

Schedulvis: Visualization of Personal Time Logs

Hung-li Chen

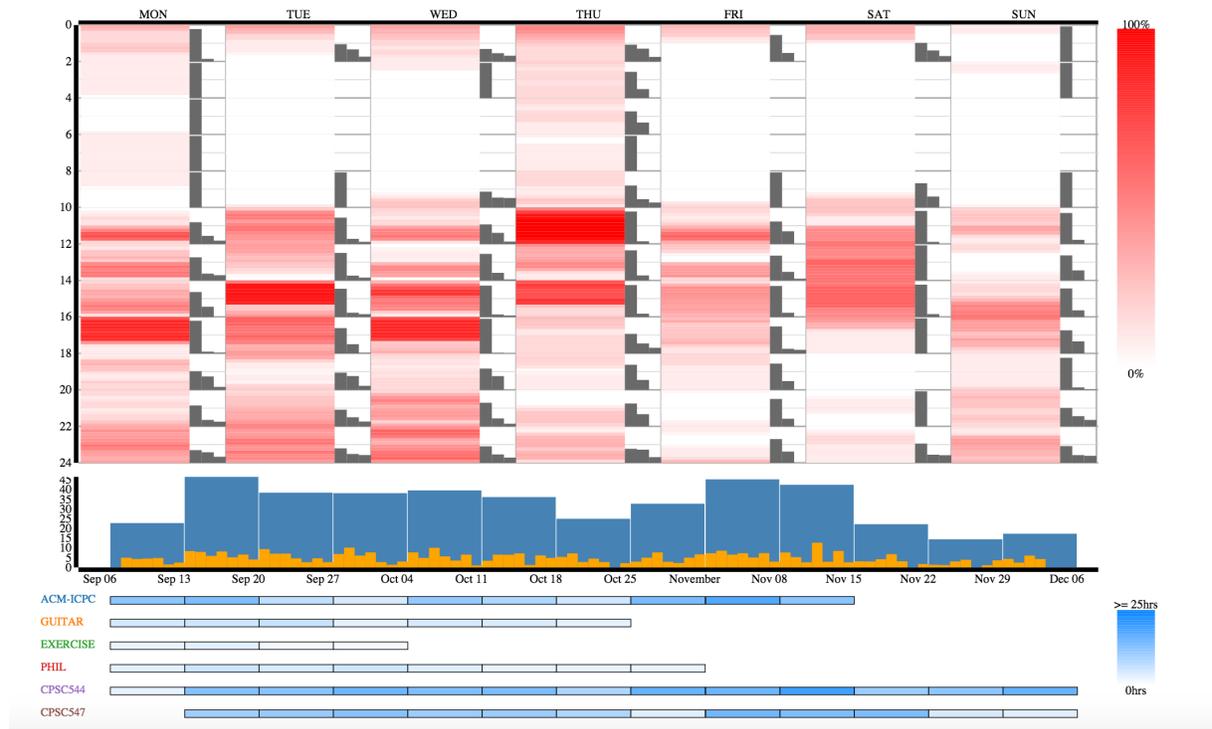


Fig. 1. User interface of Schedulvis composed of the calendar view and the timeline view. The calendar view is the top half with red colour map and the timeline view contains the blue components at the bottom.

Abstract— Personal time logs are collected by many people to provide insights of their work patterns and life patterns. However, the visualizations provided by the time tracking tools do not provide easy exploration and analysis. Schedulvis is designed and implemented to help people analyze personal time log data on desired long term activities. The system contains two main views, the calendar view and the timeline view. The calendar view shows data indexed by the time of the week, and the timeline view shows data indexed by intervals on the timeline. Easy exploration of activity entries and detailed information on demand are provided through interactions.

1 INTRODUCTION

The scarcity of time leaves people with many decisions to make. With our limited time endowment in a day, in a week, and in a life time, we need to decide, explicitly or implicitly, which activities are of the best worth of our time. In trying to use their time better, some people decide to use time tracking tools such as toggl [9] or Timesheet [8]. These tools provide timers or other input methods to record the time spent on tasks, projects, or anything of interest. By gathering data about time usage, they want to gain insights about their working or living patterns and potentially find ways to improve.

Gathering and analyzing personal data is not something new. The quantified self community has been advocating for self discovery through accurate observations [7] for more than five years, and there has been research in the HCI community about personal informatics systems [6]. However, there are not many visualization solutions for personal log data, and while reporting features are common among the

tracking tools, the visualizations presented usually lacks interactions that allow users to explore and learn more about their data.

In this project, a visualization system, Schedulvis, is designed and implemented to improve the analysis process of personal time log data on desired long term activities. Schedulvis contains a calendar view and a timeline view. The calendar view allows the users to interactively explore the raw time log entries and provides visualizations of utilization of different hours of the week and the projects involved, and the timeline view shows the the total hours aggregated over different time scales, and over different projects.

The following sections are organized as follows. In section 2, related work in personal visualization, personal time log, and time-varying data are discussed. In section 3 and 4, the data and the tasks considered by Schedulvis are discussed. In section 5 and 6, we present the design of Schedulvis and the design rationales. Section 7 describes the implementation of Schedulvis. In section 8, the results are presented with a walkthrough of an analysis scenario. In section 9, we discuss the strengths, limitations, and future directions of Schedulvis. Section 10 contains the conclusion.

• Hung-li Chen with the University of British Columbia. E-mail: hlc@stat.ubc.ca.

2 RELATED WORK

Schedulvis visualizes the personal time logs using the calendar metaphor and along a timeline. In the rest of this section, we discuss the growth of application of information visualization to personal data, previous work on personal time log data, and the time-varying data.

2.1 Personal Visualization

Personal visualization and analytics have been defined and discussed by Huang et al. [4] as visualization and analytics applied to the personal context. To guide further work, they provided design dimensions such as data scope, usage context, and actionability, and discussed the challenges such as fitting into personal routines and defining appropriate baselines for consideration. Examples of personal visualization include those produced by the Quantified Self community [7], in which visualization has been a common technique for gaining insights about oneself. The most common visualization produced include line charts, bar charts and custom visualizations such as infographic style website, calendar or photo grid, as identified by Choe et al. [3], in their study of the practice of quantified selfers. Schedulvis also utilizes these common elements, but includes more interactivity. Personal visualization also fit into the broader field of personal informatics. Li, Dey and Forlizzi discussed a stage-based model consisting of preparation, collection, integration, reflection and action [6]. Schedulvis covers the reflection stage.

2.2 Personal Time Log

Personal time log data has been considered in visualization. Barata, Nicolau and Gonçalves [1] presented AppInsight that visualizes computer usage history to help users answer questions about when and what they are doing with their computer, and how does the usage change over time. They encode the hourly usage in a grid representing specific hours of a time line, the list of applications as text ordered and sized by amount of time spent on the application, and the overall usage as a line chart at the bottom near the timeline. Schedulvis targets similar questions and has similar layout as AppInsight, but Schedulvis targets more general time log and shows more aggregated data.

Visualizations of time log data are also provided by the commercial time tracking tools. Toggl [9] shows the daily total hours as a bar chart, and the weekly percentage of projects as a donut chart. Timesheet [8] encodes average daily hours of each day of week as a bar chart and shows the total hour progress over time as a line chart. Instead of the static visualizations in the reports, Schedulvis provide more interactivity for exploring data over different weeks.

2.3 Time-Varying Data

Calendar and timelines are common ways of showing time-varying information, and they are both used in AppInsight [1] mentioned above. In addition to the common calendar grid, in their poster, Huang, Tory and Bartram [5] presented visualization of time-varying data on a calendar by appending an narrower column to each day column of the calendar, and encode the data in that narrower column. In Schedulvis, the calendar also uses two columns for each day of the week to encode more information. Showing time on the x-axis is also part of Gantt chart, and Schedulvis uses a similar horizontal bar layout to encode the time spent on different projects.

3 DATA

3.1 Data Description

The dataset is a personal activity log recorded over a span of three months using toggl[9], containing nearly four hundred entries. Each item in the dataset is associated with the following attributes: start date time, end date time, and project name. The time segments denoted by the entries do not overlap, and the time segments do not cover the entire time line.

Various data are derived from the raw data. The aggregated daily/weekly hours are computed as the sum of durations of every entry within that day/week. The aggregated weekly hours per project are

computed as the sum of durations of entries belonging to that project in that week. The time period utilization rate of a specific time interval of the week, is the total time spent in that period divided by the length of the period multiplied by the number of weeks in the dataset. For example, suppose we have a dataset spanned over two weeks. In the first week nothing is done during 13:00-13:05 on Monday, and in the second week four minutes of work are done during 13:00-13:05 on Monday. Then, the utilization rate of 13:00-13:05 on Monday is $(0+4)/(5*2) = 40\%$. In Schedulvis, the intervals considered are the five-minute intervals $[0:00 + (5*i \text{ minutes}), 0:05 + (5*i \text{ minutes})]$ for i from 0 to 287 for each day of the week. The project weight for a specific time interval of the week, is the time spent on that project during that time interval divided by the time spent in that interval by all projects. Using the above example, suppose out of the four minutes, three minutes are used for project A and one minute is used for project B. Then the weight for project A is $3/4$ and the weight for project B is $1/4$. In Schedulvis, the intervals considered are the two-hour intervals $[0:00 + (2*i \text{ hours}), 2:00 + (2*i \text{ hours})]$ for i from 0 to 11 for each day of the week.

3.2 Data Abstraction

Abstractly, each item in the raw dataset contains a interval on a timeline, a pair of quantitative attributes, and a categorical attribute of project name. It is guaranteed that the intervals of the items are mutually disjoint. The number of categories in the current dataset is six.

The derived aggregate daily/weekly hours contains a quantitative attribute of hours and is indexed by a categorical key of the date/week. Aggregate weekly hours per project has the same type as aggregate weekly hours and has an extra categorical attribute of project name. Utilization rate is a quantitative attribute between zero and one indexed by the time period of the week, and the project weight is also a quantitative attribute between zero and one indexed by the time period of the week and the categorical attribute of project.

4 TASK

Schedulvis is intended as a tool for reflecting and exploring past time logs to gain insights into oneself. It is designed to support the following tasks

1. Present an overview of the structure of time usage.
2. Discover projects that shares similar time frames.
3. Explore the raw log data.

For task one, the overview aims to cover both the temporal structure and the project structure of the time log. An example of temporal structure is the change of total hours over different weeks, and an example of project structure is the weeks during which the project was worked on. Task two aims to provide insights on which projects may be related. Example insight maybe that two projects tend to occur in the same week. Task three aims to provide the users an easier way to navigate the raw log data.

In more abstract terms, task one aims at presenting patterns with respect to the quantitative attribute time and with respect to the categorical attribute project. Task two aims to find relationships between entries of different value of categorical attribute project. Task three aims to allow easy exploration of the quantitative attributes start time and end time, and categorical attribute project.

5 USER INTERFACE AND VISUAL ENCODINGS

Schedulvis is composed of two views, the calendar view and the weekly timeline view, stacked vertically with the calendar view on top (fig.1). The calendar view provides the axis for encoding data of the specific time of the week such as the utilization rate and the project weight, and the timeline view provides the axis for encoding weekly and daily data such as the weekly total hours tracked. The two views are linked with interactive highlighting.

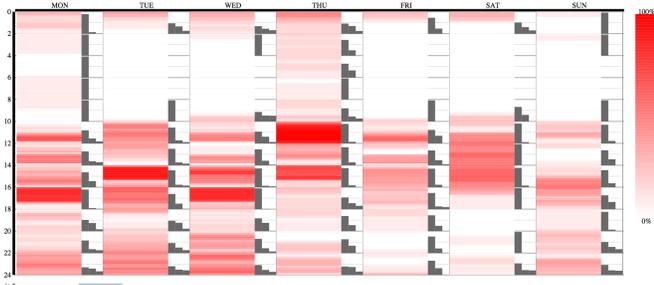


Fig. 2. Calendar View shows the utilization rate as the lightness of the bars. The colour scale used is on the right. The project glyphs are also in the calendar view, and has the same grey colour.

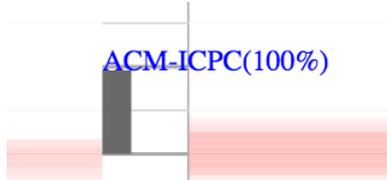


Fig. 3. Project weight glyph with the project name tooltip on hover.

5.1 Calendar View

The calendar view is composed of two parts, sharing the same axis. The first is the calendar heat map (fig.2), and the second is the project weight glyph (fig.3). The x-axis of the calendar view is split into seven bands representing the different days of the week, and the y-axis is a continuous axis from 0:00 to 24:00.

The calendar heat map are designed to present the five-minute time interval utilization rate. The utilization rate is mapped to the colour of the rectangular area mark representing that interval. The top boundary of the area mark has y-position determined by the start of the interval on the y-axis and the bottom boundary has y-position determined by the end of the interval. The left and the right boundaries is the left and right boundaries of the left sub-column of the day of week column. The colour is determined by the value of the utilization rate using a linear lightness colour map from white, representing zero, to red, representing one.

The project weight glyph is designed to help users identify specific time intervals of the week that is used by multiple projects. Spatial position is used to map the specific time period. The top and bottom boundary of the glyph has y-position determined in the same way as the heat map intervals. The left and right boundaries are the left and right boundaries of the right sub-column of the day of the week column. Within the glyph, the weight of the three projects with the highest weight are shown. The order and the x-position of the bars are sorted in descending order of the weight, and the length of the bar represents the weight of the project.

5.2 Timeline View

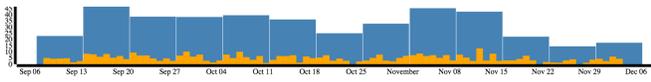


Fig. 4. Weekly and daily total hours encoded as bar charts.

The timeline view is composed of two parts, the aggregate hour bar chart (fig. 4 at the top and the project interval (fig ??) bars at the bottom. The two shares the same x-axis which represents a timeline with time increasing towards the right.

The aggregate hour bar charts has an adaptive y-axis that ranges from zero to the max total hour of any week. Weekly total hours are mapped to rectangular marks on the timeline with the width of the bar



Fig. 5. Weekly project hours mapped to the bars. The linear colour map used for encoding project weekly hour from white to blue is at the right.

corresponding to the time difference between the start of two weeks in the timeline x-axis. The total hour of the week is mapped to the height of the bars. Daily total hours are mapped similarly with the weekly bars, except with a different colour and with the width corresponding to the x-difference between the start of two days.

The project bars at the bottom presents whether any work on the project is logged in the weeks, and the total hours logged on that project in the weeks. The categorical attribute project is mapped to the vertical position on a categorical scale without a particular order. The aggregated weekly hours for the projects are mapped according to the following rules. If the number of time spent on the project is zero in that week, there will be no mark for that project in that week. If the number of time spent on the project is above zero in that week, there is a rectangular area mark, aligned to the weekly bars above, to represent the data. The colour of each mark represents the number of hours spent on that project in that week. The colour map is linear to the lightness going from white, representing zero, to blue, representing twenty five, with values over the limit clamped at blue.

5.3 Interaction

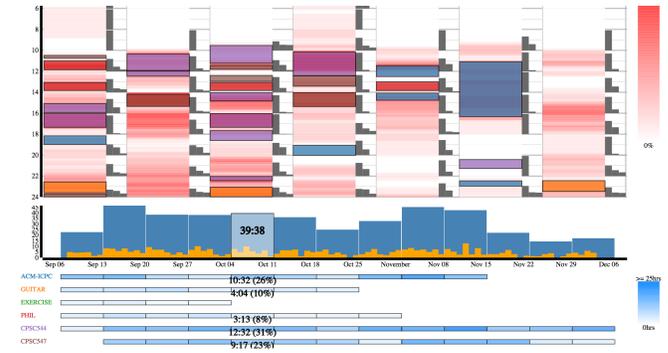


Fig. 6. Bottom bar charts.

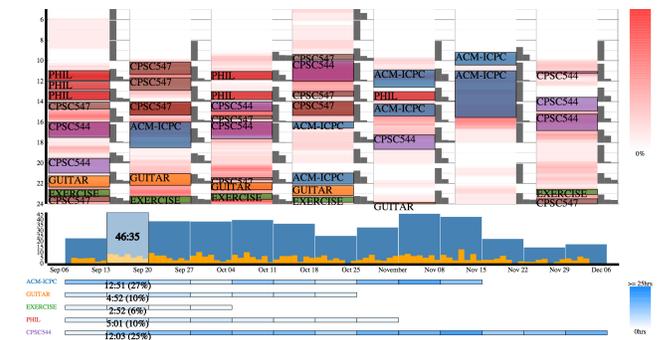


Fig. 7. Bottom bar charts.

The main interaction supported is showing raw data and detailed information of each week upon mousingentering the weekly bars in the timeline view. When the cursor moves inside the weekly hour bar, the number of hours of the week appears on top of the weekly hour bars, and the number of hours for each project in that week along with the contribution, in percentages, of each project in the total hours of that

week. At the same time, the raw data entries of the selected week are shown (fig.6) on the calendar similar to calendar events with categorical colour coding for the projects. Colour coding for the projects are consistent across different weeks.

There are also detail information shown when elements are hovered. When hovering over the bars in the glyph, the name of the project of the associated bar is shown on top of the bar (fig.3). When the raw data are shown on the calendar, hovering over any raw entry shows the project of each entry (fig.7).

6 DESIGN RATIONALE

The calendar heat map is designed to support task one by showing the overview of time usage by the utilization rate of small time intervals. Another alternative of showing the utilization rate is encoding the rate by the length of the rectangle mark, and get an area chart to show the distribution. Although the value of utilization rate maybe more accurately perceived in an area chart because humans do not distinguish lightness as well as length, the heat map has the advantage that it mimics the look of a typical calendar schedule.

The side glyph is designed to support task two by showing the weight of the top projects in the time interval. It shows a single vertical bar if the time interval is used by only one project, and it looks more similar to a horizontal bar when the top three projects have similar weights. The weights of the projects in each time interval may be encoded as pie charts or stack bar charts. The current glyph, although shows less information, takes less space, and supports the identification of the case of many projects sharing the same time interval.

Showing the weekly/daily total hours over a time axis supports the identification of temporal trends in their overall time usage, and is part of task one. Bar chart is chosen because it is customary in the domain as existing commercial time tracking software usually include temporal bar charts in their reports. It is also reasonable to use bar charts because length is perceptually easier to compare. We decide to overlay the daily charts with weekly charts because we think, because of the magnitude difference, the two will not interfere with each other.

The project bars shows the weekly total hours for each project. It is also possible to use stacked bar chart, multiple bar charts, or a single line chart. We choose the project bar instead over stacked bar chart because we want to see the temporal evolution of total hours for each project independently. Multiple bar charts are not chosen because it will take more space for each project. Line chart is not chosen because the possibility of cluttering. The project bar is also desirable because it presents the absence of a project as a salient special case.

7 IMPLEMENTATION

Data is manually transcribed from our personal toggl accounts into csv format and served with a python SimpleHTTPServer. The raw data contains the log entries with start date-time, end date-time, and associated project name. The layout, axis, scale, and visual encodings are all implemented in Javascript using D3.js [2]. The implementation involves data derivation, scale and axis definition, mark creation with placement, and interactivity incorporation. We derive the hour-utilization rate, weekly and daily hours, project weekly hours, and most used projects for an hour by restructuring the data with d3.nested and then aggregating the desirable properties. The scales and axis are built-in functions in D3.js and is created by defining the domain and range of each scale. Mark creation and interactivity incorporation are added according to the data using D3.select.

8 RESULTS

The following scenario shows how Schedulvis may be used. After tracking the time for two courses, two personal activities, and two extra-curricular commitments for thirteen week, we load the data into Schedulvis.

Upon entry, we can clearly see from the calendar heat map that the utilization rate is very high from 16:00 till 17:30 on Mondays and Wednesdays and from 14:00 till 15:30 on Tuesdays and Thursdays, and the side glyph shows that those times are mostly used by a single project (fig.8.A). We recognize that these are the lecture hours. Around

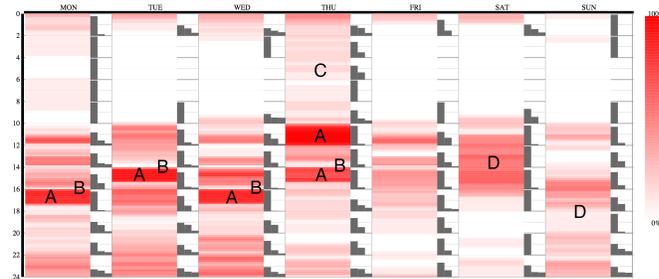


Fig. 8. A. The lecture times have the highest utilization rates. B. The utilization is low for minutes right before the lectures. C. There is a non-trivial utilization rate for early morning on Thursday. D. Work on Saturday concentrates in the afternoon while work on Sunday concentrates during the night.

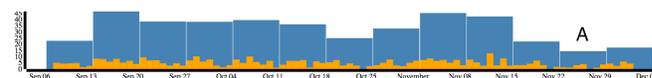


Fig. 9. A. Drop in hours in late November.

these time intervals, we notice that rarely is the time right before lectures used for the tracked projects (fig.8.B) Looking at early mornings, we see that during sleeping hours the utilization rate is low, but Thursday from 0:00 till 10:00 has a non-trivial utilization rate, signalling a bad working habit (fig.8.C). Looking at the side glyph, we see that two projects are most dominant during early morning Thursday, and after hovering over the bars, we know that the two projects are the two courses. Another interesting note is that on Saturday, the afternoon is utilized more while on Sunday the night is utilized more (fig.8.D). Arguably, all these patterns maybe recalled by the user, but the heat map provides a holistic view of the schedule and helps the user recall and recognize patterns and habits.

Looking at the temporal bar chart in the timeline view, the first notable feature is that the number of hours dropped significantly after mid December (fig.9.A). Looking at the project bars, we see that the hours dropped maybe partially explained by the inactivity of ACM-ICPC project (fig.10.A). Meanwhile, at the end of the term, the hours for CPSC544 increased (fig.10.B) and the hours for CPSC547 decreased (fig.10.C). To investigate further, we hover over last three weeks, and see that CPSC547 hours dropped by about ten hours while CPSC544 increases by five hours. The daily charts shows that the hour distribution over the week is not consistent over the weeks.

Finally, we see that the the week at the start of November has the highest number of hours, so we hover the weekly hour bar to investigate more. From the overlay, we learn that forty-five hours are tracked in that week, and they are spread over ACM-ICPC, CPSC544 and CPSC547 with ACM-ICPC being the most dominant accounting for 37% of the hours tracked. Looking at the raw data on the calendar, we see that during this week, many hours are spent on the courses in the first half, while many hours are spent on ACM-ICPC in the second half.

9 DISCUSSION AND FUTURE WORK

As demonstrated in the previous section, Schedulvis can provide insights about the time usage pattern of the person. More specifically, it provides the tools for analyzing the usage of different time of the



Fig. 10. A. Hours for ACM-ICPC dropped to zero. B. The hours for CPSC544 increased. C. The hours for CPSC547 decreased.

week, and allows for identification of temporal pattern of total hours used. It is, however, not certain that these are the patterns of interest for the users. Since there are many ways to derive new data from the raw log entries, without further investigation, we cannot be sure that these patterns are indeed relevant for personal reflection.

Using the interaction provided in Schedulvis, it is very easy to explore the raw data entries, but currently the project name labels shown on hover may be cluttered when many short log entries are next to each other. The labels for the side glyphs are sometimes unclear due to the elements underneath.

Although Schedulvis is more interactive than the static reports produced by commercial time tracking softwares, there are still far more opportunity for more interaction. Allowing the user to choose the time range and select subsets of the projects for the heat map and side glyph will provide more flexibility. It is also useful to provide ways to re-order the project bars. It is also desirable to include a view showing more information about the project composition of a time interval in addition to the glyph.

In the long term, we would also like to expand the work to cover more actionable information and to handle data tracked over longer periods and with more projects and categories. In those settings, colour may not provide enough bins for the categories.

10 CONCLUSION

Personal time logs presents opportunities for people to learn more about themselves. We provided the tasks that could be relevant in visualizing time log data, and we designed and implemented Schedulvis to improve the workflow of analyzing personal time log data. In the calendar view, Schedulvis helps identify the weekly time usage, and the projects involved in different time, using the calendar heat map and the side glyphs. In the timeline view, temporal patterns of aggregate hours are easily visible, and more details are provided on demand. Finally, the mouseenter interaction with the weekly hours bars provides simple exploration of raw data. The result is a system that helps the users learn both temporal and categorical patterns of their time usage.

REFERENCES

- [1] G. Barata, H. Nicolau, and D. Gonçalves. Appinsight: what have i been doing? In *Proc. ACM International Working Conf. Advanced Visual Interfaces*, pages 465–472, 2012.
- [2] M. Bostock, V. Ogievetsky, and J. Heer. D3: Data-driven documents. *IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)*, pages 2301–2309, 2011.
- [3] E. K. Choe, N. B. Lee, B. Lee, W. Pratt, and J. A. Kientz. Understanding quantified-selfers’ practices in collecting and exploring personal data. In *Proc. ACM conf. Human factors in computing systems (CHI)*, pages 1143–1152, 2014.
- [4] D. Huang, M. Tory, B. A. Aseniero, L. Bartram, S. Bateman, S. Carpendale, A. Tang, and R. Woodbury. Personal visualization and personal visual analytics. *IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)*, pages 420–433, 2015.
- [5] D. Huang, M. Tory, and L. Bartram. Data in everyday life: Visualizing time-varying data on a calendar. *Proc. Poster Compendium IEEE VIS*, 2014.
- [6] I. Li, A. Dey, and J. Forlizzi. A stage-based model of personal informatics systems. In *Proc. ACM conf. Human factors in computing systems (CHI)*, pages 557–566, 2010.
- [7] Quantified self. <http://quantifiedself.com>. Accessed: 2015-11-25.
- [8] F. Rauscha. Timesheet. <http://play.google.com/store/apps/details?id=com.rauscha.apps.timesheet&hl=en>. Accessed: 2015-11-25.
- [9] toggl. <http://www.toggl.com>. Accessed: 2015-11-25.