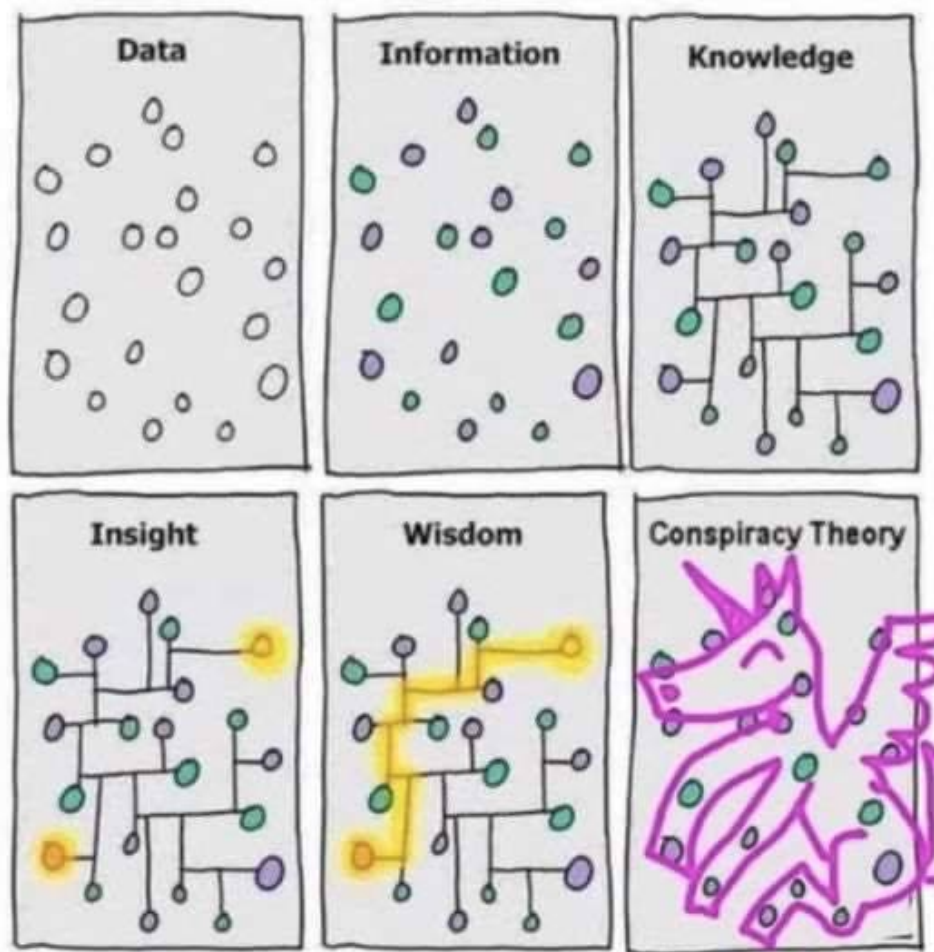


What Should you do in the Project?

You produce knowledge using the given data sets. Of course, developing data insights, professional data visualizations, and appropriate statistical models supported by simulations are the minimum requirements. However, producing knowledge in the domain of your project should be your goal. The project is open to explore ideas in any directions, you are free to design your research questions. We do not cover theory behind the statistical techniques we use in this class. We focus on applications. However, I want you to come up with a supporting evidences that justifies your arguments in your research questions via data visualizations, tabulations, hypothesis testing, and simulations. We will discuss more on hypothesis and statistical modeling during the month of April. Please understand that the best projects are those which present creativity, and independent learning beyond the syllabus. The following figure summarizes the type of work you need to do. Choice is yours!!



How do you want to present your data?

Precautions !!

1. Use Rmarkdown to do this project. Do not copy and paste the results from R or any other software to a word document.

2. Restrict yourself to five(maximum!) best tables and five best visualizations. A large number of figures do not necessarily mean that your project is better. Do not include figures or tables in the document without any reasoning behind it. If you want more than five figures or tables then create a new section called appendix at the end of your report and refer to those figures in the main document.
3. No figures and tables should be left without explanations. A paragraph (minimum!) of explanation for each graph or a table is a norm.
4. Codes, software messages, and software warnings **must** be avoided from the main body of the report. You can design a section called “appendix” (after references) to include R-codes, and other supporting arguments.
5. The main report (excluding references and appendix) should not exceed more than 15 pages. Please note: do not dump all kinds of diagrams and tables in the manuscript. In my experience, length of the report and its quality are not necessarily directly proportional.

Guidelines to write the report

1. Introduction: You must have to write about the source/s of the data, describe the variables used in the project, and clearly state your research question/s and explanations of the statistical models used for the project. Also, you need to report if other people also have done the same/similar work. Please note, you need to explain your proposal and describe any technicalities regarding the domain of your project. For example, if you are doing project in basketball analytic- do not assume that your reader has a lot of knowledge in basketball. The length of the introduction for this project should not be more than three pages.
2. Visualization: What kind of scientific questions we can address through these plots?
You need to propose research questions that your graph(s) can answer. Your graphs should raise legitimate scientific question(s). Most importantly, for full credit, you need to create new variables/information using the given variables. No one will/should tell you what to do in a research question. The questions are completely open. It is impossible to show creativity by plotting generic plots for the given sets of variables. Please note, you will NOT receive good grade by plotting given variables without data manipulation or transformations. Writing your own function(s) to create new knowledge will receive *brownie points*.
3. Descriptive statistics: In this section, you present the data in different type of tables, diagrams and find all possible descriptive statistic (for example: mean, median, mode, standard deviation, quartiles, etc.) You need to explore the data using as many different graphs and tables as possible. Choose the most relevant and informative graphs and tables for the report.

An example of descriptive statistics for the quantitative variable is given below. However, feel free to design a table of your own style to reveal the information behind the data. This is one of the most basic example of table. I would appreciate if you can find better way to present your information in different styles.

variables	minimum	Q1	median	Q3	mean	Variance	Standard Deviation
variable 1							
variable 2							
variable 3							
variable 4							

In addition to the descriptive statistics, table-you need to use different visualization techniques to formally define your research question(s). Let me reiterate: research question should be supported and justified by visualizations, tables, hypothesis testing, and simulations.

4. Analysis : I expect you to show a good understanding of the statistical procedure or simulation that you decide to use. Make sure to explain every methodologies and simulation procedures. Please note, I am not interested to see basic linear regression model or multiple regression models in this class. You would benefit yourself by writing functions, running simulations, creating stunning visualizations or running predictive models using machine learning techniques (regression tree, random forest, or neural network). Project is something that should reveal the understanding of the material covered in the class and take it to a different level.
5. Discussion: This is the place where you show your understanding of the project. Write a clear description of the results from the previous sections. Do not forget to address the research question/s that you posed in the introduction section. Research is not only about answering questions, it is also about posing relevant questions. Make sure to pose possible questions based on graphs, tables, and other inferential results. The quality of your work will also be evaluated by the strength of the questions (answered or not answered).
6. References : List all the resources (books, journals, web sites, videos etc.) that you used in the project as per the order of their appearances in the report. A proper list of references is a necessary *cheery on top* of your report.

Working together:

Each group is assigned to work on one project. Make sure to distribute your work and mention your contribution at the end (not in the body of the report). Your grade is based on the your contribution. This project is designed as a model of working together but graded separately. Make sure to to help each other. If one part of the project is weaker then the quality of the whole project will go down. Also, there should be a good compatibility and connection between your work and your partners work. It should not feel like it is written by multiple authors unless we see the contributions at the end. The project is one solid report written by multiple authors in harmony and rhythm.

Project Competition

The Undergraduate Class Project Competition (USCLAP) is organized by the American Statistical Association. I highly encourage every group to plan submitting their project to this competition. If you plan to submit in this conference- then you need to find data sets that are

not used by others or do something original that is not done previously by other researchers. Please find below some detailed information about the conference, sample projects, recorded videos and submission details.

<https://www.causeweb.org/usproc/usclap>.

Originality

You can not use the tables created by any other agencies or publications. All the tables, figures, and explanations in the project should be created by your group with accompanying R-code. If some figures, tables, or explanations from the published resources are very supportive to your arguments then you can use them by providing credit to the original source.

(Mini Project Presentation: March 30, in class)

(Project Report Due: April 28, 11:59PM)

(Final Presentation, April 25, 3:00 PM-6:00 PM)