STA440 - Final Project

Sophia Li

2024-12-16

Introduction

Background

Suicide is a significant public health issue in China, particularly among farmers, who represent a large proportion of the rural population. Although the overall suicide rate has declined substantially in the past two decades, it remains the leading cause of death for individuals aged 15 to 34 years (Jia & Zhang, 2012). Farmers face unique challenges, including economic instability, exposure to lethal means such as pesticides, and limited access to mental health services, all of which contribute to their vulnerability (Jia & Zhang, 2012). Moreover, cultural expectations, such as maintaining family harmony and collective responsibilities, can exacerbate psychological distress during periods of personal or financial conflict. These stressors, combined with educational disparities and a lack of resources, make farmers disproportionately affected by suicide risk (Ai et al., 2017). Understanding the factors contributing to suicide among farmers is essential for developing targeted interventions to address this critical public health issue.

One the other hand, the 2008 global financial crisis had far-reaching consequences for rural China, particularly its agricultural sector. While urban areas grappled with industrial layoffs, rural regions experienced disruptions in farming due to the decline in agricultural exports and increased financial strain. The crisis caused a significant reduction in the global trade of agricultural products, severely affecting the livelihoods of Chinese farmers who depended on crop sales for their income (Hung, 2020). In response, the Chinese government implemented a 4 trillion RMB stimulus package aimed at stabilizing the economy, which included measures to support agricultural productivity and rural development (Shiraishi & Yano, 2022). However, these interventions failed to address the deeper vulnerabilities of rural communities, where farming remained heavily dependent on unpredictable market and environmental factors. Against this backdrop, this project aims to investigate whether fatality rates from suicide varied across the years 2009, 2010, and 2011, and to explore how these temporal changes might reflect the broader socioeconomic challenges faced by rural farmers during and after the financial crisis.

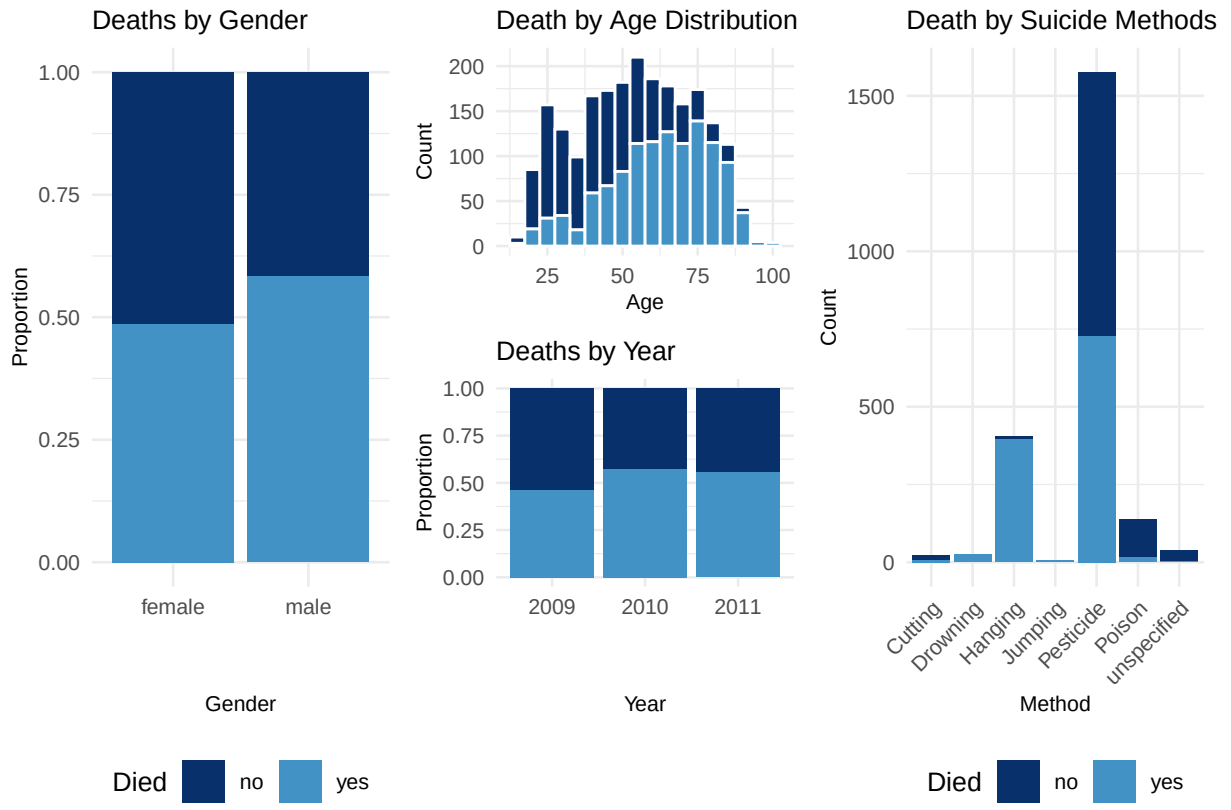
However, the definition of suicide attempt is ambiguous and hard to explicitly define. High-lethality and low-lethality suicide attempts differ significantly in terms of intention, planning, and associated risk factors. High-lethality attempts are often characterized by greater planning and intent, suggesting a deliberate and less impulsive approach, while low-lethality attempts tend to occur in more reactive, emotionally charged situations (Zhang et al., 2022). These distinctions imply that the populations engaging in these two types of attempts are heterogeneous and may require different preventative strategies. This project will build on these findings by examining the factors that distinguish high-lethality and low-lethality suicide attempts in the dataset. By modeling and comparing these two groups, the research aims to uncover critical predictors of fatality rates and contribute to more tailored intervention strategies (Zhang et al., 2022).

Data Introduction

The dataset for this project originates from a public health surveillance study conducted in three counties of Shandong Province, China, between 2009 and 2011 (Sun et al., 2015). The data were systematically

collected through a collaboration between local health authorities and hospitals, which documented all reported cases of serious suicide attempts (SSAs)—defined as suicide attempts that resulted in either death or hospitalization. Each case was recorded using standardized forms to ensure accuracy and completeness. The dataset contains 2,571 observations and includes key variables necessary to analyze suicide attempts. These variables can be grouped into three categories:

Demographic variables: Age, sex, education level (categorical variable:primary, secondary, tertiary), and occupation (categorical variable later being filtered to only farming and household). Suicide-related variables: Method of suicide (e.g., pesticide ingestion, hanging, poisoning, drowning, and other methods), outcome (whether the individual died or survived), and urban/rural location. Temporal variables: Year (2009, 2010, 2011) and month of the attempt, which were later categorized into seasons (Winter, Spring, Summer, and Fall) for further analysis.



Objectives

This project aims to examine the extent to which demographic factors such as gender, age, and education influence fatality rates among the farming and household groups in Shandong Province. Additionally, it explores the year-to-year impact following the 2008 global financial crisis and tests whether seasonality, particularly winter, increases fatality rates due to delayed discovery of suicide attempts in rural areas. Finally, the project seeks to identify key contributing factors—including suicide methods—that make fatalities more likely.

Data Mutation and Implementation

The dataset primarily comprises individuals engaged in farming or household work, who account for an overwhelming 88.68% of the total sample. The remaining occupations are distributed as follows: business/service (0.82%), students (1.36%), workers (0.23%), professionals (1.44%), unemployed (1.17%), others/unknown (10.08%), and retirees (0.12%). These groups are not only small in size but also lack clear stratification or representativeness. For instance, there are no distinct occupational categories such as teachers, lawyers, or doctors to represent highly educated individuals, nor are there adequate numbers of blue-collar workers or technicians. The insufficient sample sizes limit their ability to provide meaningful insights into the population of Shandong Province. As a result, I focus my analysis exclusively on farming and household groups, as this approach ensures a more reliable and convincing exploration of fatality rates within the dominant occupational groups in the dataset.

In handling missing data, I determined that the missing values are missing at random (MAR) and cannot be reliably predicted or assumed. Since they provide no utility for the model, these observations are excluded from the analysis to maintain data integrity.

The dataset spans the years 2009 to 2011, a period that follows the 2008 global financial crisis. I aim to investigate whether the crisis had any lingering effects on the fatality rates during this timeframe. Instead of treating year as a numeric variable, which would misrepresent its role in the analysis (since fatalities occur every year), I converted year into a categorical variable. This transformation allows me to examine year-to-year variations more effectively and interpret the results meaningfully.

Similarly, to assess whether seasonality influences fatality rates, I categorized the 12 months into four seasons aligned with the distinct seasonal patterns of Shandong Province:

Spring: February to May (3-5), Summer: June to August (6-8), Autumn: September to November (9-11), Winter: December to February (12-2).

This approach avoids incorrect implications that would arise from treating months as a numeric variable, as fatalities occur in all months.

Another critical issue involves the hospitalized variable, which indicates whether a patient received treatment following a suicide attempt. The data reveals a strong association between hospitalization and fatality: all non-hospitalized patients died, while hospitalized patients experienced mixed outcomes—some survived, and others still succumbed. Including hospitalization as a predictor would bias the model due to its direct connection with fatality outcomes. Instead, I combined the hospitalized and died variables to create a new outcome variable:

High-Fatality: Patients who died, regardless of hospitalization (e.g., immediate death, undiscovered cases like drowning, or deaths despite treatment). Medium-Fatality: Patients who survived after receiving treatment in a hospital.

I did not create a low-fatality group because there were no patients who both survived and avoided hospitalization in the dataset.

Regarding the suicide method, I identified redundancies in naming, such as “Poison,” “Other Poison,” and “Poison Unspecified.” These categories were combined into a single category labeled “Poison” for consistency and clarity. For education levels, I excluded observations with unknown educational backgrounds(7 cases) due to ambiguity and lack of clarity.

Methodology

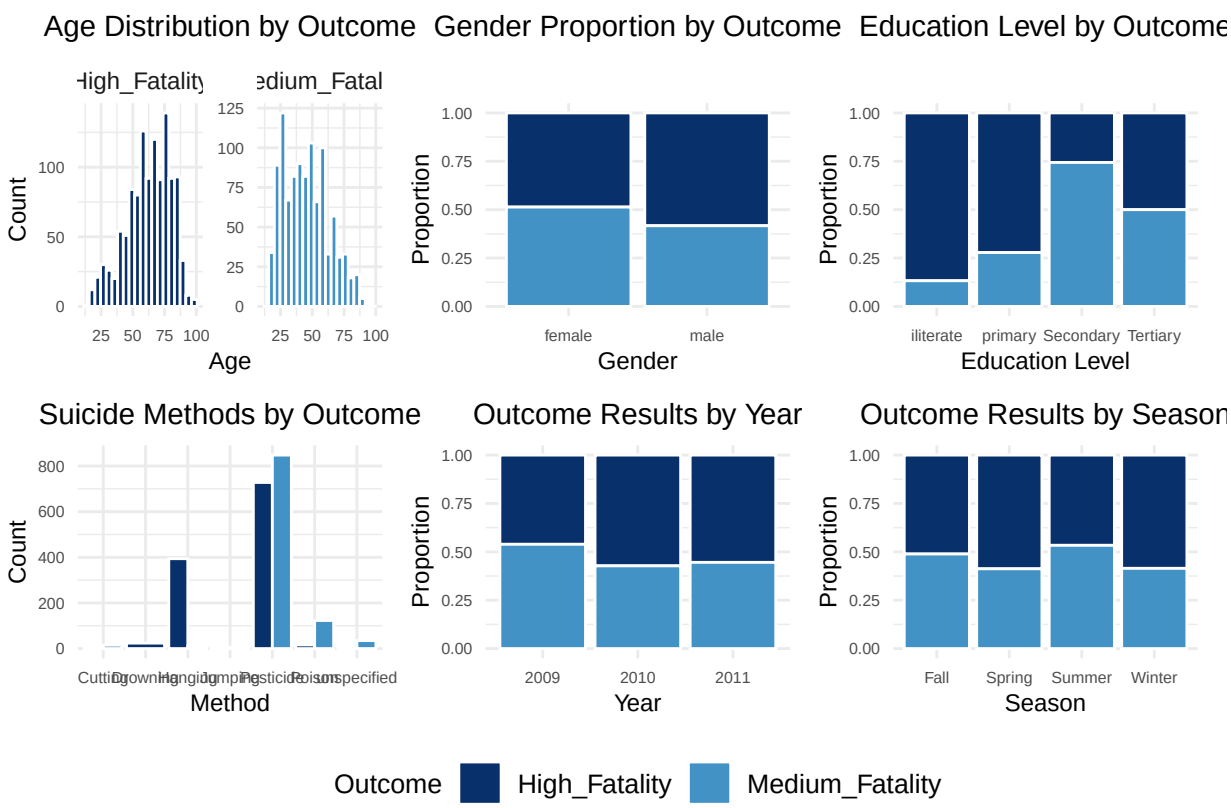
Variable Selection

Based on the exploratory data analysis (EDA) graphs, the pre-selection of variables was guided by observed trends and personal preferences to focus on factors most relevant to the study objectives. Age was selected

due to its clear association with death, as seen in the distribution where older age groups exhibited higher mortality rates. Gender was included because of notable differences in proportions of deaths between males and females. Education level was chosen as it revealed significant disparities, with lower education groups showing a higher proportion of deaths. Suicide methods, particularly pesticide use, hanging, and jumping, were retained due to their clear dominance in both frequency and association with fatality outcomes. Temporal variables, such as year and season, were included to explore potential patterns over time and seasonal variations, which might reflect environmental or social factors influencing fatality rates.

The urban versus rural classification was also considered, as it may reflect disparities in access to emergency care or resource availability. However, the exploratory data analysis (EDA) revealed no significant differences in survival outcomes between urban and rural areas. Consequently, the urban/rural variable has been excluded from the model. While the original paper that introduced the dataset does not provide a detailed explanation of the urban variable, based on my knowledge and observations, it is likely associated with the HuKou (registered residence location) system which is special to China, which may not necessarily influence fatality rates.

As mentioned above, I combined two variables(Died and Hospitalized) into one combined variable and categorize them into two levels - high fatality and medium fatality. I also generate EDA graphs to explore potential patterns and relationships between key predictors and the combined Outcome variable. The Age distribution indicates that high-fatality cases are concentrated among middle-aged individuals (40–60 years), with medium-fatality cases spanning a broader age range. Gender analysis reveals a higher proportion of high-fatality outcomes among males, while medium-fatality outcomes are more evenly distributed between genders. Education level shows a clear trend, where lower education levels, particularly among the illiterate and those with only primary education, are strongly associated with high-fatality cases. Suicide methods highlight the dominant role of pesticide ingestion in high-fatality outcomes, while methods like hanging and poisoning contribute more to medium-fatality cases. Yearly trends suggest relatively stable proportions of high-fatality outcomes from 2009 to 2011, whereas medium-fatality cases show a slight increase. Seasonal patterns indicate a slightly higher proportion of medium-fatality outcomes during spring and summer, which may reflect increased exposure to environmental or occupational stressors. Together, these findings highlight the importance of demographic, educational, temporal, and behavioral factors in understanding suicide fatality outcomes.



Model Assumption & Diagnostics

To analyze the combined variable as the response and include the selected predictors, I chose to use a multinomial logistic regression model. This approach also allows me to interpret the odds and identify which variables, or specific categories within them, are statistically significant among those considered.

The multinomial logistic regression model built for this analysis largely satisfies the key model assumptions. Independence of observations is reasonable since each individual in the dataset is a separate suicide attempt, and no clustering within households or regions is explicitly indicated. No perfect multicollinearity is confirmed as the Variance Inflation Factor (VIF) values for all predictors are well below the accepted threshold, indicating no problematic correlation among predictors. Regarding the linearity of predictors with log-odds, Age, as a continuous predictor, appears suitable for the model; however, further diagnostics like logit plots may be needed to verify its linear relationship with the log-odds. The absence of outliers has not been explicitly tested yet, but influential observations could be evaluated through diagnostic tools such as Cook's distance to ensure model robustness. Additionally, the dataset meets the assumption of adequate sample size, as there are sufficient observations for each predictor category, reducing the risk of convergence issues or bias. Finally, the non-overlapping categories assumption for the response variable is satisfied since High_Fatality and Medium_Fatality are mutually exclusive and exhaustive, providing a clear and valid outcome structure. Overall, the model assumptions appear reasonably met, with minor areas for further verification.

Model Implementation

$$\log\left(\frac{P(Y=1)}{P(Y=0)}\right) = \beta_0 + \beta_1 \text{Age} + \beta_2 \text{Sex} + \beta_3 \text{Education} + \beta_4 \text{method} + \beta_5 \text{Year} + \beta_6 \text{Season}$$

Table 1: Binary Logistic Regression Results

Predictor	Estimate	Std_Error	Z_value	P_value
Intercept	1.123417	0.607402	1.850	0.0643800
Age	-0.017616	0.003917	-4.498	6.87e-06
Sex (male)	-0.318367	0.116802	-2.726	0.0064200
Education (Primary)	0.822793	0.180521	4.558	5.17e-06
Education (Secondary)	2.481094	0.191751	12.939	2.00e-16
Education (Tertiary)	0.865654	1.084484	-0.798	0.4247400
Method (Drowning)	-17.578842	416.613360	-0.042	0.9663400
Method (Hanging)	-4.133257	0.624076	-6.623	3.52e-11
Method (Jumping)	-3.737283	1.228425	-3.042	0.0023500
Method (Pesticide)	-0.984085	0.540250	-1.822	0.0685300
Method (Poison)	1.389876	0.609575	2.280	0.0226000
Method (Unspecified)	1.684696	0.839992	2.006	0.0449000
Year (2010)	-0.356449	0.141794	-2.514	0.0119400
Year (2011)	-0.628994	0.146920	-4.281	1.86e-05
Season (Spring)	-0.421115	0.164219	-2.564	0.0103400
Season (Summer)	0.058410	0.156023	0.374	0.7081300
Season (Winter)	-0.402991	0.174818	-2.305	0.0211600

Here:

- $P(Y = 1)$: Probability of the outcome being "High_Fatality"- immediate death or death afte treatment.
- $P(Y = 0)$: Probability of the outcome being "Medium_Fatality" - survival after treatment.
- $\beta_0, \beta_1, \ldots, \beta_6$: Coefficients of the predictors.

Results

In this project, high-fatality is defined as either patients who died after attempting suicide despite receiving hospital treatment or those who died immediately after attempting suicide without receiving any treatment. While these two conditions differ slightly, they are both classified under the high-fatality standard. Additionally, this model focuses exclusively on individuals whose occupations are classified as household or farming. Year and month have been converted into categorical variables to better capture the temporal and seasonal effects on fatality outcomes.

The binary logistic regression model investigates the factors influencing the likelihood of a high fatality outcome compared to a medium fatality outcome. Age is a significant predictor ($p < 0.001$), where increasing age reduces the odds of high fatality slightly (Estimate = -0.0176). Gender also shows a significant effect ($p = 0.0064$), with males having lower odds of high fatality compared to females (Estimate = -0.318). This suggests that gender differences may play a role in survival outcomes following a suicide attempt. Education level reveals mixed effects. Primary education significantly increases the odds of high fatality ($p < 0.001$), and secondary education has an even stronger effect ($p < 0.001$). However, tertiary education is not statistically significant ($p = 0.4247$), suggesting its effect on fatality is minimal. Regarding suicide methods, hanging ($p < 0.001$) and jumping ($p = 0.0024$) are associated with significantly lower odds of high fatality, whereas poison ($p = 0.0226$) and unspecified methods ($p = 0.0449$) increase the odds of high fatality. Notably, drowning shows no significant association, likely due to data sparsity.

Temporal effects are also significant. Compared to 2009, the odds of high fatality decrease significantly in 2010 ($p = 0.0119$) and 2011 ($p < 0.001$), suggesting an overall decline in fatality rates over time. Seasonal variation reveals that spring ($p = 0.0103$) and winter ($p = 0.0212$) significantly reduce the odds of high fatality compared to fall, while summer has no significant effect ($p = 0.7081$).

The binary logistic regression model provides odds ratios that allow us to interpret the likelihood of a high fatality outcome relative to a medium fatality outcome.

For age, the odds ratio is approximately 0.98 (exp(-0.0176)), meaning that for every additional year of age, the odds of a high fatality outcome decrease by about 2%, holding all other variables constant. This suggests that older individuals have a slightly lower likelihood of experiencing high fatality compared to younger individuals. For gender, the odds ratio for males is approximately 0.73 (exp(-0.318)), indicating that males have about 27% lower odds of a high fatality outcome compared to females, holding other factors constant. This suggests that gender may influence survival rates following suicide attempts.

For suicide methods, the results vary. Hanging has an odds ratio of 0.016 (exp(-4.13)), meaning that the odds of high fatality are approximately 98.4% lower for individuals using this method compared to cutting methods. Similarly, jumping has an odds ratio of 0.024 (exp(-3.74)), indicating a 97.6% reduction in odds comparing with cutting methods. On the other hand, poisoning shows an odds ratio of 4.01 (exp(1.39)), suggesting that individuals who use poisoning have about 4 times higher odds of high fatality than cutting as baseline. Unspecified methods also increase the odds of high fatality, with an odds ratio of 5.39 (exp(1.68)).

For year, both 2010 and 2011 demonstrate significant reductions in the odds of high fatality compared to 2009. Specifically, the odds ratio for 2010 is 0.70 (exp(-0.356)) and for 2011 is 0.53 (exp(-0.629)), representing a 30% and 47% decrease in odds, respectively. This suggests a downward trend in high fatality rates over time. Finally, for seasons, spring and winter are associated with significantly lower odds of high fatality compared to fall. The odds ratios for spring and winter are 0.66 (exp(-0.421)) and 0.67 (exp(-0.403)), respectively, indicating about a 34% reduction in the odds of high fatality in these seasons. Summer, however, does not significantly differ from fall. Therefore, the consecutive decline in fatality rates from 2009 to 2011 may reflect a gradual but steady recovery of China's farming sector and overall conditions following the 2008 global financial crisis. While there may not be a direct connection between fatality rates and national economic or societal conditions, it is reasonable to speculate that they are indirectly related. The improving trends over the years could also be linked to better medical access, increased social awareness within farming and household groups, rising household incomes, and overall improved living conditions, as China experienced steady economic growth and development throughout the 2010s.

In summary, significant predictors such as age, gender, education, suicide method, year, and season demonstrate varying influences on the odds of high fatality. Methods like poisoning and temporal trends (e.g., year effects) highlight key areas for targeted prevention strategies.

Results

In this project, high-fatality is defined as either patients who died after attempting suicide despite receiving hospital treatment or those who died immediately after attempting suicide without receiving any treatment. While these two conditions differ slightly, they are both classified under the high-fatality standard. Additionally, this model focuses exclusively on individuals whose occupations are classified as household or farming. Year and month have been converted into categorical variables to better capture the temporal and seasonal effects on fatality outcomes.

The binary logistic regression model investigates the factors influencing the likelihood of a high fatality outcome compared to a medium fatality outcome. Age is a significant predictor ($p < 0.001$), where increasing age reduces the odds of high fatality slightly (Estimate = -0.0176). Gender also shows a significant effect ($p = 0.0064$), with males having lower odds of high fatality compared to females (Estimate = -0.318). This suggests that gender differences may play a role in survival outcomes following a suicide attempt. Education level reveals mixed effects. Primary education significantly increases the odds of high fatality ($p < 0.001$), and secondary education has an even stronger effect ($p < 0.001$). However, tertiary education is not statistically

significant ($p = 0.4247$), suggesting its effect on fatality is minimal. Regarding suicide methods, hanging ($p < 0.001$) and jumping ($p = 0.0024$) are associated with significantly lower odds of high fatality, whereas poison ($p = 0.0226$) and unspecified methods ($p = 0.0449$) increase the odds of high fatality. Notably, drowning shows no significant association, likely due to data sparsity.

Temporal effects are also significant. Compared to 2009, the odds of high fatality decrease significantly in 2010 ($p = 0.0119$) and 2011 ($p < 0.001$), suggesting an overall decline in fatality rates over time. Seasonal variation reveals that spring ($p = 0.0103$) and winter ($p = 0.0212$) significantly reduce the odds of high fatality compared to fall, while summer has no significant effect ($p = 0.7081$).

The binary logistic regression model provides odds ratios that allow us to interpret the likelihood of a high fatality outcome relative to a medium fatality outcome.

For age, the odds ratio is approximately 0.98 (exp(-0.0176)), meaning that for every additional year of age, the odds of a high fatality outcome decrease by about 2%, holding all other variables constant. This suggests that older individuals have a slightly lower likelihood of experiencing high fatality compared to younger individuals. For gender, the odds ratio for males is approximately 0.73 (exp(-0.318)), indicating that males have about 27% lower odds of a high fatality outcome compared to females, holding other factors constant. This suggests that gender may influence survival rates following suicide attempts.

For suicide methods, the results vary. Hanging has an odds ratio of 0.016 (exp(-4.13)), meaning that the odds of high fatality are approximately 98.4% lower for individuals using this method compared to cutting methods. Similarly, jumping has an odds ratio of 0.024 (exp(-3.74)), indicating a 97.6% reduction in odds comparing with cutting methods. On the other hand, poisoning shows an odds ratio of 4.01 (exp(1.39)), suggesting that individuals who use poisoning have about 4 times higher odds of high fatality than cutting as baseline. Unspecified methods also increase the odds of high fatality, with an odds ratio of 5.39 (exp(1.68)).

For year, both 2010 and 2011 demonstrate significant reductions in the odds of high fatality compared to 2009. Specifically, the odds ratio for 2010 is 0.70 (exp(-0.356)) and for 2011 is 0.53 (exp(-0.629)), representing a 30% and 47% decrease in odds, respectively. This suggests a downward trend in high fatality rates over time. Finally, for seasons, spring and winter are associated with significantly lower odds of high fatality compared to fall. The odds ratios for spring and winter are 0.66 (exp(-0.421)) and 0.67 (exp(-0.403)), respectively, indicating about a 34% reduction in the odds of high fatality in these seasons. Summer, however, does not significantly differ from fall. Therefore, the consecutive decline in fatality rates from 2009 to 2011 may reflect a gradual but steady recovery of China's farming sector and overall conditions following the 2008 global financial crisis. While there may not be a direct connection between fatality rates and national economic or societal conditions, it is reasonable to speculate that they are indirectly related. The improving trends over the years could also be linked to better medical access, increased social awareness within farming and household groups, rising household incomes, and overall improved living conditions, as China experienced steady economic growth and development throughout the 2010s.

In summary, significant predictors such as age, gender, education, suicide method, year, and season demonstrate varying influences on the odds of high fatality. Methods like poisoning and temporal trends (e.g., year effects) highlight key areas for targeted prevention strategies.

Discussion

The results section does not include a discussion regarding the odds of educational levels on the combined variable. Based on the model, the interpretation would be as follows: Individuals with primary education have 2.28 times higher odds of high fatality compared to those with no formal education (exp(0.8227)). For secondary education, the odds ratio is 11.94 (exp(2.4811)), indicating a substantial increase in the likelihood of high fatality compared to the baseline group. Conversely, tertiary education does not show a significant impact on fatality outcomes ($p > 0.05$). However, this interpretation may be incorrect due to a possible interaction between education level and the hospitalized variable.

The proportional bar plot indicates that individuals with tertiary education are more likely to receive treatment compared to all other groups, and those with secondary education are also more likely to be hospitalized.

In contrast, patients with no formal education or primary education are the least likely to receive hospital treatment after a suicide attempt. Consequently, when the hospitalized variable is included as a response variable in the binary logistic regression model, it causes distorted coefficients due to multicollinearity. While the graph highlights the importance of education, the table results from the final model may be somewhat misleading. Therefore, I believe it is inappropriate to interpret the impact of education levels as described earlier, and I have chosen not to discuss their effects on the outcome in detail.

For the missing data, I directly removed or dropped records with "unknown" or "unspecified" values among our predictors. However, the missing data for the occupation variable could be categorized as missing at random (MAR). There are approximately 160 cases with unknown occupations, which I believe may be influenced by geographic factors. Shandong Province is predominantly agricultural, with limited opportunities for tourism, service, or sales-related jobs. Additionally, access to education in the region is highly competitive and limited, resulting in many individuals having only primary or even no formal education. As a result, the unknown occupations are likely due to MAR, as missingness can be associated with geographic and socioeconomic factors.On the other hand, the missing data for the suicide method variable, labeled as "unknown" or "unspecified," could be missing completely at random (MCAR). This is likely because certain suicide methods may not have been accurately identified or recorded at the time of the incident.While there are different types of missing data, I chose to remove the "unknown" values directly to maintain the clarity of graphs and table results. However, alternative methods for handling missing data, such as imputation, could potentially yield better outcomes.

On the other hand, focusing solely on the farming and household groups is more convenient and allows for clearer conclusions. However, there may be differences in the odds between these two groups, which raises the question of whether it is appropriate to combine them simply because they make up the majority of the data and share some similarities compared to other very different occupations, such as students or business/service providers. It would be beneficial to obtain a larger sample with a broader representation of occupations and social groups. This would allow me to extend the model to a more diverse and comprehensive population, leading to more generalizable conclusions instead of exclusively focusing on farmers.

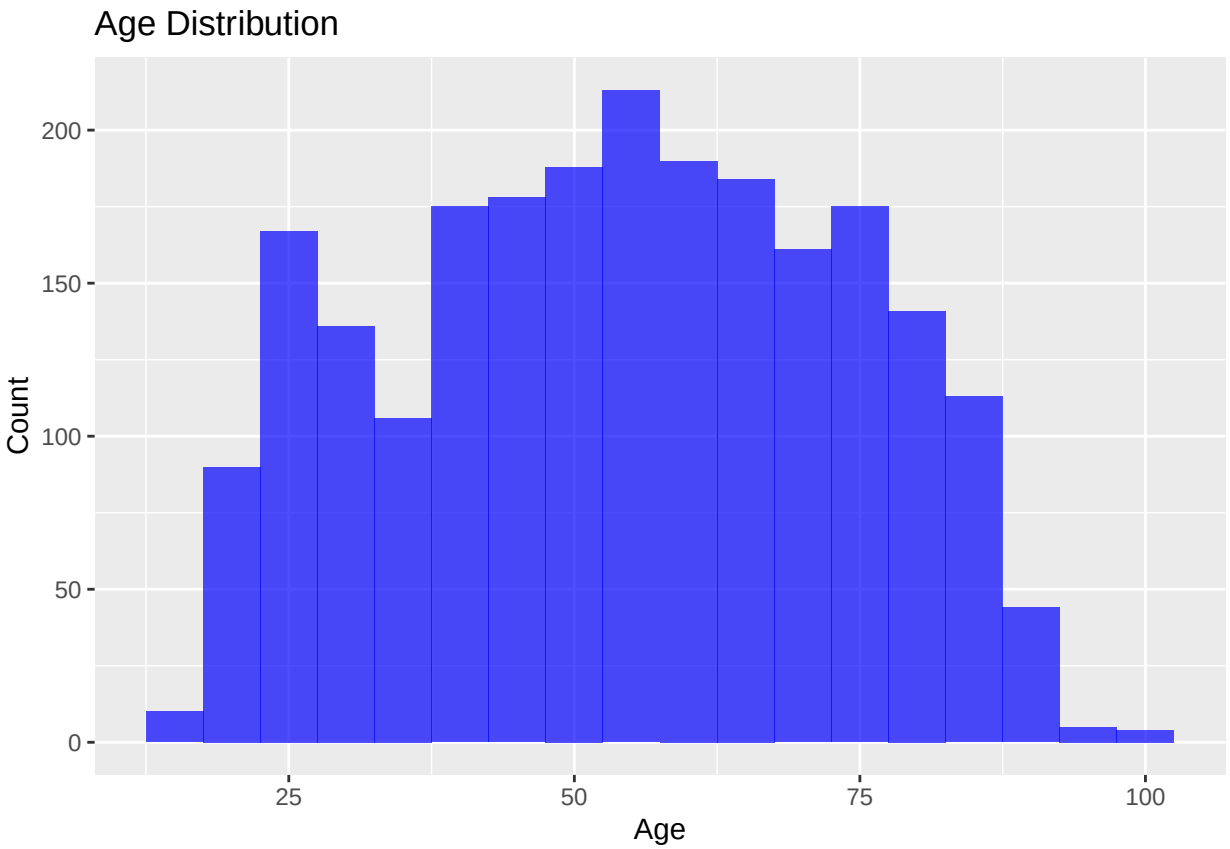
Another concern I have is about the data itself. While it contains around 2.6k observations, which is generally considered a large dataset under usual circumstances, representing China's population—especially Shandong province—could be particularly challenging due to its complexity and long history. Shandong province has 137 counties, yet the dataset is derived from only three counties. It is unclear whether these counties are sufficiently representative or diverse to reflect the entire Shandong population, let alone China as a whole. Additionally, the dataset is heavily composed of the farming population, which raises questions about the data collection process. I suspect there may be convenience bias, or the dataset appears as it does simply because farming populations were easier to access or record data from.

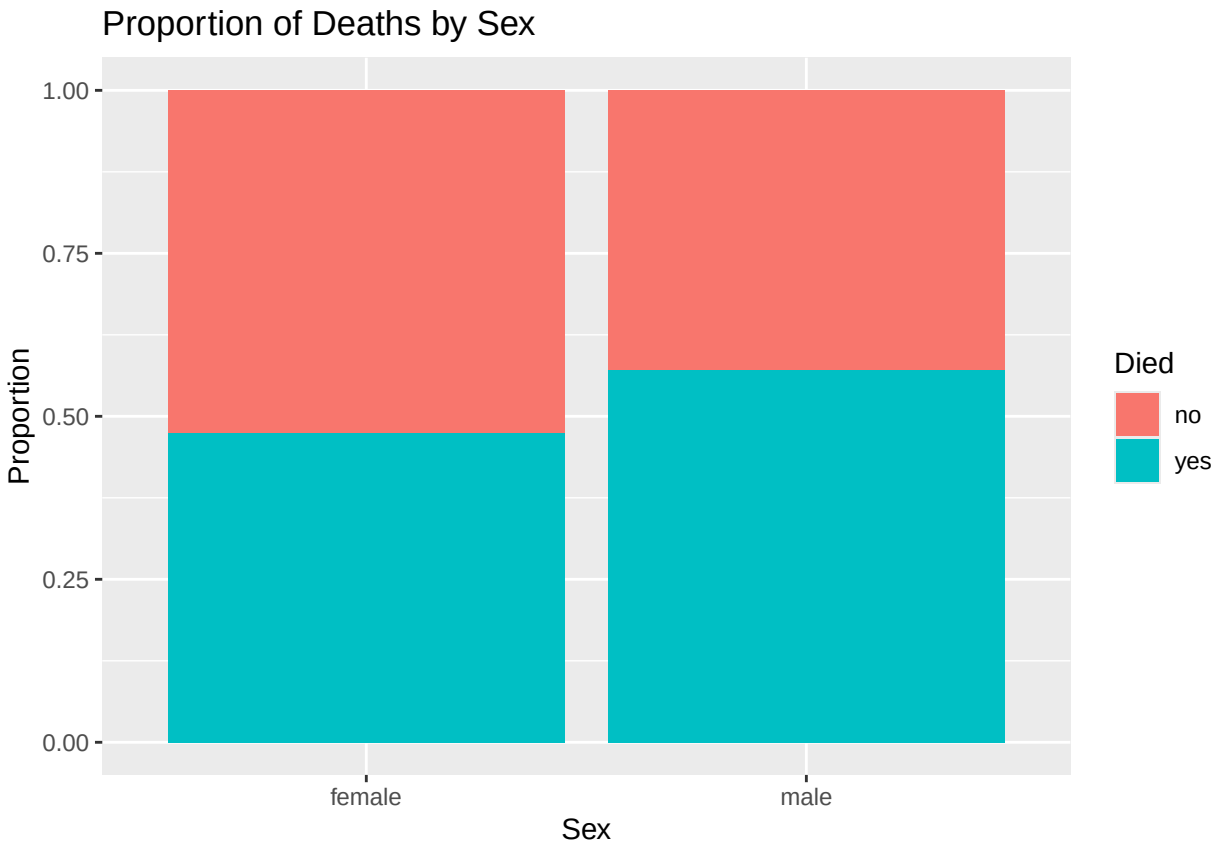
From the results, it is evident that different suicide methods have significantly varying odds of high fatality. For instance, poisoning is about four times more likely to result in high fatality compared to cutting. There may be potential collinearity between occupation and suicide methods or relationships with other demographic factors. For example, farmers are more likely to use pesticides or poisoning as suicide methods because these substances are readily accessible and familiar to them. The limited number of cases involving jumping could be attributed to the lower heights of buildings in rural areas, where high-rise structures are uncommon. If individuals jump from mountains, it is plausible that they might not be found or could be classified as missing. While hanging may initially seem less probable, traditional Chinese architecture often includes *liang* (girders), which could make this method more feasible in rural settings.

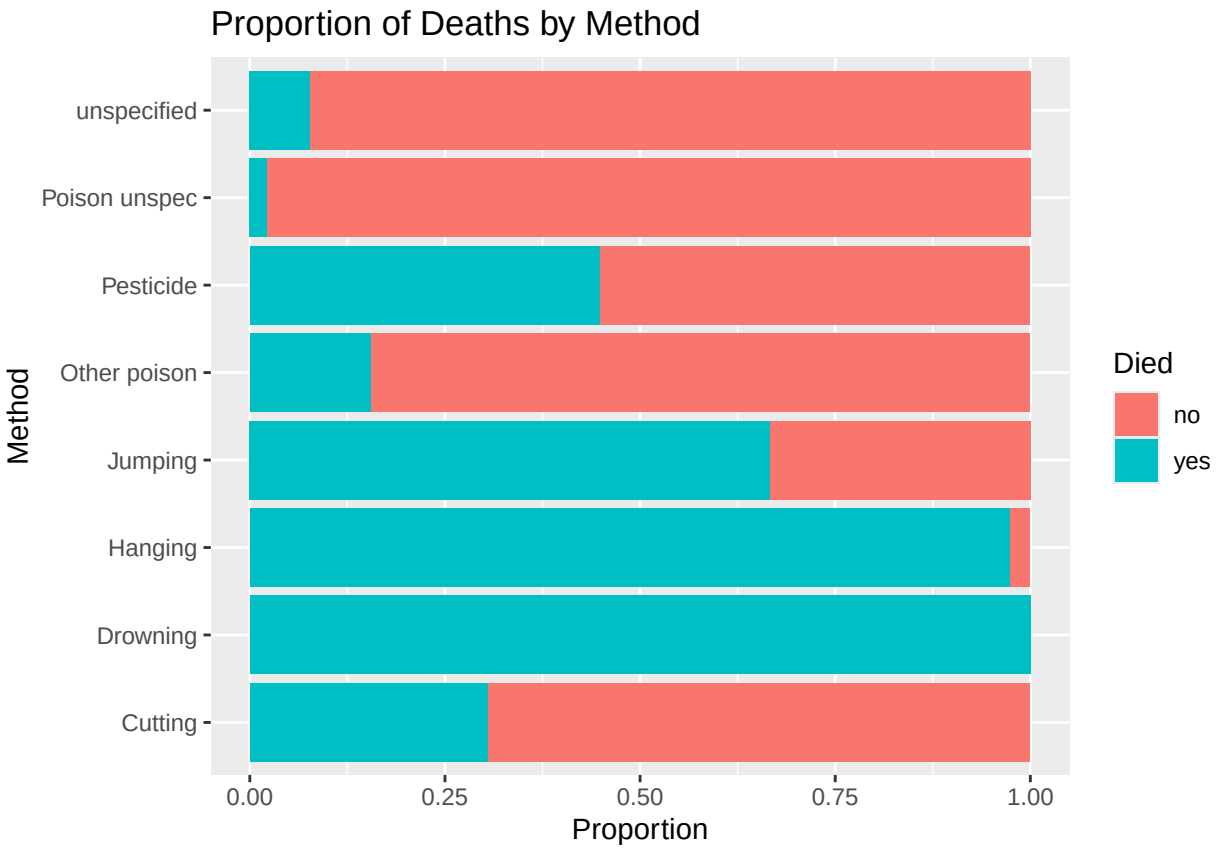
Appendix

Extra EDA

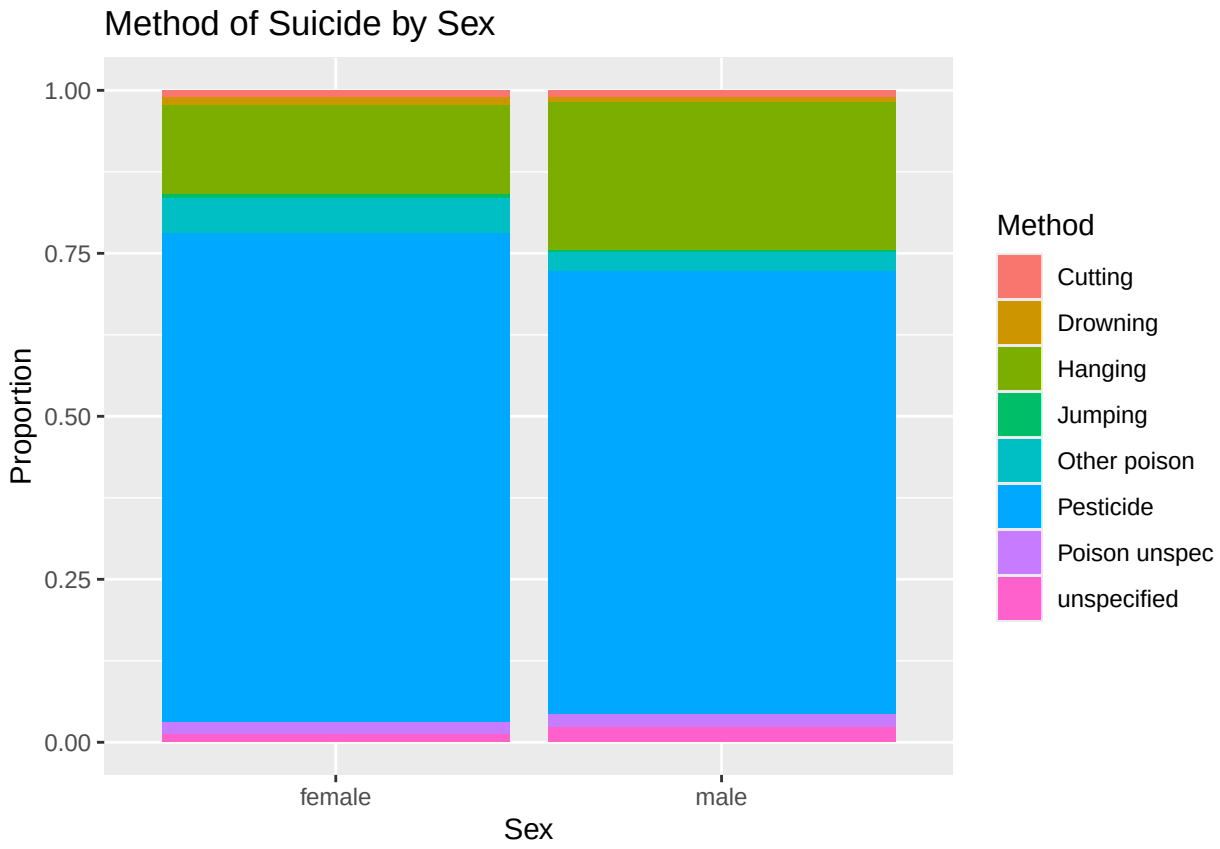
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   15.00   40.00   55.00   54.16   70.00  100.00
```

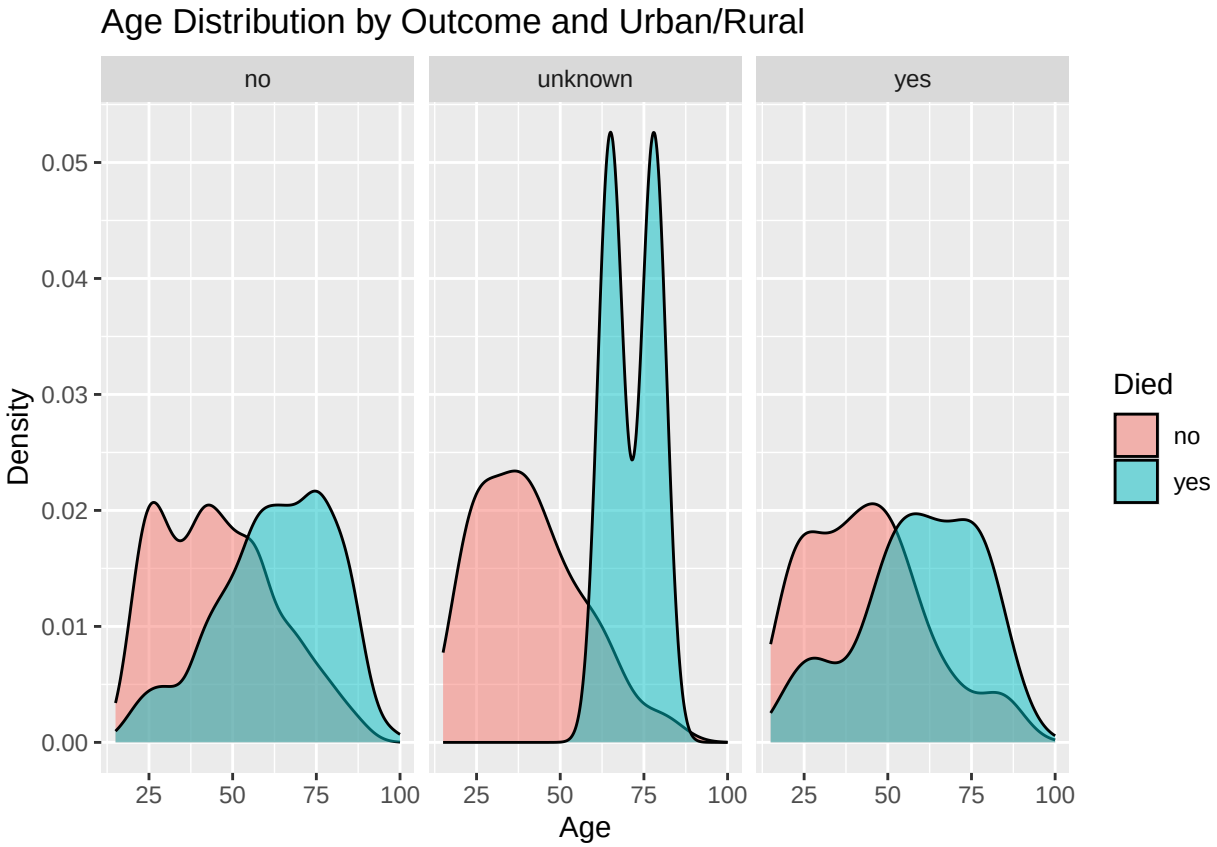

Age Distribution
Count
200
150
100
50
0
25
50
75
100
Age

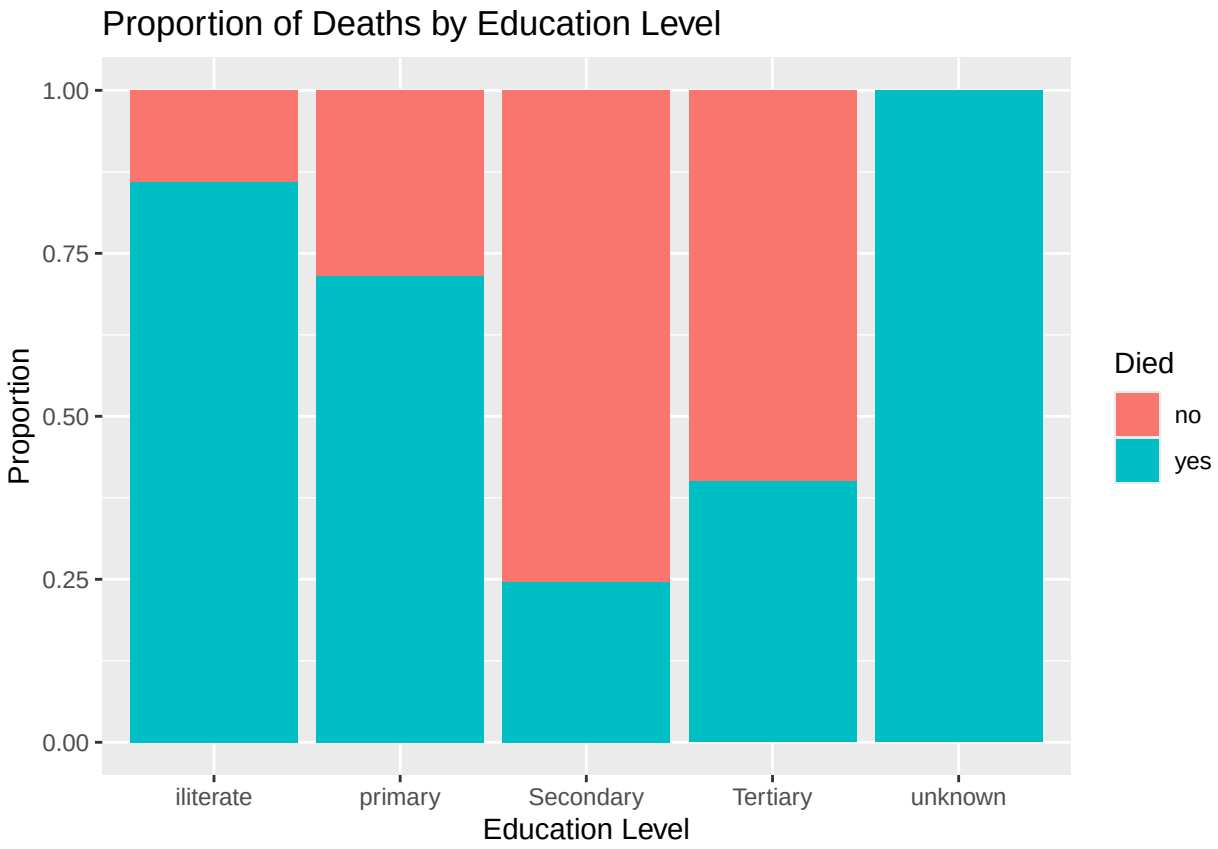

Proportion of Deaths by Sex
1.00
0.75
0.50
0.25
0.00
Proportion
female
male
Sex
Died
no
yes

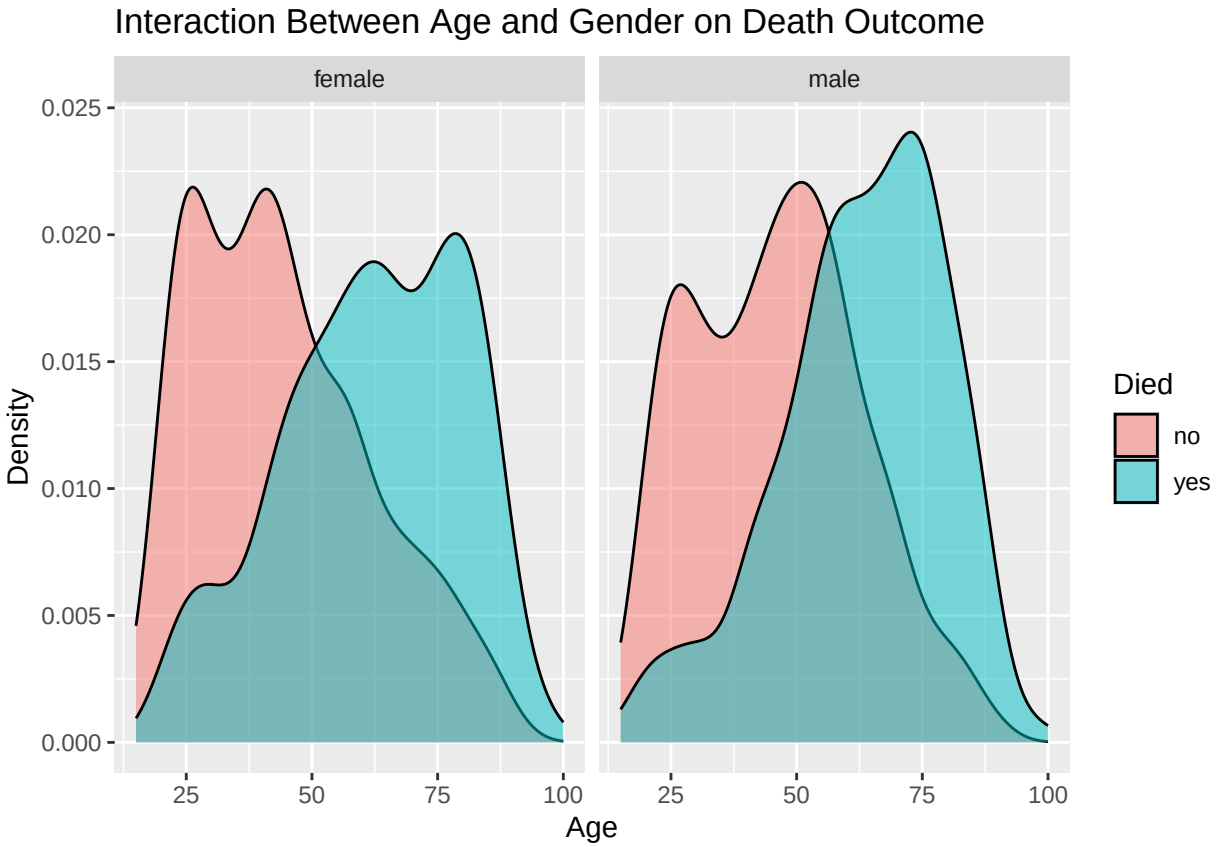

Proportion of Deaths by Method
unspecified
Poison unspec
Pesticide
Other poison
Jumping
Hanging
Drowning
Cutting
Method
0.00
0.25
0.50
0.75
1.00
Proportion
Died
no
yes

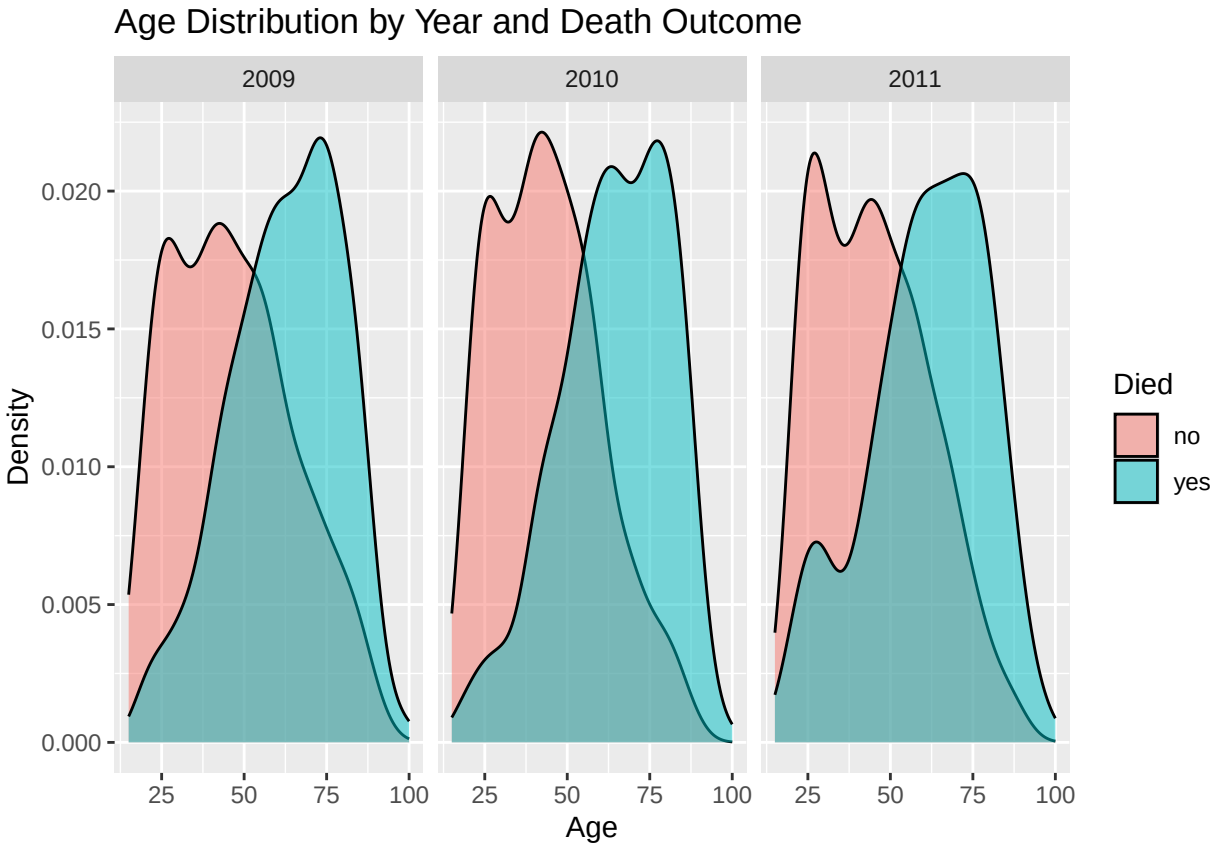

Age Distribution by Outcome
Density
0.020
0.015
0.010
0.005
0.000
25
50
75
100
Age
Died
no
yes

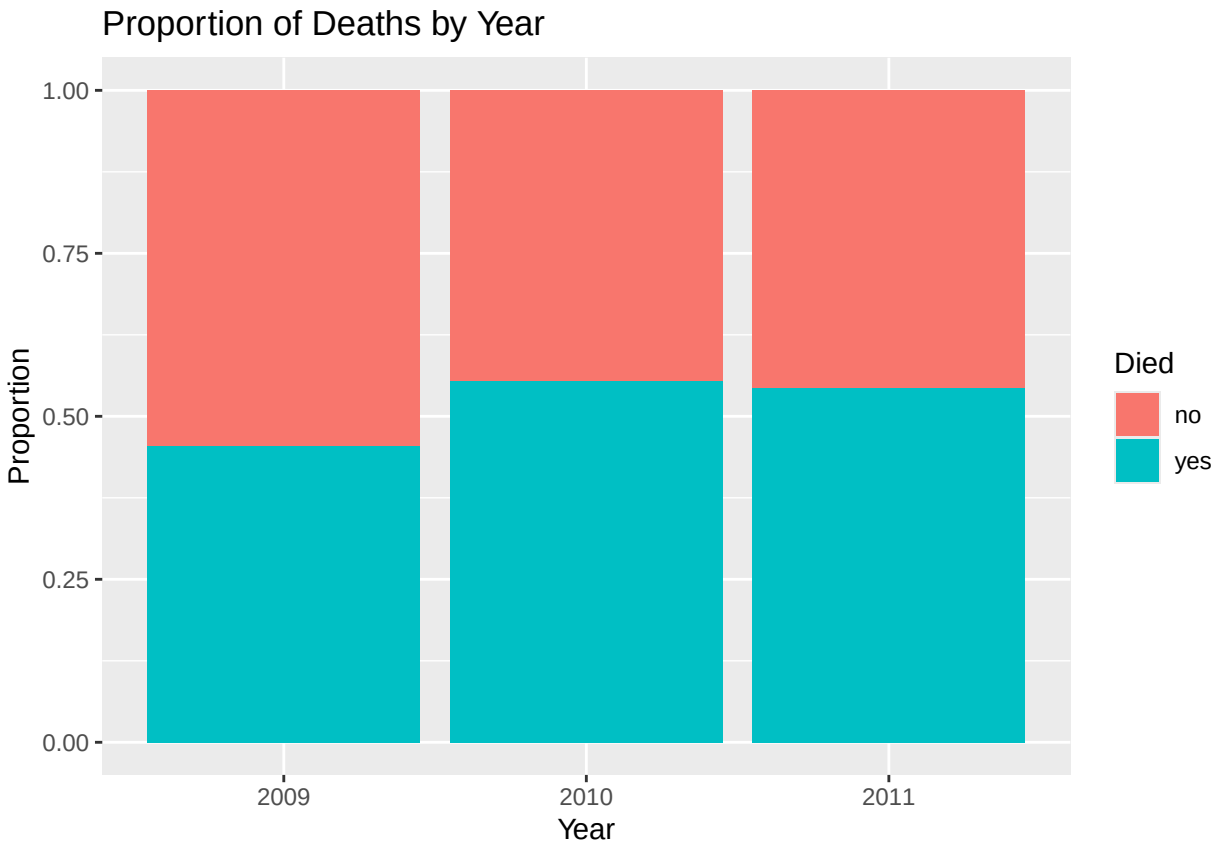

Method of Suicide by Sex
1.00
0.75
0.50
0.25
0.00
Proportion
female
male
Sex
Method
Cutting
Drowning
Hanging
Jumping
Other poison
Pesticide
Poison unspec
unspecified

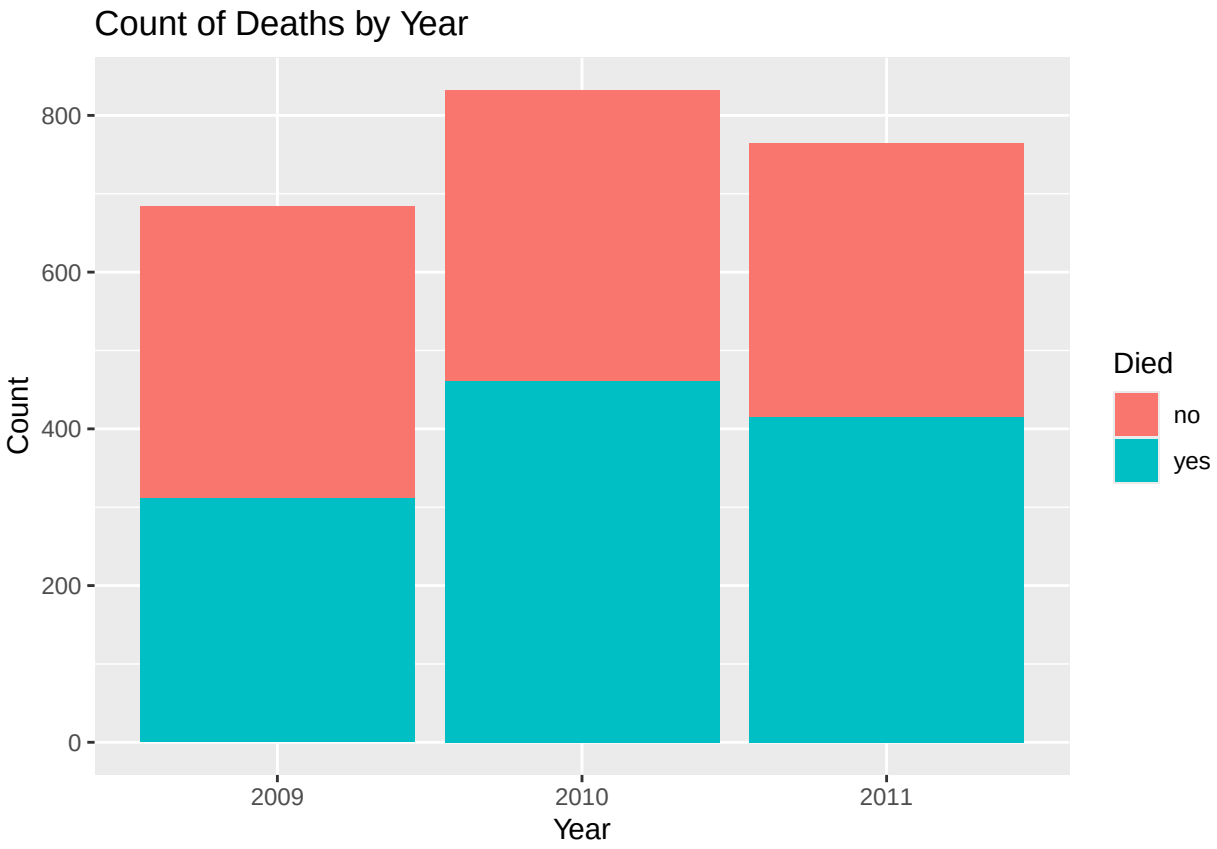

Age Distribution by Outcome and Urban/Rural
no
unknown
yes
Density
0.05
0.04
0.03
0.02
0.01
0.00
25
50
75
100
Age
Died
no
yes

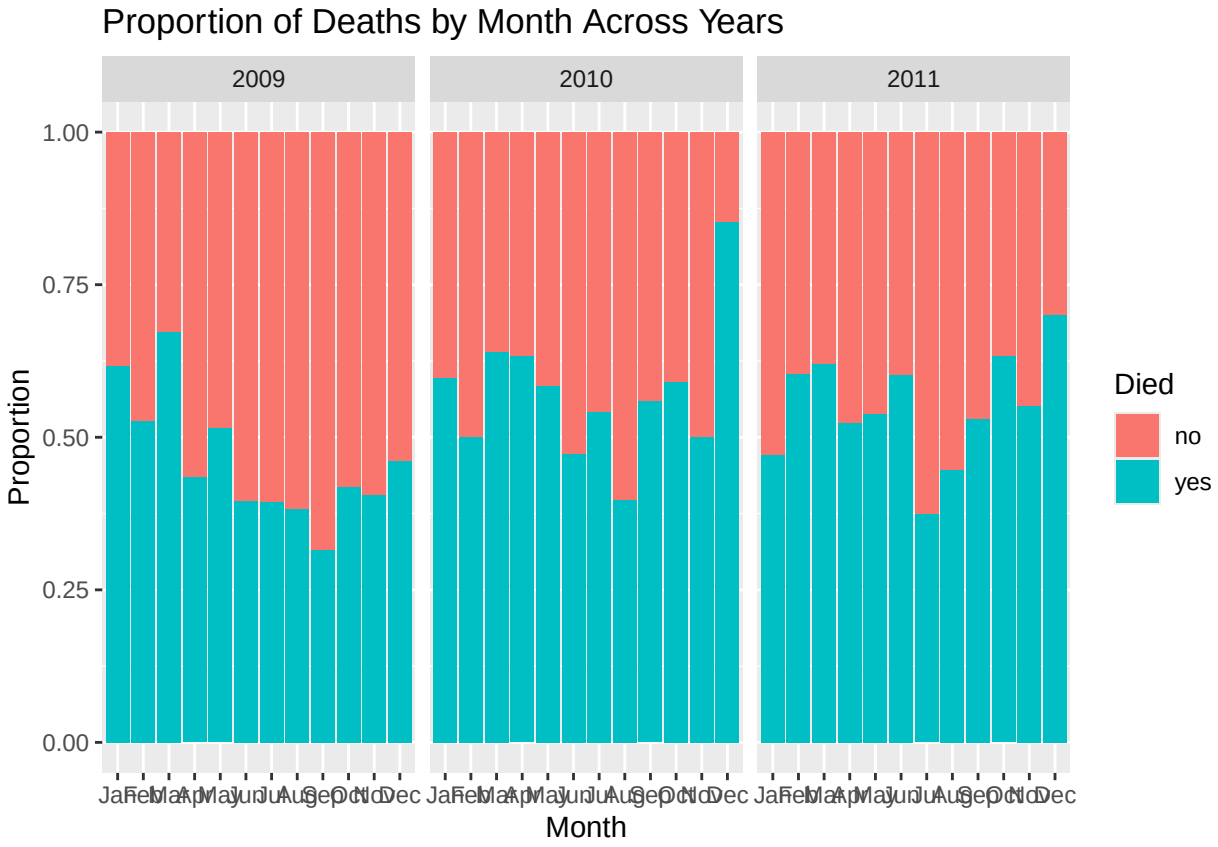

Proportion of Deaths by Education Level
1.00
0.75
0.50
0.25
0.00
Proportion
iliterate
primary
Secondary
Tertiary
unknown
Education Level
Died
no
yes

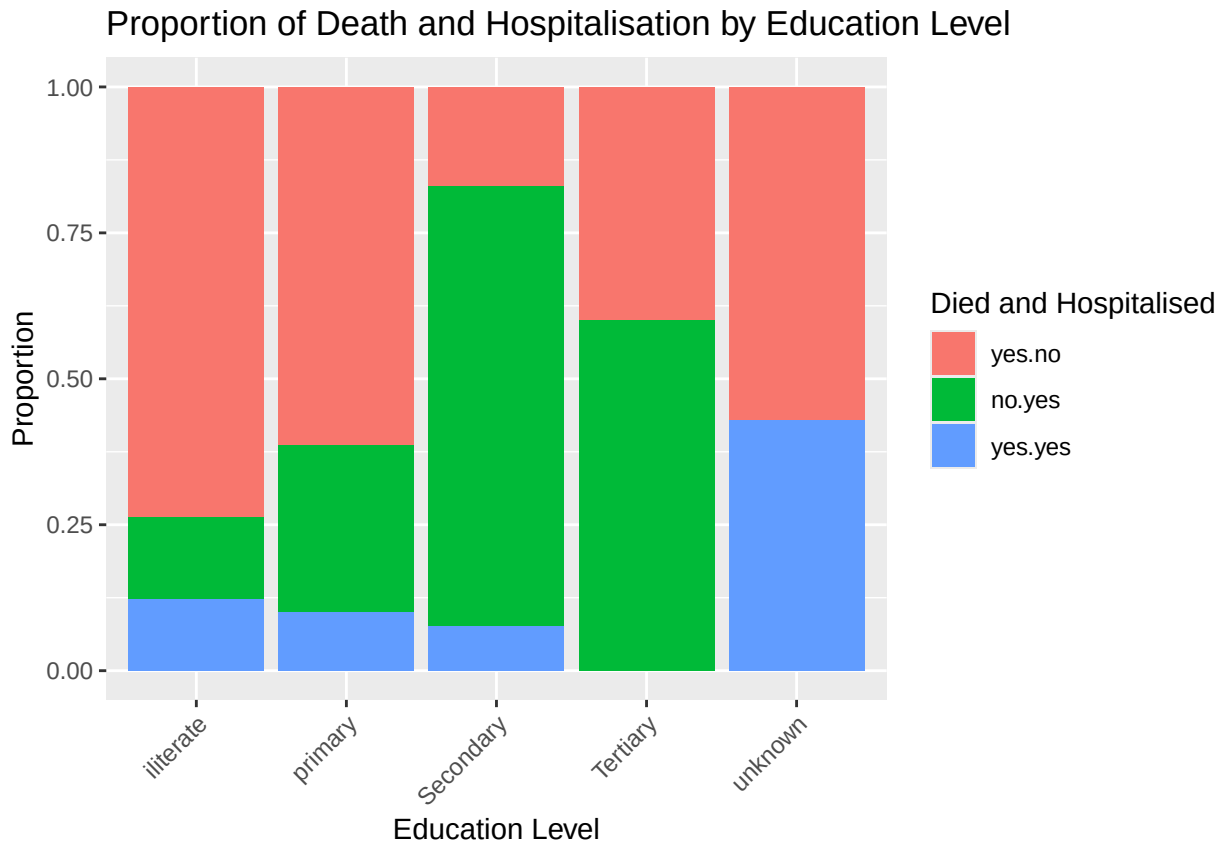

Interaction Between Age and Gender on Death Outcome
female
male
0.025
0.020
0.015
0.010
0.005
0.000
Density
25
50
75
100
Age
Died
no
yes


Age Distribution by Year and Death Outcome
2009
2010
2011
Density
0.020
0.015
0.010
0.005
0.000
25
50
75
100
Age
Died
no
yes


Proportion of Deaths by Year
1.00
0.75
0.50
0.25
0.00
Proportion
2009
2010
2011
Year
Died
no
yes


Count of Deaths by Year
800
600
400
200
0
Count
2009
2010
2011
Year
Died
no
yes


Proportion of Deaths by Month Across Years
2009
2010
2011
1.00
0.75
0.50
0.25
0.00
Proportion
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Month
Died
no
yes




```
##
## iliterate   primary Secondary  Tertiary   unknown
##       525       636      1107         5         7

##
##   1   2   3   4   5   6   7   8   9  10  11  12
## 160 188 166 184 238 257 221 203 219 189 142 113

##
##  15  16  17  18  19  20  21  22  23  24  25  26  27  28  29  30  31  32  33  34
##   1   4   5   4  13  23  26  24  37  32  38  34  26  32  26  26  23  29  22  19
##  35  36  37  38  39  40  41  42  43  44  45  46  47  48  49  50  51  52  53  54
##  26  13  26  24  31  45  33  42  48  24  32  34  40  46  32  42  33  35  31  45
##  55  56  57  58  59  60  61  62  63  64  65  66  67  68  69  70  71  72  73  74
##  37  51  49  50  37  44  33  26  41  28  44  37  34  36  29  36  27  33  29  35
##  75  76  77  78  79  80  81  82  83  84  85  86  87  88  89  90  91  92  94  95
##  43  37  31  28  27  29  29  28  26  26  23  23  15  21   6   9   2   6   1   2
##  96  97  98 100
##   1   1   3   1

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

##                   GVIF Df GVIF^(1/(2*Df))
## Hospitalised 1.997443  1        1.413309
## method       1.988658  6        1.058961
```

```
## Year         1.069229  2        1.016875
## Education    1.815731  3        1.104524
## Season       1.043236  3        1.007079
## Sex          1.028566  1        1.014182
## Age          1.488270  1        1.219947
```