##### What I’d like to do

1. Got to make report clear since the code isn’t
2. Linear models
   1. See what relation engagement with learning material has to the score
   2. < 0.5 p-value reveals significance
   3. Can convert categorical data to number and use that in y but have to be careful with that
3. Replace values inside a column without it taking a long time
4. Save the transformed CSV files and load them in transformed to speed things up
5. Merge some of the CSV files
6. Handle missing scores

A screenshot of a computer

Description automatically generated

Missing a few student scores

Because at this uni if you don't have a grade, you have a 0 grade

and then you have to get that sorted out lol

yeah either set to 0 or NA and if setting to NA, create a vector of the values excluding NA for all mathematical operations on that column (don’t need to do it when plotting)

A screenshot of a computer

Description automatically generated

Missing some values for assessment\_days\_since\_the\_module\_began\_due\_date

“If the information about the final exam date is missing, it is at the end of the last presentation week.”

A screenshot of a computer

Description automatically generated

**FIXED THAT UP!**

Have a lot of NA values for students that presumably have not unregistered yet



Have some missing values here

**Fixed this up**

A screenshot of a computer

Description automatically generated

Regions given to broad to be able to use the same Index of multiple deprivation value from one column in another column

Might go with most common value for that region {}

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A map of different colors

Description automatically generated

Needed to correct the age

**FIXED THAT UP!**

A screenshot of a computer

Description automatically generated

Missing exam due dates

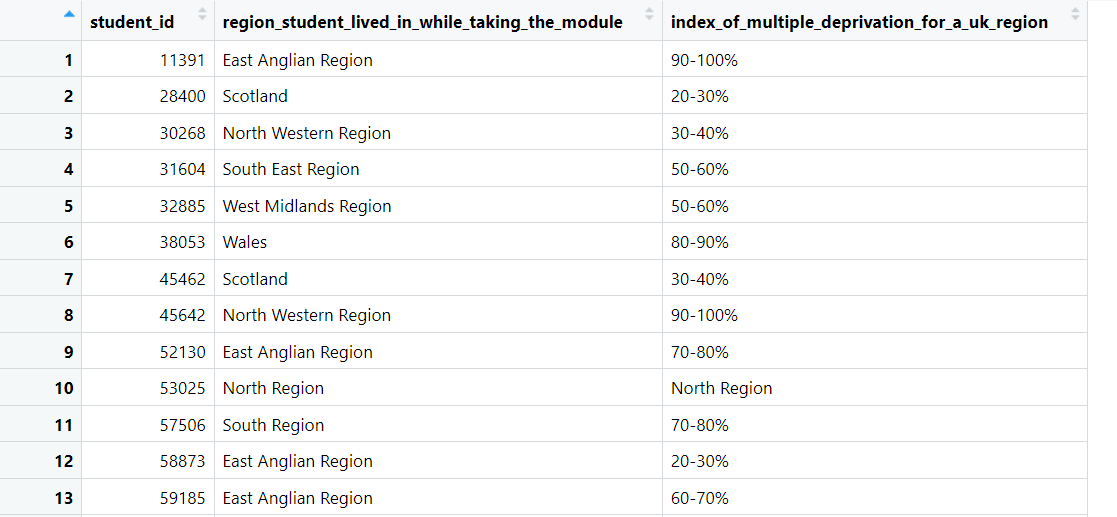
1. Comment on insights from data
2. **Answer these questions:**
   1. What is the most difficult class? Look at the class\_id that has the lowest average
   2. What class do students score the best in? Look at the class\_id that has the highest average
   3. What is the breakdown of students with a failure, pass, credit, distinction, or pass?
   4. What is the easiest class? I.e. What class do students score the highest in?
   5. What number of students have passed, failed, or achieved a credit, distinction, or high distinction (MongoDB derived column using mutate)

Show that exams only have 100% as they do in table in my code

“Typically, Exams are treated separately and have the weight 100%; the sum of all other assessments is 100%.”

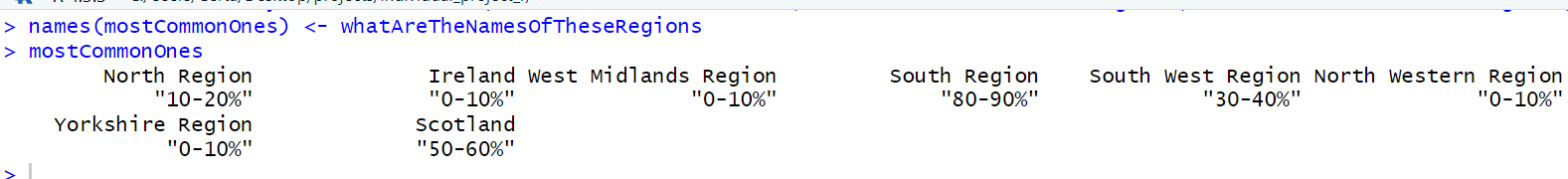
A screenshot of a computer

Description automatically generated



A close up of a screen

Description automatically generated



A screenshot of a computer

Description automatically generated

Replaced with 0 since it wasn’t too important