In [1]:
```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

In [10]:
```python
df = pd.read_csv(r'C:\Users\Sandeep Immadi\Downloads\Mcdonald .csv')
df
```

Out[10]:

	Category	Item	Serving Size	Calories	Calories from Fat	Total Fat	Total Fat (% Daily Value)	Saturated Fat	Saturated Fat (% Daily Value)	Trans Fat	...	Carbohydrates	Carbohydrates (% Daily Value)	Dietary Fiber	Dietary Fiber (% Daily Value)	Sugars	Protein	Vitamin A (% Daily Value)	Vitamin C (% Daily Value)
0	Breakfast	Egg McMuffin	4.8 oz (136 g)	300	120	13.0	20	5.0	25	0.0	...	31	10	4	17	3	17	10	0
1	Breakfast	Egg White Delight	4.8 oz (135 g)	250	70	8.0	12	3.0	15	0.0	...	30	10	4	17	3	18	6	0
2	Breakfast	Sausage McMuffin	3.9 oz (111 g)	370	200	23.0	35	8.0	42	0.0	...	29	10	4	17	2	14	8	0
3	Breakfast	Sausage McMuffin with Egg	5.7 oz (161 g)	450	250	28.0	43	10.0	52	0.0	...	30	10	4	17	2	21	15	0
4	Breakfast	Sausage McMuffin with Egg Whites	5.7 oz (161 g)	400	210	23.0	35	8.0	42	0.0	...	30	10	4	17	2	21	6	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
255	Smoothies & Shakes	McFlurry with Oreo Cookies (Small)	10.1 oz (285 g)	510	150	17.0	26	9.0	44	0.5	...	80	27	1	4	64	12	15	0
256	Smoothies & Shakes	McFlurry with Oreo Cookies (Medium)	13.4 oz (381 g)	690	200	23.0	35	12.0	58	1.0	...	106	35	1	5	85	15	20	0
257	Smoothies & Shakes	McFlurry with Oreo Cookies (Snack)	6.7 oz (190 g)	340	100	11.0	17	6.0	29	0.0	...	53	18	1	2	43	8	10	0
258	Smoothies & Shakes	McFlurry with Reese's Peanut Butter Cups (Medium)	14.2 oz (403 g)	810	290	32.0	50	15.0	76	1.0	...	114	38	2	9	103	21	20	0
259	Smoothies & Shakes	McFlurry with Reese's Peanut Butter Cups (Snack)	7.1 oz (202 g)	410	150	16.0	25	8.0	38	0.0	...	57	19	1	5	51	10	10	0

260 rows × 24 columns

In [11]:
```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 260 entries, 0 to 259
Data columns (total 24 columns):
 #   Column                         Non-Null Count  Dtype
---  ------                         --------------  -----
 0   Category                       260 non-null    object
 1   Item                           260 non-null    object
 2   Serving Size                   260 non-null    object
 3   Calories                       260 non-null    int64
 4   Calories from Fat              260 non-null    int64
 5   Total Fat                      260 non-null    float64
 6   Total Fat (% Daily Value)      260 non-null    int64
 7   Saturated Fat                  260 non-null    float64
 8   Saturated Fat (% Daily Value)  260 non-null    int64
 9   Trans Fat                      260 non-null    float64
 10  Cholesterol                    260 non-null    int64
 11  Cholesterol (% Daily Value)    260 non-null    int64
 12  Sodium                         260 non-null    int64
 13  Sodium (% Daily Value)         260 non-null    int64
 14  Carbohydrates                  260 non-null    int64
 15  Carbohydrates (% Daily Value)  260 non-null    int64
 16  Dietary Fiber                  260 non-null    int64
 17  Dietary Fiber (% Daily Value)  260 non-null    int64
 18  Sugars                         260 non-null    int64
 19  Protein                        260 non-null    int64
 20  Vitamin A (% Daily Value)      260 non-null    int64
 21  Vitamin C (% Daily Value)      260 non-null    int64
 22  Calcium (% Daily Value)        260 non-null    int64
 23  Iron (% Daily Value)           260 non-null    int64
dtypes: float64(3), int64(18), object(3)
memory usage: 48.9+ KB
```

In [12]:
```python
df.describe()
```

Out[12]:

	Calories	Calories from Fat	Total Fat	Total Fat (% Daily Value)	Saturated Fat	Saturated Fat (% Daily Value)	Trans Fat	Cholesterol	Cholesterol (% Daily Value)	Sodium	...	Carbohydrates	Carbohydrates (% Daily Value)	Dietary Fiber	D Fi
count	260.000000	260.000000	260.000000	260.000000	260.000000	260.000000	260.000000	260.000000	260.000000	260.000000	...	260.000000	260.000000	260.000000	260.0
mean	368.269231	127.096154	14.165385	21.815385	6.007692	29.965385	0.203846	54.942308	18.392308	495.750000	...	47.346154	15.780769	1.630769	6.5
std	240.269886	127.875914	14.205998	21.885199	5.321873	26.639209	0.429133	87.269257	29.091653	577.026323	...	28.252232	9.419544	1.567717	6.3
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.0
25%	210.000000	20.000000	2.375000	3.750000	1.000000	4.750000	0.000000	5.000000	2.000000	107.500000	...	30.000000	10.000000	0.000000	0.0
50%	340.000000	100.000000	11.000000	17.000000	5.000000	24.000000	0.000000	35.000000	11.000000	190.000000	...	44.000000	15.000000	1.000000	5.0
75%	500.000000	200.000000	22.250000	35.000000	10.000000	48.000000	0.000000	65.000000	21.250000	865.000000	...	60.000000	20.000000	3.000000	10.0
max	1880.000000	1060.000000	118.000000	182.000000	20.000000	102.000000	2.500000	575.000000	192.000000	3600.000000	...	141.000000	47.000000	7.000000	28.0

8 rows × 21 columns

In [13]:
```python
df.isnull().sum()
```

Out[13]:
```
Category                         0
Item                             0
Serving Size                     0
Calories                         0
Calories from Fat                0
Total Fat                        0
Total Fat (% Daily Value)        0
Saturated Fat                    0
Saturated Fat (% Daily Value)    0
Trans Fat                        0
Cholesterol                      0
Cholesterol (% Daily Value)      0
Sodium                           0
Sodium (% Daily Value)           0
Carbohydrates                    0
Carbohydrates (% Daily Value)    0
Dietary Fiber                    0
Dietary Fiber (% Daily Value)    0
Sugars                           0
Protein                          0
Vitamin A (% Daily Value)        0
Vitamin C (% Daily Value)        0
Calcium (% Daily Value)          0
Iron (% Daily Value)             0
dtype: int64
```

In [14]:
```python
df['Category'].value_counts()
```

Out[14]:
```
Coffee & Tea          95
Breakfast             42
Smoothies & Shakes    28
Chicken & Fish        27
Beverages             27
Beef & Pork           15
Snacks & Sides        13
Desserts               7
Salads                 6
Name: Category, dtype: int64
```

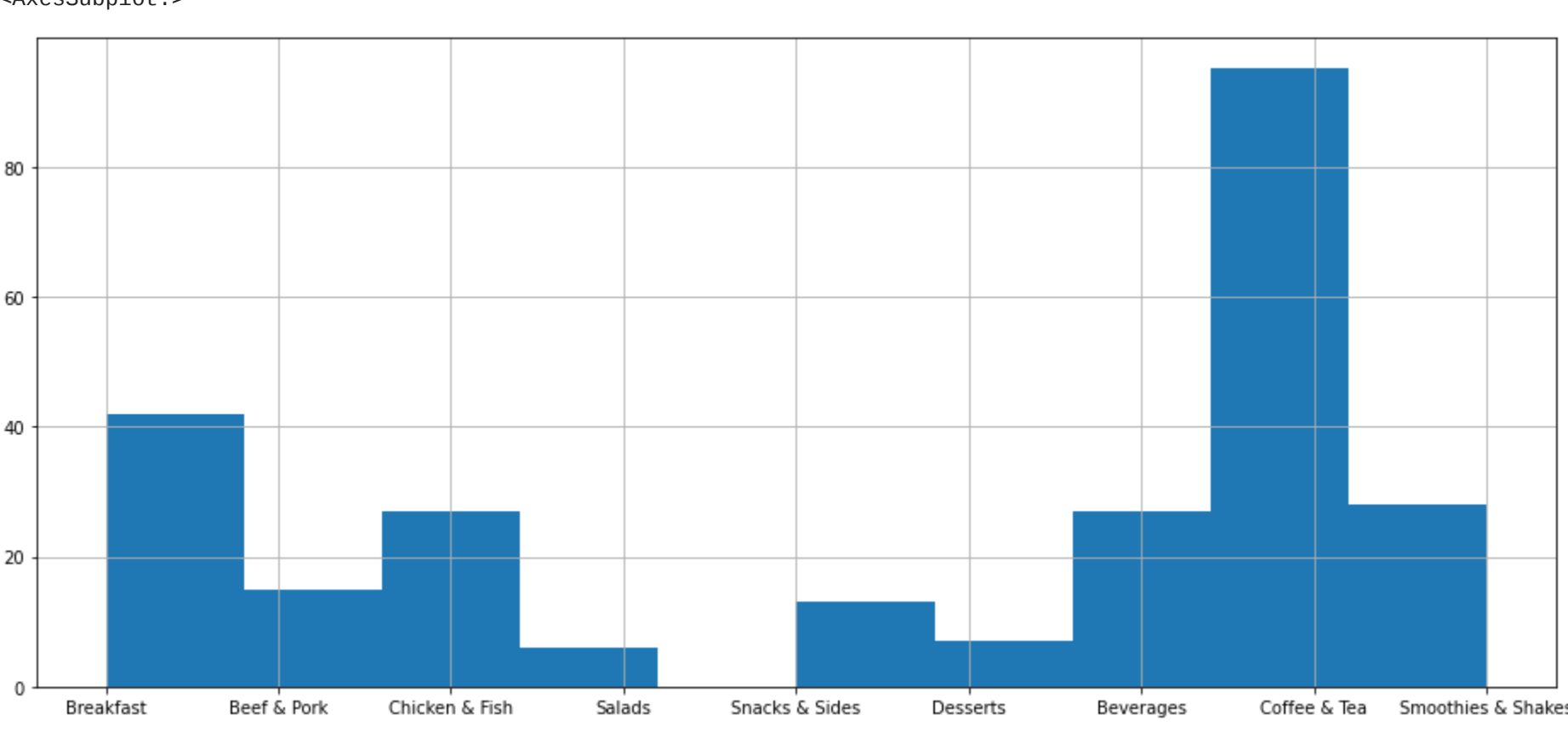
In [17]:
```python
df['Item'].value_counts()
```

Out[17]:
```
Fat Free Chocolate Milk Jug                           1
Latte with Sugar Free French Vanilla Syrup (Small)    1
Sausage Burrito                                       1
Frappé Caramel (Small)                                1
Hot Chocolate (Medium)                                1
                                                     ..
Premium Grilled Chicken Classic Sandwich              1
Latte with Sugar Free French Vanilla Syrup (Large)    1
Hotcakes                                              1
Mocha (Small)                                         1
Bacon Clubhouse Crispy Chicken Sandwich               1
Name: Item, Length: 260, dtype: int64
```

Plot graphically which food categories have the highest and lowest varieties.

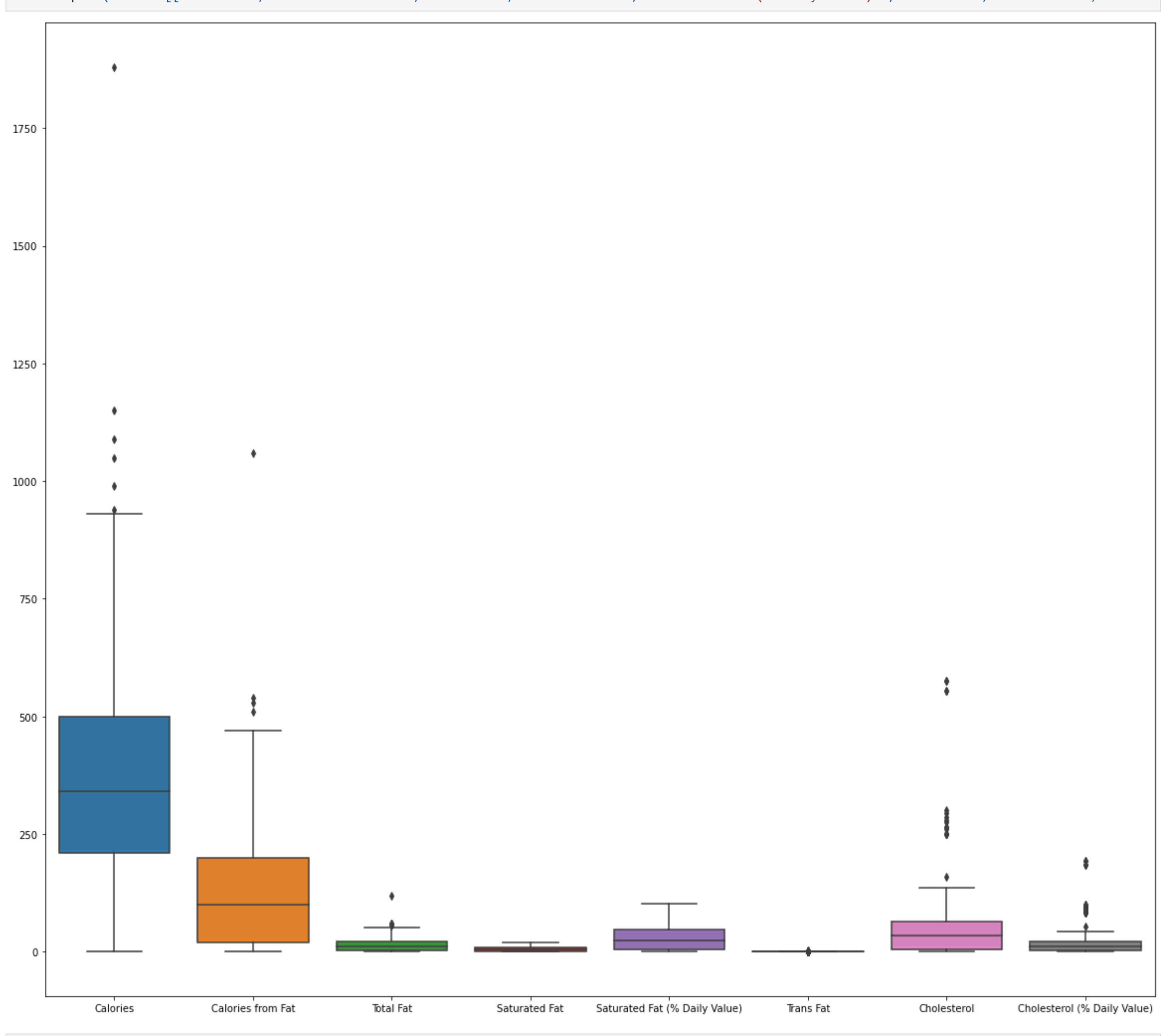
In [21]:
```python
plt.figure(figsize=(16,7))
df["Category"].hist()

# Highest sales : coffee and tea; Lowest sales :salads
```

Out[21]: <AxesSubplot:>

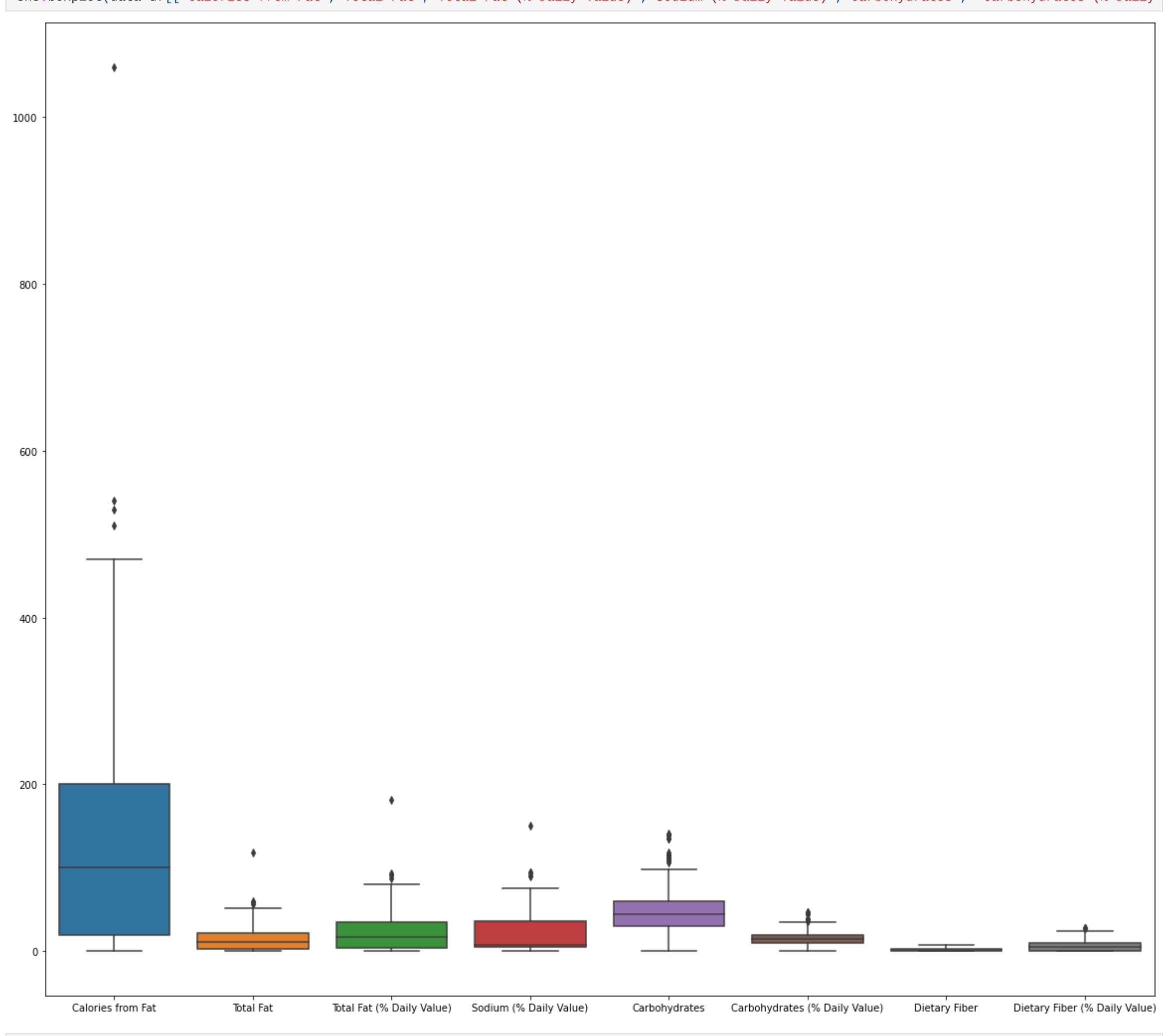


Which all variables have an outlier?

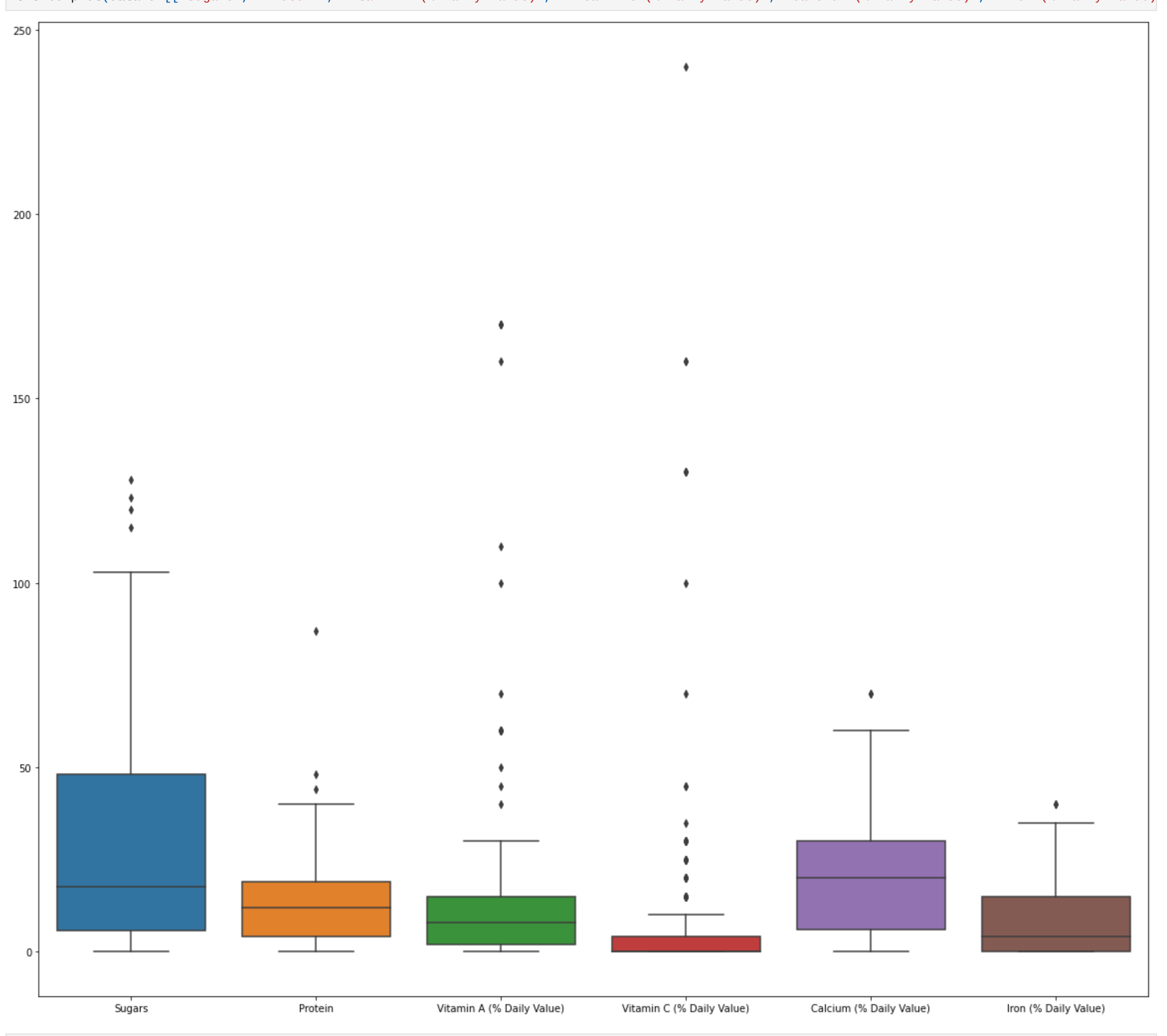
In [31]:
```python
plt.figure(figsize=(20,18))
sns.boxplot(data=df[['Calories','Calories from Fat','Total Fat','Saturated Fat','Saturated Fat (% Daily Value)' ,'Trans Fat','Cholesterol','Cholest
```



In [32]:
```python
plt.figure(figsize=(20,18))

sns.boxplot(data=df[['Calories from Fat','Total Fat','Total Fat (% Daily Value)','Sodium (% Daily Value)','Carbohydrates', 'Carbohydrates (% Daily
```



In [33]:
```python
plt.figure(figsize=(20,18))
sns.boxplot(data=df[['Sugars', 'Protein','Vitamin A (% Daily Value)', 'Vitamin C (% Daily Value)', 'Calcium (% Daily Value)', 'Iron (% Daily Value)
```



In [ ]:
```python
# columns which having an outliers :

#Calories,Calories from Fat,Total Fat,
#Saturated Fat (% Daily Value),Cholestrol,Cholestrol (% Daily Value),Total Fat,
#Total Fat (% Daily Value),Sodium (% Daily Value),Sodium,Carbohydrates,Carbohydrates (% Daily Value),
#Dietary Fiber (% Daily Value),Sugars,Protien,Vitamin A (% Daily Value),
#Vitamin C (% Daily Value),Calcium (% Daily Value),Iron (% Daily Value)
```

Which variables have the highest correlation? Plot them and find out the value?

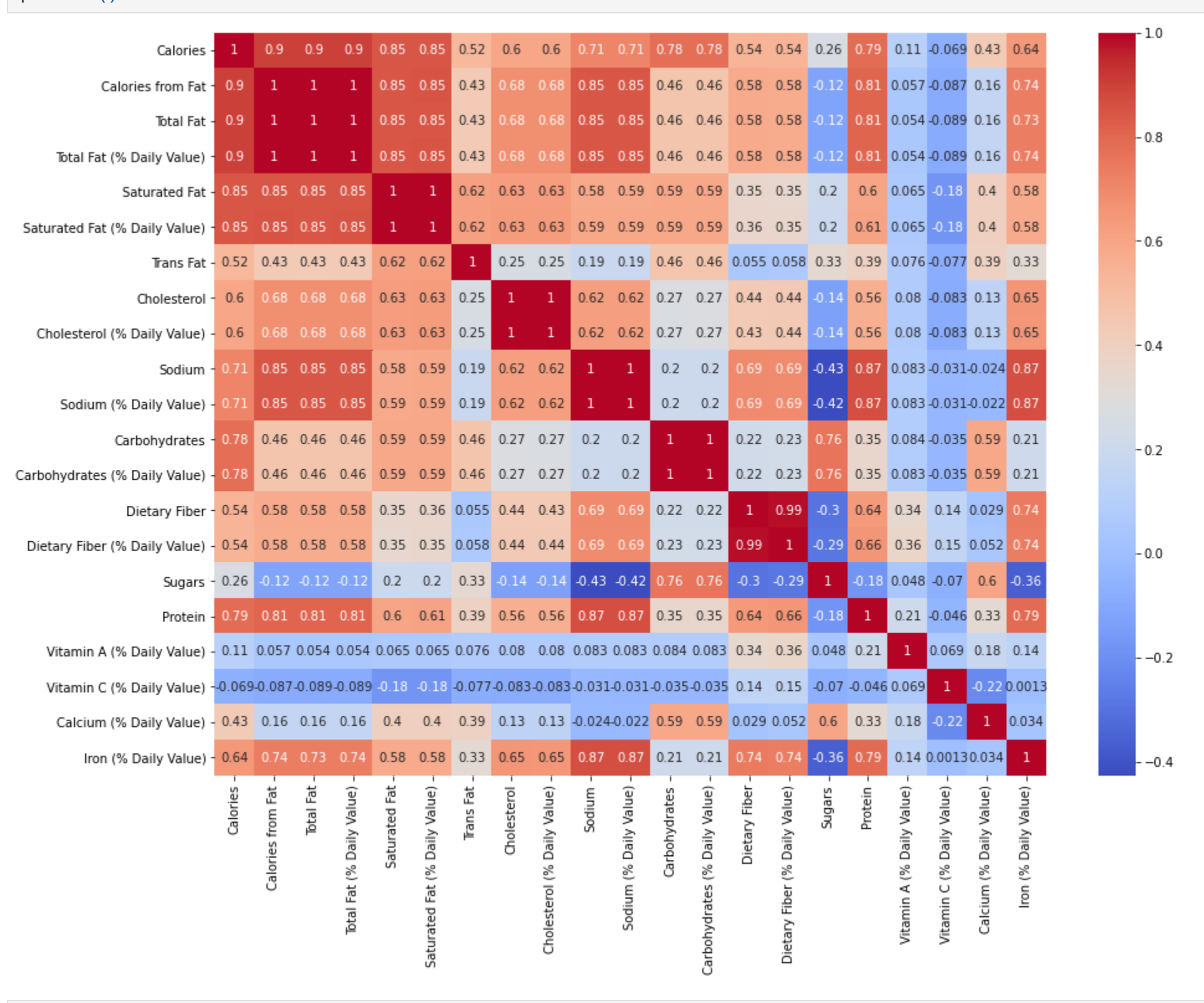
In [39]:
```python
correl = df.corr()
correl
```

Out[39]:

	Calories	Calories from Fat	Total Fat	Total Fat (% Daily Value)	Saturated Fat	Saturated Fat (% Daily Value)	Trans Fat	Cholesterol	Cholesterol (% Daily Value)	Sodium	...	Carbohydrates	Carbohydrates (% Daily Value)	Dietary Fiber	Dietary Fiber (% Daily Value)
Calories	1.000000	0.904588	0.904409	0.904123	0.845564	0.847631	0.522441	0.596399	0.595208	0.712309	...	0.781539	0.781242	0.538894	0.540014
Calories from Fat	0.904588	1.000000	0.999663	0.999725	0.847008	0.849592	0.433686	0.682161	0.681607	0.846624	...	0.461672	0.461463	0.581274	0.575621
Total Fat	0.904409	0.999663	1.000000	0.999765	0.846707	0.849293	0.431453	0.680547	0.680000	0.846158	...	0.461213	0.461005	0.580837	0.575206
Total Fat (% Daily Value)	0.904123	0.999725	0.999765	1.000000	0.847379	0.849973	0.433016	0.680940	0.680378	0.846728	...	0.460516	0.460298	0.580592	0.575033
Saturated Fat	0.845564	0.847008	0.846707	0.847379	1.000000	0.999279	0.620611	0.631210	0.630334	0.584075	...	0.591261	0.591743	0.351818	0.347152
Saturated Fat (% Daily Value)	0.847631	0.849592	0.849293	0.849973	0.999279	1.000000	0.620210	0.633603	0.632712	0.588694	...	0.591322	0.591655	0.356831	0.351797
Trans Fat	0.522441	0.433686	0.431453	0.433016	0.620611	0.620210	1.000000	0.253935	0.251502	0.187580	...	0.463250	0.462891	0.054918	0.058301
Cholesterol	0.596399	0.682161	0.680547	0.680940	0.631210	0.633603	0.253935	1.000000	0.999855	0.624362	...	0.270977	0.272662	0.435575	0.440266
Cholesterol (% Daily Value)	0.595208	0.681607	0.680000	0.680378	0.630334	0.632712	0.251502	0.999855	1.000000	0.623320	...	0.269300	0.270992	0.434940	0.439814
Sodium	0.712309	0.846624	0.846158	0.846728	0.584075	0.588694	0.187580	0.624362	0.623320	1.000000	...	0.200796	0.201032	0.694389	0.689995
Sodium (% Daily Value)	0.713415	0.847276	0.846780	0.847368	0.585323	0.589958	0.188339	0.624743	0.623720	0.999929	...	0.202426	0.202663	0.693913	0.689464
Carbohydrates	0.781539	0.461672	0.461213	0.460516	0.591261	0.591322	0.463250	0.270977	0.269300	0.200796	...	1.000000	0.999620	0.224577	0.228257
Carbohydrates (% Daily Value)	0.781242	0.461463	0.461005	0.460298	0.591743	0.591655	0.462891	0.272662	0.270992	0.201032	...	0.999620	1.000000	0.224058	0.227285
Dietary Fiber	0.538894	0.581274	0.580837	0.580592	0.351818	0.356831	0.054918	0.435575	0.434940	0.694389	...	0.224577	0.224058	1.000000	0.986350
Dietary Fiber (% Daily Value)	0.540014	0.575621	0.575206	0.575033	0.347152	0.351797	0.058301	0.440266	0.439814	0.689995	...	0.228257	0.227285	0.986350	1.000000
Sugars	0.259598	-0.115285	-0.115446	-0.115761	0.197734	0.195928	0.334756	-0.135518	-0.136459	-0.426536	...	0.762362	0.762282	-0.295178	-0.287014
Protein	0.787847	0.807913	0.807773	0.807922	0.603028	0.606581	0.388249	0.561561	0.560957	0.869802	...	0.352122	0.352178	0.641345	0.656648
Vitamin A (% Daily Value)	0.108844	0.056731	0.054434	0.054038	0.064972	0.065376	0.075833	0.080239	0.080059	0.083068	...	0.083802	0.083376	0.340518	0.361380
Vitamin C (% Daily Value)	-0.068747	-0.087331	-0.089354	-0.089353	-0.179672	-0.178059	-0.076612	-0.082978	-0.083315	-0.030769	...	-0.034724	-0.035450	0.141935	0.150019
Calcium (% Daily Value)	0.428426	0.161034	0.162860	0.162031	0.403311	0.401139	0.385331	0.132077	0.132382	-0.024074	...	0.589699	0.590263	0.028711	0.052359
Iron (% Daily Value)	0.643552	0.735894	0.734685	0.735478	0.578062	0.580488	0.325476	0.655000	0.653167	0.871593	...	0.210241	0.210643	0.740411	0.737814

21 rows × 21 columns

In [37]:
```python
plt.figure(figsize=(15,11))
sns.heatmap(correl, annot=True, cmap='coolwarm')
plt.show()
```



In [40]:
```python
# Highest correlation from each list:

#Calories and Saturated Fat (0.85)
#Total Fat and Sodium (0.85)
#Total Fat and Saturated Fat (0.85)
#Saturated Fat and Cholestrol (0.63)
#Cholestrol and Iron (0.65)
#Sodium and Iron (0.87)
#Sodium and Protien (0.87)
#Carbohydrates and Sugars (0.76)
#Dietary Fiber and Iron (0.74)
#Sugars and Calcium (0.6)
#Protien and Iron (0.79)
```

Which category contributes to the maximum % of Cholesterol in a diet (% daily value)?

In [43]:
```python
df['Category'].unique()
```

Out[43]:
```
array(['Breakfast', 'Beef & Pork', 'Chicken & Fish', 'Salads',
       'Snacks & Sides', 'Desserts', 'Beverages', 'Coffee & Tea',
       'Smoothies & Shakes'], dtype=object)
```

In [50]:
```python
chol = pd.pivot_table(df, 'Cholesterol (% Daily Value)', index=['Category'])
final = chol.sort_values(('Cholesterol (% Daily Value)'), ascending=False)
final
```

Out[50]:

Category	Cholesterol (% Daily Value)
Breakfast	50.952381
Beef & Pork	28.933333
Chicken & Fish	25.222222
Salads	17.333333
Smoothies & Shakes	14.714286
Coffee & Tea	9.378947
Snacks & Sides	6.230769
Desserts	4.857143
Beverages	0.185185

In [51]:
```python
# Breakfast category contributes to the maximum % of Cholesterol in a diet (% daily value)
```

Which item contributes maximum to the Sodium intake?

In [53]:
```python
sod = pd.pivot_table(df, 'Sodium', index=['Item'])
final2 = sod.sort_values(('Sodium'), ascending=False)
final2.head()
```

Out[53]:

Item	Sodium
Chicken McNuggets (40 piece)	3600
Big Breakfast with Hotcakes and Egg Whites (Large Biscuit)	2290
Big Breakfast with Hotcakes (Large Biscuit)	2260
Big Breakfast with Hotcakes and Egg Whites (Regular Biscuit)	2170
Big Breakfast with Hotcakes (Regular Biscuit)	2150

In [ ]:
```python
# Chicken McNuggets contributes maximum to the Sodium intake
```

Which 4 food items contain the most amount of Saturated Fat?

In [54]:
```python
sat = pd.pivot_table(df, 'Saturated Fat', index=['Item'])
final3 = sat.sort_values(('Saturated Fat'), ascending=False)
final3.head()
```

Out[54]:

Item	Saturated Fat
McFlurry with M&M’s Candies (Medium)	20.0
Big Breakfast with Hotcakes (Large Biscuit)	20.0
Chicken McNuggets (40 piece)	20.0
Frappé Chocolate Chip (Large)	20.0
Double Quarter Pounder with Cheese	19.0

In [ ]:
```python
# Most amount of Saturated Fat items are :-

# McFlurry with M&M’s Candies (Medium)
# Big Breakfast with Hotcakes (Large Biscuit)
# Chicken McNuggets (40 piece)
# Frappé Chocolate Chip (Large)
```