An illustration

As we all know, there is wide economic disparity between countries in the world. There are rich countries, poor countries, and in-between countries.

Research Q:

How much does a country's economic status affect the health and lifespan of its people?

Let's look at some data:

The Population Reference Bureau, an international, non-profit public policy development organization compiles data on a broad range of demographic, health and other characteristics of the world's population. In the following demo we will use their data to explore our research Q.

Key variables in dataset:

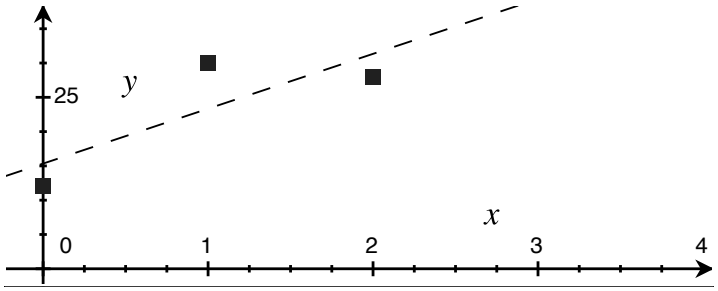
GNI = Gross National Income per capita (in US dollars)
LE = Life Expectancy (in years)

Some explorations with PNB data

- scatterplots
- explanatory and response variables.
- linear association
- correlation concept
- straight line equation
- line of best fit
- residuals

Exercise

x_i	y_i
0	12
1	30
2	28



A very small dataset consisting of only 3 observations is shown in the table, together with a scatter plot of y vs x. We want to fit a straight line approximation to the plot, as shown by the dotted line.

Consider the straight line: $\hat{y} = mx + b$.
Our goal is to find numerical values of m and b that give the "best" straight line approximation. Here are the steps:

1. For each i, find the error: $e_i = y_i - \hat{y}_i$
   Each e_i will be a function of m and b (only!).

2. Compute the function: $f = \sum (e_i)^2$
   This function is the sum of the square of the errors.

3. Next, we want to minimize f – remember calculus?!
   Find the derivative of f with respect to b.
   Then, find the derivative of f with respect to m.

4. Set $\frac{df}{db} = 0$ and $\frac{df}{dm} = 0$, and solve simultaneously for m and b.

5. Well, then that is your best straight line!