Large Dataset Information and Example Exam Questions

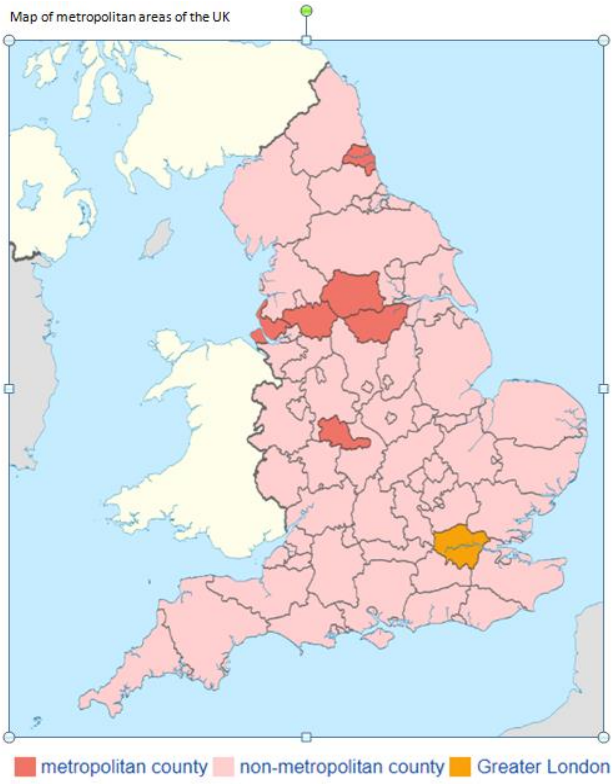
Introduction

The Large Data Set (LDS) consists of four sets of data: two each from the censuses of 2001 and 2011; two on method of travel to work and two showing the age structure of the population.

All four sets of data show information by Local Authority District (LAD), or Unitary Authority (UA), for England and Wales.

- The data is regional and comes from 348 **districts** covering all of **England** and **Wales** which are referred to as the Local Authority Districts (LAD) or Unitary Authorities (UA)
  - The unitary authorities were created by combining some LADs due to change over the 10 years between 2001 and 2011
- The 348 districts are grouped into the following **regions**

North East (12 districts)	East of England (47 districts)
North West (39 districts)	London (33 districts)
Yorkshire & The Humber (21 districts)	South East (67 districts)
East Midlands (40 districts)	South West (37 districts)
West Midlands (30 districts)	Wales (22 districts)



- Non-metropolitan district in England (E08) refers to towns and small cities
- Metropolitan boroughs (E08) refers to big cities is certain areas of England
- London borough (E09) refers to all parts of London
- Be aware that there are other Unitary Authorities in England (E06) that could be either a metropolis or a smaller city
- W06 refers to a unitary authority in Wales, these are mostly rural except for the districts of Cardiff and Newport

What data collection methods were used?

- The data were collated from the 2001 and 2011 censuses
  - The census is carried out every 10 years and collects data from every contributing member of the population across England and Wales
  - It gives the best possible estimate for an overview of the population
- The 'method of travel' data sets encompass all people who were in employment at the time of the census collection
  - The method of travel concerns the part of their journey to work where they travelled the furthest
- The age structure data set is taken from the date of birth question on the survey
- The age is taken to be the person's age on their last birthday before the time of the census collection
  - A person of less than a year old was recorded at 0
  - In 2001 a data entry of over 110 was treated as invalid
  - In 2011 a data entry of over 115 was treated as invalid

What are some of the important features?

- Take some time to review the age structure boundaries
  - They are not even and have widely varying class widths
  - Notice how the class widths for teenagers are particularly small whilst those for ages 30 – 59 are much wider
  - This means looking at a normal frequency diagram or population pyramid would be misleading, instead you should use a histogram and calculate frequency densities
- Consider the variation in age in rural areas compared to metropolises, think about why big cities and the London districts might have less variation in age
- Consider the method of travel to work in rural areas compared to metropolises, think about why big cities and the London districts might have more people who use public transport
- Look at the change in use of public transport from 2001 to 2011
  - Consider some reasons why more people may use public transport in 2011, could there have been improvements in those areas or perhaps government incentive schemes?
- Watch out for correlation, the values are given as direct numbers so as the population increases so will the method of travel to work
  - You would need to look at the proportion to get a better idea of correlation

---

Past Paper Questions

2018 (A2)

11 Christa used Pearson's product-moment correlation coefficient, r, to compare the use of public transport with the use of private vehicles for travel to work in the UK.

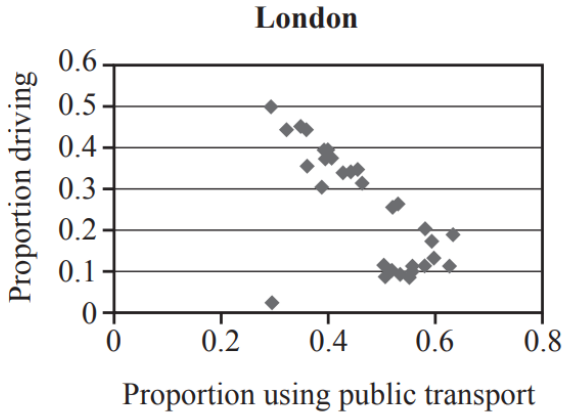
(i) Using the pre-release data set for all 348 UK Local Authorities, she considered the following four variables.

Number of employees using public transport	x
Number of employees using private vehicles	y
Proportion of employees using public transport	a
Proportion of employees using private vehicles	b

(a) Explain, in context, why you would expect strong, positive correlation between x and y. **[1]**

(b) Explain, in context, what kind of correlation you would expect between a and b. **[2]**

(ii) Christa also considered the data for the 33 London boroughs alone and she generated the following scatter diagram.



One London Borough is represented by an outlier in the diagram.

(a) Suggest what effect this outlier is likely to have on the value of r for the 32 London Boroughs. **[1]**

(b) Suggest what effect this outlier is likely to have on the value of r for the whole country. **[1]**

(c) What can you deduce about the area of the London Borough represented by the outlier? Explain your answer. **[1]**

2020 (A2)

14 Table 1 shows the numbers of usual residents in the age range 0 to 4 in 15 Local Authorities (LAs) in 2001 and 2011. The table also shows the increase in the numbers in this age group, and the same increase as a percentage.

	2001	2011	Increase	% Increase
Bolton	16 779	18 765	1 986	11.84%
Bury	11 117	12 235	1 118	10.06%
Knowsley	9 454	9 121	−333	−3.52%
Liverpool	24 840	26 099	1 259	5.07%
Manchester	24 693	36 413	11 720	47.46%
Oldham	15 196	16 491	1 295	8.52%
Rochdale	13 771	14 754	983	7.14%
Salford	12 529	16 255	3 726	29.74%
Sefton	14 896	14 601	−295	−1.98%
St. Helens	10 083	10 269	186	1.84%
Stockport	16 457	17 342	885	5.38%
Tameside	12 803	14 439	1 636	12.78%
Trafford	11 971	14 870	2 899	24.22%
Wigan	17 561	19 681	2 120	12.07%
Wirral	17 475	18 514	1 039	5.95%

Table 1

Fig. 2 shows the increase in each LA in raw numbers, and Fig. 3 shows the percentage increase in each LA.

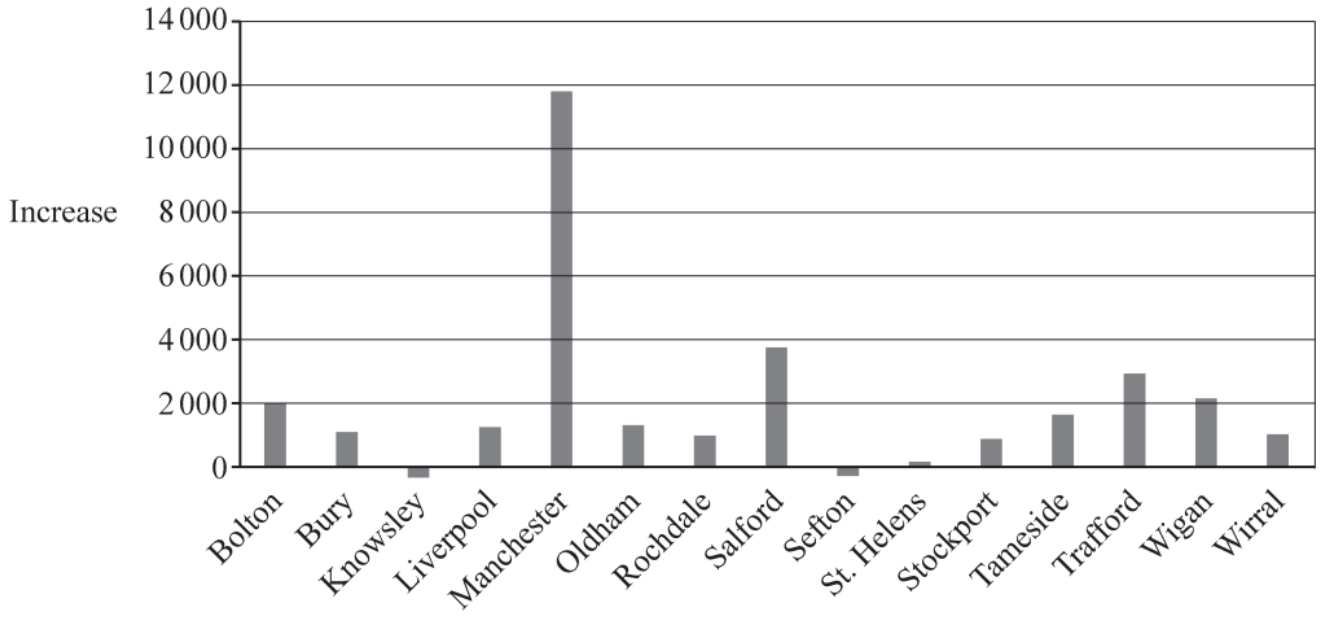


Fig. 2

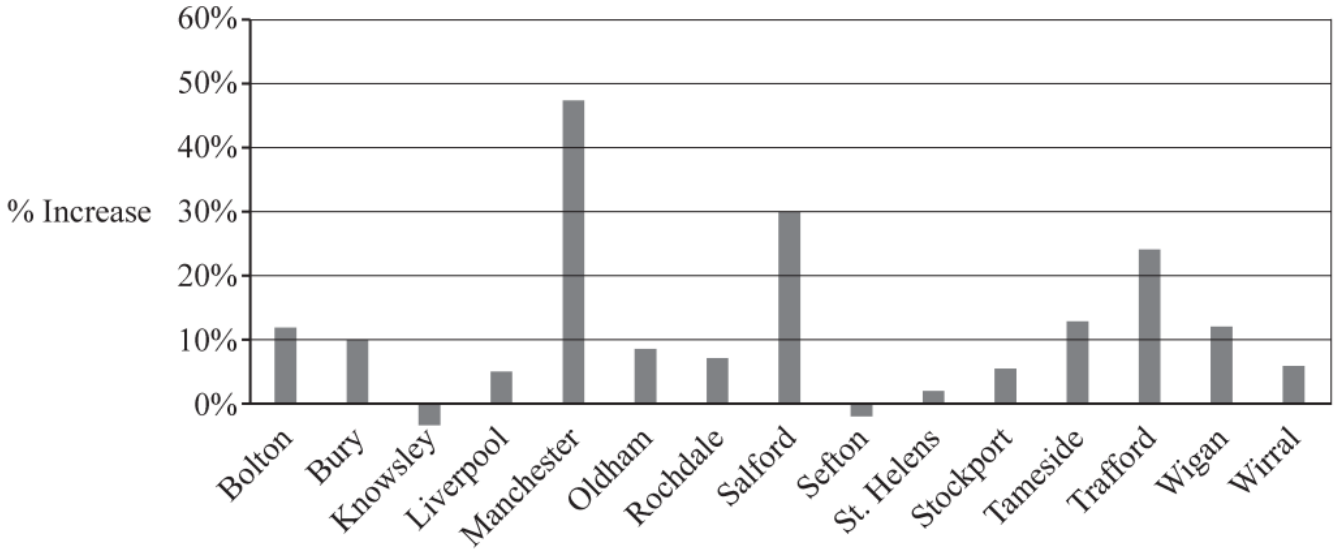


Fig. 3

(a) The Education Committees in these LAs need to plan for the provision of schools for pupils in their districts.

(i) Explain why, in this context, the increase is more important than the actual numbers. **[1]**

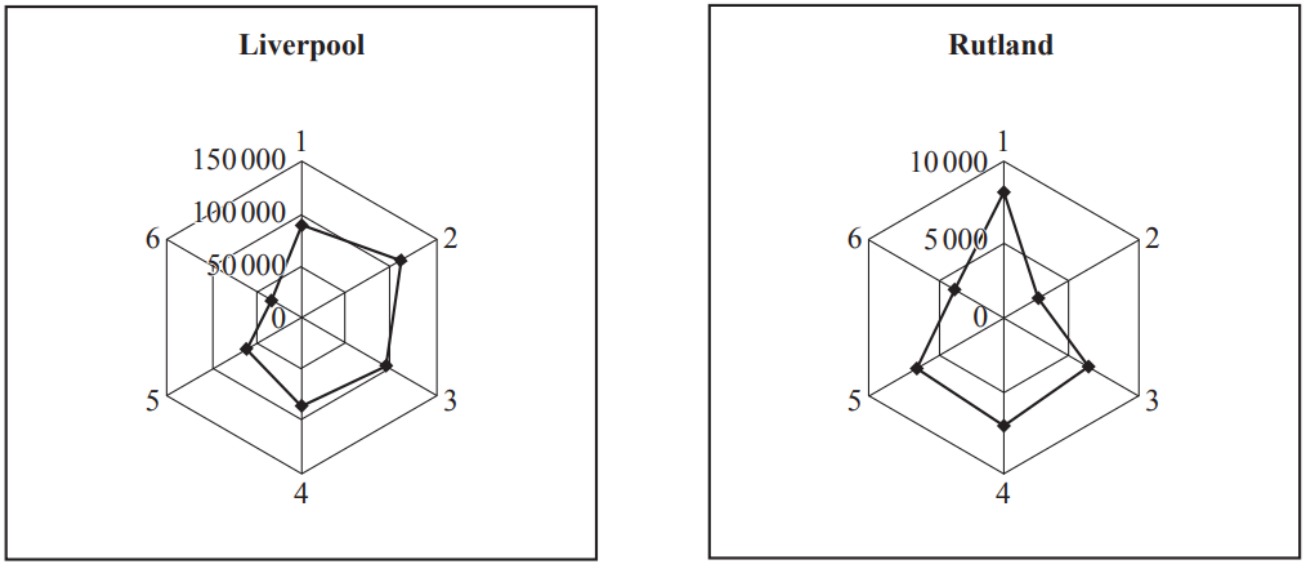
(ii) In which of the following LAs was there likely to have been the greatest need for extra teachers in the years following 2011: Bolton, Sefton, Tameside or Wigan?
Give a reason for your answer. **[2]**

(iii) State an assumption about the populations needed to make your answer in part **(ii)** valid. **[1]**

(b) In two of the 15 LAs the proportion of young families is greater than in the other 13 LAs.
Suggest, using only data from Fig. 2 and Fig. 3 and/or Table 1, which two LAs these are most likely to be. **[2]**

2018 (AS)

13 The radar diagrams illustrate some population figures from the 2011 census results.



Each radius represents an age group, as follows:

Radius	1	2	3	4	5	6
Age group	0–17	18–29	30–44	45–59	60–74	75+

The distance of each dot from the centre represents the number of people in the relevant age group.

(i) The scales on the two diagrams are different. State an advantage and a disadvantage of using different scales in order to make comparisons between the ages of people in these two Local Authorities. **[2]**

(ii) Approximately how many people aged 45 to 59 were there in Liverpool? **[1]**

(iii) State the main two differences between the age profiles of the two Local Authorities. **[2]**

(iv) James makes the following claim.

"Assuming that there are no significant movements of population either into or out of the two regions, the 2021 census results are likely to show an increase in the number of children in Liverpool and a decrease in the number of children in Rutland."

Use the radar diagrams to give a justification for this claim. **[2]**

2019 (AS)

10 The table shows extracts from the "Method of travel by LA" tabs for 2001 and 2011 in the large data set.

Local authority (LA)	All people in employment	Underground, metro, light rail, tram	Train	Bus, minibus or coach	Motorcycle, scooter or moped	Driving a car or van
LA1 2001	79226	14369	5235	20575	1227	16052
LA1 2011	118556	22486	8336	30541	1220	12445
LA2 2001	203614	190	1062	15327	1256	121690
LA2 2011	227894	323	1865	13732	1038	146644
LA3 2001	42993	35	482	4363	274	24105
LA3 2011	49014	33	828	3380	191	28981
LA4 2001	101697	65	693	21758	846	45407
LA4 2011	123218	2495	1315	24275	763	54020

(a) In one of these four LAs a new tram system was opened in 2004.

Suggest, with a reason taken from the data, which LA this could have been. **[2]**

(b) Julian suggests that the figures for "Bus, minibus or coach" for LA1 show that some new bus routes were probably introduced in this LA between 2001 and 2011.

Use data from the table to comment on this suggestion. **[2]**

(c) In one of these four LAs a congestion charge on vehicles was introduced in 2003.

Suggest, with a reason taken from the data, which LA this could have been. **[2]**

2020 (AS)

9 A researcher is studying changes in behaviour in travelling to work by people who live outside London, between 2001 and 2011.

He chooses the 15 Local Authorities (LAs) outside London with the largest decreases in the percentage of people driving to work, and arranges these in descending order.

The table shows the changes in percentages from 2001 to 2011 in various travel categories, for these Local Authorities.

Local Authority	Work mainly at or from home	Underground, metro, light rail, tram	Train	Bus, minibus or coach	Driving a car or van	Passenger in a car or van	Bicycle	On foot
Brighton and Hove	3.2	0.1	1.5	0.8	−8.2	−1.5	2.1	2.3
Cambridge	2.2	0.0	1.6	1.2	−7.4	−1.0	3.1	0.6
Elmbridge	2.9	0.4	4.1	0.2	−6.6	−0.7	0.3	−0.3
Oxford	2.0	0.0	0.6	−0.4	−5.2	−1.1	2.2	2.1
Epsom and Ewell	1.6	0.4	3.9	1.1	−5.2	−0.9	0.0	−0.6
Watford	0.7	2.0	3.1	0.4	−4.5	−1.2	0.0	−0.1
Tandridge	3.3	0.2	4.0	−0.1	−4.5	−1.1	0.0	−1.3
Mole Valley	3.3	0.1	1.9	0.3	−4.4	−0.7	0.2	−0.3
St Albans	2.3	0.3	3.4	−0.3	−4.3	−1.2	0.3	−0.2
Chiltern	2.9	1.4	1.4	0.1	−4.2	−0.6	−0.2	−0.8
Exeter	0.7	0.0	1.0	−0.6	−4.2	−1.5	1.7	3.4
Woking	2.1	0.1	3.7	0.0	−4.2	−1.3	−0.1	0.0
Reigate and Banstead	1.8	0.1	3.2	0.6	−4.1	−1.0	0.1	−0.2
Waverley	4.3	0.1	2.5	−0.5	−3.9	−0.9	−0.3	−0.9
Guildford	2.7	0.1	2.4	0.2	−3.6	−1.2	0.0	−0.3

(a) Explain why these LAs are not necessarily the 15 LAs with the largest decreases in the percentage of people driving to work. **[1]**

(b) The researcher wants to talk to those LAs outside London which have been most successful in encouraging people to change to cycling or walking to work.
Suggest four LAs that he should talk to and why. **[2]**

(c) The researcher claims that Waverley is the LA outside London which has had the largest increase in the number of people working mainly at or from home.
Does the data support his claim? Explain your answer. **[1]**

(d) Which two categories have replaced driving to work for the highest percentages of workers in these LAs? Support your answer with evidence from the table. **[3]**

(e) The researcher suggested that there would be strong correlation between the decrease in the percentage driving to work and the increase in percentage working mainly at or from home.
Without calculation, use data from the table to comment briefly on this suggestion. **[1]**