

Submitted by: Angelo Luis C. Cu

Dataset from UNICEF Child Nutrition Data: <https://data.unicef.org/topic/nutrition/child-nutrition/>

Data is easily filtered through the web service: <https://sdmx.data.unicef.org/webservice/data.html>

CSV can be downloaded thru the link:

<https://sdmx.data.unicef.org/ws/public/sdmxapi/rest/data/UNICEF.NUTRITION.1.0/BRN+IDN+KHM+LAO+MMR+MYS+PHL+SGP+THA+VNM....?format=csv-ts&labels=both&saveAs=Nutrition>

```
import pandas as pd
import numpy as np
import matplotlib as plt
```

```
nutrition = pd.read_csv('/content/data/Nutrition.csv')
nutrition
```

	REF_AREA	Geographic area	INDICATOR	Indicator	SEX	Sex	AGE	Current age	HEALTH_QUINTILE	Health Quintile	...	2015	2016	2017	2018	2019	2020	2020-01-01	2020-07-02	2021	2022	
0	BRN	Brunei Darussalam	NT_ANT_BAZ_NE2	BMI-for-age <2 SD	F	Female	Y0T4	Under 5 years old	_T	Total	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
1	BRN	Brunei Darussalam	NT_ANT_BAZ_NE3	BMI-for-age <3 SD	F	Female	Y0T4	Under 5 years old	_T	Total	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
2	BRN	Brunei Darussalam	NT_ANT_BAZ_PO2	BMI-for-age »+2 SD	F	Female	M0T5	Under 6 months old	_T	Total	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
3	BRN	Brunei Darussalam	NT_ANT_BAZ_PO2	BMI-for-age »+2 SD	F	Female	M12T23	12 to 23 months old	_T	Total	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
4	BRN	Brunei Darussalam	NT_ANT_BAZ_PO2	BMI-for-age »+2 SD	F	Female	M24T35	24 to 35 months old	_T	Total	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
...	
37972	VNM	Viet Nam	NT_CF_OTHER_FV	Food group- Other fruits and vegetables (6-23m...	M	Male	M6T23	6 to 23 months old	_T	Total	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
37973	VNM	Viet Nam	NT_CF_VITA	Food group- Vitamin-A rich fruits and vegetabl...	M	Male	M6T23	6 to 23 months old	_T	Total	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
37974	VNM	Viet Nam	NT_CF_ZEROFV	Zero vegetable or fruit consumption (6-23months)	M	Male	M6T23	6 to 23 months old	_T	Total	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
37975	VNM	Viet Nam	NT_SANT_10_19_BAZ_NE2_MOD	Prevalence of thinness among children aged 10-...	M	Male	Y10T19	10 to 19 years old	_T	Total	...	14.5	14.2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
37976	VNM	Viet Nam	NT_SANT_10_19_BAZ_PO1_MOD	Prevalence of overweight among children aged 1...	M	Male	Y10T19	10 to 19 years old	_T	Total	...	9.0	10.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	



37977 rows × 83 columns

```
nutrition.info()
```

```
27 1975      60 non-null    float64
28 1976      60 non-null    float64
29 1977      60 non-null    float64
30 1978      60 non-null    float64
31 1979      60 non-null    float64
32 1980      60 non-null    float64
33 1981      60 non-null    float64
34 1982      60 non-null    float64
35 1983      60 non-null    float64
36 1984      60 non-null    float64
37 1985      60 non-null    float64
```

```
70 2012 4595 non-null float64
71 2013 4095 non-null float64
72 2014 6509 non-null float64
73 2015 6802 non-null float64
74 2016 11700 non-null float64
75 2017 6644 non-null float64
76 2018 384 non-null float64
77 2019 6169 non-null float64
78 2020 65 non-null float64
79 2020-01-01 16 non-null float64
80 2020-07-02 2 non-null float64
81 2021 3394 non-null float64
82 2022 49 non-null float64
dtypes: float64(58), object(25)
memory usage: 24.0+ MB
```

```
# getting only data from 2017-2022 (5 years latest available)
latest_nutrition = pd.concat([
    nutrition.iloc[:, :26],
    nutrition.iloc[:, 75:]
], axis = 1)
latest_nutrition
```

	REF_AREA	Geographic area	INDICATOR	Indicator	SEX	Sex	AGE	Current age	WEALTH_QUINTILE	Wealth Quintile	...	SERIES_FOOTNOTE	CUSTODIAN	2017	2018	2019	2020	2020-01-01	2020-07-02	2021	2022	
	0	BRN	Brunei Darussalam	NT_ANT_BAZ_NE2	BMI-for-age <2 SD	F	Female	Y0T4	Under 5 years old	_T	Total	... UNICEF/WHO/World Bank Joint Malnutrition Estim...	UNICEF Division of Data Analysis Planning an...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
	1	BRN	Brunei Darussalam	NT_ANT_BAZ_NE3	BMI-for-age <3 SD	F	Female	Y0T4	Under 5 years old	_T	Total	... UNICEF/WHO/World Bank Joint Malnutrition Estim...	UNICEF Division of Data Analysis Planning an...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
	2	BRN	Brunei Darussalam	NT_ANT_BAZ_PO2	BMI-for-age +2 SD	F	Female	M0T5	Under 6 months old	_T	Total	... UNICEF/WHO/World Bank Joint Malnutrition Estim...	UNICEF Division of Data Analysis Planning an...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
	3	BRN	Brunei Darussalam	NT_ANT_BAZ_PO2	BMI-for-age +2 SD	F	Female	M12T23	12 to 23 months old	_T	Total	... UNICEF/WHO/World Bank Joint Malnutrition Estim...	UNICEF Division of Data Analysis Planning an...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
	4	BRN	Brunei Darussalam	NT_ANT_BAZ_PO2	BMI-for-age +2 SD	F	Female	M24T35	24 to 35 months old	_T	Total	... UNICEF/WHO/World Bank Joint Malnutrition Estim...	UNICEF Division of Data Analysis Planning an...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
	
	37972	VNM	Viet Nam	NT_OF_OTHER_FV	Food group- Other fruits and vegetables (6-23m...	M	Male	M6T23	6 to 23 months old	_T	Total	... UNICEF Global Infant and Young Child Feeding D...	UNICEF Division of Data Analysis Planning an...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
	37973	VNM	Viet Nam	NT_CF_VITA	Food group- Vitamin-A rich fruits and vegetabl...	M	Male	M6T23	6 to 23 months old	_T	Total	... UNICEF Global Infant and Young Child Feeding D...	UNICEF Division of Data Analysis Planning an...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
	37974	VNM	Viet Nam	NT_CF_ZEROFV	Zero vegetable or fruit consumption (6-23months)	M	Male	M6T23	6 to 23 months old	_T	Total	... UNICEF Global Infant and Young Child Feeding D...	UNICEF Division of Data Analysis Planning an...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
	37975	VNM	Viet Nam	NT_SANT_10_19_BAZ_NE2_MOD	Prevalence of thinness among children aged 10-...	M	Male	Y10T19	10 to 19 years old	_T	Total	... NCD-RisC. Worldwide trends in body-mass index ...	NCD Risk Factor Collaboration	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
	37976	VNM	Viet Nam	NT_SANT_10_19_BAZ_PO1_MOD	Prevalence of overweight among children aged 1...	M	Male	Y10T19	10 to 19 years old	_T	Total	... NCD-RisC. Worldwide trends in body-mass index ...	NCD Risk Factor Collaboration	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

37977 rows x 34 columns

```
latest_nutrition.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 37977 entries, 0 to 37976
Data columns (total 34 columns):
#   Column                Non-Null Count  Dtype
---  -
0   REF_AREA               37977 non-null  object
1   Geographic area        37977 non-null  object
2   INDICATOR              37977 non-null  object
3   Indicator              37977 non-null  object
4   SEX                   37977 non-null  object
5   Sex                   37977 non-null  object
6   AGE                   37977 non-null  object
7   Current age           37977 non-null  object
8   WEALTH_QUINTILE        37977 non-null  object
9   Wealth Quintile       37977 non-null  object
10  RESIDENCE              37977 non-null  object
11  Residence              37977 non-null  object
12  MATERNAL_EDU_LVL       37977 non-null  object
13  Mother's Education Level 37977 non-null  object
14  HEAD_OF_HOUSE          37977 non-null  object
15  Head of House          37977 non-null  object
16  REPORTING_LVL          32878 non-null  object
17  Reporting Level        32878 non-null  object
18  INDICATOR_METADATA     37957 non-null  object
19  UNIT_MULTIPLIER        37957 non-null  float64
20  Unit multiplier        37957 non-null  object
21  UNIT_MEASURE            37977 non-null  object
22  Unit of measure        37977 non-null  object
```

```
23 SOURCE_LINK          37957 non-null object
24 SERIES_FOOTNOTE      37957 non-null object
25 CUSTODIAN            37957 non-null object
26 2017                  6644 non-null float64
27 2018                  384 non-null float64
28 2019                  6169 non-null float64
29 2020                   65 non-null float64
30 2020-01-01            16 non-null float64
31 2020-07-02            2 non-null float64
32 2021                  3394 non-null float64
33 2022                   49 non-null float64
dtypes: float64(9), object(25)
memory usage: 9.9+ MB
```

```
# removes irrelevant columns
latest_nutrition = pd.concat([ # and keeps only the:
    latest_nutrition.iloc[:, 1], # country name
    latest_nutrition.iloc[:, 3], # indicator
    latest_nutrition.iloc[:, 5], # sex
    latest_nutrition.iloc[:, 7], # age
    latest_nutrition.iloc[:, 11], # residence (rural, urban, total)
    latest_nutrition.iloc[:, 19:22], # units
    latest_nutrition.iloc[:, 26:30], # 2017-2020, removes specific date
    latest_nutrition.iloc[:, 32:]
], axis = 1)
latest_nutrition
```

	Geographic area		Indicator	Sex	Current age	Residence	UNIT_MULTIPLIER	Unit multiplier	UNIT_MEASURE	2017	2018	2019	2020	2021	2022	
0	Brunei Darussalam		BMI-for-age <2 SD	Female	Under 5 years old	Total	0.0	Units	PCNT	NaN	NaN	NaN	NaN	NaN	NaN	
1	Brunei Darussalam		BMI-for-age <3 SD	Female	Under 5 years old	Total	0.0	Units	PCNT	NaN	NaN	NaN	NaN	NaN	NaN	
2	Brunei Darussalam		BMI-for-age >+2 SD	Female	Under 6 months old	Total	0.0	Units	PCNT	NaN	NaN	NaN	NaN	NaN	NaN	
3	Brunei Darussalam		BMI-for-age >+2 SD	Female	12 to 23 months old	Total	0.0	Units	PCNT	NaN	NaN	NaN	NaN	NaN	NaN	
4	Brunei Darussalam		BMI-for-age >+2 SD	Female	24 to 35 months old	Total	0.0	Units	PCNT	NaN	NaN	NaN	NaN	NaN	NaN	
...	
37972	Viet Nam	Food group- Other fruits and vegetables (6-23m...		Male	6 to 23 months old	Total	0.0	Units	PCNT	NaN	NaN	NaN	NaN	NaN	NaN	
37973	Viet Nam	Food group- Vitamin-A rich fruits and vegetabl...		Male	6 to 23 months old	Total	0.0	Units	PCNT	NaN	NaN	NaN	NaN	NaN	NaN	
37974	Viet Nam	Zero vegetable or fruit consumption (6-23months)		Male	6 to 23 months old	Total	0.0	Units	PCNT	NaN	NaN	NaN	NaN	NaN	NaN	
37975	Viet Nam	Prevalence of thinness among children aged 10-...		Male	10 to 19 years old	Total	0.0	Units	PCNT	NaN	NaN	NaN	NaN	NaN	NaN	
37976	Viet Nam	Prevalence of overweight among children aged 1...		Male	10 to 19 years old	Total	0.0	Units	PCNT	NaN	NaN	NaN	NaN	NaN	NaN	

37977 rows x 14 columns

Next steps: [View recommended plots](#)

```
latest_nutrition.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 37977 entries, 0 to 37976
Data columns (total 14 columns):
 #   Column          Non-Null Count  Dtype
---  ---
 0   Geographic area  37977 non-null  object
 1   Indicator        37977 non-null  object
 2   Sex              37977 non-null  object
 3   Current age     37977 non-null  object
 4   Residence       37977 non-null  object
 5   UNIT_MULTIPLIER 37957 non-null  float64
 6   Unit multiplier 37957 non-null  object
 7   UNIT_MEASURE    37977 non-null  object
 8   2017            6644 non-null   float64
 9   2018            384 non-null    float64
10  2019            6169 non-null   float64
11  2020            65 non-null     float64
12  2021            3394 non-null   float64
13  2022            49 non-null     float64
dtypes: float64(7), object(7)
memory usage: 4.1+ MB
```

```
# checks for duplicated values
print(latest_nutrition.duplicated().sum())
# around 8,000 duplicate values
```

8129

```
latest_nutrition.drop_duplicates(inplace = True) # removes duplicates
latest_nutrition.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 29848 entries, 0 to 37976
Data columns (total 14 columns):
 #   Column          Non-Null Count  Dtype
---  ---
 0   Geographic area  29848 non-null  object
 1   Indicator        29848 non-null  object
 2   Sex              29848 non-null  object
 3   Current age     29848 non-null  object
 4   Residence       29848 non-null  object
 5   UNIT_MULTIPLIER 29828 non-null   float64
```

```
6 Unit multiplier 29828 non-null object
7 UNIT_MEASURE 29848 non-null object
8 2017 5928 non-null float64
9 2018 364 non-null float64
10 2019 5534 non-null float64
11 2020 65 non-null float64
12 2021 2921 non-null float64
13 2022 49 non-null float64
dtypes: float64(7), object(7)
memory usage: 3.4+ MB
```

```
# as there are rows which have no data from 2017-2022
# we delete them
cleaned_nutrition = latest_nutrition.dropna(subset = [
    '2017', '2018', '2019', '2020', '2021', '2022'
], how='all')
cleaned_nutrition
```

Geographic area		Indicator	Sex	Current age	Residence	UNIT_MULTIPLIER	Unit multiplier	UNIT_MEASURE	2017	2018	2019	2020	2021	2022
336	Indonesia	Height-for-age <-2 SD (stunting)	Female	Under 5 years old	Total	0.0	Units	PCNT	NaN	29.7	NaN	NaN	NaN	NaN
337	Indonesia	Height-for-age <-2 SD and ≥-3 SD (Moderate St...	Female	Under 5 years old	Total	0.0	Units	PCNT	NaN	18.9	NaN	NaN	NaN	NaN
355	Indonesia	Height-for-age <-3 SD (Severe Stunting)	Female	Under 5 years old	Total	0.0	Units	PCNT	NaN	10.8	NaN	NaN	NaN	NaN
481	Indonesia	Weight-for-age <-2 SD (Underweight)	Female	Under 5 years old	Total	0.0	Units	PCNT	NaN	16.6	NaN	NaN	NaN	NaN
499	Indonesia	Weight-for-age <-3 SD (Severe Underweight)	Female	Under 5 years old	Total	0.0	Units	PCNT	NaN	3.3	NaN	NaN	NaN	NaN
...
36974	Thailand	Minimum diet diversity (children aged 6 to 23 ...	Male	6 to 23 months old	Total	0.0	Units	PCNT	NaN	NaN	69.4	NaN	NaN	NaN
36975	Thailand	Minimum meal frequency (children aged 6 to 23 ...	Male	6 to 23 months old	Total	0.0	Units	PCNT	NaN	NaN	87.6	NaN	NaN	NaN
36976	Thailand	Food group- Other fruits and vegetables (6-23m...	Male	6 to 23 months old	Total	0.0	Units	PCNT	NaN	NaN	53.7	NaN	NaN	NaN
36977	Thailand	Food group- Vitamin-A rich fruits and vegetabl...	Male	6 to 23 months old	Total	0.0	Units	PCNT	NaN	NaN	79.0	NaN	NaN	NaN
36978	Thailand	Zero vegetable or fruit consumption (6-23months)	Male	6 to 23 months old	Total	0.0	Units	PCNT	NaN	NaN	12.7	NaN	NaN	NaN

14579 rows × 14 columns

Next steps: [View recommended plots](#)

```
cleaned_nutrition['UNIT_MULTIPLIER'].unique()

array([ 0.,  3., nan])
```



```
cleaned_nutrition.query('UNIT_MULTIPLIER == 3')
```

Geographic area		Indicator	Sex	Current age	Residence	UNIT_MULTIPLIER	Unit multiplier	UNIT_MEASURE	2017	2018	2019	2020	2021	2022
8151	Singapore	Height-for-age <-2 SD (Stunting) Modeled Estim...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	7.2	7.4	7.6	7.6	7.4	7.1
8495	Singapore	Weight-for-height ≥+2 SD (Overweight) Modeled ...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	7.9	8.3	8.8	9.2	9.2	9.1
9001	Thailand	Height-for-age <-2 SD (Stunting) Modeled Estim...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	470.4	452.9	435.7	419.1	403.2	388.4
9718	Thailand	Weight-for-height ≥+2 SD (Overweight) Modeled ...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	349.1	335.5	321.3	307.9	295.4	283.2
10203	Malaysia	Height-for-age <-2 SD (Stunting) Modeled Estim...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	527.3	540.8	548.9	553.9	557.6	560.9
10345	Malaysia	Weight-for-height ≥+2 SD (Overweight) Modeled ...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	156.8	155.6	153.4	150.6	147.8	145.1
10843	Philippines	Height-for-age <-2 SD (Stunting) Modeled Estim...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	3575.7	3529.8	3541.7	3515.5	3459.8	3456.1
11560	Philippines	Weight-for-height ≥+2 SD (Overweight) Modeled ...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	484.7	503.5	513.7	529.0	546.4	555.5
11972	Brunei Darussalam	Height-for-age <-2 SD (Stunting) Modeled Estim...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	4.7	4.4	4.1	3.9	3.6	3.4
11980	Brunei Darussalam	Weight-for-height ≥+2 SD (Overweight) Modeled ...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	3.1	3.1	3.0	3.0	2.9	2.9
12258	Indonesia	Height-for-age <-2 SD (Stunting) Modeled Estim...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	7688.0	7517.3	7363.9	7212.9	7048.7	6896.6
12604	Indonesia	Weight-for-height ≥+2 SD (Overweight) Modeled ...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	2761.6	2749.1	2699.8	2617.0	2503.7	2357.6
13479	Lao People's Democratic Republic	Height-for-age <-2 SD (Stunting) Modeled Estim...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	261.5	252.8	244.0	235.2	226.4	218.4
14196	Lao People's Democratic Republic	Weight-for-height ≥+2 SD (Overweight) Modeled ...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	22.4	24.1	25.8	27.6	29.6	31.6
15103	Viet Nam	Height-for-age <-2 SD (Stunting) Modeled Estim...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	1680.6	1632.5	1571.2	1518.1	1462.3	1406.8
15820	Viet Nam	Weight-for-height ≥+2 SD (Overweight) Modeled ...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	463.3	491.3	519.8	545.1	568.7	590.1
16726	Cambodia	Height-for-age <-2 SD (Stunting) Modeled Estim...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	443.5	424.9	406.0	388.3	371.3	355.6
17443	Cambodia	Weight-for-height ≥+2 SD (Overweight) Modeled ...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	43.0	45.3	47.9	51.1	55.1	60.0
18354	Myanmar	Height-for-age <-2 SD (Stunting) Modeled Estim...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	1248.2	1218.8	1186.1	1151.9	1114.2	1075.4
19085	Myanmar	Weight-for-height ≥+2 SD (Overweight) Modeled ...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	49.6	45.5	41.9	38.8	36.1	33.7

```
cleaned_nutrition['Unit multiplier'].unique()

array(['Units', 'Thousands', nan], dtype=object)
```

```
cleaned_nutrition.query('`Unit multiplier` == "Thousands"')
```

	Geographic area		Indicator	Sex	Current age	Residence	UNIT_MULTIPLIER	Unit multiplier	UNIT_MEASURE	2017	2018	2019	2020	2021	2022	
8151	Singapore	Height-for-age <-2 SD (Stunting) Modeled Estim...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	7.2	7.4	7.6	7.6	7.4	7.1		
8495	Singapore	Weight-for-height >+2 SD (Overweight) Modeled ...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	7.9	8.3	8.8	9.2	9.2	9.1		
9001	Thailand	Height-for-age <-2 SD (Stunting) Modeled Estim...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	470.4	452.9	435.7	419.1	403.2	388.4		
9718	Thailand	Weight-for-height >+2 SD (Overweight) Modeled ...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	349.1	335.5	321.3	307.9	295.4	283.2		
10203	Malaysia	Height-for-age <-2 SD (Stunting) Modeled Estim...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	527.3	540.8	548.9	553.9	557.6	560.9		
10345	Malaysia	Weight-for-height >+2 SD (Overweight) Modeled ...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	156.8	155.6	153.4	150.6	147.8	145.1		
10843	Philippines	Height-for-age <-2 SD (Stunting) Modeled Estim...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	3575.7	3529.8	3541.7	3515.5	3459.8	3456.1		
11560	Philippines	Weight-for-height >+2 SD (Overweight) Modeled ...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	484.7	503.5	513.7	529.0	546.4	555.5		
11972	Brunei Darussalam	Height-for-age <-2 SD (Stunting) Modeled Estim...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	4.7	4.4	4.1	3.9	3.6	3.4		
11980	Brunei Darussalam	Weight-for-height >+2 SD (Overweight) Modeled ...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	3.1	3.1	3.0	3.0	2.9	2.9		
12258	Indonesia	Height-for-age <-2 SD (Stunting) Modeled Estim...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	7688.0	7517.3	7363.9	7212.9	7048.7	6896.6		
12604	Indonesia	Weight-for-height >+2 SD (Overweight) Modeled ...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	2761.6	2749.1	2699.8	2617.0	2503.7	2357.6		
13479	Lao People's Democratic Republic	Height-for-age <-2 SD (Stunting) Modeled Estim...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	261.5	252.8	244.0	235.2	226.4	218.4		
14196	Lao People's Democratic Republic	Weight-for-height >+2 SD (Overweight) Modeled ...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	22.4	24.1	25.8	27.6	29.6	31.6		
15103	Viet Nam	Height-for-age <-2 SD (Stunting) Modeled Estim...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	1680.6	1632.5	1571.2	1518.1	1462.3	1406.8		
15820	Viet Nam	Weight-for-height >+2 SD (Overweight) Modeled ...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	463.3	491.3	519.8	545.1	568.7	590.1		
16726	Cambodia	Height-for-age <-2 SD (Stunting) Modeled Estim...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	443.5	424.9	406.0	388.3	371.3	355.6		
17443	Cambodia	Weight-for-height >+2 SD (Overweight) Modeled ...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	43.0	45.3	47.9	51.1	55.1	60.0		
18354	Myanmar	Height-for-age <-2 SD (Stunting) Modeled Estim...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	1248.2	1218.8	1186.1	1151.9	1114.2	1075.4		
19085	Myanmar	Weight-for-height >+2 SD (Overweight) Modeled ...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	49.6	45.5	41.9	38.8	36.1	33.7		

"""

it can be noticed that the unit multiplier 3 and thousands are directly related with the same rows, and it is also noticed that these describe the modeled estimate for stunting and overweight, so I decided to place this on a separate dataframe

"""

```
modeled_estimate_nutrition = cleaned_nutrition.query('Unit multiplier' == "Thousands" and UNIT_MULTIPLIER == 3')
cleaned_nutrition = cleaned_nutrition.drop(modeled_estimate_nutrition.index)
modeled_estimate_nutrition
```

	Geographic area		Indicator	Sex	Current age	Residence	UNIT_MULTIPLIER	Unit multiplier	UNIT_MEASURE	2017	2018	2019	2020	2021	2022	
8151	Singapore	Height-for-age <-2 SD (Stunting) Modeled Estim...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	7.2	7.4	7.6	7.6	7.4	7.1		
8495	Singapore	Weight-for-height >+2 SD (Overweight) Modeled ...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	7.9	8.3	8.8	9.2	9.2	9.1		
9001	Thailand	Height-for-age <-2 SD (Stunting) Modeled Estim...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	470.4	452.9	435.7	419.1	403.2	388.4		
9718	Thailand	Weight-for-height >+2 SD (Overweight) Modeled ...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	349.1	335.5	321.3	307.9	295.4	283.2		
10203	Malaysia	Height-for-age <-2 SD (Stunting) Modeled Estim...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	527.3	540.8	548.9	553.9	557.6	560.9		
10345	Malaysia	Weight-for-height >+2 SD (Overweight) Modeled ...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	156.8	155.6	153.4	150.6	147.8	145.1		
10843	Philippines	Height-for-age <-2 SD (Stunting) Modeled Estim...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	3575.7	3529.8	3541.7	3515.5	3459.8	3456.1		
11560	Philippines	Weight-for-height >+2 SD (Overweight) Modeled ...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	484.7	503.5	513.7	529.0	546.4	555.5		
11972	Brunei Darussalam	Height-for-age <-2 SD (Stunting) Modeled Estim...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	4.7	4.4	4.1	3.9	3.6	3.4		
11980	Brunei Darussalam	Weight-for-height >+2 SD (Overweight) Modeled ...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	3.1	3.1	3.0	3.0	2.9	2.9		
12258	Indonesia	Height-for-age <-2 SD (Stunting) Modeled Estim...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	7688.0	7517.3	7363.9	7212.9	7048.7	6896.6		
12604	Indonesia	Weight-for-height >+2 SD (Overweight) Modeled ...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	2761.6	2749.1	2699.8	2617.0	2503.7	2357.6		
13479	Lao People's Democratic Republic	Height-for-age <-2 SD (Stunting) Modeled Estim...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	261.5	252.8	244.0	235.2	226.4	218.4		
14196	Lao People's Democratic Republic	Weight-for-height >+2 SD (Overweight) Modeled ...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	22.4	24.1	25.8	27.6	29.6	31.6		
15103	Viet Nam	Height-for-age <-2 SD (Stunting) Modeled Estim...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	1680.6	1632.5	1571.2	1518.1	1462.3	1406.8		
15820	Viet Nam	Weight-for-height >+2 SD (Overweight) Modeled ...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	463.3	491.3	519.8	545.1	568.7	590.1		
16726	Cambodia	Height-for-age <-2 SD (Stunting) Modeled Estim...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	443.5	424.9	406.0	388.3	371.3	355.6		
17443	Cambodia	Weight-for-height >+2 SD (Overweight) Modeled ...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	43.0	45.3	47.9	51.1	55.1	60.0		
18354	Myanmar	Height-for-age <-2 SD (Stunting) Modeled Estim...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	1248.2	1218.8	1186.1	1151.9	1114.2	1075.4		
19085	Myanmar	Weight-for-height >+2 SD (Overweight) Modeled ...	Total	Under 5 years old	Total	3.0	Thousands	NUMBER	49.6	45.5	41.9	38.8	36.1	33.7		

Next steps: [View recommended plots](#)

```
cleaned_nutrition['UNIT_MULTIPLIER'].unique()
```

```
array([ 0., nan])
```

```
cleaned_nutrition.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 14559 entries, 336 to 36978
Data columns (total 14 columns):
#   Column              Non-Null Count  Dtype

```

```
--- -----
0 Geographic area 14559 non-null object
1 Indicator      14559 non-null object
2 Sex            14559 non-null object
3 Current age    14559 non-null object
4 Residence      14559 non-null object
5 UNIT_MULTIPLIER 14539 non-null float64
6 Unit multiplier 14539 non-null object
7 UNIT_MEASURE   14559 non-null object
8 2017           5908 non-null float64
9 2018           344 non-null float64
10 2019          5514 non-null float64
11 2020          45 non-null float64
12 2021          2901 non-null float64
13 2022          29 non-null float64
dtypes: float64(7), object(7)
memory usage: 1.7+ MB

# removes the unit multiplier columns
cleaned_nutrition = pd.concat([
    cleaned_nutrition.iloc[:, :5],
    cleaned_nutrition.iloc[:, 7:]
], axis = 1)
cleaned_nutrition.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 14559 entries, 336 to 36978
Data columns (total 12 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   Geographic area  14559 non-null  object
1   Indicator        14559 non-null  object
2   Sex              14559 non-null  object
3   Current age      14559 non-null  object
4   Residence        14559 non-null  object
5   UNIT_MEASURE     14559 non-null  object
6   2017             5908 non-null   float64
7   2018             344 non-null    float64
8   2019             5514 non-null   float64
9   2020             45 non-null     float64
10  2021            2901 non-null   float64
11  2022            29 non-null     float64
dtypes: float64(6), object(6)
memory usage: 1.4+ MB

# checks for NaN values in categorical data
cleaned_nutrition['Geographic area'].unique()

array(['Indonesia', 'Cambodia', 'Lao People's Democratic Republic',
       'Myanmar', 'Malaysia', 'Philippines', 'Thailand', 'Singapore',
       'Brunei Darussalam', 'Viet Nam'], dtype=object)

cleaned_nutrition.Indicator.unique()

array(['Height-for-age <-2 SD (stunting)',
       'Height-for-age <-2 SD and 2 -3 SD (Moderate Stunting only), Survey Estimates',
       'Height-for-age <-3 SD (Severe Stunting)',
       'Weight-for-age <-2 SD (Underweight)',
       'Weight-for-age <-3 SD (Severe Underweight)',
       'Weight-for-height <-2 SD (wasting)',
       'Weight-for-height <-2 SD and 2 -3 SD, Survey Estimates',
       'Weight-for-height <-3 SD (severe wasting)',
       'Weight-for-height >+2 SD (overweight)',
       'Continued breastfeeding (12-15 months)',
       'Continued breastfeeding (12-23 months)',
       'Continued breastfeeding (20-23 months)',
       'Ever breastfed (0-23 months)',
       'Early initiation of breastfeeding',
       'Exclusive breastfeeding (0-5 months)',
       'Exclusively Breastfed for the First Two Days After Birth',
       'Mixed Milk Feeding (0-5 months)',
       'Egg and/or flesh foods consumption (6-23months)',
       'Food group- Breastmilk (6-23months)',
       'Food group- Dairy (milk, infant formula, yogurt, cheese) (6-23months)',
       'Food group- Eggs (6-23months)',
       'Food group- Flesh foods (meat, poultry, fish and organ meats) (6-23months)',
       'Food group- Grains, roots, tubers and plantains (6-23months)',
       'Introduction to solid, semi-solid foods (6-8 months)',
       'Food group- Pulses (beans, peas, lentils), nuts and seeds (6-23months)',
       'Minimum acceptable diet (children aged 6-23 months)',
       'Minimum diet diversity (children aged 6 to 23 months)',
       'Minimum meal frequency (children aged 6 to 23 months)',
       'Food group- Other fruits and vegetables (6-23months)',
       'Food group- Vitamin-A rich fruits and vegetables (6-23months)',
       'Zero vegetable or fruit consumption (6-23months)',
       'Mean BMI-for-age', 'BMI-for-age <-1 SD', 'BMI-for-age <-2 SD',
       'BMI-for-age <-3 SD', 'BMI-for-age >+1 SD', 'BMI-for-age >+2 SD',
       'BMI-for-age >+3 SD', 'BMI-for-age (Standard Deviation)',
       'Height-for-age <-2SD & Weight-for-height <-2SD',
       'Height-for-age <-2SD & Weight-for-height >+2SD',
       'Height-for-age <-3SD & Weight-for-height <-2SD',
       'Height-for-age <-3SD & Weight-for-height >+2SD',
       'Mean Height-for-age', 'Height-for-age <-1 SD',
       'Height-for-age >+1 SD', 'Height-for-age >+2 SD',
       'Height-for-age >+3 SD', 'Height-for-age (Standard Deviation)',
       'Weight-for-age (Mean)', 'Weight-for-age <-1 SD',
       'Weight-for-age >+1 SD', 'Weight-for-age >+2 SD',
       'Weight-for-age (>+3 SD)', 'Weight-for-age (Standard Deviation)',
       'Mean Weight-for-height', 'Weight-for-height <-1 SD',
       'Weight-for-height >+1 SD', 'Weight-for-height >+3 SD',
```

```
'Weight-for-height (Standard Deviation)',
'BMI-for-age <-1 SD and ≥ -2 SD, Survey Estimates',
'BMI-for-age <+2 SD and ≥ -3 SD, Survey Estimates',
'BMI-for-age ≥+1 SD and ≤+2 SD, Survey Estimates',
'BMI-for-age ≥+2 SD and ≤+3 SD, Survey Estimates',
'Height-for-age <-1 SD and ≥ -2 SD (Mild Stunting only), Survey Estimates',
'Height-for-age ≥+1 SD and ≤+2 SD, Survey Estimates',
'Height-for-age ≥+2 SD and ≤+3 SD, Survey Estimates',
'Weight-for-age <-2SD & Weight-for-height <-2SD',
'Weight-for-age <-3SD & Weight-for-height <-2SD',
'Weight-for-age <-1 SD and ≥ -2 SD, Survey Estimates',
'Weight-for-age ≥+2 SD and ≥ -3 SD, Survey Estimates'
```

```
cleaned_nutrition.Sex.unique()
```

```
array(['Female', 'Total', 'Male'], dtype=object)
```

```
cleaned_nutrition.Residence.unique()
```

```
array(['Total', 'Urban', 'Rural'], dtype=object)
```

```
cleaned_nutrition.UNIT_MEASURE.unique()
```

```
array(['PCNT', 'NUMBER'], dtype=object)
```

```
# getting the unique values
a = cleaned_nutrition['Geographic area'].unique()
b = cleaned_nutrition.Indicator.unique()
c = cleaned_nutrition.Sex.unique()
d = cleaned_nutrition.Residence.unique()
e = cleaned_nutrition.UNIT_MEASURE.unique()
```

```
# creates the dictionaries
a_key = [i for i in a]
a_value = [i for i in range(len(a))]
a_dict = {a_key[i] : a_value[i] for i in range(len(a_value))}
```

```
b_key = [i for i in b]
b_value = [i for i in range(len(b))]
b_dict = {b_key[i] : b_value[i] for i in range(len(b_value))}
```

```
c_key = [i for i in c]
c_value = [i for i in range(len(c))]
c_dict = {c_key[i] : c_value[i] for i in range(len(c_value))}
```

```
d_key = [i for i in d]
d_value = [i for i in range(len(d))]
d_dict = {d_key[i] : d_value[i] for i in range(len(d_value))}
```

```
e_key = [i for i in e]
e_value = [i for i in range(len(e))]
e_dict = {e_key[i] : e_value[i] for i in range(len(e_value))}
```

```
cleaned_nutrition['Geographic area'].replace(a_dict, inplace = True)
cleaned_nutrition.Indicator.replace(b_dict, inplace = True)
cleaned_nutrition.Sex.replace(c_dict, inplace = True)
cleaned_nutrition.Residence.replace(d_dict, inplace = True)
cleaned_nutrition.UNIT_MEASURE.replace(e_dict, inplace = True)
```

```
cleaned_nutrition
```

	Geographic area	Indicator	Sex	Current age	Residence	UNIT_MEASURE	2017	2018	2019	2020	2021	2022	
	336	0	0	0	Under 5 years old	0	0	NaN	29.7	NaN	NaN	NaN	
	337	0	1	0	Under 5 years old	0	0	NaN	18.9	NaN	NaN	NaN	
	355	0	2	0	Under 5 years old	0	0	NaN	10.8	NaN	NaN	NaN	
	481	0	3	0	Under 5 years old	0	0	NaN	16.6	NaN	NaN	NaN	
	499	0	4	0	Under 5 years old	0	0	NaN	3.3	NaN	NaN	NaN	
	
	36974	6	26	2	6 to 23 months old	0	0	NaN	NaN	69.4	NaN	NaN	
	36975	6	27	2	6 to 23 months old	0	0	NaN	NaN	87.6	NaN	NaN	
	36976	6	28	2	6 to 23 months old	0	0	NaN	NaN	53.7	NaN	NaN	
	36977	6	29	2	6 to 23 months old	0	0	NaN	NaN	79.0	NaN	NaN	
	36978	6	30	2	6 to 23 months old	0	0	NaN	NaN	12.7	NaN	NaN	

14559 rows × 12 columns

Next steps:

[View recommended plots](#)

```
# fills NaN values with the mean
years = ['2017', '2018', '2019', '2020', '2021', '2022']

row_means = cleaned_nutrition[years].mean(axis=1)

for index, row in cleaned_nutrition.iterrows():
    cleaned_nutrition.loc[index, years] = row[years].fillna(row_means[index])
```

cleaned_nutrition

	Geographic area	Indicator	Sex	Current age	Residence	UNIT_MEASURE	2017	2018	2019	2020	2021	2022
336		0	0	0	Under 5 years old	0	0	29.7	29.7	29.7	29.7	29.7
337		0	1	0	Under 5 years old	0	0	18.9	18.9	18.9	18.9	18.9
355		0	2	0	Under 5 years old	0	0	10.8	10.8	10.8	10.8	10.8
481		0	3	0	Under 5 years old	0	0	16.6	16.6	16.6	16.6	16.6
499		0	4	0	Under 5 years old	0	0	3.3	3.3	3.3	3.3	3.3
...
36974		6	26	2	6 to 23 months old	0	0	69.4	69.4	69.4	69.4	69.4
36975		6	27	2	6 to 23 months old	0	0	87.6	87.6	87.6	87.6	87.6
36976		6	28	2	6 to 23 months old	0	0	53.7	53.7	53.7	53.7	53.7
36977		6	29	2	6 to 23 months old	0	0	79.0	79.0	79.0	79.0	79.0
36978		6	30	2	6 to 23 months old	0	0	12.7	12.7	12.7	12.7	12.7

14559 rows × 12 columns

Next steps: [View recommended plots](#)

```
cleaned_nutrition.info()
# all data are clean
# it should be noted that age is still categorical
# as I had no idea how to convert to numerical with months and years
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 14559 entries, 336 to 36978
Data columns (total 12 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Geographic area  14559 non-null  int64
1   Indicator        14559 non-null  int64
2   Sex              14559 non-null  int64
3   Current age      14559 non-null  object
4   Residence        14559 non-null  int64
5   UNIT_MEASURE     14559 non-null  int64
6   2017             14559 non-null  float64
7   2018             14559 non-null  float64
8   2019             14559 non-null  float64
9   2020             14559 non-null  float64
10  2021             14559 non-null  float64
11  2022             14559 non-null  float64
dtypes: float64(6), int64(5), object(1)
memory usage: 1.9+ MB
```



cleaned_nutrition.describe()

	Geographic area	Indicator	Sex	Residence	UNIT_MEASURE	2017	2018	2019	2020	2021	2022
count	14559.000000	14559.000000	14559.00000	14559.000000	14559.000000	14559.000000	14559.000000	14559.000000	14559.000000	14559.000000	14559.000000
mean	3.224741	39.368981	1.00000	0.331960	0.125558	16.662856	16.657634	16.660077	16.660250	16.654143	16.654603
std	2.193025	22.652023	0.61822	0.665347	0.331362	23.050041	23.039329	23.047134	23.045781	23.036756	23.037958
min	0.000000	0.000000	0.00000	0.000000	0.000000	-1.800000	-1.800000	-1.800000	-1.800000	-1.800000	-1.800000
25%	2.000000	21.000000	1.00000	0.000000	0.000000	2.000000	2.000000	2.000000	2.000000	2.000000	2.000000
50%	2.000000	42.000000	1.00000	0.000000	0.000000	6.400000	6.400000	6.400000	6.400000	6.400000	6.400000
75%	6.000000	57.000000	1.00000	0.000000	0.000000	21.500000	21.500000	21.500000	21.500000	21.500000	21.500000
max	9.000000	92.000000	2.00000	2.000000	1.000000	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000

a_dict

```
{'Indonesia': 0,
 'Cambodia': 1,
 'Lao People's Democratic Republic': 2,
 'Myanmar': 3,
 'Malaysia': 4,
 'Philippines': 5,
 'Thailand': 6,
 'Singapore': 7,
 'Brunei Darussalam': 8,
 'Viet Nam': 9}
```



```
indonesian_data = cleaned_nutrition.query('`Geographic area` == 0')
indonesian_data
```


	Geographic area	Indicator	Sex	Current age	Residence	UNIT_MEASURE	2017	2018	2019	2020	2021	2022	
	336	0	0	0	Under 5 years old	0	0	29.7	29.7	29.7	29.7	29.7	
	337	0	1	0	Under 5 years old	0	0	18.9	18.9	18.9	18.9	18.9	
	355	0	2	0	Under 5 years old	0	0	10.8	10.8	10.8	10.8	10.8	
	481	0	3	0	Under 5 years old	0	0	16.6	16.6	16.6	16.6	16.6	
	499	0	4	0	Under 5 years old	0	0	3.3	3.3	3.3	3.3	3.3	
	
	31301	0	26	2	6 to 23 months old	0	0	52.6	52.6	52.6	52.6	52.6	
	31302	0	27	2	6 to 23 months old	0	0	72.3	72.3	72.3	72.3	72.3	
	31303	0	28	2	6 to 23 months old	0	0	26.6	26.6	26.6	26.6	26.6	
	31304	0	29	2	6 to 23 months old	0	0	77.3	77.3	77.3	77.3	77.3	
	31305	0	30	2	6 to 23 months old	0	0	19.5	19.5	19.5	19.5	19.5	

694 rows × 12 columns

Next steps: [View recommended plots](#)

```
myanmar_data = cleaned_nutrition.query('`Geographic area` == 3')
myanmar_data
```



	Geographic area	Indicator	Sex	Current age	Residence	UNIT_MEASURE	2017	2018	2019	2020	2021	2022	
	3041	3	43	0	Under 5 years old	0	1	-1.3	-1.3	-1.3	-1.3	-1.3	
	3095	3	0	0	Under 5 years old	0	0	26.8	26.8	26.8	26.8	26.8	
	3113	3	1	0	Under 5 years old	0	0	19.6	19.6	19.6	19.6	19.6	
	3131	3	2	0	Under 5 years old	0	0	7.2	7.2	7.2	7.2	7.2	
	3239	3	48	0	Under 5 years old	0	1	1.2	1.2	1.2	1.2	1.2	
	
	34057	3	54	2	Under 5 years old	0	1	1.0	1.0	1.0	1.0	1.0	
	34075	3	55	2	Under 5 years old	0	1	-0.6	-0.6	-0.6	-0.6	-0.6	
	34147	3	6	2	Under 5 years old	0	0	6.3	6.3	6.3	6.3	6.3	
	34219	3	8	2	Under 5 years old	0	0	1.1	1.1	1.1	1.1	1.1	
	34273	3	59	2	Under 5 years old	0	1	1.0	1.0	1.0	1.0	1.0	

207 rows × 12 columns



Next steps: [View recommended plots](#)

It can be noticed that more developed countries such as Indonesia have higher mean compared to Myanmar

```
indonesian_data.describe()
```

	Geographic area	Indicator	Sex	Residence	UNIT_MEASURE	2017	2018	2019	2020	2021	2022	
count	694.0	694.000000	694.000000	694.000000	694.0	694.000000	694.000000	694.000000	694.000000	694.000000	694.000000	
mean	0.0	18.802594	1.000000	0.253602	0.0	51.251739	51.253791	51.251978	51.251202	51.250016	51.248575	
std	0.0	10.880298	0.299109	0.599863	0.0	26.418627	26.419323	26.419693	26.420527	26.421759	26.423519	
min	0.0	0.000000	0.000000	0.000000	0.0	2.400000	2.400000	2.400000	2.400000	2.400000	2.400000	
25%	0.0	12.000000	1.000000	0.000000	0.0	29.400000	29.400000	29.400000	29.400000	29.400000	29.400000	
50%	0.0	19.000000	1.000000	0.000000	0.0	53.700000	53.700000	53.700000	53.700000	53.700000	53.700000	
75%	0.0	25.000000	1.000000	0.000000	0.0	71.275000	71.275000	71.275000	71.275000	71.275000	71.275000	
max	0.0	92.000000	2.000000	2.000000	0.0	97.200000	97.200000	97.200000	97.200000	97.200000	97.200000	

```
myanmar_data.describe()
```

	Geographic area	Indicator	Sex	Residence	UNIT_MEASURE	2017	2018	2019	2020	2021	2022	
count	207.0	207.000000	207.000000	207.000000	207.000000	207.000000	207.000000	207.000000	207.000000	207.000000	207.000000	
mean	3.0	35.338164	1.000000	0.217391	0.415459	12.290395	12.189803	12.321195	12.259828	12.255522	12.252141	
std	0.0	30.004557	0.355266	0.562663	0.493996	23.716129	23.405290	23.840913	23.645942	23.644051	23.642298	
min	3.0	0.000000	0.000000	0.000000	0.000000	-1.600000	-1.600000	-1.600000	-1.600000	-1.600000	-1.600000	
25%	3.0	4.000000	1.000000	0.000000	0.000000	0.750000	0.750000	0.750000	0.750000	0.750000	0.750000	
50%	3.0	43.000000	1.000000	0.000000	0.000000	1.400000	1.400000	1.400000	1.400000	1.400000	1.400000	
75%	3.0	55.000000	1.000000	0.000000	1.000000	12.755928	12.744582	12.743633	12.742151	12.746573	12.746573	
max	3.0	92.000000	2.000000	2.000000	1.000000	93.000000	89.900000	93.000000	89.900000	89.900000	89.900000	

```
c_dict
{'Female': 0, 'Total': 1, 'Male': 2}
```

```
female_data = cleaned_nutrition.query('`Sex` == 0')
female_data
```

	Geographic area	Indicator	Sex	Current age	Residence	UNIT_MEASURE	2017	2018	2019	2020	2021	2022
336		0	0	0	Under 5 years old	0	0	29.7	29.7	29.7	29.7	29.7
337		0	1	0	Under 5 years old	0	0	18.9	18.9	18.9	18.9	18.9
355		0	2	0	Under 5 years old	0	0	10.8	10.8	10.8	10.8	10.8
481		0	3	0	Under 5 years old	0	0	16.6	16.6	16.6	16.6	16.6
499		0	4	0	Under 5 years old	0	0	3.3	3.3	3.3	3.3	3.3
...
6440		6	26	0	6 to 23 months old	0	0	69.0	69.0	69.0	69.0	69.0
6441		6	27	0	6 to 23 months old	0	0	84.9	84.9	84.9	84.9	84.9
6442		6	28	0	6 to 23 months old	0	0	56.6	56.6	56.6	56.6	56.6
6443		6	29	0	6 to 23 months old	0	0	74.8	74.8	74.8	74.8	74.8
6444		6	30	0	6 to 23 months old	0	0	15.3	15.3	15.3	15.3	15.3

2782 rows × 12 columns

Next steps: [View recommended plots](#)

```
male_data = cleaned_nutrition.query('`Sex` == 2')
male_data
```

	Geographic area	Indicator	Sex	Current age	Residence	UNIT_MEASURE	2017	2018	2019	2020	2021	2022
30902		0	0	2	Under 5 years old	0	0	31.7	31.7	31.7	31.7	31.7
30903		0	1	2	Under 5 years old	0	0	19.6	19.6	19.6	19.6	19.6
30921		0	2	2	Under 5 years old	0	0	12.1	12.1	12.1	12.1	12.1
31047		0	3	2	Under 5 years old	0	0	18.8	18.8	18.8	18.8	18.8
31065		0	4	2	Under 5 years old	0	0	4.5	4.5	4.5	4.5	4.5
...
36974		6	26	2	6 to 23 months old	0	0	69.4	69.4	69.4	69.4	69.4
36975		6	27	2	6 to 23 months old	0	0	87.6	87.6	87.6	87.6	87.6
36976		6	28	2	6 to 23 months old	0	0	53.7	53.7	53.7	53.7	53.7
36977		6	29	2	6 to 23 months old	0	0	79.0	79.0	79.0	79.0	79.0
36978		6	30	2	6 to 23 months old	0	0	12.7	12.7	12.7	12.7	12.7

2782 rows × 12 columns

Next steps: [View recommended plots](#)

It can also be observed that the mean is higher on males than females

```
female_data.describe()
```

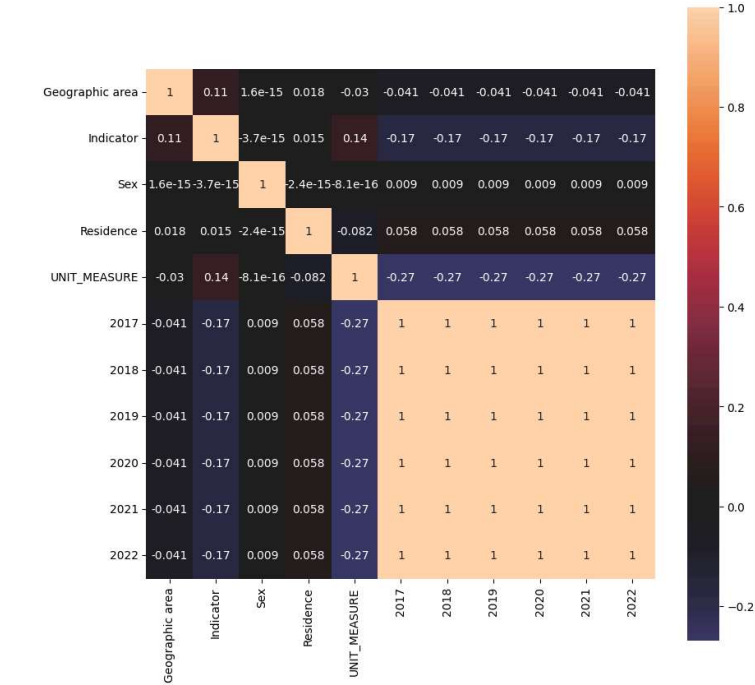
	Geographic area	Indicator	Sex	Residence	UNIT_MEASURE	2017	2018	2019	2020	2021	2022
count	2782.000000	2782.000000	2782.0	2782.0	2782.000000	2782.000000	2782.000000	2782.000000	2782.000000	2782.000000	2782.000000
mean	3.251258	41.610712	0.0	0.0	0.172538	9.621031	9.621031	9.621031	9.621031	9.621031	9.621031
std	2.139622	21.641080	0.0	0.0	0.377915	14.159363	14.159363	14.159363	14.159363	14.159363	14.159363
min	0.000000	0.000000	0.0	0.0	0.000000	-1.700000	-1.700000	-1.700000	-1.700000	-1.700000	-1.700000
25%	2.000000	33.000000	0.0	0.0	0.000000	1.400000	1.400000	1.400000	1.400000	1.400000	1.400000
50%	2.000000	45.000000	0.0	0.0	0.000000	3.900000	3.900000	3.900000	3.900000	3.900000	3.900000
75%	6.000000	58.000000	0.0	0.0	0.000000	11.800000	11.800000	11.800000	11.800000	11.800000	11.800000
max	6.000000	76.000000	0.0	0.0	1.000000	97.000000	97.000000	97.000000	97.000000	97.000000	97.000000

```
male_data.describe()
```

	Geographic area	Indicator	Sex	Residence	UNIT_MEASURE	2017	2018	2019	2020	2021	2022	
count	2782.000000	2782.000000	2782.0	2782.0	2782.000000	2782.000000	2782.000000	2782.000000	2782.000000	2782.000000	2782.000000	
mean	3.251258	41.610712	2.0	0.0	0.172538	10.288640	10.288640	10.288640	10.288640	10.288640	10.288640	

```
import matplotlib.pyplot as plt
import seaborn as sns
plt.figure(figsize=(10,10))
sns.heatmap(
    cleaned_nutrition.sort_index().corr(),
    annot=True, center=0, square=True
)

<ipython-input-234-e04bb3e3448e>:5: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.
  cleaned_nutrition.sort_index().corr(),
<Axes: >
```



It can be noticed that the years have perfect positive correlations with each other, which may be due to the `fillna()` method earlier.

However, there are only weak correlations with the years and the categorical (as numbers) data.

- 1
- 1
- 1
- 1
- 1