

Singapore Management University

## Institutional Knowledge at Singapore Management University

---

Research Collection School Of Information Systems

School of Information Systems

---

12-2018

### An essential Applied Statistical Analysis course using RStudio with Project-Based Learning for Data Science

Aldy GUNAWAN

Singapore Management University, [aldygunawan@smu.edu.sg](mailto:aldygunawan@smu.edu.sg)

Michelle L. F. CHEONG

Singapore Management University, [michcheong@smu.edu.sg](mailto:michcheong@smu.edu.sg)

Johnson POH

Follow this and additional works at: [https://ink.library.smu.edu.sg/sis\\_research](https://ink.library.smu.edu.sg/sis_research)



Part of the [Databases and Information Systems Commons](#), and the [Numerical Analysis and Scientific Computing Commons](#)

---

#### Citation

GUNAWAN, Aldy; CHEONG, Michelle L. F.; and POH, Johnson. An essential Applied Statistical Analysis course using RStudio with Project-Based Learning for Data Science. (2018). *2018 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE): Wollongong, Australia, December 4-7: Proceedings*. 581-588. Research Collection School Of Information Systems. Available at: [https://ink.library.smu.edu.sg/sis\\_research/4243](https://ink.library.smu.edu.sg/sis_research/4243)

This Conference Proceeding Article is brought to you for free and open access by the School of Information Systems at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Information Systems by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email [library@smu.edu.sg](mailto:library@smu.edu.sg).

# An Essential Applied Statistical Analysis Course using RStudio with Project-Based Learning for Data Science

Aldy Gunawan  
School of Information Systems  
Singapore Management University  
Singapore  
aldygunawan@smu.edu.sg

Michelle Lee Fong Cheong  
School of Information Systems  
Singapore Management University  
Singapore  
michcheong@smu.edu.sg

Johnson Poh  
School of Information Systems  
Singapore Management University  
Singapore  
johnsonpoh@outlook.com

**Abstract**—This paper presents a new postgraduate level course, named **Applied Statistical Analysis with R**. We present the course structure, teaching methodology including the assessment framework and student feedback. The course covers the basic concepts of statistics, the knowledge of applying statistical theory in analyzing real data and the skill of developing statistical applications with R programming language. The first half of each lesson is dedicated to teaching students the statistical concepts while the second half focuses on the practical aspects of implementing the concepts within the RStudio console. The Project-Based Learning (PBL) approach is adopted to encourage students to apply the knowledge gained to solve real world problems, answer complex questions and generate high-quality results. We present various interesting projects to show how the students have implemented their statistical knowledge in solving real problems. It is concluded that combining hands-on experience using RStudio and PBL is an effective teaching methodology to achieve the desired learning outcomes. More than 80% of students agree with the quality and usefulness of projects and there is a strong positive correlation of 0.74 between the final project grade and the final grade obtained by the students.

**Keywords**—*applied statistical analysis; project-based learning; RStudio*

## I. INTRODUCTION

Recent advances of technologies have enabled many seamless ways of generating and collecting larger volume and variety of data. Applied Statistics is hence the relevant branch of Mathematics that is used to visualize, analyze, interpret, and predict outcomes from these data. Kirk [1] mentioned that the content of statistics course has been shifted due to more sophisticated concepts and technological tools. Instead of focusing on statistical computations, the focus moves towards more fundamental meaning of the statistics constructs.

The study of statistics is categorized into two main branches: descriptive statistics and inferential statistics [2], both of which are considered very dry for students to learn. To stimulate interest in students, improve students' motivation and confidence in statistic courses, Fawcett [3] used case-based learning method with real-life, discipline-specific case studies. Landrum and Smith [4] suggested to implement some hands-on experience with a research project in statistical courses to achieve effective learning in statistics.

In this paper, we present a new postgraduate level course, named **Applied Statistical Analysis with R (ASAR)**, first introduced in Term 1 Academic Year (AY) 2017 – 2018 at the Singapore Management University (SMU). We implement the Project-Based Learning (PBL) approach in this course to ensure that students are able to apply statistical knowledge they learnt to develop their own statistical applications in RStudio to solve real world problems. This aligns with one of the learning outcomes of the curriculum, as listed in Section II. Graduates will be equipped with statistical knowledge and analytics technology skills to solve real life problems. Students may be bored if statistical courses are conducted with standard or passive approaches, such as answering questions from textbooks or tutorials.

Students will present their group projects and get feedback from their peers and the course instructor. By using PBL, students are given the opportunities to use technology for finding resources and information, create products and collaborate more effectively [5]. This paper aims to present the course structure, teaching methodology including assessment framework and student feedback. We present various interesting student projects to demonstrate how the students have implemented their statistical knowledge in solving real problems. Our results show that the course design and adopting PBL are effective to achieve the desired learning outcomes in applied statistics, and more than 80% of our students agree with the quality and usefulness of the projects. There is also a positive correlation of 0.74 between the project grade and the final grade obtained by students.

The paper is organized as follows. In Section II, the teaching methodology of ASAR is described. Section III focuses on the short literature review and the description of PBL. We then present the implementation of PBL through various interesting student projects, the findings and student feedback are presented in Section IV. Section V presents the insights and course evaluation. Finally, conclusions and some ideas for future developments of the course and research work are summarized in Section VI.

## II. TEACHING METHODOLOGY

### A. Objectives

The objectives of ASAR are to allow students to have a working knowledge and appreciation of:

- 1) Concepts in probability theory and its computation
- 2) Ways and techniques for describing data
- 3) Methods for making statistically sound inferences from sample data
- 4) Using R programming for describing data, making statistical inferences and perform rigorous statistical analysis
- 5) Using RStudio integrated development environment to develop an interactive and insightful data application (<https://www.rstudio.com/>)

Face-to-face lectures, practical activities, and the PBL approach are integrated and used in delivering this course. The theoretical aspects (related to objectives (1) to (3)) are mainly taught via face-to-face lectures. For the practical aspect (objective (4)), ASAR focuses on the ability to use R programming for both descriptive and inferential statistics. Objective (5) links to using RStudio to develop an interactive and insightful data application where the PBL approach is adopted.

### B. Course Structure

The 14-week ASAR course has a typical class size of 40-45 students with three contact hours per week. The course topics are classified into three different groups, as summarized in Figure 1. Weeks 1 to 10 cover both the theoretical aspects of descriptive statistics and inferential statistics, and the practical aspects of using R programming language to perform the computation and presentation.

From Weeks 10 to 13, students will learn how to analyze data using advanced R packages as follows:

- **manipulate** is used for generating interactive plotting functions, such as slider, check-box and button.

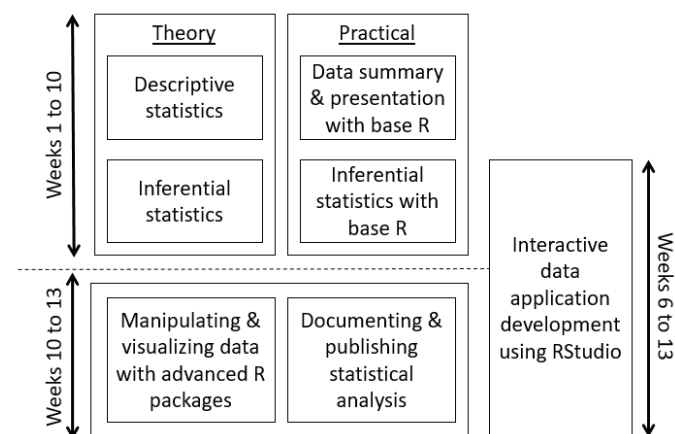


Fig. 1. ASAR course structure.

- **dplyr** is mainly for manipulating big data, such as filtering for rows, selecting specific columns, re-ordering rows and summarizing data.
- **ggplot2** is an elegant plotting system for R, such as histogram, density plot and scatter plot.

From Week 6 onwards, students will form their own groups to define a real problem, collect the data needed, and develop an interactive data application, using R Markdown documents and R Shiny package, and present the application in Week 13.

### C. Course Learning Outcomes and Assessments Framework

ASAR is structured with the following specific learning outcomes:

- Understand data types and probability distributions
- Implement statistical computations using R
- Create charts using R
- Interpret charts
- Apply central limit theorem
- Estimate population parameters from sample and implement hypothesis testing and Chi-squared testing
- Understand and compute correlation and regression relationship
- Create dynamic documents and reports using R
- Input data set and manipulate data frame using advanced R packages
- Apply statistical knowledge in real problem and build a statistical application using R studio

We create an assessment framework to continuously monitor students' performance as well as to identify students' learning needs. The assessments include two individual tests, in-class participation, one group project, and a final examination. Table I summarizes the assessment framework.

The first two assessments are formative assessments while the last two assessments are summative assessments. The objective of a formative assessment is to monitor students' learning progress, while a summative assessment is to evaluate students' learning at the end by comparing their results against some standard or benchmark. Two individual tests are conducted in Weeks 5 and 12 to test students' understanding on theories, fundamental and practical concepts of descriptive and inferential statistics, respectively. Class participation includes in-class discussions during lectures, lab exercises, and occasional group discussions throughout the entire term. The group project holds a high weightage since it involves processing and analyzing a real-world dataset of the students' choice.

TABLE I. ASSESSMENTS FRAMEWORK

Assessment Categories	Weightage	Schedule
2 × Individual tests	2 × 15%	Weeks 5 and 12
Class participation	10%	Entire term
Group project	30%	Weeks 6 onwards
Final examination	30%	Week 14

The project is first proposed by a formal proposal, graded on 10%, stating the objective of the project and the techniques that will be used. The other 20% weightage is carried by the final report (10%) and the final presentation in class (10%). The group project will be explained in more details in Section IV. The final examination which is conducted in Week 14 is an in-class written final exam covering the major concepts in the course to assess the students' knowledge gained and achievement of the course objectives.

### III. PROJECT BASED LEARNING

Project Based Learning (PBL) is defined as students working collaboratively to design solutions for authentic and meaningful questions and problems in the real world [11]. PBL is a teaching method in which students gain knowledge and skills by working for an extended period of time to investigate and respond to an authentic, engaging, and complex question, problem, or challenge. PBL offers students an opportunity to build confidence, solve problems, work in teams, communicate ideas, and manage themselves more effectively.

In a PBL environment, students tend to be more motivated to study and learn the relevant information for their projects in a dynamic approach [6]. PBL is student-centric as students are given the opportunity and flexibility to decide their own projects and solve real-world problems. The instructors will guide the students' learning to achieve the project objectives. In PBL, an in-depth and active learning process with systematic learning management is included in order to get applicable results, create motivation and strengthen necessary skills [7]. PBL has been successfully applied in several fields, e.g. power electronics [8], web applications [9], mathematics [10] and statistics [3].

PBL is applied as a complement to classical teaching methods of the power electronics course at the Regional University of Blumenau, Brazil [9]. Based on the students' perceptions, PBL serves as a basis for learning. Buj-Corral et al. [12] applied PBL method to evaluate the subject Project I at the School of Industrial Engineering of Barcelona. Students develop a project related to manufacturing and industrial statistics throughout a semester. The format of Project I is suitable to convey some transversal competences, such as independent learning, written communication and teamwork. In conclusion, students are satisfied with the way the course has been conducted.

Koparan and Güven [13] studied the effect of Project-Based Learning on the Statistical Literacy Levels of 8<sup>th</sup> Grade students in Trabzon, Turkey. Students were divided into two groups, control and intervention groups. In this context, the former experienced the traditional method of teaching while the latter

was taught using PBL. In conclusion, the PBL increased students' statistical literacy levels in the intervention group. Cobb and Moore [14] also concluded that statistics is taught most effectively with real data, and it would be even more practical if students collect their own data. Núñez-del-Prado and Gómez [15] implemented the PBL methodology in two courses: the Web Analytics and Data Mining courses, by developing projects such as a Web mining platform and a Library Analysis System. In both courses, students have the challenge to handle big amounts of heterogeneous data.

### IV. STUDENTS' PROJECTS

In this course, we choose to integrate both traditional and PBL approaches to achieve different purposes. The traditional approach emphasizes on the mastery of basic statistical knowledge through drill and practice, while the PBL requires students to apply their knowledge via the group project based on their own selected problems that engage their personal interests. Many studies including Yesilcay [16] have concluded that students actually learned more and improved their motivation by solving real-world projects. However, the truth is that many instructors still do not incorporate projects into their statistics courses to enhance students' learning [13]. PBL will allow students to demonstrate their skills in data acquisition, data cleaning, data enrichment, modeling, evaluation, and deployment.

Students need to form a group of four to five students on their own. The decision of the size of student teams is based on some considerations, such as the number of students in the entire cohort who takes the course, the number of weeks per semester, students' load in a semester and the course length (3.5 hours) in a week. They are given the opportunity to choose their own project topics, find relevant resources, define an analytics problem and apply the techniques to implement an interactive data application. In Week 6, students are required to submit their group proposal and the instructor will then assess the validity of the topics, evaluate the proposals, and discuss with the students for possible extensions, if any. The main role of the instructor is to develop the grading rubrics, manage the process and conduct weekly discussions in order to evaluate the progress of students' projects.

In Week 13, students have to submit their final reports and present their findings. During the presentation, each group is asked to share several key points of the project, such as: problem description, data, statistical concepts used and findings. The peer assessment involves each group's presentation being evaluated by students from the other groups. To ensure fairness, each group is required to evaluate and rank another three groups which are randomly selected. Individuals in each group are allowed to highlight some issues faced during the project discussions, if any. At the end of the session, the instructor summarizes the overall presentations, e.g. strengths and weaknesses for each group and the quality of the applications.

We briefly present several interesting projects completed by students in Term 2 AY 2017-2018. Many groups have decided to use high quality public datasets, such as <https://www.kaggle.com/datasets>. In order to prevent students

using existing codes from the website and ensure the originality of the project, instructors and teaching assistants were actively involved by having regular meetings or discussions with the students. By doing so, we can ensure that the students align their work with the objectives stated in their project proposals.

The main purpose of presenting student’s projects in this section is to show that the implementation of PBL is promising for the first round of offering this course. Three outstanding projects: “UK Accident Analysis”, “Kick Start Your Success” and “Airbnb”, will be described below.

### A. UK Accident Analysis

In this project (Fig. 2), students built a Shiny web application, which enables users to explore and predict traffic accidents. The data is extracted from <https://data.gov.uk/dataset/road-accidents-safety-data>. The objective is to provide users with the analysis of historical accidents in UK by regions. It also predicts the probability of accidents happening given several conditions, and predicts the probability of the accident being fatal or non-fatal. As shown in

Fig. 2, there are four different tabs: Interactive Map, Descriptive Statistics, Chi-Square Test and Logistic Regression. The Interactive Map tab covers the choropleth map that shows each area’s casualty rate and the accidents heat map that shows the hotspots for traffic accidents in a certain area.

The Descriptive Statistics tab mainly shows bar charts of selected variables. Fig. 3 shows a bar chart of the number of accidents for different road type. It shows that the number of accidents is very high on single carriageway type. Chi-Square test is conducted in order to test whether there is any association between a selected variable, e.g. speed limit and the accident severity (Fig. 4). It is observed that the speed limit of 60 mph has a strong association with both fatal and serious accident levels, as shown with the biggest bubbles. The last tab is Logistics Regression, for predicting the probability of a fatal accident happening given several conditions, such as speed limit (Fig. 5). In this project, the accuracy rate for each variable is found to be at least above 70%.

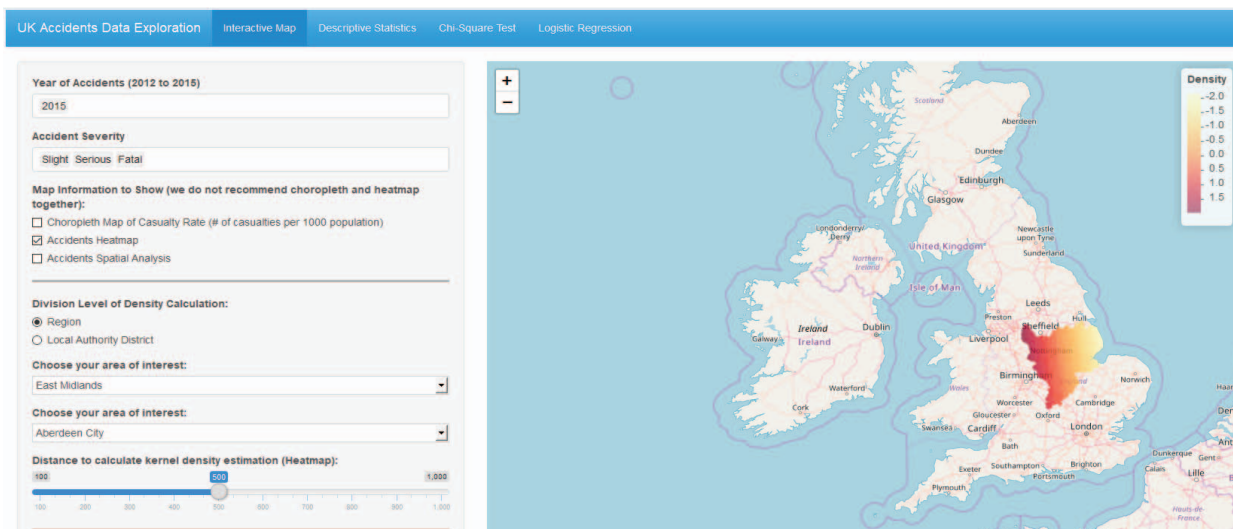


Fig. 2. UK Accident Analysis Dashboard

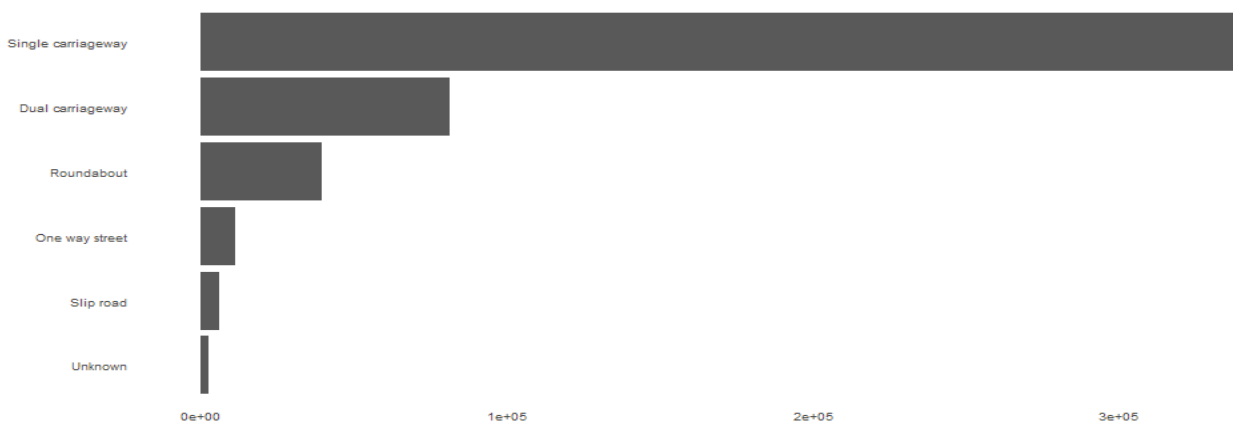


Fig. 3. Bar chart of road type variable

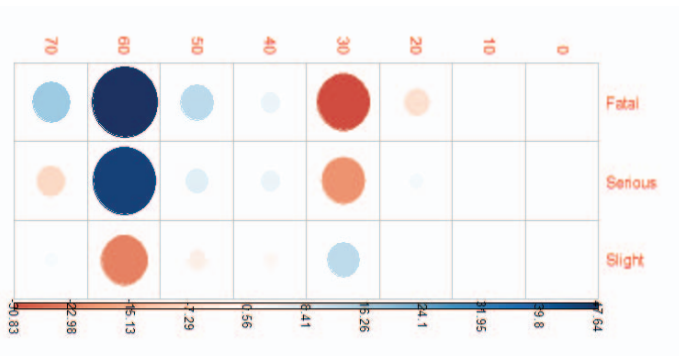


Fig. 4. Chi-Square Test of Speed Limit vs Accident Severity Level

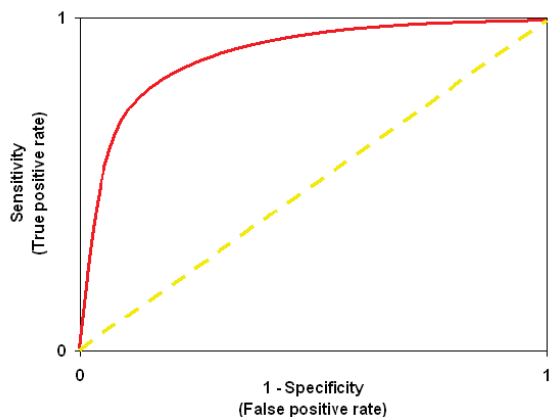


Fig. 5. Receiver Operator Characteristic (ROC)

### B. Kick Start Your Success

In this project, students created a Shiny web application for Kickstarter. Kickstarter is a global crowdfunding platform focused on creativity and merchandising. The Kickstarter dataset was collected from <https://www.kaggle.com/>. Fig. 6 displays the screenshot of the app with five different tabs: Exploration, Relationships, Comparison, Best Decision Tree and Best Word Cloud.

The Exploration tab focuses on data exploration by using simple descriptive statistical techniques. The Relationships tab builds the linear relationship between selected variables. In the Comparison tab, One-Way Analysis of Variance (ANOVA) is proposed to compare different levels of nominal variables. Fig. 7 illustrates an output for comparing percentage of goal reached for non-successful projects across 15 different project categories.

The Best Decision Tree tab helps the user to define the chance of success given selected conditions of inputs. Fig. 8 illustrates one example with the selected input values as shown, and the probability of success is estimated to be only 31.5%.

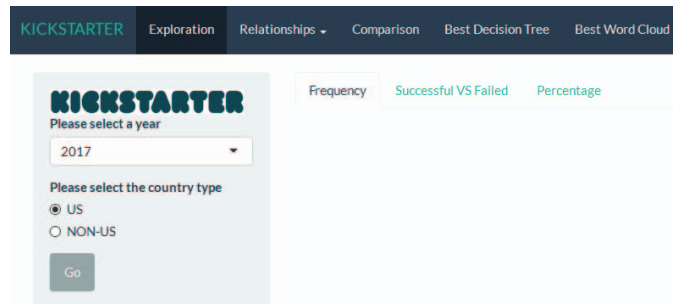


Fig. 6. Kick Start Your Success Dashboard



Fig. 7. ANOVA Results

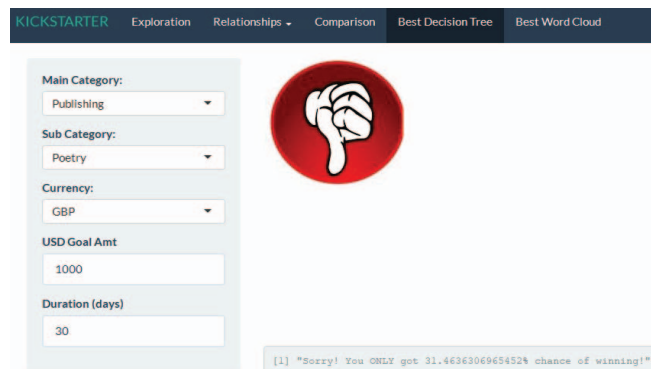


Fig. 8. Best Decision Tree illustration

The Best Word Cloud tab helps the users to identify popular words used in the titles of successful projects. Here, students apply some text mining techniques in order to generate the popular words. An example is shown in the Fig. 9 for Crafts category, where words such as “handmade” and “custom” are popular. The user can vary the number of words in the word cloud and the minimum frequency of the words such that only words which appear more than the minimum frequency will be displayed. For the last two tabs, students use techniques which are not covered in this course. They actually proposed them during discussions in order to gain more insights from the data. It proves that with PBL, the project itself is the learning.

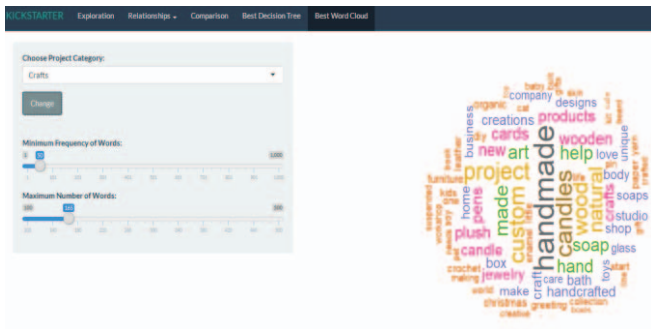


Fig. 9. Best Word Cloud Illustration for Kickstarter

### C. Airbnb

Students developed a web application that visualizes Airbnb data to assist users to know more about properties'

availability and their overall pricing throughout the year. Students added additional features, such as sentiment analysis and recommendation system, to provide more services to the users. The data is collected from <https://www.kaggle.com/>. The screenshot of the input page is shown in Fig. 10.

Fig. 11 lists the top 10 accommodations, which are located in Central London. The list is sorted according to a particular criterion, such as prices, number of reviews or scores. Some other explanatory statistical results, such as the distribution of property types and room types in London, can also be shown. The recommender system allows returning guests to decide quickly on the next best accommodation to stay in, and the sentiment analysis analysed the words used in reviews, which allows guests to identify if past guests had an overall good or bad experience with the property listing.

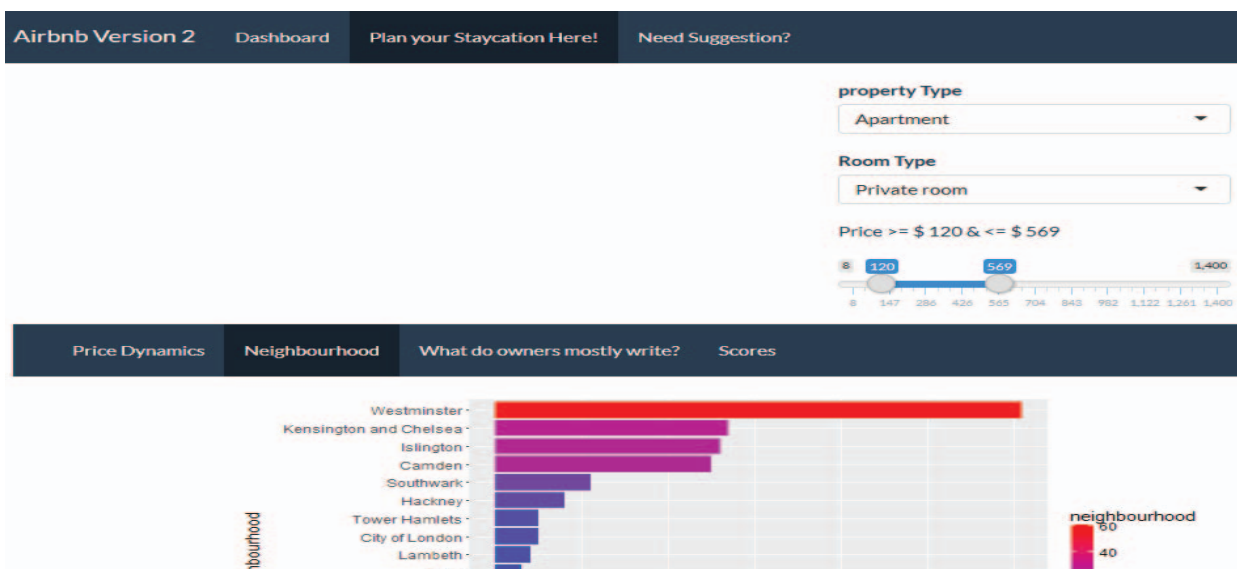


Fig. 10. Airbnb Dashboard

Airbnb Version 2								
Dashboard   Plan your Staycation Here!   Need Suggestion?								
property Type: Apartment								
Room Type: Private room								
Price >= \$ 120 & <= \$ 569								
Price Dynamics   Neighbourhood   What do owners mostly write?   Scores								
neighbourhood								
Westminster								
Kensington and Chelsea								
Islington								
Camden								
Southwark								
Hackney								
Tower Hamlets								
City of London								
Lambeth								
Brent								
neighbourhood score: 80 to 40								
Top 10 Listings   Overall View of Property   Inferential Statistics   Regression								
id	image	name	neighbourhood_cleansed	room_type	accommodates	price	number_of_reviews	review_sc
1		SOHO W1/2 THEATRELAND DESIGNER FLAT	Westminster	Private room	2	80	396	
2		You Will Save Money Here	Barnet	Private room	2	45	395	

Fig. 11. List of Accomodations

## V. INSIGHTS AND COURSE EVALUATION

### A. Insights and Findings

From the three projects described earlier, we can observe that students were able to apply and implement what they have learnt in class to analyse real data from real problems. Several groups have even gone beyond the topics and techniques covered in class, to include sentiment analysis, word cloud and recommender system. By implementing successful PBL, it shows that students were able to identify problems, apply their knowledge effectively, and even acquired learn-to-learn skill.

For the first round, there were 8 groups of students. We observed that 6 out of 8 groups (52.5%) explored and used techniques which were not covered in the course. For example, *Kick Start Your Success* and *Airbnb* projects included word cloud and recommendation systems. Word cloud is a novelty visual representation of text data [17]. This is commonly used in other courses, such as text analytics course. Recommendation systems is an important research area since they help users to find the right content, products or services [18]. This topic is part of another customer analytics course.

Table II summarizes the performance of three group projects on the individual components of the project assessment rubrics, based on the following criteria: Completeness of solutions (C1), Statistical concepts applied (C2), Innovative use of data (C3), Presentation delivery (C4), Quality of application (C5) and Peer assessment (C6). The score ranges from 0 – 10.

The final score is the average score of the six criteria. For these three projects, their final scores of 8.6 to 8.9 indicate that the students performed well in all criteria. Other group projects obtained lower final scores with the lowest of 7.2. All in all, our students were able to apply what they have learned by presenting and producing good quality applications. We also identified some local issues faced by students, especially related to the group projects: 1) to find authentic problems, and 2) time management among team members since students take different courses.

### B. Course Evaluation

At the end of the term, we conducted a survey to measure the success of the course, including the teaching techniques. The findings will provide us with directions to improve the course for future deliveries. Table III summarizes the course evaluation for Term 2 AY 2017-2018 with a total of 31 students who provided their feedback for five questions. Each question is scaled from 1 (extremely poor) to 7 (excellent). All questions have average values of above 5.0 which reflect good ratings. The highest average value is related to the quality and usefulness of assignments/projects. It is supported with the lowest values of standard deviation and Coefficient of Variance (CV).

TABLE II. TEACHING EVALUATION

Project	C1	C2	C3	C4	C5	C6	Final Score
<i>UK Accident Analysis</i>	7.5	9	9	9	9	8	8.6
<i>Kick Start Your Success</i>	9.5	9	9	9	9	8	8.9
<i>Airbnb</i>	9	8.5	9	8.5	9	8	8.7

Fig. 12 illustrates the total percentages of responses based on three different categories: 1-3 (extremely poor – poor), 4 (neutral) and 5 – 7 (good – excellent) for the five questions. In general, the histograms show a graphical representation of the students' responses on all five questions which were negatively skewed, meaning that the medians were higher than the means. This proves that there were only a few students who provided low scores compared to those who provided high scores. The low scores ranged from 3.23% to 12.90%. In general, at least 64.52% gave a value of five or higher on all questions. The highest percentage (80.65%) is for Q4 which is related to the projects and PBL.

By conducting interviews with students, we conclude that students develop skills through PBL. They learnt from their processes, group discussions and final products (statistical applications). They also learnt how academic work can connect to real problems and build their interest in applying statistics. They were able to reflect on how well they worked in a group via peer evaluation. By adopting PBL through group projects, students tend to be more motivated and self-directed. This aligns with the course objectives and expected course outcomes, such as to implement statistical computations using R, to create charts using R, to apply statistical knowledge in real problem and to build a statistical application using R studio. Students have demonstrated their learning through their projects.

During their presentations, students also welcomed other group members' ideas and suggestions. In addition to the five questions, the survey also asked the students to provide other information. Most of the students spent between 4 and 7 hours a week outside of class on this course, and they also felt that this course was very challenging. We list down several interesting comments from the students which are valuable for us to improve the course for future deliveries.

TABLE III. TEACHING EVALUATION

No	Question	Average	Std dev	CV
Q1	The learning experience in this course	5.172	1.284	24.83
Q2	The clarity of objectives and requirements	5.345	1.078	20.17
Q3	Quality and value of the course material	5.241	1.215	23.18
Q4	Quality and usefulness of course assignments/projects	<b>5.517</b>	<b>1.022</b>	<b>18.52</b>
Q5	Overall rating of the course	5.172	1.338	25.87



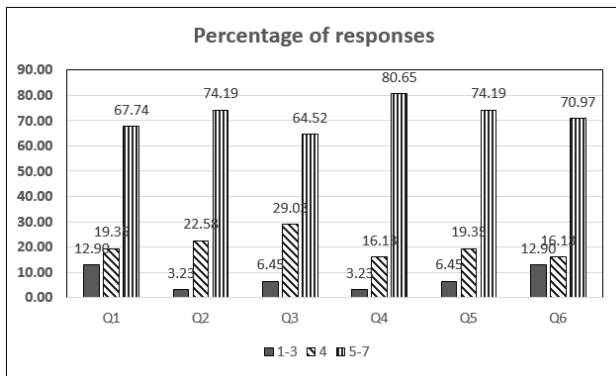


Fig. 12. Percentage Distribution of Students' Responses

We evaluate the effectiveness of the PBL by calculating the correlation value between the project grade and the final grade obtained by students. We found that the positive correlation is around 0.74 with significance value  $p\text{-value} = 0.035$ . The key point we take away is that the course structure is generally fine. However, some students still felt that the course covered too much theory and would prefer to have more content on practice in using the R Studio. Some comments from students are: *the course is very important for our field and it is nicely formulated, this course is very good because it leads me into the field of R, the course is very useful for us to do the statistical analysis and the course should teach, cover, and instruct more on the use of R for making our final project (fully functional shiny app).*

## VI. CONCLUSIONS AND FUTURE WORK

We presented a graduate level course, Applied Statistical Analysis using R (ASAR). ASAR covers the basic concepts of statistics and teaches students how to apply them in analyzing real data using a statistical software, the R console. In order to encourage students to solve real world problems, answer complex questions and generate high-quality results, we apply Project-Based Learning (PBL). Students choose a real problem with real data and develop their own applications in RStudio. Several interesting projects were also summarized. Student surveys showed that they were more motivated and displayed critical thinking skills in their projects.

Since this is a new course, there is ample room for improvement. The overall rating (Q5 in Table II) indicates that the average score is still lower than other questions. Some possible reasons that need further investigations could be related to facilities, contents, delivery methods, need to be conducted. The context of PBL can be extended to engage students. One possible active learning strategy is a flipped classroom. In order to ensure the effectiveness of this strategy, students projects should be linked to local issues which have not been explored. Therefore, we would like to extend the projects to the Singapore context so they can be implemented.

We suggest to conduct survey on the students' competencies both at the start and end of the term to measure the knowledge gain to draw more robust conclusions. More comprehensive approaches can be considered to evaluate the effectiveness of PBL, such as students' comments, assessment rubrics of the individuals in group projects, the correlation

between project assessment and the performance of the students on relevant topics. The course content also needs to be improved especially related to the content on R programming skill.

## REFERENCES

- [1] R. E. Kirk, "Changing topics and trends in introductory statistics," in Best practices for teaching statistics and research methods in the behavioral sciences, D. S. Dunn, R. A. Smith and B. C. Beins, Eds. NJ: Lawrence Erlbaum, 2007, pp. 25-44.
- [2] D. C. Montgomery and G. C. Runger, Applied Statistics and Probability for Engineers. Hoboken, NJ: Wiley, 2007.
- [3] L. Fawcett, "The CASE Project: evaluation of case-based approaches to learning and teaching in statistic service courses," Journal of Statics Education, vol. 25 (2), pp. 79-89, 2017.
- [4] R. E. Landrum and R. A. Smith, "Creating syllabi for statistics and research methods courses," in Best practices for teaching statistics and research methods in the behavioral sciences, D. S. Dunn, R. A. Smith and B. C. Beins, Eds. NJ: Lawrence Erlbaum, 2007, pp. 45-57.
- [5] J. Larmer, J. Mergendoller and S. Boss, Setting the standard for project based learning: a proven approach to rigorous classroom instruction. Alexandria, VA: ASCD, 2015.
- [6] J. Dunlap, "Changes in students' use of life long learning skills during problem based learning project," Performance Improvement Quarterly, vol. 18, pp. 5-33, 2005.
- [7] V. Boondee, P. Kidrakarn and W. Sa-Ngiamvibool, "A learning and teaching model using project-based learning (PBL) on the web to promote cooperative learning," European Journal of Social Science, vol. 21 (3), pp. 498 – 506, 2011.
- [8] Z. Zhang, C. T. Hansen and M. A. E. Andersen, "Teaching power electronics with a design-oriented, Project-Based Learning method at the technical university of Denmark," IEEE Transactions on Education, vol. 59 (1), pp. 32-38, 2016.
- [9] L. O. Seman, R. Hausmann and E. A. Bezerra, "On the students' perceptions of the knowledge formation when submitted to a Project-Based Learning environment using web applications," Computers & Education, vol. 117, pp. 16-30, 2018.
- [10] V. -L. Homes and Y. Hwang, "Exploring the effects of project-based learning in secondary mathematic education," The Journal of Educational Research, vol. 109 (5), pp. 449-463, 2016.
- [11] D. Gijbels, F. Dochy, P. Vanden Bossche and N. Segers, "Effect of problem based learning: a meta-analysis from the angle of assessment," Review of Educational Research, bol. 75, pp. 27-61, 2005.
- [12] I. Buj-Corral, L. Marco-Almagro, A. Riba, J. Vivancos-Calvet and X. Tort-Martorell, "Evaluation of project based learning in the area of manufacturing and statistics in the degree of industrial technology," Journal of Technology and Science Education, vol. 5 (2), pp. 130-140, 2015.
- [13] T. Koparan and B. Güven, "The effect of Project Based Learning on the Statistical Literacy levels of student 8<sup>th</sup> grade," European Journal of Educational Research, vol. 3 (3), pp. 145-157, 2014.
- [14] G. W. Cobb and D. Moore, "Mathematics, statistics and teaching," The American Mathematical Monthly, vol. 104, pp. 801-823, 1997.
- [15] M. Núñez-del-Prado and R. Gómez, "Learning data analytics through a Problem Based Learning course", Proceedings of 2017 IEEE World Engineering Education Conference, 19-22 March 2017, Santos, Brazil.
- [16] Y. Yesilcay, "Research project in statistics: implications of a case study for the undergraduate statistics curriculum," Journal of Statistics Education, vol. 8 (2), 2000  
<http://ww2.amstat.org/publications/jse/secure/v8n2/yesilcay.cfm>.
- [17] D. Helic, C. Trattner, M. Strohmaier and K. Andrews, "Are tag clouds useful for navigation? A network-theoretic analysis," Journal of Social Computing and Cyber-Physical Systems, vol. 1 (1), pp. 33-55, 2011.
- [18] D. H. Park, H. K. Kim, I. Y. Choi and J. K. Kim, "A literature review and classification of recommender system research," Expert Systems with Applications, vol. 39 (11), pp. 10059-10072, 2012.