```
In[190]:= (* Adam Beck *)
(* Calculate the correlation coefficient between the volume V and mortality M
 of the heart transplants based on the data in hospheart.nb *)

(* hospheart.np data *)
(* {M,V} M = one year mortality rate,
percentage of patients that died within one year of the
 transplant operation,
V = average annual number of transplants at that center during the same 4 years *)
heart = {{17.9, 27}, {23.1, 4}, {40, 3}, {6.5, 35}, {14.9, 17}, {12.5, 4}, {15.7, 45},
   {9.8, 28}, {24, 6}, {5.0, 10}, {15.4, 13}, {4.8, 7}, {0, 1}, {19.1, 47}, {4.5, 6},
   {15, 56}, {12.5, 4}, {33.9, 8}, {10.7, 9}, {13, 14}, {28.3, 12}, {57.1, 2}, {6.3, 4},
   {10, 3}, {8.3, 12}, {17.5, 10}, {20, 3}, {29.3, 10}, {21.4, 7}, {27.3, 8}, {13.6, 6},
   {21.8, 30}, {36.4, 3}, {18.2, 11}, {33.3, 2}, {20, 4}, {38.5, 7}, {20.8, 18}, {12.2, 19},
   {22.2, 18}, {29, 8}, {0, 9}, {5.7, 9}, {50, 2}, {21.7, 15}, {66.7, 4}, {29.4, 17},
   {12.1, 27}, {10.7, 14}, {6.3, 4}, {16.2, 9}, {21.1, 5}, {17.4, 33}, {23.9, 17},
   {42.9, 2}, {40, 2}, {6.7, 15}, {44.4, 3}, {18.7, 34}, {14.7, 24}, {7.4, 7}, {12.6, 24},
   {9.7, 26}, {44.4, 2}, {16.7, 6}, {15.8, 14}, {83.3, 2}, {10.9, 22}, {13.3, 5},
   {11.1, 5}, {75, 2}, {19, 20}, {14, 13}, {60, 1}, {21.2, 8}, {9.7, 8}, {50, 2}, {25, 14},
   {18.6, 15}, {0.0, 1}, {35.3, 9}, {23.5, 85}, {15.6, 11}, {37.5, 2}, {14.3, 28},
   {14.3, 4}, {16.7, 6}, {20.0, 15}, {13.0, 17}, {9.6, 26}, {66.7, 3}, {30.8, 3},
   {14.0, 13}, {27.5, 10}, {37.5, 8}, {18.9, 13}, {0.0, 4}, {12.2, 44}, {57.1, 4},
   {21.4, 35}, {23.4, 16}, {10.9, 12}, {15.6, 8}, {16.7, 2}, {13.9, 9}, {18.2, 11},
   {11.5, 26}, {18.4, 13}, {16.7, 3}, {20.4, 14}, {40.0, 5}, {20.7, 56}, {19.6, 13},
   {13.5, 9}, {29.9, 36}, {8.4, 21}, {28.4, 24}, {7.7, 23}, {19.3, 29}, {0.0, 1},
   {22.2, 20}, {30.0, 5}, {7.0, 11}, {23.8, 7}, {18.8, 29}, {14.5, 16}, {17.0, 16},
   {20.0, 15}, {6.7, 15}, {11.4, 20}, {100.0, 1}, {31.4, 9}, {17.6, 26}, {19.6, 14}};

(* Split this M and V data into separate
 lists via Transpose[] in order to parse through *)
heartTranspose = Transpose[heart];
MData = heartTranspose[[1]];
VData = heartTranspose[[2]];

(* To calculate the correlation coefficient,
we need to define functions to find the means of a data set *)
mean[x_] := Sum[x[[i]], {i, 1, Length[x]}] / Length[x];
(* Sum elements, divide by length *)

meanM = mean[MData]

Out[194]= 21.9045

In[195]:= meanV = N[mean[VData]]

Out[195]= 13.8657
```

```
In[196]:= (* Create a function to sum (m-meanM)*(v-meanV),
          where m and v are elemnts of M and V respectively. *)
          differenceMeanSum[m_, mBAR_, v_, vBAR_] :=
           Sum[ ((m[[i]] - mBAR) * (v[[i]] - vBAR)), {i, 1, Length[m]}]
          (* Sum the product of (elementInM - meanOfM) (elementInV - meanOfV) *)


          (* find the sum of the mean difference as noted above,
          this is the numerator of our correlation coefficient equation *)
          MVdifferenceMeanSum = differenceMeanSum[MData, meanM, VData, meanV]

Out[197]= -7238.42

In[203]:= (* The denominator of the correlation coefficient equation is
           the square root of: sum of (x-xBAR)^2 times sum of (y-yBAR)^2 *)
          (* Create a function to Sum a data sets elemnts, by taking an element,
          subtracting the mean from it, and squating the value *)
          squaredSum[a_, aMean_] := Sum[(a[[i]] - aMean)^2, {i, 1, Length[a]}];

In[204]:= (* Now find the the squared difference sum for the M and V data *)
          squareDifferenceM = squaredSum[MData, meanM]

Out[204]= 35996.9

In[205]:= squareDifferenceV = squaredSum[VData, meanV]

Out[205]= 22305.6

In[206]:= (* Take the root of the product of these sums,
          and that is the denominator of the correlation coefficient equation *)
          rootOfSums = Sqrt[squareDifferenceM * squareDifferenceV]

Out[206]= 28336.1

In[207]:= (* Now take the numerator and denominator,
          find the decimal of that fraction and we have the correlation coefficient *)
          coefficient = MVdifferenceMeanSum / rootOfSums

Out[207]= -0.255449

          (* This is negative, so as x increases,
          y decreases. As is is close to zero, it is not a strong correlation *)

          (* Now, we need to make a scatter plot of this data *)
```

```
(* We also need the line of best fit,
which will be calculated manually. Assuming the V data is y axis, M data is x axis*)
(* The numerator for the slope of our
 best fit line is the variable MVdifferenceMeanSum *)
(* The denominator for the slope of our best fit line is the sum of x -
 xMean squared *)

(* This denominator is as follows *)
denom = Sum[(MData[[i]] - meanM)^2, {i, 1, Length[MData]}];
slope = MVdifferenceMeanSum / denom
```

Out[217]= -0.201084

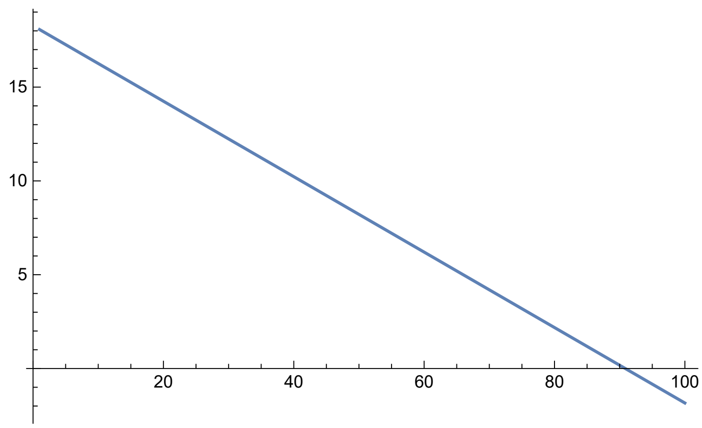
```
In[218]:= (* The b value is calculated as the y mean, minus slope * x mean *)
b = meanV - slope * meanM
```

Out[218]= 18.2703

```
In[219]:= (* The best fit equation is as follows *)
bestFit[x_] := slope * x + b;
```

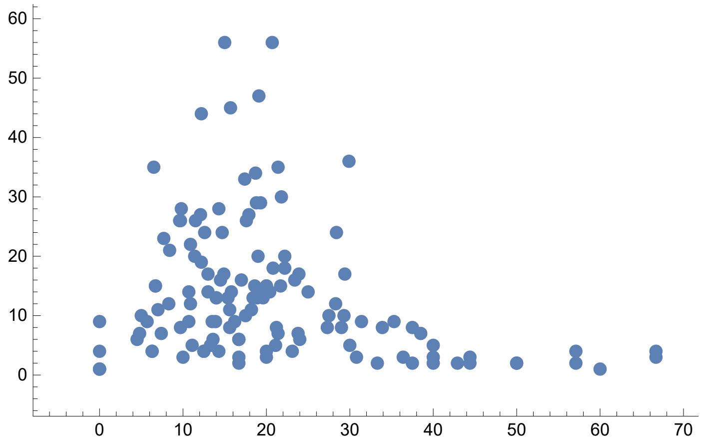
```
In[238]:= (* Now, superimpose this regression line on the scatter plot *)
bestFitGraph = Plot[bestFit[x], {x, 1, 100}]
```

Out[238]=



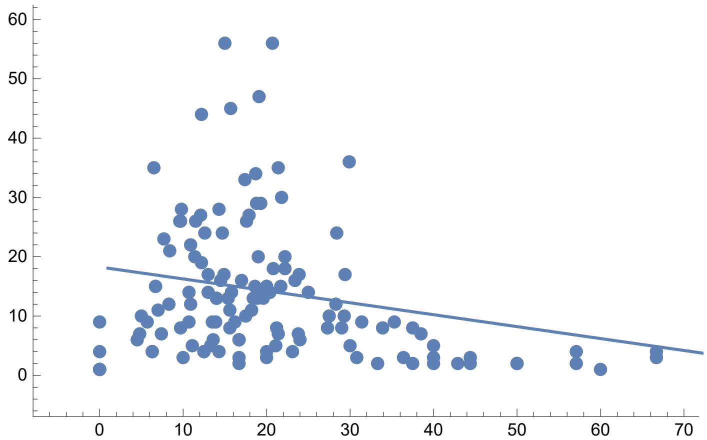
In[268]:=
```
scatterPlot = ListPlot[heart, PlotStyle → PointSize[.02],
  PlotRangePadding → Scaled[0.1], Axes → False, Frame → {True, True, False, False}]
(* Extra paramets on this scatter plot to best show the data *)
```

Out[268]=



In[269]:=
```
superimpose = Show[scatterPlot, bestFitGraph]
```

Out[269]=



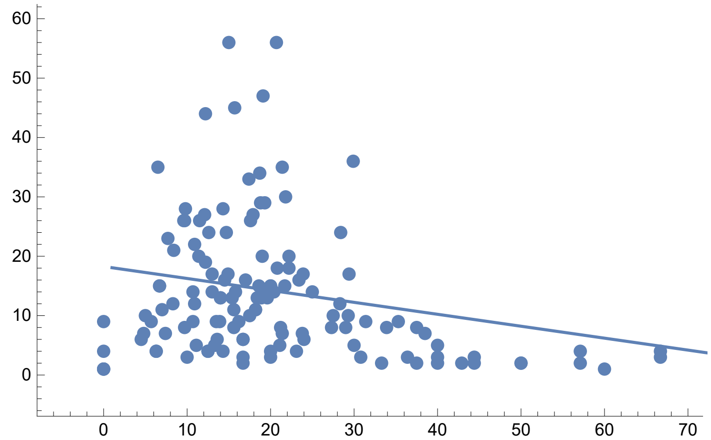
```
(* To double check my calculations are correct,
I will let Mathematica generate a scatter plot and best fit line *)
```

In[277]:=
```
Fit[heart, {1, x}, {x}]
```

Out[277]= 18.2703 - 0.201084 x

In[305]:=
```
bestLine[x_] := 18.270320917115342` - 0.20108442453442926` x;
Show[ListPlot[heart, PlotStyle → PointSize[.02], PlotRangePadding → Scaled[0.1],
  Axes → False, Frame → {True, True, False, False}], Plot[bestLine[x], {x, 1, 100}]]
```

Out[306]=




```
(* Conclusions *)
(* This is negative, so as x increases,
y decreases. As is is close to zero, it is not a strong correlation *)
(* The one year mortality rate,
and average annual number of transplants at that center during the same 4 years, are
 not correlated strongly. The correlation is very weak *)
```