

# Viz6 - Extra Credit

Hayden Ginman

2023-10-07

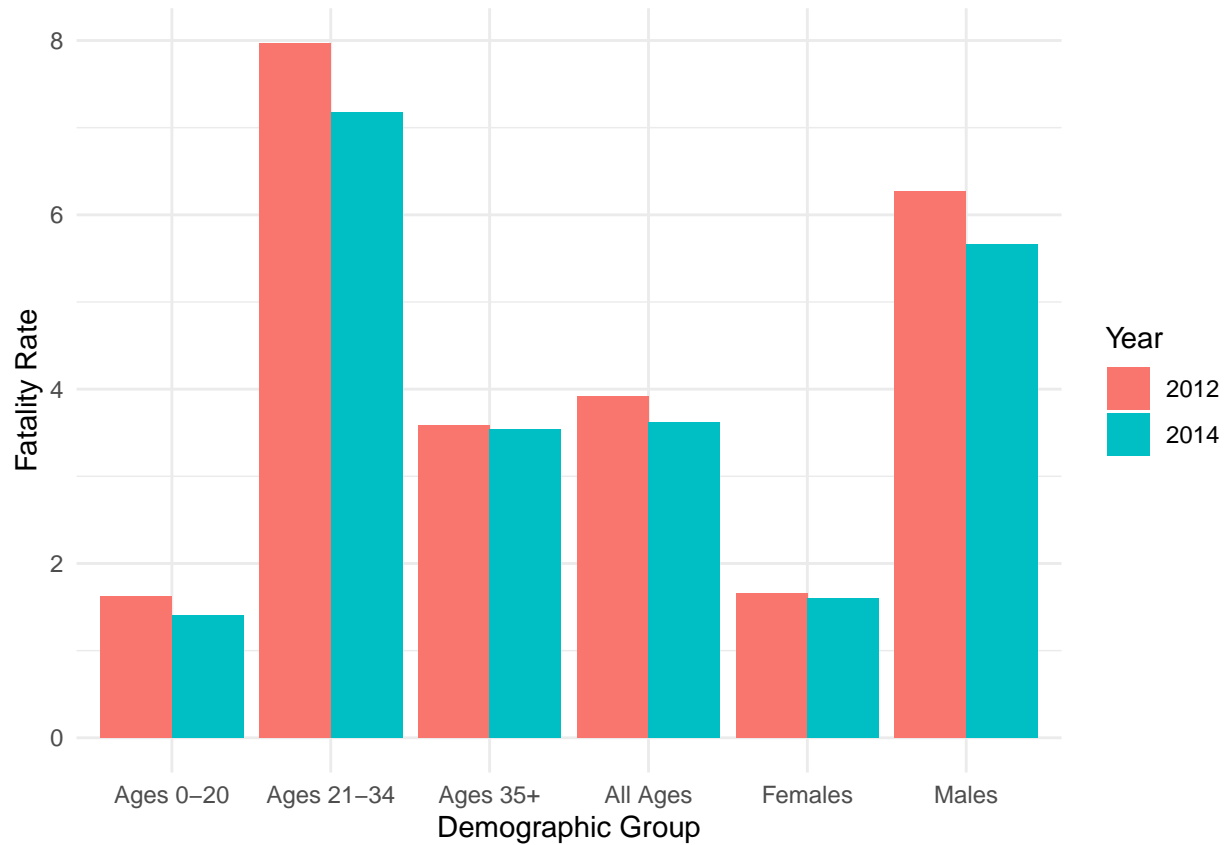
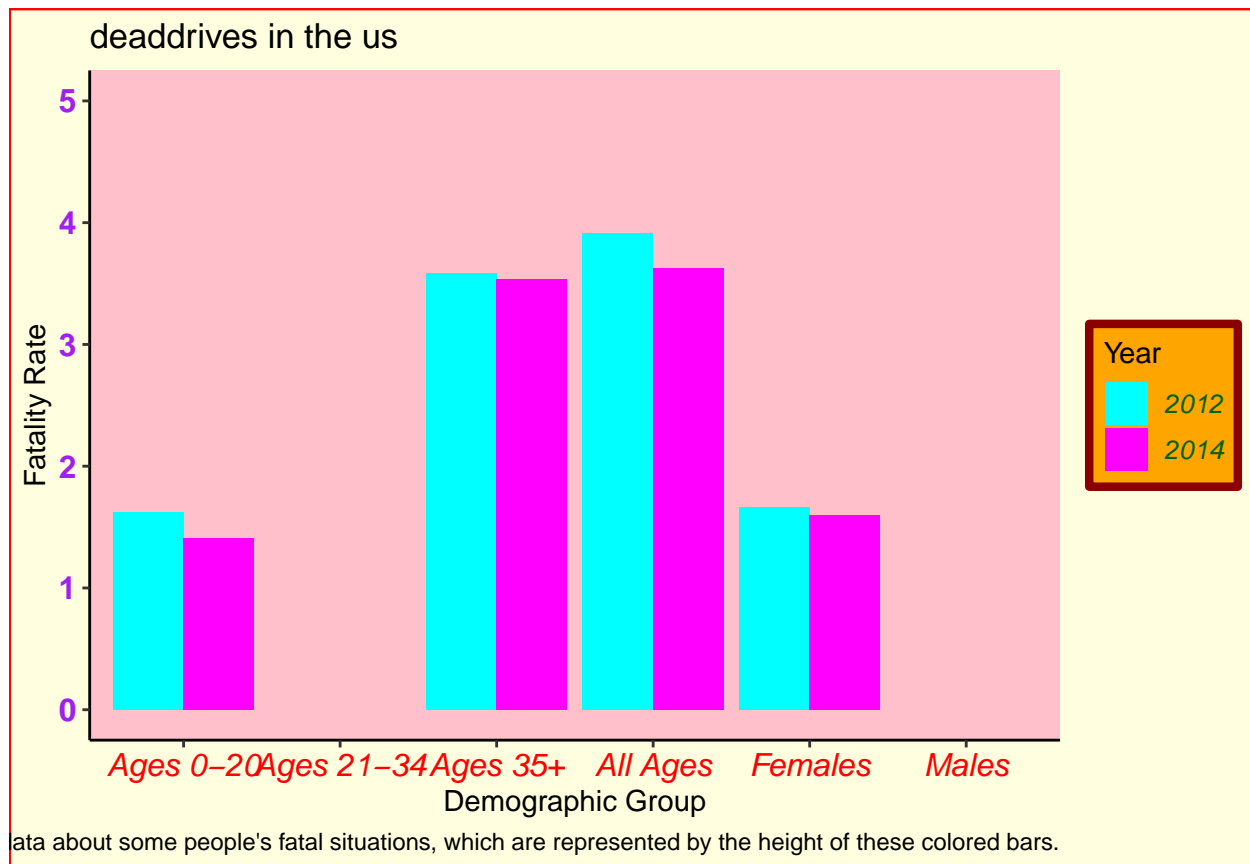


Figure 1: Comparison of impaired driving fatality rates by demographic group for the years 2012 and 2014 in the United States. The fatality rate represents the number of deaths per 100,000 population from crashes involving a driver with a Blood Alcohol Concentration (BAC) of 0.08% or higher. Data sources: Fatality Analysis Reporting System (FARS) for 2012 and the National Highway Traffic Administration's (NHTSA) FARS 2014 Annual Report File. Note: Rates based on fewer than 20 deaths are suppressed.



## Viz 2 Assignment

### Questions:

Which age group experienced the highest impaired driving fatality rate in the United States for the years 2012 and 2014?

How did the impaired driving fatality rate for males and females in the United States change between 2012 and 2014?

### Source:

<https://data.cdc.gov/Motor-Vehicle/Impaired-Driving-Death-Rate-by-Age-and-Gender-2012/ebbj-sh54>

### Data Description:

The data presents an analysis of the number of deaths per 100,000 individuals in age and gender groups for each state in the US. These fatalities occurred in car accidents where one driver had a Blood Alcohol Concentration (BAC) of 0.08% or higher which is considered the legal limit for being intoxicated in many areas. The data comes from the years 2012 and 2014. The data for 2012 is obtained from the Fatality Analysis Reporting System (FARS) while the data for 2014 is sourced from the Annual Report File of the National Highway Traffic Administrations (NHTSA).

```

# Load Packages into coding environment
library("ggplot2")
library("dplyr")
library("tidyr")
library("reshape2")
#Read data into coding environment

driving_death <- read.csv("Impaired_Driving_Death_Rate__by_Age_and_Gender__2012___2014__All_States.csv")

#rename annoying column names

new_dd <- driving_death %>%
  rename(
    State = State,
    Location = Location,
    All_2012 = `All.Ages..2012`,
    Age0to20_2012 = `Ages.0.20..2012`,
    Age21to34_2012 = `Ages.21.34..2012`,
    Age35plus_2012 = `Ages.35...2012`,
    Male_2012 = `Male..2012`,
    Female_2012 = `Female..2012`,
    All_2014 = `All.Ages..2014`,
    Age0to20_2014 = `Ages.0.20..2014`,
    Age21to34_2014 = `Ages.21.34..2014`,
    Age35plus_2014 = `Ages.35...2014`,
    Male_2014 = `Male..2014`,
    Female_2014 = `Female..2014`)

#Take the average for each age group column to aggregate data for entire US

US_dd <- new_dd %>%
  summarise(
    All_2012 = mean(All_2012, na.rm = TRUE),
    Age0to20_2012 = mean(Age0to20_2012, na.rm = TRUE),
    Age21to34_2012 = mean(Age21to34_2012, na.rm = TRUE),
    Age35plus_2012 = mean(Age35plus_2012, na.rm = TRUE),
    Male_2012 = mean(Male_2012, na.rm = TRUE),
    Female_2012 = mean(Female_2012, na.rm = TRUE),
    All_2014 = mean(All_2014, na.rm = TRUE),
    Age0to20_2014 = mean(Age0to20_2014, na.rm = TRUE),
    Age21to34_2014 = mean(Age21to34_2014, na.rm = TRUE),
    Age35plus_2014 = mean(Age35plus_2014, na.rm = TRUE),
    Male_2014 = mean(Male_2014, na.rm = TRUE),
    Female_2014 = mean(Female_2014, na.rm = TRUE))

#Transform data into long format with category and year as columns

long_dd <- US_dd %>%
  tidyr::pivot_longer(
    everything(),
    names_to = c("Category", "Year"),
    names_sep = "_",
    values_to = "Value")

```

*#Visualize data for pretty plot*

```
ggplot(long_dd, aes(x = Category, y = Value, fill = Year)) +  
  geom_bar(stat = "identity", position = "dodge") +  
  labs(x = "Demographic Group", y = "Fatality Rate") +  
  scale_x_discrete(labels = c(All = "All Ages", Age0to20 = "Ages 0-20", Age21to34 = "Ages 21-34", Age35to44 = "Ages 45-54", Age55to64 = "Ages 65-74", Age75plus = "Ages 75+"))
```

*#Visualize data for ugly plot*

```
ggplot(long_dd, aes(x = Category, y = Value, fill = Year)) +  
  geom_bar(stat = "identity", position = "dodge") +  
  labs(  
    title = "deaddrives in the us",  
    x = "Demographic Group",  
    y = "Fatality Rate",  
    caption = "This is a chart that shows data about some people's fatal situations, which are represented by the bars in the chart." ) +  
  scale_x_discrete(labels = c(All = "All Ages", Age0to20 = "Ages 0-20", Age21to34 = "Ages 21-34", Age35to44 = "Ages 45-54", Age55to64 = "Ages 65-74", Age75plus = "Ages 75+")) +  
  scale_y_continuous(limits=c(0, 5)) +  
  theme_classic() +  
  theme(  
    plot.background = element_rect(fill = "lightyellow", color = "red"),  
    panel.background = element_rect(fill = "pink"),  
    axis.text.x = element_text(face = "italic", color = "red", size = 12),  
    axis.text.y = element_text(face = "bold", color = "purple", size = 12),  
    legend.background = element_rect(fill = "orange", color = "darkred", size = 1.5, linetype = "solid"),  
    legend.text = element_text(face = "italic", color = "darkgreen", size = 10)) +  
  scale_fill_manual(values = c("2012" = "cyan", "2014" = "magenta"))
```