An empirical study of improving age-specific rates projection accuracy on small populations

18th International Longevity Risk and Capital Markets Solutions Conference

September 8, 2023

Hsin-Chung Wang, Aletheia University

Jack C. Yue, National Chengchi University

Hong-Chih Huang, National Chengchi University

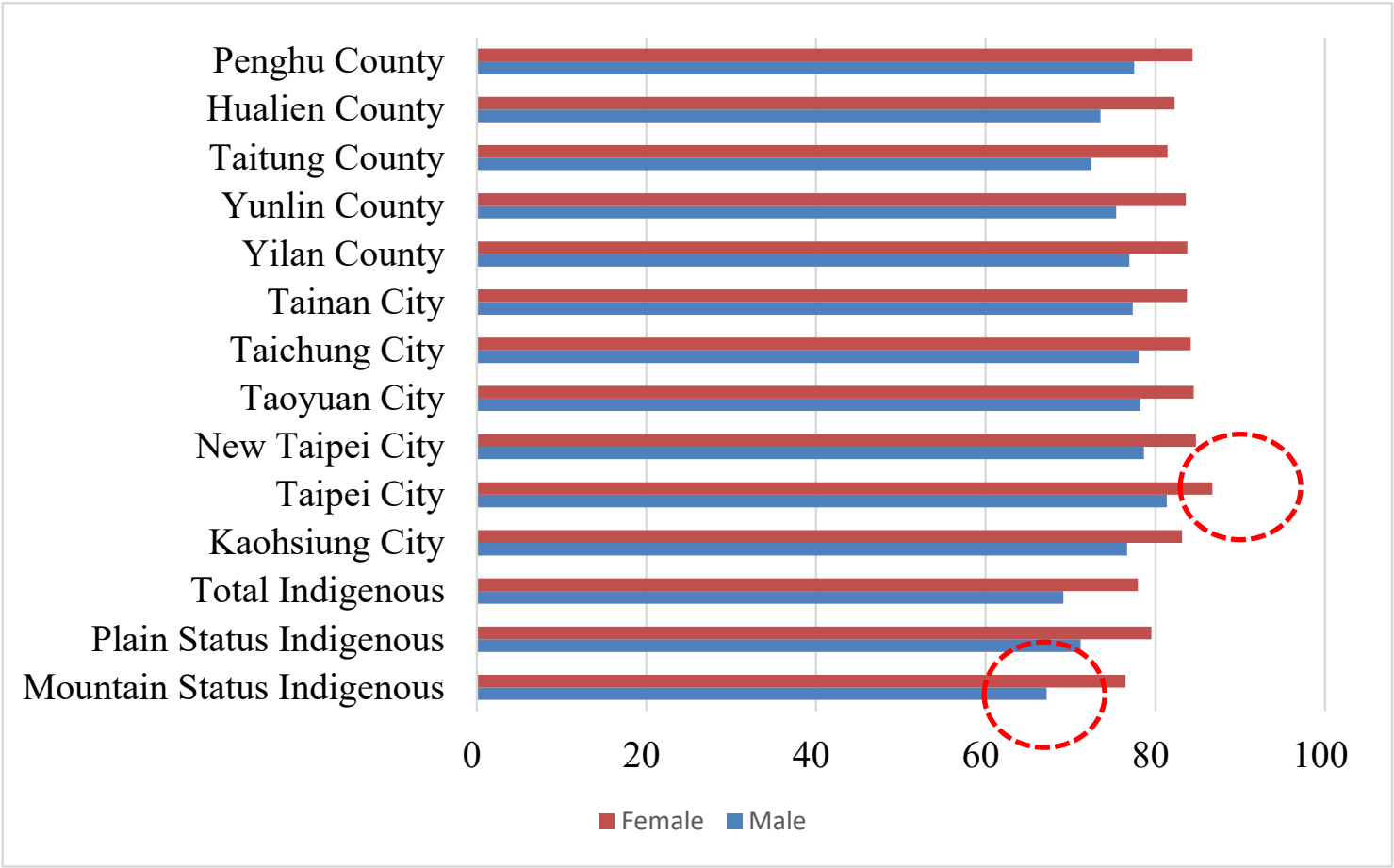
Outline

- Motivation
- Methodology
- Exploratory Data Analysis (EDA)
- Classification
- Construct Districts/ Townships Life tables
- Conclusion

Motivation

- Geographic inequalities in life expectancy among Taiwan's Townships/Districts

The life expectancy at birth among the metropolitan area and selected counties with indigenous populations. ➔The life expectancy(2019-21 Average) of females in Taipei City exceeds that of male mountain indigenous people by 20 years.

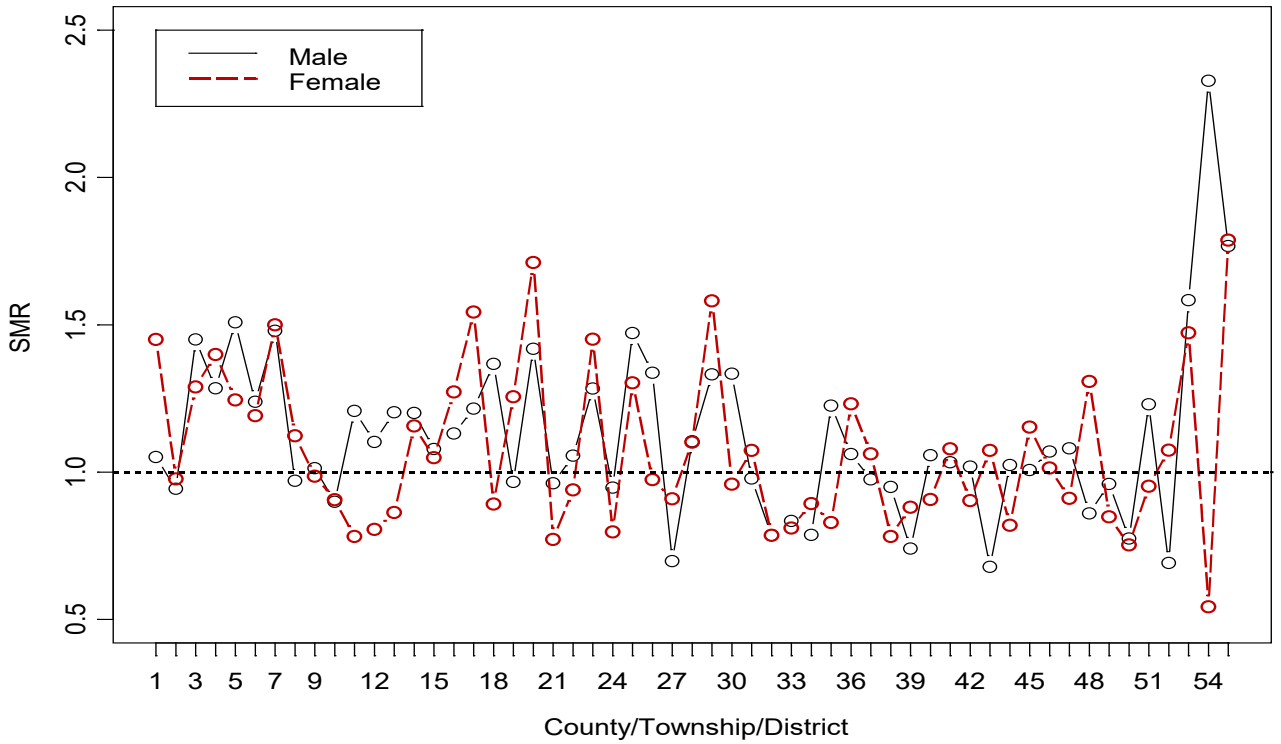


Motivation

- Geographic inequalities in life expectancy among Taiwan Taiwan's Townships/Districts

➢ Half of the 55 indigenous townships have a higher mortality risk than the overall indigenous population.

Note: An SMR value less than 1 indicates a mortality rate lower than that of the overall indigenous population.



Motivation

- Life Expectancy Prediction on Small population

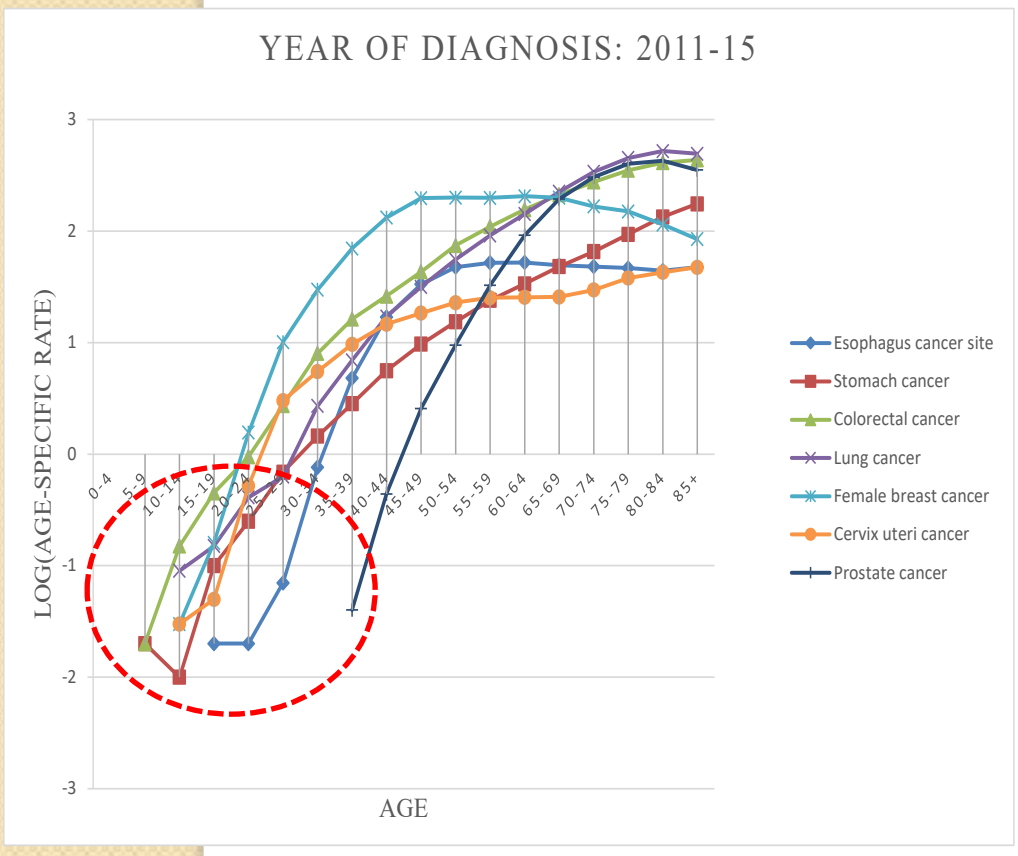
- Alexander et al. (2017) pointed out that reliable estimates of mortality rates for small populations are crucial when studying health inequalities in a country.
- Lee (2003) indicated that, for instance, in counties or townships, specific age-specific mortality rates can become unstable, leading to the disruption of any meaningful patterns or trends that might exist in the age-specific mortality curve due to random noises
- Chen et al. (2017) found that changes in population size significantly influence model parameter estimation and prediction.

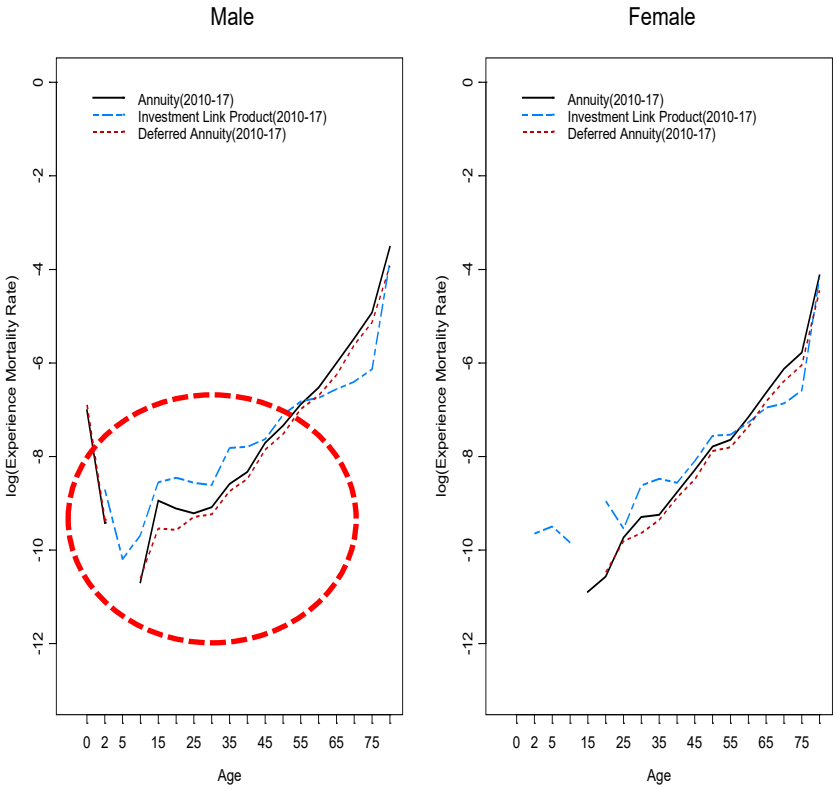
The county-level population sizes often are small and their observed age specific mortality rates fluctuate a lot for consecutive ages.

Motivation

- Life expectancy prediction on Small population

The same challenges of age-specific mortality curve fitting is observed in various counties and townships in Taiwan, incidence of different cancers (left panel), incidence experience rate for Taiwan Life Insurance Policies (right panel),… and so on.





Motivation

- How to solve the problem ?

Menzietti et al. (2019) compiled several papers that discuss borrowing of a reference population to enhance the forecasting accuracy of smaller populations.

➔Indeed, it is feasible to increase the sample size of the small population and reduce variability by utilizing reference populations.

- How to select reference populations?

Usually, we use methods that involve combining data across years or age groups, or grouping populations with the same administrative regions.

- However, neighboring townships or regions might possess significantly different socioeconomic characteristics, and their age-specific occurrence rate curves may also differ.

- In our empirical research process, we classify many small populations using big data classification methods based on three mortality indexes: crude death rate (CDR), standardized death rate (SDR), and Standardized Mortality Ratio (SMR).
- Lee (2003) heterogeneity index will be used to validate the reasonability of tween reference populations with small populations.

Menzietti, M., Morabito, M. F., and Stranges, M. (2019). Mortality Projections for Small Populations: An Application to the Maltese Elderly. Risks, 7(2), pp. 1–35.

Methodology

Mortality Indexes:

- Crude Death Rate(CDR)

CDR = Total Resident Deaths / Total population

- Crude Death Rate is the total number of deaths to residents in a specified geographic area(city, district, county, township) divided by the total population for the same geographic area.

- Standardized Death Rate(SDR)

$$SDR = \frac{\sum(m_x \times P_x^S)}{\sum P_x^S}$$

-m_x age-specific death rate of study population for age group x

-P_x^S No. of persons of standard population for age group x

Methodology

- The standard mortality ratio (SMR)

$$\text{SMR} = \frac{\sum_x d_x}{\sum_x n_x \cdot u_x^*}$$

- d_x is the death number of small area at age of x
- n_x is the population of small area at age of x
- u_x^* is the mortality rate of standard population at age of x.

- The SMR is calculated using the same age structure
- The smaller the SMR, is usually the larger the life expectancy is.

.

Methodology

- The Whittaker method is to minimize the following objective function, i.e., weighted sum of fit function F and smoothness function S:

$$\mathrm{M} = \mathrm{F} + hS = \sum_{x=1}^{\mathrm{n}} w_x (v_x - u_x)^2 + h \sum_{\mathrm{x}=1}^{\mathrm{n-z}} (\Delta^z v_x)^2$$

- u_x and v_x are observed and graduated mortality rates for age x
- w_x is the weight for age x
- n is population size
- h and z are the parameters to be decided.
- Δ^z is the z time difference.

The Whittaker ratio is an extended version and we plug into the mortality ratio of small and reference populations for graduation, instead of mortality rates.

Methodology

- The partial SMR: is one way to deal with estimating mortality rates of small populations, by adding information from other (large) population to correct possible bias

$$v_x = u_x^* \times \exp\left(\frac{d_x \times \hat{h}^2 \times \log(d_x / e_x) + (1 - d_x / \sum d_x) \times \log(\text{SMR})}{d_x \times \hat{h}^2 + (1 - d_x / \sum d_x)} \right)$$

- u_x^* and v_x are observed and graduated mortality rates for age x
- d_x and e_x where and are the observed and expected numbers of deaths at age x for the small population
- $\hat{h}^2$ is the estimate of parameter h^2 for measuring the heterogeneity (in mortality rates) between the small population and large population.

--Lee, W. (2003), "A Partial SMR Approach to Smoothing Age-specific Rates." *Annals of Epidemiology* 13(2), 89-99.

Methodology

- $\hat{h}^2$ was the estimated value of the heterogeneity parameter h^2 via:

$$\hat{h}^2 = \max\left(\frac{\sum\left((d_x - e_x \times SMR)^2 - \sum d_x\right)}{SMR^2 \times \sum e_x^2} \;,\; 0 \right)$$

- Higher value implies more heterogeneity in the rate ratios of the study and the standard populations (or stated differently, more dissimilarity in shape between the two age curves)
- Lower value, less rate-ratio heterogeneity (more similarity between the two curves).
- When it is zero, i.e. no rate-ratio heterogeneity detected in the data

➔We use it as a criterion for choosing a suitable reference population.

Exploratory data analysis (EDA)

- Data:

We utilizes population data from 22 counties and cities spanning from 1997 to 2019, population data of 368 townships and districts in Taiwan, and age-specific mortality data of townships and districts from 2007 to 2018.

➔These data are used to calculate various indicators, such as crude death rate, standardized death rate, and standardized mortality ratio.

Note: World (WHO 2000-2025) Standard - Standard Populations for SDR

Note: We can aggregate the Taiwan historical mortality data for the target population (age-wise), treating the aggregation as the reference population for SMR.

Note: The division of administrative regions in Taiwan is mainly based on governance convenience, including electoral population, to determine the scope of government authority, responsibility attribution, jurisdiction of residents, and fiscal acquisition.

Exploratory data analysis (EDA)

● CDR, SDR, SMR: For Example

➢ Shiding District in NewTaipei City:

From 2016 to 2018, the male CDR, SDR, and SMR were approximately 1.8, 1.2, and 1.2 times higher, respectively, than the urban center Banqiao District in NewTaipei City.

➢ Changbin Township in Taitung County:

The male CDR, SDR, and SMR from 2016 to 2018 approximately 5.5,2.3,1.8 times higher than Luzhu District in Taoyuan City , respectively.

➔ There are significant differences in mortality rates among various townships and districts, with notably higher death rates in the eastern and mountainous areas compared to the western flatlands.

Since the crude death rate (CDR) does not consider the weight of the population size in different age groups, and the standardized death rate (SDR) and Standardized Mortality Ratio (SMR) have similar Heatmap, the next analysis will only use SDR and SMR indexes.

Exploratory data analysis (EDA)

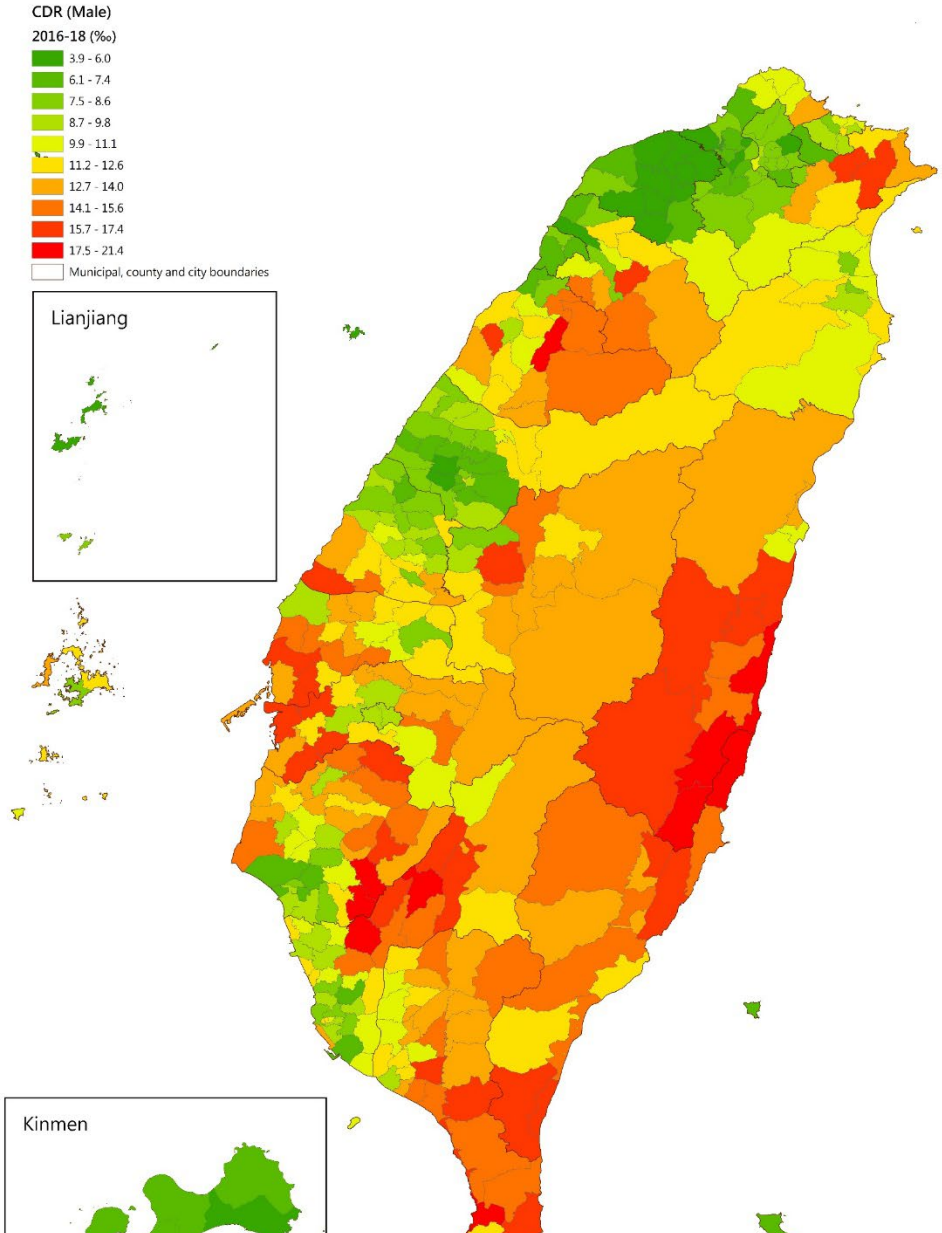
- CDR, SDR, SMR: Heatmap of 368 Districts/ Townships in Taiwan

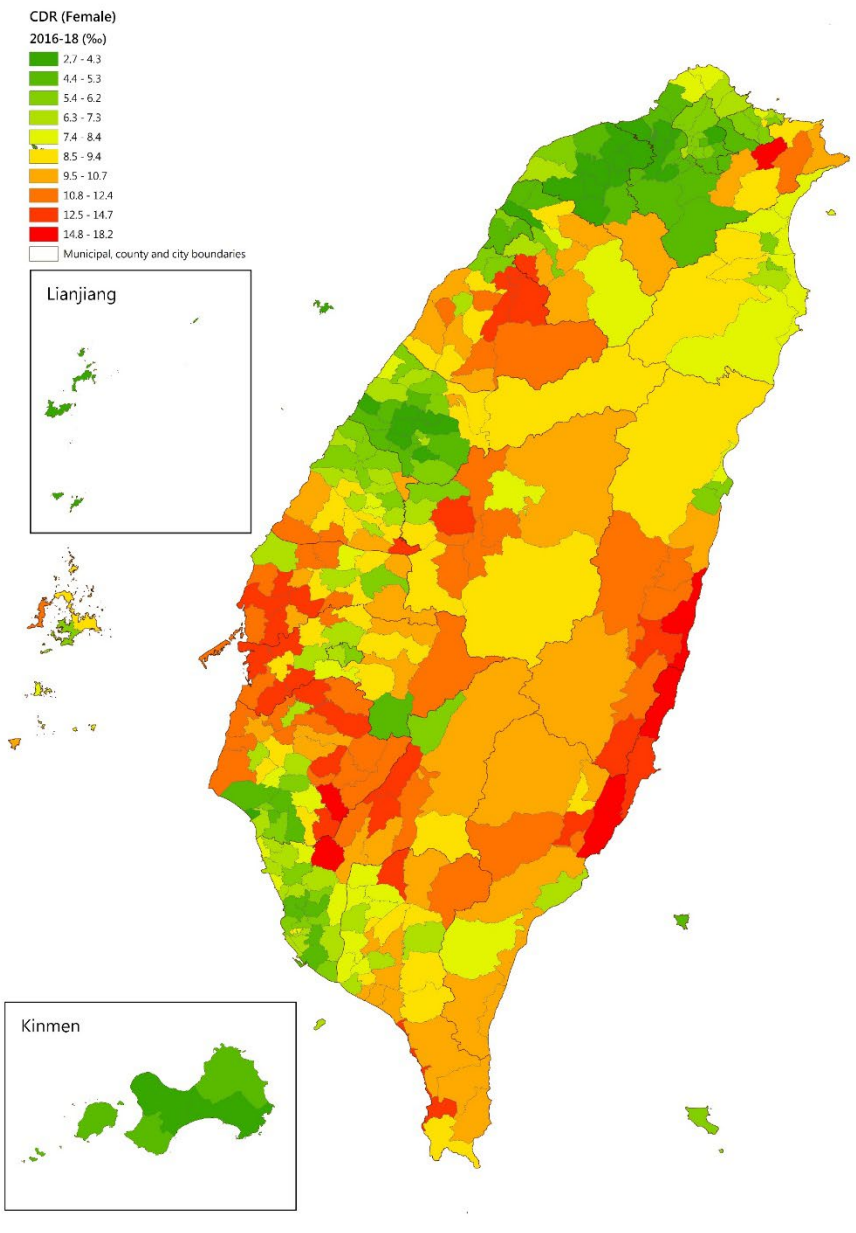
  - Regardless of the CDR, SDR, and SMR maps, it can be observed that the female mortality rate maps is generally lighter, indicating that the mortality rate is lower than that of males.
  - The color distribution presented on the CDR map is not consistent with the SDR and SMR maps. This discrepancy may be attributed to the fact that the CDR does not consider the population structure weighting.
  - The death rate is higher in the eastern and mountainous areas, as well as in the southwest.

(Next three slides)

Exploratory data analysis (EDA)

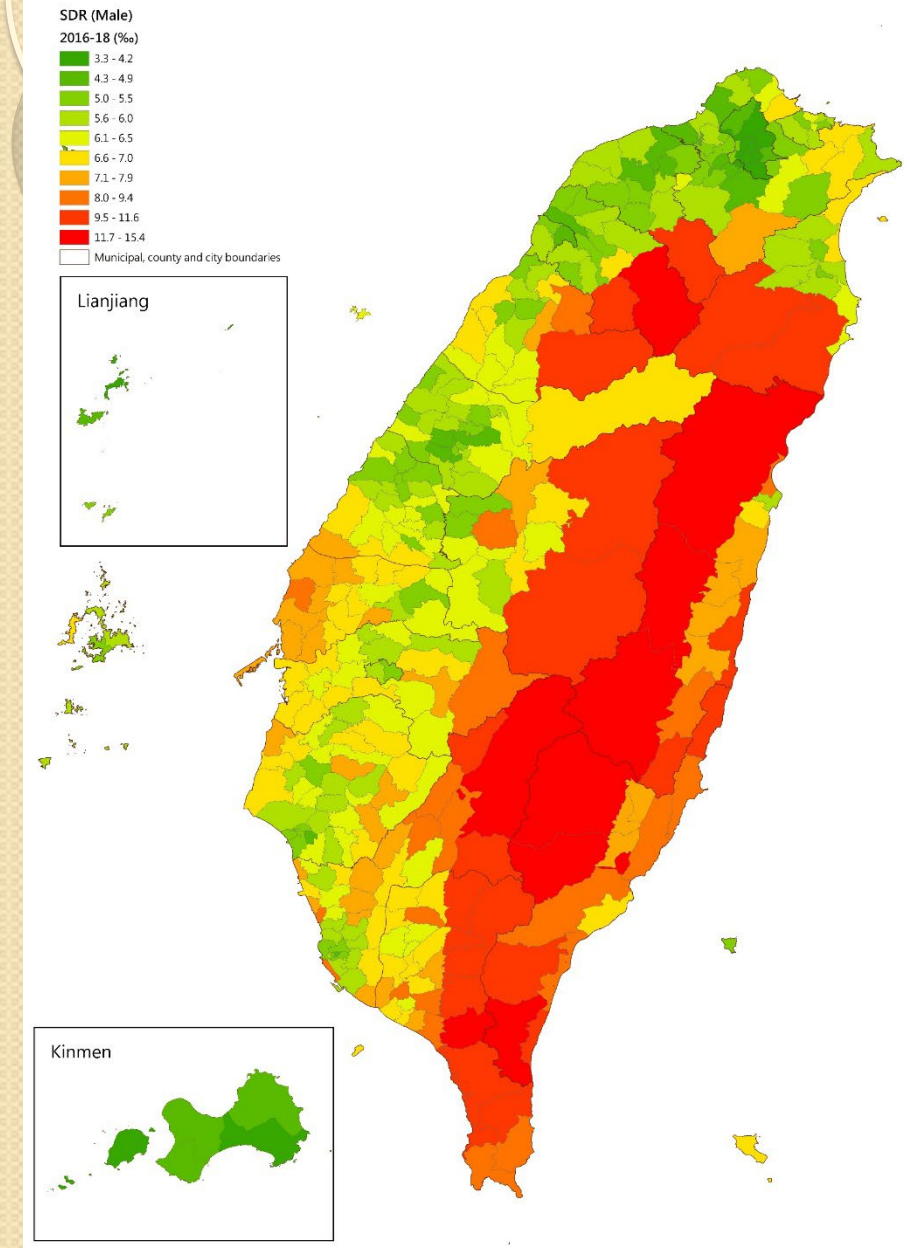
- CDR: Heatmap of 368 Districts/ Townships in Taiwan

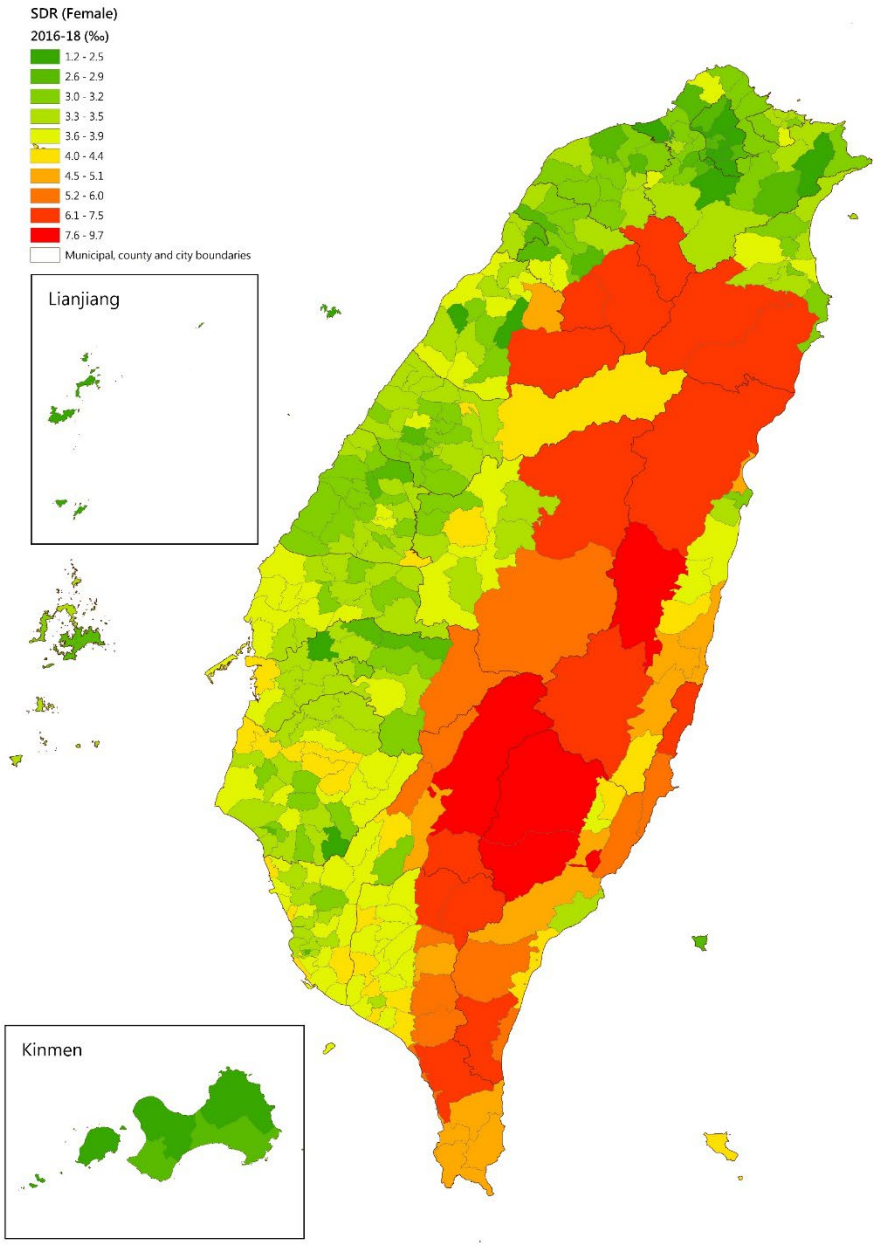




Exploratory data analysis (EDA)

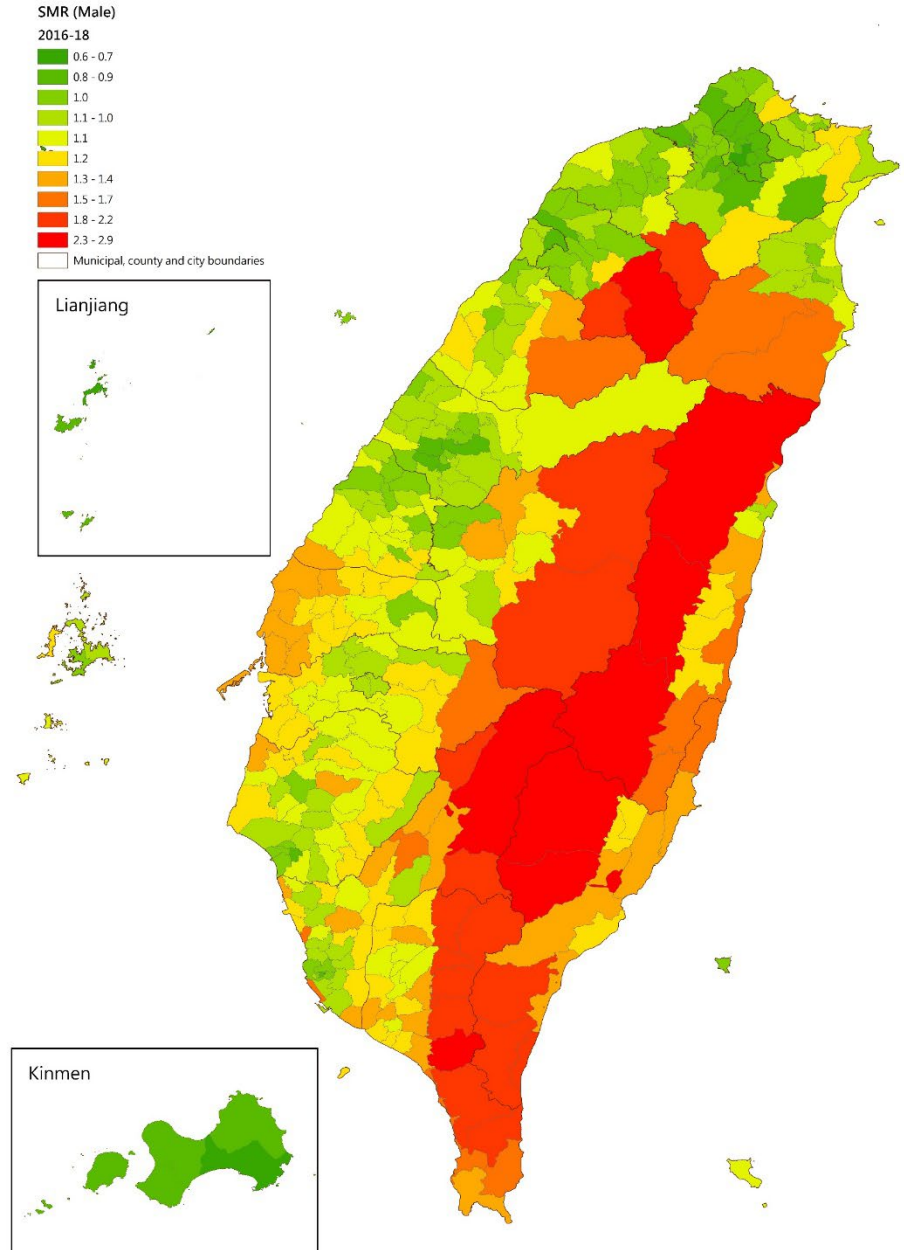
- SDR: Heatmap of 368 Districts/ Townships in Taiwan

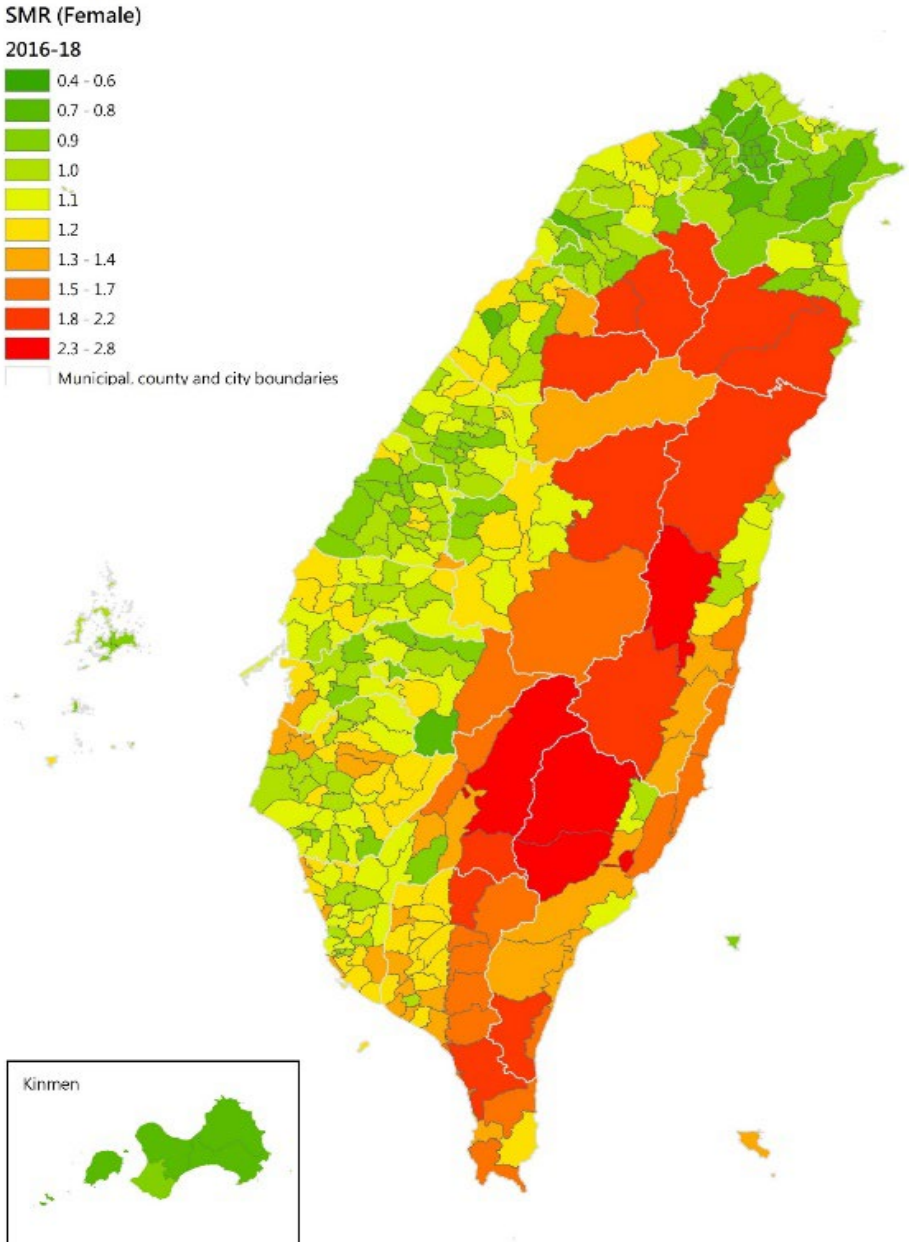




Exploratory data analysis (EDA)

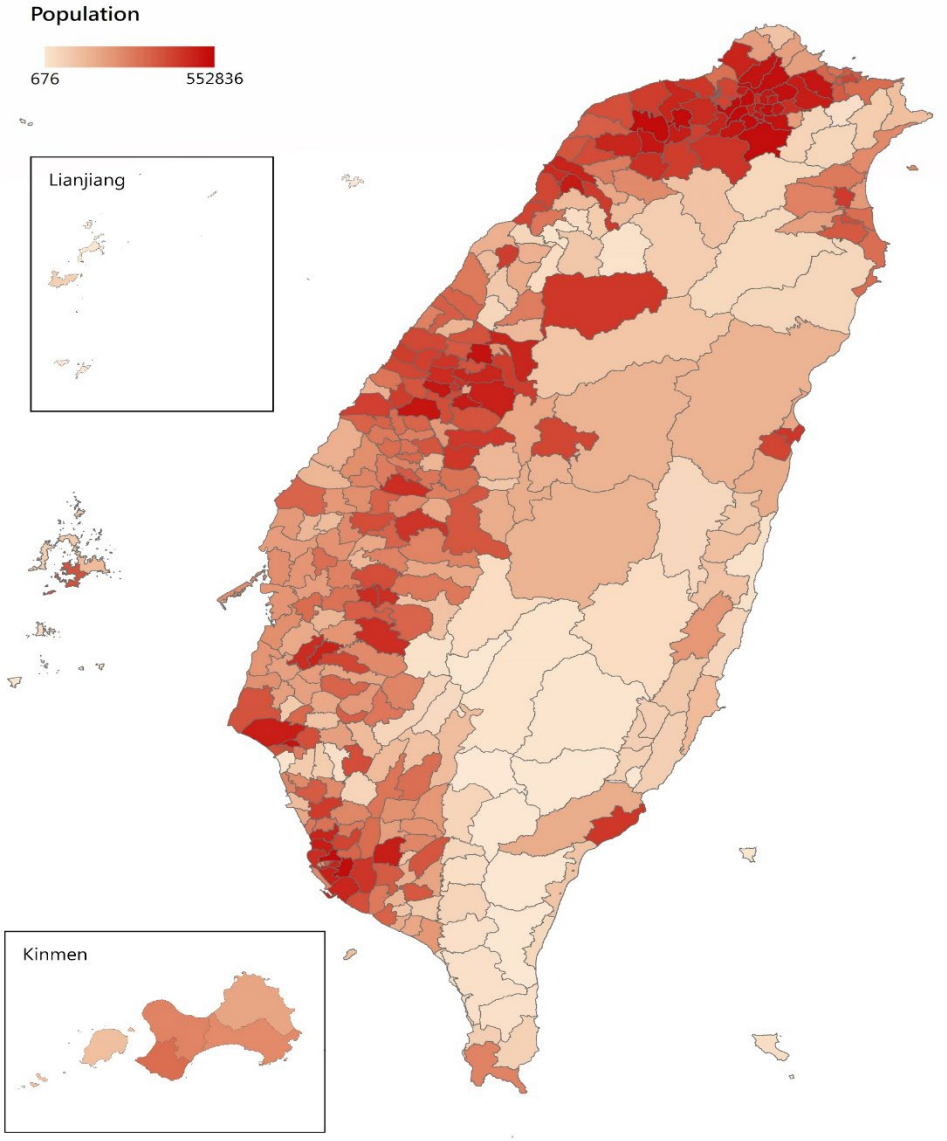
- SMR: Heatmap of 368 Districts/ Townships in Taiwan





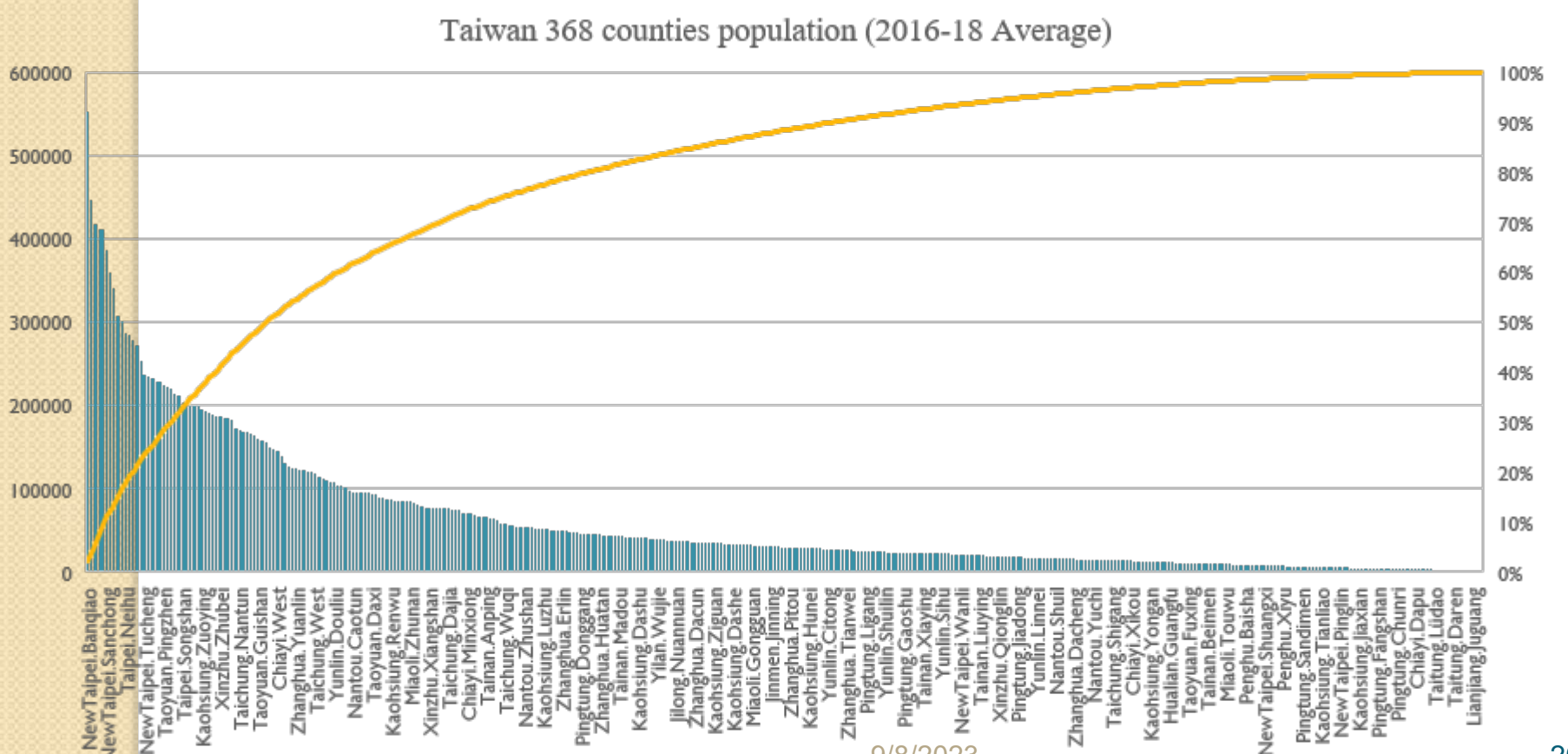
Exploratory data analysis (EDA)

- The population of each township in Taiwan varies greatly, with the population of both sexes ranging from a minimum of 676 to 554,373.



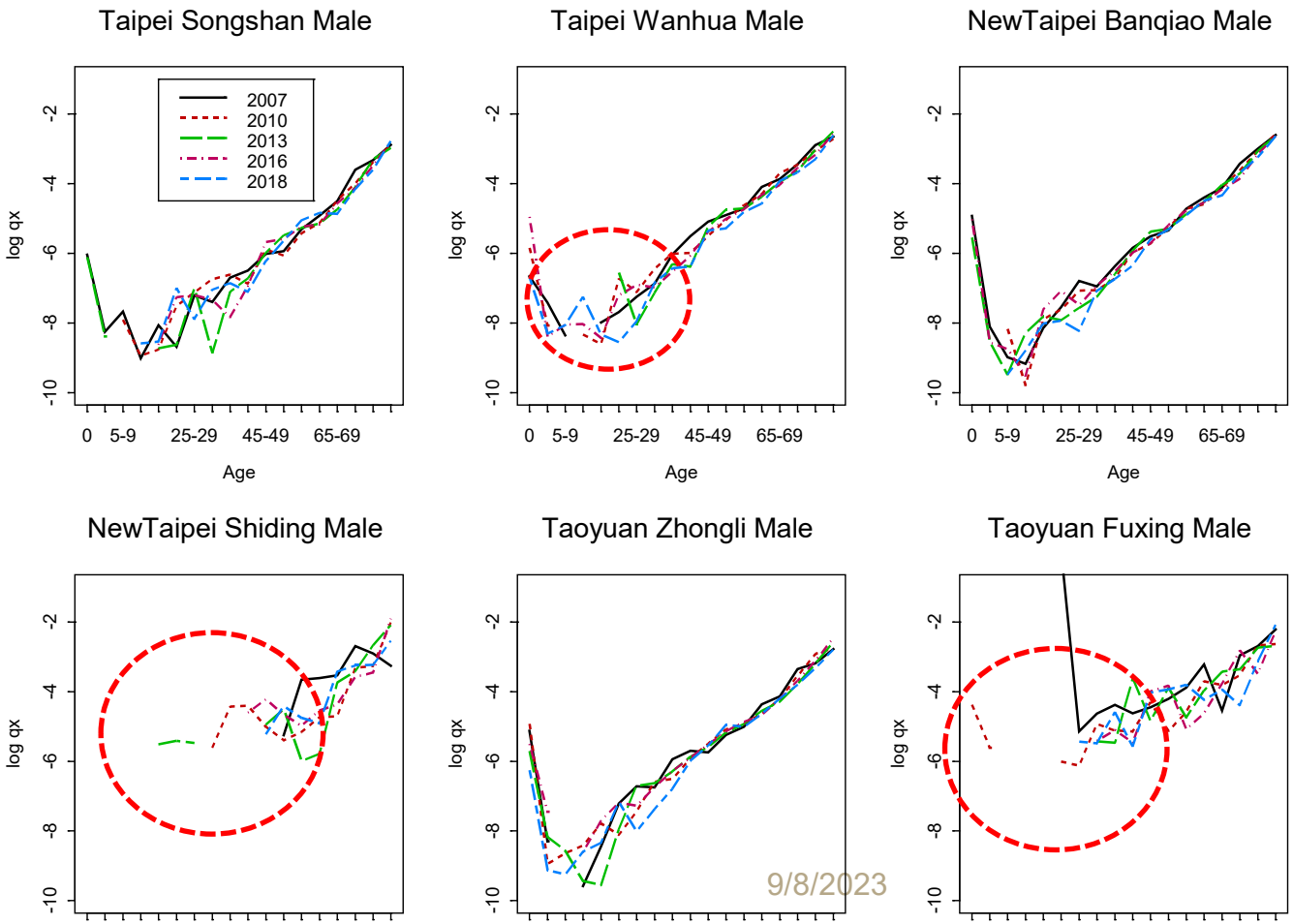
Exploratory data analysis (EDA)

- No. of Taiwan 368 Districts/ Townships Population
  - There are 177 Townships and Districts with populations between 10,000 and 50,000 and 64 Townships and Districts populations less than 10,000.



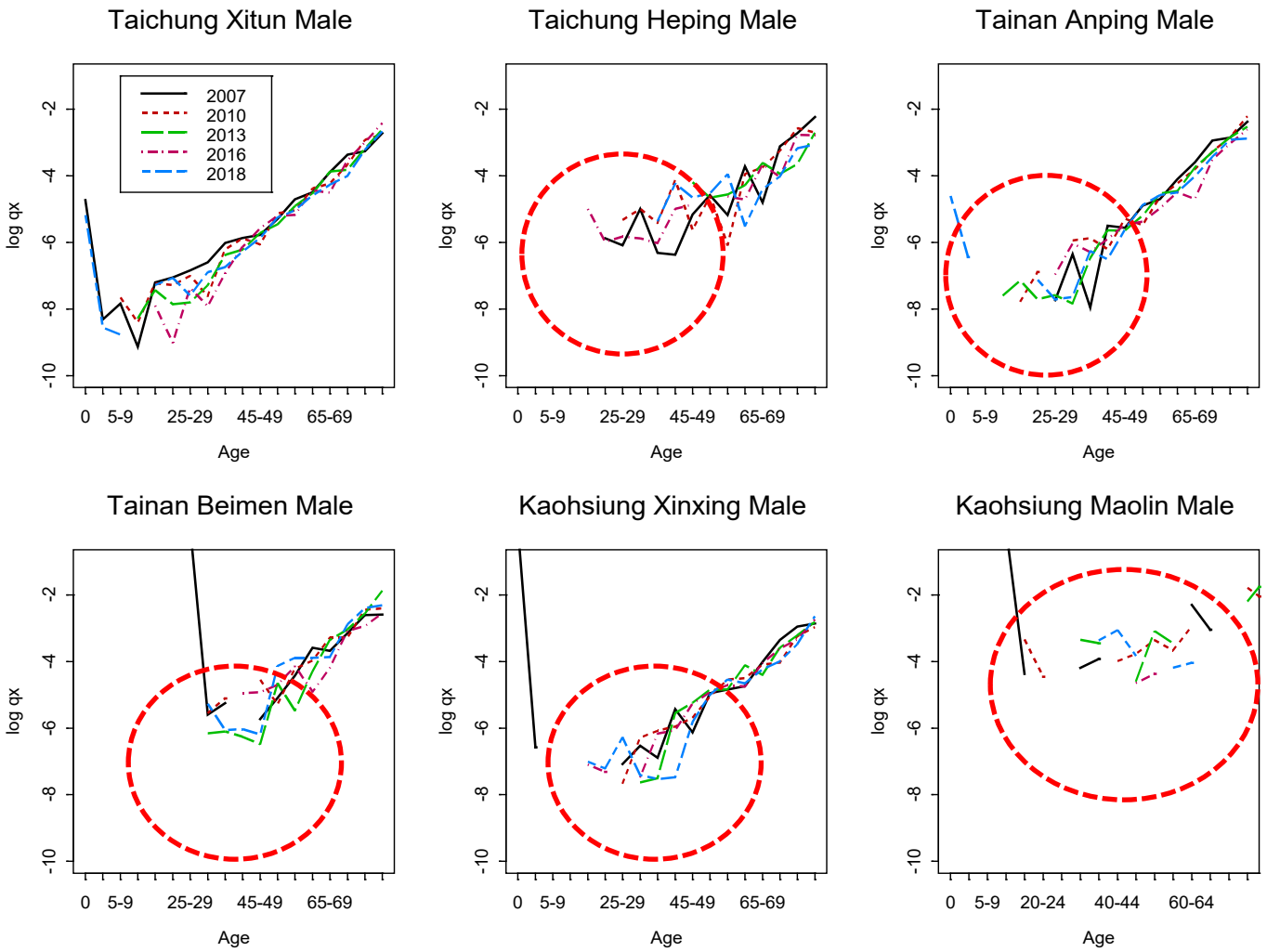
Exploratory data analysis (EDA)

- Ex: Male's Experience Mortality Rates of District in Six Metropolitan Area.
  - Some age-specific mortality rates are unstable and have missing vales in some age groups in Taiwan.



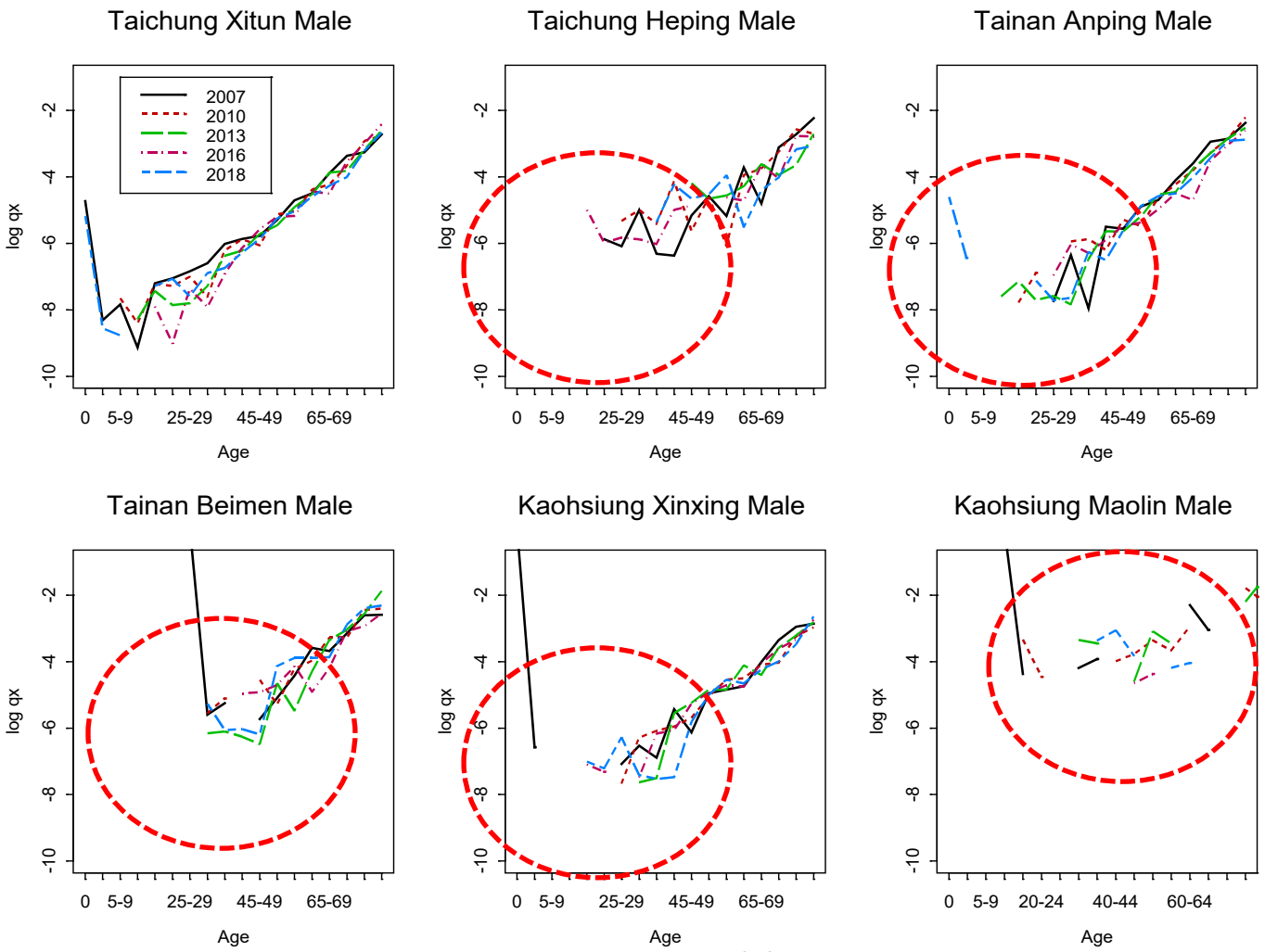
Exploratory data analysis (EDA)

- Ex: Male's Experience Mortality Rates of District in Six Metropolitan Area.



Exploratory data analysis (EDA)

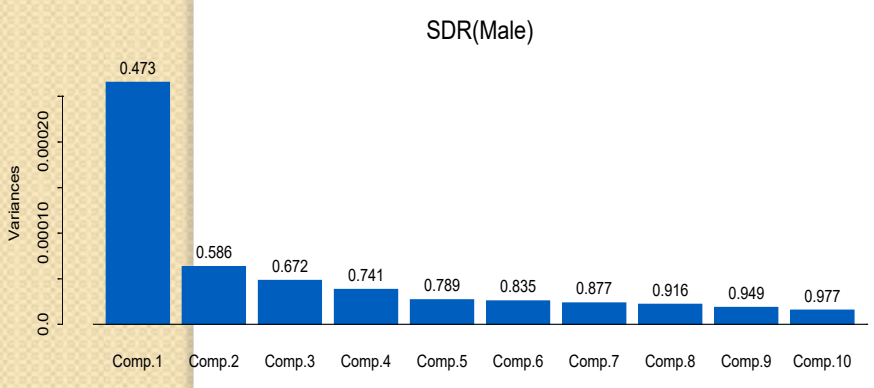
- Ex: Male's Experience Mortality Rates of District in Six Metropolitan Area.

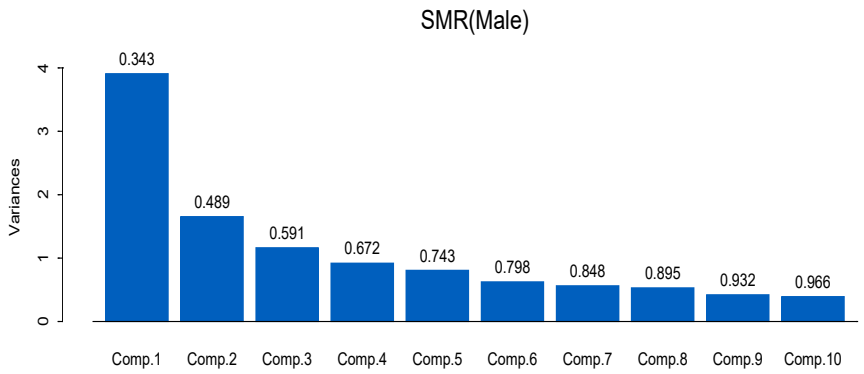


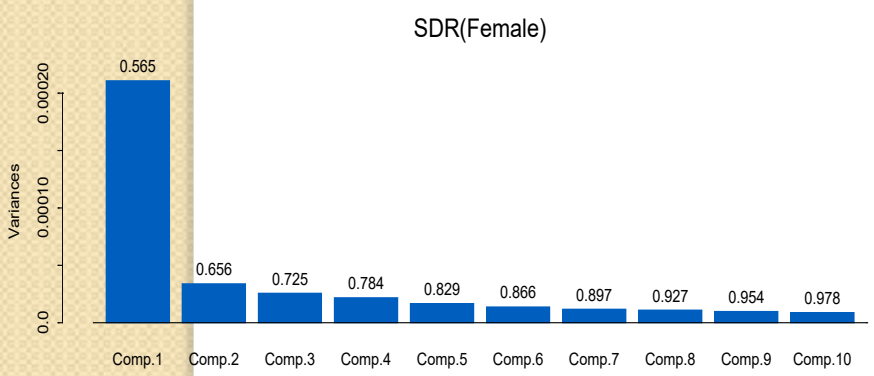
Classification

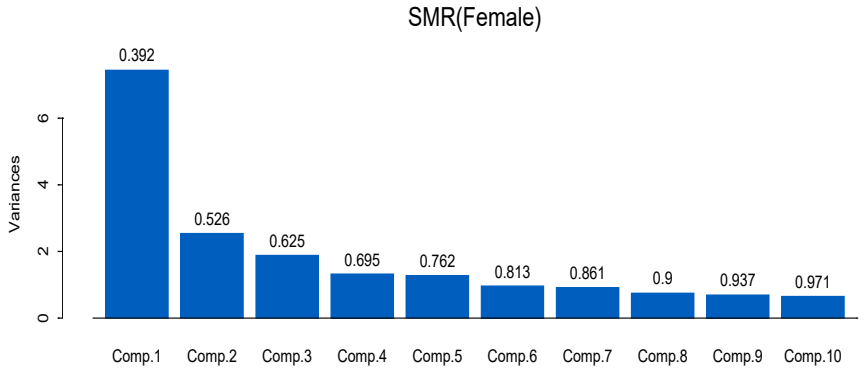
- Principal Component Analysis(PCA)

➢ Ex (Male SDR): The cumulative proportion of explained variance corresponding to the three, five, and seven principal components is 67%, 79%, and 88%, respectively.









Classification

- K-means cluster : We classified the data into three groups, as shown in the example, to compare the difference between the SDR and SMR indexes.

➔ We found that both SDR and SMR can generate similar clustering results. Ex: For males, 27 townships/districts are classified in the same Group 1.

3-Groups		SMR	SDR	Same Classification
Male	Group 1	27	31	27
	Group 2	191	190	183
	Group 3	150	147	139
Female	Group 1	312	333	312
	Group 2	44	34	23
	Group 3	12	1	1

Conveniently, in the next analysis, we will only use SMR as the mortality index to present our research results.

Classification

- K-means clustering: Classified to 3,5,7 Groups.

➔Ex: 3 Groups (Male):

Group1: Kaohsiung Jiaxian, Kaohsiung Maolin, KaohsiungTaoyuan, Kaohsiung Namaxia,Yilan Datong, …etc 27 counties

Group 2: Taipei Songshan, Taipei Xinyi, Taipei Daan, Taipei Zhongshan, Taipei Zhongzheng, …etc 191 counties

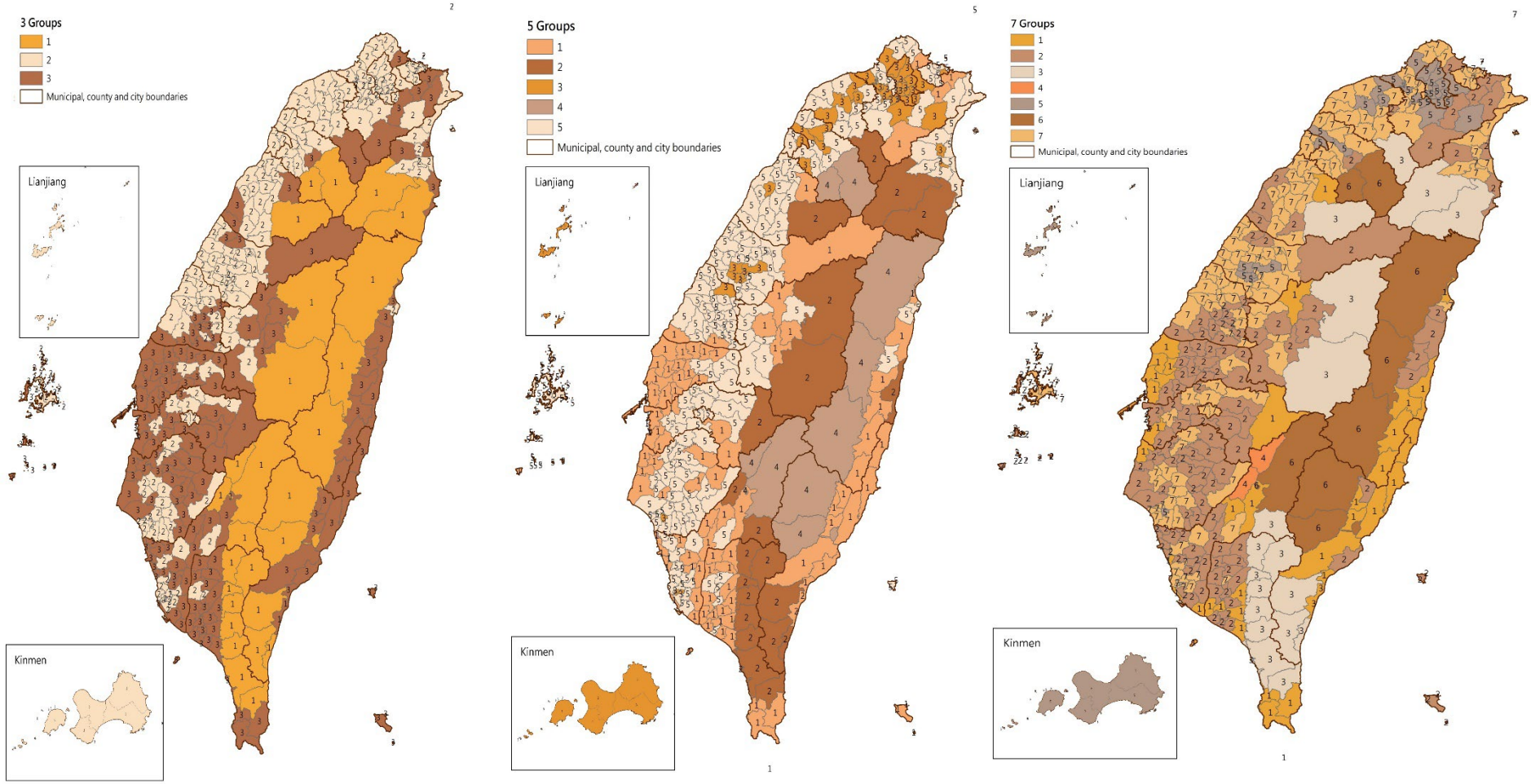
Group 3: NewTaipei Ruifang, NewTaipei Shiding, NewTaipei Pingxi, NewTaipei Shuangxi, NewTaipei Wanli, …etc 150 counties.

	3-Groups		5-Groups		7-Groups	
	Male	Female	Male	Female	Male	Female
Group 1	27	312	92	35	36	21
Group 2	191	44	22	11	121	1
Group 3	150	12	56	120	19	43
Group 4			9	2	2	11
Group 5			189	200	43	17
Group 6					8	130
Group7					139	145

Classification

- K-means clustering: Classified to 3,5,7 Groups (Male's SMR)

From the map, it can be observed that even the townships in the same county are divided into different groups.



Classification

- Based on the clustering results, we can determine the cluster to which each of the 368 townships belongs among clusters 3, 5, and 7.
- We show 28 Districts/ Townships as following tables

➢ Ex: The male population of Taipei Songshan District is about 96,327, and it belongs to Group 2 in the 3-Groups clustering, Group 3 in the 5-Groups clustering, and Group 5 in the 7-Groups clustering.

	Male Population(2016-18)	3-Groups	5-Groups	7-Groups
Taipei Songshan District	96,327	2	3	5
Taipei Wanhua District	92,944	2	5	7
NewTaipei Banqiao District	270,217	2	3	5
NewTaipei Shiding District	4,204	3	5	2
Taoyuan Zhongli District	202,412	2	3	5
Taoyuan Fuxing District	6,443	3	2	3
Taichung Xitun District	109,679	2	3	5
Taichung Heping District	5,882	3	1	2
Tainan Anping District	31,729	2	5	7
Tainan Beimen District	5,568	3	1	1
Kaohsiung Xinxing District	24,508	2	3	5
Kaohsiung Maolin District	981	1	2	3

Classification

- Ex: The male population of Yilan Datong Township is about 3,309, and it belongs to Group 1 in the 3-Groups clustering, Group 2 in the 5-Groups clustering, and Group 3 in the 7-Groups clustering.

	Male Population(2016-18)	3-Groups	5-Groups	7-Groups
Yilan Datong Township	3,309	1	2	3
Xinzhu Wufeng Township	2,501	1	4	6
Miaoli Shitan Township	2,419	2	5	7
Zhanghua Tianwei Township	14,050	3	5	2
Nantou Guoxing Township	10,026	3	1	1
Yunlin Taixi Township	12,690	3	1	1
Chiayi Alishan Township	2,935	3	2	1
Pingtung Majia Township	3,429	1	2	3
Taitung Daren Township	1,918	1	2	3
Hualian Fengbin Township	2,520	3	2	1
Penghu Qimei Township	1,983	3	1	2
Jilong Nuannuan District	19,340	2	5	7
Xinzhu Xiangshan District	39,758	2	5	7
Chiayi West District	72,026	2	5	7
Lianjiang Dongyin Township	806	2	3	5
Jinmen Wuqiu Township	347	2	3	5

Construct Districts/ Townships Life tables

- Each district/township will have four reference populations: the administrative district it belongs to (county or city), the 3-Groups clustering, the 5-Groups clustering, and the 7-Groups clustering.

➔We use Lee (2003) heterogeneity index as criteria for choosing a suitable reference population:

Ex1: The heterogeneity value of Taipei Wanhua District in the 5-Group clustering is the smallest. Therefore, the reference population will be the 5-Group clustering.

Ex2: The heterogeneity value of NewTaipei Shiding District in the 3-Group clustering is the smallest. Therefore, the reference population will be the 3-Group clustering

Construct Districts/ Townships Life tables

	Heterogeneity(2007-18 Average)				Reference Population	
	County/ City	3 Groups	5 Groups	7Groups		
Taipei Songshan District	0.007299	0.017499	0.011605	0.010606	Taipei City/ 7 groups	Group 5
Taipei Wanhua District	0.015735	0.008619	0.007423	0.007975	5 groups	Group 5
NewTaipei Banqiao District	0.081269	0.004348	0.003638	0.003886	5 Groups	Group 3
NewTaipei Shiding District	0.044059	0.027977	0.032203	0.028552	3 groups	Group 3
Taoyuan Zhongli District	0.005510	0.006156	0.007841	0.008433	3 groups/Taoyuan City	Group 2
Taoyuan Fuxing District	0.107205	0.066984	0.025733	0.026174	5 groups	Group 2
Taichung Xitun District	0.007209	0.007465	0.005552	0.005419	7 groups	Group 5
Taichung Heping District	0.061363	0.040106	0.036879	0.043127	5 groups	Group 1
Tainan Anping District	0.016075	0.016189	0.018939	0.017658	Tainan City/3 groups	Group 2
Tainan Beimen District	0.000000	0.002364	0.003749	0.005977	3 groups	Group 3
Kaohsiung Xinxing District	0.004085	0.006093	0.008930	0.009846	3 groups	Group 2
Kaohsiung Maolin District	0.000570	0.000000	0.000000	0.000000	3 groups	Group 1

Construct Districts/ Townships Life tables

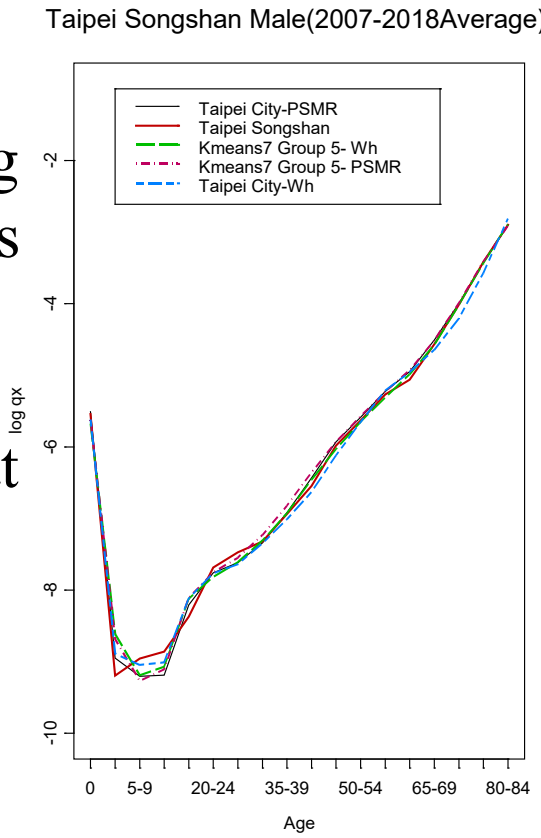
	Heterogeneity(2007-18 Average)				Reference Population	
	County/ City	3 Groups	5 Groups	7Groups		
Yilan Datong Township	0.090929	0.015330	0.018852	0.019928	3 groups	Group 1
Xinzhu Wufeng Township	0.096000	0.007396	0.004646	0.005481	5 groups	Group 4
Miaoli Shitan Township	0.046143	0.065520	0.057525	0.061702	Miaoli County/5 groups	Group 5
Zhanghua Tianwei Township	0.015123	0.012401	0.015017	0.012211	7 groups	Group 2
Nantou Guoxing Township	0.014597	0.015856	0.015347	0.014952	7 groups	Group 1
Yunlin Taixi Township	0.006268	0.006565	0.006227	0.006508	5 groups	Group 1
Chiayi Alishan Township	0.053674	0.047194	0.000219	0.023119	5 groups	Group 2
Pingtung Majia Township	0.053479	0.017979	0.019768	0.020388	3 groups	Group 1
Taitung Daren Township	0.036948	0.037294	0.036095	0.036113	5 groups	Group 2
Hualian Fengbin Township	0.005150	0.010353	0.006497	0.005850	Hualian County/7 groups	Group 1
Penghu Qimei Township	0.002237	0.000000	0.000000	0.000476	Penghu County/ 7 groups	Group 2
Jilong Nuannuan District	0.019683	0.019274	0.017193	0.018278	5 groups	Group 5
Xinzhu Xiangshan District	0.015354	0.015204	0.014878	0.014963	5 groups	Group 5
Chiayi West District	0.004158	0.006789	0.005657	0.006142	Chiayi City/5 groups	Group 5
Lianjiang Dongyin Township	0.005906	0.000000	0.005474	0.005790	5 groups	Group 3
Jinmen Wuqiu Township	0.003482	0.000000	0.000000	0.000000	5 groups	Group 3

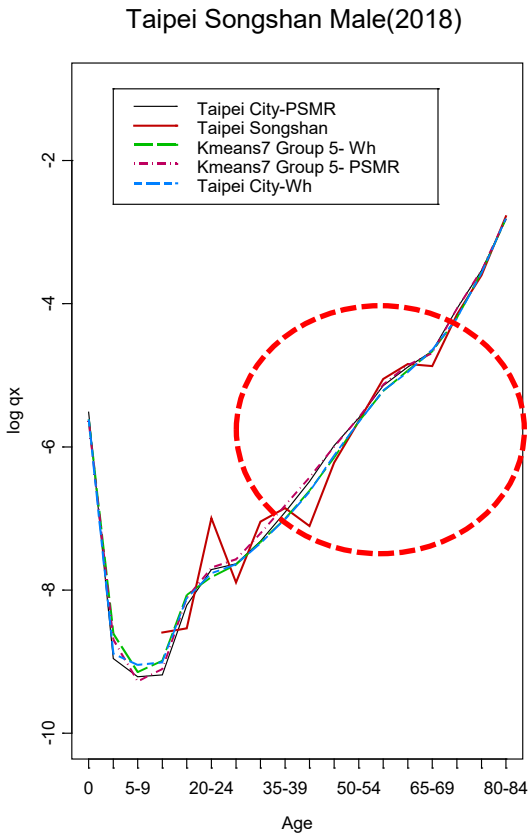
Construct Districts/ Townships Life tables

- Different townships and districts within the same county or city may be situated in diverse terrains.

Ex: Songshan District:

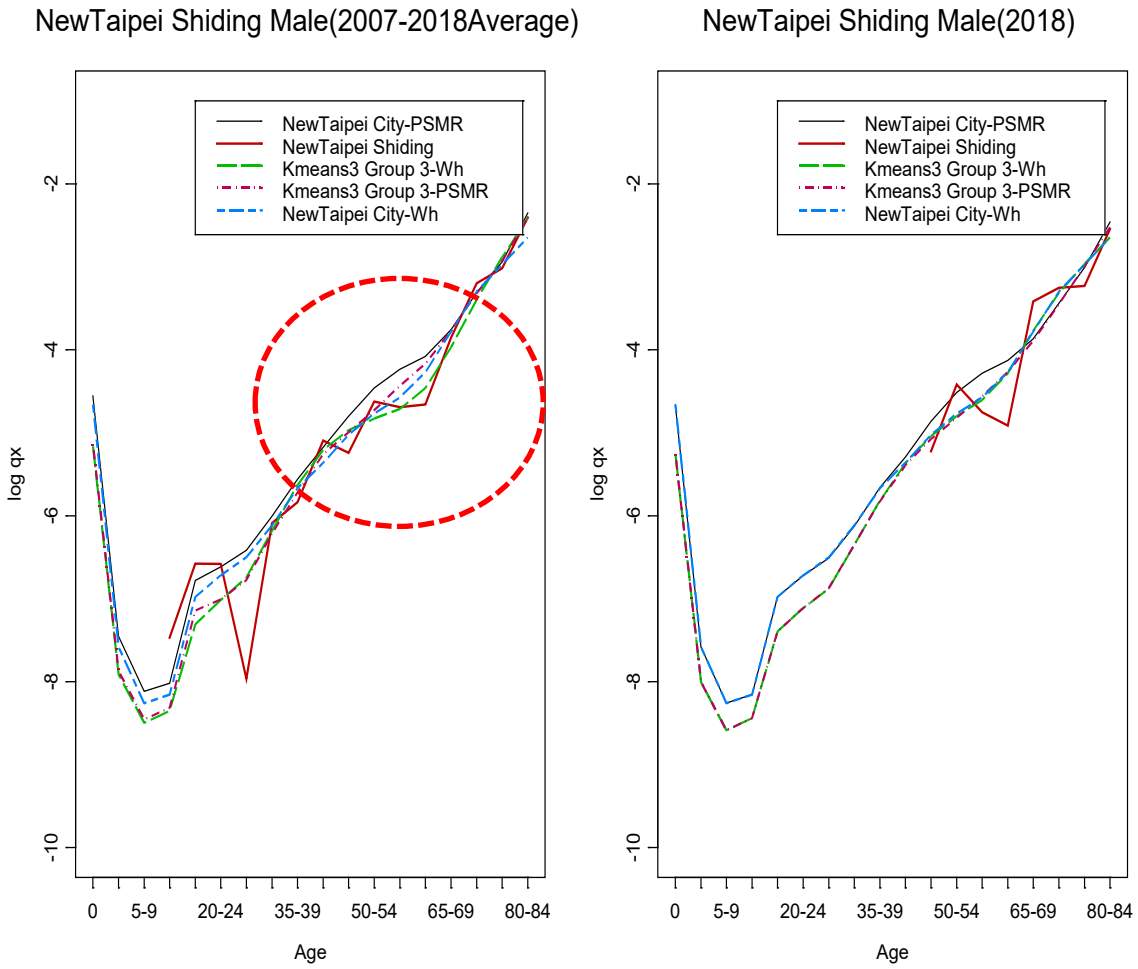
- The accuracy of estimating age-specific mortality rates using Taipei City as the reference population is approximately close to that obtained using 7-Groups clustering.
- Whittaker ratio method or PSMR are acceptable





Construct Districts/ Townships Life tables

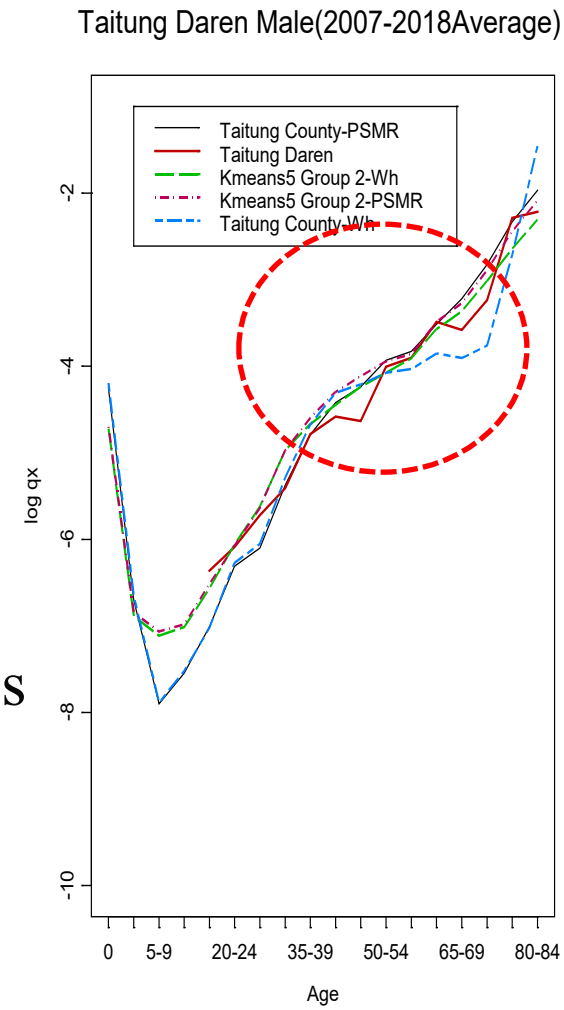
- Ex: Shiding District:
  - Estimating age-specific mortality rates using NewTaipei City as the reference population is less accurate than using 3-Groups clustering.
  - Whittaker ratio method is better than PSMR

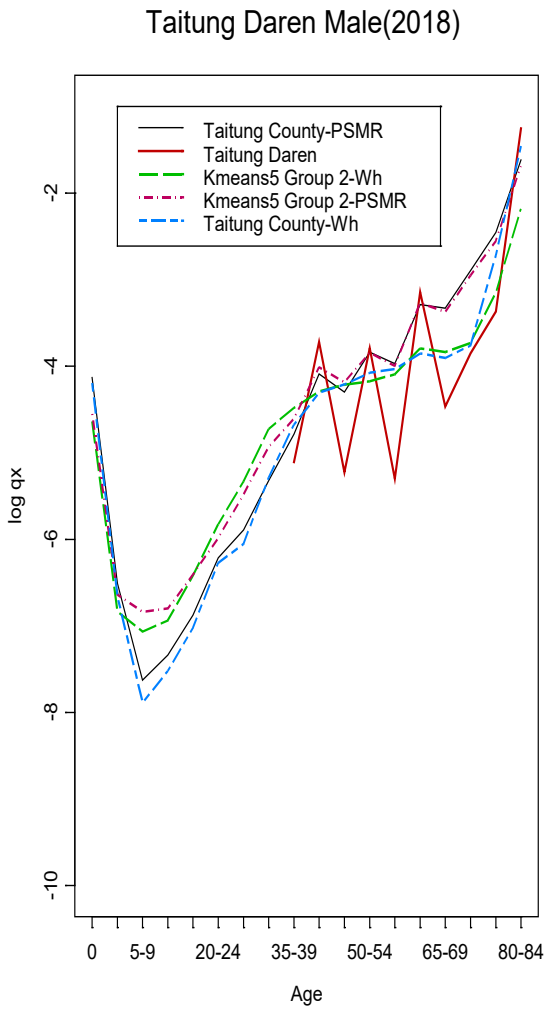


Construct Districts/ Townships Life tables

Ex: Daren Township:

- Estimating age-specific mortality rates using Taitung County as the reference population is less accurate than using 5-Groups clustering.
- Whittaker ratio method is better than PSMR



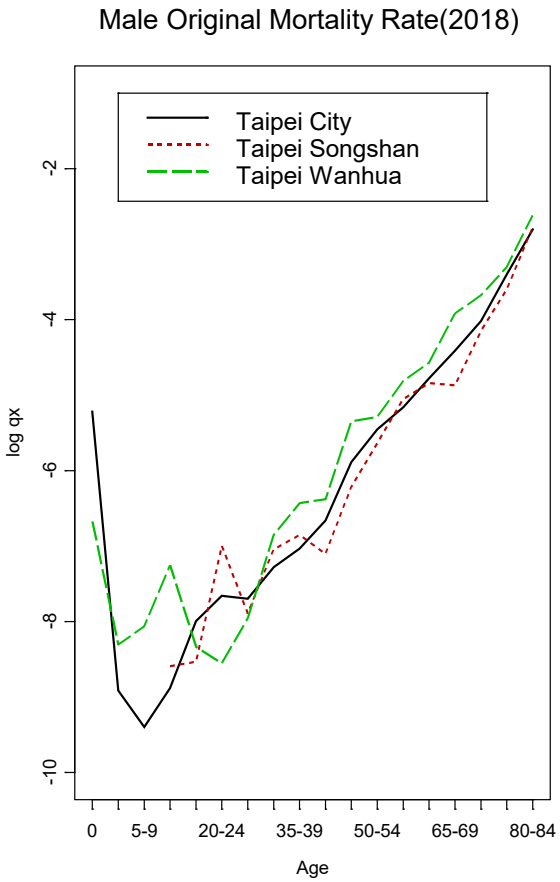


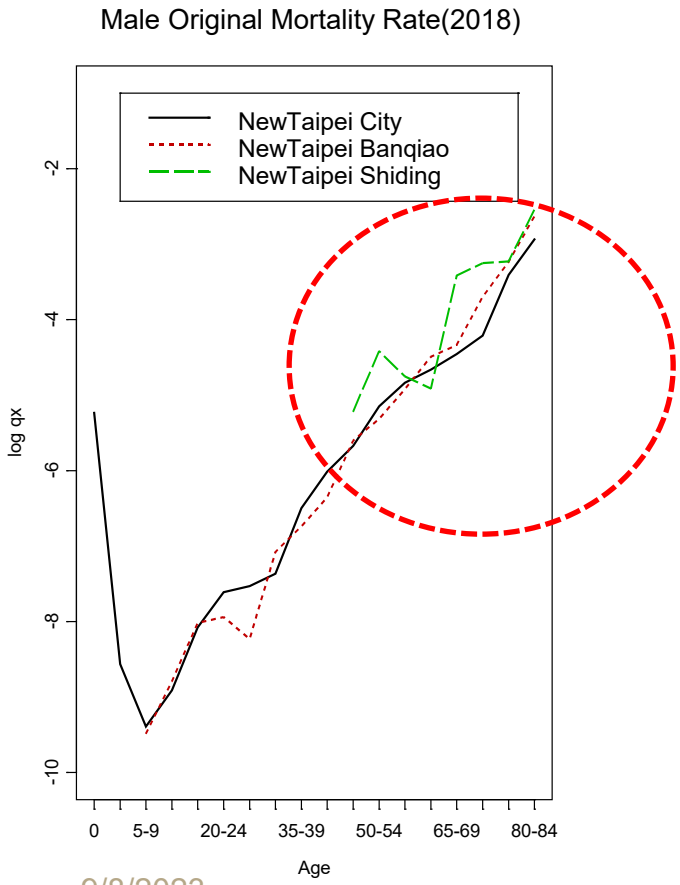
Construct Districts/ Townships Life tables

- Comparison the mortality rates

➔ Roughly, the 2018 original mortality rate of Wanhua District is higher than that of Taipei City and Songshan District.

➔Shiding District has missing data for individuals under the age of 50. However, it has a higher mortality rate than NewTaipei City and NewTaipei Banqiao after the age of 50.



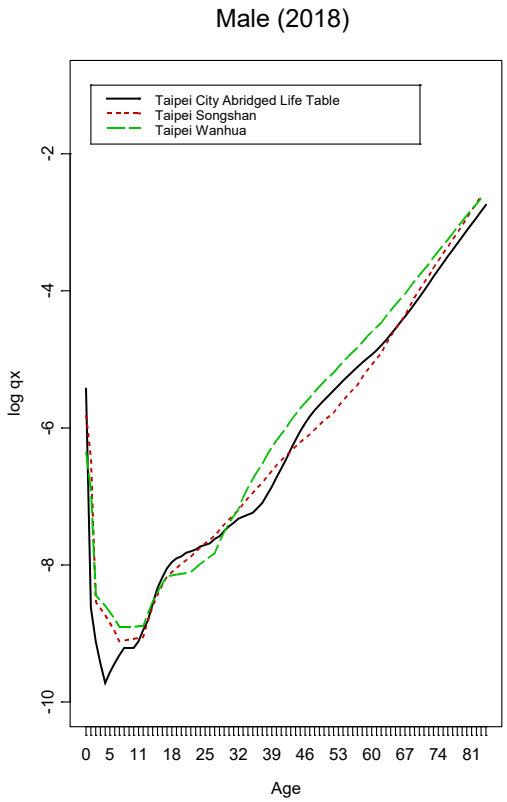


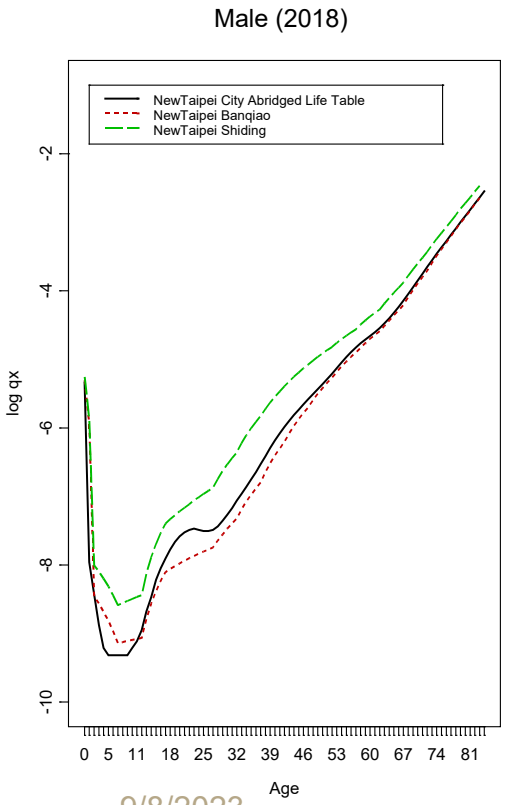
Construct Districts/ Townships Life tables

- To construct a single-age life table, we calculate the single-age death rate using both interpolation and Gompertz extrapolation methods after applying Whittaker ratio or PSMR graduation.

➔ The mortality rate of Wanhua District is higher than that of both 2018Taipei City Abridged Life Table and Songshan District.

➔Shiding District has the highest mortality rate compared to the 2018 New Taipei City Abridged Life Table and Banqiao District.





Constructing Districts/ Townships Life Tables

- Comparison the Life expectancy at Birth of 2018 abridged life table with District/Township

Ex: The life expectancy at birth of Wanhua District is (78.34 Year) less than Taipei City Abridged Life Table (80.93 Year) and Songshan District (80.2 Year)

	Life expectancy at Birth
Taipei City 2018 abridged life table	80.93
Taipei Songshan District	80.20
Taipei Wanhua District	78.34
NewTaipei City 2018 abridged life table	78.33
NewTaipei Banqiao District	78.72
NewTaipei Shiding District	74.30
Taoyuan City 2018 abridged life table	77.86
Taoyuan Zhongli District	79.43
Taoyuan Fuxing District	68.36
Taichung City 2018 abridged life table	77.61
Taichung Xitun District	79.00
Taichung Heping District	75.61
Tainan City 2018 abridged life table	76.79
Tainan Anping District	76.61
Tainan Beimen District	70.12
Kaohsiung City 2018 abridged life table	76.13
Kaohsiung Xinxing District	79.69
Kaohsiung Maolin District	60.27

Conclusion

- The idea of increasing sample size can be used in small area estimations.
- CDR is not suitable for comparing two populations with different age distributions. Instead, we propose using SMR or SDR as the indexes, and PCA + K-means is an appropriate clustering method for our analysis.
- Lee (2003) heterogeneity index could be used to determine whether the reference populations is suitable for small populations.
- If there are too many age groups with a death count of 0, the estimation error will be larger. Ex: In Jinmen Wuqiu Township, there are only 347 male residents, and on average(2007-18), there are only 1.6 deaths per year.
- Whitter ratio method is suitable for small population than PSMR
- In the future, we will compare other cluster methods and mortality indexes to verify the biological reasonableness of age-specific rates and the similarity of the long-term trends and curves between the reference population and the small population.

Thank you for your attention.