OJNI FEATURE ARTICLE

Data Visualization and Large Nursing Datasets

Deborah Vincent PhD, APRN-BC, FAANP,

Marie Hastings-Tolsma, PhD, CNM

and Judith Effken, PhD, RN, FACMI, FAAN

Citation:

Vincent, D., Hastings-Tolsma, M. & Effken, J. (June, 2010). Data Visualization and Large Nursing Datasets. *Online Journal of Nursing Informatics (OJNI),14* (2). Available at http:ojni.org/14_2/vincent.pdf

Abstract

New data visualization software applications allows for the discovery of patterns and trends in large datasets that might not be apparent with traditional data visualization approaches. The purpose of this research was to data mine an existing large dataset utilizing software developed for high-end visual applications to determine whether new patterns or trends that were not evident in prior analyses might be uncovered (Hastings-Tolsma, Vincent, Emeis, & Francisco, 2007). Nurse midwifery data (N=510) had previously been examined to identify factors related to perineal trauma in childbirth; this research reanalyzed the data using Tableau software. Lessons learned about the use of data visualization are presented and challenges and opportunities for using this tool to explore large data sets are discussed.

Key Words: data visualization, information visualization, large datasets, data mining, midwifery

OJNI FEATURE ARTICLE

Introduction

Nurse researchers are increasingly likely to utilize large datasets to examine health care outcomes as they work to solve problems affecting human health. Large datasets present opportunities to explore data in new and creative ways. A significant challenge, however, is how to understand the patterns in the data that are both novel and important, as well as how to present those findings in a more powerful fashion. Data visualization tools may assist researchers in presenting relationships in a more meaningful way.

Humans perceive most information in life in a visual form and rely on pattern recognition to process information. It is, therefore, preferable to use visualization tools for understanding a problem and assessing available data, especially in the case of large datasets (Popov, 2006). Data visualization is a tool that is receiving increased attention for its potential to visually clarify important patterns in research data. Using creative ways to present the complex patterns and relationships in data has the potential to engage others in understanding the meaning of data. This creative synthesis promotes data comprehension, especially in datasets with high dimensionality (Popov, 2006; Wang & Yang, 2005), such as patient medical record data (Harrison, 2008).

Data visualization has been defined as the graphical representation of data to facilitate the process of information extraction (Sánchez & Plassmann, 1999; Vande Moere, 2007) and the generation of hypotheses (Patterson & Basham, 2003). Researchers are familiar with standard data visualization in the form of charts and graphs; however, many are unfamiliar with more complex visualizations. Powerful software can transform data into 2 dimensional or 3 dimensional images or animations that are fixed or dynamic (Educause, 2007). Such visualizations demand software applications capable of creating sophisticated images and animations. This approach of using esthetics or beauty in understanding data, sometimes called *infosthetics,* has the potential to invent new visual metaphors for presenting data and new ways to manipulate these metaphors to make sense of the information (Manovich, 2000).

An example of a data visualization software application is Gapminder. Gapminder was developed to present information about social, economic and environmental development at local, national and global levels (Gapminder, n.d.). G*apminder World* allows for the dynamic visualization of data as can be seen in the global maternal mortality data (Gapminder World, November 19, 2007). These visualizations stimulate reconsideration of statistical data and

OJNI FEATURE ARTICLE

subsequent relationships, bringing order to apparent chaos (Homeland Security, 2009). Cross-disciplinary work and collaborations (Educause, 2007) may also be enhanced by the use of visualization tools because they can assist the research team in uncovering hidden connections. For example, work by Bollen and colleagues (2009) created a map of scientific literature from citation data to visualize the structure of scientific activity. The map provided a novel and comprehensive view of scientific activity, clarifying the use of scholarly works in varied disciplines (Bollen, Van de Sompel, Hagberg, Bettencourt, Chute, et al., 2009). A similar effort might detail scholarly activity within nursing.

Nurse researchers can use visualization tools to present data in a way that helps generate new knowledge and understanding. These tools can be particularly useful when data mining has been conducted to extract useful information from large datasets (Hand, Mannila, & Smyth, 2001). The collection of large amounts of data can be mined, then synthesized and analyzed to find patterns, trends and associations that would otherwise be unappreciated. Visualization of the data allows for enormous amounts of information to be quickly processed, emphasizing visual reasoning (Sánchez & Plassmann, 1999) through the use of design metaphors which exploit the ability of humans to organize objects in space (Vande Moere, 2005).

The purpose of this paper is to describe the process and outcomes of a secondary analysis of an existing large dataset using data visualization software. By using data visualization, we were able to identify patterns not appreciated on previous standard analysis. This example of data visualization may stimulate other nurse researchers to consider using these tools.

Materials and Methods

Exploring the Data

Data were obtained from the Nurse Midwifery Clinical Data Set (NMCDS), which is a standardized data collection instrument that describes patient characteristics, nurse midwifery antepartum and intrapartum interventions and birth outcomes. Developed by the American College of Nurse Midwives, the instrument has demonstrated content and criterion related validity; overall agreement on items was more than 86.6% (Greener, 1991).

The record was completed by the nurse midwife birth attendant for patients receiving prenatal care either at a university-based practice site or one of three community

OJNI FEATURE ARTICLE

health clinics. Births took place at a university tertiary care facility, and all were attended by experienced nurse midwives who were prepared at or above the master's level. Data were gathered related to the prenatal, intrapartal, postpartal, and neonatal care of 510 patients receiving care with a large nurse midwifery service. The study was approved by the Institutional Review Board at the University of Colorado Denver.

These data were entered into SPSS for traditional statistical analysis. Results have been previously reported (Hastings-Tolsma, Vincent, Emeis, & Francisco, 2007). Following this conventional approach to displaying data (i.e., charts, tables) data were re-examined using data visualization software

Choosing Visual Analytic Software

For researchers new to data mining and visual analysis, visual analytic tools that are relatively easy to use, have a simple user interface, and are reasonably priced, are highly desirable. Because data visualization is a relatively new area, especially in nursing, there may be relatively little guidance available from information technology support staff or clinical informaticist. We began our exploration on the Internet, read widely about various data visualization tools, and explored several that were commercially available. Many data visualization software tools are costly or otherwise unsuitable for novices who are beginning to conduct data mining and visualization.

While there are a variety of data visualization software products available, we chose Tableau Desktop Software (v 5.0) (Tableausoftware, 2009) to re-examine the NMCDS dataset. The free trial period was used to explore the software and conduct analyses. Additionally, the software package provided understandable directions, examples, visual analysis tools, and used a familiar drag and drop interface. Tableau also has server automation, the ability to assemble related data in different views (also referred to as *dashboards*), and to highlight and filter data (also referred to as *brushing* and *linking*). This later feature allows for the exploration of large data sets to filter rather than to highlight the data.

New to data visualization, we initially analyzed relationships found in an earlier investigation, examining factors related to perineal trauma in a low-risk nurse midwifery population (Hastings-Tolsma et al., 2007). In particular, perineal laceration and parity were examined, using variables that had previously been analyzed using SPSS to conduct t-tests and generate charts and graphs. These data were compared to

FEATURE ARTICLE

visualizations generated by Tableau. Results of the initial study had shown that parity was protective against all types of laceration and that multiparous women were less likely to experience perineal tears (Hastings-Tolsma et al., 2007). Finding the same relationships using the visualization tool reassured us that we were using the software correctly and correctly interpreting the data visualization findings. We then undertook additional visualization of the data in an attempt to discover patterns or relationships that our previous analysis may have missed.

Results and Discussion

First, the relationship between perineal lacerations (tears), parity and BMI was examined using a standard graphics format (Figure 1). The degree of laceration is shown in the far left column and ranges from one (least severe) to five (most severe). More than half (55.7%) of women who had a spontaneous vaginal birth sustained some sort of perineal laceration. As expected, women with lower parity were more likely to sustain more severe perineal tears. And, as in the previous analysis, there was no apparent relationship between BMI and severity of perineal tears.

Figure 1. Perineal lacerations (tears), BMI, and parity (N=510)

Midwifery Data Set

Sheet 1

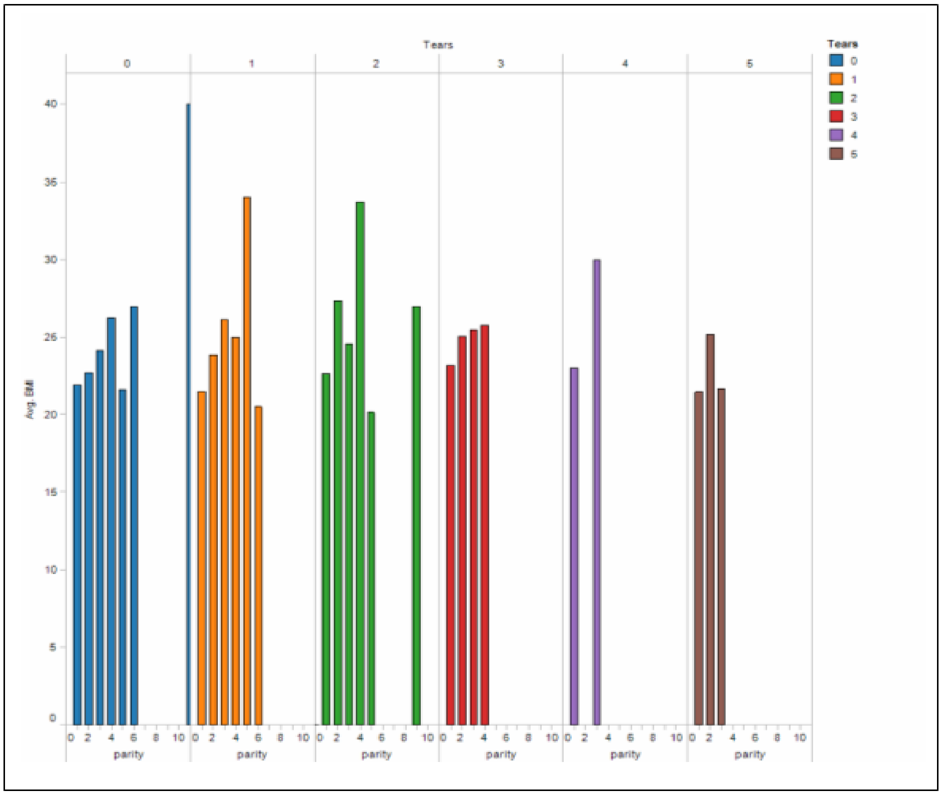
tears	Avg. bmi	Avg. parity
Null	23.47	1.43
5	22.89	1.29
4	24.58	1.23
3	25.23	2.01
2	23.74	2.04
1	23.19	2.40

FEATURE ARTICLE

Next, the relationship between BMI and perineal laceration (tears) were explored adding parity to the visualization technique (Figure 2). The inverse relationship between low parity and more severe perineal lacerations is evident. More severe lacerations, represented by the purple or brown colors, are more common in women with parity under four and are absent in women where parity is four or greater. Again, BMI demonstrated no significant relationship with either severity of lacerations or parity and age was not associated with lacerations.

Figure 2. BMI, Parity, and Perineal Lacerations (tears) (N=510).

Coding of tears: 0 =none, 1 = minor, 2= first or second degree, 3 = third degree, 4 = fourth degree, 5 = other (cervical/vaginal).

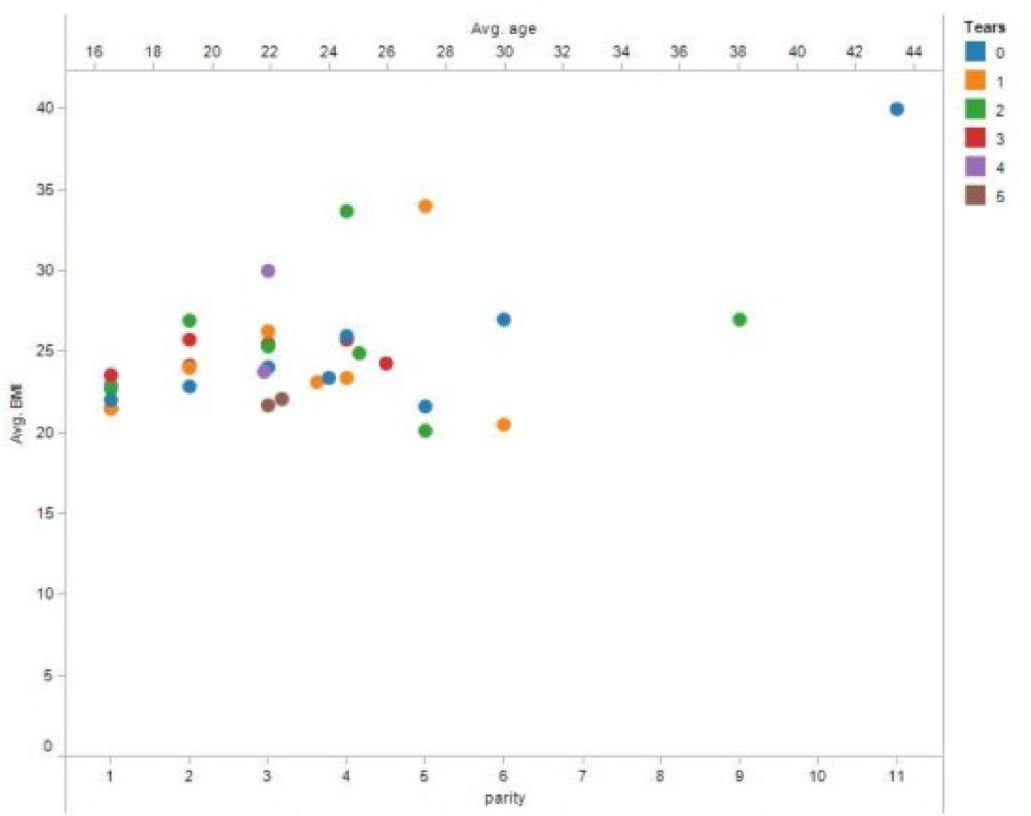


Additionally, a scatterplot was done to determine if previously unappreciated relationships among the variables might exist (Figure 3).

OJNI FEATURE ARTICLE

Figure 3. Parity, Age, BMI, and Perineal Lacerations (tears) (N=510).

Coding of tears: 0 =none, 1 = minor, 2= first or second degree, 3 = third degree, 4 = fourth degree, 5 = other (cervical/vaginal).

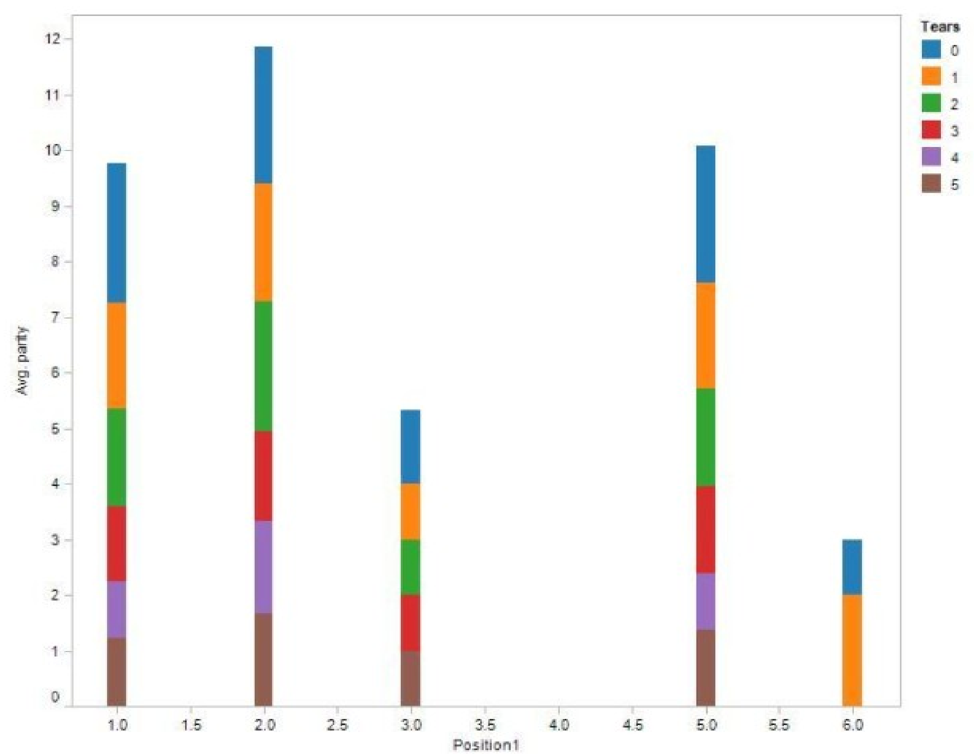


Parity and average of age vs. average of BMI. Color shows details about Tears. The data is filtered on Tears and ethnicity. The Tears filter ranges from 0 to 5. The ethnicity filter keeps 1, 2, 3, 4 and 5. The view is filtered on average of age, which ranges from 15.000 to 40.000.

OJNI FEATURE ARTICLE

Although no new patterns were noted, outliers in the data are readily apparent and these may be the most interesting findings. Reproductively older women (over 40) with high parity and a BMI greater than 30 were without lacerations whereas women older than 35 but younger than 40 years of age with high parity but a normal BMI, had significant second degree lacerations. Older, morbidly obese multiparous women may be more likely to have intact perinea related to factors not accounted for in the original investigation such as underlying medical co-morbidities that were not previously examined. The finding of second degree lacerations in women of high parity but normal

Figure 4. Maternal Position, Age, and Perineal Lacerations (tears) (N=510)



The plot of average of parity for Position1. Color shows details about Tears. The data is filtered on average of bmi, which ranges from 15 to 40. The view is filtered on Tears and Position1. The Tears filter keeps 6 members. The Position1 filter ranges from 1.000 to 6.000.

OJNI FEATURE ARTICLE

BMI is in contradiction to our original analysis, which found that parity was protective against laceration. According to that analysis, multiparous women were less likely to tear than were nulliparous women. It is unclear what might account for this finding and it warrants further exploration. While no association was found in the original analysis between birth weight and laceration (Hastings-Tolsma, Vincent, Emeis & Francisco, 2007), it may be an important factor to re-examine to explain why women with high parity but normal BMIs had significant second degree lacerations.

Finally, the relationships among position during birth, parity and perineal lacerations were examined (Figure 4 on previous page). The coding for positions was as follows: 1 = ambulatory, 2= sitting, 3= hands and knees, 4= squatting and 5 = lying down.

This analysis enabled exploration of relationships and the identification of outliers; color coding of position at birth facilitated pattern recognition. The visual analysis tool clearly depicts that women who gave birth to their first child were more likely to sustain severe perineal tears than women who had given birth three or more times. Further, position may be related to perineal lacerations. Fewer perineal lacerations were sustained in women who were kneeling or squatting during birth. However, this may be an artifact as few women in our study gave birth in these positions. Findings such as these must always be carefully reviewed by the researchers doing the analysis.

Lessons Learned

Proficiency in the use of data visualization software requires patience to learn – even for simple visualization analyses. For many faculty researchers, information technology support services for visual analytic software packages are few or non-existent. Even though software user guides offer some assistance, the terminology may be unfamiliar; so using a visual analysis tool can require a significant amount of experimentation and exploration.

A goal of this research was to implement data visualization for a known nursing dataset. Time-limited access to a publically available software tool (i.e., Tableau Software) to conduct data visualization for the existing dataset allowed the researchers to explore its capabilities. However, there were several significant constraints as the process of data visualization is both complex and technically challenging.

FEATURE ARTICLE

The Tableau Visual Analysis (v 5.0) provides on-line real-time and on demand streaming video training sessions. However, participating in on-demand sessions and developing basic skills in using the data visualization software are time consuming. It is possible to watch the training videos without using the analysis tool, but we found that it is more helpful to engage in interactive training sessions using the sample data provided in the free trial down-load. Participating in the tutorials or training sessions offered with commercial software applications facilitates the skill needed to most effectively conduct data visualization.

In addition to a learning curve and the monetary expense, data visualization presents other challenges. For example, the data visualization software utilized in this research was found to be incompatible with SPSS, which meant our data needed to be first exported from SPSS into Excel, creating a duplicate data base. Additionally, the SPSS coding format used for statistical analysis for some variables made display of data in an understandable graphical format difficult. Data for perineal lacerations needed to be entered in a different format than that used in SPSS.
Our exploratory data visualization failed to add anything to our understanding of the data that was not discovered in the simple tables developed using traditional statistical analyses.

It is possible there were no further significant relationships, although the outliers noted were not further analyzed and may warrant further consideration (Patterson & Basham). A known limitation of data visualization is the potential for showing nothing of consequence (Educause, 2007) or misinterpretation of the visualizations (Rushmeier, Dykes, Dill & Yoon, 2007). Lack of expertise related to the software can contribute to a failure to uncover new patterns or relationships. Collaboration with a skilled visual analyst, as well as software training sessions, is an important consideration to maximize results.

The possibilities for analyses presented by data visualization are intriguing. It can be a dynamic and engaging method for presenting statistics and may be highly useful in the education of students, colleagues, policy makers and the public. Using data visualization tools, researchers can produce graphical displays of large amounts of information, from multiple perspectives and sources, in an engaging manner that result in lasting impressions. Data visualization allows for networking more clearly and effectively with other researchers through graphical means that are both aesthetically pleasing and functionally useful. Data can be mined to uncover new relationships between variables and previously undiscovered patterns. However, data visualization can also exaggerate the significance of certain trends or patterns and lead to flawed or misleading conclusions and there may be difficulties where data sets are incomplete or faulty (Educause, 2007; Popov, 2006).

FEATURE ARTICLE

Conclusions

While having the potential to transform large, multivariate data into images of the phenomena measured, the application of data visualization in the social and human sciences has been limited. A relative lack of familiarity with 3D representations has been one impediment to the use of more advanced visual analytic techniques. Use of newer graphical software applications require time, experience and education (Patterson & Basham, 2003) but should be considered by nurse researchers and should be included in programs preparing nurse scientists (Rushmeier, Dykes, Dill & Yoon, 2007; Skiba & Barton, 2009).

Data visualization presents exciting new possibilities for exploring large datasets. It can facilitate presentation of information in an engaging and informative manner. Improved software makes it possible for nurse researchers to gain additional insight into a dataset that might otherwise be less apparent. However, data visualization can be time consuming, exaggerate patterns and lead to spurious or erroneous conclusions (Educause, 2007). Therefore, as with any other type of analysis it should be considered within the context of theory, and surprising results should be rechecked.

As an important tool for researchers to gain additional insight into a dataset, data visualization has the potential to uncover previously unseen patterns in data and allow for researchers to connect in an exciting way, consistent with the concept of Web 2.0 (Skiba, 2006a, 2006b). Such applications have enormous potential for nursing science though research is needed to determine how nurses extract meaning from data visualization applications and the subsequent utility in understanding varied phenomena (Bakken, Stone & Larson, 2008). Use of novel data visualization techniques should foster greater understanding of more complex datasets in a way that is both aesthetic and functional.

Acknowledgments

The authors would like to thank Diane Skiba, PhD, FAAN for her support and encouragement in the preparation of this work. We also wish to thank Jeff Mills of Tableau Software for his help with the figure preparation. Portions of this paper were presented at the Western Institute for Nursing 53rd Annual Conference in Salt Lake City, UT on April 23, 2009.

FEATURE ARTICLE

References

Bakken, S., Stone, P.W., & Larson, E.L. (2008). A nursing informatics research agenda for 2008 18: Contextual influences and key components. *Nursing Outlook*, 56(5), 206-214, 214e1-e3.

Bollen, J., Van de Sompel, H., Hagberg, A., Bettencourt, L., Chute, R., Rodriguez, M.A., & Balakireva, L. (2009). Clickstream data yields high-resolution maps of science. *PLoS ONE*, 4(3), 1-11, e4803.

Educause. (2007). 7 things you should know about data visualization. Educause Learning Initiative. http://net.educause.edu/ir/library/pdf/ELI7030.pdf

Greener, D. (1991). Development and validation of the nurse midwifery clinical data set. *Journal of Nurse Midwifery, 36,* 174-183.

Harrison, J.H. (2008). Introduction to the mining of clinical data. *Clinics in Laboratory Medicine, 28, 1-7.*

Hand, D., Mannila, H., & Smyth, P. (2001). *Principles of Data Mining.* Cambridge MA: MIT Press.

Hastings-Tolsma, M., Vincent, D., Emeis, C., & Francisco, T. (2007). Getting through birth in one piece: Protecting the perineum. *MCN The American Journal of Maternal Child Nursing*. 32(3), 158-164.

Homeland Security. (2004). DHS launches national center focusing on data visualization. *Journal of Environmental Health.* 67(3), 38.

Manovich, L. (2000). Info-aesthetics. Retrieved from http://www.manovich.net/IA

Patterson, D.A., & Basham, R.E. (2003). Visualizing change: Spreadsheets and graphical representation across domains in human service practice. *Journal of Technology in Human Services*, 21(4). 1-16. doi: 10.1300/J017v21n04_01

FEATURE ARTICLE

Popov, S. (2006). Large-scale data visualization with missing values. *Technological and Economic Development of Economy*, 12(1), 44-49.

Rushmeier, H., Dykes, J., Dill, J., & Yoon, P. (2007). Revisiting the need for formal education in visualization. *IEEE Computer Graphics and Applications*, 27(6), 12-16. doi:10.1109/MCG.2007.156

Sánchez, A., & Plassmann, P. (1999). Dynamic medical data fusion and data mining in the wound-healing process. *Health Informatics*, 5, 213-16.

Skiba, D.J. (2006a). Collaborative Tools for the Net Generation. *Nursing Education Perspectives*, 27(3), 162-163.

Skiba, D.J. (2006b). Web 2.0: Next great thing or just marketing hype? *Nursing Education Perspec*tives, 27(4), 212-213.

Skiba, D.J., & Barton, A.J. (2009). Using social software to transform informatics education. *Studies in Health Technology and* Informatics, 146, 608-12.

Vande Moere, A. (2005). Form follows data: The symbiosis between design & information visualization. *CAADfutures, OKK Verlag,* Vienna, Austria (31-40). Retrieved August 21, 2009 from http://web.arch.usyd.edu.au/~andrew/publications/caadfutures05.pdf

Vande Moere, A. (2007). Aesthetic data visualization as a resource for educating creative design. In A. Dong, A. Vande Moere, and J. S. Gero (Eds.), *Computer-Aided Architectural Design Futures 2007: Proceedings of the 12th International CAAD Futures Conference (pp. 71-84)*.

Wang, W., & Yang, J. (2005). Mining high dimensional data. *Data mining and knowledge discovery handbook: A complete guide for practitioners and researchers*. Norwell, MA: Kluwer Academic Publishers.