Health Disparities by Socio-Economic Class: Insights from UK Census Data: Visual Analytics Coursework

2542000

May 8, 2025

***abstract*–This report investigates the impact of socio-economic class on health outcomes and disability prevalence in England and Wales. Utilising data from the 2011 and 2021 censuses, the analysis uses dimensionality reduction techniques such as PCA and t-SNE to uncover clustering patterns. Stacked bar charts, choropleth maps, and credible interval plots encode class-specific health and disability distributions visually. The findings highlight significant associations between certain socio-economic classes and elevated rates of poor health and disability, emphasising the need for targeted health support in high-risk districts.**

1. Introduction

The idea for this project was born out of a personal observation during the COVID-19 pandemic, which exposed the long-standing inequalities in healthcare access. Across the world, it became clear that socio-economic status played a decisive role in determining who had access to critical healthcare resources, from hospital beds and oxygen cylinders to even basic testing. Communities struggling economically were affected the most, not only in terms of infection and mortality rates but also in their inability to recover socially and financially. What stood out most was how invisible structural inequalities suddenly became visible and urgent, and how access to health was never just about medical systems, but about where we live, what we earn, and the work we do.

Understanding the distribution of general health across England and Wales provides crucial context for evaluating the broader socio-economic landscape. Census data from both 2011 and 2021 reveal notable regional variations in general health distribution. Health is not distributed evenly across the population, and these differences are not random. They closely align with markers of deprivation, access to healthcare, education, and housing quality. The Marmot Review (2010) emphasised how people in the most deprived areas of the UK live shorter lives and spend more of those years in ill health compared to those in more affluent regions [1].

Disparities in health outcomes are also related to occupational classifications. Lower-paid, more physically demanding, and less secure jobs are more likely to expose individuals to occupational hazards and stress, contributing to cumulative health disadvantage over time (Black, 2008).[2] One of the most significant and underexplored contributors to these disparities is occupation, more precisely, the socioeconomic class associated with the type of occupation. This project investigates the relationship between occupational class and health status across England and Wales using census data from 2011 and 2021. By visualising this data across regions and time, and focusing on categories such as managerial/professional roles, routine occupations, and long-term unemployment, the goal is to reveal how deeply health outcomes are intertwined with class. Through this, the study not only quantifies inequality but also aims to surface insights that are actionable for public health planning and policy.

Addressing this issue requires targeted interventions from policymakers, local government, healthcare providers, organisations, and employment agencies. By focusing on districts with poor health distribution, targeted health programs can be developed to mitigate health risks associated with economic instability. Employment support initiatives can provide training and job opportunities aimed at improving socio-economic status, which in turn can reduce health disparities. Public health officials can also implement health interventions to address specific vulnerabilities in districts identified as high-risk and strategically allocate resources to the most affected populations.

2. What: Data preparation and Abstraction

The dataset used in this project is taken from the UK national censuses of 2011 and 2021 for England and Wales, administered by the Office for National Statistics (ONS)[3]. This study uses data from both years to investigate the intersection of occupational classification and health inequality, focusing on general health status and disability among various socio-economic groups. As defined in Munzner's framework, data abstraction involves mapping raw data to high-level data types such as items, attributes, and values, where items are the entities being studied (e.g., geographic regions), and attributes are variables describing those items. According to this model, each row in the table corresponds to a local authority (**items**), identified by a unique name and a government-issued area code. The columns are the **attributes** describing that authority, counts of people as **values** within different NS-Sec classes, classified by general health or disability status.

Table 1: LC3601EW – General Health by NS-SeC (2011) This table presents general health levels - Very good, Good, Fair, Bad, Very bad - by socioeconomic class, using the National Statistics Socioeconomic Classification (NS-SEC). These classes include "Higher managerial and professional", "Routine occupations", and "Never worked or long-term unemployed", among others.

Each data entry represents a quantitative value (a count) describing the number of individuals within a given class and health category in a particular local authority. From Munzner's perspective, this is quantitative tabular data: a matrix where each cell is a value defined by two attributes — socio-economic class and health category — for each local authority.

My first step was to extract this table from the Nomis platform, filtering for only English and Welsh authorities. This required excluding rows with Scottish or Northern Irish area codes, using the prefix convention (e.g., E for England, W for Wales). I retained only those columns describing NS-Sec health status distributions, removing redundant metadata and ensuring alignment of column headers with consistent variable names.

- **Items:**

  Lower tier local authorities represent unique geographic districts or authorities. These are nominal items and act as spatial containers for the socio-economic and health variables.

- **Attributes:**

  National Statistics Socio-Economic Classification (NS-Sec) defines social class computed based on occupation (Higher managerial, Routine occupations, Long-term Unemployed, etc). This is a categorical semantic dimension but carries an ordinal interpretation when class hierarchy is implied (e.g., L1–L3 is socially higher than L13–L14).

  General health categories (Very good, Good, Fair, Bad and Very Bad) represent an ordinal semantic scale, where the order reflects increasing severity of health situation.

- **Items:**

  Observation values represent aggregate population counts for each class–health pair per region. These are quantitative semantics, used for computing proportions, trends, and inequalities.

Table 2: LC3602EW – Long-Term Health Problem or Disability by NS-Sec (2011) This table focuses on functional health limitations, specifically those whose day-to-day activities are limited due to long-term disability. It follows the same NS-Sec breakdown, as this data is also segmented by local authority. Unlike general health, this data gives insights into chronic conditions.

- **Items:**

  Lower tier local authorities represent unique geographic districts or authorities. These are nominal items and act as spatial containers for the socio-economic and health variables.

- **Attributes:**

  National Statistics Socio-Economic Classification (NS-Sec) defines social class computed based on occupation (Higher managerial, Routine occupations, Long-term Unemployed, etc). This is a categorical semantic dimension but carries an ordinal interpretation when class hierarchy is implied (e.g., L1–L3 is socially higher than L13–L14).

  Disability - Equality Act disabled represents a categorical semantic attribute, as it includes categories such as limited a lot, limited a little, and not disabled, indicating the presence or absence of significant long-term impairment.

- **Items:**

  Observation values represent aggregate population counts for each class–health pair per region. These are quantitative semantics, used for computing proportions, trends, and inequalities.

Table 3: General Health by NS-Sec (2021) This dataset is similar to Table 1, but reflects post-pandemic census data. It captures updated general health assessments by occupational class and geography, using the same NS-Sec framework. Temporal comparison with 2011 allows the detection of changes in health inequality, especially following COVID-19 disruptions.

Preparing the data for projection and visualisation, I performed data cleaning using Python, focusing on structuring the dataset for consistency and comparability. The dataset was originally stored in multiple Excel sheets, with each sheet representing a single general health category. These were individually read and annotated with a corresponding ordinal code (1–5), and then concatenated into a single dataframe. The original LC3601EW dataset was structured in wide format, as is typical for many official statistical tables. This format is not well-suited for analytical tasks such as dimensionality reduction, clustering, or interactive visualisation. I reshaped this into long format using pandas.melt() function. In the long format, each row represents a single observation consisting of three core identifiers: the local authority code, the socio-economic class, and the general health category, along with the corresponding count of individuals as observations. This restructuring transforms the dataset from a two-dimensional matrix into a flexible, relational structure compatible with dimensionality reduction techniques and Tableau visualisations.

Next, I modified the labels to align occupational class names with the updated 2021 schema. For example, multiple sublevels (e.g., L1, L2, L3) were grouped under a single class name, and each NS-Sec category was assigned a unique numeric code. To ensure a uniform data matrix for all combinations of local authority and occupational class, I created rows with a general health code of -8 ("Does not apply") for any missing combinations. This preserves alignment across datasets and maintains compatibility with Dimensionality reduction techniques. Finally, I filtered all the rows with null and metadata, and standardised the column headers. With this, the dataset was ready for visual analysis and comparative study across 2011 and 2021.

To understand the underlying structure and spatial patterns in the health and disability data, I performed dimensionality reduction techniques, specifically Principal Component Analysis (PCA) and t-Distributed Stochastic Neighbour Embedding (t-SNE). These techniques require that data be formatted as a complete matrix of numerical features, where each row represents a unit of analysis (lower-tier local authority) and each column represents a distinct combination of health category and socio-economic class. To achieve this, I performed a pivot on the long-form census data such that each unique class–health or class–disability pairing became a separate feature column. The result was a wide-format matrix encoded multi-dimensional population distributions across NS-Sec groups and health outcomes. Then, I standardised these values to ensure that variables with larger magnitudes did not dominate those with smaller counts. PCA was applied to reduce the data to two principal components, allowing for exploratory visualisation of global variance structure. Separately, for another dashboard, t-SNE was used to model local similarities and cluster regions with similar class-health distributions. I then performed K-Means clustering on the t-SNE output to classify local authorities into health inequality profiles.

I saved the resulting 2D projections from both PCA and t-SNE as .csv files and used them for the Tableau visualisations. These helped to reveal regional patterns, clusters, and outlier authorities with unusual health or disability insights that would be less accessible through raw tabular comparisons alone.

To estimate the proportion of people with severe disabilities across socio-economic classes in each district, I used a Bayesian approach to generate posterior distributions and compute credible intervals. I focused on the category "Disabled under the Equality Act: Day-to-day activities limited a lot" and grouped the data by NS-Sec class within each local authority. For each district, I calculated the observed counts of individuals in each socio-economic class and applied Laplace smoothing to avoid zero probabilities. I then treated the smoothed class counts as parameters of a Dirichlet distribution, which served as a conjugate prior for the categorical data. By sampling 5,000 times from this Dirichlet posterior, I was able to estimate the posterior mean and the 95% credible interval (2.5th and 97.5th percentiles) for the class-wise disability proportions. This method allowed me to represent uncertainty in a principled way, rather than relying on point estimates alone. I saved the results, including the posterior means and bounds, as a CSV file for further visualisation and interpretation.

3. Why : Task Definition

This report is designed to support multiple analytical tasks related to understanding General health levels and Socio-economic classes and their interrelationships across different geographic authorities. According to Munzner's task taxonomy [4], These tasks can be categorised as follows:

Dashboard 1: General Health Distribution Across England and Wales (2011)
Action: Discover

- **Target: Spatial Distribution and Trends** This dashboard allows users to explore how general health categories are distributed geographically across local authorities. The aim is to discover regional health patterns and disparities using 2011 census data. This supports hypothesis generation about potential public health inequalities.

Action: Locate

- **Target: Extremes** This allows users to identify specific districts with the highest or lowest proportion of a selected health category. For example, the bar chart and slider tools highlight the top N districts with the best or worst reported health outcomes.

Action: Present

- **Target: Overview** This presents a summarised national view of health status using an interactive map, bar chart, and a donut chart. This helps users understand the overall health landscape across England and Wales in 2011.

Dashboard 2: Health Inequality and NS-Sec Class (2011)
Action: Discover

- **Target: Correlations and Patterns** This task explores whether certain socio-economic classes are associated with poorer health outcomes. The goal is to identify patterns in how NS-Sec class composition correlates with health across districts. The PCA scatterplot helps reveal structure in the data by projecting high-dimensional class-health distributions into a 2D space.

Action: Cluster

- **Target: Similar profiles of Local Authorities** By applying clustering to the PCA results, districts are grouped based on shared class-health profiles. This allows users to interpret health inequality not as isolated outliers but as part of broader socio-economic patterns (e.g, districts dominated by routine jobs show poor health).

Action: Compare

- **Target: Attributes Across Clusters** The stacked bar chart compares class distributions in the top N districts with the selected category of health. This lets users contrast the socio-economic makeup of poorly performing districts and relate it to the clusters in the PCA.

Dashboard 3: Change in Health and Class Distribution (2011–2021)
Action: Compare

- **Target: Temporal Changes** This task enables users to compare general health distributions and NS-Sec class compositions across two census years, 2011 and 2021. To study how the class proportions have shifted, and reflect changes in the spatial health distribution by categories. Users can directly compare the evolution of health outcomes and socio-economic classes over time.

Action: Discover

- **Target: Correlations Between Attributes** By viewing poor health percentages, alongside class changes, users can explore whether increases in specific occupational groups, like routine or semi-routine classes, correspond to worse health outcomes in the same districts. This supports the hypothesis around structural inequality or the impact of occupational shifts.

Action: Present

- **Target: Summarise** Provides a comprehensive overview of how both socio-economic structure and health indicators have evolved over the past decade. It helps stakeholders identify key patterns in population health and relate them to economic or occupational change.

Dashboard 4: Disability Distribution by NS-Sec Class
Action: Discover

- **Target: Correlation and Patterns** This task supports exploratory analysis of how disability prevalence aligns with socio-economic class. The user's goal is to identify whether certain occupational groups experience higher levels of disability, and whether this pattern is consistent across districts.

Action: Compare

- **Target: Proportional Differences and Statistical Uncertainty** Here,users can to compare the burden of disability between NS-Sec classes, assessing not only raw percentages but also the stability of these estimates. This supports comparison-based reasoning, which class is more affected and how confident can we be in this difference. The task assumes the user has multiple options and must make informed comparisons to conclude.

Action: Categorise

- **Target: Cluster groups** If users seek to group local authorities based on similarity in their disability-class profiles, this task aims to uncover regional similarities that may not be obvious from raw figures, allowing users to reason about groupings or anomalies in the dataset. It helps identify whether certain social or geographical contexts share systemic patterns of inequality.

Action: Present

- **Target: Significant Cases** The task also involves highlighting the most affected classes or districts, to support prioritisation or communication. This may involve selecting the top N areas with the greatest burden of disability, based on a user-defined filter, to contextualise the results in policy or planning decisions.

4. How: Visualisation Justification

Dashboard 1: General Health Distribution Across England and Wales (2011) This dashboard intends to provide an intuitive understanding of the distribution of general health across England and Wales in 2011. By providing insights into health distributions, policymakers can better understand the regional disparities in health outcomes and develop targeted health interventions. The 'Top N' filter allows users to quickly identify districts with the highest or lowest percentages of a specific health category, further enabling targeted policy recommendations. **Justification of Plots and Visual Encodings:** The dashboard includes a choropleth map, bar chart, donut chart, and a time series trend graph. According to Munzner's taxonomy, these visualisations align with distinct analysis tasks:

- **Choropleth Map:** The Choropleth map conveys spatial distributions of health categories across districts. Spatial encoding utilises human perceptual capabilities to quickly recognise geographic clusters or outliers. The map serves as a focal point for observing health disparities, enabling the user to interactively explore regional health trends.

- **Bar Chart:** The bar chart displays the top N districts based on the selected health category, providing a clear ranking and allowing for rapid comparison. The length encoding effectively communicates proportional differences across districts, facilitating comparative analysis of health percentages.

- **Donut Chart :** The donut chart provides a percentage breakdown of health categories for a selected district, enabling users to identify the dominant health category and observe how each category contributes to the overall health composition. Although they are less precise than bar charts, and provide quick visual summary of the overall proportions.

- **Trend Graph:** The time series trend graph depicts health trends across multiple health categories from 2011 to 2021, allowing users to identify temporal patterns and potential emerging trends in health distributions. The use of a continuous line graph capitalises on the human's ability to detect slopes and trajectories, making it easier to detect fluctuations across health categories.

The use of consistent colour encoding across all plots ensures visual consistency, aiding in cognitive recognition and retention of health categories.

Functionality: Users can click on any district on the map or type a district name in the search filter to view the health distribution for that specific area. To display the districts with the highest percentages in the selected health category, we can use the 'Top N' filter that dynamically adjusts the bar chart to focus on specific health outcomes and identify critical areas for intervention.

Inference and Insights: Initial observations from the dashboard reveal that districts with higher percentages of 'Good health' and 'Very good health' are predominantly located in the southern and eastern areas, whereas districts with higher percentages of 'Bad health' are concentrated in specific southern areas. The trend graph further indicates a steady increase in 'Very good health' percentages over the years, suggesting potential improvements in overall health outcomes across the dataset timeframe. This comprehensive visual analysis underscores the need for targeted health interventions in districts exhibiting persistently low health percentages, as well as potential policy focus on sustaining health improvements in high-performing districts.

Dashboard 2: Health Disparities among different NS-SEC classes in Districts(2011) The second dashboard provides a comprehensive understanding of the health distribution across various socio-economic classes (NS-Sec) for the selected health categories. It aims to identify the socio-economic classes that bear a disproportionate health burden, particularly focusing on the prevalence of poor health among routine and semi-routine occupations. This is crucial for policymakers who need to identify vulnerable districts and target specific groups that exhibit higher levels of poor health, thereby enabling targeted policies and resource allocation.

- **Stacked bar chart:** This represents the distribution of health categories across various socio-economic classes for the selected top N districts. Each segment within the bar represents a different socio-economic class, distinguished by colour gradient. This takes advantage of the human ability to perceive and differentiate colours, facilitating quick visual recognition and comparison. Using colour and length serves as an effective visual encoding, utilising the man's ability to process and interpret these visual cues. It also enables viewers to quickly perceive and distinguish different values or categories without conscious effort. By stacking the segments, the chart not only highlights the proportion of each class within the selected health category but also provides a cumulative view, making it easy to compare between districts and health categories simultaneously. This holistic approach supports the identification of patterns and anomalies.
- **Scatter plot:**The scatter plot visualises the PCA projections derived from health distribution and socio-economic class data, reducing dimensional complexity while retaining the dominant data patterns. The PCA projections were created using a matrix that combined six health categories and ten NS-Sec socio-economic classes, effectively capturing the variance between health distributions across different socio-economic groups. By plotting the first two principal components, the scatter plot leverages spatial positioning to identify clusters, correlations, and outliers intuitively. Additionally, Tableau's built-in clustering algorithm was employed to highlight distinct groups of districts with similar health-class profiles, aiding in the identification of vulnerable regions. The use of colour to distinguish clusters and spatial separation to highlight outliers facilitates rapid recognition of patterns, reinforces the scatter plot as an exploratory and diagnostic visualisation tool[5]. Scatterplots are particularly effective for revealing clusters and trends as they leverage human perception to detect spatial patterns and variations quickly, enhancing the overall interpretability of multivariate data projections. [6]

  The user can adjust the Top N filter and Health Category to view health-class distributions, and click a cluster point in the scatter plot to see the class distribution for that specific district reflected in the stacked bar chart.

Inference and Insights: From the stacked bar chart, it is evident that routine and semi-routine occupations consistently show higher proportions of poor health, particularly in districts like Birmingham, Bradford, County Durham, Cornwall, Leeds, etc. Additionally, the PCA projection further validates these observations, with Cluster 1 predominantly comprising districts with high levels of poor health concentrated in routine and semi-routine occupations, while Cluster 2 clusters economically active and relatively healthier districts and Cluster 3 with mixed social classes performing in moderate.

Dashboard 3: Assessing Changes in Health Distribution and Class Composition (2011 vs 2021) This dashboard provides policymakers and analysts with a comparative visualisation of the changes in health distribution and socio-economic class composition between 2011 and 2021. By enabling users to compare health distribution across districts and visualise the shifting class compositions, the dashboard helps in identifying how socio-economic restructuring may influence health outcomes.

- **Choropleth map:** This map uses spatial encoding to display the percentage of each health category across districts for both 2011 and 2021, and is crucial for identifying spatial disparities in health distribution, allowing users to instantly observe regions with high concentrations of poor health. The use of a gradient colour scheme aids in visually distinguishing districts with higher vs. lower health percentages, utilising the human perceptual system's ability to differentiate shades and detect areas of concern. The Hover-over interaction provides percentage comparisons for 2011 and 2021, allowing the user to directly assess changes over time in health distribution at the district level, supporting comparative analysis and granular inspection.

- **Diverging Bar Chart:** The diverging bar chart is used to compare class composition percentages between 2011 and 2021 for the selected district. By arranging bars in a diverging layout, users can quickly perceive increases and decreases in class proportions. The use of contrasting colours for 2011 and 2021 enhances visual differentiation, aligning with Munzner's task taxonomy under the "Compare" and "Present" categories, as it effectively illustrates temporal changes in socio-economic composition [4]. The diverging layout enables easy detection of outliers, such as significant reductions in routine and semi-routine occupations that correlate with declining poor health percentages in certain districts.

- **Horizontal Bar Chart:** The bar chart at the bottom contextualises the relationship between class composition and health distribution. By isolating the percentage of poor health within each class, users can directly assess whether changes in socio-economic class proportions align with changes in health outcomes. The use of colour-coded bars based on class categories provides intuitive comparison, facilitating rapid recognition of high-risk classes. The quantitative length of the bar allows easy comparison between classes.

The dashboard interactively allows users to hover over the choropleth map to access specific health distribution data for both 2011 and 2021, with direct comparison of health percentages over time. Class distribution for the selected local authority is displayed in the diverging bar chart. The Bar chart can be used to compare class compositions and assess their potential impact on poor health percentages in the selected district. This interactive design enables multi-faceted analysis, encouraging users to investigate potential causal links between class structure and health outcomes over time.

Inference and Insights: The analysis reveals that shifts in socio-economic class composition are associated with changes in poor health distribution, indicating a potential link between class structure and health outcomes. For instance, in County Durham, the percentage of bad health decreased from 7.1% in 2011 to 5.78% in 2021, along with a decline in the share of routine occupations from 15.5% to 11.2%. This pattern also repeats in districts like Birmingham, Powys, Salford, Warwick, and many more, indicating that areas with higher percentages of Routine and Semi-Routine occupants see a higher percentage of poor health amongst their population.

Dashboard 4: Disability Distribution by NS-Sec Class : The objective of this dashboard is to assess the distribution of disability across various socio-economic classes in England and Wales, examining whether socio-economic class influences disability prevalence. This aligns with the broader objective of identifying socio-economic disparities in health outcomes, providing policymakers with insights into which socio-economic classes and districts may require targeted interventions to address disability-related inequalities.

- **Stacked Bar chart:** Similar to the previous dashboard, this chart effectively illustrates the distribution of disability across socio-economic classes by visually encoding proportions using bar length. The use of distinct colours for each class enables rapid comparison and pattern recognition, facilitating the identification of classes with higher disability prevalence. This visualisation aligns with Munzner's "Compare" and "Summarise" categories, as it allows users to contrast disability distribution across multiple classes within districts. The use of bar length as a quantitative encoding effectively emphasises the proportions, making differences in disability distribution more visually distinct.

- **t-SNE Scatter plot :** This scatter plot reduces high-dimensional data into a 2D plane, with the x-axis representing the first t-SNE component that captures the primary variance in disability prevalence and the y-axis representing the second t-SNE component, emphasising district differentiation based on socio-economic class distribution. This projection helps in identifying natural clusters of districts with similar disability profiles, as spatial positioning encodes the degree of similarity or divergence. The integration of K-Means clustering further reinforces this visual encoding, using distinct colours to differentiate groups of districts with varying levels of disability impact. This visual strategy aligns with Munzner's "Discover" and "Present" tasks, enabling the identification of distinct clusters, potential outliers.

- **Gantt bar chart** This Gantt bar chart represents Bayesian credible intervals for each socio-economic class. The bar length indicates the estimated range of disability prevalence, with the start and end points indicating the lower and upper bounds of the credible interval. The mean value is represented as the central bar, visually conveying both the average prevalence and its variability. This approach adheres to Munzner's "Annotate" task by explicitly showing the uncertainty associated with each class. The use of distinct colours for each class assists in quick identification and comparison, reinforcing perceptual grouping and making it easier for users to detect classes with higher uncertainty in disability rates.

Inference and Insights: This dashboard indicates that districts with higher disability prevalence tend to be concentrated in socio-economic classes such as routine and semi-routine occupations, similar to the pattern observed in poor general health distribution. This alignment is further substantiated by the Bayesian credible intervals, which highlight that the routine and semi-routine occupational classes exhibit higher probabilities of disability prevalence. For instance, the routine occupation class has a mean disability percentage of 63.36%, while semi-routine occupations show 62.25%, both significantly higher than other classes.

5. Evaluation

The visualisations in this project were evaluated using Munzner's four levels of validation and were conducted through peer discussion, and feedback was integrated to refine the dashboards.

Domain Situation: The visualisations effectively addressed the socio-economic health disparities in England and Wales, specifically focusing on the impact of socio-economic classes on health and disability status. The dashboards were perceived as highly relevant for identifying target areas where health interventions could be more impactful, such as routine and semi-routine occupations, which consistently exhibited the highest rates of poor health and disability across multiple districts.

Abstraction The chosen data attributes, such as health status percentages, class distributions, and health category prevalence, were deemed appropriate for analysing disparities across socio-economic groups. The use of data projection techniques like PCA and t-SNE further enabled the identification of clusters and outliers, allowing for the detection of districts with similar socio-economic health profiles.

However, the group provided several suggestions for improvement. The use of colour-coded stacked bar charts effectively represented the class distribution by health category, enabling intuitive comparison across socio-economic classes. But, the labels for health percentages were noted to be slightly cluttered, particularly when displaying multiple districts simultaneously. To improve this, I adjusted the font size to minimise text clutter and enhance readability. Secondly, while the clusters were distinguishable, feedback suggested that the colour scheme in the PCA plot was not sufficiently distinct, making it challenging to differentiate between clusters. Hence, I enhanced the colour contrast to differentiate the clusters and added instructions to guide users in interacting with cluster points. Overall, this feedback improved the clarity, interactivity, and interpretability of the visualisations, thus enhancing their effectiveness.

6. Conclusion

This project provided a comprehensive analysis of the socio-economic disparities in health and disability distribution across districts in England and Wales, focusing particularly on how socio-economic classes influence health outcomes and disability prevalence. The findings across the dashboards show a clear pattern: lower socio-economic classes, such as routine and semi-routine occupations, consistently exhibit higher rates of poor health and disability. This trend was first observed in the initial dashboard, where the distribution of health across socio-economic classes highlighted that districts like Birmingham, Bradford, Cornwall, County Durham, and many more had a disproportionately high percentage of poor health concentrated in routine and semi-routine occupations. This insight suggests that socio-economic class not only influences general health but may also act as a determinant of health inequalities within districts. Building on this, the third dashboard compares health distribution in 2011 and 2021, allowing for an in-depth analysis of how class composition shifts over time may correlate with changes in health outcomes. The diverging bar chart revealed that in districts like County Durham and Powys, a reduction in semi-routine occupations from 2011 to 2021 coincided with a slight decrease in poor health prevalence, indicating a potential link between class composition and health disparities. This pattern was consistently observed in other districts such as Warwick and Salford, reinforcing the idea that socio-economic class transitions can significantly impact health outcomes over time. The final dashboard extended the analysis to disability prevalence, using similar visualisation techniques to draw comparisons with general health findings. Here, the analysis revealed that the same socio-economic classes—routine and semi-routine occupations—also exhibited the highest disability rates. The Bayesian credible intervals further validated these findings, showing higher uncertainty in disability prevalence within these classes, emphasising that districts with higher percentages of routine and semi-routine workers are more vulnerable to disability. This alignment between general health and disability underscores the broader socio-economic challenge of health inequality, driven by occupational disparities.

Overall, the findings suggest that policymakers should target interventions towards routine and semi-routine occupations to address health and disability disparities effectively. By identifying districts with high concentrations of vulnerable socio-economic classes, targeted health programs can be implemented to mitigate the socio-economic determinants of health and disability. Although the analysis underscores the impact of socio-economic class on health disparities, other influential factors such as housing, economic activity, and access to healthcare must also be considered, as they are intricately linked to socio-economic status. For policymakers and health analysts, focusing on these occupations to provide essential resources, targeted health support, and economic opportunities could potentially alleviate health disparities and foster more equitable health outcomes across England and Wales.

This project has significantly deepened my understanding of the power of information visualisation in conveying complex socio-economic narratives effectively. This systematic approach ensured that each visualisation was not only aesthetically coherent but also analytically informative. I learned the importance of choosing visual encodings that align with the intended task, such as using bar charts to encode quantitative differences, choropleth maps to convey spatial distributions, and scatter plots to detect clusters and patterns. The iterative design process was important in refining the visualisations, allowing me to adjust colour schemes, visual encodings, and layout to better communicate the data's narrative. This analysis highlighted the importance of user-centred design, as feedback and peer review provided valuable insights into how visualisation choices could either clarify or obscure the underlying data patterns. The data projection techniques, such as PCA and t-SNE, helped immensely in simplifying complex, multidimensional data into two-dimensional representations, making it easier to identify patterns and relationships across different variables.

Ultimately, this coursework taught me how effective visualisation is more than just data representation, it is about crafting a compelling narrative that guides the viewer towards actionable insights. By aligning data selection, visual encodings, and interaction design, I was able to present a cohesive story of socio-economic health disparities that is both informative and actionable for policymakers.

7. References

[1] Michael Marmot, Peter Goldblatt, Jessica Allen , et al. (2010). Fair Society, Healthy Lives: The Marmot Review. UCL Institute of Health Equity.

[2]Black, C. (2008). Working for a Healthier Tomorrow: Dame Carol Black's Review of the health of Britain's working age population. Department for Work and Pensions.

[3] Office for National Statistics (2022). Socioeconomic inequalities in avoidable mortality: England analysis

[4] T. Munzner, Visualisation analysis and design. CRC Press, 2014

[5] Zhang Z, Cui X, Jeske DR, Li X, Braun J, Borneman J. Clustering Scatter Plots Using Data Depth Measures. J Biom Biostat. 2011;Suppl 5:001. doi: 10.4172/2155-6180.S5-001. PMID: 24883225; PMCID: PMC4038101.

[6] Ware, C. (2019). Information Visualisation: Perception for Design. Morgan Kaufmann.