

# Report

Abhinav H Kamath(IMT2019001)

R Maruthi Sriram(IMT2019068)

Vignesh Bondugula(IMT2019092)

## Introduction:

In this report, we wrote about the remodelling which was done to show different visualizations. We wrote about what inferences can be made from the visualizations.

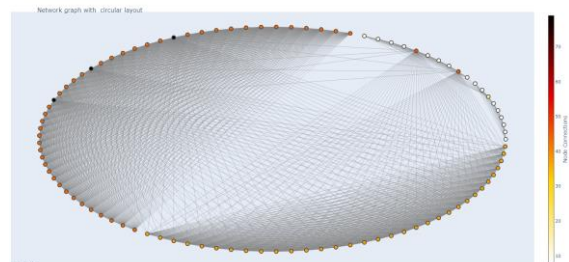
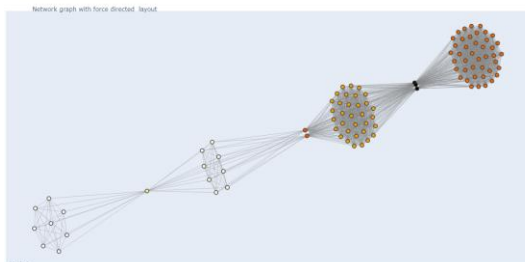
What is the remodelling done here? Describe in detail, including the rationale behind the remodelling choices.

The dataset we took was converted into an adjacency list to show Matrix Visualization and Force directed layout and Circular layout.

Force Directed Layout:

Node – used to represent a student

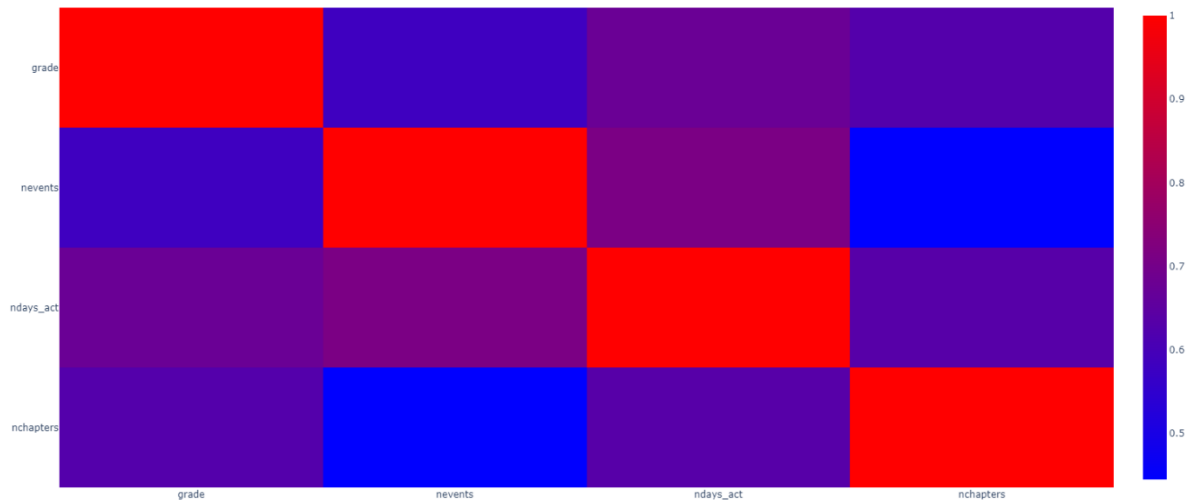
Edge - Edges exist between students that share a course



The force directed layout was done to see how students took different courses. This was done country wise. It was observed that from every country there were less people who took multiple courses. The aim being trying to find the trend in the choice of courses students make.

We end up with cliques equal to the number of courses connected together by students who took courses represented by both the clusters.

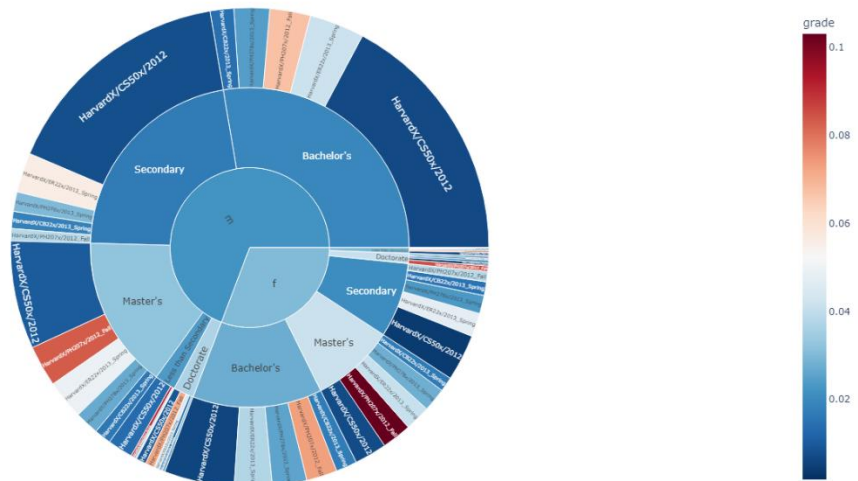
Matrix Visualization:

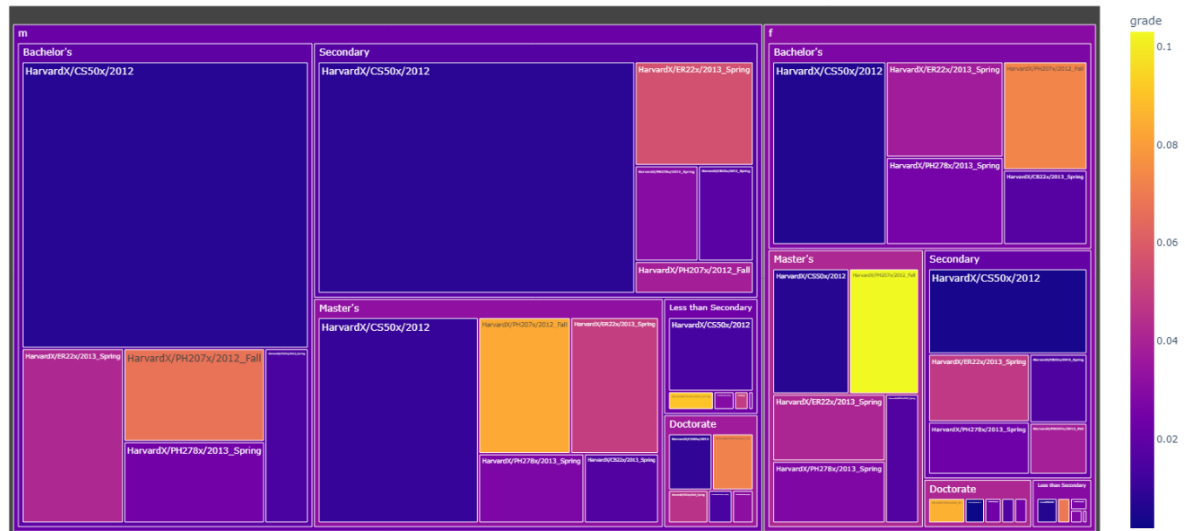


In the matrix visualization, a correlation matrix visualization was produced. The correlation between grade, nevents, ndays\_act and nchapters was drawn. It was observed that grade was most related to ndays\_act. This correlation was seen course wise to see how grade was dependent on different factors.

## Tree Map and SunBurst:

For the tree map and Sunburst visualization, the dataset was converted into a hierarchical model. We used gender, the highest degree achieved by students, and course as hierarchy.

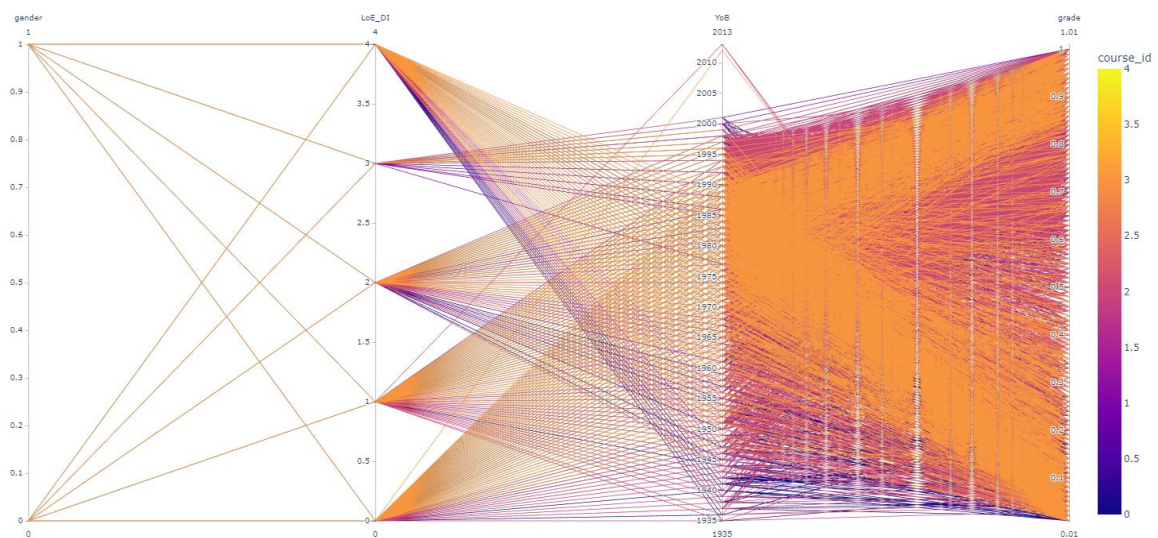




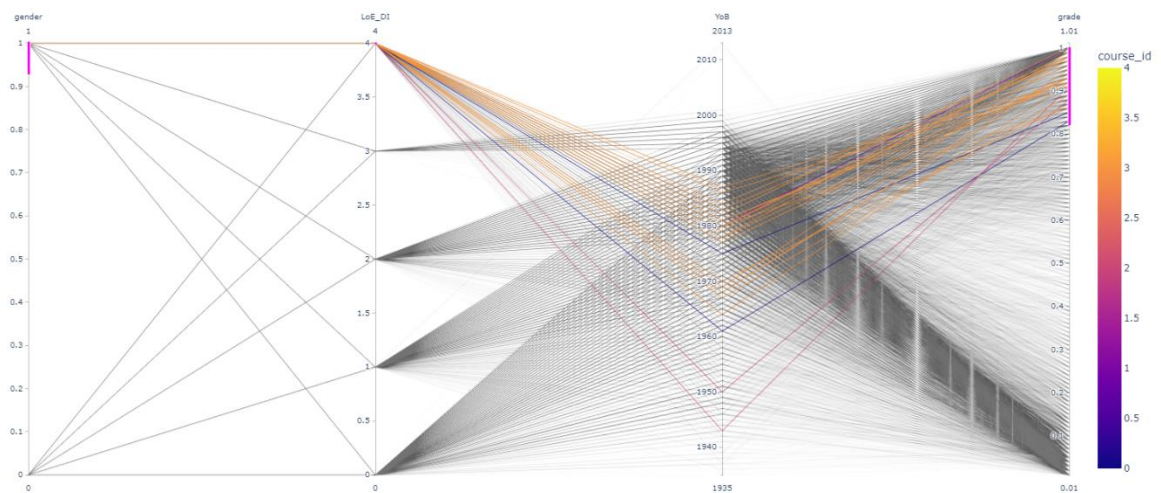
For the Sunburst Visualization, the color indicates the average grade got by students in that course. This was done to see what course was taken by majority of students and what course was preferred by what domain of students. For tree map the same remodelling was done.

## PCP Visualization:

For the PCP visualization, all categorical columns were changed to numerical, two user functions which we can apply to PCP visualization were Brushing and Swapping. These were applied to the visualization to get some more inferences. We took Gender, Level of Education, Date of Birth and grade as the columns for visualizing. Color of the lines indicates the course taken by students.

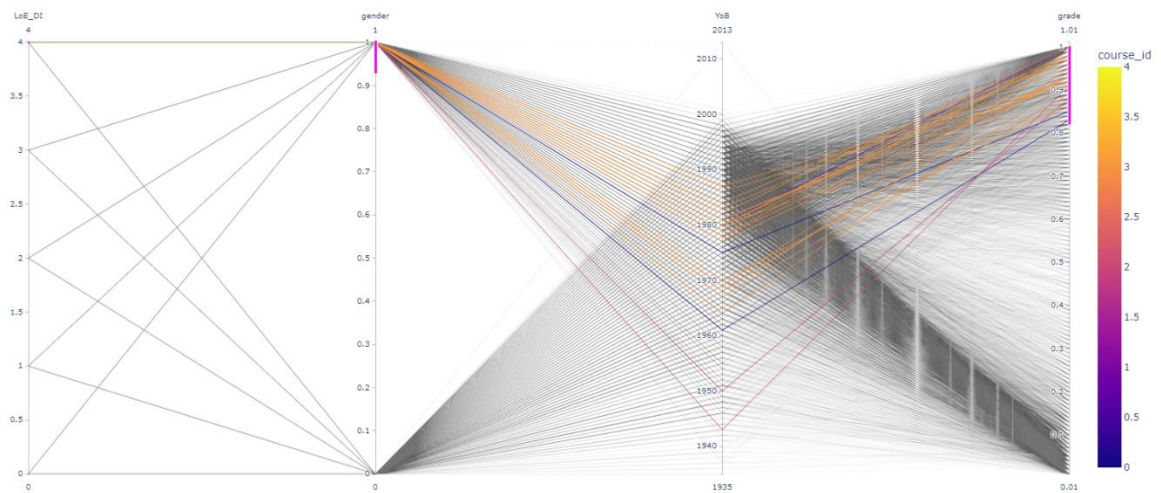


With Brushing:



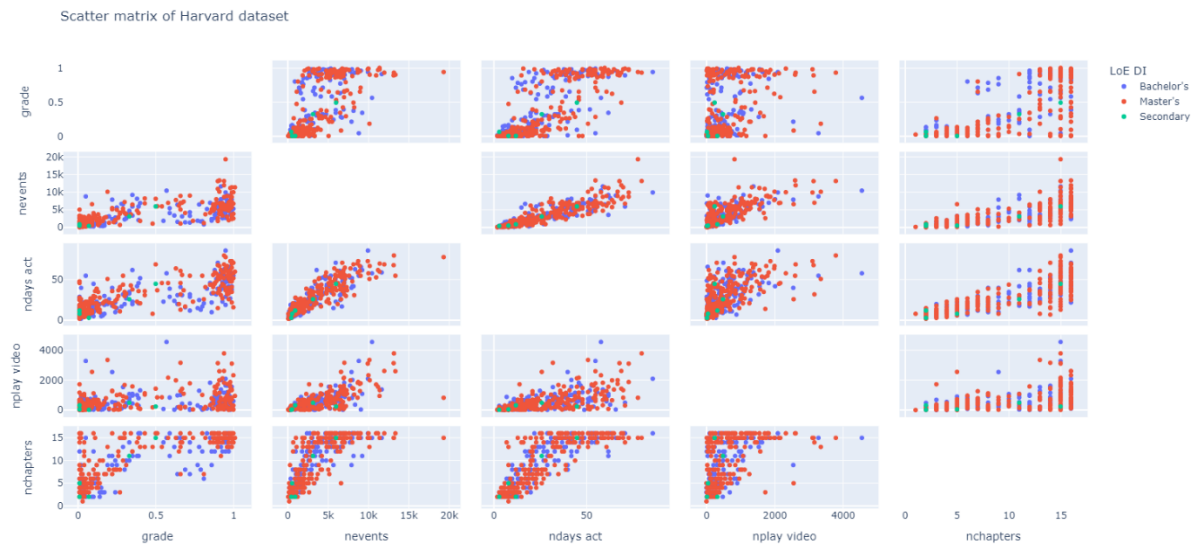
Using Brushing, we could narrow down the age group who get above 0.8 in their courses by filtering out one domain of students and gender.

With Swapping and Brushing:

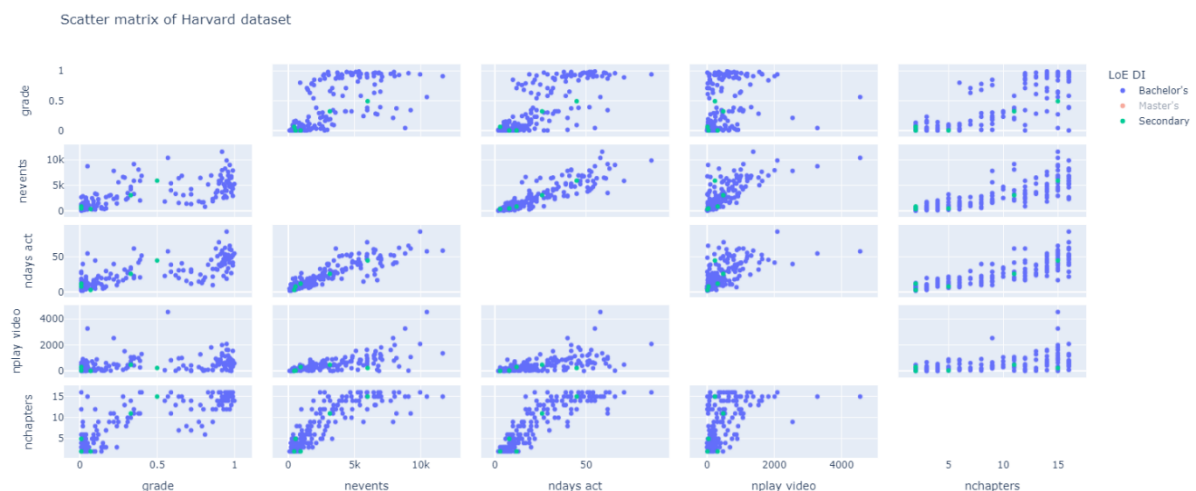


Axis swapping also gave the same results. No new Inferences were made from axis swapping

## Scatter Plot Matrix:



Scatter Plot was done to see how grade was varying with different variables. The color of the circles indicates the domain of the students. It was observed that ndays act was best for seeing the variation in grade.



In the above visualization, Master's students were removed and only Bachelor's and Secondary were kept. This is one of the user functions which can be performed on the Scatter matrix.

Was the remodelling and subsequent visualization done for the entire or part of the dataset? Explain in detail the decision made here, and if a partial dataset is used, how it was determined:

The remodelling was done for part of the dataset. Remodelling was done according to what type of visualization we wanted. We took only part of the dataset and only the columns and features which were required for the visualization were taken. For example for tree map, categorical

variables were only take for visualization as visualizing hierarchy in it was easier. For the Force directed layout and Matrix visualization columns with a one-to-one mapping were considered.

What is the joint inference that can be made from all six visualizations?

For making joint inference only one country was selected and the visualizations were made. We took Nigeria as the country to visualize and analyse how grades varied and what courses were taken by students. It was seen that Bachelor's were the majority of students who registered for the courses and their grade mostly was dependent on n\_days act variable.