Data Article

A dataset of community perspectives on living conditions and disaster risk management in informal settlements: A case study in KwaZulu-Natal Province, South Africa

Maria Christina Georgiadou [a,*], Doreen Gyamfi [b], Stephanie Hirmer [c], Claudia Loggia [d]

[a] School of Applied Management, Westminster Business School, University of Westminster, 35 Marylebone Road, NW1 5LS, London, United Kingdom
[b] Computer Science and Engineering, University of Westminster, 309 Regent Street, London, W1B 2HW, United Kingdom
[c] Energy and Power Group, Department of Engineering Science, University of Oxford, United Kingdom
[d] Housing Programme, School of Built Environment and Development Studies, University of KwaZulu-Natal, Durban, South Africa

ARTICLE INFO

ABSTRACT

This article describes a dataset of community perspectives on living conditions and disaster risk management in Khan Road, a non-serviced informal settlement, located in Pietermaritzburg, the capital of KwaZulu-Natal province in South Africa. The data were collected by local community researchers via a structured questionnaire of 159 informal dwellers conducted between August and September 2022, using mobile phones via KoboToolbox. The dataset was analysed using exploratory data analysis (EDA) techniques. This household survey is part of a research project aiming to develop an evidence base of opportunities, risks and vulnerabilities related to housing construction and resource management in incremental upgrading of informal settlements in South Africa. This dataset can be used by local practitioners and policymakers involved in decision-making for informal settlement upgrading and help them prioritise re-

* Corresponding author.
E-mail address: c.georgiadou@westminster.ac.uk (M.C. Georgiadou).
Social media: 🐦 mcg1283 (M.C. Georgiadou), 🐦 @StephiHi (S. Hirmer), 🐦 @ClaudiaLoggia (C. Loggia).

sources and upgrading interventions based on what informal dwellers need. Furthermore, this cleaned dataset could support the analysis of further South African data guiding the development of digital platforms as a real-time resource management tool or guide the enhancement of existing theoretical frameworks in the field of participatory design and co-production used by academic scholars.

Specifications Table

Subject	Social Sciences
Specific subject area	Planning and Development
Data format	Raw, analysed, filtered
Type of data	CSV file
	Power BI dashboard
	Survey Excel file
	Tables
	Graphs
	Charts
Data collection	Data was collected through a face-to-face structured survey questionnaire conducted in a non-serviced informal settlement (Khan Road in Pietermaritzburg, the capital of KwaZulu-Natal province in South Africa) from trained local informal dwellers
Data source location	Country: South Africa
	District: Pietermaritzburg, KwaZulu-Natal Province
Data accessibility	In data repository: Zenodo
	Data identification number:
	doi.org/10.5281/zenodo.10108308 (v7)
	Direct URL to data: https://zenodo.org/records/10108308 (v7)

1. Value of the Data

- This dataset will feed into further research on developing an evidence base of local knowledge, resources and skills that informal communities in South Africa have. The aim is to integrate this data into participatory and inclusive upgrading strategies that would empower informal dwellers in the decision-making towards urban resilience, self-reliance and environmental sustainability. An important consideration is to produce a dataset incorporating household energy use, namely, access to electricity and heating sources, to better understand living conditions.
- This data will be presented and discussed with the local informal community, to empower them. This dataset can also be used by local Non-Governmental Organisations (NGOs), Nonprofit Organisations (NPOs) and Community Based Organisations (CBOs) to negotiate with Msunduzi municipality.
- This dataset can be used to better understand the key priorities of informal dwellers in South Africa and could further support the analysis of datasets on barriers and drivers for resource management and community resilience against existing climatic and social risks and vulnerabilities in incremental upgrading.
- The data reveal the integrated role of housing and services on community perceptions and vulnerabilities around living standards, climate impacts and skills development.
- This dataset can be used to inform (and accelerate) future municipal interventions for this or similar settlements enabling local practitioners and policymakers involved in decision-making to prioritise resources based on the most urgent needs of the community, namely

water and sanitation, electricity and waste management programs, as similarly done by Georgiadou et al. [1] and Hirmer et al. [2].

- This dataset can also complement studies on skills training and capacity building of local communities, endorsing how social education could be valuable as a tool for disaster risk management, community's disaster preparedness and response.
- Furthermore, this cleaned dataset could support the analysis of further South African data reflecting the needs of informal settlements or guide the enhancement of existing theoretical frameworks in the field of community resilience, participatory design and co-production used by academic scholars.

2. Data Format

The dataset described in this article provides information on community perspectives around urban resilience and particularly household demographics, living conditions, income level and disaster experiences in Khan Road informal community; a non-serviced community in Pietermaritzburg, the capital of KwaZulu-Natal province, in South Africa between August and September 2022.

The data was gathered through a household survey of 159 informal dwellers using mobile phones and the KoboToolbox app to enable digital data collection. Table 1 shows that the full community population is estimated to 363 informal dwellers, including children, according to the latest enumeration completed in November 2023. The total number of households in the settlement is 121. The intention was to survey the full community (estimated total number of 121 households and 363 informal dwellers, as shown in Table 1). The response rate is, therefore, 44 %. Informal settlements are dynamic places of living and with the number of informal dwellers continuously changing; hence, a random sampling method was followed. KoboToolbox is used by development professionals and researchers in the field of urban planning and development to design and implement primary data collection protocols in various settings around the world [3].

Table 1
Key demographics of Khan Road settlement.

Key Demographic	Metric
Age of Settlement	1996
Estimated Population	+- 363
Number of Households	+- 121
Number of survey participants	159 (adults only), response rate 44 % of total population (363 inhabitants).
Number of Taps	01 stand pipes
Number of Toilets	5
Population Ratio to Taps	1 in 300
Population Ratio to Toilets	5 to 300
Structure count (Field Survey)	40 plus old structures. 33 new structures. More structures to follow
Structure count (Satellite Imagery- not updated)	Around 110
Type of Structure	Informal structure / wendy type wood & Iron
Common Building materials	Walls [zinc & wood]. Sheet metal. Roof [sheet metal]

The survey data was gathered using a structured questionnaire. It was provided as a comma-separated values file (CSV) which reports the answers to a set of 54 questions. The CSV file reports the interviewees' answers to a range of 11 topics affecting household characteristics, including: demographics, employment, community organisation and leadership, floods, fires, solid waste management, farming, water sanitation, transport, housing construction and self-building, communication and skills development. This was to ascertain qualitative data from informal

dwellers in relation to livelihood development, skills training and impact of natural disasters in Khan Road informal settlement.

Table 1 presents the key demographics of the settlement. Identifying variables such as names, GPS coordinates, and names are anonymised. In the sections below, first we describe how the dataset was built, and then we provide a complete description of the experimental design and methods used to acquire and analyse the dataset.

3. Dataset Building

The dataset was collected in three stages, as presented in Table 2:

The interviews were conducted by local informal dwellers, trained as Community Researchers (CRs). Participants were selected based on diversity, including large families, male and/or female head of households. Fifteen CRs were chosen by three local NGOs (Community Organisation and Resource Centre, CORC), Rooted Wings and Youth Mentors and Developers (YMAD) working with the community who were responsible for collecting information using direct observations and the survey questionnaire. The 15 CRs took part in three research training workshops led by Rural Senses Ltd and the academic leads to understand how to undertake the fieldwork activities and eliminate bias based on the ethics protocol issued by the University of KwaZulu-Natal.

Table 2
Key stages of dataset building.

Data Collection	August to September 2022 – following two preparatory workshops and site visits done in collaboration with three local NGOs and two academics from UKZN.
	Location: Khan Road informal settlement in Pietermaritzburg.
	Data Collectors: Khan Road dwellers (fluent in isiZulu and English) who participated in a three-day training workshop.
	Methodology: 159 completed interviews with adults (informal dwellers) were conducted using mobile phones provided by UKZN. Each interviewee completed one survey and the responses were recorded (where permission was granted). KoboCollect app was used for the primary data collection. Data was entered from interviews (offline) during fieldwork, including audio files and photos that were saved on the mobile phones.
Data Translation	The Translations were carried out between October and November 2022 by three research assistants based at the University of KwaZulu-Natal who are highly experienced in development studies in the informal space and they were both native isiZulu and English speakers.
Data Analysis	From February to May 2023 the dataset was analysed by a Nigerian Mst Data Science student based at the University of Westminster.

4. Survey

4.1. Data format

CR carried out 160 interviews, but 159 participants fully answered the questionnaire; hence, 159 responses were entered into a spreadsheet for analysis in an anonymised and disaggregated form. One interview was eliminated due to an incomplete response (i.e. handling missing data). The data from the 159 participants was then cleaned and checked for errors. Null answers (no answers) were removed and common answers ('there was nothing', 'none', 'not at all') were grouped under the same category and all renamed the same (e.g. 'nothing'). A subset dataset was created from the questionnaire consisting of 20 closed-ended questions using MS Excel. The closed-ended questions were advantageous for data analysis as they could provide clear, standardised, and quantifiable responses, facilitating data processing, comparison, and statistical analysis. Each question of the questionnaire was renamed as a new entry using the keywords as seen in Table 3. This data was then imported to Power BI for Dashboarding as a tool to aid

Table 3

Survey questions and data entries for household characteristics and community perspectives.

Question	Key
Can I take a few pictures of your house?	HOUSE_PIC
What training have you received, if any, from the municipality or other organizations active in the community?	EDU_RECEIVED
What is your main way to receive information?	INFO_RECEIVING
What is your main way to communicate with others?	COMMUNICATION_WAY
What community organization and/or social activity group you are part of? (e.g. church, stokvel, NGO, NPO, Community based organisation, etc.)	COMMU_ORG
What is your average monthly household income?	AVG_MON_SAL
Whom would you define as the head of your household?	HOUSE_HEAD
Number of people living in the same household?	NO_HOUSEHOLD
How has your economic situation changed since you moved here?	CHANG_ECO_SITUA
What best describes your accommodation status?	ACC_STATUS
What is the highest level of education have you completed?	EDU_LEVEL
What is your current occupation?	OCCUPATION
What is the employment status of the participant?	PARTICIPANT_JOB
Did you build this house yourself?	SELF_BUILT_HOUSE
Do you grow your own vegetables?	VEGES
Have you ever experienced flooding in your home?	FLOOD_EXP
How many times do you experience flooding in your home every year?	YRLY_FLOOD
In which months have you experienced floods in your home?	FLOOD_MONTH
Who would you contact in case of a flood?	FLOOD_CONTACT
Have you experienced fire in your home?	FIRE_EXP

the storytelling of the analysed data. The two main dashboards used were living Standard and Disaster analysis.

5. Data Description

Regarding household demographics and living standards analysis, the data was divided into male-headed and female-headed households, as shown in Figs. 1 and 2. For the income levels analysis, households were divided into the four categories: ZAR 500 and below, the ZAR 1000s, the ZAR 2000s, and the ZAR 3000 s, as illustrated in Figs. 3-6. The data was categorised using the thematic areas as presented below:

- The male-headed household consisted of households with fathers, uncles, sons, husbands, and grandparents.
- The female-headed consisted of households with mothers, daughters, and wives as household heads.
- The households were either fully owned or in rental accommodation.
- The households were either provided by the Municipality or self-built.
- For water infrastructure either community taps or household taps were used.
- Toilet types were communal toilets, personal toilets, ablutions blocks or open defecation.
- The educational levels varied between no education, primary, secondary, or tertiary (higher) education.
- Participants were unemployed, engaged in unpaid household work, or employed.
- Participants had received training from the municipality or other community-based or community-support organisations (e.g. NGOs) working with the community or had received zero social education to date.
- For the income analysis, the economic situation of participants had improved a lot, had improved a bit, had not changed, had not improved a lot, or had become worse.
- In terms of natural disasters, participants could have experienced either a flood, or a fire, or both, as shown in Fig. 7.
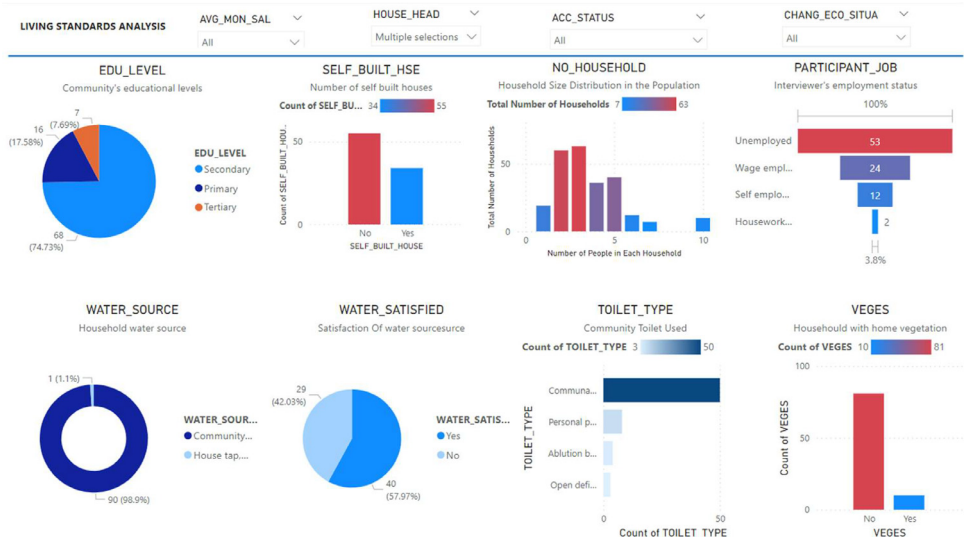
Fig. 1. Gender-based Living Standard Dashboard – Male-Headed Households.

Fig. 2. Gender-based Living Standard Dashboard – Female-Headed Households.

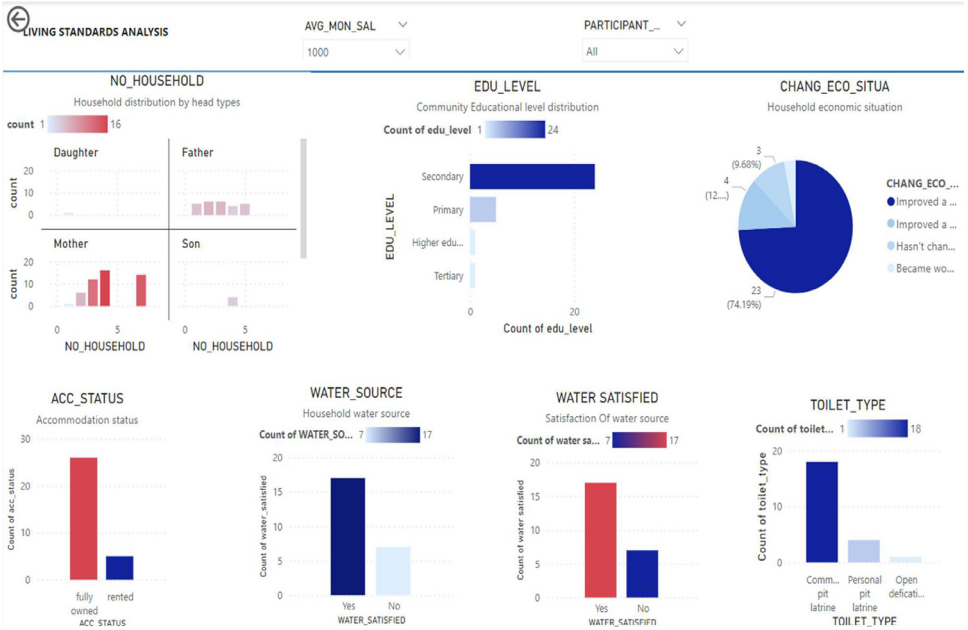**Fig. 3.** Income Analysis ZAR 500 and below.
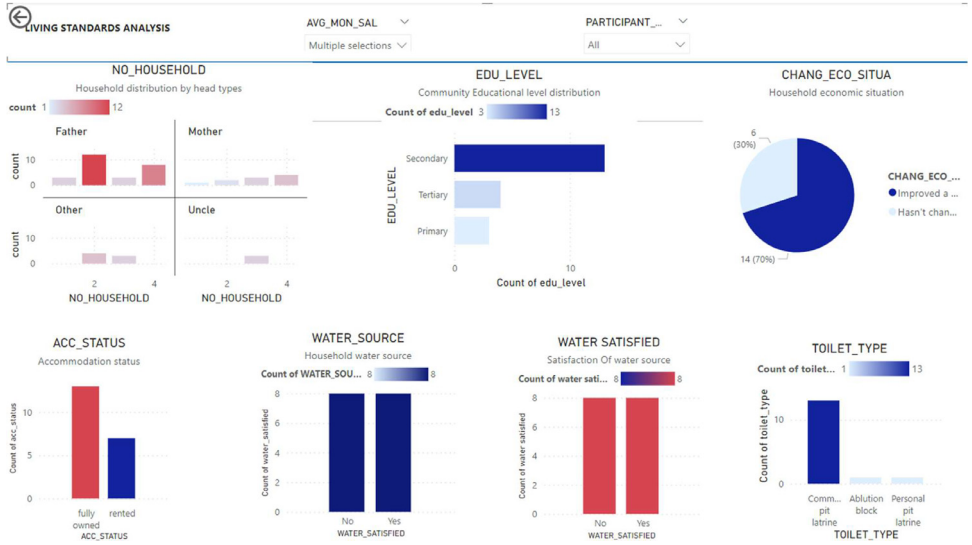

Fig. 4. Income analysis ZAR 1000s.

Fig. 5. Income analysis ZAR 2000s.

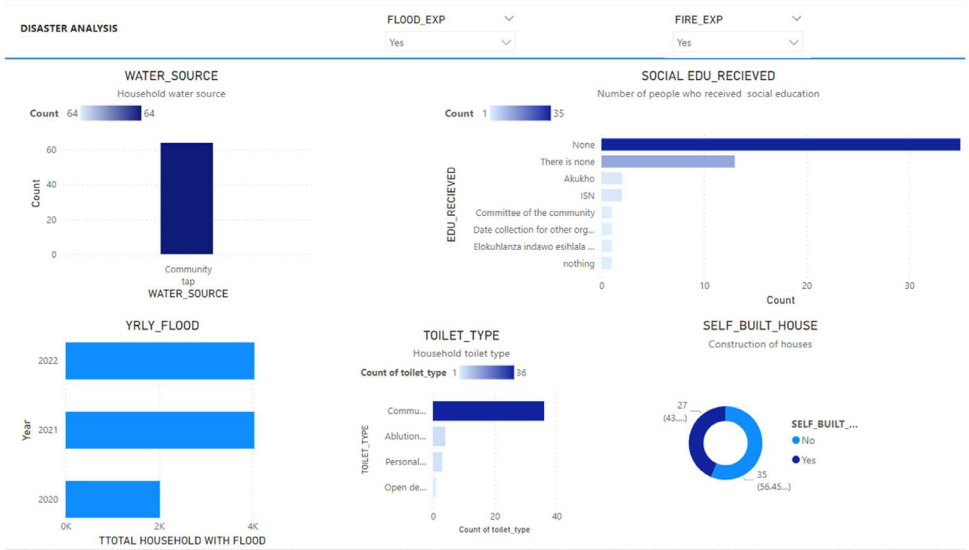
Fig. 6. Income analysis ZAR 3000 s.

Fig. 7. Disaster risk analysis dashboard – Both fire and floods.

6. Experimental Design, Materials and Methods

6.1. Sampling

Khan Road is a non-serviced informal settlement of approximately 121 households and 363 informal dwellers, as shown in Table 1. However, as the nature of informal settlements is dynamic, the total number of households is constantly changing. In this study, random sampling was applied, and data was collected by the 15 CRs (trained local inhabitants) living in the settlement. The original intention was to interview the full community. The 159 participants were adults (18 years old and older) who consented to be interviewed during fieldwork and provided complete answers to the survey questionnaire. No children took part in this study. Some informal dwellers were unavailable during data gathering, and a few were not interested in participating. The structured questionnaire was administered in person (using mobile phones) to participants by the trained CRs. The interviews were conducted orally in isiZulu and English. During household interviews, CRs took notes, photographs and audio recordings were also made. The trained CRs undertook the data collection with the supervision of local NGOs. They experienced some technical challenges (while collecting data using the smartphones) and sometimes they had to reschedule the interviews. Three translators were also employed; this was necessary to ensure consistency in fieldwork activities.

6.2. Case study selection

Khan Road was selected as a representative case study of a non-serviced informal community living in a central urban area, in proximity of many economic activities and characterised by low-density housing. The Khan Road settlement shows high vulnerability because of the poor living conditions of the community, which occupied a flood-prone area (in a wetland), with no municipal waste removal and a high fire risk.

6.3. *Exploratory data analysis*

The research method used for analysing this dataset is Exploratory Data Analysis (EDA) using Power BI for dashboard development [4]. EDA is a process of examining data to uncover patterns, relationships, and insights that can inform further research. EDA facilitates cleaning data to identify missing values, outliers, and inconsistencies in the dataset [5]. As the analysis focused on close-ended questions, there were no additional rounds of interviews to clarify any strange data. Null answers were removed were grouped under the same category and all renamed as the same (see 3.1 Data format).

EDA is particularly useful in situations where little is known about a topic or where the data is complex and difficult to analyze using more traditional statistical methods. Dasu et al. [6] add that the determination of variable selection can be examined with the use of EDA analysis by identifying correlations in the dataset. Moreover, it helps in the communication of findings for a clear and concise graphical representation of data [7].

EDA was used to analyse the responses to the survey questionnaire from the 159 informal dwellers in Khan Road informal settlement. EDA was considered as an effective method for this dataset as it allowed for a detailed analysis of the survey dataset through methods that can process and represent data visually [4]. The main objective of the questionnaire was to gain insights into the living conditions and experiences of households around housing, infrastructure, disaster risks and vulnerabilities and social service provision. This method helped to identify potential areas for intervention and further research to improve community resilience and, therefore, well-being and livelihood development of the community. Descriptive statistics were used to summarise the dataset to know the overview of the data for an understanding of EDA analysis to identify trends and patterns. Graphs, bar charts, and tables were used to visualise the data and help identify relationships between the variables. In particular, the focus was on analysing the data based on household heads, income levels, and experiences with natural disasters.

Limitations

EDA was helpful in comprehending the dataset for this investigation; however, its subjectivity makes it unsuitable for a comprehensive analysis that would include answers to semi-structured questions [8]. To mitigate bias and improve reliability, inputs from the research team in highlighting the key themes (i.e. living standards, economic conditions, and disaster experience) for analysis were taken into consideration in interpreting the dataset. EDA does not directly handle missing data [9]. When missing data is present, EDA could be performed using a subset of data without missing values. This approach may limit the generalisability of the data as it may introduce bias by excluding cases with missing data, which may have enhanced the interpretation of the data further. Despite these drawbacks, EDA offered useful insights, and future studies may consider merging it with other cutting-edge methods for a more thorough comprehension of the data.

Ethics Statement

To ensure the study's integrity, a risk and ethics assessment following the University of KwaZulu-Natal Research Ethics Committee in accordance with the procedures laid down by the University for Ethical Approval for all research involving human participants was completed and approved with Reference: HSSREC/00004460/2022. To protect the participants' identity, all names were removed.

Data Availability

A dataset of community perspectives on living conditions and disaster risk management in informal settlements: A case study in KwaZulu-Natal Province, South Africa (Original data) (Zenodo)

CRediT Author Statement

Maria Christina Georgiadou: Conceptualization, Methodology, Investigation, Visualization, Writing – original draft, Project administration, Funding acquisition; **Doreen Gyamfi:** Data curation, Formal analysis, Visualization, Writing – original draft; **Stephanie Hirmer:** Investigation, Methodology, Visualization, Resources, Validation; **Claudia Loggia:** Conceptualization, Writing – review & editing, Project administration.

Acknowledgements

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have or could be perceived to have influenced the work reported in this article.

References

[1] M.C. Georgiadou, C. Loggia, P., I. Bisaga, P. Parikh, Towards sustainable informal settlements: a toolkit for community-led upgrading in Durban, South Africa, Proc. Inst. Civil Eng. – Eng. Sustainabil. 174 (2) (2021) 83–93.

[2] S.A. Hirmer, A. Mazzone, A. Leonard, C. Conforti, The power of language: exploring values, empowerment dynamics and communication strategies for gender-inclusive energy service design in rural Uganda, Energy Res. Soc. Sci. 85 (2022) 102379.

[3] Kobo Toolbox making data collection accessible to everyone everywhere, 2022 https://www.kobotoolbox.org/about-us/ (Accessed: 05 July 2023).

[4] S. Palanisamy, Predictive analyitics with data visualization, J. Ubiquitous Comput. Commun. Technol 4 (2022) 75–96.

[5] J.W. Tukey, Exploratory Data Analysis, Addison-Wesley, Reading, Mass. And London, 1977 first ed.

[6] T. Dasu, T. Johnson, Exploratory Data Mining and Data Cleaning, John Wiley & Sons, Hoboken, New Jersey, 2003.

[7] S. Few, Now You See it: Simple Visualization Techniques for Quantitative Analysis, Analytics Press, Burlingame, California, 2009.

[8] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, Data Mining, Inference, and Prediction, Springer, New York, 2009 second ed.

[9] G. Shmueli, To explain or to predict? Stat. Sci. 25 (3) (2010) 289–310.