Temporal Visualization of Diabetes Mellitus via Hemoglobin A1c Levels

Eugenia McPeek Hinz[1], David Borland[2], Hina Shah[3], Vivian L. West[3], W. Ed Hammond[3]
[1]Duke Health Technology Solutions, Duke University; [2]RENCI, The University of North Carolina at Chapel Hill;
[3]Duke Center for Health Informatics, Duke University

Abstract

Diabetes mellitus is a chronic long-term disease requiring consistent medical treatment to achieve glucose control and prevent complications. Time of diabetes diagnosis can be variable and delayed years beyond disease onset. The spectrum of glycemic trajectories for a general population over an entire diabetes disease course is not well defined. Aligning disease course on death enables coherent data visualization. Our temporal visualization tool uses a parallel-sets inspired technique that illustrates the complicated and varied trajectories of hemoglobin A1c levels for a general diabetic population. A consistent glucose normalization trend for the majority of patients is seen over the course of their disease, especially in the six months prior to death. This tool permits discovery of population-level Hemoglobin A1c trends not otherwise evident without disease phase synchronization. These findings warrant further investigation and clinical correlation. Visualizations such as this could potentially be applied to other chronic diseases and spur further discoveries.

Introduction

Diabetes mellitus is a chronic disease that affects millions worldwide, resulting in numerous cardiovascular and renal complications, and subsequently is a major cause of death. Age of onset, duration of diabetes, and poor glycemic control are well-defined risk factors for the development of complications associated with increased mortality in persons with diabetes mellitus.[1] To decrease the development of complications associated with diabetes, tightly controlled glucose is the standard of care.[2] Notably some large prospective trials have found either worse outcomes or lack of benefit for some patients at high risk for complications under tight treatment control regimens.[3,4] Hemoglobin A1c (HbA1c), a marker of glucose control over the two to three months preceding the test, is a validated predictor of diabetes-related complications.[2] Using HbA1c to understand trajectories and temporal patterns of glycemic control over an entire diabetes disease course could be an important factor in improving treatment and reducing overall complications.

Data visualization techniques offer opportunities to explore large datasets and identify clinical patterns that might otherwise not be obvious. In this study we present a cohort of patients with diabetes (via ICD9 codes) from Duke University's data warehouse, visualizing their HbA1c levels over time, aligned by death, to explore trends of glycemic control. To the best of our knowledge, temporal visualization of glycemic control for a diabetic population standardized on death has not previously been presented. Our visualization groups HbA1c values into ordered categories of glycemic control (Normal, Borderline, Controlled, and Uncontrolled), utilizing a method based on parallel sets[5] and Sankey diagrams[6] to view temporal patterns in HbA1c values. We incorporate a number of features to facilitate interactive data exploration, such as viewing the progression of values either forwards or backwards in time, the ability to change the temporal sampling and range of the data being viewed, highlighting of multiple subpopulations, coloring based on the category along each path in the data or at the beginning/end of each path, and the incorporation of demographic data, such as gender.

Related Work

Analysis of diabetes indicators

A reduction in HbA1c levels lowers the risk of diabetes-related complications and mortality, especially for patients earlier in their disease course.[7] Counterintuitively, intensive treatment of glucose to reach near-normal levels for patients already experiencing diabetes-related complications has failed to lower all-cause mortality.[3] While large cross-sectional studies of populations such as the National Health and Nutrition Examination Survey find a temporal trend toward improving glycemic control over time, less well-established is the temporal trajectory of glycemic control for diabetic patients in general.[8] The only other work the authors are aware of looking specifically at glycemic control trajectories for a large diabetic cohort followed patients prospectively to the end point of death.[9] The study correlated initial glucose control to outcome of death, but did not report specifically on the population glucose trajectories.

Visualization methods

Our visualization tool is based on parallel sets[5] and Sankey diagrams.[6] Parallel sets were originally developed for visualizing relationships in multivariate categorical data, whereas Sankey diagrams, introduced by M. H. P. R. Sankey, are typically used for describing the flow of quantities such as energy, material, or cost. The original parallel sets user interface enables user-defined classification definitions, statistical analysis information, and various sorting methods. Parallel sets combines the concepts of parallel coordinates[10] and mosaic plots[11], enabling an aggregation of data points within visualization elements, as opposed to showing each individual element, which is typical of parallel coordinates. Multiple systems aggregate data points for summary.[5,12-14] For example, EventFlow enables the search and visualization of interval data, such as periods of medication treatment, to examine the order of sequences of events in the data.[12] OutFlow facilitates analyses of temporal event data in the form of pathways with relevant statistics.[14] All of these visualization tools look at event occurrences and their order, without placing these events on time axes. Our diabetes visualization uses the parallel sets paradigm, with each axis representing a temporal sample of HbA1c levels instead of a separate variable, similar to von Landesberger et al.[13] Although our current dataset is relatively small (121 patients), we chose a parallel sets representation in part due to its ability to aggregate many data points. The visual complexity is bounded by the number of axes and categories per axis, not by the number of data points, making it suitable for the exploration of larger datasets in the future. This representation also easily incorporates additional non-temporal variables, such as demographic data.

Methods

Data extraction and preprocessing

Data from Duke University's data warehouse were extracted using DEDUCE, an on-line query tool developed at Duke to assist researchers in human subjects research and departments seeking quality improvement data.[15] Beginning with over 4.4 million patients, we first queried by 23 IDC9 codes for diabetes mellitus, with and without complications. The query was further refined by querying on patient death indicator and laboratory tests for glycosylated hemoglobin (HbA1c), and finally by including only patients prescribed anti-hyperglycemics. This search returned data from 208 patients. From this cohort of 208, we eliminated four that did not have a year of death recorded, one whose date of death was documented but continued to have laboratory results recorded after that date, and 82 who did not have at least 10 years of HbA1c laboratory values. Our final cohort includes data from 121 patients.

We average HbA1c values, given as a percentage of total hemoglobin, over 6 month time intervals. In the case of missing HbA1c values within a 6 month period we first attempt to impute an HbA1c value from the average glucose (AG) values over that period of time, via the formula $HbA1c = (AG + 46.7) / 28.7$.[16] If no glucose values exist in that time period, the previous HbA1c value (measured or imputed) is carried forward. HbA1c values are then classified into four categories based on the severity of diabetes: Normal < 5.7, Borderline $[5.7, 6.5)$, Controlled $[6.5, 8)$, and Uncontrolled ≥ 8.

The sampled data is time-aligned by the death event for each patient. The visual representation of diabetes progression propagates backwards in time initially. Time is represented as number of years before death.

Visualization

Our visualization tool was developed using the D3 Javascript library.[17] The aim of this visualization is to investigate temporal trajectories of HbA1c levels for a large cohort of diabetes patients over a number of years prior to death. Since parallel sets is effective for showing relations between categories using aggregated frequencies of paths through categories at each dimension, it is a reasonable choice for showing HbA1c summary trajectories. The visualization tool shows a total of five categories: four representing glycemic control, and one optional Missing category for patients with data greater than 10 years before death. . Each vertical axis is a time step. The user can choose the frequency of these time steps, with a minimum sampling frequency of six months. The user can also select the maximum number of years before death.

The death event axis is placed at the right with all other time steps moving backwards in time to the left (Figure 1). Each vertical axis is split into the four HgA1c categories (Normal in green, Borderline in blue, Controlled in orange, and Uncontrolled in red), and a Missing category in grey prior to 10 years before death. The height of each axis category represents the proportion of the patients in that category at that point in time. Paths moving between axes recursively split moving backwards from death to show the trajectories of similar groups of patients. The visualization can show trends either starting at the death event i.e. going backwards in time (dividing

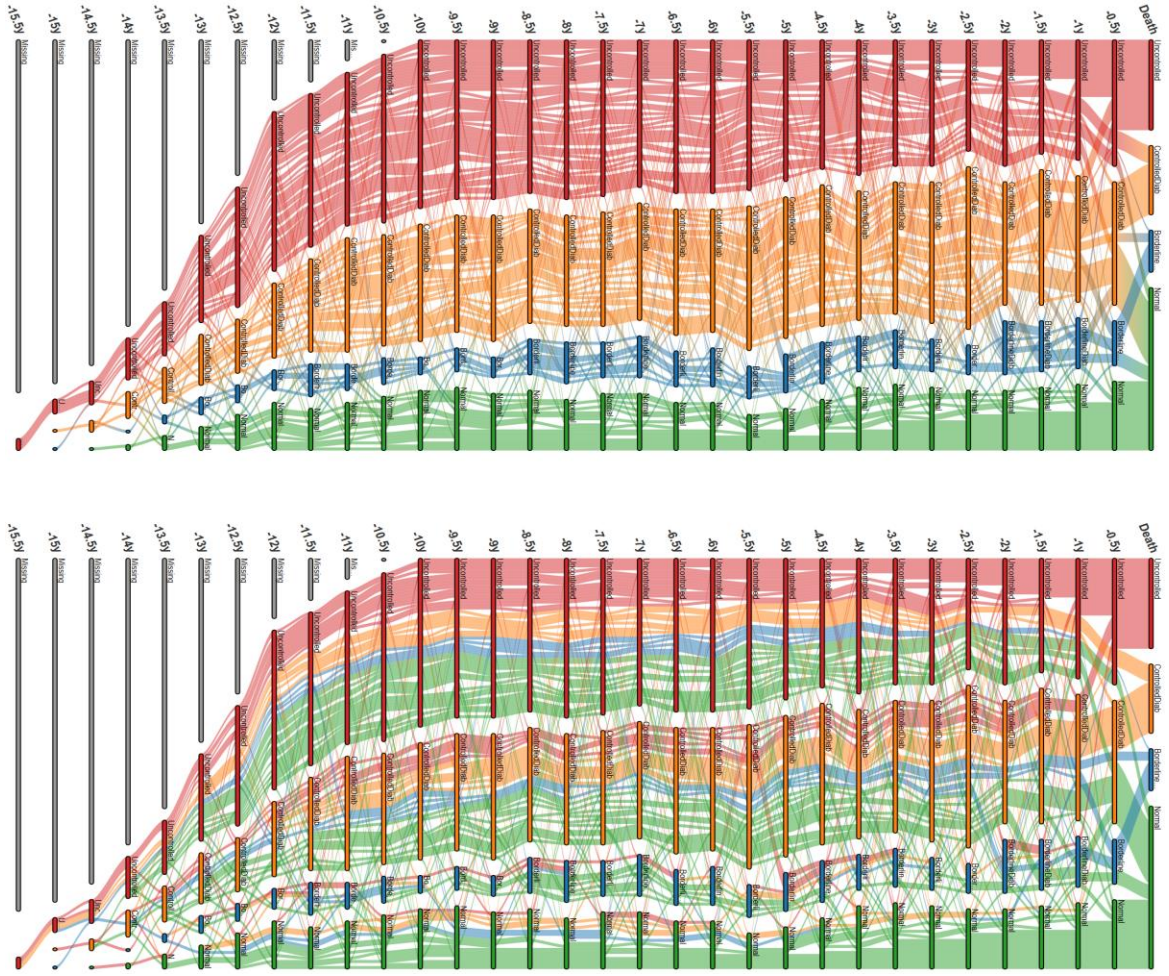

Figure 1. Diabetes progression overview visualizations. The top image colors paths by the current HbA1c at each time step, which is useful for emphasizing overall temporal trends. The bottom images colors paths by the HbA1c level at death, showing at each time step where each path will end.

recursively right to left), or starting at the last year in the visualization, i.e. going forward in time (recursive division from left to right). Going backwards and coloring by death shows at any time point the relationships between patients in a given category and their categories at death, while going forward in time shows the relationship between patients in a given category and their categories at a user-defined earlier point in time (Figure 2). Following Shneiderman's Mantra[18] of first overviewing and then filtering, the user can highlight one or more groups of patients by clicking on categories or trajectories to highlight the behavior of that group of patients going backward and forward in time, reducing visual clutter (Figure 2). A tooltip also shows the actual number of patients in each group and their percentage of the total population.

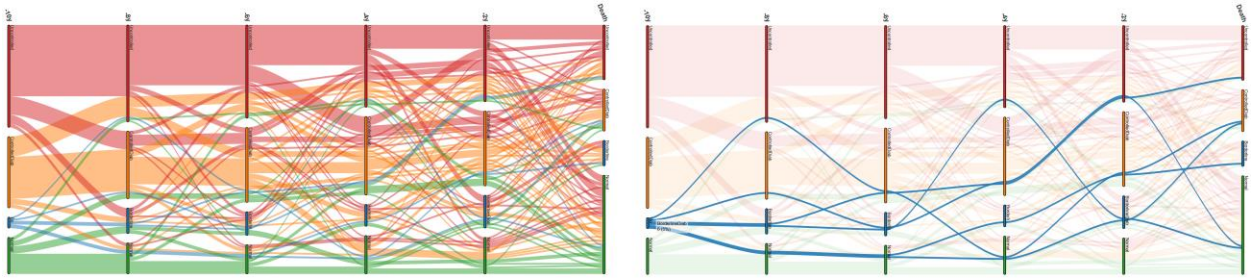

Figure 2. A 10-year range of data, sampled every two years, with forward propagation to show how the trajectories of patients change moving forward in time (left). Highlighting enables a focused view of a single category, reducing visual clutter (right).

The user can also chose between different types of coloring schemes for the paths: 1) color by the category at the first or last year (depending on the propagation direction), which shows the level of variation for a category over the length of the visualization, 2) color by transition, where the transition has a gradient from the source to target category color, which is useful for showing overall trends, and 3) color by reverse transition, where the transition path has a gradient from the target category to the source category, which is useful for category-level analysis of the distribution of source and target categories at a particular time step's category (Figure 3). To reduce visual clutter there is also an option to look at only static transitions (i.e. no change in category between time steps), and to look at only variations (i.e. only changes in the categories).

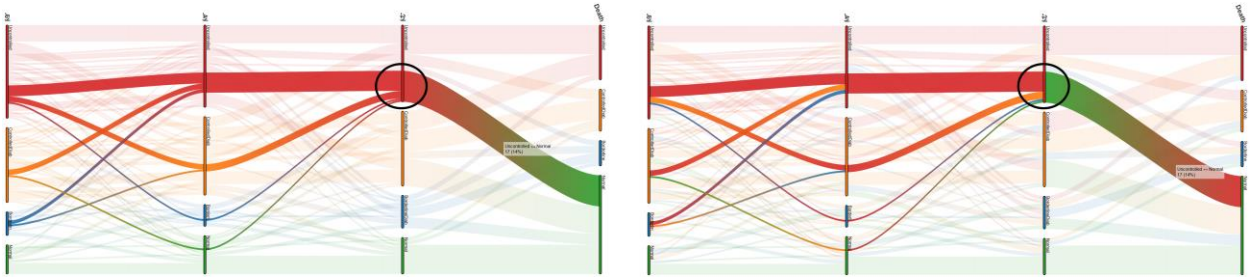

Figure 3. The user can observe separate groups by selecting individual trajectories. In addition to coloring by the starting category, paths can be colored by a gradient from source to target category (left), which redundantly encodes the category at each axis to emphasize overall trends, or by target to source category (right), which enables a rapid analysis of where paths are moving to/from at each category. The circled regions highlight this difference. On the right, it is immediately obvious what category this trajectory came from at death (Normal in green) and how this group is distributed at the previous time step.

We also include the ability to incorporate demographic data, such as gender, as additional axes (Figure 4). This feature enables the comparison of trajectories for different subpopulations based on data other than just HbA1c levels.

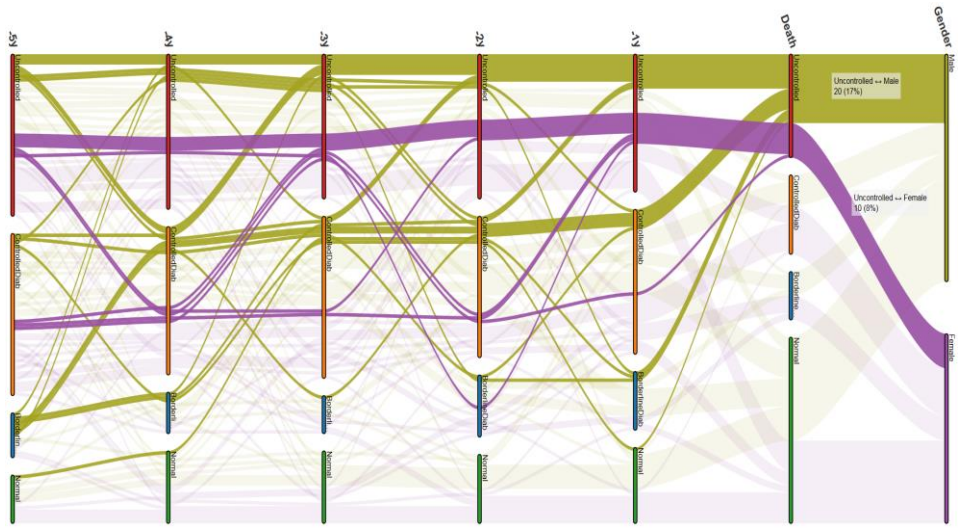

Figure 4. By adding a gender axis and selecting two groups we can compare the variability of males who were Uncontrolled at death (olive) to women who were Uncontrolled at death (purple). Men appear to have more variability over the 5-year period being visualized, as shown by the large number of transitions between different categories.

Findings

In the 10 years before death, there is a consolidating trend to improved glucose control across all diabetes control categories from uncontrolled to normal. Overall diabetes control shifts from uncontrolled diabetes for 46% of the cohort to 25% at death utilizing HbA1c and imputed glucose values. A reciprocal increase in combined borderline and normal range glucose control goes from 25% at 10 years out to 57% at death (Table 1). The trend for better glucose control is most visible in the last six months before death. The overall final glycemic trajectory is also evident in the bottom image from Figure 1 where the control category at death is colored retrospectively. Notably a

small minority of the uncontrolled sub-group remains poorly controlled over the entire disease course. By including category temporal transitions this visualization also illustrates the complexity of the underlying data, with many trajectories exhibiting a large degree of variation in HbA1c categorization over time.

Table 1. Percent of patients by Diabetes control category over 10 years prior to death using HbA1c with imputed glucose results.

Glycemic Control by HbA1c	*-10 years years to death*	*-5 years years to death*	*At Death*
Uncontrolled Diabetes	46 %	39 %	25 %
Controlled Diabetes	32 %	39 %	19 %
Borderline	5 %	11 %	12 %
Normal	17 %	12 %	45 %

Discussion

The progression of diabetes with accumulating end organ complications is well recognized. There is a clinical presumption that diabetes-related complications are also associated with worsening glycemic control for patients with end stage diabetes mellitus. Since most prospective cohort studies are organized by a patient's clinical presentation, treatment or demographics, they tend to be cross sectional studies of a population and include patients across a disease continuum. By creating a cohort organized by a death criterion with 10 or more years of diabetes lab data, we have sub-selected a general but presumably more ill diabetic population. Phasing HbA1c values by death allows data coherence that translates into the visualization of glycemic trajectories that would be less evident in cross sectional studies of diabetic patients. Understanding the course of diabetes control is important to discerning differences in outcomes, treatments and identifying sub-phenotype populations.

Death event as an organizing point for temporal data visualization permits a clear starting point to observe the course of medically treated diabetes. Cause of death is not defined, so further characterization of subpopulations visualized in the cohort, like the always uncontrolled diabetes subgroup, warrants further clinical investigation to see if they are representative of the cohort overall. All patients in this cohort had data for at least 10 years, as such our population is specific for patients under some manner of regular medical care, and interpretation of the data with respect to populations with less regular medical care should be limited. Using the imputed average glucose and average HbA1c values aligned on the cohort's endpoint enables capture of all glycemic values, including those potentially before even the diagnosis of diabetes is made.

We observed a trend to normalization of HbA1c in the last year of life. The reasons behind improved diabetes control near the end of life could include multiple factors, such as increased insulin half-life due to impaired renal and hepatic metabolism, decreased dietary intake related to anorexia or nausea, and falsely low HbA1c secondary to uremia or anemia.[19] Interestingly, the goals for end-of-life treatment in diabetic patients are generally to limit side effects of either hyper or hypoglycemia and often entail a scaling back of treatment which would be expected to be associated with more hyperglycemia not less. By using visualization tools to see the progression of HbA1c values in diabetic patients in the years before their death, our findings of glucose normalization in light of this paradigm highlight the need for further clinical investigation and interpretation.

Our data visualization tool displays temporal patterns of diabetes metric across a population and for the last years of this disease continuum. Tools such as these can only display patterns that can potentially illuminate findings that need further clinical validation and statistical investigation to determine clinical significance if any.

Future work

The visualizations we have shown here represent a small number of patients in the dataset. This has enabled us to test and refine the visualization before using large amounts of data. Next we will include diabetes-related co-morbidities, e.g. cardiovascular, neurological, and renal manifestations of prolonged diabetes illness, and additional demographic variables, e.g. age and ethnicity. We plan to link this temporal visualization to other multivariate visualizations highlighting selected groups of patients, helping to show factors related to diabetes. We are also working toward a better statistical analysis of the data, and its representation in this tool. In particular, we wish to incorporate information regarding the amount of imputed and extrapolated data in the visualization.

Conclusion

Exploring the natural disease course of diabetes control with data visualization tools permits identification of potentially clinically important trends that would be difficult to recognize otherwise. Further investigation and definition on the clinical significance of the normalization of HbA1c in the final years of life are warranted.

Acknowledgments
This work is supported by the US Army Medical Research and Materiel Command (USAMRMC) under Grant No. W81XWH-13-1-0061. The views, opinions and/or findings contained in this report are those of the authors and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation. We acknowledge the assistance from Meghana Ganapathiraju in helping to refine the visualization. Our implementation was adapted from the d3.parsets reusable chart by Jason Davies. We also acknowledge the assistance of Mark Massing MD PhD MPH and Susan Spratt MD for comments and discussions on graphical representations of diabetes control.

References

1. The Diabetes Control and Complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. N Engl J Med.1993;329(14):977-986.
2. American Diabetes Association. Standards of Medical Care in Diabetes. Diabetes Care. 2009;32(S1):S13-S61.
3. The Accord Study Group. Long-term effects of intensive glucose lowering on cardiovascular outcomes. N Engl J Med. 2011;364(9):818-828.
4. The UK Prospective Diabetes Study Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). Lancet. 1998;352(9131):837-853.
5. Bendix F, Kosara R, Hauser H. Parallel sets: visual analysis of categorical data. IEEE Symp on Info Vis. 2006;12(4):133-140.
6. Riehmann P, Hanfler M, Froehlich B. Interactive Sankey diagrams. IEEE Symp on Info Vis. 2005;233-240.
7. Holman R, Paul S, Bethel M, Matthews D, Neil H. 10-year follow-up of intensive glucose control in type 2 diabetes. N Engl J Med. 2008;359(15):1577-1589.
8. Ford E, Li C, Little R, Mokdad A. Trends in A1C concentrations among U.S. adults with diagnosed diabetes from 1999 to 2004. Diabetes Care. 2008;31(1):102-104.
9. Gebregziabher M, Egede LE, Lynch CP, Echols C, Zhao Y. Effect of trajectories of glycemic control on mortality in type 2 diabetes: a semiparametric joint modeling approach. Am J Epidemiol. 2010;171(10):1090-1098.
10. Inselberg A, Dimsdale B. Parallel coordinates. Human-Machine Interactive Systems. 1991;199-233.
11. Hoffman H. Exploring categorical data: Interactive mosaic plots. Metrika. 2000;51(1):11-26.
12. Monroe M, Wongsuphasawat K, Plaisant C, Shneiderman B, Millstein J, Gold S. Exploring point and interval event patterns: Display methods and interactive visual query. HCIL Tech Report, University of Maryland. 2012.
13. von Landesberger T, Bremm S, Andrienko N, Andrienko G, Tekusova M. Visual analytics methods for categoric spatio-temporal data. IEEE Conf on Vis Anal Sci and Tech(VAST) 2012;183(192):14-19.
14. Wongsuphasawat K, Gotz D. Exploring flow, factors, and outcomes of temporal event sequences with the outflow visualization. IEEE TransVis Comput Graph, 2012;18(12):2659-2668.
15. Horvath MM, Winfield S, Evans S, Slopek S, Shang H, Ferranti J. The DEDUCE guided query tool: Providing simplified access to clinical data for research and quality improvement. J Biomed Info. 2011;2:266-276.
16. Nathan DM, Kuenen J, Borg R, Zheng H, Schoenfeld D, Heine RJ. Translating the A1C Assay into estimated average glucose values. Diabetes Care. 2008;31(8):1473-1478.
17. Bostock M, Ogievetsky V, Heer J. D^3: Data-driven documents. IEEE Trans Visualization & Comp Graphics. 2011;17(2):2301-2309.
18. Shneiderman B. The eyes have it: A task by data type taxonomy for information visualization. Proc. 1996 IEEE Symp Vis Lang. 1996;336-343.
19. Kalantar-Zadeh K, Derose SF, Nicholas S, Benner D, Sharma K, Kovesdy CP. Burnt-out diabetes: Impact of chronic kidney disease progression on the natural course of diabetes mellitus. J Renal Nutrition. 2009:19(1):33-37.