Visual Analytics for Engagement: an Urban Analysis Task

"Visual Analytics" course, a.y. 2023/24

Professor: Giuseppe Santucci

Candidates:

Alessio Palma, palma.1837493@studenti.uniroma1.it

Antonio Andrea Gargiulo, gargiulo.1769185@studenti.uniroma1.it

VAST Challenge and Task

Challenge Context: We tackled the **VAST Challenge 2022** focused on Urban Planning of the fictitious city of Engagement, Ohio, USA.

Objective: The city council seeks the best visual analytics application to decode community patterns for strategic allocation of a large city renewal grant.

Intended User: Town council members responsible for making informed decisions on economic interventions for city improvement.

Focus Areas: Our solution addresses the following crucial city challenges:

1. Demographic Patterns Analysis

2. Characterization of Different City Areas

3. Understanding Business Landscape

4. Assessment of Living by Quality and Costs

Dataset & Domain

VAST Dataset

- **Dataset Source**: VAST Challenge 2022, focused on analysis of urban planning for Engagement.

- **Data Generation**: Artificially generated by authors, tracking the information and actions of 1011 residents over a period of 15 months.

- **Dataset Size**: Totals 18 GBs of data, divided into three main folders:

  - **Attributes folder**: Contains 9 CSV files with static information about key city entities (participants, buildings, businesses, etc.).

  - **Activity Logs folder**: Includes 72 CSV files detailing participants' financial, hunger, sleep status, and location every 5 minutes.

  - **Journals folder**: Consists of 4 CSV files summarizing activity logs into financial, social, traveling, and check-in information.

Data Preprocessing
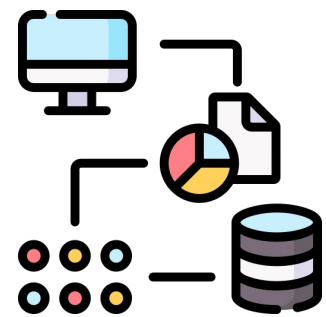
- Before diving into analytics, data preprocessing is crucial.

- So we performed extensive data preprocessing using Python to shrink the size of the huge dataset and make it computationally tractable.

- Our final dataset has an **AS index** of 63386 and is organized into 5 CSV files.

- Through data aggregations and derivations, we've structured our dataset for insightful analytics, laying the foundation for our subsequent analyses.

Data Preprocessing

- **ParticipantsAugmented:** contains static information about the 880 active participants remaining after cleaning and is used for all the plots in the participants view. Some attributes were aggregated from the Journals (e.g. total expenses for each category for each person) or derived, like the Engel's coefficient which is an indicator of the standard of living and is computed for each person as:

$$\text{Engel's coefficient} = \frac{\text{Total food expenses}}{\text{Total personal expenses}}$$

Data Preprocessing
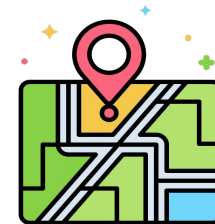
- **BuildingsAugmented:** contains GeoJSON formatted data about city buildings shapes, used to draw the map, and the type of the building (Commercial, Residential, School).

- **ActivitiesAugmented:** static information about recreational activities, it's the main source of data for the activities view.

- **MonthlyLog:** contains monthly earnings data for each activity.

- **VisitsLog:** includes participants visits count to each activity and L2 distance between their residence and the activity location.

Related Works

Related Works

1. **<u>ClusBridges</u>** focuses on analyzing traffic data and patterns of life, clustering individuals with similar routines. *However, it skips city economics and demographics.*

2. **<u>Comprehending City Economics from Heterogeneous Data</u>** emphasizes economic analysis of participants and activities separately, *but lacks spatial brushing and demographics information, obstructing detailed insights.*

3. **<u>EconomicVis</u>** offers a comprehensive visual analytics tool for city economics and daily life patterns, showing a lot of data *but lacking visual efficiency and spatial brushing.*

Related Works

4. **Visual Analytics for Demographics, Social Networks, and Business Base Pattern** focuses on the demographic characteristics, social relationship patterns and main business area patterns, *but only allows the user to apply some predefined filters on the data.*

5. **Visual Analytics for Urban Data Analysis** provides a view with the map of the city and detailed dashboards covering demographics, social networks, business, and gastronomy, *yet it lacks map brushing in most dashboards and ignore earnings data in gastronomy.*

6. **Analyzing demographics and patterns-of-life using SAS Visual Analytics** offers a comprehensive analysis about demographics, traffic data, life patterns and more, *but uses automatic and closed-source tools and lacks multiple coordinated views interactions.*

Application overview

Dual Views: Our solution offers a dual-view approach, enabling focused analysis on demographics and expenses (*"participants view"*) or earnings and recreational activities (*"activities view"*) as per user demand.

- **Participants View:** Offers a panoramic of the city's demographic landscape, revealing diverse factors such as age, income, education, happiness, and family composition.
- **Activities View:** Describes economic and sociality trends, providing a snapshot of citizens' recreational preferences, eating habits, and entertainment choices.

Application Overview
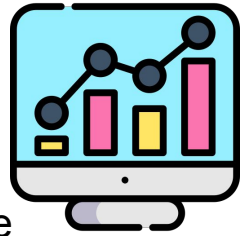
Insights: By analyzing these patterns, the user gain an understanding of the city's economic vitality and consumer behavior.

Technology Stack: Developed using Node.js, D3.js, and a Python backend for analytics.

Interactivity: Highly interactive interface allowing a pop-up tooltip, brushing on both map and parallel coordinates plots, and the ability to select clusters for analysis.

Smooth Updates: Each interaction triggers plot updates, ensuring fluid transitions and preventing change blindness with them.

Visualizations

Map Chart

Engagement, Ohio, USA
Participants selected: 880
Avg Engel's coeff: 0.26
Avg Joviality: 0.47

- Participant
- School
- Commercial
- Residential
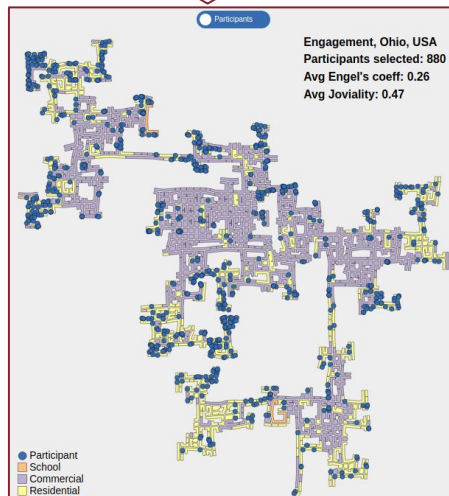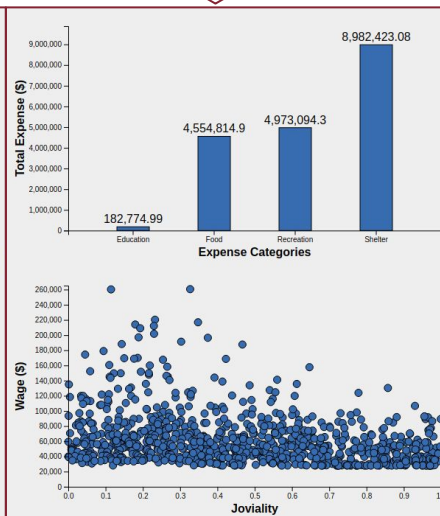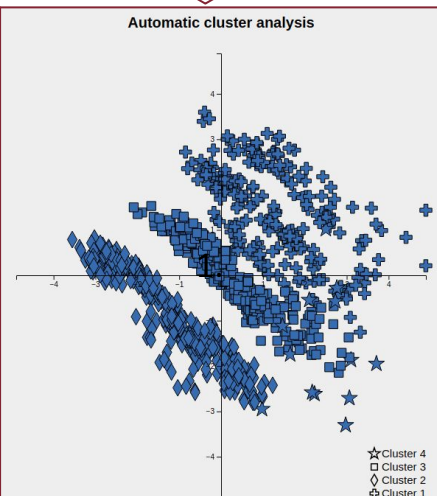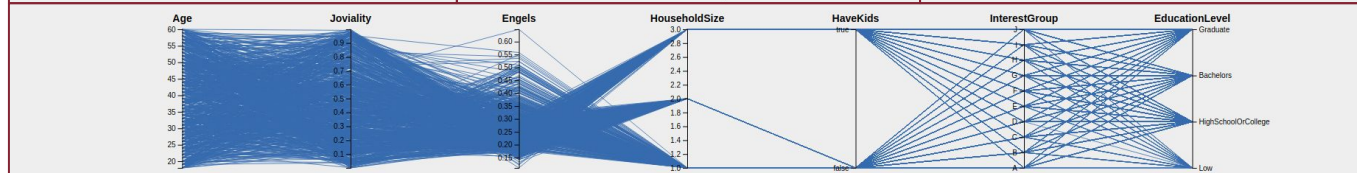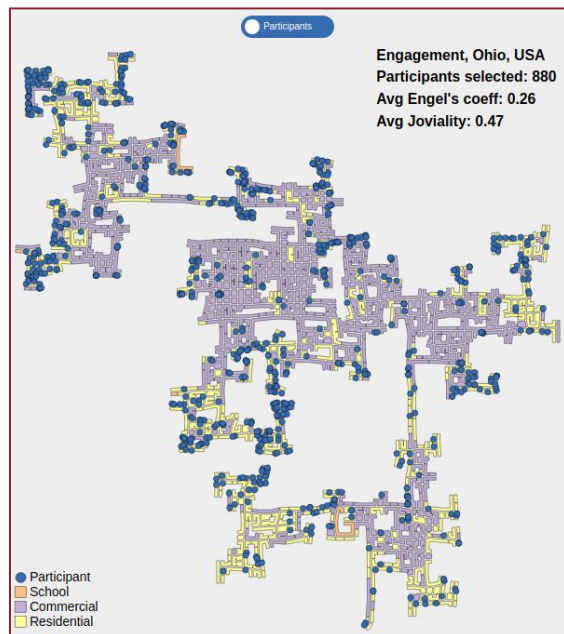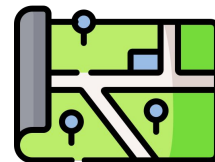
Participants view

Provides geographic information about the city.

Color + Shape Encoding for easy identification of buildings/subjects. Legend for clarification.

Toolbox: displays aggregated statistics about the current selection.

Completely **brushable:** enables users to perform focused analyses on specific areas of the city.

Tooltip: Hovering over dots reveals a tooltip with the entity ID for easy reference.

Engagement, Ohio, USA
Activities selected: 32
Percentage of Turnover: 100.0%

- Pub
- Restaurant
- School
- Commercial
- Residential
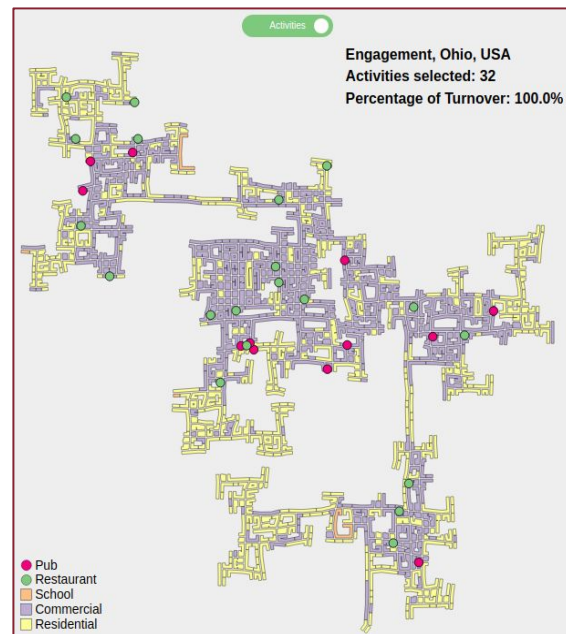
Activities view

Central Area Charts

Two Central Plots per view provide aggregated economic data, offering focused insights into specific behaviors.
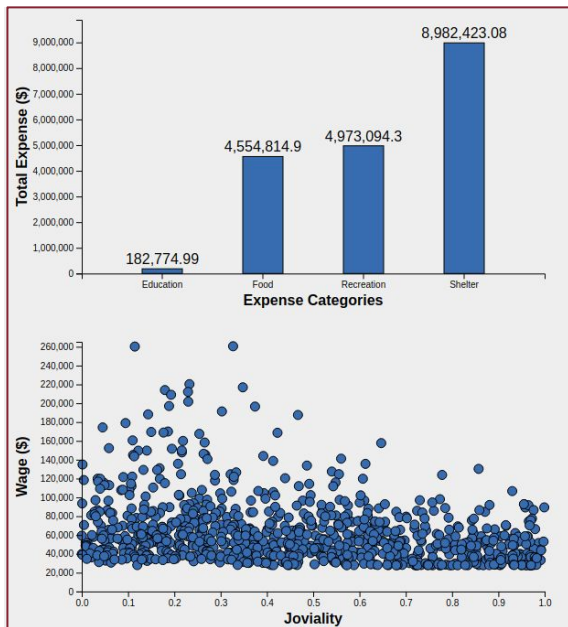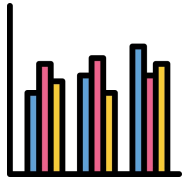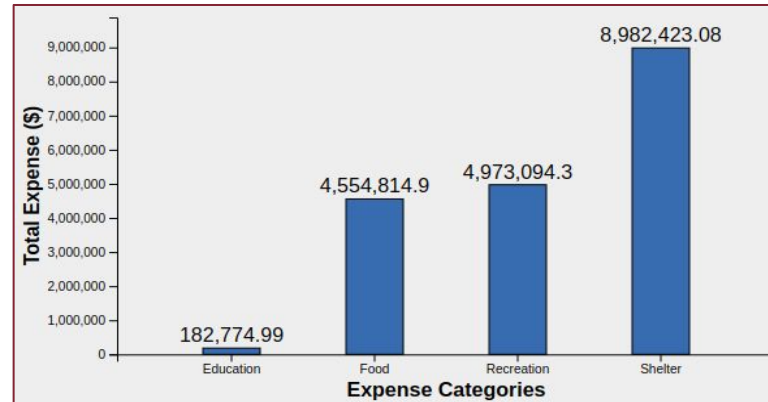

Participants view
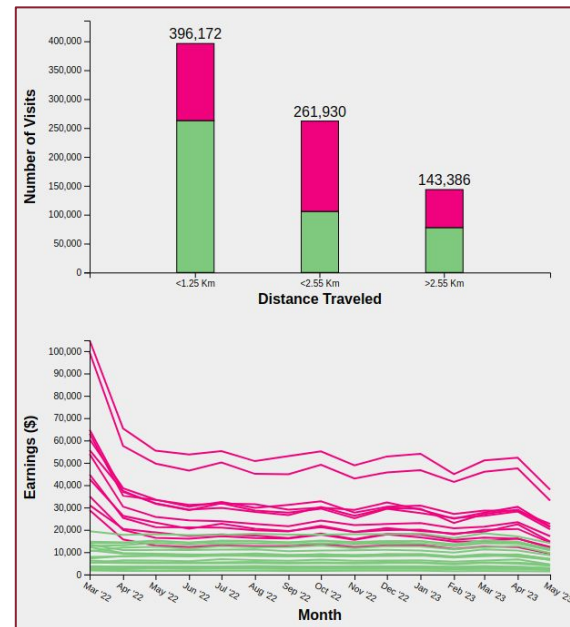
Activities view

Participants view - Bar Chart

The **Bar Chart** displays total expenses of participants across many categories:

- Provides an overview of money flow and spending preferences in the city.
- Numerical values on top of each bar to aid the understanding.
- Updates dynamically when brushing on other plots.
- Allows insights about users preferences across different activities, neighborhood spending habits and saving tendencies.

Participants view - Scatterplot
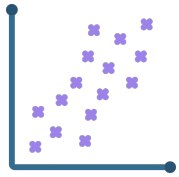
The **Scatter Plot** compares joviality with total income of participants over the 15 months:

- Explores the relationship between well-being and income.
- **Highlights**:
  - Suggests a minimum wage benchmark.
  - Indicates a potential negative correlation between income and happiness.
- Dynamically updates when brushing on other plots, displays entity's ID on mouse hovering.
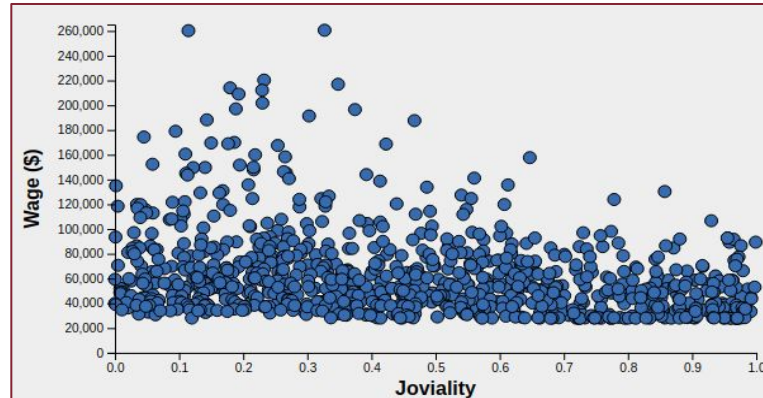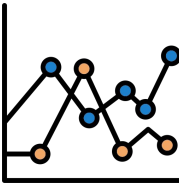
Activities view - Stacked Bar Chart
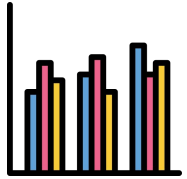
The **Stacked Bar Chart** displays distribution of distances traveled by customers to reach businesses.

- L2 distance from participant home to activity location computed for each activity visit.

- Stacked format reveals percentage contribution of restaurant and pub visits to each recreational activity.

- Indicates popular activities based on distance traveled, aiding in determining investment areas.

- Updates dynamically when brushing on other plots, providing insights into trending activities.
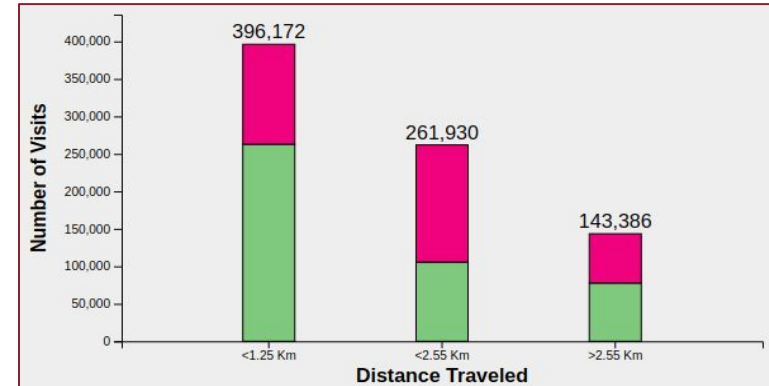
The **Multiple Lines Chart** illustrates monthly income variation for each recreational activity.

- Provides insights into sales trends over time, facilitating correlation between activities.
- Highlights earnings disparity between pubs and restaurants, suggesting potential revisions to pricing systems.
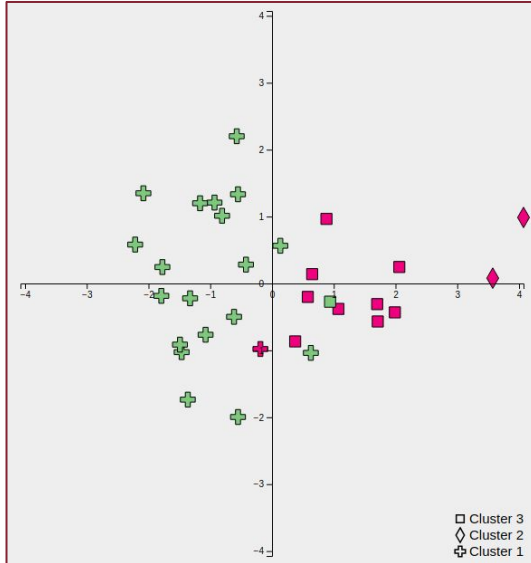- Updates dynamically when brushing on other plots, offering real-time analysis of income trends.
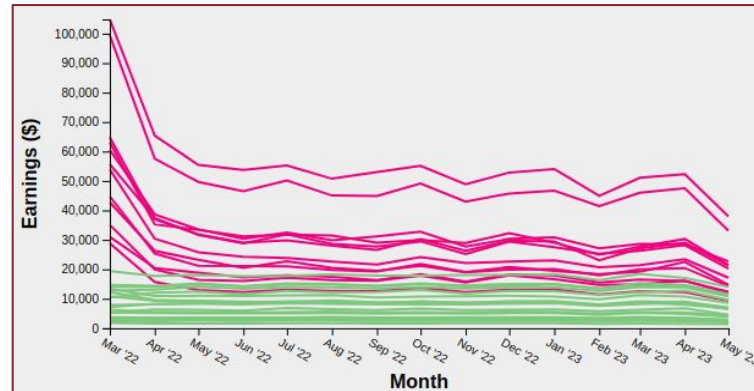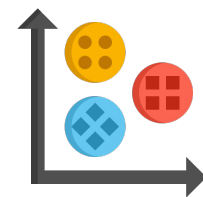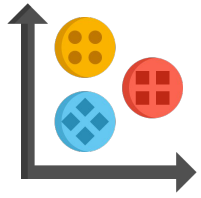
Real-time **clustering** and **dimensionality reduction** of entities in the current selection, used to draw a 2D scatter plot.

- **Gaussian Mixture Model (GMM)** clustering algorithm applied on attributes like household size, age, education level, etc.
- **Principal Component Analysis (PCA)** applied on the z-score normalized attributes, for dimensionality reduction.
- **Color encoding** consistent with the rest of the view, **shape encoding** for cluster representation.
- Clicking on a point **highlights its entire cluster**, allowing for further analysis.
- **Tooltip** displays entity ID on mouse hovering.
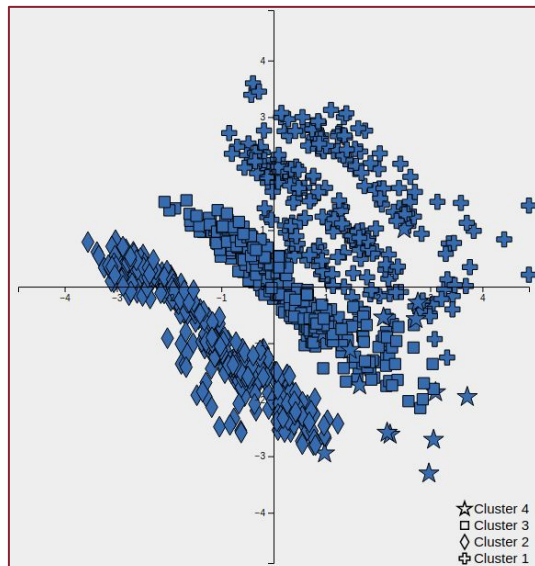
Real-time **clustering** and **dimensionality reduction** of entities in the current selection, used to draw a 2D scatter plot.

- **K-means** clustering algorithm applied on cost, maximum capacity, visits, earnings.
- **Principal Component Analysis (PCA)** applied on the z-score normalized attributes, for dimensionality reduction.
- **Color encoding** consistent with the rest of the view, **shape encoding** for cluster representation.

Benefits: aids in discovering similar groups and making decisions for enhancing city's earnings and entertainment offering.
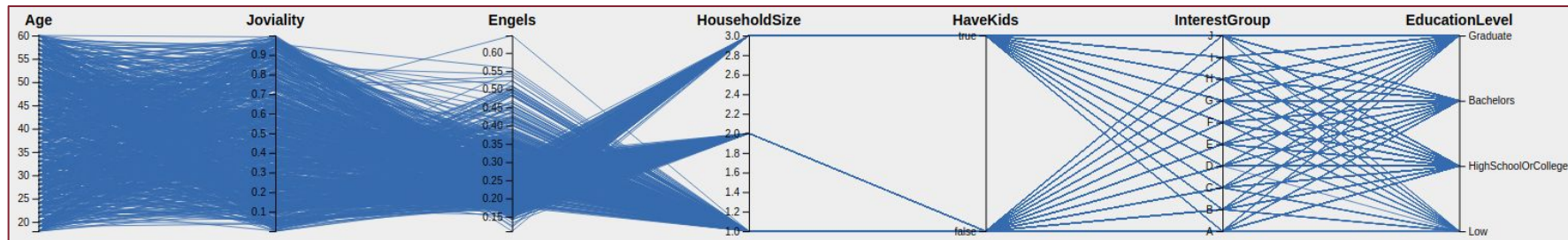
Parallel Plot
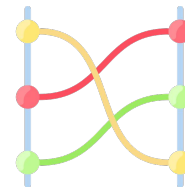
The **Parallel Plot** allows the user to have a versatile visualization method and to filter multiple attributes simultaneously through parallel axes. Each line encodes an entity.

Benefits:

- Enables easy discovery of trends, similarities, correlations and patterns for informed analysis.
- Brushing feature facilitates detailed filtering and mitigates over-plotting.

Participants view

Parallel Plot

The **Parallel Plot** allows the user to have a versatile visualization method and to filter multiple attributes simultaneously through parallel axes. Each line encodes an entity.
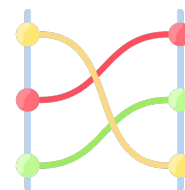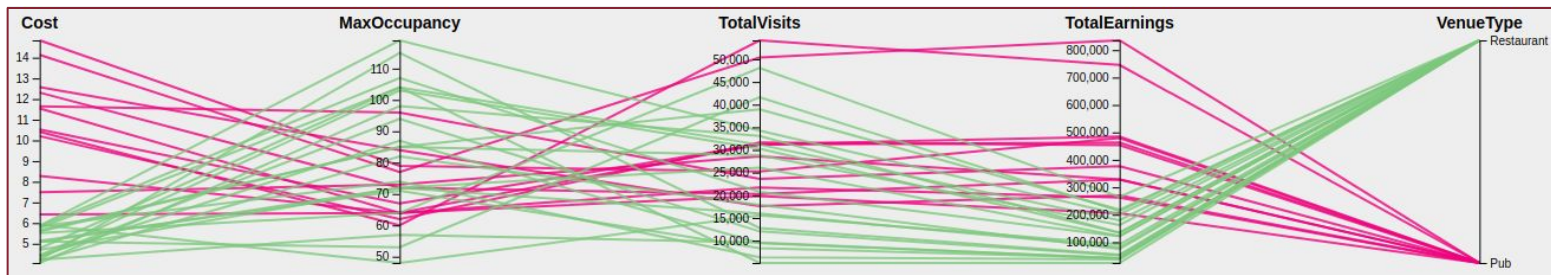
Benefits:

- Enables easy discovery of trends, similarities, correlations and patterns for informed analysis.
- Brushing feature facilitates detailed filtering and mitigates over-plotting.

Activities view

Thank you for listening!

And now let's go to the live demo for interaction and insights…

Full code available on:
https://github.com/alesspalma/VAproject