American Youth Tobacco Use: an Analysis of the American National Youth Tobacco Survey 2014

Vinky Wang

30/12/2020

Introduction

Tobacco use in any form is unsafe for youth as nicotine exposure during adolescence may have lasting adverse consequences for brain development, causes addiction, and may lead to sustained tobacco use. [1] Smoking activity tends to differ across demographic characteristics including age, sex, race, and rurality. Additionally, variation may exist between schools and/or across U.S states. This report aims to investigate the prevalence of tobacco use in U.S middle and high schools across the variables of age, sex, and rurality for White, Black, and Hispanic American youths.

Methods

The data is available through the 2014 National Youth Tobacco Survey (NYTS) which is a national pencil-and-paper questionnaire including responses on tobacco-related beliefs, attitudes, behaviours, and exposure to pro- and anti-tobacco influences. Through an exploratory analysis in Figure 1 on the total number of responses across age, only the 11 to 18 year old age groups were analysed for reliability purposes. The data contains 17421 observations in total for White, Black, and Hispanic American youths once removing the 9, 10, and 19 year old age groups.

A hiearchical logistic model with interaction effects of the age, sex, race, and rurality variables was fit to the binary response of smoking cigarettes for individual k of school j in US state i. The random effects are at the school and state level to account for their potential variations. The model was implemented in the INLA, or Integrated Nested Laplace Approximation software in R which is a computationally efficient method for implementing Bayesian models.[2] The strategy='laplace' and fast=FALSE specifications were set to approximate the conditional modes.

Model

$$Y_{ijk} \sim Bernoulli(\theta_{ijk})$$
$$logit(\theta_{ijk}) = X_{ijk}^T\beta + U_{ij} + U_i$$
$$\beta \sim N(0, 10^2)$$
$$U_{ij} \overset{iid}{\sim} N(0, \sigma_1^2)$$
$$U_i \overset{iid}{\sim} N(0, \sigma_2^2)$$

$$\sigma_1^2 \sim \text{Exponential prior such that } Pr(\sigma_1^2 > \frac{log(1.2)}{2.6}) = 0.5$$

$$\sigma_2^2 \sim \text{Exponential prior such that } Pr(\sigma_2^2 > \frac{log(3)}{2.6}) = 0.5$$

where

- i = 1,. . . ,33 US states, j = 1,. . . , 188 schools, k= 1,. . . ,17421 individuals
- Y_{ijk}: Response of smoking cigarettes for individual k of school j in state i; 0=No, 1=Yes
- θ_{ijk}: Probability of smoking cigarettes
- U_{ij}: Random effect for school j in state i
- U_i: Random effect for state i

Prior Specification

The following information was provided:

(1) It is expected that the 'worst' states will have double to triple the rate of smoking than the 'healthiest'
(2) Within a given state, the 'worst' schools will have a 10-20% greater smoking rate than the 'healthiest'
(3) 'Worst' corresponds to a 90th percentile and 'healthiest' corresponds to a 10th percentile

A default weakly informative zero-mean Gaussian prior was assigned to the regression parameters β. The penalized complexity prior is an exponential of the square root of the Kullback-Leibler distance which gives preference to less complex models unless the data informs otherwise.[3] The prior for σ_1^2 such that $Pr(\sigma_1^2 > \frac{log(1.2)}{2.6}) = 0.5$ encodes line (2) of the provided information that the median of the school-level variability is expected to be no greater than 20% between the 'worst' and 'healthiest' schools. The prior for σ_2^2 such that $Pr(\sigma_2^2 > \frac{log(3)}{2.6}) = 0.5$ encodes line (3) that the median of the state-level variability is expected to be no greater than triple the occurrence between the 'worst' and 'healthiest' states.

Results

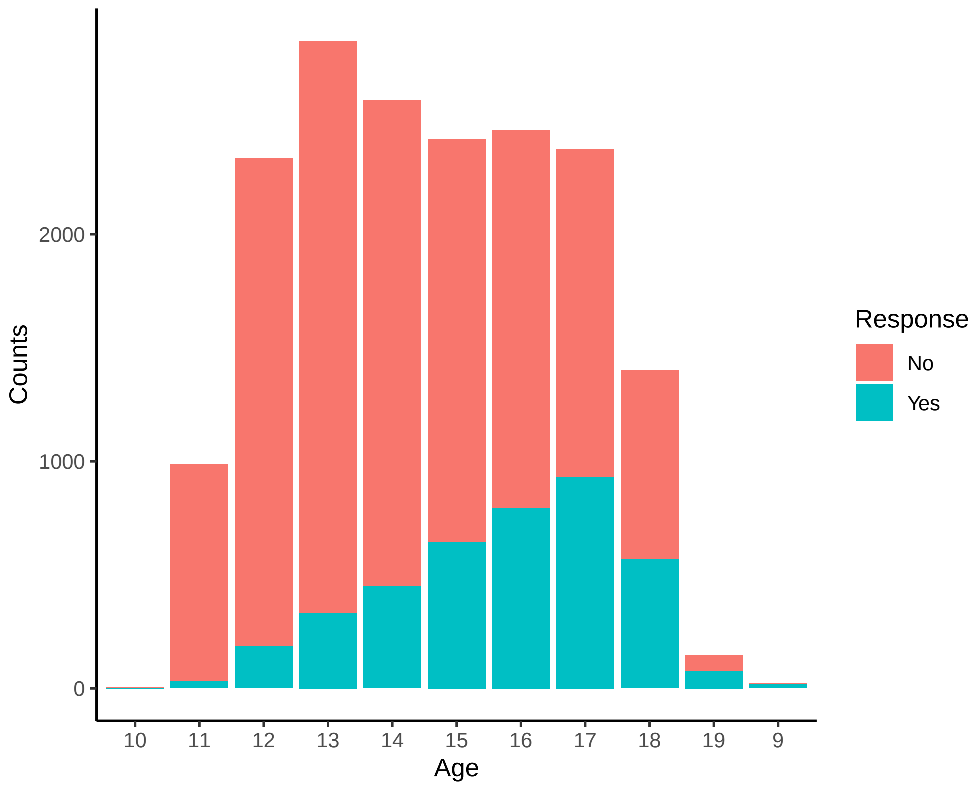
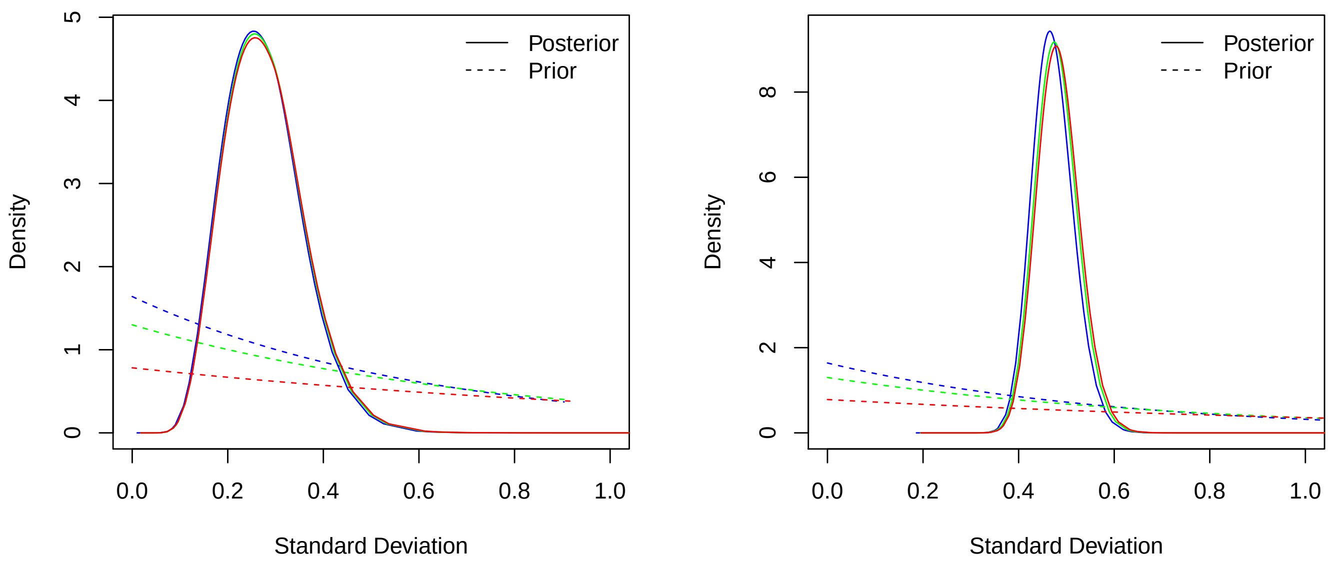


Figure 1: Exploratory analysis of cigarette smoking activity for White, Black, and Hispanic American youths by age. Ages 9, 10, and 19 will be excluded in the analysis as responses appear to be unreliable.



(a) by State

(b) by School

Figure 2: Prior Sensitivity Analysis. Posterior distributions closely overlap suggesting that priors are not sensitive.

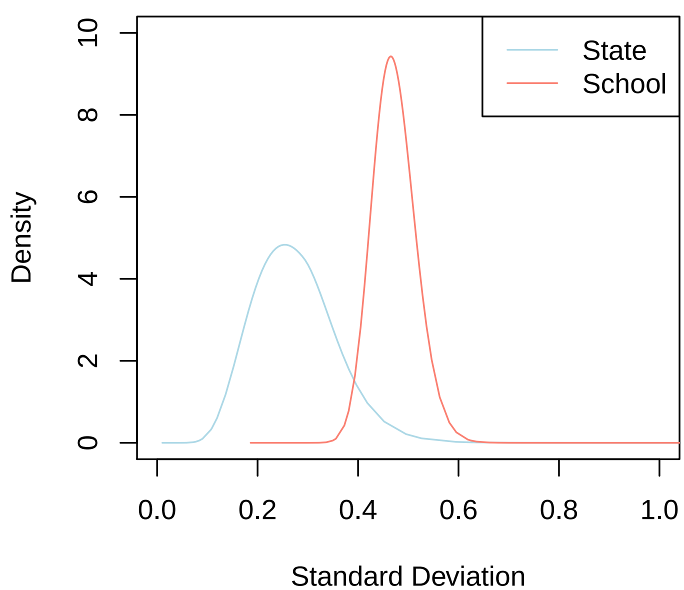
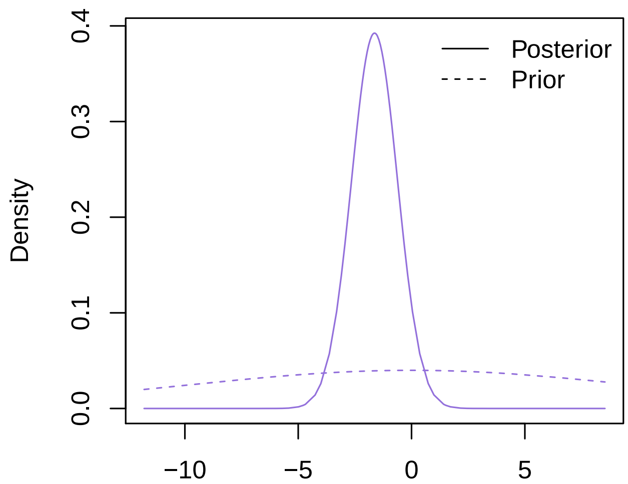
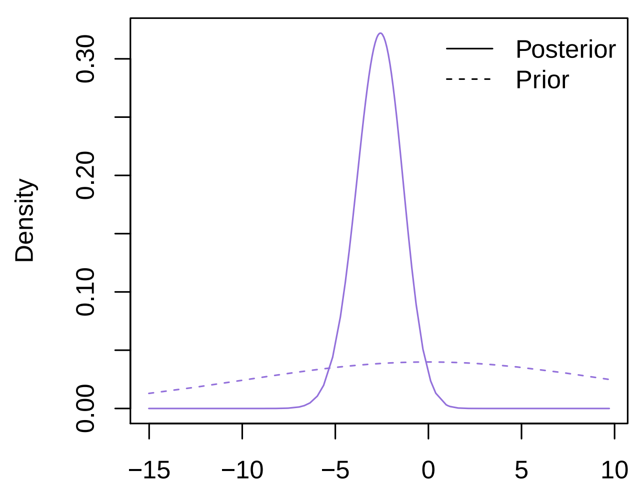


Figure 3: Posterior distributions of the fitted model.



(a) Intercept



(b) Age 11, Urban, Male, White American youth

Figure 4: Marginal Effects.

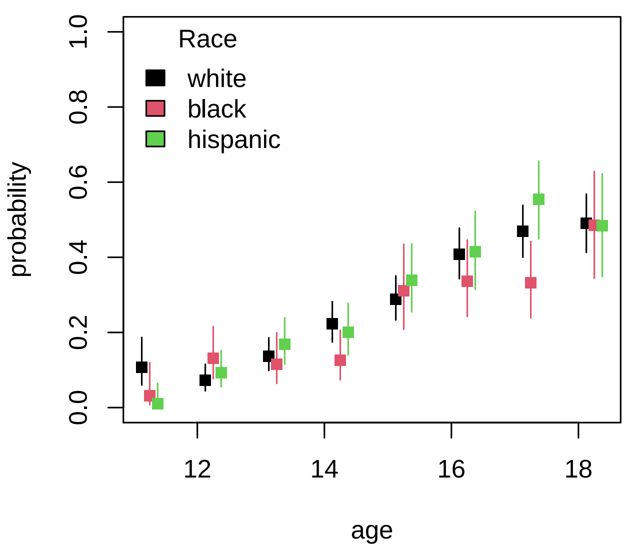


Figure 5: Smoking Activity by Age.

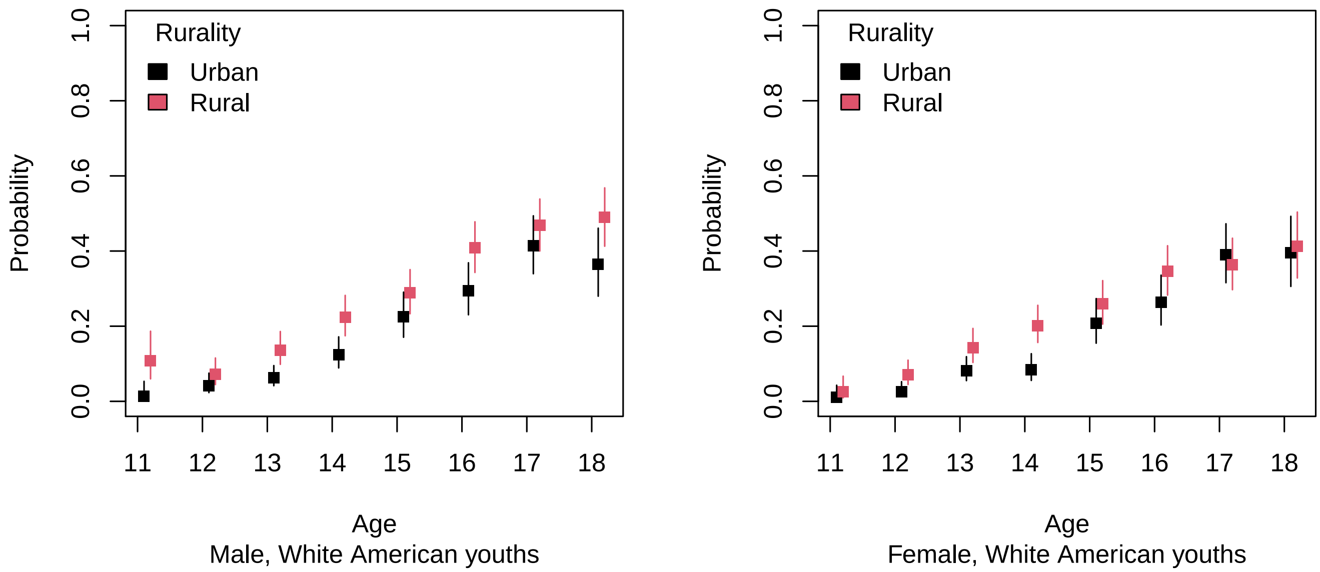


Figure 6: Smoking Activity by Age.

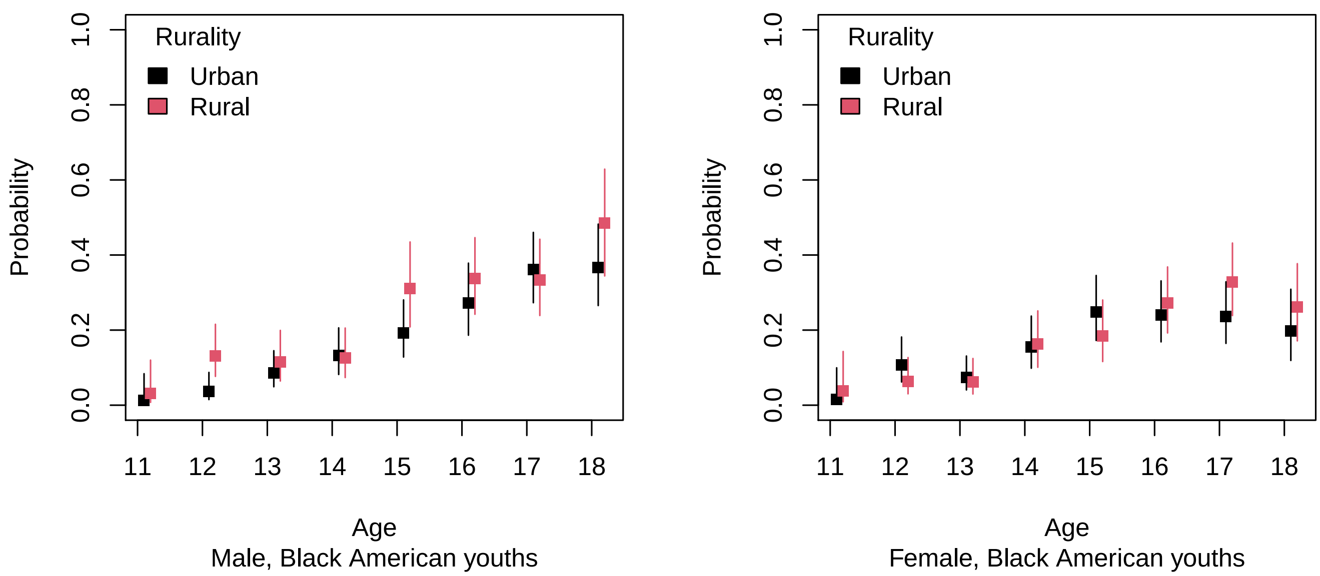


Figure 7: Smoking Activity by Age.

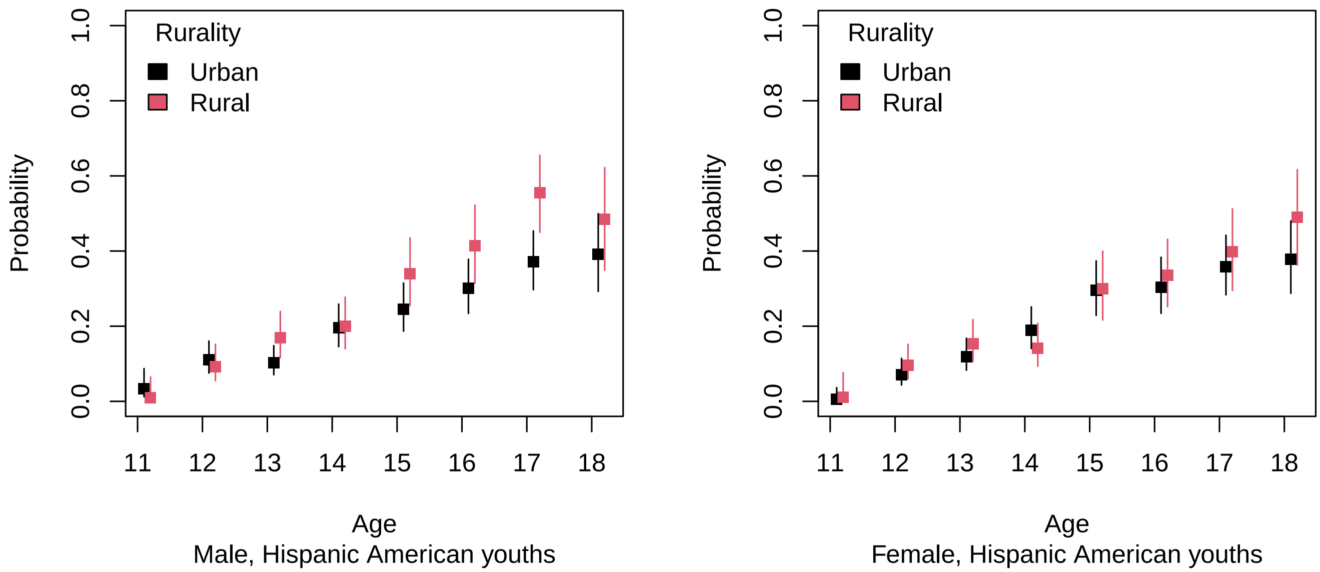


Figure 8: Smoking Activity by Age.

Table 1: Estimated Probability of Smoking by Age and Rurality for White Male American Youths

Age	RuralUrban	0.5quant	0.025quant	0.975quant
11	Urban	0.0136021	0.0033630	0.0532982
11	Rural	0.1078550	0.0598358	0.1866251
12	Urban	0.0420158	0.0231568	0.0750955
12	Rural	0.0722691	0.0444371	0.1153595
13	Urban	0.0631709	0.0415150	0.0951603
13	Rural	0.1364335	0.0984372	0.1859275
14	Urban	0.1243047	0.0889004	0.1716353
14	Rural	0.2237226	0.1744271	0.2819987
15	Urban	0.2245333	0.1704965	0.2906198
15	Rural	0.2884096	0.2330453	0.3506384
16	Urban	0.2943686	0.2303595	0.3687327
16	Rural	0.4087850	0.3429463	0.4778514
17	Urban	0.4139608	0.3395610	0.4938036
17	Rural	0.4688531	0.4001225	0.5385196
18	Urban	0.3652510	0.2797866	0.4609205
18	Rural	0.4903647	0.4128071	0.5681795

Table 2: Estimated Probability of Smoking by Age and Rurality for White Female American Youths

Age	RuralUrban	0.5quant	0.025quant	0.975quant
11	Urban	0.0108956	0.0026986	0.0428795
11	Rural	0.0253997	0.0093599	0.0670002
12	Urban	0.0258056	0.0125730	0.0522351

Age	RuralUrban	0.5quant	0.025quant	0.975quant
12	Rural	0.0706490	0.0447946	0.1096242
13	Urban	0.0813972	0.0550927	0.1188995
13	Rural	0.1429184	0.1034152	0.1940425
14	Urban	0.0845226	0.0555584	0.1267993
14	Rural	0.2016776	0.1565615	0.2555864
15	Urban	0.2075468	0.1547076	0.2734869
15	Rural	0.2592974	0.2054490	0.3213475
16	Urban	0.2637390	0.2031648	0.3357704
16	Rural	0.3455873	0.2826446	0.4141589
17	Urban	0.3907283	0.3156618	0.4725964
17	Rural	0.3629698	0.2970305	0.4342051
18	Urban	0.3949438	0.3058600	0.4924311
18	Rural	0.4133739	0.3282521	0.5037964

Table 3: Estimated Probability of Smoking by Age and Rurality for Black Male American Youths

Age	RuralUrban	0.5quant	0.025quant	0.975quant
11	Urban	0.0126376	0.0017846	0.0838234
11	Rural	0.0309243	0.0073964	0.1200530
12	Urban	0.0366441	0.0148900	0.0873286
12	Rural	0.1313710	0.0766839	0.2155281
13	Urban	0.0855380	0.0490353	0.1451530
13	Rural	0.1158313	0.0643975	0.1992689
14	Urban	0.1318907	0.0817630	0.2060015
14	Rural	0.1254801	0.0736335	0.2054685
15	Urban	0.1930191	0.1282156	0.2803865
15	Rural	0.3104666	0.2083831	0.4347344
16	Urban	0.2717136	0.1863959	0.3783698
16	Rural	0.3368191	0.2422080	0.4462412
17	Urban	0.3610502	0.2729309	0.4603203
17	Rural	0.3328952	0.2388788	0.4420277
18	Urban	0.3669119	0.2653508	0.4822932
18	Rural	0.4854746	0.3442770	0.6287531

Table 4: Estimated Probability of Smoking by Age and Rurality for Black Female American Youths

Age	RuralUrban	0.5quant	0.025quant	0.975quant
11	Urban	0.0151800	0.0021406	0.0995759
11	Rural	0.0374650	0.0089806	0.1430531
12	Urban	0.1078270	0.0618110	0.1815365
12	Rural	0.0631418	0.0302608	0.1268743
13	Urban	0.0738846	0.0405564	0.1309134
13	Rural	0.0619841	0.0297694	0.1243998
14	Urban	0.1555424	0.0984423	0.2373173
14	Rural	0.1626273	0.1008995	0.2512071
15	Urban	0.2488481	0.1723575	0.3455306
15	Rural	0.1845971	0.1162247	0.2801536
16	Urban	0.2404443	0.1686390	0.3310951
16	Rural	0.2716417	0.1923840	0.3682954

Age	RuralUrban	0.5quant	0.025quant	0.975quant
17	Urban	0.2367976	0.1645632	0.3287386
17	Rural	0.3282346	0.2387068	0.4318763
18	Urban	0.1972516	0.1191139	0.3088531
18	Rural	0.2613723	0.1712748	0.3769461

Table 5: Estimated Probability of Smoking by Age and Rurality for Hispanic Male American Youths

Age	RuralUrban	0.5quant	0.025quant	0.975quant
11	Urban	0.0331003	0.0121026	0.0872639
11	Rural	0.0096343	0.0013627	0.0647523
12	Urban	0.1107059	0.0750592	0.1605825
12	Rural	0.0927072	0.0549705	0.1520149
13	Urban	0.1030134	0.0705920	0.1481574
13	Rural	0.1689537	0.1159496	0.2393773
14	Urban	0.1956945	0.1452999	0.2589617
14	Rural	0.2000346	0.1399400	0.2774797
15	Urban	0.2449574	0.1868218	0.3151074
15	Rural	0.3393141	0.2546205	0.4355385
16	Urban	0.3007796	0.2338747	0.3783407
16	Rural	0.4145523	0.3143627	0.5222720
17	Urban	0.3713533	0.2968577	0.4537000
17	Rural	0.5547483	0.4496361	0.6551176
18	Urban	0.3907989	0.2924188	0.4995685
18	Rural	0.4843837	0.3486211	0.6223433

Table 6: Estimated Probability of Smoking by Age and Rurality for Hispanic Female American Youths

Age	RuralUrban	0.5quant	0.025quant	0.975quant
11	Urban	0.0054233	0.0007789	0.0366899
11	Rural	0.0114742	0.0016226	0.0764379
12	Urban	0.0709997	0.0432794	0.1144480
12	Rural	0.0966139	0.0598742	0.1520501
13	Urban	0.1188174	0.0828163	0.1679477
13	Rural	0.1535129	0.1055742	0.2177461
14	Urban	0.1894720	0.1404322	0.2513722
14	Rural	0.1409270	0.0931627	0.2074381
15	Urban	0.2957335	0.2288272	0.3738388
15	Rural	0.3001311	0.2161572	0.3999652
16	Urban	0.3033229	0.2342101	0.3835920
16	Rural	0.3357009	0.2517949	0.4313254
17	Urban	0.3583403	0.2832935	0.4422173
17	Rural	0.3987717	0.2949308	0.5124419
18	Urban	0.3786786	0.2873741	0.4802321
18	Rural	0.4892600	0.3628976	0.6168972

Table 7: Posterior Median for Random Effects

	0.5quant	0.025quant	0.975quant
sd for state	0.2676926	0.1363356	0.4519037
sd for school	0.4698432	0.3938673	0.5625812

Discussion

The following additional priors specifications were explored for a prior sensitivity analysis:

(1)

$$Pr(\sigma_1^2 > \frac{log(1.4)}{2.6}) = 0.5 \text{ and } Pr(\sigma_2^2 > \frac{log(4)}{2.6}) = 0.5$$

(2)

$$Pr(\sigma_1^2 > \frac{log(10)}{2.6}) = 0.5 \text{ and } Pr(\sigma_2^2 > \frac{log(1.8)}{2.6}) = 0.5$$

Model (1) represents the prior belief that the smoking activity will be 40% greater for the 'worst' schools relative to the 'healthiest' and greater by a factor of four for the 'worst' states relative to the 'healthiest'. Model (2) represents the prior belief that the smoking activity will be 80% greater for the 'worst' schools relative to the 'healthiest' and greater by a factor of 10 for the 'worst' states relative to the 'healthiest'. As there is much overlap between the posterior distributions of the fitted model with model (1) and model (2) shown in Figure 2, this suggests that the priors are not sensitive and that the data is mainly informing the posteriors rather than the priors. The default weakly informative Gaussian priors assigned to the regression parameters is sensible and covers a wide range of possible values for the marginal effects as shown in Figure 4.

The posterior median for the state standard deviation is 0.273 and for school is 0.471 with 95% credible intervals of (0.142, 0.459) and (0.391, 0.560) respectively, suggesting evidence of a statistically significant difference across schools and between states. For a 1 standard deviation increase, we expect double the rate of cigarette smoking activity for the 'worst' states relative to the 'healthiest' and about triple the rate for the 'worst' schools relative to the healthiest. Hence, tobacco prevention control programs should target schools where smoking is a problem. Figure 3 and Table 7.

The greatest difference in cigarette smoking activity between urban and rural individuals of the same gender, race, and age group shown in Figure 7 to 8 and Table 1 to 6 are the following: (1) 12.5% between urban and rural male, White American youths corresponding to the 18 year age old group; (2) 9.8% between urban and rural female, White American youths corresponding to the 14 year age old group; (3) 11.8% between urban and rural male, Black American youths corresponding to the 15 year age old group; (4) 9.2% between urban and rural female, Black American youths corresponding to the 17 year age old group; (5) 18.2% between urban and rural male, Hispanic American youths corresponding to the 17 year age old group; (6) 11.2% between urban and rural female, Hispanic American youths corresponding to the 18 year age old group.

The rural-urban difference in probability of cigarette smoking among youths of the same age, gender, and race is 18.2% which is observed for the age 17, male Hispanic American group. As the posterior median for the state is 0.273 which corresponds to double the rate of cigarette smoking for the 'worst' states relative to the 'healthiest' for a 1 standard deviation increase, the differences between states is greater than the differences between rurality for comparable individuals. Figure 7 to 8 and Table 1 to 6.

Furthermore, the probability of smoking for both rural and urban youths of all gender and race tends to increase with age. The level of cigarette smoking activity across race is most consistent for the 18 year olds, and most varied for the 17 year olds. The probability of smoking cigarettes is most likely at age 18 and least at age 12 for White-, most likely at age 18 and least at age 11 for Black-, and most likely at age 17 and least at age 11 for Hispanic- American youth. Figure 6.

Conclusion

Tobacco prevention control programs should identify schools where smoking is a problem rather than targeting select U.S states. The variability between states is greater than the variability between the rurality of schools for comparable individuals. Finally, these programs should be implemented for youths early on as the probability of smoking tends to increase with age across all demographics.

References

[1] Arrazola, R. A. (2015, April 17). Tobacco Use Among Middle and High School Students - United States, 2011–2014. Retrieved November 01, 2020, from https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6414a3.htm

[2] Bayesian computing. (2015). Spatial and Spatio-temporal Bayesian Models with R-INLA, 75-126. doi: 10.1002/9781118950203.ch4

[3] Simpson, D., Rue, H., Riebler, A., Martins, T. G., & Sørbye, S. H. (2017). Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors. Statistical Science, 32(1), 1-28. doi:10.1214/16-sts576

Appendix

```r
kable(theEstimatedProb, format="pandoc", caption="The Estimated Probabilities")
```

Table 8: The Estimated Probabilities

	ageFac	RuralUrban	Sex	Race	Age	0.5quant	0.025quant	0.975quant	theSd
lc01	11	Urban	M	white	11	0.0136021	0.0033630	0.0532982	0.7170803
lc02	12	Urban	M	white	12	0.0420158	0.0231568	0.0750955	0.3136705
lc03	13	Urban	M	white	13	0.0631709	0.0415150	0.0951603	0.2260194
lc04	14	Urban	M	white	14	0.1243047	0.0889004	0.1716353	0.1918707
lc05	15	Urban	M	white	15	0.2245333	0.1704965	0.2906198	0.1757489
lc06	16	Urban	M	white	16	0.2943686	0.2303595	0.3687327	0.1703678
lc07	17	Urban	M	white	17	0.4139608	0.3395610	0.4938036	0.1631956
lc08	18	Urban	M	white	18	0.3652510	0.2797866	0.4609205	0.2010072
lc09	11	Rural	M	white	11	0.1078550	0.0598358	0.1866251	0.3267006
lc10	12	Rural	M	white	12	0.0722691	0.0444371	0.1153595	0.2626799
lc11	13	Rural	M	white	13	0.1364335	0.0984372	0.1859275	0.1879812
lc12	14	Rural	M	white	14	0.2237226	0.1744271	0.2819987	0.1578845
lc13	15	Rural	M	white	15	0.2884096	0.2330453	0.3506384	0.1463848
lc14	16	Rural	M	white	16	0.4087850	0.3429463	0.4778514	0.1429610
lc15	17	Rural	M	white	17	0.4688531	0.4001225	0.5385196	0.1423967
lc16	18	Rural	M	white	18	0.4903647	0.4128071	0.5681795	0.1596147
lc17	11	Urban	F	white	11	0.0108956	0.0026986	0.0428795	0.7151227
lc18	12	Urban	F	white	12	0.0258056	0.0125730	0.0522351	0.3733115
lc19	13	Urban	F	white	13	0.0813972	0.0550927	0.1188995	0.2138071
lc20	14	Urban	F	white	14	0.0845226	0.0555584	0.1267993	0.2302300
lc21	15	Urban	F	white	15	0.2075468	0.1547076	0.2734869	0.1837571
lc22	16	Urban	F	white	16	0.2637390	0.2031648	0.3357704	0.1743979

	ageFac	RuralUrban	Sex	Race	Age	0.5quant	0.025quant	0.975quant	theSd
lc23	17	Urban	F	white	17	0.3907283	0.3156618	0.4725964	0.1692102
lc24	18	Urban	F	white	18	0.3949438	0.3058600	0.4924311	0.2011093
lc25	11	Rural	F	white	11	0.0253997	0.0093599	0.0670002	0.5167297
lc26	12	Rural	F	white	12	0.0706490	0.0447946	0.1096242	0.2458999
lc27	13	Rural	F	white	13	0.1429184	0.1034152	0.1940425	0.1874319
lc28	14	Rural	F	white	14	0.2016776	0.1565615	0.2555864	0.1566023
lc29	15	Rural	F	white	15	0.2592974	0.2054490	0.3213475	0.1540506
lc30	16	Rural	F	white	16	0.3455873	0.2826446	0.4141589	0.1488386
lc31	17	Rural	F	white	17	0.3629698	0.2970305	0.4342051	0.1519476
lc32	18	Rural	F	white	18	0.4133739	0.3282521	0.5037964	0.1862602
lc33	11	Urban	M	black	11	0.0126376	0.0017846	0.0838234	1.0026313
lc34	12	Urban	M	black	12	0.0366441	0.0148900	0.0873286	0.4701707
lc35	13	Urban	M	black	13	0.0855380	0.0490353	0.1451530	0.3036628
lc36	14	Urban	M	black	14	0.1318907	0.0817630	0.2060015	0.2724827
lc37	15	Urban	M	black	15	0.1930191	0.1282156	0.2803865	0.2482477
lc38	16	Urban	M	black	16	0.2717136	0.1863959	0.3783698	0.2489611
lc39	17	Urban	M	black	17	0.3610502	0.2729309	0.4603203	0.2091267
lc40	18	Urban	M	black	18	0.3669119	0.2653508	0.4822932	0.2414145
lc41	11	Rural	M	black	11	0.0309243	0.0073964	0.1200530	0.7407266
lc42	12	Rural	M	black	12	0.1313710	0.0766839	0.2155281	0.3047651
lc43	13	Rural	M	black	13	0.1158313	0.0643975	0.1992689	0.3274197
lc44	14	Rural	M	black	14	0.1254801	0.0736335	0.2054685	0.3005348
lc45	15	Rural	M	black	15	0.3104666	0.2083831	0.4347344	0.2731249
lc46	16	Rural	M	black	16	0.3368191	0.2422080	0.4462412	0.2355624
lc47	17	Rural	M	black	17	0.3328952	0.2388788	0.4420277	0.2358571
lc48	18	Rural	M	black	18	0.4854746	0.3442770	0.6287531	0.2983598
lc49	11	Urban	F	black	11	0.0151800	0.0021406	0.0995759	1.0044992
lc50	12	Urban	F	black	12	0.1078270	0.0618110	0.1815365	0.3092861
lc51	13	Urban	F	black	13	0.0738846	0.0405564	0.1309134	0.3237791
lc52	14	Urban	F	black	14	0.1555424	0.0984423	0.2373173	0.2668319
lc53	15	Urban	F	black	15	0.2488481	0.1723575	0.3455306	0.2370291
lc54	16	Urban	F	black	16	0.2404443	0.1686390	0.3310951	0.2272945
lc55	17	Urban	F	black	17	0.2367976	0.1645632	0.3287386	0.2320606
lc56	18	Urban	F	black	18	0.1972516	0.1191139	0.3088531	0.3045675
lc57	11	Rural	F	black	11	0.0374650	0.0089806	0.1430531	0.7422828
lc58	12	Rural	F	black	12	0.0631418	0.0302608	0.1268743	0.3918976
lc59	13	Rural	F	black	13	0.0619841	0.0297694	0.1243998	0.3904579
lc60	14	Rural	F	black	14	0.1626273	0.1008995	0.2512071	0.2789654
lc61	15	Rural	F	black	15	0.1845971	0.1162247	0.2801536	0.2763985
lc62	16	Rural	F	black	16	0.2716417	0.1923840	0.3682954	0.2279976
lc63	17	Rural	F	black	17	0.3282346	0.2387068	0.4318763	0.2255809
lc64	18	Rural	F	black	18	0.2613723	0.1712748	0.3769461	0.2736217
lc65	11	Urban	M	hispanic	11	0.0331003	0.0121026	0.0872639	0.5234956
lc66	12	Urban	M	hispanic	12	0.1107059	0.0750592	0.1605825	0.2184949
lc67	13	Urban	M	hispanic	13	0.1030134	0.0705920	0.1481574	0.2110914
lc68	14	Urban	M	hispanic	14	0.1956945	0.1452999	0.2589617	0.1836020
lc69	15	Urban	M	hispanic	15	0.2449574	0.1868218	0.3151074	0.1769574
lc70	16	Urban	M	hispanic	16	0.3007796	0.2338747	0.3783407	0.1758172
lc71	17	Urban	M	hispanic	17	0.3713533	0.2968577	0.4537000	0.1724050
lc72	18	Urban	M	hispanic	18	0.3907989	0.2924188	0.4995685	0.2247232
lc73	11	Rural	M	hispanic	11	0.0096343	0.0013627	0.0647523	1.0004282
lc74	12	Rural	M	hispanic	12	0.0927072	0.0549705	0.1520149	0.2867384

	ageFac	RuralUrban	Sex	Race	Age	0.5quant	0.025quant	0.975quant	theSd
lc75	13	Rural	M	hispanic	13	0.1689537	0.1159496	0.2393773	0.2229595
lc76	14	Rural	M	hispanic	14	0.2000346	0.1399400	0.2774797	0.2187716
lc77	15	Rural	M	hispanic	15	0.3393141	0.2546205	0.4355385	0.2075636
lc78	16	Rural	M	hispanic	16	0.4145523	0.3143627	0.5222720	0.2213593
lc79	17	Rural	M	hispanic	17	0.5547483	0.4496361	0.6551176	0.2149371
lc80	18	Rural	M	hispanic	18	0.4843837	0.3486211	0.6223433	0.2865083
lc81	11	Urban	F	hispanic	11	0.0054233	0.0007789	0.0366899	0.9908346
lc82	12	Urban	F	hispanic	12	0.0709997	0.0432794	0.1144480	0.2674691
lc83	13	Urban	F	hispanic	13	0.1188174	0.0828163	0.1679477	0.2049752
lc84	14	Urban	F	hispanic	14	0.1894720	0.1404322	0.2513722	0.1835730
lc85	15	Urban	F	hispanic	15	0.2957335	0.2288272	0.3738388	0.1781397
lc86	16	Urban	F	hispanic	16	0.3033229	0.2342101	0.3835920	0.1809762
lc87	17	Urban	F	hispanic	17	0.3583403	0.2832935	0.4422173	0.1773517
lc88	18	Urban	F	hispanic	18	0.3786786	0.2873741	0.4802321	0.2112526
lc89	11	Rural	F	hispanic	11	0.0114742	0.0016226	0.0764379	1.0013636
lc90	12	Rural	F	hispanic	12	0.0966139	0.0598742	0.1520501	0.2637042
lc91	13	Rural	F	hispanic	13	0.1535129	0.1055742	0.2177461	0.2185449
lc92	14	Rural	F	hispanic	14	0.1409270	0.0931627	0.2074381	0.2382372
lc93	15	Rural	F	hispanic	15	0.3001311	0.2161572	0.3999652	0.2248365
lc94	16	Rural	F	hispanic	16	0.3357009	0.2517949	0.4313254	0.2069995
lc95	17	Rural	F	hispanic	17	0.3987717	0.2949308	0.5124419	0.2347066
lc96	18	Rural	F	hispanic	18	0.4892600	0.3628976	0.6168972	0.2647472

```r
#load libraries
library("INLA", verbose=FALSE)
library(knitr)
library(kableExtra)
library(knitr)
library(bookdown)
library(ggplot2)
library(FSA)
library(dplyr)
library(reshape2)


#load data
dataDir = "../data"
smokeFile = file.path(dataDir, "smoke2014.RData")
if (!file.exists(smokeFile)) {
download.file("http://pbrown.ca/teaching/appliedstats/data/smoke2014.RData", smokeFile)
}


#create dataframe
load(smokeFile)
smoke[1:3, c("Age", "ever_cigarettes", "Sex", "Race",
"state", "school", "RuralUrban")]

forInla = smoke[, c("Age", "ever_cigarettes", "Sex",
"Race", "state", "school", "RuralUrban")]

#drop NAs
```

```r
forInla = na.omit(forInla)

#recode variables
forInla$y = as.numeric(forInla$ever_cigarettes) #convert response into 0/1 encoding

#races of interest for our study
forInla_sub <- Subset(forInla,Race=="white" | Race=="black"| Race=="hispanic")

#subset for age
age_df = forInla_sub %>%
  select(Age, y)

age_df = as.data.frame.matrix(table(age_df))
age_df$age <- rownames(age_df)
names(age_df)[1] <- "No"
names(age_df)[2] <- "Yes"

age_df <- melt(age_df, "age")

p1 <- ggplot(age_df, aes(x=age,y=value, fill=variable)) +
  geom_bar(stat='identity') + labs(x="Age", y="Counts", fill="Response") + theme_classic()
p1


#remove ages 9, 10, and 19 since appears to be unreliable
#recode age as a factor
forInla_sub = subset(forInla_sub, Age >= 11 & Age<19)

forInla_sub$ageFac = factor(forInla_sub$Age)

#view the data
str(forInla_sub)
head(forInla_sub)

#create grid
toPredict = expand.grid(
  ageFac = levels(forInla_sub$ageFac),
  RuralUrban = levels(forInla_sub$RuralUrban),
  Sex = levels(forInla_sub$Sex),
  Race = levels(forInla_sub$Race)
  )

forLincombs = do.call(inla.make.lincombs,
  as.data.frame(model.matrix( ~ ageFac:RuralUrban:Sex:Race,
    data=toPredict)))


dim(toPredict) #108 items to predict, 4 terms

#model 1: priors corresponding to expected observations
fits1 = inla(y ~ ageFac:RuralUrban:Sex:Race +
    f(state, model='iid', hyper=list(
        prec=list(prior='pc.prec', param=c(log(3)/2.6, 0.5)))
```

```r
    ) + f(school, model='iid', hyper=list(
        prec=list(prior='pc.prec', param=c(log(1.2)/2.6, 0.5)))
    ),
  data=forInla_sub, family='binomial',
  control.fixed = list(mean=0, mean.intercept=0, prec=10^(-2), prec.intercept=10^(-2)),
  lincomb = forLincombs,
  control.compute =  list(config=TRUE),
  control.inla = list(strategy='laplace', fast=FALSE))
save(fits1,file ="../data/fits1.RData")

#model 2: priors corresponding to upperbound of expected observations
fits2 = inla(y ~ ageFac:RuralUrban:Sex:Race +
    f(state, model='iid', hyper=list(
        prec=list(prior='pc.prec', param=c(log(4)/2.6, 0.5)))
    ) + f(school, model='iid', hyper=list(
        prec=list(prior='pc.prec', param=c(log(1.4)/2.6, 0.5)))
    ),
  data=forInla_sub, family='binomial',
  control.fixed = list(mean=0, mean.intercept=0, prec=10^(-2), prec.intercept=10^(-2)),
  lincomb = forLincombs,
  control.inla = list(strategy='laplace', fast=FALSE),
  list(openmp.strategy="huge"))
save(fits2,file ="../data/fits2.RData")

#model 3: priors corresponding to unexpected observations
fits3 = inla(y ~ ageFac:RuralUrban:Sex:Race +
    f(state, model='iid', hyper=list(
        prec=list(prior='pc.prec', param=c(log(10)/2.6, 0.5)))
    ) + f(school, model='iid', hyper=list(
        prec=list(prior='pc.prec', param=c(log(1.8)/2.6, 0.5)))
    ),
  data=forInla_sub, family='binomial',
  control.fixed = list(mean=0, mean.intercept=0, prec=10^(-2), prec.intercept=10^(-2)),
  lincomb = forLincombs,
  control.inla = list(strategy='laplace', fast=FALSE),
  list(openmp.strategy="huge"))
save(fits3,file ="../data/fits3.RData")

#prior and posteriors of SD for each model
theSd1= Pmisc::priorPost(fits1)
theSd2= Pmisc::priorPost(fits2)
theSd3= Pmisc::priorPost(fits3)

#psa for state
{plot(theSd1$`sd for state`$posterior, type='l', xlab='Standard Deviation', ylab='Density',
     xlim = c(0,1), col='blue')
lines(theSd1$`sd for state`$prior, col='blue', lty=2)

lines(theSd2$`sd for state`$posterior, col='green')
lines(theSd2$`sd for state`$prior, col='green', lty=2)

lines(theSd3$`sd for state`$posterior, col='red')
lines(theSd3$`sd for state`$prior, col='red', lty=2)
```

```r
legend('topright',
  lty=1:2, lwd=1, legend = c('Posterior','Prior'),
  bty='n')
}

#psa for school
{plot(theSd1$`sd for school`$posterior, type='l', xlab='Standard Deviation', ylab='Density',
     xlim = c(0,1), col='blue')
lines(theSd1$`sd for school`$prior, col='blue', lty=2)

lines(theSd2$`sd for school`$posterior, col='green')
lines(theSd2$`sd for school`$prior, col='green', lty=2)

lines(theSd3$`sd for school`$posterior, col='red')
lines(theSd3$`sd for school`$prior, col='red', lty=2)

legend('topright',
  lty=1:2, lwd=1, legend = c('Posterior','Prior'),
  bty='n')
}

#Posterior for state and school
{plot(theSd1$`sd for state`$posterior, type='l', xlab='Standard Deviation', ylab='Density',
     xlim = c(0,1), col='lightblue', ylim=c(0,10))

lines(theSd1$`sd for school`$posterior, type='l',
     xlim = c(0,1), col='salmon')

legend('topright',
  lty=1, legend = c('State','School'), col=c("lightblue", "salmon"))
}

#plot beta0
{plot(fits1[["marginals.fixed"]][["(Intercept)"]], type='l', col='medium purple', xlab="", ylab="Densit
lines(fits1[["marginals.fixed"]][["(Intercept)"]][,'x'],
dnorm(fits1[["marginals.fixed"]][["(Intercept)"]][,'x'],mean=0, sd=10), col='medium purple', lty=2)
legend('topright',
  lty=1:2, lwd=1, legend = c('Posterior','Prior'),
  bty='n')
}


#plot beta1
{plot(fits1[["marginals.fixed"]][["ageFac11:RuralUrbanUrban:SexM:Racewhite"]], type='l', col='medium pu
lines(fits1[["marginals.fixed"]][["ageFac11:RuralUrbanUrban:SexM:Racewhite"]][,'x'],
dnorm(fits1[["marginals.fixed"]][["ageFac11:RuralUrbanUrban:SexM:Racewhite"]][,'x'],mean=0, sd=10), col=
legend('topright',
  lty=1:2, lwd=1, legend = c('Posterior','Prior'),
  bty='n')
}

#estimated probability
#create matrix of estimated probabilities
```

```r
theCoef = exp(fits1$summary.lincomb.derived[,
    c('0.5quant','0.025quant','0.975quant')])
theCoef = theCoef/(1+theCoef)

theSd = fits1$summary.lincomb.derived[,'sd']

#change age from factor into numeric
toPredict$Age = as.numeric(as.character(toPredict$ageFac))

theEstimatedProb = cbind(toPredict, theCoef, theSd)

#effect of age on smoking
toPredict$shiftX_race = as.numeric(toPredict$Race)/8
toPredict$x2 = toPredict$Age + toPredict$shiftX_race

{toPlot_age = toPredict$Sex == "M" & toPredict$RuralUrban ==
"Rural"
plot(toPredict[toPlot_age, "x2"], theCoef[toPlot_age, "0.5quant"],
xlab = "age", ylab = "probability", ylim = c(0,
1), pch = 15, col = toPredict[toPlot_age, "Race"])
segments(toPredict[toPlot_age, "x2"], theCoef[toPlot_age, "0.025quant"],
y1 = theCoef[toPlot_age, "0.975quant"], col = toPredict[toPlot_age,
"Race"])
legend("topleft", fill = 1:nlevels(toPredict$Race),
legend = levels(toPredict$Race), bty = "n",
title = "Race")
}

plot comparable individuals differing by rurality

#create an x axis, shift by rurality
toPredict$shiftX_rurality = as.numeric(toPredict$RuralUrban)/10
toPredict$x1 = toPredict$Age + toPredict$shiftX_rurality

#age, males, white- on rural vs urban
toPlot_mw = toPredict$Sex == "M" & toPredict$Race == "white"

{plot(toPredict[toPlot_mw, "x1"], theCoef[toPlot_mw, "0.5quant"], sub="Male, White American youths",
xlab = "Age", ylab = "Probability", ylim = c(0,
1), pch = 15, col = toPredict[toPlot_mw, "RuralUrban"])
segments(toPredict[toPlot_mw, "x1"], theCoef[toPlot_mw, "0.025quant"],
y1 = theCoef[toPlot_mw, "0.975quant"], col = toPredict[toPlot_mw,
"RuralUrban"])
legend("topleft", fill = 1:nlevels(toPredict$RuralUrban),
legend = levels(toPredict$RuralUrban), bty = "n",
title = "Rurality")
}


#age, females, white- on rural vs urban
toPlot_fw = toPredict$Sex == "F" & toPredict$Race == "white"

{plot(toPredict[toPlot_fw, "x1"], theCoef[toPlot_fw, "0.5quant"], sub="Female, White American youths",
```

```r
xlab = "Age", ylab = "Probability", ylim = c(0,
1), pch = 15, col = toPredict[toPlot_fw, "RuralUrban"])
segments(toPredict[toPlot_fw, "x1"], theCoef[toPlot_fw, "0.025quant"],
y1 = theCoef[toPlot_fw, "0.975quant"], col = toPredict[toPlot_fw,
"RuralUrban"])
legend("topleft", fill = 1:nlevels(toPredict$RuralUrban),
legend = levels(toPredict$RuralUrban), bty = "n",
title = "Rurality")
}

#age, males, black- on rural vs urban
toPlot_mb = toPredict$Sex == "M" & toPredict$Race == "black"

{plot(toPredict[toPlot_mb, "x1"], theCoef[toPlot_mb, "0.5quant"], sub="Male, Black American youths",
xlab = "Age", ylab = "Probability", ylim = c(0,
1), pch = 15, col = toPredict[toPlot_mb, "RuralUrban"])
segments(toPredict[toPlot_mb, "x1"], theCoef[toPlot_mb, "0.025quant"],
y1 = theCoef[toPlot_mb, "0.975quant"], col = toPredict[toPlot_mb,
"RuralUrban"])
legend("topleft", fill = 1:nlevels(toPredict$RuralUrban),
legend = levels(toPredict$RuralUrban), bty = "n",
title = "Rurality")
}

#age, females, black-on rural vs urban
toPlot_fb = toPredict$Sex == "F" & toPredict$Race == "black"

{plot(toPredict[toPlot_fb, "x1"], theCoef[toPlot_fb, "0.5quant"], sub="Female, Black American youths",
xlab = "Age", ylab = "Probability", ylim = c(0,
1), pch = 15, col = toPredict[toPlot_fb, "RuralUrban"])
segments(toPredict[toPlot_fb, "x1"], theCoef[toPlot_fb, "0.025quant"],
y1 = theCoef[toPlot_fb, "0.975quant"], col = toPredict[toPlot_fb,
"RuralUrban"])
legend("topleft", fill = 1:nlevels(toPredict$RuralUrban),
legend = levels(toPredict$RuralUrban), bty = "n",
title = "Rurality")
}

#age, males, hispanic- on rural vs urban
toPlot_mh = toPredict$Sex == "M" & toPredict$Race == "hispanic"

{plot(toPredict[toPlot_mh, "x1"], theCoef[toPlot_mh, "0.5quant"], sub="Male, Hispanic American youths",
xlab = "Age", ylab = "Probability", ylim = c(0,
1), pch = 15, col = toPredict[toPlot_mh, "RuralUrban"])
segments(toPredict[toPlot_mh, "x1"], theCoef[toPlot_mh, "0.025quant"],
y1 = theCoef[toPlot_mh, "0.975quant"], col = toPredict[toPlot_mh,
"RuralUrban"])
legend("topleft", fill = 1:nlevels(toPredict$RuralUrban),
legend = levels(toPredict$RuralUrban), bty = "n",
title = "Rurality")
}

#age, females, hispanic- on rural vs urban
```

```
toPlot_fh = toPredict$Sex == "F" & toPredict$Race == "hispanic"

{plot(toPredict[toPlot_fh, "x1"], theCoef[toPlot_fh, "0.5quant"], sub="Female, Hispanic American youths
xlab = "Age", ylab = "Probability", ylim = c(0,
1), pch = 15, col = toPredict[toPlot_fw, "RuralUrban"])
segments(toPredict[toPlot_fh, "x1"], theCoef[toPlot_fh, "0.025quant"],
y1 = theCoef[toPlot_fh, "0.975quant"], col = toPredict[toPlot_fh,
"RuralUrban"])
legend("topleft", fill = 1:nlevels(toPredict$RuralUrban),
legend = levels(toPredict$RuralUrban), bty = "n",
title = "Rurality")
}

#find the age group where the largest variation between urban and rural exist, for male White American
estimatedprob_mw = Subset(theEstimatedProb, toPlot_mw)
estimatedprob_mw <- estimatedprob_mw %>%
  group_by(Age, RuralUrban) %>%
  summarise(`0.5quant`, `0.025quant`, `0.975quant`)
knitr::kable(estimatedprob_mw, caption="Estimated Probability of Smoking by Age and Rurality for White

#The greatest difference in cigarette smoking activity between urban and rural male, White American you

#find the age group where the largest variation between urban and rural exist, for female White America
estimatedprob_fw = Subset(theEstimatedProb, toPlot_fw)
estimatedprob_fw <- estimatedprob_fw %>%
  group_by(Age, RuralUrban) %>%
  summarise(`0.5quant`, `0.025quant`, `0.975quant`)
knitr::kable(estimatedprob_fw, caption="Estimated Probability of Smoking by Age and Rurality for White
#The greatest difference in cigarette smoking activity between urban and rural female, White American y

#find the age group where the largest variation between urban and rural exist, for male Black American
estimatedprob_mb = Subset(theEstimatedProb, toPlot_mb)
estimatedprob_mb <- estimatedprob_mb %>%
  group_by(Age, RuralUrban) %>%
  summarise(`0.5quant`,`0.025quant`, `0.975quant`)
knitr::kable(estimatedprob_mb, caption="Estimated Probability of Smoking by Age and Rurality for Black
#The greatest difference in cigarette smoking activity between urban and rural male, Black American you

#find the age group where the largest variation between urban and rural exist, for female Black America
estimatedprob_fb = Subset(theEstimatedProb, toPlot_fb)
estimatedprob_fb <- estimatedprob_fb %>%
  group_by(Age, RuralUrban) %>%
  summarise(`0.5quant`,`0.025quant`, `0.975quant`)
knitr::kable(estimatedprob_fb, caption="Estimated Probability of Smoking by Age and Rurality for Black
#The greatest difference in cigarette smoking activity between urban and rural female, Black American y

#find the age group where the largest variation between urban and rural exist, for male Hispanic Americ
estimatedprob_mh = Subset(theEstimatedProb, toPlot_mh)
estimatedprob_mh <- estimatedprob_mh %>%
  group_by(Age, RuralUrban) %>%
  summarise(`0.5quant`,`0.025quant`, `0.975quant`)
knitr::kable(estimatedprob_mh, caption="Estimated Probability of Smoking by Age and Rurality for Hispan
#The greatest difference in cigarette smoking activity between urban and rural male, Hispanic American
```

```
#find the age group where the largest variation between urban and rural exist, for female Hispanic Amer
estimatedprob_fh = Subset(theEstimatedProb, toPlot_fh)
estimatedprob_fh <- estimatedprob_fh %>%
  group_by(Age, RuralUrban) %>%
  summarise(`0.5quant`,`0.025quant`, `0.975quant`)
knitr::kable(estimatedprob_fh, caption="Estimated Probability of Smoking by Age and Rurality for Hispani
#The greatest difference in cigarette smoking activity between urban and rural female, Hispanic America

#The largest rural-urban difference in probability of cigarette smoking among youths of the same age, g
#As the posterior median for the state is 0.273 which corresponds to double the rate of cigarette smoki

kable(theSd1$summary[,c('0.5quant','0.025quant','0.975quant')], caption= "Posterior Median for Random E
```