VAHC 2012

Proceedings of the IEEE VisWeek

Workshop on Visual Analytics in Healthcare: *Open Health Data*

October 17th, 2012
Seattle, WA
www.visualanalyticshealthcare.org

Sponsors:

NATIONAL INSTITUTES OF HEALTH

NATIONAL LIBRARY OF MEDICINE

IBM Research

Preface

Following a very successful workshop last year in which VisWeek participants had the opportunity to discuss and showcase their visualization techniques to leading clinicians, we propose to organize a follow-up workshop to discuss how visual analytics can play a central role in current national Open Health Data initiatives.

Given the challenges facing the healthcare system across the world, a number of recent government initiatives, including those in the United States and the EU, have dramatically increased the accessibility to a wide range of public health datasets. Many datasets are now available electronically, often via easy-to-use online APIs. Governments are looking to the innovation of data scientists and visualization experts to help extract insights from this data that can help influence policy, drive efficiency, and improve overall health and wellness. But governments are not alone, non-governmental medical institutions are also starting to embrace open data access in the hopes that new solutions emerge that can improve efficiency and cost effectiveness.

The 2012 Workshop on Visual Analytics in Healthcare provides an opportunity for participants to discuss visual analysis applications in the healthcare domain. The focus of this year's workshop will be on Open Health Data initiatives. In addition to sharing and discussing use cases, techniques, and applications in this area, participants will be able to meet with a program officer involved in major open health data initiatives and learn about funding opportunities. The workshop will allow participants to showcase their ongoing work on visualization and data mining techniques and to learn more about emerging sources of open health data.

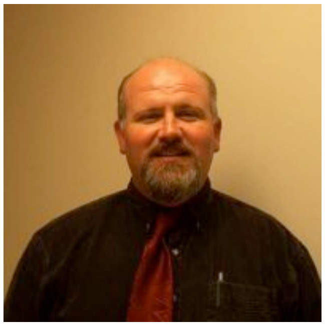
Jesus J Caban,
NICoE / Naval Medical Center
CC / National Institutes of Health

David Gotz
IBM Research

Keynote Speakers:

Will Pugh
CTO, Socrata

Mr. Will Pugh is the Chief Technology Officer Socrata. Socrata is the leading developer and provider of Open Data Services that enable federal, state, and local governments to improve the reach, usability and social utility of their public information assets. Mr. Pugh began his career with Microsoft as a Software Design Engineer before founding SourceLabs, a provider of tools and services used to support open source software environments. In past positions he has led teams for VMWare, EMC, and BEA Systems. Mr. Pugh graduated from Cornell University with a degree in computer science.



Rick Barnhill
Deputy Chief and
Program Manager,
Clinical Informatics,
WRMC

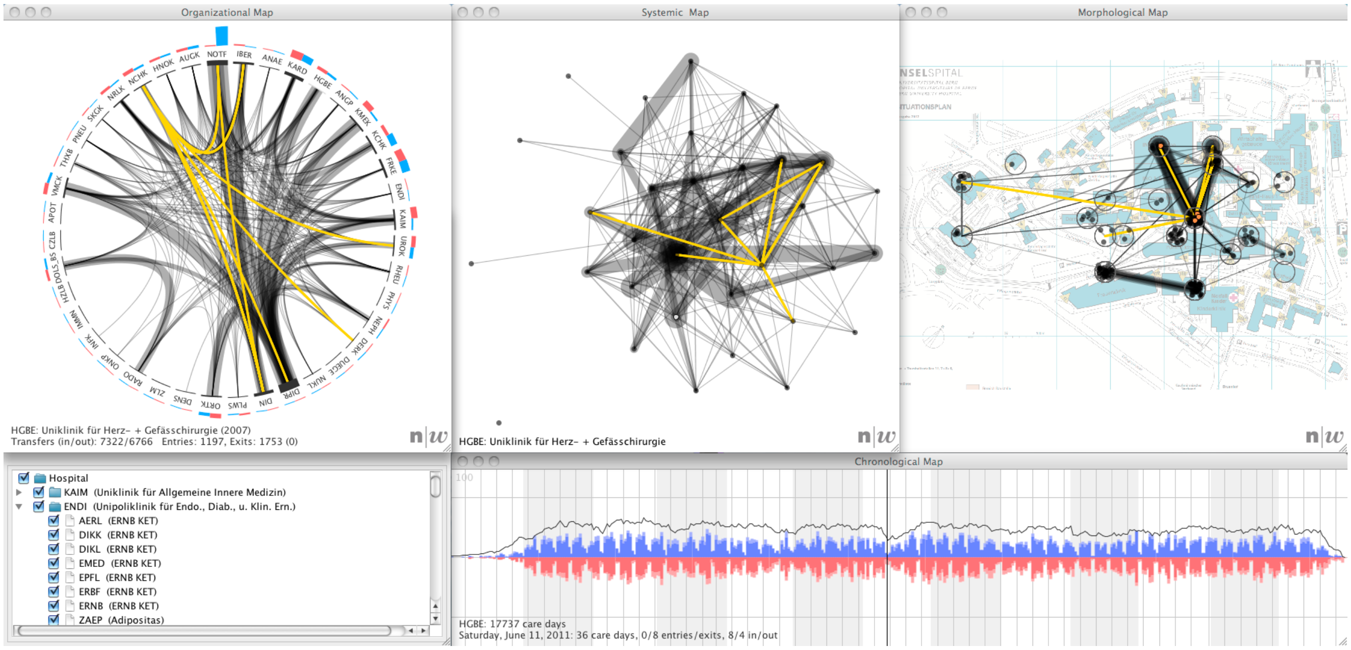
Rick Barnhill is the Deputy Chief and Program Manager for the Regional Clinical Informatics division for the Army's Western Regional Medical Command and Madigan Army Medical Center. In this capacity he develops and manages electronic medical applications for all U.S. Army facilities in a twenty state area reaching from Alaska to west Texas. Over the last seven years his office has worked as well on sharing electronic medical information with the Veterans Administration, supporting tactical medicine and many efforts focused on personal health information.

Agenda:

Session I: Visual Analysis of Electronic Health Records Chair: Jesus J. Caban, NICoE / Naval Medical Center	
2:00 - 2:10	Welcome
2:10 - 2:40	**Keynote:** Will Pugh Chief Technology Officer, Socrata Topic: Open Data Initiatives
2:45 – 3:45	Paper Presentations
	"*Masterplan: A different view on electronic health records*" Dominique Brodbeck, Markus Degen and Andreas Walter
	"*Designing an Open Source Presentation Layer for the Patient-Centered Medical Home*" Christopher Goranson and Jihoon Kang
	"*Interactive Visual Patient Cohort Analysis*" Zhiyuan Zhang, David Gotz and Adam Perer
	"*ReportViz: Interactive Visualization and Exploration of Topics and Key Words in Public Health Reports*", Wei Zhuo
3:45 – 4:15	Coffee Break
Session II: Visualization of Temporal Health Data Chair: David Gotz, IBM Research	
4:15 - 4:45	**Keynote:** Rick Barnhill Deputy Chief and Program Manager, Clinical Informatics, WRMC Topic: Funding Opportunities
4:50 – 5:50	Paper Presentations
	"*Challenges of Time-oriented Data in Visual Analytics for Healthcare*" Wolfgang Aigner, Paolo Federico, Theresia Gschwandtner, Silvia Miksch and Alexander Rind
	"*Interactive Visualization of Prescriptions of Drugs to Individuals within Large Populations - Analyses of Temporal Relationships of Events*" Jimmy Johansson, Morten Andersen, Alexander Fridlund and Mikael Hoffmann
	"No country for fat men; the young could not even round their arms" Aaron Lai, Thomas Ho and Ryan Walker
	"*No Country for Fat Men - Investigating Obesity with Visual Analytics*" Hui Zhang, Mike Boyles and Masatoshi Ando
5:50 – 5:55	Closing remarks

Masterplan: A Different View on Electronic Health Records

Dominique Brodbeck, Markus Degen, and Andreas Walter



HGBE: Uniklinik für Herz– + Gefässchirurgie (2007)
Transfers (in/out): 7322/6766 Entries: 1197, Exits: 1753 (0)

HGBE: Uniklinik für Herz– + Gefässchirurgie

☑ Hospital
☑ KAIM (Uniklinik für Allgemeine Innere Medizin)
☑ ENDI (Unipoliklinik für Endo., Diab., u. Klin. Ern.)
☑ AERL (ERNB KET)
☑ DIKK (ERNB KET)
☑ DIKL (ERNB KET)
☑ EMED (ERNB KET)
☑ EPFL (ERNB KET)
☑ ERBF (ERNB KET)
☑ ERNB (ERNB KET)
☑ ZAEP (Adipositas)

HGBE: 17737 care days
Saturday, June 11, 2011: 36 care days, 0/8 entries/exits, 8/4 in/out

Fig. 1. The flow of patients through a large hospital in four different views: organizational, systemic, topographic (top row), and chronological (bottom right). The yellow lines show the trajectory of a single case treated for abdominal metastasis during a four-week stay at the hospital.

Abstract— Hospitals collect large amounts of data during their daily operation. Next to its immediate primary purpose, this data also contains implicit information that can be used to improve clinical and administrative processes. We present a case study of how strategic infrastructure planning can be supported by the analysis of enriched patient flow through a hospital. Data from various hospital information systems was collected, enriched with topographical and organizational data, and integrated into a coherent data store. Common analysis tools and methods do not support exploration and sense-making well for such large and complex problems. We therefore developed a highly interactive visual analytics application that offers various views onto the data. The analysts were able to validate their experiences, confirm hypotheses and generate new insights. As a result, several sub-systems of clinics were identified that will play a central role on the future hospital campus. In this case study we show that a different view on electronic health records can provide not only clinical insights but also help improve operational efficiency.

Index Terms—Clinical informatics, visual analytics, patient flow, hospital planning, case study.

1 Introduction

Health care institutions such as hospitals collect and manage large amounts of data. Next to its eventual use in supporting clinical care, the data is also largely used for administrative purposes such as billing, scheduling, or resource planning [1]. While there are common hospital information systems that perform the core of the data management, there are typically many additional independent systems that are designed to support a specific medical procedure.

Integrating this data into a coherent data store, opens up the possibility to have a different view than what can be gained by just looking at the individual systems and their intended purpose. Interesting questions for strategic hospital planning for example require queries across several of the systems and along different dimensions: Which departments have many transfers between each other? Are there unusual transfers into other departments that deviate from this trend, and how are the involved patients characterized? Is this effect seasonal? If the data is enriched even more, for example with external data like geographic location of departments, weather conditions, etc. then further investigations become feasible.

Hospital sites are developed in an evolutionary manner over a long time span. This leads to physical and organizational layouts of the facilities that are usually not optimal anymore after a certain while. Strategic planning with time horizons of 25 years and more provides the opportunity to correct this degeneration, and optimize the layout when the campus is enlarged, new facilities are built, or old ones replaced.

- *Dominique Brodbeck is with University of Applied Sciences and Arts Northwestern Switzerland, e-mail: dominique.brodbeck@fhnw.ch*
- *Markus Degen is with University of Applied Sciences and Arts Northwestern Switzerland, e-mail: markus.degen@fhnw.ch*
- *Andreas Walter is with Inselspital, Bern University Hospital, e-mail: andreas.walter@insel.ch*

.

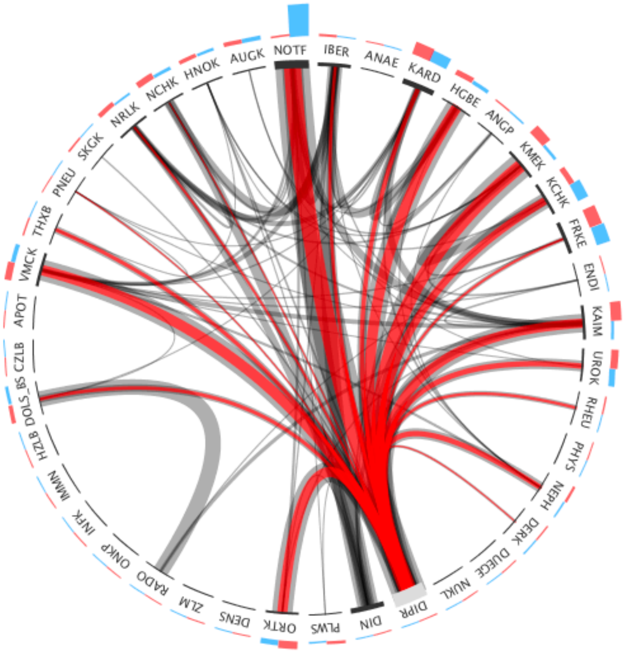


Fig 2. Incoming or outgoing (red) transfers can be highlighted to reveal relationships between clinics. The radiology department (DIPR) plays the role of a service center and is highly connected to the other clinics.

The optimal configuration of departments, their organizational units and technical facilities is not always evident. Questions such as "where should the emergency department be placed, and if we locate it in a new building, do we need an additional radiology facility?" should be answered based on evidence and insights rather than intuition, subjective opinions, or obsolete experience [2]. The idea therefore was to use past real data to identify existing clusters of organizational units that are related based on what they actually do, and not on where they are placed in the organization chart. With these insights, it should be possible to define future sub-systems of organizational units and medical functions, optimized for efficiency. These new sub-systems can then be characterized again with the past data for further analysis and communication to stakeholders.

In this paper, we present a case study where we collected, combined, and enriched data from a large university hospital, and used interactive visualization to access, analyze, and interpret the data to support strategic infrastructure planning.

2 Methods

2.1 Data Aggregation

Large hospitals, and in particular university hospitals, typically have a heterogeneous IT-infrastructure due to the fact, that the different clinics are rather autonomous and have different needs for the type of data to store. Often data of one clinic (i.e., orthopedics) is stored in a specific IT system only used by that specific clinic. In our project, information from several sources was used and linked:

- Inter-organizational transfer histories of stationary patients
- Case attributes (e.g., diagnosis, treatment, diagnosis related group (DRG)) of stationary cases
- Times of surgeries (timestamps at the cut and at the end of suturing)
- Transfers to ambulatory facilities (i.e., radiology department)
- Hierarchical organization of the hospital
- Physical layout of the organizational units of the hospital

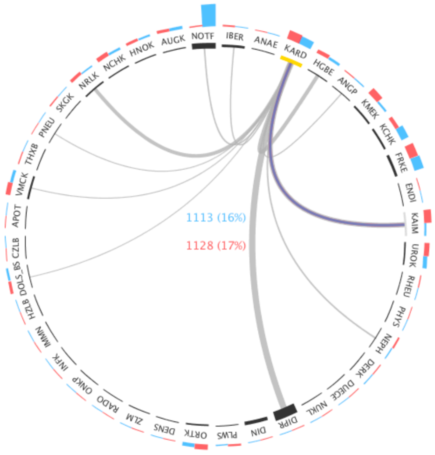


Fig. 3. Limiting the view to only transfers that go in or out of one specific organizational unit show that the cardiology clinic (KARD) is also exhibiting a service center characteristic.

The design of our data store was heavily influenced by the two central dimensions of expected queries:

Multi-Scale: Case data has to be aggregated into several layers to allow drill-down from the hospital level (e.g., number of patient-days per year) to specific organizational units (e.g., is there a seasonal pattern in patients' visits to the pneumology clinic?), and to individual patient cases (e.g., chronology of visits to the radiology department for one specific case).

Multi-Aspect: View the data from different aspects, ranging from the linear temporal view (e.g., chronological view of all events in a patient case), to a two-dimensional geographic map (e.g., where should the radiology department be placed on a campus to minimize travel distances for patients?), to the network topology of relationships between organizational units (e.g., which units transfer the most patients between each other?).

Overall, we collected one full year of data from 40 clinics comprising 300 organizational units that treated 40'000 cases from 30'000 stationery patients, with 320'000 transfers between the organizational units.

2.2 Visual Analytics

With all the data integrated and available, the next challenge was to render it usable for the planning experts. For the type of problems found in our case study, analysts often only have vague notions of what they are looking for ("I know it when I see it"). It is therefore crucial to make the data visible from various angles, and to provide highly interactive tools to identify interesting patterns and access details in context. We developed a visual analytics application to support analysts in making sense of the collected data. The application offers four principal views (Figure 1):

- Organizational: shows the organizational structure and how the actual medical activities shape the administrative space
- Systemic: reveals the operational structure as it emerges from patients flowing through the hospital
- Topographical: shows the actual physical situation as a structure that evolved through many individual decisions
- Chronological: adds the dynamic view on how events and quantities change over time

2.2.1 Organizational View

The organizational view (Figure 1, top left) uses a circular layout to arrange all the major clinics of the hospital. Circular layouts have proven effective to show genetic sequences and relationships between genomic positions [3]. We adapted this technique to show the flow of patients in relation to the organizational structure of the hospital.

The outer circle shows how many patients enter (blue bars) or exit (red bars) a clinic from outside of the hospital. The inner circle

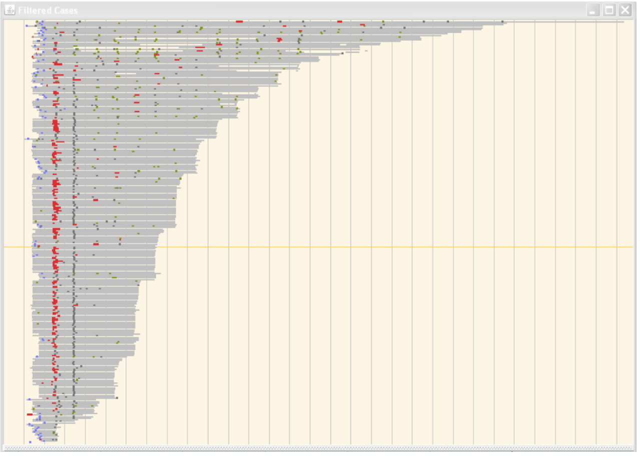


Fig. 4. All the filtered cases can be displayed in parallel to find patterns of similarity, or outliers. This image shows the variation in procedures and duration of all cases for a particular DRG, in this example “craniotomy”.

represents the size of the clinic, as measured by the number of individual cases that passes through that clinic, mapped to the thickness of the black bar. The combined bi-directional flow of patients between two clinics is shown as a curved line, whose thickness is proportional to the number of transfers.

If a clinic is probed (hover with the mouse cursor) then all the incoming or outgoing transfers from that clinic are overlaid (Figure 2). If a clinic is selected then only the transfers to and from that clinic are shown and all the others suppressed (Figure 3). Probing now shows quantitative information about the number of transfers from the selected clinic to the probed one.

2.2.2 Systemic View

The movement of patients between clinics effectively creates a network of relationships, where clinics that move more patients between them are closer, or more similar, than clinics with fewer or no transfers.

To make this network visible, we employ a multidimensional scaling algorithm [4]. Multidimensional scaling is a family of methods that turns information about the similarity of objects into geometric positions in such a way that, as best as possible, similar objects are close together and dissimilar ones far apart. It is particularly well suited for our data because it is able to reproduce non-linear high-dimensional structures in a lower-dimensional (i.e., two-dimensional) geometric representation (i.e., points on a plane).

Once the positions of the clinics on what we now call the systemic view are determined, the transfers are represented analogous to the organizational view in order to emphasize their complementary aspect (Figure 1, top center).

2.2.3 Topographical View

The topographical view (Figure 1, top right) shows the patient transfers on a geographical representation of the current hospital campus. The clinics cannot be represented as single units like in the other views, since the various organizational sub-units of a clinic are not typically located in a single physical location. Distinct locations are therefore symbolized as circles and the organizational sub-units at this location are represented as filled dots within this circle. The layout within a circle is randomized and the dots drawn transparently. This visualization scales well with the greatly varying number of units at a single location.

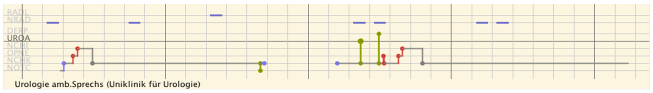


Fig. 5. Individual cases in the filter result can be examined in detail in a visualization of the whole case history.

2.2.4 Chronological View

The time-dependent behavior of the system is shown in the chronological view (Figure 1, bottom right). The in- and out-transfers for each day are shown as a mirrored stacked bar chart. The dark hues in the center close to the time axis show the transfers from (blue, pointing up) and to (red, pointing down) other clinics, whereas the external entries and exits are stacked on top and shown in light hues. The mirroring makes it easy to spot imbalances between in- and out-flows. The net flow for each day is cumulated and over-plotted as a black line. This essentially shows the number of patients that are present in a clinic on a particular day.

2.2.5 Interaction

All the views are coordinated through brushing and linking. In order to rationalize and interpret the insights and hypotheses generated with the four principal views, it is necessary to drill-down to the level of individual cases. Cases can be filtered either by organizational unit that they have visited on their journey through the hospital, or by various categorical or numerical case attributes (e.g., destination after discharge, diagnosis, length of stay). When a case is selected from the list of filtered results, its details are shown both as a table of transfers, as well as a visualization of the whole case history (Figure 5), showing admission, surgeries, radiology procedures, ambulatory visits, and transfers between organizational units (vertical axis) on a time-line (horizontal axis). At the same time, the transfers of the selected case are highlighted (yellow) in the main views (Figure 1).

In a separate view, we show all the filtered cases at the same time. In order to display several hundred case histories in parallel, their representation is condensed to a single line that is only one pixel high, but still preserves the essential information about the case history (Figure 4).

3 Results

The analysis starts with the organizational view (Figure 1, top left) that shows the overview of how patients flow through the hospital system. The outer ring shows that by far the most patients enter the hospital (blue bars) through the emergency department (NOTF), followed by gynecology (FRKE), pediatric surgery (KCHK), and cardiology (KARD). The clinic for pediatric surgery has a negative balance for entries/exits (red bars), which means that like in the emergency department, patients enter the hospital through this unit but then get transferred into other units. The opposite is the case for the children’s clinic, suggesting a typical path for pediatric patients. This is indeed the case, since the clinic for pediatric surgery also operates an emergency room for children, which accounts for most of these transfers. Another feature that pops out is that while the clinic for gynecology (FRKE) has many external entries and exits, it has very few internal transfers. This is due to most of the women giving birth without further complications.

Looking at the transfers on the inside of the circle, it becomes obvious that the institute for diagnostic, interventional and pediatric radiology (DIPR) plays an important role. Highlighting all the transfers that go out of DIPR (Figure 2) shows that it is a service center for many of the hospital’s clinics. Further examination shows a less pronounced but similar pattern for the cardiology clinic (Figure 3). About 17% of its transfers are from and to the clinic for internal medicine (KAIM) and about the same for the clinics of cardiovascular surgery (HGBE), and neurology (NRLK).

The systemic view reveals a number of interesting features of the way that the clinics are related based on the actual flow of patients. In Figure 6 we can see a dense core of highly related clinics in the center, surrounded by a ring of clinics with a less central role, and finally followed by a number of clinics that are very peripheral (e.g., clinics of hematology, osteoporosis, oncology, infectiology).

If we limit the view to clinics that have at least 1000 transfers with any of the other clinics, four groups emerge (Figure 7). At the core we have the emergency department (intersection between

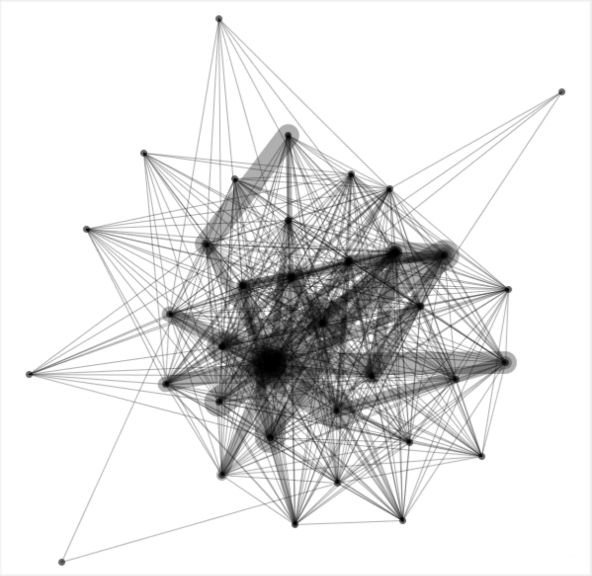


Fig. 6. The systemic view shows a core of highly connected clinics surrounded by six clinics that only play a peripheral role.

groups A and B). Group A encompasses the radiology at the center, and clinics such as cardiology, abdominal medicine, orthopedic surgery, or internal medicine surrounding it. Group B also connects to the emergency department, but groups around clinics such as neuroradiology, intensive medicine, neurology, neurosurgery, or immunology and allergology. Groups C (radio-oncology and oncology wards) and D (children's clinic and pediatric surgery) are somewhat separate and less central.

The chronological view provides a view of the patient flow across time. Looking at the whole hospital (Figure 8, top), it can be seen that the number of patients who stay at the hospital is quite constant (black line), with only minimal seasonal effects. The oscillation pattern is due to the fact that surgeries tend to take place at the beginning of the week (peaks), and patient discharge takes place preferentially before the weekend (lows).

If we look at the emergency department shown in Figure 8 (bottom left), we can see that only a small number of patients stay at this department for a long time (black line) but there are a lot of patients entering directly from outside the hospital (light blue bars) and are transferred to other clinics within the hospital (dark red bars). Figure 8 (bottom right) shows the radiology department. The black line stays at zero because no patients stay overnight in this clinic. This department adds to the weekly pattern found in the overall hospital view, since planned interventions are not performed during the weekend.

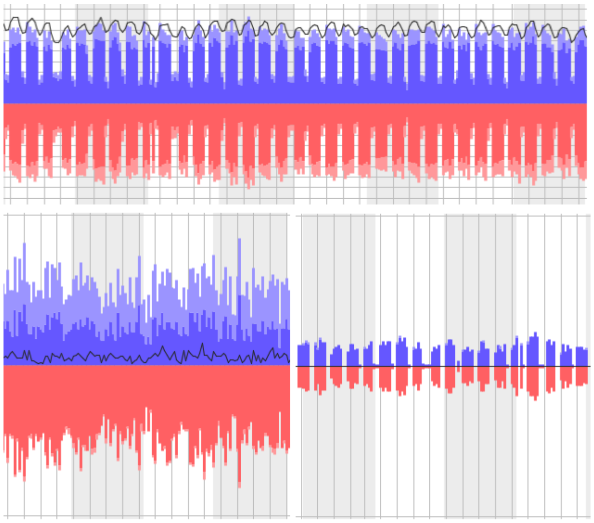


Fig. 8. Extracts from the chronological view. Grid lines denote weeks, with months shown as alternating shaded backgrounds.

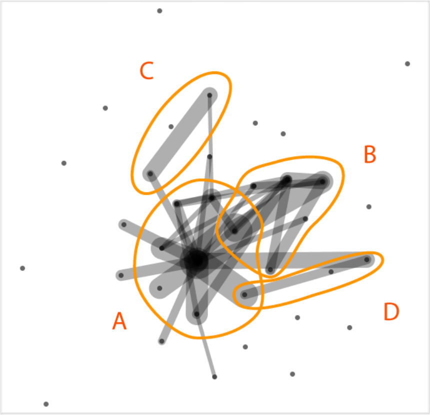


Fig. 7. Four groups of clinics make up the core of the hospital system, as measured by the number of patients that they exchange (threshold at >1000 transfers per clinic per year).

4 Conclusion

In their daily work, people develop an intuition for the relations in a system. The view of the observers however is often limited to their sphere of action. The dependencies on the next or next-to-next source of influence are not taken into account sufficiently. Our analysis with this application however allowed us to gain an overview of the big picture of the hospital system.

By making the flow of patients visible, we were able to contrast the hierarchical organizational structure with the actual implemented working relationships. This showed the difference between the operational structures that developed through medical consequences, and the theoretically defined organizational structure. Based on this difference, we were able to describe new sub-systems and identify an organizational form that corresponds to the current actual needs.

It was not really a surprise that the core functions of a hospital such as emergency department, operating rooms, and diagnostic functions appeared in the center of the system, but it was not expected to be so pronounced. A new insight was the role of the cardiology clinic as an important service center for diagnostics. This led to the decision to also assign it a central role on the campus. Also new was the interpretation of the role of the clinic for internal medicine as being primarily a receiving station for the emergency room, with the further distribution into the specialized clinics taking place only one or two days later.

In summary, our case study has shown that there is a wealth of interesting information in the data that is collected in large hospitals, beyond their immediate and intended use. We took a different view on electronic health records to support strategic infrastructure planning.

References

[1] Chaudhry, B., Wang, J., Wu, S., Maglione, M., Mojica, W., Roth, E., Morton, S. C., and Shekelle, P. G. (2006). Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med, 144(10):742–752.

[2] van der Aalst, W. M. P. (2012). Process mining. Communications of the ACM, 55(8):7683.

[3] Krzywinski, M. I., Schein, J. E., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., and Marra, M. A. (2009). Circos: An information aesthetic for comparative genomics. Genome Research.

[4] Chalmers, M. (1996). A linear iteration time layout algorithm for visualising high-dimensional data. In Proceedings of the 7th conference on Visualization '96, VIS '96, pages 127 ff., Los Alamitos, CA, USA. IEEE Computer Society Press.

Designing an Open Source Presentation Layer for the Patient-Centered Medical Home

JIHOON KANG, MFA, ASSOCIATE DIRECTOR,
PARSONS INSTITUTE FOR INFORMATION MAPPING

ABSTRACT

The Parsons Institute for Information Mapping (PIIM) through funding provided by the Telemedicine & Advanced Technology Research Center (TATRC) is developing a widget-based prototype for the Patient-Centered Medical Home (PCMH) environment. Through a collaboration with the Walter Reed National Military Medical Center (WRNMMC), the prototype provides both a patient and provider portal. This presentation layer is streamlined for easy access to medical information and interaction between patients and healthcare providers. We hope to show how through an Open Source environment and a better mapping of data sources and providers, better design can contribute to more manageable EHR environment. This presentation will show some results of our recent usability testing efforts and demonstrate the prototype. The designs, code base, and documentation are planned for an Open Source release in December 2012 through the OSEHRA framework.

INTRODUCTION

The Parsons Institute for Information Mapping (PIIM), of The New School is developing a widget-based prototype for the Patient-Centered Medical Home (PCMH) Environment. The prototype, Healthboard, is being developed in Flex/Flash and will serve to provide a one-stop visual dashboard for patients and providers. Developed specifically to enhance the communication of information through better visualization of medical data, the tool strives to make difficult data much easier to understand by utilizing best-practice design principles and standardization of information provided through the system.

Healthboard is scheduled for a planned release within the Open Source Electronic Health Record Agent (OSEHRA) platform. We hope to establish Healthboard as an alternative presentation layer for electronic medical records (EMR) or for use directly within the PCMH environment. The prototype and design documents will be provided to assist developers in providing an enhanced level of support, better user experience and a streamlined interface between patients and their healthcare providers. Because the system follows a modular structure, some or all of Healthboard may be implemented depending on the need of the particular program.

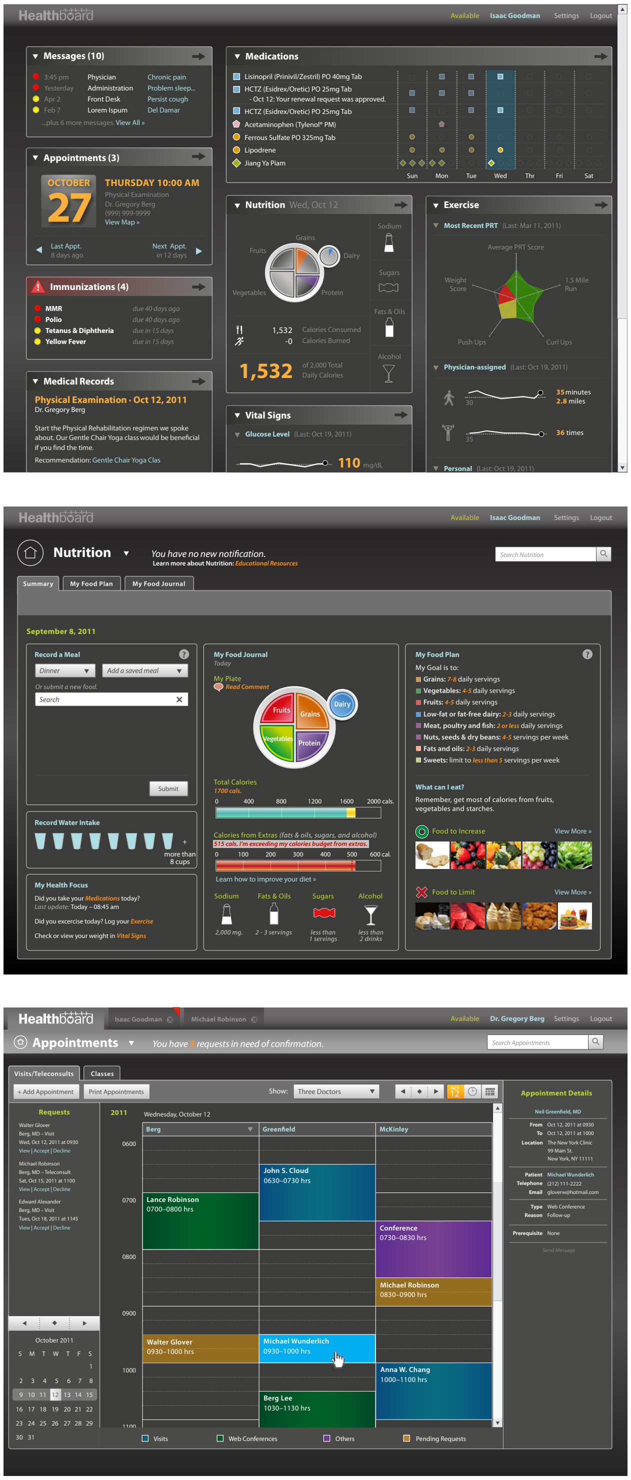


FIGURE 1: *Screen shots from Patient Portal and Providers Portal of Healthboard*

BACKGROUND

In 2008, PIIM began a project with the Telemedicine & Advanced Technology Research Center (TATRC) to develop an enhanced graphic user interface (GUI) for the Armed Forces Health Longitudinal Technology Application (AHLTA).[1] PIIM's work was to identify opportunities to improve the GUI and the overall visualization of medical information available in AHLTA. PIIM was also responsible for performing a review of the usability of the system and provide recommendations for improving the system's use amongst end users. Finally, PIIM developed a prototype that represented all of these areas to demonstrate the effect that a redesigned GUI could have on the system.

In 2009, PIIM began working on the Visual Dashboard and Heads-up Display of Patient Conditions award, which ultimately created the Healthboard system described herein. The project involved developing a visual style for the patient and provider dashboards used within a Patient-Centered Medical Home Environment, developing a user experience strategy, engineering a prototype, and finally performing usability testing and redesigning the interface based on user feedback.

DEVELOPMENT OF THE HEALTHBOARD PROTOTYPE

The goal of Healthboard was to ultimately provide a way for active duty military personnel and their spouses to interact with their own personal health information and electronic medical records. Furthermore, Healthboard was to provide a streamlined mechanism for patients and providers to interact. Our development effort began with the following high-level requirements:

- To enhance communication between the patient and providers
- To allow the patient to gain easy assess to his/her own health records
- To enable heath data self-reported by the patient
- To help the patient gain health literacy and awareness

By integrating best-practice design principles and data visualization strategies, the enhanced GUI design and its attention to user friendliness and usability targets should provide a better overall user experience for patients as they interact with medical information. As a result, medical information could be streamlined. To some degree, the intimidation factor of viewing such information could be partially mitigated for non-medically trained users.

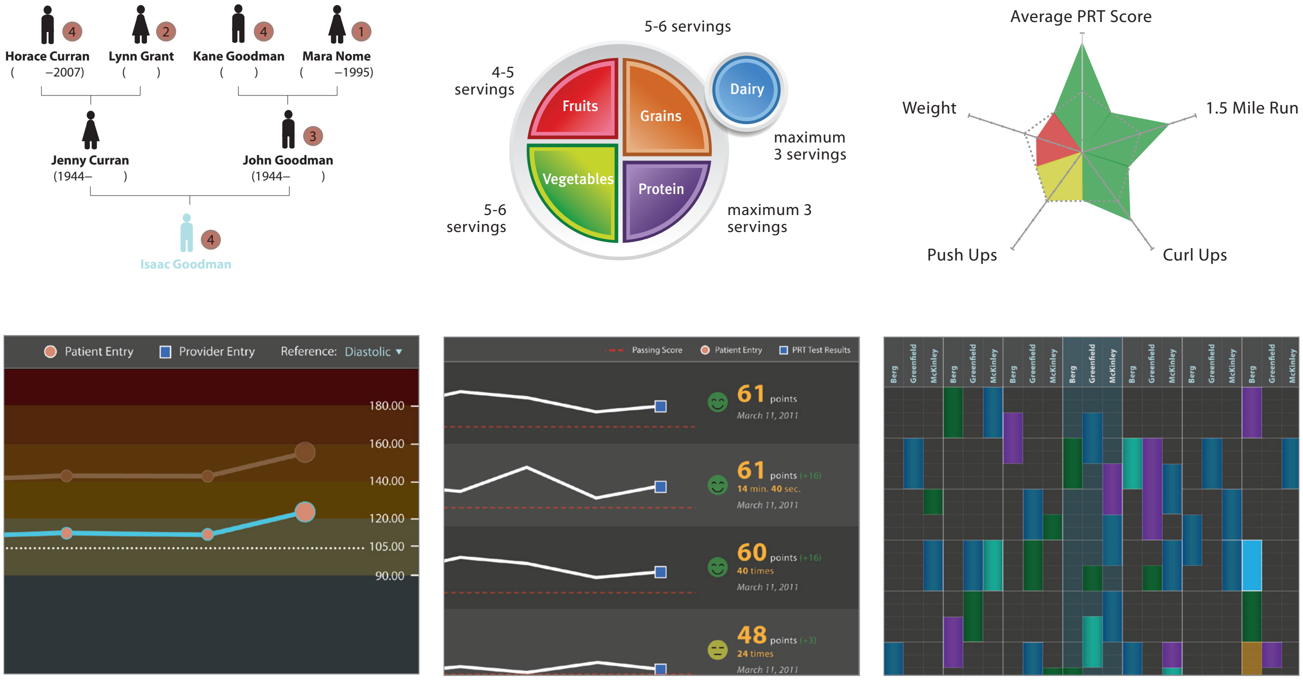


FIGURE 2: *UI components related to data visualization*

COLLABORATION

Working with project stakeholders and Military Health System (MHS) representatives, PIIM developed a Product Requirements Document (PRD) and Project Management Plan (PMP). Developing iterative versions of three primary documents, the *Detailed GUI Design Volume,*[2] *Information Strategy Volume,*[3] and *Engineering Volume,*[4] these projects served to provide the basis and backb one for the eventual prototype.[5] As the Patient-Centered Medical Home Environment within Walter Reed National Military Medical Center (WRNMMC) was identified as the primary use case for the development of the Healthboard prototype, PIIM worked with the medical home program at WRNMMC and Telemedicine & Advanced Technology Research Center (TATRC). The following experts in each organization have contributed towards the successful completion of this project:

Parsons Institute for Information Mapping (PIIM)

- *GUI Designers*
- *UX Designers*
- *Information Designers*
- *Usability Specialists*
- *Medical Informatics Specialists*
- *Engineers*

Walter Reed National Military Medical Center

- *Physicians*
- *Nurses*
- *Dietitians*
- *Pharmacists*
- *IT Specialists*
- *Hospital Administrators*

Telemedicine & Advanced Technology Research Center

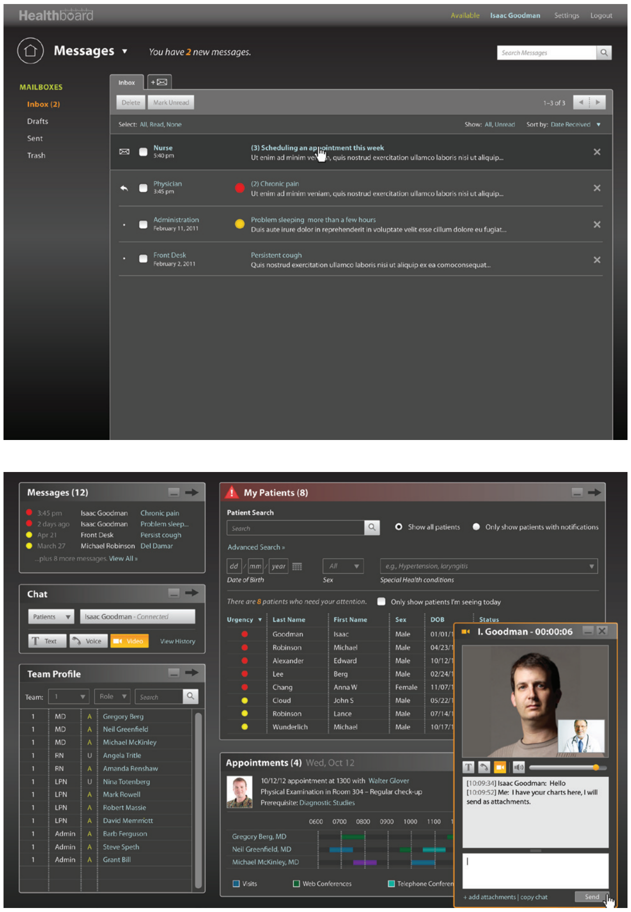
- *MHS Subject Matter Experts*
- *Program Mangers*
- *Grant Managers*

PIIM developed user requirements and use cases based off of feedback from collaborators. In addition, feedback was received through weekly teleconferences held with a team of reviewers, where the iterative GUI designs and other works were shared. Each module was presented as it was being developed, and additional medical expertise was consulted as necessary to inform the design team.[6]

DESIGN OUTCOMES

PIIM successfully delivered GUI/UX models satisfying the high-level requirements listed above through the design iterations. The following modules/UI features have been identified and designed *to enhance communication between the patient and providers:*

- Messages
- Live Chat (video, audio, text)
- Appointments
- Reminders
- Medications (requesting Renewals



You have 2 new messages.

Select: All, Read, None · Show: All, Unread · Sort by: Date Received ▾

✉	Nurse 5:40 pm		(3) Scheduling an appointment this week Ut enim ad minim ve[illegible]n, quis nostrud exercitation ullamco laboris nisi ut aliquip...
↩	Physician 3:45 pm	🔴	(2) Chronic pain Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip...
·	Administration February 11, 2011	🟡	Problem sleeping more than a few hours Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat...
·	Front Desk February 2, 2011		Persistent cough Quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea comoconsequat...

🔴	3:45 pm	Isaac Goodman	Chronic pain
🔴	2 days ago	Isaac Goodman	Problem sleep...
🟡	Apr 21	Front Desk	Persist cough
🟡	March 27	Michael Robinson	Del Damar

...plus 8 more messages. View All »

Isaac Goodman - Connected

1	MD	A	Gregory Berg
1	MD	A	Neil Greenfield
1	MD	A	Michael McKinley
1	RN	U	Angela Tritle
1	RN	A	Amanda Renshaw
1	LPN	U	Nina Totenberg
1	LPN	A	Mark Rowell
1	LPN	A	Robert Massie
1	LPN	A	David Memmott
1	Admin	A	Barb Ferguson
1	Admin	A	Steve Speth
1	Admin	A	Grant Bill

There are 8 patients who need your attention. □ Only show patients I'm seeing today

Urgency ▾	Last Name	First Name	Sex	DOB	Status
🔴	Goodman	Isaac	Male	01/01/1	
🔴	Robinson	Michael	Male	04/23/1	
🔴	Alexander	Edward	Male	10/12/1	
🔴	Lee	Berg	Male	02/24/1	
🔴	Chang	Anna W	Female	11/07/1	
🟡	Cloud	John S	Male	05/22/1	
🟡	Robinson	Lance	Male	07/14/1	
🟡	Wunderlich	Michael	Male	10/17/1	

10/12/12 appointment at 1300 with Walter Glover
Physical Examination in Room 304 – Regular check-up
Prerequisite: Diagnostic Studies

[10:09:34] Isaac Goodman: Hello
[10:09:52] Me: I have your charts here, I will send as attachments.

FIGURE 3: *Examples of modules and components designed to enhance communication between patients and providers*

The following modules/UI features have been identified and designed *to allow the patient to gain easy assess to his/her health records:*

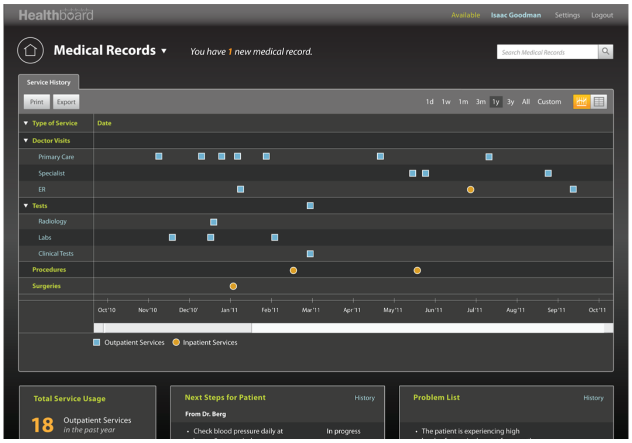
- Medical Records (visit summary, test results, procedure/surgery records)
- Immunization
- Medications (prescription medications)
- Vital Signs (taken during visits)

The following modules/UI features have been identified fied and designed *to enable heath data self-reported by the patient:*

- *Vital Signs (taken by the patient)*
- *Nutrition*
- *Medications (over-the-counter, supplements, herbal medicines)*
- *Exercise*

The following modules/UI features have been identified and designed *to help the patient gain health literacy and awareness:*

- *Educational Resources*
- *Nutrition (provider-entered recommendations and nutrition guides)*
- *Context-sensitive help and tips*



You have 1 new medical record.

Total Service Usage

18 Outpatient Services *in the past year*

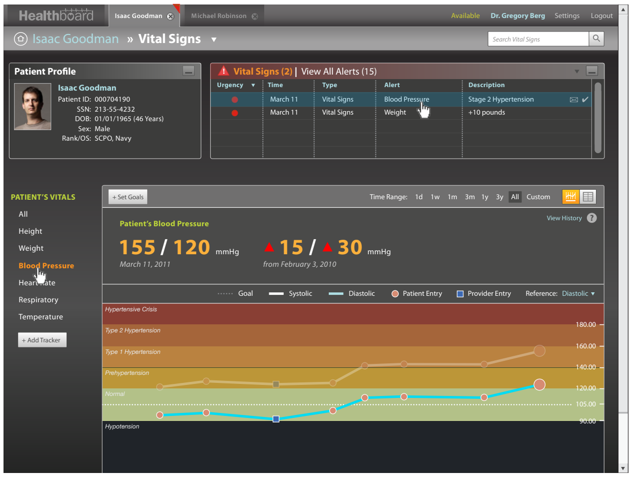
Next Steps for Patient

From Dr. Berg

- Check blood pressure daily at — In progress

Problem List

- The patient is experiencing high



Isaac Goodman
Patient ID: 000704190
SSN: 213-55-4232
DOB: 01/01/1965 (46 Years)
Sex: Male
Rank/OS: SCPO, Navy

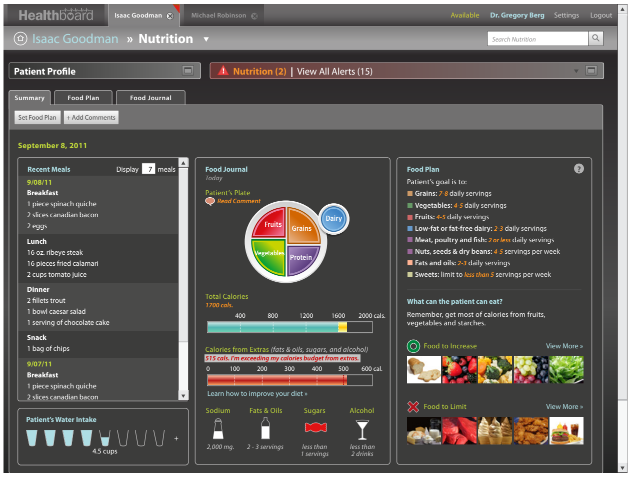
Vital Signs (2) | View All Alerts (15)

Urgency	Time	Type	Alert	Description
●	March 11	Vital Signs	Blood Pressure	Stage 2 Hypertension
●	March 11	Vital Signs	Weight	+10 pounds

Patient's Blood Pressure

155 / 120 mmHg
March 11, 2011

▲15 / ▲30 mmHg
from February 3, 2010



Nutrition (2) | View All Alerts (15)

September 8, 2011

Recent Meals — Display 7 meals

9/08/11
Breakfast
1 piece spinach quiche
2 slices canadian bacon
2 eggs

Lunch
16 oz. ribeye steak
16 pieces fried calamari
2 cups tomato juice

Dinner
2 fillets trout
1 bowl caesar salad
1 serving of chocolate cake

Snack
1 bag of chips

9/07/11
Breakfast
1 piece spinach quiche
2 slices canadian bacon

Patient's Water Intake
4.5 cups

Food Journal
Today

Patient's Plate
Read Comment

Total Calories
1700 cals.

Calories from Extras *(fats & oils, sugars, and alcohol)*
515 cals. I'm exceeding my calories budget from extras.

Learn how to improve your diet »

Sodium	Fats & Oils	Sugars	Alcohol
2,000 mg.	*2 - 3 servings*	*less than 1 servings*	*less than 2 drinks*

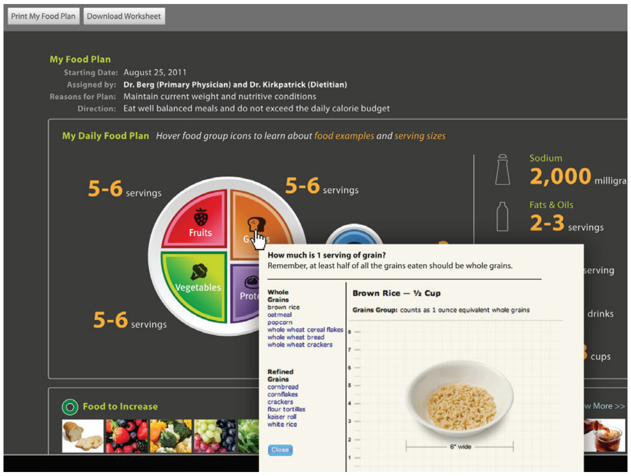
Food Plan

Patient's goal is to:

- Grains: 7-8 daily servings
- Vegetables: 4-5 daily servings
- Fruits: 4-5 daily servings
- Low-fat or fat-free dairy: 2-3 daily servings
- Meat, poultry and fish: 2 or less daily servings
- Nuts, seeds & dry beans: 4-5 servings per week
- Fats and oils: 2-3 daily servings
- Sweets: limit to less than 5 servings per week

What can the patient can eat?

Remember, get most of calories from fruits, vegetables and starches.



My Food Plan
Starting Date: August 25, 2011
Assigned by: Dr. Berg (Primary Physician) and Dr. Kirkpatrick (Dietitian)
Reasons for Plan: Maintain current weight and nutritive conditions
Direction: Eat well balanced meals and do not exceed the daily calorie budget

My Daily Food Plan *Hover food group icons to learn about food examples and serving sizes*

How much is 1 serving of grain?
Remember, at least half of all the grains eaten should be whole grains.

Whole Grains
brown rice
oatmeal
popcorn
whole wheat cereal flakes
whole wheat bread
whole wheat crackers

Refined Grains
cornbread
cornflakes
crackers
flour tortillas
kaiser roll
white rice

Brown Rice — ½ Cup
Grains Group: counts as 1 ounce equivalent whole grains

FIGURE 4: *Screen shots from modules supporting self-reported data and context-sensitive help and tips*

MOVING TOWARDS AN OPEN SOURCE ENVIRONMENT

Following conversations with project stakeholders, PIIM made a decision in early 2012 to begin planning for the eventual move of the Healthboard prototype to an open source environment. In August 2012, PIIM began conversations with representatives of the Open Source Electronic Health Record Agent project (OSEHRA). OSEHRA is built on the notion of facilitating the development, improvement, and maintenance of EMR platforms, making them freely available for medical beneficiaries.[7]

In December 2012, PIIM plans to release the prototype and design documents through the OSEHRA framework, making them available as an alternate presentation layer. This presentation layer may ultimately prove useful as an informative tool to the development and designs of future MHS systems and initiatives, as well as to the development of systems outside the U.S. Military.

CONCLUSION

Healthboard is a unique system designed to serve both patients and providers with a better user experience among other things through intuitive interface, data visualization (for better comprehension of information and decision-making), and end-user support while respecting the workflow of the remote patient-care process. As PIIM continues with the usability testing and related research on Healthboard, we hope that the final prototype and its supporting documents will help establish a benchmark for how patients and healthcare providers can interact through better design by December 2012. The system has met most of the requirements that were defined at the initiation of the project, as well as the requirements stated during the iterations. We expect the health professionals will continuously provide quality healthcare while limiting the number of doctor's visits once this system is successfully deployed and utilized in reality. The system would help patients stay healthy, prevent illnesses, gain health literacy and awareness, and monitor their own health records, as well as allow them to self-report health-related activities and educate themselves through online resources. The key to achieving such meaningful goals is a collaborative effort; the designers play a significant role in an EMR system's design or redesign, they cannot take the task alone, and neither can the engineers nor the clinical experts. It was successful only through the successful formation of a team of clinical experts, designers, usability experts, engineers, IT experts, and administrative experts. Ultimately, we hope that the collaborative effort contributed by PIIM, WRNMMC, and TATRC will become a major precedence for the practice of designing the EMR system.

ACKNOWLEDGEMENTS

The author would like to thank our TATRC and WRNMMC project stakeholders who have provided invaluable guidance during our design and development process. We would also like to thank the PIIM Design and Engineering teams for their continued support and hard work during the development of this project.

This work is supported by the Telemedicine and Advanced Technology Research Center (TATRC) at the U.S. Army Medical Research and Materiel Command (USAMRMC) through award W81XWH-11-2-0025.

NOTES

[1] Award No. W81XWH-09-1-0456: Leveraging the PIIM Process to Advance Health IT. The Telemedicine and Advanced Technology Research Center (TATRC).

[2] Jihoon Kang et al. The Visual Dashboard & Heads-up Display of Patient Conditions: GUI Design Volume. New York: Parsons Institute for Information Mapping, The New School, 2012.

[3] J Kang and D Bendersky. The Visual Dashboard & Heads-up Display of Patient Conditions: Information Strategy Volume. New York: Parsons Institute for Information Mapping, The New School, 2012.

[4] D Bendersky and J Kang. The Visual Dashboard & Heads-up Display of Patient Conditions: Engineering Volume. New York: Parsons Institute for Information Mapping, The New School, 2012.

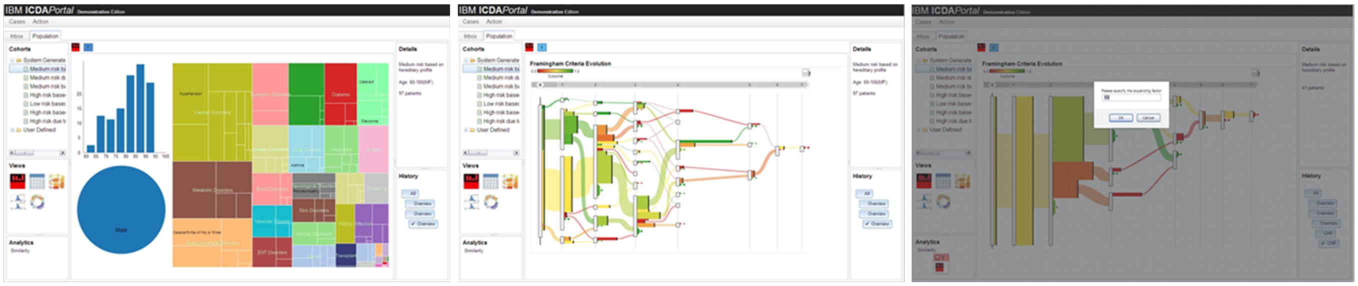
[5] B. Willison. "Quarterly Report: Leveraging the PIIM Process to Advance Health IT." Quarterly Report, Parsons Institute for Information Mapping, New York, NY, June–September, 2009.

[6] J Kang, S Yoshida, A Ina. The Visual Dashboard & Heads-up Display of Patient Conditions: Assessment Volume. New York: Parsons Institute for Information Mapping, The New School, 2012.

[7] OSEHRA, "About OSEHRA," About Us, http://www.osehra.org/page/about-us, 2012.

Interactive Visual Patient Cohort Analysis

Zhiyuan Zhang, David Gotz, and Adam Perer



(a) (b) (c)

Figure 1. (a) Summary view of a cohort of males over 60, visual filtered from an initial cohort of "medium risk of heart failure" patients. (b) The same sub-cohort visualized for symptom progression. (c) After further filtering reduces the cohort below a statistically significant size, similarity analytics can be used to grow the cohort by querying for additional similar patients.

Abstract—Retrospective patient cohort analysis is widely used in many healthcare studies. Due to its data intensive nature, the traditional analytical pipeline requires expertise from several areas, such as databases, data mining, software development, statistics, and domain knowledge. As a result, domain experts often rely on a team of technologists to help perform such studies which can make the process slow and cumbersome. To allow domain experts to perform faster and more flexible analyses, we designed an integrated system that combines visual exploration and data analytics with an intuitive user interface. Our system lets clinicians interactively visualize and refine cohorts, request analytics on those cohorts, and make new discoveries.

Index Terms—Cohort Study, Retrospective Patient Cohort Analysis, Visual Analytics, Interactive Cohort Definition and Refinement.

1 Introduction

Retrospective patient cohort analysis [2] is the analysis of patients' medical and diagnostic histories to make healthcare discoveries. In the traditional pipeline, analysts work manually to define specific cohort constraints (e.g., "female patients over age 70") or apply specialized batch analytics to computationally determine a meaningful group of patients (e.g., high-utilization cohorts [1]). Unfortunately, both methods have limitations. For the definition of the cohort constraints, it's difficult to select the attributes that are to be queried from a list of hundreds or thousands of patient attributes. For batch analytics that behave like a "black box", users have few ways to apply their domain expertise to influence the process.

Once a patient cohort has been defined, the next step in the analysis process is to apply specific statistics or data mining techniques to the cohort and look at the results to uncover insights. These steps can often be unintuitive for clinical users. As a result, exploratory analysis can often require several iterations of work between domain experts and computational staffs. This can significantly slow down the process and limit the clinical analyst's flexibility to explore.

To address these challenges, we have designed an integrated system that combines visual exploration and data analytics with an intuitive user interface. The system empowers clinical users to quickly and efficiently perform interactive visual patient cohort analyses. Using our system:

- Patient cohorts can be interactively defined and modified at any step of an analysis;
- Cohorts can be visualized in various ways and users can pivot easily between different visualization metaphors; and
- Analytics can be applied to cohorts for on-demand processing at any time in an analysis,

The system's user interface supports these tasks by allow direct manipulation of three key artifacts: (1) ***cohorts***, (2) ***views***, and (3) ***analytics***. Cohorts represent sets of patients and their associated information. Views are visualization components used to graphically represent and interactively refine the cohorts. Analytics operate on cohorts and are used to generate new cohorts, produce additional data for a specific cohort, or to otherwise modify (e.g., expand or segment) an existing cohort. The remainder of this paper describes our system design and user interaction model, and presents a use case that demonstrates the utility of our approach.

2 System Design

Our system design manages three types of artifacts—cohorts, views, and analytics—and connects them within a single integrated user interface. This section describes the artifact types in more detail and discusses how they connect within our architecture.

2.1 Cohorts

Patient cohorts are groups of patients and their associated information, such as gender, age, diagnoses, and treatments. A cohort serves as the underlying data structure that is used to pass data throughout the system's pipeline. They are the objects on which the other two artifacts—analytics and views—operate. Included within each patient cohort representation is a list of individual patient identification numbers that can be used to connect cohort members with more detailed clinical data located in a remote data store.

2.2 Analytics

Analytics are computational components that operate on cohorts in various ways. Our system supports two main types of analytics: (1) ***batch analytics*** and (2) ***on-demand analytics***. Batch analytics are components that are executed automatically in the background by our system (e.g., nightly as new patient data is imported to the system). They process an entire patient population and identify groups of interest. For example, a batch analytic may be used to perform risk stratification, generating lists of patients that have common sets of risk factors. The batch analytics components generate new cohorts that can serve as starting points for a user's exploratory analysis.

• Zhiyuan Zhang is with Stony Brook University.
• David Gotz and Adam Perer are with IBM Research.

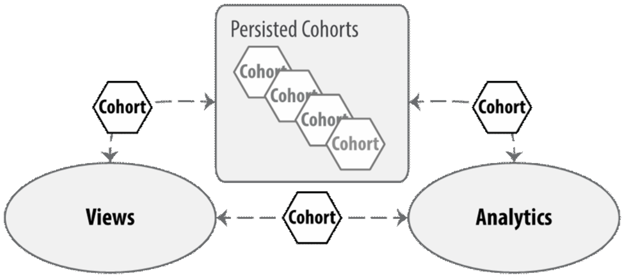


Figure 2. Patient cohorts are passed between views and analytics via drag-and-drop interactions as an analysis unfolds. Cohorts can also be persisted for future reference.

On-demand analytics, in contrast, are performed in ad-hoc fashion at the specific request of a user. On-demand analytics take as input a specific patient cohort, plus an optional set of input parameters. In response, an on-demand analytics can produce additional information about patients in the cohort (e.g., calculate risk scores) and/or refine the membership of the cohort (e.g., query for additional similar patients).

2.3 Views

Views are visualization components that offer specific targeted ways to graphically depict and interact with a patient cohort. Each view is designed to take a single cohort as input and render a specific subset of patient features. Views also provide interactive capabilities through which users can selectively brush and filter to explore and refine the set of patients in the cohort.

For example, our system includes a patient cohort summary view that depicts general information about a group of patients such as age and gender distributions along with Treemap [3] summarizing diagnosis code statistics. This view provides multiple coordinated visualizations through which users can refine the set of patients in a cohort (e.g., "filter to only male patients over age 50 with specific classes of cancer"). Our system also provides a generic table view to look at a detailed list of patients in a cohort including individual patient identification numbers.

Beyond these generic views, additional components are provided for use-case specific visualizations. For example, another view provided by our prototype is the Outflow visualization for exploring patient symptom evolution [4].

Each view has the additional ability to export the set of patients being visualized at any given point in time. Therefore, from a data perspective, views are very much like on-demand analytics in that they both take a cohort as input and produce a cohort as output.

2.4 Interactions

The similarity in data flows for both on-demand analytics and views are critical to the design of our system. This commonality allows users to chain together views and analytics—both serving as operators on cohorts—into arbitrary sequences. This lets users interactively perform complex and ad hoc exploratory analysis processes that mix visual interactions and filtering with computational analysis routines. The approach is illustrated in Figure 2.

Users interact with our system primarily via a drag and drop model that connects our three types of artifacts. Users drag cohorts to views to visualize them, and drag cohorts to analytics to process them. Users can select a cohort from either the current view or from a list of saved cohorts in a sidebar. The sidebar contains both system generated cohorts (via batch analytics) and those that have been manually defined (via prior user interaction).

In addition, individual views allow selection, brushing, and filtering. Callouts and a dedicated sidebar panel are also used to provide more information about moused-over elements. A user's analytic history is summarized in a sidebar, capturing the provenance of the currently viewed cohort and allowing a user to revisit prior stages of his/her investigation.

3 Use Case Demonstrating Typical Workflow

We have developed a prototype implementation of our design as a web-based interactive visualization application. This section reviews an example use case (see Figure 1) to demonstrate the typical workflow that our system supports.

A session starts with a user selecting a cohort from the sidebar located on the left side of the user interface, such as a group of patients flagged as being at risk of developing heart disease. The user can drag and drop the cohort from the sidebar onto the view icon for our cohort summary visualization. The user can interactively explore the aggregate information about the cohort and apply filters to modify the cohort for a specific analytic task. For example, Figure 1(a) shows the summary view after filtering to only male patients over the age of 60.

The cohort created by the filtering step can then be dragged-and-dropped onto an Outflow view to visualize the variations in disease progression for the set of identified patients. It is clear from the view shown in Figure 1(b) that significant variations have been observed. A user can then select a specific pathway from the Outflow visualization and apply additional filters. For example, a user might select the largest pathway, which has mixed outcomes to perform further analysis.

This iterative filtering process using multiple views of the data can let users quickly and intuitively identify a population of interest. However, additional information is often needed that was not contained in the original cohort. For example, in the sequence described here the user has applied several filters that have reduced the size of the cohort population significantly. This reduces its statistical power.

In order to re-grow the patient population, which would allow the user to draw more meaningful conclusions, the user can take advantage of our system's on-demand analytics to retrieve additional patients that are similar to those in the current cohort but that were left out of the initial cohort that was first used to start the investigation.

On-demand analytics are initiated when a user drags the cohort from the current view (or a persisted cohort from the sidebar) to the analysis component of their choice. All available on-demand analytics are listed in the lower left sidebar. After a cohort is dropped on a specific analytic component, the system immediately begins the analysis process. If additional input parameters are required by a given analytics, a dialog box is displayed to gather the needed user input. For example, Figure 1(c) shows the dialog box shown for our system's patient similarity analytic component. This analysis algorithm needs to know an "expansion factor" that specifies how many similar patients to retrieve as it grows the input cohort. For example, an expansion factor of 0.2 will grow the size of a cohort by 20%.

After the similarity computation completes, an expanded cohort is returned and immediately visualized using the same view that was active prior to the analytics request.

References

[1] J. Hu, F. Wang, J. Sun, R. Sorrentino, and S. Ebadollahi. "A Healthcare Utilization Analysis Framework for Hot Spotting and Contextual Anomaly Detection." AMIA, 2012.

[2] M. Porta (editor). A Dictionary of Epidemiology. 5th. ed. New York: *Oxford University Press*, 2008.

[3] B. Shneiderman. "Tree Visualization with Tree-Maps: 2-d Space-filling Approach." *ACM Transactions on Graphics*, 11(1), pp. 92-99, 1992.

[4] K. Wongsuphasawat and D. Gotz. "Exploring Flow, Factors, and Outcomes of Temporal Event Sequences with the Outflow Visualization." *IEEE Information Visualization,* 2012.

ReportViz: Interactive Visualization and Exploration of Topics and Keywords in Public Health Reports

Wei Zhuo*

Graphics, Visualization and Usability Center
College of Computing
Georgia Institute of Technology

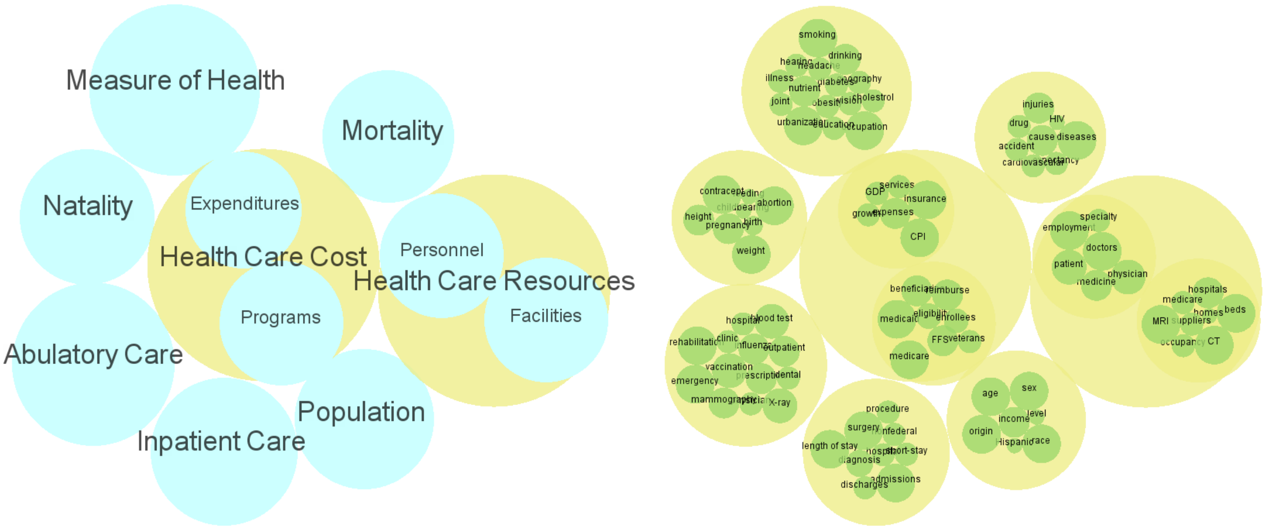


Figure 1: We show topics in a health report as gathered bubbles on the left. Expanding all topic bubbles spawns groups of key word bubbles within each topic bubble (right). The bubble radius encodes the term weight.

Abstract

Public health documents contain a rich set of topics and keywords that cover various aspects of a nation's health system. Each topic is characterized as a distribution over vocabulary from which we extract the keywords. Visualizing this topic structure allows the reader to have an intuitive and convenient overview of a usually lengthy report. In this paper, we describe *ReportVis*, an interactive utility that uses nested bubbles to represent topics, key words and their weights. We use a collection of US Health Reports issued by CDC from 2009 to 2012 as our exemplar corpus to demonstrate a interactive exploration of topics and their keywords in public health. We further introduce an animation mode to allow a dynamic view of topic evolution.

Keywords: Topic Visualization, Key Words in Health

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Animations; I.3.6 [Computer Graphics]: Methodology and Techniques—Interaction techniques

1 Introduction

Health organizations regularly issue reports that contain a rich set of subjects covering various aspects of a health system. As the collection continues to grow, it becomes more challenging to mine topics of interest. We search with keywords to find target texts in a document. In addition, we might want to ① explore texts associated with keywords within the same topic and ② have an overview of the topic structure in this document. All these tasks call for an intuitive interface to visualize the topics/keywords and their weights mined from a collection of reports.

We build ReportViz that aim to address these tasks. We choose a collection of comprehensive health reports issued yearly by Centers for Disease Control and Prevention. A nice feature of this collection is its consistency: the set of topics are well-defined. Although different reports may have different focuses and sometimes multiple topics are nested under a broader theme, the total number of topics is fixed and is shown in Tab. 1. We first extract, for each topic, a list of key words with their weights, and then visualize the topic structure. We restrict our focus to visualization in describing ReportViz. Our work claims the following contributions. First, we represent each topic as a tree node and keywords as leaf nodes. This allows convenient operations for topic merge and split. Second, we use physics-inspired layout and interaction techniques that provide engaging experiences for topic exploration and an animation mode for visualizing topic evolution. Last but not least, we show lists of extracted key words as well as the revision made by medical students. The result could serve as future references or set the priors

*wzhuo3@cc.gatech.edu

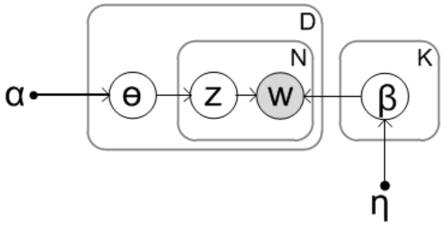


Figure 2: graphical model of Latent Dirichlet Allocation

for topic modeling in health.

2 Prior Art

2.1 Statistics and Text Analysis

Probabilistic topic modeling aims to automatically extract topics and keywords from a collection of documents. The Latent Dirichlet Allocation (LDA [3]) assumes prior distributions (θ, β) over topics and vocabulary, and estimates these parameters from the observed text (w), as shown in Fig. 2.

Tractable implementations includes Gibbs sampling [11] and variational inference [1]. Dynamic Topic Models [2] are a family of probabilistic time-series models developed to analyze the time evolution of topics in a collection of documents ordered by time. Its state space models inspire us to design a dynamic view of the time-varying weights evaluated from texts.

2.2 Visualization

Off-the-shelf visualization models such the tag cloud in ManyEyes [12] can provides web-based visualization of word frequencies: the bubble chart displays a set of numeric values as circles, hence can be used to represent key terms with their frequencies. For compactness, we use physics-inspired techniques to layout the bubbles.

Note that the CVT energy function can be exploited for general icon layout [4]. Cui et al. [5] point out that topic could merge or split over time, and present a static view of topic evolution as a flow graph. Liu et al. [6] use stacked graph to visualize topic evolution by summarizing emails in different times with the output of LDA on the whole corpus.

3 Visualization

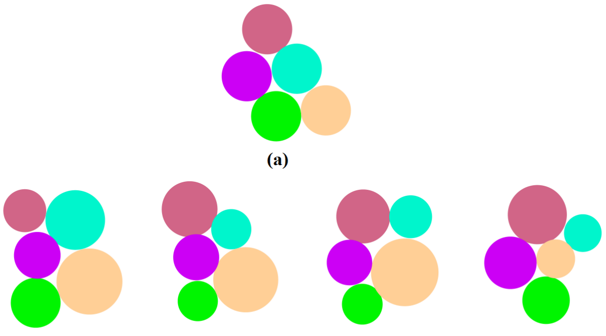
3.1 Health Document

Our text corpus consists of yearly health reports issued by CDC from 2009 to 2012 [7] [8] [9] [10]. There are about 2300 pages presenting analysis and tables by topics. The major topics are: *Population*, *Fertility and Natality*, *Mortality*, *Measure of Health*, *Ambulatory Care*, *Inpatient Care*, *Personnel*, *Facilities*, *Expenditures* and *Coverage and Programs*. In addition to lists of automatically extracted keywords, we ask two medical students to input their revision. The revision tasks are:

(1) **Selection**: within each topic and subtopic, we let our colleagues choose 3-10 terms which they think are most related to the topic.

(2) **Addition**: we also ask them to add any term which they think is as relevant as those selected in (1), but is not in the list of extracted key terms.

The resulting group of topics and lists of key terms are shown in Tab. 1.



(b)

Figure 3: (a) disks are initially of the same radius. (b) disks are changing their radius while remain repelled and gathered.

3.2 Representation and Layout

Each topic has a group of keywords. Also, each topic may have subtopics or belong to a broader theme in health reports. Motivated by these observations, we use a tree-node to depict each topic. The root node denotes the whole document and has a array of child nodes corresponding to a group of topics in the document. A leaf node represents a key word and has no child nodes. The root node is not visualized. Each non-root node is visualized as a circle with its weight encoded as the radius. To keep a group of circles together without overlapping, we compute the disk centers iteratively depending on the following geometric relations:

- **Repel**: For each pair of disks (e.g. D_1 and D_2) within a group, if the distance between their centers is smaller than the sum of their radii, we compute a scaled vector V from D_1's center to D_2's center and translate D_2 by V and D_1 by $-V$. This would prevent overlapping among a group of disks.

- **Sink**: we translate each disk D towards the center of D's parent disk by a scaled amount. This would keep the group of disks together.

We apply the "sink" and "repel" steps at each frame update until convergence. The scaling factor in "repel" is larger than that in "sink" as overlapping is less desirable than being off-centered.

Physics-based vs. Physics-inspired
An important difference of our approach from a force-directed layout is that we directly manipulate on the positions instead of velocities. In an typical physically-based approach, each frame update requires time-integration[1]: ① computing forces based on geometric relations, ② updating velocities w.r.t. forces, ③ updating positions w.r.t. velocities. We found that in general, directly modifying the positions converges faster and hence produces more visually stable layout with less oscillations. Hence, we make a distinction from alternative physically-based approaches and refer to ours as *physics-inspired*.

3.3 Changing Radius Over Time

As the topic/keyword weight may change over time, we allow the circles representing terms to inflate or shrink in an animation mode. Specifically, the animation requires T N-tuple vectors as inputs, where T is the number of time slices. These are weight vectors for all N terms over time.

[1] E.g. explicit Euler method

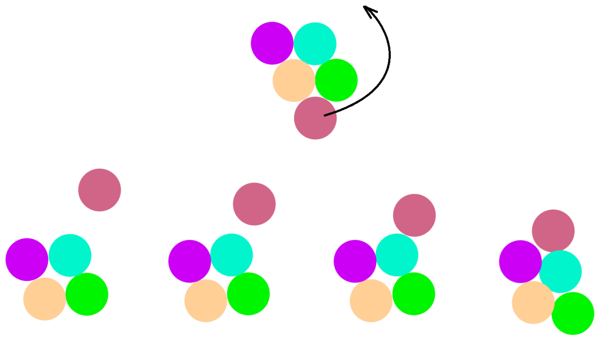


Figure 4: Drag a disk around and release it: the disk join the group in a different spot.

An issue is that the number of time slices may not match the number of animation frames. Usually, even an short animation of 5 seconds (with a framerate of 30 fps) requires far more frames than slices available. Therefore, we use closed, cubic interpolation to fill the gap and produce a smooth, periodic animation.

4 Interactions

ReportViz aims to provide engaging experiences for topic exploration. Currently, it allows a viewer to:

- **Drag and release**: The viewer can interfere with the layout by dragging a disk around and releasing it. Then the disk joins the group in a different spot driven by the physics-inspired algorithm, as shown in Fig. 4.
- **Expand and collapse**: The viewer can select a tree node to expand or collapse it. Expanding a topic node spawns a group of key term nodes, as shown in Fig. 1.

Also the user can switch to the animation mode to play a synthesized topic fluctuation where disks are changing their radius periodically while they remain repelled and gathered. (Bias can be adjusted in a default range for faster or slower convergence of the layout).

5 Feedbacks

Preliminary experiments with ReportViz suggest it as a utility providing an overview of topics and key words in public health. We summarize feedbacks from users of ReportViz as follows:

- **Topic complexity**: As shown in Fig. 5, a topic could split into subtopics, or multiple topics could be nested in a broader theme. For example, in Health 2011 [10], "Expenditure" and "Programs" are bundled as "Health Care Cost", while they remain separated in Health 2008 [7], 2009 [8] and 2010 [9]; In particular, the topic "Expenditure" is splitted into "National Expenditure" and "State Expenditure" in Health 2008 [7].
- **Word choices**: words represent the same concept are used across different topics. For example, "cancer" is a key word in the topic "Measure of Health" while "maglignant neoplasm" is used in the topic "Mortality", as shown in Fig. 6.

6 Conclusion

In this paper we present ReportViz, a utility that integrates text mining and topic visualization. We propose to use the tree data structure for topic merge and split, physics-inspired algorithms for the layout, and animation for incorperating time-varying features. In the future, we would like to build ReportViz as a full-fledged navigation tool for health documents. Specifically, we would like to encode more attributes, such as word-to-topic specificity and topic-to-topic correlation into the visualization. It would also be interesting to design use cases to evaluate the effectiveness of ReportVis for understanding topic complexity and word choices in public health.

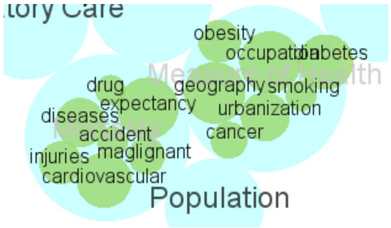


Figure 6: different word choices in the two topics "Measure of Health" and "Mortality".

Acknowledgements

I wish to thank Professor Jarek Rossignac, Jacob Esenstein, PhD students Zhuhong (Jonah) Chen, Zhen Chen, and Yangfeng Ji for their discussions.

References

[1] D. M. Blei. Variational inference. Technical report, Princeton University, 2011.

[2] D. M. Blei and J. D. Lafferty. Dynamic topic models. In *Proceedings of the 23rd international conference on Machine learning*, ICML '06, pages 113–120, New York, NY, USA, 2006. ACM.

[3] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[4] N. Cao, D. Gotz, J. Sun, and H. Qu. Dicon: Interactive visual analysis of multidimensional clusters. *IEEE Trans. Vis. Comput. Graph.*, 17(12):2581–2590, 2011.

[5] W. Cui, S. Liu, L. Tan, C. Shi, Y. Song, Z. Gao, H. Qu, and X. Tong. Textflow: Towards better understanding of evolving topics in text. *IEEE Trans. Vis. Comput. Graph.*, 17(12):2412–2421, 2011.

[6] S. Liu, M. X. Zhou, S. Pan, Y. Song, W. Qian, W. Cai, and X. Lian. Tiara: Interactive, topic-based visual text summarization and analysis. *ACM Trans. Intell. Syst. Technol.*, 3(2):25:1–25:28, Feb. 2012.

[7] National Center for Health Statistics. *Health, United States, 2008: With Special Feature on the Health of Young Adults*, 2009.

[8] National Center for Health Statistics. *Health, United States, 2009: With Special Feature on Medical Technology*, 2010.

[9] National Center for Health Statistics. *Health, United States, 2010: With Special Feature on Death and Dying*, 2011.

[10] National Center for Health Statistics. *Health, United States, 2011: With Special Feature on Socioeconomic Status and Health*, 2012.

[11] P. Resnik and E. Hardisty. Gibbs sampling for the uninitiated. Technical report, University of Maryland, 2010.

[12] F. B. Viegas, M. Wattenberg, F. van Ham, J. Kriss, and M. McKeon. Manyeyes: a site for visualization at internet scale. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1121–1128, Nov. 2007.

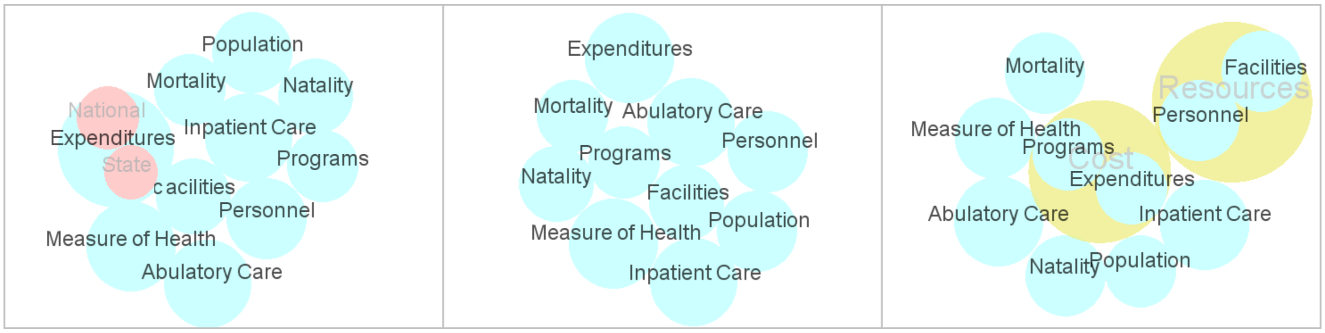


Figure 5: The topic “Expenditure” is splitted into “National” and “State” on the left. On the right, “Expenditure” and “Programs” are bundled as one theme.

Topics	Lists of Key terms
Population	level, **age**, **sex**, **race**, **resident**, poverty, Hispanic, **income**, **origin**
Fertility and Natality	**weight**, **height**, childbearing, prenatal, birth, breastfeeding, **pregnancy**, **abortion**, **contracept**, marital
Mortality	**cause**, **injuries**, homicide, suicide, **cardiovascular**, **malignant neoplasm**, trachea, bronchus, breast, HIV, drug, expectancy, fetal, **diseases**, **accidental death**
Measure of Health	**occupation**, illness, industry, condition, geography, survival, heart, stroke, diabetes, headache, pain, joint, activity, cancer, limitation, vision, hearing, assess, disability, **urbanization**, distress, **smoking**, education, **drinking**, hypertension, cholesterol, **nutrient**, leisure, obesity, dental
Ambulatory Care	**prescription**, urbanization, access, visit, clinic, influenza, **vaccination**, coverage, pneumococcal, dental, mammography, pap smears, procedures, **emergency**, **X-ray**, physician, hospital, outpatient, primary, dietary, supplement, **blood tests**, **rehabilitation**
Inpatient Care	hospital, **admissions**, discharges, nonfederal, short-stay, **diagnosis**, **length of stay**, procedure, **surgery**, **specialty**
Personnel	**physician**, **patient**, **doctors**, medicine, primary, **specialty**, dentists, **employment**, wages, enrollment, graduates, schools
Facilities	**hospitals**, **beds**, occupancy, ownership, organization, treatment, community, nursing, homes, medicare, certified, providers, suppliers, **MRI (Magnetic resonance imaging)**, **CT (computed tomography)**
Expenditures	**GDP (gross domestic product)**, national, **CPI (Consumer Price Index)**, growth, services, annual, **expenses**, payment, out-of-pocket, **insurance**
Coverage and Programs	Insurance, private, **medicaid**, **medicare**, enrollees, FFS (fee-for-service), **beneficiaries**, eligibility, veterans, state, poverty, fiscal, **reimbursement**

Table 1: We show in the left column a list of topics addressed in health reports. On the right column we show lists of key words associated with each topics. The terms shown in black are summarized from text. The words highlighted in bold are selected by medical students as terms with higher specifity. Words shown in red are considered related, but not reported from text analysis.

Challenges of Time-oriented Data in Visual Analytics for Healthcare

Wolfgang Aigner[*] Paolo Federico[†] Theresia Gschwandtner[‡] Silvia Miksch[§] Alexander Rind[¶]

Institute of Software Technology and Interactive Systems
Vienna University of Technology

Abstract

The visual exploration and analysis of time-oriented data in healthcare are important yet challenging tasks. This position paper presents six challenges for Visual Analytics in healthcare: (1) scale and complexity of time-oriented data, (2) intertwining patient condition with treatment processes, (3) scalable analysis from single patients to cohorts, (4) data quality and uncertainty, (5) interaction, user interfaces, and the role of users, and (6) evaluation. Furthermore, it portrays existing and future work by the authors tackling these challenges.

Index Terms: H.5.m [Information Systems]: Information Interfaces And Presentation (e.g., HCI)—Miscellaneous I.3.6 [Computing Methodologies]: Computer Graphics—Methodology and Techniques; J.3 [Computer Applications]: Life and Medical Sciences—Medical information systems

1 Introduction

Utilizing the huge volumes of heterogeneous data resources and collections is one of the greatest challenges of our computerized society. This holds in particular for healthcare, where different user groups are collecting, assessing, exploring, and analyzing such kinds of data and information. Visual Analytics denotes "the science of analytical reasoning facilitated by visual interactive interfaces" [22] and aims to make complex information structures more comprehensible, facilitate new insights, and enable knowledge discovery. Visual Analytics methods focus on the information discovery process exploiting both the computational power of computers and the human's visual information processing capabilities. Therefore, it aims to enable the exploration and the understanding of large and complex data sets intertwining interactive visualization, data analysis, and human-computer interaction.

2 Challenges

In the last years, different articles summarized open problems and main challenges for Visual Analytics (cp. [12, 13, 21, 22, 23]). We surveyed the state-of-the-art of information visualization approaches for exploring and querying Electronic Health Record systems (EHRs) in a recent article [20] and collected visualization methods of time-oriented data and information [2]. According to these references, we illustrate the most important open problems and challenges for Visual Analytics in Healthcare and present possible solutions to some issues.

Scale and Complexity of Time-oriented Data. Usually, the heterogeneous data resources and collections in healthcare are not only large and complex, but also time-oriented. In contrast to other quantitative data dimensions that are usually "flat", time has inherent semantic structures, contains natural cycles and re-occurrences (as for example seasons), but also social (often irregular) cycles, like holidays or school breaks. For example, the time span between check-up examinations of a chronic patient may vary between weeks and years. Therefore, time-oriented data need to be treated differently from other kinds of data and demand appropriate interaction, visual and analytical methods to analyze them.

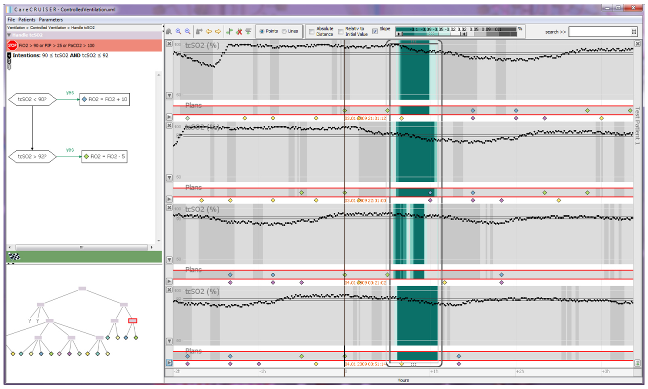


Figure 1: CareCruiser [9]: The temporal view (on the right) arranges patient parameters together with applied clinical actions along a horizontal time axis. In this screenshot the turquoise color marks the falling of the $tcSO_2$ values. Multiple instances of applying the same clinical action to one patient are aligned on a vertical axis.

Intertwining Patient Condition with Treatment Processes. Healthcare data, such as in EHRs, cover not only observations about the patients' condition, but also information about the various treatment actions over time. All these data and information need to be analyzed intertwinedly. The medical staff usually does not examine a single patient parameter, but observes the correlations of multiple parameters to assess the patient's health condition. Moreover, the parameter value at a single point in time is less meaningful than its evolution over time. In particular, the identification of changes in a patient's condition in reaction to applied treatments, demands for an intertwined view. The following tasks are high-level tasks in medical care and require a representation of the patient's parameters (i.e., the health condition of the patient) in tight combination with the applied treatment actions:

1. Monitoring the treatment progress (i.e., which treatment action is being applied at the moment, which treatments have been applied so far, and which actions may be applied in the near future),
2. Monitoring the overall success or failure of applied treatments,
3. Seeing the effects of different treatment actions on the individual patient's condition,
4. Getting a comprehensive picture about the possible reasons for changes of the patient's condition (i.e., the bettering or worsening of single patient parameters),

[*] e-mail: aigner@ifs.tuwien.ac.at
[†] e-mail: federico@ifs.tuwien.ac.at
[‡] e-mail: gschwandtner@ifs.tuwien.ac.at
[§] e-mail: miksch@ifs.tuwien.ac.at
[¶] e-mail: rind@ifs.tuwien.ac.at

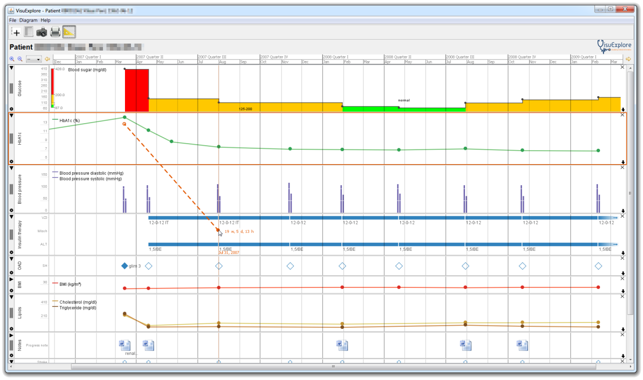


Figure 2: VisuExplore [19]: Overview visualization of a patient's medical history predominantly using well-know and easy to read visual representations (e.g., line plots).

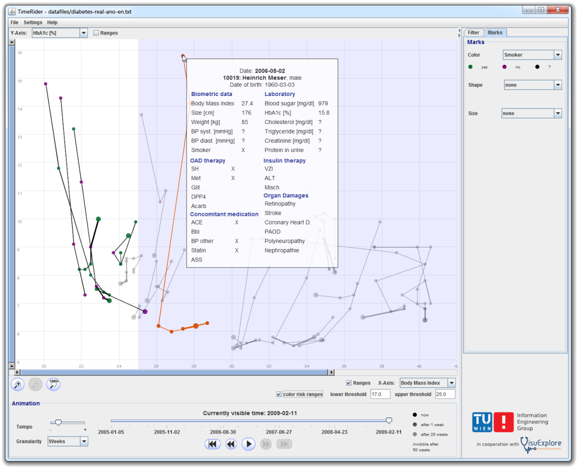
5. Identifying sub-optimal treatment choices, and thus,
6. Optimizing treatment for individual patients.

Considering the information needs of different user groups, the following visualization prototypes all tackle specific aspects of these tasks.

CareVis [1] is an interactive visualization that displays patient data in combination with computer-interpretable medical guidelines and protocols (CGP), which capture the complex structural and temporal constraints of applied and future treatment. CareCruiser [9] is a conceptual extension of CareVis with a special focus on the exploration of the effects of clinical actions on a patient's condition. It provides several features to support a step-wise interactive exploration: (1) aligning clinical actions, (2) color-coding curve events, (3) filtering color-coded information, and (4) a focus & context window for the detection of patterns of effects (Figure 1). VisuExplore [19] is more powerful regarding patient data but less regarding the structural and temporal constraints of treatment. It can display various aspects of an EHR by supporting different visualization methods in parallel panels along a common time axis. For example, in Figure 2, medical test values are represented by line plots, bar charts, and a step chart with color-coded qualitative abstractions. Treatment performed over a period of time is shown in a timeline chart through horizontal bars.

Scalable Analysis from Single Patients to Cohorts. Healthcare requires scalable visualization and analysis methods. Besides the complexity and the scale of the time dimension and the multivariate nature of healthcare data, an additional dimension to consider is the number of patients to be analyzed simultaneously. Indeed, while a system focusing on the analysis of a single patient might be sufficient to provide appropriate care tailored to the needs of that specific patient, multiple-patients systems can be useful to compare the response of diverse patients, to follow the development of an entire cohort, or to assess the effectiveness of a therapy on a larger scale.

We have proposed different solutions for the visual analysis of multiple patients. CareCruiser [9] (Figure 1) enables the exploration of two or more patients, providing collapsible facets each showing the evolution of one patient along the time axis; to support a better comparison, the data can be interactively aligned using a relative time (e.g., calendar date, time since start of therapy, or time since any other event). TimeRider [18] and Gravi++ [11] exploit animation and traces to show the evolution of multiple patients. TimeRider (Figure 3) enables bi-variate analysis of cohort trends by the means of animated scatter plots: marks representing patients are laid out according to two categorical or numerical axes and animated to show their temporal evolution; data wear is encoded to transparency, to take into account different sampling rates.


TimeRider - datafiles/diabetes-real-ano-en.txt
File Settings Help
Filter Marks
Animation
TU WIEN
Information Engineering Group
In cooperation with VisuExplore

Y-Axis: HbA1c [%]
X-Axis: Body Mass Index
☑ Ranges
☑ color risk ranges lower threshold 17.0 upper threshold 25.0
Color: Smoker
Shape: none
Size: none
Currently visible time: 2009-02-11
Tempo
Granularity: Weeks

Date: **2006-05-02**
10019: Heinrich Meser, male
Date of birth: 1960-03-03

Biometric data		Laboratory	
Body Mass Index	27.4	Blood sugar [mg/dl]	979
Size [cm]	176	HbA1c [%]	15.8
Weight [kg]	85	Cholesterol [mg/dl]	?
BP syst. [mmHg]	?	Triglyceride [mg/dl]	?
BP diast. [mmHg]	?	Creatinine [mg/dl]	?
Smoker	X	Protein in urine	?
OAD therapy		**Insulin therapy**	
SH	X	VZI	
Met	X	ALT	
Glit		Misch	
DPP4		**Organ Damages**	
Acarb		Retinopathy	
Concomitant medication		Stroke	
ACE	X	Coronary Heart D.	
Bbl		PAOD	
BP other	X	Polyneuropathy	
Statin	X	Nephropathie	
ASS			

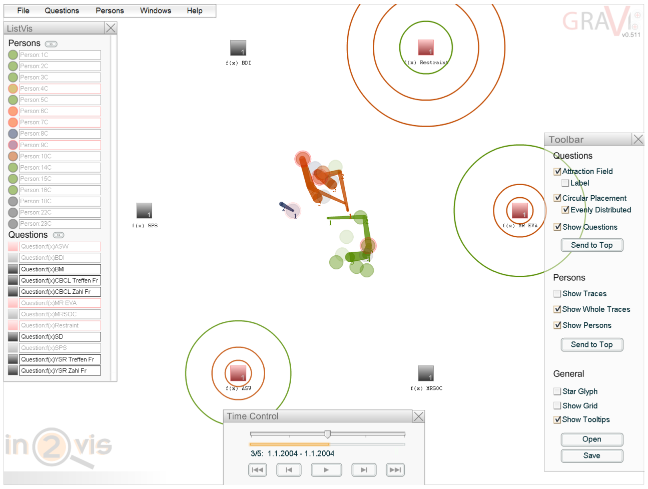
Figure 3: TimeRider [18]: Animated scatter plot for bi-variate analysis of patient cohort trends. Optional traces show the complete trajectory of the patients. The blue background denotes elevated parameter values ($BMI > 25$).

Gravi++ (Figure 4) enables multi-variate analysis: different patients are spatially clustered by a dynamic spring-based layout taking into account several variables. An analogous incidence-model could be used to visualize patients' cohorts as dynamic networks. In ViENA [7] we have integrated different static visualizations for dynamic networks, namely juxtaposition (small multiples), superimposition, and 2.5D views (Figure 5). The benefits and limitations of animation and static visualizations of patients' parameters need further investigations also concerning the different medical users.

Knowledge-based temporal abstractions, besides supporting specific user needs in the medical domain by combining quantitative and qualitative aspects, also enable more compact visualizations. With Midgaard [6], a visualization that combines raw data and abstractions as well as a semantic zoom changing the level of abstraction has been introduced (Figure 6). Moreover, we have evaluated its effectiveness in supporting tasks involving parameters of single patients [4]. Such compact visualizations based on temporal abstractions can be useful when dealing with multiple patients as a mean to optimize the display space occupancy, but their application in the case of multiple variables and multiple patients should be researched further. Furthermore, a promising research topic is the development of context-based temporal abstractions that pursue a closer interaction with knowledge, adapting to the patient's context and reacting dynamically to its modifications.

Data Quality and Uncertainty. A central issue in Visual Analytics is to avoid misinterpretation by the analysts. However, in real-life data there are several issues that may lead to misinterpretation or wrong results, such as missing data, uncertain data, ambiguous data, or simply wrong data. Especially in the discipline of healthcare, data sets may contain an unavoidable amount of uncertainty, errors, and ambiguity. To assure the reliability of any data analysis step, quality problems within the data set have to be detected and – if possible – resolved first. Several taxonomies of general data quality problems exist, but they do not consider the very special characteristics of time (in healthcare, data sets are highly time-oriented). To this end, we have provided a taxonomy of time-oriented data quality problems [10]. On the one hand, it gives a unified view on the various existing taxonomies of general data quality problems. On the other hand, it provides an important reference when formulating quality checks of time-oriented data.

However, there may be data issues that cannot be corrected, such



Questions

☑ Attraction Field

□ Label

☑ Circular Placement

☑ Evenly Distributed

☑ Show Questions

Persons

□ Show Traces

☑ Show Whole Traces

☑ Show Persons

General

□ Star Glyph

□ Show Grid

☑ Show Tooltips

3/5: 1.1.2004 - 1.1.2004

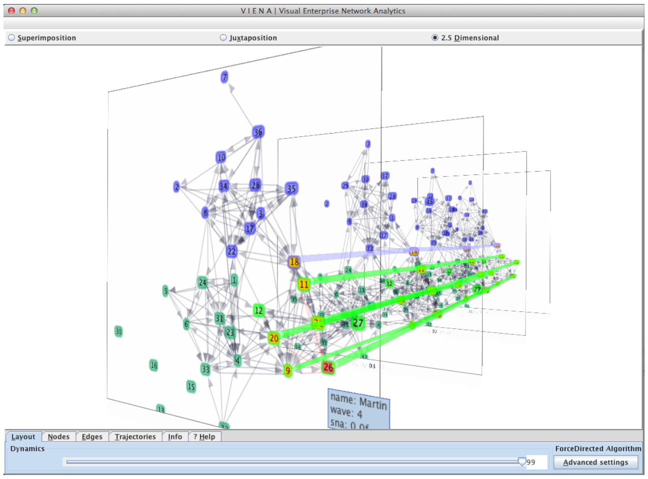
Figure 4: Gravi++ [11]: A spring-based layout is used for spatial clustering by multiple parameters, which are represented by six squares in the screenshot. Animation and traces show evolution over time.

as an uncertain starting time of a future event. These issues have to be communicated appropriately to the user in order to ensure an informed interpretation of the data at hand. PlanningLines [3], for instance, use novel glyphs to visualize temporal uncertainties. The glyph visually communicates the earliest and latest possible starting time of a task, the earliest and latest possible ending time, as well as the minimum and the maximum duration of the task. It was designed to represent complex time annotations for CGPs.

Interaction, User Interfaces, and the Role of Users. Large and complex data sets cannot be visualized and analyzed as a whole at once. Exploration of these data sets is an interactive, multistep process that involves trial-and-error, human judgment, and exchange with colleagues. Therefore, task-specific interaction methods and user interfaces are required. Furthermore, Visual Analytics methods need to account for the different backgrounds and usage contexts of the user groups involved in healthcare. While physicians or nurses are driven by tight schedules and frequent interruptions towards simple interfaces that deliver overview at a glance, clinical researchers and quality analysts need flexibility and support for their reasoning process. Non-professionals such as patients, family member carers, or other intermittent users play an important role in healthcare, but need to be addressed more specifically by Visual Analytics methods.

We follow a user/data/task-centered design approach in all our application-oriented projects and we develop task-specific and user-specific interaction methods and user interfaces. For example, CareCruiser [9] is tailored for CGP-based care as it allows for an active investigation of the development of the patient's condition and the detection of effects of applied CGPs. VisuExplore [19] relies on well-know and easy to read visual representation techniques such as line plots and timeline charts to provide a clear and unambiguous overview of a patient's medical history. Furthermore, it allows personalization of the user interface either by interaction or through a configuration file.

There is a need for well-defined process models in Visual Analytics, in order to better understand the analytical reasoning process and develop more suitable Visual Analytics methods. We empirically analyzed interaction logs collected from user studies of Gravi++ and VisuExplore and identified common interaction patterns and transition probabilities [17]. In related work, we describe a Visual Analytics process that uses the structure of time to build hypotheses and statistical models on time-series data [14].



○ Superimposition ○ Juxtaposition ● 2.5 Dimensional

name: Martin
wave: 4
sna: 0.0

Dynamics 99

ForceDirected Algorithm

Figure 5: ViENA [7]: 2.5D visualization of dynamic networks. Traces represent change of network metric for person nodes over time.

Evaluation. The Visual Analytics process is complex, comprised of multiple steps of computation and human reasoning, and produces outcomes that are hard to measure. Thus, it is very difficult to assess the quality and effectiveness of Visual Analytics methods, in particular in an interdisciplinary domain like healthcare. Nevertheless, evaluation is essential both for adoption in clinical practice and advancing Visual Analytics for Healthcare as a scientific community.

Evaluation methods can be categorized by the threats they address in the design process [16]. Evaluation against certain threats requires the involvement of domain experts (e.g., physicians), but their tight schedules make it hard to recruit more than a few subjects. The combination of different methods can alleviate these problems and strengthen the evidence on Visual Analytics methods (e.g., Gravi++ [11]).

To compare and assess various Visual Analytics solutions, large benchmark data sets of de-identified patient records, relevant tasks, and gold standard solutions would be necessary. On the other hand, if a user-centered design process is followed and concrete tasks and data of the involved users are addressed, established categorizations can be used to make the results better comparable. For that purpose, we regularly apply the task framework by Andrienko and Andrienko [5], the user intents by Yi et al. [24], and the heuristics by Forsell and Johansson [8]. Furthermore, we have proposed a categorization for time-oriented data [2] and a task framework that is extended along the structure of time [15].

Finally, many steps of an evaluation study such as task display, time keeping, and data collection can be automated. We are working on a general evaluation framework that can be plugged into Visual Analytics prototypes. It has been tested successfully in several user studies.

Other Open Problems and Challenges. We are aware that the above list does not cover all issues, for example, we did not elaborate about infrastructures, hardware, display and interaction devices, data streams, patient safety, data security, personalization, or privacy. However, we aim to address the most important issues specific to healthcare, first.

3 Outlook

Our research group will tackle these challenges in two current research projects and future work. In the course of the Laura Bassi Centre of Expertise *CVAST*,[1] we aim to develop novel, user-

[1] http://www.cvast.tuwien.ac.at/, cited Aug 31, 2012.

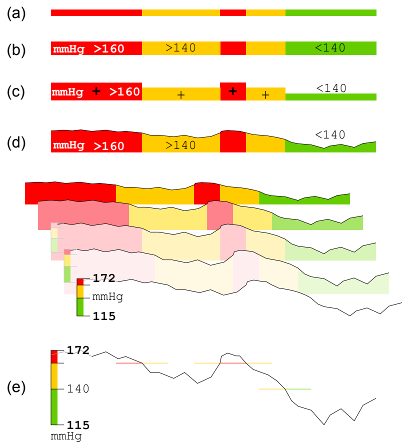


Figure 6: Midgaard [6]: Visualization technique for numerical variables with semantic zoom to one of five levels of detail: (a) colored background, (b) colored background with labels, (c) colored bars, (d) colored area charts, and (e) augmented line charts.

oriented, and task-specific Visual Analytics methods that foster new insights and enable knowledge discovery. Through our participation in the *MobiGuide* project,[2] we aim to design and develop Visual Analytics methods for the patients' data and the guideline processes, focusing on and their compliance and modifications over time, also addressing uncertainty and incompleteness.

Acknowledgements

The Laura Bassi Centre of Expertise CVAST is funded by the Austrian Federal Ministry of Economy, Family and Youth (project number: 822746). MobiGuide is funded by the European Commission under the 7th Framework Program (grant number: 287811).

References

[1] W. Aigner and S. Miksch. CareVis: Integrated visualization of computerized protocols and temporal patient data. *Artificial Intelligence in Medicine*, 37(3):203–218, 2006.

[2] W. Aigner, S. Miksch, H. Schumann, and C. Tominski. *Visualization of Time-Oriented Data*. Springer, London, 2011.

[3] W. Aigner, S. Miksch, B. Thurnher, and S. Biffl. PlanningLines: Novel glyphs for representing temporal uncertainties and their evaluation. In *Proc. 9th Int. Conf. Information Visualisation (IV 2005)*, pages 457–463. IEEE, 2005.

[4] W. Aigner, A. Rind, and S. Hoffmann. Comparative evaluation of an interactive time-series visualization that combines quantitative data with qualitative abstractions. *Computer Graphics Forum*, 31(3):995–1004, 2012.

[5] N. Andrienko and G. Andrienko. *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach*. Springer, Berlin, 2006.

[6] R. Bade, S. Schlechtweg, and S. Miksch. Connecting time-oriented data and information to a coherent interactive visualization. In *Proc. ACM SIGCHI Conf. Human Factors in Computing Systems (CHI)*, pages 105–112, 2004.

[7] P. Federico, W. Aigner, S. Miksch, F. Windhager, and L. Zenk. A visual analytics approach to dynamic social networks. In *Proc. 11th Int. Conf. Knowledge Management and Knowledge Technologies (i-KNOW '11)*, pages 47:1–47:8. ACM, 2011.

[8] C. Forsell and J. Johansson. An heuristic set for evaluation in information visualization. In G. Santucci, editor, *Proc. Int. Conf. Advanced Visual Interfaces (AVI 2010)*, pages 199–206. ACM, 2010.

[9] T. Gschwandtner, W. Aigner, K. Kaiser, S. Miksch, and A. Seyfang. CareCruiser: exploring and visualizing plans, events, and effects interactively. In *Proc. IEEE Pacific Visualization Symp. (PacificVis 2011)*, pages 43–50, 2011.

[10] T. Gschwandtner, J. Gärtner, W. Aigner, and S. Miksch. A taxonomy of dirty time-oriented data. In G. Quirchmayr, J. Basl, I. You, L. Xu, and E. Weippl, editors, *Multidisciplinary Research and Practice for Information Systems, Proc. CD-ARES 2012*, LNCS 7465, pages 58–72. Springer, 2012.

[11] K. Hinum, S. Miksch, W. Aigner, S. Ohmann, C. Popow, M. Pohl, and M. Rester. Gravi++: Interactive information visualization to explore highly structured temporal data. *Journal of Universal Computer Science*, 11(11):1792–1805, 2005.

[12] D. Keim, G. Andrienko, J. Fekete, C. Görg, J. Kohlhammer, and G. Melançon. Visual analytics: Definition, process, and challenges. In A. Kerren, J. T. Stasko, J. Fekete, and C. North, editors, *Information Visualization: Human-Centered Issues and Perspectives*, LNCS 4950, pages 154–175. Springer, Berlin, 2008.

[13] D. Keim, J. Kohlhammer, G. Ellis, and F. Mansmann, editors. *Mastering The Information Age – Solving Problems with Visual Analytics*. Eurographics, Goslar, Germany, 2010.

[14] T. Lammarsch, W. Aigner, A. Bertone, S. Miksch, and A. Rind. Towards a concept how the structure of time can support the visual analytics process. In S. Miksch and G. Santucci, editors, *Proc. Int. Workshop on Visual Analytics (EuroVA 2011) in conjunction with EuroVis 2011*, pages 9–12. Eurographics, 2011.

[15] T. Lammarsch, A. Rind, W. Aigner, and S. Miksch. Developing an extended task framework for exploratory data analysis along the structure of time. In K. Matkovic and G. Santucci, editors, *Proc. Int. EuroVis Workshop on Visual Analytics (EuroVA 2012)*, pages 31–35. Eurographics, 2012.

[16] T. Munzner. A nested process model for visualization design and validation. *IEEE Trans. Visualization and Computer Graphics*, 15(6):921–928, 2009.

[17] M. Pohl, S. Wiltner, S. Miksch, W. Aigner, and A. Rind. Analysing interactivity in information visualisation. *KI – Künstliche Intelligenz*, 26:151–159, May 2012.

[18] A. Rind, W. Aigner, S. Miksch, S. Wiltner, M. Pohl, F. Drexler, B. Neubauer, and N. Suchy. Visually exploring multivariate trends in patient cohorts using animated scatter plots. In M. M. Robertson, editor, *Ergonomics and Health Aspects of Work with Computers, Proc. Int. Conf. held as part of HCI International 2011*, LNCS 6779, pages 139–148. Springer, 2011.

[19] A. Rind, W. Aigner, S. Miksch, S. Wiltner, M. Pohl, T. Turic, and F. Drexler. Visual exploration of time-oriented patient data for chronic diseases: Design study and evaluation. In A. Holzinger and K. Simonic, editors, *Information Quality in e-Health, Proc. USAB 2011*, LNCS 7058, pages 301–320. Springer, 2011.

[20] A. Rind, T. D. Wang, W. Aigner, S. Miksch, K. Wongsuphasawat, C. Plaisant, and B. Shneiderman. Interactive information visualization to explore and query electronic health records. *Foundations and Trends in Human-Computer Interaction*, 2012. In review.

[21] J. Thomas and J. Kielman. Challenges for visual analytics. *Information Visualization*, 8(4):309–314, 2009.

[22] J. J. Thomas and K. A. Cook, editors. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE, 2005.

[23] P. C. Wong, H.-W. Shen, C. R. Johnson, C. Chen, and R. B. Ross. The top 10 challenges in extreme-scale visual analytics. *IEEE Computer Graphics and Applications*, 32(4):63–67, Aug. 2012.

[24] J. S. Yi, Y. A. Kang, J. T. Stasko, and J. A. Jacko. Toward a deeper understanding of the role of interaction in information visualization. *IEEE Trans. Visualization and Computer Graphics*, 13(6):1224–1231, 2007.

[2] `http://www.mobiguide-project.eu/`, cited Aug 31, 2012.

Interactive Visualization of Prescriptions of Drugs to Individuals within Large Populations—Analyses of Temporal Relationships of Events

Jimmy Johansson, Morten Andersen, Alexander Fridlund and Mikael Hoffmann

Abstract—This paper reports on work in progress on interactive visualization of prescriptions of drugs within large populations. A visualization prototype for interactive analysis is presented. The prototype has been developed in close collaboration between visualization researchers and domain experts within the area of pharmacoepidemiology. To illustrate the functionality of the prototype, data on treatment of hyperlipidemia with statins is used. The data set includes inhabitants from 40 to 85 years of age in Östergötland county, Sweden during the period of 2007 through 2010. The goal of the treatment with statins is to decrease the risk of future cardiovascular events among patients at an increased risk. Although no formal evaluation has been performed at this early stage of the development phase, feedback from end users has been positive. Many additional features have been proposed so there are several interesting and challenging directions for future work.

Index Terms—Interactive Visualization, Pharmacoepidemiology, Multivariate and Temporal Data

1 Introduction

During the last decade the volume of information gathered in large databases about health care has increased dramatically. The data can be generated from administrative health care records, electronic health records and quality-assurance tools, and even from claims databases. The accumulation of such large and complex data sets, combined with the possibility of record-linkage studies through the national personal identity number, creates many new opportunities for the growing field of pharmacoepidemiology.

The goal of this paper is to introduce and discuss a novel way to interactively visualize time-dependent multivariate data for large populations such as the dispensing of prescribed drugs. More specifically, the paper presents work in progress and discusses the development of an initial prototype for interactive visualization of prescriptions of drugs to individuals within large populations. Today descriptive, and thus also analytical, pharmacoepidemiology of large data sets is hampered by non-intuitive interfaces for basic variables. The prototype presented here is the result of collaboration between visualization researchers and domain experts in the field of pharmacoepidemiology.

2 Background

Pharmacoepidemiology is the discipline of the research on the use of drugs and the effects of drugs in large numbers of people [10]. Both diseases caused by drugs (adverse drug reactions) and patterns of drug use (drug utilization research, DUR) are of interest. Relationships between treatment, predictors of treatment and outcome are analysed. Even the basic variables of DUR such as period- and point prevalence, incidence and simultaneous treatment are however sensitive to assumptions about the behaviour of patients, prescribers and pharmacists. These behaviours can be influenced by external factors such as for instance changes in the pharmaceutical benefit scheme (co-payment, inclusion/exclusion of drugs etc.) and/or organization of the health care. Specific challenges are the multiple categories of drugs and diagnoses, recurrent events, derivation of treatment duration from information on dosing, complex treatment sequences with intermittent drug use and switches, and the use of multiple simultaneous drugs. Established classifications of drugs (such as Anatomical Therapeutical Chemical Classification, ATC, established by WHO) and diagnoses (ICD) [3, 10] are important for analysing and disseminating the results internationally.

Since July 2005 all dispensings of prescribed drugs in Sweden (9.5 million inhabitants) have been gathered for each individual, and have been made available for research and for generating statistics by the Medical Board of Health and Welfare through the Swedish Prescribed Drug Registry. Similar databases exist in the other Nordic countries, as well as some other countries or regions, in order to facilitate research. Insurers or providers of health care also collect data sets with a similar structure.

Currently the access to data and the resources and time necessary for data management and analysis using database software, statistics applications and spreadsheets limit the use of these health data for exploratory analyses or rapid evaluation. Traditionally, data from different registers are retrieved and linked separately for each project, and conventional bar charts, histograms etc. are produced as part of the usual statistical analysis. Sensitivity analysis of the assumptions made during the analysis of the data are cumbersome to perform. Thus, the possibilities to discover possible confounders caused by external factors such as changes in the organization and provision of health care, including pharmaceutical benefits, are limited. Exploratory analyses follow the same path, and are therefore rarely done. Similarly, the use of interactive analyses that could add to the interpretation of patterns discovered in these data remains a huge challenge in these large multidimensional data sets.

Time-dependent, highly multivariate data, such as that exemplified by medical data of this nature, is one of the major challenges in data analysis [2]. Already involving the records of, potentially, millions of individuals, the data continue to expand with time, becoming ever larger and more difficult to manage but also, and more importantly, more difficult to represent in an understandable way. This creates new demands on tools and techniques to analyse and visualize the data sets. Traditional analysis methods suitable for hypothesis testing fall short as such tools, when applied to complex data, generates too many hypotheses that must be examined in the exploration process. Many excellent visual representations have been developed for high dimensional data [11] but the addition of the time factor renders most of them unhelpful since they are unable to show many time steps in a coher-

- *Jimmy Johansson is with C-Research, Linköping University, Sweden, e-mail: jimmy.johansson@liu.se.*
- *Morten Andersen is with the Centre for Pharmacoepidemiology, Karolinska Institutet, Sweden, e-mail: morten.andersen@ki.se*
- *Alexander Fridlund is with C-Research, Linköping University, Sweden, e-mail: alexander.fridlund@liu.se.*
- *Mikael Hoffmann is with the Department of Medical and Health Services, Linköping University, Sweden, e-mail: mikael.hoffmann@nepi.net.*

Manuscript received 31 March 2012; accepted 1 August 2012; posted online 14 October 2012; mailed on 5 October 2012.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

ent way. Specific tools for visualization of temporal data exist [1] but need to be extended and customized to explore correlations and time-trends in order to support these new and important medical research and monitoring processes.

3 Interactive Visualization Prototype

This section describes the developed prototype and the case used in the examples.

The case considered is the use of statins for the treatment of hyperlipidemia. The goal of the treatment with statins is to decrease the risk of future cardiovascular events among patients at an increased risk. The use of statins has undergone major changes in the last years. The number of patients treated has increased, major patents have expired, and the reimbursement of statins has been revised in Sweden. The data set used for the pilot consists of all pharmacy-dispensed statins to inhabitants from 40 to 85 years of age in Östergötland county during the period of 2007 through 2010. In addition, the dispensation of any anti-diabetic agent to an individual during the period was also included as a proxy for diabetes mellitus treated with drugs. The data set was extracted as a part of a larger research project describing the impact of reimbursement changes during that period in Sweden. Data has been anonymized by replacing the unique patient identifier with a case number, as well as moving all birth-dates to the beginning of the corresponding year. Two different subsets of the database with 10,000 and 50,000 events respectively were used for the visualization.

The developed prototype consists of an interface based on coordinated and multiple linked views [9]. The primary visualization technique used is the Lexis diagram [7] which is a tool used in epidemiology to visualize relationships between events in time and a person's age. Figure 1 shows an illustration of the Lexis diagram where each point represents an event in time, in this case each event describes a drug prescription and each line connects all drug prescriptions made by a single individual over time.

After the data is loaded, the user can choose an arbitrary number of graphical views. Each view can be separately customized via a number of filters and visualization options. Standard interaction techniques such as zooming and panning are performed separately on each view. Figure 2 shows screen shots of possible configurations. Examples of interaction and visualization options are:

- down-scaling the data by sampling in order to focus the analysis on a specific subset.
- selecting data based on the Anatomical Therapeutic Chemical (ATC) Classification System for in-depth analysis of one or several substances, isolated or in relationships to each other.
- splitting the population based on sex in order to study differences between males and females.
- specifying a range in age to analyse differences or similarities between various age groups.
- selecting starting and end dates for analysis of a specific period in time.
- identifying patients that at some point in time have been dispensed an antidiabetic drug as a risk factor for cardiovascular disease.
- colour coding the events and individuals based on ATC codes to analyse patterns in drug prescriptions over time.

The prototype has been developed in C++ and OpenGL to ensure efficient rendering. The interface is implemented using wxWidgets. The user interface as well as the specification of visualization and interaction techniques have been made in close collaborations with domain experts to ensure that the prototype supports as many tasks as possible and that the developed functionality is actually useful. Although no formal evaluation has been performed at this early stage of the development phase, feedback from end users has been positive. Many additional features have been asked for and there are many possible directions for future work.

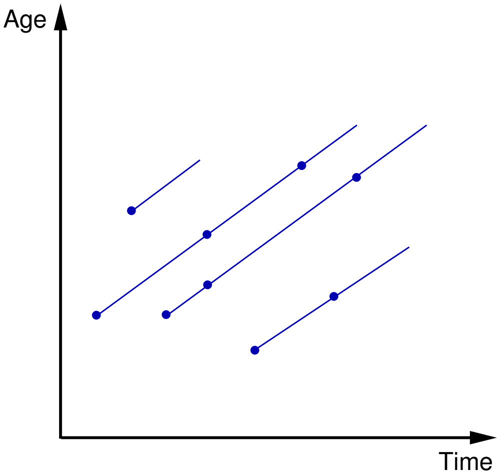


Fig. 1. A Lexis diagram is a representation of the relationship between events in time and a person's age. The x-axis shows calendar time and the y-axis shows the age of the patients. Each point represents an event in time, in this case each event describes a drug prescription. Each line connects all drug prescriptions made by a single individual over time.

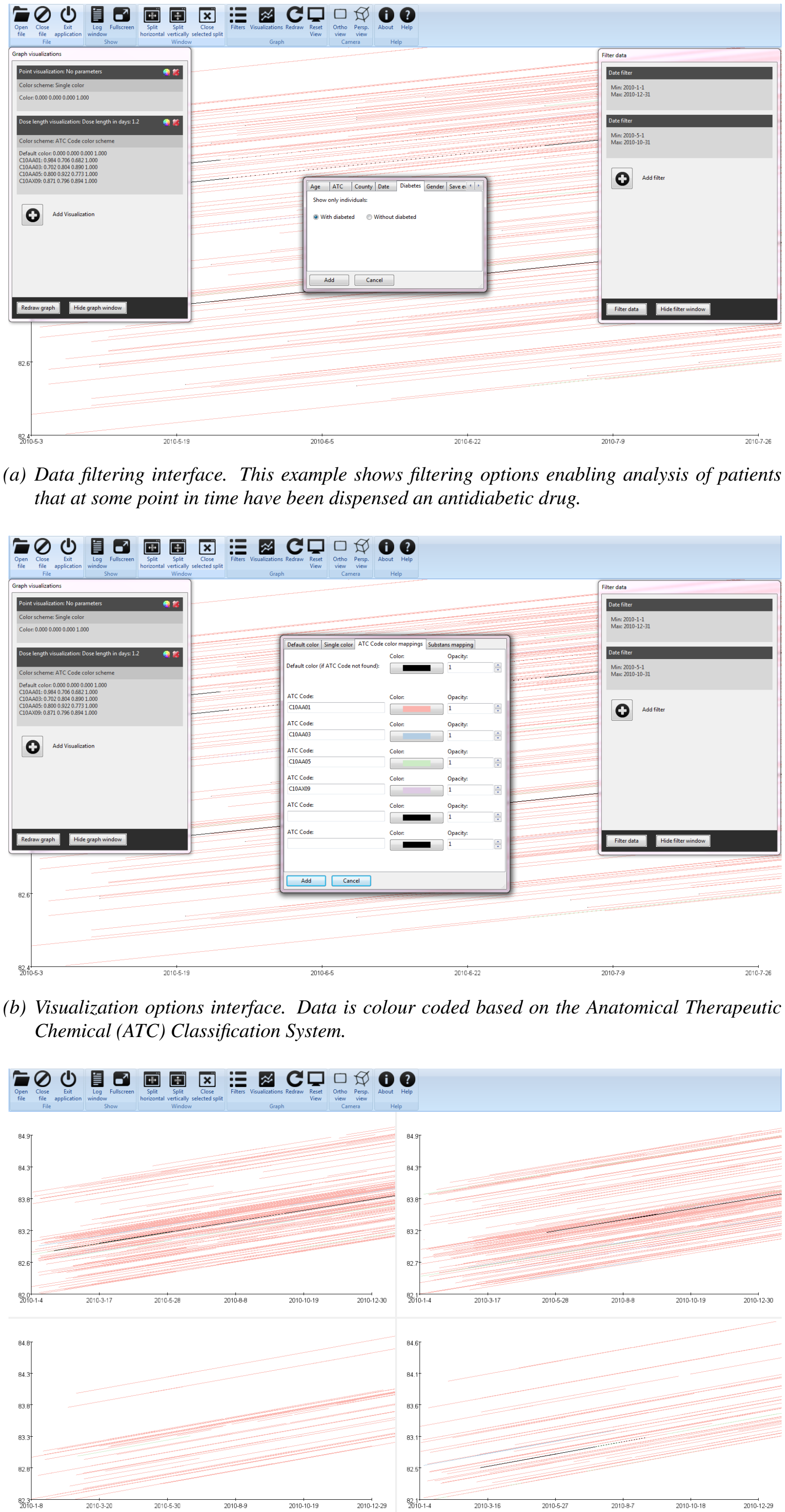
4 Conclusions and Future Work

This paper reports on work in progress on interactive visualization in the area of pharmacoepidemiology. A visualization prototype has been developed in close collaboration with domain experts within the area of pharmacoepidemiology. Preliminary results (supported by domain experts) suggest that the tool facilitates interactive analysis of this type of data and can be used for a number of different tasks that today are performed using traditional statistics software packages.

One area in which the developed prototype might be of particular value is in the identification of sequences and temporal relationships between dispensings of one or several different pharmaceuticals, which today is a time-consuming and cumbersome task. Possible confounders in the data set such as changes in reimbursement systems, influence of media 'scares' and/or introduction in the market-place of substitute treatments within another drug class have to be recognized in order to be included in the analyse. The developed prototype might be used as a tool for pattern recognition in large temporal data sets in order to identify and handle such confounders.

Further development aims to visualize different measures important in pharmacoepidemiology. For a single drug such measures include point- and period prevalence, incidence, adherence over time including overlap in dispensed amounts, persistence of treatment including duration of treatment and interruptions. For multiple drugs other measures such as combination of treatment, treatment sequence and interactions are important measures. In both cases temporal relationships to other important events such as diagnosis or in-patient treatment episodes are possible to analyse.

The prototype has so far been used to visualize 50,000 events. For larger data sets the issue of visual clutter needs to be considered. The existing Lexis diagram could be enhanced by using advanced blending schemes [6] in order to support efficient rendering of much larger data sets. For exploratory analyses of data, it is often valuable to use a multitude of visual representations combined with data mining approaches. A possible approach is to use clustering algorithms to group individuals based on different parameters. In addition, other visualization techniques, such as parallel coordinates [4], multi-relational parallel coordinates [5] and pixel-oriented visualization techniques [8], will be evaluated for use within this area.



Graph visualizations

Point visualization: No parameters

Color scheme: Single color

Color: 0.000 0.000 0.000 1.000

Dose length visualization: Dose length in days: 1.2

Color scheme: ATC Code color scheme

Default color: 0.000 0.000 0.000 1.000
C10AA01: 0.984 0.706 0.682 1.000
C10AA03: 0.702 0.804 0.890 1.000
C10AA05: 0.800 0.922 0.773 1.000
C10AX09: 0.871 0.796 0.894 1.000

Show only individuals:

● With diabeted ○ Without diabeted

Filter data

Date filter

Min: 2010-1-1
Max: 2010-12-31

Date filter

Min: 2010-5-1
Max: 2010-10-31

(a) Data filtering interface. This example shows filtering options enabling analysis of patients that at some point in time have been dispensed an antidiabetic drug.

Graph visualizations

Point visualization: No parameters

Color scheme: Single color

Color: 0.000 0.000 0.000 1.000

Dose length visualization: Dose length in days: 1.2

Color scheme: ATC Code color scheme

Default color: 0.000 0.000 0.000 1.000
C10AA01: 0.984 0.706 0.682 1.000
C10AA03: 0.702 0.804 0.890 1.000
C10AA05: 0.800 0.922 0.773 1.000
C10AX09: 0.871 0.796 0.894 1.000

Default color (if ATC Code not found): Color: ■ Opacity: 1

ATC Code: C10AA01 Color: 🟥 Opacity: 1

ATC Code: C10AA03 Color: 🟦 Opacity: 1

ATC Code: C10AA05 Color: 🟩 Opacity: 1

ATC Code: C10AX09 Color: 🟪 Opacity: 1

ATC Code: Color: ■ Opacity: 1

ATC Code: Color: ■ Opacity: 1

Filter data

Date filter

Min: 2010-1-1
Max: 2010-12-31

Date filter

Min: 2010-5-1
Max: 2010-10-31

(b) Visualization options interface. Data is colour coded based on the Anatomical Therapeutic Chemical (ATC) Classification System.

(c) An example of multiple views. Right to left: males and females. Top row: patients not treated with anti-diabetic drugs. Bottom row: patients treated with anti-diabetic drugs.

Fig. 2. Screen shots from the visualization prototype illustrating possible analyses of patterns in drug prescriptions over time using the Lexis diagram [7]. The data consists of patients with treatment of hyperlipidimia with statins.

Acknowledgments

This work was supported by the NEPI Foundation—Swedish Network for Pharmacoepidemiology, Norrköping Science Park, Sweden and the Swedish Research Council in the Linnaeus Centre CADICS.

References

[1] W. Aigner, S. Miksch, W. Müller, H. Schumann, and C. Tominski. *Visualization of Time-Oriented Data*. Springer, 2011.

[2] C. Chen. Top 10 unsolved information visualization problems. *Computer Graphics and Applications*, 25(4):12–16, 2005.

[3] Guidelines for ATC classification and DDD assignment. WHO Collaborating Centre for Drug Statistics Methodology, Oslo, 2011.

[4] A. Inselberg. The plane with parallel coordinates. *The Visual Computer*, 1(4):69–91, 1985.

[5] J. Johansson, M. Cooper, and M. Jern. 3-dimensional display for clustered multi-relational parallel coordinates. In *Proceedings IEEE International Conference on Information Visualization, IV05*, pages 188–193, 2005.

[6] J. Johansson, P. Ljung, M. Jern, and M. Cooper. Revealing structure within clustered parallel coordinates displays. In *Proceedings IEEE Symposium on Information Visualization 2005*, pages 125–132, 2005.

[7] N. Keiding. *Encyclopedia of Biostatistics*, volume 4, chapter Lexis Diagram, pages 2767–2769. Wiley, 2nd edition, 2001.

[8] D. A. Keim. Pixel-oriented visualization techniques for exploring very large data bases. *Journal of Computational and Graphical Statistics*, 5(1):58–77, 1996.

[9] C. North and B. Shneiderman. Snap-together visualization: a user interface for coordinating visualizations via relational schemata. In *Proceedings of the working conference on Advanced visual interfaces*, pages 128–135, 2000.

[10] B. L. Strom, S. E. Kimmel, and S. Hennesy, editors. *Pharmacoepidemiology*. John Wiley and Sons Ltd, 5th edition, 2012.

[11] P. C. Wong and R. D. Bergeron. 30 years of multidimensional multivariate visualization. In *Scientific Visualization, Overviews, Methodologies, and Techniques*, pages 3–33, 1997.

No Country for Fat Men — Investigating Obesity with Visual Analytics

Aaron Lai[1,*], Thomas Ho[2], and Ryan Walker[3]

Abstract

Visual analytics is gaining importance due to the explosion of data availability and processing capabilities. In this example, we demonstrated the power of visual analytics to investigate various aspect of obesity using a readily available commercial product called Tableau on the CHIS (California Health Interview Survey). A recent JAMA article claimed that there was no time to waste in doing obesity research and a broad-based effort was needed [1]. Since CHIS tracked responses to hundreds of questions, our demonstration provided an excellent example of how visual analytic tools could empower end-users to find interesting relationships within a morass of data.

Index Terms— obesity, interaction visualization, Tableau Software, race, lifestyle, sugar, CHIS

1 Introduction

According to CDC, more than one-third of U.S. adults were obese and it costed us $147 billion in 2008 [2]. Diabetes, which was highly comorbid with obesity, had also reached an epidemic level [3]. Because these were enormously complicated problems and their prevalence was related to so many different factors, it was often difficult to uncover the relationships these diseases share with environmental and lifestyle factors. However, using appropriate data visualization techniques to analyze multiple dimensions simultaneously, we could efficiently sift through many factors to identify potentially important relationships. In this article, we used a readily available commercial product called Tableau to perform visual analysis of factors that were potentially related to obesity and diabetes prevalence. This powerful and easy to use software allowed us to focus on the visual analytics rather than programming as was done in [4].

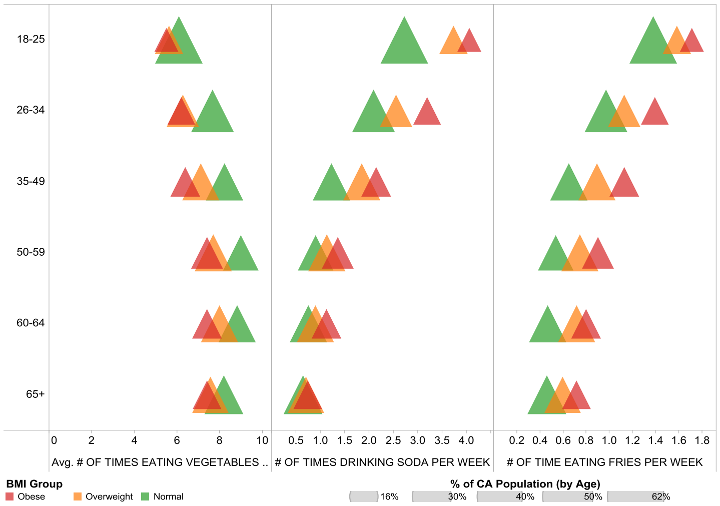


Figure 1. Food Preference by Age.

For example, in Figure 1, we used triangles to represent the “Normal”, “Overweight”, and “Obese” populations with different demographic segmentations. We put a age group on the Y-axis and lifestyle variables (number of vegetable servings per week, number of sodas per week, number of times eating fries) on the X-axes. The size of each triangle corresponded to the proportion of Californians in that weight group. The horizontal position of the triangle’s center gave the average value of the corresponding lifestyle variable.

We could see from that older people ate more vegetables, drunk fewer sodas, and ate fewer fries than younger people. Obese and overweight percentages tended to converge with age. This chart clearly illustrated those relationship in a straightforward and intuitive way. This also showed that the soda drinking gap between normal and obese was most prominent in young adult.

[1,2,3] *are with Market Insights and Intelligence, Blue Shield of California, San Francisco.*

[1] *is also a current MSc student in Evidence-based Healthcare, University of Oxford.*

[3] *is also a current PhD student in Applied Mathematics, University of Kentucky.*

[*] *the corresponding author (aaron.lai@blueshieldca.com).*

All discussions are their personal opinions and they do not represent those of their employers or affiliations.

2 Background

2.1 Public Health Issues and CHIS Data

California Health Interview Survey (CHIS) is a random-dial telephone survey [5] conducted by UCLA Center for Health Policy Research in collaboration with the California Department of Public Health and the Department of Health Care Services. The study covered a wide range of topics including demographics, lifestyle factors, health status, access to healthcare, and health insurance status … etc. In this article, we focused on the problem of obesity from the angle of general health status, dietary intake, physical activity, insurance status, and demographics. Given the high dimensionality of the CHIS data, this survey is an ideal candidate for visual analytic analysis. Using Tableau to generate quick visualizations of key variables, anyone could replicate our results with minimal software skills.

2.2 Related Work on Obesity

The authors in [6] found a week link between the physical food environment on BMI and obesity in California. It was known that different ethnic groups had different dietary behaviors [7]. These results affected the prevalence of chronic diseases such as diabetes and cardiovascular disease [8]. Sugar intake and soda consumption played a main role in obesity [9]. Since obesity was contagious in a social network [10][11][12][13], the fluency of English might also be related to lifestyle.

3 Visualization Solutions

We downloaded the 2009 adult CHIS dataset from the askCHIS site and converted the SAS dataset into a Tableau data extract. Note that the CHIS data set included a series of weights. The first of these RAKEDW0 was the final survey weight, and must be used to obtain frequencies of the California population.

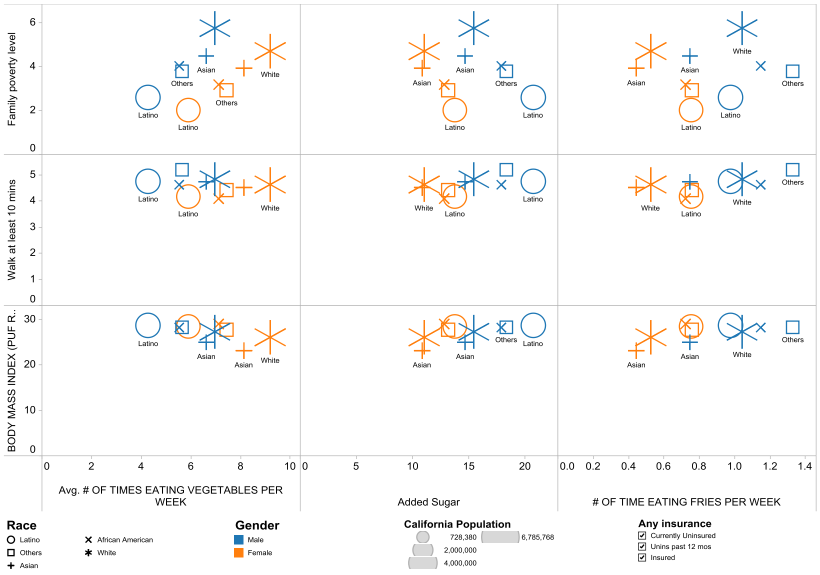


Figure 2. Food Preference, BMI, Physical Activity, and Poverty by Race

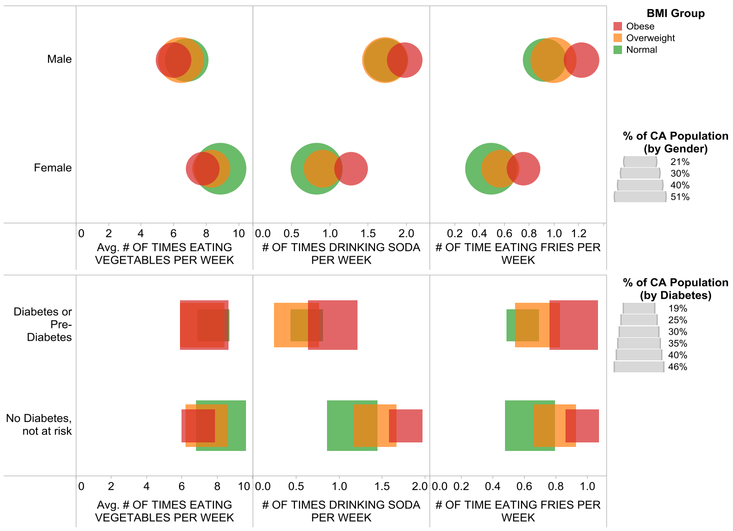


Figure 3. Food Preference, Gender and Diabetes by Weight Group

3.1 Approach

Since our purpose was to perform quick visual analytics and not detailed statistical analysis, we approached the problem in a simple way. We grouped some variables (e.g. race) for simplicity. We also introduced some combined variables (e.g. combined leisure and work walk).

Figure 2 is an example of displaying multi-dimensional data. In this graph, we focused on how people in different ethnic groups would exhibit different patterns in food preference, given their poverty level, physical activity level, and BMI. It was obvious that White women consumed more vegetable than any other groups. We could observe a linear relationship between income and vegetable consumption and the men/women group differences were consistent across all races. A similar but opposite relationship could be seen between sugar and income. However, all groups shared similar average physical activity level and BMI while Latino men tended to add more sugar in their diet than other groups.

3.2 Design

Our target audience were healthcare business people thus our focus was to reveal relationship in a self-explanatory way. We added as much information to the charts as possible while maintaining a clean design.

3.3 Results

The top figure of Figure 3 indicated that men eat fewer vegetables and drunk more soda. The percentage of obese men was only slightly higher than the percentage of obese women 23% vs. 21%, but the percentage of overweight men (40%) was much larger than the percentage of obese women. Obese and overweight people of both genders ate fewer vegetables than the normal weight group. Overweight and normal and women drunk nearly the same quantity of soda, but there was a notable shift towards higher soda consumption for both genders in the obese group.

The bottom figure of Figure 3 showed that obese people made up the majority of California diabetics (43%), trailed by 37% overweight, and 20% normal weight. We noted that diabetics, fortunately, drunk significantly less soda than non-diabetics. Nevertheless, obese people drunk more soda than normal, whether they were diabetics or not.

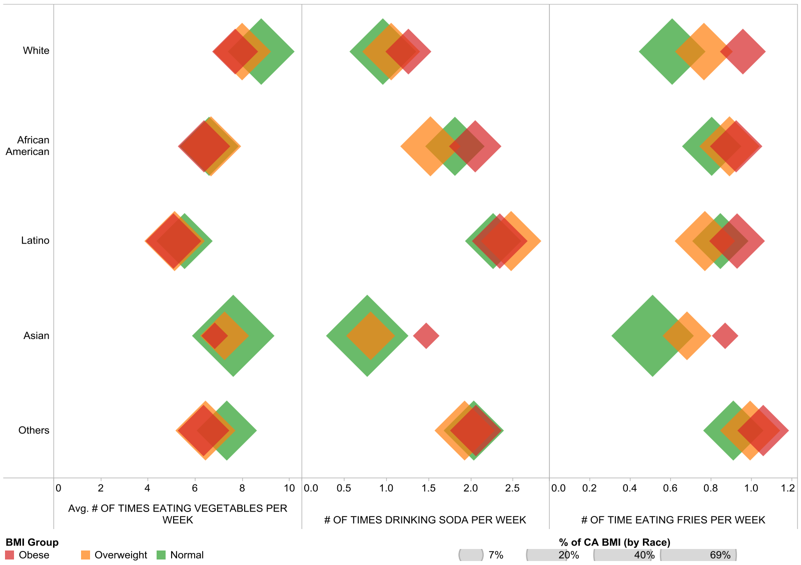


Figure 4. Obesity and Food Preference by Race

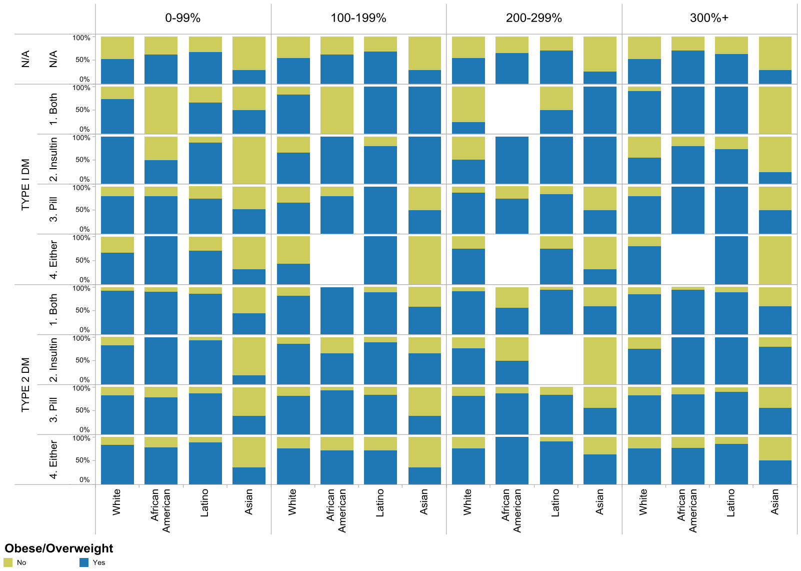


Figure 5. Diabetes and Intervention by Race and Poverty

In Figure 4 we saw that Asians and Whites ate more vegetables and drunk more sodas. Correspondingly, obese and overweight groups formed a relatively larger share of the non-Asian and non-White population. Asian has a much smaller obese population.

We created Figure 5 to investigate the relationship between obesity and race. The proportion of obese/ overweight by race was consistent across all income level. It is interesting to note that Asian was the only group that being obese/overweight beared no relationship to diabetes.

4 Discussion

Visual analytics open up the possibility of rapid and intuitive analysis of complex, high-dimensional data to all users. In seeking important relationships within high dimensional datasets, visualizations exercises such as this enable users to rapidly identify useful connections worthy of deeper analysis. In this exercise, we have visually identified various factors that related to obesity that are worthy of further investigation. It would be very difficult to accomplish that using traditional approach.

5 Acknowledgements

We appreciate the help and support of our employer, Blue Shield of California, and our leaders, Kimball Wilkins and Doug Biehn. We are not compensated for this article and declare no Conflict of Interest. All discussions are our own personal opinions.

6 References

[1] G. P. Rodgers and F. S. Collins, “The next generation of obesity research: no time to waste.,” *JAMA : the journal of the American Medical Association*, vol. 308, no. 11, pp. 1095–6, Sep. 2012.

[2] “Obesity and Overweight for Professionals: Data and Statistics: Adult Obesity - DNPAO - CDC.”

[3] A. L. Diamant, S. H. Babey, T. A. Hastert, and E. R. Brown, “Diabetes: the growing epidemic.,” *Policy brief (UCLA Center for Health Policy Research)*, no. PB2007–9, pp. 1–12, Aug. 2007.

[4] S. Al-Hajj, R. Arias, and B. Fisher, “Interactive Visualization for Understanding and Analysing Medical,” *Proceedings of the IEEE VisWeek Workshop on Visual Analytics in Healthcare: Understanding the Physicians Perspective*, 2011. [Online]. Available: http://interaction-science.iat.sfu.ca/files/f/Interactive Visualization for Understanding and Analysing Medical Data_SamarAlHajj.pdf. [Accessed: 02-Sep-2012].

[5] “About CHIS.” [Online]. Available: http://askchis.com/about.html. [Accessed: 02-Sep-2012].

[6] J. Lopez-Zetina, H. Lee, and R. Friis, “The link between obesity and the built environment. Evidence from an ecological analysis of obesity and vehicle miles of travel in California.,” *Health & place*, vol. 12, no. 4, pp. 656–64, Dec. 2006.

[7] D. H. Sorkin and J. Billimek, “Dietary Behaviors of a Racially and Ethnically Diverse Sample of Overweight and Obese Californians.,” *Health education & behavior : the official publication of the Society for Public Health Education*, Mar. 2012.

[8] T. C. Harjo, A. Perez, V. Lopez, and N. D. Wong, “Prevalence of diabetes and cardiovascular risk factors among California Native American adults compared to other ethnicities: the 2005 California Health Interview Survey.,” *Metabolic syndrome and related disorders*, vol. 9, no. 1, pp. 49–54, Mar. 2011.

[9] S. H. Babey, M. Jones, H. Yu, and H. Goldstein, “Bubbling over: soda consumption and its link to obesity in California.,” *Policy brief (UCLA Center for Health Policy Research)*, no. PB2009–5, pp. 1–8, Sep. 2009.

[10] N. A. Christakis and J. H. Fowler, “The spread of obesity in a large social network over 32 years.,” *The New England journal of medicine*, vol. 357, no. 4, pp. 370–9, Jul. 2007.

[11] N. A. Christakis and J. H. Fowler, “Social Network Visualization in Epidemiology.,” *Norsk epidemiologi = Norwegian journal of epidemiology*, vol. 19, no. 1, pp. 5–16, Jan. 2009.

[12] E. Cohen-Cole and J. M. Fletcher, “Is obesity contagious? Social networks vs. environmental factors in the obesity epidemic.,” *Journal of health economics*, vol. 27, no. 5, pp. 1382–7, Sep. 2008.

[13] A. J. O’Malley, S. Arbesman, D. M. Steiger, J. H. Fowler, and N. A. Christakis, “Egocentric social network structure, health, and pro-social behaviors in a national panel study of Americans.,” *PloS one*, vol. 7, no. 5, p. e36250, Jan. 2012.

3D-Time Series Analysis of Caries Lesion Activity for Oral Health Care

Hui Zhang, Masatoshi Ando, and Michael J. Boyles

Abstract— This paper presents a framework to analyze 3D-time series caries lesion activity based on collections of SkyScan® μ-CT images taken at different times during the longitudinal study. Whereas most current methods for detection and assessment of caries lesions are relying on subjective clinical methods (such as visual or tactile inspection), our work exploits multi-core computing, storage, and interactive visualization to facilitate the segmentation of high-resolution μ-CT images, the construction of 3D models, and the visual analysis of 4D (3D + time) dental images. Our workflow enables quantitative analysis as well as three-dimensional comparison of multiple temporal datasets from the longitudinal dental research studies. Such quantitative assessment and visualization can help us to understand and evaluate the underlying processes that arise from dental treatment, and therefore can have significant impact in the clinical decision-making process and caries diagnosis.

Index Terms—3D-time series analysis, 4D images, Interactive visualization

1 Introduction

Dental caries is an infectious, communicable disease which causes the destruction of teeth via acid-forming bacteria (such as Streptococcus mutans and lactobacilli) found in dental plaque [1]. The acids produced by these bacteria diffuse through the plaque and into the tooth, leaching calcium and phosphate from enamel (*demineralization*), and eventually causing the destruction of tooth structure (cavity). The caries process is preventable and interruptible [15]. The replacement of the minerals lost during the demineralization process is called *Remineralization*. The development of dental caries is dynamic: demineralization of the hard tissue by acid-producing bacterial metabolism (pathological factors) alternates with periods of remineralization by fluoride treatment and plaque control (protective factors). Therefore, caries progression or reversal (demineralization / remineralization, respectively) is determined by the balance of this dynamic process between the biofilm and environmental factors resulting in the loss / gain of tooth mineral (calcium and phosphate). Tooth mineral loss is a direct indication of an overlaying "active" demineralizing bacterial biofilm. If mineral loss outweighs gain, the caries process will eventually progress to the clinical cavitation of the tooth surface.

Assessing caries activity in 3D has traditionally been limited to evaluating the patient "live" in the chair and by subjective clinical methods (i.e., visual or tactile inspection). New 3D imaging technologies (such as Microfocus Computed Tomography or μ-CT images) are just beginning to be introduced into the dental communities in order to evaluate the three-dimensional volumetric relationship of the patient's oral health care data [6]. In this paper, we report one such research study where we developed a workflow to enable the visual exploration of caries lesions and the objective analysis of caries lesion's volumetric changes in longitudinal studies.

2 Our Motivation

3D-Time Series Analysis The use of 3D-time series analysis of volumetric models reconstructed from longitudinal medical image data have a long history, and its importance in validating medical treatment is undisputed [14, 13, 10, 9]. Our longitudinal caries development study uses a demineralization/remineralization model. Extracted human teeth have been collected from dental practitioners in the State of Indiana and transported in 0.1% thymol solution to the Oral Health Research Institute. The collection of human teeth for use in dental laboratory research studies has been approved by the Indiana University, Institutional Review Board (IRB) study $\#0306-64$. Sound (intact) permanent incisors were sterilized with ethylene oxide gas. From these teeth, 195 $3 \times 3 \times 2mm$ blocks were prepared (Figure 1). μ-CT images were scanned from each tooth specimen at several time frames during the caries development process. Specimens were divided into three demineralization group: $group^1$ was demineralized for 3 days, $group^2$ for 6, and $group^3$ for 9 days. All specimen were then remineralized for 4, 6 or 9 days with 1100 ppm F as NaF. For longitudinal evaluation of mineral content change, the μ-CT images of sound, demineralized and remineralized enamel and that of the phantom were acquired using the SkyScan® 1172. The scanning procedure was performed using $65kV$, $153\mu A$, and 2.25 μm pixel size. Tooth mineral loss is a direct indication of demineralized enamel (e.g., determined between 87 and 0 wt% of the sound enamel), and the cross-section of demineralized enamel (i.e., the Region-Of-Interest, or, ROI) shows an observable gray-scale difference from that of sound enamel (e.g., between 87 and 0% of sound enamel's average pixel value). Therefore, by segmenting the ROIs from stacks of μ-CT images, we can measure the tooth mineral loss and reconstruct the caries lesion volume during the dynamic caries process (see Figure 2).

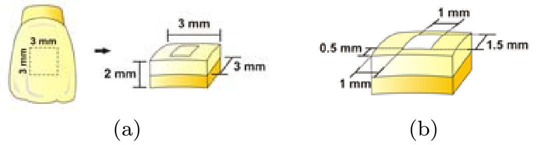


Fig. 1. (a) Schematic diagrams showing specimen dimension. (b) Region of interest (ROI).

Heavy Computation A typical μ-CT scan image in our study is a 16-bit gray-scale of resolution 1636×1120 which requires about 4 MB of disk space. Around 1000 μ-CT images were acquired from every specimen at each point in time. Over the

- *Hui Zhang is with Pervasive Technology Institute at Indiana University, e-mail: huizhang@iu.edu.*
- *Masatoshi Ando is with Indiana University School of Dentistry, e-mail: mando@iupui.edu.*
- *Michael J. Boyles is with Pervasive Technology Institute at Indiana University, e-mail: mjboyles@iu.edu.*

Manuscript received 31 March 2012; accepted 1 August 2012; posted online 14 October 2012; mailed on 5 October 2012.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

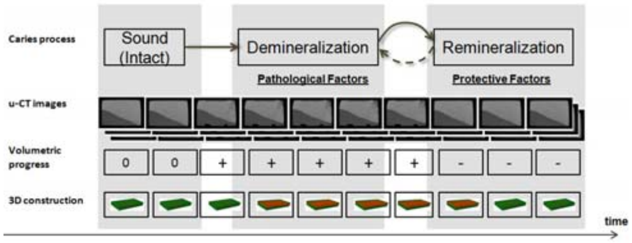


Fig. 2. The flow chart of our longitudinal study procedure. (1^{st} row) Our study uses a demineralization / remineralization model. Demineralization is done by acid-producing bacterial metabolism (pathological factors) and remineralization by fluoride treatment and plaque control (protective factors). (2^{nd} row) During the caries process, a large amount of high-definition high-quality SkyScan® μ-CT images are scanned from tooth specimens. (3^{rd} row) By segmenting the region-of-interest from each cross-sectional image, we can identify the tooth mineral loss / gain during the entire caries process. When a lesion is locally larger than that identified in the previous phase, the lesion is actively growing. The lesion can be stopped and reversed at some point in the artificial caries process. (4^{th} row) Temporal volumetric data can be reconstructed from segmented image stacks, and can be used to visualize the lesion's temporal variation.

entire 5-phase longitudinal evaluation, around 5000 such images were acquired from each single specimen. Applying traditional frame-by-frame image processing methods to our data is a time-consuming process. Furthermore, the constructed geometric models in this study each contain approximately 200,000 facets. Visualizing the development of longitudinal caries lesion activity requires an array (sometimes a matrix) of these large 3D models. Attempting to perform such visualization on a single workstation can be prohibitively expensive.

Toward More Efficient and Effective Caries Lesion Analysis Our workflow consists of three steps: (1) image segmentation, (2) 3D model construction and visualization, and (3) qualitative and quantitative analysis. HPC resources are used throughout each step. Figure 3 shows the architecture of our time series analysis of caries lesion activity. Our main workflow is based on a family of image processing operations for ROI segmentation, with a parallel computing extension that utilizes the computing and storage capabilities of IU's HPC resources. We achieve an order-of-magnitude higher computational power and speed in processing huge dental image collections. We used ParaView to support interactive visualization of large 3D models in our study.

3 Identifying the Cross-section of Demineralized Enamel

An essential step when studying caries lesion activity based on μ-CT image information is to identify the cross-section of demineralized enamel in every image from each image sequence. This process of identification is broadly referred to as *segmentation*. Figure 4(a) shows our initial image, a cross-section of human enamel that has been chemically demineralized for 5 days. The image can be partitioned into 4 segments (or, sets of pixels), labeled in the figure as *air*, *dentin*, *sound enamel*, and *demineralized enamel* (i.e., the ROI in our study). In principle, one can segment the demineralized enamel from such an image by manually drawing an ROI bounding rectangle and then using a thresholding operation. Since our study, however, involves identifying ROIs from a large collection of μ-CT image sequences we discarded algorithms that require significant manual input and created an automated workflow that is suitable for parallel, batch processing. This improved workflow still relies on thresholding but then supplements with despeckling and clustering operations (see Figure 4).

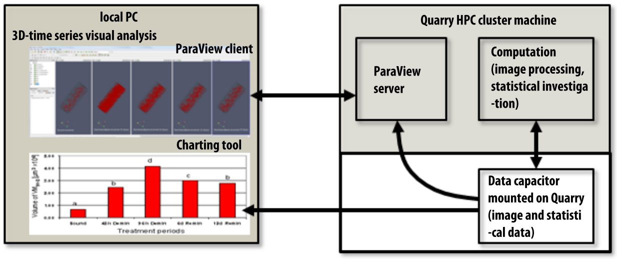

local PC
3D-time series visual analysis
ParaView client
Charting tool
Treatment periods
Sound
a
b
d
c
b
Quarry HPC cluster machine
ParaView server
Computation (image processing, statistical investiga -tion)
Data capacitor mounted on Quarry (image and statisti -cal data)

Fig. 3. Exploiting HPC resources in 3D-time series visualization and analysis. The three main components in our system are (1) parallelized image processing computations on IU's Quarry HPC cluster machine, (2) data storage enabled by IU's Data Capacitor, mounted on Quarry, and (3) the use of the parallel visualization package *ParaView* for visualizing large temporal volumetric data in our study.

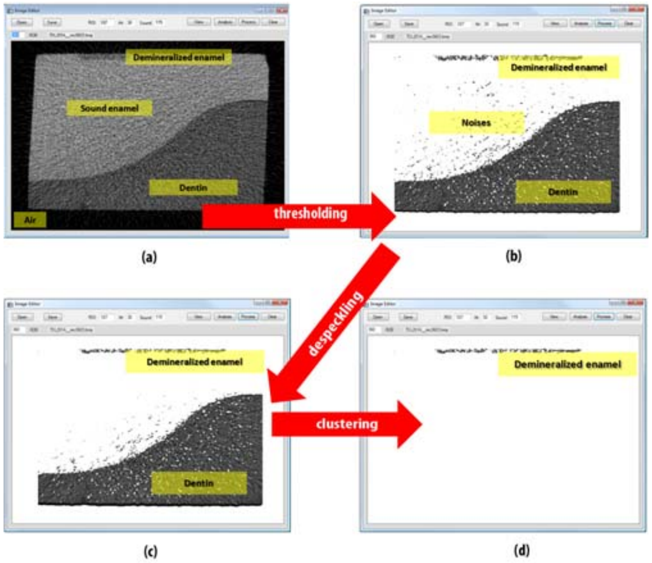
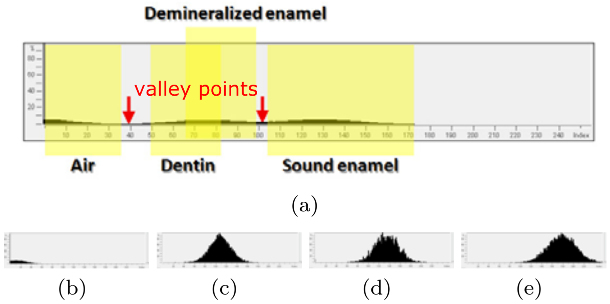


Fig. 4. An example of our image segmentation workflow that consists of three phase: thresholding, despeckling, and clustering. (a) A slice of SkyScan® μ-CT images with four segments labeled. Pixels with the same label share similar visual characteristics (such as intensity and texture). (b) Processed image after thresholding operation contains two segments, as well as a number of residual noises. (c) A cross section of μ-CT images after despeckling. (d) ROI identified after clustering operation.

Thresholding Our image processing operations begin with a thresholding [3] to correct unwanted data. As shown in Figure 5, the whole image has a distribution of gray-scale levels ranging from black (0) to white (255). The gray scales of pixels belonging to demineralized enamel are substantially different from the gray scales of the pixels belonging to air and sound enamel but overlap with the gray scales of the pixels belonging to dentin area. Thresholding is a simple but effective tool to separate demineralized enamel and dentin from air and sound enamel. One method is to first select the two "valley points" ($T_1 = 38, T_2 = 102$, see Figure 5) that separate these two groups. If we assume that the gray-scale histogram corresponds to the μ-CT image, $f(x, y)$, then any point (x, y) for which $f(x, y) > T_2$ is segmented as sound enamel, and any point (x, y) for which $f(x, y) > T_1$ is segmented as air. After this thresholding operation, only the demineralized enamel and dentin, as well as a number of residual noises remain (see Figure 4(b).)

The overlapping of dentin's gray-scale range and deminerlized enamel's gray-scale range make it difficult to segment the region of interest (i.e., the demineralized enamel) using only thresholding. After some experimentation, we determined that performing despeckling and clustering operations post thresholding allowed us to identify the cross-section of the demineralized enamel.



(a)

(b) (c) (d) (e)

Fig. 5. (a) A gray-scale frequency distribution of the μ-CT image in Figure 4(a). The four segments occupy different ranges of gray levels in the histogram. Two valley points can be identified in the gray-scale distribution. (b) Gray-scale distribution of air. (c) Gray-scale distribution of dentin. (d) Gray-scale distribution of lesion area. (e) Gray-scale distribution of healthy enamel.

Despeckling Speckles in our μ-CT images are usually caused by noises. If their size is not comparable to the real features of the demineralized enamel, they mostly can be removed with a despeckling operation. Our despeckling operation is designed to use a connected grid (pixel^2) of a predefined size [8] and removes observable pixels from the dataset if the surrounding grid's averaged gray scale is different from that of demineralized enamel. A good setting for this in our dataset is a 7×7 grid. Figure 4(c) shows the processed image after despeckling.

Clustering After despeckling, the remaining observable pixels can be clustered into two regions in which the upper observation is considered the cross-section of demineralized enamel. We use the constrained K-means clustering algorithm [4, 12] to cluster the remaining observable pixels based on their spatial features in 2D Euclidean space into two groups. Figure 4(d) shows an example of the region-of-interest we classified.

Parallelizing the Segmentation Process Sequentially processing each image from each specimen image stack presents a compelling challenge. For example, we typically acquire 5000 digital images from a single tooth specimen. Using a desktop computer with an Intel Pentium 4 3.20GHz processor, it takes approximately 7.1 seconds to fully process a single μ-CT image. Using the same computer, it would take nearly one hour to process all 5000 images from each tooth. Fortunately, this processing is easily parallelized. The research being reported in this paper used an Indiana University supercomputer called *Quarry* [1], where we were able to acquire 32 nodes (i.e., 256 cores). Segmenting a single image using one core on *Quarry* took about 15 seconds. Despite this significantly slower segmentation execution time for a single image, we achieved much improved overall performance by processing 5000 images within 200 seconds.

4 From Imaging to Visual Analysis

Once the images from each specimen image stack are processed, we then construct a 3D geometric model by "stacking" the slices together and applying a Marching Cube algorithm. We used the $CTan$ software and the "marching cubes 33" as the model generating algorithm. Most of the resulting 3D models have $O(10^5)$ facets. Interacting with such large data requires significant memory resources and imposes large data transfer between the CPU and GPU. This can be problematic for a single local desktop computer. To overcome this problem, we utilize the ParaView high-performance visualization software. ParaView's client/server architecture facilitates parallel computation and rendering [2]. In our study , both the "*data server*" and "*render server*" of ParaView program are configured on IU's Quarry HPC cluster [16].

3D-Time Series Visualization A visualization based on lesion models extracted from the segmented image data at different time frames, in conjunction with possible critical statistical investigation result, is used to illustrate the monitored dynamic caries process. ParaView allows us to explore multiple datasets synchronously by giving the user the ability to present them in multiple views. This functionality is used to visualize the temporal sequence of lesion volumes as a 4D image and analyze the lesion volumes as 4D connected components [7]. Figure 6 shows such an example. This 3D-time series analysis is based on the full three-dimensional comparison of the caries lesion volume of the three specimen (specimen#007 was demineralized for 3 days, #085 for 6 days, and #145 for 9 days). This objective and quantitative assessments of caries lesion activity helps validate the caries progress and caries reversal in our longitudinal study.

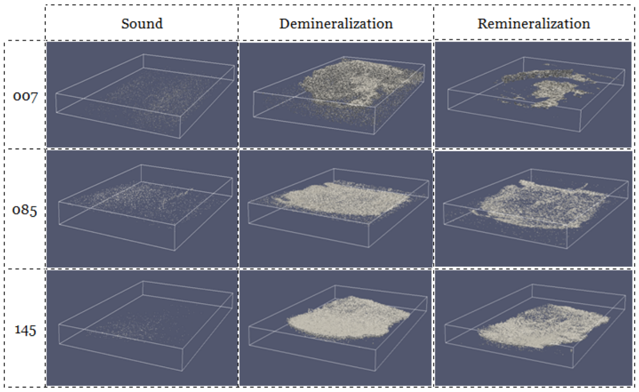


Fig. 6. Examples of 3-Dimensional images of all groups for three periods. Light gray line box indicates the region of interest (ROI: 3.72.80.6mm). Darker gray inside (or within) the ROI represent mineral content between 87 to 30% of sound gray scale.

Table 1. Average and standard errors of mineral volume computed using the μ-CT images ($\times 10^{-3}$ mm^3). $Group^1$ was demineralized for 3 days, $Group^2$ was demineralized for 6 days, and $Group^3$ for 9 days.

	Sound	Demineralization	Remineralization
$Group^1$	1.83(0.44)	15.60(1.30)	2.87(0.81)
$Group^2$		22.86(4.92)	6.67(1.64)
$Group^3$		25.01(2.22)	10.28(1.98)

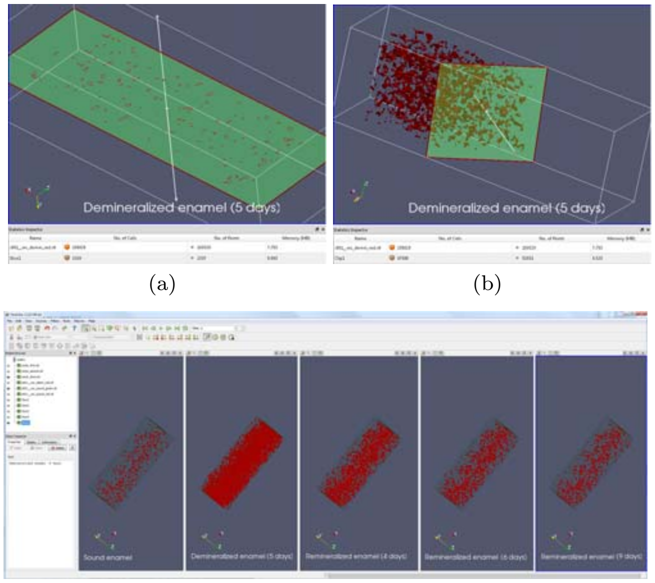
Table 1 shows the artificial lesion volume grows significantly in the demineralization phase, and is reversed in the remineralization phases. 3D-time series analysis can thus serve as an objective and quantitative measure to clinically assess caries lesion activity at the time of examination and is very important for clinical decision making and for patient follow-up (see e.g., [5]).

Interacting with Volumetric Models Although 3D visualization of volumetric models can provide a qualitative understanding of the data (e.g., the density and volume size), gaining quantitative insights from large 3D models can be challenging because the high data density makes it difficult to view all the data at once. In many tasks, both 3D and 2D visualization strategies are needed to better interpret data. For example, Springmeyer et al. observed that 2D techniques are often used to establish precise relationships between parameters, whereas 3D views are typically used to gain a qualitative understanding of the data and sometimes comparative understanding between multiple datasets [11].

We can apply various filters to ParaView's visualization pipeline to combine 2D slices (cross-sections) with a 3D overview. Figure 7(a) shows one way to peer inside of a volume is to perform a *Slice* on it. The *Slice* filter will intersect a

[1] http://pti.iu.edu/hps/quarry/

volume with a plane and allow you to see the data in the volume where the plane intersects. *Slicng* is a good mechanism to validate features on some specific locations inside the lesion volume. Figure 7(b) shows another example where a *Clip* filter is applied to obtain a subset of the original volumetric data. This is especially useful if we want to extract and compare sub-regions across a large data model. Figure 7(c) shows a 3-dimensional model of portion of ROI over all experiment phases.


Demineralized enamel (5 days)
Demineralized enamel (5 days)
Sound enamel
Demineralized enamel (5 days)
Remineralized enamel (4 days)
Remineralized enamel (6 days)
Remineralized enamel (9 days)

(a) (b)

Fig. 7. (a) Intersects the lesion volume with a plane. The effect is similar to clipping except that all that remains is the geometry where the plane is located, in this way we can examine the distribution and density of demineralized enamel inside a volumetric model. (b) Intersects the lesion volume with a half space. The effect is to remove all the geometry on one side of a user-defined plane. (c) 3D visualization, that allows a time series analysis based on μ-CT images. Time series analysis stands for comparing image data sets from the same person or specimen taken at different times to show the changes.

5 Conclusion

Compare to visual and tactile examinations that are the most common but subjective ways to examine caries lesion activity, 3D-time series analysis based on μ-CT images can evaluate lesion activity by assessing the caries lesion developed with an active biofilm, stopped and reversed by means of a remineralization protocol. We are most interested in exploiting imaging technology, computational algorithm, and visualization methods to make caries activity assessment fast and accurate. To sum up, we report our first steps to exploit high performance computing resource in the 3D-time series analysis of caries lesion activity, based on a large collection of acquired μ-CT images. The current results have suggested that multi-core computation on HPC platforms will have the best chance to enable such scientific computation and analysis faster and more accurate than ever before. In addition, we introduce the use of ParaView for the visual analysis of large 3D models constructed from processed images.

6 Acknowledgments

This work was supported in part by NIH/NIDCR grant R21 DE018390-01A2. We gratefully acknowledge Eduardo K. Kohara's contribution to the μ-CT image data acquisition, and Huian Li's contribution to the implementation of multi-core image processing functionality used in this paper.

References

[1] `http://en.wikipedia.org/wiki/Dental_caries`.

[2] A. Cedilnik, B. Geveci, K. Moreland, J. P. Ahrens, and J. M. Favre. Remote large data visualization in the paraview framework . In A. Heirich, B. Raffin, and L. P. P. dos Santos, editors, *EGPGV*, pages 163–170. Eurographics Association, 2006.

[3] R.-M. Chao, H.-C. Wu, and Z.-C. Chen. Image segmentation by automatic histogram thresholding. In *Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human*, ICIS '09, pages 136–141, New York, NY, USA, 2009. ACM.

[4] T.-W. Chen, Y.-L. Chen, and S.-Y. Chien. Fast image segmentation based on k-means clustering with histograms in hsv color space. In *Multimedia Signal Processing, 2008 IEEE 10th Workshop on*, pages 322 –325, oct. 2008.

[5] G. Gerig, D. Welti, C. Guttmann, A. Colchester, and G. Szkely. Exploring the discrimination power of the time domain for segmentation and characterization of lesions in serial mr data. In W. Wells, A. Colchester, and S. Delp, editors, *Medical Image Computing and Computer-Assisted Interventation MICCAI98*, volume 1496 of *Lecture Notes in Computer Science*, pages 469–480. Springer Berlin Heidelberg, 1998.

[6] H. Hong, H. Lee, Y. G. Shin, and Y. H. Seong. Three-dimensional brain ct-dsa using rigid registration and bone masking for early diagnosis and treatment planning. In *Proceedings of the Third Asian simulation conference on Systems Modeling and Simulation: theory and applications*, AsiaSim'04, pages 167–176, Berlin, Heidelberg, 2005. Springer-Verlag.

[7] D. Metcalf, R. Kikinis, C. Guttmann, L. Vaina, and F. Jolesz. 4d connected component labelling applied to quantitative analysis of ms lesion temporal development. In *Engineering in Medicine and Biology Society, 1992 14th Annual International Conference of the IEEE*, volume 3, pages 945 –946, 29 1992-nov. 1 1992.

[8] O. Michailovich and A. Tannenbaum. Despeckling of medical ultrasound images. *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*, 53(1):64 –78, jan. 2006.

[9] D. Rey, G. Subsol, H. Delingette, and N. Ayache. Automatic detection and segmentation of evolving processes in 3d medical images: Application to multiple sclerosis. In *Proceedings of the 16th International Conference on Information Processing in Medical Imaging*, IPMI '99, pages 154–157, London, UK, UK, 1999. Springer-Verlag.

[10] J.-P. Thirion and G. Calmon. Measuring lesion growth from 3d medical images. In *Nonrigid and Articulated Motion Workshop, 1997. Proceedings., IEEE*, pages 112 –119, jun 1997.

[11] M. Tory, T. Moller, M. S. Atkins, and A. E. Kirkpatrick. Combining 2d and 3d views for orientation and relative position tasks. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '04, pages 73–80, New York, NY, USA, 2004. ACM.

[12] K. Wagstaff, C. Cardie, S. Rogers, and S. Schrödl. Constrained k-means clustering with background knowledge. In *Proceedings of the Eighteenth International Conference on Machine Learning*, ICML '01, pages 577–584, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc.

[13] A. Wismller, O. Lange, D. R. Dersch, G. L. Leinsinger, K. Hahn, B. Ptz, and D. Auer. Cluster analysis of biomedical image time-series. *International Journal of Computer Vision*, 46:103–128, 2002. 10.1023/A:1013550313321.

[14] G. Wollny. Analysis of changes in temporal series of medical images, 2004.

[15] D. Zero, M. Fontana, E. Martnez-Mier, A. Ferreira-Zandon, M. Ando, C. Gonzlez-Cabezas, and S. Bayne. The biology, prevention, diagnosis and treatment of dental caries: scientific advances in the united states. *J Am Dent Assoc*, 140 Suppl 1, 2009.

[16] H. Zhang, H. Li, M. J. Boyles, R. Henschel, E. K. Kohara, and M. Ando. Exploiting hpc resources for the 3d-time series analysis of caries lesion activity. In *Proceedings of the 1st Conference of the Extreme Science and Engineering Discovery Environment: Bridging from the eXtreme to the campus and beyond*, XSEDE '12, pages 19:1–19:8, New York, NY, USA, 2012. ACM.

Program Committee

Organizers:

Jesus J Caban, PhD
NICoE / Naval Medical Center
CC / National Institutes of Health

David Gotz, PhD
IBM T.J. Watson Research Center

Program Committee	
Catherine Plaisant, PhD	University of Maryland, College Park
Terry S. Yoo, PhD	National Library of Medicine
Adam Perer, PhD	IBM Research
Dan Mollura, MD, PhD	Center for Infectious Disease Imaging, NIH
Ketan Mane, PhD	University of North Carolina, Chapel Hill
Paul Nagy, PhD	John Hopkins University

Author Index

Walker	Ryan	Blue Shield of California	25
Walter	Andreas	Bern University Hospital	5
Zhang	Zhiyuan	Stony Brook University	11
Zhang	Hui	Indiana University	29
Zhuo	Wei	Georgia Institute of Technology	13