MSIN0164 Data Visualisation - Group Assignment
Food Consumption Landscape in the UK

Jupyter Notebook for static visualisations:
https://drive.google.com/file/d/1LSQ3UQ9Xb8Hz-oC9YRs_tfIYbRabU-wg/view?usp=sharing

Github repository for interactive visualisations:
https://github.com/rasyidanr/data_visualisation

Table of Contents

Word count: 4000

1.0 Formulating the brief

1.1 Curiosity: Why This Topic

After our initial group meeting, we came to the agreement that we did not want to limit our dataset search to only topics that the team found interesting, instead we wanted to create something that is impactful and relevant. We also agreed to not limit our dataset search to specific sources, such as Kaggle and gov.uk. Therefore, to give our group a direction on what to look for, our team formulated the following search prompts.

1.1.1 Recency and relevance towards today's society

This prompt essentially encourages the team to get inspired by current news and become aware of what is going on in the rest of the world. The following are examples of where the team's inspirations came from:

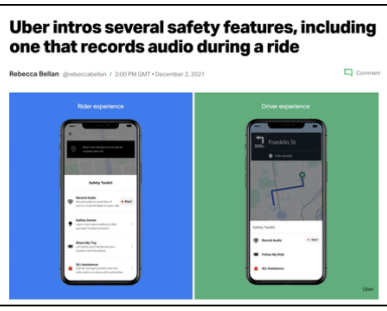
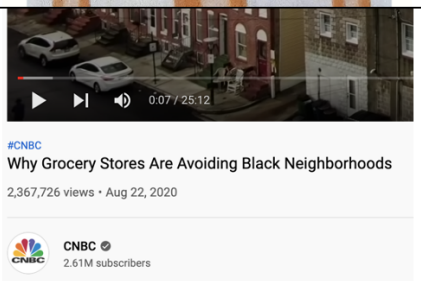
Article/News	Notes
	• Gig-economy development • Tech advancement • Transportation safety • Uber data • Security, data and privacy concerns
	• Art vandalism • Museums • Anna Leporskaya • Art History and preservation process or significant art pieces • Russian art history
	• Food deserts in the US • Socio-economic impact on household food consumption • Cyclical poverty • Number of accessible grocery stores per neighborhoods

Fig. 1: Except of our topic brainstorming process using current articles, news, documentaries and research papers

1.1.2 Creative potential

This prompt mainly refers to the potential of the dataset and/or topics to be presented in a unique and novel way. This ranges from the number of dimensions within the dataset,

potential colour scheme for the visualisations and other unique aspects of the dataset and/or topics.

From the two prompts, our team managed to gather a total of 12 datasets from a range of different topics. However, after some individual evaluation and voting, we shortlisted 7 datasets that received at least 1 vote (figure 2). We then proceeded by evaluating the pros and cons of each of the 7 datasets and came to the conclusion that the Family Food Datasets received the most votes and have no cons at initial evaluation.

Data Visualisation

Part A: Design Report
Challenge: Produce a detailed design specification report
Work in groups, choose your topic/dataset (to be approved)
Explore the data, find the angles, and conceive the visualisation
Explain and justify all your decisions: things you do/don't do

No	Topic	Submitted by	Pros	Cons	Ghea	Ziva	Marfa	Rasyida	Total
1	Uber Traffic	Rasyida		novelty	1	1			2
2	Food delivery time	Ghea	plenty of dimensions	lots of missing values that were replaced by average		1		1	2
3	Pirate attacks on the High Seas	Ziva	novel	topic not of interest			1		1
4	Graduate Employment by Major	Ziva		limited period/scope				1	1
5	Art museum datasets	Ziva		covers specific museums only			1		1
6	Top NFT Collections	Marfa	novel	image datatype is not something we are familiar with	1	1	1		3
7	Family Food Datasets	Rasyida	plenty of dimensions, sufficient datapoints, interesting topic.		1	1	1	1	4

Fig. 2: Screenshot of "DataViz Group - Dataset" Spreadsheet

1.1.3 Importance of chosen topic

Our team unanimously voted for the Family Food Dataset for its relevance towards today's society, creative potential, and the quality and range of the dataset. However, after further research, our team found that nutrient intake has been an ongoing problem in the UK, with 50% of Britons not knowing what nutrients are (Hall, 2019).

Researchers found that in the UK, only 8% of adolescent of low socioeconomic status (SES) meets the recommended daily fruits and vegetable intake. Although this diet is often passed down by parental figures and emphasised by the lack of support for healthy food, behaviours and preferences in adolescence tend to be carried out into adulthood – leading to serious health consequences (Aftomes-Tobio et al., 2016). The NHS found that adults with poor diet are at significantly higher risk of diabetes, cancer and cardiovascular diseases. Furthermore, the NHS reported observing younger and younger illnesses and deaths arising from diabetes, obesity and other illnesses stemming from poor nutrition (nhs.uk, 2022).

1.2 Circumstances

1.2.1 People

Stakeholders: As data analysts at CEDAR (Communicating Diet and Activity Research), a public health advisory organisation focusing on health & diet in the UK, our main stakeholders are our organisation and our partner, the UKCRC (UK Clinical Research Collaboration). As our main stakeholders are established organisations, it is our team's duty to ensure that the data is represented within the visualisation with high accuracy and without bias or misinformation. We also intend to ensure that decisions relating to the final draft are appropriate to the organisations' colours, composition, and tone.

Our team also identified The Economist as a stakeholder. As our current plan is to publish our visualisations with The Economist, we need to consider the direction of our visualisation and evaluate whether the content is appropriate for the media company. Furthermore, with The Economist as a stakeholder, we acknowledge the need for a visualisation that is both attractive and professional

Audience: In 2016, The Economist reached 35 million readers across its print media, online subscription, and website (The Economist Group, 2021). Although the age demographics of readers are not explicitly elaborated by The Economist, their reluctance to converge into the millennial and Gen-Z market suggests that its primary customers are within the generational bands of late millennial, boomer II, and boomer I (30 – 65 years old) (Stibbs,2015). Audiences of The Economist also tend to be working professionals in industries including politics, education, business, finance and similar and are 70% male.

For these reasons, our team expects that the audience of our visualisation project will be highly educated, medium-high income individuals, with basic background knowledge of economics, business, and the markets. Our team also expects that readers will be sound with slightly more complex charts and diagrams without feeling overwhelmed. Despite these, our team will ensure that our visualisation is digestible for the majority public, to encourage information sharing.

1.2.2 Consumption

Setting: We assume that the audience can spend the time to consume our product. As many of our audience are working professionals, media like The Economist tend to be consumed on their commute to work, during their lunch break and on their commute home.

Despite the leisurely setting, for our static visualisation, we recognised that studies found that 55% of people only spend about 15 seconds on an article and up to 40% of readers decide how long they will spend on the article just based on the header information

(Read,2016). Hence, we want to ensure that the static visualisation is in an exhibitory style, where it is attractive enough to grab the attention of a busy person, who wants to spend time reading it.

For our dynamic visualisation, we aim to give our audience the independence to explore the topic from different angles through filters and thresholds. Hence, we choose an exploratory style of visualisation.

Frequency: One-off.

1.2.3 Constraints and deliverables

Pressures: time constraint of 2 months to work on 2 visualisations and a thorough report.
Resources: four consultants with a similar level of technical skills.
Format: static visualisation will be in printed magazines. Interactive visualisation will be in The Economist Online.

1.3 Purpose

The main purpose of our visualisations is to raise awareness and put urgency towards the disparities in food consumption in the UK. As we will be publishing these visualisations on The Economist, we aim to grab the attention of the highly-educated, well-established decision-makers that are audiences of The Economist, with hopes to accelerate change in the system.

1.4 Ideas: research and inspiration

After dedicating some time to looking at the datasets, our team came up with curiosities and questions we could use to explore the dataset. For instance, “How does the rise of income affect food consumption and consequently, nutrient intake?” and “To what extent did the consumption of fresh food reduce during the pandemic?”. These curiosities are listed in a shared spreadsheet, along with details of how these curiosities can be represented in a visualisation.

Ideas

Theme	No	Visualisation Type	Curiosity / Question to Answer	Details	Media Type	Submitted by
Comparing proportion of food consumption between income deciles	1	Pie chart	How does the UK household food consumption pattern differ between income deciles?	A 'pie chart' of food proportion in a plate, representing a region/income deciles		Ida
	2		What types of food are consumed more by higher income decile compared to the lower income decile?			Ida
	3		What is the distribution of 'healthy' food consumption & nutrient intake between income deciles?			Ida
	4	General Layout	What is the distribution of calorie intake between income deciles?		Static	Ida
	5	Euler diagram	Highlighting the extremes - Food consumption "outliers" among UK cities according to ~5 parameters, e.g. smallest/largest amount of consumed food, healthiest/unhealthiest consumption, most/least diverse food eaters.	The main big bubble is representative of all UK cities. The highlighted (or picked out by a smaller) icons of cities are outliers. This will show the proportion of certain trends in comparison to the whole nation.	Static	Marfa
	6	Beeswarm	What is the distribution of nutrient intake/healthy food consumed by a certain age	Percentage of 'healthy food' consumed by a certain age		Ida
	7	Interactive Graphic	Overall UK food proportion / nutrient intake	Interactive nested pie chart to illustrate the proportion of food and its nutrient content		Ida
	8	Beeswarm plot	How does dietary pattern differ between areas <25 miles away from London and areas >25 miles from it?	London is the threshold presented by a line. All the bubbles before it are cities far away from London (i.e. 25 miles away from it), bubbles after the line - cities that are within the 25 miles radius. Supposedly, bubbles after the London line will be much larger because of better food access and higher incomes.	Interactive/Stat	Marfa
	9	Bubble chart	How does the rise of income affect food consumption and consequently, nutrient intakes?	X axis: income decile, Y axis: total nutrient, bubble: food types, bubble size: expenditure on food type	Interactive	Ghea
Time Series Trend: Before & After Pandemic	10	Trend	How does the food consumption pattern in the UK evolve over time?	Illustrating food consumption trend (if any) by increase/decrease of consumption (please hover below, the one with watermelon, oil, etc)		Ida
	11	Trend	How does the food consumption pattern in the UK evolve over time?	We can focus on a certain type of food over time (e.g. what is the main source of carbohydrates for people in the UK over time?) Year-binning, highlight top consumed food Highlight trends happening during those years	Interactive	Ida
	12	Stacked bar chart	How does food expenditure change over time and its effect on food composition?	Food expenditure is measured as a proportion of income and analyzed year over year. We deep dive into how the changes in expenditure affect the type of food consumed by households (e.g. we can take a cutoff point, e.g. Feb 2020, as the start of pandemic. In the time-series graph, we can also point out the time where people stock up their food expenditures during lockdowns)	Interactive/Stat	Ghea
	13		Expenditures: How does the UK food expenditure pattern shifted before and after pandemic?			Ida
	14		How does the UK food consumption pattern shifted over time (household vs eating out) purchases ?	(e.g. Do people eat more healthy products after pandemic, considering increasing awareness towards healthy lifestyle?)	Interactive/Stat	Ida
	15		To what extent does consumption of fresh food (vegetables, milk, fruits) reduced during the pandemic?	Inspired by the research done by behavioural scientists that found people in Denmark, Slovenia and Germany went grocery shopping less, leading to the less fresh food consumed. Essentially want to see if the same thing can be observed in the UK. link to research: https://www.frontiersin.org/articles/10.3389/fnut.2021.635859/full	Interactive/Stat	Ziwa
Regional Analysis	16	Choropleth map	How does dietary pattern differ accross regions in the uk / Colour-blind map about nutrition intake among UK regions	Map to show eating habit in different regions. Create health index based on composition of food consumption (e.g. % of meat, % of vegetable, % of milk, % of confectionary, etc). Feature engineering: Health index based on type of food	Interactive	Ghea & Marfa
	17	Scatter plot	How does food expenditure change between regions in the UK?	A comparison of expenditure change in % for different cities	Static	Ida
	18	Proportional symbol map	What are the main cities-consumers of food in the UK?	The bigger the bubble on the map, the bigger the food consumption in the city.		Marfa
	19	not sure what its called	Comparing nutrient intakes between people from different ethnic origin, rural/urban, occupation	Each line represents each ethnic origin in the dataset, and the color would be representative of the ethnicity (i.e in Mona Chalabi's style)	Static	Ziwa
	20	Cartogram	Comparing veg vs meat intakes between people regions	Veg vs Meat Ratio	Interactive	Marfa
Others	21	Sankey diagram	Food distribution	Total food produced/imported to the UK vs how much is wasted, recycled, eaten, given to animals (if the data exists)		Marfa
	22	Marimekko chart	How much food waste does the UK generate and recycle?	We can compare it sideways to a similar plot about the same topic on EU countries.		Marfa

Fig. 3: Screenshot of "DataViz Group - Ideas" Spreadsheet

1.4.1 Static: Dashboard of a single topic, supported by different perspectives

Dashboard allows us to present different information in a single visualisation. Despite the convenience, we need to ensure that the dashboard is not overwhelming for consumption. Hence, if we choose to create a dashboard, our group will visualise one topic, but support it with different perspectives.

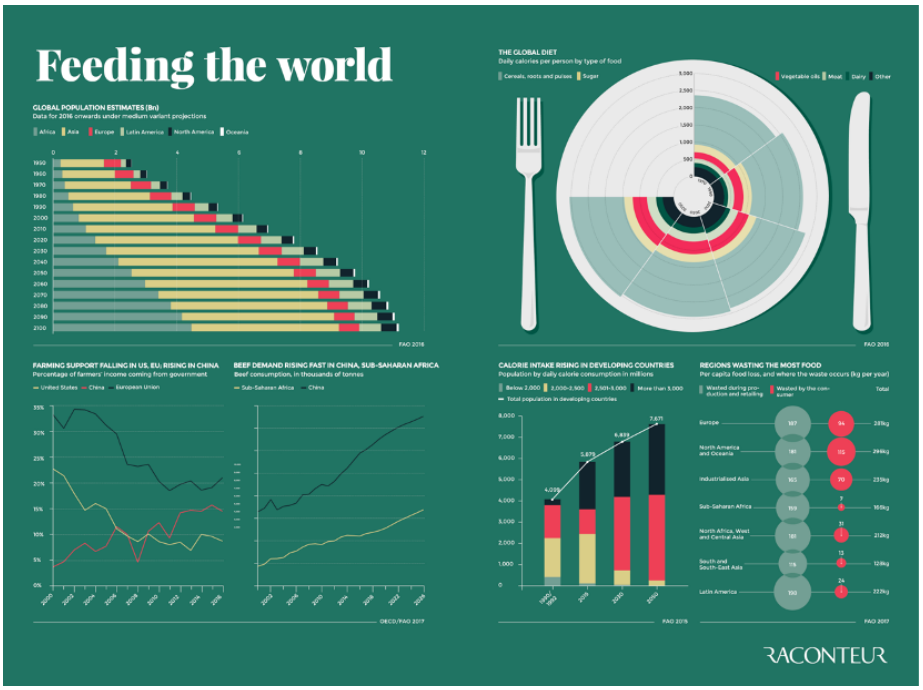


Fig. 4: Sample dashboard-style infographic (source: Raconteur, 2022)

1.4.2 Static: Beeswarm plot

For instance, we can use a beeswarm to plot average calorie intake in different parts of the UK. We can also create multiple beeswarms for different dimensions, such as calorie intake by race, by gender, by profession, etc.

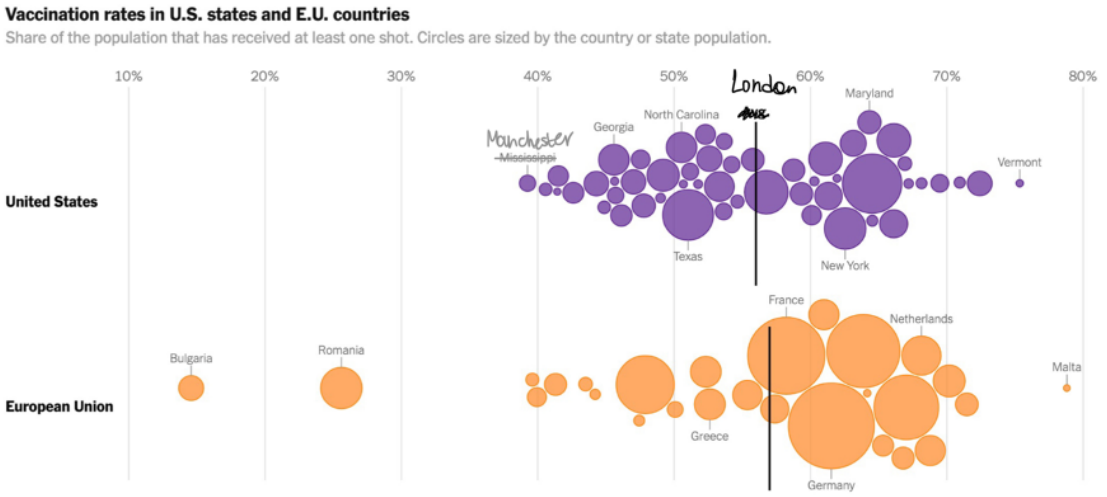


Fig. 5: Sample of static Beeswarm plot (source: lecture slide week 3)

1.4.3 Static and/or dynamic: Map plot

As our team has access to location data, we will be able to utilise this and perhaps highlight the 3 areas in the UK with the best diet and the worst diet. If we decide to create a map plot for a static visualisation, it is important to only highlight 3 – 5 important points to mitigate the possibility of the map being too busy. On the other hand, using a dynamic visualisation we can plot every single data point and only show the details when a user hovers over the map.

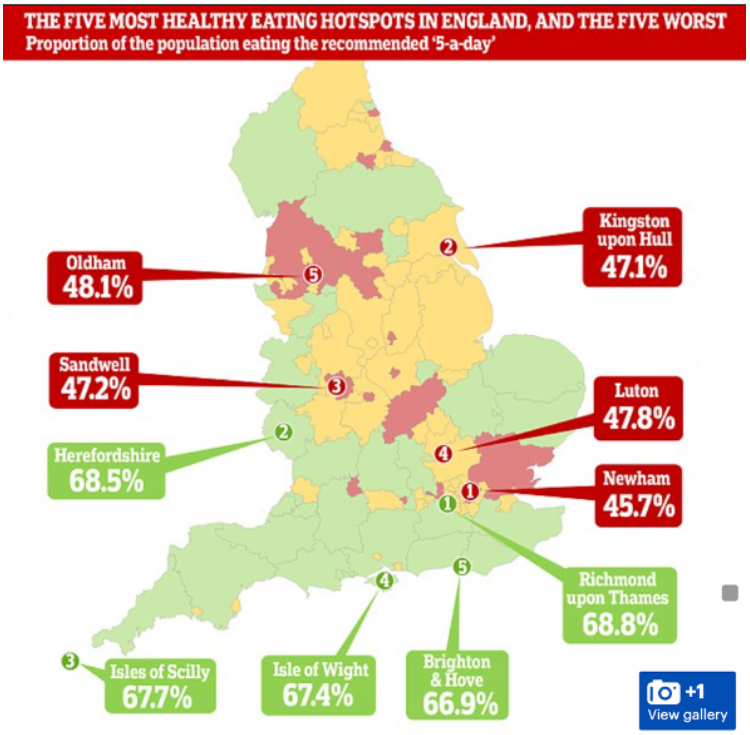


Fig. 6: Sample of map plot (source: Daily Mail, 2018)

1.4.4 Dynamic: Interactive diagram used to view different time periods in the data

Our initial idea of how we will use an interactive visualisation like figure 4 is to present the diet composition or nutrient intake by groups (i.e, proteins, carbohydrates, fats, etc.), over different periods of time. However, it is important to note that this would only work if changes between different time periods are significantly different.

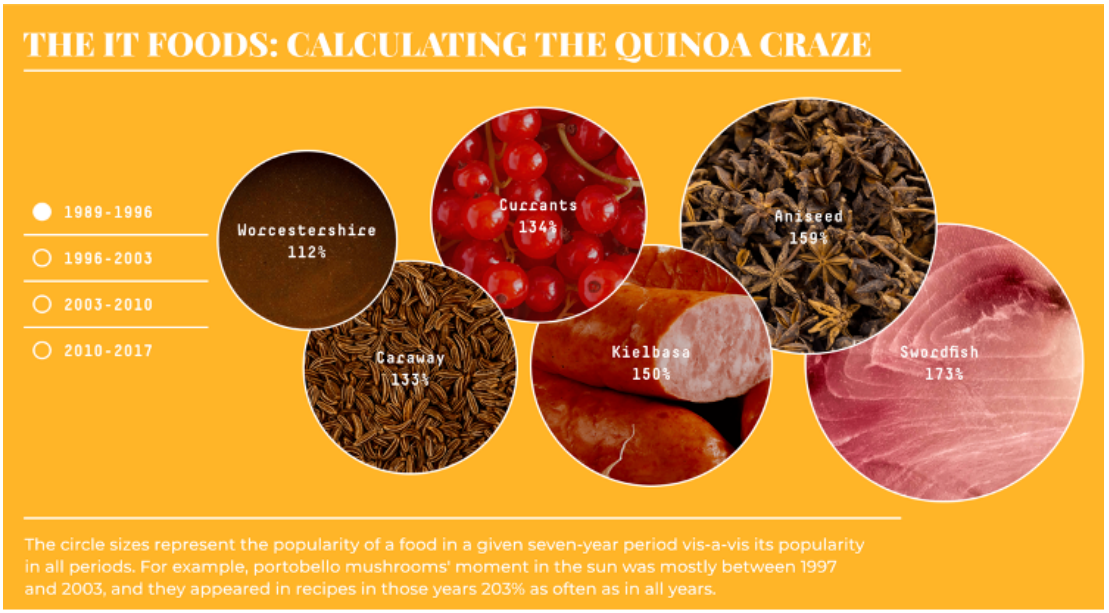


Fig. 7: sample of interactive bubble chart with time series (source: Ceros, 2022)

1.4.5 Dynamic: Gamified/user input visualisation

This visualisation will allow users to select foods that would normally be in their plate, and the corresponding amounts. After the selections are made, users will then be able to see a score representing how healthy their diet is. Furthermore, their input will then be plotted against the average in different parts of the UK in a simple line graph.

Fig. 8: Sample of gamified visualisation (source: Fruhlinger and Mcnair, 2022)

1.5 Rough Sketches

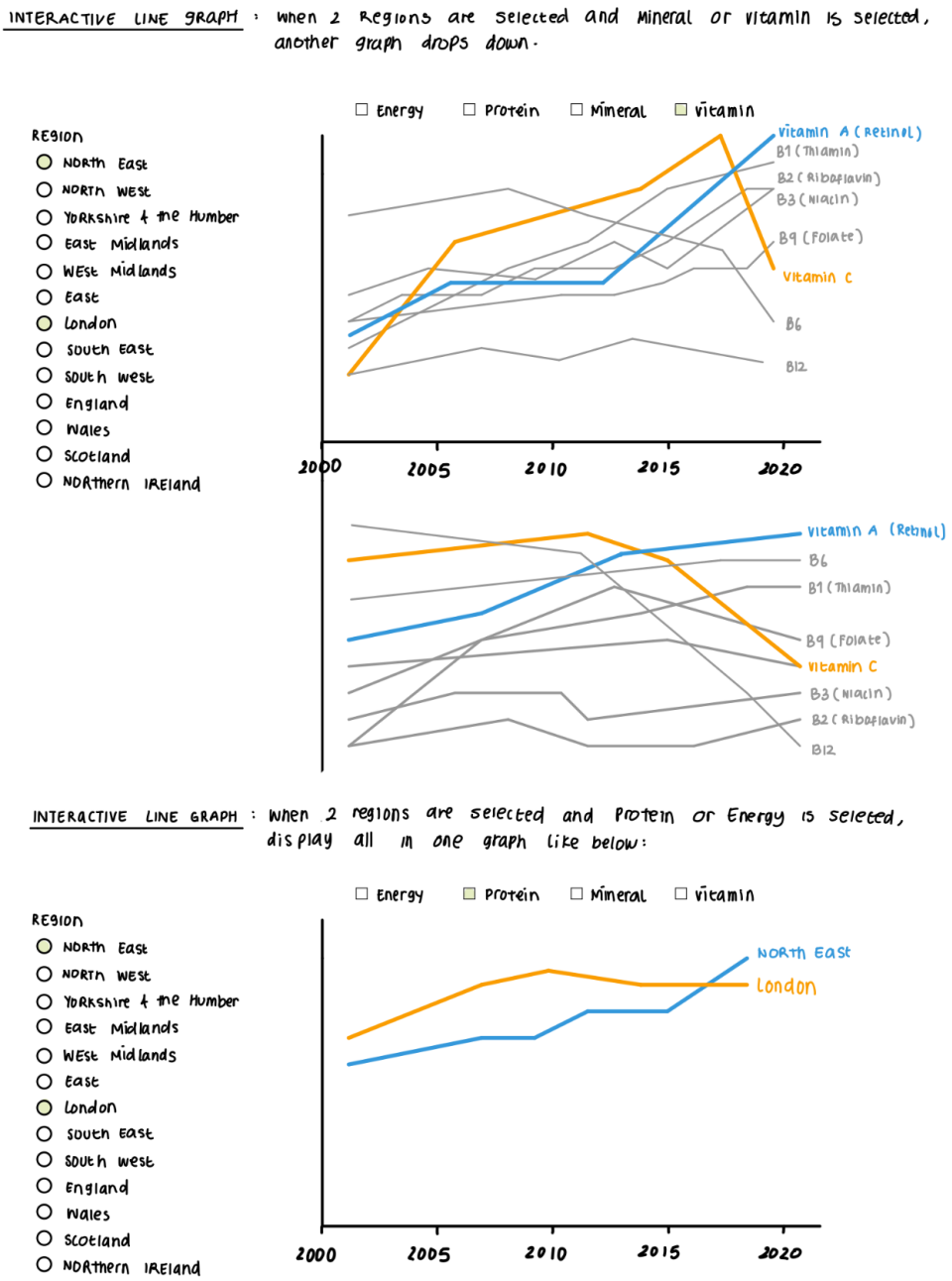


Fig. 9: Rough sketch of interactive line graph

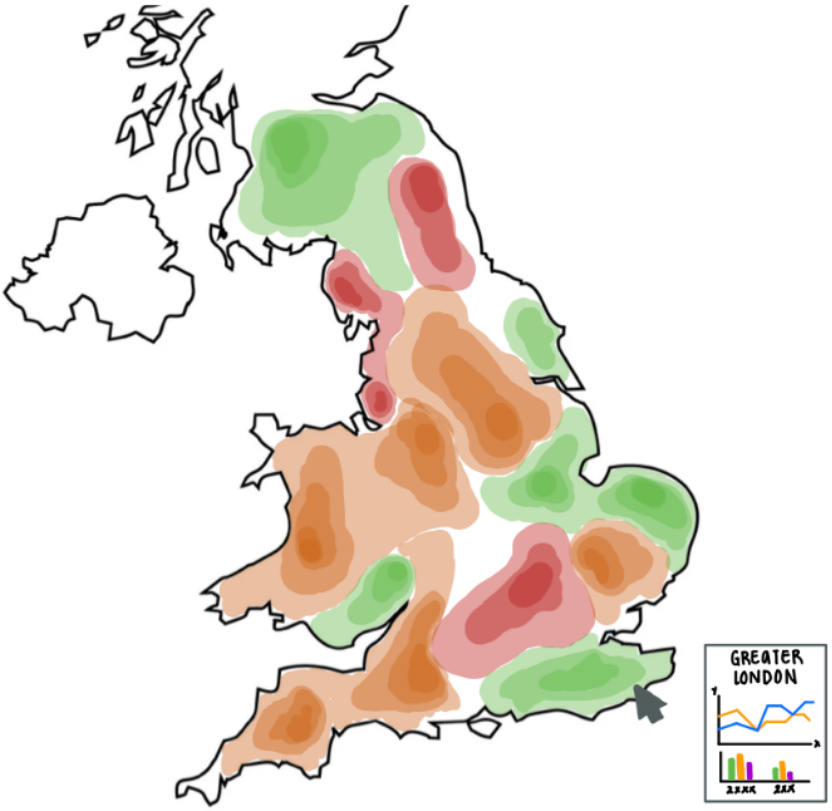


Fig. 10: Rough sketch of choropleth map

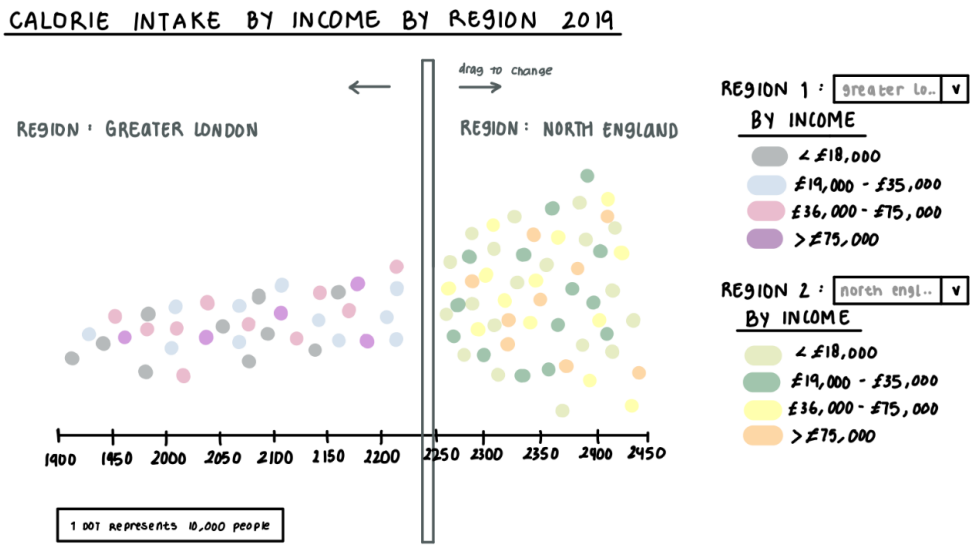


Fig. 11: Rough sketch of interactive beeswarm plot

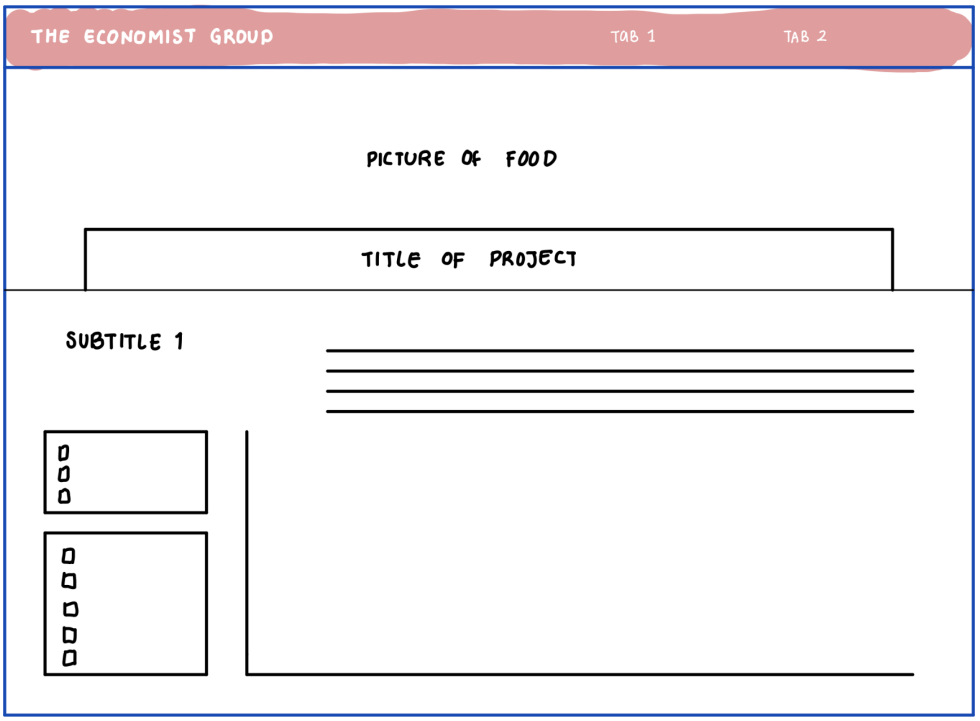


Fig. 12: Rough sketch of website template

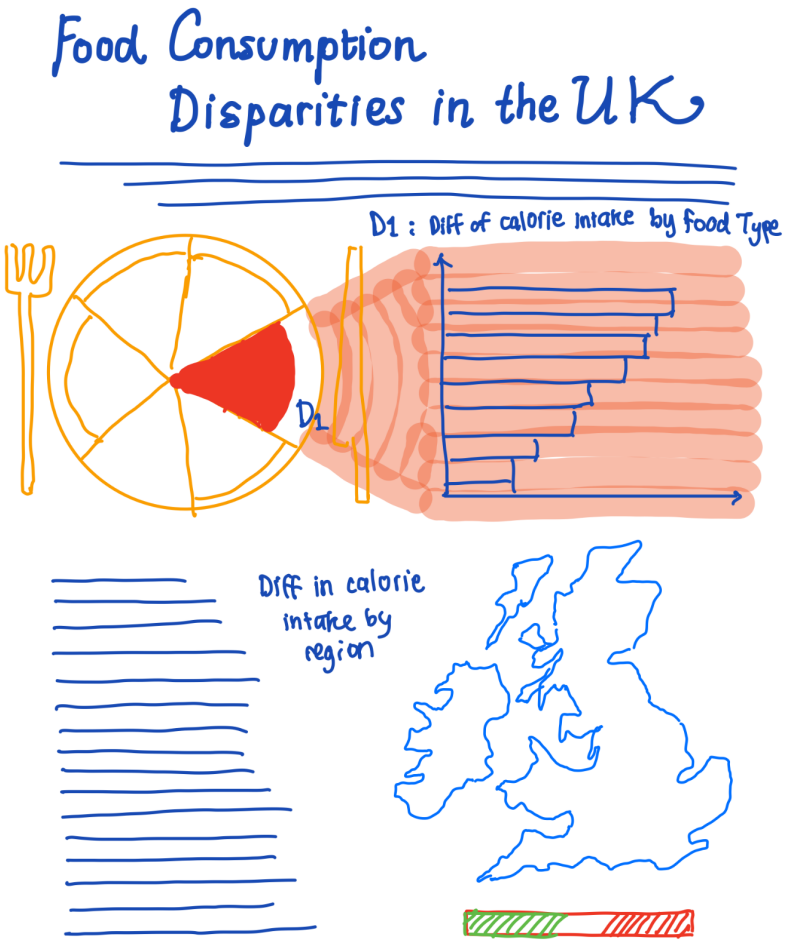


Fig. 13: Rough sketch of infographic

2.0 Working with data

2.1 Data preparation

2.1.1 Raw data

All data is sourced from the Family food datasets by the UK's Department for Environment, Food & Rural Affairs. The data comprises several tables, both time-series data from 2001 to 2019 grouped by Equivalised Income Decile Group (EID), age, household composition, ethnic origin, economic status, rural versus urban, Gross Income Quintile Group (GIQ), age of the household reference person ceased full-time education, and occupation, as well as cross-sectional data of different regions in the UK.

Do we need to collect more data?

Based on the curiosity section, it is likely that we will find the following sources helpful:

- Gross Domestic Product
- Consumer Price Index on Food
- Coordinators of the UK regions and countries

Do we need to clean our existing data?

Although the underlying data provides comprehensive information on family food expenditure and consumption, further data processing is required to ensure a suitable format for visualisations. As the family food dataset compilation consists of a number of different datasets, we will divide it into three general themes; namely income deciles, time series and regional analysis. Data preparation steps for each theme are presented below and in our notebook attached. .

2.1.2 Income deciles

There are some inconsistencies in the naming of the food group on the more granular level of sub food groups. To solve this, some sub food groups are merged into one comparable food group. In addition, for the purpose of comparing consumption across different groups, some metrics are converted to one uniform metric. For example, liquid type food groups like milk and soft drinks are measured in litres, are converted to grams so they can be measured in proportion to other food groups that are measured in grams like meat, fruit, and vegetable (see notebook: chapter 1).

2.1.3 Time-series analysis

The time-series analysis focuses on two datasets, namely the nutrient intake and food expenditure data, with a horizon of year 2001 until 2020. Firstly, for the nutrient intake

dataset, we need to standardise the values according to the food reference intake baseline, so all types of nutrients are comparable to each other. Secondly, for the food expenditure data, we aggregate the values into the second level of granularity to simplify the analysis. This includes, for example, combining dairy products into one category and meat products into another. Thirdly, as the unit of each food product may differ, we convert the unit of analysis into one uniform metric (see notebook: chapter 2)

2.1.4 Regional analysis

As part of data preparation for regional analysis, we merged 2 datasets: one is created by searching UK region coordinates, and another is "regions household nutrient intakes". For a general overview, we only considered the last tab of the Excel sheet. It presents a 3-year average intake as a percentage of food energy excluding energy from alcohol. Due to the complexity of the geographical visualisation in Tableau, we only kept the essential rows from the dataset: total average energy per person in kcal, and the percentage of that energy that is consumed by fats, carbohydrates, and proteins - all per UK region and country (see notebook: chapter 3).

What constraints do we need to consider?

After data cleaning, there are remaining constraints that we need to address. First, in the expenditure dataset, for a new type of food product introduced in recent years, the data is only available from that year onwards. To overcome this, we impute the moving average of corresponding next five years' data.

Second, as there is a 20 years span of data, occasionally there is an introduction of a new food group that was not relevant before. For example, vegetable-based 'meat' products have only been available for the last 10 years. In this case, as the contribution to the total expenditure is relatively small, we exclude it from the analysis of time-series.

Lastly, as we want to examine both the relationship between food consumption pattern and expenditure, it is also interesting to see how this dynamic is affected by price. In the scope of this project, however, price data is not available. Furthermore, there may be other factors influencing the rise or decline of expenditure that may not be captured by the current project scope, such as inflation, a changing consumer preference towards a certain type of food or a change in public policy in general. Hence, we will proceed with the analysis by taking into account the above three constraints.

2.2 Data exploration

2.2.1 Income deciles

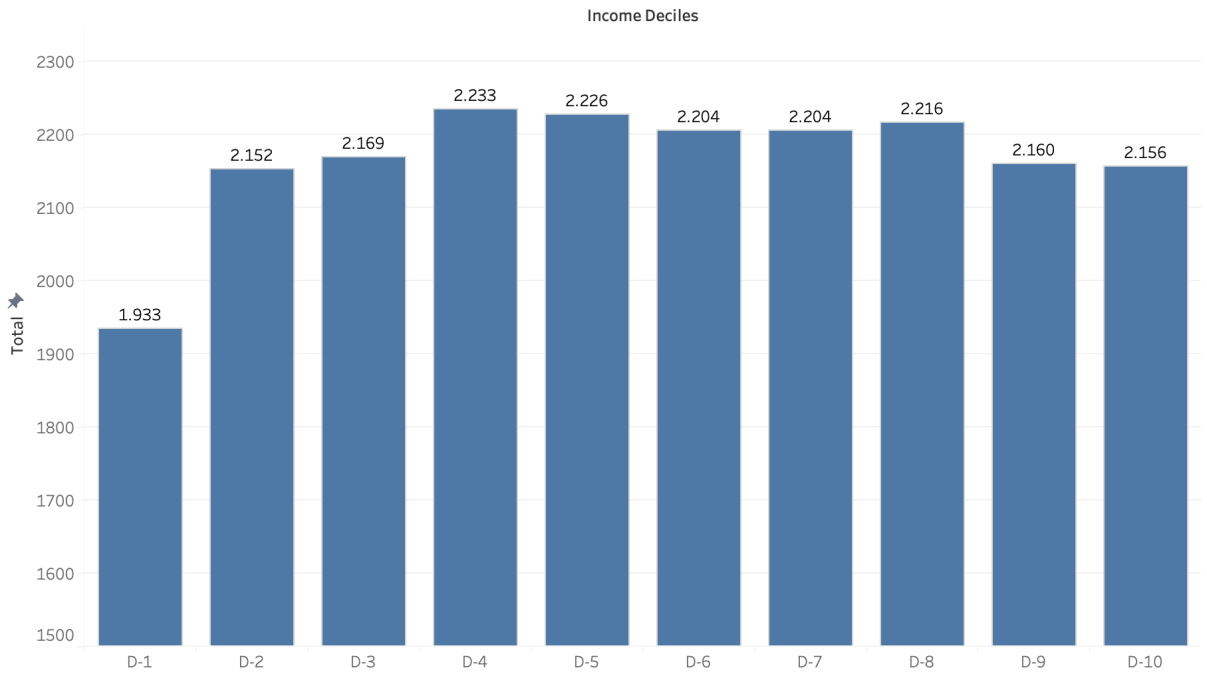


Fig. 14: Energy consumption by income deciles (avg. 2017-2019)

Based on three years' average energy consumption per day, measured in kcal, people in the lowest income category consumed significantly less amount of calories. However, the relationship is not linear. People in the fourth income decile consumed the most calories, higher than people in higher-income categories. The wealthiest people in decile 10 are ranked 7th in the number of calories consumed per day.

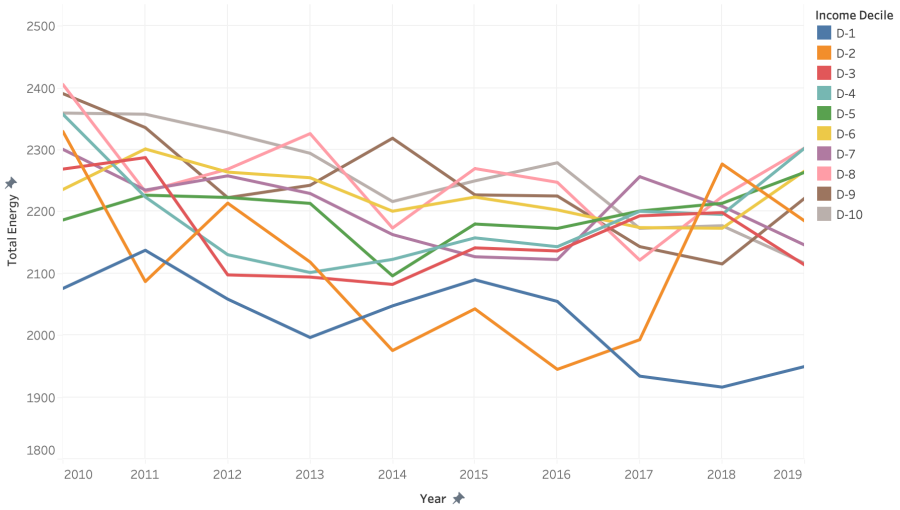


Fig. 15: Total energy consumed by income deciles

Over the years, the average calories consumed per day is relatively stable for most income deciles. However, the average calories for the poorest families in decile 1 showed a

deteriorating trend. Income decile 2 also showed a declining trend until 2016 before increasing and reaching its peak in 2018.

Fig. 16: Proportion of food consumption by income deciles in % (avg. 2017-2019)

There are some interesting patterns that we observe in the difference of food type composition, measured in grams, among income deciles. Although fruit and vegetables make up for the highest proportion across all income deciles, the food group accounts for the highest proportion compared to other food groups for the higher income deciles. On the other hand, lower-income deciles tend to eat a higher proportion of milk and milk produce. Wealthier individuals also consume a higher proportion of soft drinks compared to their lower-income counterparts. Interestingly, lower-income individuals consume a higher portion of alcohol compared to higher-income individuals.

2.2.2 Time Series Model

Apart from the analysis from the income decile perspective, the food intake dataset can also be analysed from a time-series perspective. There are four different approaches explored to gain more insights; namely the nutrient intake yearly trend, total expenditures on food products over the years, comparison between household spending vs eating out and lastly the trend of fresh vs processed food consumption across different years.

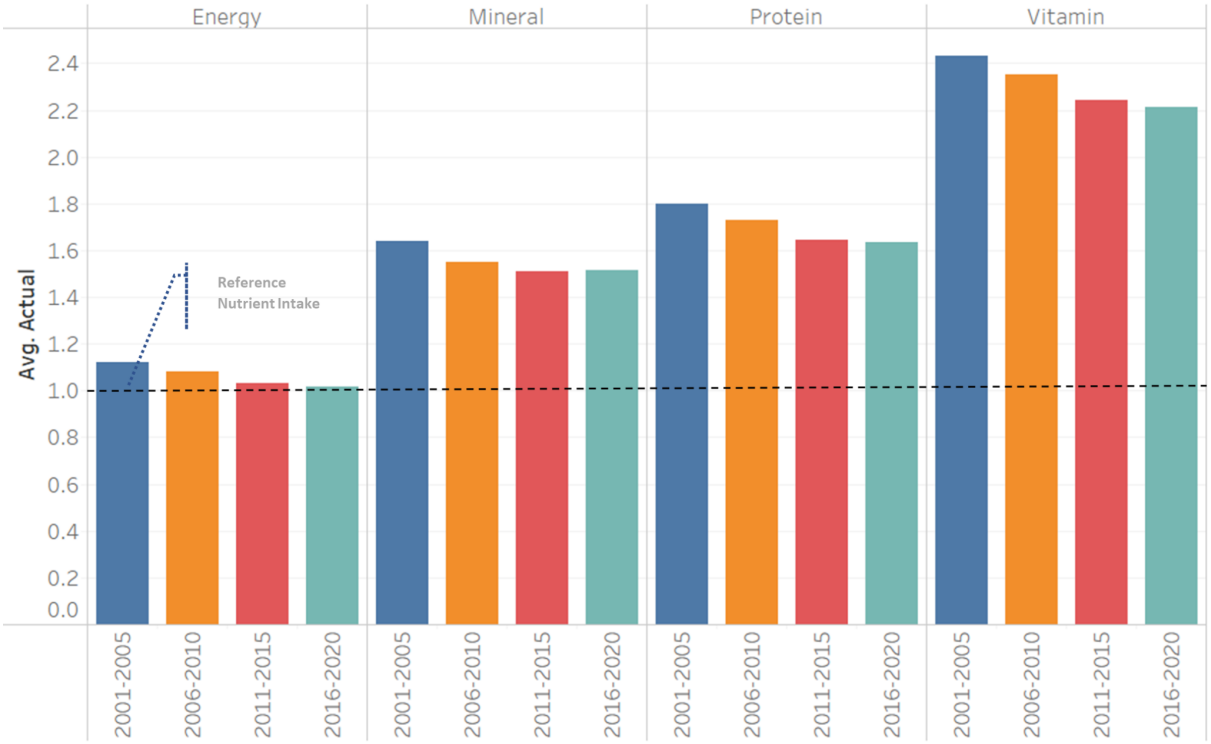


Fig. 17: Average intake as a percentage of weighted reference nutrient intake

Figure 17 shows the relative nutrient intake for each nutrient group compared to its reference value based on the suggested 2000 kcal daily average nutrient intake. For simplicity, the reference value has been standardised as "1". An actual value of 1.5 means that there is 50% more nutrient intake than the suggested daily average. From the graph we can infer that there is generally a declining trend in average nutrient intake across different nutrient groups, however, it is important to note that this decline is still within the acceptable range of daily nutrient intake. It is interesting to deep dive into what causes this decline and whether it is also reflected in the household food expenditure pattern.

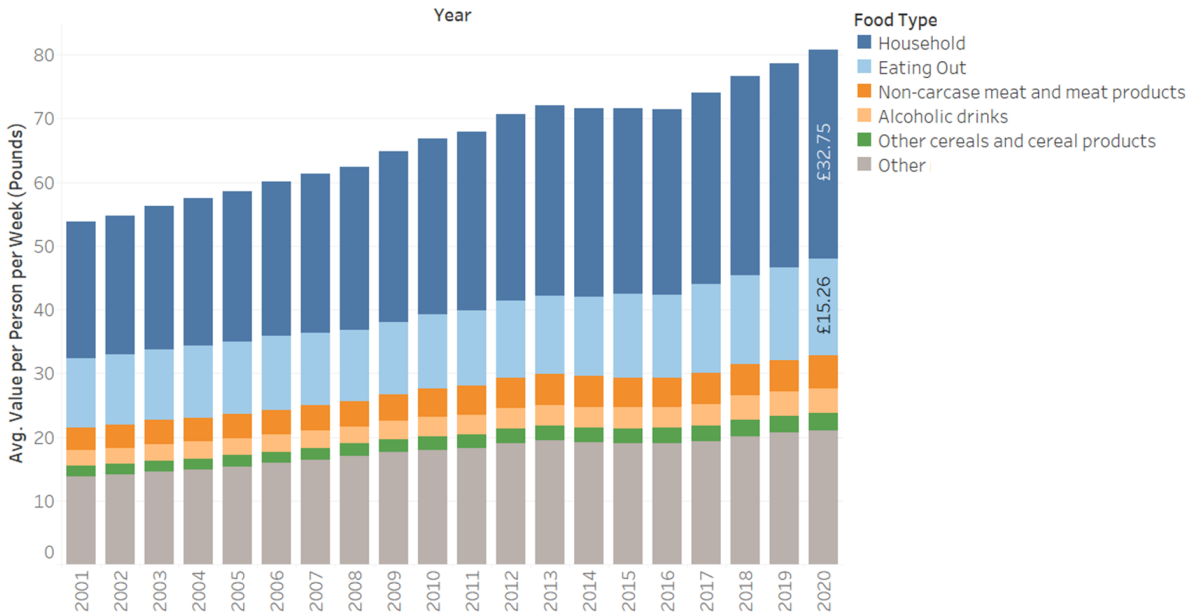


Fig. 18: Yearly average household expenditure for top product spend

Figure 18 shows the trend of UK total expenditure on food and drink which consists of both household and eating-out expenditures. The graph highlights that in general there is a stable increase in food expenditure. From 2001 until 2020, there is a 25% increase in the average value spent on food and drinks alone. The top categories contributing to the expenditure trend are meat products, alcoholic drinks and cereal products. It is intriguing to see that figure 18 seems to contradict the findings in figure 17. Nevertheless, the expenditure pattern is far more complex than the trend of average nutrient intake, as we need to take into account the inflation rate, changes in people's buying power and the change in food composition over time; which are out of the scope of this project. It is possible, however, to analyse the breakdown of household vs eating-out patterns across the years, which is presented in figure 19.

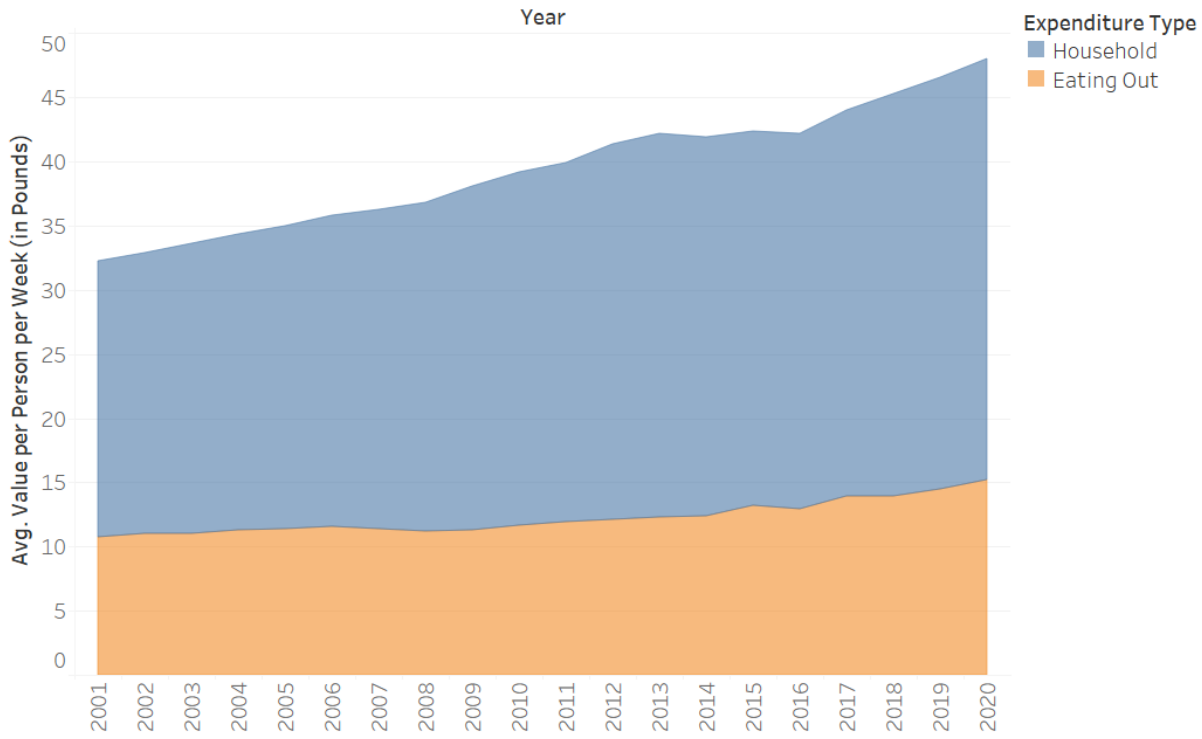


Fig. 19: household vs eating out expenditures

We can infer that the increase of expenditures over the years is not uniform between household vs eating out expenditures. While eating out expenditure has a steady increase up to 42% in 20 years, household expenditure has significantly increased by 62%.

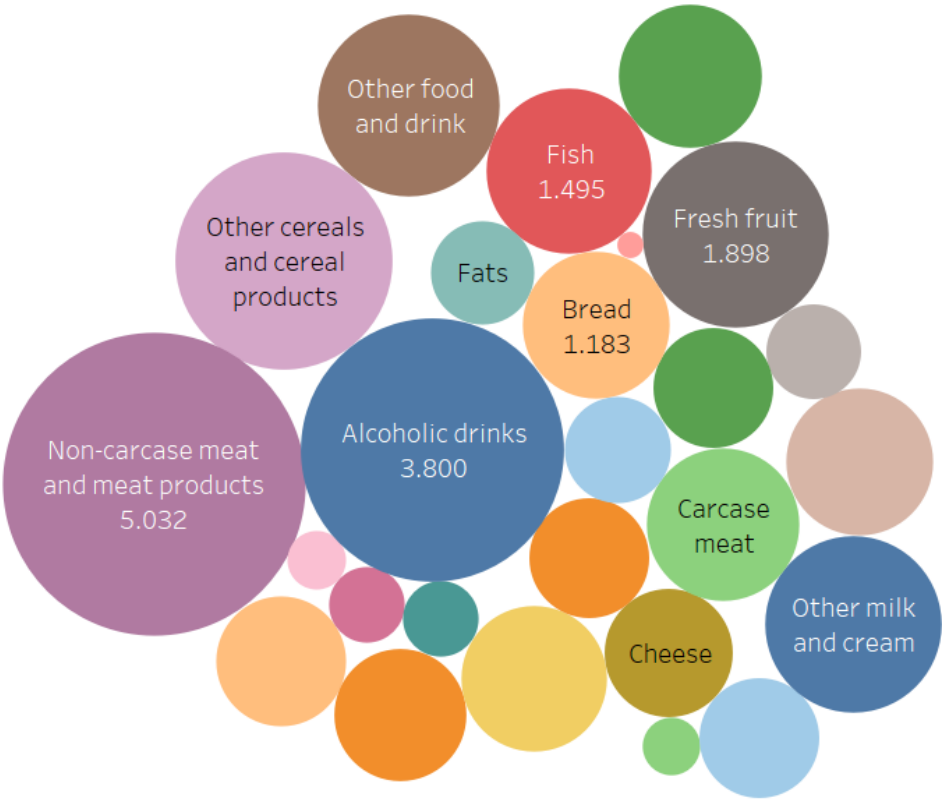


Fig. 20: proportion of average household expenditure in a year

If we deep dive into the components of food expenditures each year, there is a steady increase for each food type. Meat and meat products remain the top category based on spending, assuming that the price per volume for meat products is generally high. The second-largest contributor is alcoholic drinks, which is driven by high-value alcoholic drinks such as wine and champagne (44% of total alcoholic drinks expenditure). The third-largest contributor is cereal products, consisting of carbohydrate-rich foods such as breakfast cereals, pasta, rice but also pizzas. Furthermore, after analysing the main contributors of expenditures, it is interesting to see whether there is a significant shift towards fresh food consumption compared to processed food, which is presented in Figure X5.

2.2.3 Regional analysis

The figure below shows a deviation from the recommended calorie intake. For men, it's 2500, for women - 2000, so we considered the average of the two as the recommended, 2250 kcal. Interestingly, no UK region had a positive deviation, i.e., no region ate more than the recommended calorie intake. Greater London has the biggest deviation from the recommended calorie intake, so they consume the least. Meanwhile, in the East Midlands, the difference from the threshold is the smallest, so they consume the most. Such a drastic difference is surprising, as the 2 areas are very close to each other geographically.

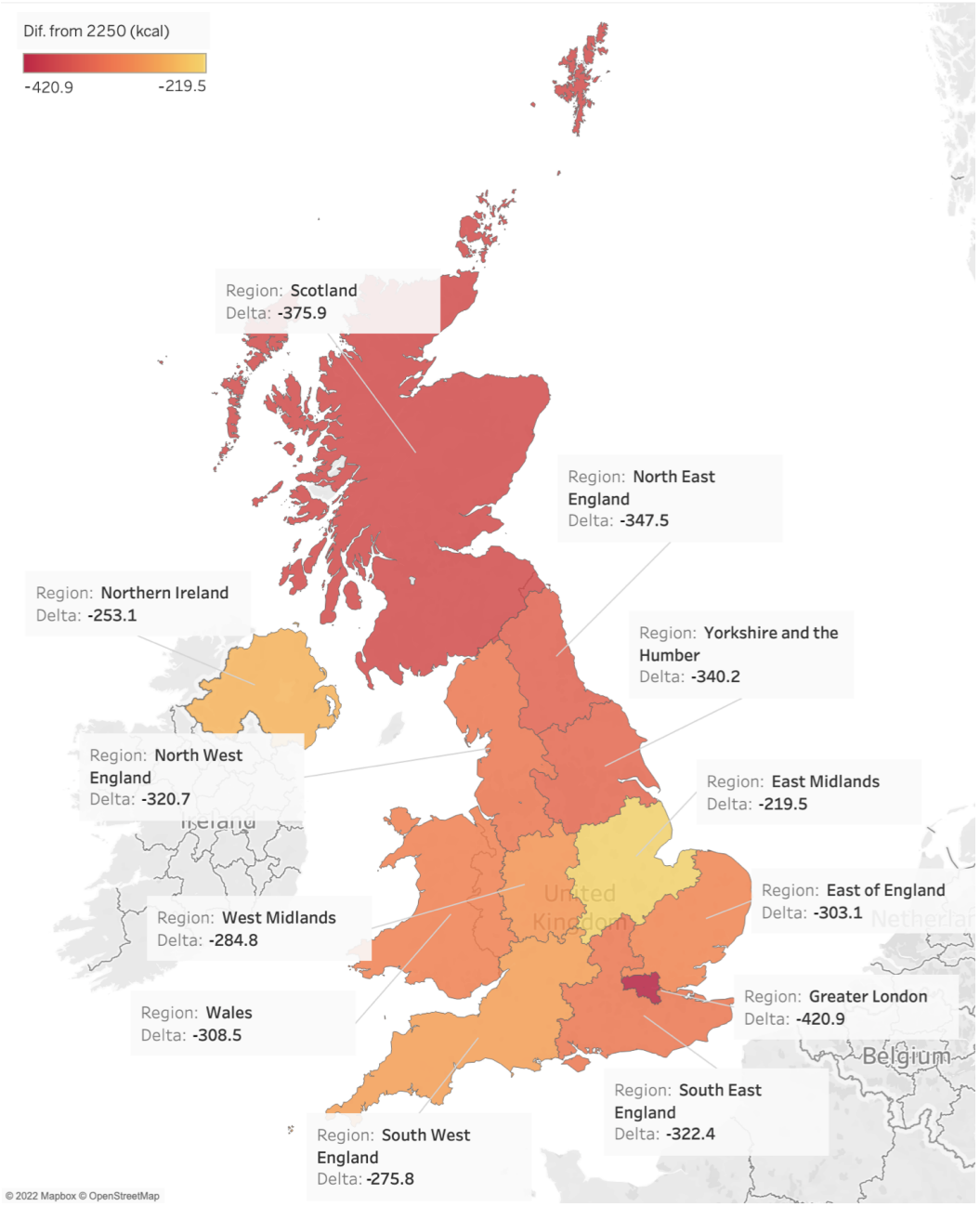


Fig. 21: deviation from the recommended energy intake, per UK region

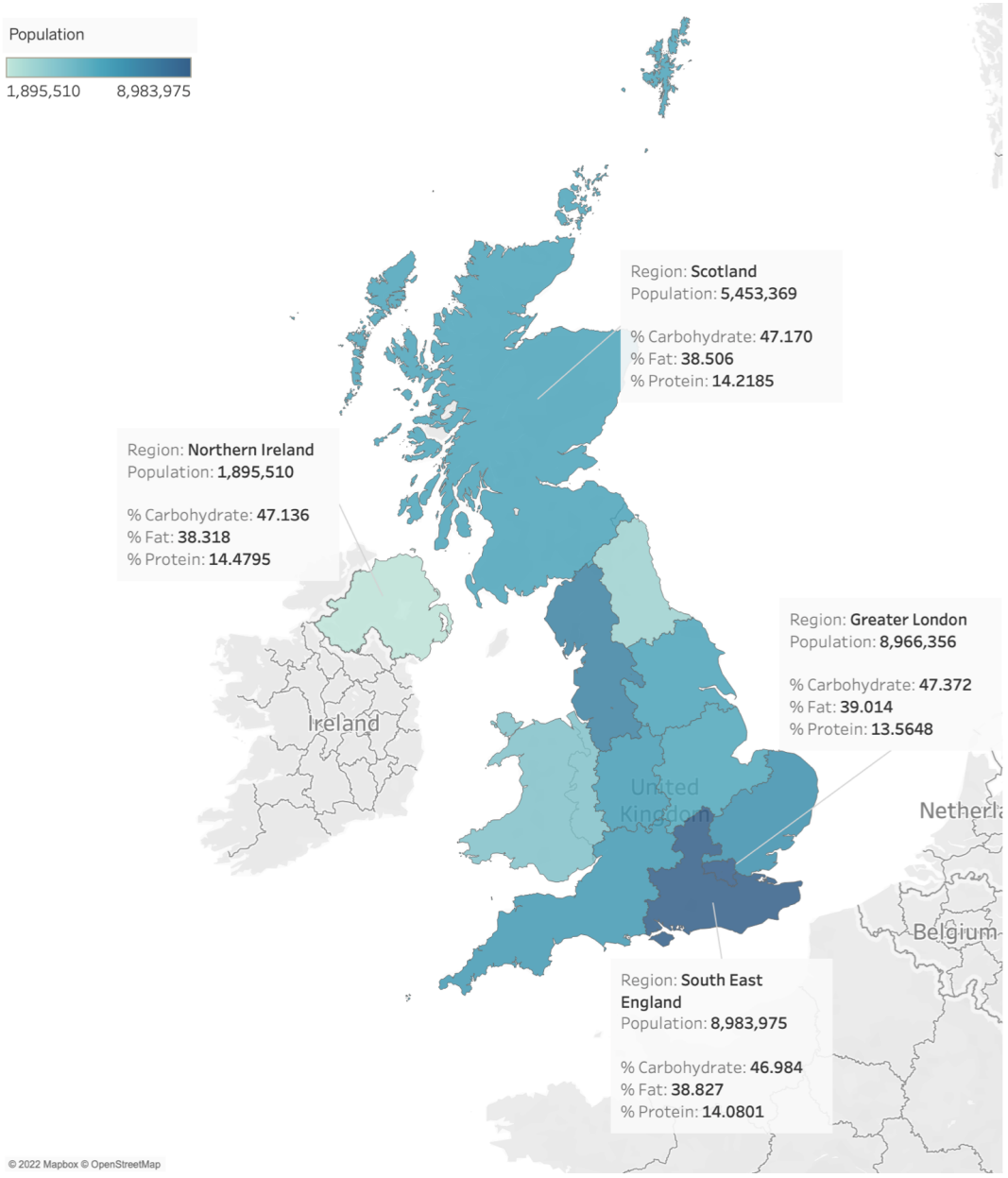


Fig. 22: population per UK region vs. nutrient percentage intakes

2.3 Key messages

1. Calories consumed increase simultaneously as income increases to a certain point. Beyond that, an individual does not consume more calories.
2. People in the lower-income deciles consumed fewer calories over the years.
3. There is a decline in the actual nutrient intake compared to the reference nutrient intake, although the level is still in a normal range.

4. There is a steady increase in household and eating out expenditures in the UK. In terms of fresh vs processed food comparison, the increase of processed food consumption drives this change.
5. Over 3 years, East Midlands has the smallest deviation from the recommended energy intake, whilst Greater London has the greatest. This is surprising as the 2 regions are nearly neighbouring each other and thus should have relatively same food distribution channels and routes.
6. The more the region is to the North, the higher its average proportion of carbohydrates consumed, making Wales a protein leader, and North West England leader in fats.

3.0 Editorial thinking

3.1 Initial editorial thoughts

There are three themes that we are covering in this report: food consumption across income deciles, food consumption trend over time, and food consumption across the UK regions. For the first one, we are interested in the relationship between income and calories consumed, specifically how it differs across different income deciles. The second theme focuses on food consumption patterns and how it changes from one year to another for different food groups and regions. In the last theme, we will explore the differences in nutrient intake in different regions in the UK.

What angles are we covering?

For the first theme on income and calories, we found that the bottom 10% of earners in the UK consumed significantly fewer calories compared to the overall average. Based on the NHS, the recommended calorie intake per day is between 2,000 and 2,500 (*What should my daily intake of calories be?,* 2022). With this in mind, the bottom 10% are consuming less than the recommended amount. To understand better where this discrepancy is coming from we will deep dive on the food composition of different income deciles and find the food groups that individuals in decile one consumed significantly less than the overall average.

The second theme takes on two angles. The first angle focuses on time-series data of food nutrient intake in the UK by region. The second angle is on food expenditure patterns over time by food type. From the initial analysis, we found that there is a contradicting trend between nutrient intake and food expenditure pattern, therefore we would like to explore further whether this difference is consistent for each time period.

For our last theme on calories and region, we will focus on the relationship between the region's population and its deviation from the recommended daily calorie intake.

What messages are we communicating to our audience?

In our first theme, we want to raise the urgency on the issue of the calories deficiency happening with the bottom 10% of earners. Furthermore, the breakdown on food types will inform policymakers on the difference in diet between the bottom ten percent and the other income deciles that can potentially eradicate the calorie deficiency.

The objective of our last theme is to inform stakeholders in the food and beverage industry about the regional disparities in calories consumption. By raising this issue to the general audience, we hope to encourage the government to look into regions with the most calorie deficiency to develop better local infrastructure like roads for better food deliveries.

Is our portrayal fair and complete?

We sourced our data from the UK government's official website so we can ensure that there is minimal bias. However, there are a few limitations that we are aware of. Since the data only contains information on food consumption, there is no explanation of what factors are driving the trends and causing the increase and decrease in a particular year compared to another. This can potentially be explained by incorporating other data sets that capture the changes in food prices and income.

We also need to be mindful of the characteristics of self-reported dietary surveys, which is the source of our dataset, because respondents are known to under-report their purchase. Consequently, any derived nutrient intake that is based on the purchased quantity will be underestimated as well. This problem is even more prominent for food types such as alcohol, in which empirical comparisons of sales and duty data for alcohol suggest that reported alcohol consumption could be 40-60% lower than the reality.

Will there be any difference in our design choices?

For the final output, we will have two types of visualisation, one static and one dynamic. The static visualisation will be an infographic featured in *The Economist* printed magazine. The dynamic visualisation will be presented on *The Economist* website.

The first theme on income and calories will be presented as a static visualisation. A part of theme three, the deviation from recommended calorie intake by region, will also be presented as static visualisation. The reason for this design choice is that both themes are focused specifically on one angle. The first theme focuses on decile one's calorie deficiency while the third theme focuses on the deviation in calorie intake by region. This information is straightforward and can easily be shown in a static visualisation.

Dynamic visualisation will present food expenditure and nutrient intake over time, which can be sliced by region and nutrient types. Due to the versatile nature of the data with many angles, we decided that the presence of a filtering feature will be useful to display a vast amount of information, simultaneously keeping the simplicity of the graphs.

4.0 Design solution

Since both the static and dynamic visualisations are designed to be featured in *The Economist*, magazine for the prior and website for the latter, the colours and font used have to be consistent.

4.1 Colour

The choice of colours is intended to convey professionalism and trustworthiness. Since the visualisations will be featured in *The Economist* magazine and website, the entire visualisation will be using four colour palettes that are taken from the official *The Economist* colour guideline. For static visualisations for the magazine, the main palettes will be Greyscale and Tertiary. For the dynamic visualisations, which will be featured on the web page, the main palettes will be Greyscale, Canvas and Tertiary. Both static and dynamic visualisations will incorporate elements of red, which is *The Economist* Brand palette. The aforementioned colour palettes will be elaborated below.

The first colour palette is The Economist Brand which consists of two shades of red, The Economist Red (#E3120B) and The Economist Red 60 (#F6423C). These shades will be used to draw users' attention to a specific aspect of the data. Since red is the colour of The Economist logo, it has to be used very carefully and only when necessary in order to avoid conflict with the brand logo.

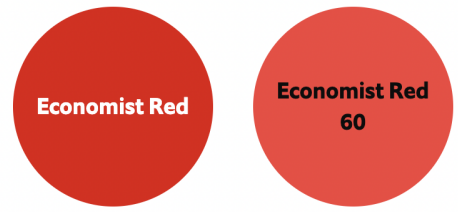


Fig. 23: The Economist brand colour palette (source: The Economist Design System, 2022)

The second colour palette is Greyscale. This will be used for general visualisation and to provide a contrast to the highlighted aspect which be coloured in red. The grey scale colour palettes consist of London 5 (#0D0D0D), London 10 (#1A1A1A), London 20 (#333333), London 35 (#595959), London 70 (#B3B3B3), London 85 (#D9D9D9), London 95 (#F2F2F2), and London 100 (#FFFFFF).

Fig. 24: The Economist Greyscale colour palette (ibid.)

The third colour palette is Canvas which contains shades of light blue and grey. This palette will also be used for general visualisation as well as background. The palette consist of Los Angeles 85 (#E1DFD0), Los Angeles 90 (#EBE9E0), Los Angeles 95 (#F5F4EF), Paris 85 (#D0E1E1), Paris 90 (#E0EBEB), Paris 95 (#EFF5F5).

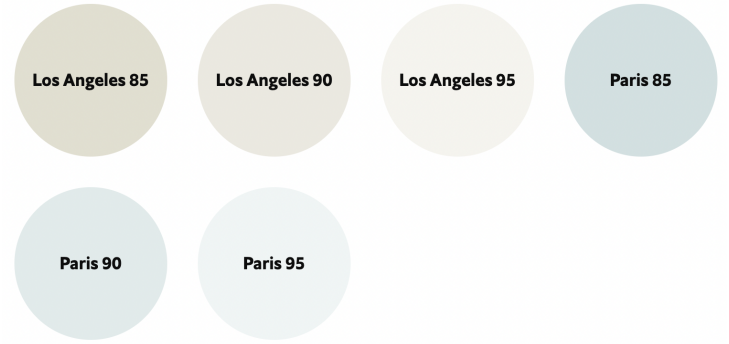


Fig. 25: The Economist Canvas colour palette (ibid.)

The last colour palette is Tertiary which contains shades of orange and yellow. This palette will be used to highlight specific data points. The palette consists of Singapore 55 (#F97A1F), Singapore 65 (#FB9851), Singapore 75 (#FCB583), Singapore 90 (#FEE1CD), New York 55 (#F9C31F), New York 65 (#FBD051), New York 75 (#FCDE83), and New York 90 (#FEF2CD)

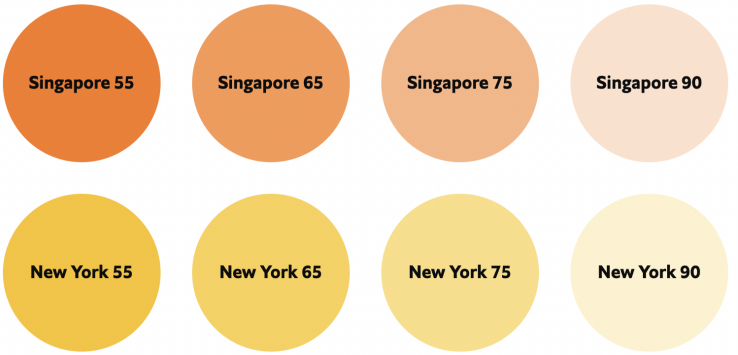


Fig. 26: The Economist Tertiary colour palette (ibid.)

4.2 Font

We decided to refer to the fonts used on *The Economist*'s official website. The writings are a mix of Times New Roman, DejaVu Sans and Econ Sans while the annotations in the graphs are using DejaVu Sans.

4.3 Data Representation

4.3.1 Topic 1: Calorie Intake by Income Deciles (Static Visualisation)

In the first topic we want to emphasise the difference in average calorie intake, measured in kcal, of individuals in the bottom ten percent of income in the UK compared to the other nine income deciles. Since the aim of this visualisation is to compare one decile to another, we want to show all calorie intake of all ten income deciles. We chose a circular bar plot in which polar coordinates are used instead of the regular cartesian coordinates. The circle is divided into ten even sections with identical angles. The thickness of the section represents the calorie intake, the thicker the section is the higher the average calorie intake for that particular decile. In order to make the differences more visible, data is transformed to the power of three.

Fig. 27: Circular bar chart

4.3.2 Topic 2: Percentage Difference in Food Consumption of Decile One (Static Visualisation)

In the second topic we want to show how the diet of individuals in income decile one differs from the average individuals. For the most part, decile one individuals consume less food than average. Since we are only concerned about the deviation, the differences are shown in absolute percentage. Due to limited space, taking into account the infographic is to be shown in one page, we are only showing the top 15 food groups with the highest absolute deviation from average. We chose a bar chart for this data because it can clearly showcase all fifteen food groups and their absolute percentage difference. The x-axis represents the absolute percentage difference while the y axis is the list of food group names. So readers can easily find the percentage difference value, we draw a line every five units ranging from 0-60%.

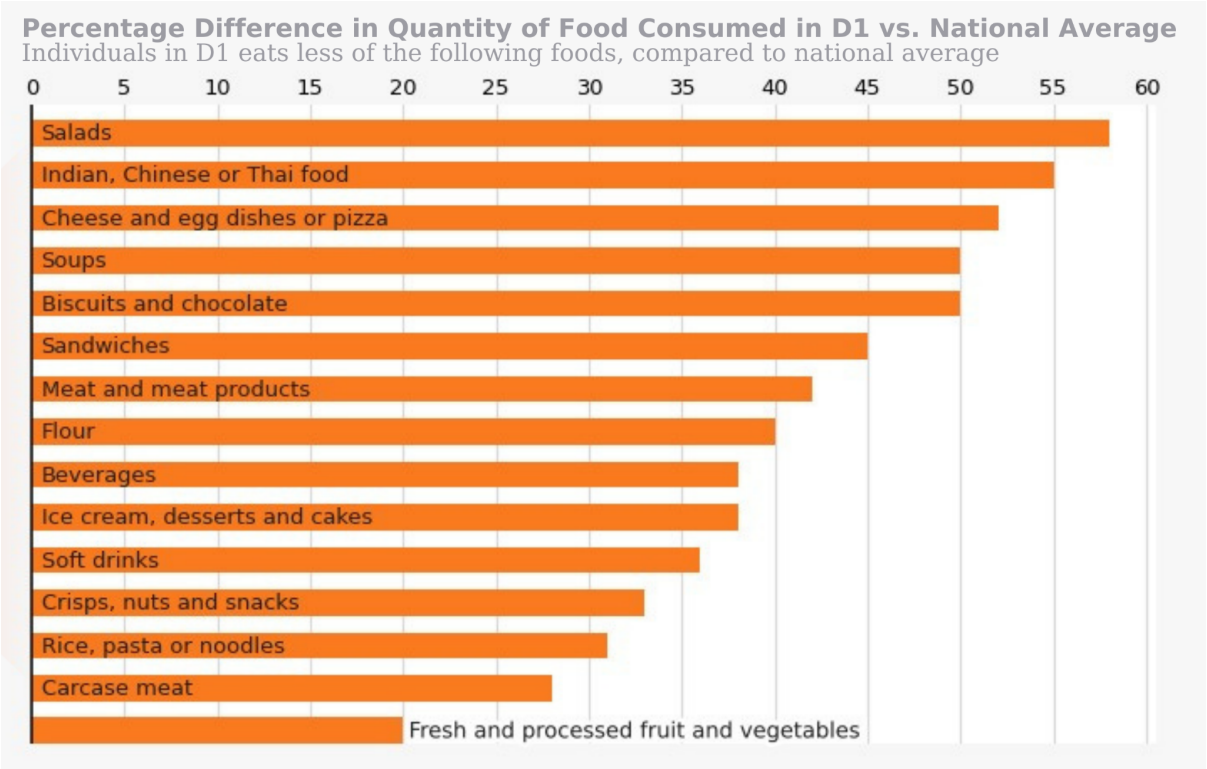


Fig. 28: Percentage difference in food consumption of decile one

4.3.3 Topic 3: Deviation From Average Calorie Intake by Region (Static Visualisation)

In the third topic we want to show how the average calorie intake of the population in each region in the UK deviates from the recommended calorie intake. We will display a map of the UK in which each region is coloured by different shades of red that represents the magnitude of the deviation. The darker the red the further away it deviates from the recommended calorie intake.

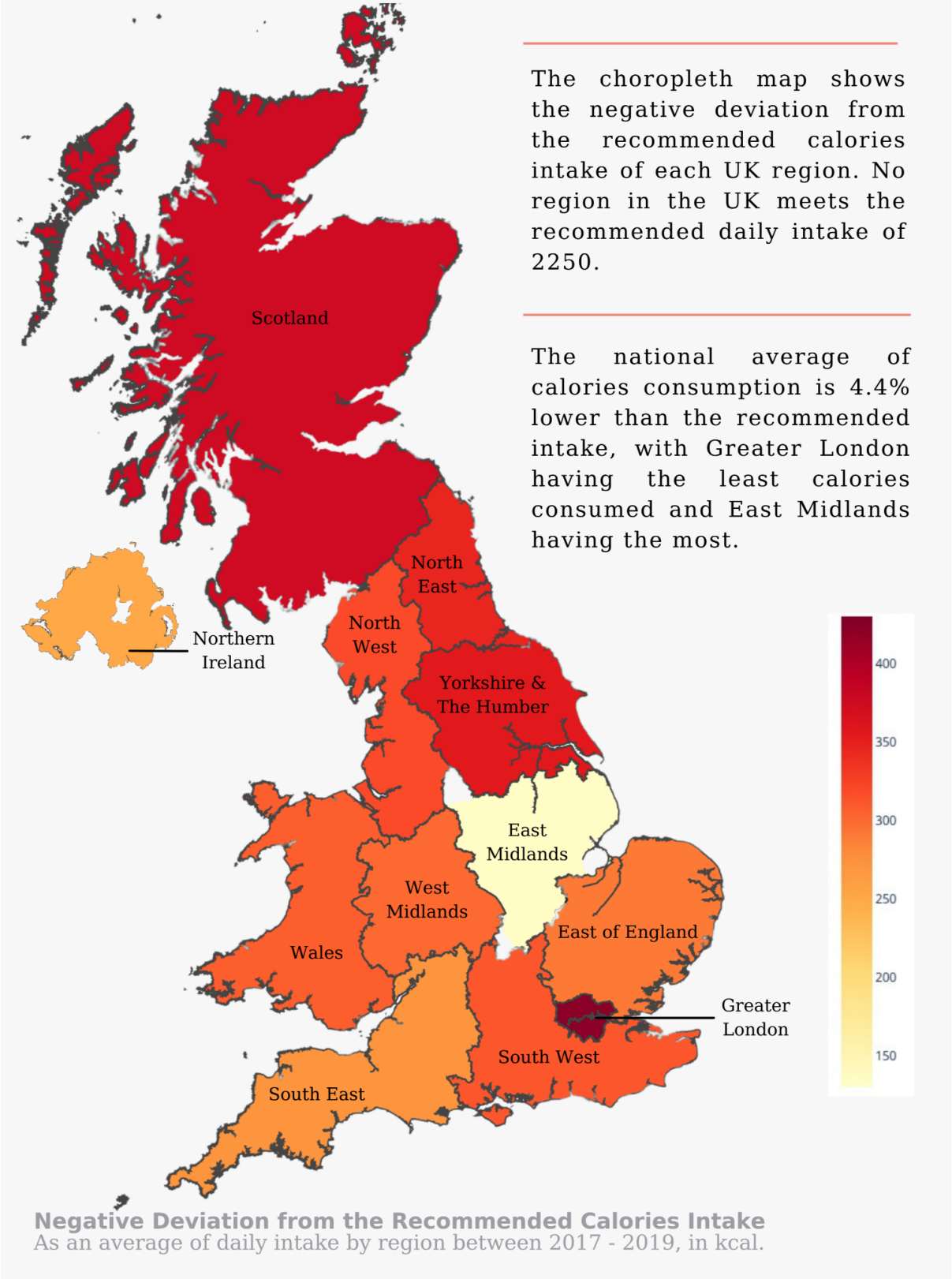


Fig. 29: Static heatmap

4.3.4 Topic 4: Time-series Data of Food Expenditure Patterns in The UK (Dynamic Visualisation)

The fourth topic covers the changes in food expenditure in the UK from 1976 to 2020. The years are grouped into nine brackets each containing a range of five years. The expenditures are categorised into different food groups so readers can see expenditure changes for a specific type of food. We can also filter the graph to only show expenditure values of a specific amount. When the graph is filtered, the data points that lie beyond the range specified will have a different colour than the rest of the data points.

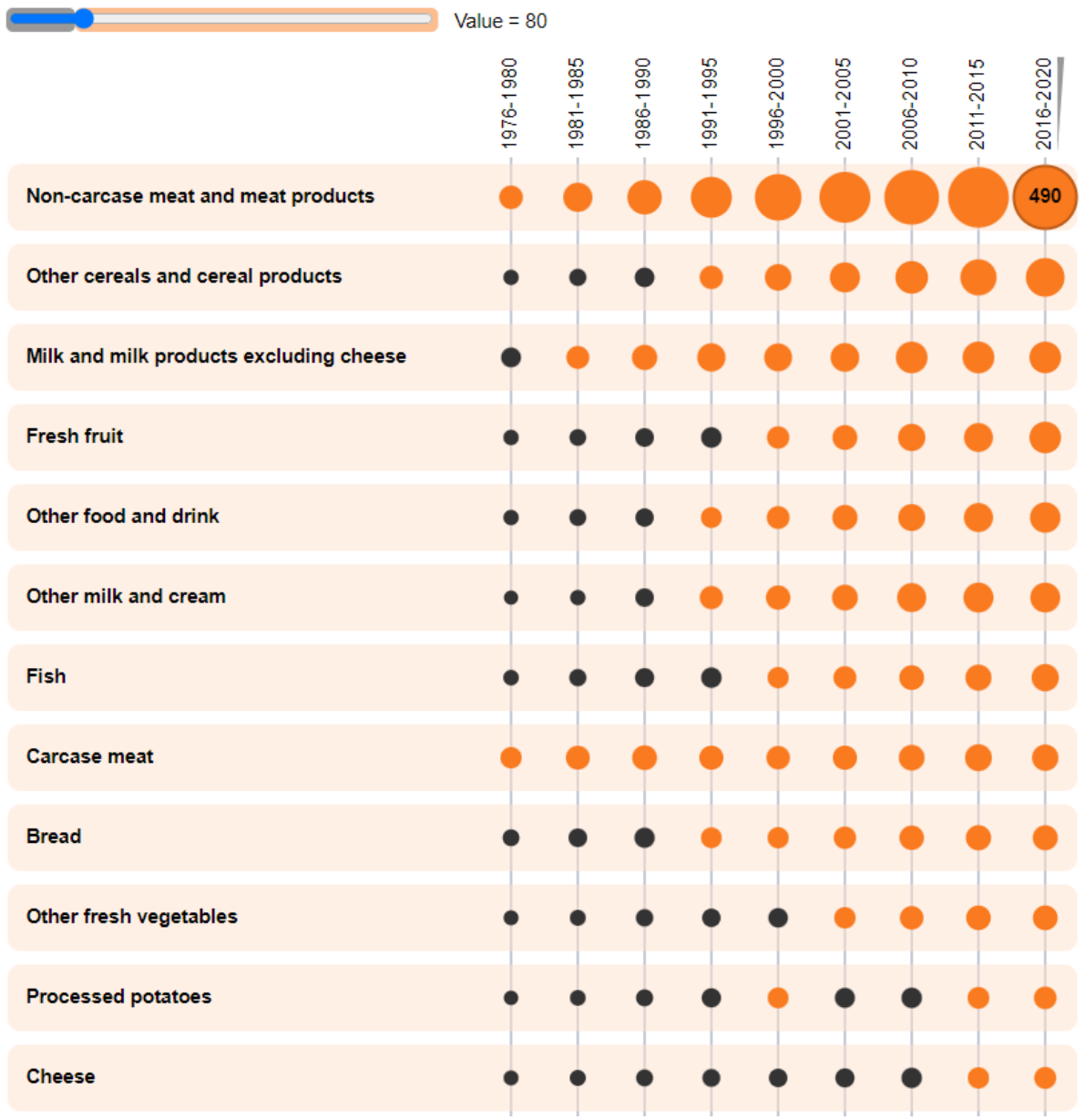


Fig. 30: Interactive bubble chart: Average food expenditure trend by food type (in pounds)

4.3.5 Topic 5: Time-series Data of Food Nutrient Intake by Regions in The UK (Dynamic Visualisation)

In the last topic we want to show, for different nutrient types, what is the average intake compared to the recommended nutrient intake for each year from 2001 to 2019. All nutrient intake is measured as a proportion to the recommended intake at 100%. The graph can be filtered by nutrient type and region so readers can observe the trend for a combination of different food groups and regions.

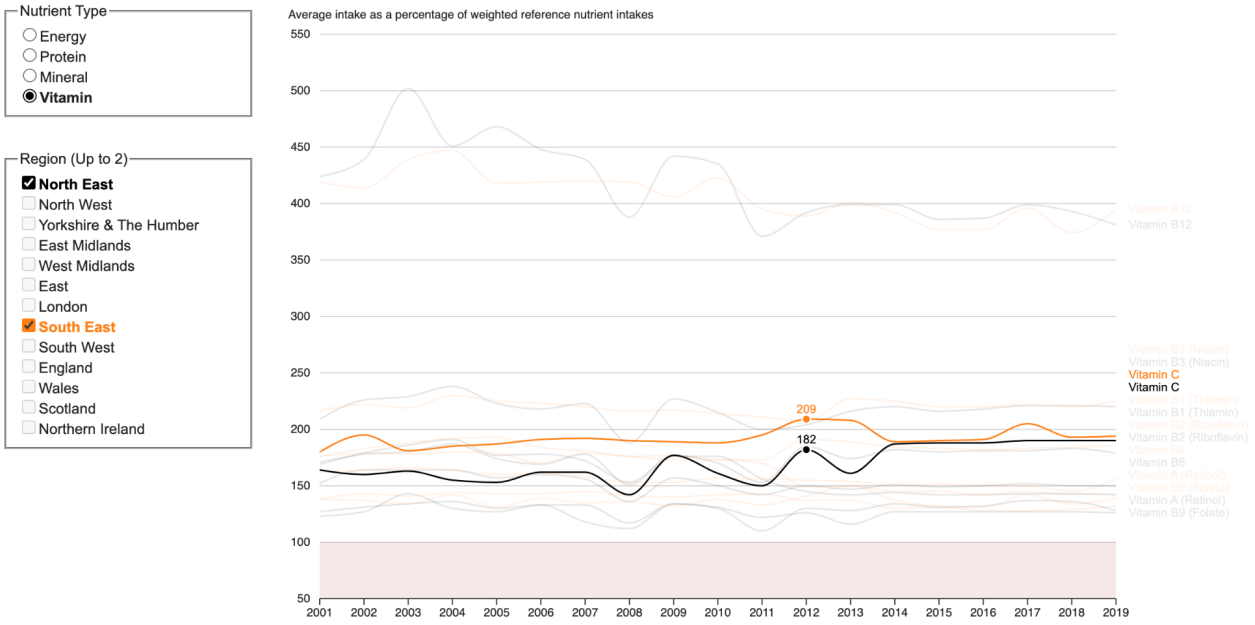


Fig. 31: Interactive Line Chart: Average nutrient intake compared to the suggested reference nutrient intake (in percentage)

4.4 Annotation

4.4.1 Static (magazine)

This visualisation is intended to take place in an infographic featured in *The Economist* magazine, so it has to be self-explanatory. We will have appropriate titles, subtitles and introductory messages.

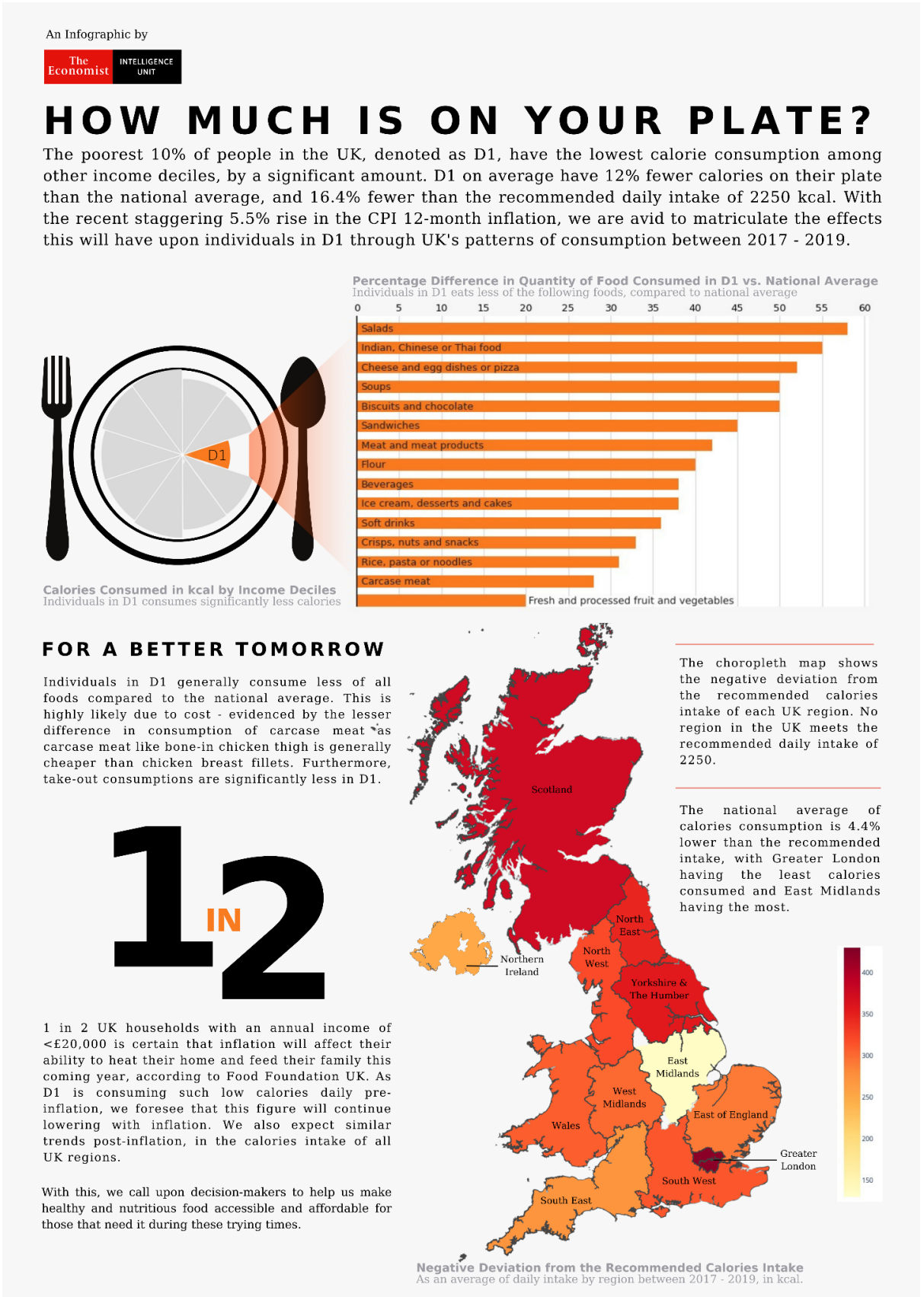


Fig 32: Final Static Visualisation

4.4.2 Dynamic (website)

Our dynamic visualisation will be featured on *The Economist* website. There will be three tabs on the website, the first one is “Home”, the second is “Nutrient Intake”, and the third one is “Expenditure”. The first tab, on which users will initially land on, contains a brief introduction of our overall report describing the topics and data used. The second tab features the line chart on time-series data of food nutrient intake by regions in the UK. The last tab contains a bubble chart on time-series data of food expenditure patterns in the UK.

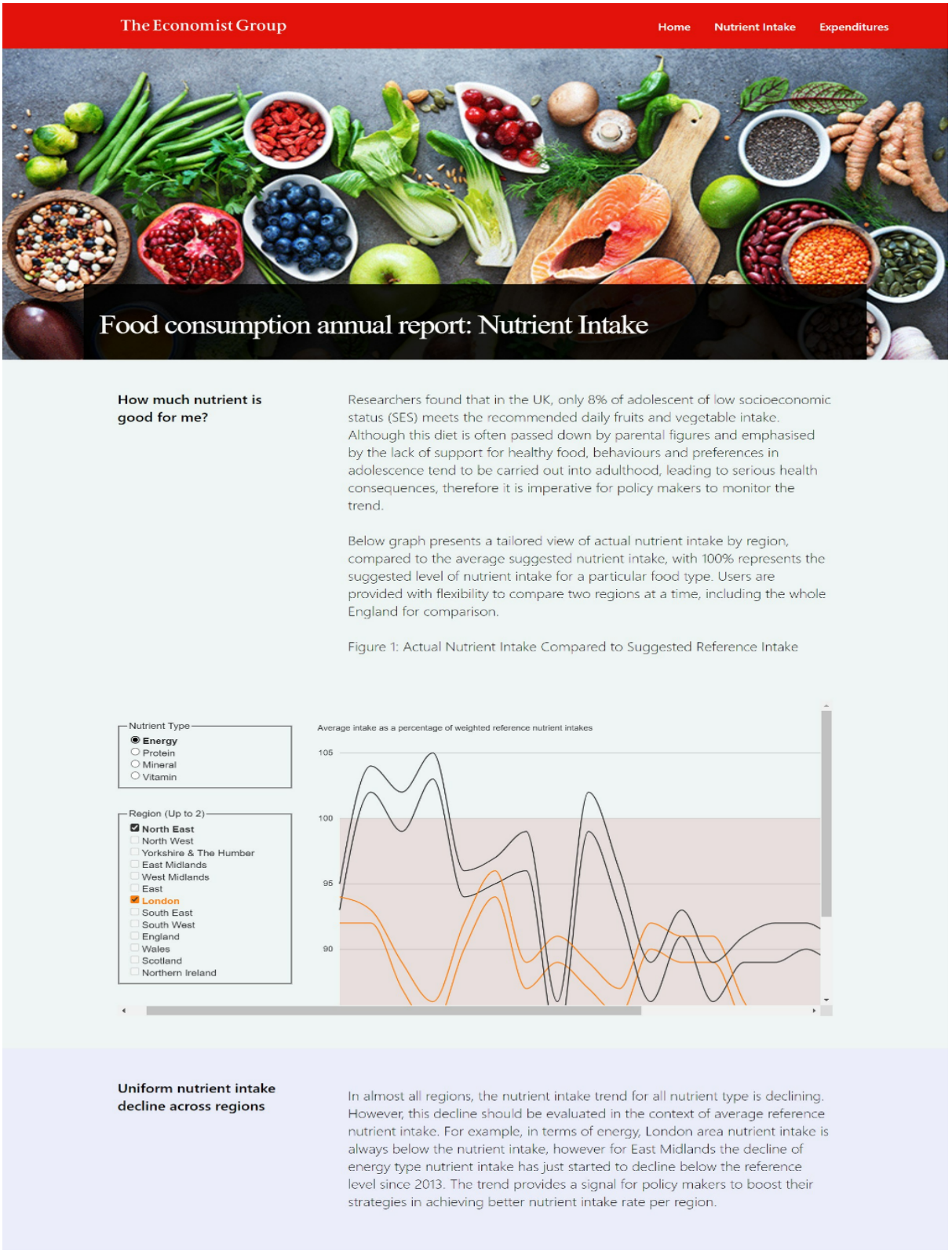


Fig 32: Final Interactive Visualisation on a Website: Nutrient Intake

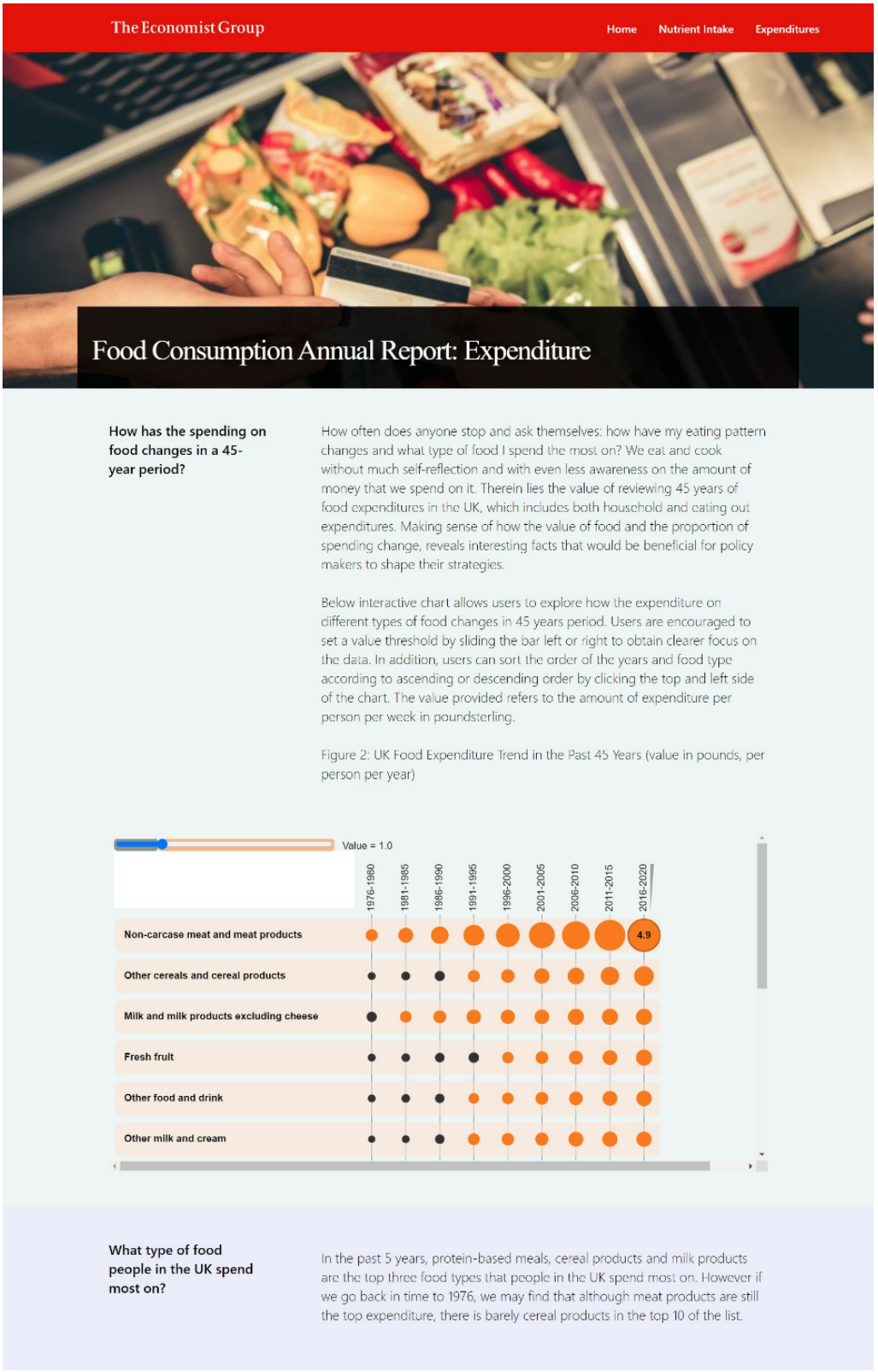


Fig 33: Final Interactive Visualisation on a Website: Expenditure

5.0 References

Aftosmes-Tobio, A., Ganter, C., Gicevic, S., Newlan, S., Simon, C.L., Davison, K.K. and Manganello, J.A. (2016). A systematic review of media parenting in the context of childhood obesity research. *BMC Public Health*. [Online] 16(1). Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4831097/ [Accessed 13 Mar. 2022].

Chorley, M. (2014). *UK-GeoJSON*. [Online] Github.io. Available at: https://martinjc.github.io/UK-GeoJSON/ [Accessed 28 Mar. 2022].

Clark, D. (2021). *UK: population by country*. [Online] Statista. Available at: https://www.statista.com/statistics/294729/uk-population-by-region/ [Accessed 13 Mar. 2022].

CNBC (2020). *Why Grocery Stores Are Avoiding Black Neighborhoods*. *YouTube*. Available at: https://www.youtube.com/watch?v=Rd8J-9uUnfc [Accessed 13 Mar. 2022].

Daily Mail (2018). *The Five Most Healhy Hotspots in England, and the Five Worst*. [Online image] *www.dailymail.co.uk*. Available at: https://www.dailymail.co.uk/health/article-6154557/How-people-5-day-area.html [Accessed 30 Mar. 2022].

Flores, M. (2018). *Flavor of the Day: Analyzing Epicurious.com Recipes - Ceros Inspire*. [Online image] Ceros Inspire: Create, Share, Inspire. Available at: https://www.ceros.com/inspire/originals/epicurious-recipes-incredients/ [Accessed 30 Mar. 2022].

Hall, A. (2019). *Half of British adults have 'no idea' about nutrition, research claims*. [Online] The Independent. Available at: https://www.independent.co.uk/news/uk/home-news/diet-balanced-nutrition-healthy-food-groups-a9202986.html [Accessed 13 Mar. 2022].

Judge, J. (2021). *UK Postcodes*. [online] GitHub. Available at: https://github.com/academe/UK-Postcodes/blob/master/uk-regions.csv [Accessed 28 Mar. 2022].

Mcnair, J. (2019). *Sweetgreen, Shake Shack, and Their Big-Money Backers*. [Online image] Ceros Inspire: Create, Share, Inspire. Available at: https://www.ceros.com/inspire/originals/fast-casual-restaurants-steve-case-danny-meyer/ [Accessed 30 Mar. 2022].

Nations, U. (2019). *UNSD — Methodology*. [Online] Un.org. Available at: https://unstats.un.org/unsd/methodology/m49/ [Accessed 15 Mar. 2022].

nhs.uk. 2022. *What should my daily intake of calories be?*. [online] Available at: <https://www.nhs.uk/common-health-questions/food-and-diet/what-should-my-daily-intake-of-calories-be/> [Accessed 30 March 2022].

Office for National Statistics (2021). *Estimates of the population for the UK, England and Wales, Scotland and Northern Ireland - Office for National Statistics*. [Online] Ons.gov.uk. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland [Accessed 25 Mar. 2022].

Plotly Graphs (2019). *Mapbox County Choropleth*. [Online] plotly.com. Available at: https://plotly.com/python/mapbox-county-choropleth/ [Accessed 28 Mar. 2022].

Postcodebyaddress. (n.d.). *England Postcodes Overview - Postcode by Address*. [Online] Available at: https://postcodebyaddress.co.uk/country/england [Accessed 13 Mar. 2022].

Raconteur (2022). *Feeding the World*. [Online image] *www.raconteur.net*. Available at: https://www.raconteur.net/infographics/feeding-the-world/ [Accessed 30 Mar. 2022].

Read, A. (2016). *55% of Visitors Read Your Articles For 15 Seconds or Less: Why We Should Focus on Attention Not Clicks*. [Online] Buffer Resources. Available at: https://buffer.com/resources/55-visitors-read-articles-15-seconds-less-focus-attention-not-clicks/ [Accessed 13 Mar. 2022].

Scott, H., Craddock, C. and Craig, L. (2020). What features of a nutrition resource are important to adolescents of a low socioeconomic status? *Public Health Nutrition*, 23(10), pp.1716–1725.

Stibbs, C. (2015). There's a reason why at The Economist we don't chase the millennial generation. *The Guardian*. [Online] Available at: https://www.theguardian.com/media-network/2015/mar/04/economist-ceo-chris-stibbs-millennial-generation [Accessed 13 Mar. 2022].

The Economist Design System (2022). *The Economist Design System*. [Online] design-system.economist.com. Available at: https://design-system.economist.com/foundations/colour/palettes [Accessed 30 Mar. 2022].

Wikipedia. (2022). *Crown Dependencies*. [Online] Available at: https://en.wikipedia.org/wiki/Crown_Dependencies [Accessed 13 Mar. 2022].

6.0 Appendix

6.1 Previous considerations

In the second theme, we wanted to communicate the importance of food nutrient intake and alert readers if a certain projection does not seem to be in the right direction. We also wanted to inform the audience about the trend of processed versus fresh food consumption over the years and use this information to drive more action towards a healthier lifestyle.

Additionally, by providing the breakdown of the nutrient groups per region, we can see the difference in food consumption between regions. This would complement our first message by highlighting which food should be delivered to certain regions in order to ensure all regions can meet the recommended calories intake.

The second angle focuses on exploring the relationship between the region's geographical position and nutrients consumed on average per region. The nutrients are categorised into four groups: protein, carbohydrates, fats, and vitamins.

MSIN0167_Data Visualisation_Group Coursework

ORIGINALITY REPORT

2%	1%	1%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.gov.uk Internet Source	1%
2	Submitted to Buckinghamshire Chilterns University College Student Paper	<1%
3	Submitted to Asia Pacific University College of Technology and Innovation (UCTI) Student Paper	<1%
4	"Appendix A: Household Expenditure Tables", Family Spending, 2010 Publication	<1%
5	acikbilim.yok.gov.tr Internet Source	<1%
6	digitalassets.lib.berkeley.edu Internet Source	<1%
7	e-space.mmu.ac.uk Internet Source	<1%
8	www.cambridge.org Internet Source	<1%

9	www.citethisforme.com Internet Source	<1%
10	prev-radicalites.org Internet Source	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		

MSIN0167_Data Visualisation_Group Coursework

GRADEMARK REPORT

FINAL GRADE

71 /100

GENERAL COMMENTS

Instructor

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11

PAGE 12

PAGE 13

PAGE 14

PAGE 15

PAGE 16

PAGE 17