Ancient Tree Inventory

Project Lifecycle

Source
Identify opensource datasets

Details

Dataset
Feature Layer

21 August 2023
Info Updated

21 August 2023
Data Updated

1 March 2023
Published Date

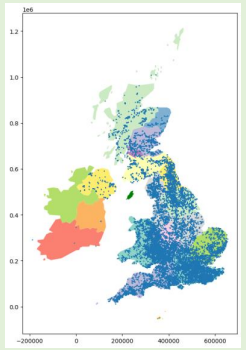
Records: 79,933
View data table

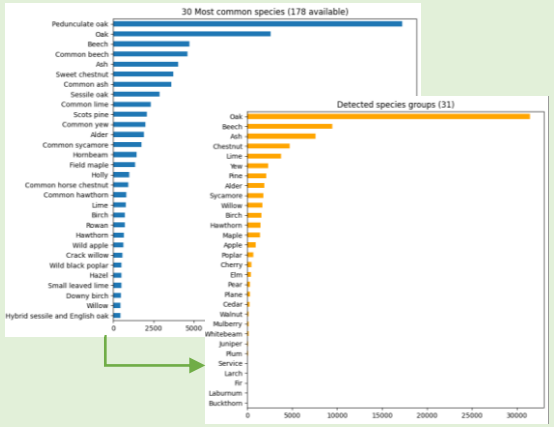
Public
Anyone can see this content

Custom License
View license details

Raw dataset attributes at source

EDA
Explore data quality, granularity, hierarchies & patterns





Including trials for different geospatial mapping approaches and study of species patterns and groupings

ETL
Design target tables, build cleaning / transformation pipeline

```python
def groupSpecies(rowstring, grouplist):
    # if available, take intersecting term, if not available, take species raw, if
    intersect=[word for word in rowstring.lower().split(' ') if word in grouplist]
    if len(intersect)>1:
        intersect=[intersect[-1]]
    elif len(intersect)==0:
        intersect=[rowstring]
    return str(intersect[0]).capitalize()


def createBoolFlag(data, fields):
    for each in fields:
        colname=each+'_flag'
        data[colname]=np.where(data[each].isnull(), 0, 1)
    print ('Boolean flags generated for '+', '.join(markerDict.keys()))


def livingStatusFlags(row):
    testcase=row.LivingStatus.lower()
    if 'alive' in testcase:
        row['LivingGroup']='Alive'
    elif 'dead' in testcase:
        row['LivingGroup']='Dead'
    else:
        row['LivingGroup']='Unknown'

    if 'chalara fraxinea' in testcase:
        if 'confirmed' in testcase:
            row['AshDieback']='Confirmed'
        else:
            row['AshDieback']='Suspected'
    elif 'Ash' in row.species_group:
        row['AshDieback']='Unknown'
    else:
```

Including setup of modular python file structures and functions

Visualise
Build Tableau charts & interactive narrative

Showing front page overview focussed on species and regional patterns, with user interactivity to study alternative metrics

tabular18 helen.k8608 info@helenkinneydata.com

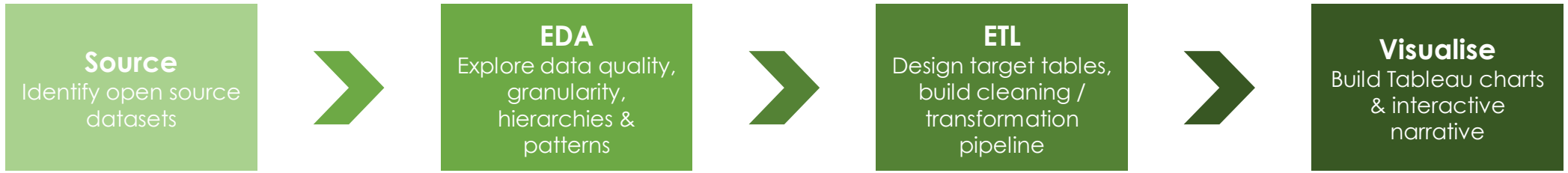
Ancient Tree Inventory

Mission Statement & Context

The Ancient Tree Inventory is a citizen science dataset project managed and owned by the Woodland Trust, aiming to collect information on the oldest and most important trees in the UK & Ireland region.

What can we learn from this dataset? My aim is to compile interactive visualisations which address:

- ***Rate of recording*** – can we identify areas where perhaps awareness of the inventory could be improved, to encourage further recording?
- ***Species patterns*** – do certain species favour certain areas, or show clusters in certain habitats? Do we see more ancient specimen within one species or another?
- ***Accessibility*** – do we see patterns in accessibility of our oldest trees geographically, be it on private/public land or in given surroundings?
- ***Symbiotic/other relationships*** – can we infer which species favour or create favourable conditions for other organisms such as epiphytes or fungi? Do we see an impact to the growth of those trees which are associated with other organisms?



tabular18 helen.k8608 info@helenkinneydata.com

Ancient Tree Inventory

Exploratory Data Analysis

Source

Data was sourced from https://opendata-woodlandtrust.hub.arcgis.com/datasets. This data is available free-of-charge for non-commercial use, and in this instance data in the Tableau backend is reproduced with the permission of The Woodland Trust. The data was accessed in September 2023. and a live connection will not be used for this project.

Initial dataset analysis

Granularity

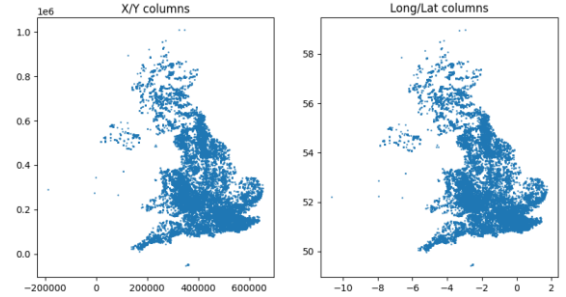
- Data was found to be at the individual-tree granularity, with two primary key / unique identifiers available per tree (one being an index, ObjectId, the other a unique number, Id). No logical/contextual information behind Id was found, therefore these two are treated as redundant.

Duplicates

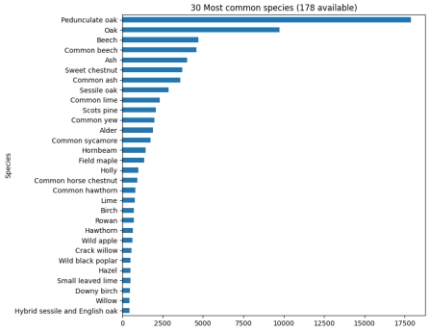
- It was found that multiple records could have identical latitude/longitude
  - In some cases this is evidently dense woodland/orchard areas – e.g. a case with 45 apple/pear trees at the same location with various height/girth
  - In other cases, all tree attributes appear to be identical. Scenarios could exist in which two neighbouring related trees (e.g. from same parent tree) are recorded at the same time, or indeed the same tree may be revisited or verified twice.
- Given the fact that data has been manually verified and maintained, for the purposes of this project I will assume that any apparent duplicates are in fact separate trees, and assume that the data is maintained in a Type-1 fashion (e.g. history is overwritten rather than new columns/rows added). In a client scenario, this detail would be verified during requirements gathering sessions!

Data Quality / Opportunity

- Latitude and Longitude are fully populated, but other geographic information has lower coverage and low consistency.
- Species appears to be free-form text entry, so standardisation could improve this, along with higher level grouping of trees within the same umbrella species (e.g. 'Oak', 'Chestnut'). We will opt to use a data-driven rather than manual grouping approach, meaning for instance there may be some groups which are missed (an example being the grouping of Maidenhair Tree and Gingko Biloba – which I know anecdotally is the same, however without scraping lookups it is not possible to combine these in a data-driven fashion).
- X and Y is a different CRS (coordinate reference system) projection of Lat and Long – we will keep these as they are useful for polygon mapping.
- A number of fields can have multiple concatenated values (e.g. multi-tick box) from a pre-defined list. These will go into a separate marker table to enable full analysis
- There are some null dates and these are in US format – null handling and transformation required here.




```
LivingStatus
Alive                          75800
Dead                            3484
Chalara fraxinea suspected       218
Felled                           135
Unknown                          128
Chalara fraxinea confirmed        71
Collapsed                         57
Other                             22
Test living status                11
Monitor AOD/COD                    7
Name: count, dtype: int64
```




```
Common words in the species listed

array(['oak', 'beech', 'common', 'cedar', 'white', 'lime', 'walnut',
       'wych', 'ash', 'sessile', 'alder', 'hawthorn', 'willow', 'larch',
       'wild', 'silver', 'elm', 'large', 'hybrid', 'monterey', 'european',
       'poplar', 'cherry', 'black', 'service', 'apple', 'red', 'indian',
       'juniper', 'caucasian', 'mulberry', 'lombardy', 'english', 'birch',
       'grey', 'weeping', 'monkey', 'sycamore', 'maple', 'norway',
       'purple', 'chestnut', 'pear', 'plane', 'cypress', 'plum', 'yew',
       'horse', 'laburnum', 'pine', 'tree', 'golden', 'whitebeam', 'of',
       'leaved', 'or', 'fir', 'puzzle', 'buckthorn', 'and', 'var'],
      dtype=object)

Following review, the below terms are appropriate for an initial higher-level grouping of species names.
Note this is NOT based on taxonomy, but is a more informal grouping of the tree types listed in the inventory

array(['oak', 'beech', 'cedar', 'lime', 'walnut', 'ash', 'alder',
       'hawthorn', 'willow', 'larch', 'elm', 'poplar', 'cherry',
       'service', 'apple', 'juniper', 'mulberry', 'birch', 'sycamore',
       'maple', 'chestnut', 'pear', 'plane', 'cypress', 'plum', 'yew',
       'laburnum', 'pine', 'whitebeam', 'fir', 'buckthorn'], dtype=object)
```


tabular18 helen.k8608 info@helenkinneydata.com

Ancient Tree Inventory

Geospatial Handling Trials

From first principles, we know that some form of geospatial analysis will be useful. The relevant fields available are: Latitude, Longitude, Town, County, Country (grid ref is available but more useful to on-the-ground consumers).
The first two are directly of use to map individual trees and perhaps density mapping in Tableau. To create roll-up aggregates however, the categorical fields will be useful.

Town – highly variable level of detail – sometimes a city, sometimes a hamlet – therefore not pursued for analysis
County – wellpopulated, but includes a lot of abbreviations / misspellings / entities not recognised as counties by vis softwares. This would be a useful level of detail to aggregate by to see regional variations
Country – incorrect level of detail in some cases so needs to be standardised. Best to have one very high level (UK/RoI) and one lower level (Scotland/Wales)

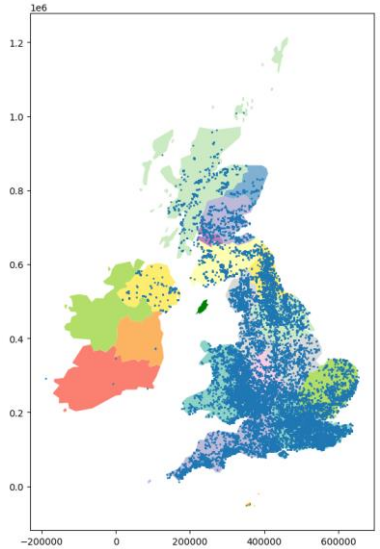
Important to note is the dependency on end visualisation software – for instance, Tableau uses OpenStreetMap, therefore the auto-detected level of detail available is NUTs codes (Europe-wide) or Country (UK).

Trialled approaches:

Reverse geo-coding (GeoPy (Nominatim or OpenCage locators), reverse_geocoder
This approach uses Lat/Long as input and returns a location json-like response. The attributes available in the response differ based on the location, but county is usually available. However, output still did not fully align with Tableau autodetection, therefore this would require some lookup procees to match to a given list. In addition, when covering a large number of records this becomes computationally expensive, or APIs may have limits on number of requests per minute/day.

Text lookup mapping
This would allow standardisation of the counties, for instance used to map from commonly misspelled or noncompliant names. However, the greated difficulty is identifying one confluent list of counties which covers all ATI areas and aligns with visualisation softwares. This is not available via OpenStreetmap (used by Tableau), although some county names align with NUTs regions which are used in Tableau. Considerable research into structured or non-structured data was performed, including trials using BeautifulSoup to scrape from Wikipedia and other sites. Unfortunately this did not align with Tableau visualisation.

Shape file in-polygon checks (Geopandas)
This approach enables us to specify our own geometry regions and classify every record based on the geometry it sits within. This allows us to ensure that we have the level of detail we want and to ensure full coverage. To ensure this aligns with Tableau visualisation we will use NUTs which is a Europe-wide system with multiple layers. We will use level 2 for Ireland and Scotland and level 1 for the remainder of the UK. In addition, we will add single shapes corresponding to the Isle of Man and Guernsey, though these do not have NUTs classifications (but are recognised as counties in Tableau).
This final approach is used alongside some text processing/lookup mapping to ensure the correct country is available.
Polygon sources used: UK&I, Isle of Man, Guernsey



Colors showing each of the NUTs / other regions with ATI lat/long overlayed. CRS shown is EPSG:27700

tabular18 helen.k8608 info@helenkinneydata.com

Ancient Tree Inventory

Target Table Design

Design Factors

- The raw data has multiple columns which contain concatenated lists in text form – some of which appears to be free-form input, therefore non standardised text. These columns are unwieldy for the viewer and not suitable for calculations.
- As a result, we treat each column as a marker type – e.g. *Surroundings* and *Protection*. For each marker type, a tree can have any number of marker values – e.g. *Field* or *Churchyard* for Surroundings, *Fencing* or *Conservation Area* for Protection.
- To be most suitable for both storage and visualisation purposes, this marker information will be in the form of an unpivoted table which is long and narrow – with one row per marker value, per marker type, per tree.
- The remainder of information will be retained in a base record table form, with one row per tree, in the form of a Fact table

Base Record/Fact Table
Granularity: one row per tree (ID)
34 fields, <100k records

ID	Height	Girth	Survey Date	…

ID

Marker Table
Granularity: one row per marker value, per marker type, per tree (ID)
3 fields, <500k records

ID	Marker Type	Marker Value

Each row contains the base attributes for a tree, with higher-level grouping columns added for those which are irregular or highly detailed – e.g. species and living status. Concatenated marker columns (e.g. Fungus) are converted to binary flags indicating whether the tree has any entries for that marker type.

Each row represents one of the listed values within one of the designated marker columns. These are: Surroundings, Condition, Protection, SpecialStatus, Fungus, Epiphyte.
Trees with no listed markers do not exist in this table

tabular18 helen.k8608 info@helenkinneydata.com

Ancient Tree Inventory

Target Tables - ETL

Transformation stages

1. Creating Boolean Marker flags
2. [*First run only*] Generating customised shapefile with target Regions
3. Assigning regional polygons per Long/Lat
4. Null handling
5. Type conversion
6. Datetime formatting / replacement values
7. Creating Species groups
8. Creating grouped fields for Living Status, Public Accessibility, and Ash Dieback
9. Splitting out marker text lists, expanding across columns, and pivoting to long form
10. Correcting text in markers (pre-identified symbol replacements)
11. Archiving previous datasets
12. Saving Base and Marker tables to output location

BASE RECORD TABLE

*Stored as **ATI_Base_table_DD-MM-YYYY_HHMM***

Granularity: one row per tree (Id)
34 fields, <100k records

Column Number	Column Name	DataType	DataType (Python)	Description/Notes	Can be null?	Null Handling
0	**OBJECTID**	Integer	int64	Unique per Id, useful as contiunous Index field	N	N/A
1	**Id**	Integer	int64	**Primary Key**	N	N/A
2	**SurveyDate**	DateTime	object	Stored as datetime, visualised as date	N	"12/31/9999 12:00:00 AM"
3	**VerifiedDate**	DateTime	object	Stored as datetime, visualised as date	N	"12/31/9999 12:00:00 AM"
4	**MeasuredGirth**	Decimal (5, 2)	float64	Measured girth (circumference) of tree in meters	Y	N/A
5	**MeasuredHeight**	Decimal (5, 2)	float64	Measured height of tree in meters	Y	N/A
6	**EstimatedGirth**	Boolean	bool	Has girth been estimated (True) or measured (False)?	N	N/A
7	**Latitude**	Decimal (10, 8)	float64	Reported latitude, seen between 2 and 8 decimal places. Standardised (zero-padded) to 8dp	N	N/A
8	**Longitude**	Decimal (10, 8)	float64	Reported longitude, seen between	N	N/A

The data pipeline uses a stripped-back set of popular modules

- Throughout the data pipeline we treat the original data in DataFrame form (Pandas), and the transformations applied are either at the DataFrame or row (lambda function) level.
- For assigning regions based on polygon geometries, we use GeoPandas, which has *within* and *distance* functions.
- Regex used to extract/normalise country information
- No significant text preprocessing is required, instead using dictionaries to perform regex replacement of known symbol issues in the raw text. As a result, no nlp packages such as ntlk/spacy are imported.

MARKER TABLE

*Stored as **ATI_Marker_table_DD-MM-YYYY_HHMM***

Granularity: one row per marker value per marker type per tree (Id)
(e.g. one tree with two types of fungus listed and 5 types of protection listed, with no other markers, would have 7 total records. A tree with no markers will not exist in this

3 fields, <500k records

Column Number	Column Name	DataType	DataType (Python)	Description/Notes	Can be null?	Null Handling
0	Id	Integer	int64	**Primary Key**	N	N/A
1	MarkerType	String	object	Name of the marker column from which the value originated (e.g. Protection)	N	N/A
2	Markervalue	String	object	Individual marker values within the given marker type. These have been extracted from a concatenated list in the original	N	N/A

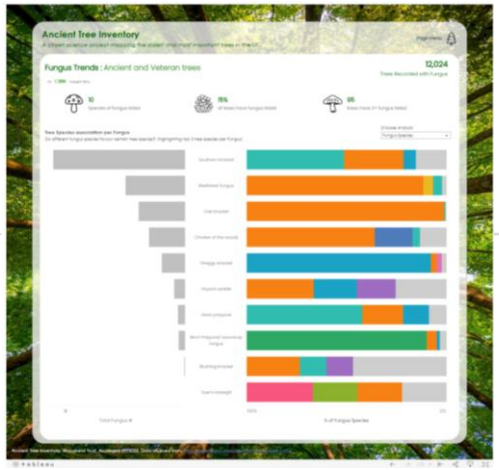
tabular18 helen.k8608 info@helenkinneydata.com

Ancient Tree Inventory

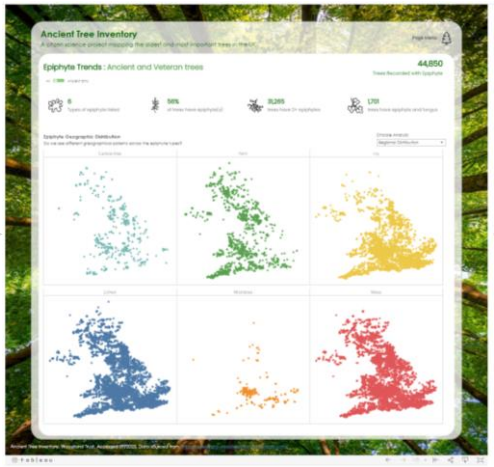
Tableau : User Journey

Overview Trends

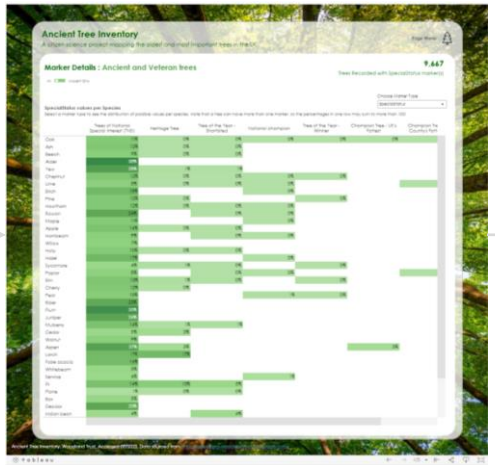
Deep-dive Trends: Fungus



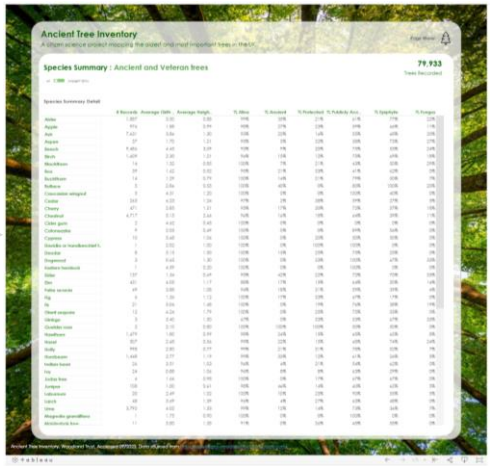
Deep-dive Trends: Epiphytes



Marker Detail Matrix



Species Summary Statistics



High level of detail — *Low level of detail*

tabular18 helen.k8608 info@helenkinneydata.com

Ancient Tree Inventory

Example Insights

Species Distribution

- Species counts are heavily skewed with a very long tail. Given 178 original non-standardised species names, 31 common names were identified and assigned to trees. The result was 67 species groups.
- The majority of records belong to a limited number of species groups - 40% of trees being Oak and the top 5 species (Oak, Beech, Chestnut, Ash, Lime) accounting for over 70% of records.
- Low number of trees considered 'ancient' – just under 17%. This aligns with the fact that ancient trees are by definition rare, and different trees have different criteria to being 'ancient'. Monitoring veteran trees allows us to better prepare and protect to ensure more live to an ancient age!

Regional

- We see coverage across not only the UK but also Republic of Ireland and Guernsey (albeit low numbers). The vast proportion of recorded trees are in the South of the UK
    - This may show that citizen awareness / project coverage needs to be improved, but it may also reflect the true variation of tree coverage
    - The Highland Clearance in Scotland led to loss of considerable forest cover, likely including older trees. In addition, trends in timber forestry (e.g. felling native trees or underplanting them with fast-growing species) and burning of gorse for game had additional impact on the ability of tree populations to regenerate
    - Patterns of long-term agriculture and grazing have had an impact on coverage in a number of regions – for instance, Northumberland national park
- Some tree species have very geo-specific recordings – for instance:
    - Aspen, which is clustered in the North of England and the Highlands of Scotland. This is consistent with the growing habits of this tree
    - Juniper, recordings of which are isolated to southern Scotland, Yorkshire and the Highlands
    - Pine, which favours the Highlands area of Scotland
    - Plane, which has many clustered records in London (as expected) and is more sparsely spread across the South of England
    - Yew, which is much more commonly reported in Wales, South West and South East England

Protection

- Only 20% of trees in the inventory have at least one form of Protection listed, and we see much more protection in the south of England. The greatest extent of protection is seen in the South West of England – with 40% of all trees in the 50-51° band being protected, and 47% respectively for Ancient trees.
- This is consistent for Ancient trees specifically too – with protection increasing from 5-12% in the 56° /57° band (Scotland) to 20-30% across the 51-53° band (South England). Of the protections listed, being on Uncultivated Land is the most dominant form (accounting for 86% of protected trees – note some trees have multiple protections).

tabular18 helen.k8608 info@helenkinneydata.com

Ancient Tree Inventory

Example Insights

Epiphytes

- 56% of trees recorded have epiphytes listed, and 40% have 2+ listed
- Whilst Mistletoe may be anecdotally known to impact apple trees, a larger portion of all recorded mistletoe affects Lime or Hawthorn. This epiphyte shows the most discernment regarding host tree species – other epiphytes show a tree species distribution which echoes that of the general ATI population (e.g. Oak, Beech and Ash being most common)
- Mistletoe also shows a distinctive regional distribution, with clusters in the West Midlands and bordering Wales/South West England
- Trees with epiphytes are likely to have a larger girth/diameter on average (22% larger, though this reduces to 8% when only tree species with at least 10 epiphyte records are included)
- Trees with epiphytes are likely to have lower height on average (18% smaller, increasing to 23% smaller when only tree species with at least 10 epiphyte records are included)

Fungus

- 15% of trees have fungus listed, increasing to 21% of ancient trees
- Different species of fungus show very different distribution of tree species – for instance:
    - Southern Bracket and Giant Polypore frequent Beech trees (49% and 58% respectively)
    - Shaggy Bracket is dominated by Ash (92%)
    - Chicken of the Woods is linked with Yew (19% - whilst Yew rarely features across other fungus species)
- Beefsteak and Chicken of the Woods are the most popular combination of fungi species on one host tree
- Southern Bracket (fungus), Moss (epiphyte) and Lichen (epiphyte) show frequent co-occurrence on the same host trees
- Trees with fungus are likely to have a slightly larger girth/diameter on average (3% larger, though this increases to 10% when only tree species with at least 10 fungus records are included)
- Trees with fungus are likely to have lower height on average (11% smaller, 14% smaller when only tree species with at least 10 fungus records are included)

tabular18 helen.k8608 info@helenkinneydata.com

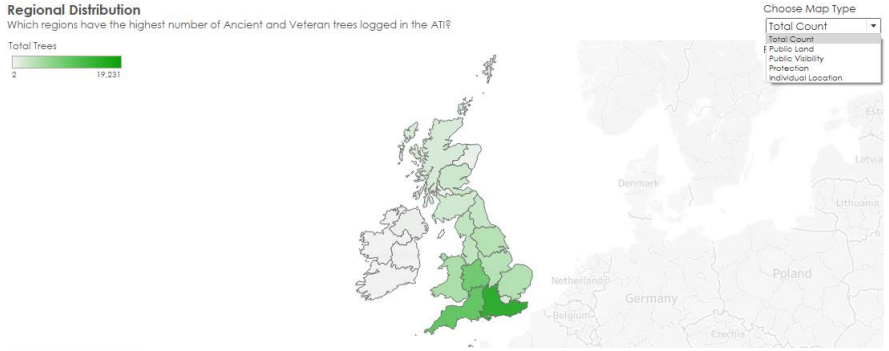
Ancient Tree Inventory

Tableau : Features

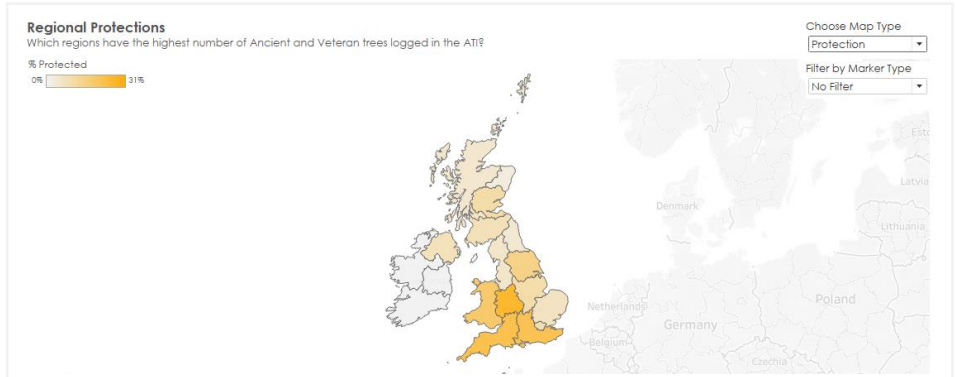
Overview Trends – Geographical Distribution Analysis

High level of detail

Low level of detail

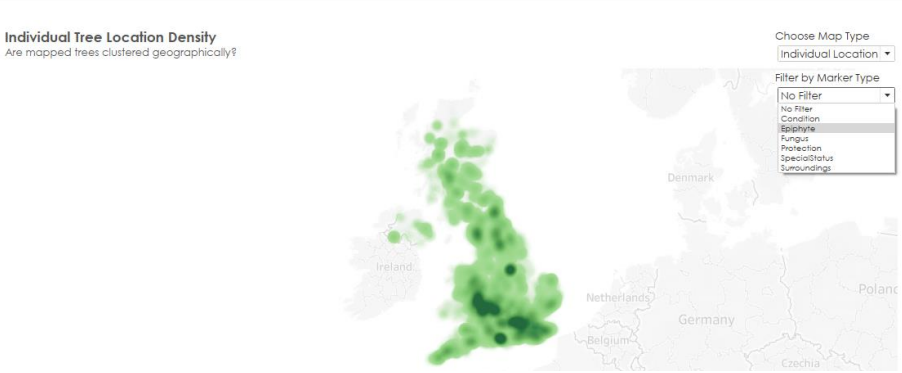


Default: Regional Distribution, Total Record Count

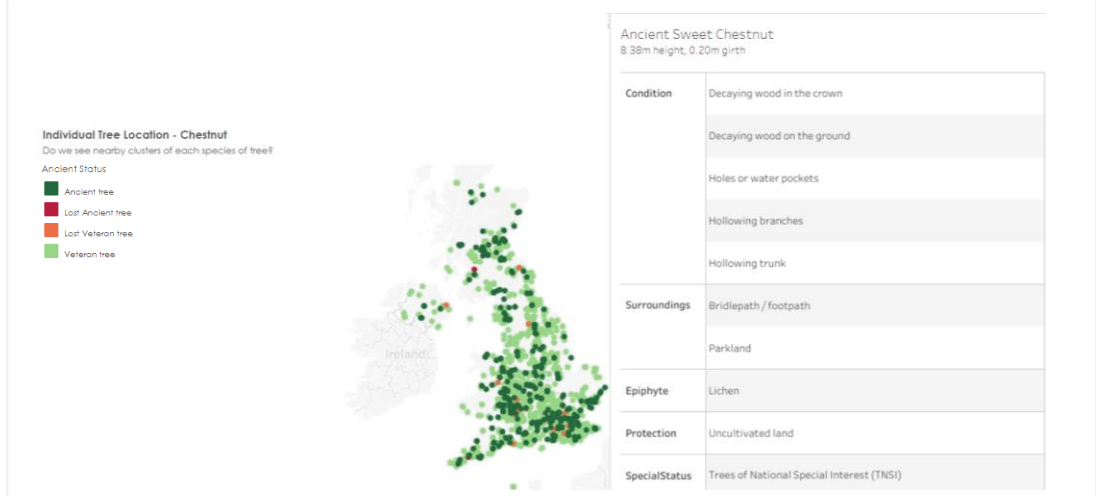


Regional Distribution, % Protected

Other regional metrics available:
% on Public Land
% Visible from Public land



Specific Location – Hotspots (record count)



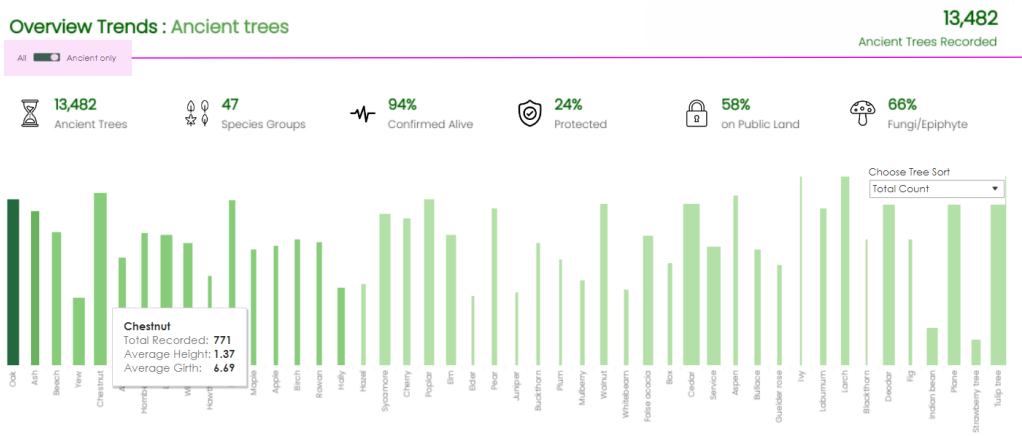
Specific Location – Per Tree (one species only)
With tooltip showing all markers per tree

tabular18 helen.k8608 info@helenkinneydata.com

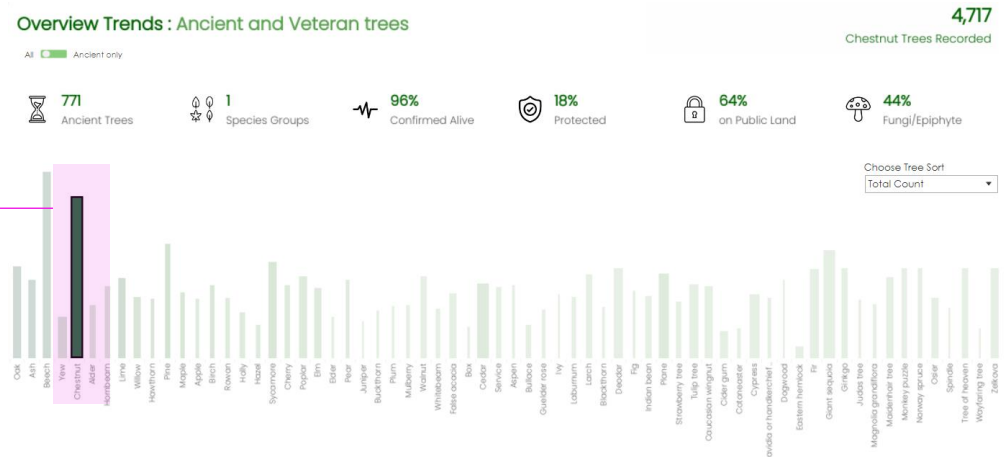
Ancient Tree Inventory

Tableau : Interactivity

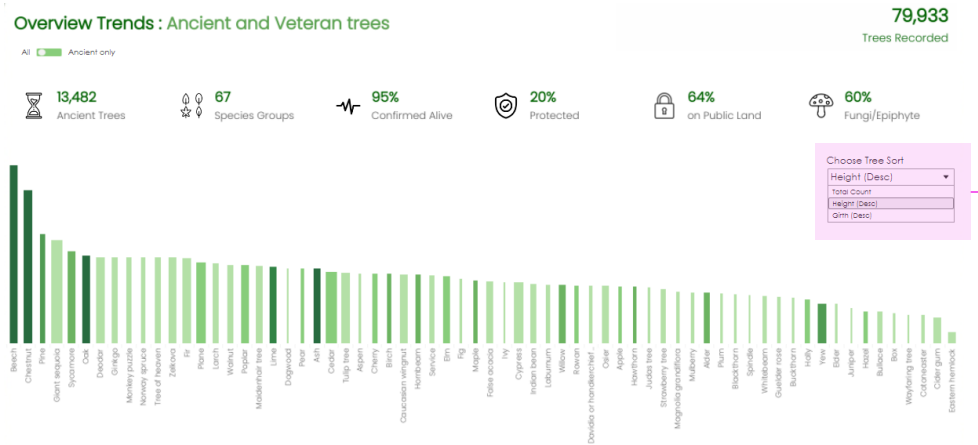
Overview Trends – User Interactivity



Toggling the Ancient-Only Slider *filters all metrics and charts to only Ancient, or Ancient and Veteran trees*



Clicking a species bar *filters the rest of the page (including maps not shown here) for the given species – here, Chestnut*

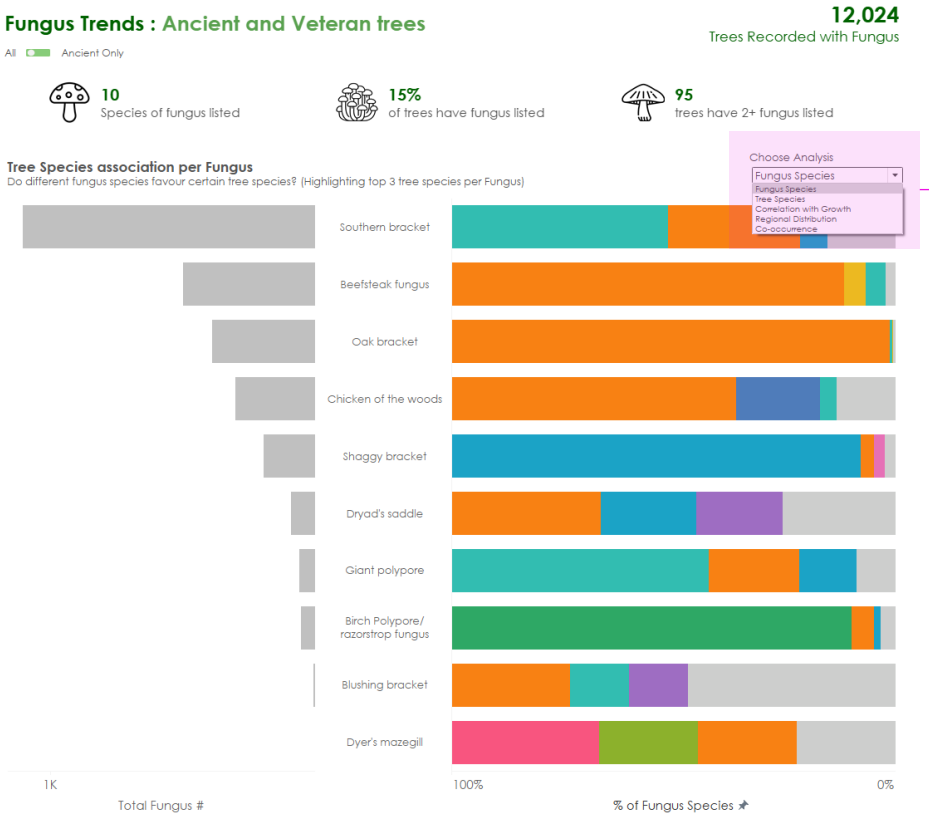


Selecting a Tree Sort *changes the order of the species shown (x-axis) in the height/girth/count chart*

Ancient Tree Inventory

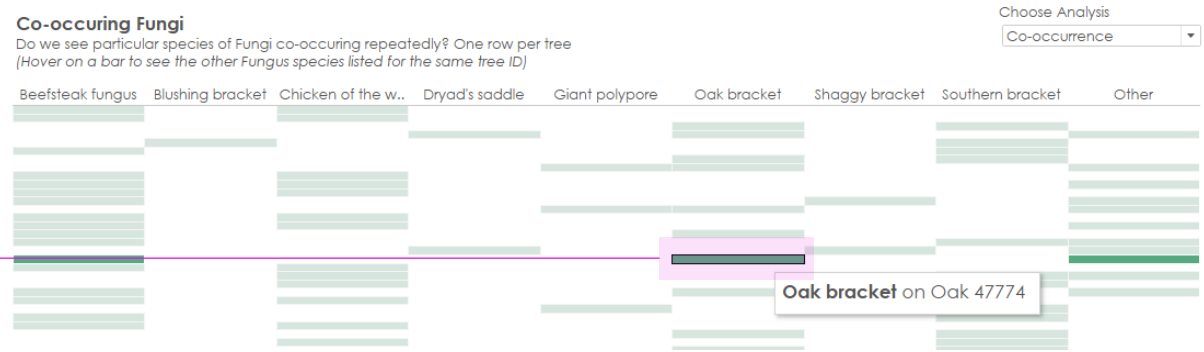
Tableau : Interactivity

Deep-dive Trends – User Interactivity



Selecting Analysis Approach *changes the main Fungus/Epiphyte chart display*

Hovering over a fungus record *in the co-occurrence plot highlights the other fungus species on that individual tree*



tabular18
helen.k8608
info@helenkinneydata.com

Ancient Tree Inventory

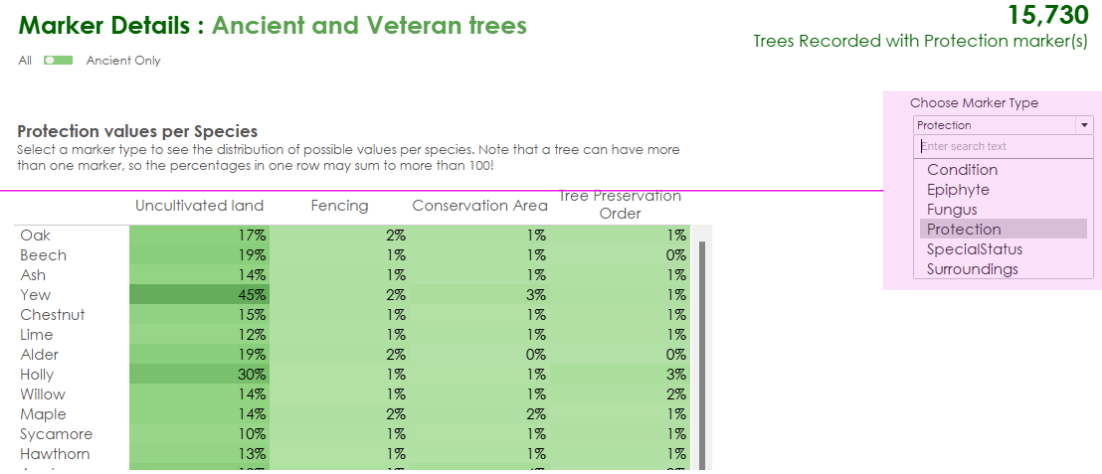
Tableau : Interactivity

Other User Interactivity

Clicking 'Page Menu' on any page *activates the page navigation menu*

Selecting a Page button *navigates to the chosen page, but the 'Close' button still needs to be used to de-activate the menu*

Selecting a Marker Type *on the Marker Details page changes the marker value columns*



tabular18 helen.k8608 info@helenkinneydata.com