1. 2016 United States presidential election: Who voted for Donald Trump in Wisconsin?

Summary

We used data about the result of Wisconsin voters in the 2016 U.S. presidential election collected provided by Professor Patrick Brown to determine whether Trump supporters have something in common. We found that there is a large spatial variation for the distribution of Trump voters. Specifically, Trump had a small ratio of voters in all cities, but his support rate is high in many rural areas. It is obvious that the indigenous people do not like Trump. However, we can not certainly conclude that most White supported Trump. We can conclude that the areas with high Trump votes are basically the areas with a high percentage of white people.

Introduction

We used data provided by Professor Patrick Brown which is about the basic information, such as the living place and the race, of the voters in Wisconsin for 2016 United States presidential election to analyze the election result, which is who voted Trump, in specific. We wanted to figure out whether there exists a spatial pattern in Trump support.

Method

Since we do not have too much information about the data, we try to build a model based on the given parameters. Since the parameter $trump$ is a yes or no question, which provides the number of votes for Trump in regions, we used a logistic regression model to explore the factors influencing the voters in Wisconsin. We would like to use $propWhite$ and $propInd$ which are the proportion of white and indigenous people, because it can check if there are racial differences among Trump supporters. We preferred to use $logPdens$ rather than $pdens$ since our model is logistic. Since our covariates are based on the location in two dimension, there is a spatial random effect. Thus, We could apply the Besag, York and Mollie model (BYM) model for our analysis. Thus, we decided to use the provided model, which is

$$Y_i \sim Binomial(N_i, \rho_i)$$
$$log(\frac{\rho_i}{1 - \rho_i}) = \mu + X_i\boldsymbol{\beta} + U_i$$
$$U_i \sim \text{BYM}(\sigma^2, \phi)$$

with priors

$$\sigma \sim PC.param(\log(2.5), 0.5) \text{ where } Pr(\sigma > \log(2.5)) = 0.5$$
$$\phi \sim PC.param(0.5, 0.5) \text{ where } Pr(\phi > 0.5) = 0.5$$

- Y_i indicates the number of Trump voters in the ith region in Wisconsin;
- N_i represents the number of voters in the ith region in Wisconsin; - ρ_i is the proportion of voters in the ith region that supports Trump;
- μ is the intercept;
- X_i represents a collection of personal information for voters in the ith region including the proportion of White, the proportion of Indigenous and the log ratio of the total population and the surface area;
- $\boldsymbol{\beta}$ are coefficients for all covariates X_i and the intercept, which revels the influences of each factor;

- U_i is the residual spatial variation for the ith region, the difference between actual number of Trump voters and what the covariates predict, which depends on

σ variability in residual variation

ϕ the spatial proportion

Result

	0.5quant	0.025quant	0.975quant
(Intercept)	0.56964	-0.82716	-0.29674
logPdens	0.92215	-0.08979	-0.07232
propWhite	4.13212	1.15241	1.68307
propInd	0.45410	-1.13430	-0.44628
sd	1.37479	0.30419	0.33446
propSpatial	2.61212	0.91715	0.98591

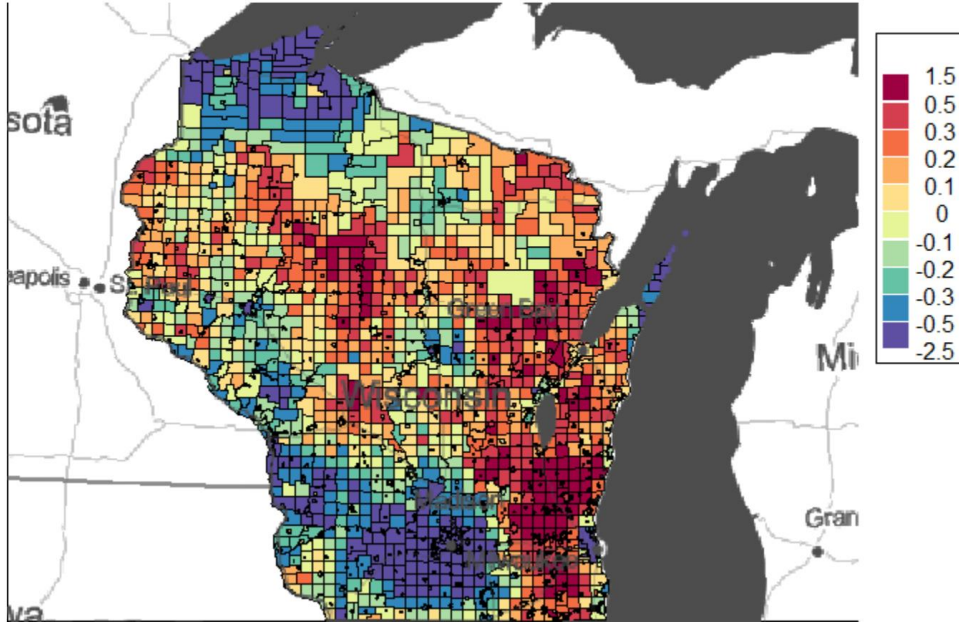Figure 1: The table of odds of the resTrump model with confidence intervals.

Figure 2: The posterior means of the spatial random effect.

Figure 1 shows the exponential coefficients with its confidence interval of our model. We can easily do comparison by computing the odds ratio between different covariates. Figure 2 and 3 show the results from fitting the BYM model to the Trump data, where the first one represents the spatial random effect $E[U_i \mid Y]$ and the second one represents fitted effect $E[\rho_i \mid Y]$. According to plots, we can clearly
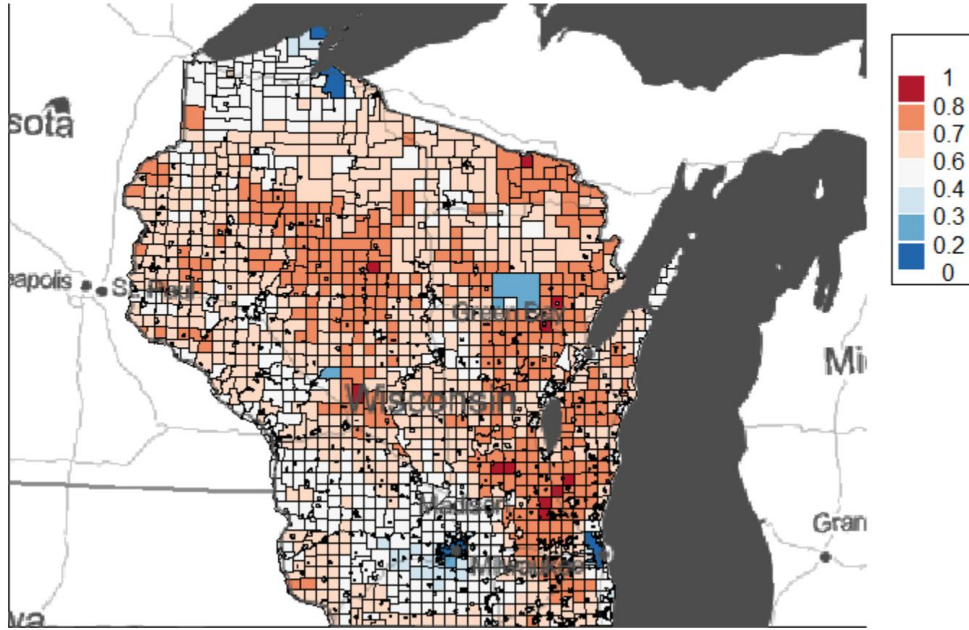
Figure 3: The predicted probability.

see the colors in different regions are differ, which reveals that there exists a large spatial variation. We can get the same result by checking the coefficient about the spatial effect in Figure 1 as the value and its confidence interval are high. Thus, we want to see the spatial plot for each covariate.
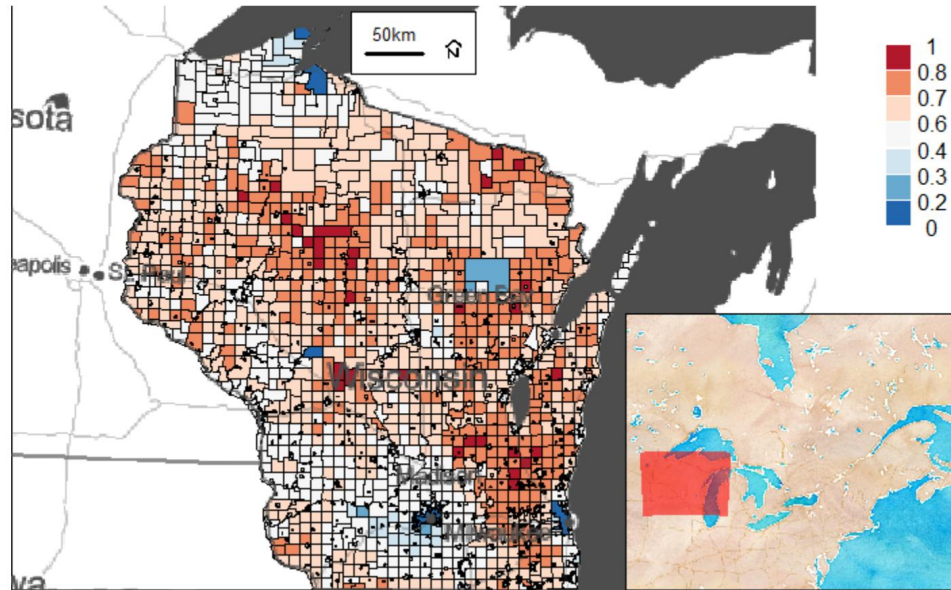

Figure 4: The proportion of Trump voters.

Figure 4 shows the spatial plot about the proportion of Trump voters. We used it as the reference. According to Figure 5, which gives information about the population density, we can clearly see that the greater the population density in the area it is, the lower Trump's support rate would be. Corresponding to its coefficient, since both the coefficient and its confidence interval are negative, it supports our conclusion and we can also get that as the log ratio of the population density increase 1 unit, the vote
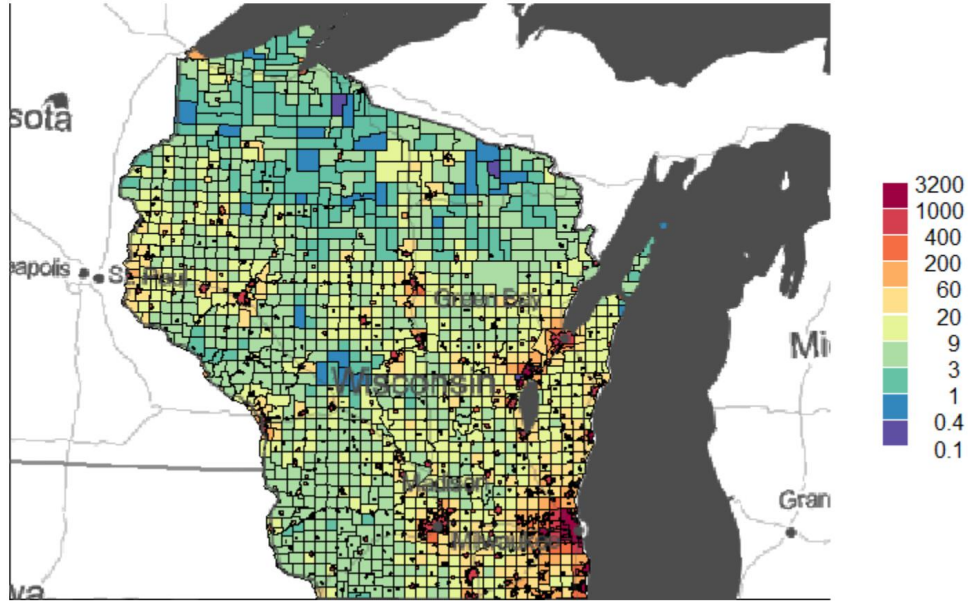
Figure 5: The proportion of the population density.

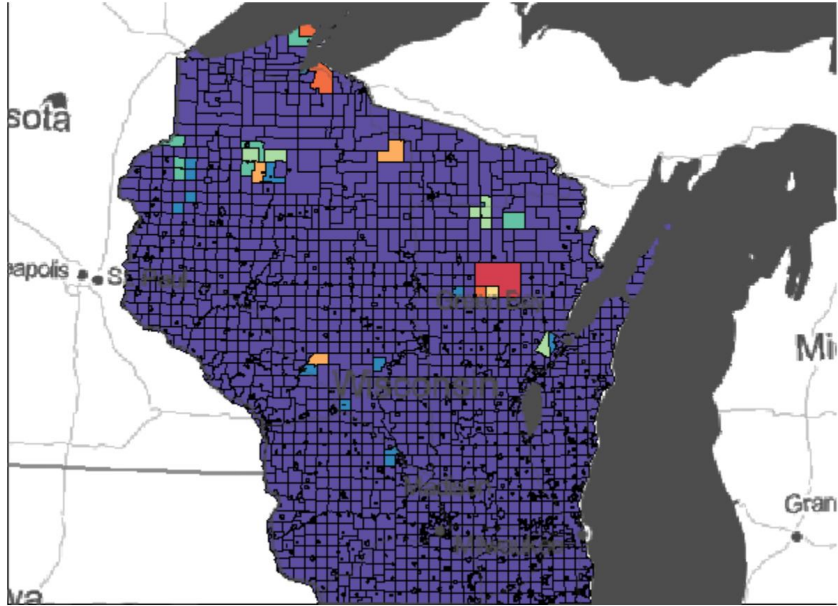rate for Trump would decreased about 13.6% ($\frac{-0.56276}{-0.56276-0.08105}$).

Figure 6: The proportion of indigenous people.

Figure 6 and 7 illustrate the proportion of the indigenous and white people respectively. It is clear that indigenous people do not like Trump as the red area in Figure 6 is exactly the blue area in Figure 4. According to Figure 1, we found that both the coefficient and its confidence interval are negative. For the White people, we can see that Trump's support rate in areas with sparse whites are not high by Figure 7. It is difficult to identify the explicit difference for areas with majority whites. However, we found that as the proportion of white people increase 1 unit, the vote rate for Trump would increase about 52.1% ($\frac{1.41879-0.56276}{-0.56276}$). Thus, we can conclude that the areas with high Trump
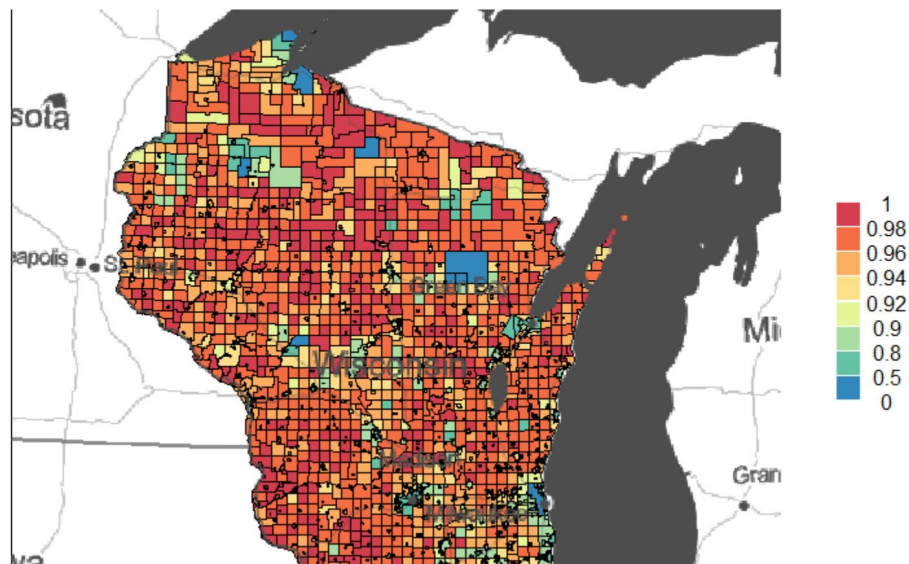
4

Figure 7: The proportion of white people.

votes are basically the areas with a high percentage of white people.

Appendix

We used R to built the models, which can be accessed in Appendix 1.

2. COVID-19 in England: Who is more likely to get COVID-19 in England?

Summary

We used the data about the number of COVID-19 cases in England to analyze whether people's living environment, working condition, and ethnicity influenced the probability of getting risk. We found that there is a large spatial variation for the distribution of the case counts. There is no evidence to conclude that high pollution places have more people to get COVID-19. However, we can certainly conclude that places with high unemployment rate usually have more cases. Although some evidence shows that the pandemic is more serious in areas with many ethnic minorities, the impact is not significant.

Introduction

We used the data about the number of COVID-19 cases up to October 15, 2020 as well as the corresponding information in regions about the air pollution, the unemployment rate, and the percentage of ethnic minorities in England provided by Professor Patrick Brown to analyze whether those factors could affect individuals to get COVID-19. In other words, we wanted to figure out whether there exists a spatial pattern for the COIVD-19 cases.

Method

Since we do not have too much information about the data, we built a model based on the given parameters. Since the parameter *cases* represent the count number, which provides the number of COVID-19 cases in regions, we used a logistic regression model with Poisson distribution to explore the factors influencing the case number in England. We set the *logExpected* as the offset term for our model since we want to check whether the actual number of cases fits our expectation. We would like to use *modelledpm*25 and *Ethnicity* which are concentrations of fine particulate matter and the proportion of ethnic minorities, because it can check if there are environmental or racial differences among individuals who got positive for the COVID-19. We also like to use *Unemployment* since we want to figure out whether the working condition is one of the factors. Since our covariates are based on the location in two dimension, there is a spatial random effect. Thus, we applied the BYM model for our analysis. Thus, we decided to use the provided model, which is

$$Y_i \sim \text{Poisson}(E_i \lambda_i)$$
$$\log(\lambda_i) = \mu + X_i \beta + U_i$$
$$U_i \sim \text{BYM}(\sigma^2, \tau^2)$$
$$\theta_1 = \sqrt{\sigma^2 + \tau^2}$$
$$\theta_2 = \frac{\sigma}{\sqrt{\sigma^2 + \tau^2}}$$

with priors

$$\theta_1 \sim PC.param(0.5, 0.5) \text{ where } Pr(\theta_1 > 0.5) = 0.5$$
$$\theta_2 \sim PC.param(0.5, 0.5) \text{ where } Pr(\theta_2 > 0.5) = 0.5$$

- Y_i indicates the case counts in the ith region;
- E_i represents the offset term, the expected case counts in the ith region;
- λ_i is the proportion of the number of COIVD-19 cases in the ith region;
- μ is the intercept;
- X_i represents a collection of personal information for individuals in the ith region including the

6

concentrations of fine particular matter, the unemployment rate and the percentage of ethnic minorities;
- β are coefficients for all covariates X_i, which revels the influences of each factor;
- U_i is the residual spatial variation for the ith region, the difference between actual number of COVID-19 cases and what the covariates predict, which depends on

σ variability in residual variation

τ the spatial proportion

- θ_1 is the marginal standard deviation;
- θ_2 is the ratio of the variability in residual variation and the marginal standard deviation.

Result

	mean	0.025quant	0.5quant	0.975quant
(Intercept)	0.365	-1.523	-1.007	-0.494
Ethnicity	1.012	0.008	0.012	0.016
modelledpm25	1.057	-0.004	0.056	0.116
Unemployment	1.120	0.058	0.113	0.169
sd	1.342	0.259	0.293	0.336
propSpatial	2.455	0.768	0.907	0.975

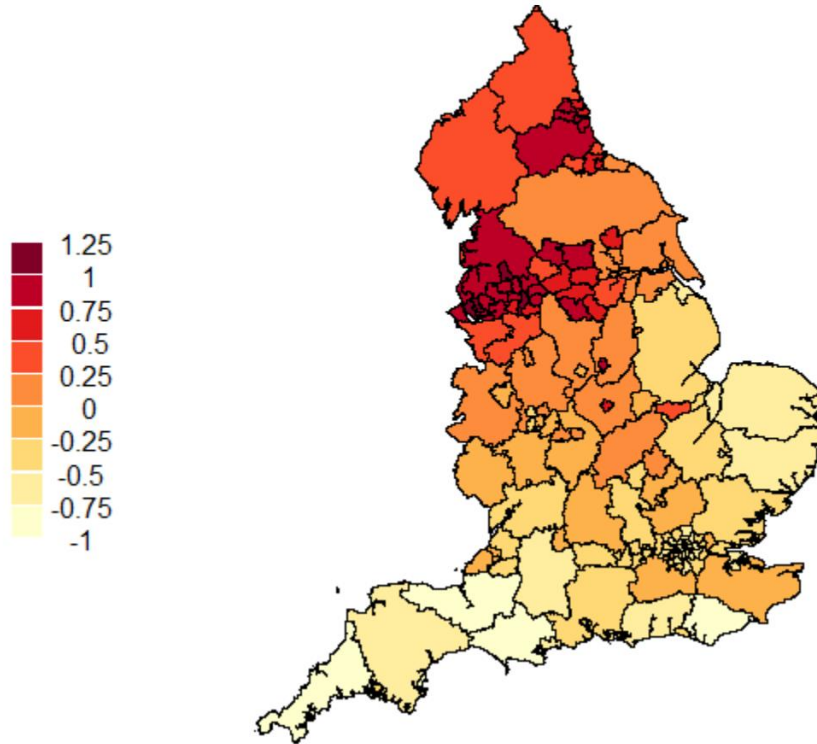Figure 8: The table of odds of the englandRes model with confidence intervals.

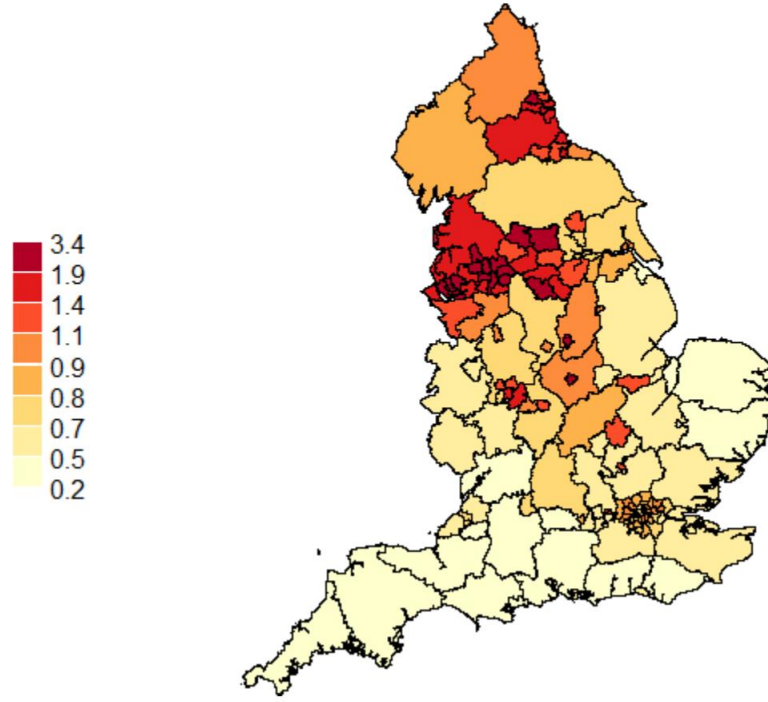Figure 9: The posterior means of the spatial random effect.
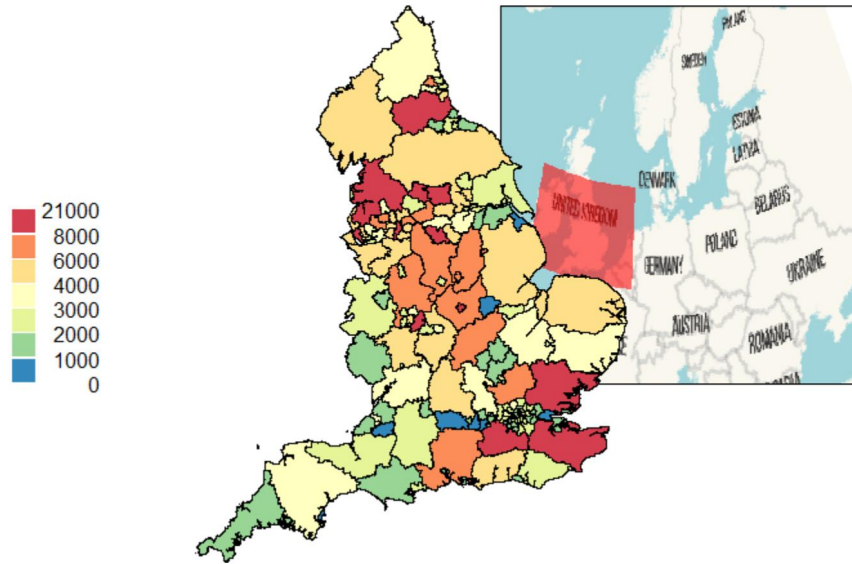
Figure 10: The predicted probability.

Figure 11: The actual number of COVID-19 cases.

Figure 8 shows the exponential coefficients with its confidence interval of our model. We can easily do comparison by computing the odds ratio between different covariates. Figure 9 and 10 show the results from fitting the BYM model to the COVID-19 data, where the first one represents the spatial random effect $E[U_i \mid Y]$ and the second one represents fitted effect $E[\frac{\lambda_i}{E_i} \mid Y]$. According to plots, we can clearly see the colors in different regions are differ, which reveals that there exists a large spatial variation. We can get the same result by checking the coefficient about the spatial effect in Figure 8 as the value and its confidence interval are high. Thus, we want to see the spatial plot for each covariate.
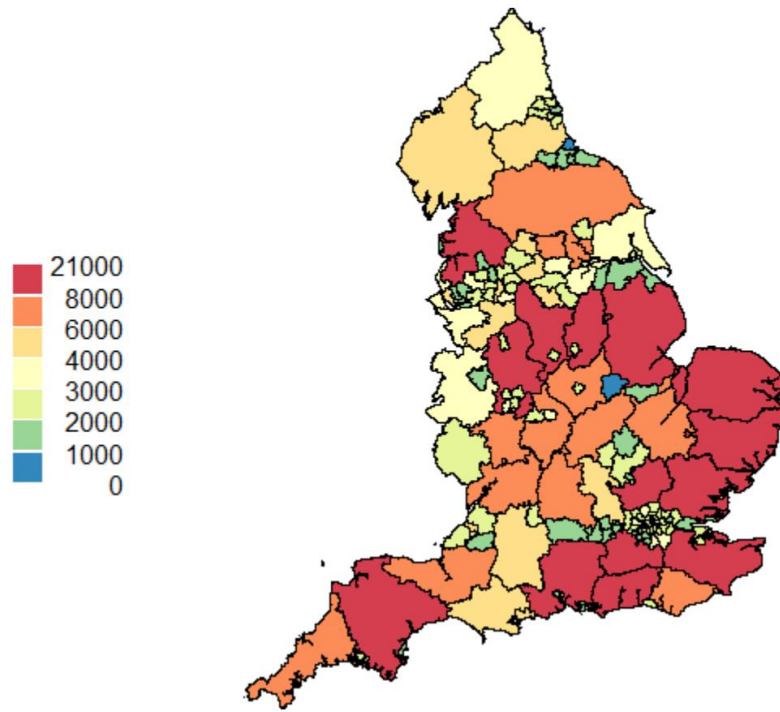
8

Figure 12: The expected number of COVID-19 cases.
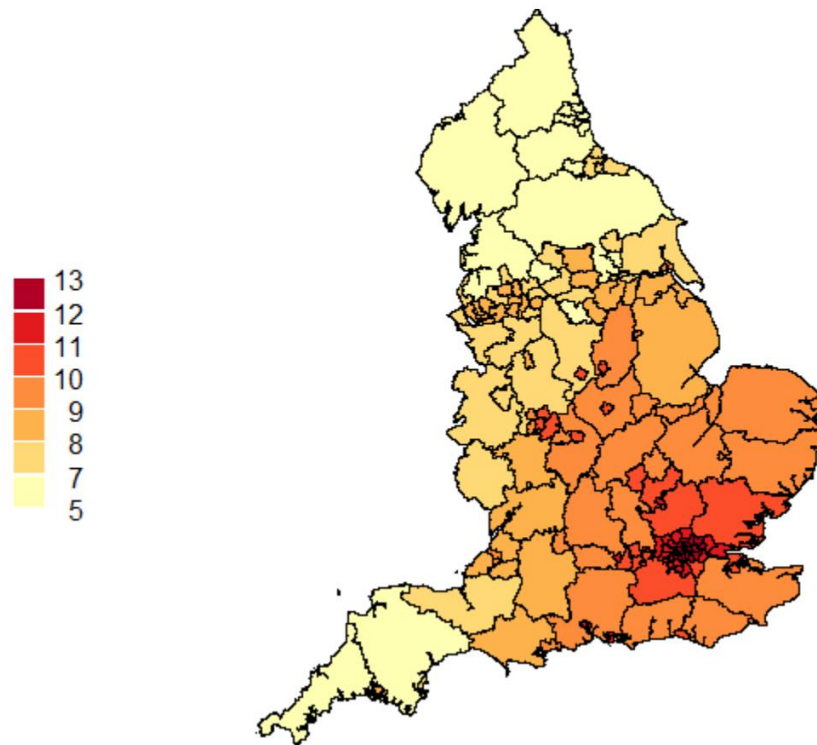

Figure 13: The concentrations of fine particulate matter.
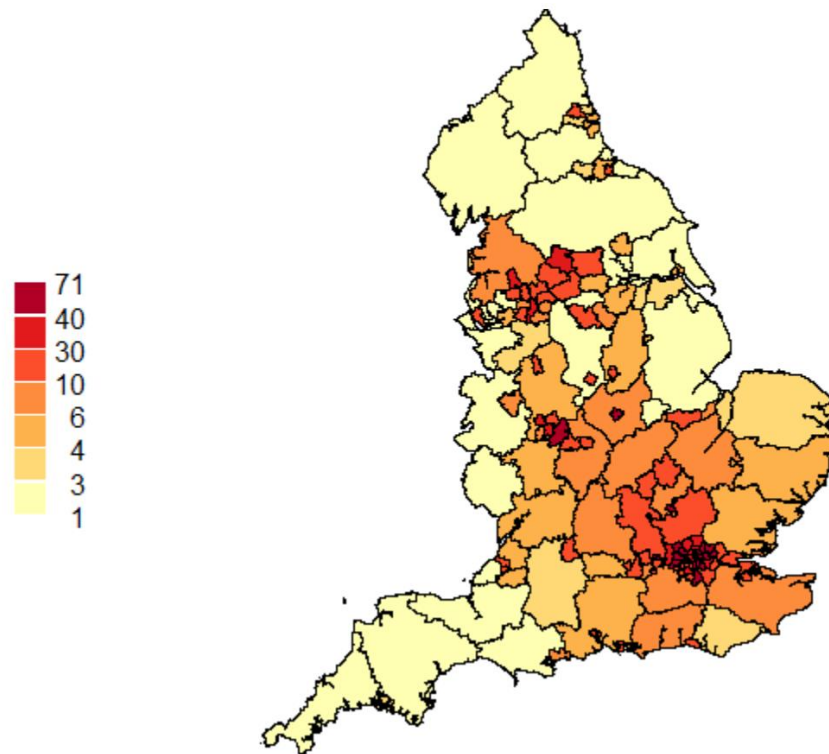
Figure 14: The proportion of individuals who are ethnic minorities.
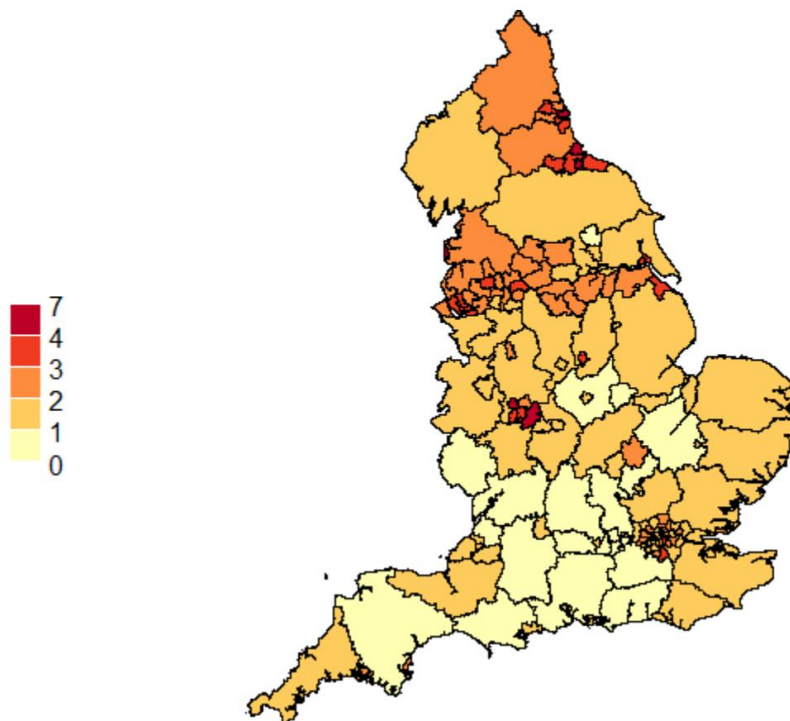

Figure 15: The proportion of individuals who are unemployed.

Figure 11 and 12 show the spatial plot about the actual number and the expected number of COVID-19 cases. We used them as the reference. According to Figure 13, which gives information about the air pollution, we can not see an explicit pattern since the number of cases in the most polluted places is much lower than that in other areas. Corresponding to its coefficient, although the coefficient ,1.057, is positive, the confidence interval includes 0, which reveals that it is not significant. Thus, it supports our conclusion that we cannot certain that air pollution put stress on the lungs and resulted in large number of cases.

Figure 14 shows the spatial plot about the proportion of ethnic minorities in each region. Similarly, we can not see an explicit pattern since the number of cases in the place that has the most diversity communities is much lower than that in other areas. For its coefficient, it is 1.012 with the confidence interval above 0, which reveals that it has effects on the ratio of the actual number and the expected number of cases. As the proportion of ethnic minorities increase 1 unit, the ratio would increase about 3.7 times more ($\frac{0.365+1.012}{0.365}$), which means that more real cases as the expected number is stable. Thus, we can conclude that the race might be one of the factors influencing the health.

Figure 15 shows the spatial plot about the unemployment rate. Except for the bottom right region, it seems like the place that has high unemployment rate is more likely to get high risk for the COVID-19. According to Figure 8, as the proportion of unemployed people increase 1 unit, the ratio would increase about 4 times more ($\frac{0.365+1.120}{0.365}$), which means that more real cases as the expected number is stable. The positive lower bound of its confidence interval illustrates that we cannot ignore this factor. Thus, we can conclude that the unemployed people have higher potential to get risk of COVID-19.

Appendix

We used R to built the models, which can be accessed in Appendix 2.

Appendix

Appendix 1 R code for Question 1

```r
library("Pmisc")
library("INLA")
library("sp")
library("raster")

(load("./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")
# save(resTrump, file = "./resWisconsin.RData", compress = "xz")

(load("./resWisconsin.RData"))

parTableVote <- resTrump$parameters$summary[, paste0(c(0.5, 0.025, 0.975), "quant")]
parTableVote[1:4, 1] <- exp(parTableVote[1:4, 1])
knitr::kable(parTableVote, digits = 5)

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)

mapmisc::map.new(wisconsinCsubm, 0.85)
# trump
sp::plot(wisconsinCsubm, col = theColTrump$plot, add = TRUE, lwd = 0.2)
raster::plot(theBg, add = TRUE, maxpixels = 10^7)
mapmisc::insetMap(wisconsinCsubm, "bottomright", theInset, outer = TRUE, width = 0.35)
mapmisc::scaleBar(wisconsinCsubm, "top", cex = 0.8)
mapmisc::legendBreaks("topright", theColTrump, bty = "n", inset = 0)
mapmisc::map.new(wisconsinCsubm, 0.85)
# pdens
plot(wisconsinCsubm, col = theColPop$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)
# ind
plot(wisconsinCsubm, col = theColInd$plot, add = TRUE, lwd = 0.2)
plot(theBg, add = TRUE, maxpixels = 10^7)
mapmisc::legendBreaks("right", theColInd, bty = "n", inset = 0)
mapmisc::map.new(wisconsinCsubm, 0.85)
# white
plot(wisconsinCsubm, col = theColWhite$plot, add = TRUE, lwd = 0.2)
plot(theBg, add = TRUE, maxpixels = 10^7)
```

```
56  mapmisc::legendBreaks("right", theColWhite, bty = "n", inset = 0)
57
58  theColRandom = mapmisc::colourScale(resTrump$data$random.mean, col = "Spectral",
59                                      breaks = 11, style = "quantile", rev = TRUE, dec =
                                            1)
60  theColFit = mapmisc::colourScale(resTrump$data$fitted.invlogit, col = "RdBu", rev =
        TRUE,
61                                      breaks = sort(unique(setdiff(c(0, 1, seq(0.2, 0.8, by
                                              = 0.1)),
62                                                                      0.5))), style = "fixed")
63
64  mapmisc::map.new(wisconsinCsubm, 0.85)
65  # random
66  plot(resTrump$data, col = theColRandom$plot, add = TRUE, lwd = 0.2)
67  plot(theBg, add = TRUE, maxpixels = 10^7)
68  mapmisc::legendBreaks("topright", theColRandom)
69  mapmisc::map.new(wisconsinCsubm, 0.85)
70  # fit
71  plot(resTrump$data, col = theColFit$plot, add = TRUE, lwd = 0.2)
72  plot(theBg, add = TRUE, maxpixels = 10^7)
73  mapmisc::legendBreaks("topright", theColFit)
```

Appendix 2 R code for Question 2

```
1   (load("./England_shp.RData"))
2
3   UK_shp$logExpected = log(UK_shp$E)
4   # remove an island
5   UK2 = UK_shp[grep("Wight", UK_shp$Name, invert = TRUE), ]
6   # englandRes = diseasemapping::bym(cases ~ offset(logExpected) + Ethnicity +
7   #                                    modelledpm25 + Unemployment,
8   #                                    prior = list(sd = c(0.5, 0.5), propSpatial = c(0.5,
        0.5)),
9   #                                    family = "poisson", data = UK2)
10  # save(englandRes, file = "./englandRes.RData")
11
12  (load("./englandRes.RData"))
13
14  parTableCOVID <- englandRes$parameters$summary[, c(1, 3:5)]
15  parTableCOVID[1:4, 1] <- exp(parTableCOVID[1:4, 1])
16  knitr::kable(parTableCOVID, digits=3)
17
18  casesCol = mapmisc::colourScale(UK2$cases, dec = -3, breaks = 12, col = "Spectral",
19                                      style = "quantile", rev = TRUE)
20  Ecol = mapmisc::colourScale(UK2$E, breaks = casesCol$breaks, col = casesCol$col, style
        = "fixed")
21  pmCol = mapmisc::colourScale(UK2$modelledpm25, breaks = 9, dec = 0, style = "quantile")
22  ethCol = mapmisc::colourScale(UK2$Ethnicity, breaks = 9, digits = 1, style = "quantile"
        )
23  uCol = mapmisc::colourScale(UK2$Unemployment, breaks = 12, dec = 0, style = "quantile")
24  rCol = mapmisc::colourScale(englandRes$data$random.mean, breaks = 12, dec = -log10
        (0.25),
25                                      style = "quantile")
26  fCol = mapmisc::colourScale(englandRes$data$fitted.exp, breaks = 9, dec = 1, style = "
        quantile")
27  insetEngland1 = mapmisc::openmap(UK2, zoom = 3, fact = 4, path = "waze",
28                                      crs = CRS("+init=epsg:3035"))
29  insetEngland = raster::crop(insetEngland1, extend(extent(insetEngland1),
30                                                      -c(25, 7, 4, 9.5) * 100 * 1000))
31
32  mapmisc::map.new(UK2)
33  mapmisc::insetMap(UK_shp, "topright", insetEngland, width = 0.4)
34  plot(UK2, col = casesCol$plot, add = TRUE, lwd = 0.2)
35  mapmisc::legendBreaks("left", casesCol, bty = "n")
36  mapmisc::map.new(UK2)
37  plot(UK2, col = Ecol$plot, add = TRUE, lwd = 0.2)
```

```r
38  mapmisc::legendBreaks("left", casesCol, bty = "n")
39  mapmisc::map.new(UK2)
40  plot(UK2, col = pmCol$plot, add = TRUE, lwd = 0.2)
41  mapmisc::legendBreaks("left", pmCol, bty = "n")
42  mapmisc::map.new(UK2)
43  plot(UK2, col = ethCol$plot, add = TRUE, lwd = 0.2)
44  mapmisc::legendBreaks("left", ethCol, bty = "n")
45  mapmisc::map.new(UK2)
46  plot(UK2, col = uCol$plot, add = TRUE, lwd = 0.2)
47  mapmisc::legendBreaks("left", uCol, bty = "n")
48  mapmisc::map.new(UK2)
49  plot(UK2, col = rCol$plot, add = TRUE, lwd = 0.2)
50  mapmisc::legendBreaks("left", rCol, bty = "n")
51  mapmisc::map.new(UK2)
52  plot(UK2, col = fCol$plot, add = TRUE, lwd = 0.2)
53  mapmisc::legendBreaks("left", fCol, bty = "n")
```