

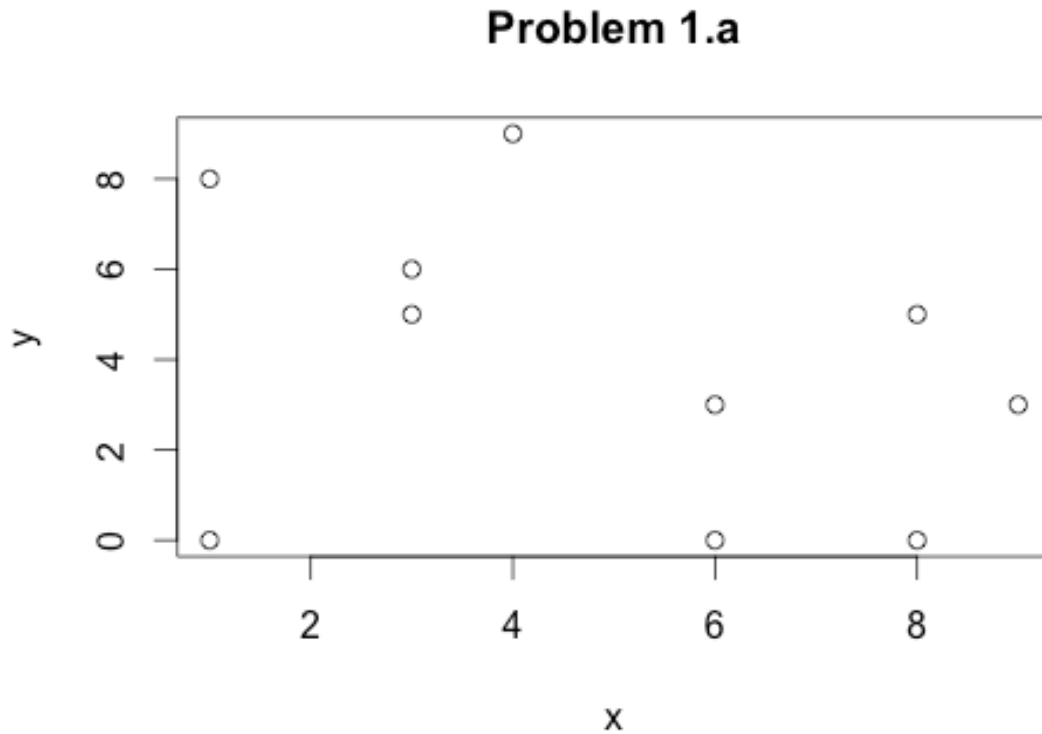
Problem1

Code

```
# Numbers from random.org
x = c(6,3,6,8,1,3,8,4,9,1)
y = c(0,5,3,0,0,6,5,9,3,8)

### Problem 1
problem1 = function(){
  cat("PROBLEM 1")
  # a.
  plot(x, y, main = "Problem 1.a")
  # b.
  cat("\nMean of x is", mean(x))
  cat("\nMean of y is", mean(y))
  # c.
  cat("\nVariance of x is", var(x))
  cat("\nVariance of y is", var(y))
  # d.
  cat("\nStandard deviation of x is", sd(x))
  cat("\nStandard deviation of y is", sd(y))
  # e.
  cat("\nCovariance of x and y is", cov(x,y))
  # f.
  cat("\nCorrelation of x and y is", cor(x,y))
  # g.
  slope = cov(x,y)/var(x)
  cat("\nThe slope of the regression line is cov(x,y) /
var(x) =", slope)
  new.x = 13
  pred.y = mean(y) + slope * (new.x - mean(x))
  cat("\nFor x =", new.x, "the point estimate for y is y
=", pred.y)
}
problem1()
```

Plot



Output

PROBLEM 1

Mean of x is 4.9

Mean of y is 3.9

Variance of x is 8.544444

Variance of y is 10.76667

Standard deviation of x is 2.923088

Standard deviation of y is 3.28126

Covariance of x and y is -3.233333

Correlation of x and y is -0.3371071

The slope of the regression line is $\text{cov}(x,y) / \text{var}(x) = -0.3784135$

For x = 13 the point estimate for y is $y = 0.8348505$

Problem 2.

- a. Discrete. The number of rides taken by a resident in a month is a count variable that takes values on the set of natural numbers, which is discrete.
- b. Continuous. Bicyclist's speed can take any real value bounded from below by 0 and from above by the speed of light. No particular value of speed has probability mass, only density.
- c. Continuous. Minute variations in lamps and current suggest a continuous variable. If the minute variations are not of interest, then the luminosity measurements can be converted to a discrete variable.
- d. Discrete. Assuming bankers cannot receive arbitrarily small fractions of a dollar, the salaries are discrete, however a continuous variable may be more convenient.
- e. Discrete. The number of hotels in Manhattan is a count variable.
- f. Continuous. Loudness is continuous.
- g. Discrete. Coffee can be 'good or bad' or 'very good, good, average, bad, very bad'. One could devise other metrics that would yield a continuous variable.

Problem 3.

There are several things to consider given the data.

- If we are arguing that Wall St. bankers have higher salaries compared to general population or some other specific professions, then we need to obtain data about salaries, IQs, and education.
- If we are trying to explain variation of salaries among Wall St. bankers, then we can try to do it with the given data by looking at scatter plots of salary against IQ and salary against education, we can examine correlations, or we can build a linear model. In doing so we should check that IQ and education are not too highly correlated, because high correlation would give unstable estimates. We can do the correlation check manually, or do model comparison using adjusted R^2 or ANOVA.

Intuition suggests that these variables might not explain much of the salary variation. We could postulate that many other variables like years of experience, types of banks and types of commodities traded, different descriptors of trading behavior would have predictive value and try to collect more data.