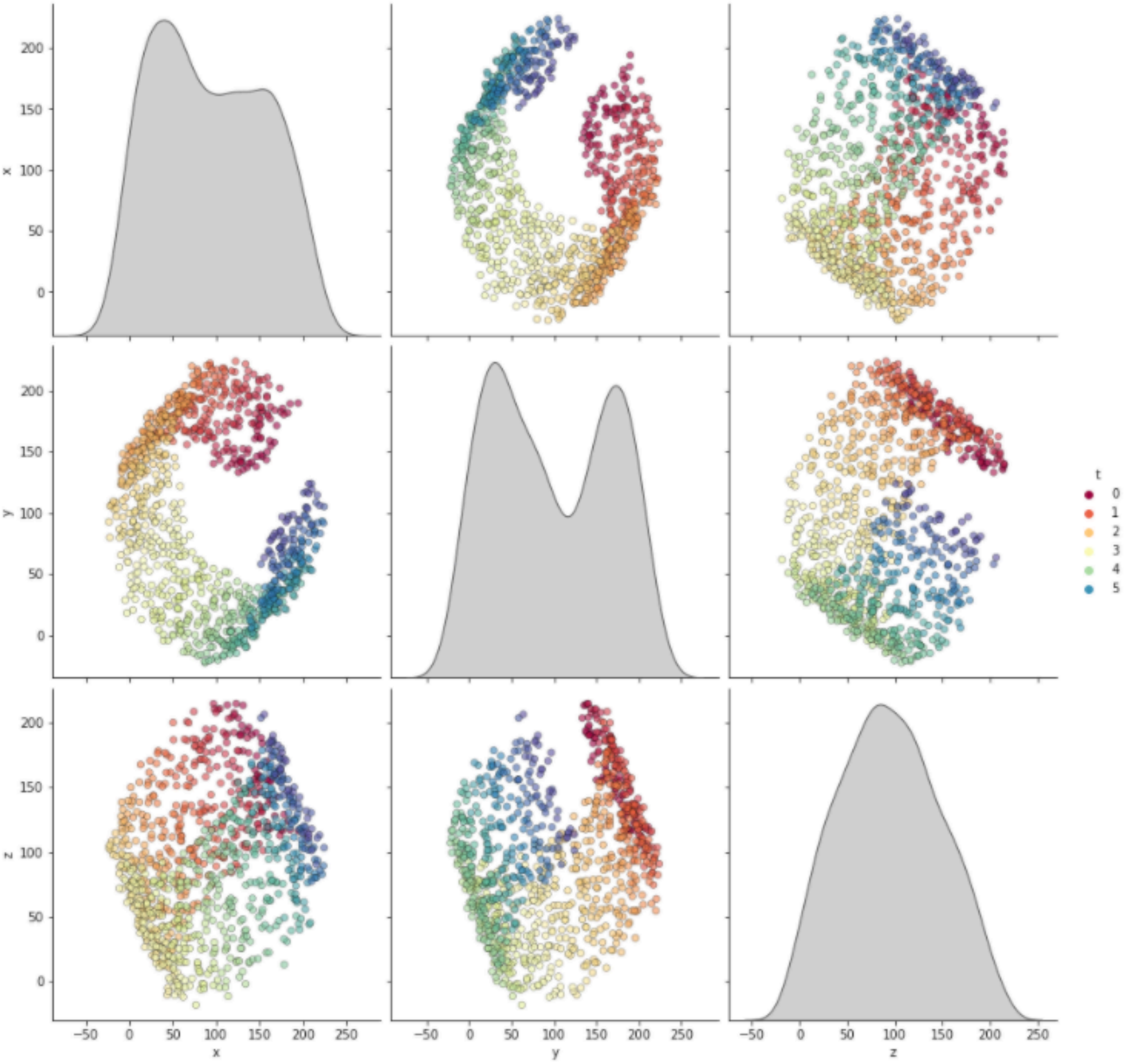
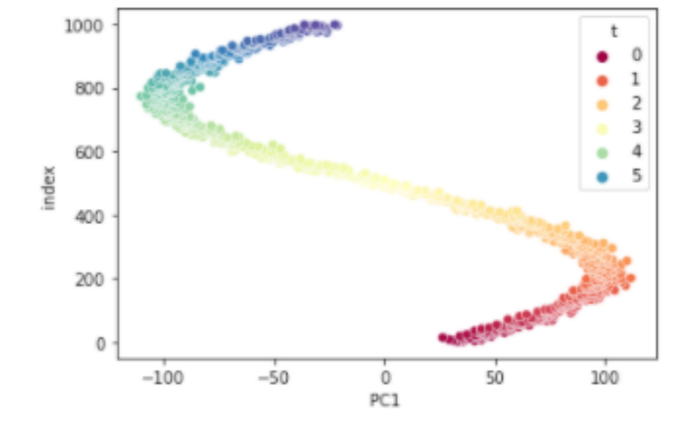
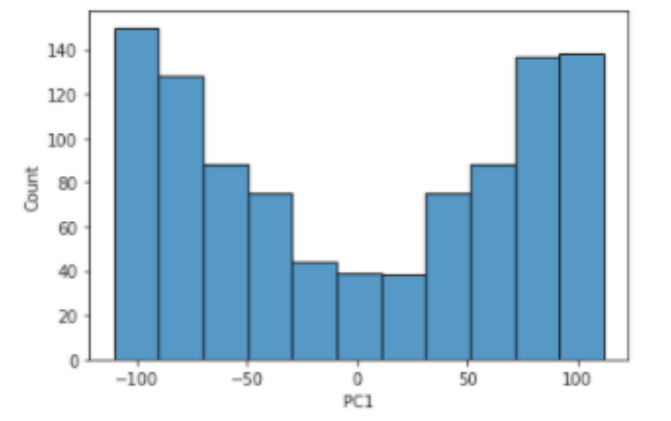
Exercise1

(a)

Each datapoint has unique 't' value in the dataset, which will be used as data point color. I map these 't' value to a spectral palette, therefore each point has unique color. This leads to the difficulty of legend representation. My solution is to make the legend as an index for reader to infer the rough 't' values instead of accurate 't'.



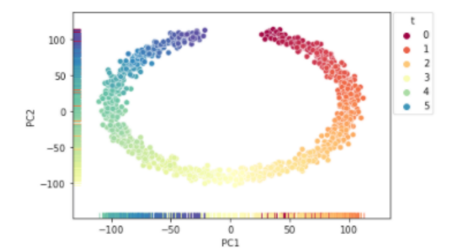
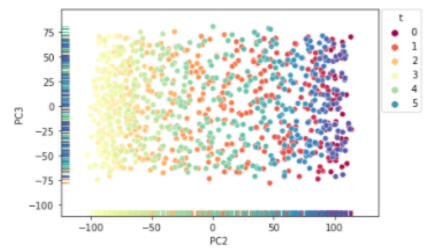
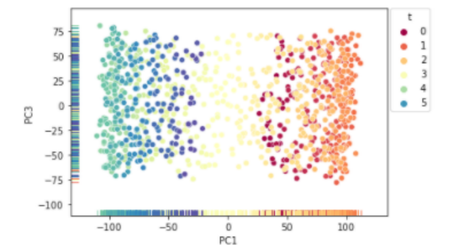
(b) PCA method

1D mapping plot	1D Histogram
	

We need to center the data first.

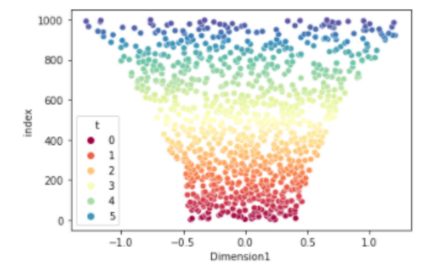
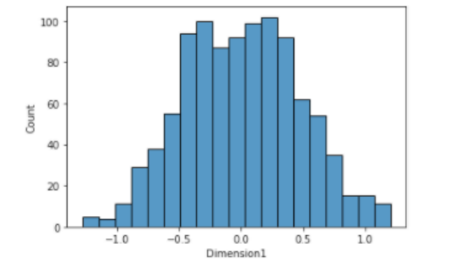
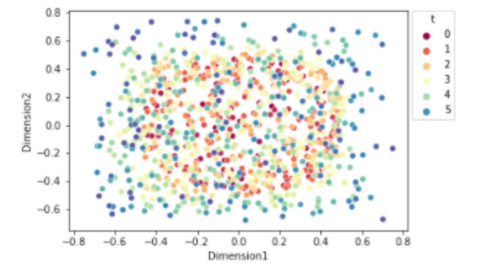
Reason: suppose we have a matrix $\mathbf{X}$ need to do PCA, if we don't apply centering, then the corvariance matrix will become $\mathbf{X}\mathbf{X}^T/(n-1)$. However, according to the definition of corvariance matrix, it should be $(\mathbf{X}-\mu)(\mathbf{X}-\mu)^T/(n-1)$. Essentially, PCA will do singular value decomposition (SVD) on corvariance matrix. So centering the data ensure we get the correct principal components. If not, principal compoents will be affected by the variables' means, which against the maximizing variance rule.

(c) PCA method

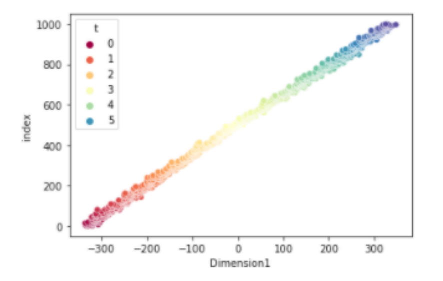
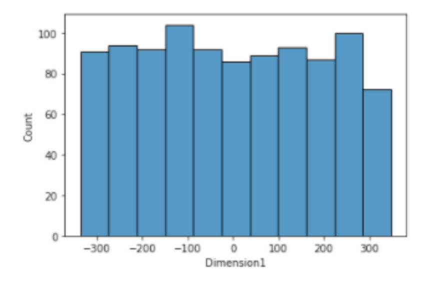
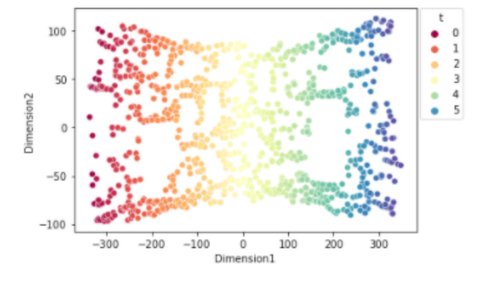
The first and second PCA components	The second and third components	The first and third components
		

The shape is like a gap ring. From the top view, it is a circle but doesn't close(gap ring). From the front and side view, it is a rectangular, which shows the height of the gap ring.)

(d) Nonmetric MDS mapping

1D mapping plot	1D Histogram	2D mapping plot
		

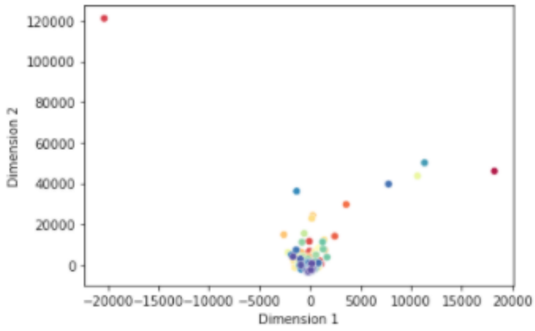
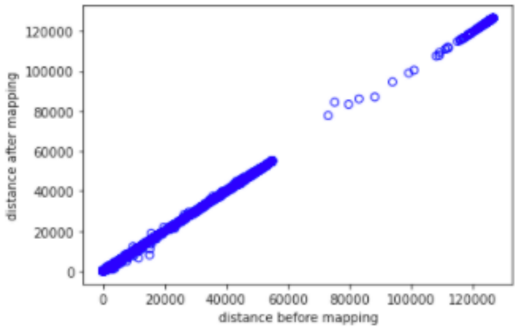
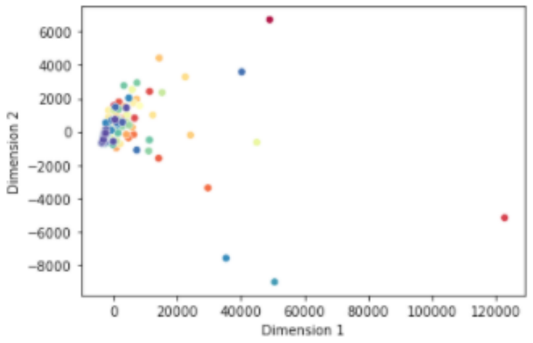
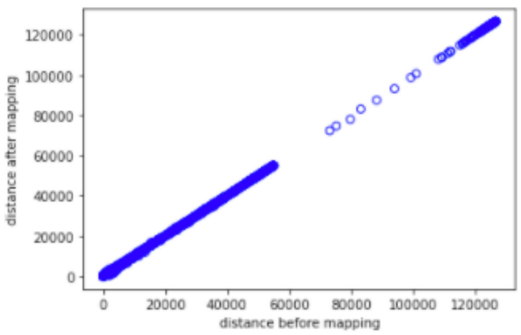
(e) ISOMAP mapping

1D mapping plot	1D Histogram	2D mapping plot
		

Exercise2

I applied simple data preprocessing over the given data. In brief, I remove the first row(the whole country population in different ages), the first two columns(area and total), because there are not comparable with the other data. Apart from these removal, I preserve all the other data[data shape is(309,19)], meaning 309 areas and 19 age groups.

In two mapping methods, I used Euclidean distance to measure the distance between two points.

Mapping method	2D plot	Shepard plot
MMDS		
Sammon		

In 2D mapping plot, each point represents one area, similar colors means area names are close in alphabetical order. According to shepard plot, we could conclude Sammon mapping is slightly better, because true distance and apparent distance are almost the same. And we can visually infer the conclusion by the collinear points in Sammon's shepard plot, meaning zero distortion. MDS has some distortion in short distance(non-collinear points). As compared to MDS, Sammon mapping should be more accurate for shorter distances but less accurate for longer, because Sammon scales the square distance difference by the true distance.

Exercise3

I chose circle to represent nodes and lines as edges because they are the most common and intuitive way. According to the given data, it is clear that we can have two separate graphs, meaning no edges between them.

Design principle:

- Nodes and edges are evenly distributed.
    - In Fig.2, nodes and edges are highly symmetric.
- Edge-crossings should be minimized.
    - In Fig.1, no edge crossings.
- Depict symmetric sub-graphs in the same way

    - In Fig.1, I noticed that there are same sub-structure in the big graph(abc, efg and dfg). So they are presented in the same/symmetric way.
    - In Fig.2, M nodes and F nodes are fully connected. So M nodes should be gathered together and F nodes are the same.
- Minimize the edge bending ratio
    - I chose to use direct lines here. 0 curves.
- Minimize the edge lengths.
    - Given enough space to separate nodes, I tried to minimize edge lengths and also keep the symmetry.
- Maximize data-ink and semantic subgraphs.
    - I tried my best to make elements(nodes, labels and directed edges) distinguishable and tried to discover semantic sub-structure(abc, efg dfg, jklhi, fully connected network) so that readers can notice them in the shortest time.

Conclusion:

I didn't apply all the principles in the course in two graphs. For example, there are edge crossing in Fig.2. The key principle I comply with is try to visualize semantic sub-structure information with minimized data-ink.

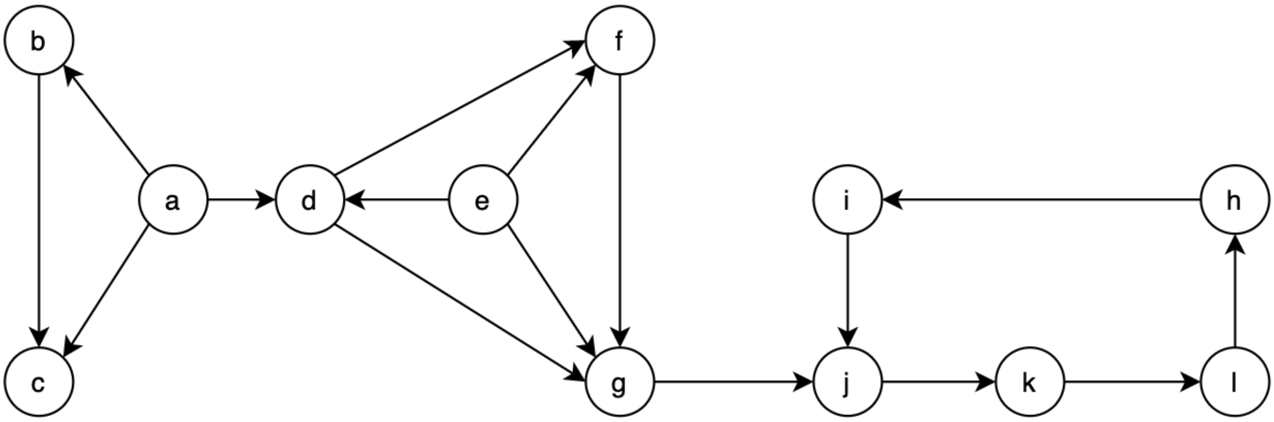


Fig.1

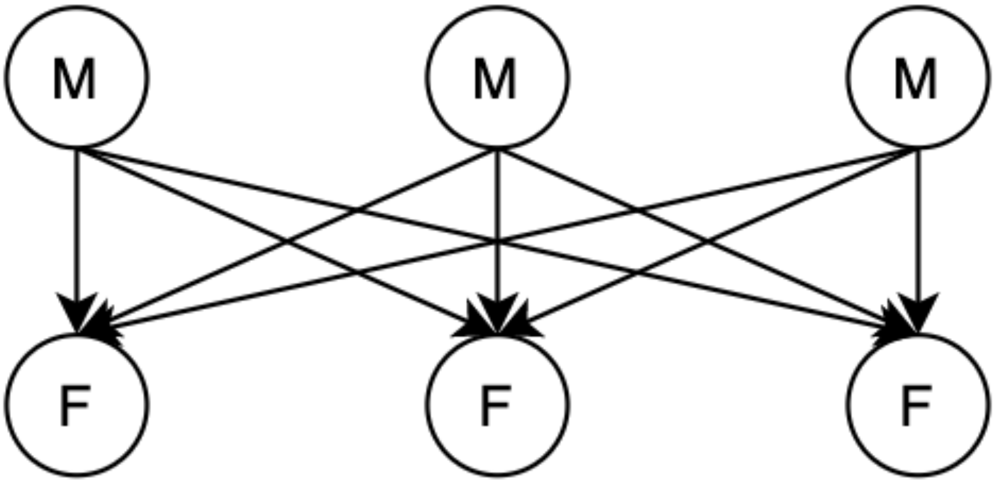


Fig.2