Mathematics 327
More principal component analysis

Shown below is the average weekly consumption of 17 types of food in the four countries of the United Kingdom. The units are grams per person per week.

	England	N Ireland	Scotland	Wales
Alcoholic drinks	375	135	458	475
Beverages	57	47	53	73
Carcase meat	245	267	242	227
Cereals	1472	1494	1462	1582
Cheese	105	66	103	103
Confectionery	54	41	62	64
Fats and oils	193	209	184	235
Fish	147	93	122	160
Fresh fruit	1102	674	957	1137
Fresh potatoes	720	1033	566	874
Fresh Vegetables	253	143	171	265
Other meat	685	586	750	803
Other vegetables	488	355	418	570
Processed potatoes	198	187	220	203
Processed vegetables	360	334	337	365
Soft drinks	1374	1506	1572	1256
Sugars	156	139	147	175

Notice that there are four data points in $\mathbb{R}^{17}$. The web page `http://gvsu.edu/s/0Yl` has some Sage cells, the first of which has the data in list form along with two functions:

- `findmean(data)` returns the mean of `data`.
- `demean(data)` returns the matrix of demeaned data points.

Here is another helpful Sage command that you should feel free to use when appropriate:

`A.matrix_from_columns( list )` returns the matrix obtained from A by picking out the columns in `list`. For instance, `A.matrix_from_columns([3,5,8])` returns the matrix obtained from A by putting the fourth, six, and ninth columns of A into a matrix (counting begins at zero).

Looking at the data, does anything stand out?

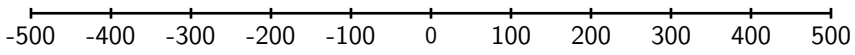
Find the mean and state the average consumption of cereals across the four countries.

Now find the demeaned data matrix A and the covariance matrix C.

Find the eigenvalues of C and list the three nonzero eigenvalues.

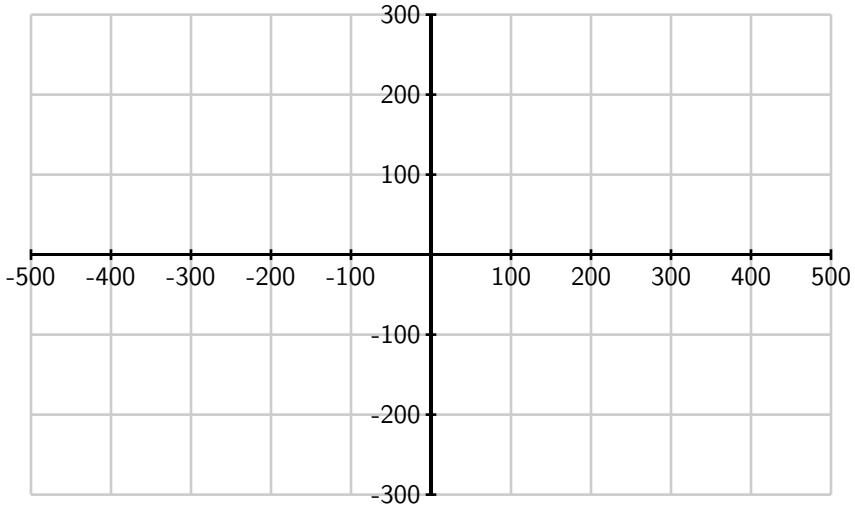
For what percentage of the total variance does the first eigenvalue account?

Let's project the data points onto the direction $\mathbf{u}$ with the greatest variance. For each demeaned data point $\mathbf{x}_i$, find the dot product $\mathbf{x}_i \cdot \mathbf{u}$ and plot the projected points on the line defined by $\mathbf{u}$.



What stands out about this plot?

Let's now project the data point onto the two-dimensional subspace, defined by vectors $\mathbf{u}_1$ and $\mathbf{u}_2$, having the greatest variance. For each demeaned data point $\mathbf{x}_i$, find the dot products $\mathbf{x}_i \cdot \mathbf{u}_j$ and plot the projection below.



Why is this plot wider than it is tall?

For what percentage of the total variance does this two-dimensional subspace account?