

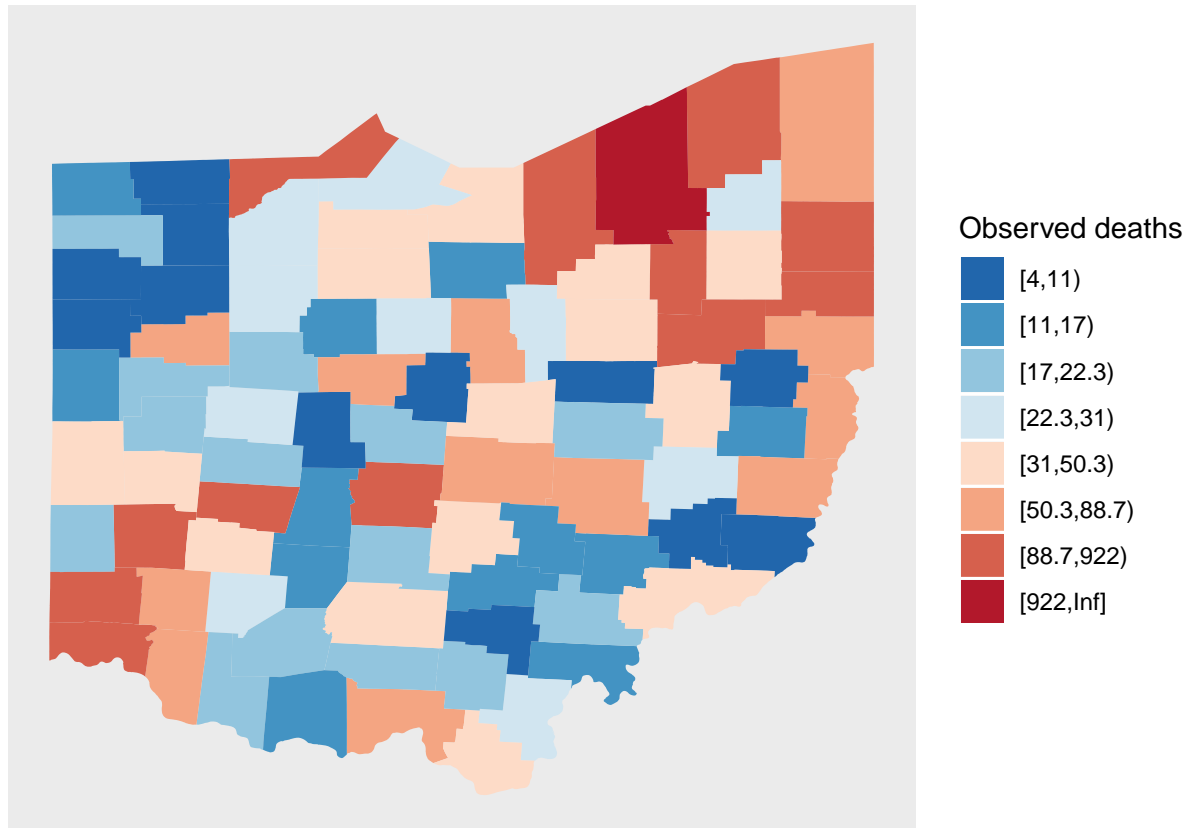
Spatial Epi HW 2

David Coomes

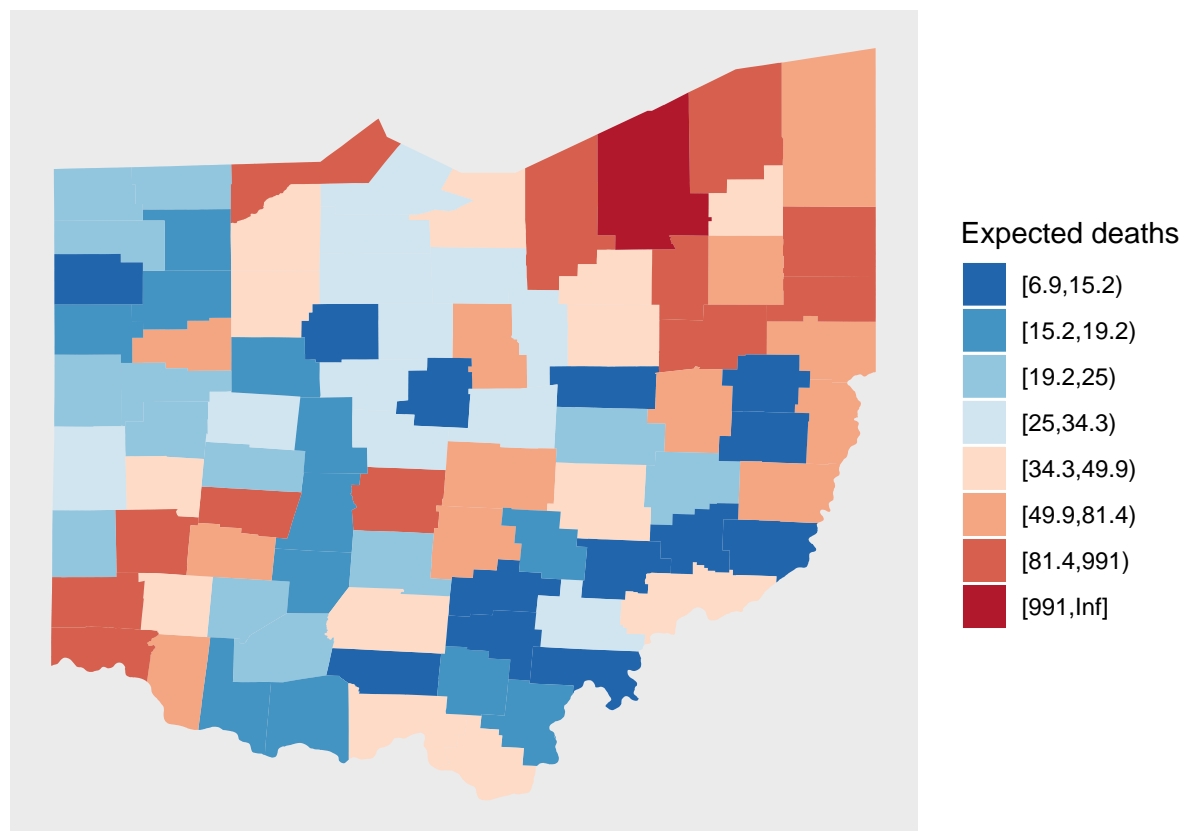
2/5/2020

Question 1

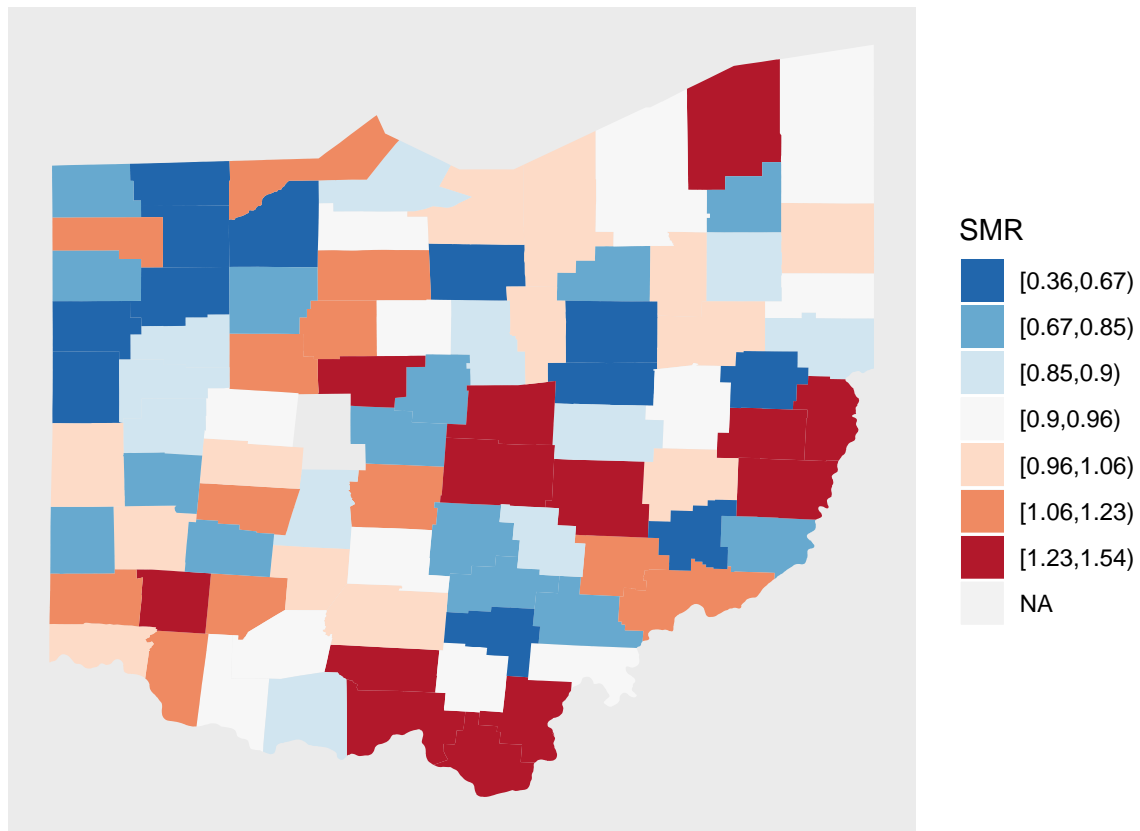
(a) Map of observed lung cancer mortality counts in Ohio (1988).



(b) Map of expected lung cancer mortality counts in Ohio (1988). Expected counts are adjusted for gender, race, and age using internal standardization.



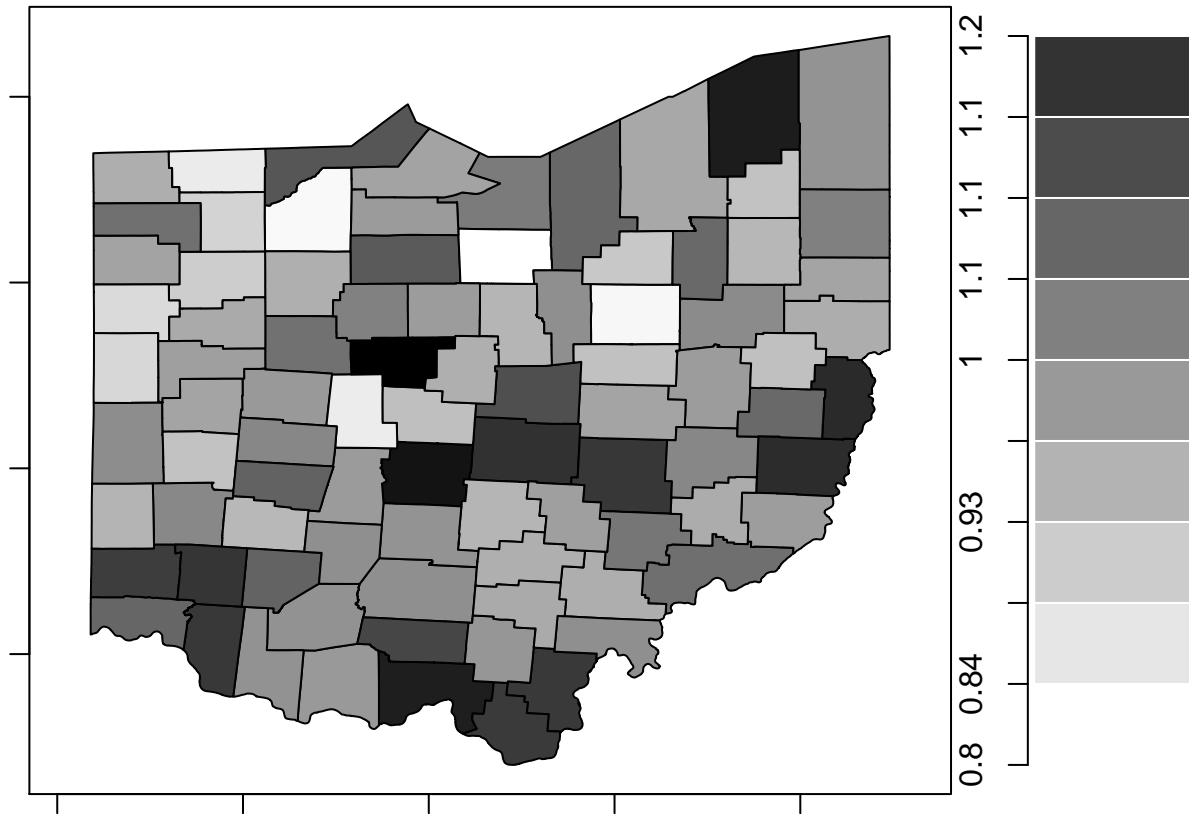
(c) Map of standardized mortality ratios (SMR) in Ohio (1988).



The SMRs seems somewhat normally distributed, with a mean and median slightly below 1 (0.94 and 0.93 respectively). The minimum SMR is 0.36 and the maximum is 1.54.

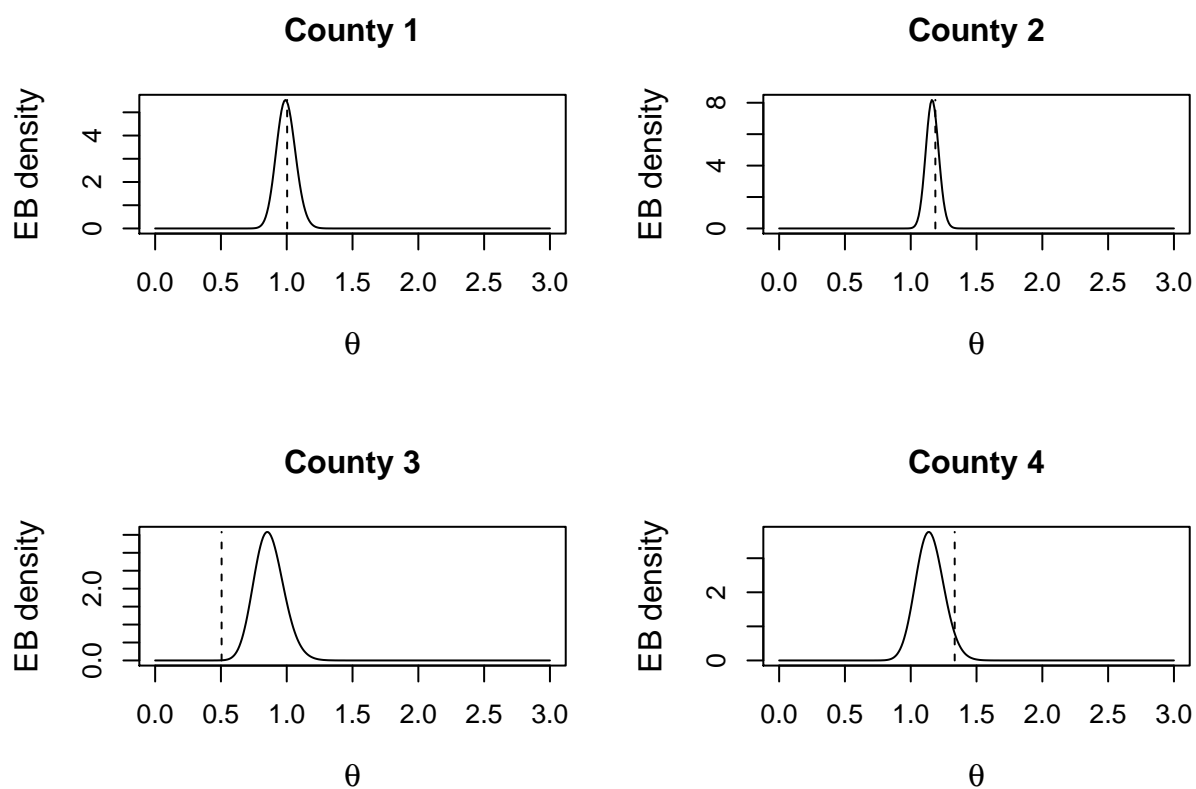
Question 2

(a) Map of posterior means of RR for lung cancer mortality in Ohio (1988).

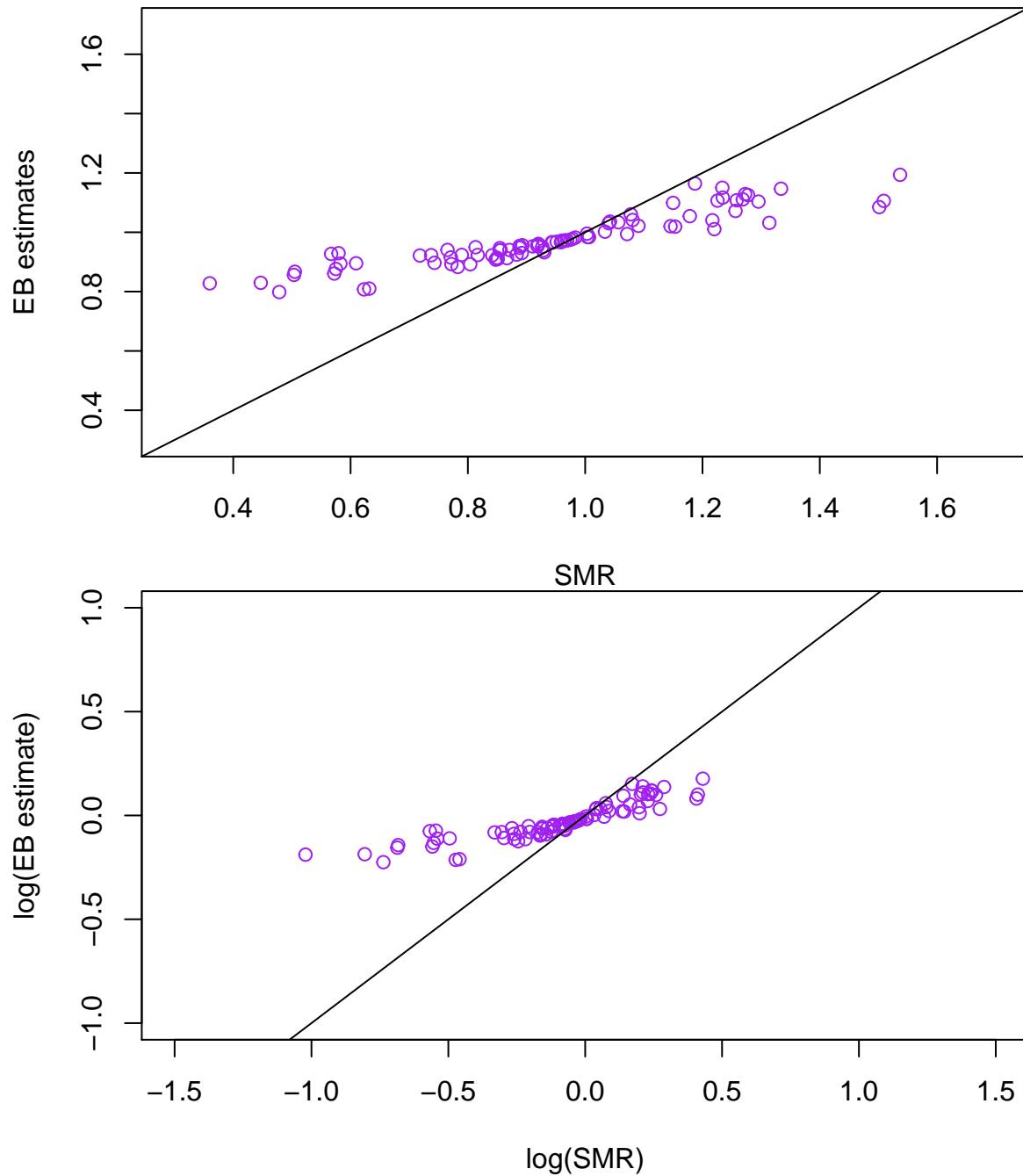


The distribution of the posterior means is similar to that of the non-smoothed SMR estimates in that they are generally normally distributed, and the SMRs correspond to the posterior means (those counties with high SMR tend to have high posterior RR and the same with low SMR). However, the distribution of the posterior RR estimates is substantially smaller than that of the SMRs. The minimum RR is 0.798 as compared to 0.36 for the SMR, and the maximum RR is 1.194 as compared to 1.54 for the SMR estimates. This shift of estimates towards the mean is what we would expect in our spatially smoothed model.

(b) Gamma posterior densities of the relative risks for the first four counties in Ohio.

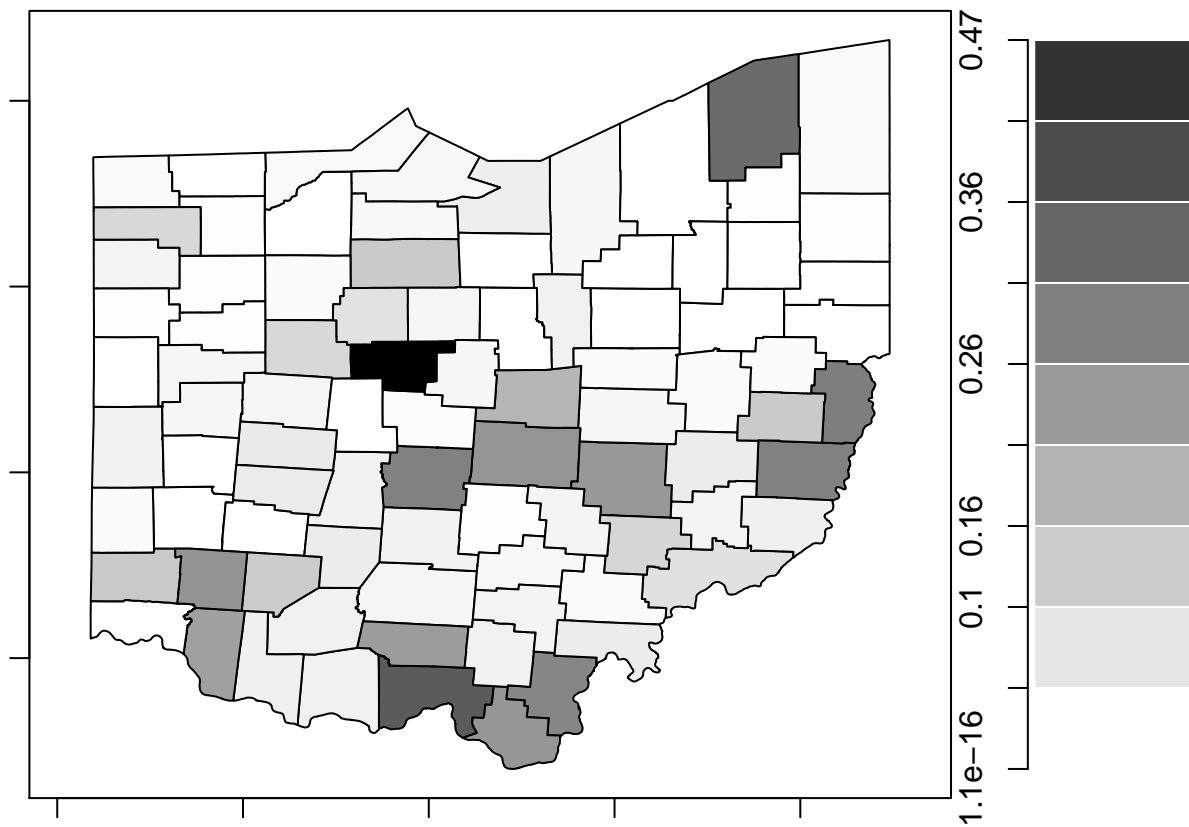


Plot of empirical Bayes estimates and SMRs.



We can see that the plot of the empirical Bayes estimates against the SMRs are tilted slightly from the abline that has a slope of 1 on both the normal and the log scale. This is to be expected as our smoothing model tends to push those values that are higher or lower than the average towards the average, especially if they have a relatively low sample size. Because of this, counties with a low estimated SMR will have a higher EB estimate, and those with a high estimated SMR will have a lower EB estimate.

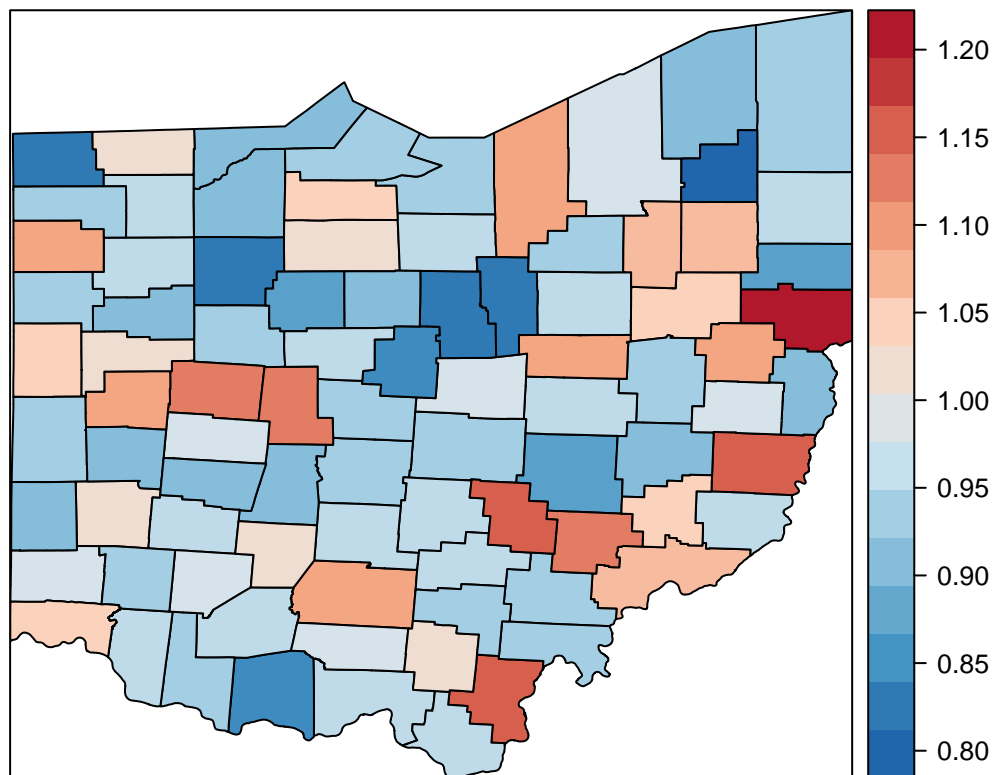
(c) Posterior probabilities that an RR exceeds 1.2 in each county.

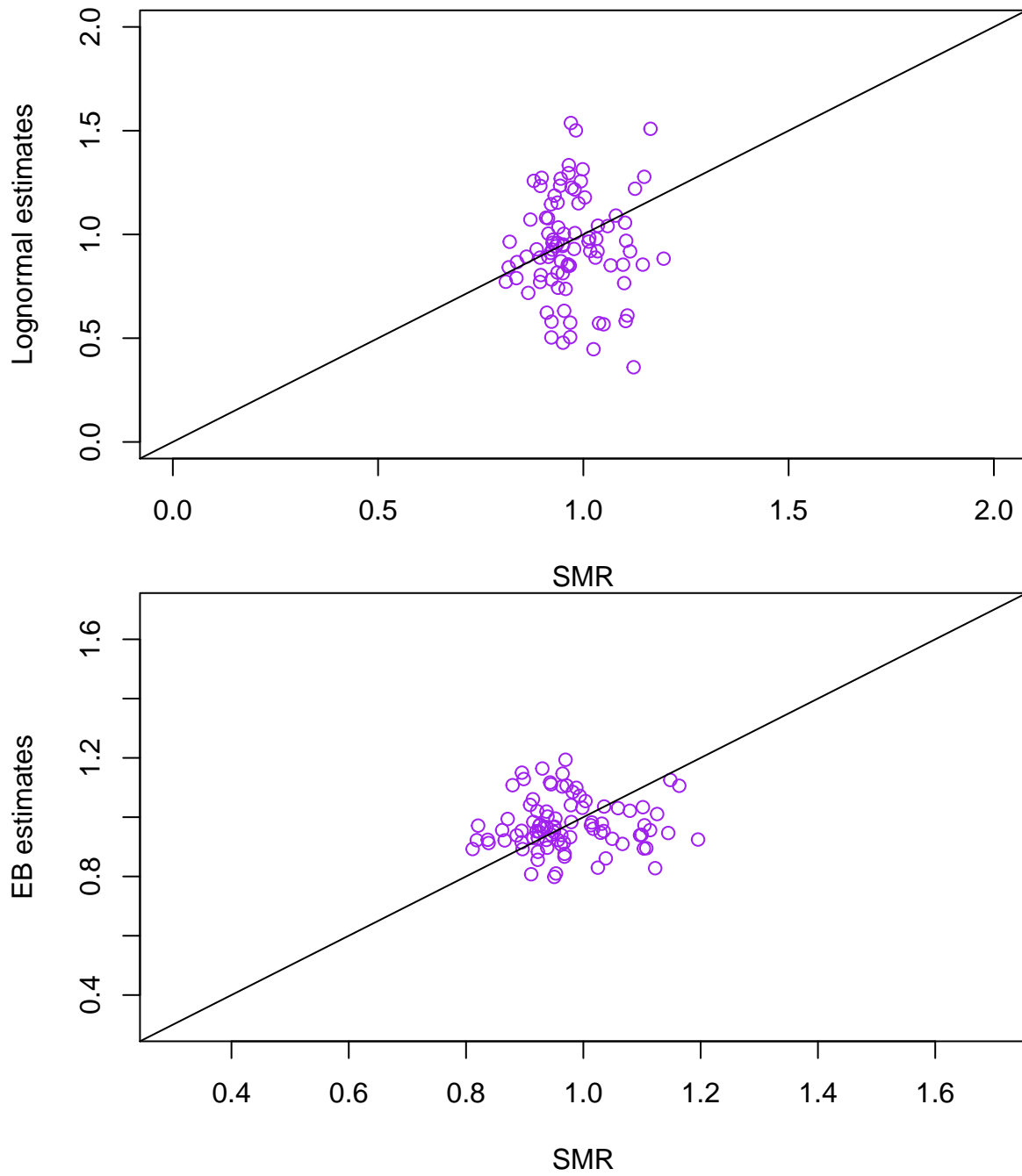


There are no counties where the probability of exceeding a RR of 1.2 is over 50%. There is one county that is close to 50% probability (right in the middle of the state), and several more counties that have around a 25-35% of exceeding a RR of 1.2.

Question 3

- (a) The posterior mean for β_0 is -0.033 and for σ_e it is 0.139.
- (b) RR of lung cancer mortality in Ohio using Poisson Lognormal smoothing





The RRs from my Lognormal model do not align well with either the RRs from the Gaussian model or the estimated SMRs. This makes me think that my Lognormal model was not fit correctly - I wouldn't expect it to be this far off.

Appendix

```
knitr::opts_chunk$set(echo = FALSE, warning=FALSE)

rm(list=ls())
library(haven)
#install.packages("INLA", repos=c(getOption("repos"), INLA="https://inla.r-inla-download.org/R/stable")
library(INLA)
library(rgdal)
library(maptools)
library(sp)
library(spdep)
library(SpatialEpi)
library(RColorBrewer)
library(ggplot2)
library(maps)
library(broom)
library(raster)
library(leaflet)
library(dplyr)

link = "https://github.com/dmccoomes/Spatial_epi/raw/master/HW%202/ohio_2019_version.txt"
ohio_canc <- read.table(url(link), header=TRUE)

#data
#Obs= observed deaths
#Exp = expected deaths

zip_ohmap_SHP = "https://github.com/dmccoomes/Spatial_epi/raw/master/HW%202/Ohio%20shape%20files%20(2).."

library(utils)
temp=tempfile()
download.file(zip_ohmap_SHP, temp)
unzip(temp)

#DMC - showing which tempfiles are in the directory

(maps=list.files(pattern = 'shp'))

# notice the parameters use in the chunk!!

#ohMap <- readOGR("ohio_map.shp", stringsAsFactors=F)

#The above code doesn't seem to work - loading data manually.
```

```

#On pc
ohio_map <- readOGR(dsn="/Users/david/Desktop/Dave/school/Spatial Epi/HW 2/HW 2_data", layer = "ohio_map")

#on laptop
#ohio_map <- readOGR(dsn="C:/Users/dcoomes/Documents/GitHub/Spatial_epi/HW 2/Ohio shape files", layer="ohio_map")

#creating SMRs
ohio_canc$smr <- ohio_canc$Obs/ohio_canc$Exp

ohio_map$fips <- as.numeric(as.character(ohio_map$CNTYIDFP00))
ohio_map <- merge(ohio_map, ohio_canc, by="fips")

#set breakpoints for plotting
num_bins <- 8
quants <- round(quantile(ohio_map@data$Obs, seq(0,1,length.out=num_bins)),2)
ohio_map@data$obs_cut <- cut(ohio_map@data$Obs, breaks=c(quants, Inf), include.lowest=TRUE, right=FALSE)

ohio_df <- tidy(ohio_map)

ohio_map$polyID <- sapply(slot(ohio_map, "polygons"), function(x) slot(x, "ID"))
ohio_df <- merge(ohio_df, ohio_map, by.x="id", by.y="polyID")

observed_ct_map <- ggplot() +
  geom_polygon(
    data=ohio_df,
    aes(x=long, y=lat, group=group,
        fill = obs_cut)) +
  scale_fill_brewer("Observed deaths", palette="RdBu", direction=-1) +
  theme(line=element_blank(),
        axis.text=element_blank(),
        axis.title=element_blank()) +
  coord_equal()

#save the map
ggsave("ohio_map_obs_ggplot.pdf", width=8, height=8)

plot(observed_ct_map,
     border='blue',
     lwd=1)

#set breakpoints for plotting
num_bins <- 8
quants <- round(quantile(ohio_map@data$Exp, seq(0,1,length.out=num_bins)),2)
ohio_map@data$exp_cut <- cut(ohio_map@data$Exp, breaks=c(quants, Inf), include.lowest=TRUE, right=FALSE)

```

```

ohio_df <- tidy(ohio_map)

ohio_map$polyID <- sapply(slot(ohio_map, "polygons"), function(x) slot(x, "ID"))
ohio_df <- merge(ohio_df, ohio_map, by.x="id", by.y="polyID")

expected_ct_map <- ggplot() +
  geom_polygon(
    data=ohio_df,
    aes(x=long, y=lat, group=group,
        fill = exp_cut)) +
  scale_fill_brewer("Expected deaths", palette="RdBu", direction=-1) +
  theme(line=element_blank(),
        axis.text=element_blank(),
        axis.title=element_blank()) +
  coord_equal()

#save the map
ggsave("ohio_map_exp_ggplot.pdf", width=8, height=8)

plot(expected_ct_map)

#set breakpoints for plotting
num_bins <- 8
quants <- round(quantile(ohio_map@data$smr, seq(0,1,length.out=num_bins)),2)
ohio_map@data$smr_cut <- cut(ohio_map@data$smr, breaks=c(quants, Inf), include.lowest=TRUE, right=FALSE)

ohio_df <- tidy(ohio_map)

ohio_map$polyID <- sapply(slot(ohio_map, "polygons"), function(x) slot(x, "ID"))
ohio_df <- merge(ohio_df, ohio_map, by.x="id", by.y="polyID")

smr_ct_map <- ggplot() +
  geom_polygon(
    data=ohio_df,
    aes(x=long, y=lat, group=group,
        fill = smr_cut)) +
  scale_fill_brewer("SMR", palette="RdBu", direction=-1) +
  theme(line=element_blank(),
        axis.text=element_blank(),
        axis.title=element_blank()) +
  coord_equal()

#save the map
ggsave("ohio_map_smr_ggplot.pdf", width=8, height=8)

plot(smr_ct_map,
     border='blue',
     lwd=1)

```

```

summary(ohio_map$smr)

hist(ohio_map$smr, breaks=10)

ebresults <- eBayes(ohio_map$Obs, ohio_map$Exp)
ebresults$alpha
#alpha = 51.73209
exp(ebresults$beta)
#beta = 0.9749581

mean(ohio_map$Obs/ohio_map$Exp)
#mean is 0.9432857

summary(ebresults$RR)

hist(ebresults$RR, breaks=10)

par(mar = c(1,1,1,1))
mapvariable(ebresults$RR, ohio_map)

par(mfrow = c(2,2))
EBpostdens(ohio_map$Obs[1], ohio_map$Exp[1], ebresults$alpha, ebresults$beta,
            lower = 0, upper = 3, main = "County 1")
EBpostdens(ohio_map$Obs[2], ohio_map$Exp[2], ebresults$alpha, ebresults$beta,
            lower = 0, upper = 3, main = "County 2")
EBpostdens(ohio_map$Obs[3], ohio_map$Exp[3], ebresults$alpha, ebresults$beta,
            lower = 0, upper = 3, main = "County 3")
EBpostdens(ohio_map$Obs[4], ohio_map$Exp[4], ebresults$alpha, ebresults$beta,
            lower = 0, upper = 3, main = "County 4")

egamma <- qgamma(seq(0.5, length(ohio_map$Obs), 1)/length(ohio_map$Obs),
                 ebresults$alpha, ebresults$alpha)
par(mfrow = c(1,2))

#first plot SMR estimates
plot(egamma, exp(-ebresults$beta) * sort(ohio_map$smr), xlim = c(0.4, 1.6), ylim = c(0.4, 1.6), xlab = "SMR",
     abline(0,1))

#second plot is estimates from gamma model
plot(egamma, exp(-ebresults$beta) * sort(ebresults$RR),
     xlim = c(0.4, 1.6), ylim = c(0.4, 1.6), xlab="Exp Order Stat",
     ylab = "Obs Order Stat (Gamma)")
abline(0,1)

```

```

#par(mfrow = c(2,1))
plot(ohio_map$smr, ebresults$RR, xlab="SMR", ylab="EB estimates",
     xlim = c(0.3, 1.7), ylim = c(0.3, 1.7),
     pch = 1, col="purple")
abline(0,1)

plot(log(ohio_map$smr), log(ebresults$RR),
     xlim= c(-1.5, 1.5), ylim = c(-1, 1),
     xlab="log(SMR)", ylab="log(EB estimate)",
     pch = 1, col="purple")
abline(0,1)

par(mfrow = c(1,1), mar=c(1,1,1,1))
thresh_1.2 <- EBpostthresh(ohio_map$Obs, ohio_map$Exp, alpha = ebresults$alpha,
                           beta = ebresults$beta, rrthresh=1.2)
mapvariable(thresh_1.2, ohio_map)

lnprior <- LogNormalPriorCh(1,5,0.5,0.95)
lnprior

plot(seq(0,7,0.1), dlnorm(seq(0,7,0.1),
                           meanlog=lnprior$mu, sdlog=lnprior$sigma),
     type="l", xlab=expression(theta),
     ylab="LogNormal Density")

spplot(ohio_map, c("smr"), at = c(0,0.2,0.4,0.6, 0.8, 1, 1.2, 1.4, 1.6), col.regions = rev(brewer.pal(8

ohio_map2 <- as.data.frame(ohio_map)

pcprec <- list(theta = list(prior = "pc.prec",
                           param = c(1, 0.05)))
ohio.fit1 <- inla(Obs ~ 1 + f(fips,
                           model = "iid", hyper= pcprec), data = ohio_map2,
                 family="poisson", E = Exp, control.predictor = list(compute=TRUE))

summary(ohio.fit1)

names(ohio.fit1)

expbeta0med <- ohio.fit1$summary.fixed[4]
sdmed <- 1/sqrt(ohio.fit1$summary.hyperpar[4])

expbeta0med
sdmed

#summary(ohio.fit1)

```

```

ohio.fit1$summary.fixed

lnorminter <- ohio.fit1$summary.fixed[4]
lnormREs <- exp(ohio.fit1$summary.random$fips[5])

lnormRRs <- as.double(exp(lnorminter)) * lnormREs[,1]

ohio_map$RRlnorm <- lnormRRs
spplot(ohio_map, c("RRlnorm"), col.regions=colorRampPalette(rev(brewer.pal(8, "RdBu")))(50))

par(mfrow = c(1,1))

plot(ohio_map$RRlnorm, ohio_map$smr, xlab="SMR", ylab="Lognormal estimates",
     xlim = c(0,2), ylim = c(0,2),
     pch=1, col="purple")
abline(0,1)

plot(ohio_map$RRlnorm, ebresults$RR, xlab="SMR", ylab="EB estimates",
     xlim = c(0.3, 1.7), ylim = c(0.3, 1.7),
     pch = 1, col="purple")
abline(0,1)

```