USDA FDC Data Analysis

Team 58 - Payton Furr, Nayla Henriquez,
Andreas Mayr & Lianne Tal

Our Team

Lianne Tal

Payton Furr

Andreas Mayr

Nayla Henriquez

Agenda

01 Exploration & Research Questions

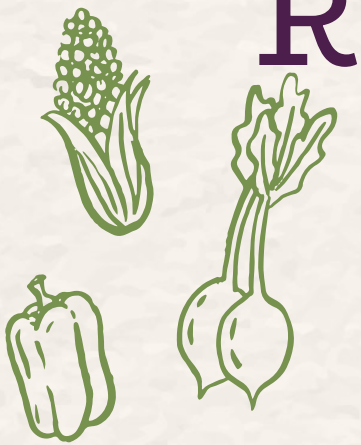
02 Data Preparation & Processing

03 Our Findings

04 Conclusion

01 Exploration & Research Questions

USDA FDC Data Set



Exploring the Data Set

- Data from USDA's Food Data Central (FDC)
  - Launched in 2019 for transparent food analysis
- Focus on two data types:
  - Foundation Foods (current dataset - 2023)
  - SR Legacy (historical data - 2013)
- Foundation Foods includes detailed values and metadata, while SR Legacy offers comprehensive historic data
- Manipulation of the dataset is necessary to group like food items together so they can be clearly visualized

Objective: Explore data to understand changes and constants between foods to analyze their nutrients and components for future food updates

Problem Statement

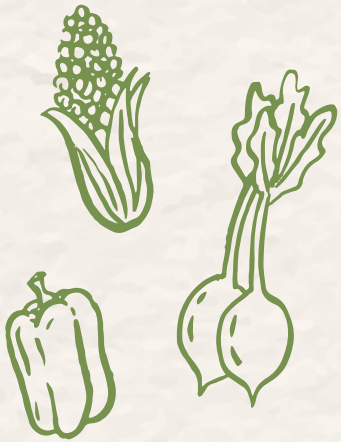
Addressing the challenge of understanding the dynamic changes in comparable foods over time is crucial for optimizing our analysis approach. Determining which components require ongoing examination and identifying those that remain constant will streamline our food evaluation process during updates.

https://www.nist.gov/image/usda-logo

Research Questions

1. What is the average change in nutrients for all the different food groups?
2. Which food groups have had the biggest change?
3. What are the top 10 nutrients that have changed the most between the SR and FF Mean values?
4. Does the mean value for SR Legacy fall between the minimum and maximum values of the FF Data for nutrients/components in equivalent foods?

02
Data Preparation & Processing



Data Cleaning and Preparation

We noticed that our data was hard to look at and make sense of because it did not contain the food categories or nutrient names.

```
main_df.head()
```

	FF_NDB	SR_NDB	food_category_id	FF Food description	SR Food description	Nutrient_id	rank	FF_Component	SR_Component	unit_name	...	SR Min	SR Max	Std_Error	SR Num_Data_pts	FF Mean per 100g	FF Min	FF Max	FF Median
0	16158	16158	16	Hummus, commercial	Hummus, commercial	1090	5500	Magnesium, Mg	Magnesium, Mg	MG	...	NaN	NaN	NaN	2	71.10	56.60	82.00	70.40
1	16158	16158	16	Hummus, commercial	Hummus, commercial	1258	9700	Fatty acids, total saturated	Fatty acids, total saturated	G	...	NaN	NaN	NaN	0	2.22	NaN	NaN	NaN
2	16158	16158	16	Hummus, commercial	Hummus, commercial	1089	5400	Iron, Fe	Iron, Fe	MG	...	NaN	NaN	NaN	2	2.41	1.87	2.96	2.33
3	16158	16158	16	Hummus, commercial	Hummus, commercial	1051	100	Water	Water	G	...	NaN	NaN	NaN	2	58.70	56.10	65.60	57.60
4	16158	16158	16	Hummus, commercial	Hummus, commercial	1103	6200	Selenium, Se	Selenium, Se	UG	...	NaN	NaN	NaN	0	16.20	0.00	32.30	16.20

The ID numbers corresponded to a food group or nutrient name in the provided lookup tables, so we substituted them in.

Data Cleaning & Preparation (Cont.)

By replacing the IDs for both food categories and nutrient IDs with their respective names based on the provided lookup tables, it became far easier to read and locate items, especially when exported to Tableau.

```
# Replace food category IDs with names
food_id_to_desc = pd.Series(categories_df.description.values,index=categories_df.food_category_id).to_dict()
main_df['food_category_id'] = main_df['food_category_id'].map(food_id_to_desc)

# Replace nutrient IDs with names
nutrient_id_to_name = pd.Series(nutrient_df.name.values,index=nutrient_df.nutrient_id).to_dict()
main_df['Nutrient_id'] = main_df['Nutrient_id'].map(nutrient_id_to_name)
```

	FF_NDB	SR_NDB	food_category_id	FF Food description	SR Food description	Nutrient_id	rank	FF_Component	SR_Component	unit_name	...	SR Max	Std_Error	Num_Data_pts	SR Mean per 100g	FF Min	FF Max	FF Median
0	16158	16158	Legumes and Legume Products	Hummus, commercial	Hummus, commercial	Magnesium, Mg	5500	Magnesium, Mg	Magnesium, Mg	MG	...	NaN	NaN	2	71.10	56.60	82.00	70.40
1	16158	16158	Legumes and Legume Products	Hummus, commercial	Hummus, commercial	Fatty acids, total saturated	9700	Fatty acids, total saturated	Fatty acids, total saturated	G	...	NaN	NaN	0	2.22	NaN	NaN	NaN
2	16158	16158	Legumes and Legume Products	Hummus, commercial	Hummus, commercial	Iron, Fe	5400	Iron, Fe	Iron, Fe	MG	...	NaN	NaN	2	2.41	1.87	2.96	2.33
3	16158	16158	Legumes and Legume Products	Hummus, commercial	Hummus, commercial	Water	100	Water	Water	G	...	NaN	NaN	2	58.70	56.10	65.60	57.60
4	16158	16158	Legumes and Legume Products	Hummus, commercial	Hummus, commercial	Selenium, Se	6200	Selenium, Se	Selenium, Se	UG	...	NaN	NaN	0	16.20	0.00	32.30	16.20

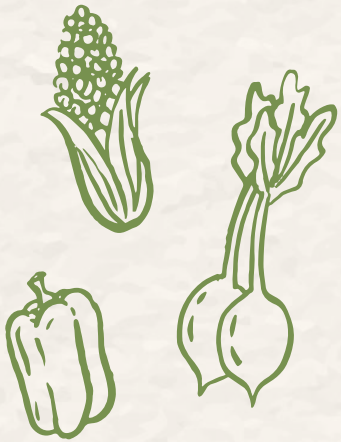
Data Cleaning: Filtering of Erroneous Values

	food_category_id	Nutrient_id	unit_name	Mean per 100g percentage change
2	Baked Products	Lactose	G	inf
3	Beef Products	Carbohydrate, by difference	G	inf
4	Beverages	Fatty acids, total saturated	G	inf
5	Beverages	SFA 14:0	G	inf
6	Beverages	SFA 16:0	G	inf
7	Beverages	SFA 18:0	G	inf
8	Beverages	Vitamin B-12	UG	inf
9	Dairy and Egg Products	Lycopene	UG	inf
10	Dairy and Egg Products	Tocotrienol, gamma	MG	inf
11	Fats and Oils	SFA 14:0	G	inf
12	Fruits and Fruit Juices	Lactose	G	inf
13	Nut and Seed Products	Tocopherol, gamma	MG	inf
14	Nut and Seed Products	Tocotrienol, alpha	MG	inf
15	Spices and Herbs	Galactose	G	inf
16	Spices and Herbs	Tocopherol, beta	MG	inf
17	Sausages and Luncheon Meats	Carbohydrate, by difference	G	2745.47
18	Sausages and Luncheon Meats	Vitamin C, total ascorbic acid	MG	2642.86
19	Dairy and Egg Products	Tocopherol, beta	MG	1940
20	Nut and Seed Products	Carotene, beta	UG	850
21	Legumes and Legume Products	Selenium, Se	UG	749.39

Removed due to infinite values caused by denominator of 0

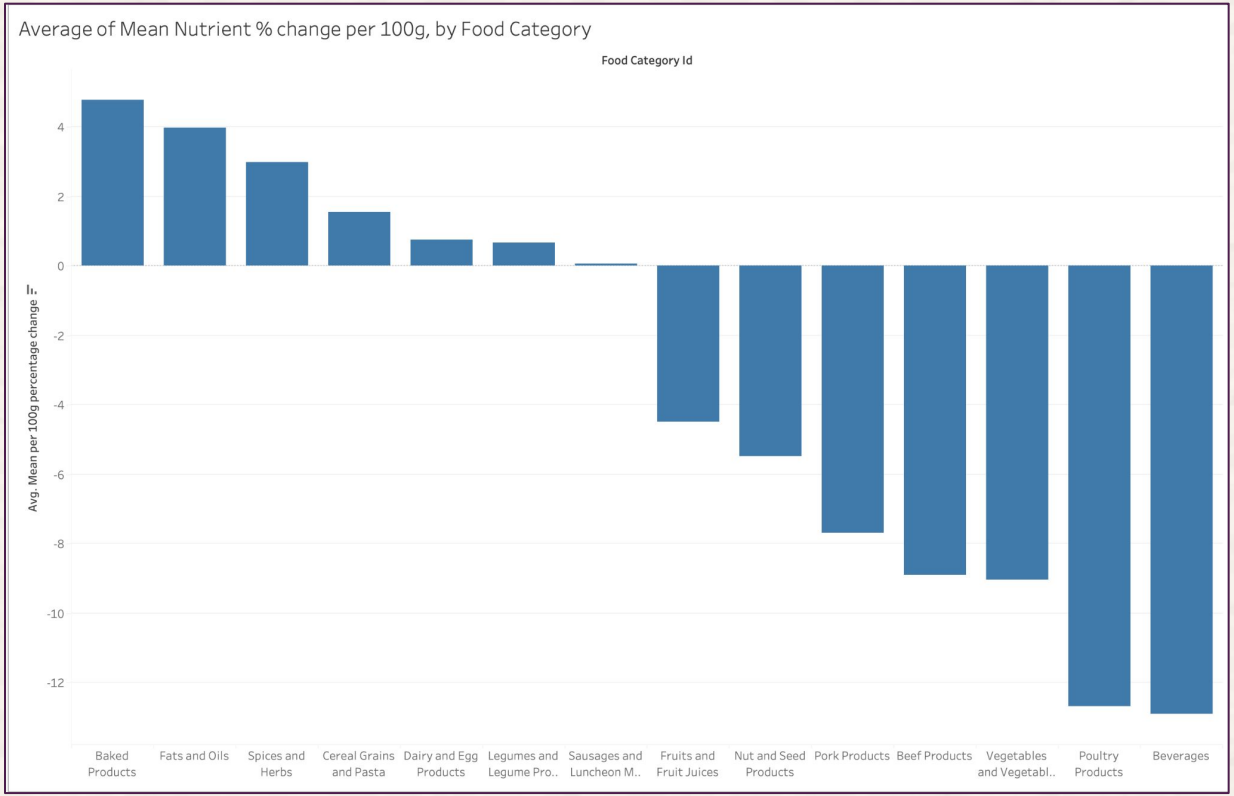
Removed due to outlier values well above 4 standard deviations of mean (above 99.99% of values)

03
Our Findings



Question 1

What is the average change in nutrients for all the different food groups?

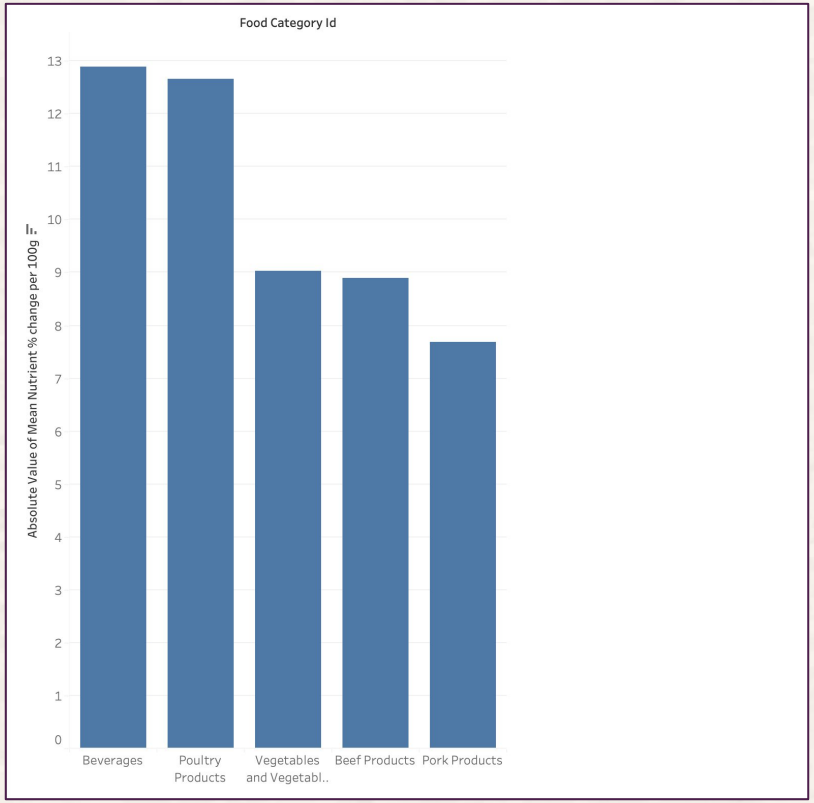


Question 1 Analysis

- We created a graphic showcasing the average change in nutrients for each of the food categories presented in our data set
    - Nutrients in fruits and fruit juices, nut and seed products, pork, beef, vegetables, poultry, and beverages have decreased over time
    - Nutrients in baked products, fats and oils, spices and herbs, cereal grains and pasta, dairy and egg products, legumes, sausage and lunch meats have increased over time
- We see a drastic comparison between the categories related with high protein levels versus the categories with high carbohydrates
- To further explore why, we created a graph highlighting the food categories with the biggest change, those categories that decreased significantly in average mean nutrient levels

Question 2

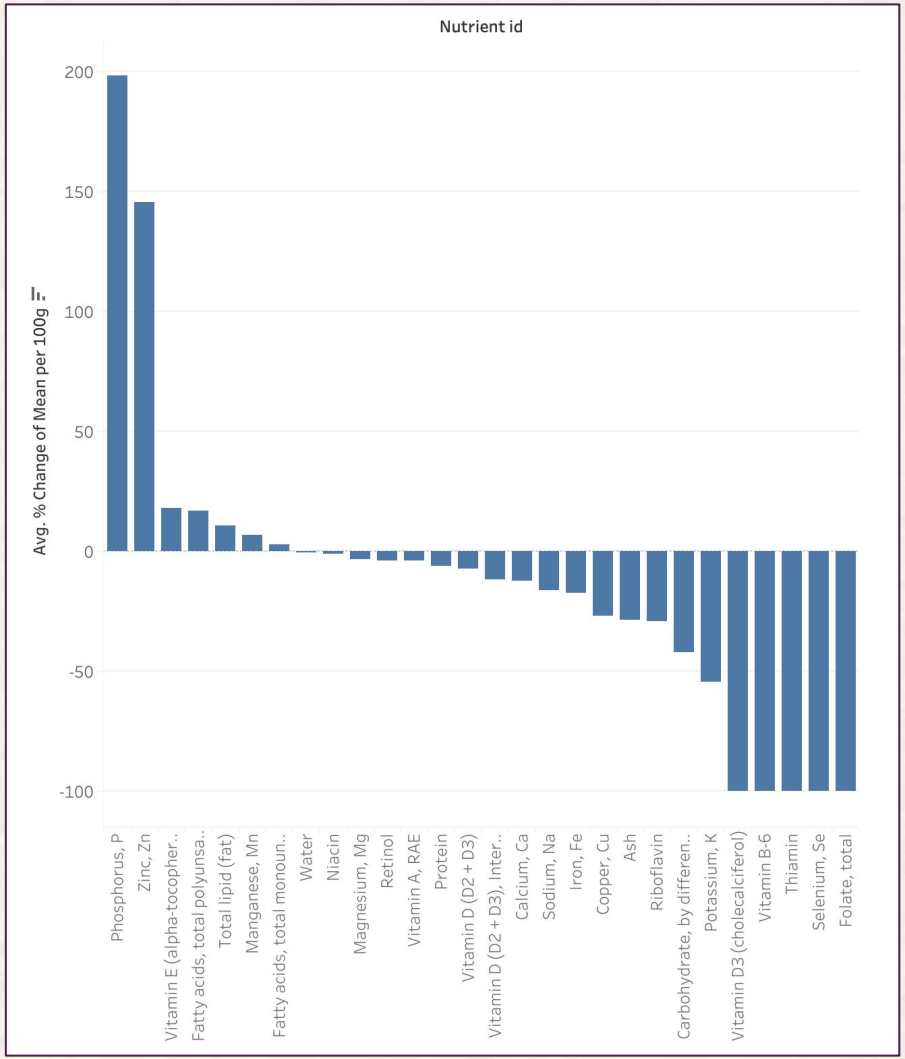
Which food groups have had the biggest change & why?



Beverages

Increase in Phosphorus and Zinc and we see a decrease in Vitamin D3, Vitamin B, Thiamin, Selenium, and Folate.

The potential reason for this increase is that the implications/benefits are interchangeable with all the nutrients that have shown a decrease in beverages.



Implications of Phosphorus and Zinc:

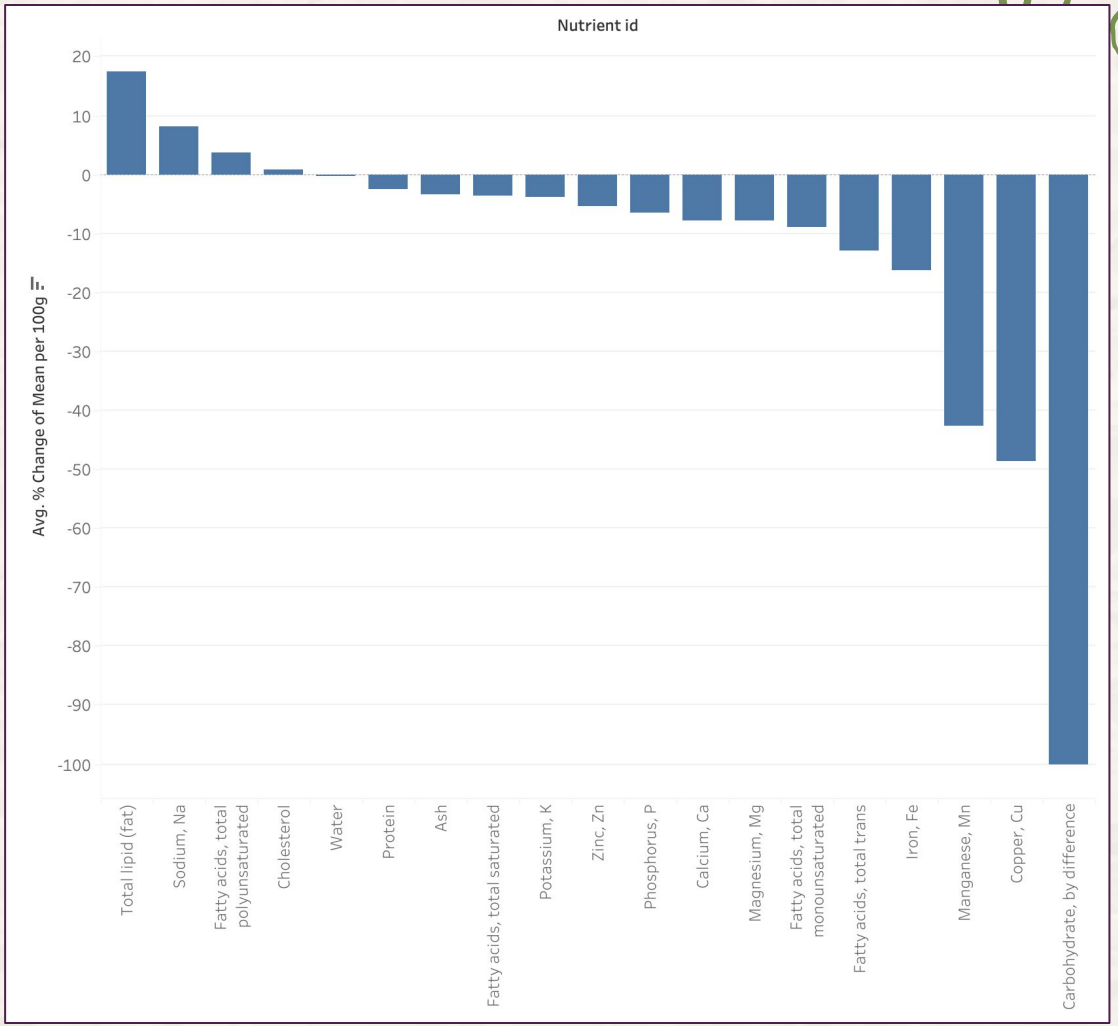
- Bone health, immune function, and metabolism.

Poultry Products

Largest nutrient decline:
- Carbohydrates

Role in Poultry:
- Primary energy source

Decline Impact:
- Effects the overall energy content, potentially influencing growth rates, poultry health, and egg production.



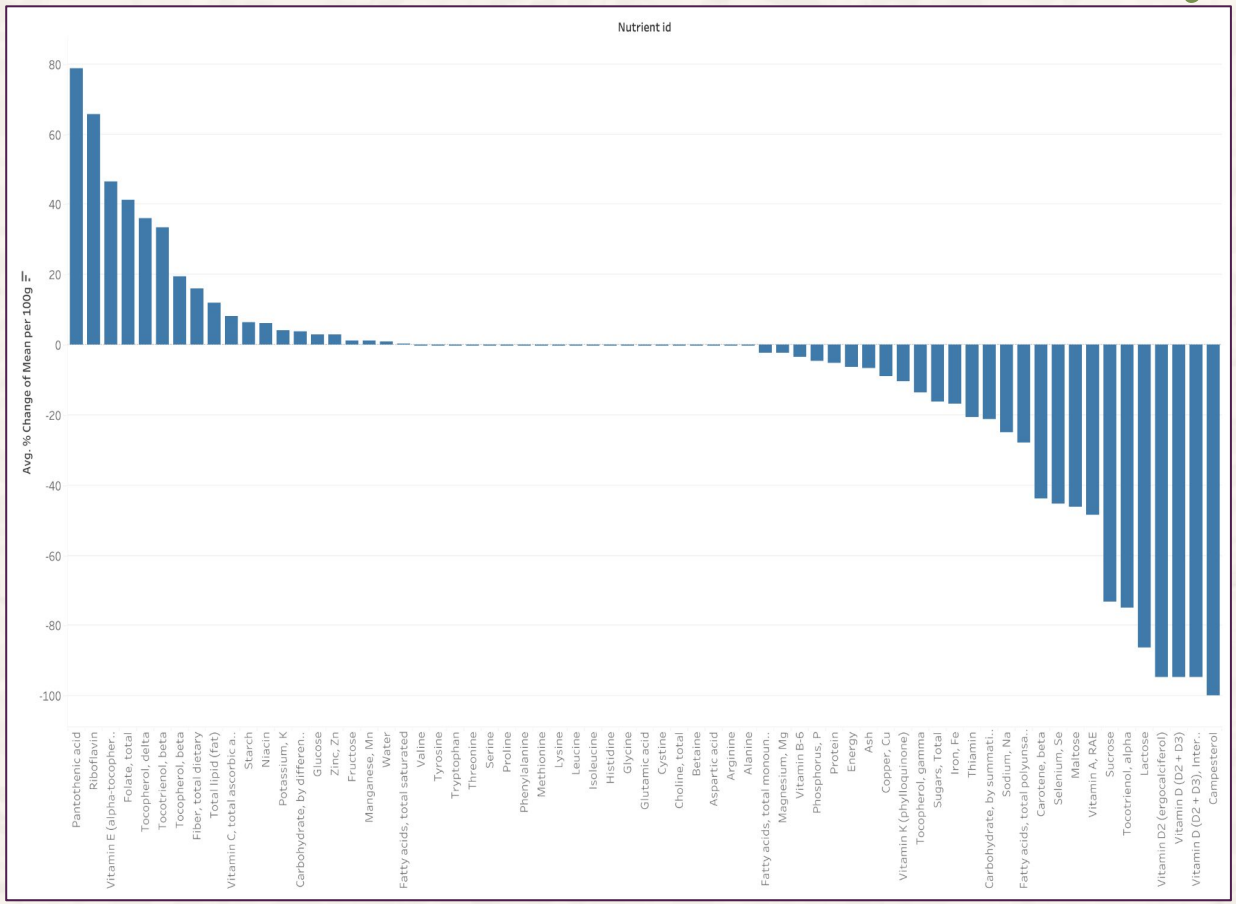
Vegetable & Vegetable Juices

Largest Increase:

- Pantothenic Acid (aka Vitamin B5)
- Contributes to the nutritional quality of the vegetables. Higher content → higher value for consumers

Largest Decrease:

- Campesterol
- A plant sterol that consumers can be sensitive to
- A decrease can be related to types of vegetables used or agricultural practices.



Beef Products

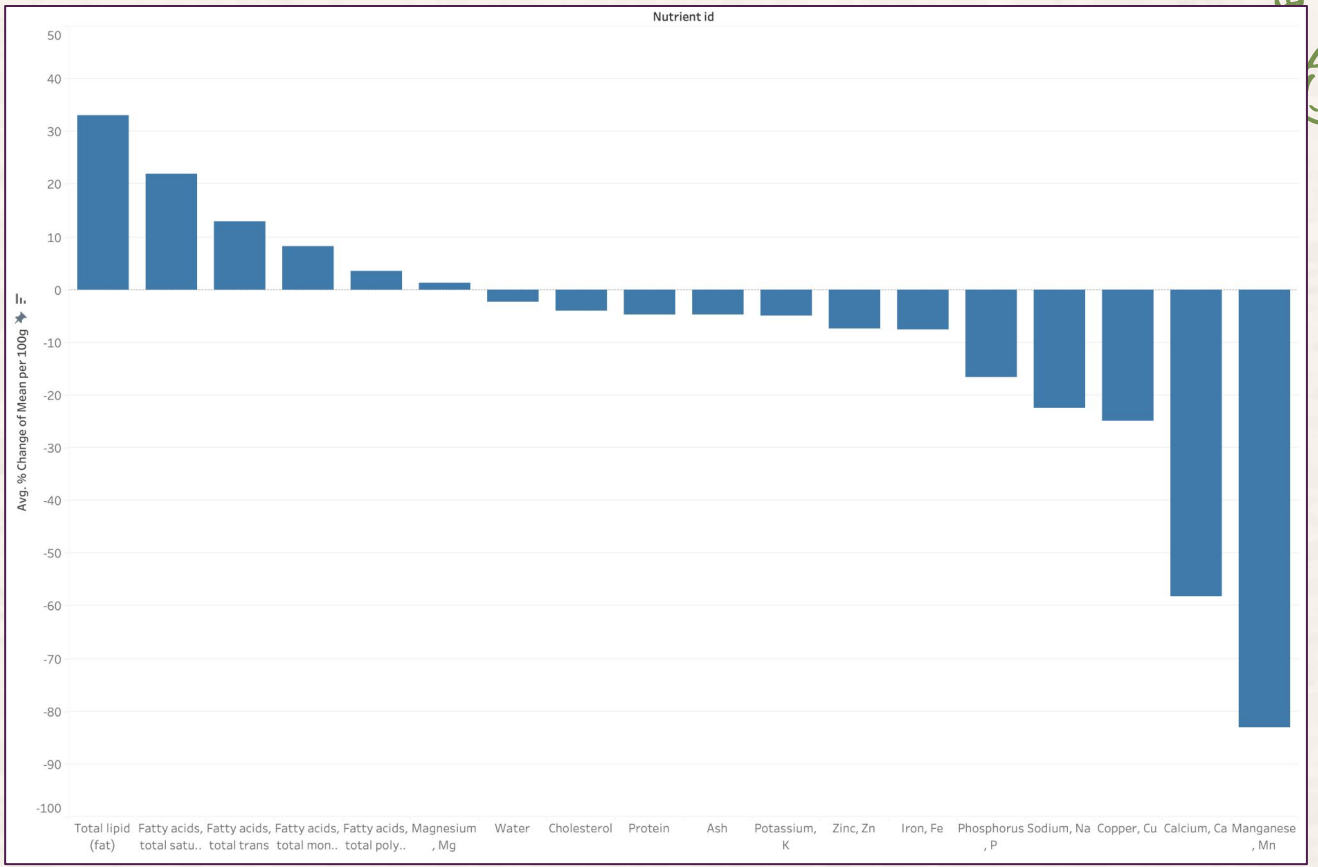
Significant decrease in Manganese and Calcium

Potential Impact of Manganese

- Decrease can be influenced by animal feed composition or soil quality

Potential Impact of Calcium

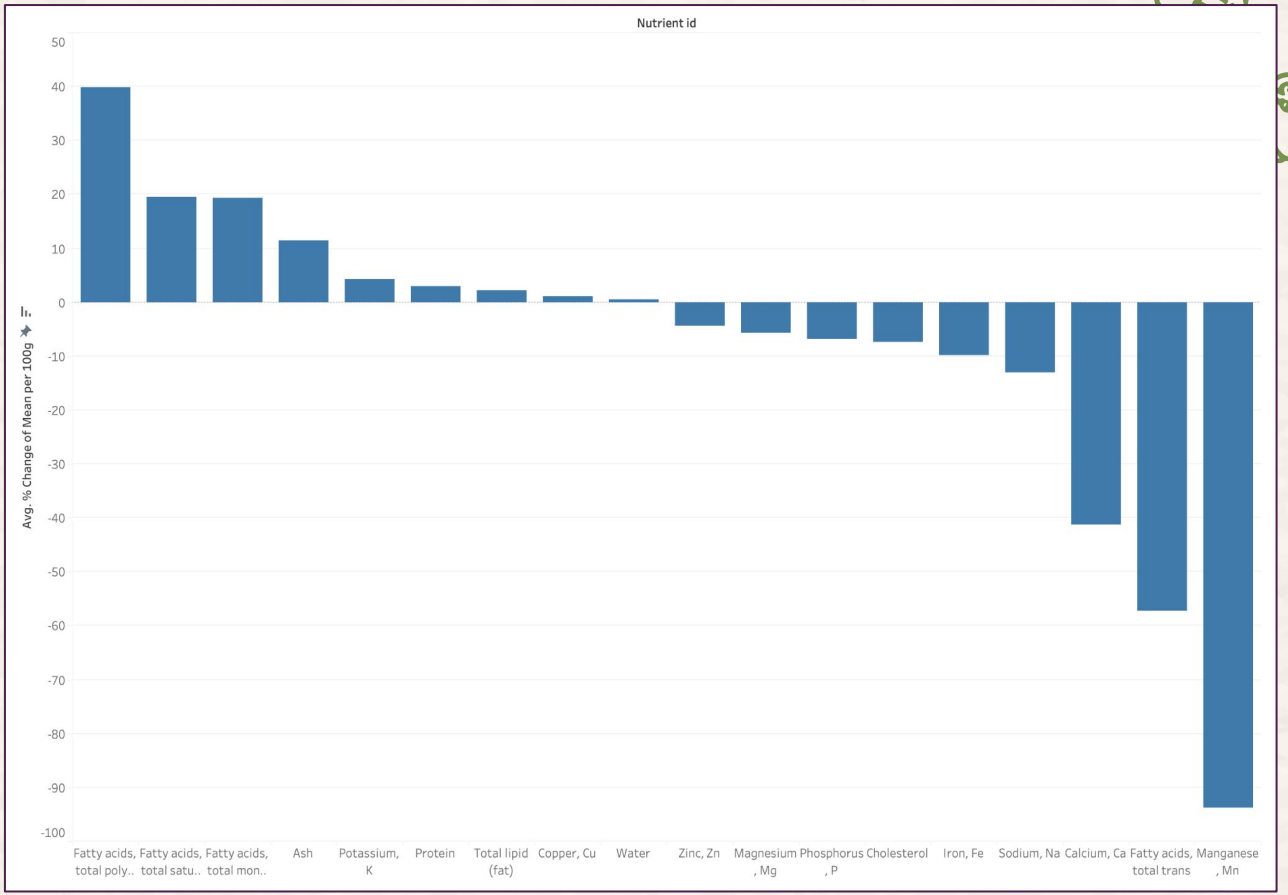
- Decrease can be influenced by animal diets or bone structure



Pork Products

We see a commonality in the largest nutrient decrease for both beef and pork categories:
Manganese

- Altercations in Feed Composition
- Soil Manganese Levels
- Processing Methods
- Mineral Supplementation
- Animal Genetics
- Market Trends and Consumer Preferences

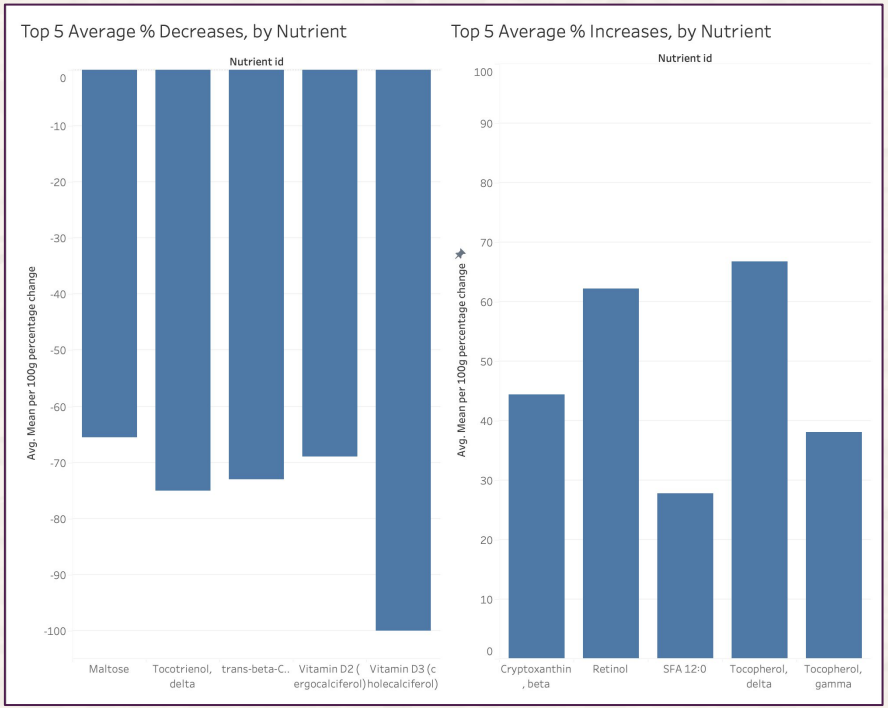


Question 2 Analysis

- Food groups with the biggest change:
  - Beverages
  - Poultry Products
  - Vegetables and Vegetable Juices
  - Beef Products
  - Pork Products
- The drastic change in these nutrients can come from a variety of non-scientific factors such as:
  - **Consumer Trends**: consumer preferences can shift contributing to nutrient fluctuations
  - **Regulatory Factors:** changes in regulations occur frequently and can impact the use of specific nutrients in certain product categories
  - **Health Trends**: consumers increasingly become more health-focused impacting nutrient composition

Question 3

What are the top 10 nutrients that have changed the most between the SR and FF Mean Values?



Question 3 Analysis

- Top 5 **increasing** nutrients
  - Cryptoxanthin* — found in **fruits** and **vegetables**
  - Retinol — found in **dairy products**
  - SFA 12-0 — found in **oils**
  - Tocopherol delta — found in **oils, nuts,** and **seeds**
  - Tocopherol gamma — found in **grains, oils, nuts,** and **seeds**
- Top 5 **decreasing** nutrients
  - Maltose — found in sugary and alcoholic **beverages**
  - Tocotrienol delta — found in **nuts** and **seeds**
  - Trans-beta-c (beta carotene) — found in **vegetables** like carrots
  - Vitamin D2 — found in **animal products**
  - Vitamin D3 — found in orange juice and other **beverages**

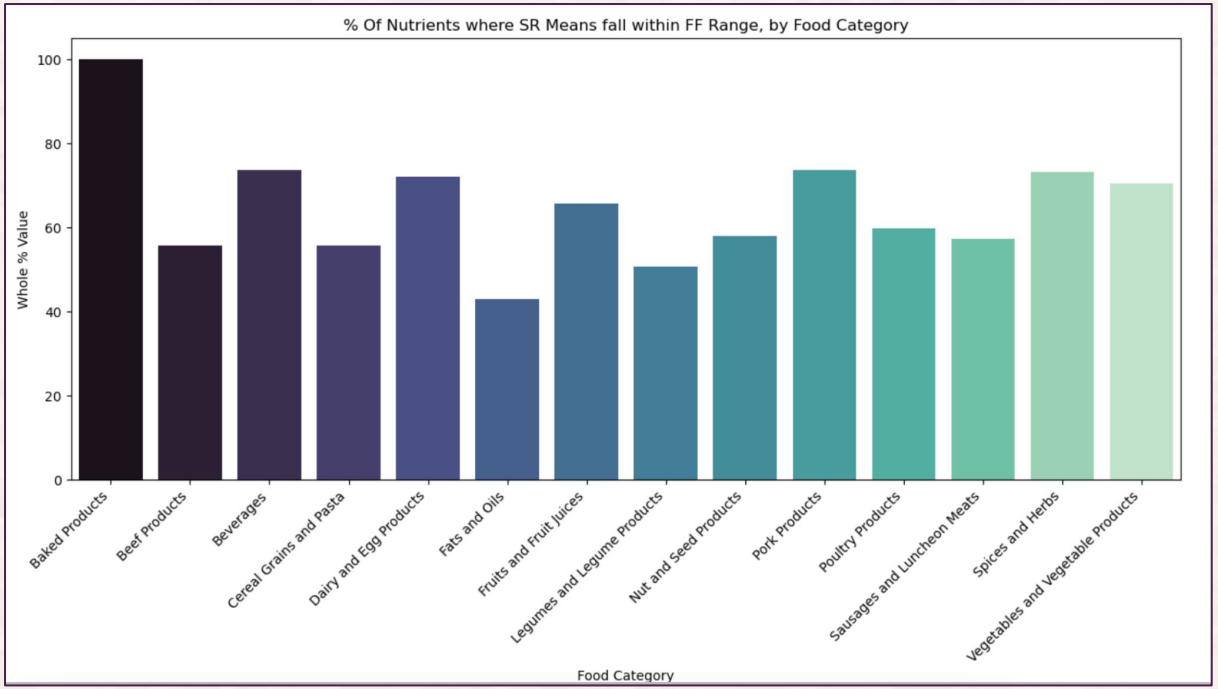
*Outlier for these purposes

Q3 Analysis

- Can connect our results to our original research question related to the food groups with the biggest change in nutrients overall:
  - Beverages, poultry, vegetables, nut and seed products
- The top 10 nutrients that increased and decreased (respectively) were ingredients found in these food groups
  - No processed food nutrients really changed! It was mostly protein items, carbohydrates, fats, and plant based foods.

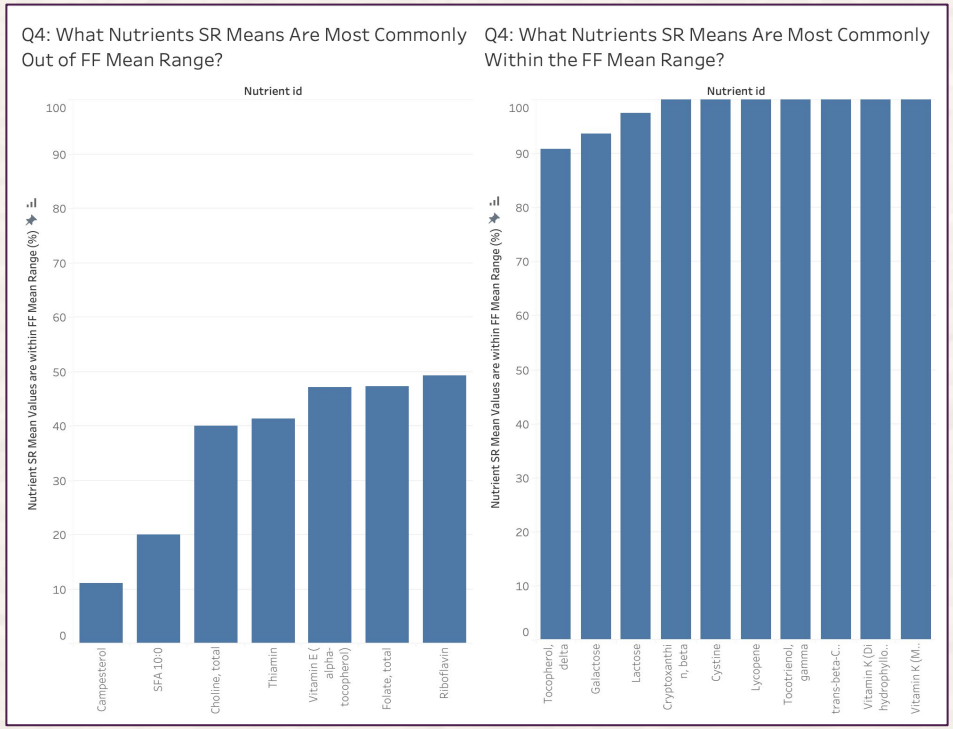
Question 4

Does the mean value for SR Legacy fall between the minimum and maximum values of the FF Data for nutrients/components in equivalent foods?



Question 5

Which nutrients/components fall out of range the most and which fall in range more frequently?



Analysis Question 5

A significant number of SR nutrient means **do** fall within the FF mean range. One possible reason is they are mainly fruits, vegetables and protein which include all very commonly tested nutrients

The Nutrients listed below have 100% or more of their SR values within the FF Mean range

- Cryptoxanthin, beta- Found in tangerines, persimmons and oranges.
- Cystine- Found in high protein foods
  - Including yogurts, chicken, pork
- Lycopene- carotenoid hydrocarbon found in red,pink and orange fruits and vegetables
  - Including tomatoes, apricots, melons, watermelons, and cranberries
- Tocotrienol,gamma- Found in Cereal and in vegetables
  - Including rice grains and palm oil
- This list also includes trans-beta-Carotene, Vitamin K (Dihydrophylloquinone), Vitamin K(Menaquinone-4)

Analysis Question 5 (Cont.)

A significant number of SR nutrient means **do not** fall within the FF mean range.

The 3 most notable are:

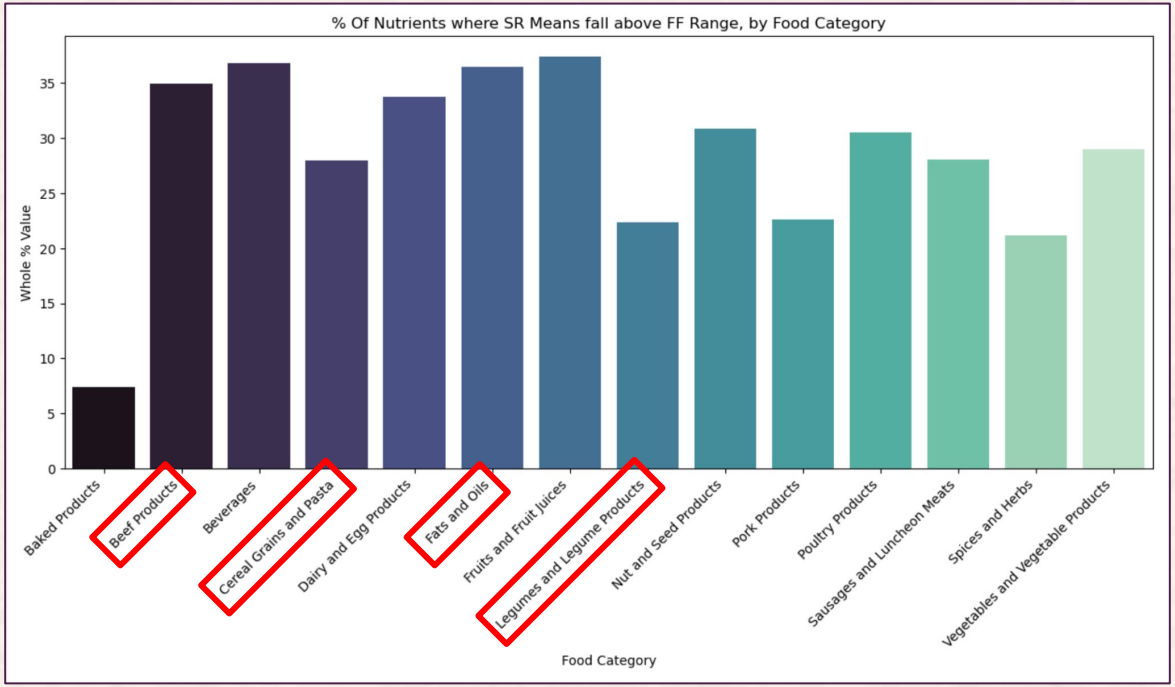
1. Riboflavin: Which is a supplement also known as Vitamin B2
    a. Found in eggs, milk and meats (specifically kidneys and liver)
    b. Possible causation: Higher exposure to UV light can diminish Riboflavin.
2. Folate: Important in red blood cell formation
    a. Found in dark green leafy vegetables, peas, nuts and beans
    b. Possible causation: Can destroy vegetables by 40% by cooking and in grains/cereals 70% can be destroyed by milling and baking
3. Vitamin E (Alpha Tocopherol)
    a. Found in sunflower seeds and almonds
    b. Possible causation: Can diminish overtime with exposure to oxygen and freezing

Follow up Question:

What % Nutrient SR mean values fall **above** the FF Range in various food groups?

Using the FF Max Instead of Mean:

- Indicate which past values *modern foods don't reach at all*

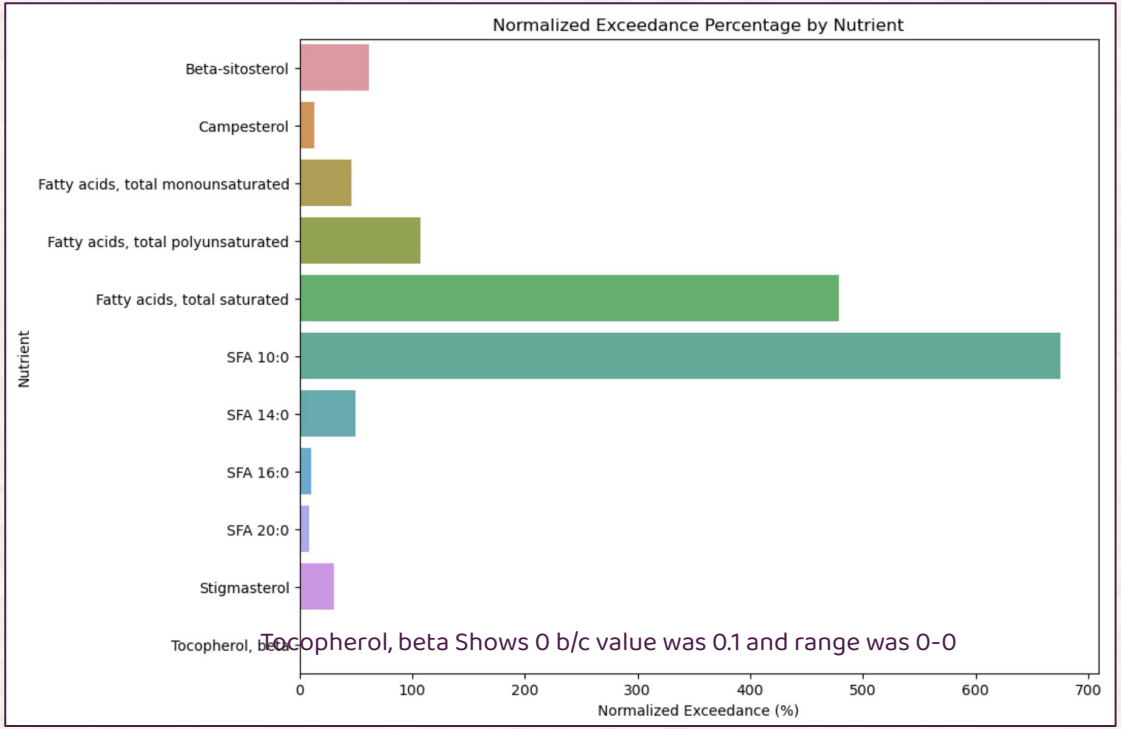


Nutrient Exceedance: Fats and Oils

SR values displayed more:

- SFA 10:0
- Fatty acids, saturated

This is good! These have decreased.



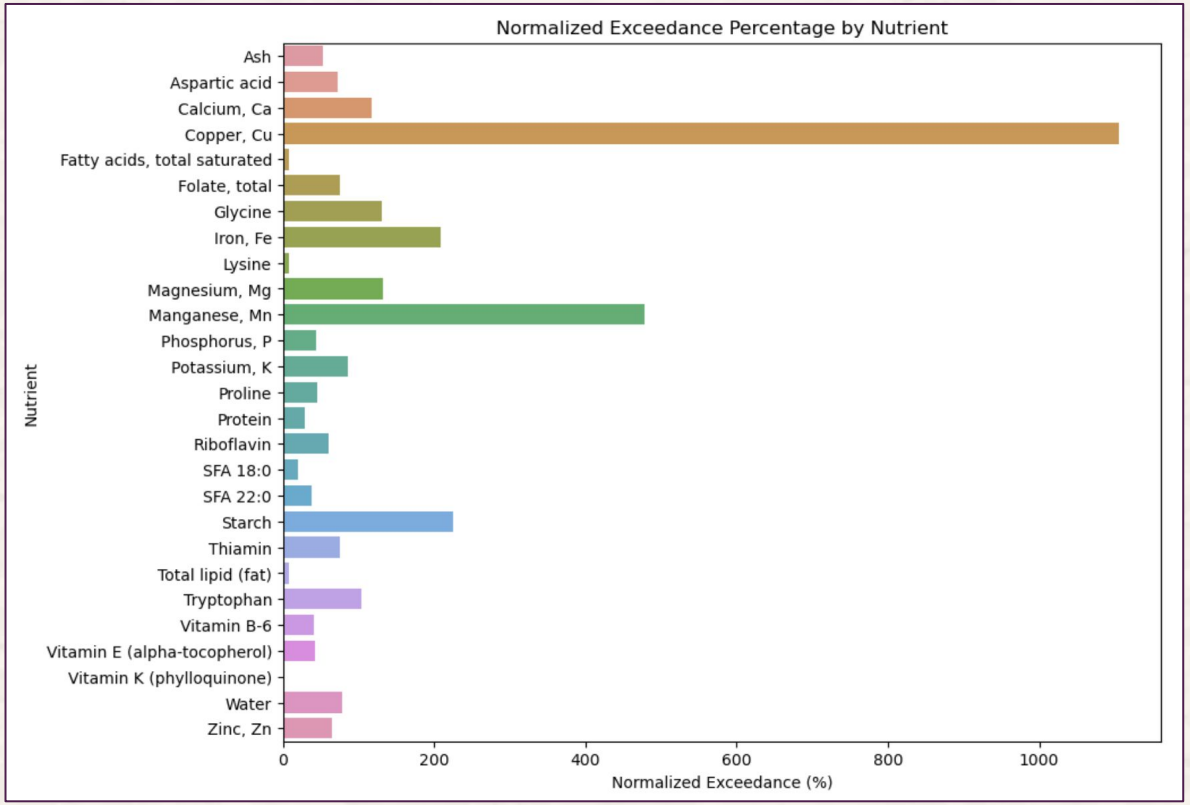
Tocopherol, beta Shows 0 b/c value was 0.1 and range was 0-0

Nutrient Exceedance: Legumes and Legume Products

SR values displayed more:

- Copper
- Manganese
- Iron

This is not good! These essential nutrient levels have decreased.

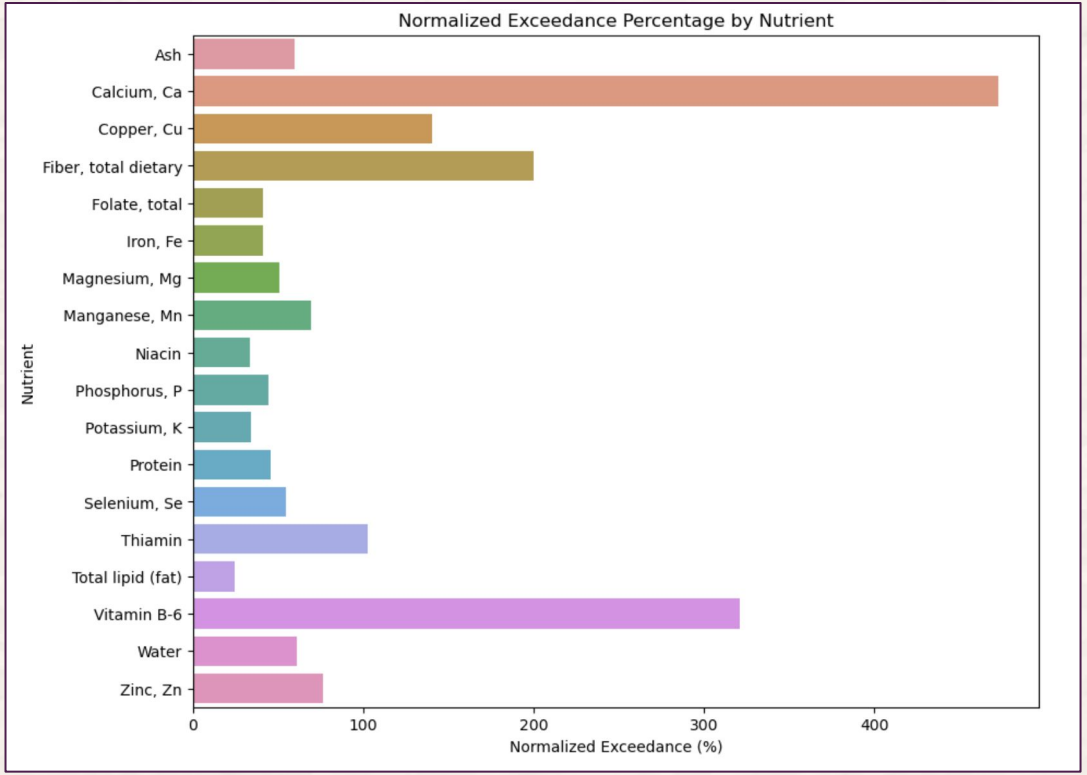


Nutrient Exceedance: Cereal Grains and Pasta

SR values displayed more:

- Calcium
- Vitamin B-6
- Dietary Fiber
- Copper

This is not good! These essential nutrient levels have decreased.

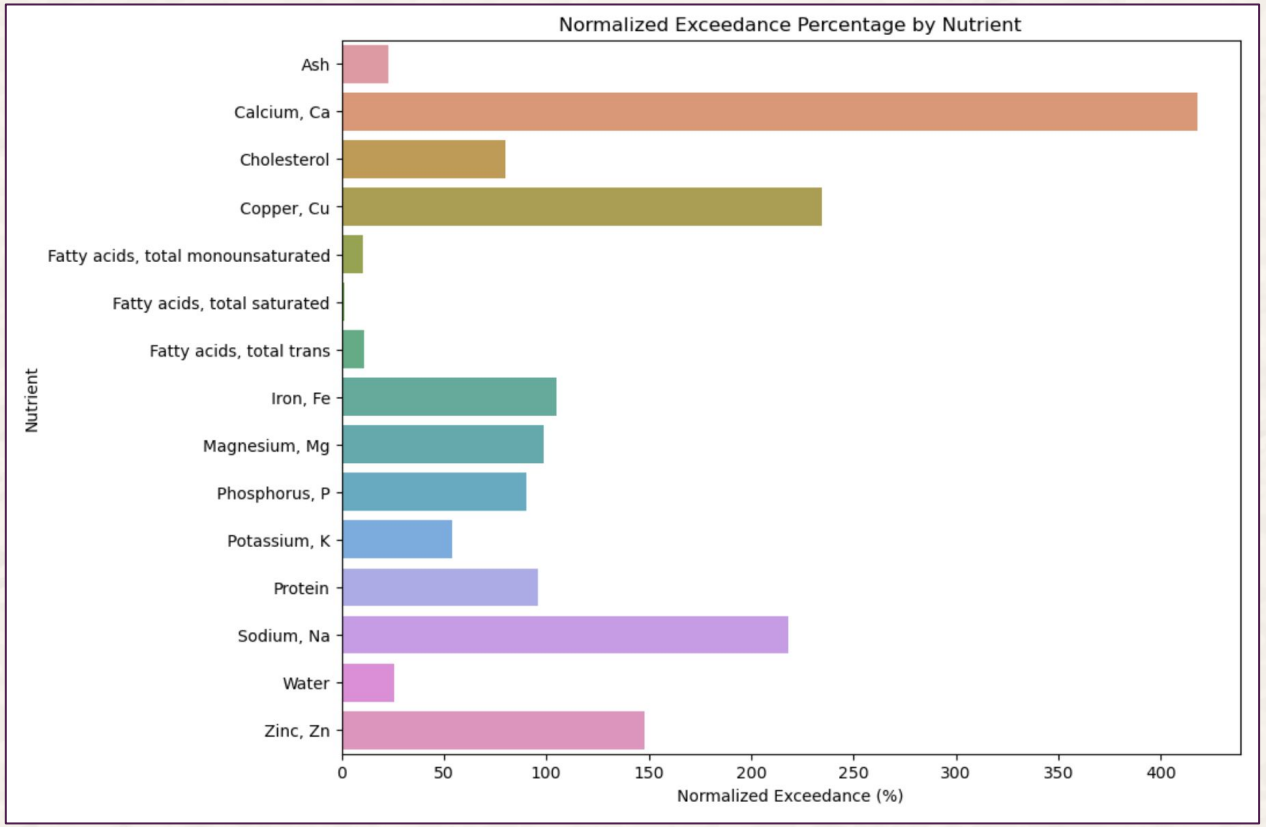


Nutrient Exceedance: Beef Products

SR values displayed more:

- Calcium
- Copper
- Sodium
- Zinc

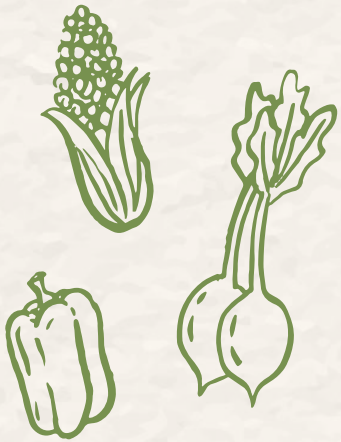
This is not good! These essential nutrient levels have decreased.



What Do We Find?

- Certain decreases in nutrient values are generally good!
  - Decrease in Saturated Fats/Fatty Acids
- Significant decreases in essential nutrients in categories like grains, legumes, and beef.
  - Decrease in vital nutritional value

04
Conclusion



Conclusion

- The nutrients of non-processed foods are more likely to change because of changes in health trends and agricultural practices
  - Processed foods generally stay the same because you know exactly what you get!
  - They are not supposed to be "healthy" and their lack of nutrients explains that.
- The nutrients that should be more in depth analyzed are beverages, poultry, vegetables, nut and seed products since they had the biggest changes in individual nutrients and overall
- Not all nutrient changes are good!
  - Many important nutrients are diminishing like vital vitamins and minerals.
- **We recommend that these decreases in essential nutrients are further looked into to find the cause, whether due to agricultural practices, lack of regulation, or deliberate ingredient changes.**

Thank You!

Questions?

(Slide deck from SlidesGo)

Resources

Nickle, M. *Info Challenge FoodDataCentral-QuickOverview* [PowerPoint Slides]. https://docs.google.com/presentation/d/1DzH_dC6MJ-ZCRVjf0Ck1zusb6hq0XWWI/edit?usp=sharing&ouid=114016100017749737350&rtpof=true&sd=true

Office of Dietary Supplements - Riboflavin. (n.d.). Ods.od.nih.gov. Retrieved March 1, 2024, from https://ods.od.nih.gov/factsheets/Riboflavin-HealthProfessional/#:~:text=Riboflavin%20is%20yellow%20and%20naturally

All data from fdc.nal.usda.gov