

Clustering

Principal Component Analysis (PCA)

The first principal component follows a ``best fit'' through the data points. The data also has maximum variance for PC1.

The two categories that feature most prominently are fresh potatoes and soft drinks. Wales has less consumption of soft drinks, and Scotland has less consumption of fresh potatoes. These two factors cause the most variance along the PC2 axis, which mainly shows the differences between Scotland, England, and Wales.

Class 7: Machine Learning 1

Clustering

We will start with k-means clustering, one of the most prevalent clustering methods.

Let's make some data up:

```
hist(rnorm(10000, mean=3))
```

Histogram of rnorm(10000, mean = 3)

