



UNITED STATES MILITARY ACADEMY  
**WEST POINT**

# Mathematical Considerations for Precision Nutrition

Diana Thomas, Ph.D., F.T.O.S



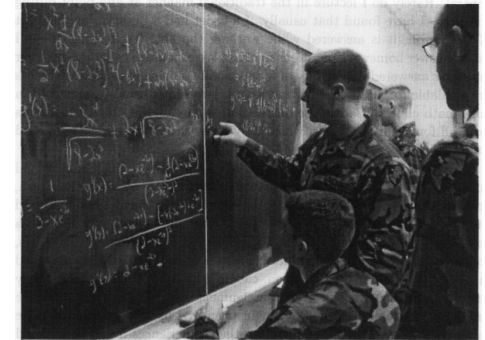
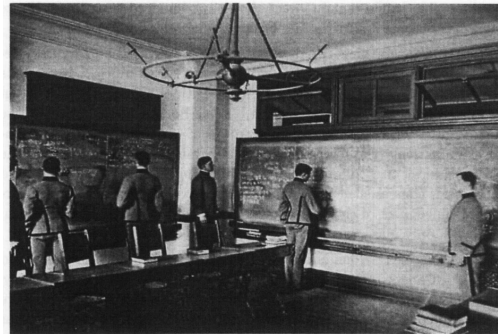


**Thank you to our Nutrition for Precision Health funding NIH  
AOD22022001**

To my NPH West Point team



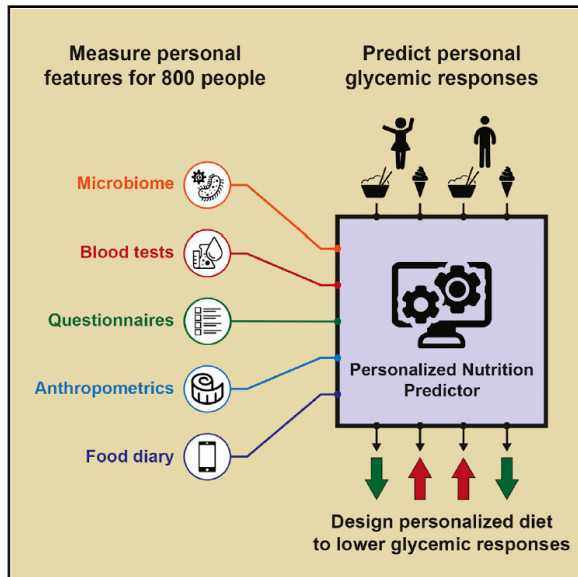
**MATH**





## Personalized Nutrition by Prediction of Glycemic Responses

David Zeevi,<sup>1,2,8</sup> Tal Korem,<sup>1,2,8</sup> Niv Zmora,<sup>3,4,5,8</sup> David Israeli,<sup>6,8</sup> Daphna Rothschild,<sup>1,2</sup> Adina Weinberger,<sup>1,2</sup> Orly Ben-Yacov,<sup>1,2</sup> Dar Lador,<sup>1,2</sup> Tali Avnit-Sagi,<sup>1,2</sup> Maya Lotan-Pompan,<sup>1,2</sup> Jotham Suez,<sup>3</sup> Jemal Ali Mahdi,<sup>3</sup> Elad Matot,<sup>1,2</sup> Gal Malka,<sup>1,2</sup> Noa Kosower,<sup>1,2</sup> Michal Rein,<sup>1,2</sup> Gili Zilberman-Schapira,<sup>3</sup> Lenka Dohnalová,<sup>3</sup> Meirav Pevsner-Fischer,<sup>3</sup> Rony Bikovsky,<sup>1,2</sup> Zamir Halpern,<sup>5,7</sup> Eran Elinav,<sup>3,9,\*</sup> and Eran Segal<sup>1,2,9,\*</sup>



## nature medicine

Explore content ▾ About the journal ▾ Publish with us ▾ **Subscribe**

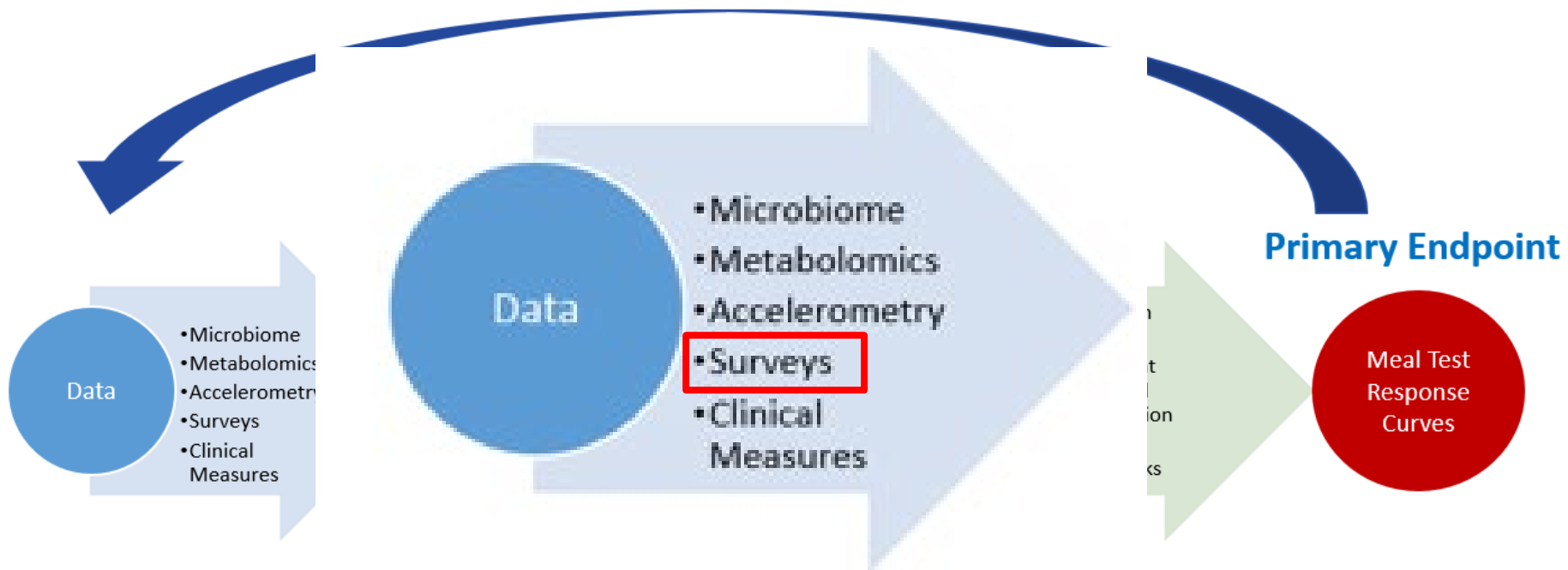
[nature](#) > [nature medicine](#) > [articles](#) > article

Article | [Published: 11 June 2020](#)

## Human postprandial responses to food and potential for precision nutrition

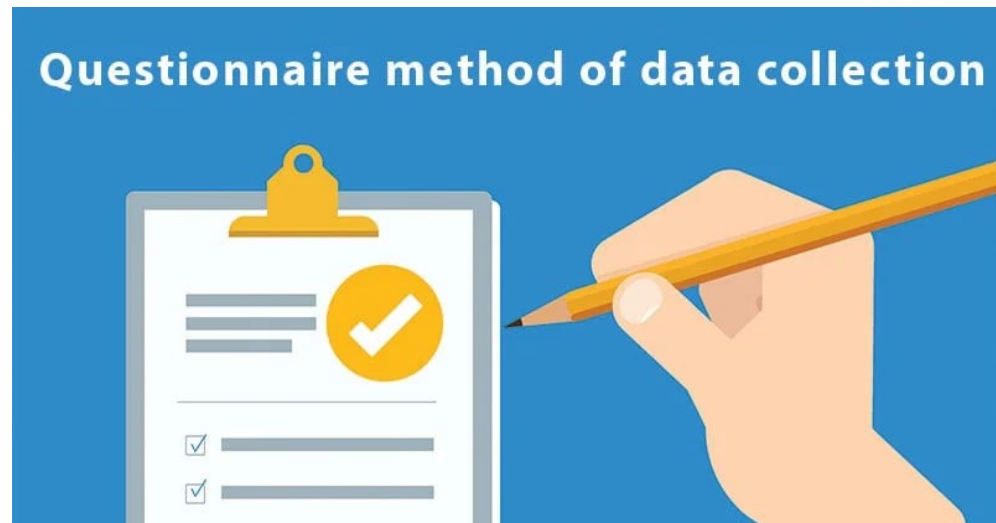
[Sarah E. Berry](#), [Ana M. Valdes](#), [David A. Drew](#), [Francesco Asnicar](#), [Mohsen Mazidi](#), [Jonathan Wolf](#), [Joan Capdevila](#), [George Hadjigeorgiou](#), [Richard Davies](#), [Haya Al Khatib](#), [Christopher Bonnett](#), [Sajaysurya Ganesh](#), [Elco Bakker](#), [Deborah Hart](#), [Massimo Mangino](#), [Jordi Merino](#), [Inbar Linenberg](#), [Patrick Wyatt](#), [Jose M. Ordovas](#), [Christopher D. Gardner](#), [Linda M. Delahanty](#), [Andrew T. Chan](#), [Nicola Segata](#), [Paul W. Franks](#) & [Tim D. Spector](#)

[Nature Medicine](#) **26**, 964–973 (2020) | [Cite this article](#)





- The algorithm for clustering needs to be one developed for Likert type survey data which is ordinal and categorical.
- The usual methods (k-means, hierarchical) do not apply here.
- Discrete data is often more challenging to work with.





## STATISTICAL ANALYSIS **AND** DATA MINING

RESEARCH ARTICLE

 Open Access



# An efficient *k*-modes algorithm for clustering categorical datasets

Karin S. Dorman, Ranjan Maitra 

First published: 18 September 2021 | <https://doi.org/10.1002/sam.11546> | Citations: 3

**Funding information:** National Institute of Food and Agriculture, IOW03617; IOW03717; United States Department of Agriculture





July 2023

## Under the hood of k-modes

Diana Thomas

$$X = \{X_1, X_2, \dots, X_n\}$$

$$X_i = [x_{i,1}, x_{i,2}, \dots, x_{i,m}]$$

each  $x_{i,j}$  has a range in the categorical values

I believe

So I think there are  $m$  questions on the survey and  $X_i$  represents the  $i^{\text{th}}$  respondent and  $x_{i,j}$  represent the  $i^{\text{th}}$  respondent's response on the  $j^{\text{th}}$  survey question. There are  $n$  respondents.





UNITED STATES MILITARY ACADEMY  
**WEST POINT**

# Step 2 Program the method on an example dataset



International  
**Weight Control**  
Registry

**August 2023**



**Dr. James O Hill, PhD**



**Dr. Susan B Roberts, PhD**

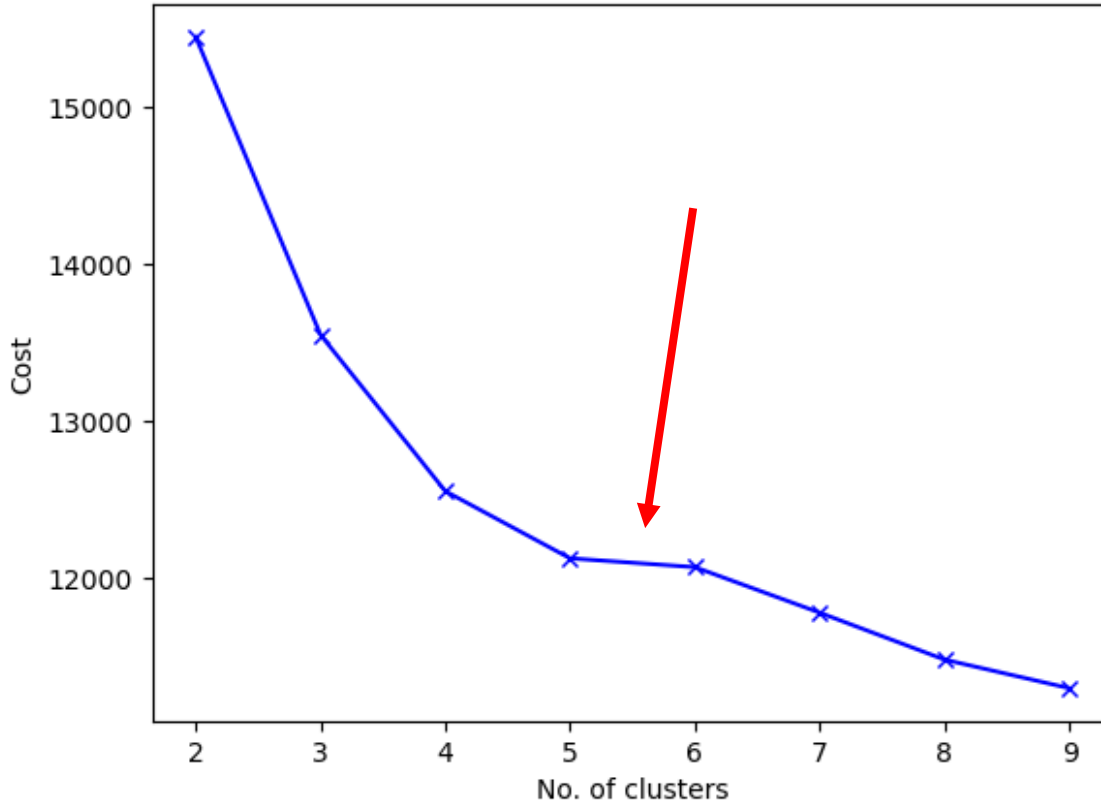






# Step 2 Program the method on an example dataset

Elbow Method For Optimal k



N=1652  
The elbow of the scree plot is 5 or 6.

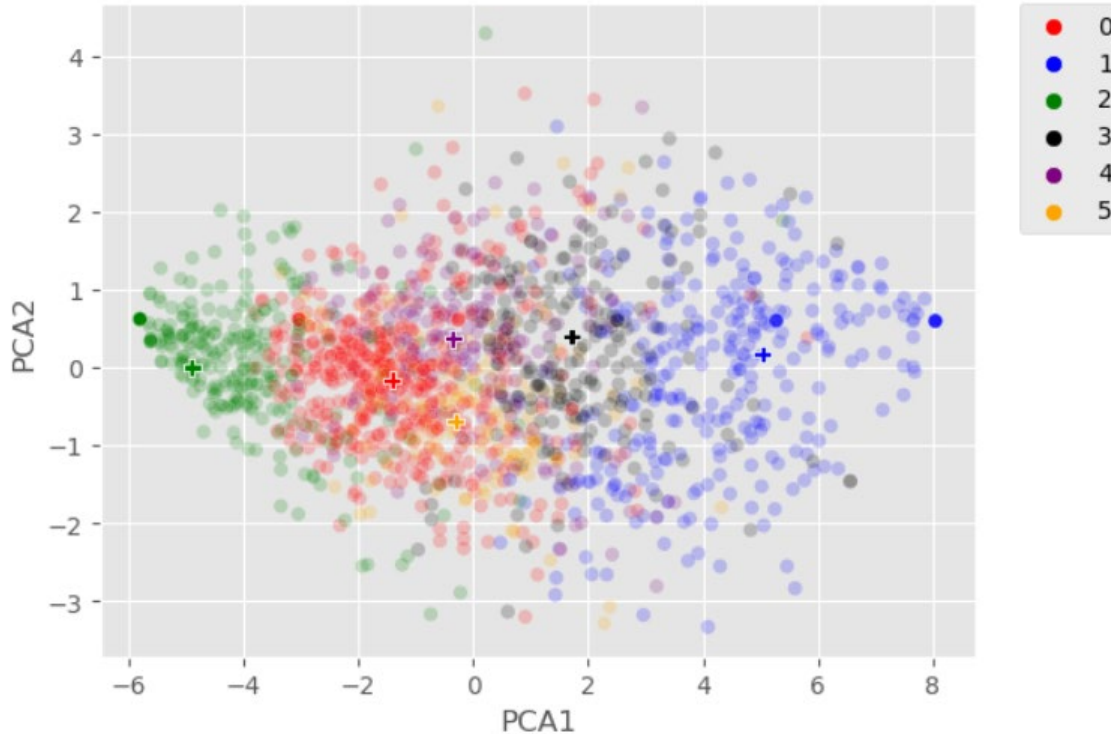
I used 6.  
This means we grouped 1652 people into 6 clusters.





# Step 3: Sniff checks visualization

October 2023



Sniff check-do we visually see distinct clusters?

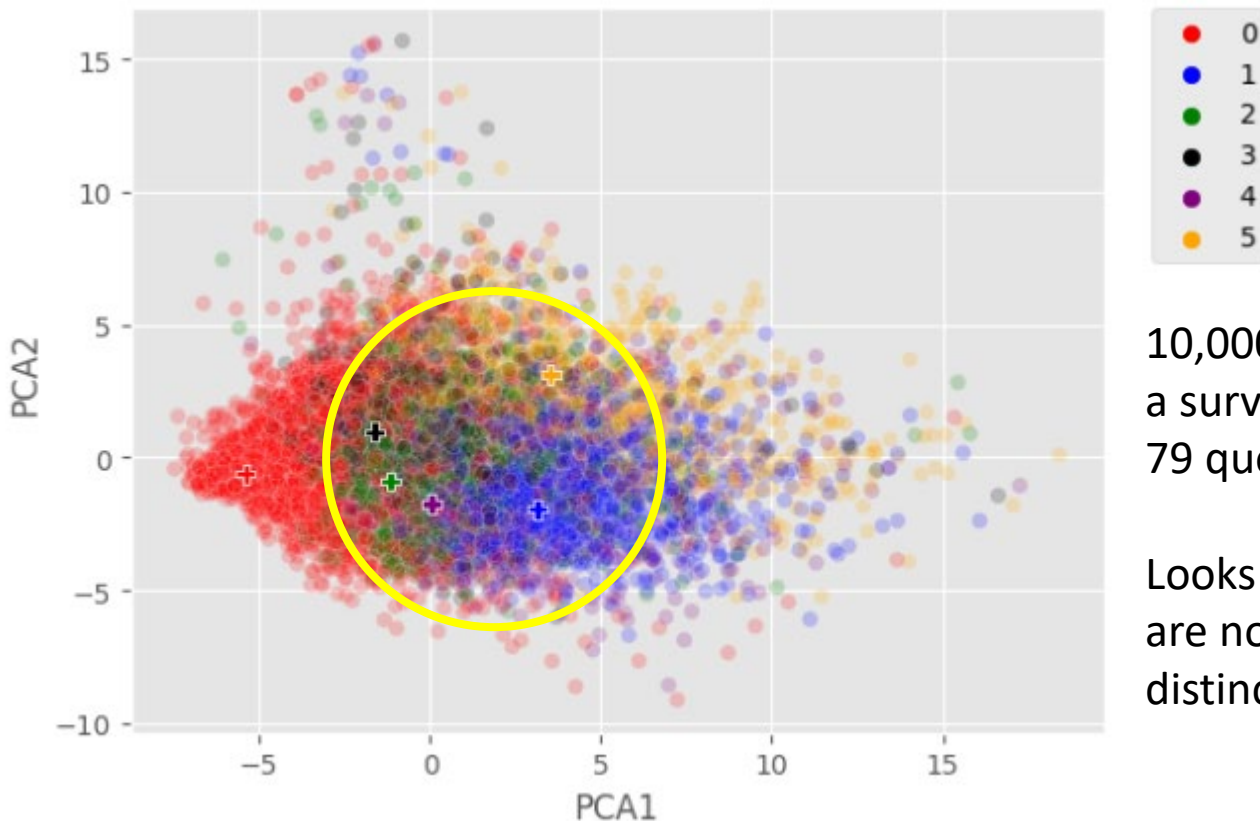
Yes: but it looks like we have 5 not 6 clusters.





# Step 4: Expand SDH Questionnaire AoU Workbench

December 2023



10,000 observations on a survey consisting of 79 questions.

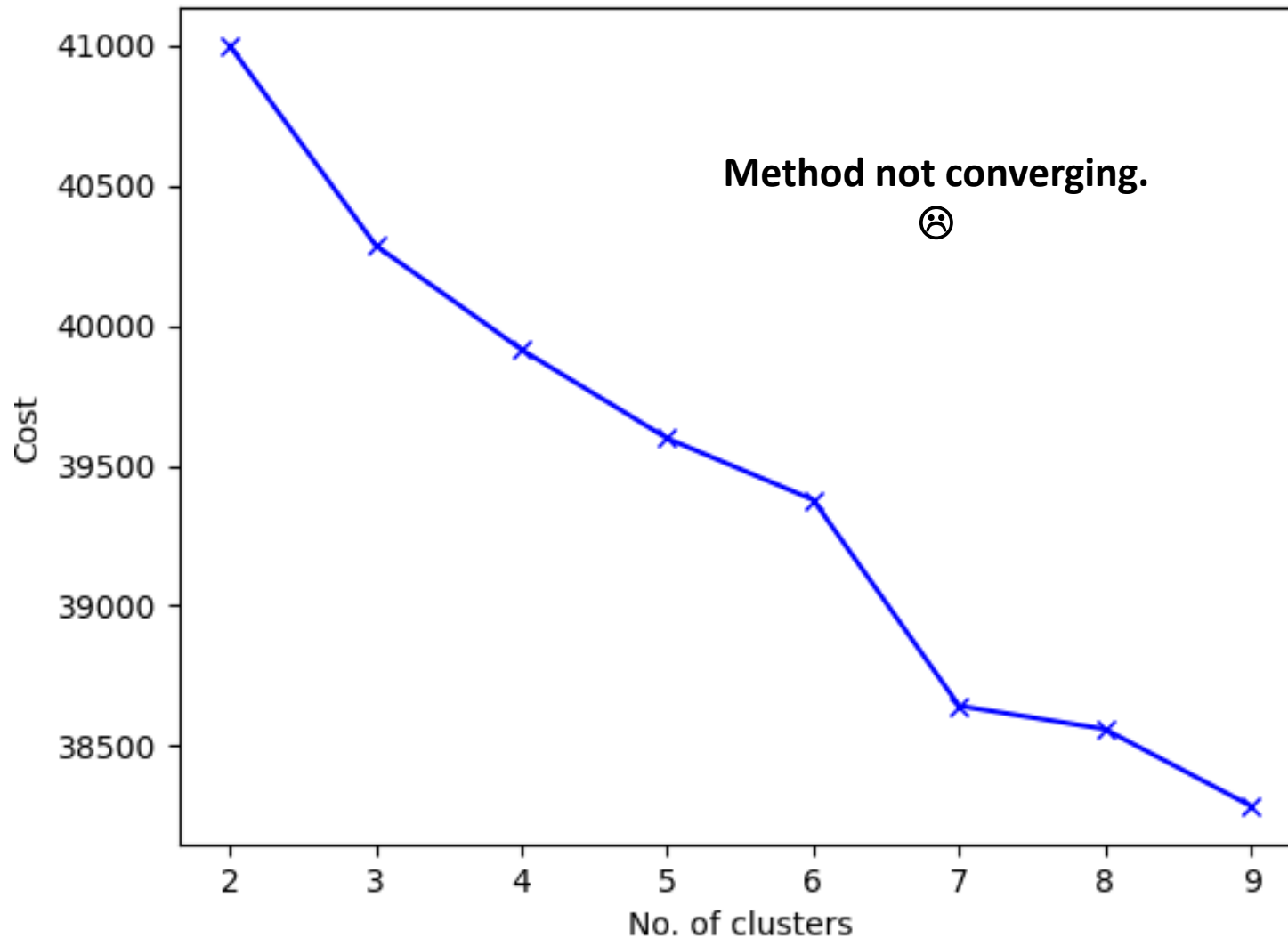
Looks like the clusters are not settling into distinct groups.





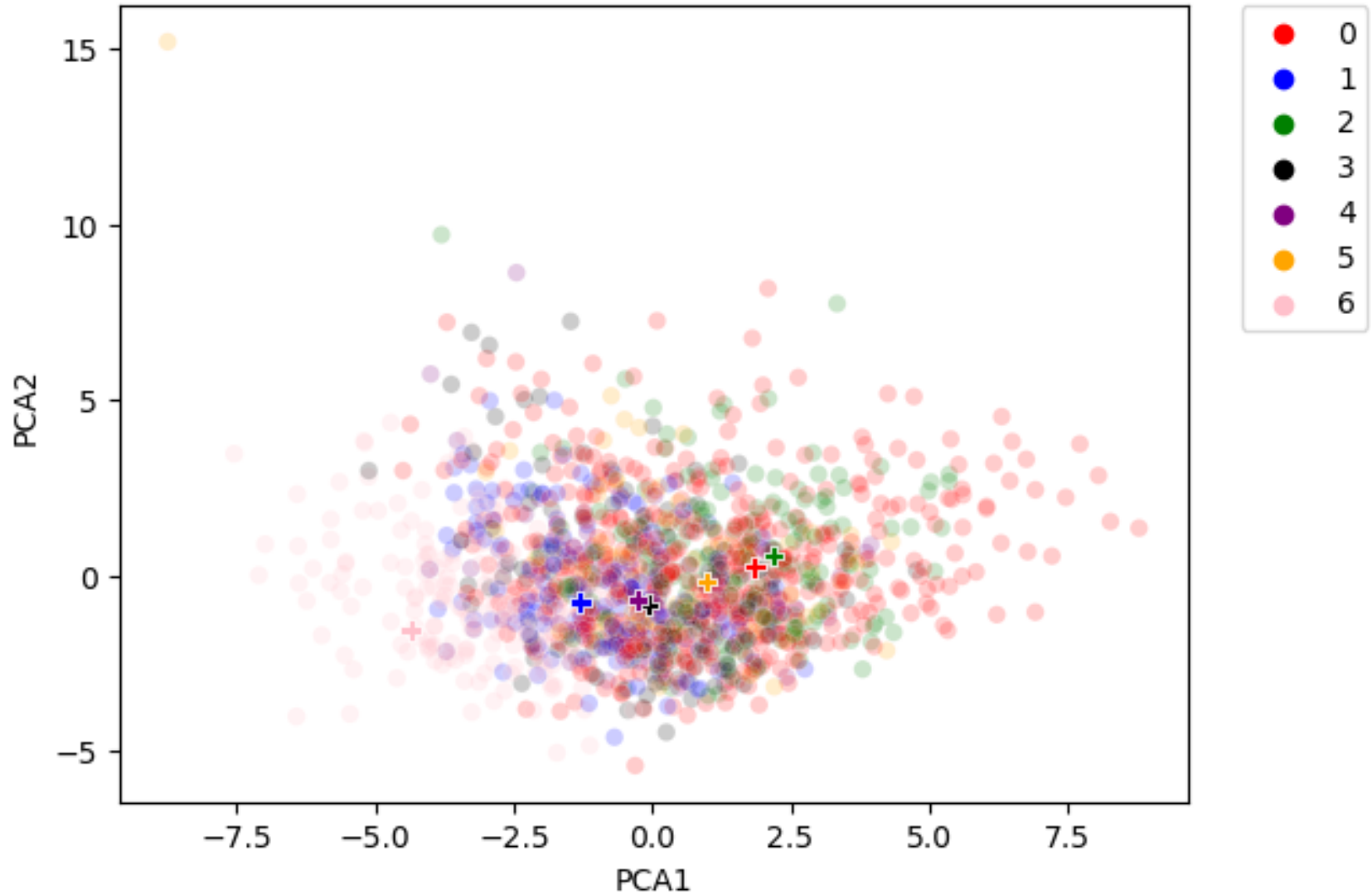
Jan-Feb 2024

### Elbow Method For Optimal k





Jan-Feb 2024





## Step 5: Investigate Why is this happening?

Maybe the survey number of questions is too much.  
Maybe the discrete nature causes non-convergence.

- Solution: Break the survey down into smaller groups.
- Solution: Start over with other methods. ☹️

**Now**





- Always communicate with your AI people from the start. Writing Group: Homer Simpson, Seymour Skinner, Stat Person, Troy McClure

“most people think of statisticians as Taco Bell...go to the window and order a crunch wrap t-test and a Baha regression”. The chef can’t come back and say I think you really prefer a burrito wrapped in a proportions test.

- Allow time for data preparation, human learning, human judgement, restarting, unforeseen results (not hypothesis testing, not inference), innovation (e.g., NLP and Likert clustering).

“You ordered the crunch wrap, but I think you need the burrito for the data you have.”

- Understand that data science is a big umbrella. What content matter expert do you wish to work with? (e.g., my strengths are NLP and differential equations. Rob Knight’s strengths are ML associated to the microbiome).
- Take into consideration publishing and granting ethics. How much time/effort will this take? Should the first/last authors be content matter experts? Journals can consider a first clinical author and first AI author.
- **Overcommunicate** For AI to be used in precision nutrition-we can’t be thought of as support staff.

“It's important because you need to have the intellectual space between friends to ask those what may seem as 'dumb' questions (on both sides). We accept that each other is the expert of our own domains and we lean heavily on each other to bring that expertise to the table. We also don't ride each others coat tails!”







1. Analyze a complex computing problem and to apply principles of computing and other relevant disciplines to identify solutions.
2. Design, implement, and evaluate a computing-based solution to meet a given set of computing requirements in the context of the program's discipline.
3. Communicate effectively in a variety of professional contexts.
4. Recognize professional responsibilities and make informed **and equitable** judgments in computing practice based on legal and ethical principles.
5. Function effectively as a member or leader of a team engaged in activities appropriate to the program's discipline.





# Different categories that have similar job ads

## Data Analyst



### Differentiating Competencies:

- Storytelling & Influence
- Descriptive Analysis
- Impactful Data Visualization
- Business Acumen

### Representative Titles:

- Data Analyst
- Business Analyst
- Analytics Specialist
- Analytics Translator

## Data Scientist



### Differentiating Competencies:

- Python Programming
- Statistical Modeling & Machine Learning
- Descriptive & Inferential Analysis

### Representative Titles:

- Data Scientist
- Data Science Specialist
- Machine Learning Scientist
- Statistical Data Scientist

## Analytics Engineer



### Differentiating Competencies:

- Building data structure and data pipelines
- Technology Deployment & Operationalization
- Data Strategy

### Representative Titles:

- Analytics Engineer
- Data Engineer
- Machine Learning Engineer
- Data Engineering Specialist

## Data Visualization Engineer



### Differentiating Competencies:

- Data & Analytics Visualization
- Human Centered Design
- Application Development
- Dashboarding

### Representative Titles:

- BI Developer
- Analytics Interface Dev.
- Visualization Engineer
- Front-end Developer

## Analytics Manager



### Differentiating Competencies:

- People Management
- Program Management
- Thought Leadership
- Business Acumen

### Representative Titles:

- Analytics Manager
- Analytics Translator
- Business Analytics Mgr.
- Data Science Manager

