The large data set (Edexcel)

What is the large data set?
- Edexcel's LDS looks at weather data across 5 locations within the UK and 3 locations abroad.
- Edexcel will test knowledge and familiarity of the data set.
- You will **not** be required to take copies of the LDS into the exam and you will not be expected to have a detailed knowledge of the actual data within the data set.

The Edexcel specification states for questions that use the LDS:
- *The expectation is that these questions should be likely to give a material advantage to students who have studied and are familiar with the data set.*

In particular, it makes the following remarks about questions testing the LDS:
- Questions may assume familiarity with the terminology and contexts of the data and may not explain them. This is so that students that have not seen or studied the data set do not have the same opportunities to access marks as students that have seen and studied the data set.
- Questions may use summary statistics or selected data from, or statistical diagrams, based on the data set – these might be given in the question/task, or as stimulus material;
- Questions may be based on samples related to the contexts in the data set where students' work with the data set will help them understand the background context.
- Questions may require students to interpret data in ways that would be too demanding in an unfamiliar context.

Locations of the large data set
The LDS contains data for 5 UK weather stations and 3 weather stations overseas for May to October 1987 and May to October 2015. This is important that you know which months the LDS contains.
The UK weather stations are:
- Camborne
- Heathrow
- Hurn
- Leeming
- Leuchars

Advice: You will need to have a rough idea of where these locations are in terms of how north and south they are within the UK.

Leuchars is a lot further north than any other location and it is by the sea so we would expect it to be colder than other locations like Hurn and Heathrow.

The great storm of 1987: This happened between 15th October 1987 and 16th October 1987. This brought heavy rain, as well as heavy wind speeds and devastating damage. Check out the videos below for context:
- https://www.youtube.com/watch?v=NnxjZ-aFkjs
- https://www.youtube.com/watch?v=ciBox3QHYq4

The overseas locations are:
- Jacksonville (Florida, USA)
- Beijing (China)
- Perth (Australia)

Data set variables

The LDS features 11 variables:
- Daily Mean Temperature
- Daily Total Rainfall
- Daily Total Sunshine
- Daily Maximum Relative Humidity
- Daily Mean Windspeed
- Daily Maximum Gust
- Daily Mean Wind Direction
- Daily Maximum Gust Direction
- Cloud Cover
- Visibility
- Pressure

If a reading is not available, then it is listed as 'n/a'.

Daily Maximum Temperature

This is measured in Degrees Celsius. Data values are given to one decimal place. A negative value indicates a temperature less than $0°C$.

Daily Total Rainfall

All totals are given in millimetres. **If the total amount of rainfall recorded is less than 0.05mm, then it is recorded as 'tr' (trace)**. **A "tr" value is treated as 0 if we have to do a calculation with it.**

Daily Total Sunshine

Values for this are given in hours and recorded to one decimal place. For example, an entry of '4.5' would indicate that there were 4.5 hours (e.g. four and a half hours) of sunshine on that particular day in that particular location.

Daily Maximum Relative Humidity

Relative humidity is a measure of the saturation of water vapour in the atmosphere. Higher relative humidities indicate that the air contains more water vapour. The values are given as a percentage. Values above 95% are associated with mist and fog. If the relative humidity is 100%, then the air is fully saturated and condensation can occur.

Daily Mean Windspeed

Daily mean windspeed is measured in knots in one column. **1 knot is 1.15 miles per hour.** The daily mean windspeed is also recorded using the Beaufort scale in another column. This is a non-numerical and empirical scale that maps windspeeds to a number and then a descriptive term. The descriptive terms are **light, moderate and fresh**. Light happens the most in most locations.

Daily Maximum Gust

This is the maximum instantaneous speed that occurred over a 24 hour period. It is measured in knots.

Daily Mean Wind Direction

The daily mean wind direction is averaged over the 24 hours of the day. This value is given in degrees relative to true North and is the direction the wind was blowing **from**. The corresponding cardinal direction is also given. Values are rounded to the nearest 10 degrees.

Daily Maximum Gust Direction

This is the direction the wind was blowing **from** in the hour that the corresponding daily maximum gust occurred. Values are given in degrees relative to true North. The corresponding cardinal direction is also given. The direction of the maximum gust is that direction from which the wind was blowing.

Cloud cover

This is a measure of the fraction of the celestial dome covered by cloud. It is measured in eighths. The technical unit in this case is the okta. A value of 0 oktas represents a clear sky, while a value of 8 indicates complete overcast **(discrete variable with 9 values)**.

Visibility

This is measured horizontally. Readings are given in decametres **(1dm=10m)**. Unavailable data is indicated by a dash. Usually between 1000dm and 7000dm. Less than 100dm is fog and less than 200dm is mist.

Pressure

- The measurement is taken at the location and converted into the equivalent amount for the pressure at sea level (to remove the effect of altitude at location).
- Pressure is measured in hectopascals (hPa) to the nearest whole number.
- This value is always around 1000 hPa.
- Higher pressure is associated with good weather
- Lower pressure is associated with rain, cold, cloud.
- The UK locations have pressures of 980 to 1040 hPa.
- The overseas locations have pressures of 1000 to 1040 hPa.

Analysis of the data

UK locations

For the UK locations in 2015, the LDS gives (rounded to the nearest whole):

UK location	Temperature range	Rainfall range	Windspeed range
Camborne	10-20	0-34	3-18
Heathrow	8-29	0-52	3-19
Hurn	6-24	0-26	2-19
Leeming	4-23	0-26	3-17
Leuchars	4-19	0-23	3-23

You can see that Heathrow has the highest recorded value of temperature. This is influenced by the airport and its black asphalt runways and airport buildings which naturally absorb more heat.

Heathrow has an extremely high rainfall range. The highest rainfall value of 51.6ml was recorded on 26th August 2015 and was down to the fact of extreme rainfall within the UK at the time. The next highest was 37.6ml and after this it goes down to 19ml. This suggests that we have two outliers (could be an exam question in the future).

The windspeed ranges are fairly similar amongst the UK locations with Leuchars having some higher values. We can suggest that this is because of its close proximity to the sea.

The locations of Leuchars and Leeming have the lowest temperatures due to their locations being further north than the others.

Overseas locations
The overseas locations have less variables. They only have daily average temperature, daily rainfall, daily mean pressure, and daily mean windspeed in knots and then in the **Beaufort scale (descriptive terms and not continuous data).**

For the overseas locations in 2015, the LDS gives (rounded to the nearest whole):

UK location	Temperature range	Rainfall range	Windspeed range
Jacksonville	15-31	0-80	1-12
Beijing	8-33	0-49	2-9
Perth	8-25	0-102	4-14

Jacksonville
This is located in Florida in the USA, close to the sea. It is in the Northern hemisphere. We have high temperatures here. There is more rainfall here than in the UK due to being nearer the equator and close to where hurricanes form. The UK's rainfall pattern is more spread out as Jacksonville's rainfall is more intense. Jacksonville had 102 days without rainfall in 2015 (May 1st - October 31st). There are moderate windspeeds at this location.

Beijing
This is located in China and is in the Northern hemisphere. High temperatures are expected here with values similar to Jacksonville, however, we see lower temperatures than Jacksonville due to the phenomenon of monsoon circulation. This effect is where cold currents from the North West prevail in the winter and then hot and humid currents from the tropics replace these in the summer. Monsoon circulation also brings **intense** rainfall at times during the summer months. Despite this, Beijing still has periods of time without rain and this is why it had 115 days without rain in 2015 (May 1st - October 31st).

Perth
This is located in Australia. A key difference is that Perth is in the Southern Hemisphere. The significance of this is that the seasons are reversed in the Southern Hemisphere, so between June-August, it is winter. It is located next to the sea, so we expect higher windspeeds here. Perth has two high rainfall values of 63ml and 102ml. The rainfall range would have been 0-38ml if these two values were excluded (probably outliers).