Tools for Analyzing the Built Environment: Spatial Temporal Aggregated Predictors

Adam Peterson Brisa Sanchez

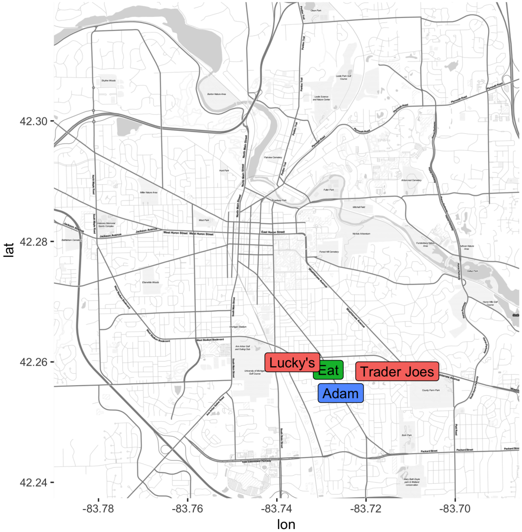
Ann Arbor R User's Group

April 11, 2019

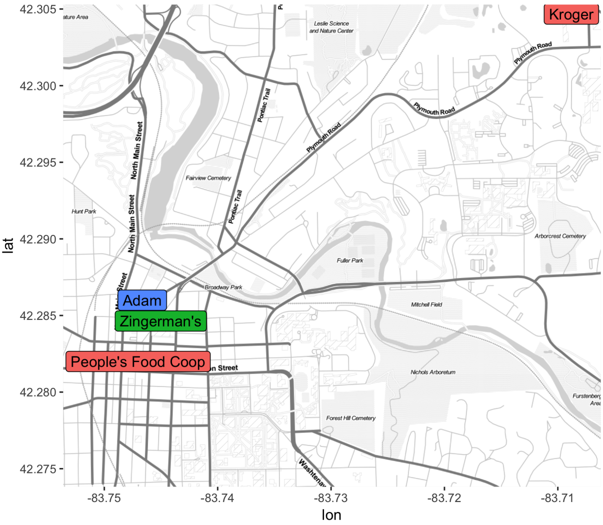
Once upon a time...

A young biostatistics graduate student moved to Ann Arbor, MI

He lived in the south-east corner of AA

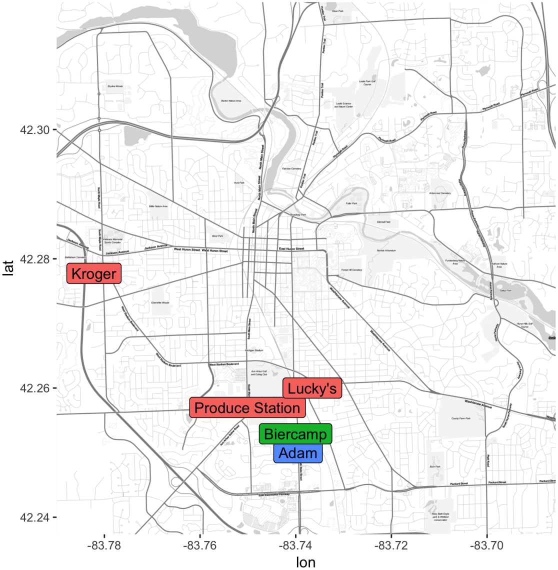


And then he moved...



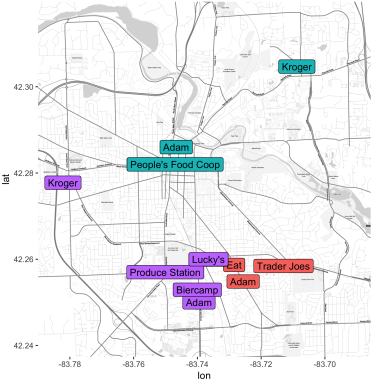
he stayed there a while...

but then he eventually moved again...



Scientific questions:

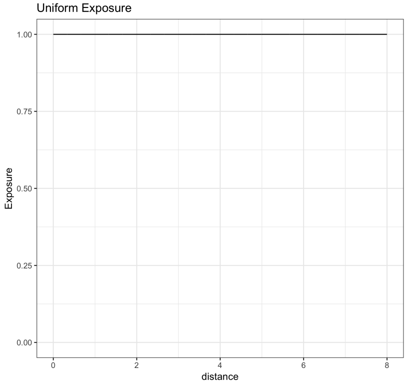
- Do these differing food environments effect Adam's health at all?
- If so,
  - What role does distance play in this effect?
  - How about time?

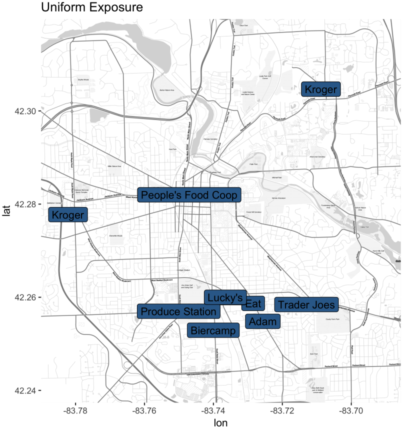


How would distance mediate the effect of a single store?

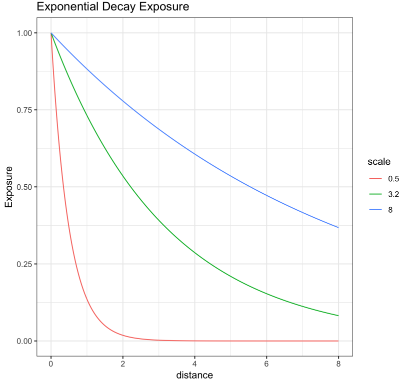
- What would the null look like?
- What might a more likely scenario look like?

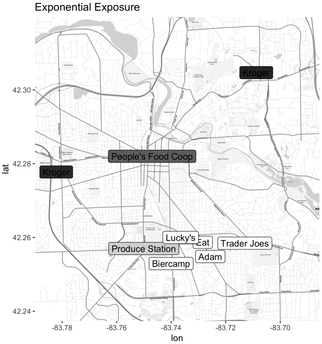
The Null


Uniform Exposure
1.00
0.75
0.50
0.25
0.00
Exposure
0
2
4
6
8
distance


Uniform Exposure
Kroger
People's Food Coop
Kroger
Lucky's
Eat
Trader Joes
Produce Station
Adam
Biercamp
42.30
42.28
42.26
42.24
lat
-83.78
-83.76
-83.74
-83.72
-83.70
lon

Alternative





The Model: Built Environment Point Pattern Exposure

Spatial Aggregated Covariate

$$X_i := \sum_{d \in \mathcal{D}} \mathcal{K}_s\left(\frac{d}{\theta^s}\right)$$

- d: a *known* distance between subject i and a built environment feature
- $\mathcal{D}$: set of distances between subject i and all built environment features of a certain class.
- θ: *unknown* scale describing the distance - exposure relationship
- $\mathcal{K}_s$: "known" spatial weighting function - e.g. e^{-x}

Explicitly

Business	Class	Distance (mi)
Eat	Restaurant	0.20
Trader Joes	Grocery Store	0.88
Lucky's	Grocery Store	0.75

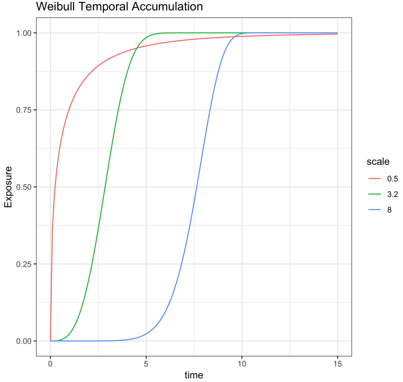
$$X_{Adam,R} = e^{-\frac{.2}{\theta_R^s}} = .81 \qquad \text{(theta = 1)}$$

$$X_{Adam,G} = e^{-\frac{.88}{\theta_G^s}} + e^{-\frac{.75}{\theta_G^s}} = 1.33 \qquad \text{(theta = 2)}$$

$$E[BMI_{Adam,Year1}] = \alpha + \text{male}\delta_{male} + \text{Age}\delta_{Age} + ... + X_{Adam,G}\beta_G + X_{Adam,R}\beta_R$$

What about time?

- Would it be that much harder to look at the temporal relationship?



Temporal Aggregated Covariate

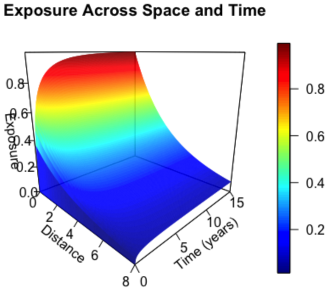
$$X_i := \sum_{t \in \mathcal{T}} \mathcal{K}_t(\frac{t}{\theta^t})$$

- t: The *known* time subject i has been exposed to a built environment feature at the time of measurement
- $\mathcal{T}$: set of times between subject i and all built environment features of a certain class.
- θ: *unknown* scale describing the time - exposure relationship
- $\mathcal{K}_t$: "known" temporal weighting function - e.g. $1 - e^{-x}$

Put them together and what have you got?

Spatial-Temporal Aggregated Covariate

$$X_i = \sum_{(d,t)\in\mathcal{S}} \mathcal{K}_s(\frac{d}{\theta^s})\mathcal{K}_t(\frac{t}{\theta^t})$$



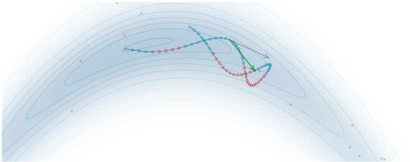
Some technical details

Final Model: Spatial Temporal Aggregated Predictors

$$g(E[Y_{i,j}]) = \boldsymbol{Z}_{i,j}^T\boldsymbol{\delta} + X_{i,j,k}(\theta_k^s, \theta_k^t)\beta_k + X_{i,j,l}(\theta_l^s)\beta_l + X_{i,j,p}(\theta_p^t)\beta_p + \boldsymbol{W}_{i,j}^T\boldsymbol{b}_i$$

- size of $\mathcal{S}, \mathcal{D}, \mathcal{T}$
- Space-Time Assumption: $\mathcal{K}_s(\frac{d}{\theta^s})\mathcal{K}_t(\frac{t}{\theta^t})$
- No features? $X_{i,j} = 0 + \sum_{(d,t)\in\mathcal{S}_j} \mathcal{K}_s(\frac{d}{\theta^s})\mathcal{K}_t(\frac{t}{\theta^t})$
- Some other assumptions
  - direct vs. indirect exposure
  - point process vs. continuous spaces

Estimation



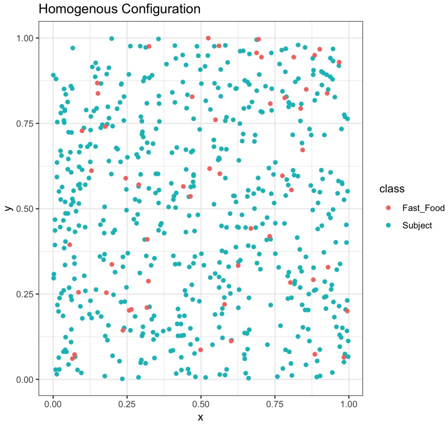
- Motivation
  - STKAP - MH/Gibbs
- Technical Details
  - HMC-NUTS

Applications

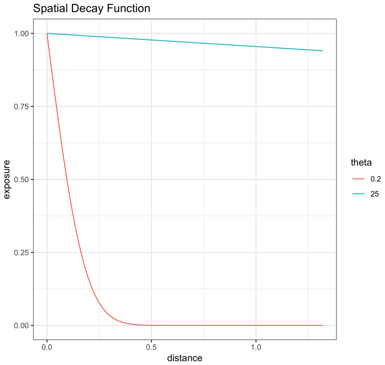
- Simulation - inclusion distance
- Real data
- rstap

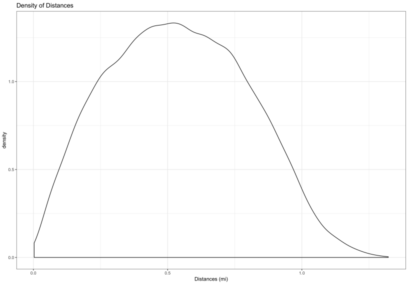
Simulation: Set-Up

$$E[\text{“BMI”}] = \alpha + \text{sex}\delta + FF(\theta_s)\beta$$



Simulation: Descriptives

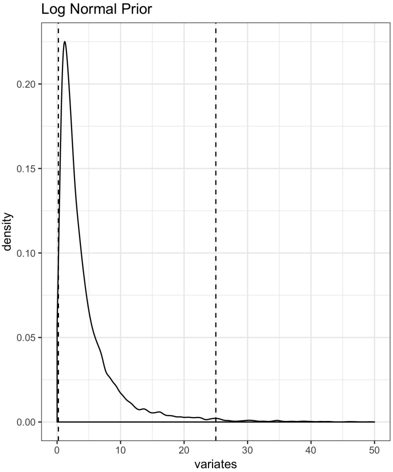


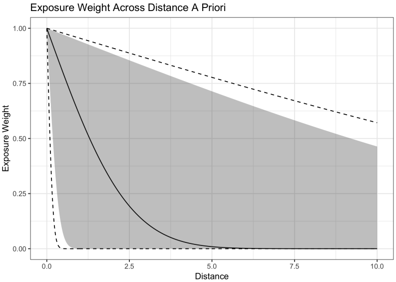


Model Fitting

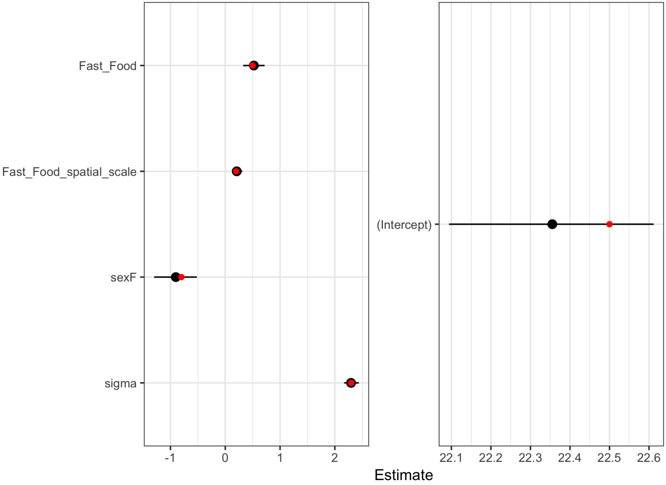
```
fit <- stap_glm(BMI ~ sex + sap(Fast_Food),
                 subject_data = subject_data,
                 distance_data = distance_data,
                 family = gaussian(link = "identity"),
                 id_key = "subj_id",
                 prior = normal(location = 0, scale = 5),
                 prior_intercept = normal(26,5),
                 prior_stap = normal(0, 3),
                 prior_theta = log_normal(1, 1),
                 prior_aux = cauchy(location = 0,scale = 5)
                 max_distance = max(dists), chains = 4,
                 iter = 2E3, cores = 4, warmup = 1E3)
```

Model fitting, priors

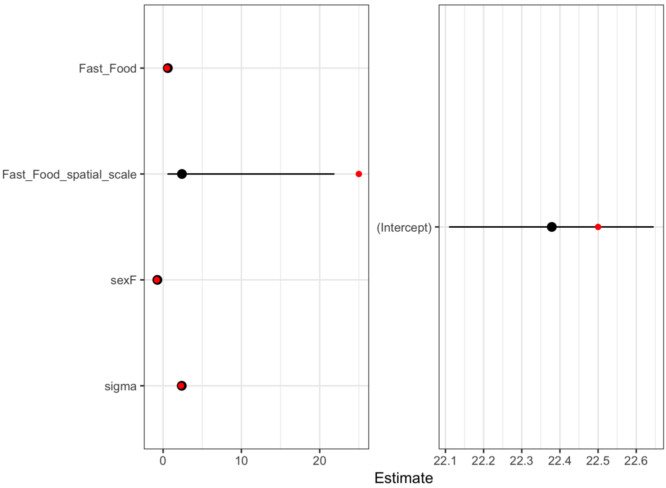




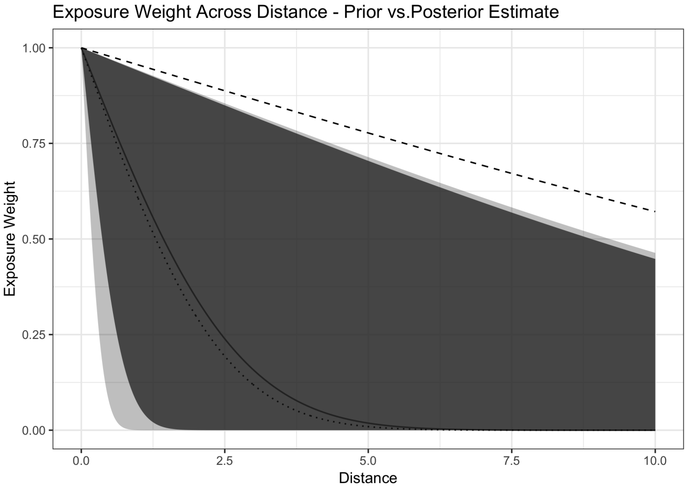
Results - "sharp decay"



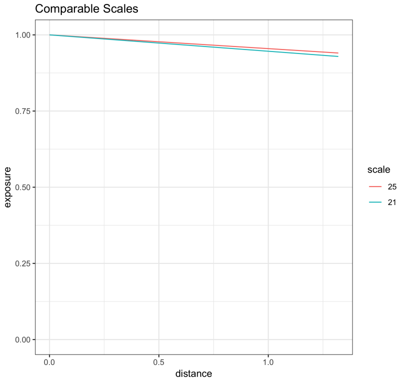
Results - “slow decay”

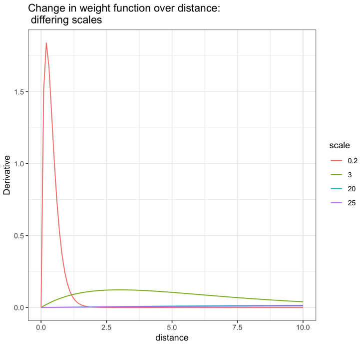


Results - “slow decay”



Estimating the Main Effect - Explanation



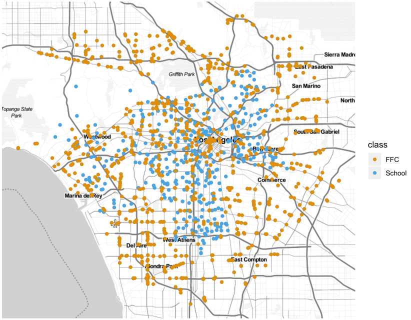


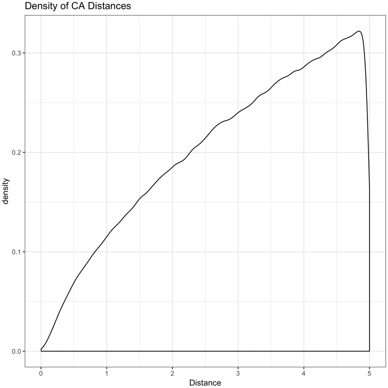
Real Data: Description

- CA Dept. of Ed. Fitnessgram project
- Academic Year 2010
- Built Environment Data: NETS database

Strata	Urban
Total Schools	2386
Mean # of Obese Students	33.1
Mean # of Total Students	86.6

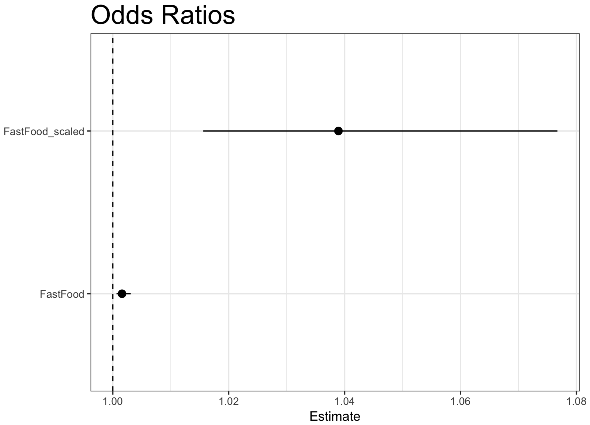
Real Data: Descriptives



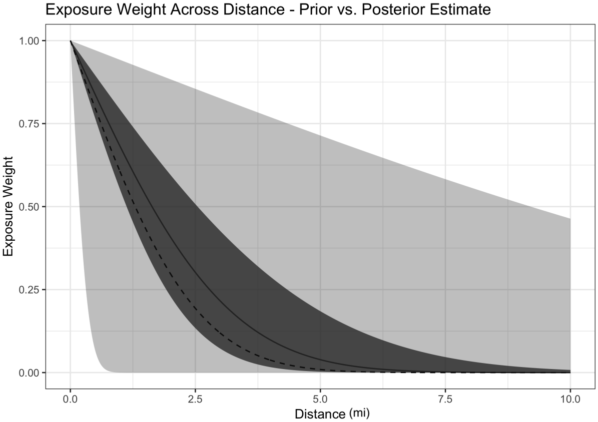


Real Data: Model and Estimates

$$\text{logit}(\mu_i) = \alpha + \boldsymbol{Z_i}^T \delta + X_{i,FF}^T(\theta_s)\beta$$



Real Data: Spatial Scales



Distance at which Exposure Effect is Negligible (mi)

	2.5%	50%	97.5%
Urban	4.27	6.28	9.62

Future Directions

- Causal Inference - Difference in Differences
- Faster sampling - custom C++
- Non-monotonic Spatial scales

References

- https://biostatistics4socialimpact.github.io/rstap/
- Neal, Radford M. "MCMC using Hamiltonian dynamics." Handbook of Markov Chain Monte Carlo 2.11 (2011): 2.
- Betancourt, Michael. "A conceptual introduction to Hamiltonian Monte Carlo." arXiv preprint arXiv:1701.02434 (2017).
- Heaton, Matthew J., and Alan E. Gelfand. "Spatial regression using kernel averaged predictors." Journal of agricultural, biological, and environmental statistics 16.2 (2011): 233-252

Questions?