

# Investigating Healthcare Spending vs Adult Obesity (State-Level)

Cai Lin

```
library(readr)
library(readxl)
library(dplyr)
library(ggplot2)
library(scales)
library(stringr)

obesity <- read_excel("adult_obesity.xlsx", sheet = "Sheet1") |>
  rename(state = State, obesity_prev = `Obesity %`) |>
  mutate(
    state = str_trim(as.character(state)),
    obesity_prev = as.numeric(obesity_prev)
  ) |>
  filter(!is.na(state), !is.na(obesity_prev))

# Download from KFF: "Health Care Expenditures per Capita by State of Residence"

spend <- read_csv(
  "kff_per_capita_spending.csv",
  skip = 2,
  col_types = cols(
    Location = col_character(),
    `Health Spending per Capita` = col_character()
  )
) |>
  transmute(
    state = str_trim(as.character(Location)),
    spend_pc = as.numeric(str_replace_all(`Health Spending per Capita`, "[\\$,]", ""))
  ) |>
  filter(!is.na(state), !is.na(spend_pc)) |>
  # Optional: drop national total; keep or drop DC depending on your assignment convention
  filter(state != "United States")

## Warning: One or more parsing issues, call 'problems()' on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)

dat <- obesity |>
  inner_join(spend, by = "state") |>
  mutate(
    quintile = ntile(spend_pc, 5),
    quintile = factor(
```

```

    quintile,
    labels = c("Lowest 20%", "Second 20%", "Middle 20%", "Fourth 20%", "Highest 20%")
  )
)

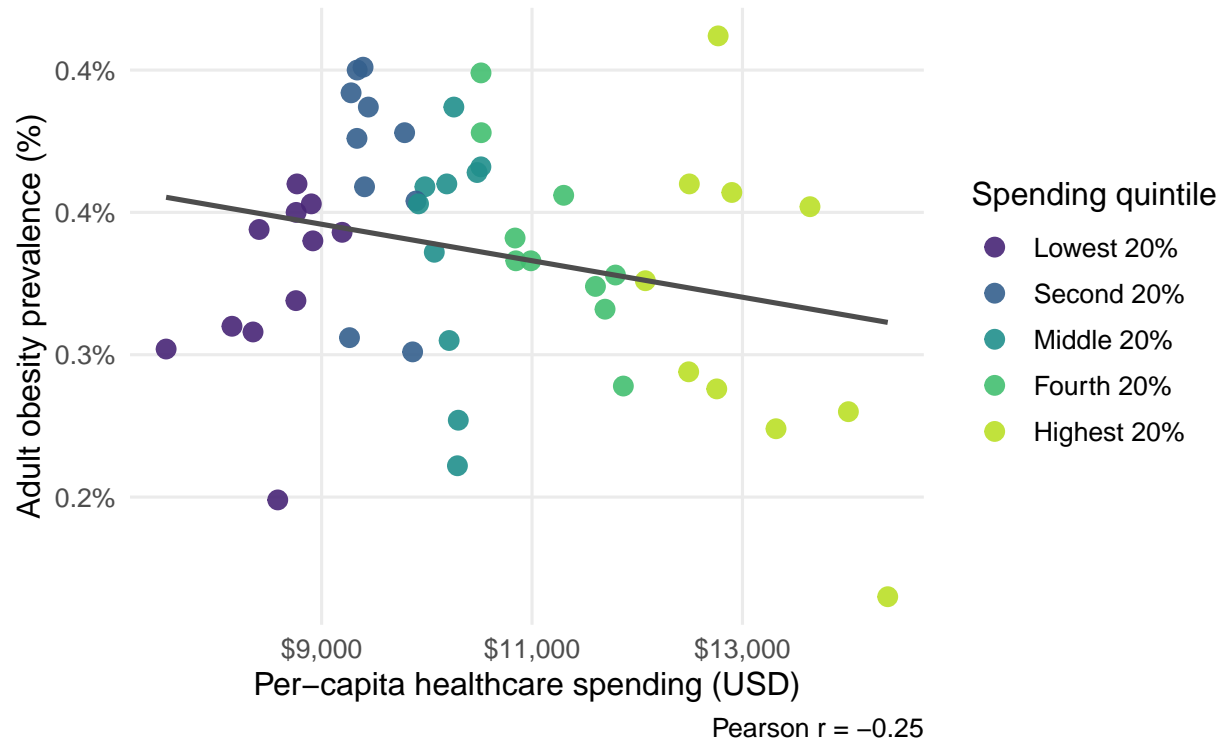
stopifnot(nrow(dat) >= 45)
r <- cor(dat$spend_pc, dat$obesity_prev, use = "complete.obs", method = "pearson")
corr_label <- paste0("Pearson r = ", number(r, accuracy = 0.01))

ggplot(dat, aes(spend_pc, obesity_prev, color = quintile)) +
  geom_point(size = 3, alpha = 0.9) +
  geom_smooth(method = "lm", se = FALSE, linewidth = 0.9, color = "grey30") +
  scale_color_viridis_d(begin = 0.1, end = 0.9) +
  scale_x_continuous(labels = label_dollar(accuracy = 1)) +
  scale_y_continuous(labels = label_percent(accuracy = 0.1, scale = 1)) +
  labs(
    title = "Do higher per-capita healthcare outlays align with adult obesity rates?",
    subtitle = "States colored by spending quintile (lowest → highest)",
    x = "Per-capita healthcare spending (USD)",
    y = "Adult obesity prevalence (%)",
    color = "Spending quintile",
    caption = corr_label
  ) +
  theme_minimal(base_size = 12) +
  theme(legend.position = "right",
        panel.grid.minor = element_blank(),
        plot.title.position = "plot")

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

```

States colored by spending quintile (lowest → highest)



## Summary and analysis

Across states, the linear fit is negative with Pearson  $r = -0.25$ , indicating that higher per-capita health spending is associated with lower adult obesity prevalence. The scatter directly addresses the research question by mapping spending (x) to obesity (y) and encoding spending quintiles with an accessible palette to show distribution across spending levels. Outliers are visible where states pair relatively low spending with low obesity or high spending with high obesity, which merits brief discussion.