Winnow: Interactive Visualization of Temporal Changes in Multidimensional Clinical Data

Hsueh-Chien Cheng
Department of Computer Science, University of Maryland
College Park, Maryland 20742
hccheng@umd.edu

Rainer von Coelln
Department of Neurology, University of Maryland School of Medicine
Baltimore, Maryland 21201
rvoncoelln@som.umaryland.edu

Ann L. Gruber-Baldini
Department of Epidemiology and Public Health, University of Maryland School of Medicine
Baltimore, Maryland 21201
abaldin@som.umaryland.edu

Lisa M. Shulman
Department of Neurology, University of Maryland School of Medicine
Baltimore, Maryland 21201
lshulman@som.umaryland.edu

Amitabh Varshney
Department of Computer Science and UMIACS, University of Maryland
College Park, Maryland 20742
varshney@umiacs.umd.edu

ABSTRACT

Over the last two decades, an emerging paradigm of Parkinson's disease (PD) research that considers diverse clinical and biological markers in evaluating disease severity has emerged. This can lead to important insights including patterns of disease subtypes and progression. However, identification of disease subtypes and progression across multidimensional longitudinal variables representing diverse biomarkers, outcomes, and demographics remains a challenge. In this paper we present our interactive visual analytics tool, *Winnow*™, to help clinicians understand and interpret inter-relationships between a broad range of outcome measures and patient subgroups. Based on a close collaboration between clinical experts and computer scientists, Winnow is designed to intuitively visualize data from multiple outcome measures to enable users to easily filter and compare patient subgroups. We summarize the challenges in multidimensional clinical data analysis, the visual representations and design of Winnow, and present case studies based on a public PD dataset.

KEYWORDS

Interactive visualization; parallel coordinates; multidimensional visualization; web-based application; clinical visual analytics

1 INTRODUCTION

Parkinson's disease (PD) is a chronic neuro-degenerative disorder characterized by gradual progression. Whereas motor impairments are the most recognized symptoms of PD, there is increasing recognition of the importance of a range of non-motor symptoms including cognitive decline, sleep disturbance, and depression [20]. For example, the presence of inter-relationships between clinical subgroups and disease progression is supported by the discovery that patients with greater postural instability and gait difficulty tend to have faster rates of progression and greater cognitive decline [3]. This paradigm shift has resulted in large datasets with diverse clinical and biologic markers [10]. For example, at the University of Maryland PD Center, we have collected 15 years of heterogeneous and multidimensional data (e.g. clinical and genomics) on 2, 500 PD patients across 20, 000 office visits.

The expanded scope of disease beyond traditional motor symptoms raises the complexity of analysis significantly. These large multidimensional longitudinal datasets need new tools to identify disease subtypes and patterns of disease progression across diverse biomarkers, outcomes, and demographic subgroups. Conducting longitudinal studies of multidimensional data is challenging because of the limited human capacity to comprehend numerous interactions among a large number of variables. The immense cognitive load of such complex tasks causes these analyses to be time-consuming and ineffective. To address this challenge, many computational techniques such as data clustering and dimensionality reduction have been developed [1, 2]. Clustering techniques group similar data points (patients) into clusters based on specific criteria defined in a high-dimensional space. Dimension reduction techniques transform high-dimensional data points into low-dimensional representations while preserving crucial information required to proceed with the intended analysis.

Although these computational techniques provide a data-driven, objective perspective generated mathematically and statistically, they do not generate clinically meaningful results since clinical insight and experience with the domains is disconnected from the analysis. In addition, in some cases the stability of these results is questionable. For example, PD subtypes generated by cluster analysis were found to be inconsistent in various studies [22]. The results generated by pure data-driven approaches can therefore be misleading without proper clinical interpretation. We had a similar experience when we applied computational techniques in our preliminary studies, where the results were inaccessible to clinicians and not clinically relevant. This observation led us to believe that the most useful tools should be intuitive, interactive,

ACM-BCB '17, August 20-23, 2017, Boston, MA, USA

ACM ISBN 978-1-4503-4722-8/17/08…$15.00
https://doi.org/10.1145/3107411.3107449

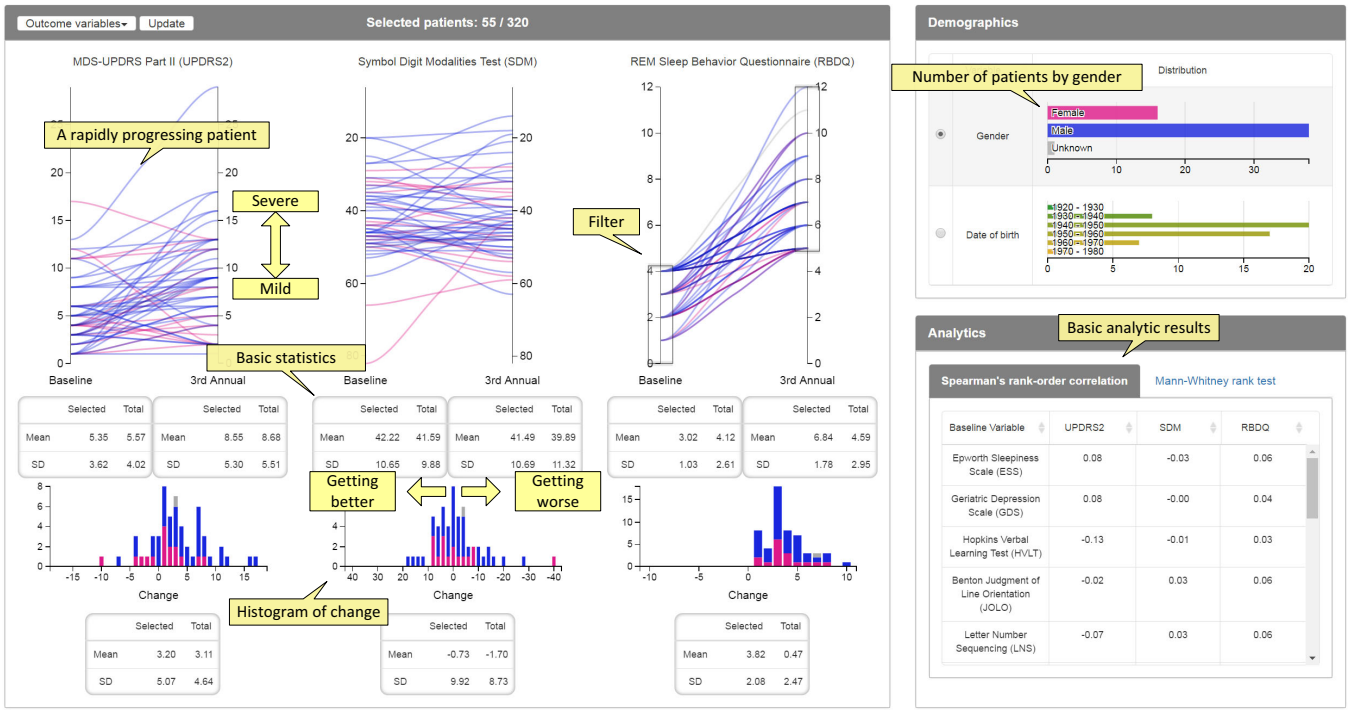


Figure 1: This screenshot of Winnow shows the results based on the Parkinson's Progression Markers Initiative (PPMI) dataset (Section 4). Winnow consists of three panels: the Outcomes panel (left), the Demographics panel (top right), and the analytics panel (bottom right). Here 55 patients (out of 320) with low baseline REM sleep behavior disorder (RBDQ < 5) and high RBDQ (RBDQ ≥ 5) at the third annual visit are selected. The colors represent the gender of patients (magenta for females and blue for males).

and generate results interpretable by clinicians. This tool should allow clinicians to efficiently sift through arrays of longitudinal data, identify patterns, and generate hypotheses, which can then be subjected to rigorous statistical analysis.

In this paper, we present a preliminary analysis of a public PD dataset with our tool Winnow (Figure 1) to study patterns of disease progression between time points. As a visual analytics tool, Winnow presents complex data in a way that facilitates detection of patterns and anomalies by the human eye. In contrast to the fully-automatic analysis conducted with computational techniques, visual analytics tools allow users to interactively explore the data through dynamic filtering and subgrouping. By interacting with intuitive visual representations, users can generate clinically relevant hypotheses and insights that are not easily accessible with conventional approaches.

Winnow visualizes multiple selected outcome measures simultaneously to enable the investigation of inter-relationships across outcome measures in various domains. By representing each patient as a line defined by the values of a selected outcome measure recorded at the first and second time points, the slope of lines indicates the rate of disease progression of the corresponding outcome. We also show a histogram of disease progression to facilitate the selection of fast- and slow-progressing patients. All visualizations are color-coded by demographic characteristics of patients (e.g. gender) to foster understanding of demographic-related questions such as whether female and male patients progress differently with respect to specific outcomes. The panels in Winnow are linked together to provide consistent visualizations during data exploration.

We have made our design choices in Winnow based on the feedback obtained from data scientists and clinicians in a long-term collaboration following the guidelines of Multi-dimensional In-depth Long-term Case studies (MILC) [19]. One example of the design choices we collaboratively made is reducing the complexity of visualization by choosing clinically relevant features, thus improving the usability of the tool. Clinical investigators want the opportunity for "hands-on" experience with clinical datasets to explore data based on experiential intuition. We repeatedly found the interpretability of results to be the top priority for the clinicians.

We make the following contributions in this work:

- A visual analytics tool, Winnow, to help clinicians understand the interactions across multiple disease features and their changes over time.
- A MILC analyzing the needs of clinicians to facilitate the design of a clinician-friendly tool for visualizing multidimensional clinical data.
- A preliminary analysis of a public PD dataset based on 16 outcome measures in multiple domains.

In the following, we first review the work on computational and visualization-based techniques in Section 2. We then summarize our MILC in Section 3. We introduce the dataset used in our study

in Section 4, the visual analytics tool Winnow in Section 5, and the example case studies conducted with Winnow in Section 6. Finally, we present conclusions and future extensions of Winnow in Section 7.

2 RELATED WORK

2.1 Data Clustering

Clustering techniques reveal patterns by grouping similar data points together. These techniques create another layer of abstraction on top of the raw high-dimensional data, thus allowing users to inspect the clustering results without directly addressing the high dimensionality of the data. Because the grouping of data points is based on an explicitly defined similarity function, defining a reasonable function is key to the success of clustering analysis. Typical k-means algorithm [11] defines dissimilarity of two high-dimensional points as the Euclidean distance between them. Density-based techniques such as OPTICS [1] rely on a neighborhood function to determine whether two data points are (similar) neighbors. Spectral clustering [18] requires a similarity function in constructing the adjacency graph. In practice, different similarity functions may lead to distinct results.

Nevertheless, the similarity of two data points is subjective, application-dependent, and usually difficult to define especially for high-dimensional data. This difficulty is fundamental because typical distance functions (e.g. Euclidean distance) fail to measure similarity precisely in high-dimensional space, a known problem referred to as the "curse of dimensionality". In an exploratory clinical analysis that lacks a well-defined disease model, the limited interpretability of clustering results also aggravates the generation of clinically relevant, actionable results.

2.2 Dimension Reduction

Dimension reduction techniques reduce data dimensionality while maintaining certain relationships among data points with respect to specific criteria. Principal component analysis (PCA) finds a series of mutually orthogonal principal components that account for the most variance in the data. Locally linear embedding (LLE) [17] and Laplacian Eigenmaps [2] first construct a neighborhood representation, typically based on pairwise Euclidean distances, and then derive a low-dimensional space that preserves local distances among neighboring data points. A neural network that learns to reconstruct the original high-dimensional training data from the derived low-dimensional latent variables can also be used for reducing data dimensionality [7].

All these techniques share similar limitations with clustering techniques such as requiring an explicitly-defined similarity function and limited interpretability of results. Whereas each principal component generated by PCA represents a linear combination of variables, the clinical meaning of adding (or subtracting) two clinically-unrelated features from different domains is unclear. The results generated by LLE and Laplacian Eigenmaps provide little clue for further interpretation. In fact, even extending a low-dimensional space derived by LLE or Laplacian Eigenmaps to include a new data point is non-trivial. The latent variables found by neural-network-based techniques are usually cryptic unless the training data are already labeled, which is not the case for exploratory analysis.

2.3 High-dimensional Data Visualization

High-dimensional data visualization creates a visual abstraction of high-dimensional data such that patterns and anomalies can be detected by the human eye. Several successful visualizations of high-dimensional data such as the scatter plot matrix (SPLOM) [4] and parallel coordinates [8] have been widely applied. A SPLOM organizes scatter plots generated for each dimension pair as the elements inside a matrix; this layout facilitates efficient scanning of correlations between dimension pairs. A parallel coordinates visualization represents dimensions as individual parallel lines (axes); a data point is therefore represented as a segmented line connecting each point in order from the first axis to the last. Clusters of data points can therefore be visually detected as clusters of lines. Spreadsheet-based approaches that associate groups of cells with various types of visual representations can also help understand complex inter-relationships in heterogeneous datasets [23]. In contrast to the automatic analysis of computational approaches, data visualization can be easily combined with user interactions (e.g. brushing and linking) to allow interactive data exploration via filtering and zooming in and out of specific subsets of data.

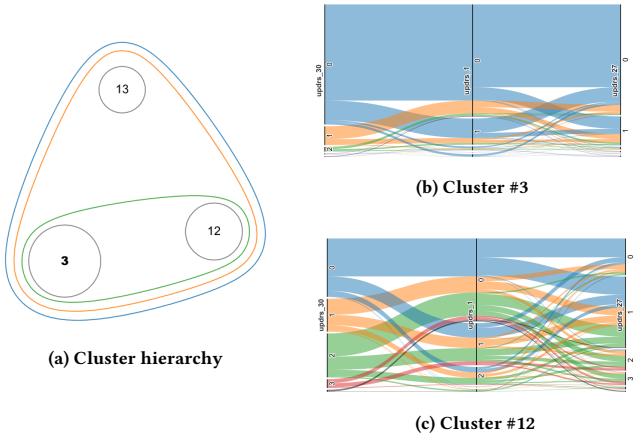
The amount of information that can be displayed in a single visualization is, however, limited by the screen size and the perceptual capability of the human visual system. Such limitations can cause usability issues that adversely affect the efficacy of visual analytics tools. For example, visualizing all pairs of dimensions in a SPLOM is increasingly unrealistic as dimensionality grows. The parallel-coordinates visualization is incomprehensible even with a moderate number of dimensions. Better visual design strategies such as reordering dimensions [16], subsampling [6], and edge bundling [25] improve visual quality, but in general visualizing high-dimensional data remains a challenging task.

3 PRELIMINARY STUDY: MILC

The Multi-dimensional In-depth Long-term Case studies (MILC) method [19] evaluates the efficacy of visualization tools using various approaches, such as interviews and observations, while collaborating closely with expert users. Here we describe the MILC we conducted as a preliminary study in the development of a clinician-friendly visual analytics tool.

We formed a multi-disciplinary research group of data scientists, database specialists, neurologists, and biostatisticians in 2015, targeting the analysis of the PD data collected at the University of Maryland PD Center. Group members included movement disorder specialists (neurologists) with extensive expertise in PD but little experience with visualization tools. In contrast, other group members with computational backgrounds were only familiar with computational and visualization techniques.

During each group session, a visualization tool or a new version of a tool based on previous group discussion was presented. Users then attempted to apply the tool for routine analysis tasks or confirming/rejecting PD-related hypotheses. The comments and feedback from users were recorded by the observers while they interacted with the tool.



(a) Cluster hierarchy

(b) Cluster #3

(c) Cluster #12

Figure 2: (a) Our early attempt created a hierarchy of clusters to allow splitting and merging of clusters following a pre-generated hierarchy. Nevertheless, the parallel set visualization of two sibling clusters: cluster #3 (figure (b)) and cluster #12 (figure (c)), are difficult to interpret even with three dimensions.

Early in the project, we focused on identifying novel multi-domain PD subtypes to extend the recognized motor-based subtypes [9]. Our first attempt (April, 2015) was an interactive hierarchical clustering method based on OPTICS [1]. We implemented an interactive tool that enables users to explore alternative clustering results in a pre-generated hierarchy of clusters (Figure 2a). We then tried a series of dimension reduction techniques including PCA, LLE [17], Laplacian Eigenmaps [2], and t-Distributed Stochastic Neighbor Embedding [21] to reveal possible low-dimensional representations that lead to visually apparent clusters of patients.

Common issues in all the above attempts were the limited interpretability of the results and the lack of flexibility for clinicians to steer and manipulate the analysis as desired. The goals from the clinician's perspective were to 1) visualize data on patient signs and symptoms in ways that align with their experience and 2) enable a simple hands-on interrogation of data to pose questions and generate hypotheses. We tried to improve user-friendliness by visualizing the constituent patient clusters but without success because of the complexity introduced by high data dimensionality. For example, one of our early attempts created parallel sets visualizations for each cluster. The visualizations containing overlapping lines are, however, difficult to read even with only three dimensions (Figure 2b and 2c). The lack of flexibility is related to the limited capacity for user interventions in these techniques–the only way users can modify the results is through reconfiguring the parameters of the applied computational techniques. Furthermore, these two issues of interpretability and flexibility are related and may exacerbate each other. For example, one would not know how to modify the parameters of OPTICS without a proper interpretation of the results.

Acknowledging the obstacles in our group's process, we developed two key strategies: 1) switching to a simpler PD dataset and 2) organizing two half-day group retreats (July and August, 2016). We chose to use the Parkinson's Progression Markers Initiative (PPMI) dataset (Section 4) because it is a smaller, simpler, and more structured dataset with a well-defined protocol, thus allowing us to find patterns more easily with less interference from noise. Unlike the University of Maryland PD dataset, PPMI patients are more homogeneous (recently diagnosed and untreated at enrollment) and are assessed at more standardized intervals (every six months).

A consultant with expertise in applied biostatistics and modeling (Dr. Søren Bentzen) was invited to attend the retreats for a fresh perspective and to recharge the group dynamic. After a lively discussion, the major result of the first retreat was a preliminary sketch of Winnow, a clear breakthrough to achieve our goals. As compared to the previously developed methods, Winnow's data visualization is clinically intuitive and invites the hands-on experience that clinicians seek. In addition to Winnow's approach, we also discussed the following computational techniques for future extensions.

- Clustering analysis that groups patients by their temporal characteristics (e.g. multiple cross-sectional clustering or clustering the change in clinical features).
- Supervised learning methods that predict a user-defined rate of disease progression based on multiple outcome measures.
- Statistical methods, for example multivariate regression analysis, that prove or disprove hypotheses generated by users.

These techniques provide complementary strength to the visual analytics approach of Winnow. We also reached consensus to add a biomarkers panel in the future to include data from genomics, imaging, serology, and cerebrospinal fluid. Group members had mixed reactions on the importance of interactivity to a successful tool; although most members ranked interactivity high on the list of key components, some prioritized a robust data model or user-friendly interfaces over interactivity.

Based on the results of the first retreat, we designed an interactive visualization tool with the features found to be useful for clinicians. The second retreat was held in August, 2016 to present the first version of Winnow and to review its strengths and weaknesses. The group was unanimous in their positive assessment of the intuitiveness of the data visualization and the capacity for simple interaction with the data by users without intensive training (e.g. selecting patient subgroups and outcomes for analysis). We later developed new features proposed in the second retreat, including statistical tests for group comparisons. We describe the design of Winnow in Section 5.

4 DATA

We present the following examples and case studies using data obtained from the Parkinson's Progression Markers Initiative (PPMI) database [14] (www.ppmi-info.org/data). The PPMI dataset contains patient and clinician-reported outcome measures as well as genetics, imaging and serologic data collected over a five-year period since 2010. As of now, more than 400 PD patients have been enrolled in the study. The PPMI is an ongoing study with patient records being updated on a rolling basis. In this paper, we use the data downloaded on March 2nd, 2017. For up-to-date information on the PPMI study, visit www.ppmi-info.org.

Table 1: List of the 16 outcome measures and six domains of PD in our study.

Domain	Outcome Measure	Description	Range
PD Severity: Motor & Non-Motor	MDS-UPDRS Part I (UPDRS1)	Non-motor aspects of experiences of daily living	0–52
	MDS-UPDRS Part II (UPDRS2)	Motor aspects of experiences of daily living (disability)	0–52
	MDS-UPDRS Part III (UPDRS3)	Motor examination for signs of PD	0–132
	MDS-UPDRS Total Score (T-UPDRS)	General PD severity: sum of MDS-UPDRS Part I-III	0–236
Autonomic	Autonomic Scale for Outcomes in PD (SCOPA-AUT)	Autonomic symptoms in PD	0–75
Cognitive	Hopkins Verbal Learning Test (HVLT)	Verbal short-term memory and new learning	0–12
	Benton Judgment of Line Orientation (JOLO)	Visuospatial perception and orientation	0–30*
	Semantic Fluency (SFT)	Semantic and phonetic memory	> 0
	Letter Number Sequencing (LNS)	Attention, working memory, and visuospatial ability	0–21
	Symbol Digit Modalities Test (SDM)	Processing speed	0–110*
	Montreal Cognitive Assessment (MoCA)	Global cognitive evaluation	0–30*
Sleep	Epworth Sleepiness Scale (ESS)	Daytime sleepiness	0–24
	Rapid Eye Movement Sleep (REM) Behavior Disorder Questionnaire (RBDQ)	Abnormal behaviors during REM sleep	0–13
Behavior	Geriatric Depression Scale (GDS)	Depressive symptoms in the elderly	0–15
	State-Trait Anxiety Inventory (STAI)	State and traits of anxiety in adults	40–160
Disability	Modified Schwab & England Activities of Daily Living (SEADL)	Independent and dependent function of daily activities	0–100*

* Lower score indicates more severe symptom. For all others, higher score indicates more severe symptoms.

We selected the 16 outcome measures listed in Table 1 that were collected when patients were enrolled (BL) and at their third annual follow-up visit (Y3) to study disease progression. The MDS-UPDRS (Movement Disorders Society-Unified PD Rating Scale) measures PD severity and the other measures are selected from the autonomic, cognitive, sleep, behavior, and disability domains. We chose the third instead of the fifth annual visit to generate a representative set of data with a sufficient number of patients. We excluded patients with missing or incomplete data (e.g. enrolled for less than three years) from our analysis. In summary, a total of 320 patients (90 female, 211 male and 19 unknown) born between 1927 and 1979 were included. Each patient is represented by 34 features (16 outcome measures collected at BL and Y3, and two demographic attributes). The outcome measures are aggregated scores calculated from sets of items from validated questionnaires.

5 WINNOW

Figure 1 shows a screenshot of Winnow consisting of three panels: the outcomes panel (left), the demographics panel (top right), and the analytics panel (bottom right). The number near the top shows the total number of selected patients (55 in Figure 1). We also provide the option for descriptions of the selected ranges of the currently applied filters in a tooltip (not shown here).

5.1 Outcomes Panel

Each selected outcome measure corresponds to two plots in the outcomes panel: one showing the values at BL and Y3 (top) for each individual patient, and the other showing the number of patients with a certain amount of change between BL and Y3 as a stacked histogram (bottom). In the top plot, each patient is represented by a line connecting the values of that outcome measure at BL and Y3 for that patient. Using this visual representation, changes in values correspond to the slope of lines, whose differences can be detected efficiently by human eye [13]. We reverse the positive direction of axes in the plots, if needed, such that a positive slope always indicates PD progression (from mild to severe symptoms). Therefore, a line with a steep upward slope shows the corresponding patient is experiencing rapid progression on that outcome measure. This axis reversal is applied to Benton Judgment of Line Orientation (JOLO), Symbol Digit Modalities Test (SDM), Montreal Cognitive Assessment (MoCA), and Modified Schwab & England Activities of Daily Living (SEADL) where higher scores indicate milder symptoms (marked with an asterisk in Table 1).

Although the slope of lines can be easily seen for a moderate number of lines (patients), the difficulty in locating individual lines increases significantly with the degree of occlusion. We therefore also show the number of patients with a certain amount of change between BL and Y3 in a stacked histogram in the bottom plot to provide a summary of disease progression. A similar axis reversal is applied to the histograms such that a bar on the right represents

more rapid progression. In both the top and bottom plots, the colors of lines and bars are determined by the selected demographic attribute in the demographics panel (Section 5.2).

Users can select a subset of patients by dragging the target interval on the target axis. During selection, all plots, including the ones in the demographics panel, are updated interactively to reflect the latest selected patients through brushing and linking. We show the mean and standard deviation of the values at BL, Y3, and the changes in between the two for the total sample and for the selected patients to allow for easy comparison.

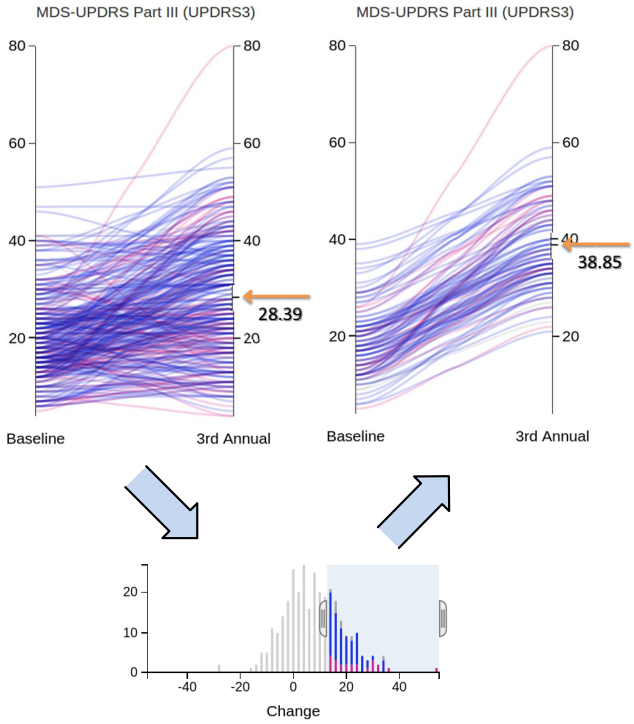


Figure 3: After using the bottom histogram for selecting patients with the most rapid progression in the motor domain (Δ(UPDRS3) > 12), the corresponding plot of UPDRS3 in the outcomes panel shows only the 109 matching patients, represented by the lines with steep upward slopes. The mean UPDRS3 at Y3 of the group of selected patients (right arrow) is greater than that of the total patient sample (left arrow).

For example, users can select the rapidly progressing patients whose UPDRS3 (motor exam) score increased by more than 12 over three years in the bottom histogram (bottom of Figure 3). After applying the filter, the 109 selected patients are represented by lines with steep upward slopes in the top plot (right of Figure 3). The mean UPDRS3 at Y3 for the selected patients (38.85, right arrow in Figure 3), is higher than the mean UPDRS3 at Y3 for the total patient sample (28.39, left arrow in Figure 3).

5.2 Demographics Panel

The demographics panel shows the gender and the year of birth in two individual histograms (top right of Figure 1). Instead of showing every year as a separate bar in the histogram, which would result in a crowded visual display, the years are grouped into decades to facilitate interpretation of age distribution and efficient use of the limited screen space.

The color scheme used in Winnow is determined by the selected demographic attribute. For example, when gender is selected, the lines (top plot) and bars (bottom plot) in the outcomes panel are colored in magenta and blue for female and male patients, respectively. Users select a subset of patients with respect to a particular demographic attribute by clicking the corresponding bar.

For example, users can select the 24 female patients from the previously selected 109 rapidly progressing patients using the demographics panel (left of Figure 4). After applying the gender filter, the mean value of Δ(SDM) is −3.33 (comparable to the decline of −3.36 before filtering by gender); this shows that the female patients in the selected cohort progress similarly to the mean total sample in terms of cognitive function as measured by SDM.

Users can also identify female patients with the most rapid progression in SDM. For example, they can select the four patients with Δ(SDM) < −16.62, which is more than one standard deviation from the mean, represented by lines with significantly steeper slopes when compared with other female patients (center of Figure 4). If we select those four patients and switch from gender to year of birth in the demographics panel, we see that these four patients were born between 1940 and 1960, corresponding to ages between 56 to 76 years (right of Figure 4).

5.3 Analytics Panel

The analytics panel shows the relationships between pairs of variables through statistical analysis (bottom right of Figure 1). The first tab shows the correlation of a pair of outcomes evaluated by the Spearman's rank correlation; a high correlation coefficient for a pair of outcomes indicates that patients with severe symptoms in one outcome are likely to have severe symptoms in the other outcome; similarly, those with low on one outcome are likely to be low on the other. The second tab shows the p-values of the Mann-Whitney test comparing the distributions of outcomes in the selected patients and the remainder of the patients in the total sample; an outcome has a low p-value when its distributions are different in the selected patients and the remainder of the patients. Both the Spearman's rank correlation and the Mann-Whitney test are non-parametric. The variable tested can be selected from the values at BL, Y3, or the change in values from BL to Y3.

In the following example we use the Mann-Whitney test to compare the changes in values between two groups: The 109 fast-progressing patients with respect to UPDRS3 (selected in Figure 3) and the remainder of the patients ($n = 211$). The result shows, beside the trivial case comparing Δ(UPDRS3) in the two groups, the other two UPDRS sub-scales (Δ(UPDRS1) and Δ(UPDRS2)–non-motor symptoms and disability) have low p-values ($p < 0.05$). Other variables with low p-values in increasing order are Δ(GDS), Δ(SDM), Δ(SCOPA-AUT), Δ(SEADL), Δ(STAI), and Δ(LNS), associated with a range of domains (behavior, cognitive, autonomic, and disability). In summary, disease progression in motor functions is associated with progression of autonomic dysfunction, cognitive decline, depression, anxiety, and disability.

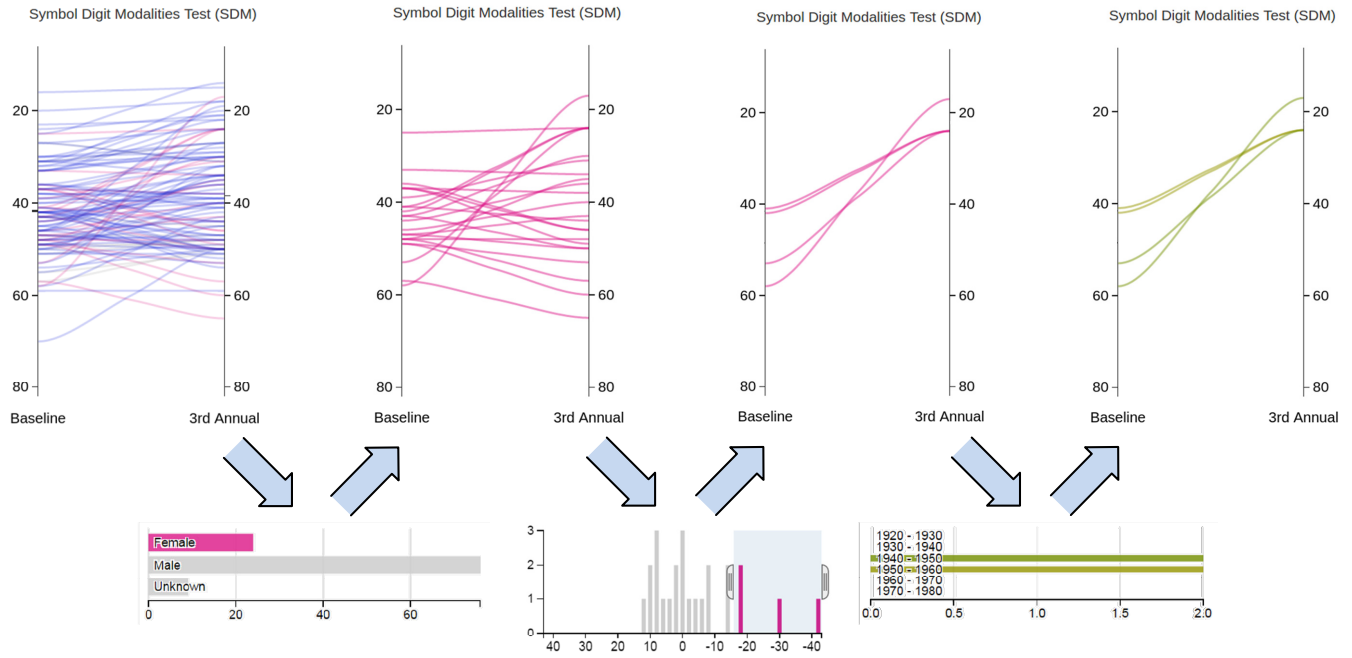


Figure 4: Users can filter the 109 previously selected patients (in Figure 3) by gender by clicking the corresponding bar in the demographics panel. After selecting female patients (left) with unusually rapid changes in cognitive function as measured by SDM (center), users can then change the color scheme to explore the year of birth of the selected patients (right).

5.4 Implementation

Winnow is implemented in JavaScript (frontend) and Python (backend). Users can conveniently run Winnow on modern browsers without installation through a link to the website[1].

6 CASE STUDIES

We now summarize two case studies we conducted with Winnow. The results shown in this section are for demonstration only. Analyses with Winnow are intended to uncover promising relationships between a range of outcome measures for selected patient subgroups. These analyses are for the purpose of generating hypotheses and need to be reproduced in future studies.

6.1 Questions

We have identified two clinically relevant questions about disease progression in PD with input from our clinical experts.

Q1 Does gender affect Parkinson's disease severity at year three?
Q2 Does the baseline severity of REM sleep behavior disorder (RBDQ) affect year three outcomes?

In the following we assess the effect of grouping using the Cohen's d [5], which is the difference between two sample means divided by the pooled standard deviation:

$$d = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}}, \quad (1)$$

where x_i and σ_i are the mean and standard deviation of the i-th group. A Cohen's d of value 0.2, 0.5, or 0.8 suggests a small, medium, or large magnitude of relationship, respectively [5]. We use Cohen's d to provide a different but complementary perspective on group differences in addition to the rank-based Mann-Whitney test.

6.1.1 *Q1: Does gender affect Parkinson's disease severity at year three?*

Gender differences in PD-related symptoms have been studied in the past. For example, RBD (REM sleep behavior disorder) was shown to be more prevalent in male patients than female patients [24]. A thorough review of gender differences in cognitive functions can be found in [12]. In the following analysis, we first applied a filter in the demographics panel and then used the plots in the outcomes panel to investigate various outcome measures at Y3 (Figure 5).

For UPDRS3 in the motor domain, female patients have a mean value of 25.89 at Y3 (first plot in the top of Figure 5) and males have a mean value of 29.36 at Y3 (first plot in the bottom of Figure 5). Therefore, comparing female with male patients, females have less severe motor symptoms than males at Y3 ($d = 0.27$). The Mann-Whitney test shows a significant difference ($p = 0.003$, shown in Table 2) between the female and male patients. In contrast, female patients have more severe symptoms on semantic fluency (SFT in the cognitive domain) at Y3. Females have mean SFT of 53.13 whereas males have mean SFT of 45.98 at Y3. The difference in mean SFT at Y3 for females and males shows a medium to large magnitude ($d = 0.59$ and $d = 0.63$). The Mann-Whitney test confirms the significance of gender difference in SFT at Y3. Other measures that significantly differ by gender at Y3 are listed in Table 2, including measures in the

[1] http://hccheng.pythonanywhere.com/vis/

cognitive, sleep, and disability domains. Male patients have more severe symptoms than females in seven out of the nine outcome measures (marked in bold in Table 2). These seven outcomes are UPDRS2, UPDRS3, and T-UPDRS (general PD severity), MoCA and SDM (cognition), ESS (sleep), and SEADL (disability).

Table 2: Outcome measures that are significantly different ($p < 0.05$) by gender at year three (Y3)

Domain	Outcome Measure	Mean value at Y3		p-value
		Female	male	
PD Severity: Motor & Non-Motor	UPDRS2	6.98	**9.31**	< 0.001
	T-UPDRS	41.07	**46.71**	0.001
	UPDRS3	25.89	**29.36**	0.003
Cognitive	SFT	**53.13**	45.98	< 0.001
	JOLO*	**11.80**	12.98	< 0.001
	MoCA*	27.08	**25.94**	< 0.001
	SDM*	42.48	**38.19**	0.006
Sleep	ESS	6.49	**7.49**	0.013
Disability	SEADL*	89.06	**87.49**	0.021

* Lower score indicates more severe symptom.

6.1.2 Q2: Does the baseline severity of REM sleep behavior disorder (RBDQ) affect year three outcomes?

In this question, we investigate how the baseline severity of REM sleep behavior disorder (measured by RBDQ) affects the severity of outcomes in other domains over three years. We approached this question by selecting patients with increasingly severe REM sleep behavior disorder (greater values of RBDQ) at BL and comparing the severity of the other outcomes at Y3.

After applying a filter on baseline RBDQ with a cut-off value of four (Figure 6a), the mean values of T-UPDRS (PD severity) at Y3 is 50.17 for the top 50% of patients for RBDQ at BL (left of Figure 6b), and 39.96 for the bottom 50% of patients for RBDQ at BL (not shown here). These results show that for patients with greater baseline REM sleep behavior disorder, their general PD severity (T-UPDRS) is more severe at the third annual follow-up visit ($d = 0.57$).

In fact, if we use the analytics panel to compare the top and bottom 50% of baseline RBDQ ratings, the Mann-Whitney test shows that the distributions of nearly all outcome measures at Y3 are significantly different ($p < 0.05$) between the two groups except for a single cognitive measure, the semantic fluency test (SFT). These results show that the baseline severity of REM sleep behavior disorder is an important predictor of three-year outcomes across all the domains. Further studies are needed to investigate this important finding.

The difference in outcomes at Y3 between patients with more and less baseline REM sleep behavior disorder becomes even more apparent if we adjust the filter to select increasingly greater RBDQ ratings at BL (Figure 6a). The patient groups with the top 50%, 25%, and 10% baseline RBDQ ratings have increasingly greater mean values of T-UPDRS at Y3 (50.17, 53.65, and 58.54 from left to right in Figure 6b). In summary, these results show a positive correlation between the severity of baseline REM sleep behavior disorder and the severity of multiple other domains at year three, indicating that the baseline severity of REM sleep behavior disorder is an important predictor of Parkinson's disease progression.

6.2 Discussion

Winnow enables clinicians and scientists to interactively and flexibly query complex and high-dimensional datasets with minimal training. This hands-on experience is critical for clinicians to generate hypotheses and to discover actionable, clinically-relevant results. Currently hypotheses are investigated with basic statistics; in the future we plan to use Winnow as an interface for complex analytic methods, such as supervised learning. For example, users can define patient subgroups with distinct profiles of progression in Winnow and then use these defined subgroups as the targets for supervised learning methods.

Based on the features of the dataset used in our examples, our analysis is limited to three years of longitudinal data. Whereas Winnow currently supports comparing two groups of patients (selected and unselected), future modification will enable selecting and switching between multiple patient groups to allow more sophisticated comparisons. In addition to the outcome measures used in our preliminary study, biologic data (e.g. genomics and serologic markers) may enable more precise and reliable groupings.

7 CONCLUSIONS

Over the last 25 years, chronic medical conditions have been redefined with an expanding scope of symptoms and biologic disease markers. These advances have resulted in complex multidimensional clinical datasets. Such datasets pose substantial opportunities for discovery, but also pose unique challenges for traditional analysis. Parkinson's disease (PD) is a key example of these challenges–a complex progressive neurodegenerative disorder with rapidly expanding biologic data including genetics, imaging, and serologic markers. Since the diverse symptoms of PD progress over different timeframes, and are likely to vary in different patient subgroups, novel tools and approaches are needed to capitalize on the large, multidimensional, and longitudinal datasets available today.

In this paper we present Winnow, a visual analytics tool that is intuitive, interactive and insightful for both scientists and clinicians. The design of Winnow is based on an intensive long-term collaboration between experts from diverse backgrounds. Winnow is a significant step to conduct complex analysis on existing databases that are necessary to advance the study of chronic medical conditions.

We are designing more concise and scalable visualizations by grouping similar patients in the outcomes panel and creating a graph-based representation summarizing the changes in outcomes [15]. We are also working on extending Winnow to incorporate new emerging biomarkers including genetics, imaging, serology, and biosensor metrics. The extension will also include a machine learning module to perform automatic data analysis.

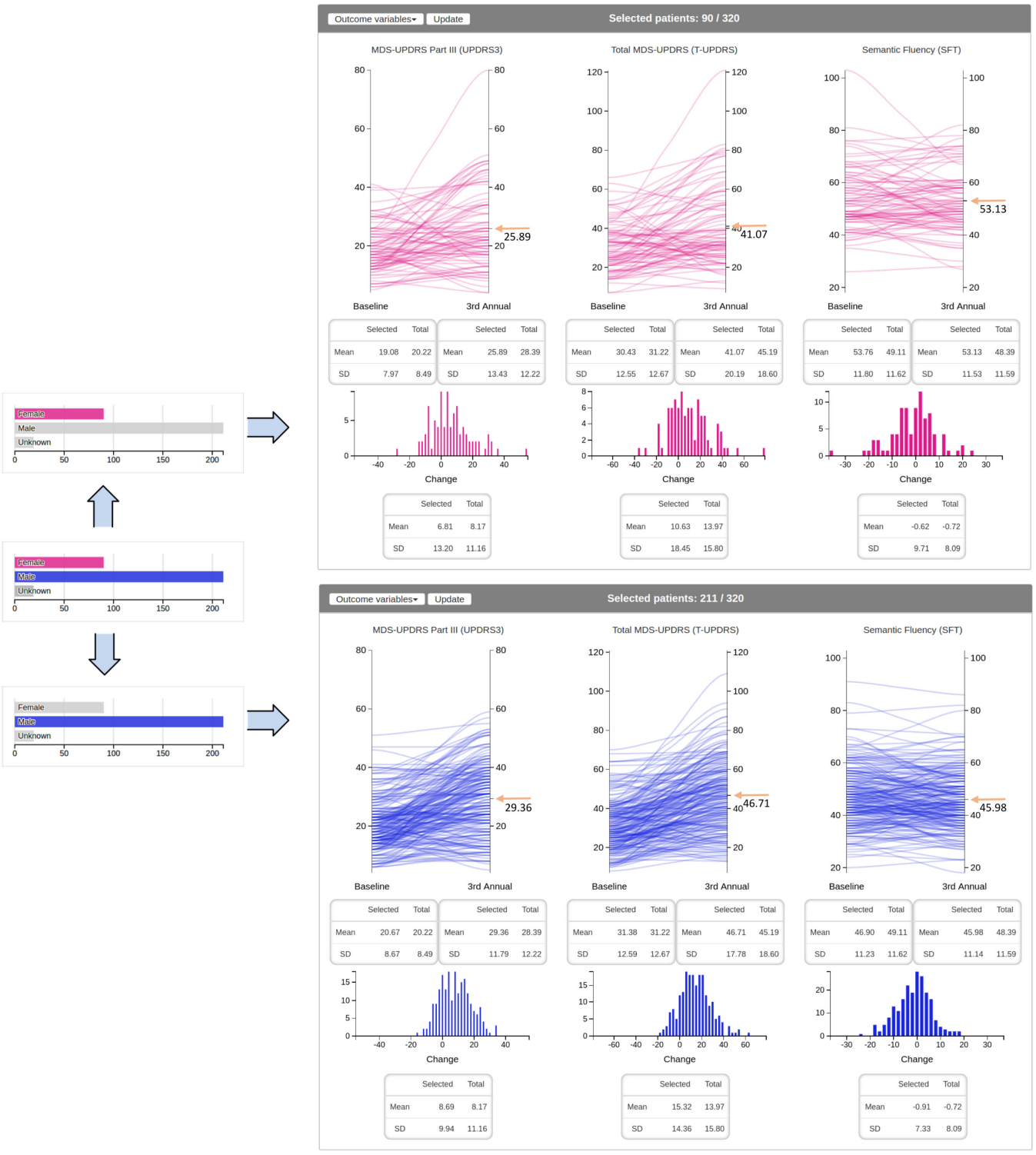
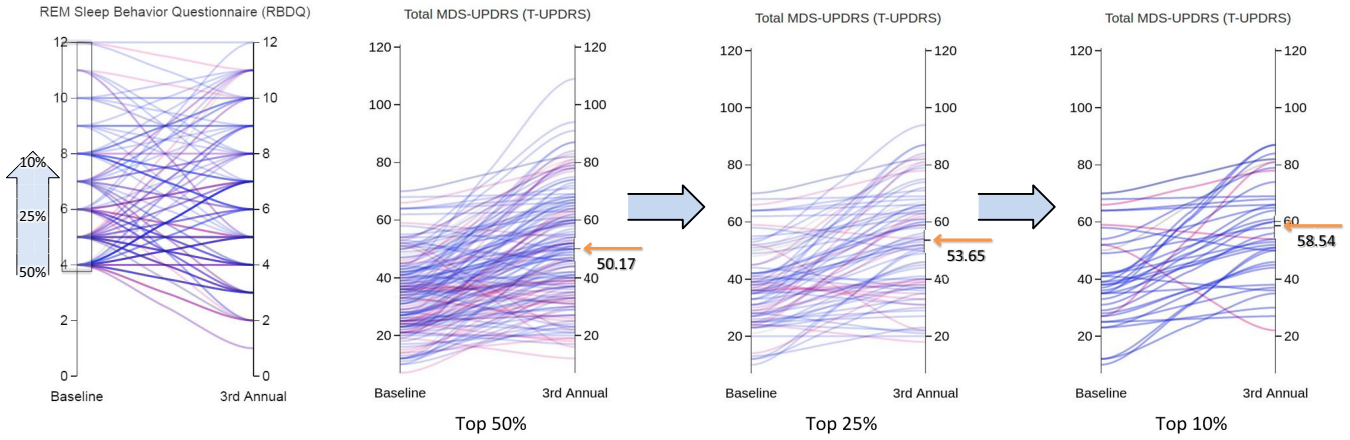


Figure 5: After applying the gender filter, the plots show that female patients (top) have comparable motor symptoms (UPDRS3) and general PD severity (T-UPDRS) at BL and less severe symptoms at Y3 than male patients (bottom). In contrast, female patients have more severe symptoms on semantic fluency (SFT) at Y3. arrows mark the mean values at Y3 for the males and females.



(a) Filtering baseline RBDQ

(b) Increasing mean T-UPDRS at Y3 from 50% to top 10% of patients (left to right)

Figure 6: (a) A filter that selects patients with increasingly greater baseline RBDQ. (b) As the filter moves to select patients with more severe baseline RBDQ (top 50%, 25%, and 10% from left to right), the mean T-UPDRS at Y3 increases from 50.17 to 53.65 and 58.54.

ACKNOWLEDGMENTS

This research is supported in part by the Seed Grant Program of the University of Maryland. We thank Dr. Søren Bentzen for useful advice on data analysis. We also thank PPMI for providing the data used in the case study. The PPMI dataset is a public-private partnership, funded by the Michael J. Fox Foundation for Parkinson's Research and the Fox funding partners- AbbVie, Avid, Biogen, Bristol-Myers Squibb, Covance, GE Healthcare, Genentech, GlaxoSmithKline, Lilly, Lundbeck, Merck, Meso Scale Discovery, Pfizer, Piramal, Roche, Sanofi Genzyme, Servier, Teva, and UCB.

REFERENCES

[1] Mihael Ankerst, Markus M. Breunig, Hans-Peter Kriegel, and Jörg Sander. 1999. OPTICS: Ordering Points to Identify the Clustering Structure. In *Proceedings of ACM SIGMOD International Conference on Management of Data*. ACM, New York, NY, USA, 49–60.

[2] Mikhail Belkin and Partha Niyogi. 2003. Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. *Neural Computation* 15, 6 (2003), 1373–1396.

[3] D. J. Burn, E. N. Rowan, L. M. Allan, S. Molloy, J. T. O'Brien, and I. G. McKeith. 2006. Motor Subtype and Cognitive Decline in Parkinson's Disease, Parkinson's Disease with Dementia, and Dementia with Lewy Bodies. *Journal of Neurology, Neurosurgery, and Psychiatry* 77, 5 (May 2006), 585–589.

[4] William C. Cleveland and Marylyn E. McGill. 1988. *Dynamic Graphics for Statistics* (1st ed.). CRC Press, Inc., Boca Raton, FL, USA.

[5] Jacob Cohen. 1992. Statistical Power Analysis. *Current Directions in Psychological Science* 1, 3 (June 1992), 98–101.

[6] Geoffrey Ellis and Alan Dix. 2006. Enabling Automatic Clutter Reduction in Parallel Coordinate Plots. *IEEE Transactions on Visualization and Computer Graphics* 12, 5 (2006), 717–724.

[7] Geoffrey E. Hinton and Ruslan R. Salakhutdinov. 2006. Reducing the Dimensionality of Data with Neural Networks. *Science* 313, 5786 (July 2006), 504–507.

[8] Alfred Inselberg. 1985. The Plane with Parallel Coordinates. *The Visual Computer* 1, 2 (Aug. 1985), 69–91.

[9] J. Jankovic, M. McDermott, J. Carter, S. Gauthier, C. Goetz, L. Golbe, S. Huber, W. Koller, C. Olanow, I. Shoulson, M. Stern, and C. Tanner. 1990. Variable Expression of Parkinson's Disease: A Base-Line Analysis of the DATATOP Cohort. *Neurology* 40, 10 (Oct. 1990), 1529–1534.

[10] Roxanne E. Jensen, Nan E. Rothrock, Esi M. DeWitt, Brennan Spiegel, Carole A. Tucker, Heidi M. Crane, Christopher B. Forrest, Donald L. Patrick, Rob Fredericksen, Lisa M. Shulman, David Cella, and Paul K. Crane. 2015. The Role of Technical Advances in the Adoption and Integration of Patient-Reported Outcomes in Clinical Care. *Medical Care* 53, 2 (Feb. 2015), 153–159.

[11] Stuart Lloyd. 1982. Least Squares Quantization in PCM. *IEEE Transactions on Information Theory* 28, 2 (1982), 129–137.

[12] Ivy N. Miller and Alice Cronin-Golomb. 2010. Gender Differences in Parkinson's Disease: Clinical Characteristics and Cognition. *Movement Disorders* 25, 16 (Dec. 2010), 2695–2703.

[13] Hans-Christoph Nothdurft. 1993. The Role of Features in Preattentive Vision: Comparison of Orientation, Motion and Color Cues. *Vision Research* 33, 14 (Sept. 1993), 1937–1958.

[14] Parkinson Progression Marker Initiative. 2011. The Parkinson Progression Marker Initiative (PPMI). *Progress in Neurobiology* 95, 4 (Dec. 2011), 629–635.

[15] Robert Patro, Cheuk Yiu Ip, Sujal Bista, Samuel S. Cho, D. Thirumalai, and Amitabh Varshney. 2011. MDMap: A System for Data-Driven Layout and Exploration of Molecular Dynamics Simulations. In *Proceedings of IEEE Symposium on Biological Data Visualization*. 111–118.

[16] Wei Peng, Matthew O. Ward, and Elke A. Rundensteiner. 2004. Clutter Reduction in Multi-Dimensional Data Visualization Using Dimension Reordering. In *Proceedings of IEEE Symposium on Information Visualization*. 89–96.

[17] Sam T. Roweis and Lawrence K. Saul. 2000. Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science* 290, 5500 (Dec. 2000), 2323–2326.

[18] Jianbo Shi and Jitendra Malik. 2000. Normalized Cuts and Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, 8 (2000), 888–905.

[19] Ben Shneiderman and Catherine Plaisant. 2006. Strategies for Evaluating Information Visualization Tools: Multi-Dimensional in-Depth Long-Term Case Studies. In *Proceedings of the International Conference on Advanced Visual Interfaces*. ACM, 1–7.

[20] Lisa M. Shulman, Robin Leifert Taback, Judy Bean, and William J. Weiner. 2001. Comorbidity of the Nonmotor Symptoms of Parkinson's Disease. *Movement Disorders* 16, 3 (2001), 507–510.

[21] Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing Data Using T-SNE. *Journal of Machine Learning Research* 9, Nov (2008), 2579–2605.

[22] Stephanie M. van Rooden, Willem J. Heiser, Joost N. Kok, Dagmar Verbaan, Jacobus J. van Hilten, and Johan Marinus. 2010. The Identification of Parkinson's Disease Subtypes Using Cluster Analysis: A Systematic Review. *Movement Disorders: Official Journal of the Movement Disorder Society* 25, 8 (June 2010), 969–978.

[23] Amitabh Varshney and Arie Kaufman. 1996. FINESSE: A Financial Information Spreadsheet. In *Proceedings of IEEE Symposium on Information Visualization*. IEEE, 70–71.

[24] Asako Yoritaka, Hideki Ohizumi, Shigeki Tanaka, and Nobutaka Hattori. 2009. Parkinson's Disease with and without REM Sleep Behaviour Disorder: Are There Any Clinical Differences? *European Neurology* 61, 3 (2009), 164–170.

[25] Hong Zhou, Xiaoru Yuan, Huamin Qu, Weiwei Cui, and Baoquan Chen. 2008. Visual Clustering in Parallel Coordinates. *Computer Graphics Forum* 27, 3 (2008), 1047–1054.