International Journal of Research in E-learning
Vol. 10 (1), 2024, pp. 1–24

DOI 10.31261/IJREL.2024.10.1.03

Justyna Marcinkowska
Poznan University of Medical Sciences, Department of Computer Science and Statistics
https://orcid.org/0000-0002-6510-5538

Magdalena Roszak
Poznan University of Medical Sciences, Department of Computer Science and Statistics
https://orcid.org/0000-0001-6495-6771

Interactive RShiny Reports – Independence and Autonomy in Medical Data Visualization

Abstract

The article discusses the use of the R language, especially the RShiny tool, to create interactive reports in the field of medical data analysis. The authors emphasize the need for an interdisciplinary approach to teaching statistics among medical students. An alternative to traditional static reports was presented, proposing the creation of interactive web applications that enable exploration, analysis and visualization of medical data changing in real time. The use of the R language as an open source tool allows for the development of medical students' competences in the field of data analysis and the adaptation of research tools to individual needs. A draft lesson plan using sample medical data on cervical cancer was also presented, along with a proposal for specific analyzes of these data and their visualization using interactive RShiny reports. The article ends with a discussion on the role of learning the R language in the education of students of Polish medical universities and the need to expand the educational offer for them with courses in data analysis in an open source environment.

K e y w o r d s: RShiny, R, data visualization, interactive report, statistics, medical teaching

The motivation for creating an interactive teaching project

Teaching data analysis to medical students requires an interdisciplinary approach. Understanding how complex mathematical formulas and data analysis techniques impact topics in biology and chemistry (fields central to our students' interests) is crucial. Typically, biostatistics courses illustrate data analysis examples using commercial programs available at a given university. Of course, every instructor aims to minimize the time spent on specific software during classes, choosing packages with an intuitive interface and sufficient capabilities for implementing the teaching material. This is the right approach: students who comprehend the discussed issues and learn how to calculate them using the software used in classes will likely adapt to using another statistical package in the future and perform similar calculations.

However, it is important to recognize the abundance of available software with varying capabilities. Learning how to use each new package consumes valuable time, particularly for students or clinicians. Furthermore, as we advance in the scientific field, it becomes apparent that even mastering and acquiring the latest software may still be insufficient to perform all the analyses related to our study. So, what alternatives do we have, and in what direction should education progress in order to become a true guide on the path of development for future scientists?

The present state of biostatistics instruction

The current approach to teaching biostatistics at the Poznan University of Medical Sciences is tailored to the allocated hours for this subject within a specific field. Some classes also incorporate elements of online teaching. Whether conducted remotely or in-person, the focal point remains practical application (Roszak et al., 2020). For medical majors, the emphasis is on the hands-on aspect of performing analyses frequently employed in scientific research within the medical domain.

This involves delving into the fundamental principles governing individual statistical tests, allowing students to interpret results obtained through dedicated software. Historically, this methodology has sufficed for most students, involving executing specific commands according to discussed frameworks and in alignment with the capabilities of the software in use. The prevalent form of presenting results has been static reports, essential for conveying findings in scientific publications. Naturally, during the initial phases of learning basic analyses, students document and describe each result covered in classes.

Currently, at Poznan University of Medical Sciences, four commercial software licenses are available for teaching biostatistics: Statistica, PQStat, MedCalc, and JMP. Illustrated in Figure 1 is a typical report of a task calculated using the Statistica program during biostatistics classes:

H0 : the data distribution follows a normal distribution

H1 : the data distribution does not follow a normal distribution

Aggregated results
Normality tests (Sheet in descriptive)

variable	gender	N	W	p
height	female	100	0,983807	0,259734
height	male	102	0,966666	0,011156

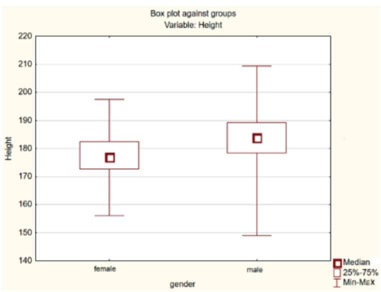
α= 0,05 Mann-Whitney tesst α -> I accept H1

H0 : ϕ1 = ϕ2 – women and men do not differ statistically significantly in height

H1 : ϕ1 ≠ ϕ2 – women and men differ statistically significantly in height

U Manna-Whitneya test (with continuity correction) (athletes in description)
Relative to the variable: Gender
The marked results are significant with p <.05000

variable	sum of ranks	sum of ranks male	U	Z	p	Z corrected	p	N female	N male
Height	8212,000	12291,00	3162,000	-4,66428	0,000003	-4,66448	0,000003	100	102



Women and men differ statistically significantly in height.

Figure 1. An example static report from biostatistics classes generated in the Statistica program

S o u r c e: Authors' own work.

However, real data mining requires calculating many simple analyzes and deciding whether and what more advanced models we will need in our study. Efficient analysis of basic results is crucial here. This is noticed by young researchers who analyze their data within scientific clubs, or when analyzing their own or team research work. We think about them with the greatest care, because they are the future of science. Providing them with optimal data analysis tools is the goal that every teaching and research team should set.

Choosing R as an option

This article aims to emphasize the existence of a tool that can serve as a robust foundation for the development of students or doctoral candidates in the field of data analysis and scientific research. Just as proficiency in English serves as a crucial communication tool in the realm of science, similarly, in the domain of data analysis, mastery of the R language is indispensable. As open-source software, available for free to everyone, R offers numerous benefits and advantages. Designed with statistics in mind, the R language excels as a tool for performing calculations and processing data. It is an excellent choice for students outside the IT field who primarily want to use programming skills for data analysis and processing. Moreover, learning R prepares students to acquire future competencies valued in the world of science. The R language boasts a vast community of users and package developers(*CRAN: Manuals,* n.d.). providing numerous built-in statistical functions and eliminating the need to search for them in additional libraries. The availability of free educational materials, online courses, support forums, and extension packs facilitates students' learning journey, enabling them to expand their skills. University classes can be organized both on-site and on e-learning platforms, offering a synchronous or asynchronous teaching. It is crucial to note that the R language allows the creation of non-standard solutions and tools tailored to specific research needs (Biecek, 2017), making it more flexible than commercial programs with limitations in adapting to unusual tasks.

Researchers using R are not constrained by the capabilities of specific software, allowing exploration of the latest data mining and statistical analysis techniques, consultation, and discovery of better, more tailored solutions in their research area. Programming in R enables the creation of custom tools that empower medical scientists to explore data and test hypotheses, expediting discoveries in clinical medicine. Proficiency in programming basics automates routine tasks related to data analysis and clinical trials, enhancing the efficiency of scientists' work and saving valuable time. When sharing test results with other specialists and patients, interactive applications and reports offer virtually unlimited possibilities for visualizing test results.

Interactive resources

Facilitating interactivity and visualization is the heart of good pedagogy (SL Wang et al., 2021). Simulations, in turn, are a very effective teaching method because computer software offers them the opportunity to review concepts multiple times (Jamie, 2002). All this means that today's educational materials, especially e-learning (Grześkowiak et al., 2020), have so many interactive elements (Kumar et al., 2017). Graphical representations of data, such as interactive charts, allow you to quickly communicate key information. The human brain is more likely to process visual information quickly than text. The ability to manipulate data allows users to experiment with different perspectives and scenarios. This, in turn, helps you understand the context of your data, eliminating the need to analyze large (and boring) results tables.

Among many tools that enable the creation of such materials, the most popular are those that create web applications. They only require a browser, which makes them more accessible. Hence, these technologies can replace the capabilities of applications created on traditional platforms that are no longer fully supported, such as Java or Flash.

Perhaps the best-known collection of online applications for teaching statistics is the Rossman-Chance applet collection, available to educators at www.rossmanchance.com/applets/. It is a software written mainly in JavaScript and is considered as an example of a well-designed educational application.

One of many tools related to the R language, giving unlimited possibilities for creatively presenting the results of your own research, are interactive dashboards and web applications. Since RShiny is an R package, researchers can leverage their R coding knowledge and experience to create RShiny applications without additionally learning another programming language. Of course, advanced interactive RShiny reports can be expanded with additional CSS themes, HTML elements, JavaScript scripts, etc. But this is not necessary, and they can successfully be used only at the basic stage.

RShiny

Since the creation of RShiny interactive reports, they have become a popular tool for creating web applications. Thanks to them, you can replace hundreds of pages of traditional static reports with interactive applications that provide self-service data analysis and easy sharing of analysis results. And all of this without advanced programming skills. A chart with „independence and autonomy" characteristics is one that is self-contained and entirely understandable to the audience. It does not require additional explanations. Individual elements of the chart (such as axes, legends) are clear enough for the chart to operate autonomously. Through interactive reports, we have a greater chance of choosing a medical data illustration that is fully comprehensive, allowing, for example, to shorten the time associated with selecting the direction for further data exploration by the research team.

Many useful and easy-to-use applications of RShiny web application have been implemented so far in the field of medicine and health sciences. These include:

- HPVTIMER: A shiny web application for tumor immune estimation in human papillomavirus-associated cancers (Liu et al., 2023), created at the University of Houston
- RIMeta – an RShiny application for estimating the reference interval from a meta-analysis (Jiang et al., 2023), University of Minnesota, USA.

- ShinyBioHEAT – an RShiny application enabling the identification of the gene driving the phenotype in two commonly used model bacteria, E. coli and Bacillus subtilis (C. Wang et al., 2023), Baylor College of Medicine, Houston.
- RNA-Seq Ontology Graphic User Environment – an RShiny application for RNA sequencing analysis and biomarker discovery (Farel et al., 2023), written by scientists from the USA, National Institutes of Health, Bethesda

The number of such applications is increasing rapidly. Thanks to RShiny, specialists can visualize and share the results of their work with a wide audience. This tool is also used by startups and their commercial implementations. Noteworthy, for example, is the Polish application Mbaza AI, which supports the protection of endangered animal species. It was recognized in 2023 as one of the 10 best applications in the world by the International Center for Research on Artificial Intelligence (IRCAI) under the auspices of UNESCO (‚Polish application Mbaza AI one of the best in the world according to UNESCO', 2023).

RShiny's interactive reports can also be used to support teaching during data analysis classes for medical university students. They can constitute a reliable teaching base that facilitates creative exploration and analysis of data. The purpose of such statistical education applications is to enrich the teaching of both introductory and upper-level statistics courses. To date, many RShiny applications have been developed around the world, and the collection is constantly growing. As an example goes the collection of RShiny applications created for preliminary statistics at Grand Valle State University (*Daniel Adrian | IntroStat Shiny Apps,* n.d). Another example is University Park, Pennsylvania, which recognized the benefits of RShiny's interactive reports and used them to try to incorporate research-based learning into a senior-year undergraduate statistics course (SL Wang et al., 2021). Interactive graphics and dynamic visualizations built in the RShiny environment used for working with analyzed data were also appreciated by veterinary sciences from the University of Veterinary Medicine Hannover (Liebig et al., 2022).

At the Poznan University of Medical Sciences, there are plans to introduce classes using the RShiny application, enabling future scientists to integrate medical competences with data analysis skills. As a proposal to use the possibilities of R in biostatistics classes, the authors present the project whose aim is to create a web application allowing the exploration, analysis and visualization of medical data. It will generate interactive RShiny reports. Web applications during classes will be built using the RShiny package, while on the server side interactive data processing will be made using R script (version 4.0.2, https://www.r-project.org/).

The classwork should revolve around examples closely related to the students' field of interest (Pfannkuch, 2011). Ideally, this involves using real datasets from specific clinical trials or medical scenarios. In this project, a publicly available database containing information like age, sex, diagnoses, and laboratory test results was utilized. Through RShiny, we developed an interactive user interface that facilitated data viewing and analysis. These classes provide a valuable opportunity for students to independently explore data, adjusting variables and parameters to observe their impact on diagnoses. Importantly, advanced programming skills are not required, as only a basic knowledge of the R language is sufficient to design tables, interactive charts, and dashboards. The added advantage of RShiny lies in its ability to visualize and share results online without the need for extensive web coding.

The three basic parts needed to create an RShiny application (Wickham, 2021), then:

- **ui**: nested R functions that assemble an HTML user interface for your application
- **server**: a function with instructions on how to build and rebuild the R objects displayed in the UI
- **shinyApp**: combines ui and server into an application. Wrap with *runApp()*.

Below (Figure 2) is an example code of the simplest application that generates a histogram based on random data with a normal distribution generated using the *rnorm()* function (Wickham, 2023).

```
# This is a Shiny web application

library(shiny)
# Define UI for application that draws a plot histogram
ui <- fluidPage(

    numericInput(inputId ="size",
                 "Sample size",
                 value=25),
    plotOutput(outputId="our_histogram")
)

# Define server logic required to draw a histogram
server <- function(input, output) {

    output$our_histogram <- renderPlot({
      hist(rnorm(input$size))
        })
}
# Run the application
shinyApp(ui = ui, server = server)
```

add input to the UI with functions – numericInput()
add R object/output to the UI with fucntions – plotOutput()

tell howto build the output from the input in functions renderPlot()
assign renderPlot() to the output

Figure 2. Sample R code for an application that generates a histogram based on random normally distributed data generated using the *rnorm()* function

Source: Authors' own work.

The RShiny application interface may contain tabs or panels that allow you to select specific views and analyzes (Wickham, 2021). In the prepared tool, using the ggplot2 package or other graphic packages related to R, we will show how individual variables are related to each other and what relationships can be noticed in the analyzed medical data. For example, you can simulate the impact of variable X on outcome Y depending on various scenarios and parameters. Providing students with the opportunity to explore various statistical models, manipulate the assumptions and parameters of these models, and even generate data from the model can improve understanding of the medical problem being analyzed as well as the statistical issues being discussed (Garfield & Ben-Zvi, 2007).

Data used in the project

The project in question used the so-called open medical data. Students participating in classes should be made aware of the legal aspects of data use. Users must be sure that we meet all conditions for using data found online, provided by research institutions or organizations, and respect the principles of privacy and ethics. The most popular data sources that can be used to work with students are: Kaggle, UCI Machine Learning Repository and GitHub. These platforms offer access to a truly diverse range of medical data.

The project presented in the article used a data set collected at the „Hospital Universitario de Caracas" in Caracas, Venezuela, and made available in 2017 under the CC BY 4.0 license (Fernandes & Fernandes, 2017). This dataset was created to analyze indicators related to cervical cancer diagnosis and includes, among others: demographic information, patient habits and medical records (858 records). Some patients did not decide to answer all the questions, so we also have missing values in the analyzed database.

Cervical cancer is closely associated with human papillomavirus (HPV) infection, which is transmitted sexually. Infection with this virus can lead to chronic infections, increasing the risk of genital and anal cancer. HPV is a common sexually transmitted infection, especially among young people. In addition to HPV infection, there are several risk factors associated with cervical cancer, such as age, early sexual initiation, number of sexual partners, long-term smoking, low socioeconomic status, and history of cervical intraepithelial neoplasia (CIN). It is also worth taking into account other risk factors, such as long-term use of hormonal contraceptives, a diet low in antioxidants and untreated vaginal inflammation caused by Chlamydia trachomatis, Neisseria gonorrhea and HSV-2.

The biopsy is the so-called the gold standard in the diagnosis of cervical cancer, therefore it was taken as a dependent variable in the analyzes performed.

Examples developed in the project

Descriptive statistics

Below (Figure 3) is an example screenshot of the RShiny application with a description of the interactively selected continuous variable. You can discuss during class what elements of descriptive statistics (Guzik & Więckowska, 2023) we want to present for specific data and how it would be best to visualize them.

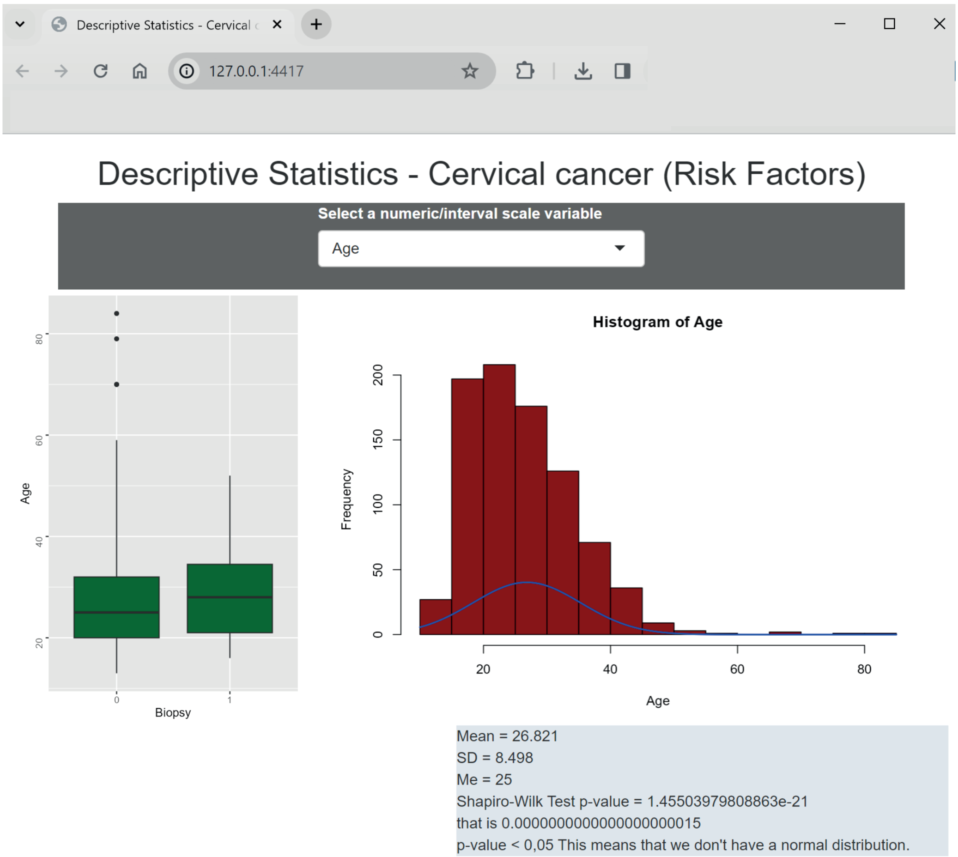


Figure 3. Sample screenshot of the RShiny app with descriptive statistics

S o u r c e: Authors' own work.

In the presented example, the application running in a browser is designed to select a numeric variable using a selection list. For the selected variable, the simplest descriptive statistics were automatically determined and charts were generated: in this case, a *boxplot* and a *histogram*. Class participants can independently explore subsequent variables to learn the specifics of the analyzed database.

Comparison of two groups

The next image (Figure 4) shows a screenshot of the application that presents interactive results of comparing a selected variable in two independent groups that received a positive and negative biopsy result. Then, depending on the results of the test checking the compliance of the selected variable with the normal distribution (Guzik & Więckowska, 2023), a comparison is made using the parametric or non-parametric method (Milewski & Roszak, 2023).

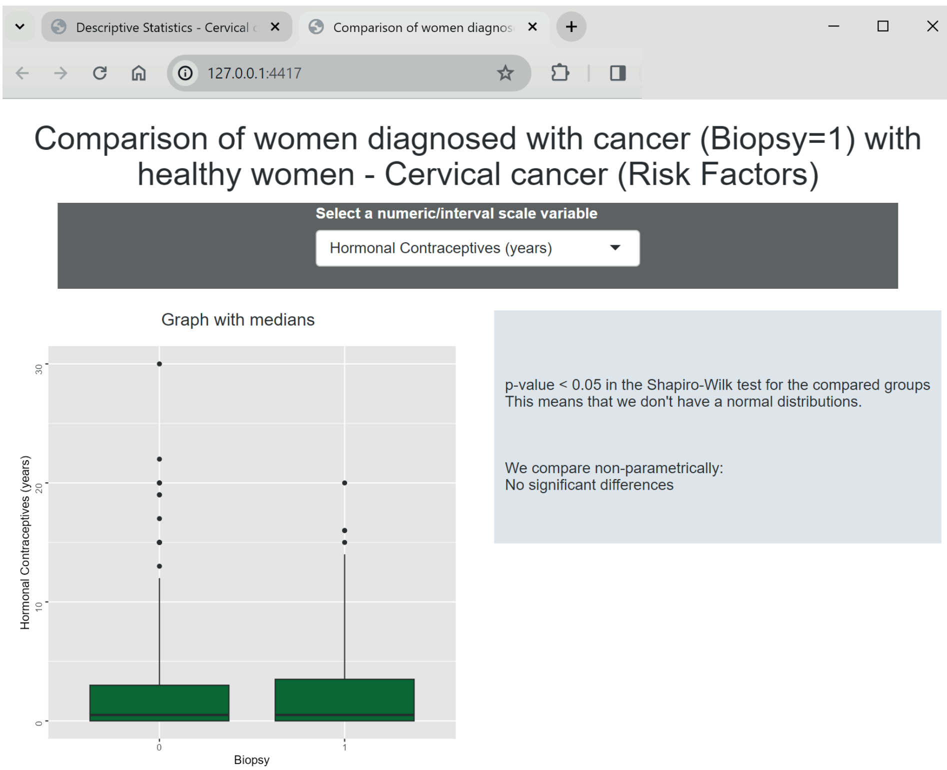


Figure 4. Sample screenshot of the RShiny application with the results of comparing two variables

Source: Authors' own work.

ROC curve

One example of statistical analysis very often used for medical data is ROC (Receiver Operating Characteristic) curve analysis. It is a tool used in medical diagnostics to evaluate the performance of a diagnostic test based on the chosen cut-off point set for a specific variable. The ROC curve shows the relationship between sensitivity (true positive rate) and specificity (1 – false positive rate) depending on the different thresholds used in the diagnostic test. In practice, the higher the ROC curve, the better the diagnostic test, and the Area Under the Curve (AUC) measures the overall effectiveness of the test (Biecek, 2013). This method is widely used in fields such as oncology, cardiology, and neurology to evaluate diagnostic tests such as imaging tests, laboratory tests, and diagnostic questionnaires.

Below (Figure 5) an example of an application drawing a ROC curve for selected variables. By analyzing the displayed results one by one, you can try to answer the question: Will a given variable be useful in detecting cancer?

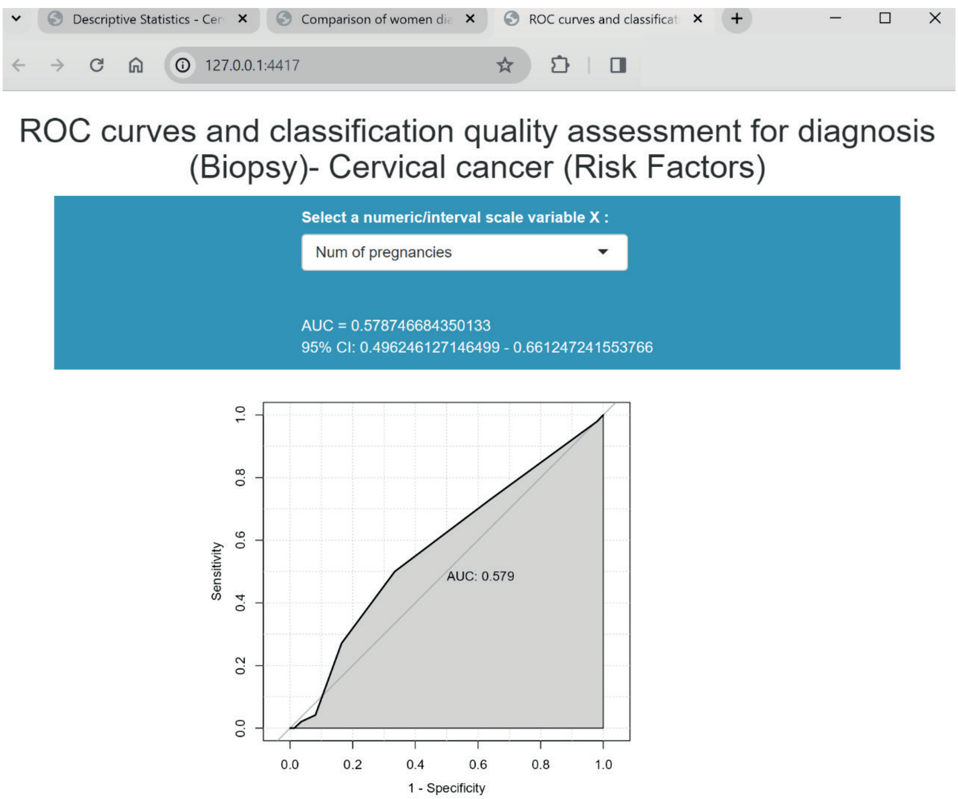


Figure 5. Sample screenshot of the RShiny application with ROC curve analysis results

Source: Authors' own work.

Correlation study

Another way to look at the data analyzed in the project is to analyze their interconnections. Studying the correlation between selected variables can shed new light on the interactions between them. (Hegarty, 2004) Below (Figure 6) an example operation of the application is presented, in which the user selects two variables and interactively observes the changing calculations and a scatterplot regarding the analyzed relationship.

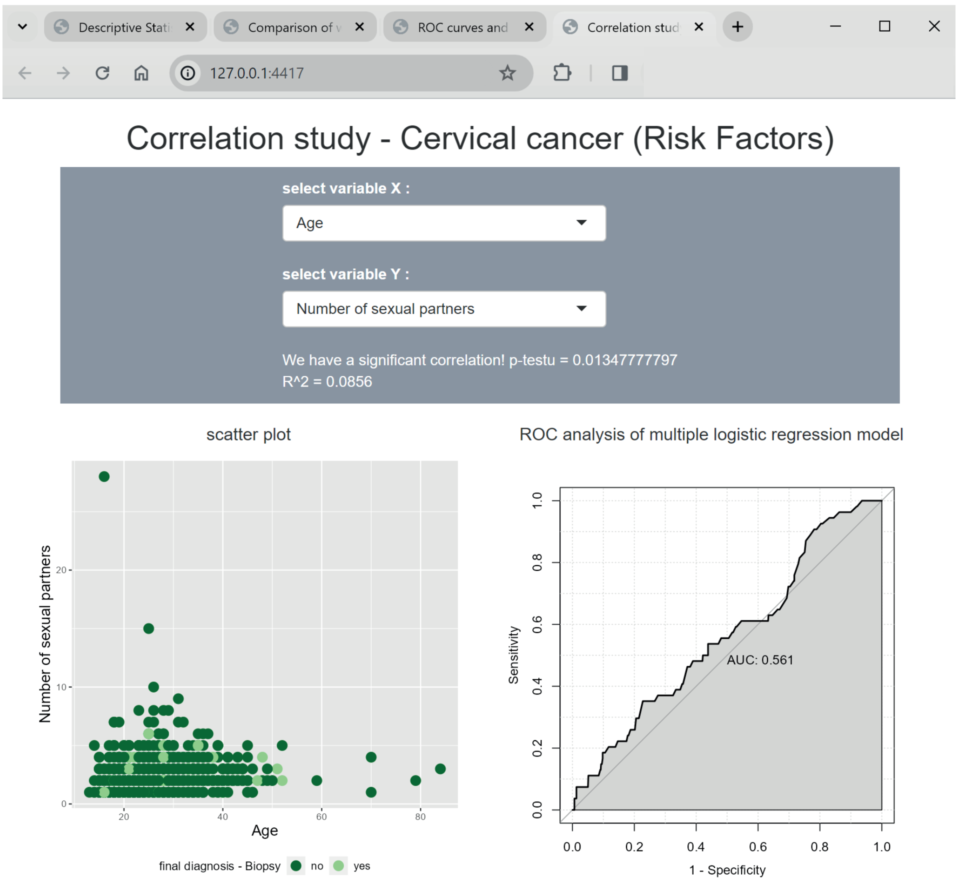


Figure 6. Sample screenshot of the RShiny application with correlation analysis results

Source: Authors' own work.

In the above example (Figure 6), patients with positive and negative biopsy results are additionally marked in color. Such an overview of the relationships between various variables may provoke further exploration of the data.

Advanced models

An additional advantage of using RShiny's interactive reports may also be the designed simulation of single- and multi-dimensional models, the changing parameters of which will suggest the advantage of a specific research concept. You can encourage class participants to discuss the specificity of the analyzed measurements and explore various possibilities for designing a model to support cancer diagnosis. Figure 6 additionally shows the ROC plot related to the evaluation of the logistic model (using the *glm()* function) for two selected independent variables. This is, of course, a version for students learning more advanced methods during compulsory or optional classes. A logistic model is a statistical tool used to model the relationship between one or more independent variables and a dependent variable (Fawcett, 2018). In this case, the dependent variable is the biopsy, which takes binary values (yes/no). This model is based on a logistic function that has the shape of a sigmoidal curve. In practice, the logistic model is used when we predict the probability of belonging to one of two categories (biopsy). This model is often used in the analysis of medical data, disease risk prediction or subgroup classification.

In the case of an extended timeframe for working with an advanced group of students, it is also possible to introduce visualization for multidimensional logistic models. However, it is crucial to ensure that the assumptions of each analysis are met. Students are expected to determine which parameters of the resulting model are essential and should be presented in the research report.

Implementation of the project in classes

This section introduces two distinct proposals for implementing the lesson plan outlined in the preceding part, contingent on the allocated number of class hours for a specific subject.

A Suggestion for Classes with Limited Hours

When dealing with classes of shorter duration, a viable approach is to provide students with templates. These templates could be files containing a pre-prepared operational R code structure, ready for completion during class. This way, students are not required to code the entire RShiny application from scratch; instead, they focus on the specific sections related to direct data analysis. Furnishing students with code snippets for independent work can be an effective means of teaching, offering them exposure to programming basics, even if they have no prior experience. Suggested code additions for students may include invoking relevant statistical tests or modifying visualizations.

For instance:

- Modifying the chart type,
- Introducing additional geometries to an existing plot in the ggplot() function, like points, lines, bars, etc.
- Altering the scale of the x or y-axis,
- Adjusting parameters related to the chart background, font color, and including titles and labels.

In the R environment, we can craft charts that are precisely tailored to the specific requirements of the data analysis author.

Recommended Classes for Advanced Students

For more advanced students, this example can be used to delve into the fundamentals of programming. In the presented R code example below (Figure 7), BMI coefficients are calculated. Subsequently, they are described within individual groups using an appropriate measure of descriptive statistics, depending on whether their distribution adheres to the normal distribution (Guzik & Więckowska, 2023). The decision is made using the logical function if() or ifelse(). This function is a fundamental tool in many programming languages, including R, for creating conditions in code (Biecek, 2017). It allows the execution of a specific instruction or block of instructions only when a specific condition is met. Another very useful instruction is ‚for', which enables performing specific analyses for the entire list of variables – in the code below (Figure 7), for each of the many subgroups of patients treated with different methods.

For students interested in coding, there is the opportunity to create personalized functions or packages, which are sets of functions. One of the significant advantages of advanced coding in R is the capability to perform multiple analyses simultaneously within a written script. This script, essentially a set of instructions or commands compiled in a single file, serves a specific task. Another undeniable advantage is the independence to generate reports and visualize data according to the specific guidelines set by the code author (researcher) and easily modify them.

```
# Function to calculate BMI
calculate_BMI <- function(weight, height) {
  bmi <- weight / (height^2)
  return(bmi)
}

# A for loop to calculate BMI for different subgroups
unique_groups <- unique(data$patient_group)


# Initialization of empty lists into results
description_BMI <- list()

# A for loop to calculate average BMI for different subgroups
for (group in unique_groups) {
  subgroup <- data[data$patient_group == group, ]
  BMI_column_name <- paste("BMI", group, sep = "_")
  subgroup[[BMI_column_name]] <- calculate_BMI(subgroup$weight, subgroup$height)

   # Checking the normality of distribution using the Shapiro-Wilk test
  p_value <- shapiro.test(subgroup[[BMI_column_name]])$p.value

  # If p-value less than 0.05, calculate the median instead of the mean
  if (p_value < 0.05) {
    description_BMI[[group]] <- median(subgroup[[BMI_column_name]])
  }else{
    # Calculation of the average BMI for a given subgroup
    description_BMI[[group]] <- mean(subgroup[[BMI_column_name]])
  }

  # Display the results
  print(paste("Mean BMI (or median if irregularly distributed) for the group", group, ":",
              description_BMI[[group]]))
}
```

Figure 7. Sample R code for an application that calculates BMI in patient subgroups

Source: Authors' own work.

Discussion

Contemporary medicine and clinical research heavily depend on data analysis. Evidence-Based Medicine (EBM) represents an approach to medical practice where clinical decisions hinge on the best available scientific evidence. It has become the standard in medical research, striving to guarantee that healthcare is evidence-based and personalized to meet patients' needs. Currently, 18 Regional Digital Medicine Centers are being established in Poland, aiming to aggregate comprehensive clinical data from local hospitals within a span of 5 years. (Creation and development of Regional Digital Medicine Centers–2023–Medical Research Agency, 2023). This nationwide initiative is modeled on European practices. The establishment of these centers necessitates individuals with competencies in both medicine and a broad range of data analysis skills. Graduates from medical universities equipped with enhanced proficiency in working with data across various tools, including open source platforms, will be well-positioned to contribute to research teams, including those associated with these centers.

Students undergoing education at medical universities should not limit their learning experience solely to the use of commercial software, as has been the case until now. While paid statistical programs offer advantages, such as a user-friendly interface, they should not be entirely excluded from educational practices. However, acquiring knowledge in biostatistics and working with data in an open-source environment (e.g., R) can serve as a valuable complement to students' skills, providing them with opportunities to engage with entities like University Clinical Research Support Centers or the aforementioned Regional Digital Medicine Centers. The amalgamation of scientific knowledge enables future doctors to approach medical issues more precisely and effectively. Therefore, if there is a discernible interest among medical students in scientific pursuits, it is crucial to support and facilitate their usage and development of these skills. The willingness of medical university students to enhance their expertise in areas such as biostatistics, bioinformatics, or medical physics is an invaluable phenomenon. This allows them to participate in interdisciplinary research teams, contributing to scientific discoveries and innovations in the realms of medicine and health sciences.

Numerous medical universities in Poland not only provide courses that cultivate skills related to fundamental data processing, the utilization of statistical tools, and the interpretation of scientific research findings but also offer classes aimed at extending this knowledge for independent data exploration using open-source tools. A noteworthy example is the bachelor's studies program in Clinical Biostatistics at the Medical University of Bialystok, which stands out for its uniqueness and deserves significant recognition (Laureaci_szp_2023_2. Pdf, n.d). The field of Clinical Biostatistics at the Medical University of Bialystok serves as an alternative to both theoretical sciences like mathematics or computer science and practical fields associated with medical, pharmaceutical, or health sciences. Researchers from this university emphasize that engaging in biostatistics would be unattainable without operating in an open-source environment. As they assert, „In R, you can find libraries with all the methods we are looking for," a level of convenience unmatched by commercial software.

Medical universities in Poland are progressively broadening their curriculum at the intersection of exact sciences. The preeminent Polish institution, the Jagiellonian University Medical College, presently provides elective classes in the R environment within the Department of Bioinformatics and Telemedicine (the JU MC Syllabus, n.d.). Tailored for medical and dietetics students, these classes cover the assessment of tools for handling and analyzing data in scientific projects. Additionally, the university extends biostatistics classes, augmented with fundamental programming skills in R, to participants in the doctoral school.

The Poznan University of Medical Sciences also provides optional classes in data analysis using R and Python, serving as a valuable resource for aspiring scientists. Beyond the primary audience of medical students, these classes also attract participants from pharmacy, physiotherapy, and medical analytics. While recognizing the need for straightforward statistics classes with a user-friendly interface for all medical majors, there is a call to expand educational offerings. This expansion also enhances the competencies of medical university students, allowing them to integrate knowledge from diverse fields. Ideally, students should be introduced to both R and Python. However, due to constraints in time and budget, the decision on the language of instruction for biostatistics classes for medical students should be made judiciously.

The feedback from medical students who have participated in optional R language classes is overwhelmingly positive. They highlight the ‚practical side' of these classes and express the belief that the skills acquired ‚may prove useful when writing scientific papers in the future.' Among the aspects that garnered the most satisfaction is the independent design of graphic concepts for visualizing analyzed data. We concur with the notion that ‚clear and communicative data visualization strengthens the results, while ill-considered or illegible data visualization undermines the reliability of the analyses performed' (Discover! Reveal! Explain! A collection of essays on the art of presenting data – PDF, 2014). Below is an illustrative chart (Figure 8) created by a student from optional classes during the 'Data Analysis in R' class in the academic year 2023/24:

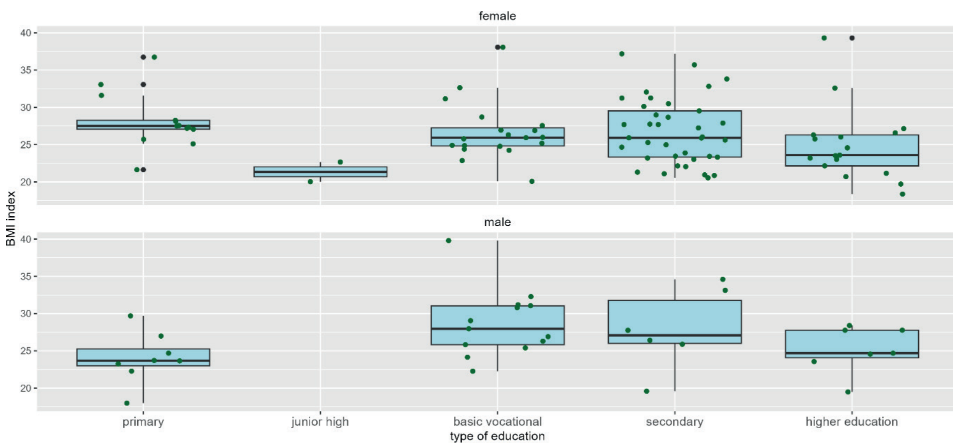


Figure 8. Graph generated during the optional class „Data analysis in R"

S o u r c e: Authors' own work.

At Poznan University of Medical Sciences, a distinctive, unique teaching initiative has been established, centered on interacting with dynamic RShiny reports in an open-source environment. In the information society, cultivating critical thinking skills and the proficiency to manipulate data through visual representations are paramount for fostering research competencies, facilitating decision-making, and ensuring effective communication (Murray, 2014). Today, information is available at our fingertips, so the ability to approach it critically becomes a key tool in developing knowledge (Central Statistical Office / Thematic Areas / Health, n.d.) (How to Gain Access to SRP – Ministry of Digitization – Gov.Pl Portal, n.d.) (Creation and Development of Regional Digital Medicine Centers–2023–Medical Research Agency (2023).

Visualizations play a crucial role in identifying trends, relationships, and patterns, thereby prompting the formulation of new research questions. Simultaneously, the capacity to generate or interpret visual representations enhances the effectiveness of conveying information to others and engaging in evidence-based discussions.

The meticulously crafted scenario is poised for execution, with due consideration for the intended recipients within medical universities, whether they be students or doctoral candidates. This culmination is the product of several years of research and pedagogical observations, pointing to a growing demand for interdisciplinary scientists. In the future, these individuals will integrate into research teams in academic units, both within universities (public sector) and the private sector operating in the medical domain. Suggesting classes where future scientists actively participate in programming and autonomously create data mining solutions within a web application embodies an innovative approach. It can be implemented in various formats, depending on the allocated class hours and the proficiency level of the participant group. Polish medical universities currently lack implementations in this didactic domain. Developing such a class concept requires high statistical and IT competencies, including programming, experience in teamwork on medical projects, and extensive teaching experience with medical students. Future publications will present the results of evaluating such classes to validate the aforementioned assumptions. The ultimate decision between R and Python hinges on the specific project needs and user experience, encompassing IT competencies. In practice, both platforms are frequently employed depending on the context of the task. It is crucial to tailor the tools to the educational objectives and understand the unique nature of the student group. For medical school students aiming to concentrate on medical data analysis, R may be a more immediate choice. Consequently, R was deliberately selected as the tool for this project.

Conclusions

The presented project, designed for medical university students, outlines a curriculum illustrating the utilization of the RShiny tool to develop interactive web applications, eliminating the necessity for advanced programming skills, particularly in JavaScript. Students will learn to construct RShiny applications, engaging in data visualization, statistical analysis, and model simulation, offering valuable skills for their future research and scientific endeavors. Unlike static reports limited to predefined parameters, these applications will produce dynamic visualizations and reports based on selected parameters. The results will be accessible through interactive charts and tables in the web browser. RShiny facilitates the creation of applications where users can dynamically choose variables and parameter settings through buttons, sliders, and other interactive elements. This approach fosters a more flexible and interactive exploration of data. The resulting series of visualizations allows for the selection of a graphical interpretation of the discussed research problem that is self-contained, requiring no additional explanations for the audience's full understanding.

Given the limited time available for medical students, this project offers a means to autonomously create dynamically evolving visuals that captivate attention, stimulate creative thought, and facilitate pattern recognition. Visual representations, including charts, graphics, and infographics, prove to be highly effective tools in analyzing and interpreting medical data. The innovative character of the presented research project (Table 1) sets it apart as a unique concept ready for integration into medical education, especially for the facilitation of engaging interactive sessions.

The proficiency in utilizing diverse tools, beyond just commercial solutions, can enhance students' adaptability and preparedness for the varied challenges they may encounter in their future scientific endeavors within the realm of medicine. The gradual incorporation of programming elements into classes and courses focused on data analysis within medical university curricula becomes imperative. This evolution is a natural response to the dynamic advancements in technology and the progress within biomedical sciences, where data analysis plays an increasingly pivotal role in enhancing healthcare and driving innovative medical research. Clinician scientists equipped with fundamental programming skills will be better equipped to tackle the future challenges in medicine.

The faculty engaged in the convergence of mathematical and biological domains should inspire medical students to broaden their perspectives within the realm of exact sciences. This presents an enticing opportunity, offering not only intellectual fascination but also a chance to enhance one's medical practice. Proficiency in areas such as mathematics, statistics, or computer science can serve as a meaningful contribution to advancements in diagnostics, therapy, and clinical research. Embracing this approach brings notable advantages, including the potential for more efficient utilization of modern technologies, precise analysis of medical data, and active involvement in innovative research initiatives. Developing competencies in interactive data visualization within this context will undoubtedly provide students with the gratification of expanding their knowledge while unlocking new vistas and challenges in the dynamic realm of evidence-based medicine.

Table 1.
Identified Needs and Challenges at Medical Universities in Poland with Innovative Solutions Proposals

Identified Needs and Challenges at Medical Universities in Poland	Innovative Solutions Proposals
1. Pioneering the development of research competencies in the contemporary information society.	• Crafting graphics that capture attention, inspire creative thinking, and encourage pattern recognition, thus extending memory traces through associations.
2. Enhancing the efficiency of interdisciplinary teamwork within the medical domain.	• Providing data manipulation capabilities, allowing users to experiment with various perspectives, scenarios, and the formulation of new research questions.
3. Equipping future medical researchers for evidence-based discussions (EBM) skills.	• Proficiency in real-time visualization creation using the RShiny package, without the need for in-depth programming understanding.
4. Initiating a discourse on the significance of introducing programming fundamentals into the curriculum of Polish medical university students.	• Teaching fundamental programming skills in the R environment within the education of medical university students.
5. Broadening the educational spectrum by offering courses in data analysis within an open-source environment, catering to the needs of exceptionally talented students.	• Developing a missing implementation concept in the discussed educational area for medical university students by a qualified team with statistical and IT competencies, including programming skills.
6. Fostering interdisciplinary collaboration by combining expertise across scientific and medical disciplines among all medical university students.	• Utilizing experience in collaborative work with medical projects and extensive teaching experience with students in medical disciplines.
7. Cultivating high competencies and interdisciplinary skills among teaching and research staff.	• Implementing classes in various formats, depending on the number of instructional hours and the proficiency level of the participant group.

Source: Own work.

References

ABM, Creation and development of Regional Digital Medicine Centers–2023–Medical Research Agency, konkurs ABM. https://www.abm.gov.pl/pl/konkursy/archiwalne-nabory-1/2023/1918, Tworzenie-i-rozwoj-Regionalnych-Centrow-Medycyny-Cyfrowej.html.

Biecek, P. (2013). *Data analysis with R program.* https://ksiegarnia.pwn.pl/Analiza-dani-z-programem-R,68467550,p.html.

Biecek, P. (2017). *Guide to the R package – Publishing House of the Warsaw University of Technology*. http://www.wydawnictwopw.pl/index.php?s=karta&id=3217.

Central Statistical Office / Thematic areas / Health. (n.d.). Retrieved 7 January 2024, from https://stat.gov.pl/obszary-tematyczne/zdrowie/.

CRAN:Manuals. (n.d.). Retrieved December 28, 2023, from https://cran.r-project.org/manuals.html

Daniel Adrian | IntroStat Shiny Apps. (n.d.). Retrieved 28 December 2023, from https://facweb.gvsu.edu/adriandl/215apps.html.

Farrel, A., Li, P., Veenbergen, S., Patel, K., Maris, J. M., & Leonard, W. J. (2023). ROGUE: an R Shiny app for RNA sequencing analysis and biomarker discovery. *BMC Bioinformatics, 24* (1), 303. https://doi.org/10.1186/s12859-023-05420-y.

Fawcett, L. (2018). Using Interactive Shiny Applications to Facilitate Research-Informed Learning and Teaching. *Journal of Statistics Education, 26* (1), 2–16. https://doi.org/10.1080/10691898.2018.1436999.

Fernandes, Kelwin, Cardoso, Jaime, & Fernandes, J. (2017). *Cervical cancer (Risk Factors).* Published: UCI Machine Learning Repository. https://archive.ics.uci.edu/dataset/383/cervical+cancer+risk+factors.

Garfield, J. & Ben-Zvi, D. (2007). How Students Learn Statistics Revisited: A Current Review of Research on Teaching and Learning Statistics. *International Statistical Review, 75* (3), 372–396. https://doi.org/10.1111/j.1751-5823.2007.00029.x.

Gov.pl: How to get access to SRP - Ministry of Digitization–Gov.pl portal. (n.d.). Retrieved 7 January 2024, from https://www.gov.pl/web/cyfryzacji/jak-uzyskac-dostep-do-srp.

Grześkowiak, M., Chudzicka-Strugała, I., Zwoździak, B., Swora-Cwynar, E., Nijakowski, K., Jokiel, M., & Roszak, M. (2020). E-learning During the Coronavirus Pandemic – Creating Educational Resources for Teaching Medical Students. *Studies in Logic, Grammar and Rhetoric, 64* (1), 77–97. https://doi.org/doi:10.2478/slgr-2020-0041.

Guzik, P. & Więckowska, B. (2023). Data distribution analysis – a preliminary approach to quantitative data in biomedical research. *Journal of Medical Science,* 92(2):e869. https://doi.org/10.20883/medical.e869.

Hegarty, M. (2004). Dynamic visualizations and learning: Getting to the difficult questions. *Learn Instr*, *14.* https://doi.org/10.1016/j.learninstruc.2004.06.007.

Discover! Disclose! Explain! A collection of essays on the art of presenting data – PDF, (2014). https://www.wuw.pl/product-pol-6322-Odkryc-Ujawniac-Objasniac-Zbior-esejow-o-sztuce-prezentowania-anych-PDF.html.

Jamie, D. M. (2002). Using Computer Simulation Methods to Teach Statistics: A Review of the Literature. *Journal of Statistics Education, 10* (1), null-null. https://doi.org/10.1080/10691898.2002.11910548.

Jiang, Z., Cao, W., Chu, H., Bazerbachi, F., & Siegel, L. (2023). RIMeta: An R shiny tool for estimating the reference interval from a meta-analysis. *Research Synthesis Methods, 14* (3), 468–478. https://doi.org/10.1002/jrsm.1626.

Kumar, A., Kumar, P., Palvia, S. C. J., & Verma, S. (2017). Online education worldwide: Current status and emerging trends. *Journal of Information Technology Case and Application Research, 19.* https://doi.org/10.1080/15228053.2017.1294867.

Laureaci_szp_2023_2.pdf. (n.d.). Retrieved 22 December 2023, from https://www.umb.edu.pl/photo/pliki/aktualnosci/2023/laureaci_szp_2023_2.pdf.

Liebig, P., Pröhl, H., Sudhaus-Jörn, N., Hankel, J., Visscher, C., & Jung, K. (2022). Interactive, Browser-Based Graphics to Visualize Complex Data in Education of Biomedical Sciences for Veterinary Students. *Medical Science Educator, 32* (6), 1323–1335. https://doi.org/10.1007/s40670-022-01613-x.

Liu, L., Xie, Y., Yang, H., Lin, A., Dong, M., Wang, H., Zhang, C., Liu, Z., Cheng, Q., Zhang, J., Yuan, S., & Luo, P. (2023). HPVTIMER: A shiny web application for tumor immune estimation in human papillomavirus-associated cancers. *IMeta, 2* (3), e130. https://doi.org/10.1002/imt2.130.

Milewski, R. & Roszak, M. (2023). The Issue of Selection of Appropriate Methodology and Methods of Graphical Representation of Data in Biomedical Research. *Studies in Logic, Grammar and Rhetoric*, 68 (1), 123–131. https://doi.org/10.2478/slgr-2023-0006.

Murray, S. (2014). Interactive data visualization (1–1 online resource). Helion; WorldCat. http://www.myilibrary.com?id=610946.

Pfannkuch, M. (2011). The Role of Context in Developing Informal Statistical Inferential Reasoning: A Classroom Study. *Mathematical Thinking and Learning, 13* (1–2), 27–46. https://doi.org/10.1080/10986065.2011.538302.

The Polish Mbaza AI application is one of the best in the world according to UNESCO. (2023, December 28). *MamStartup.* https://mamstartup.pl/polska-aplikacja-mbaza-ai-chodzi-z-najlepszych-na-swiecie-wg-unesco/.

Roszak, M., Leszczyński, P., Starowicz, K., Wiktorzak, P., Torres, K., Świniarski, P., & Kononowicz, A. (2020). *E-learning in medical education.* Warsaw: Warsaw University of Technology. https://ruj.uj.edu.pl/xmlui/handle/item/266266.

Jagiellonian University Medical College syllabus. Retrieved 28 December 2023, from https://sylabus.cm-uj.krakow.pl/pl/6/1/7/1/1?masterElement.

Wang, C., Govindarajan, H., Katsonis, P., & Lichtarge, O. (2023). ShinyBioHEAT: an interactive shiny app to identify phenotype driver genes in E.coli and B.subtilis. *Bioinformatics, 39* (8), btad467. https://doi.org/10.1093/bioinformatics/btad467.

Wang, S. L., Zhang, A. Y., Messer, S., Wiesner, A., & Pearl, D. K. (2021). Student-Developed Shiny Applications for Teaching Statistics. *Journal of Statistics and Data Science Education, 29* (3), 218–227. https://doi.org/10.1080/26939169.2021.1995545.

Wickham, H. (2021). *Welcome | Mastering Shiny* (2021st ed.). https://mastering-shiny.org/.

Wickham, H. (2023). *R for Data Science, 2nd Edition [Book].* https://www.oreilly.com/library/view/r-for-data/9781492097396/.

Justyna Marcinkowska, Magdalena Roszak

Interaktywne raporty RShiny – czyli niezależność i samodzielność wizualizacji danych medycznych

S t r e s z c z e n i e

W artykule omówiono wykorzystanie języka R, a szczególnie narzędzia RShiny, do tworzenia interaktywnych raportów w dziedzinie analizy danych medycznych. Autorzy podkreślają potrzebę interdyscyplinarnego podejścia do nauczania statystyki wśród studentów medycyny. Przedstawiono alternatywę dla tradycyjnych statycznych raportów, proponując tworzenie interaktywnych aplikacji webowych, które umożliwiają eksplorację, analizę i wizualizację danych medycznych zmieniającą się w czasie rzeczywistym. Wykorzystanie języka R, jako narzędzia open source, pozwala na rozwijanie kompetencji studentów kierunków medycznych w dziedzinie analizy danych oraz dostosowywanie narzędzi badawczych do indywidualnych potrzeb. Przedstawiono także projekt scenariusza zajęć wykorzystujących przykładowe dane medyczne dotyczące raka szyjki macicy wraz z propozycją konkretnych analiz tych danych i ich wizualizacji przy pomocy interaktywnych raportów RShiny.

Artykuł kończy się dyskusją na temat roli nauki języka R w edukacji studentów polskich uczelni medycznych i potrzeby rozbudowy oferty edukacyjnej dla nich o kursy z zakresu analizy danych w środowisku open source.

S ł o w a k l u c z o w e: RShiny, R, wizualizacja danych, interaktywne raporty, statystyka, nauczanie studentów medycyny

Justyna Marcinkowska, Magdalena Roszak

Informes interactivos de RShina – independencia y autosuficiencia de la visualización de datos médicos

R e s u m e n

El artículo analiza el uso del lenguaje R, especialmente la herramienta RShiny, para crear informes interactivos en el campo del análisis de datos médicos. Los autores enfatizan la necesidad de un enfoque interdisciplinario para la enseñanza de la estadística entre los estudiantes de medicina. Se presentó una alternativa a los informes estáticos tradicionales, proponiendo la creación de aplicaciones web interactivas que permitan la exploración, análisis y visualización de datos médicos cambiantes en tiempo real. El uso del lenguaje R como herramienta de código abierto permite el desarrollo de competencias de los estudiantes de medicina en el campo del análisis de datos y la adaptación de herramientas de investigación a las necesidades individuales. También se presentó un borrador de plan de lección utilizando datos médicos de muestra sobre el cáncer de cuello uterino, junto con una propuesta para análisis específicos de estos datos y su visualización mediante informes interactivos RShiny. El artículo finaliza con una discusión sobre el papel del aprendizaje del lenguaje R en la educación de los estudiantes de las universidades médicas polacas y la necesidad de ampliar la oferta educativa para ellos con cursos de análisis de datos en un entorno de código abierto.

P a l a b r a s c l a v e: RShiny, R, visualización de datos, informes interactivos, estadística, enseñanza a estudiantes de medicina

Юстина Марцинковска, Магдалена Рошак

Интерактивные отчеты РШина – независимость и самодостаточность визуализации медицинских данных

А н н о т а ц и я

В статье рассматривается использование языка R, в частности инструмента RShiny, для создания интерактивных отчетов в области анализа медицинских данных. Авторы подчеркивают необходимость междисциплинарного подхода к преподаванию статистики студентам-медикам. Была представлена альтернатива традиционным статическим отчетам, предлагающая создание интерактивных веб-приложений, позволяющих исследовать, анализировать и визуализировать изменяющиеся медицинские данные в режиме реального времени. Использование языка R в качестве инструмента с открытым исходным кодом позволяет развивать компетенции студентов-медиков в области анализа данных и адаптации исследовательских инструментов к индивидуальным потребностям. Также был представлен проект плана урока с использованием выборочных медицинских данных о раке шейки матки, а также предложение по конкретному анализу этих данных и их визуализации с использованием интерактивных отчетов RShiny. Статья завершается обсуждением роли изучения языка R в образовании студентов польских медицинских вузов и необходимости расширения образовательного предложения для них курсами по анализу данных в среде с открытым исходным кодом.

К л ю ч е в ы е с л о в а: RShiny, R, визуализация данных, интерактивные отчеты, статистика, обучение студентов-медиков