

Have budget spendings on environmental factors helped decrease children obesity in UK?

Introduction

Obesity in children is a growing global problem which according to the UN has reached epidemic levels. In England more than 20% of 6 years old kids are obese. Children obesity is associated with poor physical and mental health, cardio-vascular and other diseases and overall poorer well-being. Immediate action on both government and local levels is needed to fight this problem. It is generally assumed that healthy diet and physical activity are key factors affecting this health condition. However improving environmental aspects could lead to excessive weight too. This report evaluates the relationship between the environmental factors investment and children obesity levels.

Methodology

First, we explore the data presented for the report, check for outlier and normalise it.

Second, we check if there is linear relationship between the money spent on environment and the change in obesity cases using Pearson's correlation.

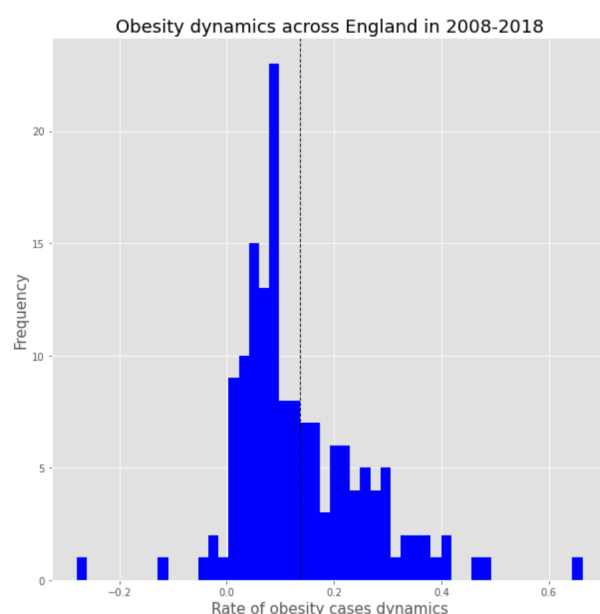
Third, multiple linear regression was used to analyse if change in obesity cases can be predicted based on expenditures allocation.

Data

The dataset provided for this assessment was collected from the London datastore and additionally manipulated for the sake of exemplariness. Information on 152 local authority areas from 9 regions in England is presented in the dataset. For each geographic unit there is information related to the amount of childhood obesity cases: number of cases in 2008, 2013 and 2018, both total and split by genders; and information about average annual spendings in related areas: air quality, cleaning public spaces, training health professionals, awareness in schools, awareness in media, and subsidising counselling services. There are no missing values in the dataset.

Descriptive statistics on obesity cases count per 100,000 people

quantity	152,00
mean	0,14
median	0,10
min	-0,28
max	0,66
LQ	0,06
UQ	0,20
IQR	0,14
variance	0,01
st.dev	0,12



To perform the regression analysis we need to choose one dependent variable representing the children obesity dynamics. Since the budget spendings, being the independent variables, are given as a mean for 2008-2018 years, it is reasonable to use the same time range for the obesity dynamics. Also it is important to normalise cases count per population.

As a result, we get the rate of obesity cases dynamics for 2008-2018. We see that most areas have an increase in normalised obesity cases with a mean increase 0,14 (+14% change from 2008 to 2018).

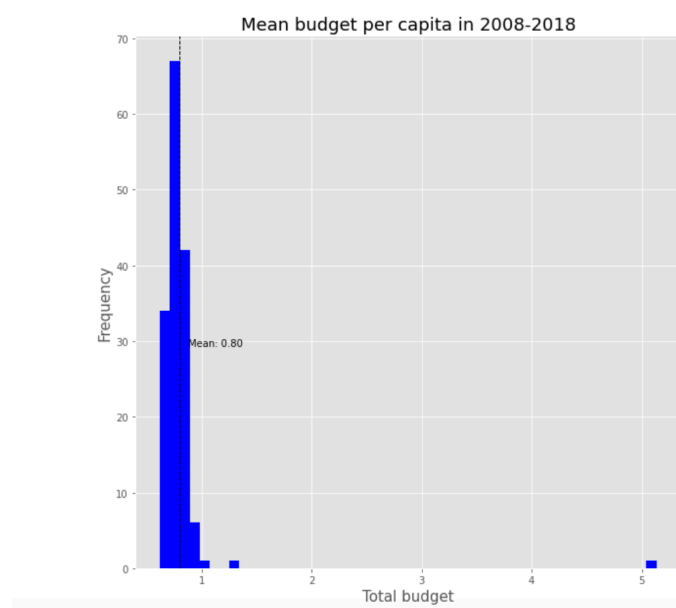
There is one lower outlier (Isles of Scilly with a rate of -0,28) and three upper outliers (Richmond upon Thames, Kingston upon Thames, Kensington and Chelsea, 0,66; 0,47 and 0,47 respectively).

However the outliers are not excluded at this stage due to uncertainty of their origins. It is not obvious if the outliers do not follow the general pattern, are crucial for the analysis or even are caused by errors.

Data about budget spending was normalises per population too.

Descriptive statistics on total budgets per capita

quantity	152,00
mean	0,80
median	0,77
min	0,62
max	5,13
LQ	0,72
UQ	0,82
IQR	0,10
variance	0,13
st.dev	0,36



The Tukey's fences shows there are three upper outliers with one of them being far out from others. City of London's average spending was 5,13 which is more than 6 times higher that average (Fig 1). As we are later using Pearson's correlation which is not robust to outliers, the extreme observation was excluded from the analysis.

We also looked at the distribution of specific expenditures. There is one area where City of London is not ranked the first in spendings. London borough of Haringey has highest media awareness spendings among all other locations. However it is not as far from other observations, so the data point was not excluded.

Next, we examined the distribution again, after excluding City of London and found another outlier in the school awareness distribution (Fig 2). The investment of Isles of Scilly in media was 4.6 times higher than average. Additionally this area did not spend money on anything else. Such a data point would violate the quality of regression model. Therefore it was excluded too. Overall, 150 observations were left for further analysis.

Results

First, we test if there is linear relationship between total budget expenditures and children obesity cases. We use Pearson correlation because both variables are ratio data types.

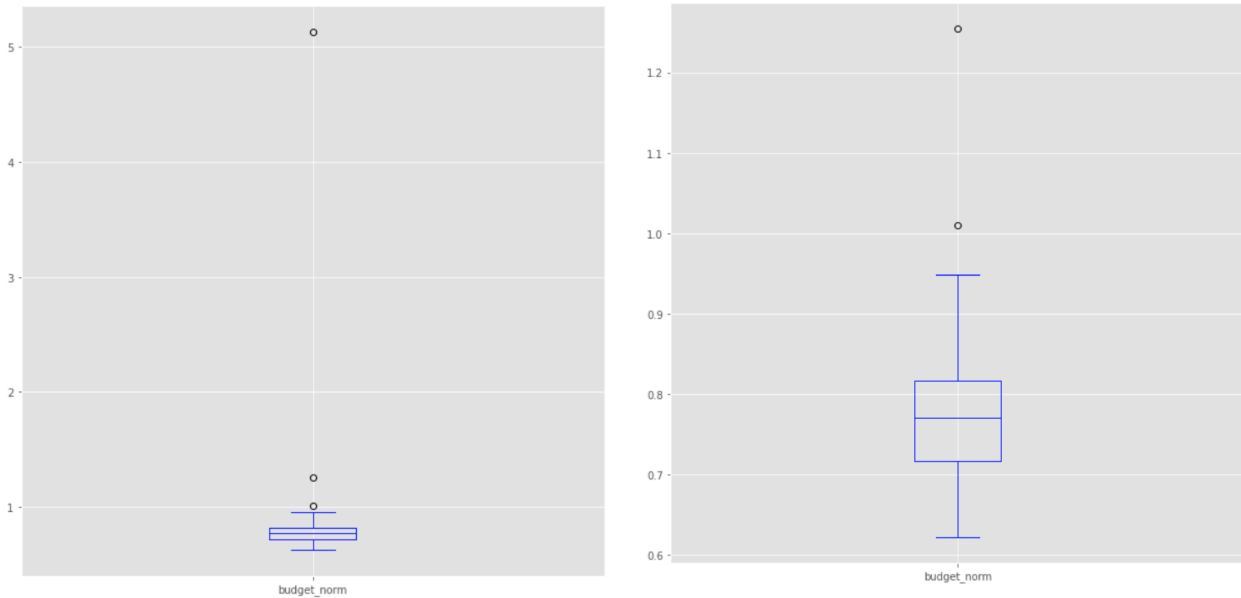


Fig 1. Box plots of total budgets per capita before (a) and after (b) the exclusion of City of London

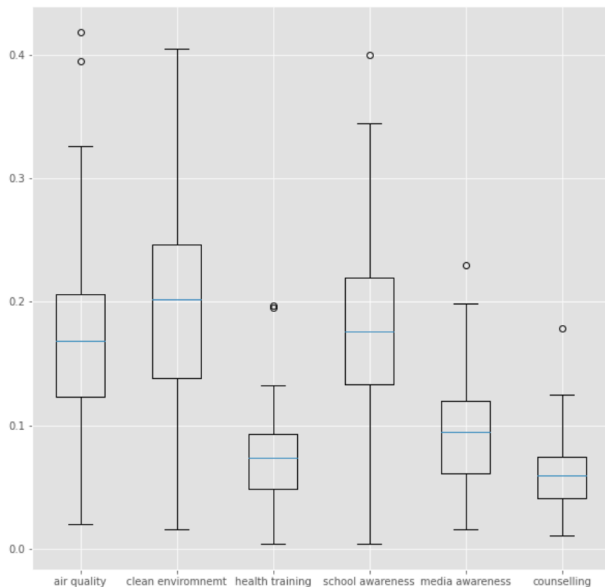


Fig 2. Box plots of six area spendings per capita

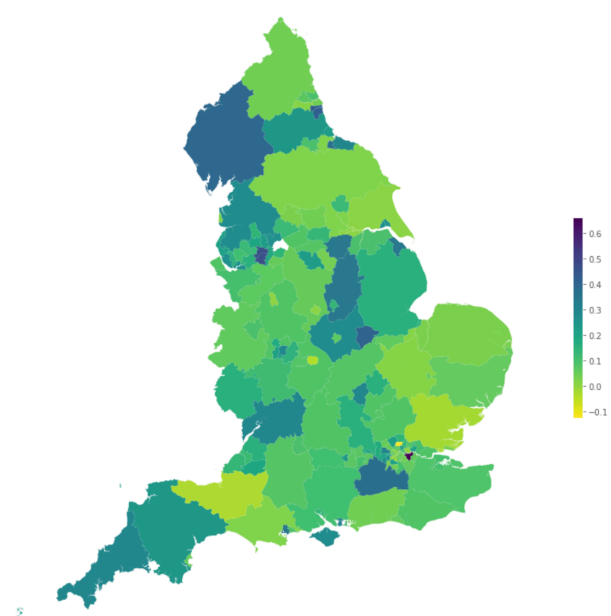


Fig 3. Spatial distribution of obesity change rates

Pearson's correlation r -value is -0.05 , which is very close to zero and indicates very low negative linear association between the two variables (Fig 3). It means there is no linear correlation between overall environment spendings and obesity cases change.

Then multiple regression was fit. Since total budget is the sum of other variables and therefore related to them, it is not used for the regression analysis to avoid multicollinearity. An analysis of independent variables was made before fitting a linear model. To avoid multicollinearity we created a correlation matrix for independent variables to check if there is a strong relationship between any two of them (Fig 4). As none of the correlation coefficients was larger than 0.9 all six variables were used.

The results were the following. The R -squared value is 0.043 , which means that only 4.3% of the total variance of the obesity cases can be explained by the six variables used in the model. The p -values of six variables are all very high and thus statistically insignificant.

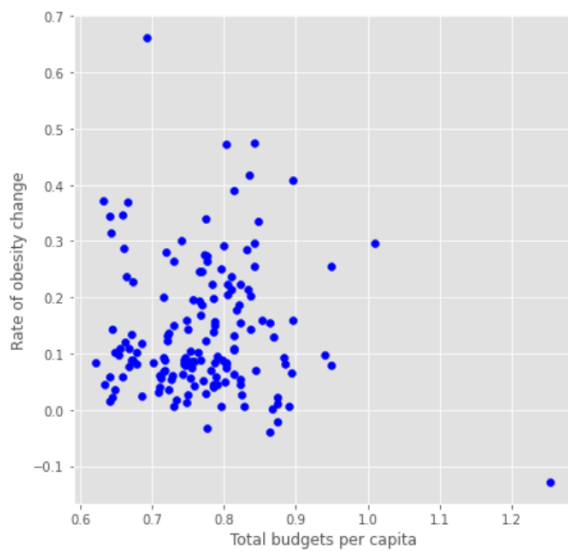


Fig 4. Scatterplot for total budgets and obesity change rate

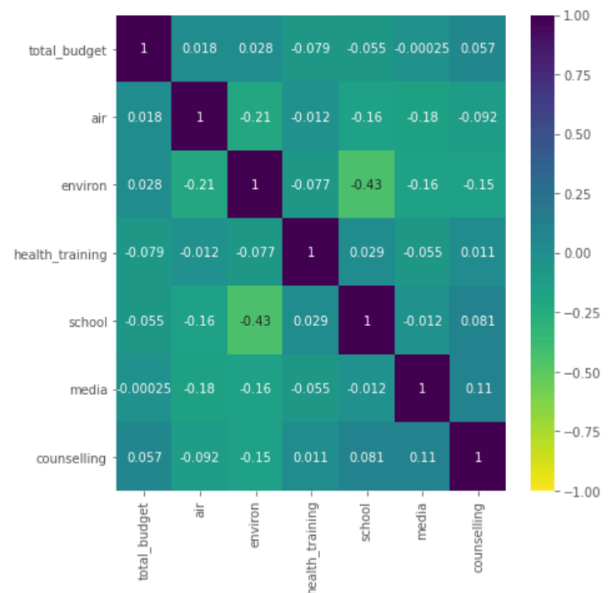
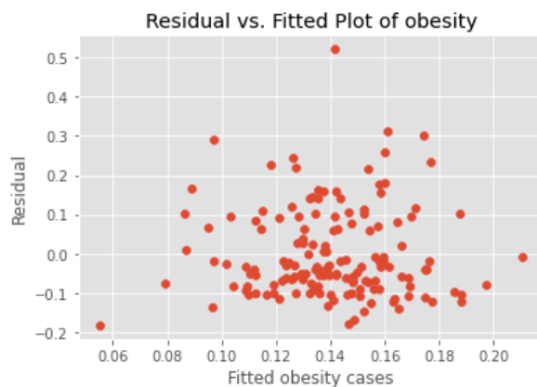


Fig 5. Correlation matrix for independent variables

Variable	p-value
air quality	0.35
cleaning public spaces	0.726
training health professionals	0.629
awareness in schools	0.204
awareness in media	0.202
subsidising counselling services	0.918



Conclusion

This report examined the relationship between the environmental investment and children obesity levels in England during 2008-2018 years. Correlation and linear regression analyses were performed. According to the results, there is not enough evidence to reveal linear associating between budget allocation and the change in obesity levels. Further investigation is needed to find relationship in the data. Additionally, more temporal granularity in data could be of use to estimate if change in investment could affect children obesity.

Word count: 901

Jupyter notebook: https://github.com/alvova/CASA_QM_coursework/blob/master/CW_1.ipynb