Yao Yao
Professor Patrick Brown
STA442
03, December, 2021

Question 1--Donald Duck

The Potential Factors and Laws of Voting for Trump in 2016

Introduction

The election of the President of the United States is held every four years. There was an election for President of the United States in 2016. After that, Donald Trump became the man who served as the 45th president of the United States from 2017 to 2021. This report will analyze the factors that influenced the voters in Wisconsin to vote for Trump in 2016 and will examine the distribution of votes to Trump as well as the phenomena that existed in the distribution.

Method

Our response variable is the number of votes of Trump, and we assume it follows the binomial distribution since this variable is discrete and a voter's decision is binary, whether to vote for Trump or not. Hence I used INLA to establish a generalized mixed model. The model is established as follows:

$$Y_i \sim Binomial(N_i, p_i)$$

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

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

where $[U(s1), U(s2)...U(sn)]' \sim MVN[(\mu(s1), \mu(s2)...\mu(sn))', \Sigma]$

$$cov[U(s+h), U(s)] = \sigma^2\rho(\frac{h}{\phi}, v)$$

where Y_i represents the number of votes for Trump in sub-country i.
N_i denotes the total number of votes in sub-country i.
p_i denotes the probability of voting for Trump in corresponding sub-country i.
Therefore, $\frac{p_i}{1-p_i}$ represents the odds and $log(\frac{p_i}{1-p_i})$ represents the result after taking log transformation.
X_i are the explanatory variables(fixed effects), which are log of the ratio of total population and surface area(square km), proportion of each region which is White and proportion of each region which is Indegenous in sub-country i.
μ is the fixed intercept.

U_i is the spatial Random Effect in sub-country i, follows BYM model with spatially variance parameter σ^2 and observation variance τ^2
σ^2 is spatially variance parameter and $\tau^2$is observation variance.
$cov[U(s+h), U(s)]$ represents covariance of the residual spatial variation at location s+h and location s.
ρ is the correlation function.
ϕ denotes the range parameter which is equivalent to θ_2, and υ denotes the shape parameter.
Then we define:

$$\theta_1 = \sqrt{\sigma^2 + \tau^2}$$

θ_1 represents standard deviation(sd) and

$$\theta_2 = \frac{\sigma}{\sqrt{\sigma^2 + \tau^2}}$$

θ_2 represents spatial proportion.
with prior distribution:

$$pr(\theta_1 > log(2.5)) = 0.5$$
$$pr(\theta_2 > 0.5) = 0.5$$

Bayesian inference allows us to give a distribution to parameters and analyze the posterior to get a more precise conclusion. It is reasonable to believe that the prior distributions have a prior median of log(2.5) for standard deviation $\theta_1$and 0.5 for the spatial proportion θ_2.
After building the model, I plotted 6 diagrams which demonstrated the population in Wisconsin and the situation of voting with the corresponding scale.

Result

From the summary table (table 1), we could observe that at 0.5quant, which is the median, both logPdens and propInd are smaller than $e^0 = 1$, while propWhite is greater than 1. This demonstrates if holding all other explanatory variables in the model fixed, the ratio of total population and surface area(square km) in each sub-country is increasing by 1 leads to odds of voting to Trump would decreasing 7.8%, and the proportion of each region which is Indegenous is increasing by 1 leads to odds of voting to Trump would decrease approximately 54.6%. These two factors would have a negative impact on voting for Trump. But if holding all other explanatory variables in the model fixed, the proportion of each region which is white is increasing by 1 leads to odds of voting to Trump would be 4.132 times than the original.

Besides, we could observe the values of 0.025quant and 0.975quant. There is no such variable with 95% confidence interval containing $e^0 = 1$. This indicates population density, population of white and indegenous all have a significant impact on voting for Trump in 2016. Furthermore, at 0.5 quant, propSpatial is 0.96 which is close to 1. Hence there is a strong correlation between each spatial location.

Consider Figure 1, 3, and 4. In Figure 1, we could see the color of each square is different. Blue indicates the population voting for Trump is approximately 0% while red indicates the population is almost 100%. The population is actually not evenly distributed, thus there might have a spatial variation. In Figure 3, there is serval sub-countries with a high population of indegenous (the red square). And we could see in Figure1 and 4, the corresponding squares are mostly blue. This denotes there may be a racial phenomenon. White people are more willing to vote for Trump while indegenous are not. Besides, from Figure2, most of the areas with a high density(red square) would also have a low proportion of voting for Trump in Figure1. Thus people living in rural areas are the majority supporting Trump. Moreover, the scale of spatial random effect is larger than fixed effect, which indicates there might be spatial random effect.

	0.5quant	**0.025quant**	**0.975quant**
(Intercept)	0.56964	0.43729	0.74324
logPdens	0.92215	0.91413	0.93023
propWhite	4.13212	3.16582	5.38206
propInd	0.45410	0.32165	0.64001
sd	0.31830	0.30419	0.33446
propSpatial	0.96016	0.91715	0.98591

Table1

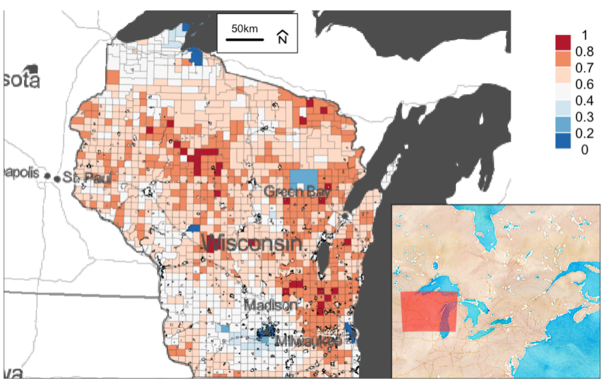

Figure1: Population of voting for Trump

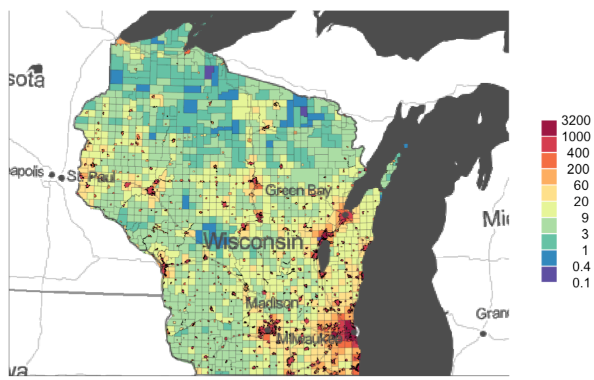

Figure2: Population density

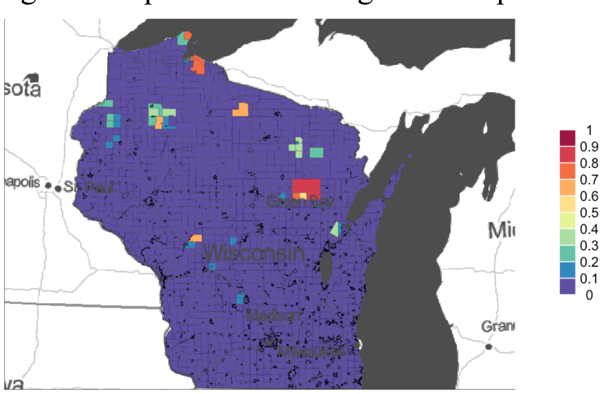

Figure3: Population of Indegenous

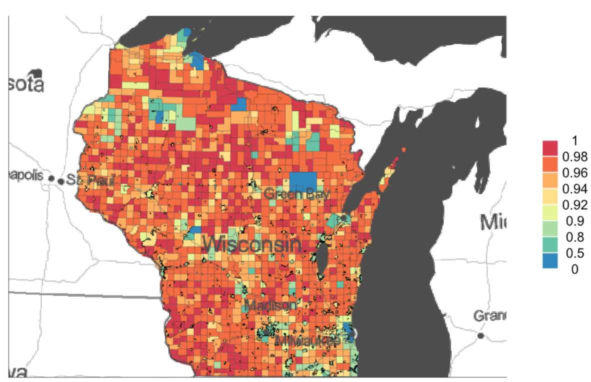

Figure4: Population of White

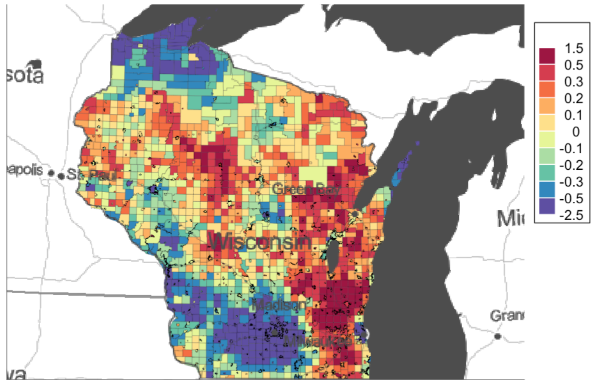

Figure5: Spatial Random Effect

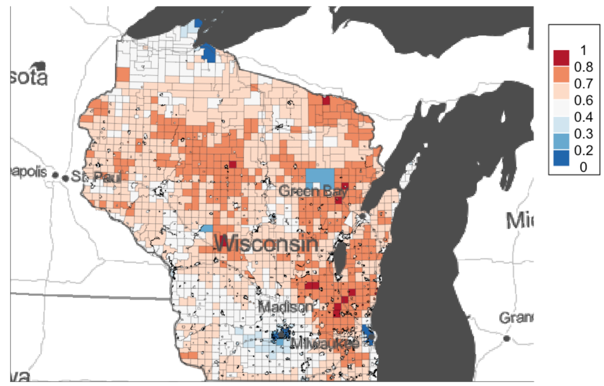

Figure6: Fixed Effect

In conclusion, there are some factors that would affect the voter turnout. The ratio of total population and surface area(square km) in each sub-country, population of white and indegenous would have a significant impact on voting for Trump in 2016. This may also be proved throughout Figure 1-6. Besides, from the diagrams, they demonstrate that white people living in rural areas will be the majority supporting Trump. And since the votes are not evenly distributed, there might be a spatial variation. Furthermore, there might be other factors that influenced the voters in Wisconsin such as the policy in each sub-country, which are not shown in this report.

Appendix

```r
```{r}
library(diseasemapping)
(load("/Users/yaoyao/Downloads/wisconsin.RData"))
resTrump = diseasemapping::bym(trump ~ logPdens + propWhite +
propInd, data = wisconsinCsubm, prior = list(sd = c(log(2.5),
0.5), propSpatial = c(0.5, 0.5)), Ntrials = wisconsinCsubm$Total,
family = "binomial")
```
```

```r
```{r}
library(gtsummary)
(load("/Users/yaoyao/Downloads/resWisconsin.RData"))
a = resTrump$parameters$summary[5:6, paste0(c(0.5,
0.025, 0.975), "quant")]

b = exp(resTrump$parameters$summary[1:4, paste0(c(0.5,
0.025, 0.975), "quant")])

knitr::kable(rbind(b,a), digits = 5)
```
```

```r
```{r}
library(mapmisc)
library(raster)

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)
```

```r
library("sp")

#propTrump
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)
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)
```