Mapping Accessibility: Using R5py and GTFS Data to Optimise Women's Health Hub Locations in the Sussex

By Karrie LIU
University of Sussex

Who is Karrie?

- **Experienced Mathematician and Data Scientist**: Over 15 years of expertise in healthcare analytics, data-driven strategy, and digital transformation.
- **Founder of Hypatia Analytics**: Established in 2019 to balance corporate consultancy with pro bono projects, delivering impactful solutions for both sectors.
- **Extensive NHS/Pharma Background**: A decade of experience working with NHS & Pharma organisations, driving operational efficiency and service optimisation.
- **Committed to Social Impact**: Trustee roles at AgeUK Wandsworth and Maths on Toast, focusing on promoting numeracy and empowering communities.
- **Proven Leader and Strategist:** Skilled in governance, mentoring, and bridging academic and industry needs to drive innovation and foster collaboration.
- **Advocate for Diversity and Inclusion**: Mentor for underrepresented groups in STEM, fostering equity through initiatives like the "Discover Data" programme.

Achievements:

- *Winner for* WISE Outstanding Woman in Science Award 2024,
- Mulan Award : Contribution to Science 2024,
- *Winner of* She Inspires awards *(Women in STEM),*
- *40 Under Forty Maths and statistics),*
- *Finalist of* Asian women achievement awards: Science
- *Finalist of* British Data award *(Rising Star of the Year)*
- *Top 50 Future List of* Northern Power Women Award *in 2023*
- *Winner of* Top 100 Women in Tech 2019 #TechWemon100

HYPATIA ANALYTICS

Karrie:

Hypatia

- For corporate clients, such as **NHS providers and pharmaceutical companies**, I lead projects that use **predictive modelling and operational analytics** to optimise services.
- During the COVID-19 pandemic, I helped the **NHS tackle patient backlogs**, building data models that improved resource allocation.
- Applying **forecasting and segmentation** to enhance brand planning and marketing strategies.

Data for Good

Through **pro bono projects**, I support charities like **AgeUK Wandsworth**, helping them **improve services for vulnerable older adults** by leveraging data insights.

Data Volunteer Team, where professionals contribute their skills to **non-profits and charities**, offering data solutions they wouldn't otherwise have access to.

Discover Data, a programme aimed at **inspiring young people** to explore careers in data science through hands-on workshops.

PhD Maths Student

My research focuses on **mathematical modelling and operational research** to improve healthcare delivery.

My current project examines **Women's Health Hubs in Sussex**, aim to **reduce health inequalities** and improve access to care.

Through **data-driven simulations**, I aim to provide **actionable insights for policymakers** to enhance healthcare services and patient outcomes.

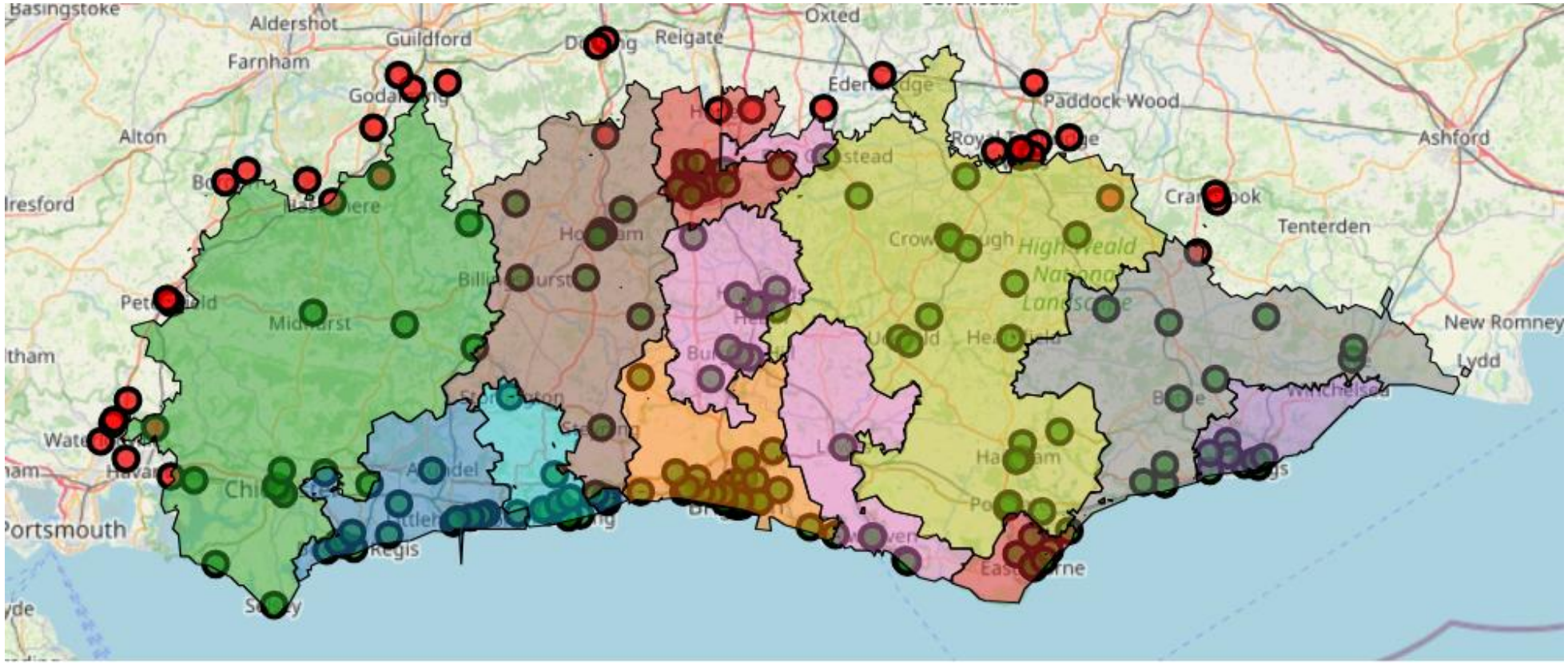
My academic work reinforces my belief that **data science can be a force for good**, solving real-world problems.

HYPATIA ANALYTICS

Placement of Women's Health Hubs

Joint work with the NHS

Question : where should Women's Health Hubs be located to serve communities most effectively?



HYPATIA ANALYTICS

Consider : What is the demand?

One consideration regarding demand is patients' ability to reach WHH. Rather than basing analysis solely on current GP attendance, we aim to identify potential demand, patients who can reasonably access a GP within a given travel time.

Patient Population:
All postcodes within a 30-minute public transport journey of a GP practice.

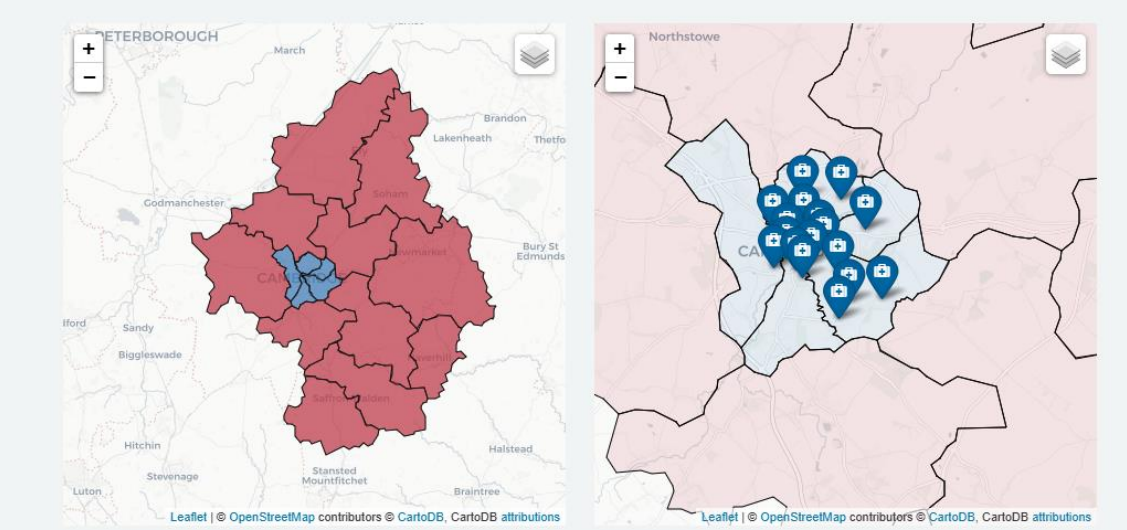
Assumptions:
Travel times are calculated using an algorithm assuming a typical Wednesday at 10 a.m.

HYPATIA ANALYTICS

Research Topic :

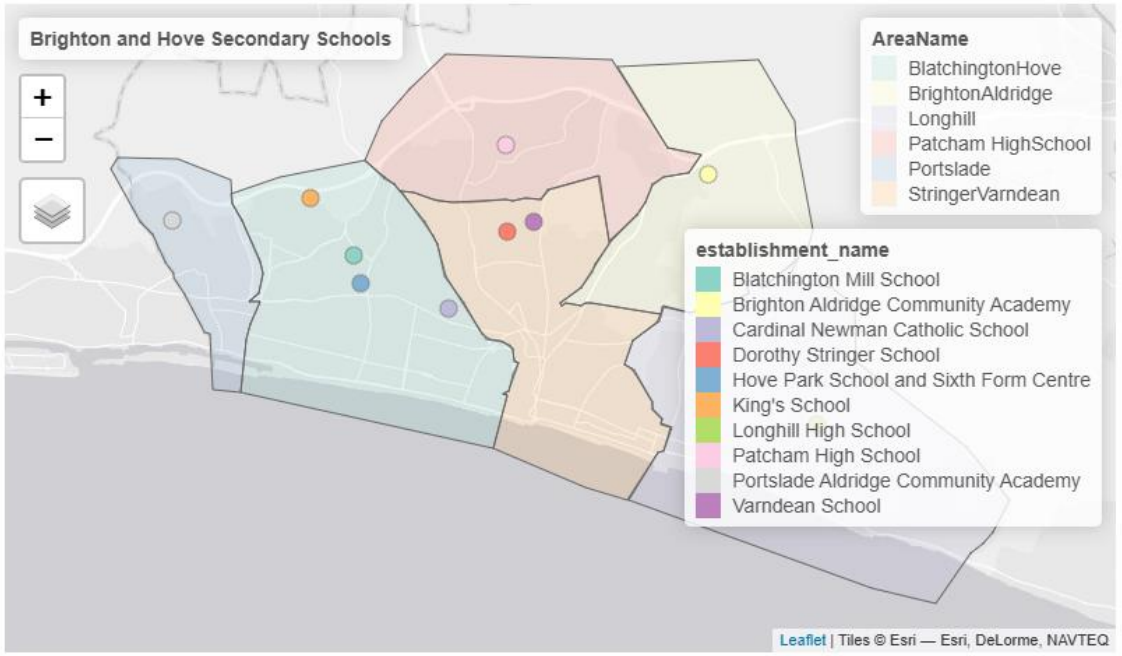
Travel time to NHS organisations

by NHS England Mattia Ficarelli, Data Engineer, and Paul Carroll, Senior Data Scientist,



Brighton secondary school admissions

by Adam Dennett Professor of Urban Analytics, UCL

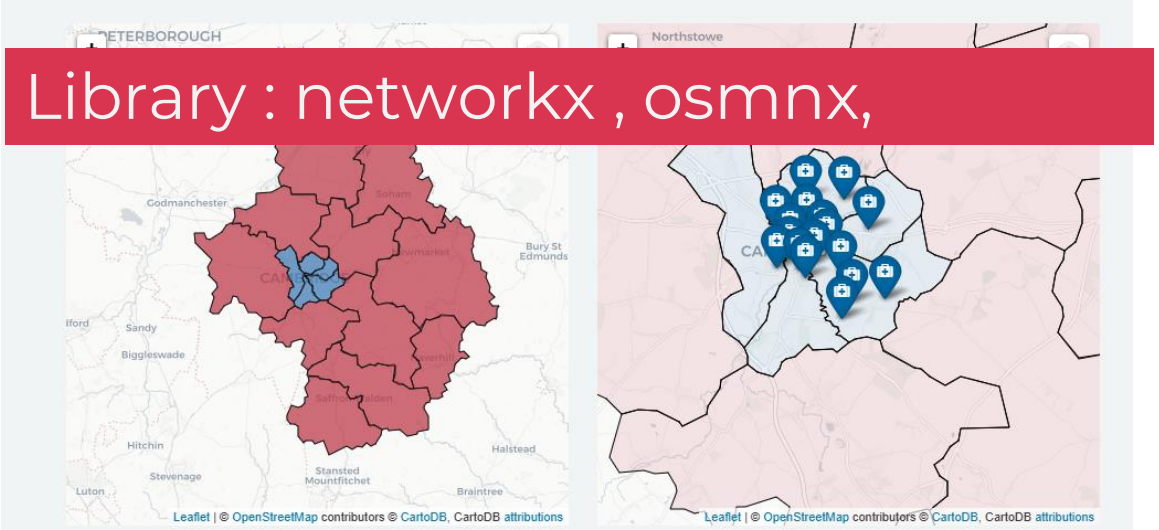


HYPATIA ANALYTICS

Research Topic :

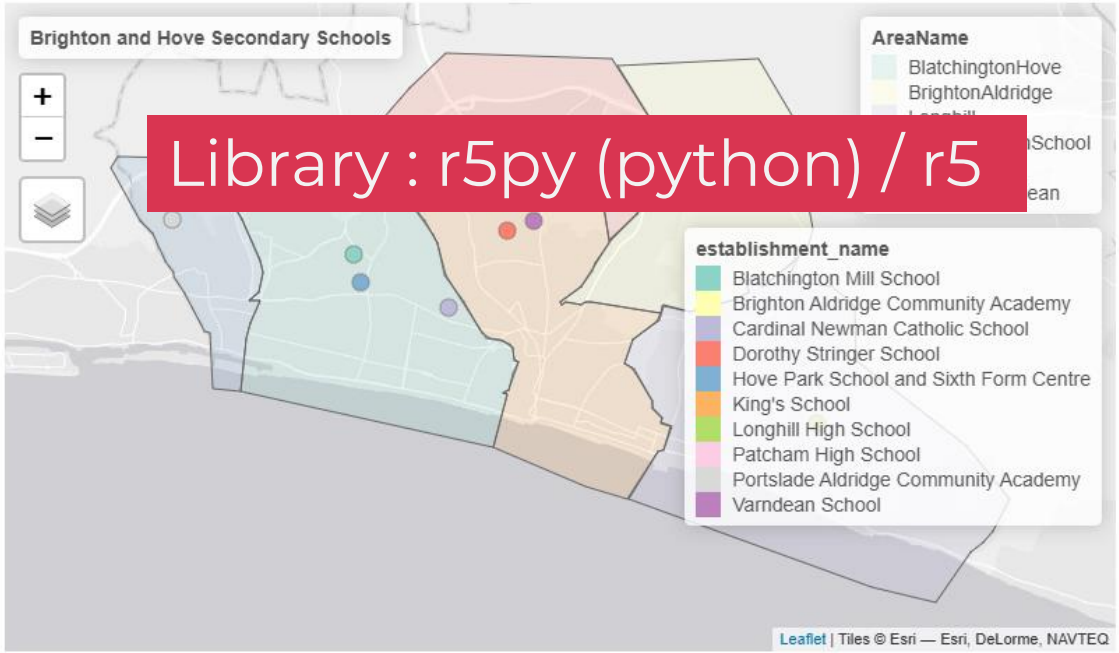
Travel time to NHS organisations

by NHS England Mattia Ficarelli, Data Engineer, and Paul Carroll, Senior Data Scientist,



Brighton secondary school admissions

by Adam Dennett Professor of Urban Analytics, UCL



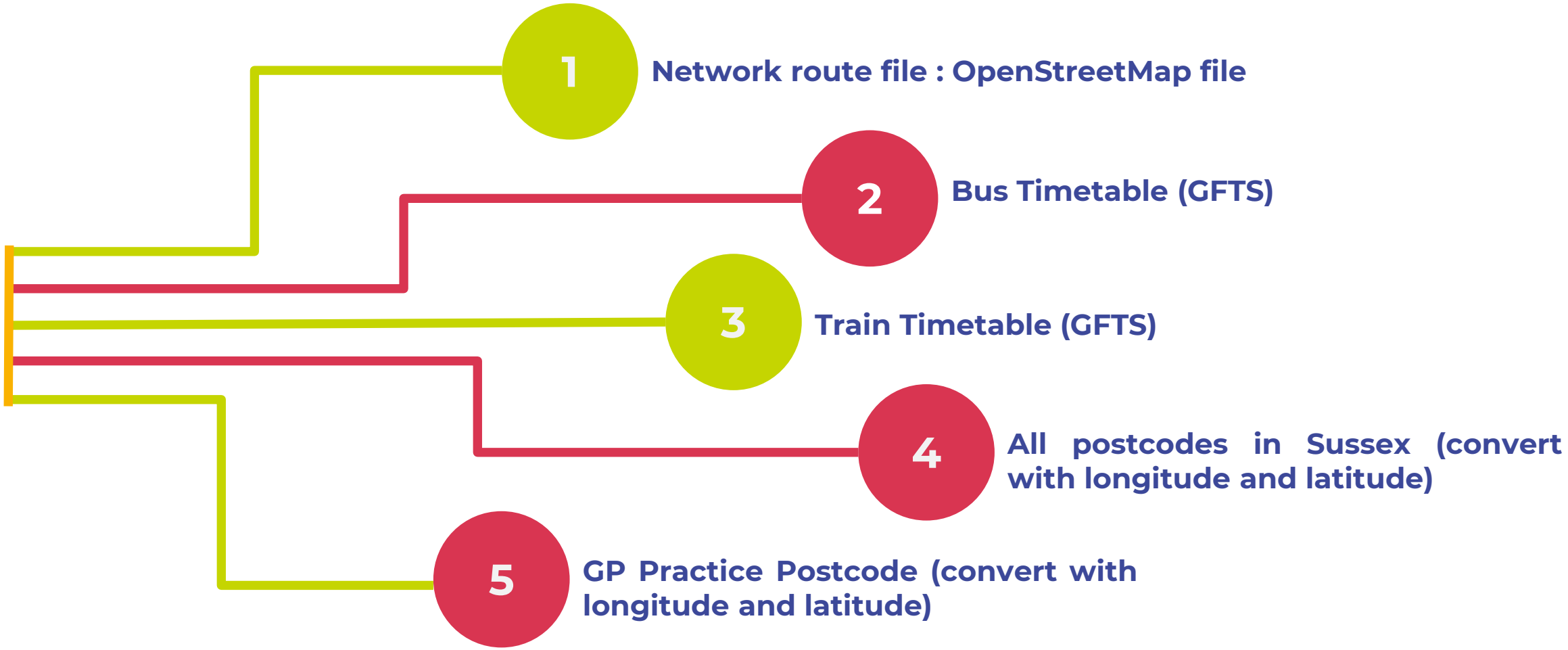
HYPATIA ANALYTICS

Library : r5py

- **R⁵py** is a Python library for rapid realistic routing on multimodal transport networks (walk, bike, public transport, and car). It provides a simple and friendly interface to R⁵, the Rapid Realistic Routing on Real-world and Reimagined networks, a routing engine developed by Conveyal. *r5py* is inspired by r5r, a wrapper for R, and it is designed to interact with GeoPandas GeoDataFrames.

- *R5py* offers a simple way to run R⁵ locally with Python, allowing the users to calculate travel time matrices and accessibility by different travel modes

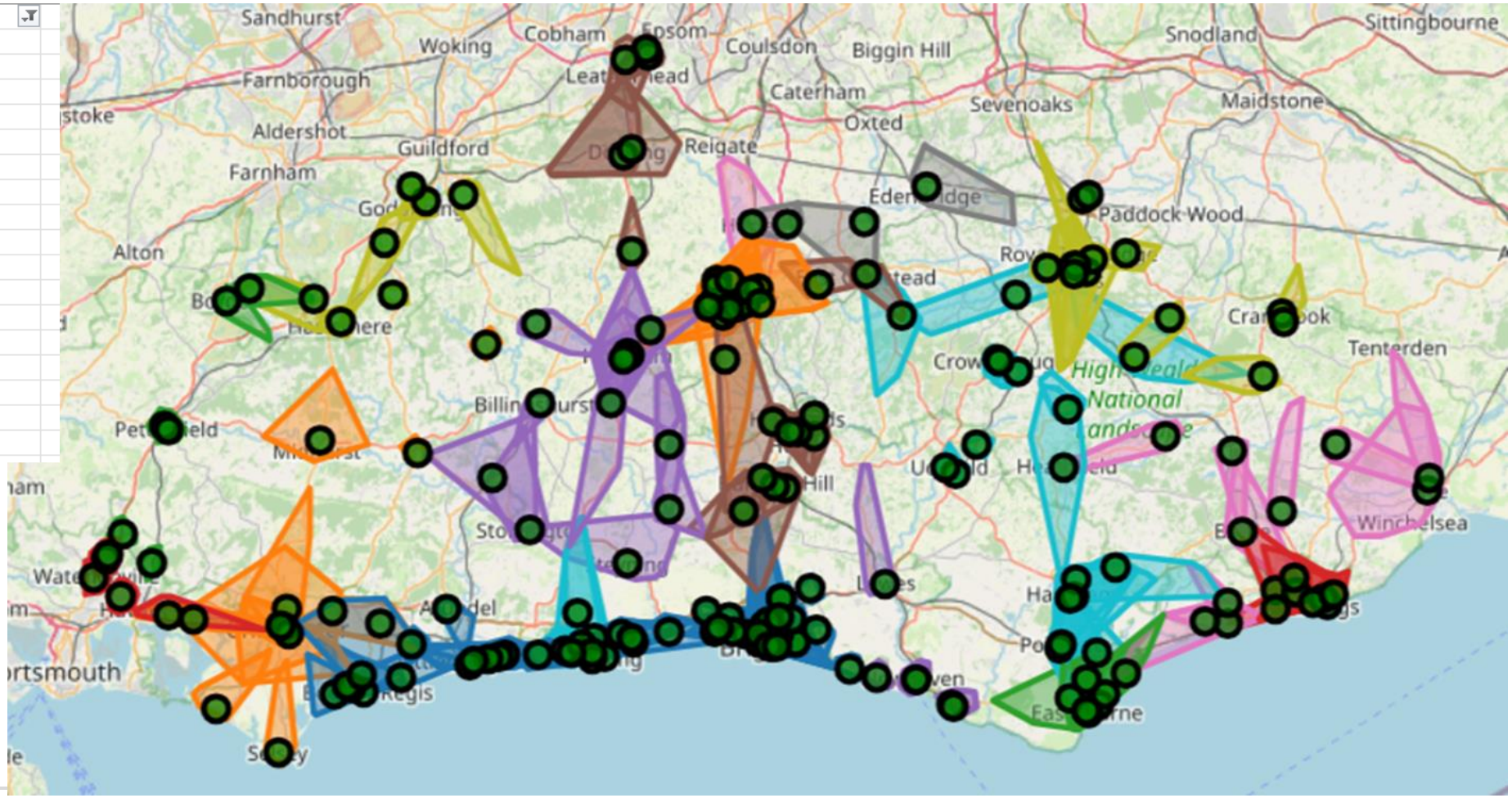
What kind of data we need?



HYPATIA ANALYTICS

Output : 30 minute transit time

from_id	to_id	travel_time
PO10 7AA	H82043	101
PO10 7AA	H82020	87
PO10 7AA	H82038	99
PO10 7AA	H82077	79
PO10 7AA	H82058	98
PO10 7AA	H82099	101
PO10 7AA	H82016	84
PO10 7AA	H82048	117
PO10 7AA	H82021	103
PO10 7AA	H82037	87
PO10 7AA	H82006	103
PO10 7AA	H82095	93
PO10 7AA	H82032	110
PO10 7AA	H82051	64
PO10 7AA	H82078	27
PO10 7AA	H82067	82
PO10 7AA	H82013	44



HYPATIA ANALYTICS

Library : r5py

- Pros:
  1. Multimodal routing: Supports walking, cycling, driving, and especially public transport using GTFS.
  2. Detailed travel time & accessibility analysis: Supports time-dependent routing, isochrones, and accessibility maps.
  3. Built for transport planning: Ideal for cities, transit authorities, or accessibility researchers.
  4. High performance: Backed by R5 (Java) engine, optimized for large networks.

- Cons:
  1. Complex setup: Requires Java and understanding of GTFS data formats.
  2. Less customizable graphs: Focused on routing and accessibility, not general graph theory.
  3. Smaller community: Less general-purpose usage than networkx, and fewer tutorials/examples.

Library : networkx / osmnx

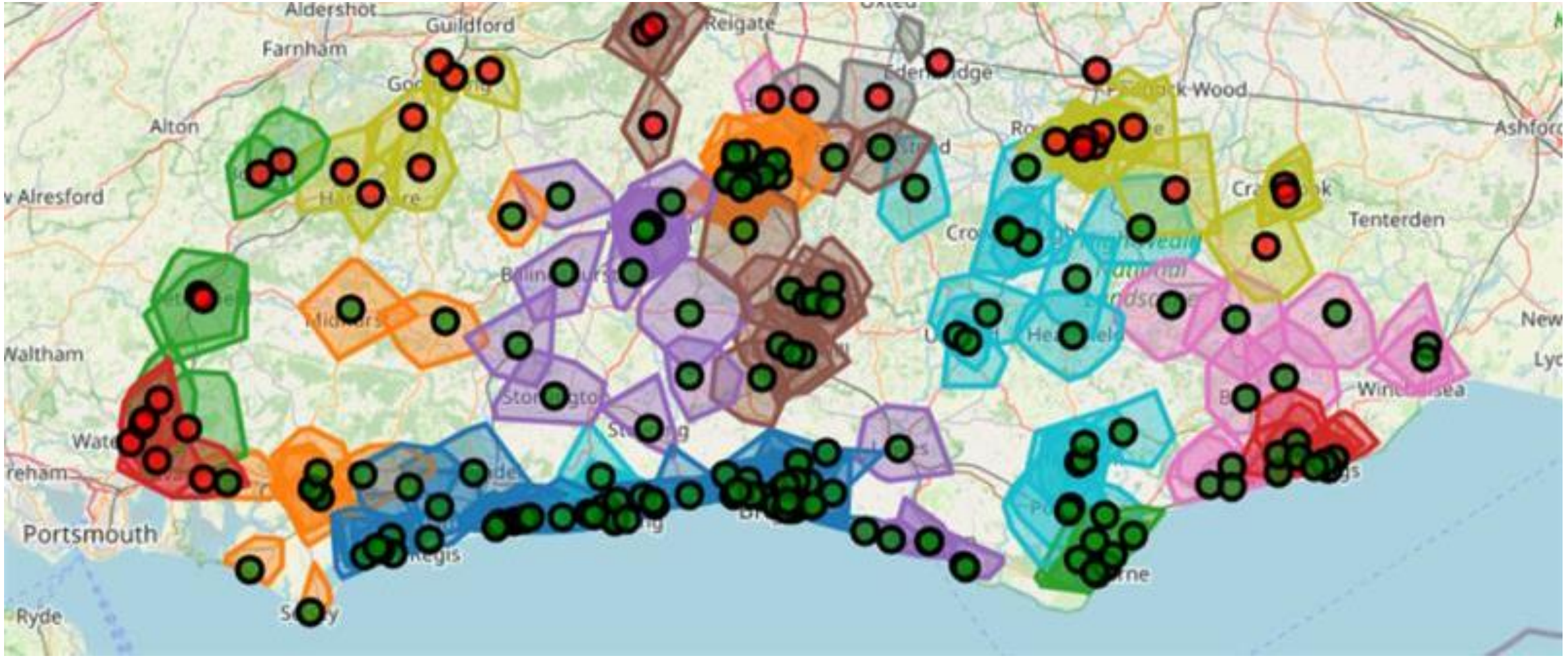
- **NetworkX** is a Python library for the creation, manipulation, and study of complex networks and graphs. It provides a powerful and flexible framework for analyzing the structure, dynamics, and functions of networks such as social connections, transportation systems, or communication networks.

- With **NetworkX**, users can easily construct graphs (directed, undirected, or multigraphs), compute network statistics, and run a wide range of algorithms — from shortest paths and centrality measures to community detection and flow analysis.

- **OSMnx** is a Python library for acquiring, constructing, analyzing, and visualizing street networks and other spatial data from **OpenStreetMap**. It provides a simple and intuitive interface for downloading real-world geospatial networks — such as roads, walking paths, and bike lanes — and representing them as **NetworkX** graph objects.

- With **OSMnx**, users can perform advanced spatial and network analysis directly in Python, including routing, network centrality computation, and accessibility studies. It also supports visualization of networks, building footprints, and urban forms with just a few lines of code.

GP Practice catchment : 10 mins drive



HYPATIA ANALYTICS

Library : networkx / osmnx

- Pros:
  1. Easy to use and well-documented: Pure Python with a rich ecosystem and readable syntax.
  2. Custom graph modeling: networkx allows full flexibility to build and analyse any type of graph.
  3. OSM integration: osmnx simplifies downloading and visualizing street networks directly from OpenStreetMap.
  4. Great for visualization: Built-in plotting, strong integration with matplotlib and folium.

- Cons:
  1. Limited transport mode support: Mostly road networks—public transit, schedules, or multimodal analysis are not handled.
  2. Lacks high-performance routing: Suitable for small to medium datasets, but slow on large-scale or time-sensitive routing.
  3. No support for GTFS data: Can't natively analyze transit schedules or frequency-based service.

Library : networkx / osmnx vs r5py

Aspect	networkx / osmnx	R5py
Main Focus	Graph theory, urban network analysis	Multimodal transport routing & accessibility
Supported Modes	Walking, driving, cycling (static)	Walking, cycling, driving, **public transit (GTFS)**
Routing Type	Static, shortest-path based	Time-dependent, multimodal, schedule-aware
GTFS Support	Not supported	Full GTFS support
Performance	Python-native, slower on large networks	High-performance (R5 Java backend)
Use Case Fit	Custom graph analysis, network visualization	Accessibility analysis, transit planning, large-scale models

Recommended
Use networkx / osmnx for:Urban network modeling, visualization, general-purpose graph analysis.
Use R5py for:Realistic, multimodal transport routing and accessibility studies with GTFS.

HYPATIA ANALYTICS

Abstract

- Background: Women in the UK experience substantial health inequalities, with more years spent in ill health and disability. Establishing optimally located Women's Health Hubs (WHHs) is essential for improving service delivery and reducing these disparities. In order to determine the optimal location, we needed to quantify the patient catchment area of each GP in the ICB.

- Methods: We defined the patient catchment as the set of postcodes that were within 30 minutes using public transport. In order to do this, we employed the package R5py, which is a free package in Python based on an equivalent package in R (called R5R) , that uses General Transit Feed Specification (GTFS) data, also publicly available , to find the optimal route from a start node to an end node using public transport journeys and walking. In this talk, I will demonstrate what GTFS data is, how it is obtained, how to use R5py and compare the methodology to other well-known route-optimising tools.

- Findings: R5py's transit-time modelling demonstrated marked differences between accessibility by public transport (by bus) and by car, highlighting disparities in provision across areas such as Hastings, Rother, and Chichester. The limitations of the current work are that we only consider the bus network, due to challenges with train data that will be explained, and it is computationally heavy. In terms of computational performance, the simulation was relatively quick: once the map was loaded (Which takes about an hour in the first time, loading the GTFS bus data was significantly faster about 2 mins, then create the network map taking around 2 minutes, and the bus simulation itself required 60 minutes to compute each catchment.

- Impact and Next Steps : This R5py-enabled approach provides a rigorous, data-driven basis for strategic service planning, emphasising genuinely accessible catchments. It is the main component of an optimisation algorithm which aims to mathematically determine the next optimal locations of the Women's Health Hubs in Sussex.

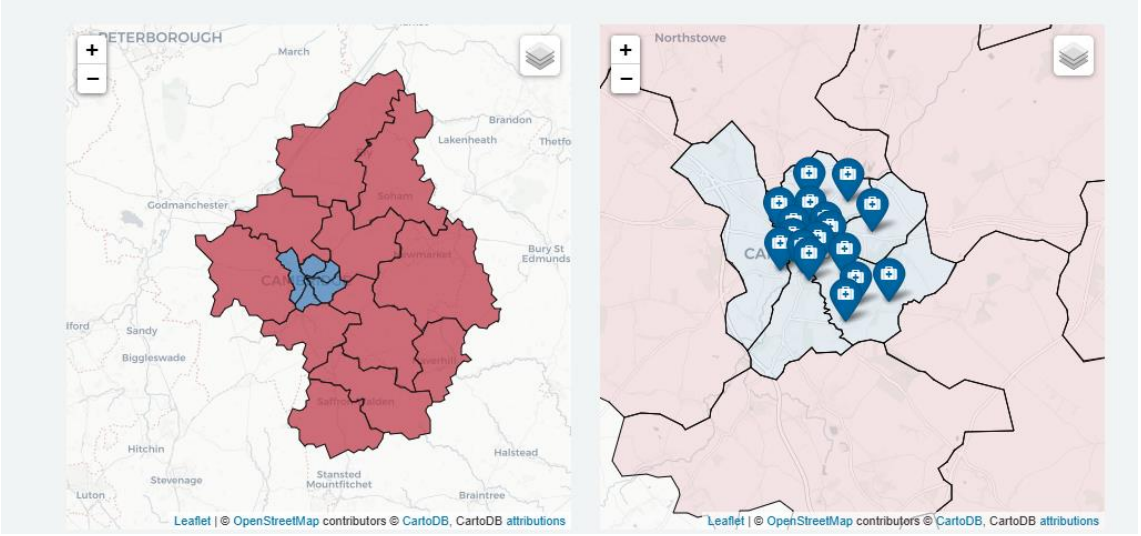
HYPATIA ANALYTICS

Travel time to NHS organisations

by NHS England Mattia Ficarelli, Data Engineer, and Paul Carroll, Senior Data Scientist,
https://github.com/nhsx/nhs_time_of_travel/blob/main/walking_gp_practice_cambridge.ipynb

- It is a proof-of-concept / exploratory analytical tool, built using open source software and publicly available NHS and transport data, not an official policy site.

- The project models and visualises patient travel times to NHS services across London, Cambridge, and Lincolnshire. It maps 5-, 10-, and 20-minute walking zones around Cambridge GPs to reveal access gaps and refines London driving times using Uber Movement data to reflect real-world traffic conditions accurately.



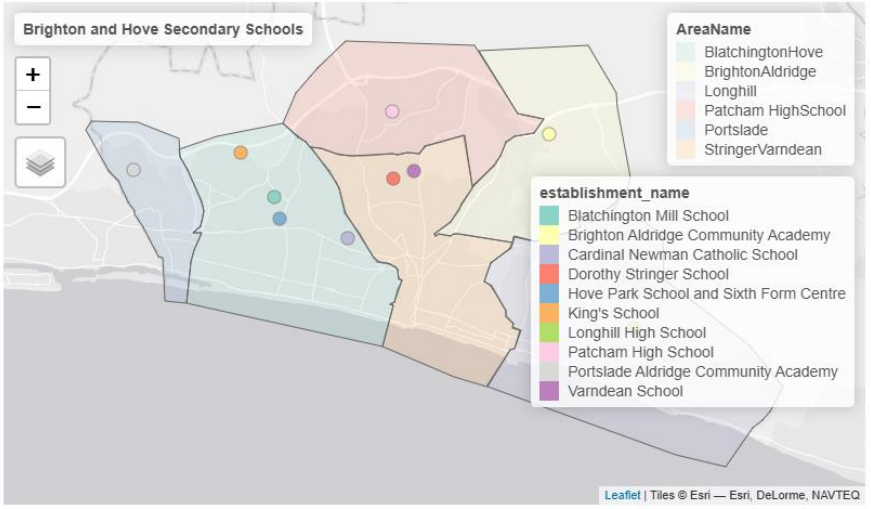
HYPATIA ANALYTICS

Brighton secondary school admissions

by Adam Dennett Professor of Urban Analytics, UCL

https://adamdennett.github.io/BH_Secondary_Admissions_Analysis/BH_Sec_Sch_Analysis.html

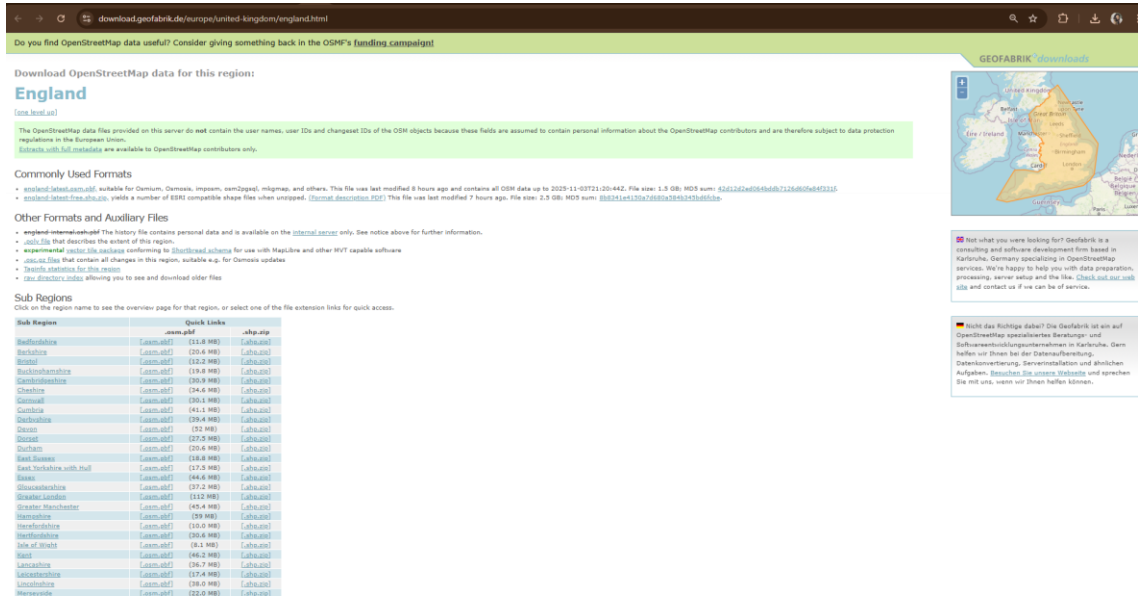
- The analysis critiques the Council's proposed reductions in PANs (Published Admission Numbers) across Brighton & Hove secondary schools, arguing many cuts are unjustified given current demand.
- The author argues there is no need to redraw catchment boundaries in 2026; instead, better redistribution of existing capacity (via PAN adjustments) could meet projected pupil numbers.



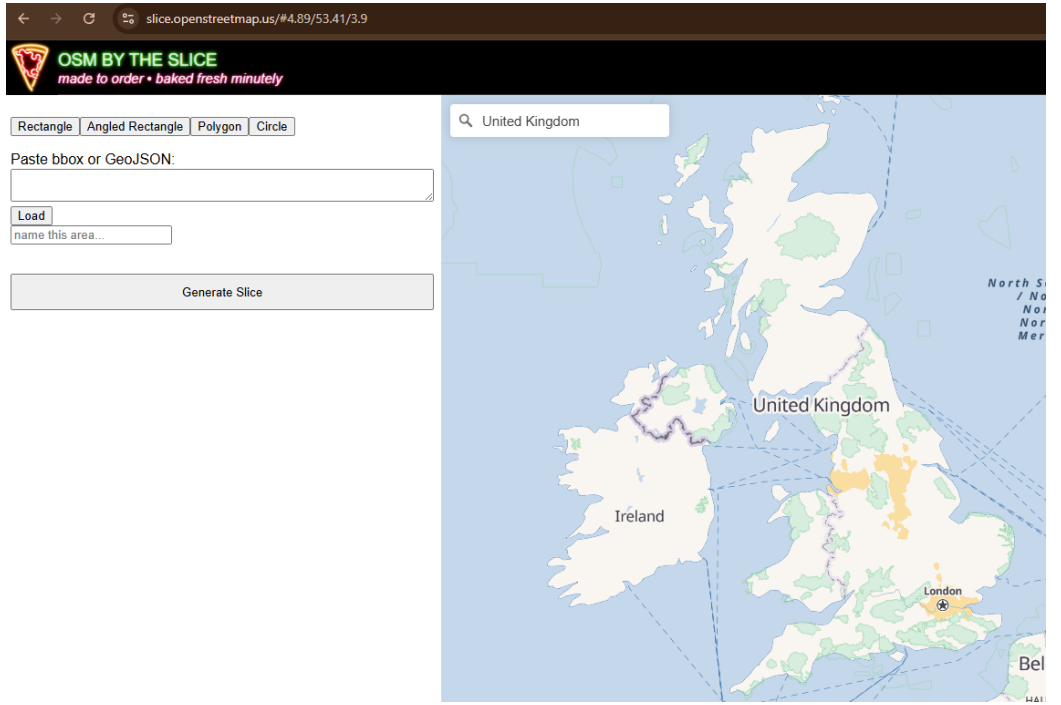
HYPATIA ANALYTICS

Dataset : OpenStreetMap file (osm_file)

https://download.geofabrik.de/europe/united-kingdom/england.html



https://slice.openstreetmap.us/#4.89/53.41/3.9



HYPATIA ANALYTICS

Dataset : Bus Timetable (GTFS data)

https://data.bus-data.dft.gov.uk/timetable/download/

Download regional and national data sets in GTFS format

You can download timetables data in GTFS format for the whole of GB, or filter for specific regions. This dataset is created using compliant data published to the Bus Open Data Service. Where data is not yet published on Bus Open Data, or is published but not compliant, this data is supplemented with Traveline National Dataset (TNDS).

All - Download timetables data in GTFS format

England - Download timetables data in GTFS format

East Anglia - Download timetables data in GTFS format

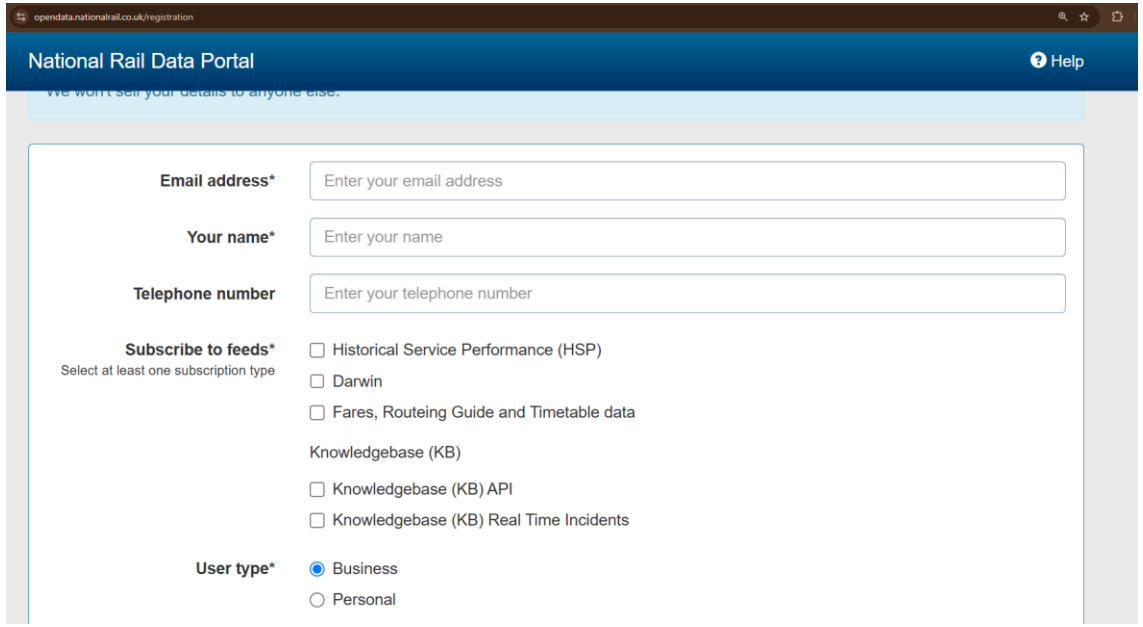
East Midlands - Download timetables data in GTFS format

HYPATIA ANALYTICS

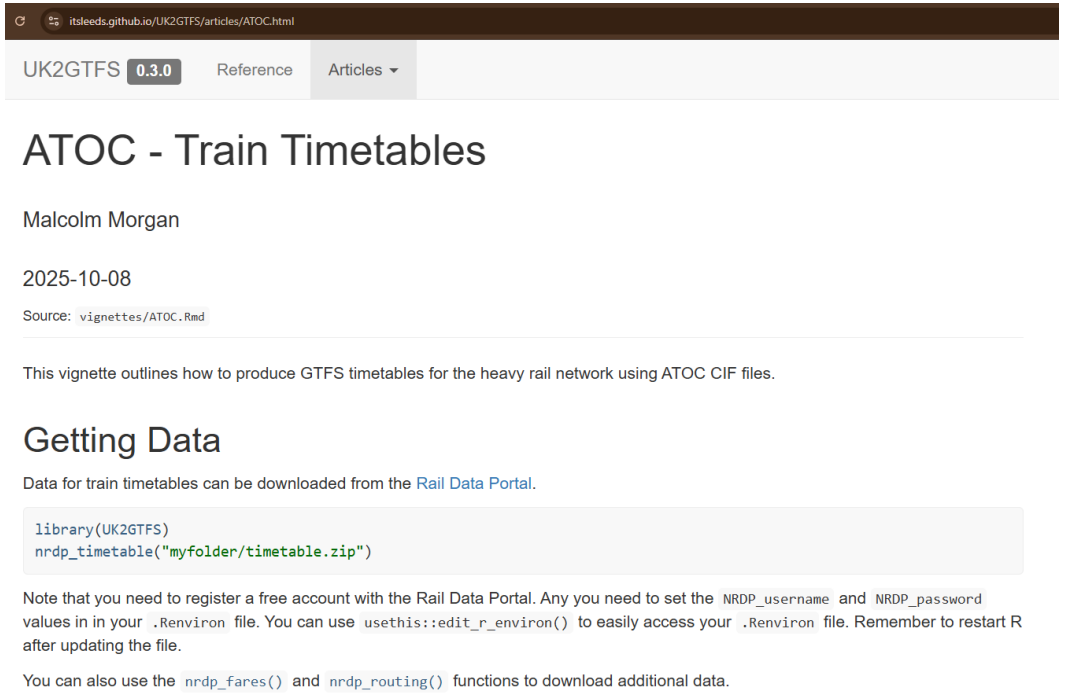
Dataset : Train Timetable

https://opendata.nationalrail.co.uk/registration

National Rail API:



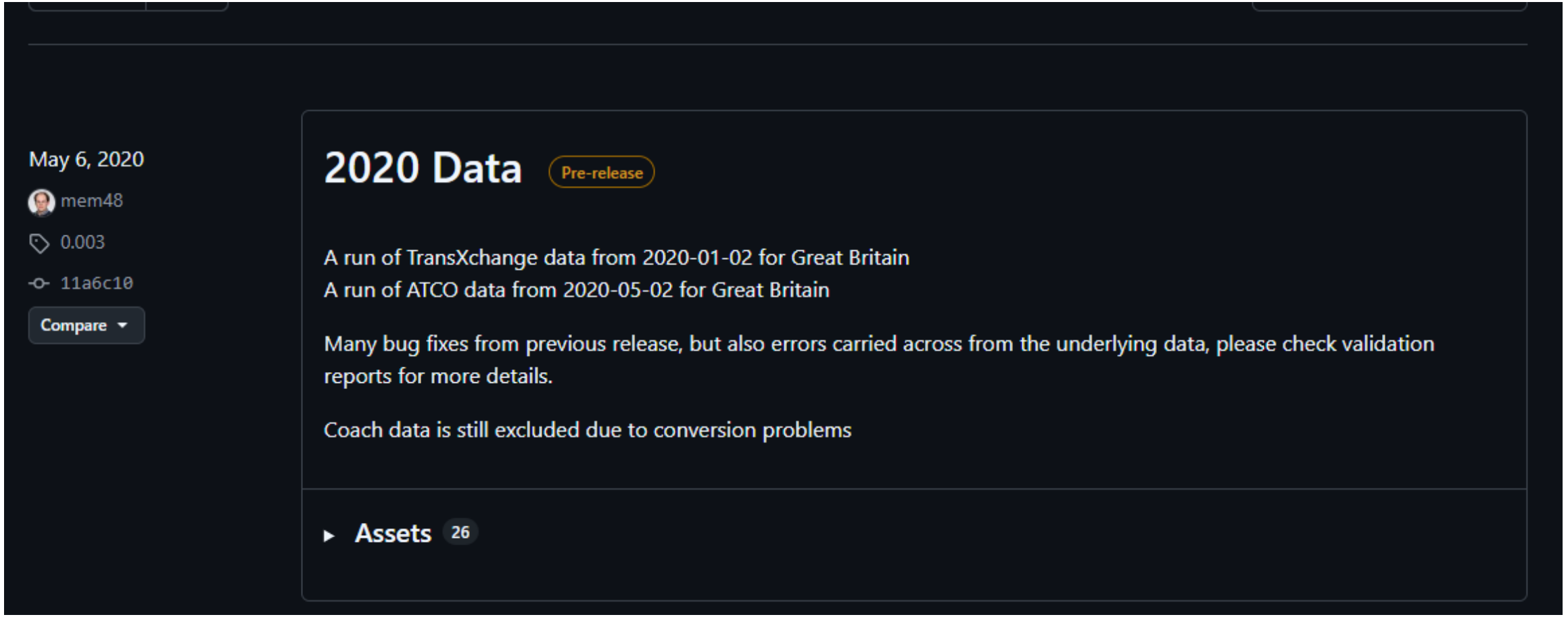
Should be convertible following UK2GTFS guide



HYPATIA ANALYTICS

Dataset : Train Timetable (GTFS data)

https://github.com/ITSLeeds/UK2GTFS/releases



HYPATIA ANALYTICS

Analysis: TravelTimeMatrix

```
# Compute travel time matrix
travel_times = TravelTimeMatrix(
    transport_network=tn,
    origins=origins,
    destinations=destinations,
    transport_modes=[TransportMode.TRANSIT, TransportMode.WALK],
    departure=datetime.datetime(2025, 6, 25, 10, 0),  # Optional: specify departure time
    snap_to_network=True,
)
```

1. To create the Transport Network with the map note, and the public transport file.

- **tn = TransportNetwork(osm_file, [bus_GTFS, Train_GTFS])**

  - osm_file ="east-sussex-latest.osm.pbf" #- 1m
  - osm_file ="west-sussex-latest.osm.pbf" #- 3m 12s
  - osm_file ="united-kingdom-latest.osm.pbf" #- 57m7s

Analysis: TravelTimeMatrix

```
# Compute travel time matrix
travel_times = TravelTimeMatrix(
    transport_network=tn,
    origins=origins,
    destinations=destinations,
    transport_modes=[TransportMode.TRANSIT, TransportMode.WALK],
    departure=datetime.datetime(2025, 6, 25, 10, 0),  # Optional: specify departure time
    snap_to_network=True,
)
```

2. **Define origins:** Include all postcodes within the Sussex area (and some neighbouring border areas) and convert them into a GeoDataFrame, using longitude and latitude coordinates derived from each postcode. In total, there are 82,842 postcodes.

Analysis: TravelTimeMatrix

```
# Compute travel time matrix
travel_times = TravelTimeMatrix(
    transport_network=tn,
    origins=origins,
    destinations=destinations,
    transport_modes=[TransportMode.TRANSIT, TransportMode.WALK],
    departure=datetime.datetime(2025, 6, 25, 10, 0),  # Optional: specify departure time
    snap_to_network=True,
)
```

3. **Define destinations**: In this project, we have included 156 GP practices within the Sussex area, along with 40 additional GP practices located near the border. These have been converted into a GeoDataFrame using longitude and latitude coordinates derived from each postcode.

Analysis: TravelTimeMatrix

```
# Compute travel time matrix
travel_times = TravelTimeMatrix(
    transport_network=tn,
    origins=origins,
    destinations=destinations,
    transport_modes=[TransportMode.TRANSIT, TransportMode.WALK],
    departure=datetime.datetime(2025, 6, 25, 10, 0),  # Optional: specify departure time
    snap_to_network=True,
)
```

4. **Transport_Modes**: Using *TransportMode.TRANSIT* and *TransportMode.WALK*, this project models travel from a patient's postcode to a GP practice. It includes walking to transit stops, the transit journey itself, or walking directly to the GP. The quickest overall travel time between the two points is used for analysis.

5. **Departure :** This is useful because different people require different travel times. In this project, we focus on patients who prefer to travel outside school hours, when children are not unavailable, so we use non-peak times to better reflect their likely travel behaviour.

Analysis: DetailedItineraries

```
detailed_itineraries = DetailedItineraries(
    transport_network = tn ,
    origins=origins,
    destinations=destinations,
    #departure=datetime.datetime(2022, 2, 22, 8, 30),
    transport_modes=[TransportMode.TRANSIT, TransportMode.WALK],
    snap_to_network=True,
)
```

✓ 307m 48.3s

- Recording every single step, however it was mentioned that they are not using, unless completely necessary. It tooks more 300 mins to run area of Arun.
-

	from_id	to_id	option	segment	transport_mode	departure_time	distance	travel_time	wait_time	feed
0	BN14 9JU	H82023	0	0	TransportMode.WALK	NaN	6607.434000	0 days 01:50:55	0 days 00:00:00	NaN
1	BN14 9JU	H82023	1	0	TransportMode.WALK	2025-06-20 11:46:06	2992.278000	0 days 00:50:23	0 days 00:00:00	NaN
2	BN14 9JU	H82023	1	1	TransportMode.BUS	2025-06-20 12:39:00	1258.921973	0 days 00:04:00	0 days 00:06:00	itm_south_east_gtfs

Thank you!

/Karrie.Liu

@HypatiaAnalytics

Karrie@HypatiaAnalytics.com

HYPATIA ANALYTICS

HYPATIA ANALYTICS