ilouz_david final project

AUTHOR
David Ilouz

Summary and Introduction

Is access to broadband a function of wealth? This analysis explores the relationship between county-level economic status and broadband access, using data from the U.S. Census Bureau and the Federal Communications Commission (cited below). Below, I include code loading data, cleaning data, and performing initial analysis and visualizations.

Loading Data

Importing Census data, cited below:

- U.S. Census Bureau. 2025. “American Community Survey (ACS) 5-Year Data (2018–2022).” Census Bureau API. Accessed May 4, 2025. https://www.census.gov/data/developers/data-sets/acs-5year.html.

```
library("tidycensus")
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
```

```
library("tidyr")
library("stringr")
library("ggplot2")
library("ggtext")
library("glue")
library("knitr")

# To run this code, you'll need to load your Census API key:
# Replace "YOUR_API_KEY" with your actual key (do not share it publicly)
# census_api_key("YOUR_API_KEY", install = TRUE)

income_raw <- get_acs(
  geography = "county",
  variables = c(
    median_income = "S1901_C01_012E",
    poverty_rate = "S1701_C03_001E"),
  year = 2022)
```

```
Getting data from the 2018-2022 5-year ACS
```

```
Using the ACS Subject Tables
```

Importing Broadband data, cited below:

- Federal Communications Commission. 2024. “Fixed Broadband Summary by Geography Type.” National Broadband Map. Accessed May 4, 2025. https://broadbandmap.fcc.gov/data-download/nationwide-data.

```
#importing raw broadband access data from FCC source
broadband_raw <- read.csv("bdc_us_fixed_broadband_summary_by_geography_J24_29apr2025.csv")
```

Cleaning Data

```
#to be used for cleaning county/state names
state_lookup <- tibble::tibble(
  abbr = state.abb,
  full = state.name
```

```r
)

#cleaning income table
income_clean <- income_raw %>%
  select(NAME, variable, estimate) %>%
  distinct() %>%
  pivot_wider(
    names_from = variable,
    values_from = estimate
  ) %>%
  rename(poverty_rate = S1701_C03_001,
         median_income = S1901_C01_012) %>%
  mutate(NAME = str_replace_all(NAME, c(" County," = ",", " Parish" = ""))) %>%
  filter(str_detect(NAME, paste0(", (", paste(state_lookup$full, collapse = "|"), ")$"))) #remove

#cleaning broadband data
broadband_clean <- broadband_raw %>%
  filter(
    geography_type == "County",
    biz_res == "R",
    technology == "Any Technology"
  ) %>%
  select(
    geography_desc,
    geography_desc_full,
    speed_25_3,
    speed_100_20
  ) %>%
  mutate(
    county_name = str_remove(geography_desc_full, ", [A-Z]{2}$"),
    state_abbr = str_extract(geography_desc_full, "[A-Z]{2}$")
  )

#standardizing county and state names in broadband data
broadband_clean <- broadband_clean %>%
  left_join(state_lookup, by = c("state_abbr" = "abbr")) %>%
  mutate(NAME = paste0(county_name, ", ", full)) %>%
  select(-county_name, -state_abbr, -full, -geography_desc, -geography_desc_full) %>%
```

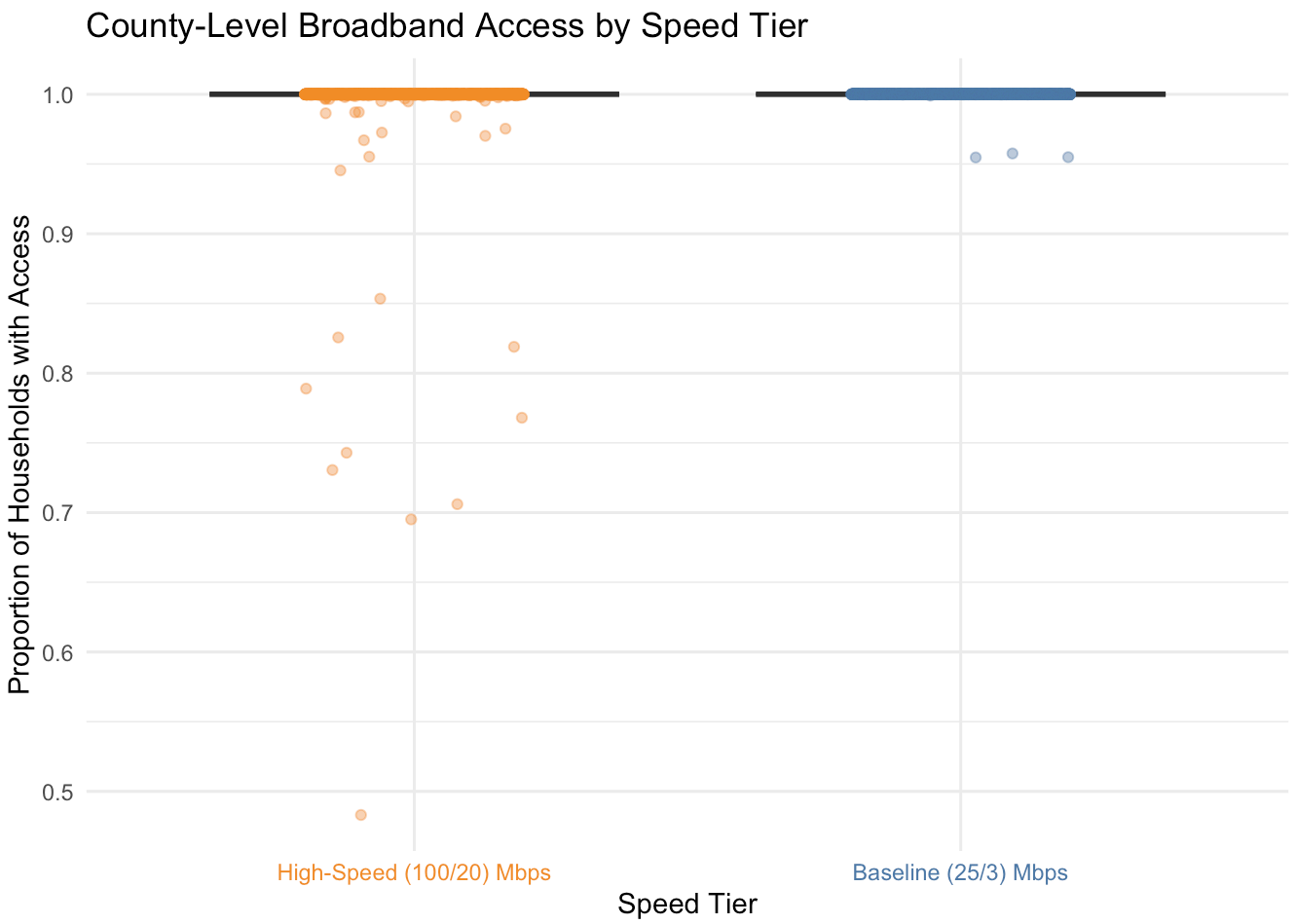
```
  group_by(NAME) %>%
  summarize(
    speed_25_3 = mean(speed_25_3, na.rm = TRUE),
    speed_100_20 = mean(speed_100_20, na.rm = TRUE) #averaging in case of duplicates
  ) %>%
   filter(str_detect(NAME, paste0(", (", paste(state_lookup$full, collapse = "|"), ")$"))) #remov

#merging datasets
merged_data <- income_clean %>%
  filter(!str_detect(NAME, ", Connecticut$")) %>%
  left_join(broadband_clean, by = "NAME") %>%
  mutate(poverty_rate = poverty_rate / 100) #turn to percentage rather than whole number

#creating a wealth score for each county
merged_data <- merged_data %>%
  mutate(
    poverty_rate = poverty_rate / 100,
    inverse_poverty = 1 - poverty_rate,
    z_income = scale(median_income)[,1],
    z_inverse_poverty = scale(inverse_poverty)[,1],
    wealth_score = z_income + z_inverse_poverty
  ) %>%
  select(-inverse_poverty, -z_income, -z_inverse_poverty)%>%
  relocate(wealth_score, .after = median_income)

#deleting unnecessary data: income_raw, broadband_raw, income_clean_broadband_clean, state lookup
rm(income_raw, broadband_raw, income_clean, broadband_clean, state_lookup)
```

Summary and Analysis of Data

Distribution of baseline and high-speed broadband access

```
# Define custom colors
tier_colors <- c("speed_25_3" = "#4E79A7",  # blue
                 "speed_100_20" = "#F28E2B")  # orange
```

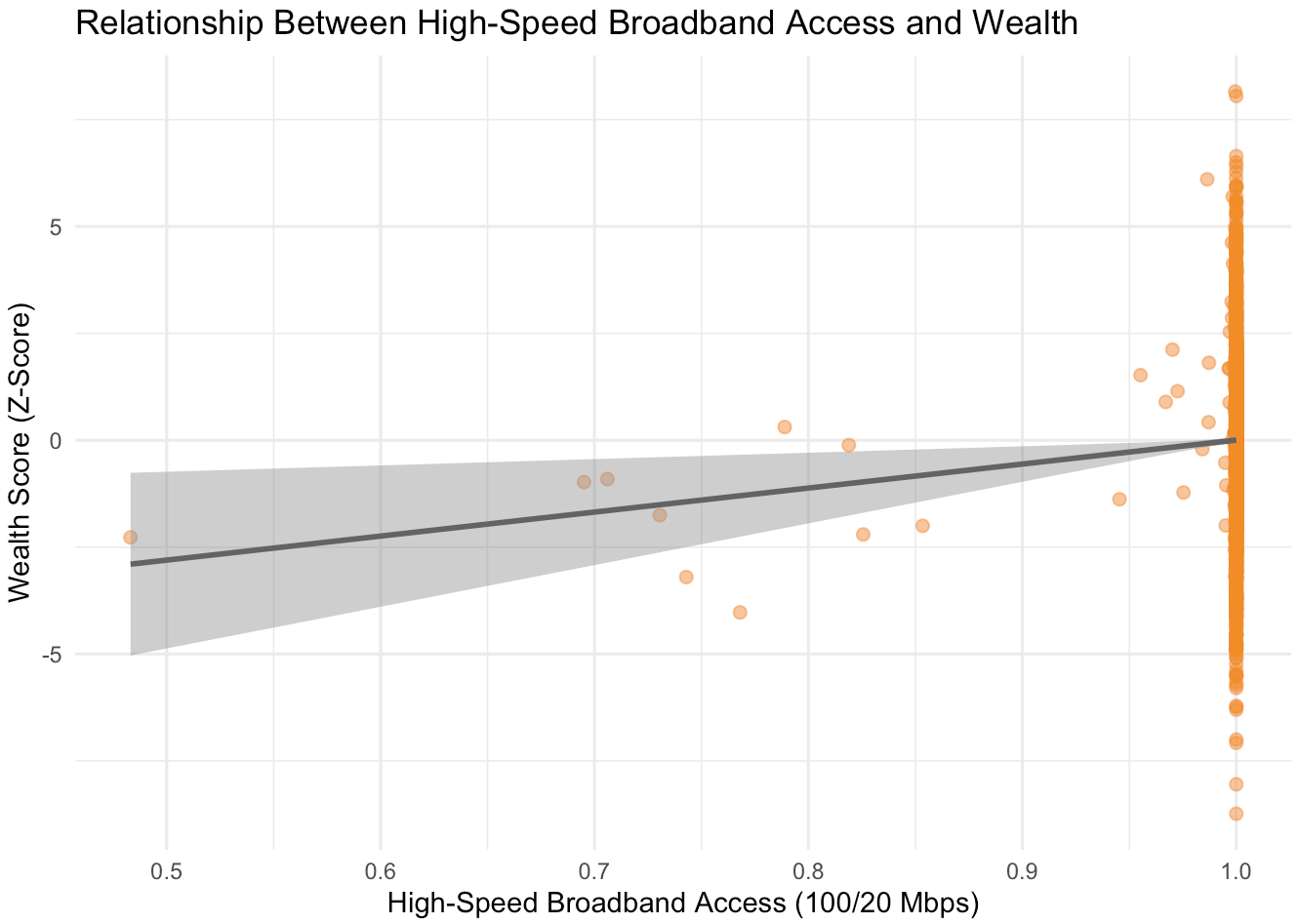
```r
long_data <- merged_data

long_data %>%
  pivot_longer(cols = c(speed_25_3, speed_100_20),
               names_to = "speed_tier", values_to = "coverage") %>%
  ggplot(aes(x = speed_tier, y = coverage, fill = speed_tier)) +
  geom_boxplot(outlier.colour = NA, alpha = 0.7) +
  geom_jitter(aes(color = speed_tier), width = 0.2, alpha = 0.4, size = 1.5) +
  scale_fill_manual(values = tier_colors,
                    labels = c("Baseline (25/3Mbps)", "High-Speed (100/20Mbps)")) +
  scale_color_manual(values = tier_colors) +
  scale_x_discrete(labels = c(
    "speed_25_3" = "<span style='color:#4E79A7;'>Baseline (25/3) Mbps</span>",
    "speed_100_20" = "<span style='color:#F28E2B;'>High-Speed (100/20) Mbps</span>"
  )) +
  labs(
    title = "County-Level Broadband Access by Speed Tier",
    x = "Speed Tier",
    y = "Proportion of Households with Access"
  ) +
  theme_minimal() +
  theme(
    legend.position = "none",
    axis.text.x = element_markdown()
  )
```



Scatter plot of wealth score vs high-speed (100/20Mbps) broadband access

```
ggplot(merged_data %>% drop_na(speed_100_20, wealth_score), aes(x = speed_100_20, y = wealth_scor
  geom_point(alpha = 0.5, color = "#F28E2B", size = 2) +
  geom_smooth(method = "lm", se = TRUE, color = "gray40", linetype = "solid", alpha = 0.5) +
  labs(
    title = "Relationship Between High-Speed Broadband Access and Wealth",
    x = "High-Speed Broadband Access (100/20 Mbps)",
    y = "Wealth Score (Z-Score)"
  ) +
  theme_minimal()
```

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



Correlation between high-speed broadband access and different wealth metrics

```
#calculate correlations
cor_income <- cor(merged_data$speed_100_20, merged_data$median_income, use = "complete.obs")
cor_poverty <- cor(merged_data$speed_100_20, merged_data$poverty_rate, use = "complete.obs")

#print formatted output
asis_output(glue("
Correlation between high-speed broadband access and **median income:** *r* = {round(cor_income, 3
```

```
Correlation between high-speed broadband access and **poverty rate:** *r* = {round(cor_poverty, 3
"))
```

Correlation between high-speed broadband access and **median income:** *r* = 0.047

Correlation between high-speed broadband access and **poverty rate:** *r* = -0.041