

FinalProjectReport

Chang Lu

2024-11-7

Abstract

This report delves into the analysis of the U.S. Chronic Disease Indicators (CDI) – 2023 Release, a comprehensive dataset containing 1,185,676 rows and 34 columns on chronic disease metrics across the United States. The primary focus is on alcohol-related issues, particularly chronic liver disease mortality, exploring geographic, demographic, and temporal patterns through extensive exploratory data analysis (EDA), modeling, and visualization.

Introduction

For my final project, I have chosen the dataset titled “U.S. Chronic Disease Indicators (CDI) – 2023 Release.” This comprehensive dataset encompasses a significant volume of information related to various chronic diseases across the United States. While the dataset’s extensive nature presents challenges, it also offers rich opportunities for insightful analysis, which is promising for my development in data analysing skills.

The Data Source

I got access to the data(“U.S. Chronic Disease Indicators (CDI) – 2023 Release.”)([Link Text](#)) in the website of [data.gov](#)([Link text](#)). As the page described, this dataset is intended for public access and use.

```
# read the data
chronicdata <- read.csv("C:/Users/beiq1/Desktop/678Finalproject/chronicdata.csv")
dim(chronicdata)
```

```
## [1] 1185676      34
```

EDA

data cleaning

```
# remove the columns that only contain NA
chronicdata_filtered <- chronicdata %>% select_if(where(~ !all(is.na(.))))

dim(chronicdata_filtered)
```

```
## [1] 1185676      24
```

Several columns in the dataset contain only NA values, so I excluded those columns from the analysis.

```
summary(chronicdata_filtered$DataValueAlt)

# Remove rows where DataValueAlt is NA
chronicdata_filtered <- chronicdata_filtered %>%
  filter(!is.na(DataValueAlt))

summary(chronicdata_filtered$DataValueAlt)
```

Upon examining the dataset, I noticed that the “DataValueAlt” column contains a value for each valid row. Therefore, I removed rows with NA in this column to proceed with the analysis.

Alcohol issue

The original dataset encompasses a variety of chronic disease topics. For this analysis, I focused on issues related to alcohol, creating a subset of the data by filtering rows where the “Topic” is “Alcohol.”

```
# Filter dataset for the topic "Alcohol"
alcohol_data <- chronicdata_filtered %>%
  filter(Topic == "Alcohol")
```

```
# Check dimensions of the filtered dataset
dim(alcohol_data)
```

```
## [1] 42969    24
```

```
# View unique questions related to alcohol
unique(alcohol_data$Question)
```

```
## [1] "Chronic liver disease mortality"
## [2] "Alcohol use before pregnancy"
## [3] "Alcohol use among youth"
## [4] "Binge drinking prevalence among youth"
## [5] "Amount of alcohol excise tax by beverage type (beer)"
## [6] "Amount of alcohol excise tax by beverage type (wine)"
## [7] "Amount of alcohol excise tax by beverage type (distilled spirits)"
## [8] "Per capita alcohol consumption among persons aged >= 14 years"
## [9] "Heavy drinking among adults aged >= 18 years"
## [10] "Heavy drinking among women aged 18-44 years"
## [11] "Binge drinking intensity among adults aged >= 18 years who binge drink"
## [12] "Binge drinking frequency among adults aged >= 18 years who binge drink"
## [13] "Binge drinking prevalence among adults aged >= 18 years"
## [14] "Binge drinking prevalence among women aged 18-44 years"
```

```
# View unique DataValueType for alcohol data
unique(alcohol_data$DataValueType)
```

```
## [1] "Number"                "Crude Prevalence"
## [3] "Age-adjusted Rate"      "Crude Rate"
```

```
## [5] "US Dollars"                "Per capita alcohol consumption"
## [7] "Age-adjusted Prevalence"    "Mean"
## [9] "Age-adjusted Mean"
```

The “Question” column outlines various alcohol-related issues. For my analysis, I focused specifically on “Chronic liver disease mortality” as the area of interest.

```
cldm_data <- alcohol_data %>% filter(Question == "Chronic liver disease mortality") # cldm means chron
summary(cldm_data)
```

Understand Different Datatypes

The dataset includes three types of data: “Number,” “Age-adjusted Rate,” and “Crude Rate.” I began by focusing on the “Number” data to draw initial conclusions and build models to predict mortality counts.

```
unique(cldm_data$DataValueType)
```

```
## [1] "Number"                "Age-adjusted Rate" "Crude Rate"
```

```
number_data <- cldm_data %>%
  filter(DataValueType == "Number")

adjusted_data <- cldm_data %>%
  filter(DataValueType == "Age-adjusted Rate")

crude_data <- cldm_data %>%
  filter(DataValueType == "Crude Rate")
```

```
type_summary <- cldm_data %>%
  group_by(DataValueType) %>%
  summarize(
    Mean = mean(DataValueAlt, na.rm = TRUE),
    Median = median(DataValueAlt, na.rm = TRUE),
    SD = sd(DataValueAlt, na.rm = TRUE),
    Min = min(DataValueAlt, na.rm = TRUE),
    Max = max(DataValueAlt, na.rm = TRUE),
    Count = n()
  )

print(type_summary)
```

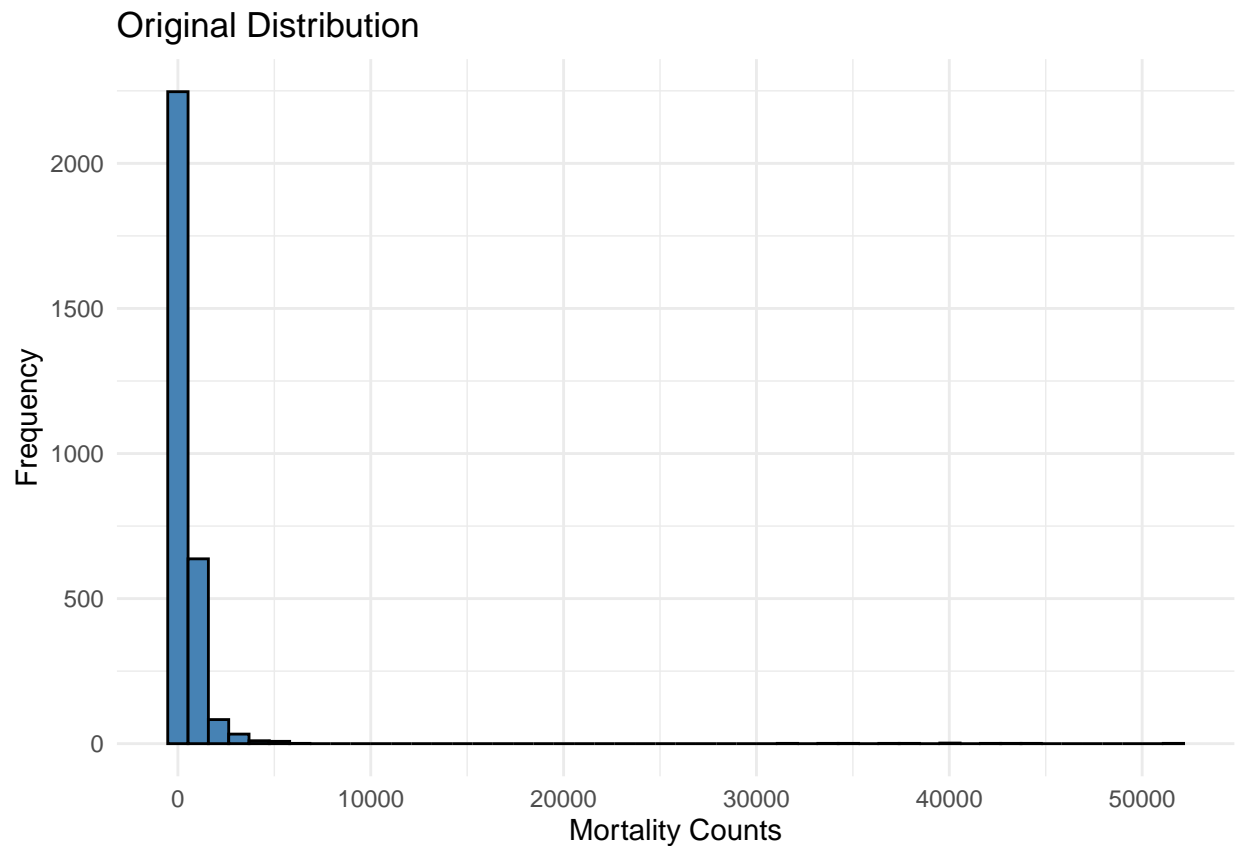
```
## # A tibble: 3 x 7
##   DataValueType      Mean Median      SD   Min    Max Count
##   <chr>          <dbl> <dbl> <dbl> <dbl> <dbl> <int>
## 1 Age-adjusted Rate  13.0  10.4  13.4   2.1  291.  3030
## 2 Crude Rate        13.8  11.6  11.5    2   264.  3030
## 3 Number           573.  220  2462.   20 51642  3030
```

Number data analysis

```
# Log-transform the DataValue to reduce skewness
number_data <- number_data %>%
  mutate(LogDataValueAlt = log10(DataValueAlt + 1)) # Adding 1 to avoid log(0)

# Check summary statistics for the log-transformed data
summary(number_data$LogDataValueAlt)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  1.322   1.914   2.344   2.335   2.737   4.713
```



As shown in Figure 1, the distribution is highly skewed. To address this, I applied a logarithmic transformation to the “DataValueAlt” column and plotted the transformed data to observe the new distribution.

```
##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##      combine
```

As is shown in figure 3, the log-transformed mortality counts exhibit a nearly normal distribution. This transformation makes it easier to analyze and model the data, particularly if working with highly skewed data in its raw form.

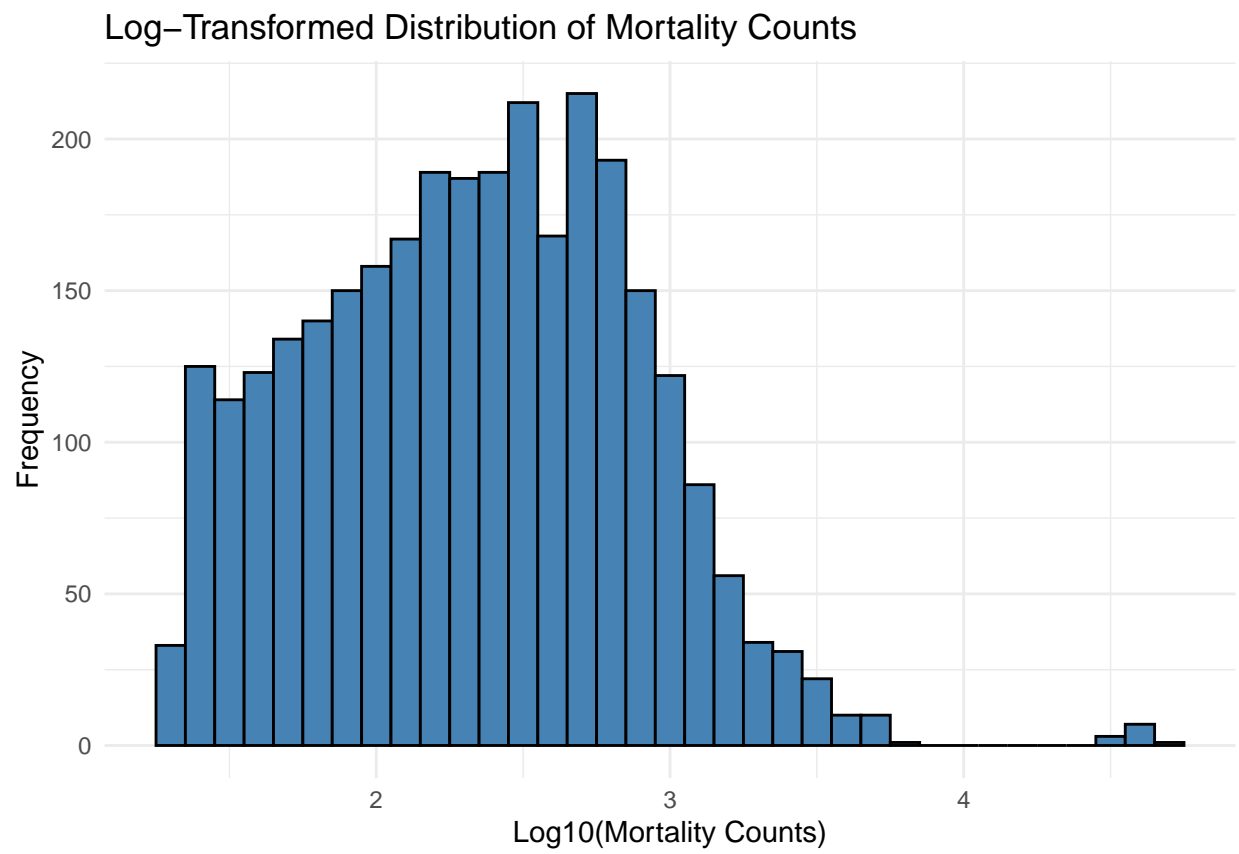


Figure 1: Log-Transformed Distribution of Mortality Counts

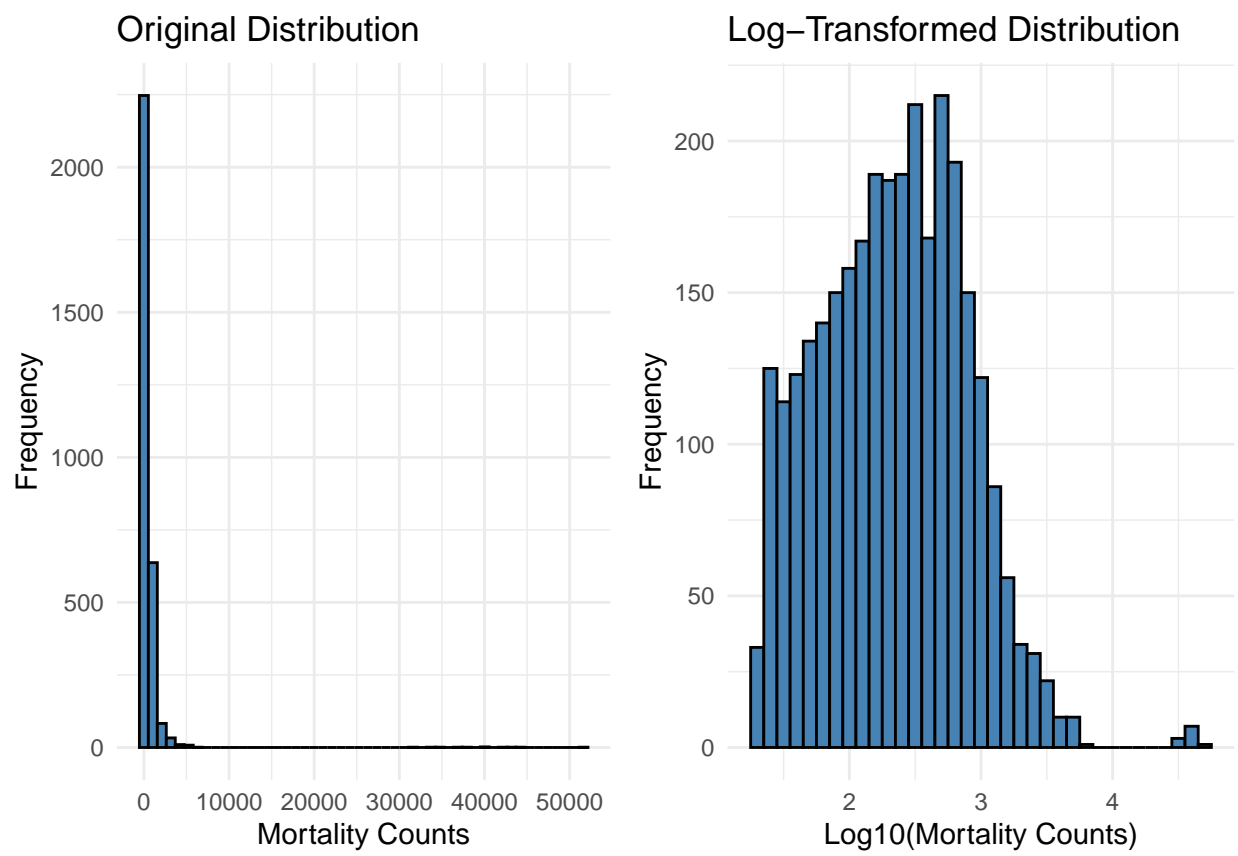


Figure 2: Comparison of the Two Plots

location Analysis

Firstly, I used “LocationDesc” and the transformed “LogDataValueAlt” to analyze variations in mean mortality counts across different states and regions in the USA.

| ## | State | MeanLog10 | Rank |
|-------|----------------|-----------|------|
| ## 1 | California | 2.983876 | 1 |
| ## 2 | Texas | 2.929862 | 2 |
| ## 3 | Florida | 2.872766 | 3 |
| ## 4 | Ohio | 2.609235 | 4 |
| ## 5 | Tennessee | 2.609161 | 5 |
| ## 6 | Illinois | 2.602255 | 6 |
| ## 7 | New York | 2.600771 | 7 |
| ## 8 | Pennsylvania | 2.588096 | 8 |
| ## 9 | Michigan | 2.578160 | 9 |
| ## 10 | Arizona | 2.568671 | 10 |
| ## 11 | North Carolina | 2.564395 | 11 |
| ## 12 | Georgia | 2.535597 | 12 |
| ## 13 | Alabama | 2.533876 | 13 |
| ## 14 | South Carolina | 2.518685 | 14 |
| ## 15 | Virginia | 2.446438 | 15 |
| ## 16 | Colorado | 2.419550 | 16 |
| ## 17 | Missouri | 2.418375 | 17 |
| ## 18 | Kentucky | 2.416092 | 18 |
| ## 19 | New Jersey | 2.401072 | 19 |
| ## 20 | Louisiana | 2.394003 | 20 |
| ## 21 | Indiana | 2.393380 | 21 |
| ## 22 | New Mexico | 2.367382 | 22 |
| ## 23 | West Virginia | 2.338112 | 23 |
| ## 24 | Iowa | 2.335715 | 24 |
| ## 25 | Massachusetts | 2.322349 | 25 |
| ## 26 | Maryland | 2.317300 | 26 |
| ## 27 | Oregon | 2.292387 | 27 |
| ## 28 | Mississippi | 2.291833 | 28 |
| ## 29 | Washington | 2.276119 | 29 |
| ## 30 | Minnesota | 2.259431 | 30 |
| ## 31 | Arkansas | 2.250638 | 31 |
| ## 32 | Wisconsin | 2.224935 | 32 |
| ## 33 | Oklahoma | 2.224410 | 33 |
| ## 34 | Kansas | 2.204795 | 34 |
| ## 35 | Nevada | 2.193624 | 35 |
| ## 36 | Connecticut | 2.140123 | 36 |
| ## 37 | Idaho | 2.138109 | 37 |
| ## 38 | Maine | 2.095670 | 38 |
| ## 39 | Nebraska | 2.088648 | 39 |
| ## 40 | New Hampshire | 2.044730 | 40 |
| ## 41 | Utah | 2.023016 | 41 |
| ## 42 | Rhode Island | 1.983369 | 42 |
| ## 43 | Montana | 1.943366 | 43 |
| ## 44 | South Dakota | 1.897582 | 44 |
| ## 45 | Wyoming | 1.837564 | 45 |
| ## 46 | Hawaii | 1.831779 | 46 |
| ## 47 | Delaware | 1.829108 | 47 |
| ## 48 | Alaska | 1.765589 | 48 |

| | | | |
|-------|----------------------|----------|----|
| ## 49 | North Dakota | 1.731846 | 49 |
| ## 50 | Vermont | 1.689300 | 50 |
| ## 51 | District of Columbia | 1.589086 | 51 |



Figure 3: Mean Log10 Mortality Count Rank by State

Secondly, we can defer several results from figure 4:

1. States with High Mean Mortality (Log10 Scale):

States like California, Florida, and Texas have the highest mean log-transformed mortality values, as seen from their positions at the bottom of the chart with the largest bars. This suggests that these states, on average, experience higher mortality counts (in original scale) compared to others.

2. States with Low Mean Mortality (Log10 Scale):

States like District of Columbia, Vermont, and North Dakota have the lowest mean log-transformed mortality values, indicating relatively lower average mortality counts.

3. Possible Drivers of Variation:

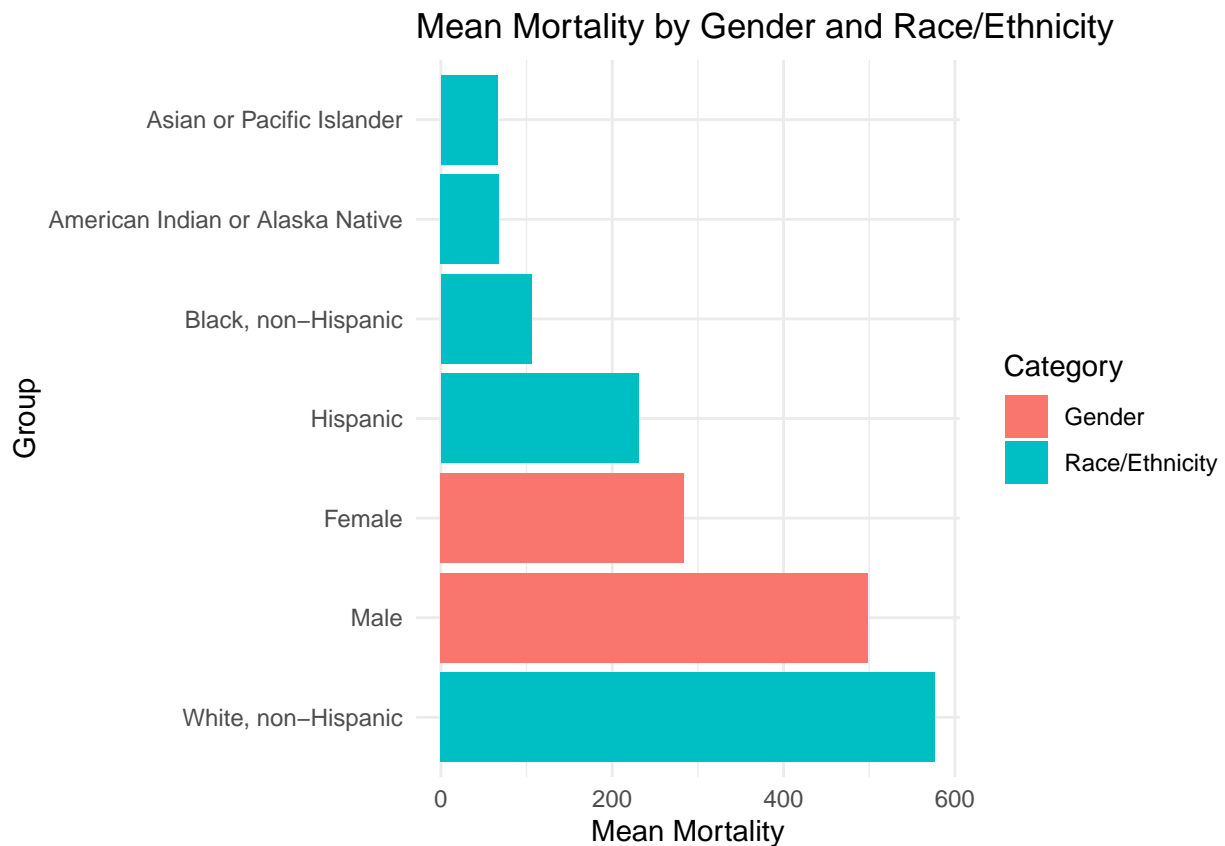
Higher mortality counts in populous states like California and Texas might reflect their larger populations or other factors such as age distribution, healthcare access, or environmental factors and low mortality counts in smaller states like Vermont and North Dakota might be due to their smaller populations or other demographic characteristics.

Gender and Race Analysis

In the next step, I divided the “Stratification1” and “StratificationCategory1” columns to extract additional predictors, such as gender and race, for further analysis.

```
## # A tibble: 2 x 4
##   Gender MeanValue MaxValue Count
##   <chr>      <dbl>    <dbl> <int>
## 1 Female      284.    2221  554
## 2 Male       498.    3959  560

## # A tibble: 5 x 4
##   Race                               MeanValue MaxValue Count
##   <chr>                               <dbl>    <dbl> <int>
## 1 American Indian or Alaska Native    67.9      263   132
## 2 Asian or Pacific Islander           66.0      345    75
## 3 Black, non-Hispanic                 106.      357   314
## 4 Hispanic                           230.     2489   273
## 5 White, non-Hispanic                 577.     2928   550
```



According to the figure 5, we can defer that:

1. Gender-Based Mortality Differences:

Males have a higher mean mortality rate compared to females. This suggests that mortality is disproportionately higher for males in the population under study.

2. Race/Ethnicity-Based Mortality Differences:

The highest mean mortality rate is observed for the White, non-Hispanic group, which significantly exceeds other racial/ethnic groups.

Model Part

```
# Build a linear model for mortality count prediction (Complete Pooling)
complete_pooling_model <- lm(DataValueAlt ~ LocationDesc + Stratification1, data = model_data)

# Build a mixed-effects model for partial pooling
partial_pooling_model <- lmer(DataValueAlt ~ (1 | LocationDesc) + (1 | Stratification1),
                             data = model_data)

# Create an interaction term for Stratification and Location
model_data$Group <- interaction(model_data$LocationDesc, model_data$Stratification1)

# Fit a no pooling model using the interaction term
no_pooling_model <- lm(DataValueAlt ~ Group, data = model_data)
```

Data Preparation

Complete Pooling The complete pooling model assumes that all groups share the same underlying structure, with variations captured through location, demographic stratifications, and binary gender variables. Here is the detailed interpretation:

1. Overall Model Fit Residual Standard Error (RSE): 471.3 The typical deviation of the observed mortality counts from the predicted counts is approximately 471.3 units.

$R^2=0.9641$, Adjusted $R^2=0.9634$ which means the model explains 96.41% of the variance in mortality counts, indicating a strong fit. The adjusted R^2 suggests the model remains robust despite having many predictors.

2. Coefficients

Each coefficient represents the expected change in DataValueAlt (mortality count) for a one-unit increase in the predictor, holding all other predictors constant.

2.1 Intercept (23.389):

The baseline mortality count for the reference group (LocationDesc = Alabama and Stratification1 = Asian or Pacific Islander) when gender is not considered.

2.2 LocationDesc (e.g., California = 1730.417):

The estimated increase or decrease in mortality count for each state compared to the reference state (Alabama). There are some significant States ($p < 0.05$): Examples include California (+1730.417), Texas (+1398.532), and New York(+454.623).

California shows a large positive impact, suggesting higher mortality counts.

2.3 Stratification1 (e.g., Female = 273.802):

The difference in mortality count based on demographic stratifications compared to the reference group (Asian or Pacific Islander): Female (+273.802): Being female is associated with an increase in mortality count compared to the reference group.

Male (+492.281): Being male is associated with a larger increase in mortality count compared to the reference.

White, non-Hispanic (+564.773): Shows significantly higher mortality compared to the reference demographic.

Partial Pooling The partial pooling model accounts for group-level variations in LocationDesc (locations) and Stratification1 (demographics), while estimating global fixed effects for IsFemale and IsMale. Here is the detailed explanation:

1. Overall Model Fit

Residual Standard Error (RSE): 471.3: The typical deviation of the observed mortality counts from the predicted counts is approximately 471.3 units.

$R^2=0.9641$, Adjusted $R^2=0.9634$: The model explains 96.41% of the variance in mortality counts, indicating a strong fit. The adjusted R^2 suggests the model remains robust despite having many predictors.

2. Coefficients

Each coefficient represents the expected change in DataValueAlt (mortality count) for a one-unit increase in the predictor, holding all other predictors constant.

2.1 Intercept (23.389):

The baseline mortality count for the reference group (LocationDesc = Alabama and Stratification1 = Asian or Pacific Islander)

2.2 LocationDesc

The estimated increase or decrease in mortality count for each state compared to the reference state (Alabama). Significant States ($p < 0.05$): Examples include California (+1730.417), Texas (+1398.532), and New York (+454.623).

2.3 Stratification

The difference in mortality count based on demographic stratifications compared to the reference group (Asian or Pacific Islander): Female (+273.802): Being female is associated with an increase in mortality count compared to the reference group.

Male (+492.281): Being male is associated with a larger increase in mortality count compared to the reference.

White, non-Hispanic (+564.773): Shows significantly higher mortality compared to the reference demographic.

No Pooling

```
poisson_model <- glm(DataValueAlt ~ LocationDesc + Stratification1,
                      family = poisson(link = "log"),
                      data = model_data)
summary(poisson_model)
```

```
# Calculate dispersion statistic
dispersion_stat <- summary(poisson_model)$deviance / summary(poisson_model)$df.residual
dispersion_stat
```

Poisson Model

```
## [1] 38.04049
```

A dispersion statistic of 38 is a clear indication of severe overdispersion in the Poisson regression model. This means that the variance of the dependent variable (DataValueAlt) is much larger than the mean, violating the Poisson assumption that variance equals the mean. So it is necessary for a change in model.

```
# quasi_poisson model
quasi_poisson_model <- glm(DataValueAlt ~ LocationDesc + Stratification1,
                           family = quasipoisson(link = "log"),
                           data = model_data)
summary(quasi_poisson_model)
```

```
# negative binomial model
neg_binomial_model <- glm.nb(DataValueAlt ~ LocationDesc + Stratification1,
                             data = model_data)
summary(neg_binomial_model)
```

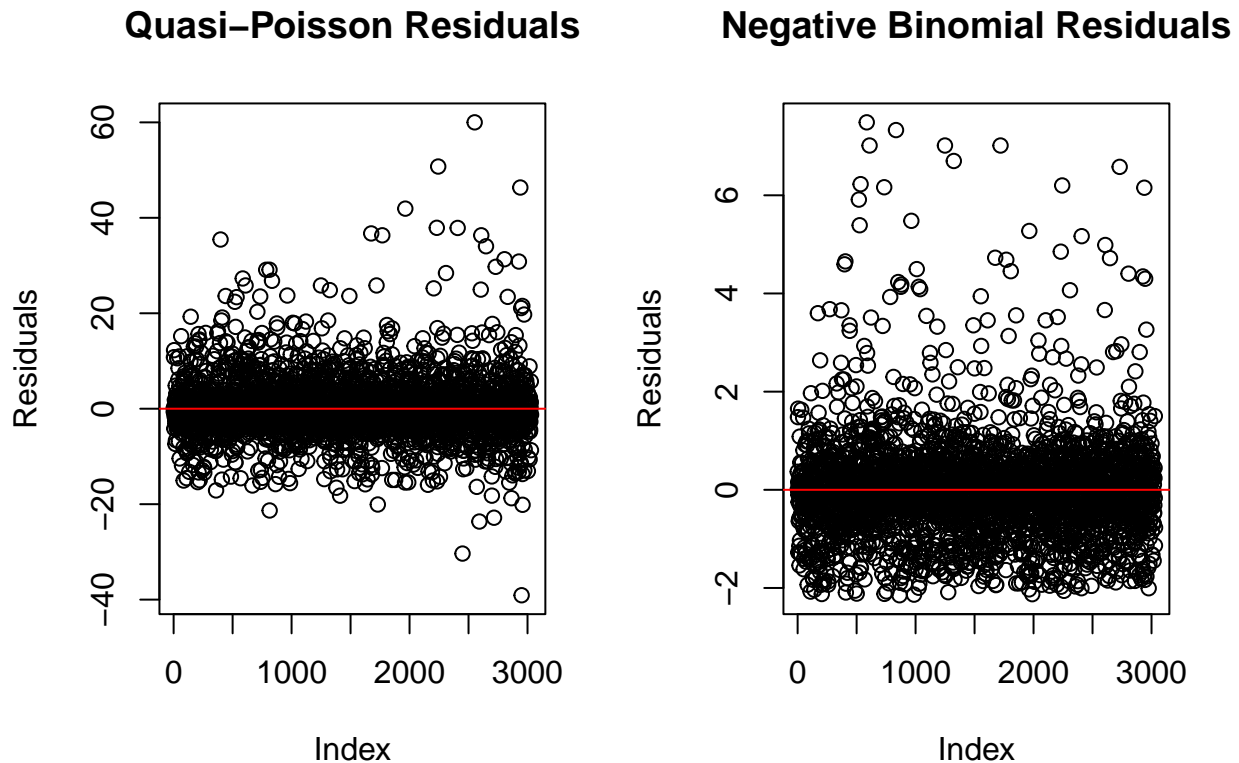
```
# Compare AIC
AIC(poisson_model, neg_binomial_model)
```

```
##              df          AIC
## poisson_model    59 134969.56
## neg_binomial_model 60  36471.35
```

The result indicates that using negative binomial model is much better than original poisson model.

```
# Extract Pearson residuals
quasi_residuals <- residuals(quasi_poisson_model, type = "pearson")
neg_bin_residuals <- residuals(neg_binomial_model, type = "pearson")

# Plot residuals
par(mfrow = c(1, 2))
plot(quasi_residuals, main = "Quasi-Poisson Residuals", ylab = "Residuals")
abline(h = 0, col = "red")
plot(neg_bin_residuals, main = "Negative Binomial Residuals", ylab = "Residuals")
abline(h = 0, col = "red")
```



According to the residuals plots, the negative binomial model is better than quasi-poisson model for it has fewer outliers and a more even distribution around 0 (negative binomial: -2~6, quasi-poisson: -40~60)

```
AIC(complete_pooling_model, partial_pooling_model, no_pooling_model, neg_binomial_model)
```

Comparison of the models

| ## | | df | AIC |
|----|------------------------|-----|----------|
| ## | complete_pooling_model | 60 | 45961.07 |
| ## | partial_pooling_model | 4 | 46394.58 |
| ## | no_pooling_model | 300 | 44426.70 |
| ## | neg_binomial_model | 60 | 36471.35 |

1. Interpretation

1.1 Complete Pooling (AIC = 45961.07):

The AIC is relatively high, suggesting this model has the weakest fit compared to the others. This is expected because the complete pooling model oversimplifies by assuming all groups have the same structure, ignoring group-level differences.

1.2 Partial Pooling (AIC = 46394.58):

The AIC is higher than the complete pooling model, indicating a slightly worse trade-off between fit and complexity. However, the partial pooling model may still be preferred in hierarchical data when overfitting is a concern or when interpretability and generalizability are priorities.

1.3 No Pooling (AIC = 44426.70):

This model has the lowest AIC, suggesting the best fit to the data among the three models. However, no pooling models can overfit because they do not share information between groups, and their complexity is reflected in the high degrees of freedom (300).

1.4 Negative Binomial Model (AIC = 36471.35)

This is the best model due to its smallest AIC and an appropriate degrees of freedom.

2. Model Selection

The Negative Binomial model is the best choice due to its:

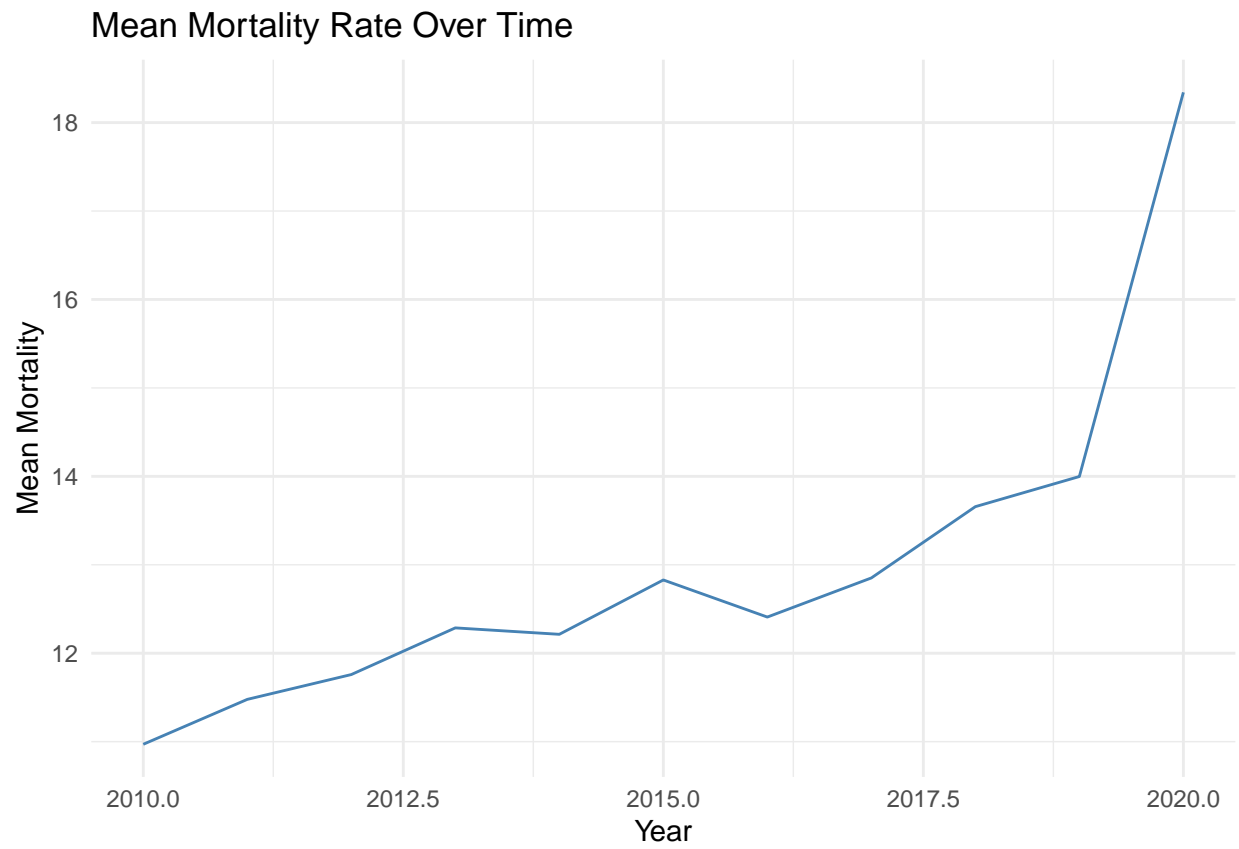
2.1 Lowest AIC, indicating the best balance between fit and complexity.

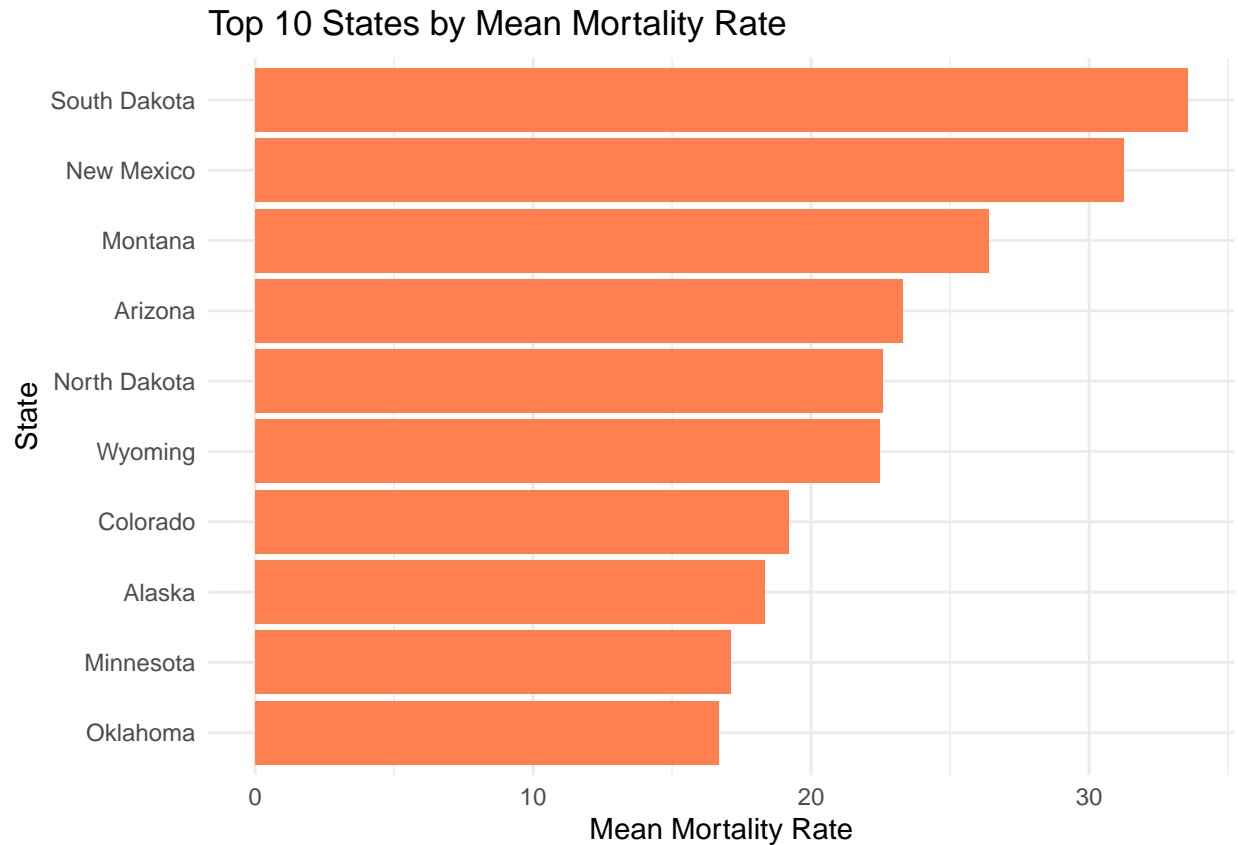
2.2 Ability to explicitly model overdispersion, which is a key issue in the dataset.

Rate Data Analysis

Basic Visualization

```
ggplot(adjusted_data, aes(x = YearStart, y = DataValueAlt)) +  
  geom_line(stat = "summary", fun = "mean", color = "steelblue") +  
  labs(  
    title = "Mean Mortality Rate Over Time",  
    x = "Year",  
    y = "Mean Mortality"  
  ) +  
  theme_minimal()
```





According to figure 6, South Dakota has the highest mean mortality rate among the top 10 states (Over 30). New Mexico and Montana follow closely behind, indicating significant health-related challenges in these states.

According to figure 7, Virginia, Connecticut, and Vermont have the lowest mean mortality rates among all states in the dataset, suggesting relatively better health outcomes or lower incidence of the measured event.

Model Part

Data Preparation

Linear Model

Logistic Model Logistic regression is only applicable if the outcome have a binary version of the mortality rate (e.g., “high” vs. “low”). So I choose the median of the mortality rate as the threshold.

```
# Ensure the dataset is clean
adjusted_data <- adjusted_data %>%
  filter(!is.na(DataValueAlt)) # Remove missing values

# Create a binary target
adjusted_data$HighMortality <- ifelse(adjusted_data$DataValueAlt > median(adjusted_data$DataValueAlt, na.rm=T), 1, 0)

# Check the balance of the target variable
print(table(adjusted_data$HighMortality))
```




Figure 4: Bottom 10 States by Mean Mortality

```

# Split into training and testing sets
set.seed(123)
train_index <- createDataPartition(adjusted_data$HighMortality, p = 0.8, list = FALSE)
train_data <- adjusted_data[train_index, ]
test_data <- adjusted_data[-train_index, ]

# Logistic Regression
logistic_model <- glm(HighMortality ~ LocationDesc + Stratification1, family = binomial, data = train_data)

# Summary of the model
summary(logistic_model)

# Predictions on test data
logistic_predictions <- predict(logistic_model, newdata = test_data, type = "response")
logistic_class <- ifelse(logistic_predictions > 0.5, 1, 0)

confusionMatrix(as.factor(logistic_class), as.factor(test_data$HighMortality))

```

Interpretation of the Confusion Matrix:

The model achieves high accuracy (84.32%) with balanced sensitivity (84.28%) and specificity (84.36%).

High precision values for both classes indicate that the model makes reliable predictions for both class “0” and class “1.”

Poisson Model

```
## Dispersion Parameter: 0.8883045
```

The overdispersion coefficient is 0.8883045, indicating no need for change of poisson model.

```
AIC(linear_model, logistic_model, poisson_model)
```

Comparison of the Models

```
##           df      AIC
## linear_model  60 20691.262
## logistic_model 59  1696.282
## poisson_model  59 15442.640
```

The logistic model is the best of the three models.

Conclusion

Results Summary

The analysis revealed significant geographic and demographic disparities in chronic liver disease mortality across the United States. States such as South Dakota, **New Mexico, and Montana exhibited the highest

mean mortality rates, while Vermont, the District of Columbia, and Hawaii consistently reported the lowest. Demographic analysis highlighted that males and White, non-Hispanic populations had notably higher mortality rates compared to other groups.

Models Summary

Modeling efforts demonstrated that negative binomial regression effectively addressed overdispersion in the data, with an AIC of 36,471.35, outperforming the Poisson model. Logistic regression achieved an accuracy of 84.32% in classifying high and low mortality rates, with balanced sensitivity and specificity. These findings underscore the importance of tailored interventions to address disparities, particularly in high-risk states and demographic groups, emphasizing the need for public health programs targeting alcohol-related behaviors and chronic disease prevention.

Validation

1. Geographic Disparities in Liver Disease Mortality

1.1 “Geographic Variability in Liver Disease-Related Mortality Rates in the United States”

“While chronic liver disease is the 12th leading cause of death in all Americans, it is the fourth leading cause of death in those 45-54 years of age and the sixth leading cause of death in Hispanic Americans. In the United States, liver disease mortality has been attributed to individual characteristics such as ethnicity, race, obesity, and alcohol consumption.”

“Figure 1 and Table 1 show significant variability in age-adjusted liver disease mortality at a state level. Age-adjusted liver disease mortality ranges from 6.4 to 17.0 per 100,000. In the northeastern United States, rates of age-adjusted liver disease mortality are the lowest in the country. New Hampshire and New York have the lowest rates in the United States (6.4 and 6.6/100,000, respectively). In contradiction to this general assumption, West Virginia’s rate is in the highest quartile, with a rate of 10.7/100,000. States in the west and central southwest carry some of the highest liver disease mortality rates, with New Mexico reporting the highest liver disease mortality, 17.0/100,000. The southern state of Georgia does not fit this assumption and falls into the lowest quartile, with a rate of 7.5/100,000. State size did not impact the variability in mortality rates.”

This study examines interstate variability in liver disease mortality, highlighting significant geographic differences that may inform public health policies. ([link Text](#))

2. Demographic Disparities in Alcohol-Related Mortality

2.1 “Racial and Ethnic Disparities in Alcohol-Attributed Deaths in the United States, 1999–2020”

“Between 1999 and 2020, a total of 605,948 individuals died from alcohol-related causes in the US. The highest AAMR was observed among American Indian/Alaska Natives, who were 3.6 times as likely (95% CI: 3.57, 3.67) to die from alcohol-related causes compared to Non-Hispanic Whites. Non-Hispanic Blacks, Asians/Pacific Islanders, and Hispanics showed lower AAMRs compared to Non-Hispanic Whites. These results were similar when stratified by sex, with American Indian/Alaska Native males being 3.2 times as likely (95% CI: 3.09, 3.21) and females being 4.8 times as likely (95% CI: 4.73, 4.96) to die from alcohol compared to Non-Hispanic Whites.”

This study examines the burden and trends in alcohol-attributed mortality rates by race and ethnicity, revealing significant disparities among different groups. ([Link Text](#))

3. Trends in Alcohol-Related Mortality

3.1 “Alcohol-Related Deaths in the U.S. More Than Double from 1999 to 2020”

“The 85 and older age group saw a possible but nonsignificant increase. Additionally, individuals aged 55-64 had both the steepest rise in mortality and the highest absolute rates in both 1999 and 2020. Both men and women experienced significant increases in alcohol-related deaths, but men had the highest rates in both years and saw the steepest increase overall. Women, however, saw the largest proportional rise, with deaths increasing from 4.8 per 100,000 in 1999 to 12 in 2020.”

“Deaths in women increased two-and-a-half times, while Asian and Pacific Islander communities experienced the steepest rise of 2.4 times. Regionally, the Midwest experienced the greatest jump, with an increase of 2.5 times in alcohol-related mortality, followed by the Northeast, West and South.”

This study reports a significant increase in alcohol-related deaths over two decades, with notable rises among younger adults and women.([Link Text](#))