

# PCA\_FoodConsumption

September 2, 2021

The data food-consumption.csv contains 16 countries in Europe and their consumption for 20 food items, such as tea, jam, coffee, yogurt, and others. We will perform principal component analysis to explore the data.

	Real coffee	Instant coffee	Tea	Sweetener	Biscuits	\
Country						
Germany	90	49	88	19	57	
Italy	82	10	60	2	55	
France	88	42	63	4	76	
Holland	96	62	98	32	62	
Belgium	94	38	48	11	74	
Luxembourg	97	61	86	28	79	
England	27	86	99	22	91	
Portugal	72	26	77	2	22	
Austria	55	31	61	15	29	
Switzerland	73	72	85	25	31	
Sweden	97	13	93	31	61	
Denmark	96	17	92	35	66	
Norway	92	17	83	13	62	
Finland	98	12	84	20	64	
Spain	70	40	40	18	62	
Ireland	30	52	99	11	80	

	Powder soup	Tin soup	Potatoes	Frozen fish	Frozen veggies	\
Country						
Germany	51	19	21	27	21	
Italy	41	3	2	4	2	
France	53	11	23	11	5	
Holland	67	43	7	14	14	
Belgium	37	23	9	13	12	
Luxembourg	73	12	7	26	23	
England	55	76	17	20	24	
Portugal	34	1	5	20	3	
Austria	33	1	5	15	11	
Switzerland	69	10	17	19	15	
Sweden	43	43	39	54	45	
Denmark	32	17	11	51	42	
Norway	51	4	17	30	15	
Finland	27	10	8	18	12	

Spain	43	2	14	23	7
Ireland	75	18	2	5	3

	Apples	Oranges	Tinned fruit	Jam	Garlic	Butter	Margarine \
Country							
Germany	81	75	44	71	22	91	85
Italy	67	71	9	46	80	66	24
France	87	84	40	45	88	94	47
Holland	83	89	61	81	15	31	97
Belgium	76	76	42	57	29	84	80
Luxembourg	85	94	83	20	91	94	94
England	76	68	89	91	11	95	94
Portugal	22	51	8	16	89	65	78
Austria	49	42	14	41	51	51	72
Switzerland	79	70	46	61	64	82	48
Sweden	56	78	53	75	9	68	32
Denmark	81	72	50	64	11	92	91
Norway	61	72	34	51	11	63	94
Finland	50	57	22	37	15	96	94
Spain	59	77	30	38	86	44	51
Ireland	57	52	46	89	5	97	25

	Olive oil	Yoghurt	Crisp bread
Country			
Germany	74	30	26
Italy	94	5	18
France	36	57	3
Holland	13	53	15
Belgium	83	20	5
Luxembourg	84	31	24
England	57	11	28
Portugal	92	6	9
Austria	28	13	11
Switzerland	61	48	30
Sweden	48	2	93
Denmark	30	11	34
Norway	28	2	62
Finland	17	21	64
Spain	91	16	13
Ireland	31	3	9

**0.0.1 Performing PCA on countries by treating each country's food consumption as their feature vectors. Extracting the first 2 principal components and plotting a scatter plot of results.**

Original data matrix (16, 20)

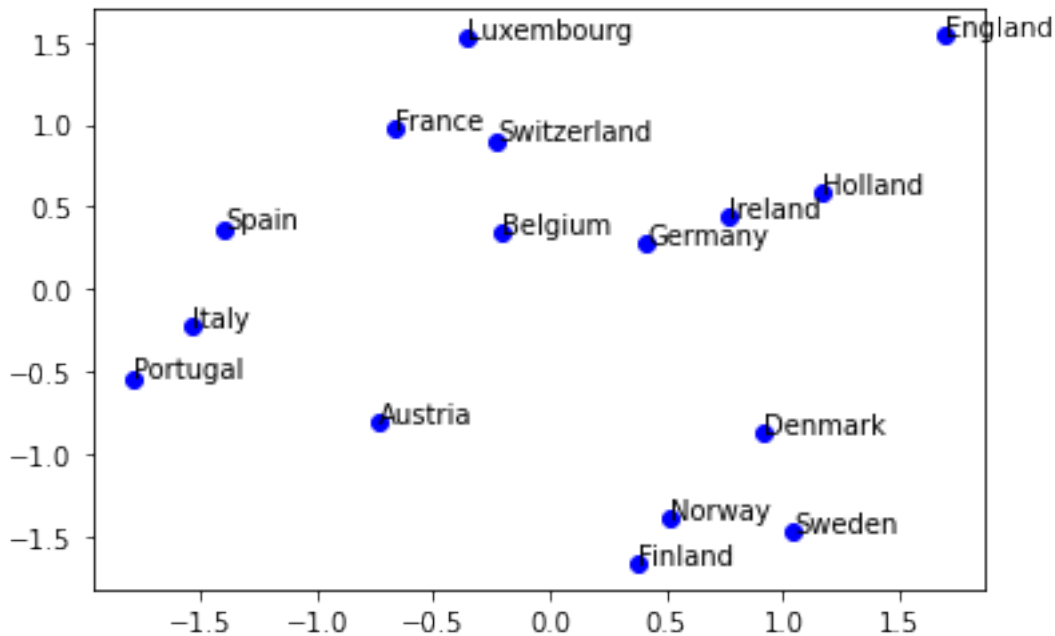
Covariance Matrix is : (20, 20)

(16,)

(16,)

The csv file was imported as a pandas data frame and then converted to a numpy array for applying PCA. We can observe that there are 16 rows in the data, each corresponding to one country and 20 columns, each representing a food item. To this analysis of performing PCA by treating each country's food consumption as their "features", below are the data and covariance matrix used. For this, the features are the food items, which are along the rows and the countries are along the columns in the data matrix. The covariance matrix that captures distance between each point in the data matrix as a  $d \times d$  dimension (with  $d$  = number of features). Features = Food items

Data points = Countries



**Observation and results** From the scatter plot, we can make some interpretation based on the direction along a specific principal component. In this case, the features are the food items. For eg, the X-axis corresponds to the subspace of the first principal component, and based on the position of Portugal and Sweden, we can infer that they have opposite food consumption patterns since they are in the opposite directions of the subspace.

Similarly, when we observe along the direction of the second principal component (Y axis), we can infer that Luxembourg and Austria have opposite food item consumption patterns.

We will not be able to exactly state which food item is mostly consumed in which country since the principal component is a linear combination of all food items, and can help with the direction of food consumption only.

### 0.0.2 Performing PCA analysis on the data by treating country consumptions as feature vectors for each food item. Extracting the first 2 principal components and plotting a scatter plot of results.

Shape of data matrix is: (20, 16)

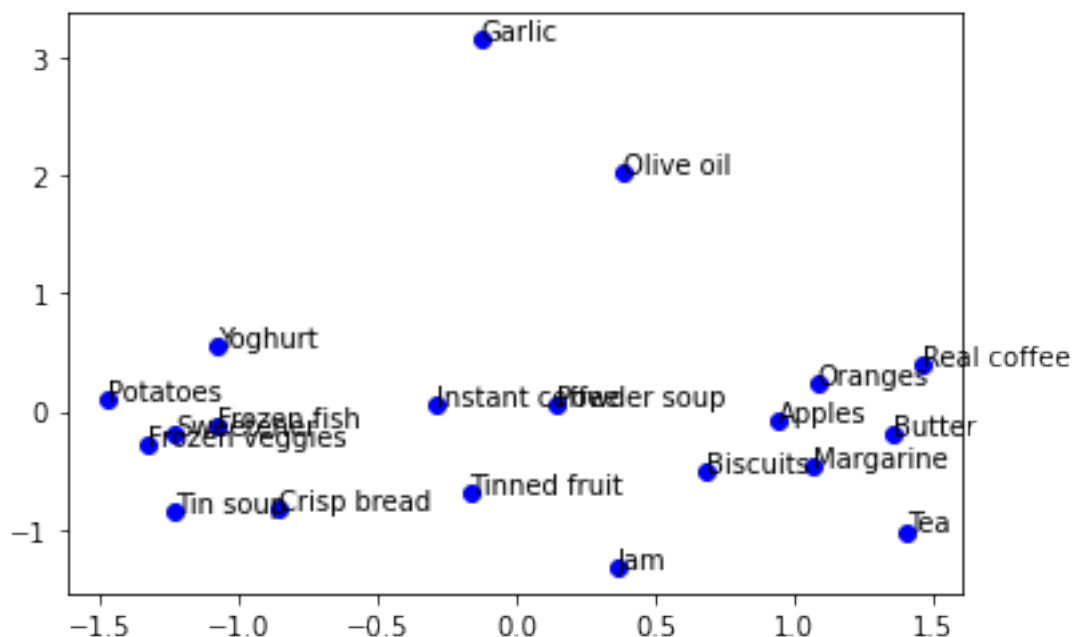
Covariance Matrix is : (16, 16)

(20,)

(20,)

To this analysis of performing PCA by treating each country of consumption as “features”, below are the data and covariance matrix used. For this, the features are the countries that consume the food item, which are along the rows and the food items are along the columns in the data matrix. The covariance matrix that captures distance between each point in the data matrix as a  $d \times d$  dimension (with  $d$  = number of features). Features = Countries

Data points = Food items



**Observation and results** Similar to the above, we can make some inference based on the direction along a specific principal component. In this case, the features are countries.

For eg, the X-axis corresponds to the subspace of the first principal component, and based on the position of Tea and Tin Soup, we can infer those countries these food items are consumed in opposite patterns (countries that consume Tea would not consume Tin soup).

Similarly, when we observe along the direction of the second principal component (Y axis), we can infer that Garlic and Jam have opposite food consumption patterns among the countries.

We will not be able to exactly state which country consumes which food item since the principal component is a linear combination of all countries in this case, and can help with the direction of food consumption only.