DSC 680 Week 1-4

Vineet mehta

2024-09-24

```
knitr::opts_chunk$set(echo = FALSE)
# Load necessary libraries
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
# Load the dataset
file_path <-
"Normal_weight__overweight__and_obesity_among_adults_aged_20_and_over__by_sel
ected_characteristics__United_States.csv"
df <- read.csv(file_path)

# Convert YEAR to numeric format for easier plotting
df$YEAR <- as.numeric(substr(gsub("-", "", df$YEAR), 1, 4))

# ----------------------------
# 1. Obesity Trends by Age Group
# ----------------------------
# Check if AGE data is available
print(unique(df$AGE))  # Diagnostic check
```

```
## [1] "20 years and over" "20-34 years"       "35-44 years"
## [4] "45-54 years"       "55-64 years"       "65-74 years"
## [7] "75 years and over"
```

```
obesity_by_age <- df %>%
  filter(PANEL == "Obesity (BMI greater than or equal to 30.0)" &
!is.na(AGE)) %>%
  select(YEAR, AGE, ESTIMATE)

# Check if the data was filtered correctly
print(head(obesity_by_age))  # Diagnostic check
```

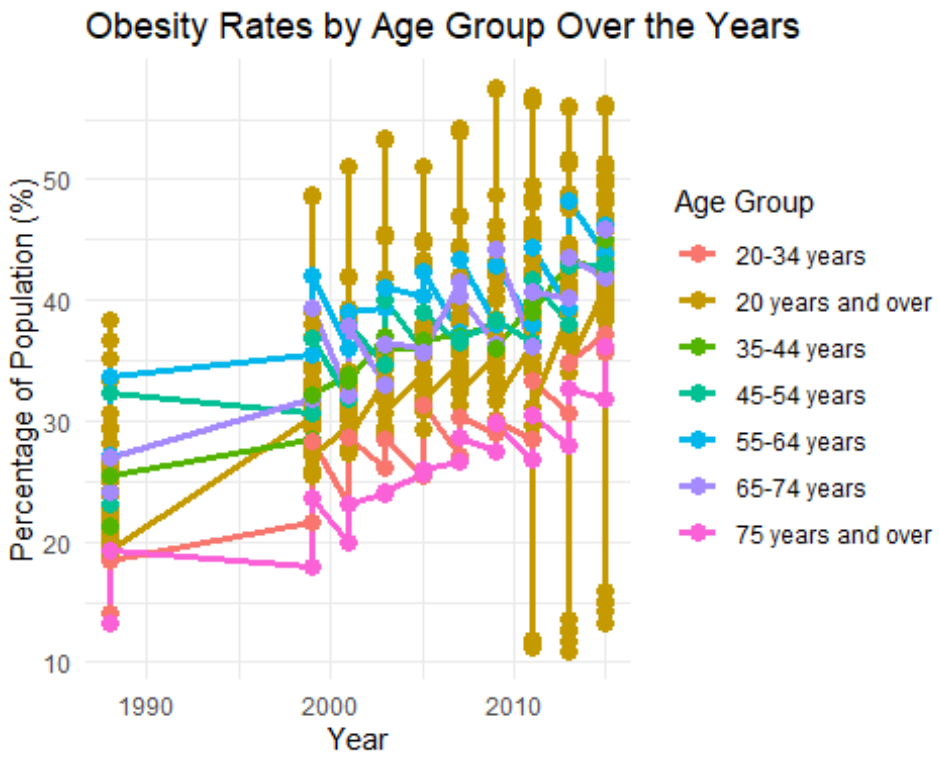
```
##   YEAR               AGE ESTIMATE
## 1 2001 20 years and over     29.5
## 2 1988 20 years and over     22.9
## 3 1999 20 years and over     30.4
## 4 2001 20 years and over     31.4
## 5 2003 20 years and over     33.4
## 6 2005 20 years and over     34.0
```

```
if (nrow(obesity_by_age) > 0) {
  p1 <- ggplot(obesity_by_age, aes(x = YEAR, y = ESTIMATE, color = AGE, group
= AGE)) +
    geom_line(size = 1.2) +
    geom_point(size = 3) +
    labs(title = "Obesity Rates by Age Group Over the Years",
         x = "Year",
         y = "Percentage of Population (%)",
         color = "Age Group") +
    theme_minimal()

  print(p1)  # Display the first plot
} else {
  print("No data available for Age Group.")
}
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## ℹ Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
## Warning: Removed 72 rows containing missing values (`geom_point()`).
```




```
# ----------------------------
# 2. Obesity Trends by Gender
# ----------------------------
# Check if gender data is available
print(unique(df$STUB_NAME))  # Diagnostic check
```

```
## [1] "Total"                                "Sex"
## [3] "Race and Hispanic origin"             "Sex and race and Hispanic origin"
## [5] "Percent of poverty level"             "Sex and age"
```

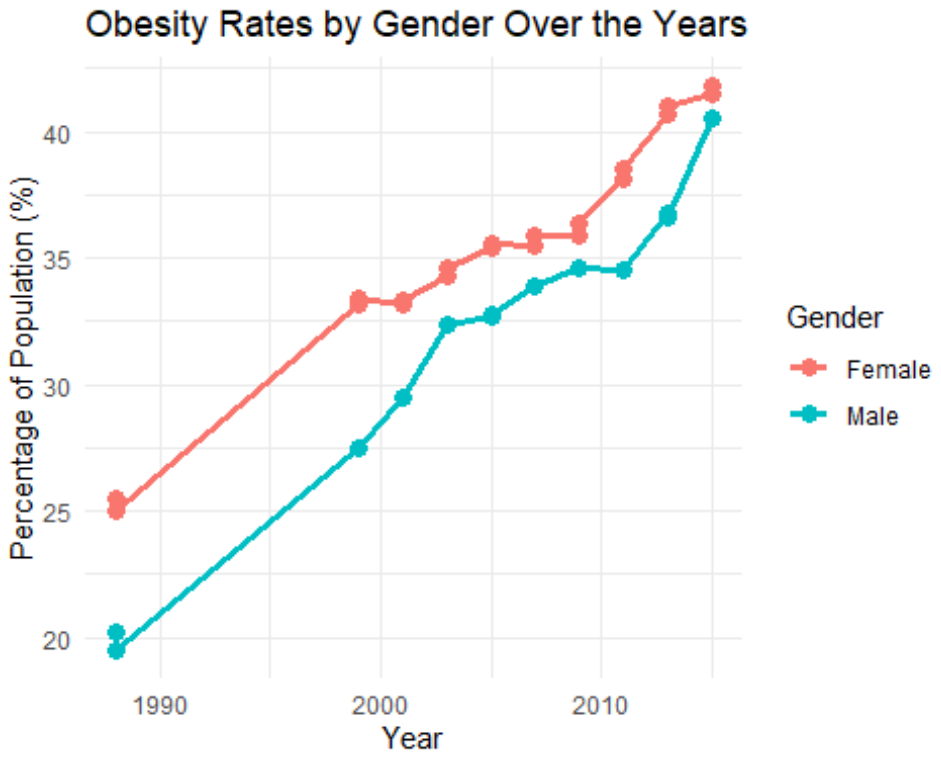
```
obesity_by_gender <- df %>%
  filter(PANEL == "Obesity (BMI greater than or equal to 30.0)" & STUB_NAME
== "Sex") %>%
  select(YEAR, STUB_LABEL, ESTIMATE)

# Check if the data was filtered correctly
print(head(obesity_by_gender))  # Diagnostic check
```

```
##   YEAR STUB_LABEL ESTIMATE
## 1 2001       Male     29.5
## 2 1988       Male     20.2
## 3 1999       Male     27.5
## 4 2001       Male     29.5
## 5 2003       Male     32.4
## 6 2005       Male     32.7
```

```r
if (nrow(obesity_by_gender) > 0) {
  p2 <- ggplot(obesity_by_gender, aes(x = YEAR, y = ESTIMATE, color =
STUB_LABEL, group = STUB_LABEL)) +
    geom_line(size = 1.2) +
    geom_point(size = 3) +
    labs(title = "Obesity Rates by Gender Over the Years",
         x = "Year",
         y = "Percentage of Population (%)",
         color = "Gender") +
    theme_minimal()

  print(p2)  # Display the second plot
} else {
  print("No data available for Gender.")
}
```




```r
# ----------------------------
# 3. Obesity Trends by Ethnicity
# ----------------------------
# Check if ethnicity data is available
print(unique(df$STUB_LABEL))  # Diagnostic check
```

```
##  [1] "20 years and over"
##  [2] "Male"
##  [3] "Female"
##  [4] "Not Hispanic or Latino: White only"
##  [5] "Male: Not Hispanic or Latino: White only"
```

```
##  [6] "Female: Not Hispanic or Latino: White only"
##  [7] "Not Hispanic or Latino: Black or African American only"
##  [8] "Male: Not Hispanic or Latino: Black or African American only"
##  [9] "Female: Not Hispanic or Latino: Black or African American only"
## [10] "Not Hispanic or Latino: Asian only"
## [11] "Male: Not Hispanic or Latino: Asian only"
## [12] "Female: Not Hispanic or Latino: Asian only"
## [13] "Hispanic or Latino: All races"
## [14] "Male: Hispanic or Latino: All races"
## [15] "Female: Hispanic or Latino: All races"
## [16] "Hispanic or Latino: Mexican origin: All races"
## [17] "Male: Hispanic or Latino: Mexican origin: All races"
## [18] "Below 100%"
## [19] "Female: Hispanic or Latino: Mexican origin: All races"
## [20] "100%-199%"
## [21] "200%-399%"
## [22] "400% or more"
## [23] "Male: 20-34 years"
## [24] "Male: 35-44 years"
## [25] "Male: 45-54 years"
## [26] "Male: 55-64 years"
## [27] "Male: 65-74 years"
## [28] "Male: 75 years and over"
## [29] "Female: 20-34 years"
## [30] "Female: 35-44 years"
## [31] "Female: 45-54 years"
## [32] "Female: 55-64 years"
## [33] "Female: 65-74 years"
## [34] "Female: 75 years and over"
```

```
obesity_by_ethnicity <- df %>%
  filter(PANEL == "Obesity (BMI greater than or equal to 30.0)" & STUB_NAME
== "Race and Hispanic origin") %>%
  select(YEAR, STUB_LABEL, ESTIMATE)

# Check if the data was filtered correctly
print(head(obesity_by_ethnicity))  # Diagnostic check
```

```
##   YEAR                         STUB_LABEL ESTIMATE
## 1 1988 Not Hispanic or Latino: White only     21.6
## 2 1999 Not Hispanic or Latino: White only     29.4
## 3 2001 Not Hispanic or Latino: White only     30.5
## 4 2003 Not Hispanic or Latino: White only     32.0
## 5 2005 Not Hispanic or Latino: White only     32.8
## 6 2007 Not Hispanic or Latino: White only     33.3
```

```
if (nrow(obesity_by_ethnicity) > 0) {
  p3 <- ggplot(obesity_by_ethnicity, aes(x = YEAR, y = ESTIMATE, color =
STUB_LABEL, group = STUB_LABEL)) +
    geom_line(size = 1.2) +
```

```
    geom_point(size = 3) +
    labs(title = "Obesity Rates by Ethnicity Over the Years",
         x = "Year",
         y = "Percentage of Population (%)",
         color = "Ethnicity") +
    theme_minimal()

  print(p3)  # Display the third plot
} else {
  print("No data available for Ethnicity.")
}
```

```
## Warning: Removed 24 rows containing missing values (`geom_line()`).
```

```
## Warning: Removed 24 rows containing missing values (`geom_point()`).
```