

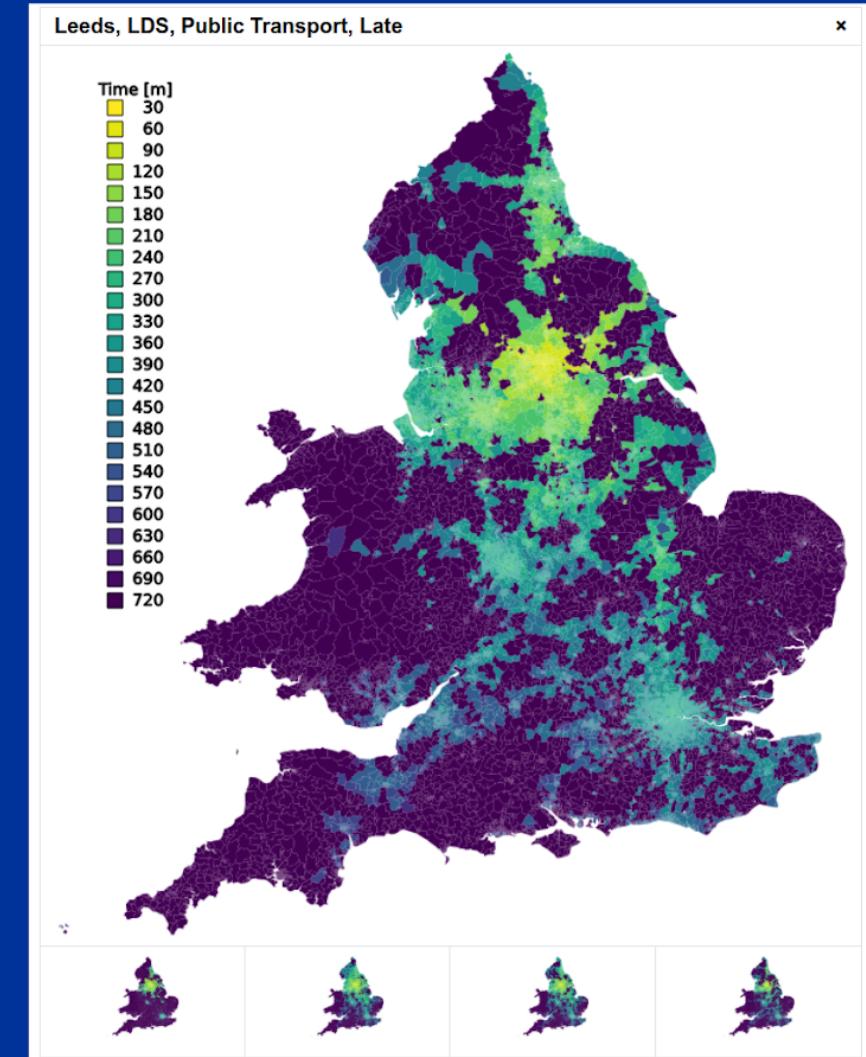
# My Love Grows (Where My Railway Goes)

on the equinox  
lets make sense of open rail data

2021-09-22T19:00/20:00+01:00



LEEDS DIGITAL FESTIVAL | 2021



# Introduction

Will Deakin

Twitter: @WillDeakin1

[will.deakin@crinstitute.org.uk](mailto:will.deakin@crinstitute.org.uk)



LEEDS DIGITAL FESTIVAL | 2021

# **Blah, blah, blah**

**'Our first speaker will guide us through the various different initiatives...'**

- Introduction
- Early Forays into data
- Visualisation and a Global Map of Wayne
- British Rail Data
- Q&A



**LEEDS DIGITAL FESTIVAL | 2021**

# Rosemary...

**'Madame Helena, the timetable  
is the most perfect product of  
the human mind'**

- R.U.R. (Rossum's Universal Robots) by  
Karel Capek



LEEDS DIGITAL FESTIVAL | 2021

# Railway (Not Rosemary)

**'He who neglects what is done  
for what ought to be done,  
sooner effects his ruin than his  
preservation'**

- The Prince, Niccolò Machiavelli



LEEDS DIGITAL FESTIVAL | 2021

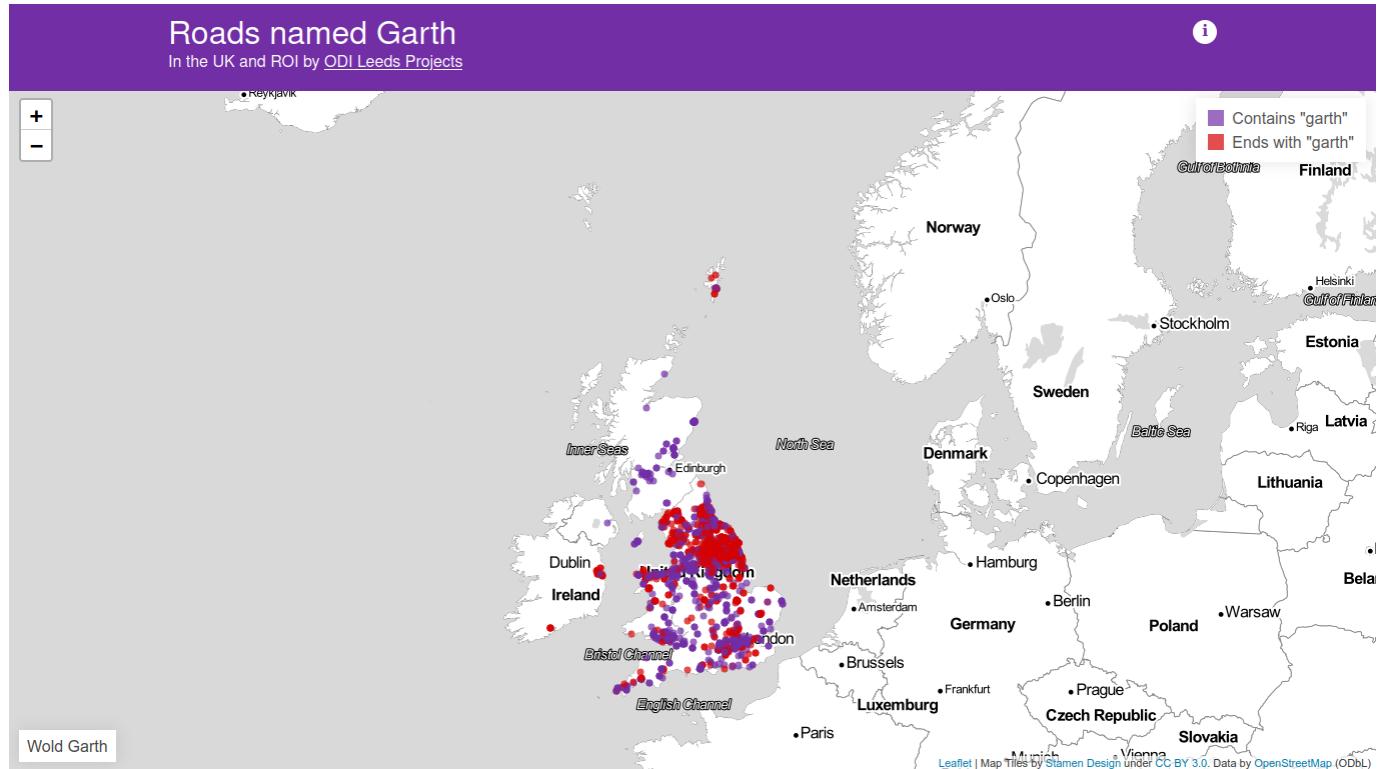
# Early Forays into Data



LEEDS DIGITAL FESTIVAL | 2021

# In the Beginning

So here are some of  
the things I do to get  
the things I do done



LEEDS DIGITAL FESTIVAL | 2021

# Early Forays into Data

OSEMN OpenStreetMap and rail

Data Science at the Command Line  
by Jeroen Janssens

The elements of data science are

- Obtaining data
- Scrubbing data
- Exploring data
- Modeling data
- iNterpreting data

OSEM pronounced ‘awesome’



LEEDS DIGITAL FESTIVAL | 2021

O'REILLY



Data  
Science  
at the  
Command Line

FACING THE FUTURE WITH TIME-TESTED TOOLS

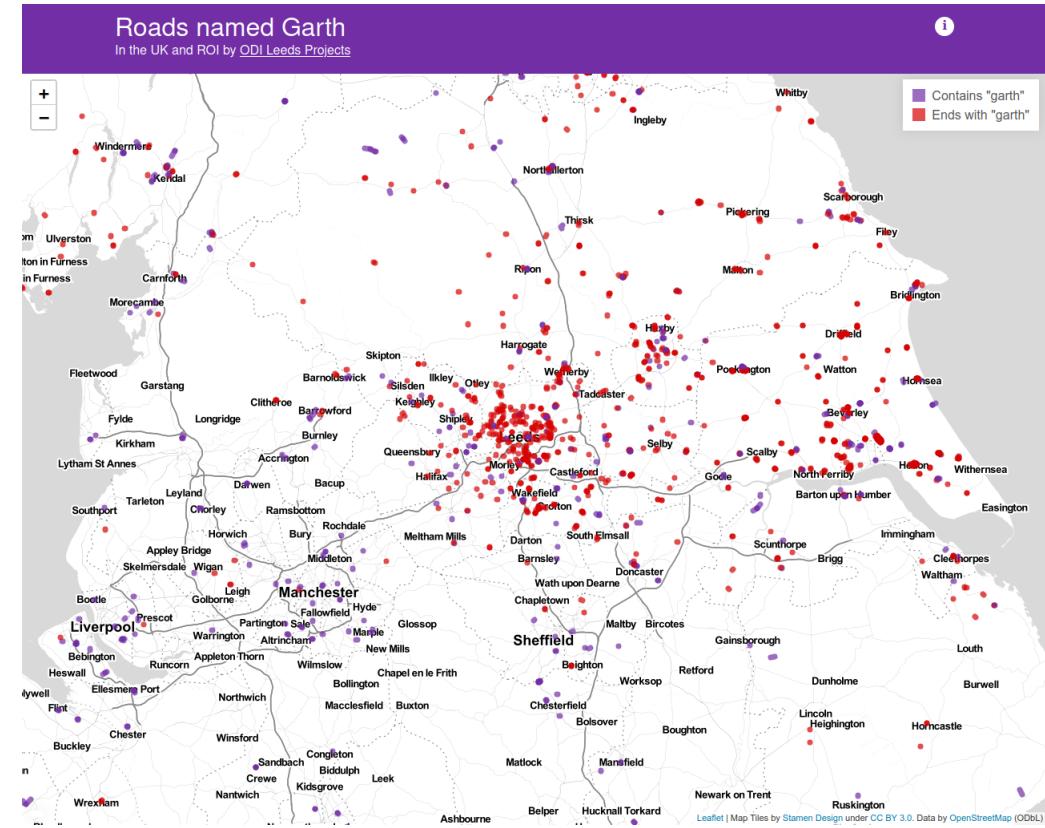
Jeroen Janssens

# The Garthiness of Leeds

'It turns out **Leeds** is a bit of a **Garth** hotspot and their distribution generally correlates with regions settled by the Vikings around 1000 years ago.

Party on!

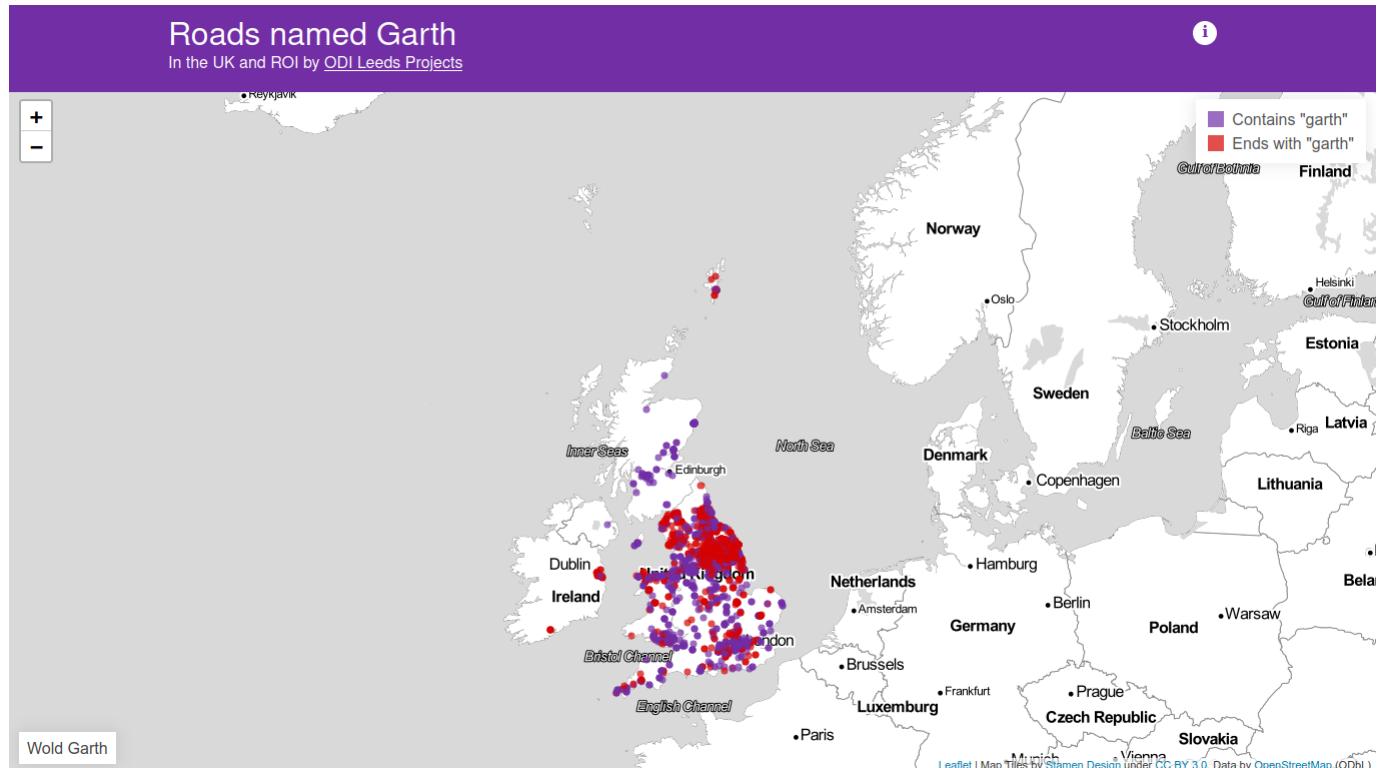
Stuart, ODI Leeds



LEEDS DIGITAL FESTIVAL | 2021

# Hiraeth of Irfon Valley C.P. School

**Hiraeth: a mixture of longing, yearning, nostalgia, wistfulness or an earnest desire for the Wales of the past**

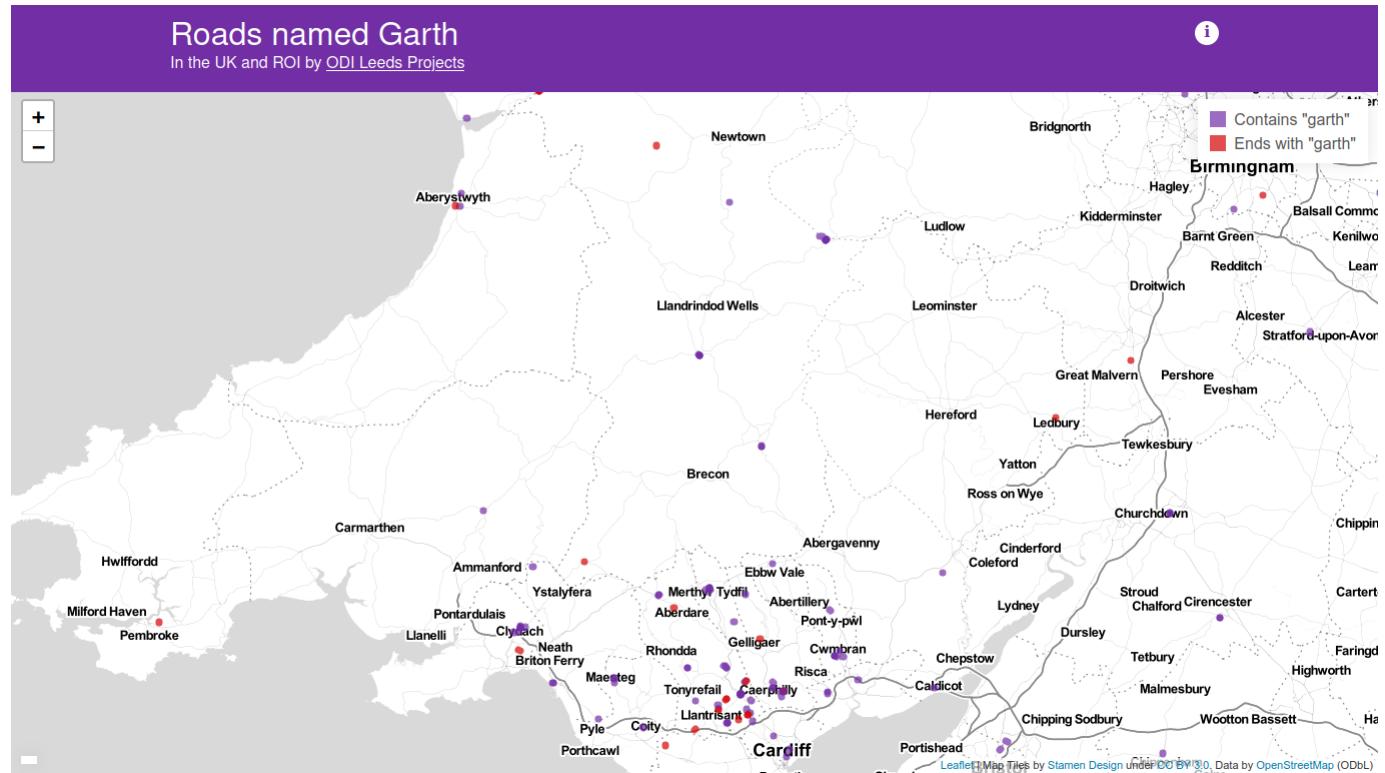


LEEDS DIGITAL FESTIVAL | 2021

# Party On!

# But where was Garth?

**...but more importantly,  
what about Wayne?**



**LEEDS DIGITAL FESTIVAL** | **2021**

# Visualisation and a Global Map of Wayne



LEEDS DIGITAL FESTIVAL | 2021

# British Map of Wayne

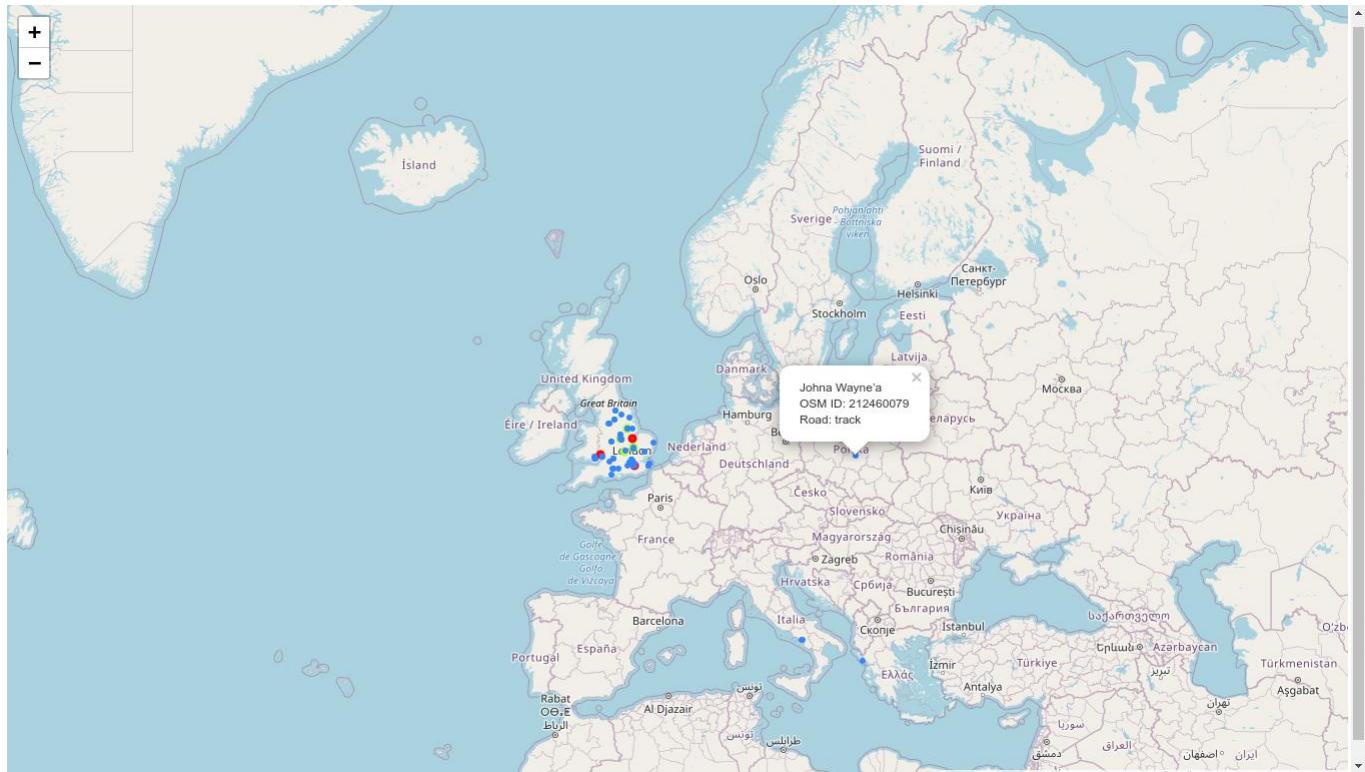
First a British Wayne



LEEDS DIGITAL FESTIVAL | 2021

# European Map of Wayne

Lest we forget  
European Wayne

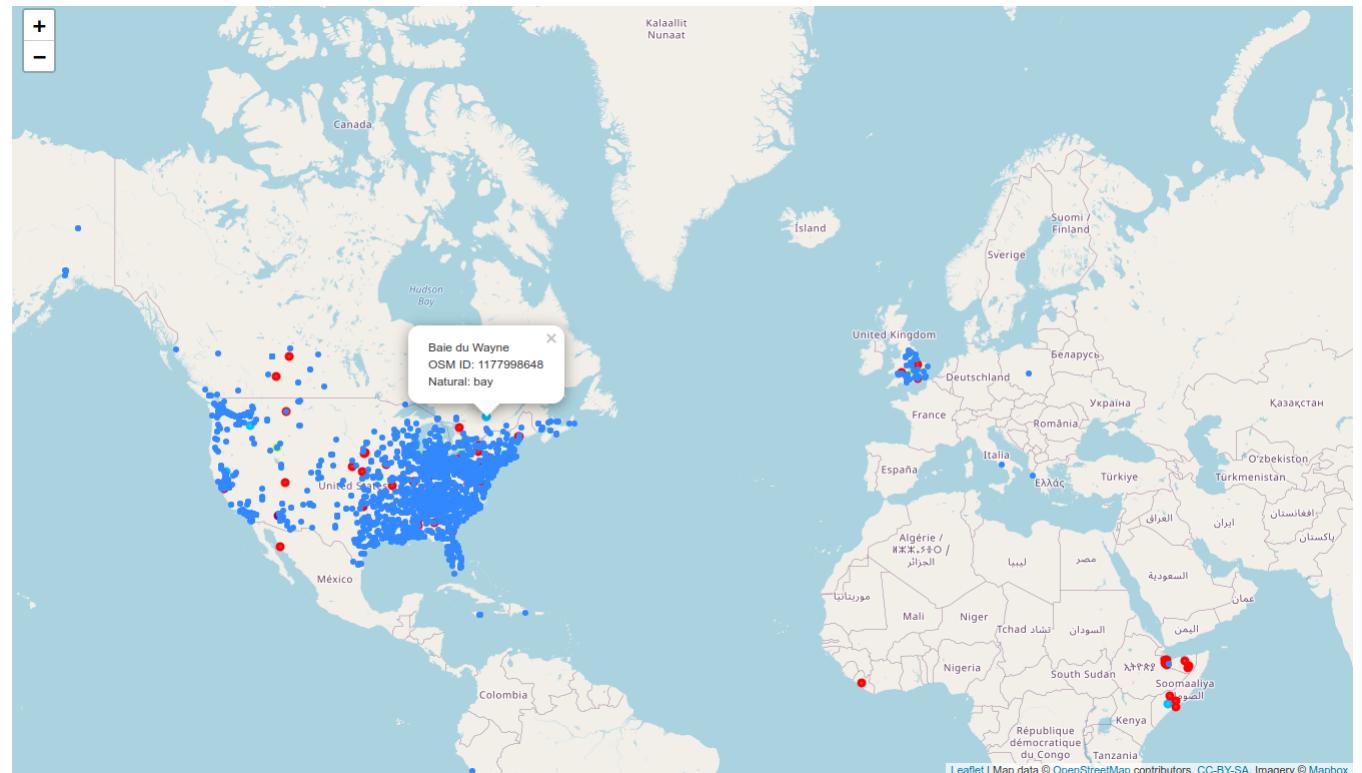


LEEDS DIGITAL FESTIVAL | 2021

# International Map of Wayne

# A whole lot of Wayne\*

## \* but no fountains

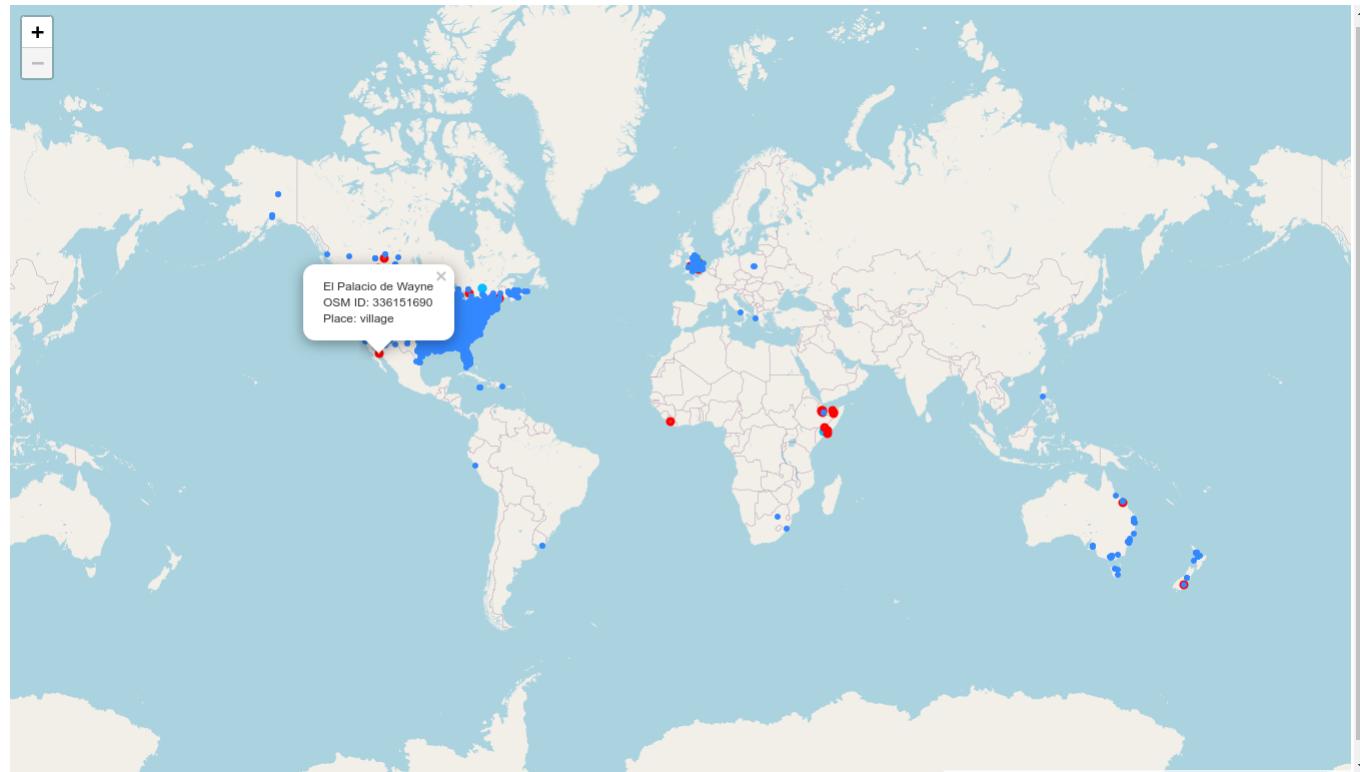


LEEDS DIGITAL FESTIVAL | 2021

# The Global Map of Wayne

'In tribute to the work of the Open Data Institute Leeds ([HERE](#)) I bring you global mapping of Wayne, or as it were, **Wayne's World**'

Excellent! Excellent!



LEEDS DIGITAL FESTIVAL | 2021

# Party on Wayne!

It took about 20 hours to develop and three hours to run  
On an ancient desktop machine with about 36.8GB of disk

- Running Linux Mint (a Debian based Linux distribution)
- Using the global OpenStreetMap data from [geofabrik.de](http://geofabrik.de)
- The Geospatial Data Abstraction Library (GDAL) ogr2ogr tool
- Plus the lightweight and flexible command-line JSON processor jq

To produce a GeoJSON data file

- Rendered in 62 lines of JavaScript
- Using the Mike Bostock d3.js library
- Code and data hosted on [gist.github.com/anisotropi4](https://gist.github.com/anisotropi4)
- Published via bl.ocks.org [HERE](#)

```
<script type="text/javascript">
    var radius = 3;
    var weight = 3;
    var linewidth = 6;
    var log2 = Math.log(2.0);
    var minZoom = 2;
    var maxZoom = 18;
    var map = L.map('map').setView([53.533, -0.53], 6);
    var colours15 = {"highway": "GreenYellow", "natural": "DeepSkyBlue",
"place": "Red", "railway": "Purple", 6: "Orange", 7: "DeepPink", 8: "Magenta", 9:
"DarkSlateBlue", 10: "GreenYellow", 11: "Lime", 12: "Cyan", 1: "Blue", 14:
"DarkBlue", 15: "DarkOrange", 16: "DarkOrchid"};
    L.tileLayer("http://{s}.tile.openstreetmap.org/{z}/{x}/{y}.png", {
        maxZoom: maxZoom,
        minZoom: minZoom,
        attribution: 'Map data &copy; <a href="https://www.openstreetmap.org/">OpenStreetMap</a> contributors, ' + '<a href="https://creativecommons.org/licenses/by-sa/2.0/">CC-BY-SA</a>, ' + 'Imagery
        © <a href="https://www.mapbox.com/">Mapbox</a>',
        id: 'osm.standard'
    }).addTo(map);

    d3.json("output-all.json").then(function(d) {
        function onEachFeature(feature, layer) {
            var popupContent = feature.name;
            var lookup = {"osm_id": "OSM ID", "place": "Place",
"natural": "Natural", "place": "Place", "railway": "Railway", "highway": "Road",
"ref": "CRS"}
            var k = Object.keys(feature).filter(i => (i != "type" && i !=
"geometry" && i != "name" && i != "is_in"))
            for (var i = 0; i < k.length; i++) {
                popupContent += '<br>' + lookup[k[i]] + ': ' +
feature[k[i]];
            }
            layer.bindPopup(popupContent);
        }
        L.geoJSON(d, {
            style: function(feature) {
                var k = Object.keys(feature).filter(i => (i == "highway"
|| i == "natural" || i == "place" || i == "railway"))[0];
                switch(feature.geometry.type) {
                    case 'Point': return {color: colours15[k], radius:
radius, weight: weight};
                    case 'LineString': return {weight: linewidth};
                    default: return {weight: weight};
                }
            },
            onEachFeature: onEachFeature,
            pointToLayer: function(feature, latlng) {
                return L.circleMarker(latlng, {
                    opacity: 1,
                    fillOpacity: 0.8
                });
            }
        }).addTo(map);
    });
</script>
```



LEEDS DIGITAL FESTIVAL | 2021



LEEDS DIGITAL FESTIVAL | 2021

# British Rail Data



LEEDS DIGITAL FESTIVAL | 2021

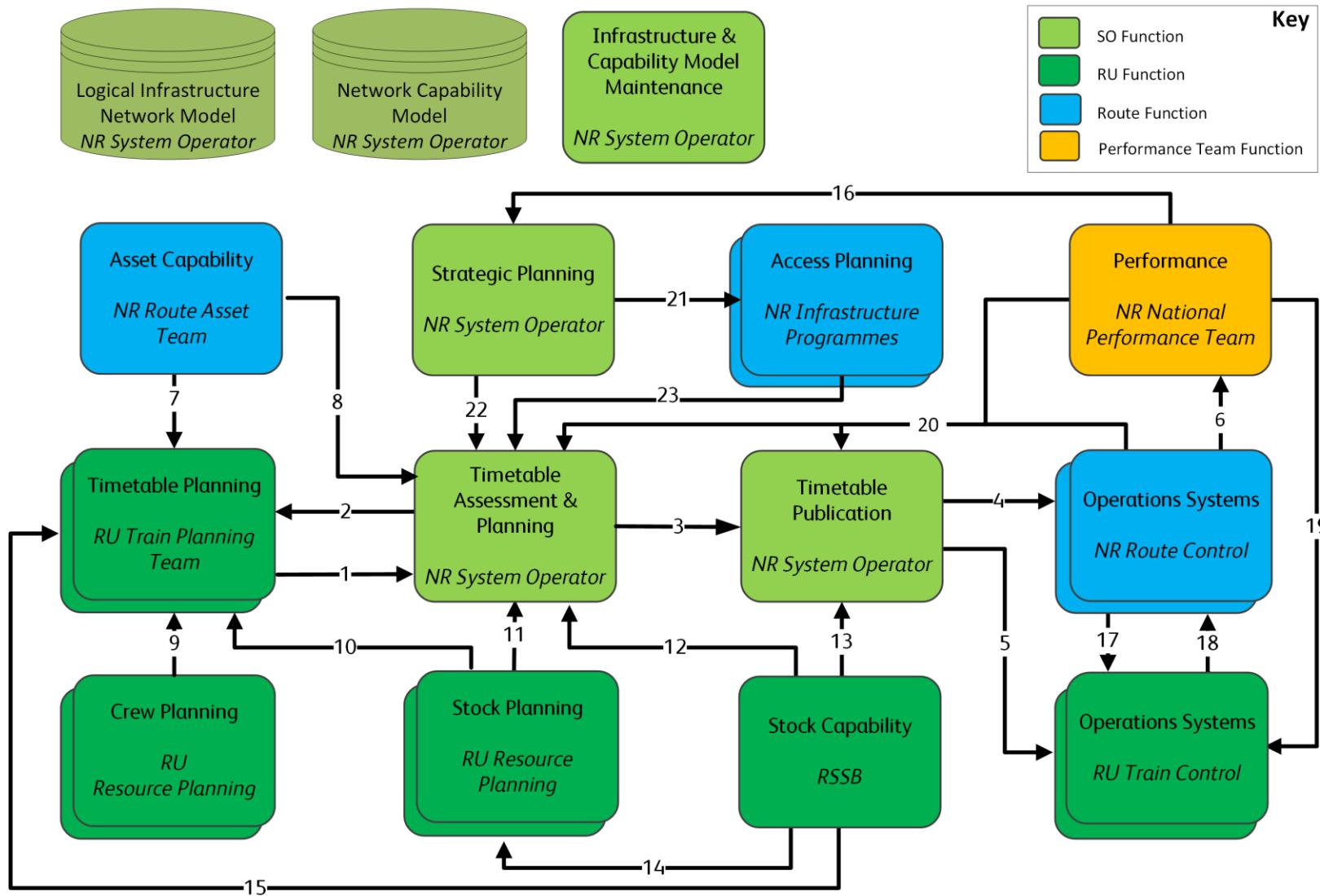
# British Rail Data

1. Public Transport Travel statistics
2. National Electronic Sectional Appendix
3. Rail Electrification
4. A Strategic Rail Network



LEEDS DIGITAL FESTIVAL | 2021

# British Rail Data



TINLRT69100799603KNORTHLALLERTON SIG. 691	15888	0
TINLRTDBW00787705PDANBY WISKE	15875	0
TINLRTEJN00799621XNORTHLALLERTON EAST JN	15880	0
TINLRTN 00799600HNORTHLALLERTON	15883	0
TINLRTNCH48799615PCASTLE HILLS LOOP	15884	0
TINLRTNU00799601WNORTHLALLERTON UP SDGS	15886	0
TINLRTREV00799605MNORTHLALLERTON REV LINE	15878	0
TINLRTWDR08886100JNORTHLALLERTON WENSLEYDALE	15901	0
TINLSN 00256500KNELSON	29033	0
TINMBRLPK16697100BNORTHUMBERLAND PARK	51929	0
TINMILTON01588100HNEW MILTON	86911	0
TINMPT05000106922QNORTHAMPTON SIGNAL RY1050	70115	0
TINMPT21600106920XNORTHAMPTON SIGNAL RY1216	70114	0
TINMPT21800106919TNORTHAMPTON SIGNAL RY1218	70113	0
TINMPTBAY56106915QNORTHAMPTON BAY	70100	0
TINMPTBST00106800DNORTHAMPTON BRIDGE ST CE	70111	0
TINMPTCMD56106811HNORTHAMPTON C.M.D.	70112	0
TINMPTCYG00106923RNORTHAMPTON CASTLE YRD GBF70109	0	
TINMPTEMA00106921PNORTHAMPTON EMU ARRIVALS	70098	0
TINMPTEMD00106918ZNORTHAMPTON EMD	70099	0
TINMPTHHT00106917YHUNSBURY HILL TUNNEL	70120	0
TINMPTMLJ56106916RNORTHAMPTON MILL LANE JN	70097	0
TINMPTN 00106900WNORTHAMPTON	70100	0
TINMPTNCY48106904MNORTHAMPTON CASTLE YARD	70102	0
TINMPTNN416106015RNORTHAMPTON DGL	70107	0
TINMPTNNJ48106906XNORTHAMPTON NORTH JN	70103	0
TINMPTNRS48106001KNORTHAMPTON RIVER SDGS	70101	0
TINMPTNSJ48106907PNORTHAMPTON SOUTH JN	70105	0
TINMPTNTC48106000JNORTHAMPTON T.C.	70104	0
TINMPTNU548106002LNORTHAMPTON TC UP SIDINGS	70104	0
TINMPTUR 24106005XNORTHAMPTON UP RECEPTION	70108	0
TINNRYMLJ56667813JNUNNERY MAIN LINE JN	25635	0

```
{"header":{"msg_type":"0003","source_dev_id":"","user_id":"","original_data_source":"SMART","msg_queue_timestamp":"1533122009000","source_system_id":"TRUST"},"body":{"event_type":"ARRIVAL","gbtt_timestamp":"1533118740000","original_loc_stanox":"","planned_timestamp":"1533118740000","timetable_variation":"1","original_loc_timestamp":"","current_train_id":"","delay_monitoring_point":"false","next_report_run_time":"2","reporting_stanox":"00000","actual_timestamp":"1533118680000","correction_ind":"false","event_source":"AUTOMATIC","train_file_address":null,"platform":"","division_code":"80","train_terminated":"false","train_id":"882D36MF01","offroute_ind":"false","variation_status":"EARLY","train_service_code":24657005,"toc_id":"80","loc_stanox":"88402","auto_expected":"true","direction_ind":"UP","route":"0","planned_event_type":"ARRIVAL","next_report_stanox":"88401","line_ind":""}}
```

```
{"header":{"msg_type":"0003","source_dev_id":"","user_id":"","original_data_source":"SMART","msg_queue_timestamp":"1533122010000","source_system_id":"TRUST"},"body":{"event_type":"DEPARTURE","gbtt_timestamp":"1533118680000","original_loc_stanox":"","planned_timestamp":"1533118680000","timetable_variation":"1","original_loc_timestamp":"","current_train_id":"","delay_monitoring_point":"true","next_report_run_time":"3","reporting_stanox":"00000","actual_timestamp":"1533118620000","correction_ind":"false","event_source":"AUTOMATIC","train_file_address":null,"platform":1,"division_code":88,"train_terminated":"false","train_id":881F25MG01,"offroute_ind":"false","variation_status":"EARLY","train_service_code":24746000,"toc_id":88,"loc_stanox":88288,"auto_expected":"true","direction_ind":DOWN,"route":2,"planned_event_type":DEPARTURE,"next_report_stanox":88285,"line_ind":""}}
```

```
{"header":{"msg_type":"0003","source_dev_id":"","user_id":"","original_data_source":"SMART","msg_queue_timestamp":"1533122010000","source_system_id":TRUST,"body":{"event_type":DEPARTURE,"gbtt_timestamp":,"original_loc_stanox":,"planned_timestamp":1533118710000,"timetable_variation":0,"original_loc_timestamp":,"current_train_id":,"delay_monitoring_point":true,"next_report_run_time":4,"reporting_stanox":87903,"actual_timestamp":1533118680000,"correction_ind":false,"event_source":AUTOMATIC,"train_file_address":null,"platform":,"division_code":88,"train_terminated":false,"train_id":629T93MD01,"offroute_ind":false,"variation_status":ON TIME,"train_service_code":22720000,"to":}}
```



LEEDS DIGITAL FESTIVAL | 2021

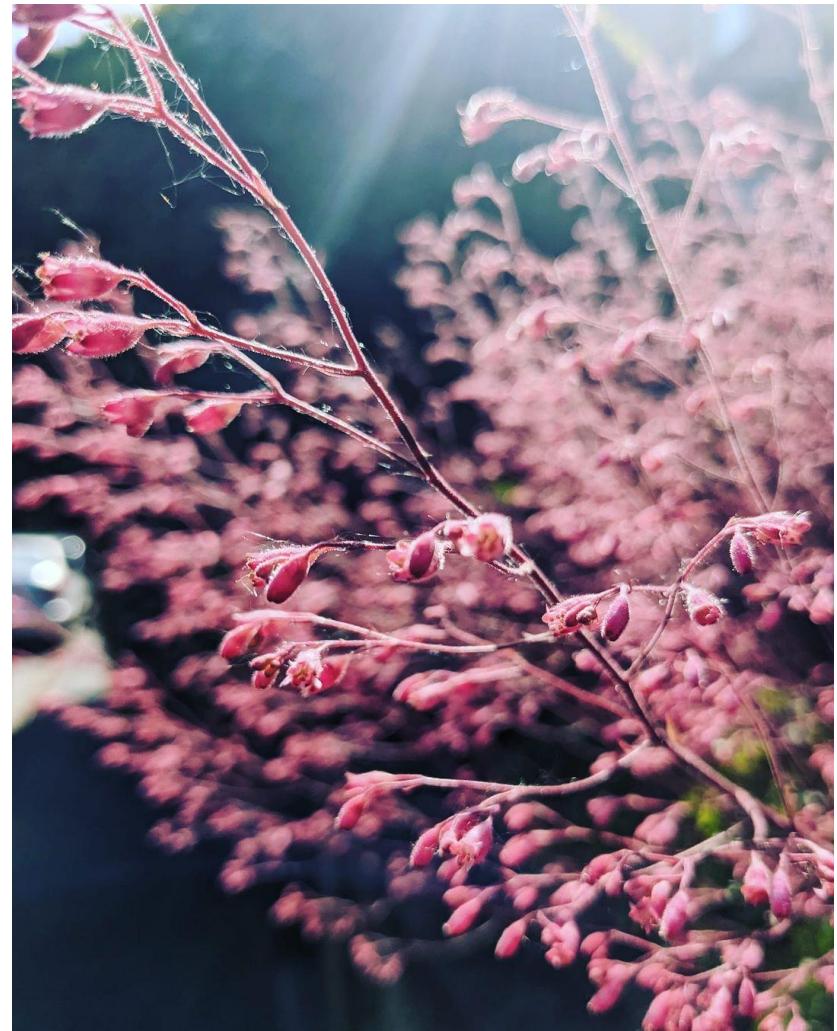
# British Open Transport and Rail Data

1. Public Transport Travel statistics
2. National Electronic Sectional Appendix
3. Rail Electrification
4. A Strategic Railway



LEEDS DIGITAL FESTIVAL | 2021

# Public Transport Times

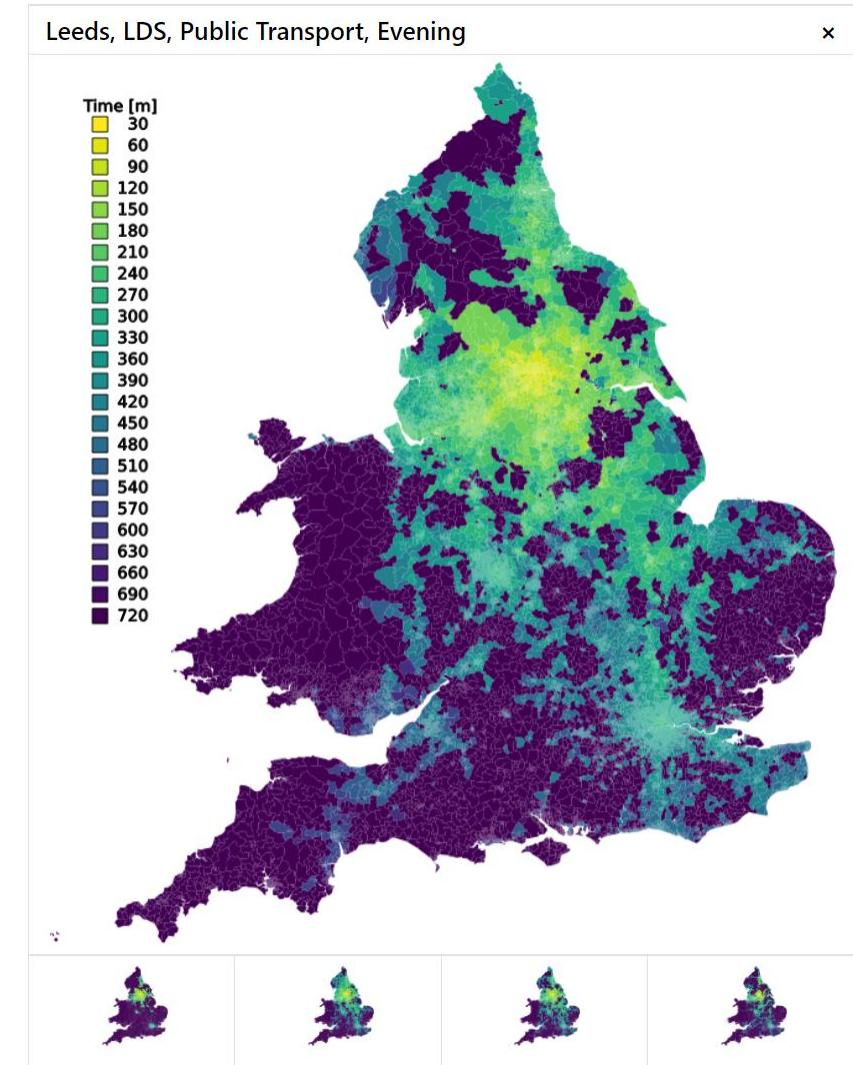


LEEDS DIGITAL FESTIVAL | 2021

# Public Transport Travel

Visualisation of Department for Transport (DfT) 2001/2011 experimental travel-time data to script **travel-time cloropleths** for public transport via train-stations for England and Wales

Based on Office for National Statistics (ONS) Lower Layer Super Output Areas (LSOA) geography



LEEDS DIGITAL FESTIVAL | 2021

# Public Transport Travel

This took far, far too long to develop and 180 minutes to run

- Based the excellent Mike Bostock Command-Line Cartography articles [HERE](#)

Use d3.js [HERE](#) and nodejs [HERE](#) on the command line

- Produce projected **GeoJSON** and **topoJSON** format LSOA geography files
- Which are combined with the travel times on the command-line
- Using the Mike Bostock **d3.js** library to colour using the Viridis colour scheme
- Code and data hosted on [gist.github.com/anisotropi4](https://gist.github.com/anisotropi4)

The JavaScript framework pulls static pre-created image files from [github.com](https://github.com)

- As file-size limits meant these to be as a separate [github.com](https://github.com) repository

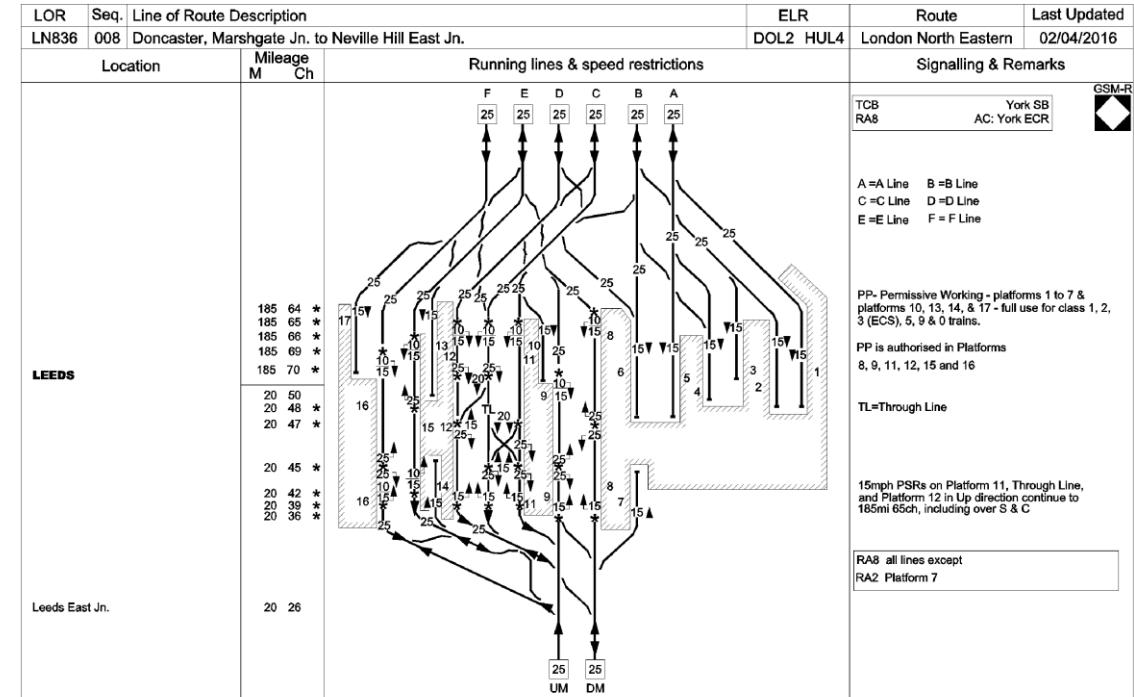
Published via [blocks.org](https://blocks.org) [HERE](#)



LEEDS DIGITAL FESTIVAL | 2021

# NESA\*

## \*Route Clearance Tables



LEEDS DIGITAL FESTIVAL | 2021

# National Electronic Section Appendix

The NESA Route Clearance tables are used to describe the Loading Gauge\* on the British rail network

Currently they are published as PDF files on the Network Rail website [HERE](#)

Using python with its Pandas and pdfplumber module an Excel format report is available to download [HERE](#)

\*maximum height and width for railway vehicles and their loads to ensure that they can pass safely through tunnels and under bridges, and keep clear of trackside buildings and structures



LEEDS DIGITAL FESTIVAL | 2021

# Why PDF\* is evil

To extract text from the PDF text-object elements, formatting issues and use of grey-scale background in key route-clearance columns break the text extraction

- To overcome this, the PDF files are converted to an uncompressed CMYK PDF/A format, the grey background removed by deleting the call and graphic state for the embedded grey background image. While it seems to work, this is in no way a recommended approach as
- It breaks the PDF files, as the PDF checksums no longer match
- It assumes the grey colour is encoded as 110 rg and rendered via a call to f\*
- Were the PDF rendering software used by Network Rail, or Ghostscript, or qpdf to change this would just break

\*and PostScript



LEEDS DIGITAL FESTIVAL | 2021

# Rail Electrification



LEEDS DIGITAL FESTIVAL | 2021

# Rail Electrification

‘Once upon a time on a hot day in May 2020 amidst a pandemic lockdown, a rail engineer tweeted...’



Garry Keenor (He/Him)  
@25kV

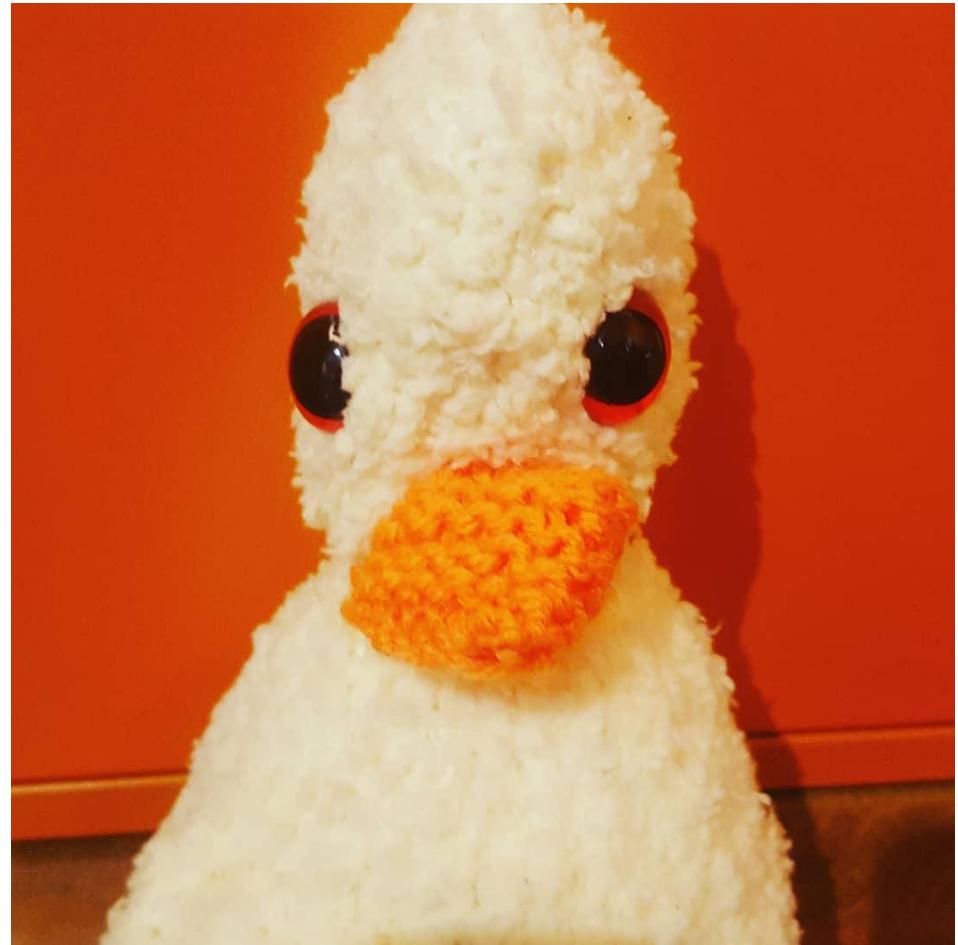
...

If anyone out there has experience of building an instance of [@openstreetmap](#), could you get in touch? I've had a dangerous idea, but I don't have all the necessary technical skills to implement

10:42 am · 25 May 2020 · Twitter for Android

---

1 Retweet 12 Likes



LEEDS DIGITAL FESTIVAL | 2021

# Fizzy Knitting

Following Garry Keenor's request I wanted to improve on Wayne's World

- That is to stop downloading 36GB of data, or at least the 6-8GB for GB

To achieve this

- Took 5-10 hours coding as I was already familiar with the approach
- Using the python OSMnx module developed by Geoff Boeing [HERE](#)
- To download OpenStreetMap rail-data using the Overpass API [HERE](#)
- To a GeoJSON format file that is rendered using d3.js

The example implementation is on bl.ocks.org [HERE](#)

The resulting UK Electric Railway Map is [HERE](#)

Please Note: my role was only to provide early OSM data and an interactive map framework



LEEDS DIGITAL FESTIVAL | 2021

# A Strategic Railway



LEEDS DIGITAL FESTIVAL | 2021

# My Love Grows

Where do you want your railway to go?

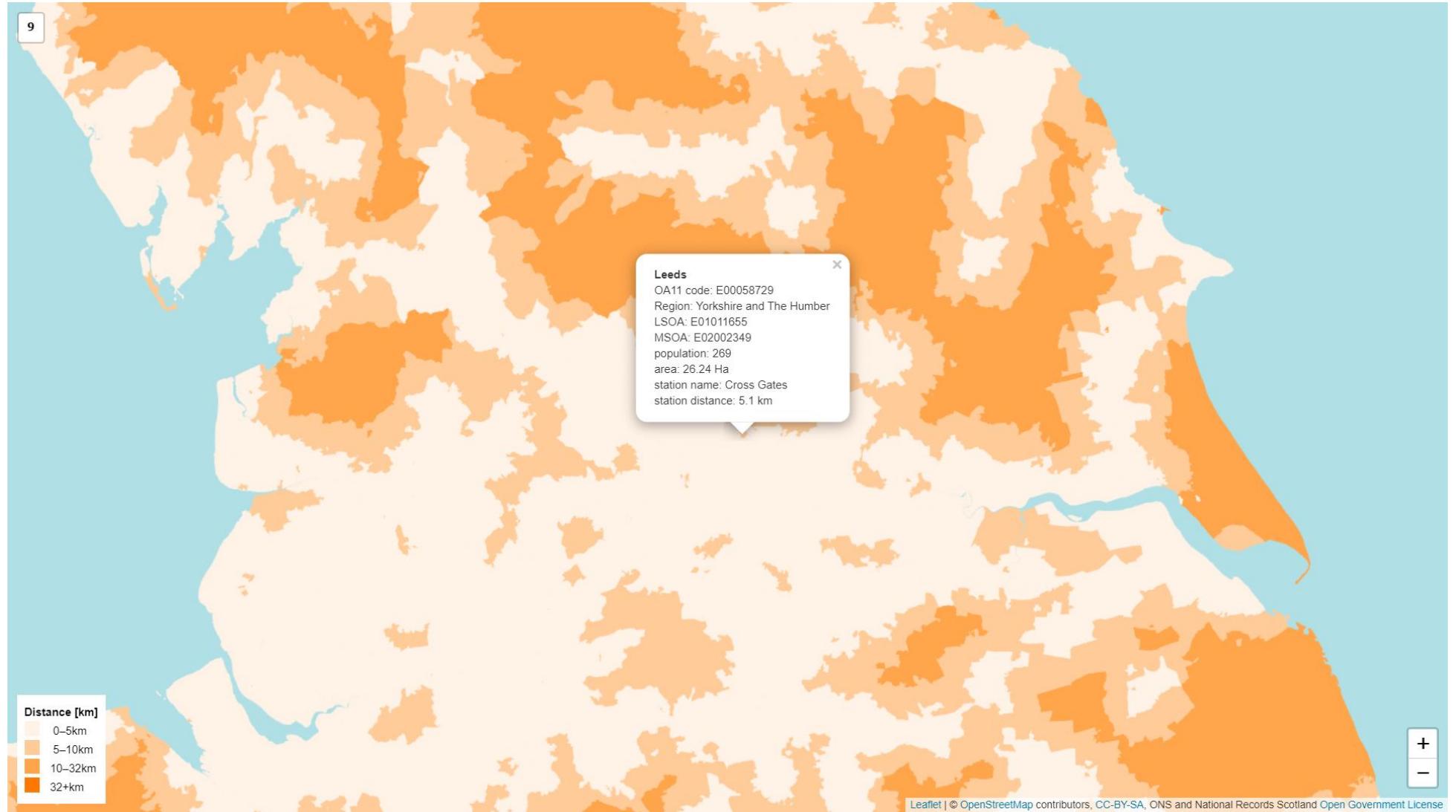
- Where people are
- Where there is no railway

Do you then want to change behaviour

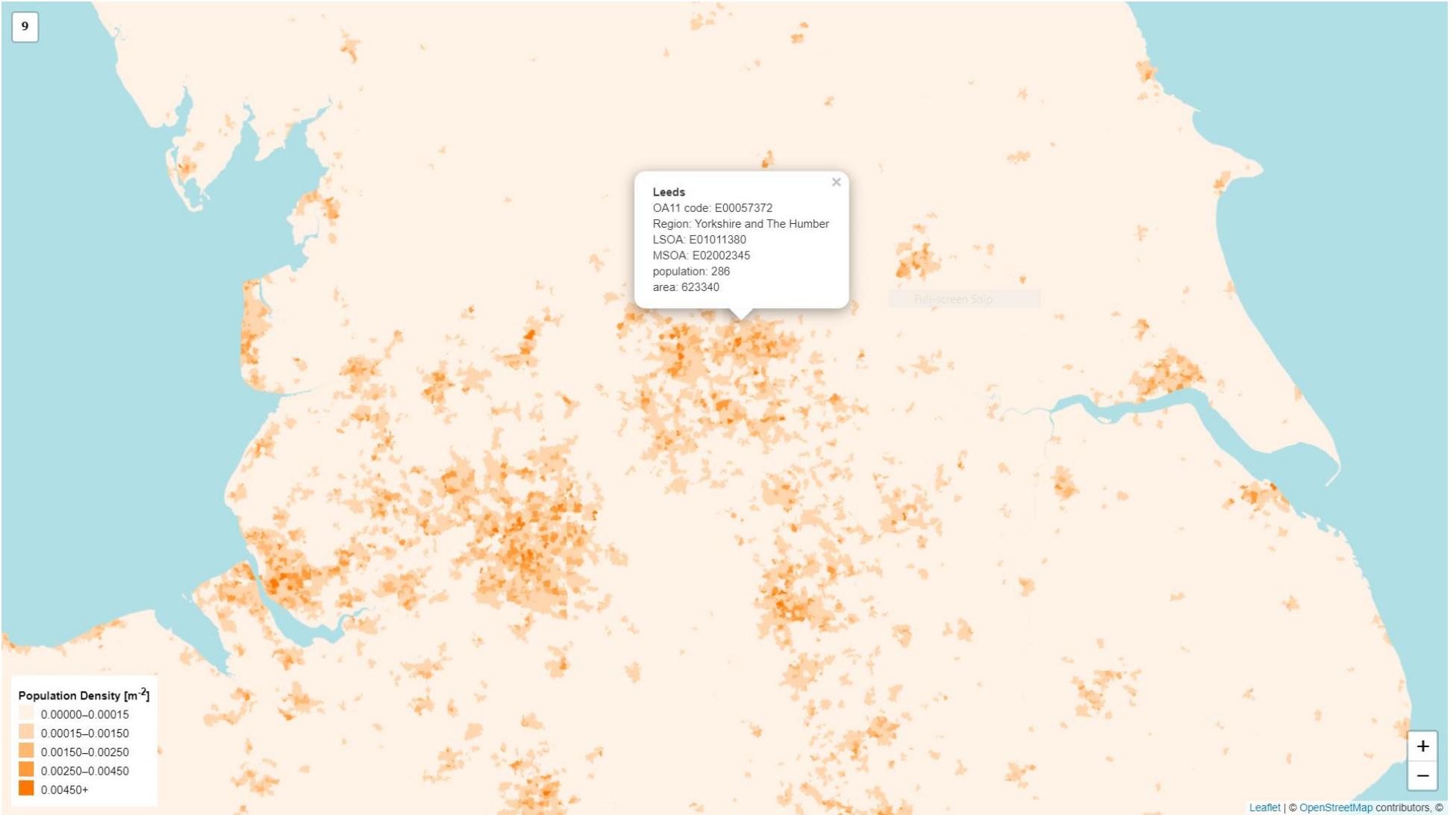
- Where do people go now?
- Where are the main roads?
- Where are there problems?



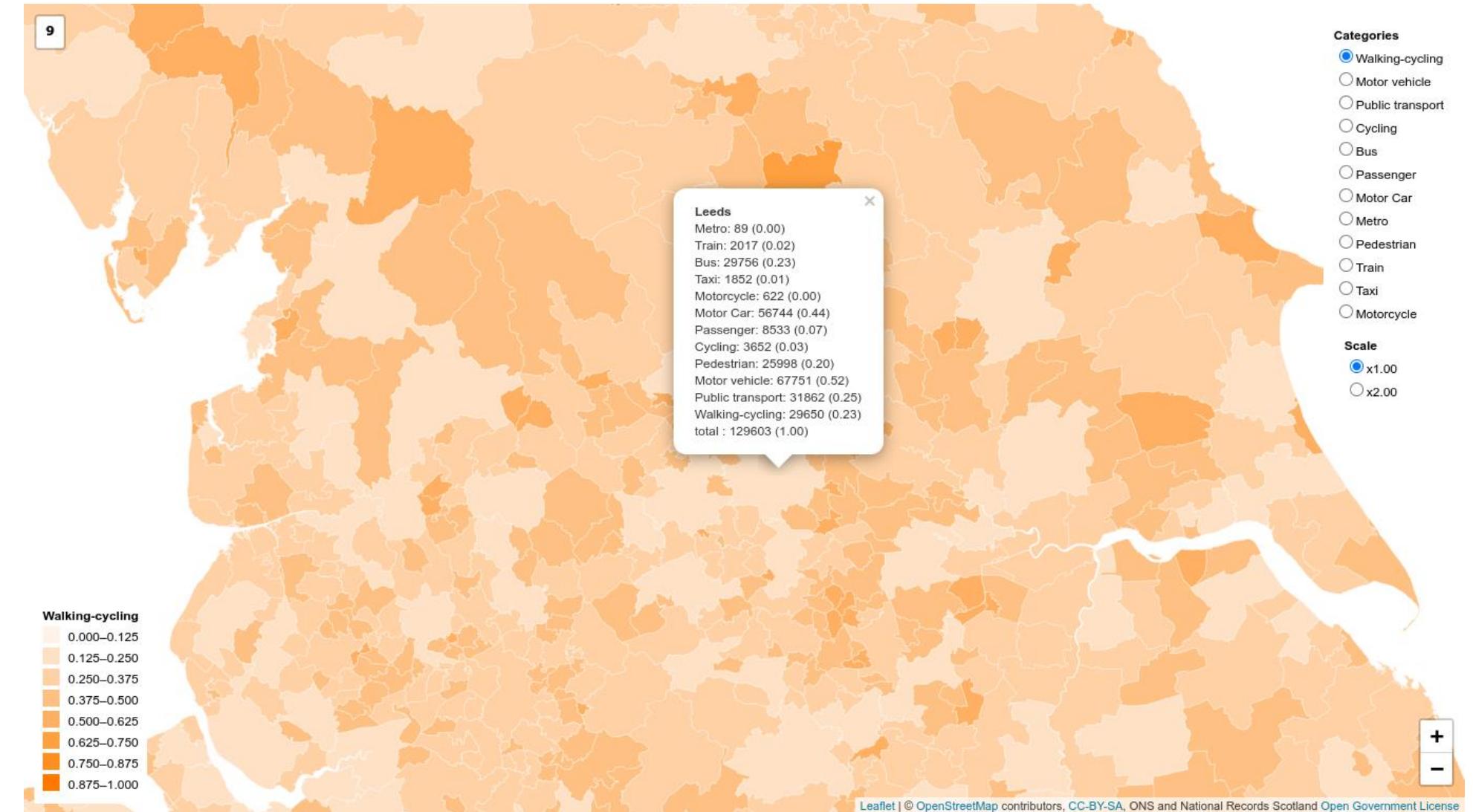
LEEDS DIGITAL FESTIVAL | 2021



LEEDS DIGITAL FESTIVAL | 2021



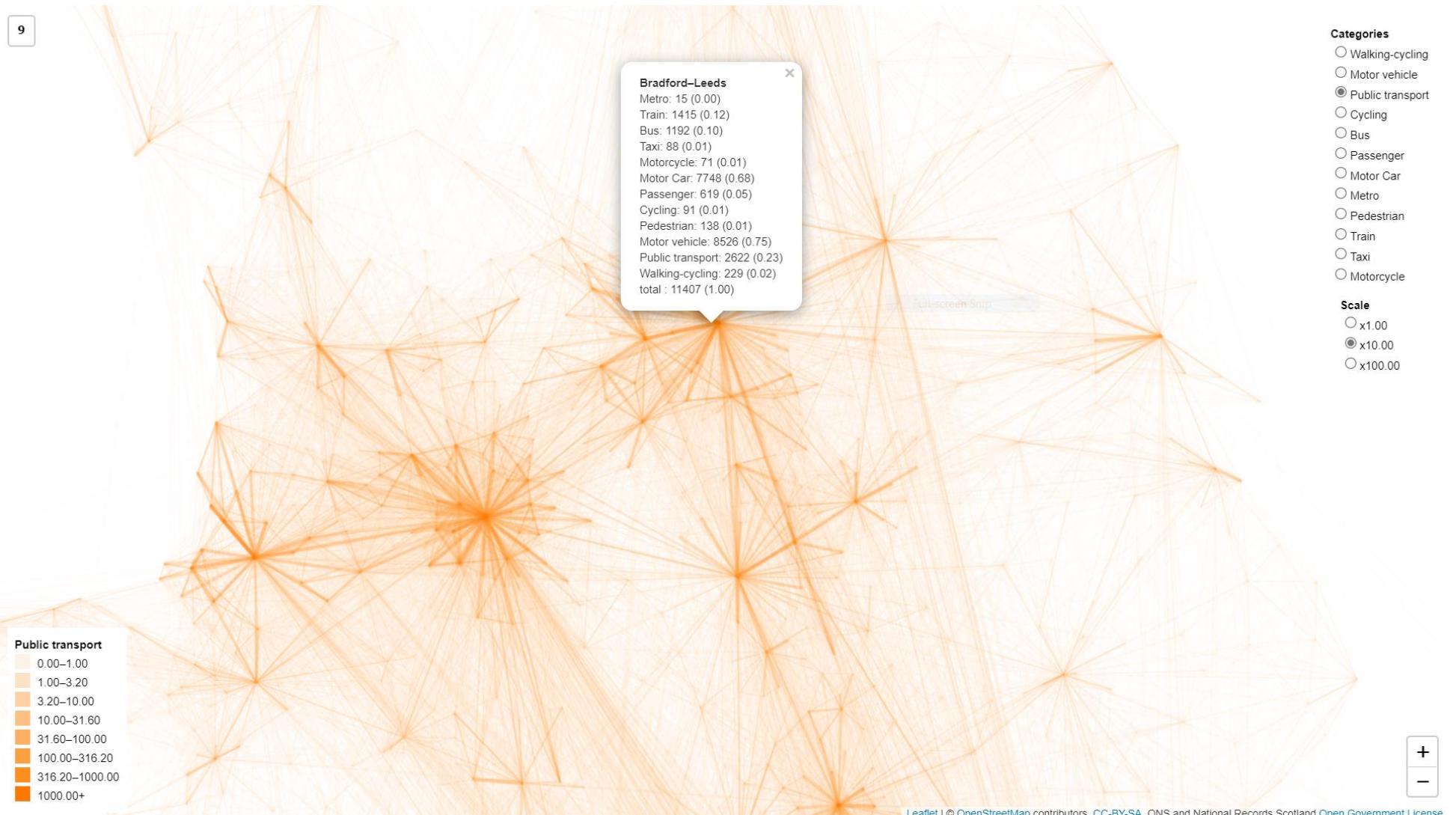
LEEDS DIGITAL FESTIVAL | 2021



Leaflet | © OpenStreetMap contributors, CC-BY-SA, ONS and National Records Scotland Open Government License



LEEDS DIGITAL FESTIVAL | 2021



# Where my Railway Goes

**Using 2011 Census geography data**

- station-location data from NaPTAN
- census SOA population data
- travel-flow MSOA population data

**Using python GeoPandas**

- To create datasets in GeoJSON and  
ESRI ShapeFile

These can be overlayed using QGIS



**LEEDS DIGITAL FESTIVAL | 2021**

# Where my Railway Goes

Alternatively these can be converted to **VectorTiles**

- Using **MapBox tippecanoe**
- Hosted on [github.com](#)
- Rendered using **Leaflet.js JavaScript**

To create interactive maps

- Density map is [HERE](#)
- Distance map is [HERE](#)
- Local travel flow map is [HERE](#)
- Inter-area travel flow map is [HERE](#)

A highway and rail map is [HERE](#)

This is work in progress



LEEDS DIGITAL FESTIVAL | 2021

# Questions and Answers



LEEDS DIGITAL FESTIVAL | 2021

# Open Data Feeds

NaPTAN:

- National Public Transport Access Nodes

Network Rail

- Open data-feeds (and mirror)

- National Electronic Sectional Appendix

Rail Development Group

- National Rail Enquiries data feeds

Department for Transport

- Journey Time Statistics

OpenStreetmap

- Editable geographic data

Ordnance Survey

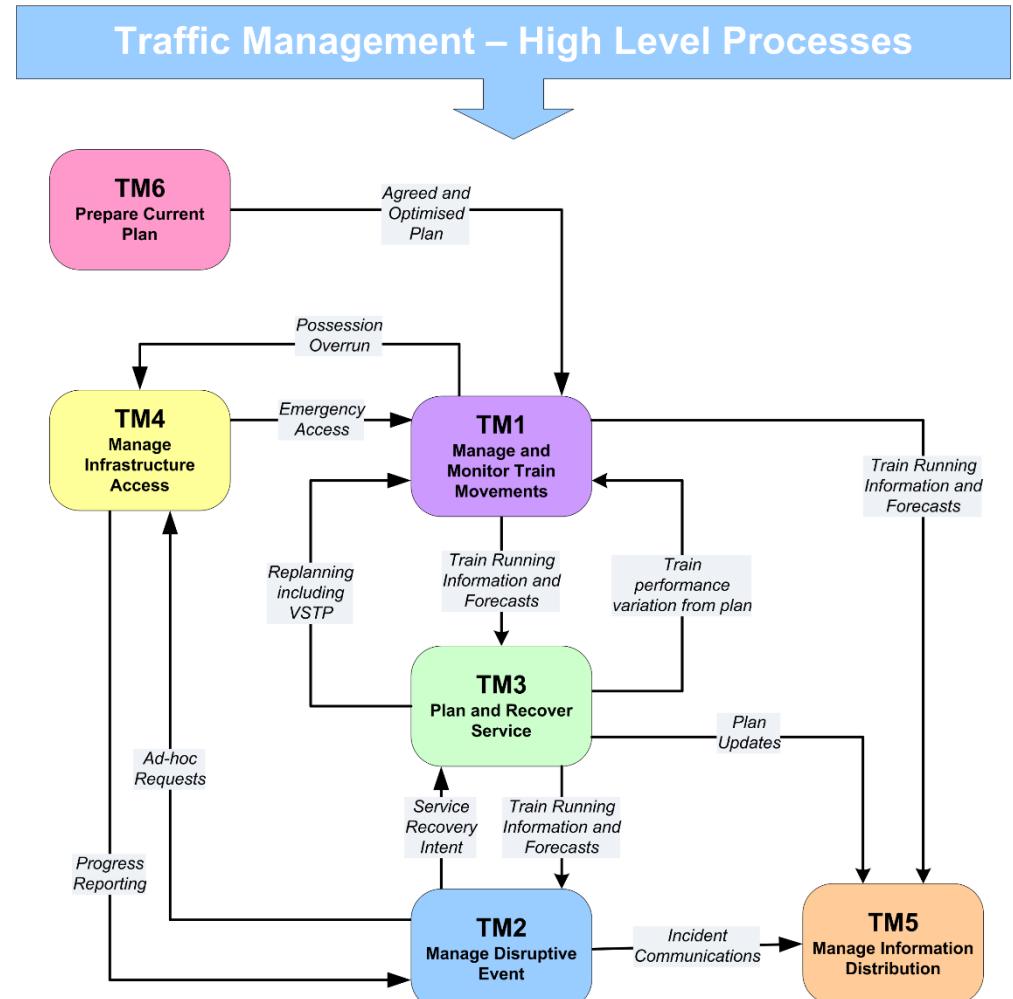
- Open data license data

Highways England

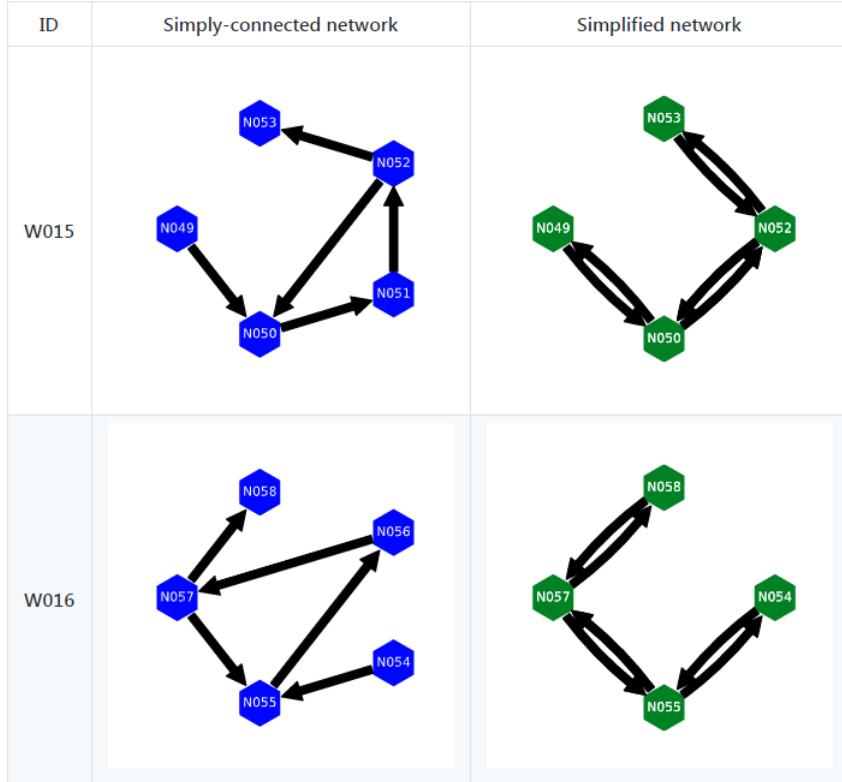
• Web API speed and flow statistics



LEEDS DIGITAL FESTIVAL | 2021



# Technology



## Tools

- [github.com](https://github.com)
- [gist.github.com](https://gist.github.com)
- [bl.ocks.org](https://bl.ocks.org)
- [d3js.org](https://d3js.org)
- [python.org](https://python.org)
- [gdal.org](https://gdal.org)
- [leaflet.js](https://leaflet.js)
- [./jq](#)
- [nodejs.org](https://nodejs.org)
- [Tesseract-OCR](https://Tesseract-OCR)
- [MapBox tippecanoe](https://MapBox/tippecanoe)
- [arangodb.com](https://arangodb.com)
- [postgresql.org](https://postgresql.org)
- [tabula.technology](https://tabula.technology)
- [tesseract-OCR](https://tesseract-OCR)
- [solr.apache.org](https://solr.apache.org)
- [qgis.org](https://qgis.org)
- [gdal.org](https://gdal.org)
- [python Pandas](https://python Pandas)
- [python GeoPandas](https://python GeoPandas)
- [python osmnx](https://python osmnx)



LEEDS DIGITAL FESTIVAL | 2021



LEEDS DIGITAL FESTIVAL | 2021





**Build it and  
they will  
come!**



**LEEDS DIGITAL FESTIVAL | 2021**

# Thank You



Will Deakin  
@WillDeakin1



LEEDS DIGITAL FESTIVAL | 2021

# Reference

- Image of 'rosemary in bloom' [picture](#), attributed to Margalob under [CC A-SA 4.0](#)
- International Klein Blue (#002FA7), Wikipedia article [here](#)
- Rail Alphabet: description and download [here](#)
- OSMnx: Boeing, G. 2017. [OSMnx: New Methods for Acquiring, Constructing, Analyzing, and Visualizing Complex Street Networks](#). *Computers, Environment and Urban Systems* 65, 126-139. doi:10.1016/j.compenvurbsys.2017.05.004

'It is better to ask some of the questions than to know all of the answers'

- James Thurber

My thanks go to Stuart Lowe, Garry Keenor, Gareth Dennis, my ever and long suffering family, all friends, colleagues and contributors to all the projects I have wilfully used for their help, support and advice

**Copyright 2021 Will Deakin**

**Licensed under the Apache License, Version 2.0 (the ‘License’);  
you may not use this file except in compliance with the License.  
You may obtain a copy of the License at**

**<http://www.apache.org/licenses/LICENSE-2.0>**

**Unless required by applicable law or agreed to in writing, software  
distributed under the License is distributed on an ‘AS IS’ BASIS,  
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.  
See the License for the specific language governing permissions and  
limitations under the License.**



**LEEDS DIGITAL FESTIVAL | 2021**