"It’s a non-linear pattern with outliers……but for some reason I’m very happy with the data."
Warm-up:
1. \[ \hat{\text{price}} = 38257 - 0.1629(\text{miles driven}) \]

What is the slope? Interpret it.

For every additional mile driven, the price decreases by about $0.16.
Warm-up:

2. \( \hat{\text{price}} = 38257 - 0.1629(\text{miles driven}) \)

What is the y-intercept? Interpret it.

The predicted price when the car is brand new (0 miles driven), is about $38,257.
Warm-up:

3. \( \hat{\text{price}} = 38257 - 0.1629(\text{miles driven}) \)

Use the regression line to predict price for a Ford F-150 with 100,000 miles driven.

\( \hat{\text{Price}} = 38257 - 0.1629(100000) \)

\( \approx \$ 21,967 \)

Show work!
Warm-up:

4. Use the regression line to predict price for a Ford F-150 with 200,000 miles driven.

Since 200,000 miles is outside the domain of the data, we cannot use this equation to make predictions.

\[ \hat{\text{price}} = 38257 - 0.1629(\text{miles driven}) \]
3.2 Least-Squares Regression

When the bivariate, quantitative data follows a linear relationship, use a...

Regression Line: to describe how $y$ changes as $x$ changes

$\hat{y} = a + bx$  

"$y$ hat"  
Predicted $y$-value  
$\hat{y}$ intercept

\[ \text{Slope} = \text{change in } \hat{y} \text{ when } x \text{ increases by 1} \]

Ex. \[ \text{HEIGHT} = 140 + 5.2 \times \text{HAND} \]

*Then we will use the regression line to predict $y$ for a given $x$.*

Extrapolate: predict when $x$ is outside data set

Interpolate: predict when $x$ is inside data set
(aka Line of Best Fit)

**Least-Squares Regression Line**: the regression line that makes

$$\sum (y - \hat{y})^2$$ as small as possible

$$\sum \text{Residual}^2$$
Residual: the difference between the actual $y$ and predicted $\hat{y}$

$y - \hat{y}$

$A - P$

Actual - Predicted

Residual Plot: a scatterplot that displays the explanatory variable and residual

Residual $= 0$

TBA
\[ S = \text{Standard deviation of residuals} \]
\[ = \sqrt{\frac{\sum \text{(residuals)}^2}{n-2}} = \sqrt{\frac{\sum (y - \hat{y})^2}{n-2}} \]

= the approximate size of a typical error prediction

\[ S = 483 \]

Ex: If we use my equation for \( \hat{y} \) to predict total points scored using the number of minutes played, our prediction will typically be off by 483 points.

**Least-Squares Regression Line (LSR line)**

the line of regression that has the smallest possible SSR (sum of squared residuals = \( \sum \text{(residuals)}^2 \))