Time Series Analysis and Models Homework 1 Report

Edison Murairi

September 21, 2023

Problem 1

We are given

$$
\begin{aligned}
x &= [1, 2, 3, 4, 5] && (1)\\
y &= [1, 2, 3, 4, 5] && (2)\\
z &= [-1, -2, -3, -4, -5] && (3)\\
g &= [1, 1, 0, -1, -1, 0, 1] && (4)\\
h &= [0, 1, 1, 1, -1, -1, -1] && (5)
\end{aligned}
$$

1. The correlation coefficient between x and y is 1.0
2. The correlation coefficient between x and z is -1.0
3. The correlation coefficient between g and h is 0.0

Problem 2

Correlation coefficient:

$$r = \frac{\sum(x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum(x_t - \bar{x})^2}\sqrt{\sum(y_t - \bar{y})^2}} \tag{6}$$

1. The correlation coefficient between x and y:

$$
\begin{aligned}
\bar{x} &= \frac{1+2+3+4+5}{5} = \frac{15}{5} = 3\\
\bar{y} &= \frac{15}{5} = 3 && (7)
\end{aligned}
$$

   since we have $x_t = y_t$ for all t, we obtain

$$
\begin{aligned}
r_{x,y} &= \frac{\sum(x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum(x_t - \bar{x})^2}\sqrt{\sum(y_t - \bar{y})^2}}\\
&= \frac{\sum(x_t - \bar{x})^2}{\sum(x_t - \bar{x})^2} = 1 && (8)
\end{aligned}
$$

2. The correlation coefficient between x and z Now, we have $x_t = -z_t$ for all t. Hence $\bar{z} = -\bar{x} = -3$.

Therefore,

$$
\begin{aligned}
r_{x,z} &= \frac{\sum(x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum(x_t - \bar{x})^2}\sqrt{\sum(y_t - \bar{y})^2}} \\
&= \frac{\sum(x_t - \bar{x})\,(-x_t + \bar{x})}{\sum(x_t - \bar{x})^2\ \sum(-x_t + \bar{x})^2} \\
&= \frac{-\sum(x_t - \bar{x})(x_t - \bar{x})}{\sqrt{\sum(x_t - \bar{x})^2}\ \sqrt{\sum\left[-1\,(x_t - \bar{x})\right]^2}} \\
&= -\frac{\sum(x_t - \bar{x})^2}{\sum(x_t - \bar{x})^2} \\
&= -1
\end{aligned} \tag{9}
$$

3. The correlation coefficient between g and h.

$$
\begin{aligned}
\bar{g} &= \frac{1+1+0-1-1+0+1}{7} = \frac{1}{7} \\
\bar{h} &= \frac{0+1+1+1-1-1-1}{7} = 0
\end{aligned}
$$

$$
\sum(g_t - \bar{g})(h_t - \bar{h}) = (6/7) + (-1/7) + (-8/7) + (8/7) + (1/7) + (-6/7) = 0 \tag{10}
$$

Therefore, the correlation coefficient is $r_{g,h} = 0$.

All the calculations agree with the result from the codes (Problem 1).

Problem 3

Figure 1 shows the scatterplot of GDP (y-axis) and Sales (x-axis), with the correlation coefficient. The correlation coefficient is $r = -0.64$. This answer is consistent with the scatter plot because we see a negative correlation between the two variables, and the scatter plot is more or less around a straight line. Therefore, we expect some level of correlation, and the correlation is negative. Hence, $r = -0.64$ makes sense.

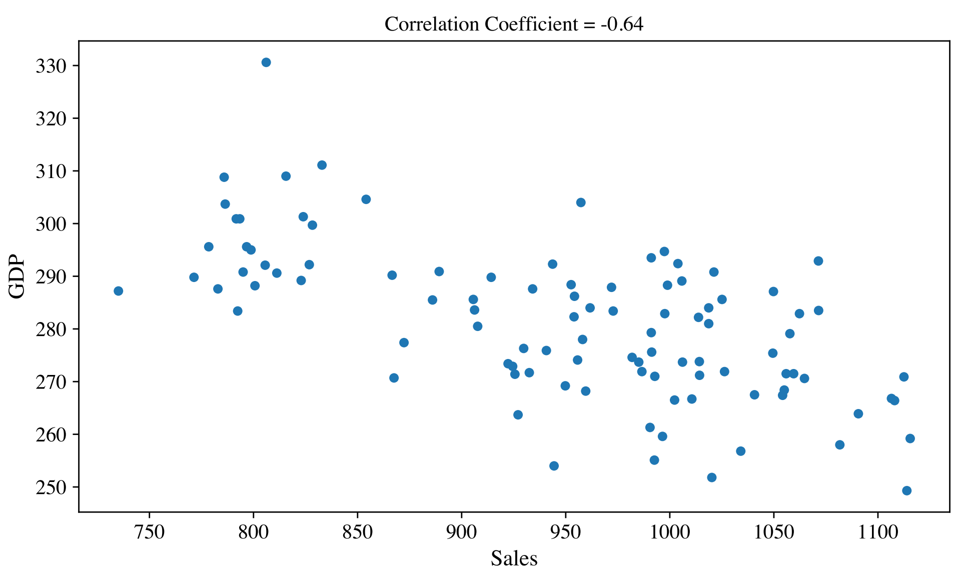

Figure 1: Scatter plot of Sales and GDP in the Tute 1 dataset.

Problem 4

Figure 3 shows the scatterplot of Ad Budget (y-axis) and Sales (x-axis), with the correlation coefficient. The correlation coefficient is $r = 0.91$. This value is consistent with the plot because the plot shows a strong positive correlation. As Sales increase, the GDP also increases, and the data are more concentrated on a straight line.

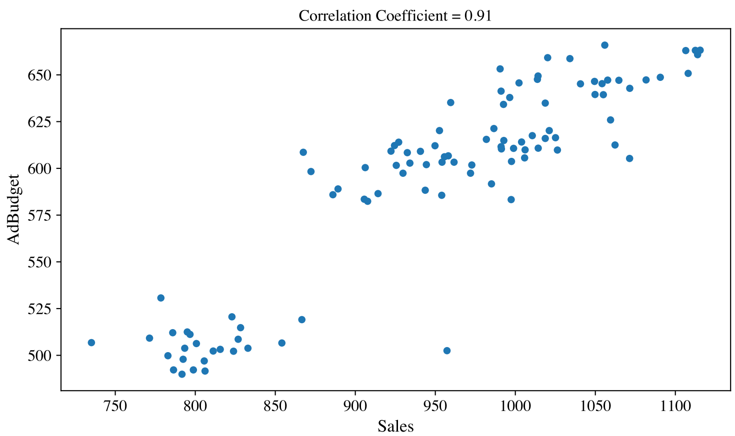

Figure 2: Scatter plot of Sales and AdBudget in the Tute 1 dataset.

Problem 5

Figure 3 shows the scatterplot of Ad Budget (y-axis) and Sales (x-axis), with the correlation coefficient. The correlation coefficient is $r = -0.77$. This answer is consistent with the scatter plot because we see a negative correlation between the two variables, and the scatter plot is more or less around a straight line. Therefore, we expect some level of correlation, and the correlation is negative. Hence, $r = -0.77$ makes sense.

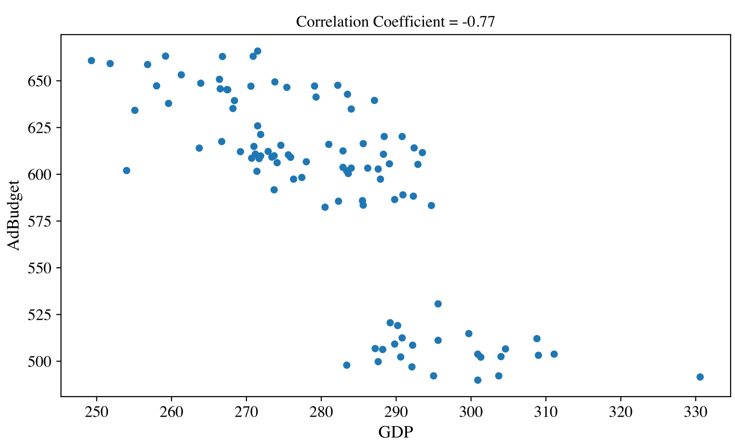

Figure 3: Scatter plot of GDP and AdBudget in the Tute 1 dataset.

Problem 6

1. kind = 'kde': The plot in Figure 4 shows the relation between every pair of variables. By choosing the option *kind = 'kde'*, each grid in the plot shows the Kernel Density Estimate (KDE) between the corresponding variables. In particular, the diagonal grids show the univariate KDE, while the off diagonal grids show the KDE for the two different variables. These KDE plots are consistent with the correlation coefficients we calculated in Problem 1. We can see that Sales and AdBudget are strongly and positively correlated. Sales and GDP are somewhat correlated, and the correlation is negative. Finally, Ad Budget and GDP are also correlated and the correlation is negative.

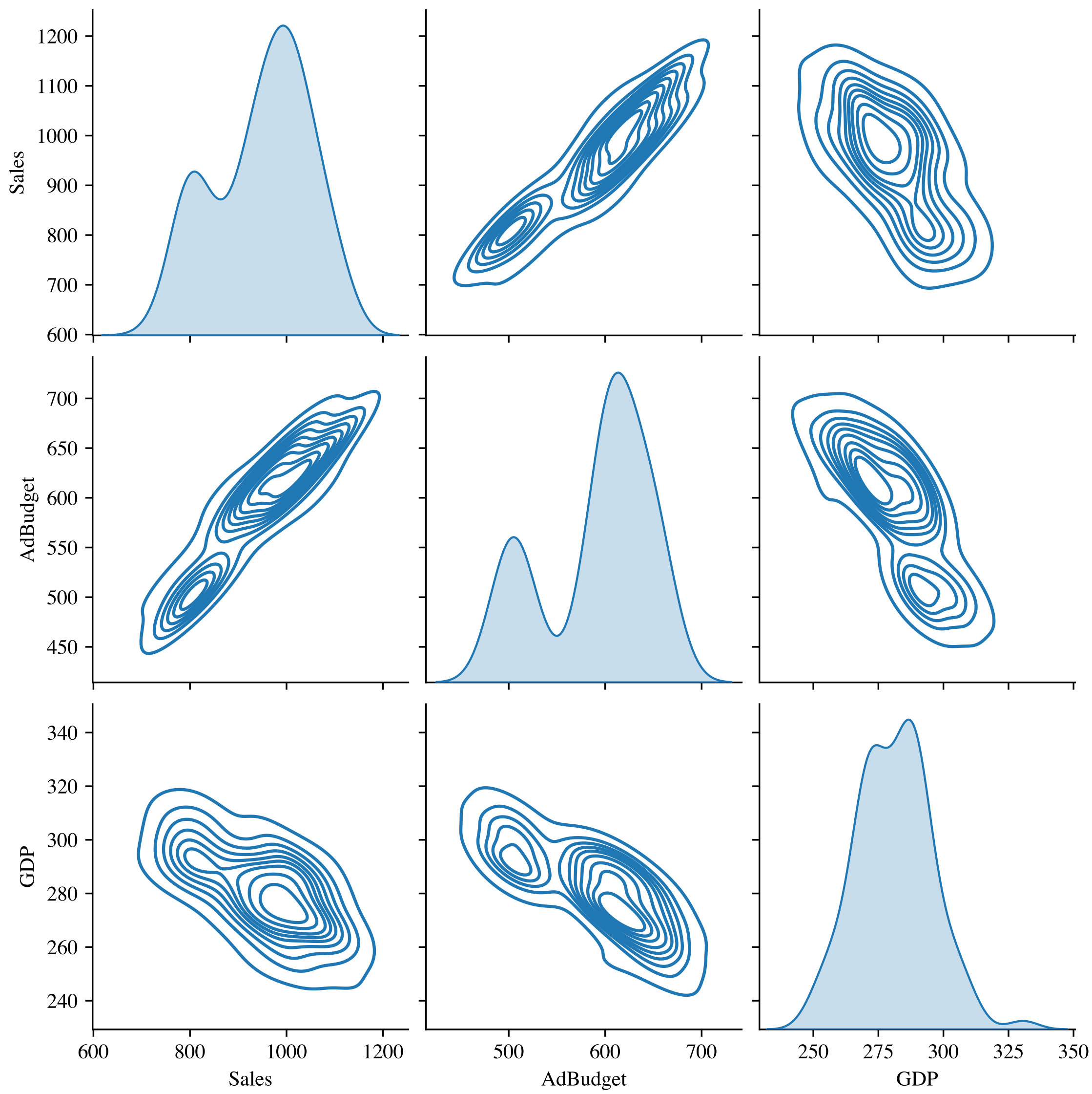

Figure 4: Pairplot of Tute 1 dataset using Kind = 'kde'

2. Kind = 'hist' The plot in Figure 5 shows the relation between every pair of variables. By choosing the option *kind = 'hist'*, each grid in the plot shows the joint histogram between the corresponding variables. In particular, the diagonal grids show the univariate histogram, while the off diagonal grids show the joint histogram for the two different variables. These plots are consistent with the correlation coefficients we calculated in Problem 1. We can see that Sales and AdBudget are strongly and positively correlated. Sales and GDP are somewhat correlated, and

the correlation is negative. Finally, Ad Budget and GDP are also correlated and the correlation is negative.

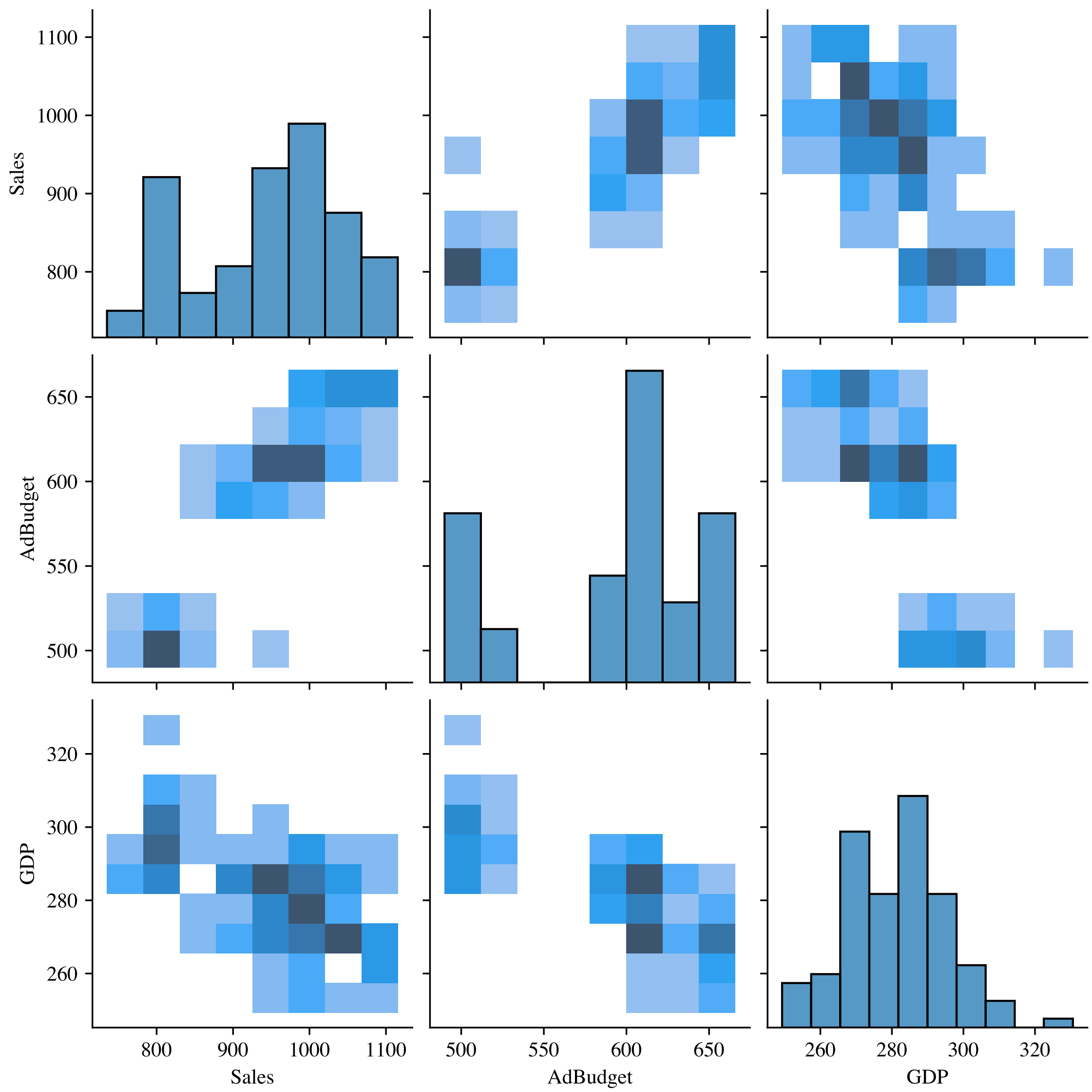

Figure 5: Pairplot of Tute 1 dataset using Kind = 'hist'

3. diag_kind = 'hist' The plot in Figure 6 shows the relation between every pair of variables. By choosing the option *diag_kind = 'hist'*, we set the diagonal grids to histograms and the off diagonal grids are left to scatterplot (the default in Seaborn). These plots are consistent with the correlation coefficients we calculated in Problem 1. We can see that Sales and AdBudget are strongly and positively correlated. Sales and GDP are somewhat correlated, and the correlation is negative. Finally, Ad Budget and GDP are also correlated and the correlation is negative.

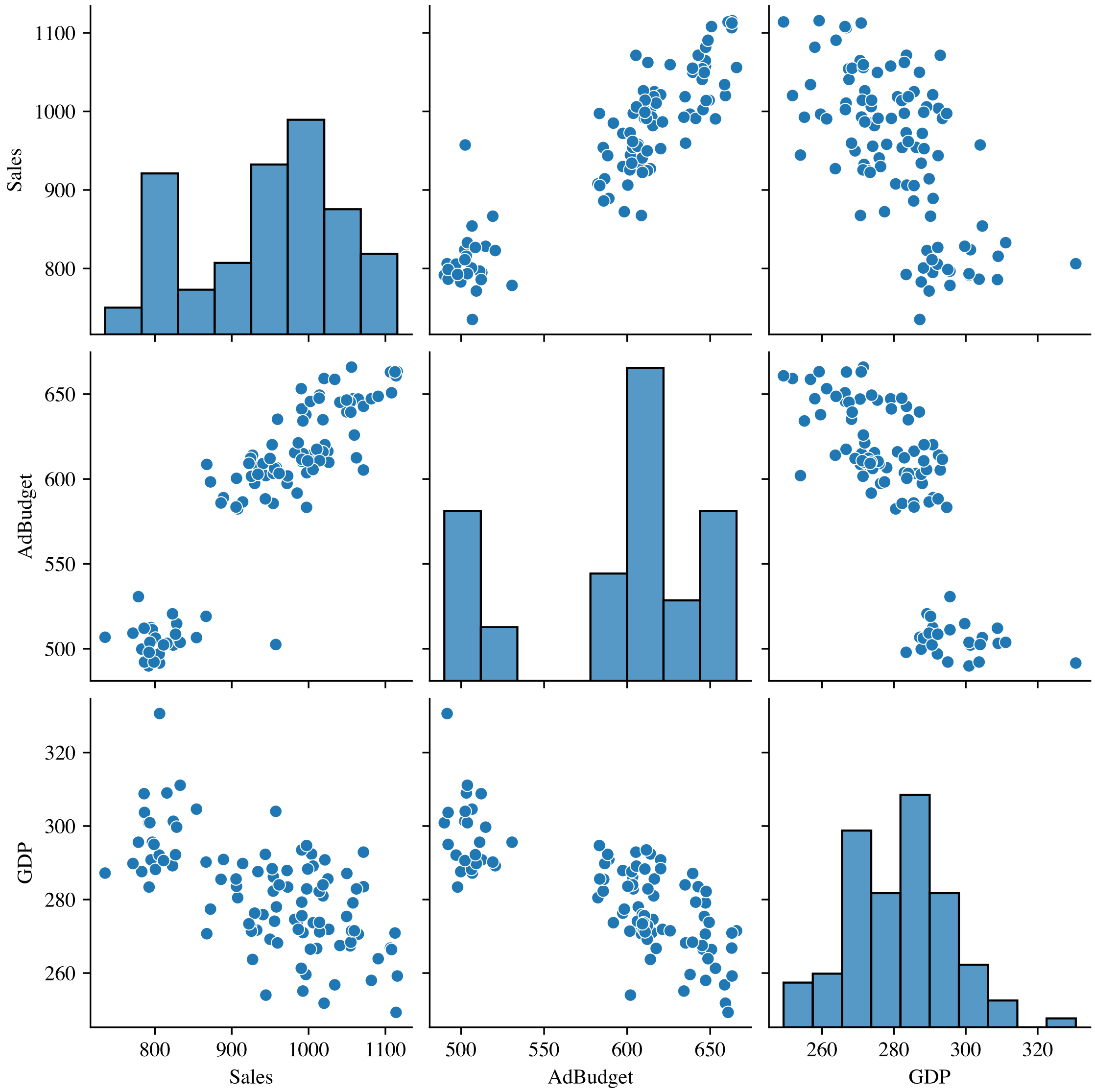

Figure 6: Pairplot of Tute 1 dataset using Diagonal Kind = ‘hist’

Problem 7

Figure 7 shows the correlation plot. In this plot, every grid represents the correlation coefficient between the corresponding pair of variables. The diagonal entries have value 1 because they represent the correlation coefficient between a variable and itself. The off diagonal values are what we calculated in Problem 1.

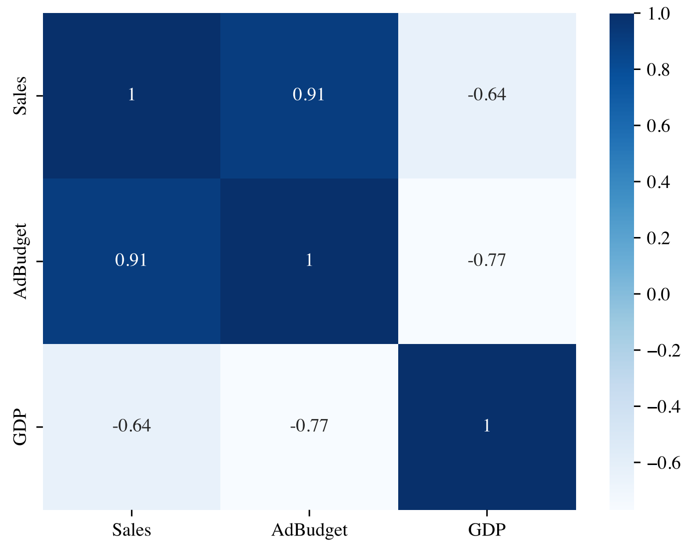

Figure 7: Correlation Plot for the Tute 1 dataset.

Problem 8

We display the result using the default values: Mean = 0 and variance = 1. The seed is set for the results to be reproducible.

1. The correlation coefficient between x and y = 0.003
2. The correlation coefficient between x and z = 0.824

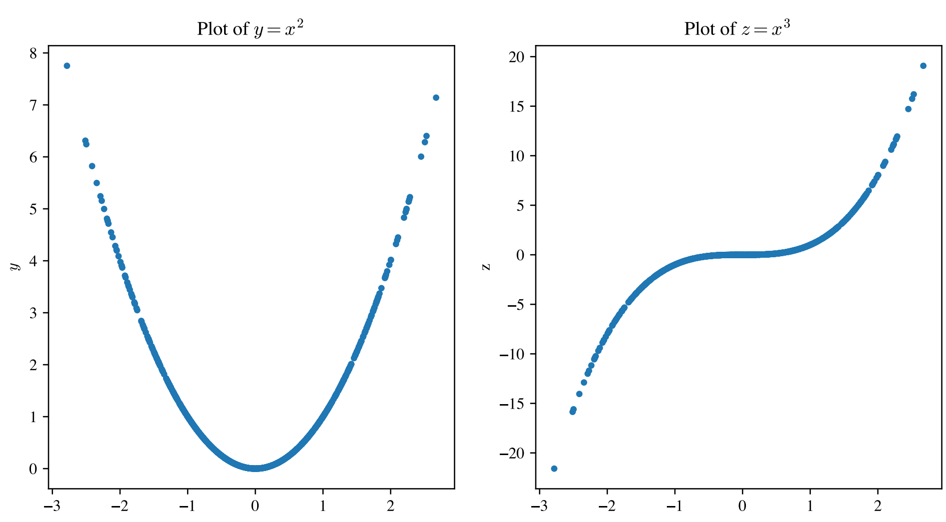

Figure 8: Plots of $y = x^2$ and $z = x^3$.

- The correlation coefficients found are consistent with the plots in Figure 8. For the case $y = x^2$, we see the data do not fall around a straightline. Therefore, the coefficient of correlation will be close to 0. However, for the case of $z = x^3$, we see that the data are close to a straightline. Therefore, the coefficient of correlation will be high, and the correlation is a positive correlation (since as x increases, z also increases).

- y and x are not independent because there is a relation between them: $y = x^2$. However, they are not correlated since the coefficient of correlation is close to 0.
- z and x are not independent because there is a relation between them: $z = x^3$. They are also correlated since the coefficient of correlation is high: $r = 0.824$.