Adam Dennett, Dissertation Topics 2020/21

Topic 1 – UK Housing Research

Sub-Projects

1. Machine Learning Approaches to estimating small-area house prices for different types of property in England and Wales.
   * Data on house prices in the UK are rich but sparse at the small-area level over space and time which makes it difficult to draw useful comparisons where stock heterogeneity is high and
   * Using a newly linked dataset of price and attribute data (see Chi et al. 2020), this project will explore methods for accurately estimating small-area (neighbourhood) level representative house prices over a 10 year period, both in terms of raw house prices and price per square metre.
   * Machine learning approaches are less concerned with interpretation, more with the accuracy of the estimates. We would expect this project to compare and evaluate the accuracy of different methods in different geographical contexts, making recommendations for new approaches that can be adopted in the future.
   * One other idea could be to compare machine Learning VS spatio-temporal modelling of small-area data in R – which method performs the best?
   * Data: <https://reshare.ukdataservice.ac.uk/854240/>
2. How does the quality of housing stock interact with dimensions of social and economic deprivation to compound multiple deprivation?
   * The English indices of deprivation

* Understanding the interaction between housing quality and aspects of deprivation - EPC data. To what extent is multiple deprivation or Health exacerbated by poor quality (energy inefficient) housing in England and Wales?

1. Visualising Housing Affordability and England and Wales
   * This project will work with a new linked dataset of price and attribute data (see Chi et al. 2020), and will explore methods for parameterising and visualising the geography of housing affordability over time in England and Wales
   * Parameters that vary by small area might include:
     1. Average Household Income
     2. Size of mortgage available based on income (a multiplier – x4 etc.)
     3. Deposit (amount of money in the bank available to contribute towards the cost of the purchase – a fixed number)
     4. Interest rates (which affect the level of monthly housing costs)
     5. Proportion of monthly income spent on housing costs
   * The main outputs for the project will be an interactive visualisation tool, however this will be use to explore substantive issues around affordability and access to housing in England and Wales.
   * Can we improve on this - <https://www.bbc.co.uk/news/business-23234033> and this <https://www.ons.gov.uk/peoplepopulationandcommunity/housing/articles/alternativemeasuresofhousingaffordability/financialyearending2018> (figure 8)

Data: <https://reshare.ukdataservice.ac.uk/854240/>

1. To what extent are renters disadvantaged more than owners in the UK? What are the associations between property values and rental prices? Normalise by size - Rent and Price per square metre.
   * Urban Big Data Centre Zoopla Data- <https://www.ubdc.ac.uk/data-services/data-catalogue/housing-data/zoopla-property-data/>
   * House Price Data: <https://reshare.ukdataservice.ac.uk/854240/>

Topic 2 –Geographies of retail, groceries, food security and nutrition

1. How does the diversity of grocery supermarket choice within towns and cities interact with online grocery shopping behaviours?

* A new Index of grocery diversity for the whole of England and Wales. Using data from Geolytix, Open Street Map and other sources, can a new national index of grocery diversity be devised

1. Food Deserts and Food Insecurity in London. Partner Project with Feast With Us -

<https://www.feastwithus.org.uk/our-mission> -

* + What are the geographies of food poverty, nutrition, malnutrition and access to food within London (revealed by Tesco data) and where are the potential areas of most need when trying to tackle this issue?
  + What are the most important characteristics associated with those individuals experiencing severe food poverty who access Feast’s services and how do these relate to nutrition geographies revealed by Tesco data?
  + Where are Feast’s services most likely to be required in the future?
  + Feast will provide access to a detailed service evaluation survey of people using their services in London (50+ respondents, 92 variables). Contains demographic information and data around diet, cooking skills, eating habits, etc.
  + Data from Tesco on clubcard-linked purchases from 411 stores in London in 2015 - <https://www.nature.com/articles/s41597-020-0397-7> - detailed nutritional information by LSOA

1. A new Food Consumption and Nutrition Classification for London
   * This project will make use of the extensive LSOA-level Tesco grocery dataset recently published by Aiello et al. <https://figshare.com/collections/Tesco_Grocery_1_0/4769354/2> - the data contain detailed information derived from some 420m food items purchased from 411 Tesco stores **in** London in 2015. Information includes average nutrient profiles from baskets of goods purchased as well as categories of goods purchased.
   * The project may also wish to incorporate other demographic data from Census or other sources as well as information on fast food outlets from public health England or open street map.
   * Interesting validation work could be carried out by comparing the results of the classification with instances of diet or nutrition-related diseases such as obesity and type 2 diabetes.

****Topic 3 – Health Geographies

1. Social and Spatial determinates of sexual and Reproductive health service Usage
   * What are the underlying socio-economic and spatially varying contextual factors that drive demand and usage of Brook’s remote (video/phone/online) health services?

Objectives

* + 1. To compare the profiles of face-to-face versus remote (video/phone/ online) patient demand in Brook’s existing clinics, and how personal socio-economic and demographic characteristics interact with other contextual factors such as area-level deprivation, ethnic and family group membership in relation to specific areas of clinical need.
    2. To use this analysis to predict the latent face-to-face, and remote (video/phone/online) demand in other areas not currently served by Brook clinics, but with comparable geodemographic profiles to those that are.
  + Partner - Brook - <https://www.brook.org.uk/>

Background

* + Brook is a long-established and leading charity in the provision of confidential SRH advice and education as well as access to contraceptive and STI testing services to young people. Accessing SRH services may be the first-time young people use a health service independently. Brook supports young people to develop trust in services, empowers them to seek care and manage their health behaviours throughout their life.
  + During the COVID-19 enforced lockdown, the majority of SRH services closed/had limited access. The number of clients accessing Brook clinics per day dropped by 69%. Brook responded by rapidly adopting digital technology and transitioning to a mixed model of face-to-face and remote delivery. The COVID-19 restrictions on movement, work and social interaction lead to a substantial (3.6 times) increase in video and phone consultations accompanying a considerable reduction in face-to-face consultations.
  + Descriptive statistical analysis conducted by Brook suggests a shift in the ethnic and social composition as well as the geographic location of client/patients using their remote services compared to their client-base accessing face-to-face services in pre-COVID-lockdown times, suggesting a latent demand amongst some sub-groups of the population.
  + Brook maintain a large client/patient database containing data on client/patients and their consultations over the last 3-5 years. Brook will provide anonymised access to these data in order to facilitate this research. Brook is the data controller for the information contained within the electronic patient records (EPR) and the information is collected and held according to Brook’s Privacy and Confidentiality statement and processed according to Brook’s Information Governance Policies and Procedures. Brook’s Privacy statement informs clients that their non-identifiable data may be analysed for planning services to ensure the needs of local young people are met. A pseudonymised core dataset will be extracted from the EPR database for the purpose of the analysis which will take place within Brook’s secure IT infrastructure. No directly identifiable data will be included in the extracted dataset and no line level client data will leave Brook equipment.
  + Data – Brook will provide access to a time-series of data from their patient database

Classify food categories by nutrition, connect food consumption with geographic information in London (by a specific radius of area, or borough), get the diet or nutrition-related diseases data in the area/borough, linked the food consumption and nutrition-related diseases.

The research is for creating a new food and nutrition classification. What is more, it can be used to discover the relationship between the consumption and the diseases which are caused by it.

As the previous research result, there are 7 major parts in nutrient classifications which are Carbohydrates, Proteins, Fats, Vitamins, Minerals, Dietary fibre, and Water. And two well-known diet or nutrition-related diseases are obesity and diabetes.

Research also shows 50% fat intake with 50% Carbohydrates intake will more easily absorb by the body and turn to fat then causes obesity. The food which includes the such Carbohydrates-Fat ratio will be consider as the high risk food of causing obesity. What is the definition of obesity? Body mass indices (BMIs) are used to define it. The indices are calculated as the weight in kilograms divided by the square of height in meters. The BMIs of the subjects are compared with the federal standards for BMI categorization: less than 18.5 = underweight, 18.5 to 24.9 = normal, 25.0 to 29.9 = overweight, and 30.0 and greater = obese.

And Highly Processed Carbohydrates, Sugar-Sweetened Drinks, Saturated and Trans Fats, and Red and Processed Meats will greatly increase Diabetes Risk. Highly Processed Carbohydrates are also known as the refined carbs which have two main types: Sugars and Refined grains. Trans fats appear in packaged baked goods and fried foods, while saturated fats can be found in fatty meats, butters, and full-fat milk and cheese. The food contains such things above will be considered as high-risk food for diabetic.

By the previous result of the way that the body absorbs fat and carbohydrates, the food has a similar Carbohydrates-Fat ratio (which is close to 1) can be selected out as the main consumption that causes obesity. The range of ratios can be adjusted depends on the amount of selected food. After finding the classification, the next step is to find out the consumption of these kinds of food in different Tesco market locations. By assuming a part of the local people consumes their daily food in Tesco, comparing the nutrition-related diseases which are obesity in this case. Using linear regression on the data to find if the relationship is significant. The definition of high-risk food for a diabetic is more complicated, for example, a standard of Highly Processed Carbohydrates needs to be defined. By investigating the consumers' data, compare the consumption of Highly Processed Carbohydrates and the diabetic in the same area, and using the same method for the rest of the other 3 kinds of food. In the end, also by linear regression, the relationship can be found out between the consumption and the number of diabetics.

The extensive LSOA-level Tesco grocery dataset recently published by Aiello et al.

London data store: obesity and diabetes.

The data from customers has to be informed consent and voluntary participation. While there will be data of nutrition-related diseases patients who might not want personal information to be exposed. Therefore, confidentiality and anonymity is especially important to keep patients or consumers personal information. Based on this, the research and investigation is going to only assess relevant components.