Public Transportation Usage and COVID-19 Infection Rates

ChE 2410 Project 1 – Madison Butler

Motivation

A positive relationship might **encourage awareness of hygiene** in public shared spaces

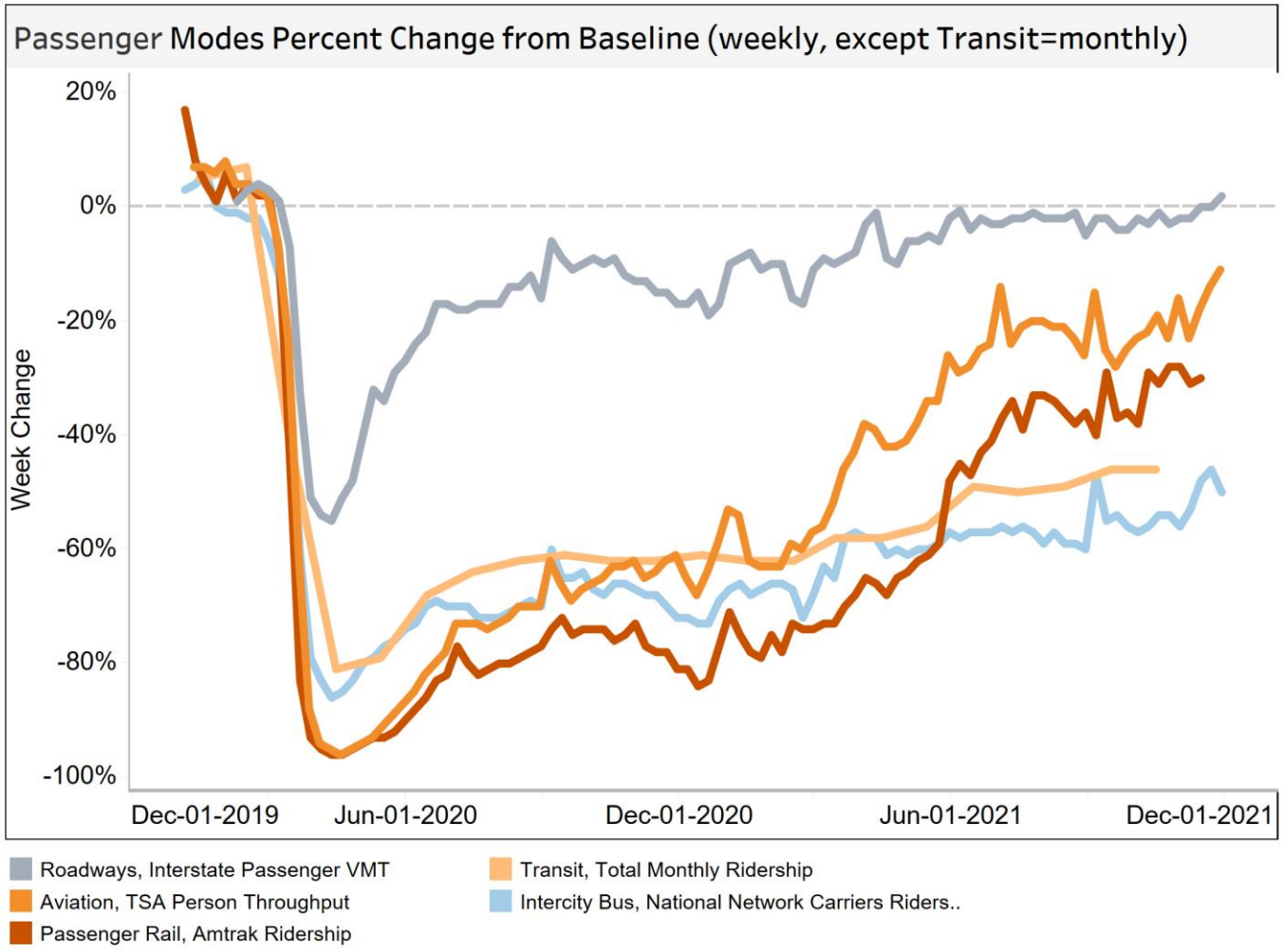
A negative relationship or a lack of relationship might **help reduce unnecessary avoidance** of public transportation

Previous relevant work

American public transportation workers are at higher risk of death by COVID compared to all industries (2021) [1]

Increased weekly bus and train usage correlated with increased infection rates[2]

[3]



Data sources

COVID cases, county level data from NYT 2021

Census.gov information on county level rates of:

- **Public transportation usage**
  - *“Percent of resident workers who commute by transit”*
- **Population density**
- **Population**

Assumptions

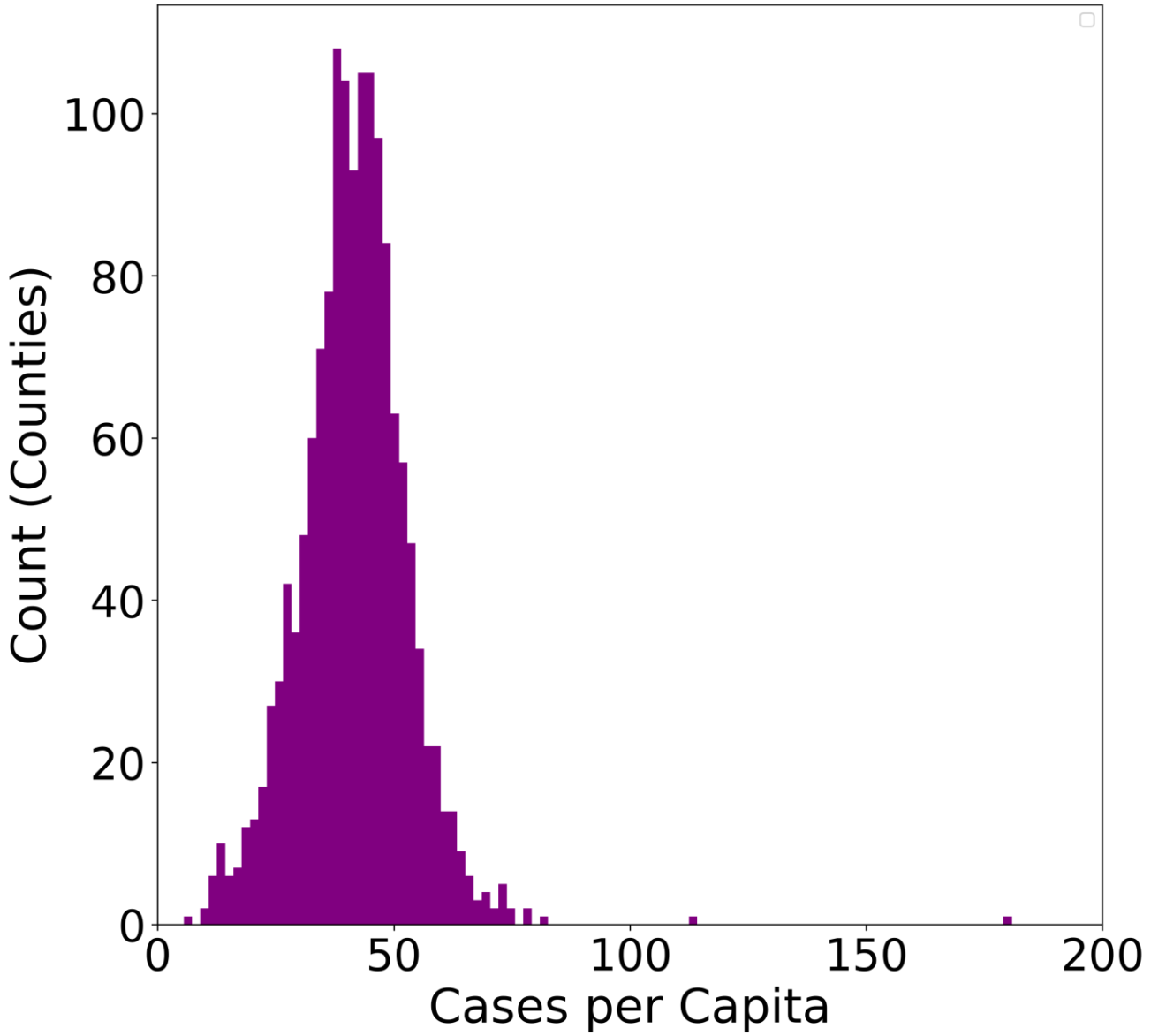
1. Non-working transit trips are uniform per capita across counties
2. COVID transit avoidance was uniform per capita across counties
3. Sum of cases across 2021, a continuous year of COVID, should show a holistic assessment of a county's ability to fight infection

Data overview

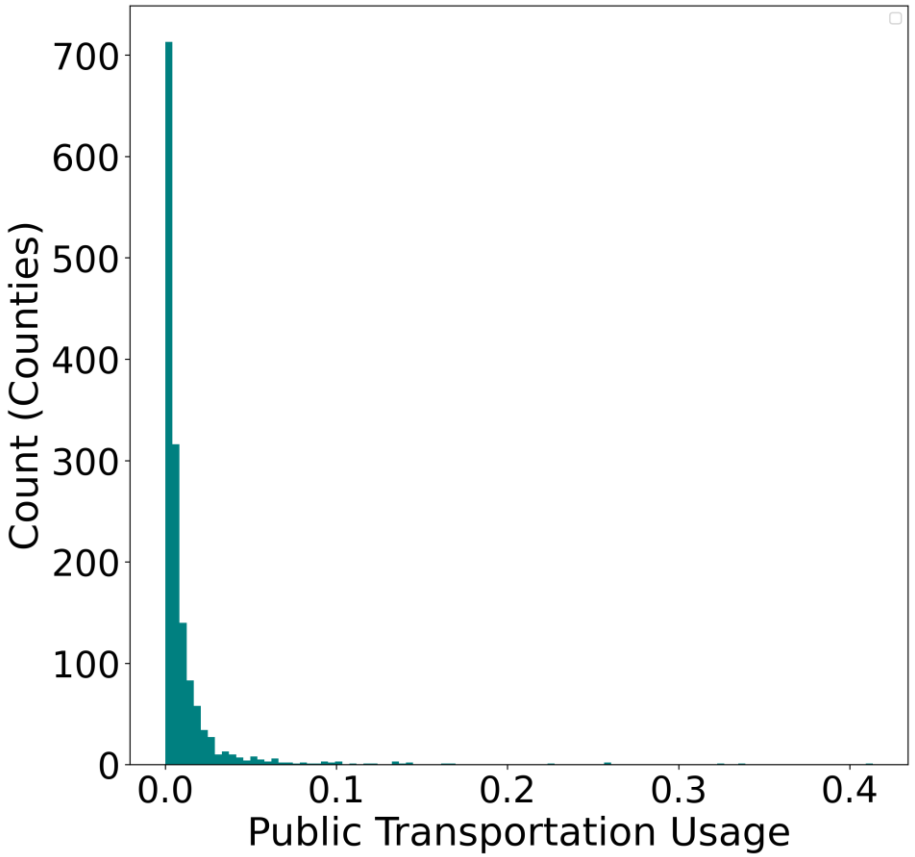
Cases per capita data appears normally distributed with distant outliers removed

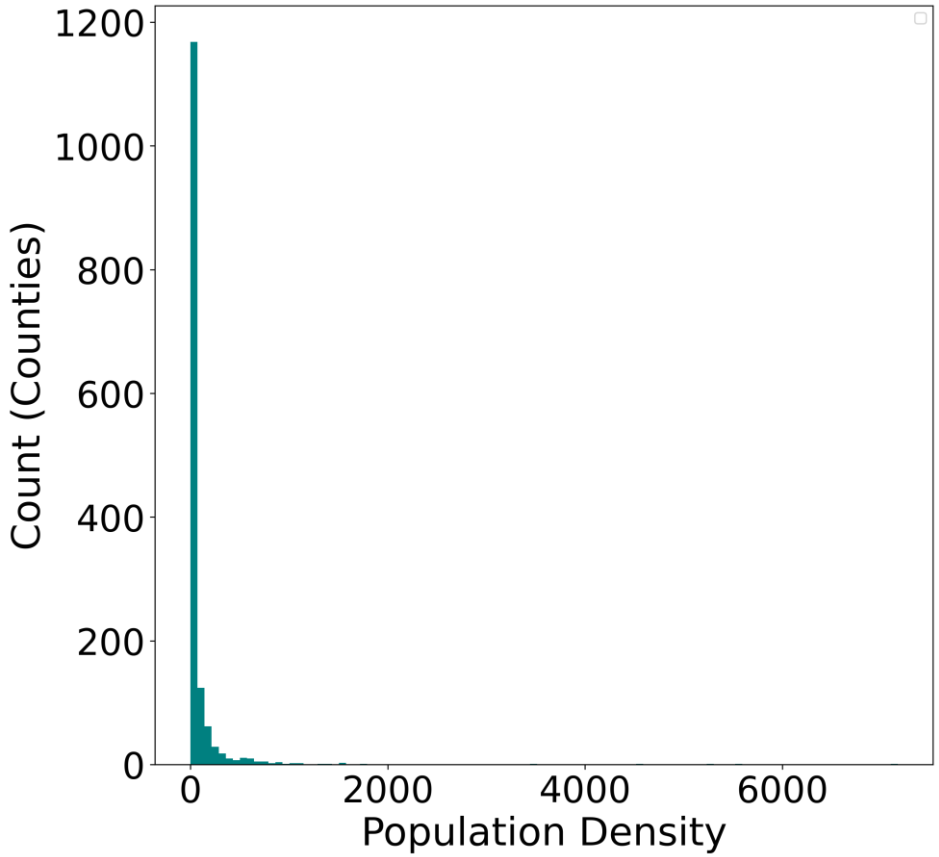
Asymmetrical outliers: there were high-case counties, but the opposite direction is impossible

$\mu = 41.5 \quad \sigma = 11.6 \quad CV = 0.279$

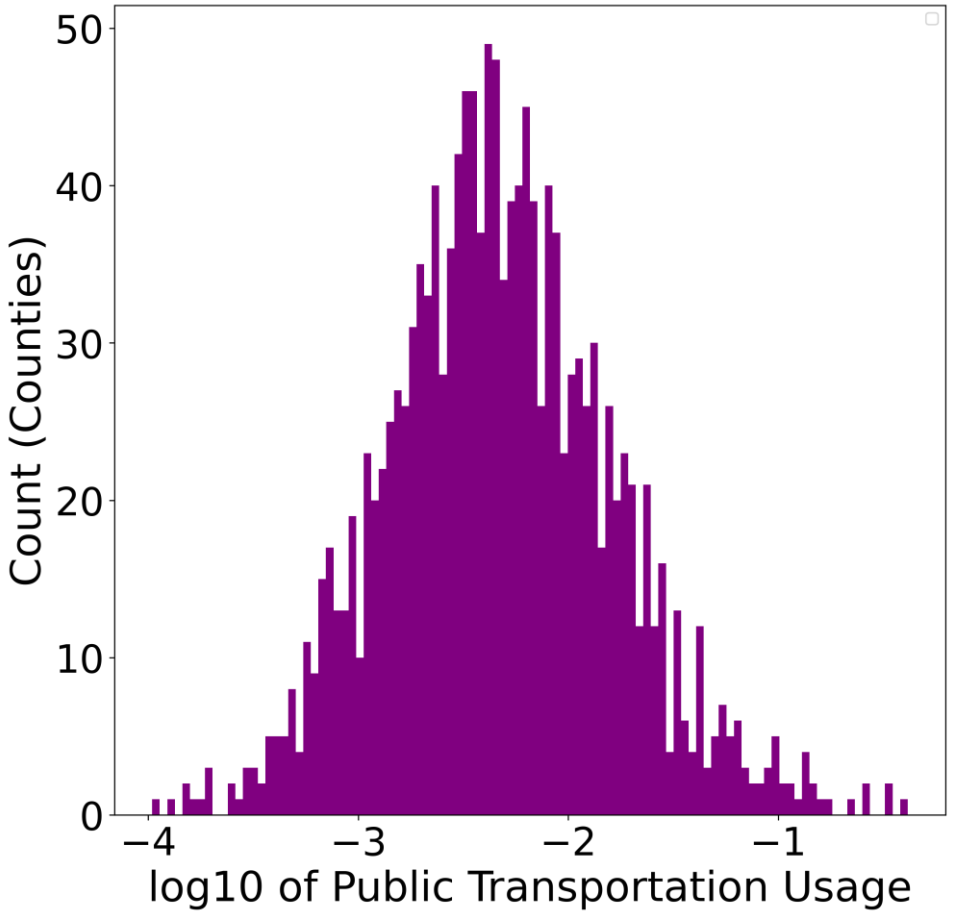


Data overview

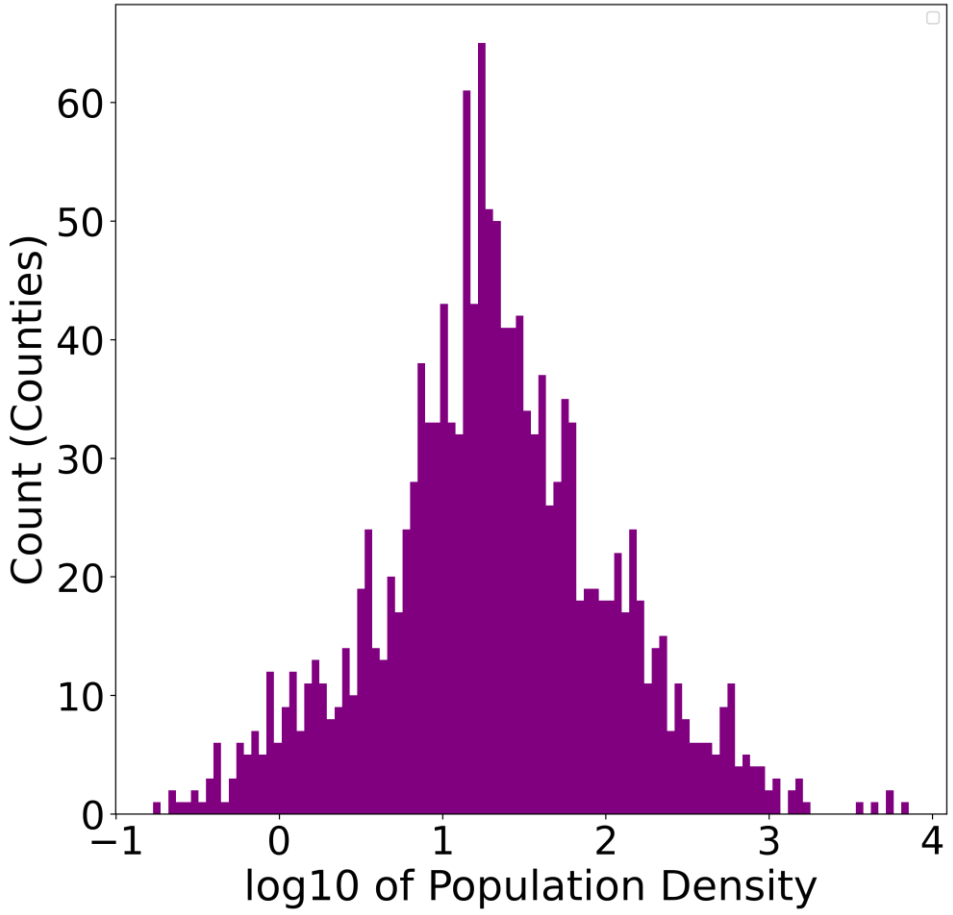




Data overview



μ = -2.33 σ = 0.536 CV = 0.230

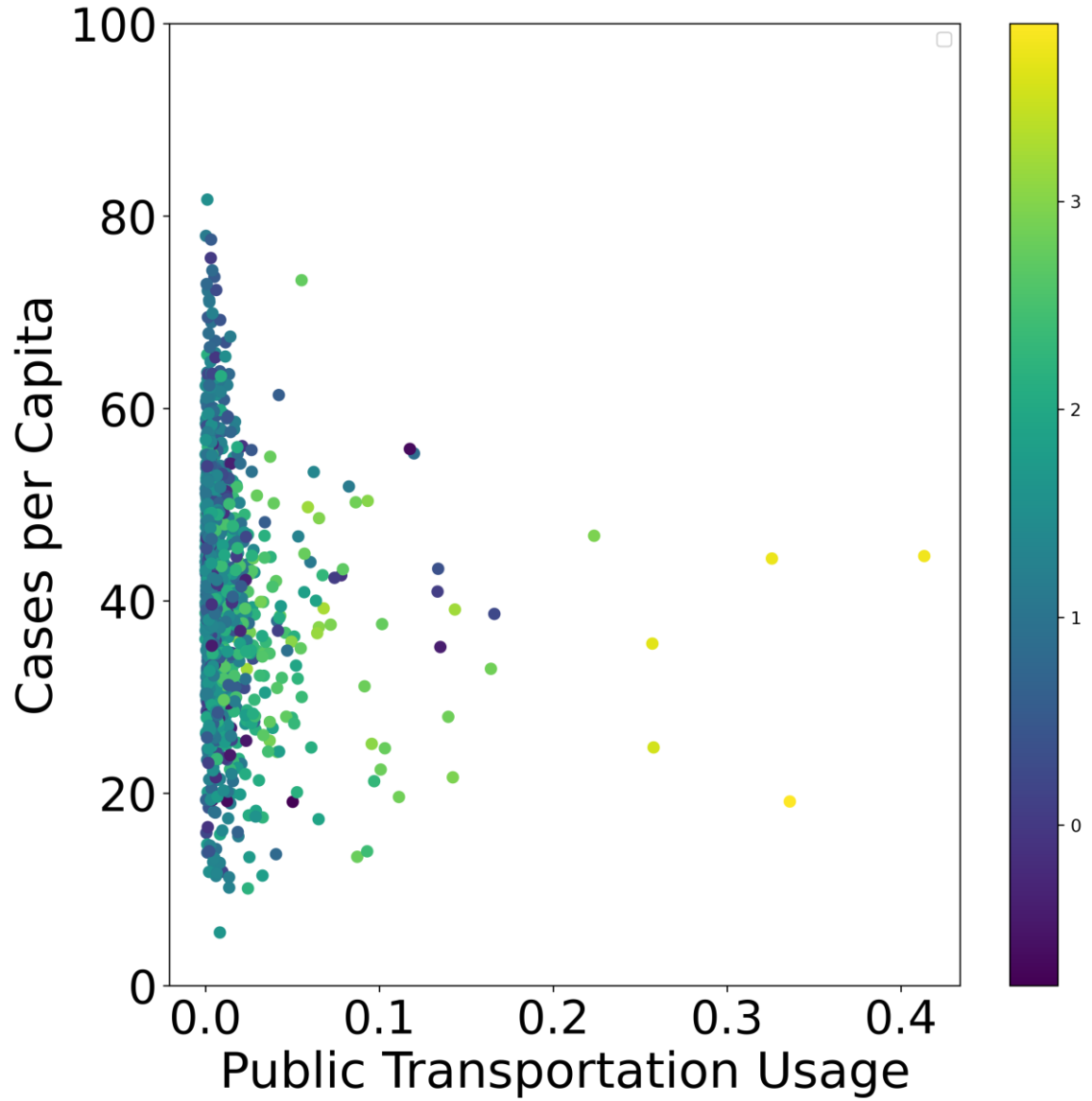


μ = 1.32 σ = 0.710 CV = 0.540

Data overview

Public transportation usage and cases per capita

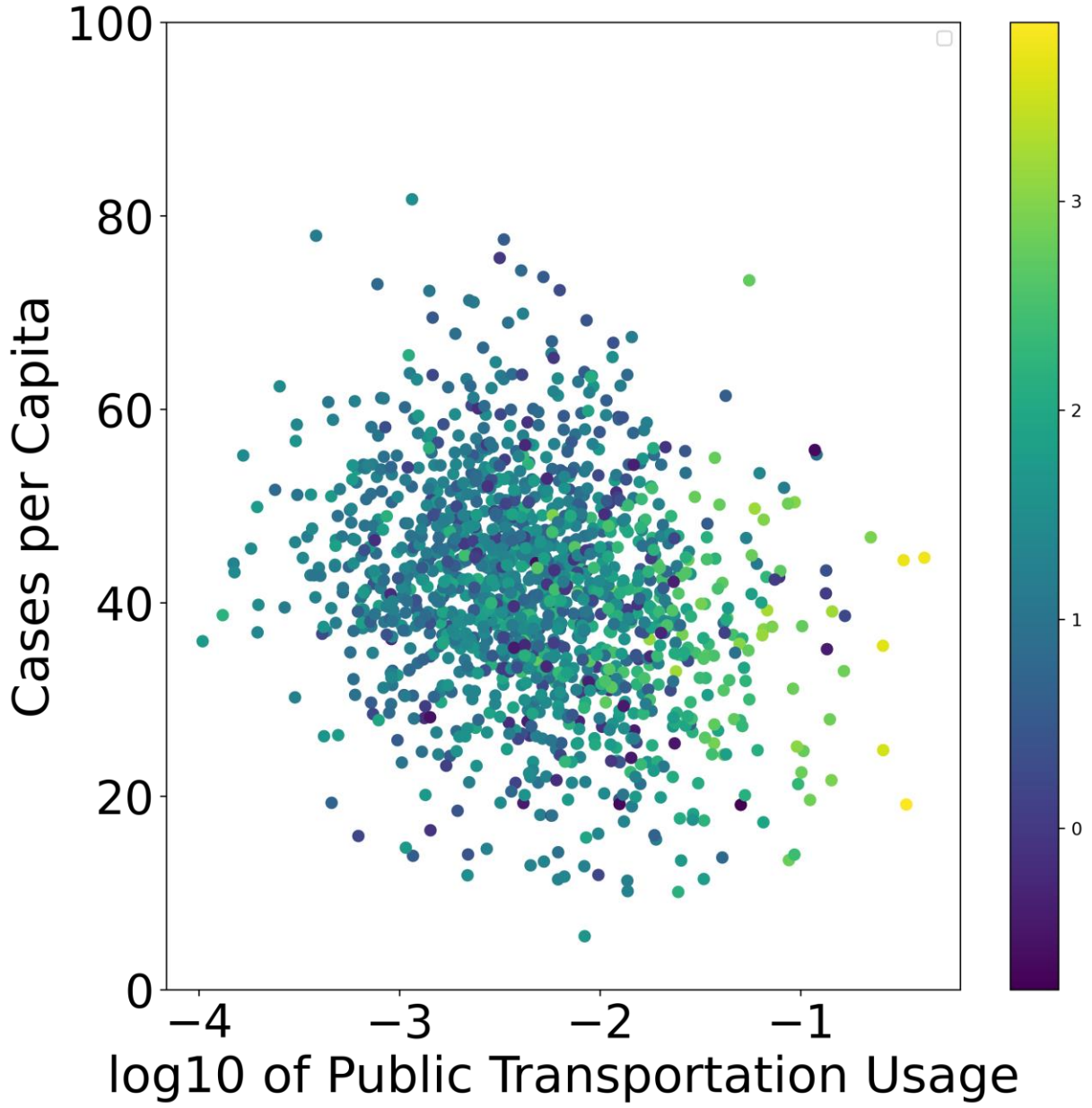
Population density mapped to color



Data overview

Log 10 of public transportation usage

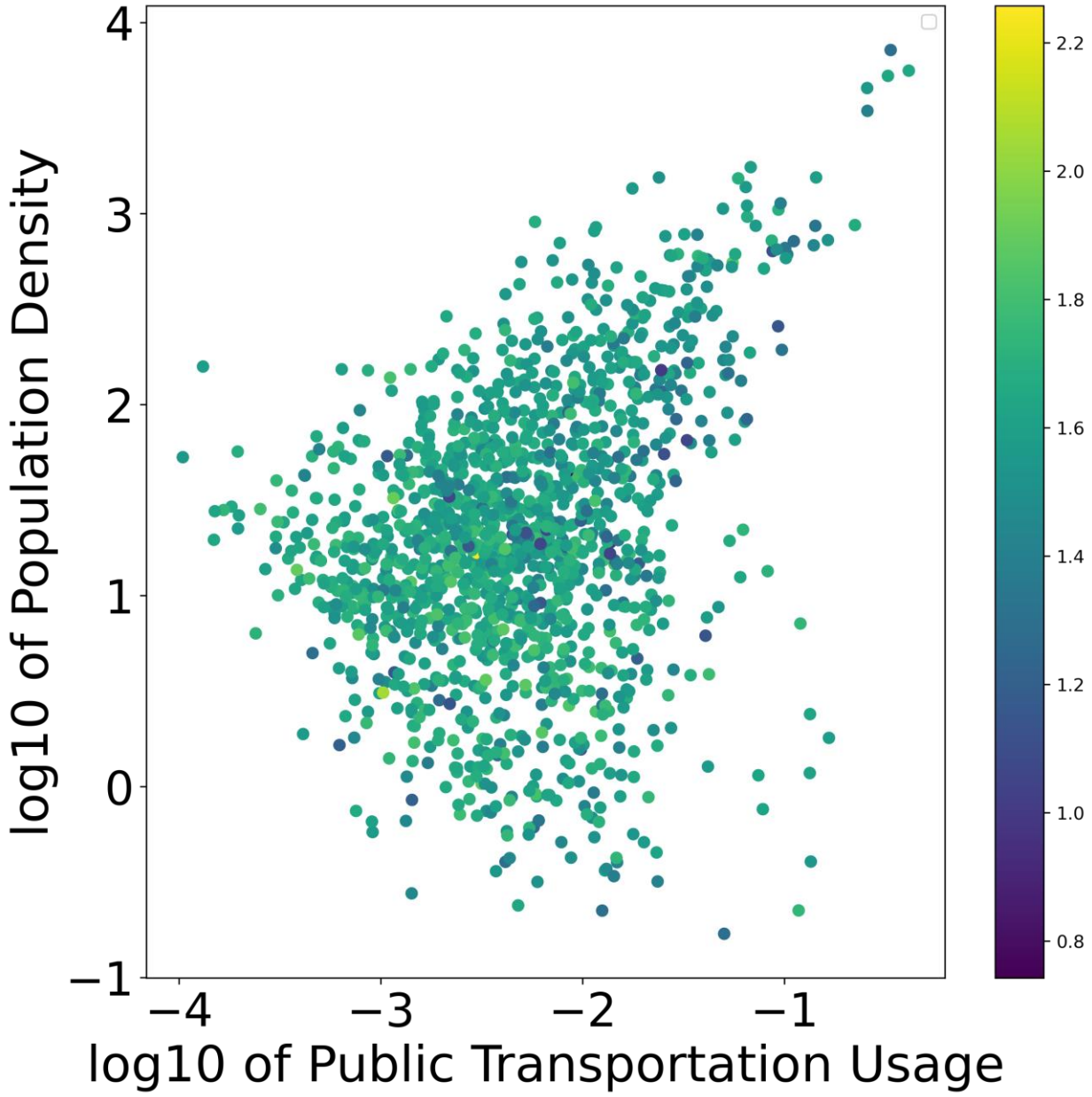
Population density mapped to color



Data overview

Public transportation usage and population density may share relationship

Per capita cases in color



Data analysis

$$y = a * x_1 + b * x_2 + c * x_1 * x_2 + d$$

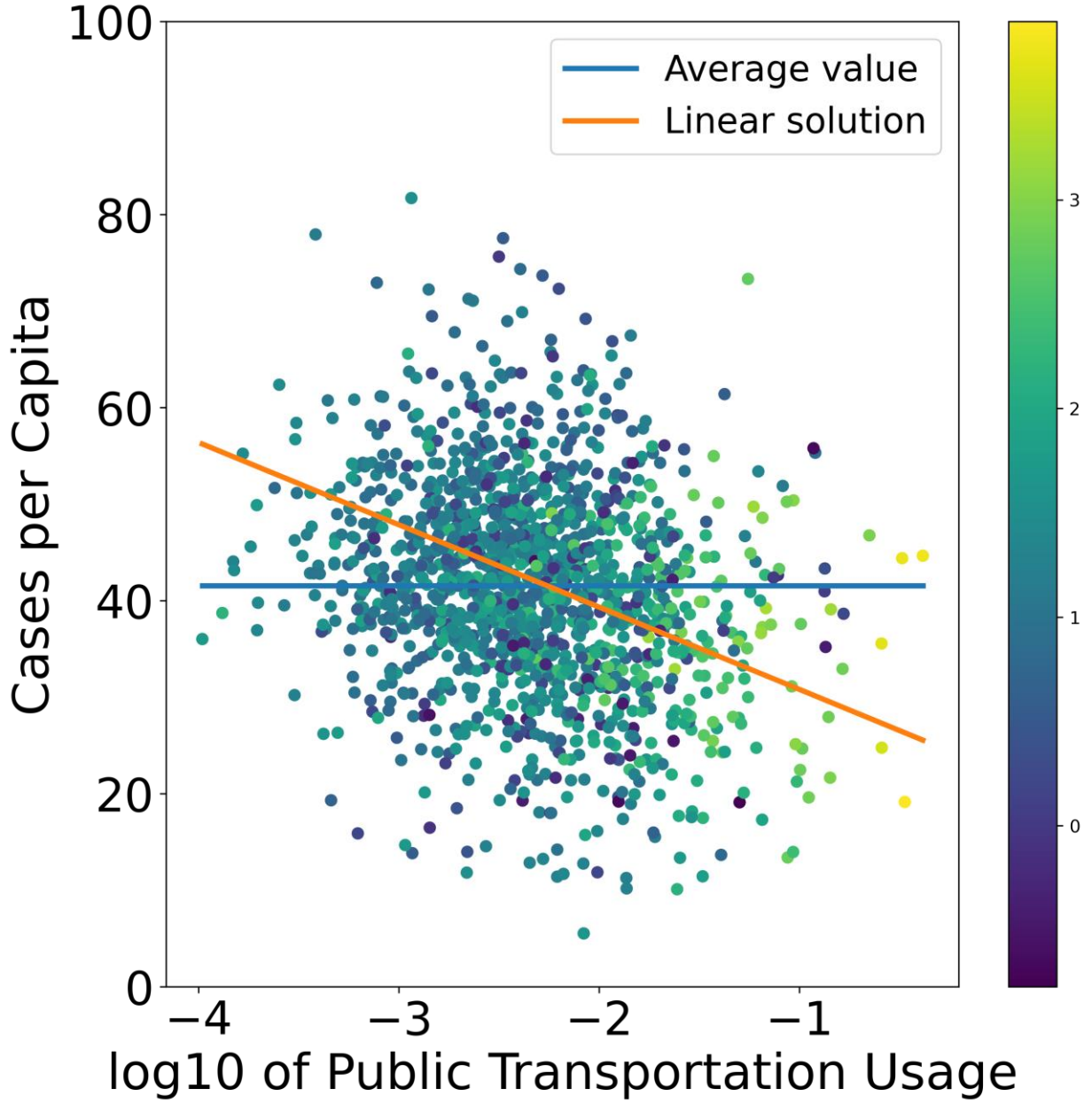
Where:

x_1 = log10(transportation usage)

x_2 = log10(population density)

y = cases

Model	SSE
Average	1.9812 E15
Linear	1.8798 E15



Data analysis

$$y = a * x_1 + b$$

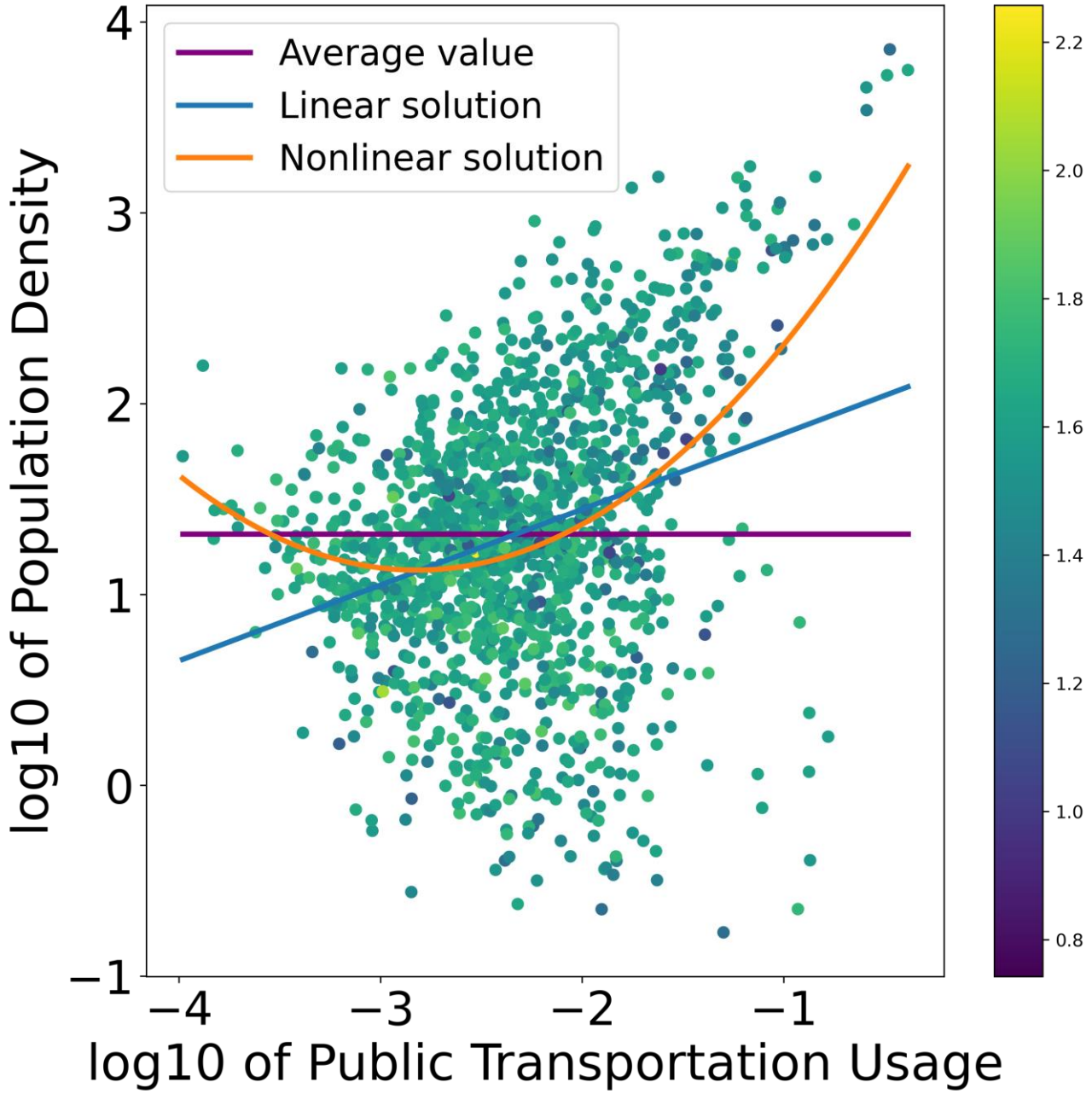
$$y = a * {x_1}^2 + b * x_1 + c$$

Where:

$x_1 = \log_{10}$(transportation usage)

y = cases

Model	SSE
Average	743
Linear	676
Nonlinear	641



Summary of findings

No significant relationship between public transportation usage and COVID cases per capita, including when controlling for population density and interactions

Relationship identified between population density and public transportation usage

Limitations

All county data may have obscured relationship by including too many counties without much public transportation usage

America does not have a lot of counties with robust public transportation and was bad at stopping infections

Next Steps

Pick case studies of metro areas with high public transportation – how did those counties do with cases over time

Try country level data, controlling for PPE use

Find better source for total cases

Maybe use relationship between population density and transit use to create a new model for prediction

Sources

[1] Heinzerling A, Vergara XP, Gebreegziabher E, et al. COVID-19 Outbreaks and Mortality Among Public Transportation Workers — California, January 2020–May 2022. MMWR Morb Mortal Wkly Rep 2022;71:1052–1056. DOI: http://dx.doi.org/10.15585/mmwr.mm7133a4.

[2] Michael M. Thomas, Neda Mohammadi, John E. Taylor,Investigating the association between mass transit adoption and COVID-19 infections in US metropolitan areas, Science of The Total Environment, Volume 811, 2022, 152284, ISSN 0048-9697, https://doi.org/10.1016/j.scitotenv.2021.152284.

[3] https://www.bts.gov/covid-19/week-in-transportation

Images from unsplashed