

chhiring_analysis

Chhiring Lama

2026-02-17

Contents

The following states' data are missing: AK, HI, MA, WY. Alaska and Hawaii are not included since the dataset is specifically focusing on contiguous US states.

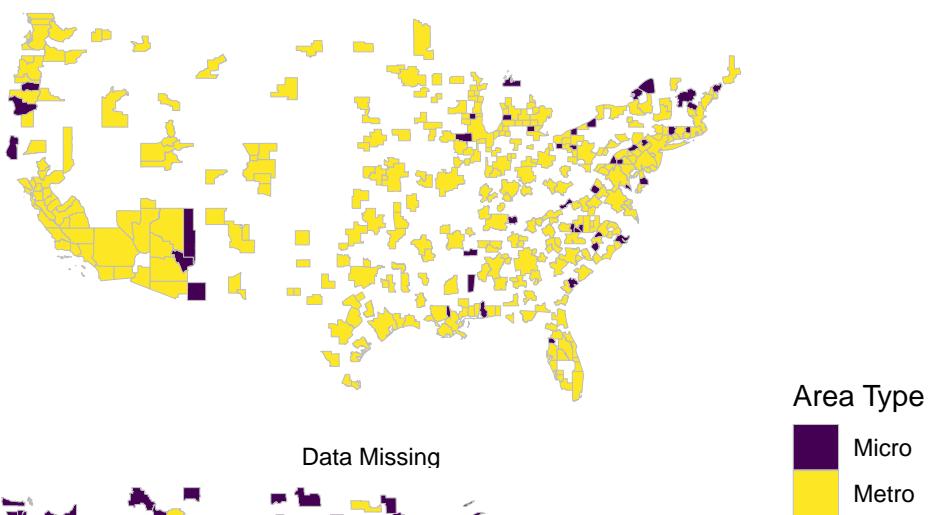
What about MA and WY?

In the EDA, we see that neighboring states have similar % white residents. Similarly, northern states generally have higher % of white residents.

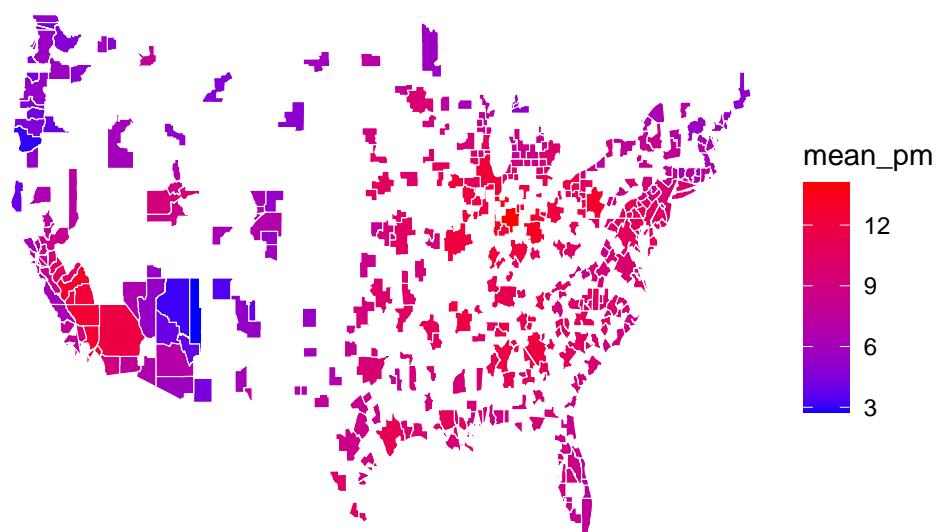
There are only 380 CBSA included out of ~900 out of them in contiguous US. This only 337 Metropolitan areas (missing 26) (population: 50k+) and 43 micropolitan areas. **Why?** There are 4 census tracts without CBSA allocation. One Census tract in CA missing PM2.5 and BC information.

US Core Based Statistical Areas (CBSAs)

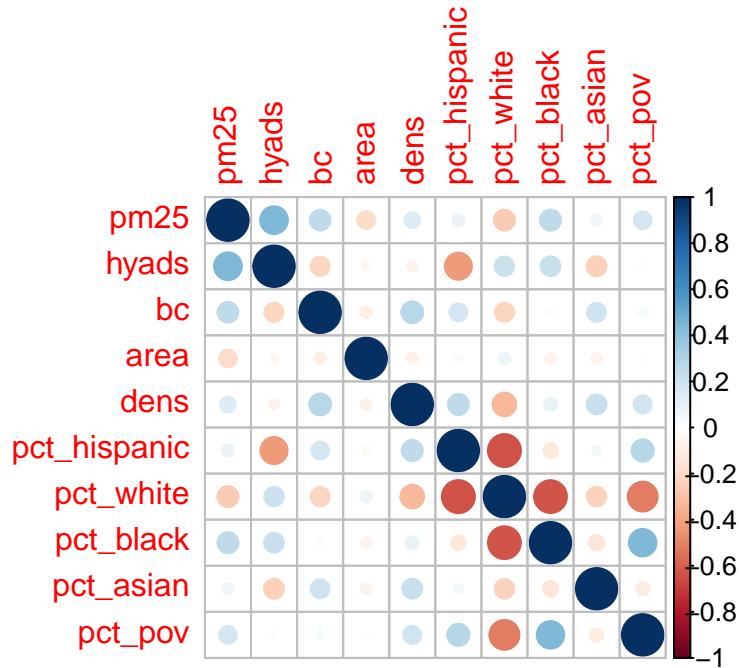
Data Present



Data Missing



Caveat: there are census tracts where total percentage of different groups don't add up. The data potentially is missing other racial/ethnic groups.



Percentage of white and percentage of population under poverty line are correlated. We need to account for this when modeling.

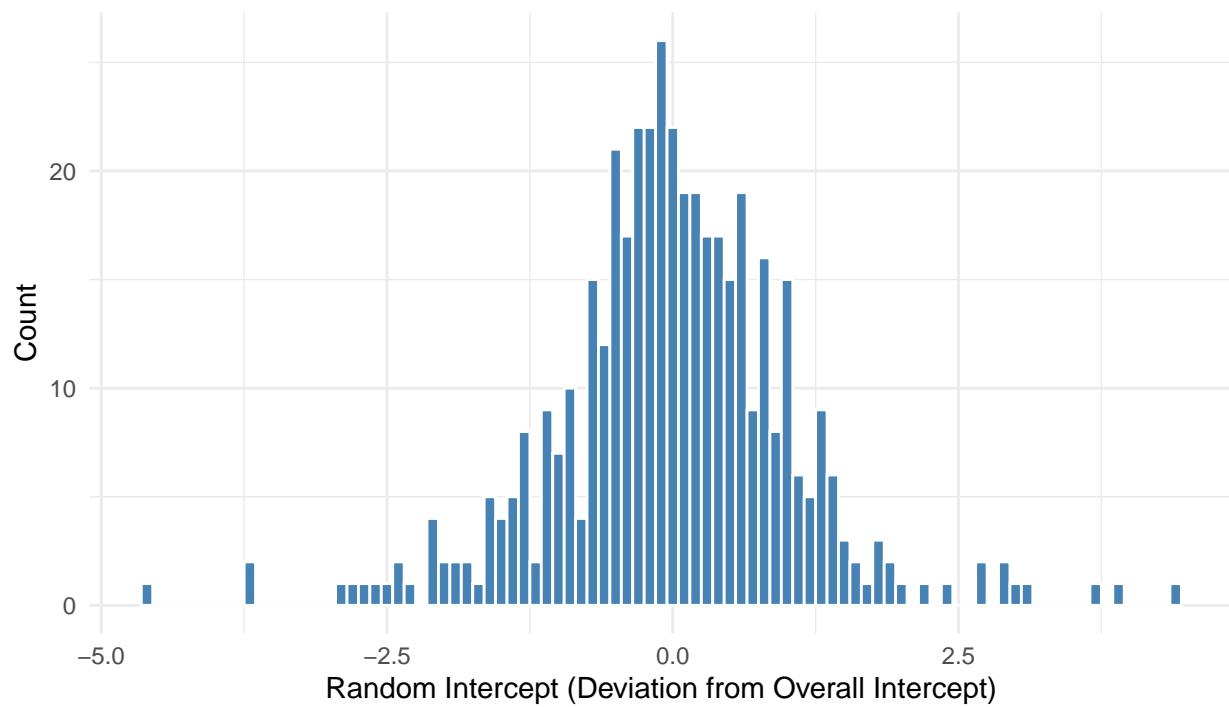
Running LMM with hierarchical structure

Characteristic	Beta	95% CI	p-value
(Intercept)	2.5	1.7, 3.2	<0.001
pct_white	-0.01	-0.01, -0.01	<0.001
pct_pov	0.00	0.00, 0.00	<0.001
ruca_agg			
rural	—	—	
suburban	0.35	0.29, 0.41	<0.001
urban	1.1	1.0, 1.1	<0.001
dens	0.00	0.00, 0.00	<0.001
bc	0.67	0.66, 0.68	<0.001
hyads	2.1	1.8, 2.3	<0.001
pct_white * pct_pov	0.00	0.00, 0.00	<0.001
bc * hyads	-0.25	-0.27, -0.23	<0.001

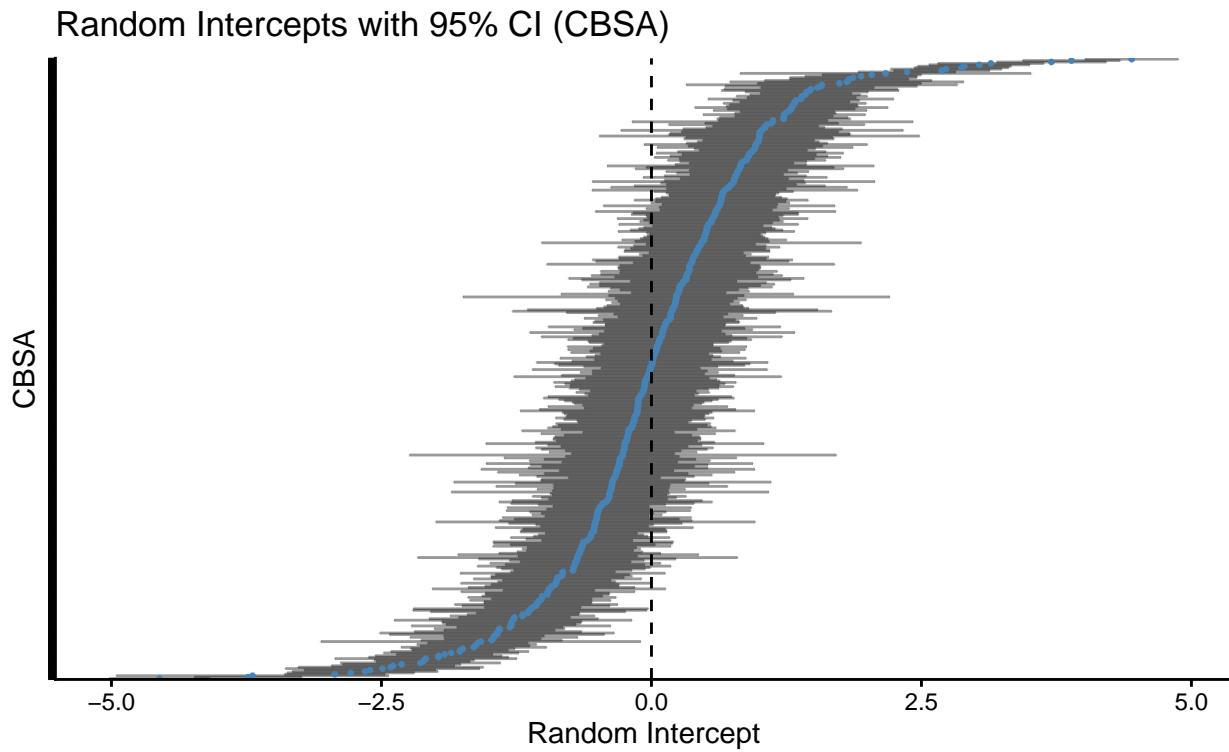
Abbreviation: CI = Confidence Interval

Checking random effects too

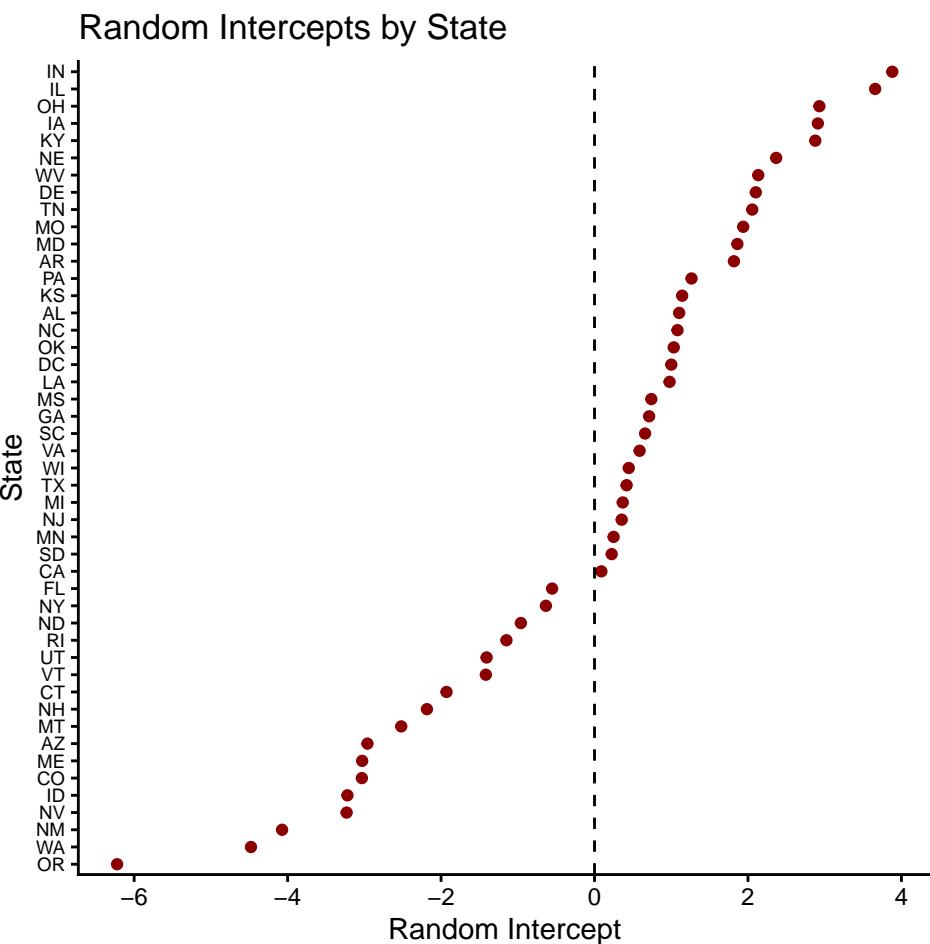
Distribution of Random Intercepts (CBSAs)



CBSA RI within each state

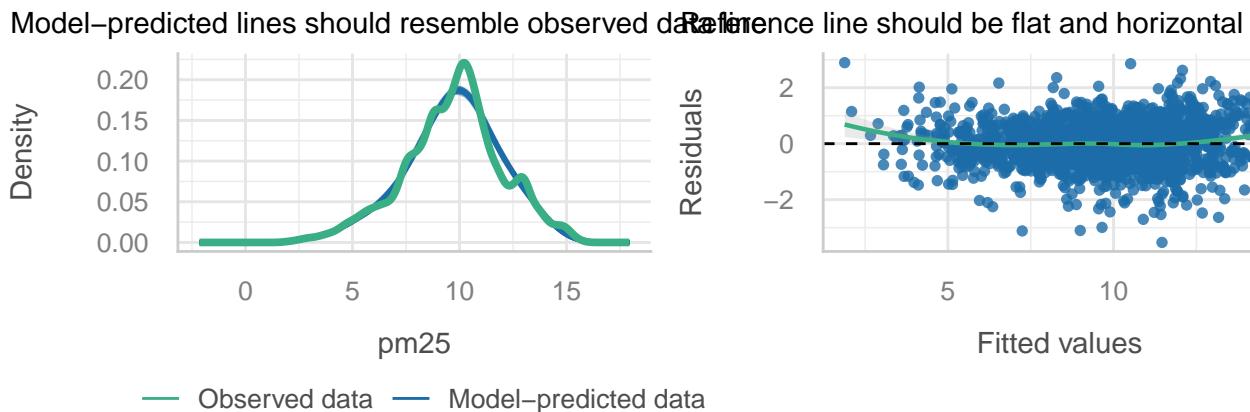


State-level RI

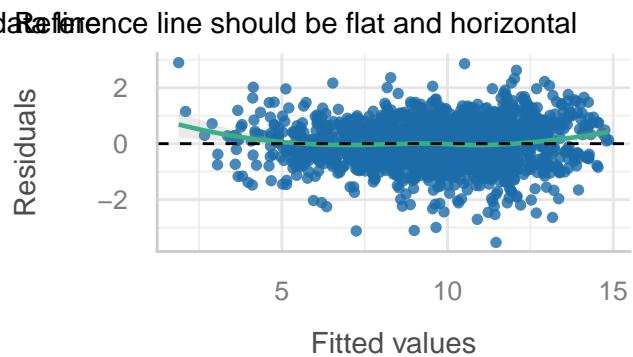


Model Diagnostics

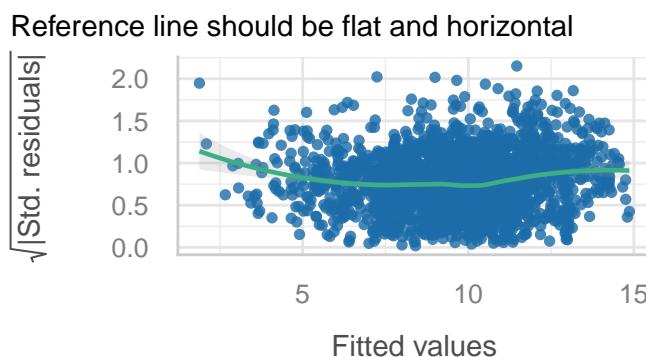
Posterior Predictive Check



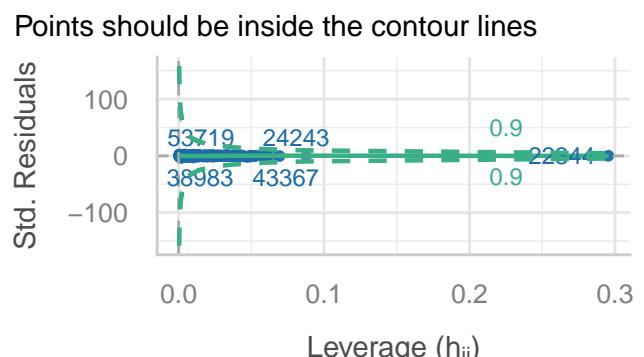
Linearity



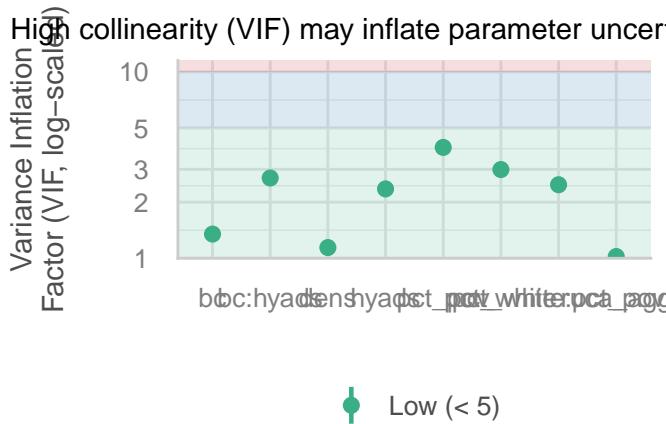
Homogeneity of Variance



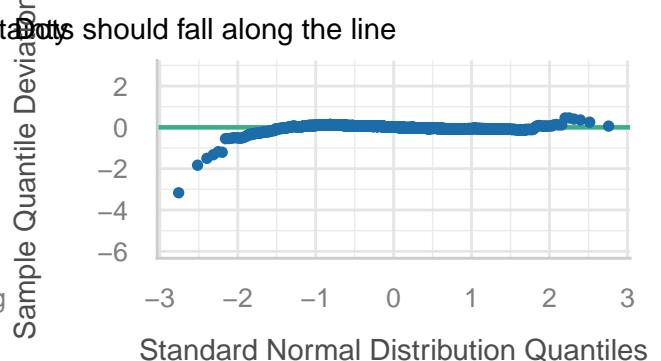
Influential Observations



Collinearity

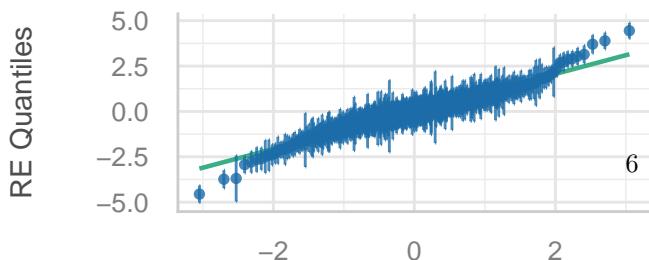


Normality of Residuals

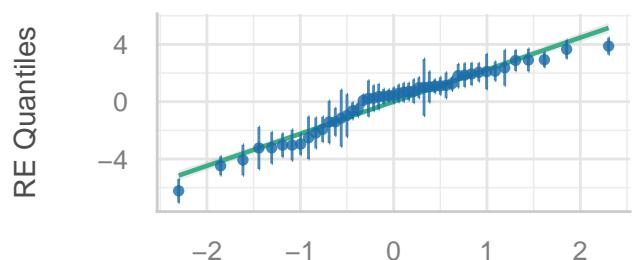


Normality of Random Effects (cbsa_id:state_abbr)

Dots should be plotted along the line

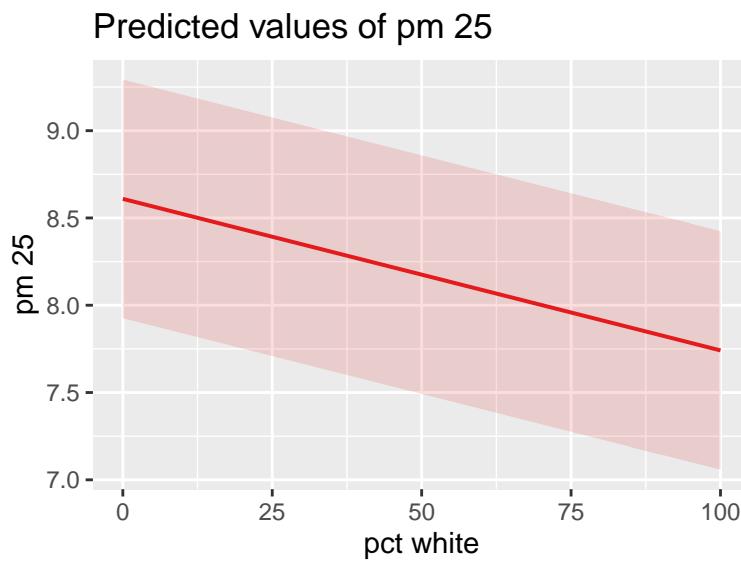


Dots should be plotted along the line

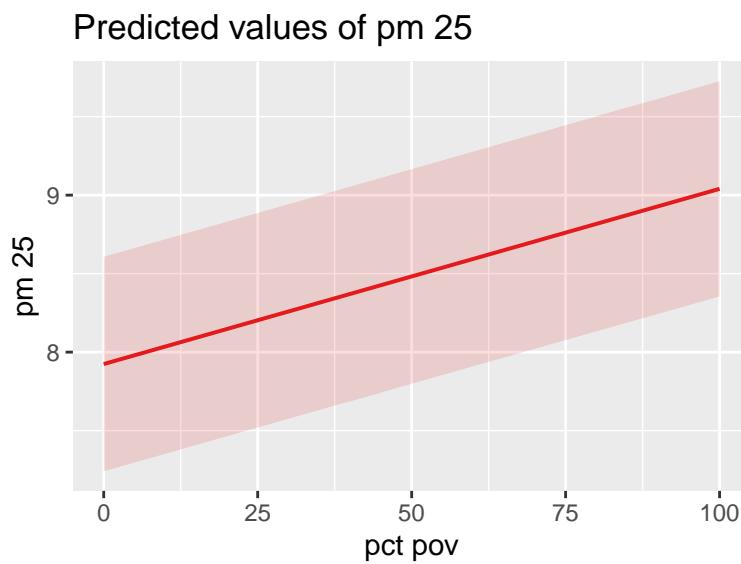


Prediction

```
## $pct_white
```

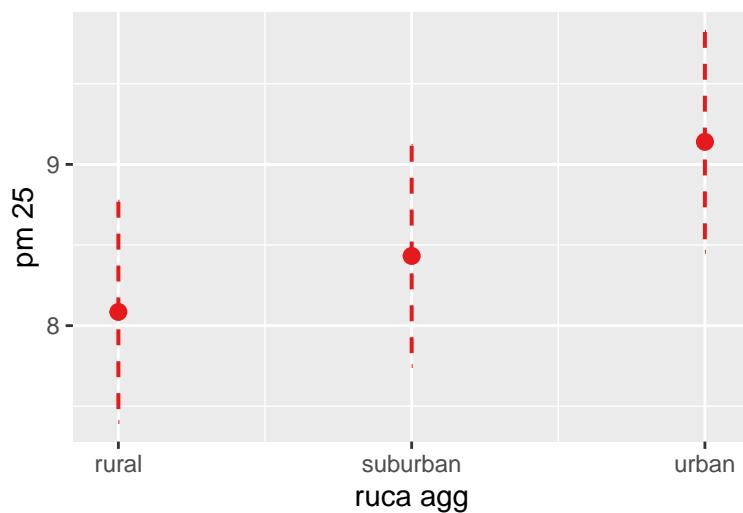


```
##  
## $pct_pov
```



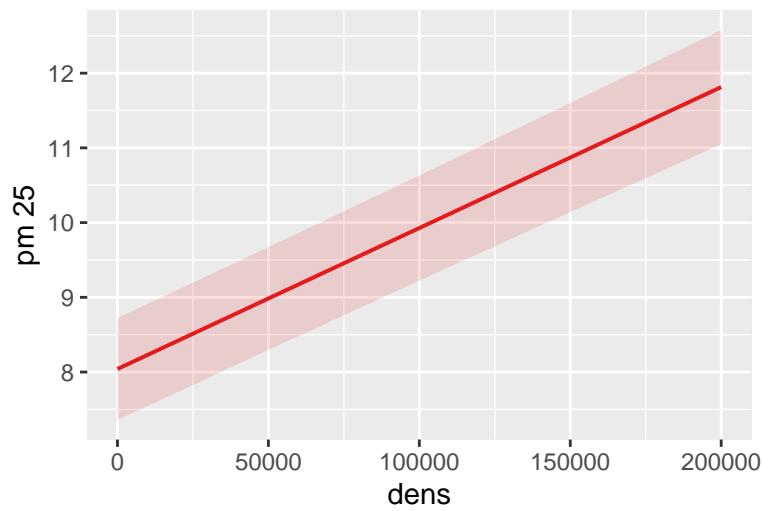
```
##  
## $ruca_agg
```

Predicted values of pm 25



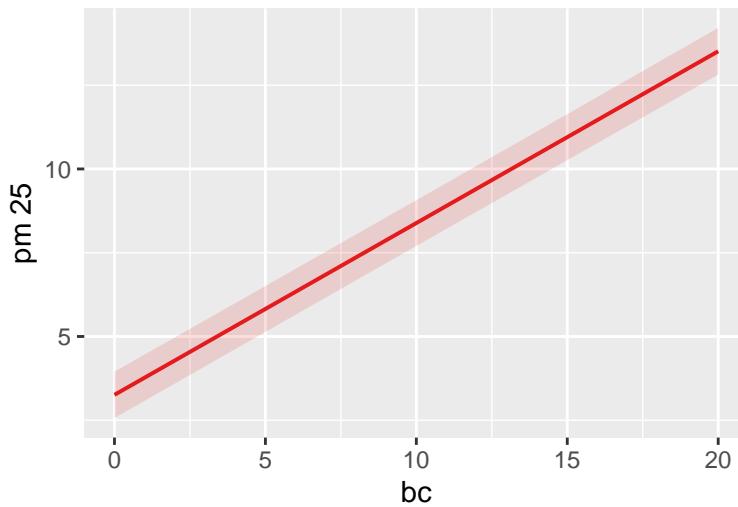
```
##  
## $dens
```

Predicted values of pm 25



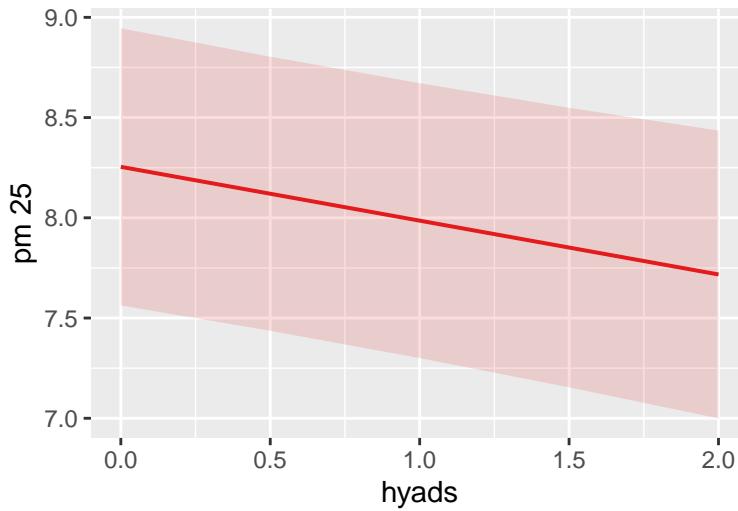
```
##  
## $bc
```

Predicted values of pm 25



```
##  
## $hyads
```

Predicted values of pm 25



Interaction Term

```
## NULL
```

```
## Data: tract
## Models:
## nested_mod2: pm25 ~ pct_white * pct_pov + ruca + dens + area + bc + hyads + (1 | cbsa_id) + (1 | state_abb/cbsa_id)
## nested_mod: pm25 ~ pct_white * pct_pov + ruca_agg + dens + bc * hyads + (1 | state_abb/cbsa_id)
##          npar    AIC    BIC logLik -2*log(L) Chisq Df Pr(>Chisq)
## nested_mod2   12 132905 133013 -66441     132881
## nested_mod    13 135440 135557 -67707     135414      0  1           1
```