

論文 / 著書情報
Article / Book Information

Title	Guiding Readers to the Focus and Context of Industrial Statistical Reports
Authors	Ken Wakita, Kohei Arimoto
Citation	Workshop on Visualization for Communication 2019 (VisComm19)
Pub. date	2019, 10
Creative Commons	See next page.

License



Creative Commons : **CC BY**

Guiding Readers to the Focus and Context of Industrial Statistical Reports

Ken Wakita*
Tokyo Institute of Technology

Kohei Arimoto†
Teikoku Databank The University of Tokyo

ABSTRACT

The poster proposes a bi-modal presentation method of industrial statistical reports to assist readers in comprehending the text by guiding readers attention to the context and the focus on the statistical data presented in the associated tables and charts. This technique is expected to assist the reader to grasp content of the report through good understanding of the statistics.

Index Terms: Human-centered computing—Visualization—Information Visualization;

1 INTRODUCTION

1.1 Dataset description

Teikoku Databank (TDB) is a Japan-based, business intelligence corporation founded in 1900. It owns Japan's largest corporate database and provides corporate credit research services. This research utilizes TDB's online reports, which are available from the following page: <https://www.tdb.co.jp/report/watching/>. The reports are targeted to industry leaders who can read Japanese, edited by professional editors hired by TDB, and published almost every working day. Each issue of the report covers specialized topics related to Japanese companies, such as “company capital investment incentives”, “changes in industrial structure”, “latest financial trends”, “market surveys”, and “employment trends”.

Though the reports are written by human editors, they are described in a standardized format. Reports are about eight pages long, single-column, and divided into sections containing paragraph sequences, followed by data illustrations, such as tables or charts (of various kind). The paragraphs refer to illustrated data points and report on facts deduced from the data.

1.2 References and Claims

To investigate the potential of a bimodal presentation of the report by linking the report text and illustration content, we extracted all the *data references* in the text content of the report and categorized them. Table 1 presents two examples of such data references found in a June 24, 2019, report on “Female leaders in Japanese companies”. The dataset they refer is a collection of quantitative values, indexed by years (1989, 1999, 2009, 2014, 2018, and 2019) and by seven industry types (“construction”, “manufacturing”, “wholesale”, “retail”, “transport and communication”, “service industry”, “real estate”, and “others”). The first example (1) in Table 1 is a simple dataset lookup, $D[2019][\text{RealEstate}]$, where D denotes the dataset. Example (2) indicates that the temporal change of the value for the “real estate” industry can be expressed using simple arithmetic: $(D[2019] - D[2019 - 30])[\text{RealEstate}]$.

It was also observed that the text contained *claims* based on various types of statistical reasonings. The lower part of Table 1 lists claims from the two paragraphs presented in Fig. 1. Claims (c) and (a) point out the lower extreme value of the column, and an upper extreme value of the temporal changes, respectively. Claims (d) and

(e) are regarded as the unique existential and universal properties, respectively. These claims can easily be expressed through statistic and logical functions. The editor's intention behind somewhat unclear phrases such as ‘high’ and ‘well below’ in claims (b) and (f) is explicitly formulated in the corresponding formulae.

1.3 Focus and Context

The reader must know that these claims are supported by a set of data values under consideration (the *focus*), the value distribution of other data items (the *context*), and the method of statistical reasoning being employed. For example, claim (a) discusses the increase in the proportion of female CEOs over the past 30 years and thus, the data points under consideration are the proportions in years 1989 and 2019 for all industry types, while the focus is the 7.5% increase in “real estate” industry, specifically. We can identify focuses and contexts of all other claims in the table and formulate their intention as described in the **Assertions** column of the table.

The focus and context in our proposal vary from one claim to another. The focus in our case refers to the data points and not variables, while the context corresponds to the scope of comparison under discussion. Latif and Beck [3] introduce a similar notions of *focus variable* and *context variable* in their interactive map report proposal. They assumed that the focus and context variables were embedded in the problem domain and do not change.

2 RESEARCH AIM: CONTEXT AND FOCUS OF STATISTICAL ANALYSIS

The aim of the research is to induce reader attention from report text content to the associated data illustrations (tables and charts). We hope to evoke a more self-motivated, critical, and explorative reading attitude by encouraging readers to

- locate illustrated data items corresponding to the text;
- verify “claims” described in the text through statistic testing; understand the statistical standards that lead to factual descriptions in the text;
- learn to discover undocumented insights on their own.

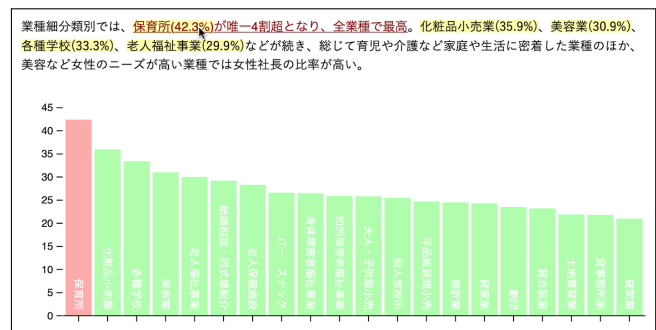
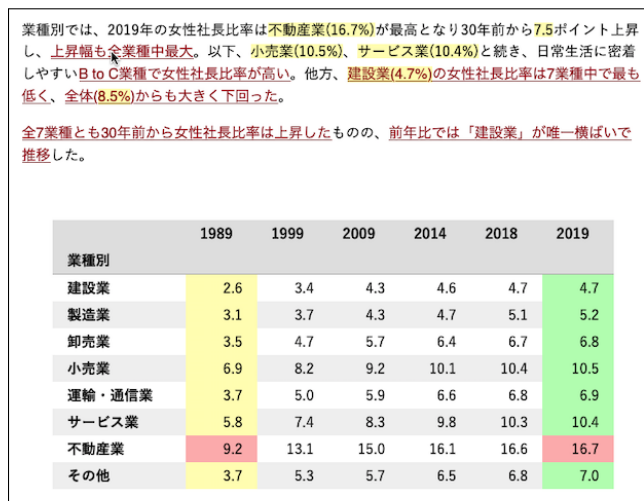
To realize this aim, we propose a simple visual interaction mechanism that links text and illustration, that directing reader interest to tables and charts, while highlighting data items shown on the tables and charts to let the reader comprehend the focus and the context of the text.

Fig. 1 and Fig. 2 depict our first Web-based prototype of this idea. The Web-based interface carries content identical to the original PDF-based report, but it also **highlights the text** that refers to data in yellow, while the claims supported by statistical reasoning are indicated with **red underlined font**.

The reader can move a pointing device over these highlighted text areas. Hovering over a data reference in the text makes a data item referred to in a table or a chart as the focus of the text and gives it a pink background color. Additionally, when a pointing device hovers over a claim in the text, its focus area is given a pink background color and the context area is given background color according to its degree of importance.

*e-mail: wakita@is.titech.ac.jp

†e-mail: kohei.a.19870908@gmail.com



REFERENCES

- [1] R. Brath and M. Matusiak. Automated annotations. In *An IEEE VIS workshop on visualization for communication (VisComm)*, 2018.
- [2] A. Kumar, M. Burch, I. van den Brand, L. Castelijns, F. Ritchi, F. Rooks, H. de Smeth, N. Timmermans, and K. Mueller. Eye tracking for exploring visual communication differences. In *An IEEE VIS workshop on visualization for communication (VisComm)*, 2018.
- [3] S. Latif and F. Beck. Interactive map reports summarizing bivariate geographic data. *Visual Informatics*, 2019. doi: 10.1016/j.visinf.2019.03.004