From Deprivation to Degree

Exploring the Impact of Local Income Levels on Youth Education in England

Group R you OK?

Table of contents

1 Executive summary **1**

2 Introduction **1**

3 Research Question **1**

4 Data Loading **1**

5 Dataset introduction **2**

5.1 Variables Description 2

5.2 Dataset Description 4

6 Methodology **5**

7 Result **8**

8 Discussion **10**

9 Conclusion **11**

10 Recommendations **11**

11 Citation **12**

1 Executive summary

This report investigates **how town-level income deprivation affects educational attainment among English youth aged 18 to 22**. Using national data, we analyze academic outcomes at three life stages: Level 3 qualification at 18, university entry at 19, and Level 6+ attainment by 22. Results show a clear upward trend—youth from lower deprivation towns consistently outperform those from higher deprivation areas, especially by age 22. These findings highlight structural inequalities and suggest the need for targeted educational support in deprived communities.

2 Introduction

Regional inequality in education remains a pressing issue in England, particularly among young people transitioning from secondary to higher education (The Times, 2024). This study examines **whether income deprivation at the town level impacts academic outcomes for individuals aged 18 to 22**. We focus on three key milestones: Level 3 qualifications at age 18, full-time university participation at age 19, and Level 6 or higher qualification by age 22. These stages reflect national benchmarks for educational progression.

To maintain analytical consistency, cities are excluded due to their distinctive socio-demographic structures, such as dense populations and diverse education systems. We also remove entries with missing or unclear income classifications to ensure accurate results. The dataset is grouped by income level and town size, and analyzed using both summary statistics and visualizations. Trends are observed across large, medium, and small towns.

Findings reveal a consistent association: **as income deprivation decreases, educational attainment increases**. This analysis deepens our understanding of structural disadvantage and supports the development of more equitable, data-driven education policies.

3 Research Question

This study will focus on:

Does the income level of a town affect the educational outcomes of young people?

4 Data Loading

In this study, our group chose to **download the dataset locally** rather than relying on an online dataset. This approach **improves the reproducibility of the analysis** because local data

ensures continuous access to a stable version of the data and avoids remote sources updating or deleting the data resulting in different analysis results. Also, **local datasets are more stable** as they do not rely on a network connection, which avoids interruptions due to application program interface limitations or network problems. It also improves processing speed and supports offline analysis, which is essential for **a reliable and efficient research workflow**.

We use the english_education.csv dataset downloading from the TidyTuesday, originally sourced from The UK Office for National Statistics (ONS). The dataset contains education and deprivation data for over 1,000 English towns from 2012 to 2013 school year, offering town-level observations across key variables such as `income_flag`, `size_flag`, `level_3_at_age_18`, `activity_at_age_19_full_time_higher_education`, and `highest_level_qualification_achieved_by_age_22_level_6_or_above2`. These indicators provide a robust basis for analyzing **educational progression** in relation to **local income deprivation**.

```
edu_df <- read_csv("../data/english_education.csv")
```

```
selected_vars <- c(
  "income_flag",
  "level_3_at_age_18",
  "activity_at_age_19_full_time_higher_education",
  "highest_level_qualification_achieved_by_age_22_level_6_or_above",
  "size_flag"
)

edu_df_selected <- edu_df %>%
  select(all_of(selected_vars))%>%
  rename(
    level3_at_18 = level_3_at_age_18,
    uni_at_19 = activity_at_age_19_full_time_higher_education,
    level6_at_22 = highest_level_qualification_achieved_by_age_22_level_6_or_above
  )
```

5 Dataset introduction

5.1 Variables Description

From Table 1, the variables listed here represent a selected subset of those most relevant to our analysis. The dataset is suitable for **examining how structural income deprivation may**

affect academic progression among young people. The original data has been archived in the project repository and here is the location of the data.

```
edu_variable_desc <- data.frame(
  Variable = c(
    "income_flag",
    "level3_at_18",
    "uni_at_19",
    "level6_at_22",
    "size_flag"
  ),
  Description = c(
    "Categorical income deprivation classification of the town (lower/mid/higher). Used
    ↪  as the independent grouping variable.",
    "Proportion of young people in the 2012/13 cohort who achieved Level 3 qualifications
    ↪  at age 18.",
    "Proportion of the cohort who entered full-time higher education by age 19.",
    "Proportion of the cohort who attained Level 6 or above qualifications (e.g.,
    ↪  Bachelor's degree) by age 22.",
    "Town/city size classification based on 2011 census data (used to identify and
    ↪  exclude cities)."
  )
)

knitr::kable(
  edu_variable_desc,
  col.names = c("Variable Names", "Variable Description"),
  booktabs = TRUE
)
```

Table 1: This table contains the key variables used in the analysis along with their descriptions

Variable Names	Variable Description
income_flag	Categorical income deprivation classification of the town (lower/mid/higher). Used as the independent grouping variable.
level3_at_18	Proportion of young people in the 2012/13 cohort who achieved Level 3 qualifications at age 18.

```
The dataset contains a total of `r nrow(edu_df)` observations and `r ncol(edu_df)`variables.
```

Figure 1: Inline R code

Table 2: First 2 rows of the data

```
head(edu_df_selected, 2)
```

```
# A tibble: 2 x 5
  income_flag              level3_at_18 uni_at_19 level6_at_22 size_flag
  <chr>                           <dbl>     <dbl>        <dbl> <chr>
1 Higher deprivation towns         50.8      30.8         NA   Small Towns
2 Mid deprivation towns            60.1      41.8         33.3 Small Towns
```

Table 1: This table contains the key variables used in the analysis along with their descriptions

Variable Names	Variable Description
uni_at_-19	Proportion of the cohort who entered full-time higher education by age 19.
level6_-at_22	Proportion of the cohort who attained Level 6 or above qualifications (e.g., Bachelor's degree) by age 22.
size_flag	Town/city size classification based on 2011 census data (used to identify and exclude cities).

5.2 Dataset Description

The dataset contains a total of 1104 observations and 31variables.

See Figure 1 for the detail of the *Inline R code* used.

See Table 2 for the detail of the first 2 rows of the data and the types of variables in the dataset.

6 Methodology

```r
edu_df_selected %>%
  count(income_flag) %>%
  mutate(percentage = round(n / sum(n) * 100, 1)) %>%
  knitr::kable(
    col.names = c("Income Flag", "Count", "Percentage (%)"),
    digits = 1
  )
```

Table 3: Proportion of Each Income Flag Category

Income Flag	Count	Percentage (%)
Cities	18	1.6
Higher deprivation towns	433	39.2
Lower deprivation towns	444	40.2
Mid deprivation towns	205	18.6
NA	4	0.4

While reviewing the data (Table 3), our group discover that there is **missing data** (`NA` value) in the income categories `income_flag`. In order **not to affect the accuracy of the subsequent analysis**, this missing data will not be included in the analysis. Also, since the sample in `Cities` of `income_flag` is **too small and less representative**, it is removed in the subsequent analysis.

```r
edu_df_clean <- edu_df_selected %>%
  filter(!is.na(income_flag) & income_flag != "Cities") %>%
  mutate(income_flag = factor(income_flag,
                              levels = c("Higher deprivation towns",
                                         "Mid deprivation towns",
                                         "Lower deprivation towns")))
```

```r
table1 <- edu_df_clean %>%
  group_by(income_flag) %>%
  summarise(
    mean_level3 = mean(level3_at_18, na.rm = TRUE),
```

```
    sd_level3 = sd(level3_at_18, na.rm = TRUE),
    mean_uni19 = mean(uni_at_19, na.rm = TRUE),
    sd_uni19 = sd(uni_at_19, na.rm = TRUE),
    mean_level6 = mean(level6_at_22, na.rm = TRUE),
    sd_level6 = sd(level6_at_22, na.rm = TRUE)
  ) %>%
  rename(
    "Income Group" = income_flag,
    "Mean: Level 3 at 18" = mean_level3,
    "SD: Level 3 at 18" = sd_level3,
    "Mean: HE at 19" = mean_uni19,
    "SD: HE at 19" = sd_uni19,
    "Mean: Level 6 at 22" = mean_level6,
    "SD: Level 6 at 22" = sd_level6
  )

kable(table1, digits = 2) %>%
  kable_styling(full_width = FALSE)
```

Table 4: Mean and SD of Educational Outcomes by Income Group(Cities Exclud

Income Group	Mean: Level 3 at 18	SD: Level 3 at 18	Mean: HE at 19	SD: HE at 19	Mean:
Higher deprivation towns	41.59	7.35	27.13	6.69	
Mid deprivation towns	47.82	7.59	31.32	7.20	
Lower deprivation towns	57.96	9.68	40.78	10.33	

Table 4 shows that **education outcomes improve significantly at all measured stages as income levels rise**.

For instance, the mean proportion of individuals achieving Level 3 qualifications by age 18 rises from **41.59%** in **higher deprivation** towns to **57.96%** in **lower deprivation** towns. Similar upward trends are observed for higher education participation at age 19 (**27.13%** to **40.78%**) and for achieving Level 6 qualifications by age 22 (**23.99%** to **35.68%**).

The standard deviations are relatively consistent across groups. This suggests that while variation exists, the overall improvement in outcomes is not due to a small number of high performers.

```r
edu_long <- edu_df_clean %>%
  pivot_longer(
    cols = c(level3_at_18,
             uni_at_19,
             level6_at_22),
    names_to = "education_stage",
    values_to = "value"
  )

edu_long$income_flag <- recode(edu_long$income_flag,
  "Higher deprivation towns" = "Higher\ndeprivation\ntowns",
  "Mid deprivation towns" = "Mid\ndeprivation\ntowns",
  "Lower deprivation towns" = "Lower\ndeprivation\ntowns"
)

edu_long$education_stage <- recode(edu_long$education_stage,
  "level3_at_18" = "Level 3 at 18",
  "uni_at_19" = "HE at 19",
  "level6_at_22" = "Level 6 at 22"
)

ggplot(edu_long, aes(x = income_flag, y = value, fill = income_flag)) +
  geom_boxplot() +
  facet_wrap(~education_stage, scales = "free_y") +
  labs(x = "Income Group", y = "Proportion") +
  theme_minimal() +
  theme(axis.text.x = element_text(size = 9, lineheight = 1.1))
```

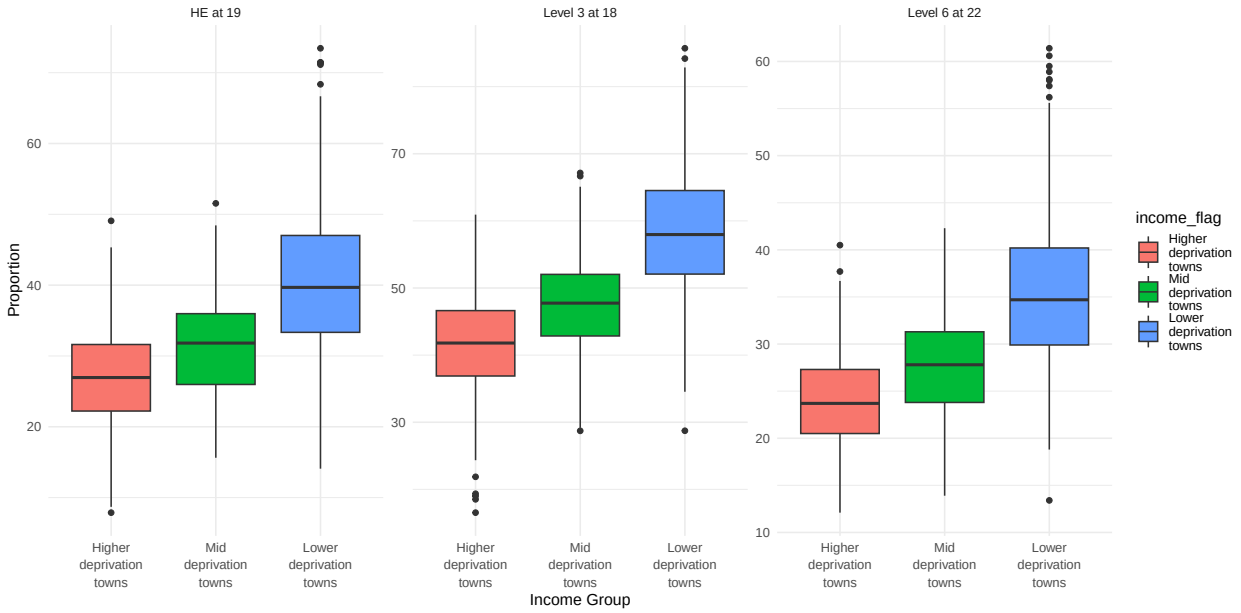


Figure 2: Educational Outcomes by Income Group(Cities Excluded)

Figure 2 illustrates **a clear upward trend in educational attainment as town income levels increase**. In all three stages, young people in lower deprivation towns consistently outperform their peers in higher deprivation towns.

Specifically, the **median** proportions of youth achieving Level 3 qualifications at age 18, participating in full-time higher education at age 19, and attaining Level 6 qualifications by age 22 are **all higher in towns with lower deprivation**. The **interquartile ranges** are also **narrower** in **lower deprivation** towns, suggesting more consistent performance within these areas.

In contrast, **higher deprivation** towns exhibit not only **lower median outcomes** but also **greater variability**, as reflected by the wider spread of the boxplots and more outliers.

7 Result

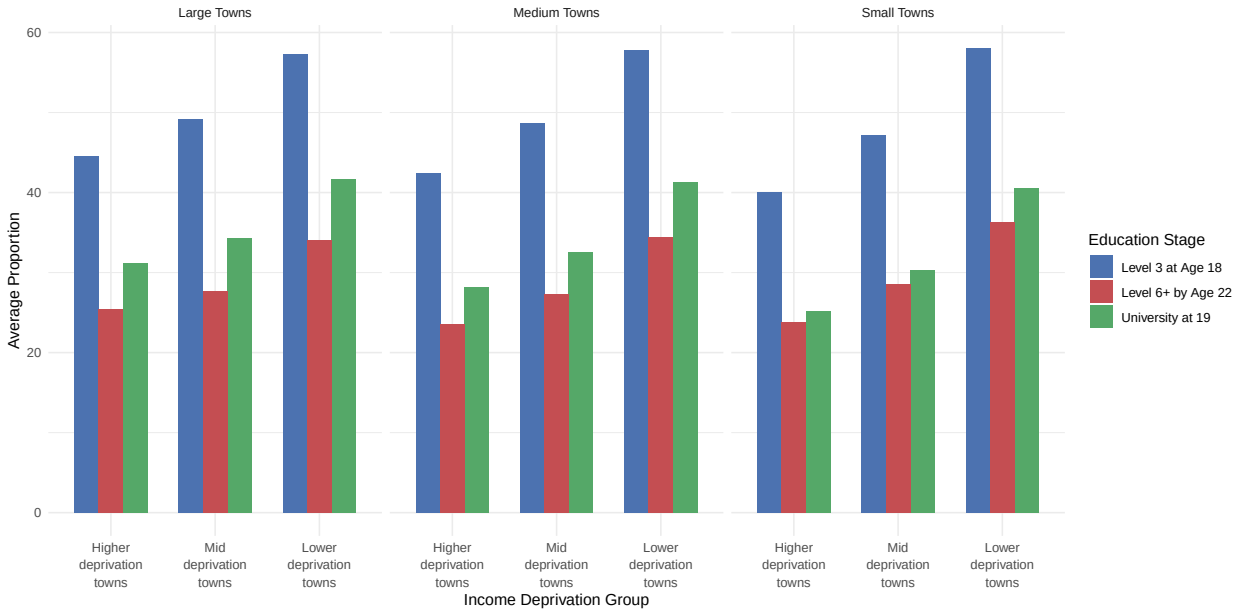
```
summary_df <- edu_df_clean %>%
  group_by(income_flag, size_flag) %>%
  summarise(
    level3 = mean(level3_at_18, na.rm = TRUE),
    uni19 = mean(uni_at_19, na.rm = TRUE),
    level6 = mean(level6_at_22, na.rm = TRUE)
  ) %>%
```

```r
  pivot_longer(
    cols = c(level3, uni19, level6),
    names_to = "education_stage",
    values_to = "mean_value"
  )

summary_df$education_stage <- recode(summary_df$education_stage,
  "level3" = "Level 3 at Age 18",
  "uni19" = "University at 19",
  "level6" = "Level 6+ by Age 22"
)

summary_df$income_flag <- recode(summary_df$income_flag,
  "Higher deprivation towns" = "Higher\ndeprivation\ntowns",
  "Mid deprivation towns" = "Mid\ndeprivation\ntowns",
  "Lower deprivation towns" = "Lower\ndeprivation\ntowns"
)

ggplot(summary_df, aes(x = income_flag, y = mean_value, fill = education_stage)) +
  geom_col(position = "dodge", width = 0.7) +
  facet_wrap(~size_flag) +
  labs(
    x = "Income Deprivation Group",
    y = "Average Proportion",
    fill = "Education Stage"
  ) +
  scale_fill_manual(
    values = c(
      "Level 3 at Age 18" = "#4C72B0",
      "University at 19" = "#55A868",
      "Level 6+ by Age 22" = "#C44E52"
    )
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(size = 9, lineheight = 1.1))
```



(a) Comparison of educational outcomes across income levels and town sizes

Figure 3: Educational Attainment by Income Deprivation Level and Town Size (Cities Excluded)

As shown in Figure Figure 3, **educational attainment increases consistently as income deprivation decreases, across all town sizes**. Towns with **lower deprivation levels** demonstrate the **highest average proportions** of students achieving Level 3 qualifications at age 18, entering university at age 19, and obtaining Level 6 or higher qualifications by age 22.

This pattern holds for large, medium, and small towns, though the gap between high and low deprivation groups appears most pronounced in small towns, particularly at the Level 6+ stage. In small towns, the difference in university attainment between high and low deprivation groups exceeds 10 percentage points.

These results **reinforce a strong association between local income levels and youth educational outcomes**, and suggest that deprivation may have an even stronger impact in smaller towns.

8 Discussion

Our analysis shows a clear trend: **towns with lower levels of income deprivation consistently achieve better educational outcomes at age 18, 19 and 22**. This pattern applies to towns of all sizes and is **particularly evident in smaller towns**, where the gap in Level 6 qualifications is most pronounced. These results are **consistent with recent ONS research** suggesting that

towns with lower levels of deprivation may provide a more conducive environment for academic success.

However, this study has **several limitations** that cannot be ignored. First, as an observational analysis, it **cannot establish a clear causal relationship**. Although there is a clear strong link, we cannot conclude that income poverty is a direct factor in lower educational attainment. Second, the measurement of poverty relies on broad categorical groupings, which may **obscure important local-level variations and nuances**. Third, we exclude cities from the analysis. While this is for consistency, this choice **limits the generalizability of the findings to urban populations**, since they often face unique educational challenges. Finally, the dataset represents a single cohort of students from the 2012/13 school year and may **not reflect evolving trends or long-term patterns in educational progress**. These limitations highlight the need for a more comprehensive, longitudinal and contextualize

Despite these limitations, the findings highlight clear structural inequalities. This points to the **need for targeted policy interventions in poorer towns**. Future research should incorporate more detailed longitudinal data to better understand the underlying causes.

9 Conclusion

This study found a clear relationship between **income deprivation** and **young people's educational attainment** at town level in England. At all three stages of education, young people from **less deprived towns** and cities consistently **outperformed** those from more deprived areas. This trend exists across towns and cities of all sizes, and is particularly evident in **smaller towns**. While the findings are consistent with national statistics and highlight structural inequalities, **caution should be exercised in interpreting the results** due to the observational nature and limited scope of the study.

10 Recommendations

To address the disparities observed, education policy should **prioritize support for students in high-deprivation towns**. This may include:

- Enhance educational funding and targeted support in high-deprivation towns
- Expand access programs to encourage university progression
- Better access to higher education pathways
- Integrate deprivation indicators into education planning and policy to ensure equitable resource distribution

In addition, **future research** should expand the study's dataset. There is a need to include **multiple clusters and urban areas** and to **integrate qualitative indicators** such as school quality and student support services as study variables. These efforts will contribute to **a more comprehensive understanding of how income poverty affects educational trajectories** and inform the development of more equitable education strategies.

11 Citation

R and R Package

1. R Core Team (2025). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
2. Wickham H, François R, Henry L, Müller K, Vaughan D (2023). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.4, https://CRAN.R-project.org/package=dplyr.
3. H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.
4. Wickham H, Hester J, Bryan J (2024). *readr: Read Rectangular Text Data*. R package version 2.1.5, https://CRAN.R-project.org/package=readr.
5. Wickham H, Vaughan D, Girlich M (2024). *tidyr: Tidy Messy Data*. R package version 1.3.1, https://CRAN.R-project.org/package=tidyr.
6. Xie Y (2024). *knitr: A General-Purpose Package for Dynamic Report Generation in R*. R package version 1.49, https://yihui.org/knitr/.

   Yihui Xie (2015) Dynamic Documents with R and knitr. 2nd edition. Chapman and Hall/CRC. ISBN 978-1498716963

   Yihui Xie (2014) knitr: A Comprehensive Tool for Reproducible Research in R. In Victoria Stodden, Friedrich Leisch and Roger D. Peng, editors, Implementing Reproducible Computational Research. Chapman and Hall/CRC. ISBN 978-1466561595
7. Zhu H (2024). *kableExtra: Construct Complex Table with 'kable' and Pipe Syntax*. R package version 1.4.0, https://github.com/haozhu233/kableExtra, http://haozhu233.github.io/kableExtra/.

Data Used

1. Office for National Statistics. (2023, July 25). *Why do children and young people in smaller towns do better academically than those in larger towns?* https://www.ons.gov.uk/peoplepopulationandcommunity/educationandchildcare/articles/whydochildrenandyoungpeopleinsmallertownsdobetteracademicallythanthoseinlargertowns/2023-07-25

2. The Times. (2024, August 14). University offers at record high as inequality in degree entry grows. Retrieved from https://www.thetimes.co.uk/article/university-offers-at-record-high-as-