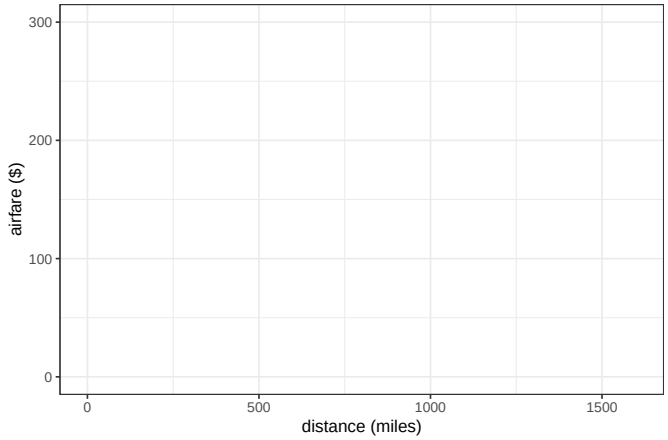
MATH 140 **Name** _____ **Date:**
Worksheet: Regression I

1. The following table reports the distance (in miles) as well as the airfare from Baltimore, MD to twelve destinations. On the axes provided, make a scatterplot of the distribution.

Destination	Distance	Airfare
Atlanta	576	178
Boston	370	138
Chicago	612	94
Dallas	1216	278
Detroit	409	158
Denver	1502	258
Miami	946	198
New Orleans	998	188
New York	189	98
Orlando	787	179
Pittsburgh	210	138
St. Louis	737	98



2. Based on the scatterplot, does it seem that knowing the distance to a destination would be useful for predicting the airfare? Explain.

3. With a straightedge, eyeball what you think is the best-fit line for these data.

4. Here is the equation of the least-squares regression line, courtesy of RStudio.

$$\hat{y} = 83.267 + 0.1174x$$

   (a) What airfare does the least squares line predict for a destination that is 300 miles away?

   (b) What airfare does the least squares line predict for a destination that is 1500 miles away?

(c) By how much does the regression line predict airfare to rise for each additional 100 miles that a destination is further away? Explain.

(d) What airfare would the regression line predict for a flight to San Francisco, which is 2842 miles from Baltimore? Would you consider this prediction as reliable as the one for 900 miles? Explain.

(e) Atlanta is 575 miles from Baltimore. What airfare does the best-fit line predict for a flight to Atlanta? What is the difference between the actual Atlanta airfare and the predicted Atlanta airfare? (This difference is called the **residual** of the observation).

(f) What is the actual airfare to Dallas? What is the airfare to Dallas predicted by the line? What is the residual value for Dallas?

(g) Which city do you think has the largest (in absolute value) residual? Circle this observation on the scatterplot.

(h) For observations with positive residual values, was their actual airfare greater or less than the predicted airfare?