CSE 578: Data Visualization (2023 Fall)

2022 VAST Challenge: Engagement, Ohio
Challenge 2: Patterns of Life

Amey Bhilegaonkar
abhilega@asu.edu

Apoorv Kakade
akakade@asu.edu

Rajat Mahajan
rmahaj13@asu.edu

Sanket Dumbare
sdumbare@asu.edu

Abir Bhattacharjee
abhatt88@asu.edu

Nishtha Wagh
nwagh2@asu.edu

Problem Description:

The VAST 2022 Mini Challenge 2 considers the patterns of daily life throughout the city of Engagement, Ohio USA. The objective of this challenge is to analyze the provided data sources through visual analytics to successfully complete three designated tasks. These tasks involve characterizing distinct areas within the city based on participants' daily routines, visualizing the busiest areas and identifying traffic bottlenecks, and describing the daily patterns of two selected participants with supporting evidence. Additionally, the challenge requires an examination of how city patterns change over time and seasons, with a focus on identifying significant transformations.

To address these tasks, we will utilize both a Nested Design Model (NDM) and a 5-design sheet methodology. This comprehensive approach will involve a range of analytical techniques, including clustering, heatmap generation, time series analysis, and trend analysis. Through these methods, we aim to extract valuable insights regarding daily life patterns, traffic dynamics, and the city's changing patterns over time.

Domain Abstraction:

Who are the target users?

The target users are as follows:

a. Digital delivery/hailing services - Companies like Uber, Lyft, Doordash, Walmart rely heavily on real time geolocation data (often along with demographic data) to provide end to end services. These patterns help them with estimation of delivery time, anticipating incoming requests and timely fulfillment of service.

b. Law enforcement agents - Law enforcement officials need to be cognizant of population distribution at any given time to manage traffic or congestion as well as to respond better to criminal activities. The visualizations we plan on implementing will help them show population density across different zones at different time frames.

c. Urban Planners (Policy makers Construction Companies) - Setting up establishments, be it an apartment or a restaurant or a flyover, requires understanding of the geographic data coupled with user patterns in and around the site. The visualizations will help them understand their potential consumers' behavior and thus strategize better.

Data & Task Abstraction

Task Abstraction

Why is the target user looking at the visualization?

Task 1: Characterize Distinct Areas in Engagement City
- Discover distincts areas within the city based on daily routines of representative individuals to understand their activities and lifestyle.

Task 2: Identify Busiest Areas of City and Traffic Bottlenecks
- Determine the busiest areas in the city and traffic bottlenecks to expedite the deliveries, the arrival of emergency services, manage traffic, etc.

Task 3: Represent Daily Patterns of Two Participants
- Recognize daily activity patterns of two representative individuals with different routines to showcase diversity in lifestyle and daily activities within the city.

Task 4: Inspect Temporal Patterns

- Analyze the patterns in the city that change over time to understand dynamics of daily life, mobility and trends.

Data Abstraction

This data abstraction serves as a foundation for understanding how to leverage the provided data to address the tasks discussed in task abstraction. For instance, locations mentioned corresponding to pubId, restaurantId, apartmentId are utilized to recognize distinct areas in the city.

The dataset provided is a static tabular form as it's collected over a time period of 15 months and provided in multiple CSV files.

ParticipantStatusLogs<n>.csv

This csv file contains information on the status of ~1000 participants over a 15 month time period at an interval of 5 minutes. As there are 72 files, <n> in the filename is replaced with a number in range (1, 72).

=> Number of DataPoints : Approximately 114M rows

Attributes:

Attributes	**Type**	**Range/ Cardinality**
timestamp	Ordered - quantitative	[2022-03-01, 2023-05-24]
currentLocation	Categorical	26753
participantId	Categorical	[1, 1010]
currentMode	categorical	5
hungerStatus	categorical	5
sleepStatus	categorical	3
financialStatus	categorical	3
availableBalance	Ordered - quantitative	[-1000, 250000]
dailyFoodBudget	Ordered - quantitative	[12, 20]
weeklyExtraBudget	Ordered - quantitative	[0, 90000]

Apartments.csv : contains information about residential areas in the city.

Datapoints: 1517

Attributes:

Attributes	Type	Range/Cardinality
apartmentId	Categorical	1517
location	Categorical	1214

Restaurants.csv : contains information about restaurants within the city.

Datapoints: 12

Attributes:

Attributes	Type	Range/Cardinality
restaurantId	Categorical	20
location	Categorical	20

Pubs.csv : contains information about pubs within the city.

Datapoints: 12

Attributes:

Attributes	Type	Range/Cardinality
pubId	Categorical	12
location	Categorical	12

Schools.csv : contains information about schools within the city.

Datapoints: 4

Attributes:

Attributes	Type	Range/Cardinality
schoolId	Categorical	4
location	Categorical	4

TravelJournal.csv : contains information and motivation about participants' movement in the city
#Datapoints: ~21M
Attributes:

Attributes	Type	Range/Cardinality
participantId	Categorical	1011
purpose	Categorical	5
travelStartLocationId	Categorical	1114
travelEndLocationId	Categorical	1114

Idiom Abstraction

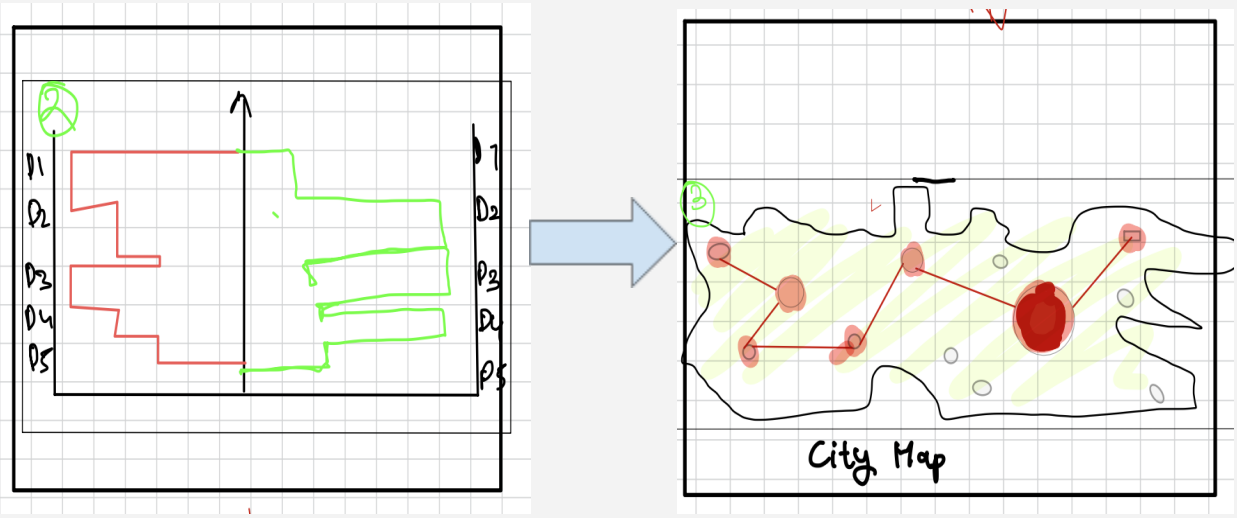
How is it shown? How to draw and manipulate ?

We will display a base map of the city divided into zones based on its density of establishments. A zone which is predominantly just apartments will be marked as a residential zone whereas a zone with a lot of schools will be classified as a school zone, so on and so forth. The legend to this chart will mention the area covered by the zone. On clicking on a particular zone, an interactive multi-line chart will be displayed showcasing trends in behavioral patterns of participants in that particular zone across time (hourly) A calendar will be added to check patterns on a different date.

Next a heat map will be displayed to demonstrate the population density of participants across the city at any given time, this is to visualize traffic patterns and bottlenecks as well as to observe hourly, daily, seasonal trends in traffic patterns. On clicking on the map, a pie chart will be displayed that breaks down the current activity of the population at that specific geographic area, thus helping to further analyze the dense areas - whether it's a traffic bottleneck or just a corporate area where a lot of employees are located.

Finally we analyze two users' daily patterns using back-to-back bar charts - this visualizes time spent in each activity by the users. This contrast helps analyze their daily behavior. A calendar will also help adjust time so that behavior on different days can be compared. On clicking a specific user's bar, an enhanced version of Bubble Map Chart will be displayed which shows

their trajectory across a specific time period and how much time they have spent in a specific location, the amount of time will be visualized by the size of bubbles.



Algorithm Abstraction

We can divide Algorithm Abstraction into various stages:

Stage 1: **Data Pre-processing**

a. **Data Cleaning**: Remove missing or irrelevant data, handle outliers, ensure usable format for improved quality and reliability.
b. **Feature Engineering**: Create or modify features for better analysis performance, including time interval aggregation and attribute extraction.

Stage 2: **Characterizing Distinct Areas using Clustering**

To identify distinct areas within the city, we employ clustering algorithms like k-means or DBSCAN. These algorithms group participant locations based on their daily routines, helping us discover regions with similar activity patterns

Stage 3: **Identifying Busiest Areas by Heatmap Generation**

We identify the city's busiest areas by creating heatmaps using participant's location data over time. Heatmaps visually depict population density, highlighting areas with high traffic. In parallel, traffic analysis algorithms like Shortest Path or Network Analysis assist in recognizing traffic

bottlenecks, optimal routes and congestion spots. This enables us to pinpoint congested locations and potential traffic concerns.

Stage 4: **Time Series Analysis - Representing Daily Patterns**
We pick two participants with unique daily routines and analyze their patterns using time series analysis. This includes monitoring their visited places, transportation modes, and activity status (like work or leisure). Visualizations, such as daily routine charts, depict their activities and transitions throughout the day.

Stage 5: **Visualizing Temporal Patterns**
To monitor changes in city patterns over time, we employ trend analysis techniques on various attributes. This includes using methods like moving averages and regression to quantify and understand these trends. For instance, we can track shifts in population density, transportation behaviors, or specific attribute variations across seasons. These methods allow us to effectively describe significant changes in the city's patterns over time.

In conclusion, our focus revolves around visualizing and analyzing the data to extract meaningful insights. We employ techniques such as clustering, heatmap generation, time series analysis, and trend analysis to provide a comprehensive understanding of the city's daily routines, traffic dynamics, individual behavior patterns, and their evolution over time.

Innovative View:

We have an innovative view that consists of an enhanced version of a Bubble Map Chart which not only highlights the amount of time spent in a specific area by a user but also connects those geolocations to demonstrate the trajectory of the user in the given timeframe. This gives a good understanding of the two facets of a person’s pattern space (where was the user in that time frame) and time (how much time did the user spend there)

SHEET 1

Ideas:

1. Using travel.csv can show what are the most frequent visited locations? → circle shape graph depending on density!
2. What are the type of people visit these locations? → Bar charts showing status of people
3. How does visits to location vary in day →
4. Hourly filter showing heatmap with tooltip showing counts of people type.
5. Zone analysis, using squares what are type of buildings in that area → mark that zone!
   ↳ tooltip no of building types
6. Onclick zone, we can show in new graph, using heatmap / donut chart pie chart showing population density / building types / people type
7. Innovative / Novel graph, → plotting two users data on Base map using size, color and position. to show, daily routine pattern of two people.
8. We can show, spider-graph of two people, on click on trajectory graphs of two.
9. Fibonacci series / golden number graph
10. Stacked bar chart to show time spent in a day.
11. Scatter plot to show daily routine.

---

filter:

1) Discard scatter plot as there are problems with overlapping data points.

2) Pie chart only give vague overview of person routine.

Combine and Refine:

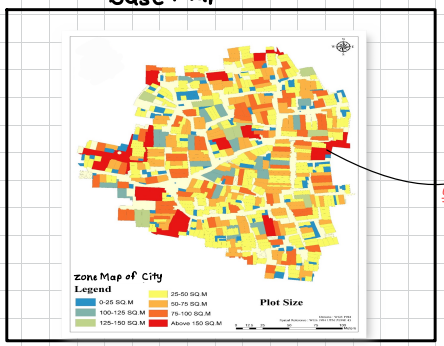
1) Combine bar chart & pie chart to make stacked bar chart.

Questions:

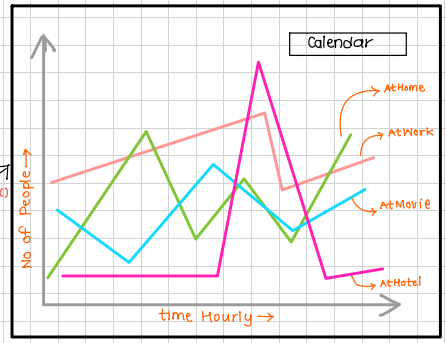
1. To what level do we need categorize people?
2. How are we going categorize them?
3. How to connect trajectory plots with novel plot?
4. How to do zonal analysis?
5. What are heatmaps for population density or distribution?
6. What would be the minimum unit for time scale for animation?

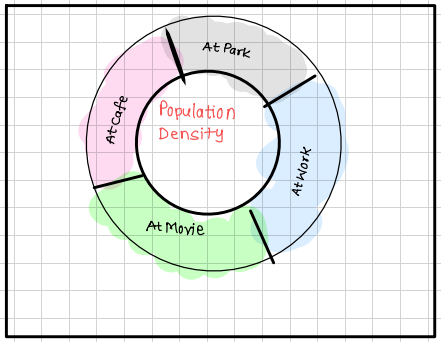
Layouts :

Base Map

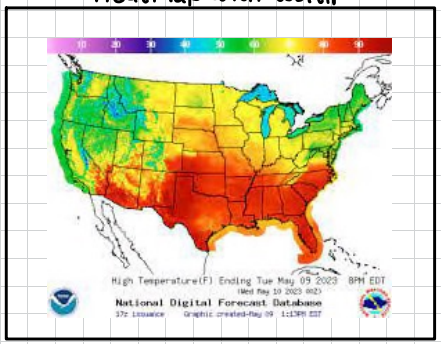


onClick()

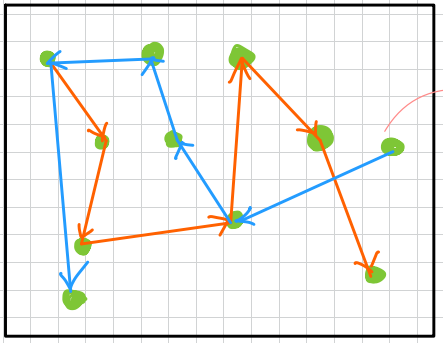


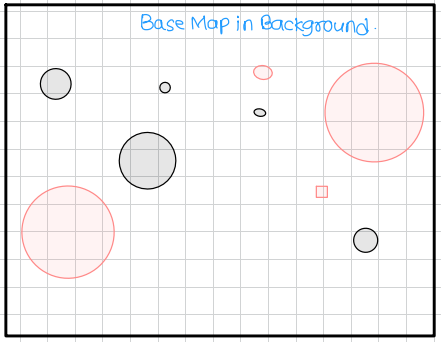


HeatMap with tooltip (No. of people)



Two user trajectory





Information :

SHEET 2

Analysis for user patterns for life

Operation :

1. Onclick on BaseMap which is a zone map, a multiline graph of No of people v/s hourly time should be plotted.

2. A tooltip on heatmap should display no. of people in the concentrated area.

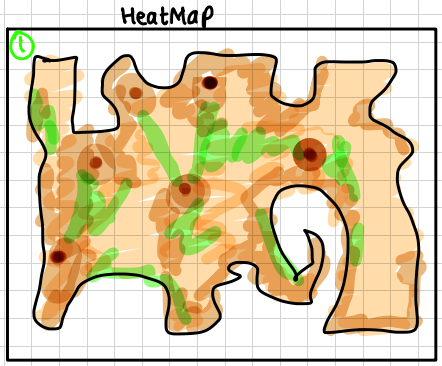
Discuss :

1. Discussed about multiple base maps & selected the zone map as the final base map as it provides a bird eye view of the problem.

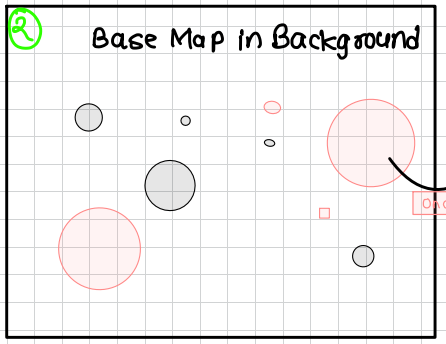
Focus and Partitions :

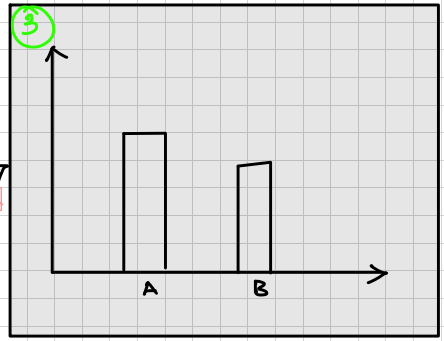
Trajectory & line graph should deliver more information.

Layouts:

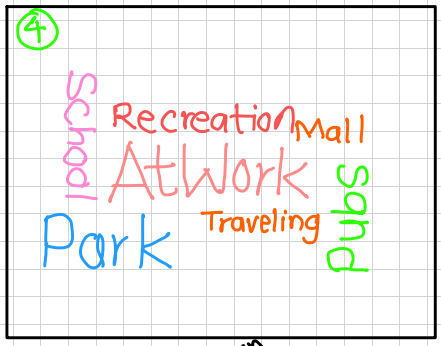


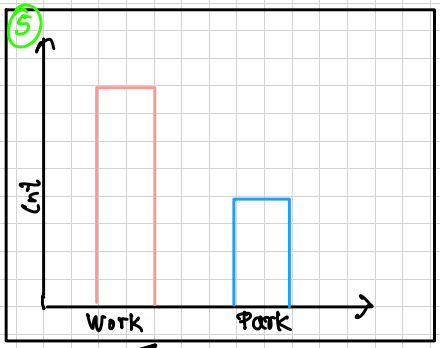
Distribution of population at Time T for x category.



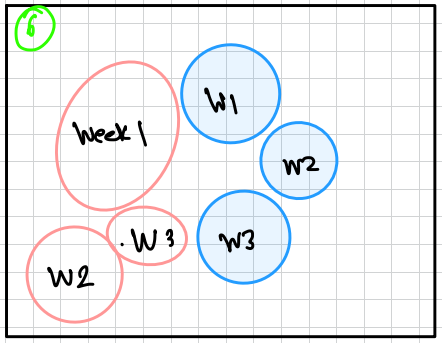


Bar Graph of two users of category 'X'





OR



Focus and Partitions:

On clicking the bar graph, weekly distribution would be shown using bubble chart.

Information:

Analysis for user patterns for life

SHEET 3

Operation:

On click chart 2, chart 3 should appear.

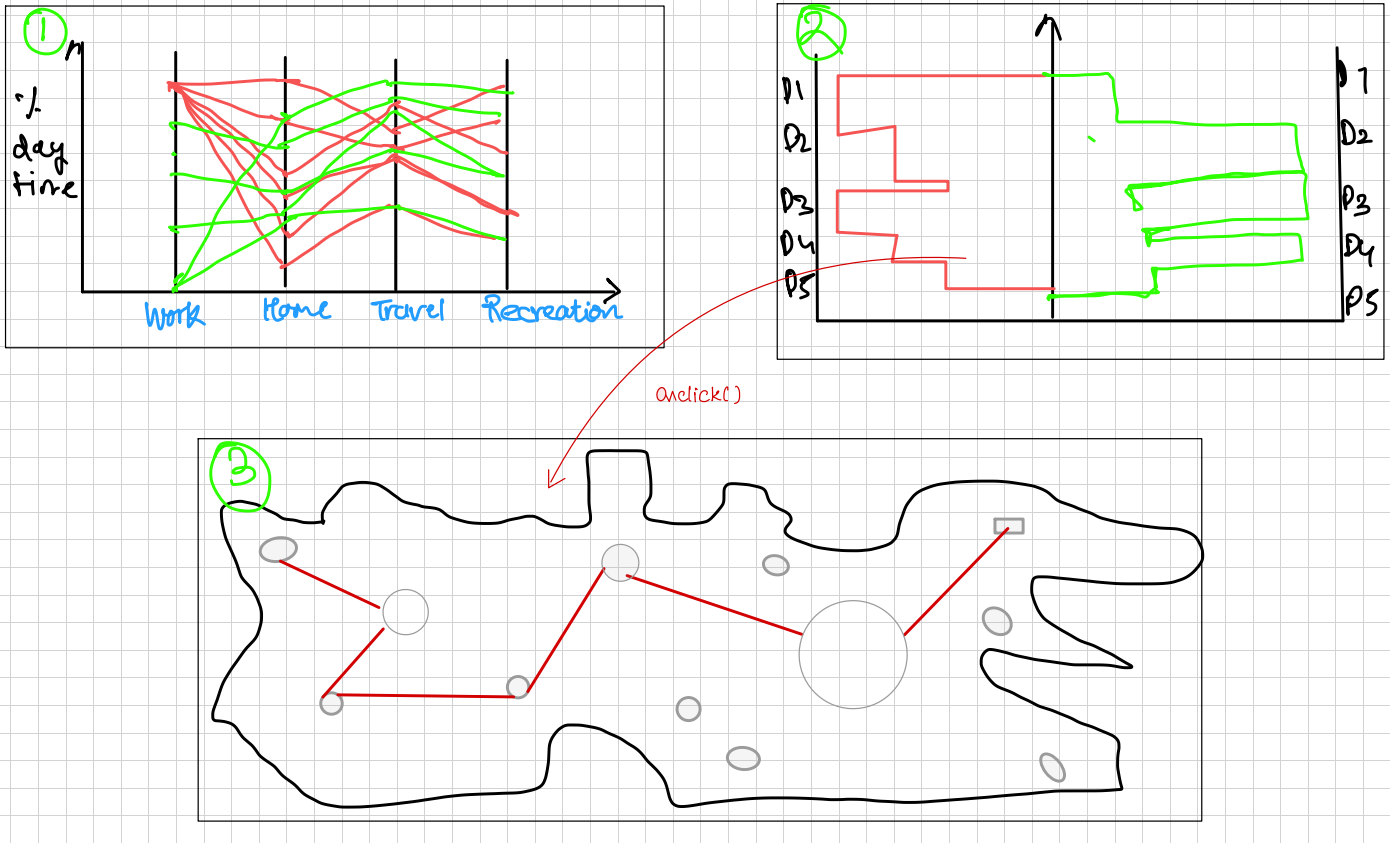
On click chart 5, chart 6 should appear.

Discuss:

1) We can either have word cloud or bar graph to represent where the participant spends their time in a week with the UI element of varying months.

2) We can also have on-click on a bar or word to give more details on the selected category.

Layouts:



Information:

SHEET 4

Analysis for user patterns for life

Operation:

1) On click graph 1, one line gets activated which describes activity of that participant on any given day.

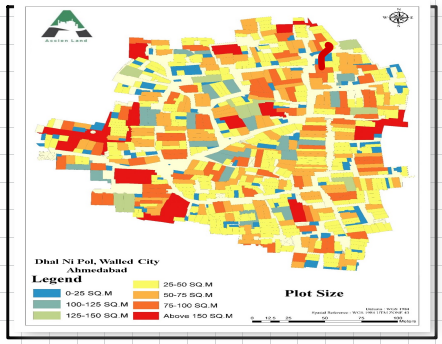
Discuss:

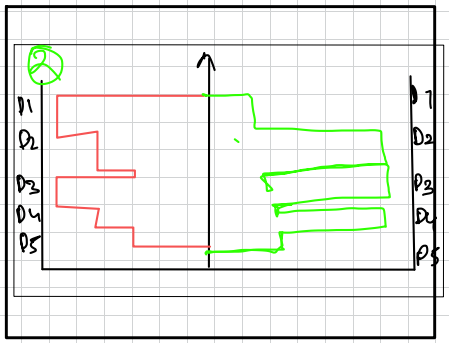
In graph 3 the size of the circle represents time spent by the user selected in graph 2. (Animated).

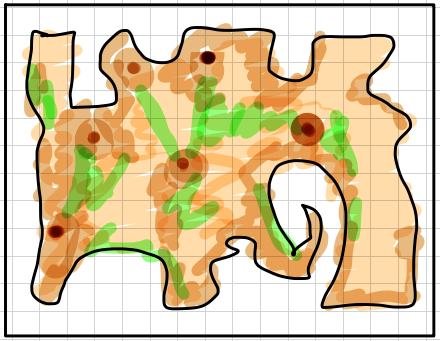
Focus and Partitions:

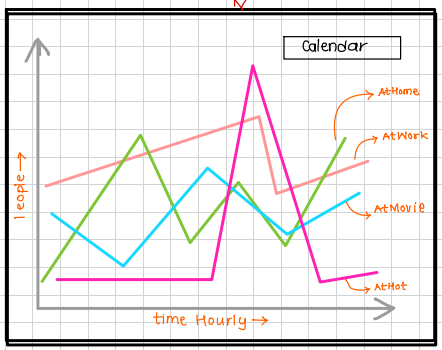
Focus is on participant trajectory & size of circles (Time spent). This will give good understanding about person's routine.

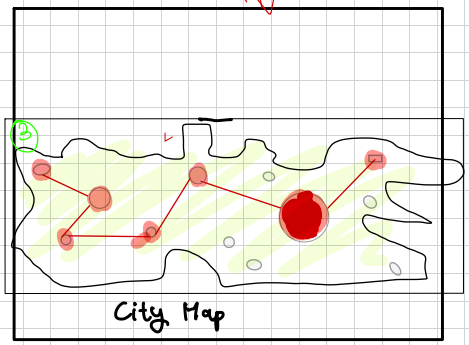
Layouts :

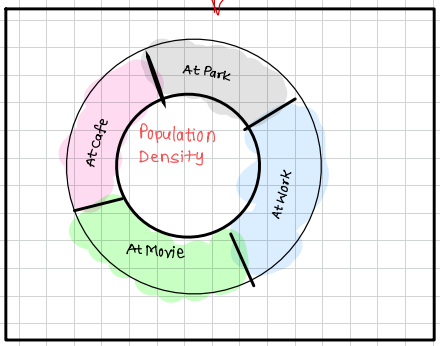












Information :

SHEET 5

Final Patterns of life of people in a city.

Operation :

1) on click on row 1 graphs generates respective row 2 graphs;

Detail :

1. Finding sections / zones for city based on geography.
2. Comparison of patterns for 2 persons with different routines for a week.
3. Analysing bottlenecks for the city at specific time with the help of population density.