```python
# Program 4
# MSBA 605
# This program explores the use of files for
# data storage, statistical correlation, and
# matplotib for simple data visualization. It
# demonstrates Correlate Arcade Revenue and CS Doctorates
# of Arcade CSV file and shows a graph.

import math #imports the math functions
import matplotlib.pyplot as plt

def mean(alist): #defines the mean
    mean = sum(alist) / len(alist)
    return mean

def standardDev(alist): #defines the standard deviation
    theMean = mean(alist)

    sum = 0
    for item in alist: #sum the difference
        difference = item - theMean
        diffsq = difference ** 2
        sum = sum + diffsq

    sdev = math.sqrt(sum/(len(alist)-1))
    return sdev

def correlation(xlist, ylist): #defines variables int he graph
    xbar = mean(xlist)
    ybar = mean(ylist)
    xstd = standardDev(xlist)
    ystd = standardDev(ylist)
    num = 0.0
    for i in range(len(xlist)): #beginning of the for loop here
        num = num + (xlist[i]-xbar) * (ylist[i]-ybar)
    corr = num / ((len(xlist)-1) * xstd * ystd)
    return corr
```

```
39
40 # Start making edits here
41
42 def correlateArcadeDoctorates(filename): #gets the file
43     arcadeFile = filename
44
45     headerLine = arcadeFile.readline() # Read headers
46
47     headerList = headerLine.split(',')
48     arcadeIndex = headerList.index("Arcade Revenue") # Find Arcade Revenue column
49     doctoratesIndex = headerList.index("CS Doctorates\n") #pull in CS Doctorates from file
50
51     arcadeList = []
52     doctoratesList = []
53     for aLine in arcadeFile:
54         rowData = aLine.split(',')
55         arcadeList.append(float(rowData[arcadeIndex])) #data must be type float
56         doctoratesList.append(float(rowData[doctoratesIndex]))
57
58     arcadeFile.close() #remember to close the file when done
59
60     corr = correlation(arcadeList, doctoratesList)
61
62     plt.scatter(arcadeList, doctoratesList) #makes the final scatter plot
63     plt.title("Arcade Revenue vs CS Doctorates\nr = %.5f" % (corr))
64     plt.xlabel("Arcade Revenue")
65     plt.ylabel("CS Doctorates")
66     plt.show()
67
68     return corr
69
70 # Test correlation function
71 filename = open("Arcade-CSphd.csv", "r") #pulls specifically from this location
72 corr = correlateArcadeDoctorates(filename)
```

Results