Data 608 Story 3

Kevin Kirby

2025-03-04

```
library(tidyverse)
library(readxl)
library(httr)
library(ggplot2)
library(dplyr)
library(viridis)
```

Overview

The assignment states: "Do differences in per capita healthcare spending correlate with adult obesity prevalence? Your task is to analyze the provided obesity data and source additional data on healthcare spending to address this question."

I am to collect state-level per capita health spending and combine it with a provided obesity dataset to answer the above question.

Data preparation

I will import the datasets and create dataframes based on them. I used the Kaiser dataset 'Health Care Expenditures per Capita by State of Residence' to get state health expenditures. This data is refreshed every five years and the below is based on 2020 data.

```
obesity_xls <- 'https://storage.googleapis.com/data_science_masters_files/adult%20obesity.xlsx'
statehealth_csv <- 'https://storage.googleapis.com/data_science_masters_files/statehealth_percapita.csv

obesity_t <- tempfile(fileext = ".xlsx")
GET(obsesity_xls, write_disk(obsesity_t, overwrite = TRUE))
```

```
## Response [https://storage.googleapis.com/data_science_masters_files/adult%20obesity.xlsx]
##   Date: 2025-03-04 17:13
##   Status: 200
##   Content-Type: application/vnd.openxmlformats-officedocument.spreadsheetml.sheet
##   Size: 10.7 kB
## <ON DISK>  /var/folders/z2/2cg34qq110n7hgx5hd3b2rnr0000gn/T//Rtmp1CZIAQ/file7571768e7c4.xlsx
```

```
obesity_df <- read_excel(obsesity_t)

statehealth_df <- read_csv(statehealth_csv)
```

Data merging and quintiles

The two dataframes are merged based on state name, with anything not matching being dropped. I used `ntile()` from `dplyr` to create the quntiles, which are assigned as values 1-5.

```r
state_health_df <- merge(obesity_df, statehealth_df, by.x = "State", by.y = "Location")

state_health_df <- state_health_df %>%
  rename(
    state_name = State,
    obesity_rank = Rank,
    obesity_pct = `Obesity %`,
    per_cap_health_spend = `Health Spending per Capita`
  ) %>%
  mutate(spend_quintile = ntile(per_cap_health_spend, 5))

state_health_df
```

```
##              state_name obesity_rank obesity_pct per_cap_health_spend
## 1               Alabama            5       0.392               $9,280
## 2                Alaska           17       0.352              $13,642
## 3               Arizona           27       0.319               $8,756
## 4              Arkansas            3       0.400               $9,338
## 5            California           39       0.277              $10,299
## 6              Colorado           42       0.249               $8,583
## 7           Connecticut           35       0.294              $12,489
## 8              Delaware           13       0.357              $12,899
## 9  District of Columbia           51       0.215              $14,381
## 10              Florida           34       0.301               $9,865
## 11              Georgia           18       0.350               $8,758
## 12               Hawaii           41       0.261              $10,291
## 13                Idaho           29       0.310               $8,148
## 14             Illinois           11       0.360              $10,190
## 15              Indiana            7       0.378              $10,517
## 16                 Iowa            7       0.378               $9,789
## 17               Kansas           12       0.359               $9,408
## 18             Kentucky           50       0.387              $10,257
## 19            Louisiana            2       0.399              $10,515
## 20                Maine           26       0.326              $12,077
## 21             Maryland           21       0.341              $10,839
## 22        Massachusetts           40       0.274              $13,319
## 23             Michigan           15       0.354               $9,897
## 24            Minnesota           24       0.333              $10,846
## 25          Mississippi            2       0.401               $9,394
## 26             Missouri           16       0.353               $9,921
## 27              Montana           32       0.305              $10,212
## 28             Nebraska            9       0.366              $10,514
## 29               Nevada           30       0.308               $8,348
## 30        New Hampshire           25       0.328              $11,793
## 31           New Jersey           36       0.289              $11,868
## 32           New Mexico           16       0.353               $8,902
## 33             New York           38       0.280              $14,007
## 34       North Carolina           22       0.340               $8,917
## 35         North Dakota           14       0.356              $11,301
## 36                 Ohio           10       0.364              $10,478
## 37             Oklahoma            6       0.387               $9,444
## 38               Oregon           23       0.336              $10,071
## 39         Pennsylvania           31       0.324              $11,603
## 40         Rhode Island           28       0.316              $11,694
```

```
## 41       South Carolina          11       0.360             $8,766
## 42         South Dakota          11       0.360            $12,495
## 43            Tennessee           8       0.376             $9,336
## 44                Texas          19       0.344             $8,406
## 45                 Utah          33       0.302             $7,522
## 46              Vermont          37       0.288            $12,756
## 47             Virginia          20       0.343             $9,195
## 48           Washington          31       0.306             $9,265
## 49        West Virginia           1       0.412            $12,769
## 50            Wisconsin          12       0.359             $9,982
## 51              Wyoming          24       0.333            $10,989
##    spend_quintile
## 1               4
## 2               3
## 3               4
## 4               5
## 5               1
## 6               4
## 7               2
## 8               3
## 9               3
## 10              5
## 11              4
## 12              1
## 13              3
## 14              1
## 15              1
## 16              5
## 17              5
## 18              1
## 19              1
## 20              2
## 21              1
## 22              3
## 23              5
## 24              2
## 25              5
## 26              5
## 27              1
## 28              1
## 29              3
## 30              2
## 31              2
## 32              4
## 33              3
## 34              4
## 35              2
## 36              1
## 37              5
## 38              1
## 39              2
## 40              2
## 41              4
## 42              2
```

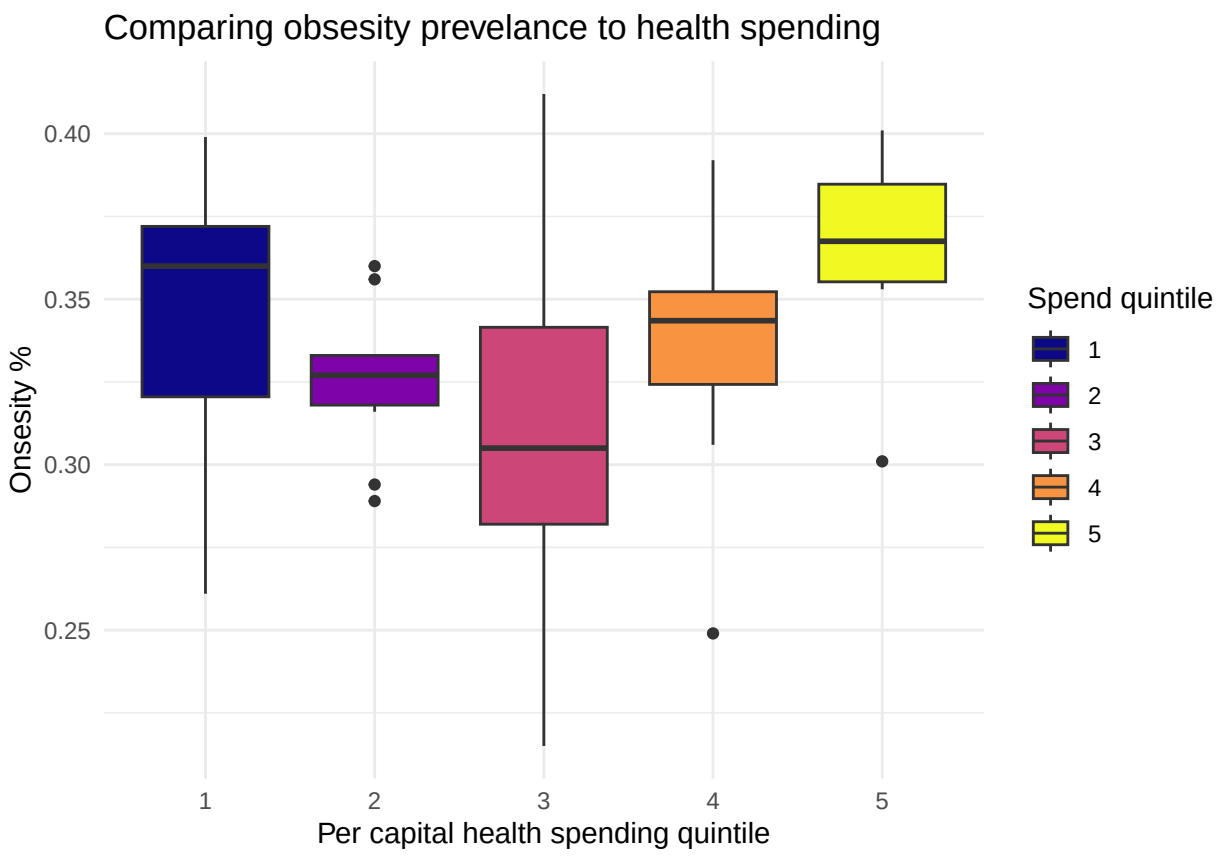
```
## 43              5
## 44              4
## 45              3
## 46              3
## 47              4
## 48              4
## 49              3
## 50              5
## 51              2
```

Analysis and visualization of data

I used obesity_pct as the health outcome metric to analyze differences across the quintiles.

Below are five box plots, one for each of the five health spending quintiles. It's interesting to see the basically non-existent relationship between health spending and obesity percent. It speaks to the generally inefficient state of American healthcare spending, where there often seems to be very little connection between spending and outcomes. If anything, you would expect there to be pressure in the reverse direction, where states are spending less because their population is generally healthy. I can determine this by seeing the seeing very similar 50% markers for both the first and last quintile. The color shading also makes it easier to scan across as accessible colors are also easier for everyone to understand.

```
ggplot(state_health_df, aes(x = factor(spend_quintile), y = obesity_pct, fill = factor(spend_quintile))
  geom_boxplot() +
  scale_fill_viridis(discrete = TRUE, option = "C") +
  labs(
    title = "Comparing obsesity prevelance to health spending",
    x = "Per capital health spending quintile",
    y = "Onsesity %",
    fill = "Spend quintile"
  ) +
  theme_minimal()
```


Comparing obsesity prevelance to health spending
Onsesity %
0.40
0.35
0.30
0.25
1
2
3
4
5
Per capital health spending quintile
Spend quintile
1
2
3
4
5