Birthweight EDA

Orion Bowers

2024-01-10

Exploratory Data Analysis

1. Summarize, in your own words, the main reasons for analyzing the data. Discuss the goals of the analysis in the context of the problem and why those goals are important

   - The main reason for analyzing this dataset is working out how a baby's birth weight is affected by mother's age, mother's race, gestational length, and the baby's gender. If we can accurately estimate a baby's birth weight, we can be ready with additional health resources if needed. We'd also like to know if there's a significant risk to low birth weight for mothers of a particular age range or race.

2. Summarize the main features and patterns in the data in words and graphics.

```
head(baby)
```

```
##   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
```

```
baby_mage = ggplot(data=baby)+
  geom_point(aes(x=Mage,y=BirthWeight, color=Race))

baby_gage = ggplot(data=baby)+
  geom_point(aes(x=Gage,y=BirthWeight, color=Race))

baby_race = ggplot(data=baby, aes(x=Race,y=BirthWeight))+
  geom_boxplot()

baby_race_count = ggplot(data=baby) +
  geom_bar(aes(x=Race))

baby_mage_vs_gage = ggplot(data=baby)+
  geom_point(aes(x=Gage,y=Mage))
```
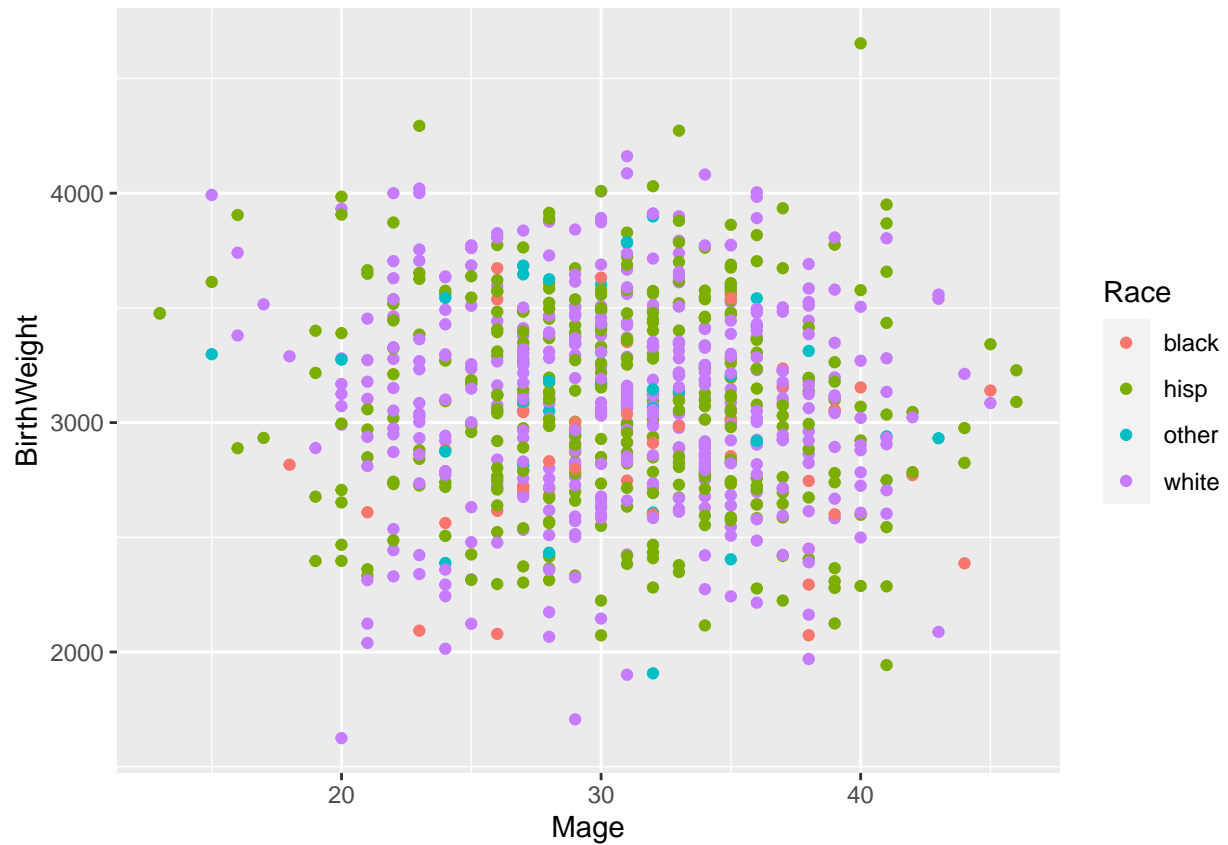
```
baby_mage #That's my next D&D character right there: "Baby Mage"
```

I was looking for anything interesting between Mother's age and birth weight, but there doesn't seem to be anything at first glance. It's difficult to tell between races, but I suspect there's a slight negative trend in Birthweight as Mother's age increases.

```
baby_gage #And if that character dies, their sibling "Baby Gauge" the Barbarian.
```
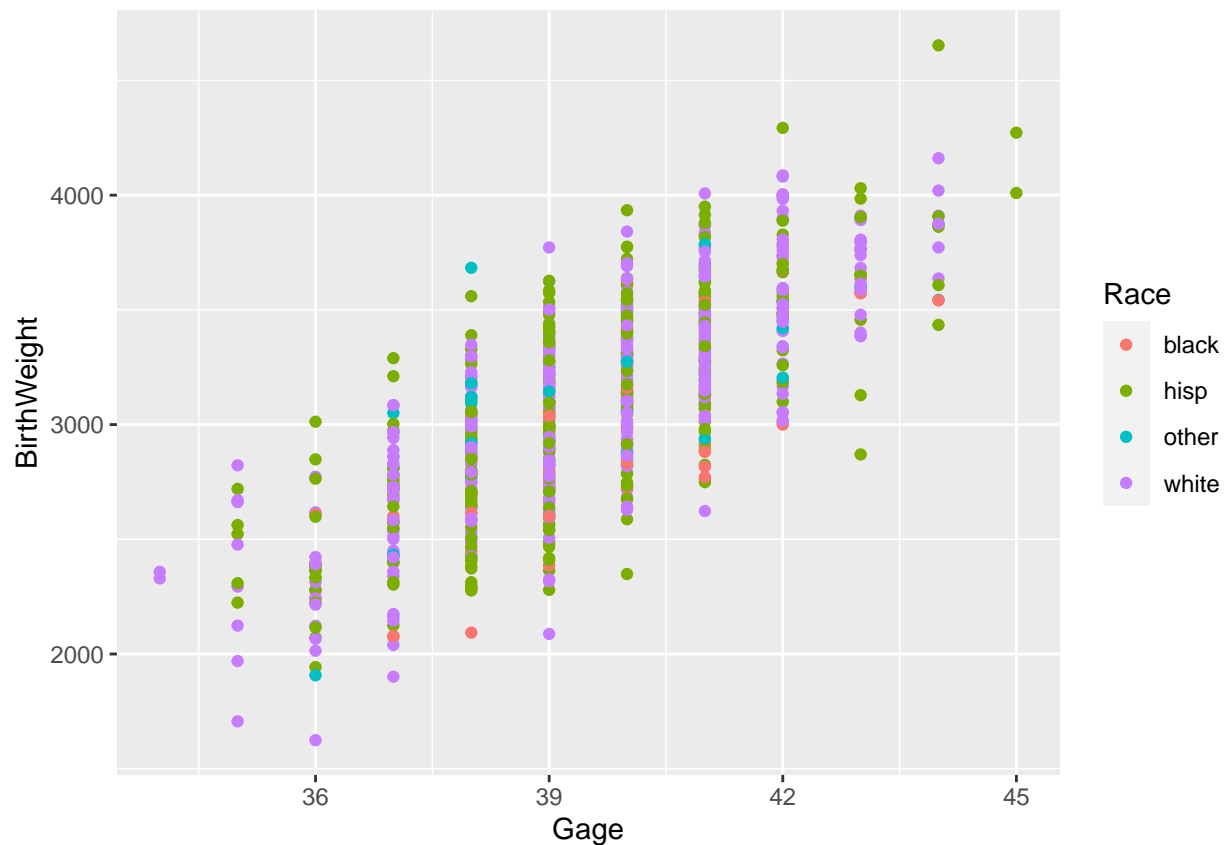
There's definitely a positive correlation between birth weight and gestation age, but that's fairly obvious without the data. For the most part, the longer the baby is in the womb, the bigger it'll be. It looks a little more interesting looking at black babies, but that may be due to the small number of black babies in the dataset.

```
baby_race #An entertaining TV show if I've ever seen one
```
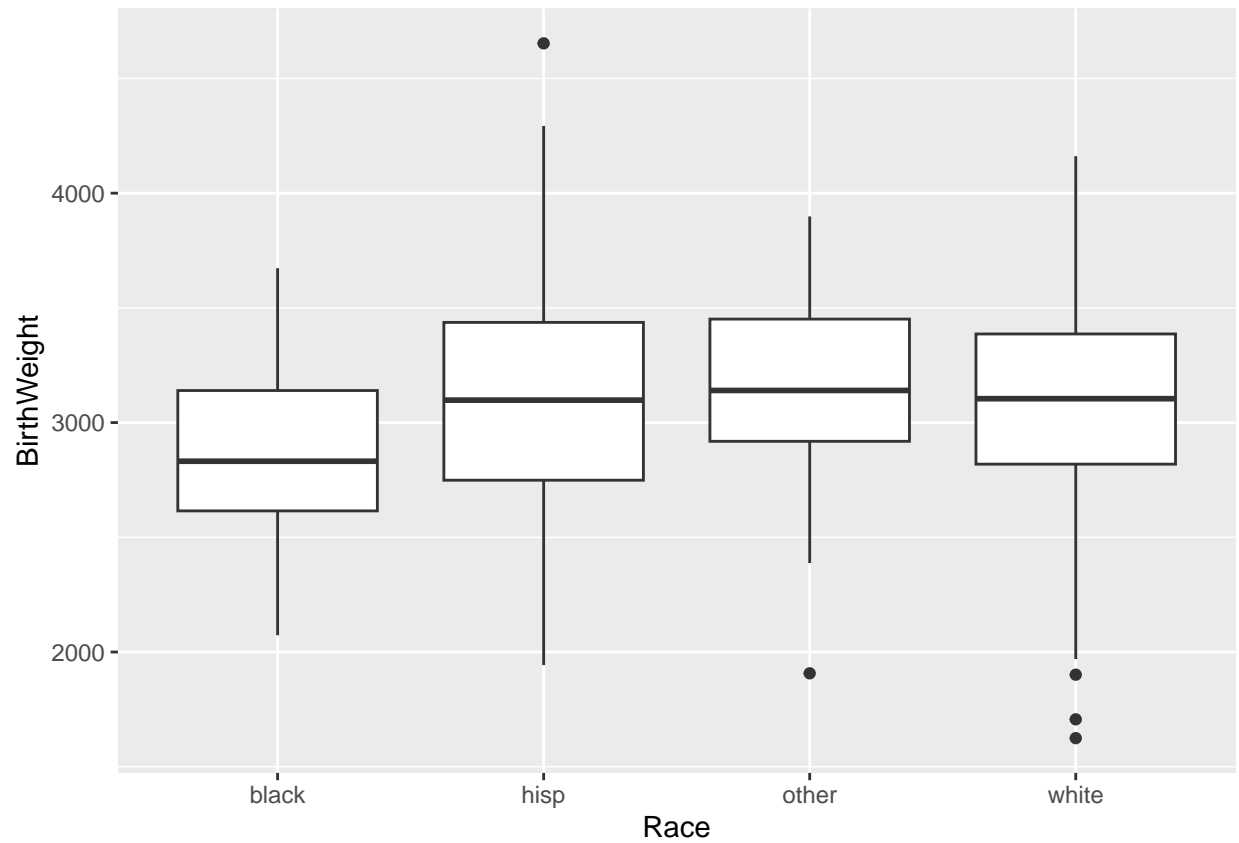
The averages for each race seem fairly similar to each other. Black babies may have a smaller birth weight, but once again, that could be the smaller count we have for that data.

```
baby_race_count #Hopefully these comments are helpful.
```
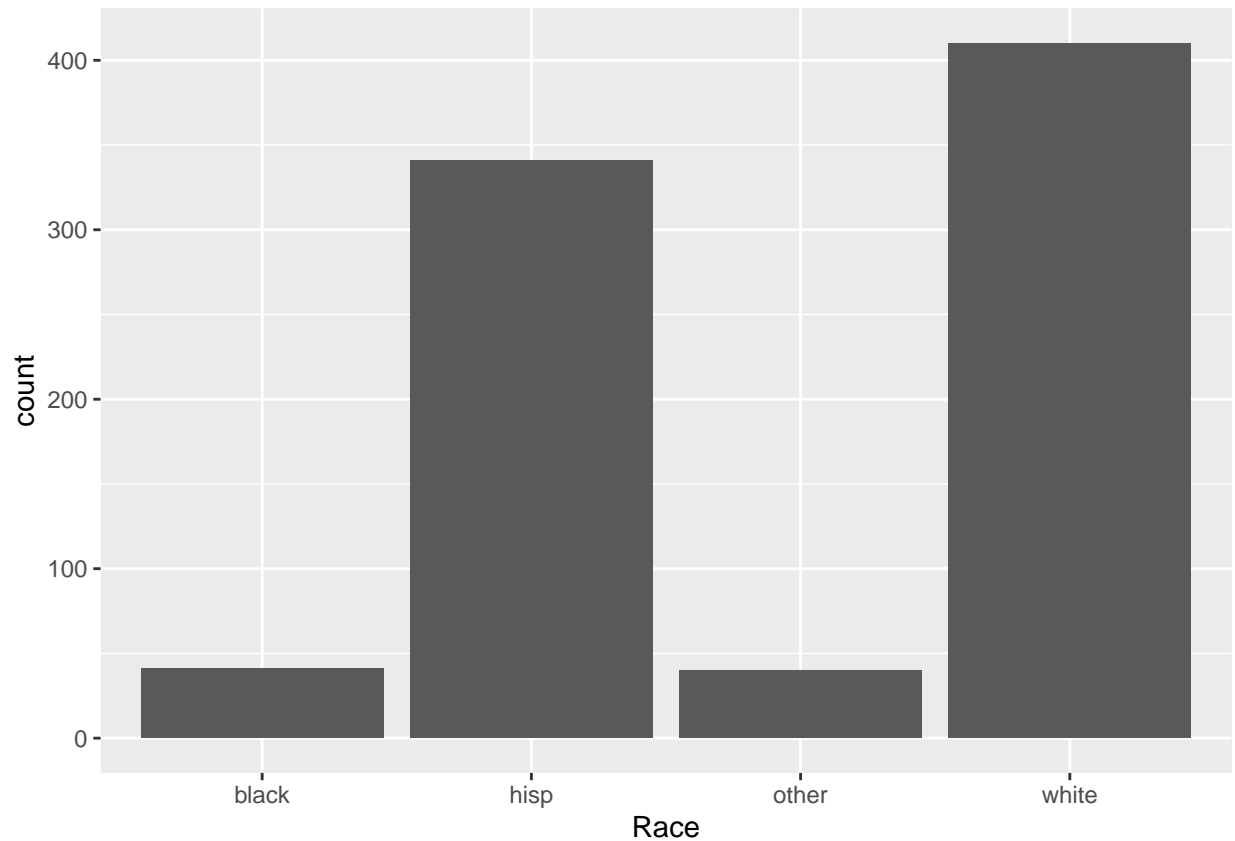
Looking at these makes me really mistrust any trends I see in the data with black babies. There's far fewer observations so I'd be careful making generalizations for just black babies. I do wonder where this data was collected. Western US? Maybe California, Arizona, or Utah?

```
baby_mage_vs_gage #An epic battle after one is raised from the dead.
```
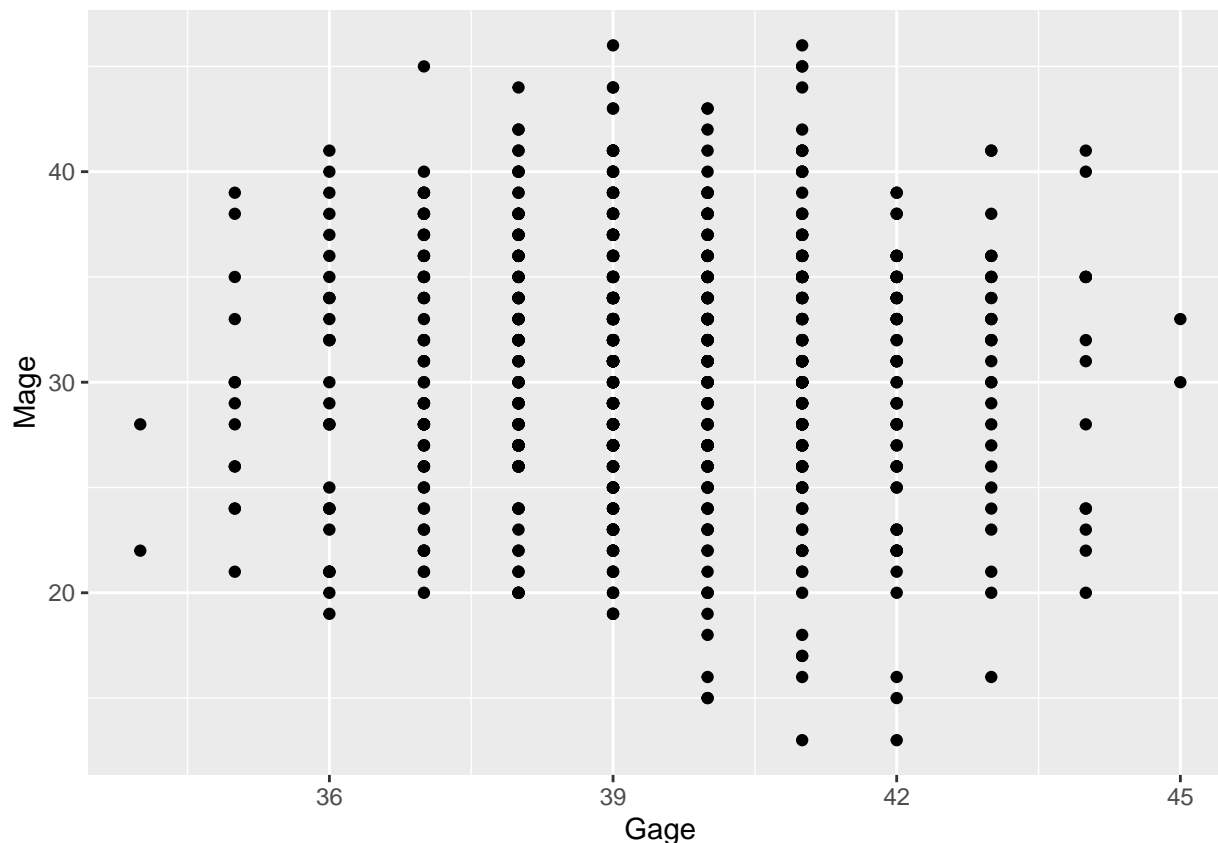
I don't see too much of a relationship between a mother's age and the gestational period (one the questions we're looking to answer). Maybe the model will reveal more, but all I see is a slightly negative relationship.

3. Discuss aspects of the data that would create correlation between observations. If applicable, quantify how much cross-observation correlation is present in the data.

   - I'm not sure what would cause correlations between observations. Indiviudal babies are fairly independent from other babies born. Perhaps if the same mother is recorded multiple times, there could be correlation?

4. Posit an appropriate statistical model that could be applied to analyze the data. Discuss why the proposed statistical method would be useful in achieving the goals mentioned in point #1.

   - I'll be honest, I do not remember much from Stat 330 or 340, so almost every model I remember has been wiped clean. With that in mind, couldn't we just go with simple linear regression? (As long as there is low correlation between observations like I think there is)

5. Identify one aspect of the analysis that you don't know how to do.

   - I'm not sure how to start the analysis. Much of Stat 330 killed me with theory so by the time coding came up, I was copying down code without knowing what it did. I'm hoping this class will help out in that department. I know how to code, but what to code is a different ballgame.