Midterm Qs1

Anous Ahmed

2025-10-17

Part I: Civic Issue – Obesity in the United States

Obesity is a major health concern in the United States. It is reported that obesity contributes to over $170 billion in annual medical costs. This is a grave strain on the healthcare system that deepens inequalities and increases healthcare costs (CDC, 2024). This includes billions in extra costs to the Medicare and Medicaid programs. According to the World Health Organization (WHO), obesity is defined as having a body mass index (BMI) of 30 kg/m2 or higher for adults (Chusan et al., 2025). The prevalence of this disease is a significant health concern, contributing to major complications like diabetes, heart disease, high blood pressure, sleep apnea, and cancer. It also contributes to poor mental health, reduced quality of life, and greater levels of disability (Mayo, 2023). According to the Centers for Disease Control and Prevention (CDC), the prevalence of obesity among adults aged 20 and older in the U.S. was 41.9% from 2017 to 2020. This equates to over 100 million individuals (CDC, 2024). It is further projected that by 2035, obesity rates will rise even further, reaching 25% among adults and 19% among youth (Chusan et al., 2025).

One of the major causes of obesity in the United States is aggressive marketing of unhealthy food and beverages. It is reported that major Food & Beverage companies spend nearly $14 billion annually on advertising fast food, sugary drinks, candy, and processed snacks (The State of Obesity, 2023). On the other hand, socioeconomic barriers are another fundamental reason as to why there is an obesity epidemic in the U.S. Individuals in low-income communities are deprived of reliable nutritional information. Hence, they cannot access a healthy nutrition plan. Moreover, areas with limited or no public transportation prevent individuals from accessing grocery stores with fresh produce or local fitness centers. Hence, leading to weight gain, and in excessive cases, obesity.

Children and adolescents with obesity are likely to remain obese into adulthood, perpetuating a cycle of poor health outcomes and social disadvantages. This has a negative impact on the national productivity and workforce capacity. national productivity and workforce capacity. Therefore, addressing this civic problem requires an indoctrination of healthcare measures and technological solutions to cultivate sustainable lifestyles.

Part II: Web scraping

```
library(rvest)
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.4.3
```

```
## Warning: package 'ggplot2' was built under R version 4.4.2
```

```
## Warning: package 'tidyr' was built under R version 4.4.2
```

```
## Warning: package 'dplyr' was built under R version 4.4.3
```

```
## Warning: package 'forcats' was built under R version 4.4.2
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ----------------------------------------- tidyverse_conflicts() --
## x dplyr::filter()         masks stats::filter()
## x readr::guess_encoding() masks rvest::guess_encoding()
## x dplyr::lag()            masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
url <- "https://www.scrapethissite.com/pages/simple/"
html <- read_html(url)
```

```r
block <- html |>
  html_elements("div.col-md-4.country")
block
```

```
## {xml_nodeset (250)}
##  [1] <div class="col-md-4 country">\n                    <h3 class="count ...
##  [2] <div class="col-md-4 country">\n                    <h3 class="count ...
##  [3] <div class="col-md-4 country">\n                    <h3 class="count ...
##  [4] <div class="col-md-4 country">\n                    <h3 class="count ...
##  [5] <div class="col-md-4 country">\n                    <h3 class="count ...
##  [6] <div class="col-md-4 country">\n                    <h3 class="count ...
##  [7] <div class="col-md-4 country">\n                    <h3 class="count ...
##  [8] <div class="col-md-4 country">\n                    <h3 class="count ...
##  [9] <div class="col-md-4 country">\n                    <h3 class="count ...
## [10] <div class="col-md-4 country">\n                    <h3 class="count ...
## [11] <div class="col-md-4 country">\n                    <h3 class="count ...
## [12] <div class="col-md-4 country">\n                    <h3 class="count ...
## [13] <div class="col-md-4 country">\n                    <h3 class="count ...
## [14] <div class="col-md-4 country">\n                    <h3 class="count ...
## [15] <div class="col-md-4 country">\n                    <h3 class="count ...
## [16] <div class="col-md-4 country">\n                    <h3 class="count ...
## [17] <div class="col-md-4 country">\n                    <h3 class="count ...
## [18] <div class="col-md-4 country">\n                    <h3 class="count ...
## [19] <div class="col-md-4 country">\n                    <h3 class="count ...
## [20] <div class="col-md-4 country">\n                    <h3 class="count ...
## ...
```

```r
Country = block |>
  html_element("h3.country-name") |>
  html_text2()
head (Country, 10)
```

```
##  [1] "Andorra"              "United Arab Emirates" "Afghanistan"
```

```
##  [4] "Antigua and Barbuda"  "Anguilla"              "Albania"
##  [7] "Armenia"               "Angola"                "Antarctica"
## [10] "Argentina"
```

```
Capital <- block |>
  html_element("span.country-capital") |>
  html_text2()
head (Capital, 10)
```

```
##  [1] "Andorra la Vella" "Abu Dhabi"        "Kabul"            "St. John's"
##  [5] "The Valley"       "Tirana"           "Yerevan"          "Luanda"
##  [9] "None"             "Buenos Aires"
```

```
Population <- block %>%
  html_element("span.country-population") %>%
  html_text2()
head (Population, 10)
```

```
##  [1] "84000"    "4975593"  "29121286" "86754"    "13254"    "2986952"
##  [7] "2968000"  "13068161" "0"        "41343201"
```

```
Area <- block %>%
  html_element("span.country-area") %>%
  html_text2()
head(Area, 10)
```

```
##  [1] "468.0"     "82880.0"   "647500.0"  "443.0"     "102.0"     "28748.0"
##  [7] "29800.0"   "1246700.0" "1.4E7"     "2766890.0"
```

```
df = tibble(
  Country,
  Capital,
  Population,
  Area
)
```

```
head(df, 10)
```

```
## # A tibble: 10 x 4
##    Country              Capital         Population Area
##    <chr>                <chr>           <chr>      <chr>
##  1 Andorra              Andorra la Vella 84000     468.0
##  2 United Arab Emirates Abu Dhabi       4975593    82880.0
##  3 Afghanistan          Kabul           29121286   647500.0
##  4 Antigua and Barbuda  St. John's      86754      443.0
##  5 Anguilla             The Valley      13254      102.0
##  6 Albania              Tirana          2986952    28748.0
##  7 Armenia              Yerevan         2968000    29800.0
##  8 Angola               Luanda          13068161   1246700.0
##  9 Antarctica           None            0          1.4E7
## 10 Argentina            Buenos Aires    41343201   2766890.0
```