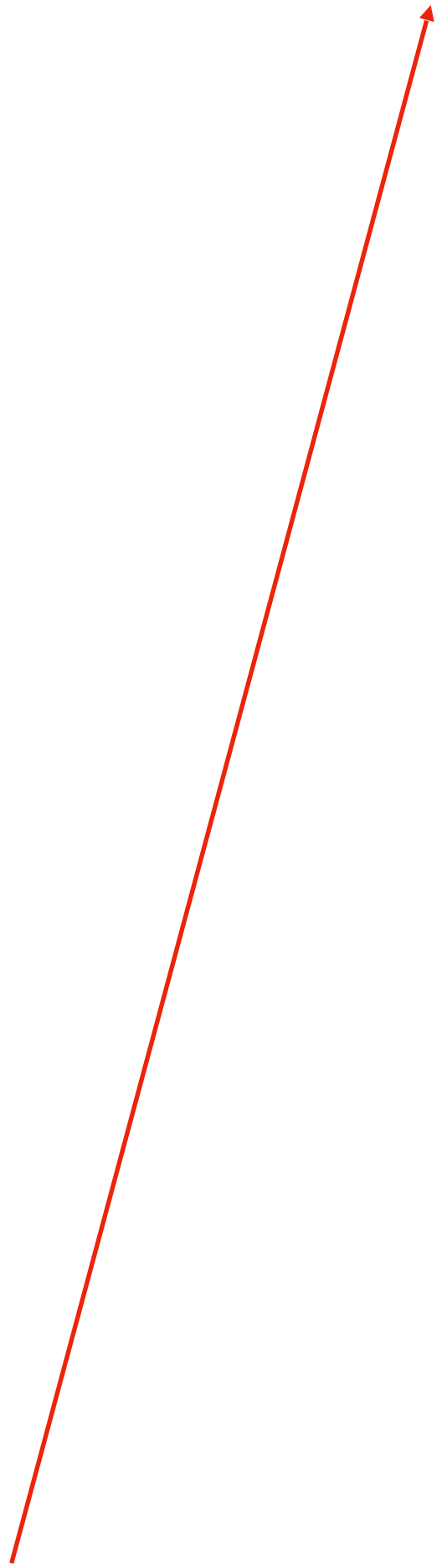
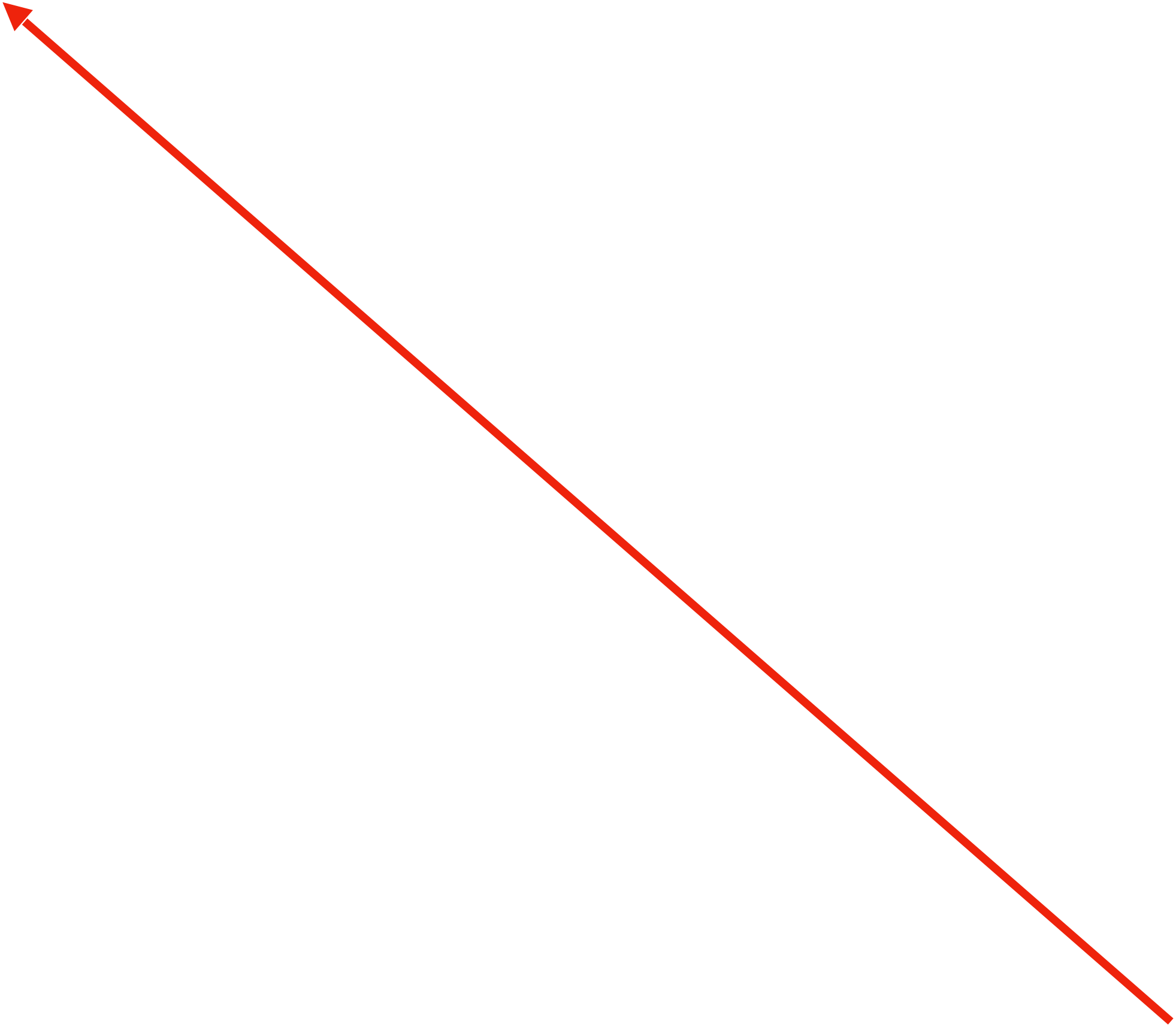
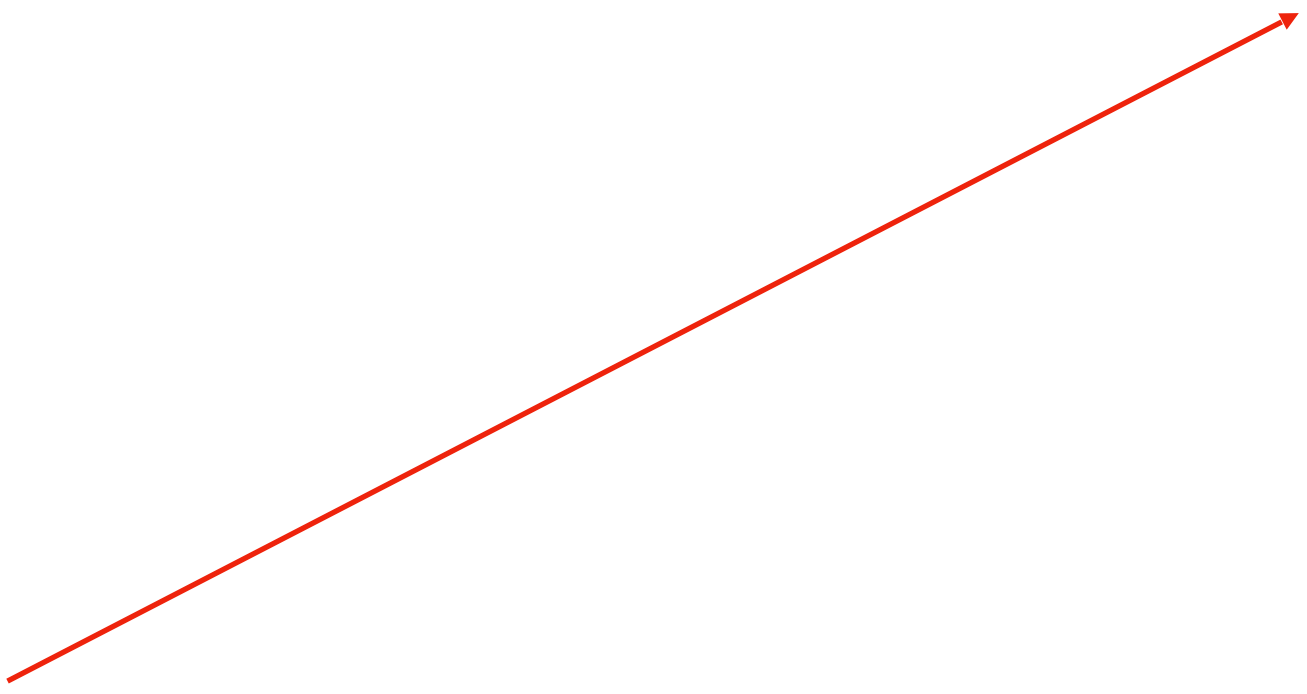
# Defining Derivations

1. Identifying the consumption patterns that **differentiate** regional and geodemographics groups

- We want to find the **Surprising** rules (contrast-set) that are **Significant** and **Large** across groups, while neglecting rules that are redundant and universal.





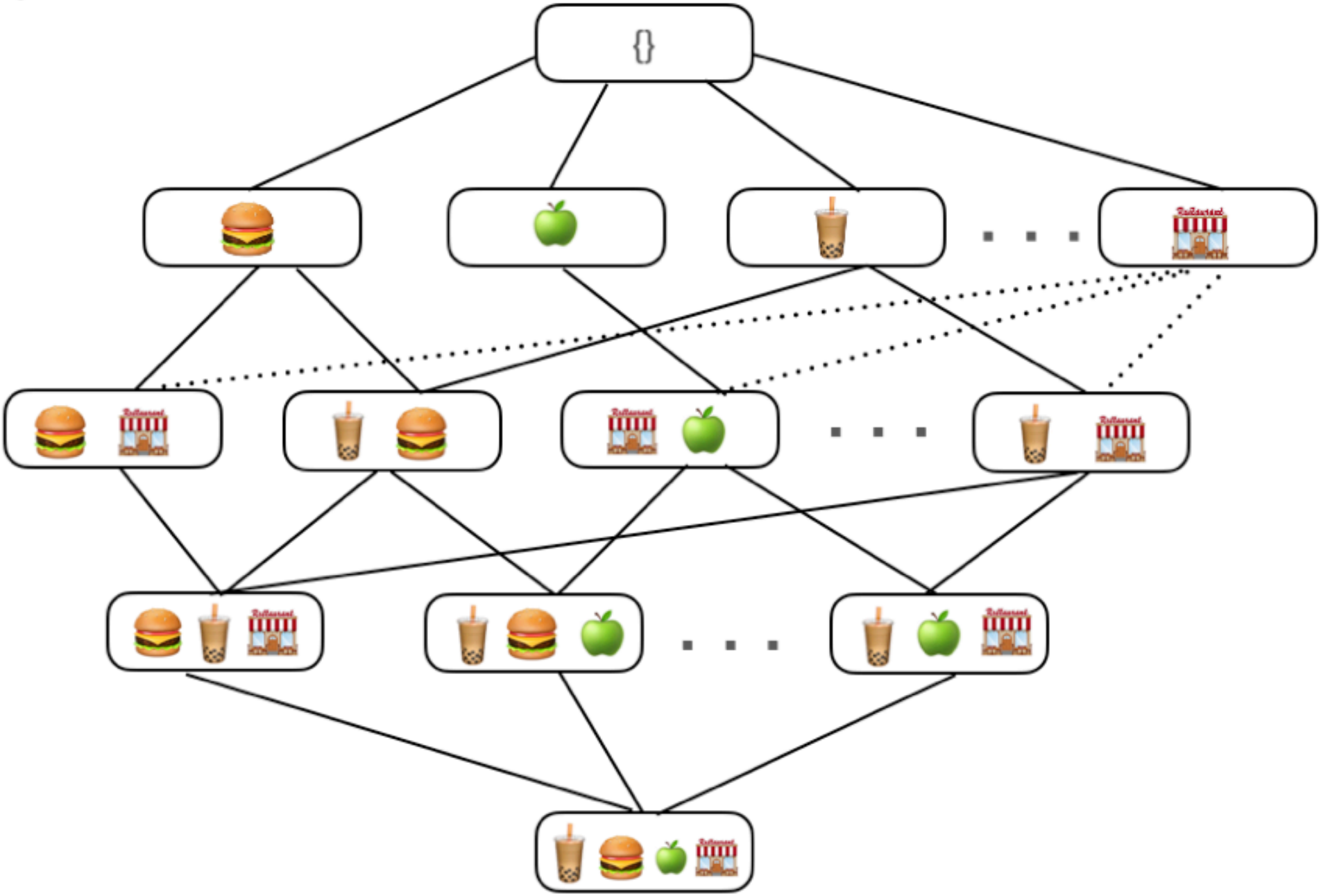


# How can we determine this?

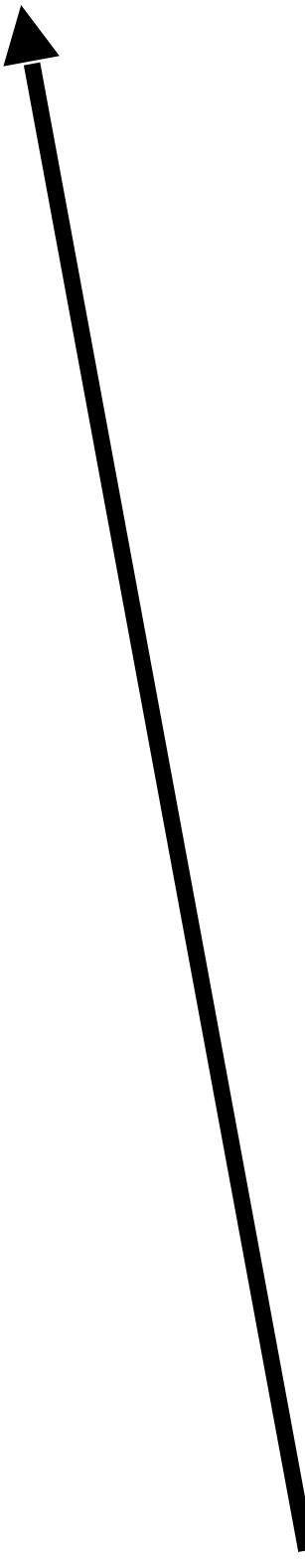
Consumer
Data
Research
Centre

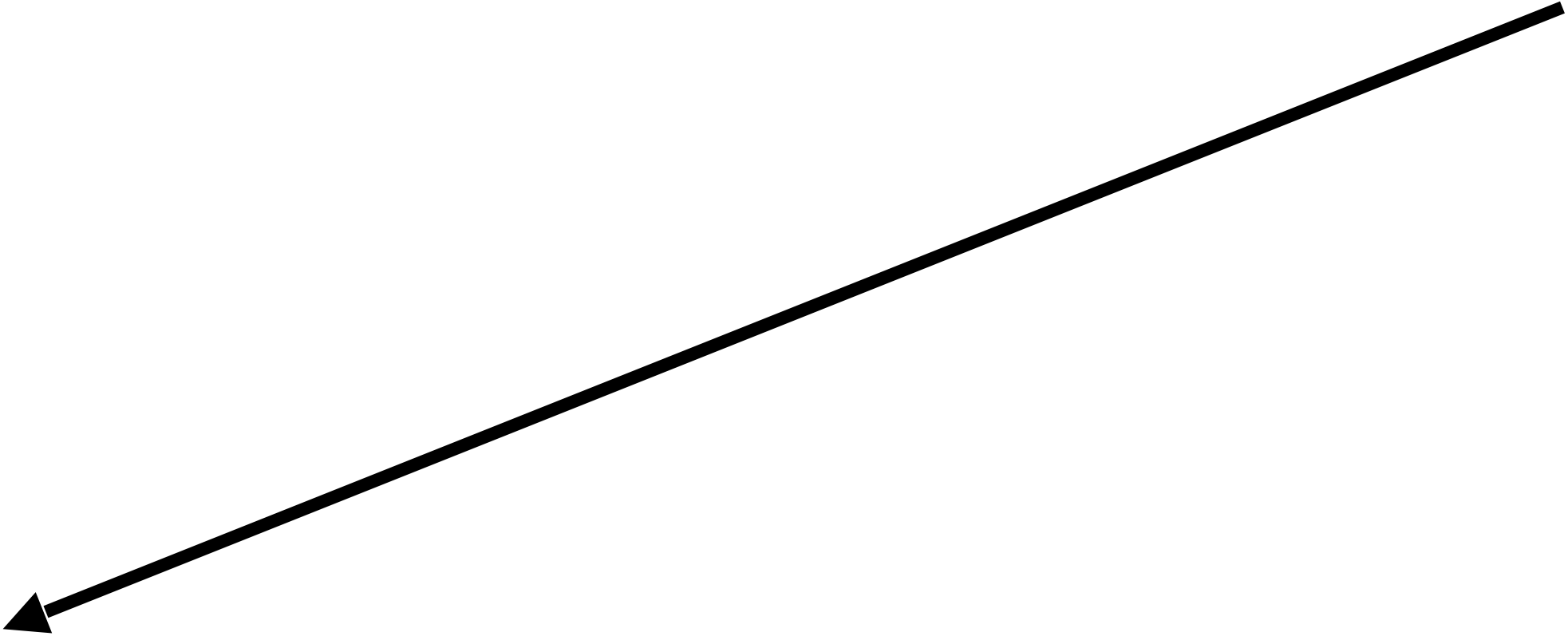
An ESRC Data Investment

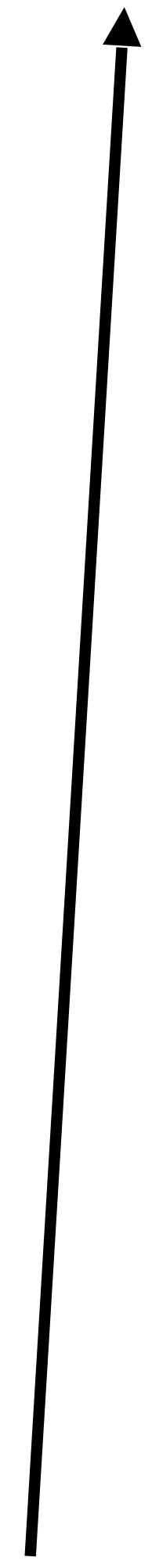
KANTAR WORLDPANEL



| Groups | Frequency |
| --- | --- |
| Inner City | 1300 |
| Countryside | 400 |
| .... | |
| Student Group | 3000 |







| Groups | Frequency |
| --- | --- |
| Inner City | 300 |
| Countryside | 4000 |
| .... | |
| Student Group | 200 |

| Groups | Frequency |
| --- | --- |
| Inner City | 1500 |
| Countryside | 660 |
| .... | |
| Student Group | 112 |

| Groups | Frequency |
| --- | --- |
| Inner City | 540 |
| Countryside | 243 |
| .... | |
| Student Group | 685 |

- Contrast set mining (Bay and Pazzani, 2001;Hilderman and Peckham, 2005; Simeon and Hilderman, 2011 )
  - Modelled as tree-search

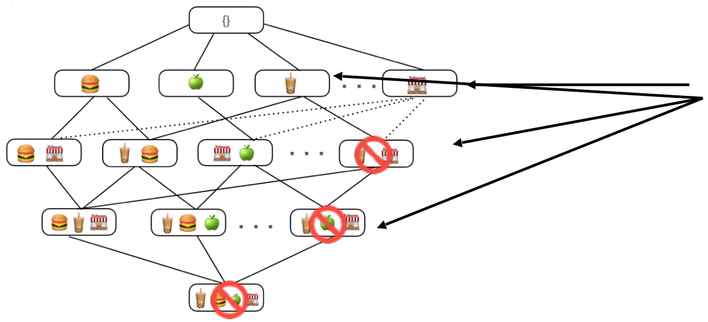
# Defining Derivations

1. Identifying the consumption patterns that **differentiate** regional and geodemographics groups

- We want to find the **Surprising** rules (contrast-set) that are **Significant** and **Large** across groups, while neglecting rules that are redundant and universal.

- Contrast set mining (Bay and Pazzani, 2001;Hilderman and Peckham, 2005; Simeon and Hilderman, 2011 )

  - Modelled as tree-search



| Groups | Frequency |
|---|---|
| Inner City | 540 |
| Countryside | 243 |
| .... | |
| Student Group | 685 |

Consumer Data Research Centre

An ESRC Data Investment

KANTAR WORLDPANEL

# Finding Derivations: Large

- A contrast-set for which the *maximum difference* between supports is greater than a *minimum support difference threshold*, is called ***Large*.**

**Example**

For the contrast set c1: "$Product = burger \wedge Store = Fastfood$" and mindev = 5%

support(c1|Inner City Cosmopolitan) = 13%

support(c1|Industrious communities) = 9%

Support(c1|Countryside Living) = 7%

**Deciding if a contrast set is large is straightforward:**

Max difference = 13% - 7% = 6%

With mindev = 5%, c1 is **Large**

To decide if a contrast set is **significant**, we use and statistical test

Consumer Data Research Centre

An ESRC Data Investment

KANTAR WORLDPANEL