Mortality Analysis in the U.S. and Related Factors

by

M.S., Wenxiao Zhou

University of Connecticut, 2021

Final Project

STAT 5225—Data Management and Programming in R&SAS

Advisor: Haim Bar

Spring 2021

Contents

Abstract

Social progress is accompanied by the development of drugs and technological progress. People use better methods to control diseases and reduce mortality. Thanks to learning the time trends of mortality statistics with scientific models, we can get information on human life and health problems and help improve the public's health by using systematic methods to monitor the whole community.

In this study, descriptive analysis and factor analysis are done basing on the Mortality Multiple Cause Data collected from the CDC. Data of Mortality Multiple Causes in the US within the year 2015-2019 aims to make a review and summary of the mortality rate variation trend, personal characteristics of the death cases. The main factors of causing mortality within different groups are also being discussed.

The overall analysis helps to detect a smooth decreasing trend of age-adjusted mortality rates among years. Race and sex categories affect the tendencies of mortality rates, i.e. Males and the Black group of people hold high mortality rates among years. Also, education level is another important factor influencing the mortality rate: people under higher education have a lower proportion of mortality rates.

Analysis in the year 2019 is focusing on two age groups: adults and elders with different top causes of mortality rates. A state-level comparison of mortality rates is included, and we find a significant association between education levels and heart mortality rate, i.e., states with higher education ratings have relatively lower heart mortality rates in general.

Keywords: Mortality Rate, Age-Adjusted, Descriptive Analysis, One-way ANOVA

1 Introduction

With the progress of medical treatment and science and technology, people can control diseases, avoid risks and accidents, and reduce mortality through reasonable and efficient ways. These advances have been made thanks to the efficient use of mortality statistics for scientific data analysis. Mortality statistics contain information on individuals' life and health problems that can help improve the public's health by using systematic methods to monitor the whole community. Although the abrupt increase in COVID-19 mortality in 2020 affects the overall trend of mortality, a long-term analysis of mortality trends over the past five years is of great significance for long-term prevention and monitoring.

Based on the Mortality Multiple Cause Data collected from the CDC (Centers of Disease Control and Prevention) [1], mortality cases are provided information during different years, with more than 20 correlated factors. The factors including general information of mortality status, causes of deaths, the basic background of death cases.
In this study, a work basing on the Data of Mortality Multiple Causes in the US within the year 2015-2019 aims to make a review and summary of the mortality rate variation trend, personal characteristics of the death cases, the main factors of causing mortality within different groups.

What's more, a specific analysis of the year 2019 is focusing on the main causes of mortality in both adults and elders. All the work is mainly the combination of descriptive analysis by EDA and frequency tables. Further data visualization basing on the geographical dataset will be included.

2 Description of Data

The Mortality Multiple Cause Data can be accessed from website of CDC (Centers of Disease Control and Prevention) [1]. Here, we choose mortality cases from year 2015-2019, which contains 13995924 cases in total. In order to make the analysis accurate and targeted, we screened only 10 related factors for analysis. See the details of data below:
Table 1-3 are the basic introduction of data using in this report.

Table 1: Mortality Cases Summary in 2015-2019

Year	Mortality Counts
2019	2861523
2018	2846305
2017	2820034
2016	2749864
2015	2718198

Table 2: Factors Explanations [2]

Factors	Type	Details
Resident Status	Categorical (4 levels)	Resident status for United States, Puerto Rico, and Virgin Islands Occurrence all includes residents, intrastate nonresidents, interstate nonresidents, foreign residents as four categories
Education (2003 Revision)	Categorical (9 levels)	Education levels include 8th grade or less, high school degrees, college degrees, graduate degrees
Month of Death	Categorical (12 levels)	Month includes January to December
Sex	Categorical (2 levels)	F(Female), M(Male)
Age	Categorical (52 levels)	Age levels contain 52 levels from infants to olders
Place of Status	Categorical (9 levels)	Place of death and decedents' status include hospital, home, hospice facility, nursing home, other and unknown. Mortality status includes inpatient or outpatient for hospital.
Martial of Status	Categorical (5 levels)	S(Single), M(Married), W(Widowed), D(Divorced), U(Unknown)

Manner of Death	Categorical (8 levels)	Manner includes Accident, Suicide, Homicide, Pending Investigation, No determine, Self-inflicted, Natural Deaths, Not Specified
39 Cause Recode	Categorical (42 levels)	CDC provides the details of categories of cause [2]
Race	Categorical (3 levels)	There are three basic levels of race: White, Black, Others

Annotation: Education levels have been redefined on the overall analysis and year 2019 analysis, see more details in analysis sections.

Age levels have redefined to separate adults (20-59 years old) and elders (60+ years old) into two groups in the below study.

Table 3: Auxiliary Datasets

Datasets	Variables	Details
age_mortality.csv	Year, Race, Sex, Average Life, Mortality	Use for overall and sex group age-adjusted mortality rate analysis from year 2015-2019
state2019.xlsx	State, Month (Jan.-Dec.), Total, Population	Use for mortality analysis under state levels in year 2019
heart2019.csv	YEAR, CODE, RATE, DEATHS, STATE	Use for detecting heart disease mortality by states in year 2019
educate2019.xlsx	Overall Rank, State, Total Score, Educational, Quality	Use for detecting relationship between states education levels and mortality rates

3 Overall Analysis

3.1 Age-adjusted Mortality Rate Analysis

First of all, we detect the overall time trend of mortality rate in the U.S. from Year 1999-2018 by sex categories. Here, an age-adjusted mortality rate is introduced to allow different age structures to be compared.[3] According to the U.S. mortality data collected by CDC [4], the age-adjusted

death rates for year 1999-2009 are basing on 2000 U.S. standard population, while rates for 2011-2018 are basing on 2010 U.S. standard population. Despite the different standards of data adjustment, we can still see a smooth decreasing trend under different age groups. Males have a higher age-adjusted rate than females. We can observe a greater decline in males' mortality rates than females.

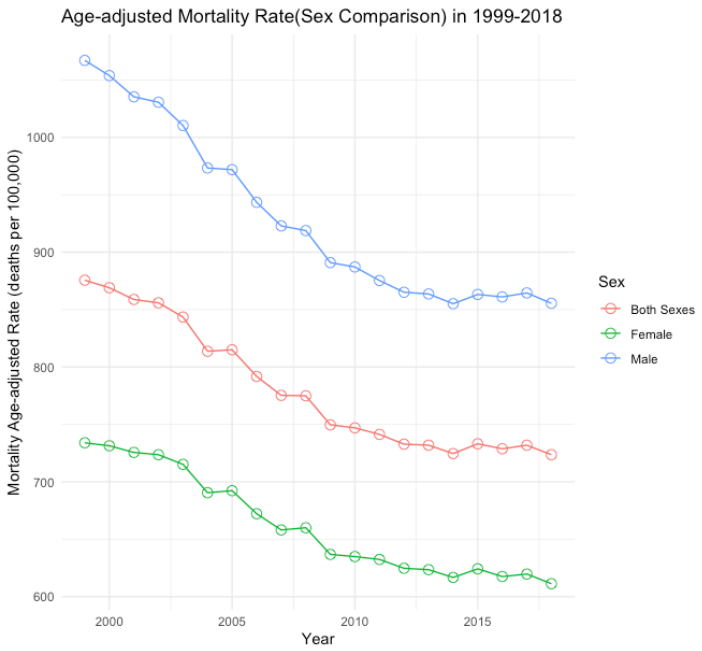

Figure 1: Age-adjusted Mortality Rate Under Sex Categories

In more detail, we can detect the trends of mortality in the U.S. from the Year 1999-2018 by sex and race categories. Here, we only compare age-adjusted mortality rates among white and black groups since these two are the main population groups in the U.S. When controlling the sex categories, black people have a higher mortality rate than whites. Both race groups under the male category have higher mortality rates than females. Also, we detect black males have a sharp decline from the year 2002 to 2010, while three other groups hold similar decline tendencies.

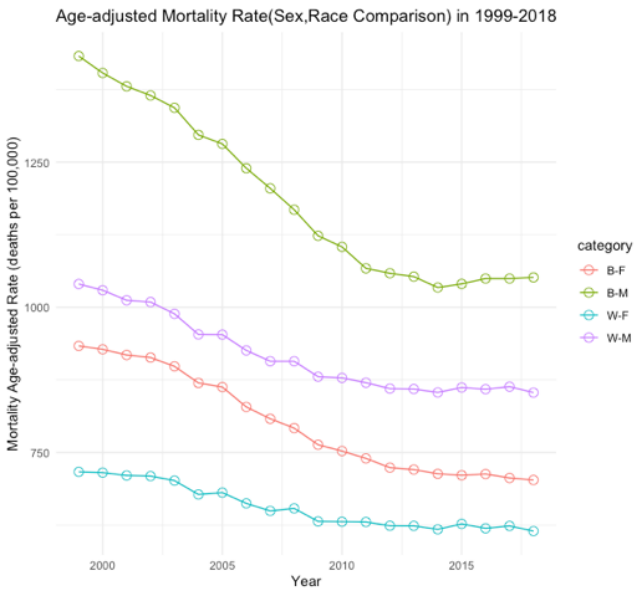

Figure 2: Age-adjusted Mortality Rate Under Sex & Race Categories

3.2 Marital Status, Resident Status, Education Levels Detection

Under the overall analysis, marital status, resident status, and education levels are the basic information of mortality cases. We concern whether the variation of levels in these three fields influences the mortality counts.

From Census Bureau Statistics of 2019 U.S. Social Characteristics [5], we can calculate the average proportions of marital status for single, married, widowed divorced groups are 33.95%, 49.5%, 5.65%, and 10.9%. Suppose there are no sharp changes among the year 2015-2019, we can compare the whole population's proportions of marital status in U.S. and mortality cases' proportions of marital status among years. We believe marital status influenced the proportion of mortality: Divorced and Widowed groups have higher proportions to mortality.

Table 4.1 Martial Status Summary in Year 2015-2019

	2015	2016	2017	2018	2019
Divorced	15.536%	15.851%	16.100%	16.442%	16.724%
Married	36.880%	36.721%	36.496%	36.338%	36.283%
Single	12.925%	13.503%	13.700%	13.867%	14.092%
Unknown	0.740%	0.793%	0.777%	0.784%	0.809%
Widowed	33.920%	33.132%	32.927%	32.569%	32.093%

Similar, we can analysis on the resident status. See the appendix Table 4.2 for summary details of resident status. Under different education levels, we first detect the proportions under basic 9 categories as discussed in Table 2. See the appendix Table 4.3 for summary details of education levels.

From Census Bureau Statistics of 2015-2019 U.S. School Enrollment [6], we redefined the education levels to fit the enrollment total data for better analysis. Here, we combine "No Formal Education" and "8th Grade or Less" as a group of junior, "9-12th Grade, No Diploma" and "High School Graduate" as a group of high, "College Credit, No Degree", "Associate Degree", and "Bachelor's Degree" as a group of college, "Master's Degree" and "Doctorate Degree" as group grad, "Unknown" group is deleted. According to the estimated counts of school enrollment within the year 2015-2019, we can calculate the education rates for every level. See the appendix Table 4.4 for a count summary of different school enrollment levels. From figure 3 below, we see a trend that people under high school education level have the highest ratio of deaths, junior-level follows it. As the level of education rises, the death rate falls.

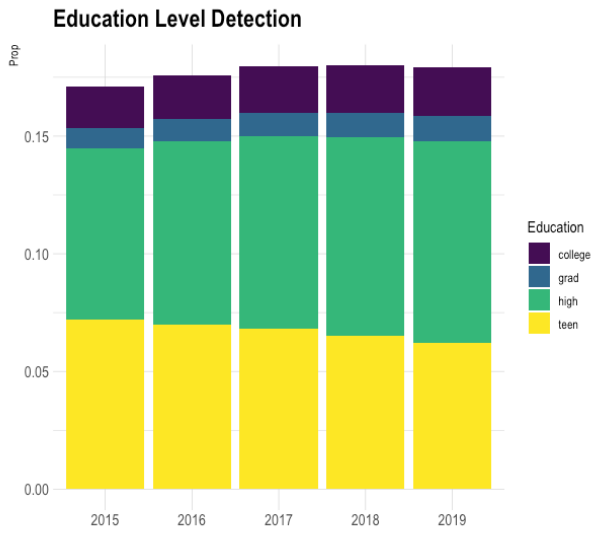

Figure 3: Rates of Education Levels in Year 2015-2019

Annotation: Since the Total numbers of School Enrollment from the U.S. Census Bureau are estimated, the rates for each category may not be accurate.

3.3 Time Tendency of Mortality Causes

First of all, we define adults as ages between 20 to 59 years old, includes the two boundaries. We detect suicide and Ischemic Heart Disease as two main categories of causes of mortality. Here, we introduce month categories as a new factor and detect the differences of mortality counts in 12 months under these two causes.

3.3.1 Suicide Analysis Within 12 Months

For a balanced design, here we define 12 levels of the treatment month, replicates 5 times in each treatment. See Table 4.5 for more details of years frequency statistics. To detect whether suicide counts are different among different months, we perform the one-way ANOVA. To perform Bartlett's Test for equal variances first, we have p-value=0.7034 and cannot reject the null hypothesis, i.e., all 12 groups' population variances are equal. Then, both Bartlett's Test and overall ANOVA results can be found in appendix 2. We have p-value=2.35e-12 for one-way ANOVA, i.e., we reject the null hypothesis and conclude suicide deaths are different among months.

A pairwise comparison using the LSD method can be performed on the 12 groups. We find the results that suicide rates are higher in March to August, i.e., during the spring and summer, while suicide rates in February, November, December are the lowest.

3.3.2 Ischemic Heart Disease Analysis Within 12 Months

Similarly, we defined a balanced design with 12 levels of the treatment month, replicates 5 times in each treatment. See Table 4.6 for more details of years frequency statistics. To perform Bartlett's Test for equal variances first, we have p-value=0.9796 and cannot reject the null hypothesis, i.e., all 12 groups' population variances are equal. Then, both Bartlett's Test and overall ANOVA results can be found in appendix 2. We have p-value=3.23e-06 for one-way ANOVA, i.e., we reject the null hypothesis and conclude heart disease deaths are different among months.

After performing a pairwise comparison among 12 months, we get the results that Ischemic Heart Disease has a winter peak and a summer peak within December, January, and July.

4 Year 2019 Analysis

4.1 Top Mortality Causes Analysis

4.1.1 Analysis on Adults

Basing on the age categories standard in section 3.3, we first analysis on Adults. According to the standard of 39 Causes of Death [2], we summarize the top 5 causes of mortality in overall and by sex:

Table 5.1 Top Overall Mortality Causes in Year 2019

Rank	Causes
1	All other and unspecified accidents and adverse effects
2	All other diseases
3	Ischemic heart diseases
4	Intentional self-harm (suicide)
5	Other malignant neoplasms

After removing the “all other diseases” and “accidents” from the list, we have heart disease, suicide, malignant neoplasms are the three main categories of mortality causes. Similarly, tables of mortality caused by sex can be found in appendix table 5.2, 5.3. Males hold higher suicide frequency than females, while females have higher frequencies in diseases related to the heart.

4.1.2 Analysis on Elderly

Here, we define elderly as age exceed 60 years old, include the boundary. According to the standard of 39 Causes of Death [2], we summarize the top 5 causes of mortality in overall and by sex:

Table 5.4 Top Overall Mortality Causes in Year 2019

Rank	Causes
1	All other diseases
2	Ischemic heart diseases
3	Other diseases of heart
4	Chronic lower respiratory diseases
5	Other malignant neoplasms

Diseases to mortality in the elderly group are focused on heart diseases and respiratory diseases. Similarly, tables of mortality caused by sex can be found in appendix table 5.5, 5.6. Especially, the female elderly group has Alzheimer's disease and Cerebrovascular disease as two high-frequency categories of mortality causes, while the male elderly group doesn't have.

4.2 State Level Mortality Analysis

To detect the mortality rates among states in the year 2019, we first collect the monthly number of deaths by state from CDC [7] and population by state [8]. The mortality rate is defined as mortality counts divide the total population for every state. Below is the heatmap for 2019 U.S. mortality rate levels by states, the deeper the color is, the higher the mortality rate is. From the plot, the below states have relatively high mortality rates: West Virginia, Maine, Kentucky, Alabama, Mississippi, Arkansas.

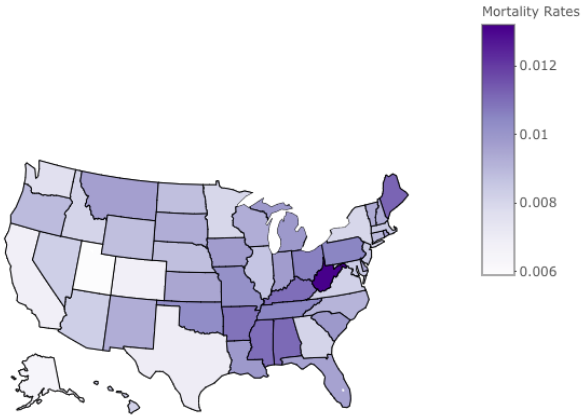

Figure 6: 2019 U.S. Mortality Rate Levels by States

4.3 Relationship Detection on Manner Status and Education Levels

Recall that the overall analysis on the relationship between education levels and mortality rates, we conclude that junior and high-school groups have the top mortality rates. From table 5.7 and table 5.8 in the appendix, we have accidents and natural deaths are two main categories of manner status, we only focus on these two categories.

Repeat the frequency analysis on two education groups of causes of deaths, we get similar results from the 2019 years comparison in section 4.1. The heart-related disease is always the top 3 mortality causes among different education levels, different races as well as age groups. Here,

we use this disease as a representative of the whole trend of mortality rates among states. To detect the state's differences in heart disease mortality rates, we cite data from CDC [9] to further analyze.

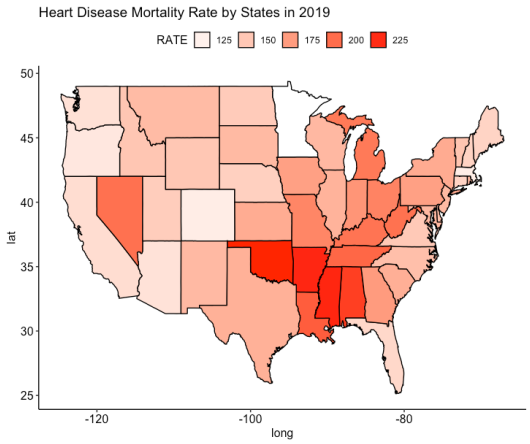

Figure 7: Heart Disease Mortality by State in 2019

From figure 7, the death rate is recorded as per 100k total population. States are categorized from highest rate to the lowest, the deeper the color is, the higher the rate is. Oklahoma, Arkansas, Louisiana, Mississippi, Alabama, Tennessee, Nevada are the top echelons in heart disease mortality rate.

According to a study of increasing education levels on rising life expectancy [10], there is a significant association between education and overall population health. Thus, we make comparisons among states on the education levels and try to verify whether there is the relationship between education levels and mortality rates in states. To clear, education rank data are collected from WalletHub, comparing all 50 states across 18 metrics to evaluate the state education levels.[11]

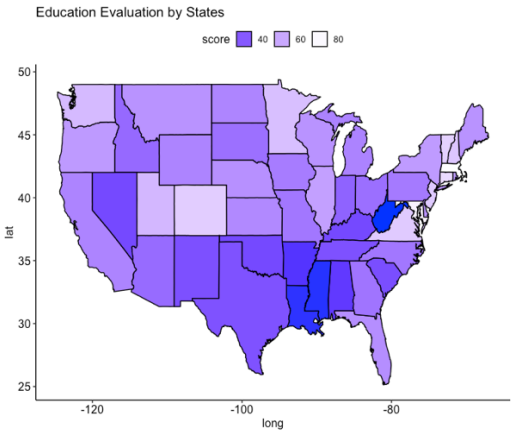

Figure 8: Education Evaluation by State

To make the color of Figure 8 contrast with Figure 7, we set the higher the education level of a state, the lighter the color on the map, so you can see West Virginia, Oklahoma, Arkansas, Louisiana, Mississippi, Alabama, Nevada are states with poor education. The results corresponding to the high levels of mortality rates in heart disease in these states, i.e., the relationship between education evaluation and mortality rates is reasonable.

5 Summary

Through this study, we aim to analyze the mortality rates in the U.S. among years. Descriptive analysis and visualization are two main parts of the report. In the overall analysis, we focus on the overall trend and detect a smooth decreasing trend among years. Males and a Black group of people hold high mortality rates among years. Education level is an important factor influencing the mortality rate: people under low levels of education have a high proportion of mortality rates. Further discussions are found in the year 2019 analysis. To detect the time tendency of mortality causes, we focus analysis on adults (age between 20 to 59 years old) and two main categories of causes of mortality (suicide and Ischemic Heart Disease). Spring and summer are the seasons with a high incidence of suicide, while winter and summer are the seasons with a high incidence of Ischemic Heart Disease.

In the 2019 analysis section, we discuss the top mortality causes analysis of adults (age between 20 to 59 years old) and elders (age larger than 60 years old). Comparison of mortality rates among states is included and a relationship between education levels and the heart mortality rate is also being verified.

Most of the analysis in this report is only a macro-level discussion, and there is no detailed discussion on the cause of death and the method of death. Further analysis will require more auxiliary data.

Reference

[1] *CDC Mortality Multiple Cause Files*. (n.d.). Centers for Disease Control and Prevention (CDC). https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm#Mortality_Multiple

[2] Centers for Disease Control and Prevention. (2018, January). *Underlying and Multiple Cause of Death Codes*. https://www.cdc.gov/nchs/data/datalinkage/underlying_and_multiple_cause_of_death_codes.pdf

[3] *Age-Adjusted Rates - Statistics Teaching Tools*. (1999, April). THE OFFICIAL WEBSITE OF NEW YORK STATE. https://www.health.ny.gov/diseases/chronic/ageadj.htm

[4] *NCHS - Death rates and life expectancy at birth*. (1900–2018). [Dataset]. Centers for Disease Control and Prevention. https://data.cdc.gov/NCHS/NCHS-Death-rates-and-life-expectancy-at-birth/w9j2-ggv5/data

[5] *SELECTED SOCIAL CHARACTERISTICS IN THE UNITED STATES*. (2019). [Dataset]. United States Census Bureau. https://data.census.gov/cedsci/table?q=dp02&hidePreview=false&tid=ACSDP1Y2019.DP02&vintage=2018

[6] *SCHOOL ENROLLMENT*. (2019). [Dataset]. United States Census Bureau. https://data.census.gov/cedsci/table?q=school&tid=ACSST1Y2019.S1401

[7] *State and National Provisional Counts 2019 Data*. (2019). [Dataset]. Centers for Disease Control and Prevention. https://www.cdc.gov/nchs/nvss/vsrr/provisional-tables.htm

[8] *2019 U.S. Population Estimates by State*. (2019). [Dataset]. United States Census Bureau. https://www.census.gov/search-results.html?searchType=web&cssp=SERP&q=population%20by%20states

[9] Heart Disease Mortality by State
https://www.cdc.gov/nchs/pressroom/sosmap/heart_disease_mortality/heart_disease.htm

[10] Luy, M., Zannella, M., Wegner-Siegmundt, C. *et al.* The impact of increasing education levels on rising life expectancy: a decomposition analysis for Italy, Denmark, and the USA. *Genus* 75, 11 (2019). https://doi.org/10.1186/s41118-019-0055-0

[11] McCann, A. (2021, December 16). *Most & Least Educated States in America*. WalletHub. https://wallethub.com/edu/e/most-educated-states/31075

Appendix A

The original datasets are download from CDC website—Mortality Multiple Cause Files (https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm#Mortality_Multiple). Here, data from year 2015-2019 are collected. Every year, only 12 variables are chosen for the analysis: Resident Status, Education Levels, Month of Death, Sex, Age, Place of Status, Martial Status, Manner of Death, 39 Cause Recode, Race, Race_Detailed.

The cleaned and accessible datasets can be found on my Github webpage: https://github.com/WenxiaoZhou/USMortality_project/tree/main/Data

A file named as "USall.RData" is overall dataset for 2015-2019. This file as well as the "US19.RData" are the main datasets used for the overall analysis and 2019 analysis. What's more, you can find four other Excel files corresponding to the Table 3 that are auxiliary datasets for detailed analysis.

Appendix B

1.Appendix in Section 3.2

Table 4.2 Resident Status Summary in Year 2015-2019

	2015	2016	2017	2018	2019
Residents	80.829%	80.611%	80.518%	80.502%	80.298%
Intrastate Nonresident	15.877%	16.064%	16.117%	16.128%	16.317%
Interstate Nonresident	3.089%	3.120%	3.134%	3.120%	3.152%
Foreign Residents	0.205%	0.204%	0.232%	0.249%	0.234%

Table 4.3 Education Levels Summary in Year 2015-2019

	2015	2016	2017	2018	2019
No Formal Education	2.687%	0.800%	0.530%	0.000%	0.000%
8th Grade Or Less	10.681%	10.569%	10.145%	9.801%	9.385%
9-12th Grade, No Diploma	10.261%	10.631%	10.379%	10.172%	9.992%
High School Graduate	39.644%	41.796%	42.049%	42.465%	42.467%
College Credit, No Degree	11.602%	12.197%	12.400%	12.498%	12.632%
Associate Degree	5.678%	6.087%	6.207%	6.340%	6.451%
Bachelor's Degree	9.606%	10.180%	10.442%	10.635%	10.805%
Master's Degree	3.825%	4.139%	4.271%	4.443%	4.565%
Doctorate Degree	1.607%	1.737%	1.796%	1.843%	1.878%
Unknown	4.408%	1.865%	1.780%	1.804%	1.824%

Table 4.4 School Enrollment Summary in Year 2015-2019

	2015	2016	2017	2018	2019
Junior	41874863	41857908	41828973	41747569	41703677
High School	17086446	17108087	17150113	17100878	16932635
College	18620217	18436671	18105312	17792902	17507427
Graduate	4036762	4169611	4188939	4275426	4321881

2.Appendix in Section 3.3

Table 4.5 Suicide Data for One-factor Experiment

Treatment	2015	2016	2017	2018	2019
January	2471	2523	2536	2632	2573
Febuary	2263	2370	2338	2423	2403
March	2724	2663	2598	2698	2812
April	2540	2581	2609	2613	2669
May	2784	2656	2876	2851	2770
June	2660	2606	2835	2964	2735
July	2740	2816	2889	2973	2820
August	2718	2779	2947	2919	2797
September	2498	2596	2735	2806	2754
October	2538	2607	2771	2733	2752
November	2295	2334	2586	2419	2437
December	2383	2371	2439	2526	2500

```
> bartlett.test(Freq~Month,data=ti1)

        Bartlett test of homogeneity of variances

data:  Freq by Month
Bartlett's K-squared = 8.1099, df = 11, p-value = 0.7034

> comp1<-aov(Freq~Month, data = ti1)
> summary(comp1)
             Df  Sum Sq Mean Sq F value   Pr(>F)
Month        11 1504113  136738   15.64 2.35e-12 ***
Residuals    48  419610    8742
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 4: Results Of Bartlett Test and One-way ANOVA For Suicide Data

Table 4.6 Heart Disease Data for One-factor Experiment

Treatment	2015	2016	2017	2018	2019
January	4472	4166	4156	4278	3804
Febuary	3864	3999	3704	3647	3579
March	4182	4244	4077	3945	3743
April	3921	3819	3730	3620	3608

May	4069	3951	3817	3720	3629
June	3806	3685	3737	3599	3382
July	3952	3816	3785	3773	3582
August	3864	3833	3648	3593	3330
September	3584	3602	3664	3391	3348
October	3751	3829	3715	3587	3354
November	3822	3744	3546	3726	3509
December	4094	4148	4023	3836	3871

```
> bartlett.test(Freq~Month,data=ti3)

        Bartlett test of homogeneity of variances

data:  Freq by Month
Bartlett's K-squared = 3.6265, df = 11, p-value = 0.9796

> summary(comp3)
            Df  Sum Sq Mean Sq F value   Pr(>F)
Month       11 2035648  185059   6.192 3.23e-06 ***
Residuals   48 1434626   29888
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 5: Results Of Bartlett Test and One-way ANOVA For Heart Disease Data

3.Appendix in Section 4.1

Table 5.2 Females Top Overall Mortality Causes in Year 2019

Rank	Causes
1	All other diseases
2	All other and unspecified accidents and adverse effects
3	Other malignant neoplasms
4	Ischemic heart diseases
5	Malignant neoplasm of breast

Table 5.3 Males Top Overall Mortality Causes in Year 2019

Rank	Causes
1	All other and unspecified accidents and adverse effects
2	All other diseases
3	Ischemic heart diseases
4	Intentional self-harm (suicide)
5	Other malignant neoplasms

Table 5.5 Females Top Overall Mortality Causes in Year 2019

Rank	Causes
1	All other diseases
2	Ischemic heart diseases

3	Other diseases of heart
4	Alzheimer's disease
5	Cerebrovascular diseases

Table 5.6 Males Top Overall Mortality Causes in Year 2019

Rank	Causes
1	All other diseases
2	Ischemic heart diseases
3	Other diseases of heart
4	Other malignant neoplasms
5	Chronic lower respiratory diseases

4.Appendix in Section 4.3

Table 5.7 Manner Summary of Junior Mortality Cases in Year 2019

Manner	Not Specified	Accidents	Suicide	Homicide	Pending Investigation	Could not determine	Natural
Counts	33679	12909	1730	1680	650	2533	215383

Table 5.8 Manner Summary of High-School Mortality Cases in Year 2019

Manner	Not Specified	Accidents	Suicide	Homicide	Pending Investigation	Could not determine	Natural
Counts	116194	98941	24715	14037	2464	5537	1239258