University of East London

Assignment Title: Socio-economic Determinants of COVID-19 Mortality Across England

Course: DS7006 – Quantitative Data Analysis

Instructor: Dr. Yang Li

Student Name: Keshav Poudel

Student ID: 3064995

Submission Date: Jan 5, 2026

Acknowledgement

Abstract

The research examines how English local authority COVID-19 death rates relate to social economic factors and population characteristics through analysis of combined COVID-19 death records and census-based statistics. The research used a standardized analytical process which started with data exploration followed by correlation and partial correlation analysis and then Principal Component Analysis (PCA) and multiple linear regression. The analysis of COVID-19 deaths per 1,000 residents showed normal distribution patterns which validated the application of parametric statistical methods.

The research data showed that COVID-19 death rates strongly linked with social economic status through low educational levels. The PCA results showed two main factors which included professional and educated advantages and labour-market disadvantages. The percentage of residents without qualifications emerged as a significant factor which predicted higher death rates even when researchers controlled for age and occupation and deprivation levels. The population aged 85 and above demonstrated significant mortality risk because of their biological susceptibility to the disease. The research results showed that housing challenges produced minimal yet positive effects on the data.

The research results confirm that COVID-19 spread through existing social economic gaps instead of impacting all communities at the same rate. The research faces two main limitations because it uses ecological methods and depends on census data from 2011. The research demonstrates that COVID-19 death rates in England were primarily influenced by permanent social economic factors which included educational levels and employment status.

Contents

\

List of Figures

List of Tables

Section 1: Introduction

The SARS-CoV-2 virus which causes COVID-19 primarily spreads through respiratory droplets and aerosols that escape from infected people during breathing and speaking and when they cough or sneeze. The risk of transmission becomes extremely high when people stay in closed spaces with inadequate ventilation while being in close proximity to each other. The majority of transmission happens through physical contact between people who live in the same household and work together and share common areas. The first pandemic waves in England produced high death numbers which showed significant differences between local areas despite all national control measures and public health guidelines (World Health Organization, 2021).

The Office for National Statistics (ONS) reported that COVID-19 deaths occurred at elevated levels in areas with lower socioeconomic status than in affluent regions which demonstrated significant geographic health disparities across England (Office for National Statistics, 2020). The pandemic revealed existing structural inequalities between communities instead of creating equal harm for all populations. Research indicates that people aged 70 and older face elevated risks of developing severe COVID-19 complications which result in death with the oldest group (85+) showing the highest mortality rate. The way people worked at their jobs determined their chances of getting infected because remote work options were unavailable to them. People who worked in healthcare and social care and transport and retail and public-facing roles faced higher exposure risks because their jobs did not allow remote work. People from disadvantaged economic backgrounds became more susceptible to harm because of their social status. Residents of low-income neighbourhoods face elevated health dangers because they earn less money and have restricted employment options which increases their COVID-19 risk and decreases their ability to protect themselves (Office for National Statistics, 2022).

The level of educational achievement between people strongly influenced the results. People with lower education levels tend to experience worse health results while facing higher economic risks and needing to work in dangerous jobs which increases their risk of dying from COVID-19. The Deprivation index combines multiple disadvantage factors which include poverty levels and educational limitations and housing difficulties and work availability issues. Studies have proven that areas with higher deprivation levels consistently reported elevated death rates which demonstrates why researchers should focus on deprivation as their primary variable for COVID-19 outcome analysis (Public Health England, 2020).

The research investigates how COVID-19 mortality rates relate to population characteristics and poverty levels and educational attainment and employment patterns across 296 English local authorities. The research team performed local authority code matching to unite datasets before executing population rate standardization through calculations per 1,000 residents. The analysis method allows researchers to evaluate different population-sized areas through standardized evaluation procedures (The King's Fund, 2020).

The research project determines which social economic factors most affected COVID-19 death rates among 1,000 population members through exploratory data analysis and correlation tests and principal component analysis (PCA) and multiple regression modelling.

Section 2: Literature Review

The research investigates how age and education levels and deprivation status and work activities affect COVID-19 results in England. The research variables follow the same pattern as the guiding factors which include age ranges and education levels and deprivation indicators and occupational distribution across local authority areas. The research evidence demonstrates that COVID-19 outcomes in England followed specific social and demographic patterns which stem from existing social inequalities.

Age

The Office for National Statistics data shows that age stands as the primary factor which determines COVID-19 death rates in England. The Office for National Statistics reports that death rates rose dramatically starting at age 65 before reaching their peak in people who were 75 years old or older. The pandemic data shows that children and young adults faced minimal risk of death which proves that age remains the main factor determining survival rates (Office for National Statistics, 2022). The UK Health Security Agency together with the Department of Health and Social Care conducted a joint study which demonstrated that people aged 75 and above accounted for most confirmed deaths during all major waves starting from 2020 until 2022 including waves that happened after vaccine distribution began. The study demonstrates that older adults maintained elevated risk levels despite advances in medical treatments. (Department of Health and Social Care and UK Health Security Agency, 2022).

The pattern of clinical data matches the overall trend. The Intensive Care National Audit and Research Centre reported that patients aged 75 and older who entered critical care facilities had significantly lower survival rates than patients under 65 years old (Intensive Care National Audit and Research Centre, 2022). Public Health England identified age as the main risk factor for COVID-19 during the initial pandemic period because it most affected people who were 80 years or older (Public Health England, 2020).

The census age bands have been organized into four categories which align with established clinical and population trends. This evidence substantiates the incorporation of aggregated age-group proportions as explanatory variables in the regression analysis.

Education

The level of education people achieve determines their COVID-19 risk because it affects their employment stability and earnings and workplace environment and their capacity to follow health guidelines. The Marmot Build Back Fairer Review demonstrated that workers with lower educational attainment tend to perform dangerous jobs which require physical contact with others while receiving minimal compensation. The working environments of these communities restricted their ability to self-isolate while making them more susceptible to virus transmission throughout England (Institute of Health Equity, 2020).

The Health Foundation documented identical employment patterns. Workers who held lower qualifications ended up in jobs which demanded physical attendance at work sites while

blocking remote work possibilities. The combination of increased workplace exposure and restricted protective measures made these groups more vulnerable to infection (The Health Foundation, 2021).

Research conducted at local authority level demonstrated that areas with higher numbers of unqualified adults showed elevated COVID-19 death rates. The study maintained its findings about education vulnerability after researchers controlled for age distribution and additional variables which confirmed education plays a distinct role in determining risk levels (Sun, 2021). The ONS disability and COVID-19 mortality study applied education level as a main factor to discover that people with lower educational qualifications faced elevated mortality risk. The study supports existing research which demonstrates that England experiences social inequality (Office for National Statstics, 2020) .

The research uses educational attainment levels per 1,000 residents to measure variations between local authorities. Education variables are incorporated into the regression model to elucidate structural socio-economic disparities in COVID-19 mortality among local authorities.

Deprivation

The English population shows that COVID-19 mortality rates directly correlate with the level of deprivation. The national review by Public Health England revealed that death rates in areas with the highest deprivation levels exceeded those in areas with the lowest deprivation by more than two times. The distinct social pattern demonstrates how environmental conditions and available resources affect health results (Public Health England, 2020). The Office for National Statistics validated these results through their analysis of death statistics based on local areas and levels of deprivation. The most deprived local authorities showed elevated mortality rates which persisted when researchers analysed small neighbourhoods from the same geographic area. (Office for National Statistics, 2020).

The Health Foundation explained why people in deprived areas face higher risks of infection. People who live in crowded homes and work without job security and earn low wages while using public transportation face higher risks of infection because these conditions increase their exposure and decrease their ability to stay safe from infection. The organization stated that these patterns stem from enduring social inequalities across England instead of personal decisions (The Health Foundation, 2020). The Marmot Review revealed that deprived areas across England experienced the most severe impact from the pandemic throughout its duration (Institute of Health Equity, 2020).

The study uses census data from education and occupation fields because the Index of Multiple Deprivation data is incomplete. Education and occupation indicators are used as proxy measures of deprivation in the regression analysis.

Occupation

The risk of COVID-19 exposure depends on occupation because essential and lower-paid positions force workers to interact with others in person. The Office for National Statistics

conducted an analysis which revealed that male workers in elementary and caring and leisure and service occupations faced higher mortality rates than workers in managerial and professional positions. The research demonstrates how different workplaces create unequal risks for COVID-19 exposure throughout England (Office for National Statistics, 2021). The ONS conducted research to determine which jobs exposed workers to the virus at the highest risk. The study revealed that healthcare personnel along with social care providers and cleaning staff and transportation personnel and retail sales personnel faced the highest virus exposure. The study showed that most professional positions had minimal virus exposure because employees could work from home (Office for National Statistics, 2020).

The Institute for Fiscal Studies demonstrated that frontline positions and non-remote work sectors employed most of the lower-wage staff. The ability to work from home became more prevalent among higher-paid professionals which resulted in two distinct groups facing different levels of infection risk (Institute for Fiscal Studies, 2020). The Health Foundation demonstrated that workers in entry-level positions received the lowest financial protection while facing maximum pandemic risk which worsened social disparities (The Health Foundation, 2021).

The research uses local authority occupational structure variations to explain why COVID-19 mortality rates differ between areas. So, occupational structure variables are used as predictors in the multivariate regression model.

Summary

The research indicates that England experiences the highest risk of severe illness and death from COVID-19 based on age factors. The ability to follow protective measures and exposure to the virus depends on education level and social status while occupation determines the amount of contact people have with others. The different factors created unequal mortality rates throughout England. The analysis depends on age data and education levels and deprivation indicators and occupational data per 1,000 residents because medical indicators and complete deprivation indexes are absent from the dataset. These variables constitute the foundation of the regression analysis employed to elucidate the variation in COVID-19 mortality among English local authorities.

Section 3: Research Questions

The project follows these research questions:

1. Age: How does COVID-19 mortality, measured as deaths, per 1,000 residents relate to the proportion of residents in the four age groups used in this project: Children (0–14) Adults (15–64) Elders (65–74) and Seniors (75+)? Which age group shows the association with COVID-19 mortality, across local authorities?

2. Education: whether the authorities, with a share of residents who have no qualifications experience higher COVID-19 death rates? Do the local authorities with a proportion of people, with Level 4 or higher qualifications have mortality?

3. Deprivation: The indicators that link to deprivation include education level and more people working low-pay jobs. How do the indicators help explain the differences, in COVID-19 deaths, between authorities?

4. Occupation: How does the mix of occupations the number of residents who work in caring, elementary or other public-facing jobs compared with the number of residents who work in managerial jobs affect COVID-19 death rate?

5. Combined Socio-economic Effect: When we look at the age the education, the deprivation-related factors and the occupational structure, in the regression model, which of the variables age, education, deprivation-related factors and occupational structure remain significant predictors of the COVID-19 deaths, per 1,000 residents?

Section 4: Methodology

4.1. Data Sources & Data Preparation

The research design of this project follows an ecological approach because it analyses each English local authority as an individual analytical unit. The DS7006 Moodle site provided COVID-19 mortality data through a CSV file which contained total deaths from COVID-19. The socio-economic information retrieved from NOMIS through Census 2011 tables which included data about age distribution, deprivation levels, occupational categories and educational qualifications. The LA_code field in each dataset allowed to link local authorities between different tables. The individual CSV files created for each thematic category:

- The research data included age distribution information which showed resident numbers for five-year age ranges starting from 0–4 years old to 85 years and above.
- The research data included information about households that faced deprivation in one or more categories.
- The research data included resident numbers who worked in specific occupational categories that followed standard classification systems.
- The research data presented resident numbers based on their highest educational achievement which spanned from no qualifications to Level 4 and above.

Variables and Measurement

The English local authority is the main thing this study looks at. The dependent variable is the mortality rate of COVID-19, quantified as deaths per 1,000 residents. Independent variables encompass age distribution (Children 0–14, Adults 15–64, Elders 65–74, Seniors 75+ per 1,000 residents), educational qualifications (No qualifications and Level 4 or above per 1,000 residents), occupational categories (specific occupational groups per 1,000 residents), and deprivation-related metrics obtained from census data. To make sure that all local authorities can be compared, all variables are continuous numeric rates per 1,000 residents.

Harmonising Local Authority Names and Codes

The research required extensive data cleaning because multiple English local authorities underwent boundary changes since the 2011 Census. The SQL system enabled researchers to update local authority names and codes for achieving uniformity between all datasets. The research included multiple local authority reorganization events which affected the following areas:

- Somerset West and Taunton,

- North and West Northamptonshire,

- Dorset,

- Bournemouth, Christchurch and Poole,

- North Yorkshire and Cumbria,

- Several smaller district merges.

The research used contemporary LA_code values which matched the same geographic areas in all datasets. The Appendix has full SQL scripts that were used to make local authorities more similar.

Merging the Datasets

The COVID-19 death table underwent harmonization before researchers performed a left-join operation with the age and deprivation and occupation and education tables.

The harmonized COVID-19 death data were combined with data on age, poverty, job, and education, using local authority codes as the common link. The process resulted in a single combined dataset named covid_19_merged which included all variables for each local authority.

Age Aggregation into Analytical Groups

The raw 18 age bands received SQL processing to create four interpretable categories:

- **Children (0–14)**

- **Adults (15–64)**

- **Elders (65–74)**

- **Seniors (75+)**

The research used these combined age categories instead of specific age ranges for all following studies. This grouping makes it easier to understand and lowers the correlation between age groups that are next to each other.

Standardisation of All Variables

The analysis required population size standardization through rate calculation per 1,000 residents for all person-based counts. The analysis applied this standardization to all data points including:

- COVID-19 deaths,
- Children, Adults, Elders and Seniors,
- All deprivation indicators,
- All occupation groups,
- All education categories.

Final Modelling Dataset

The final dataset (u3064995_DS7006_CW2_data.csv) included the following information:

- deaths per 1,000 residents
- Children, Adults, Elders and Seniors per 1,000

- deprivation levels per 1,000
- occupation structure per 1,000
- education structure per 1,000

The data served as input for R statistical modelling.

4.2. Techniques

The following section explains the research methods which follow the DS7006 Session 9 and 10 teaching guidelines:

Exploratory Data Analysis (EDA)

The analysis of COVID-19 death rates per 1,000 residents used Exploratory Data Analysis to discover their distribution patterns and behavioural patterns. We used Exploratory Data Analysis to look at the distributional characteristics, find possible outliers, and see if the assumptions for parametric regression were correct. The analysis established to check if linear modelling requirements were valid for the data.

Statistical Tests

Multiple statistical tests were conducted to analyse both group variations and data relationships in the collected information. Spearman's rank correlation served to evaluate the relationship between COVID-19 mortality and all predictor variables. The analysis employed this method due to its efficacy in managing data that does not conform to a normal distribution. The analysis utilized partial correlations to assess the persistence of education-mortality relationships after the introduction of age as a controlling variable, while employing Spearman's rank correlation for data that deviated from normal distribution or exhibited monotonic patterns rather than linear relationships. Using a variety of statistical tests, test are able to identify trends in the distribution of deaths across various regions and determine the strength of the associations between COVID-19 deaths and social and economic parameters. These statistical tests were not independent inferential judgments; rather, they were exploratory tools to help pick variables and then perform multivariate modelling.

KMO and MSA Checks for PCA

Prior to doing Principal Component Analysis, the Kaiser-Meyer-Olkin (KMO) measurement and assessment for each Measure of Sampling Adequacy (MSA) were required. The tests evaluated if the variables shared enough common structure to perform PCA effectively. The analysis excluded variables with low MSA values to achieve better results.

Principal Component Analysis (PCA)

The analysis employed PCA to diminish the dimensions of socio-economic data while uncovering essential patterns. The analysis followed these steps:

1. The analysis standardized all variables before starting.

2. The analysis performed PCA with varimax rotation as its method.

3. The analysis used scree plots together with parallel analysis to determine the optimal number of components.

4. The analysis interpreted each component through its variable loading values.

Multiple Regression Modelling

Multiple linear regression was used in the study to determine which social characteristics predicted COVID-19 death rates per 1,000 people the best. Multiple regression modelling was used to examined the impact of socioeconomic characteristics on COVID-19 deaths while accounting for any potentially ambiguous correlations between the predictors. For the modelling procedure, the following task were required:

- PCA results, correlation data, and theoretical knowledge were utilized to select non-redundant predictors.
- Variance Inflation Factor (VIF) computations were used in the analysis to search for multicollinearity.
- To exclude predictors from the model that weren't required, the analysis employed an iterative procedure.

Section 5: Implementation

5.1. Data loading and preparation

All analyses were performed in R utilizing a normalized dataset exported from a SQLite database. The analytical approach employed standard statistical tools such as dplyr for data manipulation, GGally for statistical correlation analysis, psych for principal component analysis, ppcor for partial correlation analysis, and car for regression diagnostics. The analytical workflow employed standard statistical tools such as dplyr to perform data manipulation, GGally for exploratory correlation analysis, psych for principal component analysis, ppcor for partial correlation analysis, and car for regression analyses.

The final dataset has the number of COVID-19 deaths per 1,000 people living in each English local authority. It also has socio-economic indicators that show the age structure, levels of deprivation, the types of jobs people have, and how well they do in school. There are 296 local authorities in the dataset, which means that it covers the whole country at the local authority level. To get a basic idea of how COVID-19 death rates were spread out among local authorities, summary statistics were used. These numbers showed that the death rates were only slightly different and there were no signs of extreme outliers. This means that parametric analysis will work well. The analysis employs the following variables:

- COVID-19 mortality rate per 1,000 residents (dependent variable),
- Age distribution metrics (Children, Adults, Elders, Seniors per 1,000 residents),
- Indicators of deprivation,
- Indicators of occupational structure,
- Indicators of educational attainment.

To make sure that local authorities with different population sizes could compare, all person-based variables were standardized to 1,000 residents.

5.2. Exploratory Data Analysis and Normality Assessment

Exploratory data analysis was conducted to examine the distributional characteristics of the dependent variable and to determine if the prerequisites for parametric statistical methods were met. Visual diagnostic tools and formal normality tests were used to look at the shape, distribution, and symmetry of COVID-19 death rates among different local governments. Graphical inspection and formal tests both showed that the number of COVID-19 deaths per 1,000 residents was not very different from what was expected. These results supported the use of parametric techniques such as Pearson correlation and multiple linear regression in subsequent phases of the analysis. The Discussion section goes into great detail about the test results and visual outputs.

5.3. Correlation Analysis

Conducted a correlation analysis to examine the relationship between COVID-19 mortality rates and socio-economic parameters. We employed both Pearson and Spearman correlation coefficients to investigate potential non-linear or monotonic correlations among variables. Partial correlation analysis was employed to distinctly isolate the effects of education from those of age. This approach facilitated the analysis of the relationship between educational attainment and COVID-19 mortality rates, while concurrently considering the percentage of senior residents. Using partial correlations made sure that the relationships we saw weren't just because the age structure of different local authorities was different. Section 6 talks about the most important correlation results.

5.4. Group-Based Comparisons

Group comparison tests were employed to evaluate if COVID-19 mortality rates exhibited systematic differences among local authority groups with varying levels of deprivation. Both parametric and non-parametric tests were employed to ensure the robustness of the results across various distributional assumptions. These comparisons offered further evidence about the influence of deprivation-related factors in elucidating the spatial variance in COVID-19 mortality. The Discussion section elucidates the implications of the statistical data obtained from the tests.

5.5. Principal Component Analysis

Principal Component Analysis (PCA) was employed on a set of socio-economic variables to reduce dimensionality and mitigate multicollinearity in subsequent regression models. Before PCA, the sample was checked to make sure it was good enough and that the variables chosen had enough common variance. PCA was performed multiple times using various groups of variables and thematic subsets, including those centered on education and those centered on employment. The results consistently supported a two-component structure throughout these iterations. Extra parts did not add much to the explanation and did not meet the standards for retention that had already been set. Because of this, PCA outputs were used to help choose variables instead of being used directly in the model. The Discussion section contains the full PCA results.

5.6 Multiple Regression Modelling

Also used multiple linear regression to look at how age structure, education level, deprivation-related indicators, and occupational composition all affected COVID-19 deaths per 1,000 residents. Used theoretical ideas, correlation analysis, and PCA results to help us figure out how to set up the model. An iterative modeling approach was utilized to improve predictor

selection and increase model interpretability. Variance Inflation Factors were used to check for multicollinearity, and predictors that had too much redundancy were treated with care. We did diagnostic checks to see if the data was linear, homoscedastic, and normal, as well as to see how high-leverage observations affected the results. This modeling strategy facilitated the identification of socio-economic predictors that retained significance when analyzed collectively, while controlling for confounding variables such as age, education, occupation, and deprivation. Section 6 shows and explains the results of the regression and the diagnostic tests.

5.7. Reproducibility

It used R scripts that could be repeated for all of the statistical analyses. The Appendix has detailed code, numerical outputs, and intermediate diagnostics, as well as links to the underlying datasets. This is to make sure that the work is open and can be repeated.

Section 6: Discussion

6.1 Distribution of COVID-19 mortality across local authorities

COVID-reportedly is a moderate but significant difference in the number of deaths per 1,000 residents across English local authorities. The median mortality rate of 2.06 deaths per 1,000 residents is very close to the mean value of 2.07. This means that the distribution is balanced and not very skewed. The interquartile range, which goes from 1.65 to 2.50 deaths per 1,000 residents, shows that most local governments had death rates that were pretty close to each other.

Statistic	Value
Minimum	0.00
1st Quartile (Q1)	1.65
Median	2.06
Mean	2.07
3rd Quartile (Q3)	2.50
Maximum	4.07

Table 1: Summary statistics for COVID-19 deaths per 1,000 residents

The minimum value of zero means that at least one local government said there were no deaths during the time of observation. This result is likely attributable to a limited population, a diverse demographic, or successful local containment strategies, rather than an error in the data. On the other hand, the fact that there were only about 4 deaths per 1,000 residents shows that some areas had much worse outcomes than others. This shows how the pandemic affected different areas in different ways.
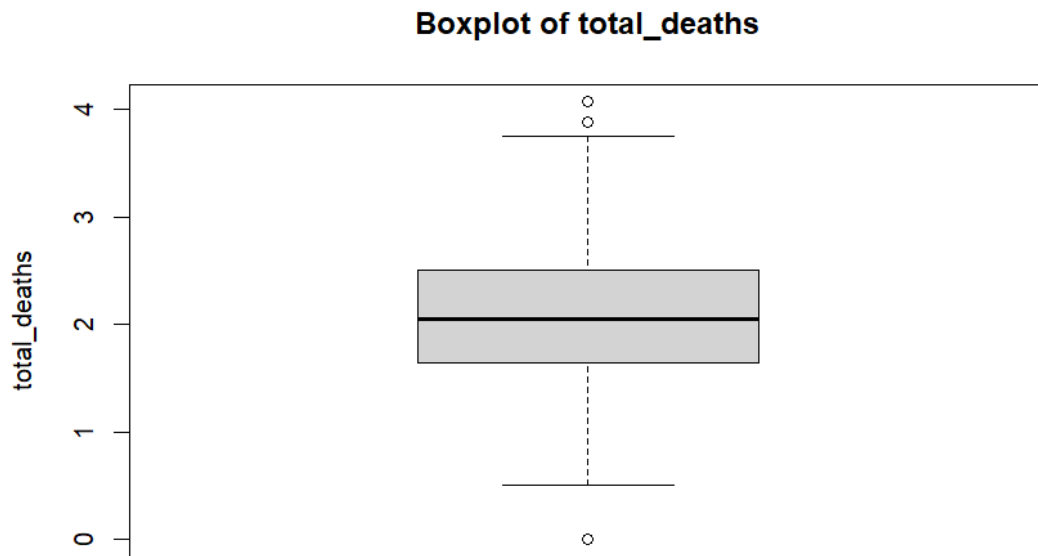
Figure 1: Boxplot of Total COVID-19 Deaths per 1,000 Residents

Explanation:

A boxplot shows how many people died from COVID-19 for every 1,000 people living across all of the 296 English local authorities. The median value is close to the middle of the interquartile range, which shows that the data is evenly spread out and not heavily skewed. The interquartile range isn't very wide, which means that most local governments had death rates that were pretty close to each other. The whiskers go up and down in a way that is symmetrical above and below the box. This supports the idea that the distribution is about symmetrical. Some of the data points are outside the whiskers, which show that some local governments have death rates that are either very low or very high. These outliers are typical in public health statistics at the population level and indicate genuine spatial variations, rather than errors in the data. The boxplot indicates that COVID-19 fatalities are distributed very uniformly throughout local authorities, with no single value dominating the whole range. This makes future analyses more reliable because the results are unlikely to be affected by a small number of extreme local authorities. This enhances confidence that the observed relationships stem from enduring socio-economic trends rather than mere chance occurrences in certain regions.

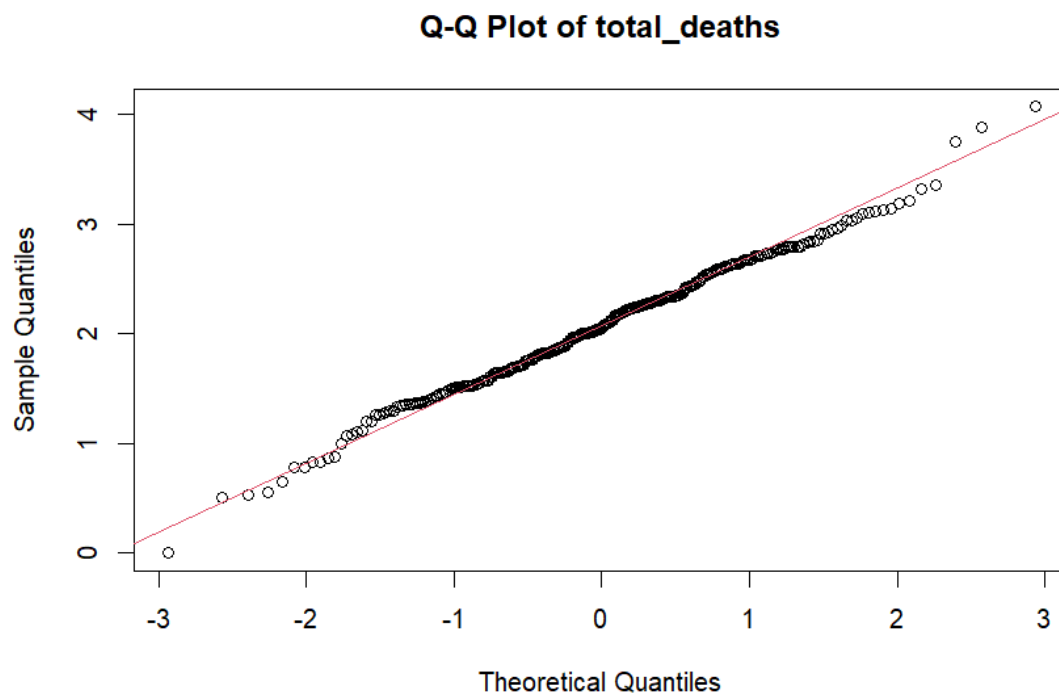## 6.2 Normality assessment and suitability for parametric analysis

Figure 2: Q–Q Plot of Total COVID-19 Deaths per 1,000 Residents

Explanation:

The actual number of COVID-19 deaths per 1,000 residents is compared to a theoretical normal distribution using a quantile–quantile (Q–Q) plot in this picture. Most of the points, particularly those in the middle of the distribution, are extremely close to the 45-degree reference line. This indicates that the majority of people support normalcy. A few local authorities deviate from the expected typical quantiles at the lower and upper tails. These kinds of discrepancies are typical in actual epidemiology data and do not indicate a significant violation of the normalcy assumptions. The plot's lack of clustering or systematic curvature indicates that there isn't any noticeable skewness or distinct tails in the dependent variable. Formal normality tests and this visual proof demonstrate that COVID-19 death data can be used for linear regression. For correlation analysis, Spearman's approach was employed to ensure its validity even in cases when the data did not adhere to rigorous normality.
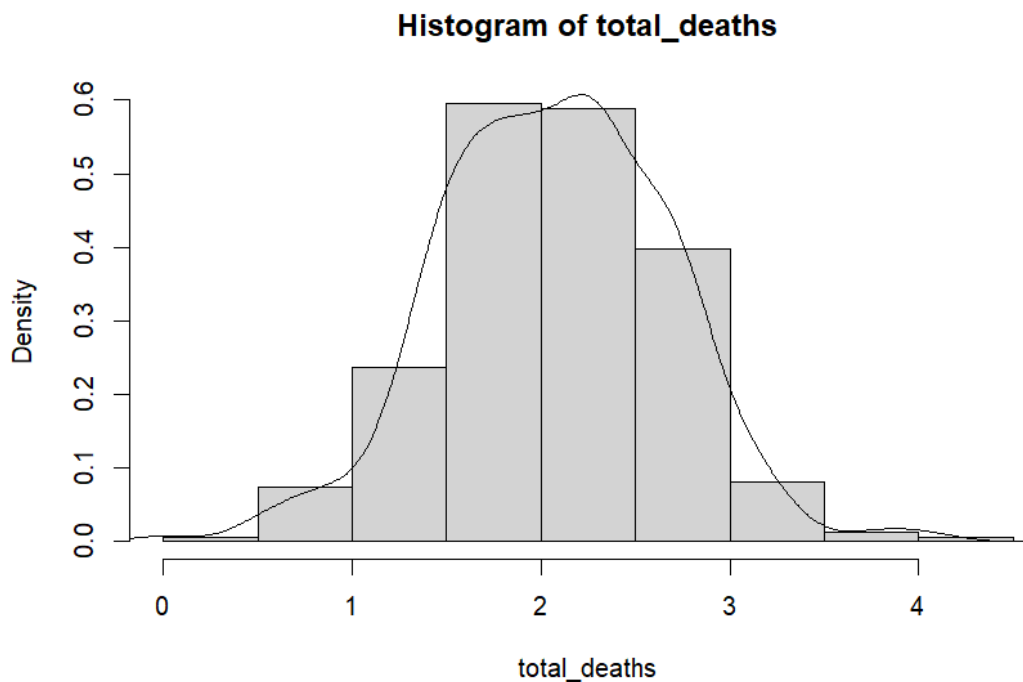
Histogram of total_deaths

Figure 3:Histogram of Total COVID-19 Deaths per 1,000 Residents

Explanation:

The graph displays the number of COVID-19 fatalities per 1,000 residents across all local administrations. The histogram shows a distinct single peak at about two fatalities per 1,000 persons, which is extremely near to the sample's mean and median. The distribution looks pretty even on both sides of the central peak, with no signs of more than one mode or extreme skewness. Although the right tail extends somewhat farther than the left, this discrepancy is negligible and consistent with the Q–Q plot's findings. This figure supports the hypothesis that, at the local government level, COVID-19 death rates are about regularly distributed. The assumptions of linear regression are successfully validated when the dependent variable is consistently shown to satisfy the normalcy assumptions by the formal tests and graphical diagnostics. Spearman's method is used to report correlations because it is more reliable. Formal tests also show that the data is close to normal (Shapiro–Wilk $p = 0.257$; KS $p = 0.911$).

6.3 Socio-economic correlates of COVID-19 mortality

Correlation analysis shows clear socio-economic trends linked to COVID-19 deaths. Areas with higher death rates have a higher proportion of people without formal qualifications. On the other hand, areas with a higher percentage of the population has Level 4 or higher qualifications typically have lower death rates.
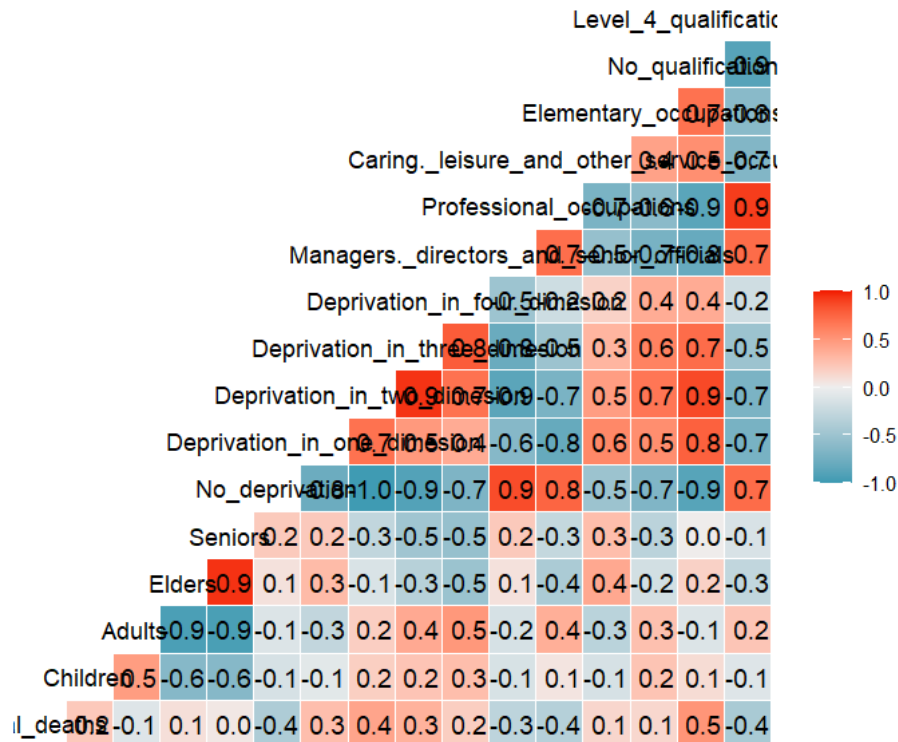
Level_4_qualificatio

No_qualification 0.8

Elementary_occupatio 0.7 0.8

Caring._leisure_and_other_servic 0.4 0.5 0.7

Professional_occupatio 0.7 0.6 0.9 0.9

Managers._directors_and_senior_officials 0.7 0.5 0.7 0.8 0.7

Deprivation_in_four_dimensio 0.5 0.2 0.2 0.4 0.4 -0.2

Deprivation_in_three_dimensio 0.8 0.8 0.5 0.3 0.6 0.7 -0.5

Deprivation_in_two_dimensio 0.9 0.7 0.9 -0.7 0.5 0.7 0.9 -0.7

Deprivation_in_one_dimensio 0.7 0.5 0.4 -0.6 -0.8 0.6 0.5 0.8 -0.7

No_deprivation 0.8 -1.0 -0.9 -0.7 0.9 0.8 -0.5 -0.7 -0.9 0.7

Seniors 0.2 0.2 -0.3 -0.5 -0.5 0.2 -0.3 0.3 -0.3 0.0 -0.1

Elders 0.9 0.1 0.3 -0.1 -0.3 -0.5 0.1 -0.4 0.4 -0.2 0.2 -0.3

Adults 0.9 -0.9 -0.1 -0.3 0.2 0.4 0.5 -0.2 0.4 -0.3 0.3 -0.1 0.2

Children 0.5 -0.6 -0.6 -0.1 -0.1 0.2 0.2 0.3 -0.1 0.1 -0.1 0.2 0.1 -0.1

l_deaths 0.2 -0.1 0.1 0.0 -0.4 0.3 0.4 0.3 0.2 -0.3 -0.4 0.1 0.1 0.5 -0.4

Figure 4: Spearman correlation heatmap of COVID-19 mortality and socio-economic variables

Explanation:

A heatmap of the Spearman association between COVID-19 deaths and socioeconomic characteristics is displayed in the figure. It is simple to see how objects are connected in a variety of ways thanks to the colour gradients. The population living in areas without formal education or working in elementary or caring occupations is positively correlated with the COVID-19 death rate. These correlations demonstrate that individuals are more likely to die in places with lower levels of education and greater exposure to frontline or low-wage labor. On the other hand, a protective link is indicated by negative correlations between mortality and higher educational attainment. Deprivation-related variables exhibit a robust correlation with one another and with lower-skilled occupations, underscoring a more extensive aggregation of disadvantage. Age shares exhibit diminished bivariate correlations at the local authority level, presumably due to the co-variation of socio-economic variables with age structure across regions. The heatmap demonstrates how socioeconomic factors, particularly employment and education play a significant role in explaining why COVID-19 fatalities differ among local governments. At the local authority level, these bivariate patterns show that structural socioeconomic factors, particularly education and occupational exposure have a higher link with COVID-19 mortality than does age composition alone. This clearly motivates us to investigate whether these correlations persist when we consider them in conjunction with other variables in multivariate models. Other occupational factors are mostly examined in the multivariate regression framework and show poorer bivariate correlations with mortality ($|\rho| <$ 0.25).

Partial correlation analysis was employed to delineate the influence of education while accounting for the percentage of older residents. After controlling for the percentage of seniors (75+), the correlation between COVID-19 mortality and the fraction of residents lacking qualifications persists as positive and statistically significant.

Outcome (y)	Predictor (x)	Control	Partial Spearman ρ	p-value
total_deaths	No_qualifications	Seniors	0.5039	2.1086e-20
total_deaths	Level_4_qualifications_or_above	Seniors	-0.4254	2.1519e-14

Table 2: Partial Spearman correlations between COVID-19 mortality and education, controlling for Seniors

Explanation:

Mortality still shows a significant negative correlation with the percentage of residents with Level 4+ qualifications ($\rho = -0.4254$, $p < 0.001$) and a significant positive correlation with the percentage of residents without qualifications ($\rho = 0.5039$, $p < 0.001$) after controlling for the percentage of Seniors (75+). This demonstrates that educational disadvantage correlates with elevated mortality, irrespective of local age demographics.

6.5 Group-based differences by deprivation level

Tests of group comparison offer further evidence of differences in COVID-19 outcomes. More deprived local authorities showed considerably higher death rates compared to less deprived areas, according to both parametric (t-test) and non-parametric (Mann–Whitney) tests.

In highly deprived areas, there were about 2.17 deaths per 1,000 residents, while in less deprived areas, there were about 1.97 deaths per 1,000 residents. There were statistically significant differences between the two testing methods. The fact that the results are the same across different methods makes this finding more reliable. These results support the hypothesis that deprivation serves as a contextual risk factor, increasing vulnerability to COVID-19 mortality at the regional level, rather than merely reflecting random variations among local authorities. Results are consistent across both parametric (Welch t-test) and non-parametric (Mann–Whitney U) methodologies, thereby diminishing sensitivity to distributional assumptions. Welch t-test: $t = -2.93$, $df = 292.64$, $p = 0.0037$; Mann–Whitney: $W = 8718$, $p = 0.0024$.

6.6 Principal Component Analysis and iterative validation

Also did Principal Component Analysis with different sets of variables, such as education-only, occupation-only, combined education-occupation, and education-occupation-deprivation.

Scree plots and parallel analysis consistently supported a two-component structure across all iterations.

Parallel Analysis Scree Plots
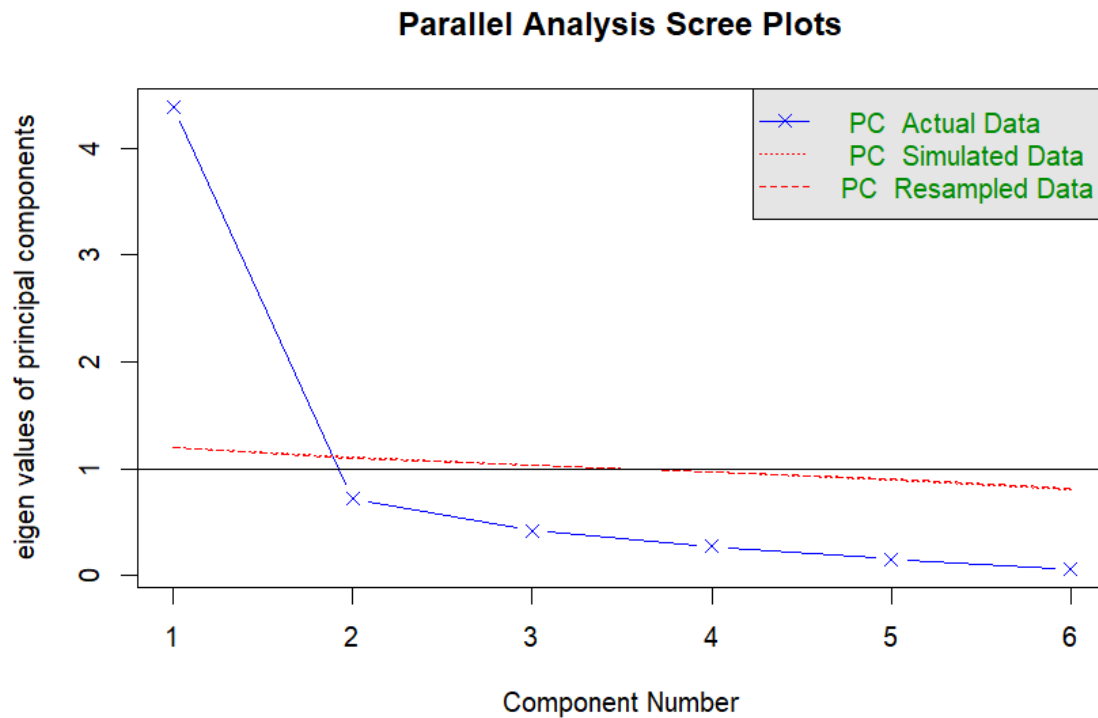
Figure 5:Parallel analysis and scree plot for principal components.

Explanation:

The parallel analysis and scree plot used to determine the appropriate number of components to retain in the Principal Component Analysis are displayed in Figure 5. The first component has a great deal of explanatory power because it is obviously higher than the eigenvalues derived from random data. Additionally, the second section crosses the simulation threshold or comes very near to it, supporting its retention. Parts after the second one fall below the simulated eigenvalue line and don't add much more variance. This pattern shows that keeping more parts will give you less and less value. The stability of this two-component structure for the chosen socio-economic variable set indicates that the socio-economic variables influencing COVID-19 mortality can be encapsulated within two principal dimensions. These results validate the application of PCA as a diagnostic instrument for elucidating multicollinearity and guiding regression model specification. PCA was utilized as a diagnostic and interpretative instrument rather than a replacement for theoretically significant regression variables, thereby preserving substantive interpretation from excessive abstraction. Overall, KMO was acceptable (MSA = 0.63), but some age-share variables had low individual MSA, so the results of PCA should be taken with a grain of salt.

6.7 Regression modelling, iteration, and model improvement

Multiple linear regression analysis shows that a large part of the difference in COVID-19 deaths between English local authorities can be explained by a combination of socio-economic and demographic factors. The final model accounts for approximately 40% (R^2 = 0.402; adjusted R^2 = 0.389) of the variance in deaths per 1,000 residents, demonstrating significant explanatory capacity for an ecological analysis of intricate public health outcomes. The very significant F-statistic ($p < 0.001$) shows that the model works much better than a null model with no predictors. Variance Inflation Factors showed that there was a lot of multicollinearities between age groups that were next to each other, especially between Elders (65–74) and Seniors (75+). This is because they are very similar in terms of demographics. So, the individual age coefficients in the full model should be taken with a grain of salt because their estimated effects are not always stable. This shows that age-related effects happen on a large scale, not just in small age groups.

Term	Estimate	Std. Error	t value	p-value	Direction
(Intercept)	2.7174	0.6364	4.2699	2.6557e-05	—
Elders	-0.0060	0.0016	-3.8636	1.3799e-04	Negative
No_qualifications	0.0152	0.0018	8.4169	1.8174e-15	Positive
Level_4_qualifications_or_above	-0.0031	0.0008	-3.8232	1.6138e-04	Negative
Managers_directors_and_senior_officials	0.0069	0.0029	2.3998	1.7039e-02	Positive
Caring_leisure_and_other_service_occupations	-0.0158	0.0061	-2.6051	9.6610e-03	Negative
Elementary_occupations	-0.0259	0.0038	-6.7265	9.3030e-11	Negative

Table 3:Final multiple linear regression coefficients predicting COVID-19 deaths per 1,000 residents

Explanation:

The table shows the final regression model after checking for model iteration and multicollinearity. Keeping other factors the same, areas with more residents who don't have any qualifications have higher COVID-19 death rates, while areas with more residents who have Level 4+ qualifications have lower death rates. Several occupational variables continue

to exhibit statistical significance, suggesting that the structure of the local labor market provides supplementary explanatory power beyond education. The negative coefficient for Elders (65–74) necessitates careful interpretation due to the high correlation of age shares with other demographic and socio-economic indicators at the area level; in multivariate models, coefficients may change sign when overlapping structures are accounted for.

Key regression outputs to reference in Discussion

- **R² = 0.402**
- **Adjusted R² = 0.389**
- **Model F-statistic significant at p < 0.001**

Education and occupation variables prove to be the most dependable indicators. Higher death rates are linked to areas with more residents who don't have any qualifications, while lower death rates are linked to areas with more residents who have higher levels of education. The demographics of the workforce is also very important because it shows how exposed the structure is to risks. When education and occupation are considered, deprivation indicators lose their statistical significance. This implies that their effects operate via these pathways. Because of overlapping socio-economic structures and multicollinearity control, some coefficients change sign when compared to bivariate correlations.

6.8 Model diagnostics and validity of inference

Regression diagnostic plots confirm that model assumptions are broadly satisfied.
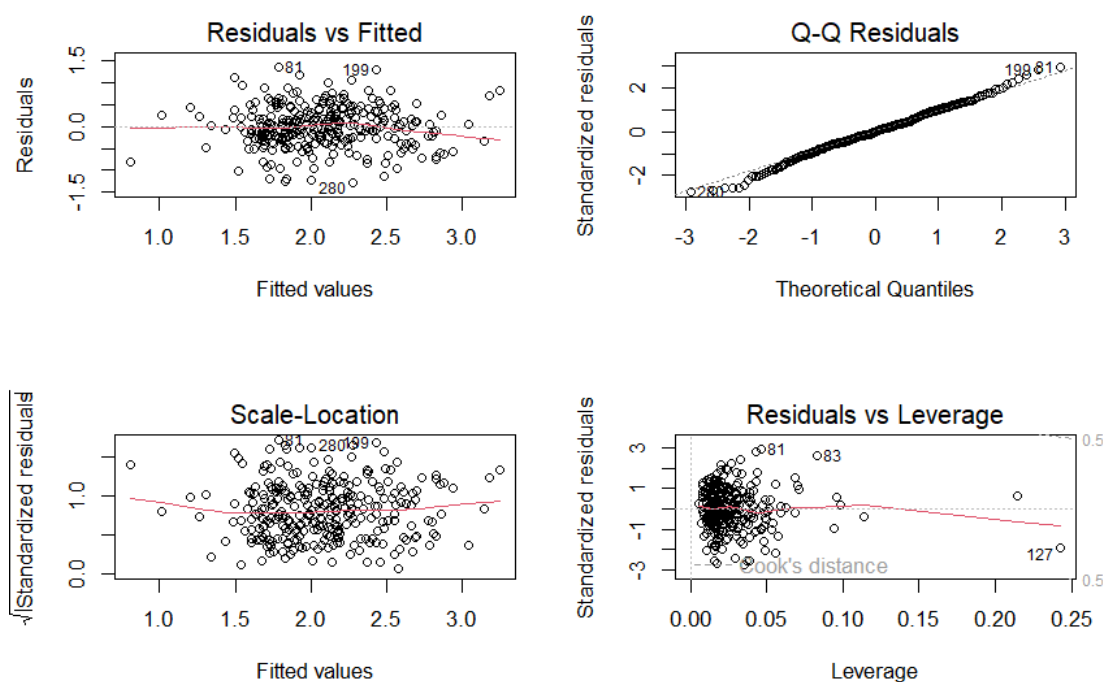
Figure 6:Regression diagnostic plots for the multiple linear regression model.

Explanation:

This indicates that the linearity assumption is satisfied. The residuals are fairly normally distributed, as indicated by the Q–Q plot of residuals, which nearly resembles the theoretical normal line. The distribution of residuals around fitted values appears to be quite stable, according to the Scale–Location plot. Heteroskedasticity is therefore not a significant issue. A few locations with larger leverage are visible in the Residuals vs. Leverage plot, but none of them exceed Cook's distance constraints.

These tests demonstrate the validity of the regression model's underlying assumptions as well as the accuracy of the statistical inference and coefficient estimates. These tests demonstrate the reliability of the regression results and the fact that results remain unchanged when key model assumptions are broken.

n	MAE	MSE	RMSE	R^2	Adjusted R^2
296	0.3630	0.2172	0.4661	0.4017	0.3893

Table 4: Model performance metrics for the final regression model

Explanation:

The last model accounts for about 40% of the differences in COVID-19 deaths per 1,000 people across local authorities ($R^2 = 0.4017$; adjusted $R^2 = 0.3893$). The prediction error is moderate (RMSE = 0.466 deaths per 1,000), which is reasonable for an ecological model where outcomes are affected by other unmeasured factors (health status, care-home exposure, local policies, and unmeasured factors).

6.9. Limitations and scope:

However, a number of limitations that need to be recognized. The analysis is ecological, therefore it can't analyse risk factors for each person. This could lead to an ecological fallacy. Second, unobserved factors like the capacity of the health system, local policy responses, and pre-existing health conditions that aren't fully captured in the data affect the death rate from COVID-19. Third, even though diagnostic checks and variable aggregation helped with multicollinearity, some socio-economic indicators are still conceptually linked. Consequently, the model elucidates considerable yet incomplete variation in mortality, leaving a segment of spatial disparities unaccounted for. Associations represent regional characteristics and should not be construed as individual-level risk.

Section 7: Conclusion

This study addressed the spatial variation in COVID-19 mortality across 296 English local authorities and evaluated whether this variation could be elucidated by disparities in age distribution, educational achievement, deprivation-related metrics, and occupational composition. The analysis employs exploratory analysis, correlation and partial correlation tests, Principal Component Analysis, and multiple linear regression to illustrate that COVID-19 mortality was not randomly distributed in England but adhered to distinct socio-economic patterns. The findings indicate that education is the most reliable and strong predictor of COVID-19 mortality at the local authority level. Areas with a higher percentage of residents who did not have formal qualifications had much higher death rates. On the other hand, areas with a higher percentage of residents who had Level 4 or higher qualifications had lower death rates. These relationships remained statistically significant even after controlling for age structure. This means that education is related to COVID-19 death rates in its own way, not just because older people are more likely to die from it.

Some of the observed variation can be explained by the occupational structure. Even after adjusting for age and education, the regression analysis shows that differences in the makeup of the local labour market are still linked to death rates. However, in bivariate analyses, but not in the multivariate model, indices of deprivation were associated with increased death rates. This trend implies that deprivation does not function as a fully independent component, but rather indirectly, mostly through interrelated routes such as educational disadvantage and occupational exposure.

The multivariate regression model explains about 40% of the differences in COVID-19 death rates per 1,000 residents between local governments. This shows that an ecological study of a complex public health outcome can explain a lot, but it also shows that a lot of the variability is still unexplained. But since there is multicollinearity between age groups that are next to each other, individual age coefficients should be looked at carefully and as part of a bigger demographic structure, not on their own.

There are still some important boundaries. The analysis is ecological, which means it can't find risk factors or causes for individuals. Second, the data do not include important factors that affect COVID-19 death rates, such as pre-existing health conditions, the number of care homes, the timing of vaccinations, the capacity of the local health system, policy responses, and differences in behaviour between areas. Third, using Census 2011 socio-economic data means that changes in the job and population markets that have happened since then are not fully shown. These limitations clarify the factors contributing to the substantial portion of spatial mortality variation that remains unexplained. Despite these limitations, the analysis provides clear evidence that structural socio-economic factors, particularly education and occupation, significantly influenced the disparities in COVID-19 mortality rates across England, even when accounting for age distribution. The findings substantiate the conclusion that the pandemic intensified pre-existing inequalities rather than affecting all sectors uniformly, highlighting the importance of persistent socio-economic factors in understanding population-level vulnerability to public health emergencies.

Appendix

The complete R scripts and cleaned datasets and additional diagnostic results for this analysis exist on OneDrive. The supplementary materials serve to enhance transparency and reproducibility but readers can understand the report's essential results without accessing them.

OneDrive link: Covid_19

Boxplot of Children
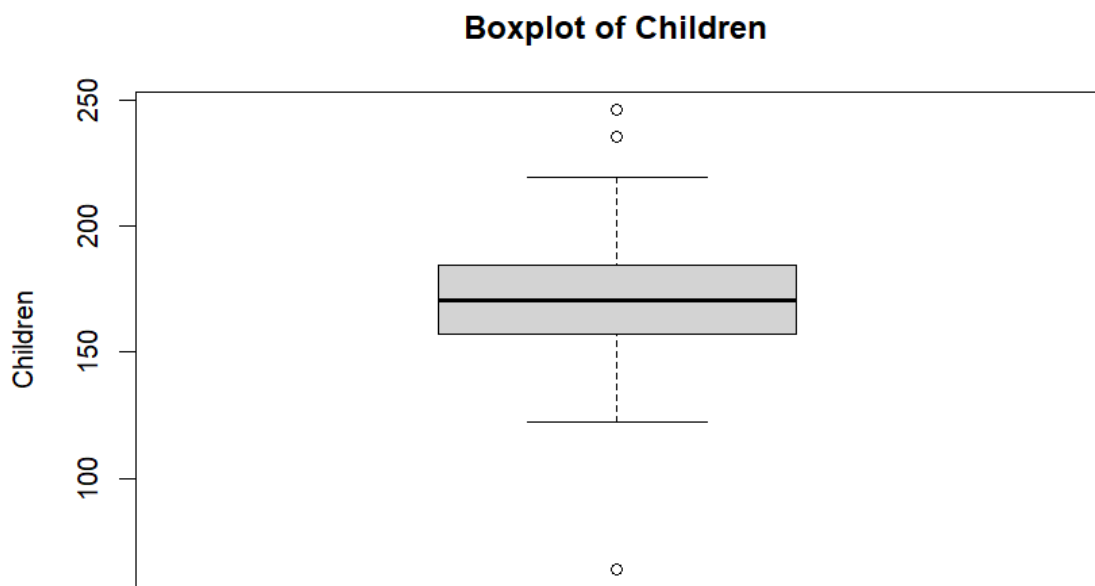
Figure 7:Boxplot of Children per 1,000 residents
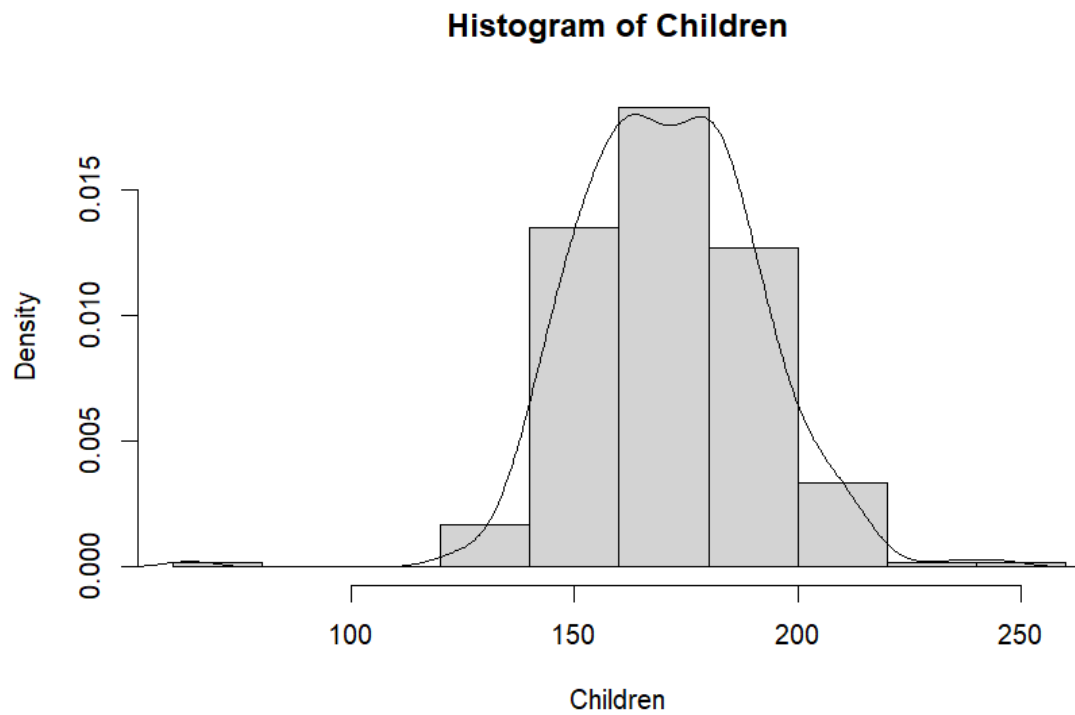
Figure 8:Histogram of Children per 1,000 residents
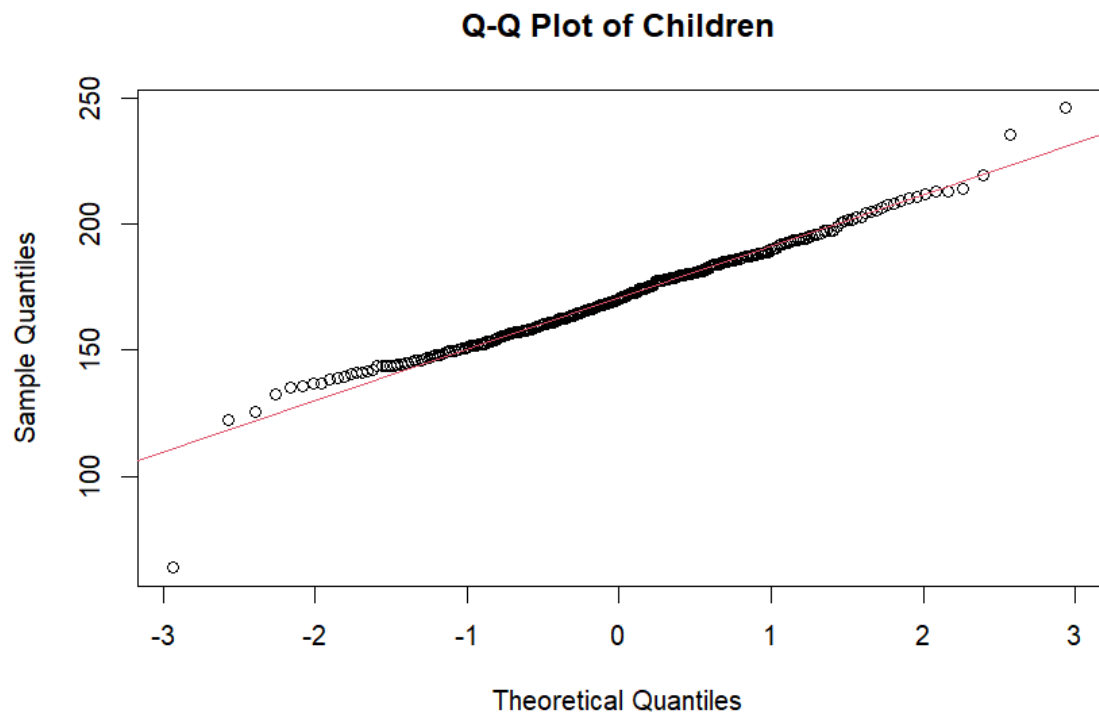
Figure 9:Q–Q plot of Children per 1,000 residents
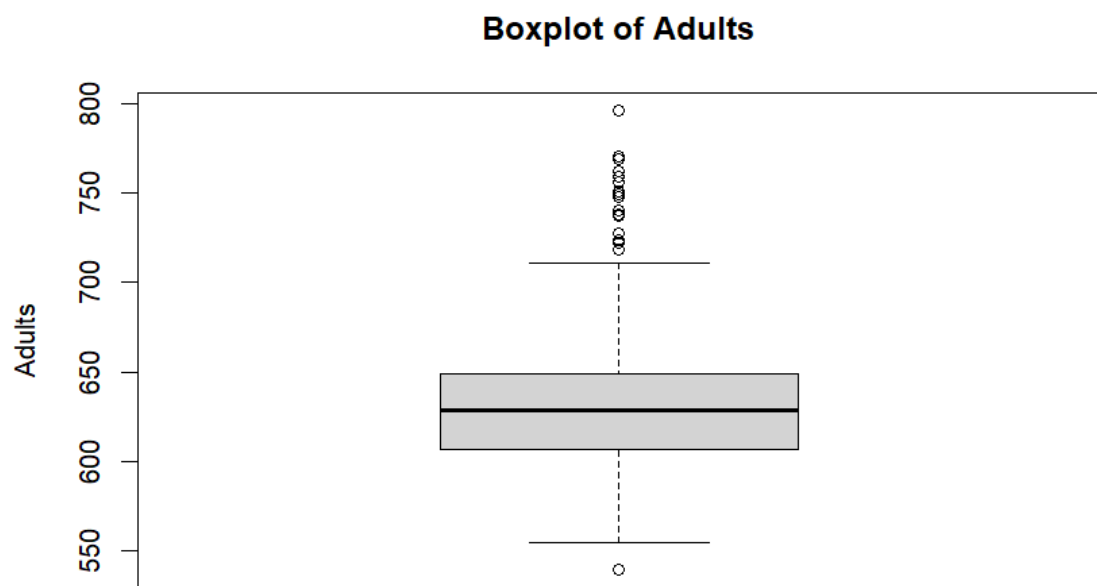
Boxplot of Adults

Figure 10:Boxplot of Adults per 1,000 residents

Histogram of Adults
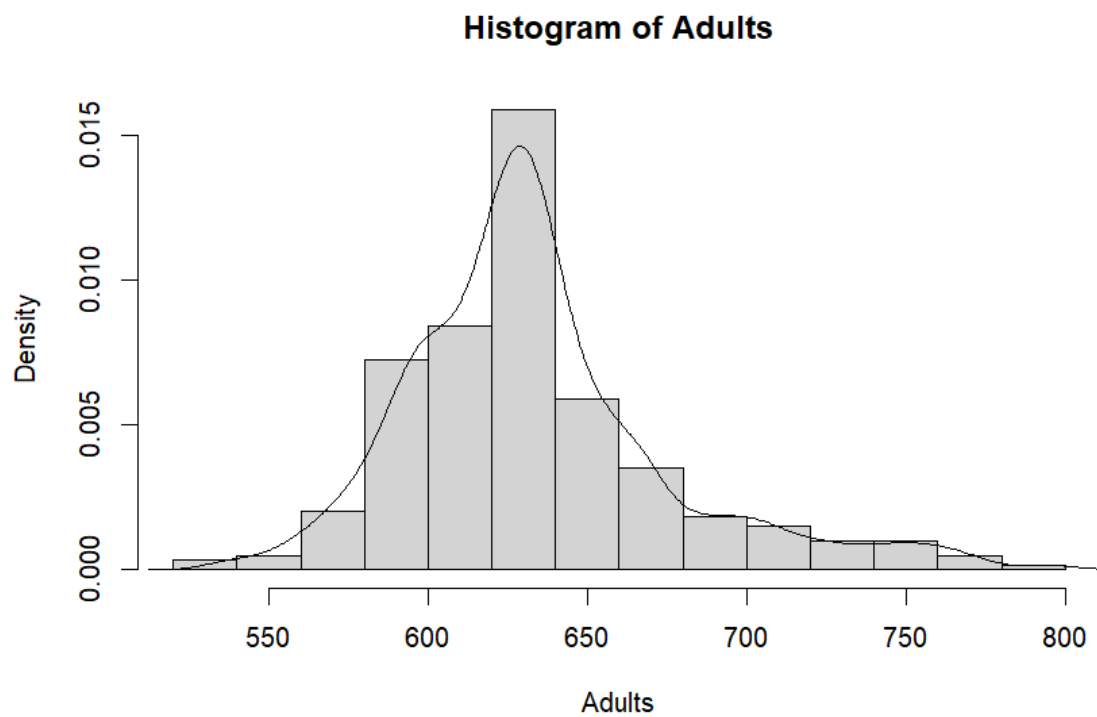
Figure 11:Histogram of Adults per 1,000 residents
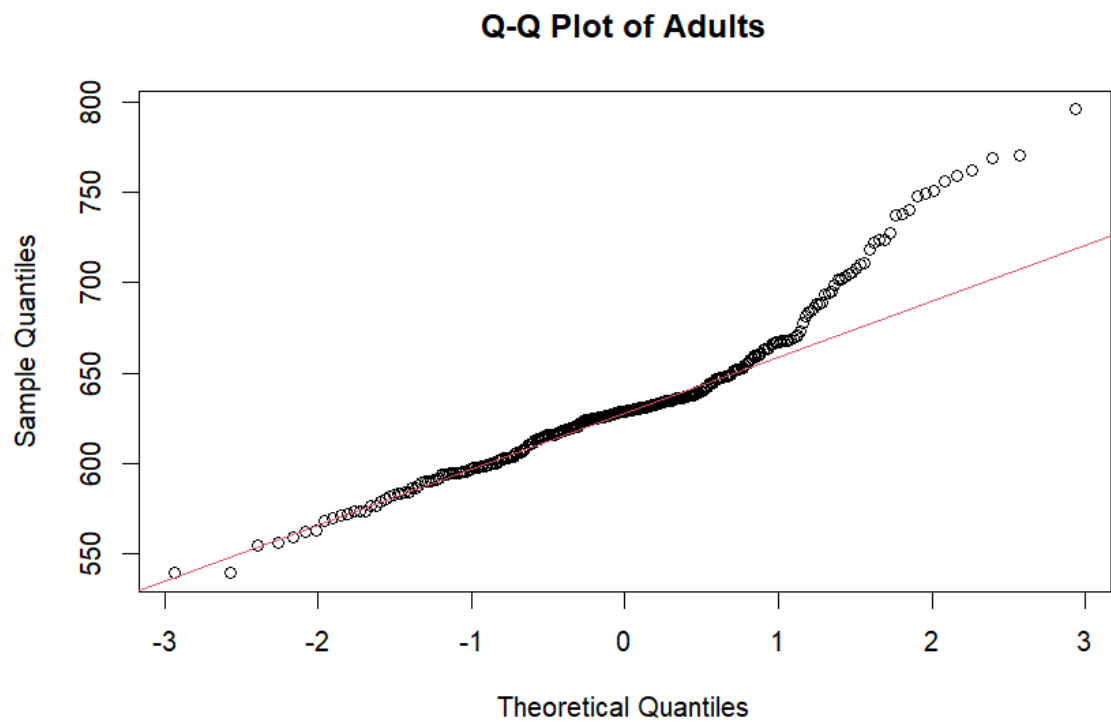
Q-Q Plot of Adults

Figure 12:Q–Q plot of Adults per 1,000 residents
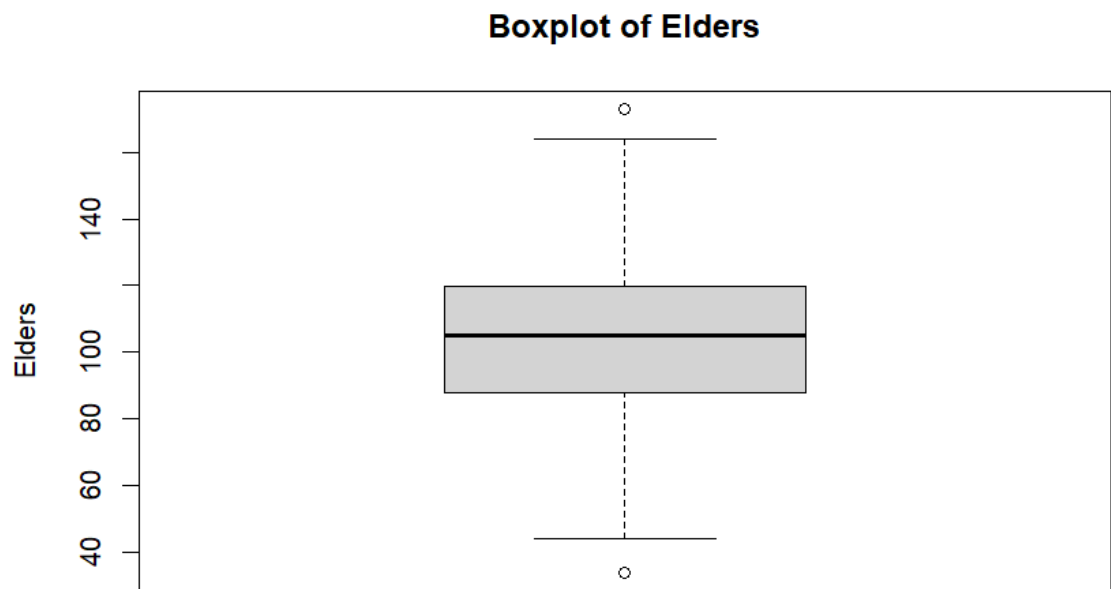
Boxplot of Elders

Figure 13:Boxplot of Elders per 1,000 residents

Histogram of Elders
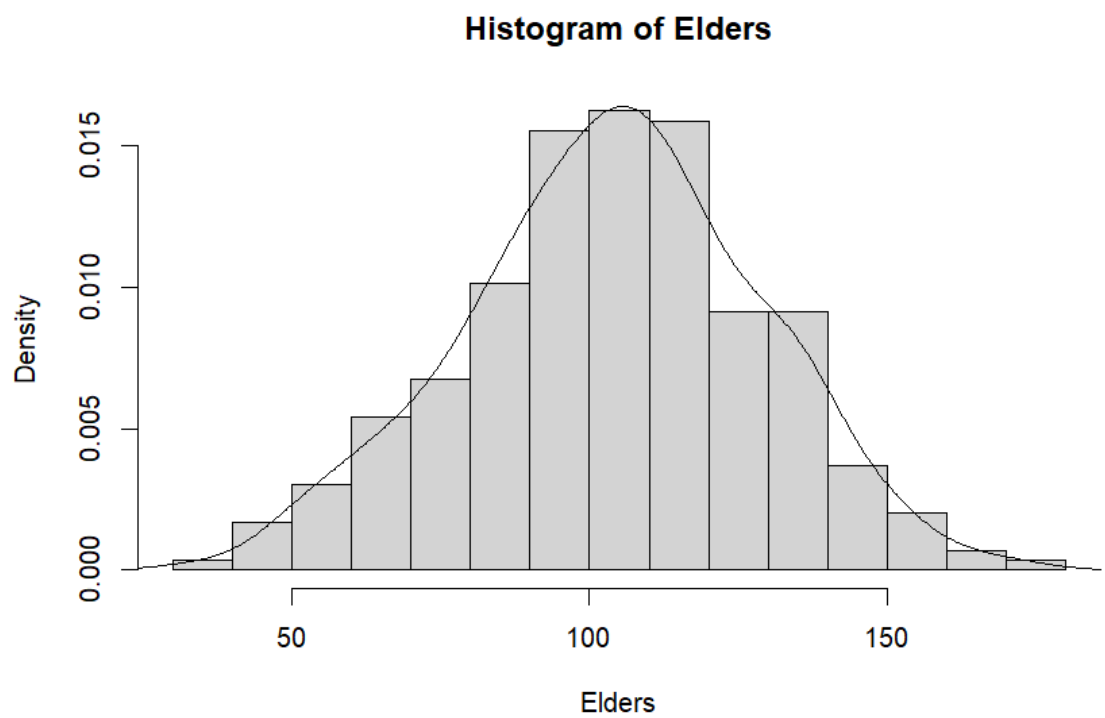
Figure 14:Histogram of Elders per 1,000 residents
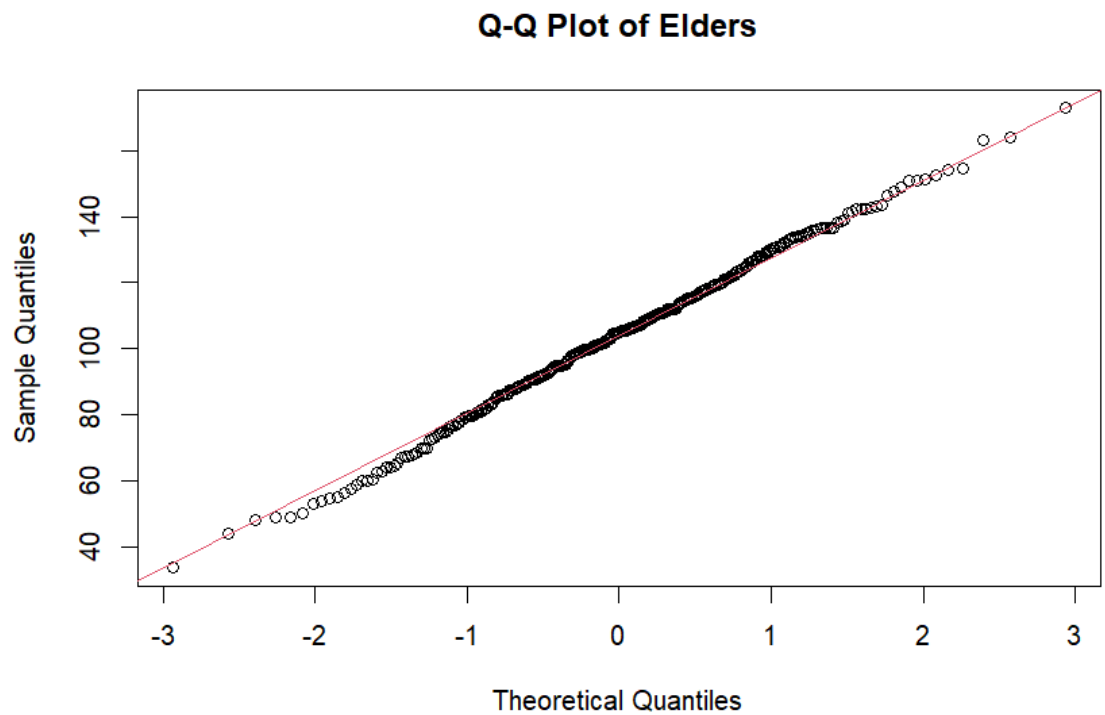
Q-Q Plot of Elders

Figure 15:Q–Q plot of Elders
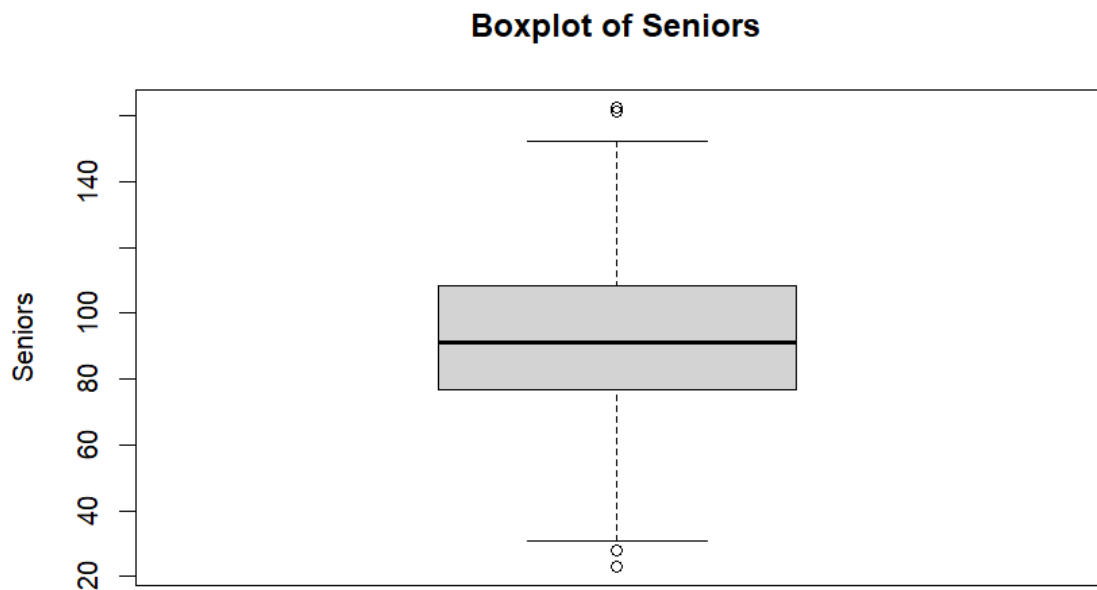
Boxplot of Seniors

Figure 16:Boxplot of Seniors per 1,000 residents

Histogram of Seniors
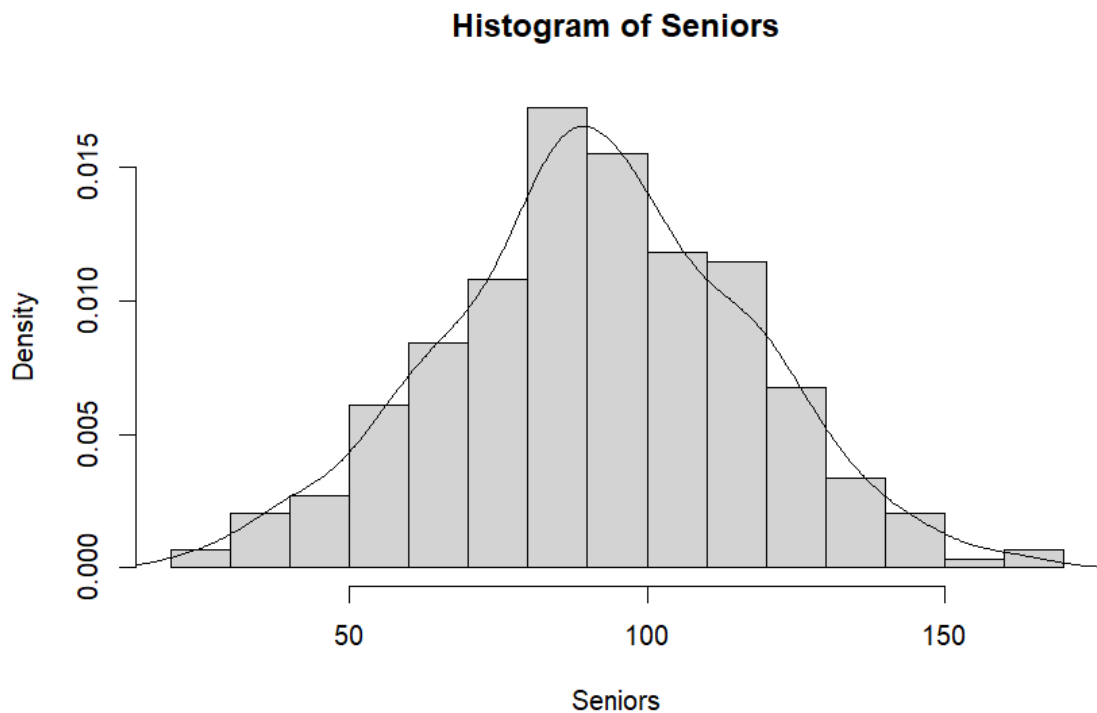
Figure 17:Histogram of Seniors (75+) per 1,000 residents
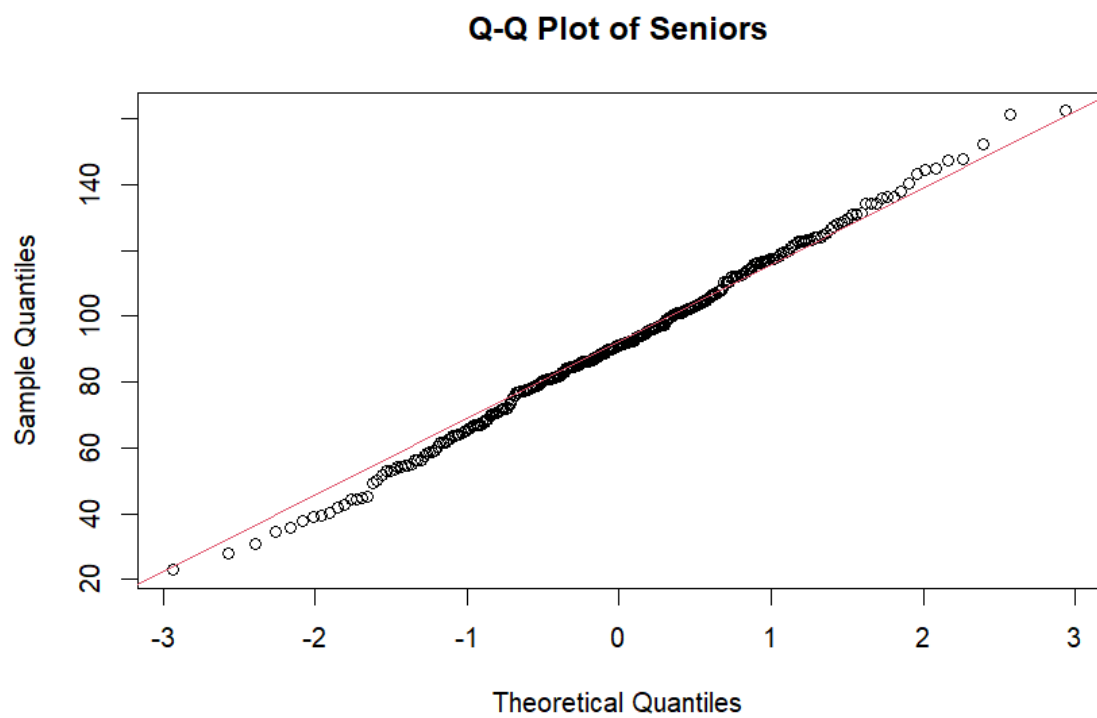
Q-Q Plot of Seniors

Figure 18:Q–Q plot of Seniors

Boxplot of No_deprivation
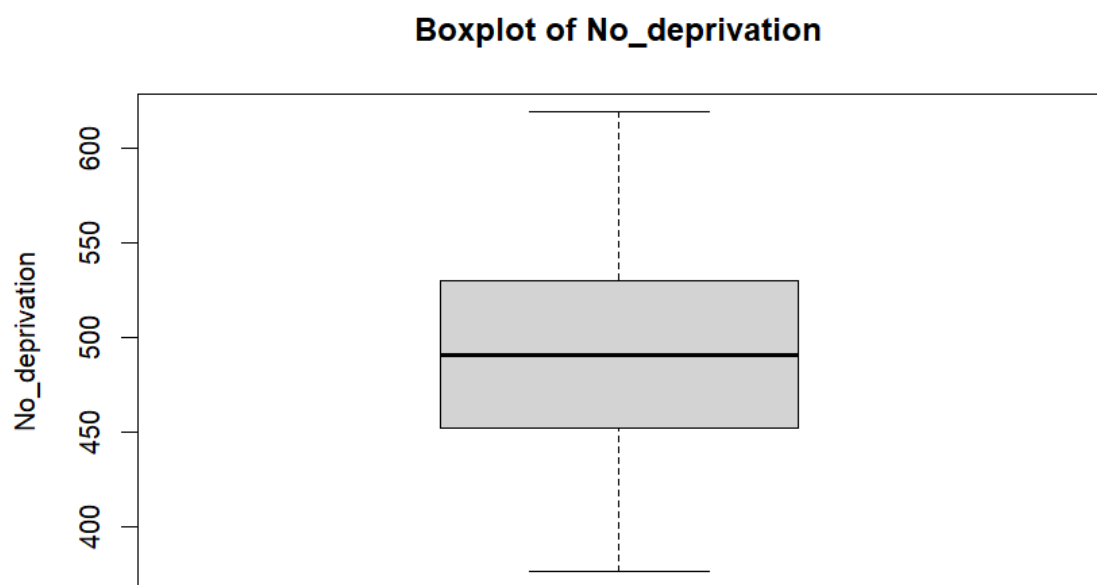
Figure 19:BoxPlot of no deprivation

Histogram of No_deprivation
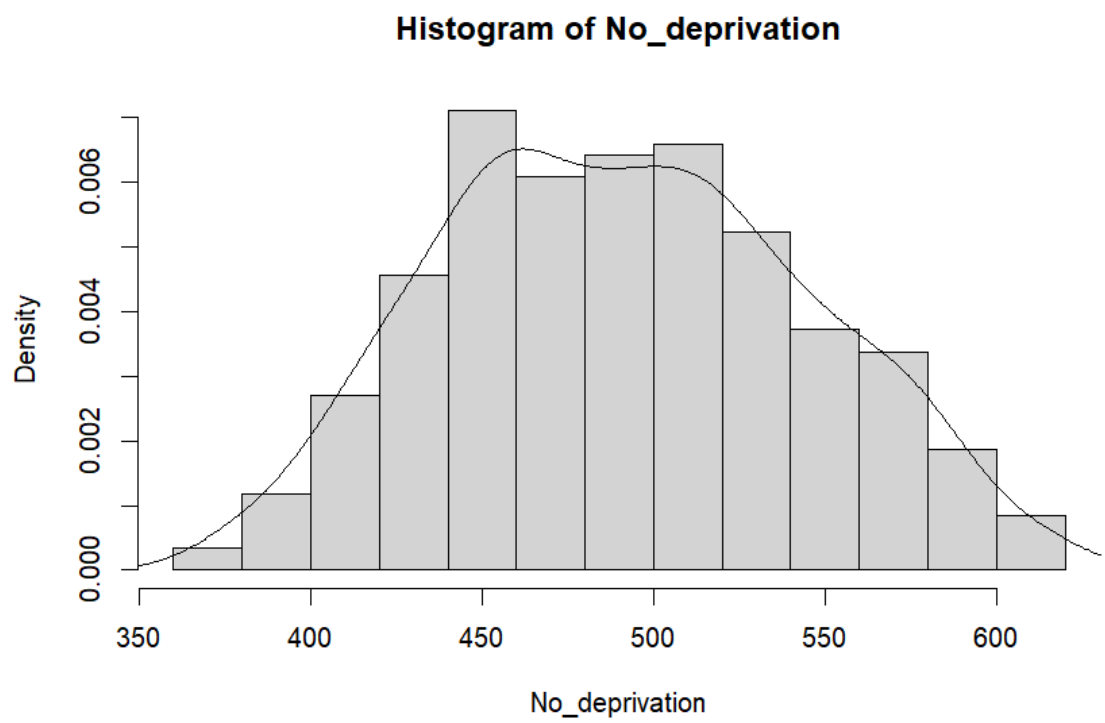
Figure 20:Histogram of no deprivation

Q-Q Plot of No_deprivation
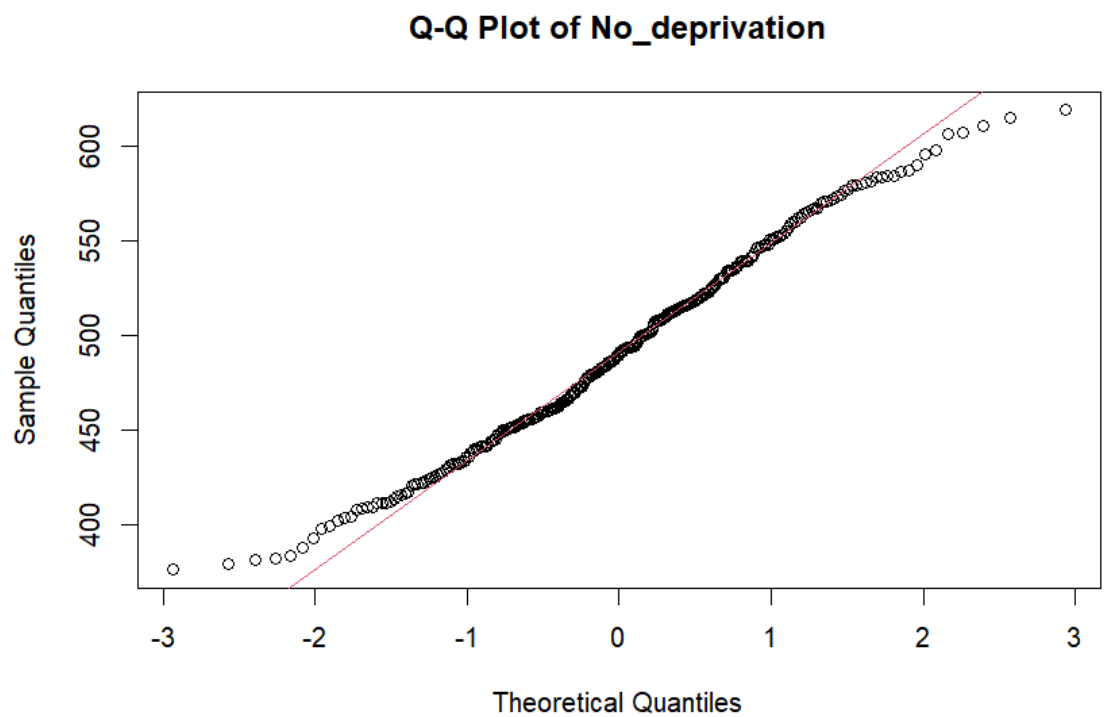
Figure 21:Q-Q Plot of no deprivation

Boxplot of Deprivation_in_one_dimesion
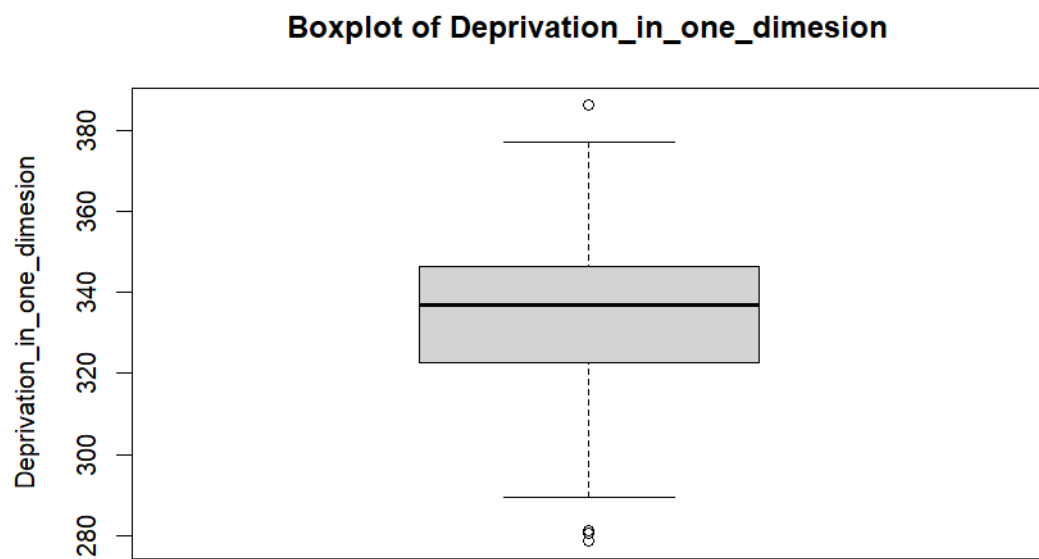
Figure 22: Boxplot of Deprivation in one dimension

Histogram of Deprivation_in_one_dimesion
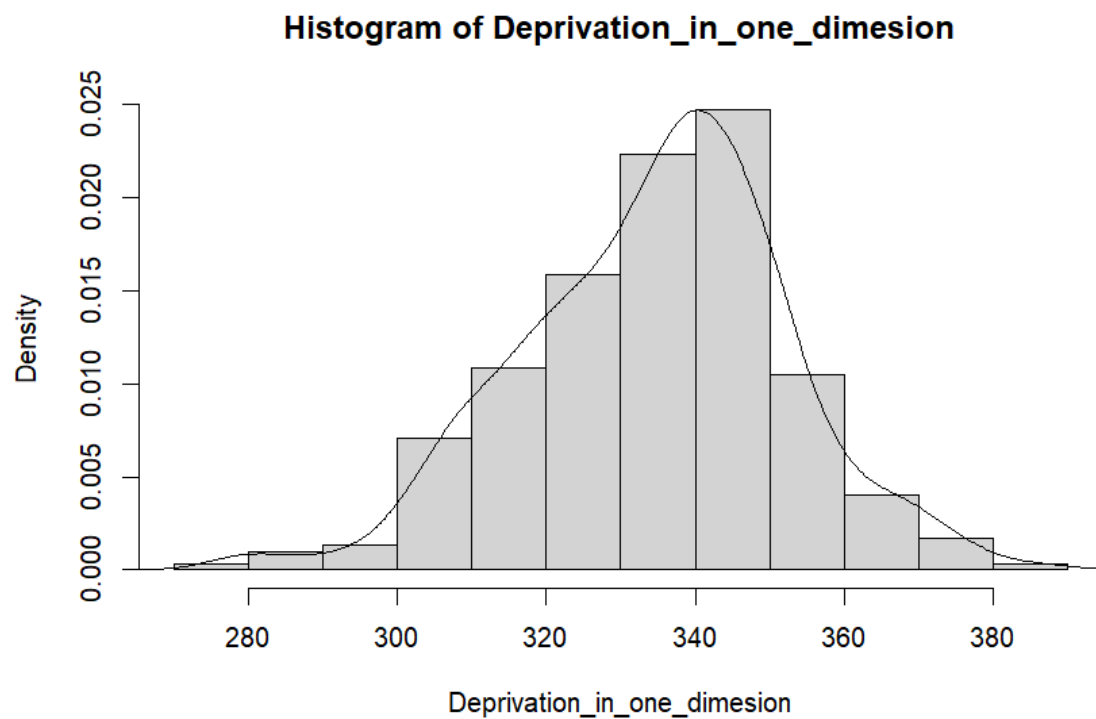
Figure 23: Histogram of Deprivation in one dimension

Q-Q Plot of Deprivation_in_one_dimesion
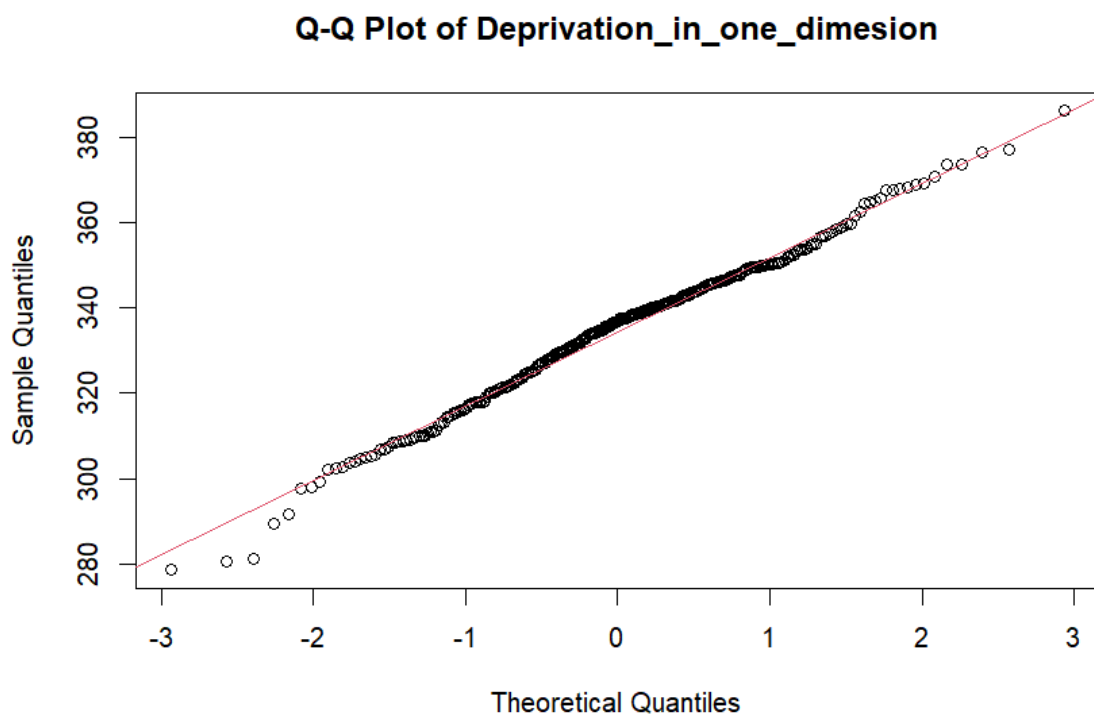
Figure 24:Q-Q plot of Deprivation in one dimension

Boxplot of Deprivation_in_two_dimesion
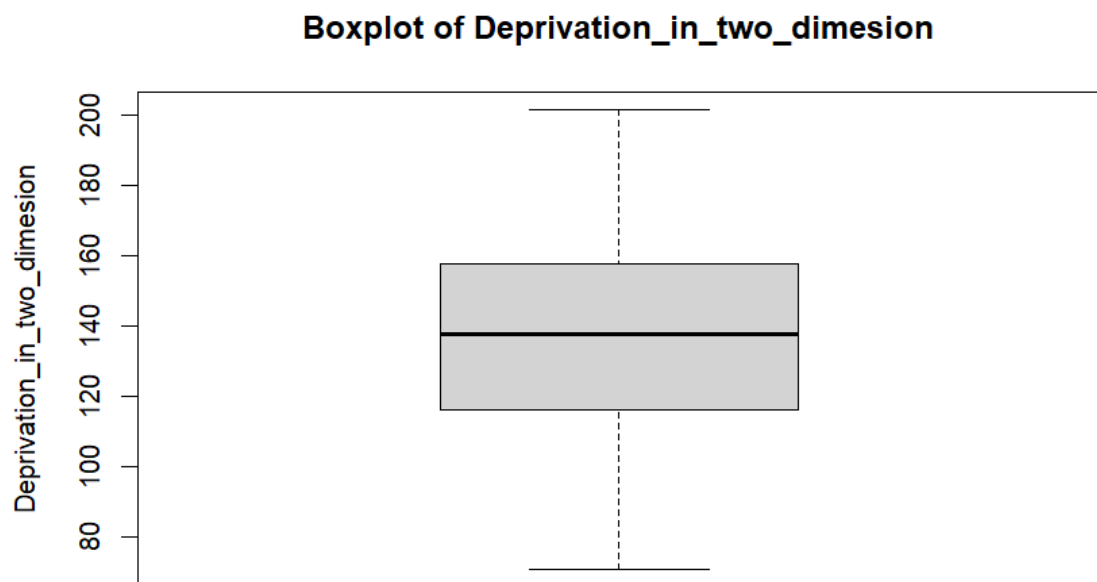
Figure 25: Boxplot of Deprivation in two dimensions

Histogram of Deprivation_in_two_dimesion
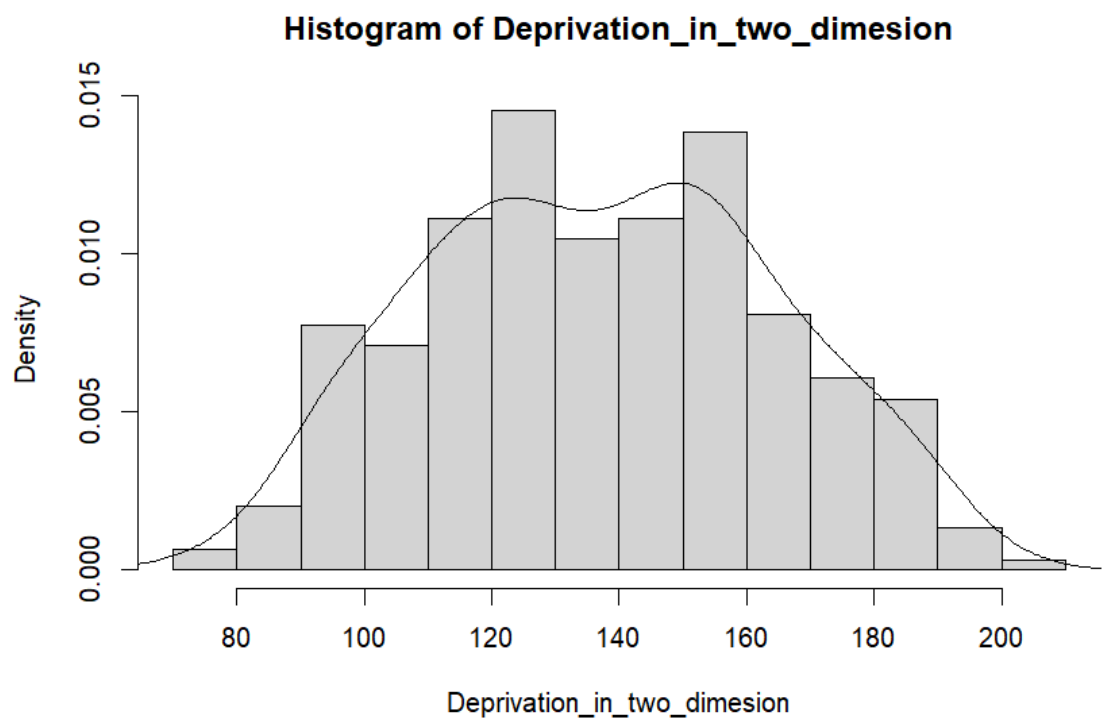
Figure 26:Histogram of Deprivation in two dimensions

Q-Q Plot of Deprivation_in_two_dimesion
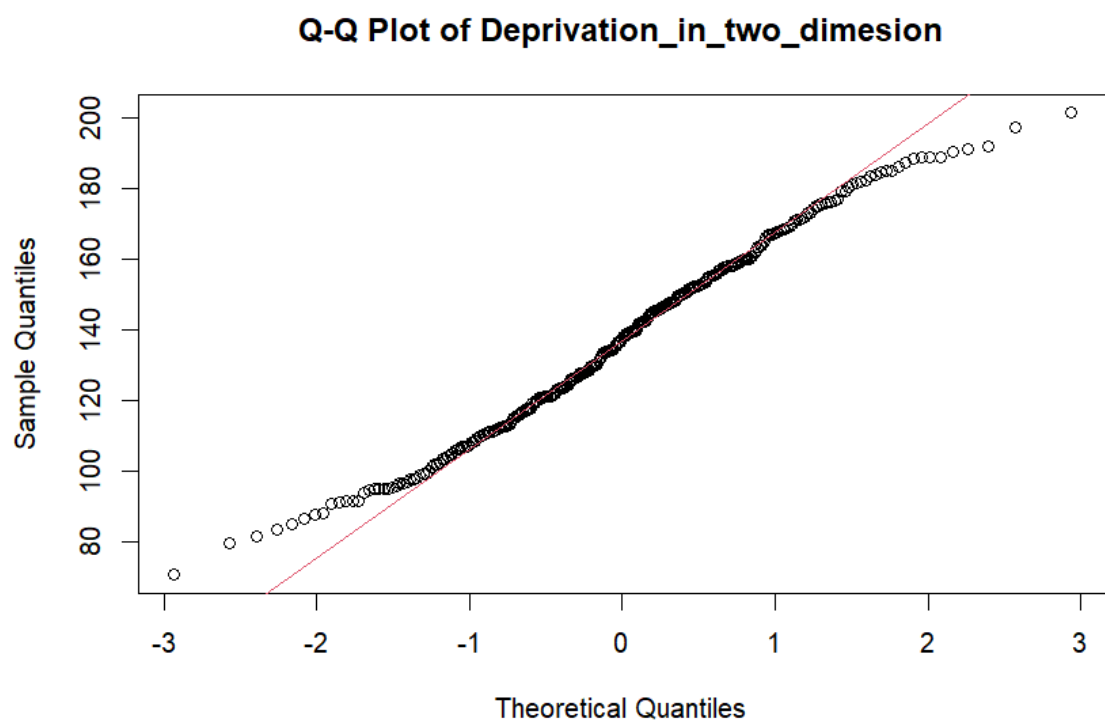
Figure 27: Q-Q plot of Deprivation in two dimensions

Boxplot of Deprivation_in_three_dimesion
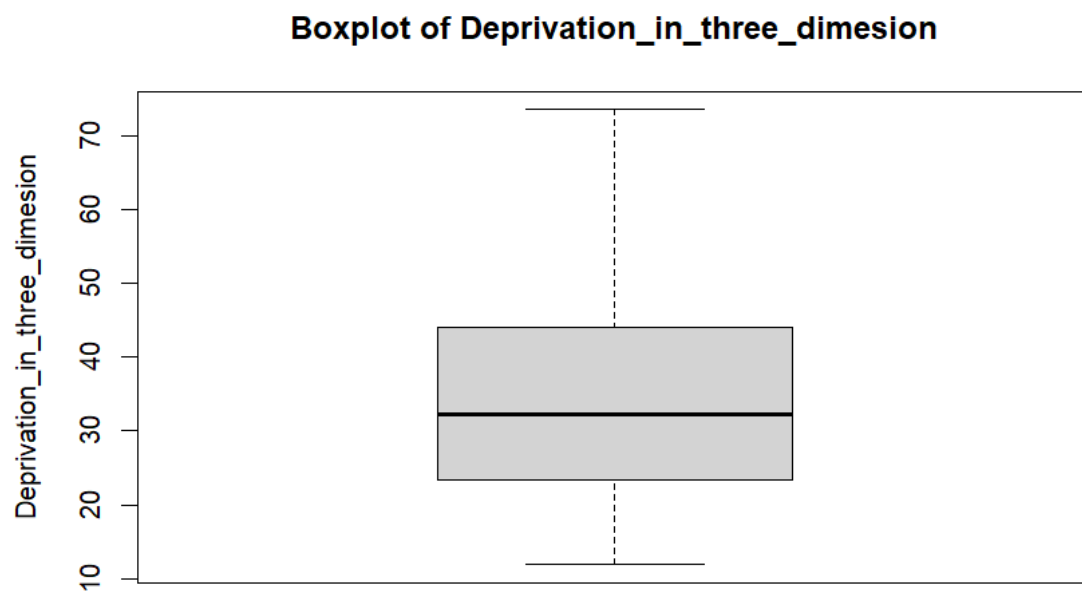
Figure 28:Boxplot of Deprivation in three dimensions

Histogram of Deprivation_in_three_dimesion
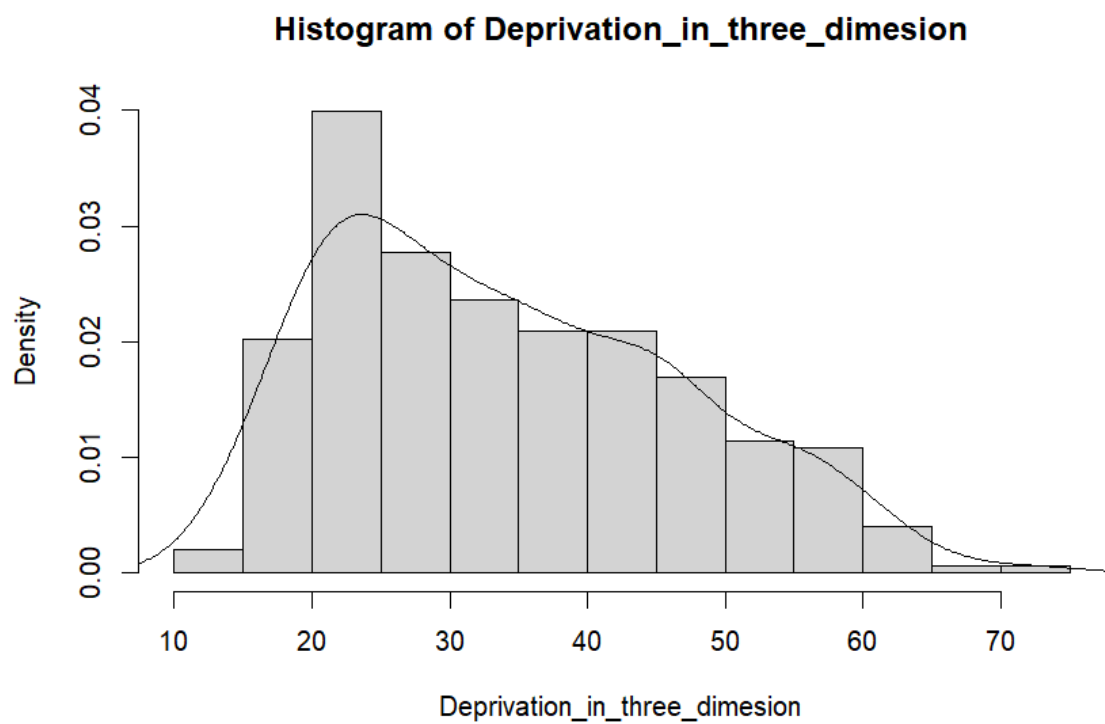
Figure 29: Histogram of Deprivation in three dimensions

Q-Q Plot of Deprivation_in_three_dimesion
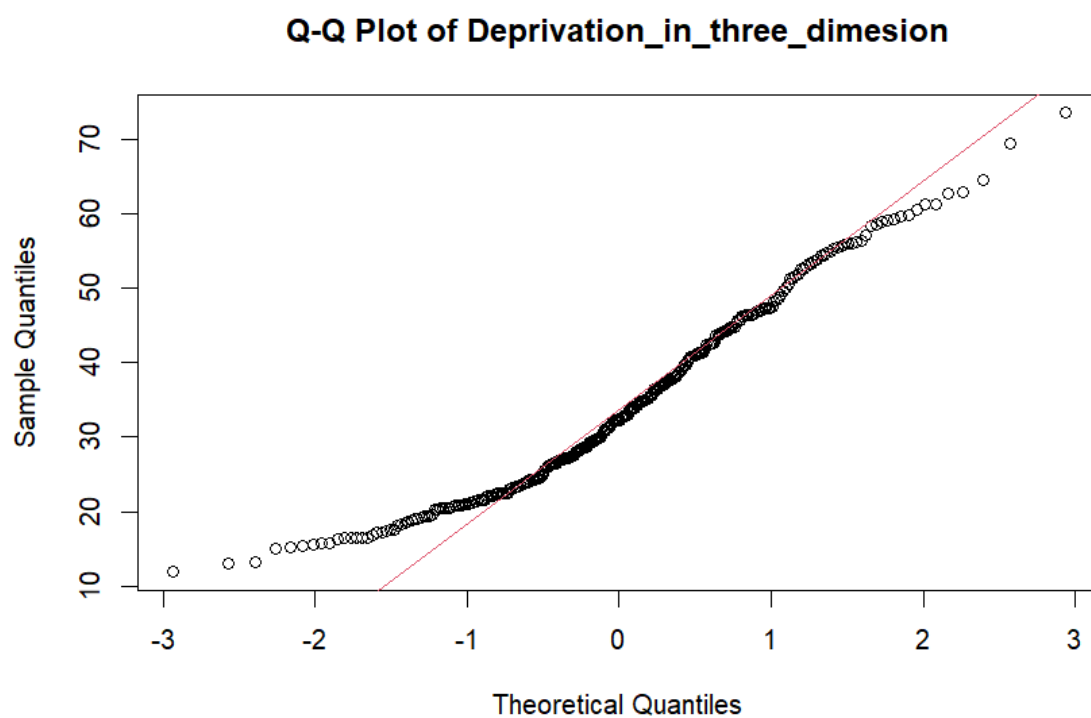
Figure 30: Q-Q plot of Deprivation in three dimensions

Boxplot of Deprivation_in_four_dimesion
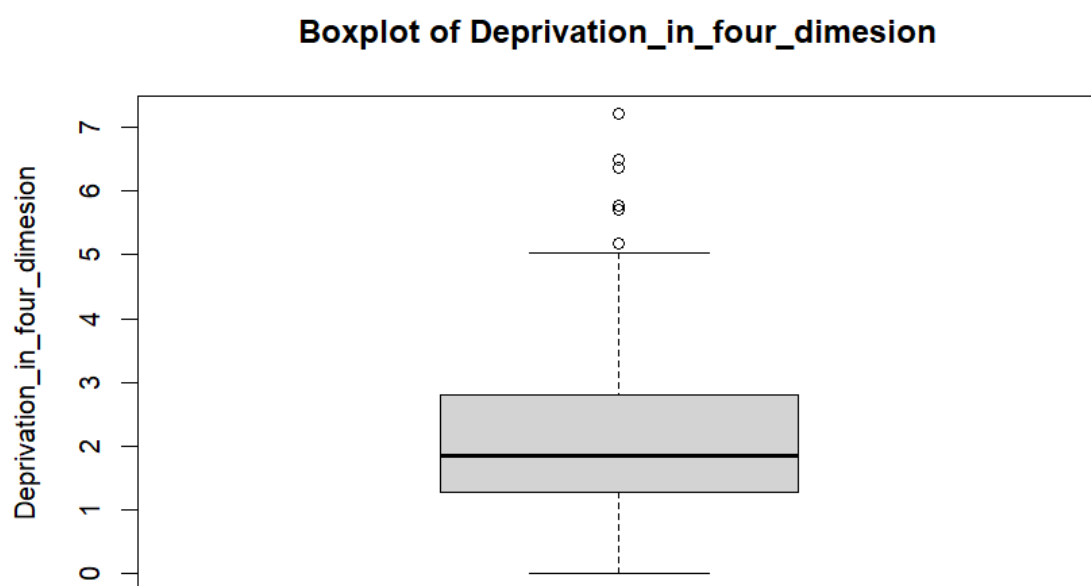
Figure 31: Boxplot of Deprivation in four dimensions

Histogram of Deprivation_in_four_dimesion
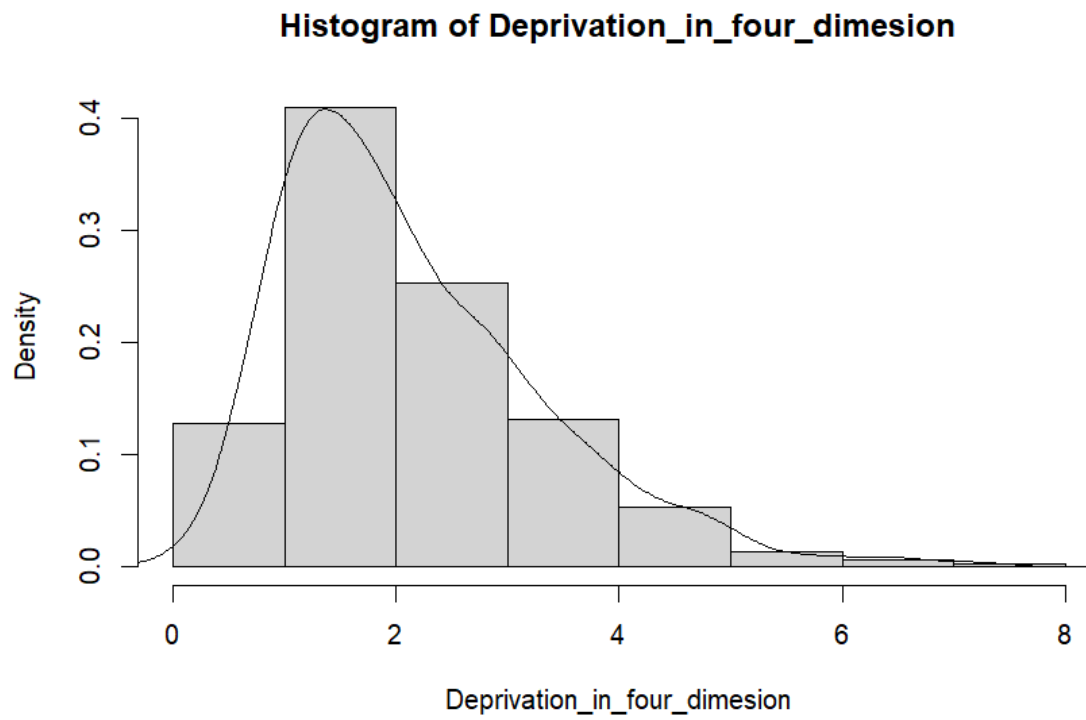
Figure 32:Histogram of Deprivation in four dimensions

Q-Q Plot of Deprivation_in_four_dimesion
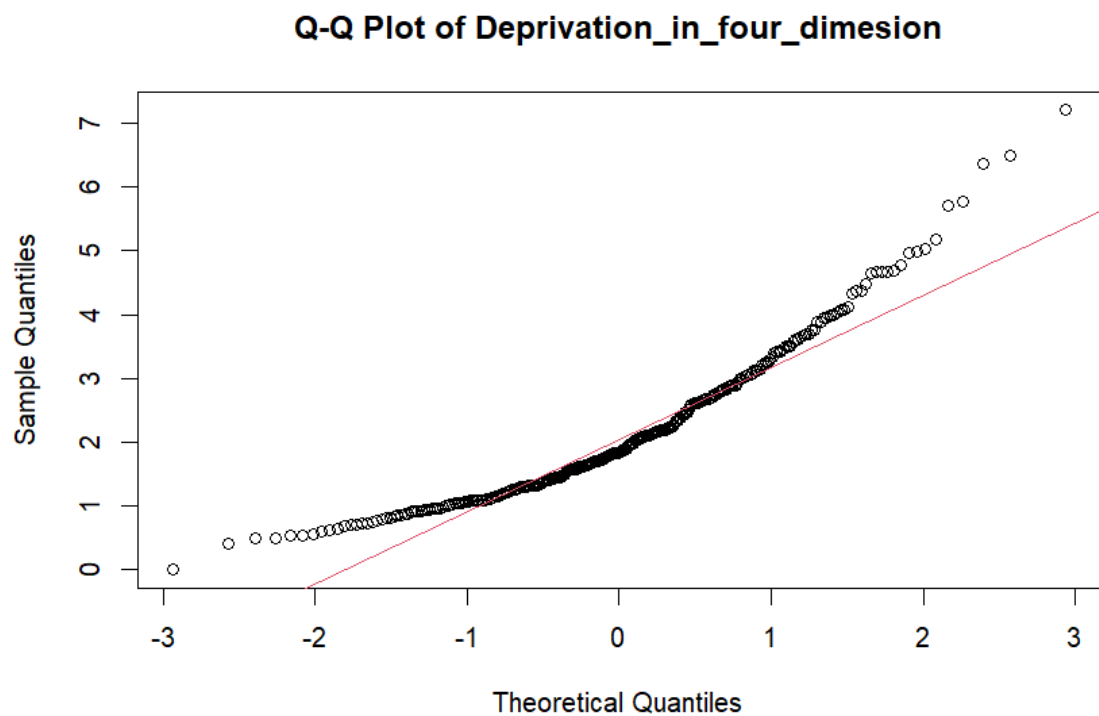
Figure 33: Q-Q plot of Deprivation in four dimensions

Boxplot of Managers._directors_and_senior_officials
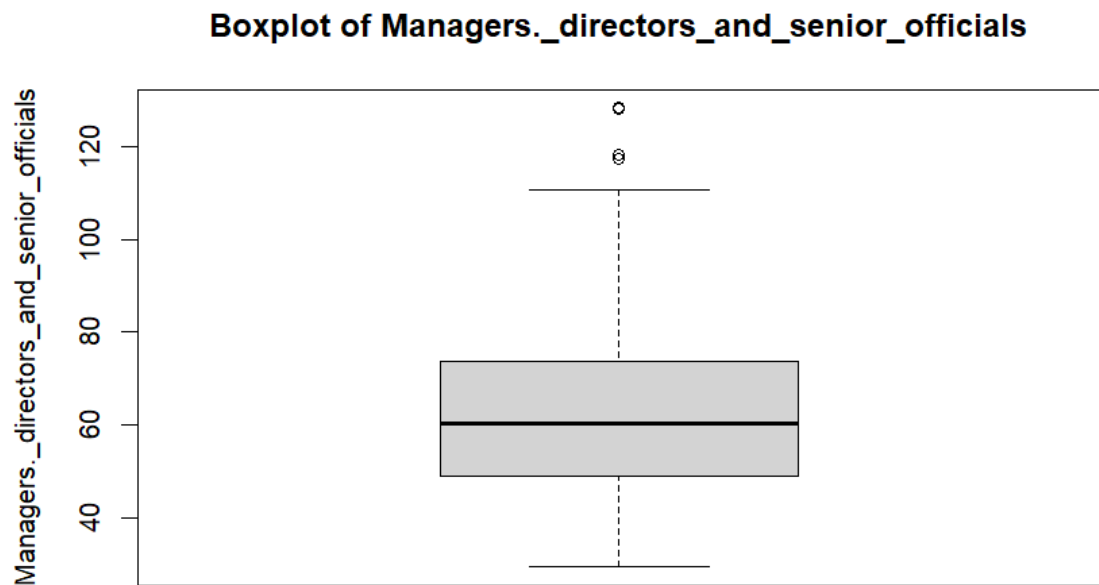
Figure 34: Boxplot of Managers, directors and senior officials

Histogram of Managers._directors_and_senior_officials
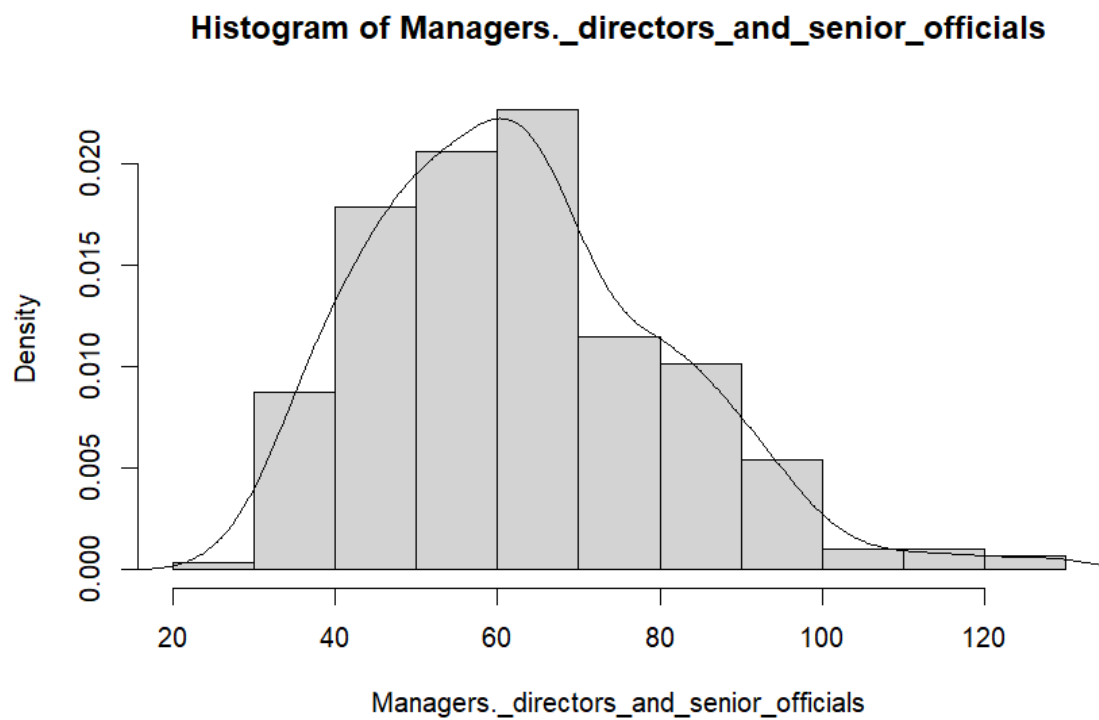
Figure 35:Histogram of Managers, directors and senior officials

Q-Q Plot of Managers._directors_and_senior_officials
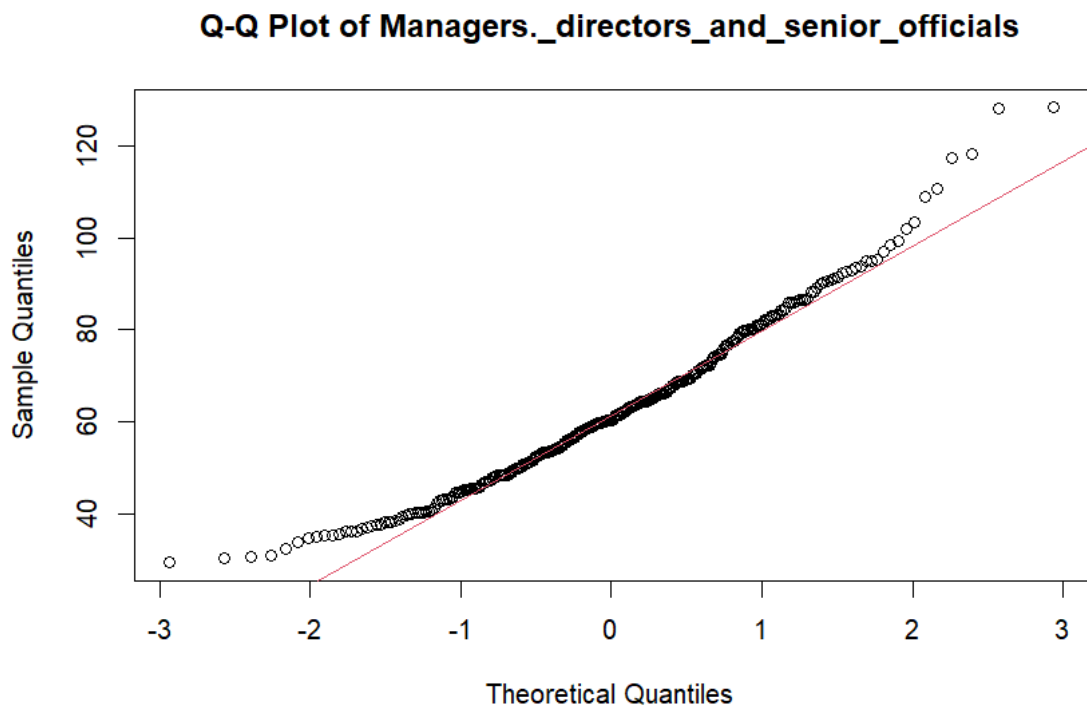
Figure 36: Q-Q plot of Managers, directors and senior officials

Boxplot of Professional_occupations
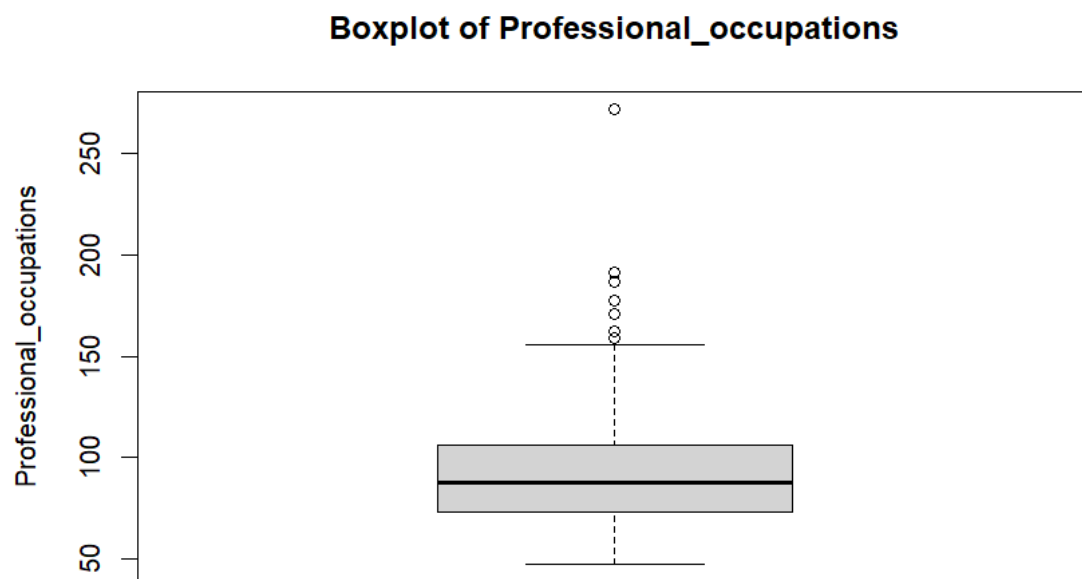
Figure 37: Boxplot of professional occupations

Histogram of Professional_occupations
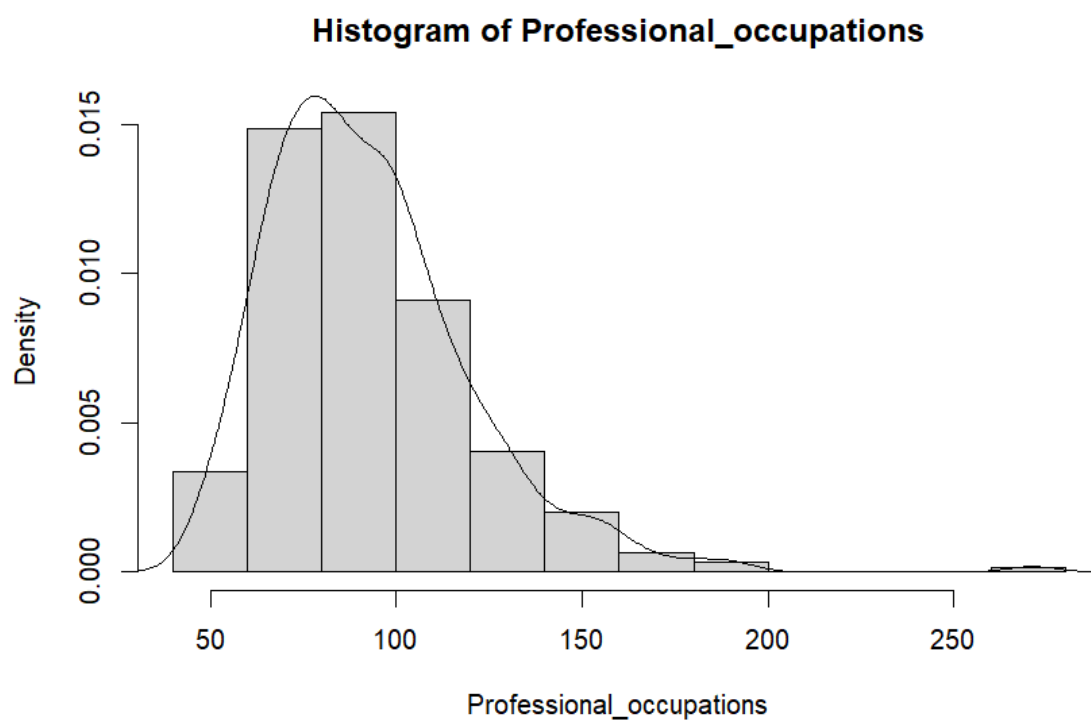
Figure 38:Histogram of professional occupations

Q-Q Plot of Professional_occupations
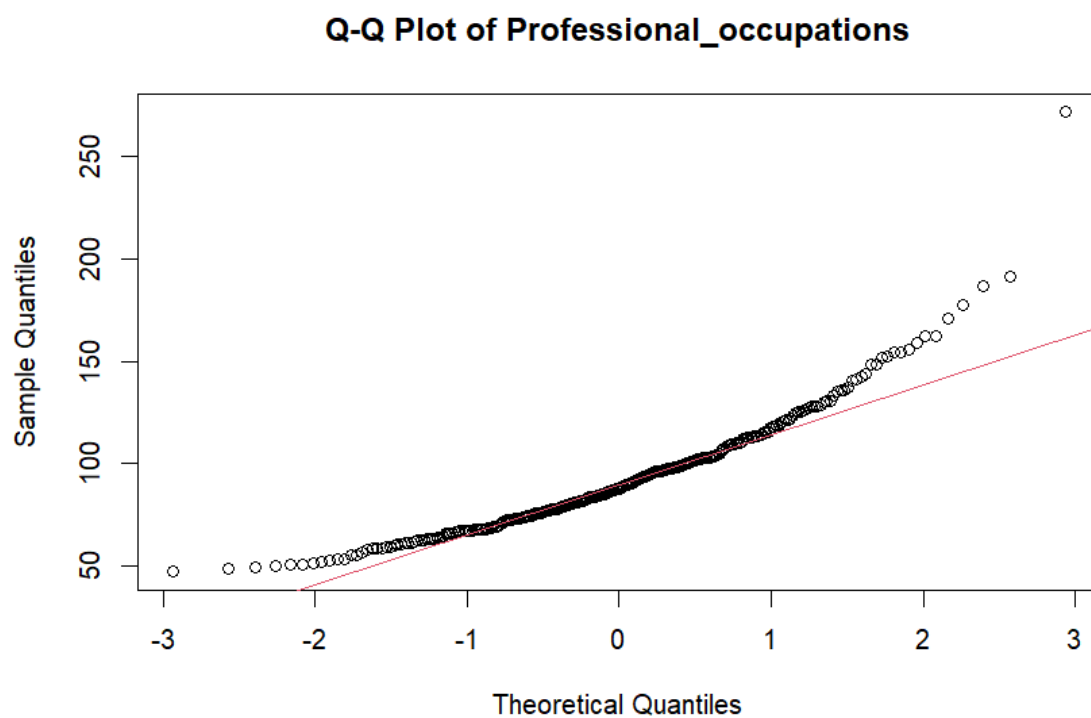
Figure 39: Q-Q plot of professional occupations

Boxplot of Associate_professional_and_technical_occupations
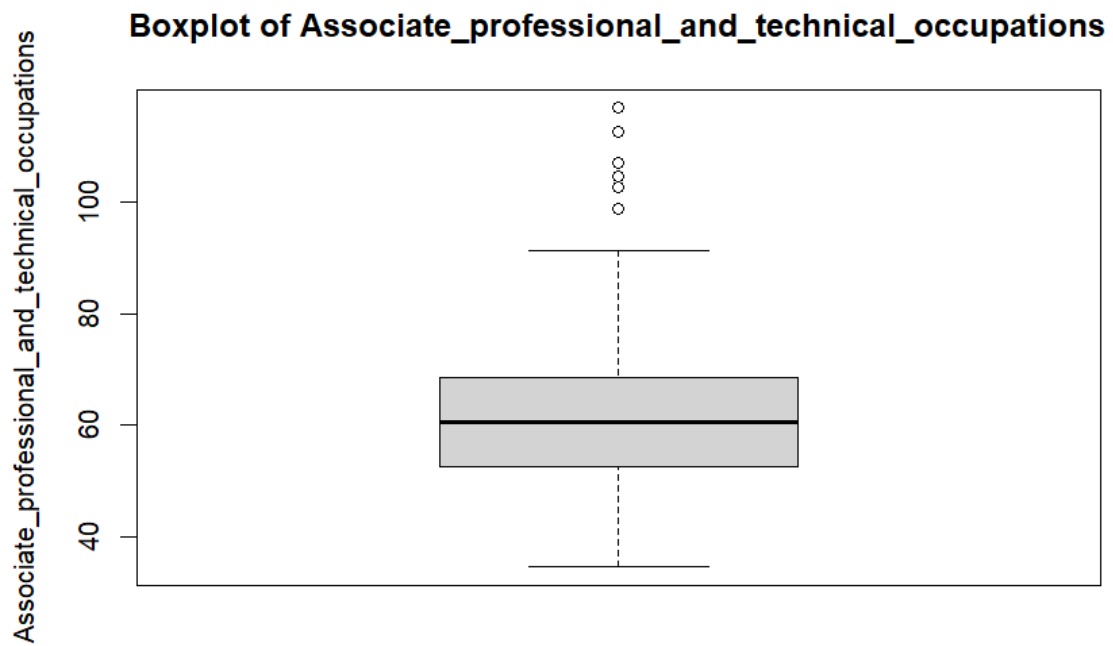
Figure 40: Boxplot of associate Professional

Histogram of Associate_professional_and_technical_occupations
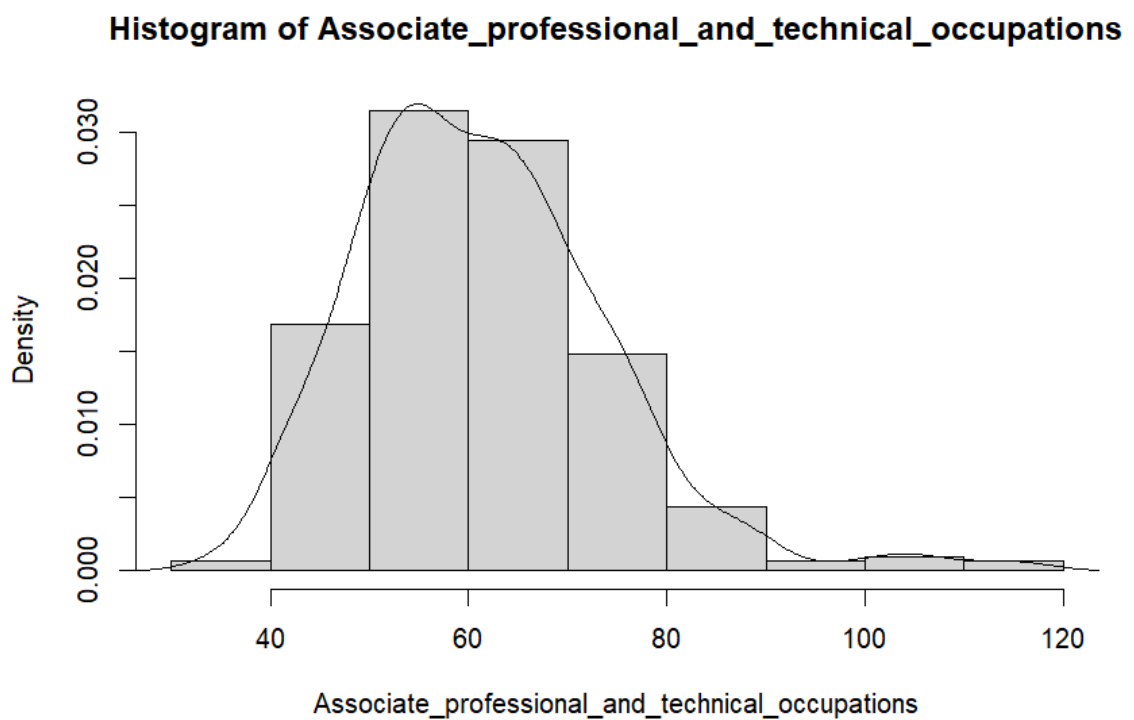
Figure 41: Histogram of Associate Professional

Q-Q Plot of Associate_professional_and_technical_occupations
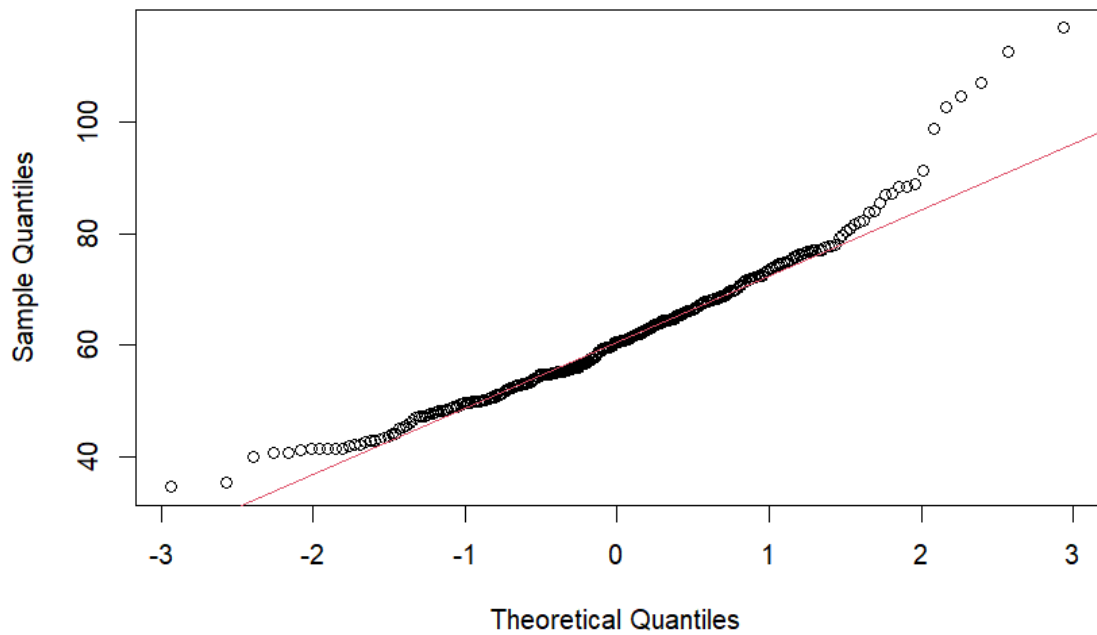
Figure 42: Q-Q plot Associate Professionals

Boxplot of Administrative_and_secretarial_occupations
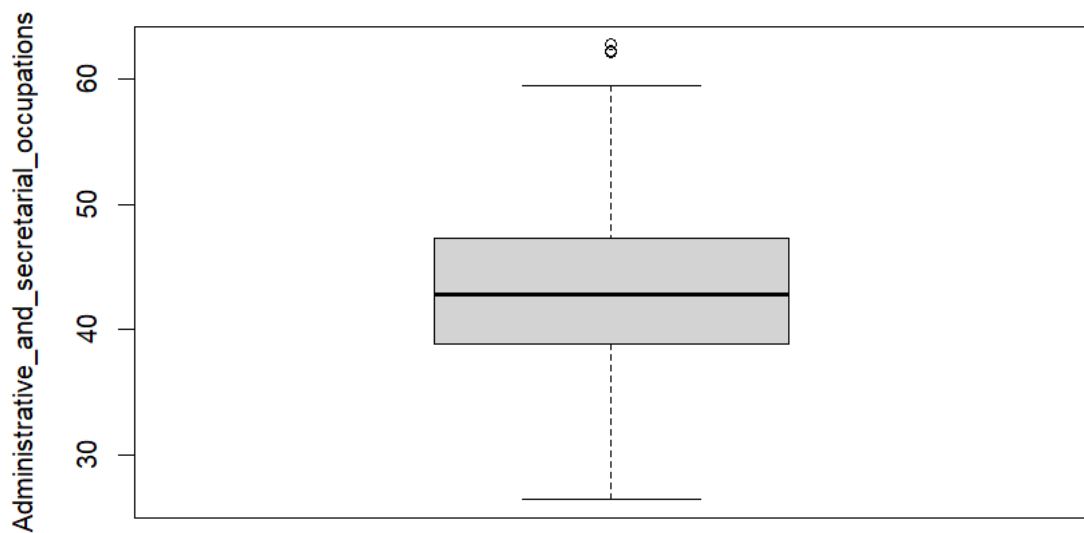
Figure 43: Boxplot of Administrative Roles

Histogram of Administrative_and_secretarial_occupations
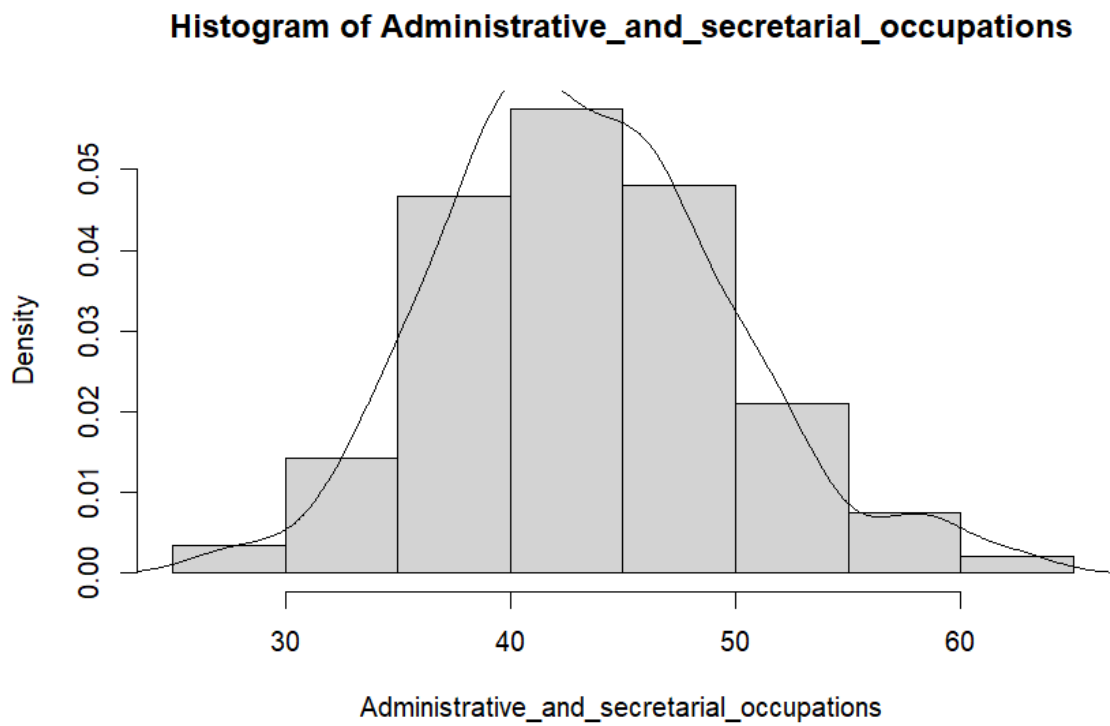
Figure 44: Histogram of Administrative Roles

Q-Q Plot of Administrative_and_secretarial_occupations
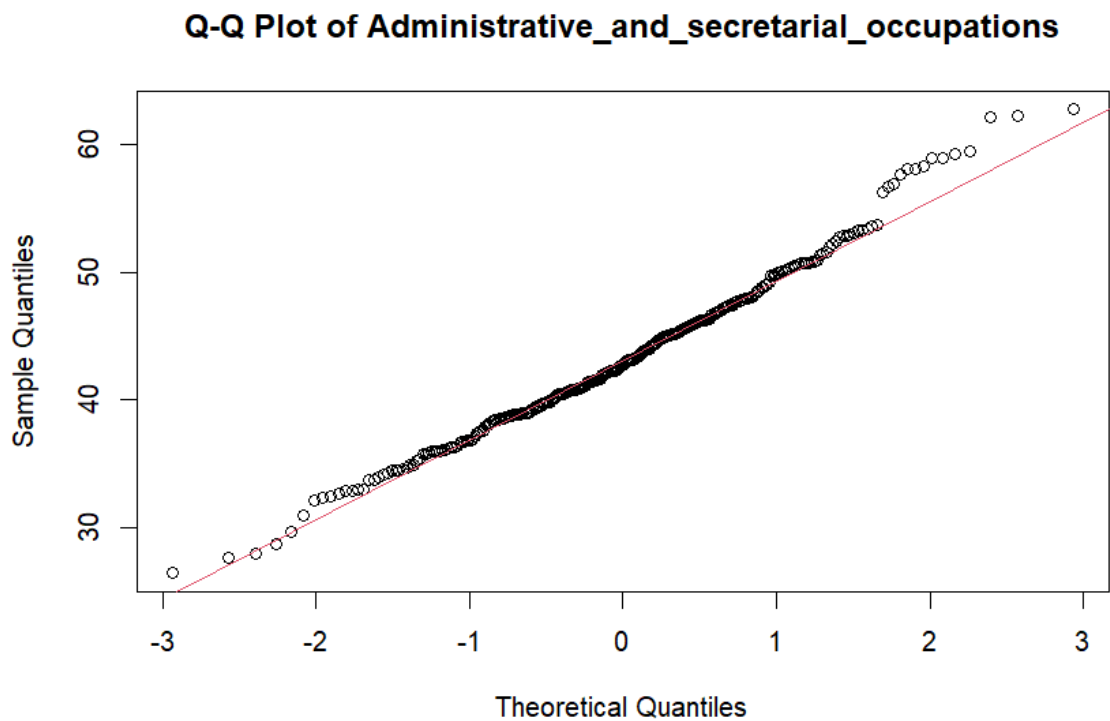
Figure 45: Q-Q plot of Administrative Roles

Boxplot of Skilled_trades_occupations
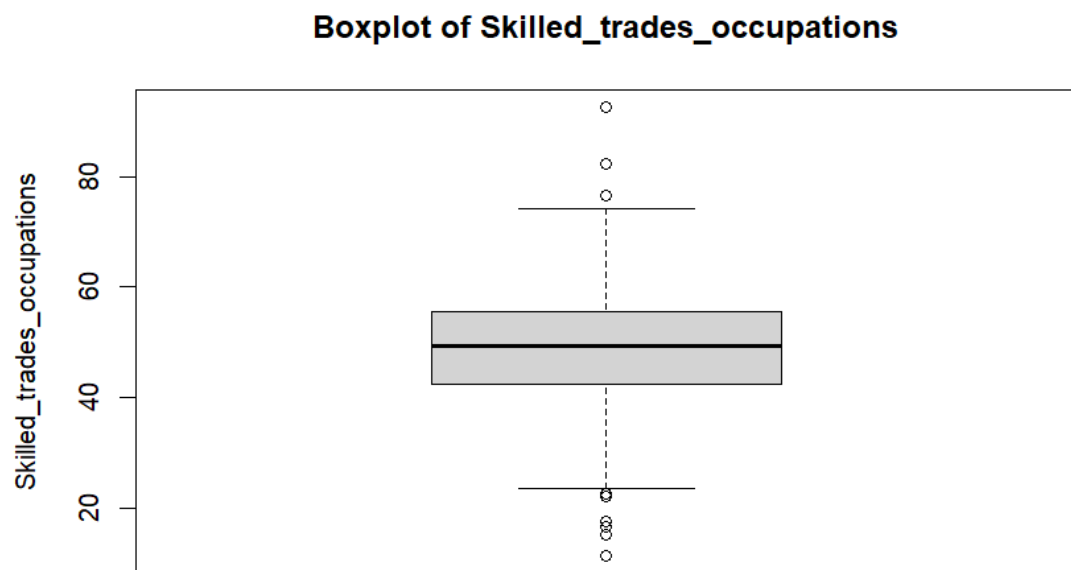
Figure 46: Boxplot of Skilled Trades

Histogram of Skilled_trades_occupations
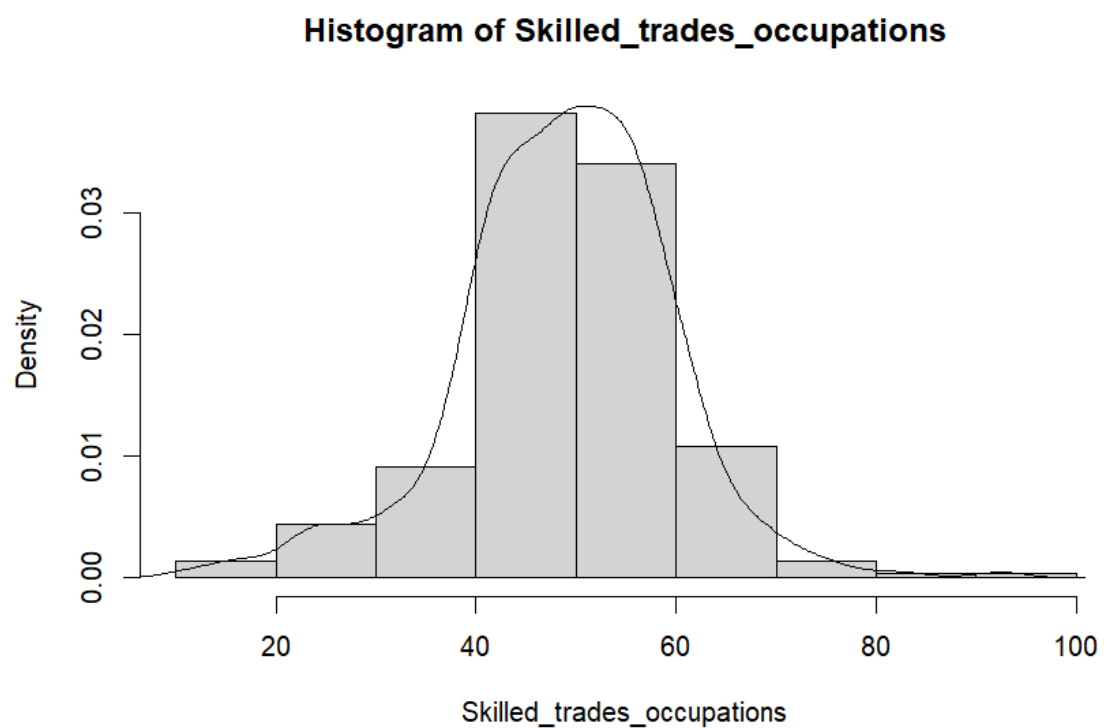
Figure 47: Histogram of Skilled Trades
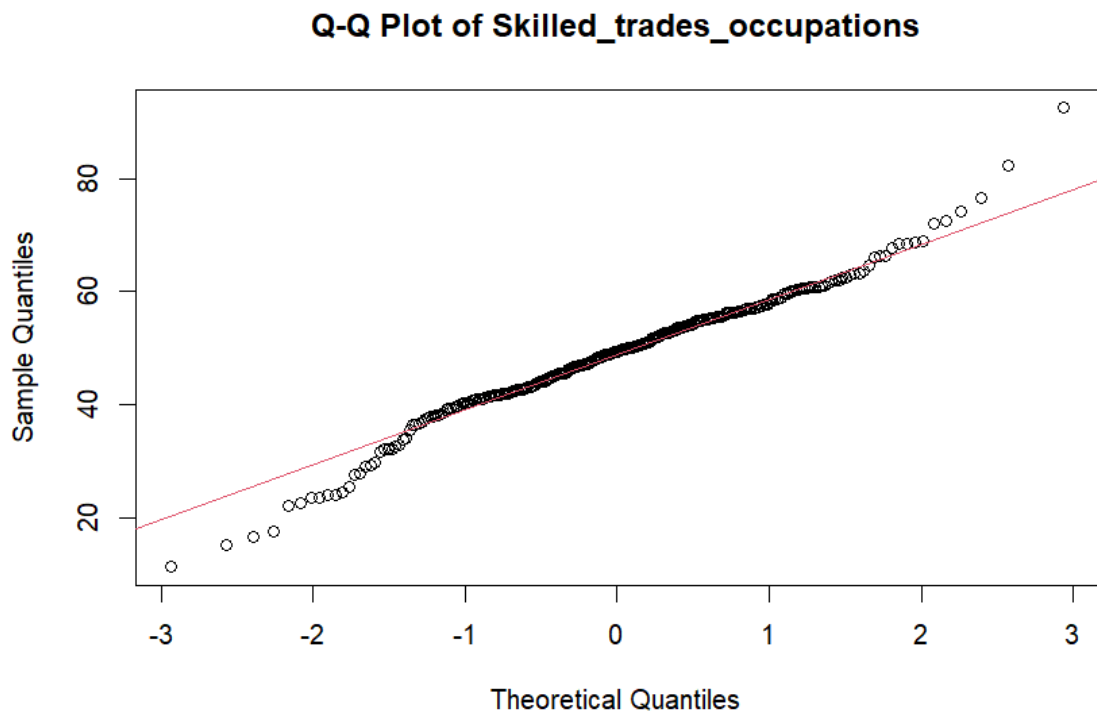
Figure 48: Q-Q plot of Skilled Trades
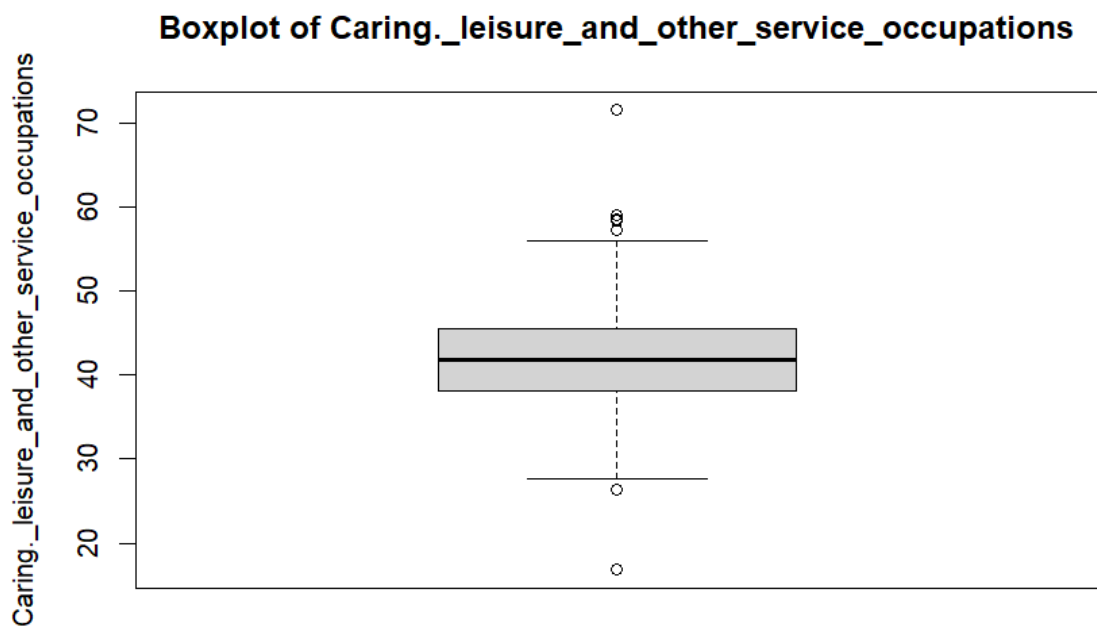
Figure 49: Boxplot of Caring & Leisure
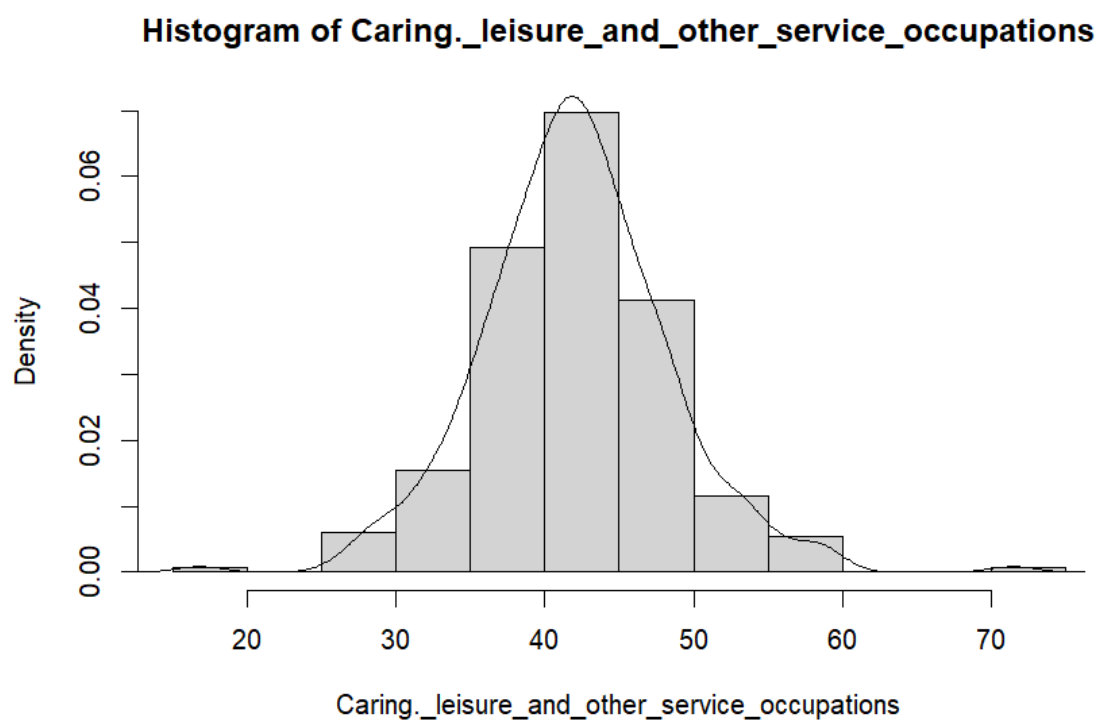
Figure 50: Histogram of Caring & Leisure
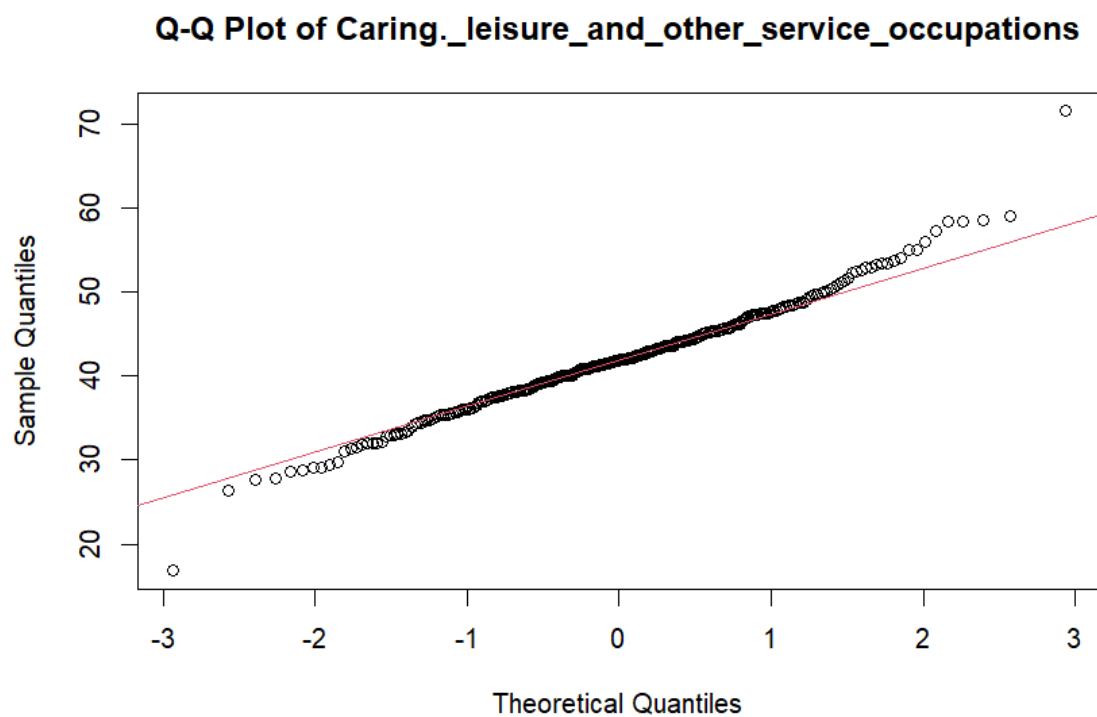
Figure 51: Q-Q plot of Caring & Leisure

Boxplot of Sales_and_customer_service_occupations
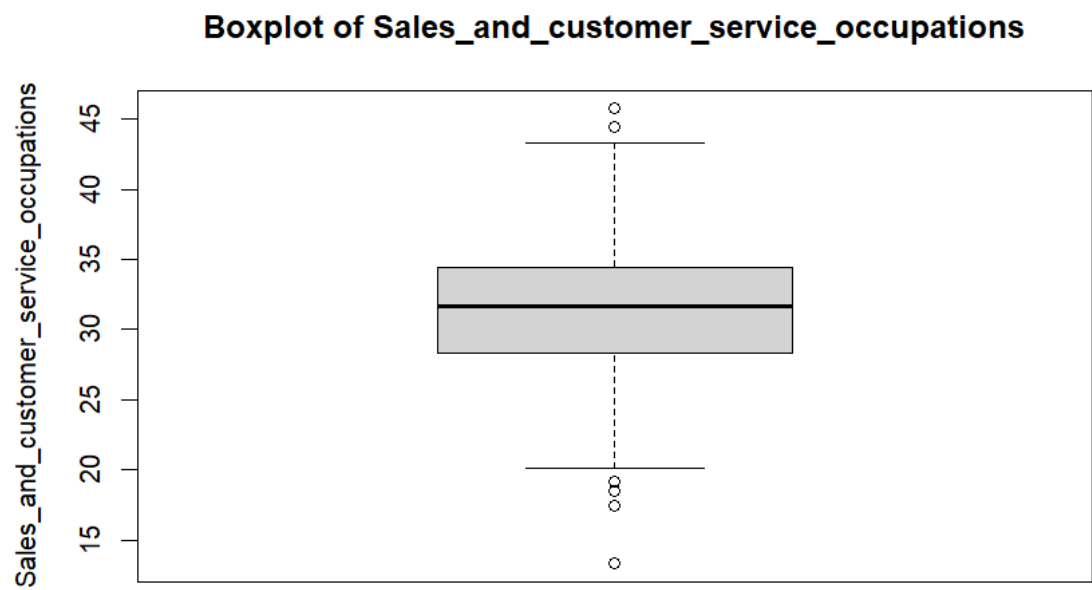
Figure 52: Boxplot of Sales

Histogram of Sales_and_customer_service_occupations
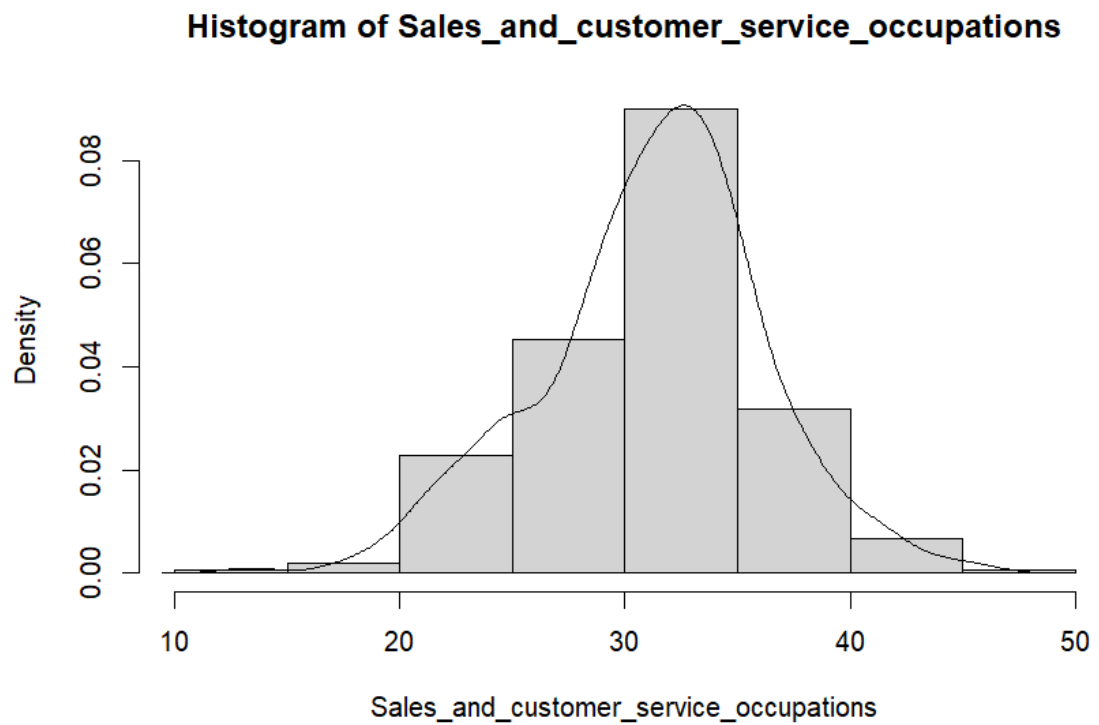
Figure 53: Histogram of sales

Q-Q Plot of Sales_and_customer_service_occupations
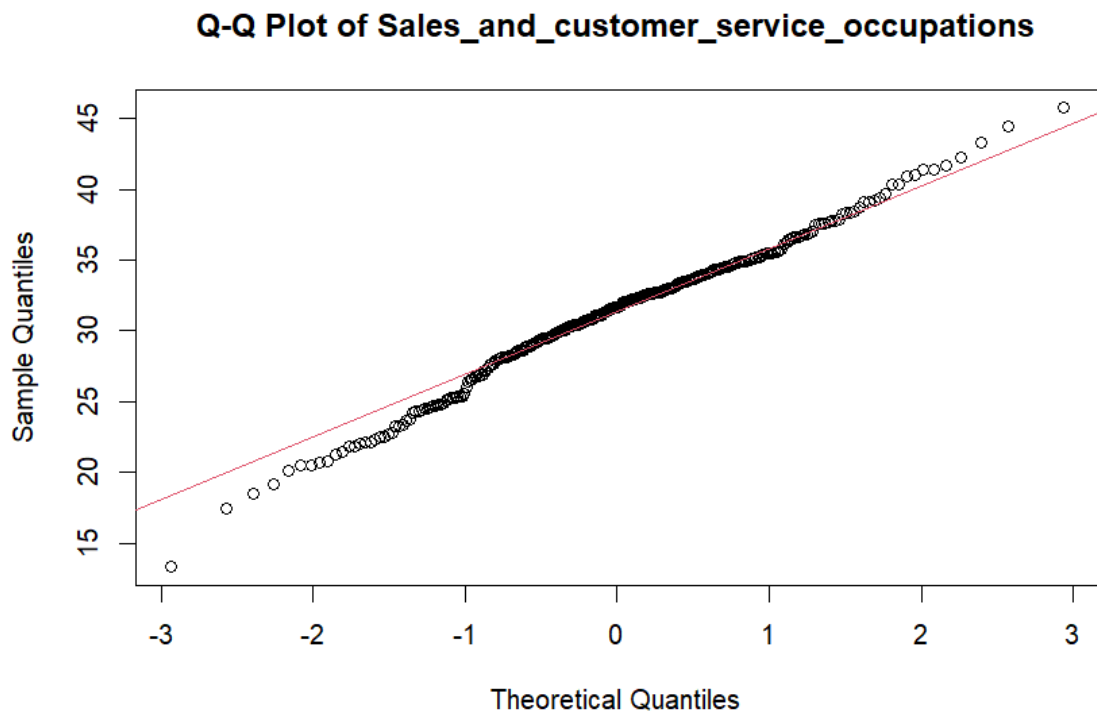
Figure 54: Q-Q plot of sales

Boxplot of Process._plant_and_machine_operatives
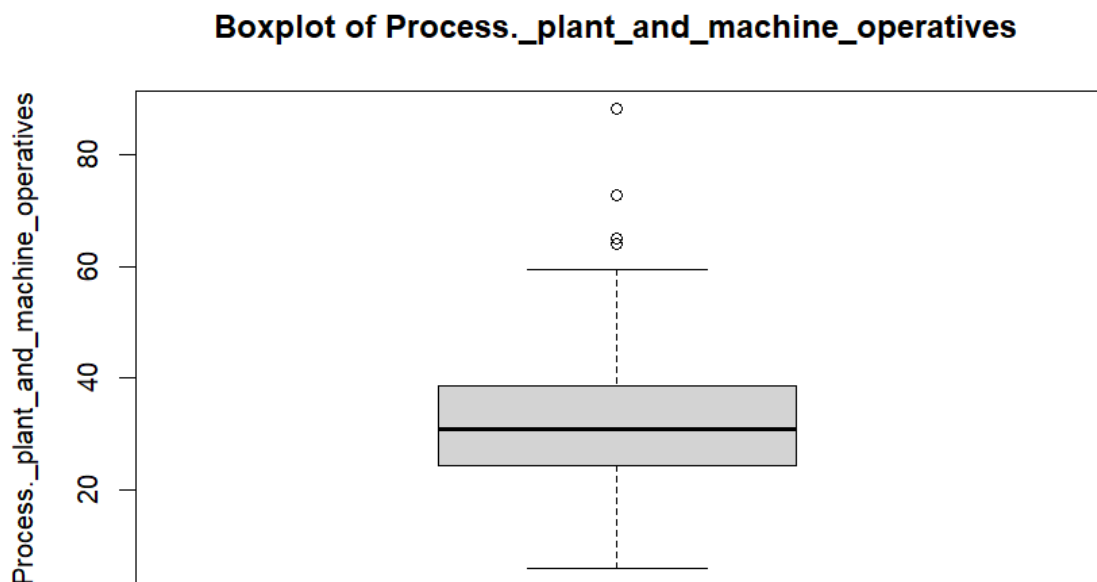
Figure 55: Boxplot of Process & Machine

Histogram of Process._plant_and_machine_operatives
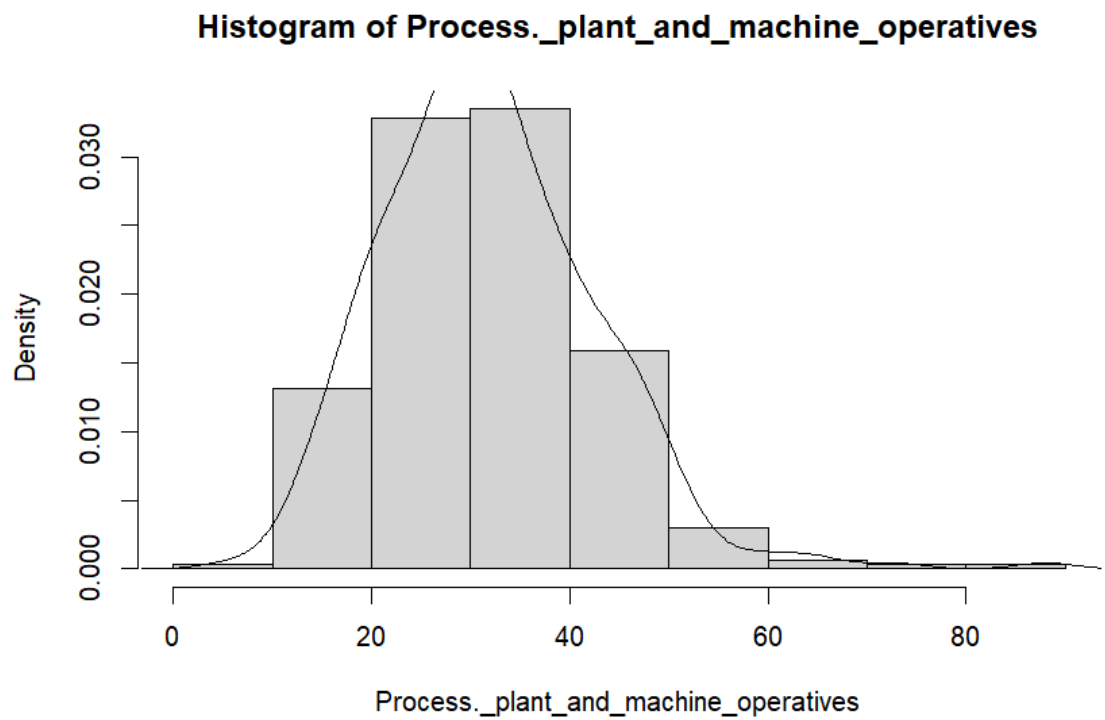
Figure 56: Histogram of Process & Machine

Q-Q Plot of Process._plant_and_machine_operatives
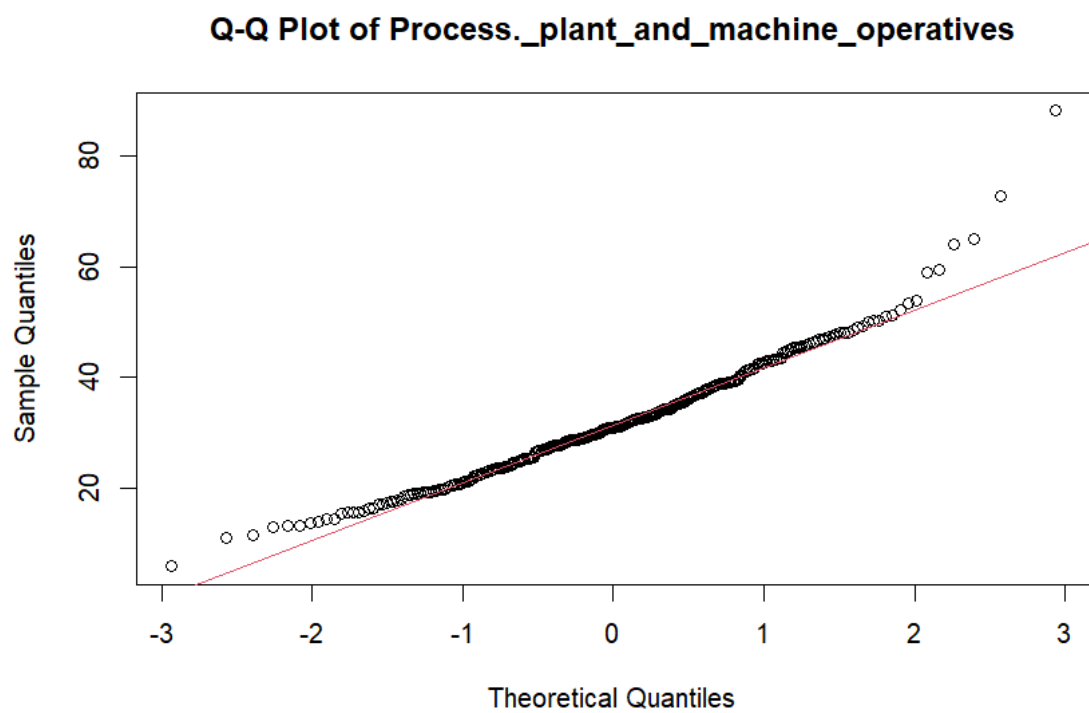
Figure 57: Q-Q plot of Process & Machine

Boxplot of Elementary_occupations
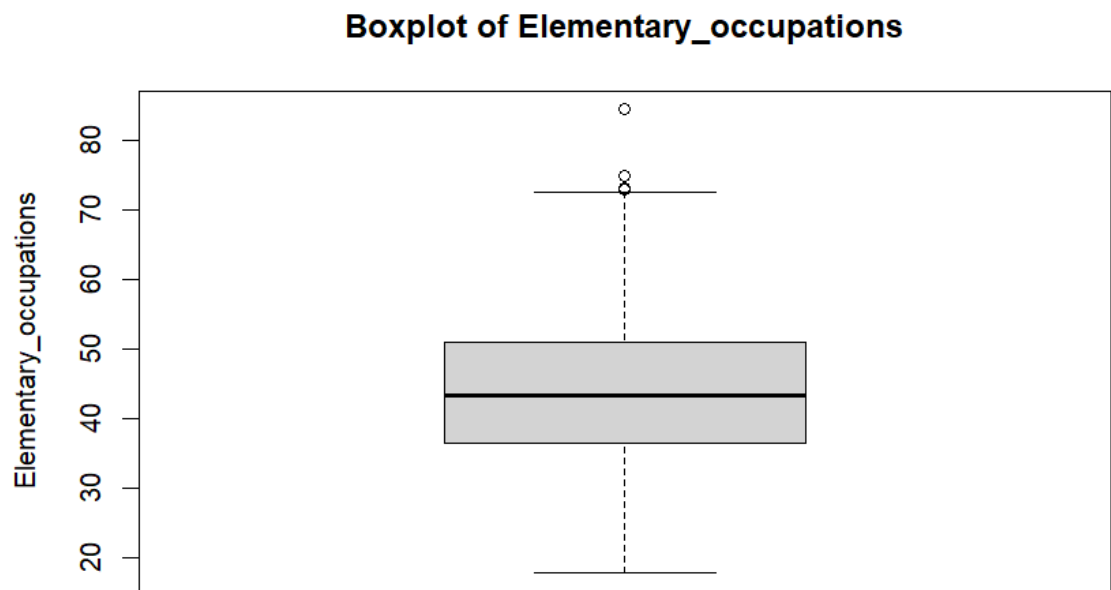
Figure 58: Boxplot of Elementry occupation

Histogram of Elementary_occupations
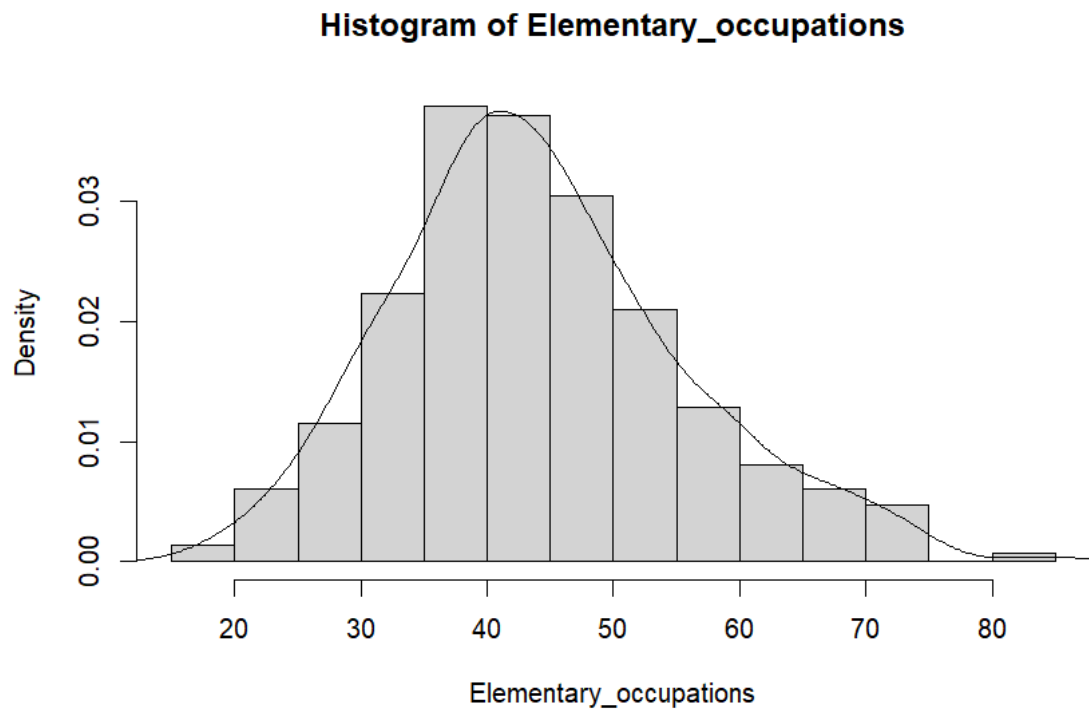
Figure 59: Histogram of Elementry Occupation

Q-Q Plot of Elementary_occupations
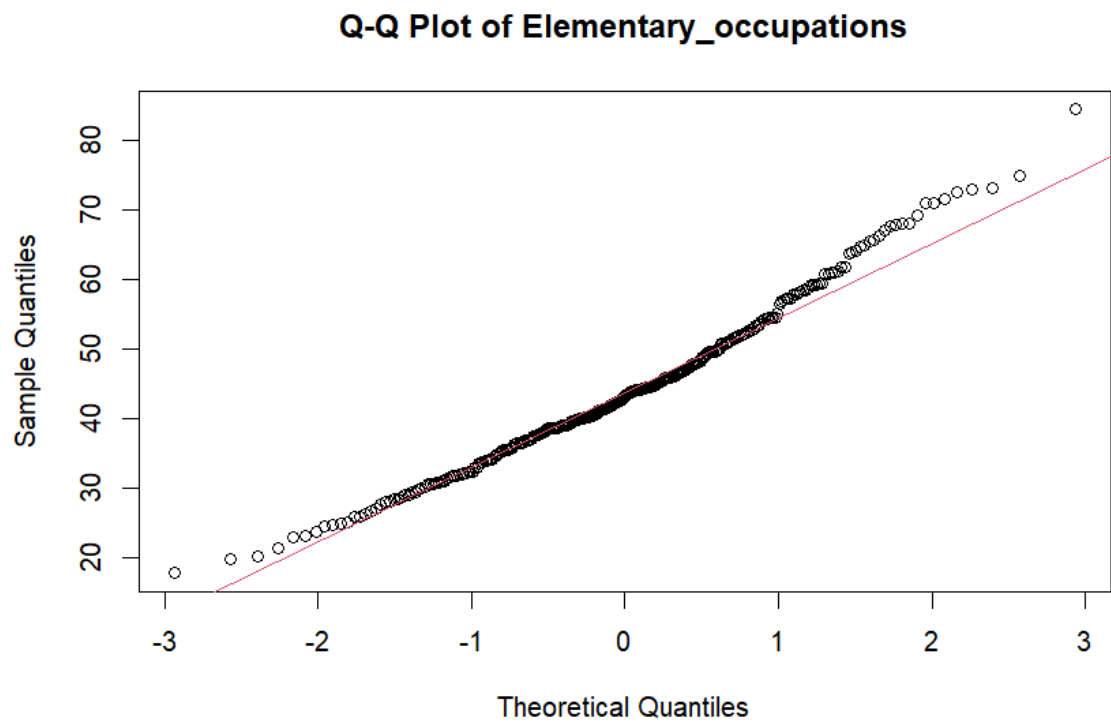
Figure 60: Q-Q plot of Elementry Occupation

Boxplot of No_qualifications
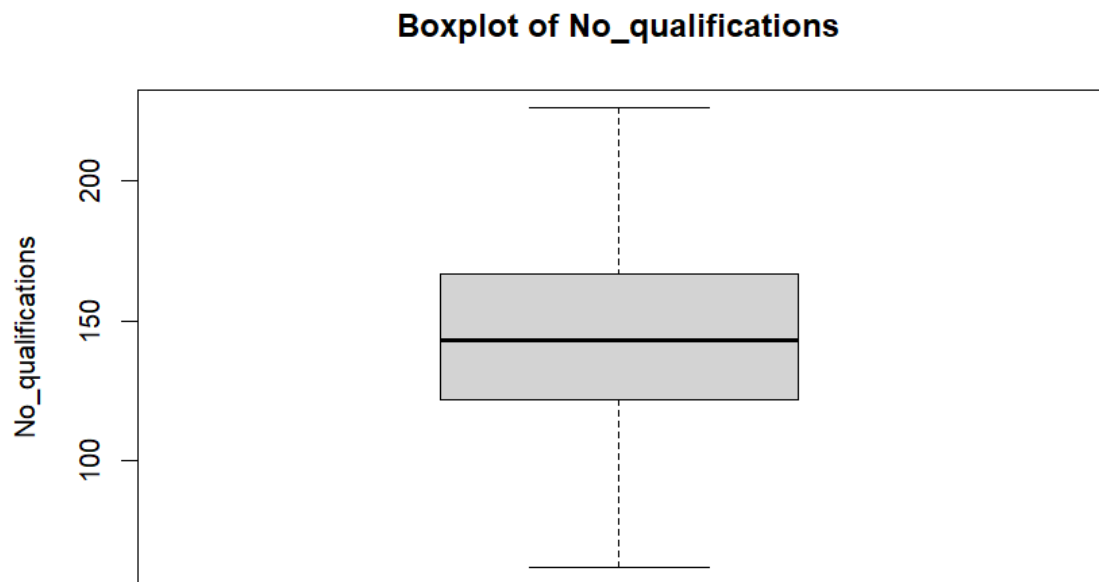
Figure 61: Boxplot of No qualifications

Histogram of No_qualifications
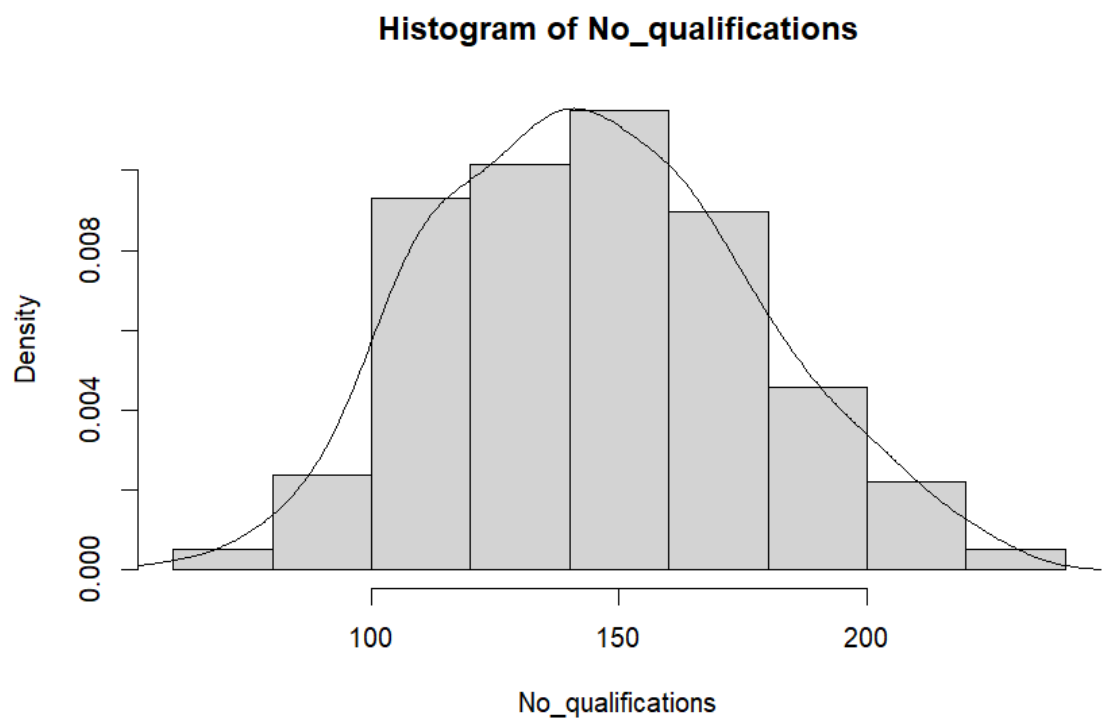
Figure 62: Histogram of No qualifications

Q-Q Plot of No_qualifications
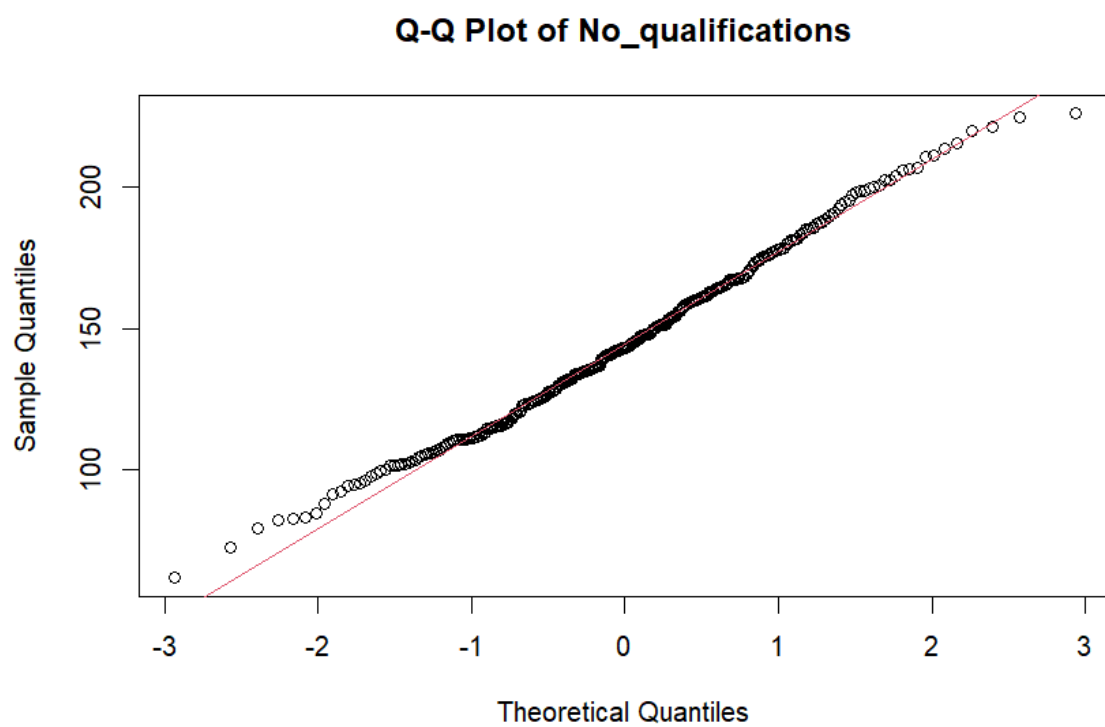
Figure 63: Q-Q plot of No qualifications

Boxplot of Level_1_and_entry_level_qualifications
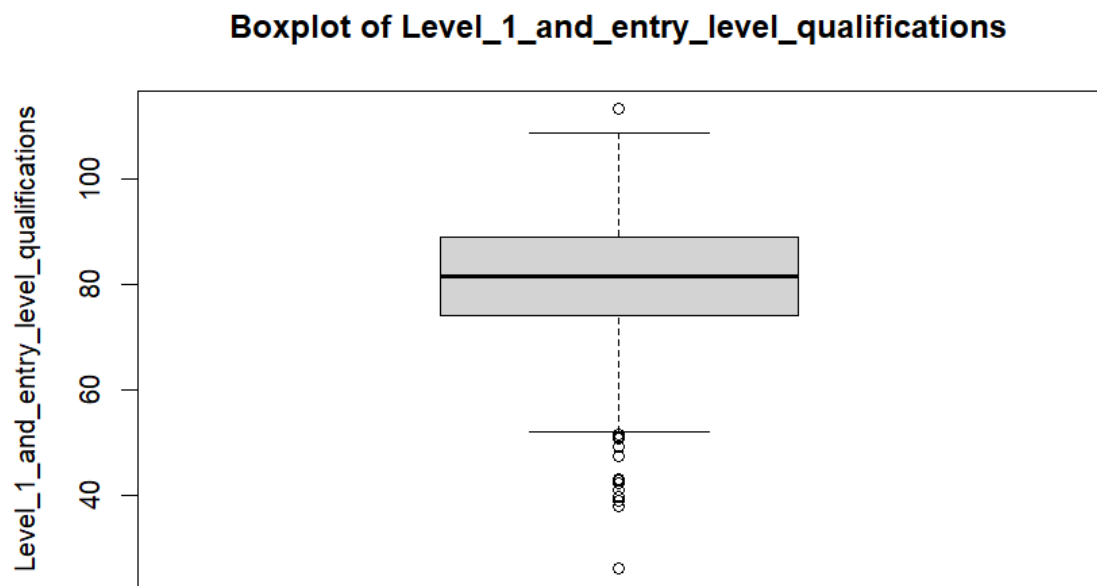
Figure 64: Boxplot of Level 1

Histogram of Level_1_and_entry_level_qualifications
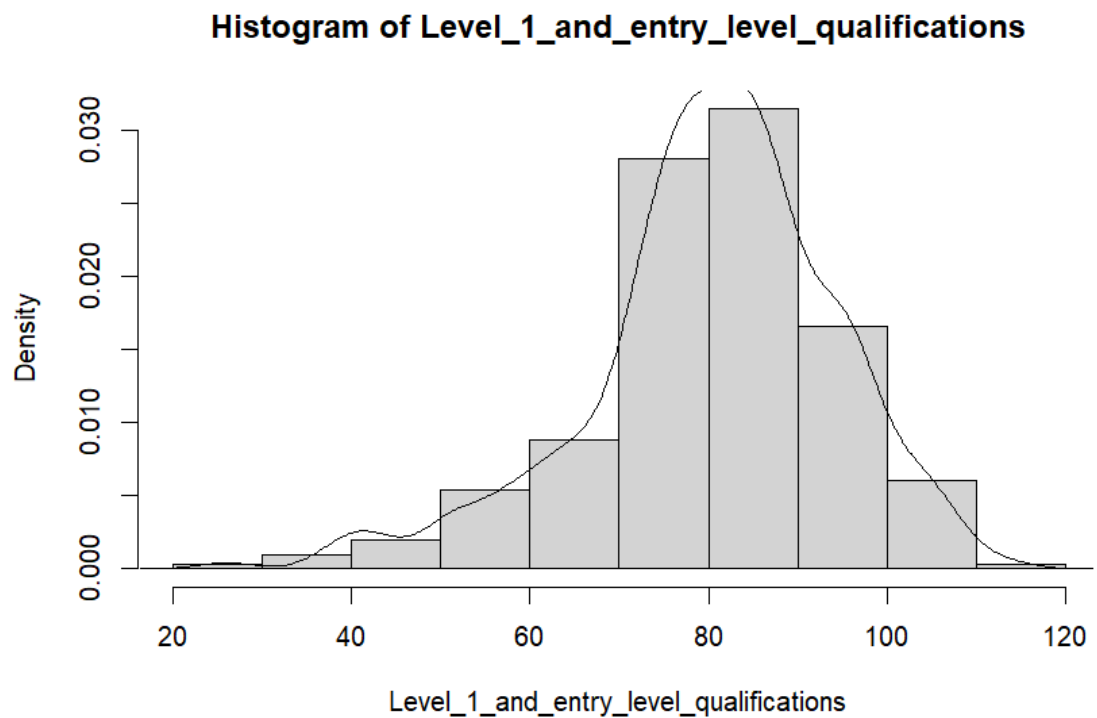
Figure 65: Histogram of Level 1

Q-Q Plot of Level_1_and_entry_level_qualifications
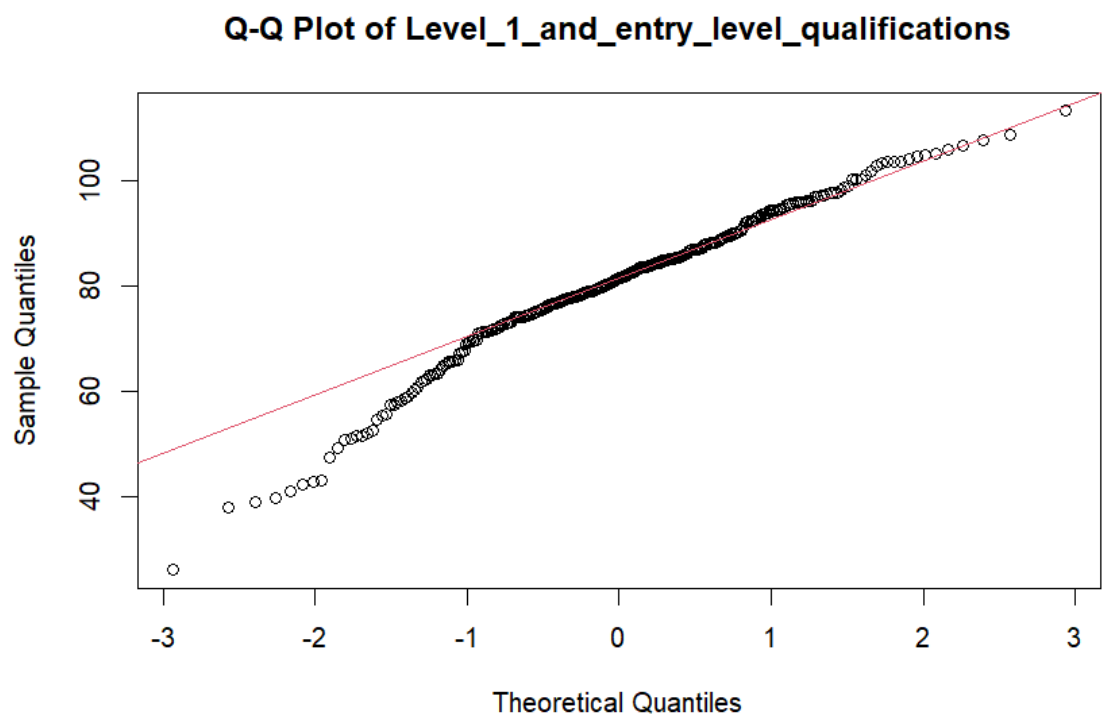
Figure 66: Q-Q plot of Level 1

Boxplot of Level_2_qualifications
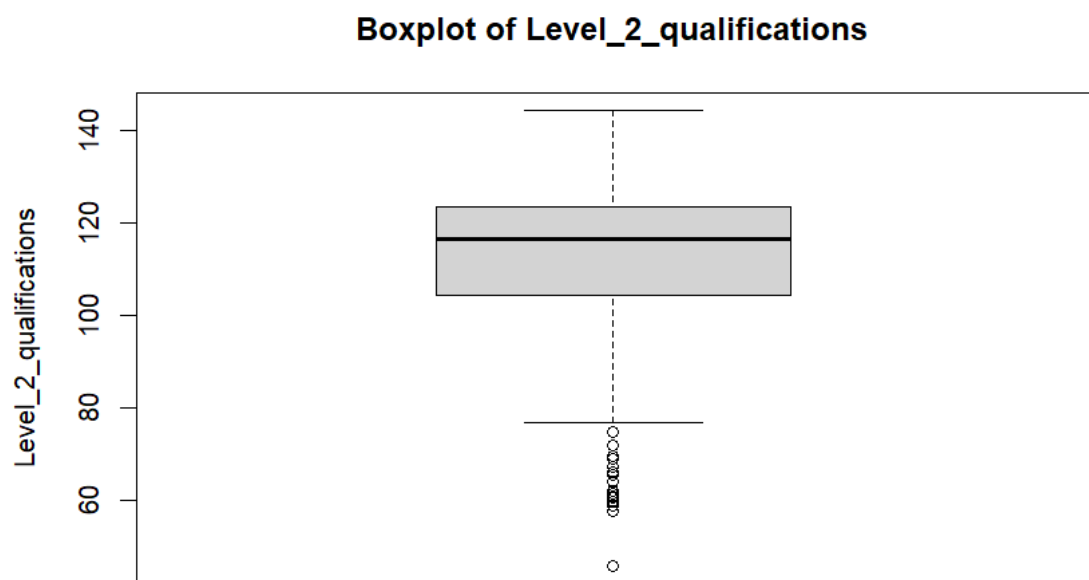
Figure 67: Boxplot of Level 2

Histogram of Level_2_qualifications

Figure 68: Histogram of Level 2

Q-Q Plot of Level_2_qualifications
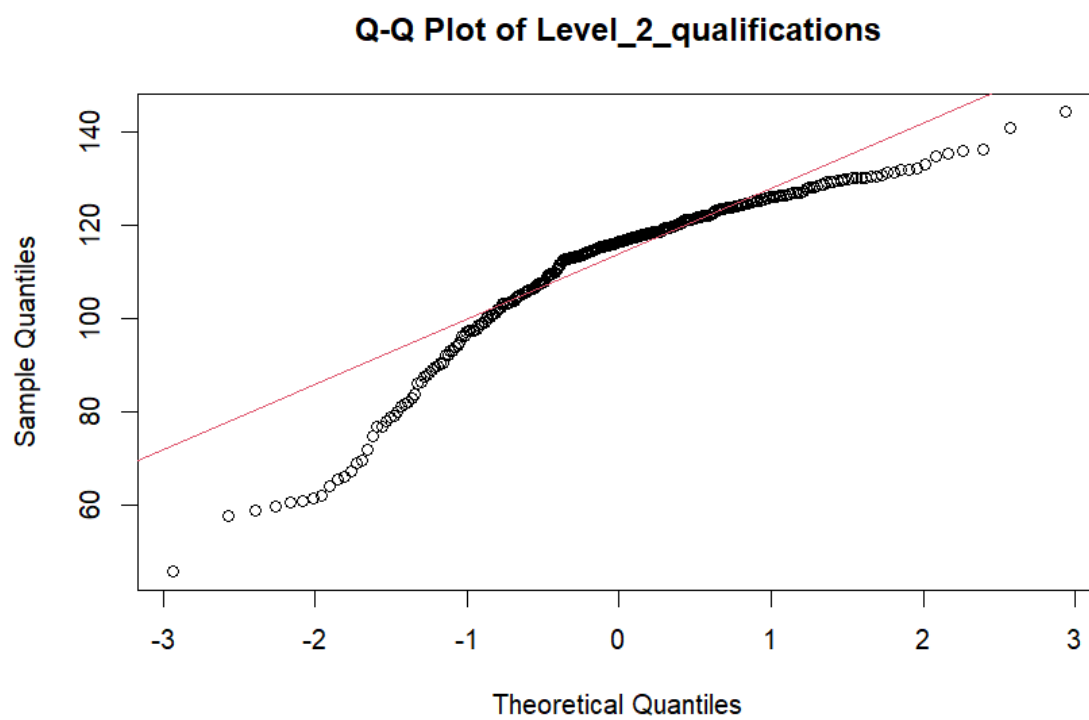
Figure 69: Q-Q plot of Level 2

Boxplot of Apprenticeship

Figure 70: Boxplot of Apprenticeship

Histogram of Apprenticeship

Figure 71:Histogram of Apprenticeship

Q-Q Plot of Apprenticeship
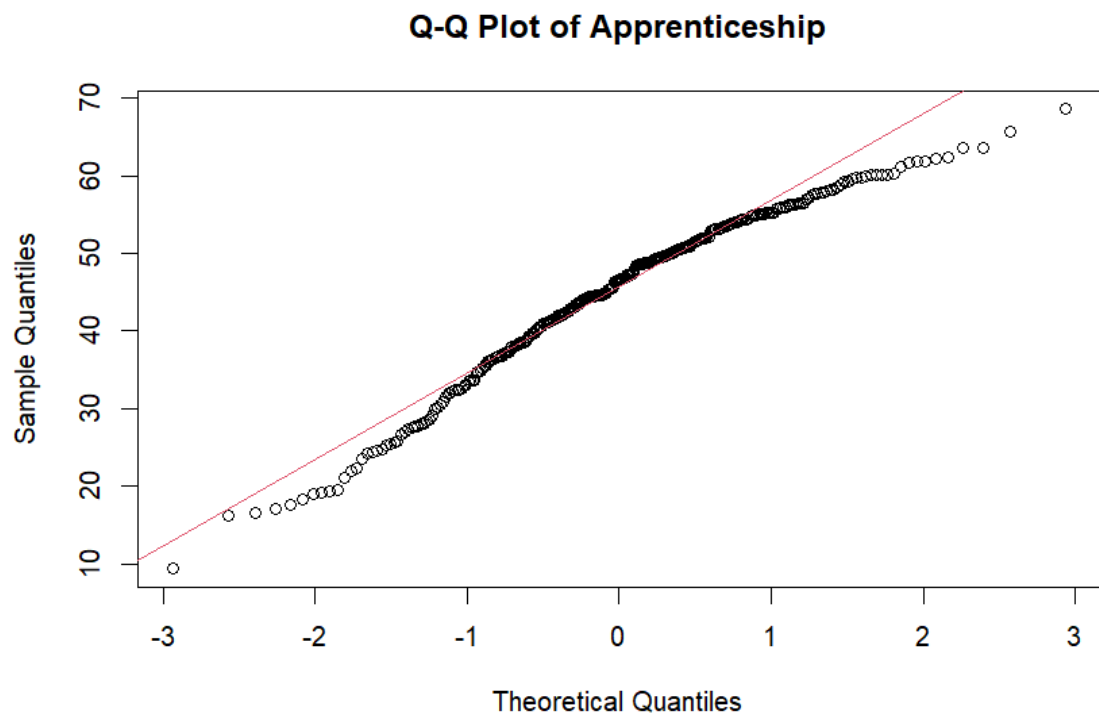
Figure 72: Q-Q plot of Apprenticeship

Boxplot of Level_3_qualifications

Figure 73: Boxplot of level 3
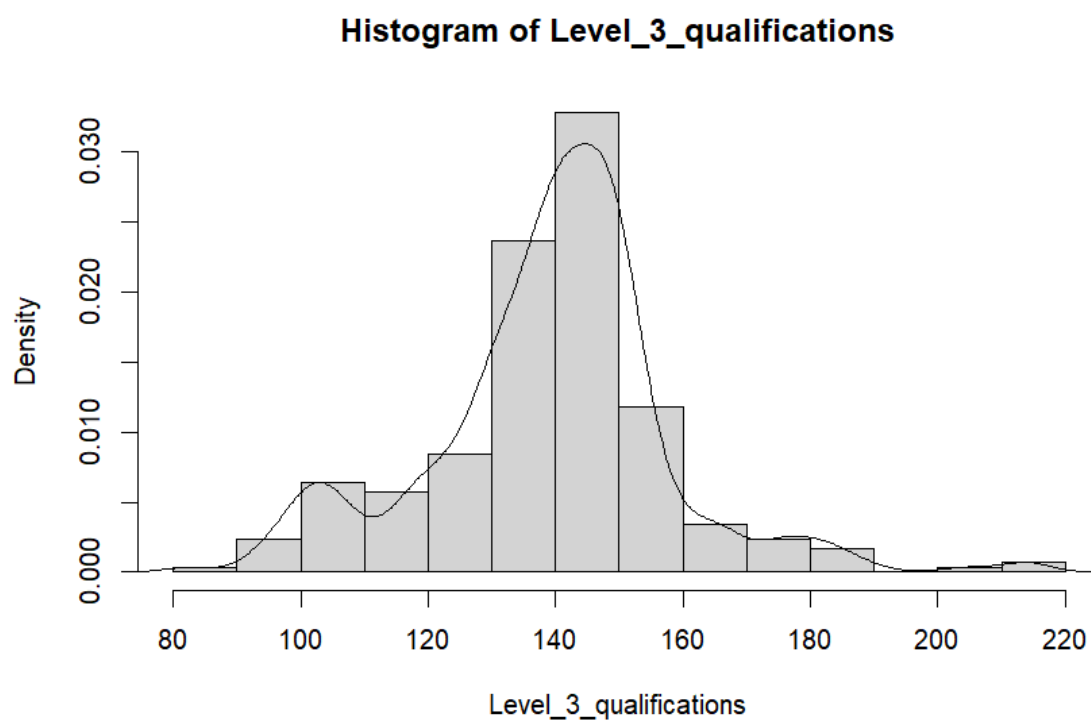
Figure 74: Histogram of level 3
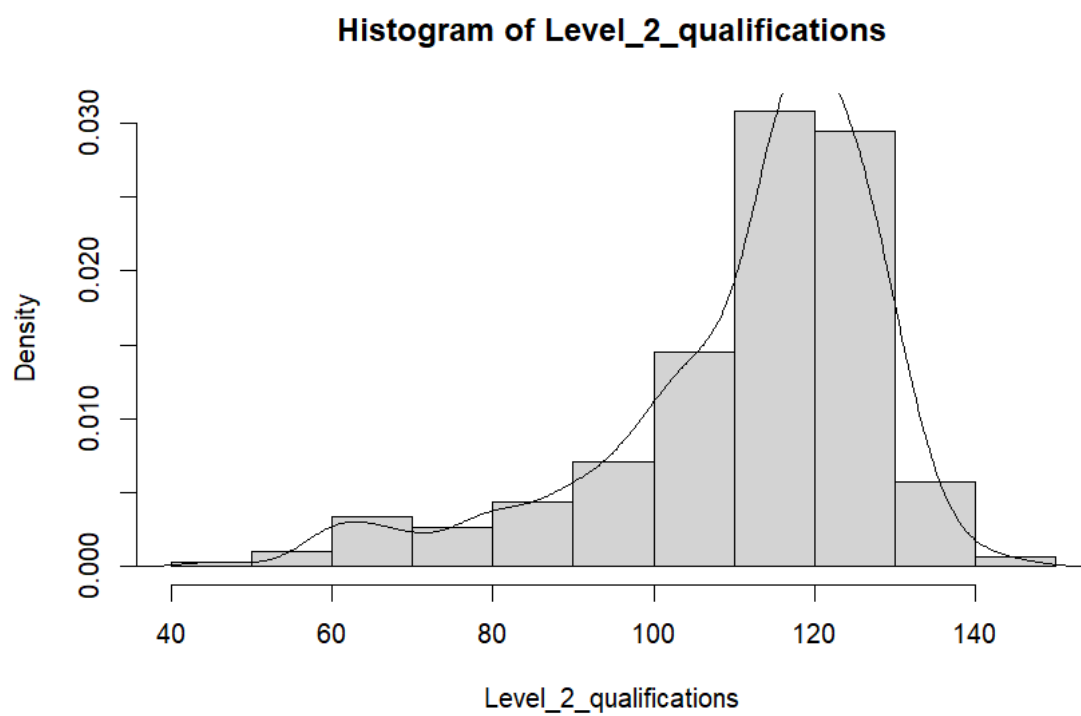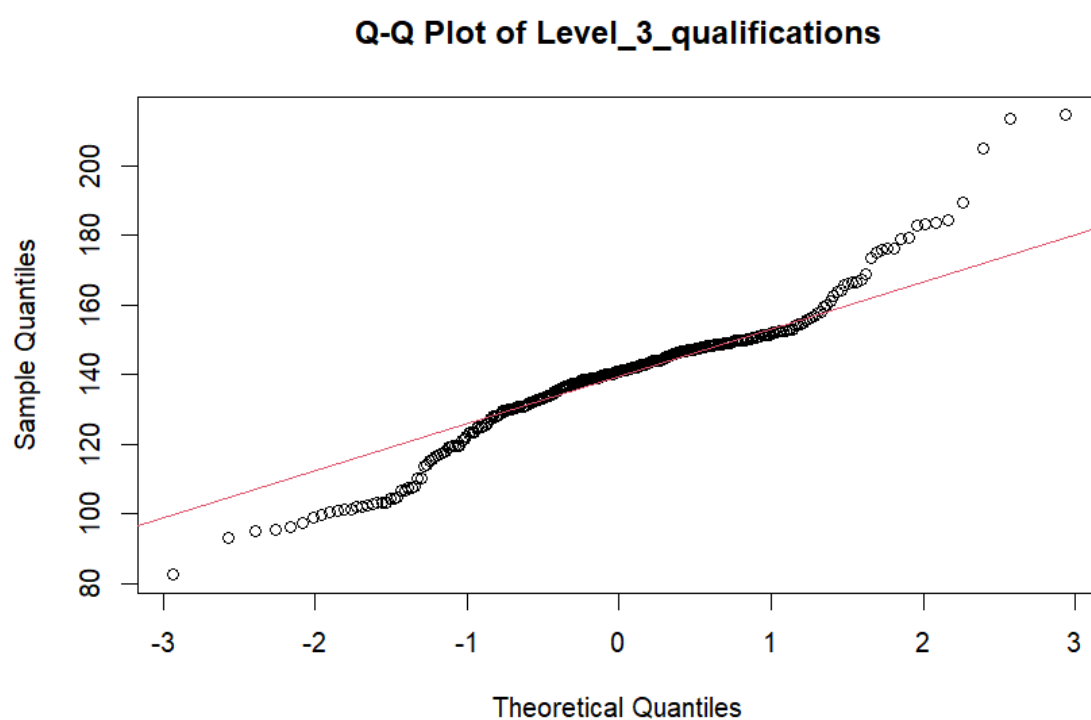

Figure 75: Q-Q plot of level 3
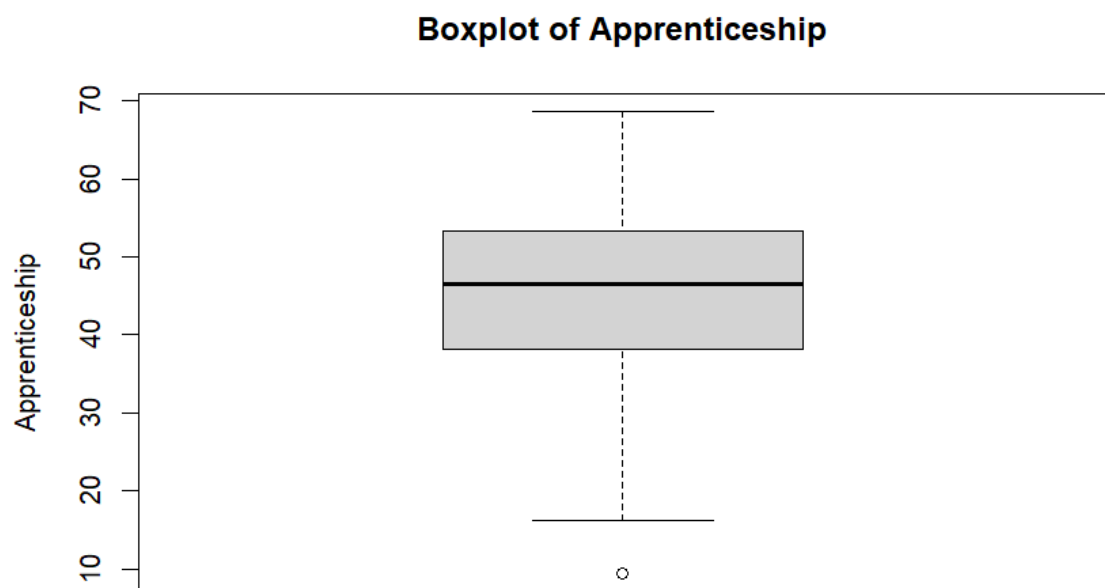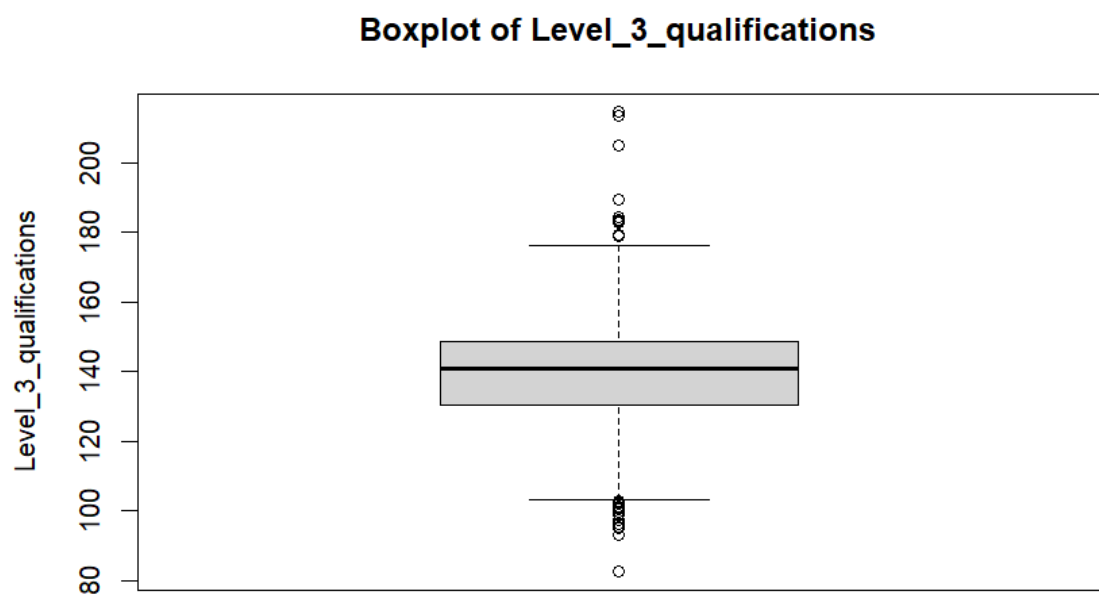
Boxplot of Level_4_qualifications_or_above
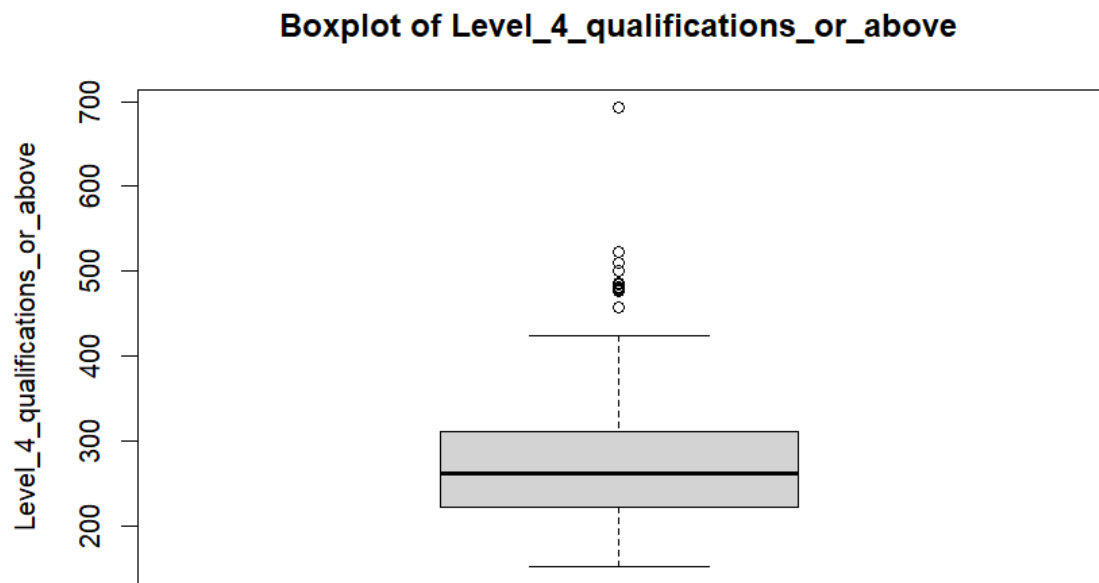
Figure 76: Boxplot of level 4 or above

Histogram of Level_4_qualifications_or_above
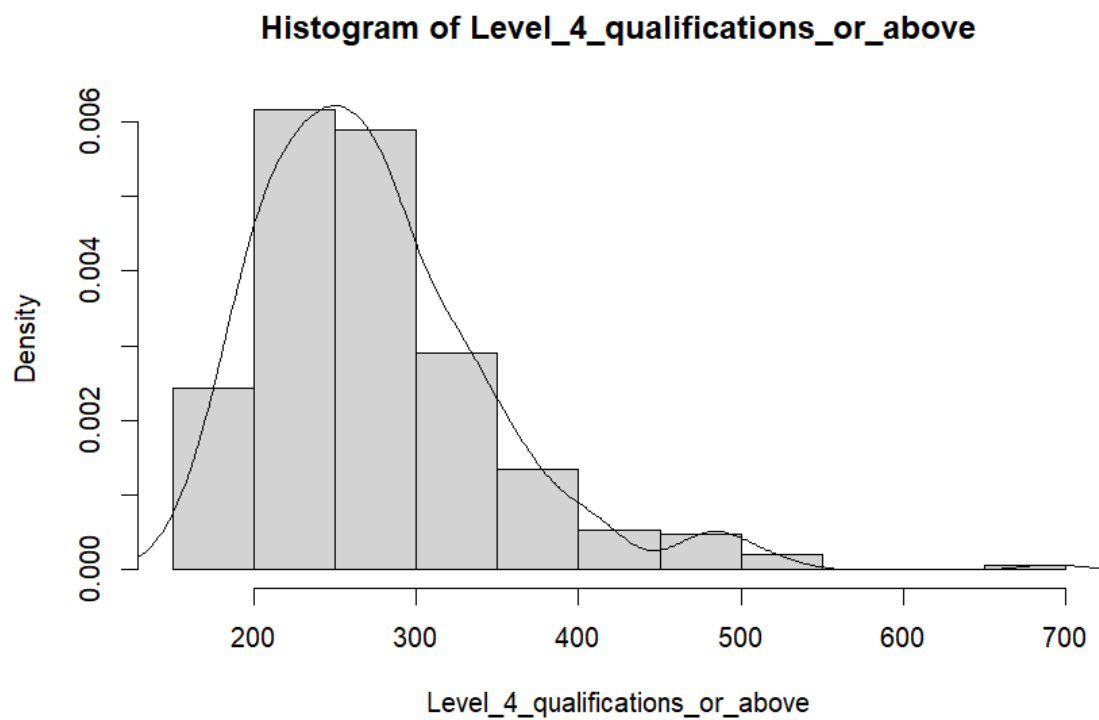
Figure 77: Histogram of level 4 or above

Q-Q Plot of Level_4_qualifications_or_above
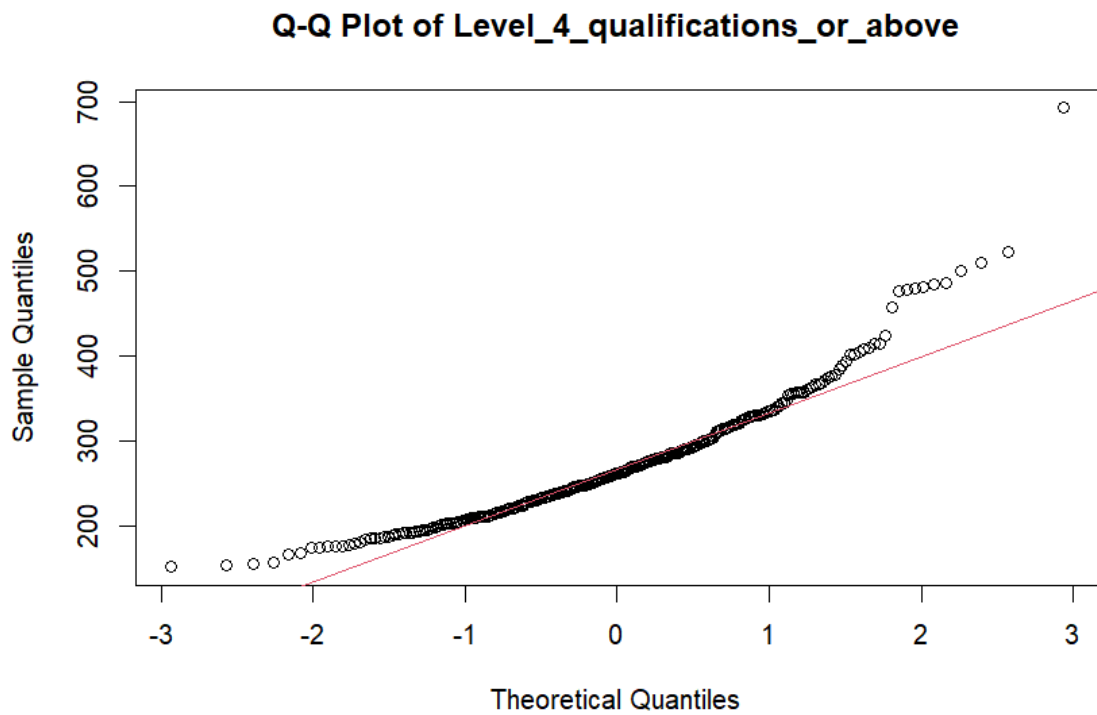
Figure 78: Q-Q plot of level 4 or above

Boxplot of Other_qualifications
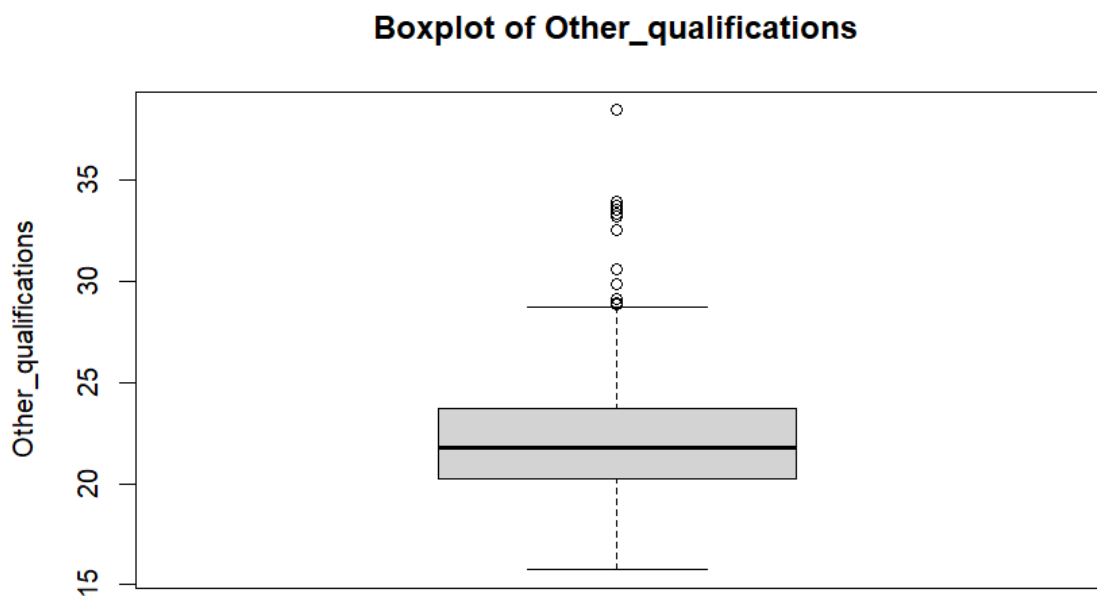
Figure 79: Boxplot of other qualifications
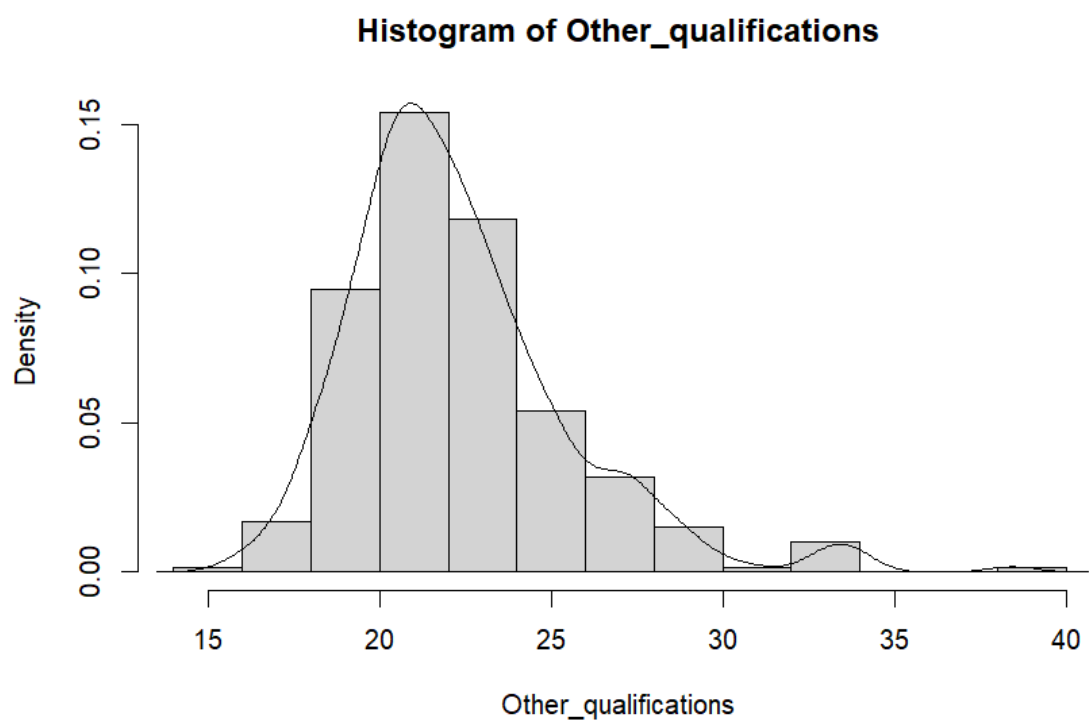
Figure 80: Histogram of other qualifications
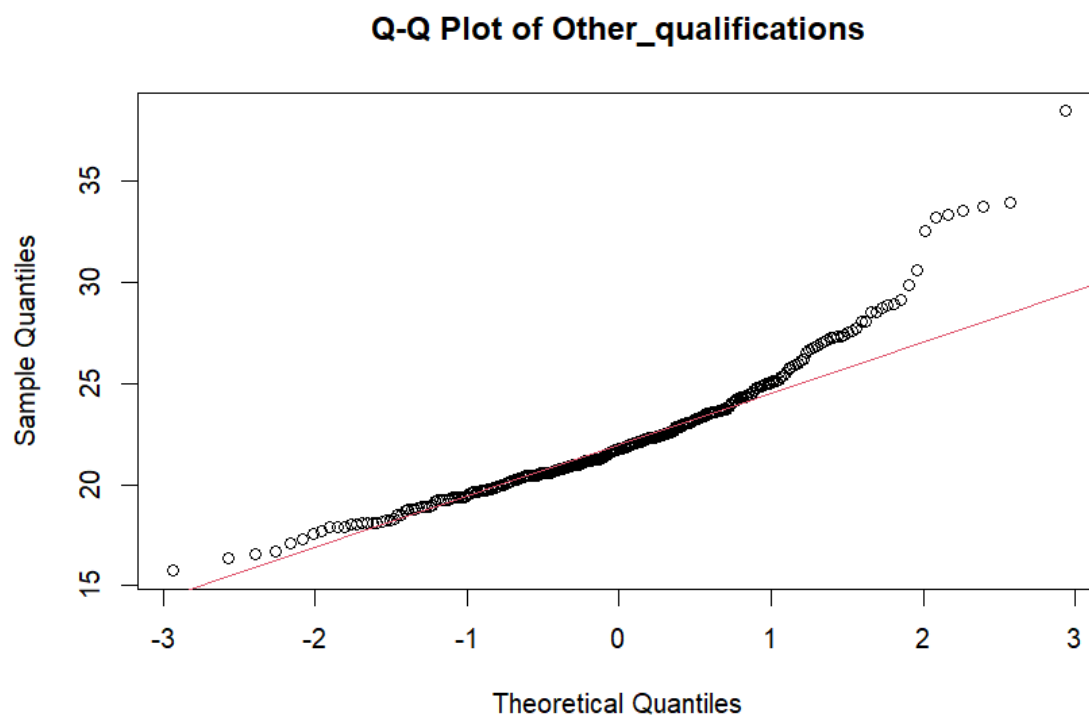
Figure 81: Q-Q plot of other qualifications

Row	actual	predicted	diff	abs_diff	abs_diff_sq
1	1.35	2.4249	−1.0749	1.0749	1.1555
2	2.42	2.2806	0.1394	0.1394	0.0194
3	1.94	2.1007	−0.1607	0.1607	0.0258
4	2.99	2.4813	0.5087	0.5087	0.2587
5	2.61	2.1417	0.4683	0.4683	0.2193
…	…	…	…	…	…
296	2.32	1.7368	0.5832	0.5832	0.3401

Table 5: Predicted and observed COVID-19 deaths per 1,000 residents for all 296 English local authorities

Bibliography

Department of Health and Social Care and UK Health Security Agency, 2022. *COVID-19 confirmed deaths in England to 31 December 2022.* [Online] Available at: https://www.gov.uk/government/publications/covid-19-reported-sars-cov-2-deaths-in-england/covid-19-confirmed-deaths-in-england-to-31-december-2022-report [Accessed 03 December 2025].

Institute for Fiscal Studies, 2020. *Sector shutdowns during the coronavirus crisis: which workers are most exposed.* [Online] Available at: https://ifs.org.uk/publications/sector-shutdowns-during-coronavirus-crisis-which-workers-are-most-exposed [Accessed 03 December 2025].

Institute of Health Equity, 2020. *Build Back Fairer: The COVID-19 Marmot Review.* [Online] Available at: https://www.health.org.uk/reports-and-analysis/reports/build-back-fairer-the-covid-19-marmot-review [Accessed 03 December 2025].

Institute of Health Equity, 2020. *Build Back Fairer: The COVID-19 Marmot Review.* [Online] Available at: https://www.health.org.uk/reports-and-analysis/reports/build-back-fairer-the-covid-19-marmot-review [Accessed 03 December 2025].

Intensive Care National Audit and Research Centre, 2022. *COVID-19 critical care reports.* [Online] Available at: https://www.icnarc.org/audit/cmp/reporting/ [Accessed 03 December 2025].

Office for National Statistics, 2020. *Deaths involving COVID-19 by local area and deprivation.* [Online] Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/datasets/deathsinvolvingcovid19bylocalareasanddeprivation [Accessed 03 December 2025].

Office for National Statistics, 2020. *Health and social care: causes of death statistics.* [Online] Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/causesofdeath [Accessed 02 December 2025].

Office for National Statistics, 2020. *Which occupations have the highest potential exposure to the coronavirus (COVID-19)?.* [Online] Available at: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/articles/whichoccupationshavethehighestpotentialexposuretothecoronaviruscovid19/20

20-05-11
[Accessed 03 December 2025].

Office for National Statistics, 2021. *Coronavirus related deaths by occupation, England and Wales.* [Online]
Available at:
https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/causesofdeath/bulletins/coronaviruscovid19relateddeathsbyoccupationenglandandwales/deathsregisteredbetween9marchand28december2020
[Accessed 03 December 2025].

Office for National Statistics, 2022. *Coronavirus (COVID-19) related deaths by occupation, England and Wales.* [Online]
Available at:
https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/causesofdeath/bulletins/coronaviruscovid19relateddeathsbyoccupationenglandandwales/latest
[Accessed 02 December 2025].

Office for National Statistics, 2022. *Deaths due to COVID-19, registered in England and Wales: 2021.* [Online]
Available at:
https://www.ons.gov.uk/releases/deathsduetocovid19registeredinenglandandwales2021
[Accessed 03 December 2025].

Office for National Statstics, 2020. *Updated estimates of coronavirus (COVID-19) related deaths by disability status, England: 24 January to 20 November 2020.* [Online]
Available at:
https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/articles/coronaviruscovid19relateddeathsbydisabilitystatusenglandandwales/24januaryto20november2020
[Accessed 03 December 2025].

Public Health England, 2020. *COVID-19: review of disparities in risks and outcomes.* [Online]
Available at: https://www.gov.uk/government/publications/covid-19-review-of-disparities-in-risks-and-outcomes?
[Accessed 02 December 2025].

Public Health England, 2020. *COVID-19: review of disparities in risks and outcomes.* [Online]
Available at: https://www.gov.uk/government/publications/covid-19-review-of-disparities-in-risks-and-outcomes
[Accessed 03 December 2025].

Public Health England, 2020. *COVID-19: review of disparities in risks and outcomes.* [Online]
Available at: https://www.gov.uk/government/publications/covid-19-review-of-disparities-in-risks-and-outcomes
[Accessed 03 December 2025].

Sun, Y., 2021. *Spatial inequalities of COVID-19 mortality rate in relation to socioeconomic and environmental factors across England.* [Online]
Available at: https://www.sciencedirect.com/science/article/pii/S0048969720371266
[Accessed 03 December 2025].

The Health Foundation, 2020. *Will COVID-19 be a watershed moment for health inequalities.* [Online]
Available at: https://www.health.org.uk/reports-and-analysis/briefings/will-covid-19-be-a-watershed-moment-for-health-inequalities
[Accessed 03 December 2025].

The Health Foundation, 2021. *The COVID-19 impact inquiry report.* [Online]
Available at: https://www.health.org.uk/reports-and-analysis/reports/unequal-pandemic-fairer-recovery
[Accessed 03 December 2025].

The Health Foundation, 2021. *Unequal pandemic, fairer recovery.* [Online]
Available at: https://www.health.org.uk/reports-and-analysis/reports/unequal-pandemic-fairer-recovery
[Accessed 03 December 2025].

The King's Fund, 2020. *Deaths from COVID-19 (coronavirus): how are they distributed across England?.* [Online]
Available at: https://www.kingsfund.org.uk/insight-and-analysis/long-reads/deaths-covid-19
[Accessed 02 December 2025].

World Health Organization, 2021. *Coronavirus disease (COVID-19): How is it transmitted?.* [Online]
Available at: https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-covid-19-how-is-it-transmitted
[Accessed 02 December 2025].