

```
In [18]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
In [2]: df=pd.read_csv('Nutrition_Physical_Activity_and_Obesity.csv')
```

```
In [3]: df
```

Out[3]:	YearStart	YearEnd	LocationAbbr	LocationDesc	Datasource	Class	Topic	Question	Data_Value_Unit	Data_Value_Type	...	GeoLocation	ClassID	TopicID	QuestionID	DataValueTypeID	LocationID	StratificationCategory1
0	2020	2020	US	National	Behavioral Risk Factor Surveillance System	Physical Activity	Physical Activity - Behavior	Percent of adults who engage in no leisure-tim...	NaN	Value	...	NaN	PA	PA1	Q047	VALUE	59	Race/Ethnicity
1	2014	2014	GU	Guam	Behavioral Risk Factor Surveillance System	Obesity / Weight Status	Obesity / Weight Status	Percent of adults aged 18 years and older who ...	NaN	Value	...	(13.444304, 144.793731)	OWS	OWS1	Q036	VALUE	66	Education
2	2013	2013	US	National	Behavioral Risk Factor Surveillance System	Obesity / Weight Status	Obesity / Weight Status	Percent of adults aged 18 years and older who ...	NaN	Value	...	NaN	OWS	OWS1	Q036	VALUE	59	Income
3	2013	2013	US	National	Behavioral Risk Factor Surveillance System	Obesity / Weight Status	Obesity / Weight Status	Percent of adults aged 18 years and older who ...	NaN	Value	...	NaN	OWS	OWS1	Q037	VALUE	59	Income
4	2015	2015	US	National	Behavioral Risk Factor Surveillance System	Physical Activity	Physical Activity - Behavior	Percent of adults who achieve at least 300 min...	NaN	Value	...	NaN	PA	PA1	Q045	VALUE	59	Income
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
88624	2021	2021	ND	North Dakota	Behavioral Risk Factor Surveillance System	Fruits and Vegetables	Fruits and Vegetables - Behavior	Percent of adults who report consuming vegetab...	NaN	Value	...	(47.47531977900047, -100.11842104899966)	FV	FV1	Q019	VALUE	38	Education
88625	2021	2021	PR	Puerto Rico	Behavioral Risk Factor Surveillance System	Physical Activity	Physical Activity - Behavior	Percent of adults who engage in no leisure-tim...	NaN	Value	...	(18.220833, -66.590149)	PA	PA1	Q047	VALUE	72	Race/Ethnicity
88626	2021	2021	WI	Wisconsin	Behavioral Risk Factor Surveillance System	Physical Activity	Physical Activity - Behavior	Percent of adults who engage in no leisure-tim...	NaN	Value	...	(44.39319117400049, -89.81637074199966)	PA	PA1	Q047	VALUE	55	Income
88627	2021	2021	UT	Utah	Behavioral Risk Factor Surveillance System	Fruits and Vegetables	Fruits and Vegetables - Behavior	Percent of adults who report consuming fruit L...	NaN	Value	...	(39.360700171000474, -111.58713063499971)	FV	FV1	Q018	VALUE	49	Education
88628	2021	2021	US	National	Behavioral Risk Factor Surveillance System	Physical Activity	Physical Activity - Behavior	Percent of adults who engage in no leisure-tim...	NaN	Value	...	NaN	PA	PA1	Q047	VALUE	59	Income

88629 rows × 33 columns

```
In [23]: df.describe()
```

Out[23]:	YearStart	YearEnd	Data_Value	Data_Value_Alt	Low_Confidence_Limit	High_Confidence_Limit	Sample_Size	LocationID
count	88629.000000	88629.000000	79851.000000	79851.000000	79851.000000	79851.000000	79851.000000	88629.000000
mean	2016.011362	2016.011362	31.236493	31.236493	26.907732	36.135953	3656.809044	30.906475
std	3.121080	3.121080	10.113829	10.113829	9.907896	11.061591	18706.392637	17.485456
min	2011.000000	2011.000000	0.900000	0.900000	0.300000	3.000000	50.000000	1.000000
25%	2013.000000	2013.000000	24.300000	24.300000	20.000000	28.600000	516.000000	17.000000
50%	2016.000000	2016.000000	31.200000	31.200000	26.800000	35.900000	1109.000000	30.000000
75%	2019.000000	2019.000000	37.000000	37.000000	32.900000	42.200000	2408.000000	45.000000
max	2021.000000	2021.000000	77.600000	77.600000	70.200000	87.700000	476876.000000	78.000000

```
In [24]: df.all()
```

Out[24]:	YearStart	True
	YearEnd	True
	LocationAbbr	True
	LocationDesc	True
	Datasource	True
	Class	True
	Topic	True
	Question	True
	Data_Value_Type	True
	Data_Value	True
	Data_Value_Alt	True
	Data_Value_Footnote_Symbol	True
	Data_Value_Footnote	True
	Low_Confidence_Limit	True
	High_Confidence_Limit	True
	Sample_Size	True
	Total	True
	Age(years)	True
	Education	True
	Gender	True
	Income	True
	Race/Ethnicity	True
	GeoLocation	True
	ClassID	True
	TopicID	True
	QuestionID	True
	DataValueTypeID	True
	LocationID	True
	StratificationCategory1	True
	Stratification1	True
	StratificationCategoryID1	True
	StratificationID1	True
	dtype: bool	

```
In [11]: df.duplicated()
```

Out[11]:	0	False
	1	False
	2	False
	3	False
	4	False
	...	True
	88624	False
	88625	False
	88626	False
	88627	False
	88628	False
	Length: 88629, dtype: bool	

```
In [25]: df.all()
```

Out[25]:	YearStart	True
	YearEnd	True
	LocationAbbr	True
	LocationDesc	True
	Datasource	True
	Class	True
	Topic	True
	Question	True
	Data_Value_Type	True
	Data_Value	True
	Data_Value_Alt	True
	Data_Value_Footnote_Symbol	True
	Data_Value_Footnote	True
	Low_Confidence_Limit	True
	High_Confidence_Limit	True
	Sample_Size	True
	Total	True
	Age(years)	True
	Education	True
	Gender	True
	Income	True
	Race/Ethnicity	True
	GeoLocation	True
	ClassID	True
	TopicID	True
	QuestionID	True
	DataValueTypeID	True
	LocationID	True
	StratificationCategory1	True
	Stratification1	True
	StratificationCategoryID1	True
	StratificationID1	True
	dtype: bool	

```
In [13]: df.drop_duplicates()
```

Out[13]:	YearStart	YearEnd	LocationAbbr	LocationDesc	Datasource	Class	Topic	Question	Data_Value_Unit	Data_Value_Type	...	GeoLocation	ClassID	TopicID	QuestionID	DataValueTypeID	LocationID	StratificationCategory1
0	2020	2020	US	National	Behavioral Risk Factor Surveillance System	Physical Activity	Physical Activity - Behavior	Percent of adults who engage in no leisure-tim...	NaN	Value	...	NaN	PA	PA1	Q047	VALUE	59	Race/Ethnicity
1	2014	2014	GU	Guam	Behavioral Risk Factor Surveillance System	Obesity / Weight Status	Obesity / Weight Status	Percent of adults aged 18 years and older who ...	NaN	Value	...	(13.444304, 144.793731)	OWS	OWS1	Q036	VALUE	66	Education
2	2013	2013	US	National	Behavioral Risk Factor Surveillance System	Obesity / Weight Status	Obesity / Weight Status	Percent of adults aged 18 years and older who ...	NaN	Value	...	NaN	OWS	OWS1	Q036	VALUE	59	Income
3	2013	2013	US	National	Behavioral Risk Factor Surveillance System	Obesity / Weight Status	Obesity / Weight Status	Percent of adults aged 18 years and older who ...	NaN	Value	...	NaN	OWS	OWS1	Q037	VALUE	59	Income
4	2015	2015	US	National	Behavioral Risk Factor Surveillance System	Physical Activity	Physical Activity - Behavior	Percent of adults who achieve at least 300 min...	NaN	Value	...	NaN	PA	PA1	Q045	VALUE	59	Income
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
88624	2021	2021	ND	North Dakota	Behavioral Risk Factor Surveillance System	Fruits and Vegetables	Fruits and Vegetables - Behavior	Percent of adults who report consuming vegetab...	NaN	Value	...	(47.47531977900047, -100.11842104899966)	FV	FV1	Q019	VALUE	38	Education
88625	2021	2021	PR	Puerto Rico	Behavioral Risk Factor Surveillance System	Physical Activity	Physical Activity - Behavior	Percent of adults who engage in no leisure-tim...	NaN	Value	...	(18.220833, -66.590149)	PA	PA1	Q047	VALUE	72	Race/Ethnicity
88626	2021	2021	WI	Wisconsin	Behavioral Risk Factor Surveillance System	Physical Activity	Physical Activity - Behavior	Percent of adults who engage in no leisure-tim...	NaN	Value	...	(44.39319117400049, -89.81637074199966)	PA	PA1	Q047	VALUE	55	Income
88627	2021	2021	UT	Utah	Behavioral Risk Factor Surveillance System	Fruits and Vegetables	Fruits and Vegetables - Behavior	Percent of adults who report consuming fruit L...	NaN	Value	...	(39.360700171000474, -111.58713063499971)	FV	FV1	Q018	VALUE	49	Education
88628	2021	2021	US	National	Behavioral Risk Factor Surveillance System	Physical Activity	Physical Activity - Behavior	Percent of adults who engage in no leisure-tim...	NaN	Value	...	NaN	PA	PA1	Q047	VALUE	59	Income

88629 rows × 33 columns

```
In [14]: df.isnull()
```

Out[14]:	YearStart	YearEnd	LocationAbbr	LocationDesc	Datasource	Class	Topic	Question	Data_Value_Unit	Data_Value_Type	...	GeoLocation	ClassID	TopicID	QuestionID	DataValueTypeID	LocationID	StratificationCategory1
0	False	False	False	False	False	False	False	False	True	False	...	True	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	True	False	...	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	True	False	...	True	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	True	False	...	True	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	True	False	...	True	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
88624	False	False	False	False	False	False	False	False	True	False	...	False	False	False	False	False	False	False
88625	False	False	False	False	False	False	False	False	True	False	...	False	False	False	False	False	False	False
88626	False	False	False	False	False	False	False	False	True	False	...	False	False	False	False	False	False	False
88627	False	False	False	False	False	False	False	False	True	False	...	False	False	False	False	False	False	False
88628	False	False	False	False	False	False	False	False	True	False	...	True	False	False	False	False	False	False

88629 rows × 33 columns

```
In [15]: del df['Data_Value_Unit']
```

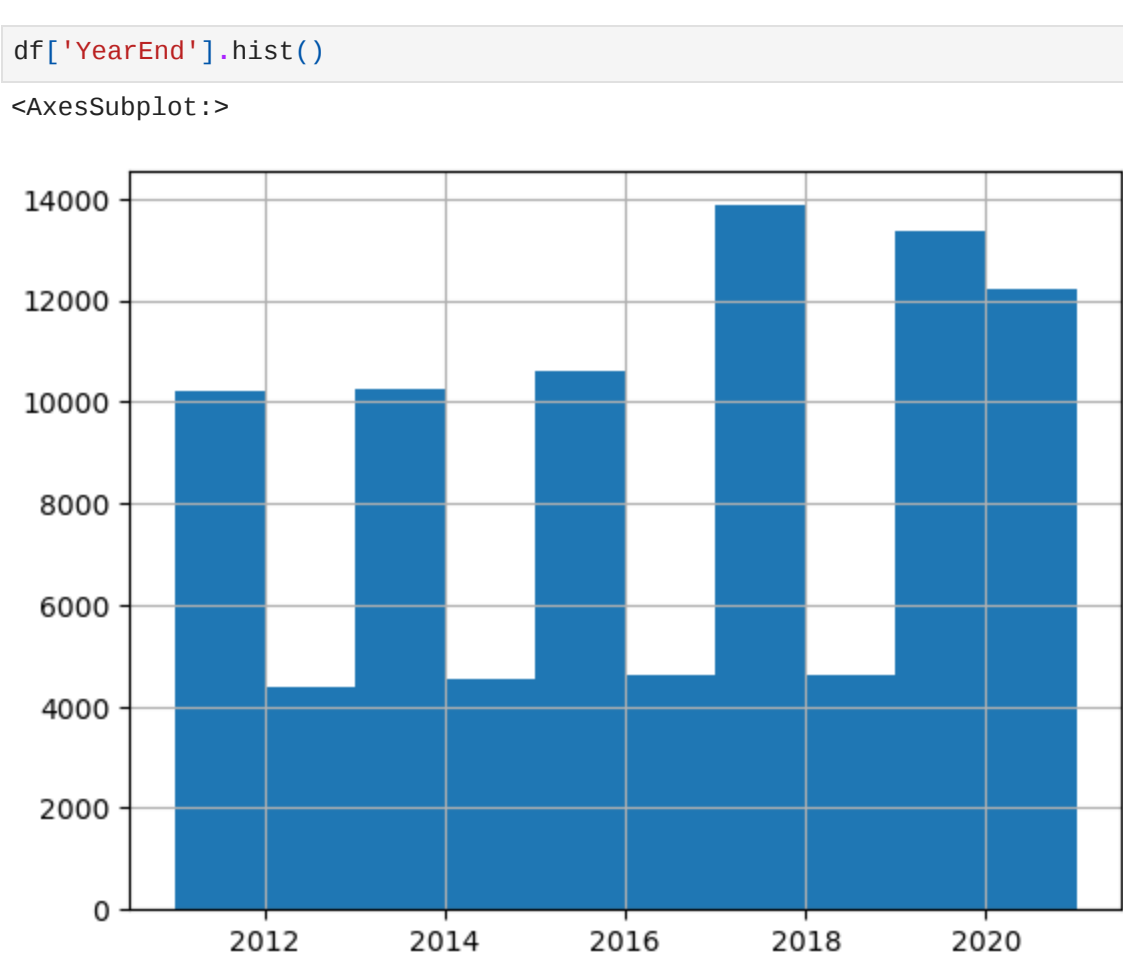
```
In [17]: df.head()
```

Out[17]:	YearStart	YearEnd	LocationAbbr	LocationDesc	Datasource	Class	Topic	Question	Data_Value_Type	Data_Value	...	GeoLocation	ClassID	TopicID	QuestionID	DataValueTypeID	LocationID	StratificationCategory1
0	2020	2020	US	National	Behavioral Risk Factor Surveillance System	Physical Activity	Physical Activity - Behavior	Percent of adults who engage in no leisure-tim...	Value	30.6	...	NaN	PA	PA1	Q047	VALUE	59	Race/Ethnicity
1	2014	2014	GU	Guam	Behavioral Risk Factor Surveillance System	Obesity / Weight Status	Obesity / Weight Status	Percent of adults aged 18 years and older who ...	Value	29.3	...	(13.444304, 144.793731)	OWS	OWS1	Q036	VALUE	66	Education
2	2013	2013	US	National	Behavioral Risk Factor Surveillance System	Obesity / Weight Status	Obesity / Weight Status	Percent of adults aged 18 years and older who ...	Value	28.8	...	NaN	OWS	OWS1	Q036	VALUE	59	Income
3	2013	2013	US	National	Behavioral Risk Factor Surveillance System	Obesity / Weight Status	Obesity / Weight Status	Percent of adults aged 18 years and older who ...	Value	32.7	...	NaN	OWS	OWS1	Q037	VALUE	59	Income
4	2015	2015	US	National	Behavioral Risk Factor Surveillance System	Physical Activity	Physical Activity - Behavior	Percent of adults who achieve at least 300 min...	Value	26.6	...	NaN	PA	PA1	Q045	VALUE	59	Income

5 rows × 32 columns

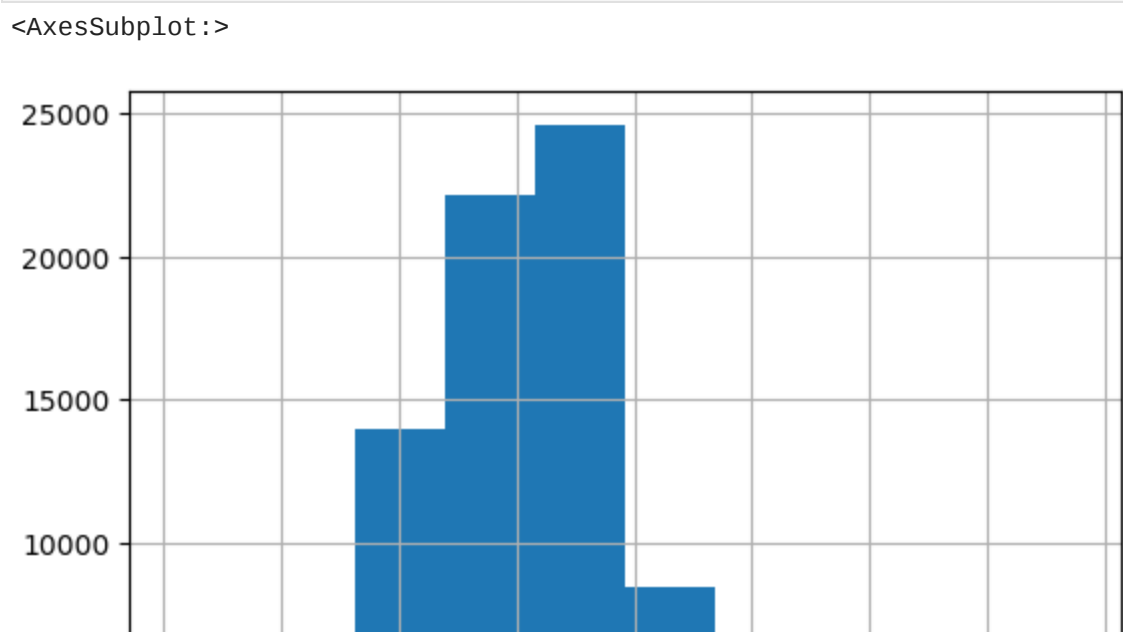
```
In [18]: df['YearEnd'].hist()
```

```
Out[18]: <AxesSubplot:~>
```



```
In [22]: df['Data_Value'].hist()
```

```
Out[22]: <AxesSubplot:~>
```



```
In [ ]:
```