

Thermostats: An Open Source Shiny App for Your Open Data Repository

Dasapta Erwin Irawan[#], Muhammad Aswan Syahputra^{*}, Prana Ugi^{**}, Deny Juanda Puradimaja[#]

[#] Applied Geology Research Group, Faculty of Earth Sciences and Technology, Institut Teknologi Bandung, Bandung, Indonesia

^{*} Sensolution ID, Kab. Bandung Barat, Indonesia

^{**} Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sumatera Utara, Medan, Indonesia

E-mail: erwin@fitb.itb.ac.id

Abstract— Hydrochemical analysis has emerged as a powerful methodology in geothermal system profiling. Indonesia is the capital of geothermal energy with its more than 100 active volcanoes. Therefore we need to have an analytical, data-driven, and user-focused online application of geothermal water quality. Proudly we introduce Thermostats (<https://aswansyahputra.shinyapps.io/thermostats/>). We collected water quality from 416 geothermal sites across Indonesia. Three main objectives are to provide an online open-free to use data repository, to visualize the dataset to suit user's needs, and to help users understand the geothermal system of each particular site. At the end, we hope they like this system and donate their own dataset to make it better for future users. We designed this online app using Shiny, because it's open source, lightweight and portable. It's very intuitive to load our descriptive, bivariate and multivariate statistics. We selected Principal Component Analysis and Cluster Analysis as two strong statistics for water sample classification. Users could add their own dataset by making a pull request on Github (<https://github.com/dasaptaerwin/thermostats>) or sending it to us by email to make it visible in the application and included in the visualization. We make this application portable, so it can be installed on a local computer or a server, to enable an easy and fluid way of data sharing between collaborators.

Keywords— Open data repository, Reproducible research, Shiny App, Geothermal dataset.

INTRODUCTION

Indonesia is associated with vast geothermal potential [1]. However, that situation hasn't been thoroughly studied. A short bibliometric study using Scopus database with 218 papers concluded that there were not many papers software. We could see at least five clusters presented in colors: purple (hydrochemistry), red (geology and geophysics), blue (well engineering), and yellow (geothermal locations). All of those colors are connected to green cluster (geothermal global terms) (Fig. 1).

Currently, we don't have a proper public geothermal data system. Therefore this paper may send out some insights to the generation of public data repository. However recent years are bringing a higher number of open source programming languages, like R, Python, Julia, and also open data repository such as: Github, Gitlab, OSF, and Zenodo, which allow a better understanding of the data using descriptive and spatial statistics, and also data archiving.

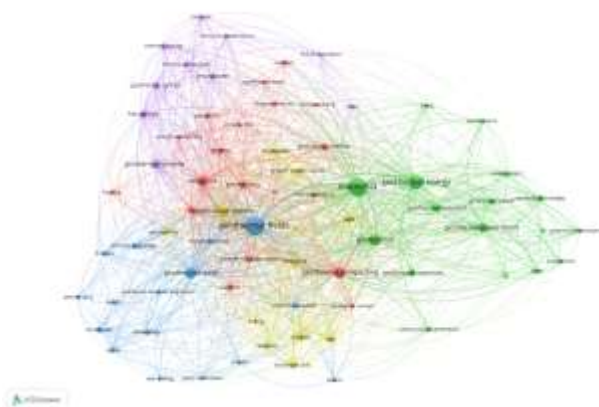


Fig. 1 The keywords connection from 218 papers taken from Scopus

REVIEW

A. Open data as a way to extend data lifetime

Data sharing could in some way extend data lifetime. Previously, placed as part of a research, data had been embedded in reports. They were printed in the same way the reports were presented, usually in PDF format and in a non machine-readable form. Researchers had long been brain-washed to share only the resume of the data, but certainly not the raw data. The classic reason was, they don't want their ideas to be scooped or their data to be stolen, which was unreasonable reason, especially in the 21st century and also if the research was financially paid by public money (eg from Ministry of Research, Technology, and Higher Education/Kemenristekdikti).

Therefore we need to set up an example way on how to share data, without losing the claim at the same time. Our proposed answer is by using the internet and by generating an open data repository

B. Shiny app to support open data initiative

The Shiny App is a web-based application written entirely in the open-source R programming language[2]. It is a cross-platform infrastructure that can be launched locally or installed in a server for public access. The basic setup of Shiny app is not complicated, involving the use of mid range computing power[3][4].

We used R as our code-based programming language to ensure the reproducibility. Reproducibility has been the main issue internationally in the last three years[5][6][7][8][9]. More blended technologies are now also available using other code-based language such as Python[10] and Julia[11].



Fig. 2 The basic setup of data driven R Shiny app[12]

DATA DECLARATION

For the pilot version of **Thermostats**, we used 416 thermal water quality data from four provinces: West Java (Jabar), Central Java (Jateng), East Java (Jatim), and East Nusa Tenggara (NTT). All data are all public data taken from several papers and reports.

METHOD

Thermostats was built using the R language as a backend and Shiny framework to create front-end interactive web applications. Data and application source code are stored in local repositories and online repositories on GitHub[13]. Whereas the application is hosted on the Shinyapps.io server[14], so that it can be accessed by users using a browser (preferably Google Chrome or Firefox). Figure 3 presents the design and architecture of **Thermostats**.

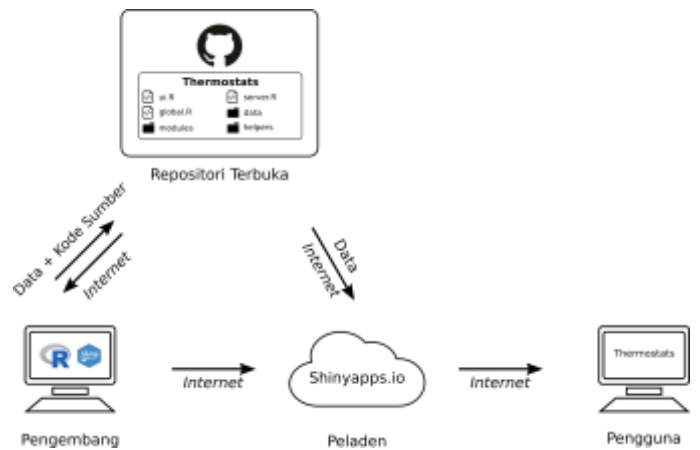


Fig. 3 The setup of Thermostat Shiny app

The **Thermostats** is composed of two main components, namely user interface component (UIC) and server component (SC). Inside UIC we put layout, content and interface sub components, whereas in SC we provide logic, calculations, and analysis procedures. The application's basic source code is as follows.

```
library(shiny)
ui <- navbarPage()
server <- function(input, output) {}
shinyApp(ui = ui, server = server)
```

We developed **Thermostats** in a modular manner, so both UIC and SC could be divided into various analytical modules, and it could be developed with more analytical tools in the future. Each analytical module was stored in the modules directory as shown in Figure 4. Due to using a modular system, the basic source code of the application used is as follows.

```
library(shiny)
ui <- tagList(
  modul1UI("m1"),
  modul2UI("m2"),
  modul2UI("m3"),
  modulNUI("mn")
)
server <- function(input, output) {
  callModule(modul1, "m1"),
  callModule(modul2, "m2"),
  callModule(modul3, "m3"),
  callModule(moduln, "mn")
}
shinyApp(ui = ui, server = server)
```

We provided three basic features, each was stored as individual module:

1. Data exploration and filtering: to showcase the raw data.
2. Plots: scatter plot and correlation plot,
3. Data analysis:
 - a. Descriptive
 - b. Correlation
 - c. Bivariate regression
 - d. Multivariate

```

Terminal
aswansyahputra@notebook:~/Projects/GitHub/thermostats$ tree
.
├── data
│   ├── data_copy.csv
│   └── dataDescriptor.csv
├── global.R
├── helpers
│   ├── plot_eigen.R
│   ├── plot_parameter.R
│   └── plot_profile.R
├── LICENSE
├── modules
│   ├── plotCorrelation.R
│   ├── plotScatterplot.R
│   ├── showData.R
│   ├── statCorrelation.R
│   ├── statDescriptive.R
│   ├── statMultivariate.R
│   └── statRegression.R
├── README.md
├── rconnect
│   ├── shinyapps.io
│   └── aswansyahputra
│       └── thermostats.dcf
├── server.R
├── thermostats.Rproj
├── ui.R
├── www
│   └── custom-summarytools.css
└── 7 directories, 26 files
aswansyahputra@notebook:~/Projects/GitHub/thermostats$

```

Fig. 4 The setup of Thermostat Shiny app

OVERVIEW OF THE APPLICATION

The main feature provided by the **Thermostats** is an interactive application that allows to carry out the complete descriptive, **bivariate**, and multivariate analysis based on the available. Here we offer data filtering feature to see the distribution of data using sorting and filtering button (Figure 5). There are 65 columns in the dataset as listed in the data descriptor[15].

The first feature is scatter plot. We provide three controls: province selection, **parameter X**, and parameter Y (Figure 6). The second feature is correlation plot for easy determination of the closeness of some parameters. It spits a simple correlation plot to determine the strength of correlation (more red color means stronger correlation) (Figure 7).



Fig. 5 The setup of Thermostat Shiny app



Fig. 6 Sample of scatter plot of EC and Ca for all province



Fig. 7 Sample of correlation plot between variables and all province

The next feature is Descriptive Statistics to make a statistical summary of the dataset. Again we could filter the data based on locations (province and field site) (Figure 8).



Fig. 8 Sample of descriptive statistic analysis from several major elements

We also provide regression and correlation feature for more detail regression analysis among parameters. The last feature which we promote as the novel feature in **Thermostats** is Multivariate analysis. Here we apply Principal Component Analysis (PCA) to reduce the number of parameters by converting them into groups of parameters[16]. Using this feature, users could customize their analysis based on locations and/or parameters (Figure 9 and Figure 10). The plots and dimensions are also changeable according to user's needs.

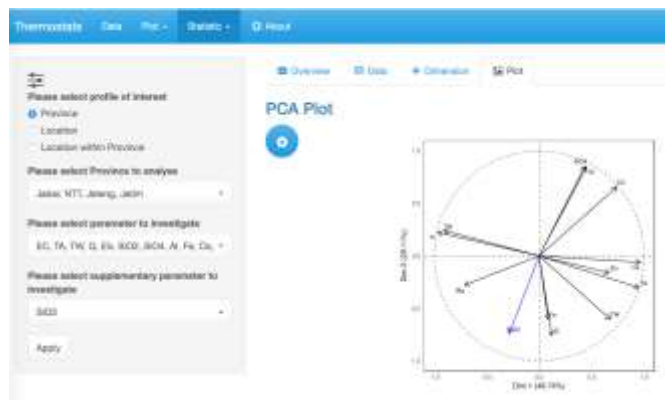


Fig. 9 Sample of PCA analysis to show grouping of parameters



Fig. 10 Sample of PCA analysis to show grouping of locations

REMARKS

This Shiny app is a prototype of both online open data repository and statistical analysis. Here we point out the importance of providing raw data accessible for the public to support reproducibility of the analysis, and also to provide a seamless user experience to make their own statistical analysis to the provided dataset. To support the raising citizen science movement, the user could also contribute their data to this project by making a pull request on the Github platform, or by emailing their dataset to the first author by stating their scope of data, the rationale of the acquisition, how they acquired the data and what open license they would use.

ACKNOWLEDGMENT

We would like to thank Institut Teknologi Bandung (ITB) Research Group Grant 2017 and ITB P3MI Research Grant 2018. We also appreciate suggestions from the users that we have received since we put the application online in early 2018.

REFERENCES

- [1] "Indonesia: Vast geothermal potential, modest but ... - ScienceDirect." https://www.sciencedirect.com/science/article/pii/B97800810033740_00218. Accessed 13 Aug. 2019.
- [2] "Shiny." <https://shiny.rstudio.com/>. Accessed 16 Aug. 2019.
- [3] "The basic parts of a Shiny app - R Shiny - RStudio." 28 Jun. 2017, <https://shiny.rstudio.com/articles/basics.html>. Accessed 16 Aug. 2019.
- [4] "Database basics - dplyr and DBI - R Shiny - RStudio." 14 Jun. 2017, <https://shiny.rstudio.com/articles/overview.html>. Accessed 16 Aug. 2019.
- [5] "Challenges in irreproducible research - Nature." 18 Oct. 2018, <https://www.nature.com/collections/prbfkwmwvz>. Accessed 16 Aug. 2019.
- [6] "1,500 scientists lift the lid on reproducibility : Nature News & Comment." 25 May. 2016, <https://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970>. Accessed 16 Aug. 2019.
- [7] "Opinion: Is science really facing a reproducibility crisis, and do ... - PNAS." 6 Jul. 2018, <https://www.pnas.org/content/115/11/2628>. Accessed 16 Aug. 2019.
- [8] "A proposal of generic university-level Research Data Management Plan." 6 Jul. 2018, <https://riojournal.com/article/28163/>. Accessed 16 Aug. 2019.
- [9] "How to Extend your Data Lifetime: Research Data ... - Atlantis Press." <https://www.atlantis-press.com/proceedings/iclick-18/125913290>. Accessed 16 Aug. 2019.
- [10] "How to Build a Reporting Dashboard using Dash and Plotly." 11 Mar. 2019, <https://towardsdatascience.com/how-to-build-a-complex-reporting-dashboard-using-dash-and-plotly-4f4257c18a7f>. Accessed 16 Aug. 2019.
- [11] "Shiny for Julia [work in progress, early request for ... - Julia Discourse." <https://discourse.julialang.org/t/shiny-for-julia-work-in-progress-early-request-for-feedback/14038>. Accessed 16 Aug. 2019.
- [12] "Creating a Business Intelligence Dashboard with R and ASP.NET MVC." 8 Jul. 2013, <https://www.red-gate.com/simple-talk/dotnet/asp-net/creating-a-business-intelligence-dashboard-with-r-and-asp-net-mvc-part-1/>. Accessed 16 Aug. 2019.
- [13] "dasaptaerwin/thermostats: It contains our project the ... - GitHub." <https://github.com/dasaptaerwin/thermostats>. Accessed 16 Aug. 2019.
- [14] "Thermostats | Muhammad Aswan Syahputra." <https://aswansyahputra.com/project/thermostats/>. Accessed 16 Aug. 2019.
- [15] "thermostats/global.R at master · dasaptaerwin/thermostats · GitHub." <https://github.com/dasaptaerwin/thermostats/blob/master/global.R>. Accessed 16 Aug. 2019.
- [16] "PCA - Principal Component Analysis Essentials - Articles - STHDA." 23 Sep. 2017, <http://www.sthda.com/english/articles/31-principal-component-methods-in-r-practical-guide/112-pca-principal-component-analysis-essentials/>. Accessed 16 Aug. 2019.