**GEOGM0054 – Introduction to Scientific Computing**

**Assessment 1 – Analysing Geospatial Datasets: the 2022 Heatwaves**

**Introduction**

In recent decades, the frequency and strength of heatwaves has increased on a global scale, with particular heat extremes being observed across the European continent (Rousi, et al., 2022). In 2022 much of the northern hemisphere experienced multiple extreme heatwaves. Record temperatures of 40.3°C were recorded in the UK during its second heatwave in July (Kendon, 2022). This report, along with the accompanying code and analysis, puts these recent extreme weather events into context.

Python is a suitable and efficient tool for this analysis as it a high-level programming language with a number of easy to install packages, specifically designed for analysing and manipulating large datasets. It is easily reproducible, and the results can be re-created by following the same process detailed in this report and the notebook. The packages used are detailed in the Methods section of this report.

**Methods**

In order to undertake this analysis, the python programming language was used and written in a Jupyter Notebook, within the Jupyter Lab environment. Jupyter Lab was accessed via the Anaconda-Navigator, which is a freely available software that can be run on almost any computer operating system. As a starting point, a new environment was created in the Anaconda-Navigator and by using the ‘conda-forge’ channel, the following packages were installed:

* Python3.9
* NumPy
* Matplotlib
* JupyterLab
* Geopandas
* Xarray
* NetCDF4
* Cartopy
* Regionmask

Once this environment has been created and the Jupyter notebook is opened within it, the relevant packages can be imported as detailed in the first cell of the notebook that accompanies this report.

There are several parameters that can be adjusted in the code so that temperature analysis can be carried out for periods and regions that differ from those used in this particular evaluation. These are detailed moreover in the notebook, but a few notable parameters are:

* The months used to select the analysis (summer) period
* The years used for the reference period
* The countries chosen to compare against the UK

In order to understand the summer heatwaves from the UK against a global perspective, the northern hemisphere summer months of June, July and August were chosen. A reference period of 29 years (1960-1989) was chosen to compare against the long-range average of 2 metre-temperatures. These parameters can be adjusted in order to draw differing outcomes or conclusions from the data.

Two main datasets were used as inputs for this analysis, which are:

* The ERA5 data, which was downloaded from the web interface of the Climate Data Store and is provided by the Copernicus Climate Change and Atmosphere Monitoring Services and distributed under the Copernicus Product Licence. (https://cds.climate.copernicus.eu/api/v2/terms/static/licence-to-use-copernicus-products.pdf).
* The GPWv4 dataset has been obtained from the Socioeconomic Data and Applications Center (SEDAC) at https://sedac.ciesin.columbia.edu/data/set/gpwv4-population-count-rev11 and is licensed under the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0>).

**Results & Discussion**

One method of quantifying the strength of a heatwave or season is to compare the difference between the measured surface air temperatures and a long-term seasonal average. To follow this approach, the 2-metre air temperatures from the ERA5 reanalysis were utilised. Due to the size of this dataset, it would be computationally expensive to run our analysis whilst continuously referring to it as a whole. Thus, the first step was to clean the data and process it into a file size and shape more appropriate for the subsequent data evaluation tasks.

First, the summer months for each year were selected and by using xarray’s ‘groupby’ the summer mean temperature for these months was calculated. This provided the respective summer average for each year and grid point, whilst reducing the time dimension of the dataset to 64. Summer anomalies were then calculated by defining a reference period – years 1960 to 1989 – calculating the mean and subtracting it from the seasonal averages. The resultant dataset is one that has a global grid with 64 timesteps and 2 metre-temperature means for each grid point. This output dataset was used for the remainder of this assessment. The file size of this new dataset has a reduction from 755.4 MB to 63.3 MB. This will help to speed up any subsequent analysis. There is also the benefit of the new dataset being specific to the needs of the assessment. The next step was to use this anomaly dataset to put the heatwaves into context. The years 1976, 2020, 2021 and 2022 were selected from the temperature anomaly data and plotted on four subplots using the matplotlib function. **Figure 1** shows the June, July and August (JJA) anomalies for these years.

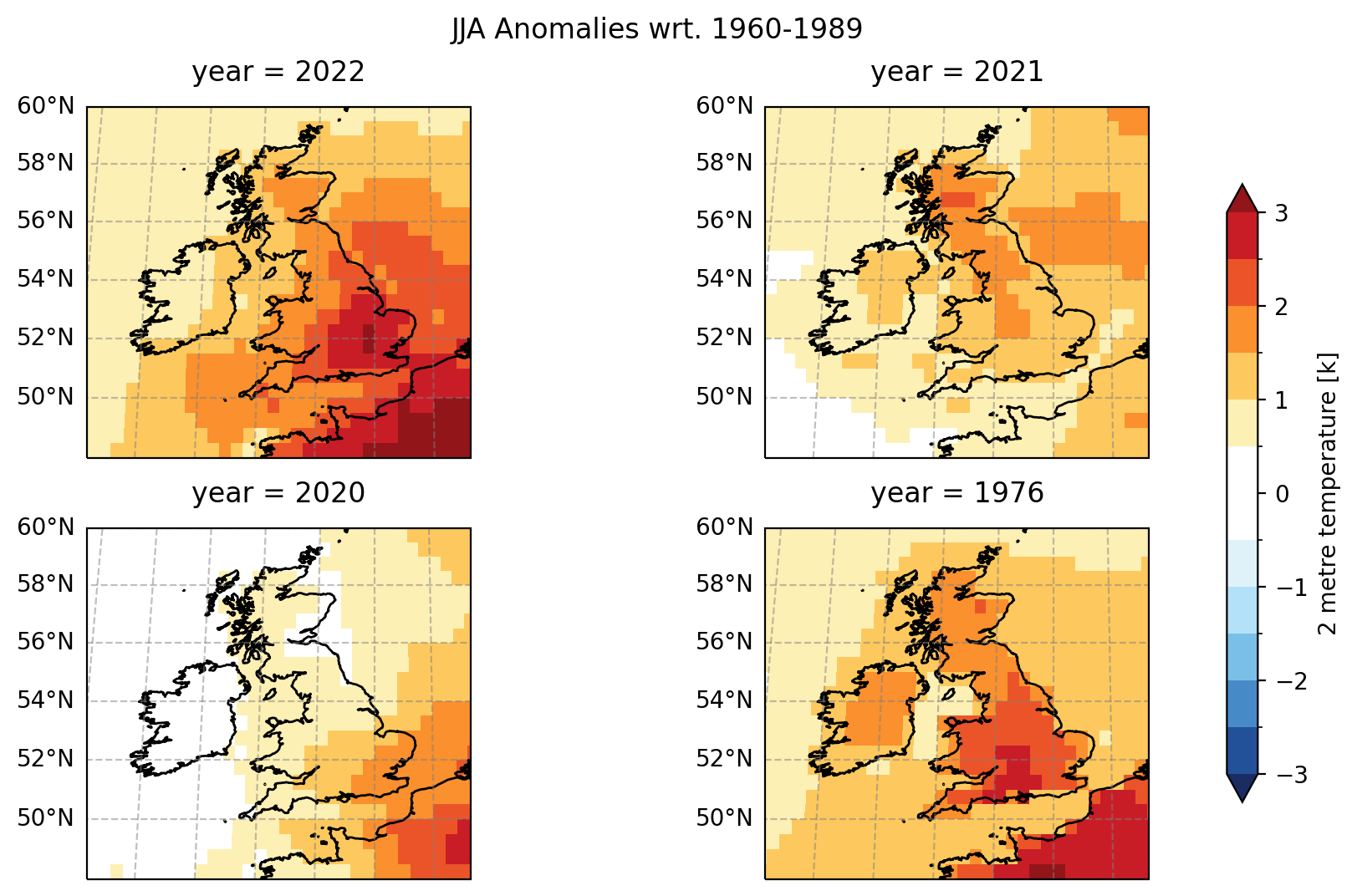


Figure 1: 2 metre air temperature anomalies for the months June to August over the last three years and for the 1976 heatwave. Data is from the ERA5 reanalysis with anomalies relative to the 1960-1989 mean.

As is apparent in **Figure 1**, the summer heatwave of 2022 is largely comparable to that of the 1976 heatwave in the UK. However, by plotting the temperature anomalies on a global map, it is clear that the UK was an outlier in the summer of 1976.

**A picture containing chart

Description automatically generated**

Figure 2: 2 metre air temperature anomalies for the months June to August over the last three years and for the 1976 heatwave, on a global scale. Data is from the ERA5 reanalysis with anomalies relative to the 1960-1989 mean.

As shown in **Figure 2**, the global temperature anomalies are much higher in 2022 (and the two years preceding it) than in 1976. **Figure 2** suggests a trend of increasing global temperatures over the last few years which is expected to continue in the future. By plotting the global mean anomalies on a timeseries graph, this trend becomes explicitly obvious.

By using xarray’s built in ‘mean’ function and passing the latitude and longitude dimensions of the dataset to it, the mean can be calculated. However, doing this calculation and giving equal weight to each to each grid point isn’t a true representation of the global mean. This is due to the fact that each "pixel" (or grid point) does not represent an equal area of Earth's surface. On a global grid, the relative area of each grid point is proportional to cos(𝜆), where 𝜆 is latitude. In respect of this, a weighted average must be calculated for the global mean by using xarray’s weighted array reductions, details of which can be found in the accompanying notebook.

**Figure 3** shows the weighted global mean values of the anomaly temperature field, for each year between 1959 and 2022.

**Chart, line chart

Description automatically generated**

Figure 3: Timeseries graph showing the weighted Global Mean Anomalies for years 1959 to 2022, with the chosen reference period highlighted. Data is based on the processed ERA5 anomaly dataset.

The JJA global temperature anomaly in 2022 is +0.818k and the 1976 temperature anomaly is -0.226k, with the difference between the two computed as 1.04k. Whilst this is useful in providing an understanding of the difference in global temperature anomalies between the 2022 and 1976 summer months, it is more suitable to look at the land temperatures only. This provides a more accurate representation of what people would experience during a heatwave as the values are not skewed by temperatures over the sea.

It is possible to quantify the temperature anomalies over land areas only by using the ERA5 land-sea mask file and masking out the areas over the ocean. This is achieved by loading in the data using xarray’s ‘load’ operator and then using the ‘where’ operator to mask out values below a certain threshold. In this case we mask out the temperature values in our anomaly dataset against values less than 0.5 in the ‘lsm’ field. This provides the land temperature anomalies which have been added to our previous timeseries and is shown on **Figure 4** below.

**Chart, histogram

Description automatically generated**

Figure 4: Timeseries graph showing the weighted Global Mean Anomalies compared with the weighted mean over land areas only.

By taking the land temperature anomalies and plotting them against the global temperature anomalies, there is a greater variance away from the mean for most years. The land temperature anomalies in years 1976 (-0.425k) and 2022 (1.336k) provide a difference of 1.761k.

It is also possible to calculate the temperature anomalies for each country. To do this, cultural vector shapefiles were loaded in from Natural Earth[[1]](#footnote-1) using the geopandas package, converted it into a GeoDataFrame (GDF) using the geopandas ‘readfile’ method and inspected the data. There are three resolutions available for the cultural vector regional shapefiles – 1:10m, 1:50m and 1:100m. There is further detail in the notebook on each of these. For this analysis, the 1:10m resolution GDF was converted into a three dimensional country mask using the regionmask package and using two fields for the names and abbreviations for each region (“ADMIN” and “ADMO\_3”, respectively). By passing the anomaly ERA5 dataset’s latitude and longitude dimensions to the package, a 3D mask is then created on a global grid with 195 regions. Using this region mask, the temperature anomaly for each region can be calculated by passing the latitude and longitude to the mean function and applying the weights calculated previously. The weighted 2-metre temperature anomalies for the United Kingdom, Spain and Italy can be seen on the graph in **Figure 5** below.

**Chart, line chart, histogram

Description automatically generated**

Figure 5: Timeseries graph showing the JJA anomalies for selected countries for years 1959 to 2022. Data is based on the processed ERA5 anomaly dataset and calculated regional anomaly means.

Using this timeseries data, a function can be defined to extract the ten warmest summers for any country name that is passed to it. When passing ‘Italy’ to the function, the ten warmest summers in ascending order are 1994, 2018, 1998, 2021, 2015, 2019, 2012, 2017, 2022 and 2003, which can be verified by checking against **Figure 5**. To gain further context, it would be beneficial to compare the global and regional temperature anomalies to areas of high population density. This can be achieved by once again using a mask and defining a population threshold based on the GPWv4 dataset.

The GPWv4 dataset can be loaded into the notebook and the population count for 2020 selected by choosing the required index. In this case, the dataset contains 20 raster layers, with the fifth raster containing the count needed, so index [4] is selected (as python uses zero indexing). The population data variable can be renamed using xarrasy’s ‘rename’ operator. In this case, the variable was renamed from 'Population Count, v4.11 (2000, 2005, 2010, 2015, 2020): 30 arc-minutes' to something more practical, such as ‘population\_2020’. This makes it easier to select and analyse the data variable. Following this, a population threshold of 50,000 persons is specified and regions with a population count lower than the threshold are masked out using the ‘where’ operator. By using this threshold, we are using 96% of the global population count for the analysis.

A function was created to extract the three warmest summers and coolest summer of our masked temperature field, on a global scale. The three warmest summers based on 2-metre temperature anomalies in regions with over 50,000 persons, in ascending order, are 2022, 2021 and 2019. The coolest summer was in 1976. These years have been plotted on **Figure 6** with the main focus being on Europe. It should be noted that an additional package – Cartopy – was imported in order to add the Borders feature to the figure.

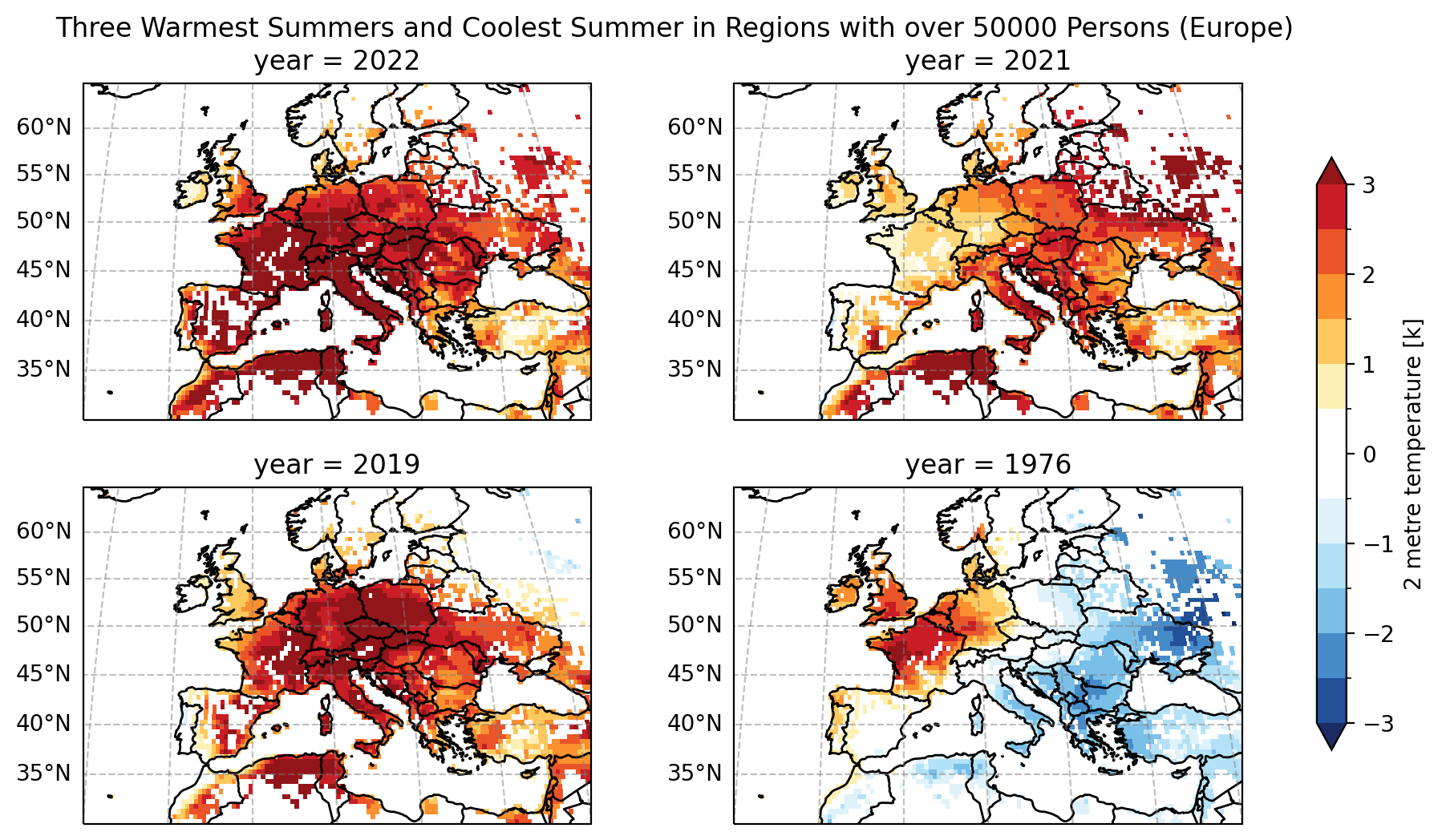


Figure 6: 2 metre air temperature anomalies for the months June to August in Europe, for the three warmest recorded years and coolest year. Areas with a population count of less than 50,000 are masked. The data is from the GPWv4 2020 population count.

Another way in which the global population data can be used in conjunction with the temperature anomaly dataset is to calculate the number of people exposed to a certain temperature anomaly. **Figure 7** shows the number of people that experience more than at least 1, 2, 3 and 4 kelvin temperature thresholds.

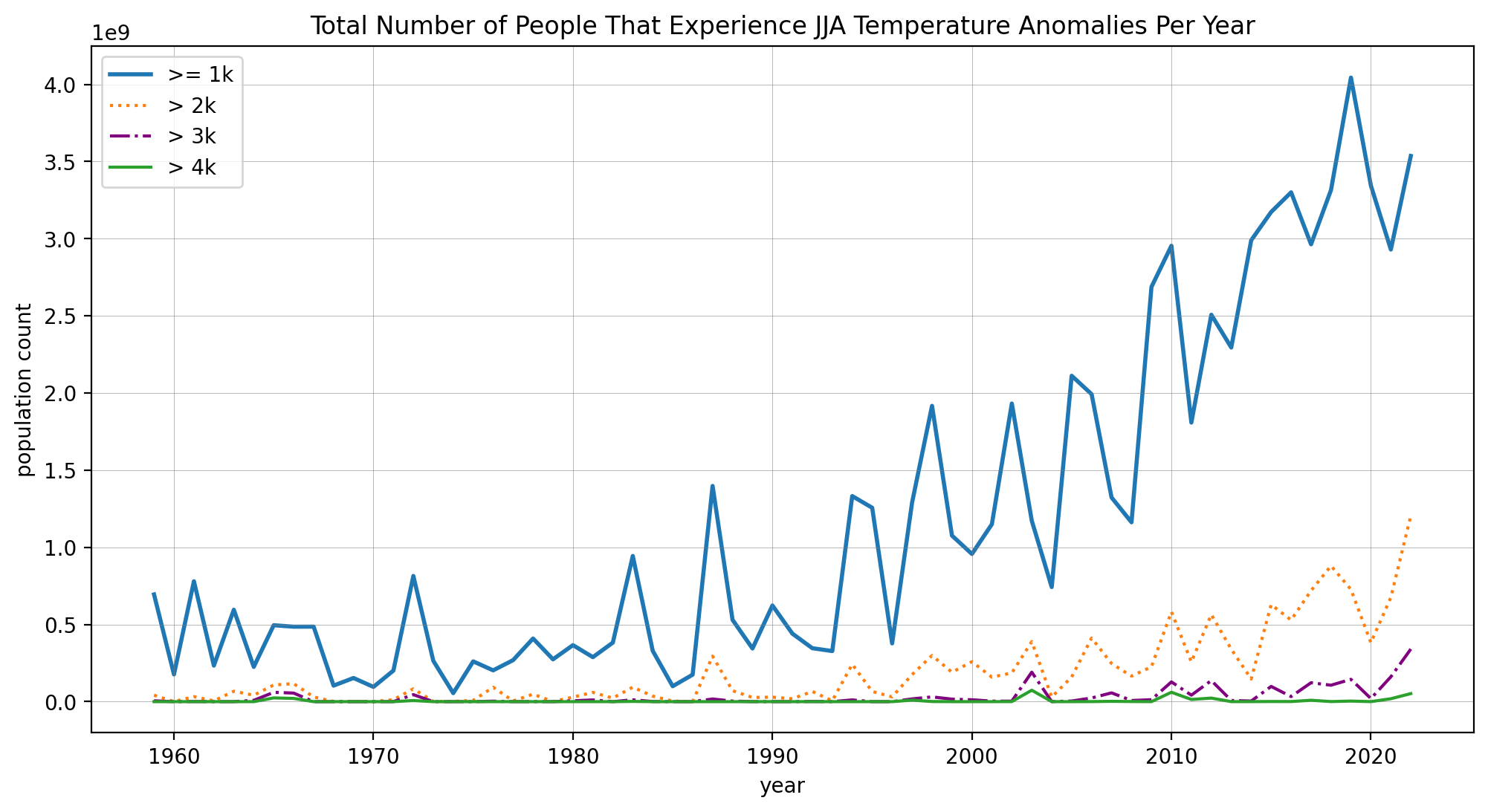


Figure 7: The number of people exposed to JJA Temperature Anomaly thresholds, between 1959 and 2022, in areas with a population count of over 50,000.

As is shown on **Figure 7**, the number of people that experience at least a 1k temperature anomaly was over 4 billion in 2019. Although the number drops off quite rapidly as the temperature threshold increases, there is a clear trend that more people are experiencing increasing heat extremes as time progresses.

By combining code used to create the above figures and undertake the analysis, an infographic can be produced to compare temperature anomalies in a chosen country against the global mean. An example of this is shown in **Figure 8** below.

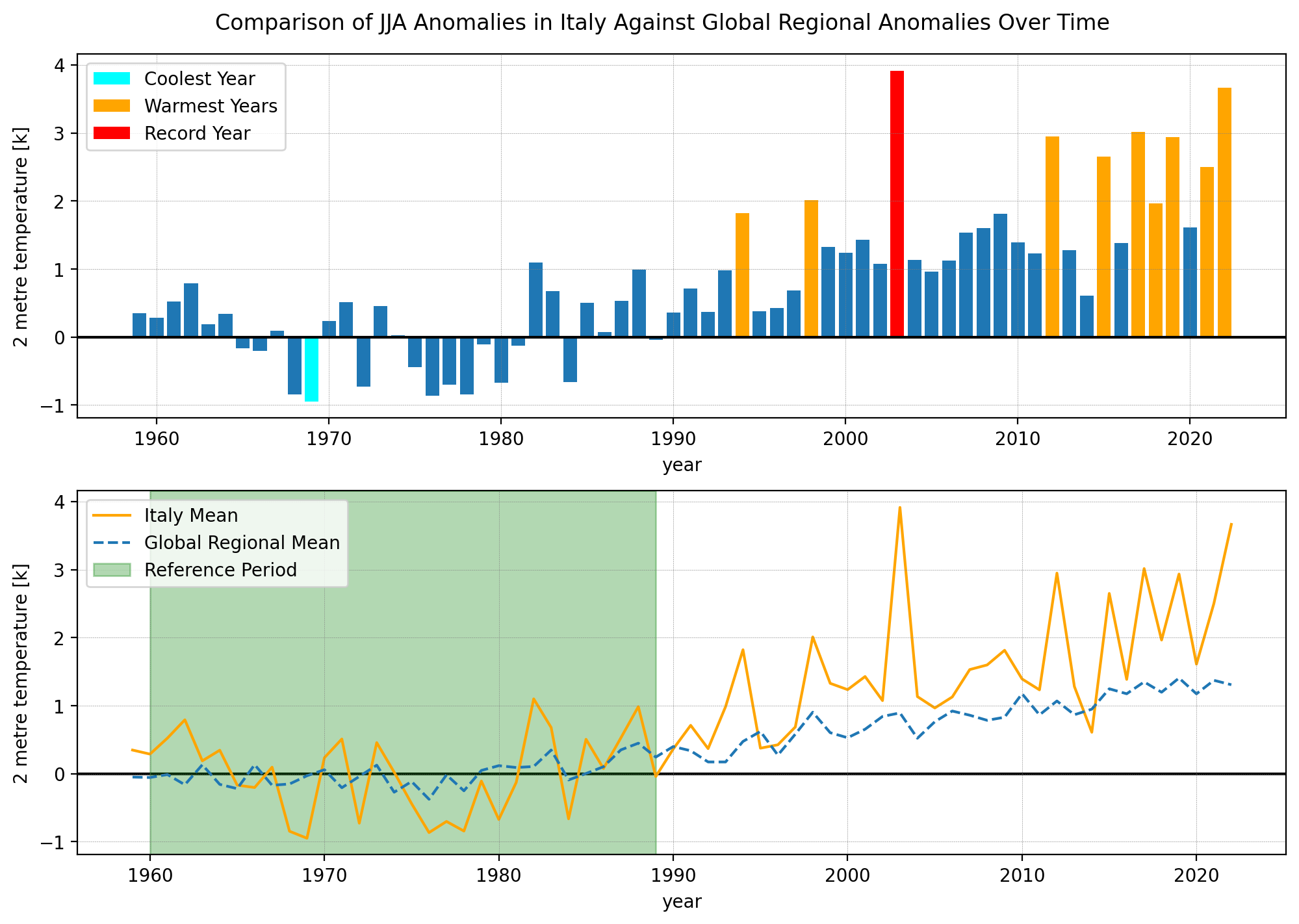


Figure 8: Example infographic for Italy showing (top) the evolution of historic summer temperature anomalies averaged across Italy and (bottom) a comparison of the summer temperature anomalies for Italy against global temperature anomalies.

The infographic takes a country name as input from the user and generates two figures for that chosen country. One weakness of the code here is that if a country name is misspelled an error is raised. This has been somewhat accounted for by accepting country names with a lower case first letter. If a country name is completely misspelled, or not contained within the masked dataset, the error message displays a custom message which lists available country names and their correct spelling.

Word Count: 1978

# Bibliography

Kendon, M., 2022. *Unprecedented extreme heatwave, July 2022.* [Online]   
Available at: https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/weather/learn-about/uk-past-events/interesting/2022/2022\_03\_july\_heatwave.pdf  
[Accessed 19 November 2022].

Rousi, E. et al., 2022. Accelerated western European heatwave trends linked to more-persistent double jets over Eurasia. *Nature Communications,* 13(1), pp. 1-11.

1. All versions of the Natural Earth vector data are in the public domain and have been downloaded from https://www.naturalearthdata.com/downloads/ in accordance with the terms of use. (https://www.naturalearthdata.com/about/termsof-use/). [↑](#footnote-ref-1)