Research Article

EpigenomeViewer: A Shiny Web Application for Visualizing Methylation Patterns Across Biological Samples

Nurul Syakima Ab Mutalib[1] *

[1] UKM Medical Molecular Biology Institute (UMBI), Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia; syakima@ukm.edu,my; 0000-0001-6914-2224

* Correspondence: syakima@ukm.edu.my; +603-9145 9073.

Abstract: *DNA methylation plays a crucial role in gene regulation and is involved in the development of numerous diseases including cancer. However, researchers often face challenges in visualising and analysing methylation data because of limited accessibility to user-friendly software that integrates interactive visualisation with bioinformatics analysis. EpigenomeViewer is a Shiny web application developed to address this gap by providing a comprehensive platform for visualising DNA methylation patterns across tumour and normal biological samples. Users can upload their differentially methylated probes and explore methylation patterns using publication-ready figures, such as violin plots, heatmaps, and correlation matrices. Additionally, the app's real-time filtering and correlation matrix analysis allowed researchers to identify co-methylation patterns, facilitating insights into gene regulation and potential biomarkers for disease. EpigenomeViewer is particularly useful in the study of drug resistance and early stage cancer biomarkers, where understanding methylation differences between tumour and normal samples is vital. With its export features, researchers can download high-quality visualisations for publication and presentation. This application is an invaluable tool for both experienced researchers and educators in bioinformatics and epigenetics, supporting a wide range of studies from academic research to clinical applications. Future developments include expanded support for multi-omics data and additional gene ontology features.*

Keywords: DNA methylation; web application; Shiny app; bioinformatics visualization; EpigenomeViewer; epigenetics; cancer research; biomarker discovery

DOI: 10.5281/zenodo.14822076

1. INTRODUCTION

DNA methylation is a critical epigenetic modification that influences gene expression without altering DNA sequence. In mammals, DNA methylation typically occurs at cytosine residues within CpG dinucleotides, where a methyl group is added, effectively silencing gene expression by hindering transcription factor binding or by attracting repressive protein complexes (Bird, 2002). This regulatory mechanism plays a fundamental role in normal development, genomic stability and cellular differentiation. Aberrant DNA methylation patterns are frequently observed in various diseases, particularly cancer, where hypermethylation of tumour suppressor genes or hypomethylation of oncogenes is common (Esteller, 2008). Therefore, understanding DNA methylation patterns is critical for uncovering disease mechanisms, identifying biomarkers, and developing novel therapeutic strategies.

Despite its importance, the analysis of DNA methylation data presents significant challenges. High-throughput technologies, such as microarrays and next-generation sequencing (NGS), generate large amounts of methylation data and require tools that can process, analyse, and visualise these data effectively (Jones & Baylin, 2002). The need for accessible bioinformatics tools has led to the development of software that allows researchers to visualise methylation patterns, identify differentially methylated regions, and explore the relationships between methylation sites. However, many existing tools are either difficult to use without extensive programming knowledge or lack interactive and flexible visualisation features, limiting their accessibility to researchers with minimal bioinformatic training (O'Donoghue, 2021).

Several bioinformatic tools have been developed to address these challenges. For instance, MethVisual offers visualisation of methylation data but requires extensive data formatting, whereas the UCSC Genome Browser provides static methylation views but lacks dynamic, interactive features for in-depth exploration (Haeussler et al., 2019). Another tool, MethyLumi in R, provides extensive options for methylation data normalisation and preprocessing but is primarily command-line-based, creating a steep learning curve for non-programmers (Du et al., 2010). The minfi package also provides powerful options for preprocessing, but it requires substantial bioinformatics knowledge, further highlighting the need for an intuitive tool, such as EpigenomeViewer (Aryee et al., 2014).

To address this gap, we developed EpigenomeViewer, a Shiny-based web application that allows researchers to intuitively visualise and analyse methylation patterns across biological samples. EpigenomeViewer is specifically designed for ease of use, enabling users to upload their methylation data and explore them through interactive, publication-ready visualisations, including violin plots, heatmaps, and correlation matrices. This paper describes the development and features of EpigenomeViewer, its advantages over existing tools, and its potential impact on methylation research, particularly in terms of understanding disease mechanisms and identifying epigenetic biomarkers.

2. METHOD & MATERIAL

2.1 Development Environment and Tools

Utilising R and Shiny, the EpigenomeViewer application was created by harnessing R's extensive statistical and graphical libraries to produce an interactive and accessible Web interface. The application was constructed using RStudio, an integrated development environment (IDE) that is particularly suitable for R-based applications. To enable web-based access without requiring local R installation, the app was subsequently hosted on Shinyapps.io.

2.2 Libraries and Packages

The shiny package forms the core of the EpigenomeViewer application, providing the web application framework required to build an interactive user experience (Chang et al., 2021). The Shinythemes package was used to apply the "United" theme, ensuring a visually cohesive and professional aesthetic across the app's interface (Wickham & Miller, 2021). To render data tables with interactive filtering and search functionality, the DT package was incorporated, allowing users to efficiently navigate large datasets (Xie et al., 2021).

The readxl and openxlsx packages support the reading of Excel files (. xlsx), making it possible for users to upload different data formats and thus enhancing the app's accessibility and flexibility (Wickham & Bryan, 2021). Additionally, ggplot2 provides powerful graphing capabilities that are used to generate violin plots and other publication-ready visualisations essential for methylation data analysis (Wickham, 2016). For efficient data manipulation, dplyr and tidyr handle tasks such as data

filtering, grouping, and tidying, which are essential for preparing datasets for visualisation (Wickham et al., 2019). The data.table package further enhances data handling efficiency, especially for large datasets, ensuring smooth functionality and performance.

To visualise methylation distributions across samples, a pheatmap was utilised to create heatmaps, allowing researchers to observe methylation intensity across tumour and normal samples (Kolde, 2019). The ggcorrplot package facilitates the generation of correlation matrix visualisations, enabling the exploration of co-methylation patterns between probes, and providing insights into possible regulatory relationships (Kassambara, 2019). To calculate correlations with additional statistical options, the Hmisc package was used to support a robust correlation analysis across probes (Harrell, 2020). Finally, shinycssloaders add loading animations, improving the user experience by indicating processing times for various visualisations, such as large datasets or complex analyses (Attali, 2021).

2.3 User Interface Design

The user interface (UI) of EpigenomeViewer was constructed with Shiny's fluidPage layout and sidebarLayout panels, aiming to create an intuitive, accessible platform. The UI includes a sidebar panel for uploading data and setting parameters as well as a main panel for displaying outputs, which are organised in multiple tabs. The application employs the "United" theme from shinythemes, giving the interface a clean, visually consistent look. Key UI elements include the following.

- Data Input Options: The app accepts data uploads in CSV, Excel (.xlsx), or Text (.txt) file formats. Users can also access the sample data via a dedicated button for immediate analysis.

- Data Refinement Tools: Users can narrow their analysis through dropdown menus, which enables gene filtering, selection of differentially methylated probes (DMPs), and choice of methylation patterns (such as hyper- or hypomethylated areas).

- Display Customisation: The interface offers sliders and selection boxes, allowing users to modify the correlation threshold, pick CpG island (CGI) characteristics, and choose gene features. These tools provide an enhanced analytical flexibility.

2.4 Data Processing and Visualization

Upon data upload, EpigenomeViewer processes the data to facilitate exploration and visualisation. Key features include:

- Data Table: Displays the full dataset in an interactive table (using DT) where users can search and filter specific genes or methylation probes.

- Violin Plot: Created with ggplot2, the violin plot tab visualises the distribution of methylation values across tumour and normal samples, with box plots overlaid to highlight quartiles and median values.

- Heatmap: Built with pheatmap, which visualises methylation intensity across samples using colour gradients to indicate methylation levels. The heatmap tab enables researchers to detect patterns and variations in methylation across samples.

- Summary Statistics: Displays bar charts for key metrics, such as counts of hypermethylated versus hypomethylated probes, and the distribution of probes across gene regions.

- Correlation Matrix: Using ggcorrplot and Hmisc for correlation calculations, the correlation matrix tab enables researchers to analyse the relationships between methylation probes and detect co-methylation patterns, which can indicate potential regulatory interactions between probes.

2.5 Correlation Matrix - Spearman Correlation

The Correlation Matrix in EpigenomeViewer uses Spearman's correlation rather than Pearson's correlation to calculate the relationships between methylation probes. Spearman correlation is a non-parametric measure that assesses the monotonic relationship between variables rather than assuming a linear relationship, as required by the Pearson correlation. This approach is beneficial in epigenetic data analysis because methylation data often do not meet the normality or linearity assumptions required for Pearson correlation. Using Spearman's correlation, EpigenomeViewer provides a more robust analysis of co-methylation patterns, capturing a broader range of relationships that may indicate co-regulatory behaviours or functional associations between probes.

2.6 Example Data and Workflow

To demonstrate the functionality of EpigenomeViewer, an example dataset from a colorectal cancer study is provided. This dataset includes the methylation values for genes related to drug resistance and other cancer-related pathways. Users can explore the dataset interactively, apply filters, and generate visualisations without preprocessing.

2.7 Accessibility and Availability

EpigenomeViewer is freely accessible for non-profit use at https://syakimalab.shinyapps.io/EpigenomeViewer/, providing researchers with an intuitive platform for methylation data visualization and analysis.

3. FINDINGS

EpigenomeViewer effectively facilitates the visualisation and exploration of DNA methylation data, providing users with multiple interactive tools to examine methylation patterns across tumour and normal samples. Using an example dataset of colorectal cancer, we investigated the capabilities of EpigenomeViewer to visualise, filter, and analyse methylation data through several interactive plots and data tables. Significant findings in methylation patterns were observed, including differential methylation between tumour and normal samples and strong correlations among certain methylation probes, indicating co-methylation that may reflect the underlying regulatory relationships.

3.1 Data Table and Filtering Results

The input data for EpigenomeViewer consists of differentially methylated probes (DMPs) obtained from bioinformatics packages such as CHAMP (Tian et al., 2017) or minfi, and formatted as a CSV, Excel (. xlsx), or text (.txt) file, making the data upload straightforward for researchers. The required columns included gene names, DMPs, and methylation values across samples, with each sample identified by a unique ID. To distinguish between sample types, IDs for tumour samples should end with "T" (e.g., C345T), while IDs for normal samples should end with "N" (e.g., C342N). Additionally, users may include optional metadata columns, such as delta beta ($\Delta\beta$) and p-values, to support more detailed statistical analysis. This flexible data format allows for a variety of analyses, from exploring differential methylation to examining co-methylation patterns across tumour and normal samples, thus enhancing the applicability of the app in both basic and clinical research.

Once uploaded, the Data Table features provided an initial overview of the methylation dataset. The dataset included multiple genes and probes, with methylation values provided for both tumor (T) and normal (N) samples. By filtering for genes associated with chemotherapy resistance, such as ABCC5, the Data Table allowed us to focus on relevant gene-probe combinations. Filtering options allowed rapid customisation of the dataset to isolate specific methylation trends and to investigate hypermethylated or hypomethylated regions. For instance, we isolated ABCC5-related probes and identified hypomethylation patterns in tumour samples, suggesting a possible role in chemotherapy resistance, consistent with findings in the literature (Berdasco & Esteller, 2010).

3.2 Violin Plot Analysis of Methylation Distribution

The Violin Plot provides an intuitive visualisation of the distribution of methylation values across tumour and normal samples. Figure 1 shows the methylation distribution of *ABCC5*, a gene associated with drug resistance in colorectal cancer. The tumour samples displayed lower methylation levels for *ABCC5*, whereas normal samples exhibited higher methylation levels, suggesting hypomethylation in the tumour samples. This hypomethylation pattern could indicate gene activation, potentially leading to increased drug efflux and chemotherapy resistance, which is a known challenge in colorectal cancer treatment. The violin plot, enhanced with box plot overlays, allowed for a detailed comparison, highlighting the differences in quartiles and median methylation levels between the two groups.

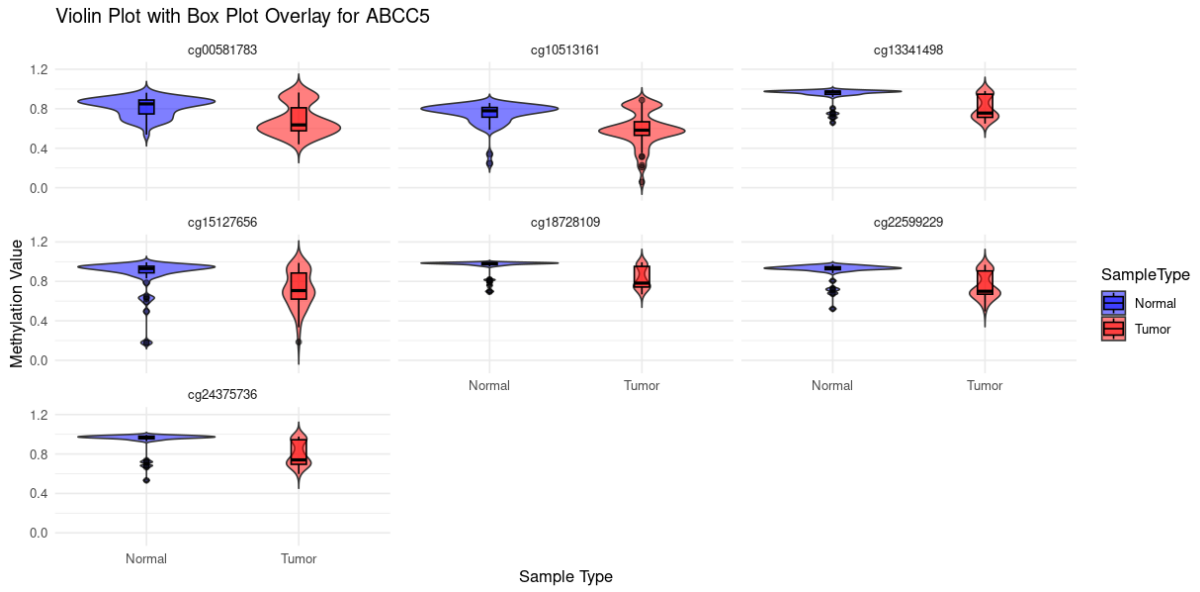

Figure 1. Violin plot of *ABCC5* methylation values across tumour and normal samples.

3.3 Heatmap Visualization of Methylation Intensity

Heatmap visualisation, constructed using the pheatmap package, effectively highlighted methylation intensity across samples. We generated a heatmap showing methylation intensities across both tumour and normal samples by selecting a subset of genes related to the cancer dataset. Tumour samples demonstrated distinct clustering patterns in the hypomethylated regions, which were visually distinguished by colour gradients, indicating varying methylation levels. Such clustering suggests possible epigenetic signatures that could serve as biomarkers for tumour classification. The heatmap provides an immediate visual comparison, identifying sample-specific methylation changes that could correlate with cancer progression or resistance mechanisms.

3.4 Summary Statistics for Hypo- and Hypermethylated Probes

The Summary Statistics tab further quantifies methylation trends, displaying a bar chart with counts of hypermethylated and hypomethylated probes across samples. Tumour samples exhibited a higher count of hypomethylated probes than normal samples, consistent with the established literature on the epigenetic downregulation of tumour suppressor genes in cancer (Esteller, 2008). In addition to the overall counts, a bar chart illustrating the top genes with the most methylation probes provided insights into gene coverage within the dataset, underscoring genes such as *ABCC5* that show significant methylation alterations. This feature offers a quantitative overview, helping researchers to quickly assess the prevalence of hypo- and hypermethylation patterns within their datasets.

3.5 Correlation Matrix of Co-Methylation Patterns

One of the most powerful features of EpigenomeViewer is its Correlation Matrix for analysing co-methylation patterns, which uses Spearman's correlation to accommodate non-linear relationships often found in methylation data. The correlation matrix for selected probes in *ABCC5* (Table 1) identified strong correlations between probes, with correlation values over 0.8 and statistically significant p-values. High correlations between probes such as cg18728109 and cg13341498 (0.876, $p < 1.27E-80$) suggested possible co-methylation patterns, which may indicate joint regulatory control. This feature is particularly valuable for understanding coordinated methylation patterns, providing insights into potential regulatory networks, and supporting hypotheses regarding functional interactions among methylation sites. These findings align with epigenetic research, suggesting that co-methylation across multiple probes can influence gene expression, impacting pathways relevant to disease progression and drug response.

Table 1. Correlation matrix for ABCC5 gene methylation probes.

Probe1	Probe2	Correlation	P.Value
cg18728109	cg13341498	0.876	1.27E-80
cg18728109	cg24375736	0.864	2.46E-60
cg18728109	cg22599229	0.839	2.19E-57
cg13341498	cg22599229	0.82	1.27E-55
cg24375736	cg22599229	0.81	2.07E-54

Correlation analysis reinforces the application of EpigenomeViewer in identifying and exploring co-methylation patterns, aiding researchers in understanding the potential co-regulatory mechanisms at play in complex diseases such as cancer. By utilising Spearman correlation, EpigenomeViewer accommodates non-parametric relationships, making it a more reliable measure for methylation data where nonlinearity is common.

3.6 Export Options for Publication-Ready Figures

An essential aspect of EpigenomeViewer's functionality is its image features, which allow users to download publication-ready figures in PNG formats. This capability enables researchers to include high-quality visualisations in manuscripts, presentations, and reports without additional formatting in the graphic design software. The ability to copy or save images of violin plots, heatmaps, and correlation matrices directly from the application simplifies the publication process, making it easier for researchers to document and share their findings.

4. DISCUSSION

EpigenomeViewer offers a unique combination of accessibility, interactivity, and functionality, which addresses the limitations of existing methylation analysis tools. Tools such as MethVisual and the UCSC Genome Browser are well regarded for visualising methylation data but are limited by static visualisations, which can hinder exploratory analysis (Haeussler et al., 2019). In contrast, EpigenomeViewer provides interactive, real-time filtering, and data exploration options, allowing researchers to dynamically adjust visualisation parameters to focus on specific genes or methylation probes. This feature is particularly useful for researchers working with large datasets, as they can explore different subsets of their data without requiring extensive reformat or preprocessing, which is often required by other tools.

Compared with MethyLumi in R, which requires command-line knowledge for preprocessing and normalising methylation data, EpigenomeViewer offers a user-friendly interface that accommodates researchers with limited programming experience (Du et al., 2010). While MethyLumi is highly robust for data preprocessing, it lacks interactive visualisations and requires additional coding to generate publication-ready figures, which EpigenomeViewer provides natively. By streamlining the data upload and visualisation process, EpigenomeViewer reduces the time and complexity associated with data preparation, enabling researchers to focus on analysis rather than technical adjustments.

The correlation matrix feature in the EpigenomeViewer distinguishes it from other tools, allowing researchers to identify co-methylation patterns that may indicate joint regulatory functions across methylation sites. This feature is absent in many other platforms, which generally focus on individual methylation site analysis rather than exploring the relationships between sites (Jones & Baylin, 2002). By enabling users to explore these correlations, EpigenomeViewer provides insights into potential epigenetic interactions and regulatory networks that could be critical for understanding diseases such as cancer. For example, co-methylation analysis can identify clusters of probes associated with gene regions that may be co-regulated during cancer progression or drug resistance.

One of the notable strengths of EpigenomeViewer is its ability to produce publication-ready figures without requiring extensive customisation or export to other software. While many tools require researchers to perform post-processing in graphic design or statistical software to obtain high-quality images for publication, EpigenomeViewer provides downloadable visualisations that meet publication standards. This capability aligns with recent shifts in bioinformatics towards platforms that offer reproducible, easily exportable figures, reducing the barrier to sharing results with the broader scientific community (O'Donoghue, 2021).

EpigenomeViewer also aligns with the United Nations Sustainable Development Goals (SDGs), specifically, SDG 4 (Quality Education) and SDG 9 (Industry, Innovation, and Infrastructure). By providing an accessible, user-friendly tool for methylation data visualisation, EpigenomeViewer supports SDG 4 by promoting inclusive and equitable access to advanced bioinformatics tools for researchers and students worldwide. Intuitive design enables students and early career researchers to learn methylation analysis techniques without requiring extensive programming knowledge, thereby democratising access to cutting-edge scientific methodologies. Additionally, the application supports SDG 9 by fostering innovation within bioinformatics and promoting a resilient research infrastructure. The open-access nature of EpigenomeViewer encourages collaboration and innovation in epigenetics, facilitating discoveries that can drive advancements in healthcare and genomics.

Despite these strengths, EpigenomeViewer is not without its limitations. The current version focuses primarily on visualisation rather than extensive preprocessing, such as normalisation or batch correction. While this keeps the application lightweight and accessible, researchers may still need to preprocess their data using other tools such as MethyLumi, minfi, or CHAMP in R for accurate

downstream analysis (Aryee et al., 2014). Additionally, the application currently supports only methylation data and does not yet integrate multi-omics data such as transcriptomics or proteomics, which could provide a more comprehensive view of gene regulation. Future developments will address these limitations by incorporating preprocessing options and expanding the compatibility with other data types.

5. CONCLUSION

EpigenomeViewer is a robust, user-friendly tool for visualising and analysing DNA methylation data, and is tailored to researchers in epigenetics and cancer biology. It combines interactive features with advanced visualisation capabilities, such as violin plots, heatmaps, and correlation matrices, to address challenges, such as co-methylation pattern identification and the creation of publication-ready figures. Prioritising accessibility requires minimal programming skills and supports data formats from popular analysis pipelines, such as CHAMP and minfi, making it versatile for various studies. Its ongoing development enhances features, contributing to a more inclusive and efficient research environment, aligning with global efforts to advance education (SDG 4) and promote innovation in scientific infrastructure (SDG 9).

Acknowledgements: The authors acknowledge the support of several grants for this project, including FRGS/1/2020/SKK0/UKM/02/11 from the Ministry of Higher Education Malaysia and GP-K020141 from Universiti Kebangsaan Malaysia.

References

Aryee, M. J., Jaffe, A. E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A. P., Hansen, K. D., & Irizarry, R. A. (2014). Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics, 30(10), 1363-1369.

Attali, D. (2021). shinycssloaders: Add loading animations to 'shiny' outputs. R package version 1.0.0. https://CRAN.R-project.org/package=shinycssloaders

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes & Development, 16(1), 6–21.

Chang, W., Cheng, J., Allaire, J. J., Xie, Y., & McPherson, J. (2021). shiny: Web application framework for R. R package version 1.6.0. https://CRAN.R-project.org/package=shiny

Du, P., Zhang, X., Huang, C. C., Jafari, N., Kibbe, W. A., Hou, L., & Lin, S. M. (2010). Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics, 11, 587.

Esteller, M. (2008). Epigenetics in cancer. New England Journal of Medicine, 358(11), 1148–1159.

Haeussler, M., Zweig, A. S., Tyner, C., Speir, M. L., Rosenbloom, K. R., Raney, B. J., ... & Kent, W. J. (2019). The UCSC Genome Browser database: 2019 update. Nucleic Acids Research, 47(D1), D853-D858.

Harrell, F. E. (2020). Hmisc: Harrell miscellaneous. R package version 4.4-2. https://CRAN.R-project.org/package=Hmisc

Jones, P. A., & Baylin, S. B. (2002). The fundamental role of epigenetic events in cancer. Nature Reviews Genetics, 3(6), 415–428.

Kassambara, A. (2019). ggcorrplot: Visualization of a correlation matrix using 'ggplot2'. R package version 0.1.3. https://CRAN.R-project.org/package=ggcorrplot

Kolde, R. (2019). pheatmap: Pretty heatmaps. R package version 1.0.12. https://CRAN.R-project.org/package=pheatmap

O'Donoghue, S. I. (2021). Grand challenges in bioinformatics data visualization. Frontiers in Bioinformatics, 1, 669186.

R Core Team. (2021). data.table: Extension of data.frame for fast data manipulation. R package version 1.14.0. https://CRAN.R-project.org/package=data.table

Tian, Y., Morris, T. J., Webster, A. P., Yang, Z., Beck, S., Feber, A., & Teschendorff, A. E. (2017). ChAMP: Updated methylation analysis pipeline for Illumina BeadChips. Bioinformatics, 33(24), 3982–3984.

Wickham, H. (2016). ggplot2: Elegant graphics for data analysis. Springer-Verlag. https://ggplot2.tidyverse.org

Wickham, H., & Bryan, J. (2021). readxl: Read Excel files. R package version 1.3.1. https://CRAN.R-project.org/package=readxl

Wickham, H., & Henry, L. (2021). tidyr: Tidy messy data. R package version 1.1.3. https://CRAN.R-project.org/package=tidyr

Wickham, H., & Miller, E. (2021). shinythemes: Themes for shiny. R package version 1.2.0. https://CRAN.R-project.org/package=shinythemes

Wickham, H., François, R., Henry, L., & Müller, K. (2019). dplyr: A grammar of data manipulation. R package version 0.8.3. https://CRAN.R-project.org/package=dplyr

Xie, Y., Cheng, J., & Tan, X. (2021). DT: A wrapper of the JavaScript library DataTables. R package version 0.18. https://CRAN.R-project.org/package=DT