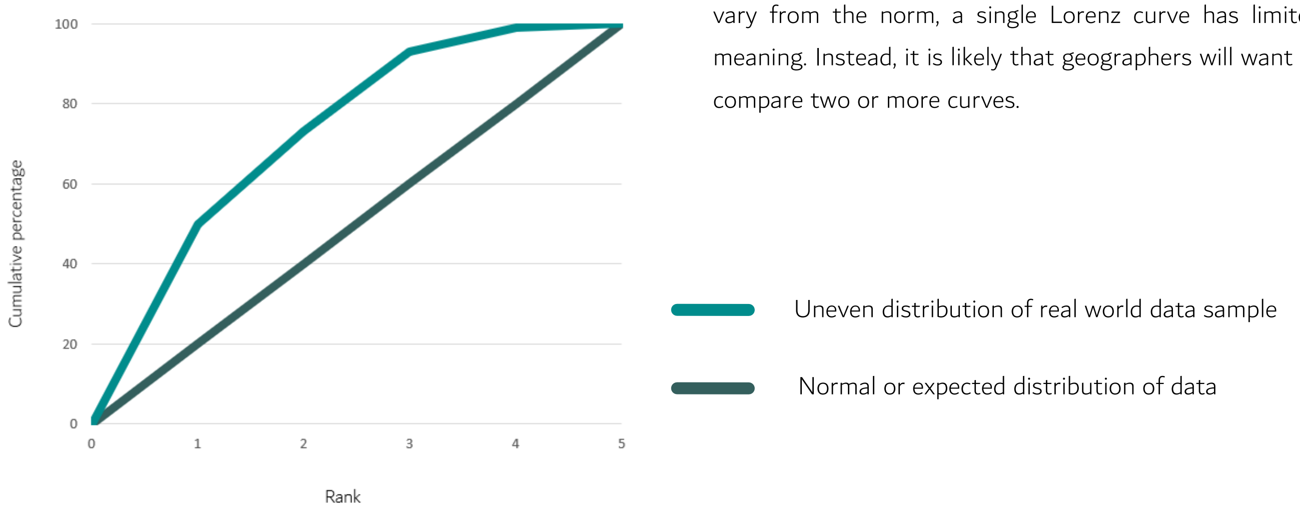
A Lorenz Curve is a graphical form of data presentation that uses basic statistics to show the level of diversity in a data set when compared to a norm or expected distribution. Lorenz curves are designed to work with data put into categories. An expected or 'normal' distribution of that data would place an even amount of data in each category and would be represented by a straight line graph. In reality, a sample of data is likely to fall more in one category than another and diverge or curve away from this line. The greater the divergence from the line, the less evenly distributed the data is in those categories. This divergence is known as the Lorenz curve.

As it is somewhat predictable that real world data will vary from the norm, a single Lorenz curve has limited meaning. Instead, it is likely that geographers will want to compare two or more curves.



Lorenz curves are traditionally used to show economic inequality between different regions or countries. However they can be used more widely in geographical studies:

- Demography - the distribution of a population amongst different **age categories**.
- Rivers - the distribution of bedload samples amongst the different **categories of the Powers scale** at two locations.
- Weather - the distribution of wind records in the **different cardinal and ordinal compass directions** over the course of a month.
- Development - the distribution of wealth among **different sections of a population**.
- Settlements - the distribution of residential dwellings amongst **different Burgess model zones** in a settlement.
- Industry - the distribution of workers of a country amongst **different employment categories**.
- Life systems - the distribution of carbon content levels amongst tree populations in **different categories of woodland** within an area
- Coasts - the distribution of beach sediment recorded in **different geological categories**.

How to carry out a Lorenz Curve calculation

For this example, a geographical researcher wished to investigate how precipitation levels in the month of April had changed over the course of the twenty years, covering 1995 and 2015. They believed that the impact of climate change could be felt more acutely in 2015, with more extreme single rain events, than in 1995 where precipitation was more evenly distributed throughout the month.

Daily Precipitation Total (mm)		
Date	Apr 1995	Apr 2015
1st	0.00	3.55
2nd	0.00	1.84
3rd	0.00	2.58
4th	0.00	0.08
5th	0.00	0.06
6th	0.00	0.05
7th	0.00	0.04
8th	0.00	0.01
9th	0.00	0.09
10th	0.00	0.67
11th	0.00	0.69
12th	0.00	0.05
13th	0.00	0.13
14th	0.00	0.01
15th	0.67	0.03
16th	0.00	0.00
17th	1.15	0.01
18th	0.48	0.00
19th	0.58	0.06
20th	0.00	0.01
21st	6.44	0.05
22nd	4.03	0.00
23rd	3.36	0.08
24th	0.96	2.35
25th	1.83	5.42
26th	0.00	0.17
27th	0.00	0.06
28th	0.00	0.83
29th	0.00	3.10
30th	0.00	0.06

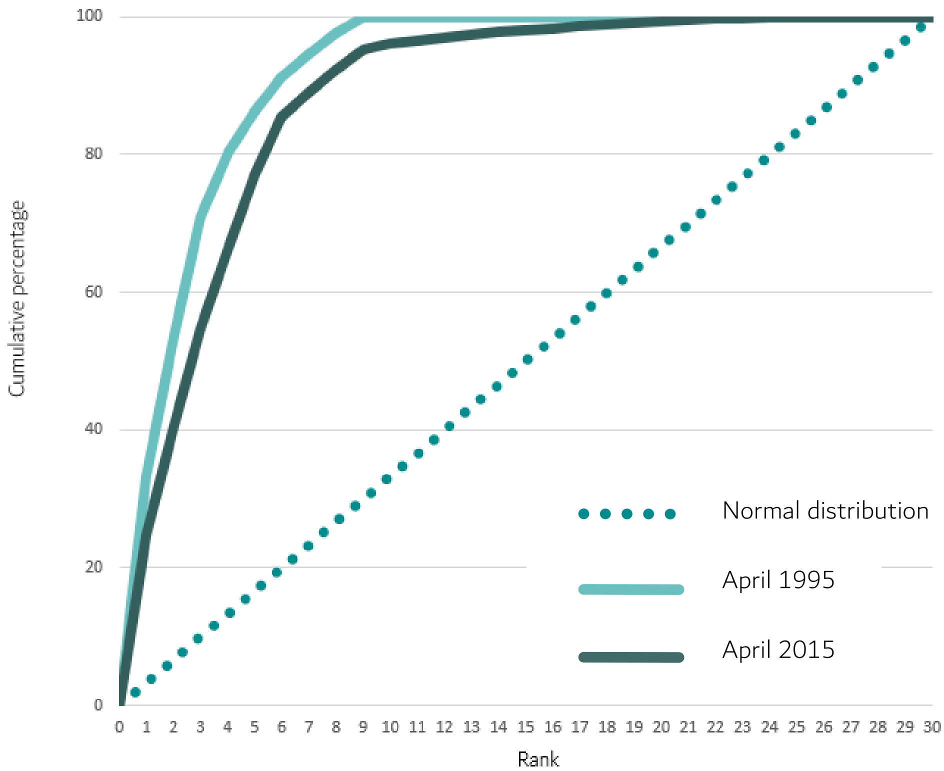
This created the following hypothesis:

"Precipitation data for April 2015 will show a greater degree of inequality across the month than that of April 1995."

The researcher started by creating a table of the raw data from the Met Office Hadley Centre weather station (left).

For each year, the data was placed in size order and ranked from largest to smallest. In this example, the categories necessary for a Lorenz curve to be calculated are thought of as the individual days: an even distribution amongst the categories would be represented by an even amount of precipitation falling on each day of the month. The precipitation for each date was converted into a percentage of the total for that month, and further calculated into a cumulative percentage, from the greatest value to the least (see overleaf).

The rankings and cumulative percentages were then used to draw the Lorenz curves for each year (below).



In this example we can see that the Lorenz curve for 2015 sits at a closer distance to the normal distribution than that of 1995, indicating that the precipitation in April 2015 was more evenly distributed throughout the month. In contrast the precipitation in April 1995 was slightly less evenly spread out and more episodic. Therefore the researcher has to reject their hypothesis.

	April 1995				April 2015		
Rank	Volume (mm)	As % of total	Cum. %	Rank	Volume (mm)	As % of total	Cum. %
1	6.44	33.0	33.0	1	5.42	24.5	24.5
2	4.03	20.7	53.7	2	3.55	16.1	40.6
3	3.36	17.2	70.9	3	3.1	14.0	54.7
4	1.83	9.4	80.3	4	2.58	11.7	66.3
5	1.15	5.9	86.2	5	2.35	10.6	77.0
6	0.96	4.9	91.1	6	1.84	8.3	85.3
7	0.67	3.4	94.6	7	0.83	3.8	89.1
8	0.58	3.0	97.5	8	0.69	3.1	92.2
9	0.48	2.5	100.0	9	0.67	3.0	95.2
10	0	0.0	100.0	10	0.17	0.8	96.0
11	0	0.0	100.0	11	0.13	0.6	96.6
12	0	0.0	100.0	12	0.09	0.4	97.0
13	0	0.0	100.0	13	0.08	0.4	97.4
14	0	0.0	100.0	14	0.08	0.4	97.7
15	0	0.0	100.0	15	0.06	0.3	98.0
16	0	0.0	100.0	16	0.06	0.3	98.3
17	0	0.0	100.0	17	0.06	0.3	98.6
18	0	0.0	100.0	18	0.06	0.3	98.8
19	0	0.0	100.0	19	0.05	0.2	99.0
20	0	0.0	100.0	20	0.05	0.2	99.3
21	0	0.0	100.0	21	0.05	0.2	99.5
22	0	0.0	100.0	22	0.04	0.2	99.7
23	0	0.0	100.0	23	0.03	0.1	99.8
24	0	0.0	100.0	24	0.01	0.0	99.9
25	0	0.0	100.0	25	0.01	0.0	99.9
26	0	0.0	100.0	26	0.01	0.0	100.0
27	0	0.0	100.0	27	0.01	0.0	100.0
28	0	0.0	100.0	28	0	0.0	100.0
29	0	0.0	100.0	29	0	0.0	100.0
30	0	0.0	100.0	30	0	0.0	100.0