Лабораторна робота №6

з навчальної дисципліни

«Математичні методи обробки інформації»

на тему:

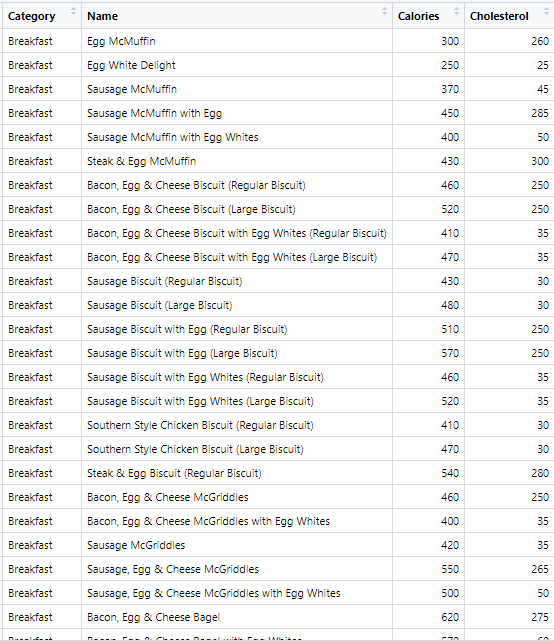
«КЛАСТЕРНИЙ АНАЛІЗ»

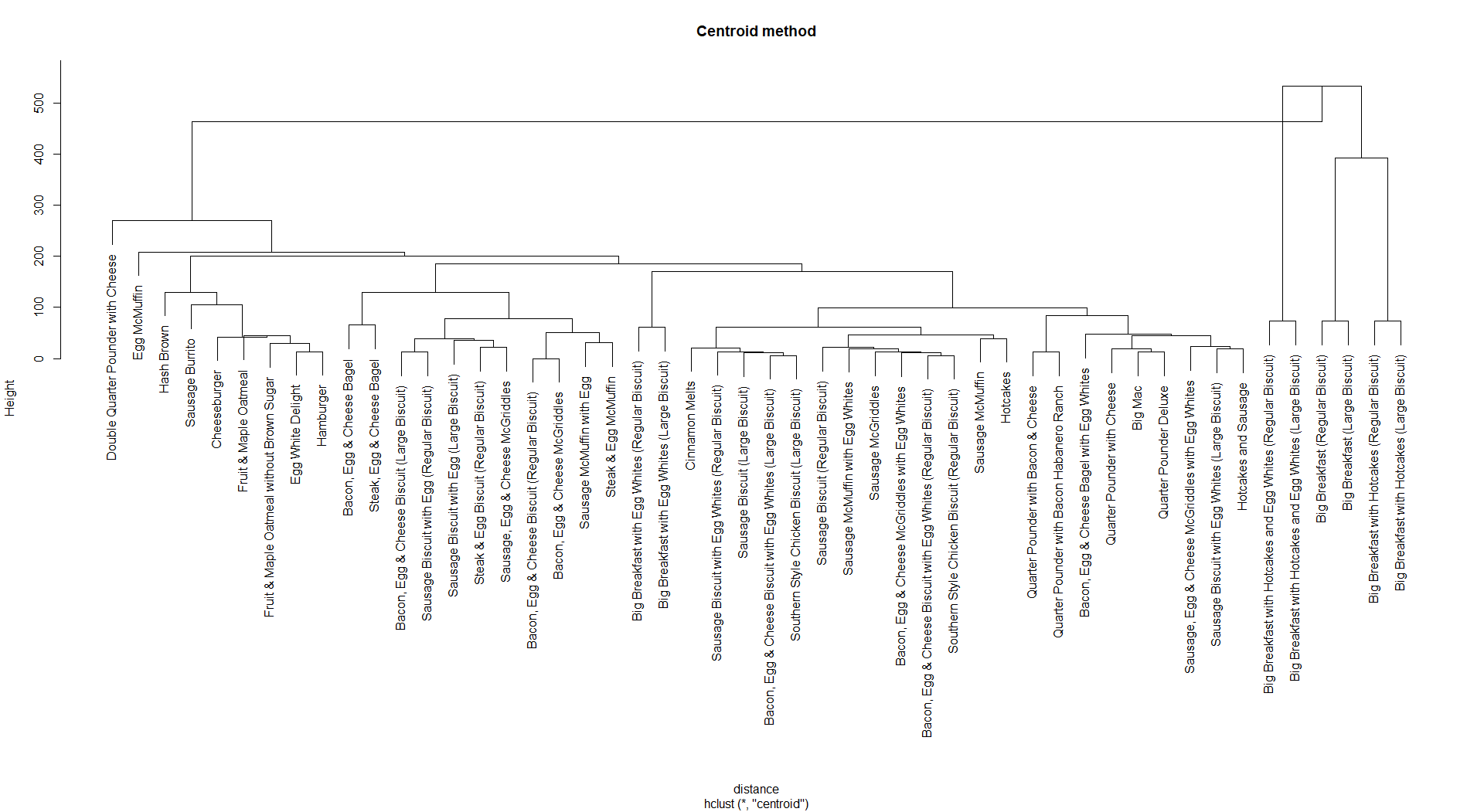
студента 3 курсу групи Т-31

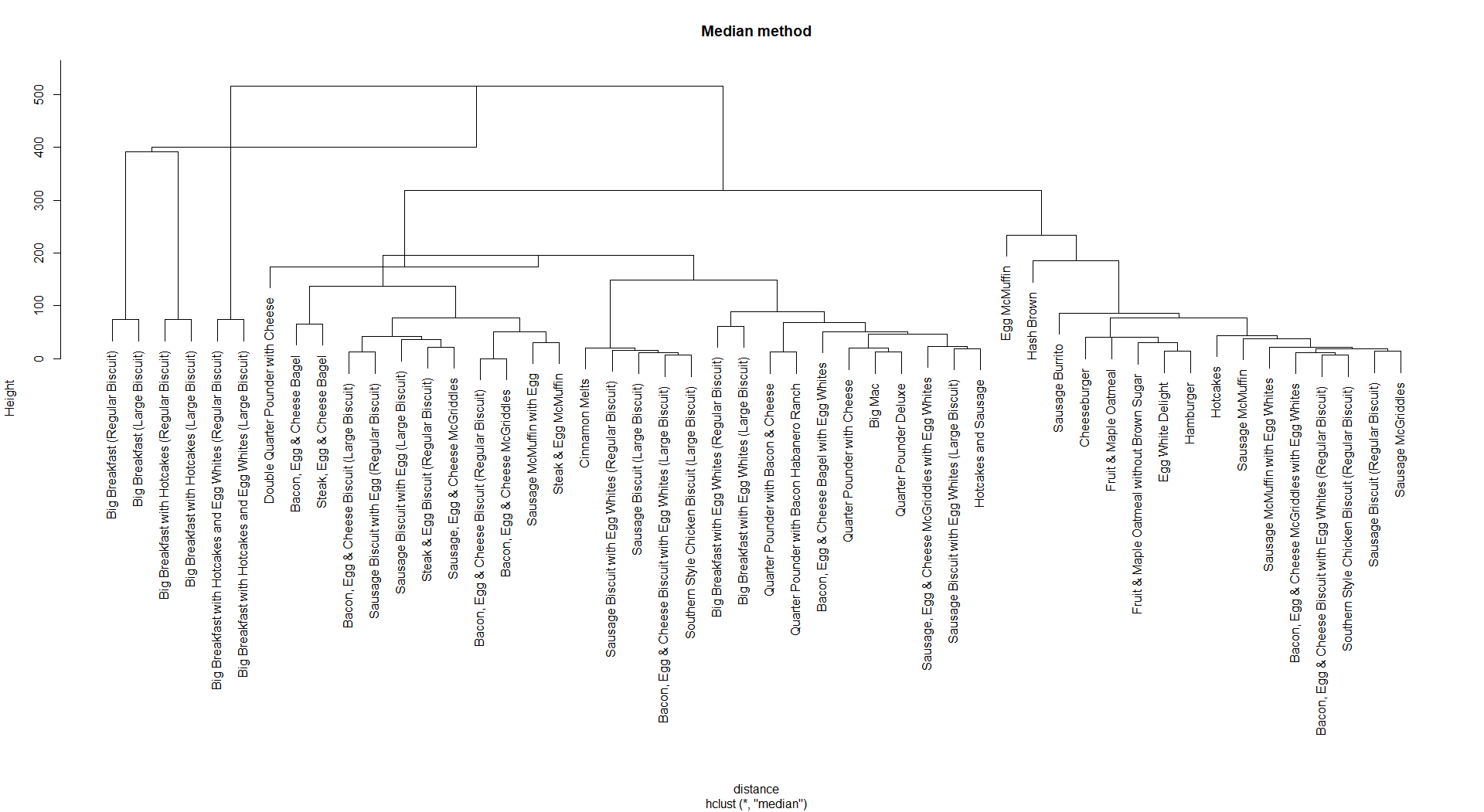
Бондаря Дениса

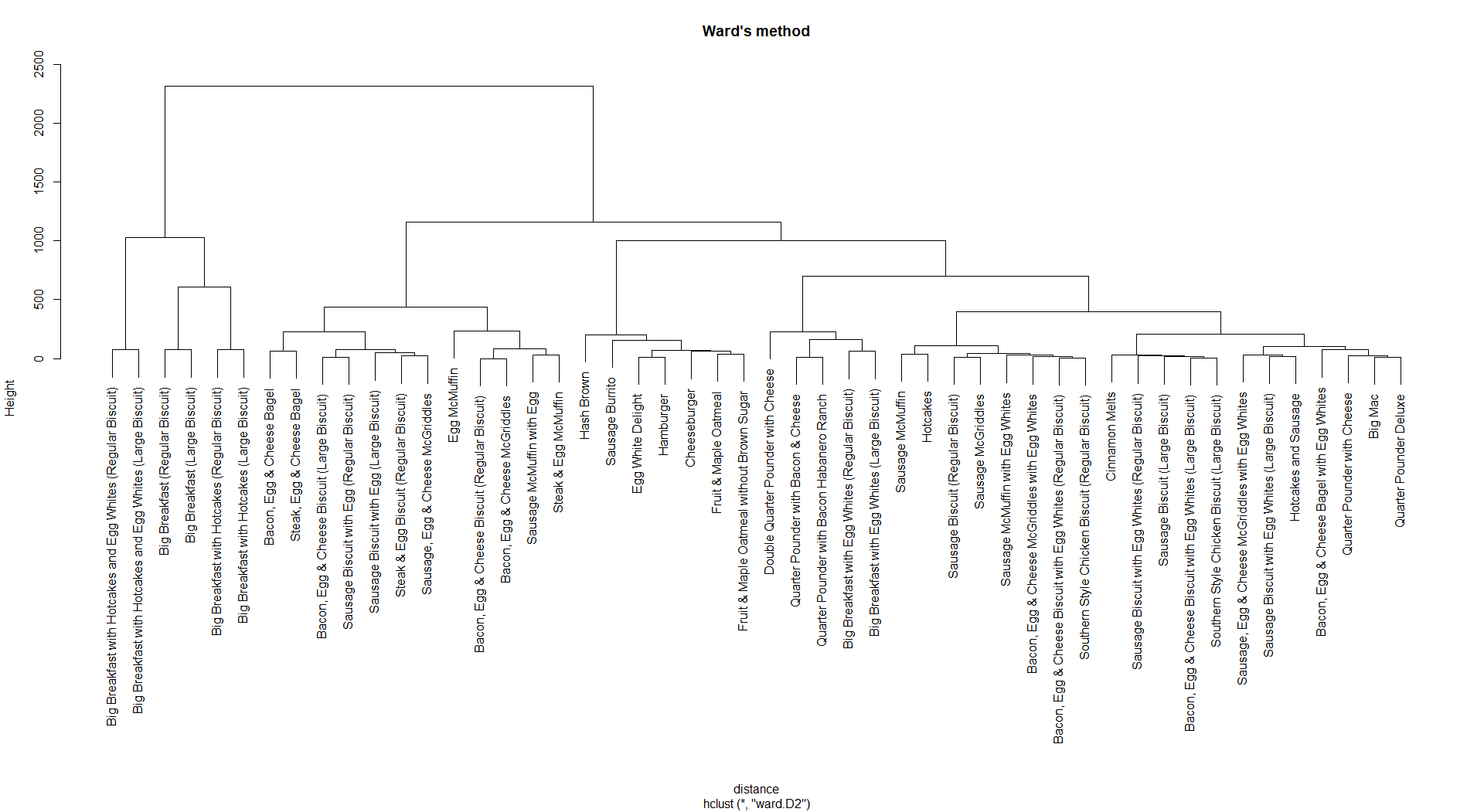
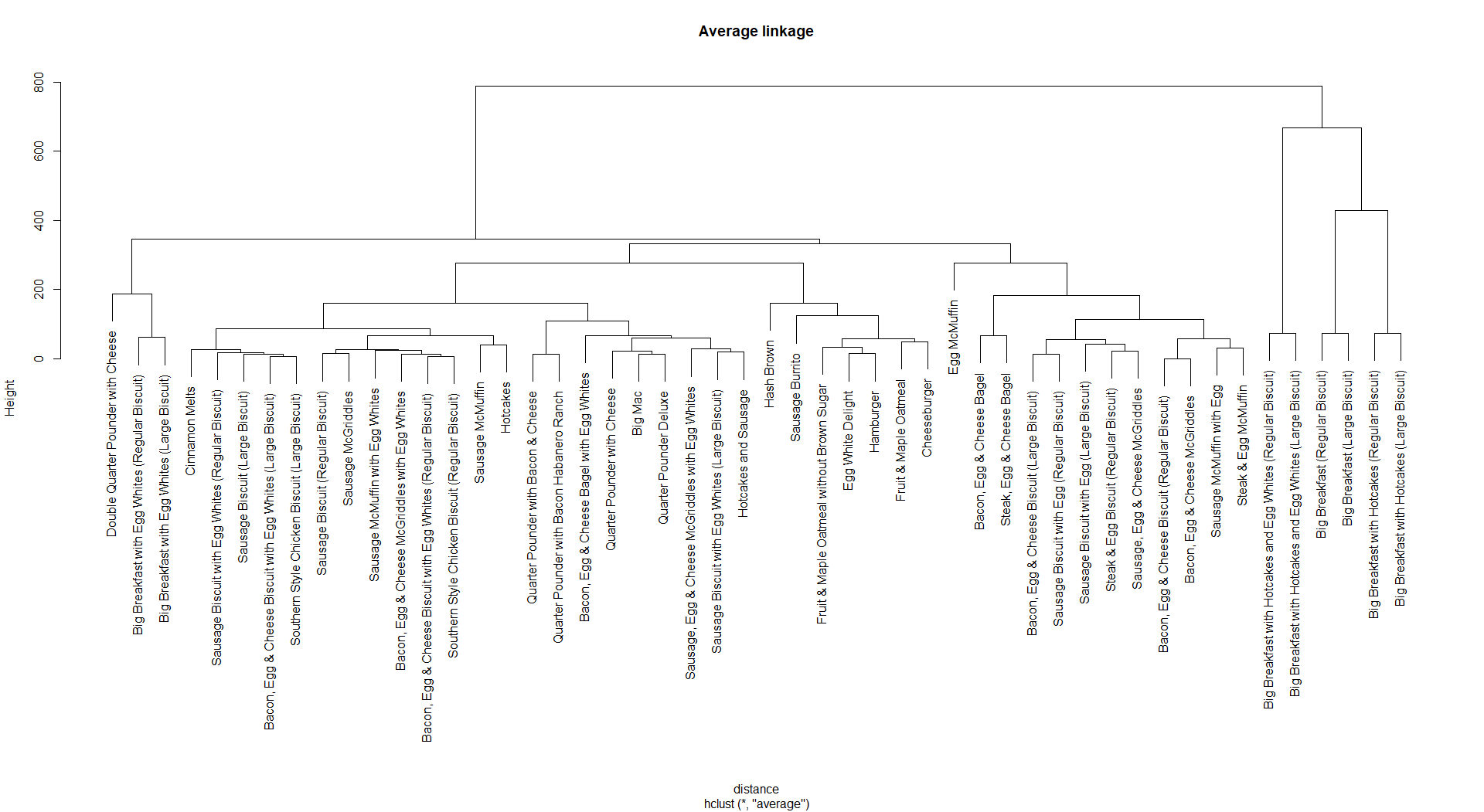
**Вибірка**

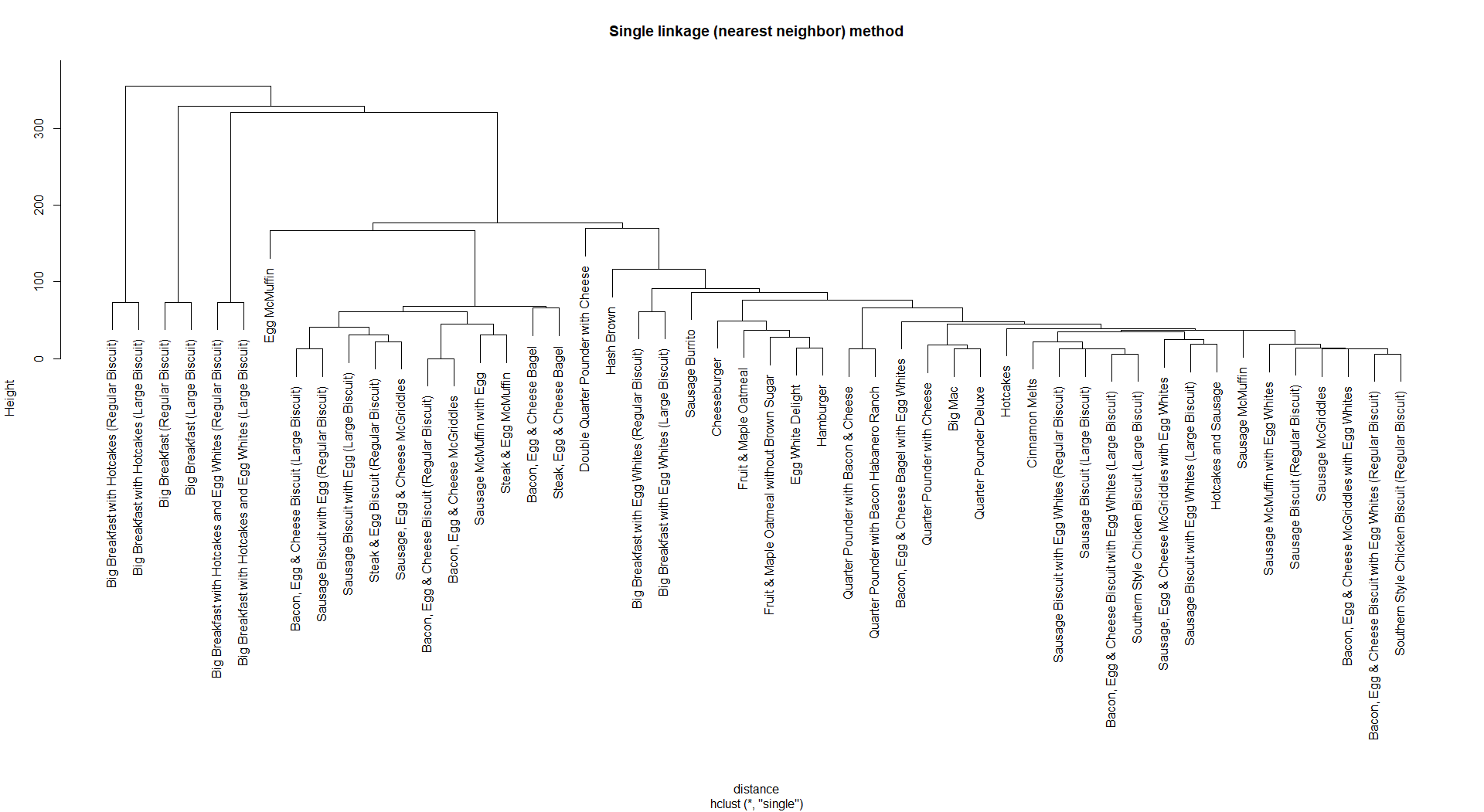
Датасет являє собою меню ресторану. Спостереженими величинами блюда будуть калорії та холестерин

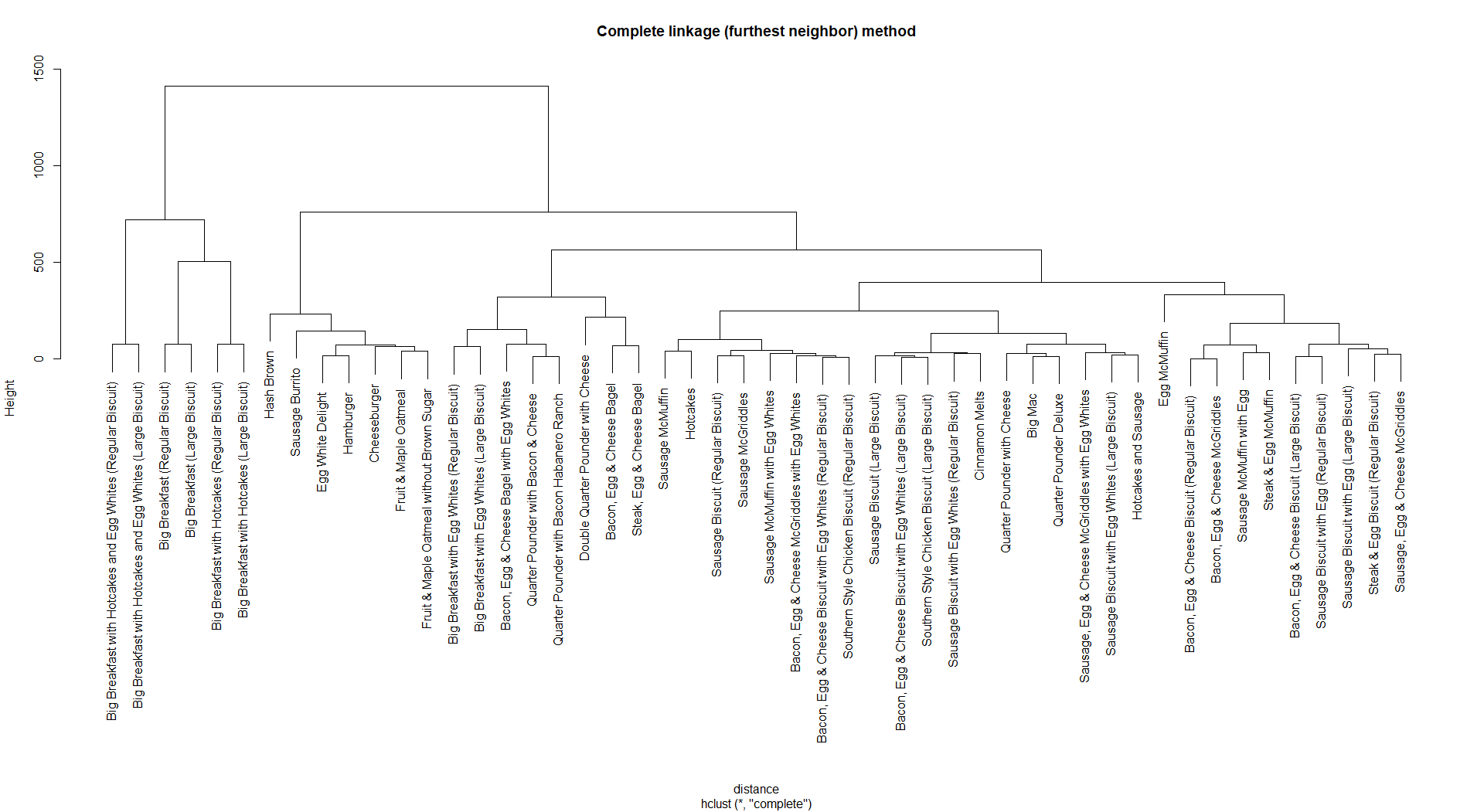


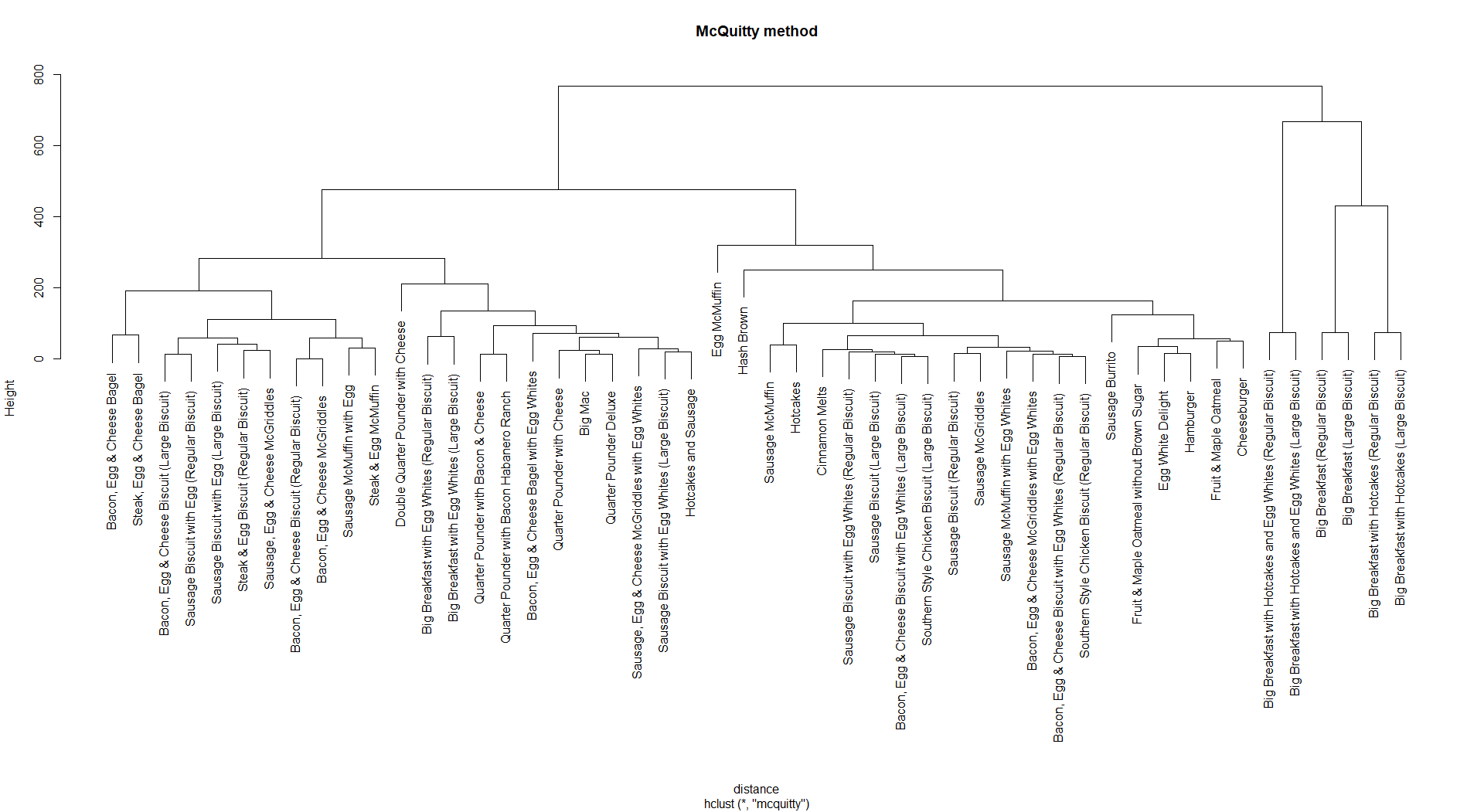


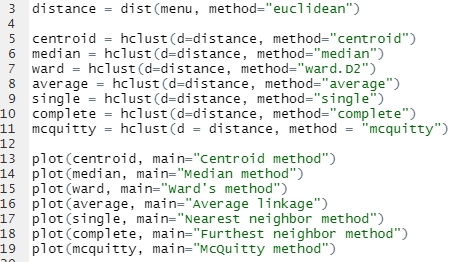


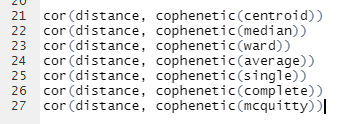


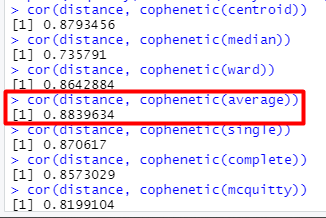








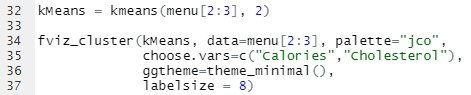


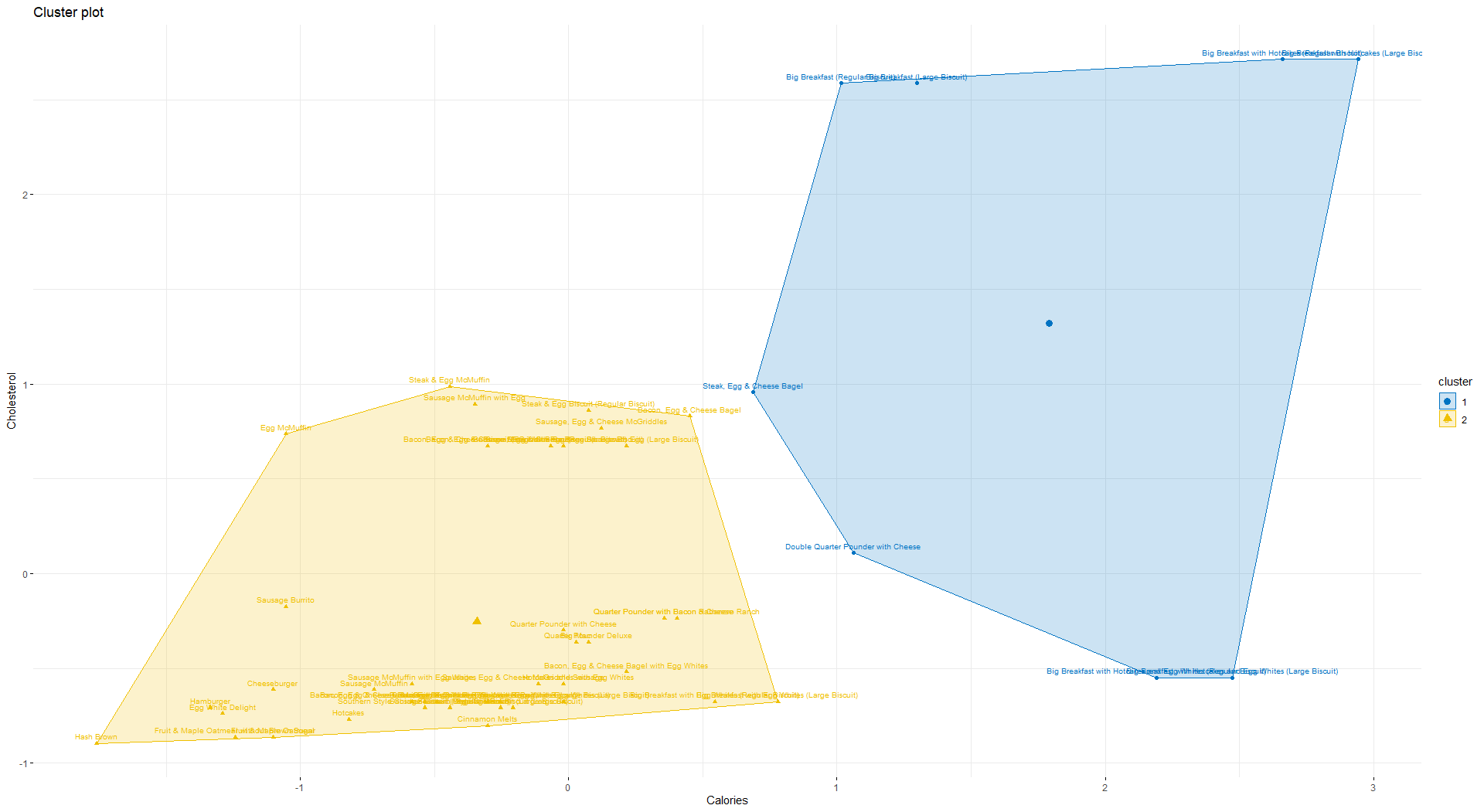


Найкраще значення кореляції – Average Linkage.

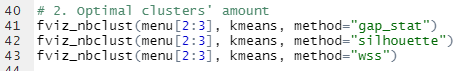
Друге найкраще – Centroid

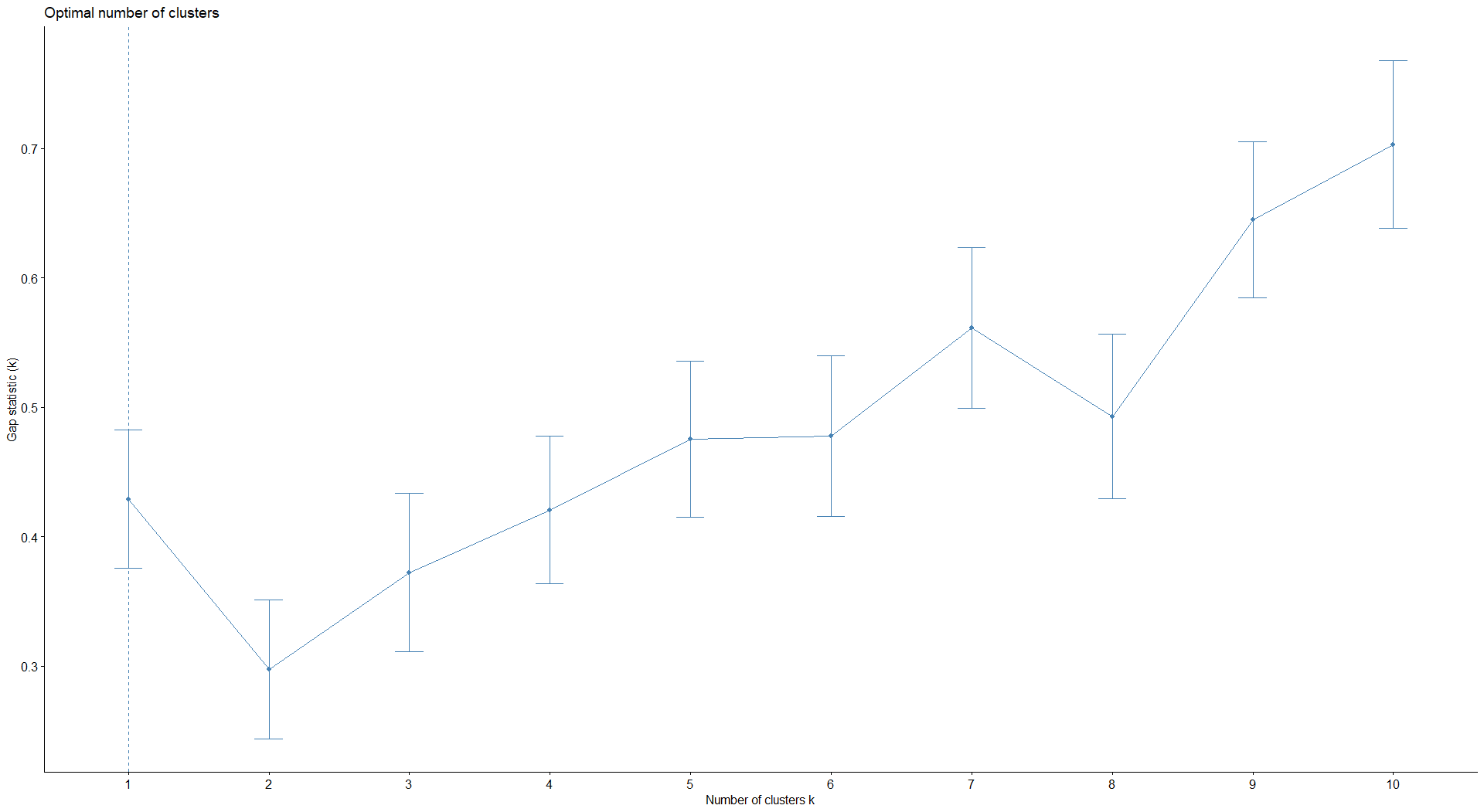
2. Метод k середніх

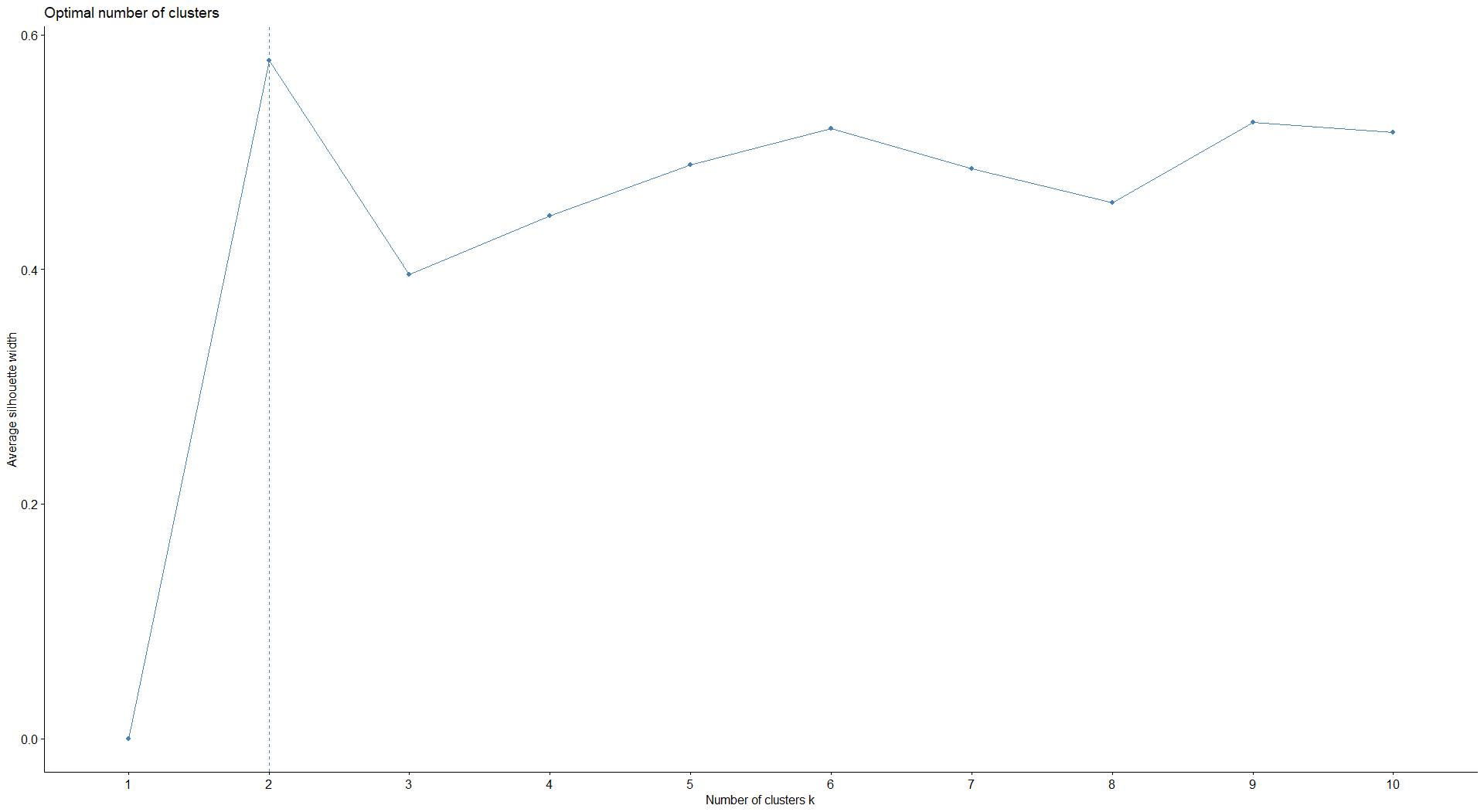


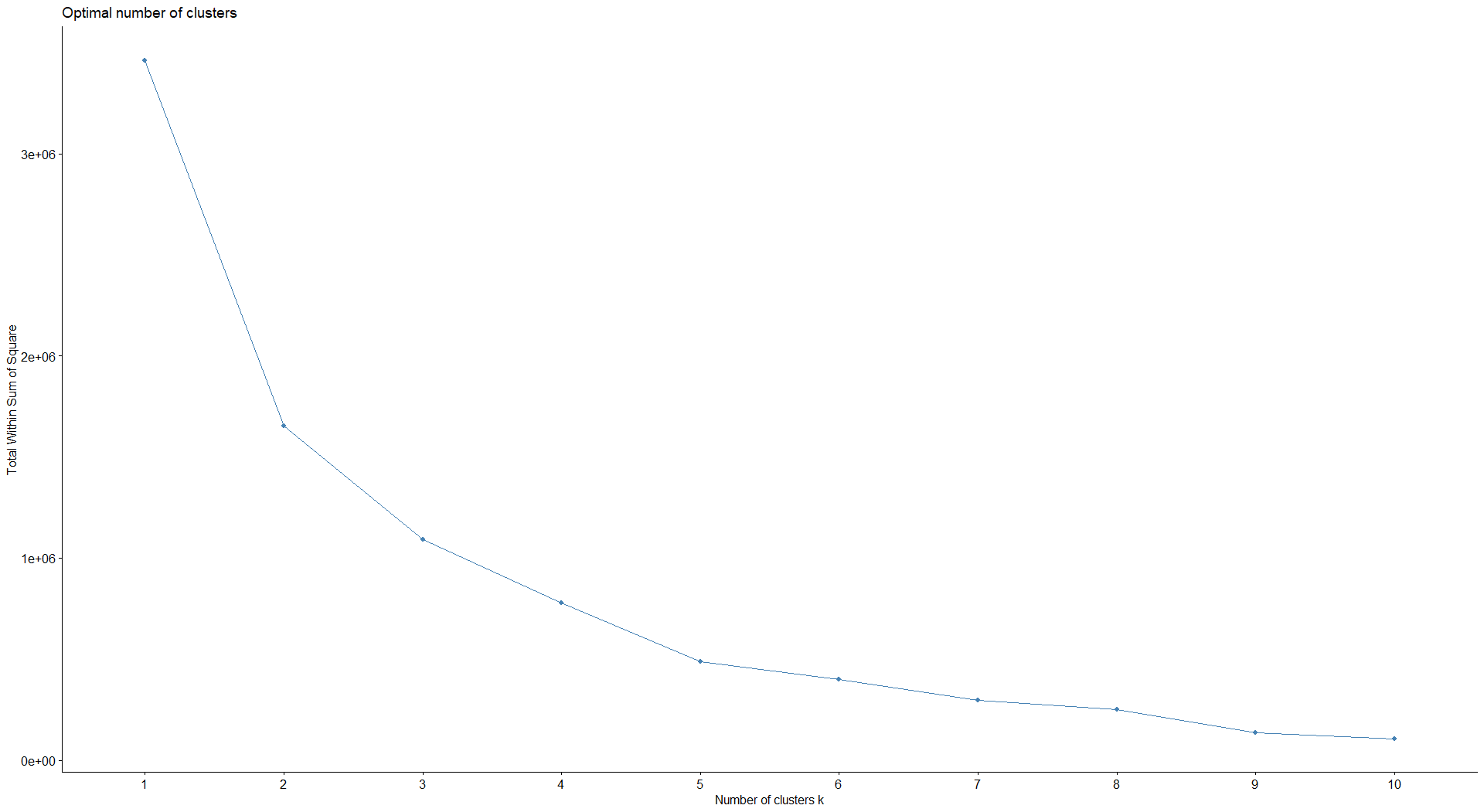


3. Оптимальна кількість кластерів

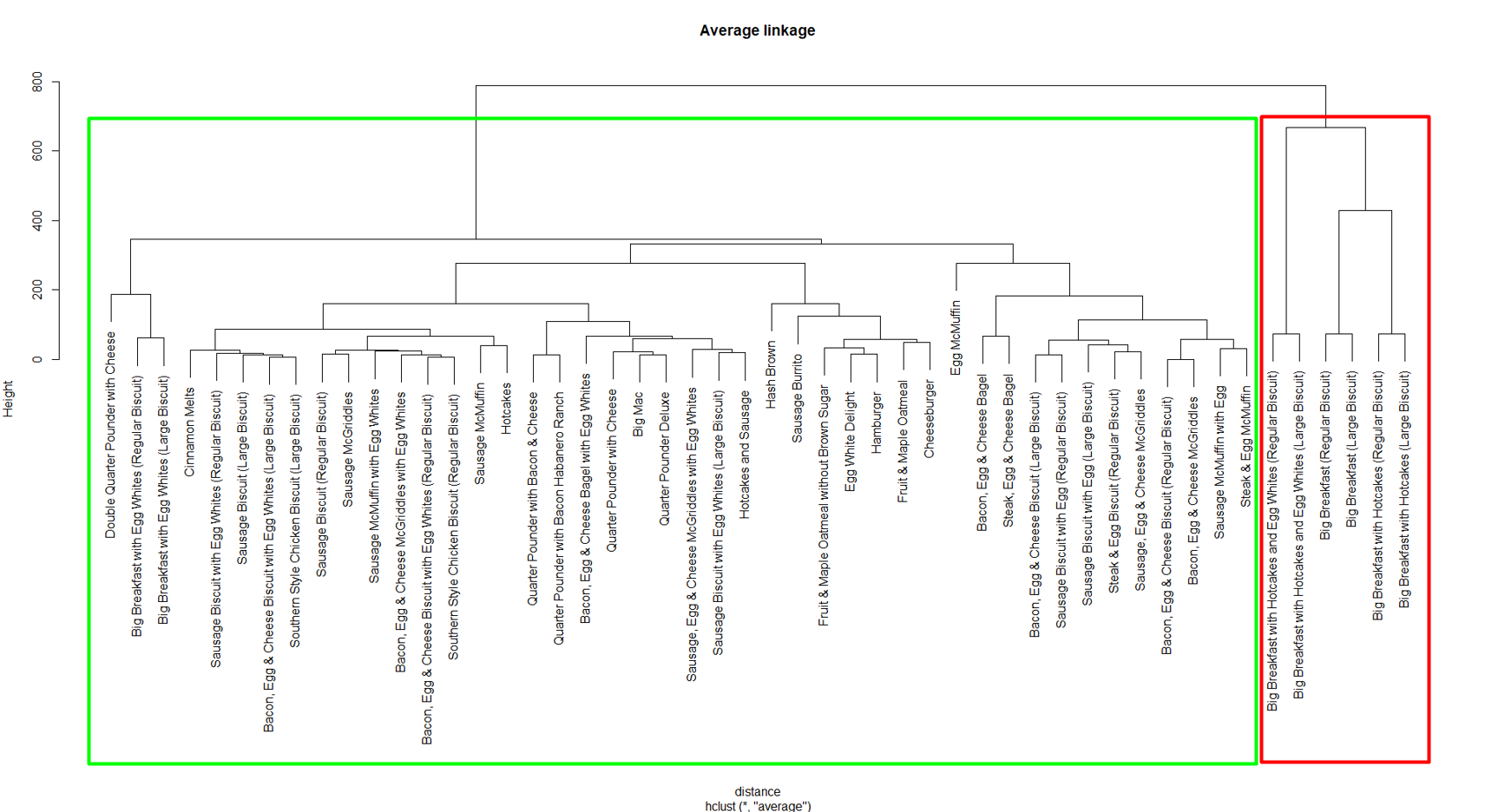


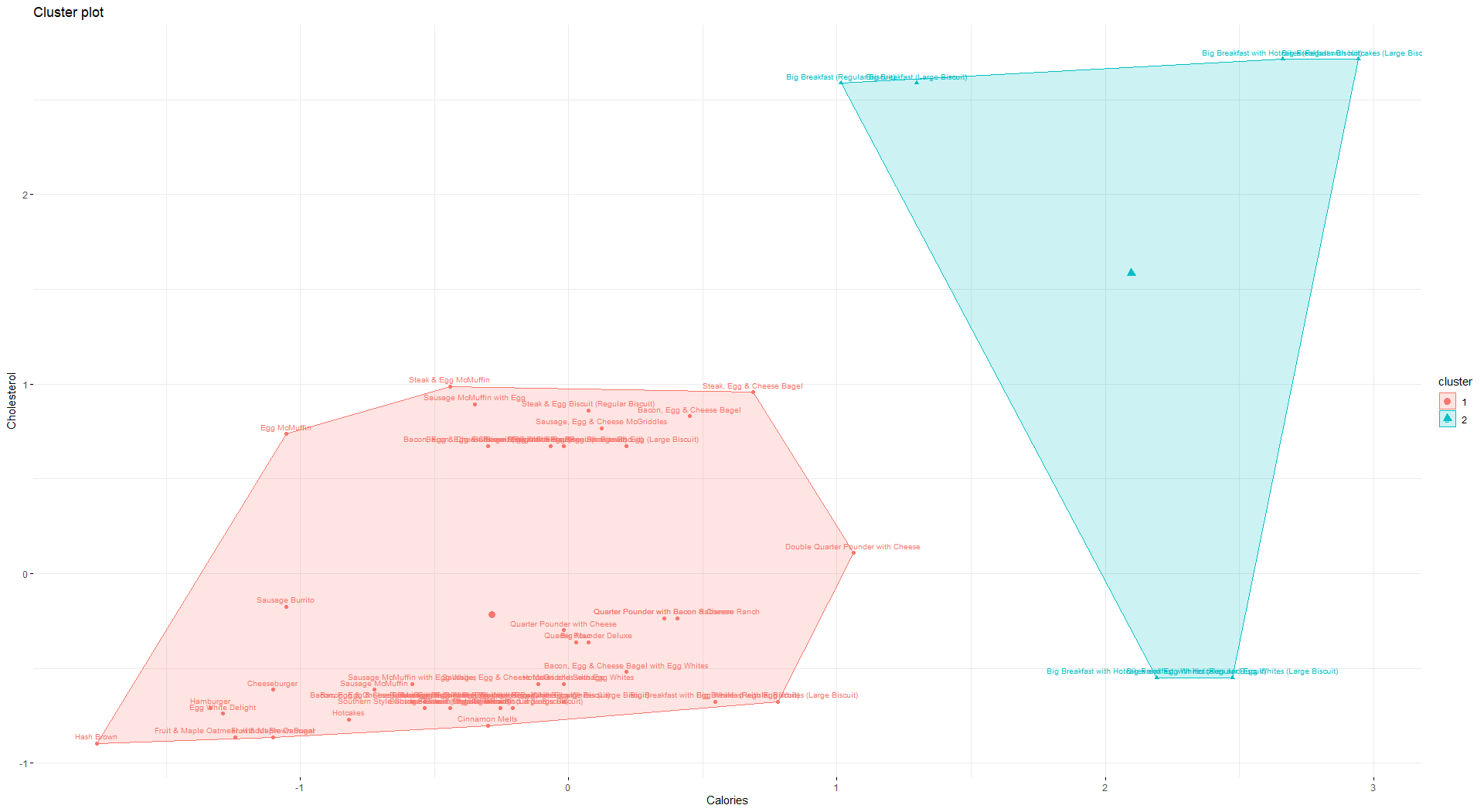


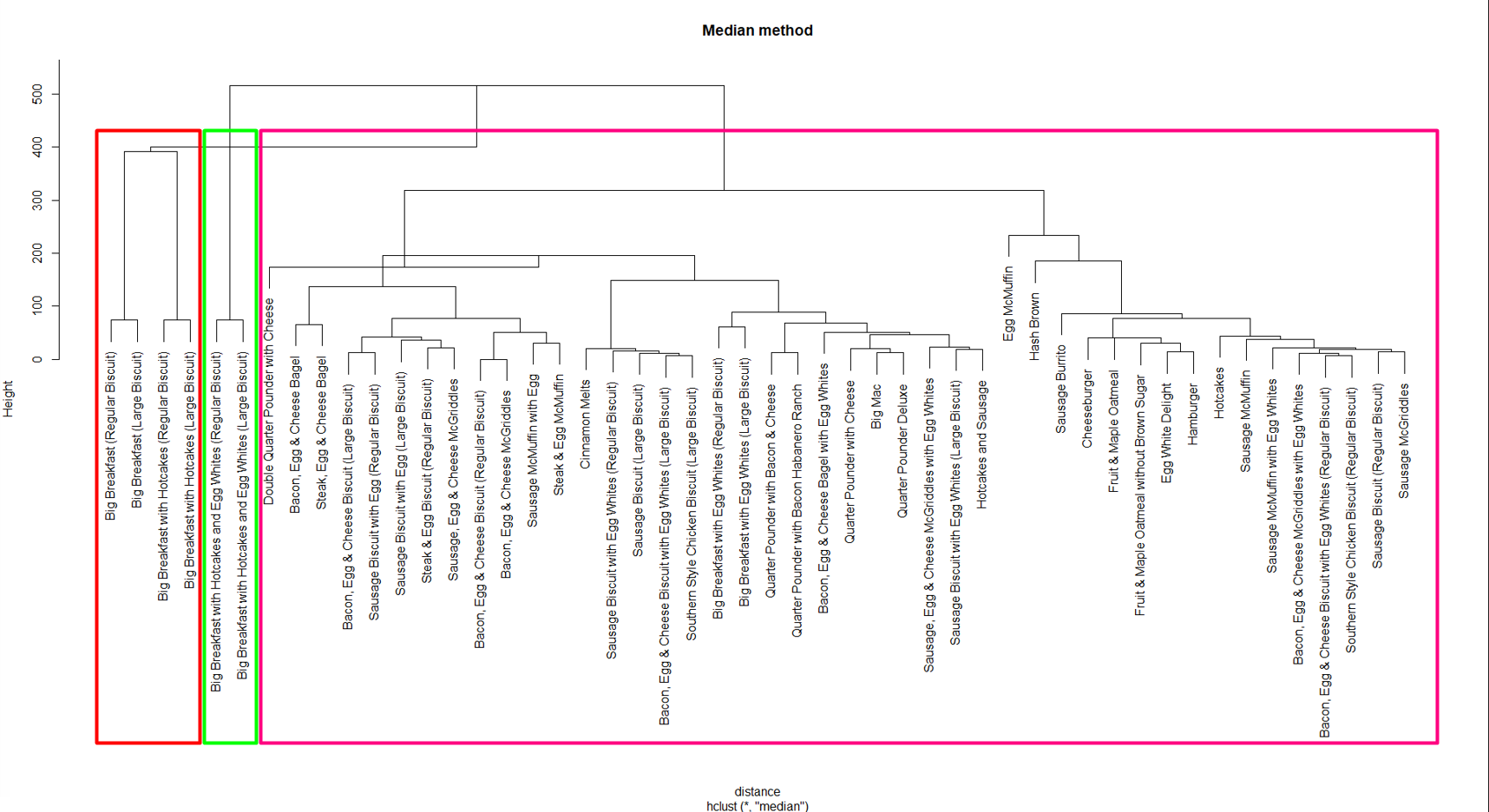


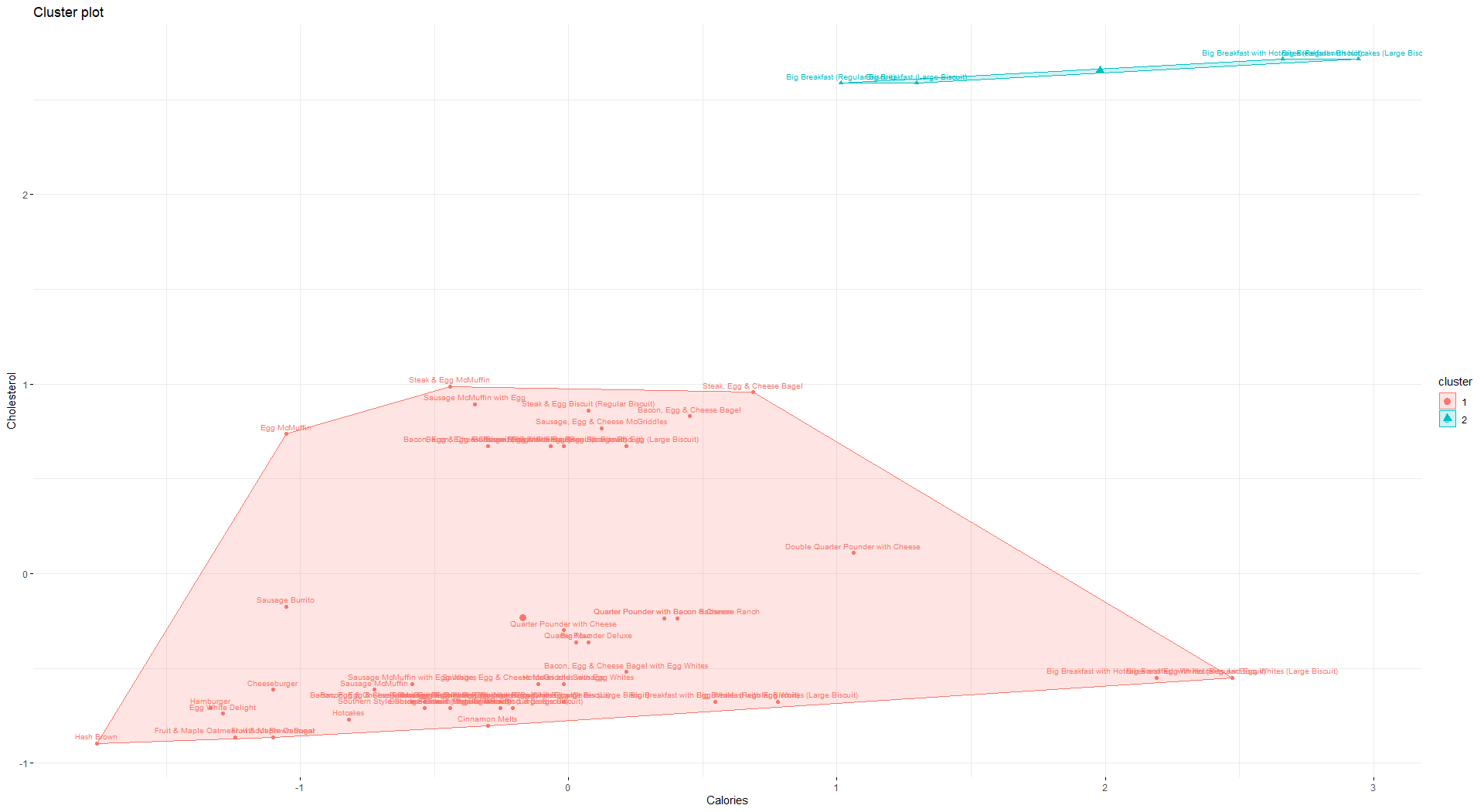


3. Найкращі результати кластеризації



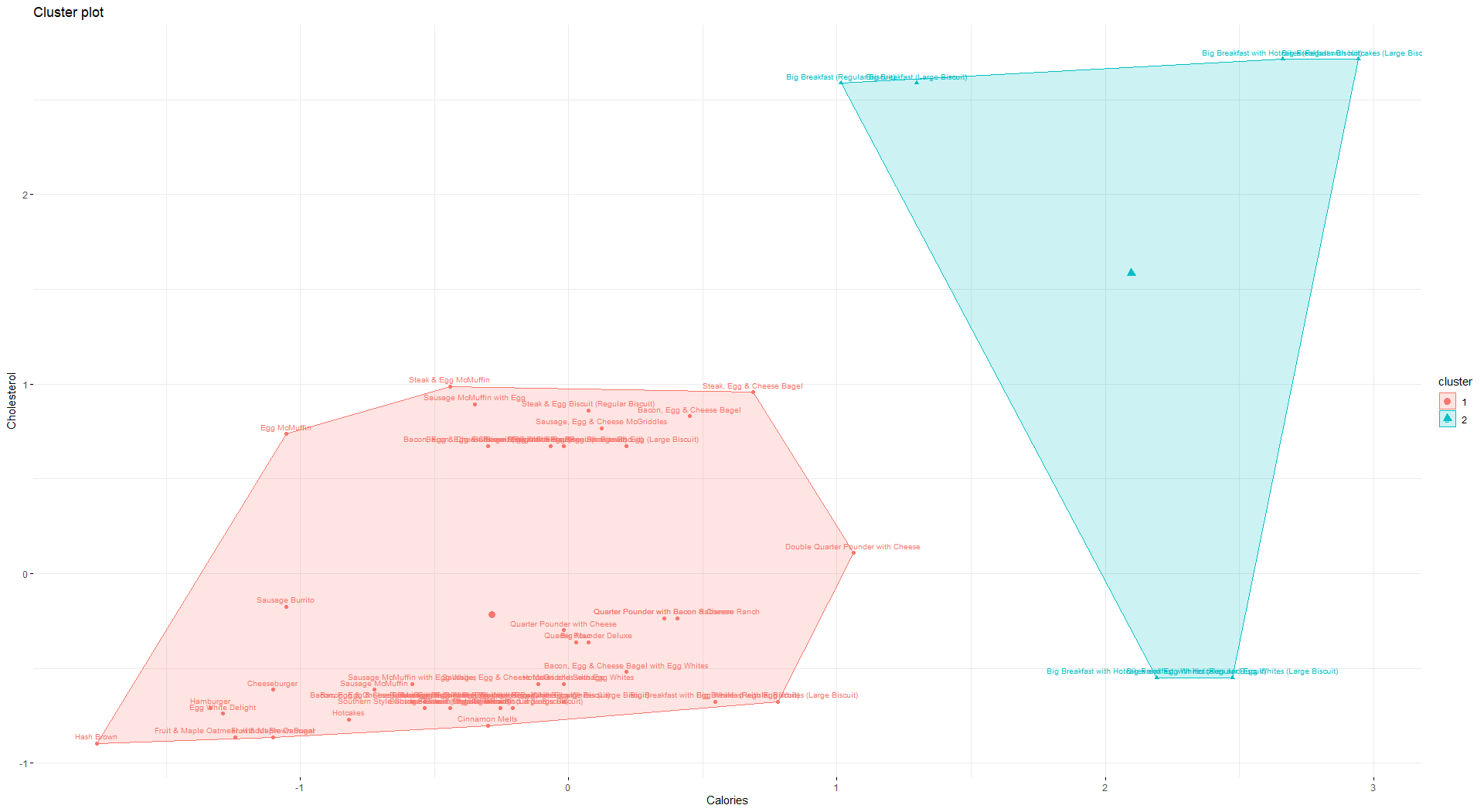






5. Остаточний варіант кластеризації

Підсумовуючи, варто відзначити, що ми надаємо перевагу результатам методу Average Linkage, тому що для нього було отримано значно більшу кореляцію відстаней (а отже, й більшу значущість кластеризації), до того ж, переважна більшість методів дала ідентичні результати, що дає підставу значно більше довіряти результатам даної кластеризації. До того ж, розбиття міст на категорії безпеки було доволі специфічним і могло відображати дещо викривлену картину. Тому остаточний варіант кластеризації такий:



4. Індекс Ренда



