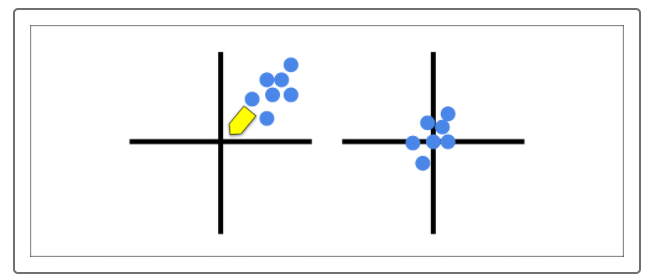
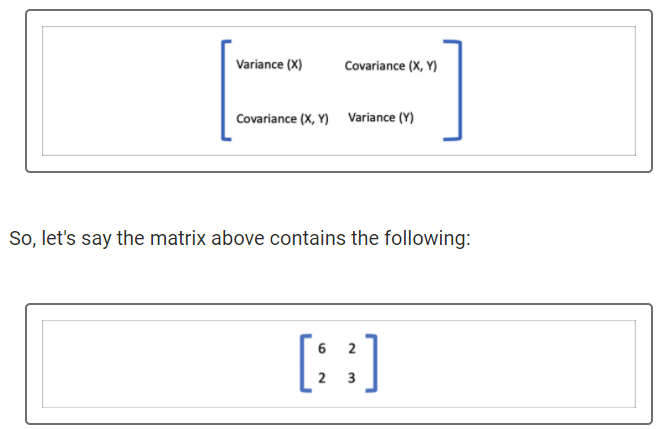
Say we have a set of points on a graph. We want to center these points by taking the average of the coordinate, both X and Y. Find the balance point and move that to zero:

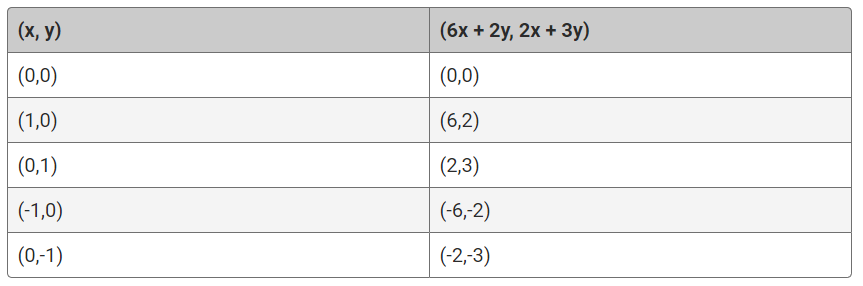


Once the points are centered, we're going to create a 2x2 matrix that consists of the variance and covariances that we found in the previous step:



This matrix will be used to transform the points from one graph to another by using the numbers to create a formula for our transformation. The top two values of the matrix will correspond to one point and the bottom two values to another.

In our example, the formula for the points becomes (6x + 2y, 2x + 3y). Let's plug some coordinates into the formula:



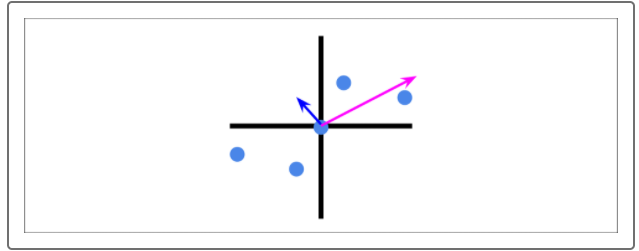
Now, let's plot the new points from the right side of the matrix to create a linear transformation:

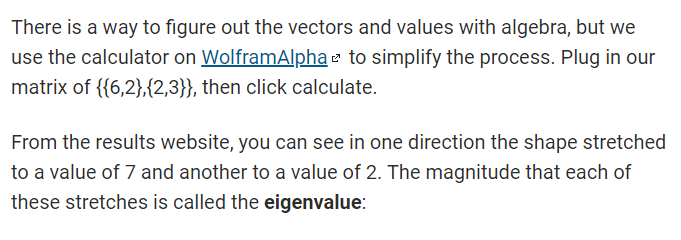
****

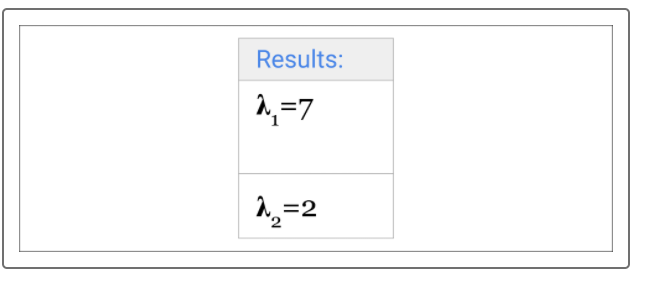
**NOTE**

Eigenvectors and eigenvalues can be complicated subjects rooted in linear algebra. We cover these at a very high level, but if you wish to explore more on your own, you can read more about [Eigenvalues and eigenvectors (Links to an external site.)](https://en.wikipedia.org/wiki/Eigenvalues_and_eigenvectors) and watch this [video (Links to an external site.)](https://www.youtube.com/watch?v=PFDu9oVAE-g)

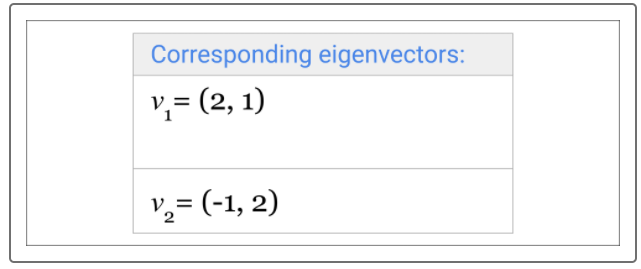
As you can see, the points stretch out in our graph in two directions. One direction moves from southwest to northeast direction while another direction moves from southeast to northwest. These are called **eigenvectors**, as indicated by the arrows in the graph below:



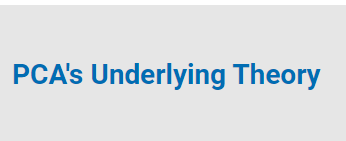




We also see the direction that stretched with the eigenvectors of (2, 1) and (-1, 2):



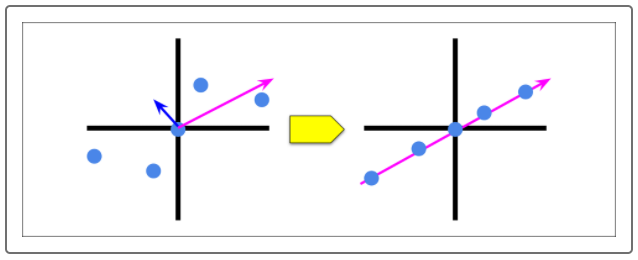
The big takeaway from eigenvectors and eigenvalues is that they show us the spread of the dataset and by how much.



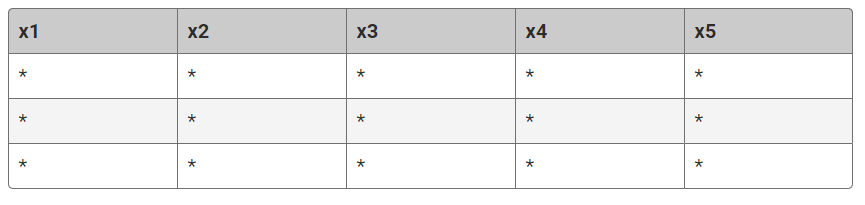
Now it's time to put everything together and show how PCA works. Given our two eigenvalues from before, 7 and 2, take the greater eigenvalue, 7, and eliminate the other since it's less important. The higher eigenvalue is the axis that carries the most amount of information.

We'll also take the corresponding eigenvector, which is (2,1).

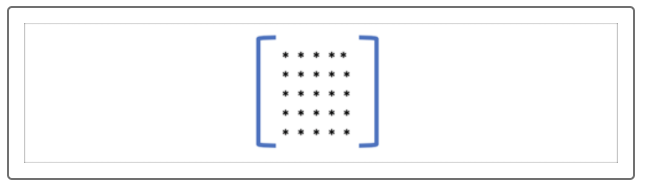
Next, extend that eigenvector with the higher value to a line and project all our points onto that line:



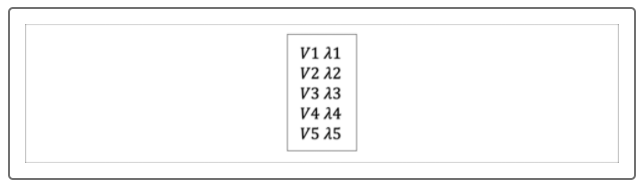
Now let's put everything together and show what PCA is doing. We'll up the ante a little bit and expand from two to five columns of data. First, take our data that consists of five columns, or features. Note, the asterisk (\*) will represent a number as we'll avoid using numbers to simplify the exercise:



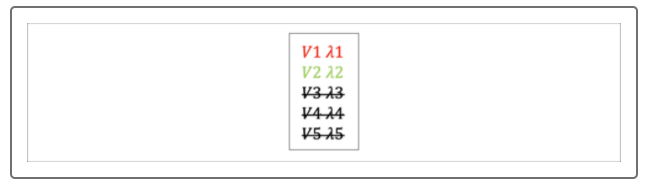
Put all the data points into a 5x5 covariance matrix. The eigenvectors and eigenvalues are calculated for each of those five columns in the matrix. Again, the asterisk (\*) represents a number:



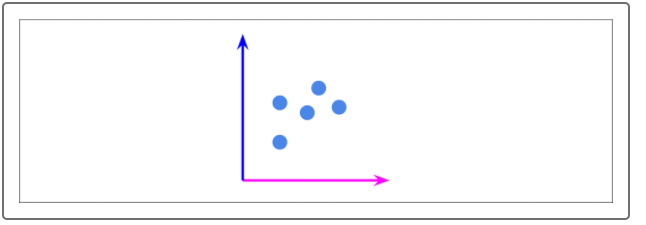
From the matrix we can produce a list of eigenvectors and corresponding eigenvalues:



We pick how many eigenvalues we want to keep and which to drop. For this example, we'll keep the top two eigenvalues and drop three:

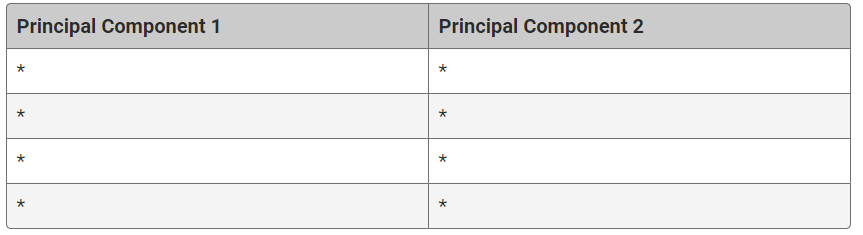


Taking two will allow us to plot on a 2D plane. The two eigenvalues and eigenvectors will create a plane on which all the points can be plotted:



This now narrows down our five features to two and gives us a good snapshot of what the data should look like because we chose the directions the data spread the most.

Finally, these data points will give us a table of two columns, where the asterisk (\*) is a number. Remember, when we coded PCA, the end result was two columns of principal components:



**IMPORTANT**

The statistics, linear transformations, and eigenvalues and eigenvectors all illustrate how PCA works. As you saw earlier, it is much easier to code than do all of this math. So, don't worry if this is confusing—remember, you've already coded it! It is important to understand, on some level, what PCA is doing in case you're ever asked in an interview.

That wraps up how PCA works. (Wow! That was a lot of fancy jargon!) Thankfully, code has made our work easier and now you have a better understanding of how to reduce dimensions yet still keep the values.