseout is also deactivated when a category was clicked. This allows even the usage of devices with touch screens.

THE DETAILED (ELEMENT) BAR CHART shows the amount of the elements which belong to the chosen category for the chosen time-frame (via brush). In general, this chart is similiarly constructed as the other bar chart. The main difference is, that the bars height scales according to the amount of the elements, so that both diagrams always have the same height. This shall make them nice looking when displaying them next to each other. Moreover, they adapt their color to their category's color.

Evaluation

Concept Map

IDEA OF USING THIS CONCEPT MAP is found to be very interesting and useful for the users, as it gives really nice insight into the paintings and elements shown in each of the paitings. It was very difficult to show all the data in this kind of chart, and this is solved by using brush. Brush is bounded to use maximum two seasons - so that the number of paintings is not too large to fit into this chart.

SO WE CAN CONCLUDE that the downside of this chart is that it can not show too many paintings. But beside of this downside, we find that the chart gives us really useful information about the statistics of the paintings and that usage of the brush solves the problem of too many pain.

Bar Charts

The solution with the two depending bar charts could also have been replaced by a typical graph. But the bar charts have some advantages compared to a graph. First of all the total number of elements in categories as well as the total number of appearances of these elements differ a lot. For example, in the category «Humans» are only two elements whereas in the category «Landscapes» are thirteen elements. Also the total amount of appearances of elements in the «Human» category are much less than the appearances of elements from the «Landscapes» category. So if we would have used a typical graph with linear scaling and the total amount of appearances on the y-axis and the seasons on the x-axis to visualize this, the values shown on the y-axis would have such a big difference that the graph might not be readable anymore (just take a look at «Humans» with four appearances and «Plants» with 1424 appearances). Moreover, by using a graph as described above we would not be able to visualize the categories as we did it in the bar chart. With the bar chart and the brush we can explore the portion of each category in the chosen seasons, and we realize that they never change during 31 episodes: the majority of the elements belong always to the category «Plants», followed by the category «Landscapes». Furthermore, in the detailed bar chart we are able to

directly compare the elements belonging to the same category and explore for example, which kind of tree Bob prefers in the several seasons. A possible disadvantage of the bar chart are the categories. By the categorization it is not possible to directly compare two elements from different categories.

Further Work

Concept Map

FURTHER WORK IN THE CONCEPT MAP would be exploring what is the best way to utilize this map with large number of central nodes - in this case paintings. For now we solved this problem by using brush which we bounded on maximum two seasons. But this should be improved.

ANOTHER POSSIBLE IMPROVEMENT could be done when showing the element gallery. At this point it is shown on the right side of the window, but it might be good to place it under the brush, in a line as shown in Figure 5.

Bar Charts

As mentioned above we could improve the bar charts in a way that allows the user to choose elements from different categories he or she wants to compare. We could implement this by adding a third bar chart. By clicking on the element in the detailed bar charts this element will be added to the third bar chart in order to allow the user the choose the important elements by him- or herself.

Conclusion

FINDING THE RIGHT WAY TO VISUALIZE the statistics in paintings was very interesting and challenging. We explored the related work in analysis of Bob Ross's paintings and we found one related website. This was additional motivation for us to succeed and to be among the first to gain some interesting insights into the statistics of elements painted, distribution of colors during the time, and the ways that the number of different elements painted was changing during the time. All statistics can be viewed on our website, by trying out the visualization charts. Here we show some of the interesting conclusions we found for the paintings:

- Elements that were drawn the most are trees, mountains and lakes.
- Categories mostly drawn are plants and landscapes, which is confirming the previous conclusion.
- Regarding the colours, burnt amber vanishes during the end of eight season and it almost does not appear again.
- Usage of phthalo green color also disappears during the tenth season and it appears but in much lower percentage than at the beginning.
- Dark sienna appears beginning of fifth season and it is continuously used until the last season.
- Midnight black is gaining on popularity during the sixth and seventh season.

WE ARE VERY SATISFIED with what have we learned not just about the good practices in data visualization area, but also about the Bob Ross's paintings.

Peer reviews

In general we had very good repartition of the tasks and subtasks, by splitting the work in such way that each of the team members was working on different tasks. The communication has improved through the weeks and we helped each other very reactively and were also learning from each other.

Peer review by Lara Wietschorke

- **Preparation - were they prepared during team meetings?** Yes.
- **Contribution - did they contribute productively to the team discussion and work?** Yes.
- **Respect for others' ideas - did they encourage others to contribute their ideas?** Yes.
- **Flexibility - were they flexible when disagreements occurred?** Yes.

Peer review by Robin Clerc

- **Preparation - were they prepared during team meetings?** Yes.
- **Contribution - did they contribute productively to the team discussion and work?** Yes.
- **Respect for others' ideas - did they encourage others to contribute their ideas?** Yes.
- **Flexibility - were they flexible when disagreements occurred?** Yes.

Peer review by Kristina Satara

- **Preparation - were they prepared during team meetings?** Yes.
- **Contribution - did they contribute productively to the team discussion and work?** Yes.
- **Respect for others' ideas - did they encourage others to contribute their ideas?** Yes.
- **Flexibility - were they flexible when disagreements occurred?** Yes.

