EveryWeek - A Location Visualization/Sonification of Personal Weekly Habits

RJ Marsan
University of Illinois
Urbana, Illinois
me@rj.io

ABSTRACT

Human habits are insightful patterns of an individual, but they do not tend to be well-defined. Within the context of location habits, many visualizations exist to show basic patterns and habits, but few attempt to dive into the details. EveryWeek is a visualization and sonification that explores people's location habits within the context of a week, who's goal is to provide a medium where habits can be more easily identified and contemplated, but not necessarily quantified. Particular care was taken to eliminate excess information, both to make a more aesthetically pleasing visualization that feels more personal, but also to engage the individual more, and force them to build connections themselves. The visualization attempts to shine light on a specific class of habits, location habits based around a week, particularly weekend vs. weekday habits, and time-of-day habits. We found the visualization highlighted a significant amount of our target habits, but was limited by the computational abilities of the web browser.

Categories and Subject Descriptors

[**Human-centered computing**]: Information visualization; [**Information systems**]: Clustering; [**Networks**]: Location based services

General Terms

Visualization

Keywords

Visualizations, Location Services, HTML5

1. INTRODUCTION

A habit, as defined by the American Journal of Psychology, is, "from the standpoint of psychology, is a more or less fixed way of thinking, willing, or feeling acquired through previous repetition of a mental experience" [7]. Habits, by definition, are not something that are consistently paied attention to. They are patterns of behavior that may be fixed, but also may fluxuate, skip cycles, and so on. If habits are considered signals, there can be plenty of noise.

Visualization is a broad term, and for EveryWeek, the best context would be Michael Friendly's 2009 description of both statistical graphics and thematic cartography, who "share the common goals of visual representation for exploration and discovery"[10]. Visualizations present the ideal medium to present such complex and arbitrary information, as they can display highly complex and dense information, while still allowing the user to readily identify patterns and turn them back into information about their habits.

1.0.1 Rational

The rise of portable computing devices has greatly expanded the amount of data gathered about an individual . Within this mass amount of data is significant information about that individual. Among the many things gathered by a personal computing device, location data may be the most unique, as there were virtually no other ways of collecting this information previously

The choice to use a week as the central aspect for the visualization was fairly straightforward: The data gathered by modern smartphones isn't terribly precise, as it must balance battery life with location precision. At the moment, that means the data gathered wasn't precise enough to pinpoint where someone was in a building, but more importantly, the location events were fairly sparse. Location events might be logged every 5 to 10 minutes, but the location 'fixes' were done much more sporadically. On a daily scale, there was too much noise. On a longer scale, the data worked quite well. Months didn't hold much useful patterns in preliminary testing, but weeks definitely did.

Existing visualizations of location history, works like Google Latitude[3], Foursquare[2], and others utilize location history in a very specific way, and the goals of visualizing them back to the user are not for finding complex patterns, but to give quick summaries, or to gather specific bits of information on one's past. No visualization has been made that focuses on the complex habits a person may have on the weekly/daily/monthy scale. This is the space EveryWeek attempts to fill.

1.0.2 Goals

With all these concepts in mind, EveryWeek is a visualization and sonification that allows users to explore their weekly

location habits. Contemplation and discovery are key goals, as opposed to summarizing information.

To accomplish these goals, several concedements had to be made. Overwhelming the user with information would dillute the overall goal of giving users insights, and encouraging them to contemplate their location habits. This lead to a minimal approach to the visualization that also took on a large aesthetic component, placing the user in an isolated context for them to contemplate without distraction.

Another tangential goal of EveryWeek is to demonstrate a medium for visualizations and sonifications that is not used as much as it should, a modern web browser. Many of the technologies uses in EveryWeek are not particularly new, but faster computers and browsers allows complex and intensive visualizations like this to be done. The modern web browser has it's limits, however, particularly with application to data processing.

2. AESTHETIC GOALS

EveryWeek was designed with an aesthetic component as well as the informational component described above. The role of aesthetics frequently overlaps with the practical aspects of the project, but there's a clear overall goal: provide a clean and intimiate environment for users to contemplate their habits.

2.1 Past Work

Aesthetically, EveryWeek is heavily inspired by several key works. Nicholas Felton's Felton Biennial Report[9] is particularly significant for it's simple yet informative display of dense information. Martin Wattenberg and Fernanda Viegas work on Wind Map[12] also was significant for presenting large amounts of raw data in a simple manner to demonstrate a highly complex pattern.

2.2 Abstract

Immediately the question of "Should a map be used?" was asked. The answer to it is fairly intricate. A balance of practicality, privacy, and engagement needed to be struck, as this was a visualization of highly personal information. Leaving out the map removes information, reduces clutter, and forces the user to engage in the experience further. At first glance, it may appear like cruical information and backing would be removed, but this presents an opportunity: Much of the information is instantly recovered from the user's memory, and the rest of it is recovered by the user doing further research. This goes along quite well with Kevin Lynch's work on the way people visualize a city: Several key locations are used to provide the vast majority of context[11]. If another individual is not able to use the presented locations to provide context, they are unable to interpret much of the visualization. This creates a personal and intimate experience.

Removing the map ultimately gives the impression that the visualization is floating in space. Rather than attempt to ground the experience in anything, that was turned into a central theme, and built upon. Translucensy was used heavily, along with fading effects and many aspects of the sonification, to aid in the surreal aspect of it.

2.3 Lack of control

Fundamentally, the user's experience is simple: Hand the computer access to your personal data, and watch it be played back. No controls are exposed, neither in playing, pausing, slowing it down, nor the framing of the data. It forces the user to remain attentive, as the only way for them to re-watch a specific moment is to restart the simulation. An atmosphere of 'I have virtually no control over this' is induced, similar to the data being gathered about them. This aspect is not unlike Aaron Zinman's work on Personas[13].

2.4 Randomness

Randomness plays a heavy part in the visualization, from the way the visualization is built, to the way the data is processed. Some of the reasons for this are purely practical, but another part of it is the concept of multiple interpretations of the same data, with the same algorithm. From an aesthetic aspect, the randomness highlights the lack of control on the user's part, adding to the atmosphere of lack of control, as well as the intimacy of every outcome being completely unique and unreplicatable. The technical aspect randomness are explored more later, from the picking of bounds and scaling, to the visualization and highlighting of locations.

3. IMPLEMENTATION

EveryWeek has two basic components: A canvas for spacial location, and a chart for temporal information. The canvas has two major aspects: concentric circles on visited locations, and markers and lines to indicate the current location of the individual. The chart shows the accumulated activity for all the weeks in the visualization. EveryWeek was not designed to be a static image, but an animation.

3.1 Google Latitude

Google Latitude was the data source of choice. At the time of implementation, they had a public API[4] for accessing a user's location history in JSON format, but it would only give up to 1000 entries before a certain timestamp. In order to gather enough data, requests would need to be made 1000 at a time, at which point, the earliest timestamp was found, and a new request was made for 1000 events before that. This would continue for enough times (50 was decided to be a good number), before the process would continue. During this process, the text 'going back in time...' is displayed on screen, a fitting title, given that's what's actually happening. On average, between 10,000 and 40,000 individual location events were gathered, the timeframe being 5-15 weeks.

3.2 Data Preparation

Parsing the data, and getting it into a usable form was a particular challenge. Dealing with 40,000 location events isn't the most difficult dataset in the world, but doing it in javascript in a web browser made it considerably more challenging. The limiting factors in parsing through this data were the memory allocation given to a webpage, and the processing power of javascript.

The first step of parsing was to add additional information to every location event. In particular, all timestamps were in 'milliseconds since epoch', and thus contained no infor-

mation about days, hours, etc, so those were added. After which, the data was organized into weeks.

Picking a bounding rectangle for the visualization was a challenge. If too small of a bounding rectangle was chosen, there would be too much information loss, and places that an individual visits rather frequently may be totally cut off. However, if a bounding rectangle is too large, it will diminish the quality of the visualization, especially with respect to the dominant locations. A balance had to be struck that maximizes the amount of relevant information displayed to the user. This aspect heavily utilized many aspects of randomness, and the law of large numbers to come to consistently acceptable results.

Removing outliers proved to be a bigger issue than one would expect. Individuals tend to go on road trips, vacations, and other events that lay greatly outside a person's usual area. However, EveryWeek's intention was to visualize habits, and by definition of vacation, they're not normally considered weekly habits. In addition, individuals tend to make many one-time trips to rather distant places - data that forces the bounding rect to be much larger, but doesn't provide much relevant information to the user.

The first step in the process is finding the most frequent location. This is not trivial: locations are almost never repeated, and many locations are within a very small distance of eachother - indicating that they're the same 'place'. Clustering was considered, and attempted several times, but the results were less than satisfactory, especially with the processing constraints. RANSAC (Random Sample And Consensus) provided a better alternate. Random datapoints were picked from the set, and then cross checked across all of the other points. After running it a certain number of times, the best result is picked. With high probability, the algorithm will pick something in the most dominant area, and that one will be the 'best' result.

Once the most frequent location had been found, we now had an easy way of removing virtually all of the outliers. A distance (50 miles) was picked, and all points farther than that were thrown out. Admittedly, this is completely arbitrary, and may not work for everyone. Now the task was to find the best bounding rectangle. Once again, RANSAC proved to be a good implementation. Random points in the dataset were selected, and a rectangle was calculated that bound them. After collecting many of these rectangles, a pass through all the data was done, computing whether a point lay inside or outside the rectangle. The rectangle that was the smallest area, yet picked some threshold of points, was selected.

Lastly, one last pass would throw out all of the weeks where no points lay inside the bounding rectangle.

3.3 Technology

Processing.js[5] was used to implement the canvas. Processing[6] is an open source graphics programming library that normally runs as a java program, and Processing.js is the javascript version. It's based off of HTML5 canvas, and is highly cross-platform.

Audiolet[1] is the library used for sonification. It uses standard web browser sound apis to generate full-featured audio synthesis in a modern web browser.

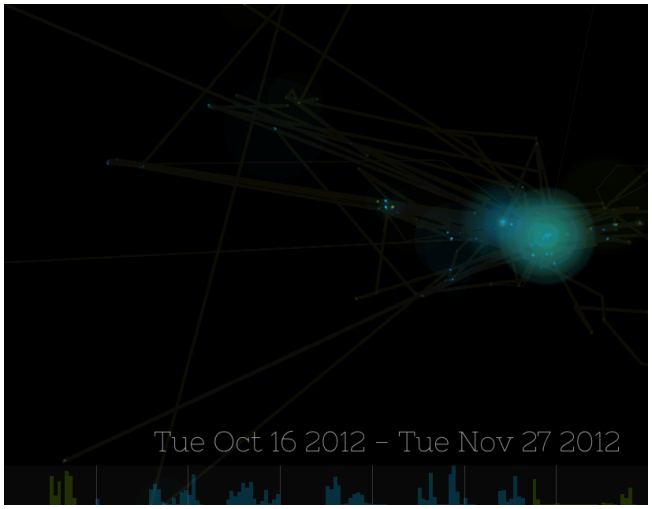


Figure 1: top: canvas with spacial information, bottom: time graph

3.4 Visualization

The visualization is obviously the main focus of the work. The aesthetic and informational goals have already been outlined, but the implementation was still quite open. The first thing the user notices when they visit the webpage is a completely black background, with a centered title and a few lines of text about the project. The visualization expands to fill the user's screen, but unfortunately can't react to changes without reloading the page, the reasons will become clear soon. The minimal start page was done to set the stage for the visualization: a clean slate. Once the user logs in, the text informs them that the visualization is 'going back in time' as it gathers past events.

Once the browser has finished gathering the data and parsing it, the visualization fades in. There are two main components: A top canvas to represent physical and spacial information, and a bottom graph to represent temporal information. The top canvas is falls into the thematic cartography of visualizations, and the bottom bar falls into the statistical graphs category. The current week is displayed in text at the bottom, although the day of the week and week number is specified, the exact date is not. Time scrolls by starting from week 0, to the current week. The canvas itself is broken up into two overlapping components, a layer that permanently holds places you visited, building a heatmap of sorts, and a layer that draws dots and lines atop exactly where you've been, creating a sketch of your travels.

3.4.1 Heatmap

The first section of the canvas, the heatmap, is a static layer. Every frame, a new bit of information is drawn, but nothing is ever cleared. This was done for performance reasons, as it would be infeasible to redraw the entire 40,000 item history every frame.

In the heatmap layer, several parameters are displayed; the

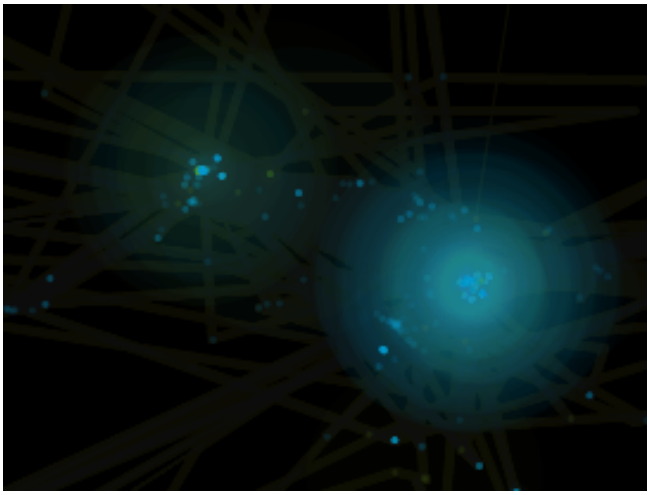


Figure 2: Example result of heatmap

most obvious one is location, where a translucent circle of several pixels is placed on top of that location. Another parameter was weekend vs. weekday, so if this event happened on a weekday, it'd be colored blue, and otherwise it'd be colored green. However, this proved to be a conflict: many points were drawn on top of eachother, and very quickly the pixels would get completely saturated. In addition, drawing dots of the same size does not do enough to highlight the places the user visits the most, while still making the places they visit occasionally visible. This is the last parameter, frequency.

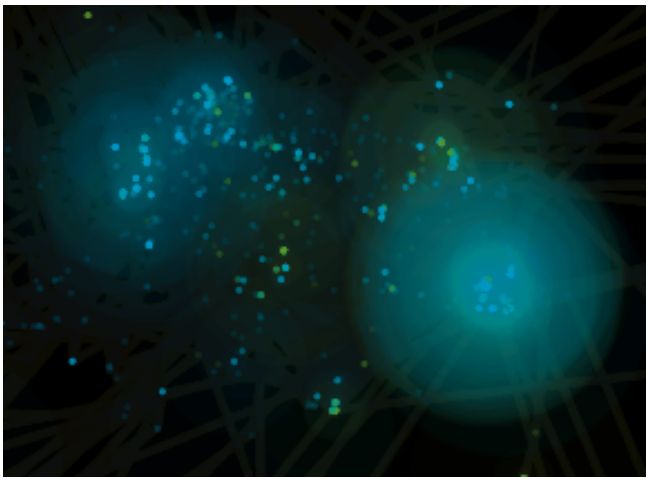


Figure 3: Too many points

In order to show how frequently a certain location was visited, concentric circles are drawn from that point. The functionality is very similar to a heatmap, but the aesthetic component is more pronounced, as all of the circles are centered around the same point, it attempts to indicate that the point is significant, moreso than the surrounding area. The only issue was deciding every frame how large to draw the circle.

A long time ago (3 weeks ago), a professor said "When you don't know what to do, do something random", and this was taken to heart. On every frame, the browser picks a random number, from -100 to 20. If the number is less than 1, a normal 3-pixel radius circle is drawn. If the number is greater than 80, a circle is drawn with a 2^k radius. This means with high probability, a place that the user visits frequently will have lots of large circles drawn around it, and places visited less frequently will have less. The circles are nearly transparent, so many of them need to be accumulated before it becomes noticeable. This technique is significant in two ways: It signifies the power of probability, and how our actions tend toward the mean, and how predictable we are, yet it also signifies how slightly unpredictable we can be - every rendering will turn out slightly different. As can be seen below, multiple runs of the same data produce different results, but they give very similar impressions.

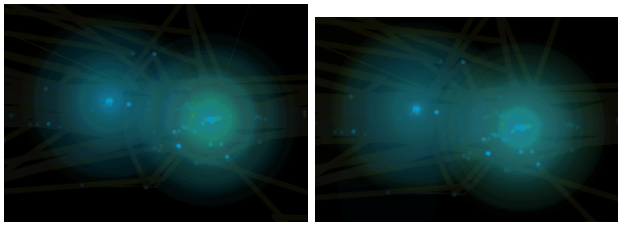
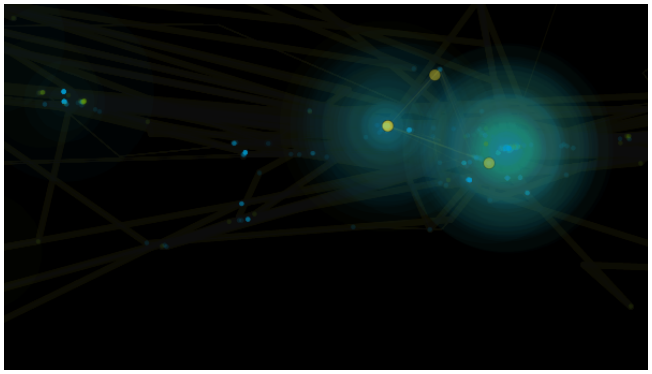


Figure 4: Two results from the same dataset

3.4.2 Paths

The next section of the canvas actually consists of two layers. The first is a circle to indicate the user's current position at the visualization's current time. After the first week, a 'ghost' circle is drawn for every previous week, to indicate where the user was at that time in the previous weeks. This allows the user to analyze their weekly habits in great detail - to watch when a large series of dots moves from home to work, or other patterns. This section is highlighted in the sonification as well.



An underlying layer connects the dots behind the user's locations. It provides a permanent record of the paths they have traveled during the simulation. Connecting lines start out fully colored, but never quite fade away. [1]

3.5 Time Graph

A graph at the bottom of the visualization highlights the user's movements throughout the week. A bar graph representing every hour of the week spans the bottom, and is

[1] Actually they sometimes do, depending on the browser, but that's not the intention

Figure 5: Two time graphs

filled in as the visualization progresses. In the beginning, the graph is empty, but scales as new information is added. The way data is added is quite simple: every timestep, the distance from the user's last location to their current location is added to that hour-bucket. Initially, the data is sparse, and difficult to make sense of any patterns, but as the visualization progresses, clearer and clearer habits can be distinguished.

3.6 Sonification

The choice to make a sonification to accompany the visualization was not done lightly: The browser was already being taxed with a computationally heavy visualization. However, only so much information could be displayed on the canvas, and audio provided a great way to both display new information, as well as highlight existing ones. This boiled down into two main parts: provide the user with a time-of-day context, and to sonify the paths layer of the visualization.

The overall aesthetics of the sonification were identical to the visualization. As such, much care was taken to avoid harsh sounds, and reverb/delay was added to give the user a sense of space.

Sonifying the time of the day was done with care, and accomplished mainly with a white noise generator sent through a band-pass filter. These sounds evoke many different thoughts, from an ocean, to rain and wind, and much more. In many ways, white noise represents randomness which itself blends into something predictable, which fits the theme perfectly. This system followed the time of day, getting higher pitched when the sun was highest. This provides the user with a lot more context into their daily habits. In addition, a faint synthesizer can be heard, running up a major scale during the day time, and running down the scale at night.

The main focus of the sonification is on the user paths. One of the weaker points in the canvas was displaying the movement of the user, so this was designed to complement that the most. The concept is simple: The screen is broken into a 24 by 24 grid of rectangles. When the user moves from point A to point B, the midpoint is found. Where that midpoint is in the grid corresponds to a note in a major scale, and that note is 'chimed' briefly, before fading away. Many times, either point A or point B will be an extremely common point, and thus taking the midpoint helps to avoid a mess of all of the same tones.

Every week, the grid of 24x24 shifts down one. This causes the same habit in week 1 to have the same rhythm as in week 2, but on a different note. The scale was picked such that common habits would make chords, a much more pleasant sound than a monotone note.

Much more intricate audio was intended, but unfortunately proved to be much too computationally intensive at the time.

4. CONCLUSIONS

The website is live at `http://weeks.rj.io`. The end result of EveryWeek is a beautiful image that's entirely unique to an individual and a moment, and is both intimate and informative.

EveryWeek aimed to provide a clean, minimal and unobtrusive space for users to explore, discover, and contemplate their weekly location habits. It's tough to quantify it's success in this area, but the small amount of testing that has been done has been quite positive.

4.1 What was learned

Many personal habits of mine were exposed through this visualization. The demo on the website is my own personal data. As a first piece of data-driven artwork, attempting many things I hadn't before, I'm quite pleased with the outcome. One of the difficulties of an artistic yet informative piece is walking the line between useful and artistic, but I feel it succeeded for the most part. It would be nice in the future to include more data analysis tools after the animation is complete.

Using randomness too heavily can be a problem. RANSAC and other randomized algorithms for finding bounds, etc. work fairly well in practice, but only if you can tolerate unpredictability. Using randomness to draw the visualization had fairly positive results, as the information tended to dominate a few points, and those points got highlighted accordingly. In retrospect, a less random method could have been conceived that gave much more consistent results with the same aesthetic component. However, a mixture of willingness to experiment, and computational technicalities made the chosen technique quite viable.

In terms of the success of the visualization, I feel it accomplished it's job well. The difficulties of exposing the ways habits can manifest themselves was definitely exposed in this project. If I could do the time graph again, I would look at doing a chart similar to Lee Byron's Stream Graph[8], and focus more on how and when you spend your time at given locations.

Technically, HTML5 is a very new medium for high performance visualizations, but provides virtually all of the tools required. Fewer libraries exist for HTML5, and they're less mature. However, the benefit of immediate feedback and engagement for anyone on the internet is an irreplacable quality. Hopefully, this paper provides some of the pros and cons of this approach.

This was my first time using Audiolet, and Processing.js. They're both roughly as simple and intuitive as their desktop equivalent, but not without some roadblocks. In the future, I hope web technologies improve to rival native visualization tools. As successful as this was, many barriers were reached, especially computationally.

Sharing is a big component that's missing. It'd be excellent for the user to take the information they saw and share it with others, but at the moment that's not an easy task, as

neither the video nor the audio can be recorded inside the browser. All that can be taken from the experience is the final image.

5. REFERENCES

[1] Audiolet. `http://oampo.github.com/Audiolet/`. [Online; accessed 13-December-2012].

[2] Foursquare. `http://foursquare.com`. [Online; accessed 13-December-2012].

[3] Google latitude. `http://latitude.google.com`. [Online; accessed 13-December-2012].

[4] Google latitude api. `https://developers.google.com/latitude/`. [Online; accessed 13-December-2012].

[5] Processingjs. `http://processingjs.org/`. [Online; accessed 13-December-2012].

[6] Processing.org. `http://processing.org/`. [Online; accessed 13-December-2012].

[7] B. Andrews. Habit. *The American Journal of Psychology*, 14(2):121–149, 1903.

[8] L. Byron and M. Wattenberg. Stacked graphs–geometry & aesthetics. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6):1245–1252, 2008.

[9] N. Felton. The 2010/2011 feltron biennial report. `http://feltron.com/ar11_01.html`, Jan. 2012. [Online; accessed 12-December-2012].

[10] M. Friendly and D. Denis. Milestones in the history of thematic cartography, statistical graphics, and data visualization. *Accessed: March*, 18:2010, 2001.

[11] K. Lynch. *The image of the city*, volume 1. MIT press, 1960.

[12] F. Viegas and M. Wattenberg. wind map. `http://hint.fm/wind/`, Sept. 2012. [Online; accessed 12-December-2012].

[13] A. Zinman, A. Dragulescu, Y. Assogba, and J. Donath. Personas - how does the internet see you? `http://personas.media.mit.edu/`, Feb. 2009. [Online; accessed 12-December-2012].