What is a sandwich?

A data analysis in $\mathbb{R}^{43}$

Vincent Macri David White Matthew Stuart

Group D

June 6, 2018

Abstract

In this paper we set out to examine the sandwich views of students and teachers at William Lyon Mackenzie C.I.

Our study consisted of a survey conducted in person with 140 participants. We asked participants basic demographic information about where they fit into the William Lyon Mackenzie Collegiate Institute community, and examined correlations between sandwich views and demographics.

We propose the creation of a two-dimensional sandwich alignment chart, inspired by two-dimensional political axes. The sandwich alignment chart has a dimension for "sandwich purity", and "sandwich orthodoxy".

We hypothesize that:

1. Sandwich purity and sandwich orthodoxy will be positively correlated.
2. Students in the MaCS or Gifted program are more likely to have a low orthodoxy score.

We believe the first hypothesis to be true because we believe those with a pure definition of a sandwich will also have an orthodoxy definition. We believe the second hypothesis to be true because we believe that students in those programs tend to challenge societal norms more so than most.

Contents

Part I

Introduction

Chapter 1

Purpose

1.1 The failure of the dictionary

We find dictionary definitions to be insufficient, as they are either too restrictive, or too vague.

The Oxford English Dictionary [10] restrictively defines a sandwich as:

> An item of food consisting of two pieces of bread with a filling between them, eaten as a light meal.

Whereas The Free Dictionary [31] more broadly defines a sandwich as:

> a. Two or more slices of bread with a filling such as meat or cheese placed between them.
>
> b. A partly split long or round roll containing a filling.
>
> c. One slice of bread covered with a filling.

Since the Oxford English Dictionary definition requires two pieces of bread, this excludes sub sandwiches, which most would consider a sandwich. This makes the Oxford definition too restrictive.

Also, both definitions fail to adequately define “filling”. The Free Dictionary gives meat and cheese as examples, but many people put lettuce and tomato in their sandwiches, neither of which are meat or cheese.

So, dictionary definitions of “sandwich” are insufficient to determine what a sandwich is.

1.2 Legal background

The question of what is a sandwich has been the centre of several legal publications. We believe that these publications have failed to provide a strong definition of what a sandwich is, and they contradict each other.

For tax purposes, the New York State Department of Taxation and Finance [28] says:

> *Sandwiches* include cold and hot sandwiches of every kind that are prepared and ready to be eaten, whether made on bread, on bagels, on rolls, in pitas, in wraps, or otherwise, and regardless of the filling or number of layers. A sandwich can be as simple as a buttered bagel or roll, or as elaborate as a six-foot, toasted submarine sandwich.
>
> Some examples of taxable sandwiches include:
>
> - common sandwiches, such as:
>   - BLTs (bacon, lettuce, and tomato sandwiches);
>   - club sandwiches;
>   - cold cut sandwiches;
>   - grilled cheese sandwiches;
>   - peanut butter and jelly sandwiches
>   - salad-type sandwiches (e.g., chicken, egg, ham, and tuna);
> - bagel sandwiches (served buttered or with spreads, or otherwise as a sandwich);
> - burritos
> - cheese-steak sandwiches;
> - croissant sandwiches;
> - fish fry sandwiches;
> - flatbread sandwiches;

- breakfast sandwiches;
- gyros;
- hamburgers on buns, rolls, etc.;
- heroes, hoagies, torpedoes, grinders, submarines, and other such sandwiches;
- hot dogs and sausages on buns, rolls, etc.;
- melt sandwiches;
- open-faced sandwiches;
- panini sandwiches;
- Reuben sandwiches; and
- wraps and pita sandwiches.

This is a very broad definition, but it is also quite comprehensive and informative. It is important to note that [28] defines burritos as sandwiches. However, other legal cases contradict this definition.

A case in the Commonwealth of Massachusetts Superior Court entitled White City Shopping Center, LP v. PR Restaurants, LLC dba Bread Panera [45] involved two companies in a dispute over whether or not burritos are sandwiches. In this case, the court ruled that burritos are not sandwiches. This contradicts the definition in New York State tax law. So, we can clearly see that there is no legal consensus on this matter.

Furthermore, the legal scholar Marjorie Florestal argues that the decision of the White City case is rooted in classist and racial views of sandwich cuisine [15]:

> The burrito meets resistance not just because of its class but also because of its race—and the way the two play off each other.

So, we have established that the definition of a sandwich is inconclusive among both the linguist and legal communities [10, 31, 28, 45]. The question is also of importance for better understanding class systems and race in our society [15].

Chapter 2

Terminology and definitions

This chapter will explain any terminology and definitions we use throughout this paper.

2.1 Question types

Demographic question A demographic question is any of the non-food questions asked in our survey. These questions asked about participants demographics, in order to examine any correlations between demographics and sandwich views.

Food question A food question is any question asked about sandwiches or their ingredients. We asked 43 food questions as a part of this study. Participants answered each question on a 0 to 10 scale. We purposely started the scale at 0 so as to make 5 exactly in the middle of the range of possible responses.

2.2 Variables

The demographic questions are our independent variables. We mainly look at academic stream, grade, and race as our independent variables.

The food questions are used to calculate orthodoxy and purity, which are our dependent variables.

We also consider the correlation between our two dependent variables to gain a deeper understanding of how they relate.

2.3 Population of interest

Our population of interest is the population of both students and teachers at William Lyon Mackenzie C.I.

Chapter 3

Methodology

3.1 Survey type

We conducted our survey as mix of a stratified, voluntary, and random sample. We surveyed roughly 10% of the William Lyon Mackenzie population. With a sample size this large, most bias should be eliminated.

Unfortunately, we failed to collect a perfectly stratified sample. However, an analysis of the data shows that to be inconsequential.

We made manual changes to categorical to correct for similar, blank, and inappropriate responses. As part of this, we grouped ethnicity into the following 11 categories: Caucasian, Chinese, East Asian, Filipino, Jewish, Korean, Middle Eastern, Mixed, Other, South Asian, and Vietnamese.

Demographic questions

For demographic information, we asked participants for their grade (with teacher as an option), favourite subjects, and ethnic background. We asked students for their academic stream, and teachers for their department.

Since the number of teachers surveyed was small, we do not do any analysis on teachers departments, and instead treat them as a separate grade and academic stream. We also do not analyze the data on favourite subjects since it is very noisy.

Food questions

We asked respondents 43 questions related to sandwiches and their ingredients. We use all of this data.

Chapter 4

Metrics

4.1 Axes metrics

We have two metrics used in our calculations: purity and orthodoxy.

Purity is how pure a respondent's definition of a sandwich is. The less things a respondent considers a sandwich, the greater their purity score will be. Similarly, the more things a respondent considers a sandwich, the lower their purity score will be.

Orthodoxy is a measure of how much a respondent differs from the mean set of responses. The less a respondent's answers differ from the mean set of answers, the greater their orthodoxy score will be. Similarly, the more a respondent's answers differ from the mean set of answers, the lower their orthodoxy score will be.

Both purity and orthodoxy are bound in the range $[-1, 1]$.

We describe the general scoring system in subsection 4.1.1. This scoring system is used to calculate the purity metric described in subsection 4.1.2 on the next page and the orthodoxy metric described in subsection 4.1.3 on page 10.

4.1.1 Scoring

While participants answered each food question on a 0 to 10 scale, it is more convenient to perform calculations using a -5 to 5 scale. We converted responses from the 0 to 10 scale to the -5 to 5 scale by subtracting each response from 5.

Formally, for each response to a food question, we calculate the score for the response by the passing the response through the sandwich spectrum function, defined in 1 on the next page.

Definition 1 (Sandwich spectrum function)**.** The sandwich spectrum function is defined as:

$$s : \{x \in \mathbb{R} \mid 0 \leq x \leq 10\} \rightarrow \{x \in \mathbb{R} \mid -5 \leq x \leq 5\} \text{ by } s(x) = 5 - x \tag{4.1}$$

Due to the format of our survey, all responses are integers, and are mapped to another integer by the sandwich spectrum function. Although, in principle, the sandwich spectrum function works for real numbers as well.

We can create a table for s:

x	0	1	2	3	4	5	6	7	8	9	10
$s(x)$	5	4	3	2	1	0	-1	-2	-3	-4	-5

One will note that this gives responses that were originally high a lower score. This is intentional. Subsection 4.1.2 will show it to be useful for calculating the purity metric, and subsection 4.1.3 on the next page will show it to be irrelevant for calculating the orthodoxy metric.

4.1.2 Purity

The purity score for a respondent is defined as the sum of a respondent's scores divided by the maximum possible score.

The maximum score for a question is 5, and there are 43 food questions. This means that the maximum possible score is $43 \times 5 = 215$.
Definition 2 (Sadwich purity function)**.** For a given response with a set of 43 food answers, A, we define the sandwich purity function as:

$$p : \mathbb{R}^{43} \rightarrow \mathbb{R} \text{ by } p(A) = \frac{\sum_{i=1}^{43} A_i}{215} \tag{4.2}$$

This definition illustrates why we subtract each response from 5 to get the score. The sandwich spectrum function will assign higher purity values for lower responses. Since a lower response to a question implies a more pure definition of a sandwich, 1 is a valid metric.

4.1.3 Orthodoxy

To calculate orthodoxy for each respondent, we take the score for each question as a dimension of a vector, which creates a vector in 43-dimensional Euclidean space.

We also calculate the mean response for each question, and create an additional $\mathbb{R}^{43}$ vector from that. This vector is referred to as the mean vector, and denoted as $\vec{m}$.

The orthodoxy score for a respondent is defined as the cosine similarity between the respondent's $\mathbb{R}^{43}$ vector and $\vec{m}$.

Definition 3 (Mean vector)**.** To calculate the value of the mean vector, $\vec{m}$:

Let $A \in \mathbb{R}^{43}$ be the set of response vectors. Then, $n(A)$ is the cardinality of the set A.

Then:

$$\vec{m} = \frac{\sum_{i=1}^{n(A)} A_i}{n(A)} \tag{4.3}$$

Definition 4 (Sandwich orthodoxy function)**.** To calculate the orthodoxy score for a respondent:

Let $\vec{r}$ be a vector in $\mathbb{R}^{43}$ defined as having its each of its components equal to the score for each food question.

$$o : \mathbb{R}^{43} \to \mathbb{R} \text{ by } o(\vec{r}) = \frac{\vec{r} \cdot \vec{m}}{\|\vec{r}\|\|\vec{m}\|} \tag{4.4}$$

Since we are calculating orthodoxy as the cosine of the angle between two vectors, it is useful to have some vector components be negative, as that allows respondents to have a negative orthodoxy score if they answer opposite to the mean response. It does not matter what direction the vectors are in, as we are only looking at the angle between them. This means that the sandwich spectrum function could have been defined as subtracting 5 instead of subtracting from 5 for the purposes of the orthodoxy function. Since both definitions would have worked for orthodoxy, we stick with 1 on the previous page for the sake of consistency with the purity metric.

Part II

Results Summary

Chapter 5

Results summary

All numbers in this chapter have been rounded to 5 digits after the decimal.

5.1 Data with outliers

Table 5.1: The results of all respondents with outliers included. Teachers had their stream set to “Teacher”.

#	Purity	Orthodoxy	Grade	Background	Stream
1	−0.26047	0.78743	Grade 12	Mixed	Gifted
2	0.04186	0.90389	Grade 12	Filipino	Academic
3	−0.70233	0.41500	Grade 11	Chinese	MaCS
4	−0.66512	0.46478	Grade 11	Caucasian	MaCS
5	0.04651	0.87398	Grade 12	Chinese	MaCS
6	0.12558	0.76755	Grade 12	East Asian	MaCS
7	−0.29767	0.04865	Grade 12	Chinese	MaCS
8	−0.16279	0.89784	Grade 12	Jewish	MaCS
9	−0.17674	0.62968	Grade 12	Middle Eastern	Academic
10	−0.50233	0.67579	Grade 12	Filipino	MaCS
11	−0.05581	0.92452	Grade 12	Mixed	MaCS
12	0.18605	0.83489	Teacher	Jewish	Teacher
13	0.13488	0.81080	Grade 12	Middle Eastern	Gifted
14	−0.67442	0.39360	Grade 12	South Asian	MaCS
15	0.11628	0.78018	Grade 12	Chinese	MaCS
16	0.12093	0.83977	Grade 11	Korean	MaCS
17	−0.55814	0.54878	Grade 11	Chinese	MaCS
18	−0.44186	0.60770	Grade 11	Chinese	MaCS
19	0.13023	0.82374	Grade 12	East Asian	MaCS
20	−0.25581	0.79897	Grade 12	Chinese	MaCS
21	−0.11628	0.85328	Grade 12	Vietnamese	Academic

22	−0.03256	0.90503	Grade 12	Jewish	Academic
23	−0.03256	0.81788	Grade 12	South Asian	Academic
24	−0.20000	0.72832	Grade 11	Vietnamese	Other
25	0.15814	0.85950	Grade 12	Filipino	Academic
26	−0.09767	0.75217	Grade 12	Jewish	Gifted
27	0.13953	0.80860	Teacher	Other	Teacher
28	0.22791	0.62137	Grade 12	Chinese	MaCS
29	−0.28837	0.84233	Grade 12	East Asian	MaCS
30	−0.20930	0.89615	Grade 12	Caucasian	MaCS
31	0.08372	0.80590	Grade 12	Jewish	MaCS
32	−0.42791	0.72797	Grade 12	Chinese	MaCS
33	−0.14419	0.87928	Grade 12	Caucasian	MaCS
34	0.09302	0.82863	Grade 12	Vietnamese	MaCS
35	−0.58605	0.56467	Grade 11	Caucasian	MaCS
36	−0.02791	0.79334	Grade 12	South Asian	MaCS
37	0.00465	0.85846	Teacher	Other	Teacher
38	−0.16279	0.69499	Grade 12	Caucasian	MaCS
39	−0.06047	0.91415	Grade 12	Caucasian	MaCS
40	−0.24186	0.76056	Grade 12	Jewish	MaCS
41	−0.53023	0.54403	Grade 11	Other	MaCS
42	−0.06047	0.77035	Grade 12	South Asian	Applied
43	0.08837	0.67952	Grade 12	Mixed	Academic
44	0.05116	0.87279	Grade 12	Filipino	Academic
45	−0.14419	0.64695	Grade 9	Caucasian	Academic
46	−0.24186	0.64867	Grade 12	Caucasian	Academic
47	−0.13953	0.73419	Grade 9	East Asian	Academic
48	0.27907	0.80968	Grade 12	Jewish	Academic
49	−0.04651	0.81095	Grade 12	Jewish	Academic
50	−0.30698	0.67331	Grade 12	East Asian	Academic
51	0.21395	0.30716	Grade 12	Caucasian	Academic
52	−0.80000	0.23213	Grade 12	Caucasian	Academic
53	0.38605	0.46599	Grade 12	Vietnamese	Academic
54	0.25581	0.52316	Grade 12	Other	Other
55	−0.14419	0.77929	Grade 9	Jewish	Academic
56	0.17674	0.82393	Grade 12	Mixed	Academic
57	0.24651	0.75861	Grade 12	Caucasian	Applied
58	0.40465	0.56366	Grade 12	Mixed	Academic
59	0.17674	0.13833	Grade 11	Other	Academic

60	−0.29767	0.78865	Teacher	Caucasian	Teacher
61	−0.01395	0.59180	Grade 11	Other	Applied
62	−0.04651	0.44875	Grade 12	Korean	Academic
63	0.00465	0.79811	Grade 12	Filipino	Academic
64	−0.66512	−0.13785	Grade 12	Korean	Gifted
65	−0.15349	0.52585	Grade 12	Filipino	Academic
66	−0.00930	0.86197	Grade 12	Caucasian	Academic
67	−0.07907	0.55806	Grade 12	South Asian	MaCS
68	−0.57674	0.52544	Grade 11	Caucasian	Academic
69	0.10698	0.49637	Grade 12	Jewish	Academic
70	−0.04186	0.84955	Grade 12	Caucasian	Gifted
71	−0.81395	0.29688	Grade 12	Chinese	MaCS
72	0.07907	0.77065	Grade 12	Caucasian	Gifted
73	0.25116	0.81689	Grade 11	Korean	MaCS
74	0.22791	0.79962	Grade 11	South Asian	MaCS
75	−0.03721	0.83146	Grade 12	East Asian	Gifted
76	0.12558	0.81047	Grade 11	Other	Academic
77	−0.24651	0.87283	Grade 11	Caucasian	MaCS
78	−0.41860	0.52009	Grade 11	Jewish	Gifted
79	−0.42791	0.52045	Grade 11	Other	Gifted
80	−0.37209	0.44516	Grade 11	Vietnamese	Gifted
81	−0.01395	−0.20124	Grade 11	Caucasian	Other
82	0.33023	0.38251	Grade 9	Other	Other
83	0.16279	0.01988	Grade 9	Caucasian	Other
84	0.26047	0.80100	Grade 9	Mixed	MaCS
85	0.07907	0.63674	Grade 12	Middle Eastern	Academic
86	−0.17209	0.72197	Grade 12	Other	Academic
87	0.09767	0.66355	Grade 12	Chinese	Academic
88	−0.01860	0.83655	Teacher	Other	Teacher
89	−0.12558	0.83448	Grade 11	Caucasian	Applied
90	0.08837	0.76628	Grade 10	Other	Academic
91	0.10698	0.70924	Grade 9	Mixed	Academic
92	−0.17674	0.82363	Grade 11	Mixed	MaCS
93	−0.33488	0.66008	Grade 11	Chinese	MaCS
94	−0.51163	0.54699	Grade 11	Mixed	MaCS
95	−0.23721	0.82223	Grade 12	East Asian	Academic
96	0.03256	0.87361	Grade 9	South Asian	MaCS
97	−0.05581	0.82581	Grade 12	Caucasian	Academic

98	0.39535	0.61396	Grade 10	Jewish	MaCS
99	0.15349	0.83300	Grade 10	Caucasian	MaCS
100	0.33488	0.75085	Grade 11	Korean	Gifted
101	0.04186	0.84494	Grade 11	South Asian	Gifted
102	0.03721	0.88977	Grade 12	Jewish	MaCS
103	−0.03721	−0.03783	Grade 11	Filipino	Academic
104	−0.28372	0.82859	Grade 11	Mixed	Academic
105	0.17674	0.85856	Grade 12	Other	Academic
106	0.15814	0.70685	Grade 12	Middle Eastern	Academic
107	0.04186	0.81200	Grade 12	Caucasian	Academic
108	−0.22791	0.62510	Grade 10	Middle Eastern	MaCS
109	0.00930	0.89270	Grade 9	East Asian	MaCS
110	0.23721	0.79064	Grade 9	Jewish	MaCS
111	0.85581	−0.02167	Grade 11	South Asian	Academic
112	−0.25581	0.88868	Grade 12	Chinese	MaCS
113	0.14419	0.33265	Grade 11	Other	MaCS
114	0.27442	0.74374	Grade 9	Other	MaCS
115	−0.33488	0.65545	Other	Other	Other
116	−0.34419	0.80198	Grade 10	Jewish	MaCS
117	0.13023	0.82970	Grade 11	Mixed	MaCS
118	−0.53488	0.63103	Grade 9	Chinese	Gifted
119	0.40000	0.62724	Grade 11	Middle Eastern	Academic
120	0.14884	0.22078	Grade 12	Caucasian	MaCS
121	−0.17209	0.83039	Grade 10	Mixed	MaCS
122	0.32093	0.77196	Grade 12	Filipino	Academic
123	−0.04186	0.81574	Grade 10	East Asian	MaCS
124	0.20930	0.79373	Grade 12	Other	MaCS
125	−0.06047	0.40400	Grade 9	Other	MaCS
126	0.21395	0.80349	Grade 11	Caucasian	MaCS
127	−0.01395	0.81005	Grade 12	Mixed	MaCS
128	−0.01860	0.89398	Grade 10	South Asian	MaCS
129	0.25581	0.45210	Grade 10	Korean	MaCS
130	0.23256	0.31299	Grade 10	South Asian	MaCS
131	0.19070	0.82500	Grade 11	Other	Academic
132	0.40930	0.70170	Grade 10	Filipino	Academic
133	−0.16279	0.73321	Grade 9	East Asian	Gifted
134	0.11163	0.72394	Grade 11	Caucasian	MaCS
135	−0.03721	0.77001	Grade 12	Other	Other

136	−0.07442	0.85982	Grade 10	Korean	Gifted
137	0.26047	0.83633	Grade 11	Caucasian	MaCS
138	0.44651	0.56466	Grade 11	South Asian	MaCS
139	0.22791	0.56602	Grade 11	Vietnamese	Gifted
140	0.01860	0.78616	Grade 12	South Asian	MaCS

5.2 Data without outliers

See chapter 6 on page 22 for an explanation of how we identified and removed outliers. The numbers given for removed outliers match with those in table 5.1 on page 12.

Table 5.2: The results of all respondents with outliers removed. Teachers had their stream set to "Teacher".

#	Purity	Orthodoxy	Grade	Background	Stream
1	−0.26047	0.78968	Grade 12	Mixed	Gifted
2	0.04186	0.90595	Grade 12	Filipino	Academic
3	−0.70233	0.42474	Grade 11	Chinese	MaCS
4	−0.66512	0.47101	Grade 11	Caucasian	MaCS
5	0.04651	0.87742	Grade 12	Chinese	MaCS
6	0.12558	0.77250	Grade 12	East Asian	MaCS
8	−0.16279	0.90018	Grade 12	Jewish	MaCS
9	−0.17674	0.63965	Grade 12	Middle Eastern	Academic
10	−0.50233	0.67820	Grade 12	Filipino	MaCS
11	−0.05581	0.92256	Grade 12	Mixed	MaCS
12	0.18605	0.84173	Teacher	Jewish	Teacher
13	0.13488	0.81892	Grade 12	Middle Eastern	Gifted
14	−0.67442	0.40191	Grade 12	South Asian	MaCS
15	0.11628	0.78026	Grade 12	Chinese	MaCS
16	0.12093	0.85531	Grade 11	Korean	MaCS
17	−0.55814	0.55359	Grade 11	Chinese	MaCS
18	−0.44186	0.61922	Grade 11	Chinese	MaCS
19	0.13023	0.81850	Grade 12	East Asian	MaCS
20	−0.25581	0.80421	Grade 12	Chinese	MaCS
21	−0.11628	0.84608	Grade 12	Vietnamese	Academic

22	−0.03256	0.91319	Grade 12	Jewish	Academic
23	−0.03256	0.81780	Grade 12	South Asian	Academic
24	−0.20000	0.73308	Grade 11	Vietnamese	Other
25	0.15814	0.86354	Grade 12	Filipino	Academic
26	−0.09767	0.77076	Grade 12	Jewish	Gifted
27	0.13953	0.82063	Teacher	Other	Teacher
28	0.22791	0.60112	Grade 12	Chinese	MaCS
29	−0.28837	0.84816	Grade 12	East Asian	MaCS
30	−0.20930	0.89411	Grade 12	Caucasian	MaCS
31	0.08372	0.79274	Grade 12	Jewish	MaCS
32	−0.42791	0.73105	Grade 12	Chinese	MaCS
33	−0.14419	0.88462	Grade 12	Caucasian	MaCS
34	0.09302	0.83049	Grade 12	Vietnamese	MaCS
35	−0.58605	0.56729	Grade 11	Caucasian	MaCS
36	−0.02791	0.79916	Grade 12	South Asian	MaCS
37	0.00465	0.86234	Teacher	Other	Teacher
38	−0.16279	0.69419	Grade 12	Caucasian	MaCS
39	−0.06047	0.92165	Grade 12	Caucasian	MaCS
40	−0.24186	0.75200	Grade 12	Jewish	MaCS
41	−0.53023	0.55661	Grade 11	Other	MaCS
42	−0.06047	0.77395	Grade 12	South Asian	Applied
43	0.08837	0.67583	Grade 12	Mixed	Academic
44	0.05116	0.88368	Grade 12	Filipino	Academic
45	−0.14419	0.66187	Grade 9	Caucasian	Academic
46	−0.24186	0.63411	Grade 12	Caucasian	Academic
47	−0.13953	0.73373	Grade 9	East Asian	Academic
48	0.27907	0.81327	Grade 12	Jewish	Academic
49	−0.04651	0.82498	Grade 12	Jewish	Academic
50	−0.30698	0.66416	Grade 12	East Asian	Academic
52	−0.80000	0.23584	Grade 12	Caucasian	Academic
53	0.38605	0.45733	Grade 12	Vietnamese	Academic
54	0.25581	0.52042	Grade 12	Other	Other
55	−0.14419	0.76742	Grade 9	Jewish	Academic
56	0.17674	0.82218	Grade 12	Mixed	Academic
57	0.24651	0.75231	Grade 12	Caucasian	Applied
58	0.40465	0.55844	Grade 12	Mixed	Academic
60	−0.29767	0.78683	Teacher	Caucasian	Teacher
61	−0.01395	0.57851	Grade 11	Other	Applied

63	0.00465	0.80447	Grade 12	Filipino	Academic
65	−0.15349	0.54823	Grade 12	Filipino	Academic
66	−0.00930	0.87140	Grade 12	Caucasian	Academic
67	−0.07907	0.55018	Grade 12	South Asian	MaCS
68	−0.57674	0.53140	Grade 11	Caucasian	Academic
70	−0.04186	0.86196	Grade 12	Caucasian	Gifted
71	−0.81395	0.30360	Grade 12	Chinese	MaCS
72	0.07907	0.75029	Grade 12	Caucasian	Gifted
73	0.25116	0.82084	Grade 11	Korean	MaCS
74	0.22791	0.79648	Grade 11	South Asian	MaCS
75	−0.03721	0.84123	Grade 12	East Asian	Gifted
76	0.12558	0.81990	Grade 11	Other	Academic
77	−0.24651	0.87925	Grade 11	Caucasian	MaCS
78	−0.41860	0.53666	Grade 11	Jewish	Gifted
79	−0.42791	0.53284	Grade 11	Other	Gifted
80	−0.37209	0.45348	Grade 11	Vietnamese	Gifted
84	0.26047	0.79445	Grade 9	Mixed	MaCS
85	0.07907	0.64001	Grade 12	Middle Eastern	Academic
86	−0.17209	0.70523	Grade 12	Other	Academic
87	0.09767	0.68654	Grade 12	Chinese	Academic
88	−0.01860	0.83619	Teacher	Other	Teacher
89	−0.12558	0.82854	Grade 11	Caucasian	Applied
90	0.08837	0.77318	Grade 10	Other	Academic
91	0.10698	0.70861	Grade 9	Mixed	Academic
92	−0.17674	0.82523	Grade 11	Mixed	MaCS
93	−0.33488	0.65441	Grade 11	Chinese	MaCS
94	−0.51163	0.55424	Grade 11	Mixed	MaCS
95	−0.23721	0.82289	Grade 12	East Asian	Academic
96	0.03256	0.87637	Grade 9	South Asian	MaCS
97	−0.05581	0.83830	Grade 12	Caucasian	Academic
98	0.39535	0.59620	Grade 10	Jewish	MaCS
99	0.15349	0.83942	Grade 10	Caucasian	MaCS
100	0.33488	0.74844	Grade 11	Korean	Gifted
101	0.04186	0.84523	Grade 11	South Asian	Gifted
102	0.03721	0.88892	Grade 12	Jewish	MaCS
104	−0.28372	0.83573	Grade 11	Mixed	Academic
105	0.17674	0.86396	Grade 12	Other	Academic
106	0.15814	0.69246	Grade 12	Middle Eastern	Academic

107	0.04186	0.80980	Grade 12	Caucasian	Academic
108	−0.22791	0.62789	Grade 10	Middle Eastern	MaCS
109	0.00930	0.88855	Grade 9	East Asian	MaCS
110	0.23721	0.78608	Grade 9	Jewish	MaCS
111	0.85581	−0.03661	Grade 11	South Asian	Academic
112	−0.25581	0.88833	Grade 12	Chinese	MaCS
114	0.27442	0.75634	Grade 9	Other	MaCS
115	−0.33488	0.66072	Other	Other	Other
116	−0.34419	0.80379	Grade 10	Jewish	MaCS
117	0.13023	0.83008	Grade 11	Mixed	MaCS
118	−0.53488	0.62596	Grade 9	Chinese	Gifted
119	0.40000	0.61374	Grade 11	Middle Eastern	Academic
121	−0.17209	0.83336	Grade 10	Mixed	MaCS
122	0.32093	0.77667	Grade 12	Filipino	Academic
123	−0.04186	0.80179	Grade 10	East Asian	MaCS
124	0.20930	0.78943	Grade 12	Other	MaCS
126	0.21395	0.80077	Grade 11	Caucasian	MaCS
127	−0.01395	0.80887	Grade 12	Mixed	MaCS
128	−0.01860	0.90132	Grade 10	South Asian	MaCS
129	0.25581	0.43961	Grade 10	Korean	MaCS
131	0.19070	0.81749	Grade 11	Other	Academic
132	0.40930	0.69923	Grade 10	Filipino	Academic
133	−0.16279	0.73236	Grade 9	East Asian	Gifted
134	0.11163	0.72956	Grade 11	Caucasian	MaCS
135	−0.03721	0.76334	Grade 12	Other	Other
136	−0.07442	0.86029	Grade 10	Korean	Gifted
137	0.26047	0.83846	Grade 11	Caucasian	MaCS
138	0.44651	0.54854	Grade 11	South Asian	MaCS
139	0.22791	0.55018	Grade 11	Vietnamese	Gifted
140	0.01860	0.79616	Grade 12	South Asian	MaCS

5.3 Summary with outliers

Table 5.3: A summary of all respondents with outliers included.

Purity	Orthodoxy
Min.:−0.81395	Min.:−0.20120
1st Qu.:−0.20233	1st Qu.:0.56570
Median:−0.01628	Median:0.76880
Mean:−0.04259	Mean:0.67580
3rd Qu.:0.15000	3rd Qu.:0.82650
Max.:0.85581	Max.:0.92450

5.4 Summary without outliers

Table 5.4: A summary of all respondents with outliers removed.

Purity	Orthodoxy
Min.:−0.81395	Min.:−0.03661
1st Qu.:−0.22326	1st Qu.:0.64361
Median:−0.02326	Median:0.78646
Mean:−0.05046	Mean:0.73032
3rd Qu.:0.13372	3rd Qu.:0.83514
Max.:0.85581	Max.:0.92256

Part III

Analysis

Chapter 6

Outliers

We begin our analysis by plotting the purity vs orthodoxy and performing a linear, quadratic, and locally weighted analysis in figure 6.1.

The solid red curve is the result of the locally weighted analysis, and the shaded area is the 99% confidence interval for that analysis. The dashed black line is the result of the linear line of best fit. The dotted blue line is the result of the quadratic curve of best fit.

6.1 With outliers

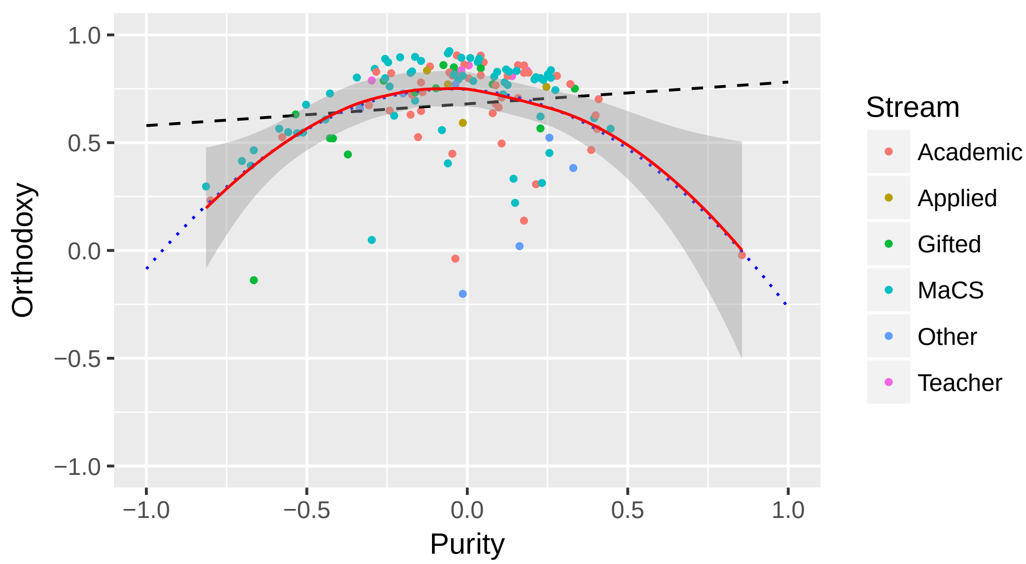

Figure 6.1: A plot of purity vs orthodoxy with data points coloured based on stream.

The linear regression has a positive slope, which proves the first hypothesis that purity and orthodoxy are positively correlated. We will investigate the strength and significance of the correlation in chapter 7 on page 25.

6.2 Identifying outliers

From the this analysis, it is clear that some points are extremely deviate. We next create a residual plot using the quadratic model in figure 6.2.

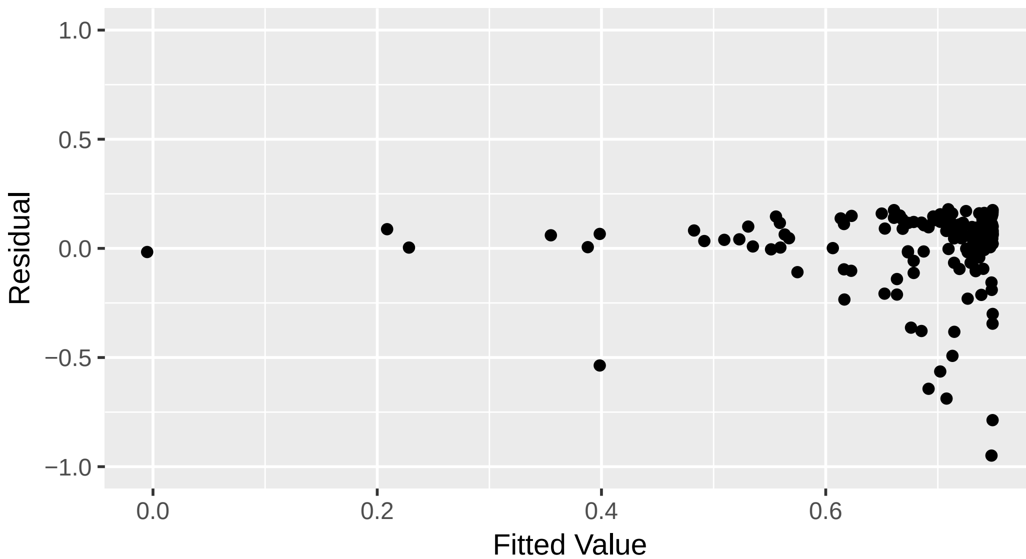

Figure 6.2: A residual plot using the quadratic regression analysis method.

We calculate the interquartile range of the residuals and remove any data that are 1.5 times the ICQ above the third quartile or 1.5 times the ICQ below the first quartile. This is done using the program in section A.8 on page 33. This program then creates a new file for analyzing with the following respondents removed: 7, 51, 59, 62, 64, 69, 81, 82, 83, 103, 113, 120, 125, and 130.

6.3 Outliers removed

We once again plot orthodoxy vs purity by stream in figure 6.3, this time with the outliers removed.

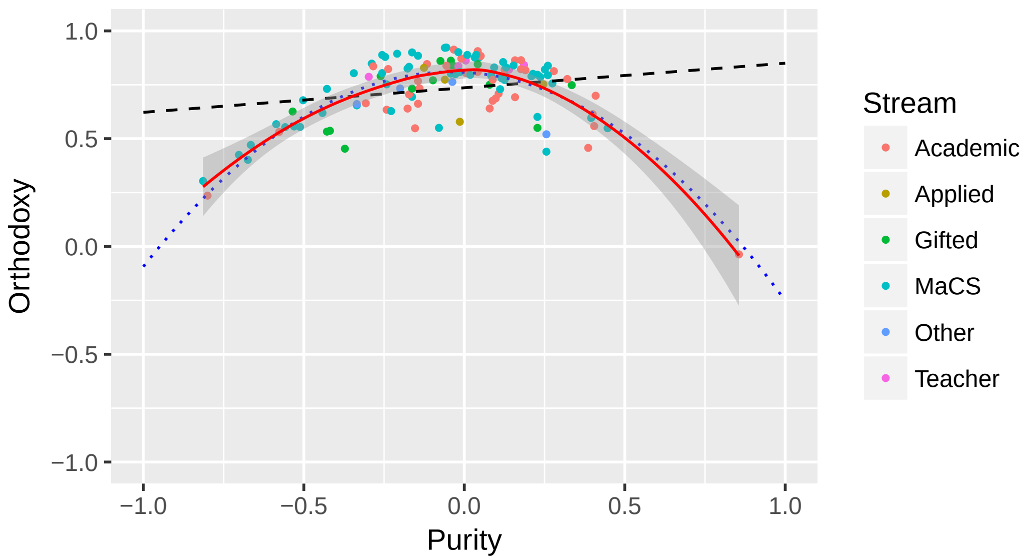

Figure 6.3: A plot of purity vs orthodoxy with data points coloured based on stream. Outliers have been removed.

The solid red curve is the result of the locally weighted analysis, and the shaded area is the 99% confidence interval for that analysis. The dashed black line is the result of the linear line of best fit. The dotted blue line is the result of the quadratic curve of best fit.

6.4 Effect of outliers on the data

Removing the outliers does not significantly effect the curves due to the high sample size. However, it does greatly increase the accuracy of the model and only a handful of points were removed.

Chapter 7

Significance

Part IV

Appendices

Appendix A

Source Code

A.1 Functions.R

```r
# Calculate the cosine similarity between two vectors.
cosineSimilarity <- function(a, b) {
        return (sum(a * b) / (sqrt(sum(a^2)) * sqrt(sum(b^2))))
}
```

A.2 Calculate.R

```r
# This is a program to calculate the orthodoxy and purity scores on the sandwich spectrum for all respondents.

source("Functions.R")

# Define the constants for what columns data is in inside the CSV file.
GRADE_COLUMN <- 1 # What column the grade data is in.
SUBJECTS_COLUMN <- 2 # What column the favourite subjects data is in.
BACKGROUND_COLUMN <- 3 # What column the ethnic background data is in.
STREAM_COLUMN <- 4 # What column the stream data is in.
QUESTIONS_START <- 6 # What column the food questions start at.
QUESTIONS_END <- 48 # What column the food questions end at.
lockBinding("GRADE_COLUMN", globalenv())
lockBinding("SUBJECTS_COLUMN", globalenv())
lockBinding("BACKGROUND_COLUMN", globalenv())
lockBinding("STREAM_COLUMN", globalenv())
lockBinding("QUESTIONS_START", globalenv())
lockBinding("QUESTIONS_END", globalenv())

```

```r
respondents <- read.csv("CleanedData.csv", check.names = FALSE) # Read in the CSV with outliers.
#respondents <- read.csv("NoOutliers.csv", check.names = FALSE) # Read in the CSV without outliers.
NUM_RESPONDENTS <- nrow(respondents)
lockBinding("NUM_RESPONDENTS", globalenv())

foodResponses <- as.matrix(respondents)[,QUESTIONS_START:QUESTIONS_END] # Convert responses to a matrix.
foodResponses <- apply(foodResponses, 1, as.numeric) # Make the matrix numeric.
foodResponses <- 5 - foodResponses

NUM_QUESTIONS <- nrow(foodResponses) # The number of sandwich questions.
lockBinding("NUM_QUESTIONS", globalenv())

totalResponse <- numeric(NUM_QUESTIONS)
averageResponse <- numeric(NUM_QUESTIONS)

# Sum up the total score for each question by respondent.
for (i in 1 : NUM_RESPONDENTS) {
        totalResponse <- totalResponse + (foodResponses[,i])
}
averageResponse <- totalResponse / NUM_RESPONDENTS # Divide by the number of respondents to find the mean.

orthodoxyScores <- numeric(NUM_RESPONDENTS)
purityScores <- numeric(NUM_RESPONDENTS)

for (i in 1 : NUM_RESPONDENTS) {
        orthodoxyScores[i] <- cosineSimilarity(foodResponses[,i], averageResponse)
        purityScores[i] <- sum(foodResponses[,i])
}
purityScores <- purityScores / (5 * NUM_QUESTIONS)

# Put data into frame.
data <- data.frame(purity = purityScores, orthodoxy = orthodoxyScores, grade = respondents[,GRADE_COLUMN], subjects = respondents[,SUBJECTS_COLUMN], background = respondents[,BACKGROUND_COLUMN], stream = respondents[,STREAM_COLUMN])
```

A.3 BoxPlots.R

```r
# This program creates box plots of the respondents.

library(ggplot2)
source("Calculate.R")

makeBoxPlot <- function(categoryData, categoryName, categoryTitle) {
        boxPlotData <- data[categoryData %in% names(table(categoryData))[table(categoryData) > 1],] # Remove all categorical data points only occurring once, as these data are not helpful for a box plot.

        # Create the purity plot.
        dataPlot <- ggplot(boxPlotData, aes_string(x = categoryName, y = "purity", fill=categoryName)) # Setup the plot.
        dataPlot <- dataPlot + coord_cartesian(ylim = c(-1, 1)) # Set the graph limits.
        dataPlot <- dataPlot + geom_boxplot() # Add the data points.
        dataPlot <- dataPlot + labs(x = categoryTitle, y = "Purity") # Give axes proper labels.
        dataPlot <- dataPlot + theme(legend.position = "none") # Remove the legend.
        ggsave(paste(categoryTitle, "Purity.pdf"), plot=dataPlot, width=9, height=8)

        # Create the orthodoxy plot.
        dataPlot <- ggplot(boxPlotData, aes_string(x = categoryName, y = "orthodoxy", fill=categoryName)) # Setup the plot.
        dataPlot <- dataPlot + coord_cartesian(ylim = c(-1, 1)) # Set the graph limits.
        dataPlot <- dataPlot + geom_boxplot() # Add the data points.
        dataPlot <- dataPlot + labs(x = categoryTitle, y = "Orthodoxy") # Give axes proper labels.
        dataPlot <- dataPlot + theme(legend.position = "none") # Remove the legend.
        ggsave(paste(categoryTitle, "Orthodoxy.pdf"), plot=dataPlot, width=9, height=8)
}

```

```r
makeBoxPlot(data$grade, "grade", "Grade")
makeBoxPlot(data$background, "background", "Ethnic Background")
makeBoxPlot(data$stream, "stream", "Stream")

boxPlotData <- data

# Create the purity plot.
dataPlot <- ggplot(boxPlotData, aes(x = 1, y = purity)) # Setup the plot.
dataPlot <- dataPlot + coord_cartesian(ylim = c(-1, 1)) # Set the graph limits.
dataPlot <- dataPlot + geom_boxplot() # Add the data points.
dataPlot <- dataPlot + labs(x = "All respondents", y = "Purity") # Give axes proper labels.
dataPlot <- dataPlot + theme(axis.text.x = element_blank(), axis.ticks.x = element_blank()) # Remove the x axis.
ggsave("PurityBoxPlot.pdf", plot=dataPlot, width=2, height=8)

# Create the orthodoxy plot.
dataPlot <- ggplot(boxPlotData, aes(x = 1, y = orthodoxy)) # Setup the plot.
dataPlot <- dataPlot + coord_cartesian(ylim = c(-1, 1)) # Set the graph limits.
dataPlot <- dataPlot + geom_boxplot() # Add the data points.
dataPlot <- dataPlot + labs(x = "All respondents", y = "Orthodoxy") # Give axes proper labels.
dataPlot <- dataPlot + theme(axis.text.x = element_blank(), axis.ticks.x = element_blank()) # Remove the x axis.
ggsave("OrthodoxyBoxPlot.pdf", plot=dataPlot, width=2, height=8)
```

A.4 PurityOrthodoxyPlot.R

```r
# This is a program to plot orthodoxy vs purity on the sandwich spectrum for all respondents.

library(ggplot2)
source("Calculate.R")

makeScatterPlot <- function(categoryName, categoryTitle) {
        # Create the plot.
```

```
        dataPlot <- ggplot(data, aes(purity, orthodoxy)) # Setup the plot.
        dataPlot <- dataPlot + xlim(-1, 1) + ylim(-1, 1) # Set the graph limits.
        dataPlot <- dataPlot + geom_point(aes_string(colour = categoryName), size = 0.75) # Add the data points.
        dataPlot <- dataPlot + geom_smooth(method = lm, fullrange = TRUE, se = FALSE, colour = "black", size = 0.5, linetype="dashed") # Add the line of best fit.
        dataPlot <- dataPlot + geom_smooth(method = lm, formula = y ~ poly(x, 2), fullrange = TRUE, se = FALSE, colour = "blue", size = 0.5, linetype="dotted") # Add the curve of best fit.
        dataPlot <- dataPlot + geom_smooth(method = loess, level = 0.99, colour = "red", size = 0.5) # Add the confidence curve.
        dataPlot <- dataPlot + labs(x = "Purity", y = "Orthodoxy", colour = categoryTitle) # Give axes and legend proper labels.
        #dataPlot <- dataPlot + geom_text(x = -0.25, y = 0.75, label = lm_eqn(data), parse = TRUE)
        ggsave(paste(categoryTitle, "PurityVsOrthodoxy.pdf", sep = ""), plot=dataPlot, width=5.5, height=3)
}

makeScatterPlot("stream", "Stream")
makeScatterPlot("grade", "Grade")
makeScatterPlot("background", "Ethnic Background")
```

A.5 Residuals.R

```
# This program creates a residual plot of the respondents using the quadratic method.

library(ggplot2)
library(broom)

source("Calculate.R")

residualPlotData <- data

# Create the quadratic model.
```

```
mod <- lm(orthodoxy ~ poly(purity, 2), data =
    residualPlotData)
df <- augment(mod)

# Create the residual plot.
dataPlot <- ggplot(df, aes(.fitted, .resid)) + geom_point()
dataPlot <- dataPlot + coord_cartesian(ylim = c(-1, 1)) # Set
     the graph limits.
dataPlot <- dataPlot + labs(x = "Fitted Value", y = "Residual
    ") # Give axes proper labels.
ggsave("QuadraticResidualPlot.pdf", plot=dataPlot, width=5.5,
     height=3)
```

A.6 Levels.R

```
# Output the levels for certain columns of interest in the
    input.

# Define the constants for what columns data is in inside the
     CSV file.
GRADE_COLUMN <- 1 # What column the grade data is in.
SUBJECTS_COLUMN <- 2 # What column the favourite subjects
    data is in.
BACKGROUND_COLUMN <- 3 # What column the ethnic background
    data is in.
STREAM_COLUMN <- 4 # What column the stream data is in.
QUESTIONS_START <- 6 # What column the food questions start
    at.
QUESTIONS_END <- 48 # What column the food questions end at.
lockBinding("GRADE_COLUMN", globalenv())
lockBinding("SUBJECTS_COLUMN", globalenv())
lockBinding("BACKGROUND_COLUMN", globalenv())
lockBinding("STREAM_COLUMN", globalenv())
lockBinding("QUESTIONS_START", globalenv())
lockBinding("QUESTIONS_END", globalenv())

respondents <- read.csv("CleanedData.csv") # Read in the CSV.

print(paste(nrow(respondents), "respondents"))

summary(respondents[,GRADE_COLUMN])
```

```
print(paste(length(levels(respondents[,GRADE_COLUMN])), " grades"))

#summary(respondents[,SUBJECTS_COLUMN])
#print(paste(length(levels(respondents[,SUBJECTS_COLUMN])), " subjects."))

summary(respondents[,BACKGROUND_COLUMN])
print(paste(length(levels(respondents[,BACKGROUND_COLUMN])), "ethnic backgrounds"))

summary(respondents[,STREAM_COLUMN])
print(paste(length(levels(respondents[,STREAM_COLUMN])), " streams"))
```

A.7 Tables.R

```
# This program creates a CSV file with the data on each respondent's metrics.

source("Calculate.R")

data$purity <- round(data$purity, digits=5)
data$orthodoxy <- round(data$orthodoxy, digits=5)
data <- data[,-4]
write.csv(data, file = "Results.csv")
write.csv(summary(data), file = "Summary.csv")
```

A.8 Outliers.R

```
# This program removes outliers.

source("Calculate.R")

outlierData <- data

# Create the quadratic model.
mod <- lm(orthodoxy ~ poly(purity, 2), data = outlierData)
resid <- unname(mod$residuals)
```

```
# Compute the cutoffs.
q1 <- unname(quantile(resid)[2])
q3 <- unname(quantile(resid)[4])
icq = q3 - q1
lowCutoff = q1 - 1.5 * icq
highCutoff = q3 + 1.5 * icq

remove <- -which(resid < lowCutoff | resid > highCutoff)
removed <- respondents[remove,]
print(remove)

write.csv(removed, file = "NoOutliers.csv", row.names = FALSE
    )
```