Abstract

For understanding how population aging detailly affects the labour market, part of the 2011 UK census data will be used for the visual analysis. The attached Tableau dashboard firstly provides an initial exploration of regionally ageing and regionally economic structure based on two ageing indicators. Then, the visualisation analysis of some labour market indicators (i.e., labour participation rate, average working hours, unpaid care hours and the net migration based on economic activity groups) is shown to evaluate the specific effects of population ageing on labour markets. In Part 3, the internal relationship between population ageing and labour market indicators will be further explored. Finally, two kinds of data projection methods (UMAP and t-SNE) analyse the similarity of local authority districts on working hours, unpaid care hours, economic activity groups and net migration situations. Through such, city administrators can find similar areas when trying to implement some measures to change the labour market so that they can learn from successful cases in similar regions.

This report records each specific step when implementing the visual analysis of the UK labour market overview in the ageing society. It contains the introduction of the population ageing and its effects, the manipulation of data preparation, the visual task summary based on Munzner's task taxonomy, visualisation justification, and the insights got from this visualisation exploration.

Contents

1. Introduction

Population ageing is a general trend that most countries are experiencing or will experience in the 21st century, which refers to the fact that the national age distribution skews towards older age due to the reduction of population fertility and the extension of life expectancy [1]. Under the combined effect of many factors, such as the improvement of public health and social ideological development, this trend is constantly intensifying. It is a sign of improving people's average living standard and public medical service facilities. Still, it also becomes one of the reasons to weak the national economy and productivity. For example, the population ageing led to changes in the proportion of social consumption and savings, a significant increase in non-productive consumption in national income such as pension, nursing and medical treatment, and a decrease in the proportion of public development investment, which may affect the scale of investment and the sustainability of economic development [2].

Apart from these, there are also many aspects that deserve to explore about the impact of population ageing on the labour market, for instance, whether the increase of the elderly population will affect the willingness of the working-age population to work, whether the migration situation has some relationship with the ageing of areas [3]. The following visual analysis mainly assesses three aspects of the impact of population ageing on the labour market: labour participation rate, average working hours and migration based on economic activity, so as to explain a part of the influence

2. Data Preparation and Abstraction

2.1 Prior Knowledge

Before describing any data manipulation, some indicators need to be explained for understanding this visualization exploration better.

Traditionally, the proportion of people aged 65 and over (or 60 and over) in the total population is usually taken as the general ageing indicator. But in order to measure ageing more exactly, this visualization exploration will apply two dependency ratios as the ageing indicators.

(1) Old-Age Dependency Ratio (OADR): $\frac{\textit{The number of people aged } 65 \textit{ and over}}{\textit{The number of people aged } 16 \textit{ to } 64}$

This formula represents how many the number of elderly people (in pensionable age, 65) need to be dependent per working-age person on average. This is a very popular and simple ageing indicator [4].

(2) Real Elderly Dependency Ratio (REDR): $\dfrac{The\ number\ of\ people\ (Age \geq (Life\ expectancy - 15))}{The\ number\ of\ poeple\ who\ is\ in\ employment}$

Compared with OADR, this indicator redefines "old", the old age becomes a dynamic number depending on what the life expectancy is in the given area. In this report, the old people are a population aged equal to or over life expectancy in the UK in 2011 minus 15. Besides, OADR is based on the assumption that all working age people are employed, but this is not rigorous. So, the denominator of REDR is the number of people who is in employment [4].

When the ageing indicators of an area exceed 0.3, it could be defined as the ageing area. Perhaps the difference between these two indicators will form a strong contrast in the degree of regional ageing.

(3) Labour Participation Rate (LPR): $\dfrac{The\ number\ of\ people\ who\ are\ economically\ active}{The\ total\ number\ of\ people\ in\ the\ given\ age\ range}$

LPR rate is the ratio of the economically active population (including employed and unemployed) to the total number of people in the investigated age range, and is an indicator used to measure people's participation in economic activities [5].

2.2 Table Description and Data Processing

The following comes into the phase of data preparation. Firstly, five highly relevant tables are found with the "2011CensusIndexofTablesandTopics_v11_4_2". The map granularity of these five tables is all local authority districts so that the connections between tables can be built up. Some brief introduction to these tables will be given firstly.

1. "**by economic activity by age by sex.csv**": The table classifies the number of usual residents in each area by economic activity, gender and age, with the aim of understanding the overall situation of the labour market in each district. Economic activity contains the following hierarchical structure.
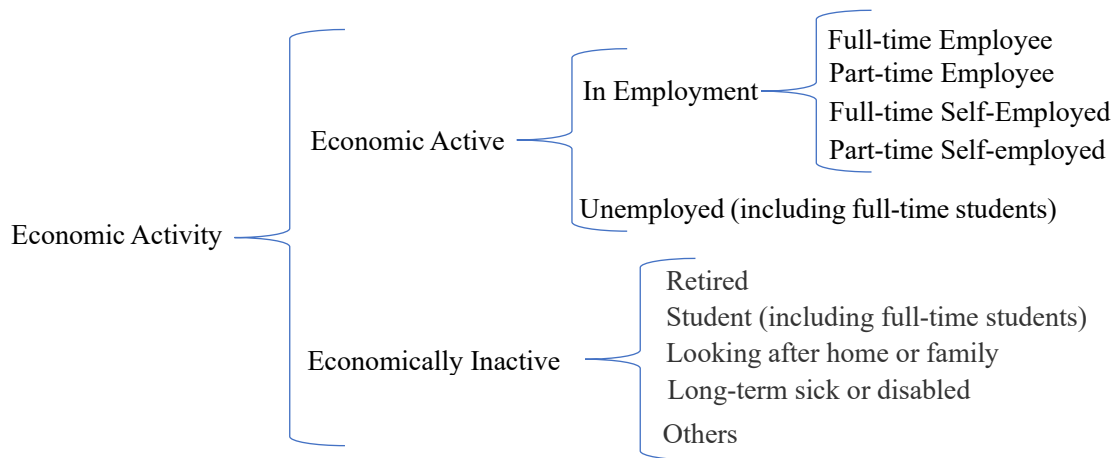
2. "**by sex by age.csv**": It records the number of usual residents with different age and different genders in each area. It is used to calculate aging indicators

3. **"by hours worked by age by sex by full or part time job type.csv"**: This table provides information that classifies usual residents aged 16 to 74 in employment by the number of hours worked for analysing the average working hours per week of working people in each region.

4. **"by provision of unpaid care by general health by sex by age.csv"**: This table records the number of usual residents with different age, different genders, different unpaid care hours and different healthy situation in each area. It is used for exploring whether the rise in average unpaid care hours due to aging is responsible for the decline in the average working hours, especially for those working-age populations.

5. **"by economic activity by migration.csv"**: The table contains specific economic activity-based migration for each region, which is used for analysing whether net migration situation would also be related to regional aging.

But the format of these tables is not suitable for importing to Tableau. Each column name of the tables contains a lot of useful information, like age, gender and economic activity, causing the table is so "wide". If putting all the raw data into Tableau and using Tableau to obtain all calculation fields, the data request time will be tremendously huge, so it is necessary to use Python for some data processing.

Cleaning Data (Data Abstraction): reduce the number of columns, i.e., divide the useful category information in the column name into separate attributes. The following picture shows the comparison of the table before and after the specific manipulation.

Figure 1: "by economic activity by age by sex.csv" Before and After Data Manipulation

Please see "cleaning data.ipynb" for specific data processing operations

Data Projection: For checking the similarity of each local authority district on migration situation, economic activity, unpaid care hours, working hours and overall situation, the data after dimensionality reduction also needs to be obtained in this phase. Because of the goal of visualizing, the number of components for three used data projection methods (t-SNE, UMAP and LDA) is all two. Thus, $3*2*5 =30$ new columns will be generated in the table named "data protection and calculation".

Some indicators: Meanwhile, for decreasing the data requesting time in Tableau, some indicators, i.e., OADR, REDR, labour participation rate and the average working hours for people aged 24 to 65 will be also calculated in Python.

Please see "data_protection.ipynb" for the manipulations about specific dimensionality reduction and generating indicators.

2.3 Data Types and Semantics

The above six tables are all spatial datasets because they all store attributes together with the location where they are measure. Here are the types of all data attributes (variables) used. The following introduction will not repeat the same attribute that exist in different tables.

Attribute Name	Type	Note
Geography	Nominal	The name of local authority districts
Geography Code	Nominal	Location Code of local authority districts
Sex	Nominal	"Male" or "Female"
Age	Nominal	It contains many age ranges, rather than just age
Economic Activity	Nominal	
Population (or Measure)	Ratio	The number of people in districts
job type	Nominal	"Full-time" or "Part-time"
Hours Worked	Nominal	It contains many the range of working hours, rather than hours
unpaid care hours	Nominal	ditto
Old-Age Dependency Ratio (OADR)	Ratio	
Real Elderly Dependency Ratio (REDR)	Ratio	
labour participation rate	Ratio	
average worked hours for 24 to 65	Ratio	Average working hours for people aged 24 to 65 in each district
All attributes for data projection	Interval	The zero point of these attributes has no absolute meaning, so they are interval attribute.

3. Task Definition

The visualization task will be based on Munzner's task taxonomy, describing and classifying from three levels: why, how and what so that the visualization tasks can be implemented from abstract to concrete [6].
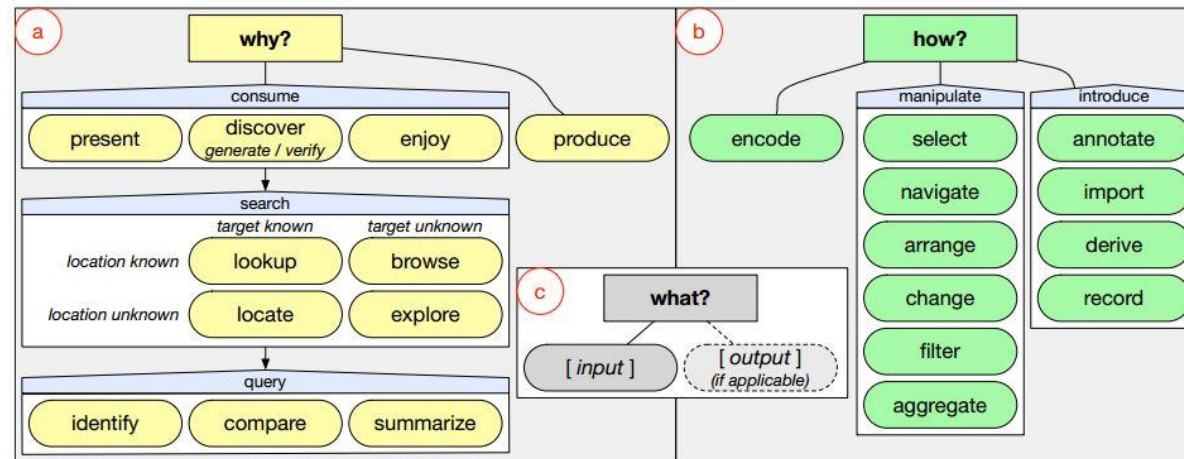

Figure 2: Munzner's task taxonomy

Task No.	Why	How	What: input	What: output
Task 1 (a)	present the degree of aging in all local authority districts and discover the differences between the two aging indicators.	After encoding, system allows users to select different ageing indicator to change visual encoding of regional aging.	"age sex LAD.csv"	A UK map based on local authority districts showing the degree of ageing and a drop-down box containing two ageing indicators.
Task 1 (b)	System can help users locate the local authority district they are	System filters out the unselected area in the map, meanwhile ties the map and the bar	"economic age sex.csv" and "age	A bar chart that shows the composition of economic activity

6

	interested in to identify the specific situation of economic activities in target are	chart about the componentization of economic activities, so the bar chart will automatically change the visual encode.	sex LAD.csv"	
Task 2	System can reflect the top N and bottom N ageing areas.	System can automatically aggregate and filter the unneeded areas and then arrange the ageing indicator about obtained areas.	"age sex LAD.csv"	A bar chart showing the top and bottom rankings
Task 3 (a)	Compare the difference between the average labour participation rate in the young and old areas.	Aggregate according to age and districts, and then arrange data according to aging indicators	"economic age sex.csv" and "age sex LAD.csv"	two line charts showing the change situation in the labour participation rate as the age range increases in the top and bottom ageing areas.
Task 3 (b)	Explore the comparison of different numbers of top N and bottom N ranking aging areas	Allow user to select the different number of Top N and bottom N, and the visual encoding will change automatically.	"economic age sex.csv" and "age sex LAD.csv"	Two input boxes
Task 4 (a)	Generate the average working hours of different age groups in different areas.	Aggregate according to age and districts and make calculation. (It is assumed that each age range follows a normal distribution in the working time range.)	"hours worked by age sex"	A calculation field
Task 4 (b)	Compare the difference between the average working hours in the young and old areas.	Arrange the above calculation field according to aging indicators	"hours worked by age sex.csv" and "age sex LAD.csv"	Two bar charts showing the working hours of different age groups, and the overall average working hours can be displayed at the same time.
Task 4 (c)	Generate the average unpaid care hours of different ages in different areas	Same as task 4 (a)	"provision unpaid care hours age sex.csv"	A calculation field
Task 4 (d)	Identify whether unpaid care time as the reflect of aging is one of the reasons for the decline of average working hours.	Change the number of districts, the visual encoding will automatically change. And select the overall average unpaid care hours by different age group in the Top N and	"provision unpaid care hours age sex.csv", "hours worked	An input box for the number of top N and Bottom N districts and two bar charts showing the unpaid care hours of different age groups in the top N and Bottom N

		Bottom N areas. And	by age sex.csv" and "age sex LAD.csv"	areas, and the overall average unpaid care hours can be displayed at the same time.
Task 5	System allows users to browse the economic activity-based net migration situation between different districts to identify the relationship between economic activity-based immigration and regional ageing.	Filter out the unneeded data by selecting the number of top N and bottom N areas. Aggregate based on net migration and economic activities.	"economic activity net migration.csv" and "age sex LAD"	Four pyramid charts showing migration situation for the economically active and economically inactive categories in young and old areas.
Task 6	Summarize the relationship between labour market indicators and ageing indicators.	Aggregate data based on districts.	"data protection and calculation.csv"	Four scatter plots showing between two ageing indicators and labour market rate and average working hours.
Task 7 (a)	Produce the visualization of data after dimensionality reduction.	Visual encode the output results after three dimensionality reduction methods, and can change the fields of exploration.	"data protection and calculation.csv"	Three scatter charts with a drop-down box containing "overall data", "working hours", "unpaid care hours", "economic activity", "migration situation"
Task 7 (b)	Users can look up the location of the districts of interest (data points), so as to compare the similarity between districts.	Users can select a district and this district will be highlighted automatically in the data projection visualization.	"data protection and calculation.csv"	A UK map showing the ageing indicator, connecting the three data projection scatter plots.

4. Visualisation Justification

4.1 the Used Visualization Techniques and Reasons

Next, I will introduce each visual encoding and interaction in units of each sub dashboard in detail, including the marks and visual channels used, then explain the reasons why these encodes are used based on theories like Visualization Principle, Human-Machine Interaction and Perception. It is recommended to open the Tableau dashboard to try some actions when checking this section for understanding this visualization work better.

Step 1: Regional Overview
Visual Encoding:

Visualization	Mark	Channel
UK map	Area	Shape as identity channel, colour saturation as magnitude channel
A bar chart	Area	Colour as identity channel, size as magnitude channel
A bar chart	Area as containment mark	Colour as identity channel, size as magnitude channel

Interaction Actions:
(1) The drop-down box in the upper left corner can change the ageing indicators used in the map and ranking.
(2) In the bar chart about ranking, two input boxes can increase or decrease the number of top N and bottom N ageing districts.
(3) When the cursor hovers in one district of the map, age composition will be shown.
(4) When some districts are selected, the composition of economic activity will only focus on those districts.
(5) Users can select some top N or bottom N districts in the rank. Then the map will show their location of them, and the composition of economic activity will only focus on the selected districts.
(6) Users also can search the districts on search box of the UK map
Reasons:
(1) For displaying the regional ageing situation, a UK map should be a good choice. But because there are many districts in the map. Even if colour saturation is used as the magnitude channel, it is still cannot reflect those youngest and oldest districts. Thus, a bar chart about ranking was made for meeting the discriminability of effectiveness principle for the UK map [7].
(2) Based on Weber's law, perceptual systems are based on relative judgments, not absolute judgments. Thus, size as magnitude channel in two bar charts are an appropriate option, especially for the composition of economic activity.

Step 2.1: Labour Participation Rate and Population Ageing

Visual Encoding:

Visualization	Mark	Channel
Two line-charts	Line	Colour as identity channel

Interaction Actions:

Two input boxes can increase or decrease the number of top N and bottom N ageing districts which are shown on the line charts.

Reasons:

This visualization aims to compare the difference of labour participation rate in young and olde areas. However, if the marks representing the two types of districts are placed on the same chart, it will violate the separability and discriminability of channel effectiveness. Thus, the overall average LPR by age groups of the UK will be added to the two figures. The size of it is thicker than others, so it will have a popout effect based on eye movement theory for achieving a comparison [8].

Step 2.2: Average Working Hours Per Week and Population Ageing

Visual Encoding:

Visualization	Mark	Channel
A table	Area	Colour saturation as magnitude channel
Four bar charts	Area	Colour as identity channel, size as magnitude channel

Interaction Actions:

Users can use the input box on the right to select the number of top N and bottom N ageing districts.

Reasons: Because the applied data is not complex, based on the simplicity of Gestalt Principle[9], visual design needs to pay attention to reducing the load on the users' brain, so this step only presents four simple bar charts as a whole component.

Step 2.3: The Migration Situation based on Economic Activity and Population Ageing

Visual Encoding:

Visualization	Mark	Channel
Four pyramid charts	Area as containment mark	Colour as identity channel, size as magnitude channel

Interaction Actions:

Users can use two input boxes on the right to increase or decrease the number of the number of top N and bottom N ageing districts which are shown on the pyramid charts.

Reasons：

There are positive and negative numbers in the number of net immigrants from different economic activity groups in each district. Using the pyramid chart is a good choice because it can clearly show groups are decreasing or increasing. More importantly, this can realise the popout effect based on eye movement [8].

Step 3: The Relationship between Labour Market Indicators and Ageing Indicators

Visual Encoding:

Visualization	Mark	Channel
Four scatter plots	Point	Colour as identity channel

Reasons：

Because every district needs to be shown in the plot, the scatter plot will be a good choice. And meanwhile in the scatter plot, the visual channel uses colour as visual popout to identify young areas and old areas.

Step 4: The Relationship between Labour Market Indicators and Ageing Indicators

Visual Encoding:

Visualization	Mark	Channel
UK map	Area	Shape as identity channel, colour saturation as magnitude channel
Four scatter plots	Point	Colour as identity channel

Interaction Actions:

(1) Users can use the drop-down box to select which field of data projection.

(2) The map is connected with three scatter plots. Users can locate the location of the districts by clicking a data point in these tree scatter plots.

(3) Similarly, users can click a district of the map to highlight the location of this district in the three scatter plots.

Reasons：

Compared with viewing the overall trend in step 3, step 4 allows users to view details, that is, to view the similarity between specific districts. Then sole scatter charts cannot help users locate districts, so a map connected with the scatter charts is needed to provide support.

4.2 the Used Data Projection Methods and Reasons

Three data projection methods are used in this visualisation, i.e., Linear Discriminant Analysis (LDA), t-Distributed Stochastic Neighbour Embedding (t-SNE) and Uniform Manual Approval and Projection (UMAP). Firstly, Real Elderly Dependency Ratio (REDR) can determine whether a district is ageing or young so that each data point in the dataset can have a binary label. When the data projection needs to consider the effect of the labels, LDA is the optimal solution because it is a supervised data projection method. But in this scenario, the label just has two values, which means the number of components after data projection is just one. When visualising, it may cause the effect of excessive overlapping data. Thus, t-SNE and UMAP were introduced in this visualization. In many cases, these two nonlinear data projection methods are used for visualization to view the structure of data points, so they are very suitable for dimensionality reduction to 2 dimensions. Besides, in this work, the purpose of the data projection visualization is only to view the similarity of data points (target data point and who its neighbours are), and more emphasis is placed on preserving the local structure information of data. So, t-SNE and UMAP are also suitable.

4.3 Limitations

There is a limitation in this visualization that needs to be declared. When calculating the average working time and average unpaid care time, the dataset only contains the number of people in the time range, not the exact time. Therefore, a hypothesis is put forward that in each time range, everyone's working time or care time obeys the normal distribution. So, we can obtain the overall average time. Although according to Central Limit Theorem, this assumption seems to hold, there are few samples in some areas, so this calculation had some imprecision.

5. Conclusion

5.1 Insights

From this visualization exploration, some valuable insights were found:
(1) The indicator which people usually used, Old-Age Dependency Ratio (OADR), seems like underestimate the ageing degree of many districts in the UK. Real Elderly Dependency Ratio (REDR) shows most districts have been in the ageing phase.
(2) The degree of regional ageing constantly deepened with the inland to coastal areas. Almost all districts where ageing is the most obvious are located in coastal areas.
(3) Compared with young areas, the labour participation rate of young people in ageing areas is higher, but that of the elderly in ageing areas is lower.
(4) People between the ages of 16 and 24 in the most ageing areas work more hours per week than their peers in younger areas. But, however, all remaining age groups in ageing areas work fewer hours per week than their peers in younger areas.
(5) The average unpaid care time may explain the insight 4. The length of unpaid care provided by people aged 0-24 in aging areas is similar to that in young areas, but people over the age of 25 in aging areas provide more care time than their peers in young areas. This shows that working-age people have to spare some working hours to take care of the elderly so that the working hours will be decreased.
(6) A very obvious sign of ageing areas is more migrants can find full-time jobs compared with that in young areas, and another is that many students migrate out from ageing areas.
(7) The last insight is the overall trend: with the deepening of aging, the average working hours and average labour participation rate will both decline.

5.2 Harvest from this Visual Exploration

In the process of making this visual work, I reviewed the two components of visual encoding, visual channel and mark, and reviewed the principles of channel effectiveness, perception theory and the pop-out effect of eye movement. More importantly, I have a deeper understanding of tableau, I learned many new built-in functions such as IIF, FIXED, RANK, etc., and mastered how to set, group, parameter and actions in Tableau.

6. References

1. D.E. Bloom, D.L. Luca (2016). The Global Demography of Aging: Facts, Explanations [online] [Viewed on 12 April]. Available from: https://www.sciencedirect.com/science/article/pii/S2212007616300062

2. J. Légaré (2015). Population Aging: Economic and Social Consequences [online] [Viewed on 14 April]. Available from: https://www.sciencedirect.com/science/article/pii/B9780080970868340417

3. A. Barrett (2015). Aging and the Labour Market. [online] [View from 10 April]. Available from: https://www.sciencedirect.com/science/article/pii/B9780080970868940145

4. J. Spijker (2020). How should we measure population ageing: using the old-age dependency ratio or is there an alternative?. International Centre on Aging [online] [Viewed on 7 April]. Available from: https://cenie.eu/en/how-should-we-measure-population-ageing-using-old-age-dependency-ratio-or-there-alternative

5. Office for National Statistics (2021). United Kingdom Economic Activity Rate. Trading Economics. [online] [Viewed on 7 April]. Available from: https://tradingeconomics.com/united-kingdom/labor-force-participation-rate

6. T. Munzner (2014) Visualization Analysis and Design. Chapter 4: Why: Task Abstraction [Viewed on 15 April].

7. T. Munzner (2014) Visualization Analysis and Design. Chapter 5: Marks and Channels[Viewed on 15 April].

8. C. Ware (2012) Information Visualization: Perception for Design. Chapter 1: Foundations for an Applied Science of Data Visualization [Viewed on 16 April]

9. D. Todorovic (2008) Gestalt principles. Scholarpedia. [online] [Viewed on 17 April]. Available from: http://www.scholarpedia.org/article/Gestalt_principles