Visualizing Online Auctions

Galit Shmueli & Wolfgang Jank

Decision & Information Technologies Dept

Robert H Smith School of Business

University of Maryland College Park MD 20742

Abstract

Online auctions have been the subject of many empirical research efforts in the fields of economics and information systems. These research efforts are often based on analyzing data from websites such as eBay.com which provide public information about sequences of bids in closed auctions, typically in the form of tables on HTML pages. The existing literature on online auctions focuses on tools like summary statistics and more formal statistical methods such as regression models. However, there is a clear void in this growing body of literature in developing appropriate visualization tools. This is quite surprising, given that the sheer amount of data that can be found on sites such as eBay.com is overwhelming and can often not be displayed informatively using standard statistical graphics. In this paper we introduce graphical methods for visualizing online auction data in ways that are informative and relevant to the types of research questions that are of interest. We start by using *profile plots* that reveal aspects of an auction such as bid values, bidding intensity, and bidder strategies. We then introduce the concept of *statistical zooming (STAT-zoom)* which can scale up to be used for visualizing large amounts of auctions. STAT-zoom adds the capability of looking at data summaries at various time scales interactively. Finally, we develop *auction calendars* and *auction scene* visualizations for viewing a set of many concurrent auctions. The different visualization methods are demonstrated using data on multiple auctions collected from eBay.com.

Key words: Bid data, eBay.com, Profile plots, STAT-zoom

1 Introduction

Almost every internet user today has heard, browsed, or used the online auction site eBay.com, a major online marketplace and currently the biggest C2C online auction place. The fascination with eBay has been documented in many recent reports and newspaper articles. From an economic point of view, eBay has been one of the few survivors of the late 1990's electronic commerce boom. In fact, eBay has not only survived but is growing faster than ever. This has led to a surge of empirical work based on data from eBay.com, typically by researchers from the fields of economics and information systems. The issues investigated in these papers range from exploring factors that affect final prices (Lucking-Reiley et al., 2000), analyzing the eBay reputation and feedback system (Dellarocas, 2001; Livingston, 2002; Resnick & Zeckhauser, 2001), finding empirical evidence for late bidding (sniping) (Roth & Ockenfels, 2002; Ockenfels & Roth, 2002), learning about commonly encountered effects such as the "Winner's curse" (Bajari & Hortacsu, 2003), detecting collusion (Kauffman & Wood, 2003a), investigating bidding strategies (Bapna et al., 2003; Ockenfels & Roth, 2002), modelling the bidder arrival process (Shmueli et al., 2004; Vakrat & Seidman, 2000), and more. Similar questions have been addressed by using data from other online auction houses such as ubid.com, amazon.com, and onsale.com. In this paper we focus on displaying data from eBay.com, but the methods could be adjusted for use with other online auction data.

eBay offers a vast amount of rich data. Besides the time and the amount of each bid placed in each auction, eBay also records plenty of information about the bidders, the seller, and the product being auctioned. On any given day, several million auctions take place on eBay and all closed auctions from the last 30 days are publicly available on eBay's website. This huge amount of information can be quite overwhelming and confusing for the user (here we refer to the user as either the seller, a potential buyer, or the auction house) who wants to incorporate this information into his/her decision making process. And of course for researchers who collect these data, it is also hard to sift through the information without appropriately visualizing it first. While standard statistical tools like summary measures and regression models are used frequently to answer specific research questions, there is a surprising void in methods that visualize the flood of information prevalent on eBay. The lack in adequate graphical displays starts at the very beginning, in describing the raw bid data. The few papers that do attempt to use graphical displays (e.g., Lucking-Reiley, 2000) tend to use over-simplified plots which in some cases even distort the information contained in the data. In this paper we make use of existing graphical displays as well as modify and develop new ones to visualize the information contained in bid data. Visualizations of historical auctions are useful as an exploratory tool for learning about bidding, selling, and winning on eBay.com or, more generally, in second-price sealed-bid online auctions. Our first aim is to expose and describe this unique type of data, which has not attracted much attention from statisticians. We point out the special features of online auction data and point out why ordinary statistical visualization methods require modification in some cases, while in other cases entirely new methods are needed. Our second aim is to highlight the need for adequate visualizations in the exploration of online data, and to introduce such graphics into the field of online auction research.

Raw eBay data come in the form of "bid histories", which are, from a technical point of view, HTML pages containing tables. These HTML pages are hard to grasp intuitively or to study directly, especially when looking at a multitude of concurrent auctions. Section 2 introduces typical examples of bid histories, their special structure and features, and the modern mechanisms that are used to collect them. In Section 3 we introduce a variety of simple visualization tools. We start by creating *Profile Plots*, a simple visualization of single or several bid histories, which preserves temporal information. We show what type of information is revealed by such displays and discuss their advantage over looking at the raw HTML pages. Several variations of the profile plot are illustrated, where additional features and enhancements can be integrated for various purposes of study (e.g., for exploring bidder behavior or bidding intensity throughout and auction). Finally, we discuss the problem of scaling profile plots for visualizing a multitude of auctions. This motivates the concept of *statistical zooming* (STAT-zoom), which we introduce in Section 4. The idea is to view data summaries at different time scales, thereby adding the capability of capturing the information contained in multiple bidding histories at a spectrum of time scales. Incorporating interactivity into the plots is known to be effective in increasing visual scalability (Eick & Karr, 2002). We implement the STAT-zoom concept for visualizing a large number of auctions. Section 5 discusses more complex types of visualizations that are useful for visualizing multiple concurrent auctions, either for a single item or for a variety of items. Two useful visualizations are *Calendars of Auctions* and *Auction Scene* maps. Section 6 concludes this paper with future directions.

2 The Data: Bid Histories on eBay.com

Understanding eBay's auction mechanism is central to understanding the special features and structure of eBay bid data. Another important factor is the special data collection mechanism which is typically used for gathering eBay data. Here we give a brief description of the auction and collection mechanisms and then explain and illustrate the structure of a bid-history for a closed-ended auction.

2.1 The eBay.com auction mechanism

Most of the auctions on eBay are second-price sealed-bid closed-ended auctions. eBay uses a proxy-bidding system where bidders are supposed to place the highest amount that they are *willing to pay* for the auctioned item. These values are usually abbreviated as WTP values (Bapna et al., 2003; Roth & Ockenfels, 2002). The system then automatically increases each bidder's bid by the minimum increment (which is relative to the highest asking price and set by eBay) until either the bidder's maximum has been reached or the bidder has the current high bid (Linoff & Berry, 2001). This guarantees that bidders will pay the minimum between their WTP value and an increment above the second highest bid. A bidder is free to place as many bids as he/she wishes.

During the ongoing auction the bidders' WTP values are not disclosed. Only the second-highest current price is displayed along with the usernames of participating bidders and the times that the bids were placed. This is in order to allow for the proxy-bidding system to work. Once an auction closes, eBay reveals the WTP values of all bidders except the winner. The complete "bid history", which includes the bid times and WTP values, is therefore available only for closed auctions.

A typical eBay closed auction page contains the sequence of bids, the bidder usernames and their rating, and the exact time & date when each bid was placed. There is also additional information about the seller (ID and rating), the prouct, shipping costs, etc. In this work we use the term "bid-history" mainly to describe the sequence of WTP values and the times they were placed.

Figure 1 displays a single closed auction page for a Palm M515 Personal Digital Assistant (PDA). Notice that the order in which eBay displays the bids is ascending in the WTP values, not chronologically! This makes it seem, at first, that the process of bidding was much more gradual and with higher intensity of bidding than actually occurred.

2.2 Data Collection Agents

Modern technologies allow for a convenient collection of large amounts of high quality data from the internet. The use of web agents or web spiders facilitate the creation of large databases of bidding data. A web agent is a software application, typically based on a programming language like Pearl or Java, that "crawls" over an internet site or a collection of web pages and gathers the desired information. In this form, data on hundreds, thousands, and even more auctions can be collected in a matter of only minutes. This modern automated collection system is much less error-prone than traditional data collection and recording. Unless the data on the website are erroneous or not sufficiently structured, the agent will usually deliver error-free data. However, pre-processing that relies on domain knowledge is still needed. For example, although most auctions are carried out in USD, occasionally a different currency is used. Kauffman & Wood (2003b) describe the revolutionary aspect of new data collection mechanisms such as software agents and discuss their impact on empirical research.

3 Displaying Raw Bid Histories

In this section we look at the raw data through informative, clear glasses. We start by displaying single auctions and then proceed to visualizing the information contained in multiple auctions.

3.1 Profile Plots: Displaying Single Bid Histories

A profile plot is a time plot of the WTP values over the duration of the auction. It is the first step in clarifying the information contained in a bid history. Looking at the data chronologically shows that many WTP values do not affect the current level of the price, since they do not exceed the highest WTP value at that time. Figure 2 displays profile plots for two 5-day auctions for a Palm M515 PDA.

home | pay | register | sign in | services | site map | help

Browse | Search | Sell | My eBay | Community

Powered By IBM

Search tips

Search titles and descriptions

eBay.com Bid History for

Palm Pilot m515 Color Handheld PDA 515 NEW NR (Item # 3074620884)

Currently	**US $200.50**	First Bid	**US $0.99**
Quantity	**1**	# of bids	**17**
Time left	**Auction has ended.**		
Started	Jan-28-04 20:30:00 PST		
Ends	Feb-02-04 20:30:00 PST		
Seller (Rating)	**uscagent (121)**		

View page with email addresses (Accessible by Seller only) Learn more.

Bidding History (Highest bids first)

User ID	Bid Amount	Date of Bid
golspice (26)	US $200.50	Feb-02-04 20:28:02 PST
audent (1)	US $198.00	Feb-02-04 17:18:14 PST
jay4blues (0)	US $196.50	Feb-02-04 14:57:54 PST
audent (1)	US $193.00	Feb-02-04 17:17:59 PST
audent (1)	US $188.00	Feb-02-04 17:17:33 PST
eagle2sc (11)	US $180.25	Feb-02-04 11:48:20 PST
audent (1)	US $179.00	Jan-31-04 09:18:51 PST
istariken (38)	US $175.25	Jan-30-04 20:45:20 PST
audent (1)	US $175.00	Jan-31-04 09:18:37 PST
audent (1)	US $169.55	Jan-31-04 09:18:16 PST
audent (1)	US $159.00	Jan-31-04 09:17:40 PST
amanda_s_brooks (14)	US $150.00	Jan-30-04 09:12:58 PST
audent (1)	US $80.05	Jan-28-04 21:54:09 PST
cscott24 (23)	US $80.00	Jan-28-04 20:35:05 PST
powergerbil (3)	US $45.01	Jan-28-04 21:05:14 PST
powergerbil (3)	US $42.99	Jan-28-04 21:04:44 PST
powergerbil (3)	US $40.00	Jan-28-04 21:04:08 PST

Figure 1: Bid History from eBay.com for closed auction of a Palm M515 PDA.

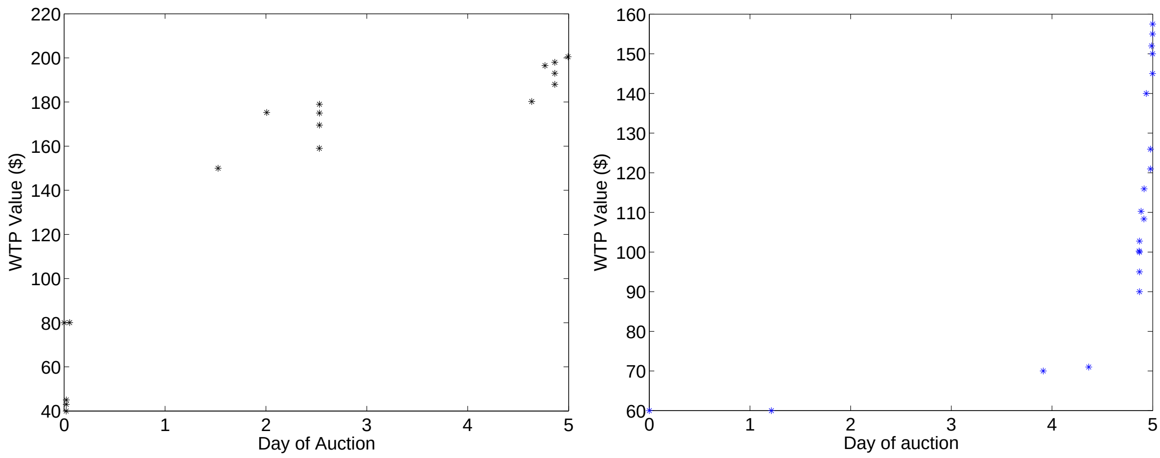

Figure 2: Profile Plots for two Palm M515 PDA auctions (The left plot corresponds to the auction in Figure 1).

The plot on the left describes the auction from Figure 1. For this auction, it can be seen that after the bid of $175.25 was submitted on day 2, three lower bids followed (of $159, $169.55, and $175). The reason for this is the proxy bidding mechanism, where the WTP of $175.25 is not displayed during the live auction. Figure 2 also gives information about the intensity of bidding over time. For example, in the right panel we see an auction that had very little or hardly any activity at the beginning of the auction, followed by very intense bidding towards the end of the auction. In comparison, the bidding activity for the auction in the left panel had a very strong start, then a spurt of bids on day 2, and a final spurt at the end of the auction.

We can integrate more information into the profile plot, depending on the research question at hand. For example, we may be interested in the final price as a function of the values that bidders saw during the live auction (rather than the WTP values). Alternatively, we could be interested in the relation between the WTP values and the values that were seen during the auction. Due to the proxy bidding mechanism that eBay uses, the WTP values are undisclosed during the auction, and therefore they can (and are very likely to) be different from the values that are displayed in the live auction, which we call *live bid values*.

To learn about the relation between the WTP values and the live bid values, we reconstruct the live auction bid values from the bid history by using a function that is based on the principles of the

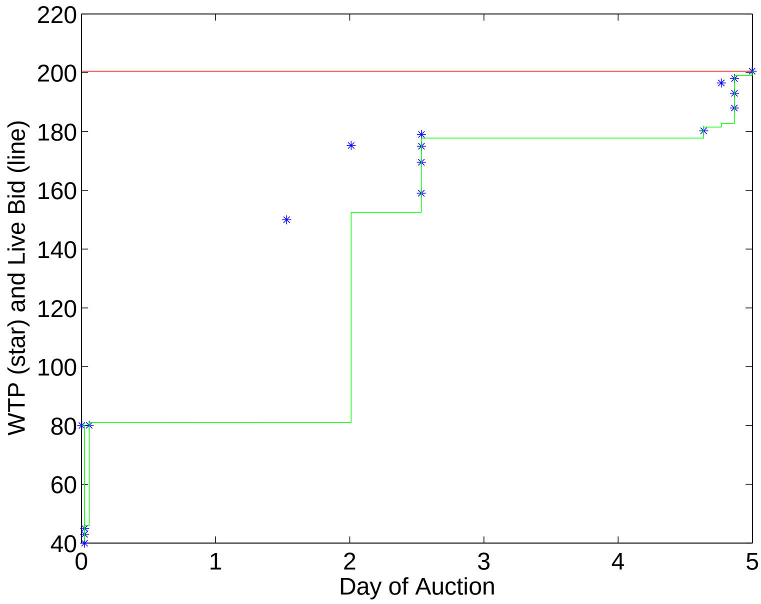

Figure 3: WTP values (blue stars) and live-bid values (green line) for a single auction. The horizontal red line (at $200.50) displays the closing price.

proxy-bidding mechanism and the increment rules that eBay uses (`http://pages.ebay.com/help/basics/g-bid-increment.html`). Figure 3 displays both types of values (WTP and live bids) and the closing price for the same Palm M515 PDA described in Figures 1 and 2 (left panel). The step-function describing the live bid values is always below the WTP values. This follows eBay's guarantee not to pay more than an increment above the highest bid. The graph shows the immediate effect of the $175.25 bid, of increasing the live bid value by an increment over the second highest WTP (from $81.05 to $152.5). However, since the bidders participating in the auction only saw the value of $152.5, it explains the arrival of the next three lower bids of $159, $169.55, and $175.

3.2 Profile Plot Variations: Integrating Additional Information

We can use color and other features to incorporate additional information into the profile plot. The type of information to be incorporated depends on the research question at hand. For example, researchers have been interested in examining bidding strategies. Various authors have observed that the number

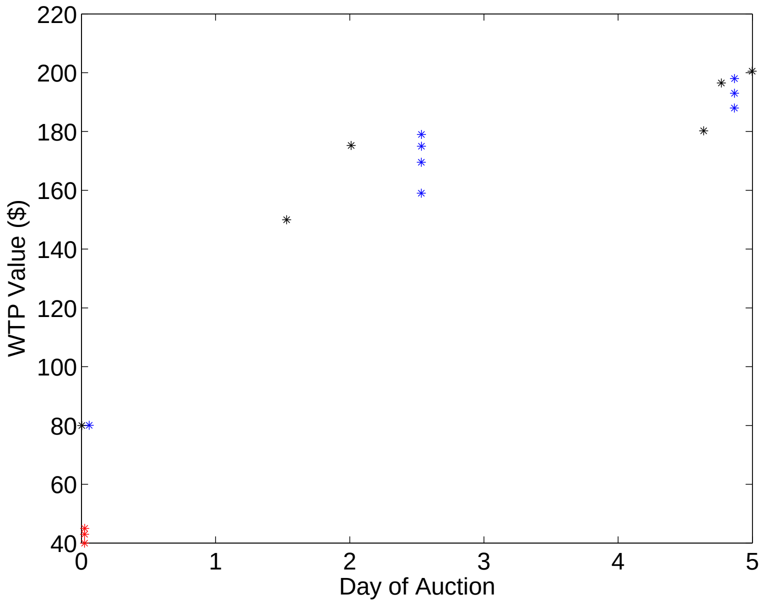

Figure 4: Profile Plot for Palm M515 PDA, with colors representing different bidders. Black denotes bids of users who placed single bids.

of bidders on eBay is usually much smaller than the number of bids placed (Bajari & Hortacsu, 2003), i.e., a few bidders submit multiple bids on the same item within the same auction. This indicates that the bids placed at each time point are not the true Willingness-to-pay values, since otherwise a bidder would not have revised his/her bid over and over again! It is therefore important to be able to visualize the behavior of different bidders, by being able to identify bids that belong to the same bidder. One option is to use different colors and/or shapes to denote different bidders. This is illustrated in Figure 4, in which an auction profile of the same Palm PDA is plotted with the addition of colors. Black is used to represent single-time bids where the user did not place any other bids. In this auction there were 8 bidders, with 2 of them placing multiple bids (represented by blue and red in the plot). These two persistent bidders placed 9 of the 17 bids. It is interesting to notice that one of the persistent bidders placed only bids at the very beginning, while the other seems to have monitored the auction and placed bids throughout its duration. The winning bid came from a single time bidder. These three types of bidding behaviors have been reported in online auctions research and are often classified as *Evaluators, Participators,* and *Opportunists* (Bapna et al., 2003). A useful addition to the bidder-specific profile plot is to integrate additional statistics on the prominent bidders. This can be implemented through a legend or by hovering over a point that corresponds to that bidder. The additional information can be taken from the same bid history, such as the bidder rating or ID. A more complicated task is to extract information on the bidder from a relational database that includes other auctions that this bidder participated in. An example of a useful statistic of this sort would be the proportion of winnings from all the auctions that the user participated in.

Another useful variation is to use color and/or shape on a profile plot to code user ratings. Bajari & Hortacsu (2003) found that experts tend to bid late in the auction relative to non-experts. Furthermore, Ockenfels & Roth (2002) posit that experienced bidders will tend to place only a single bid during the last minute of the auction. eBay bid histories also include the ratings of the users which are typically used as a measure of expertise. If this rating indeed measures expertise, then we would expect to see bids towards the end of the auction coming from bidders with high ratings, and those bids will tend to be single bids. On Figure 5 we use square size to represent the bidder rating for each bid submitted, and maintain color to represent separate bidders. If we disregard multiple bids by the same bidder, this plot shows when high-rated bidders place bids relative to low-rated bidders. It can be seen that in this auction the two persistent bidders (red and blue) have very low rating, whereas the higher rated (more experienced) bidders tended to place single bids.

In conclusion, the profile plot is easily adaptable to different research questions. With some imagination, many factors of interest can be integrated into it without clutter.

3.3 Profile Plots for Multiple Auctions

Next, we integrate the information from multiple auctions for the same item. We collected data from 10 auctions on the same item (Palm M515 PDA), each lasting 7 days, and starting between March and June, 2003. Figure 6 combines the bids from the 10 auctions. A graph of this type reveals several useful pieces of information about the auction:

- The intensity of bidding changes over time: There are two dense clusters of WTP values (blue stars) at the beginning (days 0-1) and especially at the end (day 7), while the middle of the

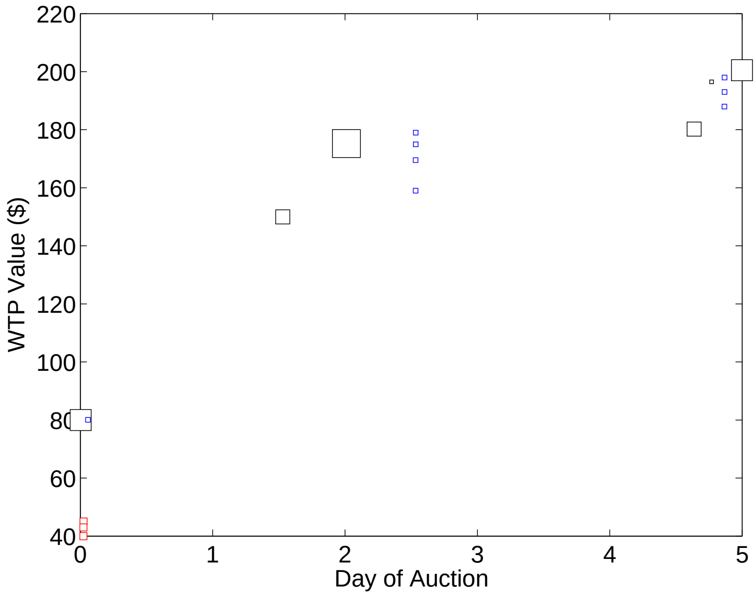

Figure 5: Profile Plot for Palm M515 PDA, with color representing bidder (black denotes single time bidders) and square size representing bidder rating. Bigger squares represent higher rating.

auctions experiences much lower bidding activity.

- The closing prices (denoted by horizontal red lines) vary between \$230-\$280, with \$280 being exceptionally high.

- Many WTP values were placed above the closing prices of other items. These are the stars in the "red zone" of the graph. This means that the valuation for this item is highly variable: there are many people who are willing to pay substantially higher prices than others!

This type of plot is useful for displaying several auctions, but it does not scale up well. Especially when conditions such as beginning price and length of auction vary, the profile chart becomes too cluttered and it is hard or impossible to track single auctions on it. Figure 7 illustrates this point by plotting the profiles of 158 7-day auctions for a Palm PDAs collected between 3-6/2003. Note the added clutter due to varying starting prices. Some characteristics of these auctions can be seen even on this profile plot:

- Activity levels are much higher on the first and last days of the auctions.

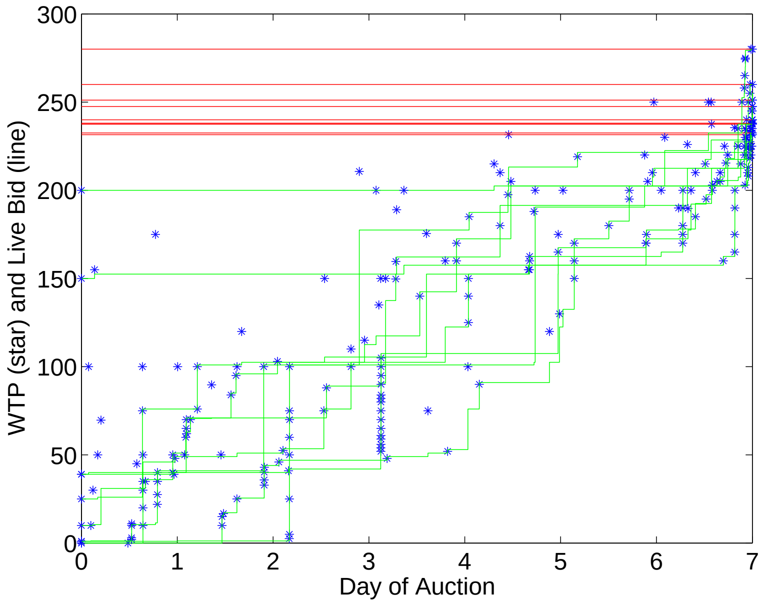

Figure 6: Profile plots for ten 7-day Palm M515 auctions.

- The closing prices (denoted by red lines) vary between approximately \$175-\$280 with the majority of auctions closing at around \$230.

- There are quite a few high red-zone bids that are placed before the last day.

If the auctions have different duration, then the profile chart is even less appealing for displaying too many auctions. Figure 8 describes the 158 7-day auctions from Figure 7 plus additional hundreds of 3-day, 5-day, and 10-day auctions for the same item. From our experience, profile plots are extremely useful for describing a single or several (< 30) auctions. The usefulness of this plot is enhanced greatly by plotting auctions that have similar starting prices, that have the same duration (eg, 7 days), and that take place in a short time period of each other. In general, any factor that is known to affect the profile should be used to separate auctions into separate profile plots.

4 Summarizing Bid-Histories

In order to learn about the characteristics of bid profiles for a certain item, bidders would ideally make use of historical data on closed auctions of the item of interest. Browsing through eBay's "bid history"

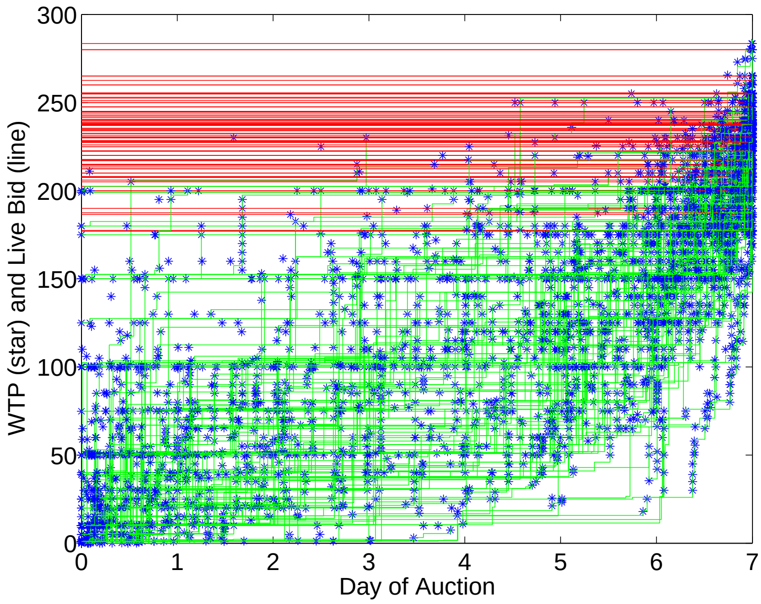

Figure 7: Bid profiles for 158 7-day Palm Pilot 515M auctions.

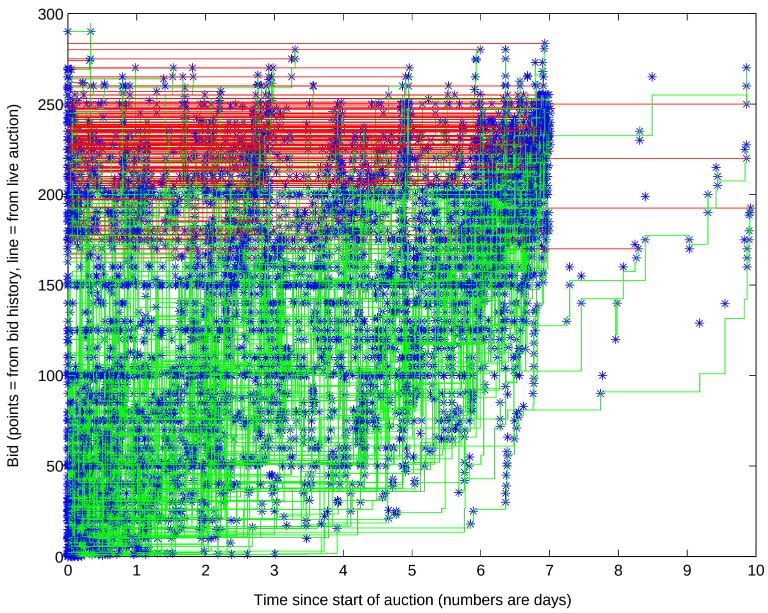

Figure 8: Bid profiles for 476 Palm Pilot M515 auctions with various durations.

pages one auction at a time can be overwhelming (since there are several million auctions taking place on eBay every day), and it is also hard to absorb the information one one single page due to the special structure of the HTML pages. Moreover, aside from an abundance of data, information is organized in a misleading way, since it is sorted by WTP values rather than chronological order.

Web tools that are aimed at supporting bidders' efforts, such as Andale.com or Hammertap.com, supply the user with aggregated information on historic (closed) auctions from eBay.com. They typically give the average selling price and the number of bids. In other words, they aggregate WTP values over time and over auctions. From graphs such as Figure 4, which display the entire WTP profile for multiple auctions, it is clear that important information is lost by such aggregation. On the other hand, as the number of auctions increases and the number of bids per auction increases, looking at the entire individual bid profiles (of both the WTP values and live bid values), might also be overwhelming.

The question is how to summarize the entire information on multiple auctions for a certain item without losing valuable information. Instead of aggregating bid values of an entire auction, we suggest to aggregate over certain time-periods within the auction so that these time intervals are affected by the bidding intensity during different periods of the auction. This intensity-dependent aggregation is described next.

4.1 Aggregating Bids Interactively

From empirical research on online auction data it is known that the bid intensity changes throughout the duration of auctions. Terms such as "last-minute bidding" or "sniping" (Roth & Ockenfels, 2002; Bajari & Hortacsu, 2003) describe the phenomenon that towards the end of second-bid online auctions there tends to be high bidding activity. In contrast, bidding is usually sparse during the middle of the auction, while bidding intensity at the start of an auction appears to vary across different items (Jank & Shmueli, 2004). Shmueli et al. (2004) developed a three-phase parametric model for the bid arrival process and showed that it can capture the bid arrival process at eBay well. Thus, an optimal time-aggregation would take into account bidding intensity, such that intense periods would be aggregated only over very short periods and less intense periods would be aggregated over longer time periods.

Since we are aggregating over multiple auctions for the same item, we rely on the user's visual ability to account for the bidding intensity in the following way: In order to find a good balance between over- and under-aggregation in time, we suggest *STAT-zoom*, a hierarchical interactive aggregation approach. This approach is more statistically advanced than techniques suggested in the context of interactivity. It has the flavor of Automatic Selection Aggregation (Eick, 2000), but it is used for continuous data rather than categorical data. In automatic aggregation, statistics are automatically recalculated for a selection of the data chosen by the user. The selection is typically a category (eg unmarried females). Thus choosing a selection of a bar chart will automatically give the statistics for the chosen selection. In our case the time scale is continuous and we treat it as a hierarchy of categories. For example, the first hierarchy could be days, then within days we have hours, then minutes, etc. The idea is not just to show, but also to *actively compute summary statistics and/or display plots* at different time scales. Figure 9 describes this: On the left we have daily boxplots of the bid values. STAT-zooming-in to the last day is achieved by clicking on the last day boxplot and selecting hourly intervals. This would instantly yield the plot on the right. We can further STAT-zoom-in by clicking on a boxplot of interest and obtain immediate summarizations for the interval and time scale of interest. The depth of STAT-zooming in and out is limited only by the units of the data. Practically, this means that we can STAT-zoom-in during periods of high activity and generate statistics and plots of the bids at frequent time intervals. During quiet periods with low activity we STAT-zoom-out, and compute averages and boxplots based on longer intervals. For summarizing the bid data we chose boxplots, which have the advantage of preserving many features of the bid distribution. It can be seen, for example, that the daily bid distribution described in Figure 9 (left) is sometimes very skewed, and thus plotting the mean or variance alone would not reveal the outliers that are of special interest in this context.

The main idea behind the STAT-zoom approach is that aggregating data at fine time resolutions will be redundant in times of low bidding activity, while aggregating at coarse time resolutions will lead to information loss during times of intensive bidding activity.

A method that is similar to STAT-zoom would be to group the data into equal-size subgroups (i.e. the intervals are chosen so that the number of observations in each interval is equal), and compute the

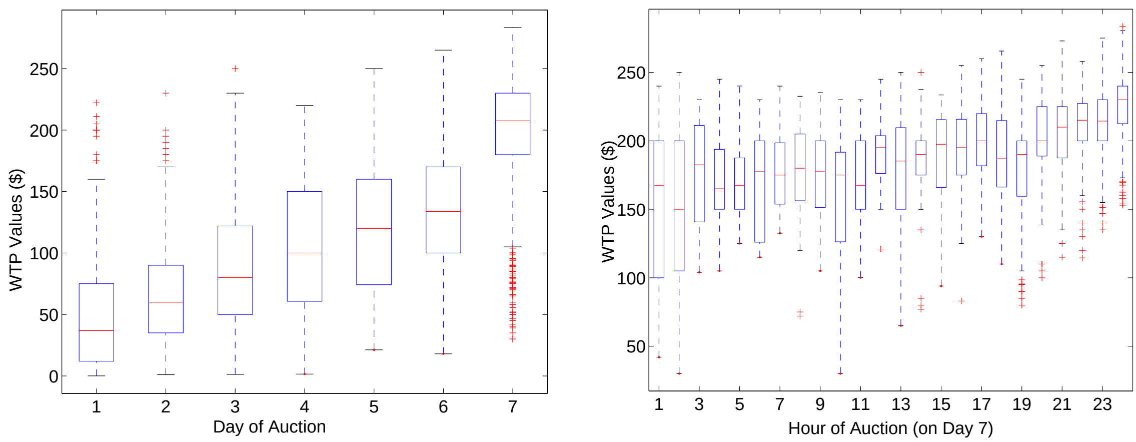

Figure 9: Aggregation of bids in 158 Palm M515 auctions: Daily boxplots (left) and last day hourly boxplots (right).

statistic/graph for each of the subgroups. This means that during low bidding activity subgroups would include bids over longer time intervals compared to high bidding activity areas. The only manipulation with this method would be to decide on the desired subgroup size. The main advantages of STAT-zoom over equal-size subgrouping are: 1. In STAT-zoom the user chooses subgroups of time intervals that are meaningful in the domain of application (such as days, or minutes), and 2. From a design and interpretation point of view, equal-size subgrouping will yield statistics/plots that are not equally spaced, whereas in STAT-zoom the intervals within a zoom level are always equal.

4.2 Displaying Bid Intensity

Although the time-aggregating boxplots account for the bidding intensity when aggregating the WTP values over time, they do not present the information on the bid intensity, that is, the amount of bidding over time. The conventional way of handling this from a statistical point of view (i.e. to describe the distribution of interest, taking into account the sample sizes), is to use boxplots with a width proportional to $\sqrt{n}$ where n is the number of aggregated bids in that boxplot (McGill et al., 1978). This method has two disadvantages in this case: First, it is useful more for the sake of comparing the boxplots (wider ones are based on more bids than narrower ones), but not for learning about the actual number of bids, which is of interest here. Second, since the display might include

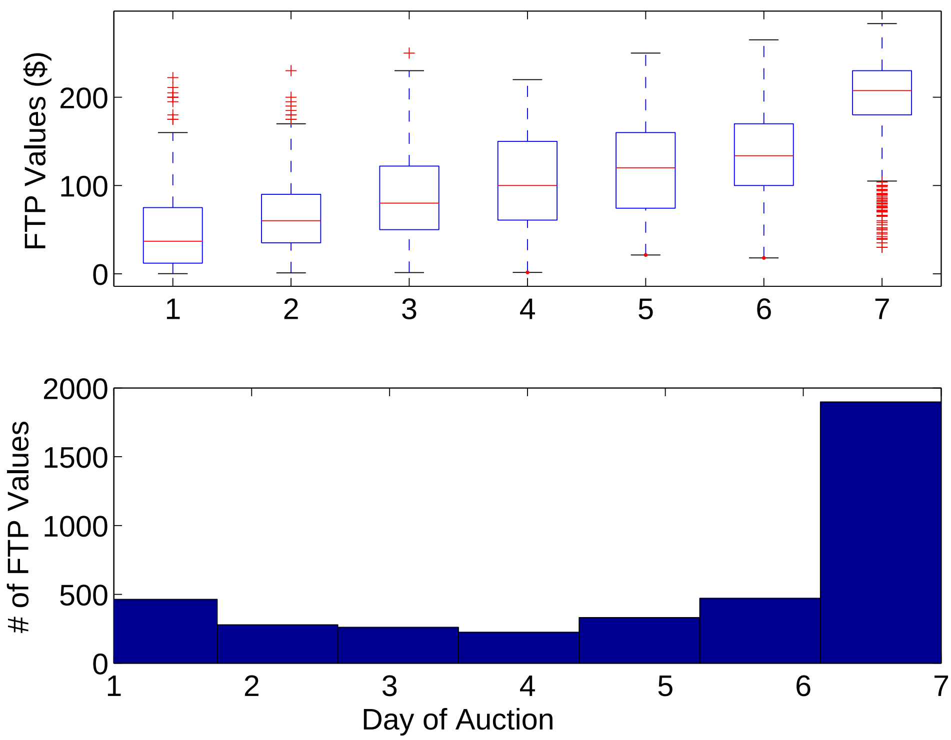

Figure 10: Daily WTP-value distribution and intensity over time for 158 Palm M515 PDA auctions.

many boxplots when refining to fine time intervals such as minutes, varying-width boxplots would cause more clutter than reveal information. We thus suggest a different enhancement to the boxplots that allows the user to browse the WTP values and bid intensity simultaneously: we add an intensity histogram at the bottom of the graph, with the bins selected to match the aggregation level used in the boxplots. The histogram can include two vertical scales to display the counts and the percentage or cumulative count/percentage. The boxplots then describe the aggregated WTP value distributions and the histogram below them reveals the number of bids in that time period. An example of a combined plot for the 158 7-day Palm M515 PDA auctions is given in Figure 10. Here we can see that the boxplots of bids during days 2-5 are based on approximately the same amounts of bids, whereas the days 1 and 6 have slightly more bids, and day 7 is based on almost 4 times the amount of bids. Combining the boxplot and intensity information we see that the multiple outliers on day 7 are not surprising, after controlling for the amount of bids placed on that day.

5 Visualizing Concurrent Auctions

Much insight can be gained from looking at concurrent auctions for the same item. Although most of the research on online auction is based on multiple auctions for the same item (or several items), only few consider the time concurrency of the different auctions in their database. For example, Zeithammer (2002) investigated the effect of the availability of multiple open auctions for the item of interest on bidding strategy and final price. Kauffman & Wood (2003a) examined the possibility of collusion through the examination of a massive dataset of concurrent auctions selling the same item. As before, we have not encountered any attempts at visualizing data from this perspective.

We suggest to start looking at concurrent auctions by creating a *Calendar of Auctions*. This is a simple visualization which displays each auction as a line that extends between its opening and closing times. On such a graph it is possible to display auctions of various durations (eg, eBay's 3,5,7, & 10 day auctions). Longer auctions are represented by longer lines. We can use different colors for different auction lengths. The second axis can be used for incorporating another factor of interest such as final price. Figure 11 displays an auction calendar for 476 auctions for the Palm M515 PDAs, where the vertical axis displays the closing price of the auction. The immediate detail that can be seen is the period over which the data were collected, extending from mid-March through June of 2003. Secondly, the auction calendar gives a sense of how many auctions were taking place on a certain day/period. It is interesting to note the "gap" around May 1, where there seem to be very few auctions taking place.

Other time related effects such as weekday/weekend effects can be examined directly from the auction calendar without the need to aggregate the data. For example, Lucking-Reiley et al. (2000) use a bar chart to describe the volume of auction closings by day-of-week. They found that more auctions tend to close on weekends relative to weekdays. Variations such as using color to represent auction length, weekend/weekday, or other classes in the data can therefore be useful for visualizing the effect of different factors.

Our second suggested visualization for concurrent auctions captures a snapshot of all the auctions in a certain time period. We call it *The Auction Scene*. The display is based on the hierarchical nature of the auction market, which is broken down to categories, sub-categories, etc. down to the item level.

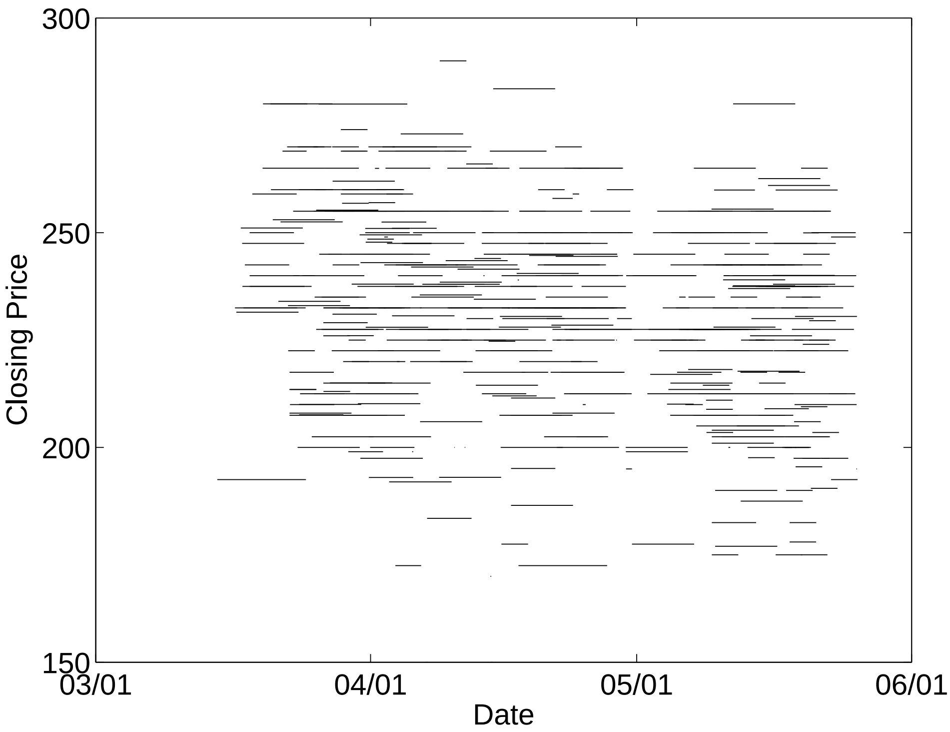

Figure 11: Auction calendar for 476 Palm M515 PDA auctions.

The visualization uses TreeMap, a space-constrained visualization of hierarchical structures designed by Shneiderman in the 1990s. TreeMap enables users to compare nodes and sub-trees even at varying depth in the tree, and help them spot patterns and exceptions. Unlike Mosaic diagrams, Treemaps are interactive and allow dynamic querying. An electronic markets application of TreeMap is the "Honeycomb" toolkit, developed by the Hive Group (`http://www.hivegroup.com/amazon_dyn.html`). It uses TreeMap to display consumer goods sold on Amazon.com.

Figure 12 displays the eBay Auction Scene for a sample of nearly 11,000 auctions that took place between August 2001 and February 2002. For further information on the data see Borle et al. (2003). The display is divided into rectangles representing categories of auctioned items (e.g., Jewelry & Watches). Each rectangle is then further divided into sub-categories (e.g., Premium wristwatches), and finally into brands (e.g., Rolex and Cartier). We can use color, size, and labels to display three variables of interest. In the figure we use color to denote seller rating (determined from feedback on previous transactions), where red denotes very low/negative rating and green very high/positive rating, and size represents the number of auctions. What is immediately apparent is that very low rated sellers are concentrated almost exclusively in the premium wristwatches sub-category. In comparison, high-rated sellers are most common in the Dell 17" monitors and Oakley sunglasses items. This can be understood if we take into account item values. Rolex watches are sold at approximately $2000, compared to other items in this sample that typically sell for less than $100. If negative seller rating is an indication of fraud, then clearly it would be worthwhile to take a risk for a $2000 watch rather than for a $50 monitor. Figure 13 explores the relation between the number of different bidders in an auction and the total number of bids in an auction (in the eBay system a bidder can place more than a single bid). Color represents the number of bids (black, green, and yellow represent few, moderate, and many bids, respectively) and size represents the number of distinct bidders. It can be seen that although the largest number of distinct bidders is in the Sports category (and especially for Golf bags), the busiest items in terms of number of bids are Oakley sunglasses and Rolex wristwatches. A plausible reason for this is that Golf bags are items of broad interest, but there is no incentive to pay more than their market value. Premium wristwatches, on the other hand, appeal to a population of bidders that is considerable smaller, but who may have a stronger interest in winning the prestigious item. Furthermore, premium watches are substantially more expensive, and therefore the price increase process is "long enough" (in the sense of bid increases) for bidders to revise their bids.

The Auction Scene maps are therefore very useful for exploring the many factors that can be measured in online auction data. They can help detect not only relations, but also outliers and unexpected patterns. Moreover, they offer a bird's-ey view of the auction scene, and thus deliver an image with is usually unavailable via standard statistical displays.

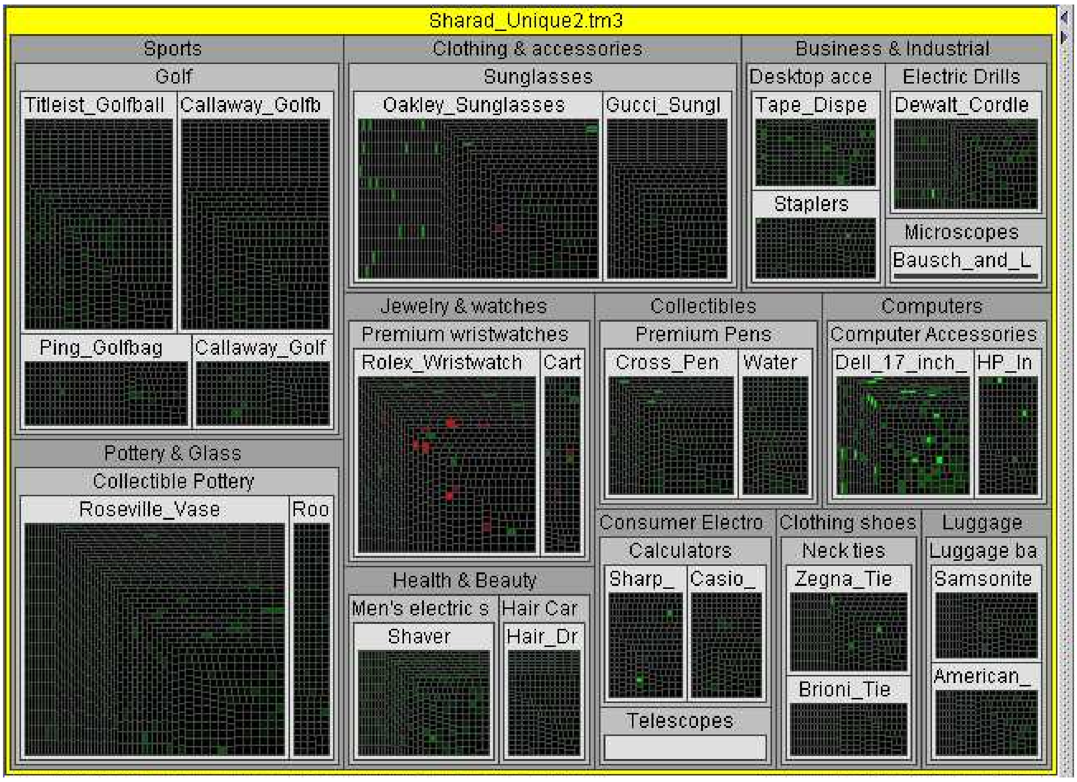

Figure 12: The eBay Auction Scene for 10078 auctions. Color represents seller rating (red=negative, black=moderate, green=high), size represents number of auctions.

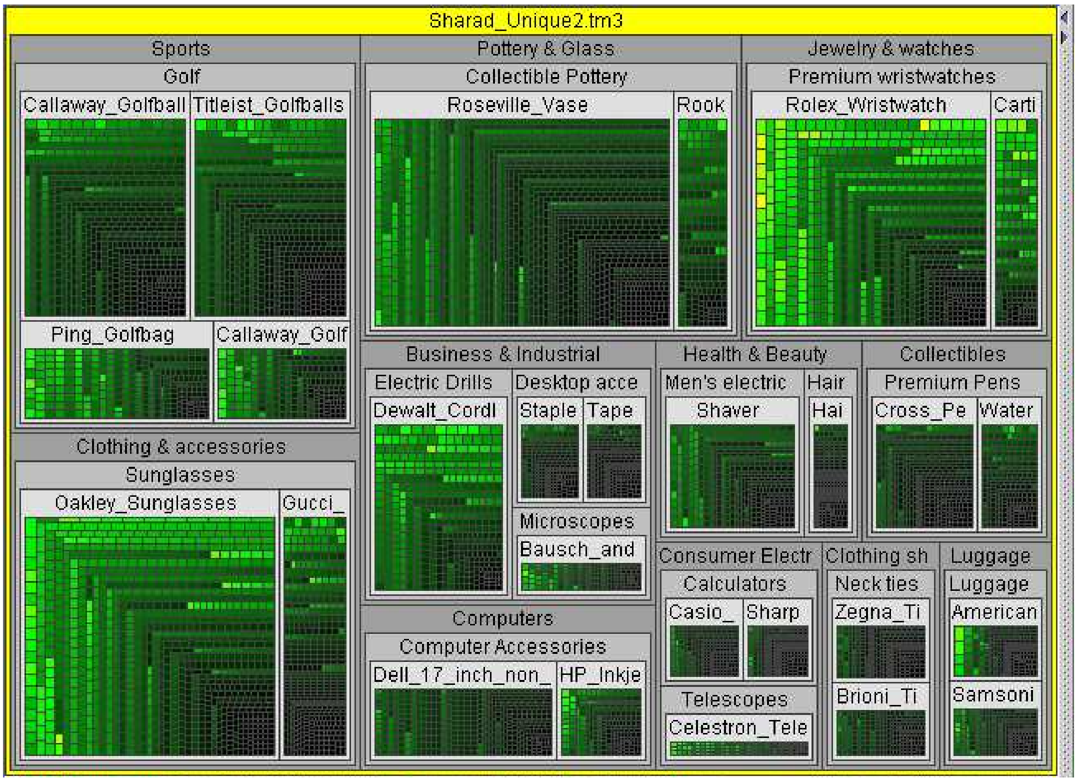

Figure 13: The eBay Auction Scene for 10078 auctions. Color represents number of bids per auction (black=low, green=moderate, yellow=high), size represents number of distinct bidders in auction

6 Future Directions

The visualizations described in this paper are meant for displaying data that have already been collected and stored. Such historic data are usually used for learning about a variety of different phenomena like bidding strategies and a seller's trustworthiness. One of the next steps is to observe and process the data in real-time. This is similar to the two phases used in control charts (in statistical quality control), where historic data are used for constructing the limits on the charts and then charts with these limits incorporated are used for monitoring real-time data. Several of the visualizations that we suggested can be used for real-time visualizations with little or no change: An auction profile can be used for monitoring an ongoing auction as long as the incoming WTP values are available. In eBay, for example, the bid history does not disclose the WTP values until the auction closes. However, by monitoring the auction using an agent, the live bids can be recorded and plotted. Since the auction duration is known at the auction start, the horizontal axis can be set accordingly. An example of a slight modification would be the calendar of auctions. In a calendar that gets updated in realtime we must show the right censoring somehow. One option is to mark an ongoing auction with a right arrow which extends to the current date. Methods based on stat-zooming require more significant modification. Finally, realtime data and their availability also call for new visualizations that would directly target their structure and the goals of monitoring them.

With respect to implementation of the proposed visualizations, most can be easily coded by using standard software. We generated all the graphs with Matlab. However, to achieve the real-time interactivity needed for STAT-zoom, a more advanced application is needed. A further complication is that the application should be able to input data with its special structure (namely, a set of unequally spaced time series of difference duration). The software package Spotfire (`www.spotfire.com`) is a tool that can handle the data structure and has many interactive options such as zooming and panning. However, since the concept of STAT-zooming is new, we have not found applications that implement it. This means that moving from one time scale to another requires, in the least, re-binning of the bid values and computing the summary statistics or graphs for the new bins. An implementation of STAT-zoom is therefore expected to be innovative.

References

[1] Bajari, P. & Hortacsu, A. (2003), "Winner's Curse,Reserve Prices and Endogenous Entry: Empirical Insights from eBay", *RAND Journal of Economics*, **34** (2), pp. 329-355.

[2] Bapna R., Goes, P., & Gupta, A. (2003), "Analysis and Design of Business-to-Consumer Online Auctions", *Management Science*, **49**, 1, pp. 85-101.

[3] Borle S., Boatwright, P., Kadane, J. B. (2003), "The Common/Private-Value Continuum: An Empirical Investigation Using eBay Online Auctions", *Working paper, Rice University.*

[4] Dellarocas, C. (2001), "Analyzing the economic efficiency of eBay-like online reputation reporting mechanisms", *Proceedings of the 3rd ACM Conference on Electronic Commerce.*

[5] Eick, S. G. (2000), "Visual Discovery and Analysis", *IEEE Transactions on Visualization and Computer Graphics*, **6** (1), pp. 44-58.

[6] Eick, S. G. & Karr, A. F. (2002), "Visual Scalability", *Journal of Computational and Graphical Statistics*, **11** (1), pp. 22-43.

[7] Jank, W., & Shmueli, G. (2003), "Dynamic Profiling of Online Auctions Using Curve Clustering", *Working Paper, Smith School of Business, University of Maryland.*

[8] Kauffman, R. J, and Wood, C. A (2003a), "Why Does Reserver Price Shilling Occur in Online Auctions?" *Proceedings of the 2003 International Conference on Electronic Commerce.*

[9] Kauffman, R. J, and Wood, C. A (2003b), "Revolutionary Research Strategies for E-Business Management: A Philosophy of Science Perspective for Research Design and Data Collection in the Age of the Internet", *MIS Research Center Working Paper #03-32, Carlson School of Management, University of Minnesota.*

[10] Linoff, G. S. & Berry, M. J. A. (2001), *Mining the Web*, Wiley and Sons.

[11] Livingston J. (2002), "How Valuable Is A Good Reputation? A Sample Selection Model Of Internet Auctions", *Workshop in Industrial Organization and Microeconomic Theory.*

[12] Lucking-Reiley, D., Bryan D., Prasad, N. & Reeves, D. (2000), "Pennies from eBay: the Determinants of Price in Online Auctions", *Working paper, University of Arizona.*

[13] McGill, R., Tukey, J. W., and Larsen, W. A. (1978), "Variations of Box Plots", *The American Statistician*, **38**, 1, pp. 12-16.

[14] Ockenfels, A. and Roth, A. E. (2002), "The Timing of Bids in Internet Auctions: Market Design, Bidder Behavior, and Artificial Agents", *Working paper #2002-33, Max Planck Institute for Research into Economic Systems, Strategic Interaction Group.*

[15] Resnick, P., & Zeckhauser, R. (2001), "Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System", *Working Paper for the NBER workshop on empirical studies of electronic commerce.*

[16] Roth, A. E. & Ockenfels, A. (2002), "Last-Minute Bidding and the Rules for Ending Second-Price Auctions: Evidence from eBay and Amazon Auctions on the Internet", *The American Economic Review*, **92**, pp. 1093-1103.

[17] Shmueli, G., Russo, R.P., & Jank, W. (2004), "Modeling Bid Arrivals in Online Auctions", *under review.*

[18] Vakrat, Y., and Seidmann, A. (2000), " Implications of the Bidders' Arrival Process on the Design of Online Auctions", *Proceedings of the 33rd Hawaii International Conference on System Sciences*,**6**, pp. 1-10.

[19] Zeithammer, R. (2002), "Sequential Auctions for Subsitituted: A Theroy of Bargain-Hunting on eBay", *Wharton working paper.*

Acknowledgement

We thank Dr. Ben Shneiderman for his useful comments and for suggesting the use of Treemaps for visualizing online auctions. We are also grateful to Sharad Borle for supplying the data for the auction scene map.