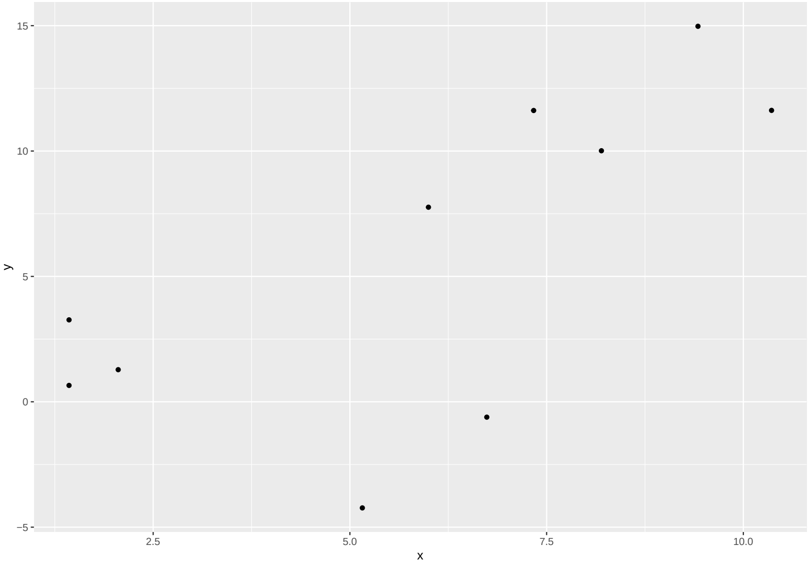
Principal Component Analysis

What it does?

How does it do it?

PCA in a view or coordinate rotation




```
Standard deviations (1, .., p=2):
[1] 6.834255 2.196776
```

Another perspective of PCA

- Consider n points of data. We look for a direction where the projections of these points to it give the maximum variance.
- Projection of a point onto a direction (vector).
- A point is presented by a pair of number $(3, 4)$. A vector/direction is also presented by a pair of number: x-axis presented by $(1, 0)$ and y-axis presented by $(0, 1)$. A vector/direction of (a, b) could be a point connecting the origin $(0, 0)$ and the point (a, b).
- A projection of a point (x_0, y_0) onto a vector (a, b) is a number (scalar), it is the dot product of the two pair or

$$ax_0 + by_0$$

- A vector also presents a subspace of one dimension. So we project a point of two dimension onto a subspace of one dimension, the projection should have only one number.

PCA as projections onto subspace

- If we have n points to a vector (one dimensional subspace), we will receive n numbers. Thus, a data of n points in two dimension (a pair), presented by a matrix, will turn into n numbers. So we can say that a matrix of $n \times 2$, projected onto a vector will be n numbers (n points in one dimensional subspace).
- So, n points projects onto a vector becomes n numbers. We want to find a vector/subspace that maximize the variance of this points. What is the variance of points. If the points have zero means, the variance is the length of the n points.
- Language of subspace and projection and the language of statistics. We should discuss this using one language only.

Projection of a point to a direction

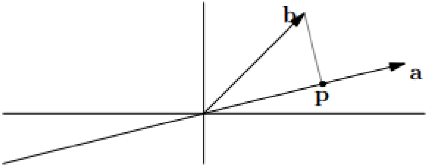
- For simplicity, we first consider 2 dimension first
- Both a point A and a vector OA (O is the origin) is presented by a column vector.
- The projection of a point A, presented by a column vector $b = [b_1, b_2]^T$ onto a vector through the origin, also presented by a column vector, $a = [a_1, a_2]^T$ (or the span of this vector which is a 1-dimensional subspace) is

$$\frac{aa^T}{a^Ta}b$$

,

Projection of a point to a direction

- $P = \frac{aa^T}{a^T a}$ is called a projection matrix.



- For example, the projection of a point $x = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ onto a direction of $v = [3, 2]^T$ (or onto span(v)) is

$$\frac{vv^T}{v^T v} x = \begin{bmatrix} 2.076923 \\ 1.384615 \end{bmatrix}$$

- We usually consider the direction vector of length 1.

Checking

- Checking only.

What does the $v^T x$ present?

- Sometime, the projection of a point $x = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ onto $span(v)$ where $\|v\| = 1$ can be presented by a dot product $x^T v$.
- This product is the length of the projected vector, which is also the coordinate of the projected point in a new axis v.
- This is because

$$u = \frac{vv^T}{v^T v} x = v \frac{v^T x}{v^T v} = vv^T x \implies \|u\| = \|v\| v^T x = v^T x$$

as $v^T v = \|v\|^2 = 1$.

Projections of n points to a direction

- The projection of vector x onto v is $x^T v$
- The projection n points $X = \begin{bmatrix} x_1^T \\ x_2^T \\ .. \\ x_n^T \end{bmatrix}$ (X is a $n \times 2$ matrix) is

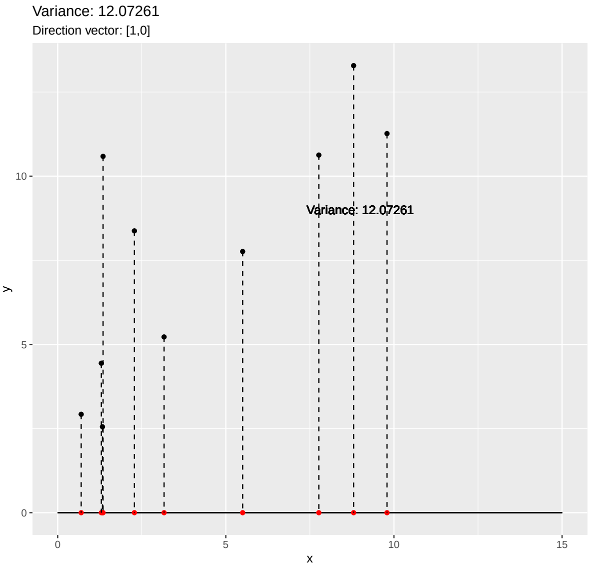
  $Xv = \begin{bmatrix} x_1^T v \\ x_2^T v \\ .. \\ x_n^T v \end{bmatrix}$, which is a $n \times 1$ matrix.
- We can say that n points $x_1^T, x_2^T, ..., x_n^T$ in R^2 are transformed to n points $x_1^T v, x_2^T v, ..., x_n^T v$ in R^1.

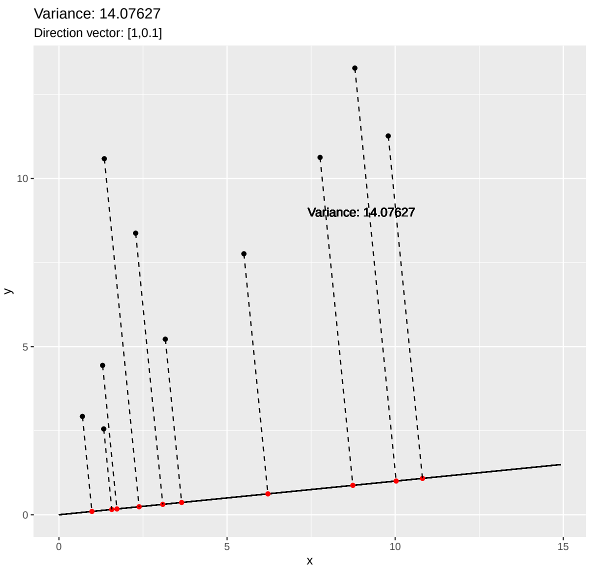
Variance and Total Variance

- In one dimension, the variance of n points/numbers $a_1, a_2, ..., a_n$ is $\frac{1}{n-1}\sum(a_i - \bar{a})^2$. If a is center, $\bar{a} = 0$, then the variance is $\frac{1}{n-1}\sum a_i^2 = \frac{1}{n-1}\|a\|^2 = \frac{1}{n-1}a^T a$.
- Let X be an $n \times 2$ data matrix. Let's X be centralized, which means both two columns have zero means. The projection of X onto v, Xv also has zero mean. Thus, the variance of the projection Xv is $\frac{1}{n-1}\|Xv\|^2$.
- The total variance of the data is the sum of the variance of the projection onto the x-axis (var(x)) and the variance of the projection onto y-axis (var(y)).
- If we rotate the xy axis into uv axis, the total variance does not change, var(u) $+$ var(v) $=$ var(x) $+$ var(y)
- PC1 is the direction that has the maximum variance of the data.

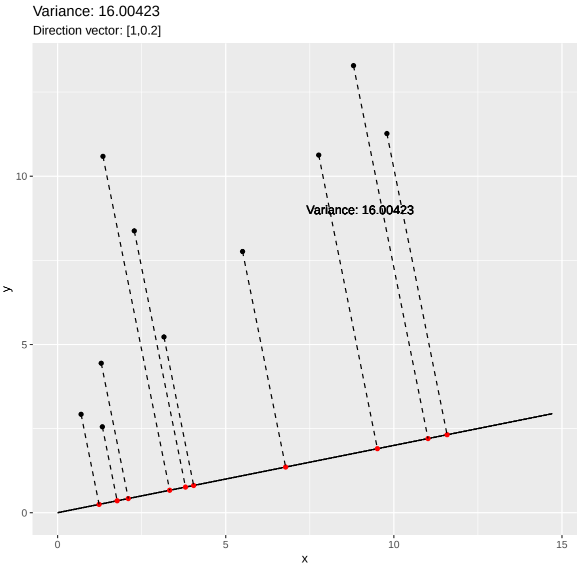
Demonstration

```
[1] 26.21958
```

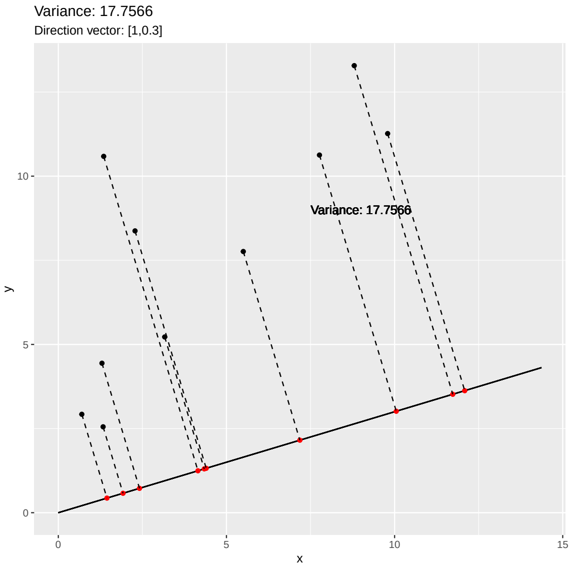

Variance: 12.07261
Direction vector: [1,0]
Variance: 12.07261
y
x
0
5
10
0
5
10
15

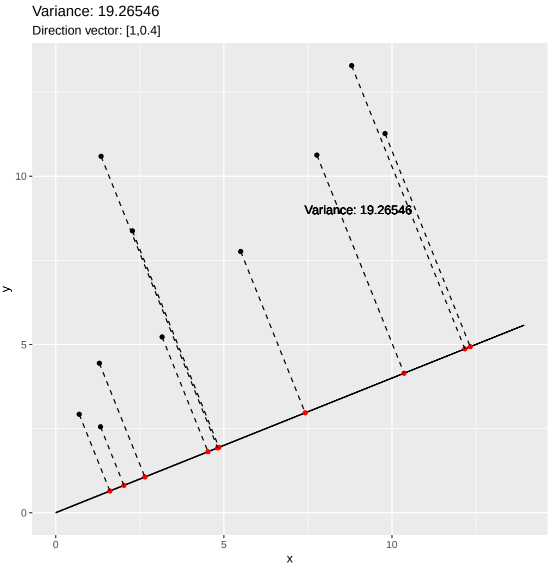


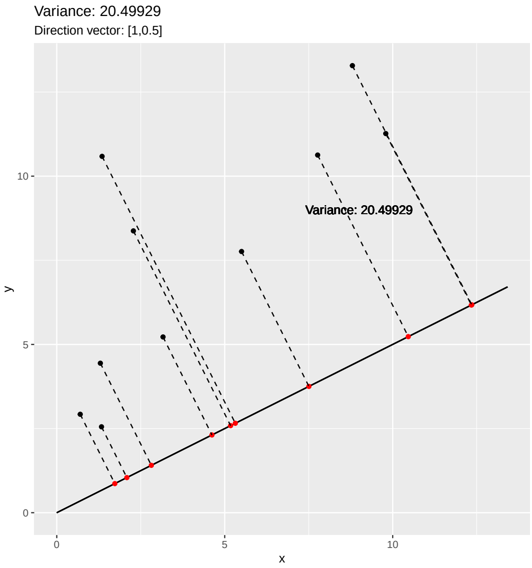
Variance: 14.07627
Direction vector: [1,0.1]
Variance: 14.07627
y
x
0
5
10
15

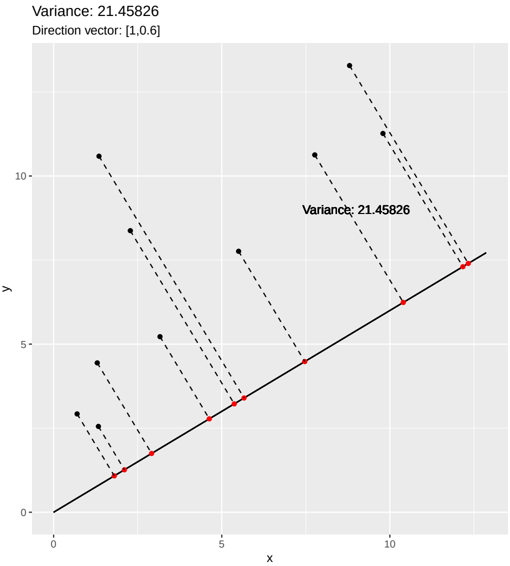


Variance: 16.00423
Direction vector: [1,0.2]
Variance: 16.00423
y
x
0
5
10
15
0
5
10

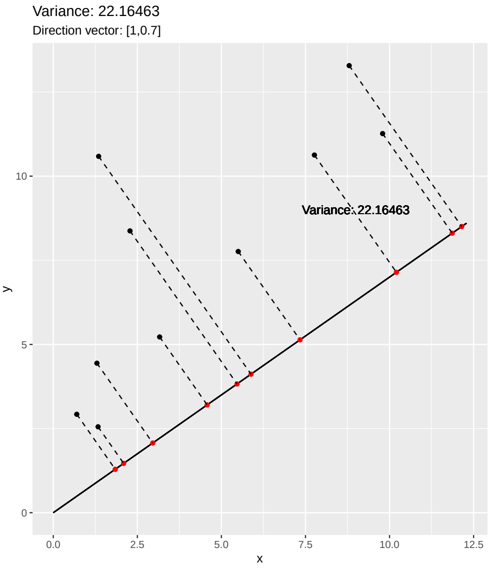

Variance: 17.7566
Direction vector: [1,0.3]
Variance: 17.7566
y
x
0
5
10
0
5
10
15

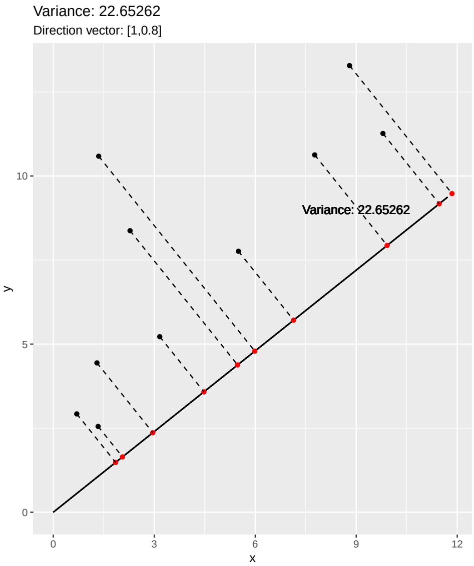

Variance: 19.26546
Direction vector: [1,0.4]
Variance: 19.26546
y
x
0
5
10
0
5
10


Variance: 20.49929
Direction vector: [1,0.5]
Variance: 20.49929
y
x
0
5
10
0
5
10

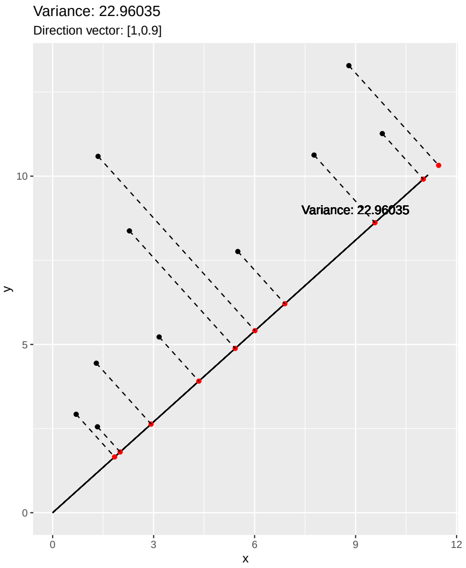


Variance: 21.45826
Direction vector: [1,0.6]
Variance: 21.45826
y
x
0
5
10
0
5
10

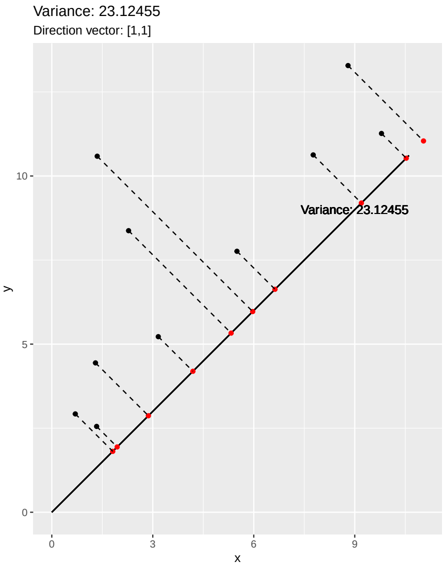

Variance: 22.16463
Direction vector: [1,0.7]
Variance: 22.16463
y
x
0
5
10
0.0
2.5
5.0
7.5
10.0
12.5

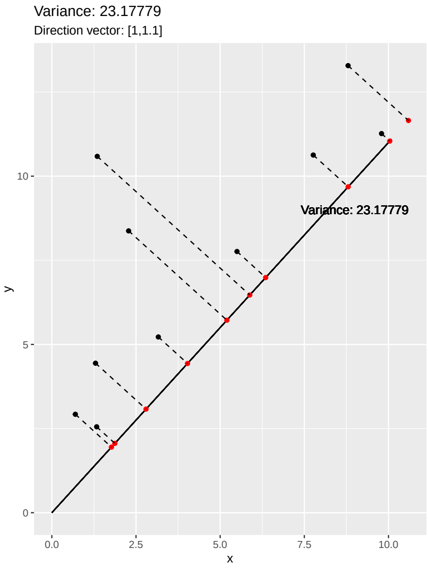


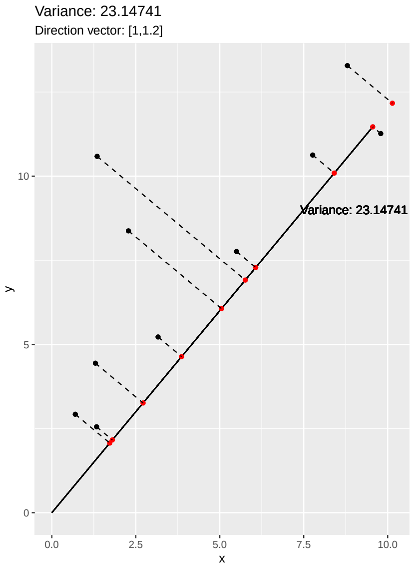
Variance: 22.65262
Direction vector: [1,0.8]
Variance: 22.65262
y
x
0
5
10
0
3
6
9
12



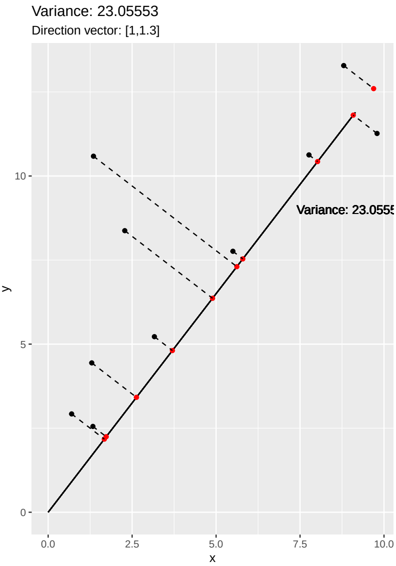
Variance: 22.96035
Direction vector: [1,0.9]
Variance: 22.96035
0
5
10
y
0
3
6
9
12
x

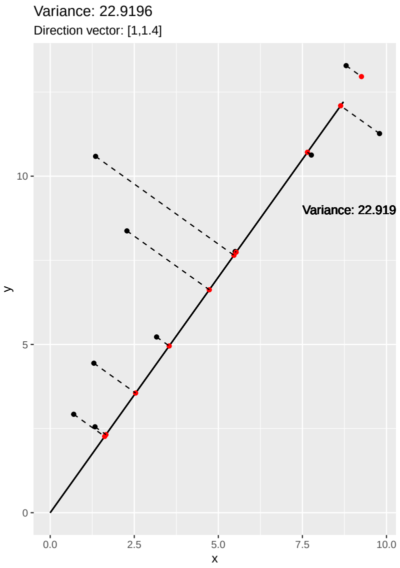

Variance: 23.12455
Direction vector: [1,1]
Variance: 23.12455
y
x
0
5
10
0
3
6
9


Variance: 23.17779
Direction vector: [1,1.1]
Variance: 23.17779
10
5
0
y
0.0
2.5
5.0
7.5
10.0
x

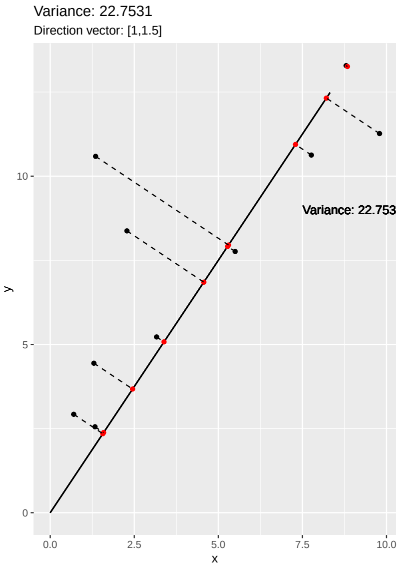


Variance: 23.14741
Direction vector: [1,1.2]
Variance: 23.14741
0
5
10
0.0
2.5
5.0
7.5
10.0
x
y

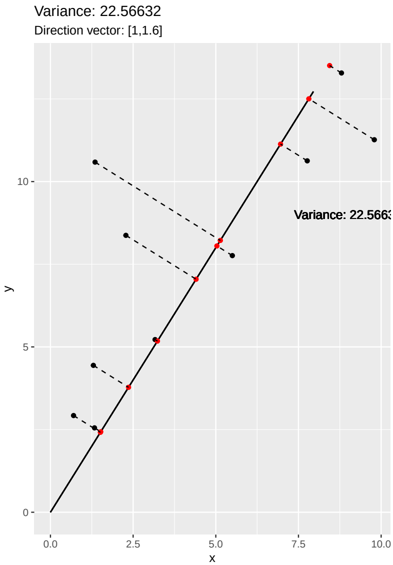

Variance: 23.05553
Direction vector: [1,1.3]
Variance: 23.0555
0
5
10
y
0.0
2.5
5.0
7.5
10.0
x

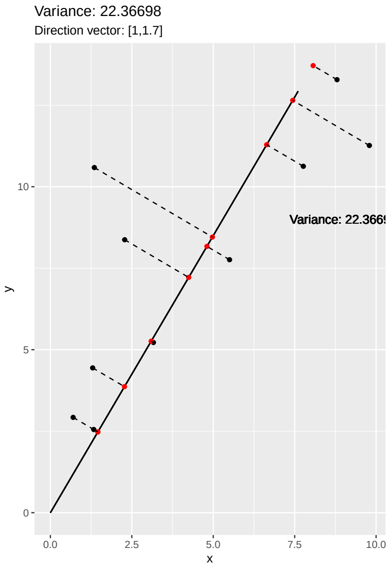


Variance: 22.9196
Direction vector: [1,1.4]
Variance: 22.919
y
x
0
5
10
0.0
2.5
5.0
7.5
10.0



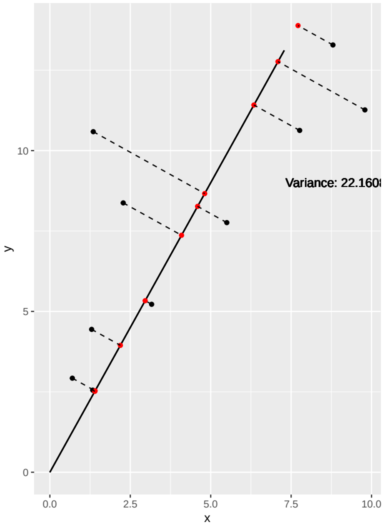
Variance: 22.7531
Direction vector: [1,1.5]
Variance: 22.753
y
x
0
5
10
0.0
2.5
5.0
7.5
10.0

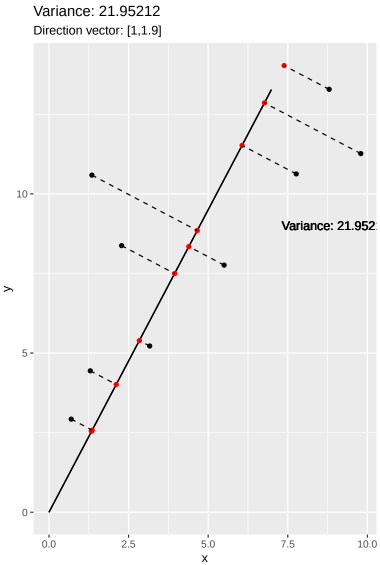

Variance: 22.56632
Direction vector: [1,1.6]
Variance: 22.5663
y
x
0.0
2.5
5.0
7.5
10.0
0
5
10


Variance: 22.36698
Direction vector: [1,1.7]
Variance: 22.3669
y
x
0
5
10
0.0
2.5
5.0
7.5
10.0

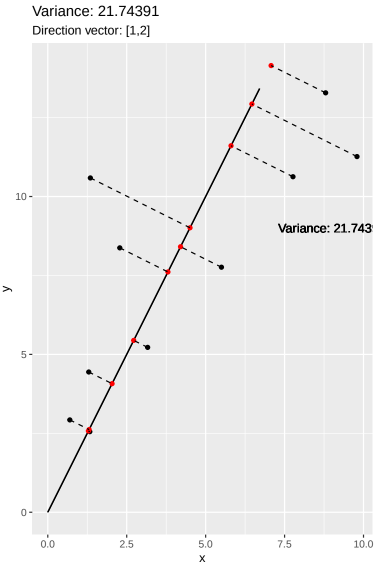
Variance: 22.16083

Direction vector: [1,1.8]

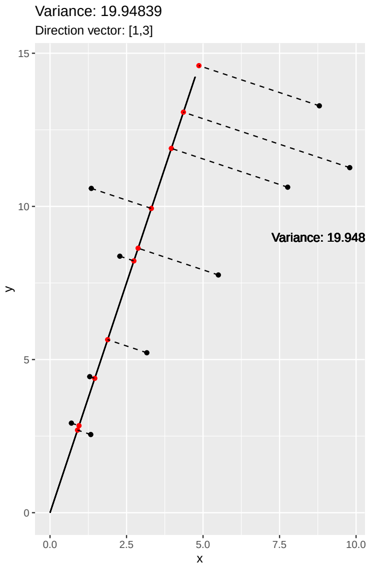




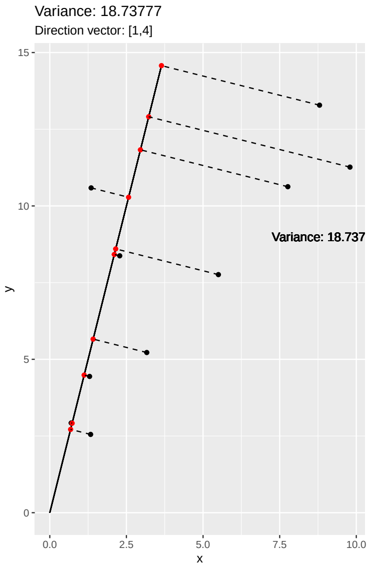
Variance: 21.95212
Direction vector: [1,1.9]
Variance: 21.952
y
x
0
5
10
0.0
2.5
5.0
7.5
10.0



Variance: 21.74391
Direction vector: [1,2]
Variance: 21.7439
10
5
0
y
0.0
2.5
5.0
7.5
10.0
x



Variance: 19.94839
Direction vector: [1,3]
Variance: 19.948
15
10
5
0
y
0.0
2.5
5.0
7.5
10.0
x


Variance: 18.73777
Direction vector: [1,4]
Variance: 18.737
y
x
0
5
10
15
0.0
2.5
5.0
7.5
10.0

Variance: 23.17812

Direction vector: [−0.669696257092749,−0.742635121197458]