

A Visual Investigation of Trends in UK Rainfall

Judith Grieves

Abstract— ‘Britain has a maritime climate that makes it one of the wettest countries in Europe’ [1] and recent years have seen more frequent media reports of flooding in the UK. Floods consume resources in rescue and recovery, cause danger to life and expense to householders and businesses. Floods are caused by a combination of events, such as more frequent building on flood plains [2], deforestation or reduction in flood prevention methods, but is usually triggered by sustained or heavy rainfall.

This paper will use visual data exploration to investigate whether there has been an increase in, or other changes to, the pattern of rainfall in the UK over the last 100 years.

Using UK Met Office data with rainfall values since 1910, we will look at patterns by region, year and annual cycles.

We will show that the data suggest a significant overall increase in rainfall and more extreme rainfall events in parts of the UK.

1 PROBLEM STATEMENT

This paper will use a visual approach to analyse geospatial time series climate data, specifically monthly UK rainfall, provided by the Met Office.

By analysing monthly rainfall data over a long period, this work will analyse whether UK rainfall patterns are changing and whether this could have contributed to the apparent rise in flood events. The main questions to be asked are:

- Has there been an overall increase in UK rainfall in recent years? If so, when did the upward trend begin and where is it happening? What other patterns have there been across the period of the data?
- Are there more extreme rainfall events than previously? That is, more occasions when the rainfall has extreme high or low values?
- What is the cyclical pattern of rainfall across a year and has this cycle changed over the period of the analysis?

The data we will analyse consists of monthly UK rainfall values (mm) for 9 regions of the UK from 1910 to the present.

These data will allow, on a month by month basis, identification of any overall trends, change in frequency of extreme months and pinpoint where and when any changes occurred.

The geospatial element of the data is high level and clearly defined so our approach will be primarily focussed on the time series analysis of the rainfall variable although we will identify spatial elements.

2 STATE OF THE ART

In this review of the geospatial and time series literature, 3 papers are considered. Two are visual data explorations of Climate data [3] [4] and a third [5] which, although analysing energy consumption, presents interesting generic methods for insight into univariate time series data that will be employed in the analysis.

The climate analyses use data from an ECHAM5 climate model run and the ERA-40 reanalysis incorporating observational data [3] and daily satellite observations of sea surface temperatures (SSTs) in the Tropical Pacific [4].

The papers investigate whether we can identify particular subsets in climate data—both in time and space—that potentially represent indicators of atmospheric climate change to narrow down the parameters for subsequent statistical analysis and also assess the influence of smoothing parameters and trend time-frames on the findings [3]. They also ask how geoscientists can be helped to gain a better understanding of geospatial time series using a visual analytics approach that allows them to extract and explore various sets of spatial situations to detect characteristic spatiotemporal patterns in the data. [4]

Each of the papers recognise that, particularly in climate analysis, a priori hypotheses are rarely available so the approach is to provide visual overviews and the ability to subsequently detect and zoom in on peculiar patterns or sub-sequences.

[3] aims at identifying promising hypotheses that are then checked in an analytical, confirmative process. These signals are compared with the climate noise to assess the significance

of the findings. The approach uses Interactive visual analysis and enables users to get into a visual dialog with the climate data. An interactive visualization, according to user input, is generated. This often leads to new questions and/or hypotheses, which can be explored and analysed in more detail in an iterative process

[4] uses the results of hierarchical clustering and groups all time steps of a geospatial time series into a hierarchy of clusters. Users can interactively explore this hierarchy to derive various sets of spatial states.

[5] and [4] point out that the problem is often to identify patterns and trends on multiple time scales (days, weeks, seasons) simultaneously and that different time scales should be investigated in repetitive data patterns.

The papers provide useful approaches that will be employed when analysing the UK rainfall data:

- Classifying the Time series according to their *similarity* to a user-defined pattern [3]: our analysis will classify the monthly rainfall data according to percentile and counts above a parameter value when analysing extreme events (fig. 6).
- Considering the temporal data as two-dimensional, for instance as f(day, hour). The dimensions can then be mapped on different axes with the third dimension used to display the data [5]: For rainfall, year will be mapped against months in a heatmap for the cyclical analysis (fig. 8).
- Employing clustering to identify periodicity patterns and visualising adjacent calendar data [4]. The rainfall year-month cyclical patterns will be clustered and visualised in a similar way (fig. 9).

3 PROPERTIES OF THE DATA

The data to be analysed was collected from the UK and Regional Series of the UK Met Office website [6]

9 regional files of identical format were downloaded. Each file covers a region of the UK and provides monthly rainfall values (mm). The cleaned files have 109 rows per file representing years 1910-2019.

A process of load and transformation takes the 9 initial raw data files to a 'tidy' dataset that will form the basis of the analysis (fig. 1).

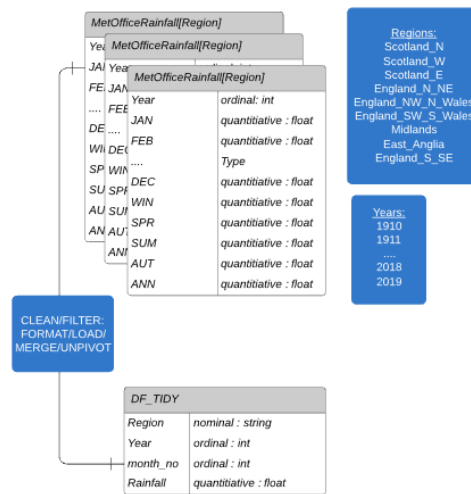


Fig. 1. Data formats: from load to transformation

LOAD – CLEAN, FILTER, FORMAT AND MERGE

- Individual raw datasets are loaded, free text headers are removed, space delimiters edited to convert to csv and these structured csv files imported.
- 2019 data row is dropped as this is not complete, a region column is added and populated from each filename.
- All 9 datasets are merged into a single DF_LOAD dataset and Pandas 'describe' on the loaded data shows that the required columns are of good quality without nulls [8].

TRANSFORMATION – DROP COLUMNS, UNPIVOT

- Subset the data columns to take only the 12 monthly columns required.
- Rename the month columns to numbers – nominal to ordinal to allow sorting
- unpivot/melt the data so that month number is a column and each row has a single rainfall (mm) value for year, month, region. This is the master dataset that will be used to create any subsequent datasets: DF_TIDY.

TRANSFORMATION - ADD FEATURES

New column 'decade' (ordinal : int) for each row so that data can be summarised to a higher level

New column 'regional percentage bin' (ordinal : int) to show the percentile of the rainfall amount within its region.

VALIDATION

Check data counts per Region (109 years) and per Year/Month (9 regions) to ensure all the data is populated [8].

A histogram of total rainfall data [8] shows it to be normally distributed but right skewed towards the higher values.

The data appears to be complete and consistent and no obvious problems are found.

CREATE TIME SERIES

From the master dataset, create various aggregate and time series datasets for the analysis.

4 ANALYSIS

4.1 Approach

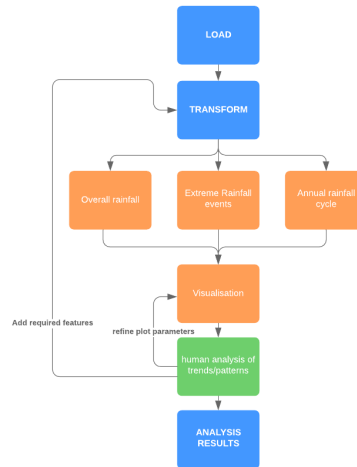
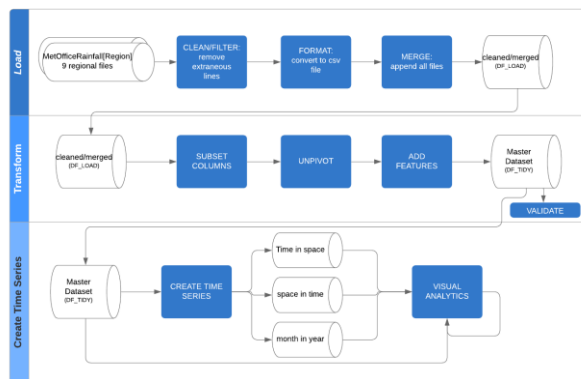


Fig. 2. (a) Data Transform and (b) Analysis Process Flow. Show the steps taken to clean, filter and merge the raw data, to transform with added features and create Time Series datasets to input into the 3 visual analyses.

The analysis approach is summarised in Fig. 2 which shows the steps to be taken.

After the data is cleaned, filtered and loaded, a transformation process adds derived new features deemed useful in the approach and found to be useful in iterations of the analysis process.

A number of Aggregated and Time Series datasets are created to provide the input for each of the analytical questions:

Master Dataset: Overall Rainfall Trend analysis / computations

Time Series: Year Region, Decade Region, Month Region, Month Year.

The following iterative process will be followed for each of our questions:

- 1) For each of the analytical questions, a visual of the data is presented, beginning with an overview of the entire data in the first instance.
- 2) Prominent trends are noted and further questions generated.
- 3) In order to answer the generated questions, a further visual is generated on a subset of the data for a given attribute or for an aggregated attribute. Where this requires new data, the feature engineering of the

attribute will be added to the appropriate data transformation section. Return to (2) and continue the iteration.

4.2 Process

Overall Rainfall Trends

To investigate the question of whether there has been an overall increase in UK rainfall, we look initially at the data for all regions [8]. This shows UK rainfall amounts increasing in the last 2 decades since 2000.

To view this trend in more detail we show regions separately for all years with a plot of a year-region time series (Fig.3).

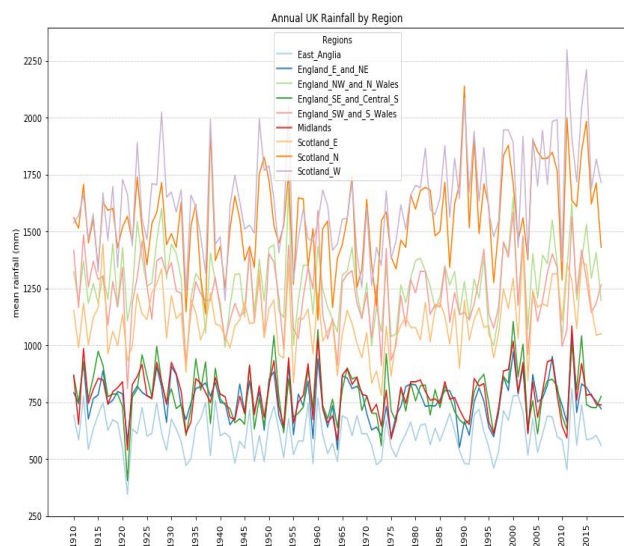


Fig. 3. Annual UK Rainfall by Region (1910-2018)

This visualisation shows that overall UK rainfall follows a relatively consistent pattern within and between regions, that is, although the rainfall amount varies from year to year, regions have high and low values in roughly the same years. This is not unexpected in a small area such as the UK. It also shows however, that the regions have quite different amounts of rain and clearly fall into 3 distinct bands of high, medium and low rainfall. The areas of highest rainfall are in the West and North and lowest in the South and East.

We can see higher and lower rainfall in particular years and the yearly rainfall looks to be trending upward, particularly in the areas of higher rainfall. This upwards trend is not definitive in Fig.3 so we investigate further by plotting each region separately with a median line for comparison [8].

The region by region visual shows a definite rise in annual rainfall for the Scottish regions and highlights that since 1990, W Scotland has had only 4 years with below median rainfall. Scotland West and North have had many instances of breaching previous maximum and upper percentile yearly values; we will do some analysis of these breaches for parameterised boundaries below (fig.6).

Scotland East has not breached previous maximum values in recent years but has had rainfall in the upper percentiles many more times since 2000.

The Midlands and East Anglia show no strong pattern over the period though both had a new historic maximum in 2012. NE/NW & N Wales/SW and S Wales appear to have had more years of high rainfall since 2000.

SE England has no discernible upward trend, though more years of high values than previously.

The visualisation of individual regions also shows that most had a new maximum value in 2012.

We now visualise mean rainfall, time binned by decade, to quantify these trends and establish when they started.

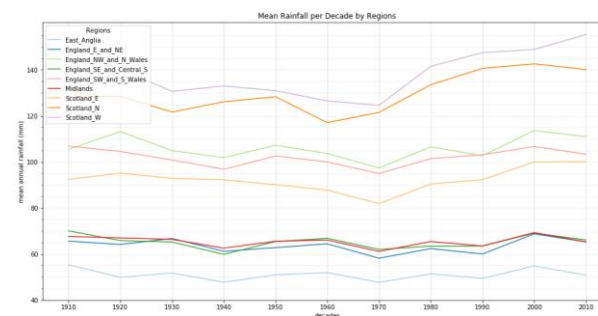


Fig. 4: Regional Mean Rainfall by Decade – all regions

Visualising the rainfall by decade (Fig. 4) smooths the graph and now shows clearly that

- for Scotland North and West from 1980, the mean rainfall per decade rose above any previous value in the last 100 years and has continued to rise. It can be calculated from the data that since 1980, Scotland N&W mean annual rainfall is 10% higher and Scotland E 3% higher than the mean of the previous 8 decades.
- SW and NW England and Wales are also showing a rise in RF, after a low in 1970, but similarly high levels were seen in the 1920s so the current values are not unprecedented and the means are virtually unchanged.
- East Anglia is showing no overall trend.

Fig. 4 does not clearly show trends for areas of low rainfall so we zoom in and plot these regions separately in fig. 5 in order to better detect any smaller increases.

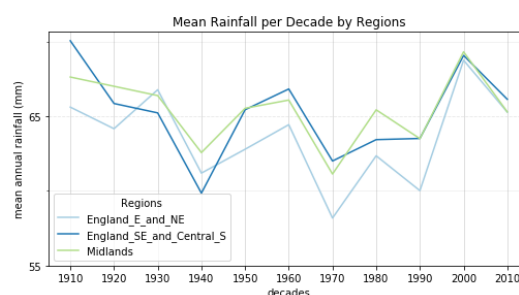


Fig.5 Rainfall by Decade – low rainfall regions

This more detailed plot (Fig. 5) shows now that E and NE England and the Midlands have seen increased rainfall since the 1990s that exceeds any previous high values. Conversely SE and Central South England has no discernible increase but had a record high in the decade beginning 1910.

In fact, calculating statistics for these Eastern and Southern regions shows that the mean annual rainfall has shown a marginal decrease [8].

‘Extreme’ Events

An extreme event is defined as a monthly rainfall value that exceeds an adjustable percentile parameter and the counts are examined by year and decade for those events equal to or over percentiles=80, 90 and 95 to analyse at what value any trends are identified. To investigate whether there has been an increase in extreme rainfall events, we add an attribute of percentile to the data and examine heatmaps of the number of these events in a time period.

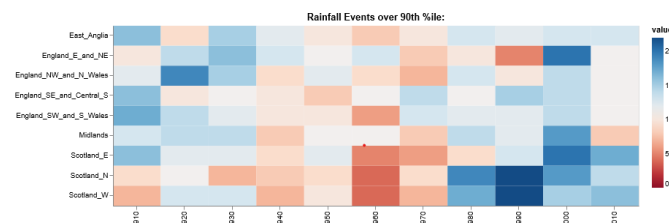


Fig. 6. Heatmap of UK Monthly Rainfall over 90th percentile

For all parameters, pre-1950 there is a wide distribution of counts, followed by some years (1950-1980) with few extreme highs; this corresponds to years of low rainfall seen in fig 4. From 1998 there is small but significant increase in high rainfall events across all regions but especially in N & W Scotland.

Analysing by decade (for k=90,95), there is a clear increase in extreme rainfall events for Scotland (N&W since 1980s, E since 2000s), the Midlands and NE England, (during the 2000s but not since). See fig. 6 which shows the increase in events over the 90th percentile.

For percentile=90, a year based heatmap also shows a small but distinct increase in extreme rainfall events for the regions identified by the decade analysis above. In addition, it also shows extreme events for East Anglia and SE England in particular years.

For k=95, the yearly picture also shows previously unseen spots of extreme monthly rainfall events, from 1990 in N&W Scotland and 2000 elsewhere.

NW & SW England and Wales do not appear to have had more frequent extreme events.

Computations on this time series yield the following conclusions:

- E and NE England and the Midlands had an unprecedented number of high rainfall months in the 2000s (circa 20) but no pattern otherwise.
- The mean number of extreme monthly rainfalls per decade in Scotland N & W has increased from 8 before

1980 to 18 since. E Scotland increased from 10 to 15 in the same time period.

- Other regions had very small increases.

Cyclical Analysis

To investigate the cyclical nature of the rainfall over a year, we first look at the monthly mean for a region over the time period (fig. 7).

The annual rainfall varies fairly consistently by monthly mean across all regions, with highs in Winter months and lows in Spring and Summer.

As previously seen, the regions fall into bands of high, mid and low rainfall and this visualisation shows that the annual cycle is consistent across regions within a band.

High rainfall occurs in October to January in all regions.

However the plot shows that the driest months differ between areas of high to mid rainfall and those with lower; Scotland and the Western areas of England have lowest values in April to June, while the South and East minimums are from February to June. East Anglia, the area of lowest rainfall overall, has minimum rainfall in Feb - April.

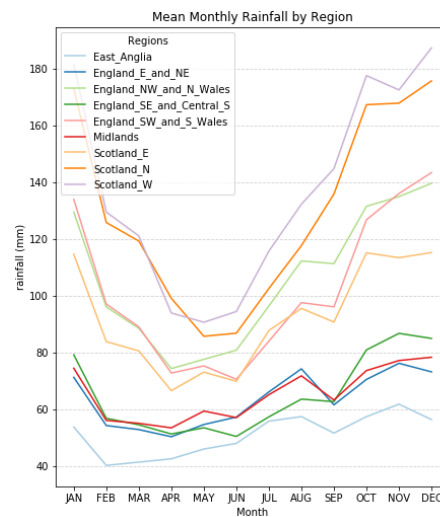


Fig. 7. Mean Monthly UK Rainfall (1910-2018)

We drill down by plotting each year's rainfall for a single region but these visuals [8] show a complex picture and no obvious trends. We will visualise year-month means on a heatmap to attempt to understand this data.

A heatmap across all regions still does not show any pattern other than the overall seasonal one we saw above of wet winters and drier summers so we visualize more heatmaps for individual regions and the regional bands we saw in fig. 3 and fig. 4.

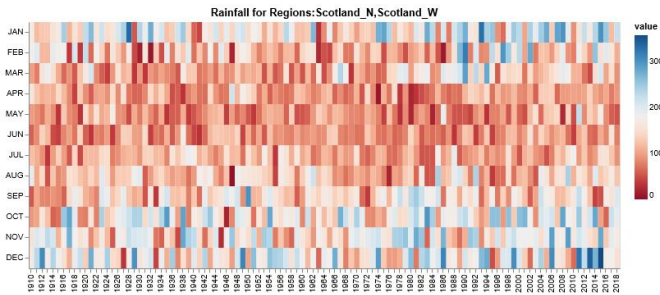


Fig. 8: Heatmap of Annual Rainfall cycles for Scotland North & West

These show some evidence (fig.8) that for N & W Scotland the Winter rainfall period is starting earlier (August when previously September) and finishing later (March rather than February). This may partially explain the overall increased rainfall in these regions.

No discernible pattern can be seen for other regional bands so we will do some K-Means clustering on the data to see if the algorithm can detect more meaningful patterns:

For all regions, we attempt to find an appropriate number of clusters, by plotting Inertias analyses, but no meaningful results were obtained.

The clustering is therefore carried out trying arbitrary values of $k=3-4-5$ to see if there are meaningful patterns. The clusters for all regions' rainfall give plots of distinct annual cycles but the years in which they are observed do not appear to follow any meaningful pattern. The clustering is repeated for individual regions.

We cluster for $k=3-4-5$ in Scotland-N & W and comment on the results.

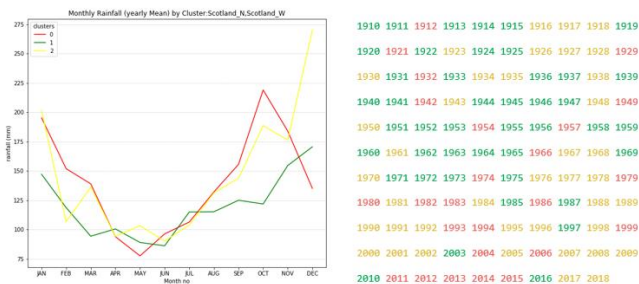


Fig. 9. K-Means Clustering ($k=3$) of Cyclical Rainfall - Scotland N&W

Scotland-N & W in 3 clusters (fig. 9) shows the only significant yearly trends. The visual shows that clusters 0 and 1 have higher October/November rainfall than cluster 2 and the calendar visual shows that these clusters have been prevalent from the 1980s onwards.

4.3 Results

The data show that UK rainfall, though varying highly from year to year, appears to have increased in recent decades.

We saw the disparity between and within regions. Scotland W (1641 mm) has over 2.5 times the mean annual rainfall of East Anglia (613 mm) and in the South East of England the

wettest year (1104 mm) was 2.7 times the driest (406 mm) over the period.

Even allowing for this variance, the data highlight a large increase in annual rainfall over the last 40 years for Scotland.

In addition to this overall increase, the UK has seen an unprecedented rise in 'extreme' rainfall events (Fig. 10). This is also highly regional, with isolated increases during the 2000s in East and NE England and the Midlands and increased extreme monthly rainfalls per decade in Scotland N & W.

Count of Monthly Rainfall over 90th %ile (with median line)

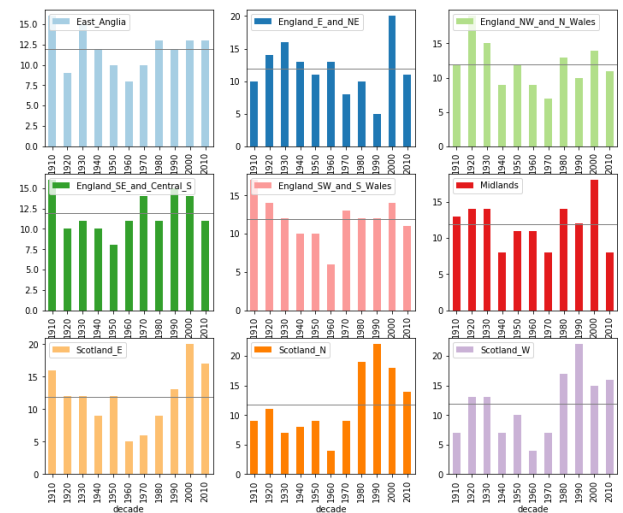


Fig. 10. Number of Monthly Rainfall Events over 90th percentile by Region

The analysis of the cyclical trends showed Scotland N & W with wetter months starting earlier and finishing later and the K-means clustering confirmed this trend.

5 CRITICAL REFLECTION

The chosen approach of a top down analysis of the data in order to identify patterns worthy of deeper study was a useful one. Many previously unexpected aspects of the data were discovered and reported. The use of visualisation was particularly helpful in this approach as it quickly and easily showed trends and exceptions suitable for more detailed analysis. The literature review provided inspiration in methodologies and visualisation styles.

Many more detailed conclusions remain undiscovered through inexperience and a lack of time and space. Further work would dive deeper into the existing analyses and investigate wider questions such as periods of low rainfall and their spatio-temporal features.

Although the 'extreme' event analysis presented some discoveries, on reflection, more granular daily data may be better suited to this investigation.

The K-means clustering analysis did not discover strong or wide-ranging conclusions and it remains to be seen whether further work with this or other data could uncover such patterns or, conversely, confirm that they do not exist. Future analysis could extend this search to include a seasonal element.

The Python code [8] resulting in the analyses presented was made as generic as possible with variable parameters to enable it to be used to identify patterns in this and other similar data for other geographies. This was reasonably successful. For example, the UK Met Office produce similarly structured data for minimum, maximum and mean temperature, sunshine and frost and a similar analysis of these data could be performed from the code with minimal changes.

Table of word counts

Problem statement	227/250
State of the art	489/500
Properties of the data	346/500
Analysis: Approach	225/500
Analysis: Process	1442/1500
Analysis: Results	176/200
Critical reflection	251/ 500

REFERENCES

- [1] <https://www.ilankelman.org/crichton/2011TCP.pdf>
- [2] <https://theconversation.com/how-englands-broken-planning-system-has-created-not-reduced-the-risk-of-floods-127287>
- [3] Kehrer, J., Ladstadter, F., Muigg, P., Doleisch, H., Steiner, A. & Hauser, H. 2008, "Hypothesis Generation in Climate Research with Interactive Visual Data Exploration", IEEE Transactions on Visualization and Computer Graphics, vol. 14, no. 6, pp. 1579-1586.
- [4] Köthür, P., Sips, M., Unger, A., Kuhlmann, J. & Dransch, D. 2014, "Interactive visual summaries for detection and assessment of spatiotemporal patterns in geospatial time series", Information Visualization, vol. 13, no. 3, pp. 283-298.
- [5] Van Wijk, J.J. & Van Selow, E.R. 1999, "Cluster and calendar based visualization of time series data", IEEE, , pp. 4.
- [6] <https://www.metoffice.gov.uk/research/climate/maps-and-data/uk-and-regional-series>
- [7] <https://www.carbonbrief.org/met-office-report-spells-out-climate-change-link-to-uk-storms-and-flooding>
- [8] Python Notebook: https://cityuni-my.sharepoint.com/:f/g/personal/judith_grieves_city_ac_uk/EtkDBxMxow9FoUHVCMckgAj4BIXAJwTuXK_sRxtQ_zbeiuQ?e=L62u8V