# INTRODUCTION

## The Researcher’s Standpoint

This paper tries to investigate the effect of using R (R Core Team, 2019) in RStudio(RStudio Team, 2016) as statistical computing tool to the performance of college students in relation to their second course in statistics. My interest on this topic started when I did my master’s thesis on rainfall prediction on agriculture using mathematical modeling. I have been looking for a software that is economical (if not free) to be used in my data analysis. That time, my wife had a collaboration research project with the UP School of Statistics headed by the former Dean Dr. Erniel Barrios. I had the opportunity to meet the members of the team who are seasoned and high-caliber statisticians. I did not waste time. I took the chance to inquire about my master’s thesis, and so Dr. Barrios introduced me to R, which UP School of Statistics uses for free – it is open source and anybody can access it without even buying a license. Dr Barrios was even so generous enough to give me e-materials for this purpose.

From there on, I took time to learn R using RStudio. In Nueva Vizcaya State University (NVSU) where I am teaching, most of the faculty and students use SPSS – a spreadsheet type statistical computing tool – in their computer laboratories. The university bought the SPSS software for more than a million, however, using the SPSS software is specially for the BS Mathematics, MST- Mathematics and PhD in Science Education-Mathematics from the College of Arts and Sciences.

So this is where the dilemma comes in. How about faculty and students from the Teacher Education, where I belong, who would like to use the said software since we are offering major in Mathematics? We might use it, perhaps, but it will take time and effort asking for permission.

On this context, I have to be resourceful and innovative. Exploring RStudio in statistics, I found out that RStudio is a very promising tool to do statistical computing since it is free and very powerful in terms of statistical computing. Aside from the programming nature of using RStudio, I don’t see any other thing that would hinder someone to learn and use RStudio to do statistical computing. If you look at it differently, its programming nature can even be an advantage since it opens possibilities to create functions for any statistical task. RStudio promotes reproducible reports in different formats (Xie, Allaire, & Grolemund, 2018). The creation of the ggplot2 package by Wickham (2016) gave way to many possibilities of elegant graphics that can be used for data analysis and reports. RStudio can be a very handy tool for mathematics teachers to create handouts and various reports (Xie et al., 2018). Certain packages can also help you with machine learning, deep learning, AI and data science. Are the students ready to use this free and very powerful software? It would be really very nice if we have a computer laboratory in the College of Teacher Education to learn and use RStudio to do statistical computing tasks and learn and explore applications of statistics.

The next section will highlight the necessity of this study in a bigger context.

## Rationale

The practice of statistics is computational. Statistical computing software are mainly used to do statistical computing. With the emergence of data science as a field due to big data, machine learning and powerful computers that are no longer expensive all brought about by the 4th industrial revolution, it is very fitting to consider how statistics education should adapt to these changes and be relevant in the practice of statistics. In this regard, in 2005, Franklin et al. (2007) put forth the Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report recommending a framework for statistics education both in the k to 12 and college level. In 2016, the GAISE College Report ASA Revision Committee (2016) revisited the effectiveness of the framework and still found it effective. The framework is now the standard in statistics education in the United States and in many countries that adopted it (Zeiffler, Garfield, & Fry, 2018).

In the GAISE report, one noteworthy recommendation is “the use of technology to explore concepts and analyze data”. Studies have shown that the use of technology can really improve statistics education; equipping learners with relevant data skills and effective powerful tools in this era where data is very much abundant (Chance, Ben-Zvi, Garfield, & Medina, 2007; Chance & Rossman, 2006; Çetinkaya-Rundel & Rundel, 2017; Doi, Potter, Wong, Alcaraz, & Chi, 2016; Harraway, 2012; Stander & Dalla Valle, 2017). However, when it comes to software used in doing statistical computing and teaching statistics, there is no single statistical computing tool that fits all statistical tasks (McNamara, 2018). Nonetheless, introductory statistics students should be taught a common statistical computing software such as SAS, SPSS, or R, enthusing them to continuously learn statistics technology since statistical tools are diverse and eventually evolve through time (Gould et al., 2018). Moreover, Gould (2010), N. J. Horton et al. (2015), N. J. Horton (2015) and Hardin et al. (2015) pointed out the importance of developing among students – with the use of statistical computing software – data management skills in introductory and second courses in statistics.

In our country, leading universities are addressing this issue and have already integrated the use of technology in their curriculum for statistics education. The University of the Philippines for example uses a number of software in its introductory statistics courses and statistics courses (eg. R with RStudio, Python, SAS, SPSS, Stata, MS Excel, QGIS, ArcGIS, Gephi, yEd Graph Editor, and more). Most of the software used are opensource.

In the province, some universities have acquired SPSS to teach statistics courses. On the other hand, some still uses calculators to do and teach statistical computing. Unfortunately, both setup do not lessen the gap between statistics education and statistical practice. In reality, most institutions to which graduates from these universities get employed cannot afford SPSS. The use of SPSS requires a paid license. One can choose MS Excel as an alternative, however its functions are limited (Biehler, Ben-Zvi, Bakker, & Makar, 2013).

In this context, in an attempt to improve statistics education in the College of Teacher Education (CTE) at NVSU and promote the use of RStudio in statistics and mathematics classes in the CTE, the researcher introduced the use of R using RStudio to BSEd Mathematics students who are taking their second course in statistics (Inferential Statistics) and looked at their performance in the course.

## Statement of the Problem

To adapt to the changes in technology brought about by the 4th industrial revolution and to be abreast with the global community when it comes to statistics education, RStudio as statistical computing tool in learning statistics was introduced to BSEd mathematics students of the CTE. The researcher then sought to answer the following questions after gathering profile variables and employing a one group pretest-posttest poor experimental design:

1. What are the profile variables of the respondents?
2. What are the calculator test scores and RStudio test scores of the respondents?
3. What are the calculator test scores and RStudio test scores of the respondents when grouped according to the profile variables?
4. Is there significant difference in the calculator test scores and the RStudio test scores of the respondents when grouped according to the profile variables?
5. Is there significant relationship between age and calculator test scores and age and RStudio test scores of the respondents?
6. Is there significant difference between the calculator test scores and RStudio test scores of the respondents?

## Statement of the Hypothesis

Is the gender of a respondent a factor when it comes to performance in the calculator test and RStudio test?

H0: There is no significant difference between the mean calculator test score of male and female.

Ha: There is significant difference between the mean calculator test score of male and female.

H0: There is no significant difference between the mean RStudio test score of male and female.

Ha: There is significant difference between the mean RStudio test score of male and female.

Is having a previous programming experience a factor to consider among the respondents when it comes to performance in the calculator test and RStudio test.

H0: There is no significant difference between the mean calculator test score of those who have previous programming experience and none.

Ha: There is significant difference between the mean calculator test score of those who have previous programming experience and none.

H0: There is no significant difference between the mean RStudio test score of those who have previous programming experience and none.

Ha: There is significant difference between the mean RStudio test score of those who have previous programming experience and none.

Is age of the respondents a factor to consider when it comes to performance in the calculator test and RStudio test.

H0: There is no significant relationship between age and the calculator test scores of the respondents.

Ha: There is significant relationship between age and the calculator test scores of the respondents.

H0: There is no significant relationship between age and the RStudio test scores of the respondents.

Ha: There is significant relationship between age and the RStudio test scores of the respondents.

Did the performance of the respondents in the test improved after learning and using RStudio as statistical computing tool?

H0: There is no significant difference between the mean calculator test score and the mean RStudio test score of the respondents.

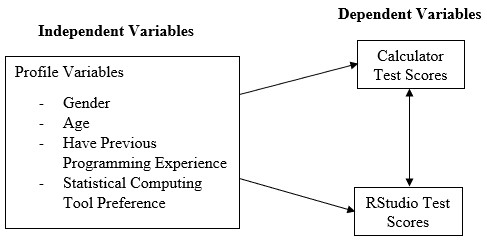
Ha: There is significant difference between the mean calculator test score and the mean RStudio test score of the respondents.

## Significance of the Study

This preliminary study is significant to universities and administrators to invest in free software which is more economical yet even at par with other leading statistical software. This is also important to faculty members who teach statistics and do research since it gives them a glimpse of what is happening in statistics education and in practice. Lastly, this study is significant to students for them to be able to see the practicality and applicability of other statistical software like R using RStudio.

## Research Framework

The independent variables are the profile variables of the respondents. There are four profile variables: gender, age, have previous programming experience and statistical computing tool preference. There are two dependent variables: calculator test scores and RStudio test scores of the respondents. Figure 1 shows the research framework of the the study. The independent variables and dependent variables were described. Significant difference in the mean calculator test scores and mean RStudio test scores when grouped according to gender and have previous programming experience where tested. Significant Correlation between age and the calculator test score and the RStudio test scores were also tested. Finally, significant difference between the mean calculator test score and the mean RStudio test score was tested. This will show if there is improvent in the test of the respondents after learning and using RStudio as a statistical computing tool.



**Figure 1.** Research Framework of the Study

## Scope of the Study

The study focused on independent and dependent variables data gathered from the BSEd Mathematics students who are taking their second course in statistics at CTE, NVSU.

## Definition of Terms

* **calculator test scores** - the scores of the respondents in the test using calculator as their statistical computing tool.
* **RStudio test scores** - the scores of the respondents in the test using RStudio as their statistical computing tool.
* **second course in statistics** - this phrase refer to inferential statistics. Descriptive statistics is the first course in statistics.
* **statistical computing** - the computing of test statistics using computer.
* **statistics education** - refers to the curriculum, instruction and assesment in the teaching and learning of statistics.
* **test** - the test referred to in the study is the same test taken in the calculator test and in the RStudio test.

# REVIEW OF RELATED LITERATURE

This part focuses on the chosen literatures which I think are most relevant in my study. After reviewing significant literatures, there are two themes I have generated: a) the necessity of statistics education and its current landscape and b) R using RStudio in teaching statistics. These will be discussed in the following sections of the paper.

## The Necessity of Statistics Education and its Current Landscape

The book entitled *International Handbook of Research in Statistics Education* edited by Ben-Zvi, Makar, & Garfield (2018) offers a wide array of topics concerning statistics which best fit in this section. The book has three main parts which are interrelated to each other: Part I: Statistics, Statistics Education, and Statistics Education Research; Part II: Major Contributions of Statistics Education Research; and Part III: Contemporary Issues and Emerging Directions.

An article entitled “What Is Statistics Education?” succinctly discusses the nature and necessity of statistics education to all level – from secondary to tertiary up to graduate level. The authors defined statistics education this way:

Statistics education is an interdisciplinary field that is focused on the teaching and learning of statistics. Evolving from the field of mathematics education, which supplied valuable theories of learning, models of conceptual development and change, and methods of qualitative research (e.g., teaching experiments, clinical interviews), statistics education has emerged as an independent area of inquiry and scholarship with its own journals, conferences, organizations, websites, and curriculum standards (Zeiffler et al., 2018).

Based from the definition of statistics education, it is said to be an evolving field from the field of mathematics wherein it is “interdisciplinary” that focused on both “the teaching and learning of statistics” and thereby “emerged as an independent area of inquiry and scholarship”.

Perhaps the next inquiry is, why is statistics education necessary in everyday life and specifically in the curricula as well as how its integration in the curricula will help students abreast with technological pedagogy and skills.

The discussion of the book article “What Is Statistics?” by Wild, Utts, & Horton (2018) directs our attention to the nitty-gritty of statistics and how this field is important in our day to day life given the advent of technological advances that produce rich data. For example, the authors argue that “In today’s data-rich world, all educated people need to understand statistical ideas and conclusions, to enrich both their professional and personal lives”. Thus, with this context of today’s world where data is accessible everywhere from mass media especially social network such as Facebook, it can be said that “the widespread availability of interesting and complex data sets and increasingly easy access to user-friendly visualization and analysis software mean that anyone can play with data to ask and answer interesting questions”. Statistics is vigorously gaining importance and recognition in today’s society.

Other authors also recognize the necessity of statistics:

“Statistics is a central tool in moving science, economics, politics, schools, and universities forward”. Quantitative information is omnipresent in media and in the everyday lives of citizens worldwide. Data are increasingly used to add credibility to advertisements, arguments, or personal and professional advice. Therefore, there is a growing public and policy consensus that being able to provide reliable and persuasive evidence-based arguments and critically evaluate data-based inferences are crucial skills that all citizens of the twenty-first century should have (Ben-Zvi & Makar, 2016).

In addition, “statistical methods are used in almost all knowledge areas and increasingly are used by businesses, governments, health practitioners, other professionals, and individuals to make better decisions” (Wild et al., 2018). This is true wherein based on my observations, most government and private agencies rely on statistical methods through surveys to make policies, laws and even programs that are geared towards communities and nation’s development.

In this case then, academic institutions also integrated statistics in the curricula thinking that “probably no academic subject is more useful to both working professionals and informed citizens on a daily basis than statistics” (Wild et al., 2018). On this note, we can say then the impending necessity of statistics in the educational system. For instance, the book article “International Perspectives on the Teaching and Learning of Statistics” posits:

Being able to provide sound evidence-based arguments and critically evaluate databased claims are important skills that all citizens should have. It is not surprising therefore that the study of statistics worldwide at all educational levels is gaining more attention. The study of statistics provides students with tools, ideas and dispositions to react intelligently to information in the world around them. Reflecting this need to improve students’ ability to think statistically, statistical literacy and reasoning are becoming part of the mainstream school and university curricula in many countries. As a consequence, statistics education is becoming a thriving field of research and curricular development (Ben-Zvi & Makar, 2016).

However, “the rapid development of data science… provides challenges for statistics educators in determining learning goals, and opportunities for statistics education researchers to explore what instructional methods can best achieve those goals” (Wild et al., 2018). Moreover, the said article also points out present challenges of statistics education especially to students:

What are the areas where statistics may need to adapt to be relevant to data science? In addition to pedagogy and content, technology is a key realm. While the Guidelines for Assessment and Instruction in Statistics Education (GAISE) K-12 (2005) and College (2016) reports encouraged the use of technology (which, on a more positive note, is now widespread in most courses), hundreds of thousands of high school students still use calculators rather than computers for their analyses, limiting their ability to move beyond simple calculations or gain any sense of realistic workflows that they might encounter in the real world. But much worse, it also narrowly constricts their vision of what statistics is and can be and neglects the huge potential of the visual sense for gaining insights from data (Wild et al., 2018).

Given these challenges of statistics education in terms of pedagogy and content, one factor to consider is technology: most students use calculator over computer in statistical analyses. Aside from this, statistics can be seen in both perspectives wherein “despite the increasing awareness of the importance of statistical literacy, statistics has been viewed by many students as difficult and unpleasant to learn” (Ben-Zvi & Makar, 2016). On the other hand, “many university instructors find statistics and research methods courses equally frustrating and unrewarding to teach” (ibid.). Not only that, “in schools, mathematics teachers often view statistics as a marginal strand in the mathematics curriculum and therefore minimize or ignore its teaching” (ibid.). Hence, it is no wonder how one article describes the dilemma of learning and teaching statistics wherein “many countries still lack sufficient resources, updated curriculum materials, effective professional development of teachers, and current technologies, infrastructure essential to carry on the reform movement in statistics education” (MacGillivray, Martin, & Phillips, 2014).

With this background on learning and teaching statistics, as well as on issues of statistics in researches, it is hope that this preliminary study will somehow fill in the gaps as pointed out in different literatures. This is where my study would like to investigate.

### A Brief History of Statistics Education

The book entitled *The Teaching and Learning of Statistics: International Perspectives* edited by Ben-Zvi & Makar (2016) and The book entitled *Third International Handbook of Research in Statistics Education* edited by Ben-Zvi et al. (2018) are useful in this section as these books trace the history of statistics education in the international perspectives. Along with the development of statistics are the books’ discussions of challenges posed in the field of statistics. These books are collaborative work where authors came from different parts of the world. Given this background of the book where author-contributors are from different countries, hence tracing the historical background of education statistics is better understood.

#### The Global Perspective on Statistics Education

The book article “What Is Statistics?” by Wild et al. (2018) present the global perspective of statistics education as it traces its beginning across Western Europe who were influenced by the Renaissance Period “by the rise of science based on observation of the natural world” wherein “the statistical analysis of data is usually traced back to the work of John Graunt”. In addition, “another fundamental thread involved in building modern statistics was the foundation of theories of probability, as laid down by Pascal (1623–1662) and later Bernoulli (1654–1705), which were developed to understand games of chance”. The probability analyses were then “later applied to social data by Quetelet (1796–1874), who with notions such as the “average man” was trying to arrive at general laws governing human action, analogous to the laws of physics”. But then, during the French Revolution “when there was a subtle shift in thinking of statistics as a science of the state with the statists, as they were known, conducting surveys of trade, industrial progress, labor, poverty, education, sanitation, and crime”.

Another significant thread in the development of statistics is the statistical graphics where “the first major figure is William Playfair (1759−1823), credited with inventing line charts, bar charts, and the pie chart” that it was considered the “period from 1850 to 1900 as the “golden age of statistical graphics””. Significant statistics organizations were formed during this century like the Royal Statistical Society in 1834 as the London Statistical Society (LSS), and the American Statistical Association, formed in 1839.

During the 1900, statistics paved way to more developments:

Another wave of activity into the 1920s was initiated by the concerns of William Gosset, reaching its culmination in the insights of Ronald Fisher with the development of experimental design, analysis of variance, maximum likelihood estimation, and refinement of significance testing. This was followed by the collaboration of Egon Pearson and Jerzy Neyman in the 1930s, giving rise to hypothesis testing and confidence intervals. At about the same time came Bruno de Finetti’s seminal work on subjective Bayesian inference and Harold Jeffreys’s work on “objective” Bayesian inference so that by 1940 we had most of the basics of the theories of the “modern statistics” of the twentieth century. World War II was also a time of great progress as a result of drafting many young, mathematically gifted people into positions where they had to find timely answers to problems related to the war effort. Many of them stayed in the field of statistics swelling the profession. We also draw particular attention to John Tukey’s introduction of “exploratory data analysis” in the 1970s; this is an approach to data analysis that involves applying a variety of exploratory techniques, many of them visual, to gain insight into a dataset and uncover underlying structure and exceptions (Wild et al., 2018).

Conversely, along with the development of statistics is the formation of local, national and international associations and conferences in different parts of the world solely dedicated on statistics “whose overall mission is to promote the understanding, development and good practice of statistics worldwide” (MacGillivray et al., 2014). Example of these are The International Association for Statistical Education (IASE) is one of the Associations of the International Statistical Institute (ISI) which was founded in 1885, then the first international roundtable on statistics education was held in 1968, and the first International Conference on Teaching Statistics (ICOTS) was held in 1982” (ibid.). Way back then, statistics continues to thrive as the IASE organizes satellite conferences to the biennial ISI World Statistics Congresses (WSC), statistical education strands within the WSC, and international roundtables every 4 years. Nonetheless, in order to promote statistics at its core, there are also options of publishing one’s researches wherein “IASE offers an optional double-blinded refereeing process and publishes proceedings, now online” (ibid.).

#### Local Perspective: The Philippine Context

One literature most relevant to this section is the paper “The Teaching of Statistics in The Philippines: Moving to A Brighter” by Bersales (2010) of the University of the Philippines. Bersales traces the development of statistics in the Philippines in the year 1953 when “Statistical Training Center was established under a bilateral agreement between the Philippine government and the United Nations” with the observation of the first board of directors of the Philippine Statistical Association that “staff doing statistical work then did not have formal training in statistics as well as college education offered only three units of elementary statistics and there were no undergraduate and graduate programs in statistics in the Philippines”. And so, the Center offered its first academic program, Master of Arts in Statistics, instituted in 1954. During the years of 1953-1969, the degree offering of Statistics in the country has gained momentum where the Center offered MS and PhD courses as well as “faculty for the center were recruited and sent to American universities to earn their MS degrees and/or PhD degrees”.

Later, the Center was formally turned over to the University of the Philippines in 1963 and in 1998, the Center was renamed The School of Statistics, which “provided more recognition in the university of statistics as a discipline separate from mathematics” and more academic programs were instituted in other universities in the Philippines such as University of the Philippines Los Baños (UPLB), Polytechnic University of the Philippines (PUP), and Mindanao State University – Iligan Institute of Technology (MSU-IIT).

The year 2000 onwards, Bersales describes as the technology years where “the importance of computer software and hardware in the practice of statistics gained recognition”. This means that learners and teachers need to be abreast with these changes yet “the teaching of statistics was modified and enhanced with technology but the enhancement did not come fast since access to facilities was available only to a few as well not all teachers were trained in the use of software and hardware”. Bersales even recognized that these challenges of statistics in the Philippines “is still existent up to the present”. The author also points out that in 2006, “nineteen academic institutions were already offering statistics programs from B.S. to PhD” and graduates became in demand in private and government institutions.

With this development of statistics in the Philippines, Bersales and some authors reports some problems of teaching statistics in the Philippines:

lack of good quality statistics books, lack of qualified teachers in statistics, inadequate facilities such as computer laboratories to aid in teaching statistics, teaching methods that do not enhance students’ learning of statistics. Additional problems were identified during the teacher trainings of the Philippine Statistical Association: lack of recognition of statistics as an important course in their respective colleges, dearth of local reference materials that have passed the review of a panel of experts, teachers’ need for more hands-on practice on handling data, unavailability of statistical software in their colleges, lack of qualified statistician as member in research/thesis advisory committee (Bersales, 2010).

These problems are also true in our university, NVSU, though we are not offering Statistics degree program but college students especially BSED major in Mathematics are required to get statistics subjects. Thus, this preliminary study is geared towards program development (i.e. existence of computer laboratories equipped with open-source statistical software dedicated to mathematics major students and policy formulations and integration (streamlining of university’s budget for statistics packages, programs and development).

## RStudio in Teaching Statistics

There are many statistical software that can be used in analyzing data. One most prominent software used in universities is the SPSS. Though, SPSS is widely used, it can be said that it is expensive that one has to buy its license. Hence, an alternative to this is looking for a statistical software that is open-source and free. One of these is the R using RStudio (RStudio Team, 2016).

One example of how RStudio is used in teaching statistics is the article “Supporting Data Science in the Statistics Curriculum” by Loy, Kuiper, & Chihara (2019). Their study “describes a collaborative project across three institutions to develop, implement, and evaluate a series of tutorials and case studies that highlight fundamental tools of data science – such as visualization, data manipulation, and database usage – that instructors at a wide-range of institutions can incorporate into existing statistics courses”. What is interesting in this study was the use of R statistical software. The authors argue that “while R is certainly not the *only* choice, we believe it is the *best* choice when adding these topics to existing statistics courses” (emphasis, original).

In addition to this, the authors also identified six (6) reasons in choosing “R tutorials and case studies to help students develop facility with statistical software for data management and visualization”:

1. R is one of the most popular programing languages in the world (Cass, 2017).
2. R was developed by statisticians for statistical analysis, making it is a natural choice for a statistics course. Additionally, there is less overhead required for tasks, such as data visualization than in Python (another popular language for data science), especially when using packages, such as mosaic and ggformula, which are designed to be easily accessible to people with no programing background.
3. R is open source, so students are learning a toolkit that will still be accessible to them after they complete the course.
4. RStudio is consistent across operating systems, eliminating the need for multiple sets of instructions. This is not the case with other software packages – even Excel is not identical across platforms. Additionally, your institution can set up an RStudio Server for your students to ensure that everyone has exactly the same version of R, the necessary R packages, and even datasets (Çetinkaya-Rundel & Horton, 2016).
5. R makes reproducibility easy. For example, if you share your dataset and R Markdown document, then your analysis can be easily rerun by another researcher.
6. Graphics, data, and RMarkdown files are easy to export into other formats.

Given these benefits of RStudio, however, the authors also emphasize two (2) common pitfalls in using R. These are the following: 1) First, the error messages produced by knitting an R Markdown file are often harder to decipher than the errors produced within code chunks; 2) Second, students often have trouble reading their own data into R within R Markdown documents if they do not save the data in the same location as the .Rmd file. Though the authors point these pitfalls, what is good about it, is they also recommend solutions by “providing zip files containing the RStudio project and all associated files so that students can simply open the project and start working through creating a GitHub repository for the labor assignment containing the necessary file and using DownGit”.

In relation to the use of R software, another interesting article “Teaching with R – A Curse or A Blessing?” by Gomes & de Sausa (2018) explains the advantages and disadvantages of R to undergraduate and graduate students of Mathematics and Social Sciences using an online module. Gomes & de Sausa (2018) argue that “although the advantages of R are well-known (free, open source, continually updated by experts), it is not the first choice among college students, especially those not majoring in mathematics or statistics”. The authors contend that “a problem that appears when teaching R is that once the great potential of the software is understood, the temptation is to focus immediately on more advanced analysis, which adds frustration for beginning learners of R”.

Aside from this, the authors also discuss two possible reason why teaching with R could be a blessing or curse: 1) possible reasons for this resistance to R, whether the learners are undergraduate or graduate students or even other teachers from a variety of areas of science, is the fact that, teachers tend to emphasize a wide range of commands and programming lines from very early on, making learning R a slow and frustrating task; 2) learning R may be seen as similar to learning a foreign language… then a student with no skills or practice in statistics or programming languages should be able to learn R.

In their discussion of their pilot study with R to undergraduate and graduate students, they found out that “the problematic issues focused more on their comprehension of the R language”. Another reason is “the individuals’ varying degree of knowledge, both technological and scientific, and in particular their overall knowledge of certain statistical principles”. Hence, in order to address these problems, the authors proposed an online module using Facebook.

Interestingly, with these innovations and variations in teaching R with students, the results show that “students served to encourage greater implementation of the use of free software, namely R, in the pedagogical practices of teachers and in their daily life”. Not only that, the way the authors present the online module through Facebook environment “contributed to promoting greater student participation in the learning process, one that is more focused not only on their autonomous work, but also on the development of their abilities to work as a team”.

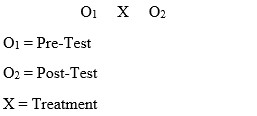
Hence, Gomes & de Sausa (2018) conclude that teaching with R could be a blessing if the right buttons of activities and even academic environments blend well. This is true as they conclude that “starting this process even at an early stage in school will most certainly contribute not only to the improvement of teaching methodologies, but also to the promotion of statistical literacy among students and teachers”.

# METHODOLOGY

The following sections will discuss the research methodology that was employed to answer the research questions.

## Research Design

The research design is a poor experimental design: One group pre-test–post-test study. This design includes a pre-test measure followed by a treatment and a post-test for a single group. The pre-test’s content is the same as the post-test’s. The difference is the statistical computating tool used by the respondents to answer the test. In the pre-test, the respondents used calculator. In the post-test, the respondents used RStudio. The treatment is introducing the basics of RStudio in statistical computing. Figure 2 shows a simple diagram of the one group pre-test–post-test study. The mean calculator test score and the mean RStudio test score were compared to see any significant improvement in the test scores of the respondents.



**Figure 2.** One Group Pre-Test–Post-Test Study

## Research Environment

The research environment is CTE, NVSU.

## Respondents

The respondents of the study are BSEd Mathematics students taking Inferential Statistics in CTE, NVSU.

## Research Instruments

The research instrument is a survey questionnaire for the profile variables (see Appendix A).

## Data Gathering Procedure

Data was gathered by conducting the calculator test and RStudio test to the respondents. The survey questionnaire was administered after the post-test or the RStudio test.

## Data Analysis

The non-parametric Wilcoxon rank-sum test – an alternative to Student’s *t* test – was performed to see whether there is significant difference in the mean calculator test score and in the mean RStudio test score when grouped according to profile variables. A Wilcoxon rank-sum test is suitable because the distribution of the calculator test scores and RStudio test scores is unknown, and the samples are small. The Pearson’s product-moment correlation was computed to see whether there is significant relationship between age and calculator test scores and age and RStudio test scores. The signed Wilcoxon signed rank-sum test was performed to see whether there is significant difference between the mean calculator test score and RStudio test score since the scores are paired. The statistical computing tool used in the study is RStudio.

# RESULTS AND DISCUSSION

## Profile Variables of the Respondents

The profile variables of the respondents considered in the study are gender: male or female, have previous programming experience: yes or no, statistical computing preference: calculator or RStudio, and age.

### Gender

Most of those who enroll BSEd mathematics are female. It is no surprise that among the respondents, 62.50% are female and 37.50% are male. Table 1 shows the frequency distribution of gender.

**Table 1:** Frequency Distribution of Gender

|  |  |  |
| --- | --- | --- |
| Gender | Freq | Percentage |
| Female | 5 | 62.5 |
| Male | 3 | 37.5 |
| Total | 8 | 100 |

### Have Previous Programming Experience

Among the respondents, half have previous programming experience while the other half have none. Table 2 shows the frequency distribution for have previous programming experience. It seems that some students still have no programming experience when they take their second course in statistics.

**Table 2:** Frequency Distribution of Have Previous Programming Experience

|  |  |  |
| --- | --- | --- |
| Experienced | Freq | Percentage |
| No | 4 | 50 |
| Yes | 4 | 50 |
| Total | 8 | 100 |

### Statistical Computing Tool Preference: Calculator or RStudio

Having learned the basics of using calculator and RStudio for statistical computing and took the examination using calculator and RStudio and then knowing their scores, half of the respondents still preferred to use calculator while the other half now preferred to use RStudio. Table 3 shows the frequency distribution for statistical computing tool preference. Among those who preferred RStudio, one said “it is easy”. Among those who preferred calculator, one found using RStudio “complicated” and that using calculator is “comfortable”. Table 3 shows the frequency distribution of statistical computing tool preference.

**Table 3:** Frequency Distribution of Statistical Computing Tool Preference

|  |  |  |
| --- | --- | --- |
| Preference | Freq | Percentage |
| Calculator | 4 | 50 |
| RStudio | 4 | 50 |
| Total | 8 | 100 |

### Age

Half of the respondents are 21 years old, two are 22 years old, one is 24 years old and one is 39 years old. Table 4 is the frequency distribution of age.

**Table 4:** Frequency Distribution of Age

|  |  |  |
| --- | --- | --- |
| Age | Freq | Percentage |
| 21 | 4 | 50 |
| 22 | 2 | 25 |
| 24 | 1 | 12.5 |
| 39 | 1 | 12.5 |
| Total | 8 | 100 |

## Calculator Test Scores and RStudio Test Scores of the Respondents

There are two sets of examination scores: the calculator test scores and RStudio test scores.

### Calculator Test Scores

The result of the calculator test shows that 62.50% failed and 37.50% passed. The mean score is 46.25% which is not even passing. The exam result is not remarkable. Table 5 shows the frequency distribution and summary of calculator test scores.

**Table 5:** Frequency Distribution and Summary of Calculator Test Scores

|  |  |  |
| --- | --- | --- |
| Scores | Freq | Percentage |
| Failed | 5 | 62.5 |
| Passed | 3 | 37.5 |
| Total | 8 | 100 |

|  |  |  |  |
| --- | --- | --- | --- |
| Min | Mean | Max | SD |
| 25 | 46.25 | 100 | 25.06 |

### RStudio Test Scores

The result of the RStudio test shows that 37.50% passed and 62.50% failed. The mean score is 54.50 which is passing. Compared to the calculator test scores, the RStudio test scores seem to have a better result. Table 6 shows the frequency distribution and summary of RStudio test scores.

**Table 6:** Frequency Distribution and Summary of RStudio Test Scores

|  |  |  |
| --- | --- | --- |
| Scores | Freq | Percentage |
| Failed | 3 | 37.5 |
| Passed | 5 | 62.5 |
| Total | 8 | 100 |

|  |  |  |  |
| --- | --- | --- | --- |
| Min | Mean | Max | SD |
| 24 | 54.5 | 86 | 21.23 |

## Calculator Test Scores and RStudio Test Scores of the Respondents When Grouped According to the Profile Variables

The calculator test scores and the RStudio test scores of the respondents are grouped according to the profile variables gender, having previous programming experience and statistical computing tool preference.

### Calculator Test Scores According to Gender

Among the female respondents, only one passed the test. Among the male respondents, only one failed the test. The mean score for female respondents is 37 which is not passing. The mean score for male respondents is 61.67 which is passing. Table 7 shows the frequency distribution and summary of calculator test scores according to gender. The calculator test scores of male respondents seems to be better than the scores of female respondents.

**Table 7:** Frequency Distribution and Summary of Calculator Test Scores According to Gender

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Gender | Failed | Passed | Min | Mean | Max | SD |
| Female | 4 | 1 | 25 | 37 | 54 | 11.94 |
| Male | 1 | 2 | 27 | 61.67 | 100 | 36.64 |

### RStudio Test Scores According to Gender

Among the female respondents, the number of respondents who failed the test is one respondent greater than the number of respondents who passed the test. All of the male respondents passed the test. The mean score for female respondents is 43.20 which is not passing. The mean score for male respondents is 73.33 which is passing. Table 8 shows the frequency distribution and summary of RStudio test scores according to gender. The RStudio test scores of male respondents seem to be better than the scores of female respondents.

**Table 8:** Frequency Distribution and Summary of RStudio Test Scores According to Gender

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Gender | Failed | Passed | Min | Mean | Max | SD |
| Female | 3 | 2 | 24 | 43.2 | 68 | 16.87 |
| Male | 0 | 3 | 61 | 73.33 | 86 | 12.5 |

### Calculator Test Scores According to Have Previous Programming Experience

Half of the respondents who don’t have previous programming experience failed the calculator test and the other half passed. Only one respondent who have previous programming experience passed the test. The mean score for those who don’t have previous programming experience is 54.25 which is passing. The mean score for those who have previous programming experience is 38.25 which is not passing. Table 9 shows the frequency distribution and summary of calculator test scores according to have previous programming experience.

**Table 9:** Frequency Distribution and Summary of Calculator Test Scores According to Have Previous Programming Experience

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Experienced | Failed | Passed | Min | Mean | Max | SD |
| No | 2 | 2 | 27 | 54.25 | 100 | 32.5 |
| Yes | 3 | 1 | 25 | 38.25 | 58 | 15.44 |

### RStudio Test Scores According to Have Previous Programming Experience

Half of the respondents who don’t have previous programming experience failed the RStudio test and the other half passed. Only one respondent who have previous programming experience failed the test. The mean score for those who don’t have previous programming experience is 52.50 which is passing. The mean score for those who have previous programming experience is 56.50 which is passing. Table 10 shows the frequency distribution and summary of RStudio test scores according to have previous programming experience.

**Table 10:** Frequency Distribution and Summary of RStudio Test Scores According to Have Previous Programming Experience

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Experienced | Failed | Passed | Min | Mean | Max | SD |
| No | 2 | 2 | 35 | 52.5 | 86 | 23.39 |
| Yes | 1 | 3 | 24 | 56.5 | 73 | 22.22 |

### Calculator Test Scores According to Statistical Computing Tool Preference

In the calculator test, only one passed among those who prefer calculator as statistical computing tool. Among those who preferred RStudio as statistical tool, half passed and half failed in the calculator test. The mean score for those who prefer calculator is 48.75 which is not passing. The mean score for those who prefer RStudio is 43.75 which is also not passing. Table 11 shows the frequency distribution and summary of calculator test scores according to statistical computing tool preference.

**Table 11:** Frequency Distribution and Summary of Calculator Test Scores According to Statistical Computing Tool Preference

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Preference | Failed | Passed | Min | Mean | Max | SD |
| Calculator | 3 | 1 | 25 | 48.75 | 100 | 35.1 |
| RStudio | 2 | 2 | 27 | 43.75 | 58 | 14.71 |

### RStudio Test Scores According to Statistical Computing Tool Preference

Among those who prefer calculator as statistical computing tool, half passed and half failed the RStudio test. Among those who prefer RStudio as statistical computing tool, only one failed. The mean score of those who prefer calculator is 53.25 which is passing. The mean score of those who prefer RStudio is 55.75 which is also passing. Table 12 shows the frequency distribution and summary of RStudio test scores according to statistical computing tool preference.

**Table 12:** Frequency Distribution and Summary of RStudio Test Scores According to Statistical Computing Tool Preference

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Preference | Failed | Passed | Min | Mean | Max | SD |
| Calculator | 2 | 2 | 24 | 53.25 | 86 | 28.74 |
| RStudio | 1 | 3 | 38 | 55.75 | 73 | 14.86 |

## Significant Difference in the Calculator Test Scores and the RStudio Test Scores of the Respondents when Grouped According to the Profile Variables

The calculator test scores and the RStudio test scores are grouped according to gender and have previous programming experience and statistical computing tool preference.

### Calculator Test Scores

For the calculator test scores of the respondents when grouped according to gender, the Wilcoxon rank-sum test gives a p value of 0.29. Since the p value is greater than the significance level of 0.05, the null hypothesis that the mean calculator test score of male is equal to the mean calculator score of female is not rejected. There is no statistical evidence that the mean score of male of 61.67 is greater than the mean score of female of 37. Gender is not a factor among the respondents when it comes to their performance in the calculator test.

For the calculator test scores of the respondents when grouped according to have previous programming experience, the Wilcoxon rank-sum test gives a p value of 0.56. Since the p value is greater than the significance level of 0.05, the null hypothesis that the mean calculator test score of the respondents who have previous programming experience is equal to the mean calculator score of the respondents who have no previous programming experience is not rejected. There is no statistical evidence that the mean score of the respondents who have no previous programming experience of 54.25 is greater than the mean score of the respondents who have previous programming experience of 38.25. Having previous programming experience is not a factor among the respondents when it comes to their performance in the calculator test.

### RStudio Test Scores

For the RStudio test scores of the respondents when grouped according to gender, the Wilcoxon rank-sum test gives a p value of 0.07. Since the p value is greater than the significance level of 0.05, the null hypothesis that the mean RStudio test score of male is equal to the mean RStudio test score of female is not rejected. There is no statistical evidence that the mean score of male of 73.33 is greater than the mean score of female of 43.20. Gender is not a factor among the respondents when it comes to their performance in the RStudio test.

For the RStudio test scores of the respondents when grouped according to have previous programming experience, the Wilcoxon rank-sum test gives a p value of 0.89. Since the p value is greater than the significance level of 0.05, the null hypothesis that the mean RStudio test score of the respondents who have previous programming experience is equal to the mean RStudio test score of the respondents who have no previous programming experience is not rejected. There is no statistical evidence that the mean score of the respondents who have previous programming experience of 56.50 is greater than the mean score of the respondents who have no previous programming experience of 52.50. Having previous programming experience is not a factor among the respondents when it comes to their performance in the RStudio test.

## Significant Relationship Between Age and the Calculator Test Scores and Between Age and the RStudio Test Scores of the Respondents

### Calculator Test Scores and Age

The computed p value and correlation of calculator test scores and age is 0.004 and 0.88 respectively. Table 13 shows that the p value of 0.004 is less than the significance level of 0.05. The null hypothesis that calculator test scores and age are linearly uncorrelated is rejected. At the 5% significance level, the data provide sufficient evidence to conclude that calculator test scores and age are positively linearly correlated. There is a very strong positive correlation. The older the respondent, the better the calculator test score.

**Table 13:** Pearson’s product-moment correlation: Age and Calculator Test

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Test statistic | df | P value | Alternative hypothesis | cor |
| 4.637 | 6 | 0.003552 \* \* | two.sided | 0.8842 |

### RStudio Test Scores and Age

The computed p value and correlation of RStudio test scores and age is 0.12 and 0.59 respectively. The p value of 0.12 is greater than the significance level of 0.05. The null hypothesis that RStudio test scores and age are linearly uncorrelated is not rejected. At the 5% significance level, the data provide no evidence to conclude that RStudio test scores and age are positively linearly correlated. In their RStudio test scores, age does not matter among the respondents.

## Significant Difference Between the Calculator Test Scores and RStudio Test Scores of The Respondents

Comparing the mean calculator test score and the mean RStudio test score, the Wilcoxon signed rank-sum test gives a p value of 0.25. Since the p value is greater than the significance level of 0.05, the null hypothesis that the mean calculator test score is equal to the mean RStudio test score is not rejected. There is no statistical evidence that the mean RStudio test score of 54.50 is greater than the mean calculator test score of 46.25. The use of RStudio as statistical computing tool in the test has not significantly improved the performance of the respondents.

# SUMMARY, CONCLUSION AND RECOMMENDATIONS

## Summary

Based on the discussion of results, here are the salient points of my study:

1. The profile variables of the respondents considered in the study are gender, previous programming experience, and statistical computing preference using calculator or RStudio. The results showed that out of 8 students enrolled in second statistics course, there are more female enrolled compared to male wherein the age of the respondents ranges from 21 to 39 years old.
2. With regard to previous programming experience and statistical computing preference using calculator or RStudio, both items gained an equal answer of yes and no as well as equal answer with using calculator or RStudio.
3. Age is a factor in calculator test, that the older the respondent, the better is the calculator test result. It seems that the result suggests that the older a respondent is, means that they are more attached in/to using the “traditional method of doing statistics” which is calculator.
4. In terms of using RStudio as a statistical tool, the results showed that gender and previous programming experience are not factors in using RStudio. This may suggest that learning RStudio as a statistical tool does not need programming experience, that it can be easily learned which can be a friendly-user software as long as students are interested in learning its features. Hence, using RStudio could be a “blessing and not a curse” as some authors coined it.
5. Most students failed the calculator test in statistics while many passed the statistics exam using the RStudio. This may suggest that using technology in teaching statistics has a “bright future” as Bersales (2010) termed it. With today’s young generation of 4th Industrial revolution, where technology is everywhere, it seems that learning and teaching statistics should be abreast with the digital age we live in.
6. On the other hand, the use of RStudio as statistical computing tool in the test has not significantly improved the performance of the respondents. This result may suggest that with the advent of digital age, students should be abreast with technological advancement, especially with courses and subjects that need mathematical and computer laboratories

## Conclusion

This study then concludes that using RStudio as a computing tool may have an impact with students taking second subject statistics. Hence, teachers and students may both benefit if technology will be used in teaching statistics. Thus, this may also serve as departure points for degree course program developments and policy formulation to statistics and mathematics courses and subjects.

## Recommendations

This study would recommend the following for future researches:

1. Increase the number of respondents as well as to sample respondents from other state universities who have taken the same statistics subject.
2. Do the survey to high school students such as Philippine Science High School Region II.
3. Consider other variables for correlation such as educational background (i.e. high school graduated from, K-12 strand and academic standing.
4. Explore other statistical software as measurement of technology readiness of students aside from RStudio.
5. Results can be channeled to the department, college and university officials for possible curriculum enhancement/ revision to suit the digital age.

# Reference

Ben-Zvi, D., & Makar, K. (Eds.). (2016). *The teaching and learning of statistics: International perspectives*. Cham: Springer.

Ben-Zvi, D., Makar, K., & Garfield, J. (Eds.). (2018). *International handbook of research in statistics education*. Cham: Springer.

Bersales, L. G. S. (2010). The teaching of statistics in the Philippines: Moving to a brighter future. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society*. Proceedings of the Eight International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia. Voorburg, The Netherlands: International Statistical Institute; International Association of Statistical Education.

Biehler, R., Ben-Zvi, D., Bakker, A., & Makar, K. (2013). Technology for enhancing statistical reasoning at the school level. In M. A. Clements, A. J. Bishop, C. Keitel, J. Kilpatrick, & F. K. S. Leung (Eds.), *Third international handbook of mathematics education* (pp. 643–690). New York: Springer.

Cass, S. (2017). The 2017 top programming languages – IEEE Spectrum. *IEEE Spectrum: Technology, Engineering, and Science News*. Retrieved from [https://spectrum.ieee.org/computing/software/the-2017-topprogramming-
languages](https://spectrum.ieee.org/computing/software/the-2017-topprogramming-%0Alanguages)

Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2007). The role of technology in improving student learning of statistics. *Technology Innovations in Statistics Education*, *1*(1). Retrieved from <https://escholarship.org/uc/item/8sd2t4rr>

Chance, B., & Rossman, A. (2006). Using simulation to teach and learn statistics. In A. Rossman & B. Chance (Eds.), *Proceedings of the seventh international conference on teaching statistics*. Voorburg, The Netherlands: International Statistical Institute.

Çetinkaya-Rundel, M., & Horton, N. J. (2016). Technology lowering barriers: Get started with R at the snap of a finger. Retrieved from <https://www.causeweb.org/cause/ecots/ecots16/breakouts/7>

Çetinkaya-Rundel, M., & Rundel, C. (2017). Infrastructure and tools for teaching computing throughout the statistical curriculum. *The American Statistician*. <https://doi.org/10.1080/00031305.2017.1397549>

Doi, J., Potter, G., Wong, J., Alcaraz, I., & Chi, P. (2016). Web application teaching tools for statistics using R and Shiny. *Technology Innovations in Statistics Education*, *9*(1). Retrieved from <https://escholarship.org/uc/item/00d4q8cp>

Franklin, C., Kader, G., Mewborn, D. S., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report: A pre-k-12 curriculum framework*. Alexandria, VA: American Statistical Association.

GAISE College Report ASA Revision Committee. (2016). Guidelines for assessment and instruction in statistics education college report 2016. Retrieved from <http://www.amstat.org/education/gaise>

Gomes, D., & de Sausa, B. (2018). Teaching with R – a curse or a blessing? In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward*. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan. Voorburg, The Netherlands: International Statistical Institute; International Association of Statistical Education.

Gould, R. (2010). Statistics and the modern student. *International Statistical Review*, *72*(2), 297–315. <https://doi.org/10.1111/j.1751-5823.2010.00117.x>

Gould, R., Wild, C. J., Baglin, J., McNamara, A., Ridgway, J., & McConway, K. (2018). Revolutions in teaching and learning statistics: A collection of reflections. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 457–472). Cham: Springer.

Hardin, J., Hoerl, R., Horton, N. J., Nolan, D., B. Baumer, Hall-Holt, O., … Ward, M. D. (2015). Data science in statistics curricula: Preparing students to “think with data”. *The American Statistician*. <https://doi.org/10.1080/00031305.2015.1077729>

Harraway, J. A. (2012). Learning statistics using motivational videos, real data and free software. *Technology Innovations in Statistics Education*, *6*(1). Retrieved from <https://escholarship.org/uc/item/1fn7k2x3>

Horton, N. J. (2015). Challenges and opportunities for statistics and statistical education: Looking back, looking forward. *The American Statistician*, *69*(2), 138–145. <https://doi.org/10.1080/00031305.2015.1032435>

Horton, N. J., Baumer, B. S., & Wickham, H. (2015). Taking a chance in the classroom: Setting the stage for data science: Integration of data management skills in introductory and second courses in statistics. *CHANCE*, *28*(2), 40–50. <https://doi.org/10.1080/09332480.2015.1042739>

Loy, A., Kuiper, S., & Chihara, L. (2019). Supporting data science in the statistics curriculum. *Journal of Statistics Education*, *27*(1), 2–11. <https://doi.org/10.1080/10691898.2018.1564638>

MacGillivray, H., Martin, M. A., & Phillips, B. (Eds.). (2014). *Topics from Australian conferences on teaching statistics: OZCOTS 2008-2012* (Vol. 81). New York: Springer.

McNamara, A. (2018). Key attributes of a modern statistical computing tool. *The American Statistician*. <https://doi.org/10.1080/00031305.2018.1482784>

R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>

RStudio Team. (2016). *RStudio: Integrated development environment for R*. Boston, MA: RStudio, Inc. Retrieved from <http://www.rstudio.com/>

Stander, J., & Dalla Valle, L. (2017). On enthusing students about big data and social media visualization and analysis using R, RStudio, and RMarkdown. *Journal of Statistics Education*, *25*(2), 60–67. <https://doi.org/10.1080/10691898.2017.1322474>

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from <https://ggplot2.tidyverse.org>

Wild, C. J., Utts, J. M., & Horton, N. J. (2018). What is statistics? In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 5–36). Cham: Springer.

Xie, Y., Allaire, J. J., & Grolemund, G. (2018). *RMarkdown: The definitive guide*. Boca Raton, Florida: Chapman; Hall/CRC. Retrieved from <https://bookdown.org/yihui/rmarkdown>

Zeiffler, A., Garfield, J., & Fry, E. (2018). What is statistics education? In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 37–70). Cham: Springer.