Birth Weight Analysis

Nick Hass

1/6/2022

EDA

1.

For this analysis, we wish to understand the effect of a mothers age on the resulting birth weight. The main reasons for analyzing this data is to know what's normal for a baby's birth weight or if the baby has some sort of other problem that's leading them to have a low birth weight. The goals are to identify and report on things that affect birth weight with the information that we now have (Mothers age, Gestational age, race, and gender). Then, when we can accurately predict the expected birth weight of a child, if the actual birth weight is significantly different from the expected, then we will know that the baby has some sort of problem.

Birth Weight Data

```
bw <- read.csv('BirthWeights.txt', sep = " ")
summary(bw)
```

```
##   BirthWeight        Mage            Gage          Race
##  Min.   :1625   Min.   :13.00   Min.   :34.0   Length:832
##  1st Qu.:2781   1st Qu.:27.00   1st Qu.:38.0   Class :character
##  Median :3094   Median :31.00   Median :39.0   Mode  :character
##  Mean   :3086   Mean   :30.75   Mean   :39.5
##  3rd Qu.:3402   3rd Qu.:35.00   3rd Qu.:41.0
##  Max.   :4653   Max.   :46.00   Max.   :45.0
##      Gen
##  Length:832
##  Class :character
##  Mode  :character
##
##
##
```

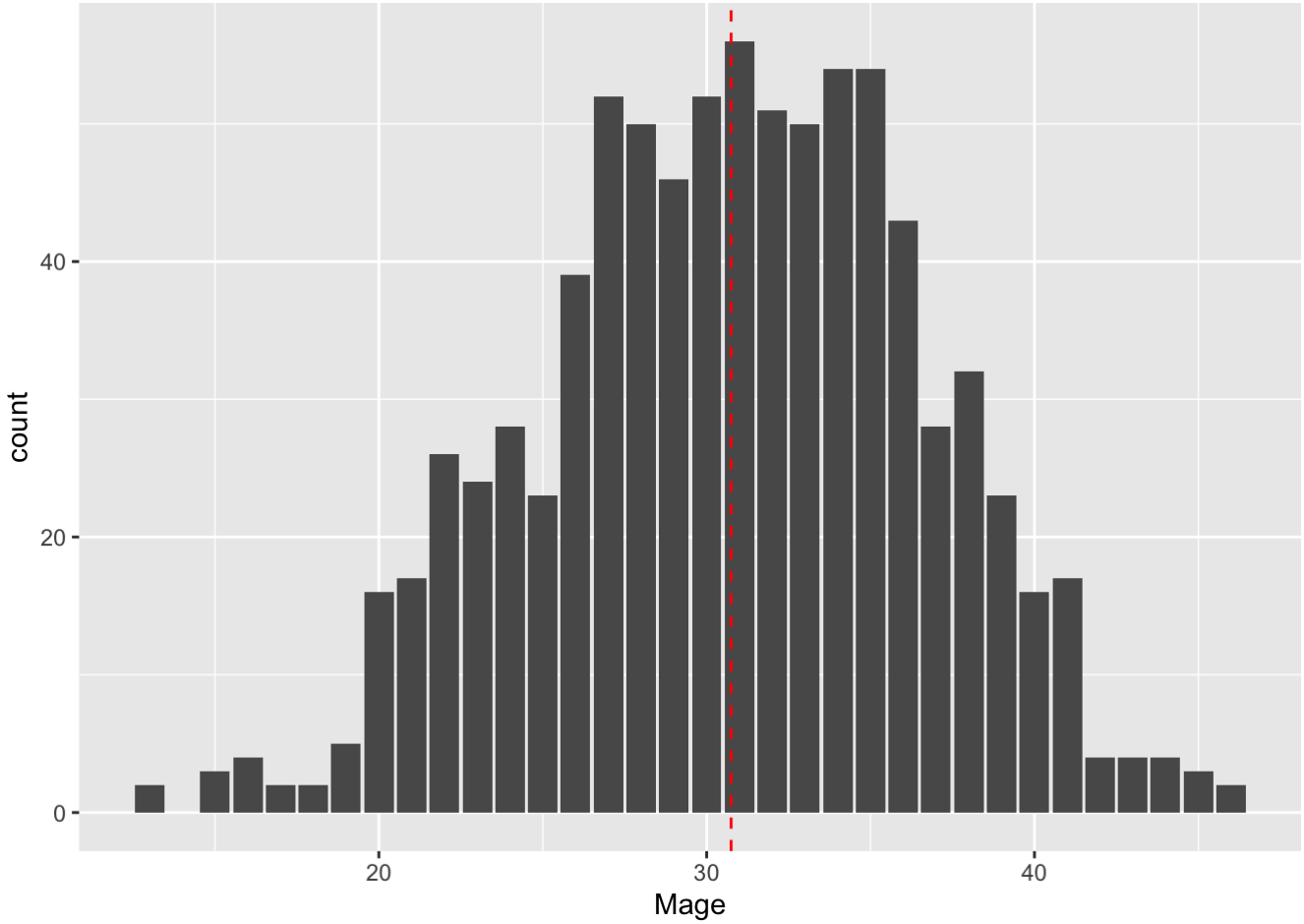
```
head(bw)
```

```
##   BirthWeight Mage Gage  Race    Gen
## 1     3026.97   23   39 white   Male
## 2     3233.11   36   39 white Female
## 3     2644.39   36   38  hisp   Male
## 4     4000.45   23   42 white   Male
## 5     3281.49   28   38 white Female
## 6     2748.88   41   41  hisp   Male
```

```
dplyr::glimpse(bw)
```

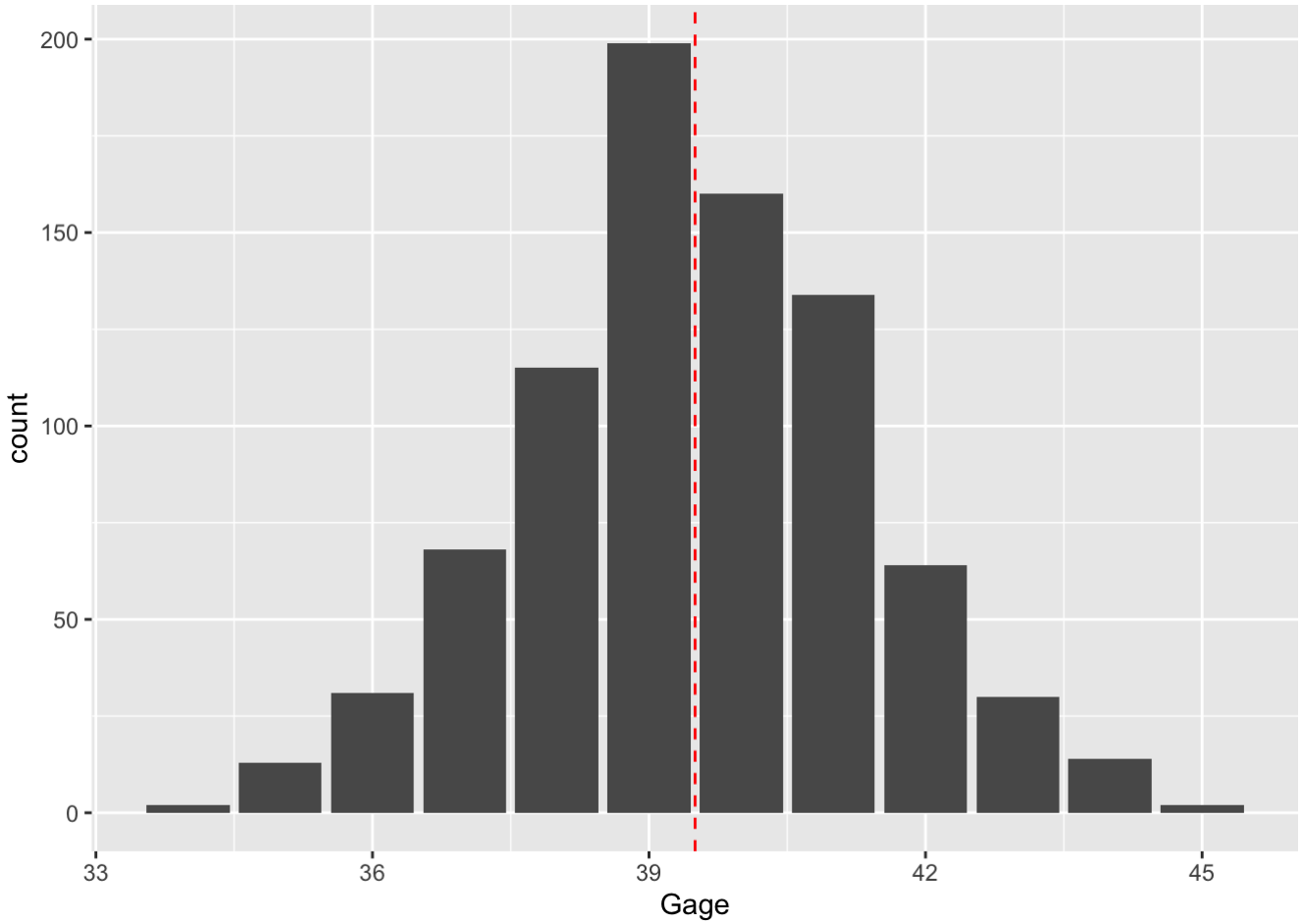
```
## Rows: 832
## Columns: 5
## $ BirthWeight <dbl> 3026.97, 3233.11, 2644.39, 4000.45, 3281.49, 2748.88, 3030…
## $ Mage        <int> 23, 36, 36, 23, 28, 41, 37, 34, 34, 33, 32, 39, 24, 26, 28…
## $ Gage        <int> 39, 39, 38, 42, 38, 41, 39, 37, 39, 40, 41, 37, 36, 35, 38…
## $ Race        <chr> "white", "white", "hisp", "white", "white", "hisp", "hisp"…
## $ Gen         <chr> "Male", "Female", "Male", "Male", "Female", "Male", "Femal…
```

```
ggplot(bw, aes(x = Mage)) +
  geom_bar() +
  geom_vline(xintercept=mean(bw$Mage), linetype="dashed", color = "red")
```



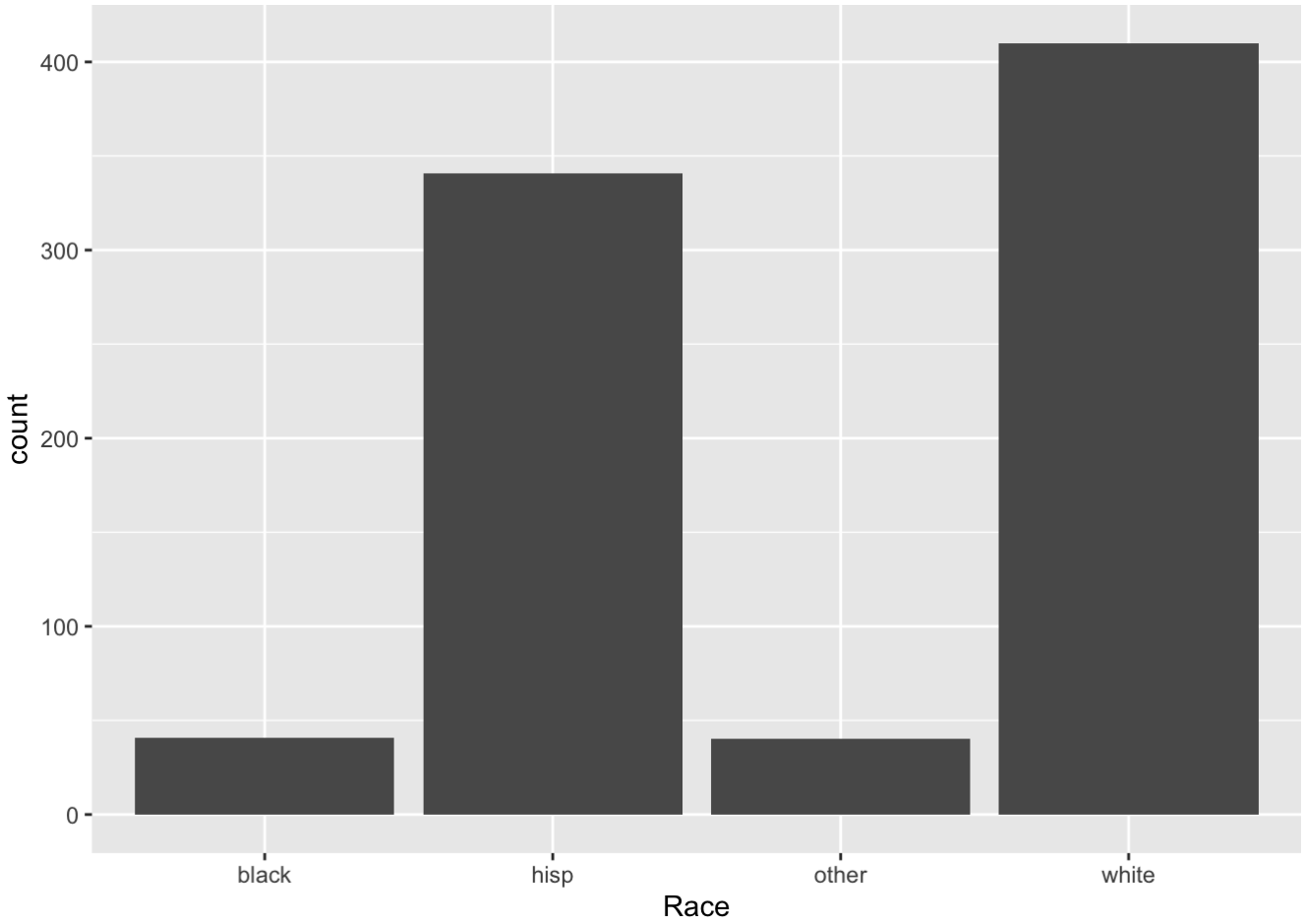
The age of mothers at the time of giving birth is centered at about 30.75 years old.

```
ggplot(bw, aes(x = Gage)) +
  geom_bar() +
  geom_vline(xintercept=mean(bw$Gage), linetype="dashed", color = "red")
```



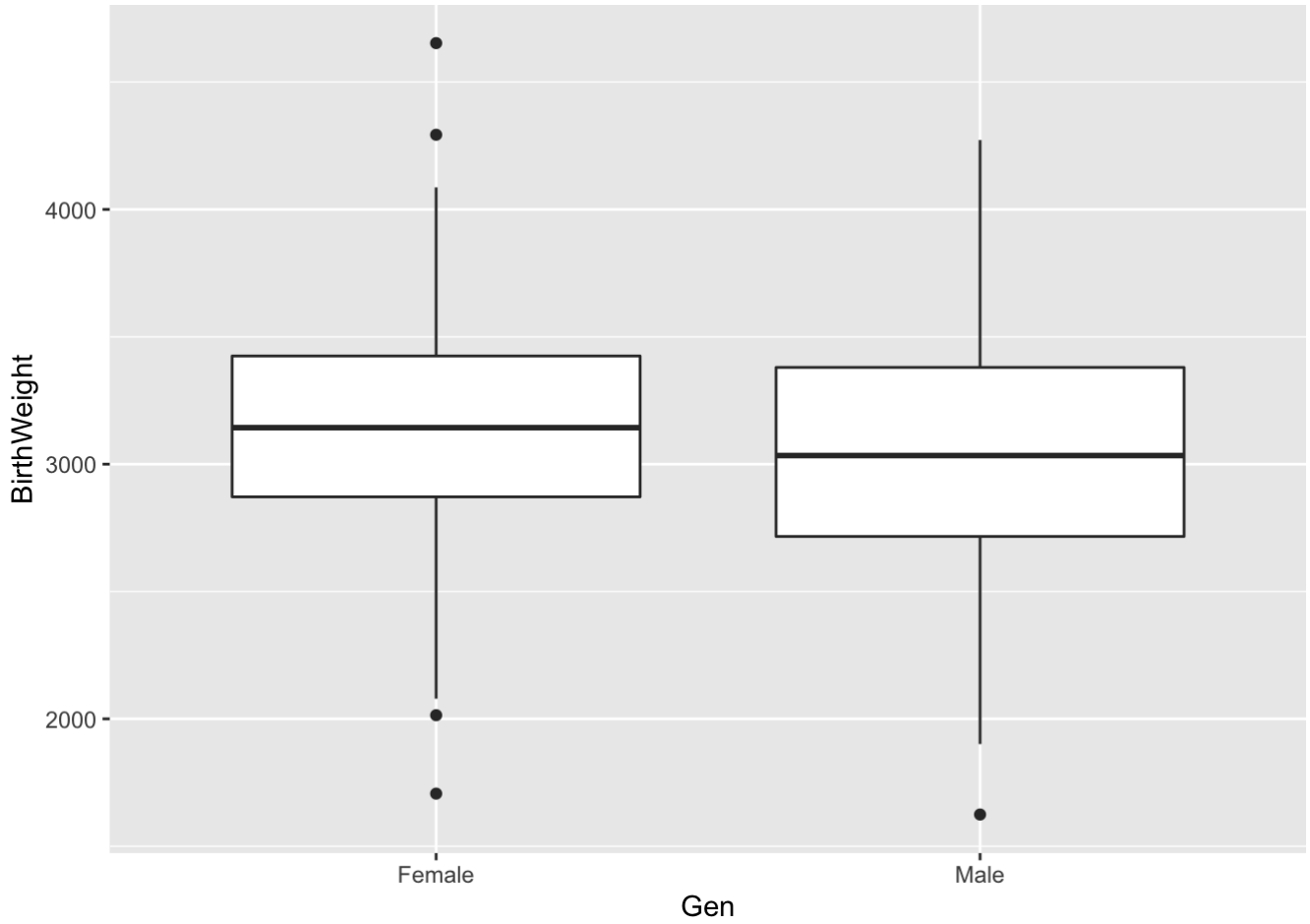
The age of gestation at the point of birth for babies in this sample is centered at 39.5 weeks.

```
ggplot(bw, aes(x = Race)) +
  geom_bar()
```



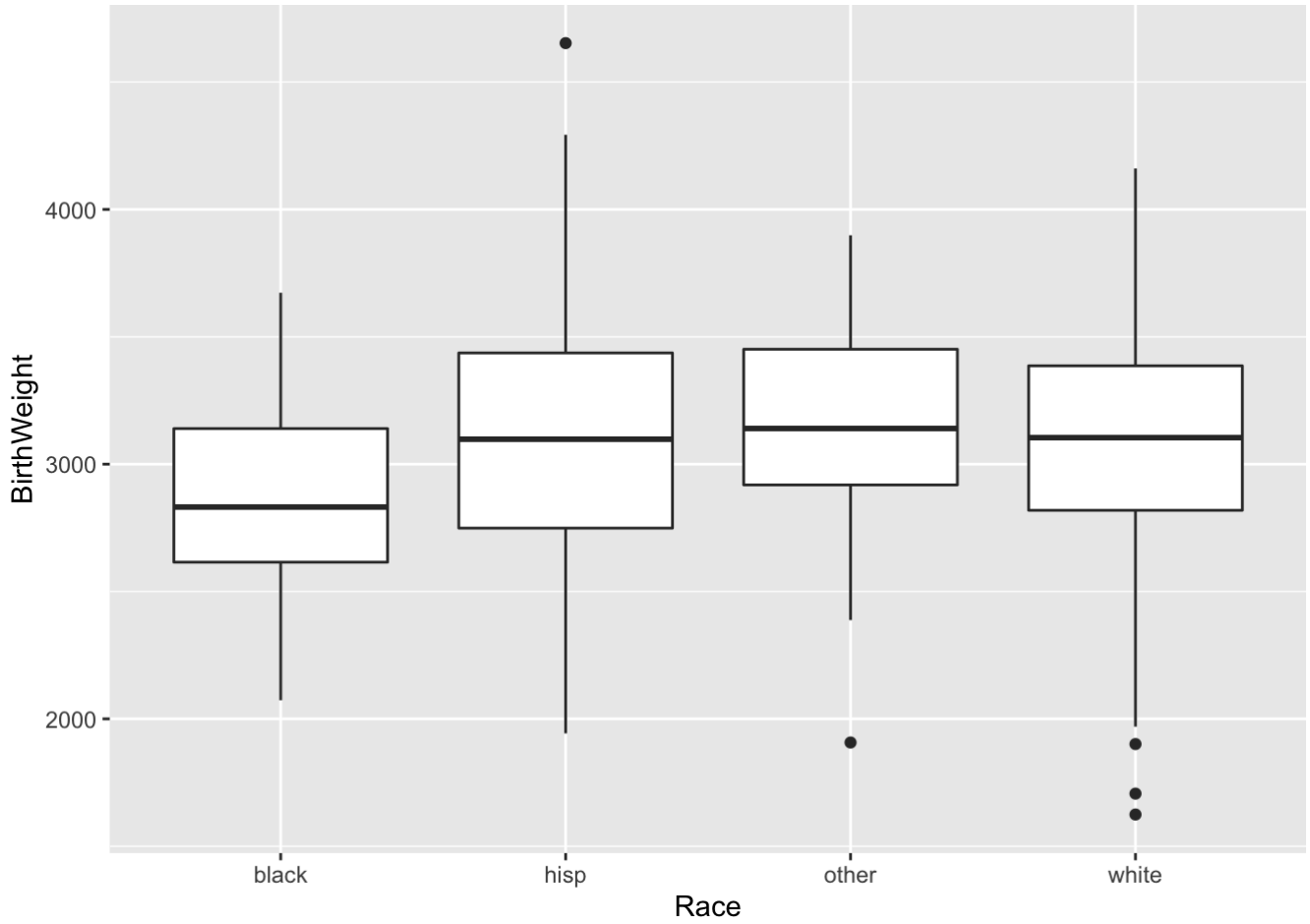
In this sample, 410 respondents were white, 341 hispanic, 41 black, and 40 other.

```
ggplot(bw, aes(x = Gen, y = BirthWeight)) +
  geom_boxplot()
```



The spread of birth weight is roughly the same for Males and Females.

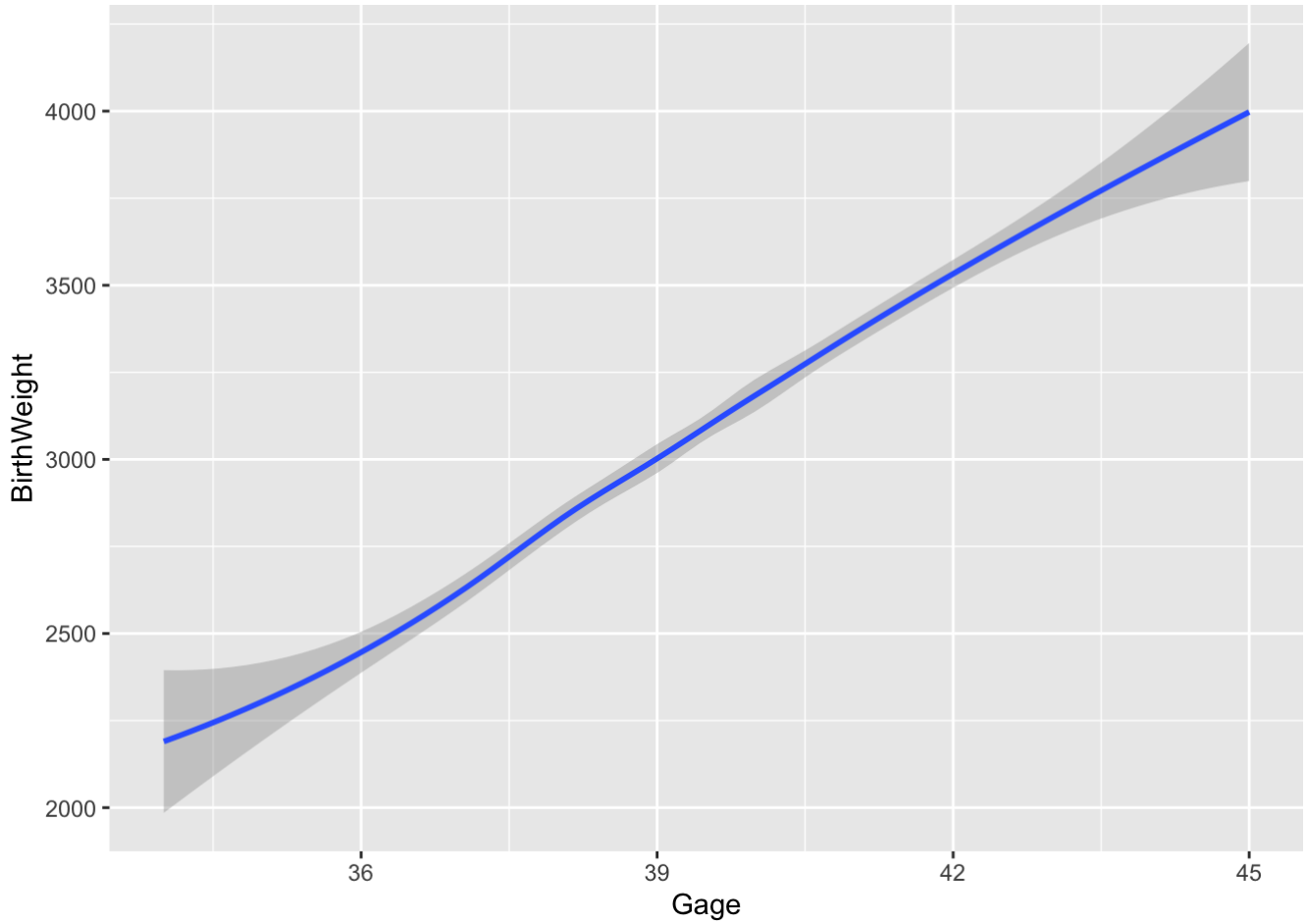
```
ggplot(bw, aes(x = Race, y = BirthWeight)) +
  geom_boxplot()
```



Black babies from this sample typically weigh less than babies of other races. White babies from this sample have the greatest spread of birth weight compared to other babies in this sample, but that is probably because there were much more observations of white babies than any other race.

```
ggplot(bw, aes(x = Gage, y = BirthWeight)) +
  geom_smooth()
```

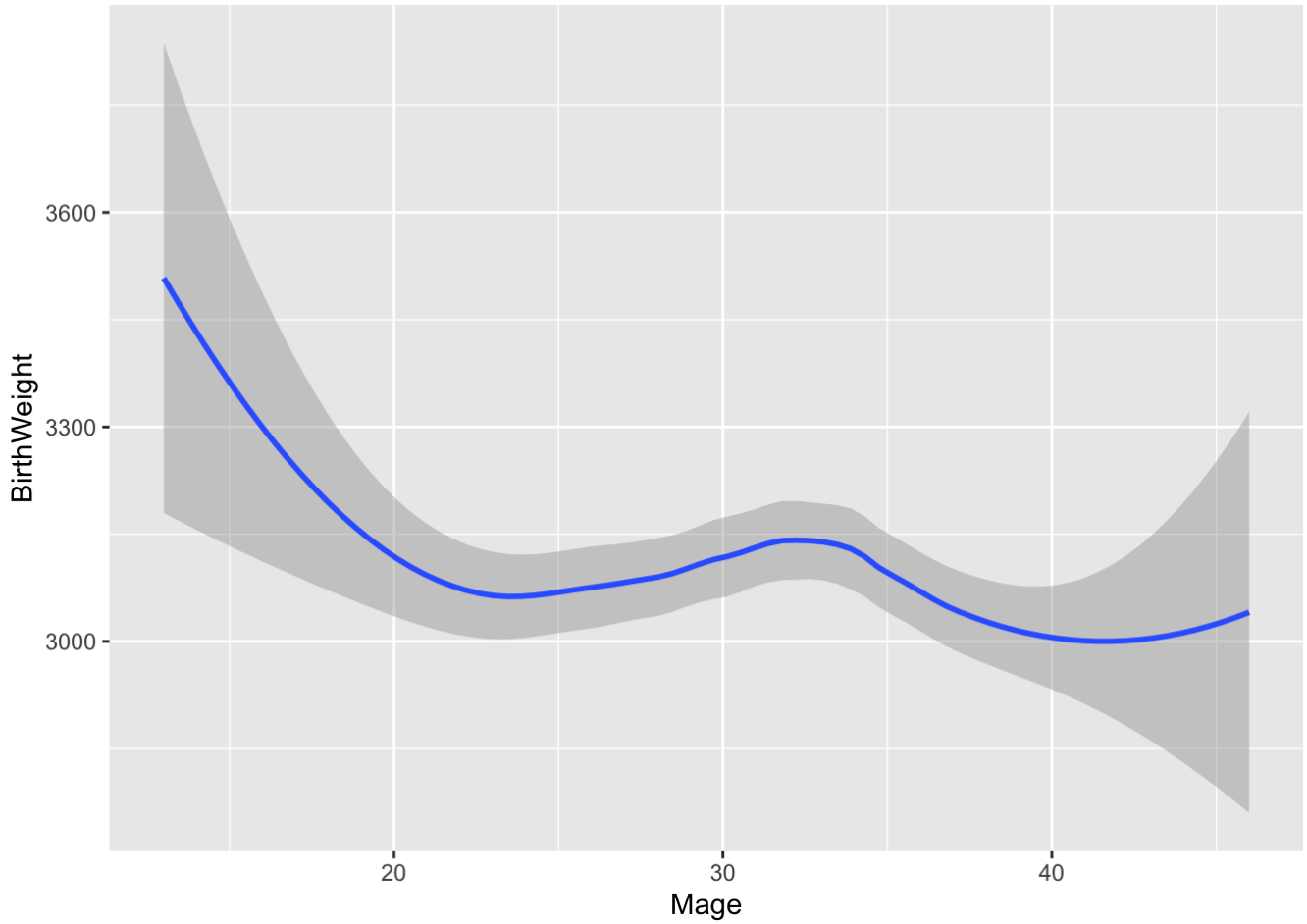
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



Babies born later in the gestational period tend to weigh more at birth.

```
ggplot(bw, aes(x = Mage, y = BirthWeight)) +
  geom_smooth()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



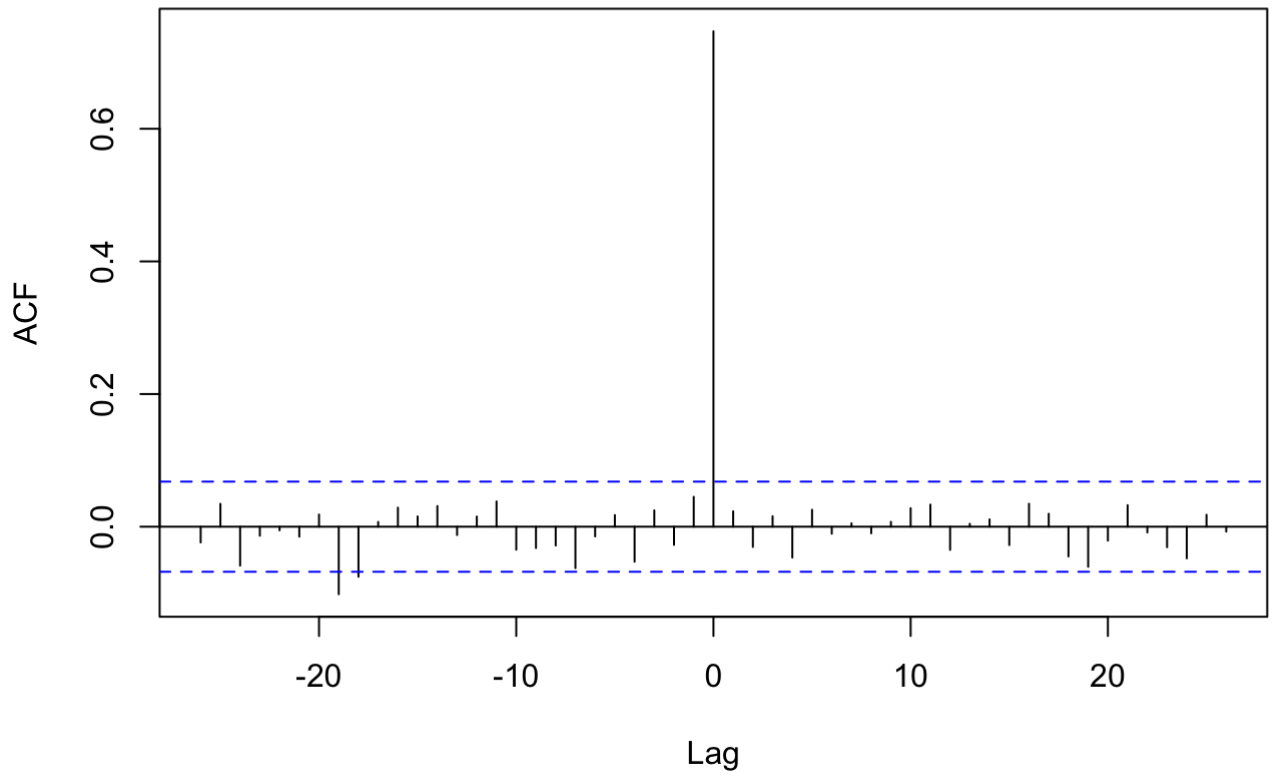
Babies with mothers younger tend to be born heavier.

3. Discuss aspects of the data that would create correlation between observations.

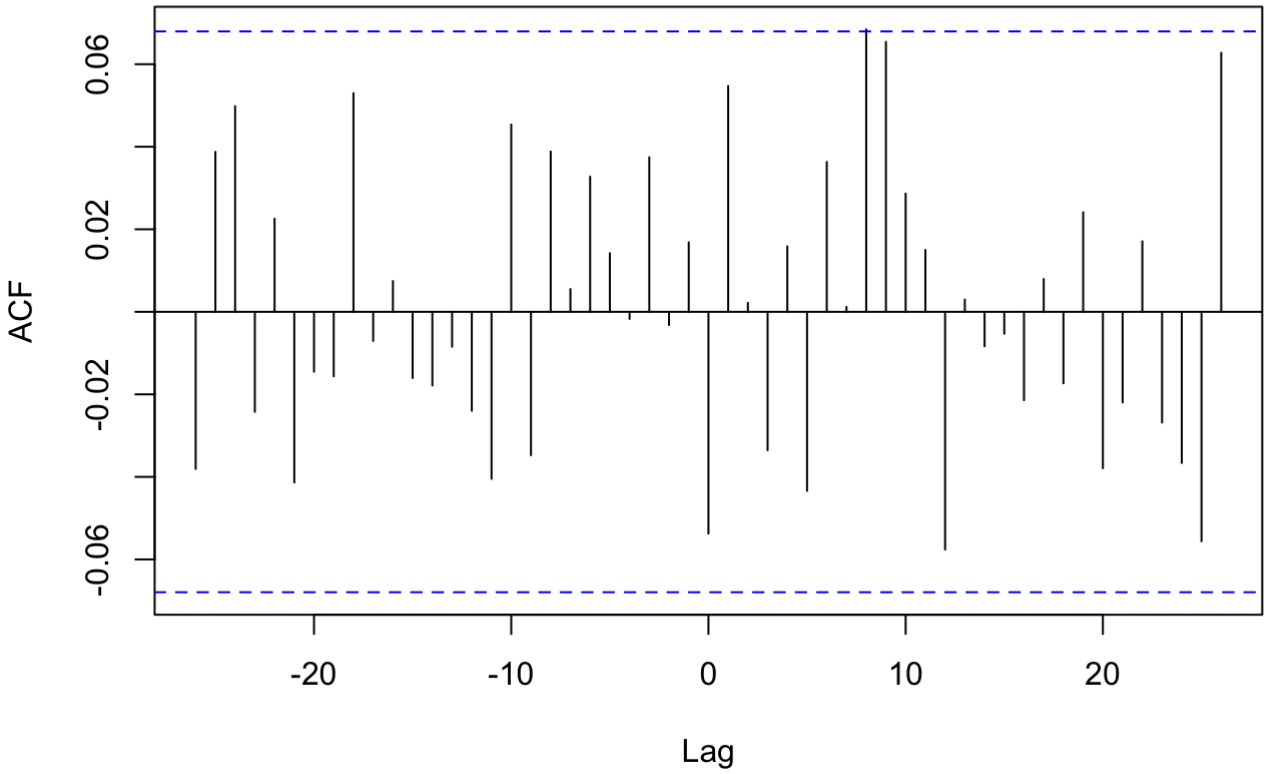
There would be correlation between observations if there were more than one baby with the same mother. This would be within subject correlation between observations.

There would be correlation between observations with mothers of the same age. There would be correlation between observations with babies of the same gestational age.

```
ccf(bw$Gage, bw$BirthWeight)
```




```
ccf(bw$Mage, bw$BirthWeight)
```



It looks like there is little to no cross-observation correlation present between Gestational Age and birth weight or Mothers age and birth weight.

4.

I propose that linear regression should be used to analyze this data. Linear regression would be useful to understand the effect of a mothers age on the resulting birth weight because we can use this to predict what an expected average birth weight to be given the age of the mother. We can also use this to quantify how abnormal the baby weight is given the mother's weight. This might help us identify if the baby has something wrong with it that is affecting its weight.

5.

One thing I do not know how to do is identify correlation between observations.