**CSE – 564 LAB -1 Documentation**

**Dataset Sources:**

* US College Dataset (<https://www.kaggle.com/yashgpt/us-college-data>)
* US States Dataset (<https://www.kaggle.com/giodev11/usstates-dataset>)

**Attributes Description:**

1. College: Names of 777 Universities/Colleges in United States
2. State: The state in which the University/College is Located
3. Popularity: Famous field of study a specific University/College
4. Type: Whether it is a Private or a Public University
5. Applications: Average number of applications received in a year
6. Accepted: Average number of applications that are accepted
7. Enrolled: Average number of students enroll in a year
8. Top10perc: The percent of new students who are from Top 10% in High School
9. Top25perc: The percent of new students who are from Top 25% in High School
10. Full Time: Number of Full-Time Undergraduates
11. Part Time: Number of Part-Time Undergraduates
12. Outstate: Out-of-state tuition fee cost
13. Room Board: Room and board cost
14. Books: Estimated book costs
15. Personal: Estimated personal spending
16. PhD: Percent of Faculty with Ph.D.’s
17. Terminal: Percent of Faculty with Terminal Degree
18. PercAlumni: Percent of Alumni who donate
19. Expend: Instructional expenditure per Student
20. Graduation Rate: The Graduation Rate be University/College

**Why Interesting?**

* I found this data interesting because I always wanted to make certain visualizations for the data that is related to colleges and do proper analysis on it
* The data here is closely related to the statistical calculations I performed back in India while applying for the U.S Universities to go to one of the best Universities for Computer Science (Which helped me get into one)
* It has a good amount of data to visualize on a particular study. For example, one of a study is between Graduation Rate of the students to the Percentage of Faculty who hold a PhD. This depicts how good the quality of education is in a particular University/College
* More attributes can be added to the data in the future as it is flexible and, I can use the same dataset in my final project
* We can find a proper relation between the number of applications to the students enrolled or accepted by the University

**Noteworthy Points:**

* The Toggle button is used for changing the X and Y axes by rotating them 90 degrees
* The Default button is used to get back to the original position by clicking on it
* Note that the Toggle button doesn’t work for the Scatter Plot (I-e if the Y-Axis button is checked). You can also see a pop-up showing the same.
* The Reset Button is just below the axis’s points selection. You can reset the graph anytime by clicking on it.
* To draw a scatter plot, you just need to select the Y-Axis so that the attribute for y-axis gets selected and the plot is drawn.
* I used Bins to arrange the histograms in a particular value range
* In scatter plot for categorical values, I used the Jittering effect as seen in the below screenshots

**Screen Shots:**

**Chart, histogram

Description automatically generated**

**Chart, histogram

Description automatically generated**

**Chart

Description automatically generated**

Graphical user interface

Description automatically generated

Chart, scatter chart

Description automatically generated