In [1]:
```python
%matplotlib inline

from numpy import arange
from matplotlib import pyplot as plt
from scipy.stats import norm
import pandas as pd

plt.rcParams['figure.figsize'] = [16, 7]
```

In [9]:
```python
df=pd.read_csv("TV_Shows.csv")
```

In [10]:
```python
df.head()
```

Out[10]:

	Unnamed: 0	Title	Year	Age	IMDb	Rotten Tomatoes	Netflix	Hulu	Prime Video	Disney+	type
0	0	Breaking Bad	2008	18+	9.5	96%	1	0	0	0	1
1	1	Stranger Things	2016	16+	8.8	93%	1	0	0	0	1
2	2	Money Heist	2017	18+	8.4	91%	1	0	0	0	1
3	3	Sherlock	2010	16+	9.1	78%	1	0	0	0	1
4	4	Better Call Saul	2015	18+	8.7	97%	1	0	0	0	1

In [11]:
```python
df.dtypes
```

Out[11]:
```
Unnamed: 0            int64
Title                object
Year                  int64
Age                  object
IMDb                float64
Rotten Tomatoes      object
Netflix               int64
Hulu                  int64
Prime Video           int64
Disney+               int64
type                  int64
dtype: object
```

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5611 entries, 0 to 5610
Data columns (total 11 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   Unnamed: 0       5611 non-null   int64
 1   Title            5611 non-null   object
 2   Year             5611 non-null   int64
 3   Age              3165 non-null   object
 4   IMDb             4450 non-null   float64
 5   Rotten Tomatoes  1010 non-null   object
 6   Netflix          5611 non-null   int64
 7   Hulu             5611 non-null   int64
 8   Prime Video      5611 non-null   int64
 9   Disney+          5611 non-null   int64
 10  type             5611 non-null   int64
dtypes: float64(1), int64(7), object(3)
memory usage: 482.3+ KB
```

In [13]:
```python
df.memory_usage()
```

Out[13]:
```
Index               128
Unnamed: 0        44888
Title             44888
Year              44888
Age               44888
IMDb              44888
Rotten Tomatoes   44888
Netflix           44888
Hulu              44888
Prime Video       44888
Disney+           44888
type              44888
dtype: int64
```

In [14]:
```python
df.memory_usage().sum()
```

Out[14]: 493896

In [15]:
```python
df.describe()
```

Out[15]:

	Unnamed: 0	Year	IMDb	Netflix	Hulu	Prime Video	Disney+
count	5611.000000	5611.000000	4450.000000	5611.000000	5611.000000	5611.000000	5611.000000
mean	2805.000000	2011.021030	7.113258	0.344145	0.312600	0.382107	0.032080
std	1619.900511	11.005116	1.132060	0.475131	0.463594	0.485946	0.176228
min	0.000000	1901.000000	1.000000	0.000000	0.000000	0.000000	0.000000
25%	1402.500000	2010.000000	6.600000	0.000000	0.000000	0.000000	0.000000
50%	2805.000000	2015.000000	7.300000	0.000000	0.000000	0.000000	0.000000
75%	4207.500000	2017.000000	7.900000	1.000000	1.000000	1.000000	0.000000
max	5610.000000	2020.000000	9.600000	1.000000	1.000000	1.000000	1.000000

Satistical Moments: 1.Mean (1st moment) 2.Variance (2nd moment) 3.Skewness (3rd moment) 4.Kurtosis (4th moment)

In [16]:
```python
df.mean()
```

Out[16]:
```
Unnamed: 0     2805.000000
Year           2011.021030
IMDb              7.113258
Netflix           0.344145
Hulu              0.312600
Prime Video       0.382107
Disney+           0.032080
type              1.000000
dtype: float64
```

In [18]:
```python
df['Netflix'].mean()
```

Out[18]: 0.344145428622349

In [21]:
```python
df.var()
```

Out[21]:
```
Unnamed: 0     2.624078e+06
Year           1.211126e+02
IMDb           1.281559e+00
Netflix        2.257496e-01
Hulu           2.149196e-01
Prime Video    2.361432e-01
Disney+        3.105626e-02
type           0.000000e+00
dtype: float64
```

skewness

df.skew()

Kurtosis

In [23]:
```python
df.kurtosis()
```

Out[23]:
```
Unnamed: 0     -1.200000
Year           15.669428
IMDb            2.130811
Netflix        -1.569853
Hulu           -1.346399
Prime Video    -1.765030
Disney+        26.229802
type            0.000000
dtype: float64
```

min/max/median

In [24]:
```python
df.min()
```

Out[24]:
```
Unnamed: 0                    0
Title          #MeToo, Now What?
Year                       1901
IMDb                          1
Netflix                       0
Hulu                          0
Prime Video                   0
Disney+                       0
type                          1
dtype: object
```

In [25]:
```python
df.max()
```

Out[25]:
```
Unnamed: 0                5610
Title          頭文字D First Stage
Year                      2020
IMDb                       9.6
Netflix                      1
Hulu                         1
Prime Video                  1
Disney+                      1
type                         1
dtype: object
```

In [26]:
```python
df.median()
```

Out[26]:
```
Unnamed: 0     2805.0
Year           2015.0
IMDb              7.3
Netflix           0.0
Hulu              0.0
Prime Video       0.0
Disney+           0.0
type              1.0
dtype: float64
```

Correlation

In [28]:
```python
df.corr()
```

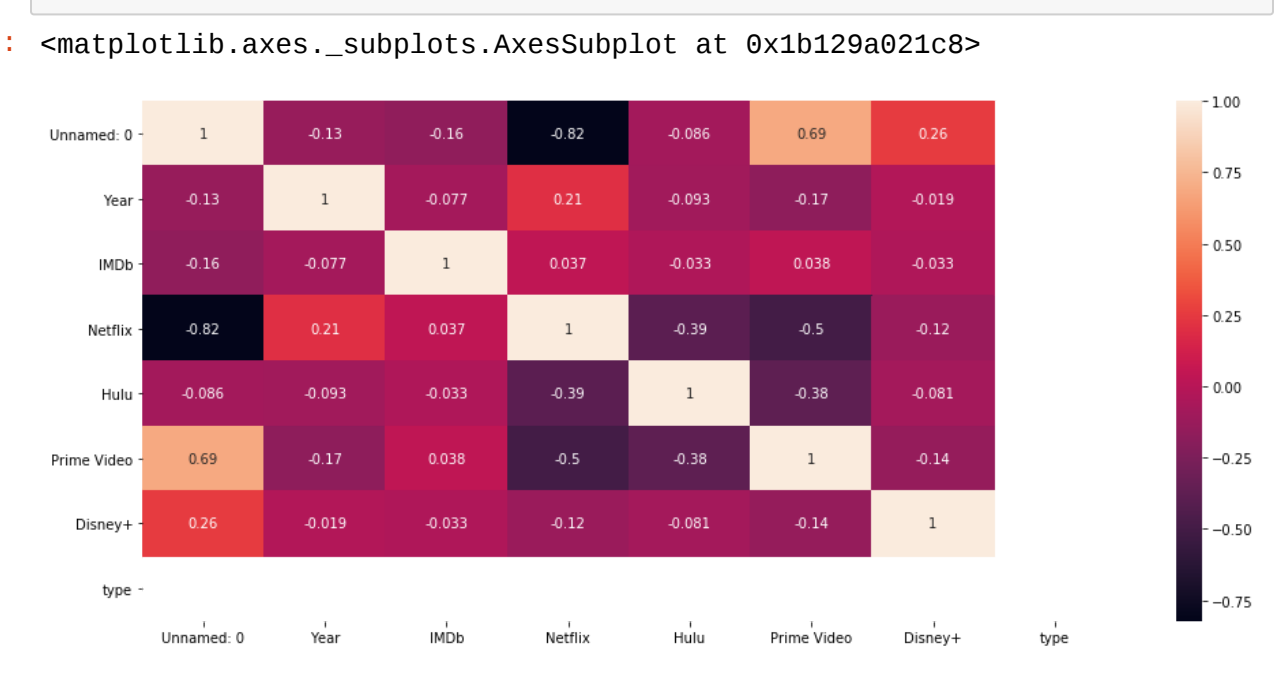
Out[28]:

	Unnamed: 0	Year	IMDb	Netflix	Hulu	Prime Video	Disney+	type
Unnamed: 0	1.000000	-0.132975	-0.163835	-0.822878	-0.085934	0.689293	0.255014	NaN
Year	-0.132975	1.000000	-0.077347	0.207759	-0.092723	-0.171060	-0.018914	NaN
IMDb	-0.163835	-0.077347	1.000000	0.036608	-0.033134	0.037909	-0.033279	NaN
Netflix	-0.822878	0.207759	0.036608	1.000000	-0.392190	-0.500160	-0.119102	NaN
Hulu	-0.085934	-0.092723	-0.033134	-0.392190	1.000000	-0.375221	-0.081313	NaN
Prime Video	0.689293	-0.171060	0.037909	-0.500160	-0.375221	1.000000	-0.143163	NaN
Disney+	0.255014	-0.018914	-0.033279	-0.119102	-0.081313	-0.143163	1.000000	NaN
type	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

In [29]:
```python
import seaborn as sns
```

In [30]:
```python
sns.heatmap(df.corr(),annot=True)
```

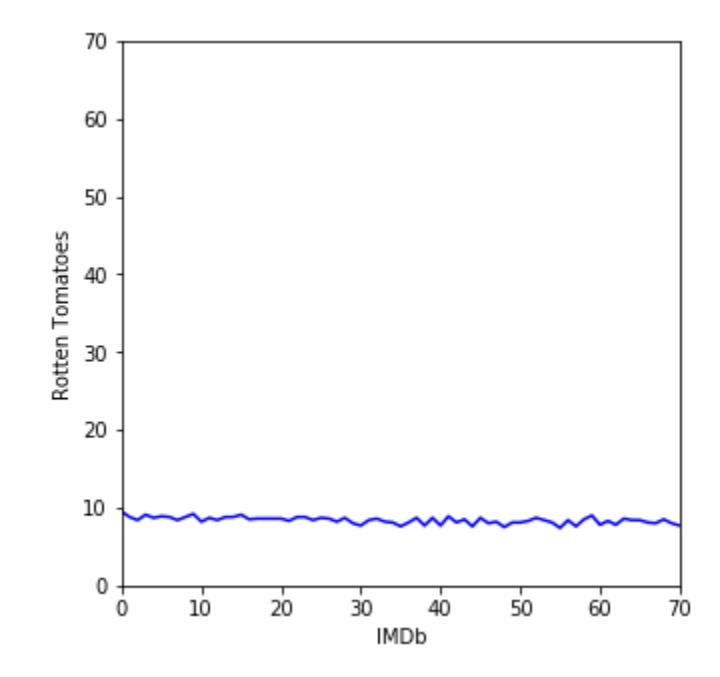
Out[30]: <matplotlib.axes._subplots.AxesSubplot at 0x1b129a021c8>



Lineplot

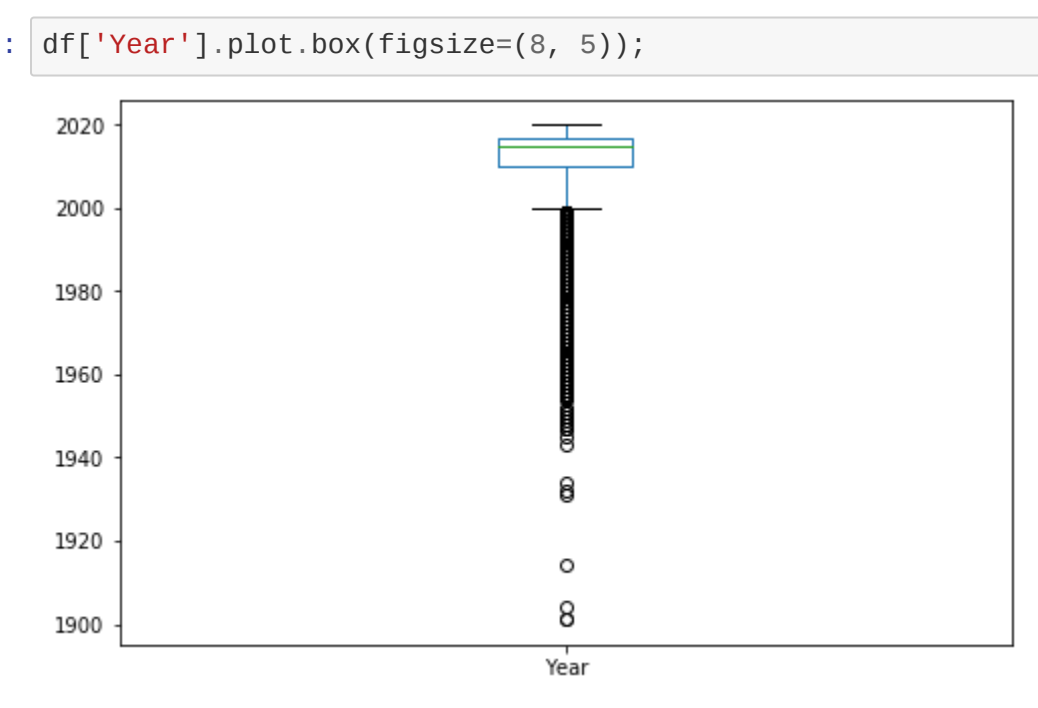
In [42]:
```python
df['IMDb'].plot(figsize=(5,5),color='blue')
plt.xlim(0,70)
plt.ylim(0,70)
plt.xlabel('IMDb')
plt.ylabel('Rotten Tomatoes')
```

Out[42]: Text(0, 0.5, 'Rotten Tomatoes')



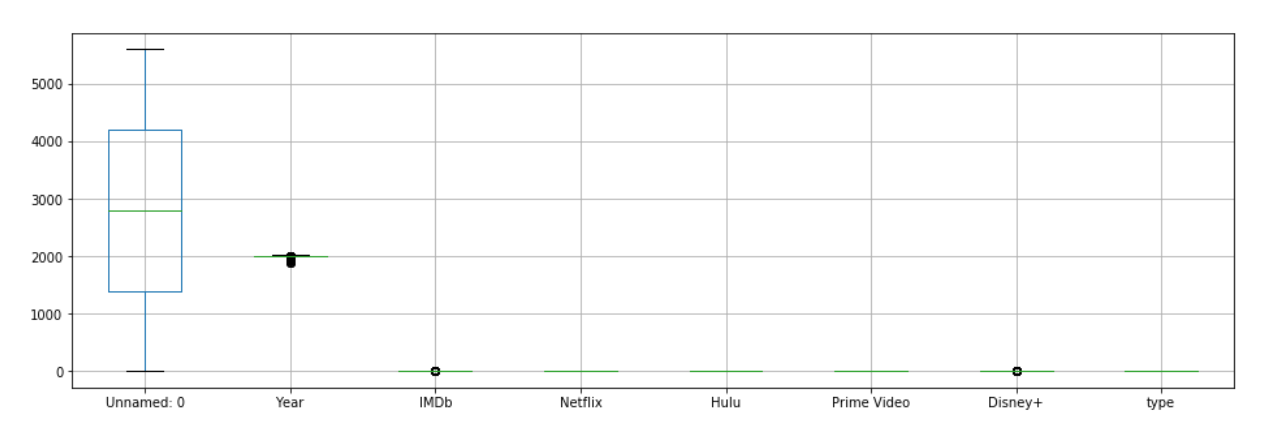
Boxplot

In [47]:
```python
df['Year'].plot.box(figsize=(8, 5));
```



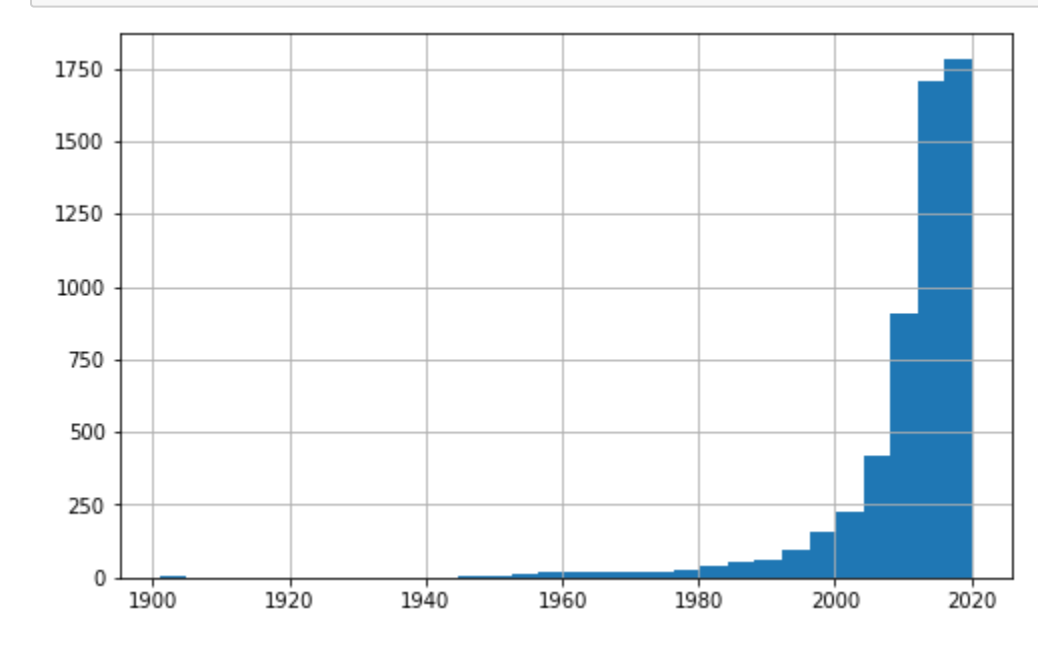
In [48]:
```python
df.boxplot(figsize=(16, 5))
```

Out[48]: <matplotlib.axes._subplots.AxesSubplot at 0x1b12bd37388>

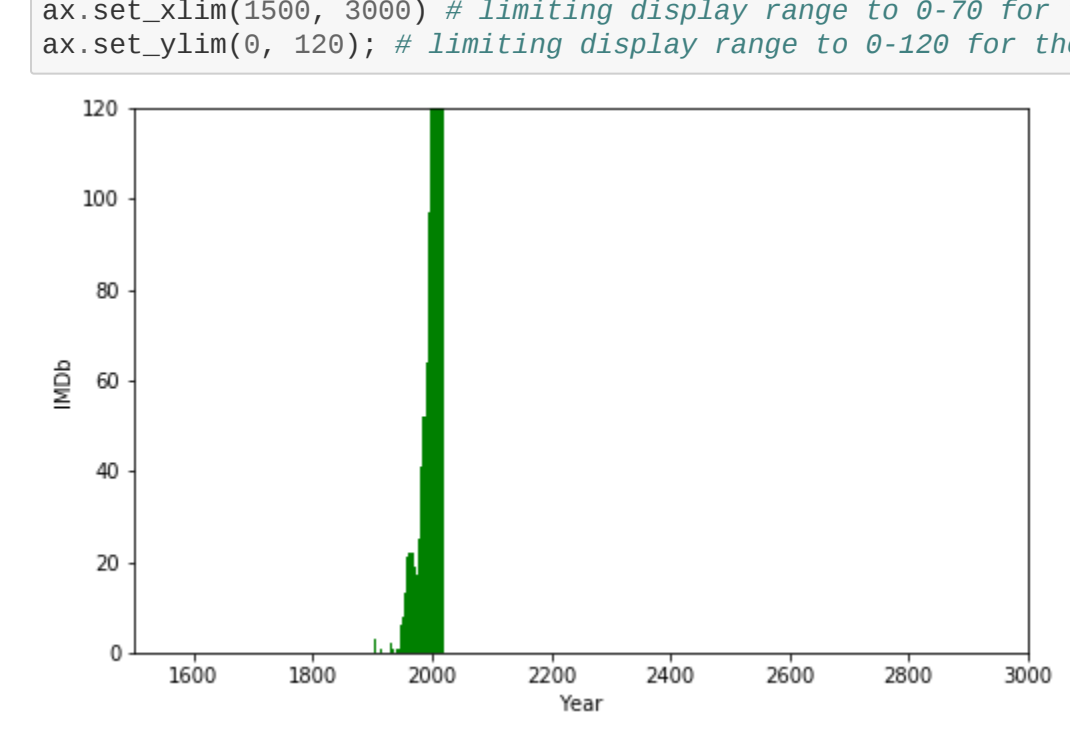


Histogram

In [49]:
```python
df['Year'].hist(bins=30, figsize=(8, 5));
```

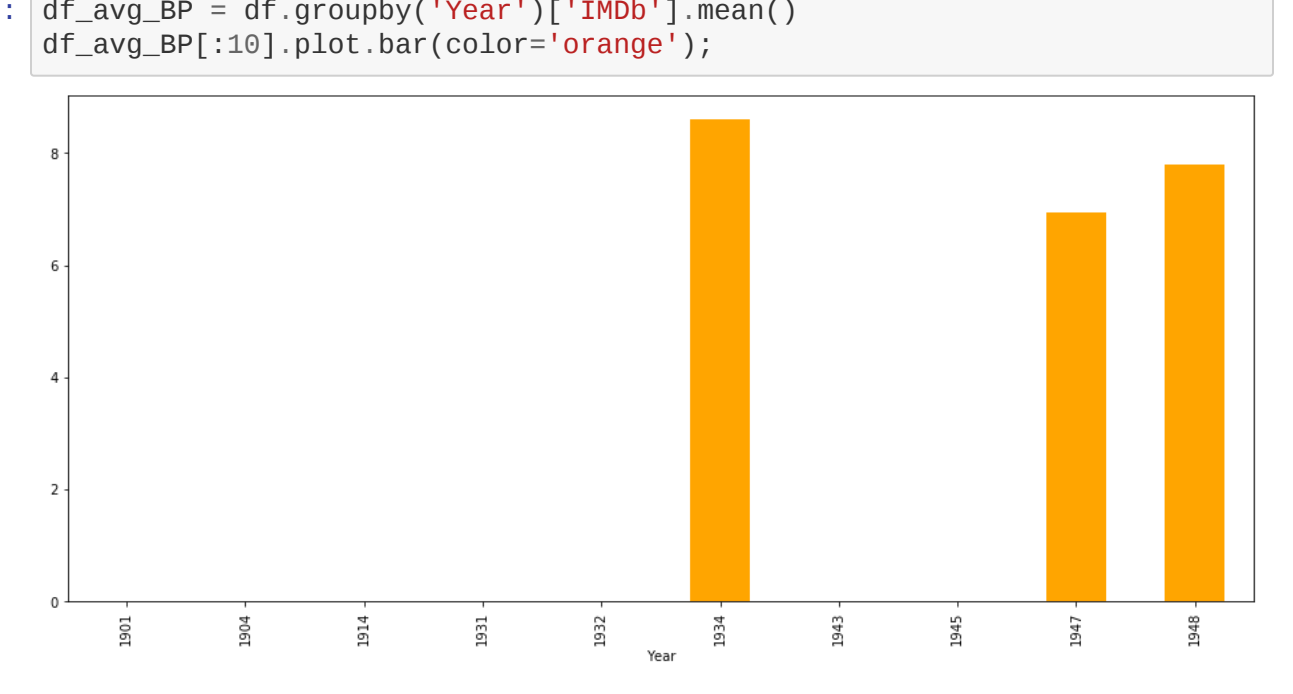


In [57]:
```python
ax = df['Year'].hist(bins=30, grid=False, color='green', figsize=(8, 5))
# grid turned off and color changed

ax.set_xlabel('Year')
ax.set_ylabel('IMDb')

ax.set_xlim(1500, 3000) # limiting display range to 0-70 for the x-axis
ax.set_ylim(0, 120); # limiting display range to 0-120 for the y-axis
```

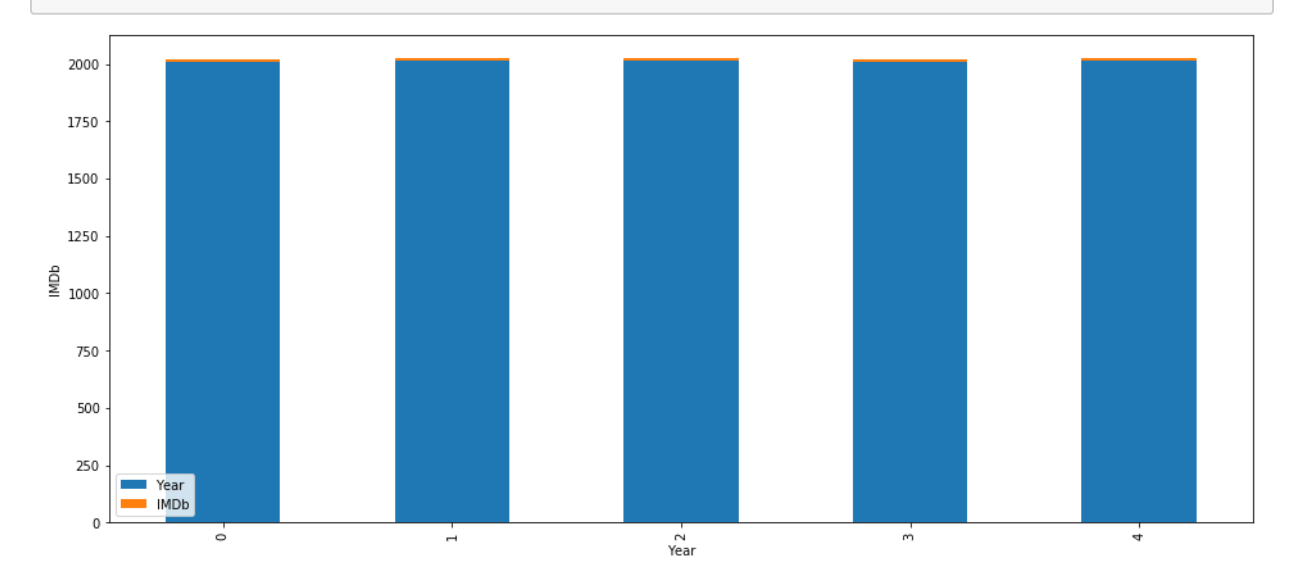


Barplot

In [59]:
```python
df_avg_BP = df.groupby('Year')['IMDb'].mean()
df_avg_BP[:10].plot.bar(color='orange');
```



In [61]:
```python
ax = df[['Year', 'IMDb']][:5].plot.bar(stacked=True)

ax.set_xlabel("Year")
ax.set_ylabel("IMDb");
```



In [ ]:

In [ ]: