Qyz: A Platform for Visual Analysis of Error, Abuse, and Fraud in Medical Bills

Noah Pepper*　Homer Strong†　Kevin Lynagh‡

Qmedtrix Systems Inc.

Abstract

Qyz is a novel domain-specific visualization tool that provides a comprehensive overview of error, fraud, and abuse in large sets of medical bills. The system supports users from a wide range of technical backgrounds and access levels. A typical analysis covers medical codes, line-level pricing, and gives broad geographic and regulatory context for a set of bills. Qyz graphically presents descriptive statistics regarding bill overpayment, medical coding error frequencies, procedure co-occurrences, and other variables of interest. The system provides interfaces for both interactive analysis and automated generation of easily-customizable, distribution-quality reports.

Index Terms: J.3 [Computer Applications]: Life And Medical Sciences—Medical information systems

H.1.2 [Information Systems]: Models and Principles—Human information processing

H.3.3 [Information Systems]: Information Storage and Retrieval—Clustering

1 Introduction

Qyz is a medical bill analysis tool which combines machine learning and data visualization tools to create focused mosaics that clearly depict patterns occurring in medical payment data. The tools discussed herein are designed for analysis within the medical domain, where extreme data sanitation problems and the complexity of regulatory environments and coding systems create significant opportunities for error, abuse, and fraud. The development of our tools is motivated by the difficulty of verifying medical claims. The resulting systems ultimately has applications in the analysis of utilization, economic trends, and risk assessment. In this paper we are entirely focused on medical data and billing patterns that occur within the United States.

We consider any billing mistake on a medical claim to be an error, and we consider patterns of these errors resulting in excess payments to providers to be abuse. The systematic and intentional abuse of medical billing procedures by individual providers we consider to be outright fraud. Many estimates put the losses from such fraud and abuse in the United States at $75 to $250 billion per year [13, 11]. While the amount lost to medical fraud and abuse is difficult to estimate with certainty, there is no doubt that it is a very expensive problem.

That such rampant abuse persists reflects the complexity of the practices and laws surrounding medical billing. While there are many potential applications for machine learning and automation in the health care cost containment industry, ultimately much of the work must be performed manually. However, prompt payment laws and other industry-specific regulations limit the time analysts have to detect errors and fraud in bill sets. Thus, it is important to create tools that allow humans to work as efficiently as possible with data. Bills must be reviewed, data sanity problems corrected, and bill sets summarized and returned to clients in a timely fashion. Sokol *et al.* [15] find that roughly 80% of time spent data mining health care data is used on exactly these formerly manual preprocessing tasks. While this paper does not discuss the actual data cleansing, the tools described here are integral to accomplishing that task. The present work exhibits the design of a tool for extracting and visualizing interesting and insightful statistics from sets of medical bills.

Qyz streamlines the detection and prevention of improper payments for health care services. The system quickly illuminates patterns and trends in very large sets of medical bills (thousands to millions of bills) to users with varying levels of technical sophistication. Our system is flexible such that it can create a report advanced enough to be useful to a skilled analyst while simultaneously constructing another report suitable for a salesperson[1] all within seconds, from the same data. There is evidence [3] that information visualization improves comprehension for such cases of synchronous group knowledge.

The process of creating such a system is non-trivial. Medical bills are structured, multidimensional data, and each set of bills tells a different story. While each individual bill may have a complex architecture, great difficulties also stem from the many disparate health care billing standards and platforms within the industry. Analysis of bills from these disparate platforms reveals heterogeneous patterns of error across bill sets. Fraud and abuse are likewise heterogeneous beasts that require distinct attention. Schemes of abuse and fraud are often specific to a medical specialty, facility type, geographic area, or even a single provider.

The goal of this system is to provide near-instant analysis of arbitrary claims dumps to authorized individuals. Such a tool must support a diverse group of users, from analysts and software developers to sales staff, doctors, and executives. The reports must be tailored to provide the appropriate level of access, detail and sophistication depending on the user's role. For example, a client executive may wish to see how a claims dump will effect her bottom line, while an analyst searches for data sanity problems and a bill reviewer looks for overpayments and fraud.

While there exist some report generation tools in the health care industry [10], Qyz is unique in that it weaves visualizations together in an educational and aesthetically pleasing fashion. Fundamentally, the novel offering of this paper is a look into the construction of an automated, domain-specific visual story telling machine. The central claim of this paper is that the proper juxtaposition of familiar and intuitive visualizations is the most powerful way to understand, explain, and explore complex medical billing data.

This is especially true given that our intended audience has a wide range of technical backgrounds and we cannot expect each

*e-mail: npepper@qmedtrix.com

†e-mail: hstrong@qmedtrix.com

‡e-mail: klynagh@qmedtrix.com, corresponding author

[1] Since we are in the business of medical bill review it is important to be able to communicate how much abuse or fraud we find in a data set. Clients or potential clients need to understand how much our services can save them. A proper representation of complex data is critical to explain the value we provide to our clients.

user to learn an entirely new visualization paradigm. Rather than create very specialized visualization tools for each use case, we have a consolidated framework for distributing specialized information to many users, each of whom has individual interests and concerns. With accessibility as a concern, Qyz judiciously chooses illuminating analyses which are appropriate for each particular use case.

Qyz employs many established visualization techniques, and draws heavily on empirical information visualization research. In line with their natural interpretations [17], we use bar charts for quantitative data that fall into distinct categories and line charts for time series. All of our colored graphics use one of Brewer's [4] qualitative, sequential, or diverging color schemes to most informatively reflect the data.

The visualizations used were selected based upon three criterion: 1) efficiency of information transfer, 2) ease of explanation to non-experts and 3) aesthetic appeal. Users quickly become familiar with the intuitive and classic visualization techniques used by Qyz; histograms, line graphs, scatter plots, chloropleths, etc. We use more rarefied techniques such as density plots, heat maps, and visualizations in semantic space when as appropriate to communicate specialized analyses. The focus on simplicity and accessibility is important because, while our system is designed to be used by trained analysts, the output often ends up in the hands of executives without specialized quantitative-visualization experience.

2 EXPOSITION

The system we describe here is a data pipeline where raw medical billing data is transformed into human readable, actionable results. The goal is to concisely expose error, then abuse, and finally highlight potential instances of fraud.

2.1 Technical Features

The Qyz system is designed using a modular architecture, which allows it to be distributed across multiple computers. Some nodes play fixed, specialized roles (i.e. a central database), while others perform general statistical calculations as needed.

This modular architecture enables multiple interfaces to the Qyz analytics software. Pre-designed PDF reports can be automatically generated on new data for marketing teams while analysts interactively explore and analyze data from a web interface. Reports on aggregate data–weekly totals, changes in medical code proportions, and other trends–can be emailed to executives automatically. Multiple users can request different kinds of analyses on several bill sets simultaneously, with Qyz transparently caching shared computational results.

2.2 Error Detection

Prior to any abuse or fraud analysis medical bills must be screened for coding errors and classified based upon origin. There are many classification schemes used throughout our bill processing engine, but the most fundamental is to assign an adjudication type to every bill [2].

The services and associated charges on a medical bill are represented by diagnosis and procedure codes. There are several varieties of codes and we must be able to process and analyze all of them. The coding systems are non-trivial; each containing several thousand codes, and sets are often amended, updated, and certain codes deprecated by standards authorities. Coding errors can originate from the provider or from software errors in 3rd party claims handling systems. The opportunity for error in a bill's codes are significant, especially given the prevalence of paper billing. Both manual bill entry workers and OCR technologies introduce errors, which are often hard to detect.

For most standard reports, Qyz describes coding errors, their associated corrections, and problems severe enough to prevent the bill from being adjudicated. In this initial error detection step, we can identify patterns of missing data by looking at macro-level distributions of codes throughout the data set (figure 1).

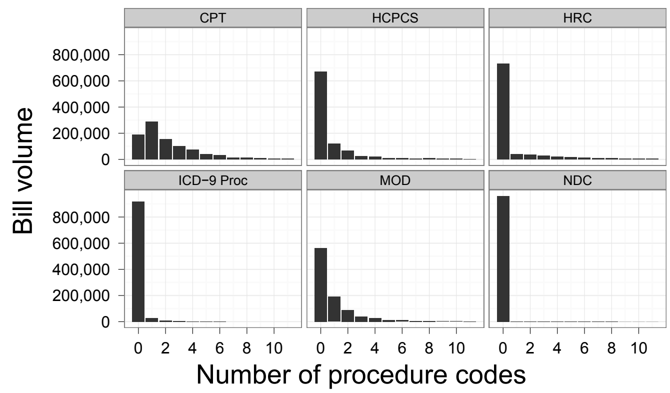

Figure 1: Histograms of the number of each kind of medical code that appears in a particular set of bills. Notice, for example, that many of these bills have no ICD-9 procedure codes. This makes inpatient hospital bills in many states impossible to audit; this simple analysis reveals critical incompleteness in the data.

This medical code analysis, performed immediately on receipt of bills, informs the decision to request additional information from the entity providing the data. Exactly what additional information to request is determined by an expert human, according to the circumstances. Using the Qyz output, the human may quickly specify the appropriate information necessary to performing nontrivial bill review tasks on a timely basis.

2.3 Preliminary Analysis

After checking (and possibly correcting) fields within the bills, the next standard step is to take a global look at the data set. Qyz satisfies the recommendations of Kang *et al.* [7] for investigative analysis tools by providing appropriate starting points for analysis and suggesting more detailed directions. This includes an assessment of the distributions of bill charges across regions, adjudication types, and procedures, which allows one to identify a bill set's largest cost factors. For example, this may involve looking at the adjudication type of a bill to determine where the most expensive bills lie (usually in hospital and surgery center bills, as in figure 2).

This can be especially useful when dealing with adjudication types or jurisdictions which require extensive manual auditing. Bill processing is often dictated by prompt pay laws; as a bill review company we must return results to clients within 24 hours of receipt of a data load. This makes it critical to prioritize those bills which will require manually auditing carefully such that precious time is not wasted.

Qyz uses several heuristics to find unusual and potentially interesting codes and providers. The preliminary reports include more detailed visualizations on these outliers, complementing the general summaries by bringing subsets of the data to analysts' attention. For the common use cases, we determine certain heuristics (such

[2] The adjudication type of a bill is a classification which indicates which laws govern the adjudication of the bill. Essentially one can think of this as the origin of the bill in terms of facility type. In the scheme showed here a bill can originate form an ambulance (AMB), ambulatory surgery center (ASC), durable medical equipment supplier (DME), emergency room (ER), inpatient hospital (IPH), outpatient hospital (OPH) or doctor's office/professional services (PRO). Each adjudication type has distinct rules regarding billing and payment and this process is required to do effective abuse detection.

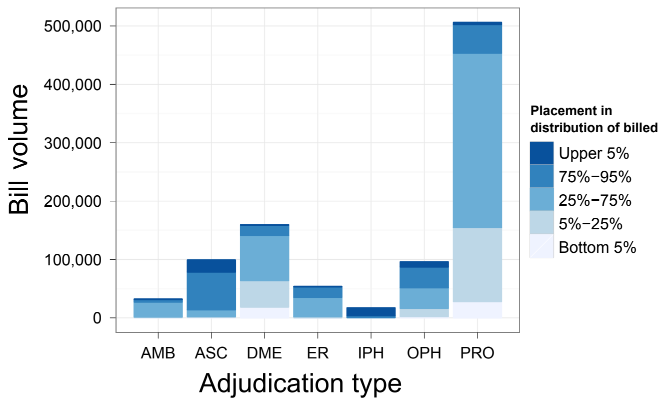

Figure 2: Histogram of bills by origin of the bill. Note that almost all of the inpatient hospital (IPH) bills are in the top 5% of the cost distribution, even though they make up the smallest proportion of this set.

as correlations or variances) that tend to bring attention to relevant aspects of the data. End users are provided with subsets of the data most applicable to their interests without having to compose a technical query.

These heuristic approaches are similar to an integration of machine learning and visualization suggested by Fuchs *et al.* [6], although there are major differences. Qyz relies more heavily on guidance from experts to suggest hypotheses, but shares the strategy in which the computer and analyst inform one another. Another major difference is that Qyz produces analyses that are accessible to a very wide audience, rather than more specialized techniques such as volume rendering. Additionally, the heuristics employed by Qyz depend on the particular user, according to user interests such as variances and correlations between certain variables. The expressions involved and the groups over which expressions are evaluated vary with the application.

2.3.1 Example: Regional Analysis

Breaking down medical data geographically illuminates not only where the bulk of the costs lie, but also where savings are likely to be found. We find chloropleths and small-multiples [16] maps to be especially useful for identifying problem areas and revealing valuable relationships in data. We can, for instance, break down the adjudication types of figure 2 geographically; see figure 3.

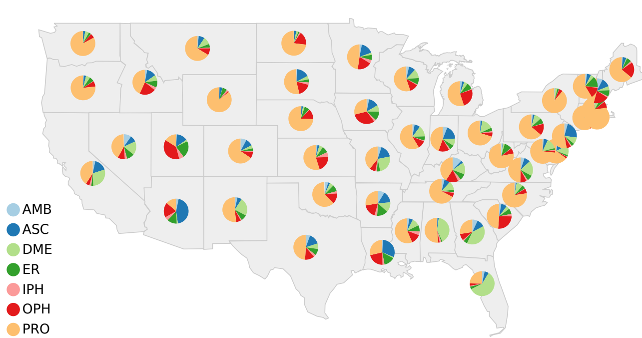

Figure 3: Map depicting the relative bill adjudication types within each state.

2.4 Exploratory fraud and abuse detection

There are two approaches to fraud and abuse detection. The first, more straightforward, method is to search for known patterns of abuse, and flag the result set as suspect. This is how the bulk of our automated system works. The second method involves using unsupervised machine learning combined with human auditors and visualization techniques to detect new patterns of fraudulent charges.

The latter method eventually informs the former. Those patterns human auditors designate as robust can be graduated to a known pattern in the first tier of abuse detection. In this way, human auditors constantly act as filters of billing tendencies, deciding which patterns consistently indicate fraud. The role of the human auditor is particularly important because they are capable of recognizing novel abuse patterns and immediately training automated detection systems.

Qyz allows auditors to efficiently recognize known patterns and explore trending billing practices. Ultimately, an auditor must decide if a pattern in the data indicates fraud. Thus Qyz's role is to highlight unusual, potentially fraudulent patterns. The strategy taken by Qyz is to display a broad assortment of summaries related to the known indicators of problematic billing, and to carefully choose subsets of data to display in greater detail.

After seeing the first report, the auditor can tailor more specific reports on certain aspects of a data set. Both the customization and the generation of reports are fast; auditors receive information as soon as Qyz processes new data. In each iteration, Qyz not only provides the analysis requested, but also suggests new directions to investigate.

2.4.1 Example: ICD-9 Analysis for DRG Up-code Detection

One extremely prevalent form of abusive billing is Diagnosis Related Group (DRG) up-coding. The DRG system was invented by Medicare to simplify hospital billing. Essentially one uses a set of rules to map a set of diagnosis and procedure codes determined by a hospital to a single DRG code. The DRG code combined with a hospital's Medicare-assigned conversion factor [3] determines how much the provider should be reimbursed for on entire inpatient stay.

While payment schedules in many ways simplify the payment process, they also introduce perverse incentives to 'game' the medical codes on a bill. A provider may include bogus diagnosis codes to trigger a 'complications' or 'major complications' flag on the DRG. This flag increases the level of reimbursement, but is only intended for use in legitimate cases of complex, difficult or especially expensive procedures [14].

While there are automated methods capable of determining upcoding in many cases where human intervention is required since a study of the hand-written operating report may be needed. A single up-coded bill can lead to tens of thousands of dollars of overpayment thus making it reasonable and worthwhile to invest some extra time ensuring no improper bills are paid.

The ICD-9 diagnosis heatmap seen in figure 4 quickly draws the auditor's attention to the pairs of ICD-9 codes which lead to the greatest payments. The pattern to search is pairs of ICD-9 codes for which their co-occurrence leads to much higher costs than either of them do with other combinations of ICD-9 codes. However, some of these are medically legitimate and some are not. Additionally some may be so expensive or rare that an auditor will see fit to examine the operating report associated with the bill. While Qyz can pick out suspicious code combinations and help auditors visualize their costs, the auditor must decide which bills are up-coded.

[3] A value that proxies for how expensive that hospital is to run; determined by Medicare after review of a hospital's 'Medicare Cost Report'.

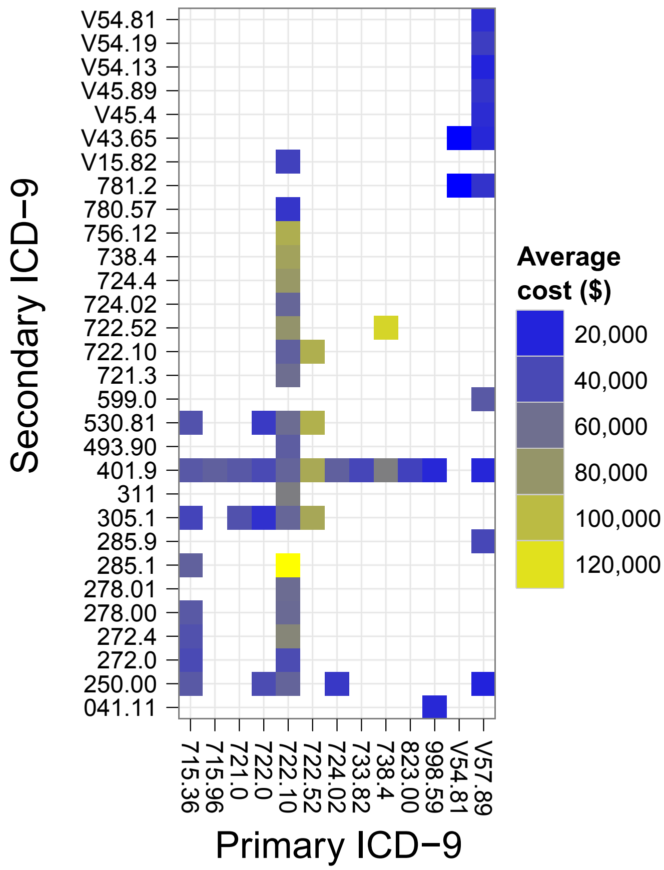

Figure 4: Heatmap of the commonly occurring ICD-9 diagnosis code pairs (from 2772 bills), colored by average costs. Note that some combinations are significantly more expensive than others, possibly indicating that those combinations are being used for up-coding their corresponding DRGs.

2.4.2 Code Clusters

Individual codes rarely appear in isolation on a medical bill. Depending on adjudication type, bills may have anywhere between five and hundreds of codes. Billing abuse often materializes in the form of unbundling or cost shifting. The former is where a procedure, P_1, is billed that should include the cost of ancillary expenses P_a and P_b but where those are billed in addition to P_1. The latter occurs when some instances of a code are billed inappropriately depending what other codes coincide with it on a bill.

More generally it can be helpful for an analyst to see the common code clusters in a data set and the distribution of charge across clusters. There are so many codes and charges, however, that any undirected detailed of such analysis by a human of such data is infeasible. We therefore have used a clustering and dimensionality reduction technique to make such an analysis tractable. The output an analyst looks at may be several pages of graphs like figure 5. This type of graph is especially illuminating when examining collections of bills from individual providers, as it can show patterns of erratic or abusive billing.

We analyze the real world usage of medical codes as observed in billing data using two techniques. The first is the well-known study of *frequent itemsets*. In the standard data mining terminology, we treat bills as *transactions* and the CPT codes as *items*. We use the Apriori algorithm [1, 2] to efficiently discover association rules

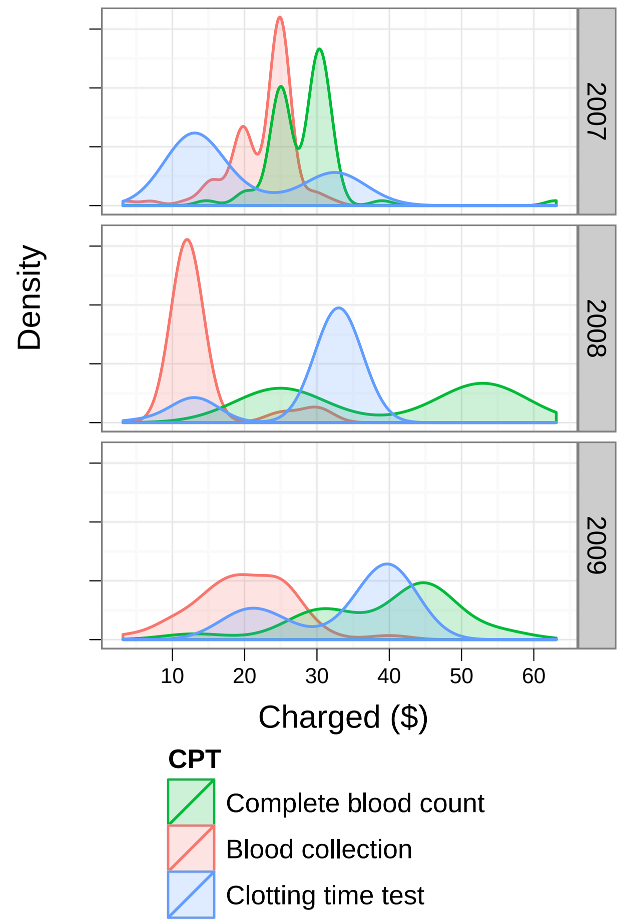

Figure 5: Density estimation of bill charges across a set of three codes over several years. This is helpful for determining if there is a significant level of unbundling or other kind of gaming of codes happening amongst the codes in these clusters. The cluster showed here was automatically created using the algorithm described later in this section.

of the form $A \Rightarrow B$, where A, B are disjoint subsets of all possible medical codes.

However, this approach can be over-specific. For instance, there are seven CPT codes pertaining to facial wounds, so the Apriori algorithm may fail to discover robust rules involving the generalized notion of "facial injury". To overcome the problem of roughly synonymous medical codes, we employ latent semantic analysis [5]. We construct the bill-code matrix $\mathbf{A} \in \{0,1\}^{m \times n}$, where the m rows represent bills and the n columns represent medical codes; $\mathbf{A}_{ij} = 1$ if bill b_i contains a charge for code c_j, and zero otherwise. We perform singular value decomposition on this bill-code matrix to give

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^{\mathrm{T}},$$

with matrices $\mathbf{U} \in \mathbb{R}^{b \times c}$ and $\mathbf{V} \in \mathbb{R}^{c \times b}$ having orthonormal columns. The matrix $\Sigma \in \mathbb{R}^{c \times c}$ is the diagonal matrix of singular values. We can then reduce the dimensionality of our dataset by constructing $\mathbf{A}_r$, the best rank-r approximation to $\mathbf{A}$ (in the least squares sense), by throwing out all but the k largest elements of Σ

and zeroing the corresponding columns in $\mathbf{U}, \mathbf{V}$. This gives

$$\mathbf{A}_r = \mathbf{U}_r \Sigma_r \mathbf{V}_r^{\mathrm{T}},$$

with $\mathbf{U}_r \in \mathbb{R}^{b \times r}$, $\Sigma_r \in \mathbb{R}^{r \times r}$, and $\mathbf{V}_r \in \mathbb{R}^{c \times r}$.

We can then measure similarity between medical codes as the cosine distance between the rows of $\mathbf{V}_r$. We construct the symmetric affinity matrix $\mathbf{A} \in \mathbb{R}^{c \times c}$, with $\mathbf{A}_{ij}$ given by the cosine distance between rows i, j of $\mathbf{V}_r$, and use it to cluster the codes according to the methodology of Ng *et al.* [12].

Spectral clustering allows us to further reduce the dimensionality of our dataset so that we can visualize clusters of medical codes. If there are k-clusters in the data, spectral clustering projects the clusters to orthogonal points on the k-sphere (see reviews by Luxburg [8], Lynagh [9]). Even with only the first three dimensions given by the spectral clustering procedure, codes form meaningful groups. Three common codes related to blood tests, for instance, appear together (see figure 6). To find clusters programmatically and in higher dimensional spaces, we use a nearest-neighbor method.

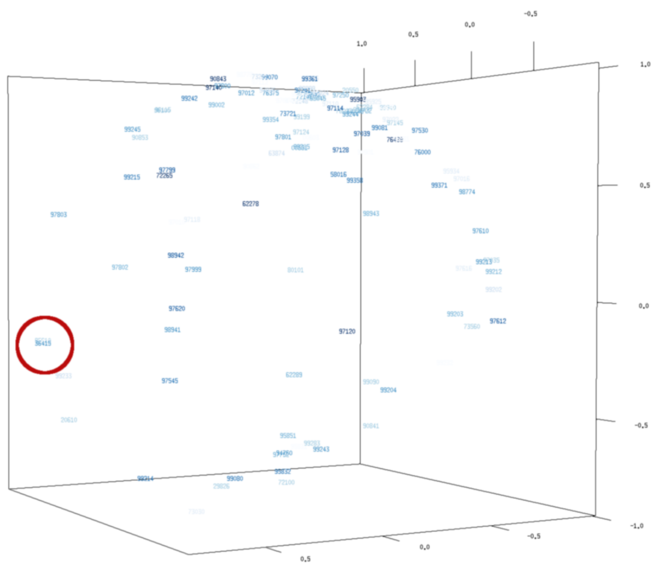

Figure 6: Screenshot of an interactive tool displaying medical codes in the first three dimensions given by spectral clustering, colored by their average line charges (darker is more expensive). A cluster of CPT codes 85025, 36415, 85610 is highlighted by the red circle, corresponding to charges for complete blood count, collection of blood by venipuncture, and Prothrombin (clotting) time tests.

2.4.3 Problem Provider Profiling

Some providers show consistent patterns of abusive or fraudulent billing. While some overpriced bills are simply errors in billing which can be negotiated amicably there are a few providers who aggressively pursue abusive billing practices and fraud.

A graphic like this is useful to an analyst since a robust pattern of increasing cost (see figure 7) demonstrates that a patter of fraud is likely, while randomly distributed high-cost charges may represent sporadic abuses, errors or multiple provider locations using a single TIN.

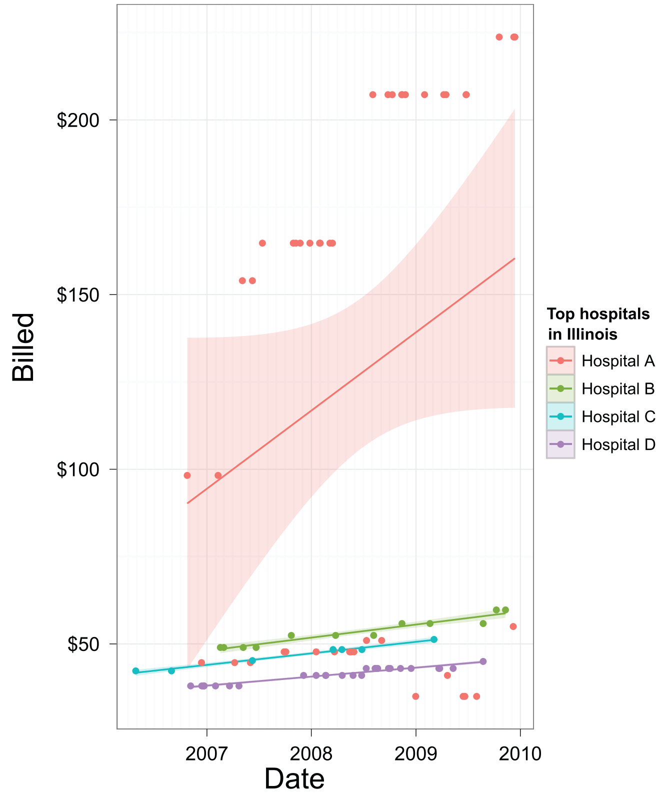

Figure 7: Time series scatter plot with fitted linear models and 95% confidence interval for the top five ER hospitals which most frequently bill for tetanus shots. This figure was automatically generated because a suspicious pattern was detected. Indeed, one hospital has been pushing for higher payments on this high frequency low dollar service for the past five years.

2.5 Audit Results

After the bills have been validated and classified they can be audited and proper payments for each code and each bill can be determined. This is a largely automated process, but humans do some of the work, especially when investigating abuse or fraud.

Once bills have been re-priced and the appropriate payment determined, we report our progress to the client using graphics like figure 8. It is also useful for us to determine the constantly changing sources of cost reductions. The state of audits can be tracked at regular and frequent intervals with privilege- and interest- specific data. This analysis usually consists of about 20 pages of tables and graphs to show various facets of data based upon interesting permutations of codes, states, and classifications.

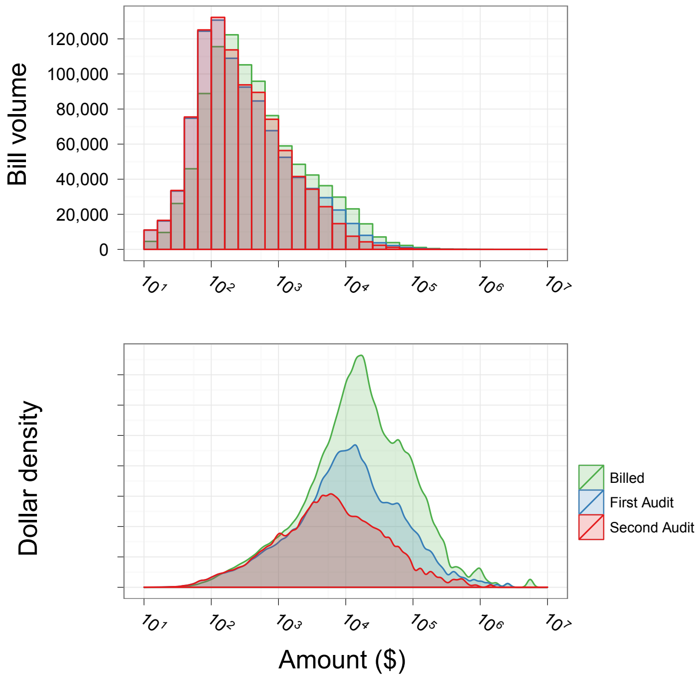

Figure 8: Density estimation of bill cost after successive audits.

3 Conclusion

In a business environment, we have found that Qyz allows a variety of users to easily and quickly traverse previously unmanageable quantities of information.

Qyz has greatly increased the volume of bills individual analysts can process, making our business processes far more efficient. Before the introduction of this systems analysts would produce reports manually using SQL and Excel. This meant that individual reports were produced in a matter of hours, or possibly days. Now report generation is a process that takes less than five minutes and can be controlled by a simple drag-and-drop web interface.

We have been able to speed up operations and provide a responsive, automated level of service to our clients. The most valuable outcome of using this engine by far is the quality of analysis. Many of the components that make up Qyz are completely automated, which guarantees that reports are created accurately and consistently. Furthermore, because the layout and graphing are automated, we can be sure that even rushed-jobs will have a consistent, visually attractive appearance. Ultimately Qyz provides not only a speed increase, but serves as a quality control engine as well.

Further work includes more customizable reports, predictive analytics, and interactive visualizations on mobile devices. We are also experimenting with machine learning techniques to learn the interests of individual users and infer customized information retrieval criteria.

Acknowledgements

We would like to thank everyone else in the R&D lab at Qmedtrix, without whom this work would not be possible. Also, we would like to thank everyone who has contributed to R, Protovis, and the open source statistical community.

References

[1] R. Agrawal and R. Srikant. Fast algorithms for mining association rules. In *VLDB'94, Proceedings of 20th International Conference on Very Large Data Bases, September 12-15, 1994, Santiago de Chile, Chile*, pages 487–499, 1994.

[2] C. Borgelt. Efficient implementations of apriori and eclat. In *Proceedings of the IEEE ICDM Workshop on Frequent Itemset Mining Implementations*, 2003.

[3] S. Bresciani and M. J. Eppler. The benefits of synchronous collaborative information visualization: Evidence from an experimental evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 15:1073–1080, 2009.

[4] C. A. Brewer. Color use guidelines for mapping and visualization. In A. MacEachren and D. Taylor, editors, *Visualization in Modern Cartography*, chapter 7, pages 123–147. Elsevier Science, Tarrytown, NY, 1994.

[5] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman. Indexing by latent semantic analysis. *Journal Of The American Society For Information Science*, 41(6):391–407, 1990.

[6] R. Fuchs, J. Waser, and M. E. Gröller. Visual human+machine learning. *IEEE TVCG*, 15(6):1327–1334, 2009.

[7] Y.-a. Kang, C. Gorg, and J. Stasko. How can visual analytics assist investigative analysis? design implications from an evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 99(PrePrints), 2010.

[8] U. Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, August 2007.

[9] K. J. Lynagh. Computational analysis of protein structures: data visualization, vector quantization, and spectral clustering. Thesis, Reed College, May 2010.

[10] C. J. Matheus, G. Piatetsky-Shapiro, and D. McNeill. An application of KEFIR to the analysis of healthcare information. In *AAAI-94 Knowledge Discovery in Databases Workshop*. GTE Laboratories Incorporated, 1994.

[11] L. Morris. Combating fraud in health care: An essential component of any cost containment strategy. *Health Affairs*, 28(5):1351–1356, 2009.

[12] A. Y. Ng, M. I. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems 14*, pages 849–856. MIT Press, 2001.

[13] S. Rosenbaum, N. Lopez, and S. Stifler. Health insurance fraud: An overview. Technical report, George Washington University Medical Center School of Public Health and Health Services, Department of Health Policy, 2009.

[14] M. A. Rosenberg, D. G. Fryback, and D. A. Katz. A statistical model to detect DRG upcoding. *Health Services & Outcomes Research Methodology*, pages 233–252, 2001.

[15] L. Sokol, B. Garcia, M. West, J. Rodriguez, and K. Johnson. Precursory steps to mining hcfa health care claims. In *HICSS*, 2001.

[16] E. Tufte. *Envisioning Information*. Graphic Press, 1990.

[17] J. Zacks and B. Tversky. Bars and lines: A study of graphic communication. *Memory and Cognition*, 27(6):1073–1079, 1999.