Imports

In [9]:
```
import pandas as pd
import numpy as np
from matplotlib import pyplot
from scipy.stats import pearsonr
```

Setting up data

In [10]:
```
dfA = pd.read_csv('p3a.csv', names=["A","B"])
dfB = pd.read_csv('p3b.csv', names=["A","B"])   #dfB and dfA are abreviations of
```

Part A

In [11]:
```
corr, pval = pearsonr(dfA.loc[:,"A"],dfA.loc[:,"B"]);
print(corr)
print(pval)
```

```
0.3808750357786363
1.0409455186803682e-83
```

Null Hypthesis: No Association Alternative Hypothesis: There is an Association alpha: 0.05

Since the p-val 1.0409455186803682e-83 is way less than alpha we **reject the null hypothesis** showing that there is strong statitical evidence to show that there is an association.

However the correlation (magnitude) is quite weak, haivng a pearsons correlation of only 0.38 Since the pearsons correlation is positive, the association is in the positive direction (positve sloped)

Part B

In [12]:
```
corr, pval = pearsonr(dfB.loc[:,"A"],dfB.loc[:,"B"]);
print(corr)
print(pval)
```

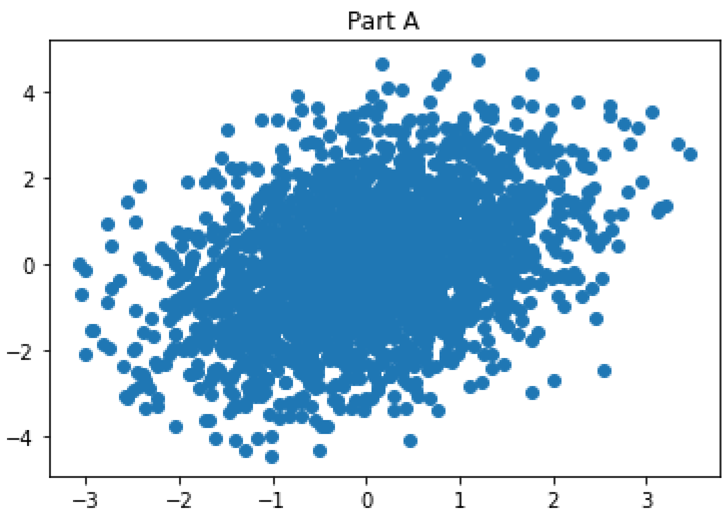
```
0.9312196333264213
3.7373210084392476e-49
```

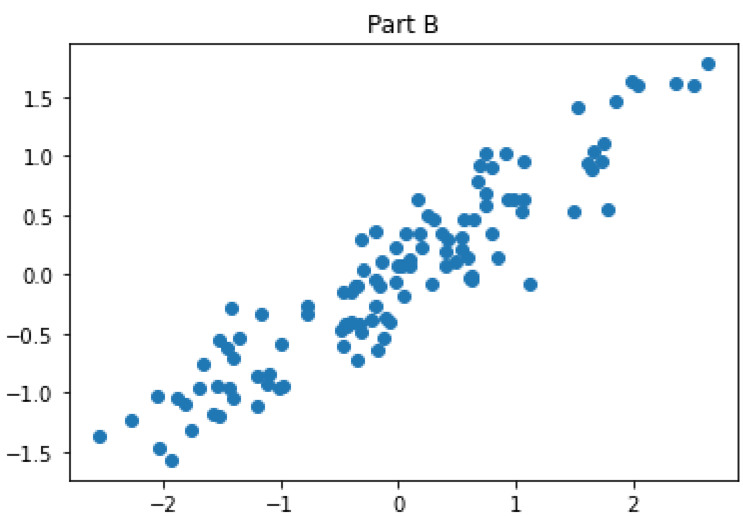
Since Part A has a lower pearsons correlations than Part B (0.38<0.93) **part B has a stronger correlation**

Since Part B has a lower Pval than Part A (0.38<0.93) **part A has a stronger association**

In [13]:

```
# Graphing data
pyplot.title("Part A")
pyplot.scatter(dfA.loc[:,"A"],dfA.loc[:,"B"])
pyplot.show()
pyplot.title("Part B")
pyplot.scatter(dfB.loc[:,"A"],dfB.loc[:,"B"])
pyplot.show()
```





By looking at the graphs, Part B seems to have a stronger association. However, since part A has just way more data points, the data shows that it has a stronger association because more points show that it is less likely for this to happen of chance thus lowering the p value.

In [ ]: