LINEAR REGRESSION PROOF by James Lee

Say we have n points on a coordinate plane.
Error $n = y_n - (mx + b)$
The objective is to minimize the squared error from each of these points to the line.
I want to find the values of m and be that define this line so that it minimizes the squared error.
Error is the vertical distance between a given point and the line.

$$SE_{line} = (y_1 - (mx_1 + b))^2 + (y_2 - (mx_2 + b))^2 + ... + (y_n - (mx_1 + b))^2$$

Find m and b that minimizes the SE_{line}!!!
Squared Error line $= (y_1 - (mx_1 + b))^2 + (y_2 - (mx_2 + b))^2 + ... + (y_n - (mx_1 + b))^2$
$= {y_1}^2 - 2y_1(mx_1 + b) + (mx_1 + b)^2$
$+{y_2}^2 - 2y_2(mx_2 + b) + (mx_2 + b)^2$
$+{y_3}^2 - 2y_3(mx_3 + b) + (mx_3 + b)^2$
$\cdots$
$\cdots$
$\cdots$
$+{y_n}^2 - 2y_n(mx_n + b) + (mx_n + b)^2$

$= ({y_1}^2 + {y_2}^2 + ... + {y_n}^2) - 2m(x_1y_1 + x_2y_2 + ... + x_ny_n) + m^2({x_1}^2 + {x_2}^2 + ... + {x_n}^2) + 2mb(x_1 + x_2 + ... + x_n) + nb^2$

******Key things to know *******

$\frac{y_1^2+y_2^2+....+y_n^2}{n} = \bar{y^2} <=> y_1^2 + y_2^2 + ... + y_n^2 = n\bar{y^2}$

$\frac{x_1y_1+x_2y_2+....+x_ny_n}{n} = \bar{xy} <=> x_1y_1 + x_2y_2 + ... + x_ny_n = n\bar{xy}$

$\frac{y_1+y_2+...+y_n}{n} = \bar{y} <=> y_1 + y_2 + ... + y_n = n\bar{y}$

$\frac{x_1^2+x_2^2+....+x_n^2}{n} = \bar{x^2} <=> x_1^2 + x_2^2 + ... + x_n = n\bar{x}$

$\frac{x_1+x_2+....+x_n}{n} = \bar{x} <==> x_1 + x_2 + ... + x_n = n\bar{x}$

$SE_{line} = n\bar{y^2} - 2m(n\bar{xy}) - 2bn\bar{y} + m^2 n\bar{x^2} + 2mb\bar{x} + nb^2$
The following equation above is the surface equation (in 3-dimension)

There are values m and b that minimizes SE_{line}

1) We have to find the partial derivative with respect to m and set it equal to 0
2) We have to find the partial derivative with respect to b and set it equal to 0