First, get the notebook setup with the key packages

In [3]:
```
import pandas as pd
import matplotlib.pyplot as plt
```

In [4]:
```
%matplotlib inline
url = "http://pbpython.com/extras/sample-salesv2.csv"
sales = pd.read_csv(url)
```

In [5]:
```
sales.head()
```

Out[5]:

	account number	name	sku	category	quantity	unit price	ext price	date
0	296809	Carroll PLC	QN-82852	Belt	13	44.48	578.24	2014-09-27 07:13:03
1	98022	Heidenreich-Bosco	MJ-21460	Shoes	19	53.62	1018.78	2014-07-29 02:10:44
2	563905	Kerluke, Reilly and Bechtelar	AS-93055	Shirt	12	24.16	289.92	2014-03-01 10:51:24
3	93356	Waters-Walker	AS-93055	Shirt	5	82.68	413.40	2013-11-17 20:41:11
4	659366	Waelchi-Fahey	AS-93055	Shirt	18	99.64	1793.52	2014-01-03 08:14:27

Rename the columns to use underscores instead of spaces

In [6]:

```
sales.columns = ['account_number', 'name', 'sku', 'category', 'quantity', 'unit_price', 'ext_price', 'date']

sales.head()
```

Out[6]:

	account_number	name	sku	category	quantity	unit_price	ext_price	date
0	296809	Carroll PLC	QN-82852	Belt	13	44.48	578.24	2014-09-27 07:13:03
1	98022	Heidenreich-Bosco	MJ-21460	Shoes	19	53.62	1018.78	2014-07-29 02:10:44
2	563905	Kerluke, Reilly and Bechtelar	AS-93055	Shirt	12	24.16	289.92	2014-03-01 10:51:24
3	93356	Waters-Walker	AS-93055	Shirt	5	82.68	413.40	2013-11-17 20:41:11
4	659366	Waelchi-Fahey	AS-93055	Shirt	18	99.64	1793.52	2014-01-03 08:14:27

Subset the data frame to include only name, category, quantity, and unit price columns

In [7]:

```
subset = sales[['name', 'category','quantity', 'unit_price']]
subset.head()
```

Out[7]:

	name	category	quantity	unit_price
0	Carroll PLC	Belt	13	44.48
1	Heidenreich-Bosco	Shoes	19	53.62
2	Kerluke, Reilly and Bechtelar	Shirt	12	24.16
3	Waters-Walker	Shirt	5	82.68
4	Waelchi-Fahey	Shirt	18	99.64

In [8]:

```
shirts = subset[subset['category']=='Shirt']
shirts.head()
```

Out[8]:

	name	category	quantity	unit_price
2	Kerluke, Reilly and Bechtelar	Shirt	12	24.16
3	Waters-Walker	Shirt	5	82.68
4	Waelchi-Fahey	Shirt	18	99.64
5	Kerluke, Reilly and Bechtelar	Shirt	17	52.82
9	Kerluke, Reilly and Bechtelar	Shirt	12	26.98

Calculate the total cost per order for shirt sales

In [9]:

```
shirts['total_cost'] = shirts.quantity * shirts.unit_price
shirts.head()
```

```
/srv/conda/envs/notebook/lib/python3.7/site-packages/ipykernel_launcher.py:1: SettingWithCopyWarning:
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/pandas-docs/stable/user_guide/indexing.html#re
turning-a-view-versus-a-copy
  """Entry point for launching an IPython kernel.
```

Out[9]:

	name	category	quantity	unit_price	total_cost
2	Kerluke, Reilly and Bechtelar	Shirt	12	24.16	289.92
3	Waters-Walker	Shirt	5	82.68	413.40
4	Waelchi-Fahey	Shirt	18	99.64	1793.52
5	Kerluke, Reilly and Bechtelar	Shirt	17	52.82	897.94
9	Kerluke, Reilly and Bechtelar	Shirt	12	26.98	323.76

In [ ]:

```
Now group the sales by company name
```

In [27]:

```
shirts_by_company = shirts.groupby('name', as_index=False).sum()
shirts_by_company.head()
```

Out[27]:

	name	quantity	unit_price	total_cost
0	Berge LLC	166	1226.54	9670.24
1	Carroll PLC	257	1098.93	13717.61
2	Cole-Eichmann	236	1226.75	14528.01
3	Davis, Kshlerin and Reilly	161	828.51	7533.03
4	Ernser, Cruickshank and Lind	262	1500.25	16944.19

Pull out the top 10 shirt sales

In [29]:

```
top_10_shirt_sales = shirts_by_company.sort_values(by='total_cost', ascending=False).head(10)
```

In [30]:

```
top_10_shirt_sales
```

Out[30]:

	name	quantity	unit_price	total_cost
11	Kihn, McClure and Denesik	288	1653.58	18956.35
19	Waters-Walker	288	1603.36	18633.71
4	Ernser, Cruickshank and Lind	262	1500.25	16944.19
7	Hegmann and Sons	278	1528.84	16774.47
14	Kunze Inc	260	1439.92	15638.87
2	Cole-Eichmann	236	1226.75	14528.01
1	Carroll PLC	257	1098.93	13717.61
10	Kerluke, Reilly and Bechtelar	269	1038.53	12958.23
17	Volkman, Goyette and Lemke	220	1136.25	12791.27
5	Gorczany-Hahn	237	1132.22	12576.83

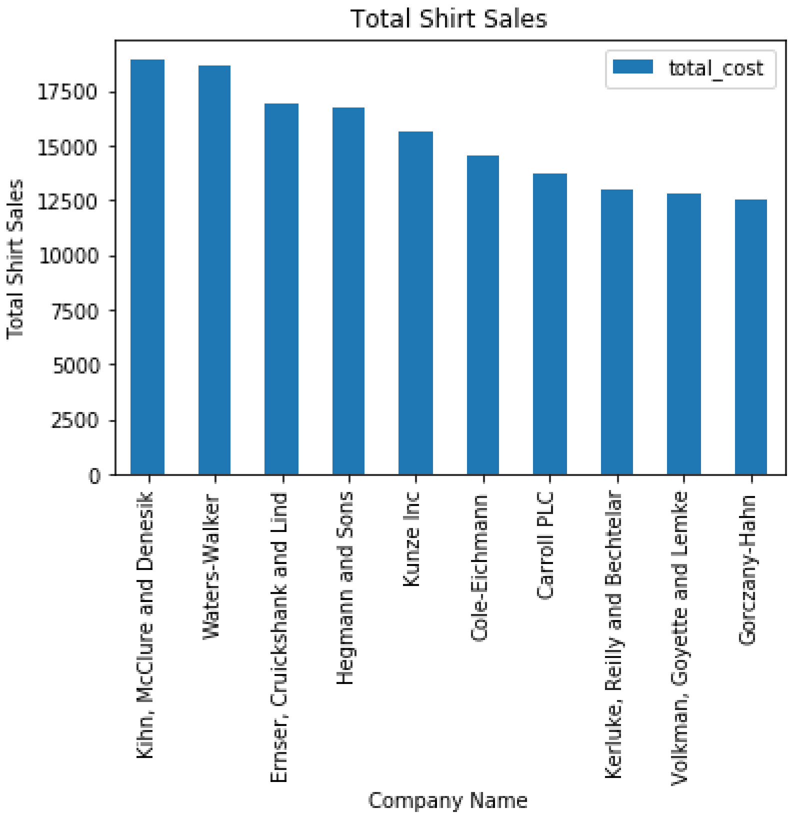
Graph the top 10 shirt sales

In [32]:

```
top_10_chart = top_10_shirt_sales.plot(kind="bar", title="Total Shirt Sales", x="name", y="total_cost")
top_10_chart.set_xlabel("Company Name")
top_10_chart.set_ylabel("Total Shirt Sales")
```

Out[32]:

```
Text(0,0.5,'Total Shirt Sales')
```


Total Shirt Sales
total_cost
Total Shirt Sales
Kihn, McClure and Denesik
Waters-Walker
Ernser, Cruickshank and Lind
Hegmann and Sons
Kunze Inc
Cole-Eichmann
Carroll PLC
Kerluke, Reilly and Bechtelar
Volkman, Goyette and Lemke
Gorczany-Hahn
Company Name