City, University of London
MSc Data Science
2019 - 2020

Project Report

Case-study of data visualization for communication of complex research findings

An appraisal of data visualizations' effectiveness in conveying inequalities in diet, using data from the UK National Diet and Nutrition Survey

Veronica Tuffrey

Supervised by Professor Jo Wood

18th January 2021

Declaration

By submitting this work, I declare that this work is entirely my own except those parts duly identified and referenced in my submission. It complies with any specified word limits and the requirements and regulations detailed in the assessment instructions and any other relevant programme and module documentation.

In submitting this work I acknowledge that I have read and understood the regulations and code regarding academic misconduct, including that relating to plagiarism, as specified in the Programme Handbook. I also acknowledge that this work will be subject to a variety of checks for academic misconduct.

Veronica Tuffrey

Abstract

This study's first component explored inequality in the UK diet, using education, occupation and income as measures of socio-economic status (SES). Using data from 6802 adults collected in the National Diet and Nutrition Survey (2008-16), patterns of dietary inequality were explored across 12 regions and nine years.

Except for red meat, associations between dietary variables and SES for the most part followed expected trends, whereby dietary variables protective for health had positive associations with SES. Geographical variation in associations between diet and SES were not consistent within or between variables. Several dietary variables showed temporal changes towards a healthier diet, and educational inequality increased in fruit and vegetable intake.

The second study component explored whether interactivity in visualizations could increase the benefits public health practitioners gain from complex datasets with respect to communication of research findings. Using a within-subjects design, three visualizations (provided in static and interactive formats) from the study's first component were evaluated by thirteen public health nutrition practitioners, and their responses gathered via an online survey.

Adding interactive components moderately increased knowledge transfer and reduced cognitive load. By integrating interactivity in visualizations provided in supplementary materials, academic public health journals could better support non-academic users.

Keywords: Nutrition, socio-economic, inequality, visualization, interaction, evaluation

Wordcount: 29,987 (pages 1 – 80)

Acknowledgements

This work is dedicated to my husband Rowland, for his support and patience throughout the process.

Also my thanks are due to:

Professor Jo Wood, for his guidance and inspiration;

The survey participants, for their time and insights;

Professor Aiden Slingsby, for adeptly directing the MSc course through these challenging times;

The MSc data science students, for their good humour and generous sharing of expertise; and

My family, for their interest and encouragement.

CONTENTS

TABLES

Supplementary Tables in Appendix L

FIGURES

Supplementary Figures in Appendix M

Abbreviations and Acronyms

AfN	Association for Nutrition (professional association for nutritionists)
AOAC fibre	Non-digestible polysaccharides, measured with methods developed by the Association of Analytical Chemists
HRP	Household reference person
IND	Individual (name of a file in the NDNS dataset)
LRNI	Lower reference nutrient intake (see footnote 10 for definition)
NDNS	National Diet and Nutrition Survey
NMES	Non-Milk Extrinsic Sugars (see footnote 4 for definition)
NS-SEC	National Statistics Socio-Economic Classification
ONS	Office for National Statistics
PHE	Public Health England
PLD	Person level dietary data (name of a file in the NDNS dataset)
PSU	Primary sampling unit
Retequiv	Retinol equivalent units (a measure of Vitamin A intake)
RP	Rolling Programme (refers to the NDNS since 2008)
SES	Socio-economic status
SFA	Saturated fatty acids
SPSS	Statistical Package for the Social Sciences
VRN	Voluntary Register of Nutritionists

1 INTRODUCTION AND OBJECTIVES

1.1 Introduction

This report documents a study undertaken to meet the requirements of the researcher's MSc in Data Science at City, University of London. The study's aim is to inform evidence-based guidance on data visualization to communicate public health research findings from studies with large samples and complex study design. The research involved analysis of data from a large publicly funded survey to examine patterns of dietary inequality in the UK, and use of visualizations derived from the survey findings to explore whether interaction in visualizations enhances the effectiveness of communication.

1.2 Background

In the twenty-first century, the increased volumes of structured and unstructured data together with growth in computational power are being exploited by scientists to their advantage, while also posing challenges to scientific inquiry (Dhar, 2013). These challenges include integration, organisation, interpretation, and generation of knowledge from huge volumes of numerical data. Data visualization has long been recognised as a valuable tool in making sense of data, since harnessing the visual system helps to identify patterns, trends, associations and anomalies that may not easily be perceived by rapidly reading raw data or long texts (Ware, 2013). Given the context of what some call the "*data explosion*" (Mabry, 2011), the potential contribution of data visualization to science has never been greater that now.

One scientific discipline tapping an ever-expanding abundance of relevant data is epidemiology[1], which involves examining health outcomes with respect to varied factors including geographic distribution, demographic characteristics, clinical risk factors, social networks, and behaviours. The resultant datasets are complex, not only due to their size and to the diversity of factors, but often also due to the multi-level nature of data, collected at individual, household, and community levels. Visualizations play an increasingly important role supporting epidemiological analysis, communication, and ultimately public health decision-making, as is evident in the ongoing Covid19 pandemic.

[1] "*The study of how disease is distributed in populations and of the factors that influence or determine this distribution*" (Gordis, 1996, p3).

Disciplines in which epidemiology is systematically applied are public health,[2] and a specialism within this field, public health nutrition.[3] Together with other risk factors including physical inactivity and smoking, poor diet has long been recognised as contributing to disease outcomes including heart disease, stroke, and some cancers, with lower socioeconomic groups having greater incidence of these outcomes (James et al., 1997). Regrettably, recent evidence indicates that health inequalities are growing in parts of England, and life expectancy has fallen over the past decade after many years of increase (Marmot et al., 2020).

Thus, the circumstances in which this study was formulated are an exponential growth in data and expansion of computational power over the past two decades, an increase in health inequality in the UK over the last decade, and a pandemic that has newly brought public health data visualizations into many peoples' consciousness.

1.3 Choice of project

The researcher is a public health nutritionist with an academic background and overseas field experience. She has contended with the process of getting work published in academic journals, as well as experiencing this system from the viewpoint of a peer reviewer. Aware of the hurdles that authors experience, as well as the effort that editors and reviewers expend in the publication process, she is frustrated that to a large extent this effort does not benefit the practitioners working at the frontline of public health. Her impression is that much material of value to non-academic public health specialists reported in academic journals is often not accessible to those practitioners, since their access is constrained in terms of electronic or physical access, and/or the expertise assumed and language used. The researcher is also puzzled that authors of epidemiological papers use so few visualizations to help communicate their messages, especially since graphics are not included in article word-counts. These thoughts lead to speculation that her belief that visualizations are helpful for communication of research findings might be a consequence of her personal preferences and is not a viewpoint that is generally held.

The choice of this research project arises from the notions described above. It allowed the researcher to explore whether the increased computational power - by enabling interactivity in visualizations - could be harnessed to increase visualizations' effectiveness with respect to communicating public health nutrition research findings to practitioners. The study provided her with the opportunity to investigate other public health nutritionists' experiences of interacting with

[2] "*The process of promoting health, preventing disease, prolonging life and improving the quality of life through the organised efforts of society*" (Vetter and Matthews, 1999, p3)

[3] *"The promotion and maintenance of nutrition-related health and wellbeing of populations through the organised efforts and informed choices of society"* (Ridgway et al., 2019)

academic journals to gather evidence relevant to their professional lives. It also enabled her to follow up a long-standing interest in health and nutrition inequalities.

For a case-study for testing if interactivity could impact on visualizations' effectiveness, the researcher sought a data analysis project relating to dietary inequalities since, as mentioned above, this is a phenomenon causing rising concern. She identified the UK government-funded annual National Diet and Nutrition Survey (NDNS) (Bates et al., 2019) as a source of relevant data, and for which cleaned data from the past nine years are available in the public domain.

1.4 Research questions and objectives

In view of the context described above, the study has two **primary research questions** as follows (where interactivity is defined as enabling the user to tailor their view of the data according to their own objectives/preferences):

- **What are the past and present patterns of dietary inequalities in the UK?**
- **Does interactivity increase effectiveness of visualizations based on complex data?**

The tasks by which these questions will be addressed are expressed as these **research objectives** separated according to the two elements of the study as follows -

1) Secondary data analysis:

- **Ascertain the patterns of variation in the UK diet with socio-economic status (SES);**
- **Explore the extent to which the associations between diet and SES differ between regions;**
- **Examine time trends in dietary inequalities in the UK.**

2) Visualization design study:

- **Test whether integration of interactivity into data visualizations aids knowledge transfer;**
- **Test whether integration of interactivity into data visualizations reduces cognitive load.**

There were two **criteria used to assess whether the objectives had been achieved** as follows:

- **The work initially judged necessary to achieve the objective has been accomplished;**
- **Findings relating to the objective have been summarised in a short paragraph.**

The work needed to achieve each objective is outlined in Section 3.1.4.3 while the paragraphs summarising the findings are included at the end of each major section of Chapter 4.

1.5 Beneficiaries

1.5.1 Potential beneficiaries of the secondary data analysis

A review of the patterns of dietary inequality in the UK, comprising text, tables, maps, and graphs, will be of potential benefit to the following groups:

- **Academics:** They can use the results for teaching or to inform their own research;
- **Public health nutrition practitioners, especially those in the public sector:** Results highlight regions and population groups at risk, and so can inform policy prioritisation and design of health-related interventions. Given the low geographical granularity of the NDNS, they are probably of greater value at national and regional levels rather than local;
- **Civil Society:** Results highlighting nutrients and foods of concern are expected to be of value to community groups working predominantly with specific demographic target audiences.

1.5.2 Potential beneficiaries of the visualization design study

Results on the impact of including interactivity in data visualizations on users' understanding of research findings will be of potential benefit to **public health professionals who use data visualization in their work**. The results can inform their practice and enhance effectiveness of their visualizations.

1.5.3 Potential beneficiaries of the whole project

The application of results both about nutrition-related inequalities by the specified groups (1.5.1), and about how more benefit could be gained from health-related research findings by public health professionals, will contribute to improving the health of **the UK population**.

1.6 Outline of methods

Although the two study elements had separate objectives, the second element was dependent on the first, in that visualizations derived from the secondary analysis were used in the design study.

1.6.1 Secondary data analysis methods

The dataset analysed is the product of the National Diet and Nutrition Survey (NDNS) for which nine years' data from the rolling programme are available up to 2016/17. The complete data sets were downloaded from the UK Data Service website, the relevant files identified, and the necessary variables extracted and manipulated.

Basic descriptive statistics were visualised and tabulated for key sample characteristics and selected dietary independent variables. Regression modelling and calculation of an index of relative inequality was used to explore the association of the dietary variables with respect to the three socio-economic variables of occupation, income, and educational attainment. Finally, geographic and temporal variation in these associations was explored, again using regression modelling and the index of relative inequality.

Visual analytics were used to support the data analysis, using line plots, ranged dot plots, heatmaps, bar charts, choropleth maps and grid maps.

1.6.2 Visualization design study methods

The evaluation had a within-subjects design thus the respondents took part in each condition of the independent variable (here interactive or static versions of the same visualizations). The participants were non-academic public health nutritionists with an interest in diet and social inequality, recruited via a professional register.

Visualizations were produced using findings from the secondary data analysis that could be presented in two formats, static and interactive. Three different methods of presenting data visually were used – maps, lines, and bars. A questionnaire was produced, which assessed effectiveness of these visual outputs adopting the following definition: *"An effective visualization transfers knowledge and reduces the cognitive load for users over non-visual representations"*. Online survey software was used for drafting and applying the online survey. The quantitative data were analysed using basic descriptive statistics, while the main themes were extracted from the free text responses, and illustrative quotations extracted.

1.6.3 Major changes of goals or methods

Following advice from the research supervisor, the eventual goals were reduced compared to those outlined in the study proposal (Appendix A) while the methods were for the most part unchanged. The amendments to the goals affected just the second element of the study, the visualization design study, and reflect a shift in approach from quantity to quality, requiring selection of constituents on which to focus.

Initially the second research question specified an exploration of the association between interactivity and both engagement and effectiveness, and ultimately engagement was omitted. The objectives were also pared down due to being unrealistically ambitious for the time available. The researcher originally intended to assess the association of effectiveness (and engagement) with both complexity and interactivity, and ultimately complexity was omitted in favour of only considering interactivity. Finally, the objective of developing outline guidelines was omitted in favour of greater focus on the other objectives.

1.6.4 Work plan

Although the two study elements had separate objectives, the second element was dependent on the first, in that visualizations derived from the analysis were used in the design study. Thus, except for the omission of the fourth stage (development of guidelines) the final workplan was obliged to abide to a large extent to the sequencing of steps outlined in the Gantt chart in the study proposal (Appendix A).

However, the overlap and length of stages did differ from that originally planned due to delays in completion of some stages. Some delays were outside of the researcher's control – those of receipt of feedback on the project proposal, receipt of informal feedback from the pilot study, and receipt of ethical approval for the survey. Other delays were due to the researcher's misestimation of timings, in that the secondary data analysis and visual analytics took longer than foreseen. As a consequence, sharing with the supervisor of findings from the visualization evaluation and discussion of the first sections of the draft dissertation, and the supervisions following these milestones, did not take place.

Table 1 is a Gantt chart, simplified from that in the project proposal, showing the approximate actual timings for the activities.

		July'20		Aug		Sep		Oct		Nov		Dec		Jan '21	
1	Preparation														
2	Download, extract, manipulate data														
	Descriptive data analysis and basic visual analytics														
	Regression modelling and complex visual analytics														
	Develop visualizations for surveys														
3	Develop and pilot-test questionnaires														
	Identify respondents														
	Distribute surveys, follow-up respondents														
	Analyse survey responses														
4	Write up dissertation														
	Submit report and additional files														

Table 1: Simplified final workplan

1.7 Structure of the report

The remainder of this report is structured as follows: Chapter 2 explains the current state of knowledge in the topics pertinent to the study; Chapter 3 describes the methods used to obtain the results described in Chapter 4; and Chapter 5 discusses the findings in the context of the existing knowledge. Finally, Chapter 6 evaluates the project, includes the researcher's reflections on the research process, identifies areas for further work, and provides general conclusions. Additional documentation relating to details of methods or results are included in Appendices.

2 CONTEXT

This chapter summarises the contextual background and current knowledge on the topics most relevant to the study. It seeks to identify the gaps in knowledge and consequently provides the rationale for the current project.

2.1 Inequality and diet in the UK

There is a considerable volume of literature about diet and inequality in UK, which is considered below. But first we examine the concept of health inequality.

2.1.1 Inequality in health and nutrition and how to measure it

Definition of health and nutrition inequality

Health and nutrition inequality is defined in this study as *"..differences in the prevalence or incidence of health and nutrition problems between individual people of higher and lower socio-economic status*" (derived from Mackenbach and Kunst, 1997). Thus, although the definition refers to individuals, the key attribute is membership of a group. It also is not intended to convey any meaning about the fairness of any differences, which would be implied by the term "inequities". In this study we define health inequalities solely with respect to socio-economic status (SES), but other categorisations are used in other contexts, such as those related to socio-demography (age, area of residence, sex, disability and ethnicity/race), social environment (housing conditions, rural versus urban), and social capital (social networks and social support) (Roberts et al., 2013).

Associations between health, nutrition and socio-economic status

The simplest model for the links between SES, diet and health is a direct causal relationship, for example, one in which low SES is positively associated with a lower consumption of both fruit and vegetables thus increasing the risk of chronic disease. However, the true associations between diet, SES and health are complex, as shown in Figure 1 (Vlismas et al., 2009), not least because the links are bidirectional, for example, poor health may cause loss of income or employment.

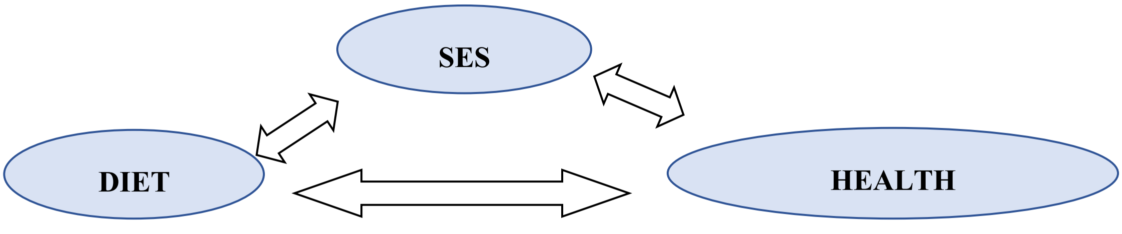


Figure 1: Model of relationships between diet, health and SES

There are several ways of measuring SES, and the associations of these varied measures with health and diet would be expected to differ. Education is linked to health and diet partly through knowledge and attitudes, while income reflects financial means, and occupation can represent one's social network and certain privileges related to health outcomes like easier access to health care and education, and better housing (Galobardes et al., 2007) .

Measurement of health and nutrition inequality

Assessment of health and nutritional inequality involves indicators, such as inequality indices (formulae that combine different characteristics of a phenomenon to provide an overall score or ranking), and decisions on what groups or areas to compare and the most appropriate form of analysis for the question being investigated. Table 2 summarises the necessary decisions, derived from (Mackenbach and Kunst, 1997, Carr-Hill et al., 2005, and Galobardes et al., 2007).

	Necessary choices	Options	This study
1	Groups between which inequalities are to be measured	1. Groups or populations of small areas 2. Countries or populations to which different socio-economic classifications have been applied 3. Similar groups or populations over time	2 and 3
2	Type of inequality of interest – Absolute v relative	1. Absolute inequality e.g. differences in mortality rate of highest and lowest SES groups 2. Relative inequality e.g. mortality rates in lowest SES group expressed as % of mortality rate in highest SES group	2
2	Type of inequality of interest – Individual risk or population outcome	1. Risk-based measures e.g. odds ratio of disease in lowest v highest SES group 2. Outcome-based measures (measure of total impact on health of whole population) e.g. population attributable risk; slope index of inequality	1
2	Type of inequality of interest – Simple v. sophisticated measurement	1. Simple measures e.g. rate differences or rate ratios for lowest v highest SES group 2. Sophisticated measures e.g. regression-based index of absolute effect; regression-based population attributable risk	1
3	Intended application for inequality analysis	Describe and monitor to: 1. Inform health policy 2. Monitor changes over time 3. Monitor differences between regions or social groups 4. Evaluate if policy targets have been reached 5. Explain causal mechanisms through which inequalities are generated 6. Statistically adjust for socio-economic circumstances when another risk factor is the main focus of interest	1
4	Indicator/marker of inequality	1. Socio-economic categories 2. A single variable or index 3. A deprivation index such as the Index of Multiple Deprivation	1 and 2
5	Source of health indicators	1. Routine health data 2. Data from existing surveys 3. Generation of new survey data	2

Table 2: Necessary choices when measuring health and nutrition inequality

Monitoring time trends in inequalities poses challenges, since the chosen SES indicator / index should measure the same construct and have the same meaning across time. It may also prove valuable to include more than one indicator. For example in a study of adults in Switzerland, using regression modelling, lower education and lower occupation were found to independently contribute to differences in diet, and that for some nutrients the two indicators had cumulative effects, suggesting both indicators should be assessed to fully describe social inequalities in dietary habits (Galobardes et al., 2001). Further evidence of the importance of assessing more than one SES indicator comes from UK obesity data, where patterns of inequality in adult obesity prevalence using education as a measure differ from patterns using income (PHE, 2012).

Measurement of socio-economic status

Indicators of individuals' SES measure some kind of individual resource or asset. Economic measures include employment, car ownership, and income, while those of social position include employment, education, occupational social class, housing and tenure status (Roberts et al., 2013). The most traditional SES indicators are occupation, education and income and these have proved very useful in describing and evaluating health inequalities (Galobardes et al., 2007).

There are several classification schemes based on occupation, and that used in UK national statistics *"The National Statistics Socio-Economic Classification"* (NS-SEC) has been in use in since 2001 (ONS, 2020). The conceptual basis for NS-SEC is in employment relations and conditions of occupations, rather than distinguishing between levels of skill in employment, and is intended to represent a social class structure within which certain classes are advantaged compared with others, and behaviours and outcomes are expected to vary by class (Rose et al., 2005).

The indicator of education is thought to capture the knowledge-related assets of an individual. It has the advantages of being relatively easy to measure in self-administered questionnaires and that response rates to educational questions tend to be high. Of the three common SES indicators, income most directly measures material circumstances, and can change most on a short-term basis. The indicator has the disadvantage that participants may be less willing to disclose income information accurately (Galobardes et al., 2007).

For adults living together in households, it has been long debated whether household SES is more relevant than individual measures (Galobardes et al., 2001). Evidence exists that household SES can be a better predictor of health than individuals' SES (Krieger et al., 1999), but the research took place more than twenty years ago, and those assigned a household classification different from their individual classification were all female. It is not clear if this finding applies independently of gender.

Measurement of dietary patterns

Individuals do not eat isolated nutrients, instead they consume meals and snacks consisting of a variety of foods, with complex combinations of micro- and macronutrients. Studies can assess the components separately or in combination, or use scores or indexes to measure the degree in which an individual's diet conforms to specific dietary recommendations, for example related to the Mediterranean Diet (Panagiotakos, 2008) or to healthy eating guidance (Patel et al., 2020).

Potential data sources include routinely collected food purchase data (though these are not generally available in the public domain), and surveys of expenditure and consumption, of which there are many in the UK, summarised by Foster and Lunn (2007). It is noteworthy that the UK has the longest-running continuous survey of household food consumption and expenditure in the world. The National Food Survey (NFS) was set up in 1940 to monitor the adequacy of urban 'working class' households' diets in wartime, and extended in 1950 to become representative of households throughout Great Britain (Foster and Lunn, 2007). The NFS was replaced by the NDNS in 1992, and the NDNS 'rolling programme' that has existed since 2008 is the source of nutrition data for the current study.

2.1.2 Association of diet and socio-economic status in the UK

Social-class gradients in nutrient intake in the UK have repeatedly been noted in the literature, starting with a desk study in the 1930s which classified the UK population into six groups according to income and then estimated the adequacy of the diet consumed in each of the groups. It was found that more than one third of the population were too poor to purchase an adequate selection of foods to maintain health (Boyd Orr, 1936). The direction of the social-class gradient is such that poor diet is consistently identified as a contributory factor to the higher rates of undesirable health outcomes in lower SES groups (e.g. James et al., 1997, Allen et al., 2017).

In 2007 there was a government-funded survey to explore the dietary habits and nutritional status of individuals living in materially deprived households. Compared to the general population average, "*low-income*" individuals were found less likely to consume vegetables and wholemeal bread, more likely to consume fat spreads and oils, non-diet soft drinks, pizza, processed meats and table sugar, and obtained relatively more energy from free sugars (NMES[4]) (Nelson et al., 2007).

An analysis of NDNS data from 2008 – 11 found that compared to the lowest SES groups, the highest SES groups consumed more fruit and vegetables, less red and processed meat, free

[4] "Free sugars" or Non-Milk Extrinsic Sugars (referred to in this report as NMES) are sugars that are in a free or readily absorbable state such as added sugars, or those released from disrupted cells, such as those from fruit juices or purees (Webster-Gandy et al., 2011).

sugars as a proportion of food energy were lower, and they were more likely to eat oily fish (Maguire and Monsivais, 2015). More recent analysis that integrated more years of NDNS data similarly showed trends for intakes of fruit, vegetables, fruit juice and oily fish to be higher, and for sugar-sweetened soft drinks to be lower, with increasing income (other SES indicators were not considered) (Bates et al., 2019).

2.1.3 Geographical variation in dietary inequality in the UK

There is a well-documented geographical variation in dietary quality in the UK, which contributes to variation in health outcomes. For example, Scarborough and colleagues (2011) noted that populations of Scotland and Northern Ireland consume a poorer diet than that of England, with respect to more saturated fat and salt, and lower consumption of fruit and vegetables. The average Welsh diet also had more saturated fat and salt, but more vegetables than the average English diet. Modelling suggested that if Wales and Northern Ireland achieved an average diet equivalent in nutritional quality to that of England, 81% of their excess cardiovascular and cancer mortality would be removed, while for Scotland the figure was 40% (Scarborough et al., 2011).

A few authors have considered geographical disparities with socio-economic variation:

- There were clear regional disparities in children’s nutrient intakes in 1946, with children in North England and Scotland consuming the worst diets especially with respect to low intakes of Vitamin C and carotene. The regional disparities were additional to socio-economic variation, and inferred as being influenced by local cultural norms (Prynne et al., 2002).
- Schofield and colleagues compared the diets of pregnant women in Edinburgh and London. They were surprised to find that regional differences in energy, protein, fat and fibre intake (intakes were consistently higher in London) outweighed social-class differences (Schofield et al., 1987). Intakes of the micronutrients iron, retinol, calcium and Vitamin C also favoured women in “non-manual” social groups in England (Schofield et al., 1989).
- Sodium intake in Great Britain was found to vary significantly across socioeconomic groups, even when adjusting for geographical variations. Dietary sodium intake was highest in people with the lowest educational attainment and in low levels of occupation, and those living in Scotland and the West Midlands (Ji and Cappuccio, 2014).
- When consumption data from Scotland and England were compared, regional dietary inequalities were more pronounced in the lower-income groups (e.g. red and processed meat consumption in the lowest-income quintile was 65 g/d in Scotland v. 58g day in England), but similar in the highest-income quintile (Barton et al., 2018).

This review indicates a gap in the literature for a study of geographical variation in inequalities that encompasses a range of foods and nutrients and includes all four UK countries.

2.1.4 Time trends in dietary inequalities in the UK

The British diet has been remarkably stable over the past 70 years. A noteworthy trend is the fall in total energy intake since the 1970s related to a shift towards a lower fat diet, with lower fat meats, such as poultry overtaking beef, pork and lamb as the most popular meats, and semi-skimmed milk dominating the milk category since its introduction in the 1980s (Foster and Lunn, 2007). An analysis of food purchase data between 2012-17 showed the trend in decreased energy intake was continuing. It indicated small declines in purchase of less healthy food products, translating to a small reduction of total energy and sugar purchased (Berger et al., 2019). Analysis reported of the dataset used in the current study showed that fruit and vegetable intake and proportion consuming oily fish hardly changed, while intakes of sodium, red meat, NMES, folate and Vitamin A decreased over the nine years (Bates et al., 2019).

Several authors have considered time trends in diet together with socio-economic variation:

- Sodium consumption in Great Britain dropped between 2001 – 2011 following the national salt reduction programme, but social inequalities in salt intake had not fallen (Ji and Cappuccio, 2014).
- Using slope and relative indices of inequality to assess trends in inequalities in household consumption in Scotland and England between 2001-09, no trends for inequalities to reduce were found (Barton et al., 2015) .
- Using the area-based Scottish Index of Multiple Deprivation to assess inequalities in children's diet between 2006 and 2010 in Scotland, again no changes were found (McNeill et al., 2017).
- When trends in adherence to UK dietary recommendations among sociodemographic subgroups from 1986 to 2012 were examined, inequalities relating to meeting the recommendations for sodium, and fruit and vegetable intake declined, while those inequalities relating to oily fish, and red and processed meat persisted across the three surveys (Yau et al., 2019).
- Time trends in a score derived to assess conformity with a diet used to prevent and treat heart disease (DASH[5]) were compared between socio-economic groups using 2008-2016 NDNS

[5] The Dietary Approaches to Stop Hypertension (DASH) diet is a set of recommendations including reduced consumption of sweets, sodium, and red and processed meats, and increased consumption of whole grains, fruits and vegetables, low-fat dairy products and nuts.

data[6]. Although the DASH score increased overall, no significant differences in the trend were observed across socio-economic position (Patel et al., 2020).

Thus except for two of the four variables in the study of Yau and colleagues, there is minimal evidence that dietary inequalities have changed with time over the last two decades. One would have expected that dietary inequalities had widened given the evidence that health inequalities are growing in parts of England (Marmot et al., 2020). One might surmise that there have indeed been increases but only in certain groups of the population, and that the dietary analysis published to date has not been sufficiently detailed to detect these changes. The researcher's own analysis of socio-economic inequalities in respiratory disease mortality found they occurred in specific demographic and geographic groups (Tuffrey, 2018).

[6] The same dataset as used for the current study, but with eight years' data rather than nine.

2.2 Effectiveness and interactivity in visualization for communication

Visualization can be defined as "*..the visual representation and presentation of data to facilitate understanding*" (Kirk, 2019, p15). This project's second component uses visualizations derived from analysis of dietary inequality data and evaluates whether integrating interactivity in visualizations can further facilitate understanding in the context of communicating research findings.

This section provides the context for the evaluation with respect first to the history of using visualizations in epidemiology, and second to the use of visualization to support communication of research findings. This moves to the questions of how visualizations' effectiveness can be assessed, and what effectiveness means in this context. Finally, we consider if there is evidence that integrating interaction in data visualizations can enhance their effectiveness.

2.2.1 Visualization in epidemiology and public health

With respect to the study's antecedents, three of the first documented uses of visualizations for analysis and communication of findings about British populations relate to health and wellbeing as follows:

- by William Playfair in the early nineteenth century who used economic data to graphically demonstrate how British workers' welfare had improved over time (Friendly, 2008);

- by John Snow who in 1854 plotted the location of water pumps and deaths from cholera in central London, and demonstrated that cases occurred close to one particular pump (Tufte, 2001); and

- by Florence Nightingale who used graphics showing deaths of British soldiers from different causes during the Crimean war, and demonstrated the importance of infectious disease as opposed to deaths at the hand of the enemy (Nightingale, 1858).

During the 19th century there was a rapid increase in the use of statistical graphics and thematic mapping, followed by the "*modern dark ages*" between 1900-1950 before the statistician John Tukey, the cartographer Jacques Bertin and others triggered data visualization to start "*..to rise from dormancy..*" in the mid-1960s (Friendly, 2008). Another surge in the popularity of data visualization occurred in the last three decades or so, partly linked to increased availability of new technologies and software products which facilitate the creation of visualizations (Strecker and Cox, 2012). Some suggest the development of dynamic graphic methods (Section 2.2.4) was key to enabling the recent growth in data visualization (Friendly, 2008). A hugely beneficial development since the 1980s to the fields of epidemiology and public health was in geographic information systems (GIS) as it redefined the capacity to work with spatial data (Moore and Carpenter, 1999).

This year's COVID-19 pandemic has emphasised the value of integrating visualizations with statistical analysis of epidemiological data (for example Dey et al., 2020).

Public health and epidemiological datasets often have a complex nature – typically several features lead to the following challenges for visual and statistical analysis (based on Preim and Lawonn, 2020):

- Datasets include a high number of dimensions;
- Datasets often have temporal and spatial dimensions;
- Often variables are measured at several levels such as person, household, school, community;
- Datasets are heterogeneous, including binary, nominal, ordinal and scalar variables;
- Data quality is often far from perfect, for example there is often a high proportion of missing data due to loss to follow-up in cohort studies;
- Data, especially when self-reported, are not completely reliable, due to "social desirability bias" for example, data relating to nutrition, alcohol and tobacco consumption is often biased towards social expectations, when people report following a healthier lifestyle than they truly do;
- Surveys often have complex sampling strategies.

2.2.2 Visualization for communication

Classifying visualizations

For this study it is not essential to distinguish between data visualization and information visualization, we will merely note that some authors do so (Kim et al., 2016). A categorisation of more practical value is exploratory v. explanatory. Explanatory visualizations are those used to transmit information, a point of view or a "story" from the designer to the audience, while exploratory visualizations are used by the designer for self-informative purposes to discover patterns, trends, or sub-problems in a dataset (Steele and Iliinsky, 2011). There is also a hybrid category, which involves a curated dataset presented to enable the reader to explore it using a graphical interface. This allows the audience to choose and constrain parameters, and discover insights for themselves (Steele and Iliinsky, 2011).

The role of visual perception in communicating information

Exponents of data visualization draw on the science of visual perception and contend that visualizing data is typically more effective for communicating information than text-based versions (Lindquist, 2011). Evidence suggests that if visualizations are designed correctly to draw on the brain's ability to detect certain properties, they can increase the speed both at which data is understood and the retention of information (Strecker and Cox, 2012).

Data visualization works by shifting the balance from thinking towards seeing – it translates the abstract into visual representations that can be meaningfully decoded, using the physical attributes of vision (length, size, shape, colour and position) (Few, 2013). Seeing, or visual perception, is handled by the visual cortex at the back of the brain and is very fast and efficient, while thinking, or cognition, is mainly handled by the cerebral cortex at the front of the brain and is much slower and less efficient (Few, 2013).

Ware distinguishes two kinds of visual perception processes: first "bottom-up", which are driven by the visual information in the pattern of light falling on the retina when "*massively parallel processing*" of the entire image occurs, and second "top-down", which are driven by active thinking, determined by the needs of the tasks. He places pattern perception in the *"flexible middle ground"* where objects are extracted from patterns of features, and translated the "Gestalt Laws", including those of proximity, similarity, connectedness, continuity, symmetry, closure, relative size, and "common fate", into a set of design principles for information displays (Ware, 2013). These laws are used below to justify design choices used in this study.

Knowledge creation in public health

Before considering the role of visualization in communication of research findings, it is helpful to break down the research process and note what is created at each stage. Here data are considered as simple measures of characteristics of people and things, with little innate value or meaning, while data analysis enables pattern identification, thereby creating information, and finally the use of information to generate recommendations, rules for action, and behaviour change signifies the creation of knowledge which can be used change behaviour and make decisions (Stansfield et al., 2006). Figure 2 shows the stages of the research process, and how knowledge derived from the final stage ideally feeds into decision-making to improve health (Tuffrey and Hall, 2016).

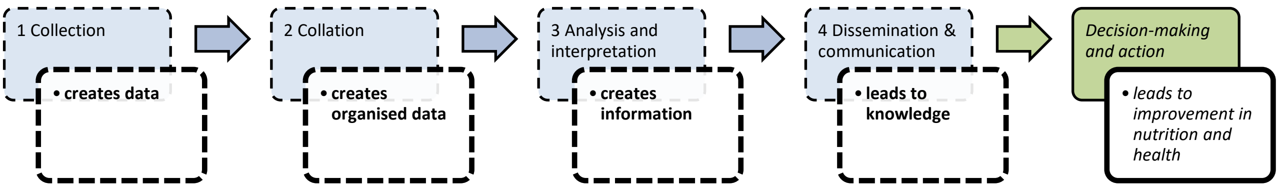


Figure 2: The research process and links with practice in public health

The role of explanatory visualizations in communication of research findings

Visualisations have an increasingly important role in exploring data and information and in generating knowledge in the context of the data "explosion" (Mabry, 2011) of the twenty-first century. Some suggest the data explosion has led to an "*information gap*" between non-scientists trying to understand very large amounts of data, and scientists and organisations who are exploring and trying to communicate the information being produced (McInerny et al., 2014). Traditionally, the scientific community has used explanatory visualizations to support scientific communication such as publications or conference talks. These visualizations are typically designed for audiences that are, to some extent, familiar with the underlying data or graphical form (Grainger et al., 2016).

While some claim that data visualizations can transform complex findings in a way that communicates research to a broader audience, and thus facilitate dialogue between academics, policy-makers, and the wider public (Gatto, 2015), it is questionable whether scientists can turn vast amounts of often complex data into outputs that are valuable to other scientists and also engage other categories of stakeholders. A promising approach being used to attempt this is "*narrative visualization*" in which interactive data visualization techniques are incorporated into story-telling (for example, Dove and Jones, 2012, Segel and Heer, 2010). The practice is being adapted in environmental science to reduce the science-to-practice gap (Arevalo et al., 2020). Given that our unwillingness or inability to apply health-related knowledge results in continuing health deficits and inequalities (Brownson et al., 2018) there is clearly value in similarly bridging the gap between researchers in public health and epidemiology, and policy-makers and public health practitioners. The only attempts identified to harness data visualization to support this goal involved stand-alone tools (for example, Isokpehi et al., 2020, Monsivais et al., 2018, Zakkar and Sedig, 2017).

2.2.3 Measuring visualizations' effectiveness

Issues around measuring effectiveness relate to its definition and the wide range of approaches used. Also findings are sensitive to users' characteristics with respect to domain knowledge, experience with visualization, and visual-spatial capability (Zhu, 2007), and to contextual factors affecting engagement with the visualization including time and emotions (Kennedy et al., 2016).

Definitions of data visualization effectiveness

There is no single established definition of effectiveness in the context of visualization. There is consensus that it cannot be considered independently of tasks - displays that are effective for one task may be ineffective for another (Hegarty, 2011).

The two most common aspects of effectiveness used to assess users' performance at tasks are accuracy (for example, Cleveland and McGill, 1984) and response times, alternatively called "efficiency" (for example, Casner, 1991). There is a recognised trade-off between these two measures, since taking more time is likely to benefit accuracy (Anderson et al., 2011). Interestingly, objectively measured performance of accuracy and response times are not always consistent with users' or designers' intuitions relating to effectiveness, thus highlighting the importance of the objective measures (Hegarty, 2011).

Tufte indicated that effectiveness was associated with “data/ink ratio” but did not define effectiveness, and while his principle to *“..maximise the data-ink ratio, within reason”* (Tufte, 2001, p96) has influenced much visualization design, there is no empirical evidence linking this ratio with accuracy or efficiency (Zhu, 2007).

The main approaches used to evaluate effectiveness, user studies and heuristics, are as follows:

User studies for assessment of effectiveness

There are several types of user studies:

1) *Objective performance measures*: Empirical data are collected on speed and accuracy.

2) *Observers’ reactions to displays:* A “think aloud” protocol giving rise to spontaneous descriptions of a graphic can reveal what information is salient, how much of the displayed information is being encoded, and what types of information a graphic communicates (Hegarty, 2011) (for example, Shah and Carpenter, 1995). Questionnaires and interviews can also be used, and as well as narrative data, observational data on behaviour can also be valuable (Brown et al., 2020). This category includes “usability studies” (for example, Cinnamon et al., 2009).

3) *Eye tracking*: Eye fixations can be interpreted as a measure of visual attention (Hegarty, 2011). Eye fixation data can complement efficiency data from reaction times, as the former can provide diagnostic information, such as areas of a display attracting attention even though they are not task relevant (for example, Ratwani et al., 2008).

4) *Logs of users’ interactions*: As use of interactive displays increases, assessment via logging users’ interactions with such displays is increasingly common (for example, Robertson et al., 2008).

5) *Longitudinal and field studies:* Studies in real-life settings may use ethnographic methods, whereby observational data are explored using structured analytic frameworks (Brown et al., 2020).

6) *Measurements of brain activity:* Measurements of electroencephalography (EEG) have been used to assess the cognitive load associated with a task, with the assumption that the EEG reflects the performance of working memory (Anderson et al., 2011).

Heuristic evaluation for assessment of effectiveness

User studies can be time-consuming and expensive, especially the empirical methods which provide objective measures. The second approach to evaluation, heuristics, is centred on visualization specialists rather than users. Experts review a system and judge how well it meets the goals of predefined guidelines. Reviews of existing lists have highlighted the difficulties of deriving an acceptable minimum set (Zuk et al., 2006, Forsell and Johansson, 2010), leading one to question if this is a worthwhile goal. Development of task-centric heuristics seems more feasible and the products more useful.

Predicting effectiveness of visual displays using cognitive approaches

Another strategy is the use of models that predict effectiveness *a priori ,* and thereby inform the design of new displays (Hegarty, 2011). This approach is based on prescriptions about display design, derived from findings from task analyses and knowledge about perception and cognition, for example findings ranking tasks with respect to accuracy were used to predict effectiveness of different types of graphs (Cleveland and McGill, 1984). Also models of visual salience or visual clutter have been validated by empirical testing and are now used to guide display design. There are challenges scaling up this approach to designs for complex tasks (Hegarty, 2011).

Evaluating effectiveness for communication of research findings

This study corresponds to the "*Evaluating Communication Through visualization*" (CTV) scenario described by Lam and colleagues. They used a binary categorisation to separate evaluation scenarios into those about understanding data analysis processes (like this study), and those about assessing visualizations themselves. They suggested that for CTV, qualitative methods like observations and interviews were suitable, or controlled experiments, or a mix (Lam et al., 2011).

A review of the literature (admittedly not systematic) only revealed one example of public health CTV studies about information systems (Zakkar and Sedig, 2017) and none relating to communication of research findings. Thus, in this sense the study fills a research gap.

2.2.4 Interaction

As mentioned above, some contend the development of dynamic graphic methods "*..allowing instantaneous and direct manipulation of graphical objects and related statistical properties"* was central to growth in data visualization over the few decades (Friendly, 2008). Interactivity has been said to make visualizations more effective by increasing the flexibility and speed of use of visualizations to "*rates resonant with the pace of human thought*" (Heer and Shneiderman, 2012). This claim is examined below after defining and characterising interaction.

Definitions of interaction and types of interactivity

Ben Schneiderman first described interaction in the 1980s when he used the term "*direct manipulation*" to describe an interaction style in which users act on displayed objects using "*.. rapid, incremental, reversible operations whose impact on the object of interest is immediately visible*" (Shneiderman, 1983). Dimara and Perin noted a lack of consensual definition of interaction for data visualization, and proposed the following definition synthesized from the visualization and human-computer interaction literature: *"Interaction for visualization is the interplay between a* *person[7] and a data interface involving a data-related intent, at least one action from the person and an interface reaction that is perceived as such"* (Dimara and Perin, 2019).

Data & View Specification	**Visualize** data by choosing visual encodings **Filter** out data to focus on relevant items **Sort** items to expose patterns **Derive** values or models from source data
View Manipulation	**Select** items to highlight, filter, or manipulate them **Navigate** to examine high-level patterns and low-level detail **Coordinate** views for linked, multi-dimensional exploration **Organize** multiple windows and workspaces
Process & Provenance	**Record** analysis histories for revisitation, review and sharing **Annotate** patterns to document findings **Share** views and annotations to enable collaboration **Guide** users through analysis tasks or stories

Table 3: Taxonomy of task groups for interactive dynamics for visual analysis

Several taxonomies of interaction operations exist (for example Figueiras, 2015, Yi et al., 2007) in addition to that shown in Table 3 which groups tasks into three high-level categories (Heer and Shneiderman, 2012). The category "View manipulation" is most relevant to this study since it focuses on explanatory visualizations (Section 2.2.2) and users use interactivity to explore them.

Does interaction lead to more effective visualizations?

There has been controversy in relation to the benefits of interaction, and concrete evidence is lacking to support claims made that interactivity can augment the user's understanding of the data and overcome some limitations of representations (Figueiras, 2015). Factors contributing to this situation seem to be the range in nature of tasks encompassed by interactive techniques (see Table 3), and the complexity of measurement of effectiveness (Section 2.2.3).

Assertions made about the benefits of interactive tools over traditional scientific outputs which show "static" visualizations are based around the assumption that by allowing users to have control over and actively explore information, web-tools can promote understanding (McInerny et al., 2014). For example, users can change parameters and uncover the data at a custom pace by clicking on interactive elements like buttons and sliders. This is said to be especially beneficial for explaining multidimensional problems, with potential to support users to develop *"personalised mental models*" of what were originally highly complex datasets (Grainger et al., 2016). Table 4 examines if research findings support these claims. It summarises the few studies identified by the researcher that compare interactive and static tools, with a symbol to indicate if interactive tools were more effective (+), less (-) or there was no / variable difference (0) compared to static.

[7] The authors replaced the usual term "user" by the term "person" that is more likely to be perceived as gender-neutral. The "data interface" is a system containing a set of controls that are perceivable by the person (e.g. a computer) linked to a data entity which is intangible.

Reference	Topic, Sample size, design and methods	Findings	
(Cinnamon et al., 2009)	Usability study of static, animated and interactive maps of injury rates and socio-demographic determinants of injury. N=8; within-subjects design, using "thinking aloud", questionnaires and interviews.	No one map type was found to be better than the others – their usefulness depended on the different purposes and for satisfying the varying skill level of the individual user.	0
(Ancker et al., 2011)	Study of perception of medical risks. Compared illustrations of health risks using static or interactive game-like graphics. N=165; between-subjects design using questionnaires.	No difference in understanding were found between the two graphics. Differences between participants w.r.t. numeracy were very important.	0
(Zikmund-Fisher et al., 2011)	Study of risk perception of risk of treatments' side- effects. Compared task performance using interactive and static "pictographs". N=3371; between-subjects design; internet survey.	The interactive graphic condition had lower completion rate, and participants were less likely to make the correct decision.	–
(Arciniegas et al., 2013)	Study comparing a set of collaborative spatial decision support tools developed to support land-use allocation. N=30, within-subjects design, users had to do group and individual tasks. Used questionnaires, video observations and task performance results.	The simplest tool was found the most useful, clearest and had highest impact on decisions. Performance with the interactive tool was the worst. The cognitive effort related to volume and format of information was key.	–
(Blasch et al., 2017)	Study comparing impact of decision-support tools related to adoption of energy-efficient appliances. Two samples of N=916 and 5015; between-subjects design using online tasks and questionnaires.	The interactive web-tool was more effective than educational slides with respect to identifying the appliance with the lowest lifetime cost.	+
(Xexakis and Trutnevyte, 2019)	Study comparing two modes of communicating information about the environmental, health, and economic impacts of scenarios relating to the Swiss electricity supply. N=313; between-subjects design; used internet survey.	Participants using the interactive web-tool to explore the database of scenarios had lower understanding compared to those using the website presenting four scenarios with storylines.	–

Table 4: Summary of studies comparing static and dynamic tools

Table 4 only shows one study of the six reporting a positive association between interactivity and effectiveness. While bearing in mind that the review to identify papers was not systematic (it is largely based on the review in the introduction to Xexakis and Trutnevyte's paper (2019)), and the papers were not critiqued for methodological robustness, the table does to a limited extent undermine claims that interactivity improves effectiveness.

There are many reasons why interactivity may not show benefits, one of which is that just providing people with an interactive visual display does not ensure that they will use it effectively, as demonstrated by a study using interaction logs (Keehner et al., 2008). With respect to the use of interaction in public health and epidemiological data visualizations, a review found that static graphics are still the main method for representing data in infectious disease visualization tools for public health professionals (Carroll et al., 2014).

3 METHODS

The first sections of this chapter describe the methods used for each of the two study components, and the third section briefly comments on aspects related to research ethics.

3.1 Data analysis

This section describes the dataset and methods for the secondary analysis, to achieve the first three research objectives. All details about the National Diet and Nutrition Survey (NDNS) come from the most recent user guide (NatCen Social Research, 2019c) and Appendix B to the report for Years 1-9 (Bates et al., 2019).

3.1.1 Dataset

The dataset used in this study is the product of the UK NDNS, a survey designed to assess the dietary habits and nutritional status of the general adult and child population in the UK. NDNS data are used by government to develop policy and assess progress against nutritional targets and exposure to chemicals in food. The programme began in 1992, and since 2008 the survey has been conducted annually as part of a rolling programme (RP) funded by Public Health England (PHE) and the Food Standards Agency. The secondary analysis reported here is based on the RP data and spans nine years' data collected between 2008 – 2017.

3.1.2 Sample

The NDNS sample is intended to represent the UK general population aged ≥1.5 years living in private households. Around 500 adults (≥ 19 years) and 500 children take part each year. Recruitment in Wales and Northern Ireland is boosted to 200 participants annually to enable representative country-specific data.

The sample is drawn from the "small users" sub-file of the Postcode Address File, a computer list prepared by the Post Office, of all UK addresses which receive fewer than 25 articles of mail daily. For efficiency, two-stage sampling is used:

- Primary Sampling Units (PSU) are created from the addresses clustered into geographical areas based on postcode sectors, and the PSUs are randomly sampled (158 PSUs in 2016)
- A random sample of addresses is drawn from each of the selected PSUs (28 addresses per PSU in 2016, so the total sample was 4,424 addresses).

If more than one household lives at a particular address visited, then one is randomly selected for inclusion, and within participating households, up to one adult and one child are randomly selected to participate. Table 5 shows the response rates and resultant total sample sizes for individual fieldwork years, grouped by "waves" (Section 3.1.4). The final column shows the number of adults, since for this study only data from adults aged ≥19 years were used.

Wave	*Response rate (proportion of individuals selected who completed ≥3 days' food diary*	*Resultant total sample size*	*Sub-sample of adults*
Years 1-4	56%, 57%, 53% and 55%	6,828	3,450
Years 5&6	54% and 53%	2,546	1,288
Years 7&8	55% and 49%	2,723	1,417
Year 9	50%	1,253	647
TOTAL		13,350	6,802

Table 5: Response rates and sample sizes for NDNS surveys, 2008-16

Notes: Sources: (Bates et al., 2014, Bates et al., 2019)

3.1.3 Primary data collection and analysis

3.1.3.1 Fieldwork

Details of the NDNS fieldwork are included in Appendix B. The analysis reported here only uses data collected using the 4-day diet diary, and face-to-face Computer Assisted Personal Interviews, collected over 3 main visits to households who had agreed to participate.

3.1.3.2 Original data analysis

Appendix B includes details of the analysis of the raw dietary data by PHE.

3.1.4 Secondary data analysis

The software used for analysis was SPSS Statistics for Windows (V25.0, Armonk, NY: IBM Corp).

3.1.4.1 Data download and extraction

For the NDNS RP, nine years' data are available, and can be downloaded from the UK data service website (https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=6533) together with supporting documentation relating to the survey protocols, and editing, manipulation and coding of the primary and derived data. The data are provided in four time-based sets, called "waves", corresponding to Years 1-4; Years 5&6, Years 7&8, and Year 9.

For each wave, the researcher selected those files that included the variables needed to address the study objectives, and downloaded the two files in SPSS format as follows:

- the "Individual" (IND) file, containing contextual data for all fully productive individuals (defined as those that completed at least three of the four food diary days)
- the "PersonLevelDietaryData" (PLD) file containing mean intakes of nutrients and foods.

For each of the four waves, the IND and PLD files were merged using individuals' serial number as the key variable. The four resulting datasets were combined and then the desired variables extracted. These variables fall into five categories:

1. Dietary dependent variables: The researcher chose a small number of dietary variables for which there is strong evidence of associations with health-related outcomes, either as protective or risk factors. These outcomes include chronic disease (such as cardiovascular disease and some cancers), immunity and cognition (Webster-Gandy et al., 2011). Study variables related to three tyeps of dietary elements:

    - Foods: Fruit and vegetables[8], red and processed meat, oily fish, alcohol and soft drinks;
    - Macronutrients: Total dietary energy, % food energy from fat, % food energy from saturated fats; % food energy from non-milk extrinsic sugars (NMES) and fibre;
    - Vitamins and minerals: Iron, folate, calcium, sodium, Vitamin A and Vitamin D[9].

2. Socio-demographic independent variables: Variables relating to age-group, sex and ethnicity were extracted because the association between socio-economic status and diet may confounded or modified by these factors.
3. Socio-economic independent variables: Data relating to the three socio-economic variables of income, occupational social class and educational attainment were extracted because there is evidence these three indicators contribute independently to dietary outcomes, and because the limitations of the measures differ (Section 2.1.1).
4. Geographical independent variables: Quality of diet is known to vary within the UK (Section 2.1.4) but it is not known if levels of dietary inequality vary between regions.
5. Temporal independent variable: Use of survey year as an independent variable enabled exploration of dietary trends over the decade, both in absolute values and inequalities.

[8] The variable for proportion achieving the "5-a-day" target was used as well as the variable for total fruit and vegetable intake, since one cannot assume their variation would follow the same patterns.

[9] One cannot assume patterns of variation in intake for the whole population will be reflected in patterns of variation in the lowest section of the intake distribution, so variables for proportion of population falling below a reference value (LRNI, see next footnote) were used, in addition to the continuous variables for intake of micronutrients (except for Vitamin D for which the binary variable was not available).

3.1.4.2 Data manipulation, coding and checking

Appendix C provides details about the original downloaded variables, and manipulation by the researcher for the secondary data analysis. Most variable names were identical in the four datasets. The location of the variable in the downloaded files is provided only for the most recent dataset (Year 9) in the second column of the table.

Binary dietary dependent variables (8)

Variables indicating dietary intake lower than the LRNI (Lower Reference Nutrient Intake[10]) only had values of 1 in the datafiles for those individuals meeting the criteria, while the remainder of the sample had "system missing" values. The latter were converted to values of 0 to enable modelling of the variables' variation using logistic regression. This manipulation was undertaken for the binary variables relating to Iron, Vitamin A, Calcium and Folate (*bloIronLRNI, bloVitAlrni, bloCalciumlrni and bloFolatelrni).*

Continuous dietary dependent variables (13)

The shape of continuous variables' distributions was checked, both visually using histograms and quantitatively using the SPSS **Frequencies** command. Three variables had a high proportion of zero values (those relating to alcohol consumption (*Alcoholg* , 46.9%), oily fish (*OilyFishg,* 71.6%) and sugar-sweetened soft drinks (*SoftDrinksNotLowCal*, 54.6%)), so instead of transforming them to correct for the high positive skew, they were converted to binary variables coded 1 if the individual had consumed the dietary element and zero otherwise.

Most other continuous variables also demonstrated positive skew. Thus new transformed variables were generated for use in the regression modelling, since an assumption of this technique is that the underlying data are normally distributed (Field, 2005). The square root transformation was found to be optimal for all dietary variables except Vitamin A consumption (*VitaminAretinolequivalents)* which was transformed using the logarithmic function, and total fat expressed as a percentage of food energy (*FatpcfoodE)* for which the distribution was near normal. Z scores of the transformed dietary variables were also generated to standardise the variables and thereby facilitate comparison of the output from regression modelling (Section 3.1.4.3).

Socio-demographic independent variables (3)

Age: Age was included in the models as a categorical variable to avoid making assumptions about the nature of its association with the dietary dependent variables. The variable *agegad2* provided had five categories (16-18y, 19–34y, 35-49y, 50-64y and 65+) so the upper four categories could be

[10] The Lower Reference Nutrient Intake is the amount of a nutrient that is enough for only the small number of people with low needs. Most people will need more than the LRNI if they are to eat enough. If individuals are habitually eating less than the LRNI they will almost certainly be deficient (Department of Health, 1991).

used unchanged for analysis.

Ethnicity: For ethnicity, again a derived variable was provided that was adopted for the analysis. For this variable the label "White" was coded as 1 and "Non-white" as 2, with the latter category having combined the following groups: "Mixed ethnic group", "Black or Black British", "Asian or Asian British", and "Any other group". For analysis, this variable was renamed "WHITE" and recoded as 1 & 0 to facilitate interpretation of the regression models.

Sex: The variable *Sex* provided was renamed "MALE" and recoded from 1 & 2 to 1 & 0, to facilitate interpretation of the regression models. When reporting findings this variable is referred to as "Gender" since it appears that the data were recorded from appearance or self-identification rather than a strict biological meaning.[11]

Socio-economic independent variables (3)

SES measures used in NDNS are "gender-neutral household" measures (Krieger et al., 1999) rather than individual, since they use the status of Household Reference Person (Section 3.1.3.1).

Occupation: The original variable was recorded in the eight NS-SEC categories shown in the leftmost column of Table 6. For the analysis, the categories were combined into four groups to obtain as equal a distribution as possible. Those who had never worked were excluded, both because the group was too small to form a category, and because if sick or disabled their dietary choices would be affected by their condition.

NS-SEC Category	*New derived category label*	*New code*
Higher managerial and professional occupations	Higher managerial and professional	4
Lower managerial and professional occupations	Lower managerial and professional	3
Intermediate occupations	Intermediate	2
Small employers and own account workers		
Lower supervisory and technical occupations	Routine and manual	1
Semi-routine occupations		
Routine occupations		
Never worked		-
Other		-

Table 6: Occupation-based social-class categories

Education: The original variable for highest attained educational qualification had eight categories as shown in the leftmost column of Table 4. For use in this study, the categories were combined into four groups to obtain as equal a distribution as possible. Those with foreign or other qualifications were excluded as the level was unknown, and full-time students were excluded because the level of their highest attained qualification was unknown.

[11] The interviewer schedule notes: "*INTERVIEWER: Ask or record sex of X - 1 Male; 2 Female.*"
.

Qualifications Category	*New derived category label*	*New code*
Degree or equivalent	Degree or higher	4
Higher education, below degree level	A level or equivalent	3
GCE, A level or equivalent		
GCSE grades A - C or equivalent	GCSE or equivalent	2
GCSE grades D-G/Commercial qualifications/apprenticeship		
Foreign or other qualifications	-	-
Still in FT education		
No qualifications	No qualifications	1

Table 7: Education-based social-class categories

Income: The original_variable provided, *eqvinc,* is "equivalised household income". Equivalisation is *"..a standard methodology that adjusts household income to account for different demands on resources, by considering the household size and composition"* (Bates et al., 2019*).* The sole description found of the process was in a footnote to a Scottish NDNS report (Bates et al., 2017) and this is included in Appendix E, since one can assume the process is applied to the income data collected across the UK (though the scores allocated to the age groups may differ).

The second original variable provided was terciles of equivalised household income, however this variable only existed in the datasets for years 5 – 9. There is no explanation for their omission in the dataset for the first wave, which is odd given that equivalised income quintiles were calculated for Scotland (Bates et al., 2017).

Given that the tercile data were incomplete in the combined dataset, the researcher derived a new variable with the same number of categories as for education and occupation. Thus, equivalised income quartiles were calculated from each year's income data, where quartile 1 is the group with lowest values and quartile 4 the highest.

3.1.4.3 Statistical analysis and visual analytics

Table 8 provides a summary of the statistical analysis and visual analytics undertaken to address each of the three first objectives. More details are provided below the table.

	Research objective	*Statistical analysis*	*Visual analytics*
1	*Ascertain the patterns of variation in the UK diet with SES*	Descriptive: Tabulation of selected dietary outcomes with respect to the three SES variables; Correlation analysis; Tabulation of relative inequality index;	Line plots of all dietary variables by SES categories, for whole sample and disaggregated by demographic groups; Ranged dotplots of relative inequality index;

		Analytic: For subset of variables, regression modelling (1) with SES alone; (2) with demographic variables as covariates; (3) with interaction of SES and demographic variables	Heatmap of correlation coefficients; Barcharts of regression coefficients
2	*Explore the extent to which the associations between diet and SES differ between regions*	Descriptive: Tabulation of relative inequality index by Region	Line plots of selected dietary variables by SES categories by Region; Ranged dotplots of Relative Inequality Index by Region; Grid maps
		Analytic: For subset of variables, regression modelling (1) with Region and SES alone; (2) with Region, SES and demographic variables as covariates; (3) with Region, SES and interaction of SES and Region	Barcharts of regression coefficients
3	*Examine time trends in dietary inequalities in the UK*	Descriptive: Tabulation of relative inequality index by Year.	Line plots of selected dietary variables by SES categories, by Year; Ranged dotplots of Relative Inequality Index by Year; Line plots of relative inequality index by Year;
		Analytic: (1) with Year and SES alone (2) with Year, SES and demographic variables as covariates; (3) with Year, SES and interaction of SES and Region	Barcharts of regression coefficients

Table 8: Summary of work necessary to address the research objectives

Weighting

Each wave's IND file includes a weighting variable to enable analysts to reduce bias resulting from differential non-response, and unequal selection probabilities due to the multi-stage design. The weights consider the age, sex and regional profiles of the participants. For example, the sample design entails over-sampling of addresses in Northern Ireland and Wales, and the weighting variable adjusts the 4 countries' data to their correct population proportions.

For this study it was necessary to combine the four datasets, and since the number of years included in each wave varied, the four sets of weights needed to be rescaled to ensure the datasets were in correct proportions (4:2:2:1). The basic code to create the new weighting variable for the combined dataset was obtained from the user guide (NatCen Social Research, 2019b) and Appendix D includes the derived code.

Adjustment for multistage sampling

Due to the multi-stage sampling design of NDNS, during data analysis statistical adjustment of estimates of the size of sampling errors is needed, which affect the sizes of confidence intervals (CI), and p values for hypothesis testing. If this adjustment is not applied, values for CI and p obtained assume simple random sampling, and so the width of confidence intervals will be under-estimated, and the statistical significance of hypothesis test will be over-estimated.

Complex sample analysis involves two steps in SPSS, first a complex sample "plan file" is created after computing a weight variable (Appendix D) and second analyses are run using the plan file through the **Complex samples** package (IBM, 2017). The Complex Samples Plan (CSPLAN) file contains information about sampling levels, weights, and strata, and Appendix F includes the script for the plan file used in this study. All descriptive and analytic statistics presented in this report were adjusted for multi-stage sampling.

Preliminary descriptive analysis

Before addressing the objectives, frequency tables were tabulated for the categorical dietary and independent variables, and summary descriptive statistics were tabulated for the continuous variables. Choropleth maps were plotted to examine geographical variation in the dietary variables, and line graphs plotted to examine temporal trends in the dietary variables, without disaggregating by SES category.

Assessment of inequality

As outlined in Table 2, assessment of inequality can be undertaken in several ways. In this study, variation in dietary variables was assessed by SES category for each of the three SES measures, and an index of relative inequality was calculated, amending the approach used by Si Hassen and colleagues (Si Hassen et al., 2016). Relative difference in dietary variables between individuals belonging to the highest SES category and those of the lowest SES category was computed as:

$$\text{Relative inequality (\%)} = \frac{(\text{mean value of highest SEP category - mean value of lowest SEP category}) * 100}{\text{average value across all four SES categories}}$$

A positive relative difference shows higher values in the highest SES category. The method used here differs from the original approach with respect to the denominator - the mean value of SES was used instead of the value for the highest SES category, since there was one dietary variable (Folate intake < LRNI) for which the value for the highest SES category was zero.

Regression modelling

A subset of six dietary variables was selected for regression modelling. Four of these corresponded to those examined in a recent analysis of dietary inequalities (Maguire and Monsivais, 2015), that is Fruit and Vegetables intake; Red and processed meat intake, NMES and Oily fish. The other variable considered in the previous study was SFA as %food energy. Instead of this variable, Alcohol and Folate<LRNI were selected for modelling here, since preliminary data analysis revealed these had much stronger associations with SES than SFA.

Multiple linear regression was used for the continuous dietary variables, and logistic regression for the binary dietary variables to examine their association with SES variables, and variation by Region and Survey Year. These were accessed respectively via the **General Linear Model** and **Logistic Regression** options within the **Complex Samples** module of SPSS. The demographic independent variables were included in the models as covariates. To compare the models outlined in Table 8 above, adjusted R squared values were used for multiple linear regression modelling, while for logistic regression Cox and Snell's Pseudo R Square was used, which is based on the log-likelihood (Field, 2005, p223). The Wald statistic was used to compare the contribution of predictors between models, this is the value of the regression coefficient divided by its associated standard error (Field, 2005, p224).

QQ and other diagnostic plots including histograms of residuals were used to evaluate if assumptions underlying the method were being met (Field, 2005, p96). Since such plots are not available in the **Complex Samples** module, these checks were undertaken using the **Linear** and **Binary Logistic** options within the **Regression** module of SPSS, before using the **Complex Samples** module to run the models for which the results are presented in Chapter 4. As described in Section 3.1.4.2, Z scores of the transformed dietary variables were used as the dependent variables, to help compare model outputs between variables measured in different units. For plotting values of regression coefficients in bar charts, reference categories were chosen as category 4 for SES (highest), male for gender, ≥65y for age-group, white for ethnicity and Northern Ireland for region.

Generally a p-value < 0.05 was deemed statistically significant. When interpreting output from multiple comparisons, such as when testing for interactions, the Bonferroni correction was used whereby each test conducted uses a significance criterion equal to the normal criterion divided by the number of tests conducted (Field, 2005, p725). Consistent with recent guidance, exact p values were provided in tables of results rather than inequalities, unless the p value was less than 0.001 (Greenland et al., 2016).

Visual analytics

In parallel with the descriptive analysis and modelling described above, visualizations were generated using data from the summaries and tests. Consideration of these visualizations aided comparison between groups and between variables, and helped build familiarity with the dataset and the relationships between dependent and independent variables. The software packages used to generate the visualizations were initially Tableau (www.tableau.com/) and then Litvis (www.gicentre.net/litvis). Litvis provides a "notebook" environment integrating live coding input, rendered output and textual narrative (Wood et al., 2018), and thus facilitated documentation of production and fine-tuning of charts for inclusion in the survey and this report.[12]

Data were prepared for import into these packages (as .csv files) using Microsoft Excel (https://office.microsoft.com/excel). The Color Brewer website was used to select colours (https://colorbrewer2.org/) when the default colours were not satisfactory.

1) Line graphs were produced in the early stages of analysis to examine trends in dietary variables over time, and differences between socio-economic groups, regions, age-groups, ethnic-groups and genders. They were also helpful in later stages to elucidate interactions between variables.

2) Maps were used to examine both cross-sectional and longitudinal geographical variation in dietary variables**.** To create maps, geographical boundary data were downloaded from the Open Geography Portal of the Office for National Statistics (https://geoportal.statistics.gov.uk/) and simplified using the free online map data editor **MapShaper** (https://mapshaper.org/). The resulting shapefiles were matched with the NDNS data and the visualizations generated.

Data that are both space- and time-referenced can be considered either as a "*spatial arrangement of local behaviours over time*", or as a "*sequence of momentary behaviours over the territory*" (Andrienko and Andrienko, 2006). Both approaches were helpful, so line-graphs showing time trends in dietary variables for each region were produced, as well as multiple series of choropleth facet maps. Also, grid maps were produced showing each region as a rectangle at its approximate geographic location, to aid comparison between regions of dietary variables' trends over time and SES categories. This type of map allocated space equally between the twelve regions (London was now the same size as Scotland) so more closely matched the population distribution within the UK.

3) Bar charts were used in early stages of analysis to help explore variation in dietary variables between regions, but their main contribution to facilitate interpretation of the outputs from regression modelling (inspired by Alkerwi et al., 2015).

4) A heat map helped explore the relative strength of associations between variables, and within them, for example, comparing the strength of correlation between dietary variables and occupation to that with education and income.

[12] A selection of visualizations is included in this report to help the reader decipher the patterns of association between the variables and support the researcher's conclusions.

3.2 Visualization design study

This section describes the methods used to address the fourth and fifth study objectives.

3.2.1 Strategy

This second component of the study can be regarded as a "design study" as described by Sedlmair and colleagues, that is, "*a project in which visualization researchers analyse a specific real-world problem faced by domain experts, design a visualization system that supports solving this problem, validate the design, and reflect about lessons learned in order to refine visualization design guidelines*" (Sedlmair et al., 2012) . Here, the domain expert and the visualization researcher were the same person, and the "real-world problem" is represented by perceived gap with respect to effective communication of research findings between researchers in public health and epidemiology, and public health practitioners. While the researcher did not design a specific "system" to help bridge the gap, she explored a general approach, that of giving some power to the practitioners to explore visualizations.

To test whether integration of interactivity into data visualizations aids knowledge transfer, and/or reduces cognitive load, a within-subjects design was chosen, whereby the respondents took part in each condition of the independent variable (here interactive or static versions of a visualization). The major advantage of this design is that it eliminates issues related to differences between individuals, which otherwise might lead to the two groups not being equivalent so that the effect (if one exists) of the condition was obscured.

The study corresponds to the second category of types of studies to assess effectiveness discussed in Section 2.2.3, that of "User studies". Given the constraints on face-to-face interaction due to the Covid19 pandemic, and the short time available for the project, the method chosen for the evaluation was an online survey using self-completion questionnaires. This technique had the advantages of being economical with respect to materials and time, and of bypassing any geographical limits to the location of respondents, and several disadvantages as outlined by Oates, such as the researcher being unable to correct misunderstandings, probe for more details, offer explanations or query disparities between answers (Oates, 2006, p230).

3.2.2 Sampling approach

The sampling approach adopted was non-probabilistic, and purposive. The target population was non-academic public health nutrition practitioners. The actual study population was professionals on the UK Voluntary Register of Nutritionists (VRN). The VRN includes nutritionist practitioners who meet certain training, competence, and professional practice criteria. The register is intended to protect the public and assure the credibility of nutrition as a responsible profession and is provided as a searchable resource on the Association for Nutrition website (https://www.associationfornutrition.org/). Together with their names, registrants can opt to provide additional information such as LinkedIn profile, website, and email and Twitter addresses. Since the VRN offered a means of identifying and contacting nutritionists working in diverse professional contexts, it was adopted as the study's sampling frame, from which the researcher compiled a list of potential survey participants using the criteria in Table 9.

	Selection Criteria	*Justification*
1	The nutritionist provides an email address on the AfN register.	The registrant had indicated a willingness to be contacted by those searching the register.
2	The nutritionist provides a LinkedIn profile on the AfN register.	The registrant had indicated a willingness for those searching the register to explore their background and experience.
3	The nutritionist's LinkedIn profile indicates a professional interest in diet and nutrition inequalities.	Those interested in the underlying subject matter would be more inclined to invest time in reading the background material and scrutinising the visualizations than other nutritionists.
4	The nutritionist does not work in an academic setting.	The project was exploring bridging the science-practice gap, and respondents needed to be from the practice side.

Table 9: Criteria for selecting potential participants from the VRN

The target sample size was 20. This number was not justified using a power calculation since the sampling strategy was non-probabilistic, instead it was justified by evidence that in a qualitative study with a fairly specific research question, little new information is gained after interviews with around 20 people belonging to one participant 'category' (Green and Thorogood, 2018, p102). The study is classified as "qualitative" rather than "quantitative" because it relates to individual cases and their subjective impressions and experiences in a natural setting, rather than a controlled one. Also, the use of a non-probabilistic sampling strategy distinguishes the study from a traditional quantitative research study which uses random sampling, and which reports numerical results together with an indication of their precision.

3.2.3 Survey design

For the within-subjects study design, the participants were exposed to each condition of the independent variable. The order in which participants encountered the static or interactive visualizations was randomised, to avoid bias due to learning effect.

The online software Qualtrics (www.qualtrics.com) was used for drafting and applying the online survey. Each participant was sent two surveys with the structure detailed in Table 10. The surveys are provided as Appendices J and K (red headings were not visible to the respondents). Two more versions of the surveys were produced, identical to those in the Appendices except for having switched the interactive visualizations into the first survey and static into the second.

Section #	*Description*	*Obligatory?*
SURVEY 1		
1	Informed consent form, binary response	Obligatory
2	Binary or multi-answer questions on participant background and experience, with options to add free-text comments	Prompted for response
3	Three visualisations (either static or interactive) and associated Likert-scale questions, with options to add free-text comments	Prompted for response
SURVEY 2		
1	Informed consent form, binary response	Obligatory
2	Three visualisations (either static or interactive) and associated Likert-scale questions, with options to add free-text comments	Prompted for response
3	Likert-scale questions on comparative effectiveness of static and interactive viz, with options to add free-text comments	Prompted for response
4	Statements on visualizations in academic journals, with request for free-text response	Prompted for response

Table 10: Survey structure

For the Likert-scale questions, respondents needed to specify their level of agreement or disagreement on a scale for a series of statements. The scale was “even-point” with four options, so that a neutral option of "neither agree nor disagree" was not available, to avoid respondents passively taking what could be considered the easy central option when they were unsure. Respondents could actively choose to express their uncertainty, by choosing the fifth option offered of “Don’t Know”. For all questions text boxes were provided for addition of comments.

The draft survey and background information sheet were informally pilot tested by two ex-colleagues of the researcher to check for ambiguity, difficulties, need for additional instructions, whether the pre-defined responses covered all likely options, and whether the estimated necessary time for completion was reasonable (10 minutes per survey). After minor amendments the participant information sheet, surveys and background information sheet were submitted for supervisor approval according to the process required by the University’s Research Ethics Framework (Section 3.3).

3.2.4 Survey content

Participant background: The first section of the first survey was designed to obtain contextual information about sources of research evidence used by participants (inspired by Brownson et al., 2018), and their experience of reporting and communicating research evidence. This section was intended to provide contextual information to help interpret the evaluation results.

Evaluations: The following definition of an effective visualization, derived from a review by Zhu (Zhu, 2007), was used to guide the question construction: "*An effective visualization transfers knowledge and reduces the cognitive load for users over non-visual representations*." The intention was to situate the participant as though viewing visualisations in an academic journal article. To this end, a single sheet of background information was provided (Appendix I) and participants were requested to view the visualizations together with this sheet.

1) Transfer of knowledge was explored in section 3 of the first survey and section 2 of the second. Each visualisation was followed by three questions with the aim of eliciting whether the visualization had generated "*insights*", that is, whether the participant had seen something that had previously not been noticed or saw something in a new light, more formally defined as *"..an individual observation about the data by the participant, a unit of discovery"* (Saraiya et al., 2005). The wording of the three questions was inspired by those used is study comparing two models of visualization use to deliver stories (Badawood and Wood, 2014). Two additional questions were included to obtain feedback on the visualization that might help interpretation of the results, with respect to design or the participant's perceived deficient expertise.

2) Reduction of cognitive load was explored in section 3 of the second survey. Cognitive load is defined here as the "*Amount of work needed to interpret a visualization*" and has been proposed as an objective measure of visualization effectiveness (Anderson et al., 2011). According to cognitive load theory, the way information is presented can affect task performance through its effect on the load placed on the working memory system (Anderson et al., 2011). Direct questions as to the relative speed and ease of gaining insights from the static or interactive versions were included, together with others to help interpret the results.

Attitudes: The final section of the second survey was designed to explore practitioners' views on the use of visualizations in academic journals, to inform discussion about practical implication of the survey findings. Two statements were provided (largely inspired by McInerny et al., 2014, , and Arevalo et al., 2020) and the survey participants were invited to comment.

3.2.5 Survey visualizations

The requirements for the visualizations included were twofold, first it should be possible to have two versions of each, that is, static and interactive; and second there should be a range of forms in case effectiveness depends on the type of visualisation. Three forms of visualizations were included, becoming progressively more complex to reduce the risk that intimidation of participants would lead to survey non-completion. The forms chosen were choropleth maps, a line graph and bar charts, as described below. See Appendix J for the static versions, while the web addresses for the interactive versions are included here as they are truncated in Appendix K.

- Choropleth maps: This visualization illustrated temporal and geographical variation in four dietary variables (two continuous and two binary categorical). Lightness of a single hue was used to differentiate quantities. The static version had a series of "small multiples" – maps of the UK for four dietary variables over nine years, so a total of 36 maps presented in four rows. In the interactive version there was just one map for each dietary variable, and the user chose the year to view using a slider bar. "Tooltips" were displayed as long as the user hovered over a region. The dietary variables selected for display were Median red meat intake, Median sodium intake, %Population consuming alcohol, %Population with low folate intake - chosen because they had statistically significant temporal variation, and contrasting patterns of geographical variation. https://tufty317.github.io/web/MapViz_forSurvey_interactive_3.html

- Line chart: This visualization illustrated socio-economic and geographic variation in one dietary variable. The line chart was the only visualization of the three with the same form in static and interactive versions. The chart plotted just one dietary variable (Median total fruit and vegetable intake) against four categories of occupation, by region, of which there were 12. Lines joined the symbols corresponding to the same region, thereby supporting the user's perception of change, following the organizing principle of Connectedness (Ware, 2013, p183). Curved not straight lines were used, following the Gestalt principle of Continuity[13] (Ware, 2013, p184). The use of shape and colour applied the Gestalt principle of Similarity (Ware, 2013, p182) to group pattern elements together. It was necessary to use the shape channel as well as colour for visual encoding because only six to twelve colours are discriminable by users (Munzner and Maguire, 2014, p226).

  In the interactive version, users could choose a single region by clicking a line in the chart or a symbol in the legend. This would highlight that region, and the others faded into light grey

13 "We are more likely to construct visual entities out of visual elements that at smooth and continuous rather than ones that contain abrupt changes in direction."

thereby using colour to harness a “figure-ground” effect[14] described by the Gestalt psychologists ((Ware, 2013, p189). “Tooltips” were displayed as long as the user hovered over a symbol. https://tufty317.github.io/web/LineViz_forSurvey_interactive_2.html

- Bar charts: This visualization illustrated socio-economic and demographic variation in two dietary variables. The bar charts plotted regression coefficients for normalised and standardised dependent variables. The researcher recognised before the survey was applied that this visualization would need some familiarity with statistical techniques and language to interpret. The static version had six charts, three (one for each of the SES variables occupation, income and education) for each of the dietary variables (Fruit and vegetable intake and Red meat intake). In the interactive version the three charts for each dietary variable were overlain, and the user chose which of the three SES variables to view by clicking on the legend. https://tufty317.github.io/web/BarViz_forSurvey_interactive.html

Litvis software (Wood et al., 2018) was used to design the visualizations because the flexibility and notebook environment aided documentation of the process of creation and fine-tuning (Section 3.1.4.3). Litvis uses the functional language Elm together with the Vega-Lite grammar of interactive graphics (Satyanarayan et al., 2016). Use of the Coblis website ensured that where possible the colours used were distinguishable by users with the most common form of colour blindness (https://www.colorblindness.com/coblis-color-blindness-simulator/).

3.2.6 Recruitment and survey administration

- The researcher emailed invitations (Appendix G) to 36 nutritionists identified from the AfN register who fulfilled the criteria shown in Table 9. The participant information sheet was included in the email rather than as an attachment, to reduce necessary actions for potential participants. Ten individuals replied favourably and were emailed links to the surveys and background information sheet (Appendices H and I). Allocation of the two sets of surveys (static visualization first or interactive first) was done alternately according to the order of receipt of acceptances, to ensure equal numbers of the two sets were distributed and bias minimised. This resulted in five completed double survey sets and one single survey.
- Recruitment was boosted by contacting another ten nutritionists identified from the VRN without an email address provided. They were sent messages on the LinkedIn platform together with invitations to connect, leading to four expressions of interest to participate, and resulting in two completed double survey sets and one single survey. More participants were recruited from the

[14] Whereby a “figure” is something objectlike that is perceived as being in the foreground.

four facilitators of the Scottish AfN regional facebook site whom the researcher asked to post a message about the study (the Facebook message itself did not provoke interest). Two expressed interest, and both contributed double survey sets.

- The researcher contacted three potential participants from her own network, with equivalent professional experience and interests to those recruited from the VRN. Two completed a double survey set.

The final sample size was 13 of which 11 completed both surveys and two contributed just one survey (no reasons were provided for this).

3.2.7 Survey data analysis

The software used for data analysis was SPSS Statistics for Windows (V25.0, Armonk, NY: IBM Corp). For this second study component the analysis was mainly descriptive. The values reported or illustrated are simply frequencies, as it was not correct to report sampling errors in terms of confidence intervals or standard errors given that these methods assume random sampling.

To compare responses to the static and interactive charts, the scores were compiled and ranked for each of the five evaluation questions and three visualization forms, and mean ranks compared between the static and interactive formats. Also, the non-parametric test *Wilcoxen matched-pairs* was applied to the data collected on equivalent static and interactive charts, because although the p and Z values cannot be interpreted literally (due to the purposive sampling), the values reflect the relative size of difference between scores for static and interactive charts, while taking the within-subjects study design into account. Simple graphics were produced using the dialog boxes of the Chart Builder facility in SPSS to create the basic bar charts, and then formatting such as bar orientation and label font size was amended using the interactive SPSS chart editor.

For open-ended questions, key themes were identified, and responses summarised by theme. Direct quotations were selected for inclusion in the report to illustrate views expressed.

3.3 Research ethics

This section describes the efforts made to treat the participants in the research with dignity, and wherever possible, to gain some benefit from the research. The first five rows of Table 11 below outlines how the five "rights" of research participants (Oates, 2006, p56) were addressed in the two study components, and the sixth and final row describes the benefits of participation. Ethics approval was obtained for Years 1-5 of the NDNS data from the Oxfordshire A Research Ethics Committee (Ref. 07/H0604/113), and for Years 6–9 from the Cambridge South NRES Committee

(Ref. 13/EE/0016). For the visualization survey, the Research Ethics review form is included in the study proposal (Appendix A), and supervisor approval was obtained for the survey according to the requirements of City University's Department of Computer Science Research Ethics Committee (CSREC) (http://www.city.ac.uk/department-computer-science/research-ethics).

	Secondary dataset (NatCen Social Research, 2019a)	*Visualization evaluation*
Right not to participate	(The interviewers' documentation does not specify that this is mentioned to participants.)	Many individuals were invited to participate, and only those who actively responded expressing interest were sent links to the survey (Appendix H).
Right to withdraw	(The interviewers' documentation does not specify that this is mentioned to participants.)	Participants were free to withdraw themselves and their data at any point in the study up until submission of the report. This was specified in the participant information sheet (Appendix G) and in the consent forms before the questionnaires (Appendices J & K).
Right to give informed consent	An advance letter, describing the purpose of NDNS is sent to all sampled addresses a few days in advance of fieldwork. The letter briefly describes the study. Participants are informed that if they would like more information about the survey they can visit the NDNS website www.nationaldiet.co.uk, email NDNS@natcen.ac.uk, or telephone NatCen. At the first visit, the interviewer answers any remaining questions, and obtains consent using initials or tick mark.	Those nutritionists selected for potential participation were emailed a participant information sheet (Appendix G). Those willing to participate were required to complete a consent form at the start of the questionnaires (Appendices J and K).
Right to anonymity	Participants' names and any other identifiable information like postcode were removed from the datasets so participants cannot be identified.	No personal data were collected during the survey except for contact details. When the data are made accessible to the supervisor and examiners it will be in anonymised format (email addresses and names removed).
Right to confidentiality	No-one outside the research team knows who has taken part, or will be able to identify an individual's results.	The survey data are only stored on the researcher's own desktop computer. The data will be retained securely only until the researcher's graduation and then destroyed.
Benefits of participation	Participants who completed at least 3 food diary recording days were asked if they would like feedback on the analysis of their records. The feedback described how their intake compared to nutrient intake recommendations, and general information on sources of healthy eating advice. There is a £30 token of appreciation for each individual taking part.	Short-term: Findings from the research will be shared with those participants who request this. Long-term: Findings from the research may inform future guidelines on visualization design and thereby potentially help participants with their work.

Table 11: Ethical aspects of the study

4 RESULTS

4.1 Secondary data analysis

4.1.1 Sample characteristics - demographic, socio-economic and geographical

Table 12 gives the demographic, socio-economic and geographical characteristics of the adult sample from the NDNS 2008-16. Population representative values are provided on the left, having corrected the data for survey design and non-response (Section 3.1.4.3), and unadjusted values derived directly from the dataset are on the right. The adjustment particularly affected the regional breakdown, for example, London was under-represented in the data, accounting for 6% of the actual adult sample and 13% of the adjusted sample. Scotland and Northern Ireland were oversampled, accounting for 16% and 13% of the actual sample, and 9% and 3% of the adjusted sample respectively. All results reported after Table 12 are derived from the adjusted sample.

		Corrected (Weighted)			Uncorrected		
		Median & mean	**N**	**%**	**Median & mean**	**N**	**%**
Gender	Female		3497	51.4		3992	58.7
	Male		3305	48.6		2810	41.3
	Total		*6802*	*100.0*		*6802*	*100.0*
Age (years)		47.0 and 48.3	6802		49.0 and 49.9	6802	
	19-34		1844	27.1		1524	22.4
	35-49		1860	27.3		1994	29.3
	50-64		1601	23.5		1705	25.1
	65+		1496	22.0		1579	23.2
	Total		*6802*	*100.0*		*6802*	*100.0*
Ethnicity	Non-white		775	11.4		494	7.3
	White		6023	88.8		6300	92.7
	Total		*6798*	*100.0*		*6794*	*100.0*
Occupation	Routine and manual		2214	34.2		2421	37.3
	Intermediate		1371	21.2		1399	21.6
	Lower managerial and professional		1703	26.3		1632	25.2
	Higher managerial and professional		1193	18.4		1032	15.9
	Total		*6481*	*100.0*		*6484*	*100.0*
Equivalised income		£27,500 and £32,672	5745		£27,500 and £31,662	5827	
	Category 1		1773	30.9		1865	32.0
	Category 2		1227	21.4		1193	20.5
	Category 3		1411	24.6		1329	22.8
	Category 4		1333	23.2		1440	24.7
	Total		*5745*	*100.0*		*5827*	*100.0*

		Corrected (Weighted)			Uncorrected		
		Median & mean	N	%	Median & mean	N	%
Education	No qualifications		1323	21.5		1586	25.3
	GCSE or equivalent		1382	22.5		1404	22.4
	A level or equivalent		1650	26.8		1618	25.8
	Degree or higher		1797	29.2		1652	26.4
	Total		*6151*	*100.0*		*6260*	*100.0*
Region	1 North East		283	4.2		235	3.5
	2 North West		754	11.1		539	7.9
	3 Yorkshire & the Humber		570	8.4		374	5.5
	4 East Midlands		491	7.2		365	5.4
	5 West Midlands		594	8.7		455	6.7
	6 East of England		632	9.3		472	6.9
	7 London		871	12.8		426	6.3
	8 South East		929	13.7		649	9.5
	9 South West		582	8.6		390	5.7
	10 Wales		330	4.9		952	14.0
	11 Scotland		578	8.5		1073	15.8
	12 Northern Ireland		188	2.8		872	12.8
	Total		*6802*	*100.0*		*6802*	*100.0*

Table 12: Descriptive statistics of sample characteristics

4.1.2 Sample dietary characteristics

Table 13 shows the basic descriptive statistics for the 16 dietary constituents considered in the study. Five constituents have two variables each, because there are alternative modes of assessing them – for fruit and vegetables, the two modes are absolute intake and % achievement of the "5-a-day" target, and for iron, calcium, folate, and Vitamin A the two modes are absolute intake and whether intake falls below recommended levels. Where two variables relate to the same food or nutrient they are included in the same row of the table. Standard deviations are provided but must be interpreted with caution because the distributions of all the continuous variables are significantly positively skewed (apart from the variable Fat%Energy) (Section 3.1.4.2).

		Continuous			Binary		
		Median	Mean	SD		N	%
Fruit and vegetables	Daily intake (g)	256	291	181	Achieve No	4766	70.1
					5 a day: Yes	2036	29.9
Red and processed meat	Daily intake (g)	57.0	65.4	51.3			
Oily fish					Consume: No	4867	71.6
					Yes	1935	28.4
Soft drinks, not low calorie					Consume: No	3715	54.6
					Yes	3087	45.4
Total energy	Daily intake (kcal)	1746	1812	576			
Fat % food energy	%	34.9	34.8	6.4			
SFA % food energy	%	12.7	12.9	3.5			
NMES % food energy	%	10.7	11.7	6.5			
Alcohol					Consume: No	3191	46.9

		Continuous			Binary		
		Median	Mean	SD		N	%
					Yes	3611	53.1
AOAC Fibre	Daily intake (g)	17.7	18.5	7.1			
Sodium	Daily intake (mg)	2031	2134	828			
Iron	Daily intake (mg)	10.0	10.4	3.7	Below LRNI: No	6051	89.0
					Yes	751	11.0
Calcium	Daily intake (mg)	775	816	322	Below LRNI: No	6347	93.3
					Yes	455	6.7
Folate	Daily intake (µg)	234	249	105	Below LRNI: No	6565	96.5
					Yes	237	3.5
Vitamin A	Daily intake (retinol equivalent units)	720	971	1087	Below LRNI: No	6226	91.5
					Yes	576	8.5
Vitamin D	Daily intake (µg)	2.40	2.90	2.13			

Table 13: Descriptive statistics for dietary variables

Note: Sample size for all variables N = 6802

4.1.3 Dietary variation by socio-economic status

The findings in this section address the first study objective: "Ascertain the patterns of variation in the UK diet with socio-economic status"

4.1.3.1 Patterns of associations – unadjusted

	Continuous variables				Binary Variables			
		Occupation	Income	Education		Occupation	Income	Education
Fruit and Veg	Intake	0.19	0.18	0.25	5 a day	0.14	0.13	0.19
Red meat	Intake	-0.05	(0.04)	-0.10	-			
Oily fish	-				Consume	0.15	0.15	0.13
Soft drinks	-				Consume	(0.01)	(0.01)	0.05
Total energy	Intake	0.12	0.16	0.15	-			
Fat %Energy	%	0.10	0.11	0.09	-			
SFA %Energy	%	0.05	0.05	(-0.01)	-			
NMES%Energy	%	-0.05	-0.06	-0.07	-			
Alcohol	-				Consume	0.19	0.22	0.14
Fibre	Intake	0.15	0.16	0.20	-			
Sodium	Intake	(0.03)	0.12	0.06	-			
Iron	Intake	0.16	0.19	0.20	< LRNI	-0.08	-0.12	(-0.04)
Calcium	Intake	0.11	0.13	0.07	< LRNI	-0.09	-0.09	-0.05
Folate	Intake	0.16	0.17	0.14	< LRNI	-0.07	-0.09	-0.07
VitaminA	Intake	0.12	0.10	0.12	< LRNI	-0.10	-0.10	-0.08
VitaminD	Intake	0.06	0.11	0.05	-			

Table 14: Heatmap of non-parametric correlation coefficients for dietary v. SES variables

Notes (1) Variables relating to the same food or nutrient are included in the same row to aid comparisons;
(2) "Listwise" rather than "pairwise" deletion was used to ensure the same sample (4794 adults) was used for all three SES variables;
(3) All correlations were highly significant (p<0.001) except for those in brackets.

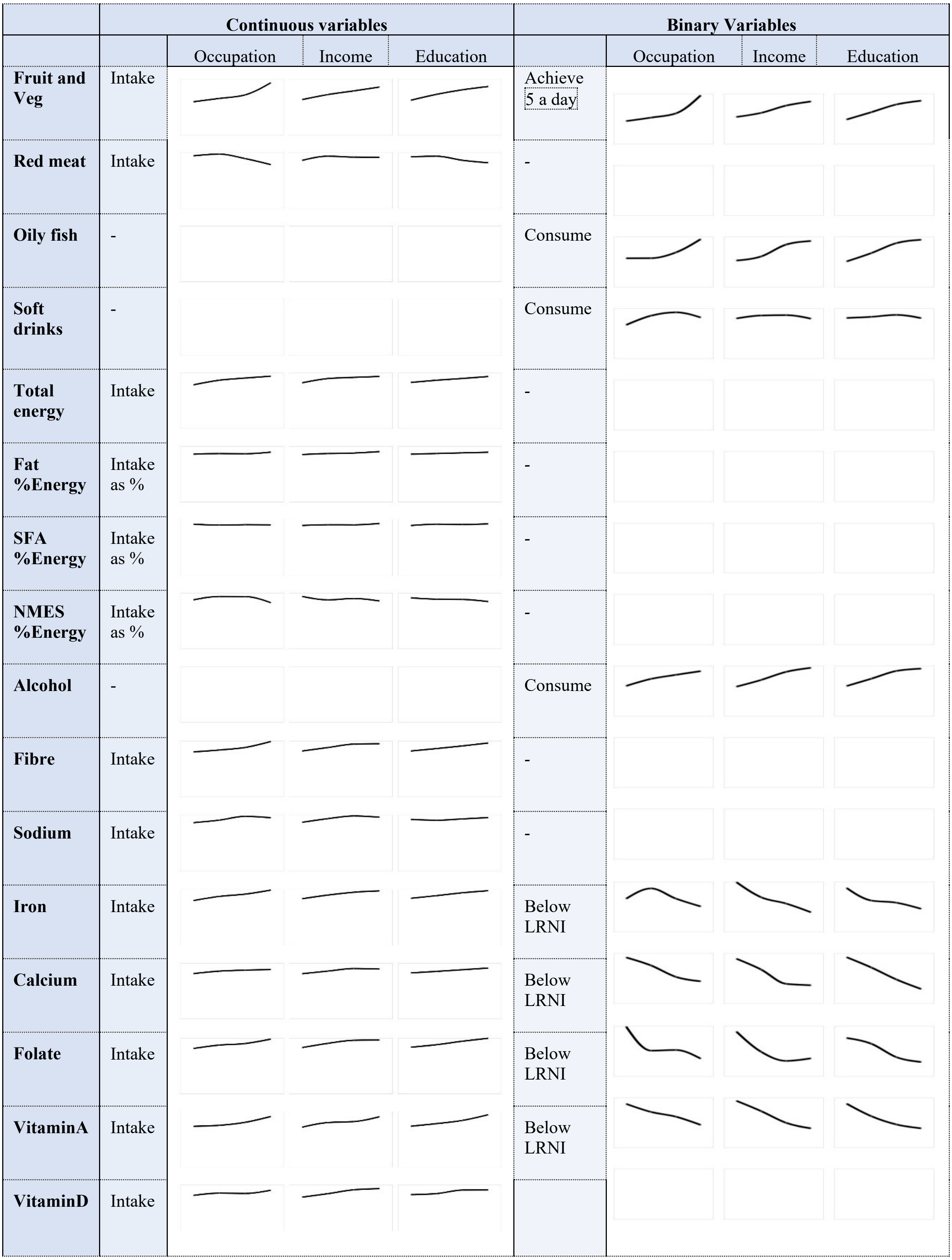


Figure 3: Sparklines showing associations between dietary and SES variables

Notes: For continuous dietary variables Y axes represent median values; for binary dietary variables Y axes represent proportion of sample; X axes represent four SES categories.

Table 14 is a heatmap enabling comparison of association strength between and within variables. Correlations between dietary variables and SES were all positive (green) except for RedMeat (for occupation and education), NMES%Energy and the four variables for micronutrient intake <LRNI. Table 14 shows the strongest associations between dietary and SES variables were for Fruit&Veg (both variables); OilyFish; TotalEnergy; Alcohol; Fibre; Iron intake and Folate intake. The table also shows that associations with SES variables differ between the dietary variables - for VitaminD and BelowLRNIIron, association with income is notably stronger than with occupation or education, while for Fruit&Veg (both variables), Fibre and SoftDrinks, correlation with education is stronger than with occupation or income.

The sparklines of Figure 3 illuminate the data of Table 14. The steep slopes of the lines for the negative associations between SES and proportions falling below LRNI for the four micronutrients are noteworthy and are due to the low proportions deficient in the highest SES category.

Relative inequality

Figure 4 plots the relative differences (Section 3.1.4.3) between highest and lowest SES categories. The continuous dietary variables were sorted in descending value of mean across the three SES inequality values, while the four variables for proportion <LRNI were plotted separately in ascending order, since their inequality values were much greater than for the other variables, and all were negative.

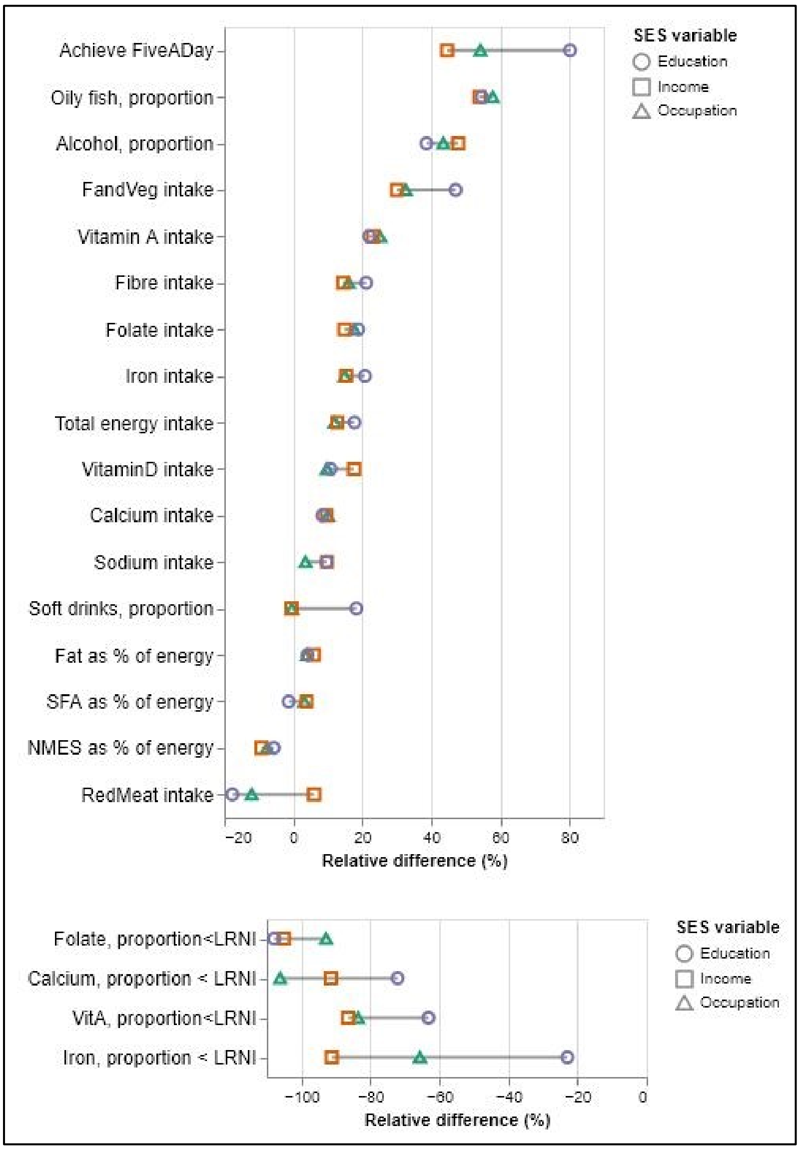


Figure 4: Ranged dotplots of relative inequality in 21 dietary variables

Note: A positive relative difference shows higher intake of the food or nutrient (for continuous variables) or higher proportions of the sample (for binary variables), for the highest SES category compared to the lowest SES category.

The values underlying Figure 4 are provided in Tables A16 and A17. Figure 4 shows the greatest RI values were for the variables relating to Fruit and Vegetables, OilyFish, Alcohol, and Folate < LRNI.

Findings from regression modelling

Based on the patterns indicated in Table 14 and Figures 3 and 4, and on previous studies, a sub-set of six variables was chosen for detailed analysis (Section 3.1.4.3). Tables A18 and A19 provide unadjusted values of these six variables by SES category, together with the values for R Squared and *p* of regression models including just the dietary variables (normalised where necessary and standardised to enable comparison of regression coefficients between dietary variables) and categorical SES variables. Values of R Squared in Tables A18 and A19 do not completely correspond to squared values of the correlation coefficients in Table 14, first because the values in Table 14 are for Spearman not Pearson coefficients, and second because sample sizes in Tables A18 and A19 are greater than for Table 14 (where the sample was restricted to those with values for all SES variables).

Figure 5 displays the regression coefficients for this modelling. Error bars show 95% confidence intervals (CI) around the values for each category in comparison with the reference categories (highest SES). Coefficient values greater than zero indicate positive associations. For Fruit and Vegetable intake on the left side of the figure, the negative values show that intakes for the first three SES groups have lower intakes than the highest SES group, and the difference is most pronounced for education. For Folate consumption, the positive values show that the undesirable outcome of prevalence of Folate intake <LRNI is highest in the lowest SES groups, and (except for occupation group 2) this is the only group that differs significantly from the highest SES group.

4.1.3.2 Dietary associations with SES adjusted for demographic characteristics

Confounding of associations with SES by demographic variables

Figure 6 displays regression coefficients for two dietary variables from regression modelling identical to that described above, except for inclusion of demographic covariates. Error bars show 95% CI for each category compared to the reference categories (highest SES, male, ≥65y and white ethnicity). Figure A25 shows the equivalent charts for the other four key dietary variables. The length of the green bars and the width of the CI in Figure 6 can be compared with those in Figure 5 to examine whether demographic variables were confounding the relationships between dietary and SES variables.

The notable difference between the charts is only for education, whereby the green bars' lengths increased, and the CI width decreased between Figures 5 and 6 for both Fruit & vegetable intake and Folate<LRNI.

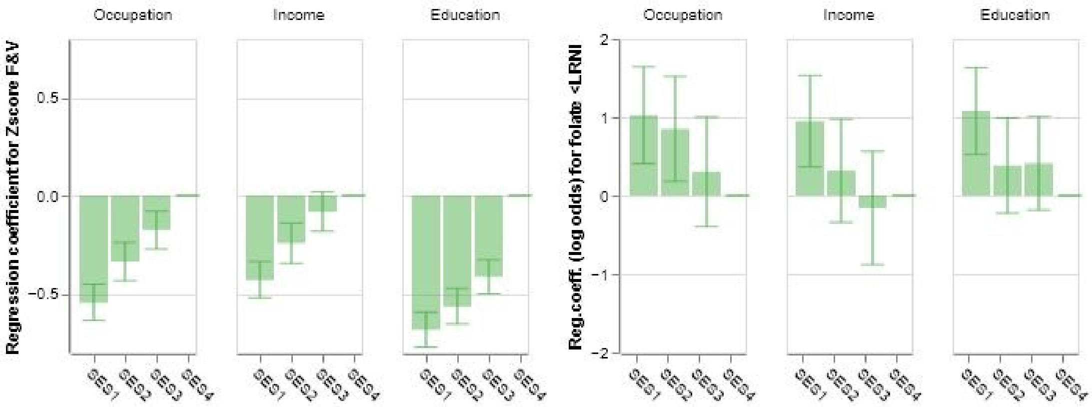


Figure 5: Barcharts of regression coefficients for Fruit & vegetable intake and Folate<LRNI, from models for each SES independent variable separately

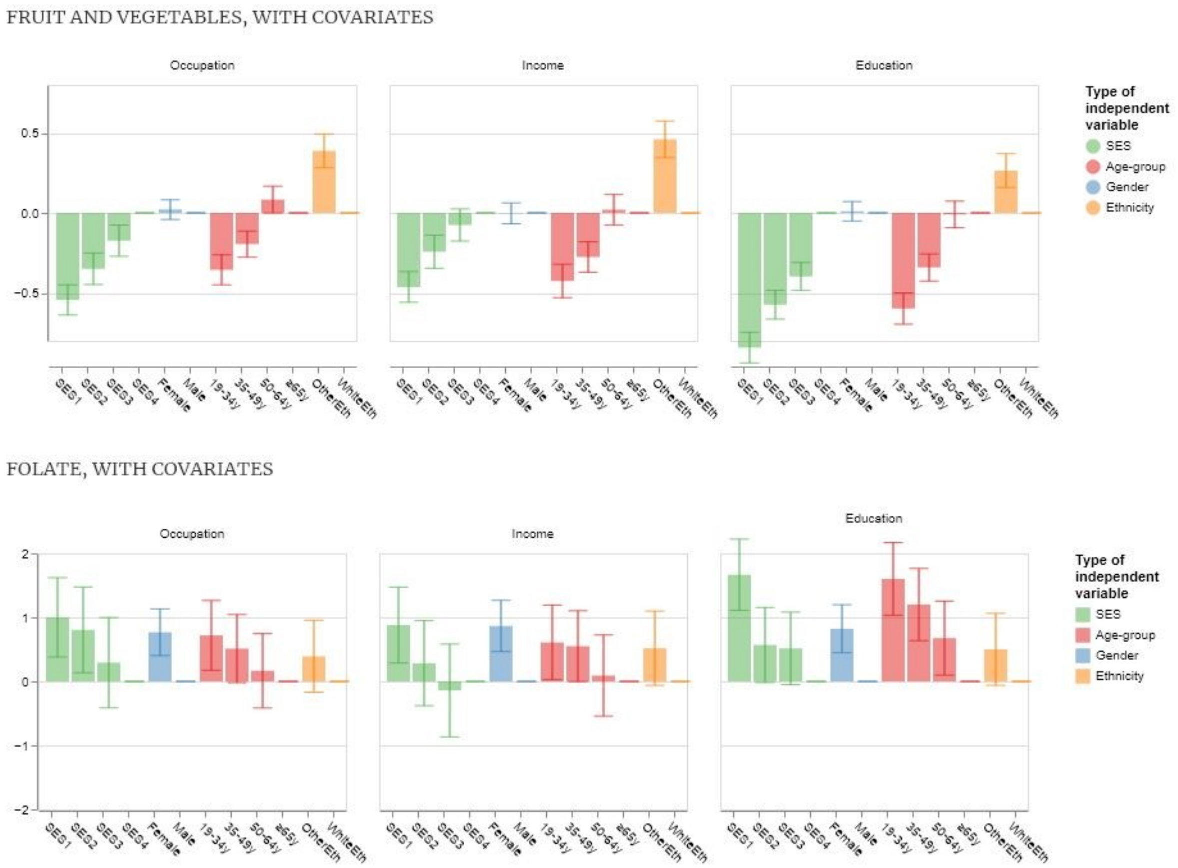


Figure 6: Barcharts of regression coefficients of Fruit & vegetable intake and Folate<LRNI, from models for each SES independent variable separately, with demographic covariates

For Folate, after adjustment for the demographic variables, the CI for education SES groups 2 and 3 include zero, so it is now only the lowest group which differs significantly from the highest group (Figure 6). Figure A25 also shows that for the other dietary variables, where differences exist in SES regression coefficients for models with and without demographic covariates, these differences are subtle, and affect associations of the dietary variables with education rather than occupation and income. The most notable changes (increases in absolute values of the education coefficients due to adjustment) are for Alcohol and Oily fish.

Figures 6 and A25 also enable comparison of strength of associations of dietary variables with demographic variables, both between and within variables. For example, fruit and vegetable intake was lower in the younger and white ethnicity adults compared to the older and non-white ethnicity groups (top row of Figure 6). Red meat intake was lower in women and adults of non-white ethnicity compared to men and adults of white ethnicity (top row of Figure A25).

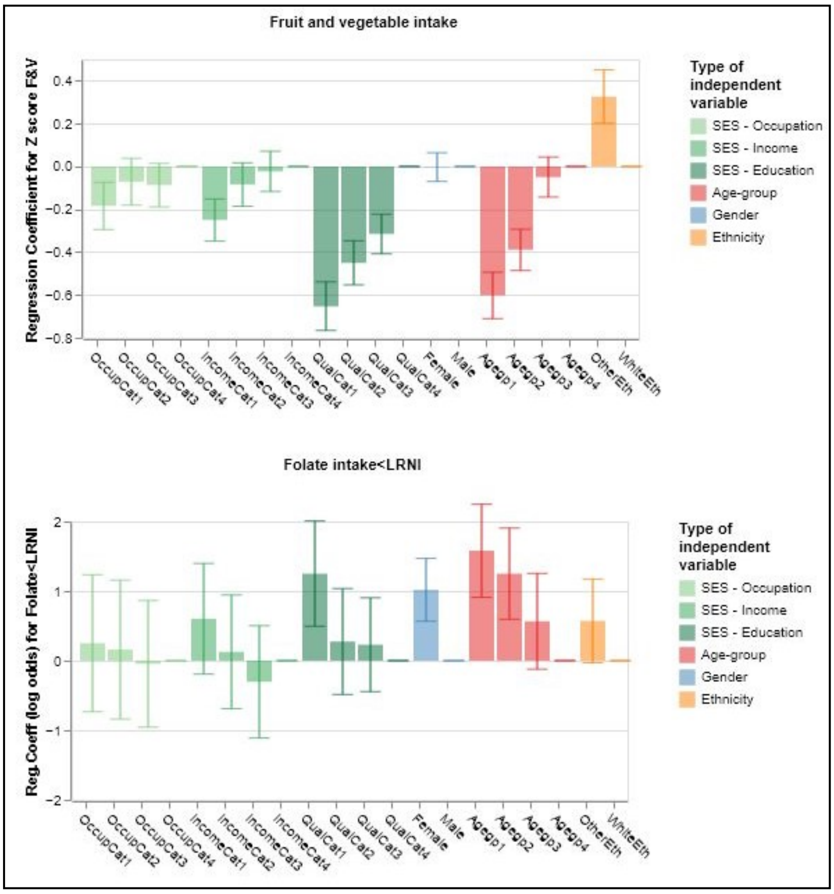
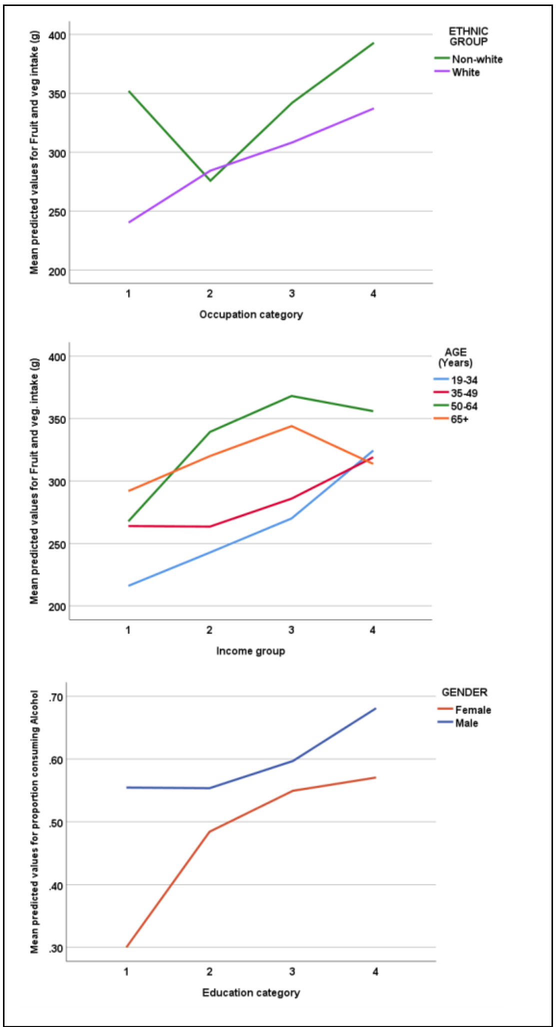


Figure 7: Bar charts of regression coefficients for Fruit and vegetable intake, and Folate<LRNI, from models with the three SES variables included simultaneously, with demographic covariates

Figure 7, and Figure A26 are bar charts for regression coefficients for models with all three SES variables included together, to examine the extent to which the three SES variables have independent effects. Comparing Figure 7 with Figure 6 shows that when education is included in models together with occupation and income, the associations of the latter two SES variables with Fruit and Vegetable intake and Folate<LRNI are no longer significant (CIs include the value of zero). In contrast, comparing Figures A25 and A26 show that for Alcohol and Oily Fish consumption, income remains significant when education is included with it in regression models.

Interactions of SES with demographic variables

As described above, Figures 6 and A25 showed that statistical adjustment for age-group, sex and ethnicity generally did not impact on the associations between key dietary variables and SES, indicating that demographic variables were not important as confounding variables. But the possibility remained they were interactors, modifying the effect of SES on dietary variables. For testing for interactions, the critical probability value was adjusted using the Bonferroni correction.[15]



18 of the 54 interactions tested were significant at the 0.05 level, but only three were significant at below the critical 0.001 level. These are shown in Figure 8 - two were for Fruit & Vegetable intake, first Ethnicity interacted with Occupation (such that the trend to increase intake with increasing income was weaker for the non-white group), and second, AgeGroup interacted with Income (such that the trend for increased intake with increasing income was weaker for the two older categories). For Alcohol, Gender interacted with Education such that the trend for proportion consuming Alcohol to increase across increased education level was greater for women than men.

Figure 8: Line plots to examine interactions between demographic and socio-economic variables for Fruit & Vegetable intake (upper and centre) and proportion consuming Alcohol (lower plot).

Summary of findings in Section 4.1.3 about dietary variation by socio-economic status

- Associations between dietary variables and SES for the most part followed expected trends, whereby dietary variables that were protective for health showed positive associations with SES while dietary variables that were risk factors were negatively associated with SES.
- However, for Red meat intake, the three SES variables had independent and opposite effects (intake increased with income and decreased with occupation and education).
- For the six dietary variables tested, there was little evidence of demographic variables confounding the dietary v. SES associations, but some evidence of interactions.

[15] 54 hypotheses were tested (3 demographic variables x 6 dietary variables x 3 SES variables, so the normal threshold p-value of 0.05 was divided by 54 and the result rounded to give a significance threshold of p <0.001, to maintain 95% confidence in the set as a whole.

4.1.4 Geographical variation in dietary inequalities

The findings in this section address the second study objective: *"Explore the extent to which the associations between diet and SES differ between regions*". First, we examine simple geographical variation in dietary variables, separately from SES.

4.1.4.1 Patterns of geographical variation in dietary variables – unadjusted

Table 15 shows findings from simple regression modelling of the 21 dietary variables, including Region as independent variable. Using the modified probability threshold value of 0.0024 (0.05/21 hypothesis tests), 14 of the 21 variables showed significant geographical variation.

	Continuous variables			Binary Variables		
		Rsquare	*p*		Cox and Snell R sq	*p*
Fruit and Veg	**Intake**	0.027	*<0.001*	**5 a day**	*0.013*	*<0.001*
Red meat	**Intake**	0.011	*<0.001*	**-**		
Oily fish	**-**			**Consume**	*0.011*	*<0.001*
Soft drinks	**-**			**Consume**	*0.003*	*0.066*
Total energy	**Intake**	0.007	*<0.001*	**-**		
Fat %Energy	**%**	0.005	*0.002*	**-**		
SFA %Energy	**%**	0.012	*<0.001*	**-**		
NMES%Energy	**%**	0.007	*0.001*	**-**		
Alcohol	**-**			**Consume**	*0.005*	*<0.001*
Fibre	**Intake**	0.014	*<0.001*	**-**		
Sodium	**Intake**	0.005	*0.027*	**-**		
Iron	**Intake**	0.007	*<0.001*	**< LRNI**	*0.003*	*0.031*
Calcium	**Intake**	0.016	*<0.001*	**< LRNI**	*0.007*	*0.003*
Folate	**Intake**	0.011	*<0.001*	**< LRNI**	*0.004*	*0.128*
VitaminA	**Intake**	0.010	*<0.001*	**< LRNI**	*0.008*	*0.003*
VitaminD	**Intake**	0.004	*0.069*	**-**		

Table 15: Findings from regression modelling of 21 dietary variables with region as independent variable
Note: N = 6802; Continuous dietary variables were normalised and standardised; p values <0.0024 are highlighted.

Choropleth maps for all variables are provided in Figure A27. Figure 9 shows maps for two of these variables that show interesting contrasting patterns – Fruit and Vegetable intake is highest in London and lowest in Northern Ireland, while NMES % food energy is low in both regions.

Comparison of the maps of Figure A27 reveals some dietary variables co-vary following expected patterns, for example, variation in Fruit and vegetable intake, % achieving 5-a-day, Fibre intake and Vitamin A intake is similar. However, some between-variable variation is inconsistent. For example, red meat intake is highest in Northern Ireland, but iron intake is not, and it has the highest percentage with iron intake below recommended levels. Also, the maps show Scotland has the highest intake of sodium, and interestingly one of the lowest proportions of the population consuming alcohol (Northern Ireland has the lowest). Overall, the South-West region appears to have the healthiest diet, showing high intakes of Fruit & Vegetables, Iron, Folate, Calcium, Vitamin A and Vitamin D.

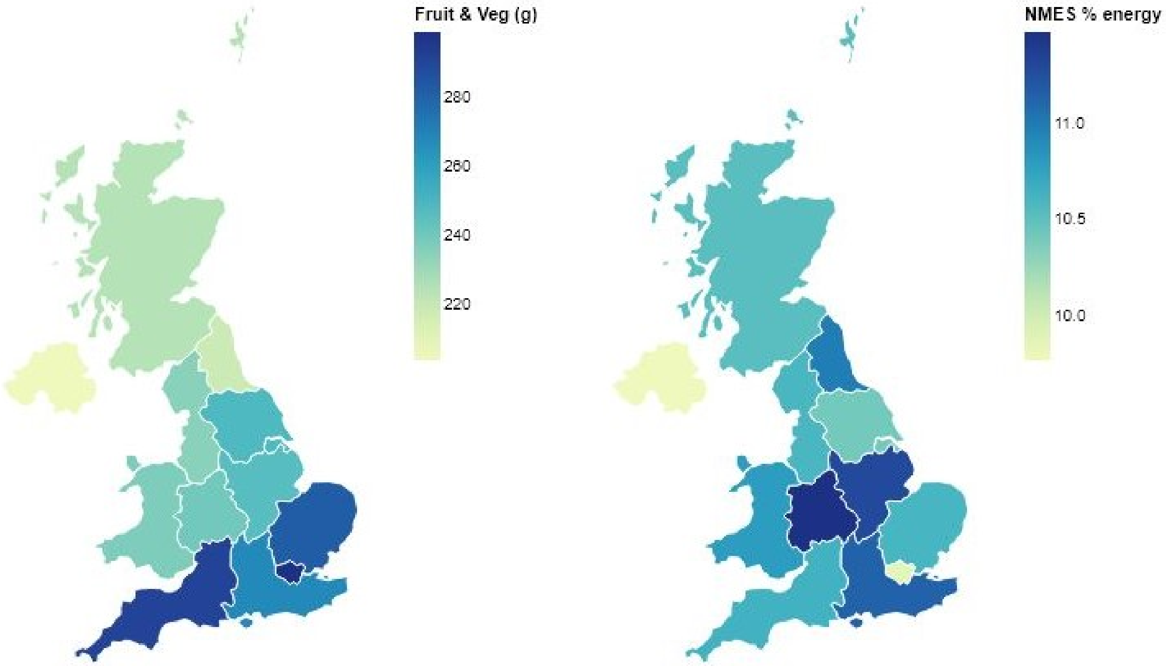


Figure 9: Choropleth maps of median values for Fruit and Vegetable intake, and NMES as % of food energy

4.1.4.2 Geographical variation in associations between dietary variables and SES status

Now moving to directly address the second research objective, Figure 10 plots RI values for each region for the same dietary variables as the choropleth maps, and Figure A29 plots RI values for the other four key dietary variables. For reference, all values are provided in Table A23. Insights gained from comparison between regions for each variable include the generally higher inequality in Fruit and Vegetable intake in the northern regions of England and in the other UK countries compared to the southern English regions (left of Figure 10), and how inequality in red meat intake in London is consistently positive in contrast to the other UK regions (top left of Figure A28).

Figures 10 and A28 also reinforce the value of considering all three socio-economic variables. For example, Figure 10 shows that within regions, inequality in Fruit and vegetable intake does not vary greatly between the three SES measures, and all three measures are consistently positive. But for Red meat, the direction of inequality depends on the SES measure, usually being positive for income and negative for occupation and education (except for London, as described above). These findings are consistent with, and provide more detail to elaborate, those shown in Table 14 and Figure 4. Figure A28 shows that within dietary variables, the three inequality values for London are consistently similar (except for Alcohol consumption), while for regions like the North-East and South-West the ranges of RI values are wider.

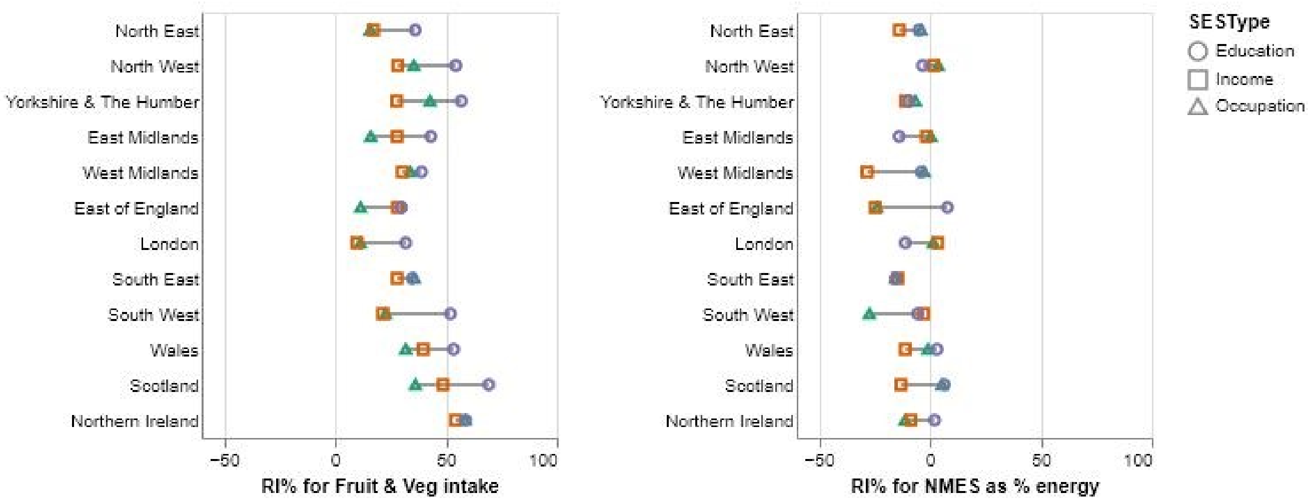


Figure 10: Ranged dotplots of relative inequality in Fruit and Vegetable intake, and in NMES as % of food energy, by geographical region

Note: A positive relative difference shows higher intake of the food or nutrient in the highest SES category.

Figure 11 shows grid maps for total Fruit and Vegetable intake, and Folate intake < LRNI, while maps for the other four key variables are included in Figure A29. These maps enable comparison within dietary variables of geographical variation by SES categories – moving left to right these are in turn occupation, income, and education. Table 14 showed that Fruit and vegetable intake was more strongly linked to education than to income and education, and that low Folate was more strongly linked to income. Figure 11 shows whether the variation occurs uniformly across the 12 regions.

The grid maps give the impression there are not well-defined deviations from expected patterns, and findings from regression modelling of interactions with geographical region with SES confirm this (Table A21). For testing for existence of interactions between region and SES, the critical probability value was adjusted using the Bonferroni correction. Three of the 18 interactions tested were significant at the 0.05 level (Table A21) but with the modified threshold value of <0.003 (0.05/18 hypothesis tests), only one was statistically significant, that of Folate <LRNI, for the interaction of Income with Region. The centre lowest grid map of Figure 11 shows that it is in the regions of East Midlands and West Midlands that the general trend for folate deficiency to fall with increasing income does not occur.

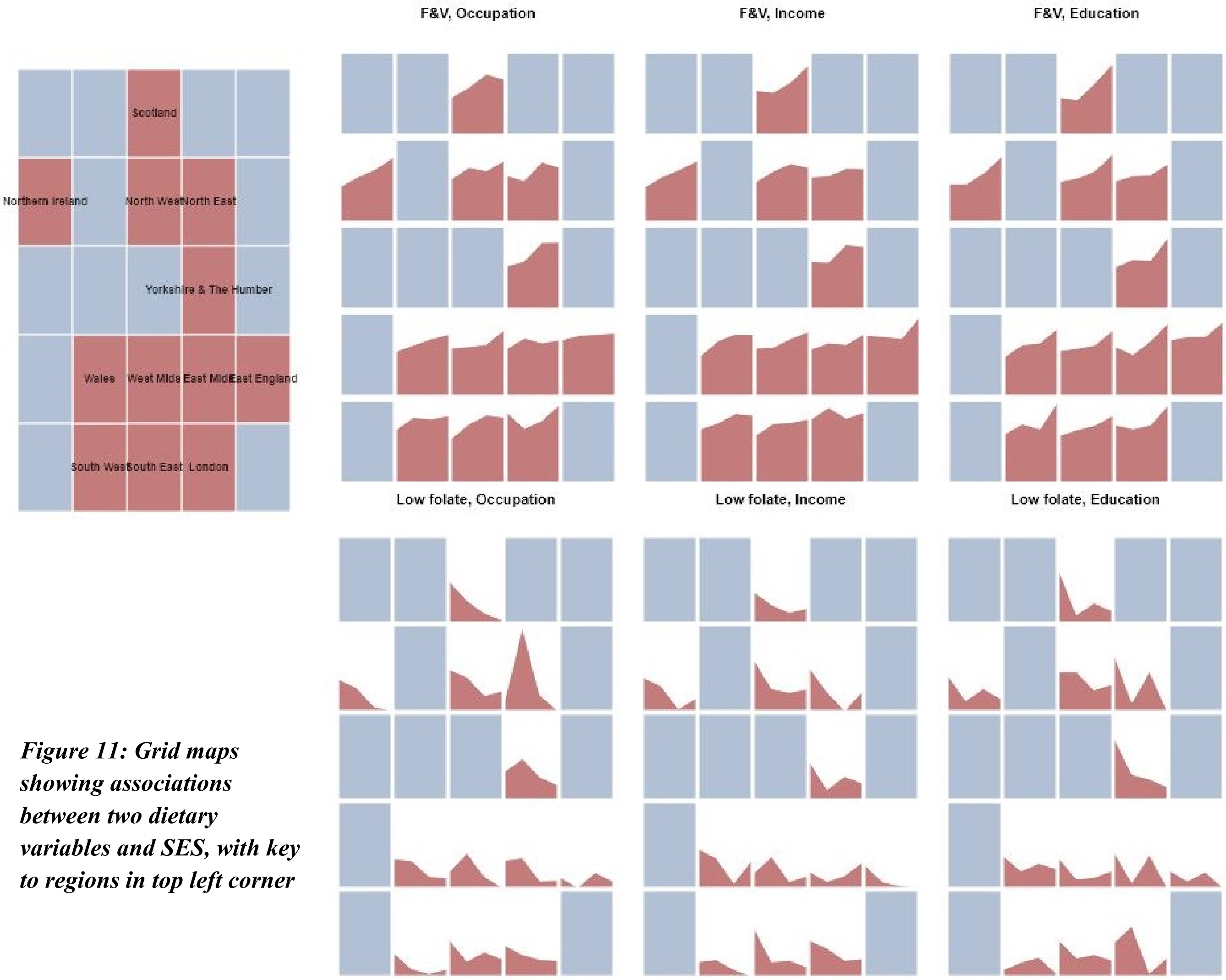


Figure 11: Grid maps showing associations between two dietary variables and SES, with key to regions in top left corner

4.1.4.3 Geographical variation in inequality adjusted for demographic characteristics

Table A21 reports findings from regression modelling, with output for the six key dietary variables from models with Region alone, Region with each SES variable, and Region with each SES variable and the three demographic variables. The analysis indicates that the demographic variables were not confounding the relationship between dietary variables and Region or SES except for Red meat (the significance of Region decreased and of SES increased when the demographic variables were included) and Alcohol (the significance of Region increased).

<u>Summary of findings from Section 4.1.4 about geographical variation in dietary inequalities</u>

- While raw dietary variables showed considerable geographic variation, geographical variation in associations between diet and SES were not consistent either within or between variables.
- Of the six dietary variables tested, there was evidence of interaction between region and SES only for the interaction Income * Region, for Folate intake <LRNI. The general trend for folate deficiency to fall with increasing income did not occur in the East Midlands and West Midlands.

4.1.5 Temporal variation in dietary inequalities

The findings in this section address the third study objective: "*Examine time trends in dietary inequalities in the UK*". First we examine simple temporal variation in dietary variables.

4.1.5.1 Patterns of temporal variation in dietary variables - unadjusted

Figure 12 shows Sparklines for the 21 dietary variables illustrating the temporal trends across the nine survey years, together with values of R squared and p from regression modelling. Using the modified probability threshold value of 0.0024 (0.05/21 hypothesis tests), 9 of the variables showed significant temporal variation. Intakes of Red meat, NMES, Sodium, Folate and Vitamin A all fell between 2008-16; the proportions of the population consuming Alcohol and non-low-calorie soft drinks also fell, and the proportions of the population with inadequate dietary intakes of Folate and Vitamin A increased.

4.1.5.2 Temporal variation in associations between dietary variables and SES status

Now directly addressing the third research objective, Figure 13 plots RI values for Fruit and vegetable intake and % population with Folate < LRNI for each year (chosen because the plots indicate a temporal trend for the RI values to increase), and Figure A30 plots the RI values for the remaining four key dietary variables. For reference, all values are provided in Table A23.

The line plots of Figure A31 enable comparison between the SES measures in temporal change of dietary variables disaggregated by SES category (from left to right the SES variables are occupation, income, and education), and Figure A32 shows temporal changes in RI values. Although Figure 12 shows there was no temporal trend in mean absolute intake of Fruit and Veg, Figure A31 (top left) shows that for education the gap between categories 1 and 4 increases, resulting in an increase in educational inequality evident in Figure A32, and confirmed by regression modelling summarised in Table 24.

Further, Figure 12 had shown that Red meat intake, NMES and proportion consuming alcohol decreased significantly over the nine years included in the survey. Figure A31 shows these trends affect all four SES categories which is why Figure A32 does not show decreases in inequality for these variables. In contrast, Figure 12 shows the proportion with Folate <LRNI increased significantly over time, and Figure A31 shows this is due to increases only in the categories 1 and 2 (low SES). One can discern a weak trend in Figure A32 for inequality in % population with Folate <LRNI to increase over time (base right).[16] The trend was also weakly discernible in the right RI plot in Figure 13, however it was not confirmed by the regression analysis summarised in Table 24.

[16] The erratic nature of the lines indicates sample size is not sufficient for this kind of sub-group analysis.

	Continuous variables				Binary Variables			
			Rsq	*p*			Rsq	*p*
Fruit and Veg	Intake		0.000		Achieve 5 a day		0.000	
Red meat	Intake		0.009	↓ *<0.001*	-			
Oily fish	-				Consume		0.001	
Soft drinks	-				Consume		0.004	↓ *<0.001*
Total energy	Intake		0.001		-			
Fat %Energy	Intake as %		0.000		-			
SFA %Energy	Intake as %		0.000		-			
NMES %Energy	Intake as %		0.006	↓ *<0.001*	-			
Alcohol	-				Consume		0.006	↓ *<0.001*
Fibre	Intake		0.001		-			
Sodium	Intake		0.018	↓ *<0.001*	-			
Iron	Intake		0.003	↓ *0.003*	Below LRNI		0.002	↑ *0.006*
Calcium	Intake		0.000		Below LRNI		0.001	↑ *0.029*
Folate	Intake		0.011	↓ *<0.001*	Below LRNI		0.005	↑ *<0.001*
VitaminA	Intake		0.008	↓ *<0.001*	Below LRNI		0.003	↑ *<0.001*
VitaminD	Intake		0.000					

Figure 12: Sparklines showing associations between 21 dietary variables and Survey year

Notes: For continuous dietary variables Y axes represent median values; for binary dietary variables Y axes represent proportion of sample; X axes represent survey year; p values below the threshold value of 0.0024 are highlighted.

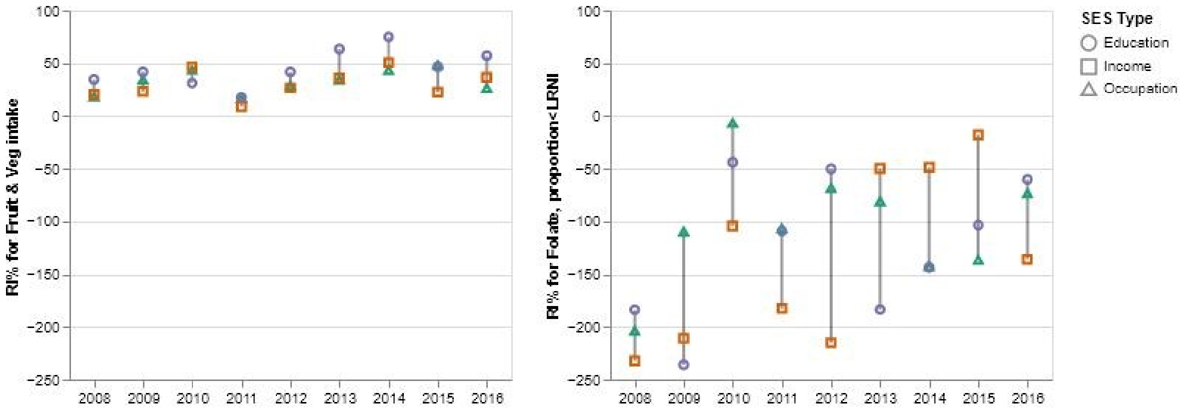


Figure 13: Ranged dotplots of relative inequality in Fruit and Vegetable intake, and proportion sample with Folate < LRNI, by year

Note: A positive relative difference shows higher intake of the food or nutrient in the highest SES category

Regression analysis with interactions

For testing for existence of interactions between year and SES, the critical probability value was adjusted using the Bonferroni correction. One of the 18 interactions tested was significant at the 0.05 level (Table A24) and this interaction of marginal significance was for Education x Fruit and vegetable intake. However, using the modified probability threshold value of 0.003 (0.05/18 hypothesis tests), none were statistically significant.

4.1.5.3 Temporal trends in inequality adjusted for demographic characteristics

Table A25 reports findings from regression modelling for the six key dietary variables, with output from models with Year alone, Year with each SES variable, and finally Year with each SES variable and the three demographic variables. The analysis indicates that the demographic variables were not confounding the relationship between dietary variables and Survey Year or SES, except for Red meat intake (the significance of Occupation and Income increased when the demographic variables were included) and for NMES as %food energy (the significance of income increased).

Summary of findings from Section 4.1.5 about temporal variation in dietary inequalities

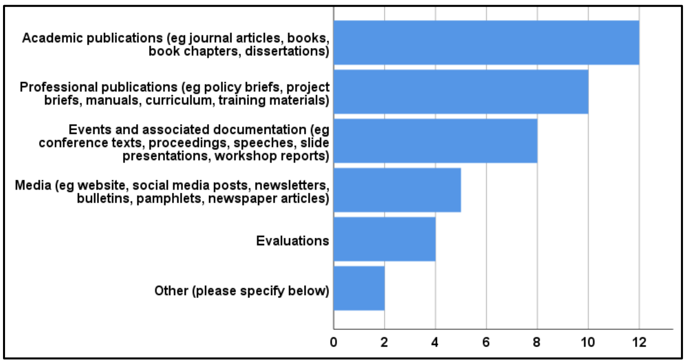
- Several dietary variables showed significant temporal changes in their raw values, including reductions in intake of red meat, sodium, NMES, folate and Vitamin A.
- However, temporal variation in associations between diet and SES was ascertained only for one of the six dietary variables tested, that of fruit and vegetable intake. There was a weak trend for educational inequality to increase between 2008-2016.

4.2 Visualization design study

This section presents the findings derived from the visualization evaluation.

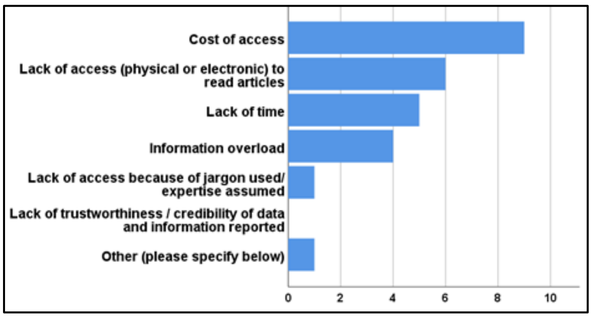
4.2.1 Participant background and experience

Below the responses about participant background and experience from section 2 of the first survey (Appendix J) are summarised. 13 nutritionists participated in the data collection, of which 12 completed this introductory section. All the free text comments are provided in Appendix L.



The participants reported using the full range of sources of research evidence specified, with all using academic publications (Figure 14 and Table A26). "Other" sources reported were "*Professional organisation websites - e.g. BDA, Nutrition Society*" and "*Food Industry reports and Market Research reports*".

Figure 14: Bar chart of responses to "Where do you obtain research evidence in your day-to-day work?"



All except one respondent (shown as the category "Other" in Figure 15) reported at least one barrier to increased use of academic journals. 75% reported cost as a barrier to increased use of academic journals, with lack of access and time as the next most common barriers.

Figure 15: Bar chart of responses to "What are the main barriers to your increased use of academic journals as a source of research evidence?"

Half the participants had reported their own findings in academic journals (bar chart not provided; see Table A27 for comments). At least half had made presentations in the workplace or at conferences (Figure 16 and Table A28). "Other" situations were to "*Provide evidence-based nutrition information and advice to families and children*" and *"To groups at other academic institutions"*

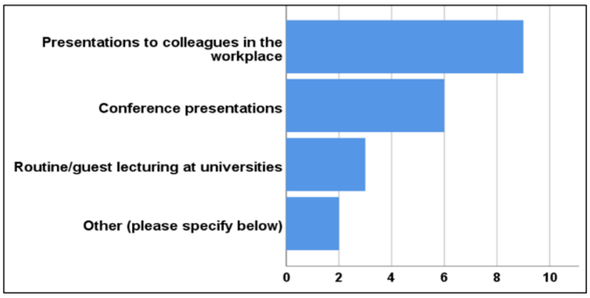


Figure 16: Bar chart of responses to "Do/have you deliver(ed) formal spoken communication(s) to an audience presenting findings from your own or others' research?"

4.2.2 Knowledge transfer with respect to integration of interactivity into visualizations

This section summarises the results from section 3 of the first, and section 2 of the second surveys. Figures 17a and 17b illustrate the responses to the evaluation questions for the maps (left columns), line chart (central columns) and bar charts (right columns). The data values themselves are provided in Table A29. Each row in these figures has charts for the equivalent questions on each visualization. The y axes range from lowest value 1 for "Strongly agree" to 4 for "Strongly disagree."[17] For each bar chart, the left-hand blue bars relate to the static versions and right-hand red bars to the interactive. Figure 18 summarises the differences between the responses to the static and interactive charts, derived from calculating the difference in mean ranks of the compiled responses (Section 3.2.7), and created to help identify differences between responses to the static and interactive versions.

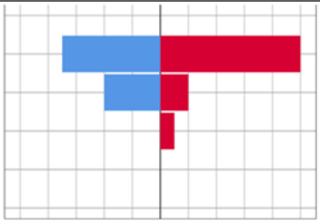
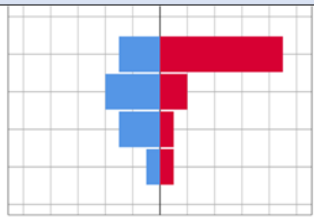
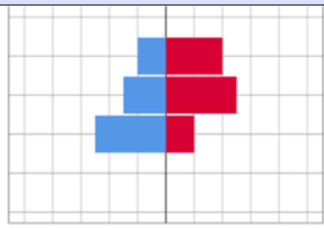
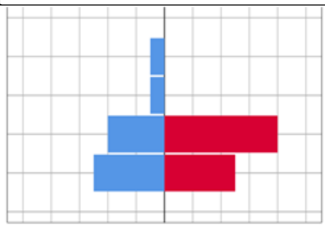
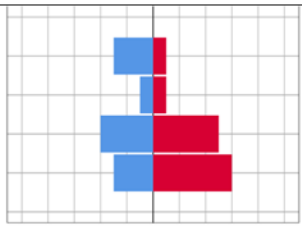
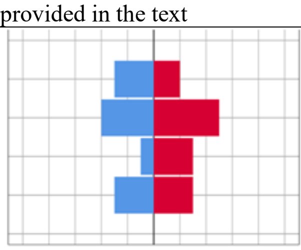
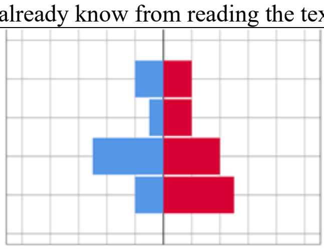
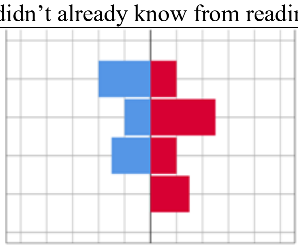
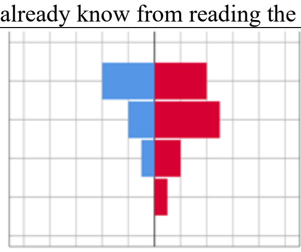
<table>
<tr><th></th><th>MAPS</th><th>LINE CHART</th><th>BAR CHARTS</th></tr>
<tr><td>1</td><td>
The maps are superfluous because the important information was already provided in the text</td><td>
The line chart is superfluous because the important information was already provided in the text</td><td>
The bar charts are superfluous because the important information was already provided in the text</td></tr>
<tr><td>2</td><td>
I learnt something about these dietary variables from the maps that I didn't already know from reading the text</td><td>
I learnt something about these dietary variables from the line chart that I didn't already know from reading the text</td><td>
I learnt something about these dietary variables from the bar charts that I didn't already know from reading the text</td></tr>
<tr><td>3</td><td>
I learnt something about these dietary variables from the maps that surprised me (eg it contradicted what I already knew)</td><td>
I learnt something about these dietary variables from the line chart that surprised me (eg it contradicted what I already knew)</td><td>
I learnt something about these dietary variables from the bar charts that surprised me (eg it contradicted what I already knew)</td></tr>
</table>

Figure 17a: Divergent bar charts of responses to visualization evaluation questions about knowledge transfer
Note: X axis is number of respondents and vertical grid indicates 2; Left blue bars relate to static charts; Right red bars relate to interactive charts; Y axis is level of agreement, 1 = Strongly agree and 4 = Strongly disagree.

[17] The option 5 "Don't know" was offered but no respondents used it, instead there were a few missing values.

Figure 17a shows responses to the first three questions. For the first question, higher y values indicate a favourable view of the charts' value, while for the second and third, higher y values are unfavourable with respect to having learnt something new or surprising. Figure 17a shows:

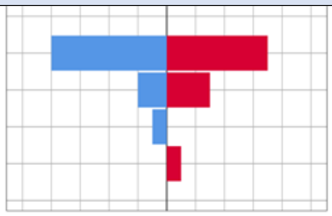
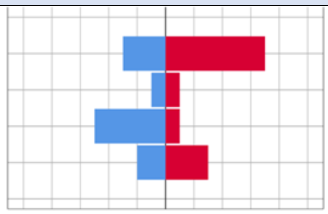
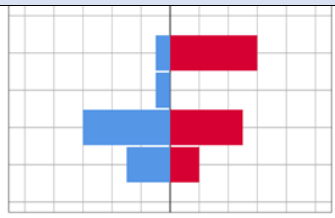
- For the maps and line charts, most respondents disagreed that the charts were superfluous, and agreed they learnt something from them;
- Compared to the maps and line chart, the bar charts had fewer high y values for the first question, and fewer low y values for the second and third, indicating a less favourable view of bar charts with respect to value for transferring knowledge.

Figure 18 illustrates the differences in mean ranks of responses about the static and interactive charts. The top three bars correspond to the three questions illustrated in Figure 17a, and show that:

- For all three forms of charts, the responses were in favour of the interactive charts over the static;
- The differences were most noteworthy for the line charts (confirmed by the values of Z and p in Table A29).

Responses to the fourth and fifth evaluation questions are shown in Figure 17b. Higher y values show stronger disagreement with the statements and therefore a more favourable view. Figure 17b shows:

- Greater difficulties were experienced in the interpretation of bar charts compared to the maps and line chart (in the right-hand column there are notably more responses for the lower y values);
- For the maps (left-hand column) there was little difference between respondents' views about barriers to interpretation of the static maps compared to those for the interactive maps;
- With respect to poor design as a barrier to interpretation (first row), for the line chart and bar charts (second and third columns), there was a bigger difference in responses to static and interactive versions than for maps (confirmed by the values of Z and p in Table A29). Responses were more favourable towards the interactive versions than the static;
- With respect to lack of expertise as a barrier to interpretation (second row), responses for the bar chart show a clearer preference for the interactive over static versions, compared to responses for the maps and line chart (confirmed by the largest Z value being for the bar charts in Table A29).

	MAPS	LINE CHART	BAR CHARTS
4	 I find the maps difficult to interpret because they are badly designed	 I find the line chart difficult to interpret because it is badly designed	 I find the bar charts difficult to interpret because they are badly designed

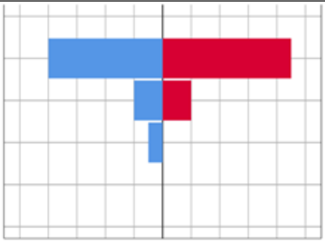
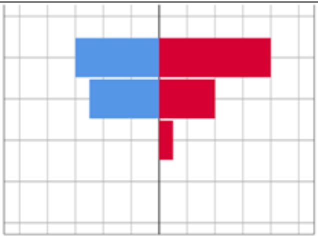
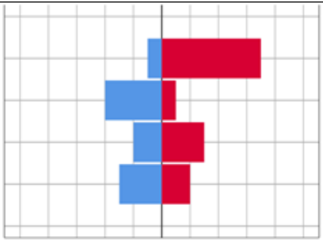
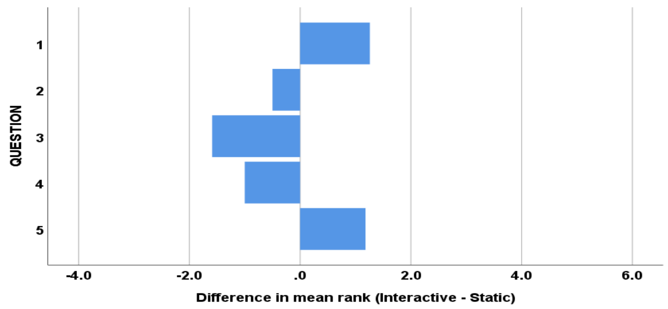
5	 I find the maps difficult to interpret because I don't have sufficient expertise	 I find the line chart difficult to interpret because I don't have sufficient expertise	 I find the bar charts difficult to interpret because I don't have sufficient expertise

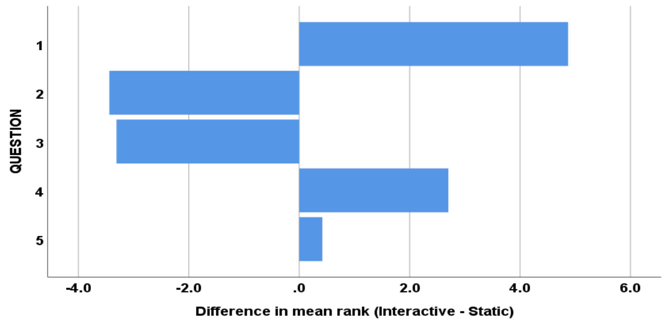
Figure 17b: Divergent bar charts of responses to evaluation questions about barriers to interpretation

Note: X axis is number of respondents and vertical grid indicates 2; Left blue bars relate to static charts; Right red bars relate to interactive charts; Y axis is level of agreement, 1 = Strongly agree and 4 = Strongly disagree.

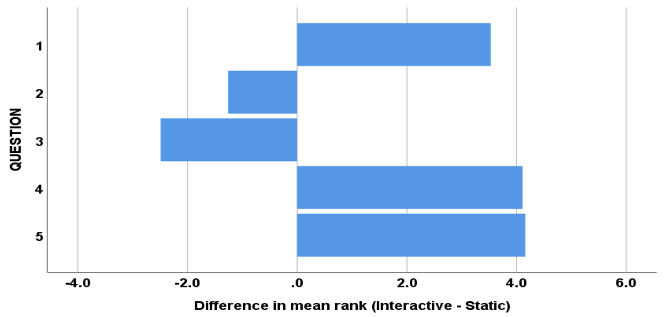
MAPS



LINE CHART



BAR CHARTS



The lower two bars of Figure 18 correspond to the two questions illustrated in Figure 17b, and here the differences in mean ranks of responses to the static and interactive charts confirm the impressions described above, that is

- For the maps, there was minimal difference between static and interactive charts with respect to expertise and design being perceived as barriers to interpretation;
- For the line charts, the design of the interactive charts was perceived as less of a barrier than for the static;
- For the bar charts, both expertise and design were perceived as lesser barriers to the interpretation of the interactive versions than for the static.

Figure 18: Bar charts of differences in mean ranks of responses to static and interactive charts

Note: A positive value favours the interactive format for questions 1, 4 and 5, and favours the static format for questions 2 and 3.

The respondents wrote several free text comments about the charts during allocation of the scores, provided in full in Table A30, and integrated into Table 16, except with respect to expertise lacking for interpretation, for which responses were only provided for the bar charts. These were "*I couldn't tell which groups had higher intakes. What was intercept? What was the Y axis?*" and expertise lacking was specified as: *"Coefficient knowledge"* and "*Knowledge of coefficients and accronyms (e.g SES)"*.

Table A31 provides in full the free text comments written after the scores had been allocated. Table 16 provides the comments categorised by major theme, which were "Value of charts over text"; "Design"; "Preference for static or interactive", and "Additional value of interactivity". Comments independent of static v. interactive format are listed first.

<table>
<tr><th>STATIC</th><th>INTERACTIVE</th></tr>
<tr><td colspan="2">MAPS</td></tr>
<tr><td colspan="2"><u>Value of charts over text;</u>
• "It's easier to see the comparisons within England and in the different nations over time"
• "The maps are useful as provide more detail than the text on regional variation and trends"
• "The maps are very useful, they provide more information and allow faster interpretation of the data"
<u>General design:</u>
• "Would be useful to also have recommended and average amounts on the key.
"Would be better to have specific ranges with specific shades rather than gradual shade difference on a continuous variable in legend.
"No indication of confidence intervals of differences between regions - suspect overlapping due to small sample numbers - therefore misleading.
"Variables/metrics displayed are not consistent i.e. mean daily grams v. % of population - the maps lend themselves better to % population than the former. Alcohol metric is unclear in its definition."</td></tr>
<tr><td><u>Design</u>: "Maps too small and not easy to specify which specific area of the country they relate to"

<u>Preference:</u> "I preferred this format than the slider and would find it easier to cut and paste selected years in to powerpoint slides for a presentation"</td><td><u>Design</u>: "Would be useful to zoom into areas"

<u>Additional value of interactivity:</u> "Useful that areas and data amounts appear when hovering over"</td></tr>
<tr><td colspan="2">LINE CHART</td></tr>
<tr><td colspan="2"><u>Value of charts over text;</u>
• "The chart is very difficult to interpret - the text could be supported visually but not like this it is unhelpful"
• "Useful to see and compare the figures of the meat intakes at different points"
<u>General design:</u>
• "Clustering and the close proximity of the data points makes the chart hard to interpret"
• "Information is not clear as there are too many categories (lines) to look at"
• "Graph was too busy. It wasn't clear which colour was which region"
• "No indication of confidence intervals - which means data could be misleading"
• "It wasn't clear if the X axis was continuous variable or not"
• "A line chart is not the appropriate method for displaying these data"
• "Discrete data points - feels like they shouldn't be connected"</td></tr>
<tr><td><u>Design:</u>
• "Hard to interpret - too many lines and discrete categories"
• "A lot of lines, and not easy to distinguish one line from the other"</td><td><u>Preference:</u> "The interactive visualisations make a big difference to interpreting the data in a meaningful way and don't require much knowledge or expertise"

<u>Additional value of interactivity</u>: "There's lots of information in the chart, but you can focus in on</td></tr>
</table>

Preference: "Including colours on the legend helped identify which region a line related to more easily than the interactive version"	narrow aspects of the data to ease interpretation and understanding of the data"
BAR CHARTS	
General design: • "Far too many categories of data to interpret in one graph" • "Overlapping error bars and colours make the graphs difficult to clearly follow/understand" • "Too small with too much detail on them. I would have found it easier to interpret if the y axis was portions of F&V or meat rather than a regression" • "3 headings income education occupation are confusing. Differences between 3 plots not clear" • "Something must be missing on axes or explanation as can't make sense of them stand alone"	
Design: "Difficult to read, and small on one screen if on a small laptop, and for those with weakened eyesight"	

Table 16: Free text comments about the survey visualizations, categorised by theme

The main noteworthy feature of Table 16 is the number of misgivings expressed about the charts' design – for example with respect to size (too small), the number of categories (too many), confidence intervals (not plotted), and explanation and labelling (insufficient). It is interesting to note that for the line and bar charts, disagreement with the statement *"I find the charts difficult to interpret because they are badly designed*" was higher for the interactive versions (Question 4 in Figure 19).

Summary of findings from Section 4.2.2 about knowledge transfer

Value of charts over text: The charts were overwhelmingly perceived as not being superfluous to the text. Most respondents reported having learnt something new and/or surprising from the charts.

Value of different chart formats: The maps were perceived as the most effective with respect to having learnt something new or surprising, followed by line charts, and the bar charts least effective.

Value of static v. interactive chart versions: The interactive versions were perceived to be more effective with respect to knowledge transfer than the static versions. The extent of difference in effectiveness depended on chart format, in that the benefit of interactivity was greatest for the line charts, and lowest for the maps.

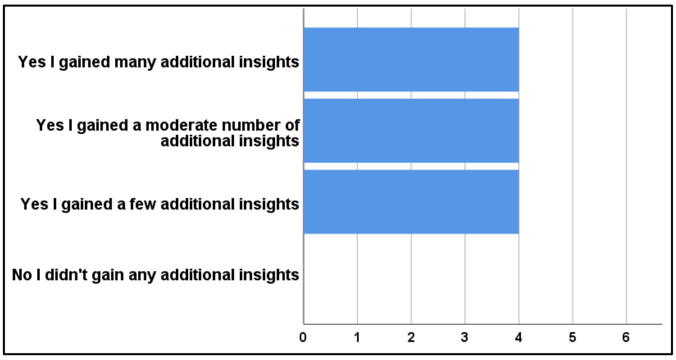
Interpretation barriers with respect to chart format: Barriers to interpretation were more often reported for bar charts, compared to maps and line charts.

Interpretation barriers with respect to static v. interactive versions: Interpretation barriers were more often reported for the interactive compared to static versions, and the difference depended on format -

- For the maps, there was minimal difference between respondents' views about barriers to interpretation of the static versions, compared to those for the interactive versions;
- With respect to design, for both the line chart and bar charts the responses were more favourable towards the interactive versions over the static;
- With respect to lack of expertise, only the bar charts showed a clear favouring of the interactive versions over the static.

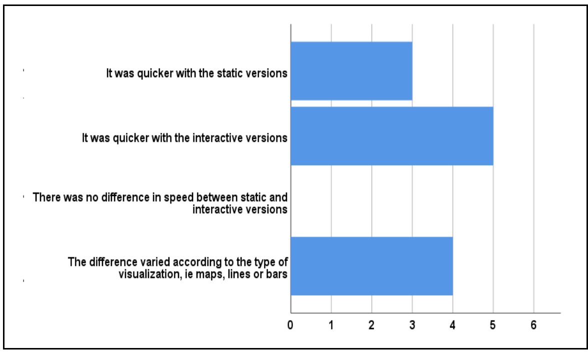
4.2.3 Cognitive load with respect to integration of interactivity into visualizations

This section summarises the results from section 3 of the second survey (Appendix K), for which responses were provided by twelve participants.



All participants reported they had gained additional insights from the charts (Figure 19).

Figure 18:Bar chart showing responses to question "Did you gain insights from the visualisations that were additional to those gained from the information provided in the one-page text?"



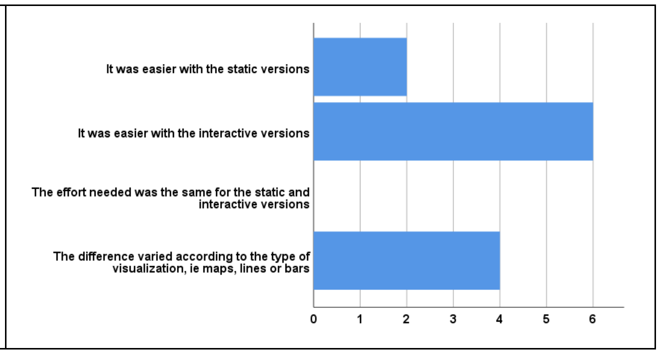
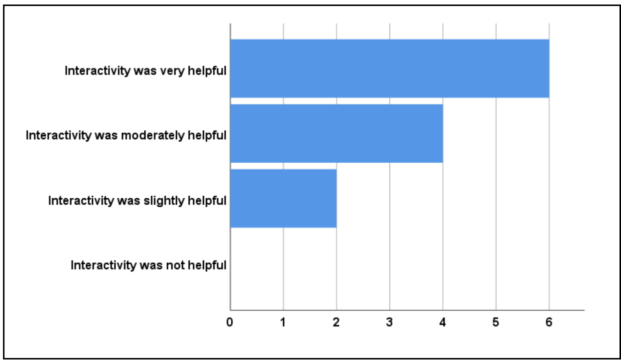


Figure 19 (left): Bar chart of responses to question "Did you gain the additional insights more quickly from the static or interactive versions?"
Figure 20(right): Bar chart of responses to question "Was it easier to gain the additional insights from the static or interactive versions?"

Figures 20 and 21 show that more participants perceived the speed and ease of obtaining insights were greater from interactive versions of the charts than static, with a slightly greater difference for ease. A third of participants reported that ease and speed depended on the format.



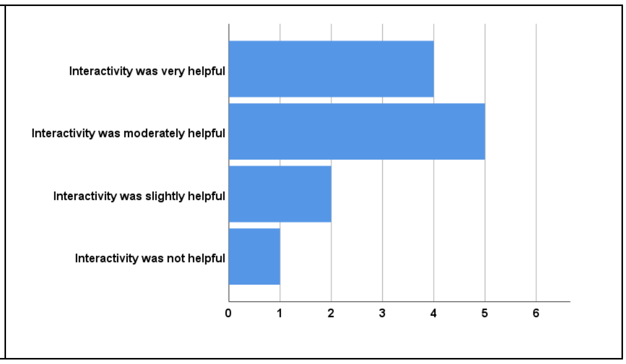


Figure 21 (left): Barchart of responses to question "To what extent was the facility to interact with the visualizations helpful in increasing your understanding of the researcher's findings described in the one-page text?"

Figure 22 (right): Barchart of responses to question "To what extent was the facility to interact with the visualizations helpful in communicating and supporting the researcher's findings described in the one page text?"

Figures 22 and 23 show that all participants felt interactivity was to some degree helpful in increasing understanding of the findings, and in communicating and supporting the researcher's findings. The support for understanding (left figure) was perceived to be greater than for communication.

Figure 24 shows that all respondents found the visualizations engaging, and of those who perceived a difference in engagement between static and interactive versions (three-quarters of the respondents), all reported in favour of the interactive charts.

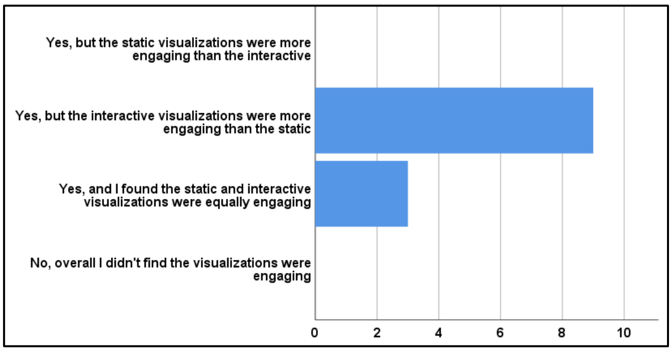


Figure 23: Barchart of responses to question "Did the visualizations engage you?"

The free text comments about gaining insight and comparison between versions are provided in full in Table A32 and demonstrate varied reactions of the respondents. One explained why they were in favour of interactive versions:

"It required far more effort to interpret the static data. I spent much longer examining the interactive data and found it more interesting. I became quickly tired and bored of the static data."

while another expressed surprise that they preferred the static versions because they would be more useful in their own presentations:

"If you had just asked me do I prefer interactive or static data without showing me examples I would have said interactive! But actually the static was more useful to me and how I would go on to use and present the data."

Summary of findings from Section 4.2.3 about cognitive load

- More respondents felt it was easier to gain insights from the interactive versions of the charts, than from the static versions, and likewise more found it quicker;
- One third of respondents felt that the difference between interactive and static charts with respect to speed and ease of gaining insights depended on the form of the chart (map, line or bar);
- All respondents found interactivity helpful in increasing their understanding of the findings, and in communicating and supporting the researcher's findings. The support for understanding was perceived to be greater than for communication;
- All respondents found the charts engaging, and all those who perceived a difference in engagement between versions (three-quarters of respondents) found the interactive versions more engaging than the static.

5 DISCUSSION

Summary results are repeated below in italics, and then discussed in the perspective of prior work.

5.1 Discussion of secondary data analysis

This first component of the research project explores patterns of inequalities in diet in the UK, to increase understanding of a critical factor underlying the unequal distribution of health problems.

5.1.1 Objective 1 - Patterns of variation in the UK diet with socio-economic status

Links between dietary variables and socio-economic status (SES) were identified by examining trends against the three SES measures (occupation, income, and education), and using the derived variable of "Relative Inequality".

- *Associations between dietary variables and SES for the most part followed expected trends, whereby dietary variables that were protective for health (fruit and vegetables' intake, fibre intake, proportion consuming oily fish and intake of the five micronutrient variables considered (iron, calcium, folate, Vitamin A and Vitamin D)) had positive associations with SES, while dietary variables that were risk factors (micronutrient intakes <LRNI for iron, calcium, folate and Vitamin A, and intake of NMES as % food energy) had negative associations.*

The government report of findings from the same full dataset considered only the SES measure of income for survey years 5-9 only (Bates et al., 2019). Results from analysis of the full 9 years' data reported here are consistent with the previous findings, and provide more detail, showing that associations of some dietary variables with SES differ according to the SES measure used.

- *For red meat intake, the three SES variables had independent and opposite effects, in that intake increased with income and decreased with occupation and education.*

The recent analysis found for adults aged 19-64 years there was a significant positive association between intake of red and processed meat and equivalised income, but for adults aged ≥65years there was no association (only survey years 5–9 included) (Bates et al., 2019). In an earlier study using data from years 1-3, there was a trend for the highest SES groups to have consumed less red and processed meat across categories of all three SES measures (Maguire and Monsivais, 2015). The current study found this trend for lower meat intake in the higher SES groups applied to occupation and education, but the association with income was positive. All associations were relatively weak, and so are perhaps due to chance, but worth following up when more data are available from the NDNS rolling programme.

- *Within-variable associations with SES measures differed between the dietary variables - for Vitamin D and Iron intake<LRNI, association with income was notably stronger than with occupation or education, while for fibre and the two variables for fruit and vegetables, correlation with education was stronger than with occupation or income.*

These findings confirm the value of using more than one measure of SES, since their effects can be independent (Alkerwi et al., 2015, Si Hassen et al., 2016). This is unsurprising if one considers education and occupation as markers of social relationships and command over life-long skills, while income as more indicative of a current standard of living (Duncan et al., 2002). Education reflects the ability to understand and act in response to health promoting messages (Patel et al., 2020) so the strong association of fibre and fruit and vegetable intake with education could reflect effectiveness of the “Five-a-day” campaign (Rayner, 1998). The inverse association between proportion of population with low iron intake and income is consistent with the positive association between total red meat intake and income, since red meat is an important dietary source of iron.

Regression modelling revealed that for the six dietary variables tested:

- *There were strong associations between dietary variables and demographic variables, for example fruit and vegetable intake was lower in the younger and white ethnicity adults compared to the older and non-white ethnicity groups, and red meat intake was lower in women and adults of non-white ethnicity compared to men and adults of white ethnicity.*

With respect to gender, the findings are consistent with those of the earlier study of dietary inequality using NDNS data (Maguire and Monsivais, 2015). With respect to ethnic group, since all non-white ethnic groups were aggregated into one group for analysis, generalisability of findings is limited. The UK has a rich mix of cultures and culturally diverse communities, with the largest minority groups being South Asian, Black African-Caribbean and Chinese, which have different dietary traditions and associated patterns of chronic disease (Leung and Stanner, 2011).

- *There was little evidence of demographic variables confounding the dietary v. SES associations*
- *There was some evidence of interactions between demographic and SES variables. Ethnicity interacted with Occupation for Fruit and vegetable intake, such that the trend to increase intake with increasing income was weaker for the non-white group. Also, Age-group interacted with Income for Fruit and vegetable intake, in that the trend for increased intake with increasing income was weaker for the older groups. For Alcohol, Gender interacted with Education such that the trend for the proportion of sample consuming alcohol to increase across increased education level was greater for women than men.*

These interactions are not discussed due to limitations on reporting space. To help determine if they are due to chance or are real effects, it would be valuable to check if these interactions still exist when more NDNS data become available.

5.1.2 Objective 2 - Extent to which the associations between diet and SES differ between regions

Geographical variation in patterns of associations between dietary variables and SES were examined using the derived variable of "Relative Inequality" (RI), and grid maps of dietary variables plotted against the three SES measures.

- *The raw dietary variables showed considerable geographic variation.*

The study's results broadly support earlier findings that populations of Scotland and Northern Ireland consume a poorer diet than that of England, with respect to higher sodium consumption and lower fruit and vegetable consumption (Scarborough et al., 2011). However, for other nutrients the findings are more nuanced, and considerable variation between the regions of England was also found to exist. As a broad generalisation, North-East and North-West England were more similar to Scotland and Northern Ireland than to the southern English regions, and inclusion of SES and demographic variables in regression models did not dent regional disparities (findings not presented in Chapter 4), demonstrating the influence of local cultural norms on dietary decisions in the UK.

The South-West region appears to have the healthiest diet, with high intakes of fruit & vegetables, iron, folate, calcium, Vitamin A and Vitamin D, and no extreme values of RI.

- *Geographical variation in patterns of associations between diet and SES were not consistent within or between variables. Of the six dietary variables tested, there was evidence of interaction between region and SES only for the interaction Income * Region, for Folate intake <LRNI. The general trend for folate deficiency to fall with increasing income did not occur in the East Midlands and West Midlands.*

Maternal deficiency in folate has been linked with increased risk of neural tube defects such as spina bifida, and so folic acid supplements are recommended before pregnancy. The fact that two-thirds of UK women do not take these supplements (Bestwick et al., 2014) makes it particularly important to identify which groups of women have low folate intake. The choropleth map of Figure A28 shows the regions with highest proportions of the population having folate intake <LRNI are Yorkshire, North-East England and North-West England. Nonetheless, the finding relating to an interaction with income - that implies those with higher income in the Midlands are still at risk of folate deficiency - is of potential value with respect to nutrition intervention design in the Midlands.

5.1.3 Objective 3 - Time trends in dietary inequalities

Temporal variation in patterns of associations between dietary variables and SES were examined using the derived variable of "Relative Inequality", and line graphs of dietary variables plotted against the three SES measures.

- *Several dietary variables showed considerable temporal changes towards a healthier diet (intakes of red meat, NMES and sodium, as did the proportions of population consuming alcohol and non-low calorie soft drinks), while others were less beneficial (intakes of folate and Vitamin A reduced).*

Not surprisingly, results from the secondary analysis reported here are consistent with the findings in the government report on the same dataset (Bates et al., 2019).

- *Temporal variation in associations between diet and SES was ascertained only for one of the six dietary variables tested - There was a weak trend for educational inequality to increase between 2008-2016.*

Consistent with findings from data collected 2001-09 in Scotland (Barton et al., 2015), and between 2001-11 in Great Britain (Ji and Cappuccio, 2014), this study revealed minimal indication that SES dietary inequalities reduced over the period 2008-16 in the UK. The single (weak) interaction between year and SES discerned was for fruit and vegetable intake, and pertained to an increase in inequality. This is consistent with recent findings from analysis of eight years' NDNS data collected from 2008 (one year fewer than the current study). All graphs relating to occupational or income inequality show no changes over time, while some graphs relating to educational inequality do show changes - educational inequality in vegetable intake widens, while it narrows in red and processed meat intake, and low-fat dietary products[18] (Patel et al., 2020).

Thus overall, this study did not find evidence for changes in diet-related inequality for the nine years after 2008 in the UK. This is consistent with the findings of Patel and colleagues, who reported a general improvement in the UK diet but no changes in its inequality between 2008-2016 (Patel et al., 2020), and with non-significant changes in relative inequality in all-cause mortality rates in England, Wales and Scotland between 1990-94, and 2005-9 (Mackenbach et al., 2016). However the findings are not consistent with evidence of widening health inequalities in England since 2010 (Marmot et al., 2020), and in any case the patterns and time-trends identified by this study and others are likely to change substantially, given evidence that certain groups have been disproportionately affected by the current COVID 19 pandemic (Health Foundation, 2020).

[18] These differences are not statistically significant, and the authors do not draw attention to them.

5.1.4 Validity, generalisability and implications of results from the secondary data analysis

5.1.4.1 Limitations of the original dataset

The NDNS is the only source of high quality nationally representative data on quantities and types of foods consumed by individuals in the UK. Estimates of population nutrient intake derived from them are used to develop policy and monitor progress on diet and nutrition objectives, and assess exposure to chemicals in food (NatCen Social Research, 2019b). Give these important roles of the data, the UK government has a vested interest in maximising their quality and representativeness. But the data still have the following limitations:

- The survey non-response rate of around 47% (Table 5) is a potential source of bias since those who do not participate may be systematically different from those who do. As in most nationwide population surveys, the most deprived groups may be under-represented, such as the unemployed, homeless, and migrants with poor English language skills. Weights were provided for the dataset in order to reduce the effect of non-response bias (Section 3.1.4.3) but some bias may still remain;
- The dietary data are self-reported and subject to both random error and systematic error due to recall bias, and misreporting. Misreporting is probably mainly due to social desirability bias, since respondents are likely to be conscious of desirable eating habits (Vlismas et al., 2009). Dietary diaries are particularly prone to under-reporting of energy intakes (Rennie et al., 2007), and analysis of NDNS data from a sub-sample of 197 adults found that for energy, the average reporting error across all adults was 32% (range 16.5% (over-reported) to 72.2% (under-reported)) (Bailey, 2018). The estimates of SES inequality in this study may be over- or under-estimates if the degree of recall bias and misreporting was socio-economically patterned.
- The SES measure of income is less trustworthy than the education or occupation measures, since personal income is a sensitive issue and individuals may be unwilling to disclose this information accurately (Galobardes et al., 2007);
- The level of geographical granularity is low, for example Scotland, Wales and Northern Ireland are single regions, while England is divided into nine;
- Sample size is not sufficient for certain sub-group estimates, so that one cannot reliably ascertain temporal trends in geographical variation and geographic variation in temporal trends;
- The data are cross-sectional rather than longitudinal – this reduces the power to detect trends since they are derived from different individuals in each annual survey.

5.1.4.2 *Limitations of the secondary data analysis*

- A proportion of the original sample were excluded for analysis involving two of the newly created SES variables - adults from a household where the HRP was economically inactive were excluded from analysis using the SES measure Occupation, and if the HRP had Foreign or other qualifications, or who were still in full-time education, adults from that household were excluded from analysis using the SES measure Education.
- The calculated inequality values may give a misleading impression of true values since they were derived using just one measure of social class. Occupation, income and education are not perfectly correlated, since they reflect different pathways through which SES can independently influence diet (Galobardes et al., 2001), and so their effects are likely to be additive.[19]
- The researcher's training and experience is in null-hypothesis significance testing. While she has tried to modify her approach towards that of "New Statistics" (Calin-Jageman and Cumming, 2019), appraisal of the output from data analysis in this document is in places over-dependent on the binary interpretation of p-values.

5.1.4.3 *Quality of findings, and their implications*

The major strengths of the secondary data analysis are that it used NDNS data which are representative and of high quality,[20] and that the statistical analysis undertaken was systematic and robust. Given these strengths, and the fact that the limitations of the data and analysis described above are not so severe to bring the general validity of the findings into question, the overall conclusions based on trends and patterns identified in this study can be considered valid and trustworthy. Nonetheless, a small degree of error may exist in estimates of effects.

The findings are generalisable to the UK population between 2008-17 since they are based on a well-constructed representative sample. But the COVID 19 pandemic since March 2020 will likely mean the patterns at the time of writing (January 2021) differ from those described above.

There is not sufficient space here to discuss the public health significance of the inequalities identified in individual dietary elements. However, the overall finding that inequalities in diet have persisted - despite a general improvement in the UK diet having occurred over the last decade - points to the continuing need for effective interventions to promote and facilitate healthy dietary choices by individuals in the most disadvantaged groups.

[19] While in theory the effects may be synergistic, this was statistically tested, and no interactions were detected between the SES variables (results not shown).
[20] This is especially true with respect to the dietary data which were gathered using a four-day diary, rather than the more commonly used survey methods of food-frequency questionnaire or 24 hour dietary recall.

5.2 Discussion of design study

5.2.1 Survey chart design

Maps: Because of the large number of maps (36), the static version differed from interactive with respect to the size of each map (much smaller in static). Of the three forms of visualizations, the maps uniquely provided a finer breakdown of findings than was described in the background information sheet (Appendix I).[21] Surprisingly, only one respondent commented on the static maps' size, even though this is an obvious weakness of the visualization's design (due to constraints of the survey software). Another respondent mentioned there was "*..no indication of confidence intervals of differences between regions*", and one of the informal pilot testers also noted "*..there is substantial variation between years in the same region e.g. salt in Scotland over 3 consecutive years for example, probably because of sampling error..*"

While one can contend the design enables adequate communication of the overall trend for increase or decrease over time (by means of the green hue on average becoming darker or lighter), a measure of uncertainty would indeed ideally have been integrated by means of another colour, pixelation or glyphs (Lucchesi and Wikle, 2017). But this was not feasible with small multiples.

Line chart: Of the three visualizations, the line chart uniquely had the same size in static and interactive versions. There are three main aspects of the line chart design to consider as follows:

1) Number and similarity of categories: Several comments expressed users' concerns with respect to the "busy-ness" of the visualization, for example "*Clustering and the close proximity of the data points makes the chart hard to interpret"; "Information is not clear as there are too many categories (lines) to look at"* and *"Graph was too busy"*. These comments are well-founded – the researcher had recognised the number of categories was at the limit of discriminability (between six and twelve (Munzner and Maguire, 2014, p226)) so used shape as well as colour to encode fruit and vegetable intake (see Section 3.2.5). The close proximity of the lines justifies the rather derogatory term "Spaghetti graph" (Knaflic, 2013). The researcher justifies plotting too many categories in one graphic to explore the possible benefits of interactivity – she had anticipated users might welcome the ability to highlight one line and eradicate the effort needed to perceive different regions.

2) No indication of uncertainty: One respondent commented "*No indication of confidence intervals - which means data could be misleading*", and an informal pilot tester said "*The two outliers for*

[21] The sheet simply stated that all four variables had statistically significant associations with survey year across the UK, while the maps provided a geographical breakdown by displaying median values for each region.

London at each end of the scale suggest to me that the confidence intervals are wide". Again, these comments are very valid. The researcher justifies not having plotted confidence intervals because they would have made the visualization even more cluttered. She recognises the graphic would be acceptable for inclusion in an academic journal without indication of the estimates' uncertainty.

3) Suitability of form, given the non-continuous independent variable: There were several comments related to this concern, such as "*Discrete data points - feels like they shouldn't be connected";* and that from an informal pilot-tester ".. *the four groups on the x-axis are independent of each other so what is the justification for joining the points unless they are regression lines?"* The users' concerns can be countered by the fact that the occupation categories, while admittedly not on interval or ratio scale, are on ordinal scale.

Bar chart: Users needed familiarity with statistical techniques and language to interpret this visualization. Little explanation was provided to replicate the situation encountered by users of academic papers. The visualization was unique among the three used to not display raw values of the dietary variables themselves, as it was based on adjusted statistics derived from their analysis.

As for the line chart, some comments related to the cluttered nature, for example "*Far too many categories of data to interpret in one graph*"; "*Too small with too much detail on them*"; and *"Difficult to read..,"* while others related to the lack of explanation, for example "*Something must be missing on axes or explanation as can't make sense of them stand alone*". Again, the researcher recognises these comments as well-founded, and rationalises providing cluttered and scantily explained graphics to explore the possible benefits of interactivity.

Complexity of visualization: Since much feedback related to "density", it is useful to consider this issue. Several factors can contribute to perception of visualizations' "busy-ness" – not only chart "clutter" (Tufte, 2001, p168) and the number of variables and categories, but also attributes and arrangement of objects. Models link aspects of visualizations' complexity with higher cognitive load (Huang et al., 2009) but this association does not uniformly exist in reality. For example a study of predictors of message effectiveness for online health communication revealed increased design complexity to be positively associated with levels of perceived message comprehensibility, ease of use and usefulness (Lazard and Mackert, 2014). Apparently, rather than automatically considering a visualization's "density" as negative, the various aspects of this characteristic must be critically examined separately, and in relation to users' skills, knowledge and interests.

5.2.2 Objective 4 – Effect of integration of interactivity into data visualizations on knowledge transfer

- *Value of charts over text:* *The charts were overwhelmingly perceived as not being superfluous to the text. Most respondents reported having learnt something new and/or surprising.*

This findings is consistent with the literature summarised in Chapter 2, since making sense of text and tables of figures requires conscious thinking, while charts and graphs can be more readily perceived due to "pre-attentive visual processing", that is, the processing that automatically occurs prior to conscious awareness (Few, 2013).

- *Value of different chart formats:* *The maps were perceived as the most effective with respect to generation of insights, followed by the line charts, and the bar charts least effective.*

Saraiya and colleagues defined insight (Saraiya et al., 2005; Section 3.2.4)) and Yi and colleagues identified four independent processes, often used together, through which users gain insight from information visualization as "*Provide Overview*", "*Adjust*", "*Detect Pattern*", and "*Match Mental Model*" (Yi et al., 2008). Recognition of these processes aids interpretation of the study's results. Of the three visualizations, the maps more closely resemble every-day, rather than scientific, tools so the gap between most users' mental model of the data and the data themselves would be smallest. While a line chart is a familiar form of visualization, some users found the categories of occupation confusing and/or did not recognise their ordinal nature, so the processes "*detect pattern*" and "*match mental model*" were less successful than for the maps. Finally, the bar charts confused several respondents since the bars represented coefficients from regression modelling of adjusted dependent variables. Not surprisingly this third visualization was least successful in the last two insight generating processes.

Another likely reason the maps were reported as generating more insights is that they provided a finer breakdown of findings than the background information sheet (Section 5.2.1).

- *Value of static v. interactive chart versions:* *The interactive versions were perceived to be more effective with respect to knowledge transfer than the static versions. The extent of difference in effectiveness depended on visualization format, in that the benefit of interactivity was greatest for the line charts, and lowest for the maps.*

This result can be viewed using the same four process framework, focusing now on *"Adjust*". Through this process users explore a dataset by adjusting the level of abstraction and/or the range of selection, and the ability to change perspective on the dataset supports users to make sense of aspects of the data and test their own hypotheses (Yi et al., 2008). Of the three visualizations the line chart uniquely had the same form and size in both versions. With the interactive version, users selected a region to highlight and the other lines were still visible in the background. But for the interactive bar chart, comparison between findings for each SES variable was less easy, because values for the non-selected SES variable were visible only if their coefficient values were lower than those for the selected variable. For the maps, respondents needed to use memory more for comparisons between years, since only one map was visible at a time.

- *Interpretation barriers with respect to chart format:* *Barriers to interpretation were more often reported for bar charts, compared to maps and line charts.*

The bars charts plotted coefficients from regression modelling, so their interpretation needed knowledge of statistical methods. The sample represented non-academic practitioners, and several recognised that their lack of understanding of these techniques as a hindrance. This experience likely reflects some encounters with articles reporting research findings in academic journals (though only one respondent mentioned this issue in response to the query about access to journals (Figure 15)).

- *Interpretation barriers with respect to static v. interactive versions*: *Interpretation barriers were more often reported for the interactive compared to static versions -*
    - *For the maps, there was minimal difference between respondents' views about barriers to interpretation of the static versions, compared to those for the interactive versions;*
    - *With respect to design, for both the line and bar visualizations the responses were more favourable towards the interactive versions over the static;*
    - *With respect to lack of expertise, only the bar charts showed a clear favouring of the interactive versions over the static.*

There were fewer barriers to overcome for maps as no specialist knowledge was needed for interpretation. But it is surprising no difference was reported between versions with respect to design barriers, given the greater size of the interactive maps than static.

It is interesting that for the line and bar charts, interactivity apparently somewhat ameliorated the design flaws (Question 4 in Figure 19), and for the bar charts, interactivity appeared to ameliorate the effect of lack of expertise (Question 5). This is consistent with the literature showing the benefits of integrating interaction appear to depend in large part on the skill level of individual users (Table 4). For example in a study of risk-perception, the interactive graphics were rated more helpful by low-numeracy than high-numeracy participants (Ancker et al., 2011).

5.2.3 Objective 5 – Effect of integration of interactivity into data visualizations on cognitive load

- *More respondents found it easier to gain insights from the interactive versions of visualizations, than from the static versions, and likewise more respondents found it quicker;*

- *One third of respondents felt that the difference between interactive and static charts with respect to speed and ease of gaining insights depended on the form of the visualization (map, line or bar);*

Separate questions were asked about speed (to explore cognitive load) and ease (to explore mental effort). *"Cognitive load"* refers to the work, or cognitive resources needed to perform a task. The performance measure of speed (as well as accuracy) is assumed to be associated with "cognitive load" and thus often used to evaluate effectiveness (Section 2.2.3). *"Mental effort"* refers to the proportion of cognitive capacity allocated to a task (Huang et al., 2009), and is germane to this study because, given two visualizations of poor and good design, a user might employ more mental effort to compensate for the greater cognitive load consequent to poor design, still maintaining the performance level as attained with the better designed visualization.

It is tempting to over-interpret the finding that interaction seemed to impact more on ease than speed. Given the sample was only twelve individuals, and only one changed their response, this finding might be due to chance and will not be emphasised. More noteworthy is the consistency of findings across objective 4 and 5.

- *All respondents found interactivity helpful in increasing their understanding of the findings, and in communicating and supporting the researcher's findings. The support for understanding was perceived to be greater than for communication.*

Given the design of the visualizations, it is not surprising interaction reportedly aided understanding. The line and bar charts were cluttered, and interaction enabled viewing of simpler versions. Interestingly, although a few respondents had already said it was quicker and/or easier to gain insights with the static versions, in answer to the question about extent of the helpfulness of interactivity in increasing understanding of findings, ALL participants responded it was helpful to some degree. This inconsistency can perhaps be explained by discerning two aspects of insight:

- the experience – the "Aha!" moment when a mental model is restructured; and
- the product of this experience - the changed mental model that represents new knowledge or understanding (Dove and Jones, 2012)

The experience may have been more vivid with the interactive versions, thus creating a stronger memory of their viewing (this question was asked after evaluation of the visualizations).

- *All respondents found the charts engaging, and the 75% of respondents who perceived a difference between versions found the interactive versions more engaging than the static.*

Objectively measures of performance are not always consistent with users' or designers' intuitions relating to effectiveness (Hegarty, 2011). Some argue technical measures like speed and accuracy do not adequately capture what users experience as an 'effective' visualization, and that perhaps "engagement" should also be considered as a facet of effectiveness (Kennedy et al., 2016). Here there is no inconsistency between responses about engagement and those from the other questions.

5.2.4 Overall validity and generalizability of design study findings

5.2.4.1 Limitations of the dataset

Small sample size: There were 13 participants (11 double survey sets and 2 single). The target was 20.

Sampling strategy: The study is identified as "qualitative" rather than "quantitative" since it relates to individual cases and their subjective impressions and experiences in a natural setting, rather than a controlled one. Potential participants were individually identified as being appropriate. The non-probabilistic approach to sampling means that measures of uncertainty in statistics derived from the coded and compiled responses would be meaningless. However, some might consider that the lack of measures of uncertainty indicates the study lacks rigour and so the findings lack validity.

Bias introduced by sampling and recruitment method: The target population, to which the researcher wished to generalise the study results, was non-academic public health nutrition practitioners and nutrition policy makers. Bias was introduced by:

- recruiting from the VRN - the annual subscription (£145) inhibits many nutritionists from joining;
- limiting the potential sample to those who provided a LinkedIn profile and email address, thus excluding those unwilling to be contacted by strangers (the latter criteria was later relaxed to access more potential participants); and
- self-selection of only those with interest in the topic and time available – 17% of invitees.

Two more participants were recruited from the researcher's network who were not on the VRN, but would meet the professional criteria to become members if they wished.

Small number of interactive techniques: Interactivity can assist with many tasks (Table 3), but the techniques integrated into the survey visualizations only aided selection.

5.2.4.2 Quality of findings

The findings are valid and useful, despite the limitations just described because:

- While the sample is small, it is sufficient to show discernible and consistent favouring of the interactive versions.

- The non-purposive sampling strategy, and omission of confidence intervals when reporting results was appropriate given the qualitative nature of the survey. Responses using exact proportions of the participants were not reported, to avoid communicating a misleading impression of precision.
- Some bias was introduced by the sampling and recruitment methods towards wealthier, highly-motivated and outgoing nutritionists, but there is little reason to think these characteristics would have impacted on views of interactivity.
- While only one aspect of the visualizations was aided by integration of interactive techniques, this was sufficient to demonstrate potential benefits. Integration of more interactive features would have added unnecessary complexity to the design, and diminished the quality of findings.

With respect to generalisability, because the sampling was non-random and small, strictly the findings cannot be generalised beyond the sample included in the study to the wider population of non-academic public health nutritionists. However, given that the bias related to sample selection is considered to have had little impact on responses, the researcher contends that the overall trend of results from a larger more controlled study would probably not differ greatly from those just reported.

5.2.5 Implications and a practical application

The key implication of the central finding - that integration of interactivity into visualizations can further enhance the benefits of visualizations - is that these enhanced benefits could be exploited to the advantage of research users. One means would be for academic public health journals to use available technology to better communicate research findings, and enable and encourage their audiences to explore the data underlying the findings.

The methods used by academic researchers in public health to disseminate their findings are not necessarily those that best connect those working in public health practice with research evidence. Academic journals and academic conferences are the most frequently used means (reported by 99% and 81%) (Brownson et al., 2018). All the participants in this study reported using academic publications to access research evidence (Section 4.2.1), with cost and lack of physical or electronic access as the main barriers to increased use. Half had reported their own findings in journals, and/or made presentations in the workplace or at conferences, so would be familiar with the challenges of summarising their own and others' research findings.

Regrettably the content, format and language of scientific publications often do not align with practitioners' needs (Arevalo et al., 2020). Lack of time, information overload and jargon were also reported as barriers here (Section 4.2.1). Decision- and policy-makers need scientific evidence but are diverse and not all have statistical expertise. Page-limited print layouts of academic journals can impose rigid formats on visualizations, and only one version offered, so outputs make assumptions about users' expertise and interests (McInerny et al., 2014).

Given the development and accessibility of new technology, it can be argued that journals are now able to cater for diverse audiences and should attempt to do so. The study results reported in Section 4.2 indicate that integration of simple interactivity into visualizations can further enhance the benefits of visualizations over tables and text. The interactive features evaluated here represent only a small proportion of those available. Integration of more features would enhance the power of interactive visualizations to enable users explore curated datasets for their own purposes. The facility could be provided as part of supplementary materials to articles, and depending on the article's subject, it may be appropriate to use the "narrative visualization" approach in the supplement (Segel and Heer, 2010).

Practitioners' views on the use of visualizations in academic journals were sought in the second survey. The comments (Table A33) show respondents broadly supported the notion that visualizations' role in communication of research findings via journals should be increased. They recognised the value of enabling users to explore the data and extract what they need, and identified practical constraints like the probably lack of authors' skills to produce interactive visualizations, so the journals would need to provide support.

Within environmental science there are initiatives related to using visualizations in academic journals to help bridge the science - practice gap (Grainger et al., 2016, McInerny et al., 2014). In public health this aim is apparently being addressed solely through stand-alone tools (for example Monsivais et al., 2018, Zakkar and Sedig, 2017). Some journals provide tips about use of social media to "*extend the reach and impact of your paper*",[22] or offer links with a company to create a video, infographic, or visual abstract.[23] But this study provides support for the concept that academic public health journals go one step further, and provide interactive interfaces. As well as benefiting practitioners and other non-academic users, this would enable additional value to be obtained from expensive survey data.

[22] https://www.biomedcentral.com/getpublished/writing-resources/author-tips

[23] https://www.springernature.com/gp/researchers/article-promotion?utm_source=publicationletters&%20utm_medium=email&utm_campaign=springer

6 EVALUATION, REFLECTIONS AND CONCLUSIONS

The first section of this chapter describes the researcher's views of the project, and so is written in the first person to maximise clarity.

6.1 Evaluation and reflections

Choice of objectives: I recognised early on that the first version of the objectives was far too ambitious (Appendix A). The final objectives (Section 1.4) were much more reasonable but still challenging to achieve. I considered reducing them down again. I'm glad now that I didn't do this, as I achieved more having these objectives as a target. Both criteria specified in Section 1.4 were satisfied for all five objectives.

Literature review: Having two components of the study from such diverse disciplines made the literature review challenging. I was much more familiar with the public health nutrition literature, so with respect to jargon the literature review for this component was easier. But it was less interesting - the articles were detailed, and I needed to extract and compare their quite dry findings carefully. The visualization literature was tougher with respect to language and methods, but the overall task more stimulating. It was exciting to get to grips with a new discipline and gain an overview of several topics within it. I probably missed some important sources, but my main critique of Chapter 2 is that I did not describe a unifying psychological framework and provide a clear graphic for it that I could refer to in Chapter 5 to help explain the findings of the design study.

Methods used: I consider the methods used for both study components are as good as possible given the time available for the project. With more time –

- I would have used other approaches to recruit more participants for the surveys because a larger sample would have strengthened the findings;
- Some design aspects of the survey visualizations could have been improved. But the versions used delivered on the principal requirement of being a good basis for the evaluation questions.

Findings: A weakness that would probably be noted by academic peer reviewers is the lack of measures of uncertainty in estimates (except for the bar charts of Figures 6, 7, A25 and A26, and several tables in the Appendices). Related to this issue, and as might be perceived from footnote 16, at times perhaps I squeezed findings from my data analysis that were not fully defensible due to the small size of sub-groups. I was trying to keep the visualizations clean and clutter free, and not be too constrained by my traditional background in null-hypothesis significance testing. Some might deem I moved too far in the other direction.

Discussion:

- The discussion of findings from the secondary data analysis only touches on possible reasons behind the patterns identified, and there is nothing about implications of absolute inequalities for health outcomes, relating to separate dietary variables (fruit and vegetables, sodium etc). This would have strengthened the project from a public health perspective, but I realised it was more important to allocate my remaining word allowance to discussion of the design study.

- When discussing the survey findings, I found I lacked a good understanding of the cognitive processes underlying visualization. By then it was too late to appraise more work.
- The excessive length of Section 5.2.5 is simply justified by the theme being close to my heart!

Planning: My original workplan (Appendix A) was adequate as a guide, although it underestimated the time needed for the statistical analysis and visual analytics. The plan was constantly amended during the research process. As described in Chapter 1, there were three delays consequent to the Covid19 pandemic. I'm grateful to the university for granting me an extension to the hand-in date, since more than half of the survey responses were submitted after the original deadline.

The final product: The instructions state that a final project report has typically 12-15,000 words with an upper limit of 30,000 words. The length of this report is close to the limit, in the main not due to verbosity[24] but because the study involved two separate components for which there was no overlap in literature review and methods. I sympathise with readers of the first parts of Chapters 4 and 5 because reports from analysis of this kind of data are quite unexciting, as I experienced during my literature review! However, this report's readers are more fortunate than I was, because here many visualizations break up the text and hopefully provide interest.

Future work:

For the secondary data analysis -

- The next stage would be publication of findings in a journal like BMC Nutrition. Or I could wait for release of the next two years' NDNS data, integrate them with my dataset and rerun the analyses to include 11 years rather than nine. A report covering years 10 and 11 was released in December 2020 (Bates et al., 2020) so release of the additional data is expected to be imminent.
- More detailed analysis of % with low folate intake by demographic sub-group is warranted.

With respect to the design study component, this initial study could usefully inform more in-depth research of the potential benefits of integrating interactivity into visualizations. I found a gap in the literature on how interactivity impacts on the association between visualizations' effectiveness and their complexity. Better understanding of this topic would benefit public health researchers such as myself, as well as those in other disciplines of course.

Main challenges:

- **My own limitations** especially with respect to speed of coding, and discovering the reasons for unsuccessful attempts to create or amend visualizations using Litvis;

[24] I recognise some might apply this term to parts of the results and discussion chapters..

- **The Covid19 pandemic** caused delays in the process as described above;
- **Lack of supervisory input at the end of the project**. My online supervisions were helpful and enlightening, and the quality of project outcome would have benefited from more expert input. The two first factors listed above meant I finished the survey analysis and report-writing late, leaving minimal opportunity to obtain guidance on issues that had arisen (I didn't try).

What I would do differently:

- Knowing now how long the statistical analysis and visual analytics took, I realise I should have started the secondary data analysis earlier, and not waited for feedback on the project proposal;
- I slightly went off on a tangent at the start of the project, spending too long using Tableau to look for evidence of dietary inequalities in sub-groups. I checked for the impact of interactions between the socio-economic, demographic, temporal and geographic variables, when this level of detail was not needed to address the objectives.

6.2 Conclusions

Diet-related chronic disease is an important public health challenge in the UK. Diet-related health problems are not evenly distributed in society – incidence of disease outcomes including heart disease, stroke, and some cancers are higher in lower socioeconomic groups. The first half of this project provides a valuable contribution to help understand and address this issue. It used high-quality representative data, and is the first to simultaneously explore inequality in the UK diet using three different SES indicators, across regions and time. By showing inequalities exist for some key components of the diet, and there has been negligeable change in inequalities over the decade prior to 2017, findings replicate those of previous studies on associations between diet and SES in the UK and underline the need for more effective interventions than currently exist.

The project's second component explored whether interactivity in visualizations could reduce the gap between researchers and practitioners with respect to communication of research findings, and thereby increase the benefits public health practitioners gain from complex datasets. There is only a weak evidence-base for claims that interaction in visualization can augment users' understanding of data. This study convincingly showed that the benefits visualizations offer with respect to knowledge transfer and cognitive load can be increased by integrating interactivity. By making the change to their systems suggested in this report, academic public health journals could better support non-academic users.

Thus, the research project generated useful results about nutrition-related inequalities, and ways more benefit could be gained from research findings. Application of health-related knowledge generated from original research helps improve the nation's health, and the project contributes towards effecting this important goal.

7 REFERENCES

ALKERWI, A. A., VERNIER, C., SAUVAGEOT, N., CRICHTON, G. E. & ELIAS, M. F. 2015. Demographic and socioeconomic disparity in nutrition: application of a novel Correlated Component Regression approach. *BMJ open,* 5.

ALLEN, L., WILLIAMS, J., TOWNSEND, N., MIKKELSEN, B., ROBERTS, N., FOSTER, C. & WICKRAMASINGHE, K. 2017. Socioeconomic status and non-communicable disease behavioural risk factors in low-income and lower-middle-income countries: a systematic review. *The Lancet Global Health,* 5**,** e277-e289.

ANCKER, J. S., WEBER, E. U. & KUKAFKA, R. 2011. Effects of game-like interactive graphics on risk perceptions and decisions. *Medical Decision Making,* 31**,** 130-142.

ANDERSON, E. W., POTTER, K. C., MATZEN, L. E., SHEPHERD, J. F., PRESTON, G. A. & SILVA, C. T. A user study of visualization effectiveness using EEG and cognitive load. Computer graphics forum, 2011. Wiley Online Library, 791-800.

ANDRIENKO, N. & ANDRIENKO, G. 2006. *Exploratory analysis of spatial and temporal data: a systematic approach.* Berlin, Springer

ARCINIEGAS, G., JANSSEN, R. & RIETVELD, P. 2013. Effectiveness of collaborative map-based decision support tools: Results of an experiment. *Environmental modelling & software,* 39**,** 159-175.

AREVALO, V. J. C., VERBRUGGE, L. N., SOOLS, A., BRUGNACH, M., WOLTERINK, R., VAN DENDEREN, R. P., CANDEL, J. H. & HULSCHER, S. J. 2020. Storylines for practice: A visual storytelling approach to strengthen the science-practice interface. *Sustainability Science***,** 1-20.

BADAWOOD, D. & WOOD, J. 2014. The effect of information visualization delivery on narrative construction and development. *Eurographics Conference on Visualization (EuroVis) (2014) Short Papers.* N. Elmqvist, M. Hlawitschka, and J. Kennedy (Editors).

BAILEY, R. 2018. *Evaluating Calorie Intake* [Online]. Office for National Statistics. https://datasciencecampus.ons.gov.uk/eclipse/ [Accessed 14th December 2020].

BARTON, K. L., CHAMBERS, S., ANDERSON, A. S. & WRIEDEN, W. L. 2018. Time to address the double inequality of differences in dietary intake between Scotland and England. *British Journal of Nutrition,* 120**,** 220-226.

BARTON, K. L., WRIEDEN, W. L., SHERRIFF, A., ARMSTRONG, J. & ANDERSON, A. S. 2015. Trends in socio-economic inequalities in the Scottish diet: 2001–2009. *Public Health Nutrition,* 18**,** 2970-2980.

BATES, B., COLLINS, C., COX, L., NICHOLSON, S., PAGE, P., ROBERTS, C., STEER, T. & SWAN, G. 2019. National Diet and Nutrition Survey, Years 1 to 9 of the Rolling Programme (2008/2009 – 2016/2017). London: Public Health England.

BATES, B., JONES, K., PAGE, P., ROBERTS, C., STEER, T. & SWAN, G. 2020. National Diet and Nutrition Survey Rolling programme Years 9 to 11 (2016/2017 to 2018/2019). London: Public Health England.

BATES, B., LENNOX, A., PRENTICE, A., BATES, C., PAGE, P., NICHOLSON, S., MILNE, A. & SWAN, G. 2017. National Diet and Nutrition Survey Rolling Programme (NDNS RP) Results from Years 1-4 (combined) for Scotland (2008/09-2011/12). London: Food Standards Agency and Public Health England.

BATES, B., LENNOX, A., PRENTICE, A., BATES, C., PAGE, P., NICHOLSON, S. & SWAN, G. 2014. National Diet and Nutrition Survey, Results from Years 1, 2, 3 and 4 (combined) of the Rolling Programme (2008/2009 – 2011/2012). London: Public Health England.

BERGER, N., CUMMINS, S., SMITH, R. & CORNELSEN, L. 2019. Recent trends in energy and nutrient content of take-home food and beverage purchases in Great Britain. *BMJ Nutrition, Prevention & Health*.

BESTWICK, J. P., HUTTLY, W. J., MORRIS, J. K. & WALD, N. J. 2014. Prevention of neural tube defects: a cross-sectional study of the uptake of folic acid supplementation in nearly half a million women. *PloS one,* 9**,** e89354.

BLASCH, J., FILIPPINI, M., KUMAR, N. & MARTINEZ-CRUZ, A. 2017. Narrowing the energy efficiency gap: The impact of educational programs, online support tools and energy-related investment literacy. *Economics Working Paper Series.* Zurich: Center of Economic Research at the Swiss Federal Institute of Technology.

BOYD ORR, J. 1936. *Report on a Survey of Adequacy of Diet in relation to Income.* London, Macmillan.

BROWN, V., TURKAY, C. & JIANU, R. 2020. Dissecting visual analytics: Comparing frameworks for interpreting and modelling observed visual analytics behavior. *Eurographics Proceedings 2020* The Eurographics Association.

BROWNSON, R. C., EYLER, A. A., HARRIS, J. K., MOORE, J. B. & TABAK, R. G. 2018. Research full report: getting the word out: new approaches for disseminating public health science. *Journal of Public Health Management and Practice,* 24**,** 102.

CALIN-JAGEMAN, R. J. & CUMMING, G. 2019. The new statistics for better science: ask how much, how uncertain, and what else is known. *The American Statistician,* 73**,** 271-280.

CARR-HILL, R. A., CHALMERS-DIXON, P., LIN, J. & BRITAIN, G. 2005. *The public health observatory handbook of health inequalities measurement.* Oxford, South East Public Health Observatory.

CARROLL, L. N., AU, A. P., DETWILER, L. T., FU, T.-C., PAINTER, I. S. & ABERNETHY, N. F. 2014. Visualization and analytics tools for infectious disease epidemiology: a systematic review. *Journal of Biomedical Informatics,* 51**,** 287-298.

CASNER, S. M. 1991. Task-analytic approach to the automated design of graphic presentations. *ACM Transactions on Graphics,* 10**,** 111-151.

CINNAMON, J., RINNER, C., CUSIMANO, M. D., MARSHALL, S., BEKELE, T., HERNANDEZ, T., GLAZIER, R. H. & CHIPMAN, M. L. 2009. Evaluating web-based static, animated and interactive maps for injury prevention. *Geospatial Health,* 4**,** 3-16.

CLEVELAND, W. S. & MCGILL, R. 1984. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association,* 79**,** 531-554.

DEPARTMENT OF HEALTH 1991. Dietary Reference Values, A Guide London: HMSO.

DEY, S. K., RAHMAN, M. M., SIDDIQI, U. R. & HOWLADER, A. 2020. Analyzing the epidemiological outbreak of COVID-19: A visual exploratory data analysis approach. *Journal of Medical Virology,* 92**,** 632-638.

DHAR, V. 2013. Data science and prediction. *Communications of the ACM,* 56**,** 64-73.

DIMARA, E. & PERIN, C. 2019. What is interaction for data visualization? *IEEE transactions on visualization and computer graphics,* 26**,** 119-129.

DOVE, G. & JONES, S. 2012. Narrative visualization: sharing insights into complex data. *Paper presented at the Interfaces and Human Computer Interaction (IHCI), 21 - 23 Jul 2012, Lisbon, Portugal*.

DUNCAN, G. J., DALY, M. C., MCDONOUGH, P. & WILLIAMS, D. R. 2002. Optimal indicators of socioeconomic status for health research. *American journal of public health,* 92**,** 1151-1157.

FEW, S. 2013. Data visualization for human perception. *The Encyclopedia of Human-Computer Interaction, 2nd Ed.* Aarhus, Denmark: The Interaction Design Foundation.

FIELD, A. 2005. *Discovering statistics using SPSS, 2nd edition.* London, Sage.

FIGUEIRAS, A. Towards the understanding of interaction in information visualization. 2015 19th International Conference on Information Visualisation, 2015. IEEE, 140-147.

FORSELL, C. & JOHANSSON, J. An heuristic set for evaluation in information visualization. Proceedings of the International Conference on Advanced Visual Interfaces, 2010. 199-206.

FOSTER, R. & LUNN, J. 2007. Food availability and our changing diet - 40th Anniversary Briefing Paper. *Nutrition Bulletin,* 32**,** 187-249.

FRIENDLY, M. 2008. A brief history of data visualization. *Handbook of Computational Statistics.* Springer Verlag.

GALOBARDES, B., LYNCH, J. & SMITH, G. D. 2007. Measuring socioeconomic position in health research. *British Medical Bulletin,* 81**,** 21.

GALOBARDES, B., MORABIA, A. & BERNSTEIN, M. S. 2001. Diet and socioeconomic position: does the use of different indicators matter? *International Journal of Epidemiology,* 30**,** 334-340.

GORDIS, L. 1996. *Epidemiology.* Philadephia, W.B.Saunders Company.

GRAINGER, S., MAO, F. & BUYTAERT, W. 2016. Environmental data visualisation for non-scientific contexts: Literature review and design framework. *Environmental Modelling & Software,* 85**,** 299-318.

GREEN, J. & THOROGOOD, N. 2018. *Qualitative methods for health research.* London, Sage.

GREENLAND, S., SENN, S. J., ROTHMAN, K. J., CARLIN, J. B., POOLE, C., GOODMAN, S. N. & ALTMAN, D. G. 2016. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol,* 31**,** 337-350.

HEALTH FOUNDATION. 2020. *COVID-19 – now is the time to redraw the UK's map of inequality* [Online]. London: The Health Foundation. https://www.health.org.uk/news-and-comment/news/covid-19-now-is-the-time-to-redraw-the-uks-map-of-ine [Accessed 06/01/2021].

HEER, J. & SHNEIDERMAN, B. 2012. Interactive dynamics for visual analysis: A taxonomy of tools that support the fluent and flexible use of visualizations. *ACM Queue,* 10**,** 1-26.

HEGARTY, M. 2011. The cognitive science of visual-spatial displays: Implications for design. *Topics in Cognitive Science,* 3**,** 446-474.

HUANG, W., EADES, P. & HONG, S.-H. 2009. Measuring effectiveness of graph visualizations: A cognitive load perspective. *Information Visualization,* 8**,** 139-152.

IBM 2017. IBM SPSS Complex Samples 25. Armonk, NY: IBM Corporation.

ISOKPEHI, R. D., JOHNSON, M. O., CAMPOS, B., SANDERS, A., COZART, T. & HARVEY, I. S. 2020. Knowledge visualizations to inform decision making for improving food accessibility and reducing obesity rates in the United States. *International Journal of Environmental Research and Public Health,* 17**,** 1263.

JAMES, W. P. T., NELSON, M., RALPH, A. & LEATHER, S. 1997. Socioeconomic determinants of health: the contribution of nutrition to inequalities in health. *British Medical Journal,* 314**,** 1545.

JI, C. & CAPPUCCIO, F. P. 2014. Socioeconomic inequality in salt intake in Britain 10 years after a national salt reduction programme. *BMJ Open,* 4**,** e005683.

KEEHNER, M., HEGARTY, M., COHEN, C., KHOOSHABEH, P. & MONTELLO, D. R. 2008. Spatial reasoning with external visualizations: What matters is what you see, not whether you interact. *Cognitive science,* 32**,** 1099-1132.

KENNEDY, H., HILL, R., ALLEN, W. & KIRK, A. 2016. Engaging with (big) data visualizations: Factors that affect engagement and resulting new definitions of effectiveness. *First Monday,* 21.

KIM, M. C., ZHU, Y. & CHEN, C. 2016. How are they different? A quantitative domain comparison of information visualization and data visualization (2000–2014). *Scientometrics,* 107**,** 123-165.

KIRK, A. 2019. *Data Visualisation: A Handbook for Data Driven Design.* London, Sage.

KNAFLIC, C. N. 2013. *Strategies for avoiding the spaghetti graph* [Online]. Storytelling with data. http://www.storytellingwithdata.com/blog/2013/03/avoiding-spaghetti-graph [Accessed 13/01/2021].

KRIEGER, N., CHEN, J. T. & SELBY, J. V. 1999. Comparing individual-based and household-based measures of social class to assess class inequalities in women's health: a methodological study of 684 US women. *Journal of Epidemiology & Community Health,* 53**,** 612-623.

LAM, H., BERTINI, E., ISENBERG, P., PLAISANT, C. & CARPENDALE, S. 2011. Empirical studies in information visualization: Seven scenarios. *IEEE transactions on visualization and computer graphics,* 18**,** 1520-1536.

LAZARD, A. & MACKERT, M. 2014. User evaluations of design complexity: The impact of visual perceptions for effective online health communication. *International journal of medical informatics,* 83**,** 726-735.

LEUNG, G. & STANNER, S. 2011. Diets of minority ethnic groups in the UK: influence on chronic disease risk and implications for prevention. *Nutrition Bulletin,* 36**,** 161-198.

LUCCHESI, L. R. & WIKLE, C. K. 2017. Visualizing uncertainty in areal data with bivariate choropleth maps, map pixelation and glyph rotation. *Stat,* 6**,** 292-302.

MABRY, P. L. 2011. Making sense of the data explosion: the promise of systems science. *American Journal of Preventive Medicine,* 40**,** S159-S161.

MACKENBACH, J. & KUNST, A. 1997. Measuring the magnitude of socio-economic inequalities in health: An overview of available measures illustrated with two examples from Europe. *Social Science & Medicine,* 44**,** 757-771.

MACKENBACH, J. P., KULHÁNOVÁ, I., ARTNIK, B., BOPP, M., BORRELL, C., CLEMENS, T., COSTA, G., DIBBEN, C., KALEDIENE, R. & LUNDBERG, O. 2016. Changes in mortality inequalities over two decades: register based study of European countries. *bmj,* 353.

MAGUIRE, E. R. & MONSIVAIS, P. 2015. Socio-economic dietary inequalities in UK adults: an updated picture of key food groups and nutrients from national surveillance data. *British Journal of Nutrition,* 113**,** 181-189.

MARMOT, M., ALLEN, J., BOYCE, T., GOLDBLATT, P. & MORRISON, J. 2020. *Health Equity in England: The Marmot Review 10 years on.* London, Institute of Health Equity.

MCCLEMENTS, L. D. 1977. Equivalence scales for children. *Journal of Public Economics,* 8**,** 191-210.

MCINERNY, G. J., CHEN, M., FREEMAN, R., GAVAGHAN, D., MEYER, M., ROWLAND, F., SPIEGELHALTER, D. J., STEFANER, M., TESSAROLO, G. & HORTAL, J. 2014. Information visualisation for science and policy: engaging users and avoiding bias. *Trends in Ecology & Evolution,* 29**,** 148-157.

MCNEILL, G., MASSON, L. F., MACDIARMID, J. I., CRAIG, L. C., WILLS, W. J. & BROMLEY, C. 2017. Socio-economic differences in diet, physical activity and leisure-time screen use among Scottish children in 2006 and 2010: are we closing the gap? *Public Health Nutrition,* 20**,** 951-958.

MONSIVAIS, P., FRANCIS, O., LOVELACE, R., CHANG, M., STRACHAN, E. & BURGOINE, T. 2018. Data visualisation to support obesity policy: case studies of data tools for planning and transport policy in the UK. *International Journal of Obesity,* 42**,** 1977-1986.

MOORE, D. A. & CARPENTER, T. E. 1999. Spatial analytical methods and geographic information systems: use in health research and epidemiology. *Epidemiologic Reviews,* 21**,** 143-161.

MUNZNER, T. & MAGUIRE, E. 2014. *Visualization analysis & design.* Boca Raton, CRC Press.

NATCEN SOCIAL RESEARCH 2019a. National Diet and Nutrition Survey (NDNS) Interviewer Project Instructions P10041.04. London: NatCen Social Research.

NATCEN SOCIAL RESEARCH 2019b. National Diet and Nutrition Survey Year 9 (2016/17) User Guide. London: NatCen Social Research.

NATCEN SOCIAL RESEARCH 2019c. National Diet and Nutrition Survey Years 1-9, 2008/09-2016/17. 15th Edition. SN: 6533. UK Data Service, MRC Elsie Widdowson Laboratory.

NELSON, M., ERENS, B., BATES, B., CHURCH, S. & BOSHIER, T. 2007. *Low income diet and nutrition survey, Summary of key findings.* London, The Stationery Office.

NIGHTINGALE, F. 1858. *Mortality of the British army: at home and abroad, and during the Russian war, as compared with the mortality of the civil population in England; illustrated by tables and diagrams* [Online]. London: Harrisons and Sons. https://wellcomelibrary.org/item/b20452433#?c=0&m=0&s=0&cv=0&z=-1.5261%2C-0.1218%2C4.0523%2C1.8128 [Accessed 3/11/2020].

OATES, B. J. 2006. *Researching information systems and computing.* London, Sage.

ONS. 2020. *The National Statistics Socio-economic classification (NS-SEC)* [Online]. Office for National Statistics (ONS). https://www.ons.gov.uk/methodology/classificationsandstandards/otherclassifications/thenationalstatisticssocioeconomicclassificationnssecrebasedonsoc2010 [Accessed 3/11/2020].

PANAGIOTAKOS, D. 2008. α-Priori versus α-posterior methods in dietary pattern analysis: a review in nutrition epidemiology. *Nutrition Bulletin,* 33**,** 311-315.

PATEL, L., ALICANDRO, G. & LA VECCHIA, C. 2020. Dietary approach to stop hypertension (DASH) diet and associated socioeconomic inequalities in the United Kingdom. *British Journal of Nutrition,* 124**,** 1076-1085.

PHE 2012. Patterns and trends in adult obesity. London: Public Health England (PHE).

PREIM, B. & LAWONN, K. A survey of visual analytics for public health. Computer Graphics Forum, 2020. Wiley Online Library, 543-580.

PRYNNE, C., PAUL, A., MISHRA, G., HARDY, R., BOLTON-SMITH, C. & WADSWORTH, M. 2002. Sociodemographic inequalities in the diet of young children in the 1946 British birth cohort. *Public Health Nutrition,* 5**,** 733-745.

RATWANI, R. M., TRAFTON, J. G. & BOEHM-DAVIS, D. A. 2008. Thinking graphically: Connecting vision and cognition during graph comprehension. *Journal of Experimental Psychology: Applied,* 14**,** 36.

RAYNER, M. 1998. Vegetables and fruit are good for us so why don't we eat more? *British journal of nutrition,* 80**,** 119-120.

RENNIE, K. L., COWARD, A. & JEBB, S. A. 2007. Estimating under-reporting of energy intake in dietary surveys using an individualised method. *British Journal of Nutrition,* 97**,** 1169-1176.

RIDGWAY, E., BAKER, P., WOODS, J. & LAWRENCE, M. 2019. Historical developments and paradigm shifts in Public Health Nutrition Science, guidance and policy actions: a narrative review. *Nutrients,* 11**,** 531.

ROBERTS, K., CAVILL, N., HANCOCK, C. & RUTTER, H. 2013. *Social and economic inequalities in diet and physical activity.* London, Public Health England.

ROBERTSON, G., FERNANDEZ, R., FISHER, D., LEE, B. & STASKO, J. 2008. Effectiveness of animation in trend visualization. *IEEE transactions on visualization and computer graphics,* 14**,** 1325-1332.

ROSE, D., PEVALIN, D. J. & O'REILLY, K. 2005. *The National Statistics Socio-economic Classification: origins, development and use.* Basingstoke, Palgrave Macmillan.

SARAIYA, P., NORTH, C. & DUCA, K. 2005. An insight-based methodology for evaluating bioinformatics visualizations. *IEEE Transactions on Visualization and Computer Graphics,* 11**,** 443-456.

SATYANARAYAN, A., MORITZ, D., WONGSUPHASAWAT, K. & HEER, J. 2016. Vega-lite: A grammar of interactive graphics. *IEEE transactions on visualization and computer graphics,* 23**,** 341-350.

SCARBOROUGH, P., MORGAN, R. D., WEBSTER, P. & RAYNER, M. 2011. Differences in coronary heart disease, stroke and cancer mortality rates between England, Wales, Scotland and Northern Ireland: the role of diet and nutrition. *BMJ open,* 1**,** e000263.

SCHOFIELD, C., STEWART, J. & WHEELER, E. 1989. The social diets of pregnant and post-pregnant women in different groups in London and Edinburgh: calcium, iron, retinol, ascorbic acid and folic acid. *British Journal of Nutrition,* 62**,** 363-377.

SCHOFIELD, C., WHEELER, E. & STEWART, J. 1987. The diets of pregnant and post-pregnant women in different social groups in London and Edinburgh: energy, protein, fat and fibre. *British Journal of Nutrition,* 58**,** 369-381.

SEDLMAIR, M., MEYER, M. & MUNZNER, T. 2012. Design study methodology: Reflections from the trenches and the stacks. *IEEE transactions on visualization and computer graphics,* 18**,** 2431-2440.

SEGEL, E. & HEER, J. 2010. Narrative visualization: Telling stories with data. *IEEE transactions on visualization and computer graphics,* 16**,** 1139-1148.

SHAH, P. & CARPENTER, P. A. 1995. Conceptual limitations in comprehending line graphs. *Journal of Experimental Psychology: General,* 124**,** 43.

SHNEIDERMAN, B. 1983. Direct manipulation: a step beyond programming languages. *Sparks of innovation in human-computer interaction,* 17**,** 1993.

SI HASSEN, W., CASTETBON, K., CARDON, P., ENAUX, C., NICOLAOU, M., LIEN, N., TERRAGNI, L., HOLDSWORTH, M., STRONKS, K. & HERCBERG, S. 2016. Socioeconomic indicators are independently associated with nutrient intake in French adults: a DEDIPAC study. *Nutrients,* 8**,** 158.

STANSFIELD, S., WALSH, J., PRATA, N. & EVANS, T. 2006. Information to improve decision making for health. *In:* JAMISON, D., BREMAN, J., MEASHAM, A., ALLEYNE, G., CLAESON, M., EVANS, D., JHA, P., MILLS, A. & MUSGROVE, P. (eds.) *Disease control priorities in developing countries, 2nd edition.* New York: Oxford University Press.

STEELE, J. & ILIINSKY, N. 2011. *Designing data visualizations.* Sebastopol, CA, O'Reilly Media, Inc.

STRECKER, J. & COX, A. 2012. *Data visualization in review: part of the strategic evaluation on communicating research for influence.* Ottawa, International Development Research Centre.

TUFFREY, V. 2018. Why is inequality increasing in mortality from respiratory disease in England? An exploratory analysis of public health and deprivation data using visual analytic techniques. *Unpublished report submitted in partial fulfilment of the requirements for MSc Data Science, City, University of London.*

TUFFREY, V. & HALL, A. 2016. Methods of nutrition surveillance in low-income countries. *Emerg Themes Epidemiol,* 13**,** 4.

TUFTE, E. R. 2001. *The visual display of quantitative information, Second Edition.* Cheshire, Connecticut, Graphics Press

VETTER, N. & MATTHEWS, I. 1999. *Epidemiology and Public Health Medicine.* Edinburgh, Churchill Livingstone.

VLISMAS, K., STAVRINOS, V. & PANAGIOTAKOS, D. B. 2009. Socio-economic status, dietary habits and health-related outcomes in various parts of the world: a review. *Central European Journal of Public Health,* 17**,** 55-63.

WARE, C. 2013. *Information visualization: perception for design (3rd edition).* Waltham, Massachusetts, Morgan Kaufmann.

WEBSTER-GANDY, J., MADDEN, A. & HOLDSWORTH, M. 2011. *Oxford handbook of nutrition and dietetics.* Oxford, Oxford University Press.

WOOD, J., KACHKAEV, A. & DYKES, J. 2018. Design exposition with literate visualization. *IEEE Transactions on Visualization and Computer Graphics,* 25**,** 759-768.

XEXAKIS, G. & TRUTNEVYTE, E. 2019. Are interactive web-tools for environmental scenario visualization worth the effort? An experimental study on the Swiss electricity supply scenarios 2035. *Environmental modelling & software,* 119**,** 124-134.

YAU, A., ADAMS, J. & MONSIVAIS, P. 2019. Time trends in adherence to UK dietary recommendations and associated sociodemographic inequalities, 1986-2012. *European Journal of Clinical Nutrition,* 73**,** 997-1005.

YI, J. S., AH KANG, Y., STASKO, J. & JACKO, J. A. 2007. Toward a deeper understanding of the role of interaction in information visualization. *IEEE transactions on visualization and computer graphics,* 13**,** 1224-1231.

YI, J. S., KANG, Y.-A., STASKO, J. T. & JACKO, J. A. Understanding and characterizing insights: how do people gain insights using information visualization? Proceedings of the 2008 Workshop on BEyond time and errors: novel evaLuation methods for Information Visualization, 2008. 1-6.

ZAKKAR, M. & SEDIG, K. 2017. Interactive visualization of public health indicators to support policymaking: An exploratory study. *Online Journal of Public Health Informatics,* 9.

ZHU, Y. Measuring effective data visualization. International Symposium on Visual Computing, 2007. Springer, 652-661.

ZIKMUND-FISHER, B. J., DICKSON, M. & WITTEMAN, H. O. 2011. Cool but counterproductive: interactive, Web-based risk communications can backfire. *Journal of medical Internet research,* 13**,** e60.

ZUK, T., SCHLESIER, L., NEUMANN, P., HANCOCK, M. S. & CARPENDALE, S. Heuristics for information visualization evaluation. Proceedings of the 2006 AVI workshop "BELIV'06 BEyond time and errors: novel evaLuation methods for Information Visualization", 2006. 1-6.

8 APPENDICES

A Project proposal

<u>Project Proposal for MSc Data Science</u>

Veronica Tuffrey
Email: Veronica.Tuffrey@city.ac.uk **Mob:** 07881 931459

Project Title: Case-study of data visualization for communication of complex research findings
An appraisal of data visualizations' effectiveness in conveying inequalities in diet, using data from the UK National Diet and Nutrition Survey
Supervisor: Dr Jo Wood

1. Introduction

"Information overload" is a common experience in daily life. For those working in science-based professions, it can endanger quality of decision-making. Exponential growth of scientific literature, facilitated by new channels of publication, means it is more important than ever that information communication has integrity and is effective. Data visualization has a central role to play here.

Scientific disciplines in which data visualization is especially valuable are epidemiology and that in which epidemiology is systematically applied, public health[1]. Research studies in these fields, as well as routine population surveillance, involve temporal, geographic, socio-demographic and behavioural data. The resultant datasets are therefore complex, and graphs and maps are essential tools both for analysis and presentation of findings.

- For subject-specialists, use of data visualizations to communicate findings from data analysis can help inform development of health policy and design of health-related interventions.
- For non-specialists, effective communication of epidemiological information is also important, as demonstrated by the COVID19 pandemic - the public hankers for reliable data presented in an easy-to-digest format enabling comparisons over time and space. Politicians and civil-servants need recent data presented in a format aiding speedy evidence-based decision-making.

Studies on the effectiveness of data visualization are scarce. There is a plethora of literature on the practice, but much guidance is based on knowledge acquired through experience (Lonsdale and Lonsdale, 2019, p.6). Within the specific fields of epidemiology and public health, studies on the visualizations' effectiveness to communicate complex research findings are lacking, as is guidance for practitioners about producing visualizations, in contrast to the field of environmental science (for example Harold et al., 2017, Grainger et al., 2016). Some generic guidance exists but would benefit from being updated (Kelleher and Wagener, 2011). Hence, this study's purpose is as follows:

<u>Research aim</u>
• To inform evidence-based guidance on data visualization to communicate public health research findings from studies with large samples and complex study design (for example, including temporal, geographic, socio-demographic and behavioural elements)

The visualizations used to investigate effectiveness in the proposed study will be products of data analysis relating to health inequalities – a phenomenon well documented in the UK over the last 50 years (Marmot, 2006), and a concern that was recently highlighted anew. A study shows that in parts of England, life expectancy has fallen over the past decade after many years of increase. Also evidence suggests that the years people spend in poor health are increasing, and that these trends particularly involve women in deprived communities and the north (Marmot et al., 2020). Socio-economic differences in diet are a potential contributor to health inequalities, and the links between diet and socio-economic status were last reviewed in 2015 using data collected during the annual National Diet and Nutrition Survey (NDNS) between 2008-11 (Maguire and Monsivais, 2015).

[1] Epidemiology is "*the study of how disease is distributed in populations and of the factors that influence or determine this distribution*"(Gordis, 1996, p.3) and Public Health is "*the process of promoting health, preventing disease, prolonging life and improving the quality of life through the organised efforts of society*" (Vetter & Matthews, 1999, p.3)

1

While a report on findings from the most recently collected NDNS data available in the public domain (2016/7) is available on the gov.uk website (Bates et al., 2019), summarised findings have not yet been published in a peer-reviewed journal, so visualizations presenting such findings are ideal for an evaluation of communication effectiveness. The research questions for this study, and objectives by which they will be achieved are as follows:

Primary research questions:
• What are the past and present patterns of dietary inequalities in the UK? • Does interactivity facilitate engagement and effectiveness of visualizations based on complex data?
Research objectives:
1. Ascertain the patterns of variation in the UK diet with socio-economic status (SES); 2. Explore the extent to which the associations between diet and SES differ between regions; 3. Examine time trends in dietary inequalities in the UK; 4. Test the extent to which users' engagement with data visualizations - and the visualizations' effectiveness – are associated with the visualizations' complexity; 5. Test the extent to which interactivity (enabling the user to tailor their view of the data according to their own objectives/preferences) impacts on users' engagement with data visualizations and visualizations' effectiveness; 6. Develop an outline of concise guidance on use of visualizations for communication of complex research findings relating to public health.

Outcomes from the study, and the main beneficiaries are tabulated below:

Deliverables / 'products'	*Beneficiaries – and how they will benefit*
A review of the patterns of association between diet, socio-economic status and geographical region, and trends over time in England. This will comprise text, tables, maps and graphs.	Public health practitioners, including analysts in in the Health and Wellbeing Directorate of Public Health England - Findings are expected to highlight regions and population groups at risk, and can therefore inform priority setting in policy.
A report on the benefits of using interactivity to increase engagement and understanding.	Researchers and practitioners who use data visualization in their work - Findings can inform their practice and enhance effectiveness of their visualizations.
An outline of guidelines for using data visualizations to communicate complex research findings relating to public health. The guidelines will be fully developed after the dissertation.	Public health researchers and practitioners - Guidelines based on research into visual perception and principles of design, as well as the findings from this study, will help this group produce visualizations that are functional, accurate and engaging, without giving up data content.

2. Critical Context

Data visualization can be defined as *"..the visual representation and presentation of data to facilitate understanding"* (Kirk, 2019, p.15). Two of the first documented uses of visualizations for analysis and communication of findings related to data from British populations are:
- First, by William Playfair who between 1786 and 1821 wrote several works with *"..many excellent examples of the line graph, circle graph, bar graph and pie diagram.."* (Funkhouser, 1937, cited in FitzPatrick, 1960). Playfair used economic data to graphically demonstrate how British workers' welfare had improved over time (Friendly, 2008).
- Second, by John Snow who in 1854 plotted the location of water pumps and deaths from cholera in central London, and demonstrated that cases occurred close to one specific pump (Tufte, 2001p.24). Removal of that pump's handle is an example of how effective communication of epidemiological findings can have immediate and valuable policy implications for public health action (Gordis, 1996, p.257).

During the 19th century there was a rapid increase in the use of statistical graphics and thematic mapping, and another surge occurred in the last twenty years or so, partly linked to the increased availability of new technologies and software products which facilitate the creation of visualizations (Strecker and Cox, 2012). There is now an *"international community of scholars"* exploring information and data visualization (Lindquist, 2011). Exponents of the practice draw on the science of visual perception, such as Colin Ware (Ware, 2013), and contend that visualizing data is typically more effective for communicating information than text-based versions

(Lindquist, 2011). There are numerous publications describing good data visualization practice (for example, Kirk, 2019, Munzner and Maguire, 2014) and providing instructions (for example, Fisher and Meyer, 2018, Suda, 2010), nonetheless, some argue the rapid expansion of visualization use has been accompanied by a general lack of understanding, which can cancel the benefits and even lead to confusion (Strecker and Cox, 2012).

One use of data visualisations is to communicate research evidence, a process which is not straightforward because users' engagement depends in large part on context such as emotive state, skills and beliefs (Allen, 2018). This dependency has implications for assessment of "effectiveness" of visualisations (Kennedy et al., 2016). Evidence on the links between the constructs of engagement and effectiveness (definitions are provided in Section 3) and the role of visualizations' complexity in this is lacking. Nonetheless, some claim that data visualizations can transform complex findings in a way that communicates research to a broader audience, and thus facilitate dialogue between academics, policy-makers, and the wider public (Gatto, 2015).

When research findings are derived from large and complex datasets, creating data visualizations that are comprehensible to users is a challenge, since humans' "*capacity limits of attention*" influence the visualizations' effectiveness (Haroz and Whitney, 2012). To address these limits, data abstraction (simplification via reduction to a set of essential characteristics) can reduce visual clutter and dataset size (Elmqvist and Fekete, 2009), and when done by the user, this is an example of "interactivity". Schneiderman described this in the 1980s when he first used the term *"direct manipulation"* to describe an interaction style in which users act on displayed objects using incremental and reversible actions for which the effects are immediately visible on the screen (Shneiderman, 1983) – now a commonplace approach. His taxonomy of information visualization tasks (Shneiderman, 1996) places abstraction operations such as filter and zoom in an overall context, which notes the importance of gaining an overview first. However there is still little evidence of interactivity's effectiveness with respect to improving understanding (Figueiras, 2015).

The visualizations used to investigate effectiveness will be products of data analysis relating to dietary inequalities[2] in the UK, specifically, the data collected as part of the National Diet and Nutrition Survey (NDNS, NatCen Social Research, 2019b). In the latest report, data collected from 2012 to 2016 revealed trends for intakes of fruit, vegetables, fruit juice and oily fish to be higher, and for sugar-sweetened soft drinks to be lower, with increasing income (Bates et al., 2019). The most recent relevant publication in a peer-reviewed journal used NDNS data collected between 2008 – 11, and socio-economic differences were reported for consumption of three food groups (fruit and vegetables, red and processed meat and oily fish) and one nutrient of public health importance (non-milk extrinsic sugars) (Maguire and Monsivais, 2015). Social gradients in the UK diet had already been identified (Mishra et al., 2004, Bates et al., 2019, Bolton-Smith et al., 1991). Also geographical variation (Barton et al., 2018, Schofield et al., 1987, Scarborough et al., 2011), and time trends (for example, Foster and Lunn, 2007, Berger et al., 2019, Yau et al., 2019) have been described. But as far as the author is aware, the proposed study will be the first to collectively examine socio-economic, demographic, geographical and temporal influences on the UK diet.

3. Approaches: Methods and tools for design, analysis and evaluation

There are four sub-sections below, respectively describing the methods used for the two main components of the study, the study methods' limitations, and their ethical implications. The software packages to be used in this project are the following:

- SPSS Statistics for Windows for most statistical analysis (V25.0, Armonk, NY: IBM Corp), together with Python programming language (Python Software Foundation (www.python.org/)
- Tableau (www.tableau.com/) and Litvis (www.gicentre.net/litvis) for visualizations.

[2] Here the term "inequalities" is defined as for health inequalities, that is, "differences … between individual people of higher and lower socio-economic status" (Carr-Hill et al, 2005, p.3)

3.1 Data analysis and visualization design (to address research objectives 1 – 3).

3.1a Design: Secondary data analysis.

3.1b Sample, and methods for primary data collection: The dataset to be analysed is the product of a government-funded survey, the National Diet and Nutrition Survey (NDNS) which is intended to assess the dietary habits and nutritional status of the general adult and child population in the UK. The NDNS programme began in 1992, and since 2008 has been conducted annually as part of a rolling programme (NatCen Social Research, 2019b). Nine years' data are available up to 2016/17, and can be downloaded from https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=6533 .

The sample represents the UK general population aged ≥1.5 years living in private households. Two-stage sampling is used - first all UK addresses are clustered into geographical areas (the Primary Sampling Unit; PSU) and PSUs randomly sampled (158 in 2016), second a random sample of addresses is drawn from each PSU (28 in 2016). Around 500 adults and 500 children take part each year. Recruitment in Wales and Northern Ireland is boosted to get representative country-specific data. Data collection consists of a 4-day diet diary, face-to-face Computer Assisted Personal interviews, physical measurements, self-completion questionnaires, and blood and urine samples (NatCen Social Research, 2019b).

3.1c Methods used for secondary data analysis: As outlined in Section 1, the objectives of this component are exploration of patterns of variation in diet with SES; of the extent to which associations between diet and SES differ between regions; and of time trends in these associations.
I Data extraction: To address these three objectives, data relating to five categories of variables (shown below, together with justifications) will be extracted from NDNS datasets.

Variable category	*Names*
Dietary	*Foods*: Fruit and vegetables; red and processed meat; oily fish; *Macronutrients*: Total dietary energy; % energy from fat; % energy from saturated fats; % energy from non-milk extrinsic sugars (NMES); *Vitamins and minerals*: Iron, sodium, Vitamin D
Reason: There is strong evidence that these dietary elements act as protective factors or risk factors with respect to chronic disease (such as cardiovascular disease and some cancers) and other health-related outcomes like immunity and cognition (Webster-Gandy et al., 2011). Their variation with socio-economic status has been examined in earlier relevant studies (Maguire and Monsivais, 2015, Ji and Cappuccio, 2014, Yau et al., 2019, Schofield et al., 1987).	
(Socio-)Demographic	Age-group (7 categories), sex, ethnicity (2 categories)
Reason: As is the case for health outcomes (Carr-Hill et al., 2005), the association between SES and diet may be modified by age-group, gender and ethnicity, so it is important to test for interactions.	
Socio-economic status	Income, occupational social class (8 categories), educational attainment (8 categories)
Reason: Three SES variables will be used, because there is evidence these three indicators contribute independently to dietary outcomes (Maguire and Monsivais, 2015). Also limitations of each measure differ (Galobardes et al., 2007)	
Geographical	Region (12 categories)
Reason: There geographical differences in the nutritional qualities of national diets within the UK, and it will be useful to test if these are partly independent of SES, as for Scotland (Barton et al., 2018).	
Temporal	Year
Reason: It is important to test if dietary inequalities are widening in parallel with those in health.	

II Data coding, checking and manipulation:

- Datasets will be merged, variable names edited, variable categories merged (to reduce the number of age and socio-economic categories[3], and enable comparison of England v. Scotland, Wales and Northern Ireland[4], and weighting variables created to enable correction for differences in sample selection and response according to instructions provided (NatCen Social Research, 2019ap.22). Weighting variables will also be constructed for analysis of data collected across all 9 years, since sample sizes included each year varied in each dataset[5].
- For continuous variables (all dietary variables and income), the shape of their distributions will be checked, and if markedly non-normal the data will be transformed using logarithmic or square root functions for positive skew, and square function for negative skew. If income is

[3] Variable NSEC8 is based on the National Statistics Socio-Economic Classification. The 8 classes can be combined into 5 or 3 classes (https://gss.civilservice.gov.uk/wp-content/uploads/2016/03/P4-Economic-Status-June-16.pdf p.16)
[4] Variable "GOR" (government office region) has 12 categories, 9 for England, 1 each for Scotland, Wales and N.Ireland
[5] Sample sizes are Years 1-4, N=6,828; Years 5&6, N=2,546; Years 7&8, N=2,723 and Year 9 N=1,253. Total N=13,350

extremely skewed, it will be converted into categories, and binary variables will be created for dietary variables if skew exists due to large proportions of non-consumers.

III Data analysis: First frequency tables will be tabulated for the categorical independent variables. Next descriptive statistics will be tabulated for the dietary variables, with respect to the three SES variables, age-group, ethnicity, sex, educational attainment, and region. Finally, stepwise multiple linear regression will be used to examine association of each dietary variable with SES variables. The other independent variables may be included in the models as covariates, using adjusted R squared values to compare models' effectiveness, using AIC and BIC[6] as error criteria, and QQ plots to evaluate if assumptions underlying the method are met (Field, 2005).

3.1d Design of data visualizations: Visualizations will be produced to further examine, and finally to communicate, patterns of variation of the dietary variables with the independent variables. For example, maps will help elucidate geographical variation in dietary inequalities both cross-sectionally and longitudinally. For this, decisions will be needed on how best to use visual attributes to communicate patterns of variation[7]. For each dietary variable, it will probably prove preferable to make mini-time graphs for each of the four countries and display on a single map, rather than make a series of maps for each year. Some of the visualizations will incorporate interactive features, such as search bars and the ability to reveal extra information by hovering over an image. Finally, a series of visualisations will be selected for use in the evaluation, and these will be annotated and user guides created for them (Kirk, 2019, p.234).

3.2 Evaluation of visualizations (to address research objectives 4 – 5).

3.2a Design: Cross-sectional survey

3.2b Sample: The extent to which complexity hinders, and interactivity facilitates engagement and effectiveness of visualizations based on complex data is likely to be affected by individual-level factors (Allen, 2018), specifically self-perceived numeracy, demographics, and attitudes. "*Decoding of meaning*" also depends on the educational and class background of the viewer (Hall, 1973) and underlying emotional state (Kennedy et al., 2016). It will be impossible in this study to explore the effect of all these elements, so two alternative approaches could be adopted – to try to control these elements by recruiting participants with as similar characteristics as possible, or to recruit participants with as diverse characteristics as possible. The strategy remains to be chosen in consultation with the project supervisor, but the first is the most likely.

A sample from the general public will be recruited (if students or academics, they should not be subject specialists). The sampling method remains to be chosen in consultation with the project supervisor. Available options range from recruitment of personal contacts, through recruitment from the pool of City University students and academics, to the extreme of recruitment through online crowdsourcing platforms such as Amazon MTurk (Sheehan, 2018). Limitations and ethical issues will be raised for any method chosen, and these will be fully discussed in the final report[8]. The target sample size is 25. Justification for this number is not a power calculation, since the data will not be quantitative in nature. Rather it is based on the evidence that in a qualitative study with a fairly specific research question, little new information is gained after interviews with around 20 people belonging to one participant 'category' (Green and Thorogood, 2018, p.102).

3.2c Methods for data collection: An online questionnaire will be produced. Questions will assess engagement with, and effectiveness of visual outputs from the study's first component as described in Section 3.1. Engagement is a complex construct, and the definition adopted for this study focuses on the aspect referred to as "affective engagement", that is the "*..user's emotional*

[6] AIC and BIC are both "penalized-likelihood criteria" used for choosing best predictor subsets in regression

[7] Space- and time-referenced data can be viewed either as a "*spatial arrangement of local behaviours over time*", or as a "*sequence of momentary behaviours over the territory*" (Andrienko and Andrienko, 2006, p.9)

[8] The page limits for this assignment mean these issues are discussed only in very general terms in Sections 3.3 and 3.4

5

involvement or investment while interacting with a visualisation" (Hung and Parsons, 2018). In contrast, a suitable definition for "effectiveness" with respect to information visualization could not be found but the construct was decomposed into three aspects as tabulated below (Zhu, 2007):

1) Principle of accuracy: For a visualization to be effective, the attributes of visual elements shall match the attributes of data items, and the structure of the visualization shall match the structure of the data set.
2) Principle of utility: An effective visualization should help users achieve the goal of specific tasks
3) Principle of efficiency: An effective visualization should reduce the cognitive load for a specific task over non-visual representations.

Element 1) is dependent on the researcher, while 2) and 3) are contingent on users and are task focused. Since the proposed study relates specifically to scientific communication (with the implicit goal that users' knowledge is increased) the following definition, derived from Zhu's review, will be adopted: "*An effective visualization transfers knowledge and reduces the cognitive load for users over non-visual representations.*" These two definitions will guide the construction of the questionnaire. An example of an online survey on engagement using the Likert scale, is at https://yahsin.github.io/VisEngage/example.html. Questions in this style will be included in the study questionnaire, together with open-ended questions. Qualtrics survey software (www.qualtrics.com) will be used.[9] The draft survey and visuals will be pilot-tested and amended before the full survey. SPSS will be used for data analysis of the quantitative responses, and for open-ended questions, key themes and illustrative quotations will be identified, and responses summarised by theme.

3.3 Limitations

3.3a Findings from secondary data analysis: Internal validity of the survey data will likely have been affected by non-response bias since those who participated may be systematically different from those who did not, such as in SES. Response rates were between 49% - 55% between 2008-16 for completion of 3 or 4 diet diary days. Also, dietary data are self-reported and may have both random and systematic error, for example, some food categories may have been under-reported due to social desirability bias. The data were collected cross-sectionally, and this design will limit any causal inference to be made between SES and the dietary patterns identified.

3.3b Findings from evaluation of visualizations: Internal validity will likely be affected by non-response bias - those who choose to participate in the survey may be more motivated and engaged than others. External validity will be affected by the design, for example, the sample will not be representative of the public so the findings will not be generalisable to them. Careful planning of the sampling strategy will enable characterisation of the limits of generalisability.

3.4 Ethics (See Research Ethics review form on pp 10-13)

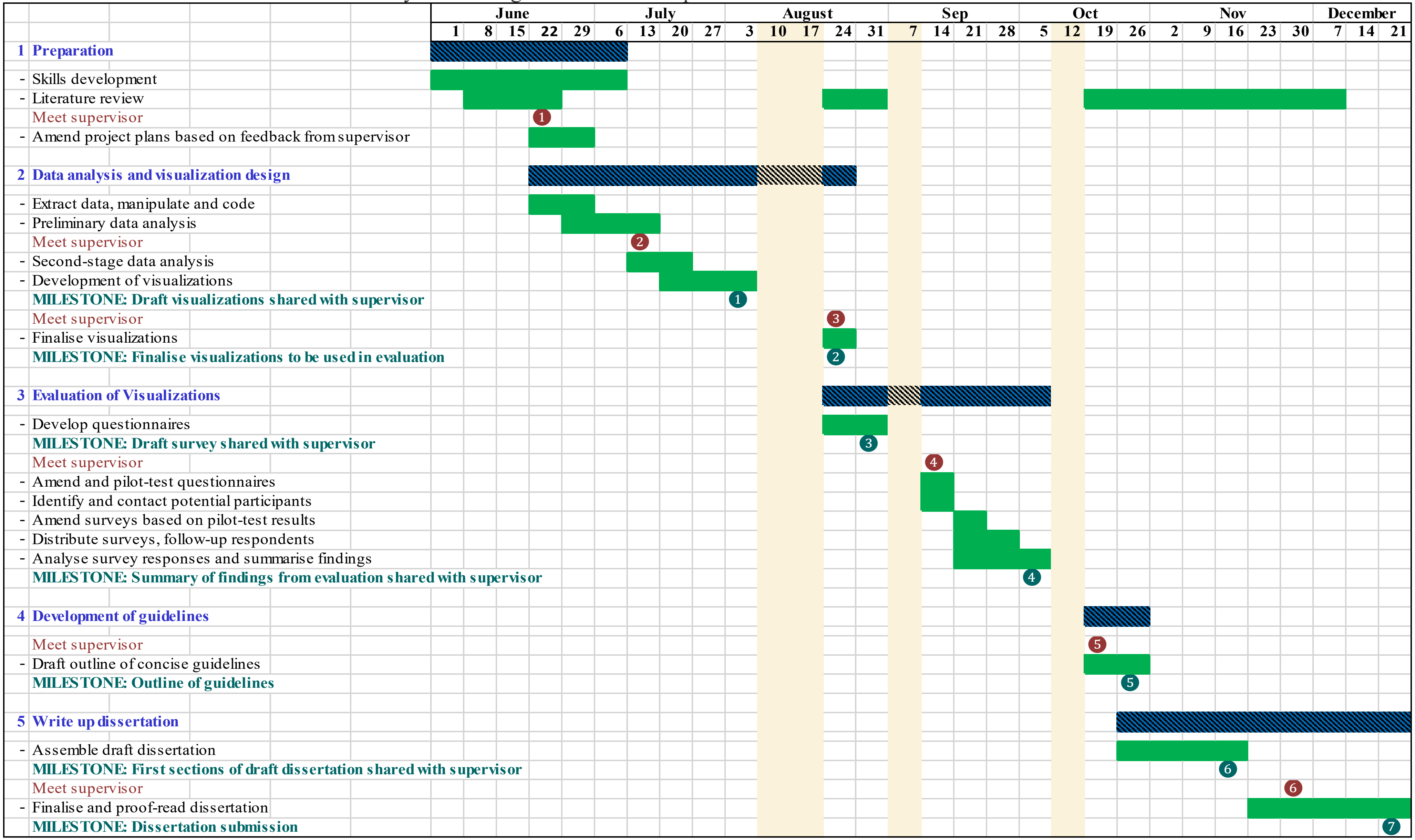
3.4a: Secondary dataset: Participants' names and any other identifiable information like postcode were removed from the datasets so participants cannot be identified. Ethics approval for Years 1-5 was obtained from the Oxfordshire A Research Ethics Committee (Ref. 07/H0604/113), and for Years 6–9 from the Cambridge South NRES Committee (Ref. 13/EE/0016).

3.4b: Evaluation of visualizations: At the start of the survey, there will be a cover page that briefly summarises the project, and a participant information sheet. At the end of the survey, before the 'Submit' button, there will be a statement reminding participants that clicking the button will constitute provision of consent to participate, in full knowledge of the information in the information sheet. Participants will be free to withdraw themselves and their data at any point in the survey, by using a "withdraw" button leading to a short debrief page confirming their data will not be retained. Participants will complete the survey in their place of work or homes, and their privacy will be protected by ensuring anonymity – no personal data will be collected. The survey data will be stored on the researcher's own laptop, which is password protected, and the data will not be accessed by anyone other than the researcher, supervisor or examiners. The data will be retained securely only until the researcher's graduation. Findings will be shared with participants.[10]

[9] This is because City University has an agreement with Qualtrics about rights & responsibilities relating to personal data

[10] using the websites for the Centre for Open Science (https://www.cos.io/osf) or Figshare (https://figshare.com)

4. Work Plan In the Gantt chart below, blue shading shows duration of main stages of project; green shading shows duration of individual tasks; cream columns indicate breaks due to holiday or awaiting feedback from supervisor.



5. Risks The table below identifies the main risks, and uses colour to indicate estimated overall severity (red for highest and green lowest). Overall risk is calculated as sum of ratings for probability of occurrence and severity of impact. The register will be reviewed and updated during the project..

#	Risk	Trigger	Occurrence probability 1-3	Impact severity 1-3	Overall 2 - 6	Action to prevent occurrence	Action to mitigate effect should it occur
1	***Slippage in schedule:*** Slippage due to personal reasons	Loss of motivation Paid work takes priority over university work Perfectionist attitude Illness of self or family member	2 Medium	1 – 3 Low to high (depends on extent of slippage)	3 – 5 Medium to high	Detailed scheduling at project start Inclusion of realistic amounts of slippage time into schedule Tracking of progress v. schedule Making timetable at start of each week Good time-management of other tasks	***Reduce project scope in consultation with supervisor and depending on stage of project e.g.*** Remove draft guidelines (If due to illness) apply for extension of deadline citing extenuating circumstances
2	Slippage due to lack of skills	Data analysis and visualizations' creation harder than envisaged	2 Medium	2 Medium	4 Medium	Improve data analysis and visualization skills while waiting for proposal feedback Use techniques that are already familiar Do analysis in stages	Use findings from early analysis for visualizations Reduce variables for, or quantity of, visualizations
3	Slippage due to external factors	Participants return responses after deadline has passed	2 Medium	1 Low	3 Medium	Schedule extra time for return of responses Send reminders to participants before deadline passes	Progress all other tasks so that can prioritise analysis of responses on arrival
4	Sample size for survey smaller than planned	Low response rate	2 Medium	2 Medium	4 Medium	Contact potential participants early Contact more participants than planned for Have list of backup participants prepared	Accept reduced sample size
5	Quality of deliverables is weak	Research questions and objectives are not well addressed Project slippage as above	1 Low	2 Medium	3 Medium	Keep aim, questions and objectives in mind when carrying out tasks Record activities and findings as they are generated, together with critical comments Avoid slippage using actions above, with additional extra slippage time for write-up Regular meetings with supervisor	Submit log of time allocated to various tasks, to show examiners enough effort was expended Spend extra time at end of project to try to rectify final deliverable
6	Findings are inconclusive	Responses are not internally consistent Responses are not consistent between respondents	3 High	1 Low	4 Medium	Request clarification from participants Recruit more participants	Accept that scale of study was too small to expect clear-cut findings
7	Loss of stored work or data; loss of access to software	Hardware failure Theft Fire Flood	1 Low	1 – 3 Low to high (depends if early/late in project)	2 – 4 Low to medium	Using cloud storage: Create backups of data, analysis outputs and written work at end of each hour and day Create complete project backup weekly Load software on laptop as well as PC	Retrieve lost data or last version of work from backup location Use laptop if PC fails Borrow laptop if own fails

References

ALLEN, W. L. 2018. Visual brokerage: Communicating data and research through visualisation. *Public Understanding of Science*, 27, 906-922.

BARTON, K. L., CHAMBERS, S., ANDERSON, A. S. & WRIEDEN, W. L. 2018. Time to address the double inequality of differences in dietary intake between Scotland and England. *British Journal of Nutrition*, 120, 220-226.

BATES, B., COLLINS, C., COX, L., NICHOLSON, S., PAGE, P., ROBERTS, C., STEER, T. & SWAN, G. 2019. National Diet and Nutrition Survey, Years 1 to 9 of the Rolling Programme (2008/2009 – 2016/2017). London: Public Health England.

BERGER, N., CUMMINS, S., SMITH, R. & CORNELSEN, L. 2019. Recent trends in energy and nutrient content of take-home food and beverage purchases in Great Britain. *BMJ Nutrition, Prevention & Health*.

BOLTON-SMITH, C., SMITH, W., WOODWARD, M. & TUNSTALL-PEDOE, H. 1991. Nutrient intakes of different social-class groups: results from the Scottish Heart Health Study (SHHS). *British Journal of Nutrition*, 65, 321-335.

CARR-HILL, R. A., CHALMERS-DIXON, P., LIN, J. & BRITAIN, G. 2005. *The public health observatory handbook of health inequalities measurement*, Oxford, South East Public Health Observatory.

ELMQVIST, N. & FEKETE, J.-D. 2009. Hierarchical aggregation for information visualization: Overview, techniques, and design guidelines. *IEEE Transactions on Visualization and Computer Graphics*, 16, 439-454.

FIELD, A. 2005. *Discovering statistics using SPSS, 2nd edition.*, London, Sage.

FIGUEIRAS, A. Towards the understanding of interaction in information visualization. 2015 19th International Conference on Information Visualisation, 2015. IEEE, 140-147.

FISHER, D. & MEYER, M. 2018. *Making data visual: a practical guide to using visualization for insight*, Sebastopol, CA, O'Reilly.

FOSTER, R. & LUNN, J. 2007. Food availability and our changing diet - 40th Anniversary Briefing Paper:. *Nutrition Bulletin*, 32, 187-249.

FRIENDLY, M. 2008. A brief history of data visualization. *Handbook of Computational Statistics*. Springer Verlag.

FUNKHOUSER, H. 1937. Historical Development of the Graphical Representation of Statistical Data *Osiris*, 3.

GALOBARDES, B., LYNCH, J. & SMITH, G. D. 2007. Measuring socioeconomic position in health research. *British medical bulletin*, 81, 21.

GATTO, M. A. 2015. Making research useful: Current challenges and good practices in data visualisation. Oxford: Reuters institute.

GORDIS, L. 1996. *Epidemiology*, Philadelphia, W.B Saunders Company.

GRAINGER, S., MAO, F. & BUYTAERT, W. 2016. Environmental data visualisation for non-scientific contexts: Literature review and design framework. *Environmental Modelling & Software*, 85, 299-318.

GREEN, J. & THOROGOOD, N. 2018. *Qualitative methods for health research*, London, Sage.

HALL, S. 1973. Encoding and decoding in the television discourse. Birmingham: Centre for Cultural Studies, University of Birmingham.

HAROLD, J., LORENZONI, I., COVENTRY, K. & MINNS, A. 2017. *Enhancing the accessibility of climate change data visuals: Recommendations to the IPCC and guidance for researchers*, Norwich, UK, Tyndall Centre for Climate Change.

HAROZ, S. & WHITNEY, D. 2012. How capacity limits of attention influence information visualization effectiveness. *IEEE Transactions on Visualization and Computer Graphics*, 18, 2402-2410.

HUNG, Y.-H. & PARSONS, P. 2018. Affective engagement for communicative visualization: Quick and easy evaluation using survey instruments. *Visualization for Communication (VisComm)*, 5.

JI, C. & CAPPUCCIO, F. P. 2014. Socioeconomic inequality in salt intake in Britain 10 years after a national salt reduction programme. *BMJ Open*, 4, e005683.

KELLEHER, C. & WAGENER, T. 2011. Ten guidelines for effective data visualization in scientific publications. *Environmental Modelling & Software*, 26, 822-827.

KENNEDY, H., HILL, R., ALLEN, W. & KIRK, A. 2016. Engaging with (big) data visualizations: Factors that affect engagement and resulting new definitions of effectiveness. *First Monday*, 21.

KIRK, A. 2019. *Data Visualisation: A Handbook for Data Driven Design*, London, Sage.

LINDQUIST, E. 2011. Surveying the world of visualization, Background paper prepared for the HC Coombs policy forum. Canberra: Australian National University.

LONSDALE, M. & LONSDALE, D. 2019. Design2Inform: Information visualisation guidelines. The Office of the Chief Scientific Advisor, Gov UK.

MAGUIRE, E. R. & MONSIVAIS, P. 2015. Socio-economic dietary inequalities in UK adults: an updated picture of key food groups and nutrients from national surveillance data. *British Journal of Nutrition*, 113, 181-189.

MARMOT, M. 2006. Introduction. *In:* MARMOT, M. & WILKINSON, R. (eds.) *Social Determinants of Health*. Oxford: Oxford University Press.

MARMOT, M., ALLEN, J., BOYCE, T., GOLDBLATT, P. & MORRISON, J. 2020. Health Equity in England: The Marmot Review 10 years on. London: Institute of Health Equity.

MISHRA, G., PRYNNE, C., PAUL, A., GREENBERG, D. & BOLTON-SMITH, C. 2004. The impact of inter-generational social and regional circumstances on dietary intake patterns of British adults. *Public Health Nutrition*, 7, 737-744.

MUNZNER, T. & MAGUIRE, E. 2014. *Visualization analysis & design*, Boca Raton, CRC Press.

NATCEN SOCIAL RESEARCH 2019a. National Diet and Nutrition Survey Year 9 (2016/17) User Guide. London: NatCen Social Research.

NATCEN SOCIAL RESEARCH 2019b. National Diet and Nutrition Survey Years 1-9, 2008/09-2016/17. 15th Edition. SN: 6533. . UK Data Service, MRC Elsie Widdowson Laboratory.

SCARBOROUGH, P., MORGAN, R. D., WEBSTER, P. & RAYNER, M. 2011. Differences in coronary heart disease, stroke and cancer mortality rates between England, Wales, Scotland and Northern Ireland: the role of diet and nutrition. *BMJ open*, 1, e000263.

SCHOFIELD, C., WHEELER, E. & STEWART, J. 1987. The diets of pregnant and post-pregnant women in different social groups in London and Edinburgh: energy, protein, fat and fibre. *British Journal of Nutrition*, 58, 369-381.

SHEEHAN, K. B. 2018. Crowdsourcing research: data collection with Amazon's Mechanical Turk. *Communication Monographs*, 85, 140-156.

SHNEIDERMAN, B. 1983. Direct manipulation: a step beyond programming languages. *Sparks of innovation in human-computer interaction*, 17, 1993.

SHNEIDERMAN, B. The eyes have it: A task by data type taxonomy for information visualizations. Proceedings 1996 IEEE symposium on visual languages, 1996. IEEE, 336-343.

STRECKER, J. & COX, A. 2012. *Data visualization in review: part of the strategic evaluation on communicating research for influence*, Ottawa, International Development Research Centre.

SUDA, B. 2010. *A Practical Guide to Designing with Data*, Penarth, Five Simple Steps.

TUFTE, E. R. 2001. *The visual display of quantitative information, Second Edition*, Cheshire, Connecticut, Graphics Press.

WARE, C. 2013. *Information visualization: perception for design (3rd edition)*, Waltham, Massachusetts, Morgan Kaufmann.

WEBSTER-GANDY, J., MADDEN, A. & HOLDSWORTH, M. 2011. *Oxford handbook of nutrition and dietetics*, Oxford, Oxford University Press.

YAU, A., ADAMS, J. & MONSIVAIS, P. 2019. Time trends in adherence to UK dietary recommendations and associated sociodemographic inequalities, 1986-2012. *European Journal of Clinical Nutrition*, 73, 997-1005.

ZHU, Y. Measuring effective data visualization. International Symposium on Visual Computing, 2007. Springer, 652-661.

Research Ethics Review Form: BSc, MSc and MA Projects

Computer Science Research Ethics Committee (CSREC)

http://www.city.ac.uk/department-computer-science/research-ethics

Undergraduate and postgraduate students undertaking their final project in the Department of Computer Science are required to consider the ethics of their project work and to ensure that it complies with research ethics guidelines. In some cases, a project will need approval from an ethics committee before it can proceed. Usually, but not always, this will be because the student is involving other people ("participants") in the project.

In order to ensure that appropriate consideration is given to ethical issues, all students must complete this form and attach it to their project proposal document. There are two parts:

PART A: Ethics Checklist. All students must complete this part.

The checklist identifies whether the project requires ethical approval and, if so, where to apply for approval.

PART B: Ethics Proportionate Review Form. Students who have answered "no" to all questions in A1, A2 and A3 and "yes" to question 4 in A4 in the ethics checklist must complete this part. The project supervisor has delegated authority to provide approval in such cases that are considered to involve MINIMAL risk.
The approval may be *provisional – identifying the planned research* as likely to involve MINIMAL RISK.
In such cases you must additionally seek *full approval* from the supervisor as the project progresses and details are established. *Full approval* must be acquired in writing, before beginning the planned research.

A.1 If you answer YES to any of the questions in this block, you must apply to an appropriate external ethics committee for approval and log this approval as an External Application through Research Ethics Online - https://ethics.city.ac.uk/		*Delete as appropriate*
1.1	Does your research require approval from the National Research Ethics Service (NRES)? *e.g. because you are recruiting current NHS patients or staff?* *If you are unsure try - https://www.hra.nhs.uk/approvals-amendments/what-approvals-do-i-need/*	**NO**
1.2	Will you recruit participants who fall under the auspices of the Mental Capacity Act? *Such research needs to be approved by an external ethics committee such as NRES or the Social Care Research Ethics Committee - http://www.scie.org.uk/research/ethics-committee/*	**NO**
1.3	Will you recruit any participants who are currently under the auspices of the Criminal Justice System, for example, but not limited to, people on remand, prisoners and those on probation? *Such research needs to be authorised by the ethics approval system of the National Offender Management Service.*	**NO**
A.2 If you answer YES to any of the questions in this block, then unless you are applying to an external ethics committee, you must apply for approval from the Senate Research Ethics Committee (SREC) through Research Ethics Online - https://ethics.city.ac.uk/		*Delete as appropriate*
2.1	Does your research involve participants who are unable to give informed consent? *For example, but not limited to, people who may have a degree of learning disability or mental health problem, that means they are unable to make an informed decision on their own behalf.*	**NO**
2.2	Is there a risk that your research might lead to disclosures from participants concerning their involvement in illegal activities?	**NO**
2.3	Is there a risk that obscene and or illegal material may need to be accessed for your research study (including online content and other material)?	**NO**
2.4	Does your project involve participants disclosing information about special category or sensitive subjects?	**NO**

10

	For example, but not limited to: racial or ethnic origin; political opinions; religious beliefs; trade union membership; physical or mental health; sexual life; criminal offences and proceedings	
2.5	Does your research involve you travelling to another country outside of the UK, where the Foreign & Commonwealth Office has issued a travel warning that affects the area in which you will study? *Please check the latest guidance from the FCO - http://www.fco.gov.uk/en/*	NO
2.6	Does your research involve invasive or intrusive procedures? *These may include, but are not limited to, electrical stimulation, heat, cold or bruising.*	NO
2.7	Does your research involve animals?	NO
2.8	Does your research involve the administration of drugs, placebos or other substances to study participants?	NO
A.3 If you answer YES to any of the questions in this block, then unless you are applying to an external ethics committee or the SREC, you must apply for approval from the Computer Science Research Ethics Committee (CSREC) through Research Ethics Online - https://ethics.city.ac.uk/ **Depending on the level of risk associated with your application, it may be referred to the Senate Research Ethics Committee.**		*Delete as appropriate*
3.1	Does your research involve participants who are under the age of 18?	NO
3.2	Does your research involve adults who are vulnerable because of their social, psychological or medical circumstances (vulnerable adults)? *This includes adults with cognitive and / or learning disabilities, adults with physical disabilities and older people.*	NO
3.3	Are participants recruited because they are staff or students of City, University of London? *For example, students studying on a particular course or module.* *If yes, then approval is also required from the Head of Department or Programme Director.*	NO
3.4	Does your research involve intentional deception of participants?	NO
3.5	Does your research involve participants taking part without their informed consent?	NO
3.5	Is the risk posed to participants greater than that in normal working life?	NO
3.7	Is the risk posed to you, the researcher(s), greater than that in normal working life?	NO
A.4 If you answer YES to the following question and your answers to all other questions in sections A1, A2 and A3 are NO, then your project is deemed to be of MINIMAL RISK. **If this is the case, then you can apply for approval through your supervisor under PROPORTIONATE REVIEW. You do so by completing PART B of this form.** **If you have answered NO to all questions on this form, then your project does not require ethical approval. You should submit and retain this form as evidence of this.**		*Delete as appropriate*
4	Does your project involve human participants or their identifiable personal data? *For example, as interviewees, respondents to a survey or participants in testing.*	YES

PART B: Ethics Proportionate Review Form

If you answered YES to question 4 and NO to all other questions in sections A1, A2 and A3 in PART A of this form, then you may use PART B of this form to submit an application for a proportionate ethics review of your project. Your project supervisor has delegated authority to review and approve this application under proportionate review. You must receive final approval from your supervisor in writing before beginning the planned research.

However, if you cannot provide all the required attachments (see B.3) with your project proposal (e.g. because you have not yet written the consent forms, interview schedules etc), the approval from your supervisor will be *provisional*. You must submit the missing items to your supervisor for approval prior to commencing these parts of your project. Once again, you must receive written confirmation from your supervisor that any provisional approval has been superseded by with *full approval* of the planned activity as detailed in the full documents. **Failure to follow this procedure and demonstrate that final approval has been achieved may result in you failing the project module.**

Your supervisor may ask you to submit a full ethics application through Research Ethics Online, for instance if they are unable to approve your application, if the level of risks associated with your project change, or if you need an approval letter from the CSREC for an external organisation.

B.1 The following questions must be answered fully. All grey instructions must be removed.		*Delete as appropriate*
1.1.	Will you ensure that participants taking part in your project are fully informed about the purpose of the research?	YES
1.2	Will you ensure that participants taking part in your project are fully informed about the procedures affecting them or affecting any information collected about them, including information about how the data will be used, to whom it will be disclosed, and how long it will be kept?	YES
1.3	When people agree to participate in your project, will it be made clear to them that they may withdraw (i.e. not participate) at any time without any penalty?	YES
1.4	Will consent be obtained from the participants in your project? Consent from participants will be necessary if you plan to involve them in your project or if you plan to use identifiable personal data from existing records. "Identifiable personal data" means data relating to a living person who might be identifiable if the record includes their name, username, student id, DNA, fingerprint, address, etc. *If YES, you must attach drafts of the participant information sheet(s) and consent form(s) that you will use in section B.3 or, in the case of an existing dataset, provide details of how consent has been obtained.* *You must also retain the completed forms for subsequent inspection. Failure to provide the completed consent request forms will result in withdrawal of any earlier ethical approval of your project.*	YES
1.5	Have you made arrangements to ensure that material and/or private information obtained from or about the participating individuals will remain confidential?	YES

B.2 If the answer to the following question (B2) is YES, you must provide details		
2	Will the research be conducted in the participant's home or other non-University location? *If YES, you must provide details of how your safety will be ensured.*	YES

B.3 Attachments **ALL of the following documents MUST be provided to supervisors if applicable.**	*YES*	*NO*	*Not Applicable*

12

All must be considered prior to final approval by supervisors. A written record of final approval must be provided and retained.			
Details on how safety will be assured in any non-University location, including risk assessment if required (see B2)			N/A
Details of arrangements to ensure that material and/or private information obtained from or about the participating individuals will remain confidential (see B1.5) *Any personal data must be acquired, stored and made accessible in ways that are GDPR compliant.*	**To be provided**		
Full protocol for any workshops or interviews**			N/A
Participant information sheet(s)**	**To be provided**		
Consent form(s)**			N/A
Questionnaire(s)** *sharing a Qualtrics survey with your supervisor is recommended.*	**To be provided**		
Topic guide(s) for interviews and focus groups**			N/A
Permission from external organisations or Head of Department** *e.g. for recruitment of participants*			N/A

***If these items are not available at the time of submitting your project proposal, then **provisional approval** can still be given, under the condition that you must submit the final versions of all items to your supervisor for approval at a later date. All such items must be seen and approved by your supervisor before the activity for which they are needed begins. Written evidence of **final approval** of your planned activity must be acquired from your supervisor before you commence.*

Changes

If your plans change and any aspects of your research that are documented in the approval process change as a consequence, then any approval acquired is invalid. If issues addressed in Part A (the checklist) are affected, then you must complete the approval process again and establish the kind of approval that is required. If issues addressed in Part B are affected, then you must forward updated documentation to your supervisor and have received written confirmation of approval of the revised activity before proceeding.

Templates for Consent and Information

You must use the templates provided by the University as the basis for your participant information sheets and consent forms. You must adapt them according to the needs of your project before you submit them for consideration.

Participant Information Sheets, Consent Forms and Protocols must be consistent. Please ensure that this is the case prior to seeking approval. Failure to do so will slow down the approval process.

We strongly recommend using Qualtrics to produce digital information sheets and consent forms.

Further Information

http://www.city.ac.uk/department-computer-science/research-ethics

https://www.city.ac.uk/research/ethics/how-to-apply/participant-recruitment

https://www.city.ac.uk/research/ethics

B NDNS Fieldwork and data analysis methods

Fieldwork

NDNS fieldwork is undertaken by a consortium comprising NatCen Social Research and the MRC Elsie Widdowson laboratory, formerly known as MRC Human Nutrition Research. Fieldwork in Northern Ireland is carried out by the Northern Ireland Statistics and Research Agency. Data collection takes place throughout the year (for example, the most recent data available were collected between April 2016 and August 2017) so that seasonal variations in food consumption are taken into account. The survey is also designed so that all days of the week are evenly represented.

Data collection consists of a 4-day diet diary, face-to-face Computer Assisted Personal Interviews (CAPI), physical measurements, self-completion questionnaires, and blood and urine samples. The analysis reported here uses only the dietary diary and CAPI data. These data were collected over 3 main visits to households who had agreed to participate.

Dietary data

Respondents were requested to complete a record of their food intake for four consecutive days which included both weekdays and weekends. For each item consumed, the respondents were asked to provide a description of the item, the quantity (using household measures and photographs) and the time of day it was consumed. The diary was explained and left with the household at the first household visit, and at the second and third visits interviewers prompted for information on potentially missing food items.

Demographic and socio-economic data

For the CAPI interviews, the interviewer read questions from a laptop screen and entered the participants' responses into designated fields. Information was obtained from the household reference person (the person with highest income in whose name the property is owned or rented) or their partner. The Household Reference Person (HRP) was asked questions to ascertain if he/she was in paid work at the time of the interview and, if not, whether they had ever had a paid job. If yes, there were further questions about their current or most recent job to enable classification into the NS-SEC groupings (Section 2.1.1 and Table 6). The HRP was asked to estimate their total household income in the last 12 months, before any deductions for tax, including income from earnings, self-employment, benefits, pensions, and interest from savings. They were asked if they had qualifications from school, college or university, or connected with work or from government schemes. If yes, the HRP was shown a card with 50 options descending in level from higher degree, from which they should identify the first one they had passed.

Quality control and training

Fieldworkers were provided with comprehensive written instructions covering survey procedures. All interviewers received feedback on the diaries from their previous assignment before working on another. All new interviewers were briefed and trained before undertaking an assignment and monitored during their assignment. "Early work" checks were undertaken on the first 2 or 3 food diaries returned, and feedback provided on any areas of concern.

Original data analysis

Intakes of individual foods and nutrients were derived using an in-house dietary assessment system based on the UK Nutrient Databank. This databank contains extensive information on the nutrient content of foods commonly consumed in the UK, it is updated each year, and is currently administered by Public Health England. Each food code in the databank has a value assigned for 56 nutrients and is disaggregated into 28 food components. The databank therefore enables derivation of intakes of individual food components and nutrients from NDNS diary data including those collected on processed foods and composite dishes, whether shop-bought or home-produced.

C Description of NDNS variables used for secondary data analysis

Original variable name	*Variable # in Y9 IND or PLD file*[1]	*Description*	*Variable manip-ulation*[2]	*Final variable type (# of categories)*
INDIVIDUAL IDENTIFICATION (Used for matching IND and PLD files)				
seriali	PLD1 and IND1	Individual serial number	-	Continuous
WEIGHTING				
wti_Y9 (also *wti_78, wtiY56* and *wti_UK1234*)	IND1,441	Weight for individual and diary-all ages, combined Y9 (also Y78, Y56 and Y1234)	-	Continuous
DIETARY - FOODS				
totalfruitandveg	PLD327	Total fruit (not including juice) and vegetables	SqRoot transf.	Continuous
Achieve5	PLD331	Consuming 5 or more portions per day of fruit and vegetables	-	Binary
totalredmeat	PLD348	Total red meat (incl from composite dishes) (g)	SqRoot transf.	Continuous
OilyFishg	PLD344	Oily fish (incl from composite dishes) (g)	Converted to binary	Binary
SoftDrinksNotLowCal	PLD258	Soft drinks not low calorie (g)	Converted to binary	Binary
DIETARY – MACRONUTRIENTS				
Energykcal	PLD11	Total energy (kcal) diet only. Includes energy from alcohol	SqRoot transf.	Continuous
FoodEkcal	PLD12	Food energy (kcal) diet only	SqRoot transf.	Continuous
FatpcfoodE	PLD146	Fat percent food energy		Continuous
SFApcfoodE	PLD150	Saturated fatty acids percent food energy	SqRoot transf.	Continuous
NMESpcfoodE	PLD164	Non-milk extrinsic sugars percent food energy	SqRoot transf.	Continuous
Alcoholg	PLD67	Alcohol (g) diet only	Converted to binary	Binary
AlcoholpctotE	PLD172	Alcohol percent total energy	Log transf.	Continuous
AOACFibreg	PLD34	AOAC Fibre (g) diet only	SqRoot transf.	Continuous
DIETARY - MICRONUTRIENTS				
Sodiummg	PLD52	Sodium (mg) diet only	SqRoot transf.	Continuous
Ironmg	PLD57	Iron (mg) diet only	SqRoot transf.	Continuous
bloIronLRNI	PLD225	Below LRNI Iron	-	Binary
Calciummg	PLD54	Calcium (mg) diet only	SqRoot transf.	Continuous
bloCalciumlrni	PLD224	Below LRNI Calcium	-	Binary
Folateµg	PLD48	Folate (µg) diet only	SqRoot transf.	Continuous

Original variable name	*Variable # in Y9 IND or PLD file*[1]	*Description*	*Variable manip-ulation*[2]	*Final variable type (# of categories)*
bloFolatelrni	PLD219	Below LRNI Folate	-	Binary
VitaminAretinolequivalents	PLD40	Vitamin A (retinol equivalents)	Log transf.	Continuous
bloVitAlrni	PLD213	Below LRNI Vitamin A	-	Binary
VitaminDµg	PLD41	Vitamin D (µg) diet only	SqRoot transf.	Continuous
SOCIO-DEMOGRAPHIC				
AGEGAD2	IND29	(D)[3] Age of respondent 16+, grouped into 5 groups	Dropped first category	Ordinal (x)
Sex	IND24 and PLD5	Sex	-	Binary
Ethgroup2	IND92	(D)[3] Ethnic group, 2 groups	-	Binary
SOCIO-ECONOMIC				
Eqvinc	IND146	(D)[3] Equivalised household income	Derived quartiles	Ordinal (4)
Nssec8	IND71	(D)[3] NS-SEC 8 variable classification	Combined categories	Ordinal (4)
Qual7	IND53	(D)[3] Qualifications gained, grouped	Combined categories	Ordinal (4)
GEOGRAPHIC				
Area	IND1446	Primary sampling unit (PSU)		158
Country	PLD6	Country	-	Nominal (4)
GOR	IND141	Government Office Region	-	Nominal (12)
TEMPORAL				
SurveyYear	PLD2 and IND145	NDNS Survey year	-	Ordinal (9)

[1] IND / PLD = Individual / PersonLevelDietary data (names of files)

[2] Transf. = Transformation, either square root or logarithmic

[3] (D) shows variable had been derived. The SPSS code for this is provided in *NatCen Social Research, 2019a*

D SPSS syntax for rescaling weights for analysis of the combined dataset excluding children

```
/* select the sub-group.
select if age>=19.
weight off.
compute x=1.
/* calculate the sum of the weights for each wave.
aggregate outfile = * mode=addvar
/break x
/n1 = sum(wti_UKY1234)
/n2 = sum(wti_Y56)
/n3 = sum(wti_Y78)
/n4 = sum(wti_Y9).
/* calculate the total of all the four waves' weights.
compute N = sum (n1, n2, n3, n4).
/* Multiply each weight variable by the total (combined) sum of the four weights (N);
   divide each weight variable by the sum of its own weights (n1, n2, n3 and n4), and
   multiply the Year 1-4 weight by 4/9, the Year 5&6 weight by 2/9, the Year 7&8 weight by
2/9 and the Year 9 weight by 1/9.
compute wti_UKY1234r = wti_UKY1234 * N / n1 * (4/9).
compute wti_Y56r = wti_Y56 * N / n2 * (2/9).
compute wti_78r = wti_Y78 * N / n3 * (2/9).
compute wti_9r = wti_Y9 * N / n4 * (1/9).
/* Combine the resulting weights into one variable.
compute wti_UKY1to9Ad = sum (wti_UKY1234r, wti_Y56r, wti_78r, wti_9r).
/* Re-scale this weight to have a mean of 1.
aggregate outfile = * mode=addvar
/break x
/mean=mean(wti_UKY1to9Ad).
compute wti_UKY1to9Ad = wti_UKY1to9Ad / mean.
execute.
```

E Calculation of equivalised income using the McClements scoring system

1) Allocate a score to each household member depending on age and relationship to household reference person as follows:

First adult (HRP)	0.61
Spouse/partner of HRP	0.39
Other second adult	0.46
Third adult	0.42
Subsequent adults	0.36
Dependant aged 0-1	0.09
Dependant aged 2-4	0.18
Dependant aged 5-7	0.21
Dependant aged 8-10	0.23
Dependant aged11-12	0.25
Dependant aged13-15	0.27
Dependant aged16+	0.36

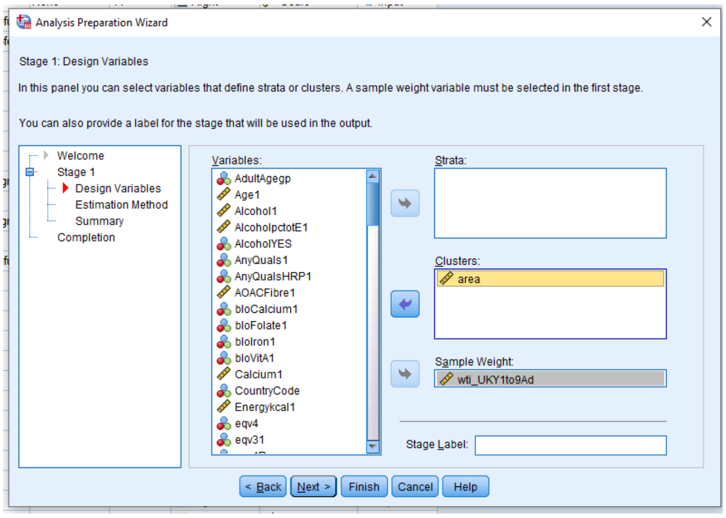
2) Add the scores for all household members to obtain an overall household McClements score.

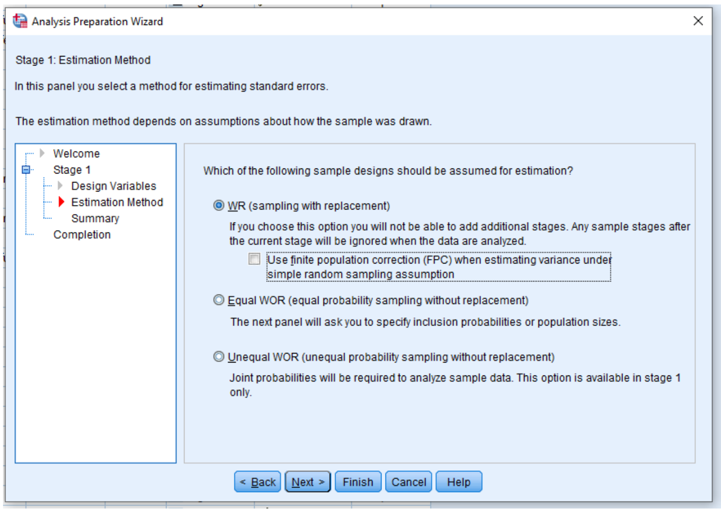
3) Divide the annual household income by the McClements score. This is the equivalised income.

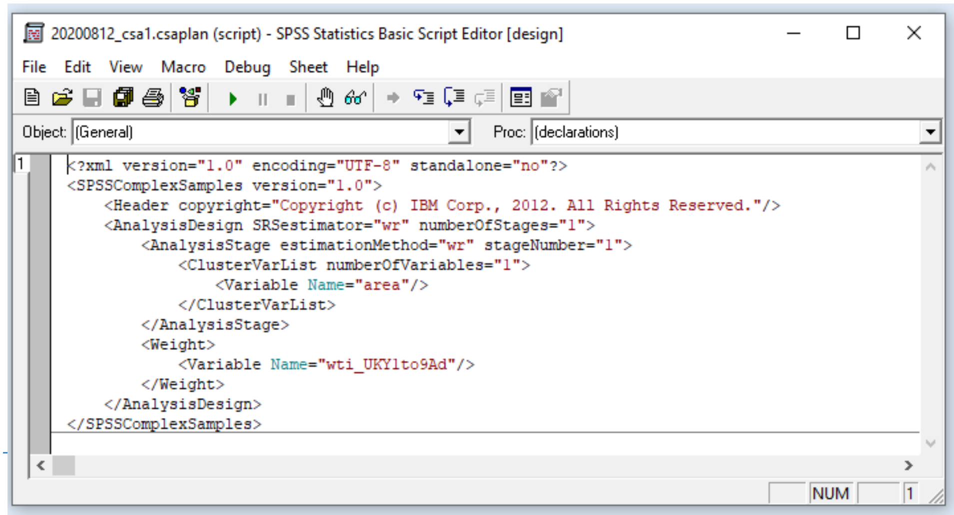
(McClements, 1977)

F Derivation of the plan file used for the Complex Samples module in SPSS

The plan file was derived using the menu-driven Complex Samples "Analysis Preparation Wizard". The options chosen are shown in the SPSS dialog boxes captured in the first two snips below, and the resulting script for the plan file 20200812_csa1 (this plan file was used for all analyses) is shown in the third snip.







G Invitation email and participant information sheet

Invitation to participate in research study

PG-Tuffrey, Veronica <Veronica.Tuffrey@city.ac.uk>

Sat 28/11/2020 19:26

To: [redacted]

Dear [redacted]

Please excuse this unsolicited contact.

I am an independent public health nutritionist registered with the Association for Nutrition, and studying part- time on the MSc Data Science at City University.

I am seeking participants for my research study on visualization of complex public health datasets, and so emailing a small number of fellow AfN registrants who have provided email addresses and weblinks for their LinkedIn profiles on the AfN website, and whose profiles indicated a professional interest in diet and socio- economic inequality.

Participants in the study would spend 15 - 20 minutes providing feedback on a small number of data visualizations I've produced from analysis of the National Diet and Nutrition Survey dataset 2008-2016. The information sheet with details of the research is included below.

I would be very grateful if you could read these details, and then let me know if you would be happy to participate in the study, or if you would like more information.

Thank you in advance, and best regards,

Veronica Tuffrey

https://www.linkedin.com/in/veronica-tuffrey-165a2a22

CITY
UNIVERSITY OF LONDON
EST 1894

Case-study of data visualization for communication of complex research findings

Participant information sheet

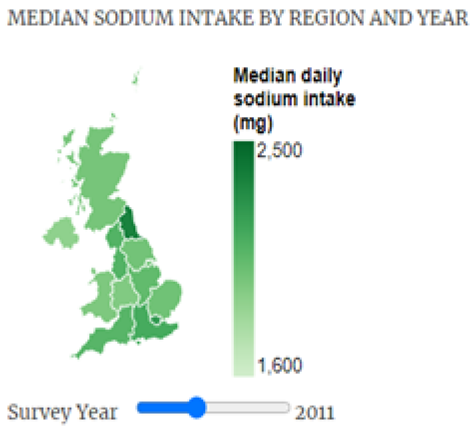
We would be grateful if you could take part in a research study. The information below explains why the research is being done and describes what it would involve for you, so that you can decide whether you wish to take part. Please read the following information, and ask us if anything is not clear or if you would like more information.

What is the purpose of the study?

- The aim of this research is to inform evidence-based guidance on data visualization to communicate public health research findings from studies with large samples and complex study design (for example, including temporal, geographic, socio-demographic and behavioural elements).
- The specific research question addressed by this study is whether interactivity (enabling the user to tailor their view of the data according to their own

objectives/preferences) enhances the effectiveness of visualizations based on complex data. See the section on context below for more details.

- Findings will form part of the researcher's dissertation for the MSc Data Science at City, University of London.



Context: *Visualizations can support data analysis, communication and decision-making. Improved tools to support these tasks would be of value to public health practitioners, given the ever-increasing volume and complexity of data being produced.*

When visualizations are interactive, users can explore charts by changing which data they see. For example, in the chart shown above (snipped from this study's survey) users can choose which year's findings are visible by sliding the circle along the bar.

Interactive graphics are rarely used for communication of findings from public health research. There are many datasets in the public domain (for example from the annual National Diet and Nutrition Survey) with under-exploited potential to help address current public health and nutrition concerns. This study will explore whether interactivity in visualizations could increase the benefits practitioners gain from such datasets.

Why have I been invited?

You have been invited because you are included on the Association for Nutrition register; the register displays an email address and a link to a LinkedIn profile for you, and your profile indicates you are working on issues related to diet and socio-economic inequality.

Do I have to take part?

- Participation in the project is voluntary - you can choose not to participate in this study.
- You may withdraw from the study at any time without an explanation or being penalized.
- You may withdraw any data you have contributed up until submission of the study report to City, University of London (January 2020).

What will happen if I decide to take part?

- The researcher will email links to two surveys.

- You provide your consent to participate via consent forms integrated into the surveys.
- The first survey will include questions about the way(s) your work involves the communication of research findings, and on your interpretation of a small set of visualizations provided
- The second survey will include questions on your interpretation of another small set of visualizations, and about the relative effectiveness of the two sets.
- The two sets of visualizations will differ only by the inclusion of interactivity in one set
    - Half the respondents will encounter interactive visualizations in the first survey, and the remaining participants will encounter static visualizations in the first survey
    - The order in which respondents complete the two surveys will be randomly determined to avoid potential bias due to learning effect.
- The visualizations will illustrate findings from the researcher's analysis of data collected during the National Diet and Nutrition Survey (NDNS) between 2008-2016.
- Each of the two surveys will take around ten minutes, or longer if you provide optional additional responses.
- Your responses will be anonymised and treated as confidential (see below).

What do I have to do?

The surveys will ask a small number of questions which you should try to answer truthfully and in a way that reflects your current professional practice.

What are the possible disadvantages and risks of taking part?

There should be no physical or psychological consequences related to participating in this study.

What are the possible benefits of taking part?

- Responding to the questions may help illuminate how you use data visualizations in your work
- Your input may help shape guidelines on visualization design which may eventually help you with your work

What will happen when the research study is completed?

- The data collected as part of the research study will be held (in an anonymized format) until the researcher's dissertation has been marked, then securely destroyed.
- Findings from the study will be published in the researcher's MSc dissertation.

Please let the researcher know if you would like to be sent the relevant sections of the dissertation.

Will my taking part in the study be kept confidential?

- Only the researcher and her supervisor will have access to the raw survey data.
- Anonymised quotes or information from the study may be presented in the final research report, in such a way that you will not be identifiable.
- Data will be stored only on the researcher's PC which is password protected.

What if there is a problem?

- If you have any problems, concerns or questions, you should speak to the researcher or her supervisor. Their contact details are provided below.
- If you remain unhappy and wish to complain formally, you can do this through City's complaints procedure.
- To make a complaint, you should telephone 020 7040 3040 and ask to speak to the Secretary to Senate Research Ethics Committee. Specify that the project name is "*Case-study of data visualization for communication of complex research findings".*
- You could also write to Anna Ramberg (the University's Research Governance & Integrity Manager) at the following address:

  Anna Ramberg, Secretary, Research Governance & Integrity Committee,
  City, University of London, Northampton Square, London EC1V 0HB
  Anna.Ramberg.1@city.ac.uk

- City holds insurance policies which apply to this study. If you feel you have been harmed or injured by participation you may be eligible to claim compensation. This does not affect your legal rights to seek compensation. If you are harmed due to someone's negligence, then you may have grounds for legal action.

Who has reviewed the study?

This study has been approved following the process specified by City Department of Computer Science Research Ethics Committee (CSREC).

For information on CSREC research governance, please visit:
http://www.city.ac.uk/department-computer-science/research-ethics

Thank you for taking the time to read this information sheet

For further information, please contact:

- Researcher - *Veronica Tuffrey MSc Data Science candidate, Department of Computer Science*

  Veronica.Tuffrey@city.ac.uk

- Supervisor - *Jo Wood, Professor of Visual Analytics, at the giCentre, Department of Computer Science*

  J.D.Wood@city.ac.uk

H Emails with links to surveys

Research on Data Visualization - Link to first survey

PG-Tuffrey, Veronica <Veronica.Tuffrey@city.ac.uk>

Sun 29/11/2020 12:55

To: [redacted]

Dear [redacted]

Many thanks for agreeing to take part in my research study.

Please follow the first link below this message for the first survey.
The second link is for the background information sheet.
Please click and open this document, as you will need it to make sense of the visualizations in the

survey. If you have any questions, please don't hesitate to ask.

With thanks and best regards,
Veronica Tuffrey

Please press Ctrl+click on this link to access the first survey:
[redacted]

Online Survey Software | Qualtrics Survey Solutions

Qualtrics sophisticated online survey software solutions make creating online surveys easy. Learn more about Research Suite and get a free account today.

cityunilondon.eu.qualtrics.com

Follow this link to the background information sheet for the survey:
https://tufty317.github.io/web/Background_Info_sheet.pdf

Second Survey for research on data visualization

PG-Tuffrey, Veronica <Veronica.Tuffrey@city.ac.uk>

Sun 29/11/2020 13:01

To: [redacted]

Dear [redacted]

Thank you for completing the first survey for my research study on data visualization.

Please follow the link at the base of this message for the second survey.
You will need the background information sheet again, that you accessed from the link in the first

email. If you have any questions, please don't hesitate to ask.

With thanks and best regards,
Veronica

Please press Ctrl+click on this link to access the second survey:
[redacted]

Online Survey Software | Qualtrics Survey Solutions

The most powerful, simple and trusted way to gather experience data. Start your journey to experience management and try a free account today.

cityunilondon.eu.qualtrics.com

Case-study of data visualization for communication of complex research findings

Veronica Tuffrey, October 2020

Background information for visualizations

DATA SOURCE -The visualizations used in this study are based on data from a government-funded survey, the National Diet and Nutrition Survey (NDNS) which is intended to assess the dietary habits and nutritional status of the general adult and child population in the UK. Since 2008 the NDNS has been conducted annually as part of a rolling programme[1]. Nine years' data are available up to 2016/17, and can be downloaded from the UK data service website[2]. Around 500 adults and 500 children take part each year. The sample represents the UK general population aged ≥1.5 years living in private households. For this secondary data analysis, only the data collected from adults aged ≥19 years were used.

DATA COLLECTION AND ANALYSIS - The survey uses two-stage sampling. Recruitment in Wales and Northern Ireland is boosted to get representative country-specific data. The fieldwork includes several components, and for the secondary data analysis the researcher only used data from the 4-day diet diaries, and the interviews. She chose to analyse a small number of dietary variables for which there is strong evidence of associations with health-related outcomes, either as protective or risk factors, and examined variation in these variables with respect to age (4 categories), sex, ethnicity (2 categories), three socio-economic variables (occupation, income and educational attainment - 4 categories each), survey year (9) and region (12). For a sub-set of the dietary variables, the researcher used non-parametric correlation, linear regression and logistic regression to explore their temporal and geographical variation, and also their association with the socio-economic variables (the other independent variables were included as covariates in the models).

SELECTED FINDINGS -

Time-trends: Using regression modelling, the four dietary variables examined had statistically significant associations with survey year ($p < 0.001$) across the UK as a whole. The consumption of sodium and red meat, and the proportion of the population consuming alcohol decreased. The proportion of the adult population with folate intake lower than the recommended levels increased. *Chart 1 displays the median values for each year for each geographical region.*

Socio-economic and regional variation: Regression modelling showed the regions with lowest fruit and vegetable intakes were Northern Ireland and the North-East, and the regions with highest intakes were London and the South-West. Non-parametric correlation analysis showed the region with strongest association between fruit and vegetable intake and occupation category was Scotland, while the region with weakest association was East of England. *Chart 2 displays the median total daily consumption of fruit and vegetables by occupation category for each geographical region.*

Associations between dietary, socio-economic and socio-demographic variables: Using regression modelling -

- daily intake of fruit and vegetables had a stronger positive association with education and occupation than with income when statistically adjusted for age-group, sex and ethnicity. On average, total fruit and vegetable intake was lower in the younger and white ethnicity adults compared to the older and non-white ethnicity groups.
- daily intake of red meat had a negative association with education and occupation and no association with income when statistically adjusted for age-group, sex and ethnicity. On average, total red and processed meat intake was lower in women and adults of non-white ethnicity compared to men and adults of white ethnicity.

Chart 3 displays regression coefficients from this modelling. The dietary variables had been transformed to correct for skew, and then standardised to enable comparisons between variables. Error bars show 95% confidence intervals around the values for each category in comparison with the reference categories (highest SES, male, ≥65y, and white ethnicity). Coefficient values greater than zero indicate positive associations.

---

[1] Details about the NDNS included in this document were obtained from: NATCEN SOCIAL RESEARCH 2019. National Diet and Nutrition Survey Years 1-9, 2008/09-2016/17. 15th Edition. SN: 6533. UK Data Service, MRC Elsie Widdowson Laboratory.
[2] https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=6533

(1. Informed Consent)

Welcome to the research study!

We are interested in understanding whether interactivity enhances the effectiveness of visualizations based on complex public health data.

For this, the first of two short surveys, you will be asked some questions about your how your work involves communication of research findings, then you will be shown some visualizations and asked to answer some questions about them.
The survey should take you around 10 minutes to complete.

Your responses will be kept completely confidential. Your participation in this research is voluntary.
You have the right to withdraw at any point during the study.

The Principal Investigator of this study, Veronica Tuffrey, can be contacted at Veronica.Tuffrey@city.ac.uk

By clicking the button below, you acknowledge:

Your participation in the study is voluntary. You are at least 18 years of age. You are aware that you may choose to terminate your participation at any time for any reason.

- ◯ I consent, begin the study
- ◯ I do not consent, I do not wish to participate

(2. Participant background)

First some questions about communication of research findings

This section is about your work, to provide contextual background about the survey participants

Where you do obtain research evidence in your day-to-day work?

Please select all your sources from the options below

- ☐ Academic publications (eg journal articles, books, book chapters, dissertations)
- ☐ Professional publications (eg policy briefs, project briefs, manuals, curriculum, training materials)
- ☐ Events and associated documentation (eg conference texts, proceedings, speeches, slide presentations, workshop reports)
- ☐ Evaluations
- ☐ Media (eg website, social media posts, newsletters, bulletins, pamphlets, newspaper articles)
- ☐ Other (please specify below)

Optional comments about your sources of research evidence (and if you selected "Other" in Question 1A, please provide a brief description of the source(s) here)

What are the main barriers to your increased use of academic journals as a source of research evidence?

Please select all options that apply below

☐ Lack of access (physical or electronic) to read articles
☐ Lack of access because of jargon used/ expertise assumed ☐ Cost of access
☐ Lack of time
☐ Lack of trustworthiness / credibility of data and information reported ☐ Information overload
☐ Other (please specify below)

Optional comments about barriers to your increased used of academic journals to access research evidence
(and if you selected "Other" in Question 2 please provide a brief description of the barrier(s) here)

Do/have you report(ed) findings from your own research in academic journals?

◯ YES ◯ NO

Optional brief details about reporting findings from your own research in academic journals

4A Do/have you deliver(ed) formal spoken communication(s) to an audience presenting findings from your own or others' research? (This could be face-to-face or online)

◯ YES ◯ NO

4B If you answered "YES" to Question 4A, please select all contexts from the options below

☐ Routine/guest lecturing at universities
☐ Conference presentations
☐ Presentations to colleagues in the workplace

☐ Other (please specify below)

Optional comments about your oral communication of research findings (and if you selected "Other" in Question 4B please provide a brief description of the context(s) for your oral communication here)

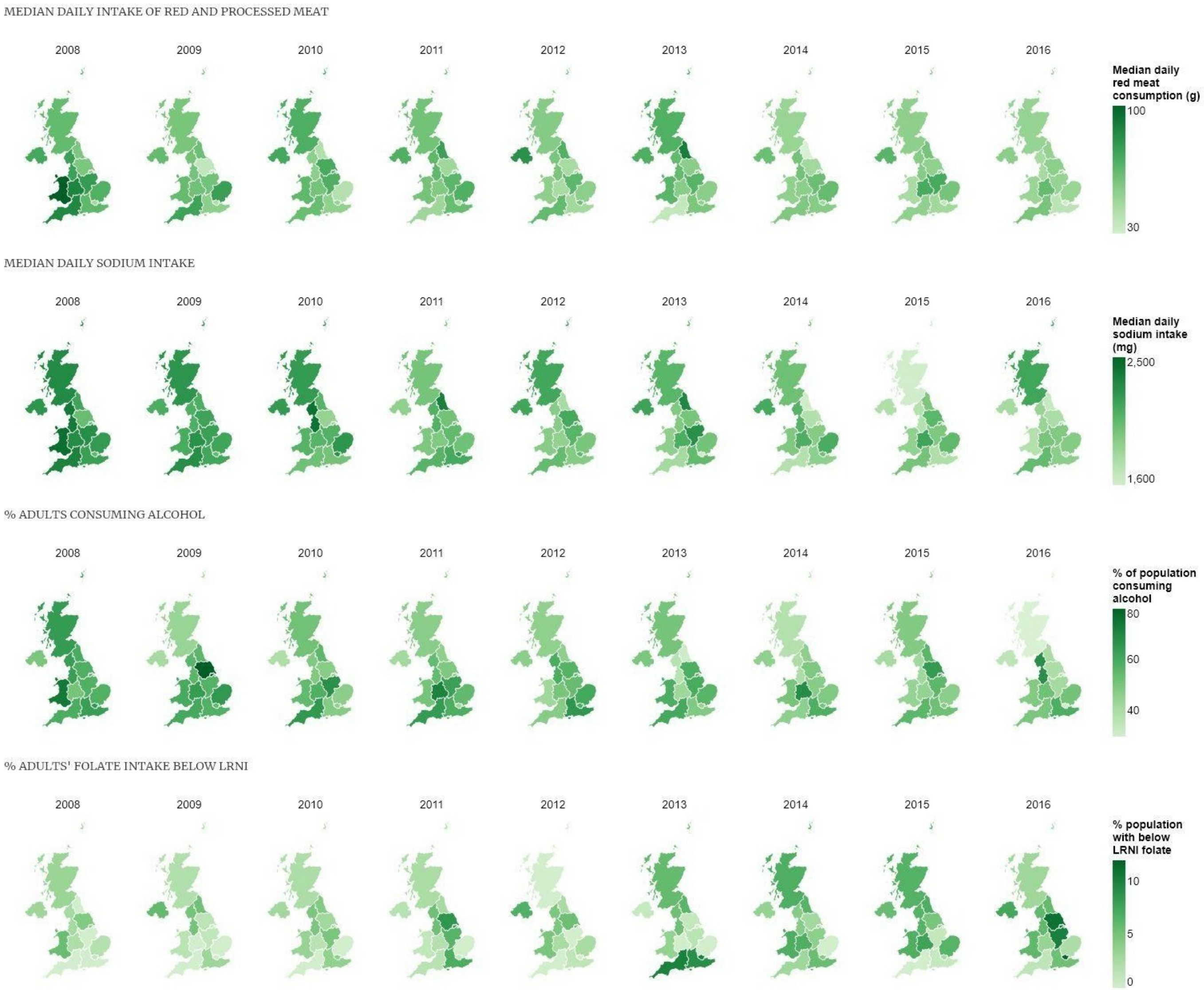
(3. Visualization evaluation)

Now we will move onto the evaluation

The visualizations below can be envisaged as part of online supplementary materials for an article in an academic journal.
In a journal, explanatory context for the supplementary materials would be included in the article's text.
You were sent a one-page document to provide equivalent background information.

The title of the first visualization is:

Maps of the UK showing temporal and geographical variation in four dietary variables between 2008-2016 (data relates to adults aged > 18 y.; sample size for each year in individual regions ranges from 8 to 120, mean 64)



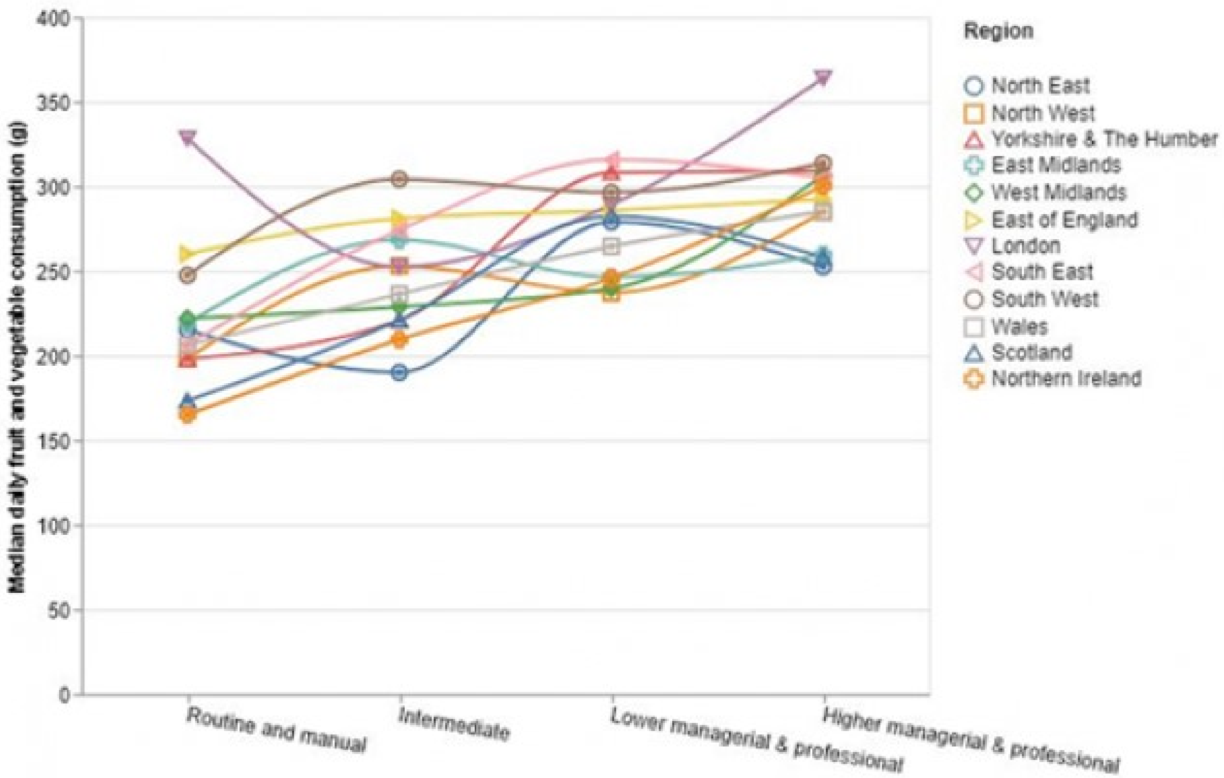
Q5 Please refer to the background document and the maps above to respond to the statements using a scale where 1 = Strongly agree, and 4 = Strongly disagree, and add comments to explain your responses if you wish.

agree	1 Strongly	2	3	4 Strongly disagree	Don't know
5A The maps are superfluous because the important information was already provided in the text	○	○	○	○	○
5B I learnt something about these dietary variables from the maps that I didn't already know from reading the text	○	○	○	○	○
5C I learnt something about these dietary variables from the maps that surprised me (eg it contradicted what I already knew)	○	○	○	○	○
5D I find the maps difficult to interpret because they are badly designed	○	○	○	○	○
5E I find the maps difficult to interpret because I don't have sufficient expertise (please use the box below to specify the type of expertise)	○	○	○	○	○

Optional additional comments about the maps

The title of the second visualization is:

Line chart showing daily fruit and vegetable intake by occupation category and region



Q6 Please refer to the background document and the line chart above to respond to the statements below using a scale where 1 = Strongly agree, and 4 = Strongly disagree, and add comments to explain your responses if you wish

agree	1 Strongly	2	3	4 Strongly disagree	Don't know
6A The line chart is superfluous because the important information was already provided in the text	○	○	○	○	○
6B I learnt something about these dietary variables from the line chart that I didn’t already know from reading the text	○	○	○	○	○
6C I learnt something about these dietary variables from the line chart that surprised me (eg it contradicted what I already knew)	○	○	○	○	○
6D I find the line chart difficult to interpret because it is badly designed	○	○	○	○	○
6E I find the line chart difficult to interpret because I don’t have sufficient expertise (please use the box below to specify the type of expertise)	○	○	○	○	○

Optional additional comments about the line chart

The title of the third visualization is:

Barcharts showing coefficients from regression modelling of fruit and vegetable intake, and red meat intake (normalised dependent variables) with socio-economic and demographic variables (independent variables)

Q7 Please refer to the background document and the bar charts above to respond to the statements below using a scale where 1 = Strongly agree, and 4 = Strongly disagree, and add comments to explain your responses if you wish

agree	1 Strongly	2	3	4 Strongly disagree	Don't know
7A The bar charts are superfluous because the important information was already provided in the text	○	○	○	○	○
7B I learnt something about these dietary variables from the bar charts that I didn't already know from reading the text	○	○	○	○	○
7C I learnt something about these dietary variables from the bar charts that surprised me (eg it contradicted what I already knew)	○	○	○	○	○
7D I find the bar charts difficult to interpret because they are badly designed	○	○	○	○	○
7E I find the bar charts difficult to interpret because I don't have sufficient expertise (please use the box below to specify the type of expertise)	○	○	○	○	○

Optional additional comments about the barcharts

Thank you, that was the last question, NOW PLEASE CLICK THE RED BUTTON ON THE RIGHT TO SUBMIT YOUR RESPONSES otherwise they will not be stored.

(End of survey1 message)

This is the end of the first survey, many thanks for your responses. Please click the link provided to the second survey now.

Powered by Qualtrics

(1. Informed Consent)

Welcome to the second survey of the research study!

For this, the second of the two short surveys, you will be shown some more visualizations and asked some questions about them, and about their effectiveness in comparison with the first set you saw.

The survey should take you around 10 minutes to complete.

Your responses will be kept completely confidential.

Your participation in this research is voluntary.

You have the right to withdraw at any point during the study.

The Principal Investigator of this study, Veronica Tuffrey, can be contacted at Veronica.Tuffrey@city.ac.uk

By clicking the button below, you acknowledge:

Your participation in the study is voluntary. You are at least 18 years of age. You are aware that you may choose to terminate your participation at any time for any reason.

- ◯ I consent, begin the study
- ◯ I do not consent, I do not wish to participate

(2. Visualization evaluation)

You will now be shown some more visualizations - these are based on the same data as those you saw in the first survey, but have a different format.

As mentioned earlier, the visualizations can be envisaged as part of online supplementary materials for an article in an academic journal. In a journal, explanatory context for the supplementary materials would be included in the article's text.
You were sent a one-page document to provide equivalent background information.

The title of the first visualization is:

Interactive maps of the UK showing temporal and geographical variation in four dietary variables between 2008- 2016 (data relates to adults aged > 18 y.; sample size for each year in individual regions ranges from 8 to 120, mean 64)

**** Please note: Slide button along bar to select years from 2008 to 2016 ****

To access the maps, please open a new tab in your browser and cut and paste this link into the address bar:

https://tufty317.github.io/web/MapViz_forSurvey_interacti

Slide the button along the bar at the base of the chart to select years from 2008 to 2016

Q1 Please refer to the background document and the maps to respond to the statements using a scale where 1 = Strongly agree, and 4 = Strongly disagree, and add comments to explain your responses if you wish.

	1 Strongly agree	2	3	4 Strongly disagree	Don't know
1A The maps are superfluous because the important information was already provided in the text	○	○	○	○	○

1B I learnt something about these dietary variables from the maps that I didn't already know from reading the text

1C I learnt something about these dietary variables from the maps that surprised me (eg it contradicted what I already knew)

1D I find the maps difficult to interpret because they are badly designed

1E I find the maps difficult to interpret because I don't have sufficient expertise (please use the box below to specify the type of expertise)

Optional additional comments about the interactive maps

The title of the second visualization is:

Interactive line chart showing daily fruit and vegetable intake by occupation category and region

**** Please note: Click on line in chart, or double click on shape in legend to select region ****

To access the line chart, please open a new tab in your browser and cut and paste this link into the address bar:

https://tufty317.github.io/web/LineViz_forSurvey_interactiv

Click on line in chart or double click on shape in legend to select region

Q2 Please refer to the background document and the line chart to respond to the statements below using a scale where 1 = Strongly agree, and 4 = Strongly disagree, and add comments to explain your responses if you wish.

	1 Strongly agree	2	3	4 Strongly disagree	Don't know
2A The line chart is superfluous because the important information was already provided in the text	○	○	○	○	○

	1 Strongly agree	2	3	4 Strongly disagree	Don't know
2B I learnt something about these dietary variables from the line chart that I didn't already know from reading the text	○	○	○	○	○
2C I learnt something about these dietary variables from the line chart that surprised me (eg it contradicted what I already knew)	○	○	○	○	○
2D I find the line chart difficult to interpret because it is badly designed	○	○	○	○	○
2E I find the line chart difficult to interpret because I don't have sufficient expertise (please use the box below to specify the type of expertise)	○	○	○	○	○

Optional additional comments about the interactive line chart

The title of the third visualization is:

Barcharts showing coefficients from regression modelling of fruit and vegetable intake, and red meat intake (normalised dependent variables) with socio-economic and demographic variables (independent variables)

**** Please note: Click button in legend to select socio-economic variable ****

To access the bar charts, please open a new tab in your browser and cut and paste this link into the address bar: https://tufty317.github.io/web/BarViz_forSurvey_intera

Click buttons on right side of chart to select socio-economic variables

Q3 Please refer to the background document and the bar charts to respond to the statements below using a scale where 1 = Strongly agree, and 4 = Strongly disagree, and add comments to explain your responses if you wish.

	1 Strongly agree	2	3	4 Strongly disagree	Don't know

3A The bar charts are superfluous because the important information was already provided in the text

3B I learnt something about these dietary variables from the bar charts that I didn't already know from reading the text

3C I learnt something about these dietary variables from the bar charts that surprised me (eg it contradicted what I already knew)

3D I find the bar charts difficult to interpret because they are badly designed

3E I find the bar charts difficult to interpret because I don't have sufficient expertise (please use the box below to specify the type of expertise)

Optional additional comments about the interactive barcharts

(3. Comparison)

Now moving on to compare the static versions (in the first survey) and interactive versions (in this survey) of the visualizations.

If you leave comments, please refer to specific visualizations.

Did you gain insights from the visualisations that were additional to those gained from the information provided in the one page text?

- ○ Yes I gained many additional insights
- ○ Yes I gained a moderate number of additional insights
- ○ Yes I gained a few additional insights
- ○ No I didn't gain any additional insights
- ○ Don't know

Optional additional comments about the value of the visualizations

If you answered yes to Question 4, did you gain the additional insights more quickly from the static or interactive versions?

- ○ It was quicker with the static versions
- ○ It was quicker with the interactive versions
- ○ There was no difference in speed between static and interactive versions
- ○ The difference varied according to the type of visualization, ie maps, lines or bars
- ○ Don't know

If you answered yes to Question 4, was it easier to gain the additional insights from the static or interactive versions?

- ○ It was easier with the static versions
- ○ It was easier with the interactive versions
- ○ The effort needed was the same for the static and interactive versions
- ○ The difference varied according to the type of visualization, ie maps, lines or bars
- ○ Don't know

Optional additional comments about the effort or speed of gaining insights

To what extent was the facility to interact with the visualizations helpful in increasing your understanding of the researcher's findings described in the one-page text?

- ○ Interactivity was very helpful
- ○ Interactivity was moderately helpful
- ○ Interactivity was slightly helpful
- ○ Interactivity was not helpful
- ○ Don't know

To what extent was the facility to interact with the visualizations helpful in communicating and supporting the researcher's findings described in the one-page text?

- ○ Interactivity was very helpful
- ○ Interactivity was moderately helpful
- ○ Interactivity was slightly helpful
- ○ Interactivity was not helpful
- ○ Don't know

Did the visualizations engage you?

- ○ Yes, but the static visualizations were more engaging than the interactive
- ○ Yes, but the interactive visualizations were more engaging than the static
- ○ Yes, and I found the static and interactive visualizations were equally engaging
- ○ No, overall I didn't find the visualizations were engaging
- ○ Don't know

Optional additional comments about the value or drawbacks of the interactive facility

(4. Discussion)

Finally, if you have time, I would be grateful for your views on visualizations in academic journals.
This will feed into the discussion section of my dissertation, and help inform recommendations for future work.

Please read the following statements and comment on either or both of them

"Several strategies exist for reducing dimensionality of information displays *(eg using interaction to alter graphical layout, as explored in this survey)* and thereby reducing the displays' complexity - Academic journals should promote these strategies to their authors."

"The current format of academic articles obliges researchers to present findings using a small number of 'explanatory' visualizations – Journals could also request authors provide (whenever possible) an interactive interface in their articles' supplementary materials, where audiences could explore summarised data and obtain personalised insights."

Optional comments in response to the statements above, or about other aspects of graphics in academic journals

Thank you, that was the last question, NOW PLEASE CLICK THE RED BUTTON ON THE RIGHT TO SUBMIT YOUR RESPONSES otherwise they will not be stored.

(Thanks and invitation)

This is the end of the second survey.
Many thanks for your participation in the research study.

Please provide your preferred email address if you would like to be sent a copy of the relevant sections of the dissertation when available.

(End of Survey 2 message)

Thanks again
Veronica Tuffrey
(Veronica.Tuffrey@city.ac.uk)

Powered by Qualtrics

L Supplementary tables

		Occupation		Income		Education	
		Abs diff	Rel Diff (%)	Abs diff	Rel Diff (%)	Abs diff	Rel Diff (%)
Fruit and vegetables	Intake (g)	86.3	32.6	79.0	30.1	119.5	47.1
Achieve 5 a day		0.169	54.3	0.139	44.6	0.231	80.3
Red meat	Intake (g)	-6.904	-12.0	3.580	6.1	-10.3	-17.6
Oily fish	Consume	0.173	57.8	0.156	54.0	0.152	54.7
Soft drinks	Consume	-0.0018	-0.4	-0.0024	-0.5	0.082	18.3
Total energy	Intake (kcal)	212.9	12.1	227.6	12.7	310.2	17.7
Fat % food energy	Intake as %	1.400	4.0	2.103	6.0	1.437	4.1
SFA % food energy	Intake as %	0.415	3.2	0.490	3.8	-0.167	-1.3
NMES % food energy	Intake as %	-0.802	-7.6	-0.984	-9.2	-0.608	-5.7
Alcohol	Consume	0.242	43.5	0.263	47.8	0.206	38.7
AOAC Fibre	Intake (g)	2.9	16.1	2.6	14.5	3.722	21.2
Sodium	Intake (mg)	72.6	3.6	201.3	9.8	195.4	9.6
Iron	Intake (mg)	1.512	15.0	1.555	15.3	2.062	20.8
Calcium	Intake (mg)	78.7	10.1	75.4	9.7	65.9	8.5
Folate	Intake (µg)	41.9	17.6	35.4	14.9	43.8	18.9
Vitamin A	Intake (retequiv)	186.4	25.1	168.6	23.1	159.0	22.1
Vitamin D	Intake (µg)	0.238	9.7	0.428	17.6	0.259	10.8

Table A16: Absolute and relative inequality values for 17 dietary variables

		Occupation			**Income**			**Education**		
		Abs diff	**Rel Diff (%)**	**p**	**Abs diff**	**Rel Diff (%)**	**p**	**Abs diff**	**Rel Diff (%)**	**p**
Iron	Below LRNI	-0.067	-65.6		-0.096	-91.2		-0.025	-22.9	
Calcium	Below LRNI	-0.063	-106.2		-0.053	-91.5		-0.048	-72.1	
Folate	Below LRNI	-0.029	-92.9		-0.032	-105.1		-0.038	-107.9	
Vitamin A	Below LRNI	-0.064	-83.6		-0.071	-86.4		-0.055	-63.2	

Table A17: Absolute and relative inequality values for four micronutrient variables

			Fruit and Veg (g/day)				Red meat (g/ day)				NMES as % of food energy			
		N (adjusted)	Median	Mean	Lower CI	Upper CI	Median	Mean	Lower CI	Upper CI	Median	Mean	Lower CI	Upper CI
All sample		6802	256	291	285	297	57.0	65.4	63.7	67.1	10.7	11.7	11.5	11.9
Occupation	Routine and manual	2214	217	252	243	262	60.0	69.1	66.0	72.2	11.0	12.2	11.8	12.6
	Intermediate	1371	255	283	272	295	60.8	68.4	64.6	72.2	10.7	11.5	11.1	12.0
	Lower manag. & prof.	1703	283	312	300	323	56.0	63.8	61.0	66.5	10.6	11.5	11.1	11.9
	Higher manag.& prof.	1193	304	343	326	360	53.1	60.6	57.0	64.2	10.2	11.1	10.7	11.6
	All	*6481*		*291*	*285*	*297*		*66.0*	*64.3*	*67.7*		*11.7*	*11.5*	*11.9*
R squared				0.042				0.003				0.002		
p				*<0.001*				*0.01*				*0.065*		
Income	Lowest quartile	1773	221	259	248	270	55.9	62.0	59.0	65.0	11.3	12.4	11.9	12.8
	Second quartile	1227	253	286	273	299	60.8	68.7	64.7	72.8	10.5	11.5	11.0	12.0
	Third quartile	1411	276	314	301	326	59.7	68.7	65.2	72.2	10.8	11.6	11.2	12.0
	Highest quartile	1333	300	329	314	345	59.5	66.9	63.6	70.1	10.3	11.1	10.7	11.5
	All	*5744*		*294*	*288*	*301*		*66.2*	*64.5*	*68.0*		*11.7*	*11.5*	*11.9*
R squared				0.029				0.004				0.003		
p				*<0.001*				*0.007*				*0.017*		
Education	No qualifications	1323	208	241	231	251	61.2	67.8	64.4	71.3	10.5	11.5	11.1	11.9
	GCSE or equivalent	1382	229	259	248	271	63.2	70.9	67.2	74.5	11.3	12.6	12.1	13.0
	A level or equivalent	1650	251	283	272	294	58.0	67.5	64.2	70.8	11.3	12.0	11.6	12.5
	Degree or higher	1797	327	357	344	370	50.9	59.1	56.1	62.1	9.9	10.8	10.4	11.2
	All	*6152*		*290*	*284*	*297*		*65.9*	*64.1*	*67.6*		*11.7*	*11.5*	*11.9*
R squared,				0.070				0.010				0.010		
p				*<0.001*				*<0.001*				*<0.001*		

Table A18: Absolute values for 3 key continuous dietary variables by SES categories with 95% confidence intervals

R squared and p are from regression modelling of transformed variables using the complex samples option in SPSS

			Folate (proportion of sample below LRNI)			Alcohol (proportion of sample consumed)			Oily Fish (proportion of sample consumed)		
		N (adjusted)	Estimate	Lower CI	Upper CI	Estimate	Lower CI	Upper CI	Estimate	Lower CI	Upper CI
All sample		6802	.0348	.0292	.0404	.531	.514	.548	.284	.270	.299
Occupation	Routine and manual	2214	.0457	.0353	.0561	.418	.392	.444	.203	.182	.225
	Intermediate	1371	.0387	.0260	.0514	.521	.486	.556	.270	.241	.300
	Lower manag. & prof.	1703	.0228	.0133	.0323	.628	.598	.658	.349	.319	.379
	Higher manag.& prof.	1193	.0169	.0074	.0264	.660	.625	.695	.377	.342	.412
	All	*6481*	*.0329*	*.0272*	*.0386*	*.539*	*.523*	*.556*	*.288*	*.273*	*.303*
Cox &Snell R^2			0.005			0.039			0.024		
p			*0.001*			*<0.001*			*<0.001*		
Income	Lowest quartile	1773	.0533	.0407	.0659	.409	.379	.439	.211	.187	.235
	Second quartile	1227	.0290	.0181	.0399	.503	.467	.539	.244	.211	.276
	Third quartile	1411	.0182	.0091	.0274	.616	.582	.649	.338	.306	.370
	Highest quartile	1333	.0213	.0103	.0323	.672	.641	.703	.367	.335	.399
	All	*5744*	*.0321*	*.0262*	*.0379*	*.541*	*.523*	*.559*	*.285*	*.269*	*.301*
Cox &Snell R^2			0.007			0.044			0.021		
p			*<0.001*			*<0.001*			*<0.001*		
Education	No qualifications	1323	.0593	.0447	.0740	.419	.385	.452	.229	.202	.256
	GCSE or equivalent	1382	.0305	.0189	.0422	.517	.482	.552	.229	.200	.258
	A level or equivalent	1650	.0313	.0208	.0418	.574	.543	.605	.275	.248	.302
	Degree or higher	1797	.0210	.0111	.0309	.625	.595	.656	.381	.351	.411
	All	*6152*	*.0341*	*.0283*	*.0400*	*.543*	*.525*	*.560*	*.286*	*.270*	*.301*
Cox &Snell R^2			0.005			0.023			0.020		
p			*<0.001*			*<0.001*			*<0.001*		

Table A19: Absolute values for 3 key binary dietary variables by SES categories with 95% confidence intervals

R squared and p are from regression modelling of transformed variables using the complex samples option in SPSS

	Region	Occupation		Income		Education	
		Abs diff	Rel Diff (%)	Abs diff	Rel Diff (%)	Abs diff	Rel Diff (%)
Fruit and vegetables Intake (g)	North East	37.2	15.9	39.9	17.4	80.5	36.2
	North West	86.5	35.6	66.7	28.3	128.5	54.5
	Yrks& Hmb	111.2	42.9	71.1	27.8	139.5	56.9
	East Mids	40.2	16.2	69.2	28.0	109.5	43.2
	West Mids	84.7	34.0	76.9	30.4	94.8	39.0
	East Engl	32.5	11.6	83.7	28.2	86.7	30.0
	London	35.6	11.6	31.9	10.0	91.8	31.9
	South East	98.7	35.8	75.0	28.0	91.7	34.9
	South West	66.3	22.8	63.0	21.7	146.6	52.1
	Wales	79.2	31.9	100.7	39.8	130.7	53.5
	Scotland	84.8	36.3	118.4	48.7	157.6	69.4
	N Ireland	135.2	58.7	122.7	54.4	130.7	58.9
Red meat Intake (g)	North East	-5.2	-8.7	3.3	5.7	10.3	17.2
	North West	-1.2	-2.0	2.2	3.5	-6.8	-10.9
	Yrks& Hmb	0.8	1.5	13.3	25.4	-8.5	-15.3
	East Mids	-17.0	-26.6	24.8	36.7	-13.0	-20.8
	West Mids	1.3	2.0	-1.1	-1.7	-19.3	-32.0
	East Engl	-6.1	-10.3	4.7	7.8	-16.5	-28.0
	London	14.3	30.2	17.8	38.3	8.4	17.6
	South East	-6.5	-12.0	-2.6	-4.6	-11.2	-21.4
	South West	-23.8	-41.4	-3.8	-6.3	-7.3	-11.9
	Wales	-1.7	-2.9	-0.1	-0.2	-12.7	-21.7
	Scotland	-20.4	-35.7	7.5	12.8	-13.2	-22.0
	N Ireland	-4.1	-6.1	-3.4	-5.0	5.2	7.5
Oily fish Proportion consume	North East	0.210	84.5	0.151	65.7	0.200	93.9
	North West	0.115	41.3	0.036	13.7	0.076	28.7
	Yrks& Hmb	0.182	68.1	0.163	66.1	0.175	68.9
	East Mids	0.083	30.6	0.091	36.3	0.065	24.5
	West Mids	0.136	53.1	0.049	20.3	0.059	24.3
	East Engl	0.097	27.1	0.179	52.3	0.077	22.1
	London	0.212	57.3	0.211	54.4	0.149	44.2
	South East	0.180	58.7	0.180	60.2	0.154	54.8
	South West	0.171	52.3	0.223	66.8	0.281	90.1
	Wales	0.030	12.6	0.093	36.7	0.092	37.6
	Scotland	0.289	98.1	0.115	44.5	0.254	101.8
	N Ireland	0.258	109.7	0.225	99.4	0.169	87.8
NMES %	North East	-0.4	-3.9	-1.6	-13.7	-0.5	-4.9

food energy	North West	0.4	3.9	0.2	1.8	-0.3	-3.2
	Yrks& Hmb	-0.7	-6.5	-1.1	-10.7	-1.0	-9.5
	East Mids	0.1	0.5	-0.2	-1.5	-1.6	-13.8
	West Mids	-0.3	-2.5	-3.3	-28.3	-0.4	-3.8
	East Engl	-2.5	-23.8	-2.7	-24.6	0.8	8.0
	London	0.2	1.6	0.4	3.7	-1.1	-11.0
	South East	-1.7	-15.4	-1.6	-14.3	-1.7	-15.4
	South West	-2.9	-27.2	-0.3	-2.8	-0.6	-5.4
	Wales	-0.1	-0.8	-1.2	-10.9	0.4	3.3
	Scotland	0.6	5.6	-1.4	-13.1	0.7	6.6
	N Ireland	-1.1	-11.1	-0.8	-8.5	0.2	2.2
Alcohol	North East	0.411	72.9	0.286	52.9	0.193	40.1
Proportion consume	North West	0.154	28.8	0.210	39.1	0.216	40.5
	Yrks& Hmb	0.168	28.9	0.282	49.2	0.294	52.0
	East Mids	0.290	49.4	0.234	41.4	0.222	38.8
	West Mids	0.300	49.5	0.198	33.9	0.014	2.4
	East Engl	0.177	32.0	0.257	47.5	0.125	22.6
	London	0.309	60.5	0.440	90.1	0.220	46.6
	South East	0.266	45.7	0.294	50.8	0.174	30.8
	South West	0.237	41.2	0.189	32.9	0.292	54.1
	Wales	0.212	37.1	0.191	34.7	0.211	38.6
	Scotland	0.290	57.7	0.159	31.2	0.350	71.8
	N Ireland	0.108	23.0	0.249	50.9	0.253	54.8
Folate	North East	-0.017	-37.2	-0.038	-116.2	-0.091	-216.2
Proportion <LRNI	North West	-0.035	-79.3	-0.048	-104.4	-0.020	-39.6
	Yrks& Hmb	-0.022	-51.5	-0.035	-100.6	-0.080	-165.5
	East Mids	-0.033	-114.9	0.016	65.1	-0.058	-193.4
	West Mids	-0.025	-102.0	-0.005	-20.7	-0.021	-78.3
	East Engl	-0.007	-54.2	-0.037	-301.2	-0.028	-174.5
	London	-0.025	-72.6	-0.030	-77.5	-0.025	-58.3
	South East	-0.032	-84.6	-0.063	-178.5	-0.033	-88.9
	South West	-0.026	-163.4	-0.023	-154.8	-0.006	-38.2
	Wales	-0.031	-100.6	-0.018	-45.3	-0.020	-54.4
	Scotland	-0.063	-210.0	-0.027	-92.6	-0.066	-180.6
	N Ireland	-0.051	-213.3	-0.034	-115.1	-0.038	-116.9

Table A20: Absolute and relative inequality values for six key dietary variables by region

		Region alone		Region with SES and Interaction factor					
				Occupation		Income		Education	
		Wald F	***p***	**Wald F**	***p***	**Wald F**	***p***	**Wald F**	***p***
Fruit & Veg	Region	12.6	*<0.001*	7.3	*<0.001*	9.9	*.000*	7.8	*<0.001*
	SES			55.6	*<0.001*	38.1	*0.000*	93.0	*<0.001*
	Region*SES			1.8	*0.005*	1.1	*0.258*	1.0	*0.411*
Red meat	Region	5.1	*<0.001*	3.1	*<0.001*	3.7	*<0.001*	4.3	*<0.001*
	SES			3.9	*0.009*	5.7	*0.001*	11.6	*<0.001*
	Region*SES			1.5	*0.029*	1.7	*0.009*	0.7	*0.861*
NMES	Region	3.0	*0.001*	2.9	*0.001*	4.2	*<0.001*	3.5	*<0.001*
	SES			2.5	*0.055*	4.1	*0.006*	12.9	*<0.001*
	Region*SES			1.2	*0.210*	1.35	*0.090*	1.0	*0.470*
Folate <LRNI	Region	1.5	*0.13*	-	-	17.6	*<0.001*	1.8	*0.073*
	SES			-	-	6.8	*0.001*	32.3	*<0.001*
	Region*SES			-	-	53.8	*<0.001*	1.1	*0.367*
Alcohol, consume	Region	3.6	*<0.001*	2.8	*0.001*	1.6	*0.092*	2.5	*0.004*
	SES			53.2	*<0.001*	51.6	*<0.001*	32.9	*<0.001*
	Region*SES			1.0	*0.505*	1.24	*0.165*	1.2	*0.185*
Oily fish, consume	Region	6.1	*<0.001*	3.1	*<0.001*	4.3	*<0.001*	4.5	*<0.001*
	SES			32.6	*<0.001*	26.6	*<0.001*	26.9	*<0.001*
	Region*SES			1.3	*0.092*	0.9	*0.614*	1.1	*0.368*

Table A21: Findings from regression modelling of interactions of SES with geographical region for six key dietary variables

Note:
(1) Some values are missing for Folate intake <LRNI because for some regions the highest occupation category (which would normally be used as the reference category) had no individuals with Folate intake <LRNI.
(2) For the interactions, p values close to the threshold value of 0.003 are highlighted

				Occupation				Income				Education			
		Region alone		Region & SES		Region SES & demogs		Region & SES		Region SES & demogs		Region & SES		Region SES & demogs	
		Wald F	p	Wald F	p	Wald F	p	Wald F	p	Wald F	p	Wald F	p	Wald F	p
Fruit & Veg	Region	12.585	<0.001	8.87	<0.001	7.53	<0.001	11.10	<0.001	8.46	<0.001	7.560	<0.001	6.79	<0.001
	SES			46.83	<0.001	47.10	<0.001	28.18	<0.001	32.35	<0.001	75.881	<0.001	99.90	<0.001
	Male					.42	.517			.04	.848			.09	.770
	Agegroup					35.21	<0.001			37.22	<0.001			66.73	<0.001
	EthWhite					28.70	<0.001			32.13	<0.001			12.99	<0.001
Red meat	Region	5.110	<0.001	4.00	<0.001	1.93	.031	4.06	<0.001	2.15	.015	4.159	<0.001	2.42	.006
	SES			2.91	.034	6.95	<0.001	4.51	.004	1.121	.339	8.546	<0.001	7.38	<0.001
	Male					300.46	<0.001			260.96	<0.001			271.39	<0.001
	Agegroup					1.42	.235			1.31	.268			2.19	.088
	EthWhite					74.88	<0.001			62.22	<0.001			61.70	<0.001
NMES	Region	2.964	0.001	2.74	.002	3.02	.001	3.65	<0.001	3.44	<0.001	3.684	<0.001	4.02	<0.001
	SES			2.15	.092	2.30	.076	2.98	.030	4.07	.007	11.090	<0.001	9.24	<0.001
	Male					17.63	<0.001			20.58	<0.001			22.19	<0.001
	Agegroup					29.57	<0.001			25.24	<0.001			30.32	<0.001
	EthWhite					46.19	<0.001			38.83	<0.001			33.72	<0.001
Folate <LRNI	Region	1.491	0.128	1.17	.304	1.12	.338	1.42	.157	1.45	.146	1.236	.257	1.38	.177
	SES			5.30	.001	4.73	.003	7.37	<0.001	5.84	.001	6.824	<0.001	14.43	<0.001
	Male					15.59	<0.001			17.71	<0.001			17.78	<0.001
	Agegroup					3.07	.027			2.32	.074			10.92	<0.001
	EthWhite					2.60	.107			4.04	.045			3.74	.053
Alcohol, consume	Region	3.628	<0.001	2.47	.005	3.76	<0.001	1.42	.156	2.43	.005	2.570	.003	3.72	<0.001
	SES			49.85	<0.001	42.24	<0.001	51.13	<0.001	37.21	<0.001	29.395	<0.001	42.94	<0.001
	Male					58.27	<0.001			50.26	<0.001			61.54	<0.001
	Agegroup					11.09	<0.001			11.59	<0.001			14.13	<0.001
	EthWhite					102.70	<0.001			83.05	<0.001			118.65	<0.001
Oily fish, consume	Region	6.063	<0.001	4.35	<0.001	4.71	<0.001	4.63	<0.001	4.75	<0.001	4.625	<0.001	4.86	<0.001
	SES			29.37	<0.001	33.20	<0.001	21.62	<0.001	28.58	<0.001	21.620	<0.001	49.04	<0.001
	Male					8.41	.004			8.23	.004			7.07	.008
	Agegroup					34.56	<0.001			34.58	<0.001			60.20	<0.001
	EthWhite					.28	.597			.00	.956			2.14	.144

Table A22: Findings from regression modelling of SES with Region, with demographic covariates for six key dietary variables

Note: Highlighted p values have changed from the previous column (see Section 4.1.4.3)

		Occupation		Income		Education	
	SurveyYear	Abs diff	Rel Diff (%)	Abs diff	Rel Diff (%)	Abs diff	Rel Diff (%)
Fruit and vegetables Intake (g)	2008	51.9	18.3	56.7	20.6	97.0	34.7
	2009	90.2	34.3	58.7	23.3	101.7	41.8
	2010	112.5	43.7	123.3	46.6	79.8	31.3
	2011	42.9	16.6	22.8	8.9	44.3	17.6
	2012	77.8	28.9	74.9	26.7	105.0	41.7
	2013	84.5	34.4	89.6	36.1	157.4	63.6
	2014	116.5	43.6	136.6	50.7	183.3	74.8
	2015	133.3	48.3	59.1	22.5	113.6	46.6
	2016	74.2	26.5	100.9	36.6	153.6	57.3
Red meat Intake (g)	2008	-11.4	-16.7	-2.6	-3.7	-11.84	-17.0
	2009	-4.0	-6.9	3.9	6.7	-14.71	-25.2
	2010	-18.2	-31.1	-9.2	-15.2	-14.04	-23.2
	2011	-10.9	-19.8	11.2	19.9	0.27	0.5
	2012	-10.5	-18.1	5.7	9.4	-20.62	-34.8
	2013	3.5	5.8	-2.3	-3.9	-18.36	-32.9
	2014	0.8	1.5	5.1	8.6	0.14	0.2
	2015	5.2	9.8	12.6	23.7	3.58	6.7
	2016	-8.1	-17.1	14.5	28.2	-17.11	-33.5
Oily fish Proportion consume	2008	0.126	43.9	0.161	61.5	0.122	45.3
	2009	0.194	64.6	0.185	64.1	0.134	47.7
	2010	0.141	50.1	0.164	57.1	0.040	15.2
	2011	0.112	41.1	0.051	18.9	0.143	54.7
	2012	0.159	55.6	0.166	60.5	0.122	44.2
	2013	0.231	76.1	0.267	95.8	0.263	96.5
	2014	0.181	56.5	0.133	42.9	0.130	44.1
	2015	0.208	64.9	0.087	27.9	0.129	44.1
	2016	0.183	56.1	0.166	51.2	0.272	93.1
NMES % food energy	2008	0.22	2.0	-1.94	-17.1	0.886	7.7
	2009	-0.77	-6.7	-1.93	-17.6	0.197	1.7
	2010	-1.63	-15.2	0.33	3.1	-1.215	-11.0
	2011	-1.44	-13.4	-0.41	-3.7	0.775	7.0
	2012	1.26	11.3	0.18	1.6	1.767	16.1
	2013	-2.45	-22.0	-1.05	-9.4	-2.093	-18.7
	2014	-1.85	-18.4	-0.07	-0.7	-2.104	-19.9
	2015	-0.70	-6.7	-1.39	-13.1	-0.613	-5.8
	2016	0.55	6.0	-0.70	-7.7	-1.273	-14.0
Alcohol Proportion	2008	0.286	44.7	0.341	54.0	0.354	57.5
	2009	0.232	37.4	0.166	27.5	0.151	24.9

consume	2010	0.222	41.0	0.280	52.6	0.241	47.9
	2011	0.258	45.5	0.282	49.1	0.315	58.8
	2012	0.234	41.2	0.303	53.0	0.187	33.4
	2013	0.217	40.9	0.237	45.2	0.126	24.4
	2014	0.244	46.7	0.275	52.2	0.179	35.4
	2015	0.295	55.4	0.252	49.2	0.165	33.4
	2016	0.249	50.5	0.272	57.9	0.227	49.9
Folate Proportion <LRNI	2008	-0.019	-203.8	-0.035	-232.1	-0.030	-183.7
	2009	-0.015	-109.7	-0.030	-210.7	-0.037	-235.8
	2010	-0.002	-6.6	-0.025	-104.6	-0.013	-44.0
	2011	-0.031	-106.7	-0.052	-182.3	-0.042	-109.1
	2012	-0.014	-68.4	-0.029	-215.0	-0.011	-50.1
	2013	-0.037	-81.0	-0.019	-49.9	-0.091	-183.5
	2014	-0.056	-142.5	-0.018	-48.6	-0.060	-143.5
	2015	-0.053	-136.4	-0.007	-18.0	-0.050	-103.6
	2016	-0.040	-73.5	-0.096	-136.1	-0.039	-60.0

Table A23: Absolute and relative inequality values for six key dietary variables by Survey year.

		Year alone		Year with SES and Interaction factor					
				Occupation		Income		Education	
		Wald F	*p*	Wald F	*p*	Wald F	*p*	Wald F	*p*
Fruit & Veg	Year	0.9	*0.336*	0.7	*0.390*	1.0	*0.312*	0.3	*0.575*
	SES			9.7	*<0.001*	6.6	*<0.000*	11.8	*<0.001*
	Year*SES			0.7	*0.545*	0.4	*0.723*	3.3	*0.020*
Red meat	Year	30.9	*<0.001*	30.5	*<0.000*	4.2	*<0.001*	3.5	*<0.001*
	SES		*(negative*	1.5	*0.207*	4.1	*0.006*	12.9	*<0.001*
	Year*SES		*trend)*	0.4	*0.788*	1.4	*0.090*	1.0	*0.470*
NMES	Year	22.6	*<0.001*	22.2	*<0.001*	25.0	*<0.001*	23.2	*<0.001*
	SES		*(negative*	0.5	*0.675*	0.8	*0.499*	2.7	*0.047*
	Year*SES		*trend)*	0.4	*0.731*	1.2	*0.306*	0.1	*0.961*
Folate <LRNI	Year	21.2	*<0.001*	14.3	*<0.001*	23.9	*<0.001*	20.7	*<0.001*
	SES		*(negative*	0.7	*0.555*	4.3	*0.005*	2.7	*0.046*
	Year*SES		*trend)*	1.4	*0.255*	1.0	*0.404*	0.5	*0.673*
Alcohol, consume	Year	22.0	*<0.001*	23.1	*<0.001*	23.6	*<0.001*	25.4	*<0.001*
	SES		*(positive*	9.9	*<0.001*	14.5	*<0.001*	10.8	*<0.001*
	Year*SES		*trend)*	.1	*0.936*	1.4	*0.237*	1.0	*0.402*
Oily fish, consume	Year	3.2	*0.076*	3.3	*0.069*	4.1	*0.042*	1.2	*0.280*
	SES			8.3	*<0.001*	5.8	*0.001*	2.4	*0.063*
	Year*SES			1.4	*0.256*	0.1	*0.951*	1.8	*0.146*

Table A24: Findings from regression modelling of interactions of SES with Survey year for six key dietary variables

		Year alone		Occupation				Income				Education			
				Year & SES		Year SES & demogs		Year & SES		Year, SES & demogs		Year & SES		Year, SES & demogs	
		Wald F	***p***	**Wald F**	***p***	**Wald F**	***p***	**Wald F**	***p***	**Wald F**	***p***	**Wald F**	***p***	**Wald F**	***p***
Fruit & Veg	Year	0.9	*0.336*	0.9	*0.350*	.31	*.581*	0.8	*0.357*	.19	*.661*	.07	*.796*	.69	*.406*
	SES			52.4	*<0.001*	51.77	*<0.001*	31.9	*<0.001*	36.19	*<0.001*	84.63	*<0.001*	105.94	*<0.001*
	Male					.43	*.512*			.01	*.929*			.10	*.754*
	Agegroup					32.62	*<0.001*			34.84	*<0.001*			65.31	*<0.001*
	EthWhite					51.06	*<0.001*			60.91	*<0.001*			24.56	*<0.001*
Red meat	Year	30.9	*0.001*	33.150	*<0.001*	34.40	*<0.001*	26.310	*<0.001*	28.33	*<0.001*	22.81	*<0.001*	24.27	*<0.001*
	SES		*(negative*	3.456	*.016*	7.07	*<0.001*	4.585	*.003*	1.22	*.300*	9.77	*<0.001*	6.73	*<0.001*
	Male		*trend)*			303.48	*<0.001*			264.84	*<0.001*			271.79	*<0.001*
	Agegroup					1.31	*.269*			1.23	*.299*			2.09	*.099*
	EthWhite					75.62	*<0.001*			63.200	*<0.001*			60.55	*<0.001*
NMES	Year	22.6	*<0.001*	21.082	*<0.001*	19.70	*<0.001*	16.653	*<0.001*	15.27	*<0.001*	19.95	*<0.001*	18.42	*<0.001*
	SES		*(negative*	2.060	*.104*	1.91	*.126*	3.033	*.028*	3.87	*.009*	11.27	*<0.001*	8.50	*<0.001*
	Male		*trend)*			18.58	*<0.001*			22.71	*<0.001*			24.10	*<0.001*
	Agegroup					29.66	*<0.001*			25.14	*<0.001*			29.98	*<0.001*
	EthWhite					46.26	*<0.001*			43.04	*<0.001*			30.61	*<0.001*
Folate <LRNI	Year	21.2	*<0.001*	28.411	*<0.001*	28.05	*<0.001*	22.590	*<0.001*	22.90	*<0.001*	23.24	*<0.001*	24.87	*<0.001*
	SES		*(negative*	5.515	*.001*	5.24	*.001*	9.001	*<0.001*	7.39	*<0.001*	7.70	*<0.001*	15.60	*<0.001*
	Male		*trend)*			16.68	*<0.001*			17.77	*<0.001*			18.38	*<0.001*
	Agegroup					3.22	*.022*			2.42	*.064*			11.59	*<0.001*
	EthWhite					1.39	*.239*			2.74	*.098*			2.52	*.113*
Alcohol, consume	Year	22.0	*<0.001*	24.821	*<0.001*	22.05	*<0.001*	22.332	*<0.001*	20.53	*<0.001*	26.19	*<0.001*	27.790	*<0.001*
	SES		*(positive*	52.015	*<0.001*	46.33	*<0.001*	55.191	*<0.001*	42.15	*<0.001*	30.37	*<0.001*	47.76	*<0.001*
	Male		*trend)*			56.68	*<0.001*			49.10	*<0.001*			61.51	*<0.001*
	Agegroup					10.50	*<0.001*			11.52	*<0.001*			14.18	*<0.001*
	EthWhite					88.92	*<0.001*			72.49	*<0.001*			109.31	*<0.001*
Oily fish, consume	Year	3.2	*0.076*	2.972	*.085*	2.22	*.136*	4.370	*.037*	3.72	*.054*	2.06	*.151*	3.72	*.054*
	SES			33.171	*<0.001*	37.59	*<0.001*	25.219	*<0.001*	32.07	*<0.001*	24.15	*<0.001*	32.07	*<0.001*
	Male					8.43	*.004*			8.79	*.003*			8.79	*.003*
	Agegroup					32.99	*<0.001*			33.37	*<0.001*			33.37	*<0.001*
	EthWhite					1.09	*.297*			3.28	*.071*			3.28	*.071*

Table A25: Findings from regression modelling of SES with Survey year, with demographic covariates for six key dietary variables
Note: Highlighted p values have changed from the previous column (see Section 4.1.5.3)

“When using media for research evidence - I would only use media if it was a reliable source of high quality evidence. e.g. a twitter post from an academic specialising in the field. Newspaper articles would not be considered sufficient evidence in isolation, I would always investigate the original data upon which the article was based.” “Data sources from ONS, Government departments etc” “Professional organisation websites - e.g. BDA, Nutrition Society” “Food Industry reports and Market Research reports”

Table A26: Free text comments about sources of evidence (to accompany Figure 14)

“I was a co author” “I have had difficulty finding funding to publish in open access journals.” “We work with the University of West England who report findings from our projects”

Table A27: Free text comments about writing

“My main experience of presenting data is presenting obesity prevalence stats to local teams. I would present this by age, ethnicity and deprivation.” “I do present at conferences and to colleagues but on the subject of diet and nutrition related to specific population groups I work with e.g. infants and young children or older people. It's not necessarily about my findings”

Table A28: Free text comments about presenting findings

		STATIC	INTER-ACTIVE	STATIC	INTER-ACTIVE	STATIC	INTER-ACTIVE
		MAPS		LINE CHART		BAR CHARTS	
KNOWLEDGE TRANSFER							
The graphics are superfluous because the important information was already provided in the text	4 Strongly disagree	7	10	3	9	2	4
	3	4	2	4	2	3	5
	2	0	1	3	1	5	2
	1 Strongly agree	0	0	1	1	0	0
	Mean rank	*11.8*	*13.1*	*9.9*	*14.7*	*9.2*	*12.7*
	Z, p (N)	*-0.45, 0.66 (11)*		*-2.12, 0.03 (11)*		*-1.41, 0.16 (8)*	
I learnt something about these dietary variables from the graphics that I didn't already know from reading the text	4 Strongly disagree	1	0	3	1	3	2
	3	1	0	1	1	4	5
	2	4	8	4	5	1	3
	1 Strongly agree	5	5	3	6	3	3
	Mean rank	*12.8*	*12.3*	*14.4*	*10.9*	*13.1*	*11.9*
	(N) Z and p	*-0.91, 0.37 (11)*		*-2.04, 0.04 (11)*		*-0.74, 0.46 (11)*	
I learnt something about these dietary variables from the graphics that surprised me (eg it contradicted what I already knew)	4 Strongly disagree	2	2	4	2	4	4
	3	1	2	2	5	2	5
	2	5	4	3	2	1	2
	1 Strongly agree	2	5	0	3	0	1
	Mean rank	*12.9*	*11.3*	*12.9*	*9.6*	*11.6*	*9.1*
	(N) Z and p	*-1.63, 0.10 (10)*		*-1.19, 0.23 (8)*		*-1.34, 0.18 (7)*	
BARRIERS TO INTERPRETATION							
I find the graphics difficult to interpret because they are badly designed	4 Strongly disagree	8	7	3	7	1	6
	3	2	3	1	1	1	0
	2	1	0	5	1	6	5
	1 Strongly agree	0	1	2	3	3	2
	Mean rank	*12.0*	*11.0*	*10.6*	*13.3*	*10.3*	*14.4*
	(N) Z and p	*0.00, 1.00 (9)*		*-2.16, 0.03 (10)*		*-2.17, 0.03 (11)*	
I find the graphics difficult to interpret because I don't have sufficient expertise	4 Strongly disagree	8	9	6	8	1	7
	3	2	2	5	4	4	1
	2	1	0	0	1	2	3
	1 Strongly agree	0	0	0	0	3	2
	Mean rank	*10.9*	*12.1*	*12.3*	*12.7*	*9.7*	*13.8*
	(N) Z and p	*-0.45, 0.66 (9)*		*0.00, 1.00 (11)*		*-1.27, 0.21(10)*	

Table A29: Table of response frequencies to questions about charts in surveys (data illustrated in Figure 17)

Note: Values in italics are Z and p (Z>2.0 and p<0.05 highlighted) from application of Wilcoxen signed ranks test to paired agreement scores, and sample size for paired scores (N differs from number of scores due to missing values)

	STATIC	INTERACTIVE
	MAPS	
Question 2: I learnt something about these dietary variables from the graphics that I didn't already know from reading the text	"Its easier to see the comparisons within England and in the different nations over time"	
	LINE CHART	
Question 1: The graphics are superfluous because the important information was already provided in the text		"The chart is very difficult to interpret. The text could be supported visually but not like this it is unhelpful."
Question 3: I learnt something about these dietary variables from the graphics that surprised me (eg it contradicted what I already knew)		"Useful to see and compare the figures of the meat intakes at different points"
Question 4: I find the graphics difficult to interpret because they are badly designed	• "The clustering and the close proximity of the data points makes the chart hard to interpret" • "The information is not clear as there are too many categories (lines) to look at."	
	BAR CHARTS	
Question 4: I find the graphics difficult to interpret because they are badly designed	"Far too many categories of data to interpret in one graph"	"The overlapping error bars and colours make the graphs difficult to clearly follow/understand"
Question 5: I find the graphics difficult to interpret because I don't have sufficient expertise	"knowledge of coefficients and accronyms (e.g SES)"	• I couldn't tell which groups had higher intakes. What was intercept. What was the Y axis? • Coefficient knowledge

Table A30: Free text comments in response to specific questions about charts

STATIC	INTERACTIVE
MAPS	
• “I preferred this format than the slider and I would find it easier to cut and paste selected years in to powerpoint slides for a presentation.” • “I find 1c difficult to answer because learning took place with first maps” • “The maps were too small and not easy to specify which specific area of the country they relate to especially for those with poor geographical knowledge. For example not easy to separate Scotland from the North on these maps” • “Would be useful to also have recommended and average amounts on the key and be able to zoom in to areas of interest.”	• “The maps are useful as provide more detail than the text on regional variation and trends. As I work with regional teams they are usually more interested in the data for their region rather than a national summary, as they would use the data to plan initiatives targeted at their local populations - so this visualization would allow me to home in on whichever region I am working with.” • “The maps are very useful, they provide more information and allow faster interpretation of the data”. • “Would be better to have specific ranges with specific shades rather than gradual shade difference on a continuous variable in legend. No indication of confidence intervals of differences between regions - suspect overlapping due to small sample numbers - therefore misleading. Unhelpful that the variables/metrics displayed are not consistent ie mean daily grams v % of population - the maps lend themselves better to % population than the former. Alcohol metric is unclear in its definition.” • “Would be useful to zoom into areas. Useful that areas and data amounts appear when hovering over.”
LINE CHART	
• “Including colours on the legend helped identify which region a line related to more easily than the interactive version.” • “there's a lot of lines and not easy to distinguish one line from the other” • “Hard to interpret - too many lines and discrete categories”	• “The graph was too busy. It wasn't clear which colour was which region. It would have been better if I could have selected the regions I was most interested in. It wasn’t clear if the X axis was continuous variable or not.” • “Very surprised by the Scottish data in the chart, information in the chart was more detailed and allows very easy comparison between regions” • “A line chart is not the appropriate method for displaying these data. The chart is very confusing and difficult to interpret. The text could be supported visually but not like this it is unhelpful. No indication of confidence intervals again - which means data could be misleading.” • “the interactive visualisations make a big difference to interpreting the data in a meaningful way and don't require much knowledge or expertise” • “Discrete data points - feels like they

	shouldn't be connected." • "There's lots of information in the chart, but you can focus in on narrow aspects of the data to ease interpretation and understanding of the data"
BAR CHARTS	
• "I lack the skills to interpret regression. I am unclear what it means when it shows the red meat intake in women 0.5 lower than the average and men intake on the middle line. The written description is clearer." • "feel I would need to brush up on stats to interpret the information on the barcharts" • "This one is more difficult to read and small on one screen - if on a small laptop (suddenly the letter after "o" on my keyboard is not working!) and for those with weakened eyesight. Also I am not sure I know what intercept is" • "It is the 3 headings that are confusing. Income education occupation. Differences between 3 plots not clear"	• "They were small with too much detail on them. I would have found it easier to interpret if the y axis was portions of F&V or meat rather than a regression." • "The interactive element of these charts suffers from same problem as before. What are the 3 occupation income and education variables. Something must be missing on axes or explanation and can't make sense of them stand alone" • "A written example of how to interpret these graphs would be useful."

Table A31: General free text comments about the three forms of charts

Comparison of visualizations with text: • "Visualisations were much easier to examine, more engaging and held my attention much longer" Comparison of interactive with static versions: • "It required far more effort to interpret the static data. I spent much longer examining the interactive data and found it more interesting. I became quickly tired and bored of the static data" • "If you had just asked me do i prefer interactive or static data without showing me examples I would have said interactive! But actually the static was more useful to me and how I would go on to use and present the data"

Table A32: Free text comments about insight, and comparison between static and interactive versions with respect to speed and ease

	Value of interactivity
+	“I agree with the second statement, it would be better to encourage authors to provide whenever possible an interactive interface allowing the opportunity for readers to explore more data.”
+	“Interactive options don't always make it simpler to view data. This is true for the bar chart example given. I would welcome journals allowing me to access the full detail of say data by region, if it exists, so that I could extract what I needed.”
+	“I agree with both comments. This would allow faster interpretation of results and also open up scientific information to make it more easily interpreted by non-academics.”
+	“I agree with both of the statements that greater efforts should be made to make the data in journal articles more accessible for busy professionals and members of the public, who have an interest but may lack the prerequisite training to interpret it.” “I agree with the second statement that interactive resources should be placed in the supplementary materials section. This is so as not to overpopulate the main body, which may make the written sections harder to follow.”
	Practical aspects
	“Journal authors may not have the skills or expertise to produce or determine the most appropriate type of visualisations. These should be created based on the target audience's preferences.”
	“Graphics may enhance readers understanding of a paper, but I would question the ability of some academics to provide useful and high-quality interactive graphics unless the journals are to provide support for this. The additional time required may put off some academics in providing this, unless the journal has a large impact factor.”
	“I prefer the first statement as I think that only a few readers will spend time on the supplementary materials. Also be mindful of small fonts and access aspects for those with eyesight limitations.”

Table A33: Free text comments in response to statements about visualizations in academic journals, categorized by theme

Note: The first column uses “+” to indicate general agreement with the statements provided.

M Supplementary charts

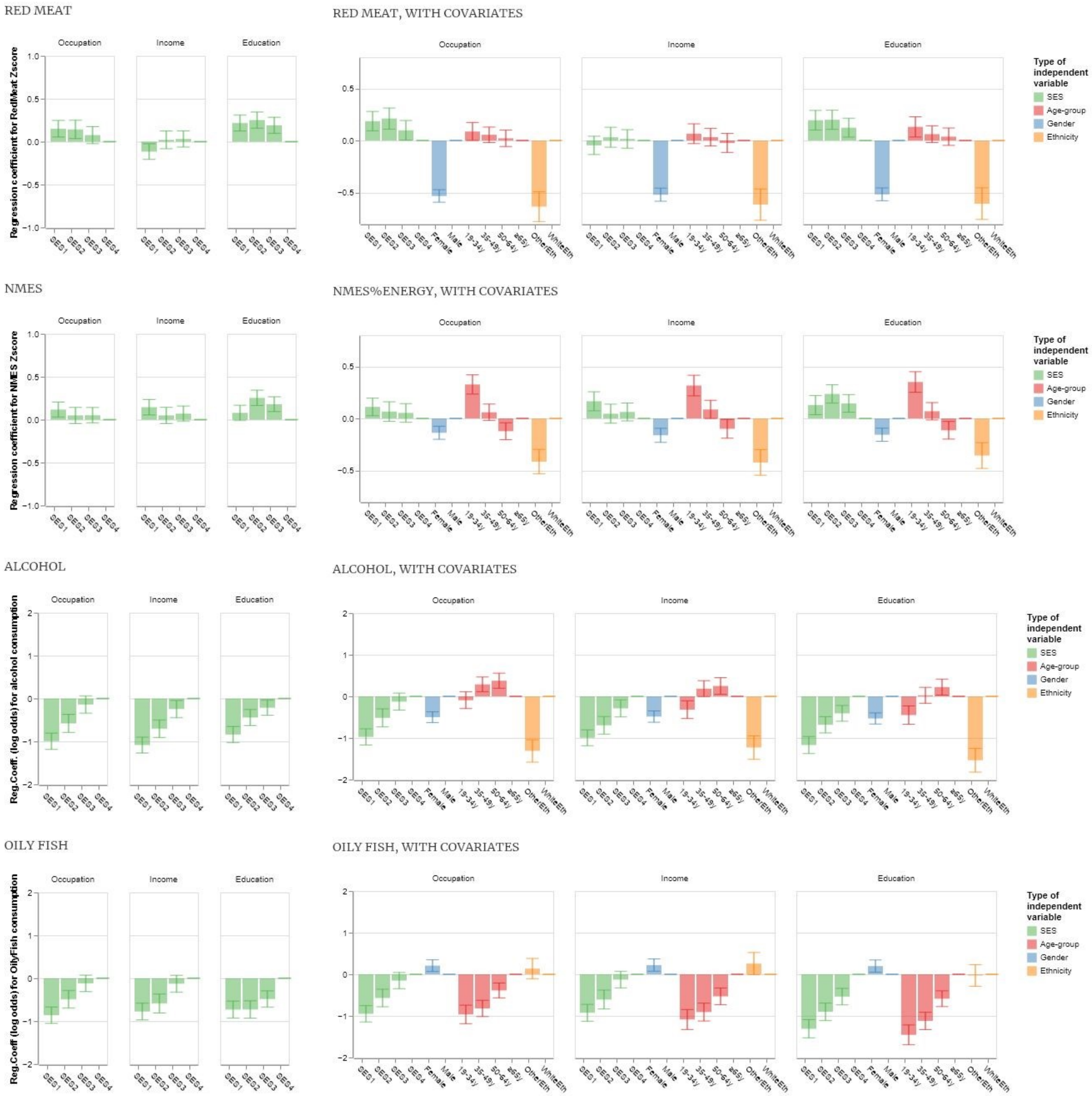


Figure A24: Barcharts of regression coefficients for dietary variables, with socio-economic independent variables only (left), and with demographic covariates added (right)

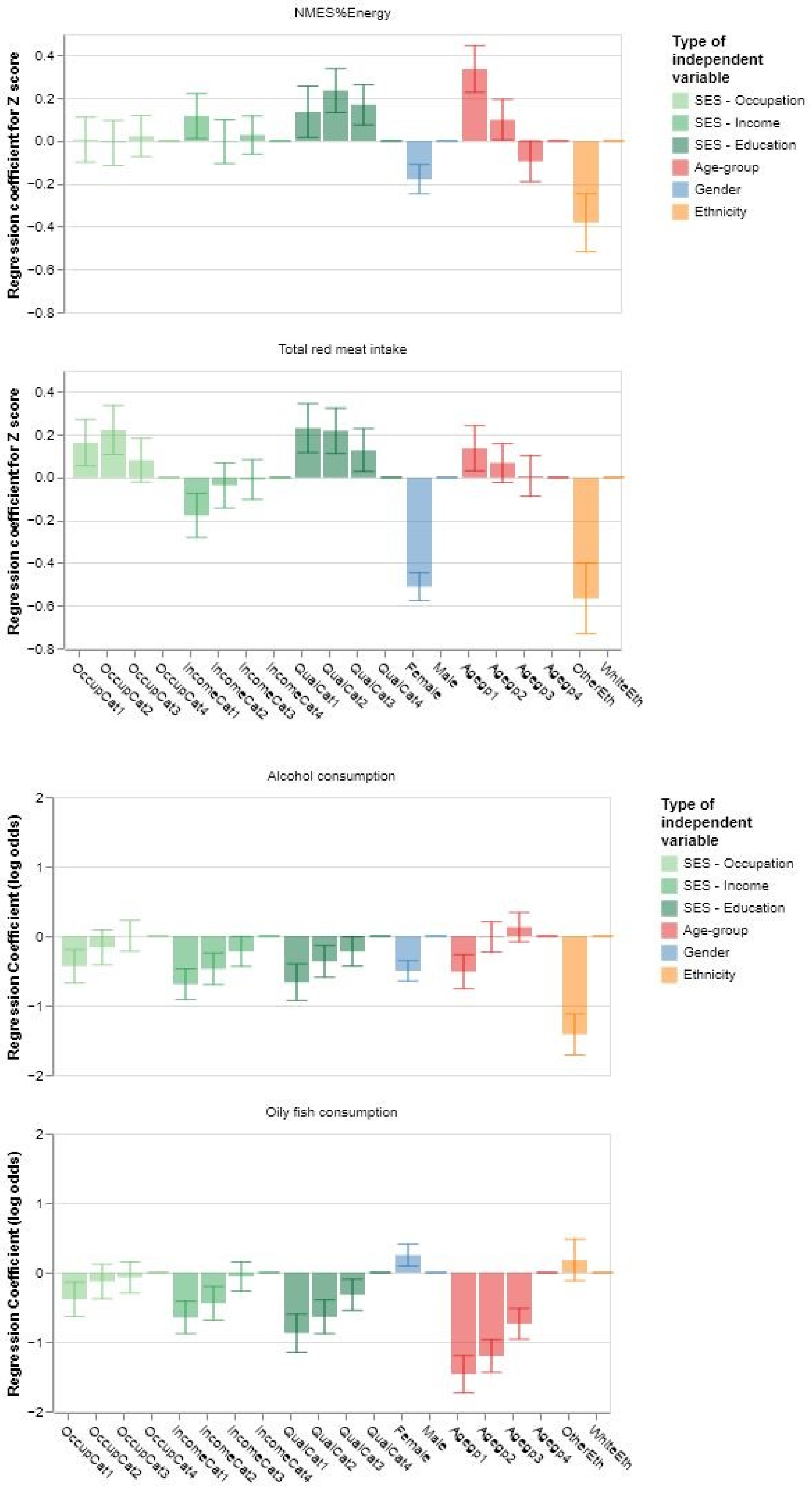


Figure A25: Barcharts of regression coefficients for dietary variables, in models with the three socio-economic independent variables together with demographic covariates


Fruit & Veg (g)
280
260
240
220
5aDay % achieve
35
30
25
20
Fibre (g)
18
17
16
Red Meat (g)
65
60
55
50
Iron (mg)
10.2
10.0
9.8
9.6
9.4
Iron %<LRNI
14
12
10
8
OilyFish % consume
35
30
25
20
VitA (retinol units)
800
750
700
650
VitA %<LRNI
12
10
8
6
4
Alcohol % consume
55
50
45
Folate (microg)
250
240
230
220
210
Folate %<LRNI
4
3
2

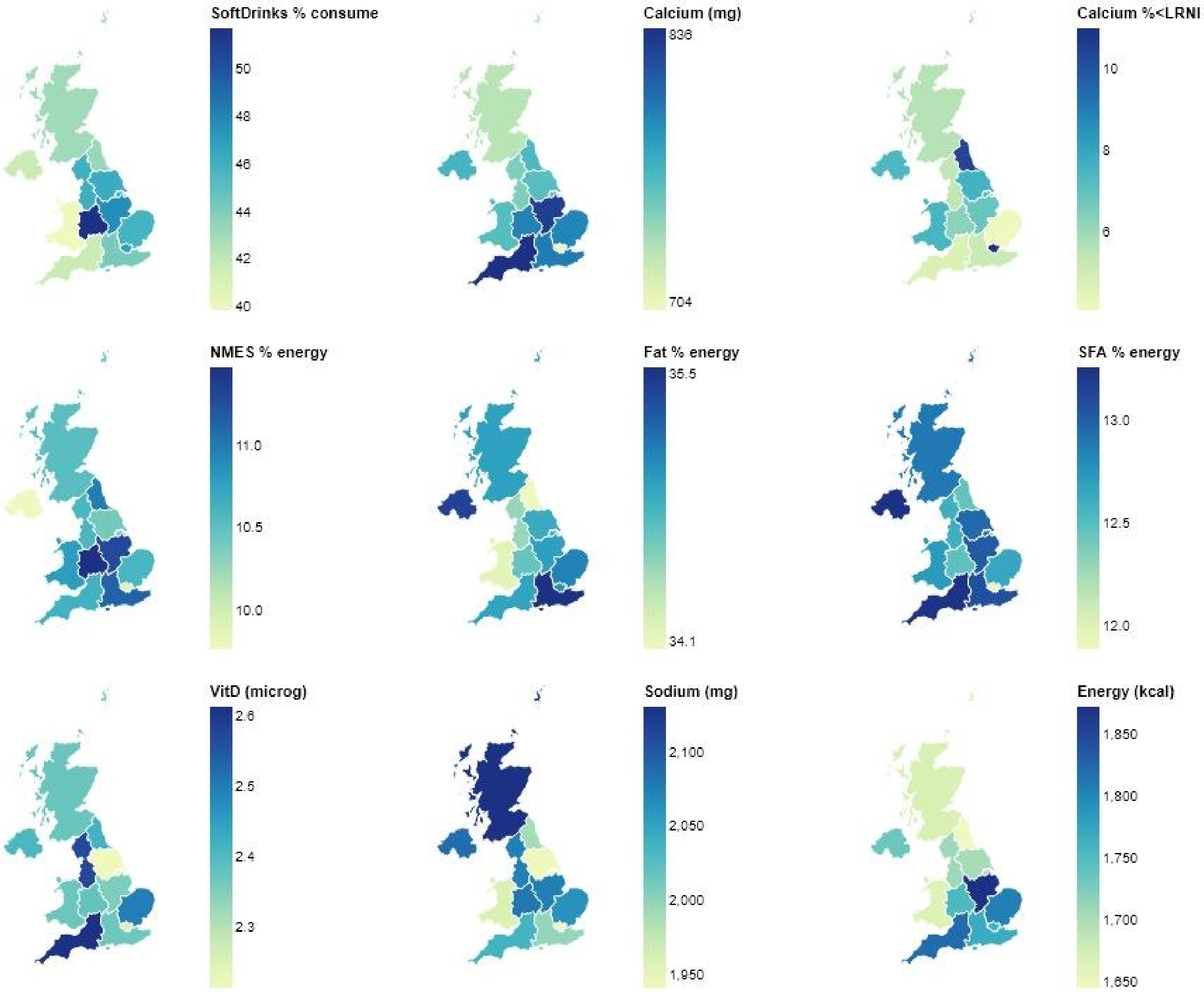
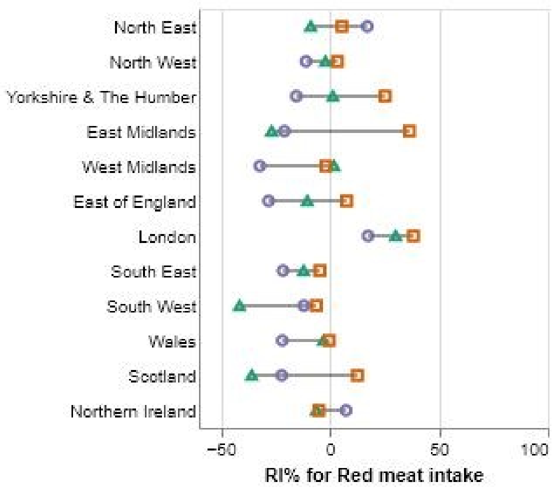
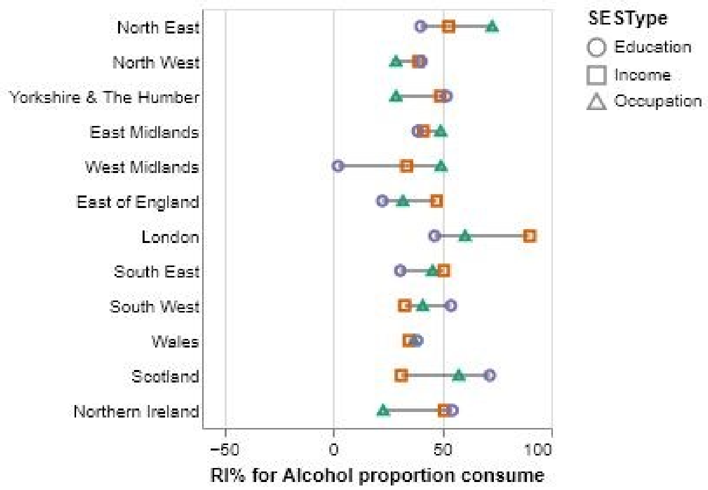
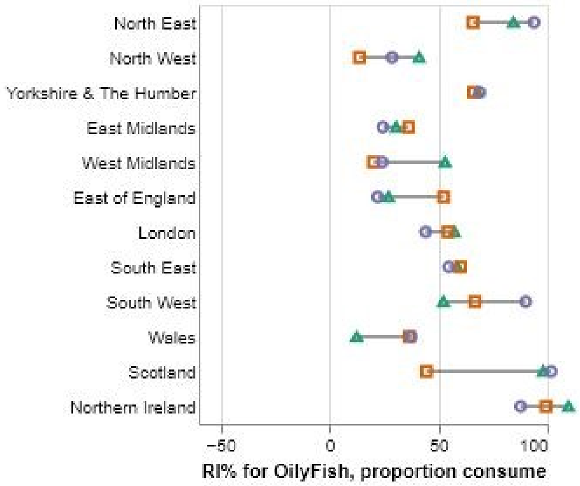


Figure A26: Choropleth maps of median values for 21 dietary variables







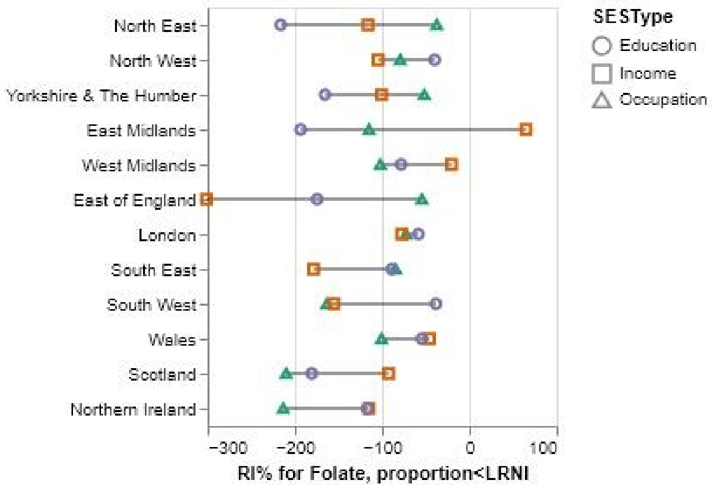
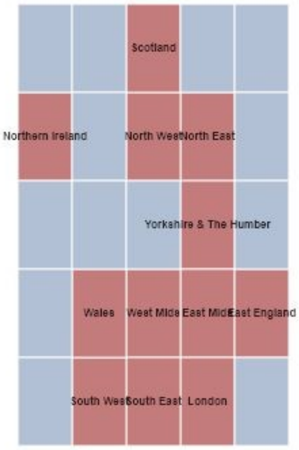


Figure A27: Ranged dotplots of SES inequality in four key dietary variables by geographical region

Note: The scale used for Folate<LRNI differs from that used for the other three variables



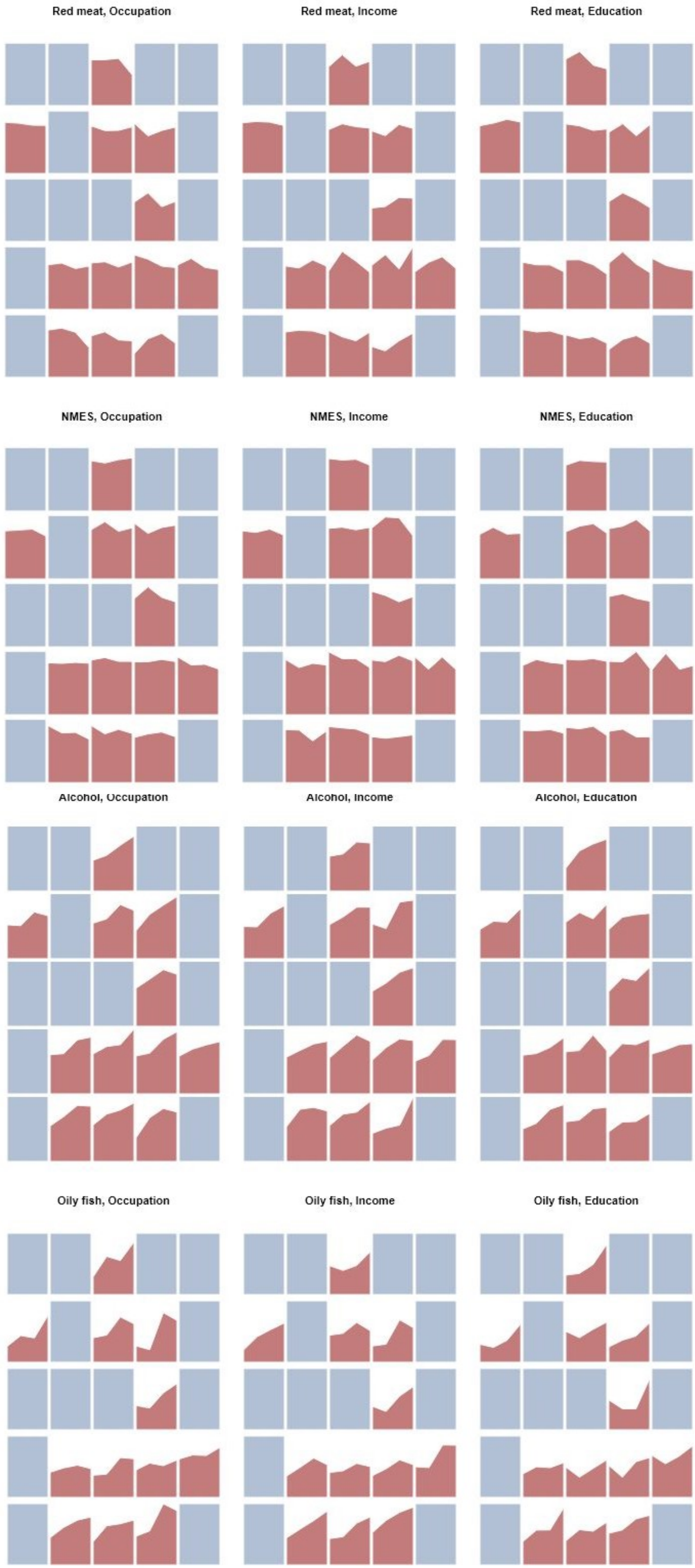


Figure A28: Grid maps showing associations between four dietary variables and SES, with key to regions in top left corner

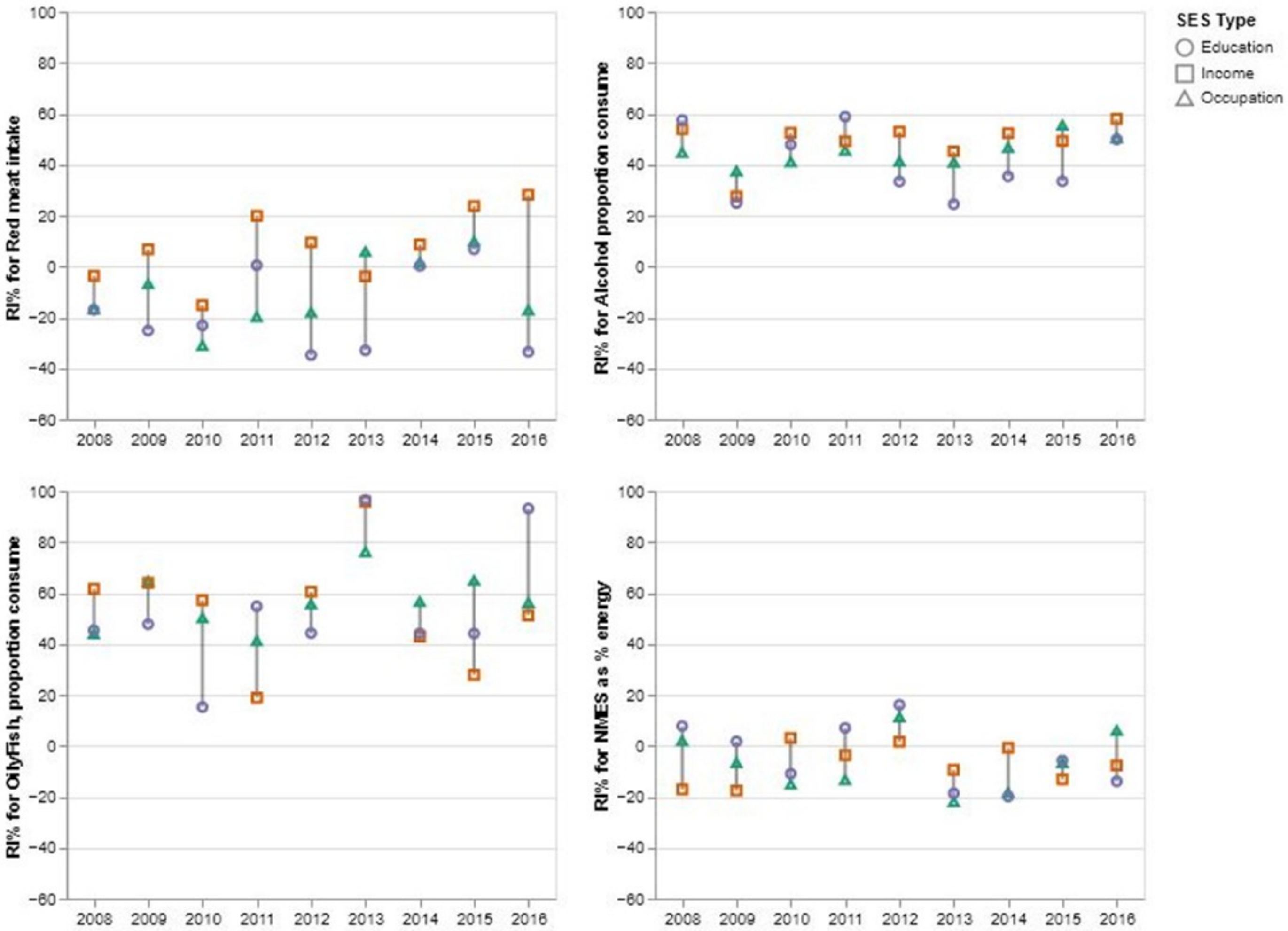
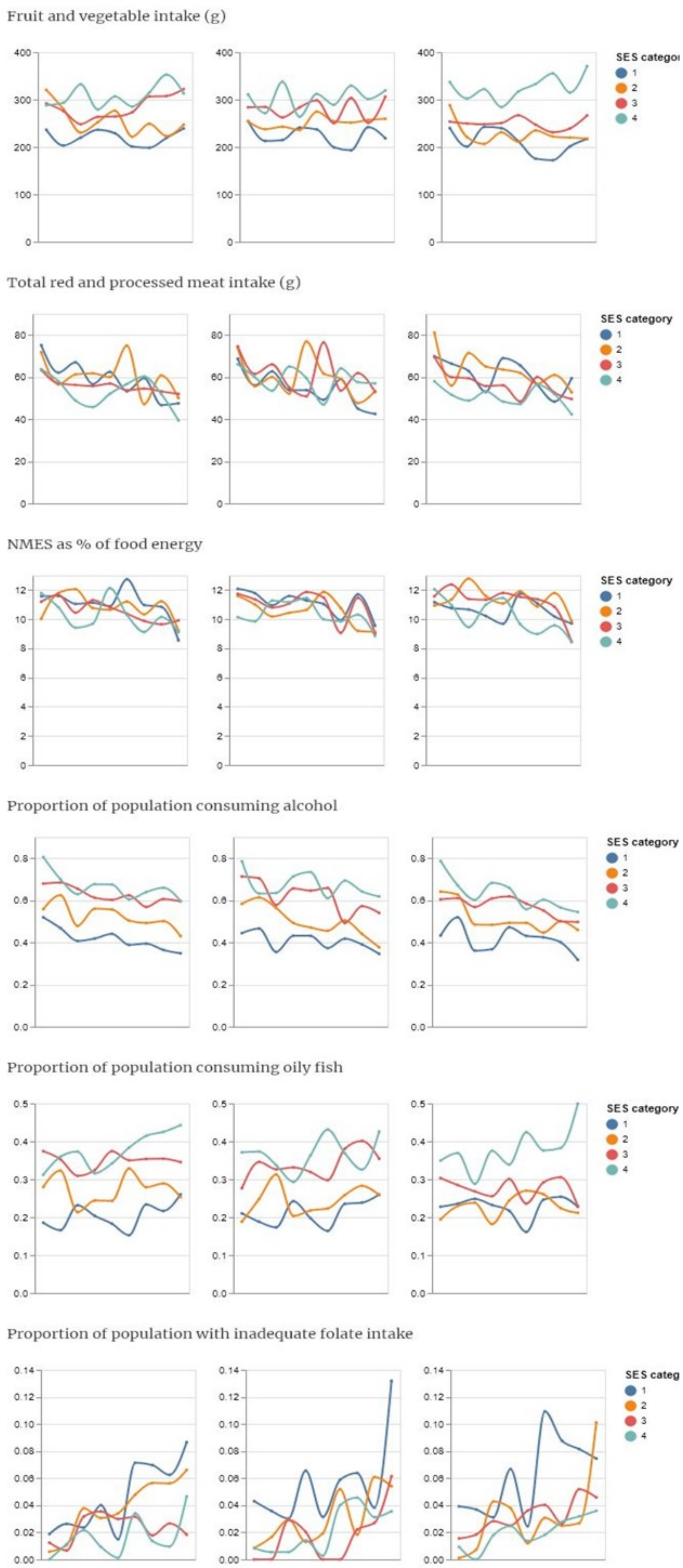


Figure A29: Ranged dotplots of SES inequality in four dietary variables by survey year



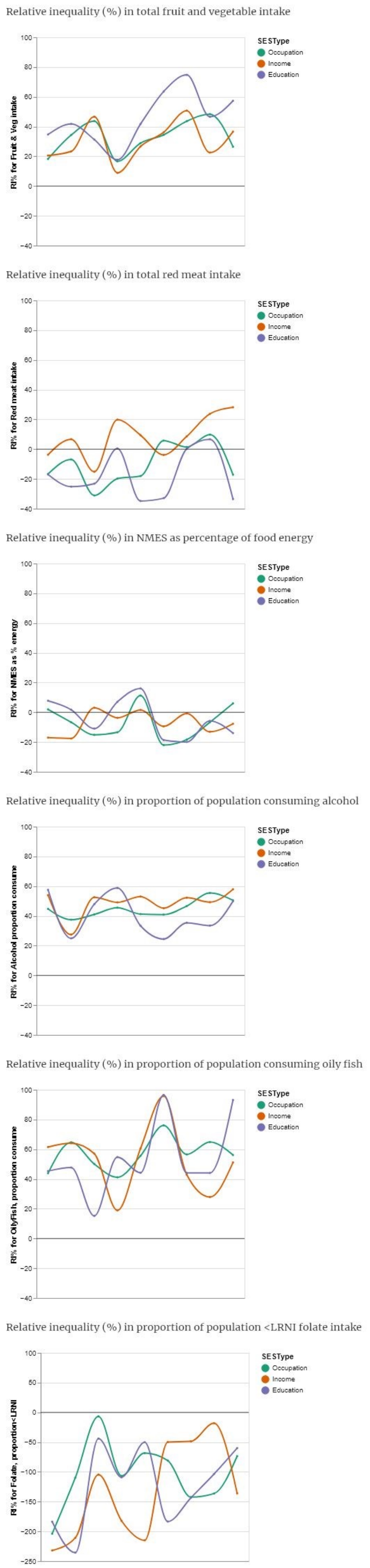


Figure A30 (left): Line plots for 6 key dietary variables by SES category (respectively for occupation, income and education from left to right) by Survey year
Figure 31 (right): Line plots for SES inequality for six key dietary variables by Survey year

CODE AND DATA FILES for VERONICA TUFFREY'S MSc DISSERTATION

1) Files for secondary data analysis component

SPSS DATA FILE	AllYears_adultsOnly_V10.sav
SPSS PLAN FILE	20200812_csa1.csaplan
JSON FILE FOR LITVIS MAPS	NUTS_Level_1__January_2018__Boundaries (6).json

Table or Figure	Content	SPSS syntax	Excel or CSV	Litvis
TABLES IN MAIN REPORT				
Table 12	Descriptive statistics of sample characteristics	20201109_01_ed.sps	n/a	n/a
Table 13	Descriptive statistics for dietary variables	20201109_01_ed.sps	n/a	n/a
Table 14	Heatmap of non-parametric correlation coefficients for dietary v. SES variables	20200922_02.sps	n/a	n/a
Table 15	Findings from regression modelling of 21 dietary variables with region as independent variable	20201203_2.sps	n/a	n/a
TABLES IN APPENDIX L				
Table A16	Dietary inequalities for 17 dietary variables	20201116_01.sps (data)	20201117_1.xlsx (calculations)	n/a
Table A17	Dietary inequalities in four micronutrient variables	20201116_01.sps (data)	20201117_1.xlsx (calculations)	n/a
Table A18	Absolute values for 3 key continuous dietary variables disaggregated by SES categories with 95% confidence intervals	20201111_01_ed.sps (data)	n/a	n/a
		20201120_01.sps (regressions)		
Table A19	Absolute values for 3 key binary dietary variables disaggregated by SES categories with 95% confidence intervals.	20201111_01_ed.sps (data)	n/a	n/a
		20201120_01.sps (regressions)		
Table A20	Dietary inequalities for 6 key dietary variables by region	20200915_06.sps (data for continuous variables)	20201119_7.xlsx (calculations for continuous variables)	n/a
		20200915_07.sps (data for categorical variables)	20201119_8.xlsx (calculations for categorical variables)	n/a
Table A21	Findings from regression modelling of interactions of SES with Region for six key dietary variables	20201118_01.sps	n/a	n/a
Table A22	Findings from regression modelling of SES with Region, with demographic covariates for six key dietary variables	20201207_01.sps	n/a	n/a
Table A23	Dietary inequalities for 6 key dietary variables by Survey year	20200918_01.sps (data for continuous variables)	20201119_1.xlsx (calculations for continuous variables)	n/a
		20200918_02.sps (data for categorical variables)	20201119_3.xlsx (calculations for categorical variables)	n/a
Table A24	Findings from regression modelling of interactions of SES with Survey year for six key dietary variables	20201118_03.sps	n/a	n/a
Table A25	Findings from regression modelling of SES with Survey year for six key dietary variables, with demographic covariates	20201204_01.sps	n/a	n/a
FIGURES IN MAIN REPORT				
Figure 3	Sparklines showing associations between dietary and SES variables	20201116_01.sps (data)	20201116_3.csv	Linecharts_21dietVars_bySEScat_catVars_Spark_noaxes.md (categorical variables)
		20201116_01.sps (data)	20201116_3.csv	Linecharts_21dietVars_bySEScat_contVars_Spark_noaxes.md (continuous variables)
Figure 4	Ranged dotplot of Relative Inequality in 21 dietary variables	See Tables 16 and 17	20201117_3_pos.csv	RelDiffs_21dietVars_bySEScat_withLines.
			20201117_3_neg.csv	
Figure 5	Barcharts of regression coefficients for Fruit & vegetable intake and Folate<LRNI, from models for each SES independent variable separately	20201120_01.sps	20201120_ed.csv	barchart_6Vars_3SESseparately_noCovariates_2vars.md
Figure 6	Barcharts of regression coefficients of Fruit & vegetable intake and Folate<LRNI, from models for each SES independent variable separately, with demographic covariates	20200925_01.sps	20200925_ed2.csv	barchart_6Vars_3SESseparately_2vars.md
Figure 7	Barcharts of regression coefficients for Fruit and vegetable intake, and Folate<LRNI, from models with the three SES variables included simultaneously, with demographic covariates	20200928_01.sps	20200928_V2_ed_2Vars.csv	barchart_6Vars_3SEStogether_F&VandFolate.md
Figure 8	Line plots to examine interactions between demographic and socio-economic variables for Fruit&Vegetable intake (upper and centre) and proportion consuming Alcohol (lower plot).	20201123_02.sps (to identify significant interactions)	n/a	n/a
		20201123_04.sps (to plot folate)	n/a	n/a
		20201123_05.sps (to plot fruit and veg.)	n/a	n/a
Figure 9	Choropleth maps of median values for Fruit and Vegetable intake, and NMES as % of food energy	20200915_02_V2.sps	20200915_1A_V3.csv	Abs_2VarsOnly_Regions_repeatMaps.md
Figure 10	Ranged dotplots of Relative Inequality in Fruit and Vegetable intake, and in NMES as % of food energy, by geographical region	See Table 20	20201119_8_V2.csv	RelDiffs_2dietVars_byGOR_withLines.md
Figure 11	Grid maps showing associations between two dietary variables and SES, with key to regions in top left corner	20200915_07.sps (data for categorical variables)	20200915_3_ed.csv	SES_3CatVars_Regions_separateGrid.md
		20200915_06.sps (data for continuous variables)	20200915_3_ed.csv	SES_3ContVars_Regions_separateGrid.md
		n/a	GORgrid_text.csv	SES_Regions_textGrid.md
Figure 12	Sparklines showing associations between 21 dietary variables and Survey year	20200915_02_V2.sps	20200915_1B_V3.csv	Linecharts_21dietVars_byYear_catVars_spark_noaxes.md (categorical variables)
		20200915_02_V2.sps	20200915_1B_V3.csv	Linecharts_21dietVars_byYear_contVars_spark_noaxes.md (continuous variables)
		20201203_03.sps (regressions)		
Figure 13	Ranged dotplots of Relative Inequality in Fruit and Vegetable intake, and proportion sample with Folate < LRNI, by Survey year	See Table 23	20201119_6_V2.csv	RelDiffs_2dietVarsSameScale_byYear_Vertical.md
FIGURES IN APPENDIX M				
Figure 25A	Barcharts of regression coefficients for dietary variables, with socio-economic independent variables only (left), and with demographic covariates added (right)	20201120_01.sps	20201120_ed.csv	barchart_6Vars_3SESseparately_noCovariates_4vars.md
		20200925_01.sps	20200925_ed2.csv	barchart_6Vars_3SESseparately_4vars.md
Figure 26A	Barcharts of regression coefficients for dietary variables, in models with the three socio-economic independent variables together with demographic covariates	20200928_01.sps	20200928_V2_ed_2Vars.csv	barchart_6Vars_3SEStogether_2_4Vars.md
Figure 27A	Choropleth maps of median values for 21 dietary variables	20200915_02_V2.sps	20200915_1A_V3.csv	Abs_18Vars_Regions_repeatMaps_first12
		20200915_02_V2.sps	20200915_1A_V3.csv	Abs_18Vars_Regions_repeatMaps_second
Figure 28A	Ranged dotplots of SES inequality in four key dietary variables by geographical region	See Table 20	20201119_8_V2.csv	RelDiffs_4dietVarsforAppendix_byGOR_withLines.md
Figure 29A	Grid maps showing associations between four dietary variables and SES, with key to regions in top left corner	20200915_07.sps (data for categorical variables)	20200915_3_ed.csv	SES_3CatVars_Regions_separateGrid.md

		20200915_06.sps (data for continuous variables)	20200915_3_ed.csv	SES_3ContVars_Regions_separateGrid.md
		n/a	GORgrid_text.csv	SES_Regions_textGrid.md
Figure 30A	Ranged dotplots of SES inequality in four dietary variables by survey year	See Table 23	20201119_6_V2.csv	RelDiffs_4dietVars_byYear_Vertical.md
Figure 31A	Line plots for dietary variables by SES category by Survey year	20200918_01.sps (data for continuous variables)	20200918_1.csv	Linecharts_6dietVars_bySESvals_byYear.md
		20200918_02.sps (data for categorical variables)	20200918_2.csv	n/a
Figure 32A	Line plots for SES inequality for six key dietary variables by Survey year	See Table 23	20201119_6_V2.csv	RelDiffs_6dietVars_byYear.md
FIGURES USED IN SURVEYS				
	Choropleth maps - Static version	20200915_08.sps	20200915_4_year_percent.csv	Maps for survey_facets.md
	Choropleth maps - Interactive version	20200915_08.sps	Maps_data.csv (on github)	Maps for
	Line chart - Static version	20200915_06.sps	Lines_data.csv	LineViz_forSurvey_static_noColourDomain
	Line chart - Interactive version	20200915_06.sps	Lines_data.csv (on github)	LineViz_forSurvey_interactive_nocolordo
	Bar chart - Static version	20200925_01.sps	20200925_ed.csv	BarViz_forSurvey_static.md
	Bar chart - Interactive version	20200925_01.sps	Bars_data.csv (on github)	BarViz_forSurvey_interactive.md

2) Files for visualization design study

Table or figure	CONTENT	SPSS SYNTAX FILE	SPSS DATA FILE	
TABLE IN APPENDIX L				
Table A29	Response frequencies to questions about charts in surveys	20210110_1 (for Wilcoxen matched pair test)	Survey_AABB1122_Jan+08+2020_anon.sav	
		20210108_1.sps (for creating long version of datafile)	Survey_AABB1122_Jan+08+2020_anon.sav	
		2021019_1 (for values)	Survey_AABB1122_Jan+08+2020_LONG_anon.sav	
FIGURES IN MAIN REPORT				
Figure 14	Bar chart of responses to "Where do you obtain research evidence in your day-to-day work?"	20210108_3.sps	Survey_AABB1122_Jan+08+2020_anon.sav	
Figure 15	Bar chart of responses to "What are the main barriers to your increased use of academic journals as a source of research evidence?"	20210108_3.sps	Survey_AABB1122_Jan+08+2020_anon.sav	
Figure 16	Bar chart of responses to "Do/have you deliver(ed) formal spoken communication(s) to an audience presenting findings from your own or others' research?"	20210108_3.sps	Survey_AABB1122_Jan+08+2020_anon.sav	
Figures 17a and b	Divergent bar charts of responses to visualization evaluation questions about knowledge transfer (a) and barriers to interpretation (b)	20210108_1.sps (for creating long version of datafile)	Survey_AABB1122_Jan+08+2020_anon.sav	
		20210109_1.sps (for values)	Survey_AABB1122_Jan+08+2020_LONG_anon.sav	
		Charts were created using drop-down menus; 20210109_3.sps has example of code for one chart, copied from output file	Survey_AABB1122_Jan+08+2020_LONG_anon.sav	
Figure 18	Bar charts of differences in mean ranks of responses to static and interactive charts	20210211_1.sps (to obtain ranks)	Survey_AABB1122_Jan+08+2020_LONG_anon.sav	
		20210111_2.sps (Bar charts)	VizEval_long_diff.sav	
Figure 19	Bar chart showing responses to question "Did you gain insights from the visualisations that were additional to those gained from the information provided in the one-page text?"	Charts were created using drop-down menus; 202109_4.sps has code copied from output file	Survey_AABB1122_Jan+08+2020_anon.sav	
Figure 20	Bar chart of responses to question "Did you gain the additional insights more quickly from the static or interactive versions?"	As above	As above	
Figure 21	Bar chart of responses to question "Was it easier to gain the additional insights from the static or interactive versions?"	As above	As above	
Figure 22	Barchart of responses to question "To what extent was the facility to interact with the visualizations helpful in increasing your understanding of the researcher's findings described in the one-page text?"	As above	As above	
Figure 23	Barchart of responses to question "To what extent was the facility to interact with the visualizations helpful in communicating and supporting the researcher's findings described in the one page text?"	As above	As above	
Figure 24	Barchart of responses to question "Did the visualizations engage you?"	As above	As above	