

User Guide

1. Introduction Page

This page has a short description of our R Shiny Application and an overview of the application layout.

Group 14 - Data Visualisation Survey Data Analysis

Introduction

EDA ▾

Cluster Analysis ▾

Association Analysis ▾

About

Where does the data visualisation stands today? Collected and organized by Data Visualization Society, Annual Data visualization Community Survey includes 50+ data visualization related questions and covers data visualization details such as salary, tool usage, demographic data, audiences and organizational structure. This survey provides valuable insights for organizations, practitioners and people who love data visualisation. We are three university students passionate about data visualisation. This project aims to design and develop a R Shiny application to allow audiences to explore data visualisation community survey data, and to maximise the insights obtained from the survey data. We demonstrate survey result from 3 dimensions, exploratory data analysis, cluster analysis and association analysis.

Comments

For Exploration Data Analysis, survey questions are categorised into 3 groups, Demography, Job and Challenge:

1. Demography - demographic questions, e.g. gender, country
2. Job - role, organization, engagement level in data visualisation
3. Challenge - willingness to improve visualisation skill, which visualisation skill to improve

For Cluster Analysis, 3 visualisation formats are available for Latent Class Analysis:

1. Line chart
2. Stacked bar chart
3. Heatmap

Application Layout

Interactive Exploratory Data Analysis

Bar Chart with error bar

- Demography
- Job
- Challenge

Cluster Analysis

Latent Class Analysis

- Line Chart
- Stacked Bar Chart
- Heatmap

Association Analysis

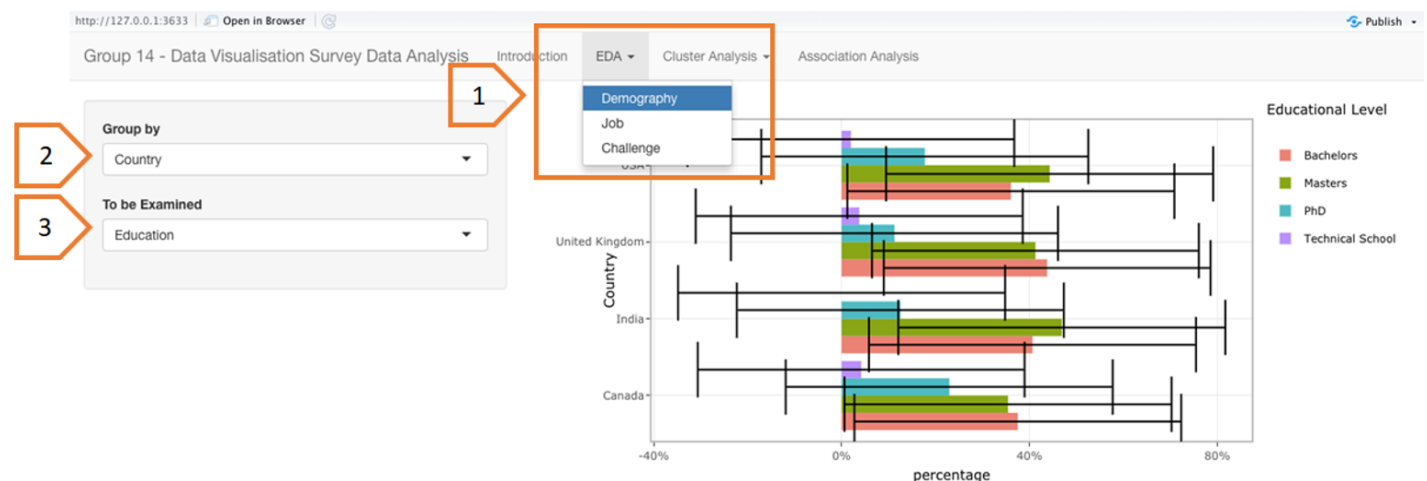
Association Rule Explorer

- Data Table
- Scatter Plot
- Matrix
- Grouped Matrix
- Graph
- Rule Export

Parallel Set Chart

2. EDA

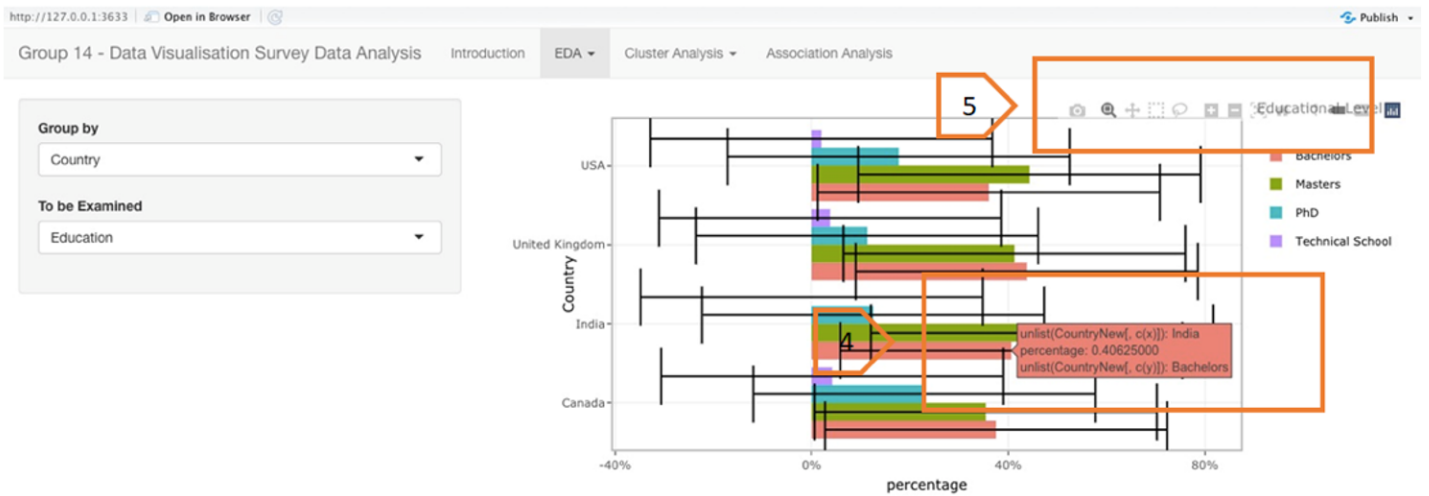
Bar chart with error bar allows for exploring categorical data with uncertainty and segregated by demographics variables e.g. country, gender, education level and etc..



[1] Select the desired subject to visualise the distribution of measure values of variables falling under the subject selected.

[2] Select the group-by variable to examine the distribution of each measure value of the to-be-examined variable grouped by that variable.

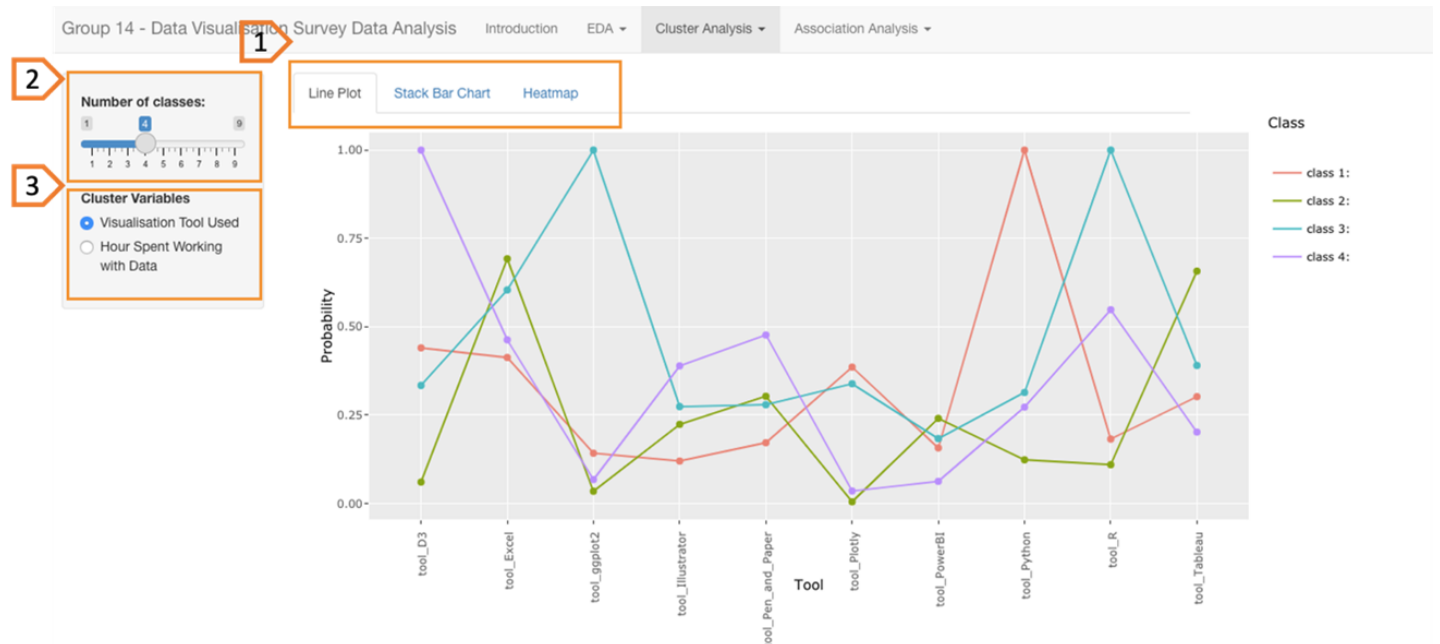
[3] Select the to-be-examined variable to visualise the distribution of each measure value of that variable.



[4] [5] Explore the handy facility provided by plotly, such as zoom in, hover and information label.

3. Cluster Analysis

Latent class analysis (LCA), which was commonly used for the analysis of multivariate categorical data. LCA offers a way to uncover hidden groupings (latent classes) in multivariate categorical data, by applying maximum likelihood method to calculate the probability that a case will fall in a particular latent class. In our application, 3 visualisation variations, line plot, stacked bar chart and heatmap are available for selection to demonstrate LCA. Same data is used for all three different formats. Line plot and stacked bar chart aims to cluster variables, in order to find out what tools belong to the same group and the relationship between varied duration of working on different visualisation tasks, whereas heatmap cluster respondents based on their behaviours (i.e. tools used and hours spent).



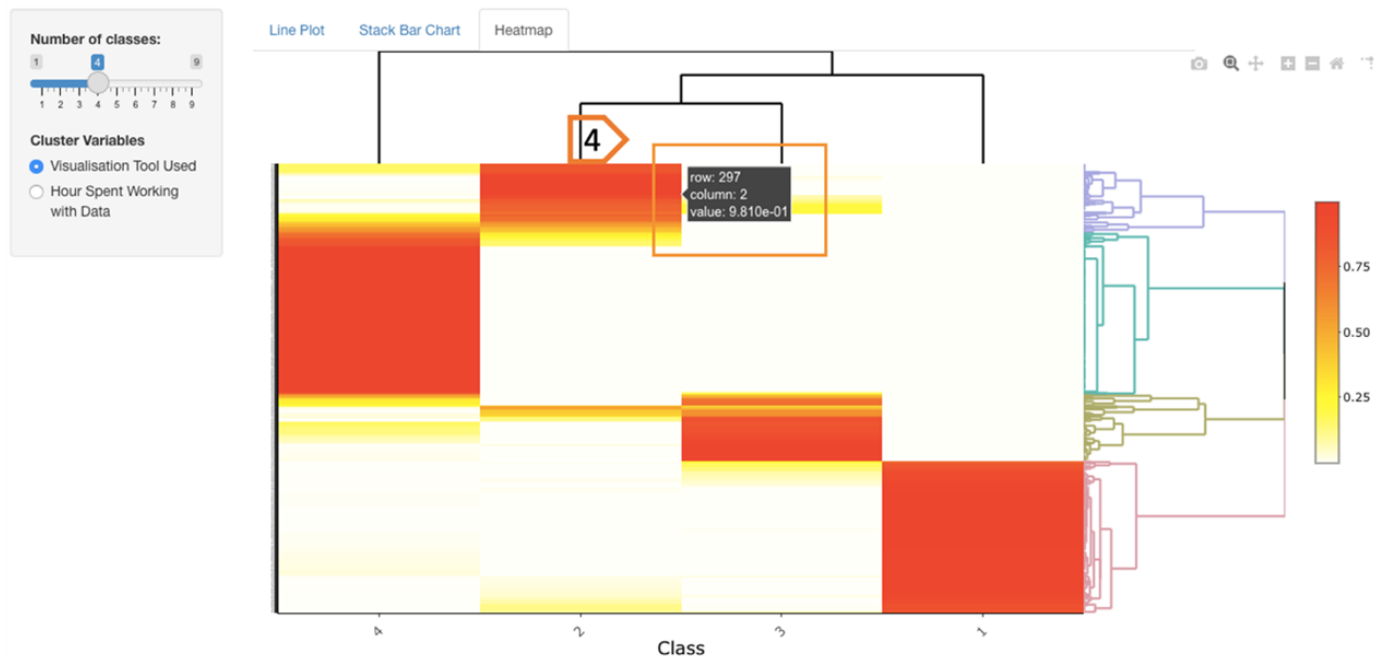
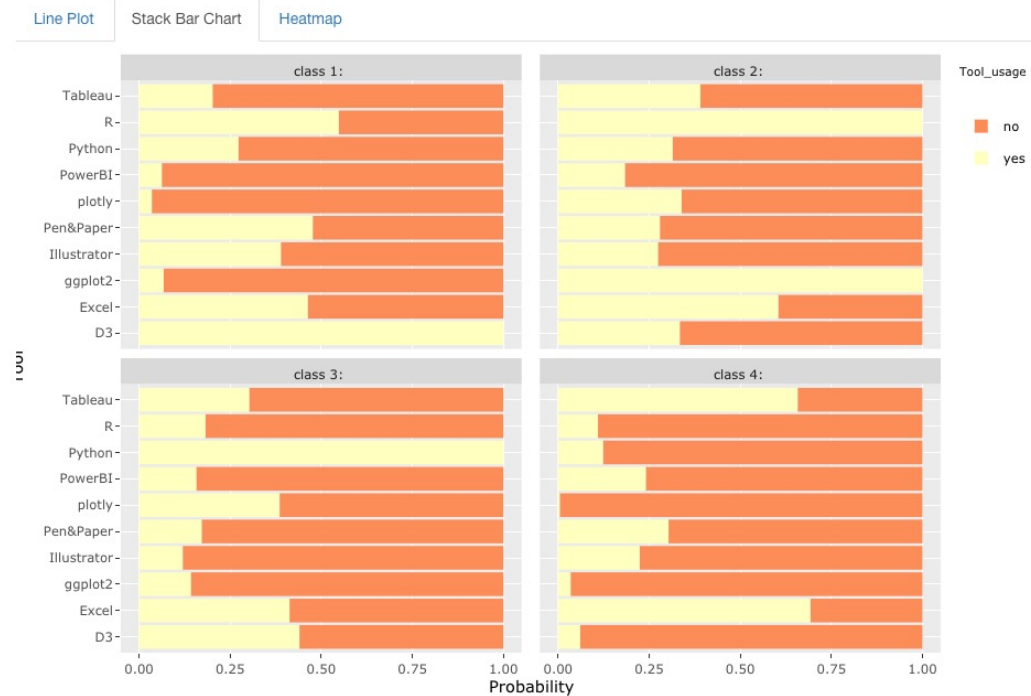
Number of classes:

1 2 3 4 5 6 7 8 9

Cluster Variables

☒ Visualisation Tool Used

☐ Hour Spent Working with Data



[1] Select the visualisation format of interest.

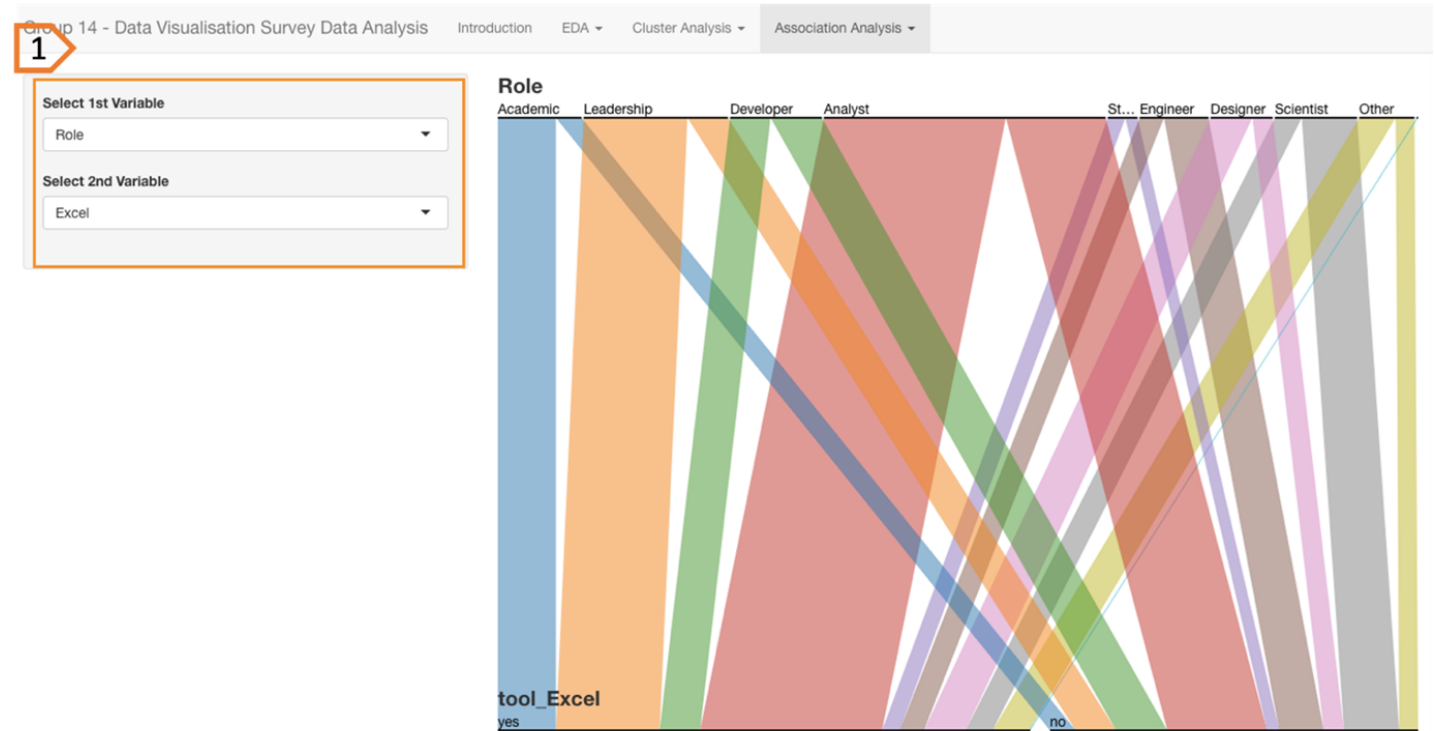
[2] Adjust number of clusters to observe the changes and find optimal number of classes, by moving the circle button.

[3] Select the cluster variables of interest. There are two groups of cluster variables available for exploration. The first group includes Excel, Tableau, R, ggplot2, D3, Python, Pen&Paper, Illustrator, PowerBI and Plotly. The second group consists of hours spent on data visualisation, data engineering, data science, design, data preparation and building portfolio.

[4] Hover over the heatmap for more information about the respondents.

4. Association Analysis

Parallel sets marry the advantages of two visualisation techniques, parallel coordinates which can display high-dimensional data and using frequencies to represent for the categories. In parallel sets, the width of box between two dimensions reflects the percentage of different category.



[1] Select variables of interest to explore the relationship between two variables, such as role of respondents in organization and usage of tool Excel. Variables are undergraduate major, organization area, role, gender, country, yearly pay, Excel, Tableau, R, ggplot2, D3, Python, Pen&Paper, Illustrator, PowerBI and Plotly.