Researchers are interested in examining the relationship between gestational age (in weeks) and length (in cm) for newborn infants from a population of low-birth weight infants, defined to be infants weighing less than 1500 gm. Here is a scatter plot of a sample of 100 infants from this population.

The R output of fitting a regression line to predict length from age is below.

Coefficients:

| Estimate | Std. Error | t value | Pr(>|t|) |
|----------|------------|---------|---------|
| (Intercept) | 9.3282 | 3.0452 | 3.063 | 0.00283 ** |
| Age | 0.9516 | 0.1050 | 9.062 | 1.31e-14 *** |

---

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.648 on 98 degrees of freedom
Multiple R-Squared: 0.4559, Adjusted R-squared: 0.4504
F-statistic: 82.13 on 1 and 98 DF, p-value: 1.311e-14

(a) What feature of the data exhibited in the scatter plot merits further investigation?
(b) Circle the number closest to the correlation coefficient: −0.7 −0.2 0 0.4559 0.7 1.0 2.4.
(c) Write the regression line as a formula in the style of the example below.

(\text{height in inches}) = 30.25 + 0.25(\text{age in months})

(d) Construct a 95% confidence interval for the slope of the regression line.
(e) Use the regression line to predict the length of an infant whose gestational age is 25 weeks. Comment on the validity of this prediction.
(f) Use the regression line to predict the length of a full-term infant with gestational age 40 weeks. Comment on the validity of this prediction.
(g) Circle TRUE or FALSE: Suppose that the regression line goes through the point (age=21.7,length=30). Then, the regression line that predicts age based on length would go through the point (length=30,age=21.7).
(h) The following questions do not relate to the data set.
    Circle TRUE or FALSE: If the correlation coefficient is $r = -1$, then the points lie exactly on a line with slope $-1$.
    Circle TRUE or FALSE: For every simple linear regression, if the $X$ value is 1.7 standard deviations above the mean, then the predicted $Y$ value is also 1.7 standard deviations above the mean.
    Circle TRUE or FALSE: If $r = 0.98$, this implies that a straight line will fit the data much better than a curve.