STA442_HW4_Q1

Jiayue_Wu

11/30/2021

Demographic and Geographic factors exploration for Donald Trump Support.

1. Introduction:

The president election in U.S has attracted a high level of attention from people all over the world. Trump won the election in 2016, and I am curious of the geographic factors that influenced the voters in Wisconsin to vote for Trump. Here in this report, our research questions are finding important demographic factors that cause spatial pattern in Trump support, whether they are urban/rural, any racial phenomenon for White voters, and so on.

2. Data:

In this study we consider number of votes for Trump and consider *propWhite*, proportion of each region which is White; *propInd*, proportion of each region which is Indigenous; *logPends*, log ratio of total population, surface area the factors that might influence voters in Wisconsin to vote for Trump. Since from bellowing *Plot 1*, as total population higher (*logPends*), i.e., as the color in (b) deeper, the greater number of voters for Trump, i.e., the color in (a) deeper. Similarly, as the proportion of Indegenous increase in (c), number of voters for Trump seems decrease; and as the proportion of White increase in (d), number of voters for Trump seems increase. For more precisely and creditable exploration, we build model in the following part.

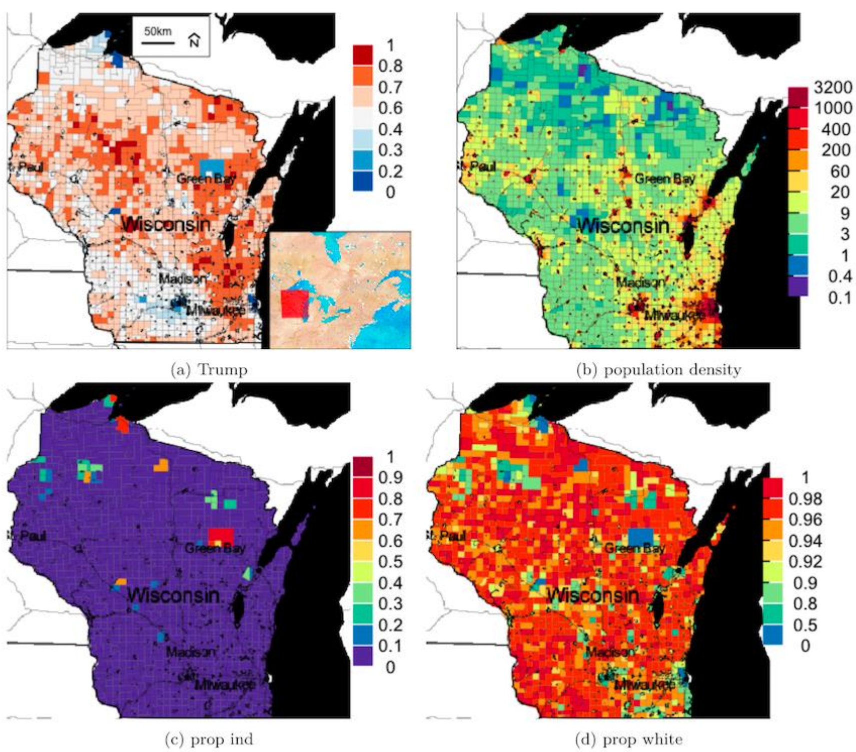


Figure 1: Figure about data info for variables

3. Method & Model

Here when fitting the model, we assume the response variable *trump* follows a Bionomial distribution, since the response variable is the number of voters, which is positive concrete. Furthering, for each value of the response variable, it is a binary decision. More precisely, the model I described as following is the Generalized Mixed model, the mathematical expression is as following.

$$Y_i \sim Bionomial(N_i, \rho_i)$$

$$log(\frac{\rho_i}{1-\rho_i}) = \mu + X_i\beta + U_i$$

$$U_i = BYM(\sigma^2, \tau^2)$$

- Y_i: response variable, which represents the number of people vote for Trump in sub-region i.
- N_i: the total number of votes in sub-region i.
- ρ_i: the probability of being the voter of Trump in sub-region i, i.e., the probability of voting for Trump in sub-region i. Here $\frac{\rho_i}{1-\rho_i}$ is the odds ratio and the $log\left(\frac{\rho_i}{1-\rho_i}\right)$ represents for the log-odds.
- μ: the overall intercept.
- X_i: the fixed effect in the model, where we set *logPdens*, *propWhite*, and *propInd* all fixed effect. More precisely, *logPdens* is the log-ratio of total population and surface area (square km), *propWhite* is the proportion of White for each sub-region, *propInd* is the proportion of Indigenous for each sub-region.
- β: the corresponding parameter for each fixed effect listed above.
- U_i: the spatial random effect. We set the spatial random effect in this model since there may exist correlation or dependence of whether voting for Trump witnin each region.
- σ: the spatially structed variance parameter σ^2.
- τ: the spatially independent variance τ^2.

More precisely, the details of the BYM are in the following.

$$U_i = W_i + V_i$$
$$V_i \sim i.i.d.\ N(0, \tau^2)$$
$$W_i|\{W_j: j \neq i\} \sim N(mean\{W_j: j \sim i\}, \frac{\sigma^2}{|j \sim i|})$$

$-$ V_i: the independent noise.
- W_i: a spatial improper Conditionally Autoregressive Model (iCAR), which involves a Markov random field whose parameters depends on the neighbours ~ i $\{W_j: j \neq i\}$.

What is more, for we can calculate θ_1 and θ_2 by following equation:
- $\theta_1 = \sqrt{\sigma^2 + \tau^2}$, the marginal standard deviation.
- $\theta_2 = \sigma/\sqrt{\sigma^2 + \tau^2}$, which is the spatial proportion prior.

And we set the prior for θ_1 and θ_2 as:

$$prob(\theta_1 > log(2.5)) = 0.5$$

$$prob(\theta_2 < 0.5) = 0.5$$

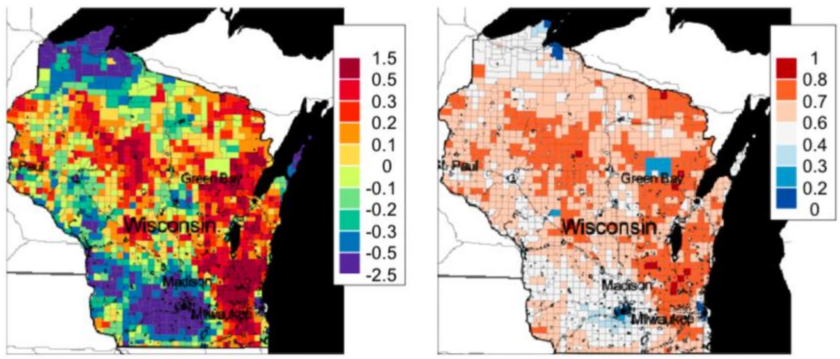
which is relatively small and will conduct little effect on the results of the model.

4. Result

Here for better understanding about the effect of factors listed above on the voting proportion for Trump, I made a summary model. Furthering, for better interpretation, I exponential the summary results for fixed effect. More precisely, for some specific parameter β_i, $log\left(\frac{\pi}{1-\pi}\right) = \mu + \beta_i x$, i.e., $\frac{\pi}{1-\pi} = e^{\mu+\beta_i x}$,for 1 unit increase in x, we will get $\frac{\pi}{1-\pi} = e^{\mu+\beta_i(x+1)}$, after the deviation for these two equation, we will the the value of

e^{β_i} is change of odds ratio. Moreover, if the 0 included in the original summary table, i.e., 1 included in the exponential result, we have no strong statistical evidence to show that there is effect.

Firstly, comparing two figures in the Figure 2, we can notice that there exist spatial difference in the probability voting for Trump. The scale of random figure is more than the scale in fitted model. Furthering, the fitted model is similar to the plot with original data shown in Figure 1: (a) Trump, which also indicate our model fit well. Further detailing about parameters will be presented in *Table 1*.



(a) random (b) fitted

Figure 2: Figures of fitted effects and random effect

Table 1: Parameter posterior means and 95% credible intervals

	0.5quant	0.025quant	0.975quant
Intercept	0.5698	0.4374	0.7435
Log Population density (square km)	0.9221	0.9141	0.9302
Proportion of White	4.1307	3.1645	5.3808
Proportion of Indegenous	0.4541	0.3216	0.6400
sd	0.3185	0.3045	0.3350
Proportion of Spatial	0.9603	0.9183	0.9861

-Population density: holing all other factors the same, as the log population density increase by 1 unit, the odd ratio of log population density is 0.9221 of the original one, i.e., the odds ratio decrease 1-0.9221 = 7.79%.
- Proportion of White: holding all other factors the same, the odds ratio of probability voting for Trump in a sub-region all White is 4.1307 times of that odds ratio probability in a sub-region with no White.
- Proportion of White: holding all others the same, the odds ratio of probability voting for Trump in a sub-region all Indigenous is 0.4541 times of that odds ratio probability in a sub-region with no Native American.
- sd: as the standard deviation of spatial U_i increase by 1, the log odds ratio of probability voting for Trump is 0.3183, i.e., increase the odds ratio of probability voting for Trump 37.48% ($e^{0.3183} - 1 = 0.3748$)
- *propSpatial*: the number of Spatial proportion is very close to 1, which indicates a strong spatial effect, i.e., there should exist a strong relationship with neighbours.

All parameters do not contain 1 in the 95% CI, which means we have strong evidence for the result.

5. Conclusion

In conclusion, after the investigation, fitting modelling, and summary table result, it demonstrates a positive effect of the proportion of White while negative effect of the proportion of Indigenous and population density in the region on the probability of voting for Trump. More precisely, for the high population density, where we assume it urban area, the probability of voting Trump decrease; as for the race, White people are more likely to support Trump while Native American are more likely to not vote for Trump. In conclusion, rural White non-Native voters are more likely to vote for Trump, and there is a spatial variation in Trump support.

6. Appendix

```
(load(file ="~/Desktop/STA442_HW4/resWisconsin.rdata" ))

theColTrump = mapmisc::colourScale(wisconsinCsubm$propTrump, col='RdBu',
    breaks=sort(unique(setdiff(c(0,1, seq(0.2,0.8,by=0.1)), 0.5))), style='fixed', rev=TRUE)
theColPop = mapmisc::colourScale(wisconsinCsubm$pdens, col='Spectral',
    breaks=11, style='equal', transform='log', digits=1, rev=TRUE)
theColWhite = mapmisc::colourScale(wisconsinCsubm$propWhite,
    col='Spectral', breaks=c(0, 0.5, 0.8, 0.9,seq(0.9, 1, by=0.02)), style='fixed',  rev=TRUE)
theColInd = mapmisc::colourScale(wisconsinCsubm$propInd,
    col='Spectral', breaks=seq(0, 1, by=0.1), style='fixed',  rev=TRUE)
theBg = mapmisc::tonerToTrans(mapmisc::openmap(wisconsinCm,
 fact=2, path='stamen-toner'), col='grey30')
theInset = mapmisc::openmap(wisconsinCm, zoom=6, path='stamen-watercolor',
    crs=mapmisc::crsMerc, buffer=c(0,1500,100,700)*1000)

library('sp')
mapmisc::map.new(wisconsinCsubm, 0.85)
sp::plot(wisconsinCsubm, col = theColTrump$plot, add=TRUE, lwd=0.2)
raster::plot(theBg, add=TRUE, maxpixels=10^7)
mapmisc::insetMap(wisconsinCsubm, 'bottomright', theInset, outer=TRUE, width=0.35)
mapmisc::scaleBar(wisconsinCsubm, 'top', cex=0.8)
mapmisc::legendBreaks('topright', theColTrump, bty='n', inset=0)

b<-resTrump$parameters$summary[5:6, paste0(c(0.5, 0.025, 0.975), 'quant')]
resTable=rbind(exp(resTrump$parameters$summary)[,paste0(c(0.5, 0.025, 0.975), 'quant')])
a<-rbind(resTable[1:4,],b)
rownames(a)=c('Intercept','Population density (square km)','Proportion of White','Proportion o
f Indegenous','sd',"Proportion of Spatial")
knitr::kable(a, digits=4 , caption = "\\label{tab:tab1} Parameter posterior means and 95% cred
ible intervals"  )

mapmisc::map.new(wisconsinCsubm, 0.85)
plot(wisconsinCsubm, col = theColPop$plot, add=TRUE, lwd=0.2)
plot(theBg, add=TRUE, maxpixels=10^7)
mapmisc::legendBreaks('right', theColPop, bty='n', inset=0)

mapmisc::map.new(wisconsinCsubm, 0.85)
plot(wisconsinCsubm, col = theColInd$plot, add=TRUE, lwd=0.2)
plot(theBg, add=TRUE, maxpixels=10^7)
mapmisc::legendBreaks('right', theColInd, bty='n', inset=0)

mapmisc::map.new(wisconsinCsubm, 0.85)
plot(wisconsinCsubm, col = theColWhite$plot, add=TRUE,lwd=0.2)
plot(theBg, add=TRUE, maxpixels=10^7)
mapmisc::legendBreaks('right', theColWhite, bty='n', inset=0)

theColRandom = mapmisc::colourScale(
    resTrump$data$random.mean, col='Spectral', breaks = 11, style='quantile', rev=TRUE, dec=1)
theColFit= mapmisc::colourScale(resTrump$data$fitted.invlogit, col='RdBu', rev=TRUE,
    breaks=sort(unique(setdiff(c(0,1, seq(0.2,0.8,by=0.1)), 0.5))), style='fixed')

mapmisc::map.new(wisconsinCsubm, 0.85)
plot(resTrump$data, col = theColRandom$plot, add=TRUE, lwd=0.2)
plot(theBg, add=TRUE, maxpixels=10^7)
mapmisc::legendBreaks('topright', theColRandom)

mapmisc::map.new(wisconsinCsubm, 0.85)
plot(resTrump$data, col = theColFit$plot, add=TRUE, lwd=0.2)
plot(theBg, add=TRUE, maxpixels=10^7)
mapmisc::legendBreaks('topright', theColFit)
```