Description of Visualization

1. We are trying to have a general idea about the data set and see how the features correlate with the grade.
2. We use treemaps to discover the distribution of “grade”. We found that there are 210569 out of 254657 cases got 0/1, which means that there are only 17.3% students got a grade over 0.
3. We use bar chart to explore the average action rates and grade in each course.
4. We looked at the scatter plots across all ten courses based whether the student is certified or not, their active days, video played times, number of chapters viewed etc.

* According to the scatter plots, we notice that there are outliers and influential data points, however we don’t have clear idea about how to deal with them. Should we exclude them or not? How to define the outliners?





