UPPSALA
UNIVERSITET

UPTEC STS 19023
Examensarbete 30 hp
June 2019

The bubble funnel

A visualisation concept designed to increase understanding of user funnels

Max Karpefors

UPPSALA
UNIVERSITET

Teknisk- naturvetenskaplig fakultet
UTH-enheten

Besöksadress:
Ångströmlaboratoriet
Lägerhyddsvägen 1
Hus 4, Plan 0

Postadress:
Box 536
751 21 Uppsala

Telefon:
018 – 471 30 03

Telefax:
018 – 471 30 00

Hemsida:
http://www.teknat.uu.se/student

Abstract

The bubble funnel

Max Karpefors

To understand user acquisition is of major importance for digital service providers. Drawing on a case study at Hedvig AB, this thesis presents how a user funnel can be visualised in, to my knowledge, a novel way to increase the understanding of the inflow of users. The design is human-centred and research methods include, among others, semi-structured interviews and evaluation sessions. The visualisation is developed for the web with JavaScript and D3. In contrast to previous solutions, this visualisation takes advantage of motion and combines the state of the art representation of a user funnel with a bubble chart. This approach is a step away from the static visualisation and a step towards a more engaging
and interactive solution to communicate and analyse data. Two main ideas of usage are presented: one for simulation of historical data and one for live data. Mainly, this visualisation provides a clear overview of a whole user funnel, it addresses user cohorts in different ways and it efficiently shows where users drop-off.

Handledare: John Ardelius
Ämnesgranskare: Mats Lind
Examinator: Elísabet Andrésdóttir
ISSN: 1650-8319, UPTEC STS 19023

Acknowledgement

This work would not have been possible without Hedvig AB. I am grateful to the team behind Hedvig AB – especially to CTO John Ardelius – for a generous welcome, constant support and insightful feedback. Thanks also to the informants for lending me their time for interviews and evaluation sessions. In addition, a special thanks to Professor Mats Lind at Uppsala University for supervision, all discussions and for helping me navigate through the endless literature on data visualisation, human perception and usability.

Sammanfattning

Det digitala och globala samhället producerar dagligen otaliga mängder data – allt ifrån email, chattmeddelanden och filuppladdningar till transaktioner, loggar och spårning – och data beskrivs ofta som ”det nya guldet”. Mitt i detta hav av data söker företag och organisationer insikt och förståelse för att kunna förbättra sina digitala tjänster, förbli konkurrenskraftiga och locka nya kunder. Det finns numera en hel yrkeskår som kallas för ”growth hackers” vars främsta uppgift är att förvärva nya kunder och användare. För att bli framgångsrik här är förståelse av kunderna och användarna avgörande: Vilka är de? Vad vill de ha? Hur beter de sig – och varför?

Ett sätt att möta det här behovet och uppnå förståelse är genom visualisering av data. Den här uppsatsen undersöker hur inflödet av användare i en digital tjänst kan visualiseras på ett nytt sätt med målet att öka förståelsen för inflödet.

Genom en teknisk designstudie på Hedvig AB, ett Stockholmsbaserat tech- och försäkringsbolag, togs ett visualiseringskoncept fram för att möta företagets växande behov att förstå sitt inflöde av användare. Designstudien är uppdelad i fyra delar: en förstudie, en design- och utvecklingsdel, en utvärdering-av-prototyp-studie och en vidareutvecklingsdel. I förstudien undersöktes behovet av en visualisering genom semi-strukturerade intervjuer, en analys av Hedvig ABs nuvarande verktyg samt en simpel explorativ dataanalys. I design- och utvecklingsdelen togs en prototyp fram genom en inkrementell utveckling från skisser till avancerade skisser till en slutgiltig prototyp. I utvärderingsstudien testades prototypen på olika användare. I den sista delen (vidareutvecklingsdelen) utvecklades konceptet vidare genom bättre kod, åtgärdade problem samt nya funktioner. Dessa delar utkristalliserar ett användarcentrerat visualiseringskoncept: *the bubble funnel*.

The bubble funnel är ett visualiseringskoncept – byggt med D3 och JavaScript – som visar ett flöde från A till B med hjälp av klustring, distribution och rörelse hos bubblor. Varje bubbla representerar en eller flera användare. Konceptet är skalbart och kan anpassas till olika förståelsebehov. Studien visar på två användningsområden för konceptet – ett för att simulera historiska data och ett för att övervaka data i realtid. Konceptet ökar förståelsen genom att enkelt, effektivt och visuellt skapa en överblick av ett användarflöde med fokus på själva flödet, användaregenskaper och distribution av användare. Framförallt visar det var användare lämnar flödet. Till skillnad från andra analytiska verktyg erbjuder detta mindre kvantitativ analys och mer kvalitativ analys.

Förhoppningsvis kan the bubble funnel bli ett solklart komplement i verktygslådan till ”growth hackers”, företag och organisationer som vill tälja guld av sin data.

Table of contents

1. Introduction: a wicked visualisation problem

> The simple graph has brought more information to the data analyst's mind than any other device – John Tukey (Jones, 1986, p. 457).

A 'simple graph' is a visualisation. Visual information as a medium for communication has been around for centuries, perhaps for as long as humans have settled the earth. Data visualisations are mainly used to communicate data or analyse data (Evergreen and Metzner, 2013), and the term data visualisation could be defined as 'computer-supported, interactive, visual representations of abstract data to amplify cognition' (Card et al., 1999 cited in Meirelles, 2013 p.13). Cairo (2013) suggests that visualisations can be seen as 'functional art' and Samsel et al. (2018) highlight that visualisation is an effective way to communicate complex scientific data. Hence, data visualisation is a tool for communicating and understanding data – 'graphics reveal data' (Tufte, 2001, p.13). Advantageously, data visualisation could be applied in any field of research or business since it simply is an efficient data-communication tool and should therefore be paid attention to.

Globally, a boom of collected data is taking place and companies and organisations have to understand their data in order to be successful:

> Today, companies are under extreme pressure to dig deeper into and connect all information and pieces of data at their disposal to find new differentiators from their competitors. They need to better understand their market, customers, competitors, and talent pool – Isson and Hwang (2018).

In parallel, customers and users are expecting more and more from digital service providers and data analysts try to keep the same pace to explore the needs and behaviour of its users (Isson and Hwang, 2018). At the same time, a new profession called 'growth hacker' has appeared. They are specialised in business growth and customer acquisition and are taking advantage of analytical tools, commonly embedded with visualisations, to understand the market and to execute appropriate business decisions. That is one example where visualisations could create business value.

This thesis will address a narrow, yet important, business problem: to understand a user flow in the context of conversion on digital platforms. In particular, it will explore new ways of interpreting the user funnel (also known as the marketing funnel, conversion funnel, purchase funnel and goal funnel, depending on the purpose) visually. I will use the term *user funnel* since it entails that the funnel constitutes of users and it could be placed in any context.

Basically, a user funnel starts in A and ends in B with different steps in between. Throughout the funnel, some users continue all the way to B while others abandon the funnel on the steps in between creating a funnel since the quantity of users is getting smaller at each step. Noteworthy, the funnel is not visible for the user – it is a (visual) representation of user behaviour in a certain digital context used by the data analyst to monitor and investigate the performance of a funnel within a digital service. For instance, the purchase funnel (see e.g. Hoban and Bucklin, 2015) typically starts in the shopping cart and ends with a completed purchase. The goal is to analyse where and when users drop the purchase procedure.

Conclusively, visual information is a way to understand data, understanding data is a key for success in the twenty-first century, and a user funnel is a vital part of a digital service's performance. This thesis explores how a user funnel can be visually designed and developed in – to my knowledge – a new way through a case study at Hedvig AB, a growing start-up in need of understanding its inflow of customers. The work is carried out in a design-thinking fashion through incremental development (IDEO, 2019). Yet, as with all design challenges, the problem could be interpreted as a *wicked problem* since the solution to a wicked problem is not true or false, but it is good or bad (Rittel and Webber, 1973). Research motives include how complex data can be visualised in order to provide an understanding of user behaviour laying ground for data-driven decision making and data-driven UX (User Experience) design.

1.1 Research aim

The aim of this thesis is to explore how the user funnel can be visualised in a novel way to help digital service providers improve their insights and knowledge of the process it depicts. To achieve this, the following question will be addressed: how could the flow of users in a user funnel be visualised in order to increase understanding for relevant user groups?

1.2 Delimitations

This project has two major delimitations. First, it is limited to the field of data visualisations since, as Karpefors and Weatherall (2018) point out: 'visualization is the key to effective, data-driven decision making in many areas, and therefore, it is important to keep inventing new visual designs' (p.1). Moreover, data visualisation is a way to understand complex data and should, therefore, be paid attention to, especially since human perception is a powerful tool for pattern seeking and: 'perception and cognition are closely interrelated, which is why words *understanding* and *seeing* are synonymous' (Ware, 2013, p. xvi).

Second, this project is limited to Hedvig AB which is an insurtech (insurance and technology) company based in Stockholm, Sweden, since they are using deterministic chats to acquire new customers, i.e. their chats are acting as user funnels (see *3. General method* for more information about Hedvig AB). Therefore, in this project, Hedvig AB symbolises a 'digital service provider' mentioned in the research aim. Thus, Hedvig AB's desired insights and knowledge will define 'understanding for relevant user groups' also mentioned in the research aim.

1.3 Disposition

The thesis is organised in the following way. First, there is an overarching, traditional thesis structure with sections including introduction, theory, general method, analysis and discussion, and conclusion. Beyond that, the empirical work is split up into four different parts – I, II, III and IV – constituting of fieldwork, design and development, evaluation and a developing-further phase, respectively. Each part contains its own aim, methodology, result and conclusion. This structure will be explained in more detail in the section *3. General method*. But first, some relevant literature will be presented.

2. Theory

This chapter will provide relevant literature on the topics of human-centred design, interactivity, human perception, visual information, and existing solutions for visualising user funnels. All this in order to provide a slim, yet comprehensive, framework for designing a good visualisation.

2.1 Human-centred design

Data visualisation is about good and effective design (Meirelles, 2013). Mike Cooley coined the concept 'human centered systems' in contrast to 'machine based systems', suggesting that the design of a system should be designed to increase the capacity of humans using it, address them instead of the machines (Cooley, 1987). The idea of a human-centred design has developed since and is now present in many fields, such as the field of UX design where communication design and information visualisation are sub-categories (Benyon, 2019).

Don Norman (2013), an advocate for human-centred design, argues that design should be so good that the intention and usage of the design imply how a product is supposed to be used. Unfortunately, that is not always the case – neither with everyday objects such as major appliance nor with technology products and software. Norman suggests that 'human error' may not exist, only *bad design*. He states that 'good design is actually a lot harder to notice than poor design, in part because good design fits our needs so well that the design is invisible' (Norman, 2013, p. 162). Norman has plenty of ideas on how to produce human-centred products, here are three examples: One, to add *constraints,* which basically means to limit the options of how a product can be used to simplify the usage. Two, to let a product provide appropriate *feedback,* i.e. communicate the results of interactivity back to the user. Three, to use concepts that the user is familiar with, such as a button to trigger an action or a folder icon to indicate that files could be placed inside. Conclusively, a well-designed product should fit the needs and capabilities of its users. That is the fundamental idea of human-centred design (Benyon, 2019).

2.2 Design and interactivity

Don Norman (2013) puts usability first, but the aspect of beauty, aesthetics and perception also seem to matter. Tractinsky (1997) sheds light on a correlation between a user's perception of aesthetics and usability. In a follow-up article, *What is beautiful is usable,* Tractinsky et al. (2000) once again prove that aesthetics matter and that interfaces perceived beautiful seem to be easier to use. However, one should be careful to take that notation – what is beautiful is usable – for granted. Tuch et al. (2012) show that the opposite occurs under certain conditions, instead suggesting that: 'the frustration of poor usability lowers ratings on perceived aesthetics' (p. 1596). Regarding visualisations, Tufte (2001) suggests that graphics and charts could still be made aesthetics, but that should not entail messy visual displays.

Regarding interactivity and data visualisation, Shneiderman (1996) found that powerful data visualisations have something in common: the information-seeking mantra. The mantra consists ‘overview first, zoom and filter, and then details-on-demand’ (p.337). Drawing on the mantra, Shneiderman suggests information-visualisation interfaces to expand in the following task-based order allowing the user to: first, get an *overview*; then to *filter and zoom*; then *details-on-demand*; then allow *relationships* between items to be displayed; then show *history* of interactions, such as supporting ‘replay’ or ‘undo’; and lastly, to provide *extraction* of data. This task-based taxonomy will allow the user to explore data quickly in a user-controlled fashion. Yet, implementation requires specialised, clever and fast data structures, user training, high screen resolution and parallel computation (Shneiderman, 1996).

2.3 Human perception

Another important aspect of data visualisation, not always considered in design challenges, is human perception and its impact on design. Perception is an umbrella term for processes that are making sense of our senses, such as our visual interpretation of the world surrounding us. In short, our eyes are constantly sending information to our brain that is processed and interpreted, meanwhile, our brain is sending back tasks instructing the eyes what to search for. In this process, *visual queries* are acts performed by the visual thinking system in order to solve a cognitive task where a visual search is a part of the solution. Visual queries can be either simple or complex. For instance, if a person is looking at a map with the intention to go from A to B. Step one would be to locate A and B and step two would be to create a visual query to locate the nearest route. Here, a cognitive task is solved with a visual query. This is also doable since most people are experienced with maps – a map is a known type of pattern. Previous experience of solving a cognitive task helps a user to solve a similar problem (Ware, 2008).

Thus, according to Ware (2008), this implies that visual information should be designed so ‘that visual queries are processed both rapidly and correctly for every important cognitive task the display is intended to support’ (p.14). Suitable design should support visual queries of the users to help in perceiving and using a product.

Accordingly, good design should target what humans can easily see. That is things that pop out from its surroundings. Basic pop-out channels include: **colour**, such as hue, lightness or simply a contrast in luminance; elementary shape and texture, such as **size**, elongation or orientation; ← *motion* →, such as a direction or vibration; and,

spatial grouping.

These are four basic pop-outs. However, there are several ways to accomplish a pop-out effect, but what is important is that the pop-out is different from its surroundings (Ware, 2008).

2.4 Excellent visual information

Edward Tufte is known for his work on data information and visualisation, partly as a pioneer within the field. This section will present some of his ideas on excellent data visualisations and his fundamental principles of analytical design.

Excellent graphics should: display the data; make the viewer think about the meaning of the visualisation instead of the design, method or technique used; not misrepresent the data; provide much information in a limited space; provide coherent data; show comparisons; show the data on both an overview, but yet, with details on different levels; have a clear aim; be in line with descriptions (verbal and statistical) of the data set. Yet, a lack of design skills, lack of belief in users' intelligence and little interest in quantitative numbers may cause bad visualisations resulting in lying or dishonest charts. Another way to ruin a graph is to add 'chartjunk'. The term refers to unnecessary colour or elements or other stuff that complicate the reading of a visualisation (Tufte, 2001).

Tufte (2006) suggests, in *Beautiful Evidence,* that excellent visualisations, both for presentation and analysis of data, should fulfil his fundamental principles of analytical design:

Principle 1 – Comparisons
An excellent piece of visualised information should show comparisons, differences and contrasts in order to explore changes, patterns and support reasoning about the data. The comparisons should be selected and displayed appropriately.

Principle 2 – Causality, mechanism, structure and explanation
An excellent piece of visualised information should provide explanation, causality, systematic structure and mechanisms. This is to allow the data visualisation to contain causal analysis to enable decision making.

Principle 3 – Multivariate analysis
An excellent piece of visualised information should not be restricted to one or two variables, instead many variables should be displayed. That is since 'reasoning about evidence should not be stuck in 2 dimensions' (p.130), especially not since data originates from the real, complex world that surrounds us.

Principle 4 – Integration of evidence
An excellent piece of visualised information should integrate, among others, charts, diagrams, tables, numbers, words and images to that extent that it provides an explanation. These so-called pieces of evidence should be included as long as they are needed to explain what is to be explained.

Principle 5 – Documentation
An excellent piece of visualised information should have documentation that contains: good titles; documented data sources; measurements and scales, authors, affiliations and sponsors (if any); and highlighted issues.

Principle 6 – Content counts most of all
An excellent piece of visualised information should contain relevant, quality content displayed with integrity. A graph is data-driven, or content-driven – it should not be something else.

In fact, the principle of causality – principle two – may be difficult to include in a visualisation since there is so much uncertainty tied to human activities and human behaviour. Yet, according to Tufte, that makes these principles, all of them, even more relevant for studying human activities and behaviour (2006, p.138). The principles that are used should be in line with the cognitive task the data visualisation is intended to solve. If the task is to compare something – a comparison should be presented (Tufte, 2006). Conclusively, these principles constitute a toolbox for displaying and analysing data.

2.5 Existing solutions

Regarding user funnels, the common practice is to visualise them with either a dashboard software (e.g. Klipfolio, 2019) or an analytical (funnel) tool (e.g. Google Analytics, see Google, 2019, or Hotjar, see Hotjar, 2019). The state of the art is to visualise these user funnels as static charts. Either as line, bar or area charts or as funnel charts (see Appendix II for an example of a funnel chart available in Klipfolio). Hotjar, a popular web analytics company, is using a stacked area chart for their conversion funnels, showing the drop-off of users at every step and ending with a total conversion rate (Hotjar, 2019). These solutions are typically showing linear user funnels, displayed step by step, in a horizontal direction. Funnel charts, on the other hand, are commonly visualised as funnels in a downward direction. All these solutions have numbers and statistics displayed along the user funnel.

Visualisation of user flows varies to a large extent. For instance, Google Analytics provides a user flow visualisation in the shape of a sankey diagram (Google, 2019). The literature, on the other hand, provides a number of different solutions to visualise flows, time and motion. These are usually very different and highly dependent on the cognitive task they are intended to solve (see e.g. Meirelles, 2013, or Tufte, 1990).

3. General method

This chapter of the thesis explains the overarching methodology. First, Hedvig AB will be introduced and why it was suitable as a case-study company. Second, design thinking will be introduced, followed by some notes on data and ethics. Lastly, the structure of the empirical body of the thesis will be presented.

3.1 The case of Hedvig AB

This project was carried out at Hedvig AB, a Stockholm-based insurance company (see Hedvig AB, 2019, for more information). It is a start-up founded in 2017 and the company is aiming to change the insurance industry into something it is not: fun, easy and transparent. The goal is to create 'the world's most remarkable insurance experience.'

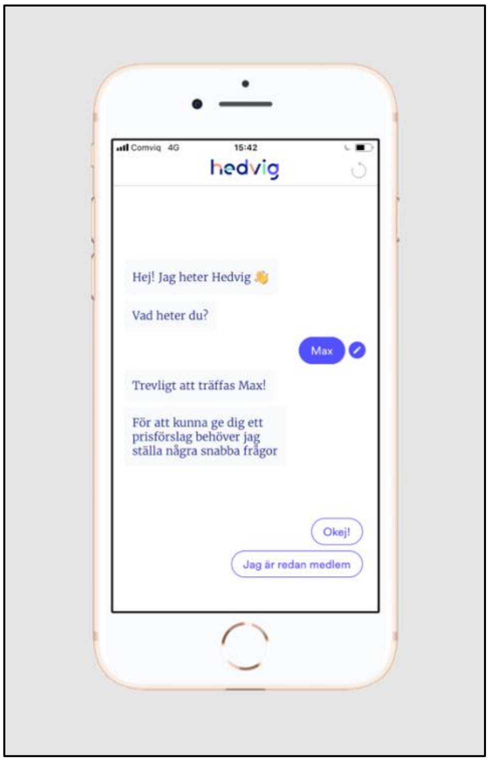


Figure 1. Hedvig AB's onboarding chat.

The service of Hedvig AB is relying on a smartphone application and customers, or members – which is the term the company is using – communicate with Hedvig AB mainly through the application. Customers download Hedvig AB's application or navigate to their website and start to chat with 'Hedvig' in order to get an insurance deal (see Figure 1). This chat is deterministic with pre-defined answers, yet some questions depend on your previous answer. Every customer or potential customer will answer questions like: what is your name? what is your social security number? do you live in an apartment or a house? how many square meters is your living space? do you currently have an insurance deal at another company? are you a student? When all questions are answered, ‘Hedvig’ will provide a customised insurance offer. This chat, from the first question until an insurance offer is signed, will hereby be addressed as the *onboarding chat*. In addition, customers and potential customers will hereby be referred to as *users*. (Noteworthy, the communication with Hedvig AB after an insurance deal is signed is also chat-based, yet this communication is not deterministic since claims and general questions are managed by Insurance Experience Specialists.)

Conclusively, Hedvig AB is providing a digital insurance service and they acquire new customers through an onboarding chat, both available on their website and in their application. These two chats act as user funnels (see 1. Introduction for definition). Therefore, Hedvig AB is suitable for this project. Especially since they are a start-up, dependent on an increasing inflow of new customers to secure growth and ease the raising of investment capital.

3.2 Data and ethics

When users go through the onboarding chat, data such as selected answers, timestamps and personal information is collected. The main data source in this visualisation project is a data view with funnel data (see Table 1).

Table 1. An example of one data point in the funnel view.

ID	Step	Timestamp	Status
673980654	First app open	2019-04-19 08:20:13.456789 + 00:00	INITIATED

The column ‘status’ refers to whether a user is a customer (‘signed’), is currently in the onboarding chat (‘initiated’), or for some reason terminated (‘terminated’). Unlike the other columns, ‘status’ is a characteristic of the user. The funnel view shows the tracking within the app; therefore, this project will exclusively target Hedvig AB’s app.

In an early stage of this project, some anonymised user data including age, gender, day of opening the Hedvig AB’s app, day of signing, previous or current insurer, living space, housing type, amount of people in household, UTM (Urchin Tracking Module) source and medium, was used while investigating the possibilities for different prototypes (see 5. Part II: The development phase – design and deliver). However, the data used was anonymised to the extent that it would have been impossible to triangulate the data to a specific individual. In addition, no pure data will be presented in this thesis, only the result in the shape of a visualisation concept.

3.3 Design thinking

To reach the aim of this project, a design-thinking approach was selected. Design thinking is basically the process of producing and designing new products with a focus on the strategy, planning and implementation. It is a creative problem-solving process (Benyon, 2019; IDEO, 2019). More specifically, in this case, the double-diamond model developed by the UK Design Council was used to plan and carry out the project. This model was selected because it has a clear structure supporting incremental development (UK Design Council, 2015). The double diamond (see Figure 2) is a model conceptualising the working process of solving a design task. It divides the work into two 'diamonds'. The first diamond aims to define the problem. It diverges to discover insights into the problem and then it converges to define a specific problem. The second diamond aims to develop and design the solution. It diverges to investigate potential solutions and then converges to one solution. This approach is suitable for this project since it is includes developing a new visualisation concept incrementally, aiming to provide new dimensions to a user funnel.

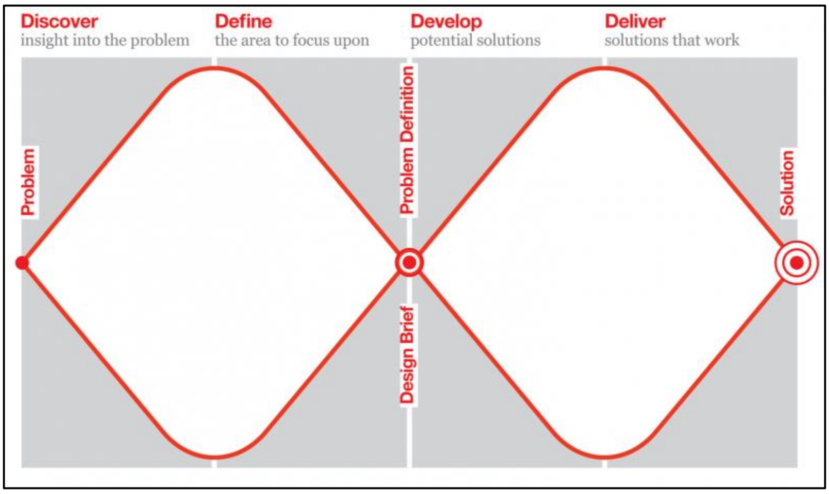


Figure 2. The double diamond (UK Design Council, 2015).

Besides the double diamond model, the working procedure was inspired by the PACT framework. Benyon (2019, p. 25) describes the PACT framework as a way to interpret a design situation. PACT refers to *people*, *activities*, *contexts* and *technologies*. The framework demands the designer to understand the users, who they are, what they do, how they work, etcetera, in their context. Thereafter, according to obtained insights develop and design a product. Lastly, the designer needs to be aware of existing technologies that are available to solve the design task. To design and develop products that are useful and user-friendly motivates this framework. Moreover, it is in line with how Don Norman sees a design challenge: 'One of my rules in consulting is simple: never solve the problem I am asked to solve. Why such a counterintuitive rule? Because, invariably, the problem I am asked to solve is not the real, fundamental, root problem. It is usually a symptom' (Norman, 2013, p. 217). Thus, the general-method idea is to go out and explore the problem and, thereafter, design and develop a visualisation accordingly.

3.4 Introduction to part I, II, III & IV

To achieve the research aim and to follow a design-thinking process, the empirical work of this project is divided into four parts in an incremental fashion. Part I symbolises diamond one and Part II, III and IV symbolise diamond two (see Figure 2). This framework is suitable since both a thesis project and the framework have a clear start and a clear end.

- *Part I: The investigative phase – discover and define* consists of a pre-study to find the needs of a visualisation of the inflow of new customers, what current tools are missing and what 'increase understanding' can be regarded as.
- *Part II: The development phase – design and deliver* consists of the development of the design and the visualisation prototype. It explains the process from qualitative data to sketches to one prototype. One prototype is selected to be tested instead of multiple due to the timeframe of this project.
- *Part III: The evaluation phase – issues revealed* consists of the evaluation of the prototype. It explains how it was carried out, results, lessons learned and issues that need to be addressed.
- *Part IV: The rethinking phase – further development* consists of further development of the prototype and how issues were addressed.

Each part's methodology section and results section will be named accordingly to simplify the mapping between methods and results. Altogether, these parts describe the process of the creation of a visualisation concept. The four parts are specifically targeting Hedvig AB. Thereafter, the visualisation will be analysed drawing on the literature on a more conceptual level. In addition, the conclusion will highlight the main results and provide ideas for future research.

4. Part I: The investigative phase – discover and define

This part of the thesis will explain the pre-study and its results. In particular, it will focus on *why, when* and *how* the pre-study was carried out and *what* it found.

4.1 Objectives

Part I aims to discover and define core intents for a visualisation of the user funnel. In particular, it aims is to understand the needs of Hedvig AB in terms of what insights and knowledge that are desired to define what 'increase understanding' is, as mentioned in the research aim. This leads to the following questions:

1) What information is desired from the onboarding chat and how could that information give Hedvig AB value?
2) What are the benefits and drawbacks from today's visualisation of the user funnel?

4.2 Methods

In order to investigate this, i.e. the needs of Hedvig AB, multiple methods have been used to increase understanding of the problem and to enhance the quality and credibility of the qualitative research (Patton, 1999). First, fieldwork including semi-structured interviews, informal conversations and simple contextual inquiry has been conducted. Second, an analysis of the current tool has been carried out. Third, an unorthodox version of exploratory data analysis was conducted.

4.2.1 Fieldwork

The fieldwork was inspired by Holtzblatt and Bayer's contextual inquiry (2015) which stresses the designer to be in the context, observe, participate and talk to people in order to gain a more comprehensive understanding of the users (the employees at Hedvig AB) and their needs. This is in line with the human-centred approach of this project. The fieldwork contains semi-structured interviews, conversations and simple contextual inquiry. All fieldwork was held in Swedish since the informants are Swedish. In addition, a framework to analyse the fieldwork is explained in this section.

Semi-structured interviews
The main fieldwork-information gathering consists of semi-structured interviews. In contrast to Holtzblatt and Bayer's contextual design interviews (2015), questions were prepared to guide the conversation since the goal – to visualise a user funnel – is quite specific. Yet, despite the prepared questions, the interviews were carried out more as conversations inspired by Holtzblatt and Bayer's contextual design interviews. The focus of every interview was the onboarding chat (the interview questions can be found

in Appendix I). Two of the interviews contained a card sort technique to enrich the interviews. In practice, words were written on post-it notes and the informant was asked to sort them according to perceived importance. The words on the post-it notes were possible features that the author considered potentially important. These were ‘user flow’, ‘signed/not signed’, ‘fraud’, ‘city’, ‘age’, ‘cohort’, ‘gender’, ‘time to answer a question’, ‘when/where do users leave the chat’ and ‘user’s response time to questions’. This technique was used to get the informant to speak (Benyon, 2019). However, this was not a problem and therefore the card-sort technique was redundant and abandoned in the last three interviews.

Table 2. Semi-structured interviews – list of informants.

Informant	Working area	Length	Date	Card sort
1	Tech & management	30 min	28/1	Yes
2	Marketing	30 min	29/1	Yes
3	Tech	30 min	4/2	No
4	Marketing & insurance	45 min	4/2	No
5	Insurance experience	30 min	1/2	No

Employees from different parts of the company who potentially had an interest in information linked to the onboarding chat were selected as informants. See Table 2 for informants’ characteristics. The interviews took place at Hedvig AB’s office. Moreover, the interviews were recorded with a voice recorder (except the interview with informant 5 due to a technical error). Notes were taken simultaneously.

Informal conversations
Another source of information was informal conversations since the author had access to Hedvig AB’s office. Notes were taken during or after these conversations. Several shorter conversations took place.

Simple contextual inquiry
To understand the product and the service of Hedvig AB, sit-by sessions with two employees were held. This, as a part of Holtzblatt and Bayer’s contextual inquiry approach (2015). See Table 3 for additional information about each session.

Table 3. Sit-by session – list of informants.

Informant	Working area	Length	Date	Description
2	Marketing	30 min	28/1	The author was shown how data is used by the marketing team and in their decision making.
6	Insurance experience	45 min	29/1	The author was shown how the insurance team work with data, insurance claims, questions and what data they use.

Analysis of the fieldwork

The analysis of the fieldwork was analysed with an affinity diagram in a similar vein to Holtzblatt and Bayer's example (2015, p. 26). This approach enables the fieldwork data to be grouped in a hierarchal structure, allowing an overview of the data. The main hierarchal groups were 'what' referring to what the informants wanted to know, 'why' referring to why that information could be useful, and 'how' referring to how a visualisation could be produced and in what way it could be used. Practically, the author carefully listened through recordings and read collected notes and simultaneously wrote notations and quotes underneath the 'what', 'why' and 'how'. Thereafter, inspired by Braun and Clarke's (2006) thematic analysis, a simple thematic analysis was conducted. A theme is defined as an important pattern or an aspect present within the qualitative data. I found four themes, these constitute the results of the fieldwork.

4.2.2 Analysis of the current visualisation

Hereby, the *current visualisation* refers to the visualisation used by Hedvig AB when this study began. A simple analysis of the current visualisation was conducted based on the literature provided above in order to investigate advantages and drawbacks. The structure of the analysis was divided into three parts, each answering whether the current visualisation is designed for interactivity, human perception and analytics, respectively. The aspect of interactivity will be analysed according to the literature in section 2.1 and 2.2, mainly Shneiderman's work (1996). The aspect of human perception will be analysed according to Ware (2008). The analytical aspect will be analysed according to section 2.4, mainly Tufte's fundamental principles of analytical design (2006).

4.2.3 Simple exploratory data analysis

Exploratory data analysis is a method created by John Tukey (1977) that requires the analyst to play with data visually to find characteristics of a data set. Unlike other data-analysis methods, this approach helps the analyst to get a feeling of the whole data set. Another motive is to back selection of suitable techniques and tools for further analysis. Inspired by Tukey, I explored the data and created a dashboard for Hedvig AB's user funnel (for the app) with Klipfolio, an online dashboard software used by Hedvig AB (Klipfolio, 2019). Klipfolio is connected to a data warehouse with PostgreSQL which Hedvig AB use as their relational database management system. Within Klipfolio, data sources are created and linked to so-called 'klips'. A klip could be pre-defined charts, text, numbers or tables. Each klip is data-driven. I call this method *simple* exploratory data analysis since it is only inspired by Tukey (1977). The result is a dashboard. The result is a dashboard and it will help the design process in two ways. First, it will provide an overall understanding of the user-funnel data that is supposed to be visualised. Second, it will briefly show what a dashboard tool, such as Klipfolio (Klipfolio, 2019), could provide in terms of insight and functionality.

4.3 Results

Below are the results of the fieldwork, analysis of the current visualisation and the simple exploratory data analysis.

4.3.1 Fieldwork

Drawing on the fieldwork, four themes are presented, all closely related to the objective of this chapter. The first theme, *overall goals*, raise overall goals of understanding the onboarding chat within Hedvig AB's application. The three remaining themes, on the other hand, target what usage a visualisation could fill and what insights and knowledge that are desired. They are *conversion*, *flow* and *user cohorts.*

Overall goals
Two main goals were found. First, to acquire more customers, i.e. converting users, because growth matters. Insights into the onboarding chat can help to increase the conversion rate, i.e. the number of acquired users divided with the total number of users entering the onboarding chat during a selected period of time. Second, to turn Hedvig AB's insurance product into a great and 'remarkable' product. Insights into the onboarding chat can help to develop the onboarding chat, the application and the overall service. The onboarding chat needs to be a smooth and pleasant experience for every user. Here, a major focus is '[to provide a] nice experience of the onboarding', as one informant mentioned, since Hedvig AB aspires to have friendly dialogues with users and customers as well as being an accommodating company. Preferably, the experience should be so 'remarkable' that people want to speak about Hedvig AB's insurance product.

Conversion
The term ‘conversion’ was mentioned frequently in every interview and refers to the acquisition of new customers, i.e. from the point that the application is initially launched until an insurance deal is signed. Conversion is interpreted as a number viewed both in terms of a distinct number but also as a percentage during a selected period of time. The latter is referred to as the ‘conversion rate’ (e.g. the conversion rate of December was XX %). All informants spoke about how important the conversion of new members is, and questions connected to conversion were frequently raised. For instance, questions like ‘who is converting and who is not?’, ‘who converts?’ and ‘why doesn’t switchers convert better?’ (‘switcher’ refers to a user who already has an insurance deal with another company, yet launched Hedvig AB’s application and goes through the onboarding chat) are questions mentioned. One informant referred to the total conversion rate as the most important number – the number that is desired to increase.

Flow
Flow, on the other hand, includes questions about what is happening in the onboarding chat, i.e. within the user funnel. The theme *flow* roughly addresses *where*, *when* and *how* questions. Questions are, among others: where do users leave the chat, which path do they take to reach the insurance offer provided at the end of the chat, and how long time do they take to answer each question. For example: ‘would be interesting to see what people [users] who presses “tell me more” [an action that provides the user with more information]?’ or ‘here [at this point in the chat] do people spend a lot of time – why is it like that?’ or ‘pathways would be interesting to see.’

User cohorts
The theme cohorts, on the other hand, refers to how different groups with different characteristics are converting or behaving in the chat. Typically, what and who questions. Characteristics could include age, gender, demography, house type, former insurers, sign-up date, whether the user uses iOS or Android, living space, what medium the users come from (such as from Facebook, Google, Snapchat or organic traffic), whether the user is a ‘switcher’, etcetera. Questions are, among others: ‘What variables among users are significant?’ or ‘What is the ultimate combination of [user] variables?’ The informants explained, in different ways, that they want to be able to compare different cohort groups going through the onboarding chat in terms of user characteristics and selected time period. One informant mentioned that he would be interested in comparing the Swedish cohort with the Norwegian cohort, once Hedvig AB has launched their insurance product in Norway.

4.3.2 Analysis of the current visualisation

When this investigative phase was conducted, Hedvig AB was using a bar chart to monitor the onboarding chat as a user funnel. In Figure 1, this bar chart is displayed in Apache Superset’s browser interface (Apache Superset, 2019). Each bar shows the total number of users who have reached a specific question in the chat over a selected period of time. The first bar is ‘First app open’ (to the left) and the last bar is ‘message.kontraktklar.ss’ which refers to acquired customers (to the right). The names of the bars derive from Hedvig AB’s application’s code (see Hedvig AB’s GitHub, 2019) used to push tracking data to their data warehouse. In fact, the bar chart is visually displaying a PostgreSQL query – modify the chart equals modify the query and vice versa (PostgreSQL, 2019). The query used is based on the funnel-view data (see 3.2 Data and ethics). Modify the query can either be done through a pre-defined setting’s panel provided to the left of the bar chart or to rewrite the query. The setting’s panel has titles like ‘Datasource & Chart Type’, ‘Time’ and ‘Query’. The selections below each title show variable names from the data warehouse and interactive elements to modify specific features of the query.

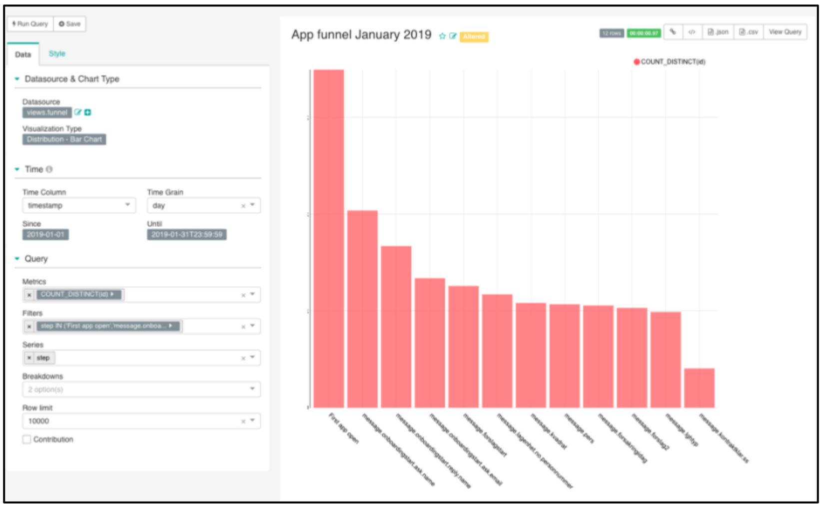


Figure 3. The current visualisation – a bar chart. Please note that amount numbers on top of each bar was removed due to the integrity of Hedvig AB’s data.

Designed for interactivity

Since modification of the bar chart requires PostgreSQL knowledge, at least to some extent (even if the setting’s panel is used), the accessibility is limited. Employees at Hedvig AB who works with marketing, business and product development are using it and some know PostgreSQL and some do not. The bar chart is displayed in an all-around environment for queries. Thus, the current visualisation is not designed to fit the capabilities of its users which is a cornerstone of human-centred design (Benyon, 2019). Yet, a person with good insight into the variable names (bar names), the data warehouse and id familiar with PostgreSQL may be able to use it well.

Regarding Shneiderman’s (1996) information-seeking mantra and task-type taxonomy, the current visualisation cut both ways. The overview is clear – it is a bar chart – but a bar chart does not intuitively imply a user funnel or a user flow. Especially not when it is surrounded by settings and diffuse naming, at least for a person inexperienced with the data warehouse and the naming of the funnel steps. Filter-and-zoom features are existing, especially if the user is familiar with PostgreSQL. Details-on-demand is more or less none existing. But, for instance, the total amount of query rows and the running time of the query are provided. The time range is also discoverable in the setting's panel. The same applies to relationships between items, they are nowhere to be seen. History of interactions, however, is available and so is a save feature. Also, data extraction is convenient. Note the '.json', '.csv' and '</>' buttons are available in the upper-right corner.

Designed for human perception
The design supports easy cognitive tasks, such as change the style or the time range. Regarding what humans can easily see, the bar chart is easily distinguished from its background because of the difference in luminance between the white background colour and the red bar colour. The area of the setting's panel and the area of the bar chart are also easily separable because they both have a white background which is in contrast towards the light grey background of the interface. The two areas are also spatially separated. Furthermore, pop-outs include the green rectangle showing the running time of the query and the yellow rectangle showing that the chart has been altered. The selections, i.e. the grey rectangles with text inside, also pop out against their surroundings. Yet, they do not cause an instant pop-out effect (Ware, 2008).

Designed for analytics
First of all, the chart contains 'chartjunk' since there is no reason for having a grid behind the bars (Tufte, 2001). Regarding Tufte's principles (2006), the current visualisation succeeds in comparison in terms of comparison between the bar lengths, yet it fails in comparison in terms of cohorts, such as characteristics among the users, time period, etcetera. Neither does it provide a clear causality or a multivariate analysis. However, if one is familiar with the data and the naming of the steps in the funnel, the naming can provide a little bit of causality. For instance, 'message.forslag2' equals 'Insurance offer', which could indicate that the price is a key issue (however, that is only one possible explanation). The integration of other pieces of evidence, such as numbers and words are poorly implemented. The title of the current visualisation 'App funnel January 2019' is the only description. Neither is any documentation about what is shown provided (such as scales), nor any of the issues highlighted. A critical issue that is present, but not explained, is that the bars are ordered in height and not in the order of the funnel. Therefore, 'message.forslag2' (the step for the insurance offer) is not next to 'message.kontraktklar.ss' (the last step, signed customers), even if that is the most logical way to display the data. This can clearly mislead the user of the current visualisation. Lastly, the content is quality data directly fetched from the data warehouse accessible through tracking. Hence, it fulfils Tufte's sixth principle; the chart is content-driven.

4.3.3 Simple exploratory data analysis

A dashboard with statistics, numbers and graphs was created with Klipfolio (Klipfolio, 2019). All the klips produced were linked to the application's user funnel. The newly created dashboard contains the following:

- A funnel chart showing the performance of the application during the last 30 days. Each step is showing how users are converting in between steps. This number is displayed in per cent with a total number of unique users presented next to it.

- Three bar charts showing the funnel for the last 30 days, 14 days, and 7 days, respectively – exactly as the bar chart Hedvig AB is currently using (see Figure 3). Each with an additional klip showing the conversion rate for the time range and an additional klip showing the number of people taking the offer in per cent.

- Two bar charts showing weekly signs (for the last 20 weeks) and monthly signs since Hedvig AB's insurance product was launched. Each bar, in both these bar charts, has its value on the top on the each bar to simplify the reading. The reason for the high number of bars is to clearly allow comparisons, as Tufte's first principle highlights (2006), and to uncover trends (if any) in customer acquisition.

- Lastly, a line chart with three lines: one for daily signs, one for daily starts of the onboarding chat and one for the conversion rate. The x-axis covers the last 75 days and one y-axis covers the amount of unique user IDs and a second y-axis covers the daily conversion rate in per cent. The orange colour of the conversion-rate line is to create a pop-out effect, to highlight that it is an important line that is different from the two other lines (Ware, 2008).

All these klips combined together form a dashboard that aims to provide the user with a wide range of evidence on how Hedvig AB's application is performing over time in regard to acquisition and conversion. Titles are selected carefully to inform the user what each klip shows. This approach addresses the idea of integration of different pieces of evidence (Tufte, 2006). Some of these figures are included in Appendix II.

4.4 Conclusion: the needs of Hedvig AB

To conclude, the fieldwork shows that Hedvig AB aspires to be a data-driven company and insights into the inflow of users from the onboarding chat could be helpful to acquire more customers and to build a better, more remarkable product (such as designing a better onboarding chat). Regarding the core intents of the visualisation that will be designed, these are mainly the themes found. In particular, the three themes linking to insights that are desired: *conversion, flow* and *user cohorts.* Therefore, the focus of the visualisation project will target the following issues in the following order:

- The conversion of customers.
- An overview of the flow. This aspect includes, for instance, questions like: how can the flow of users be understood with a quick glance? Where are users

dropping off? How long does it take to go through the onboarding or to answer specific questions?
- Distinguish user cohorts. This aspect includes, for instance, questions like: how can different cohorts be separated? How can different cohorts be understood in relation to other cohorts or towards the general flow of users?

The analysis of the current visualisation showed that the current visualisation – the bar chart – is useful in some ways: it is simple, it provides an overview, and it is easy to compare the length of the bars. But there are several questions that cannot be answered with it. In addition, the current visualisation is too simple for analytics; it contains issues; the names of the funnel steps are hard to understand; it requires the user to be familiar with PostgreSQL to interact with the visualisation: and lastly, it provides only a little bit of information. In general, a lot of tacit knowledge is required to use it and to understand it well.

Lastly, the simple exploratory data analysis gave a good insight into the characteristics of the data and broadened the perspectives of the visualisation challenge of this project. In addition, the dashboard created is an improvement of the current tool since it shows multiple graphs and key numbers with descriptions in a more accessible and user-friendly environment. This is a good way to look at basic data over a longer period of time.

Drawing on these results, the visualisation will be developed in regard to the following aspects. First, it will not show the same numbers and data as the dashboard. Second, it will target the themes and needs presented above from the fieldwork. Third, it will be more accessible and self-explaining than the current visualisation. Fourth, it will be developed both to be a visualisation that can be viewed on a computer and on a monitor. Fifth, it will target users – employees at Hedvig AB – that are familiar with insurance and Hedvig AB's product. These conclusions are acting as guidelines for the design and development presented in Part II.

5. Part II: The development phase – design and deliver

This part of the thesis will explain the development phase and its results. In particular, it will focus on the procedure from idea to sketches to one prototype.

5.1 Objectives

Drawing on the previous part and the literature, the aim of this chapter is to develop a prototype of the user funnel – a visualisation concept – that can be used for evaluation, and thereafter, be developed further.

5.2 Methods

A prototype can be anything from a sketch on a piece of paper to a working product and it is a common tool for designers to develop and design products since it is, for instance, a good way to communicate ideas and test designs (Benyon, 2019). A trade-off, among others, is that 'high-quality graphics and animation can be used to create convincing and exciting prototypes *but* may also lead to premature commitment to some design decisions' (Rosson and Caroll, 2002, cited in Benyon, 2019, p. 203). In this development phase, I aspired to develop a prototype with a well-developed concept, yet without interactivity and minor details since those could be addressed later – after the evaluation of the prototype. In short, I first developed simple sketches, then advanced sketches, and thereafter produced two prototypes. Lastly, one prototype was selected for evaluation.

The process from simple sketches, to advanced sketches, to two prototypes, and finally, to one prototype requires selection. Each sketch and prototype comes along with a motivation of why it was *kept* or why it was *dismissed*. Selections are based on the following:

1) Whether the sketch or prototype could fulfil the findings of the previous part.
2) Whether the sketch or prototype could address the literature presented in chapter two.
3) Whether the sketch or prototype was viable in terms of technical implementation and doable in regard to the timeframe of this project.

Since a visualisation is data-driven, data accessibility has been taken into account while designing the sketches and the prototypes. Moreover, all the sketches as well as the prototypes will be provided below.

5.2.1 Simple sketches

The sketches could be seen as 'lo-fi' prototypes since they act as real products while only being mediators of a concept. Lo-fi prototypes are typically in the shape of a sketch on a piece of paper drawn by hand. This is efficient when communicating or trying out new ideas (Benyon, 2019).

The simple sketches were sketched by hand with a marker on a piece of paper. Each sketch was targeting Shneiderman's 'overview first' (1996) to mediate the concept of a user funnel. This was suitable in order to communicate brief ideas and to spot issues and potential advantages at an early stage (Benyon, 2019). Six simple sketches were developed.

5.2.2 Advanced sketches

The advanced sketches were sketched with Sketch (Sketch, 2019), and, Javascript and D3 (Javascript, 2019; D3, 2019). The advanced sketches also followed the notation of 'overview first' and were designed accordingly. But in contrast to the simple sketches, each advanced sketch was developed further towards 'zoom and filter' and 'details-on-demand' (Shneiderman, 1996). In practical terms, elements for interactivity (e.g. input fields and hover effects) and/or elements for additional information (e.g. text and numbers) were added. In addition, the advanced sketches were also developed further in terms of layout. At this point, the designer played around with basic pop-out channels, such as colour, shapes, motion and spatial grouping in order to explore pop-out effects and what could easily be seen (Ware, 2008). Three advanced sketches were developed.

5.2.3 Two prototypes

In contrast to a 'lo-fi' prototype, a 'hi-fi' prototype is usually produced well in terms of design and functionality, allowing more detailed evaluation (Benyon, 2019). Therefore, the two most promising advanced sketches, in regard to the selection process described above, were developed into hi-fi prototypes. The two prototypes originated from the advanced sketches, but with two major differences. First, each prototype was driven by real data, and second, each prototype was programmed instead of simply being sketched. Hence, the main idea was to test the visualisation concepts with real data and code, in order to see if they were suitable for the aim of the visualisation, i.e. to create a visualisation of a user funnel that increases understanding.

Tools used for turning the lo-fi prototypes into hi-fi prototypes were Python, Pandas and Dash (Python, 2019; Pandas, 2019; Dash, 2019), and Javascript and D3 (JavaScript, 2019; D3, 2019), respectively. Both these were selected since they can be used to create analytical, data-driven and visually appealing web interfaces. To develop for the web is suitable since a visualisation interface created for the web can easily be accessed on a browser, either on a monitor or a private computer. (A more expanded methodology on the selection of tools, for the selected prototype, will be provided in Part IV.)

5.3 Results

Below are the results presented: simple sketches, advanced sketches and, finally, two prototypes, along with their intent and motivation why they were *dismissed* or *kept*. The numbers associated with each sketch are the same throughout this chapter.

5.3.1 Simple sketches

Every sketch has a black arrow pointing out the direction of the user funnel. The simple sketches are presented below (see Figure 4).

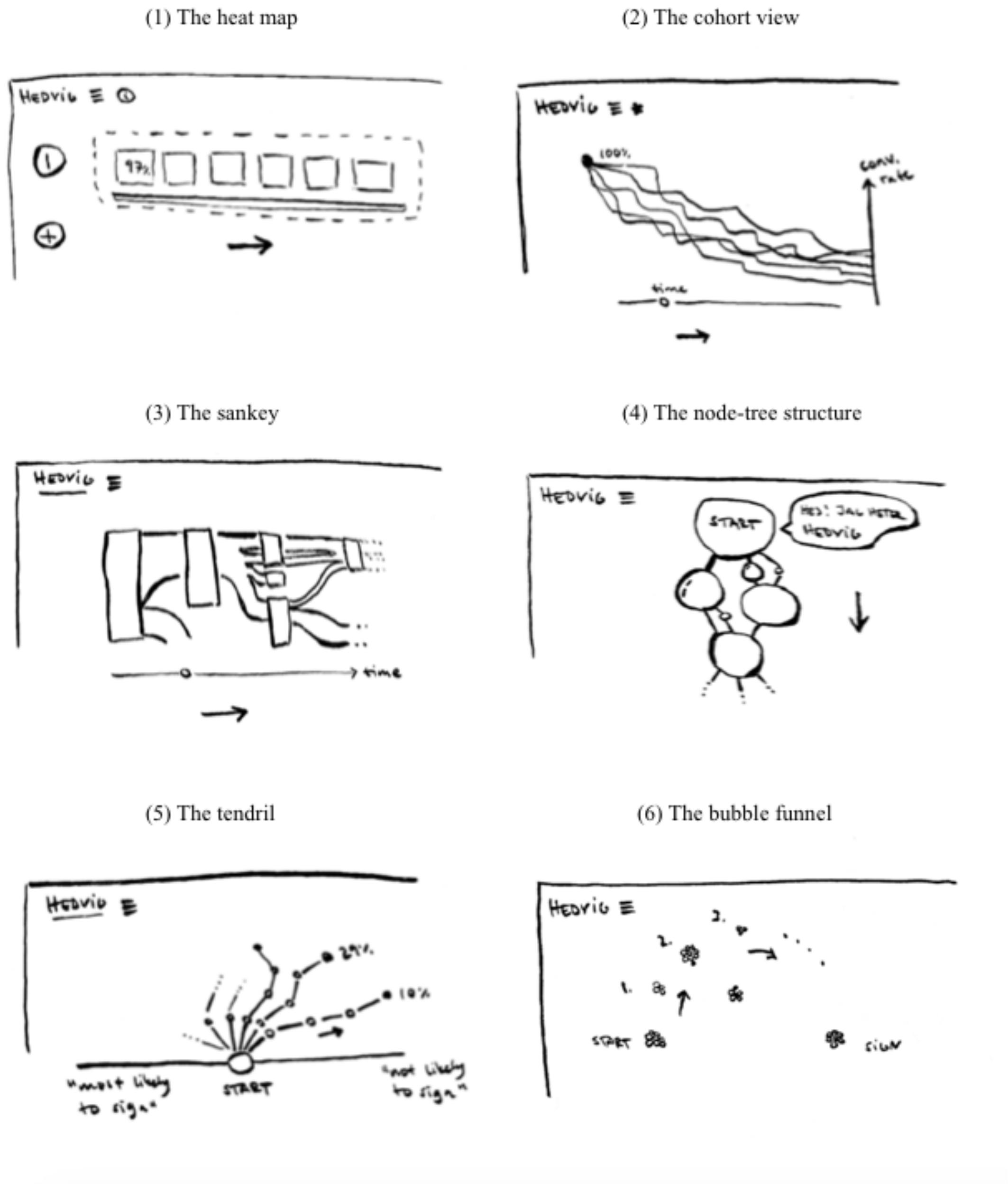


Figure 4. Simple sketches.

(1) The heat map
The flow is visualised horizontally in a heat-map style where each box is a step in the funnel. The oblong rectangle underneath symbolises the average time in the funnel with a colour gradient (e.g. red colour underneath a step equals much time spent at that certain step by the users). A ⊕ symbol to add additional funnels, e.g. funnels for different cohorts, is included. Inspiration comes from the fact that heat maps are commonly used to show user progresses among cohorts on digital platforms (see e.g. Han, 2017).

The heat map was *dismissed* since it is static, it does not utilise the richness of what a visualisation can provide, such as with motion, shapes or spatial grouping (Ware, 2008).

(2) The cohort view
Every line symbolises a cohort, such as gender, current insurer, created-on month, age, etcetera. All cohorts, together, create a big ”funnel wave” that starts in 100 per cent conversion rate (since no drop-offs have occurred at the first step) to the left, at the first step in the funnel, and ends in a y-axis showing the conversion rate for each cohort. The x-axis symbolises the funnel, each tick is a funnel step. Thus, the idea is that the conversion is decreasing at every step along the x-axis ending in the y-axis (see Figure 4). This view could make it easy to compare the conversion rate among cohorts. Inspiration comes from Ashkenas and Parlapiano's multiple line chart (2014).

The cohort view was *kept* since it could fulfil the requirements found in Part I – foremost it shows the conversion very clearly with an overview-first view (Shneiderman, 1996). In addition, it efficiently shows contrasts and differences between cohorts, hence, following Tufte's first principle well (2006).

(3) The sankey
A sankey diagram (sometimes referred to as an alluvial diagram), is a traditional way of showing a flow. This sketch shows the user funnel with multiple pathways towards a common goal, in this case, to sign an insurance deal. Below is a slider to adjust the time range. Each node (or rectangle) represent a question while the links in-between the nodes represents the flow. The funnel starts from left flowing rightwards. This sketch is partly inspired by Google's visualisations of user flows (Google, 2019).

The sankey was *dismissed* despite the fact that it generates a good overview-first view (Shneiderman, 1996). Partly since it does not utilise the richness of what a visualisation can provide, such as motion, shape or spatial grouping (Ware, 2008). Partly since a sankey showing multiple cohorts (e.g. in shape of coloured links and nodes) or an interface with multiple sankeys (one for each cohort) could easily generate a messy visualisation due to an overflow of nodes and links connecting, endanger a clear overview-first view (Shneiderman, 1996).

(4) The node-tree structure
Inspired by a network graph, each node (or circle or bubble) represents a chat question

or an answer where the size of the node symbolises the number of users reaching that question. The idea is to plot all possible pathways showing all interactions with the onboarding chat in one single graph. The flow starts in the top and with a direction downwards.

The node-tree structure was *kept* since it could fulfil the requirements found in Part I. In addition, it creates a novel overview-first view showing all questions and answers present in the chat together with all pathways. Nodes entail that spatial grouping can be used efficiently (Ware, 2008).

(5) The tendril
This approach is inspired by Karpefors and Weatherall's newly invented 'tendril plot' (2018). Each tendril (line) represents a cohort which is bending towards left or right whether the users of that cohort are likely to become customers or not. The tendrils and their nodes could change in size or colours to add features. The funnel starts in 'start' and ends at the end of each tendril.

The tendril was *dismissed* since it does not create an intuitive overview of a user funnel. Especially since there are many tendrils (lines) going in different directions – on the contrary to an 'intuitive' flow which has a certain direction. Thus, there are better ways to provide an overview-first view (Shneiderman, 1996).

(6) The bubble funnel
Inspired by Nathan Yau's animated bubble and cluster chart (2019) and Gapminder's animated bubble chart (2019), this sketch shows a user funnel constituting of bubbles and clusters. Each cluster is a step in the funnel, while each bubble is a user. The idea is to show the funnel as a clock where each bubble moves clockwise from start to sign. In contrast to the other ideas, this sketch lets the users form the visual representation of the user funnel since the users make up clusters that show a distribution of the users in the funnel. If there are no users in the funnel, the screen will more or less be a blank sheet.

The bubble funnel was *kept* since it could fulfil the requirements found in Part I. In addition, it provides a funnel that could take advantage of spatial grouping, motion, and colour – thus utilise the pop-out channels (Ware, 2008).

5.3.2 Advanced sketches

Drawing on the simple sketches, three advanced sketches were developed.

(2) The cohort view
This sketch was sketched with Sketch (see Figure 5). In contrast to the simple sketch of the cohort view, filter and zoom functions were included to follow the information-seeking mantra (Shneiderman, 1996). Features include a selection of cohorts through a search field, an average line aggregating the selected cohorts, a coloured oblong for measuring time (e.g. red colour underneath a step equals much time spent at that certain step by the users), and lastly, a slider for selection of time. Hoovering or clicking on a

line provides a tooltip or similar to provide further information in order to simplify the comparison of cohorts and to include details-on-demand.

The cohort view was *kept* with the same motivation as above (see *5.3.1 Simple sketches*). In addition, in regard to Shneiderman's (1996) taxonomy for tasks, it has the potential to be developed according to his six categories of tasks.

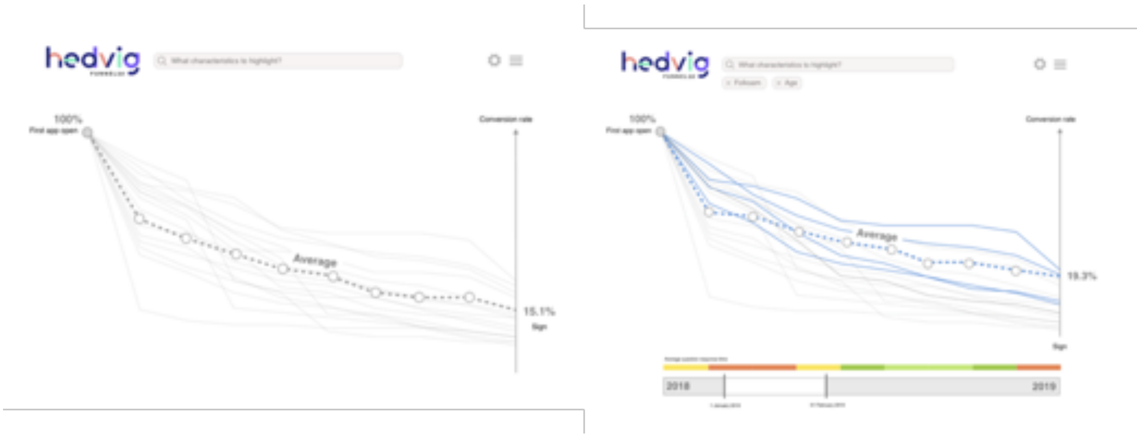


Figure 5. The cohort view as an advanced sketch. One with the overview (left) and one with details-on-demand and filter-and-zoom features.

(3) The node-tree structure
The node-tree structure was designed further into two alternatives. Alternative one (see Figure 6) was made with Javascript and D3 and alternative two (see Figure 7) was made with Sketch; alternative one form a flow horizontally while alternative two forms a flow vertically. Interaction with a bubble should generate details-on-demand (Shneiderman, 1996). Two mouse-hoovering interactions are exemplified: in alternative one, the name of the step, the conversion percentage and the total amount of users who have reached a certain node is displayed in black text; in alternative two, the name of the step, an average time number, and a performance number (green text) represented as a percentage difference in user interaction is shown. As mentioned previously, the size of the bubbles represents the number of users who have triggered that specific question or answer within the chat.

The node-tree structure was *dismissed* for one main reason: the idea to let the size of a bubble represent the number of users equals the idea to let the height of a bar (e.g. in a bar chart) represent the number of users. Hence, the node-tree structure is equal to the bar chart used in Hedvig AB's current visualisation used today (analysed in Part I). Furthermore, differences between sizes of circles are less intuitive and more difficult to perceive compared to the height of bars.

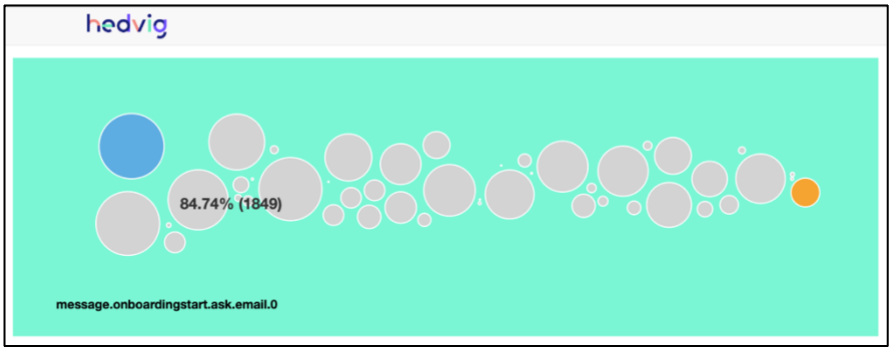


Figure 6. The node-tree structure as an advanced sketch – alternative one. The blue node represents the start of the funnel while the orange circle represents the end of the funnel. The flow is horizontal.

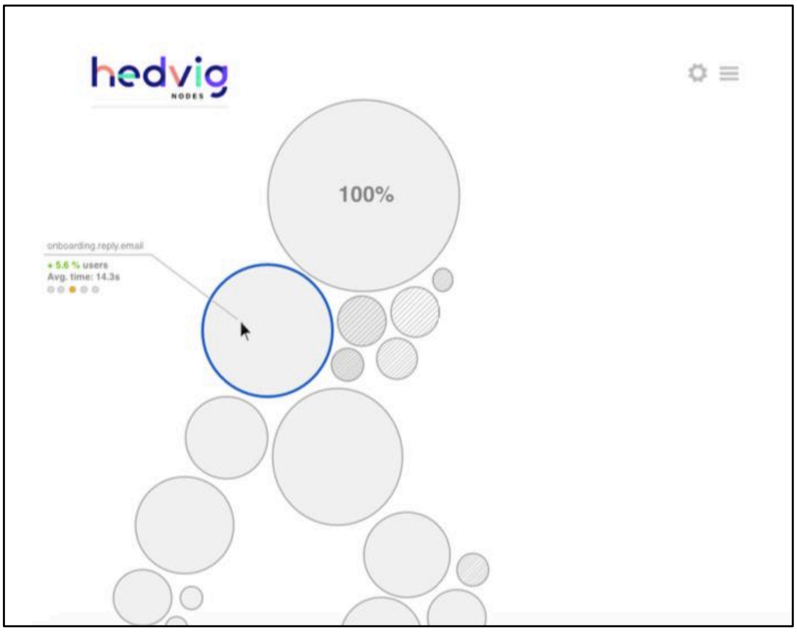


Figure 7. The node-tree structure as an advanced sketch – alternative two. The large bubble is the start of the funnel and the flow is vertical in a downwards direction.

(6) The bubble funnel

Bubbles are entering from the left moving rightwards (see Figure 8). Five steps are shown, two steps at the beginning of the funnel, one 'in-between step' symbolising multiple steps, and then, 'offer' and 'sign'. In this sketch, bubbles move randomly, and the data is randomised. Information is provided underneath each step and symbolise that statistics and numbers could be displayed. This sketch mainly shows an overview of the flow by focusing on distribution and user drop-offs. While adding motion – bubbles jump between steps – this brings life to the data.

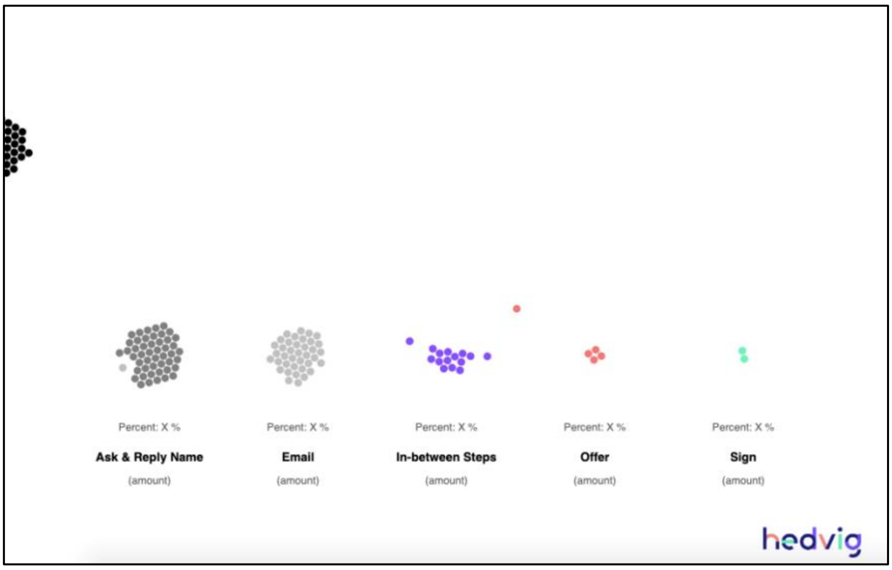


Figure 8. The bubble funnel as an advanced sketch. The flow is horizontal and the funnel starts in the black cluster (left) and ends in the green cluster (right).

The bubble funnel was *kept* with the same reason as above (see *5.3.1 Simple sketches*). Additional advantages include, first, that the bubble funnel leaves so much free space on the interface for additional numbers, text, bubbles, clusters, etcetera. This also applies to interactivity, there is much room to add filter-and-zoom features as well as details-on-demand (Shneiderman, 1996). Thus, the integration of additional evidence is possible (Tufte, 2006). As can be seen in the sketch, basic pop-out channels work well with this sketch (Ware, 2008).

5.3.3 Two prototypes

Drawing on the advanced sketches, two prototypes were created in order to test the concepts with real data. The prototypes and their motivations are described below:

(2) The cohort view
The cohort view was developed with Python, Pandas and Dash. Cohorts were filtered with Python and Pandas and displayed as light-grey lines, a title and a brief description were added in the top of the page (see Figure 9). Regarding interactivity, Dash's settings menu can be seen in the right corner offering a range of filter-and-zoom features and while hover over a line a tooltip is displayed with the conversion rate, providing detail-on-demand. A 'funnel wave', i.e. a multiple line chart, is displayed in the middle of the screen. However, the development of this prototype got an abrupt end.

The cohort view was *dismissed* since the concept failed – the visualisation was not designed for the data (as can be seen in Figure 9). The cohort data, i.e. anonymised user data, was collected too late in the onboarding chat and a large amount of the users did not go that far. This implies that users who are associated with a cohort had already been going through almost the whole funnel, meaning that the conversion rate at a majority of the steps in the funnel is about 100 per cent. Therefore, the only line that is of interest is the line ‘All users’ (the bottom line), and that line is only another representation of the bar chart which constitutes the current visualisation used by Hedvig AB (see 4.3.2 Analysis of the current visualisation). However, the cohort view would have been interesting if all user data was collected at step one in the funnel, or in a scenario where all user data already is known.

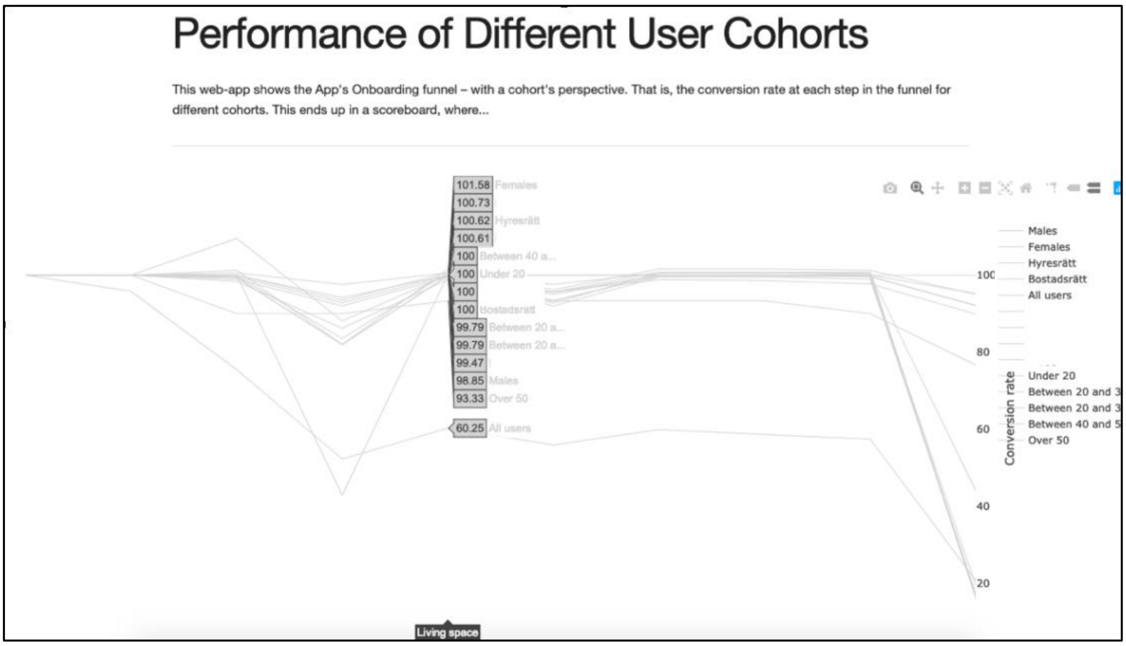


Figure 9. The cohort view as a prototype.

(6) The bubble funnel
The bubble funnel was developed with D3 and JavaScript allowing integration with the data (see Figure 10). It was developed into two different ideas: one for historical data and one for live data. Hereby, the one for historical data will be referred to as the *simulation view* and the one for real-time live data will be referred to as the *live view*.

The simulation view intended to visualise a simulation of a day in Hedvig AB’s onboarding chat. This first version had no interactivity. As before, each bubble represents one user. On start (refreshing the web page) the time is ‘00:00’ with all bubbles (users) clustered in the left outskirt – each bubble ready to move rightwards and enter the funnel. At the end of the simulation, at ‘24:00’, bubbles are distributed along the funnel in clusters. Each bubble represents a user who has had some kind of interaction with the onboarding chat that specific day. An interaction could be one click, one answer or multiple of either (but the data could, obviously, be changed to only include e.g. user interactions where all users are created on the same day or similar). Bubbles at the step ‘Sign’ are green to symbolise acquired, converted customers. Users who clicked login and are at the ‘Logged in’ step are displayed as purple bubbles to distinguish them from the bubbles in the funnel which are coloured black. Each colour is selected to be easily separable from the background, and from each other. Each funnel step’s name, the percentage of the number of bubbles, and the total amount is provided underneath each cluster. That information is there as a proof-of-concept – adding stats, numbers and text are possible. A text string underneath the time and date is shown to add some descriptions correlating to the time, such as ‘Early birds are up…’ or ‘Time to sleep… Zzz…’ correlating to the time of the simulation. Furthermore, the idea of having a clockwise flow displayed as a fan (see simple sketch, Figure 4) was abandoned since a straight flow was more intuitive. Lastly, the percentage is a distribution rate of all the bubbles.

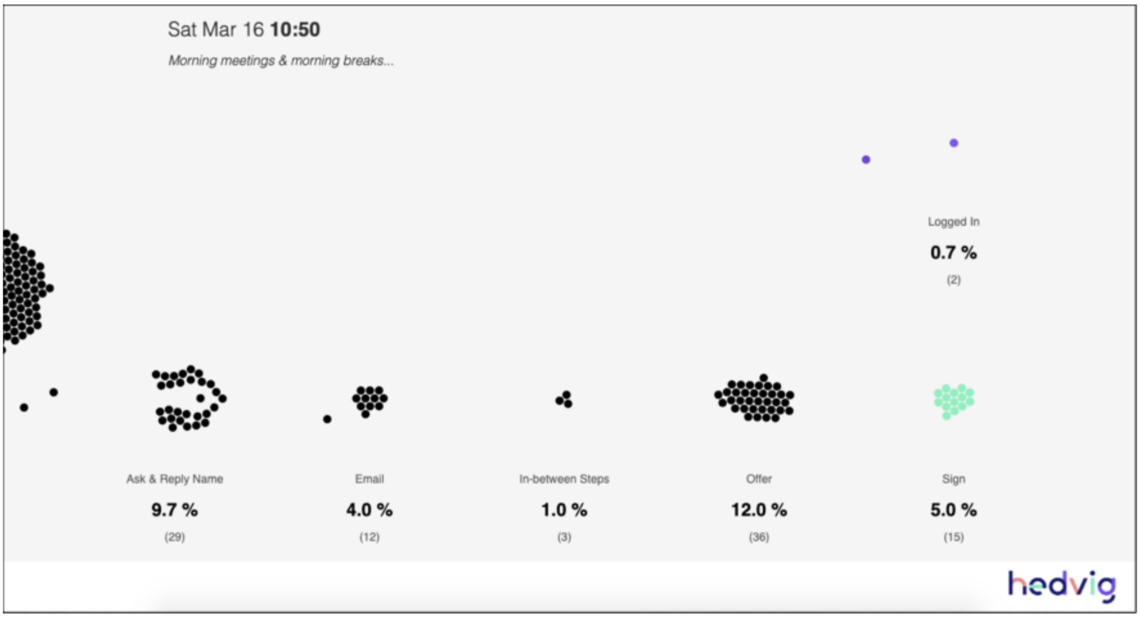


Figure 10. The bubble funnel as a prototype. The bubbles not belonging to a cluster are moving to their next destination, the next cluster.

The second one, the live view of the bubble funnel, is similar to the simulation view in layout but slightly different in usage. In contrast to the simulation view, it has a clock showing the current time and date. Moreover, the data is generated with a Python script, containing PostgreSQL, which constantly fetches data from a data warehouse. In the live view, the bubbles only move if someone is interacting with the onboarding chat somewhere, somehow. Otherwise, colours, layout and the information provided are similar to the simulation tool.

Regarding motion and the clustering in both views, each bubble’s selected cluster is calculated through a JavaScript function comparing the current time with the data. Asynchronously, the position of the bubbles in terms of x and y coordinates, both in relation to other bubbles and to the selected cluster, is re-calculated frequently to allow smooth animations. This is done with D3’s force layout (D3 Force layout, 2019). The JavaScript functions handling gravity, collision and the ‘pop’ come from an example tutorial by Nathan Yau (2019) and the appearance of the bubbles comes from an example by Mike Bostock (2019 A). Another note regarding the motion of the bubbles, once a bubble has a new destination (a cluster) it creates a ‘pop’ and adds padding towards surrounding bubbles and then it takes off to the next cluster (see Figure 11). The gravity is towards each cluster. If a bubble hits another bubble or a cluster, it collides – as if a ball hits another ball in real life – and then continuing its journey towards its next cluster.

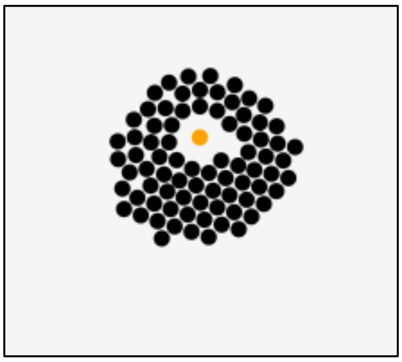


Figure 11. A 'pop'. The yellow bubble is getting a new direction, therefore, padding is added. Now, the bubble is on the way to its next cluster.

The bubble funnel was *kept* since it could fulfil the requirements found in Part I. In addition, it provides a visual representation of the user funnel that could take advantage of spatial grouping, motion, and colour – thus utilise the pop-out channels efficiently (Ware, 2008). In regard to Tufte's principles, they could all be fulfilled when the bubble funnel is developed further. This also applies to Shneiderman's taxonomy of tasks (1996), especially since there is much space left of the screen. Moreover, the bubble funnel makes data look beautiful and appealing, which may or may not provide usability (Tractinsky et al. 2000; Tuch et al. 2012).

5.4 Conclusion: the bubble funnel

Conclusively, this chapter has developed a visualisation concept in the shape of a data-driven prototype mediating a visual representation of the user funnel, as intended. The prototype – the bubble funnel – is a good alternative according to the literature (provided in chapter two) and could be developed to fulfil the needs of Hedvig AB (presented in Part I). Thirdly, the development is also doable within the timeframe of this project.

Moreover, the bubble funnel is working with different types of data (historical data and real-time data) and two main ideas of how the bubble funnel could be used are presented: the *live view* and the *simulation view.*

The live view could be used to monitor live data of the onboarding chat's performance in real time, either on a monitor or on a laptop. The simulation view, on the other hand, could be used to re-play the user flow in order to compare days and to get a feeling for the data in an exploratory fashion. In regard to the themes found in Part I, ideas on how the bubble funnel could benefit include the following:

- It could clearly show the conversion, from the first step to the last step in the funnel, i.e. from start to sign. Converted users will be displayed in their own cluster, and the same applies to users who have been presented an insurance offer.

- It could provide a visual overview of the user funnel, showing what is happening in the funnel and where users are dropping off combined with details-on-demand such as text, numbers, etcetera, targeting time aspects, conversion rates, number of users, or distribution rates.
- Cohorts could potentially be addressed by selecting different colours on the bubbles or to have multiple bubble funnels in the same interface. These could be selected through common interactive web elements such as buttons and selectors.

In addition, the bubble funnel may be able to:

- Provide hypotheses and give clues on phenomenons that should be investigated further, such as with proper data-analysis techniques or with dashboards that allows a large amount of collected data over long time periods be aggregated and displayed.
- Create interest in data among employees not working with data. Thus, being a bridge between the data and the company, especially if it is displayed on a monitor.
- Spot emergency issues. For example, if all users suddenly start dropping at a certain step in the funnel, that may be an effect of a software bug in the digital service.

The options of development are many and so are the ideas. Yet, whether this concept works will be evaluated in the next chapter.

6. Part III: The evaluation phase – issues revealed

This part of the thesis will explain the evaluation of the bubble funnel – the prototype developed in the previous part. In particular, it will focus on *why, when* and *how* the evaluation was carried out and *what* potentials and issues the prototype has.

6.1 Objectives

Drawing on the prototype developed in the previous chapter, this evaluation aims to answer the following questions:

1) What is working and what is not working with the prototype?
2) How can the prototype be improved and developed further?

6.2 Methods

In order to answer these questions and to evaluate the prototype, sit-by sessions were held where different scenarios of the prototype were shown for potential users, i.e. employees at Hedvig AB (Rubin and Chisnell, 2008). These informants were the same informants as the informants interviewed in Part I (see Table 4 for user characteristics). Noteworthy, all informants had pre-knowledge about the aim of the project – to visualise the user funnel.

Table 4. List of informants.

Informant	Working area	Length	Date
1	Tech & management	40 min	29/3
2	Marketing	35 min	28/3
3	Tech	50 min	28/3
4	Marketing & insurance	30 min	1/4
5	Insurance experience	25 min	29/3

The sessions were accompanied by semi-structured interview questions (Benyon, 2019). Each evaluation session constituted of three parts: open questions, task-based questions and a further-development conversation. The two first parts focused on the prototype, the bubble funnel, while the latter part focused on how the prototype could be expanded and improved in terms of functionality, interactivity and design.

Appendix III shows the questions and Appendix IV show static pictures of the prototype in action. The numbers of the scenarios below are mapping to pictures in Appendix IV.

6.2.1 Open questions

This part addresses the overall design and the visualisation concept: movement of bubbles between clusters in a horizontal flow from left to right. Two simplified versions of the prototype were shown for the informants accompanied with questions.

1) Bubbles from left to right in 5 steps.
2) Bubbles from left to right in 5 steps, but with more details.

The idea behind this set-up was to exclude a narrative of these scenarios, and instead, to focus on the impression of the bubbles and the clusters, and then to see whether the impression changed when more details were added. Still, without explaining any narrative.

The questions accompanied were broad, targeting the users' first impression and general thoughts while having an initial encounter with the prototype. The author – at this part of the evaluation – was focused on listening and asking neutral follow-up questions, such as: 'can you explain your thoughts in more detail?'. Throughout this part of each session, the informants were also asked to 'think out loud', i.e. say what they were thinking while visually engaging with the prototype and being asked questions in order to access informants' thoughts and concerns (Rubin and Chisnell, 2008).

6.2.2 Task-based questions

Thereafter, task-based scenarios were introduced to the informant because: 'people find it difficult to react to a prototype if it is simply placed in front of them devoid of any context; some sort of structuring narrative is required' (Benyon, 2019, p. 202). Since the prototype is designed to show information visually, the questions were addressing information that could be retrieved. Questions were, among others: 'How many per cents of the users are currently at the email-question?' (see Appendix III for more all the questions). Once again, the informants were asked to 'think out loud' (Rubin and Chisnell, 2008). And once again, the author was listening and asking neutral follow-up questions. These are the following scenarios that were tested:

3) The visualisation with live data – the *live view*.
4) The visualisation showing a simulation of a day – the *simulation view*.
5) Retrieve information from a user or a step in the funnel.

Hence, the main focus was to evaluate the two main ideas: the live view and the simulation view. The live view vas displayed first and contained information about what was happening on the screen: 'Today's current state of the onboarding chat. Every bubble represents a user at its present step in the onboarding funnel.' In contrast, the simulation view did not have a description. Instead, it was presented after the live view, to indicate that it was the same visualisation, containing short messages hinting that a simulation was taking place. Here are some example messages: 'Good morning! Time to go to work', 'Sign up for Hedvig during lunch?' or 'End of the day.' The idea here was to present both concepts – the live view and the simulation view – in a similar, yet a slightly different manner, to see if the informants grasped the concepts.

Regarding the last scenario, ‘Retrieve information from a user or a step in the funnel’, the idea was to find out where the informants expected additional information about a bubble or a cluster to be found (if interactivity was included). I.e. where would the details-on-demand most logically be included (Shneiderman, 1996).

6.2.3 Further-development conversation

Lastly, in contrast to previous sections, this part was a discussion on how the prototype can be expanded and improved in terms of functionality, interactivity and design. Here, the author took a more active role in discussing issues and ideas. The author had the prototype ready in order to be able to use it as a base for the discussion. Both the live view and the simulation view were discussed.

6.3 Results

This section compiles the results from the evaluation sessions.

6.3.1 Open questions

When asked what they thought about when facing scenario (1), three informants referred to the funnel – they said that they saw the funnel. The two remaining informants, on the contrary, said that the bubbles remind them of ‘molecules’, ‘a liquid’, ‘ants’, ‘popcorns’ or ‘an aquarium’. Everyone except informant 4 informant thought that it was appealing and nice to watch the bubbles moving from cluster to cluster. Informant 4 instead said that ‘it is happening stuff, so it is not boring to watch.’ One informant adds that it is ‘visually appealing’ but hard to see how many bubbles there are and requested actual numbers. Adjectives used to comment on the movement of the bubbles were, among others, ‘beautiful’, ‘cute’, ‘nice’ and ‘fun’. Informant 1 liked the symmetry of the bubbles while being clustered.

Otherwise, regarding scenario (1) and (2), there was a wide range of inputs and opinions. Three informants mentioned that they thought it was confusing that the bubbles appeared in the left outskirt of the screen – yet, they had qualified guesses why that was the case. One informant said that he prefers ‘function in front of finesse’, but adds that he sees a value of creative data visualisation for a wider audience, saying that it can create engagement among users of the visualisation. Another informant says that the flow of users can be better understood with a prototype as the bubble funnel compared to other data representations, such as a bar chart. Lastly, all informants think, in one way or another, that this type of data visualisation could create interest in the data.

6.3.2 Task-based questions

Below are the results from the task-based questions.

Live data
While facing scenario (3), all informants seemed to understand the visualisation directly. Three of them were confident and correct while 'thinking out loud'. While the other two said things that were correct in the beginning, but while the time passed and they were exposed for the visualisation a longer time, they began to say things that were not correct. For example, one informant expected the visualisation to be similar to the bar chart (the current visualisation used by Hedvig AB), but in the shape of bubble clusters, and therefore, got confused. The other informant who was unsure did not understand what was happening and got stressed by the task-based questions. Both these informants did not read the description provided underneath the time and date. This was also the case for informant 2, he did read the description text, but first after some time.

Regarding the task-specific questions regarding percentages, sizes of clusters and amount of bubbles, four informants answered correctly to all of them. The last informant asked correctly to a majority of the questions, but not to all of them. Yet, almost everyone was a little bit confused with the percentage value displayed underneath each cluster – was it a conversion rate between steps, a distribution rate or a completely different value? However, the number of bubbles, the funnel, the time and date, and the size of the clusters were information interpreted correctly. Another thing that was confusing according to the informants was the 'Logged in' cluster (with purple bubbles). Mainly, the location of it and whether it was a part of the funnel or not was causing confusion.

Other aspects include the following. One informant working with tech was questioning whether the bubble funnel could be used in decision making or not – instead, he would like to see a larger focus on conversion and cohorts. He adds that a visualisation like a bubble funnel entails a risk since the decision making would be drawn on a too small scope.

Simulation
While facing scenario (4), everyone grasped the concept of a simulation, except informant 4 who still was a little bit confused. Much of the feedback and criticism was similar to the one mentioned above in the live-data view. This time, however, the informants seemed to get slightly confused by the shift in colour from grey to green (a colour transition that was implemented with a duration of two seconds) for bubbles moving from one cluster to the last cluster, the signed-customer cluster. For instance: 'they get green already at offer, hmm…' and 'it is a little bit unclear when they get green.' Some informant added that the bubbles are moving fast, not really understanding that it is a consequence of the time being simulated.

Furthermore, one informant reacted spontaneously to the data: ‘Holy shit, it is so many people who get an offer!’ One informant said that he liked the concept to re-play a day and added: ‘to be able to compare cohorts is valuable, and not only snapshots [i.e. referring to the live-data view].’ Also, all informants remembered one or several of the text messages underneath time and date in the simulation view.

Retrieve information
Every informant wanted to click or hover on a bubble to retrieve information about a user. To get information about a step, every informant wanted to hover, click or interact with either the cluster of bubbles or the funnel’s name or information area underneath the cluster.

6.3.3 Further-development conversation

One informant working with marketing said that ‘[the bubble funnel is] a much more qualitative variation of a funnel than any I have ever used or seen.’ He continued and explained that the advantage of Tableau, Apache Superset, Hotjar and Google Analytics is that he can select multiple parameters of the data, but they are static pictures and he sees a value in motion, and the movement of users.

Regarding cohorts, two informants mention that it would be interesting to see differences between students, switchers, age, price of the insurance offer provided, and whether users come from special campaigns or not. One informant suggests cohorts to be in different colours and another one suggest the cohorts to be in layers, i.e. multiple bubble funnels on top of each other. Informant 5 wants to be able to distinguish students since that would be interesting when ‘we run student campaigns.’

Another input is to add the web’s funnel, instead of only looking at the app’s funnel. Two informants said that they would rather see a drop-off rate from the previous step instead of the distribution percentage. Informant 1, on the other, thinks that the distribution per cent could be useful. A third informant highlights that it is very easy to see where users drop. One informant working with tech says: ‘you get a very good idea to questions for investigating with this method.’ The same informant asks whether it is possible to add steps or not to the interface. He also asks whether it is possible for a bubble to skip steps since that was what he thought happened while watching the simulation view since the bubbles are moving from the first cluster, in the left outskirt, downwards in a slanting direction above early clusters to clusters at the end of the user funnel.

6.4 Conclusion: a concept with potential

Drawing on the results of the evaluation sessions, some conclusions are provided below. First and foremost, a majority of the informants grasped the concept of the prototype – the bubble funnel – easily. The prototype is working in regard to:

- The representation of a funnel with certain steps and a flow from left to right is interpreted correctly.
- The clustering and distribution of bubbles (users) within the funnel seems to work well. For instance, according to the answers of the task-based questions, it is easy to compare the size of different clusters, the number of users at each step, the percentage (in this case, the number is easy to read and see), easy to understand what a bubble represents.
- The conversion of users. The informant experienced no problems in finding how many users that have signed. This also applies to the drop-offs of users.
- The layout in terms of aesthetics seems to work. Especially since the movement of bubbles together with the clusters seems to be appealing. In fact, the informants were positive towards this type of data representation and thought it could generate interest in data.

On the contrary, everything was not working well. Some issues were found as well as some needs that the prototype has to address:

- The fact that the bubbles start in the left outskirt of the screen was indeed confusing and not intuitive at all.
- The informants expressed that cohorts need to be addressed, such as 'students' and 'switchers', where the users are originating from (i.e. the source of destination before entering Hedvig AB's chat), and price level of the insurance offer provided in 'Offer' in Hedvig AB's onboarding chat. One user points out that he would like to see what the 'Offer' cluster consists of.
- Interactivity has to be addressed. Especially since there is no interactivity at the moment.
- A number of smaller issues, such as the position of the 'logged in' state and whether it is a part of the funnel and not; the transition of the bubbles' colour between steps; the speed of the bubbles.
- Lift the 'business value' of the visualisation.

In the next section, these findings will be addressed, interactivity will be added, and new versions of the bubble funnel will be presented.

7. Part IV: The rethinking phase – further development

This part explains how the results from the evaluation were implemented and how the prototype – the bubble funnel– was *developed further* into a more advanced and useful visualisation.

7.1 Objectives

Drawing on the results of the evaluation and the findings of Part I, the aim of this chapter is first, to refine the prototype, and second, to develop the concept of the bubble funnel.

7.2 Methods

This section describes how the further development of the prototype was carried out. In particular, it will focus on the live view and the simulation view. First, it will describe how the sketches guiding further development were created. Second, it will describe the technical implementation and the tools used – both for further development and to turn the bubble funnel into a more sustainable visualisation product. Third, it will describe how the mock-ups providing additional ideas were created. This arrangement – i.e. the inclusion of the mock-ups – is a result of the limited timeframe of this project. Therefore, some features are implemented (i.e. programmed to work efficiently), while others are described in the mock-up section. Combining the developed state of the bubble funnel – both the live view and the simulation view – and the mock-ups compose the result of this thesis.

7.2.1 Simple sketches

New 'lo-fi' prototypes were sketched drawing on the prototype evaluated in the previous chapter (Benyon, 2019). They are, once again, sketched with a black marker on a piece of paper. The sketches aim to move further towards fulfilling the information-seeking mantra, that is, targeting filter-and-zoom features and add details-on-demand (Shneiderman, 1996). Moreover, Don Norman (2013) highlights that a powerful tool for a designer when designing interactivity is to add constraints. Basically, that refers to limit the interactivity to make a product more useable and not give the user too many options which may cause confusion and frustration. This combined with the timeframe of this project motivate why only some interactive features are included.

In total, one sketch for the live view and one for the simulation view were produced. In addition, two sketches for ideas that could be included in the bubble funnel were sketched. In total, four sketches were produced. Each sketch is accompanied by an idea and why it and its features are suitable according to the literature.

7.2.2 Implementation of the final versions

Here, *final versions* mean the final versions of the simulation view and the live view. In this part, the programming of the visualisation was done more properly. The prototype was re-programmed into an object-oriented style (A Dictionary of Computer Science, 2016). In practical terms, this entails that the bubble funnel is interpreted as an object – if the object is changed so is the bubble funnel. This approach is handy to simplify the understanding of the code, the implementation of the bubble funnel elsewhere, the re-use of code, and to enable adding new features to the bubble funnel.

The development was carried out with common web technologies such as HTML, CSS, and JavaScript. JavaScript libraries used include D3, JQuery and Bootstrap. The library of D3 was used for the visualisation to create transitions, interactivity and motion of the bubbles. D3 makes use of SVG, HTML and CSS to create data-driven documents for the web (D3, 2019). JQuery was used to simplify Javascript code, especially in terms of interactivity (JQuery, 2019). Bootstrap, a front-end library, was used for layout, style and to include interactive elements (Bootstrap, 2019) To fetch data, Python, Pandas and PostgreSQL were used (Python, 2019; Pandas, 2019; PostgreSQL, 2019).

Regarding design, the final versions were produced according to the simple sketches. But some additional elements and features were added drawing on the findings of the pre-study (see 4.4 Conclusion: the needs of Hedvig AB) and the evaluation sessions (see 6.4 Conclusion: a concept with potential). Design choices originated both from those findings and the literature, mainly Shneiderman (1996), Ware (2008) and Tufte (2001; 2006). Regarding interactivity, the final versions of the bubble funnel are developed to respond to user input instantly. Since, *feedback* of a product is important in order to perform well in terms of interactivity (Norman, 2013). Once again, elements added that have not been motivated previously are motivated. Regarding beauty, the designer aimed to develop a beautiful visualisation according to the device that beauty cannot be disadvantageous, especially not when other parts of the literature are considered too.

7.2.3 Mock-ups for improvement

In addition to the final versions, some mock-ups were developed to show how the bubble funnel – in particular, the live view and the simulation view – could be improved and developed further to show on variations of the concept. These mock-ups originate from the simple sketches and were produced with D3 and JavaScript, and, Keynote (Keynote, 2019). Regarding Keynote, it was used to merge different print screens and images produced with D3 and JavaScript. These mock-ups could be seen as a mix between 'hi-fi' and 'lo-fi' prototypes since they are so similar to the final versions, yet not fully implemented (Benyon, 2019). These mock-ups target Hedvig AB's needs and desired insights and show concepts of improvement. Moreover, the design choices are based on the literature.

7.3 Results

The results of the development phase are presented below.

7.3.1 Simple sketches

Four sketches with ideas on further development were sketched: one for the live view, one for the simulation view, and two for ideas on improvement. These are described below.

The live view

The live view of the bubble funnel is sketched, as before, with a horizontal direction of the funnel, and information underneath each cluster (see Figure 12). Bubbles are entering from the left, yet how they are doing that, remains undecided according to this sketch. The idea is to continue to keep the live view simple to avoid 'chartjunk' (Tufte, 2001) and to keep the overview-first style (Shneiderman, 1996). The latter is especially important since the live view could be useful in terms of monitoring the inflow of users in real time, such as having it displayed on a monitor in Hedvig AB's office.

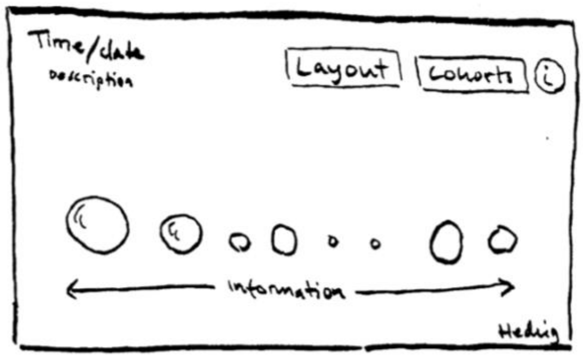


Figure 12. The simple sketch of the live view.

This time interactivity is included. First, a selection drop-down menu called 'Layout' is included to allow the live view to change the layout. Layout changes could, for instance, include changing the information underneath the funnel (e.g. change distribution rate to conversion rate, as was requested by some informants in the evaluation), to show multiple bubble funnels (this option will be discussed later in the mock-up section and was requested by one informant) or to change the steps of the funnel (also requested by one informant). To change the layout is a way to target the different needs of insights desired among employees at Hedvig AB. Second, another selection-drop-down menu called 'Cohorts' is included. If a cohort is selected, bubbles fulfilling the cohort's characteristic is then coloured. The idea is to make these bubbles to 'pop-out' from its surrounding (Ware, 2008). Both these selection options could be seen as zoom-and-filter functions while the information provided underneath each cluster could be interpreted as details-on-demand (Shneiderman, 1996). Lastly, an information symbol: ⓘ is added to provide information about the bubble funnel. Information about, for instance, how it works, the data, the creator, why it was created, and whether it contains any issues – thus, to provide documentation as an excellent chart should do (Tufte, 2006). When clicked, a modal (i.e. a pop-up window) will appear on the screen showing this information. All these interactive elements are displayed in the top of the interface – almost clustered together – to imply that they constitute a setting's area.

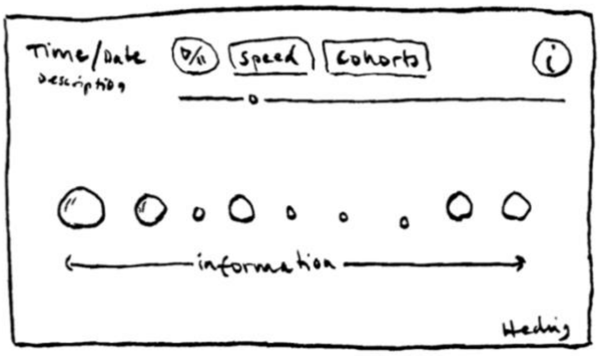


Figure 13. The simple sketch of the simulation view.

The simulation view

The simulation view is similar to the live view in regard to the location of the bubble funnel and the setting's area (see Figure 13). The placement of the setting's area in the top of the screen is to distinguish it from the funnel, i.e. use a spatial grouping of the settings to clearly separate them from each other (Ware, 2008). In contrast to the live view, the simulation view is developed to be used on a browser on a computer to explore the data. Therefore, more settings and interactivity are included. Yet, without creating a messy interface, losing the overview-first view (Shneiderman, 1996).

Interactivity includes a start-and-stop button (sketched as a play-and-pause button, see Figure 13), a speed-selection button, a cohort selection, an information button, and a time slider. The date and time are displayed in the same way as previously. The information button and the cohort selection are similar to the live view, presented above. The time slider, on the other hand, is new. When dragging in the time-slider marker the time is changed accordingly, and when released, the bubbles will instantly move to their funnel step of the selected time. This will allow the user of the simulation view to go back and forth in time easily. The play-and-pause button, as well as the speed-selection button, will allow the analyst to replay a day or parts of a day in order to explore the flow. The cohort selection will be implemented as described for the live view. Once again, all these features provide filter-and-zoom abilities and details on demand (Shneiderman, 1996). The concept of simulation also allows a 'historical' aspect of the data, moving further down in Shneiderman's (1996) taxonomy of tasks an advanced visualisation preferably should contain.

Idea sketches
In addition to the sketches of the live view and the simulation view, two ideas on how to 'lift the business value' and to face the aspect of cohorts were produced. This – to address cohorts –originates from the evaluation and it was pointed out as a major issue (i.e. idea on improvement).

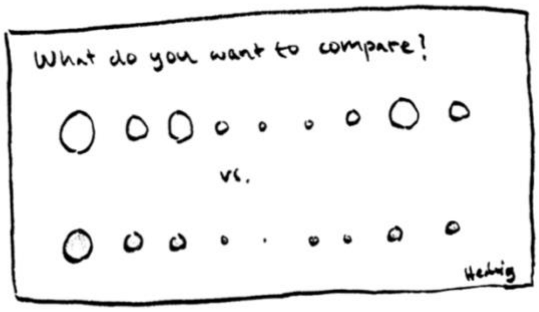


Figure 14. The comparison view.

The first sketch simply shows two bubble funnels, each for a different cohort, on top of each other to allow comparison, contrasts and to spot differences – a comparison view (see Figure 14). The comparison view could be implemented in either the live view, the simulation view or both – this is partly why the 'layout' selection is included in the sketch of the live view. In particular, this idea is in line with Tufte's first principle of comparisons (2006) allowing the pop-out channels of shape (size of the clusters) and motion (activity in the funnel) to be utilised (Ware, 2008).

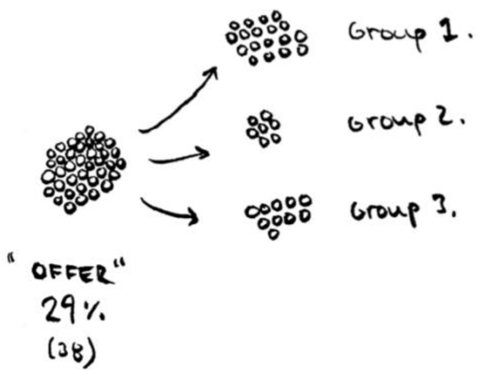


Figure 15. the division of a cluster into smaller clusters.

The second sketch is showing how a cluster of bubbles, in this case, the ‘Offer’ cluster in a bubble funnel could be re-clustered into three small clusters on top of each other (see Figure 15). The idea is to either separate bubbles (users) with different characteristics or features, such as price level of the offer as one informant suggested. This second idea is, as with the first idea sketch, suitable according to Tufte’s first principle (2006) and could also be implemented through the ‘layout’ selection included in the sketch of the live view.

7.3.2 Implementation of the final version

This section describes the live view and the simulation view in their final shapes. They are presented as ‘final’, yet much could be improved and expanded but they are ‘final’ due to the time frame of this project. The selection of steps, both in the live view and the simulation view, originates from the bar chart Hedvig AB uses (see 4.3.2 Analysis of the current visualisation).

The live view
The live view was developed with the tools presented above and according to the simple sketch (see Figure 16). Differences from the previous version include the following. First, the description underneath time and the date is changed into: ‘The current state of Hedvig’s app onboarding. One bubble represents one user’, in order to shorten the text and simplify to make it easier to read, since some informants missed to read it. Second, the background is white instead of smoky white to increase the contrasts between bubbles and background, text elements and background, and interactive elements and background. This is to increase the difference in luminance between objects on the interface and the background which will make these objects more visible (Ware, 2008). Another difference is the light grey path from the left outskirt (where the former starting cluster was located) heading to ‘First app open’ cluster branching towards a ‘Clicked logged-in’ in the one direction and the funnel steps in the other direction. This was implemented in order to solve the confusion among the informants on what constituted the funnel and what was not, but also to solve the problem of the starting point of the bubbles.

Yet another difference is the location and the naming of the former ‘Logged in’ state versus the current ‘Clicked logged-in’ and its location. The name symbolises that the users (bubbles) only have clicked logged in, and do not represents all people who are currently logged in at Hedvig AB’s application. In addition, a small path from the initial cluster, ‘First app open’, is provided to show that the ‘Clicked logged-in’ is outside of the actual funnel. This is also the reason why there is no distribution percentage rate underneath neither the ‘Clicked logged-in’ nor the ‘First app open’, only an amount of the total number of bubbles is displayed.

Figure 16. The live view, final version. In this figure, 'Switcher' is selected as cohort.

Colours were selected to distinguish the bubbles from their surroundings, i.e. the white background and adjacent clusters (see Figure 14). The colour of a 'normal' bubble is grey. In the basic view, i.e. without a cohort or layout selection, the bubbles are grey since there was no criticism against that in the evaluation sessions (nor against black). Cohort colours are orange, dark red, blue, purple, pink and green. The transition between colours (cohorts) has a duration of two seconds to make the transition more aesthetic. Regarding the bubble colours as well as the 'About' button's colour, dark grey – all colours are chosen to be easily to distinguishable from its surrounding (Ware, 2008). The cohorts 'students' and 'switchers' were selected since they were explicitly asked for in the evaluation. 'Clicked OB insurance' (where OB stands for Object Insurance), 'Clicked login' and 'Clicked house' were added since they are suitable to just provide an idea of how popular these features are. See Figure 17 for the cohort selection and see Figure 18 for an example when 'student' is selected.

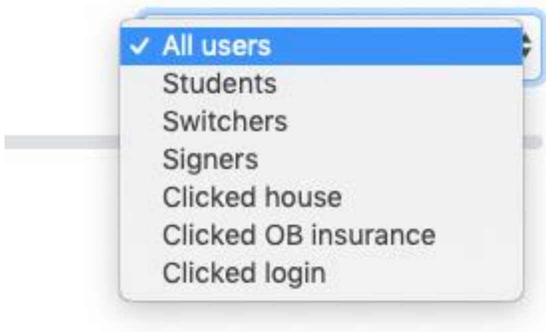


Figure 17. Cohort selection.

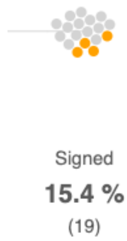


Figure 18. ‘Student’ is selected.

In the live view, the bubbles are ‘popping’ (as was shown in Figure 11), creating padding towards adjacent bubbles when moving to generate a more engaging interface, since the live view is quieter than the simulation view. This is to trigger the pop-out channel of motion (Ware, 2008).

In addition, minor things that should be highlighted includes the following. First, the ‘About’ button with the modal appears and casts a shadow over the screen, except the modal. Second, the bubbles could be clicked, and then a reddish border is added on the bubble to be able to keep track of a bubble of interest. Third, to add more details-on-demand a tooltip is added and appears while one hovers over a cluster (Shneiderman, 1996). The information here depends on the cluster since the data is collected along the user funnel (see Figure 19).

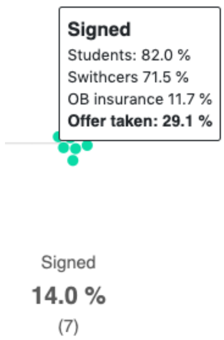


Figure 19. The tooltip feature. The information is varying slightly depending on the cluster. For instance, in ‘Signed’ this information is available.

Technically, the data structure was constructed to allow the data to be updated constantly; both allowing new data points to enter the visualisation – in the shape of bubbles – and allowing old data points, not present in the data set disappear in silence. This was solved with D3’s update pattern which utilises a number of D3 methods (Bostock, 2019 B). The data is fetched every fifth second by the Python script, dumped in a CSV file, read by D3 and managed with JavaScript. Another technical challenge was to create a clever data structure as is pointed out by Shneiderman (1996) as a challenge when developing efficient data visualisations.

Unfortunately, due to the time frame of this project, the ‘Layout’ selection is not responding to any events or actions. However, what would have had happened will be presented in the mock-up section and discussed in the conclusion of this chapter.

The simulation view

Once again, as previously, the simulation view is similar to the live view in most aspects (see Figure 18). Yet, differences include the following.

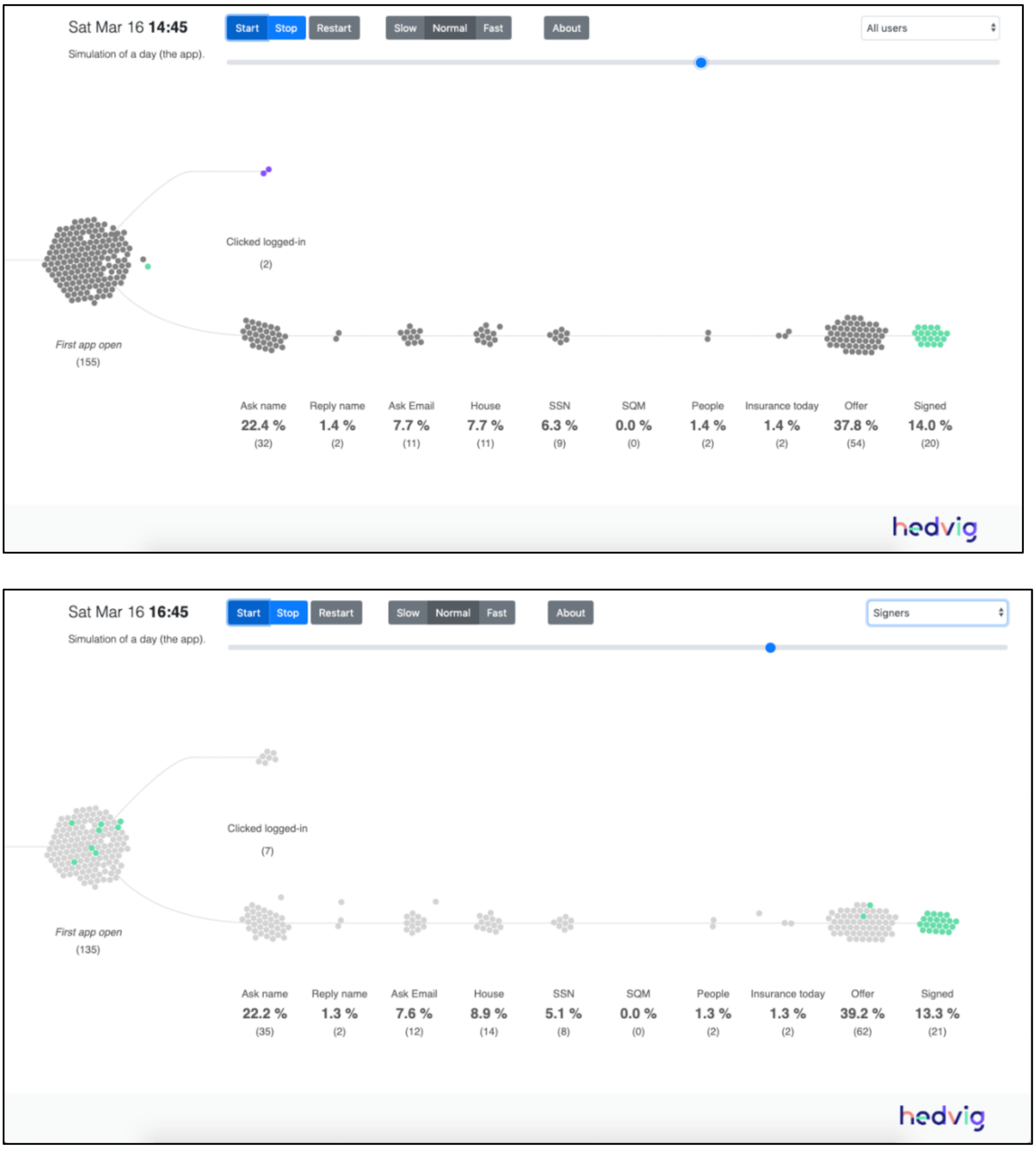


Figure 20. The final version of the simulation view. One view (above) show the default mode, and one view (below) show how ‘signers’ can be tracked in the simulation view.

The data is not fetched, it is simply loaded from a CSV file. On start, all the bubbles appear at ‘First app open’ and the time shows ‘00:00’. Underneath the time, ‘Simulation of a day (in the app)’ is written as a description to ease the understanding. Besides the time, date and description the setting's area is located. In contrast to the simple sketch, a 'Restart' button and a time-slider is added to allow a 'history' aspect of the data visualisation (Shneiderman, 1996). Regarding the button colours, 'Start' and 'Stop' is coloured blue to map to the marker of the time slider that is also blue – i.e. tell the users that they are interlinked. When 'Start' is pressed the time starts ticking, the time slider starts moving rightwards and the bubbles start moving. On 'Stop' all these features also respond and stop. This also applies to when the time-slider is dragged and released, everything changes accordingly, and the bubbles move to their expected cluster dependent on the time selected with the slider. There is a so-called radio button included in the setting's area to allow different speed sections of the simulation. All these features have instant feedback to increase usability (Norman, 2013).

Regarding the cohort selection, everything is similar to the live view except one detail – 'Signers' can be selected since the simulation view is displaying historical data. Hence, one can know whether a person signed or not, and therefore it is possible to keep track of these bubbles and see their behaviour along the funnel (see Figure 20).

In addition, the 'popping' and 'bumping' among the bubbles in this view is restricted. If they are allowed, bubbles are everywhere, especially when the time is adjusted with the time slider and multiple bubbles are aiming for new destinations simultaneously. Therefore, bubbles are only allowed to bump into bubbles from their own, current cluster. This is why there are 'empty holes' in the 'First app open' in Figure 18, those holes were previously filled with bubbles that have left for the funnel. Each bubble then moved 'over' the other bubbles without 'bumping' them. This is to limit the pop-out channel motion slightly (Ware, 2008).

7.3.3 Mock-ups for improvement

Additionally, here are some important expansions of the concept that need to be paid attention to.

Two versions of the newly introduced comparison view were developed. The first one is a comparison view implemented as a simulation showing the difference between two days, the 15th of March and the 16th of March. At '8:27' in the morning, it is a large difference in activity in the different funnels (see Figure 21, above). Later, at ”24:00” the total distribution of the funnel for each day is shown (see Figure 21, below). Comparisons between the cluster sizes, the distribution rate and the actual amount are displayed. The comparison views showed in Figure 21 are developed with D3 and JavaScript.

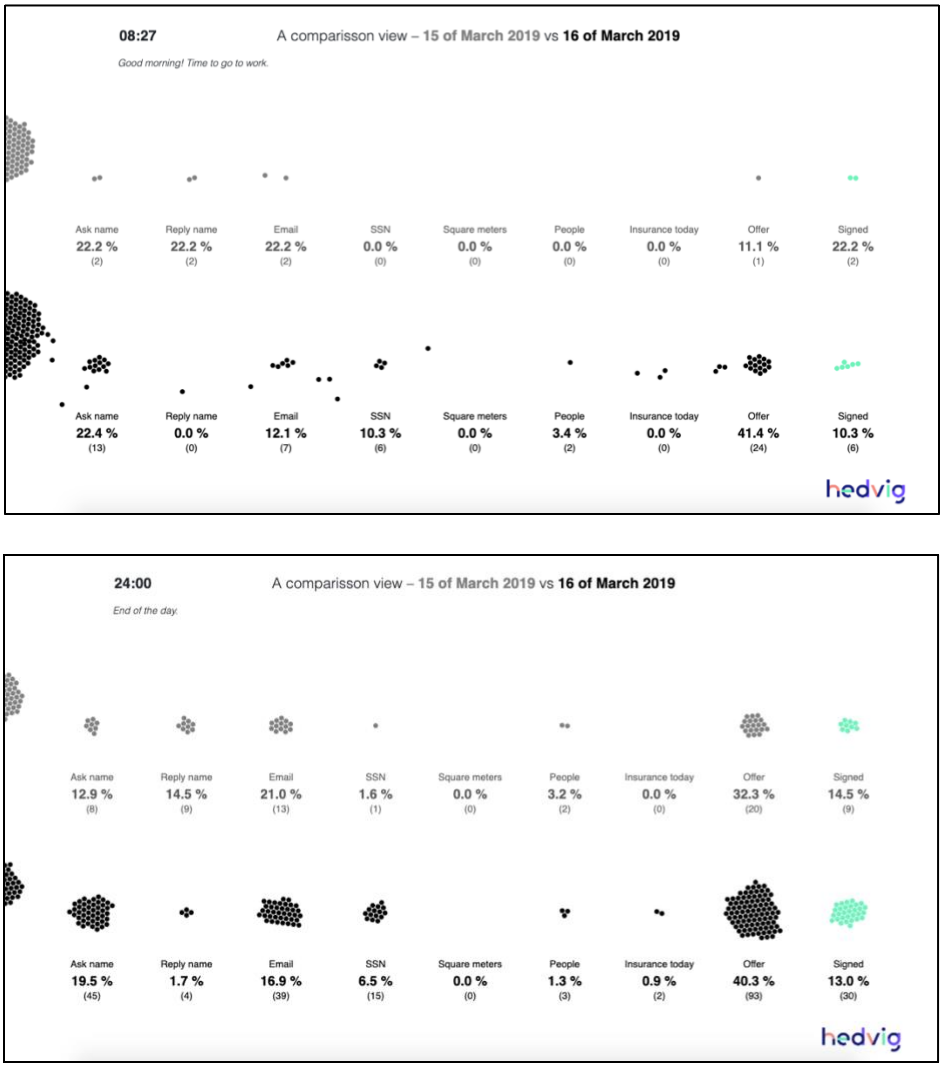


Figure 21. The comparison view – 8:27 in the morning (above) and 24:00 in the night (below.)

The second version of the comparison view is a pure mock-up. It shows the same concept, but instead of being implemented in the simulation view, it is implemented in the live view as 'Layout two'. The idea is to show the live inflow of users distinguished into separated funnels reliant on each user's UTM source before entering Hedvig AB's chat, such as Facebook, Snapchat or where else Hedvig AB's run ads (see Figure 22). This is, however, only one idea. This view could display multiple bubble funnels and could thus target days, cohort features, etcetera. The main idea is to take advantage of the fact that this view shows comparisons, contrast and differences in line with Tufte's first principle (2006).

Regarding the sketch dividing a cluster. As with the comparison view, this could also be included as a 'Layout' in the live view: 'Layout three'. The idea is again to show comparisons (Tufte, 2006). Figure 23 shows one idea of this – split up the 'Offer' cluster into three clusters for three price intervals. At this point in the funnel, pretty much is known about the users which means that 'price level' could be swapped to something else holding a business value. Regarding design, a path (as used in the live view or the simulation view) would beneficially be added to tell the viewer that the funnel is branching in three.

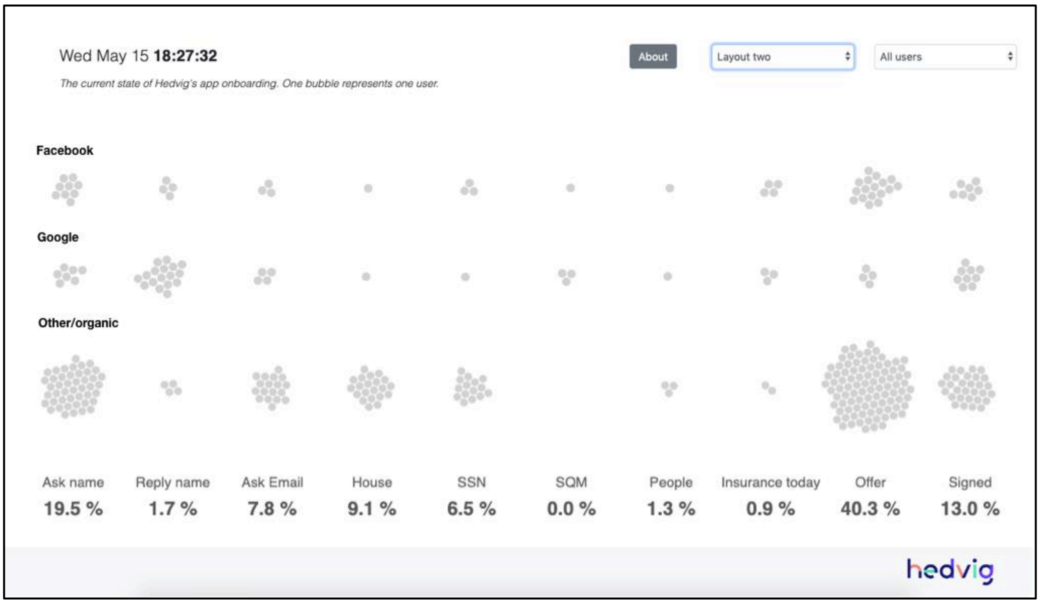


Figure 22. An example of the live view with a comparison view.

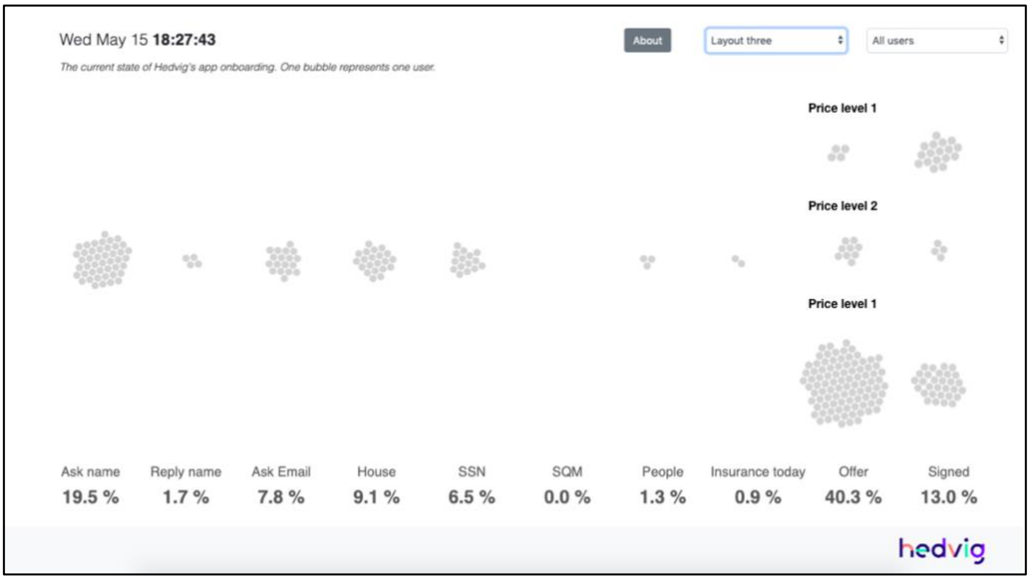


Figure 23. An example of how the offer and sign can be divided into multiple clusters

7.4 Conclusion: a whole new bubble game

This chapter has provided the final versions of the bubble funnel – the simulation view and the live view – and some mock-ups developing the concept further. Conclusively, the code has been sophisticated, interactivity has been included and the design has been refined. This conclusion discusses the bubble funnel in regard to the needs of Hedvig AB and the issues found in the previous chapter.

In regard to the results of the evaluation, the issues of the prototypes have been solved (see above for details). The bubble funnel is now present in a more clean, self-explaining interface with light grey paths showing branches. The need for lifting the 'business value' has been addressed by adding zoom-and-filter functions and details-on-demand (Shneiderman, 1996), such as tooltip, time slider, selection of layout and cohorts. The comparison view lifts the business value to a large extent due to its clear comparisons. Yet, to lift the 'business value' even more could be addressed with sophisticated data analysis techniques, more variables for multivariate analysis, and inclusion of conversion rate numbers between steps (the overall conversion rate is available in the dashboard created in Part I). But, on the other hand, the conversion rate is already addressed well in the dashboard.

In regard to the needs of Hedvig AB, the themes found in the pre-study (see 4. Part I: The investigative phase – discover and define) are addressed quite well. Practically, the conversion could be visually seen; user drop-offs are clear; cohorts are addressed through the comparison view, the division-of-a-cluster view and through colours; and the bubble funnel provides a clear overview of the user funnel in one interface. Time is only addressed partly. One can see a bubble go from question to question (both in the simulation view and the live view) and thus get a 'feeling' for the time it takes to answer a question or go through the flow. Yet, information linked to time could be added in the tooltip or underneath each step in the funnel.

Different pathways are not addressed properly with this visualisation – only one branch exists (one branch for login, away from the funnel). This could be addressed by expanding the number of clusters and making more detailed pathways and branches (i.e. adding more light grey lines). A more detailed funnel results in a more accurate data representation which depends on the accuracy of details one aspires. This could be addressed by, for example, having simple, medium and advanced versions of the bubble funnel available through the 'Layout' selection.

Noteworthy, an issue with the bubble funnel is that users are grouped in clusters depending on their latest tracking-point available in the funnel. For instance, if the funnel step 'House' is removed from the funnel, these bubbles will cluster with the 'Email' bubbles instead. This is a risk of this visualisation and needs to be highlighted in the visualisation's documentation (Tufte, 2006), in this case within the 'About' modal. Otherwise, decision making drawing on dishonest data may occur, which would be devastating. The level of details should depend on the purpose of the visualisation.

In addition, a selection option of date and time should preferably be provided in the simulation view. This could be done with a button beside time and date. For instance, when clicked it could show a drop-down container showing a calendar with selectable dates.

Conclusively, the current state of the bubble funnel is targeting the needs of Hedvig AB but could be expanded to address them even better. Some implementation remains, but the concept is developed to a large extent. Compared to the visualisation Hedvig AB is using today – the bar chart – this is a whole new bubble game. However, how the concept and the current state of the bubble funnel are performing in regard to the literature remains uncharted. That will happen in the next chapter.

8. Analysis & discussion

This chapter will provide an analysis of the visualisation concept developed – the bubble funnel – according to the selected literature. Each subtitle is mapping to a section in the literature providing a different perspective on the bubble. Noteworthy, this analysis will mainly focus on the concept of a bubble funnel since that is the main contribution of this thesis.

8.1 Designed for interactivity

As mentioned previously, due to the timeframe of this project the interactivity is limited. The interactivity added, however, show the wide range of possibilities and opportunities of the concept. It also shows how the concept can react instantly to user input and, hence, provide appropriate feedback (Norman, 2013).

What is beautiful or aesthetic is a personal opinion, perceived by people. Tractinsky et al. (2000) used layouts that were rated as aesthetic by participants in a previous study by Tractinsky (1997) to conclude that perceived aesthetics matter for usability. Yet, these results are questioned (Tuch et al., 2012). Still, in this project, the designer aimed to develop a beautiful visualisation according to the device that beauty cannot be disadvantageous. During the evaluation of the early prototype (Part III), four out of five informants perceived the bubbles funnel as appealing – using words as: 'cute', 'nice', 'beautiful' – which indicates that it is aesthetic, and hopefully, slightly more usable. In fact, a majority of the informants also answered correctly to the task-based questions, whether that is a result of the beauty, however, remains unknown. Therefore, this study cannot comment on the debate on beauty and usability. But what is clear is that beauty and aesthetics – in the shape of bubble movement and clustering – seems to generate interest among users, at least at first sight. However, this may also be due to the pop-out channel of motion that the moving bubbles are triggering (which will be discussed more in the next section) (Ware, 2008).

In addition, the transition of bubble colours, the movement of bubbles and the clustering of bubbles are especially appealing. Yet, the beauty of these features, especially movement, may decrease if there is a very large number of bubbles. This could be solved by aggregating the data. I.e. one bubble could represent five, fifty or five hundred users depending on the amount of the inflow of users, and hence the size of the digital service provider using the bubble funnel. Such a solution would also show that the bubble funnel is scalable and could be developed to manage large amounts of data.

Regarding the information-seeking mantra Shneiderman (1996), the visualisation succeeds in creating a clear overview first – the user funnel is clearly visible with its bubbles and clusters (or at least if the user funnel is not empty). Thereafter, filter-and-zoom features are allowed through interactive elements (e.g. the time slider in the simulation view and the drop-down-cohort selection). Details-on-demand are provided in the shape of visible information (e.g. numbers underneath each cluster) and through interactivity (e.g. the tooltip). Moving further, looking at the remaining task types in Shneiderman's taxonomy (1996), relationships between items is, for instance, displayed with the paths between the clusters. To show history is done through interactive elements (e.g. the restart button and the time slider allowing the user to go back and forth in time). Lastly, the extraction of data is not explicitly addressed. However, since the data is visual, a print screen of the funnel would 'extract' the data and enable it to be inserted in a report or published in any medium allowing images. Yet, this feature could also be enabled through buttons or interactive elements.

All these interactive features are only a few ideas originating from the qualitative work of this study. Thus, each task type – in particular the ones further down in Shneiderman's taxonomy – would have been addressed more properly if time was unlimited. For instance, the settings area above the bubble funnel could be expanded with more settings, options and interactivity. Yet, Shneiderman's ideas (1996) have definitely shaped this work and all the task types he suggests are included, at least to some extent, and seem to work well. His mantra and his taxonomy seem to still be relevant despite the fact that they were introduced in 1996, especially since the task types are broad and the taxonomy presents a logical, incremental way to develop an interactive visualisation designed for humans.

8.2 Designed for human perception

Since the bubble funnel is a new concept, people lack experience of using it and has to learn it – it is not like a bar chart. This was visible during the evaluation sessions; one informant stated '[the bubble funnel is] a much more qualitative variation of a funnel than any I have ever used or seen.' And two out of five informants were slightly wrong when they were asked to describe the prototype presented in the evaluation, while three were correct. Thereafter, it has been developed further to simplify visual queries and to allow cognitive tasks to be solved – i.e. retrieving information. In practical terms, this was done through the adoption of more concepts from the literature, mainly Shneiderman (1996) and Tufte (2001; 2006), and through addressing what humans can easily see, based on Ware's work (2008). This has increased its suitability to solve cognitive tasks (compare the final versions with the prototype evaluated). Despite this, the major, upcoming challenge for the bubble funnel is to be adopted and familiarised among its users – both at Hedvig AB and elsewhere. Hopefully, the bubble funnel will be as intuitive as a subway-system map in the future.

In regard to the basic pop-out channels – colour, shape and texture, motion, and spatial grouping – the bubble funnel successfully takes advantage of each channel (Ware, 2008). Colours and contrast in luminance distinguish controls (such as buttons and selection fields), bubbles and clusters from the interface's white background colour. This also applies to a selection of a cohort; bubbles fulfilling the cohort requirement (e.g. 'student') are coloured while the remaining bubbles turn from grey into light grey, yet visible. Regarding shape and texture, the clustering of the bubbles creates a symmetric pattern, almost acting as a texture, which informant 1 emphasised that he liked. Shape, on the other hand, is used to distinguish the size of clusters which is enabled since all bubbles have the same size – every cluster creates an area with a certain size allowing comparison that works for human perception. The comparison between clusters also works well since there is a spatial-grouping aspect, for example, the different versions of the bubble funnel available in 7.3 Results show that small and large clusters are efficiently separable both due to the amount and the spatial distance between them. This is probably why the cognitive tasks of comparing cluster sizes was such an easy task to solve for the informants. In addition, spatial grouping in terms of padding works well too, see Figure 11 for example.

The last pop-out channel is motion (Ware, 2008). Motion was used for the movement of bubbles – and it was indeed an eye-catcher. As one informant mentioned during an evaluation session: 'it is happening stuff, so it is not boring to watch' while other informants used words such as 'fun' and 'nice' while asked to think-out-loud. Thus, the movement is both visible and appreciated – it creates visual engagement among its users. It works indeed, but perhaps too well. In the simulation view, a movement overload can occur if multiple bubbles are moving to multiple clusters simultaneously, bumping into each other and 'popping' from clusters. Preferably, if there is a large number of bubbles this feature should be turned off (as in the simulation view) to allow a visualisation with motion, yet not chaotic motion. This is to enable the data to be in focus instead of the motion. As Tufte (2001) suggests, the meaning of the data should not be overshadowed by design – in this case, by motion. However, much motion means a lot of activity – which indeed is an aspect of a data set. Hence, motion has to be added carefully.

8.3 Designed for analytics

First of all, an excellent chart should not contain 'chartjunk' since that may hinder the quick understanding of the visualisation (Tufte, 2001). Due to the fact that bubbles build the bubble funnel, that is easily achieved. No bubbles equal, almost, a blank sheet. And in the case of Hedvig AB, all elements, paths, shapes, buttons, etcetera, fill a function.

What about the bubble funnel and Tufte's fundamental principles of analytical design (2006), well, here are some thoughts.

Principle 1 – Comparisons
The bubble funnel succeeds well in comparison, contrasts and differences in many aspects. First, comparisons between clusters in a bubble funnel work well, both according to the evaluation sessions and the literature on perception (Ware 2008). Second, comparisons between multiple bubble funnels, as in the comparison view, has unfortunately not been evaluated on informants but should also work well according to the literature since it is similar to the comparison between clusters. Third, a comparison between cohorts when bubbles are coloured works well but is slightly more difficult

since the coloured bubbles and the light-grey bubbles are mixed together within clusters. If the bubbles were sorted according to cohorts within a cluster, then it would be easier. Or if they resorted on the selection of a cohort. Fourth, if the simulation view is merged with the comparison view (see section 7.3.3 Mock-ups for improvement), a difference in activity between the bubble funnels could be seen if the data is different.

Principle 2 – Causality, mechanism, structure and explanation
This principle is, as Tufte highlights (2006), trickier to address since human activities are of complex nature and tied to uncertainty. Yet, in one way this visualisation succeeds. Partly since the visualisation, the bubble funnel is nothing but users and their behaviour in (real) time. Partly since the visualisation offers direct insight into the funnel steps, such as 'Ask name', 'Email', 'SSN' (Social Security Number) and 'Offer'. All these steps require something from the user who enters Hedvig AB's onboarding chat and hence gives a clue to causality. 'Ask name', 'Email' and 'SSN' ask for personal information and 'Offer' asks for a sign-up, i.e. money. Hence, if these clusters are growing or are becoming 'unexpectedly' large – well, then this visualisation can indicate what may bother the users (i.e. potential customers) in the user funnel provide some sort of explanation. Thus, the bubble funnel's ability to provide a causality highly depends on the impact and meaning of the funnel steps.

Principle 3 – Multivariate analysis
Multiple variables could be provided with this visualisation. In the case of Hedvig AB, distribution of users, user characteristics, time, and selected answers are included to some extent. Hence, whether a multivariate analysis is possible depends on the data that is explored with the bubble funnel. The rule should be: the more user data features that are joined onto the tracking data, the more multivariate analysis is available (if implemented). Such information could then be added in the tooltip, as cohort selections or as additional information underneath funnel steps. Several implementations are possible.

Principle 4 – Integration of evidence
Regarding integration of evidence, this principle is adopted. For instance, underneath each cluster, a name, a distribution per cent and a total amount of bubbles are provided to increase the understanding. Time and date and a simple description are also provided. To increase the display of further evidence and allow further integration, interactive elements such as buttons, selections and tooltips are provided. These are only some examples of how integration could be included in the bubble funnel. Obviously, the bubble funnel could be endlessly customised to fit different needs, activities, contexts and people. Hence, principle three and principle four allow the bubble funnel to be highly customised because variables and evidence could be added, respectively. In the case study at Hedvig AB, only some versions and ideas are shown.

Principle 5 – Documentation
This principle is important and is in the case of Hedvig AB solved with an 'About' button providing an information modal. The information modal is containing useful

information, yet this could be achieved in other ways too. For instance, visible descriptions in the interface without making the interface messy could make vital information more accessible. In addition, this feature depends on whether the users of the visualisation are using it frequently and are familiar with it. If that is the case, they can be taught how to use it, potential flaws and additional information that is good to know. Regarding titles, readable funnel names is a must. Such as ‘Offer’ and ‘Signed’ instead of ‘message.forslag2’ and ‘message.kontraktklar.ss’.

Principle 6 – Content counts most of all
A visualisation is driven by the content. The content, i.e. the data, is the fuel for this visualisation – no data, no bubbles. In the case of Hedvig AB, the data is of good quality since it is directly fetched from Hedvig AB’s data warehouse and the data is pure tracking data which makes it suitable. Importantly, if there are too many steps in the user funnel, the bubble funnel may be a bad selection of visualisation since there will not be enough horizontal space and clusters needs to be grouped. In the case of Hedvig AB (partly discussed in 7.4 Conclusion: a whole new bubble game), a few bubbles cluster at steps that are closest to their real step of presence, since their real step of presence is not available in the visualisation. On the one hand, this is problematic in terms of the dishonesty of data visualisation, on the other hand, all the important steps are included and the large majority of the bubbles are correctly displayed. One has to ask what the aim of the visualisation is for each particular case – is it to provide a clear overview and communicate patterns and trends of the data or is it to show every little detail available within the tracking data? Regardless, this has to be described in the documentation (see the previous principle). The problem is partly a result of the screen width, the bubble funnels presented in the result (see 7.3 Result) are shown on a small screen but wider screens, obviously, allow longer bubble funnels to be displayed. Otherwise, horizontal scrolling could be implemented. Noteworthy, this is the problem of Hedvig AB’s bar chart too since there are some, but less important, steps in between some of the bars. The problem is also partly a result of multiple pathways within a user funnel, i.e. a non-linear funnel. The bubble funnel could address pathways for non-linear funnels as discussed previously. One could customise it and scale it into a net of branches which is not possible with a bar chart.

Conclusively, the bubble funnel is best suited for linear and shorter funnels (about 10 steps or less) but it could be designed – if desired – to address non-linear funnels to a high level of detail. On the contrary to a bar chart.

To sum up this section, Tufte’s principles of analytical design (2006) are fulfilled to a high extent. However, much could be developed further in terms of analytics, for instance, the business value could be lifted by a number of implementations such as more variables and more integration of evidence. Yet, this is a good foundation for an analytical visualisation. Depending on the context and the data, the bubble funnel could take many different analytical shapes.

8.4 The role of the bubble funnel

In relation to existing funnels on the market, such as Hotjar's funnel visualisation (2019) or Google's (2019) user-flow sankey, the bubble funnel provides less quantitative analysis but more qualitative analysis since a single user can be monitored in real time. Yet it could be developed towards a more quantitative tool as well, such as letting one bubble represent multiple users. This argument also applies in comparison with software like Klipfolio (2019) and Apache Superset (2019). These tools are, indeed, very handy for aggregating large amount of user-tracking data with queries and display it on dashboards. However, the bubble funnel is a suitable complement to these tools, especially since it communicates a clear and intuitive overview of a whole user funnel. In particular, it is a good complement in terms of its ability to monitor real-time data, and thus, could provide instant business feedback (e.g. the performance of a social media campaign as shown in 7.3.3 Mock-ups for improvement), generate interest in the data, and provide a clear overview of the user funnel while also providing details-on-demand. Optimally, the bubble funnel (live view) is displayed on a monitor beside another monitor that show stats and numbers on long-term performance, and thus, these two together provide a comprehensive understanding.

9. Conclusion: the functional art of bubbles

This chapter will provide some concluding remarks and ideas on further research.

9.1 Concluding remarks

Universal selection theory describes the process of change beyond biological evolution with Darwin's notions: variation, selection and retention (Hodgson, 2005). In fact, a design-thinking process is – especially when it is human-centred and contains evaluation – somewhat evolutionary since it is a gradual cycle of development and change. Initial ideas and sketches evolve prototypes; variations are tested, some features retain and are passed on to future versions. Thus, a continual selection process of the most promising prototypes and sketches is taking place. Finally, one version remains: the bubble funnel.

The bubble funnel is a visualisation concept constituting of clustering, distribution and movement of bubbles from A to B with multiple steps in between. The concept is scalable and customisable, and this thesis has shown a number of variations in terms of layout and usage – i.e. simulation view, live view and comparison view – and provided ideas on interactivity to allow more information-retrieval tasks to be carried out. The bubble funnel has been developed to increase the understanding of the inflow of users within Hedvig AB's onboarding chat, addressing the narrow, yet the important business problem of understanding the acquisition of customers in a digital service. Hence, it could be used elsewhere too.

In regard to the research question, the bubble funnel helps to increase the understanding in the following ways. First, it efficiently shows conversion – a key factor for digital service providers – and the movement of users towards conversion. Second, it communicates a clear overview of the whole user funnel in one view. In particular, in regard to the distribution of user drop-offs. Third, it provides three ways of addressing user cohorts; a comparison view, a partition of clusters, and lastly, a colouring of bubbles fulfilling certain user characteristics. Fourth, it could show multiple pathways and address details on a more advanced level if designed accordingly. Importantly, the bubble funnel is a complement to existing tools that can aggregate a large amount of data and show long-term performance. A risk is, otherwise, that the decisions are based on a too narrow scope. In the case of Hedvig AB, the concept of the bubble funnel could be this complement and one of several pieces of evidence of how the onboarding chat is performing, and thus, support data-driven decision making and data-driven UX design.

These improved insights are a consequence of the human-centred approach and the choice of literature. The bubble funnel follows the information-seeking mantra (Shneiderman, 1998) and the fundamental principles of analytical design (Tufte, 2006). It addresses human perception through our basic pop-out channels (Ware, 2008) and it aspires to be aesthetic and usable (Tractinsky et al. 2000). Conclusively, the bubble funnel is a solution to a *wicked problem* and it is one way to understand user funnels better (Rittel and Webber, 1973). To lend Albert Cairo's expression, the bubble funnel is a contribution to the domain of 'functional art' (Cairo, 2013). This is hopefully only the beginning for the bubble funnel and its mission to deliver understanding to companies acquiring users within the digital world of today and tomorrow.

9.2 Further research

The concept of visualising data as a flow of bubbles from A to B, distributed over clusters in order to increase understanding of a phenomenon should be investigated further. For instance: could this concept of data representation be placed in other contexts far from a user funnel? How could this concept be developed further in order to increase its suitability in regard to human perception? Finally, according to the evaluation (Part III) and the analysis above: this is a good way to communicate data. Yet, my analysis is brief and the informants few, therefore, more in-depth research on the usability and suitability of the bubble funnel – as an analytical and communicative tool – would be interesting to investigate further.

References

A Dictionary of Computer Science (2016), 'Object-oriented Programming', 7th Edition, Oxford University Press.

Ashkenas, J. and Parlapiano, A. (2014), 'How the Recession Reshaped the Economy, in 255 Charts', New York Times, https://www.nytimes.com/interactive/2014/06/05/upshot/how-the-recession-reshaped-the-economy-in-255-charts.html [Updated June 6, 2014].

Benyon, D. (2019), Designing user experience: a guide to HCI, UX and interaction design, 4 Edition, Pearson Education, Harlow.

Bostock, M. (2019 A), 'Clustered Force Layout III' https://bl.ocks.org/mbostock/7881887 [Retrieved 2019-03-12, updated April 22, 2019]

Bostock, M. (2019 B), 'General Update Pattern, I' https://bl.ocks.org/mbostock/3808218 [Retrieved 2019-03-11, updated 2019-01-28]

Braun, V. & Clarke, V. (2006), 'Using thematic analysis in psychology', Qualitative Research in Psychology, Vol. 3, No. 2, 77-101.

Cairo, A. (2013). The functional art: an introduction to information graphics and visualizations. New Riders: The United States.

Cooley, M. (1987), 'Human centred systems: An urgent problem for systems designers', AI & Society, Vol. 1, No. 1, 37-46.

D3 Force layout. (2019). 'API References - Layouts - Force Layout', https://github.com/d3/d3-3.x-api-reference/blob/master/Force-Layout.md [Retrieved 2019-05-16].

Evergreen, S. & Metzner, C. (2013), 'Design Principles for Data Visualization in Evaluation', New Directions for Evaluation, Vol. 2013, No. 140, 5-20.

Gapminder (2019), 'Gapminder Tools', https://www.gapminder.org/tools/#$chart-type=bubbles [Retrieved 2019-04-29].

Google (2019), 'Analyze your data with Users Flow: See specific examples of how the Users Flow report can give you insights into your data', https://support.google.com/analytics/answer/1713056 [Retrieved 2019-05-06].

Han, P. (2017), 'A Beginner's Guide to Cohort Analysis: the Most Actionable (and Underrated) Report on Google Analytics', Medium Inc, https://medium.com/analytics-for-humans/a-beginners-guide-to-cohort-analysis-the-most-actionable-and-underrated-report-on-google-c0797d826bf4 [Retrieved 2019-05-5].

Hedvig AB (2019), www.hedvig.com [Retrieved 2019-02-28].

Hedvig AB’s GitHub (2019), https://github.com/HedvigInsurance [Retrieved 2019-01-25].

Hoban, P.R. & Bucklin, R.E. (2015), ‘Effects of Internet Display Advertising in the Purchase Funnel: Model-Based Insights from a Randomized Field Experiment’, Journal of Marketing Research, Vol. 52, No. 3, 375-393.

Hodgson, G.M. (2005), ‘Generalizing Darwinism to Social Evolution: Some Early Attempts’, Journal of Economic Issues, Vol. 39, No. 4, 899-914.

Holtzblatt, K. & Beyer, H. (2015). Contextual design: evolved, Morgan & Claypool Publishers: San Rafael, California.

IDEO (2019), ‘Design thinking defined’. https://designthinking.ideo.com/ [Retrieved 2019-02-28].

Isson, J.P. & Hwang, M. (2018), Unstructured Data Analytics: How to Improve Customer Acquisition, Customer Retention, and Fraud Detection and Prevention, John Wiley & Sons, Incorporated: Hoboken.

Jones, Lyle V (1986), ‘The Collected Works of John W. Tukey: Philosophy and Principles of Data Analysis 1949-1964’, 1st Edition, Vol III. Chapman and Hall; London, UK.

Karpefors, M. & Weatherall, J. (2018), ‘The Tendril Plot – a novel visual summary of the incidence, significance and temporal aspects of adverse events in clinical trials’, Journal of the American Medical Informatics Association (JAMIA), Vol. 25, No. 8, 1069-1073.

Meirelles, I. (2013), Design for information: an introduction to the histories, theories, and best practices behind effective information visualizations, Rockport Publishers: Beverly, MA.

Norman, D.A. (2013), The design of everyday things, Revis. and expanded edition, Basic Books: New York.

Patton, M.Q. (1999), ‘Enhancing the quality and credibility of qualitative analysis’, Health services research, Vol. 34, No. 5, 1189-1208.

Rittel, Horst W. J. & Webber, M.M. (1973), ‘Dilemmas in a General Theory of Planning’, Policy Sciences, Vol. 4, No. 2, 155-169.

Rubin, J. & Chisnell, D. (2008), Handbook of usability testing: how to plan, design, and conduct effective tests, 2nd Edition, Wiley Pub: Indianapolis, IN.

Samsel, F., Bartram, L. and Bares, A. (2018). ‘Art, Affect and Color: Creating Engaging Expressive Scientific Visualization’, conference paper presented at VISAP’18. Berlin, Germany.

Shneiderman, B. (1996), ‘The eyes have it: a task by data type taxonomy for information visualizations’, Proceedings 1996 IEEE Symposium on Visual Languages, 336-343.

Tractinsky, N. (1997), 'Aesthetics and apparent usability: empirically assessing cultural and methodological issues', In CHI'97: Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, 115–122.

Tractinsky, N., Katz, A.S. & Ikar, D. (2000), 'What is beautiful is usable', Interacting with Computers, Vol. 13, No. 2, 127-145.

Tuch, A.N., Roth, S.P., Hornbæk, K., Opwis, K. & Bargas-Avila, J.A. (2012), 'Is beautiful really usable? Toward understanding the relation between usability, aesthetics, and affect in HCI', Computers in Human Behavior, Vol. 28, No. 5, 1596-1607.

Tufte, E.R. (1990), Envisioning information, Graphics Press: Cheshire, Conn.

Tufte, E.R. (2001), The visual display of quantitative information, 2nd edition, Graphics Press: Cheshire, Conn.

Tufte, E.R. (2006), Beautiful evidence, Graphics Press LLC, Cheshire, Conn

Tukey, J.W. (1977), Exploratory data analysis, Addison-Wesley: Reading, Mass.

UK Design Council, (2019), 'Design methods for developing services', PDF report, available at: https://www.designcouncil.org.uk/sites/default/files/asset/document/Design%20methods%20for%20developing%20services.pdf

Ware, C. (2004), Information visualization: perception for design, 2nd edition, Morgan Kaufman: San Francisco, USA.

Ware, C. (2008), Visual thinking for design, Morgan Kaufmann, Burlington, Mass; Amsterdam.

Yau, N. (2019), 'Make a Moving Bubbles Chart to Show Clustering and Distributions', FlowingData: https://flowingdata.com/2016/08/23/make-a-moving-bubbles-chart-to-show-clustering-and-distributions/ [Retrieved 2019-03-06].

Tools and software

Apache Superset (2019), 'Apache Superset (incubating)', https://superset.incubator.apache.org/ [Retrieved 2019-05-24].

Bootstrap (2019), 'Bootstrap'. https://getbootstrap.com/ [Retrieved 2019-05-09].

Dash (2019), 'Build beautiful, web-based analytics applications with Dash', https://plot.ly/products/dash/ [Retrieved 2019-05-12].

D3 (2019), 'Data-Driven Documents'. https://d3js.org/ [Retrieved 2019-05-09].

Hotjar (2019), 'Conversion Funnels' https://www.hotjar.com/tour [Retrieved 2019-05-06].

JavaScript (2019), 'JavaScript' https://developer.mozilla.org/en-US/docs/Web/JavaScript [Retrieved 2019-05-12].

jQuery (2019), ‘What is jQuery?’ https://jquery.com/ [Retrieved 2019-05-09].

Keynote (2019), ‘Beautiful presentations for everyone. By everyone’, https://www.apple.com/keynote/ [Retrieved 2019-05-25].

Klipfolio, (2019), ‘What is Klipfolio?’ https://www.klipfolio.com/features [Retrieved 2019-05-12].

Pandas (2019), ‘Python Data Analysis Library’ https://pandas.pydata.org/ [Retrieved 2019-05-09].

PostgreSQL (2019), ‘PostgreSQL: The World's Most Advanced Open Source Relational Database’, https://www.postgresql.org/ [Retrieved 2019-05-09].

Python (2019), ‘Python’ https://www.python.org/ [Retrieved 2019-05-09].

Sketch (2019), ‘The best products starts with sketch’ https://www.sketch.com [Retrieved 2019-05-16].

Appendix I – Interview questions

Is it okay if I record the interview?

Questions on the introduction to the research project?

What data do you use in your role at Hedvig AB?

From the chat– what information do you want to retrieve?

- What is interesting?
- What information do you lack?

What is the overarching goal of understanding the onboarding chat better?

- How could that help Hedvig AB to turn into a better product?
- How could that provide additional value to Hedvig AB?
- How could that help you?
- Who else could it help?

I have heard that 'get insights' is important. What does insight into the onboarding chat means to you, in your role, at Hedvig AB?

What potential do you see in such a visualisation tool?

- For Hedvig AB?
- For you, your role?
- For other teams?

If possible – please show how you analyse or use data or draw conclusions from data today.

Who, at Hedvig AB, are looking and investigating the onboarding chat data?

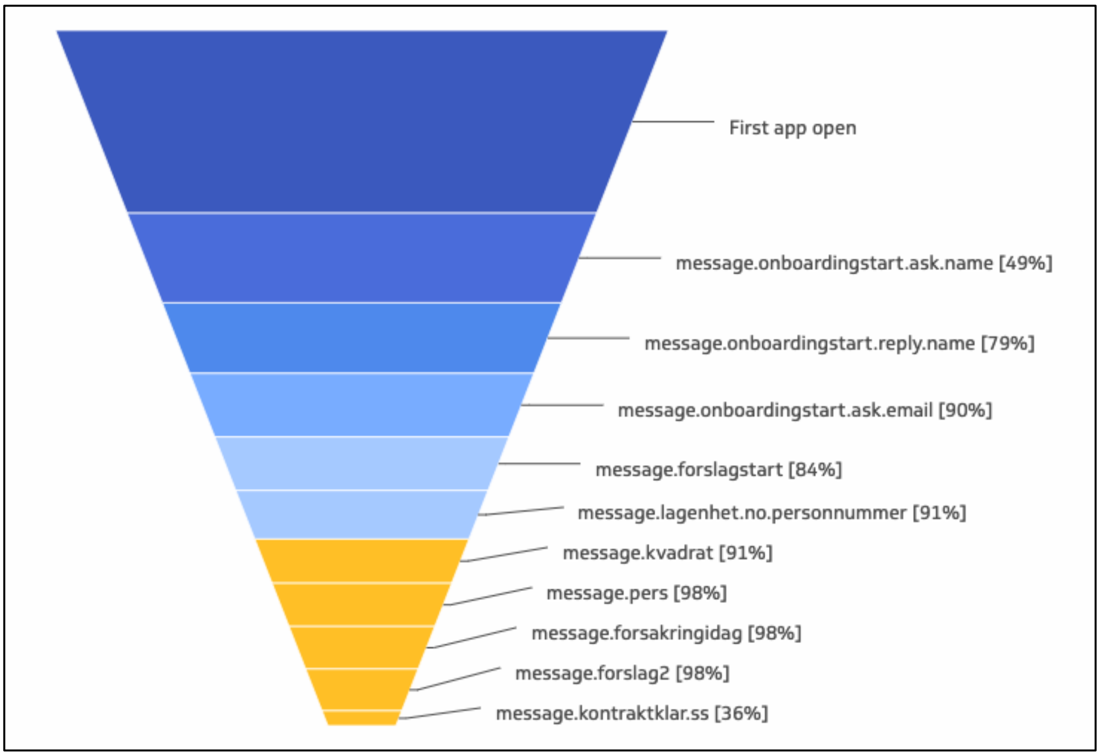
Who else should I speak to?

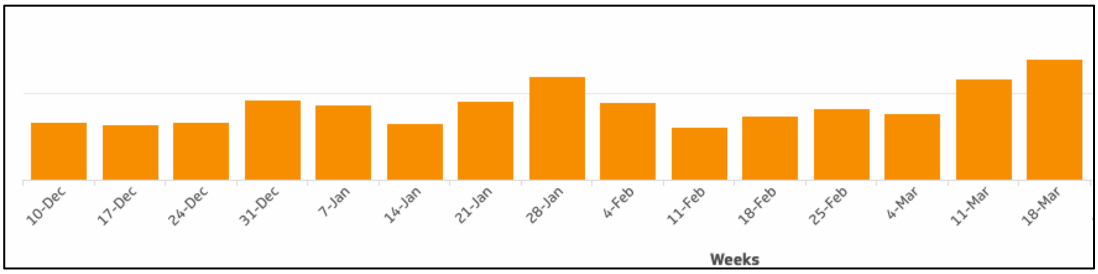
What have I forgotten to ask?

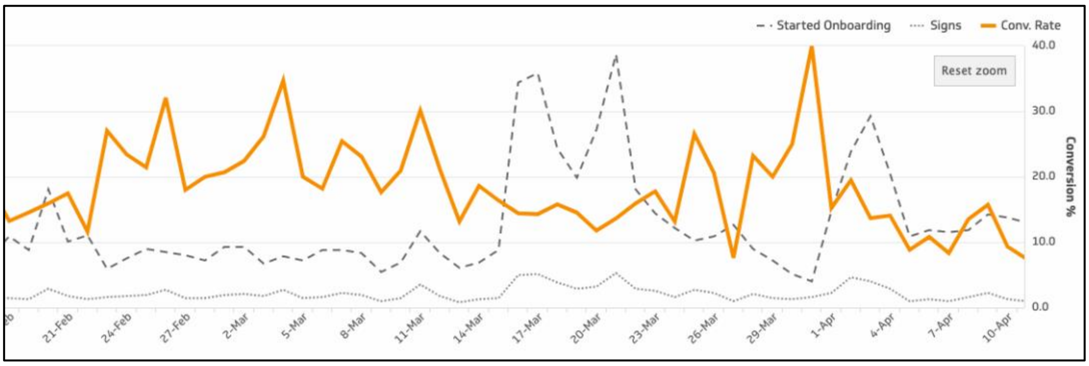
How could such a visualisation prototype be made and what could it look like? Why? Please describe how you would use it?

Appendix II – A glance at the dashboard

Some of the figures included in the dashboard can be seen below. First, a funnel chart displaying the conversion between steps in percent. Second, a bar chart showing amount of weekly signed customers. And last, a line graph showing number of users starting the onboarding chat, signing and an orange line showing the daily conversion rate.







Appendix III – Interview questions, evaluation

Is it okay if I record the interview?

Open questions

(1) Proof-of-concept

What do you think of, when you watch this visualisation? Explain the thoughts that comes to your mind.

Is this visualisation appealing – do you think it is good looking?

Is this visualisation motivating you in any way?

Is this visualisation making you curious?

(2) Proof-of-concept, little bit more details.

What about this one? Please think-out-loud.

Task-based questions

(3) Live data

What information can you retrieve? Please tell me everything that comes to your mind.

What is a bubble?

Why do they have different colours?

How many percent of the users are currently at the email-question?

Please tell me the funnel steps with the most users, starting with the step holding the largest number of users in a decreasing order?

What do you think that the ‘logged in’ means?

(4) Simulation of a day

Can you explain the visualisation to me?

How many people have signed (at 24:00)?

At what time were most users talking with Hedvig?

Why where so few bubbles moving in the beginning of the session?

What was written underneath the time/date label?

(5) Get information:

If you were going to get information about a bubble – what would you do?

If you were going to get information about 'a step' how would you do?

Further-development conversation

How could this visualisation – both the live data and the simulation – be improved and developed?

How could the design be better?

What interactivity could be nice to add/remove?

How can you get usage of this visualisation in your professional role? In your working tasks?

How could this visualisation be useful for others at the company?

How could the question of time be addressed?

How could the question of cohorts be addressed?

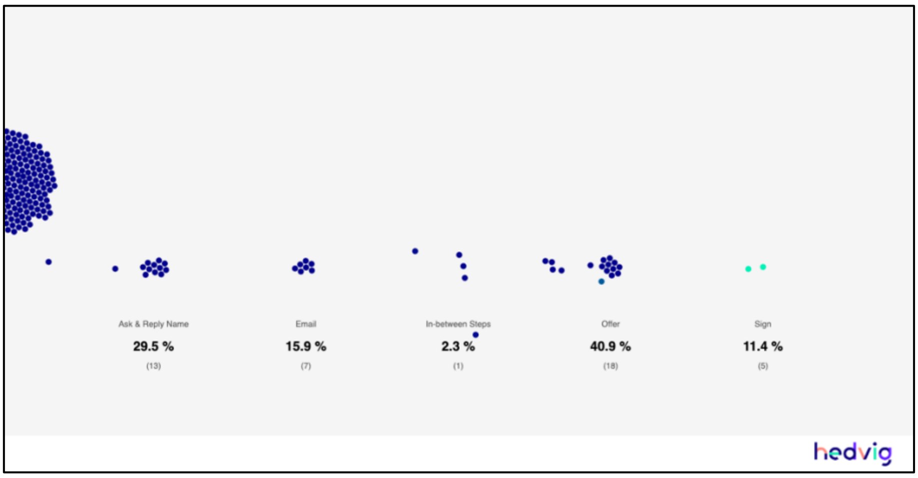
How could the question of the in-between steps or multiple flows be addressed?

Anything else? Have I forgotten to ask you something?

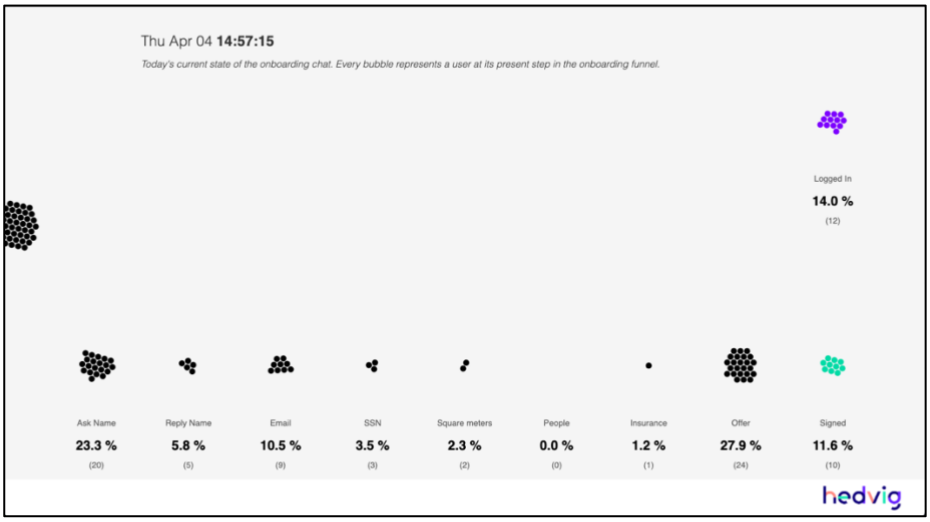
Appendix IV – Evaluation scenarios

(1) Proof-of-concept

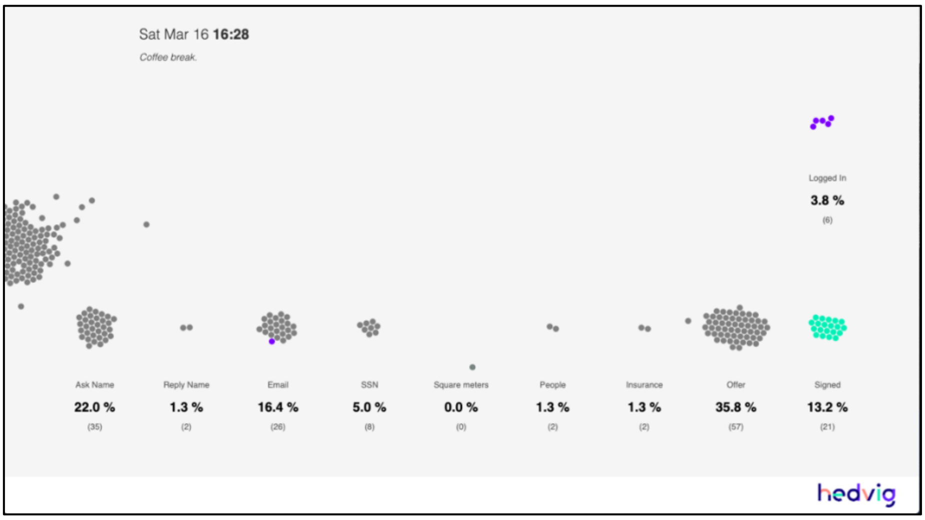
(2) Proof-of-concept, little bit more details.



(3) Live data



(4) Simulation of a day



(5) Retrieve information of a user or a cluster.

The bubbles get a reddish border if they are clicked.