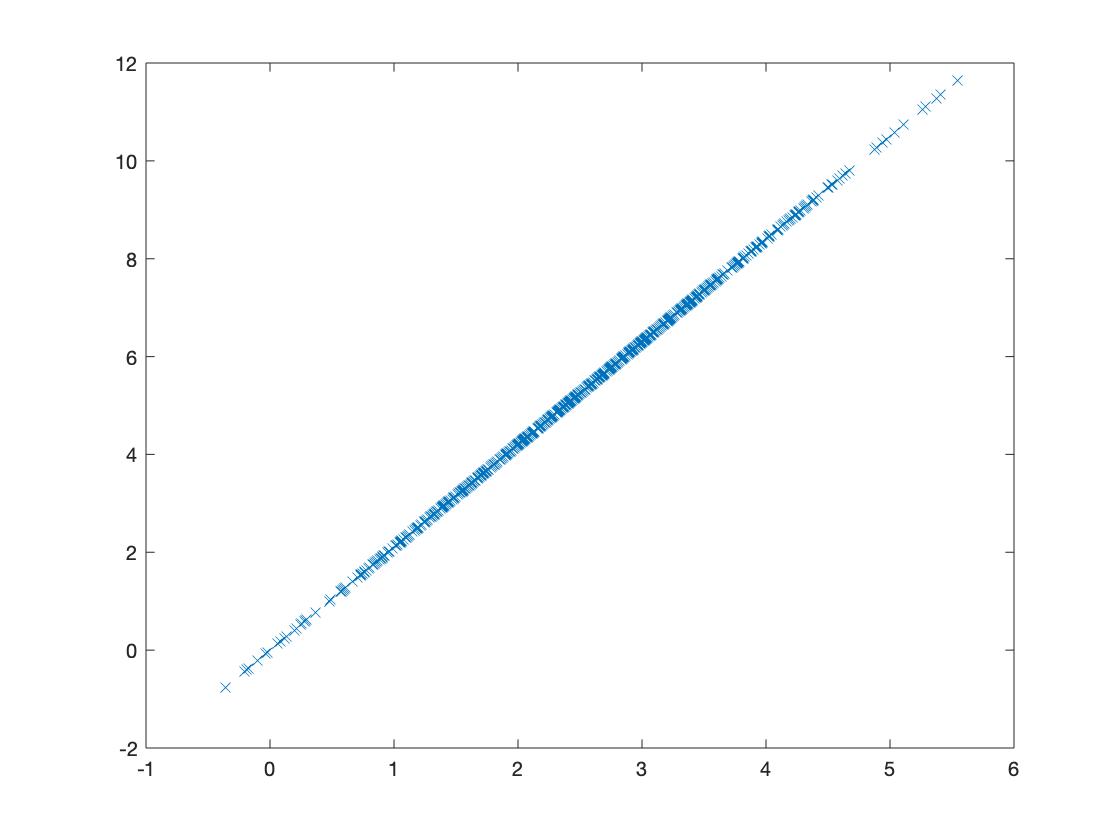
HW7 Ivan Isakov

a)

k=rank(X,X(1));

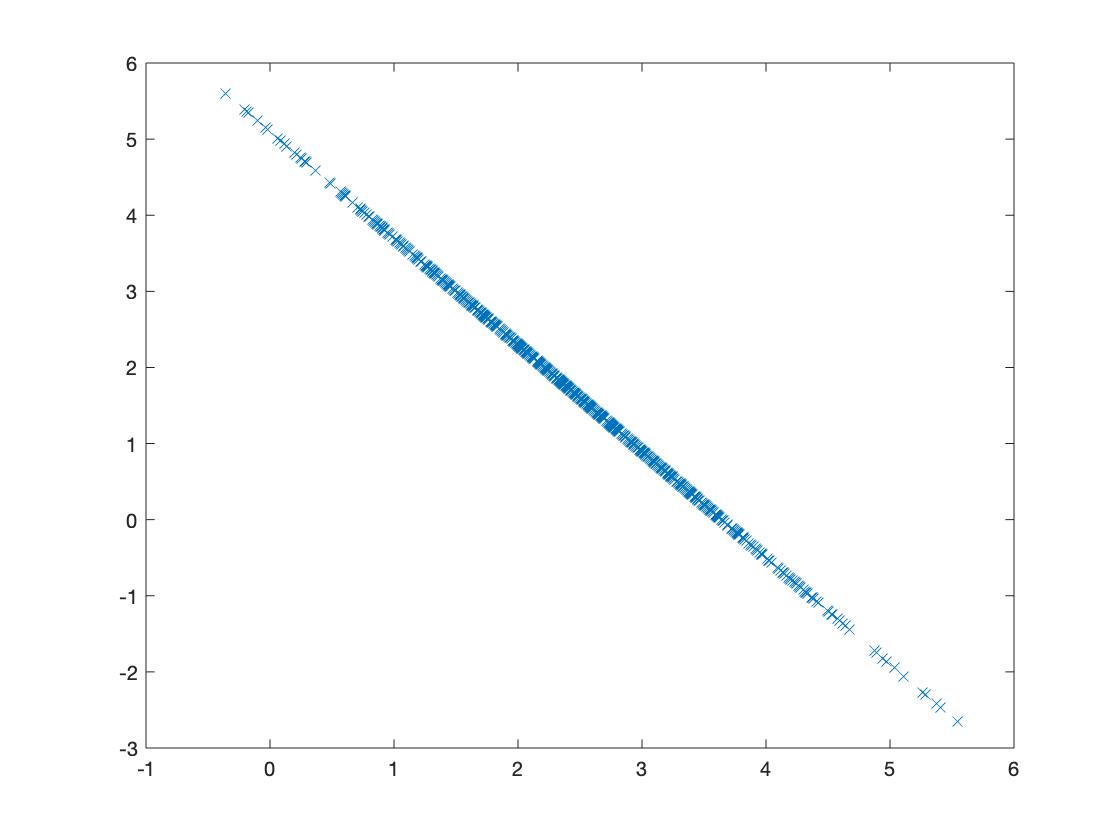
plot(X(:,1),X(:,k),'x')



b)

c=corr(X);

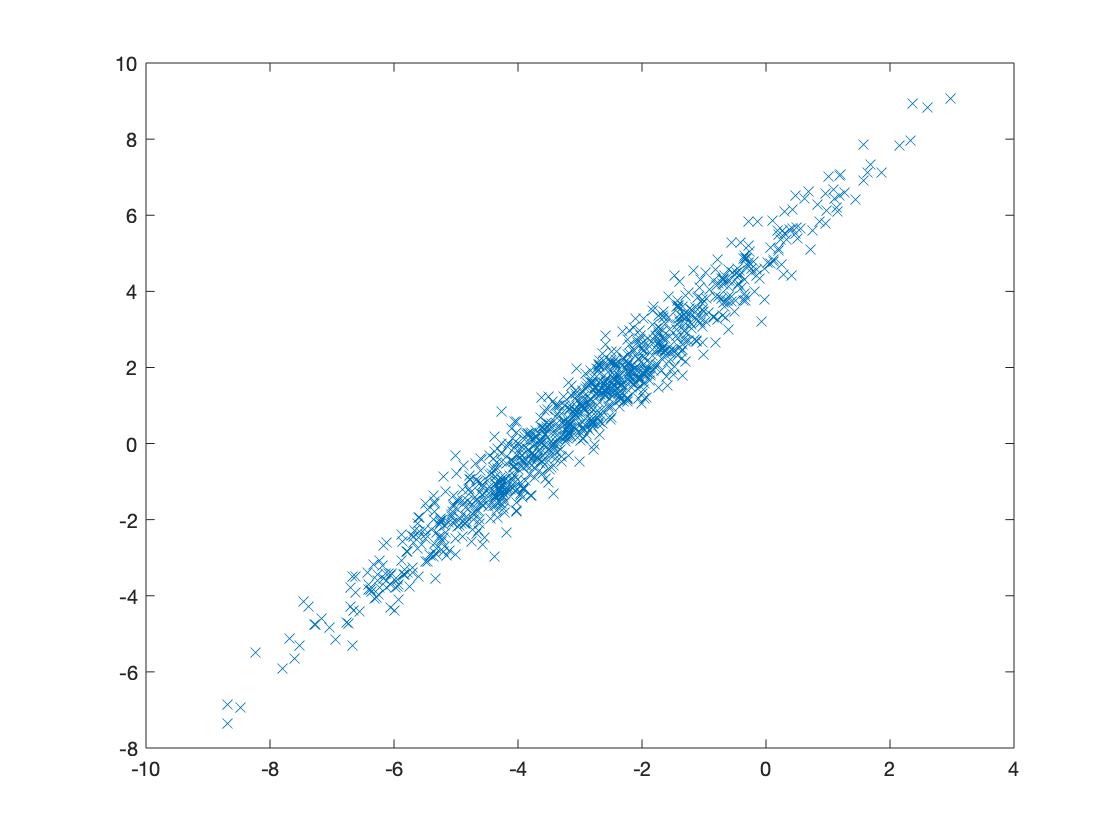
plot(X(:,1),X(:,3),'x')



The rank test missed it because the correlation coefficient is -1

c) c=corr(X);

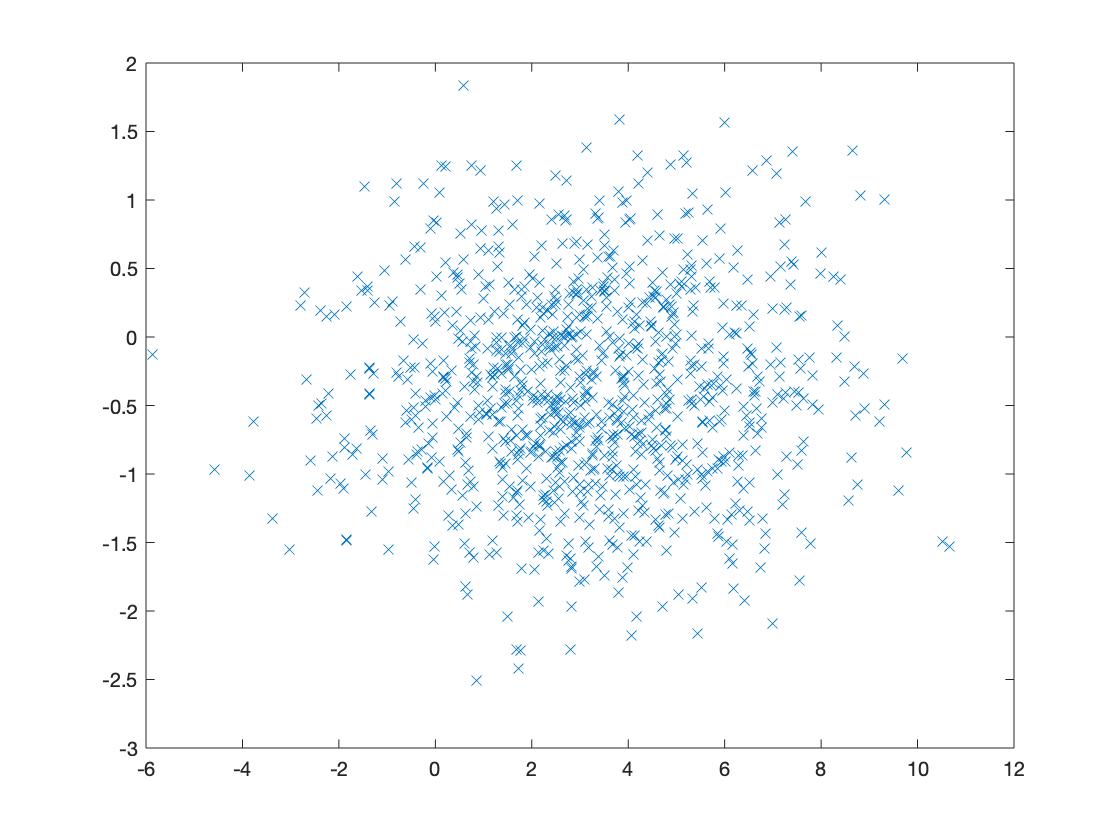
plot(X(:,5),X(:,7),'x')



The data is much more spread out from the line of best fit.

d) c=corr(X);

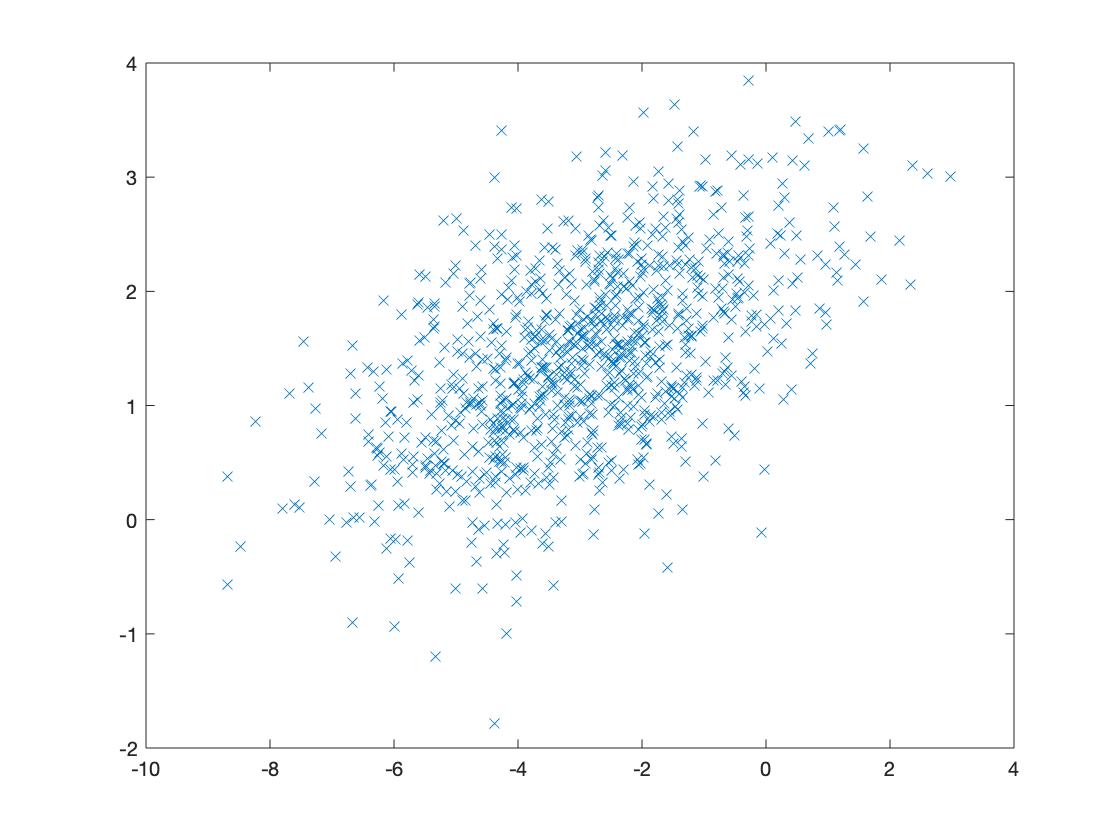
plot(X(:,2),X(:,4),'x')



As the correlation coefficient is very close to 0, the data is spread out all over the range.

e) c=corr(X);

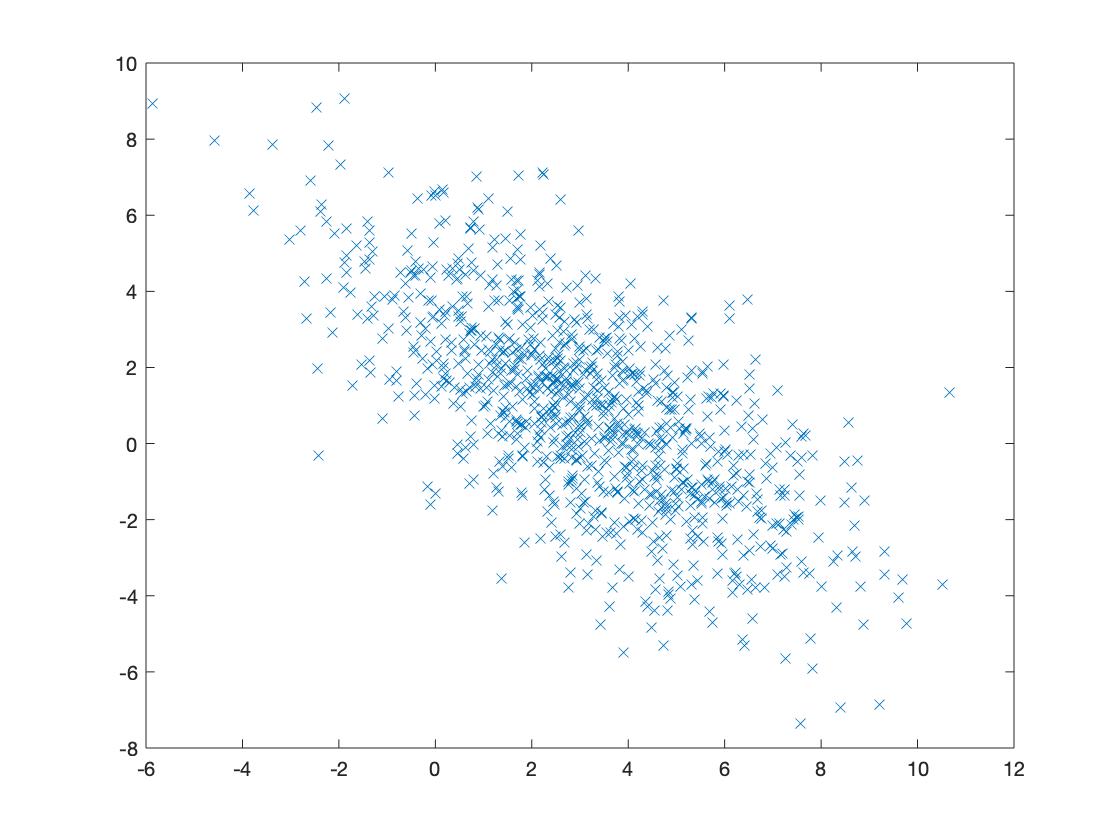
plot(X(:,5),X(:,8),'x')



The data is more concentrated around the line of best fit, but still is not very close to it.

f) c=corr(X);

plot(X(:,2),X(:,7),'x')



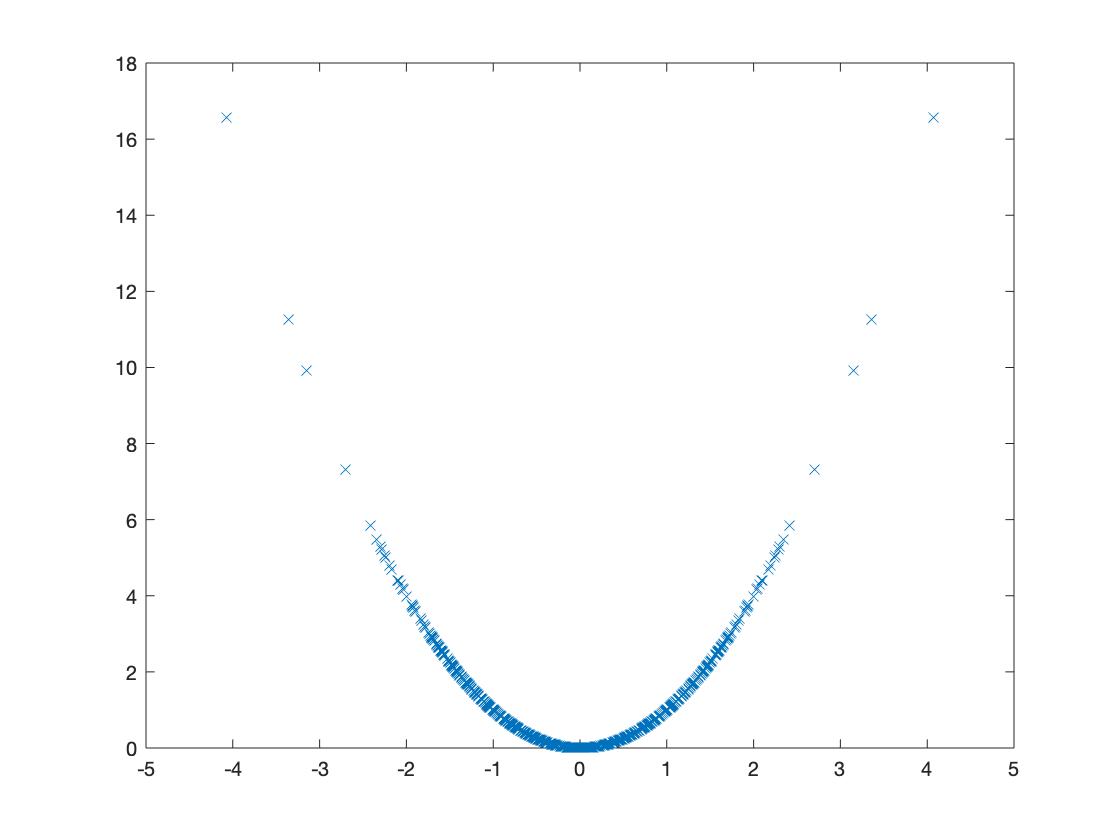
Exactly the same as e) but with a negative correlation.

g) cy=corr(Y(:,1),Y(:,2))

cy=-5.2055e-18

So we would expect to see data that is very far away from the line of best fit for a linear function. Because the coefficient is close to 0.

plot(Y(:,1),Y(:,2),'x')



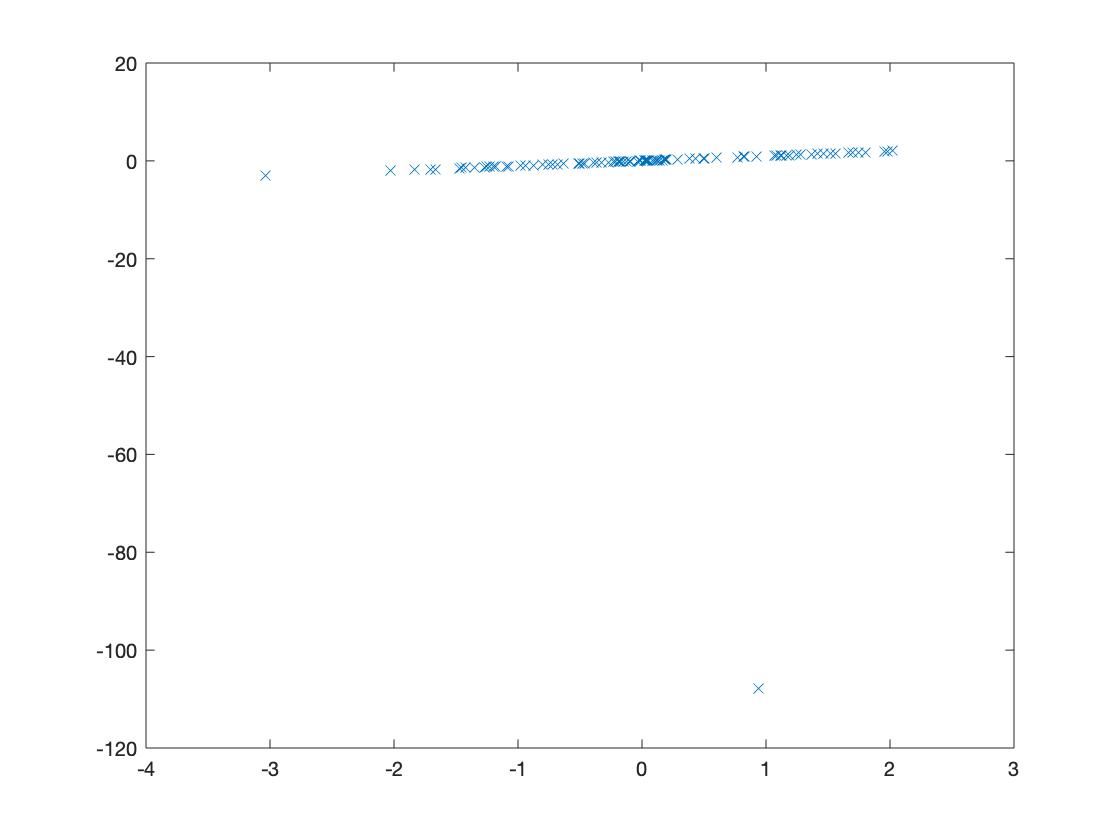
But we see a quadratic relationship between the first and the second column.

h)

cz=corr(Z(:,1),Z(:,2))

cz=-6.5309e-18 close to 0

so we would expect to see a data that is spread out a lot.



We see a very close linear relationship between the first and the second column with an outlier at (1,-105).

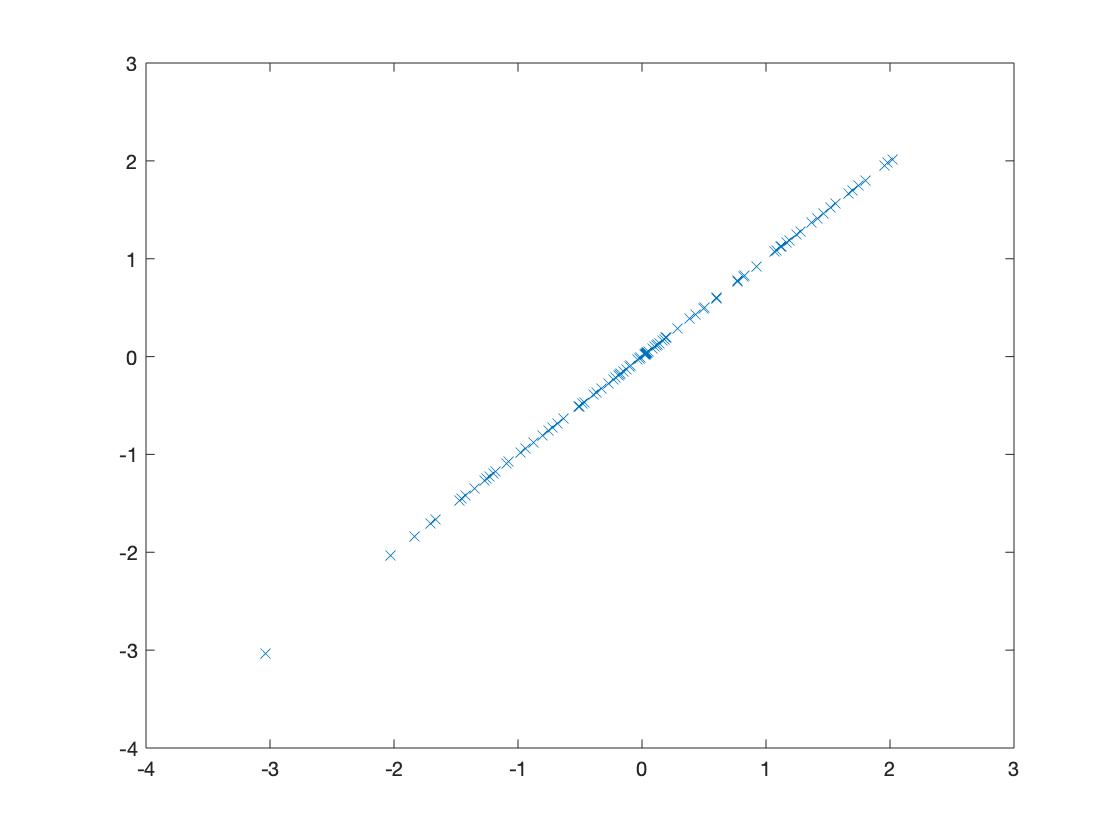
Removed the last row of Z:

Z(100,:)=[]

cz=corr(Z(:,1),Z(:,2))

cz=1

plot(Z(:,1),Z(:,2),'x')



We have removed the outlier and now the data is close to the line of best fit.