USDA National Nutrient Analysis

By:
Kriti Srivastava

Introduction

- **Background**
  - Healthy eating is an essential part of a healthy lifestyle.
  - Nutrition is the preeminent part of everyone's life starting from the new born baby to oldsters.
  - Gym can help shape physic of human but nutritions also play a vital role to achieve the goal.
  - The modern diet is mostly comprised of processed food and lacks essential nutrients that lead to : diabetes, hypertension, anemia, and obesity.
  - This study is to analyze different food products with the nutrients present and to know our food little better.
- **Data**
  - In this study we are introducing the various approaches performed on USDA database containing the food and nutrition information.
  - Sample dataset contains: 8,618 rows , 6 Categorical variables, and 38 Continuous variables.

Literature Review

- USDA's National Food and Nutrient Analysis Program: Food Sampling

  Pehrsson, P. R., Haytowitz, D. B., Holden, J. M., Perry, C. R., & Beckler, D. G. (2000). USDA's National Food and Nutrient Analysis Program: Food Sampling. *Journal of Food Composition and Analysis*, *13*(4), 379–389. https://doi.org/10.1006/jfca.1999.0867

- Dietary patterns derived from principal component- and k-means cluster analysis: Long-term association with coronary heart disease and stroke.

  Stricker, M., Onland-Moret, N., Boer, J., Schouw, Y. V., Verschuren, W., May, A., . . . Beulens, J. (2013). Dietary patterns derived from principal component- and k-means cluster analysis: Long-term association with coronary heart disease and stroke. *Nutrition, Metabolism and Cardiovascular Diseases, 23*(3), 250-256. https://doi.org/10.1016/j.numecd.2012.02.006

- Analysis of patterns of food intake in nutritional epidemiology: Food classification in principal components analysis and the subsequent impact on estimates for endometrial cancer.

  McCann, S., Marshall, J., Brasure, J., Graham, S., & Freudenheim, J. (2001). Analysis of patterns of food intake in nutritional epidemiology: Food classification in principal components analysis and the subsequent impact on estimates for endometrial cancer. Public Health Nutrition, 4(5), 989-997. doi:10.1079/PHN2001168

Research questions and Methods

- Is it possible to explain the food products based on classification of their nutrients description?

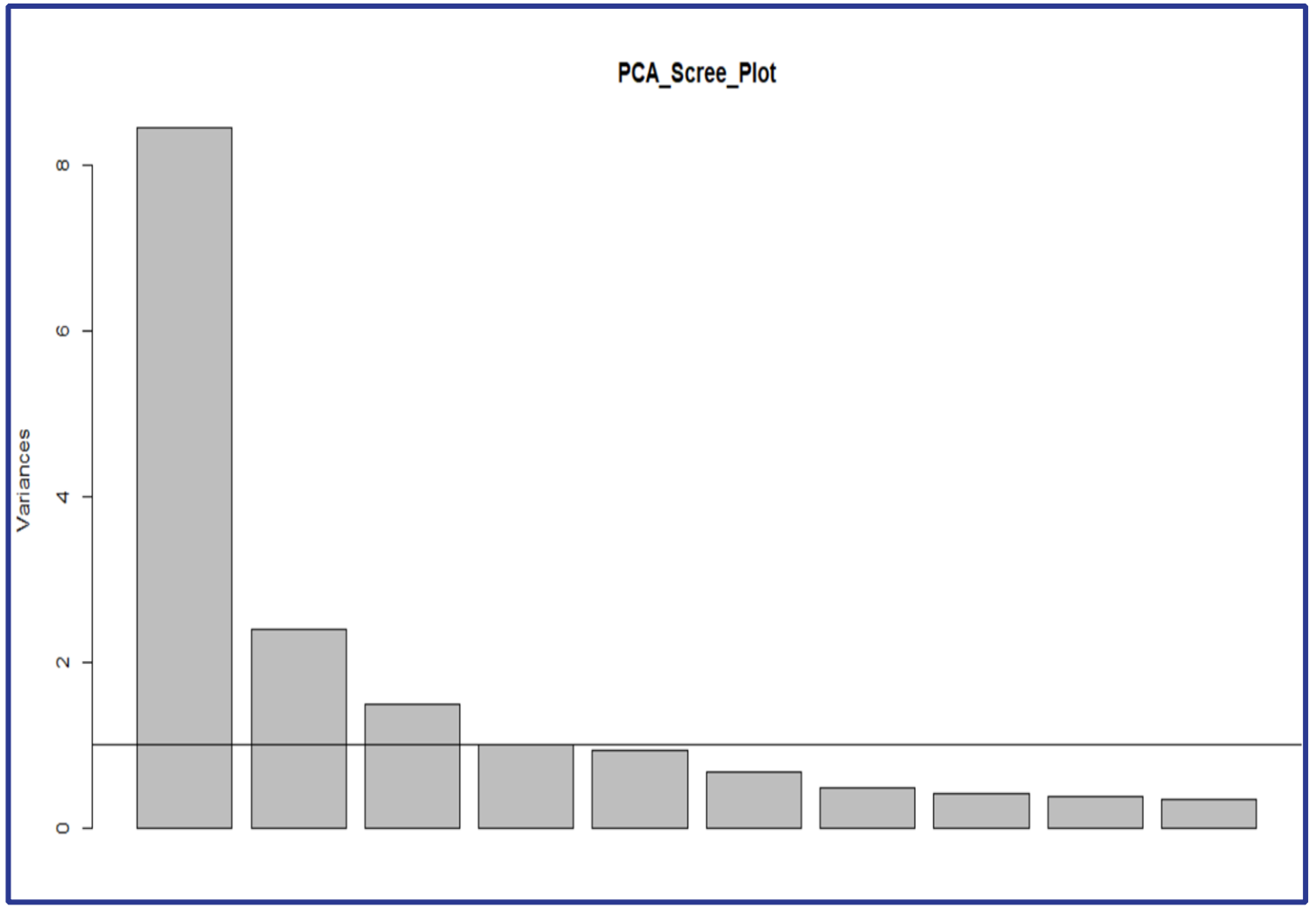
  Method: Principal Component Analysis

- If yes, is there any relationship between the nutrients ?

  Method: Canonical correlation analysis

Discussion And Results - Principal Component Analysis

- Explains the food products based on classification of their nutrients description
- Perform dimension reduction: for the next procedure i.e. Canonical Correlation Analysis on the resultant components of PCA Analysis
- The scree plot shows three components that explains most of the variance.
- Assumptions for factorability
  - **KMO Test:** Overall MSA = 0.89
  - **Bartlett's Test of Sphericity**: p-value < 2.22e-16 (very small which means we have enough variance in the data so we can perform factor analysis)
  - **Reliability Analysis using Cronbach's Alpha:** raw_alpha = 0.92 Hence, The sample data is good to perform PCA.



Discussion And Results - Principal Component Analysis

- Rotate="varimax", nfactors=3, cutoff=.48

- PCA produced 3 components:
  - RC3 : rich in fiber, Vitamin E as well as carbohydrates. Therefore, this component explains the sets of **FiberRichedCarbsFood** variables**.**

  - RC2: rich in Vitamins, iron, and zinc.
    - Foods rich in vitamin C helps in the absorption of iron
    - Good for anemic people therefore, this component explains the sets of **Anti-AnemicFood** variables.

  - RC1: rich in protein and fats therefore, this component explains the sets of **HighProteinFood** variables.

Nutrients	*FiberRichedCarbsFood*	*Anti-anemicFood*	*HighProtienFood*
	RC3	**RC2**	**RC1**
Carb_g	0.709		
fiber_g	0.826		
vitE_mg	0.547		
Copper_mcg	0.756		
Magnesium_mg	0.640		0.547
Manganese_mg	0.832		
VitA_mcg		0.737	
VitB6_mg		0.800	
VitB12_mcg		0.782	
VitC_mg		0.799	
Riboflavin_mg		0.665	
Iron_mg	0.581	0.596	
Zinc_mg		0.603	
Protien_mg			0.822
Fat_g			0.641
Calcium_mg			0.686
Phosphorus_mg			0.853
Selenium_mcg			0.672

	RC3	RC2	RC1
SS loadings	4.224	4.155	3.962
Proportion Var	0.235	0.231	0.220
Cumulative Var	0.235	0.465	0.686

Discussion And Results - Factor Analysis

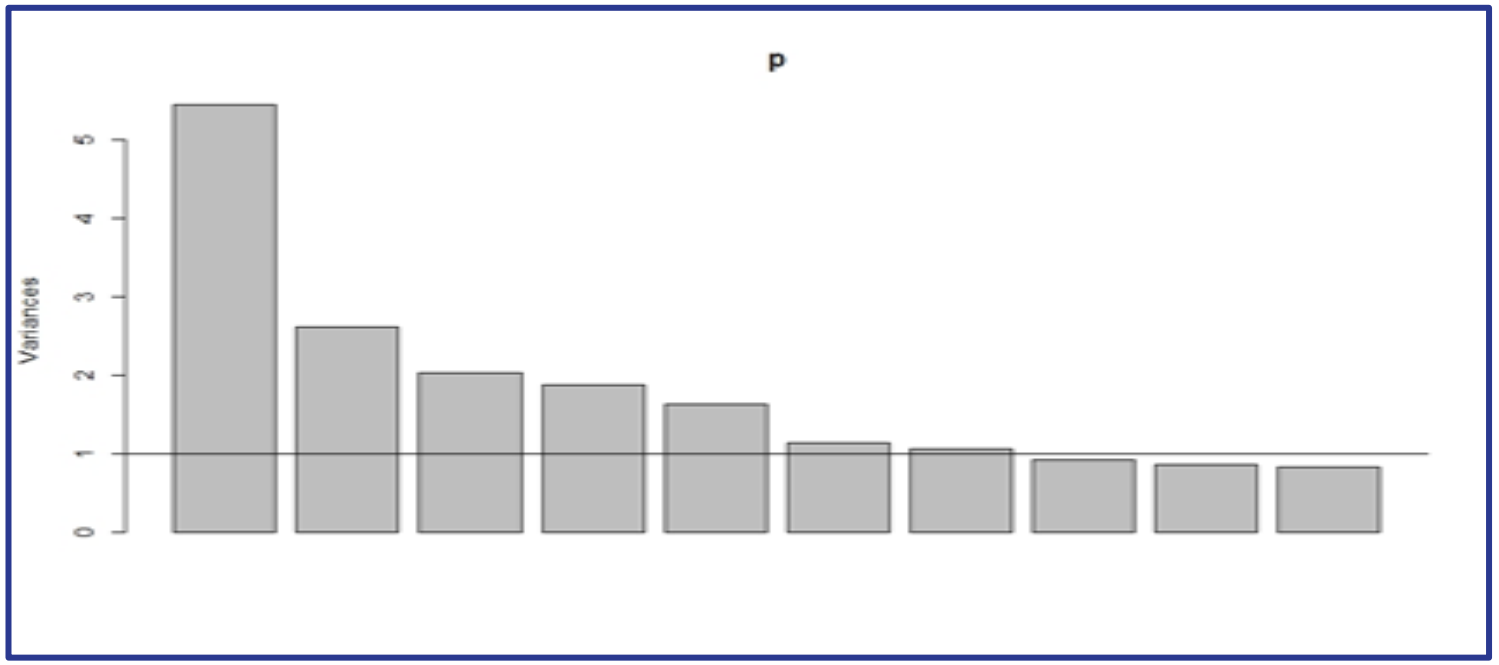
- Factor analysis applied on the USDA National Nutrient dataset having total 45 columns and 8618 entries.

- Tests:
  - Cronbach's alpha **: 0.54**
  - Kaiser-Meyer-Olkin (KMO) Test for Sampling Adequacy **: 0.64**
  - Bartlett's Test of Sphericity **: chi-square: 120430.831**

p-value < 0.001

Discussion And Results - Factor Analysis

Scree plot. A scree plot is a plot of the Eigenvalues against the number of factors in order of extraction.



Factor Matrix

Cutoff is 0.55

Displayed matrix consisted of 5 Factors.

- **Dietary**
- **Minerals**
- **Aminos**
- **Strength**
- **Disaccharides**

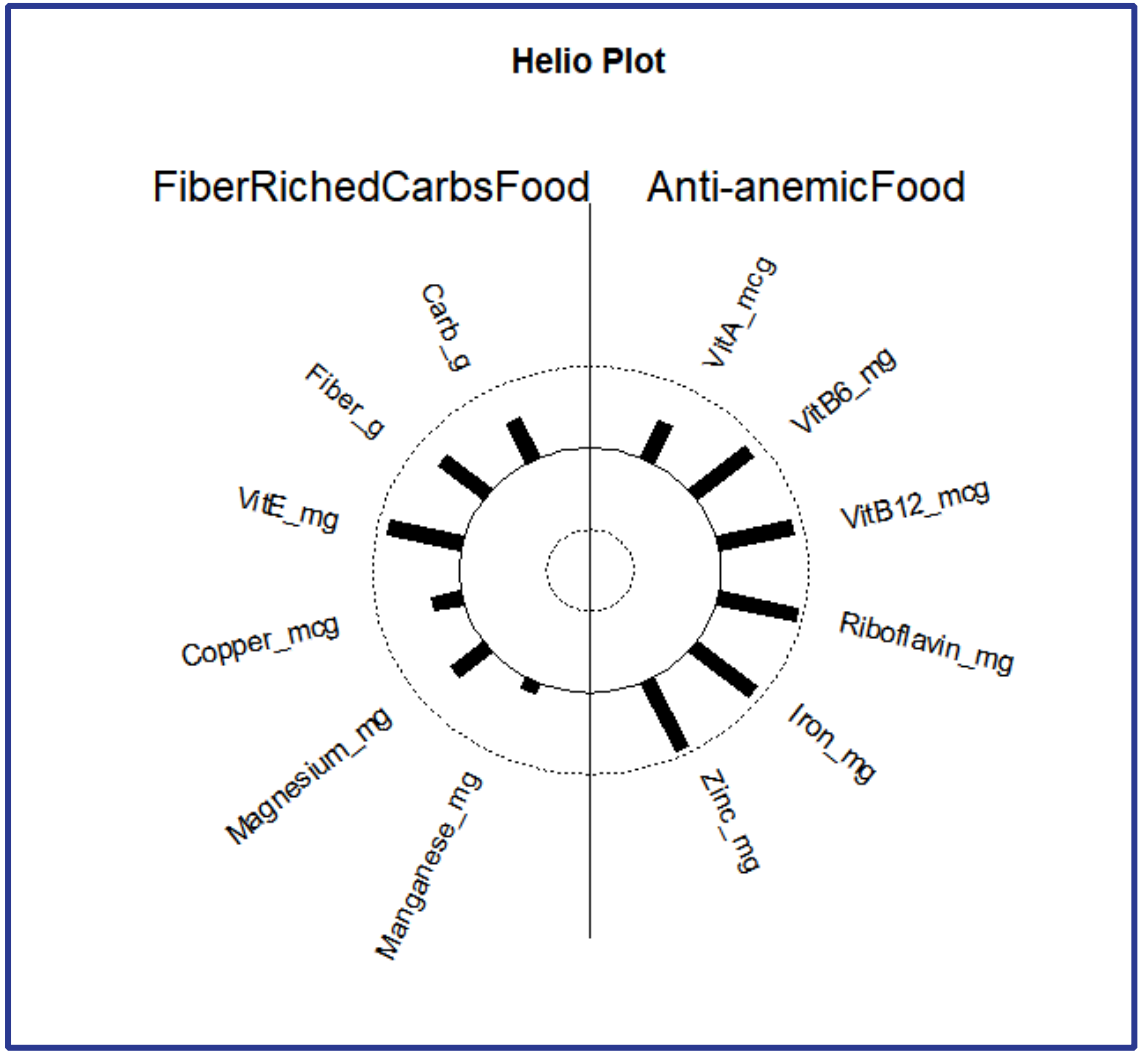
```
> print(fit$loadings, cutoff=.5, sort=T)

Loadings:
             Factor1 Factor2 Factor3 Factor4 Factor5
VitB6_mg      0.736
Folate_mcg    0.681
Niacin_mg     0.825
Riboflavin_mg 0.783
Thiamin_mg    0.674
Iron_mg       0.564
VitA_mcg              0.779
VitB12_mcg            0.733
Copper_mcg            0.724
Protein_g                     0.975
Energy_kcal                           0.834
Fat_g                                 0.987
Carb_g                                        0.959
Sugar_g                                       0.596
VitC_mg
VitE_mg
Calcium_mg
Magnesium_mg
Manganese_mg
Phosphorus_mg
Selenium_mcg
Zinc_mg

               Factor1 Factor2 Factor3 Factor4 Factor5
SS loadings      3.631   1.958   1.912   1.844   1.808
Proportion Var   0.165   0.089   0.087   0.084   0.082
Cumulative Var   0.165   0.254   0.341   0.425   0.507
```

Factors Contributing Variance

	Dietary	Minerals	Aminos	Strength	Disaccharides
SS Loadings	3.631	1.958	1.912	1.844	1.808
Proportion Variance	0.165	0.089	0.087	0.084	0.082
Cumulative Variance	0.165	0.254	0.341	0.425	0.507

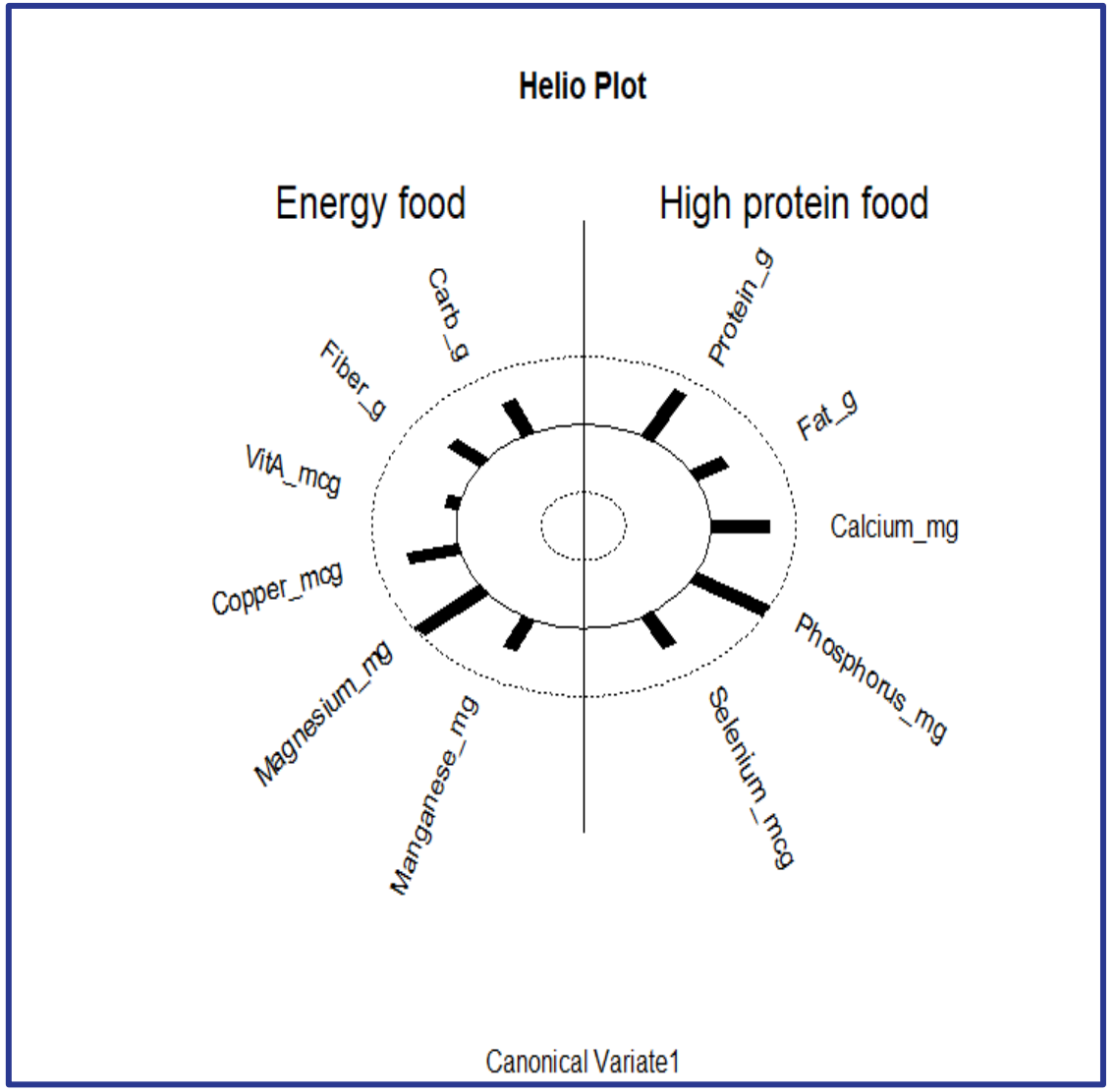
Discussion And Results - Canonical Correlation Analysis

- This makes VitE_mg and Fiber_g the most important and influential variables for its covariate FiberRichedCarbsFood.
- VitB6_mcg, VitB12_mcg, Riboflavin_mg, Iron_mg , and Zinc_mg are the most important and influential variables for its covariate Anti-Anemic Food.
- Also, the variables seem to have positive correlation among themselves.



Discussion and Results - Canonical Correlation Analysis

- Magnesium is the most important influential variable for the covariate Energy food or fiber rich carb food (RC3).
- Phosphorus is the most important influential variable for the covariate high protein food(RC1).



Future Works

- This study is done on the outdated version SD27 of USDA National Nutrient Database.

- This study only shows how the food products are rich or deficits in different good components however, it does not cluster the food products or labels the good product into categories on the basis of their nutrient factors.

- This study does not include the relationship between the components produced from the factor analysis which can be done in the future analysis.

Limitations

- The study do not identifying the amount / portion of the nutrients are needed in our diet.

- Absence of data to study the estimation of the risk of disease with nutrient patterns.

Conclusion

- Depending upon the need of the individual, it is possible to focus on food product that are rich in specific set of nutrient:
    - **Fiber - riched Carbohydrates :** Nutritious carbohydrate foods are rich source of fiber, as fiber itself is a form of carbohydrate. A high-fiber diet may protect against conditions like tips and heart disease.
    - **Anti-anemic nutrient** : includes high content of iron with vitamin C and zinc. anti-anemic diet is good for people struggling with anemia.
    - **High-Protein :** high protein diet includes proteins and good fats that is good for building muscles and are very important especially for athletes.

Conclusion

- The nutrients pattern are related to each other :
  - Fiber- rich Carbohydrates increases with the increase in Anti-anemic food nutrients also increases.
  - Similarly, as Fiber- rich Carbohydrates increases in the food items, high protein nutrients are also observed to be increased.

References

1. Pehrsson, P. R., Haytowitz, D. B., Holden, J. M., Perry, C. R., & Beckler, D. G. (2000). USDA's National Food and Nutrient Analysis Program: Food Sampling. *Journal of Food Composition and Analysis*, *13*(4), 379–389. https://doi.org/10.1006/jfca.1999.0867
2. US Department of Agriculture, Agricultural Research Service, Nutrient Data Laboratory. USDA National Nutrient Database for Standard Reference, Release 28 (Slightly revised). Version Current: May 2016. Internet: http://www.ars.usda.gov/ba/bhnrc/ndl
3. McCann, S., Marshall, J., Brasure, J., Graham, S., & Freudenheim, J. (2001). Analysis of patterns of food intake in nutritional epidemiology: Food classification in principal components analysis and the subsequent impact on estimates for endometrial cancer. Public Health Nutrition, 4(5), 989-997. doi:10.1079/PHN2001168
4. Stricker, M., Onland-Moret, N., Boer, J., Schouw, Y. V., Verschuren, W., May, A., . . . Beulens, J. (2013). Dietary patterns derived from principal component- and k-means cluster analysis: Long-term association with coronary heart disease and stroke. *Nutrition, Metabolism and Cardiovascular Diseases, 23*(3), 250-256. doi:10.1016/j.numecd.2012.02.006Uusitalo, L., Nevalainen, J., Salminen, I., Ovaskainen, M., Kronberg-Kippilä, C., Ahonen, S., . . . Virtanen, S. M. (2011). Fatty acids in serum and diet - a canonical correlation analysis among toddlers. Maternal & Child Nutrition, 9(3), 381-395. doi:10.1111/j.1740-8709.2011.00374.x https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1740-8709.2011.00374.x