

Birth Outcomes Associated with CAFOs in Different Race

<https://github.com/cwangjared/ENV872-FINAL-PROJECT.git>

Jared Wang

4/24/2020

```
#-----working directory-----  
getwd()  
  
#-----load packages-----  
library(tidyverse)  
library(ggthemes)  
library(ggplot2)  
library(tidyr)  
library(gridExtra)  
library(nlme)  
  
#-----set ggplot theme-----  
theme.hc01 <- theme_hc() +  
  theme(axis.title = element_text(family = "serif", size = (10)),  
        axis.text = element_text(family = "serif", size = (8), color = "black"))  
  
#-----load dataset-----  
df.birth.temp <- read.csv("../data/raw/birth.csv")
```

1 Rationale and Research Questions

Concentrated animal feeding operations is a significant source of water and air contaminants in southeastern North Carolina. Contamination from these facilities poses significant health threats to local communities. Residents living in areas with more CAFOs tend to have higher odds for adverse health outcomes such as respiratory disease and kidney disease. My Masters' Project investigated association between proximity to CAFOs and infant birth outcomes (gestational age and birth weight). In the past project, I identified that higher hog density and larger hog size are associated with smaller gestational age and birth weight.

While the Masters' Project focuses on impact of CAFOs on the general population, it only take race as a confounding variable. However, certain races, such as Black and Latino, are usually more strongly impacted by adverse environmental conditions than other races. For this project, I am interested in exploring if a certain race is disproportionately vulnerable to CAFOs. This project aims to use generalized linear model to explore association between hog CAFO kernel density score (explained in the next section) and birth outcomes (gestational age & birth weight), stratified by race.

The research question is: is the relationship between hog CAFO kernel density scores and birth outcomes the same among different races?

2 Dataset Information

Predictor: hog CAFO kernel density score (represents impact from hog CAFOs) Birth outcomes: birth weight & gestational age Confounding variables: infant sex, maternal age, education, smoking history, BMI, and marital status

This project uses 1) NC household level demographic and birth outcome dataset in 2016 and 2) hog CAFO kernel density scores. Both datasets are used in and wrangled for the purpose of my Masters' Project. In this project, I take the dataset wrangled for my MP and modify it to make it fit this project.

The demographic and birth outcome dataset is obtained from the NC birth certificate dataset (confidential information excluded). The dataset include the following variables: birth weight, gestational age, infant sex, mother's age, prenatal BMI, smoking history, prenatal care index level, naternal race, and maternal education.

Kernel density score reflects relative impact of hog CAFOs on each household (each household has a unique score). It is calculated from numbers of hog CAFOs in 5 mi radius region, animal count in each CAFO, and animal size in each CAFO. The computation is completed by the Kernel Density tool in ArcGIS. The GIS model was completed as a part of the MP. Results are previously merged with the birth certificate dataset (code included below).

```
#-----data wrangling completed as a part of MP-----
df.birth <- df.birth.temp %>%
  select(KID, GEST, LBS, OZS, SEX, MARITAL, CIGPN, #select explanatory variables
         CIGFN, CIGSN, CIGLN, MEDUC, MRACE, MHISP,
         KOTEL, MAGE, BMI, PLUR, Poul1, Poul2, Poul5,
         Hog1, Hog2, Hog5, gridcodehog, gridcodepol) %>%
  mutate(HOG1 = Hog1, HOG2 = Hog2, HOG5 = Hog5,
         POU1 = Poul1, POU2 = Poul2, POU5 = Poul5,
         EFFHOG = gridcodehog, EFFPOU = gridcodepol) %>%
  select(-Hog1, -Hog2, -Hog5, -Poul1, -Poul2, -Poul5,
         -gridcodehog, -gridcodepol) %>%
  filter(PLUR == 1) %>%
  mutate(SEX = ifelse(SEX == 9, NA, SEX),
         MARITAL = ifelse(MARITAL == 9, NA, MARITAL),
         CIGPN = ifelse(CIGPN == 99, NA, CIGPN),
         CIGLN = ifelse(CIGLN == 99, NA, CIGLN),
         CIGSN = ifelse(CIGSN == 99, NA, CIGSN),
         CIGFN = ifelse(CIGFN == 99, NA, CIGFN),
         MEDUC = ifelse(MEDUC == 9, NA, MEDUC),
         KOTEL = ifelse(KOTEL == 0, NA, KOTEL),
         BMI = ifelse(BMI == 999, NA, BMI)) %>%
  mutate(ERROR = ifelse(MAGE == 29 & BMI == 25.7, 1, 0),
         ERROR = as.factor(ERROR),
         ERROR.GEST40 = ifelse(ERROR == 1 & GEST == 40, 1, 0),
         ERROR.GEST40 = as.factor(ERROR.GEST40)) %>%
```

```

select(-PLUR) %>%
mutate(HOG1 = as.factor(HOG1), #create 1 mi, 2 mi, and 5 mi distance zones
       HOG2 = as.factor(HOG2),
       HOG5 = as.factor(HOG5),
       POU1 = as.factor(POU1),
       POU2 = as.factor(POU2),
       POU5 = as.factor(POU5),
       DSTHOG = ifelse(HOG5 == "1" & HOG2 == "0" & HOG1 == "0", "x5", NA),
       DSTHOG = ifelse(HOG2 == "1" & HOG1 == "0", "x2", DSTHOG),
       DSTHOG = ifelse(HOG1 == "1", "x1", DSTHOG),
       DSTHOG = ifelse(HOG5 == "0", "x0", DSTHOG),
       DSTHOG = as.factor(DSTHOG),
       DSTPOU = ifelse(POU5 == "1" & POU2 == "0" & POU1 == "0", "x5", NA),
       DSTPOU = ifelse(POU2 == "1" & POU1 == "0", "x2", DSTPOU),
       DSTPOU = ifelse(POU1 == "1", "x1", DSTPOU),
       DSTPOU = ifelse(POU5 == "0", "x0", DSTPOU),
       DSTPOU = as.factor(DSTPOU)) %>%
mutate(LBS = ifelse(LBS == 99, NA, LBS), #convert birth weight to kilogram
       OZS = ifelse(OZS == 99, NA, OZS),
       lbs_kg = LBS/2.205,
       ozs_kg = OZS/35.274,
       WTKG = lbs_kg + ozs_kg) %>%
select(-lbs_kg, -ozs_kg, -LBS, -OZS) %>% #delete used columns
mutate(NUMSMOKE = CIGPN + CIGLN + CIGSN + CIGFN,
       SMOKE = ifelse(NUMSMOKE != 0, 1, 2),
       SMOKE = as.factor(SMOKE)) %>%
select(-CIGPN, -CIGLN, -CIGSN, -CIGFN, -NUMSMOKE) %>%
mutate(MEDUC = ifelse(MEDUC %in% c(1, 2, 3), 1, MEDUC),
       MEDUC = ifelse(MEDUC %in% c(4, 5), 2, MEDUC),
       MEDUC = ifelse(MEDUC == 6, 3, MEDUC),
       MEDUC = ifelse(MEDUC %in% c(7, 8), 4, MEDUC)) %>%
mutate(SEX = as.factor(SEX), #convert variables to factors
       MARITAL = as.factor(MARITAL),
       MEDUC = as.factor(MEDUC),
       KOTEL = as.factor(KOTEL),
       MRACE = as.factor(MRACE),
       MHISP = as.factor(MHISP)) %>%
mutate(MRACE = ifelse(MHISP %in% c("C", "M", "O", "P", "S"), "H", MRACE), #collapse r
       MRACE = ifelse(MRACE == "2", "W", MRACE),
       MRACE = ifelse(MRACE == "3", "B", MRACE),
       MRACE = ifelse(MRACE %in% c("1", "4", "5", "6", "7", "8", "9"), "0", MRACE),
       MRACE = as.factor(MRACE)) %>%
select(-MHISP) %>%
mutate(GEST = ifelse(GEST == 99, NA, GEST), #clean gestational age values

```

```

PRETERM = ifelse(GEST < 37, "1", "0"),
PRETERM = as.factor(PRETERM),
WTKG = ifelse(WTKG > 7, NA, WTKG),
WTKG = ifelse(WTKG < 0.1, NA, WTKG),
LBW = ifelse(WTKG < 2.5, "1", "0"),
LBW = as.factor(LBW)) %>% #clean birth weight values
na.omit()

```

The dataset used for MP is almost ready for this analysis. To simplify the dataset, the following wrangling approach is used. The ten variables used in this analysis is selected and renamed. Then variables are recoded. The dataset covers the entire state of NC, while kernel density analysis is only applicable for households living within 5 mi from a hog CAFO. Therefore, all zero values for kernel density should be eliminated.

```

df.birth.kernel <- df.birth %>%
  select(GEST, WTKG, EFFHOG, MARITAL, SEX, MAGE,
         MEDUC, MRACE, KOTEL, BMI) %>% #select desired variables
  mutate(gestational.age = GEST, birth.weight = WTKG,
         kernel.score = EFFHOG, marital.status = MARITAL,
         infant.sex = SEX, mother.age = MAGE,
         education = MEDUC, race = MRACE,
         prenatal.care = KOTEL) %>%
  mutate(marital.status = ifelse(marital.status == 1, "married", "single"), #recode var
         marital.status = as.factor(marital.status),
         infant.sex = ifelse(infant.sex == 1, "male", "female"),
         infant.sex = as.factor(infant.sex),
         prenatal.care = as.factor(prenatal.care),
         education = ifelse(education == 1, "highschool.below", education),
         education = ifelse(education == 2, "some.college", education),
         education = ifelse(education == 3, "complete.college", education),
         education = ifelse(education == 4, "graduate", education),
         education = as.factor(education),
         race = as.factor(race)) %>%
  select(-GEST, -WTKG, -EFFHOG, -MARITAL, -SEX, -MAGE,
         -MEDUC, -MRACE, -KOTEL) %>%
  filter(kernel.score != 0)

#-----export the dataset-----
write.csv(df.birth.kernel, "../data/processed/birth.kernel.csv")

```

3 Exploratory Analysis

In general, distribution of kernel scores is skewed and data transformation may be needed. The report is associated with some other values. The total number is large.

```
summary(df.birth.kernel)
```

```
##      BMI      gestational.age  birth.weight    kernel.score
##  Min.   : 7.90   Min.    :17.0   Min.     :0.1701   Min.    :  1.0
## 1st Qu.:22.50   1st Qu.:38.0   1st Qu.:2.9762   1st Qu.: 11.0
## Median :26.40   Median :39.0   Median :3.3164   Median : 42.0
## Mean   :27.83   Mean    :38.6   Mean    :3.2774   Mean    :160.4
## 3rd Qu.:31.80   3rd Qu.:40.0   3rd Qu.:3.6565   3rd Qu.:169.0
## Max.   :66.80   Max.    :43.0   Max.    :6.0374   Max.    :4864.0
## marital.status  infant.sex    mother.age      education
## married:10693   female: 9820   Min.     :11.00   complete.college:2842
## single : 9452   male  :10325   1st Qu.:23.00   graduate         :1217
##                                     Median :27.00   highschool.below:8734
##                                     Mean    :27.29   some.college     :7352
##                                     3rd Qu.:31.00
##                                     Max.    :53.00
## race      prenatal.care
## B: 5633    1:3916
## H: 3199    2:1359
## O: 1101    3:5618
## W:10212    4:9252
##
##
```

In general, distribution of birth weight and gestational age values fit normal distribution, while that of kernel density does not. Natural logarithm is therefore applied to transform the kernel density data. Natural-log-transformed kernel density score generally fits normal distribution (Figure 1).

```
hist.bw <- ggplot(df.birth.kernel) +
  geom_histogram(aes(birth.weight), binwidth = 0.025, fill = "lightskyblue4") +
  labs(y= "# Household", x = "Birth weight (kg)") +
  theme.hc01

hist.gest <- ggplot(df.birth.kernel) +
  geom_histogram(aes(gestational.age), binwidth = 1, fill = "lightskyblue4") +
  labs(y= "# Household", x = "Gestational age (weeks)") +
  theme.hc01

hist.kern <- ggplot(df.birth.kernel) +
  geom_histogram(aes(kernel.score), binwidth = 1, fill = "lightskyblue4") +
```

```

scale_y_continuous(limits = c(0, 300)) +
labs(y= "# Household", x = "Kernel density score") +
theme.hc01

df.birth.kernel <- df.birth.kernel %>%
  mutate(kernel.score.log = log(kernel.score))

hist.kernlog <- ggplot(df.birth.kernel) +
  geom_histogram(aes(kernel.score.log), binwidth = 1, fill = "lightskyblue4") +
  labs(y= "# Household", x = "Log kernel density score") +
  theme.hc01

grid.arrange(hist.bw, hist.gest,
              hist.kern, hist.kernlog, ncol = 2)

```

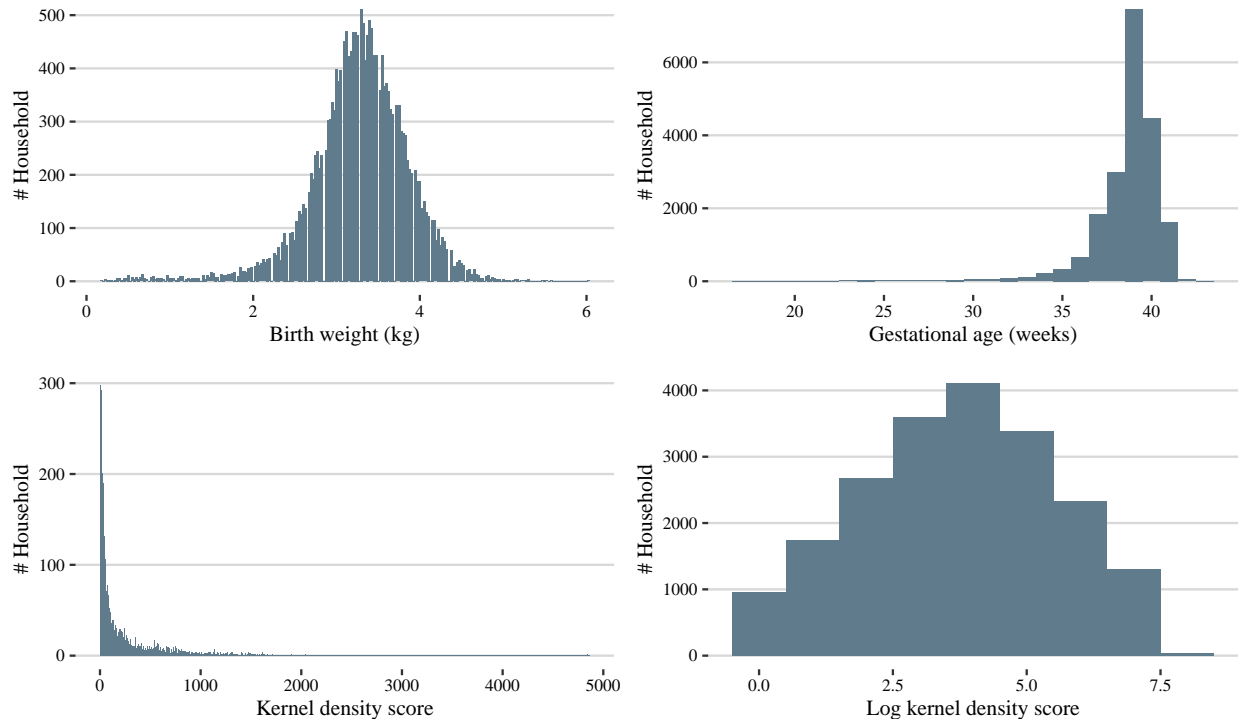


Figure 1: visualization of birth weight, gestational age, and hog CAFO kernel density scores.

It is difficult to conclude a general trend from visualized data. However, there appear to be difference between average birth weight among different race. For example, white and hispanic infants tend to have higher average birth weight than black infants (Figure 2).

```

#create four datasets
df.black <- df.birth.kernel %>%
  filter(race == "B")
df.hisp <- df.birth.kernel %>%
  filter(race == "H")

```

```

df.white <- df.birch.kernel %>%
  filter(race == "W")
df.other <- df.birch.kernel %>%
  filter(race == "O")

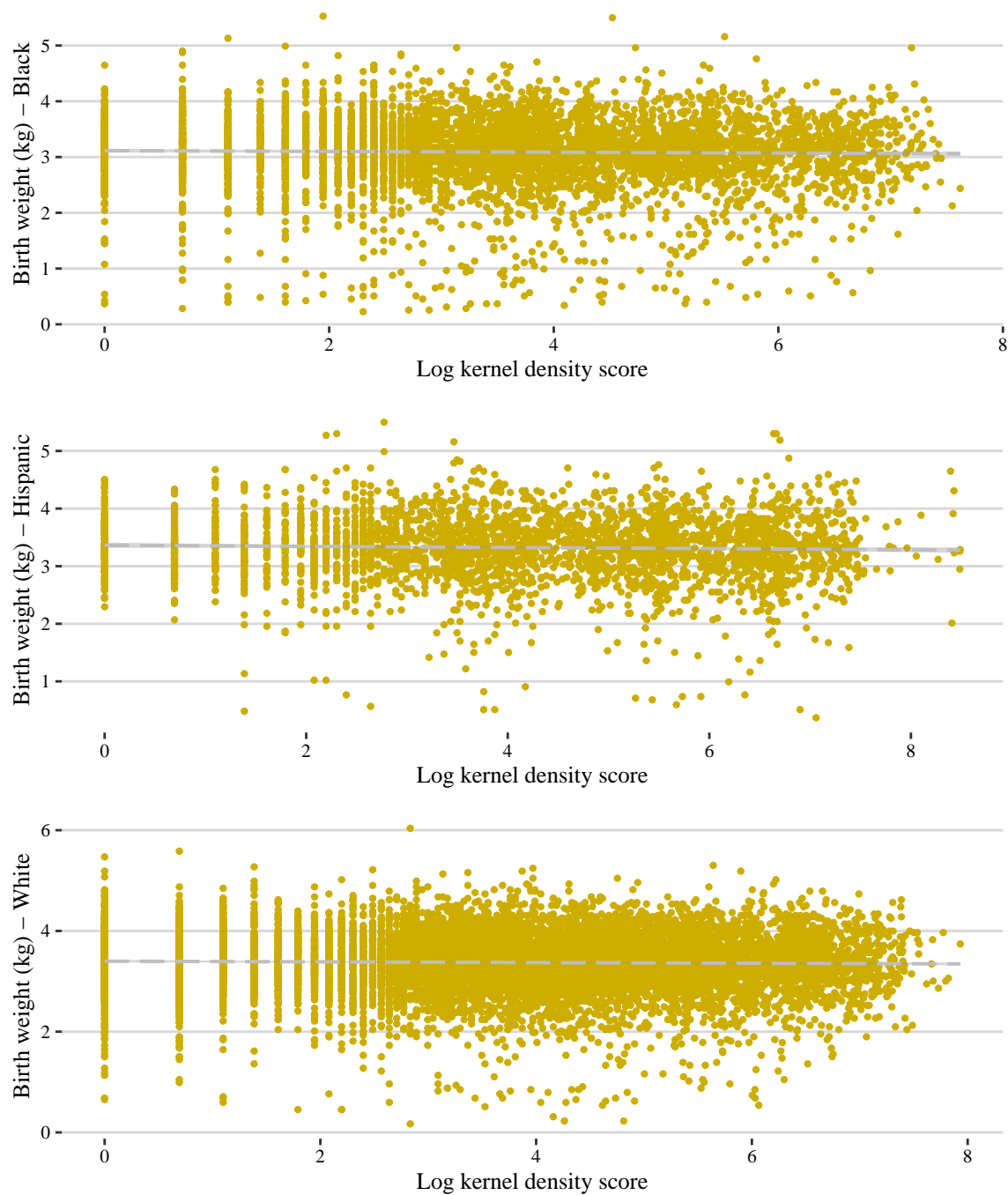
#plot
scat.bw.black <- ggplot(df.black) +
  geom_point(aes(x = kernel.score.log, y = birth.weight),
    size = 0.8, alpha = 1, color = "gold3") +
  geom_smooth(aes(x = kernel.score.log, y = birth.weight),
    method = lm,
    lty = 5, lwd = 0.7, color = "gray") +
  labs(x = "Log kernel density score",
    y = "Birth weight (kg) - Black") +
  #scale_color_manual(values = c("gold3", "lightskyblue3", "gray", "red3")) +
  theme.hc01

scat.bw.hisp <- ggplot(df.hisp) +
  geom_point(aes(x = kernel.score.log, y = birth.weight),
    size = 0.8, alpha = 1, color = "gold3") +
  geom_smooth(aes(x = kernel.score.log, y = birth.weight),
    method = lm,
    lty = 5, lwd = 0.7, color = "gray") +
  labs(x = "Log kernel density score",
    y = "Birth weight (kg) - Hispanic") +
  #scale_color_manual(values = c("gold3", "lightskyblue3", "gray", "red3")) +
  theme.hc01

scat.bw.white <- ggplot(df.white) +
  geom_point(aes(x = kernel.score.log, y = birth.weight),
    size = 0.8, alpha = 1, color = "gold3") +
  geom_smooth(aes(x = kernel.score.log, y = birth.weight),
    method = lm,
    lty = 5, lwd = 0.7, color = "gray") +
  labs(x = "Log kernel density score",
    y = "Birth weight (kg) - White") +
  #scale_color_manual(values = c("gold3", "lightskyblue3", "gray", "red3")) +
  theme.hc01

grid.arrange(scat.bw.black, scat.bw.hisp, scat.bw.white,
  ncol = 1)

```

Figure

2: Birth weight against log kernel density score for different race.

```
#plot
scat.gest.black <- ggplot(df.black) +
  geom_point(aes(x = kernel.score.log, y = gestational.age),
    size = 0.8, alpha = 1, color = "gold3") +
  geom_smooth(aes(x = kernel.score.log, y = gestational.age),
    method = lm,
    lty = 5, lwd = 0.7, color = "gray") +
```

```

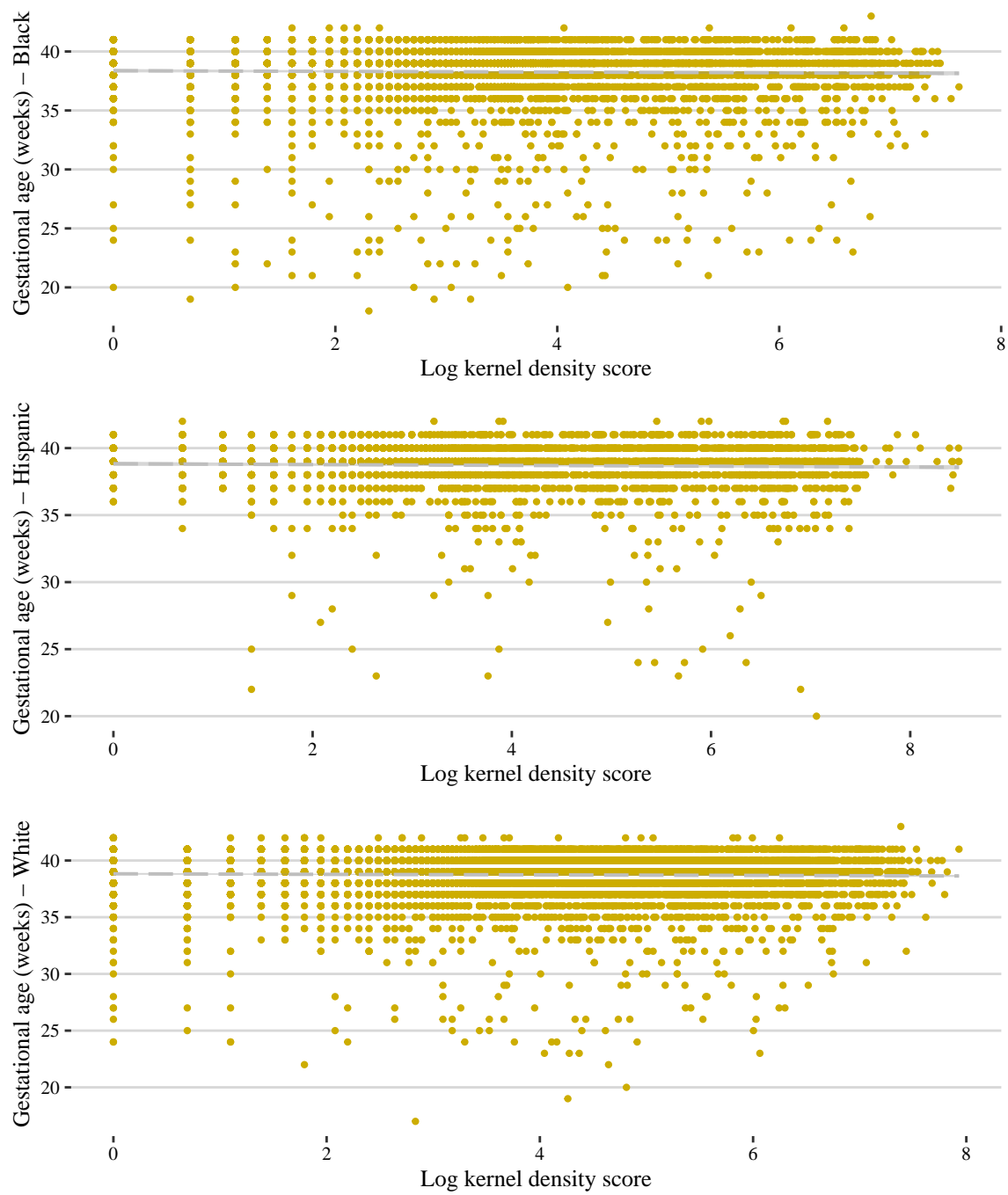
labs(x = "Log kernel density score",
     y = "Gestational age (weeks) - Black") +
#scale_color_manual(values = c("gold3", "lightskyblue3", "gray", "red3")) +
theme.hc01

scat.gest.hisp <- ggplot(df.hisp) +
  geom_point(aes(x = kernel.score.log, y = gestational.age),
            size = 0.8, alpha = 1, color = "gold3") +
  geom_smooth(aes(x = kernel.score.log, y = gestational.age),
            method = lm,
            lty = 5, lwd = 0.7, color = "gray") +
  labs(x = "Log kernel density score",
       y = "Gestational age (weeks) - Hispanic") +
#scale_color_manual(values = c("gold3", "lightskyblue3", "gray", "red3")) +
theme.hc01

scat.gest.white <- ggplot(df.white) +
  geom_point(aes(x = kernel.score.log, y = gestational.age),
            size = 0.8, alpha = 1, color = "gold3") +
  geom_smooth(aes(x = kernel.score.log, y = gestational.age),
            method = lm,
            lty = 5, lwd = 0.7, color = "gray") +
  labs(x = "Log kernel density score",
       y = "Gestational age (weeks) - White") +
#scale_color_manual(values = c("gold3", "lightskyblue3", "gray", "red3")) +
theme.hc01

grid.arrange(scat.gest.black, scat.gest.hisp, scat.gest.white,
             ncol = 1)

```



Figure

3: Gestational age against log kernel density score for different race.

4 Analysis

Generalized linear models are used to predict birth weight and gestational age with natural log of kernel score and other confounding variables. One log unit increase in hog CAFO kernel density score among black and hispanic women is associated with 0.009 and 0.010 kg lower birth weight on average ($p < 0.05$). One log unit increase in kernel score among white women may be associated with 0.005 kg lower birth weight on average, but the result is not statistically significant ($p > 0.05$).

```
glm.birthweight.black <- glm(data = df.black,
                             birth.weight ~ kernel.score.log + infant.sex + mother.age +
                             BMI + education + prenatal.care + marital.status)
glm.birthweight.hispanic <- glm(data = df.hisp,
                                birth.weight ~ kernel.score.log + infant.sex + mother.age +
                                BMI + education + prenatal.care + marital.status)
glm.birthweight.white <- glm(data = df.white,
                              birth.weight ~ kernel.score.log + infant.sex + mother.age +
                              BMI + education + prenatal.care + marital.status)
summary(glm.birthweight.black)
```

```
##
## Call:
## glm(formula = birth.weight ~ kernel.score.log + infant.sex +
##      mother.age + BMI + education + prenatal.care + marital.status,
##      data = df.black)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.08212  -0.27669   0.05475   0.36064   2.44794
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      3.117007   0.067448  46.214 < 2e-16 ***
## kernel.score.log -0.009302   0.004559  -2.040 0.041372 *
## infant.sexmale    0.078752   0.016257   4.844 1.31e-06 ***
## mother.age      -0.003813   0.001634  -2.333 0.019668 *
## BMI              0.006348   0.001029   6.167 7.44e-10 ***
## educationgraduate -0.084817   0.055878  -1.518 0.129097
## educationhighschool.below -0.098706   0.034037  -2.900 0.003746 **
## educationsome.college -0.066419   0.033392  -1.989 0.046741 *
## prenatal.care2     0.071317   0.034613   2.060 0.039407 *
## prenatal.care3     0.100128   0.022925   4.368 1.28e-05 ***
## prenatal.care4    -0.070144   0.020621  -3.402 0.000675 ***
## marital.statuussingle -0.058934   0.021309  -2.766 0.005698 **
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.3716648)
##
##      Null deviance: 2146.3  on 5632  degrees of freedom
## Residual deviance: 2089.1  on 5621  degrees of freedom
## AIC: 10424
##
## Number of Fisher Scoring iterations: 2
```

```
summary(glm.birthweight.hispanic)
```

```
##
## Call:
## glm(formula = birth.weight ~ kernel.score.log + infant.sex +
##      mother.age + BMI + education + prenatal.care + marital.status,
##      data = df.hisp)
##
## Deviance Residuals:
##      Min        1Q      Median        3Q        Max
## -3.06256  -0.27805   0.01654   0.32228   2.03273
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.682003   0.076625  35.002 < 2e-16 ***
## kernel.score.log  -0.010132   0.004900  -2.068  0.0387 *
## infant.sexmale     0.134912   0.018995   7.103 1.50e-12 ***
## mother.age         0.008962   0.001570   5.707 1.26e-08 ***
## BMI                0.012864   0.001626   7.912 3.45e-15 ***
## educationgraduate  -0.021025   0.081360  -0.258  0.7961
## educationhighschool.below -0.005523   0.043910  -0.126  0.8999
## educationsome.college  0.002226   0.047058   0.047  0.9623
## prenatal.care2       0.088074   0.037634   2.340  0.0193 *
## prenatal.care3       0.117177   0.026128   4.485 7.56e-06 ***
## prenatal.care4      -0.060220   0.024488  -2.459  0.0140 *
## marital.statuissingle -0.003660   0.020227  -0.181  0.8564
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.2877486)
##
##      Null deviance: 983.60  on 3198  degrees of freedom
## Residual deviance: 917.05  on 3187  degrees of freedom
## AIC: 5107.5
##
```

```
## Number of Fisher Scoring iterations: 2
```

```
summary(glm.birtheight.white)
```

```
##
```

```
## Call:
```

```
## glm(formula = birth.weight ~ kernel.score.log + infant.sex +  
##     mother.age + BMI + education + prenatal.care + marital.status,  
##     data = df.white)
```

```
##
```

```
## Deviance Residuals:
```

```
##      Min       1Q   Median       3Q      Max  
## -3.432  -0.295   0.013   0.344   2.568
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)  
## (Intercept)      3.225049   0.044614  72.288 < 2e-16 ***  
## kernel.score.log -0.005209   0.003042  -1.712  0.08684 .  
## infant.sexmale    0.120412   0.010875  11.072 < 2e-16 ***  
## mother.age       -0.001985   0.001151  -1.725  0.08460 .  
## BMI              0.008493   0.000801  10.603 < 2e-16 ***  
## educationgraduate 0.013000   0.021995   0.591  0.55451  
## educationhighschool.below -0.173138   0.017717  -9.773 < 2e-16 ***  
## educationsome.college -0.066819   0.015248  -4.382 1.19e-05 ***  
## prenatal.care2    0.085260   0.027067   3.150  0.00164 **  
## prenatal.care3    0.145178   0.018099   8.021 1.16e-15 ***  
## prenatal.care4   -0.019000   0.016861  -1.127  0.25983  
## marital.statuussingle -0.098531   0.013340  -7.386 1.63e-13 ***
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## (Dispersion parameter for gaussian family taken to be 0.3013919)
```

```
##
```

```
##      Null deviance: 3285.5  on 10211  degrees of freedom
```

```
## Residual deviance: 3074.2  on 10200  degrees of freedom
```

```
## AIC: 16747
```

```
##
```

```
## Number of Fisher Scoring iterations: 2
```

One log unit increase in hog CAFO kernel density score among black, hispanic, and white women is associated with 0.042, 0.034, and 0.025 week earlier delivery on average ($p < 0.05$). These values translate to 0.29, 0.27, and 0.18 days earlier. White residents living within 5 mi from a CAFO may be less severely influenced than black and hispanic residents.

```
glm.gestation.black <- glm(data = df.black,  
                           gestational.age ~ kernel.score.log + infant.sex + mother.age
```

```

        BMI + education + prenatal.care + marital.status)
glm.gestation.hispanic <- glm(data = df.hisp,
        gestational.age ~ kernel.score.log + infant.sex + mother.age
        BMI + education + prenatal.care + marital.status)
glm.gestation.white <- glm(data = df.white,
        gestational.age ~ kernel.score.log + infant.sex + mother.age
        BMI + education + prenatal.care + marital.status)
summary(glm.gestation.black)

```

```

##
## Call:
## glm(formula = gestational.age ~ kernel.score.log + infant.sex +
##     mother.age + BMI + education + prenatal.care + marital.status,
##     data = df.black)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -20.4354   -0.5467    0.4528    1.2909    5.1619
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    39.7110451   0.2832682  140.189 < 2e-16 ***
## kernel.score.log -0.0419599   0.0191478   -2.191 0.028466 *
## infant.sexmale  -0.1035856   0.0682767   -1.517 0.129287
## mother.age      -0.0353991   0.0068627   -5.158 2.58e-07 ***
## BMI             -0.0005979   0.0043230   -0.138 0.890002
## educationgraduate -0.2471556   0.2346783   -1.053 0.292309
## educationhighschool.below -0.3096997   0.1429483   -2.167 0.030314 *
## educationsome.college -0.2142490   0.1402381   -1.528 0.126630
## prenatal.care2    0.5117095   0.1453691    3.520 0.000435 ***
## prenatal.care3    0.4691546   0.0962819    4.873 1.13e-06 ***
## prenatal.care4   -0.5719678   0.0866044   -6.604 4.36e-11 ***
## marital.statussingle 0.0454597   0.0894925    0.508 0.611493
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 6.555559)
##
##      Null deviance: 38333  on 5632  degrees of freedom
## Residual deviance: 36849  on 5621  degrees of freedom
## AIC: 26592
##
## Number of Fisher Scoring iterations: 2

```

```
summary(glm.gestation.hispanic)
```

```
##
## Call:
## glm(formula = gestational.age ~ kernel.score.log + infant.sex +
##       mother.age + BMI + education + prenatal.care + marital.status,
##       data = df.hisp)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -18.5672  -0.4020   0.2875   0.8785   3.4228
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    39.474977   0.266258 148.259 < 2e-16 ***
## kernel.score.log -0.034520   0.017026  -2.027  0.0427 *
## infant.sexmale  -0.011316   0.066003  -0.171  0.8639
## mother.age      -0.012895   0.005457  -2.363  0.0182 *
## BMI             -0.003456   0.005649  -0.612  0.5407
## educationgraduate -0.140994   0.282711  -0.499  0.6180
## educationhighschool.below -0.199063   0.152577  -1.305  0.1921
## educationsome.college -0.138897   0.163518  -0.849  0.3957
## prenatal.care2    0.559363   0.130770   4.277 1.95e-05 ***
## prenatal.care3    0.470293   0.090790   5.180 2.36e-07 ***
## prenatal.care4   -0.442611   0.085093  -5.202 2.10e-07 ***
## marital.statuussingle -0.004815   0.070286  -0.069  0.9454
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 3.474352)
##
##      Null deviance: 11634  on 3198  degrees of freedom
## Residual deviance: 11073  on 3187  degrees of freedom
## AIC: 13076
##
## Number of Fisher Scoring iterations: 2
```

```
summary(glm.gestation.white)
```

```
##
## Call:
## glm(formula = gestational.age ~ kernel.score.log + infant.sex +
##       mother.age + BMI + education + prenatal.care + marital.status,
##       data = df.white)
##
```



```

## Deviance Residuals:
##      Min        1Q      Median        3Q        Max
## -21.0557   -0.4875    0.3967    0.8584    4.4272
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    39.870526   0.148101 269.211 < 2e-16 ***
## kernel.score.log -0.025653   0.010097  -2.541 0.011080 *
## infant.sexmale  -0.087127   0.036103  -2.413 0.015827 *
## mother.age      -0.019674   0.003821  -5.149 2.67e-07 ***
## BMI             -0.009226   0.002659  -3.470 0.000523 ***
## educationgraduate  0.021214   0.073014   0.291 0.771401
## educationhighschool.below -0.373938   0.058814  -6.358 2.13e-10 ***
## educationsome.college -0.159245   0.050617  -3.146 0.001660 **
## prenatal.care2     0.451236   0.089853   5.022 5.20e-07 ***
## prenatal.care3     0.530139   0.060082   8.824 < 2e-16 ***
## prenatal.care4    -0.387013   0.055973  -6.914 4.99e-12 ***
## marital.statuussingle -0.010009   0.044285  -0.226 0.821188
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 3.321356)
##
##      Null deviance: 35955  on 10211  degrees of freedom
## Residual deviance: 33878  on 10200  degrees of freedom
## AIC: 41253
##
## Number of Fisher Scoring iterations: 2

```

5 Discussion and Summary

In general, CAFO impact level quantified by the kernel density approach is a predictor of adverse birth outcomes (smaller gestational age & lower birth weight) among infants born in 2016 in NC. This project considers other health-related confounding variables. For each log unit change in kernel density score, black and hispanic mothers tend to be associated with more significant change in delivery time. Black and hispanic mothers also tend to deliver lighter infants with the same change in kernel density scores, comparing to white mothers. This finding adds to the belief that exposure to CAFOs is a potential environmental justice issue. This study indicates the possibility that, groups of minority are not only disproportionately exposed to CAFOs. At the same time, however, they may be more vulnerable given the same level of exposure.