

# ES114 Assignment 2- Data Narrative

Name: Naimkhan Shaikh

Roll no. :22110239

Discipline: Chemical Engineering

## I. OVERVIEW OF THE DATASET

This collection includes details on 777 American colleges and institutions. It was utilized during the 1995 Data Analysis Exposition, which was supported by the American Statistical Association's Statistical Graphics Section. The American Association of University Professors' 1994 Salary Survey and the U.S. News & World Report's Guide to America's Top Schools are the two sources for the dataset. Variables in the dataset include tuition, room and board costs, SAT or ACT scores, application/acceptance rates, graduation rates, student-to-faculty ratios, spending per student, average salaries, overall compensation, and the number of faculty.

## II. SCIENTIFIC QUESTIONS/HYPOTHESES

*A. Is there a relation between a college's acceptance rate and its average SAT scores?*

*B. Is there a relation between a college's acceptance rate and its graduation rate?*

*C. Are colleges with higher percentages of part-time students more likely to have lower graduation rates?*

*D. Is there a relation between the graduation rate of a college and the percentage of alumni who donate to the college?*

*E. Are public universities likely to be cheaper to attend than private universities? Does this mean that the cheaper one compromises in terms of facilities?*

*F. Are professors in research universities paid more than those in other universities?*

*G. What is the variation in the average salaries of professors from state to state? Is there any particular reason for it?*

*H. Are professors in colleges with lesser acceptance rates more likely to be paid more than other colleges with higher acceptance rates?*

*I. How does the distribution of salary vary across full profs, associate profs and assistant profs?*

*J. Are institutions with higher student-to-faculty ratios more likely to pay their faculty members less?*

*K. Do research universities provide more benefits to their faculty members than other universities?*

## III. DETAILS OF LIBRARIES AND FUNCTIONS

This contains the details of the libraries and functions that were used.

### A. Libraries:

- matplotlib: matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. <sup>1</sup>
- seaborn library: Seaborn is a Python data visualization library based on matplotlib. It

provides a high-level interface for drawing attractive and informative statistical graphics. <sup>2</sup>

- pandas library: pandas is an opensource library which provides high-performance data structures and data analysis tools. It has been released under the three-claused BSD license. <sup>3</sup>

### B. Functions:

- Pandas.read\_csv(): To read and store data from csv file to a Pandas dataframe.
- x.corr(y): Correlation between x and y.
- Series.mean(): returns the mean of values of a Series
- seaborn.boxplot(): plots the boxplot of the data given to it.
- seaborn.barplot(): plots the bar plot of the data given.
- Pandas.DataFrame(): Creates a dataframe of a dictionary
- seaborn.jointplot(): Plots two variables with bivariate and univariate graphs <sup>4</sup>
- seaborn.boxplot(): Plots a boxplot of the data given.
- Axes.set\_xticks(): Set the x-axis' tick locations and optionally labels.
- pandas.DataFrame.groupby(): Group DataFrame using a mapper or by a Series of columns. <sup>5</sup>

#### IV. ANSWERS OF THE QUESTIONS

A. *Is there a relation between a college's acceptance rate and its average SAT scores?*

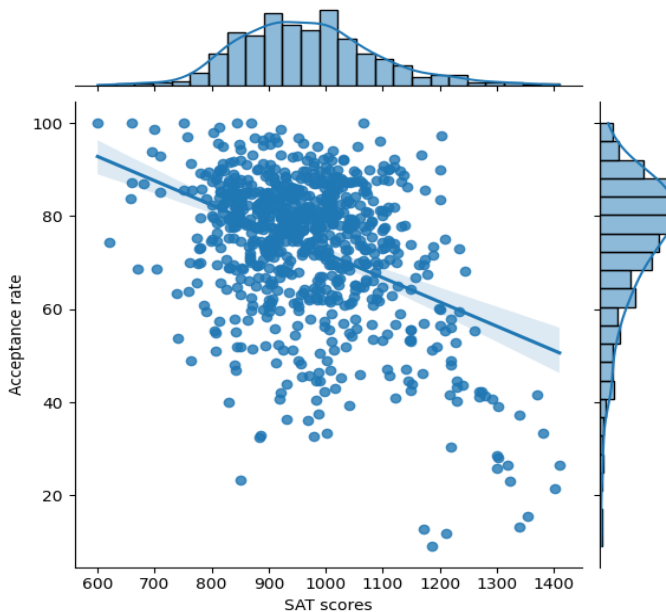


Fig.1 Scatterplot of acceptance rate of a college vs its average SAT score

None of the colleges with an average SAT score of 1200 or above have an acceptance rate above 40%. The correlation coefficient between the acceptance rate of a college and its average SAT score is -0.417, which suggests a moderately strong correlation between them. The negative sign suggests that as the average SAT score of a college increases, the acceptance rate decreases and vice-versa.

B. *Is there a relation between a college's acceptance rate and its graduation rate?*

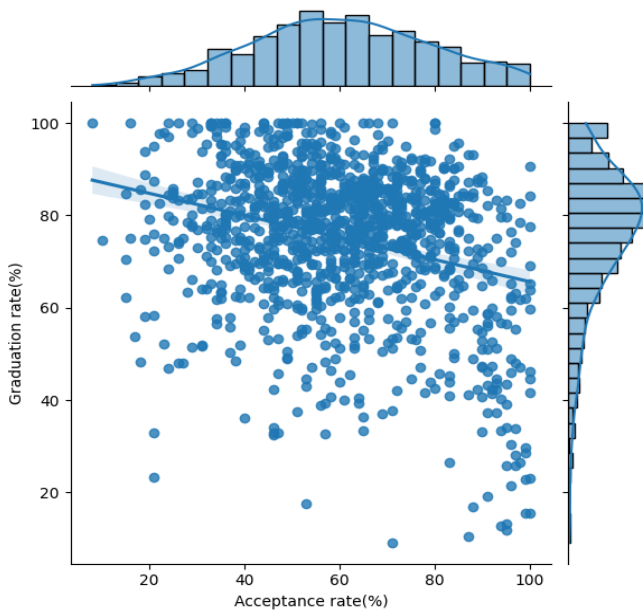


Fig.2 Scatterplot of graduation rate of a college vs. the acceptance rate of a college

The correlation coefficient here is -0.279, which suggests an inverse relation between the graduation rate and acceptance

rate. The value of 0.279 suggests that the relation between these two variables is weak. The graph is very dense between 40-60% acceptance rate and around 80% graduation rate and is less dense in other parts. There are other factors which affect the graduation rate more than the acceptance rate. Thus, we can conclude that as the acceptance rate increases, the graduation rate tends to decrease slightly and vice-versa.

C. *Are colleges with higher percentages of part-time students more likely to have lower graduation rates?*

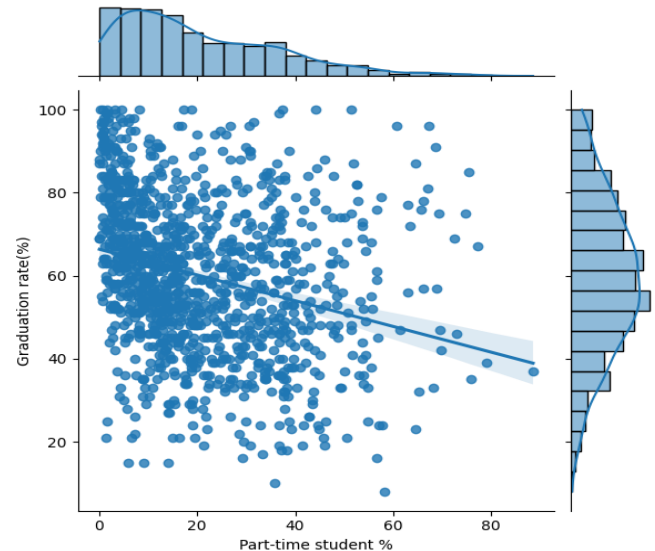


Fig.3 Scatterplot of graduation rate vs no. of part-time students

We can observe from the above graph that the graph is denser in the region where the graduation rate lies between 0-20, and the % of part-time students is between 0-20%. In other places, the graph is less dense. The correlation coefficient between the graduation rate and the percentage of part-time students is -0.27. This indicates that the relation between the two variables is weak. The negative sign is indicative of an inverse relation between the number of part-time students and the graduation rate.

Thus, we can conclude that as the no. of part-time students increases, the graduation rate of a college decreases slightly.

```
e=usnews.parttime_ugs
f=usnews.fulltime_ugs

x=[]
y=[]

for i in range(len(a)):
    if e[i]!='*' and f[i]!='*' and d[i]!='*' and int(d[i])<=100:
        x.append(((int(e[i]))/(int(e[i])+int(f[i])))*100))
        y.append(int(d[i]))
        j+=1

x1=pd.Series(x)
y1=pd.Series(y)
print("Correlation coefficient: ",x1.corr(y1))

data1 = pd.DataFrame({'Part-time student %': x, 'Graduation rate(%)': y})
sns.jointplot(data=data1, x='Part-time student %', y='Graduation rate(%)',kind='reg')
plt.show()
```

Fig. 4 Code snippet for Question3

D. Is there a relation between the graduation rate of a college and the percentage of alumni who donate to the college?

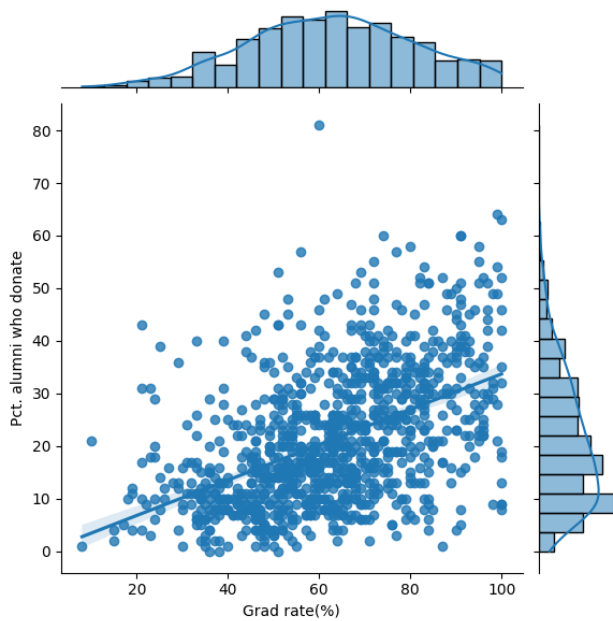


Fig. 5 Scatterplot of % of alumni who donate vs graduation rate

The correlation coefficient between the percentage of alumni who donate back and the graduation rate is 0.484. This implies a moderate positive correlation between the two variables. As the graduation rate increases, the percentage of alumni who donate back also increases.

E. Are public universities more likely to be cheaper to attend than private universities? Does this mean that the cheaper one compromises in terms of facilities given to the students?

TABLE I.

Parameter		Public University	Private University
Tuition (excluding additional costs)	In-state	5780.578	15654.516
	Out of state	9759.3021	15684.809
Average student-faculty ratio		17.651	13.354
Average % of faculty with a terminal degree		77.987	74.592
Instructional expenditure/student		6876.826	10064.064

We can observe from the table that public universities are less expensive than private universities. There is not much difference in the tuition fees of in-state and out-of-state private universities. On the other hand, there is a significant difference in the in-state and out-of-state fees of public universities.

However, despite being less expensive to attend, public universities provide a higher mean student-to-faculty ratio and have more faculty with a terminal degree.

Private universities do have more instructional expenditure per student though.

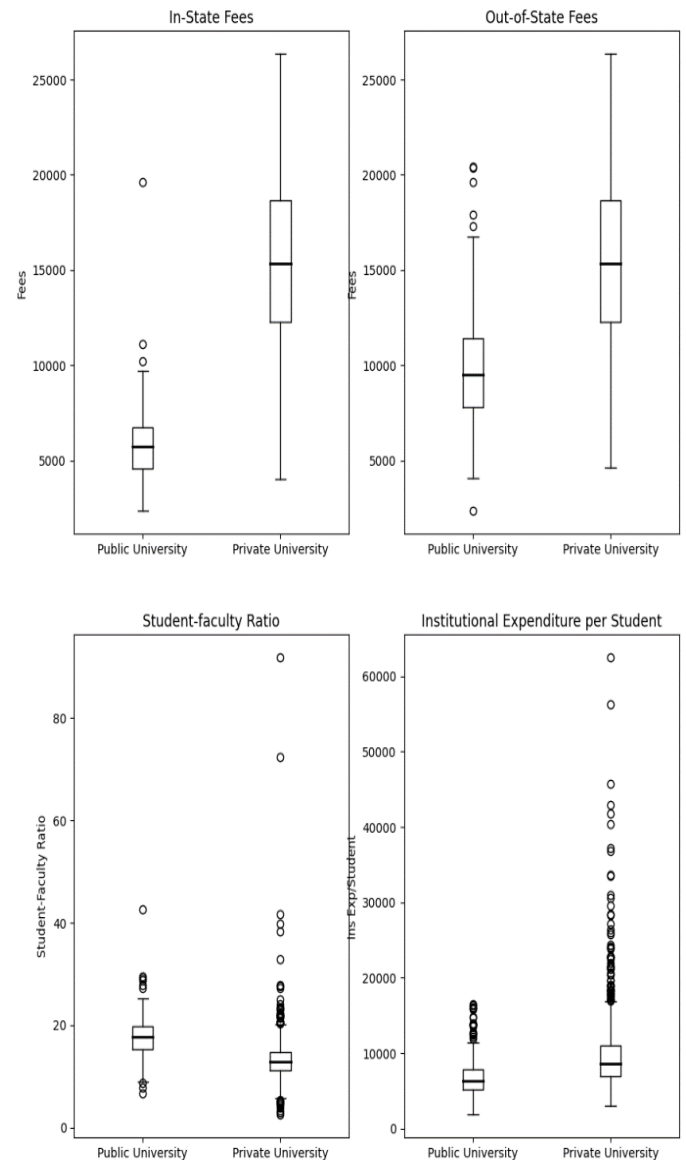


Fig. 6 Various boxplots comparing public universities to private universities

*F. Are professors in research universities paid more than those in other universities?*

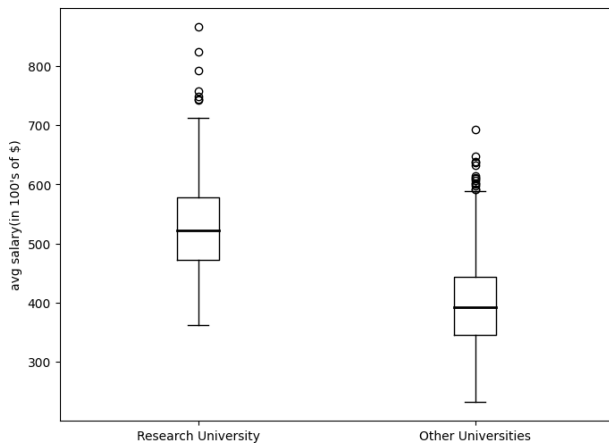


Fig. 7 Boxplot comparing the salaries of professors from research and non-research universities

We can observe from the above boxplot that professors who teach in research universities are paid more than professors who teach in non-research focused universities. The mean average salary of professors teaching in research universities is around \$53366.67, and the mean average salary of professors teaching in other universities is around \$39965.88.

*G. What is the variation in the average salaries of professors from state to state? Is there any particular reason for it?*

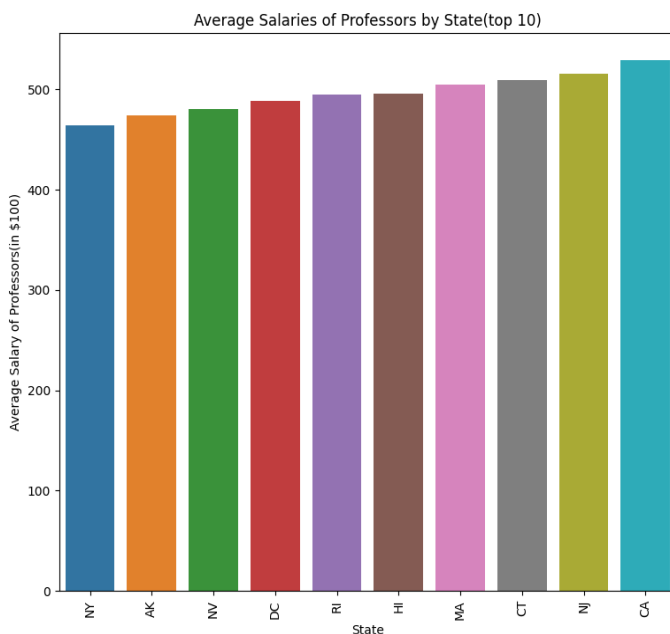


Fig. 8 Graph comparing top 10 mean salaries of professors in different states

We can observe from the above graph that the state with the code 'CA'(California) has the highest salary paid to professors. This may be because of high cost of living in the state, along with the presence of several prestigious

universities in the state, including the likes of Stanford and Caltech.

*H. Are professors in colleges with lesser acceptance rates more likely to be paid more than other colleges with higher acceptance rates?*

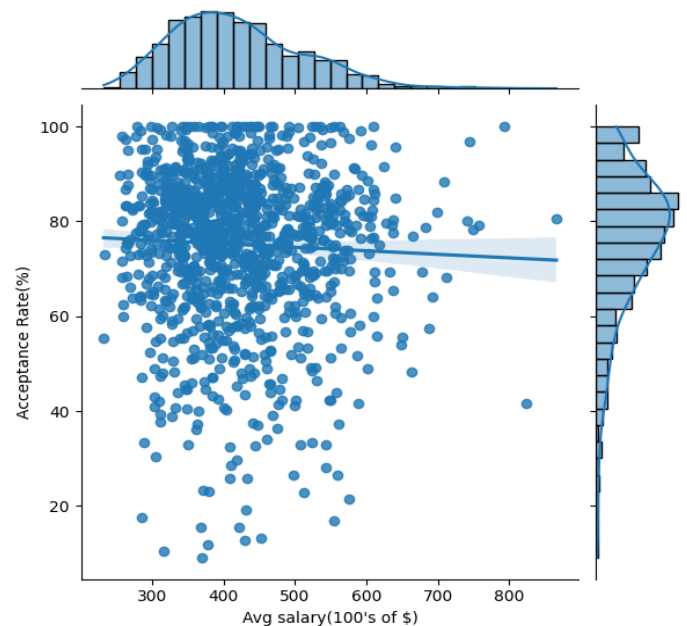


Fig.9 Scatterplot of acceptance rate of a college vs average salary of a professor

We can observe from the above graph that this is true to a very small extent. The correlation coefficient here is -0.042, which is very small. This implies a very weak (almost none) negative correlation between the two variables.

*I. How does the distribution of salary vary across full profs, associate profs and assistant profs?*

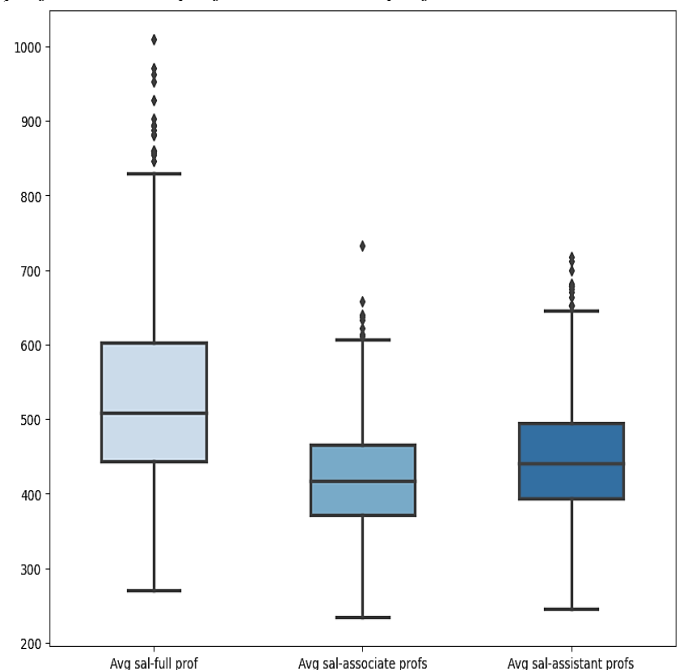


Fig. 10 Boxplot comparing salaries of full professors, associate professors and assistant professors

We can observe from the above graph that a full professor gets paid the highest, followed by assistant professors and followed by associate professors. The mean salaries of each of them are as follows:

Mean full profs avg salary: \$ 52644.734

Mean associate profs avg salary: \$ 42001.025

Mean assistant profs avg salary: \$ 44639.981

*J. Are institutions with higher student-to-faculty ratios more likely to pay their faculty members less?*

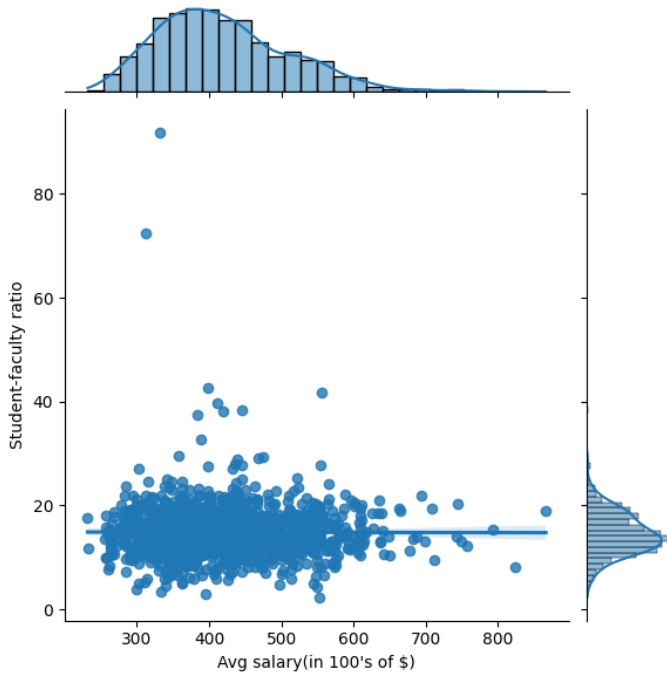


Fig. 11 Scatterplot of student-faculty ratio vs average salary of a professor

The correlation coefficient here is -0.0035. This implies a very weak negative correlation between the student-faculty ratio and the average salary of professors. Thus, we can conclude that the student-faculty ratio has a negligible effect on the average salary of a professor.

*K. Do research universities provide more benefits to their faculty members than other universities?*

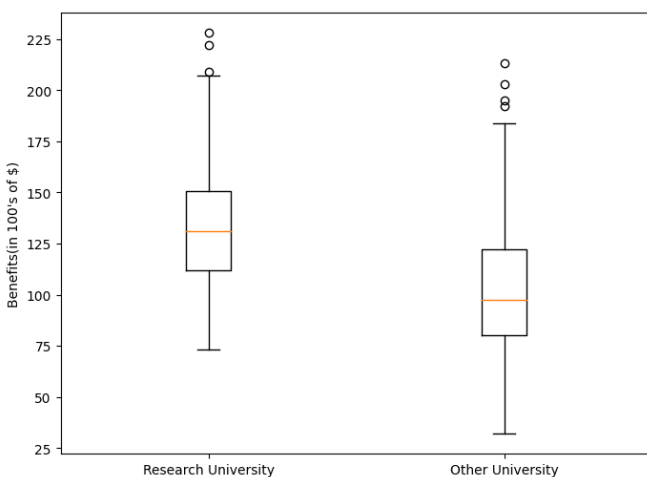


Fig. 12 Boxplot showing benefits to professors of research university and professors of other university

We can observe in the above graph that the benefits received by a professor in a research university are more than a professor in a non-research-focused university. We calculate the benefits received by the professors by subtracting the average salary from the average compensation. The mean average benefits of the professors are as follows:

Mean benefits to research university: \$13192.78

Mean benefits to other university: \$10162.14

## V. SUMMARY OF OBSERVATIONS

1. There is a moderately strong relation between a college's acceptance rate and its average SAT scores, with acceptance rates decreasing with an increase in average SAT scores.
2. As acceptance rates increase, the graduation rates tend to decrease slightly due to a weak negative correlation between them.
3. As the number of part-time students increases, the graduation rate of a college decreases slightly. This is due to a weak negative correlation between them.
4. With an increase in the graduation rate, the percentage of alumni who donate back also increases. This is due to a moderate strong positive correlation between the two variables.
5. Public universities are cheaper to attend than private universities. There is a major difference in the in-state and out-of-state tuition fees of a public university. However, this does not necessarily mean that public universities compromise on facilities, with both the mean student-faculty ratio and average percentage of faculty with a terminal degree being higher. Private universities do have a higher instructional expenditure per student though.
6. Professors in research universities are paid more than professors in non-research focused universities.
7. There is a reasonable amount of variation in the salaries of professors from state to state. The state with the highest mean salary of \$52492.592 is California(CA), while the state with the lowest mean salary of \$33893.333 is WV(West Virginia). This can be attributed to a number of factors, like the cost of living, competition from other prestigious universities in the same state etc.
8. Professors in colleges with lower acceptance rates are very likely to be paid the same as other professors who teach in colleges with higher acceptance rates.
9. Full professors get paid the highest, followed by assistant professors and then followed by associate professors.
10. The student-faculty ratio has a negligible effect on the average salary of a professor.
11. Professors teaching in research universities receive more benefits(nearly \$3000 more) than professors who teach in non-research focused universities.

## REFERENCES

1. Matplotlib documentation — Matplotlib 3.7.1 documentation. “Matplotlib Documentation — Matplotlib 3.7.1 Documentation,” n.d. <https://matplotlib.org/stable/index.html>.
2. seaborn: statistical data visualization — seaborn 0.12.2 documentation. “Seaborn: Statistical Data Visualization — Seaborn 0.12.2 Documentation,” n.d. <https://seaborn.pydata.org/>.
3. seaborn: statistical data visualization — seaborn 0.12.2 documentation. “Seaborn: Statistical Data Visualization — Seaborn 0.12.2 Documentation,” n.d. <https://seaborn.pydata.org/>.
4. seaborn.jointplot — seaborn 0.12.2 documentation. “Seaborn.Jointplot — Seaborn 0.12.2 Documentation,” n.d. <https://seaborn.pydata.org/generated/seaborn.jointplot.html>.
5. pandas.DataFrame.groupby — pandas 1.5.3 documentation. “Pandas.DataFrame.Groupby — Pandas 1.5.3 Documentation,” n.d. <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.groupby.html>.

## ACKNOWLEDGEMENTS

I would like to thank Mr. Shanmuga R, my professor for giving me the opportunity to do this project. I would also like to thank my TA, Mr. Madhav, and the instructor, Mr. Atul for helping me in this project.