SmokeID: Enabling Visual Analysis of Qualitative and Quantitative Smoking Data from the Field

THEODORE WEBER, Western Washington University, USA
SHAANJOT GILL, Western Washington University, USA
PHILLIP WU LIANG, Western Washington University, USA

Tobacco addiction is one of the most challenging behavioral health problems with successful cessation rates remaining in the single digits. With increased availability of commercial mobile and wearable devices we have a opportunity to feasibly collect data from regular smokers in the field in near real-time. Tools to process and analyze this data can be utilized by health researchers to develop data-driven applications to promote smoking cessation. Furthermore, data collected from subjects in their natural environment could be used by behavioral health experts to support precision medicine. In this work we present an iteratively designed web-based visual analytics tool to help gain insights from qualitative and quantitative data collected from regular smokers in the field (N=5). We then present a case study with a health researcher, as well as discuss usage scenarios of how our tool can be used in practice.

Additional Key Words and Phrases: data visualization, mHealth, p4 medicine, smoking cessation, wearable technology, human-computer interaction

ACM Reference Format:
Theodore Weber, Shaanjot Gill, and Phillip Wu Liang. 2018. SmokeID: Enabling Visual Analysis of Qualitative and Quantitative Smoking Data from the Field. 1, 1 (December 2018), 16 pages. https://doi.org/0000001.0000001

1 INTRODUCTION

Smoking remains as one of the most challenging behavioral health faced in society today. It is the leading cause of preventable disease and death in the U.S., being responsible for nearly 1 in every 5 deaths [for Disease Control et al. 2017]. 37.8 million U.S. adults are currently cigarette smokers, and 78.1% are considered daily smokers, smoking at least one cigarette per day [for Disease Control et al. 2012]. Nearly 68.8% of smokers report a desire to quit, but cessation success rate still remain in the single digits [(CDC et al. 2011; US Department of Health and Human Services et al. 2006]. Even with an abundance of cessation support applications tobacco addiction is still a very prevalent behavioral health problem.

Increased availability of commercial mobile and wearable devices presents a unique opportunity to continuously monitor smoking behavior in the field [Ertin et al. 2011; Hossain et al. 2017; Pantelopoulos and Bourbakis 2010; Villar et al. 2015]. With this data we can identify key factors contributing to smoking incidents, and monitor smoking behavior

Authors' addresses: Theodore Weber, Western Washington University, 516 High Street, Bellingham, WA, 98225, USA, webert3@wwu.edu; Shaanjot Gill, Western Washington University, 516 High Street, Bellingham, WA, 98225, USA, gills4@wwu.edu; Phillip Wu Liang, Western Washington University, 516 High Street, Bellingham, WA, 98225, USA, wulianp@wwu.edu.

in near real-time. Due to the volume and diversity of both contextual and sensor data, it is absolutely necessary to develop effective visualization tools to help health researchers and behavioral health experts gain insights.

Recent works have presented tools capable of visualizing smoking data collected in the field. Polack Jr. et al. developed an interactive web application for health researchers to explore discretized events generated from mHealth data [Polack Jr et al. 2018]. This tool was developed primarily for health researchers and behavioral health experts to identify common physiological event sequences preceding smoking episodes. Similarly, in [Sharmin et al. 2017] Sharmin proposes a mobile-based application enabling the analysis of aggregated smoking behavior in relation to stress. This application targets regular smokers as the primary user. While these tools are able to distill large datasets effectively, they are still heavily reliant on quantitative measurements. Qualitative data collected from regular smokers in the field (e.g. information extracted from in-person interviews and survey responses) can also yield significant insights on smoking behavior, but they have not been utilized in existing visualization tools.

To address this problem we present SmokeID—an easy-to-use visual analytics tool capable of visualizing both quantitative and qualitative data collected from regular smokers in the field. Our tool focuses on answering the following questions related to smoking behavior in the field:

- How are environmental and contextual factors (e.g. location, smoking cues, emotional state) associated with smoking behavior?
- Are there any similarities in the smoking behavior between participants (i.e. cohort analysis)?
- What are the most important features that predict smoking lapse? We define *smoking lapse* as a time in which a participant smokes a cigarette.

To address these research questions we develop a web-based application consisting of two primary interfaces: the *user* interface, which enables behavior analysis of individual users, and the *cohort* interface, which enables behavior analysis across all users. We populate our application using data collected from a field study ($N = 5$) where both physiological and contextual data was collected from regular smokers in their natural environment for 5 days. We iteratively design our application by sharing preliminary visualizations with a group of health researchers ($N = 15$) to improve our design during the development phase. Finally, we present findings from a case study with a health researcher using our tool to answer questions related to real-world smoking behavior.

2 METHODOLOGY

The data used in our tool comes from a 5-day mixed-method study in which regular smokers ($N = 5$) were asked to use mobile and wearable devices in the field. Physiological data was collected using the Hexoskin ©Smart Kit (Carré Technologies Inc., Montreal, Québec Canada)[Düking et al. 2016], a commercially available wearable smart shirt (see Figure 7). Participants also carried an Android phone with a custom application enabling self-reports for stress, urge, and smoking episodes (see Figure 7). This app also collected some quantitative data (e.g. location, noise-level), as well as prompting users with Ecological Momentary Assessments (EMAs) [Shiffman et al. 2008; Stone and Shiffman 1994] at least three times a day (see Table 1). Finally, our dataset contains some information derived from transcribed in-person interviews (see Table 2) annotated using the open coding technique [Strauss and Corbin 1990]. After post-processing our dataset contained ∽148.7 hours of sensor data, 95 self-reported smoking episodes, 38 self-reported stress values, 36 self-reported urge values, and 65 EMA responses.

To prepare the data for our visualizations we used the Python scripts to generate CSV (table data) & JSON (network data) files, depending on the visual encoding. Python is frequently used for data mining and machine learning tasks

Table 1. Sample EMA Questions

In the last ten minutes, have you had a strong temptation/urge to smoke (1 - 6 Likert scale)?
In the last 30 minutes were you engaged in social interaction?
In the last ten minutes, have you seen any of the following people smoke? (select all that apply) *Options:* i. Partner, ii. Family Member, iii. Friends, iv. Supervisor, v. Subordinates, vi. Co-workers, vii. Stranger, viii. Other.
Was there a specific reason for your most recent cigarette? *Options:* i. To reduce stress, ii. To reduce urge, iii. No specific reason, iv. Other.
Is a cigarette available to you right now?

Table 2. Sample Pre-interview Questions

How many cigarettes do you smoke per day (on average)?
How does your smoking behavior vary on a weekday vs. on the weekend?
How many hours a week do you work?
What are the things that cause you the most stress on a day-to-day basis?
How many hours a day do you spend using a smartphone device (browsing, calling/texting, playing games, etc)?

Table 3. Tools used to build our web application.

Language	Dependencies
Python	Pandas, NumPy, SciPy
JavaScript	d3 (plugins: d3-sankey, correlation-graph), node, webpack, leaflet, jStat
HTML/CSS	Paper Dashboard

operating on large quantitative datasets similar to our own. Thus, it was well suited for our preprocessing requirements. To develop our web application we employed the javascript libraries listed in Table 3. We chose to build a web application using *d3.js*[1] because it is cross-platform and enables customizable, interactive data visualizations.

3 VISUALIZATIONS

3.1 Map View

The map view (see Figure 1) uses a familiar map interface to show locations where study participants have smoked. The analyst may select between participants using the drop-down menu, and scroll or zoom to interact with the map interface. When a smoking marker is selected, additional details become visible in a tooltip. This visualization addresses the task of identifying what environmental factors affect smoking behavior. By using a map interface we are not only able to identify common smoking locations (i.e. clusters), but we can utilize map imagery to identify additional contextual information. For example, in Figure 1 we can see that participant 2 frequently smoked on a university campus. This could indicate that studying, or being in class, has an impact on their smoking behavior. Also, by plotting the smoking episodes of all participants we can see if there are any commonalities between all participants.

[1] https://d3js.org/

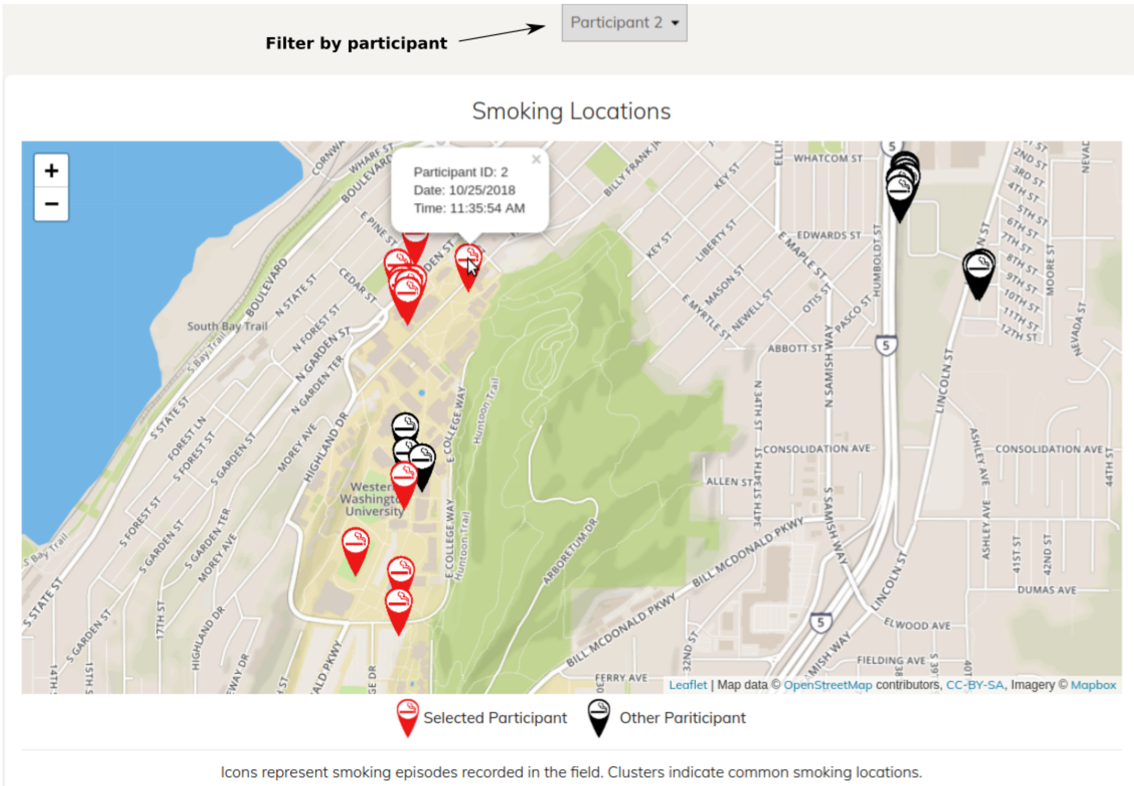


Fig. 1. The map view shows smoking locations of all participants, highlighting the participant selected in the drop-down menu.

3.2 HRV Scatterplots

Heart-rate variability (HRV) is commonly used as an indicator of stress [Kim et al. 2018], and stress is cited as a significant factor contributing to smoking lapse [Strauss and Corbin 1990]. HRV was a feature in our dataset that was derived from R-R intervals using the calculation in equation 1. Our dataset also contained self-reported stress and urge values reported on a 1 - 6 Likert scale [Allen and Seaman 2007] by participants in the field. Using this data we can analyze the relationship between stress/urge and HRV to address the task of identifying key factors contributing to smoking lapse. In Figure 2 we present a scatterplot visualization as part of the *cohort* interface. Here we compute the z-score of each self-report stress and urge value to determine whether it is high or low relative to each participant's mean response value. We do the same calculation for HRV to determine how much the 5, 10, and 15-minute HRV values preceding the self-reports deviate from the participant's mean HRV response. By analyzing the HRV directly preceding self-reported stress and urge we can see if there is a significant relationship between the variables.

We choose to use a scatterplot for our visual encoding because it is easy to identify correlation between two variables. As shown in Figure 2, we present one scatterplot showing HRV vs. self-reported stress, and another showing HRV vs. self-reported urge. Analysts may filter between time intervals using the drop-down menu. We also plot a best-fit linear regression line to help identify linear correlation between the two variables. Finally, we present both pearson's and spearman's linear correlation coefficients to assess the linear and monotonic relationship between the two variables.

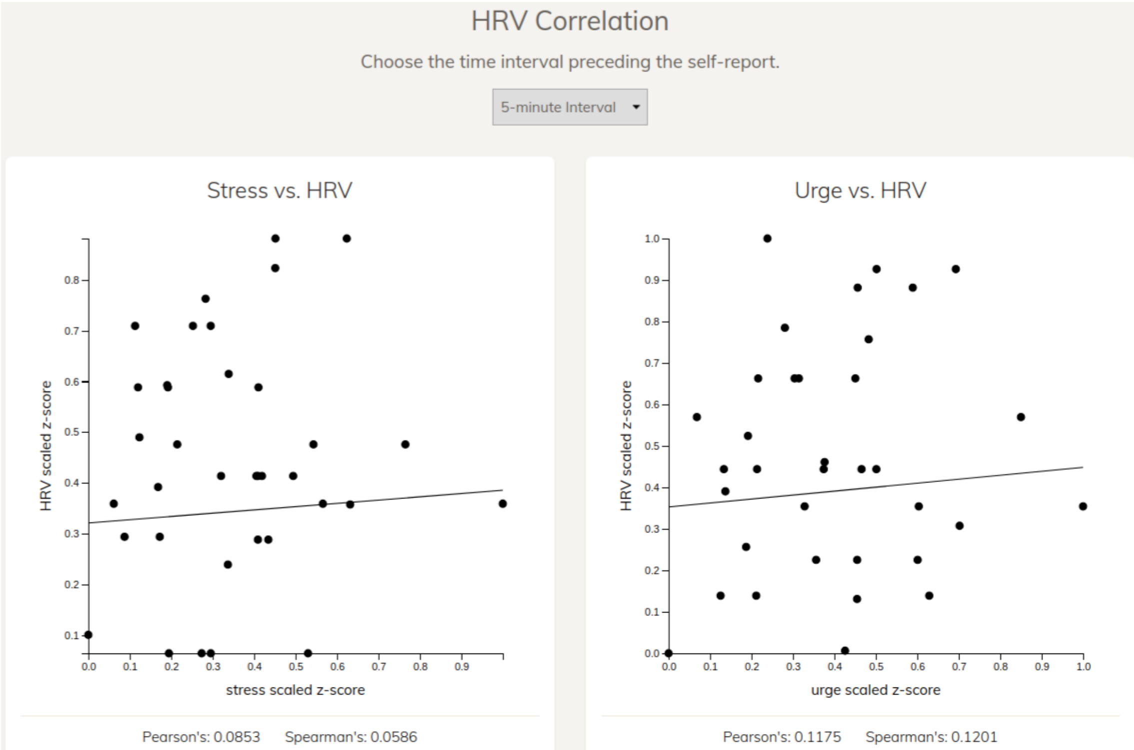
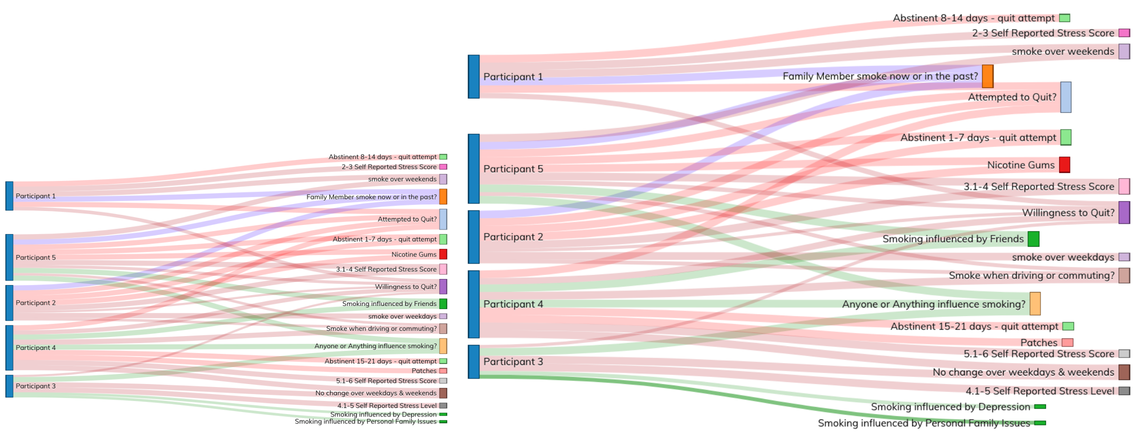


Fig. 2. Two scatterplots: one showing the relationship between self-reported stress and HRV, the other showing the relationship between self-reported urge and HRV.

3.3 Interactive Sankey Diagram



Initial rendered visualization. Visual analysis using UI features based on changing requirements.

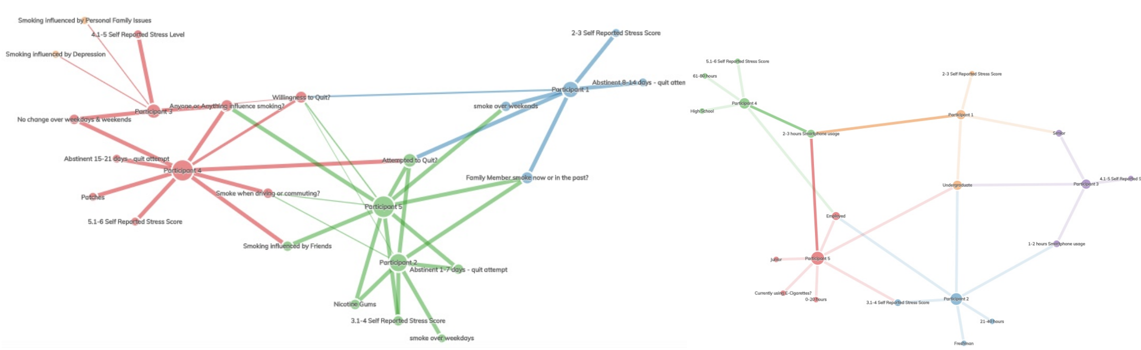
Fig. 3. The Sankey diagram visualization for mapping categorical data extracted from survey and interview data.

During the recruitment process, participants were initially interviewed to collect some pertinent data in regards to their smoking behavior. Also for the field study, participants were provided with an Android phone with a custom application deployed, which regularly prompted them with EMAs (see Figure 7). We decided to utilize this data to develop visualizations that will provide useful insights on how environmental and contextual factors are associated with smoking behavior. We also use the data to enable cohort analysis.

Our dataset contained multiple qualitative ordinal data points that were reported on a 1 - 6 Likert scale [Allen and Seaman 2007], qualitative binary data points (i.e. 'yes' or 'no' responses) and quantitative data points. To avoid a possible hairball problem [Kaski and Peltonen 2011] by incorporating all the ordinal data (Likert scale of 1-6), we apply data preprocessing to convert ordinal data into binary data whenever possible. We standardized all the EMA responses and classified them as 'yes' if the z-score is higher than 0 or 'no' if the z-score is less than 0. Another challenge was converting a flat data into a network structured JSON file that could be utilized to create visualizations. Finally, we had to normalize our data for visualizing EMA responses, so that the EMA response represent the relative fluctuation from each participant's mean Likert scale response.

Sankey diagram (also know as an Alluvial diagram) is a flow diagram (Figure 3 flows from left to right) that depicts many to many connections from one set of values to another [Riehmann et al. 2005]. Sankey diagram was chosen because it is easy to identify many to many connections between nodes. The width of the connection or link is proportional to the number of records. When visualizing participants based on their interview data, the width of the links are uniform as the data consists of unstructured and binary data points. Where as when visualizing EMA responses (see Figure 8), the width depicts the weight of the relation. As shown in Figure 11, the nodes of the visualizations can be moved around in both x-axis and y-axis and all connections to and from the node can be highlighted by clicking on the node. The colors of the links are based on the type of the attribute that it connects to, when visualizing demographic data (see Figure 10) work-status is colored blue, stress is brown, education is green, and smart phone usage is red. The tool-tip for the connections display the direction of the flow and the associated value, where as the same for the node rectangle displays it's total value and name. We utilize this kind of visualization to perform cohort analysis.

3.4 Participant Correlation Graphs



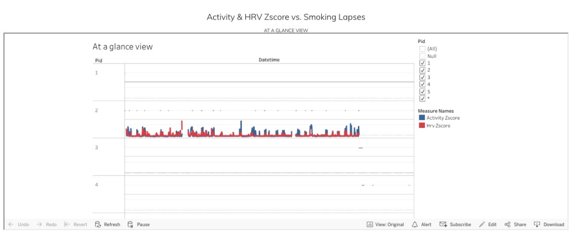
Visualizing smoking behavior connections. Visualizing demographic connections.

Fig. 4. Correlation Graphs to identify connections and cohorts among participants.

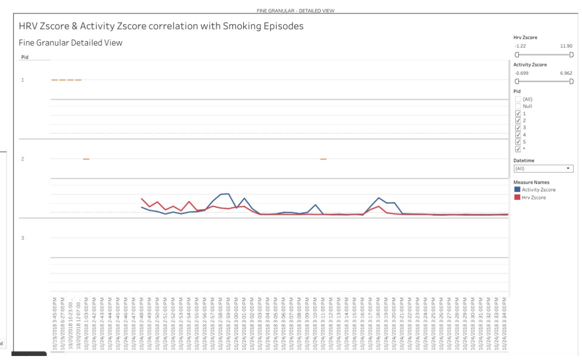
The data utilized in this visualization is similar to the one used for producing corresponding sankey visualizations (Figures 3 and 9). Though this JSON dataset has different property and field names and an additional 'id' property.

Correlation graph (see Figure 4) draws a correlation network of relationships from tabular data based on pairwise correlations. We chose this particular visual encoding because it is highly scalable and can incorporate data from hundreds of participants. Using the interactive features of the visualization, we can click target node in the graph to trace it's connections with other nodes and the weight of the connection. The nodes can also be moved around for ease of accessibility. Using this visualization, we can perform cohort identification and analysis, one of it's many application could be as a recruitment tool. As shown in Figure 13, clicking on the 'smoke when driving or commuting' node will only highlight it's corresponding connections and nodes. So based on this visualization, participant 2 and 5 sometimes smoke while driving or commuting (weight=0.5) whereas participant 4 frequently smokes in this scenario (weight=1).

3.5 Standardized HRV and Activity vs. Smoking Lapses



At a Glance View.



Fine Granular View.

Fig. 5. Plotting HRV and Activity z-score in 120 minute bins [1 hour before and 1 hour after a reported smoking incident] .

We have known stress to be a significant contributor to smoking lapse [Strauss and Corbin 1990] and HRV has been used as indicator of stress [Kim et al. 2018]. But one challenge in consistently utilizing HRV for detecting stress is that high stress situation and intense physical activity exhibit similar physiological profiles [Sharmin et al. 2017]. To address this issue, we applied a z-score transformation on HRV and activity features on an individual basis. That is, for each participant, we computed the z-score using only their data to account for individual differences in their physiological behavior. The 'activity' feature is provided through the Hexoskin API, and is based on the accelerometer intensity vector[2]. There were wide gaps of time for which sensor data of participants was not available, so we decided against using continuous time series. Instead, we focused on extracting only HRV and activity z-score values for 2 hour windows surrounding smoking lapse (i.e. ~60 minutes before the recorded smoking lapse and 60 minutes after). Thus, activity z-score was plotted on the same axis as HRV z-score against discrete date time. Also, smoking episodes were marked against same date time axis.

For this particular visualization, we utilized Tableau Online visualization tool[3]. This customized visual encoding was developed for a specific goal and is a hybrid of discrete time series and small multiples [van den Elzen and van Wijk 2013]. Two separate Tableau work sheets were created of the same visualization but one has it's view set to 'entire' for providing an 'at a glance' view and other has it set to 'standard' providing a 'detailed view'. To create this visualization,

[2]See: https://api.hexoskin.com/docs/index.html#acceleration-related
[3]See: https://www.tableau.com/products/cloud-bi

we had to set activity and HRV z-scores as attributes and put together as a dual axis in rows. The participant ID was also added as an attribute to rows providing a small multiples appearance and functionality. This visualization was integrated into the app as shown in Figure 12, directly as a HTML snippet using an <iframe>.

This visualization can be utilized to visually find a correlation between HRV [stress] and smoking lapse. It can also be used to disregard data points where HRV and Activity z-score have similar values or slope. It will also enable comparisons between different participants in regards to how their smoking lapses relate to these 2 physiological features.

4 FINDINGS

4.1 Research Questions

Here we present insights we gained from using our visualization tool, and highlight findings that address our proposed research questions.

The map view shows the exact location of where each participant self-reported a smoking incident. This helped us easily locate individual participants as well as cohorts that tend to smoke in common locations. Using this visualization we saw several clusters of smoking markers, implying that those were frequent smoking locations. This indicates that location could be a valuable contextual factor contributing to smoking lapse. Future smoking cessation systems should consider this feature when attempting to identify a user's vulnerability to smoking lapse.

Finding commonalities between participants can also provide a lot of insight on behavioral and environmental factors related to smoking. We extracted this qualitative data from the Ecological Momentary Assessments (EMAs) and in-person interviews to use in our visualizations. The Sankey diagrams allow us dive deeper and gain insights from our EMA data. We found that cigarette disposal containers (i.e. ashtrays) were the most commonly observed smoking cue, which may indicate that it is a factor contributing to lapse. The correlation graph was also a surprisingly effective tools for identifying common themes from the in-person interviews, as well as identifying user cohorts. Participants 2 & 5, for example both described anxiety as a factor contributing to lapse, and mentioned that they often smoke while commuting. These connections brought them closer together in the correlation graph, and unified them as part of the green cluster (see Figure 4). As the number of participants grow, we believe this will be an extremely effective way to identify cohorts, and the common themes connecting the participants within those cohorts.

When analyzing the visualizations to identify significant predicters of smoking lapse we found that HRV showed no significant linear or monotonic correlation with self-reported stress or smoking urge. This was realized by observing our scatterplot visualization. Although HRV is commonly used as a surrogate for stress [Kim et al. 2018], these findings could indicate that HRV alone might not be a sufficient indicator. Future researchers should err on the side of caution when using HRV in this manner.

4.2 User Evaluation

To evaluate our application we invited a health researcher at Western Washington University to answer a series of research questions using our tool. After some analysis they were asked to rate the efficacy and intuitiveness of each visualization on a scale of 1 - 6 (see Table 4), and provide feedback on how the visualization could be improved. For the map view the researcher commented that using a map interface was very intuitive for exploring geolocation, but it might not be scalable for a large number of participants or smoking episodes. To address this we could provide additional filtering options to reduce the amount of data presented. For the correlation graph the researcher said it

Table 4. Results from evaluating our visualization tool with a health researcher. Effectiveness and intuitiveness is based on a 1 - 6 scale, where 1 ⇒ poor and 6 ⇒ excellent.

Visualization	Research Question	Effectiveness	Intuitiveness
Map View	*Do these participants tend to smoke in similar locations?*	5	6
Correlation Graph	*What are some common things that influence smoking behavior among our participants?*	4	4
Sankey Diagram	*What are the most common smoking cues observed by our participants?*	5	6
Scatterplot	*How correlated is HRV with self-reported stress and smoking urge?*	5	5

was effective for identifying similar participants, but difficult to identify common traits among those participants. For an improvement they suggested using a different shape for participant and trait nodes. The Sankey diagram was well received by the researcher, and they claimed it was more effective than the correlation graph for identifying common traits among participants. They suggested that we add more instructions on how to use the highlighting feature because it was especially useful for answering questions. Finally, for the scatterplot visualization the researcher suggested we add some additional information about where you would expect the trends to be to better inform the user of what to look for. The time series activity and HRV plots were not available when we ran the user evaluation, so those visualizations were not addressed.

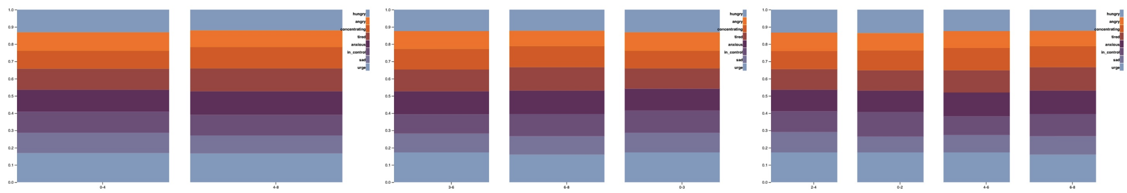
4.3 Additional Findings

From observing Figure 6 a stacked bar chart visualization, we discovered that the mean Likert scale EMA responses of participant do not change drastically with changing cigarette use. This visualization was not incorporated in our application due to time constraints, but will be availale in future version. From observing Sankey diagram (Figure 8), we discovered that out of 53 cases where participant smelled smoke, there 28 cases where no smoking cues were observed. Also, caffeinated drinks were consumed by participants 7 times (13.2%) which closely followed strong urge which appeared 9 times (17%). Alcohol beverages was only reported 1 time. Also, participants changed location 32 times (60.3%) to smoke, of these about another 60% changed location at home to smoke. Out of 5 participants, 2 smoke more often on weekends, 1 on weekdays and 2 there is no change over weekdays and weekends. There are multiple cases where HRV z-score rises in very close proximity to a similar change in slope of activity z-score. In such cases, correlation between stress and smoking lapse can be ignored.

5 DISCUSSION

There were many challenges that we encountered and overcome during the course of this project. This included learning web development fundamentals, how to deploy on and manage a node server, and how to use D3.js. Initially, the Sankey diagrams that we developed (see Figure 15) had too much information and were too cluttered making it difficult for users to follow and interpret the links and nodes. Based on feedback, we decided to classify attributes into categories and then develop a separate Sankey diagram for each of the categories. We also had to implement normalization techniques to make sure that Sankey diagram's node did not aggregate to a total greater than their actual value, which was the case in which EMA reponses were visualized. Similarly, as part of optimization, we develop a novel visualization (Figure 6) which required normalization to make each bar (sum of average response) to be equal to 1.

There were few limitation that we observed in our project. Firstly, we only had 5 participant for data collection and they only wore the Hexoskin sensor (Figure 7) for ∽6 hours against the expected 11 hours which lead to limited quantitative data. Also, we observed that participants would sometimes take off the sensors in the middle of the day, creating discontinuities in the time series data. Finally, we could only get our visualization tool evaluated by one health researcher, and we lack a formal evaluation from a behavioral health expert.



2 Bins [0-4,4-8] 3 Bins [0-3,3-6,6-8] 4 Bins [0-2,2-4,4-6,6-8]

Fig. 6. How EMA and contextual factors change with increased cigarette consumption?

In future, we envision our visualization tool as a portal to which health and behavior experts could upload specifically structured data and corresponding visualizations will get rendered. To achieve this goal, we plan on adding new visualizations, deploying Python data pre-processing scripts on web server and implementing a data-pipeline to incorporate sensor data in real-time. The sensor data can be extracted remotely via a cloud service as the sensor is synced up to a mobile phone provided to participant which is connected to internet. As a first step, we developed a new visualization (Figure 6) in D3.js that displays the mean Likert response for various EMA responses over all participants who smoked as many cigarettes per day as defined in the bin (x-axis). Thus, the stacked bar at 0-2 represents the mean Likert scale EMA responses for whether they are hungry, angry, difficulty concentrating, tired, anxious, in control of urge, sad and feeling urge (from top to bottom) over all participants who smoked 0-2 cigarettes per day.

6 CONCLUSION

In this work we designed a web-based visual analytics tool, SmokeID, capable of combining qualitative and quantitative data from smokers in the natural environment. We then used our tool to analyze data from a 5-day field study ($N = 5$) to identify key factors contributing to smoking lapse. Finally, we evaluated our results with a health researcher to assess the efficacy and usability of our application.

In future we plan on implementing more visualizations for user-level analysis (see Figure 16), as well as addressing the limitations identified during our user evaluation. Furthermore, we hope to present our tool to additional health researchers as well as behavioural health experts in a longitudinal study to better understand how our tool can be used to analyze data from smokers in the field.

ACKNOWLEDGMENTS

We would like to acknowledge our advisor, Dr. Moushumi Sharmin, as well as researcher Matthew Ferrin, from the NEAT lab at Western Washington University.

REFERENCES

I Elaine Allen and Christopher A Seaman. 2007. Likert scales and data analyses. *Quality progress* 40, 7 (2007), 64–65.

Centers for Disease Control and Prevention (CDC et al. 2011. Quitting smoking among adults–United States, 2001-2010. *MMWR. Morbidity and mortality weekly report* 60, 44 (2011), 1513.

Peter Düking, Andreas Hotho, Hans-Christer Holmberg, Franz Konstantin Fuss, and Billy Sperlich. 2016. Comparison of non-invasive individual monitoring of the training and health of athletes with commercially available wearable technologies. *Frontiers in physiology* 7 (2016), 71.

Emre Ertin, Nathan Stohs, Santosh Kumar, Andrew Raij, Mustafa Al'Absi, and Siddharth Shah. 2011. AutoSense: unobtrusively wearable sensor suite for inferring the onset, causality, and consequences of stress in the field. In *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems*. ACM, 274–287.

Centers for Disease Control, Prevention (CDC, et al. 2012. Current cigarette smoking among adults-United States, 2011. *MMWR. Morbidity and mortality weekly report* 61, 44 (2012), 889.

Centers for Disease Control, Prevention, et al. 2017. Smoking & tobacco use fast facts. 2016. *URL: https://www. cdc. gov/tobacco/data_statistics/fact_sheets/fast_facts/[accessed 2017-02-27][WebCite Cache ID 6ob55B8NG]* (2017).

Syed Monowar Hossain, Timothy Hnat, Nazir Saleheen, Nusrat Jahan Nasrin, Joseph Noor, Bo-Jhang Ho, Tyson Condie, Mani Srivastava, and Santosh Kumar. 2017. mCerebrum: A Mobile Sensing Software Platform for Development and Validation of Digital Biomarkers and Interventions. In *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*. ACM, 7.

Samuel Kaski and Jaakko Peltonen. 2011. Dimensionality reduction for data visualization [applications corner]. *IEEE signal processing magazine* 28, 2 (2011), 100–104.

Hye-Geum Kim, Eun-Jin Cheon, Dai-Seg Bai, Young Hwan Lee, and Bon-Hoon Koo. 2018. Stress and Heart Rate Variability: A Meta-Analysis and Review of the Literature. *Psychiatry investigation* 15, 3 (2018), 235.

Alexandros Pantelopoulos and Nikolaos G Bourbakis. 2010. A survey on wearable sensor-based systems for health monitoring and prognosis. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 40, 1 (2010), 1–12.

Peter J Polack Jr, Shang-Tse Chen, Minsuk Kahng, Kaya De Barbaro, Rahul Basole, Moushumi Sharmin, and Duen Horng Chau. 2018. Chronodes: Interactive Multifocus Exploration of Event Sequences. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 8, 1 (2018), 2.

Patrick Riehmann, Manfred Hanfler, and Bernd Froehlich. 2005. Interactive sankey diagrams. In *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*. IEEE, 233–240.

Moushumi Sharmin, Theodore Weber, Hillol Sarker, Nazir Saleheen, Santosh Kumar, Shameem Ahmed, and Mustafa AlâĂŹAbsi. 2017. Opportunities and Challenges in Designing Participant-Centric Smoking Cessation System. In *Computer Software and Applications Conference (COMPSAC), 2017 IEEE 41st Annual*, Vol. 1. IEEE, 835–844.

Saul Shiffman, Arthur A Stone, and Michael R Hufford. 2008. Ecological momentary assessment. *Annu. Rev. Clin. Psychol.* 4 (2008), 1–32.

Arthur A Stone and Saul Shiffman. 1994. Ecological momentary assessment (EMA) in behavorial medicine. *Annals of Behavioral Medicine* (1994).

Anselm Strauss and Juliet Corbin. 1990. Open coding. *Basics of qualitative research: Grounded theory procedures and techniques* 2, 1990 (1990), 101–121.

US Department of Health and Human Services et al. 2006. The health consequences of involuntary exposure to tobacco smoke: a report of the Surgeon General. *Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention, Coordinating Center for Health Promotion, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health* 709 (2006).

Stef van den Elzen and Jarke J van Wijk. 2013. Small multiples, large singles: A new approach for visual data exploration. In *Computer Graphics Forum*, Vol. 32. Wiley Online Library, 191–200.

Rodrigo Villar, Thomas Beltrame, and Richard L Hughson. 2015. Validation of the Hexoskin wearable vest during lying, sitting, standing, and walking activities. *Applied Physiology, Nutrition, and Metabolism* 40, 10 (2015), 1019–1024.

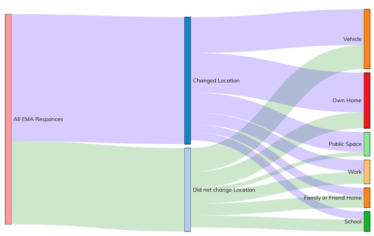
A SUPPLEMENTAL MATERIALS

$$RMSSD = \sqrt{\frac{1}{N-1}(\sum_{i=1}^{N-1}(RR_i - RR_{i+1})^2} \tag{1}$$

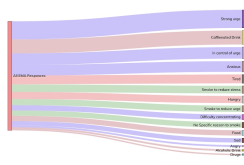
Hardware used for data collection—Hexoskin Vest and Motorola Moto G5

Android application used for self-reports.

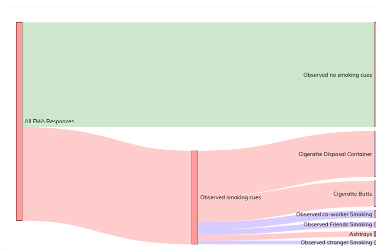
Fig. 7. Devices used for data collection in the field.



Location based analysis.

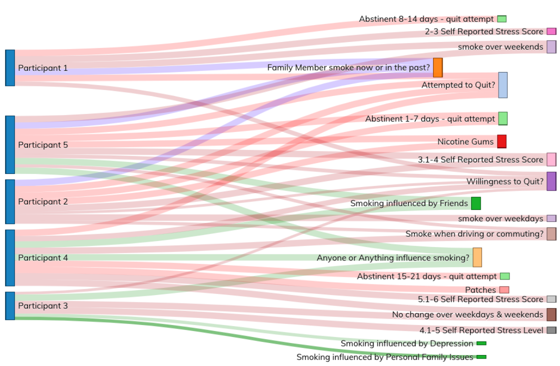


Stress related analysis.



Smoking cues related analysis.

Fig. 8. Sankey diagram to map categorical data extracted from participant EMA reports.



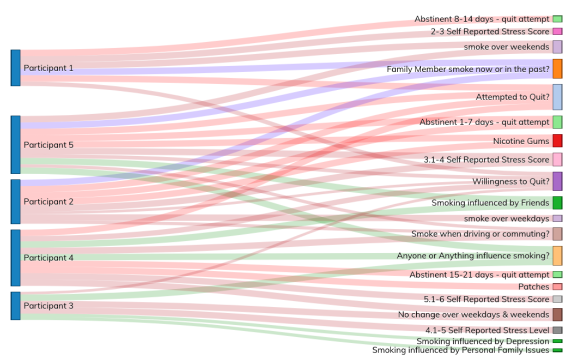


Fig. 9. Sankey Diagram to map smoking behavior categorical data extracted from survey and interview data.

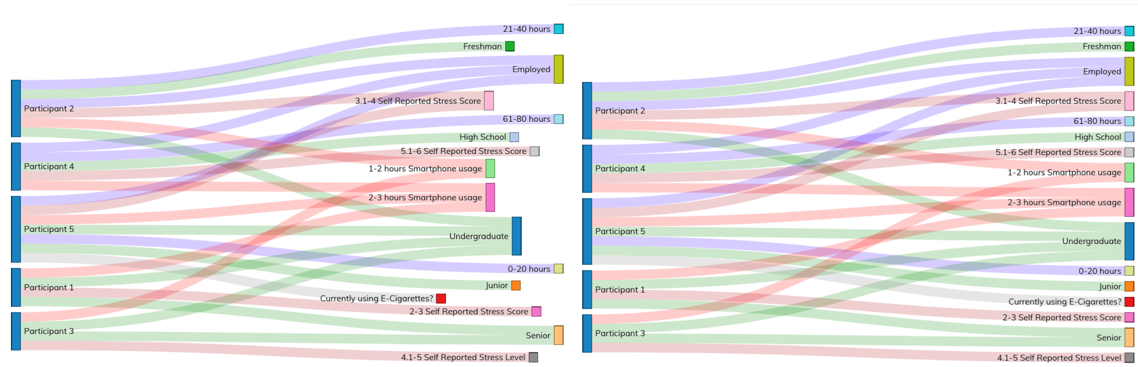


Fig. 10. Sankey Diagram to map demographic data extracted from categorical survey and interview data.

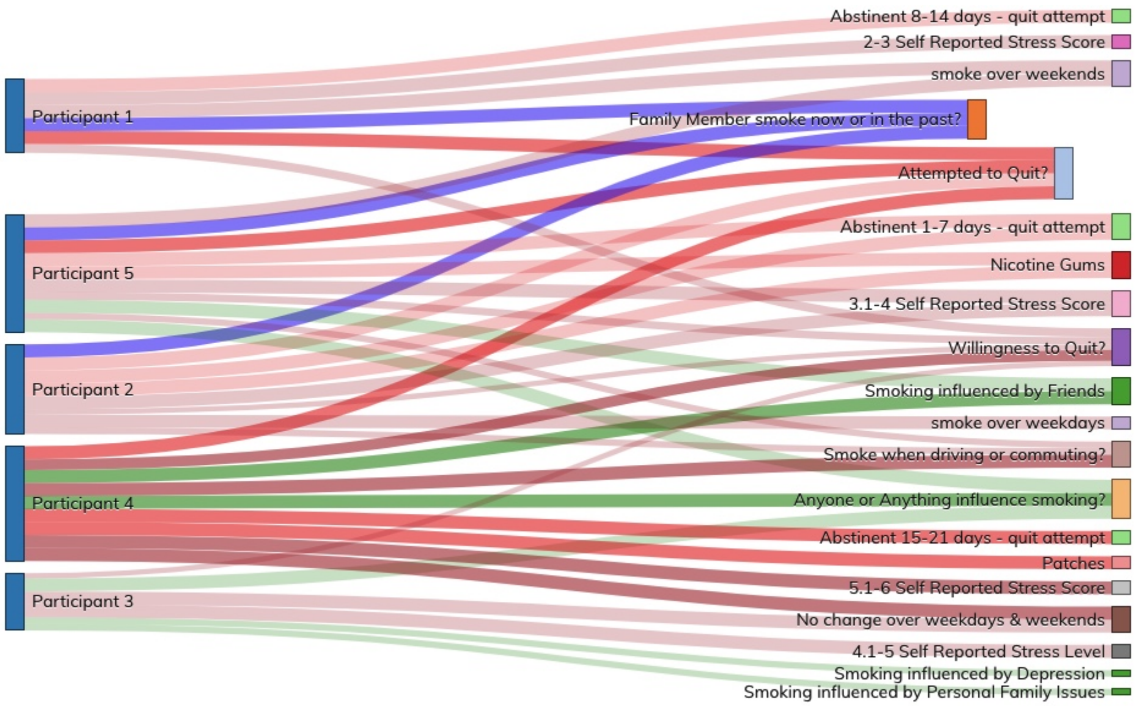


Fig. 11. Visualization supports node highlighting to trace connections which is activated by clicking on the rectangular node.

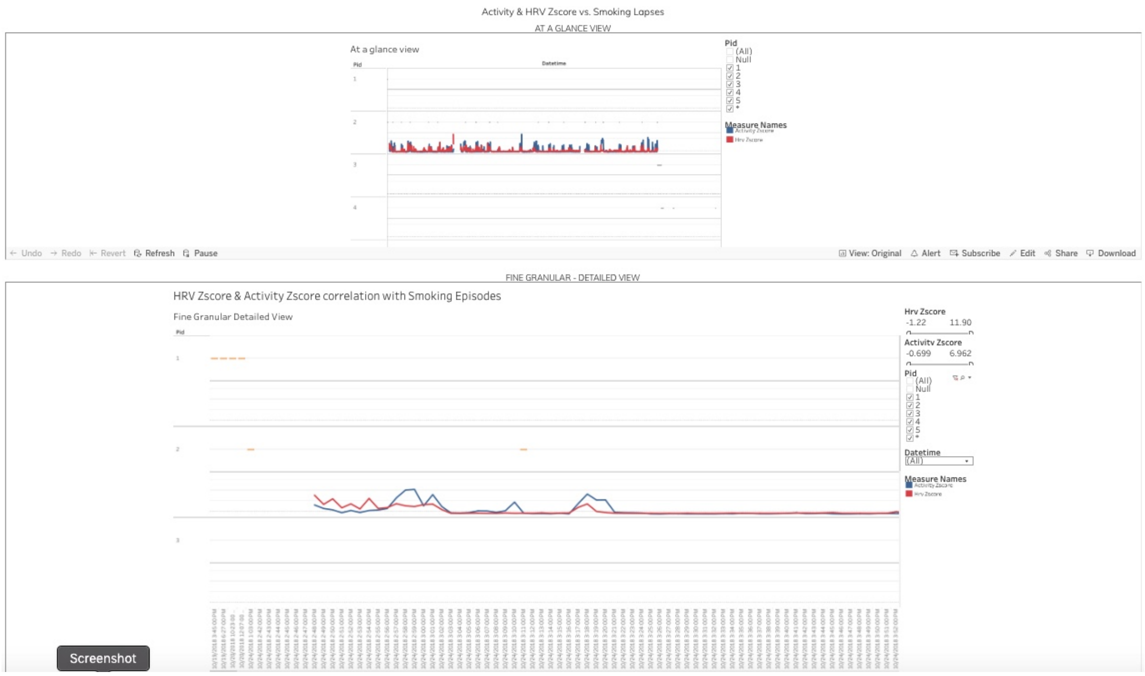


Fig. 12. Dashboard view of the Activity and HRV z-score plot in relation to recorded smoking incidents.

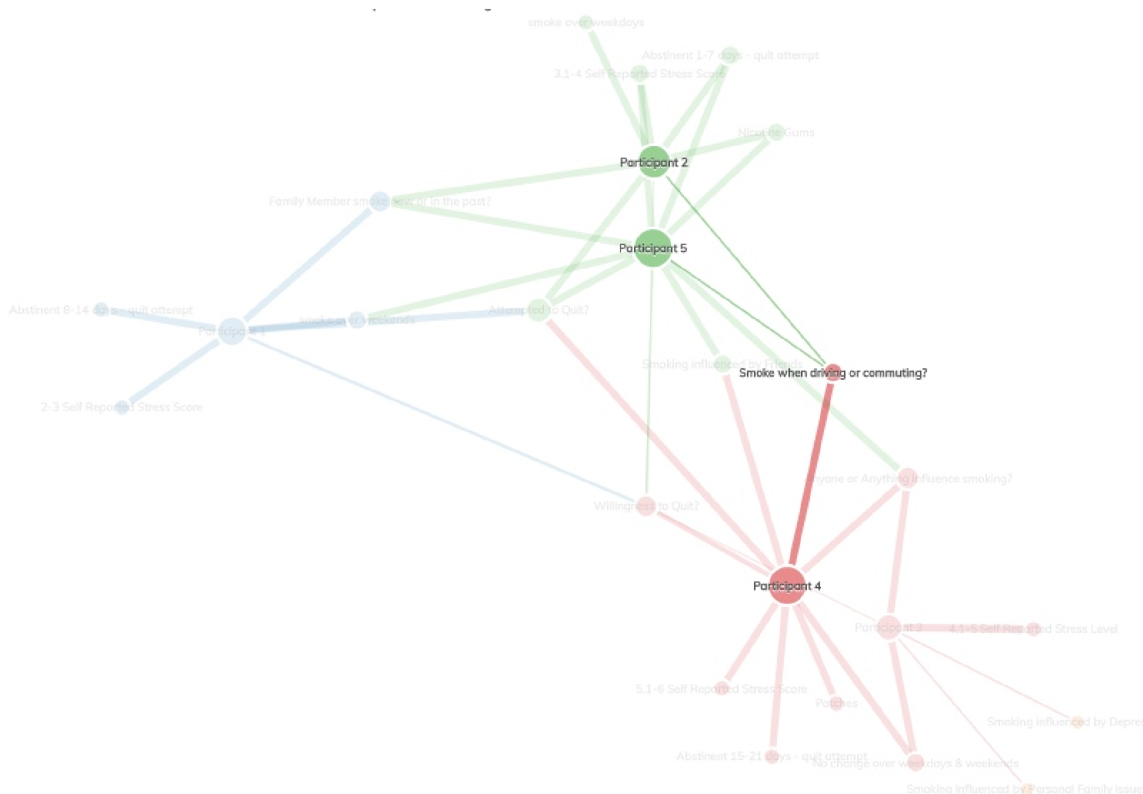


Fig. 13. Correlation graph intractability.

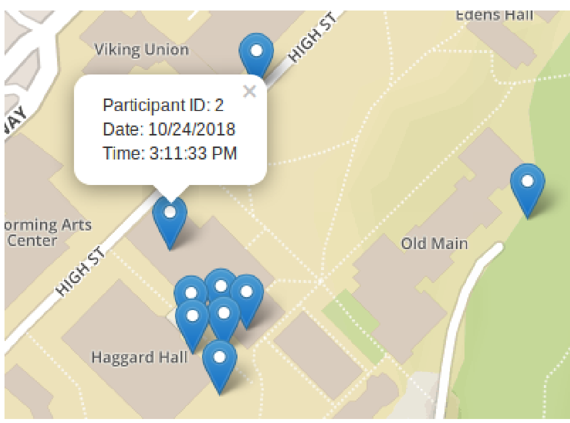


Fig. 14. Preliminary Map View

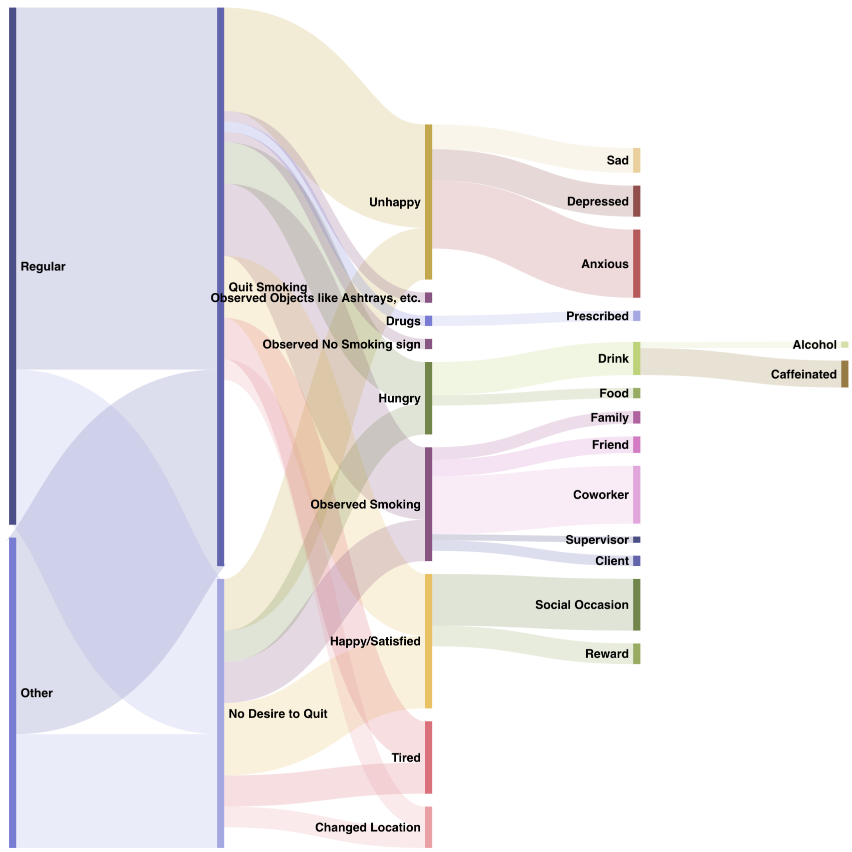


Fig. 15. Preliminary Sankey Diagram

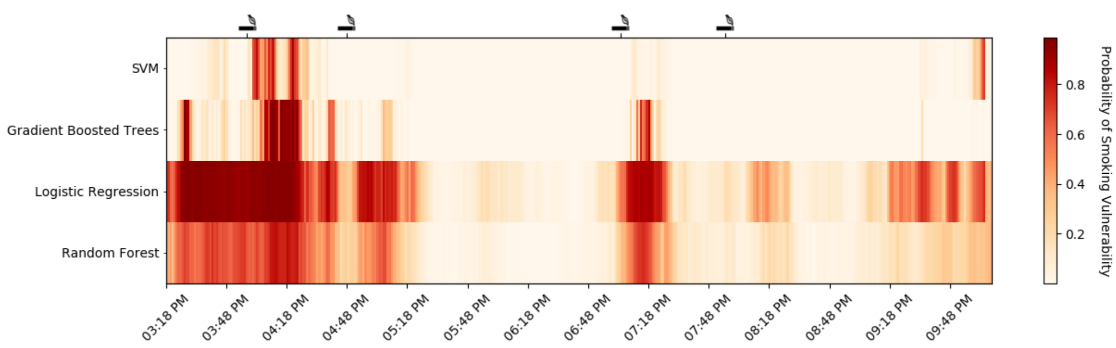


Fig. 16. Visualizing craving probability overtime for various machine learning models.