TempMunger: A Visual Analytics Approach Supporting Transformations of Time-Oriented Data

Robert Thurnher

Masterstudium:
Information & Knowledge Management

Technische Universität Wien
Institut für Softwaretechnik und Interaktive Systeme
Arbeitsbereich: Information & Software Engineering
Betreuerin: Univ.Prof. Dr. Silvia Miksch
Mitwirkung: Dr. Theresia Gschwandtner

Motivation

Applied work within data science, specifically **Visual Analytics (VA)**, *"the science of analytical reasoning facilitated by interactive visual interfaces"* [1] (p. 4), can be seen to roughly consist of three main basic building blocks:

1. **Data wrangling** a.k.a. **munging**
2. Statistics & machine learning
3. Visualization & analysis itself

While the second and last mentioned fields are constantly evolving related tools and techniques, the first one is comparatively still lacking in terms of **high-level support**. This is when it comes down to actually wrangle usually messy real-world data into a format prepping it useful for analysis.

To this day, it mostly means fiddling around with the data manually, applying hand-crafted transformation scripts. A **tedious task** discouraging analysis of data altogether, especially if the ones intending to work with the data are technically unversed (e.g., journalists or business analysts).

However, combining contemporary knowledge and technology from the domains of *Human-Computer Interaction (HCI)* & *User Experience (UX)* [2] as well as *information retrieval*, *data mining / machine learning*, and *Information Visualization (InfoVis)* [3], could yield substantial improvements in this space.

The focus of our approach is on **time-oriented data** [4].

Problem Statement

The main research question in the context of this work is:
How can we support data wrangling with VA techniques?
More particular, we are to tackle the following subquestions:

- *Which data transformations are best supported by analytical methods and for which transformations is visual support beneficial?*
- *How do concrete data wrangling workflow processes look like and how can these processes be supported by VA methods?*
- *What data wrangling tasks need to be tackled in particular when dealing with time-oriented data and how can we support them with VA methods?*

The emphasis of this thesis is put on evaluating the feasibility of corresponding concepts via iterative **design**, **implementation**, and **evaluation** of a software **prototype**.

Methods and Results

At the end of the day, the, in an agile manner, iteratively designed and implemented, prototype had to be evaluated. Therefore, we have made use of the following well-known methods:

1. **Heuristic evaluation**
2. **User/usability tests**

Qualitative evaluation proved our approach and prototype to be overall **useful**, while providing potential for future work. Furthermore, we have generally **fulfilled** the requirements list derived during our design process based on state-of-the-art tools like UX personas and wireframing with UI mockups.

Hence, we have been able to answer our research questions **satisfactorily**, indicating our approach of supporting wrangling time-oriented data with VA techniques to be **viable**.
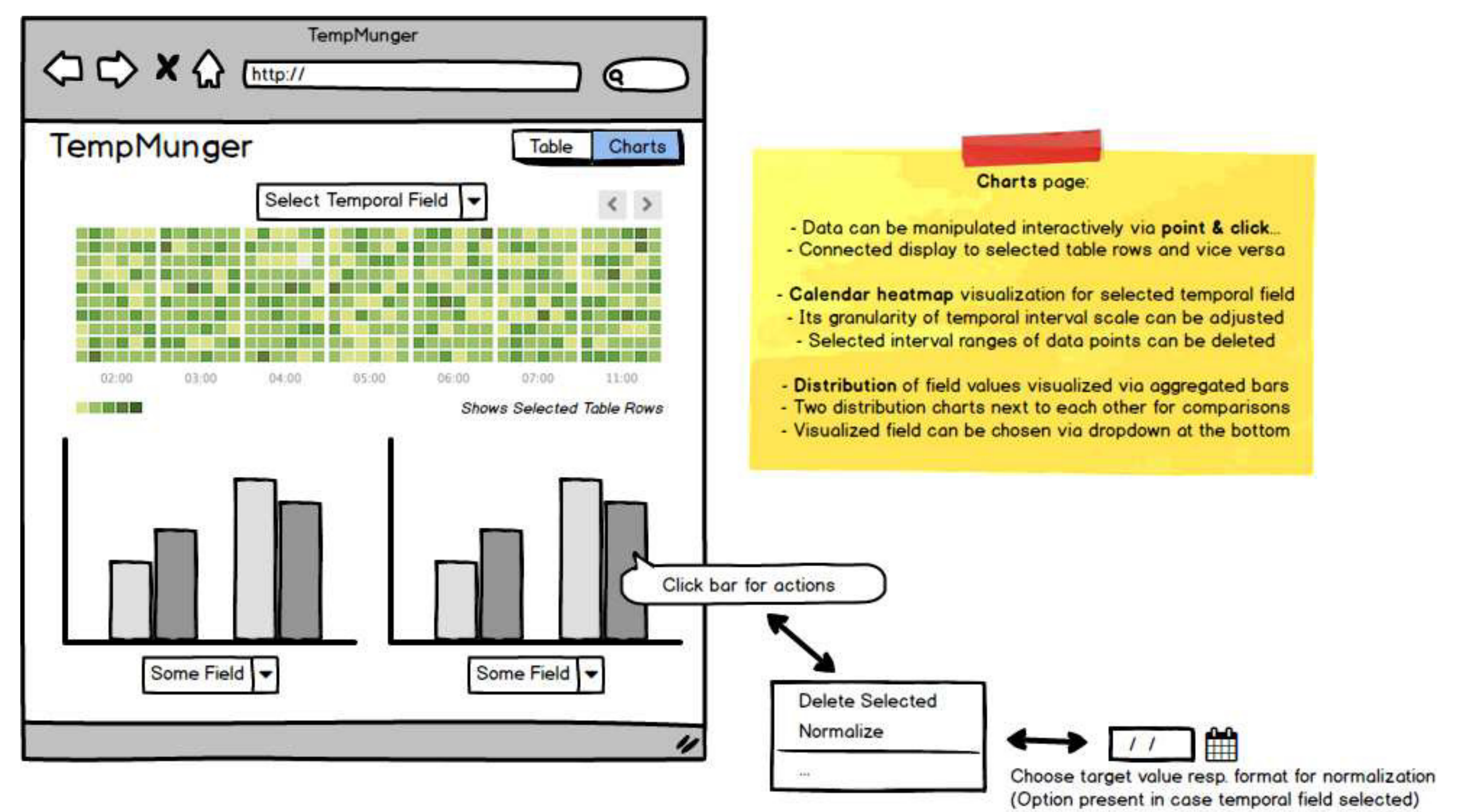
Our Approach and Prototype

Figure: UI mockup of the charts page including calendar heatmap visualization.
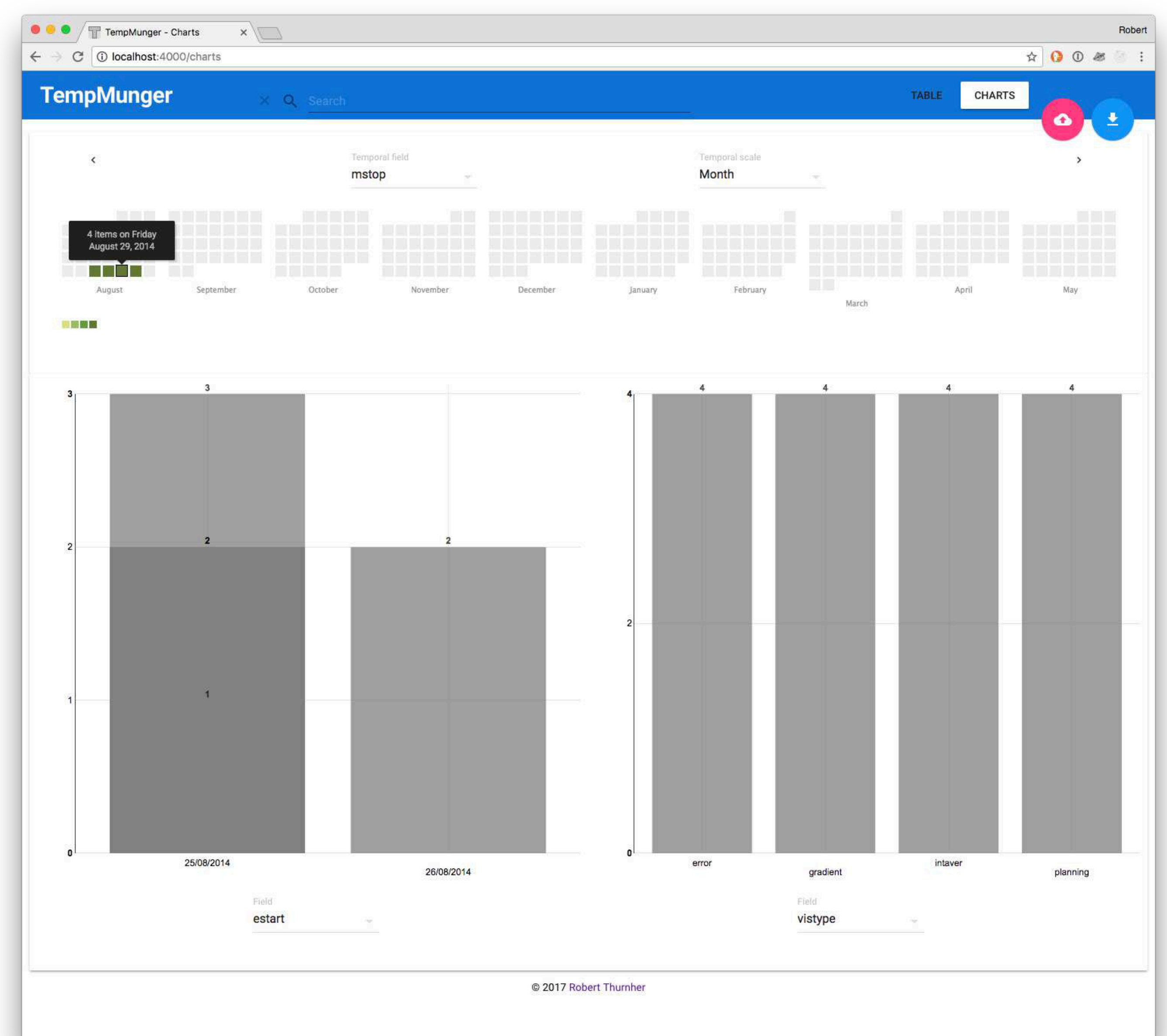

Figure: Screenshot of the resulting charts page, offering visual overview.

References

- James J Thomas and Kristin A Cook.
  Illuminating the Path: The Research and Development Agenda for Visual Analytics.
  IEEE Computer Society, 2005.

- Alan Cooper.
  The Inmates Are Running the Asylum.
  Sams - Pearson Education, 2nd edition, 2004.

- Stuart K Card, Jock D Mackinlay, and Ben Shneiderman.
  Readings in Information Visualization: Using Vision to Think.
  Morgan Kaufmann, San Francisco, CA, USA, 1999.

- Wolfgang Aigner, Silvia Miksch, Heidrun Schumann, and Christian Tominski.
  Visualization of Time-Oriented Data.
  Human-Computer Interaction. Springer, 1st edition, 2011.

Kontakt: robert@thurnher.email