

Analysis Report about Table Lens

1. The Data Set: colleges

1.1. The introduction to the data set

This data is downloaded from the StatLib library, which is maintained at Carnegie Mellon University. The original dataset contains over 1300 schools, but the number in this visualization only 240 considering the performance.

Here is their descriptions about this data:

1995 Data Analysis Exposition sponsored by the Statistical Graphics Section of the American Statistical Association. The U.S. News data contains information on tuition, etc., for over 1300 schools, while the AAUP data includes average salary, etc.

Information about the dimensions:

- College: college name
- State: State(postal code)
- Type: (1, 2A, 2B)
- avgSaFP: Average salary - full professor
- avgSaOP: Average salary - associate professor
- avgSaIP: Average salary - assistant professors
- avgSaAR: Average salary - all ranks
- avgCoFP: Average compensation - full professors
- avgCoOP: Average compensation - associate professors
- avgCoIP: Average compensation - assistant professors

- avgCoAR: Average compensation - all ranks
- numFP: Number of full professors
- numOP: Number of associate professors
- numIP: Number of assistant professors
- numIN: Number of instructors
- numFA: Number of faculty - all ranks

NULL proprieties in the data set was replaced by the average attribute of the corresponding propriety.

1.2. The analysis of the data set

1.2.1. The basic analysis of the data set

Firstly, we analyse the data set by coloring and sorting.

- Coloring according to the state attribute.

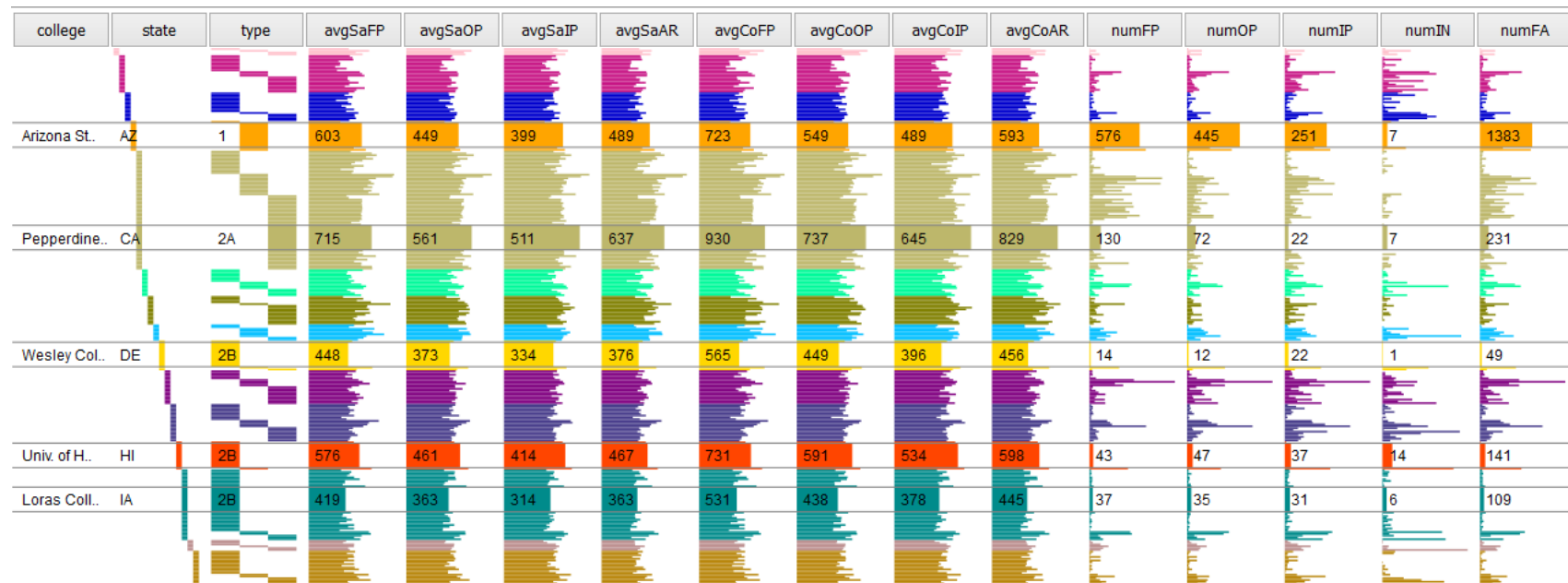


Figure.1 State(Coloring) and State(Sorting)

- ◆ Sorting by the state attribute

We could make the comparison among different states in terms of the number of schools easily. There are most colleges in CA, next is IA. The number is small in some states, such as AZ, DE and HI as shown in Figure1.

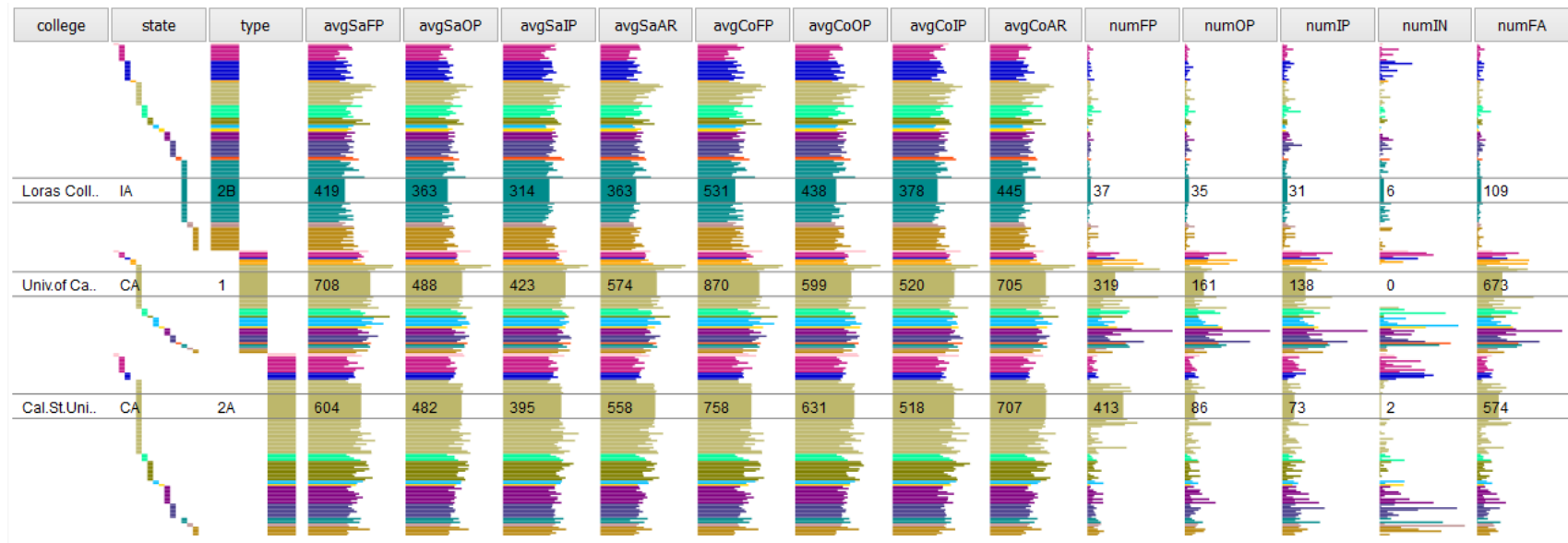


Figure.2 State(Coloring) and Type(Sorting)

◆ Sorting by the type attribute.

As we can see from the Figure2, for Type1 and Type2A, there are most colleges in CA. For type2B, the number of schools in IA is highest.

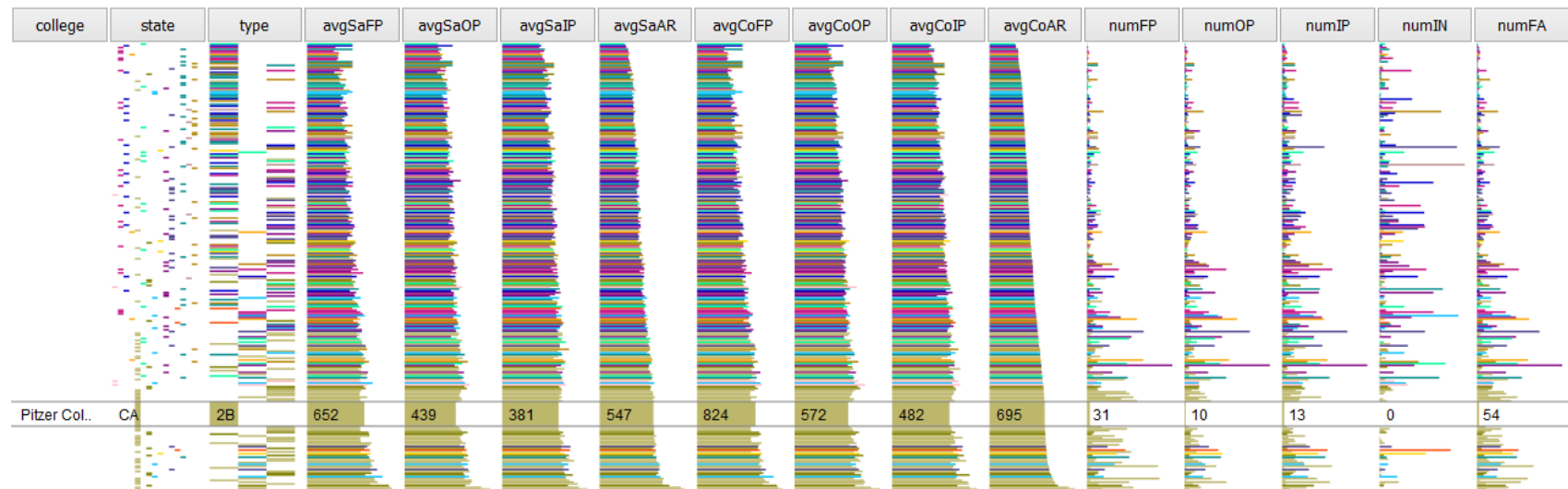


Figure.3 State(Coloring) and attributes about salary(Sorting)

- ◆ Sorting by avgSaFP, avgSaOP, avgSaIP, avgSaAR, avgCoFP, avgCoOP, avgCoIP and avgCoAR.
The pictures sorting by these 8 attributes are similar and it means that there are some dependencies between these attributes in this data set. What's more, we can see that the college in CA is relatively concentrated in the front area sorting by these 8 attributes.

- Coloring according to the type attribute

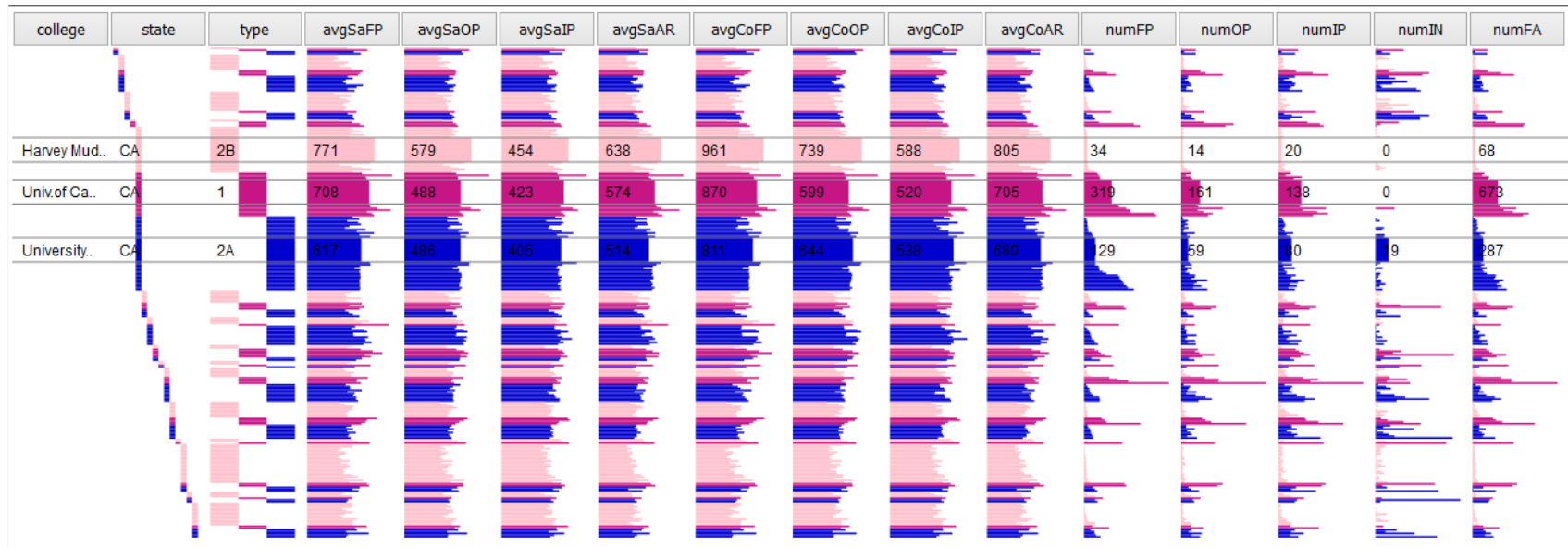


Figure.4 Type(Coloring) and State(Corting)

- ◆ Sorting by the state attribute

We could get the different proportions of 3 types in each states. Most states have the same distribution that the percentage of Type2A is maximum, next is Type2B and the minimum one is Type1.

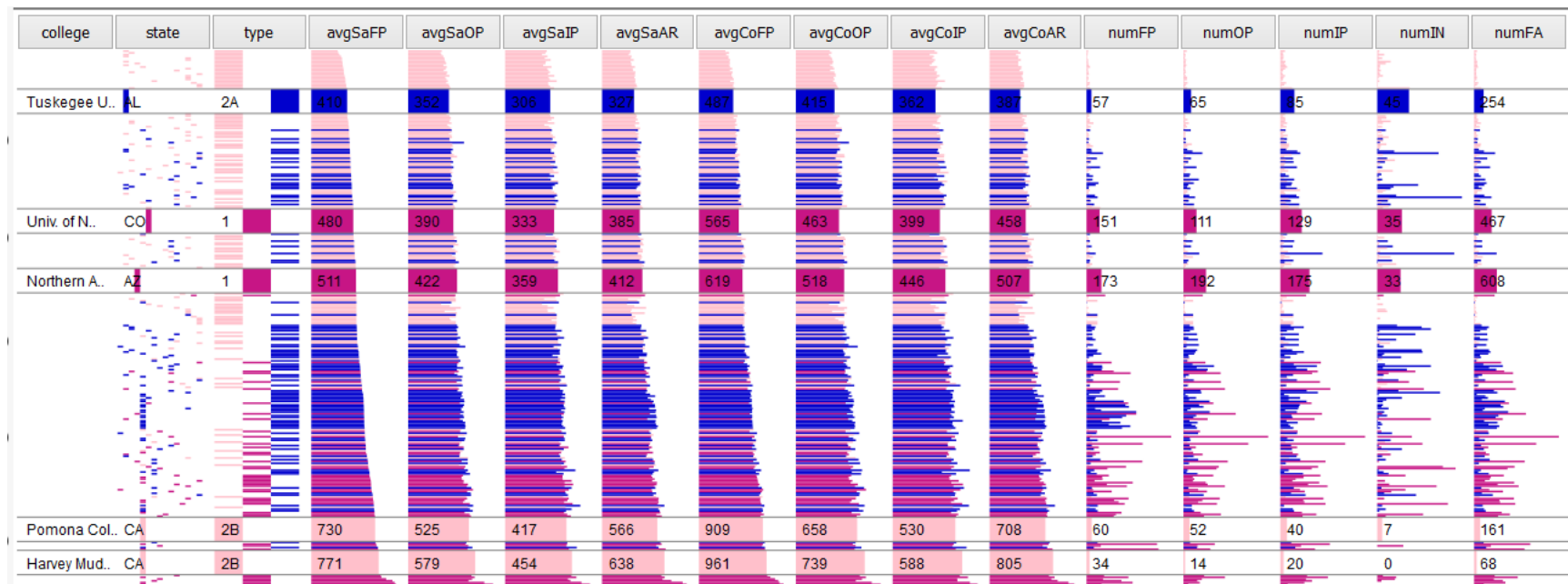


Figure.5 Type(Coloring) and avgSaOP (Sorting)

◆ Sorting by the avgSaOP attribute

We could get the distribution of different types of university in terms of salary attributes. The college of Type1 is in a relatively high level, the Type2A is in the middle and the Type2B is the lowest. Also, we could find some abnormal data easily in Figure5. There are some colleges of Type2B in the CA have high salaries, but the salaries of Type1 and Type2A are relatively low as the focus area shows.

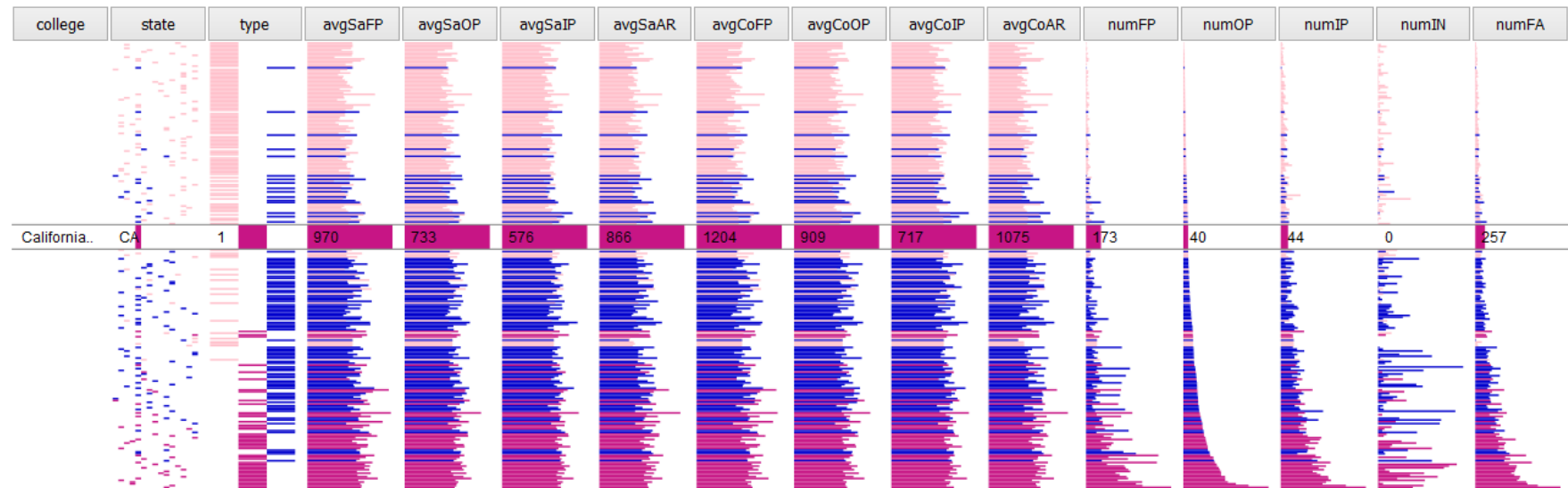


Figure.6 Type(Coloring) and attributes about numbers of teachers (Sorting)

◆ Sorting by attributes about numbers of teachers.

The difference of these attributes about the number of teachers between different types of colleges is obvious, but it is also apparent that a school does not follow this rule. As the focus area shows in the Figure6, this school is California inst. of tech, the feature of this college is that it is small but with highest salary. Also, it can be a clue to the subsequent analysis.

- Coloring according to the attributes about salary.

The color control bar is needed when coloring according to the numerical salary. Just as the Figure7 shows, red represents the high values and the blue represents the low values. We could get the distribution in different intervals of this attribute, and we also could change the interval partition in the follow analysis.

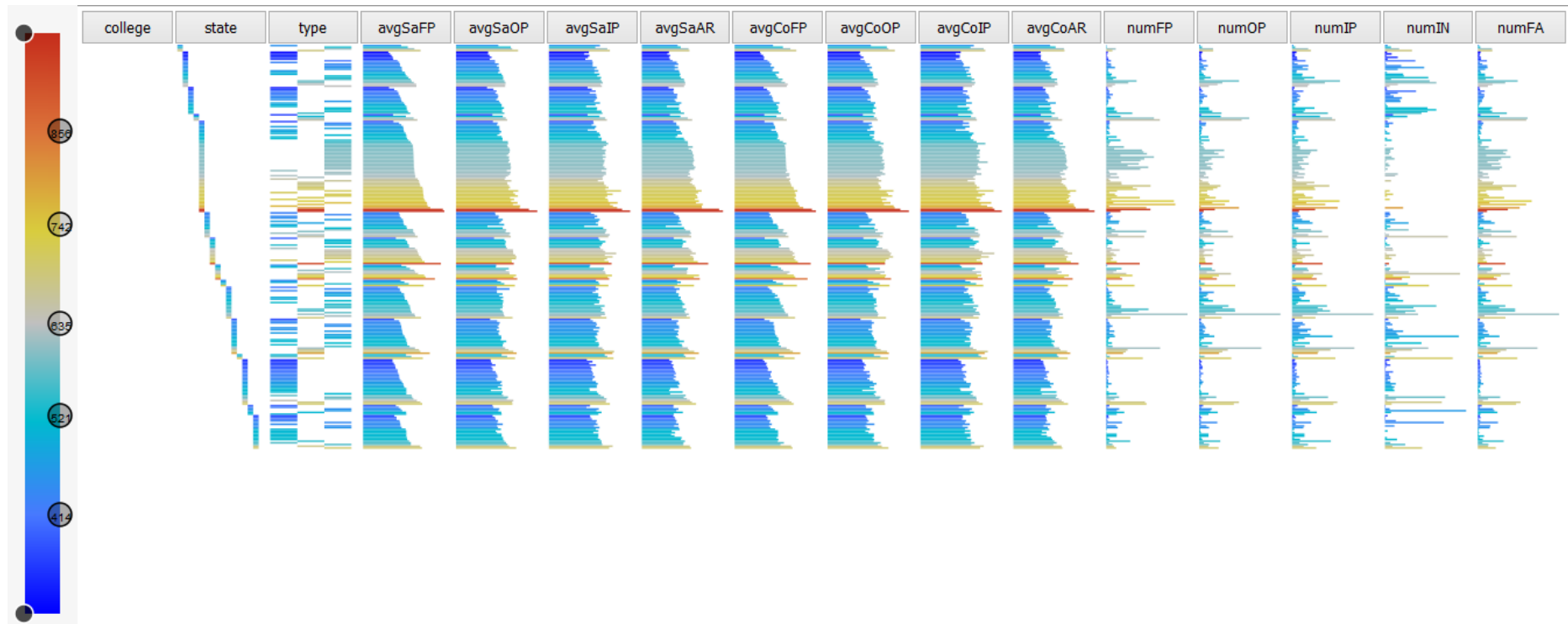


Figure.7 Attributes about salary(Coloring) and State(Sorting)

- ◆ Sorting by state attribute

From Figure7, we could see that the difference of salaries in CA is biggest among these states and most colleges with highest salary belong to this state.

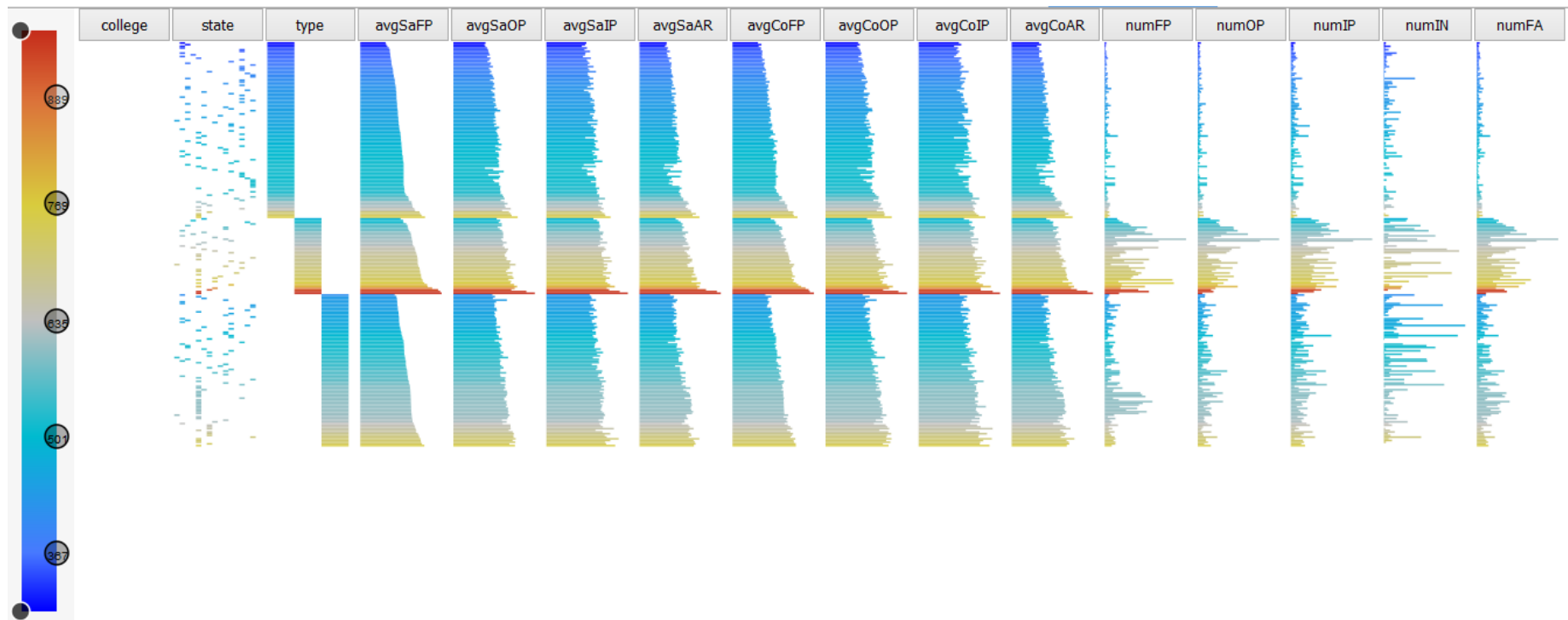


Figure.8 Attributes about salary(Coloring) and Type(Sorting)

◆ Sorting by type attributes.

From the Figure8, we could see that the difference in Type1 and Type2B is bigger than Type2A.

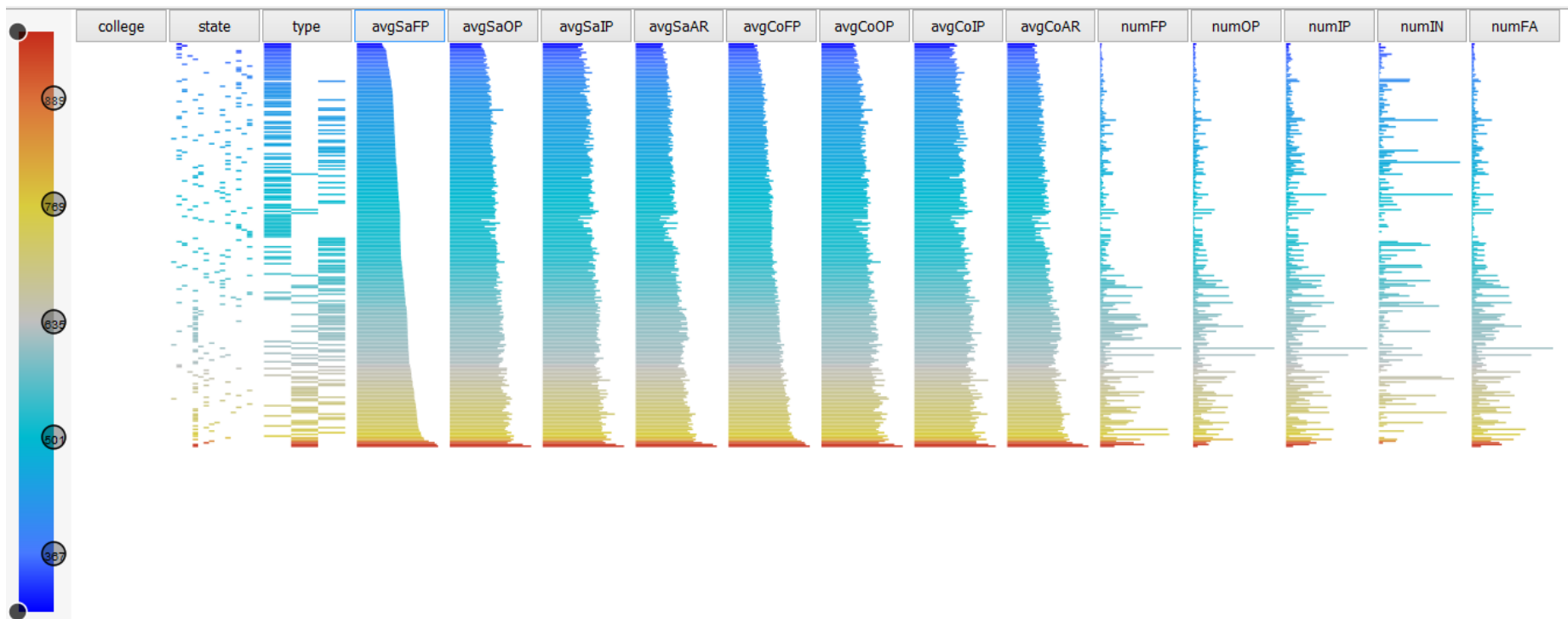


Figure.9 Attributes about salary(Coloring) and salary (Sorting)

◆ Sorting by attributes about salary.

We could see that the color scheme is different between the color control bar and the colored attribute, specifically, the proportion of corresponding interval is different, so we could get the information easily that the difference in the high salary level is bigger than the low salary level from the Figure9.

- Coloring according to the attributes about the number of teachers

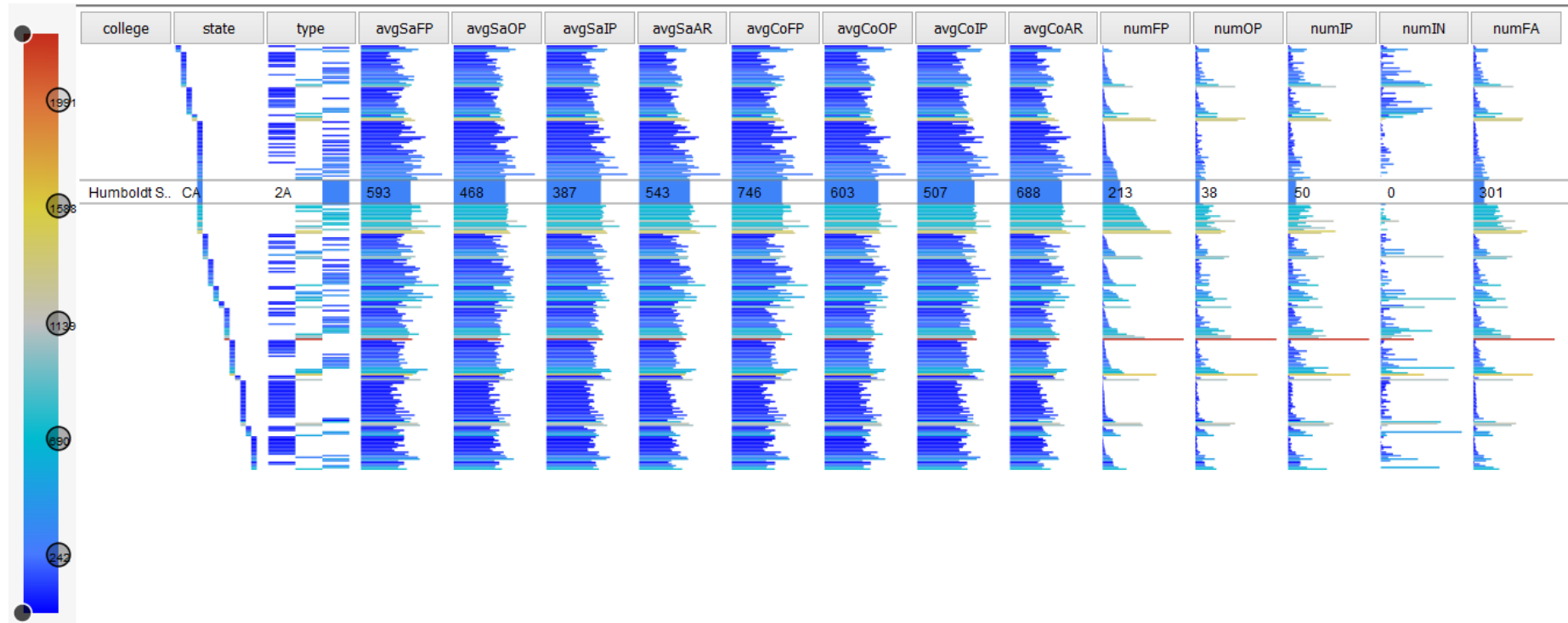


Figure.10 Attributes about the number of teachers(Coloring) and state(Sorting)

- ◆ Sorting by state attribute.

We could get the information that the difference in term of number of teachers is big and this attribute in CA is significantly higher than other states.

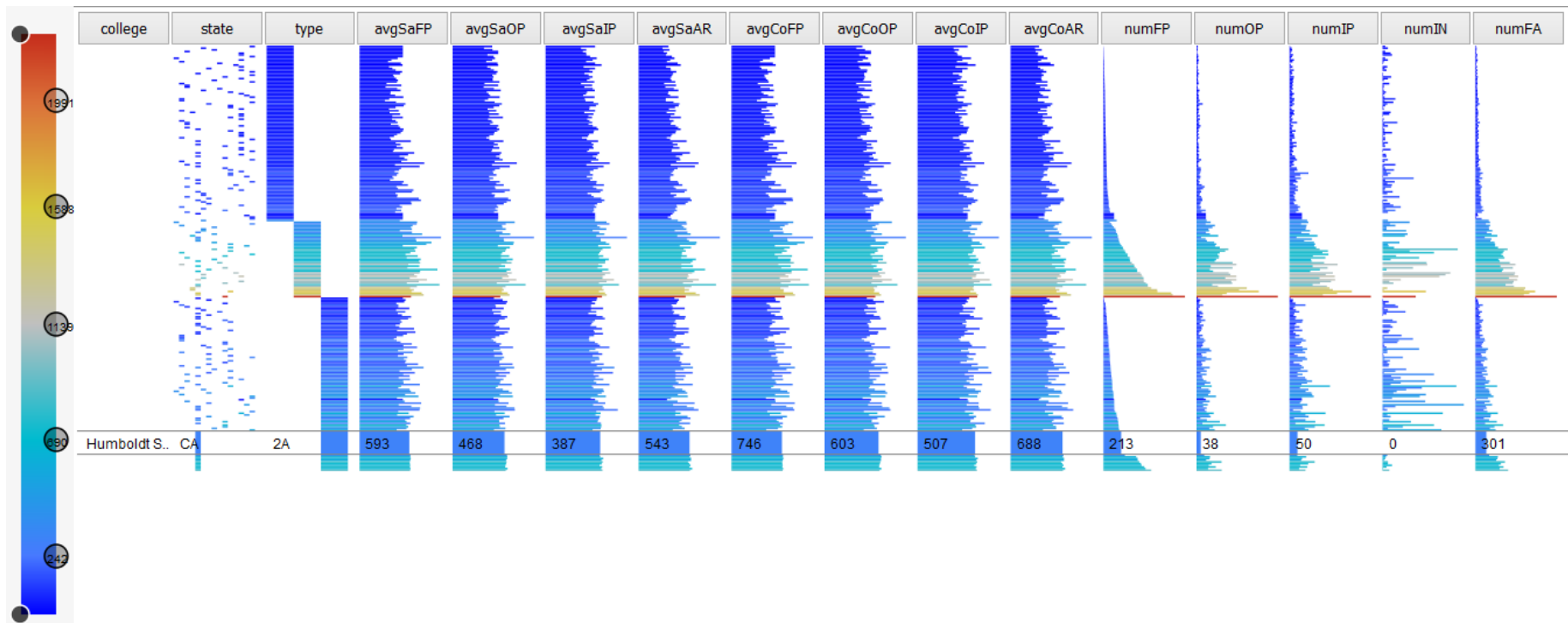


Figure.11 Attributes about the number of teachers(Coloring) and Type(Sorting)

◆ Sorting by the type attribute.

From the Figure11, we can see that the gap inside the number of Full Professors in Type1 colleges is far more than the Type2A and Type2B.

1.2.2. Some other discoveries

- **The distributions of salary and compensation**

We could found the information easily that the distribution of salary and compensation are similar from the Figure12.

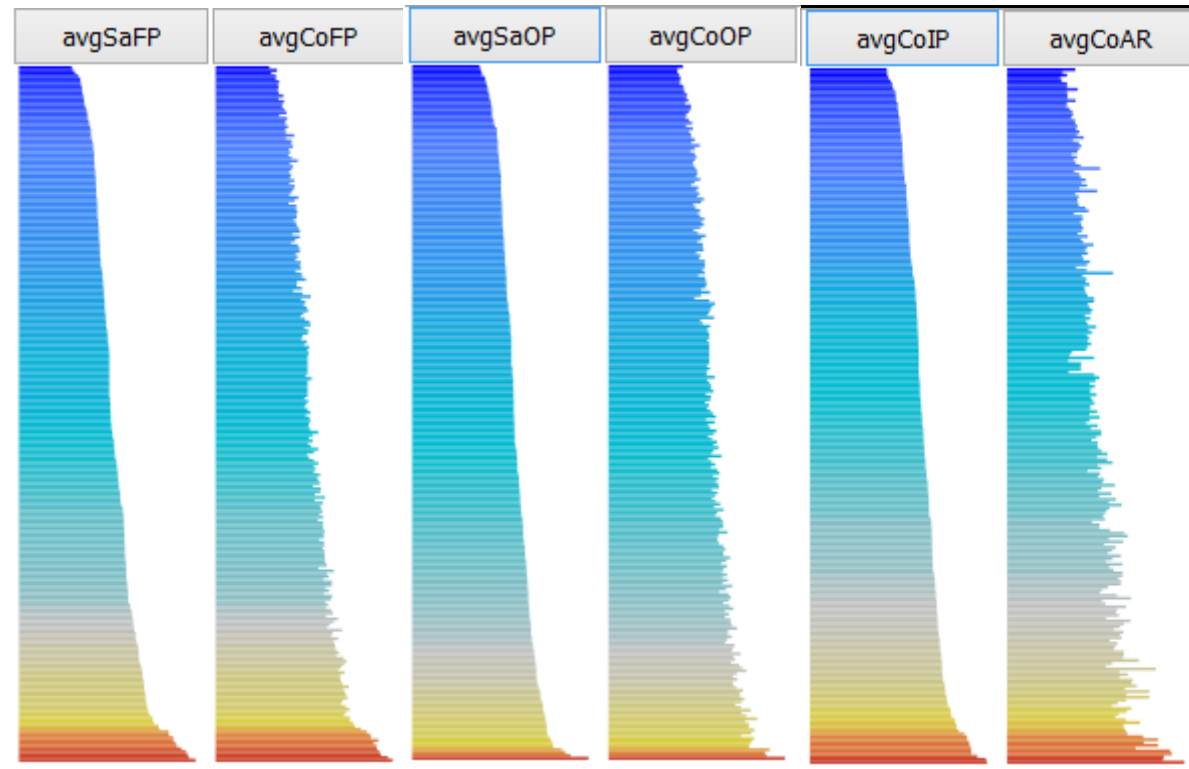


Figure.12 The comparison between salary and compensation of Full Professor, Associate Professor, Assistant Professor

1.2.3. Simulation to solve some problems

- **The First Problem: The features of best colleges**

In the previous analysis, we could see that the Type1 colleges have relatively more teachers with more salary, but we also could find that there are some counter-examples, such as California inst. of tech, the feature of this college is that it is small but with highest salary, so next we will get the real features of the best colleges.

In order to find the features of the best college, we find the top 10 universities including Princeton Univ., Harvard Univ., Yale Univ., Columbia Univ., Stanford Univ., Chicago Univ., Massachusetts Inst. of Tech., Duke Univ., California inst. of Tech., and then focus display these items under the color focus option.

Firstly, we sort data by the salary attribute, we could found that the average salary and average compensation of full professor in these 10 universities are also the highest, however the salary and compensation of associate professor and assistant professor are in the relatively high level.



Figure.13 The distribution of salary and compensation of full professor



Figure.14 The distribution of salary and compensation of associate professor

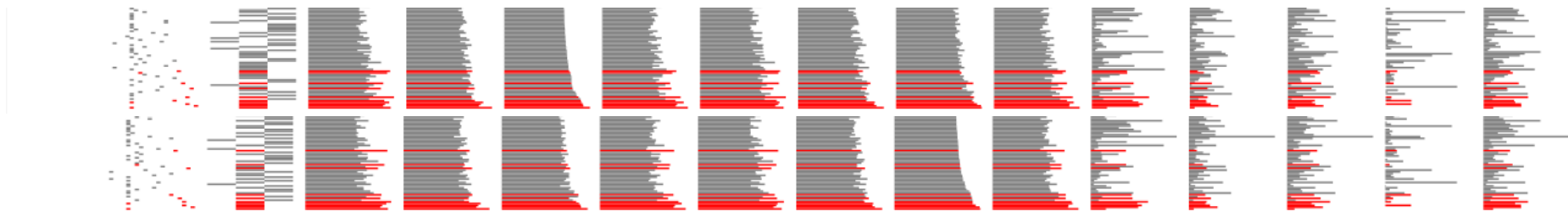


Figure.15 The distribution of salary and compensation of assistant professor

Next, we sort data by the attributes about the number of teachers and then we could found that the features is that the number of full professors and assistant professors is relatively high, however the number of associate professors is small. Also, the proportion of the full professors in these colleges is high.

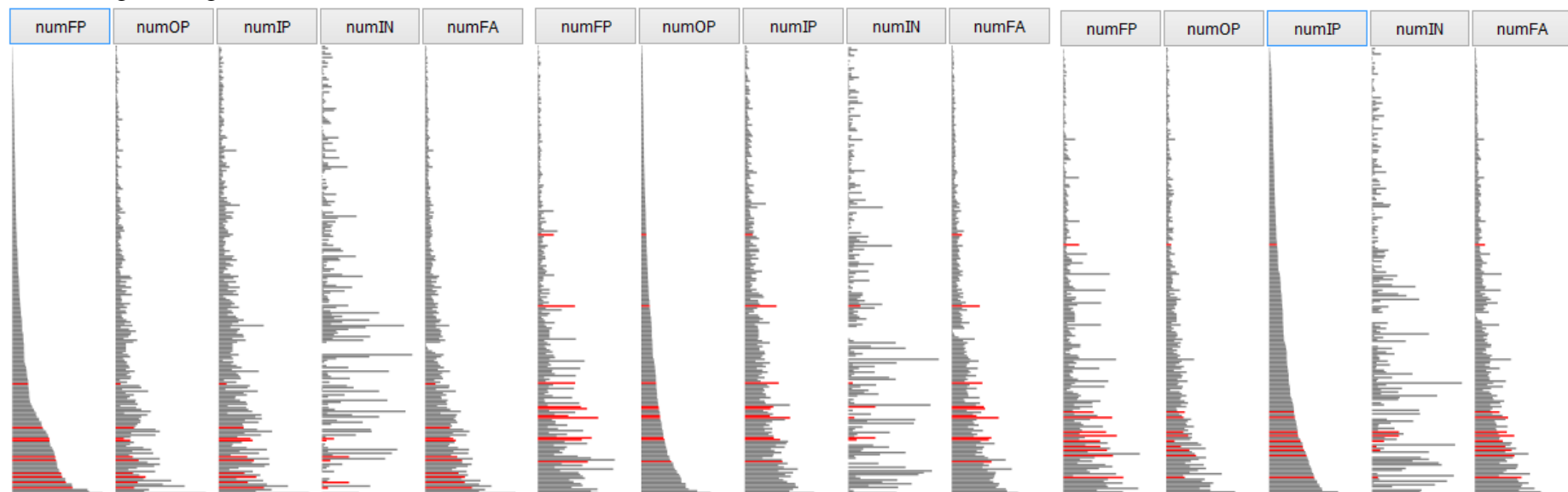


Figure.16 The distribution of number of full professors, associate professors and assistant professors.

From another perspective, we could get some other features of the best colleges. We select the items that all of the attributes about the number of

teachers is at a high level to get some information.

From the Figure17, the red items represents the colleges that all of the attributes about the number of the teachers are at a high level, we can see from the type attribute that these colleges are both belong to the Type1. It means that these colleges are also good universities, but we analyse the detail information of each individually, we found that these universities are not the best as the top 10 universities above.

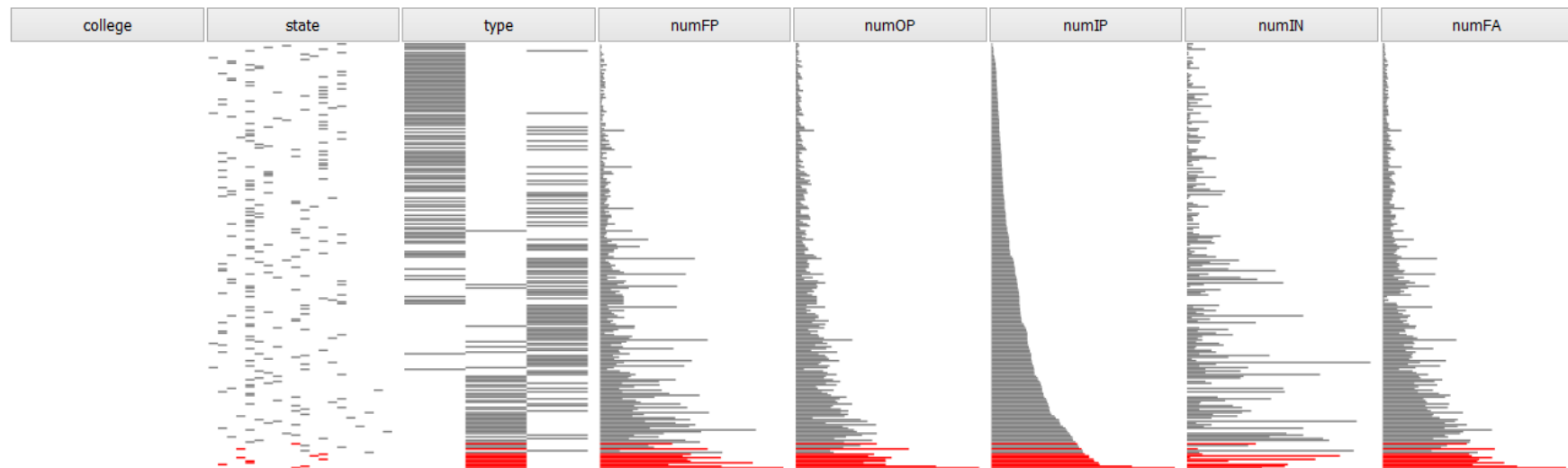


Figure.17 the colleges selected about the number of teachers

- **The Second Problem: The features of some states**

According to the previous analysis, the number of college in California state, the proportion of the Type1 and Type2A are highest, so we analyse the CA state to get its features.

In order to get some information other than that in previous, we sort this data by the average salary and type, then we could found that the colleges in the CA state in these 3 types are distributed in the upper part from Figure18.

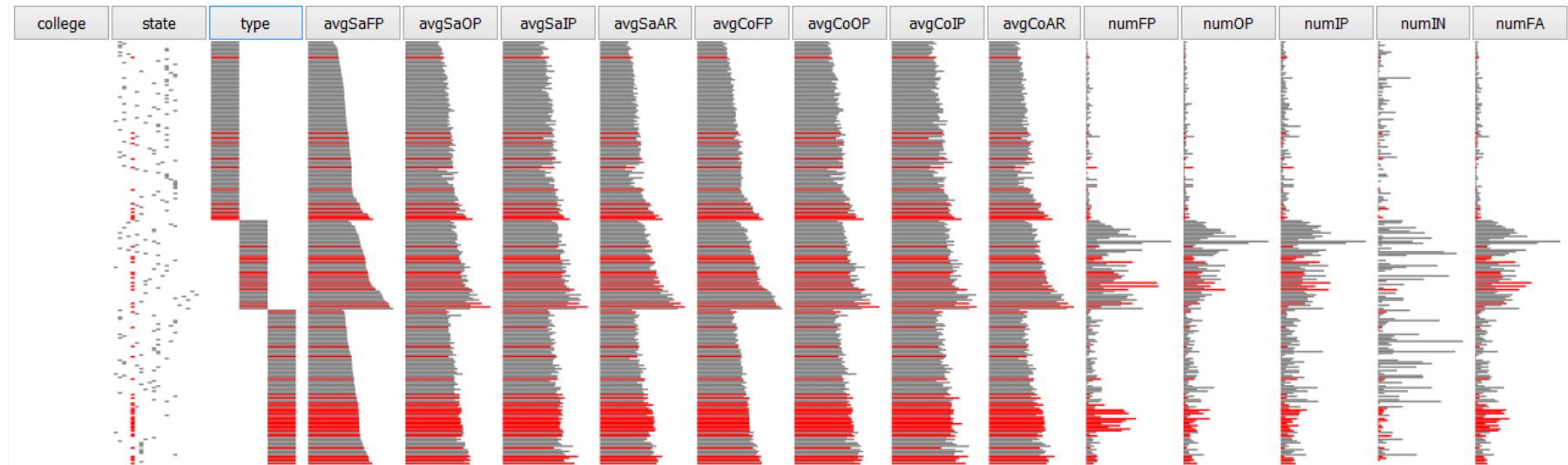


Figure.18 The salary distribution of CA state

Next, we sort the data by the attributes about the number of teachers and the type, then we could get the figure below. From this Figure19, we could get the information that for Type1 and Type2B, the universities of CA State distributed uniformly, however, for Type2A, the universities distributed mainly in the relatively high level. From the Figure19, we can see that the number of instructors in the colleges in the CA state is minimum comparing with other states.



Figure.19 The number of teachers distribution of CA state

- **The Third Problem: The features of the Inst. Colleges**

The Inst. Colleges have some common features in its name, such as including “Inst.” or “Tech.”, so we could find these items by the function of searching. Actually, this problem represents the filtering according to the features of the name, such as finding the persons with the same family name, find some places in the similar location and so on.

In order to find the features of the Inst. Colleges, we search the items whose name including “Inst” or “Tech”, then we find these 6 colleges in this data set. We could see that the most of them belong to Type1 and the teachers in these colleges are distributed in the upper parts.

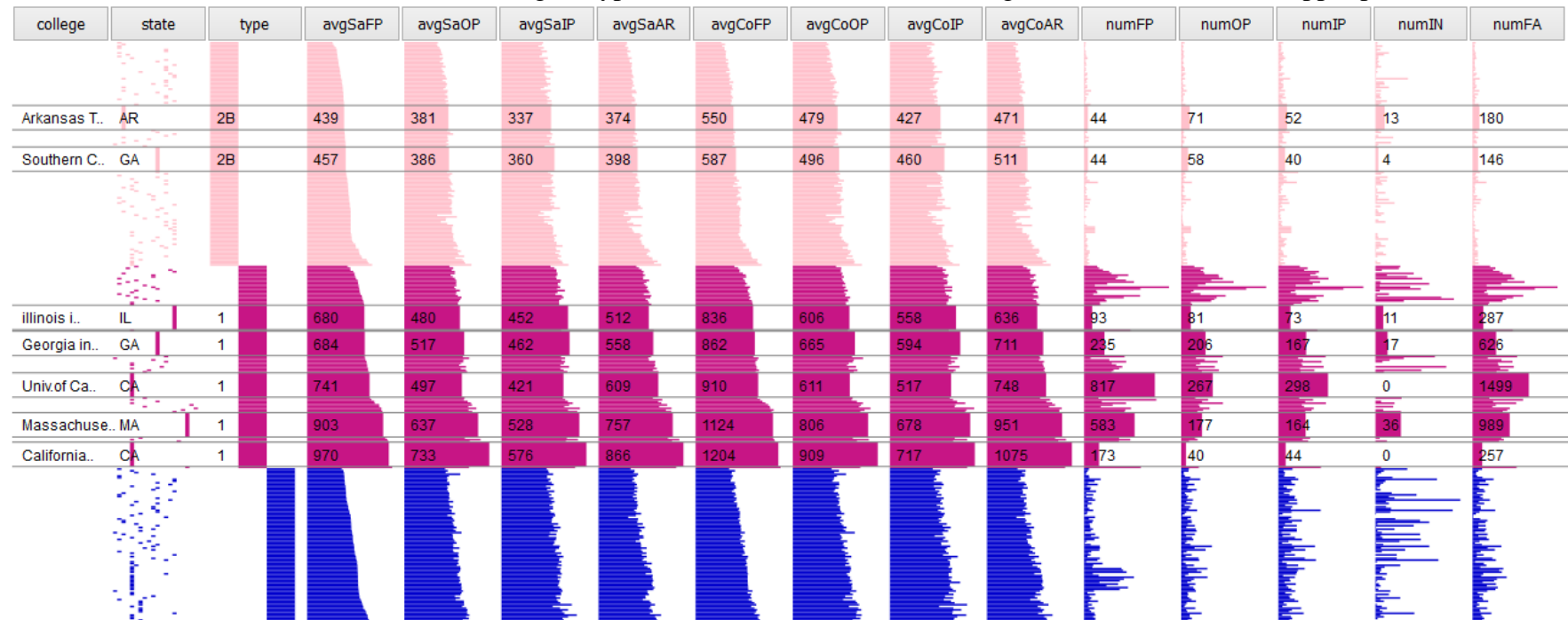


Figure.20 the information of the Inst. Colleges selected

From the Figure21, we could get that there are quite big difference among these Inst. College in the terms of number of teachers.

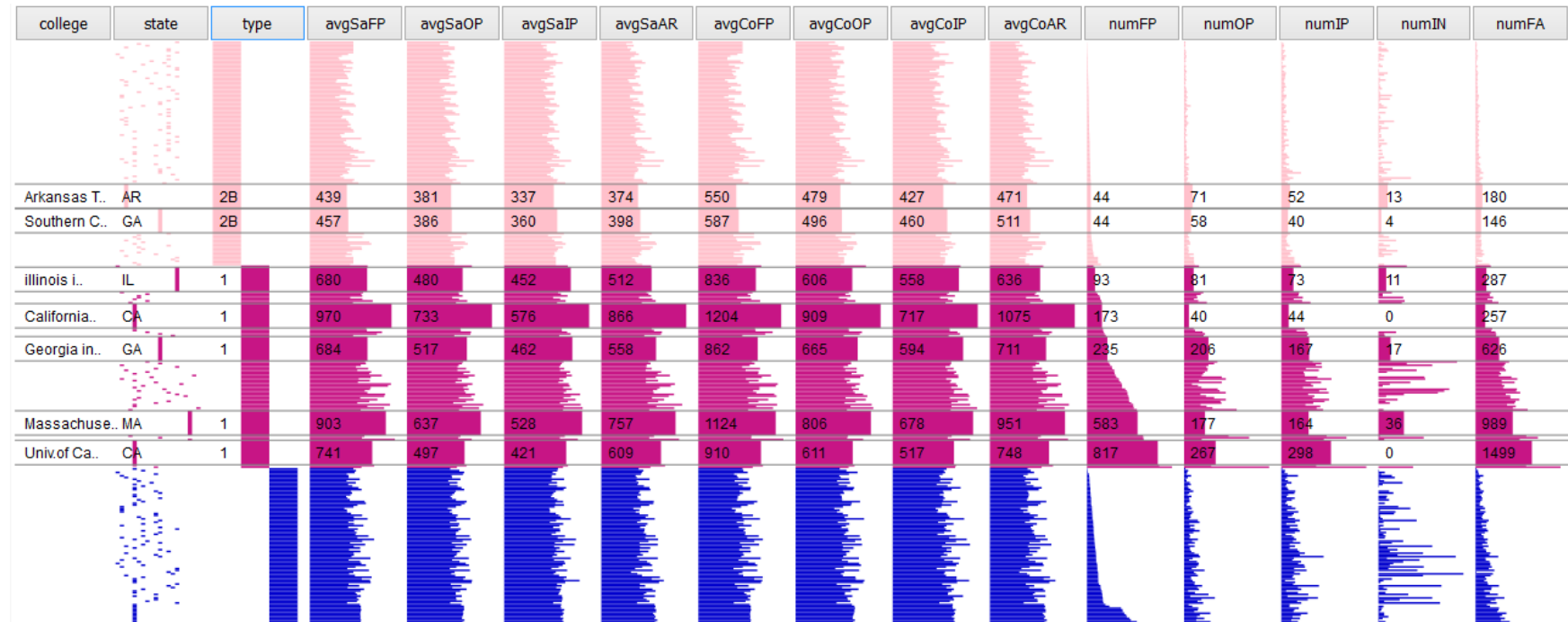


Figure.21 The distribution of Inst. College selected