Covariance

Theme

Part 1: Small Dataset

In[89]:=
```
dataSmall = {{1, 0}, {3, 2.5}, {2, 3}, {0, 2.5}};
```

In[90]:=
```
μ = Mean[dataSmall] // N
```

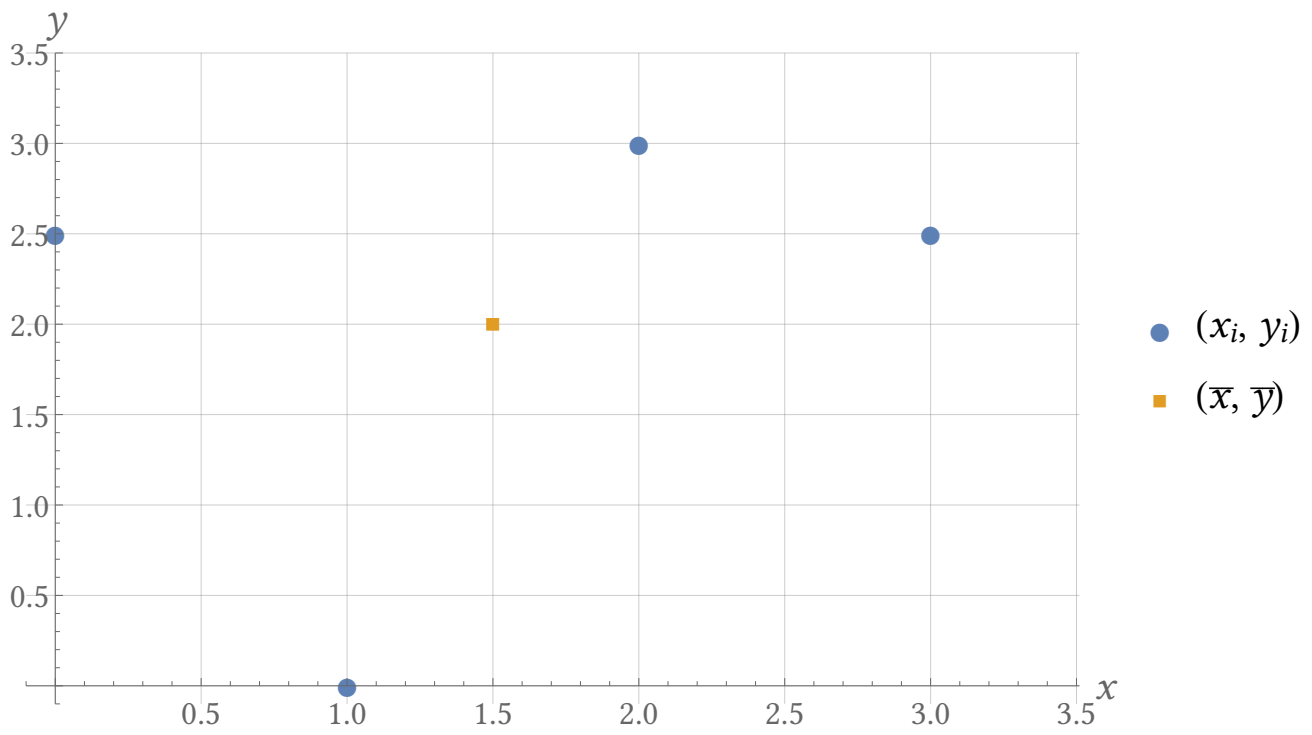
Out[90]=
```
{1.5, 2.}
```

In[91]:=
```
Covariance[dataSmall] // MatrixForm
```

Out[91]//MatrixForm=

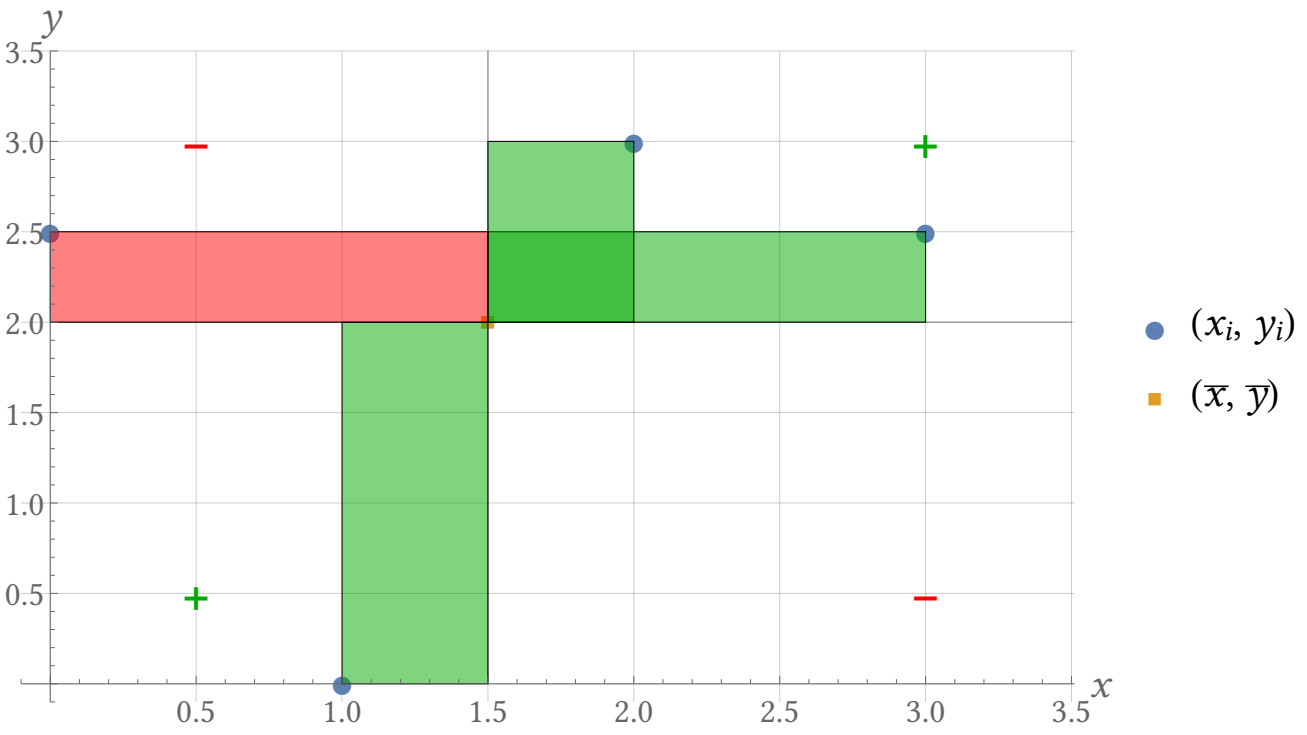
$$\begin{pmatrix} 1.66667 & 0.5 \\ 0.5 & 1.83333 \end{pmatrix}$$

In[101]:=
```
base = ListPlot[{dataSmall, {μ}},
  PlotTheme → {"myTheme", "ThickLines"},
  GridLines → Automatic,
  PlotRange → {{-0.1, 3.51}, {-0.1, 3.5}},
  AxesLabel → {it["x"], it["y"]},
  PlotMarkers → {Automatic, 15},
  PlotLegends → {"(xᵢ, yᵢ)", "(x̄, ȳ)"}
 ]
```



```
Show[
 base,

 Graphics[{
   Gray,
   Line[{{μ[[1]], 0}, {μ[[1]], 3.5}}],
   Line[{{0, μ[[2]]}, {3.5, μ[[2]]}}],

   Darker@Green,
   FontSize → 25,
   Text["+", {3, 3}],
   Text["+", {0.5, 0.5}],

   Red,
   Text["-", {3, 0.5}],
   Text["-", {0.5, 3}],

   EdgeForm[Thin],
   Opacity[0.5],
   Darker@Green,
   Table[Rectangle[{Min[p[[1]], μ[[1]]], Min[p[[2]], μ[[2]]]},
     {Max[p[[1]], μ[[1]]], Max[p[[2]], μ[[2]]]}, {p, dataSmall[[{1, 2, 3}]]}],

   Red,
   Table[Rectangle[{Min[p[[1]], μ[[1]]], Min[p[[2]], μ[[2]]]},
     {Max[p[[1]], μ[[1]]], Max[p[[2]], μ[[2]]]}, {p, dataSmall[[{4}]]}]
  }]
]
```

Out[97]=



We can calculate the covariance by weighting the rectangles:

In[102]:=
$$\frac{1}{3}\left(0.5*1+1.5*0.5+0.5*2-1.5*0.5\right)$$

Out[102]= 0.5

Part 2: Sign

The new datasets for the different signs may look like:

- $s_{XY} > 0$: positive line
- $s_{XY} = 0$: circle
- $s_{XY} < 0$: negative line

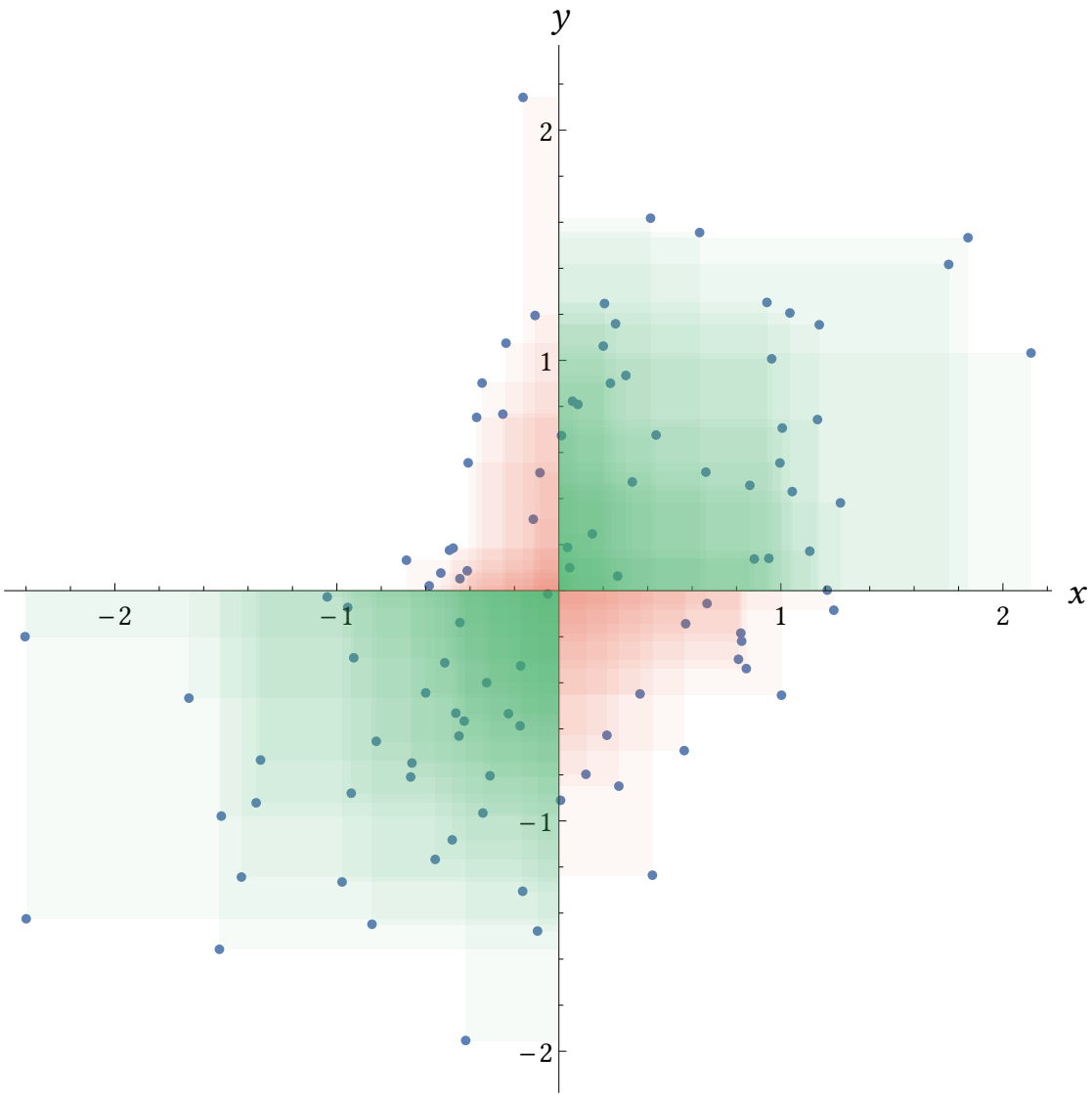
Part 3: Large Dataset

```
SeedRandom[1337];
data = RandomVariate[MultinormalDistribution[{0, 0}, {{0.8, 0.5}, {0.5, 0.8}}], 100];
data = (#1 - Mean[data] &) /@ data;
Covariance[data] // MatrixForm
```

$$\begin{pmatrix} 0.724658 & 0.37535 \\ 0.37535 & 0.71112 \end{pmatrix}$$

```
dataPos = Cases[data, {x_, y_} /; x * y ≥ 0];
dataNeg = Cases[data, {x_, y_} /; x * y < 0];
```

```
ListPlot[data,
 PlotTheme → "myTheme",
 AxesLabel → {"x", "y"},
 AspectRatio → 1,
 Epilog → {
   {■, Opacity[0.05], (Rectangle[{0, 0}, #1] &) /@ dataPos},
   {■, Opacity[0.05], (Rectangle[{0, 0}, #1] &) /@ dataNeg}
  }
]
```



- Increase: add a new point in the green area.
- Increase even further: make sure that the new point is further away from the origin than the previous one.
- The scaling of the features matters and if we change the units, the absolute value also changes (there is no normalization by the variances like for the correlation coefficient).