ZOO-Journal **5**: 23-26, 2019 **Innovation/Methods** ***ISSN: 2631-2204***

DATA ANALYSIS TOOLS: ESSENTIAL FOR SCIENTIFIC COMMUNITY

Chitra B. Baniya

Central Department of Botany, Tribhuvan university, Kirtipur

E-mail: cbbaniya@gmail.com

ABSTRACT

Data analysis tools and software have been increasing each day in the academic field. They were classified into two categories: one category of statistical tool that needs money to buy its licence to use and install into owns computer. Such software are called closed software. Another category of tool or software is called open source tool or open software. This later group comes freely and one can get licence freely. Academicians are free to choose one. Definitely 'R' and its associated tools which we can download freely and use freely has no better alternatives. 'R' users have been increasing rampantly these days in this world.

Key words: Close source, user friendly, keyboard, licence, open source.

BACKGROUND

Research is impossible without data. Data is prerequisite for all research at all four scales of measurement, nominal, ordinal, interval and ratio. Seemingly, it can be better said that nature of data differs according to scale of measurement in the research. A good quality of research begets after fully relay on a good research question, proper research design, sound data, sound analysis and finally on suitable interpretations. All sophisticated big laboratories in the world to the small microcosm study dedicated whole heartily into the generation of noble data. Data are generating at every fraction of second. People have created several new ideas, thoughts and knowledge after utilization of such data. Wiser thinking comes only after proper utilization of brain on the valid, sound and noble data. Generally, data are two types: qualitative and quantitative. Both the type of data must support a valid science.

Much if not all academics have been dreaming to do noble research in their life-time. A noble research is that which can be published in the world's reputed journal. Quality of journal based on frequency of their publication, regularity of publication, number of people cited that particular journal etc. It is sad but true to say that there are no such journals yet in Nepal so far till date. This does not mean to say here that no Nepalese researchers have published their research in such journals. Yes, there are several Nepalese researchers who are publishing their research work in such journals after corresponding from Nepal. These journal need not to be within their own home country to publish. One can correspond from other country in order to publish their results. Not all researches can be published immediately after corresponding to the journal. There are definite procedure from the beginning of quick reading from the chief editor to thoroughly reviewing by unknown reviewers. Number of reviewers varied from two to many. Thus one needs at least two years to come final print in the journal it data was really publishable.

DOI: https://doi.org/10.3126/zooj.v5i0.34778

Two aims always associated with research data are either to describe phenomena of what we are interested in or to test own desired research data with similar other data. Data analysis is one of major steps before writing research articles. Not only noble data begets quality research paper but also valid and sound analyses. Tones of statistical software popping up in market each day. Scientific communities got crapped or lost in this floods of software about which one to be used and why to be used and how to use. A very fundamental and essential insight regarding scientific dilemma of using data analysis tool intended to present. It is a message driven resulted after my several years feelings and thoughts rather well formatted research article. My main objective of this manuscript is to let our intellectual academic audiences familiarize data analysis tool or software to some extent.

WHAT ARE BASICS TO DEFINE DATA PUBLISHABLE OR NOT?

Mainly randomization, replication and precision in the data after independent experimental design are basics behind the crucial issue whether our data is publishable or not. There are many things afterwards such as nobility of research, abiding theories or testing theories, readability, language, relevancy, analytical tools and techniques, interpretation and further prospectives etc. are some of important guiding parameters behind quality of research and acceptance for publication in the high quality journal. On the top of these, application of non-pirated tools or cracked version software is taken as one of a serious issues. Propagation of sound analysis after utilizing valid statistical tool is an another good manner to regard in the field of academia.

WHAT ARE MAIN TOOLS TO ANALYSE DATA?

Two categories of data analysing tools are available in the software world. Some examples are R, Python, S-Plus, MATLAB, Minitab, STATA, SAS, SPSS, Statistix, Sigmaplot, Qualtrics, Nvivo, JMP and Gpower. This broad classification was based on whether one needs to purchase licence to use or not. Among these tools, only "R" and Python provide licence freely to all individual user. Remaining tools need to buy licence or each one is fully based on proprietary. In other sense former is known as open source tool and later are called closed source tools. Question always comes about the sustainability of the open source tools. Yes, no one got paid and no full-time service holder working for 'R'. This tool runs thoroughly by charity, free donation or free help by institutions. Individual who is contributing for 'R' nothing got paid. There are several thousand free packages uploaded freely into its authentic homepage: https://cran.r-project.org. All these packages can be downloaded freely into our personal computer and get our work done freely afterwards. What does this project get ultimately from us? Only "citation" in the form of "R Core Team", nothing related with contributor. Python analyzes business data in addition to the biological data.

While browsing this site on December 14, 2018 a total of 13,526 free packages uploaded and get accessed freely. That number has been changed and increased to 13535 after re-accessed on December 16, 2018. Each of all packages hosted freely by servers of several institutions, universities, organizations in the world. Each such host is called CRAN MIRROR ('R' Core Team, 2018). Unfortunately, Nepal has not been hosting by any organization so far yet. One very important point to know that one can not upload

their packages freely. There is a core team who judges, peer reviews each package at first. After their peer review if they find publishable then only one can publish or upload in this website. There are of course high rejection rate meaning that several people have been producing their own tool and trying to upload in this website.

In this software world there are several tools which need to pay or buy its licence before to use. Such group of software is known as closed tools. All closed software run by proprietary. Proprietary organization employed several software engineers to develop it. Thus each closed software made more user friendly, ready to extract result and use it. Technical resources found utilize more on development of keyboard button. Thus utilizing closed software gives results at finger tips. Each closed software has more or less similar philosophy of business orientated such as six months free subscription with limited access on functions. If we want to continue further then definitely asks money to renew licence. How many departments or colleges or universities bought collective licence and supplied to students, faculties or staffs in our country? I am sorry to say that there is non. Once we got tamed or software addicted close software companies forcing us to buy. Pity to say here that these companies indirectly forcing genuine academics to steal their software by cracking and installing. This is an academic insult. Who will come to check? No one. But there are places where authenticity of software has to be justified and proved. Certainly, these places are answering comments by review team of International Science Index (ISI) ranked journals, publicly delivering lectures or training etc. If one can not verify originality of their statistical tool then no such research paper can be accepted. Thus there is always fear of legal issues after using cracked version software. Academic institutions, campuses and academics targeted more about this issue than others. What will be a feeling of a professor if delivers training on the pirated software?

In contrast, whichever version of 'R' we were using, all came with free licence. Pity to say that not much easy keys or toolbox have been supplied to 'R' so far yet. People blamed this 'R' as not user friendly. Yes, I fully agreed with this statement. 'R' let's its users freely teach through search.help options. Less chances to misuse open source software as like "Garbage in and garbage out". At least we will get time to think about what will happen if 'I do like not like that' or vice versa. If one utilizes figures that were build in 'R', then there will be no meaning to ask about licence.

WHICH ONE IS BETTER AND WHY?

Of course 'R' has no alternative so far till now among academic institutions. It is a free software with the licence. More and more user communities of 'R' are increasing. Rapidity of its popularity has been making many functions more user friendly these days. 'R' has been functioning for more than 25 years in the computing field after the first inception in 1992. Of course, easiness must gain after investing money. Closed software are no more far away from that philosophy. In term of contribution towards companies that must be essential and each one should gain that satisfaction after investment. Why not be ready to bear a little bit difficulties after using publicly free software? One can proudly say that R has no other better alternative. Scientists have now guessed or estimated that there will be only two statistical tools after few years at least among all universities in the world These will be R and Python. An academic is not an advertiser of private software vendor rather than sober and wiser inventor, owner and user of software.

CONCLUSION

We academia must rise from our independent and noble thoughts. All our creations are becoming public. We must think and create less debatable society after utilizing valid tools and techniques rather than from illegal ways.

ACKNOWLEDGEMENTS

I am highly grateful to Associate Professor Kishor Kumar Pokharel, HOD, Department of Zoology, Prithvi Narayan (P.N.) Campus, Pokhara and his entire team for requesting me an article related to 'R' helpful for analysis of biological data since long. Also my sincere thanks to my dear friends, Mr. Kalyan Panta and Mr. Om Dwa, Head of the Department of Botany, P. N. Campus whom I am discussing about creating and disseminating free and open academic society.

REFERENCES

'R' Core Team (2018). *'R': A language and environment for statistical computing*. 'R' Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

https://cran.r-project.org/ accessed date: December 14 & 16, 2018.

(Received July 13, 2019; revised accepted September 5, 2019)