FIV Coursework - Visualising School Data

Adam Kulpa 4302781

April 15, 2020

Contents

1 Introduction

In this report I aim to visualise data about attendance, behaviour, and exclusions in different schools across the UK, but mainly the East Midlands. This data set is from Think For The Future, a small company that works with these schools to improve their students' attendance, behaviour, and ultimately but indirectly, grades. Before using the data for this project, I have made sure to have anonymised any personal data so that this data set is GDPR compliant. Other data sets will be used in combination to answer questions around links between behaviour, attendance, exclusions, grades, and area statistics, such as average household income.

2 Description of Raw Data

The raw data holds data about roughly 40,000 students from 38 schools around the UK (mainly the East Midlands). This data includes:

1. Attendance
2. Achievements (good behaviour)
3. Behaviours (bad behaviour)
4. Exclusions

The tables are in a not so intuitive format, and for some of the tables there is upwards of 20m rows, making the data slow and difficult to work with. **Please see the appendix** for more detail on these raw tables.

3 Initial Questions

1. Is student attendance related to their behaviour?
2. How does behaviour and attendance differ throughout the day?
3. How is attendance/behaviour affected by the average salary/household income in an area?

3.1 Fitness of Dataset

To answer Q1 and Q2, the raw data holds all the data necessary. The data includes attendance records, positive behaviour, negative behaviour, and exclusions, on each student including the time of day that these records occur. To answer the last question, I had to look to [2] to append an income column to my schools table. I used the address of each school as a lookup on the referenced website to get the local annual income (before housing costs) and this is the value I used in answering the question. To answer the follow-on question in Q3 of what the different types of schools are and whether they are religious, I used [1] to lookup the school information and append a further two more columns to the schools table. This addition can be seen in the attached dataset or in more detail in the appendix.

Please note that to answer any of the questions, a lot of data cleaning had to occur due to the nature of the raw data. I will only highlight some of these cleaning methods in this report, and will only do this for unique operations that are of interest, but **please see the R script attached** for more information.

3.2 Common Data Cleaning and Transformation

To get the data ready to use, there are a few things I did to prepare the data for the other questions.

3.2.1 Students Table

For the students table, I only selected the gender, yeargroup, school id, and student id. The yeargroup column is polluted with inconsistencies so I used:

```
...
mutate(yeargroup = as.numeric(gsub("Year", "", yeargroup))) %>%
...
```

This extracted the numeric value from the yeargroup column and where there wasn't one, for example in 'Reception', this would turn into an NA. NAs are next omitted using:

```
...
na.omit(yeargroup) %>%
...
```

Next to allow any other table to use this table to join to, I used the student id and the school id to make one school_student_id which is unique for any student. This had to be done as student id by itself is not unique across schools.

3.2.2 Schools Table

To get the schools table ready, I wanted to calculate how many students there are per school. To do this I joined the students table using a left join (to only keep students that are in a school from the school table), grouped by each school, and summarised the data, counting the records per group (school) giving the count of students. I then filter the schools to have at least 10 students (to get rid of schools with no data).

3.2.3 Attendance Tables

Attendance uses two tables. Attendance records and the attendance codes. Each attendance record uses a mark, and this mark has a meaning in the scope of the school only. This meaning is described in the attendance codes table.

Therefore I had to join the two tables. To do this I had to make a new column in each table that is a composite of the mark and the school id (for uniqueness). I also clean up the attendance category and put approved and present as the same thing (since the student did not miss school in either). This was done using:

```
...
 attendance_codes$short_meaning_description[
    attendance_codes$short_meaning_description
    == "Approved"] <- "Present"

attendance_codes$short_meaning_description <-
    factor(attendance_codes$short_meaning_description)
...
```

To join the tables I used inner join to remove any attendance records without a meaningful mark, and to remove any meanings without any records assigned to them.

The attendance table has more than 20 million rows, therefore to make it more friendly to use in the future, I summarise the data based on the category of record.

Firstly I filter out all meaningless marks. Then the data is grouped by each student. Here I create a new column to count how many attendance records each student has in total so I can next summarise and calculate the percentage for each category like so:

```
...
group_by(school_id, student_id) %>%
mutate(totalMarks = n()) %>%
group_by(school_id, student_id,
    short_meaning_description) %>%
mutate(markTypeCount = n()) %>%
summarise(markTypePerc 100*
    round(first(markTypeCount)/first(totalMarks),
    3)) %>%
ungroup() %>%
...
```

This now gives me a much smaller table, 4 records for each student, with all the attendance statistics summarised, 1 record per summary. However this is in long format, and would make more sense to be in wide format so that it is 1 record per student, with the categories as the columns. This is done using the spread dplyr function:

```
...
spread(short_meaning_description, markTypePerc) %>%
replace(is.na(.), 0) %>%
```

...

The second line is there to replace NAs with 0s.

The last stage is to join this table with the students table so I now have the attendance statistics summarised with the information about the student such as gender and yeargroup in one table.

3.2.4 Achievements and Behaviours

To clean both achievements and behaviours I needed to join their two respective tables. Achievements and behaviours both have objects/instances which are stored in one table, and assignments where students are allocated to those instances in another table. To join these I had to create a new column that combined the school_id with the behaviour_id as that makes each id unique. I then for each table group by the school id and the student id and summarise both the achievements to be left with the number of achievements/behaviours and the summed points for each student. Then these tables are joined with the student table so that the statistics on achievements/behaviours are together with the yeargroup and gender of each student.

An interesting part about cleaning the behaviour is when it came to putting the time field of the behaviour table into the correct AM or PM. This is because initially this is a field comprising of a factor with 100s of levels. Some referred to the time, some to the activity that was being done. Here I used many grep functions to decide which value should be classed as morning vs which should be classed as afternoon. As well as using strptime to get the string format into a time format so I can compare it with midday:

```
#Clean time field to AM and PM
behTimesWithYeargroups$beh_time[strptime(behTimesWithYeargroups$beh_time, "%H:%M") < strptime("12:00", "%H:%M")] <- as.factor("AM")
behTimesWithYeargroups$beh_time[strptime(behTimesWithYeargroups$beh_time, "%H:%M") >= strptime("12:00", "%H:%M")] <- as.factor("PM")
behTimesWithYeargroups$beh_time[grepl("AM", behTimesWithYeargroups$beh_time, fixed=TRUE)] <- as.factor("AM")
behTimesWithYeargroups$beh_time[grepl("Am", behTimesWithYeargroups$beh_time, fixed=TRUE)] <- as.factor("AM")
behTimesWithYeargroups$beh_time[grepl("PM", behTimesWithYeargroups$beh_time, fixed=TRUE)] <- as.factor("PM")
behTimesWithYeargroups$beh_time[grepl("Lunch", behTimesWithYeargroups$beh_time, fixed=TRUE)] <- as.factor("PM")
behTimesWithYeargroups$beh_time[grepl("lunch", behTimesWithYeargroups$beh_time, fixed=TRUE)] <- as.factor("PM")
behTimesWithYeargroups$beh_time[grepl("Break", behTimesWithYeargroups$beh_time, fixed=TRUE)] <- as.factor("AM")
behTimesWithYeargroups$beh_time[grepl("Before", behTimesWithYeargroups$beh_time, fixed=TRUE)] <- as.factor("AM")
behTimesWithYeargroups$beh_time[grepl("After", behTimesWithYeargroups$beh_time, fixed=TRUE)] <- as.factor("PM")
behTimesWithYeargroups$beh_time[grepl("Reg", behTimesWithYeargroups$beh_time, fixed=TRUE)] <- as.factor("AM")
behTimesWithYeargroups$beh_time[grepl("break", behTimesWithYeargroups$beh_time, fixed=TRUE)] <- as.factor("AM")
behTimesWithYeargroups$beh_time[grepl("[5,6,7]", behTimesWithYeargroups$beh_time)] <- as.factor("PM")
behTimesWithYeargroups$beh_time[grepl("[1,2,3,4]", behTimesWithYeargroups$beh_time)] <- as.factor("AM")
behTimesWithYeargroups$beh_time[grepl("End", behTimesWithYeargroups$beh_time)] <- as.factor("PM")
```

3.2.5 Exclusions

For some reason, when I was importing the exclusions, a lot of ghost records would find their way in to the table that are not originally there. The only cleaning I do with this table is to select only the columns I want (as there are many), and filter out any empty records as such:

```
...
select(exclusion_days, start_date, student_id,
    school_id, exclusion_type, exclusion_reason,
    start_session) %>%
filter(!start_session=="")
```

3.3 Cleaning for the rest of the questions

Explaining how I cleaned the data to get each table for each visualisation would take me way over the word count as my R script is over 400 lines long, therefore **please see the R scripts for more detail**. Each question builds on top of the common cleaning I just explained so I hope this introduction into the cleaning methods I used is satisfactory in this report.

3.4 Q1: Is student attendance related to their behaviour?

3.4.1 Visualisation Strategy

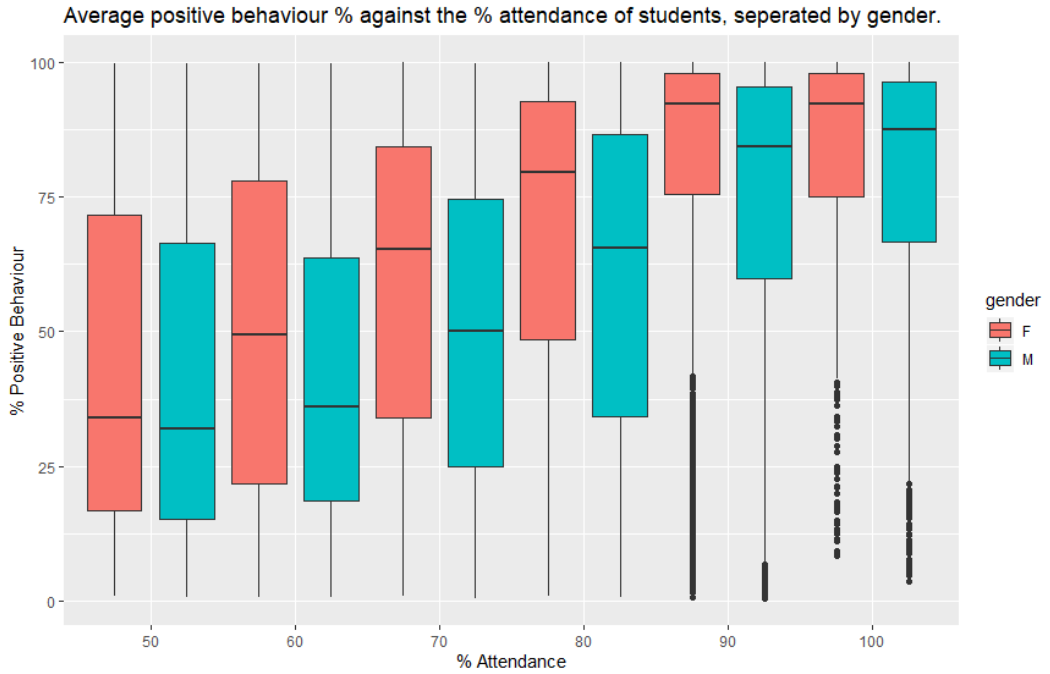
To start, I want to tackle a question with an expected answer. Here I expect to see that as attendance percentages drop, the behavior gets worse. To add an extra part to this question, I wanted to see how this changed with gender. To do this, I needed attendance and behaviour percentages, and the gender, making this a trivariate table.

The variables I want to represent:

- Percentage Attendance (Nominal - 6 Factor)
- Percentage Positive Behaviour (Ratio Quantitative)
- Gender (Nominal - 2 Factor)

As the goal is to visualise a relationship between attendance and behaviour, I will choose these two variables as most important and therefore will use position to visualise these in the most effective way. Gender will be added using colour as there are only two values in the data, and it will be easy for the interpreter to distinguish between two colour values. I want to see the mean as well as the quartiles for this data for better comparisons between % attendance groups/bins. Therefore my choice of visualisation is a boxplot.

3.4.2 Interpretation and Further Exploration



Along the x-axis is the % attendance, 10% sized bins from 50% to 100%. Along the y-axis is the % of conducts (behaviours and achievements) that are achievements. So it is the ratio of achievements to behaviours for each student.

The box plot clearly shows that for students that attend less, the positive behaviour percentage is under 50% meaning that they have more negative behaviours than positive. As the percentage attendance increases, so does the mean, lower, and upper quartiles of each box-plot, showing a strong correlation between good attendance and good behaviour.

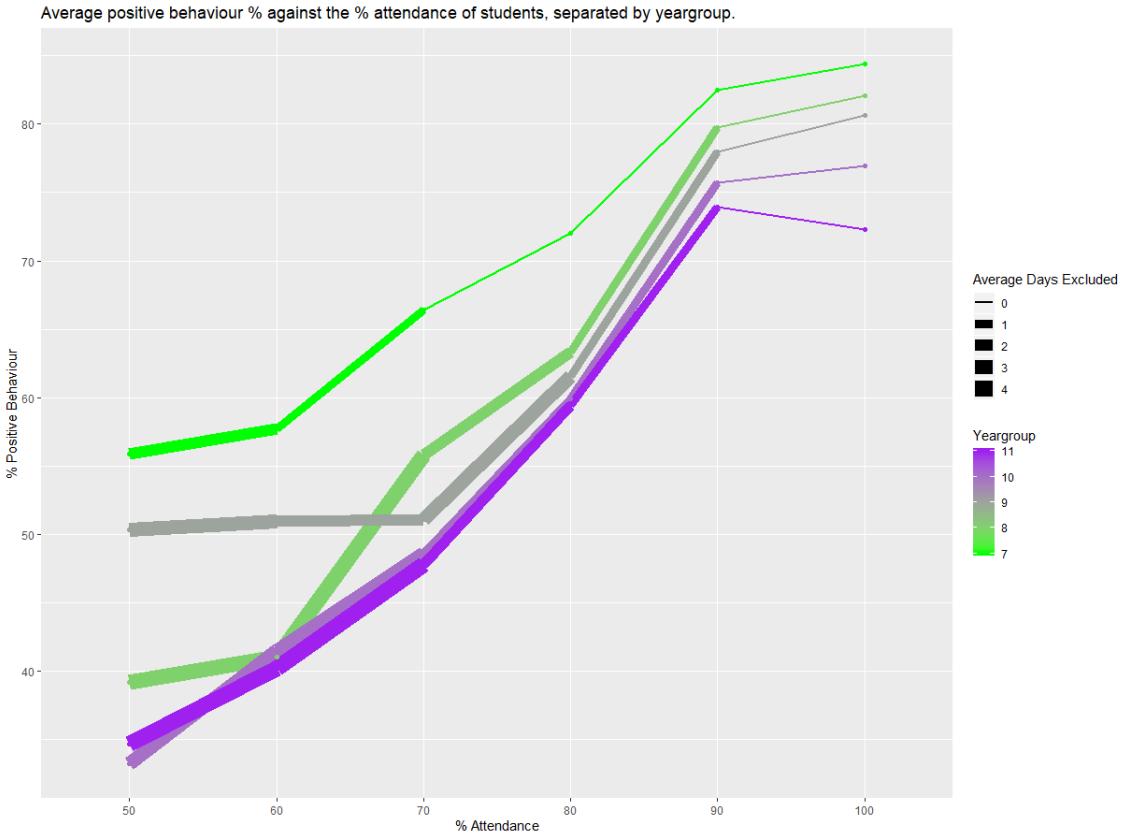
This box plot also shows that girls and boys both follow the same trend of better behaviour with better attendance, however, interestingly, for all attendance categories the girls are shown to be better behaved. This is shown quite strongly as the mean, lower, and upper quartile of every female box plot against the male in each attendance % bin is higher.

A further question that stemmed from this visualisation is whether there is any difference in attendance and behaviour between different yeargroups, and how that links to the average number of days a student is excluded.

This removes the gender variable, however introduces two more nominal variables:

- Yeargroup (Nominal - 5 Value)
- Average Days Excluded (Ratio Quantitative)

For this hypervariate data I decided to use a line plot, encoding the yeargroup as a different line colour, from green to purple, representing year 7s to 11s, and encoding the exclusion days by using the thickness of the line, hoping to see a change from thick to thin.



From this graph, it is interesting to see that the younger yeargroups are more well behaved than the older yeargroups across the attendance % range. Towards the higher % attendance, the order of yeargroup perfectly indicates that the older the students the less well behaved they are.

The thickness also shows the average number of days the students are excluded for, and as the lines get thinner towards the top right where attendance and behaviour is good, and thicker where the attendance and behaviour is poor, it is clear that attendance also is correlated to how often/ how severely a student is excluded.

3.4.3 Evaluation

It is surprising to me how clear the trend here is and how strong the correlations are. The question was answered with confidence in my opinion, especially since these visualisations are based on dozens of millions of records. Looking back, I am unsure if the boxplot adds any value since each boxplot is similarly sized so the mean value may have been enough, however, on the other hand I wouldn't have known that if it wasn't displayed.

3.5 Q2: How does behaviour and attendance differ throughout the day?

3.5.1 Visualisation Strategy

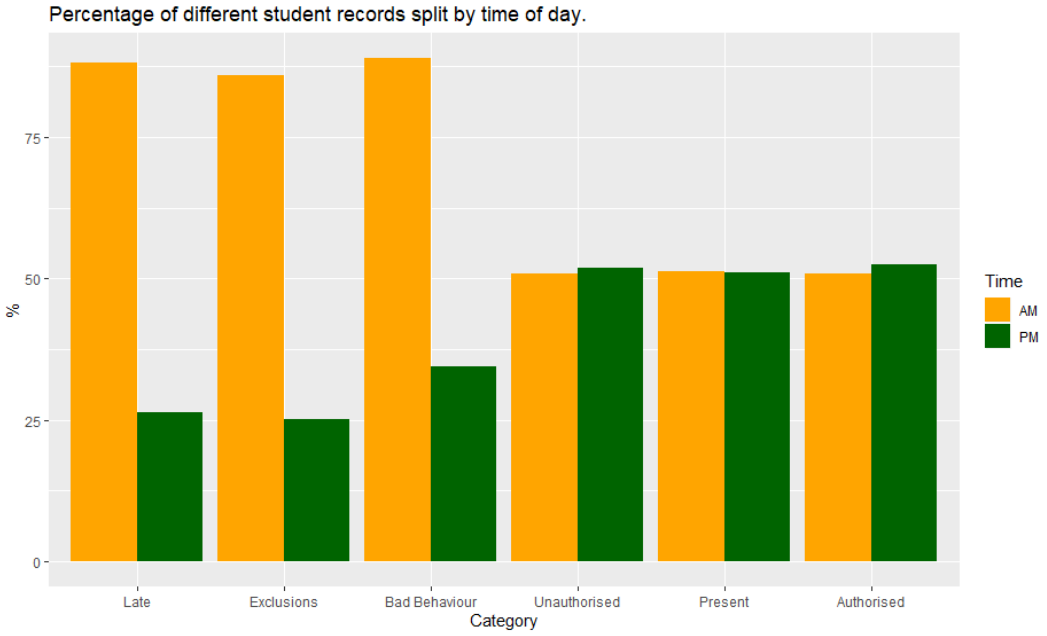
To answer this question, I wanted to gather student records from ranging categories and show how they differ between the morning and afternoon. For this I wanted to use attendance data, behaviour data, and exclusion data. In order to eliminate the bias of certain schools giving out more or less points, this had to be done using ratios, e.g. 'What percentage of all lates are in the morning?' and 'What percentage of all bad behaviour occurs in the morning?'

Variables to visualise (trivariate):

- Category (Nominal - 6 Factor)
- Percentage (Ratio Quantitative)
- Time (Nominal - 2 Factor)

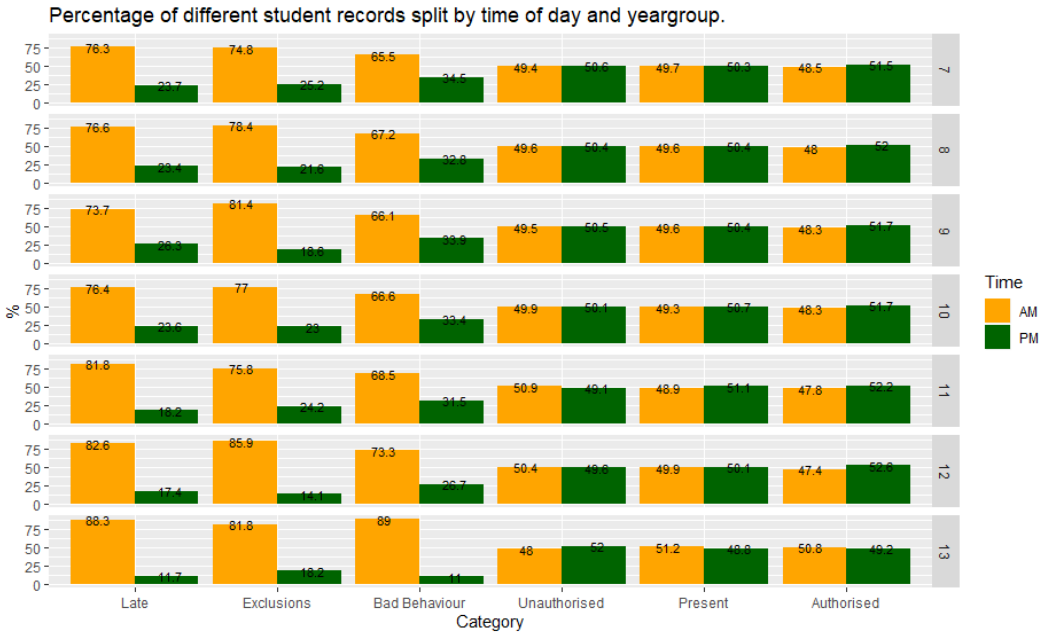
As there are 6 different values to represent for the category, this variable will be prioritised in terms of importance. Using Mackinlays principle of importance ordering, these will be encoded first using the most effective technique, therefore position, in this case I will choose this to be the x-axis. Being the only quantitative variable, the percentage will use the other position encoding, the y-axis. This leaves the time, and since it only has two values, I will use the colour to differentiate between "AM" and "PM".

3.5.2 Interpretation and Further Exploration



The graph shows that for unauthorised and authorised absences, and present marks the difference between morning and afternoon is negligible. It is expected that a lot more lates occur in the morning, for example students being late to school. However it is most interesting to see that most of the bad behaviour, and exclusions (more than 75%) occur in the mornings too. It is therefore safe to say that students behave and therefore most likely perform better in the afternoons of school.

To find more detail about this, I wanted to see if this is the case for all yeargroups, and if so if there is any trend between the differing ages of students. To do this I had to add the yeargroup data into each summary table giving me an extra nominal variable, turning this into a hypervariate table. In this case I decided to use small multiples.



3.5.3 Evaluation

It was difficult to get the data in the correct format, mainly due to the 'time' field in the behaviour table which had a lot of different nominal values which had to be summarised to AM or PM. See the data cleaning for more detail of this. This meant that some of the behaviour records would have been removed, some I may have misinterpreted, and therefore the results here could be slightly skewed, however, I still think the correlations seen here and therefore the conclusions drawn are strong and answer the question well.

3.6 Q3: How is attendance/behaviour affected by the average salary/household income in an area?

3.6.1 Visualisation Strategy

To visualise this data, I wanted to plot each school as its own data point, showing a correlation between the local area annual income of the school and how 'good' the school is. To define 'good' I wanted to define my own simple scoring system. I also wanted to encode the size of the school in terms of the number of students into the visualisation to see if there is any correlation there also.

This meant I had to visualise 3 variables per school (trivariate data).

- Score (Ratio quantitative)
- Local Annual Income (Ratio quantitative)
- Number of students (Ratio quantitative)

According to Mackinlays principle of importance ordering, the most important variables need to be encoded in the most effective and accurate way. The most accurate way of encoding quantitative data is using position. As the main objective is to see correlation between local annual income and the school score, these will be prioritised as the x and y coordinates, and the number of students will be encoded as the size of the data point using area. This is also because it is intuitive to interpret the size of the data point as the population of the school. This leads me to believe that a scatter plot, plotting each school as a data point along the score vs. income axis, will be the most effective way to visualise a correlation between the two. It will also allow me to plot a regression line through the data to clearly answer the question.

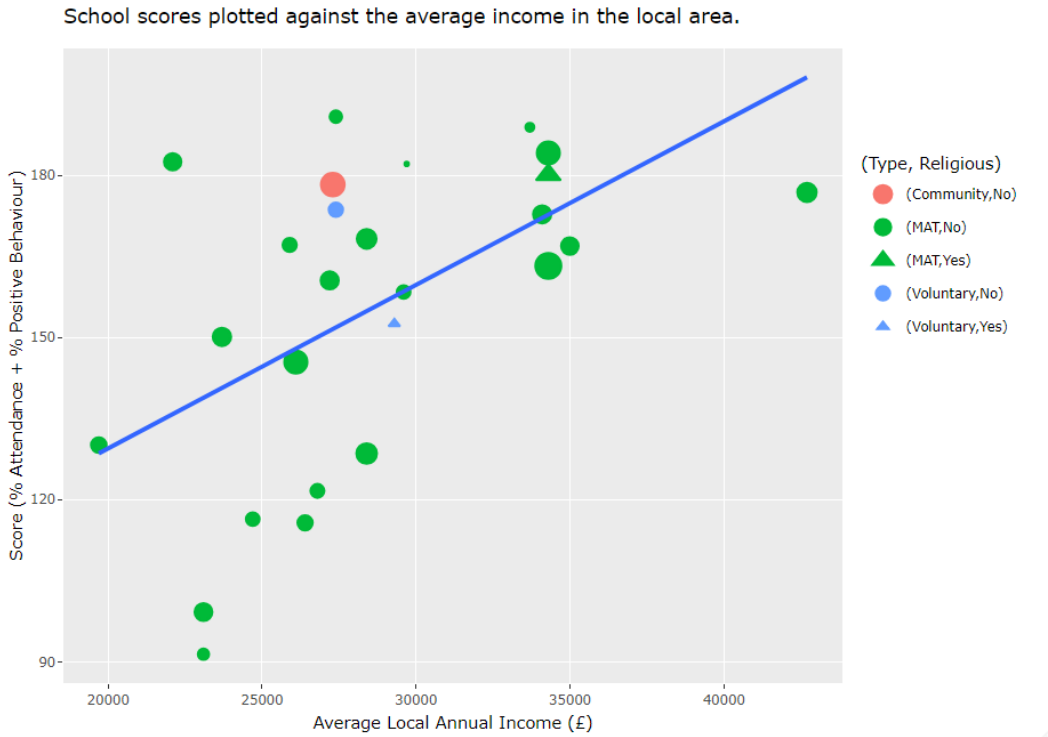
3.6.2 Interpretation and Further Exploration

The resulting graph shows that as the annual income increases, the score of the school also generally increases. In short this means that schools in richer areas are more likely to have a high score, although this does not mean you cannot find a good school in an area with a lower income. It also looks like size of the school does not correlate with the area or score with any reasonable strength.

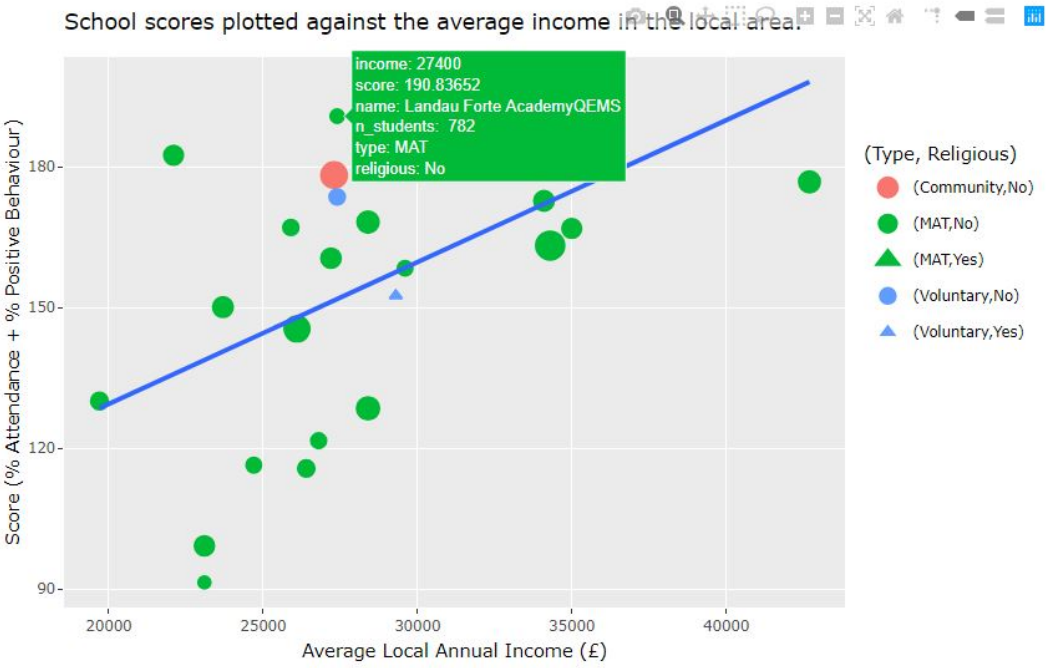
From this visualisation, to further explore correlations between schools and their performance, I wanted to differentiate between different types of schools. To do this, I wanted to consider whether the school was private, part of a multi academy trust (MAT), a community school, or ran by/ with the aid of volunteers, as well as whether the school is religious or not. This adds more variables and makes the data hypervariate. Two more variables need to be encoded:

- Type (Nominal - 4 Factor)
- Religious (Nominal - 2 Factor)

I decided to use the shape and colour encoding as they are both good for nominal data. Colour is more differentiating than shape however, and therefore it is assigned to the type as it is a larger factor and therefore more values to differentiate between. Whether a school is religious or not is therefore represented by the shape.



I wanted to include user interaction into this visualisation so that the user is able to look at the specifics of each data point, most importantly which school is which. Using the interactivity of this visualisation, you can see that the best school in my dataset is Landau Forte Academy QEMS.



Unfortunately, due to the number of schools in the sample, there isn't a large variation in the type or the religiousness of the schools, and in fact the one private school didn't make it to the visualisation as it had incomplete data. However an observation that can be made is that in both cases where a school was either one of the couple religious schools, or not a part of a MAT, it is in the upper half of schools for the score they have.

To conclude this question, any school, regardless of area or type can be a school full of good behaviour and good attendance, however schools that are a part of a multi academy trust have the most mixed scores, especially under the £30,000 annual income mark, and schools that are not part of an MAT, or are religious, all have scores in the upper end of the data. To have the best chances of finding a good school, one should look for schools that are not

a part of an MAT or that are religious, and to look in areas with an annual income of over £30,000.

3.6.3 Evaluation

The biggest issue with this question is that it is difficult to answer, and conclude correlations, with confidence. The way this answer could be improved is by adding many more data points (schools) to solidify the strength of the correlation. Other than that, I am happy with the way the data is visualised, although it is a shame that the ggplotly library does not support hypervariate data legends, leading to a somewhat confusing combined legend.

4 Reflection on Development Process

Through this coursework I have felt myself become more comfortable using R and developing a process of how to get data into the format I need it to be in. I learned a lot about data manipulation due to the difficult format the data I used was initially in. Throughout the coursework I also became a lot more familiar and comfortable with information visualisation theory such as the differences between univariate, bivariate, trivariate, and multivariate data, the differences between nominal and quantitative variables and how to encode them using Mackinlays principle of importance ordering for effective visualisations.

References

[1] Government. *GOV.UK Search For Schools*. URL: `https://www.compare-school-performance.service.gov.uk/find-a-school-in-england`.

[2] Government. *Income Estimates for small areas*. URL: `https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/bulletins/smallareamodelbasedincomeestimates/financialyearending2018`.

5 Appendix

5.1 Fact Tables

5.1.1 anon_students

Stores all the students in all schools that there is data for.

Column Name	Data Type	Example	Description
id	Int	12345	The primary key and identifier of each row (not used).
student_id	Int	54321	Foreign key to other tables that reference a student.
forename	Char	forename_hidden	Forename of the student.
surname	Char	surname_hidden	Surname of the student.
gender	Char	M	Gender of the student.
yeargroup	Char	Year 08	Year group that the student is in.
dob	Date	dob_hidden	Birthday of the student.
status	Char	OnRoll	Whether the student is currently studying or has left.
reg_group	Char	reg_group_hidden	Register group of the student.
entry_date	Date	04/09/2018	The date that the student started at this school.
school_id	Int	8304054	Foreign Key to the schools table.

5.1.2 anon_schools

Stores all the schools that there is data for.

Column Name	Data Type	Example	Description
school_id	Int	3734279	The primary key for each school, this is how other tables reference the school.
name	Char	"Greenwood Academy"	Name of the school.
income	Int	25500	The income of the local area. This value is received from [2]
type	Char	"Greenwood Academy"	Name of the school.
religious	Char	"Greenwood Academy"	Name of the school.

5.2 Attendance Tables

5.2.1 anon_attendance

Stores the attendance marks (2 a day AM/PM) that represent whether the student was present, late, absent, etc. (the marks do go into more detail). This is for all students. Must be used with the attendancecodes table to tell what the mark means.

Column Name	Data Type	Example	Description
id	Int	123456	The id for each attendance record. Used as the primary key for uniqueness.
school_id	Int	3734279	Foreign key that refers to the school this record is from.
date	Date	04/02/2019	Date of this attendance record.
session	Char	"AM"	Whether this record is for the morning or afternoon.
mark	Char	"#"	The mark that was given. Meaning of this is described in the 'attendancecodes' table.
student_id	Int	43027	The id of the student this attendance record refers to.

5.2.2 anon_attendancecodes

Stores the meanings of the marks for each school as each school has different mark to meaning mappings.

Column Name	Data Type	Example	Description
id	Int	234	The primary key which is used as the id for each mark to meaning mapping.
school_id	Int	3734279	Foreign key that refers to the school this record is from.
mark	Char	"#"	The mark code.
description	Char	"Medical/Dental appointments"	Description of what the student was doing.
meaning_description	Char	"Approved Edu. Activity"	The category that the record falls under.
short_meaning_description	Char	"Approved"	The short version of the category that the record falls under.
physical_meaning	Char	"IN"	Whether the student was physically at the school or not. Could be "IN", "OUT", or "LATE".

5.3 Achievement Tables

5.3.1 anon_achievement

Stores the achievements (good behaviour). One achievement can involve many students therefore these are one achievement per instance and students are assigned to these achievement records in the studentachievement table.

Column Name	Data Type	Example	Description
id	Int	234	The primary key which is used as the unique identifier for a particular record in this table.
school_id	Int	3734279	Foreign key that refers to the school this record is from.
achievement_id	Int	138061	The id of this achievement. This is in the scope of the school that the achievement record is from.
conduct_id	Int	138061	The id of the conduct (achievement and behaviours put together). This foreign key will uniquely identify the achievement/behaviour.
achievement_type	Char	"Independence and Initiative"	What the good behaviour is for.
date	Date	"2017-12-01"	When this achievement was observed.
activity	Char	"History"	In what activity the good behaviour was observed. Similar to the 'subject' field.
recorded_by	Char	"{recorded_by_hidden}"	Hidden field. Used to represent the name of the staff member that observed the good behaviour.
recorded_on	Date	"2017-12-01"	When the good behaviour was recorded.
description	Char	"{description_hidden}"	Description of why/how the student got the achievement. Hidden due to personal information given in these descriptions.
subject	Char	"BTEC Sport"	The subject that the achievement was observed in.
category	Char	"Ach"	A label to say that this record is an achievement record.
student_ids	Char	"30617,30648,30459"	A comma separated list of student ids. These are the students that are allocated to this achievement record.

5.3.2 anon_studentachievement

Stores the assignments of students to the achievement records. Every record links one student to one achievement.

Column Name	Data Type	Example	Description
id_big_int	Int	234	The primary key which is used as the unique identifier for a particular record in this table.
school_id	Int	3734279	Foreign key that refers to the school this record is from.
student_achievement_id	Int	138061	The id of student concatenated with the id of the achievement.
achievement_id	Int	138061	The id of the achievement object that this record links the student to.
points	Int	1	How many positive points the student received.
outcome	Char	"Bronze Certificate"	What the outcome of this achievement was, if any.
outcome_code	Char	"BC"	Shortened version of the 'outcome' field.
student_id	Int	43027	The id of the student that is linked to the achievement specified by the 'achievement_id' field.

5.4 Behaviour Tables

5.4.1 anon_behaviour

Stores the behaviour records (bad behaviour). One behaviour can involve many students, therefore these are one behaviour record per instance, students are assigned to these records in the studentbehaviour table.

Column Name	Data Type	Example	Description
id	Int	234	The primary key which is used as the unique identifier for a particular record in this table.
school_id	Int	3734279	Foreign key that refers to the school this record is from.
behaviour_id	Int	138061	The id of this behaviour. This is in the scope of the school that the behaviour record is from.
conduct_id	Cgar	"Beh.198865"	The id of the conduct (achievement and behaviours put together). This foreign key will uniquely identify the achievement/behaviour.
conduct_type	Char	"Disruption"	The type of bad behaviour this is categorised as.
date	Date	"2017-12-01"	When this bad behaviour was observed.
time	Char	"Lesson 5"	Which lesson of the day this bad behaviour occured in.
activity	Char	"History"	In what activity the bad behaviour was observed. Similar to the 'subject' field.
status	Char	"C1 - (Second Verbal Warning)"	In what activity the good behaviour was observed. Similar to the 'subject' field.
location	Char	"In Corridor"	In what activity the good behaviour was observed. Similar to the 'subject' field.
recorded_by	Char	"{recorded_by_hidden}"	Hidden field. Used to represent the name of the staff member that observed the good behaviour.
recorded_on	Date	"2017-12-01"	When the good behaviour was recorded.
description	Char	"{description_hidden}"	Description of why/how the student got the bad behaviour. Hidden due to personal information given in these descriptions.
subject	Char	"BTEC Sport"	The subject that the behaviour was observed in.
category	Char	"Beh"	A label to say that this record is a behaviour record.
student_ids	Char	"30617,30648,30459"	A comma separated list of student ids. These are the students that are allocated to this behaviour record.

5.4.2 anon_studentbehaviour

Stores the assignments of students to the behaviour records.

Column Name	Data Type	Example	Description
id_big_int	Int	234	The primary key which is used as the unique identifier for a particular record in this table.
school_id	Int	3734279	Foreign key that refers to the school this record is from.
student_behaviour_id	Char	"PL-36369Beh.267813"	The id of student concatenated with the id of the behaviour.
behaviour_id	Char	"Beh.267813"	The id of the behaviour object that this record links the student to.
points	Int	1	How many bed behaviour points the student received.
outcome	Char	"Restorative Conversation"	What the outcome of this was behaviour was, if any.
outcome_code	Char	"RESTORA"	Shortened version of the 'outcome' field.
role	Char	"Participant"	What role this student played in this bad behaviour.
student_id	Int	43027	The id of the student that is linked to the behaviour specified by the 'achievement_id' field.

5.5 Exclusion Table

5.5.1 anon_studentexclusions

Stores each student exclusion. One record represents one exclusion.

Column Name	Data Type	Example	Description
id	Int	234	The primary key which is used as the unique identifier for a particular record in this table.
school_id	Int	3734279	Foreign key that refers to the school this record is from.
forename	Char	"{forename_hidden}"	The forename of the excluded student. Hidden due to privacy reasons.
surname	Char	"{surname_hidden}"	The surname of the excluded student. Hidden due to privacy reasons.
exclusion_type	Char	"Lunchtime"	Whether this exclusion is a fixed term (spanning 1 or more days) or a lunchtime exclusion.
exclusion_type_code	Char	"LNCH"	The shorter version of the 'exclusion_type' field.
exclusion_sessions	Int	6	How many sessions the student was excluded for. Morning and afternoon are counted as two separate sessions, therefore 2 sessions a day.
start_date	Date	"2017-12-11"	The day that the student was excluded.
end_date	Date	"2017-12-13"	The day that the exclusion is finished and the student can go back to school.
start_session	Char	"AM"	Whether the student was excluded in the morning or afternoon.
end_session	Char	"PM"	Whether the student was allowed to come back in the morning or afternoon on the 'end_date'.
exclusion_days	Int	5	Number of days that the student has been excluded for.
exclusion_id	Int	345	The id of this exclusion record for the school.
exclusion_reason	Char	"Persistent disruptive behaviour"	Why the student was excluded.
exclusion_session_code	Char	"DB"	Shorter version of the 'exclusion_reason' field.
student_id	Int	43027	The id of the student that has been excluded.