In [4]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

In [5]:
```
df=pd.read_csv("Superstore.csv")
```

In [6]:
```
df.head()
```

Out[6]:

	Ship Mode	Segment	Country	City	State	Postal Code	Region	Category	Sub-Catego
0	Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Bookcase
1	Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Chai
2	Second Class	Corporate	United States	Los Angeles	California	90036	West	Office Supplies	Labe
3	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Furniture	Tabl
4	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Office Supplies	Storag

In [7]:
```
df.tail()
```

Out[7]:

	Ship Mode	Segment	Country	City	State	Postal Code	Region	Category	C
9989	Second Class	Consumer	United States	Miami	Florida	33180	South	Furniture	Fu
9990	Standard Class	Consumer	United States	Costa Mesa	California	92627	West	Furniture	Fu
9991	Standard Class	Consumer	United States	Costa Mesa	California	92627	West	Technology	
9992	Standard Class	Consumer	United States	Costa Mesa	California	92627	West	Office Supplies	
9993	Second Class	Consumer	United States	Westminster	California	92683	West	Office Supplies	Ap

In [13]:
```
print("(Rows,Columns)")
df.shape
```

```
(Rows,Columns)
```

Out[13]:
```
(9994, 13)
```

In [14]:
```
df.info() #we can get information about columns by info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9994 entries, 0 to 9993
Data columns (total 13 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Ship Mode     9994 non-null   object
 1   Segment       9994 non-null   object
 2   Country       9994 non-null   object
 3   City          9994 non-null   object
 4   State         9994 non-null   object
 5   Postal Code   9994 non-null   int64
 6   Region        9994 non-null   object
 7   Category      9994 non-null   object
 8   Sub-Category  9994 non-null   object
 9   Sales         9994 non-null   float64
 10  Quantity      9994 non-null   int64
 11  Discount      9994 non-null   float64
 12  Profit        9994 non-null   float64
dtypes: float64(3), int64(2), object(8)
memory usage: 1015.1+ KB
```

In [15]:
```
df.describe() #give mean, standard deviation, minimum and maximum value
```

Out[15]:

	Postal Code	Sales	Quantity	Discount	Profit
count	9994.000000	9994.000000	9994.000000	9994.000000	9994.000000
mean	55190.379428	229.858001	3.789574	0.156203	28.656896
std	32063.693350	623.245101	2.225110	0.206452	234.260108
min	1040.000000	0.444000	1.000000	0.000000	-6599.978000
25%	23223.000000	17.280000	2.000000	0.000000	1.728750
50%	56430.500000	54.490000	3.000000	0.200000	8.666500
75%	90008.000000	209.940000	5.000000	0.200000	29.364000
max	99301.000000	22638.480000	14.000000	0.800000	8399.976000

In [16]:
```
df.nunique()
```

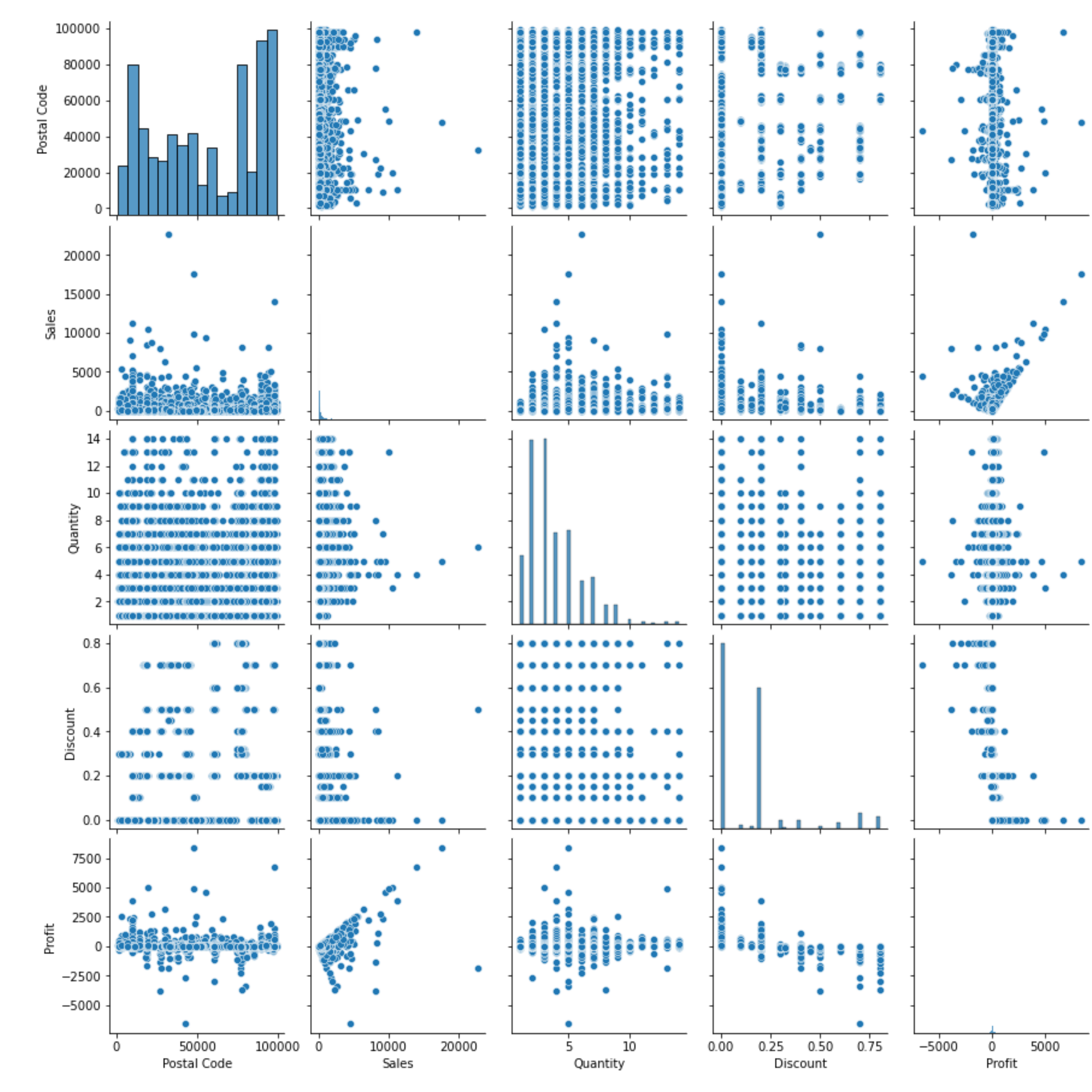
Out[16]:
```
Ship Mode          4
Segment            3
Country            1
City             531
State             49
Postal Code      631
Region             4
Category           3
Sub-Category      17
Sales           5825
Quantity          14
Discount          12
Profit          7287
dtype: int64
```

In [17]:
```
df.isnull().sum()#Data Cleaning
```

Out[17]:
```
Ship Mode       0
Segment         0
Country         0
City            0
State           0
Postal Code     0
Region          0
Category        0
Sub-Category    0
Sales           0
Quantity        0
Discount        0
Profit          0
dtype: int64
```

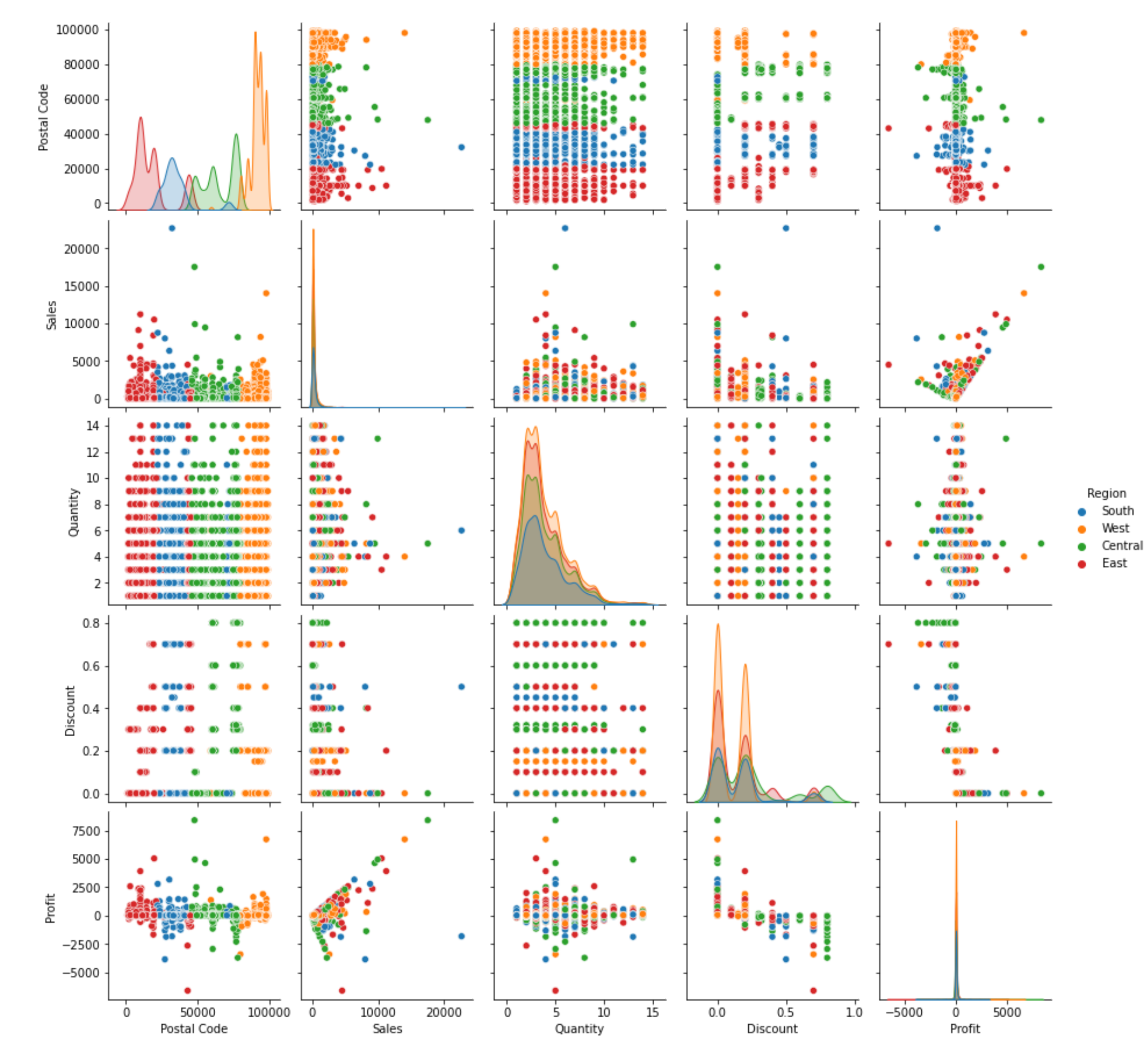
In [18]:
```
sns.pairplot(df)#Pairwise plot using Seaborn library
```

Out[18]:
```
<seaborn.axisgrid.PairGrid at 0x2b139a6ad00>
```



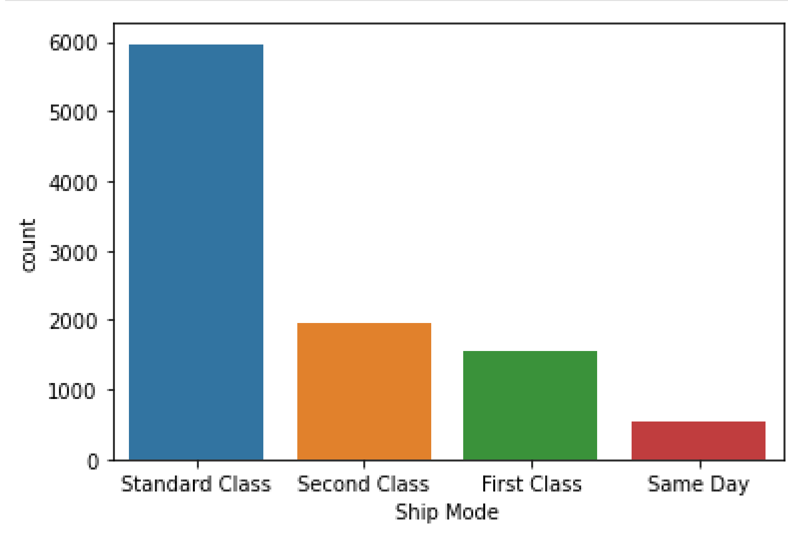
In [19]:
```
sns.pairplot(df,hue='Region')
```

Out[19]:
```
<seaborn.axisgrid.PairGrid at 0x2b13fb7e550>
```

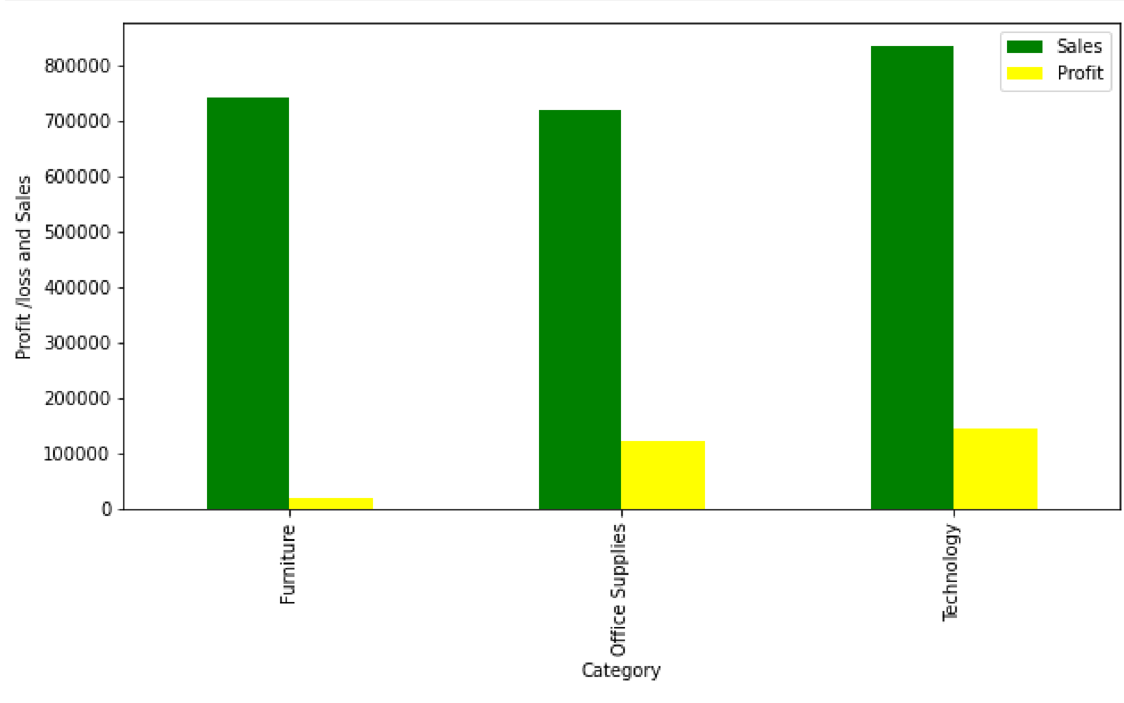


Which shipping mode is more widely used?

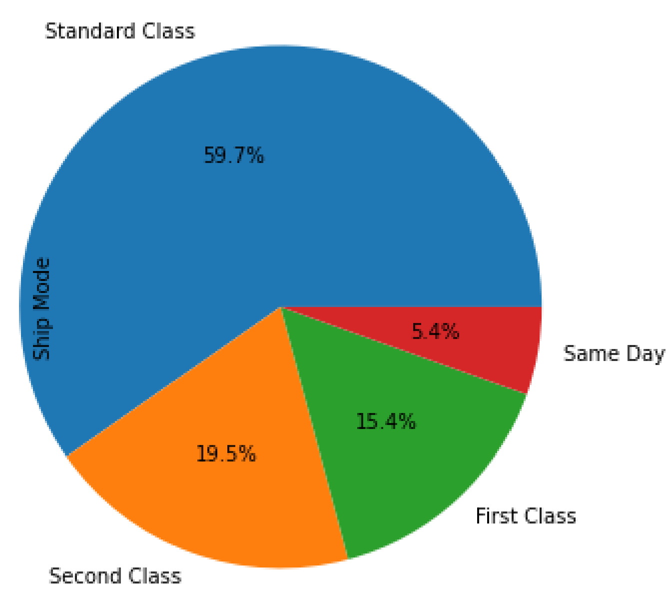
In [21]:
```
sns.countplot(x=df['Ship Mode'],order=df['Ship Mode'].value_counts().ir
plt.show()
```



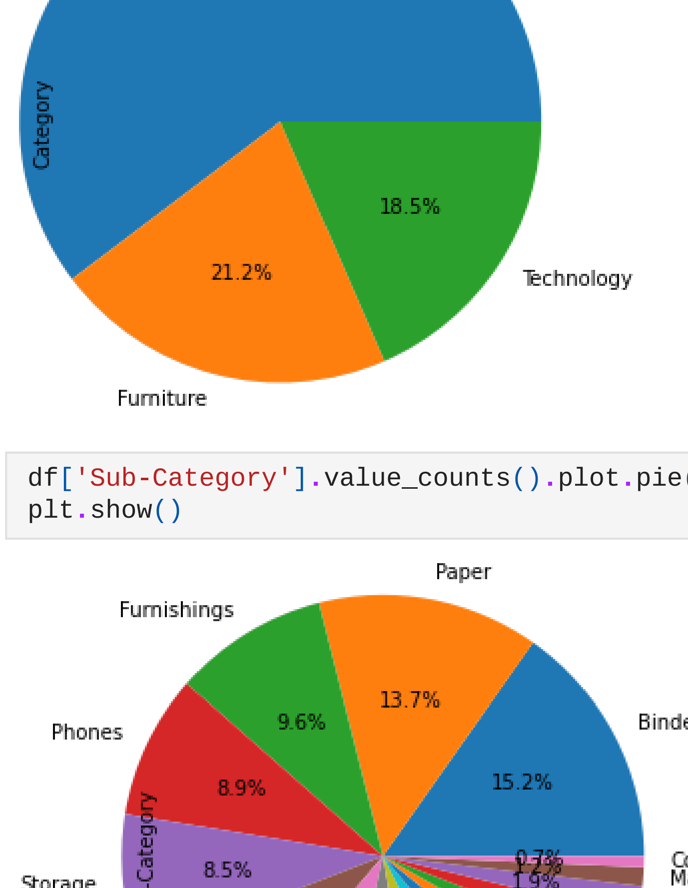
In [22]:
```
df.groupby("Category")[[ 'Sales','Profit']].sum().plot.bar(color = ["gr
plt.ylabel("Profit /loss and Sales")
plt.show()
```



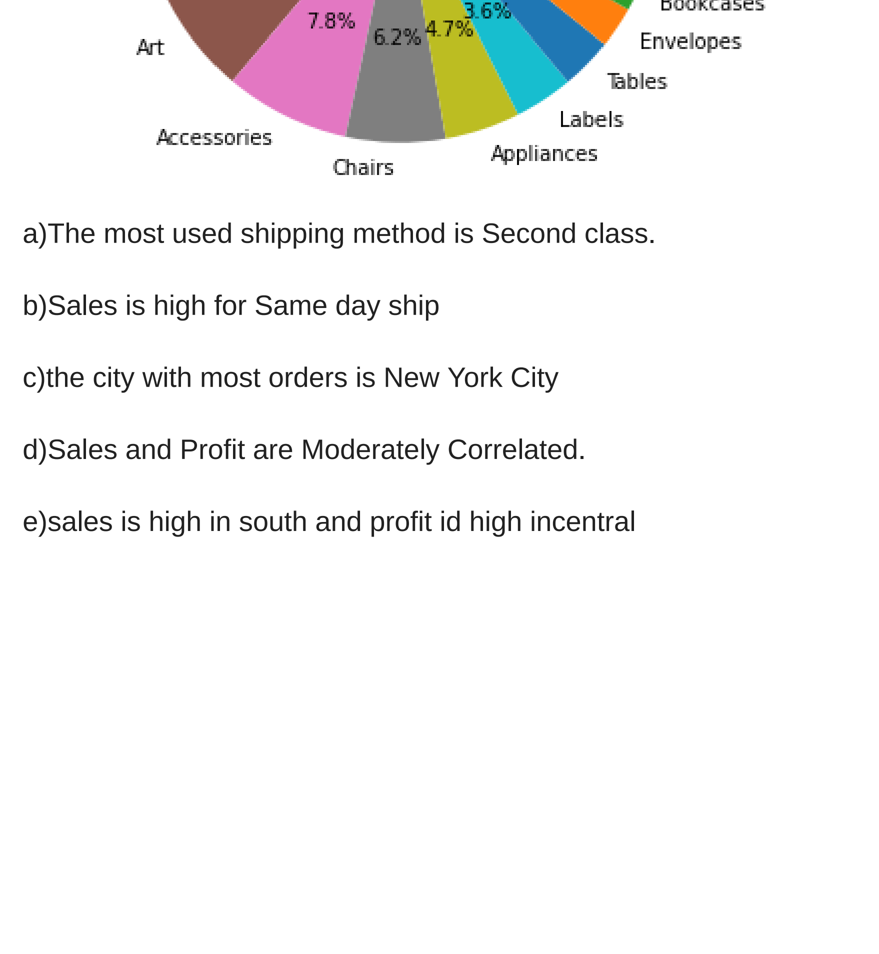
In [23]:
```
df['Ship Mode'].value_counts().plot.pie(autopct='%0.1f%%',radius=1.5)
plt.show()
```



In [24]:
```
df['Category'].value_counts().plot.pie(autopct='%0.1f%%',radius=1.5)
plt.show()
```



In [25]:
```
df['Sub-Category'].value_counts().plot.pie(autopct='%0.1f%%', radius=1.
plt.show()
```



a)The most used shipping method is Second class.

b)Sales is high for Same day ship

c)the city with most orders is New York City

d)Sales and Profit are Moderately Correlated.

e)sales is high in south and profit id high incentral