Begin by importing the necessary modules to carry out the analysis

In [9]:

```
import pandas as pd
import datetime
import re
import requests
from bs4 import BeautifulSoup
import numpy as np
import matplotlib.pyplot as plt
import matplotlib as mpl
```

Import the datafile downloaded from kaggle.com

In [10]:

```
data = pd.read_csv('./input/Airplane_Crashes_and_Fatalities_Since_1908_20190820105639.csv')
data.head()
```

Out[10]:

	Date	Time	Location	Operator	Flight #	Route	AC Type	Registration	cn/ln
0	09/17/1908	17:18	Fort Myer, Virginia	Military - U.S. Army	NaN	Demonstration	Wright Flyer III	NaN	1
1	09/07/1909	NaN	Juvisy-sur-Orge, France	NaN	NaN	Air show	Wright Byplane	SC1	NaN
2	07/12/1912	06:30	Atlantic City, New Jersey	Military - U.S. Navy	NaN	Test flight	Dirigible	NaN	NaN
3	08/06/1913	NaN	Victoria, British Columbia, Canada	Private	NaN	NaN	Curtiss seaplane	NaN	NaN
4	09/09/1913	18:30	Over the North Sea	Military - German Navy	NaN	NaN	Zeppelin L-1 (airship)	NaN	NaN

In [11]:

```
data.shape
```

Out[11]:

```
(4967, 17)
```

In [12]:

```
print(data.columns)
```

```
Index(['Date', 'Time', 'Location', 'Operator', 'Flight #', 'Route',
'AC Type',
       'Registration', 'cn/ln', 'Aboard', 'Aboard Passangers', 'Aboa
rd Crew',
       'Fatalities', 'Fatalities Passangers', 'Fatalities Crew', 'Gr
ound',
       'Summary'],
      dtype='object')
```

Selecting columns that are relevant to the analysis to be achieved, regarding the trend in plane crashes along the year and the operators that have had the most crashes (top 20 ranking)

In [13]:

```
data1 = data[["Date","Time", "Location", "Operator", "Fatalities", "Summary"]]

data1.head()
```

Out[13]:

	Date	Time	Location	Operator	Fatalities	Summary
0	09/17/1908	17:18	Fort Myer, Virginia	Military - U.S. Army	1.0	During a demonstration flight, a U.S. Army fly...
1	09/07/1909	NaN	Juvisy-sur-Orge, France	NaN	1.0	Eugene Lefebvre was the first pilot to ever be...
2	07/12/1912	06:30	Atlantic City, New Jersey	Military - U.S. Navy	5.0	First U.S. dirigible Akron exploded just offsh...
3	08/06/1913	NaN	Victoria, British Columbia, Canada	Private	1.0	The first fatal airplane accident in Canada oc...
4	09/09/1913	18:30	Over the North Sea	Military - German Navy	14.0	The airship flew into a thunderstorm and encou...

Checking for null cells and dropping the values that will not provide info towards the desired analysis

In [14]:

```
null_cols = data1.isnull().sum()
null_cols
# Seeing that around 30% of the Time data is missing, decided to do without the
 time that the accident occured, and hence this column was dropped

data2 = data1.drop("Time", axis=1)

data2.head()
```

Out[14]:

	Date	Location	Operator	Fatalities	Summary
0	09/17/1908	Fort Myer, Virginia	Military - U.S. Army	1.0	During a demonstration flight, a U.S. Army fly...
1	09/07/1909	Juvisy-sur-Orge, France	NaN	1.0	Eugene Lefebvre was the first pilot to ever be...
2	07/12/1912	Atlantic City, New Jersey	Military - U.S. Navy	5.0	First U.S. dirigible Akron exploded just offsh...
3	08/06/1913	Victoria, British Columbia, Canada	Private	1.0	The first fatal airplane accident in Canada oc...
4	09/09/1913	Over the North Sea	Military - German Navy	14.0	The airship flew into a thunderstorm and encou...

In [15]:

```
# Dropping rows containing null as they are an insignificant number

data3 = data2.dropna()

data3
```

Out[15]:

	Date	Location	Operator	Fatalities	Summary
0	09/17/1908	Fort Myer, Virginia	Military - U.S. Army	1.0	During a demonstration flight, a U.S. Army fly...
2	07/12/1912	Atlantic City, New Jersey	Military - U.S. Navy	5.0	First U.S. dirigible Akron exploded just offsh...
3	08/06/1913	Victoria, British Columbia, Canada	Private	1.0	The first fatal airplane accident in Canada oc...
4	09/09/1913	Over the North Sea	Military - German Navy	14.0	The airship flew into a thunderstorm and encou...
5	10/17/1913	Near Johannisthal, Germany	Military - German Navy	30.0	Hydrogen gas which was being vented was sucked...
...	...	...	...	...	...
4962	04/16/2019	Puerto Montt, Chile	Archipelagos Service Aereos	6.0	While the aircraft was in the initial climb, p...
4963	05/05/2019	Near Monclava, Mexico	TVPX Aircraft Solutions	13.0	The aircraft crashed while en route on a retur...
4964	05/05/2019	Moscow, Russia	Aeroflot Russian International Airlines	41.0	Forty-five minutes after taking off from Mosco...
4965	06/03/2019	Near Lipo, India	Military - Indian Air Force	13.0	Crashed about 34km WNW of Mechuka.
4966	07/30/2019	Rawalpindi, India	Military - Pakistan Army	5.0	The Pakistani military plane, on a training fl...

4886 rows × 5 columns

In [16]:

```
#Checking if any null cells are left

null_cols = data3.isnull().sum()
null_cols
```

Out[16]:

```
Date          0
Location      0
Operator      0
Fatalities    0
Summary       0
dtype: int64
```

The fatalities columns is given in float, so changing it to integers

In [17]:

```
data3['Fatalities'] = data3["Fatalities"].astype(np.int64)
data3
```

```
/usr/local/lib/python3.6/dist-packages/ipykernel_launcher.py:1: Sett
ingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/panda
s-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-cop
y
  """Entry point for launching an IPython kernel.
```

Out[17]:

	Date	Location	Operator	Fatalities	Summary
0	09/17/1908	Fort Myer, Virginia	Military - U.S. Army	1	During a demonstration flight, a U.S. Army fly...
2	07/12/1912	Atlantic City, New Jersey	Military - U.S. Navy	5	First U.S. dirigible Akron exploded just offsh...
3	08/06/1913	Victoria, British Columbia, Canada	Private	1	The first fatal airplane accident in Canada oc...
4	09/09/1913	Over the North Sea	Military - German Navy	14	The airship flew into a thunderstorm and encou...
5	10/17/1913	Near Johannisthal, Germany	Military - German Navy	30	Hydrogen gas which was being vented was sucked...
...	...	...	...	...	...
4962	04/16/2019	Puerto Montt, Chile	Archipelagos Service Aereos	6	While the aircraft was in the initial climb, p...
4963	05/05/2019	Near Monclava, Mexico	TVPX Aircraft Solutions	13	The aircraft crashed while en route on a retur...
4964	05/05/2019	Moscow, Russia	Aeroflot Russian International Airlines	41	Forty-five minutes after taking off from Mosco...
4965	06/03/2019	Near Lipo, India	Military - Indian Air Force	13	Crashed about 34km WNW of Mechuka.
4966	07/30/2019	Rawalpindi, India	Military - Pakistan Army	5	The Pakistani military plane, on a training fl...

4886 rows × 5 columns

In [18]:

```
print(set(data3['Fatalities']))
```

```
{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 1
9, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 3
6, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 5
3, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 7
0, 583, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86,
87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 1
03, 104, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117,
118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 13
1, 132, 133, 134, 137, 140, 141, 143, 144, 145, 146, 148, 150, 152,
153, 154, 155, 156, 157, 158, 159, 160, 162, 163, 166, 167, 168, 16
9, 170, 171, 174, 176, 178, 180, 181, 183, 187, 188, 189, 191, 196,
200, 213, 217, 223, 224, 225, 228, 229, 230, 234, 239, 256, 257, 25
9, 260, 261, 264, 269, 271, 275, 520, 290, 298, 301, 329, 346, 349,
71}
```

In [19]:

```
data3.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 4886 entries, 0 to 4966
Data columns (total 5 columns):
Date          4886 non-null object
Location      4886 non-null object
Operator      4886 non-null object
Fatalities    4886 non-null int64
Summary       4886 non-null object
dtypes: int64(1), object(4)
memory usage: 229.0+ KB
```

Checking the Date column and dropping the columns that represent crashes older than 50 years, change in technology and number of commercial flights mean those years are irrelevant

In [20]:

```
data3['Date'] = pd.to_datetime(data3['Date'])
data3['Year'] = data3['Date'].dt.year

# Only take into account the crashes occured in the last 50 years
data3 = data3[data3['Year'] > 1969].reset_index()
data3
```

```
/usr/local/lib/python3.6/dist-packages/ipykernel_launcher.py:1: Sett
ingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/panda
s-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-cop
y
  """Entry point for launching an IPython kernel.
/usr/local/lib/python3.6/dist-packages/ipykernel_launcher.py:2: Sett
ingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/panda
s-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-cop
y
```

Out[20]:

	index	Date	Location	Operator	Fatalities	Summary	Year
0	2435	1970-01-05	Stockholm, Sweden	Spantax	5	The plane developed trouble in the No. 4 engin...	1970
1	2436	1970-01-12	Near Villia Greece	Military - Royal Hellenic Air Force	23	The aircraft, carrying paratroopers, crashed i...	1970
2	2437	1970-01-13	Faleolo, Western Samoa	Polynesian Airlines	32	Crashed into a lagoon 400 yards past the end o...	1970
3	2438	1970-01-14	Mt. Pumacona, Peru	Faucett	28	Flew into a 10,500 ft. mountain. The mental st...	1970
4	2439	1970-01-25	Near Delhi, India	Royal Nepal Airlines	1	The aircraft crashed short of the runway after...	1970
...	...	...	...	...	...	...	...
2508	4962	2019-04-16	Puerto Montt, Chile	Archipelagos Service Aereos	6	While the aircraft was in the initial climb, p...	2019
2509	4963	2019-05-05	Near Monclava, Mexico	TVPX Aircraft Solutions	13	The aircraft crashed while en route on a retur...	2019
2510	4964	2019-05-05	Moscow, Russia	Aeroflot Russian International Airlines	41	Forty-five minutes after taking off from Mosco...	2019
2511	4965	2019-06-03	Near Lipo, India	Military - Indian Air Force	13	Crashed about 34km WNW of Mechuka.	2019
2512	4966	2019-07-30	Rawalpindi, India	Military - Pakistan Army	5	The Pakistani military plane, on a training fl...	2019

2513 rows × 7 columns

Grouping crashes per year

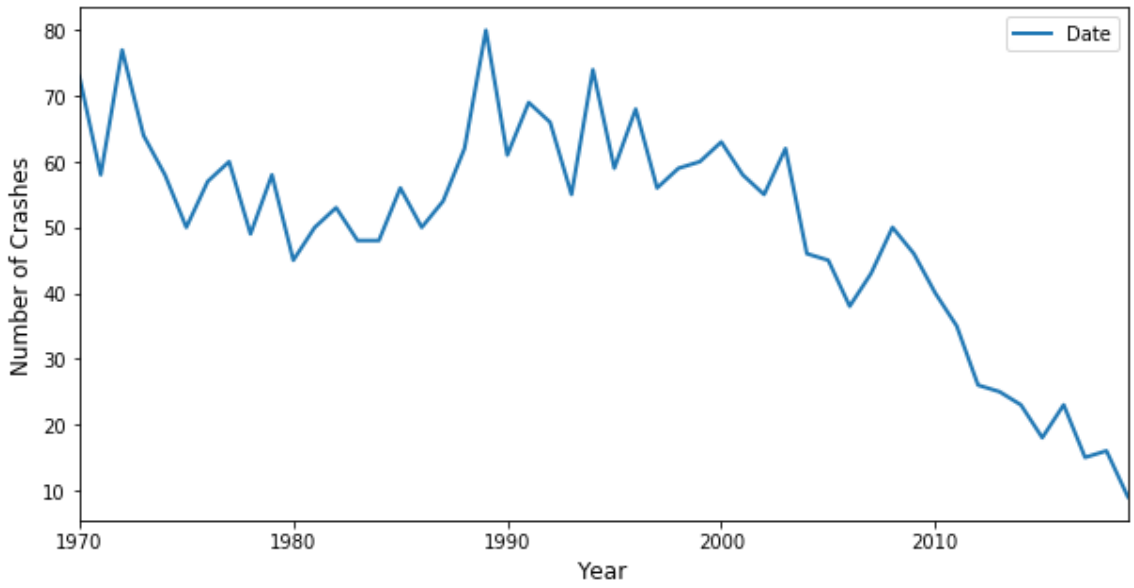
In [21]:

```
crashes_year = data3.loc[:, ["Year","Date"]].groupby(['Year']).count()
crashes_year.head()

plot1 = crashes_year.plot(lw=2, figsize=(10,5))
plot1.set_xlabel("Year", fontsize=12)
plot1.set_ylabel("Number of Crashes", fontsize=12)
```

Out[21]:

```
Text(0,0.5,'Number of Crashes')
```



Grouping accidents by operator (to check the most dangerous operator and hence the ones to avoid)

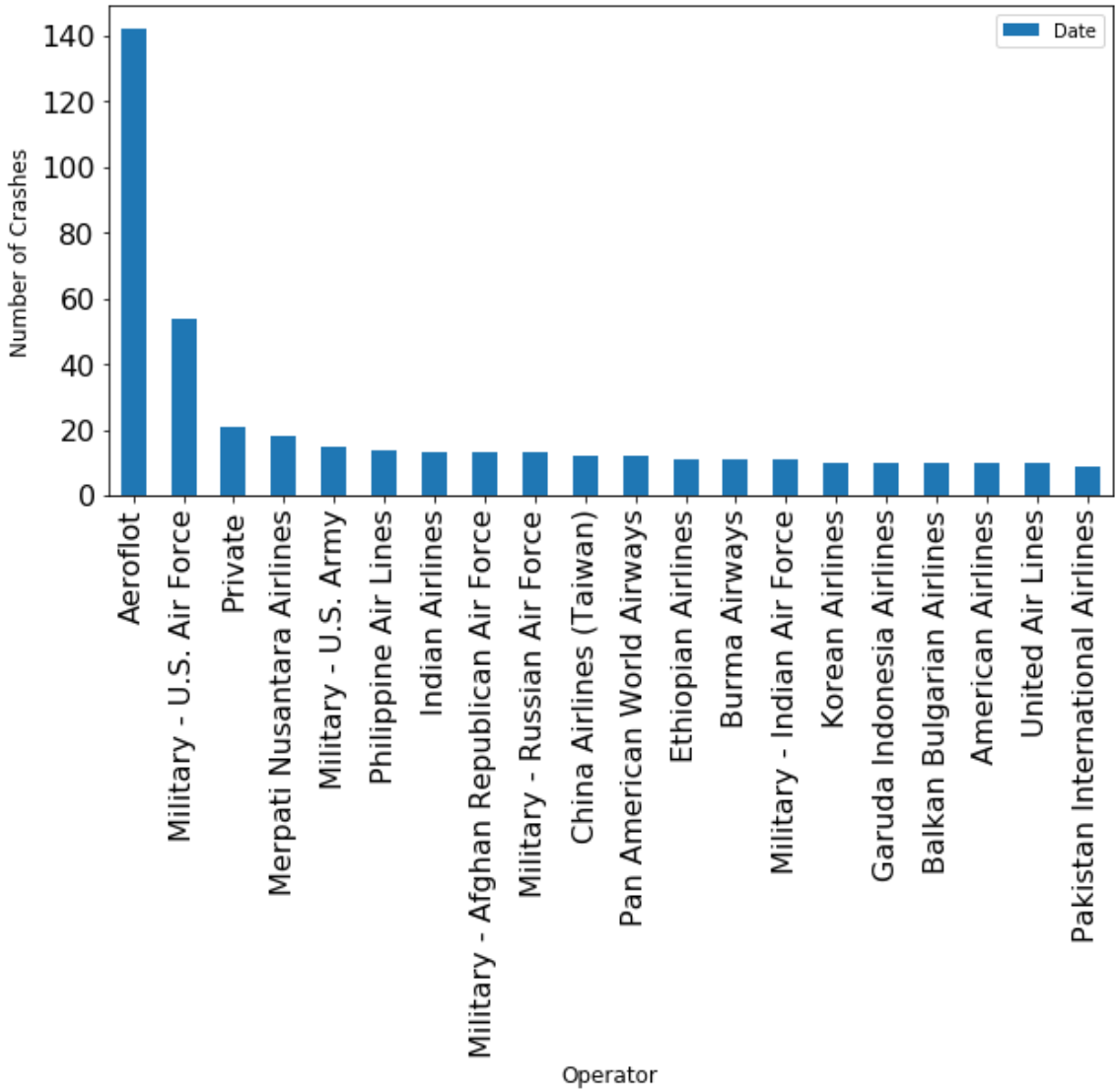
In [22]:

```
crashed_operator = data3.loc[:, ["Operator","Date"]].groupby(['Operator']).count
()
crashed_operator = crashed_operator.sort_values(by="Date", ascending=False).head
(20)

plot2 = crashed_operator.plot.bar(fontsize = 16, figsize=(10,5))
plot2.set_xlabel("Operator", fontsize=12)
plot2.set_ylabel("Number of Crashes", fontsize=12)
```

Out[22]:

```
Text(0,0.5,'Number of Crashes')
```



Conclusion

The number of plane crashes has decreased throughout the years and the most "dangerous" operator is Aeroflot.

PS - take into account that the value for 2019 is incomplete as only crashes until august have been accounted for.

In [ ]: