The Challenges of Disparate Data Formats: Analysis and Visualization in the SLIDES Project

Ang Li1, Steven Chal, MS2, Sherry Wenshun Liu1, Allison Vorderstrasse, DNSc1, Constance M. Johnson, PhD1
1 Duke University, Durham, North Carolina
2 Renaissance Computing Institute, Chapel Hill, North Carolina

Type 2 Diabetes (T2D) is a chronic disease affecting 1 in 3 adults in the U.S.

Individuals with T2D provide 65% of their own care.

Self-management includes diet, exercise, glucose testing, etc.

Self-management guidelines vary with T2D. Achieving metabolic control, which reduces the morbidity and mortality associated with this disease.

Assessing this gap through self-management is challenging for many people with T2D.

Innovative interventions to empower patients in diabetes self-management are needed.

In response to this need, we developed and tested the SLIDES Project, an intervention for adults with T2D who use a virtual environment (VE) to provide education about diabetes self-management and ongoing support.

To analyze the multidimensional data collected during a 6-month test of this intervention, we employed innovative visualization techniques.

SLIDES Project: Intervention for Adults with T2D

The SLIDES Project (Health: Second Life Impacts Diabetes Education and Self-Management) was designed to accommodate data in disparate formats.

This visualization tool was designed to accommodate data in disparate formats.

The sequence of screen shots above is from an animated browser window showing how each subject's weight, HbA1c (indicator of metabolic control), and corresponding VE activity level changed over the course of the study (6 months).

This visualization was used to see through animation how each subject's weight, HbA1c (indicator of metabolic control), and corresponding VE activity level changed over the course of the study (6 months).

The area of the circle represents the amount of that person's participation during the study.

Weekly touchpoints were represented by a circle with a unique color, so individuals can be followed over time.

When the mouse is clicked over any circle, the visualization will identify the subject.

Perceived usefulness

Perceived ease of use

Quantitative data

Survey and metabolic information at 3rd month

Qualitative data

Baseline survey administered

Graphics can be combined to compare changes over time in variables with disparate data formats.

Further Examples of Visualization

BMI Change Over 100 days

Changes in Self-Efficacy and Social Support Over 6 Months

Conclusions

We developed an internet data visualization tool for analysis of data in disparate formats collected in the SLIDES Project.

The tool generated a variety of visual representations to facilitate recognition of patterns and temporal relationships in the data.

These visualizations helped to show how data clusters, which in turn helped researchers to see new dimensions in large and diverse data sets.

Spatial transformation of data collected from SLIDES participants in disparate formats revealed trends and phenomena that could not be identified using traditional graphs.

Visualization of the data with animation made salient information more apparent, and enhanced our understanding of how time and group dynamics may affect self-management of T2D.

Major changes remain in integrating the wide variety of data gathered in the SLIDES study, including problems with rendering new audio data into a format suitable for visualization, extracting metadata from the resulting text, and, especially performance.

Finding solutions to these problems could be beneficial to researchers who use mixed methods approaches to analyze healthcare "big data." As the interactive web becomes a popular mode of delivery of health information, we need interactive research tools to better understand the outcomes of these studies.

Visualization: Why?

Visualization provides multiple views of multidimensional data beyond the traditional 2-D perspective.

Using color, shapes, volume, and animation facilitates exploration of the data through visual images.

Dynamic changes can be portrayed over a period of 6 months.

Animation allows combination of multiple types of data, highlighting temporal aspects of the data.

Visualization: How?

Initially, visualizations were created using the R environment (open-source data analysis software).

Visualization capabilities were extended by:

- Integrating a broad range of visualization packages, including those that allow for the creation of heat maps, trees, and maps;
- Applying 3D, 4D, and 5D visualizations (where color and animation can represent three or more variables);

Using color, shapes, volume, and animation facilitates exploration of the data through visual images.

Dynamic changes can be portrayed over a period of 6 months.

Animation allows combination of multiple types of data, highlighting temporal aspects of the data.

Visualization Tool

The SLIDES community

SLIDES Study Design

One group, Pre-Mid-Past Study

Phase 1: First month

- Group is the key.
- Baseline survey administered.
- Metabolic information, diet, weight, BP.
- Visit dim (week 1, classes, interactive sessions, other).

Phase 2: Next five months

- Visit site in dim (classes, interactive sessions, other).
- Survey and metabolic, information at 3rd month.

Phase 3: Sixth month

- Survey, metabolic information, and focus groups.

Results

Animated Visualization of Changes in Subjects’ Weight, HbA1c, and Activity In Virtual Environment

We collected multidimensional data over 6 weeks.

- Quantitative data included: Photos, videos.
- Quantitative variables included:
  - Body mass index
  - Metabolic risk
  - Heart rate
  - Self-efficacy
  - Quality of life
  - Satisfaction
  - Subjective indicators: DHM, HALL, SP

- Qualitative data included: Diary, 24/7 activity, exercise, weight, and group discussions.

- Dynamic drawings can uncover evolutionary paths of change over time.

- Heat maps show the locations of participants within the VE.

- An avatar’s position changes over the course of the study.

- The sequence of screen shots above is from an animated browser window showing how each subject’s weight, HbA1c (indicator of metabolic control), and corresponding VE activity level changed over the course of the study (6 months).

- The area of the circle represents the amount of that person’s participation during the study.

- Weekly touchpoints were represented by a circle with a unique color, so individuals can be followed over time.

- When the mouse is clicked over any circle, the visualization will identify the subject.

- The area of the circle represents the amount of that person’s participation in the sites.

- Each touchpoint represented by a circle with a unique color, as individuals can be followed over time.

- When the mouse is clicked on any circle, the visualization will identify the subject.

- The area of the circle represents the amount of that person’s participation during the study.

- Weekly touchpoints were represented by a circle with a unique color, so individuals can be followed over time.

- When the mouse is clicked on any circle, the visualization will identify the subject.

- The area of the circle represents the amount of that person’s participation during the study.

- Weekly touchpoints were represented by a circle with a unique color, so individuals can be followed over time.

- When the mouse is clicked on any circle, the visualization will identify the subject.

- The area of the circle represents the amount of that person’s participation during the study.

- Weekly touchpoints were represented by a circle with a unique color, so individuals can be followed over time.

- When the mouse is clicked on any circle, the visualization will identify the subject.