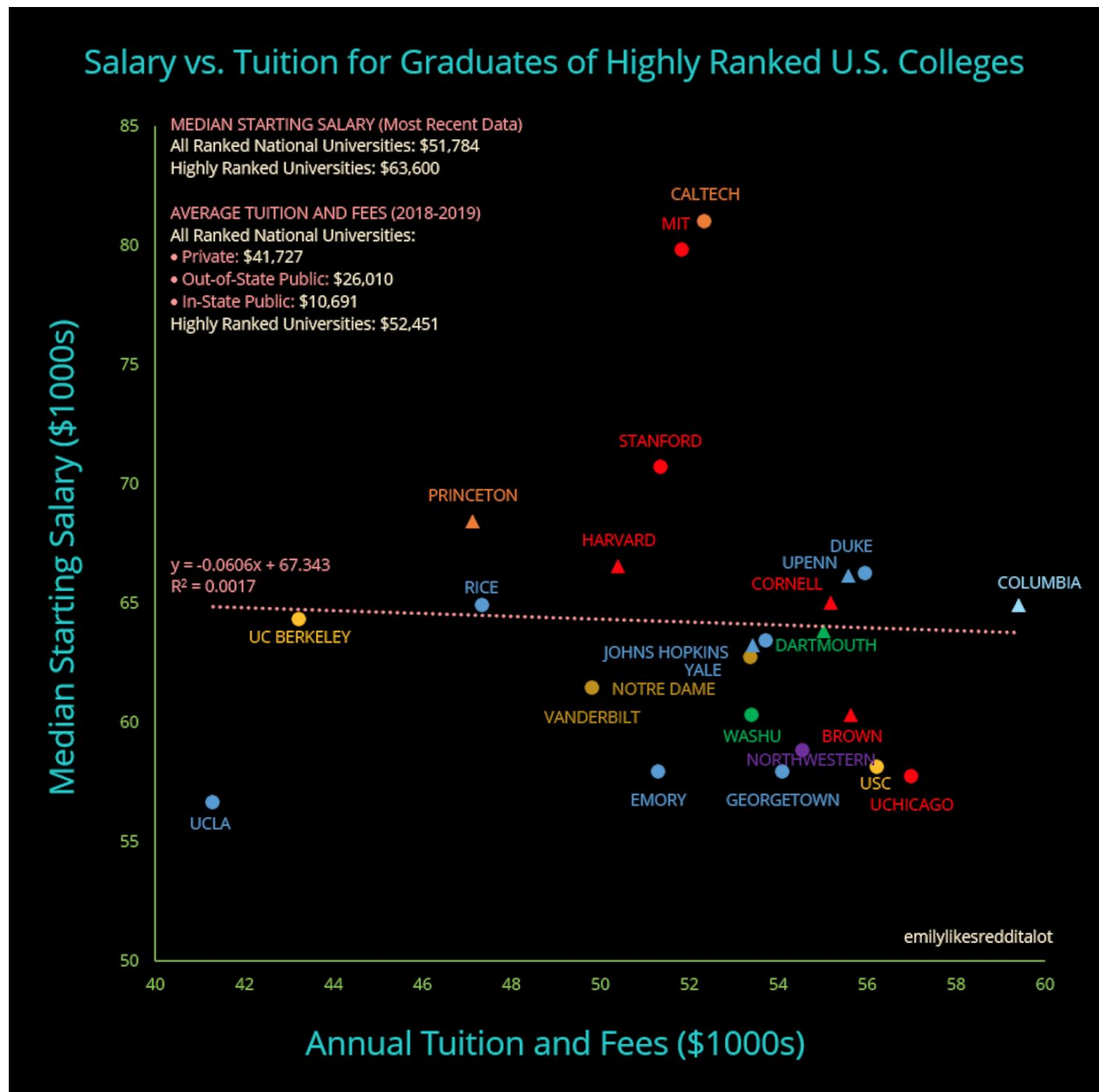


Part 1: Visualization critique (5 points)

I selected a data visualization from [r/dataisbeautiful](https://www.reddit.com/r/dataisbeautiful/), and it is a plot of salary vs. tuition for graduates of highly ranked U.S. colleges¹ (Personally, this plot attracted my eyes because where UChicago lies in the graph...).



¹ Salary vs. Tuition for Graduates of Highly Ranked U.S. Colleges
https://www.reddit.com/r/dataisbeautiful/comments/btm776/salary_vs_tuition_for_graduates_of_highly_ranked/

Before getting into my critiques to this graph, I would like to first introduce how this graph was created. According to the author, all the data she used came from U.S. News and World Report. The top 22 universities were included based on U.S. News' 2019 Best National Universities list, but, in fact, 24 schools were represented because of ties. The data points in this plot were colored according to the official university colors, and triangles represent Ivy League universities. All data, including starting salaries, tuitions, and school rankings, were just for bachelor's degrees. Graduate programs were not involved.

1. Is it truthful?

According to Cairo (p.46), a truthful visualization should be clear and sincere. At least at first glance, this plot is truthful as it satisfies Cairo's definition. All the axes were clearly labeled, and the author explained the sources of data and provided the original Excel file she used to create the plot. This extra information increased the truthfulness of the plot.

I especially appreciate that the author included the information about the median starting salary and the average tuitions and fees at the top left corner. Along with other supplement information offered by the creator of this plot, these mean and median numbers give me a baseline to make comparisons. However, I also think that there is a better way to represent these extra pieces of information on the plot so that it could be visually more straightforward and hence more truthful. For example, the median starting salary and mean tuitions could both be represented using labeled straight lines on this plot. Similarly, the author could also include a legend in the plot to explain why specific marker colors and shapes were used for different schools. Although the main body of the plot is truthful and displays necessary information clearly, these minor modifications could prevent confusions and enhance the clarity of the plot.

2. Is it functional?

Cairo claimed that a data visualization is functional if the audience could accurately interpret the information that it is designed to convey (p. 50). Overall, I think that this is a functional plot. The author aims to visualize the relationship between top U.S. colleges' annual tuitions and their graduates' starting annual salaries. Students could refer to this plot to get a rough sense of the investment/return ratio for college education in different universities.

It is interesting to note that the author included a least squares regression line (the pink dashed line) in the plot. Normally, I would think that having a regression line in a scatterplot actually makes the plot more functional and truthful. The regression line helps to better identify and understand the linear relationship between the two variables displayed in the graph. However, I do not think it is true in this particular case. As clarified by the author herself, the downward sloping regression line does not imply

that there is a negative correlation between tuitions and starting salaries, because this simplistic plot failed to take many confounding variables into account. Therefore, I think having the regression line may actually mislead people to make them draw inaccurate conclusions about school tuitions and starting salaries.

3. Is it beautiful?

Cairo evaluated whether a visualization is beautiful mainly from the aesthetic perspective. He argued that the visualization had to be elegant and clear (p. 53-54). According to this definition, I think that this plot is clear and does not contain any unwanted information that may mislead or distract the audience. Different elements in this plot (e.g., axis, labels, extra information, regression lines, and data points) were displayed using different colors so that they are visually distinguishable at first glance. However, not everyone is comfortable with this black background design. In particular, this design may not be suitable for academic use.

As I mentioned earlier, adding a legend to explain the colors and shapes of school data points may make this plot more clear. For me, personally, I was confused about what different colors represent at the first time I saw the plot. Moreover, representing schools using their official colors is not necessary and is not helpful for the audience's understanding. I originally thought that different colors might represent different rankings or geographical locations of these universities, but I realized that I was misled after reading the author's notes. I think that the author could better utilize the colors of these data points to convey more useful information.

4. Is it insightful?

Cairo described two types of insights in his book (p. 59). A visualization gives "spontaneous insights" if it is usually intuitive to the audience at the first look, and a visualization has "knowledge-building insights" if the audience could gain more knowledge through the gradual exploration of it. Because the information that the plot is trying to convey is quite straightforward, I think that it is spontaneously insightful but does not have plenty of "knowledge-building insights". It is mainly due to the nature of this plot, and I do not think that it is a shortcoming of this plot. The purpose of this plot is to give a straightforward comparison between top U.S. colleges, and this plot has done its job well.

However, I also acknowledge that further exploration of this plot could lead us to think about many other explaining factors. But these explaining factors would further weaken the relationship between tuitions and starting salaries that the author was trying to display using this plot. It is even debatable whether these schools could be compared directly. First, the author does not plot public and private colleges separately. Considering there could be huge gaps between the tuitions of public and of private colleges (also between out-of-state and in-state students), simply plotting all colleges in one graph

may not be very informative. Second, certain universities are more famous for their graduate programs while others have better undergraduate programs. Thus, comparing the bachelor's degrees of all top universities according to their rankings may be imprudent.

5. Is it enlightening?

I think that this scatterplot, though simple, is enlightening. First, it brings up a crucial question regarding the current U.S. education system: whether the investment in high-level education gives the family or the student a fair return after graduation. Although this simplistic scatterplot may not be capable of answering this question, it does make the audience to reconsider the meaning of high-level education and whether entering a top-ranking, prestigious school ensures well-paid jobs after graduation.

Moreover, as I illustrated briefly in the earlier section, this plot enlightens the audience in a way that we could come up with many other explanations to this distribution of colleges in the scatterplots. The type of universities, the geographical locations, the fields of study that these universities are famous for, and different school cultures could all lead to different starting salaries of their graduates. In addition, undergraduate students in some universities tend to choose to work directly after graduation, while those in other schools tend to continue to graduate schools. That variation could also bias the starting salaries of these schools.