

Visual Analytics project report

Høyskolen Kristiania

PGR110, Visual Analytics

Lester Lasrado & Tore Levinsen

20.05.2022

Intentionally left blank

Dataset - 1

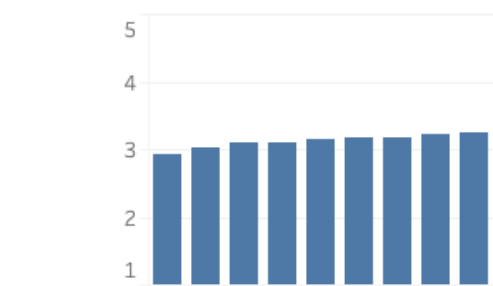
We are given the task of designing and developing a dashboard for the Kristiania Higher Management to help them evaluate how Kristiania is doing compared to other institutes on assorted quality dimensions. The dashboard will visualize Kristiania's best and worst-performing study programs in a helpful manner. With this in mind, we attempted to create the dashboard with minimalistic and aesthetic functionalities in design. The concept ought to be user-friendly for communicating with the applied data.

Dataset description

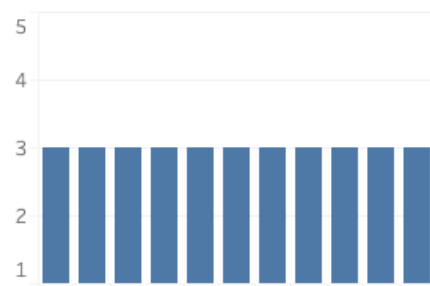
Data Type

In order to work efficiently we decided to rename all of the relevant dimensions into significant names with detailed descriptions in each comment section. Later on, we discovered it was another more intuitive way of renaming the fields by linking the source together. We have modified relevant tables to assist us in assessing the creation of our final product. In the “Survey in percentage” measure we created a calculated field with total survey divided by total enrollment, number format to percentage with 0 decimals. This data will indicate the survey rate in percent amongst all institutes. This is significant as well as how many people actually took the survey. We have to take into account that fewer answers are more likely to result in a higher averaging score. We changed the quality measurements within a scale from “1-5” to integer and aggregated them on average. We want to eliminate distractions by presenting our data in context, thus decimal was removed because we don’t need this concise information.

According to Knafllic C.N, we can eliminate data that doesn't add any values. "We don't need the level of precision or granularity that decimal points provide."(Storytelling with data, 2015, p.134) in our case we are only interested in values between 1 to 5 which are appropriate in this context.



Source: Dataset1
Datatype float(decimals)



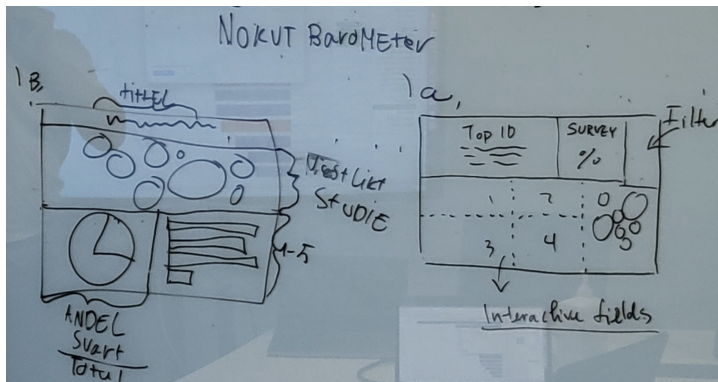
Source: Dataset1
Datatype Integer(whole)

Designing the dashboard

Sheets

The challenge itself and probably the most fun part was to obtain all the important sheets that cover the assignment task. We being the analysts are responsible to show the audience what is relevant and filter out the less relevant data, how to visualize the data and why we chose to display the elements the way they are. The initial phase was to first extract all the necessary measurements so we could iterate with sketches, essentially we had sheets for each measure.

We made the charts with the basic designing fundamentals in mind. The idea is how the audience will process the information, what is drawing their attention, and how to successfully convey the message in an easy and explanatory way.



The first sketch of the Dataset-1

Visualization, the graphical representation of information

The bar charts are very commonly used for a reason because it's easy to understand and read.

With that being said this is excellent in this case where we have to compare a lot of categories.

Referring to Knafllic, C.N, "Our eyes compare the endpoints of the bars, so it is easy to see quickly which category is the biggest, which is the smallest, and also the incremental difference between categories." (Storytelling with data, 2015, p.66). For comparing the surveys we used a horizontal bar that contains all the institutes with names, this makes it extremely easy for the western hemisphere audience that most read from left to right across the screen. Figure 2.

For the quality measures, the zero baselines in the x-axis will be at 1 and peak at 5. This is a very simple visual, but it allows for accurate comparison.

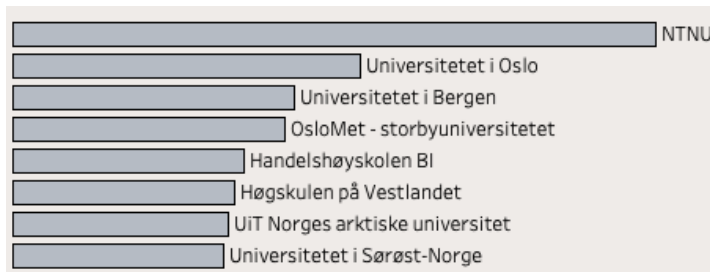


Figure 2
Source: TestBoard
Horizontal bar

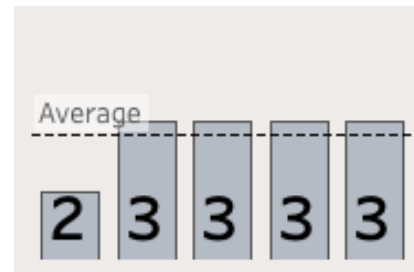


Figure 3
Source: TestBoard
Vertical bar

Color, pre-attentive attribute

The use of color is conventional and concise. The goal is to highlight the important part of our information we intentionally picked red, to be more precise, the color of Kristiania. Since we are comparing Kristiania versus the rest we want to draw attention to all data that shows anything Kristiania. Inspiration and color were taken from the Ekstern *Designmanual*. While maintaining the color aspect throughout the dashboard we reduced the color contrast to easily identify Kristiania and lead the cues on where and how Kristiania stands out in the crowd.

Clusters

To make it even easier to read we are removing all unnecessary grid lines as well as markers to make the data stand out while maintaining our principles. Data labels that are attached were removed to clean up unwanted attention, the description will be inside when speculated in detail.

Dataset - 2

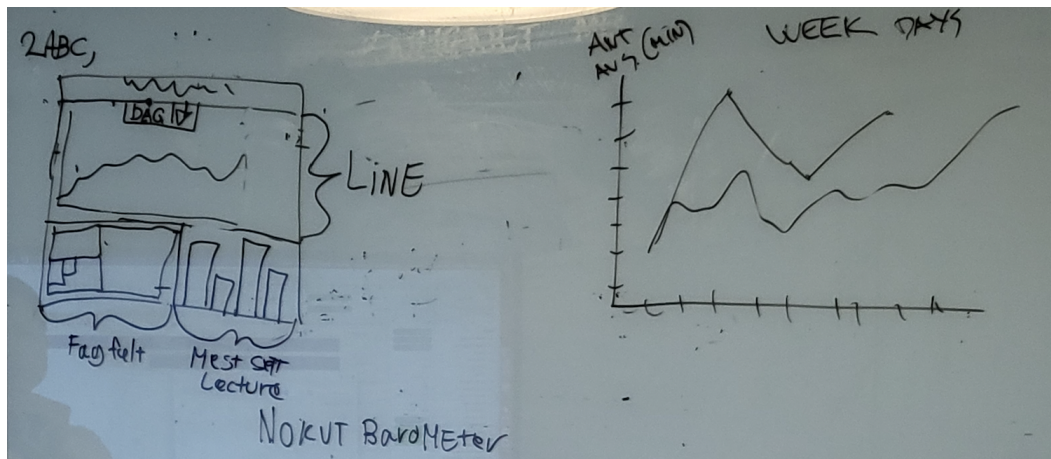
Dataset 2 is a collection of data collected from the visual analytics course at Høyskolen Kristiania/Kristiania university college. The fields contain different names and numbers ranging in different scales and types. The timestamp field contains times and dates to show at what point a student has entered the lecture/exercise for that day. The next field is the lecture/Exercise name which consists of a few titles for that particular lecture/exercise in combination with the student, and time. Session Name is the name of that particular session. Username is the student and is used to identify that student with the other fields. A study program in which program the student comes from and is studying during the course. Location is the location of that particular student in that field. Only two locations are listed and that is either Oslo or Bergen since those are the two locations the course is being taught at.

Most popular hour for exercises/lessons

One of the tasks we were to answer was: “analyze viewing patterns across the weeks, days, and time of day”. Our first thought was to solve this by using a line chart, we struggled a bit to find the perfect match of how we were to show the data and compare them. As mentioned we ended with the use of a line chart showing both Oslo and Bergen to get a general understanding of the difference in when the most popular hour for exercise/lesson was. We can clearly see a big difference between Oslo and Bergen, as shown in the dashboard you can see that the most popular time for Oslo is 9 AM while in Bergen the most popular time is 12 PM. You can shift around in the dashboard by selecting the different charts. For example, if you select April 2022 in “viewing pattern by months” you can see a big difference in when the most popular hour for exercise is, it changes from 9 AM all the way up to around 11 AM

Viewing patterns insight

We chose to display the data for viewing patterns for days and weeks with horizontal bars. Not only because horizontal bars are an easy way to spot the differences, but also because we wanted to make a good clean dashboard with data that's not going to make you bored looking at it. The most significant data we see when we open up the second dashboard is that the Thursdays stand out. When we look closer at the data, we can see that Thursdays have over 40.000 views, almost nine times as many views as Saturdays only 5000 ones. This makes perfect sense as the classes in Visual Analytics are being held on Thursdays.



The first sketch of the second dashboard

Most popular lecture/exercise

We decided on using a bar chart for the use of showing the popularity of the most popular lecture/exercise. We decided early in the sketching that we wanted a bar chart for this use, it clearly shows what the most popular lecture is and what the least popular lecture is. We also decided to show the bar chart with the popularity descending from left to right. As shown in the dashboard the data tells us that the “introduction to data visualization - part 2” is the most popular lecture while part 1 is right behind. Not too weird considering it is the introduction to the whole program and maybe the most important lecture. We can also see that “session 4 - summary

#1 (afternoon)” is the least popular lecture. This can be explained by the fact that it's the summary of the lecture so we can assume that all the students paid attention in the main lecture for session 4.

Treemap with study programs

In our first sketch we had a treemap with all of the study programs Høyskolen Kristiania offered. We decided to use this in our end product as well. The treemap is well organized and easy to navigate. When you click on the program of your choice you can see all of the different charts with a focus on that exact program, making it easy for you to see what day was the most popular, what hour was the most popular, which month was the most popular and what lecture was the most popular. When shifting through different study programs u get a good understanding of what program does well in different categories.

Pie charts

We wanted to add the different charts where they would make the most sense and where they would fit the most. Chris Zimmerman wrote in his presentation “don't use pies (..) for more than 5 values” (Chris Zimmerman, Details of Information Design, p.14). That is one of the reasons we added a pie chart to Oslo and Bergen because it was one of the few places with under five values. He also said that if it was only two values, it is better to just write it, but we think it's an easy way to navigate between the different schools. In the same lesson, Zimmerman also wrote that it's best to have the biggest value from the top, so the audience could read the pie chart clockwise, then counterclockwise.

References

Knafllic, C.N. (2015). *Storytelling with data: A data visualization guide for business professionals*, (2015), pp. 66 - 134.

Design Manual. *Ekstern designmanual V.02*. <https://designmanual.kristiania.no/#Farger>

Chris Zimmerman (Lecture & Exercises 4th March 2022), *Details of Information Design*, pp. 14