HW4

Nan Jiang

December 2020

1 Donald Duck

1.1 Introduction

This data set contains the sub-county, county and census tract election results in the 2016. The data is from the Harvard dataverse, which is from the Voting and Election Science Team form the University of Florida and Wichita Stata University. The data set is available at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/NH5S2I. In this report, we want to discuss the factors that influenced the voters in Wisconsin to vote for Trump in 2016. We want to find out the effect from race , area and urban or rural.

1.2 Method

We use a INLA method to create spatial a generalized linear model. The data follows binomial regression.

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

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

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

$$pr(\sigma > log(2.5)) = log(0.5)$$

$$pr(\tau > 0.5) = 0.5$$

The prior distributions have a prior median of log(2.5) forσ and 0.5 for the spatial proportion τ. X_1 = log of the ratios of total population, surface area (square km). X_2 = proportion of each region which is White. X_3 = proportion of each region which is indigenous. Since we know white people are mainly the supporter of Trump and the indigenous are mainly the opponents for Trump. They are two extremes. Such that we can conclude them as two important factor for the election and use other race as the base level.

1.3 Result

Posterior Quantiles

	0.5quant	0.025quant	0.975quant
(Intercept)	-0.56276	-0.82716	-0.29674
logPdens	-0.08105	-0.08979	-0.07232
propWhite	1.41879	1.15241	1.68307
propInd	-0.78943	-1.13430	-0.44628
sd	0.31830	0.30419	0.33446
propSpatial	0.96016	0.91715	0.98591

(a) estimated confidence interval

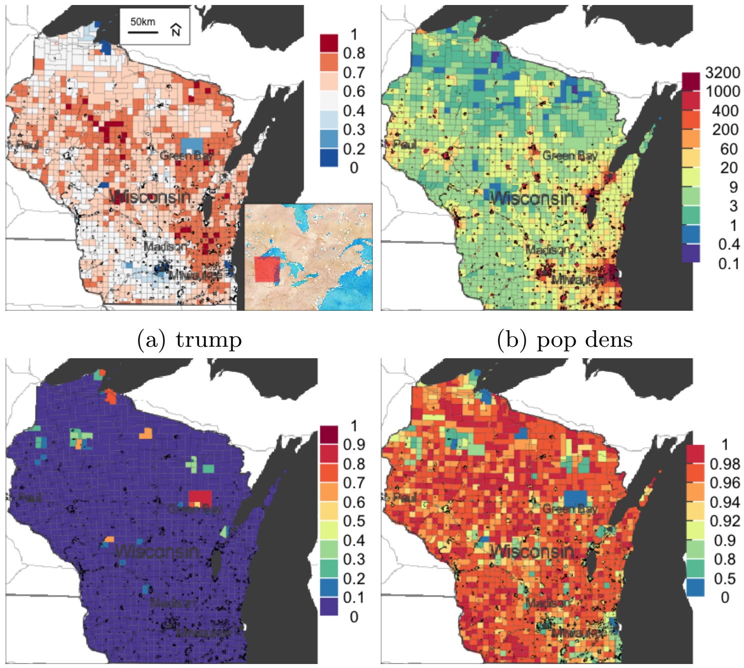
Posterior Quantiles

	0.5quant	0.025quant	0.975quant
(Intercept)	0.34948	0.21326	0.57444
logPdens	0.85949	0.84558	0.87362
propWhite	14.16116	8.60952	23.20154
propInd	0.22883	0.12015	0.43444
sd	1.81238	1.76521	1.86791
propSpatial	6.01179	5.54765	6.30797

(b) transformed estimated confidence interval

Figure 1: estimation

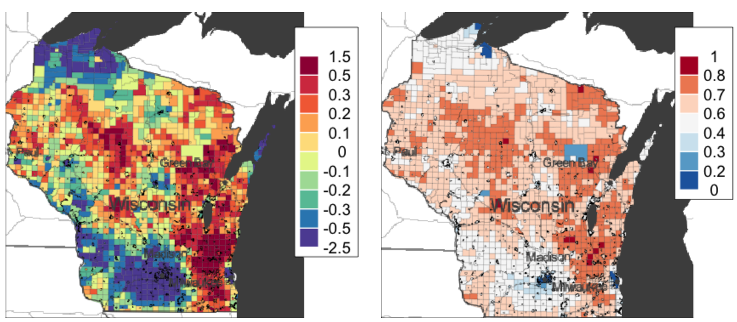
This is the estimated coefficients for the variables. And we made use the original value times the standard deviation of the log population densities, exponential it. From the original table, we can see the one unit change effect on the different log odds. In the second table, we can see the effect of one unit change of the standard deviation of log population density on the Trumps support rate. We can see the population density and being an indigenous has a negative impact on the support rate, and the race of white has an positive impact on the support rate.



(a) trump (b) pop dens

(c) ind (d) white

Figure 2: maps



(a) random (b) fitted

Figure 3: Results

looking at the graph we can see for white people, for most area have a higher support rate than 0.9. For the indigenous people, in most area, there is no support at all. Compare the support rate in high population density areas with in the low population density areas, like Milwaukee. Milwaukee has a high population density with high support rate. And in the upper area of Wisconsin, the population are low, but it also shows a high support rate. So we can conclude that the high population does not means high support rate. In fact the data shows the negative impact of the population density on the support rate. There exist a huge spatial variation.

1.4 Conclusion

From the estimated coefficients and the graph. We can make the conclusion. It it a racial phenomenon with Trump appealing to White voters. But also, there exist a huge spatial variation across Wisconsin. The rural area residence tend to have a higher probabilities to vote for trump. And for the Indigenous people, it is opposite to the white people. They usually do not vote for trump except only few people.

2 Question 2

2.1 Introduction

This data set contains the COVID-19 cases up to 15 October 2020, concentrations of fine particulate matter and the unemployment rate in different health regions in England without an island. The source is from the England_shp.RData file. There are two hypothesis. The first is, the heavy air pollution may increase infection rate of COVID-19. Since they puts stress on the lungs and respiratory tract. The second hypothesis is the area have high unemployment rate may have higher infection rate of COVID-19 due to the depravity, high population mobility and low care to the personal heath. And we want to study the factors for areas with many ethnic minorities tends to have high infection rate.

2.2 Method

we create a Besag, York and Mollié model on the data.

$$Y_i \sim Poisson(E_i\lambda_i)$$

$$log\lambda_i = \mu + X_i\beta + U_i$$

$$U_i \sim W_i + V_i$$

$$pr(\sigma > 0.5) = 0.5$$

$$pr(\tau > 0.5) = 0.5$$

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

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

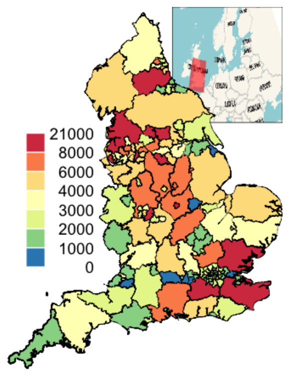
For log(expect) the expect value is number of COVID-19 cases up to 15 October 2020 and expected number, we have coefficient equal to 1. And X_1 is the ethnicity. X_2 is the concentrations of fine particulate matter (PM 2.5) in the health authority cases. X_3 is the percentage of the unemployment people.

2.3 Result

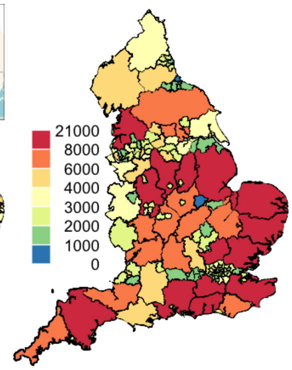
	mean	0.025quant	0.5quant	0.975quant
(Intercept)	-1.00752	-1.52329	-1.00723	-0.49380
Ethnicity	0.01205	0.00810	0.01205	0.01600
modelledpm25	0.05579	-0.00438	0.05575	0.11612
Unemployment	0.11321	0.05765	0.11321	0.16873
sd	0.29402	0.25873	0.29305	0.33554
propSpatial	0.89801	0.76785	0.90705	0.97548

(a) predicted random effects

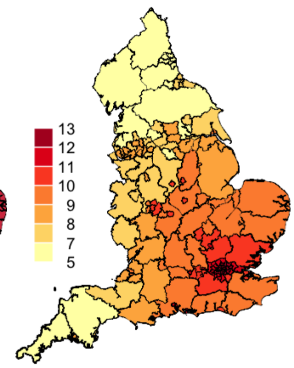
Above is the estimated coefficients from the model we built. For the concentration of PM2.5, the confidence interval contains 0. Such that we think the model is not significant. We can not predict the impact on the infection from the concentration of PM2.5 But others do. For ethnicity and unemployment rate, they both have a positive impact of 0.01205 and 0.11321 on the log probabilities.



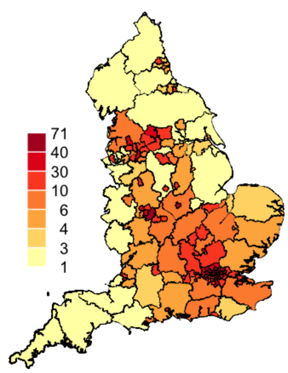
(a) number of COVID-19 cases



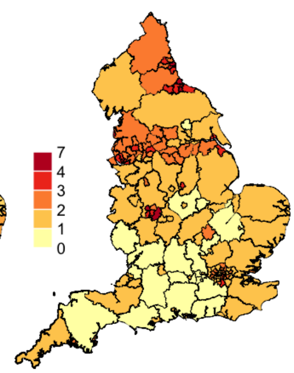
(b) Expected number of cases



(c) PM2.5 concentration

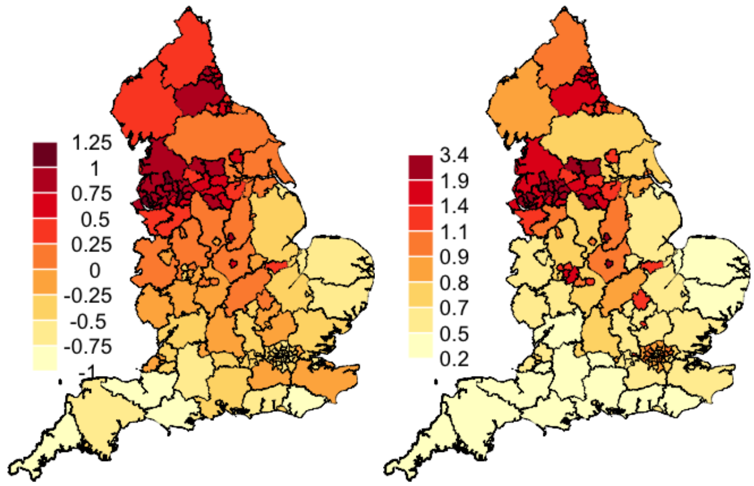


(d) Percentages of Ethnic Minorities



(e) Percentages of Unemployment

Figure 5: Overview of Sample Data



From the

(a) predicted random effects (b) predicted lambda (ratio of excess cases)

graph, it is more straightforward. The expected number of cased does not match the actual number of COVID-19 cases. And the area with high percentages of unemployment rate do show more infection cases.

Figure 6: Fitted Values from Model

2.4 Conclusion

Such that we made an conclusion. There exist spatial variation in this cases. Area with more ethnic minorities and more unemployed people tend have more infection cases. And the unemployment rate have high impact than the ethnic minorities.

3 Appendix

```
load("C:\\Users\\jn405\\Downloads\\wisconsin.RData")
load("C:\\Users\\jn405\\Downloads\\resWisconsin.RData")
load("C:\\Users\\jn405\\Downloads\\England_shp.RData")
load("C:\\Users\\jn405\\Downloads\\englandRes.RData")

knitr::kable(resTrump$parameters$summary[, paste0(c(0.5,0.025, 0.975), "quant")], digits = 5,
caption = "Posterior Quantiles")
apply(wisconsinCsubm@data[,c("logPdens","propWhite")],2,sd)
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$plot, add = TRUE,
lwd = 0.2)
raster::plot(theBg, add = TRUE, maxpixels = 10^7)
mapmisc::insetMap(wisconsinCsubm, "bottomright", theInset,
outer = TRUE, width = 0.35)
mapmisc::scaleBar(wisconsinCsubm, "top", cex = 0.8)
mapmisc::legendBreaks("topright", theColTrump, bty = "n",
inset = 0)
mapmisc::map.new(wisconsinCsubm, 0.85)
plot(wisconsinCsubm, col = theColPop$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)
knitr::kable(englandRes$parameters$summary[, c(1, 3:5)],digits = 5)
casesCol = mapmisc::colourScale(UK2$cases, dec = -3, breaks = 12,
col = "Spectral", style = "quantile", rev = TRUE)
Ecol = mapmisc::colourScale(UK2$E, breaks = casesCol$breaks,
col = casesCol$col, style = "fixed")
pmCol = mapmisc::colourScale(UK2$modelledpm25, breaks = 9,
dec = 0, style = "quantile")
ethCol = mapmisc::colourScale(UK2$Ethnicity, breaks = 9,
digits = 1, style = "quantile")
uCol = mapmisc::colourScale(UK2$Unemployment, breaks = 12,
dec = 0, style = "quantile")
rCol = mapmisc::colourScale(englandRes$data$random.mean,
breaks = 12, dec = -log10(0.25), style = "quantile")
fCol = mapmisc::colourScale(englandRes$data$fitted.exp,
breaks = 9, dec = 1, style = "quantile")
insetEngland1 = mapmisc::openmap(UK2, zoom = 3, fact = 4,
path = "waze", crs = CRS("+init=epsg:3035"))
library("raster")
insetEngland = raster::crop(insetEngland1, extend(extent(insetEngland1),
-c(25, 7, 4, 9.5) * 100 * 1000))
library("sp")
mapmisc::map.new(UK2)
mapmisc::insetMap(UK_shp, "topright", insetEngland, width = 0.4)
plot(UK2, col = casesCol$plot, add = TRUE, lwd = 0.2)
mapmisc::legendBreaks("left", casesCol, bty = "n")
mapmisc::map.new(UK2)
plot(UK2, col = Ecol$plot, add = TRUE, lwd = 0.2)
mapmisc::legendBreaks("left", casesCol, bty = "n")
mapmisc::map.new(UK2)
plot(UK2, col = pmCol$plot, add = TRUE, lwd = 0.2)
mapmisc::legendBreaks("left", pmCol, bty = "n")
mapmisc::map.new(UK2)
plot(UK2, col = ethCol$plot, add = TRUE, lwd = 0.2)
mapmisc::legendBreaks("left", ethCol, bty = "n")
mapmisc::map.new(UK2)
plot(UK2, col = uCol$plot, add = TRUE, lwd = 0.2)
mapmisc::legendBreaks("left", uCol, bty = "n")
mapmisc::map.new(UK2)
plot(UK2, col = rCol$plot, add = TRUE, lwd = 0.2)
mapmisc::legendBreaks("left", rCol, bty = "n")
mapmisc::map.new(UK2)
plot(UK2, col = fCol$plot, add = TRUE, lwd = 0.2)
mapmisc::legendBreaks("left", fCol, bty = "n")
```