Konuralp Journal of Mathematics, 12 (2) (2024) 150-157

Konuralp Journal of Mathematics

Research Paper

Journal Homepage: www.dergipark.gov.tr/konuralpjournalmath

e-ISSN: 2147-625X

Displaying Bivariate Data with Developed Cloud Based Data Visualization Tool

Emrah Altun[1*], Gökçen Altun[2] and Sümeyye İslamoğlu[1]

[1]*Department of Mathematics, Bartin University, Bartin 74100, Turkey.*
[2]*Department of Computer Technology and Information Systems, Bartin University, Bartin 74100, Turkey.*
[*]*Corresponding author*

Abstract

Data visualization is one of the hot topics of statistics. Programming languages such as R and Python are needed to create high-quality graphs. However, these programs are quite complex to use and difficult for researchers with no programming background. In this study, a cloud-based web application, named DATAVIS2, is developed to draw bivariate graphs. The developed application works independently of the operating system and the web browser. The DATAVIS2 web-tool includes fourteen different graphs for quantitative and categorical variables. The developed web-tool is available at https://beststat.shinyapps.io/datavis2/.

Keywords: *Bivariate graphs; Data visualization; R Shiny; Web-tool.*
2010 Mathematics Subject Classification: *62E15; 62J12*

1. Introduction

Data visualization is the visualization of the collected data with the help of graphics for easier understanding by the people. By visualizing the data with graphs, trends and outliers in the data can be easily revealed. Several books have been written by scientists to make data visualization easier for practitioners. Post et al. (2002) summarized the state-of-art data visualization methods such as volume and information visualization, multi-resolution methods, and interaction techniques. Anouncia et al. (2020) discussed the data visualization methods using machine learning and artificial intelligence methods. Barker and Canning (2013) used the R and JavaScript programs to visualize the spatial and time-series data sets.

The data visualization tools and software are used by researchers to summarize their own data sets to make them understandable. Two software are widely preferred to produce high-quality graphics: R and Python. However, a good level of software knowledge is required to use these programs. Several studies have been done to make data visualization easy for practitioners using the R Shiny software. Yu et al.(2018) developed a shinyCircos application to produce the Circos plot for genomic data. COVID19-world tool is developed by Tebe et al. (2020) to visualize and track the COVID-19 cases and deaths for all available countries. Similar to Yu et al. (2018), a minotaur application was developed to produce multivariate results from genome scans (see Verity et al., 2017). GENAVi is also a powerful tool for gene expression normalization, analysis, and visualization (see Reyes et al., 2019). Also, Mullan et al. (2021) developed a web-tool, ggVolcanoR, to draw the volcano plots, correlation plots, upset plots, and heatmaps for differential expression datasets. The HTPmod is developed by Chen et al. (2018) to visualize large-scale biological data using machine learning techniques. Recently, the PlotsOfData web-tool (Postma and Goedhart, 2019) is developed to visualize the data with box plot and violin plot. However, there are many graphs that can be used to visualize data. With these graphs, outliers and hidden features in the data set can be obtained. Thus, the analyst is provided with the opportunity to better understand and explain the data.

Visualization is also important field for mathematical studies, especially in graph theory. Graphical networks is used to explore the relations between units and widely used in biological and social science studies. There are some available R packages to perform graphical network analysis, but it requires well knowledge in programming. So, a web-tool can be developed in R Shiny for easy implementation of graphical network models. Graphical network models are also related to applied mathematics. Mathematical properties of the graphical networks were studied by many authors such as Do et al. (2016), Hoser and Geyer-Schulz (2005), Gürsoy (2022) and Kayaturan (2022).

In this study, we develop a web-tool to visualize the bivariate graphs for categorical-categorical, categorical-quantitative, and quantitative-quantitative variables. The developed program is named DATAVIS2. The DATAVIS2 web-tool is user-friendly and has fourteen graphs for different measurement levels of the variables. The user can upload the data easily and draw and save the desired graph. So, the users can easily get high-quality graphics without any knowledge of the R program.

Email addresses: **emrahaltun123@gmail.com (Emrah Altun), gokcenefendioglu@gmail.com (Gökçen Altun), sumislamoglu@gmail.com (Sümeyye İslamoğlu)**

The following sections are included in the remaining parts of the study. In Section 2, we present the developed web-tool, DATAVIS2. In Section 3, the development process of the web-tool is summarized. The study is concluded with Section 4.

2. DATAVIS2 web-tool

Now, we introduce the developed web-tool. The DATAVIS2 program can be easily accessible using the following website https://beststat.shinyapps.io/datavis2/. Firstly, we describe the used data set and go on with the graphs for different measurement levels of the variables.

2.1. Dataset

The data set contains the salary information of the academicians based on their rank, experience, discipline, and gender. The data set can be found in CSV format at https://r-data.pmagunia.com/dataset/r-dataset-package-car-salaries. Figure 2.1 displays the upload data section of the DATAVIS2 web-tool. The data file should be in CSV format. The delimiter can be a semicolon, comma, and tab. The data set should be chosen after clicking the *Load CSV file*. If the user wants to display the graphs for categorical-categorical variables, the *Categorical-Categorical* section should be chosen from the top of the web-tool. Once the user uploads the data, the data is shown on the right panel of the web-tool under the *Data Table* section. The useful graphs to summarize the data are produced with the developed web-tool in the next section.

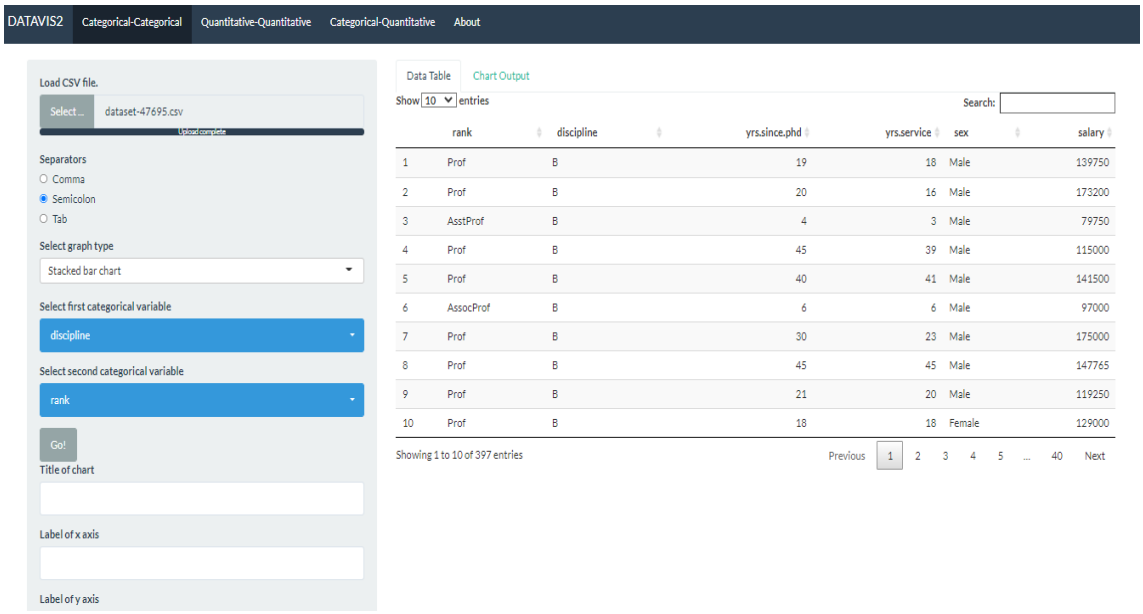

Figure 2.1: Upload data for DATAVIS2.

2.2. Categorical-categorical graphs

Here, we present the graphs for the categorical-categorical variables. Four graphs are presented in this section. These are stacked bar chart, grouped bar chart, segmented bar chart, and improved segmented bar chart. After uploading data, one can easily select the graph type using the *Select graph type* tab. Firstly, the user should select the first and second categorical variables. After selecting the graph type, the user should click the *Go!* button. The graph is displayed in the *Chart Output* section. Figure 2.2 displays the improved segmented bar chart for the discipline and rank variables. As seen from Figure 2.2, while 72.4% of those in department A are professors, 62.5% of those in department B are professors. The other available graphs can be obtained by changing the graph type. Also, the graph can be easily saved in png format by right-clicking on the figure. The other graphs are also displayed in Figures 2.3, 2.4 and 2.5.

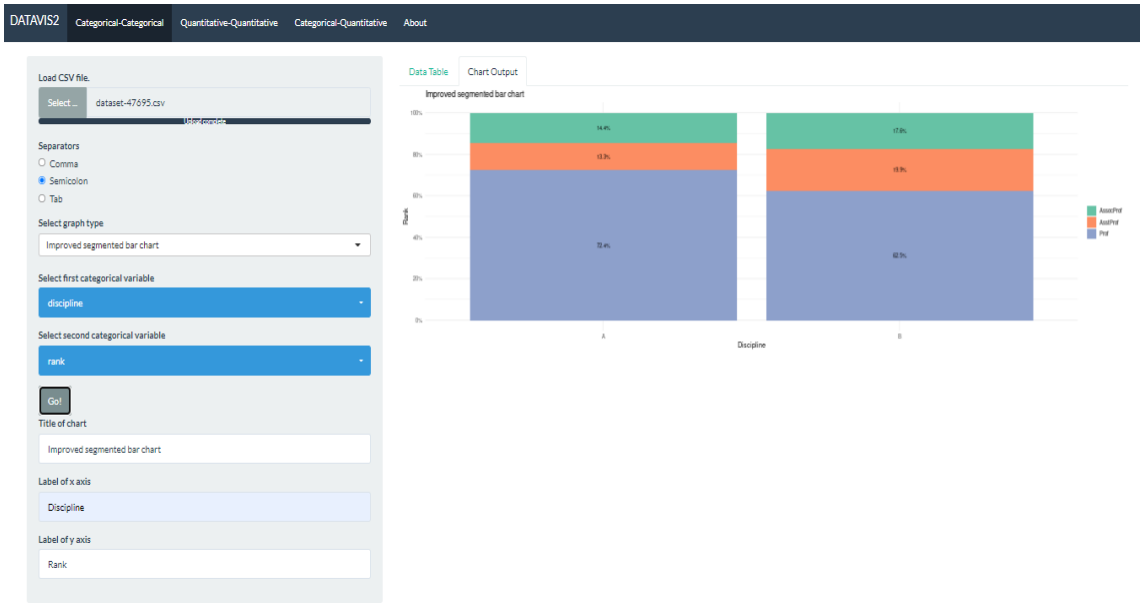

Figure 2.2: The improved segmented bar chart

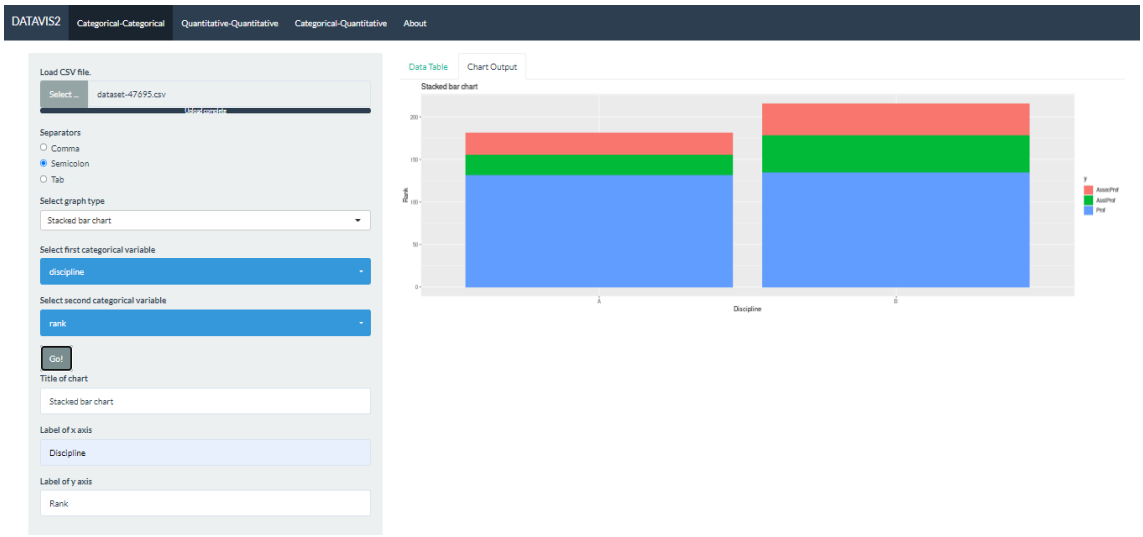

Figure 2.3: The stacked bar chart

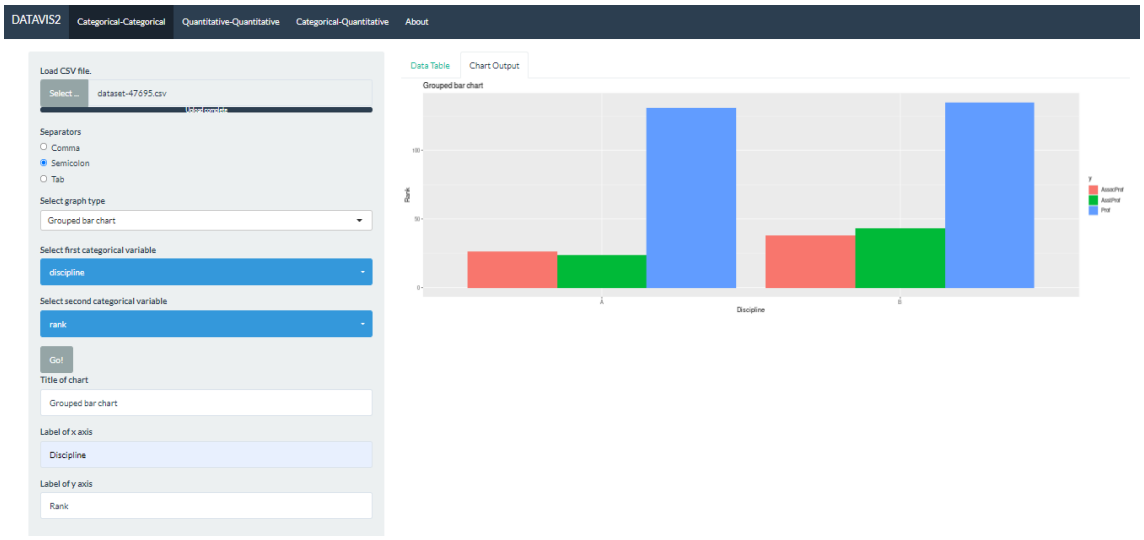

Figure 2.4: The grouped bar chart

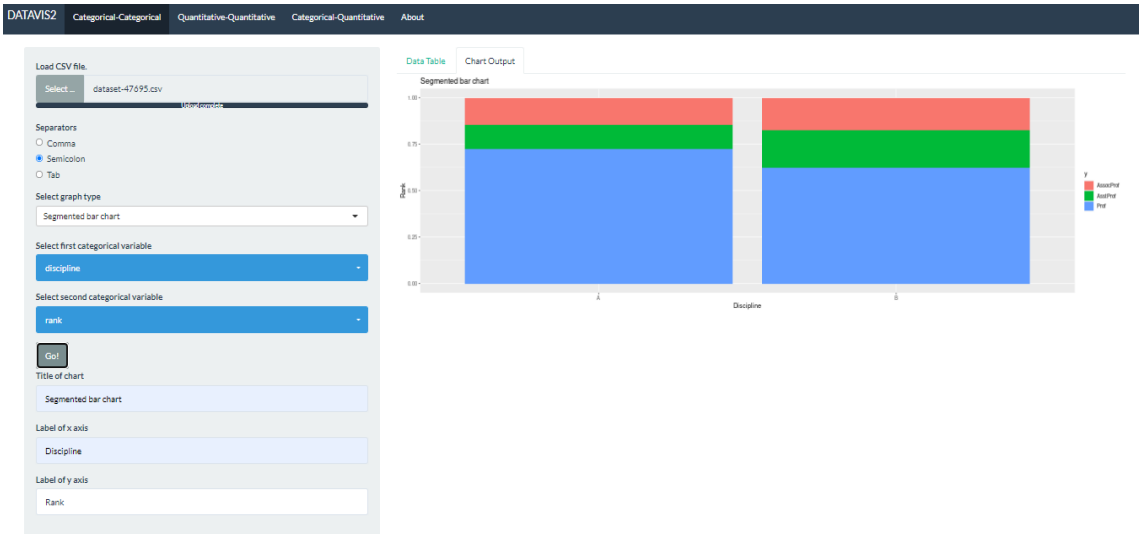

Figure 2.5: The segmented bar chart

2.3. Quantitative-quantitative graphs

Now, we present the graphs for the quantitative-quantitative variables. We have two graphs for this section. These are scatter plots and line plots. Again, the user uploads the data and selects the appropriate graph for his/her analysis. Then, the first and second quantitative variables should be selected. Then, the user clicks the *Go!* button. Then, the produced graph is displayed in the *Chart Output* section. As in the previous section, the produced graph can be easily saved and used by the users. Figure 2.6 displays the scatter plot for the number of years the professor has served the department and/or university and salary variables.

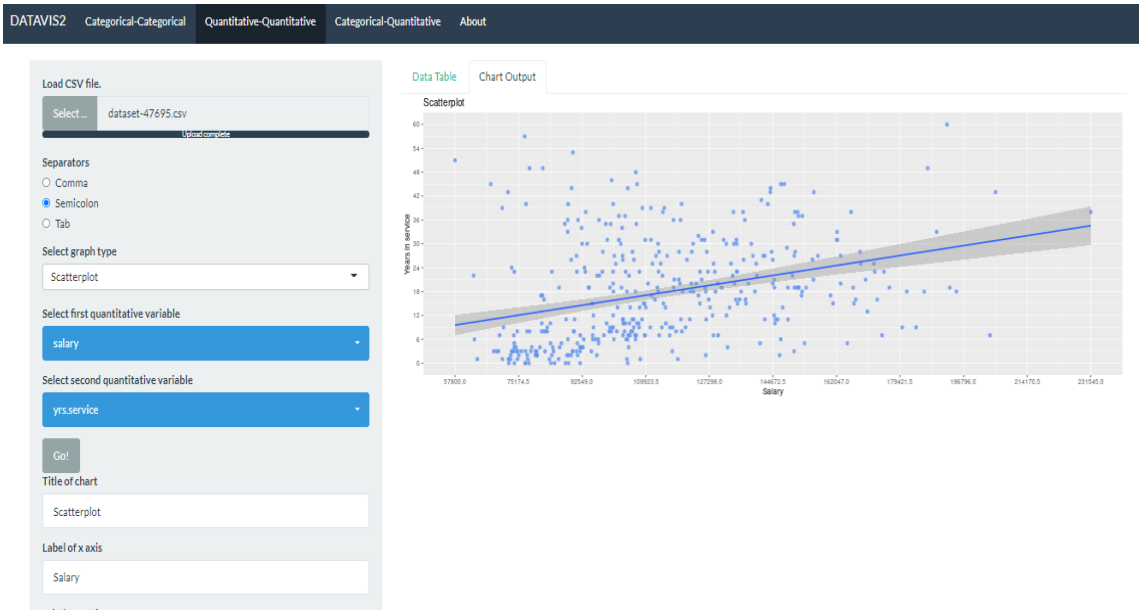

Figure 2.6: The scatter plot

Also, Figure 2.7 displays the line plot for the years number of years since the professor has obtained and salary variables.

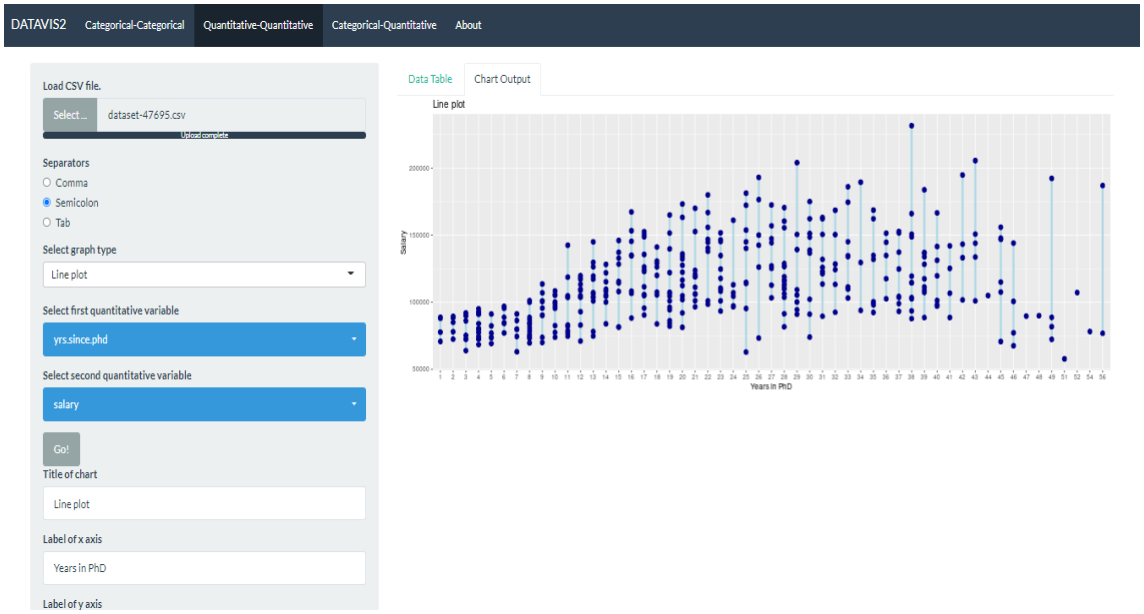

Figure 2.7: The line plot

2.4. Categorical-quantitative graphs

The categorical-quantitative graphs are given in this section. The DATAVIS2 web-tool has eight different graphs for the categorical-quantitative variables. These are bar plot, grouped kernel density plot, box plot, violin plot, ridgeline plot Mean/SEM (standard error of the mean) plot, strip plot, and beeswarm plot. The categorical and quantitative variables should be chosen. Then, the user clicks the *Go!* button to display the produced graph.

Figure 2.8 displays the bar chart and grouped kernel density plot. The bar chart can be used to visualize the mean of the quantitative variable based on each level of the categorical variable. The grouped kernel density plot is also very useful to have any information about the distribution of each level of the categorical variable.

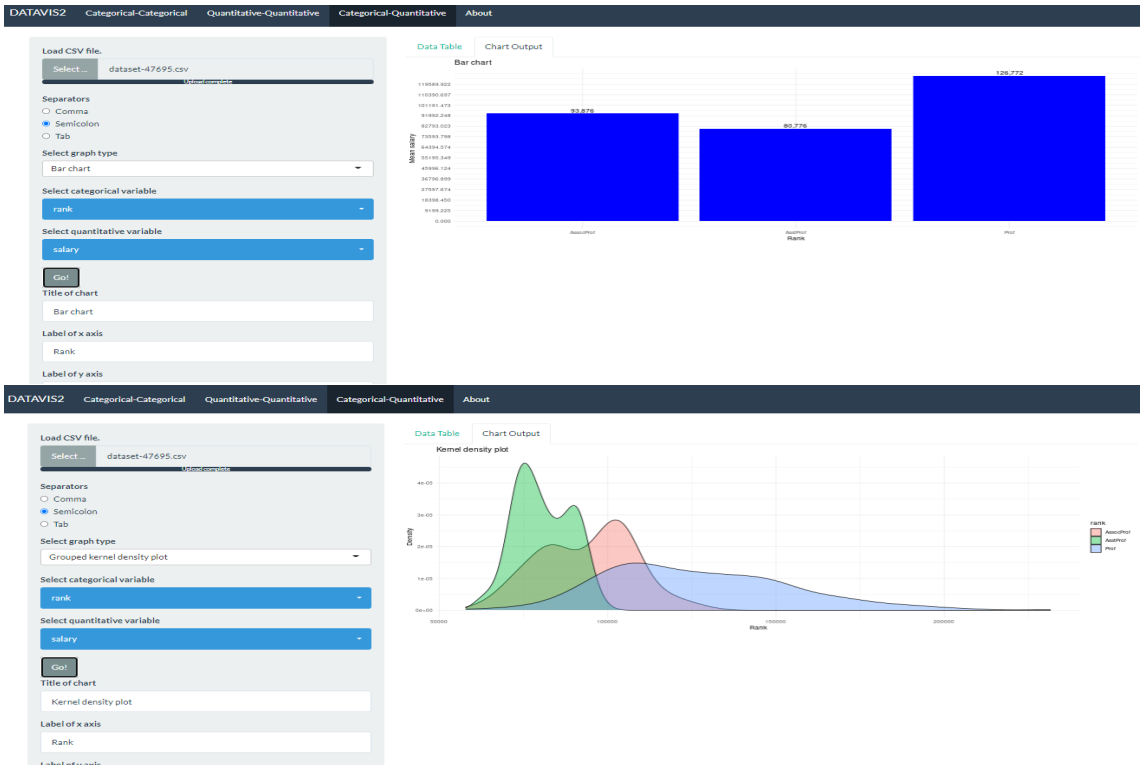

Figure 2.8: The bar chart and grouped kernel density plot.

Figure 2.9 displays the violin and mean/SEM plot. The violin plots are very similar to the kernel density plots. The violin plot can be viewed as a mirrored and rotated kernel density plot. The mean/SEM plot is preferred with many scientists to summarize their observations/measures for some categorical variables. Here, the errors bars represent the confidence intervals of the sample for each level of the categorical variable.

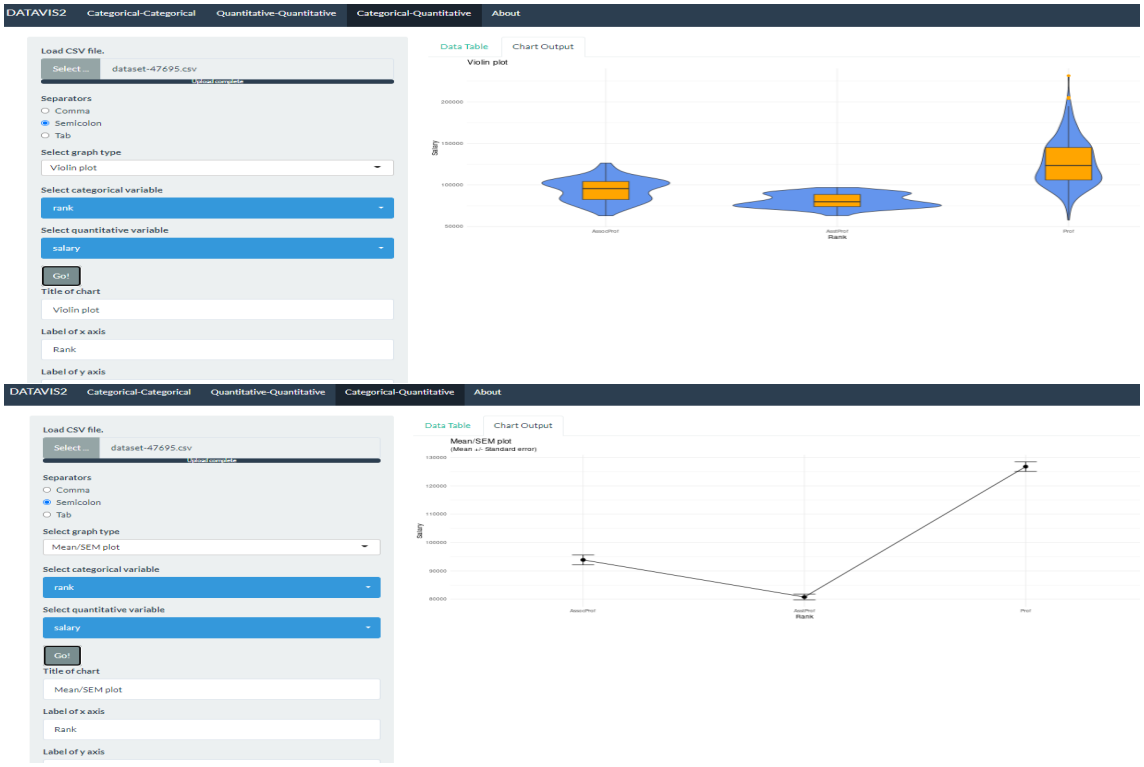

Figure 2.9: The violin and mean/SEM plot.

Figure 2.10 displays the strip and beeswarm plots. The strip can be viewed as a combined box and scatter plot. The strip plot can be used to have any information about the distribution of the quantitative variable for each level of the categorical variable. Also, the strip plot can be used to identify the possible outliers in the data set. The beeswarm plot can be also viewed as a combined scatter and violin plots. This plot also can be used to have information about the distribution of the sample.

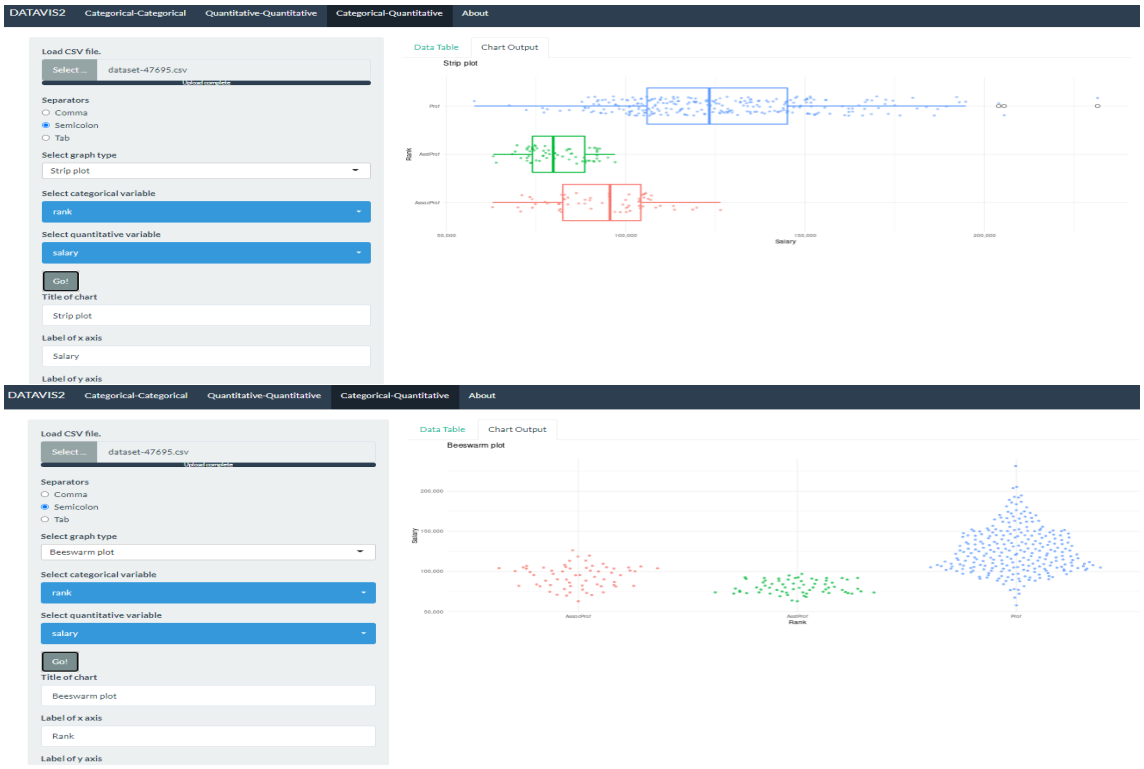

Figure 2.10: The strip and beeswarm plots.

3. Development and deployment process of the DATAVIS2 web-tool

The shiny package (see Chang et al., 2021) of R software is used to develop the DATAVIS2 web-tool. During the development process, we take the advantage of the below R packages.

- shinythemes by Chang (2021)
- shinydashboard by Chang and Ribeiro (2018)
- ggplot2 by Wickham (2016)
- treemapify by Wilkins (2021)
- ggridges by Wilke (2021)
- ggbeeswarm by Clarke and Sherrill-Mix (2017)
- scales by Wickham and Seidel (2020)
- data.table by Dowle and Srinivasan (2021)
- dplyr by Wickham et al. (2021)

After created the web-tool, we use the https://www.shinyapps.io/ server to deploy the developed web-tool. It is easy to use and do not require any installation.

4. Conclusion

This study presents the developed web-tool, DATAVIS2. Thanks to a simple and user-friendly interface, the users can easily upload their datasets and create high-quality bivariate graphs for different measurement levels of the variables. The main advantage of the web-tool is that it works independently from the web browser and operating system. The graphs are easily saved and be used for academic or reporting purposes. The DATAVIS2 is a great tool to visualize the bivariate variables without any knowledge of software such as R, Python, or Java. As a future work of the presented study, we will develop a web-tool for multivariate graphs, biplots, and heatmaps.

Article Information

Acknowledgements: The authors would like to express their sincere thanks to the editor and the anonymous reviewers for their helpful comments and suggestions.

Author's contributions: All authors contributed equally to the writing of this paper. All authors read and approved the final manuscript.

Conflict of Interest Disclosure: No potential conflict of interest was declared by the author.

Supporting/Supporting Organizations: No grants were received from any public, private or non-profit organizations for this research.

Ethical Approval and Participant Consent: It is declared that during the preparation process of this study, scientific and ethical principles were followed and all the studies benefited from are stated in the bibliography.

Plagiarism Statement: This article was scanned by the plagiarism program. No plagiarism detected.

Availability of data and materials: Not applicable

References

[1] Chang, W. (2021). shinythemes: Themes for Shiny. R package version 1.2.0. https://CRAN.R-project.org/package=shinythemes
[2] Chang, W. and Ribeiro, B.B. (2018). shinydashboard: Create Dashboards with 'Shiny'. R package version 0.7.1. https://CRAN.R-project.org/package=shinydashboard
[3] Anouncia, S. M., Gohel, H. A. and Vairamuthu, S. (Eds.). (2020). Data Visualization: Trends and Challenges Toward Multidisciplinary Perception. Springer Nature.
[4] Barker, T. and Canning, M. (2013). Pro Data Visualization using R and JavaScript (No. s 207). New York, NY: Apress.
[5] Chang, W., Cheng, J., Allaire, J.J., Sievert, C. Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A. and Borges, B. (2021). shiny: Web Application Framework for R. R package version 1.6.0. https://CRAN.R-project.org/package=shiny
[6] Chen, D., Fu, L. Y., Hu, D., Klukas, C., Chen, M. and Kaufmann, K. (2018). The HTPmod Shiny application enables modeling and visualization of large-scale biological data. Communications Biology, 1(1), 1-8.
[7] Do, A. L., Boccaletti, S., Epperlein, J., Siegmund, S. and Gross, T. (2016). Topological stability criteria for networking dynamical systems with Hermitian Jacobian. European Journal of Applied Mathematics, 27(6), 888-903.
[8] Clarke, E. and Sherrill-Mix, S. (2017). ggbeeswarm: Categorical Scatter (Violin Point) Plots. R package version 0.6.0. https://CRAN.R-project.org/package=ggbeeswarm
[9] Dowle, M. and Srinivasan, A. (2021). data.table: Extension of 'data.frame'. R package version 1.14.0. https://CRAN.R-project.org/package=data.table
[10] Gürsoy, A. (2022). Construction of networks by associating with submanifolds of almost Hermitian manifolds. Fundamental Journal of Mathematics and Applications, 5(1), 21-31.
[11] Hoser, B. and Geyer-Schulz, A. (2005). Eigenspectral analysis of hermitian adjacency matrices for the analysis of group substructures. Journal of Mathematical Sociology, 29(4), 265-294.
[12] Kayaturan, G. C. (2022). Error Elimination From Bloom Filters in Computer Networks Represented by Graphs. Fundamental Journal of Mathematics and Applications, 5(4), 240-244.
[13] Mullan, K. A., Bramberger, L. M., Munday, P. R., Goncalves, G., Revote, J., Mifsud, N. A., ... and Li, C. (2021). ggVolcanoR: A Shiny app for customizable visualization of differential expression datasets. Computational and Structural Biotechnology Journal, 19, 5735-5740.
[14] Post, F. H., Nielson, G. and Bonneau, G. P. (Eds.). (2002). Data visualization: The state of the art.
[15] Postma, M. and Goedhart, J. (2019). PlotsOfData-A web app for visualizing data together with their summaries. PLoS Biology, 17(3), e3000202.
[16] Reyes, A. L. P., Silva, T. C., Coetzee, S. G., Plummer, J. T., Davis, B. D., Chen, S., ... and Jones, M. R. (2019). GENAVi: a shiny web application for gene expression normalization, analysis and visualization. BMC genomics, 20(1), 1-9.

[17] Tebe, C., Valls, J., Satorra, P. and Tobias, A. (2020). COVID19-world: a shiny application to perform comprehensive country-specific data visualization for SARS-CoV-2 epidemic. BMC Medical Research Methodology, 20(1), 1-7.
[18] Verity, R., Collins, C., Card, D. C., Schaal, S. M., Wang, L. and Lotterhos, K. E. (2017). minotaur: A platform for the analysis and visualization of multivariate results from genome scans with R Shiny. Molecular ecology resources, 17(1), 33-43.
[19] Wickham, H. and Seidel, D. (2020). scales: Scale Functions for Visualization. R package version 1.1.1. https://CRAN.R-project.org/package=scales
[20] Wickham, H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016
[21] Wickham, H., Francois, R., Henry, L. and Muller, K. (2021). dplyr: A Grammar of Data Manipulation. R package version 1.0.7. https://CRAN.R-project.org/package=dplyr
[22] Wilke, C.O. (2021). ggridges: Ridgeline Plots in 'ggplot2'. R package version 0.5.3. https://CRAN.R-project.org/package=ggridges
[23] Wilkins, D. (2021). treemapify: Draw Treemaps in 'ggplot2'. R package version 2.5.5. https://CRAN.R-project.org/package=treemapify
[24] Yu, Y., Ouyang, Y. and Yao, W. (2018). shinyCircos: an R/Shiny application for interactive creation of Circos plot. Bioinformatics, 34(7), 1229-1231.