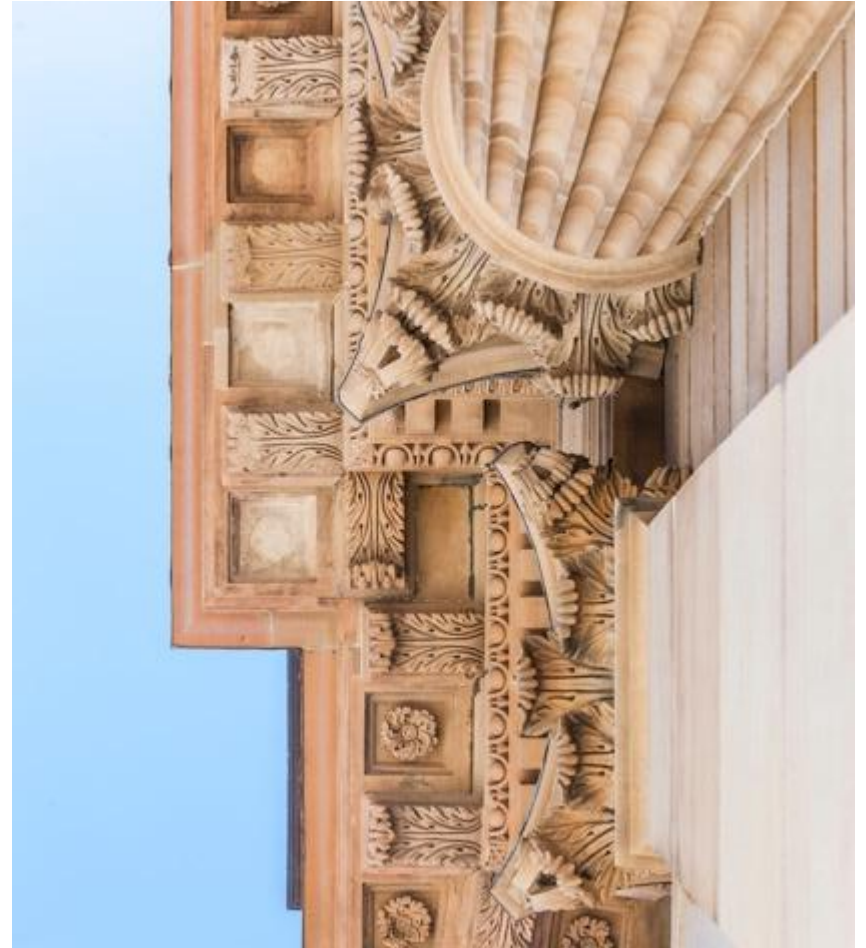# DATA/NUTM3888 PROJECT PITCH

Group 2

WEEK 7 SEMESTER 2 2021

Members: Ash Dravid, Layan Rayan, Andrew Hwang, Qi Zhang, Jackson Karrasch

RESEARCH QUESTION: **What are the effects of lifestyle factors on the BMI of people who have certain diet compositions?**

THE UNIVERSITY OF SYDNEY

# Background Information & Literature Review

BMI (Body Mass Index) - Many Contributing Factors

High Protein - Health conscious, better sleep quality, higher physical activity, low*er* energy intake.

Low Carbohydrate - Micronutrient concerns (Bs, D, E, C), low quality diet.

High Fat - Positive energy balance, low quality diet, same problems with low carbohydrate.

Higher BMI - High Fat, Low*er* Carbohydrate, Low Protein



Eat a variety of healthy foods each day

Have plenty of vegetables and fruits

Eat protein foods

Make water your drink of choice

Choose whole grain foods

# Feasibility Analysis

- Demographic Variables
    - Age
    - Sex
    - Socioeconomic Decile
- Lifestyle variables
    - Total Exercise Time
    - Sleep time
    - Smoking Status
- All variables exist in dataset

# Cluster Formation

- PCA conducted on percentage of macronutrients in diets
- K-means and hierarchical clustering used
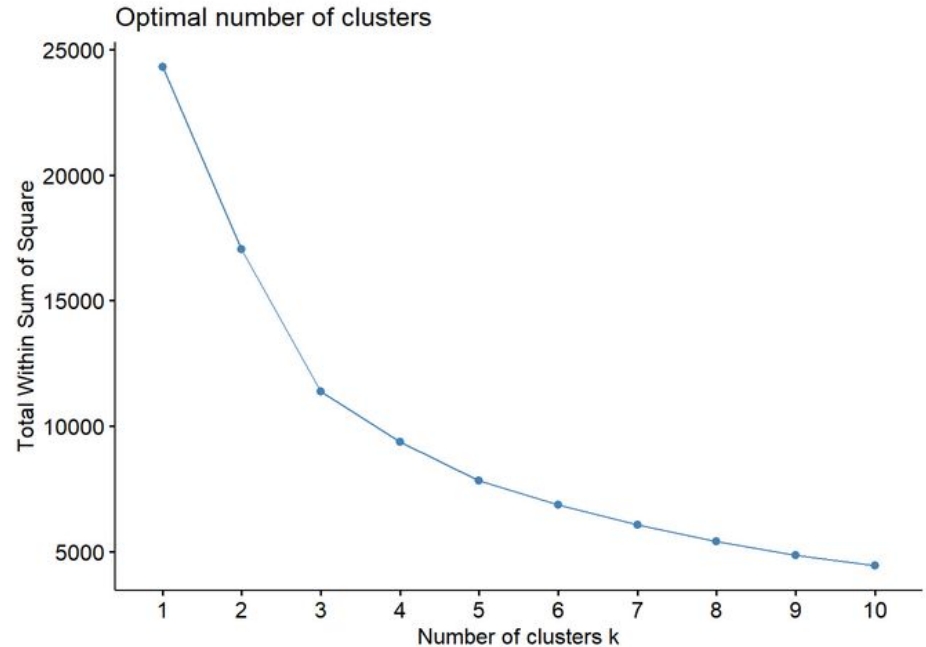- K-means with K = 4 clusters selected

Optimal number of clusters

Figure 1. Within sums of squares for each value of *k* used in k-means clustering. Elbow is roughly at k = 4 so this was used.

# Cluster Interpretation

Table 1. Interpretation of Clusters

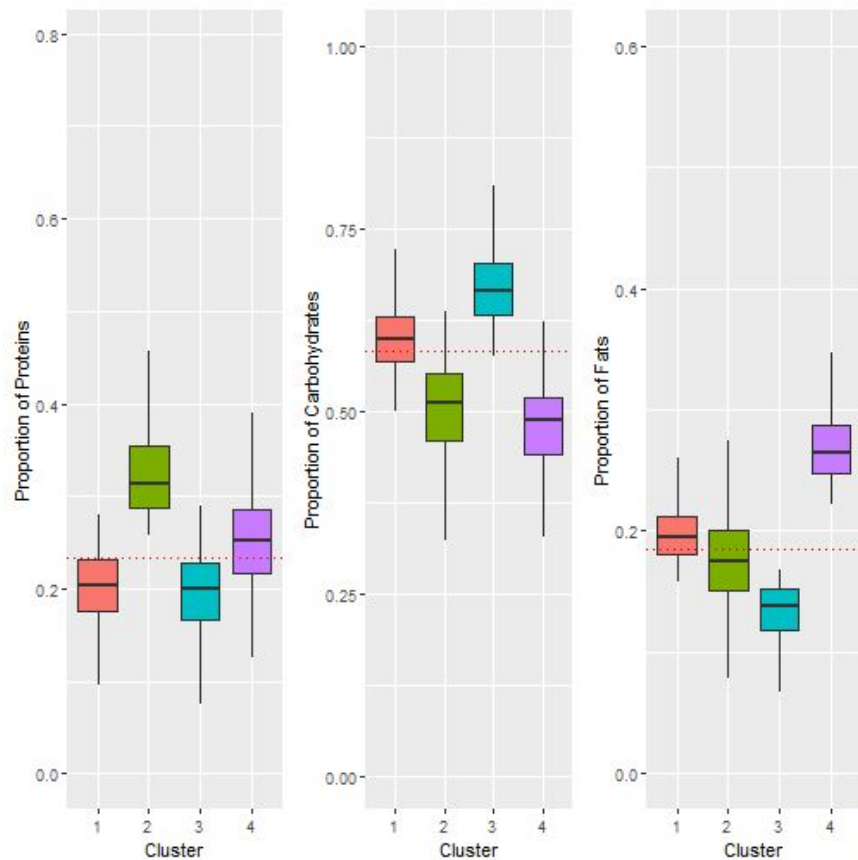| Cluster | Protein Proportion | Carbohydrate Proportion | Fat Proportion |
|---------|--------------------|-----------------------|----------------|
| 1 | Average | Average | Average |
| 2 | High | Low | Average |
| 3 | Average | High | Low |
| 4 | Average | Low | High |



Figure 2. Distribution of macronutrients in clusters
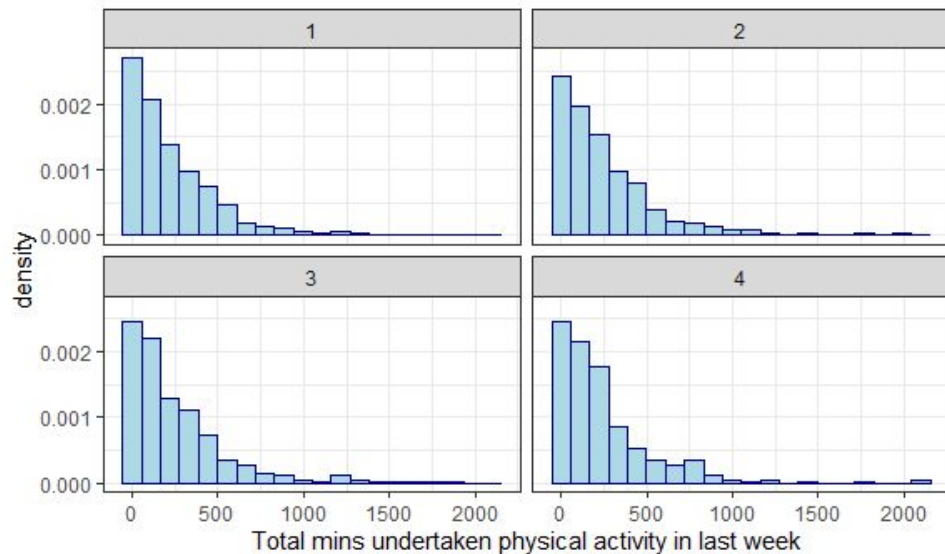
# Initial Data Visualisations



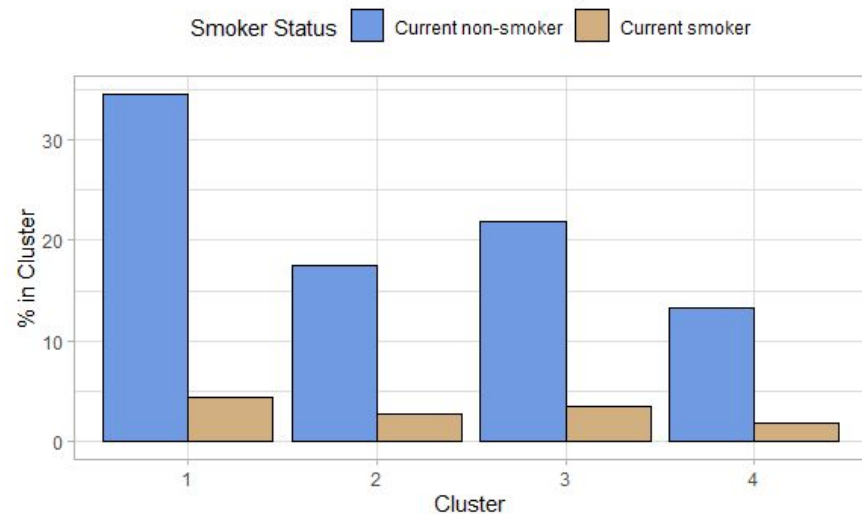Figure 3. Distribution of total physical activity (min) in clusters



Figure 4. Current smoking status percentage in each cluster

# Future Steps

- Utilise linear regression and feature selection to identify statistically significant features to predict BMI
- Cross Validation to assess model performance
- Possible other ML algorithms: Regression Trees, Neural networks
- Identify differences between dietary groups
- Use literature to compare and contrast our findings
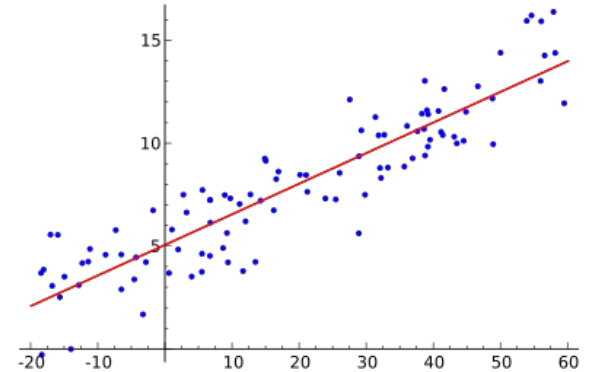
# References

- Gherasim, A., Arhire, L. I., Niță, O., Popa, A. D., Graur, M., & Mihalache, L. (2020). The relationship between lifestyle components and dietary patterns. *The Proceedings of the Nutrition Society*, *79*(3), 311–323. https://doi.org/10.1017/S0029665120006898
- Štefan, L., Čule, M., Milinović, I., Juranko, D., & Sporiš, G. (2017). The Relationship between Lifestyle Factors and Body Compositionin Young Adults. *International journal of environmental research and public health*, *14*(8), 893. https://doi.org/10.3390/ijerph14080893
- Paeratakul, S., Popkin, B. M., Keyou, G., Adair, L. S., & Stevens, J. (1998). Changes in diet and physical activity affect the body mass index of Chinese adults. *International journal of obesity and related metabolic disorders : journal of the International Association for the Study of Obesity*, *22*(5), 424–431. https://doi.org/10.1038/sj.ijo.0800603