The Large Data Set

What is a large data set?

- As part of your course there is a large data set that you can use
- It contains lots of information
- You are not expected to memorise any results from the data
- You will have an advantage if you are familiar with the large data set
  - Understand what the variables are
  - Understand the terminology used
  - Understand the context
- You will **not** get a copy of the large data set in your exam
  - if you are required to calculate anything using the large data set you will be given an extract within the question

What is the data about?

- The LDS consists of four data sets covering the age structure of the population in **England** and **Wales** and the method of travel to work from the censuses of 2001 and 2011
- 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)

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
- Look at the different types of districts and be aware of how the method of travel to work changes between them
  - 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
- 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