Healthy Aging

Dirk Hartog

2023-12-08

```
library(httr2)
library(httr)
library(jsonlite)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter()  masks stats::filter()
## x purrr::flatten() masks jsonlite::flatten()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(utils)
library(pdftools)
```

```
## Using poppler version 23.04.0
```

```
library(maps)
```

```
##
## Attaching package: 'maps'
##
## The following object is masked from 'package:purrr':
##
##     map
```

```
library(mapproj)
library(purrr)
```

INTRODUCION:

Despite a slight drop in life expectancy (LE) in the United States from 2020 - 2021, estimates for the average LE in 2023 support the return to a pre pandemic level of 79.11 years. As life expectancy continues to rise we might also expect that our aging population will continue to increase as well. As people get older, their risk of chronic diseases increases such as dementia, heart disease, type 2 diabetes, arthritis, and cancers.

QUESTIONS TO INVESTIGATE:

- What are the trends in the population of individuals 65 years or older in the United States?
- What are the leading causes of death for individuals aged 65 years or older in the United States.
- What are the trends in modifiable behaviors or risk factors that might be related to healthy aging?
- Is there any correlation between subjective cognitive decline and Alzheimer's death rates in the united states?

POPULATION TRENDS

```
# Clean and Tidy
url <- "https://raw.githubusercontent.com/D-hartog/DATA607/main/FINAL_PROJ/population_trend.csv"
population <- read_csv(url)
```

Read/clean/tidy data

```
## Rows: 69 Columns: 100
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (1): Location
## dbl (99): 2008__Children 0-18, 2008__Adults 19-25, 2008__Adults 26-34, 2008_...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
population <- population[1:52,]

head(population)
```

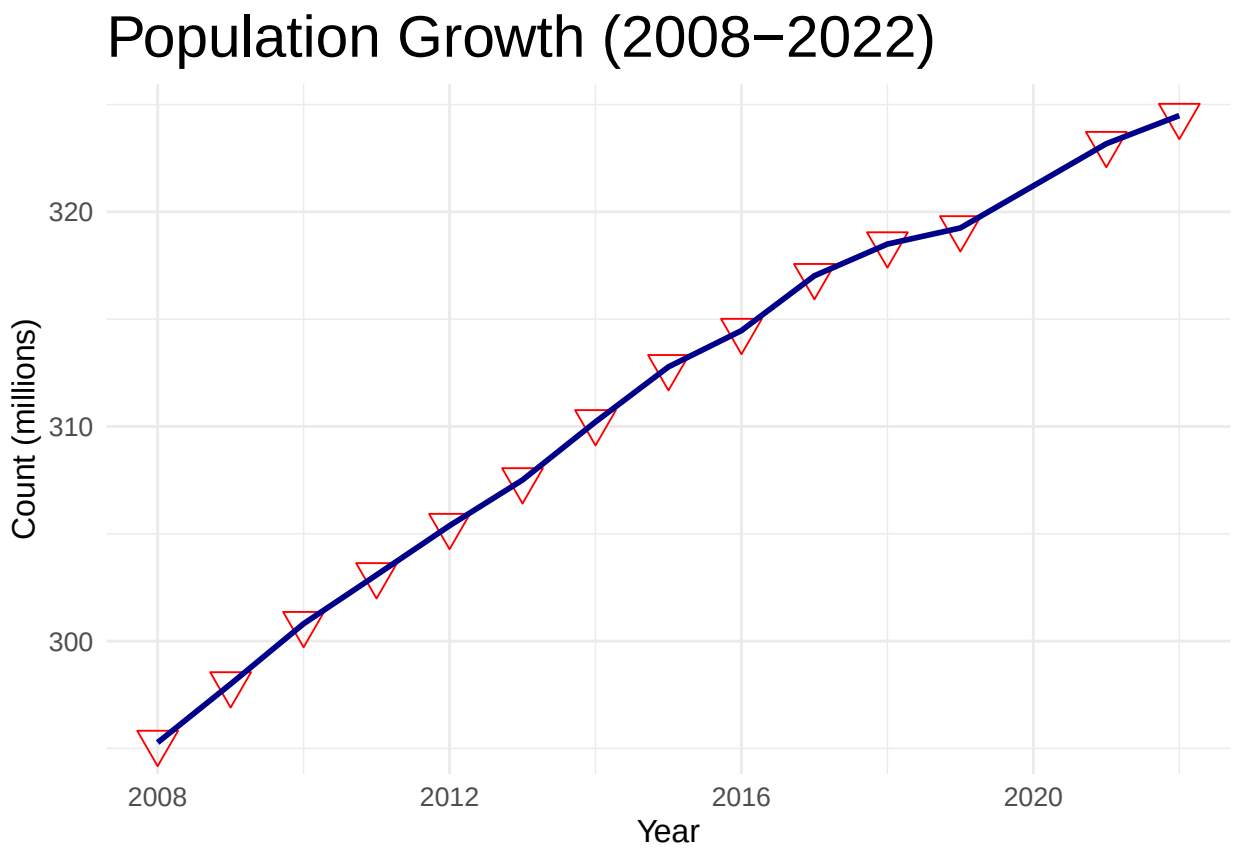
```
## # A tibble: 6 x 100
##   Location      `2008__Children 0-18` `2008__Adults 19-25` `2008__Adults 26-34`
##   <chr>                         <dbl>                <dbl>                <dbl>
## 1 United States              76770400             27241500             34823900
## 2 Alabama                     1168700               429900               509400
## 3 Alaska                       184700                74100                79700
## 4 Arizona                     1776500               572900               801300
## 5 Arkansas                     727200               250300               314800
## 6 California                  9749000              3619900              4515600
## # i 96 more variables: `2008__Adults 35-54` <dbl>, `2008__Adults 55-64` <dbl>,
## #   `2008__65+` <dbl>, `2008__Total` <dbl>, `2009__Children 0-18` <dbl>,
## #   `2009__Adults 19-25` <dbl>, `2009__Adults 26-34` <dbl>,
## #   `2009__Adults 35-54` <dbl>, `2009__Adults 55-64` <dbl>, `2009__65+` <dbl>,
## #   `2009__Total` <dbl>, `2010__Children 0-18` <dbl>,
## #   `2010__Adults 19-25` <dbl>, `2010__Adults 26-34` <dbl>,
## #   `2010__Adults 35-54` <dbl>, `2010__Adults 55-64` <dbl>, ...
```

```r
# Multiple variables stored in column names
population <- pivot_longer(population,
             cols = `2008__Children 0-18`:`2022__Total`,
             names_to = "Year_AgeGroup",
             values_to = "Count")

population$Year_AgeGroup <- str_replace_all(population$Year_AgeGroup, "__", ",")

population <- separate_wider_delim(population,
                    Year_AgeGroup,
                    delim = ",",
                    names = c("Year", "Group")
                    )
population$Footnotes <- NULL
population$Year <- as.numeric(population$Year)
head(population)
```

```
## # A tibble: 6 x 4
##   Location       Year Group            Count
##   <chr>         <dbl> <chr>            <dbl>
## 1 United States  2008 Children 0-18 76770400
## 2 United States  2008 Adults 19-25  27241500
## 3 United States  2008 Adults 26-34  34823900
## 4 United States  2008 Adults 35-54  85799700
## 5 United States  2008 Adults 55-64  33433900
## 6 United States  2008 65+           37200300
```

Visualize data 1. Time series plot of us pop

```r
# 1. Time series plot of us pop
population %>% filter(Group == "Total" & Location == "United States") %>%
  select(Year, Count) %>%
  ggplot(aes(x = Year, y = Count /1000000)) +
  geom_point(color = "red", shape = 25, size = 5) +
  geom_line(color = "darkblue", linetype = 1, linewidth = 1) +
  ggtitle("Population Growth (2008-2022)") +
  xlab("Year") +
  ylab("Count (millions)") +
  theme_minimal() +
  theme(legend.position = "none",
        axis.title.x = element_text(color="black",size=12),
        axis.title.y = element_text(color="black", size=12),
        axis.text.x = element_text(size=10),
        axis.text.y = element_text(size=10),
        plot.title = element_text(color = "black", size=22))
```

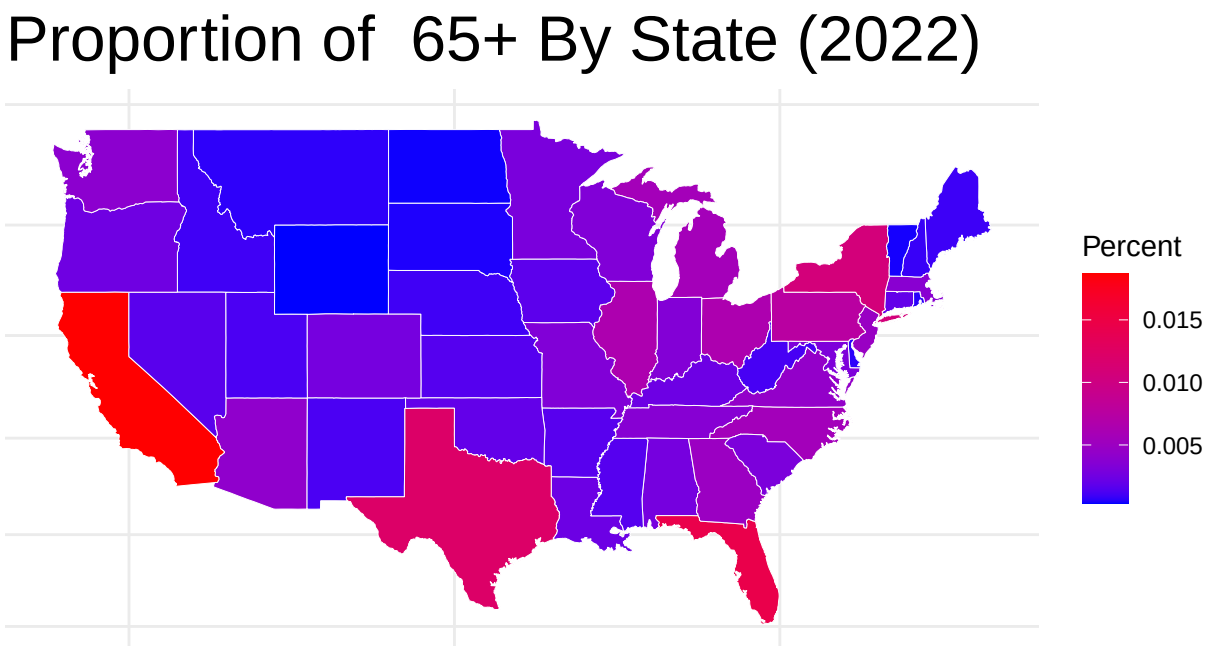



```
population %>% filter(Group == "Total" & Location == "United States") %>%
  select(Year, Count) %>% mutate(Pecent_change = Count)
```

```
## # A tibble: 14 x 3
##     Year     Count Pecent_change
##    <dbl>     <dbl>         <dbl>
##  1  2008 295269800     295269800
##  2  2009 297997800     297997800
##  3  2010 300806200     300806200
##  4  2011 303080600     303080600
##  5  2012 305376600     305376600
##  6  2013 307505500     307505500
##  7  2014 310212200     310212200
##  8  2015 312777900     312777900
##  9  2016 314466400     314466400
## 10  2017 317022500     317022500
## 11  2018 318498500     318498500
## 12  2019 319249300     319249300
## 13  2021 323175700     323175700
## 14  2022 324486000     324486000
```

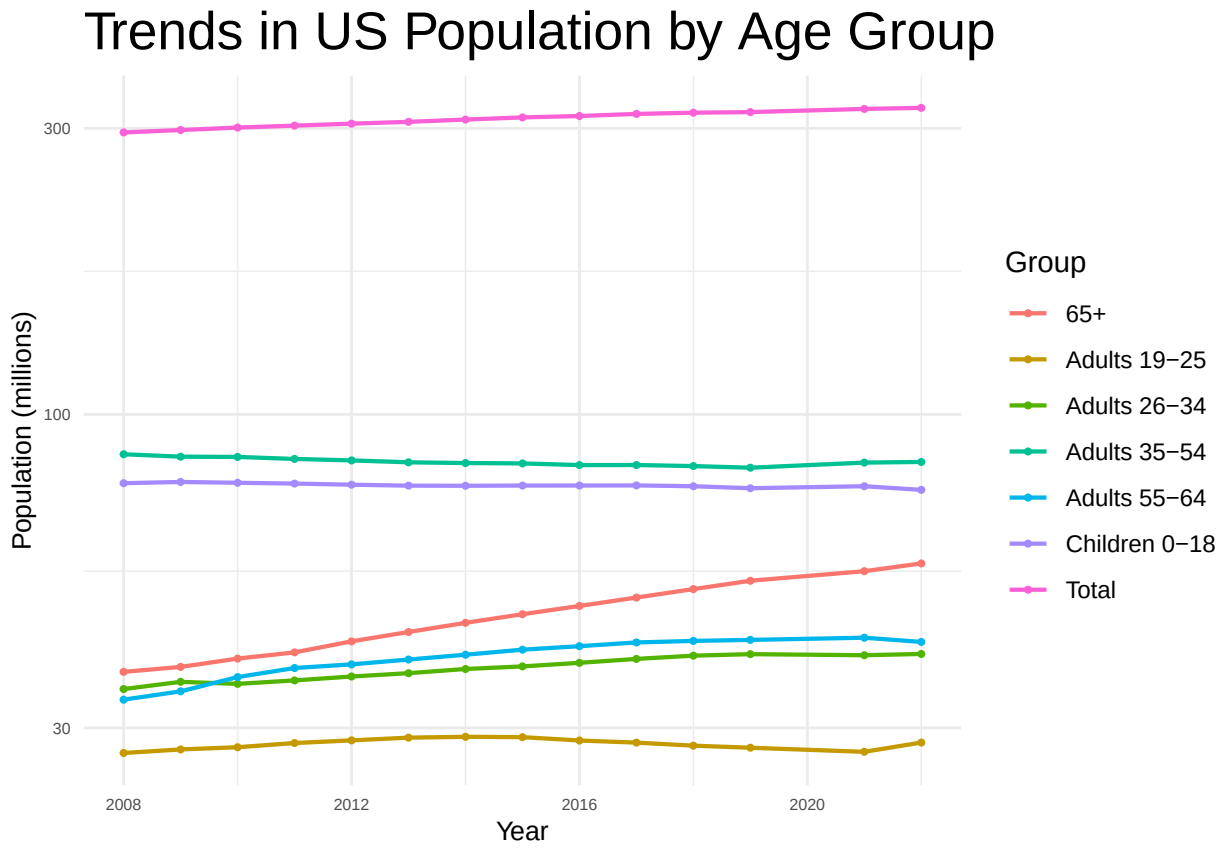
2. Proportion of population in each state

```r
states <- map_data("state")
states$region <- str_to_title(states$region)

states <- states[order(states$order),]

# Find population in 2022
#population %>% filter(Location == "United States" & Year == 2022 & Group == "Total")

# State proportions of 65+

population %>% filter(Location != "United States" & Year == 2022 & Group == "65+") %>%
  merge(states, by.x = "Location", by.y = "region") %>%
  mutate(Percent = Count /324486000) %>%
  arrange(group, order) %>%
  ggplot(aes(x = long, y = lat)) +
  geom_polygon(aes(group = group, fill = Percent), color = "white", linewidth = 0.1) +
  coord_map() +
  scale_fill_gradient(low = "blue", high = "red") +
  theme_minimal() +
  ggtitle("Proportion of  65+ By State (2022)") +
  theme(axis.title.x = element_blank(),
        axis.title.y = element_blank(),
        axis.text.x = element_blank(),
        axis.text.y = element_blank(),
        plot.title = element_text(color = "black", size=24))
```



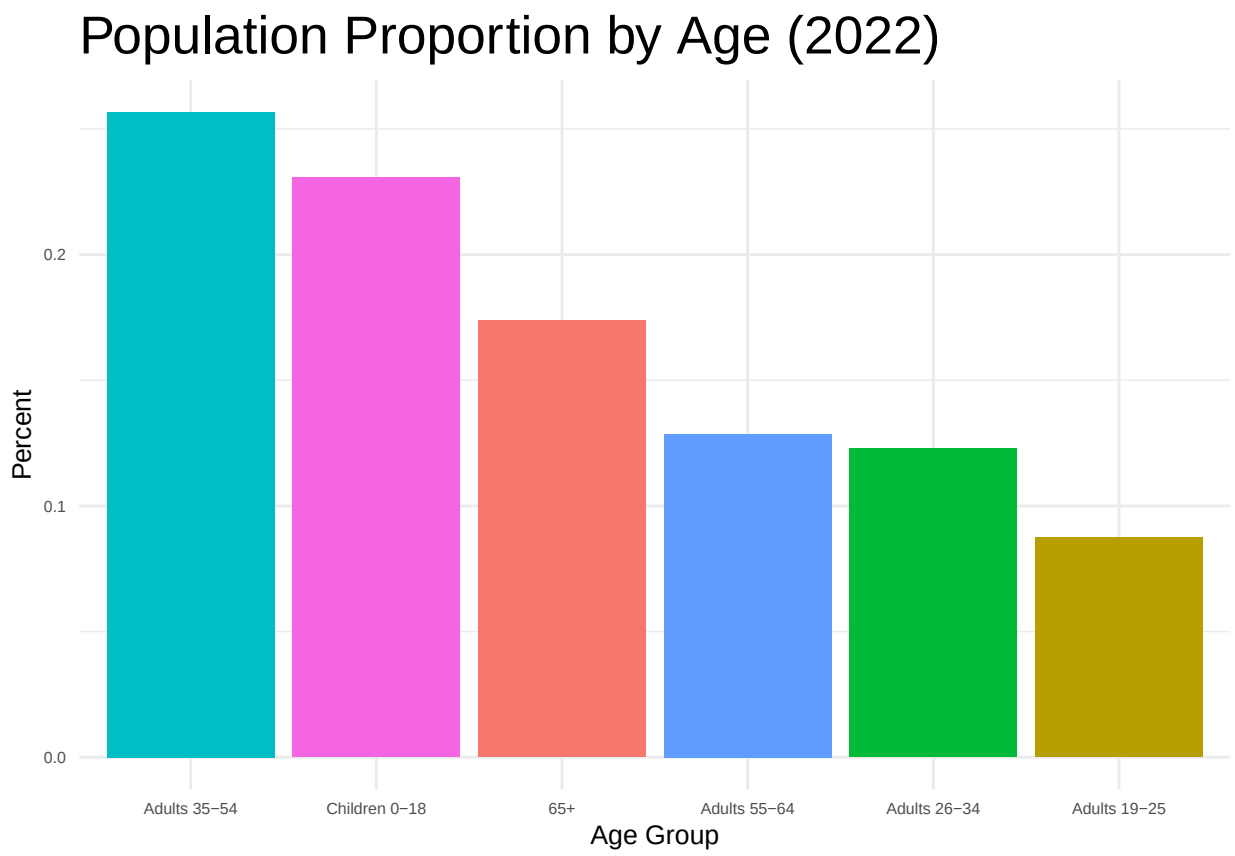
3. Trends in the proportions of the population

```
# 1. Time series plot of us population by age group
population %>% filter(Location == "United States") %>%
  group_by(Group) %>%
  ggplot(aes(x = Year, y = Count /1000000, color = Group)) +
  geom_point(size = 0.75) +
  geom_line(linewidth = 0.75) +
  scale_y_log10() +
  ggtitle("Trends in US Population by Age Group") +
  xlab("Year") +
  ylab("Population (millions)") +
  theme_minimal() +
  theme(legend.position = "right",
        axis.title.x = element_text(color="black",size=10),
        axis.title.y = element_text(color="black", size=10),
        axis.text.x = element_text(size=6),
        axis.text.y = element_text(size=6),
        plot.title = element_text(color = "black", size=20))
```




```
# 2. Proportion of the us population by age
population %>% filter(Year == 2022 &
                      Location == "United States") %>%
  mutate(Percent = Count /max(Count)) %>%
    filter(Group != "Total") %>%
  ggplot(aes(x = fct_rev(fct_reorder(Group,Percent)), y = Percent)) +
```

```
  geom_col(aes(fill = Group)) +
  ggtitle("Population Proportion by Age (2022)") +
  xlab("Age Group") +
  ylab("Percent") +
  theme_minimal() +
  theme(legend.position = "none",
        axis.title.x = element_text(color="black",size=10),
        axis.title.y = element_text(color="black", size=10),
        axis.text.x = element_text(size=6),
        axis.text.y = element_text(size=6),
        plot.title = element_text(color = "black", size=20))
```



LEADING CAUSES OF DEATH IN THE USA

Read/clean/tidy data

```
# Read in pdf file from local directory
## ISSUE - could not figure out how to load a pdf from github into R

# link: https://github.com/D-hartog/DATA607/blob/main/FINAL_PROJ/leading_causes1819.pdf
pdf1 <- c("leading_causes1819.pdf")
raw_text <- purrr::map(pdf1, pdf_text)
```

```r
#Split pages
raw_text <- purrr::map(raw_text, ~ str_split(.x,"\\n") %>% unlist)

#Concatenate the split pages
raw_text <- reduce(raw_text, c)

#Specify the start and end of the table data
table_start <- str_which((raw_text), "Data Brief 395: Mortality in the United States, 2019")
table_end <- str_which((raw_text), "... Category not applicable.")
table_end <- table_end[min(which(table_end > table_start))]

#Build the table and remove special characters
table <- raw_text[(table_start):(table_end)]
table <- str_replace_all(table, "\\s{2,}", "|")
table <- table[8:21]
table[1] <- "Rank1|Cause of death (based on International Classification of Diseases,10th Revision [ICD-
table[3] <- "AC|All causes|NA|2,839,205|100.0|723.6|2,854,838|100.0|715.2"
table[14] <- "AOC|All other causes|(residual)|744,312|26.2|...|758,167|26.6|..."
text_con <- textConnection(table)
data_table <- read.csv(text_con, sep = "|")

#Create a list of column names
colnames(data_table) <- c("Cause_of_death",
                          "ICD-10",
                          "Number_18",
                          "Percent_18",
                          "Rate_18",
                          "Number_19",
                          "Percent_19",
                          "Rate_19")
data_table <- tibble(data_table)

df <- data_table

df$`ICD-10` <- NULL

# Matching the df to what our other mortality data has
df[df == "Diseases of heart"] <- "Heart disease"
df[df == "Malignant neoplasms" ] <- "Cancer"
df[df == "Accidents (unintentional injuries)"] <- "Unintentional injuries"
df[df == "Intentional self-harm (suicide)"] <- "Suicide"
df[df == "Cerebrovascular diseases"] <- "Stroke"
df[df == "Chronic lower respiratory diseases"] <- "CLRD"
df[df == "Diabetes mellitus"] <-  "Diabetes"
df[df == "Nephritis, nephrotic syndrome and nephrosis"] <-  "Kidney disease"
df[df == "Alzheimer disease"] <-  "Alzheimer's disease"

df <- df %>% filter(Cause_of_death != "All other causes")

df$Rate_18 <- as.double(df$Rate_18)
df$Rate_19 <- as.double(df$Rate_19)

# Clean numbers column and convert the data type
```

```r
df$Number_18 <- as.double(str_remove_all(df$Number_18, ","))
df$Number_19 <- as.double(str_remove_all(df$Number_19, ","))

df$Percent_18 <-  df$Percent_18 + 0.0
df$Percent_19 <- df$Percent_19 + 0.0

yr_18 <- df[,c(1,2,3,4)]
yr_18$Year <- 2018
colnames(yr_18) <- c("Cause_of_death", "Number", "Percent", "Rate", "Year")

yr_19 <- df[,c(1,5,6,7)]
yr_19$Year <- 2019
colnames(yr_19) <- c("Cause_of_death", "Number", "Percent", "Rate", "Year")

df <- rbind(yr_18, yr_19)
df <- df %>% select(Year, Cause_of_death, Number, Percent, Rate)
df$Percent <- NULL
```

```r
# LEADING CAUSES OF DEATH 2020 - 2021 PDF
# https://github.com/D-hartog/DATA607/blob/2bb1ad36a7459bbf61e5c9d6baff1f32638af115/FINAL_PROJ/leading_
pdf2 <- c("leading_causes2021.pdf")
raw_text2 <- purrr::map(pdf2, pdf_text)
# https://www.youtube.com/watch?v=bJH-S2iaxNo

#Split pages
raw_text2 <- purrr::map(raw_text2, ~ str_split(.x,"\\n") %>% unlist)
#Concatenate the split pages
raw_text2 <- reduce(raw_text2, c)

#Specify the start and end of the table data
table_start2 <- str_which((raw_text2), "Data Brief 456. Mortality in the United States, 2021")
table_end2 <- str_which((raw_text2), "... Category not applicable\b")
table_end2 <- table_end2[min(which(table_end2 > table_start2))]

#Build the table and remove special characters
table2 <- raw_text2[(table_start2):(table_end2)]
table2 <- str_remove_all(table2, "\b")
table2 <- str_replace_all(table2, "\\s{2,}", "|")

table2 <- table2[8:24]

table2[1] <- "Rank1|Cause_of_death|Number_20|Percent_20|Rate_20|Number_21|Percent_21|Rate_21"
table2[3] <- "AC|All causes|ICD_10|3,383,729|1000|8354|3,464,231|1000|8797"
table2[7] <- "4|Accidents (unintentional injuries)|(V01-X59,Y85-Y86)|200,955|59|576|224,935|65|647"
table2[13] <- "9|Chronic liver disease and cirrhosis|(K70,K73-K74)|51,642|15|133|56,585|16|145"
table2[15] <- "10|Nephritis, nephrotic syndrome and nephrosis (kidney disease)|(N00-N07,N17-N19,N25-N27
table2[17] <- "AOC|All other causes|(residual)|879,091|260|...|882,775|255|..."
table2 <- table2[c(1,3,4,5,6,7,9,10,11,12,13,15,17)]
text_con2 <- textConnection(table2)
data_table2 <- read.csv(text_con2, sep = "|")

#Create a list of column names
colnames(data_table2) <- c("Cause_of_death",
```

```
                           "ICD-10",
                           "Number_20",
                           "Percent_20",
                           "Rate_20",
                           "Number_21",
                           "Percent_21",
                           "Rate_21")
data_table2 <- tibble(data_table2)

df2 <- data_table2

df2$`ICD-10` <- NULL

df2$Rate_20 <- as.double(df2$Rate_20)
```

```
## Warning: NAs introduced by coercion
```

```
df2$Rate_21 <- as.double(df2$Rate_21)
```

```
## Warning: NAs introduced by coercion
```

```
# Clean numbers column anc convert data type
df2$Number_20 <- as.double(str_remove_all(df2$Number_20, ","))
df2$Number_21 <- as.double(str_remove_all(df2$Number_21, ","))

df2$Percent_20 <-  df2$Percent_20 + 0.0
df2$Percent_21 <- df2$Percent_21 + 0.0

yr_20 <- df2[,c(1,2,3,4)]
yr_20$Year <- 2020
colnames(yr_20) <- c("Cause_of_death", "Number", "Percent", "Rate", "Year")

yr_21 <- df2[,c(1,5,6,7)]
yr_21$Year <- 2021
colnames(yr_21) <- c("Cause_of_death", "Number", "Percent", "Rate", "Year")

df2 <- rbind(yr_20, yr_21)
df2 <- df2 %>% select(Year, Cause_of_death, Number, Percent, Rate)
df2$Percent <- NULL

df2 %>% group_by(Year) %>% arrange(Year,desc(Rate))
```

```
## # A tibble: 24 x 4
## # Groups:   Year [2]
##     Year Cause_of_death                       Number  Rate
##    <dbl> <chr>                                 <dbl> <dbl>
##  1  2020 All causes                          3383729  8354
##  2  2020 Diseases of heart                    696962  1682
##  3  2020 Malignant neoplasms (cancer)         602350  1441
##  4  2020 COVID-19                             350831   850
##  5  2020 Accidents (unintentional injuries)   200955   576
##  6  2020 Cerebrovascular diseases (stroke)    160264   388
```

```
##  7  2020 Chronic lower respiratory diseases   152657   364
##  8  2020 Alzheimer disease                    134242   324
##  9  2020 Diabetes mellitus                    102188   248
## 10  2020 Chronic liver disease and cirrhosis   51642   133
## # i 14 more rows
```

```
df2[df2 == "Diseases of heart"] <- "Heart disease"
df2[df2 == "Malignant neoplasms (cancer)"] <- "Cancer"
df2[df2 == "Accidents (unintentional injuries)"] <- "Unintentional injuries"
df2[df2 == "Intentional self-harm (suicide)"] <- "Suicide"
df2[df2 == "Cerebrovascular diseases (stroke)"] <- "Stroke"
df2[df2 == "Chronic lower respiratory diseases"] <- "CLRD"
df2[df2 == "Diabetes mellitus"] <-  "Diabetes"
df2[df2 == "Nephritis, nephrotic syndrome and nephrosis (kidney disease)"] <-  "Kidney disease"
df2[df2 == "Alzheimer disease"] <-  "Alzheimer's disease"

df2 <- df2 %>% filter(Cause_of_death != "All other causes")


df2$Rate <- df2$Rate /10
```

Visualizations

```
# Combine data-frames into one
# ISSUE - could not figure out how to load a zipped from github into R

leading_mortality <- read_csv("https://raw.githubusercontent.com/D-hartog/DATA607/main/FINAL_PROJ/Leadi
```

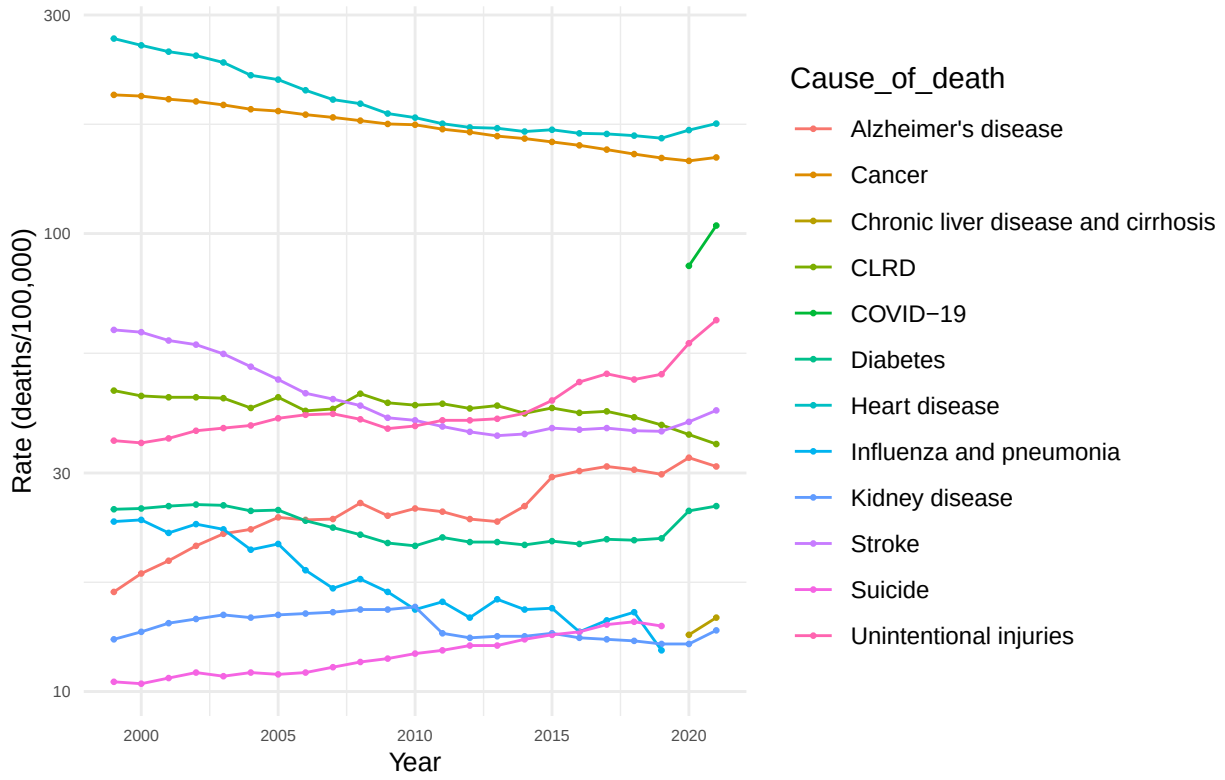
```
## Rows: 10868 Columns: 6
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (3): 113 Cause Name, Cause Name, State
## dbl (3): Year, Deaths, Age-adjusted Death Rate
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
leading_mortality <- leading_mortality %>% filter(State == "United States") %>%
  select(Year, `Cause Name`, Deaths, `Age-adjusted Death Rate`)

colnames(leading_mortality) <- c("Year", "Cause_of_death", "Number", "Rate")

leading_mortality <- rbind(df, df2, leading_mortality)

head(leading_mortality)
```

```
## # A tibble: 6 x 4
##    Year Cause_of_death          Number  Rate
##   <dbl> <chr>                    <dbl> <dbl>
## 1  2018 All causes             2839205 724.
## 2  2018 Heart disease           655381 164.
```

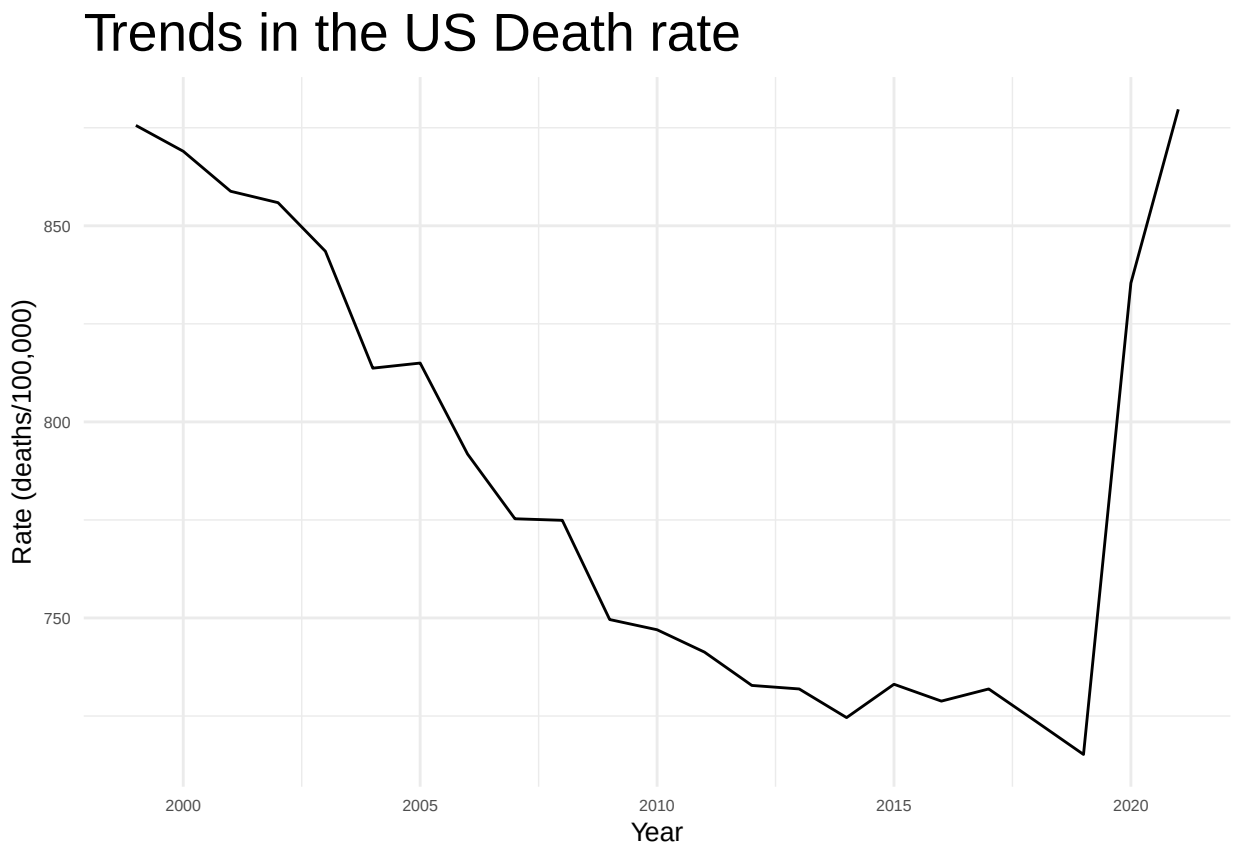
```
## 3  2018 Cancer                    599274 149.
## 4  2018 Unintentional injuries    167127  48
## 5  2018 CLRD                      159486  39.7
## 6  2018 Stroke                    147810  37.1
```

```
leading_mortality %>% filter(Cause_of_death != "All causes") %>%
  group_by(Year) %>% arrange(Year,desc(Rate)) %>%
  ggplot(aes(x = Year, y = Rate, color = Cause_of_death)) +
  geom_point(size = 0.5) +
  geom_line() +
  scale_y_log10() +
  ggtitle("Trends in the Leading Causes of Death in the US") +
  xlab("Year") +
  ylab("Rate (deaths/100,000)") +
  theme_minimal() +
  theme(axis.title.x = element_text(color="black",size=10),
        axis.title.y = element_text(color="black", size=10),
        axis.text.x = element_text(size=6),
        axis.text.y = element_text(size=6),
        plot.title = element_text(color = "black", size=20))
```




```
leading_mortality %>% filter(Cause_of_death == "All causes") %>%
  ggplot(aes(x = Year, y = Rate)) +
  geom_line() +
  ggtitle("Trends in the US Death rate") +
```

```
  xlab("Year") +
  ylab("Rate (deaths/100,000)") +
  theme_minimal() +
  theme(legend.position = "none",
        axis.title.x = element_text(color="black",size=10),
        axis.title.y = element_text(color="black", size=10),
        axis.text.x = element_text(size=6),
        axis.text.y = element_text(size=6),
        plot.title = element_text(color = "black", size=20))
```




```
# Percent change in select diseases
alz_pct_change <- ((31.0 - 16.5)/16.5) * 100
heart_pct_change <- ((173.8 - 266.5)/266.5) * 100
cancer_pct_change <- ((146.6 - 200.8)/200.8) * 100
uninten_pct_change <-((64.7 - 35.3)/35.3) * 100
tibble(Alzheimers = alz_pct_change,Heart_desiease = heart_pct_change, Caner = cancer_pct_change, Uninte
```

```
## # A tibble: 1 x 4
##   Alzheimers Heart_desiease Caner Unintentional
##        <dbl>          <dbl> <dbl>         <dbl>
## 1       87.9          -34.8 -27.0          83.3
```

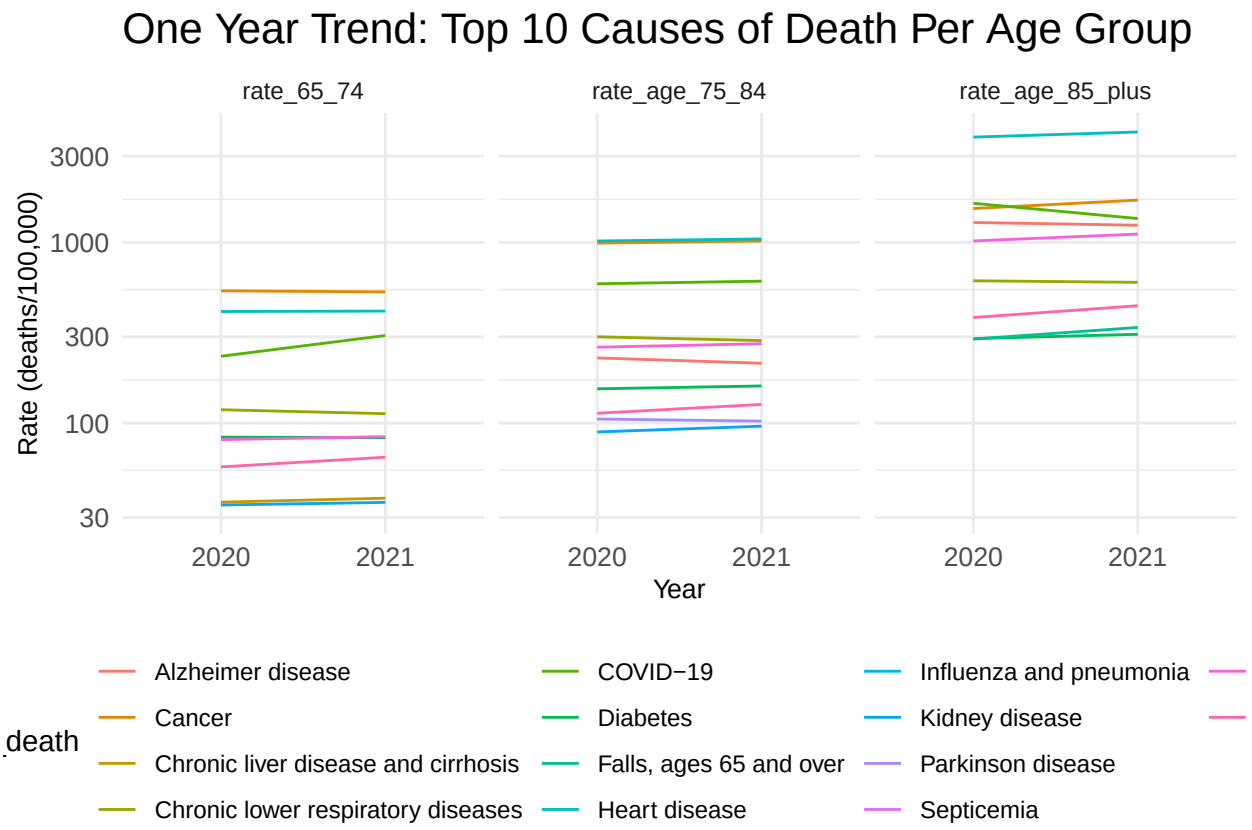
CAUSES OF DEATH IN 65+

Read/clean/tidy data

```
# read in data using - API

#https://data.cdc.gov/NCHS/NCHS-VSRR-Quarterly-provisional-estimates-for-sele/489q-934x/about_data
req <- request("https://data.cdc.gov/resource/489q-934x.json")

resp <- req_perform(req)

resp <- resp %>% resp_body_json()

# tidy data - transform to wider table

mortality65_plus <- unnest_wider(tibble(resp), resp)

# separate year and quarter to make it easier to filter data

mortality65_plus <- mortality65_plus %>% separate_wider_regex(year_and_quarter,
                                  c(year = "\\d{4}", quarter = "\\sQ\\d"))

# clean columns and convert data types
mortality65_plus$rate_overall <- as.double(mortality65_plus$rate_overall)
mortality65_plus$quarter <- str_trim(mortality65_plus$quarter)

# filter age groups, rate_type, causes, quarter and time period
# pivot age columns longer to construct visuals

mortality65_plus <- mortality65_plus %>% select(year, quarter, time_period, cause_of_death,rate_type, u
  filter(rate_type == "Crude" & cause_of_death != "All causes" & quarter == "Q4" &
           time_period == "12 months ending with quarter") %>%
  pivot_longer(cols = rate_65_74:rate_age_85_plus,
               names_to = "group",
               values_to = "rate")

mortality65_plus$rate <- as.double(mortality65_plus$rate)

head(mortality65_plus)
```

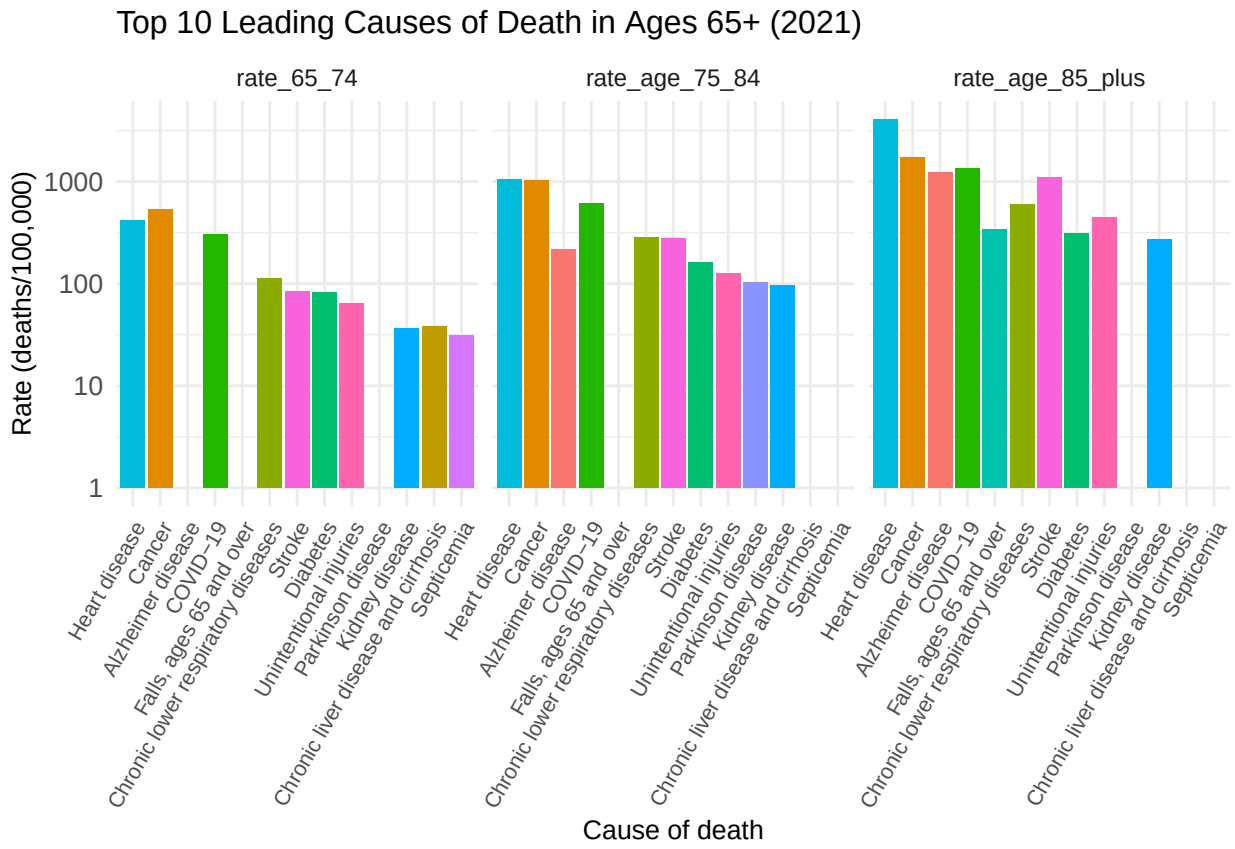
```
## # A tibble: 6 x 9
##   year  quarter time_period    cause_of_death rate_type unit  rate_overall group
##   <chr> <chr>   <chr>          <chr>          <chr>     <chr>        <dbl> <chr>
## 1 2020  Q4      12 months end~ Alzheimer dis~ Crude     Deat~         40.7 rate~
## 2 2020  Q4      12 months end~ Alzheimer dis~ Crude     Deat~         40.7 rate~
## 3 2020  Q4      12 months end~ Alzheimer dis~ Crude     Deat~         40.7 rate~
## 4 2020  Q4      12 months end~ COVID-19       Crude     Deat~        106.  rate~
## 5 2020  Q4      12 months end~ COVID-19       Crude     Deat~        106.  rate~
## 6 2020  Q4      12 months end~ COVID-19       Crude     Deat~        106.  rate~
## # i 1 more variable: rate <dbl>
```

```
mortality65_plus %>% group_by(group, year) %>% top_n(10, rate) %>%
  ggplot(aes(x = year, y = rate, color = cause_of_death)) +
  geom_line(aes(group = cause_of_death)) +
  facet_grid(~group) +
  scale_y_log10() +
  theme(legend.position = "bottom") +
  guides(color = guide_legend(ncol = 4)) +
  ggtitle("One Year Trend: Top 10 Causes of Death Per Age Group") +
  xlab("Year") +
  ylab("Rate (deaths/100,000)") +
  theme_minimal() +
  theme(legend.position = "bottom",
        axis.title.x = element_text(color="black",size=10),
        axis.title.y = element_text(color="black", size=10),
        axis.text.x = element_text(size=10),
        axis.text.y = element_text(size=10),
        plot.title = element_text(color = "black", size=16))
```




```
mortality65_plus %>% filter(year == 2021) %>% group_by(group) %>%
  top_n(10, rate) %>%
  ggplot(aes(x = fct_rev(fct_reorder(cause_of_death,rate)), y = rate, fill = cause_of_death)) +
  geom_col() +
  facet_grid(~group) +
  scale_y_log10() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1)) +
```

```
  ggtitle("Top 10 Leading Causes of Death in Ages 65+ (2021)") +
  xlab("Cause of death") +
  ylab("Rate (deaths/100,000)") +
  theme_minimal() +
  theme(legend.position = "none",
        axis.title.x = element_text(color="black",size=10),
        axis.title.y = element_text(color="black", size=10),
        axis.text.x = element_text(size=8, angle = 60, vjust = 1, hjust = 1),
        axis.text.y = element_text(size=10),
        plot.title = element_text(color = "black", size=12))
```



HEALTH INDICATORS: BEHAVIORAL RISK FACTOR SURVEILANCE SYSTEM (BRFSS)

```
brfss <- read_csv("Alzheimer_s_Disease_and_Healthy_Aging_Data_20231115.csv")
```

```
## Rows: 250937 Columns: 39
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (27): RowId, LocationAbbr, LocationDesc, Datasource, Class, Topic, Quest...
## dbl  (4): YearStart, YearEnd, Data_Value, Data_Value_Alt
```

```
## lgl  (8): Response, Sample_Size, StratificationCategory3, Stratification3, R...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
glimpse(brfss)
```

```
## Rows: 250,937
## Columns: 39
## $ RowId                       <chr> "BRFSS~2021~2021~9004~Q43~TOC11~AGE~RACE", ~
## $ YearStart                   <dbl> 2021, 2017, 2019, 2020, 2020, 2015, 2020, 2~
## $ YearEnd                     <dbl> 2021, 2017, 2019, 2020, 2020, 2015, 2020, 2~
## $ LocationAbbr                <chr> "WEST", "NRE", "MDW", "US", "NH", "MDW", "U~
## $ LocationDesc                <chr> "West", "Northeast", "Midwest", "United Sta~
## $ Datasource                  <chr> "BRFSS", "BRFSS", "BRFSS", "BRFSS", "BRFSS"~
## $ Class                       <chr> "Overall Health", "Overall Health", "Nutrit~
## $ Topic                       <chr> "Arthritis among older adults", "Arthritis ~
## $ Question                    <chr> "Percentage of older adults ever told they ~
## $ Response                    <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ Data_Value_Unit             <chr> "%", "%", "%", "%", "%", "%", "Number", "%"~
## $ DataValueTypeID             <chr> "PRCTG", "PRCTG", "PRCTG", "PRCTG", "PRCTG"~
## $ Data_Value_Type             <chr> "Percentage", "Percentage", "Percentage", "~
## $ Data_Value                  <dbl> 31.6, 50.3, 14.3, 55.5, 15.2, 59.8, 6.2, 61~
## $ Data_Value_Alt              <dbl> 31.6, 50.3, 14.3, 55.5, 15.2, 59.8, 6.2, 61~
## $ Data_Value_Footnote_Symbol <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ Data_Value_Footnote         <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ Low_Confidence_Limit        <chr> "28.8", "49.1", "13.8", "54.5", "12.8", "48~
## $ High_Confidence_Limit       <chr> "34.4", "51.6", "14.8", "56.4", "18.0", "70~
## $ Sample_Size                 <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ StratificationCategory1     <chr> "Age Group", "Age Group", "Age Group", "Age~
## $ Stratification1             <chr> "Overall", "65 years or older", "Overall", ~
## $ StratificationCategory2     <chr> "Race/Ethnicity", NA, NA, "Gender", "Gender~
## $ Stratification2             <chr> "Hispanic", NA, NA, "Female", "Female", "Na~
## $ StratificationCategory3     <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ Stratification3             <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ Geolocation                 <chr> NA, NA, NA, NA, "POINT (-71.50036091999965 ~
## $ ClassID                     <chr> "C01", "C01", "C02", "C01", "C05", "C01", "~
## $ TopicID                     <chr> "TOC11", "TOC11", "TNC02", "TOC11", "TMC01"~
## $ QuestionID                  <chr> "Q43", "Q43", "Q02", "Q43", "Q03", "Q43", "~
## $ ResponseID                  <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ LocationID                  <chr> "9004", "9001", "9002", "59", "33", "9002",~
## $ StratificationCategoryID1   <chr> "AGE", "AGE", "AGE", "AGE", "AGE", "AGE", "~
## $ StratificationID1           <chr> "AGE_OVERALL", "65PLUS", "AGE_OVERALL", "65~
## $ StratificationCategoryID2   <chr> "RACE", "OVERALL", "OVERALL", "GENDER", "GE~
## $ StratificationID2           <chr> "HIS", "OVERALL", "OVERALL", "FEMALE", "FEM~
## $ StratificationCategoryID3   <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ StratificationID3           <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ Report                      <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
```

```
brfss <- brfss %>% select(-c("Response", "Sample_Size", "StratificationCategory3", "Stratification3",
                             "ResponseID", "StratificationCategoryID3", "StratificationID3",
                             "Report", "Data_Value_Unit", "DataValueTypeID", "Data_Value_Footnote_Symbol",
                             "Datasource", "StratificationCategory1", "StratificationCategoryID1",
```

```
                        "StratificationCategoryID2", "RowId"))


brfss <- pivot_wider(brfss,
            names_from = "StratificationCategory2",
            values_from = "Stratification2")

brfss$`NA` <- NULL

# Filter out Gaum and Puerto Rico
brfss <- brfss %>% filter(LocationDesc != "Guam" & LocationDesc != "Puerto Rico")


# Clean Gender and Race columns
brfss$Gender <- as.character(brfss$Gender)
brfss$Gender[brfss$Gender == "NULL"] <- NA

brfss <- brfss %>% rename("Race_Ethnicity" = `Race/Ethnicity`)
brfss$Race_Ethnicity <- as.character(brfss$Race_Ethnicity)
brfss$Race_Ethnicity[brfss$Race_Ethnicity == "NULL"] <- NA


brfss <- brfss %>% rename("AgeGroup" = "Stratification1",
                                "AgeGroupID" = "StratificationID1",
                                "Gender_RaceID" = "StratificationID2") # new_name = old_name

head(brfss)
```

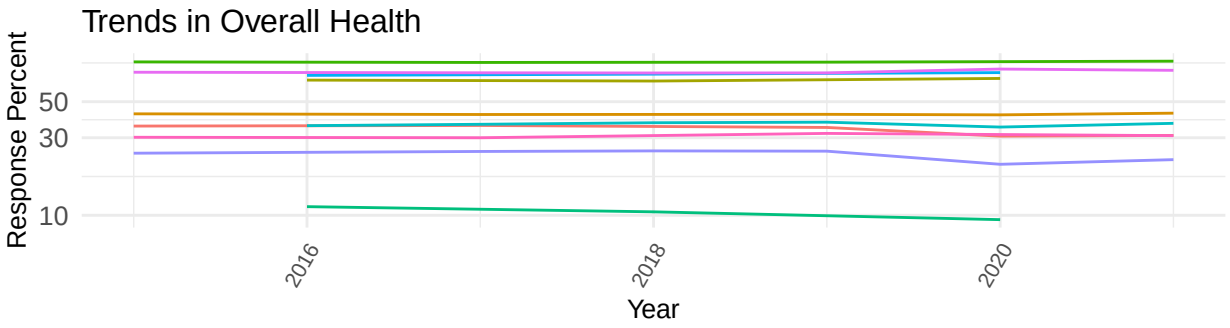
```
## # A tibble: 6 x 23
##   YearStart YearEnd LocationAbbr LocationDesc               Class Topic Question
##       <dbl>   <dbl> <chr>        <chr>                      <chr> <chr> <chr>
## 1      2021    2021 WEST         West                       Over~ Arth~ Percent~
## 2      2017    2017 NRE          Northeast                  Over~ Arth~ Percent~
## 3      2019    2019 MDW          Midwest                    Nutr~ Eati~ Percent~
## 4      2020    2020 US           United States, DC & Terri~ Over~ Arth~ Percent~
## 5      2020    2020 NH           New Hampshire              Ment~ Freq~ Percent~
## 6      2015    2015 MDW          Midwest                    Over~ Arth~ Percent~
## # i 16 more variables: Data_Value_Type <chr>, Data_Value <dbl>,
## #   Data_Value_Alt <dbl>, Data_Value_Footnote <chr>,
## #   Low_Confidence_Limit <chr>, High_Confidence_Limit <chr>, AgeGroup <chr>,
## #   Geolocation <chr>, ClassID <chr>, TopicID <chr>, QuestionID <chr>,
## #   LocationID <chr>, AgeGroupID <chr>, Gender_RaceID <chr>,
## #   Race_Ethnicity <chr>, Gender <chr>
```

```
# Rearrange columns
brfss <- brfss %>% select(YearStart, YearEnd, LocationID, LocationAbbr, LocationDesc,
                    ClassID, Class, TopicID, Topic, AgeGroupID, AgeGroup, Gender_RaceID, Race_Ethnicity
                    Gender, QuestionID, Question, Data_Value_Type, Data_Value,
                    Data_Value_Alt, Data_Value_Footnote, Geolocation)

unique(brfss$Race_Ethnicity)
```

```
## [1] "Hispanic"                NA
```

```
## [3] "Native Am/Alaskan Native" "Asian/Pacific Islander"
## [5] "Black, non-Hispanic"      "White, non-Hispanic"
```

```
# Create df of each character column with their ID and distinct value
# Maybe try to put this into table form
class <- brfss %>% select(ClassID, Class) %>% distinct()
location <- brfss %>% select(LocationID, LocationAbbr, LocationDesc) %>% distinct()
topic <- brfss %>% select(TopicID, Topic) %>% distinct()
age <- brfss %>% select(AgeGroupID, AgeGroup) %>% distinct()
gender_race <- brfss %>% select(Gender_RaceID, Gender, Race_Ethnicity) %>% distinct()
question <- brfss %>% select(QuestionID, Question, Data_Value_Type) %>% distinct()

# Trends of topics within Overall Health Class
brfss %>% filter(LocationAbbr == "US" &
                     AgeGroupID == "AGE_OVERALL" &
                     Gender_RaceID == "OVERALL" &
                   Data_Value_Type == "Percentage" &
                   Class == "Overall Health") %>%
  group_by(Class) %>% ggplot(aes(x = YearStart, y = Data_Value, color = Question)) +
  geom_line() +
  theme_minimal() +
  guides(color = guide_legend(ncol = 1)) +
  scale_y_log10() +
  ggtitle("Trends in Overall Health") +
  xlab("Year") +
  ylab("Response Percent") +
  theme(legend.position = "bottom",
        axis.title.x = element_text(color="black",size=10),
        axis.title.y = element_text(color="black", size=10),
        axis.text.x = element_text(size=8, angle = 60, vjust = 1, hjust = 1),
        axis.text.y = element_text(size=10),
        plot.title = element_text(color = "black", size=12))
```



r health among older adults with doctor−diagnosed arthritis

e of older adults ever told they have arthritis

e of older adults getting sufficient sleep (>6 hours)

e of older adults who have been told they have high blood pressure who report currently taking medication for their high

e of older adults who have fallen and sustained an injury within last year

e of older adults who report having a disability (includes limitations related to sensory or mobility impairments or a phys

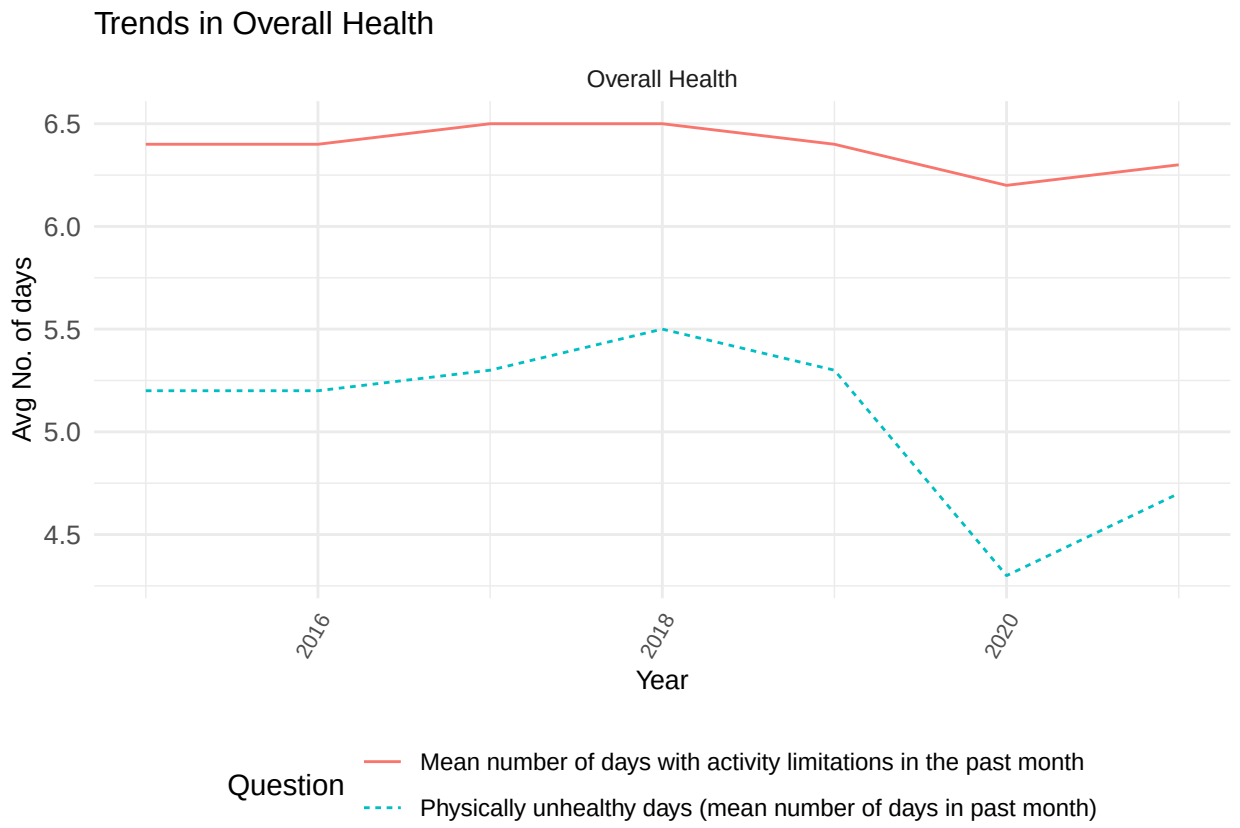
e of older adults who report having lost 5 or fewer teeth due to decay or gum disease

e of older adults who self−reported that their health is "fair" or "poor"

e of older adults who self−reported that their health is "good", "very good", or "excellent"

ıt pain due to arthritis among older adults with doctor−diagnosed arthritis

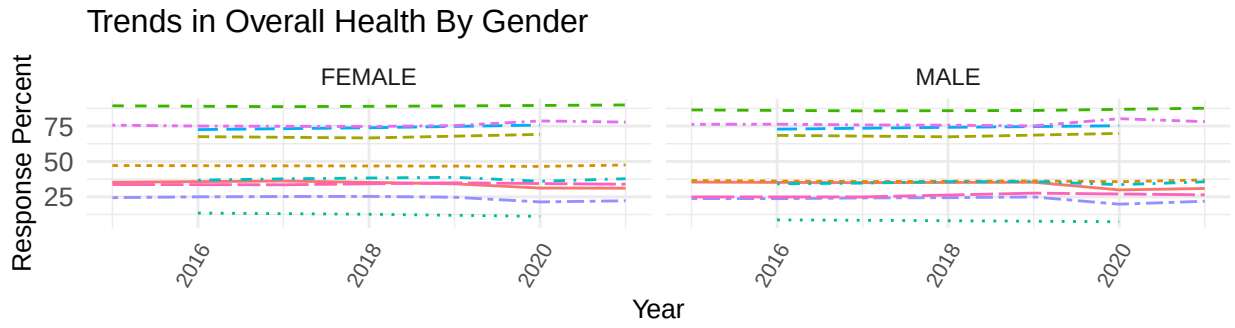
```
# Trends in questions that have a mean value type
brfss %>% filter(LocationAbbr == "US" &
                     AgeGroupID == "AGE_OVERALL" &
                     Gender_RaceID == "OVERALL" &
                     Data_Value_Type == "Mean") %>%
  group_by(Class) %>% ggplot(aes(x = YearStart,
                                 y = Data_Value,
                                 linetype = Question,
                                 color = Question)) +
  geom_line() +
  facet_grid(~Class) +
  theme_minimal() +
  guides(color = guide_legend(ncol = 1),
         linetype = guide_legend(ncol = 1)) +
  ggtitle("Trends in Overall Health") +
  xlab("Year") +
  ylab("Avg No. of days") +
  theme(legend.position = "bottom",
        axis.title.x = element_text(color="black",size=10),
        axis.title.y = element_text(color="black", size=10),
        axis.text.x = element_text(size=8, angle = 60, vjust = 1, hjust = 1),
        axis.text.y = element_text(size=10),
        plot.title = element_text(color = "black", size=12))
```




```
# Trends in overall health between genders

unique(brfss$Gender_RaceID)
```

```
## [1] "HIS"     "OVERALL" "FEMALE"  "NAA"     "MALE"    "ASN"     "BLK"
## [8] "WHT"
```

```
brfss %>% filter((Gender_RaceID == "FEMALE" | Gender_RaceID == "MALE") &
                      Class == "Overall Health" & LocationAbbr == "US" &
                      AgeGroupID == "AGE_OVERALL" &
                      Data_Value_Type == "Percentage") %>%
  ggplot(aes(x = YearEnd, y = Data_Value, color = Question, linetype = Question)) +
  geom_line() +
  facet_grid(~ Gender_RaceID) +
  theme_minimal() +
  guides(color = guide_legend(ncol = 1)) +
  ggtitle("Trends in Overall Health By Gender") +
  xlab("Year") +
  ylab("Response Percent") +
  theme(legend.position = "bottom",
        axis.title.x = element_text(color="black",size=10),
        axis.title.y = element_text(color="black", size=10),
        axis.text.x = element_text(size=8, angle = 60, vjust = 1, hjust = 1),
        axis.text.y = element_text(size=10),
        plot.title = element_text(color = "black", size=12))
```



r health among older adults with doctor−diagnosed arthritis

ɜ of older adults ever told they have arthritis

ɜ of older adults getting sufficient sleep (>6 hours)

ɜ of older adults who have been told they have high blood pressure who report currently taking medication for their higl

ɜ of older adults who have fallen and sustained an injury within last year

ɜ of older adults who report having a disability (includes limitations related to sensory or mobility impairments or a phys

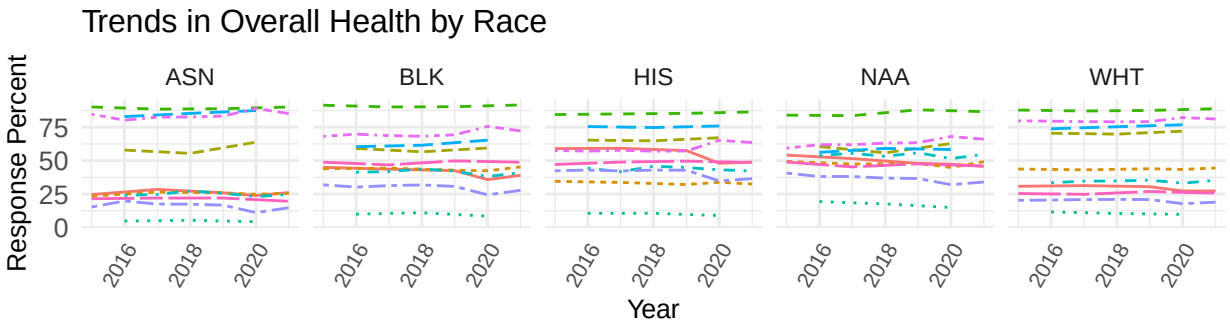
ɜ of older adults who report having lost 5 or fewer teeth due to decay or gum disease

ɜ of older adults who self−reported that their health is "fair" or "poor"

ɜ of older adults who self−reported that their health is "good", "very good", or "excellent"

ıt pain due to arthritis among older adults with doctor−diagnosed arthritis

```
# Trends in overall health between races

brfss %>% filter((Gender_RaceID == "HIS" | Gender_RaceID == "NAA" |
                        Gender_RaceID == "ASN" | Gender_RaceID == "BLK" |
                        Gender_RaceID == "WHT") &
                       Class == "Overall Health" & LocationAbbr == "US" &
                       AgeGroupID == "AGE_OVERALL" &
                       Data_Value_Type == "Percentage") %>%
  ggplot(aes(x = YearEnd, y = Data_Value, color = Question, linetype = Question)) +
  geom_line() +
  facet_grid(~ Gender_RaceID) +
  theme_minimal() +
  guides(color = guide_legend(ncol = 1)) +
  ggtitle("Trends in Overall Health by Race") +
  xlab("Year") +
  ylab("Response Percent") +
  theme(legend.position = "bottom",
        axis.title.x = element_text(color="black",size=10),
        axis.title.y = element_text(color="black", size=10),
        axis.text.x = element_text(size=8, angle = 60, vjust = 1, hjust = 1),
        axis.text.y = element_text(size=10),
        plot.title = element_text(color = "black", size=12))
```



r health among older adults with doctor−diagnosed arthritis

ɜ of older adults ever told they have arthritis

ɜ of older adults getting sufficient sleep (>6 hours)

ɜ of older adults who have been told they have high blood pressure who report currently taking medication for their higl

ɜ of older adults who have fallen and sustained an injury within last year

ɜ of older adults who report having a disability (includes limitations related to sensory or mobility impairments or a phys

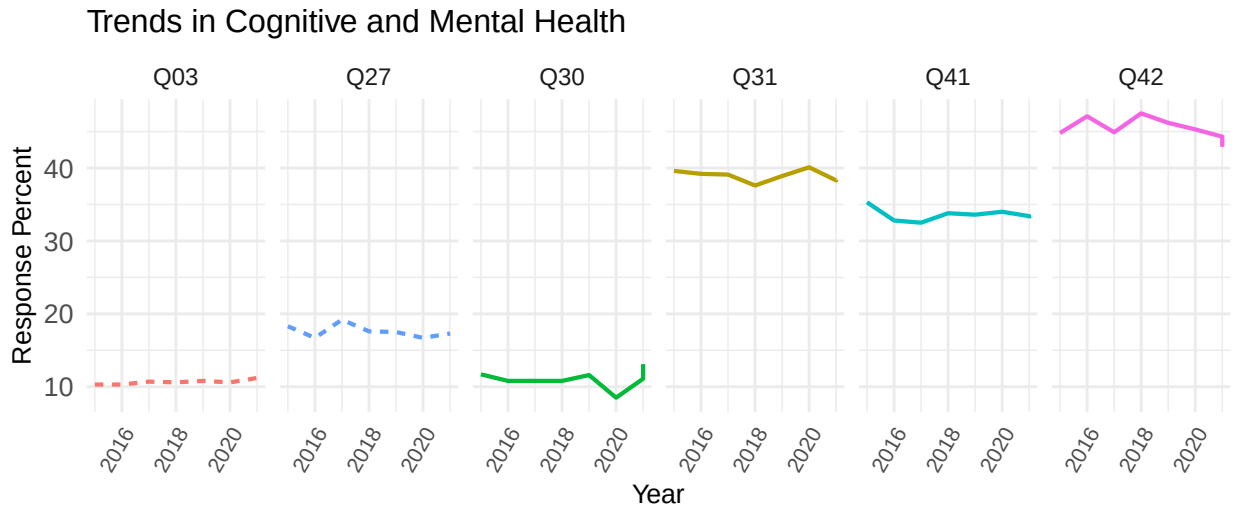
ɜ of older adults who report having lost 5 or fewer teeth due to decay or gum disease

ɜ of older adults who self−reported that their health is "fair" or "poor"

ɜ of older adults who self−reported that their health is "good", "very good", or "excellent"

ıt pain due to arthritis among older adults with doctor−diagnosed arthritis

```
# Trends in mental health and cognitive decline
brfss %>% filter(Gender_RaceID == "OVERALL" &
                      (Class == "Cognitive Decline" | Class == "Mental Health") &
                      LocationAbbr == "US" &
                      AgeGroupID == "AGE_OVERALL" &
                      Data_Value_Type == "Percentage") %>%
  ggplot(aes(x = YearEnd, y = Data_Value, color = Question, linetype = Class)) +
  geom_line(linewidth = 0.75) +
  facet_grid( ~QuestionID) +
  theme_minimal() +
  guides(color = guide_legend(ncol = 1)) +
  ggtitle("Trends in Cognitive and Mental Health") +
  xlab("Year") +
  ylab("Response Percent") +
  theme(legend.position = "bottom",
        axis.title.x = element_text(color="black",size=10),
        axis.title.y = element_text(color="black", size=10),
        axis.text.x = element_text(size=8, angle = 60, vjust = 1, hjust = 1),
        axis.text.y = element_text(size=10),
        plot.title = element_text(color = "black", size=12))
```



eriencing frequent mental distress

l subjective cognitive decline or memory loss that interferes with their ability to engage in social activities or household

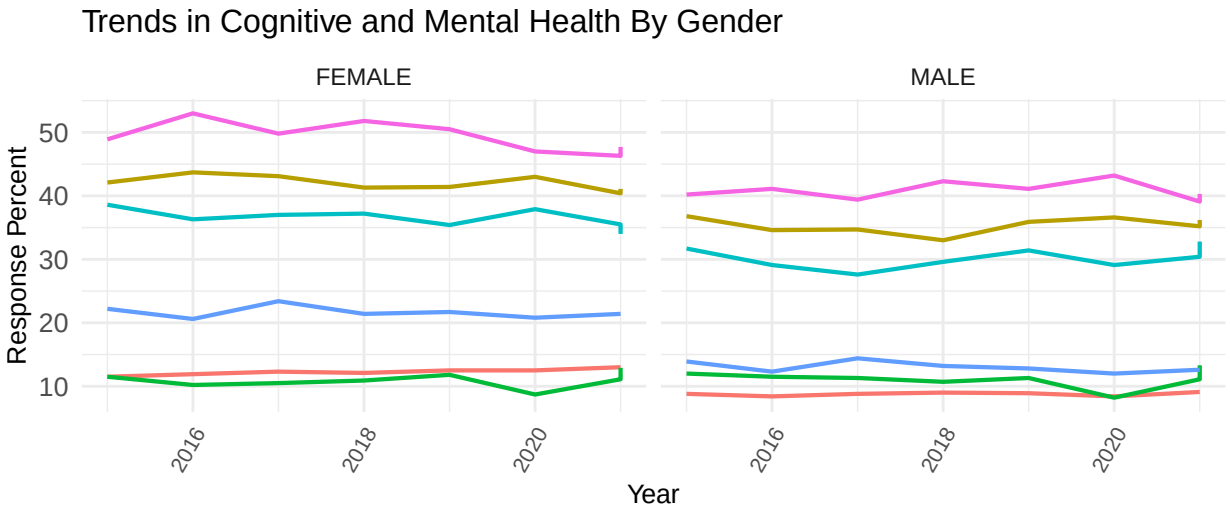
l subjective cognitive decline or memory loss that is happening more often or is getting worse in the preceding 12 mon

l that as a result of subjective cognitive decline or memory loss that they need assistance with day−to−day activities

e diagnosis of depression

ve cognitive decline or memory loss who reported talking with a health care professional about it

```
brfss %>% filter((Gender_RaceID == "FEMALE" | Gender_RaceID == "MALE") &
                     (Class == "Cognitive Decline" | Class == "Mental Health") &
                     LocationAbbr == "US" &
                     AgeGroupID == "AGE_OVERALL" &
                     Data_Value_Type == "Percentage") %>%
  ggplot(aes(x = YearEnd, y = Data_Value, color = Question)) +
  geom_line(linewidth = 0.75) +
  facet_grid( ~Gender_RaceID) +
  theme_minimal() +
  guides(color = guide_legend(ncol = 1)) +
  ggtitle("Trends in Cognitive and Mental Health By Gender") +
  xlab("Year") +
  ylab("Response Percent") +
  theme(legend.position = "bottom",
        axis.title.x = element_text(color="black",size=10),
        axis.title.y = element_text(color="black", size=10),
        axis.text.x = element_text(size=8, angle = 60, vjust = 1, hjust = 1),
        axis.text.y = element_text(size=10),
        plot.title = element_text(color = "black", size=12))
```



e of older adults who are experiencing frequent mental distress

e of older adults who reported subjective cognitive decline or memory loss that interferes with their ability to engage in

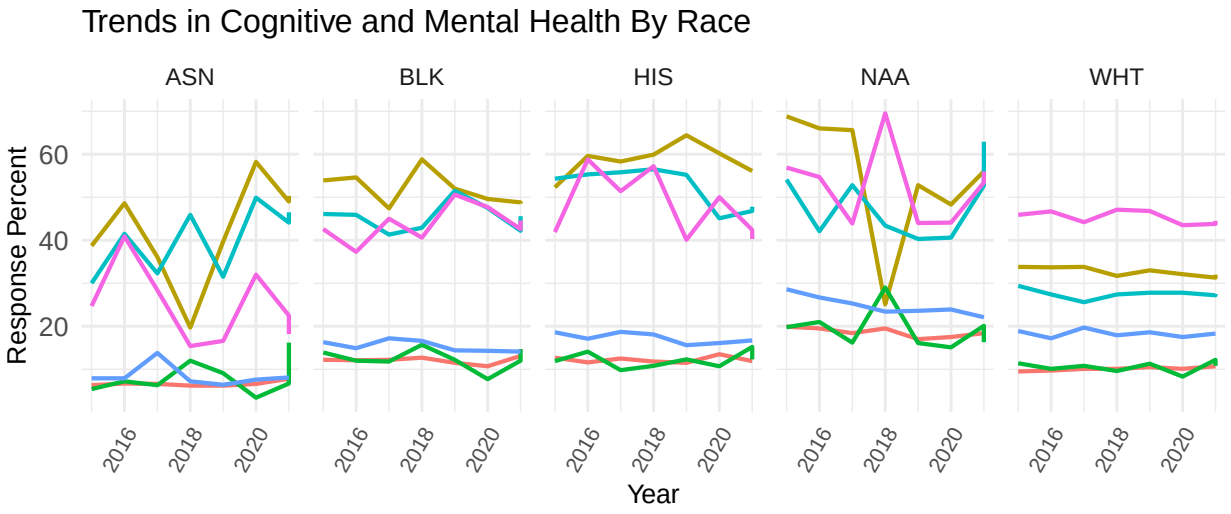
e of older adults who reported subjective cognitive decline or memory loss that is happening more often or is getting w

e of older adults who reported that as a result of subjective cognitive decline or memory loss that they need assistance

e of older adults with a lifetime diagnosis of depression

e of older adults with subjective cognitive decline or memory loss who reported talking with a health care professional a

```
brfss %>% filter((Gender_RaceID == "HIS" | Gender_RaceID == "NAA" |
                    Gender_RaceID == "ASN" | Gender_RaceID == "BLK" |
                    Gender_RaceID == "WHT") &
                   (Class == "Cognitive Decline" | Class == "Mental Health") &
                   LocationAbbr == "US" &
                   AgeGroupID == "AGE_OVERALL" &
                   Data_Value_Type == "Percentage") %>%
  ggplot(aes(x = YearEnd, y = Data_Value, color = Question)) +
  geom_line(linewidth = 0.75) +
  facet_grid( ~Gender_RaceID) +
  theme_minimal() +
  guides(color = guide_legend(ncol = 1)) +
  ggtitle("Trends in Cognitive and Mental Health By Race") +
  xlab("Year") +
  ylab("Response Percent") +
  theme(legend.position = "bottom",
        axis.title.x = element_text(color="black",size=10),
        axis.title.y = element_text(color="black", size=10),
        axis.text.x = element_text(size=8, angle = 60, vjust = 1, hjust = 1),
        axis.text.y = element_text(size=10),
        plot.title = element_text(color = "black", size=12))
```



: of older adults who are experiencing frequent mental distress

: of older adults who reported subjective cognitive decline or memory loss that interferes with their ability to engage in

: of older adults who reported subjective cognitive decline or memory loss that is happening more often or is getting w

: of older adults who reported that as a result of subjective cognitive decline or memory loss that they need assistance

: of older adults with a lifetime diagnosis of depression

: of older adults with subjective cognitive decline or memory loss who reported talking with a health care professional a

```
# Created a function for the rest of the plots
class_plot <- function(class_name){

  overall_plot <- brfss %>% filter(Gender_RaceID == "OVERALL" &
                     Class == class_name &
                     LocationAbbr == "US" &
                     AgeGroupID == "AGE_OVERALL" &
                     Data_Value_Type == "Percentage") %>%
    ggplot(aes(x = YearEnd, y = Data_Value, color = Question)) +
    geom_line() +
    facet_grid( ~QuestionID) +
    theme_light() +
    theme(legend.position = "bottom") +
    guides(color = guide_legend(ncol = 1))


  gender_plot <- brfss %>% filter((Gender_RaceID == "FEMALE" | Gender_RaceID == "MALE") &
                     Class == class_name &
                     LocationAbbr == "US" &
                     AgeGroupID == "AGE_OVERALL" &
                     Data_Value_Type == "Percentage") %>%
    ggplot(aes(x = YearEnd, y = Data_Value, color = Question)) +
    geom_line() +
    facet_grid( ~Gender_RaceID) +
    theme_light() +
    theme(legend.position = "bottom") +
```

```
    guides(color = guide_legend(ncol = 1))

  race_plot <- brfss %>% filter((Gender_RaceID == "HIS" | Gender_RaceID == "NAA" |
                       Gender_RaceID == "ASN" | Gender_RaceID == "BLK" |
                       Gender_RaceID == "WHT") &
                         Class == class_name &
                         LocationAbbr == "US" &
                         AgeGroupID == "AGE_OVERALL" &
                         Data_Value_Type == "Percentage") %>%
    ggplot(aes(x = YearEnd, y = Data_Value, color = Question)) +
    geom_line() +
    facet_grid( ~Gender_RaceID) +
    theme_light() +
    theme(legend.position = "bottom") +
    guides(color = guide_legend(ncol = 1))

  list(overall_plot, gender_plot, race_plot)
}
```

```
# Run the function to create the list
plot_list = list()

for (i in unique(brfss$Class)[c(2,4,5,6)]){
  plot_list[[i]] <- class_plot(i)
}
```

```
unique(brfss$Class)[c(2,4,5,6)]
```

```
## [1] "Nutrition/Physical Activity/Obesity" "Screenings and Vaccines"
## [3] "Smoking and Alcohol Use"             "Caregiving"
```

Uncomment below sections to see graphs

```
# plot_list[["Nutrition/Physical Activity/Obesity"]]
```

```
# plot_list[["Screenings and Vaccines"]]
```

```
# plot_list[["Smoking and Alcohol Use"]]
```

```
# plot_list[["Caregiving"]]
```

STATISTICAL ANALLYSIS

Mortality rate of Alzheimer's across the state

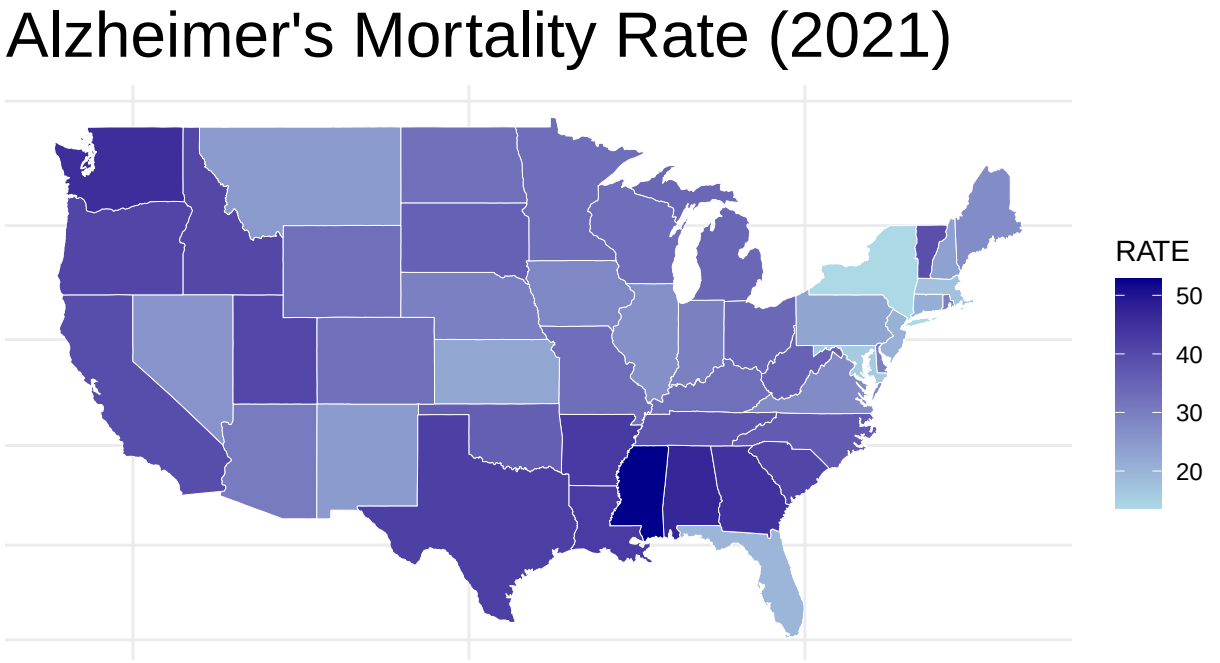
```
# https://www.cdc.gov/nchs/pressroom/sosmap/alzheimers_mortality/alzheimers_disease.htm

# https://www.alz.org/alzheimers-dementia/facts-figures

alz_state <- read_csv("https://raw.githubusercontent.com/D-hartog/DATA607/main/FINAL_PROJ/Alzheimer's_m
```

```
## Rows: 450 Columns: 5
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): STATE, URL
## dbl (2): YEAR, RATE
## num (1): DEATHS
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
alz_state$YEAR <- as.numeric(alz_state$YEAR)

alz_state <-  alz_state %>% filter(YEAR == 2021)

unique(population$Location)[c(-10, - 1)]
```

```
##  [1] "Alabama"        "Alaska"         "Arizona"        "Arkansas"
##  [5] "California"     "Colorado"       "Connecticut"    "Delaware"
##  [9] "Florida"        "Georgia"        "Hawaii"         "Idaho"
## [13] "Illinois"       "Indiana"        "Iowa"           "Kansas"
## [17] "Kentucky"       "Louisiana"      "Maine"          "Maryland"
## [21] "Massachusetts"  "Michigan"       "Minnesota"      "Mississippi"
## [25] "Missouri"       "Montana"        "Nebraska"       "Nevada"
## [29] "New Hampshire"  "New Jersey"     "New Mexico"     "New York"
## [33] "North Carolina" "North Dakota"   "Ohio"           "Oklahoma"
## [37] "Oregon"         "Pennsylvania"   "Rhode Island"   "South Carolina"
## [41] "South Dakota"   "Tennessee"      "Texas"          "Utah"
## [45] "Vermont"        "Virginia"       "Washington"     "West Virginia"
## [49] "Wisconsin"      "Wyoming"
```

```
alz_state$Location <- unique(population$Location)[c(-10, - 1)]


states <- map_data("state")
states$region <- str_to_title(states$region)

states <- states[order(states$order),]

alz_state %>%
  merge(states, by.x = "Location", by.y = "region") %>%
  arrange(group, order) %>%
  ggplot(aes(x = long, y = lat)) +
  geom_polygon(aes(group = group, fill = RATE), color = "white", linewidth = 0.1) +
  coord_map() +
  scale_fill_gradient(low = "lightblue", high = "darkblue") +
  theme_minimal() +
  ggtitle("Alzheimer's Mortality Rate (2021)") +
  theme(axis.title.x = element_blank(),
        axis.title.y = element_blank(),
        axis.text.x = element_blank(),
        axis.text.y = element_blank(),
        plot.title = element_text(color = "black", size=24))
```




```
# ALZHEIMERS MORTALITY BY STATE
#
alz_data <- read_csv("https://raw.githubusercontent.com/D-hartog/DATA607/main/FINAL_PROJ/Alzheimer's_mo
```

```
## Rows: 450 Columns: 5
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): STATE, URL
## dbl (2): YEAR, RATE
## num (1): DEATHS
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
head(alz_data)
```

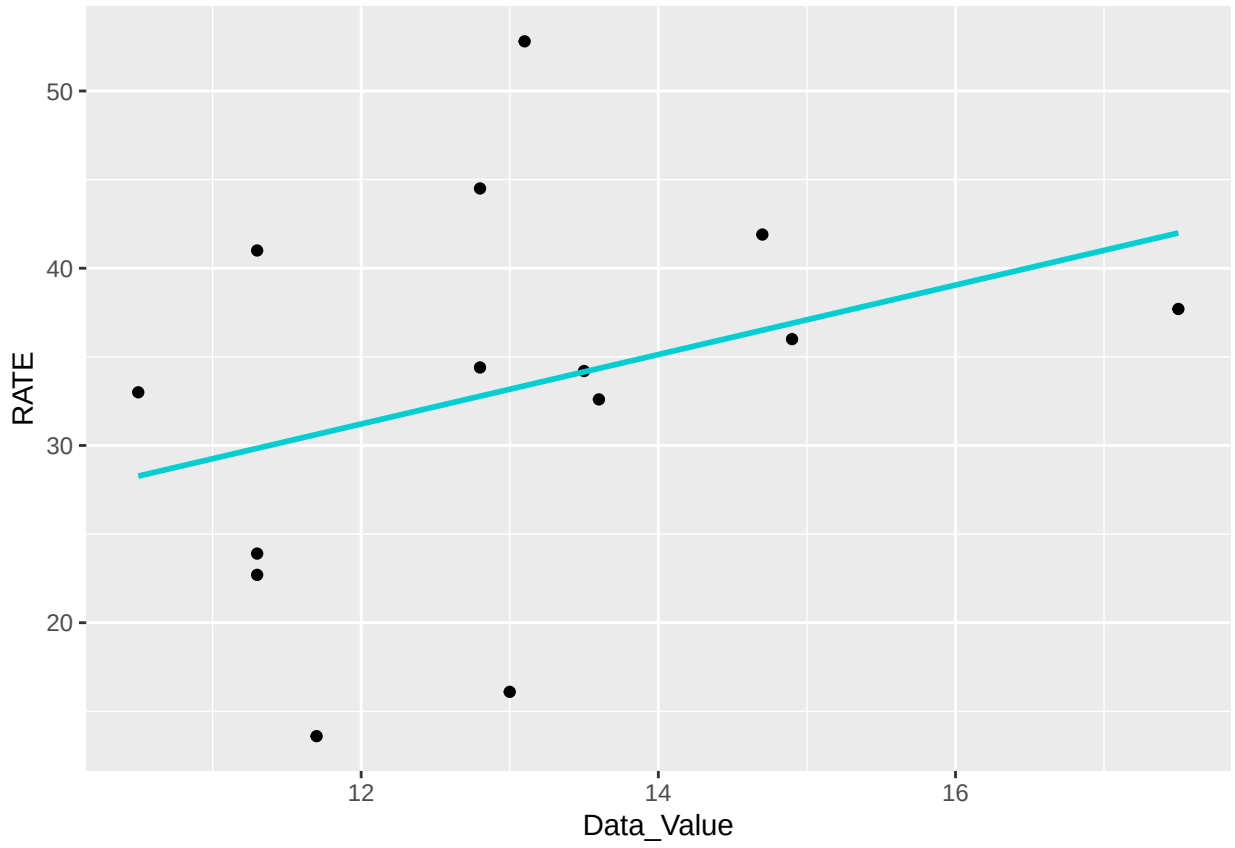
```
## # A tibble: 6 x 5
##    YEAR STATE  RATE DEATHS URL
##   <dbl> <chr> <dbl>  <dbl> <chr>
## 1  2021 AL     46.8   2725 /nchs/pressroom/states/alabama/al.htm
## 2  2021 AK     26.4    135 /nchs/pressroom/states/alaska/ak.htm
## 3  2021 AZ     30.5   2754 /nchs/pressroom/states/arizona/az.htm
## 4  2021 AR     43.2   1559 /nchs/pressroom/states/arkansas/ar.htm
## 5  2021 CA     39.5  16911 /nchs/pressroom/states/california/ca.htm
## 6  2021 CO     32.6   1778 /nchs/pressroom/states/colorado/co.htm
```

```r
# https://www.alz.org/alzheimers-dementia/facts-figures

`%nin%` <-  Negate(`%in%`)

sub_cog <- brfss %>% filter(QuestionID == "Q30", YearStart == 2021, AgeGroupID == "AGE_OVERALL",
                 Gender_RaceID == "OVERALL", LocationAbbr %nin% c("MDW", "NRE", "US", "WEST", "SOU"))

head(sub_cog)
```

```
## # A tibble: 6 x 21
##   YearStart YearEnd LocationID LocationAbbr LocationDesc ClassID Class   TopicID
##       <dbl>   <dbl> <chr>      <chr>        <chr>        <chr>   <chr>   <chr>
## 1      2021    2021 24         MD           Maryland     C06     Cognit~ TCC01
## 2      2021    2021 48         TX           Texas        C06     Cognit~ TCC01
## 3      2021    2021 28         MS           Mississippi  C06     Cognit~ TCC01
## 4      2021    2021 40         OK           Oklahoma     C06     Cognit~ TCC01
## 5      2021    2021 47         TN           Tennessee    C06     Cognit~ TCC01
## 6      2021    2021 26         MI           Michigan     C06     Cognit~ TCC01
## # i 13 more variables: Topic <chr>, AgeGroupID <chr>, AgeGroup <chr>,
## #   Gender_RaceID <chr>, Race_Ethnicity <chr>, Gender <chr>, QuestionID <chr>,
## #   Question <chr>, Data_Value_Type <chr>, Data_Value <dbl>,
## #   Data_Value_Alt <dbl>, Data_Value_Footnote <chr>, Geolocation <chr>
```

```r
alz_data %>% filter(YEAR == 2021) %>% select(YEAR, STATE, RATE) %>%
  merge(sub_cog, by.x = "STATE", by.y = "LocationAbbr") %>%
  ggplot(aes(x = Data_Value, y = RATE)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "darkturquoise")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
alz_data %>% filter(YEAR == 2021) %>% select(YEAR, STATE, RATE) %>%
  merge(sub_cog, by.x = "STATE", by.y = "LocationAbbr") %>%
  summarise(Correlation = cor(Data_Value, RATE, use = "complete.obs"))
```

```
##   Correlation
## 1   0.3301101
```

```
sub_cog2 <- brfss %>% filter((QuestionID == "Q30" | QuestionID == "Q31" | QuestionID == "Q41" | Questio
                             YearStart == 2021, AgeGroupID == "AGE_OVERALL",
                             Gender_RaceID == "OVERALL",
                             LocationAbbr %nin% c("MDW", "NRE", "US", "WEST", "SOU"))

sub_cog2 <- sub_cog2 %>% pivot_wider(id_cols = LocationAbbr, names_from = QuestionID, values_from = Data

new_data <- alz_data %>% filter(YEAR == 2021) %>% select(YEAR, STATE, RATE) %>%
  merge(sub_cog2, by.x = "STATE", by.y = "LocationAbbr")

sub_cog2_reg <- lm(RATE ~ Q30 + Q31 + Q41 + Q42, data = new_data)
summary(sub_cog2_reg)
```

```
##
## Call:
## lm(formula = RATE ~ Q30 + Q31 + Q41 + Q42, data = new_data)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.0264  -3.7591   0.1571   5.9383  15.4291
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -10.5769    40.7240  -0.260    0.801
## Q30           1.7383     2.1577   0.806    0.441
## Q31          -0.1658     0.9719  -0.171    0.868
## Q41           0.4532     0.7999   0.567    0.585
## Q42           0.3013     0.8323   0.362    0.726
##
## Residual standard error: 11.9 on 9 degrees of freedom
## Multiple R-squared:  0.1797, Adjusted R-squared:  -0.1849
## F-statistic: 0.4929 on 4 and 9 DF,  p-value: 0.7417
```

CONCLUSIONS

Observations from the data:

- There are clear trends in the growth of the aging population with over 25% of the US population in 2022 being 55 or older. We can probably expect that number to continue to rise.
- Trends the mortality rate in the US spiked during COVID and continued to rise in 2021.
- COVID-19 and Chronic liver disease over took Strokes and the Flu in the top 10 leading causes of death in the US during 2020 and 2021.
- Since 1999 there appears to be a steady climb in the rates of deaths from Alzheimer's and Unintentional injuries.
- Trends in mortality for age groups over 65 stayed relatively unchanged from 2021 to 2022.
- Alzheimer's Disease was not a top 10 leading cause of death in the age group 65 - 74 years old but historically it has been in the top 10 leading causes of death across the US. (however it was the 11th leading cause)
- Regarding overall health, 60% of respondents report that they have high blood pressure and are on blood pressure medications.
- Over 50% of the respondents in the BRFSS survey reported that their health is "good", "very good", "excellent"
- Only 50% of respondents who reported subjective cognitive decline or memory loss have spoken to a medical professional about it.
- 40-60% of Black, Hispanic, and Native American/Alaskan Natives reported that subjective cognitive decline has interfered with their social and daily activities; Compare with 25-30% of White respondents.
- Very weak to correlations exist between a states Alzheimer's death rate and the percentage of individuals who reported mild cognitive decline.

LIMITATIONS AND FUTHER CONSIDERATIONS

Limitations + Data is only up to 2021/2022

- Trends in US mortality rates overtime are age adjusted - useful for comparing to other populations
- Limited data from all states to generate any significant correlations between Alzheimer's death rates and subjective reports of cognitive decline.
- Many factors relate to cognitive decline and the onset of Alzheimer's and responses on the BFRSS might not be a good indicator of mortality

Future directions

- Since we have data on age adjusted death rates comparisons between states and other countries might be interesting
- Investigate trends from the BFRSS among the different age groups
- Comparing mortality data to other countries might shed light on effective strategies in other countries.