C-Flow: Visualizing Foot Traffic and Profit Data to Make Informative Decisions

Tiffany Chao, Ruofei Du, Jonathan Gluck, Hitesh Midadsani, Kent Wills
Department of Computer Science
University of Maryland, College Park
{tchao, ruofei, jdg, hitmai, rkw14}@cs.umd.edu

ABSTRACT
“Business intelligence”, a term first coined in 1958 [11], has been evolving for almost 60 years. With recent advancements in data collection [15, 2, 8, 13], companies are looking for new ways to gain insights in order to generate revenue. In this paper, we present C-Flow, a novel system that allows mall and store owners to reason about consumer foot traffic with respect to store profits. This is done through effective use of an integrated heatmap and overlaid footpath on a mall map. Through two rounds of usability studies, we also provide evidence that rendering abstract consumer paths may be more useful than rendering raw path data.

Author Keywords
Information Visualization; Data Mapping; Indoor Visualization; Business; Usability Test

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION
Thousands of shoppers pass through shopping malls every day [14]. With the advent of dynamic ads in the form of electronic screens (televisions, projections, touch screens, etc.), there is an increasing demand to select consumer-relevant, time-sensitive advertisements. With this demand, old technologies are being used in new ways. Closed Circuit Television (CCTV) screens, once used to monitor consumer behavior and cell phones that were once used as communication devices are now being used to track consumer movements. [13, 2].

With the increased variety of consumer data collection there are many new opportunities for visual analytics. While we were not able to locate specific works for malls, more generally, there has been some visual work of geo-spatial data with overlaid temporal heatmaps and data points. Our design and analysis of our tool, C-Flow, provides evidence that intelligent aggregation of this data may lead to more understandable and actionable business decisions.

C-Flow, a “proof-of-concept” web application, allows mall owners and supervisors to reason about temporal consumer path data. The tool incorporates the use of a heatmap, geo-spatial map, and consumer flows with modern libraries including D3.js and Kinetic.js for responsive interaction.

Mall owners can use C-Flow to visualize foot traffic data overlaid on their mall floor plan, quickly spot profit fluctuations amongst their stores for many time periods, and locate high impact areas in order to variably rate their advertisement space throughout the mall. Even store owners can use C-Flow to make informed decisions on advertisement locations, allowing them to reach their target audience. Though C-Flow is still a prototype, we firmly believe that it could boost the development of business intelligence for mall owners.

PREVIOUS WORK
Much previous work visualizes individual data points [3, 6, 4, 9]. While these visualizations provide a global understanding, it is hard to gain insight that would be helpful to a business. We instead pre-aggregate data in order to allow the user of C-Flow to quickly understand the data in order to make timely actionable business decisions.

Other previous works [12, 16, 17, 7] make effort to reduce the screen space devoted to visualizing edges in a connected network through line simplification via clustering and minimizing crossing edges.

Instead of using clustering algorithms to reduce the lines on our visualization and showing every individual data point in our data, we propose a smart aggregation and simplification of data prior to display. This simplification aims to allow the user to make more sense of the data without losing critical details that would lead to insights. To explain this simply, much in the way that [1] simplifies map data to route paths, we simplified the individual x,y coordinates with stores that were visited.

To the best of our knowledge, we are the first to visualize foot traffic data through the use of abstract connections in an indoor setting.

DATA GENERATION
For most projects MicroStrategy creates in-house synthetic data in order to test their products before releasing them to
their customers. Synthetic data is typically used in the business intelligence industry if the company is offering a tool that will be used in-house. Once you start working with real data, you have to deal with privacy concerns and interaction with the data can potentially reveal trade secrets. Furthermore, the issue of tracking consumer habits in malls [5] has become a point of contention amongst consumers.

In order to quickly develop a tool without having to get our hands on real data, MicroStrategy provided initial suggestions on data generation and some randomized data. We used these insights to create data to fit the needs of our tool. The synthetic mall data for C-Flow consists of four parts: a floor plan, shopper visiting patterns, number of shoppers in a store at a given time, and the amount of profit in a store at a given time.

In order to quickly develop a tool without having to get our hands on real data, MicroStrategy provided initial suggestions on data generation and some randomized data. We used these insights to create data to fit the needs of our tool. The synthetic mall data for C-Flow consists of four parts: a floor plan, shopper visiting patterns, number of shoppers in a store at a given time, and the amount of profit in a store at a given time.

In order to provide a realistic experience for our tool we selected Tyson’s Galleria in McLean, VA as our floor plan. Data was extracted from the website, layered in Adobe Illustrator and stores were individually exported as support vector graphic (SVG) images. In order to spend the maximum amount of time on our visualization, we did not automate the process of converting an AutoCAD file to an SVG graphic. While we did not further any efforts in this area, MicroStrategy is actively exploring this automation.

Two types of foot path data were created in the development of our tool, the raw (x,y) consumer path coordinates and a simplification of the data that only included the stores that a shopper visited. The raw (x,y) consumer data consisted of sets of points for each shopper. The points for each shopper were arranged in sequential order so that, for each time step (defined to be one second), each position represents a shopper’s location at that second. Furthermore, each shopper was instantiated with randomized preferences for store categories. The simplified foot path data was generated for 1,000 shoppers in a year time-frame with a limit of visiting all nine stores in one day. Stores were randomly chosen on a multiplicative weighting scale and number of stores they visited were randomly decided from a value of one to the maximum, nine. Restrictions placed on shoppers included not being able to visit a store outside of the given time frame, 9AM to 11PM and shoppers did not decide to go back into a store after leaving it in one time step.

Customer data was generated off of the visiting patterns described above and Profit data was randomly determined, giving each shopper present in the store at a given time-period the ability to make a purchase between $20 and $200.

The map aggregates all of the data for profit and shoppers, and the heat map shows the hourly separation of values. Both derive their data for a given filtered time period selected in the control panel. The goal behind our synthetic data generation was to create realistic synthetic data in order to potentially gain insights in areas where our tool is especially effective.

C-FLOW COMPONENTS
In this section we will describe the different components of C-Flow. More specifically we will discuss the organization of the back-end and front end components: geo-spatial map, flows, control panel and heatmap.

Overview
We can see here in Fig.1 the entire C-Flow interface. C-Flow is implemented as a web application with a Node.js backend serving the data. The tool is made up of two views and a control panel. The first view is the floor plan view, and the second is the heat map view.

Back-end
C-Flow uses a remotely hosted RESTful back-end implemented in Node.js, a modern server-side implementation in JavaScript for real-time web applications. The REST request to the server supports ten parameters, which are described as follows.

- **type**
  Indicates whether we are requesting store data or path data.

- **name**
  The name of the store we are querying.

- **value**
  Indicates whether we are requesting customer or profit data.

- **timeType (hour, day, month)**
  Indicates how to aggregate results.

- **time parameters (hour1, hour2, day1, day2, month1, month2)**
  Indicates time ranges for filtering the data specified by the time-based filters

Using these parameters, the front-end specifies the desired data from the server with a single REST request through AJAX. The server queries the database with the parameters, aggregates them based on timeType, and then returns them in a format easily usable for the front-end.

Front-end
The front-end is HTML5/CSS/Javascript (JS). We use two different JS libraries for the two main views: Kinetic.js, an HTML5/JS canvas framework for high-speed drawing and animations, for the floor plan view; D3.js, a popular JavaScript library to manipulate documents based on data, for the heat map view. We also use jQuery for general-purpose utilities and event-based coordination, and reskin jQuery UI’s range slider for the sliders in the control panel.

Control Panel
The rightmost slice is the control panel. It is the main interactive component, though it contains little information itself. From top to bottom, the controls are as follows:

- **Data type toggle**
  Toggles whether the heat map displays customer or profit data, and whether stores are colored based on profit or customer data.

- **Time selector**
  Changes time granularity (hour, day, month). Filters heat map cells and floor plan footpaths.

- **Time filters**
  Filters heat map cells and floor plan footpaths by hour, day, or month.

- **Store filters**
  Toggles whether the heat map displays data for a given store. Effectively adds or removes heat map rows.

Map View
The bulk of the interface consists of the floor plan of the mall, as well as a display of the consumer flows between stores. In Figure 2, we can see that Tyson’s Corner consists of two rows of stores. Each store has a label (its name) and a color (describing either profit level or foot traffic popularity, depending on the data toggle mentioned previously).

The grey curves with an %50 opacity represent consumer flows between stores. Figure 2 shows the latest version, which employs a simpler footpath model (motivated largely by the results from the first usability study). Initial prototypes rendered raw footpath data, which are almost random paths distributed in the entire mall. When hovering over a particular store, only the consumer flows out of that store would be displayed.

C-Flow Rendering
To generate the curve of the path flow, we used a Weighted Bézier Curve Model. Suppose there are N stores, for each store i, the leftmost and the rightmost point in the entrance

![Figure 3. Control points are distributed along the store entrance by the traffic flow.](image)
is labeled as $L_i$ to $R_i$; the number of shoppers from store $i$ to $j$ is noted as $f_{i,j}$, where $i, j \in [1, N]$. $f_{i,j}$ is dynamically loaded from the server by filtering on the control panel. We define the entrance vector $E_i$ as $\overrightarrow{L_i}$ to $\overrightarrow{R_i}$, i.e. $E_i = \overrightarrow{L_iR_i}$.

Finally, the total number of shoppers flowing out of store $i$ is defined by $F_i = \sum_j f_{i,j}$. Afterwards, we calculate the number of shoppers $\lambda$ corresponding to a unit pixel length:

\[
\lambda = \min_i \left\{ \frac{F_i}{|E_i|} \right\} = \min_i \left\{ \frac{\sum_j f_{i,j}}{|L_iR_i|} \right\}
\]

As illustrated in Figure 3, for each flow from store $i$ to $j$, the position of the flow $p_{i,j}$ is given by:

\[
p_{i,j} = \sum_{k=1}^{j-1} p_{k,j} + \lambda \cdot f(i,j)
\]

\[
M_X = \frac{1}{2} (A_X + B_X) \pm \frac{1}{2} \lambda \cdot f_{i,j}
\]

For every three continuous control points, a quadratic Bézier curve[10] is applied for C-Flow. Given points $P_0, P_1$ and $P_2$, we have

\[
B(\beta) = (1 - \beta)^2 P_0 + 2(1 - \beta)P_1 + \beta^2 P_2
\]

Finally, in order to avoid overlapping as much as possible, we propose to sort the consumer flows according to the angle $\phi$ between the store entrance $E_i$ and the simplified flow vector $\overrightarrow{M_iM_j}$ where

\[
\phi_{i,j} = \arccos \left( \frac{E_i \cdot M_iM_j}{|E_i| \cdot |M_iM_j|} \right)
\]

As illustrated in Fig.5. The layout of traffic flow is impressively improved after the sorting algorithm.

Heat Map

The bottom section is a heat map that displays profit or consumer data over time. Each row represents a store and each column represents a unit of time binned by hour, day, or month. A four-color legend sits off to the right.\(^1\)

As is illustrated in Figure 6, hovering over a store label highlights that row, fading all other rows and triggers a hover event for the store on the map. Hovering over the legend focuses on cells that fall into the hovered bucket, fading out all other cells. Hovering over a time label focuses on that column, fading all other columns. Hovering over a cell highlights that cell and fades out others and triggers a tooltip that shows the actual value of the cell. Additionally, hovering over a store in the C-Flow view, described above, highlights the relevant heat map row allowing coordination between the two views.

EVALUATION

We designed a two-round usability: the first one for the initial prototype (with raw footpaths) and the other one for the revised prototype (with abstract C-Flows). 11 volunteers were drawn from friends and family, ages ranged from 19 to 55, whose professions ranged from business student to industry computer scientist. None of them were familiar with the interface. The first six were conducted across three days. The latter five were conducted across two.

Procedure

The usability study began with a brief overview: first of the general problem (ad placement), then of the specific tasks to

\(^1\)Technically, there are five colors: white, yellow, gray, light green, dark green. The legend only describes non-zero buckets, however.
come (3 tasks, about 5 minutes each; then a brief question-naire, then time for general comments). They were told they could stop at any time. We then asked them to think aloud, and quietly took notes.

The three tasks were arranged from simple to complex, straightforward to open-ended, and designed to exercise the most important aspects of the prototype available at that time. We chose not to introduce the tool, hoping to gain some insight from the learning process. If participants took less time than allotted for any particular phase, we allowed them to proceed on at their natural pace.

**Task 1: basic navigation**

Find Safeway.
(Prompt them if they flounder: it’s towards the right.)

Do people visit Safeway?
Goal: path identification. Is the term ”path” clear?

Where’s the first store most people visit when they leave Safeway?
Goal: path direction identification.

**Task 2: detail**

Find Macy’s. Do a lot of people visit at 11:00?
(Prompt them if they linger on the map.)
Goal: heat map legibility. Heat map use.

Do more people visit Macy’s at 11:00 or 3:00?
Goal: heat map legibility. Value comparison.

**Task 3: ad placement (putting it all together)**

Pretend you own Macy’s. Suppose you want to place an ad for women’s perfume. There are two open spots: one at point P1 (center of the map, where the paths intersect; point to it), at time 4:00 PM; another at point P2 (entrance by Apple, point to it), at time 10:00 AM. Which spot would be better? Why?
Goal: data synthesis.

The first task evaluates C-Flow’s most basic visual representations: the floor plan and the footpaths. The second task evaluates the heat map: what it displays, and how it’s interpreted. The third task examines whether both can be meaningfully combined, testing how well the interface supports higher-level reasoning.

The concluding survey given, designed to gauge subjective reactions, follows.

**Basic Demographics**

Gender? M / F
Age? —

Computer use:
Once or twice a month Once or twice a week
Several times per week Several times per day

Please indicate the degree to which you agree with the following statements.

1. The interface was easy to navigate.
2. It was easy to figure out where people went.
3. It was easy to figure out when people went to various stores.
4. Buttons, text, and so on were easy to read.
5. The interface was aesthetically pleasing.

**Results: Usability Study #1**

Some elements of the first C-Flow prototype performed well; many others underperformed.

**Results: Survey**

The averaged results of the usability test survey may be seen in fig. 7. From this we may conclude that our subjects seemed to receive the overall aesthetics well. They found the interface simple to navigate, and buttons were easy to discern. On the negative side, our subjects seemed to find it difficult to understand consumer movement. This is particularly problematic, as it is one of the key components of C-Flow.

**Task One**

Task one, as previously mentioned, existed to test subjects’ understanding of the basic layout of C-Flow. The starting state of C-Flow was to display profit data; however, for this task, foot traffic data was far more useful. Many users did not notice the toggle between foot traffic and profit data in the control panel. Users tended not to use the path implementations to solve this problem, seeming to become confused when attempting to grasp their meaning. Overall, users seemed to identify stores in both views of C-Flow quickly and without trouble.

Subjects completed this task with ease; however, we had expected them to attempt to use the paths to solve problems of shopper location. Instead they resorted to using the heat map in foot traffic mode. We did not expect the heat map to be so quickly understood; however, users very quickly grasped the inherent time series data. This was encouraging. While users seemed to avoid the footpath data, they were drawn to the heat map.

**Task Two**
Task two existed to test subjects’ ability to discern details from the footpath and profit data. For this test we fully expected the subjects to utilize the heat map, as it comprised our finest granularity view. Users were very successful at task two solving all questions with ease.

We received several comments that our color palette was too subtle, making it difficult to differentiate different values. Some subjects suggested that we might adopt a more traditional red/green gradient. Subjects requested additional time controls for the paths, allowing users to filter out paths that occurred before and after a time.

Task Three

Task three was meant to test subjects’ ability to perform the actual task for which C-Flow was designed. Using all tools available, subjects were told to choose one of two advertisement placements. Again for this task we saw subjects drawn to the heat map over the raw footpath map. For this task; however, the raw footpath data was more necessary. Users had difficulty understanding the paths, stating that they were too cluttered and confusing.

All subjects eventually came to the same conclusion for ad placement, placing the ad near a large cluster of intersecting paths near the middle of the map. Several cited an uptick in nearby store foot traffic in the heat map for their choice. Other’s chose this point because it was a more central location with more paths on the raw path map. Feedback for this task suggested that the current implementation of paths was more confusing than helpful.
Subjects used the time following the three tasks to provide suggestions and helpful feedback. Almost universally, subjects requested path data be drawn directionally. Several users noticed a glitch in the initial implementation of C-Flow which caused path colors to be inconsistent. Subjects also universally reported that the current implementation of paths was overwhelming. From this feedback we decided to work on a second iteration of C-Flow incorporating several of the ideas suggested by the usability test subjects. Additionally, for this new version of C-Flow we sought out different ways to visualize path data. We will discuss our new C-Flow and the numerous changes we made in the following section.

Revisions due to Usability Study
We considered the results of our usability studies carefully, taking into account both our Subjects’ comments, and their usage patterns. We found that there were several areas that required improvement in our design. Each of our two usability studies furthered the design of C-Flow: leading to what, we believe, is a compelling visualization of consumer traffic flow between stores.

Insights from Round 1
Our first round of usability studies alerted us to three major issues. These issues were, readability of data, raw path incomprehensibility, and heat map use. The first two were negative discoveries, while the third was a positive discovery.

Readability of Data
Multiple subjects alerted us to elements in our color scheme being inconsistent and, worse, that several elements were mistakenly consistent. In particular, for the first version of C-Flow, the colors of our paths data were drawn from the same color palette as our heat map. Even though the paths and the heat map shared a color palette, there was no intended equation in values. This was a point of confusion for many subjects.

Additionally, several subjects reported that it was difficult to see the labels for stores in our data. The store labels blended in with the background colors of the stores.

In order to address these issues we decided to add a neutral grey color to our palette to indicate a lack of numerical data. We used this grey to draw our paths in the second version of C-Flow. Additionally, we use this grey as a transparent backing for the labels on our stores.

Raw Path Incomprehensibility
Universally, our subjects avoided the raw path data we presented in C-Flow. The main reasons seemed to be the fact that the data was too cluttered, and that the data was directionless. This made it impossible to tell if a shopper was arriving or departing a store. We expected users to make heavy use of these paths for Tasks one and three of our usability study, and they circumvented this element.

In order to reduce the clutter of the raw path data, we made the decision to abstract our path data down to simple curves indicating to which stores shoppers travel after leaving a given store. This data is still useful for understanding the flow of shoppers from store to store.

Heat Map Use
Our subjects seemed unable to utilize the raw footpaths presented in the original C-Flow, but they managed to solve tasks one and three despite this handicap by making use of the heat map. One example: In order to tell where most shoppers travelled after leaving safeway, users examined areas of peek Safeway traffic in the heat map, and then which locations saw an increase in foot traffic at the next time bucket. When examining the results from our usability study we were surprised at how versatile the subjects found the Heat Map, preferring it over the path view. We bore this in mind when going forward with version two of C-Flow.

Insights from Round 2
Our second round of usability studies supported the changes we made after our first round. We found that subjects began using paths to attempt to solve tasks one and two. Subjects still found issues with our path displays. Additionally, subjects found it difficult to find details in C-Flow. Lastly, subjects still showed

FUTURE WORK
There are many possible expansions to our current project, for which we envision three categories of improvements: scale, minor tweaks, and new features.

Scale
We would like to see the ability to select and filter paths based off of demographics in order to better target consumers. This requires incorporation of demographic data into our visualization, which may introduce new considerations in terms of how to show this data.

Another part of scaling would be to add larger maps and additional floor views to the visualization. Since the size of maps can be quite large and can flow outside common screen sizes, we think that implementing a tile-like visualization would help solve this issue. Basically, large maps would be split into multiple tiles. Not all tiles will be displayed on the screen, but flows from tiles towards the sides can indicate between-tile movement. These other tiles can then be viewed if needed. We would have to rework our visualization and the zooming to support tiles.

Tweaks
We would like to fix existing bugs and issues, and improve certain aesthetics. Particularly, individual cell hovering in the heat map should not hide other data cells as this creates a significant loss in data and usability. Additionally, since the visualization makes several data requests to the server, which occasionally takes some time to respond, we would like to have some loading cursors to let the user know about the request.

We would like to improve server load times by optimizing database queries and cache data. We would also like to improve coordination between the views. Also, adding titles to legends and controls, and making the legend numbering easier to understand by using fixed intervals would make the usability easier. We can also add store logos on the map can
simplify identification of stores. We would like to make path aggregation more sophisticated by using algorithms to avoid line intersection. We would like to experiment with other color schemes such as better sequential colors to see if they are effective. Additionally, we would like to try other hover behaviors.

New Features
We would like to add options to explore certain days or events such as Black Friday, Christmas Eve and others to allow businesses to target and understand those days. Also, adding specific time range filtering would be a plus. A major suggestion we have received from usability tests is to have directional paths. We can also have colored paths to signify the profit data or demographics. With directionality included in paths, the control panel can have options of selecting from bidirectional, directional, colored, or other path views. Basically, giving further customizability for path details would be helpful. We would like to add sorting features to the heat map to customize store order.

Another feature can be viewing individual person data points or nodes on the map rather than aggregated paths. This can be viewed when zooming in or in the regular view, since currently zooming doesn’t show more fine-grained details. Additionally, since the map view is the main visualization, we would like to introduce alternate visualizations and give the user the option of swapping between them. An example of an alternate visualization can be based on the visitors flow visualization for web page analytics used by Google. Additionally, we can have data import/export features.

Another important feature we would like to have is porting our visualization to mobile devices, particularly tablets such as the iPad. We designed the interface for a mouse/keyboard paradigm; however, touch gestures have new opportunities, new challenges, and perhaps design ramifications.

CONCLUSION
C-Flow is a robust means of visualizing foot traffic data and combining that data with store profit data. The changes we made between versions one and two of C-Flow were drastic; however, we believe that they were all for the better. Although we sacrificed displaying the true paths that the customers followed, we feel that the source/destination implementation we alighted upon is an elegant abstraction of that data. This tool helps solve a newly emerging problem in a world of dynamic advertisements, allowing users to reason about the demographics and shopping habits of customers via the paths they walk throughout a shopping center.

CREDITS
TODO: everyone put what you did here.

ACKNOWLEDGMENTS
We would like to thank Prof. Ben Shneiderman for his precious suggestions and instructive lectures. We are also grateful for the team at MicroStrategy for providing us with support and guidance throughout the course of this project. In particular, we would like to thank George Zaimes and Suhas Mahajan for their time and advice each week. We would also like to thank the participants in our usability studies for their time and honest opinions: C-Flow would not have evolved into its current form without your help.

OUR CUSTOMER
C-Flow was built with the guidance and support of a team at MicroStrategy lead by George Zaimes, a program manager, and Suhas Mahajan, a UX designer. MicroStrategy specializes in Business Intelligence, a space that aims to educate businesses by transforming and understanding the vast collections of data that they store from their daily interactions with consumers. Recently, they have been exploring effect visualizations for indoor traffic in malls. We had weekly meetings with the MicroStrategy team in order to get their advice and feedback.

REFERENCES
5. Gallagher, S. We are watching: Malls track shoppers cell phone signals to gather marketing data, 2011. http://goo.gl/bf0sLs.
