

Probabilistic answers to data science questions: using Bayesian hierarchical modeling in analysis

Eric Lind, Ph.D
Strategic Initiatives

Metro Transit at a Glance

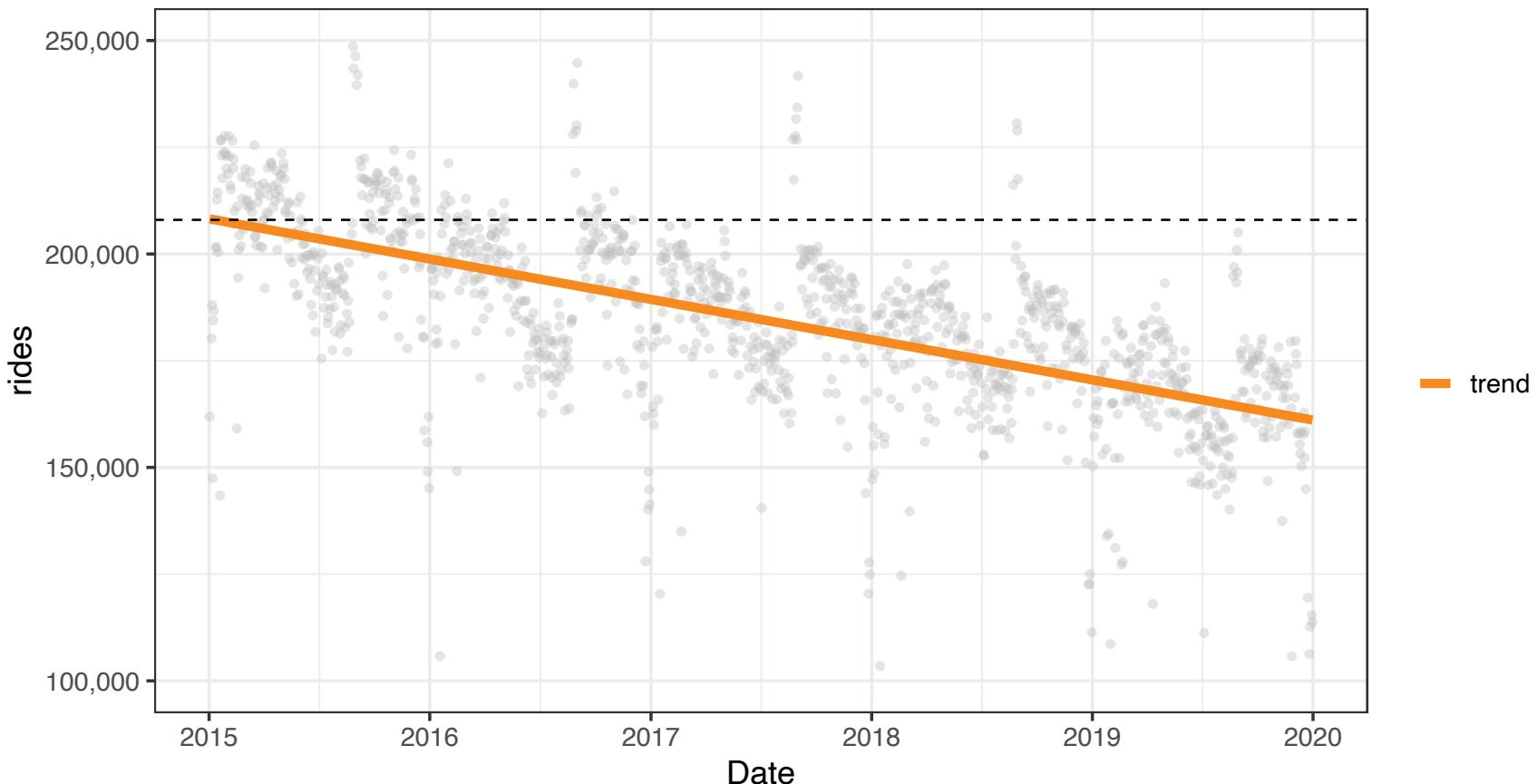
- 15th largest in U.S.
- 7 counties, 90 cities
- 122 bus routes
 - 2 arterial BRT
- 2 light rail lines
- 1 commuter rail line
- 900 buses, 86 LRVs
- 3,200+ employees
- 250,000 daily rides

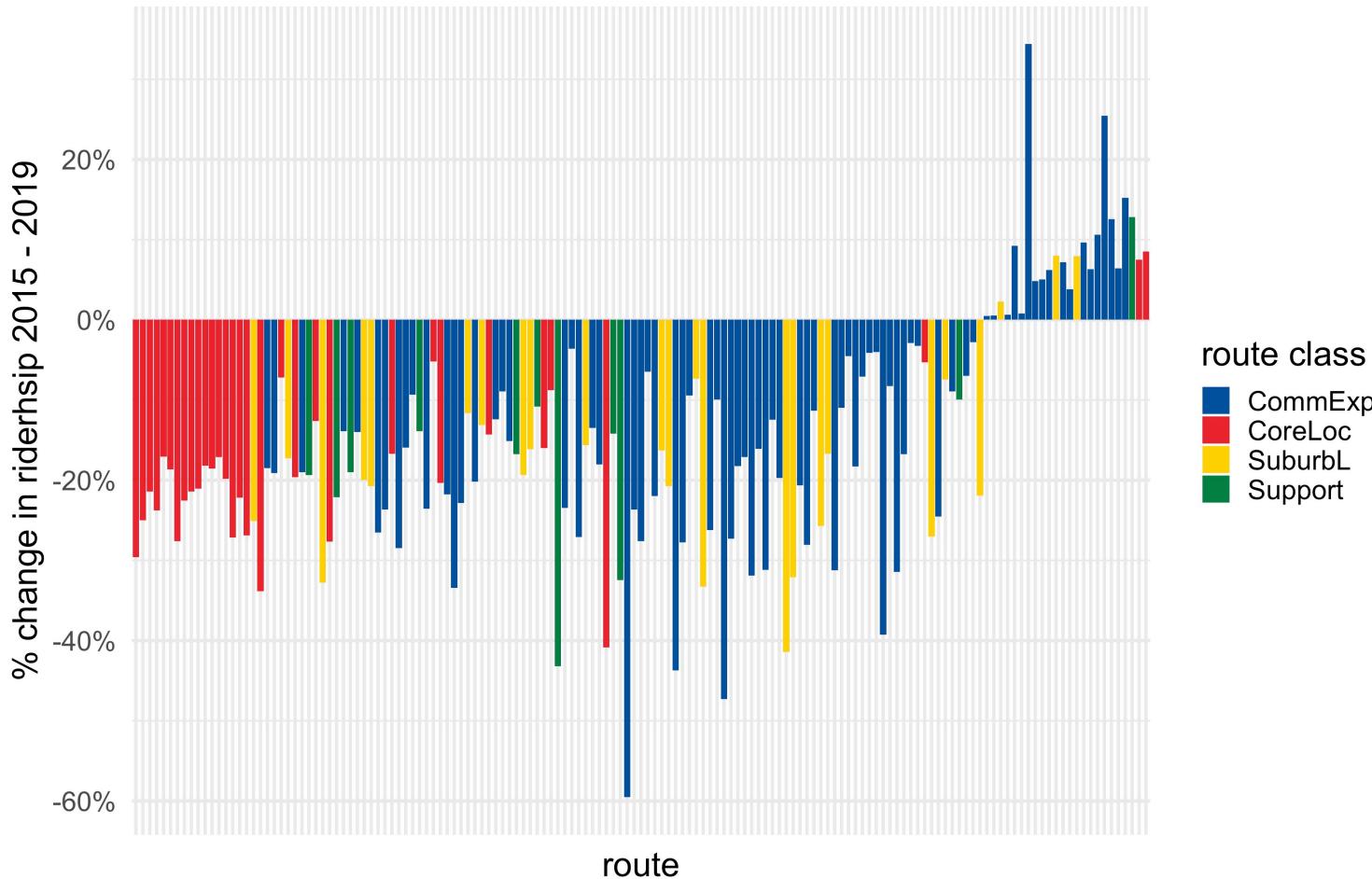


"We at Metro Transit deliver environmentally sustainable transportation choices that link people, jobs and community conveniently, consistently and safely."



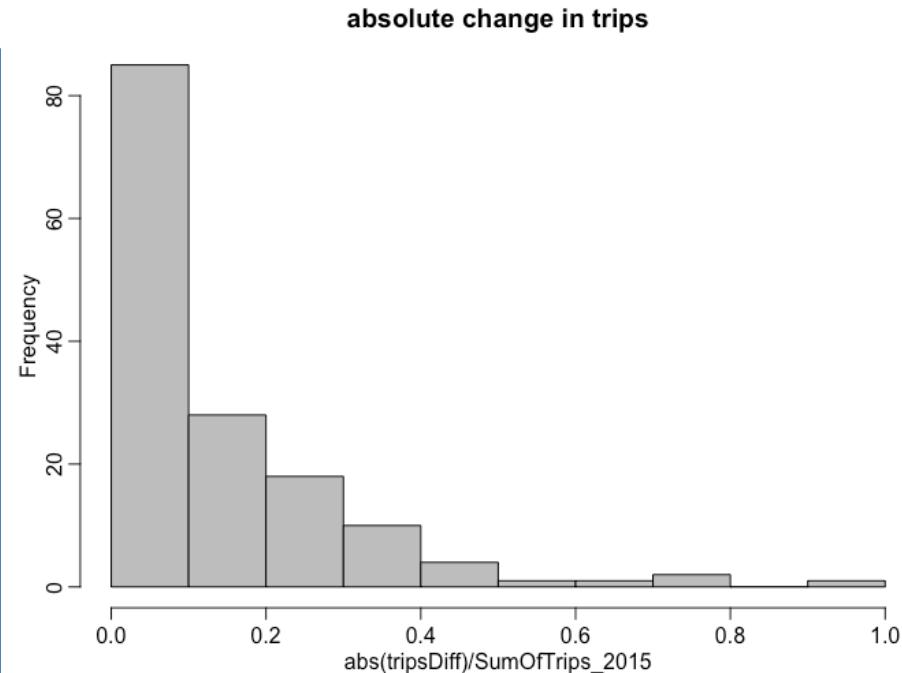
Weekday total Regular Bus rides by day





predicting change

- route class
 - Core Local
 - Suburban Local
 - Supporting Local
 - Commuter & Express
- change in service
 - trips or hours
 - proposed change cutoff 30%
 - 10 routes increased
 - 13 routes decreased
 - 127 routes not changed



goal: characterize ridership change by bus route

- capture:
 - network-wide ridership change
 - most routes share trend*hierarchical grouping*
 - differences by route class
 - classes capture type of route, ridership patterns*hierarchical grouping*
 - differences by change in amount of service
 - trips, in-service hours*predictor*



Rev. Thomas Bayes (1701-1761)

Bayes' Theorem:

The probability of variable y given x is determined by the joint probability of x and y , and the probability of x .

$$\Pr(\text{model} | \text{data}) \propto \frac{p_x(x | y) p_y(y)}{p_x(\text{data})}$$

quantifying change

$$y_i = \log(2019 \text{ rides} / 2015 \text{ rides})$$

translation: the relative rate of change of ridership for route i

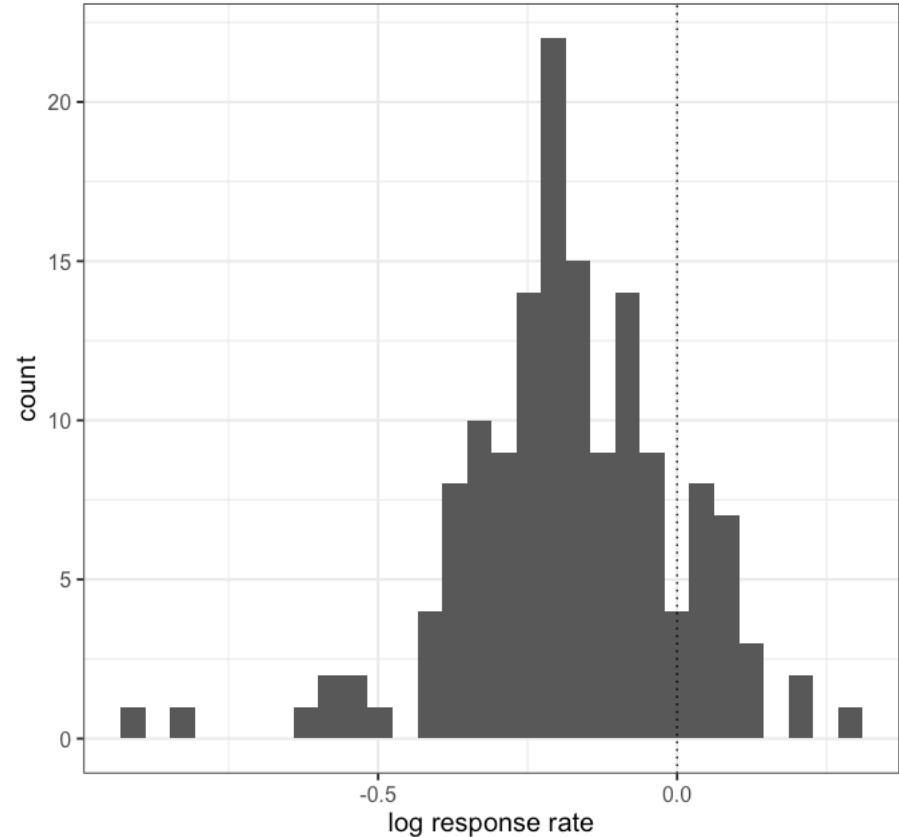
example: Route 21

4,117,347 rides in 2015

3,080,527 rides in 2019

$y = -0.290$

$e^y = 74.8\%$ of 2015 ridership



2015 to 2019 change by route

modeling change

- hierarchical model

$$y_i \sim N(\mu, \sigma^2)$$

$$\mu = \beta_{0_{class}} + \beta_{1_{class}} * \Delta hrs_i$$

$$\beta_{0_{class}} \sim N(\bar{\beta}_0, \sigma_{int}^2)$$

$$\beta_{1_{class}} \sim N(\bar{\beta}_1, \sigma_{hrs}^2)$$

route change is drawn from distribution of probable changes...

...which is determined by the expected route type change, and the change in service for that route...

...the coefficients of which can change by route type

modeling change

- hierarchical model

$$y_i \sim N(\mu, \sigma^2)$$

$$\mu = \beta_{0_{class}} + \beta_{1_{class}} * \Delta hrs_i$$

$$\beta_{0_{class}} \sim N(\bar{\beta}_0, \sigma_{int}^2)$$

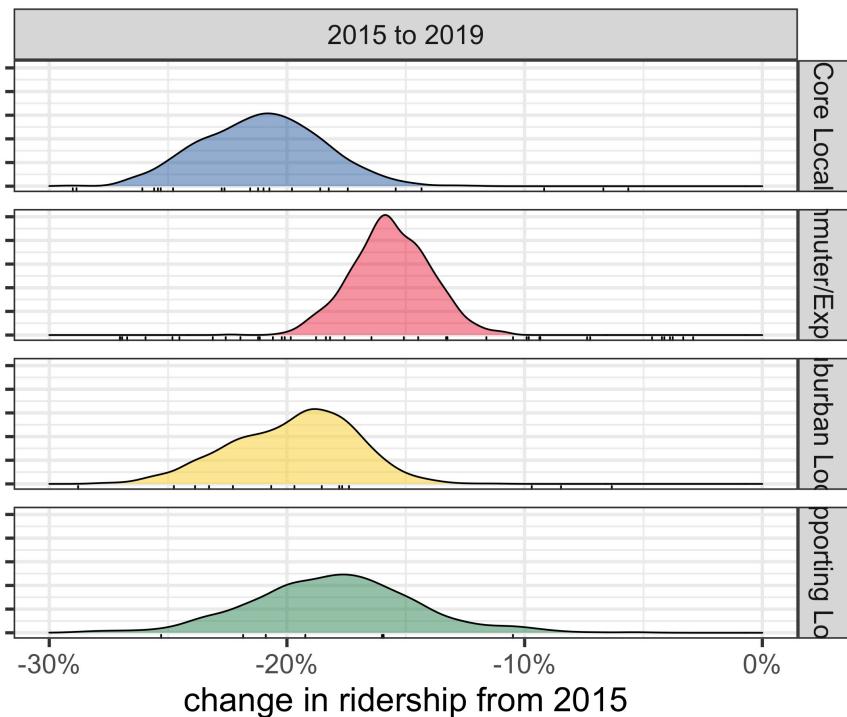
$$\beta_{1_{class}} \sim N(\bar{\beta}_1, \sigma_{hrs}^2)$$

```
m4 <- map2stan(
  alist(
    y ~ dnorm(muhat, sigma),
    muhat <- a_C[cls] + b_H[cls]*hrs_chg,
    a_C[cls] ~ dnorm(A, sigma_class),
    b_H[cls] ~ dnorm(B_H, sigma_hours),
    c(A, B_H) ~ dnorm(0, 2),
    c(sigma, sigma_class, sigma_hours) ~ dcauchy(0, 1)
  ),
  data = tomod3, chains = 4, cores = 4, debug = T,
  control = list(max_treedepth = 15, adapt_delta = 0.98)
)
```

results

expected change in ridership 2015-2019

in-service hours held constant

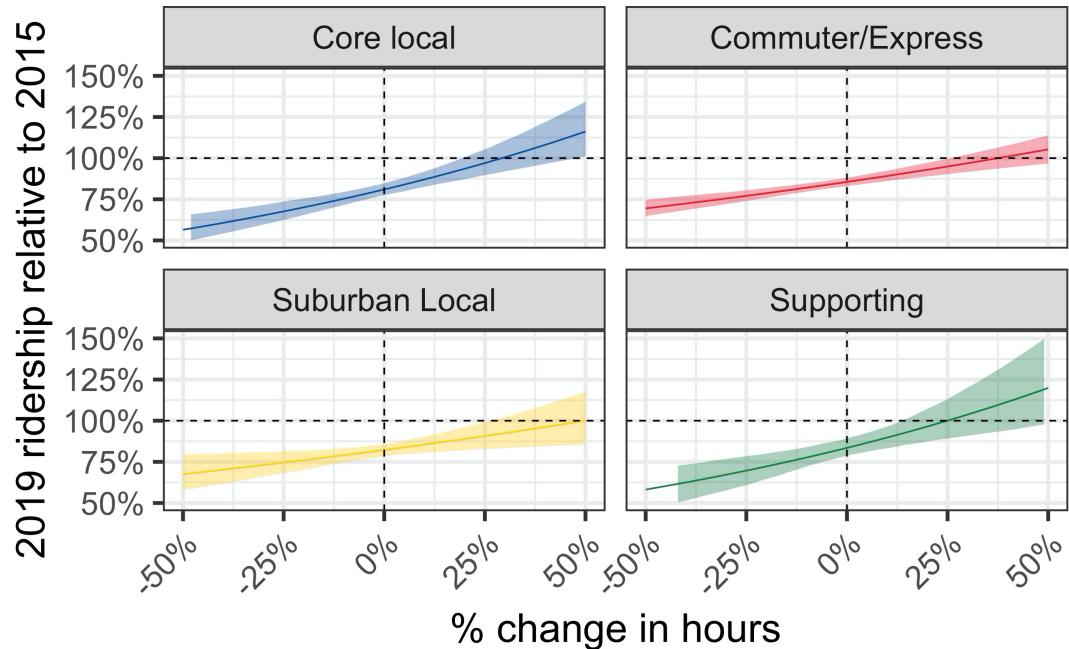


- Overall decline: 18% [12 - 24]
- Core local -23%
- Commuter -17%
- Suburban -22%
- Supporting -19%

results

- hours change varies by class
 - core local steeper
 - commuter flatter
- *vertical dashed line:*
expected ridership @ no
service change
- *horizontal dashed line:*
service change needed to
maintain 2015 ridership level

expected change in ridership 2015-2019
response to change in in-service hours



use in forecasting

scenario for Route 23: 7 additional weekday service hours

change
in annual
from
base

