

Declaration on Plagiarism

This form must be filled in and completed by the student submitting an assignment

Name:	Andrey Totev
Student Number:	20214021
Programme:	MCM-AI, part-time
Module Code:	CA682i
Assignment Title:	Data Visualisation
Submission Date:	25 Nov 2021
Module Coordinator:	Dr Suzanne Little

I declare that this material, which I now submit for assessment, is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work. I understand that plagiarism, collusion, and copying are grave and serious offences in the university and accept the penalties that would be imposed should I engage in plagiarism, collusion or copying. I have read and understood the Assignment Regulations. I have identified and included the source of all facts, ideas, opinions, and viewpoints of others in the assignment references. Direct quotations from books, journal articles, internet sources, module text, or any other source whatsoever are acknowledged and the sources cited are identified in the assignment references. This assignment, or any part of it, has not been previously submitted by me or any other person for assessment on this or any other course of study.

I have read and understood the referencing guidelines found at <http://www.dcu.ie/info/regulations/plagiarism.shtml>, <https://www4.dcu.ie/students/az/plagiarism> and/or recommended in the assignment guidelines

Name: **Andrey Totev**

Date: **25 Nov 2021**

Accommodation Supply In Points of Interest - Ireland

Abstract

Tourism is one of Ireland's most important economic sectors with an estimated yearly revenue of €9.4 billion prior to the COVID-19 pandemic [1]. The mission of the National Tourism Development Authority, Fáilte Ireland, is to develop tourism across Ireland via planning, investment, and business support [2].

In this report we use Fáilte Ireland's datasets [3] from the National Open Data Portal [4] in order to visualise the supply of accommodation in point-of-interest (POI) clusters in the Republic of Ireland. Additionally, we have enriched the POI data with review information from Google Places API [5]. The review data is then used as the basis for an interactive "significance filter" allowing the user to rebuild the visualisation by dynamically excluding data points with fewer Google reviews.

Our findings show that tourist attractions and activities in Ireland are evenly serviced by accommodation businesses with higher supply available in the east/south-east of the country and slightly lower in the inland areas adjacent to the west coast. One notable "blind spot" is North Mayo where the lack of accommodation might represent both an inhibitor to tourism and a development opportunity.

Datasets

Big Data and Scalability

The aspect of big data represented in this visualisation is clearly **variety** since we're integrating industry, topographic, and social networks data from disparate sources such as government portals, open source, and commercially available REST APIs.

Fáilte Ireland Open Data

URL: <https://data.gov.ie/organization/failte-ireland>

Dataset	Size (KB)	Row count	Attribute	Data type (NOIR)
Attractions.csv	89	545	Name URL Telephone	nominal nominal nominal
Activities.csv	1104	6319	Longitude Latitude Region	interval interval nominal

Accommodation.csv	430	2774	Locality	nominal
			Country	nominal
			Tags	nominal

Google Places API

URL: <https://developers.google.com/maps/documentation/places/web-service/search-find-place>

Sample request	<p>HTTP GET</p> <p><code>https://maps.googleapis.com/maps/api/place/findplacefromtext/json?input={name}&inputtype=textquery&locationbias=circle%3A100%40{lat}%2C{lon}&fields=name%2Crating%2Cuser_ratings_total&key={key}"</code></p> <p>(Note: API parameters printed in bold.)</p>
Sample response	<pre>{ "candidates": [{ "formatted_address": "140 George St, The Rocks NSW 2000, Australia", "geometry": { "location": { "lat": -33.8599358, "lng": 151.2090295 }, "viewport": { "northeast": { "lat": -33.85824377010728, "lng": 151.2104386798927 }, "southwest": { "lat": -33.86094342989272, "lng": 151.2077390201073 }, }, }, "name": "Museum of Contemporary Art Australia", "rating": 4.4, "user_ratings_total": 103 },], "status": "OK", }</pre>

TopoJSON - Ireland Counties

TopoJSON is an extension of GeoJSON that encodes topology [6]. TopoJSON's storage format is more efficient for features such as administrative regions.

URL: <https://github.com/deldersveld/topojson/blob/master/countries/ireland/ireland-counties.json>

Size: 38KB

Data Exploration, Processing, Cleaning and/or Integration

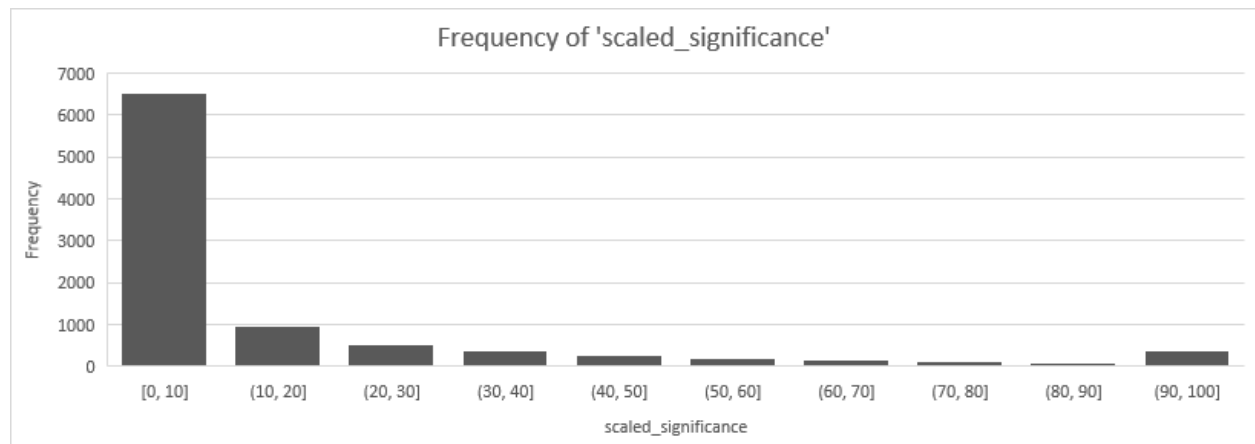
Data processing and integration

- Concatenated attractions, activities, and accommodation into a POI dataset, storing the origin of the points as a new categorical feature called “type”
- Using Google Places API, retrieve the total count of Google reviews for each POI entry, clipped at 2000 reviews
- Apply min-max scaling to the review counts and store as a new feature called “scaled_significance”
- D3 processing at runtime
 - Create a projection to fit TopoJSON data and use for translating all coordinates
 - Filter out POIs based on the position of the significance filter [7]
 - Group the data into hexagonal bins [8]
 - For each bin calculate “accommodation supply index” (ASI) using the formula:

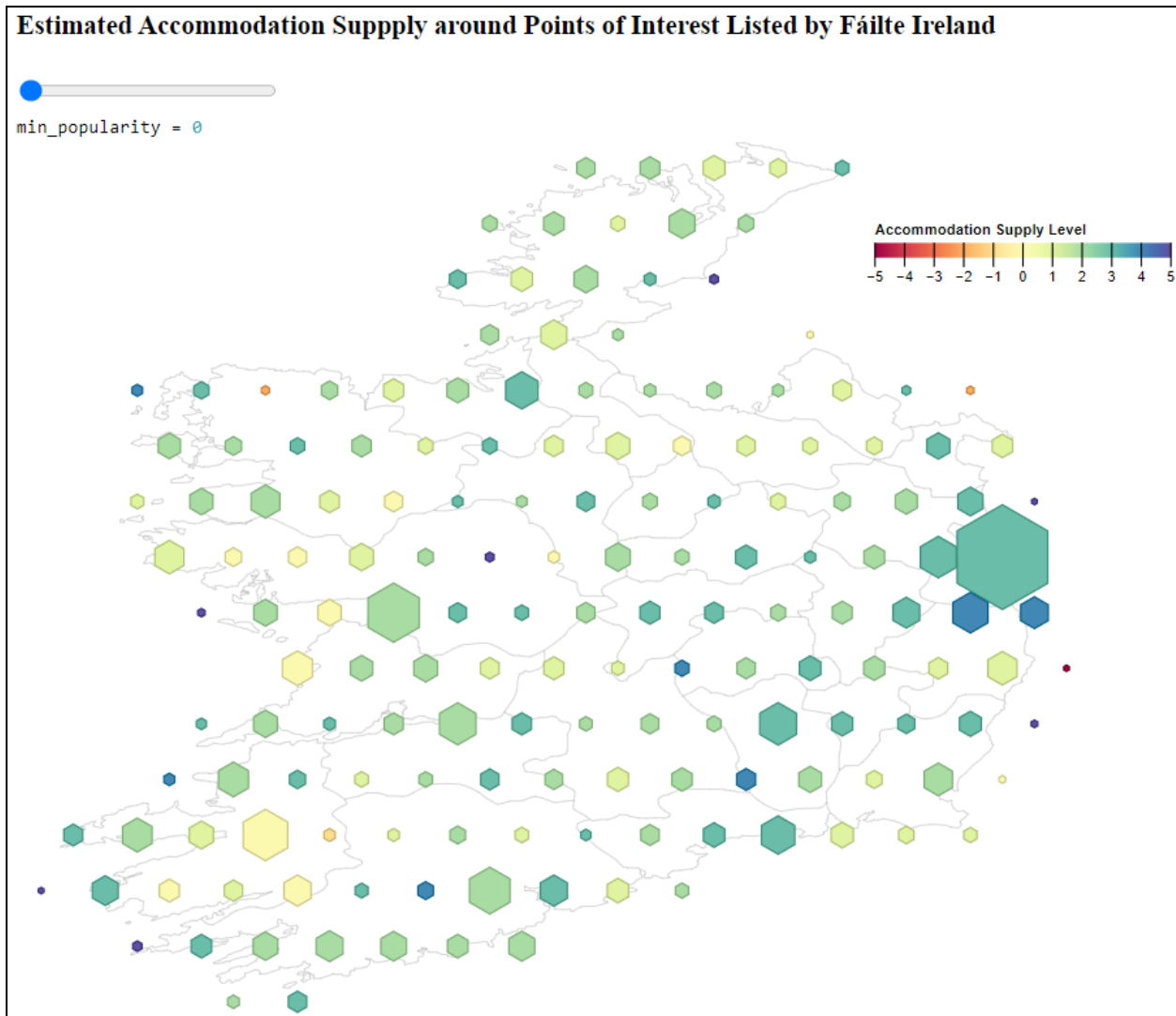
$$\text{ASI} = \text{int_round}(\text{bin_count}([\text{attractions}, \text{activities}]) / \text{bin_count_all}() * 10) - 5$$

Data exploration

The significance filtering relies on equi-width binning. Approximately 70% of the POIs fall into the first bin (<200 reviews) which makes the significance filter useful for exploring the top 30% of the POIs.



Visualisation



URL: <https://atotev.github.io/hexbin-map/index.html>

Choice of chart type

For our visualisation we choose a hexbin map with bivariate symbology [10] where the size of the bin represents the number of POIs and the color represents the ASI divergence from zero. The hexbin map leverages preattentive attributes such as spatial position, size and color in order to aid the viewer in identifying areas with fewer POIs (larger white areas), lower ASI (red color), or the opposites. Another benefit is the generous amount of blank space which helps the viewer to orientate based on the county borders.

The implementation of our chart is based on [9] from Observable where the US-centric projection is replaced and a slider is added to facilitate the interactive filtering.

Design choices

The design is mostly determined by the choice of a bivariate hexbin map. The color scheme [11] from the original example was preserved as it supports the semantics of ASI which measures

the degree of divergence from a presumably balanced state. In particular, red represents the extreme cases of lack of supply, whereas blue represents an excess of available accommodation.

Interactivity

The interactive significance filtering is mostly useful for the “top” 30% of POIs with more than 200 Google reviews while the remainder falls into the first bin (included in the default view.) This approach was taken with the assumption that importance among tourists cannot be gauged in this way before the locations achieve a “critical mass” of reviews.

Tools and libraries

This visualisation is implemented using the D3 framework via Observability - a JavaScript notebooks platform for web-based data visualisations.

Conclusion

Using the created visualisation we are able to form the following hypotheses (among others)

- Accommodation services are spread relatively evenly across the country (indicated by largely predominant green and light green clusters)
- The east/south-east region has excess supply (dense dark green to blue clusters)
- East Kerry and other areas adjacent to the west coast are relatively undersupplied (a vertical stretch of yellow clusters)
- North Mayo has a relatively limited number of POIs which are also significantly undersupplied (small orange zone in the centre of a sizeable white area)
- The density of POIs along the East/West divide (Waterford - Tipperary - West Galway - Roscommon - Sligo) is notably lower. Some of the more significant ones are hotels
- The density is also low at the crossing of Kerry, Limerick, and Cork borders. A lack of major investment in hotels in South Limerick is likely

Even though the viewer is able to derive multiple plausible-sounding hypotheses through this visualisation, the ASI and the significance score are essentially heuristics for the type of metrics we try to present. A major improvement of this work would be to research more established and sophisticated techniques for estimating accommodation supply and POI significance with the potential use of more diverse data.

Other enhancements may include using a more familiar map projection, adding the county names to the map, informational pop-ups when the user holds the mouse over a hexbin, and printing the encoded values as text in the bins themselves.

One technical limitation would be due to the Observability Runtime which has an opinionated way of tracing the state of the notebook, making it harder for embedding seamlessly in other web sites.

References

1. GOI Tourism Policy, <https://www.gov.ie/en/policy/3fcc3a-tourism/>
2. About Fáilte Ireland, <https://www.failteireland.ie/about-us.aspx>
3. Open Data available from Fáilte Ireland,
<https://www.failteireland.ie/Research-Insights/Open-data.aspx>
4. National Open Data Portal / Fáilte Ireland, <https://data.gov.ie/organization/failte-ireland>
5. Google Places API,
<https://developers.google.com/maps/documentation/places/web-service/overview>
6. TopoJSON / Ireland Counties,
<https://github.com/deldersveld/topojson/blob/master/countries/ireland/ireland-counties.json>
7. Observability - Filtering, <https://observablehq.com/@gavingosling/lab-5-filtering>
8. Observability D3 Hexbin, <https://observablehq.com/collection/@d3/d3-hexbin>
9. Observability D3 Hexbin Map,
<https://observablehq.com/@d3/hexbin-map?collection=@d3/d3-hexbin>
10. Indiemaps - Hexbins!, <http://indiemaps.com/blog/2011/10/hexbins/>
11. D3 Scale Chromatic, <https://github.com/d3/d3-scale-chromatic>