```
import pandas as pd
```

DATASET 1 : DROPOUT RATIO

```
df = pd.read_csv('/content/dropout-ratio-2012-2015.csv')
```

```
df.head()
```

·imary_Total	Upper Primary_Boys	Upper Primary_Girls	Upper Primary_Total	Secondary _Boys	Secondary _Girls	Secondar _Tota
0.68	Uppe_r_Primary	1.09	1.23	5.57	5.55	5.5
1.21	NR	1.54	0.51	8.36	5.98	7
0.51	1.44	1.95	1.69	11.47	8.16	9.8
3.18	3.21	3.51	3.36	12.21	13.25	12.7
4.35	3.46	4.12	3.78	11.95	13.37	12.6

```
df.isna()
```

Primary_Total	Upper Primary_Boys	Upper Primary_Girls	Upper Primary_Total	Secondary _Boys	Secondary _Girls	Secondar _Tota

```
df.isna().any()
```

```
State_UT              False
year                  False
Primary_Boys          False
Primary_Girls         False
Primary_Total         False
Upper Primary_Boys    False
Upper Primary_Girls   False
Upper Primary_Total   False
Secondary _Boys       False
Secondary _Girls      False
Secondary _Total      False
HrSecondary_Boys      False
HrSecondary_Girls     False
HrSecondary_Total     False
dtype: bool
```

```
drops = df[df['State_UT']=='West Bengal']
```

```
drops.head()
```

Primary_Total	Upper Primary_Boys	Upper Primary_Girls	Upper Primary_Total	Secondary _Boys	Secondary _Girls	Secondar _Tota
6.3	6.29	4.16	5.18	14.95	19.41	17
2.91	5.63	3.1	4.31	16.73	19.77	18.3
					19.06	17

Automatic saving failed. This file was updated remotely or in another tab. Show diff

```
df.groupby('State_UT').groups
```

```
radesh': [98, 99, 100], 'Uttarakhand': [101, 102, 103], 'West Bengal': [104, 105, 106]}
```

```
df['Primary_Boys'].values
```

```
array(['0.83', '1.35', '0.47', '3.3', '4.31', '6.57', '11.54', '15.84',
       '11.51', '7.02', '8.19', '16.07', 'NR', '2.38', '0.35', 'NR', 'NR',
       'NR', '4.24', '1.45', '3.08', 'NR', '1.05', '1.6', 'NR', '1.06',
       '1.8', 'NR', 'NR', 'NR', 'NR', '0.08', '0.63', '0.21', '0.5',
       '0.82', '1.48', '0.22', '5.54', '0.51', '0.57', '0.46', '6.8',
       '5.53', '6.98', '7.36', '6.89', '5.91', '3.4', '2.42', '2.03',
       'NR', 'NR', 'NR', '2.3', '0', 'NR', '5.75', '9.91', '6.48', '0.88',
```

```
       '0.51', '1.26', '10.24', '17.27', '9.5', '11.32', '11.3', '10.35',
       '24.27', '12.57', '10.17', '7.11', '19.09', '6.18', '3.63', '2.83',
       '2.91', '0.25', '0.76', '0.36', '1.99', '1.35', '2.89', '7.2',
       '7.76', '5.02', '4.78', '5.55', '3.75', '4.02', '0.53', 'NR',
       '6.04', '2.21', '2.31', '3.63', '1.37', '10.53', '7.91', '9.08',
       '1.14', '3.28', '4.37', '6.88', '3.44', '2.13', '4.68', '4.53',
       '4.36'], dtype=object)
```

```
df = df.replace('NR',0)
```

```
kerala = df[df['State_UT']=='Kerala']
```

```
kerala.head()
```

	State_UT	year	Primary_Boys	Primary_Girls	Primary_Total	Upper Primary_Boys	Primary_
51	Kerala	2012-13	0	0	0	0	
52	Kerala	2013-14	0	0	0	0	
53	Kerala	2014-15	0	0	0	0	

```
df['Primary_Boys']=df['Primary_Boys'].apply(lambda x :float(x))
```

```
df['Primary_Boys'].head()
```

```
0    0.83
```



```
4    4.31
Name: Primary_Boys, dtype: float64
```

```
df['Primary_Girls'] = df['Primary_Girls'].apply(lambda x :float(x))
```

```
df['HrSecondary_Boys'].values
```

```
array(['17.66', '18.94', '21.05', '2.66', '12.65', 0, '18.57', '7.85',
       '19.37', '4.87', '7.62', 0, 0, 0, 0, '16.32', '13.24', '12.1', 0,
       0, '1.37', '13.34', '7.07', '13.45', '6.21', '14.48', '44.38',
       '18.56', '20.28', '20.91', '16.01', '18.21', '18.27', '0.58',
       '9.06', '8.4', 0, '1.41', '6.24', '10.25', '8.44', '9.02', '11.69',
       '8.8', '13.85', 0, 0, '2.72', '19.47', 0, '5.97', 0, '6.95',
       '1.54', '2.98', '3.03', '3', 0, '0.52', 0, '2.55', '2.85', '2.02',
       '5.96', '3.3', 0, 0, 0, 0, '2.91', 0, '9', '18.67', '15.36',
```

```
       '10.36', 0, 0, 0, '6.79', '13.24', '5.4', '9.69', '7.87', '7.52',
       0, 0, 0, '11.79', '14.11', '12.48', '1.34', '4.55', '4.59',
       '13.67', '2.06', '8.4', '9.15', '8.97', 0, 0, '1.79', '1.35',
       '0.23', '3.4', '7.81', '8.03', '8.18', 0, '1.48', '0.25'],
      dtype=object)
```

```
df['Upper Primary_Boys'].values
```

```
array(['Uppe_r_Primary', 0, '1.44', '3.21', '3.46', '5.09', '4.44',
       '5.86', '5.31', '7.89', '7.6', '10.45', 0, '2.77', '4.14', 0,
       '0.72', '0.01', '6.09', '4.09', '6.47', '2.59', '3.31', '3.7', 0,
       '3.42', '3.14', 0, '3.13', '0.95', 0, 0, 0, '2.75', '3.52', '4.65',
       '0.18', '1.97', '5.5', '0.52', '0.6', '0.5', '5.51', '3.86',
       '4.98', '4.99', '7.19', '9.01', '4.96', '2.31', '3.46', 0, 0, 0,
       '0.97', '1.16', '2.37', '6.79', '9.88', '7.78', '0.89', 0, '0.89',
       '5.48', '7.48', '3.61', '8.43', '6.34', '6.77', '19.35', '6.61',
       '5.46', '10.15', '18.08', '7.87', '4.1', '3.11', '4.11', '0.33',
       '0.37', '0.44', '2.58', '2.52', '2.95', '2.86', '4.49', '2.54',
       '2.6', '6.35', '2.07', '0.38', '4.38', 0, '4.63', '2.43', '3.1',
       '3.21', '2.37', 0, 0, '0.78', 0, '1.78', '0.79', '6.29', '5.63',
       '5.84', '2.3', '3.09', '3.49'], dtype=object)
```

```
df['Upper Primary_Boys'] =df['Upper Primary_Boys'].replace('Uppe_r_Primary',0)
```

```
df['Upper Primary_Boys'].values
```

```
array([0, 0, '1.44', '3.21', '3.46', '5.09', '4.44', '5.86', '5.31',
       '7.89', '7.6', '10.45', 0, '2.77', '4.14', 0, '0.72', '0.01',
       '6.09', '4.09', '6.47', '2.59', '3.31', '3.7', 0, '3.42', '3.14',
       0, '3.13', '0.95', 0, 0, 0, '2.75', '3.52', '4.65', '0.18', '1.97',
       '5.5', '0.52', '0.6', '0.5', '5.51', '3.86', '4.98', '4.99',
       '7.19', '9.01', '4.96', '2.31', '3.46', 0, 0, 0, '0.97', '1.16',
       '2.37', '6.79', '9.88', '7.78', '0.89', 0, '0.89', '5.48', '7.48',
       '3.61', '8.43', '6.34', '6.77', '19.35', '6.61', '5.46', '10.15',
       '18.08', '7.87', '4.1', '3.11', '4.11', '0.33', '0.37', '0.44',
       '2.58', '2.52', '2.95', '2.86', '4.49', '2.54', '2.6', '6.35',
       '2.07', '0.38', '4.38', 0, '4.63', '2.43', '3.1', '3.21', '2.37',
       0, 0, '0.78', 0, '1.78', '0.79', '6.29', '5.63', '5.84', '2.3',
       '3.09', '3.49'], dtype=object)
```



```
import matplotlib.pyplot as plt
import seaborn as sns
```

```
df.head()
```

Primary_Total	Upper Primary_Boys	Upper Primary_Girls	Upper Primary_Total	Secondary _Boys	Secondary _Girls	Secondar _Tota
0.68	0	1.09	1.23	5.57	5.55	5.5
1.21	0	1.54	0.51	8.36	5.98	7
0.51	1.44	1.95	1.69	11.47	8.16	9.8

```
df.describe(include = 'all')
```

Primary_Total	Upper Primary_Boys	Upper Primary_Girls	Upper Primary_Total	Secondary _Boys	Secondary _Girls	Secondar _Tota
110.0	110.0	110.0	110.0	110.0	110.0	110
87.0	90.0	97.0	96.0	106.0	107.0	106
0.0	0.0	0.0	0.0	0.0	0.0	0
17.0	17.0	8.0	9.0	4.0	4.0	4
NaN	NaN	NaN	NaN	NaN	NaN	Na
NaN	NaN	NaN	NaN	NaN	NaN	Na
NaN	NaN	NaN	NaN	NaN	NaN	Na
NaN	NaN	NaN	NaN	NaN	NaN	Na
NaN	NaN	NaN	NaN	NaN	NaN	Na
NaN	NaN	NaN	NaN	NaN	NaN	Na
NaN	NaN	NaN	NaN	NaN	NaN	Na

Automatic saving failed. This file was updated remotely or in another tab. Show diff

```
df.info
```

```
<bound method DataFrame.info of           State_UT     year  ...  HrSecondary_Girls  H
0     A & N Islands  2012-13  ...              14.14
1     A & N Islands  2013-14  ...              15.87
2     A & N Islands  2014-15  ...              16.93
3    Andhra Pradesh  2012-13  ...               0.35
4    Andhra Pradesh  2013-14  ...              11.79
..              ...      ...  ...                ...
105     West Bengal  2013-14  ...                7.9
106     West Bengal  2014-15  ...               8.11
107       All India  2012-13  ...                  0
108       All India  2013-14  ...               1.54
109       All India  2014-15  ...                  0

[110 rows x 14 columns]>
```

```
yearkerala = kerala['year'].values
```

```
print(yearkerala)
```

```
['2012-13' '2013-14' '2014-15']
```

```
import matplotlib.pyplot as plt
```

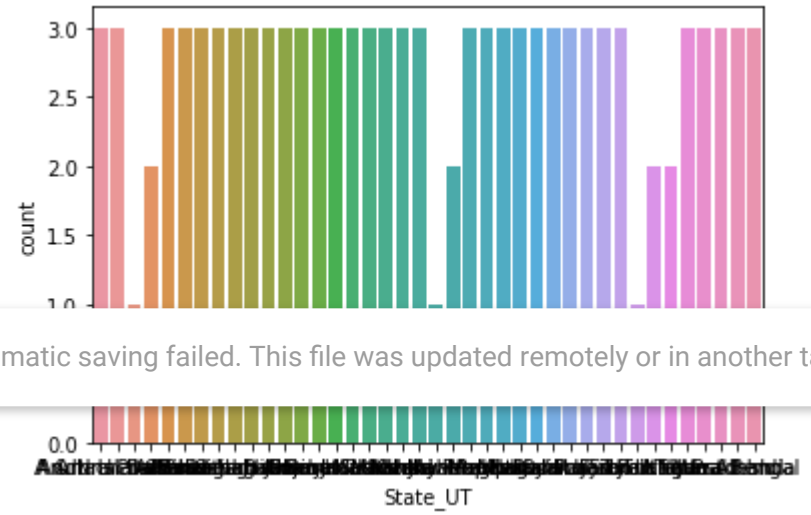
```
import seaborn as sns
```

```
plt.figure(figsize=(35,15))
```

```
<Figure size 2520x1080 with 0 Axes>
<Figure size 2520x1080 with 0 Axes>
```

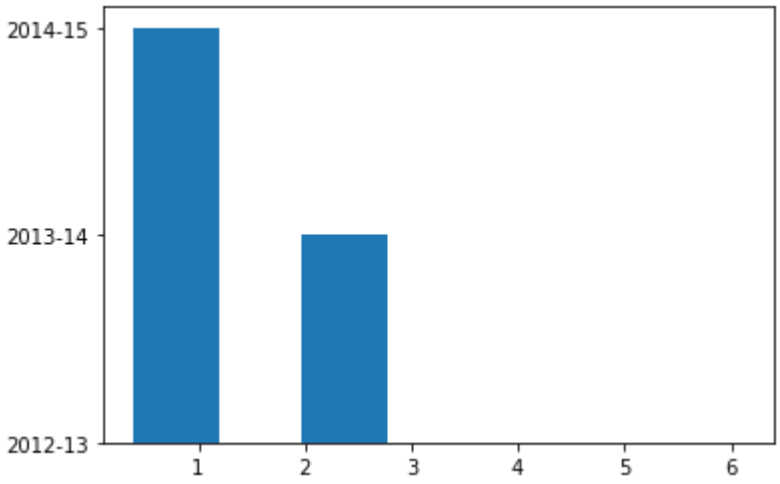
```
sns.countplot(df['State_UT'])
```

```
/usr/local/lib/python3.6/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass t
  FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7f9e3c2ce940>
```

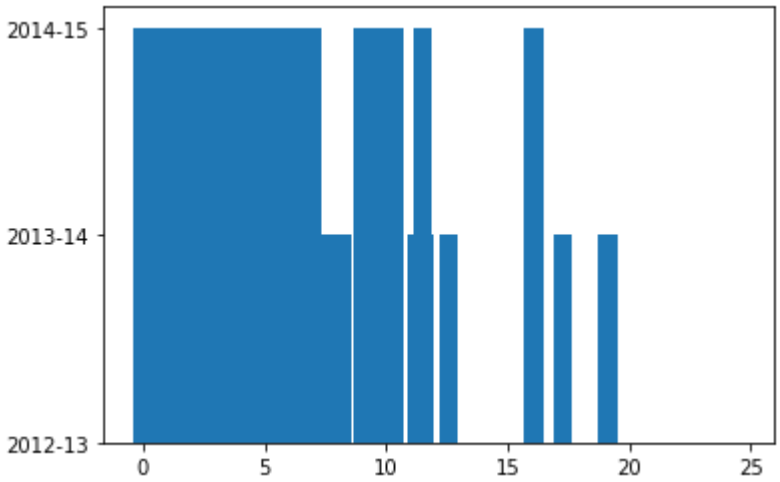



```
#DROP OUT RATE OF PRIMARY GIRLS IN TAMILNADU VS YEARS
wb = df[df['State_UT']=='West Bengal']
```

```
plt.bar(wb['Primary_Girls'],wb['year'])
plt.show()
```




```
plt.bar(df['Primary_Boys'],df['year'])
plt.show()
```




```
df['HrSecondary_Boys'].fillna(0)
```

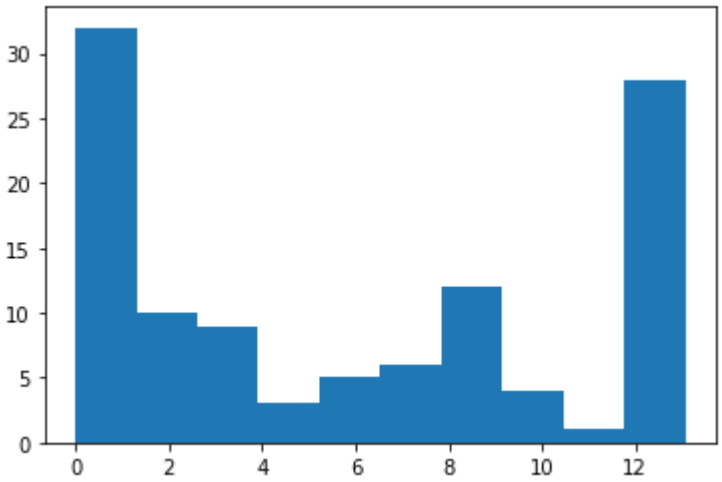


```
2      21.05
3       2.66
4      12.65
       ...
105     8.03
106     8.18
107        0
108     1.48
109     0.25
Name: HrSecondary_Boys, Length: 110, dtype: object
```

```
import numpy as np
df['HrSecondary_Boys'] = df['HrSecondary_Boys'].values.astype(np.float32)
df['HrSecondary_Girls'] = df['HrSecondary_Girls'].values.astype(np.float32)
df['Primary_Boys'] =df['Primary_Boys'].values.astype(np.float32)
df['Primary_Girls'] = df['Primary_Girls'].values.astype(np.float32)
```

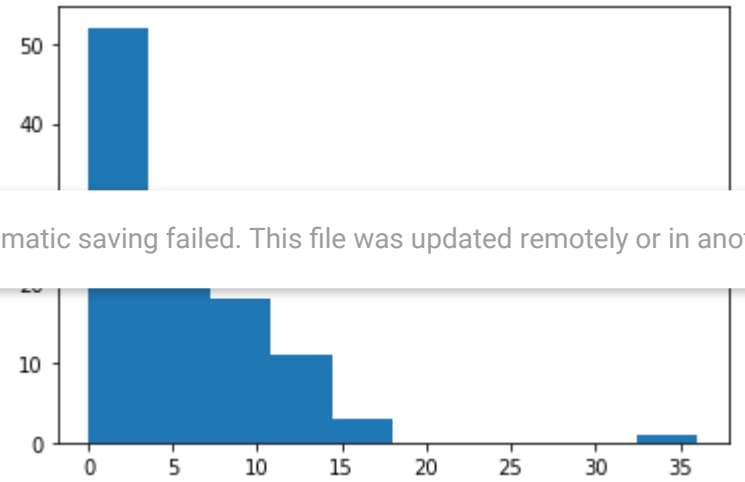
```
plt.hist(df['HrSecondary_Boys'])
```

```
(array([32., 10.,  9.,  3.,  5.,  6., 12.,  4.,  1., 28.]),
 array([ 0.     ,  1.30645,  2.6129 ,  3.91935,  5.2258 ,  6.53225,
         7.8387 ,  9.14515, 10.4516 , 11.75805, 13.0645 ], dtype=float32),
 <a list of 10 Patch objects>)
```




```
plt.hist(df['HrSecondary_Girls'])
```

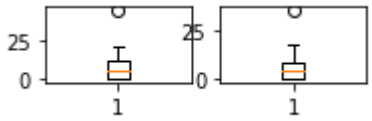
```
(array([52., 25., 18., 11.,  3.,  0.,  0.,  0.,  0.,  1.]),
 array([ 0.      ,  3.605   ,  7.21    , 10.815   , 14.42    , 18.025   ,
        21.63    , 25.234999, 28.84    , 32.445   , 36.05    ],
       dtype=float32),
 <a list of 10 Patch objects>)
```




```
plt.figsize=(20,15)
```

```
plt.subplot(5,4,1)
plt.boxplot(df['HrSecondary_Boys'])
plt.subplot(5,4,2)
plt.boxplot(df['HrSecondary_Girls'])
```

```
{'boxes': [<matplotlib.lines.Line2D at 0x7f9e3b7b70b8>],
 'caps': [<matplotlib.lines.Line2D at 0x7f9e3b7b7ac8>,
  <matplotlib.lines.Line2D at 0x7f9e3b7b7e48>],
 'fliers': [<matplotlib.lines.Line2D at 0x7f9e3b7c0588>],
 'means': [],
 'medians': [<matplotlib.lines.Line2D at 0x7f9e3b7c0208>],
 'whiskers': [<matplotlib.lines.Line2D at 0x7f9e3b7b73c8>,
  <matplotlib.lines.Line2D at 0x7f9e3b7b7748>]}
```




```
df['HrSecondary_Boys']=df['HrSecondary_Boys'].clip(lower=df['HrSecondary_Boys'].quantile(0.05
```

```
'HrSecondary_Girls']=df['HrSecondary_Girls'].clip(lower=df['HrSecondary_Girls'].quantile(0.05
```

```
df['Primary_Girls'] =df['Primary_Girls'].clip(lower=df['Primary_Girls'].quantile(0.05),upper=
```

```
df['Primary_Boys'] = df['Primary_Boys'].clip(lower=df['Primary_Boys'].quantile(0.05),upper=df
```

```
plt.figure(figsize=(90,45))
```

```
<Figure size 6480x3240 with 0 Axes>
<Figure size 6480x3240 with 0 Axes>
```

```
plt.subplot(4,4,1)
sns.boxplot(df['HrSecondary_Boys'])
plt.subplot(4,4,2)
sns.boxplot(df['HrSecondary_Girls'])
plt.subplot(4,4,3)
sns.boxplot(df['Primary_Girls'])
```

```
/usr/local/lib/python3.6/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass t
  FutureWarning
```

```
sns.pairplot(df)
```

```
<string>:6: RuntimeWarning: Converting input from bool to <class 'numpy.uint8'> for com
<string>:6: RuntimeWarning: Converting input from bool to <class 'numpy.uint8'> for com
<seaborn.axisgrid.PairGrid at 0x7f9e3b5ad080>
```

dataset 2 :analysing schools with boys toilet

```
dfs = pd.read_csv('/content/schools-with-boys-toilet-2013-2016.csv')
```

```
dfs.head()
```

ec_HrSec	U_Primary_Only	U_Primary_With_Sec_HrSec	Primary_with_U_Primary_Sec	U_Primary
100.00	0.00	100.00	100.00	
100.00	100.00	100.00	100.00	
100.00	0.00	100.00	100.00	
82.05	45.45	64.11	76.24	
96.00	75.00	93.33	91.43	

```
dfs.groupby('State_UT').groups
```

Automatic saving failed. This file was updated remotely or in another tab. Show diff

```
veli': [21, 22, 2
```

Primary_Boys Primary_Girls HrSecondary_Boys HrSecondary_Girls tamil

```
dfs.isnull()
```

	State_UT	year	Primary_Only	Primary_with_U_Primary	Primary_with_U_Primary_Sec_
0	False	False	False	False	
1	False	False	False	False	
2	False	False	False	False	
3	False	False	False	False	
4	False	False	False	False	
...	...	...	...	...	

```
dfs.isnull().any
```

```
<bound method DataFrame.any of      State_UT   year  Primary_Only  ...  Sec_with_HrSec.
0       False  False         False  ...            False       False        False
1       False  False         False  ...            False       False        False
2       False  False         False  ...            False       False        False
3       False  False         False  ...            False       False        False
4       False  False         False  ...            False       False        False
..        ...    ...           ...  ...              ...         ...          ...
105     False  False         False  ...            False       False        False
106     False  False         False  ...            False       False        False
107     False  False         False  ...            False       False        False
108     False  False         False  ...            False       False        False
109     False  False         False  ...            False       False        False

[110 rows x 13 columns]>
```

```
dfs.describe
```

```
<bound method NDFrame.describe of                     State_UT     year  ...  HrSec_O
0    Andaman & Nicobar Islands  2013-14  ...         0.00        94.52
```



```
4              Andhra Pradesh  2014-15  ...        86.54        65.34
..                        ...      ...  ...          ...          ...
105               West Bengal  2014-15  ...       100.00        89.93
106               West Bengal  2015-16  ...       100.00        97.70
107                 All India  2013-14  ...        74.36        86.56
108                 All India  2014-15  ...        88.97        88.62
109                 All India  2015-16  ...        95.67        97.02

[110 rows x 13 columns]>
```

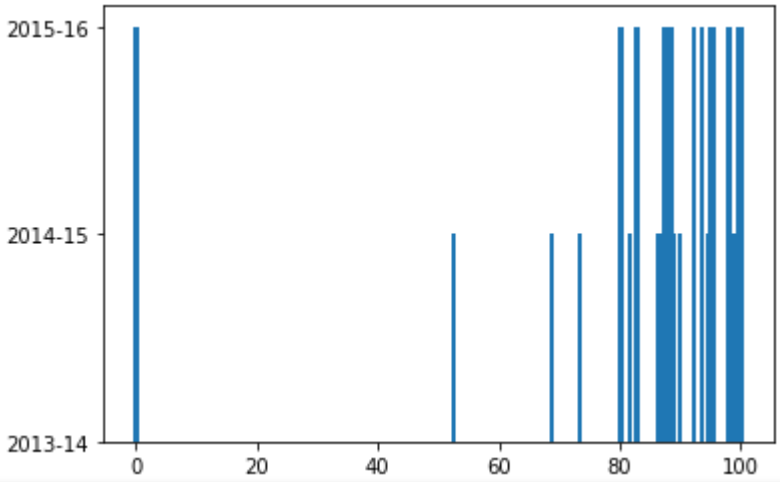
```
dfs.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 110 entries, 0 to 109
```

```
Data columns (total 13 columns):
 #   Column                            Non-Null Count  Dtype  
---  ------                            --------------  -----  
 0   State_UT                          110 non-null    object 
 1   year                              110 non-null    object 
 2   Primary_Only                      110 non-null    float64
 3   Primary_with_U_Primary            110 non-null    float64
 4   Primary_with_U_Primary_Sec_HrSec  110 non-null    float64
 5   U_Primary_Only                    110 non-null    float64
 6   U_Primary_With_Sec_HrSec          110 non-null    float64
 7   Primary_with_U_Primary_Sec        110 non-null    float64
 8   U_Primary_With_Sec                110 non-null    float64
 9   Sec_Only                          110 non-null    float64
 10  Sec_with_HrSec.                   110 non-null    float64
 11  HrSec_Only                        110 non-null    float64
 12  All Schools                       110 non-null    float64
dtypes: float64(11), object(2)
memory usage: 11.3+ KB
```

```
plt.bar(dfs['HrSec_Only'],dfs['year'])
```

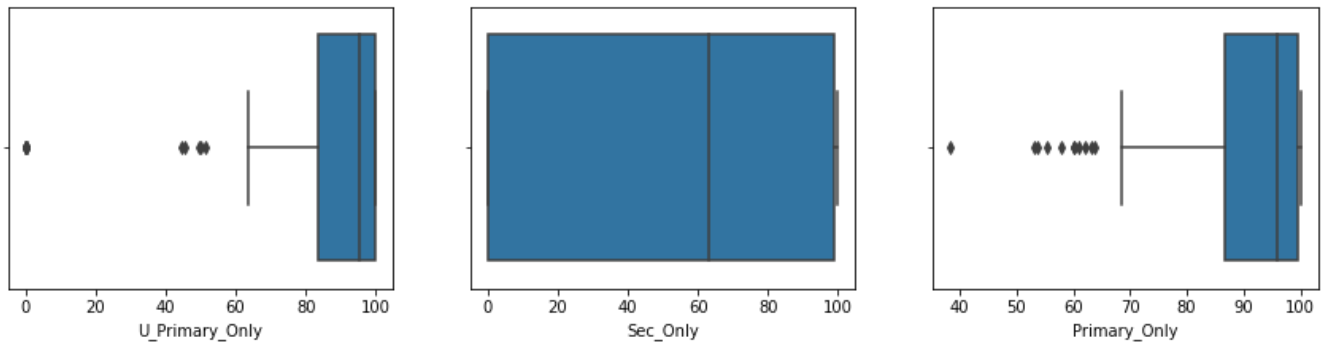
```
<BarContainer object of 110 artists>
```




```
dfs['Primary_with_U_Primary'] =dfs['Primary_with_U_Primary'].values.astype(np.float32)
dfs['Primary_with_U_Primary_Sec_HrSec'] =dfs['Primary_with_U_Primary_Sec_HrSec'].values.astyp
dfs['U_Primary_With_Sec_HrSec'] =dfs['U_Primary_With_Sec_HrSec'].values.astype(np.float32)
dfs['Sec_Only'] =dfs['Sec_Only'].values.astype(np.float32)
dfs['HrSec_Only'] =dfs['HrSec_Only'].values.astype(np.float32)
```

```
plt.figure(figsize=(20,15))
plt.subplot(4,4,1)
sns.boxplot(dfs['U_Primary_Only'])
plt.subplot(4,4,2)
sns.boxplot(dfs['Sec_Only'])
plt.subplot(4,4,3)
sns.boxplot(dfs['Primary_Only'])
```

```
/usr/local/lib/python3.6/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass t
  FutureWarning
/usr/local/lib/python3.6/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass t
  FutureWarning
/usr/local/lib/python3.6/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass t
  FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7f9e38068e80>
```

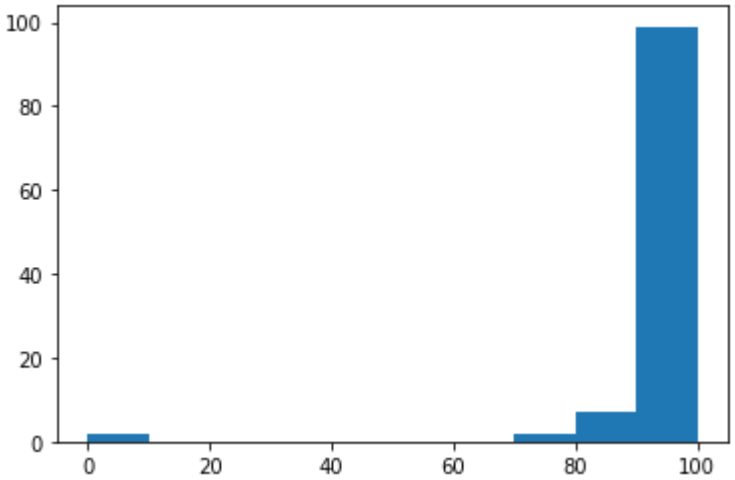



```
plt.hist(dfs['Primary_with_U_Primary'])
```

```
      (array([ 3.,  0.,  3.,  4.,  4.,  1.,  6.,  5., 11., 73.]),
       array([ 61.17    ,  65.053   ,  68.936   ,  72.819   ,  76.701996,
               80.585   ,  84.468   ,  88.351   ,  92.234   ,  96.117   ,
```

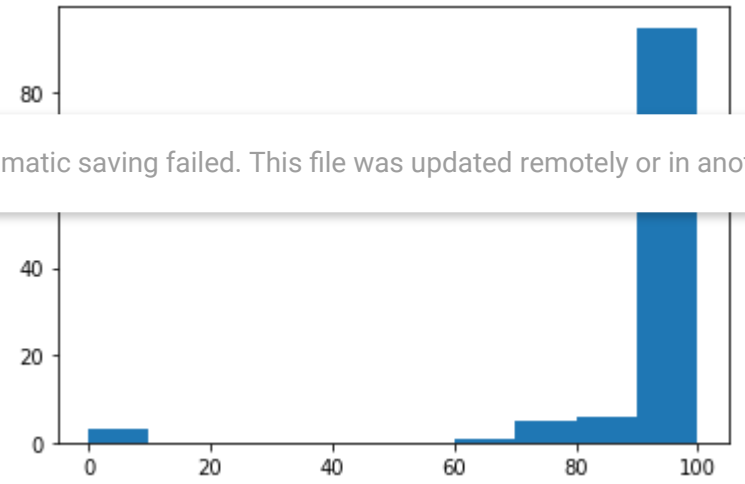
```
plt.hist(dfs['Primary_with_U_Primary_Sec_HrSec'])
```

```
      (array([ 2.,  0.,  0.,  0.,  0.,  0.,  0.,  2.,  7., 99.]),
       array([  0.,  10.,  20.,  30.,  40.,  50.,  60.,  70.,  80.,  90., 100.],
             dtype=float32),
       <a list of 10 Patch objects>)
```




```
plt.hist(dfs['U_Primary_With_Sec_HrSec'])
```

```
      (array([ 3.,  0.,  0.,  0.,  0.,  0.,  1.,  5.,  6., 95.]),
       array([  0.,  10.,  20.,  30.,  40.,  50.,  60.,  70.,  80.,  90., 100.],
             dtype=float32),
       <a list of 10 Patch objects>)
```



Automatic saving failed. This file was updated remotely or in another tab. Show diff