Exploring Demographics & Relationships in Engagement, Ohio

VAST Challenge 2022 Challenge 1: Demographics and Relationships

Ankit Basrur
Arizona State University
Tempe, USA
abasrur@asu.edu

Awani Kendurkar
Arizona State University
Tempe, USA
akendurk@asu.edu

Prasann Prashant Shimpi
Arizona State University
Tempe, USA
pshimpi@asu.edu

Rakshith Chandrashekar
Arizona State University
Tempe, USA
rchand44@asu.edu

Siddartha Gurugubelli
Arizona State University
Tempe, USA
sgurugub@asu.edu

Sidharth Dinesh
Arizona State University
Tempe, USA
sdinesh3@asu.edu

ABSTRACT

For the fictional town of Engagement in Ohio, we aim to infer a number of significant insights about the town's demographics and social relationships through respondents of a survey, which is a participatory urban planning exercise, and they are representative of the city's population. To assist with the community revitalization efforts and understand how to best allocate a huge city renewal grant, we developed visual analytics to make better sense of the city's neighborhoods and business bases. The target users of this interface will be the city planning committee as well as the city's residents. The tasks involved are to discover correlations between various attributes (like Age, Joviality, Interest Groups etc), explore trends between social interactions of the participants, characterize distribution of popular recreational spots in town, and identify the prevalence of businesses in the city. We accomplished these tasks using various visualization techniques, such as line chart, parallel coordinate chart, heat-maps, node-link graph, and an innovative radial-bubble-pie chart.

KEYWORDS

Visualization, Analytics, Data, JavaScript, D3.js

ACM Reference Format:

Ankit Basrur, Awani Kendurkar, Prasann Prashant Shimpi, Rakshith Chandrashekar, Siddartha Gurugubelli, and Sidharth Dinesh. 2018. Exploring Demographics & Relationships in Engagement, Ohio: VAST Challenge 2022 Challenge 1: Demographics and Relationships. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX).* ACM, New York, NY, USA, 7 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

This paper outlines the visualizations made for the VAST Challenge 2022 (Mini Challenge 1: Demographics and Relationships). We aim to use the data provided by the participants of a survey to visualize the demographics and relationships among the residents of the city of Engagement, Ohio [6]. We generate insights from the visualizations created using D3.js [3].

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/XXXXXXX.XXXXXXX

2 VISUALIZATION DESIGN

2.1 Line chart

Our Line chart visualizes the connection between average joviality and age among various interest groups.

- Marks: Line marks (polylines)
- Channels: X position indicates age, Y position indicates average joviality and the color of the line indicates a specific interest group
- Interactions: On clicking a particular line (interest group) updates the Parallel Coordinates Chart to represent information related to only that specific interest group

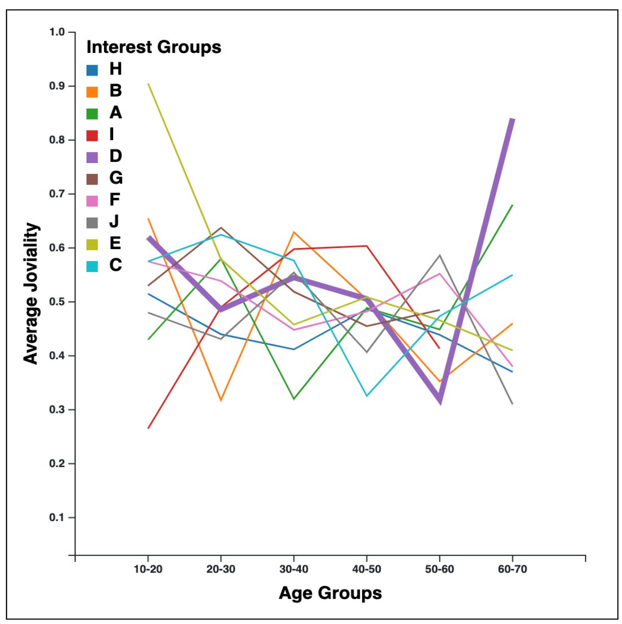

Figure 1: Line Chart

2.2 Parallel coordinates chart

This chart visualizes several distinctive trends among the participants in an interest group having attributes such as: Joviality, Age, Number of kids, Education Level and Household size, using multiple axes [2].

- Marks: Line marks (polylines)
- Channels: X and Y position for each axis represents the value for each attribute represented by the axes
- Interactions: Hovering over any line on any axis highlights the particular line (fading others behind it) and reveals a tool-tip showing relevant information about that participant. There is also a "brush" to filter points along any of the 5 axes.

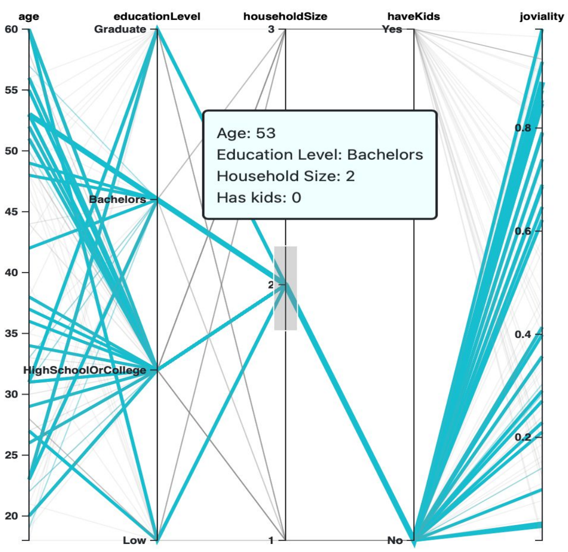

Figure 2: Parallel Coordinate Chart

2.3 Node-link graph

The Node-Link graph is a force-directed social network plot [10] which helps us understand the interactions between different interest groups during a particular date over the 15 month duration of the study and gives insight into compatible interest groups. The visualization changes with each day, showing increments/decrements in daily interactions.

- Marks: Node (or point) marks, link (or line) marks
- Channels: Color indicates a specific interest group
- Interactions: UI widgets such as interest groups in drop down and date selection using calendar icon lets user select range, and hovering over any node reveals a tool-tip with participant information

2.4 Scatter Map

A Scatter map has been implemented which visualizes the popular recreational spots in the city (restaurants and pubs) into the Engagement city's geographical map.

- Marks: Point marks

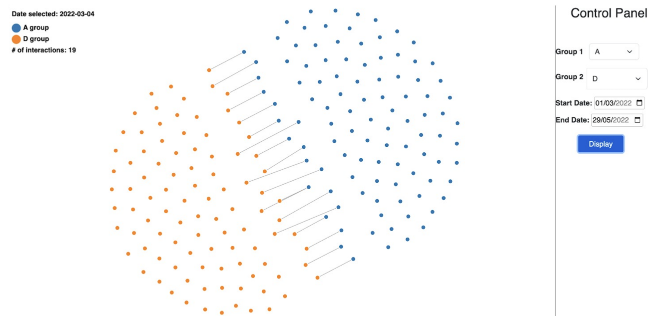

Figure 3: Node-Link Graph

- Channels: Spatial positioning roughly indicates the location of a recreation spot in the city, color luminance indicates the average cost, and the size indicates the popularity of that spot
- Interactions: Hovering over the bubble reveals a tool-tip indicating the average cost, popularity, maximum occupancy, and what type of recreational spot (pub/restaurant) it is. Clicking any bubble reveals a separate heat map view displaying further information regarding that particular spot in terms of origin of visitors in form of a Flow Map/Geographic Heat Map

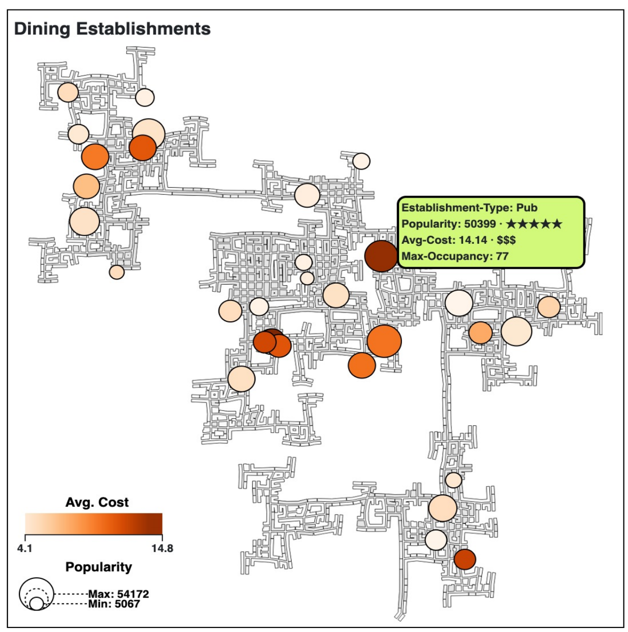

Figure 4: Scatter Map

2.5 Geographic Heat Map

This Geographic Heat Map is the first extension of the Scatter Map which visualizes visitor origin location [8]. This heat map helps us understand the population density of the visitor distribution.

- Marks : Area Marks

- Channels: Color luminance indicates the area from where most residents travel to this particular spot
- Interactions: Clicking on the Radio button for Heat Map will give us this view

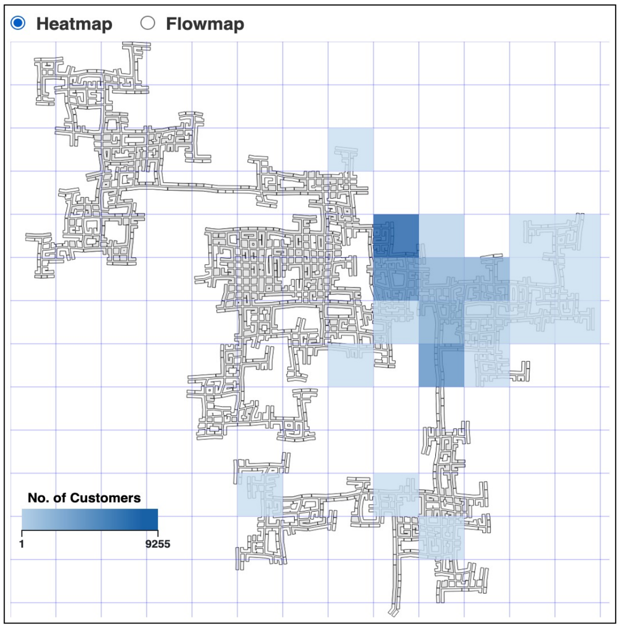

Figure 5: Geographic Heat Map

2.6 Flow Map

The second extension to the Heat Map is a Flow map visualizing the flow of the foot traffic i.e., the dots show the exact source location of the visitor's origin for that particular pub or restaurant.

- Marks: Point marks and lines
- Channels: Spatial positioning of dots represents the source location of the residents traveling to that particular pub or restaurant
- Interactions: Clicking on the Heat Map radio button will update the Flow Map to a Heat Map and vice-versa

2.7 Radial-Bubble Chart

Lastly, this innovative visualization aids in understanding the employment trends with a radial timeline [9] in the city. Each bubble is distributed and distinguished by color relative to its Education Level, along its month axis.

- Marks: Point (or area) marks
- channels: Position along the radial axis indicates the time, distance from the radial axis indicates education background with color, and the size of the bubble indicates the number of participants working under that employer
- Interactions: Hovering over the bubbles reveals a tool-tip showing more information regarding the employer. There is a radio button to select between views: either view all employers along that month axis, or view that employer only throughout the 15 months

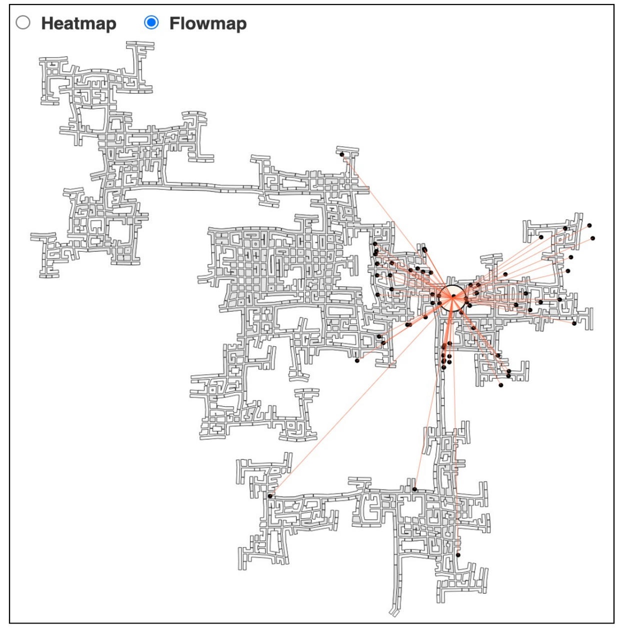

Figure 6: Flow Map

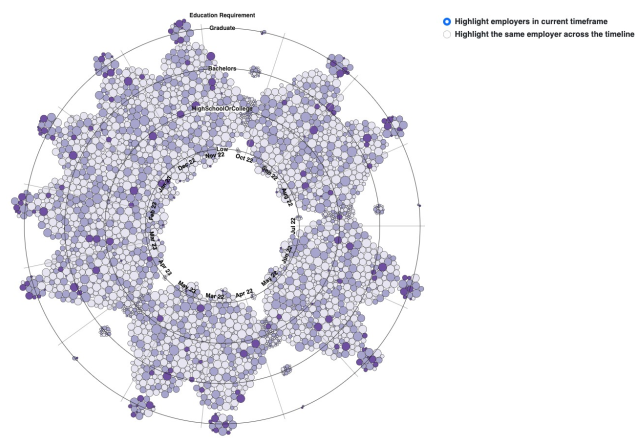

Figure 7: Radial-Bubble Chart

3 DESCRIPTION OF VAST MINI CHALLENGE

The VAST challenge we have chosen is described as follows: in the city of Engagement, Ohio, there is an anticipation of a surge in growth. People are rushing to the city having gotten word of its popularity. Amidst this, the city decided to conduct an exercise where approximately 1000 volunteers agreed to provide data using an app, which documents the places they visit, how much they spend there, whom they interact with, and what kind of jobs they work. This data is gathered to be used to decide how a large city renewal grant that they have recently received is to be allocated. Our team has supposedly joined the city planning team as visual analytic experts [4] to assist with this task. From social networks and other data, we are tasked with understanding the demographics and relationships of and among the city. We will aim to characterize the city's population, find patterns in the social networks of the city, and identify predominant business bases in the city [5].

4 DATASET DESCRIPTION

- Participants Dataset:
  Attributes:
  - interestGroup: Categorical, Cardinality = 10
    It ranges from A to J grouped with specific interests among the participants.
  - age: Ordered, Quantitative, Scale = 42
    It varies from 18 to 60.
  - haveKids: Categorical, Cardinality = 2
    It represents that the participant has kids or not.
  - educationLevel: Categorical, Cardinality = 4
    It indicates whether the participant pursued a bachelor's or high school or a graduate.
  - householdSize: Ordered, Quantitative, Scale = 2
    It indicates the number of members living in that particular house.

  Derived Attributes:
  - avgJoviality: Derived, Ordered, Quantitative, Scale = 1
- Pubs Dataset:
  Attributes:
  - pubId: Ordered, Ordinal, Cardinality = 12
    It is the ID given to that particular pub.
  - hourlyCost: Ordered, Quantitative, Scale = 8.423
    It is the amount that costs for the resident each hour, ranging from $6.417 to $14.840.
  - maxOccupancy: Ordered, Quantitative, Scale = 36
    It is the total occupancy of the pub, ranging from 60 to 96.
  - locationX: Ordered, Quantitative, Sequential, Scale= 7331.463
    It is the longitudinal value of the location.
  - locationY: Ordered, Quantitative, Sequential, Scale= 7814.385
    It is the latitudinal value of the location.
- Restaurants Dataset:
  Attributes:
  - restaurantId: Ordered, Ordinal, Cardinality = 20
    It is the ID given to that particular restaurant.
  - foodCost: Ordered, Quantitative, Scale = 1.85
    It is the cost of the food at the restaurant, ranging from $4.070 to $5.920.
  - maxOccupancy: Ordered, Quantitative, Scale = 71
    It is the total occupancy of the restaurant, ranging from 48 to 119.
  - locationX: Ordered, Quantitative, Sequential, Scale= 7331.463
    It is the longitudinal value of the location.
  - locationY: Ordered, Quantitative, Sequential, Scale= 7814.385
    It is the latitudinal value of the location.
- Jobs Dataset:
  Attributes:
  - jobId: Ordered, Ordinal, Cardinality = 1328
    It is the ID given to that particular job.
  - hourlyRate: Ordered, Quantitative, Scale = 90.00
    It is the wage given to the employee for an hour, ranging from $10 to $100.
  - employerId: Ordered, Ordinal, Cardinality = 1419
    It is the ID given to that particular employer.
- Travel Journal Dataset:
  Attributes:
  - purpose: Categorical, Cardinality = 5
    The purpose with which the participant has reached that place.
  - travelStartLocationId: Ordered, Ordinal, Cardinality = 1115
    It is an ID of the place from where the participant has departed. It can be a building, school, apartment, pub, restaurant, etc.
  - travelEndLocationId: Ordered, Ordinal, Cardinality = 1114
    It is the ID of the place where the participant has arrived.
- Social Network Dataset:
  Attributes:
  - timestamp: Ordered, Quantitative, Scale = 456
    The time at which they are communicating.
  - participantIdTo: Ordered, Ordinal, Cardinality = 963
    It is the ID of the participant who is receiving the communication (recipient).
  - participantIdFrom: Ordered, Ordinal, Cardinality = 963
    It is the ID of the participant who is starting the communication (initiator).
- ParticipantStatusLogs<n> Dataset:
  Attributes:
  - participantId: Ordered, Ordinal, Cardinality = 1011
    It is the ID given to a particular participant.
  - jobId: Ordered, Ordinal, Cardinality = 1328
    It is the job ID of a participant.
  - timestamp: Ordered, Quantitative, Scale = 456
    It is the time when an activity is logged.
  - financialStatus: Categorical, Cardinality = 2
    It describes the financial status of the participant, i.e., whether the participant is stable or not.

5 VISUALIZATION AS A SOLUTION

Our system proposes a design consisting of 5 unique visualizations, as a solution for the demographics and relationships challenge for the VAST 2022 Mini Challenge-1.
Our first visualization is a Line Chart which plots the relationship between average joviality and age for each of the 10 interest groups present in our dataset, which are differentiated using colors. On clicking any of the interest group's lines will filter out the remaining lines, by highlighting our selection in bold. This visualization helps us understand the happiness index (joviality) for the population, and how it correlates with age.

The Parallel Coordinates chart is our second visualization, which links multiple participant attributes such as: joviality, age, number of kids, education level, and household size. This chart is linked to our first visualization such that on clicking a particular line on our Line Chart, the subsequent attributes for that interest group is populated on our Parallel Coordinates plot. This linking helps us to view the backgrounds that the different age groups belong to, or

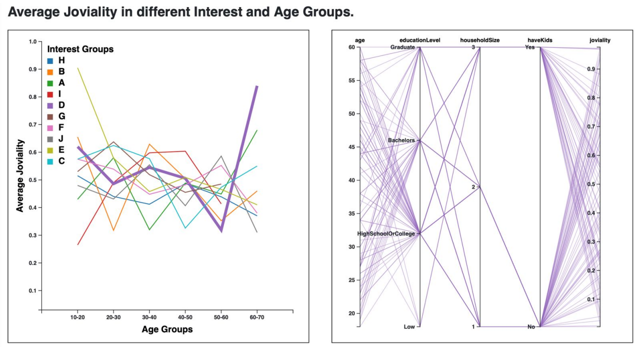

Figure 8: Joviality Trends

how their happiness may have a correlation with their household or educational background. It also opens questions for identifying other hidden patterns that exists for those participants. There is also a brush feature available which can help user filter axes points to see particular trends. This linked plot helps in answering: **characterize what you can about the demographics of the town.**

To understand the social interactions between the survey participants, we considered the social activities in the community and drew patterns in the social networks in the town, using a node-link network graph. In this force-directed network graph [1], every node denotes an individual participant, distinguished in color by its interest group, while links between nodes represent social connections logged at various timestamps. The drop-downs UI widget help us choose different interest groups, while the calendar icon widget lets the user choose the dates over the survey period of 15 months. The graph thus makes temporal trends in social gatherings visually apparent over the 15 months, allowing analysis both within and across different interest groups. It helps us in answering: **What patterns do you see in the social networks in the town?**

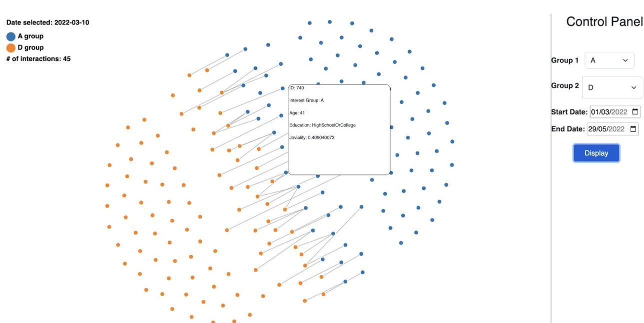

Figure 9: Social Interactions

As part of understanding the geographical aspects of the dataset, we have used several visualizations over the map of the city provided. Our primary spatial visualization is a Scatter Map which leverages the geographic layout of Engagement city to characterize distribution of popular recreational spots like restaurants and pubs. It plots visitor density patterns, where larger circles denotes higher popularity, and a darker shaded color of the circle denotes a high average cost. Clicking any of the venue circles will filter data and display a separate Flow Map OR Geographic Heat Map (user can select the view using radio buttons). This visualization helps us answering: **identify the predominant business base of the town** which is where population density is high and recreational spots are abundant (possibly Downtown of the city).

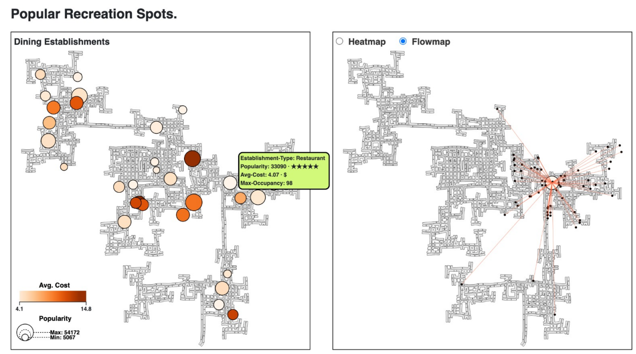

Figure 10: Scatter and Flow Map

As the first subpart of the main spatial plot, the Flow Map visualization for a selected venue, shows us the locations from where each restaurant visitor originated from, helping us understand which restaurants are popular that attracts different populations. A restaurant having longer flow map indicates its popularity that even people from far away places wish to dine out there.

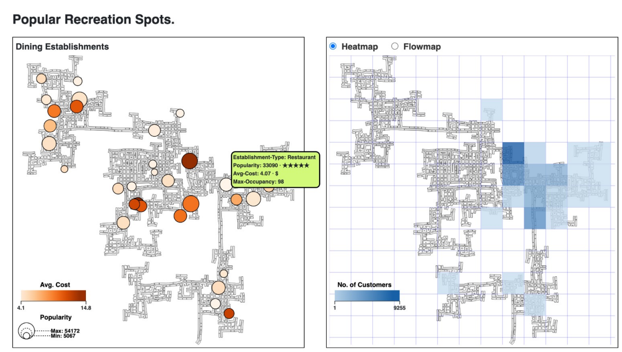

Figure 11: Scatter and Geographic Heat Map

Whereas the Geographic Heat Map provides an aggregate spatial view of the source locations from where participants travel to various recreational venues like restaurants and pubs. It is based on the same underlying map of Engagement city as the Scatter Map and Flow Map. This Heat Map visualization encodes source location frequency for all venue visits using a sequential color scheme ranging from light blue to dark blue. Areas with a higher density of source location dots appear darker, while fewer source locations yield a lighter shade.

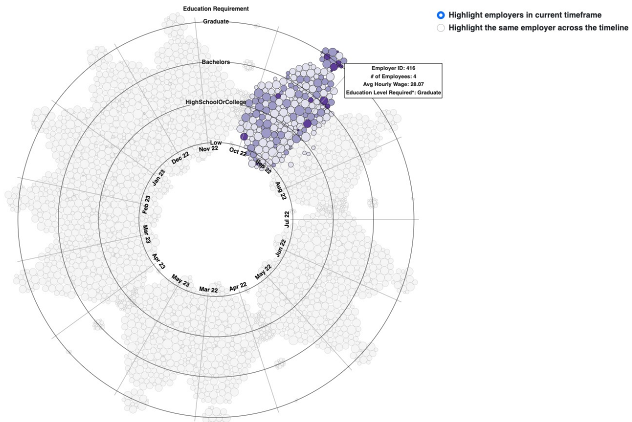

Figure 12: Employers in a particular month

Lastly, the Radial-Bubble Chart is our proposed novel visualization which represents employment and financial trends across the 15 month survey duration, with each bubbles across a month axis is an employer, and the size of the bubble denotes how many employees work under that particular employer. Throughout different months which are labelled radially, bubble sizes increase or decrease as per employees joining a new employer or leaving their current ones [7]. There are two additional views available on the choice of radio buttons: first when selected and hovered over a bubble will show all employers along that month's axis, while the second view visualizes that employer in all the months. If that employer size reduces or increases with different month, this helps understanding employment trends within that employer.

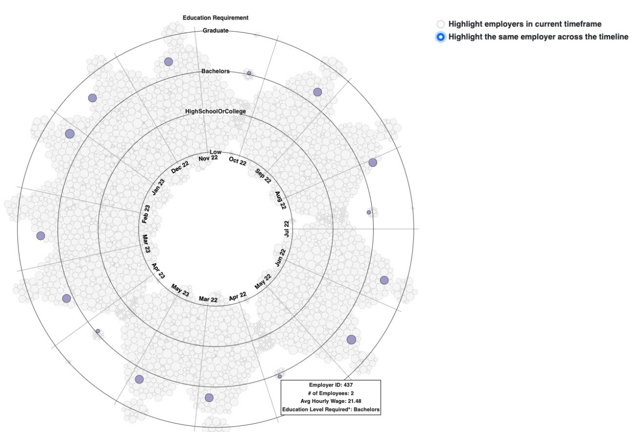

Figure 13: Employment Trends for a particular employer

6 DISCUSSION

After analysing the dataset with the help of our visualizations, we observe the predominant business base to be just north to the center of the map, with popular dining establishments as well as incoming high traffic from different corners of the city. This part of the city would be usually the Downtown area. Our Scatter Map view also helps us understand the quality and affordability of each restaurant, highlighting posh areas of the city.

With the help of our Social Network chart, we were able to understand links between different interest groups. We found that during some months of the year, the interactions were more between each interest groups, which could possibly indicate certain annual festivals or community get-together events. The timelines having lesser interactions would indicate off-season months, having lesser holidays.

Through our linked Line and Parallel Coordinate chart, we were able to view many underlying characteristics and patterns between attributes which were unexpected. The age v/s joviality plot for each interest group varied quite a lot, and visualising each interest group's attributes on the parallel coordinate chart showed us trends such that educated people were more likely to have more kids indicating better financial situation to support a family, while a majority of the high school holders had either no or 1 kid. The proportionality between education and financial situation was intuitive even though it was not an attribute, while linking it with household size was interesting.

Lastly, to understand the employment trends in the city of Engagement, our Radial-Bubble chart visualized employer population radially through the 15 month timeline. We observed that most of the month had similar employer size, certain people leaving their current employer for another, but the absolute difference remained more or less same. But it was interesting to find that certain months had unprecedented low employment. We believe this indicated possible mass layoffs due to poor market situation.

Future scope for this challenge can be improvements over our visualizations by:

- Utilizing lines of varying thickness with respect to the number of members present in that age group in the Joviality Trends Line Chart.
- Developing capability to incorporate interactions and animations other than axis brushing, enabling users the zoom in on specific subsets of data, for example, the entire visualization zooms in depicting only the plots for the subset for data in the Parallel Coordinates Plot.
- Implementing parallel axis drag and drops in the Parallel Coordinates Plot, which would give the user more control over the data presented.
- Having curved lines with thickness varying with population from the respective area in order to emphasize the influx of traffic to the recreational spots in the Flow Map.

7 TEAM MEMBERS

Our team consists of the following members:

- Ankit Basrur `abasrur@asu.edu`
- Awani Kendurkar `akendurk@asu.edu`
- Prasann Prashant Shimpi `pshimpi@asu.edu`
- Rakshith Chandrashekar `rchand44@asu.edu`
- Siddartha Gurugubelli `sgurugub@asu.edu`
- Sidharth Dinesh `sdinesh3@asu.edu`

REFERENCES

[1] Ulrik Brandes, Patrick Kenis, and Jörg Raab. 2006. Explanation Through Network Visualization. *First publ. in: Methodology 2 (2006), 1, pp. 1623* 2 (01 2006). https://doi.org/10.1027/1614-2241.2.1.16

[2] Prabhakar Chakraborty and Asoke Nath. 2014. Application of Parallel Co-ordinate System to Visualize Multivariate Data Sets. *idea* 1, 11 (2014).

[3] Wikipedia contributors. [n. d.]. *D3.js).* Technical Report. [Online; accessed 28-November-2023].

[4] Eric Hehman and Sally Y. Xie. 2021. Doing Better Data Visualization. *Advances in Methods and Practices in Psychological Science* 4, 4 (2021), 25152459211045334. https://doi.org/10.1177/25152459211045334

[5] Enamul Hoque and Maneesh Agrawala. 2019. Searching the visual style and structure of d3 visualizations. *IEEE transactions on visualization and computer graphics* 26, 1 (2019), 1236–1245.

[6] Jessica Hullman and Nick Diakopoulos. 2011. Visualization Rhetoric: Framing Effects in Narrative Visualization. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (2011), 2231–2240. https://doi.org/10.1109/TVCG.2011.255

[7] Francois Levesque and Thomas Hurtut. 2021. MuzLink: Connected beeswarm timelines for visual analysis of musical adaptations and artist relationships. *Information Visualization* 20, 2-3 (2021), 170–191.

[8] Carl M Sack, Richard G Donohue, and Robert E Roth. 2014. Interactive and multivariate choropleth maps with D3. *Cartographic Perspectives* 78 (2014), 57–76.

[9] Manuela Waldner, Alexandra Diehl, Denis Gračanin, Rainer Splechtna, Claudio Delrieux, and Krešimir Matković. 2019. A comparison of radial and linear charts for visualizing daily patterns. *IEEE transactions on visualization and computer graphics* 26, 1 (2019), 1033–1042.

[10] Hsiang-Yun Wu, Karsten Klein, and Da Yan. 2023. Effective Network Analytics: Network Visualization and Graph Data Management. *IEEE Computer Graphics and Applications* 43, 3 (2023), 10–11. https://doi.org/10.1109/MCG.2023.3267210