1.2.1 summary of the data and the questions that I want to explore:

I took the Kaggle data from a website called cardekho that has the details of used cars such as the kms driven, model, year, fuel type, mileage owner type etc., to name a few. Based on this data set, the questions that I want to explore with the data set are:

1.In which year, the cars were bought more.

2.and what type of cars were bought more that year.

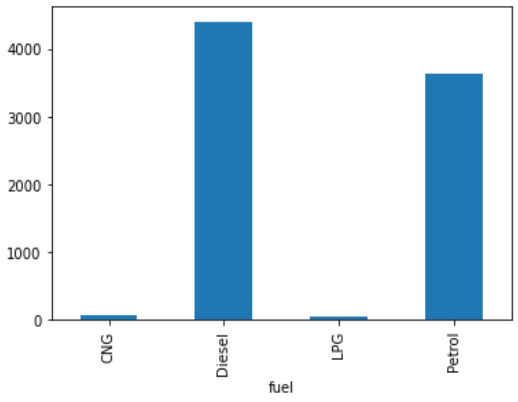
In [9]:
```
import pandas as pd
import thinkstats2
import thinkplot
df = pd.read_csv ('CarDetails.csv')
df.head()
```

Out[9]:

	name	year	selling_price	km_driven	fuel	seller_type	transmission	owner	mileage	engine	max_power	torque	seats
0	Maruti Swift Dzire VDI	2014	450000	145500	Diesel	Individual	Manual	First Owner	23.4 kmpl	1248 CC	74 bhp	190Nm@ 2000rpm	5.0
1	Skoda Rapid 1.5 TDI Ambition	2014	370000	120000	Diesel	Individual	Manual	Second Owner	21.14 kmpl	1498 CC	103.52 bhp	250Nm@ 1500-2500rpm	5.0
2	Honda City 2017-2020 EXi	2006	158000	140000	Petrol	Individual	Manual	Third Owner	17.7 kmpl	1497 CC	78 bhp	12.7@ 2,700(kgm@ rpm)	5.0
3	Hyundai i20 Sportz Diesel	2010	225000	127000	Diesel	Individual	Manual	First Owner	23.0 kmpl	1396 CC	90 bhp	22.4 kgm at 1750-2750rpm	5.0
4	Maruti Swift VXI BSIII	2007	130000	120000	Petrol	Individual	Manual	First Owner	16.1 kmpl	1298 CC	88.2 bhp	11.5@ 4,500(kgm@ rpm)	5.0

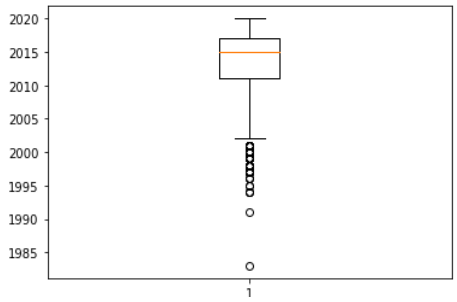
1.2.2 A Bar chart to show the fuel type and their frequency

In [32]:
```
#1.2.2 Bar chart to show fuel type and their frequency

df2_all = df.groupby(['fuel'])['fuel'].count()
df2_all.plot.bar(x="fuel", y="count");
#print(df2)
```



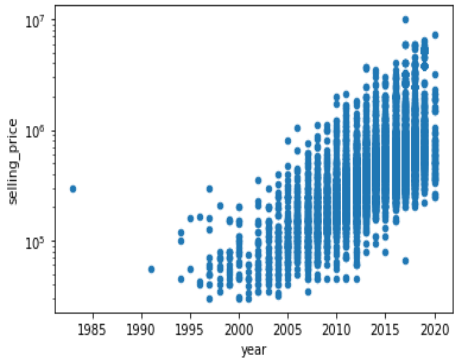
1.2.3 Box plot on the year.

In [13]:
```
# 1.2.3 Box plot on the year
import matplotlib.pyplot as plt
import numpy as np

# Creating plot
plt.boxplot(df.year)

# show plot
plt.show()
```



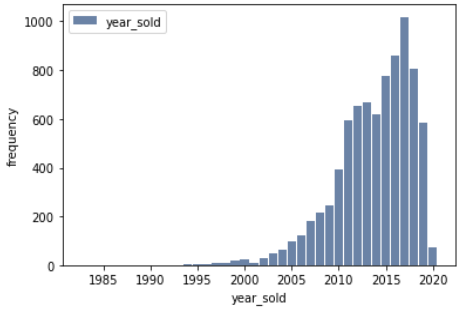
1.2.4 Bivariate plot between the year and selling price. As the vales in the Y Axis are the selling price that involves big numbers, I used log option to plot it.

In [18]:
```
# year Vs Selling pricec Scatter plot
df.plot(x='year', y='selling_price', kind='scatter',logy=True)
plt.show()
```



1.2.5 To answer my first question, which year the cars were bought more, I plotted a histogram that will help us to identify the year. Based on the year, it seems that the cars were bought more in the year 2017.

In [19]:
```
# Based on the Histogram it seems that the cars were bought more in 2017
hist2 = thinkstats2.Hist(df.year, label='year_sold')
thinkplot.Hist(hist2)
thinkplot.Show(xlabel='year_sold', ylabel='frequency')
```




```
<Figure size 576x432 with 0 Axes>
```

Also, to answer the second question, what type of cars were bought more that year. I plotted a bar chart after filtering the data for the year 2017 for the fuel type and the counts. Based on the bar chart, it seems that the Diesel cars were bought more in the year 2017.

In [30]:
```
# To answer our next question we need to filter in the year 2017 and plot a bar chart of the car Type anf frequency.

from matplotlib import pyplot as plt

df_2017=df[df.year==2017] #Filtering the ones from 2017
df2 = df_2017.groupby(['fuel'])['fuel'].count()
df2.plot.bar(x="fuel", y="count");
```