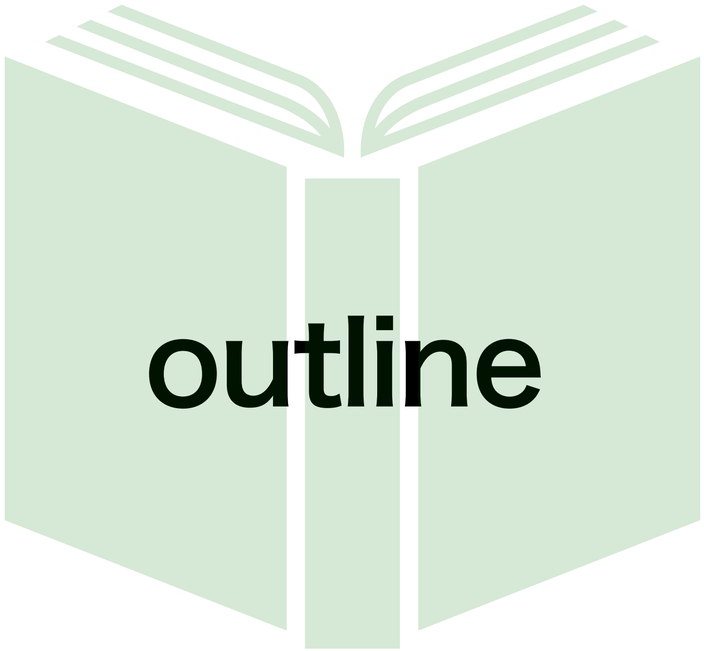
Classify Japan climatic regions using cluster analysis

Research Project A
Computer Science and Engineering, 3rd grade
Yuma Takai – 1w182187
Wednesday, 29, September 2020



1. Background
2. Process
3. Make suitable data for clustering
4. Link latitude and longitude
5. Clustering
6. Discussion
7. Plot
8. Future Work

1. Background

Many researchers classified climate regions in Japan.

A researcher [1] classified
by using a year the daily-average precipitation data.

[1] 草薙浩, (2016). 平年日降水量時系列のクラスター分析による日本の9気候地域区分の提案. 天気, 63(1), 5-12.

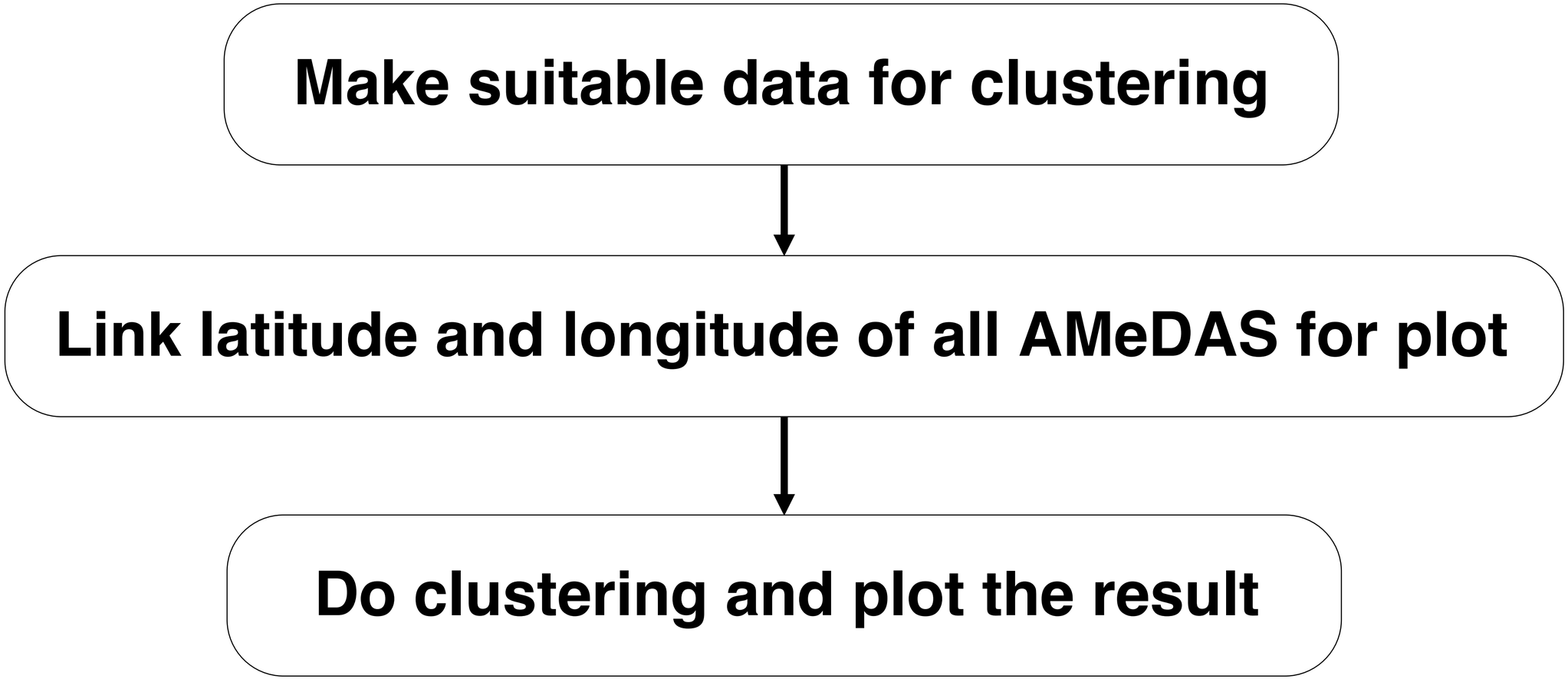
1. Background and Proposal

(1) In Japan, the climate changes significantly from month to month.
(2) The data of the daily-average temperature
also seems to be a good factor for using clustering.

↓

Analyze the data of the daily-average precipitation and temperature
with clustering every month.

2. Process


Make suitable data for clustering
Link latitude and longitude of all AMeDAS for plot
Do clustering and plot the result

3. Make suitable data for clustering

Get data (.csv) from JMA [2] (Japan Meteorological Agency)

(1) Original data have much unnecessary information.
(2) The size of the data available at one time was limited.
(It needs to extract the data for clustering)

- Make some C program and shell script
to remove unnecessary one and to extract the data of
the daily-average precipitation, temperature, date & place name.

[2] JMA, *過去の気象データ・ダウンロード*.
Retrieved 04, 2020, from https://www.data.jma.go.jp/gmd/risk/obsdl/index.php

3. Make suitable data for clustering

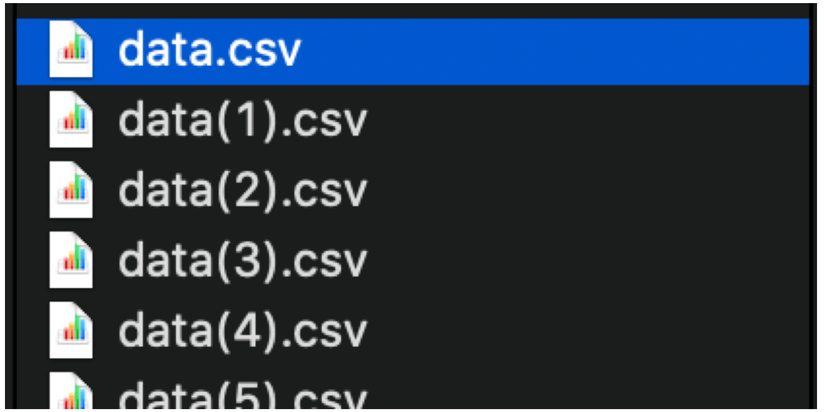


Fig.1 Downloaded CSV file

ダウンロードした時刻：2020/07/20 14:20:31													
	愛知	愛知	愛知	愛知	愛知	愛知	愛知	愛知	愛知	愛知	愛知	愛知	愛知
	一宮	一宮	一宮	一宮	一宮	一宮	一宮	一宮	一宮	一宮	豊山	豊山	
年月日	降水量の合計(mm)	降水量の合計(mm)	降水量の合計(mm)	降水量の合計(mm)	降水量の合計(mm)	平均気温(℃)	平均気温(℃)	平均気温(℃)	平均気温(℃)	平均気温(℃)	降水量の合計(mm)	降水量	
				平年値(mm)	平年値(mm)				平年値(℃)	平年値(℃)			
		品質情報	均質番号		品質情報		品質情報	均質番号		品質情報		品質情	
2019/1/1	0.0	8	1	1.7	8		0	1		0			
2019/1/2	0.0	8	1	1.8	8		0	1		0			

Fig.2 Contents of the CSV file

3. Make suitable data for clustering

Fig.3 After making suitable data for clustering

4. Link latitude and longitude

Get geographic data from JMA. [3]

- From the place name, make a list of the geographic data.

```
北見,キタミ,KITAMI,43,46.6,143,50.5,0104
境野,サカイノ,SAKAINO,43,42.4,143,38.6,0184
常呂,トコロ,TOKORO,44,06.9,144,02.2,0003
斜里,シャリ,SHARI,43,53.1,144,42.0,0015
津別,ツベツ,TSUBETSU,43,42.1,144,02.0,0100
湧別,ユウベツ,YUBETSU,44,12.8,143,37.1,0005
滝上,タキノウエ,TAKINOUE,44,10.6,143,03.7,0165
白滝,シラタキ,SHIRATAKI,43,51.9,143,09.2,0475
紋別,モンベツ,MOMBETSU,44,20.7,143,21.3,0016
網走,アバシリ,ABASHIRI,44,01.0,144,16.7,0038
美幌,ビホロ,BIHORO,43,46.2,144,10.3,0060
興部,オコッペ,OKOPPE,44,28.2,143,06.5,0008
```

Fig.4 List of the geographic data

[3] JMA, *観測概要と観測所一覧*.
Retrieved 04, 2020, from https://www.data.jma.go.jp/obd/stats/data/mdrr/man/kansoku_gaiyou.html

5. Clustering

Clustering by using scipy.

- Normalization before clustering
- Clustering by ward and euclidean.

From the dendrogram and the graph of each cluster,
decide the number of clusters.

5. Clustering

Dendrogram

- The figure that integrated similar data until all data are integrated.

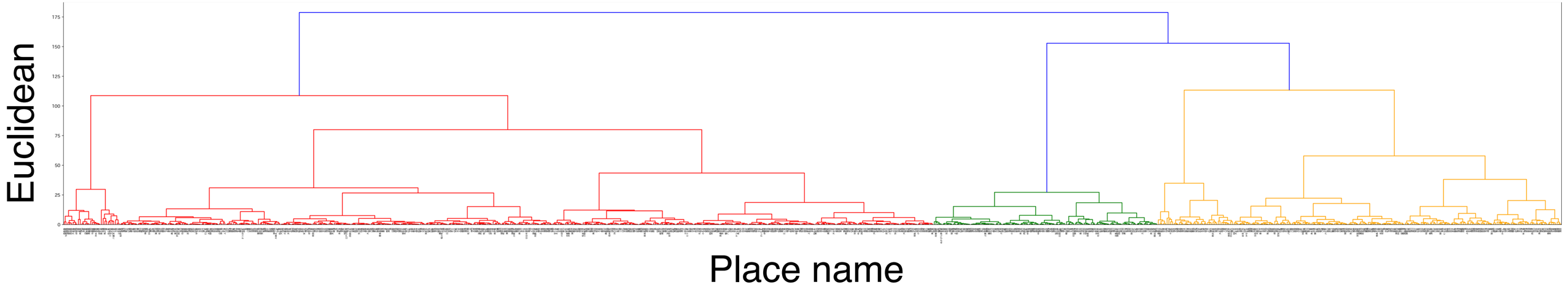


Fig.5 Dendrogram in January

5. Clustering

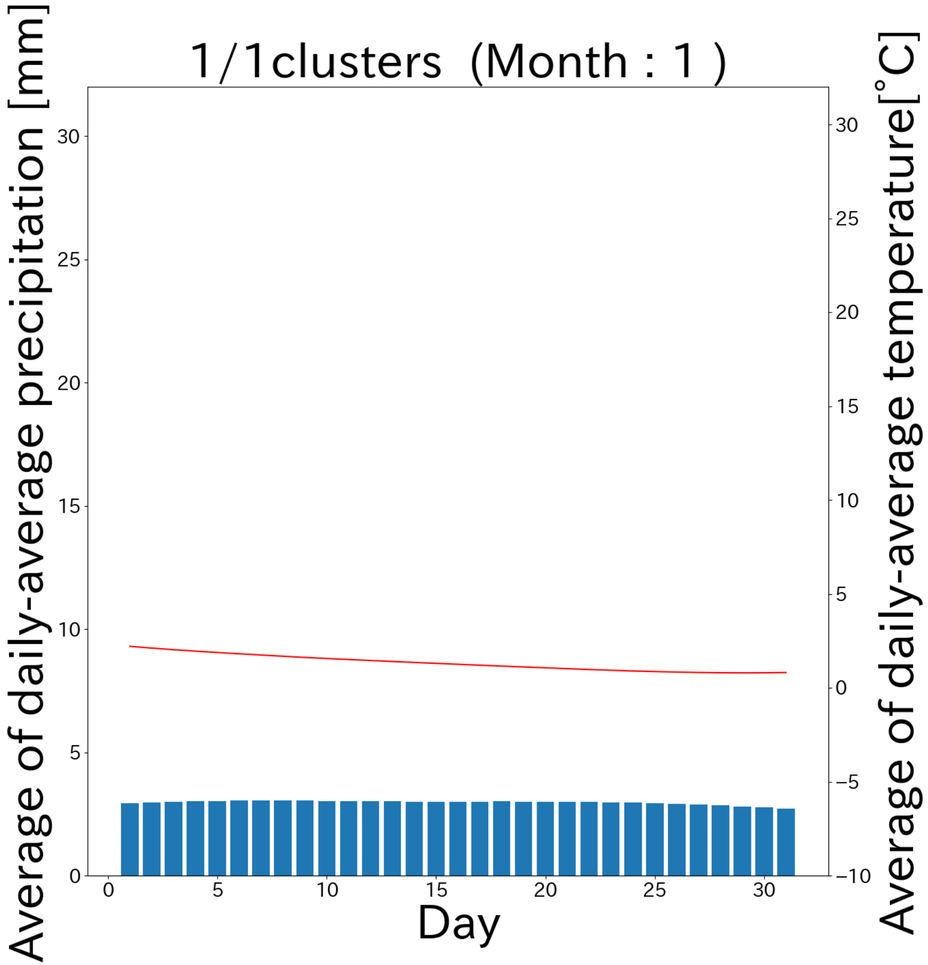
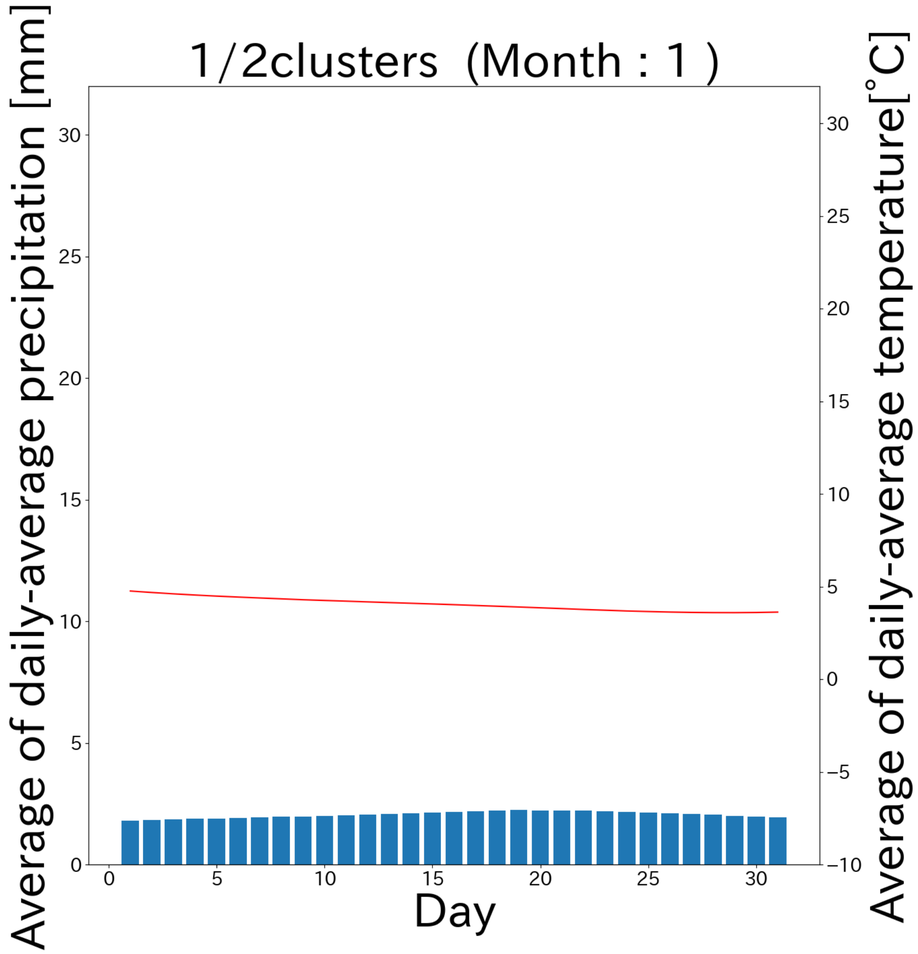


Fig.6 Average score of each cluster in January (1 cluster)

5. Clustering



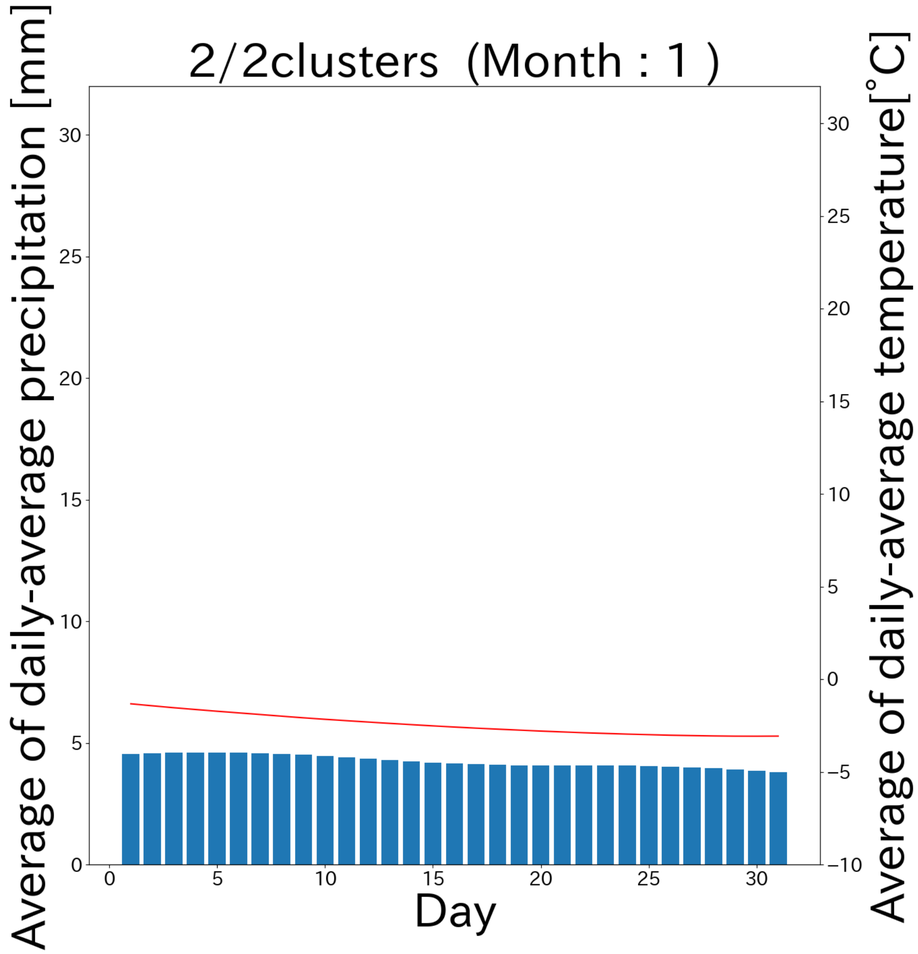


Fig.7 Average score of each cluster in January (2 clusters)

5. Clustering

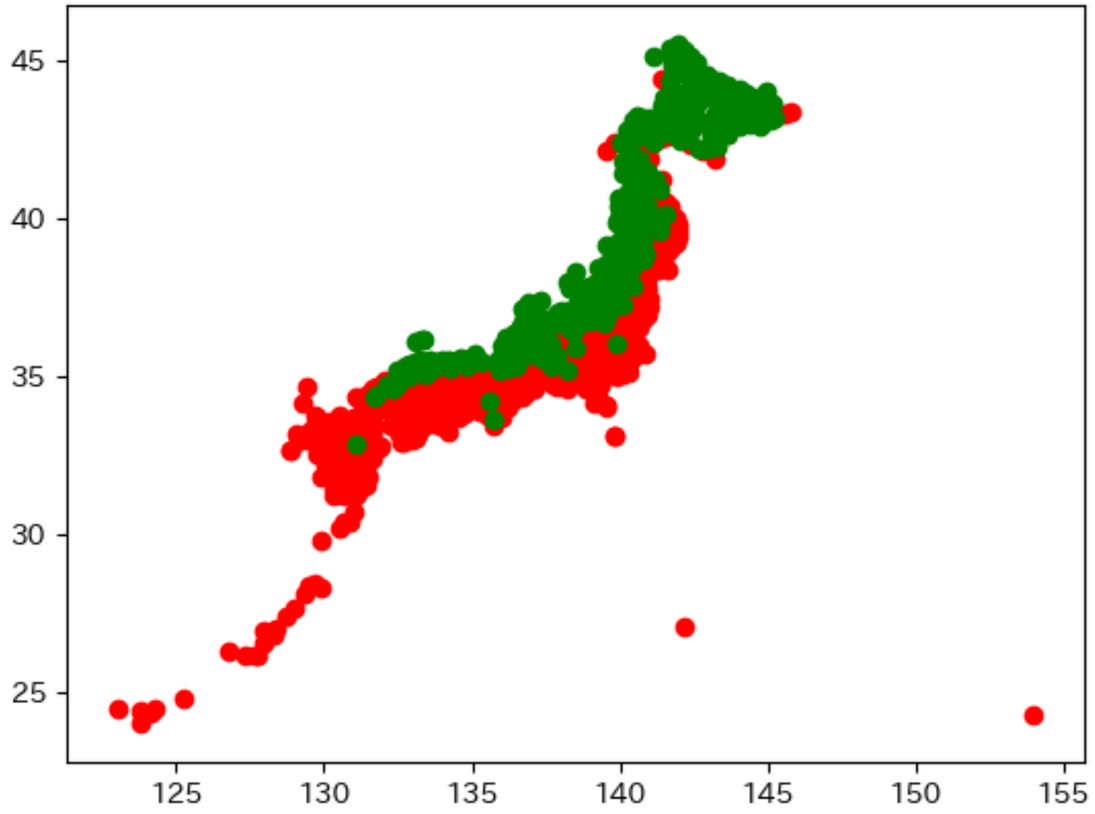
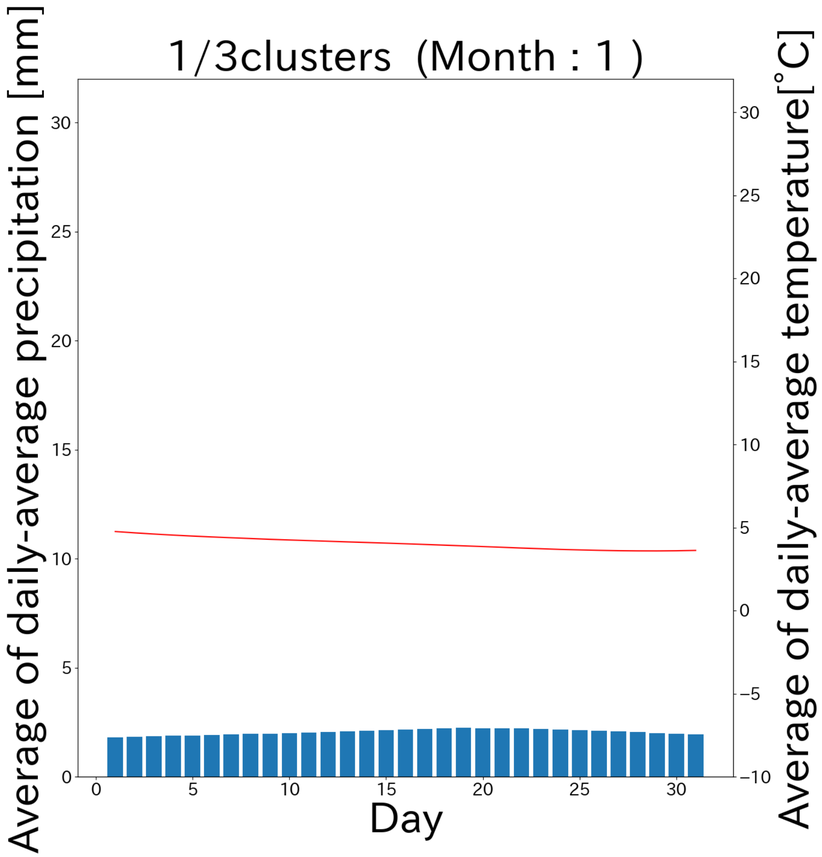
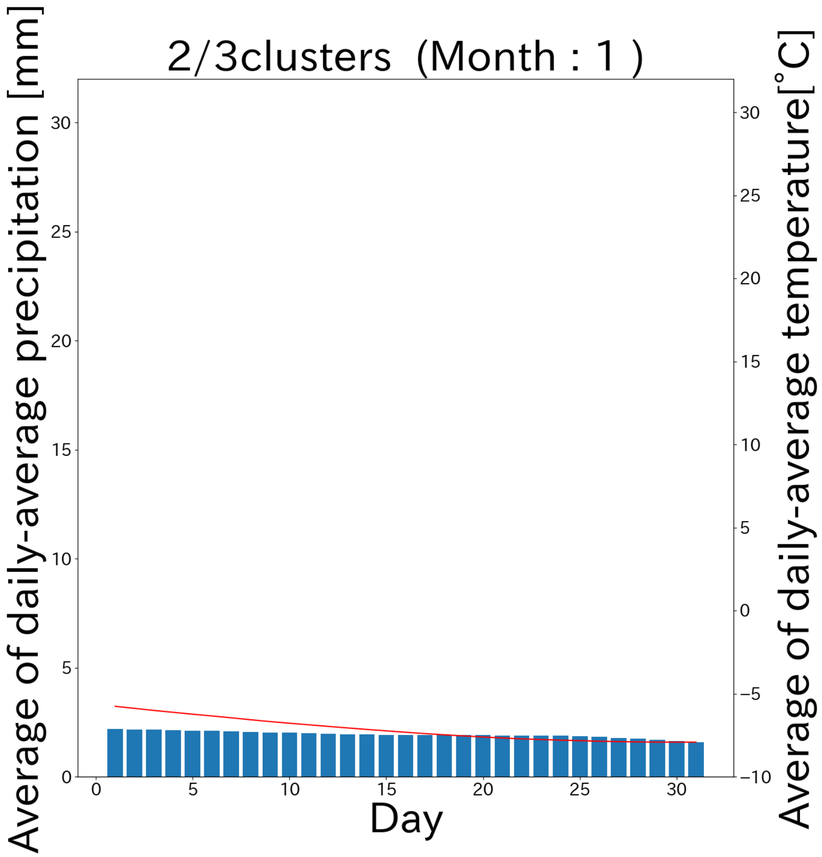


Fig.8 Scatter of each cluster in January (2 clusters)

5. Clustering





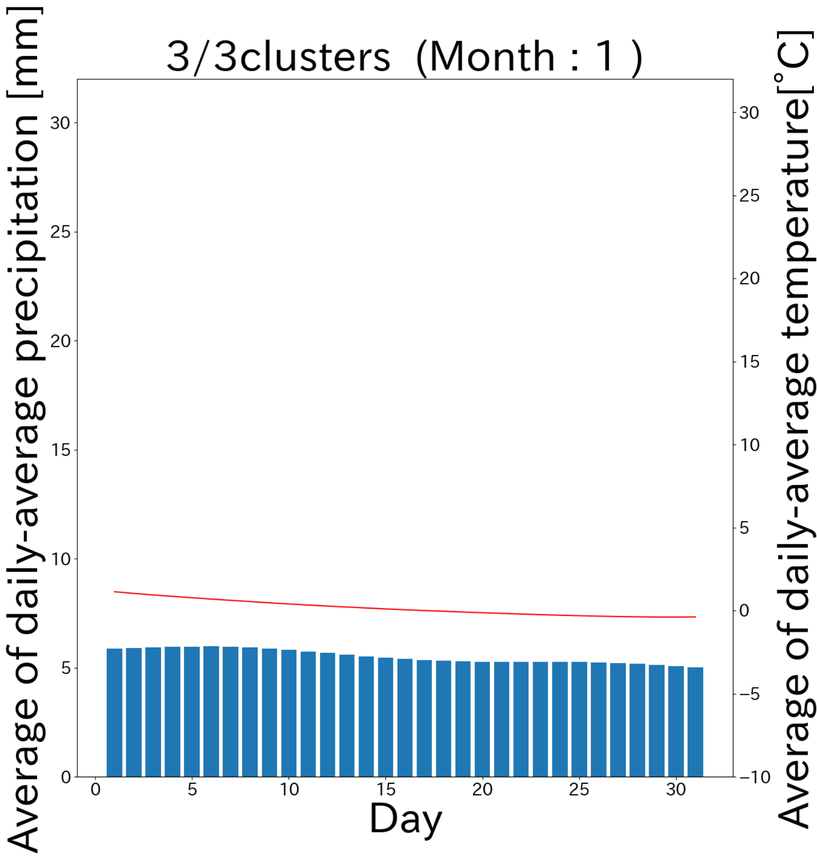


Fig.9 Average score of each cluster in January (3 clusters)

5. Clustering

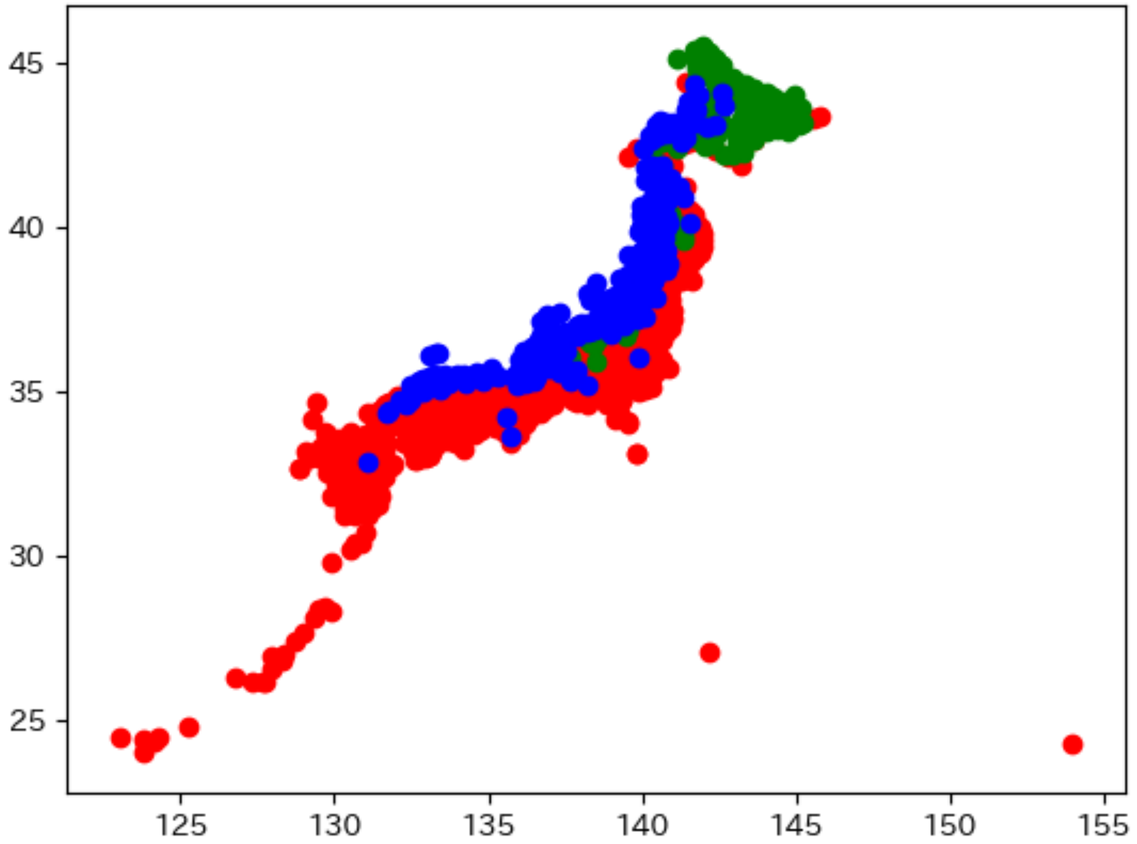


Fig.10 Scatter of each cluster in January (3 clusters)

5. Clustering

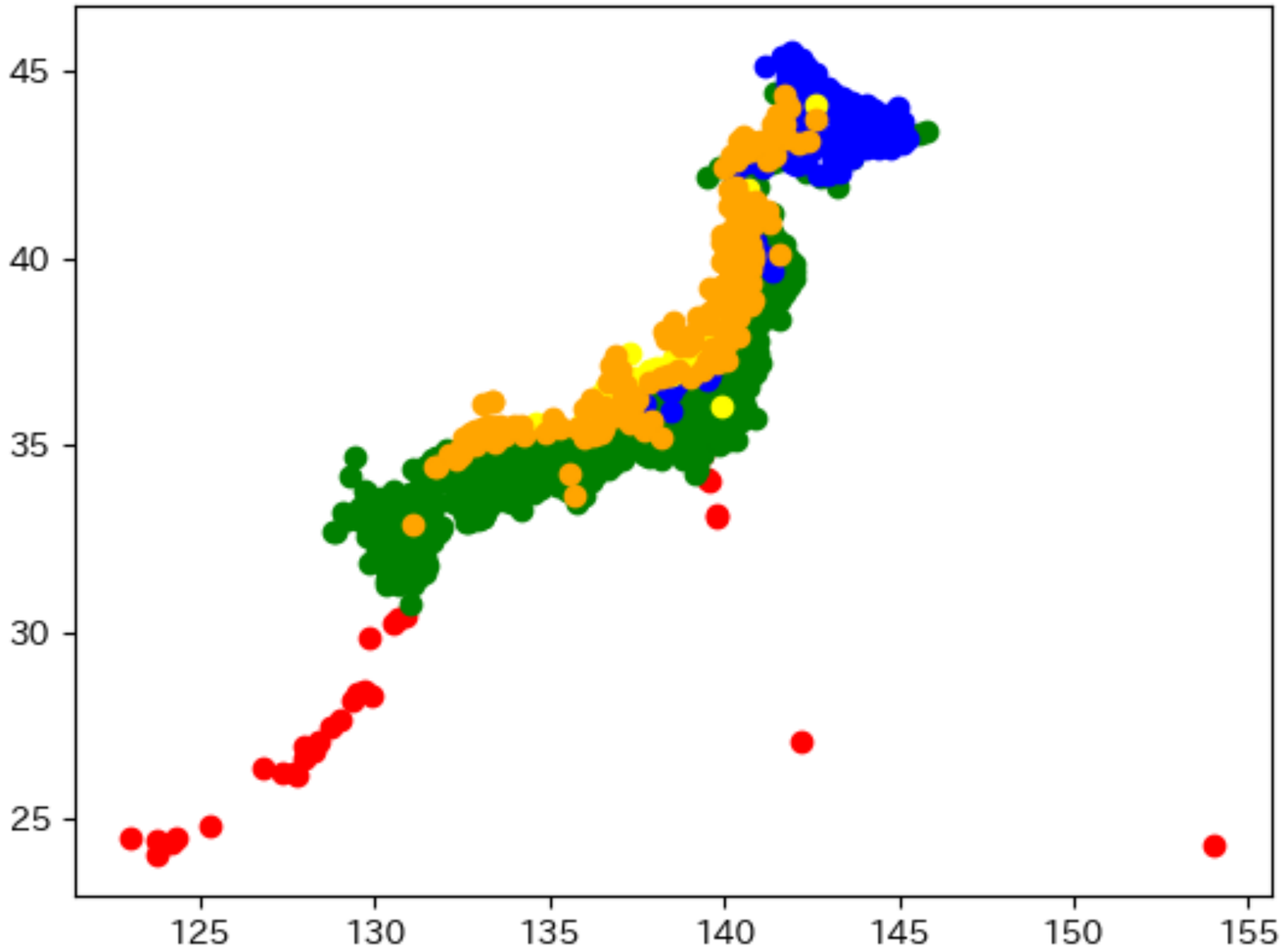


Fig.11 Scatter of each cluster in January (5 clusters)

5. Clustering

Name	Representative location	Number	precipitation	Temperature
Island	Okinawa	31	High	Very high
Pacific	Tokyo	460	Low	High
Eastern Hokkaido	Kushiro	126	Low	Very Low
Sea of Japan (Heavy Snow)	Kanazawa	42	Very High	Low
Sea of Japan	Akita	185	High	Low

Table.1 Features of each cluster in January (5 clusters)

6. Discussion

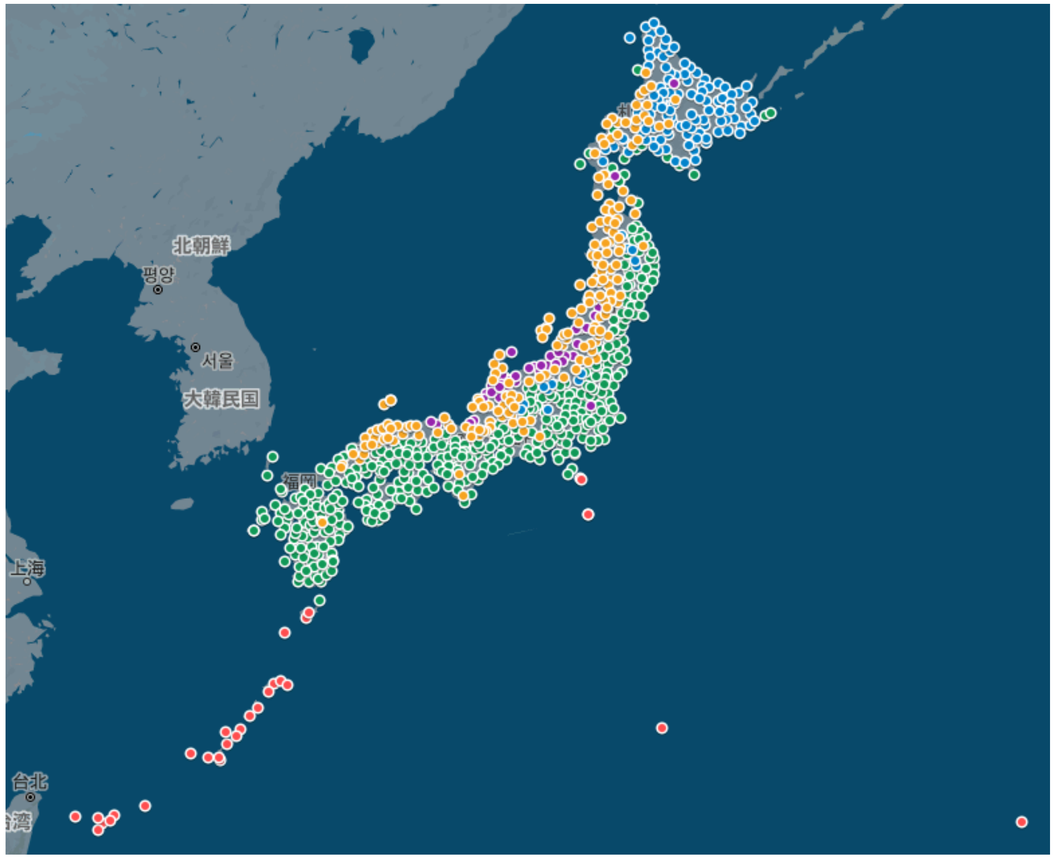
Good climate analysis on January.

- There is a lot of snow on the Sea of Japan side
  and little precipitation on the Pacific side in January.

- Warmer in the south, cooler in the north.

7. Plot

Plot the data by using Google Map



Name	Color
Island	Red
Pacific	Green
Eastern Hokkaido	Blue
Sea of Japan (Heavy Snow)	Purple
Sea of Japan	Orange

Fig.12 Result of classification in January (5 clusters)

7. Future Work

Is there a better way to decide cluster numbers ?

- I decided on a hunch.

Proposal

- The number of AMeDAS that included the cluster
- The information of each cluster (euclidean)

Reference

[1] 草薙浩, (2016). 平年日降水量時系列のクラスター分析による日本の9気候地域区分の提案. 天気, 63(1), 5-12.

[2] JMA, *過去の気象データ・ダウンロード*. Retrieved 04, 2020, from https://www.data.jma.go.jp/gmd/risk/obsdl/index.php

[3] JMA, *観測概要と観測所一覧*. Retrieved 04, 2020, from https://www.data.jma.go.jp/obd/stats/data/mdrr/man/kansoku_gaiyou.html