Angus Forbes MAT
Data Visualization of SPL data
Data from January 2007-April 2007

AverageNumItemsCheckedOut2.jpg
- Top 14 active categories (minus Drawings which tends to go off-chart due to usage)
- Top three 789, 781, 641 also more active then others. So Either normalize or exclude

TotalNumItemsPerHour.jpg
- Represents total volume of books
- 4 most active (drawings, music, etc.) excluded from chart to show subtle changes/

AverageNumItemsCheckedOutPe.jpg
- Same set of data but instead of volume shows averages

AverageRankOfCategoryByHour.jpg
- Ordered according to volume – most at top.
- Most active 398 folklore.
- Color could have represented degrees of activity
- Athletics 769 – most active at start and end of day..

Angus Comments:

I browsed through the data hoping to find a section of time containing significant variations in some aspect of the data. Presenting something changing is more interesting than presenting something static. However, when looking at a years' worth of data I was unable to find any dramatic shifts of any metric I investigated. Perhaps if looking at a larger time span (over years rather than weeks) I might have found more variation. If I had found variation, I would have naturally been led to a theory of why there was variation. An initial thought would be to look at news stories (do trends in news influence or parallel trends in library lending?) or to look at the popularity of the different types of television programs (do trends in TV viewing influence or parallel trends in library lending?).

Also, I merely "eye-balled" the data. Running some statistical tests on the data may have led me to find more significant results. (Developing a library/interface for interactive statistical analysis is something I am working on for my thesis).

How did you give the data form?

In order to make the visualization manageable, I limited the number of categories I was following to the most popular 10 or so categories. In a few of the slides I left of the very top categories as they pertained to CDs rather than books (the music categories). I started out making a simple chart with a simple legend so as to get a feel for the data, without worrying too much about how it looked. After plotting the raw data, I started to focus on what would be interesting and/or presentable in the space of a single chart. Since the focus of the charts was primarily about the way it looked over the data it represented (although ideally it would be equal amounts of both!) I felt free to choose a time-period that made a pleasing looking hump, and to search haphazardly through the data to try and find a period with the most variation. I tried different ranges of time periods -

a single week, a couple of years, and ultimately chose a 6-month period. After the first three charts, I started thinking a bit more about what would be a more interesting way to present the data. Since it seemed like it was the variation rather than the actual numbers that was most interesting, I created the version which tracked the ranking of the top categories. I think that this chart is the most successful since it allows the viewer to see in a single glance which "paths" have jumped to other "tracks". (Of course, you lose the detailed information about the actual number - or average number - of books in a particular category).

How were the colors chosen?

George provided me with a color palette, and I simply chose the colors which were most pleasing, without too much thought. Given more time I might have associated an icon with the different categories, or perhaps attempted to make a stronger connection between particular categories and colors.

How did you visualize the subtle changes in the larger?

Since the end result of the visualization was non-interactive, I simply made a few different charts to try to emphasize different aspects of the data within the time period. It would be interesting to be able to interactively "zoom" into a single hour, or out to a few years.

What worked what did not work, etc.?

Again, I think that the last chart (the average rank) worked the best, possible because it was attempting to do the least. By emphasizing only a single aspect (rank) and excising the actual numbers, the chart became more visually appealing and easier to read. I spent more time actually parsing the data than creating the graphics (probably about 8-12 hours writing functions to transform the binary data files into a useful form, and about 3-5 hours writing code to make the two different types of charts themselves). I usually code with an eye to re-using what I've written, and in fact some of the drawing code was put into a

project called the "Corpus Browser" (which lets a user interact with a large database containing samples of written and spoken language).