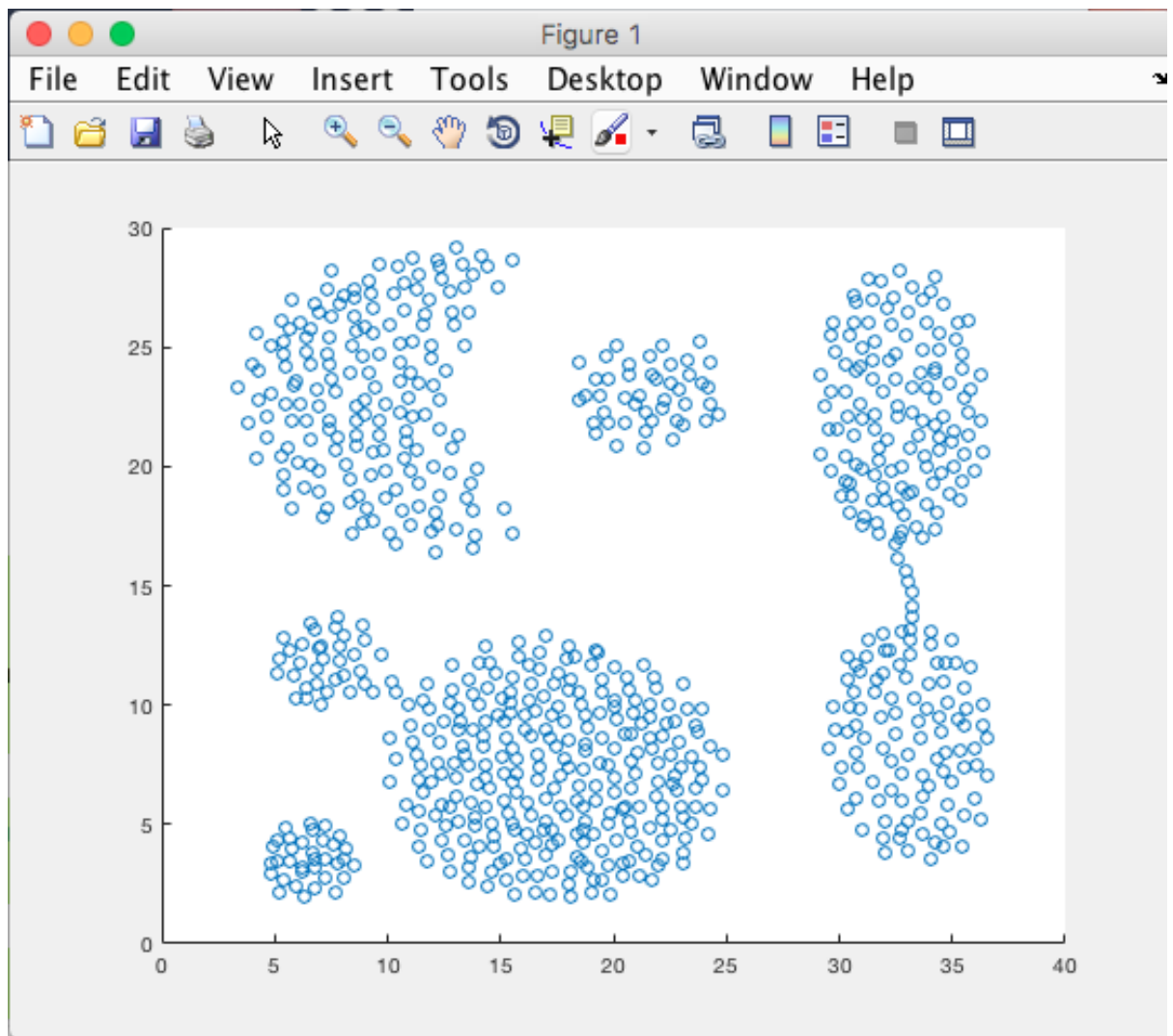


a) Hasil visualisasi data tanpe label data



b)

i. Fungsi yang dibuat

```

main.m x kMeans.m x hitungSSE.m x +
1 function [ finalCentroid,classData ] = kMeans( dataSet,centroid )
2     finalCentroid= zeros(7,2);
3     finalCentroid2 = ones(7,2);
4     c=1;
5     while(finalCentroid2~=finalCentroid)
6         finalCentroid2=finalCentroid;
7         classData = [];
8         for i=1:length(dataSet)
9             euc = [];
10            for j=1:length(centroid)
11                if(c==1)
12                    euc=[euc; norm(dataSet(i,:)-centroid(j,:))];
13                else
14                    euc=[euc; norm(dataSet(i,:)-finalCentroid(j,:))];
15                end
16            end
17            [~, label] = min(euc);
18            classData = [classData;label];
19        end
20        for k=1:length(centroid)
21            groupLabel = find(classData==k);
22            finalCentroid(k,:) = mean(dataSet(groupLabel,:));
23        end
24        c=c+1;
25    end

```

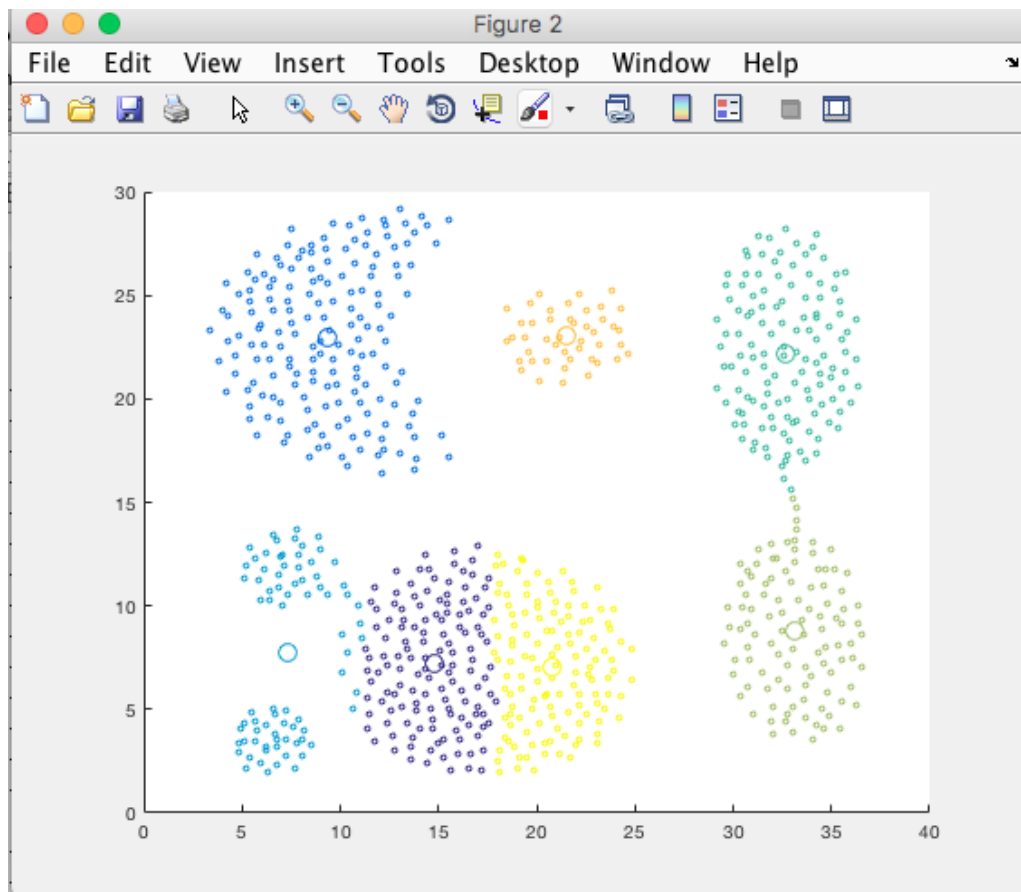
ii. Fungsi nilai SSE

```

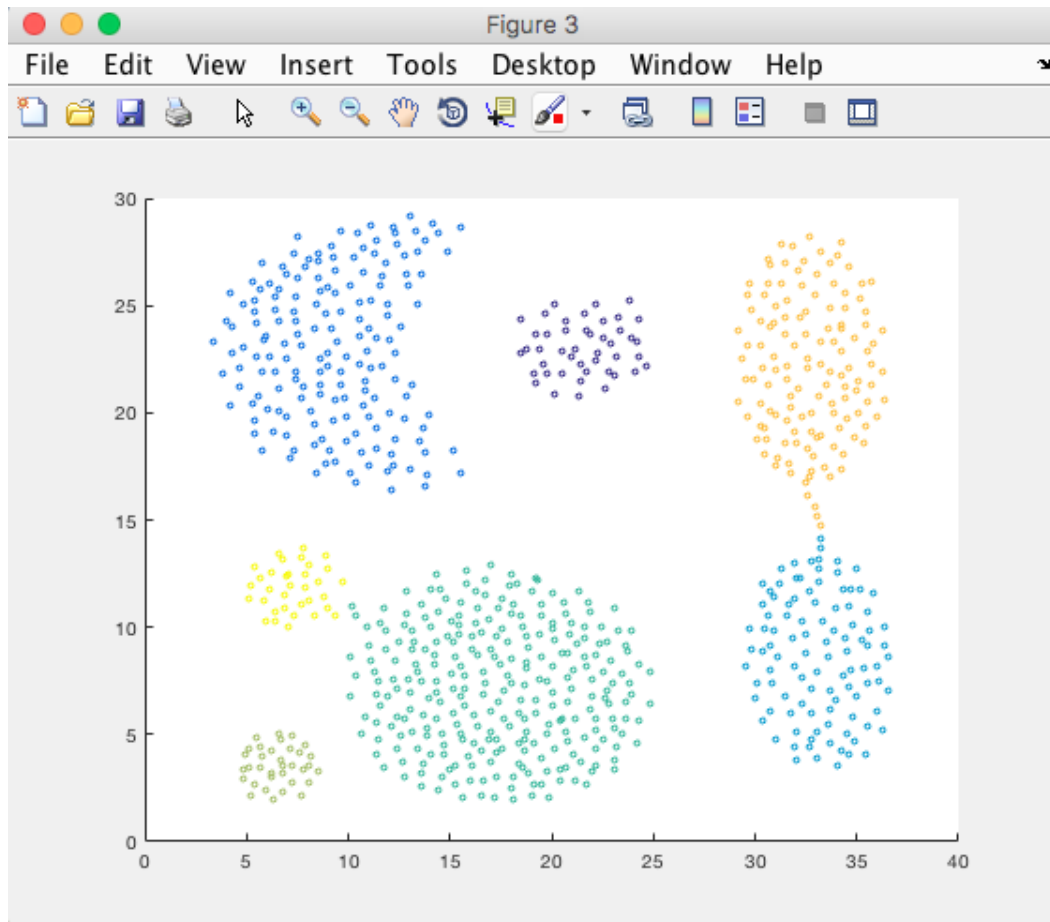
main.m x kMeans.m x hitungSSE.m x +
1 function [ SSE ] = hitungSSE( centroid,dataKluster)
2     SSE=0;
3     for i=1:length(centroid)
4         index = find(dataKluster(:,3)==i);
5         data = dataKluster(index,1:2);
6         for j=1:length(data)
7             SSE = SSE+(norm(centroid(i,:)-data(j,:)));
8         end
9     end
10

```

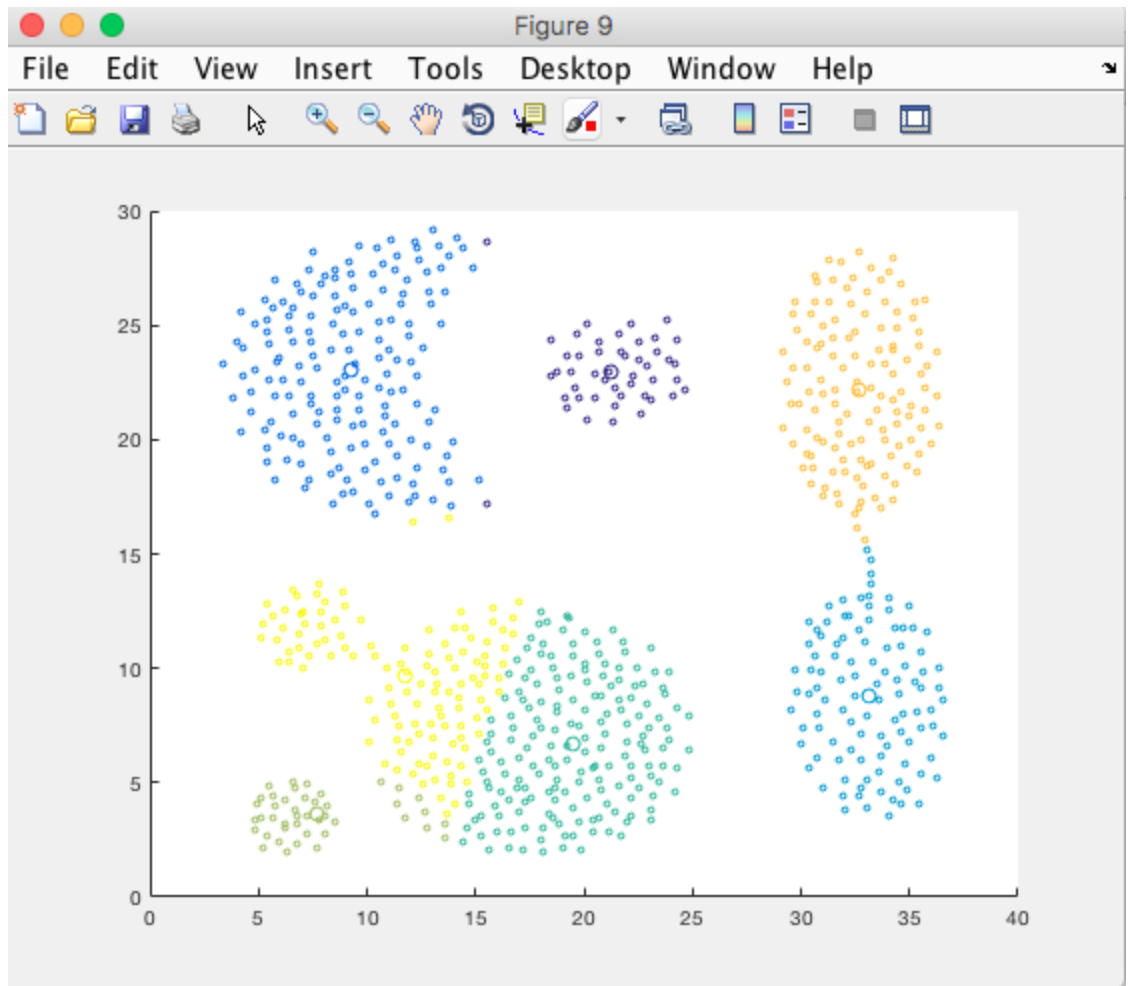
c) Hasil visualisasi dibawah didapat setelah melakukan klustering beberapa kali, dimana klustering dilakukan selama nilai SSE > 2750



d) Visualisasi dari data dengan label yang sebenarnya.



- e) Hasil visualisasi klustering menggunakan k-means dari data 1c tidak jauh berbeda dengan 1d, hal ini dikarenakan nilai  $k$  yang digunakan adalah 7 sesuai dengan jumlah kelas yang ada pada data. Dan juga karena nilai SSE yang cukup kecil sehingga kelas data yang sebenarnya tidak jauh berbeda dengan kelas data hasil klustering.
- f) Hasil visualisasi klustering, dimana *centroid* diambil dari setiap kelas data.



- g) Hasil klustering 1f lebih baik dibanding hasil klustering 1c, dikarenakan pada 1f *centroid* diambil dari setiap kelas data, tidak seperti pada 1c. akan tetapi kedua2nya memiliki hasil yang baik, dikarenakan SSEnya yang tidak terlalu besar.
- h) Klustering menggunakan *centroid* berdasarkan pada setiap kelas data lebih baik dibandingkan memilih *centroid* secara acak. Dikarenakan centroid dapat dipastikan dekat dengan data masing-masing kelas.