

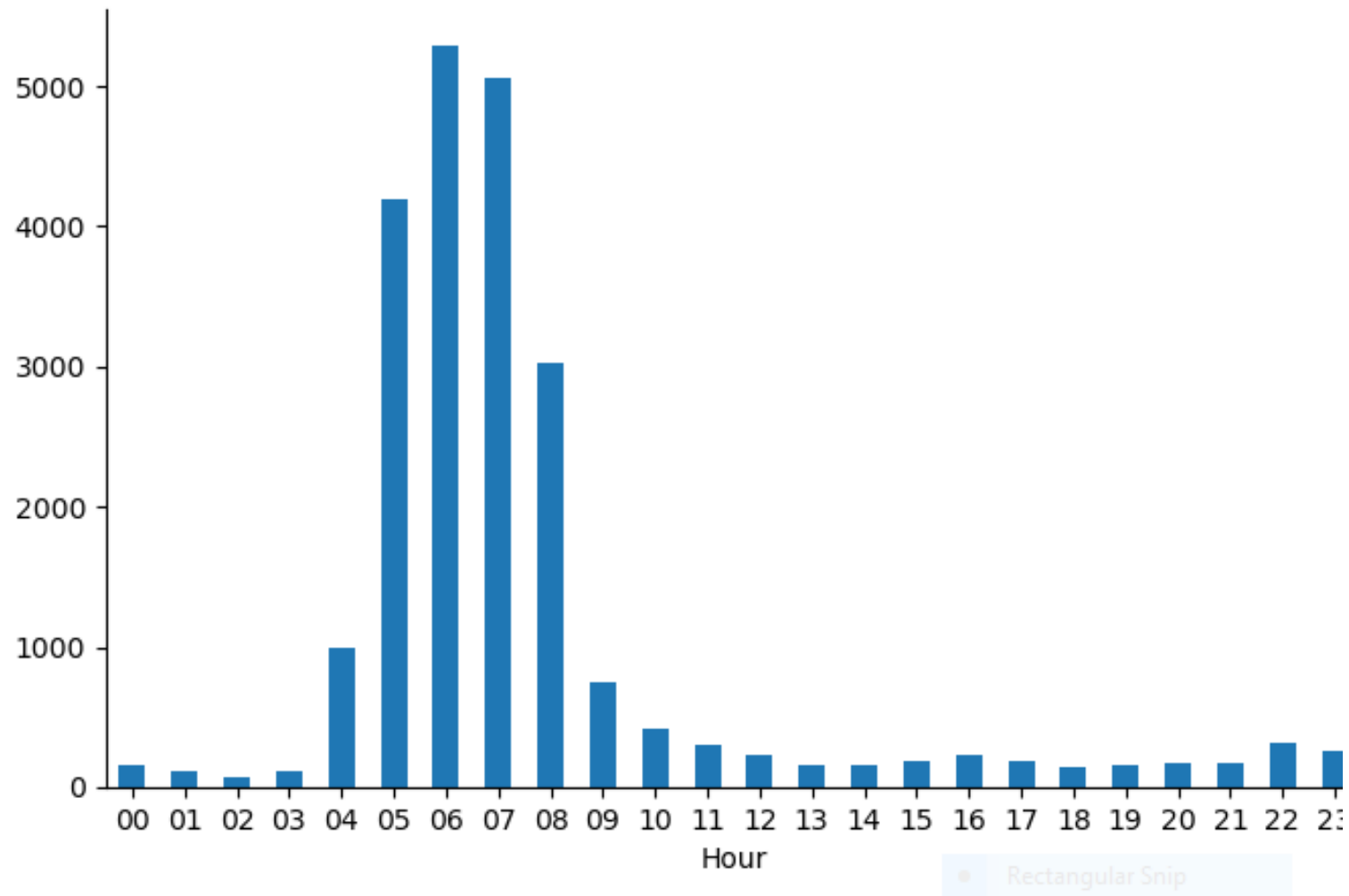
# Applied Data Science Capstone Project

# Introduction

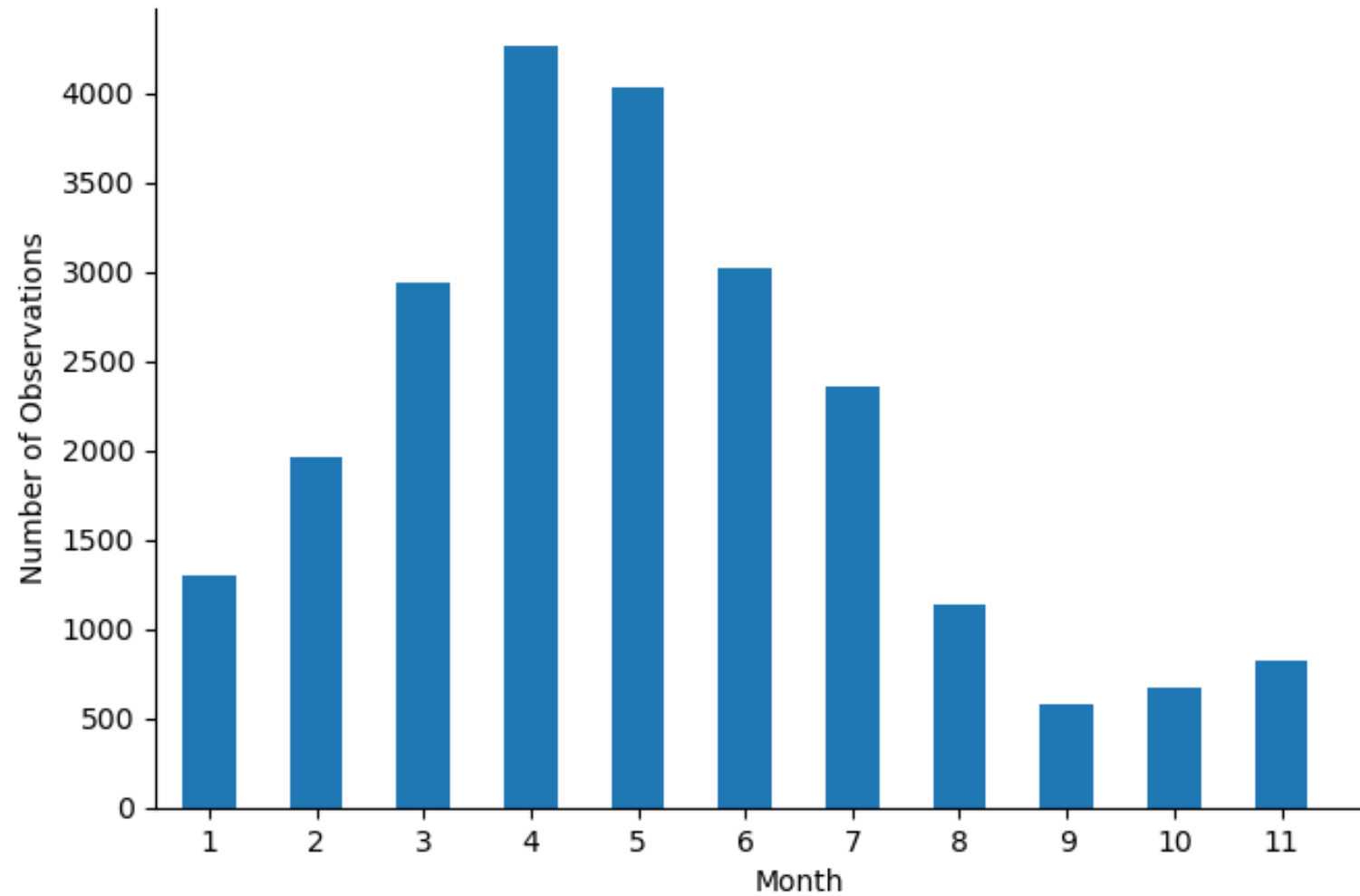
- The gain of this exercise will be to assist birdwatchers' to plan their birdwatching trips in England and Wales to see as many different species as possible by choosing the trip road through the maximum different counties.
- Another goal is to provide an overview of whether the species pattern changes with connection with the increase in the number of observations.

# Data

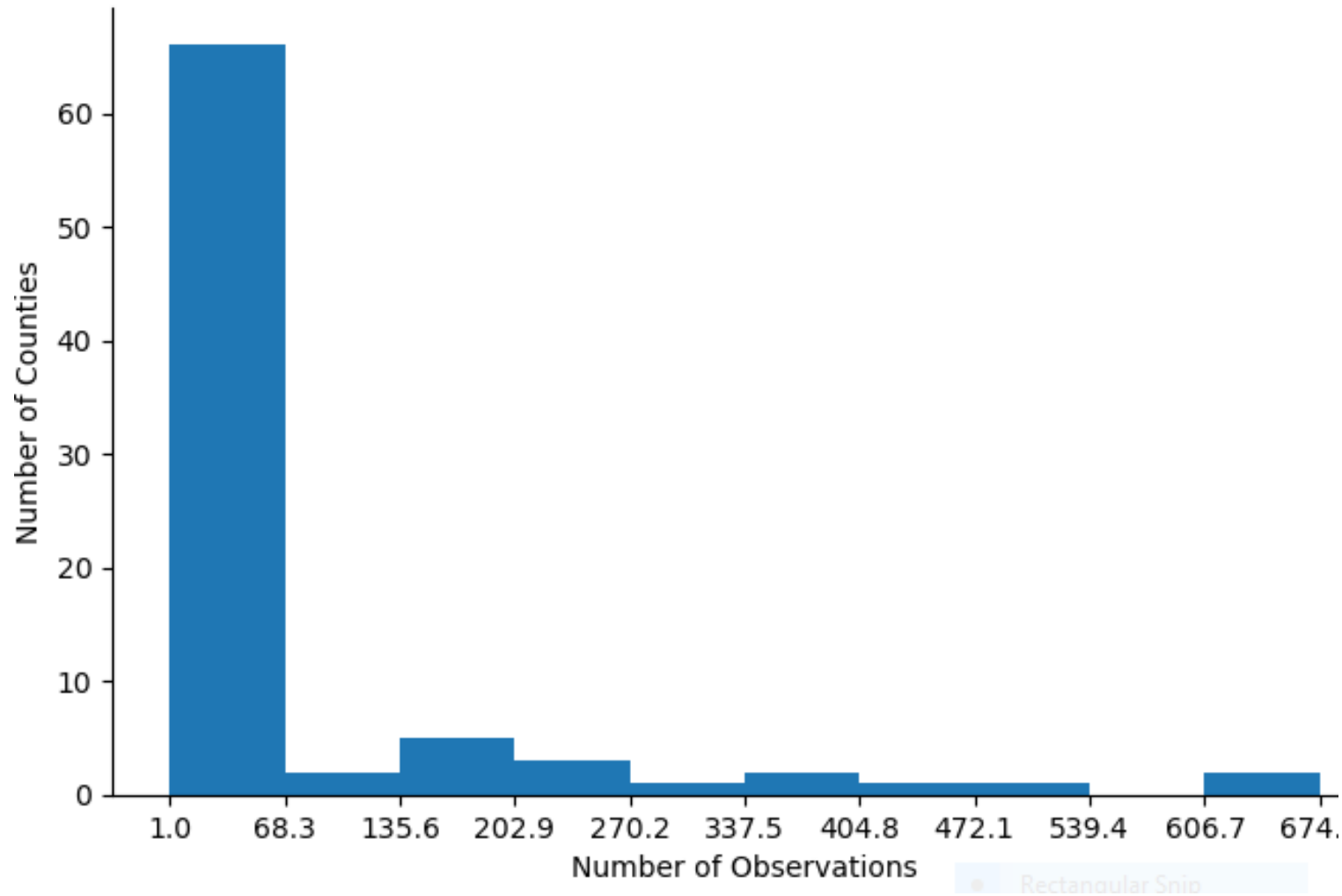
- Instead of Foursquare, this work relies on very similar the xeno-canto API.
- A geographic information system ArcGIS is also used – to retrieve England and Wales counties names that correspond the coordinates retrieved from the xeno-canto.
- For marking counties borders, the GeoJSON file from the web page <https://data.gov.uk/dataset/d6f97a1a-25dc-485c-9af3-0e5681465d77/counties-and-unitary-authorities-december-2016-full-clipped-boundaries-in-england-and-wales> have been used.
- After removing missed values a final sample remained for analysis consist of 23782 observations.



DAILY  
DISTRIBUTION  
OF  
OBSERVATIONS  
(RECORDINGS)

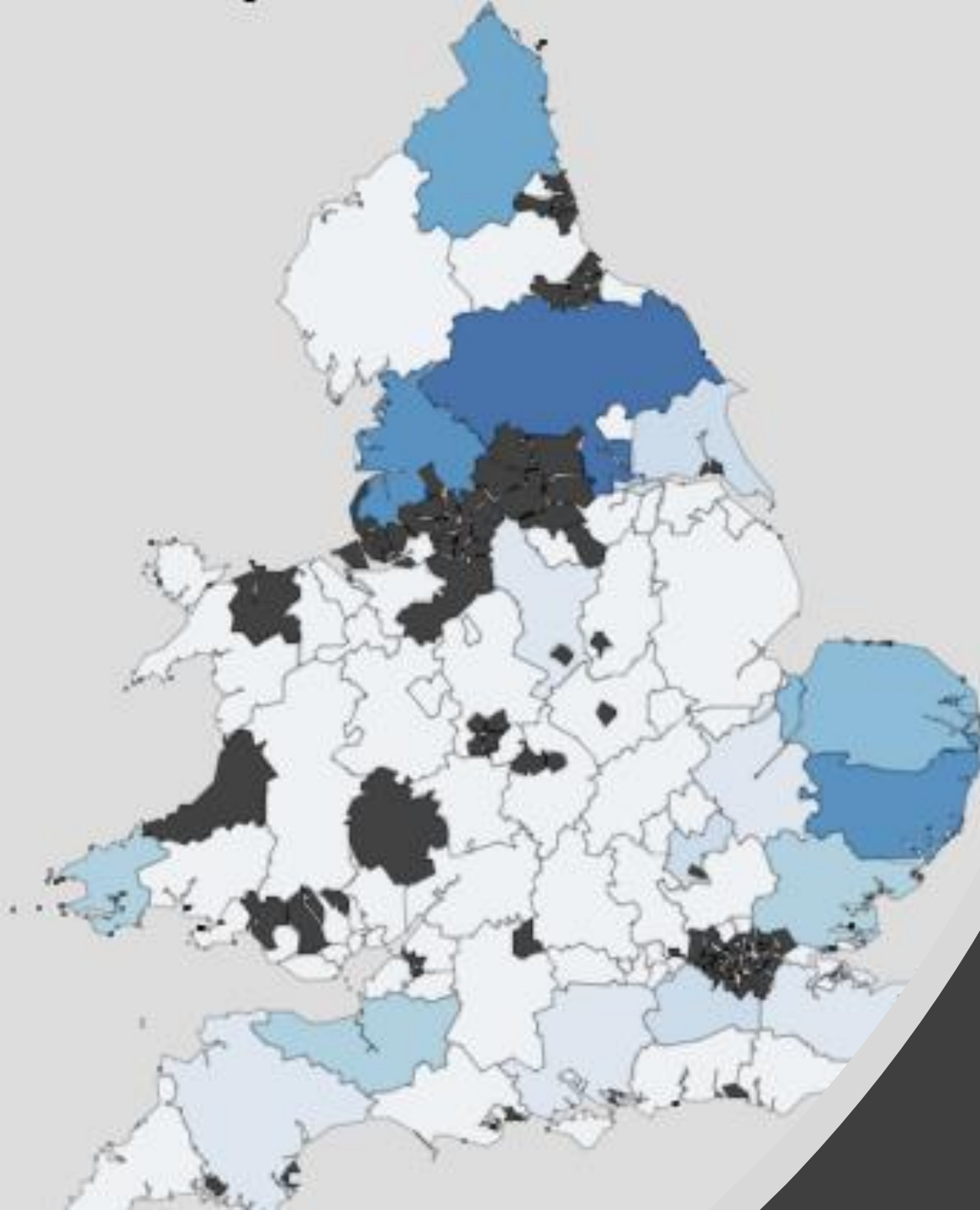


YEARLY  
DISTRIBUTION  
OF  
OBSERVATIONS  
(RECORDINGS)



## HISTOGRAM OF OBSERVATIONS (RECORDINGS) PER ENGLAND AND WALES COUNTIES.

NORTH YORKSHIRE COUNTY ARE EXCLUDED FROM HISTOGRAM AS IT HAVE 17781 OBSERVATIONS OUT OF 23782 (TOTAL NUMBER) AND SO SIGNIFICANTLY OUT OF RANGE.

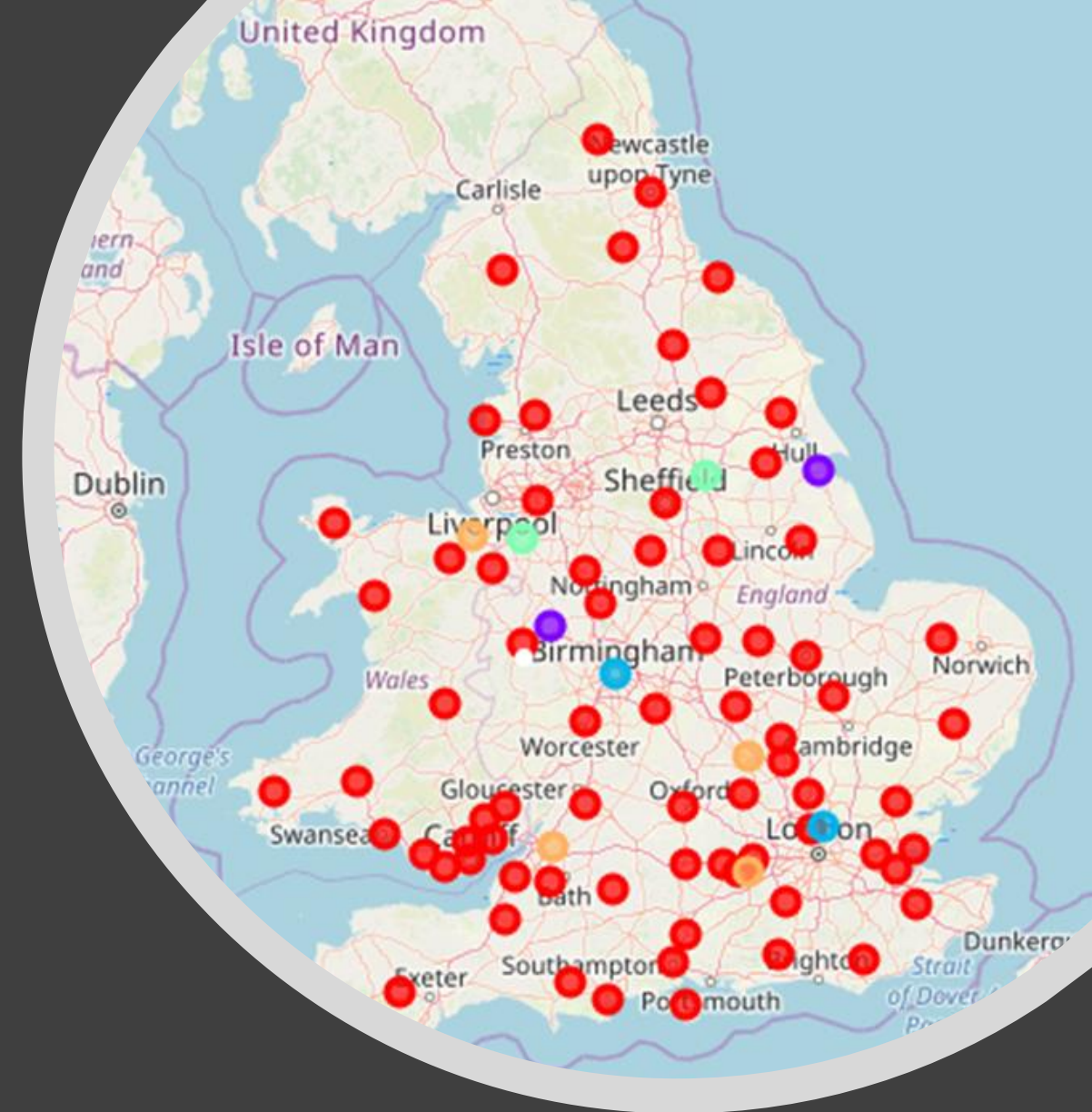


GEOGRAPHICAL DISTRIBUTION OF  
OBSERVATIONS (RECORDINGS)  
PER ENGLAND AND WALES  
COUNTIES.

DARKER BLUE MARKED MORE  
RECORDINGS PER COUNTY. BLACK  
MARKED COUNTY WITH MISSING  
OBSERVATIONS.

## GEOGRAPHICAL PRESENTATION THE ENGLAND AND WALES COUNTIES DIVIDED ON FIVE CLUSTERS ON THE BASIS OF OBSERVED (RECORDED) BIRD SPECIES PATTERN.

- There is no remarkable distinction between the counties (in recorded bird pattern), based on geographical location doesn't emerged.
- No remarkable correlations between the number of recording per county and the observed bird pattern doesn't also emerged.
- 74 of the 84 counties (where we have a data) belonged to the first cluster.
- The counties which belonged to the second or fifth cluster having only one observation (recording) per county. Therefore, it is possible to mark them as the outliers. A similar assessment is also valid for the third and fourth cluster, where is only two or three observations (recording) per county.
- If to look the cluster species pattern, then all clusters described similarly by birds which are broadly rather common to cultural landscape.
- Reason of that kind results is probably quite uneven distribution of the observation per counties.





# Discussion

- The main result of current exercise is that despite representative overall number of recordings in xeno-canto database according England and Wales, the distribution of these recordings on counties level is too uneven to analyse difference between those counties.
- Therefore, the main recommendation emerged of the current exercise is to use in further analyses about difference between England and Wales (or as the United Kingdom as whole) counties birds species pattern other public science open databases (which is also possible to find).
- The other recommendation is if to limit this exercise only in the counties level, it can be restrictive. If it possible to find a geoJSON file which describe rather borders between natural/ecological landscapes, it will be probably more useful.