

US Chronic Disease Indicators Data

Marissa Cook

2023

Analyzing US Chronic Disease Indicators (Data provided by Data.gov)

```
setwd("C://Users/maris/Downloads/Code")
data <- read.csv("U.S._Chronic_Disease_Indicators__CDI_.csv")
names(data)
```

```
## [1] "YearStart"      "YearEnd"
## [3] "LocationAbbr"   "LocationDesc"
## [5] "DataSource"     "Topic"
## [7] "Question"       "Response"
## [9] "DataValueUnit"  "DataValueType"
## [11] "DataValue"      "DataValueAlt"
## [13] "DataValueFootnoteSymbol" "DataValueFootnote"
## [15] "LowConfidenceLimit" "HighConfidenceLimit"
## [17] "StratificationCategory1" "Stratification1"
## [19] "StratificationCategory2" "Stratification2"
## [21] "StratificationCategory3" "Stratification3"
## [23] "GeoLocation"    "ResponseID"
## [25] "LocationID"     "TopicID"
## [27] "QuestionID"     "DataValueTypeID"
## [29] "StratificationCategoryID1" "StratificationID1"
## [31] "StratificationCategoryID2" "StratificationID2"
## [33] "StratificationCategoryID3" "StratificationID3"
```

```
list_of_indicators <- unique(data$Topic)
list_of_indicators
```

```
## [1] "Alcohol"
## [2] "Chronic Kidney Disease"
## [3] "Cardiovascular Disease"
## [4] "Diabetes"
## [5] "Tobacco"
## [6] "Asthma"
## [7] "Disability"
## [8] "Overarching Conditions"
## [9] "Older Adults"
## [10] "Reproductive Health"
## [11] "Oral Health"
## [12] "Nutrition, Physical Activity, and Weight Status"
## [13] "Cancer"
## [14] "Mental Health"
## [15] "Arthritis"
## [16] "Chronic Obstructive Pulmonary Disease"
## [17] "Immunization"
```

Let's clean this file, removing unnescary columns. These columns are completely blank.

```
colSums(is.na(data)| data=="")
```

```
##          YearStart          YearEnd          LocationAbbr
##              0              0              0
##      LocationDesc      DataSource          Topic
##              0              0              0
##      Question      Response      DataValueUnit
##              0      1048575      139390
##      DataValueType      DataValue      DataValueAlt
##              0      340887      342930
##      DataValueFootnoteSymbol      DatavalueFootnote      LowConfidenceLimit
##              694455      694455      452792
##      HighConfidenceLimit      StratificationCategory1      Stratification1
##              452792              0              0
##      StratificationCategory2      Stratification2      StratificationCategory3
##              1048575      1048575      1048575
##      Stratification3      GeoLocation      ResponseID
##              1048575      8822      1048575
##      LocationID      TopicID      QuestionID
##              0              0              0
##      DataValueTypeID      StratificationCategoryID1      StratificationID1
##              0              0              0
##      StratificationCategoryID2      StratificationID2      StratificationCategoryID3
##              1048575      1048575      1048575
##      StratificationID3
##              1048575
```

```
empty_columns <- colSums(is.na(data)| data=="")==nrow(data)
empty_columns
```

```
##          YearStart          YearEnd          LocationAbbr
##          FALSE          FALSE          FALSE
##          LocationDesc          DataSource          Topic
##          FALSE          FALSE          FALSE
##          Question          Response          DataValueUnit
##          FALSE          TRUE          FALSE
##          DataValueType          DataValue          DataValueAlt
##          FALSE          FALSE          FALSE
##          DataValueFootnoteSymbol          DatavalueFootnote          LowConfidenceLimit
##          FALSE          FALSE          FALSE
##          HighConfidenceLimit          StratificationCategory1          Stratification1
##          FALSE          FALSE          FALSE
##          StratificationCategory2          Stratification2          StratificationCategory3
##          TRUE          TRUE          TRUE
##          Stratification3          GeoLocation          ResponseID
##          TRUE          FALSE          TRUE
##          LocationID          TopicID          QuestionID
##          FALSE          FALSE          FALSE
##          DataValueTypeID          StratificationCategoryID1          StratificationID1
##          FALSE          FALSE          FALSE
##          StratificationCategoryID2          StratificationID2          StratificationCategoryID3
##          TRUE          TRUE          TRUE
##          StratificationID3
##          TRUE
```

```
data<- data[,!empty_columns]
```

Texas Let's get the information pertaining only to Texas.

```
texas_data <- subset(data,LocationDesc == "Texas")
```

Describe the information associated with liver disease in Texas.

```
liver <- subset(texas_data,Question == "Chronic liver disease mortality")
age_liver <- subset(liver, DataValueType== "Age-adjusted Rate")
all(age_liver$YearStart== age_liver$YearEnd)
```

```
## [1] TRUE
```

```
gender <- subset(age_liver, StratificationCategory1== "Gender")
```

```
##Convert data value from a character to a numeric value
is.numeric(gender$DataValue)
```

```
## [1] FALSE
```

```
is.character(gender$DataValue)
```

```
## [1] TRUE
```

```
gender$DataValue <- as.numeric(gender$DataValue)
is.numeric(gender$DataValue)
```

```
## [1] TRUE
```

```
is.character(gender$YearStart)
```

```
## [1] FALSE
```

```
is.numeric(gender$YearStart)
```

```
## [1] TRUE
```

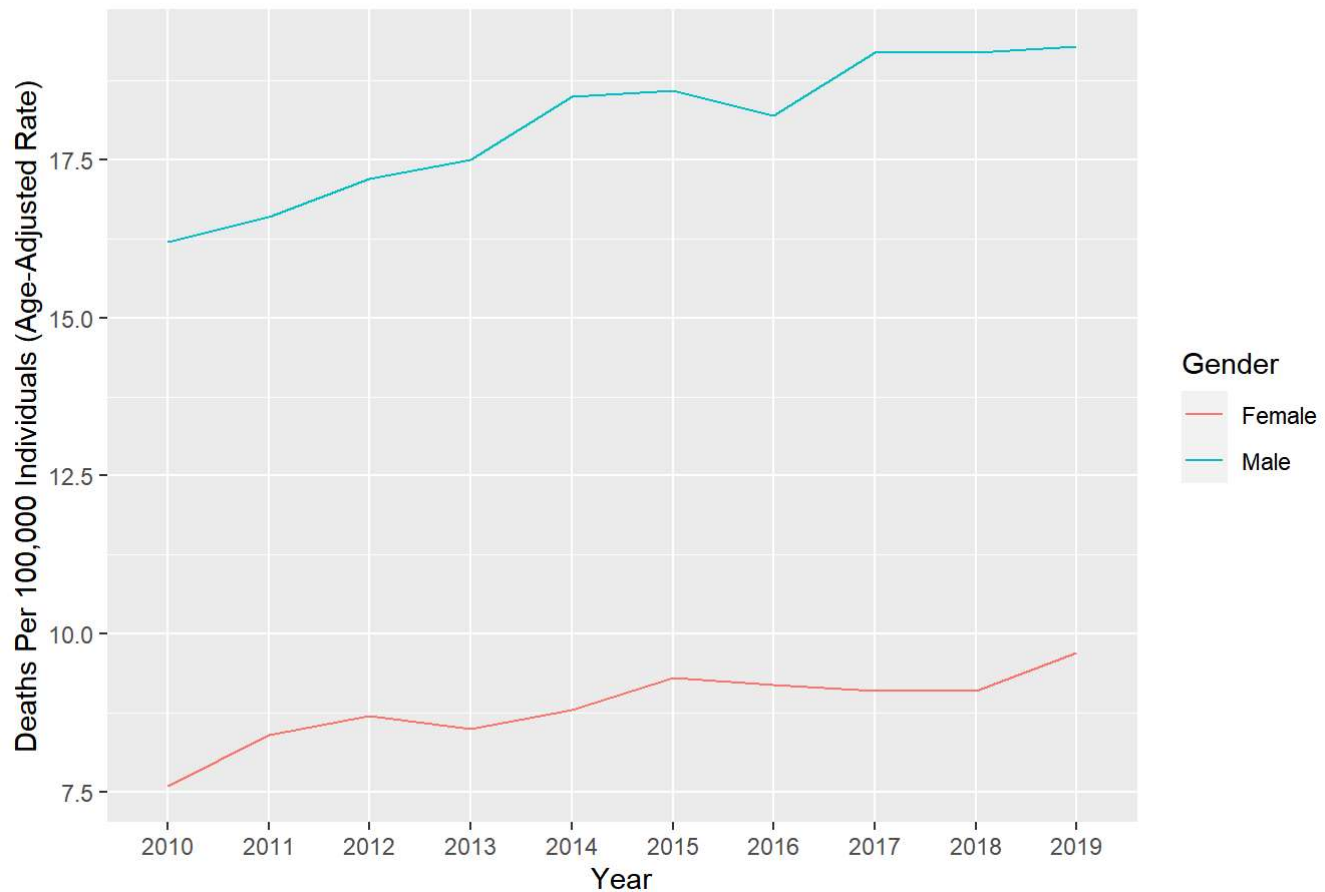
```
##Convert year from a numeric value to a character to lock it in as a string
gender$YearStart <- as.character(gender$YearStart)
is.character(gender$YearStart)
```

```
## [1] TRUE
```

```
library(ggplot2)
liver_gen <- ggplot(gender, aes(x=YearStart, y=DataValue, group=Stratification1, color= Stratifi
cation1))+
  geom_line()+
  ggtitle("Chronic Liver Disease Mortality Based on Alcohol Consumption in Texas")+
  labs(y= "Deaths Per 100,000 Individuals (Age-Adjusted Rate)", x= "Year", color= "Gender")

liver_gen
```

Chronic Liver Disease Mortality Based on Alcohol Consumption in Texas



```
print("From the graph, it appears that there is a significant difference between genders in chronic liver disease mortality caused by alcohol consumption in the state of Texas. Males have a reportedly higher age-adjusted mortality rate.")
```

```
## [1] "From the graph, it appears that there is a significant difference between genders in chronic liver disease mortality caused by alcohol consumption in the state of Texas. Males have a reportedly higher age-adjusted mortality rate."
```

How does this compare across ethnicities?

```
liver_ethn <- subset(age_liver, StratificationCategory1== "Race/Ethnicity")

##Convert data value from a character to a numeric value
liver_ethn$DataValue <- as.numeric(liver_ethn$DataValue)
is.numeric(liver_ethn$DataValue)
```

```
## [1] TRUE
```

```
##Convert year from a numeric value to a character to lock it in as a string
liver_ethn$YearStart <- as.character(liver_ethn$YearStart)
is.character(liver_ethn$YearStart)
```

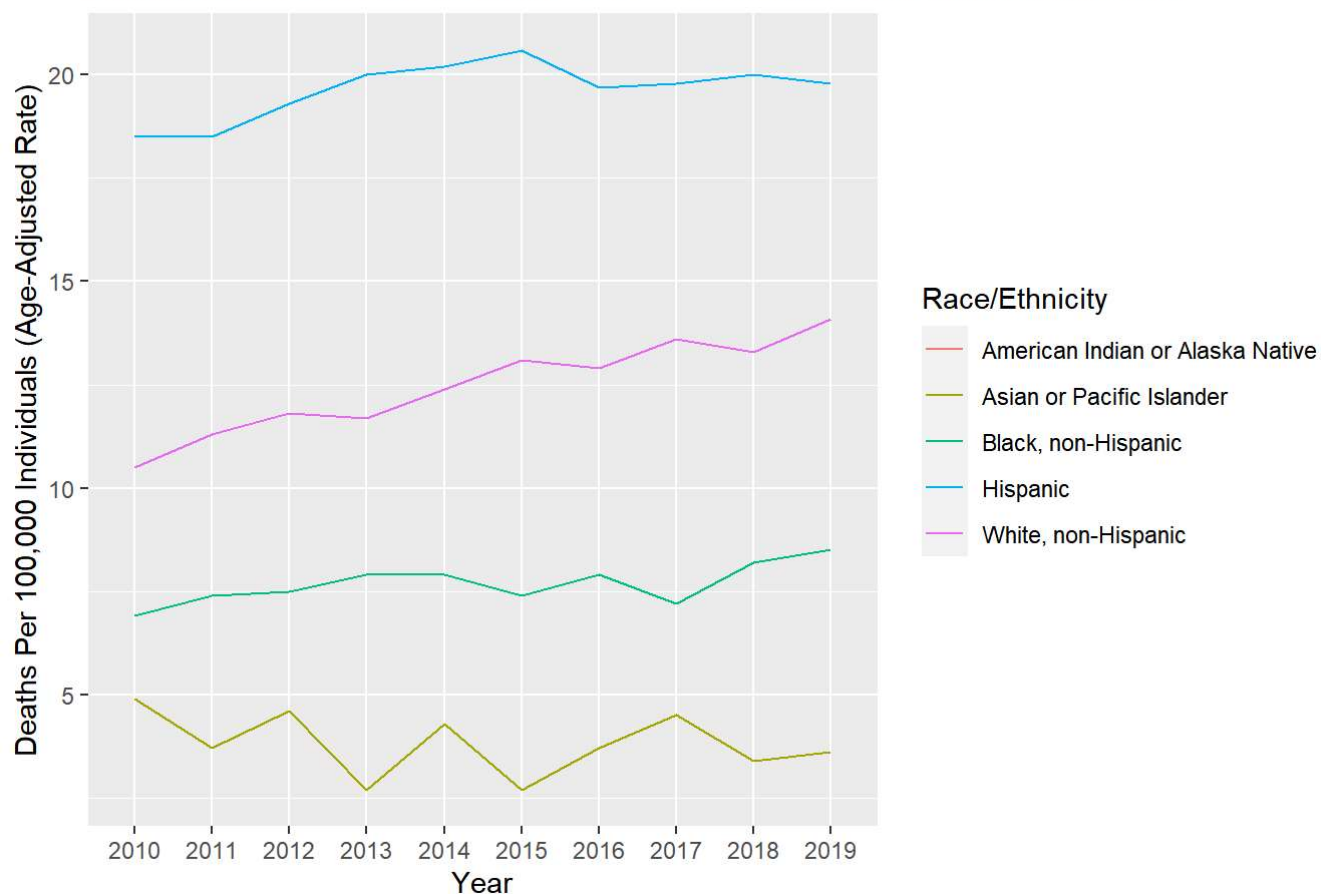
```
## [1] TRUE
```

```
liver_ethn <- ggplot(liver_ethn, aes(x=YearStart, y=DataValue, group=Stratification1, color= Stratification1))+
  geom_line()+
  ggtitle("Chronic Liver Disease Mortality Based on Alcohol Consumption in Texas")+
  labs(y= "Deaths Per 100,000 Individuals (Age-Adjusted Rate)", x= "Year", color= "Race/Ethnicity")
```

```
liver_ethn
```

```
## Warning: Removed 10 rows containing missing values (`geom_line()`).
```

Chronic Liver Disease Mortality Based on Alcohol Consumption in Texas



```
print("From the previous graph, the data suggests that of the ethnicities listed, individuals who identify as Hispanic have a higher chronic liver disease mortality rate in Texas. It was noted in the original dataset that there were attempts to gather this information for American Indians or Alaska Natives but there were either not enough occurrences in this population or very few respondents.")
```

```
## [1] "From the previous graph, the data suggests that of the ethnicities listed, individuals who identify as Hispanic have a higher chronic liver disease mortality rate in Texas. It was noted in the original dataset that there were attempts to gather this information for American Indians or Alaska Natives but there were either not enough occurrences in this population or very few respondents."
```

Let's look at the mortality rate of chronic liver disease for males in US territories in the most recent year provided in this dataset.

```
territories <- unique(data$LocationDesc)
unique(data$YearStart)
```

```
## [1] 2010 2017 2018 2011 2012 2015 2014 2016 2013 2009 2019 2001 2008 2020 2007
## [16] 2021
```

```
territories_males_2019 <- subset(data, data$Stratification1== "Male")
territories_liver <- subset(territories_males_2019, Question == "Chronic liver disease mortality")
all_male_liver_2019 <- subset(territories_liver, YearStart== "2019")
male_aged_liver_2019 <- subset(all_male_liver_2019, DataValueType== "Age-adjusted Rate")

##Convert data value from a character to a numeric value
male_aged_liver_2019$DataValue <- as.numeric(male_aged_liver_2019$DataValue)
is.numeric(male_aged_liver_2019$DataValue)
```

```
## [1] TRUE
```

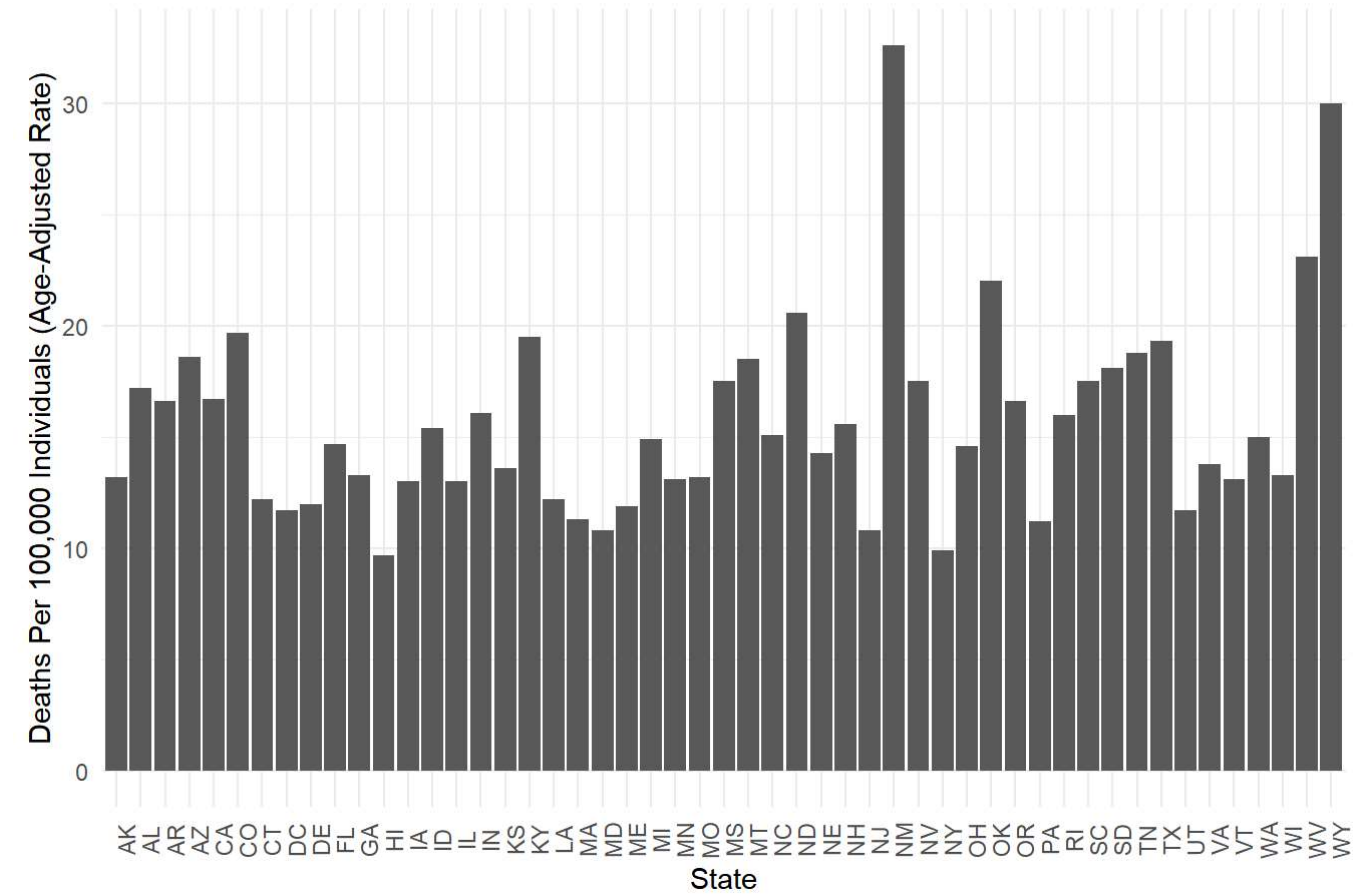
```
##Convert year from a numeric value to a character to lock it in as a string
male_aged_liver_2019$YearStart <- as.character(male_aged_liver_2019$YearStart)
is.character(male_aged_liver_2019$YearStart)
```

```
## [1] TRUE
```

```
us_territories_liver <- ggplot(male_aged_liver_2019, aes(x=LocationAbbr, y=DataValue))+
  geom_bar(stat="Identity")+
  ggtitle(" US Male Chronic Liver Disease Mortality From Alcohol Consumption in 2019")+
  labs(y= "Deaths Per 100,000 Individuals (Age-Adjusted Rate)", x= "State")

us_territories_liver <- us_territories_liver+ theme_minimal()+theme(axis.text.x = element_text(angle= 90))
us_territories_liver
```

US Male Chronic Liver Disease Mortality From Alcohol Consumption in 2019



From this graph, we can see that of all the US territories in 2019, the mortality rate of chronic liver disease by alcohol consumption was highest in the territories of New Mexico, Wyoming, and Wisconsin, respectively.