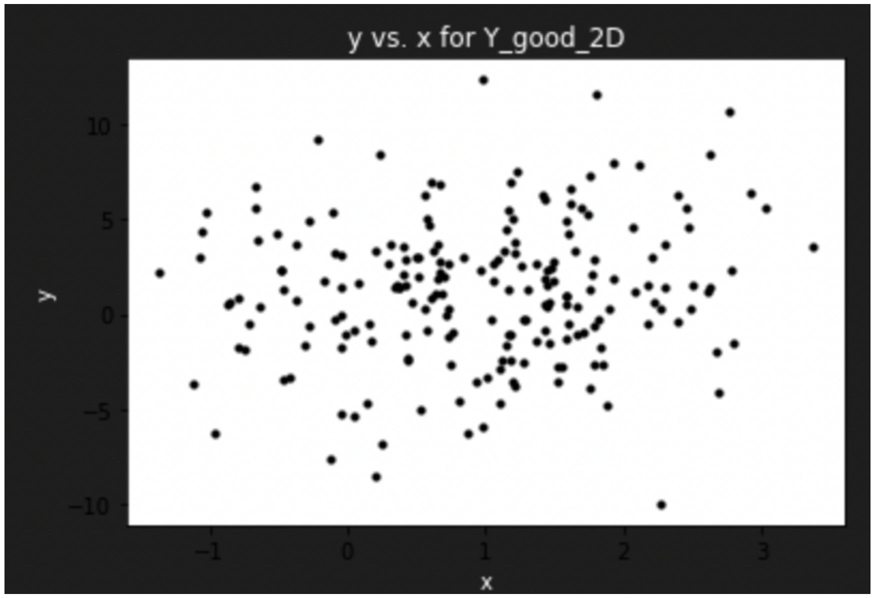
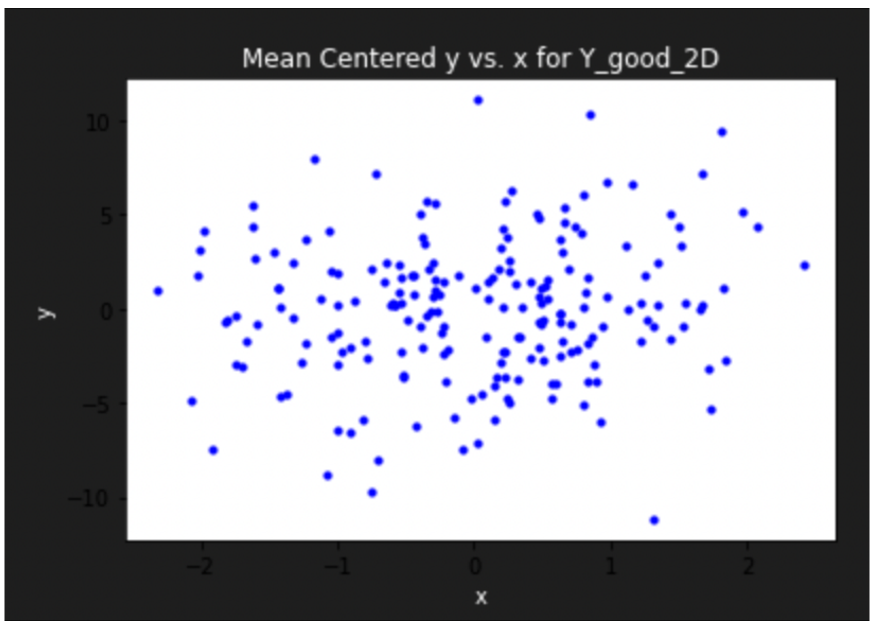
1.

a.



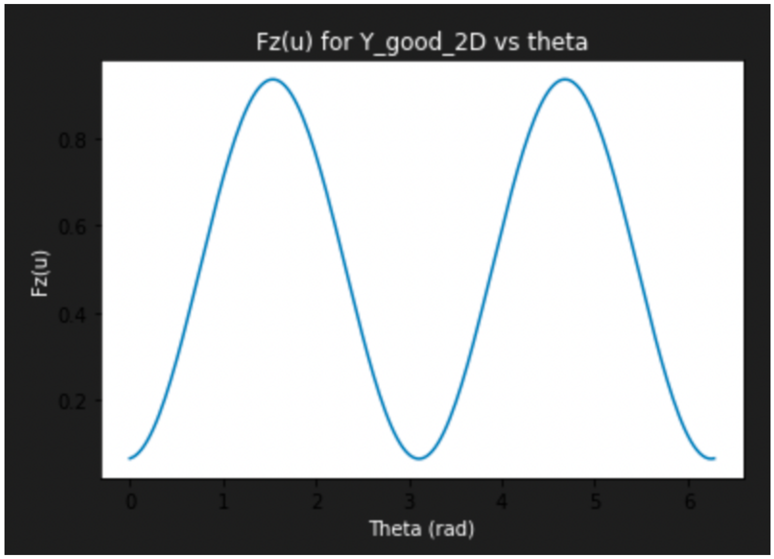

b.




c.

```
Fz(1,0):  0.06653629329652416
Fz(0,1):  0.06920533833732008
```

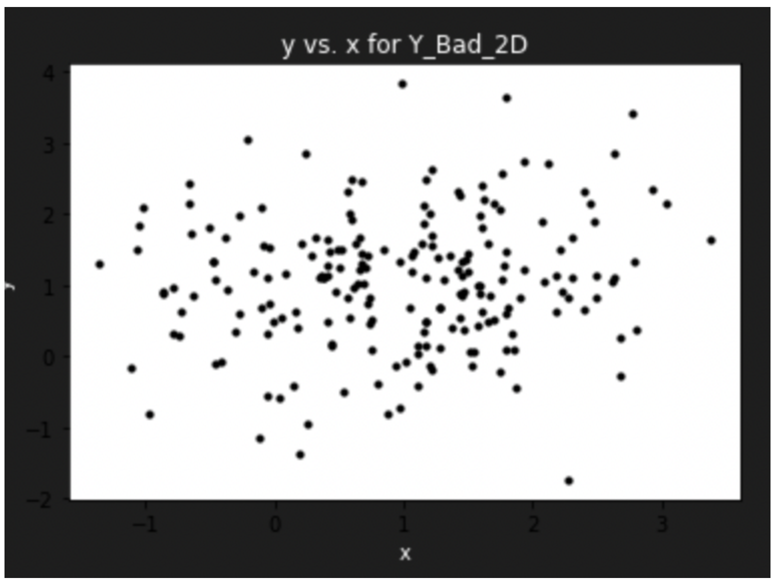
d. Y



Fz(u) for Y_good_2D vs theta
Fz(u)
0.2
0.4
0.6
0.8
0
1
2
3
4
5
6
Theta (rad)

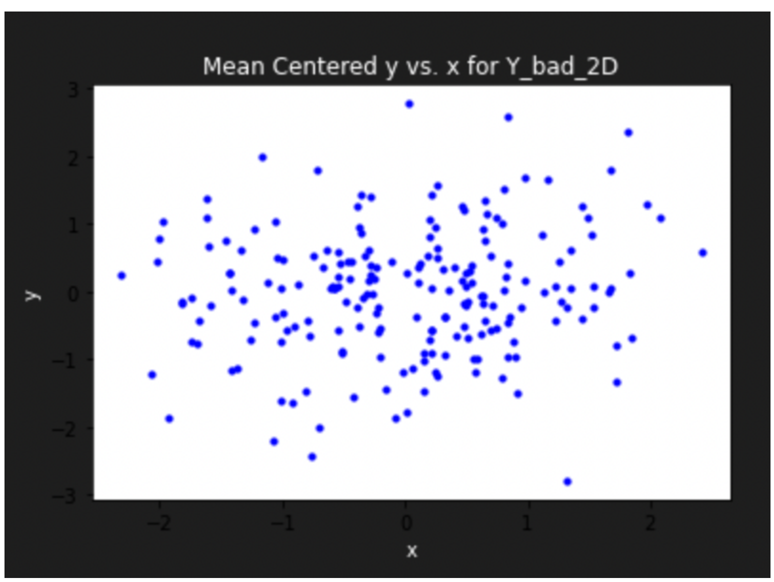

2.

a.



y vs. x for Y_Bad_2D
y
−2
−1
0
1
2
3
4
−1
0
1
2
3
x


b.

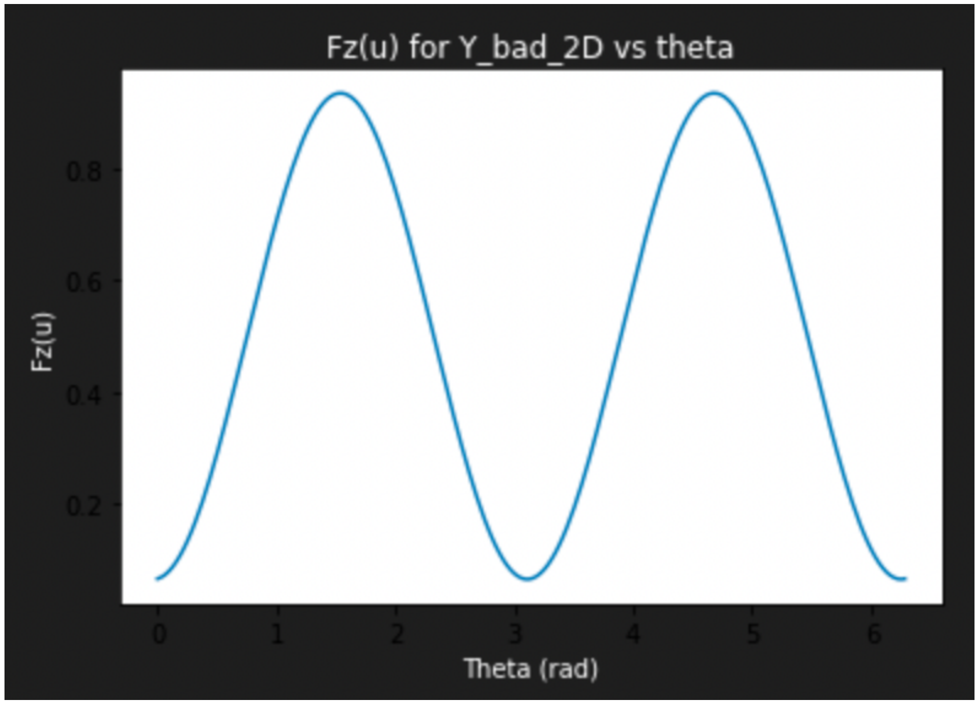


Mean Centered y vs. x for Y_bad_2D
y
−3
−2
−1
0
1
2
3
−2
−1
0
1
2
x

c.

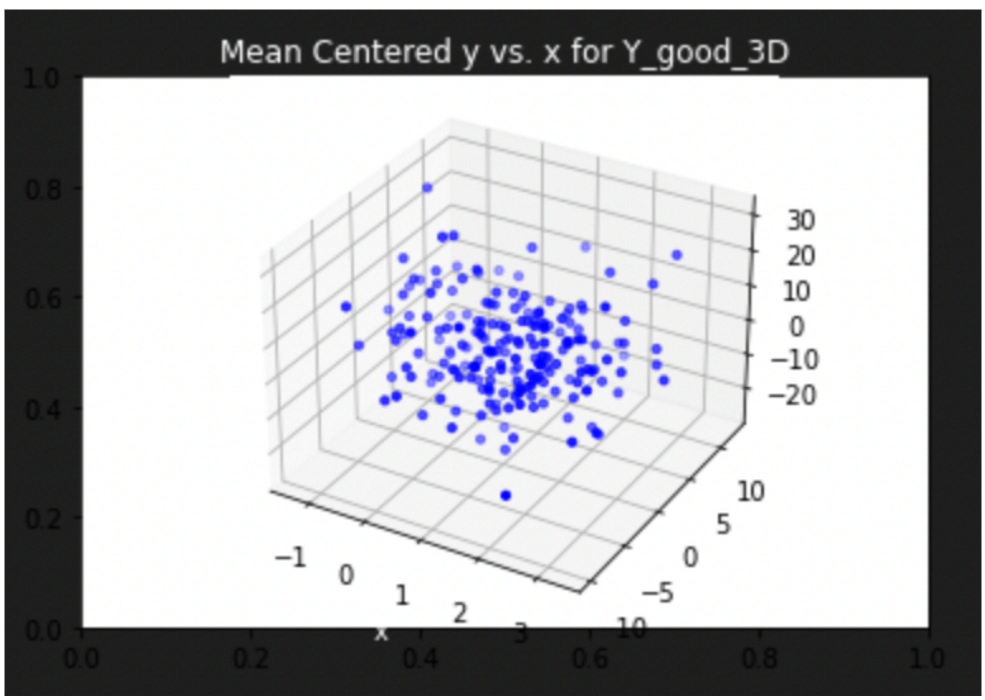
```
Fz(1,0):  0.5328118392195119
Fz(0,1):  0.536360223056589
```

d.




3.

**a. Results for data set Y_good_3D:**



Covariance Matrix:

```
[[ 1.88780262  1.58144622 -0.32499922]
 [ 1.58144622 15.12326145  4.06254879]
 [-0.32499922  4.06254879 95.70291403]]
```

Eigenvalues:

```
[ 1.69460381 15.11151903 95.90785525]
```

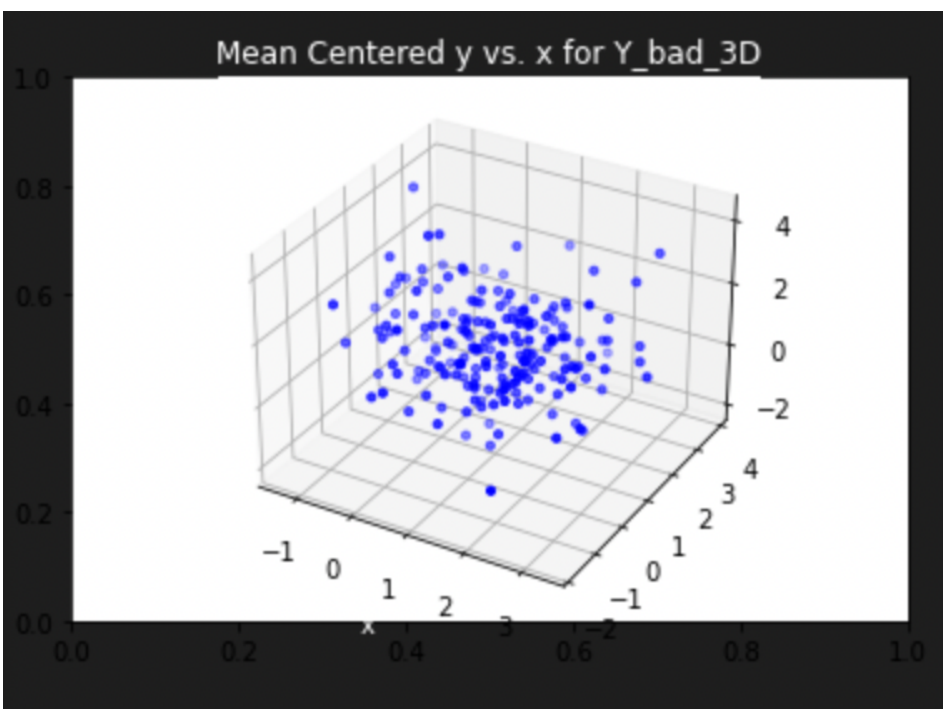
Eigenvectors:

```
[[-0.99279471  0.11979929 -0.00260839]
 [ 0.11951887  0.99156331  0.05017409]
 [-0.00859721 -0.04950081  0.99873708]]
```

A dimension reduction is appropriate because there are two large eigenvalues that account for a majority of the data out of three total eigenvalues, so we can reduce the dimension.

```
New span of the appropriate collection of principal components:
[[-1.68239369e+00  2.02537135e-01 -1.45688571e-02]
 [ 1.81034919e+00  1.49840279e+01 -7.48032494e-01]
 [-2.50165289e-01  4.81208890e+00  9.57867314e+01]]
```

**b. Results for data set Y_bad_3D:**



Covariance Matrix:

```
[[1.88780262 1.11286989 0.81427122]
 [1.11286989 1.96370323 1.11488414]
 [0.81427122 1.11488414 2.12605141]]
```

Eigenvalues:

```
Eigenvalues:  [4.02823486 1.20091988 0.74840251]
```

Eigenvectors:

```
[[-0.53801028 -0.64803782  0.53906579]
 [-0.60626989 -0.14682443 -0.78158775]
 [-0.58564644  0.7473216   0.31389277]]
```

A dimension reduction is appropriate in this case because one of the eigenvalues is significantly larger than the other two and accounts for a majority percentage of the data, so it is appropriate to reduce the data set by one or two dimensions.

```
New span of the appropriate collection of principal components:
[[-2.16723176 -2.44219752 -2.35912142]
 [-0.7782415  -0.17632437  0.89747337]
 [ 0.40343819 -0.58494223  0.23491814]]
```

c. **Results for data set Y_good_5D**

Covariance Matrix:

```
[[ 1.88780262e+00  1.58144622e+00 -3.24999225e-01  5.10811966e-01
   3.09550703e-01]
 [ 1.58144622e+00  1.51232614e+01  4.06254879e+00 -1.81578505e+00
  -5.27269425e+00]
 [-3.24999225e-01  4.06254879e+00  9.57029140e+01 -6.15901485e+00
   3.16462010e+01]
 [ 5.10811966e-01 -1.81578505e+00 -6.15901485e+00  2.59645367e+02
  -3.63947470e+01]
 [ 3.09550703e-01 -5.27269425e+00  3.16462010e+01 -3.63947470e+01
   6.03362722e+02]]
```

Eigenvalues:

```
Eigenvalues:  [609.22160365 255.90872784   1.68995082  15.01842188  93.88336319]
```

Eigenvectors:

```
[[-3.63726609e-04 -2.08762581e-03  9.92613567e-01  1.21271727e-01
  -2.64278451e-03]
 [ 8.06255191e-03  1.00931857e-02 -1.20912796e-01  9.91067094e-01
   5.47652656e-02]
 [-6.23549084e-02  1.76239932e-02  9.21669886e-03 -5.36085264e-02
   9.96414792e-01]
 [ 1.04391276e-01 -9.94200180e-01 -2.91146327e-03  7.56395605e-03
   2.45514343e-02]
 [-9.92546820e-01 -1.05589610e-01 -2.23117277e-03  1.21694873e-02
  -5.95698735e-02]]
```

A dimension reduction is also appropriate in this case because three of the five eigenvalues are significantly larger than the other two and account for a majority percentage of the data. Therefore, it would be appropriate to reduce the data set by one or two dimensions

```
New span of the appropriate collection of principal components:
[[-2.21590108e-01  4.91188080e+00 -3.79879573e+01  6.35974208e+01
  -6.04680965e+02]
 [-5.34241666e-01  2.58293430e+00  4.51013368e+00 -2.54424503e+02
  -2.70213029e+01]
 [ 1.67746811e+00 -2.04336679e-01  1.55757678e-02 -4.92022974e-03
  -3.77057226e-03]
 [ 1.82130996e+00  1.48842637e+01 -8.05115466e-01  1.13598683e-01
   1.82766494e-01]
 [-2.48113498e-01  5.14154732e+00  9.35467718e+01  2.30497122e+00
  -5.59262007e+00]]
```

d. **Results for data set Y_bad_5D**

Covariance Matrix:

```
[[1.88780262 1.11286989 0.81427122 0.92881138 0.93079415]
 [1.11286989 1.96370323 1.11488414 1.04546568 0.91072299]
 [0.81427122 1.11488414 2.12605141 0.97357443 1.02000827]
 [0.92881138 1.04546568 0.97357443 2.10742131 0.8582378 ]
 [0.93079415 0.91072299 1.02000827 0.8582378  1.76308958]]
```

Eigenvalues:

```
Eigenvalues:  [5.86567924 0.69841954 1.21775529 1.13262609 0.933588  ]
```

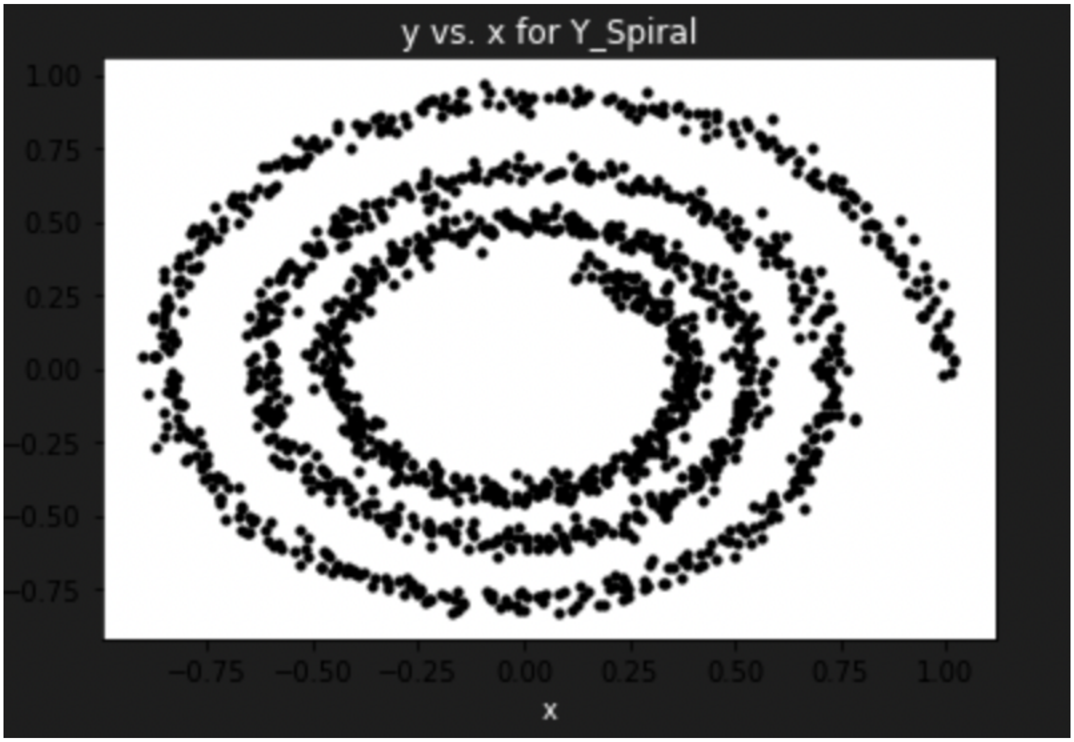
Eigenvectors:

```
[[-0.42980378 -0.56783502  0.49476559  0.49802807  0.00267452]
 [-0.47306028  0.5909886   0.10650957  0.16310677 -0.62369732]
 [-0.46503514 -0.40573686 -0.75514816 -0.11271557 -0.19017459]
 [-0.45255481 -0.02960559  0.36982969 -0.79264084  0.17106768]
 [-0.41281853  0.40348107 -0.19193746  0.29048223  0.73862302]]
```

One of the eigenvalues is significantly larger than the other four eigenvalues and accounts for a majority percentage of the data, so it would be appropriate to reduce this data set by anywhere from 1 to 4 dimensions.

```
New span of the appropriate collection of principal components:
[[-2.52109111e+00 -2.77481984e+00 -2.72774698e+00 -2.65454138e+00
  -2.42146111e+00]
 [-3.96587076e-01  4.12757983e-01 -2.83374553e-01 -2.06771233e-02
   2.81799062e-01]
 [ 6.02503409e-01  1.29702591e-01 -9.19585663e-01  4.50362059e-01
  -2.33732854e-01]
 [ 5.64079585e-01  1.84738981e-01 -1.27664592e-01 -8.97765693e-01
   3.29007749e-01]
 [ 2.49689765e-03 -5.82276331e-01 -1.77544715e-01  1.59706732e-01
   6.89569589e-01]]
```

4.)




Projecting the data onto the first principal component might be a mistake because we have to use the mean-centered x and y values, which would not be representative of the data set as a whole due to its spiral-like geometry and whacky combination of negative and positive values in this spiral pattern.

However, it might be reasonable, nonetheless to describe this dataset as being “approximately one-dimensional” because we can project this data set onto a line through the origin based on the absolute values of the distance of the points from the origin, which would make for a good projection, thus allowing us to visualize the dataset as one-dimensional.