

Data Visualization

Tyler Reed

Getting Set Up

R Packages

```
library(dplyr)
library(ggplot2)
library(readr)
```

The Data

```
nc <- read_csv("https://facweb.gvsu.edu/kapitull/sta631/nc.csv")
```

How to look at data in R

Take a glimpse

1. habit - string or character; visits - double

```
glimpse(nc)
```

```
## Rows: 800
## Columns: 13
## $ fage          <dbl> 19, 21, 18, 17, 20, 30, 21, 14, 16, 20, 18, 20, 20, 26, ~
## $ mage          <dbl> 15, 15, 15, 15, 16, 16, 16, 16, 16, 17, 17, 17, 17, ~
## $ mature        <chr> "younger mom", "younger mom", "younger mom", "younger m~
## $ weeks         <dbl> 37, 41, 37, 35, 37, 45, 38, 40, 24, 40, 37, 40, 39, 38, ~
## $ premie        <chr> "full term", "full term", "full term", "premie", "full ~
## $ visits        <dbl> 11, 6, 12, 5, 13, 9, 15, 12, 5, 8, 10, 17, 9, 11, 10, 1~
## $ marital       <chr> "married", "married", "married", "married", "married", ~
## $ gained        <dbl> 38, 34, 76, 15, 52, 28, 75, 9, 12, 20, 39, 38, 36, 30, ~
## $ weight        <dbl> 6.63, 8.00, 8.44, 4.69, 6.94, 7.44, 7.56, 5.81, 1.50, 8~
## $ lowbirthweight <chr> "not low", "not low", "not low", "low", "not low", "not~
## $ gender        <chr> "female", "male", "male", "male", "female", "male", "fe~
## $ habit         <chr> "nonsmoker", "nonsmoker", "nonsmoker", "nonsmoker", "no~
## $ whitemom      <chr> "white", "white", "not white", "not white", "white", "w~
```

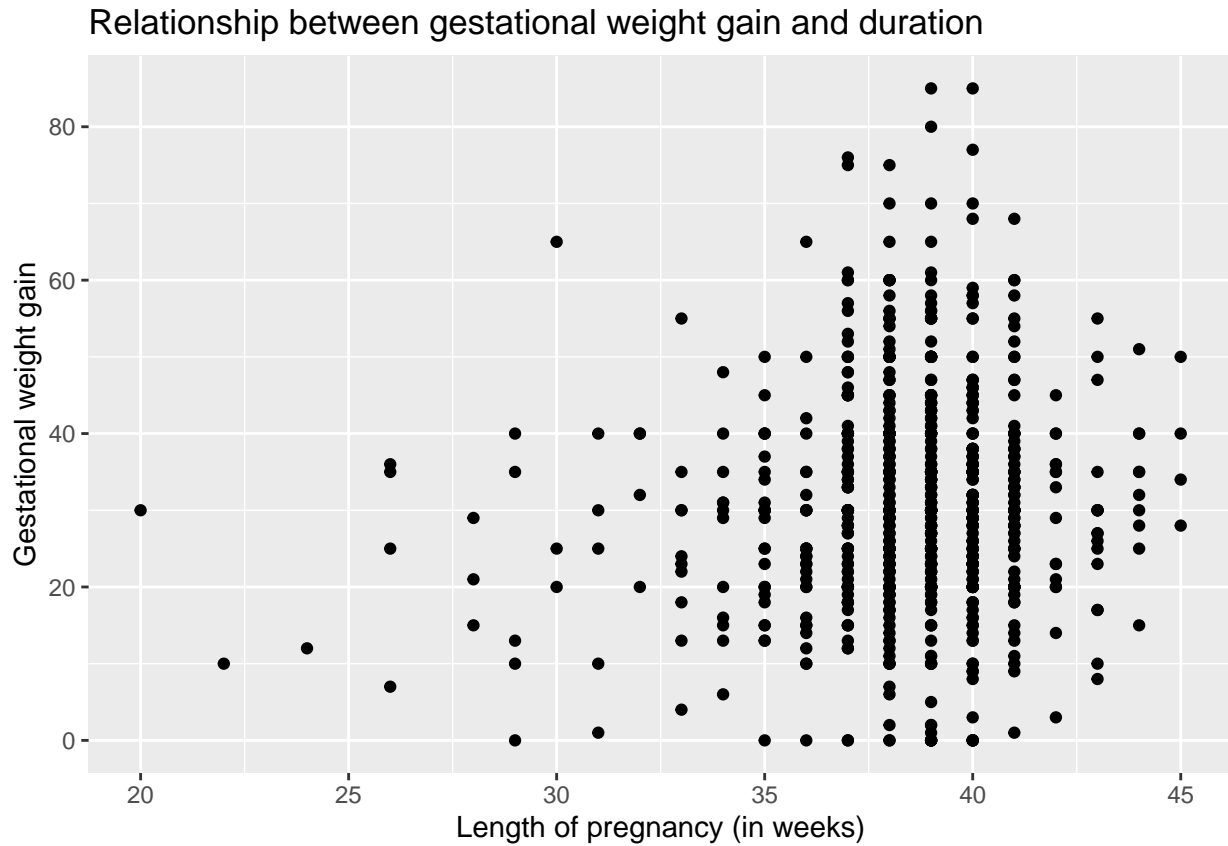
Types of Graphs

Scatterplots

1. positive relationship between the variables
- 2.

```
ggplot(data = nc, aes(x = weeks, y = gained)) +
  geom_point() +
```

```
labs(x = "Length of pregnancy (in weeks)", y = "Gestational weight gain",
     title = "Relationship between gestational weight gain and duration")
```

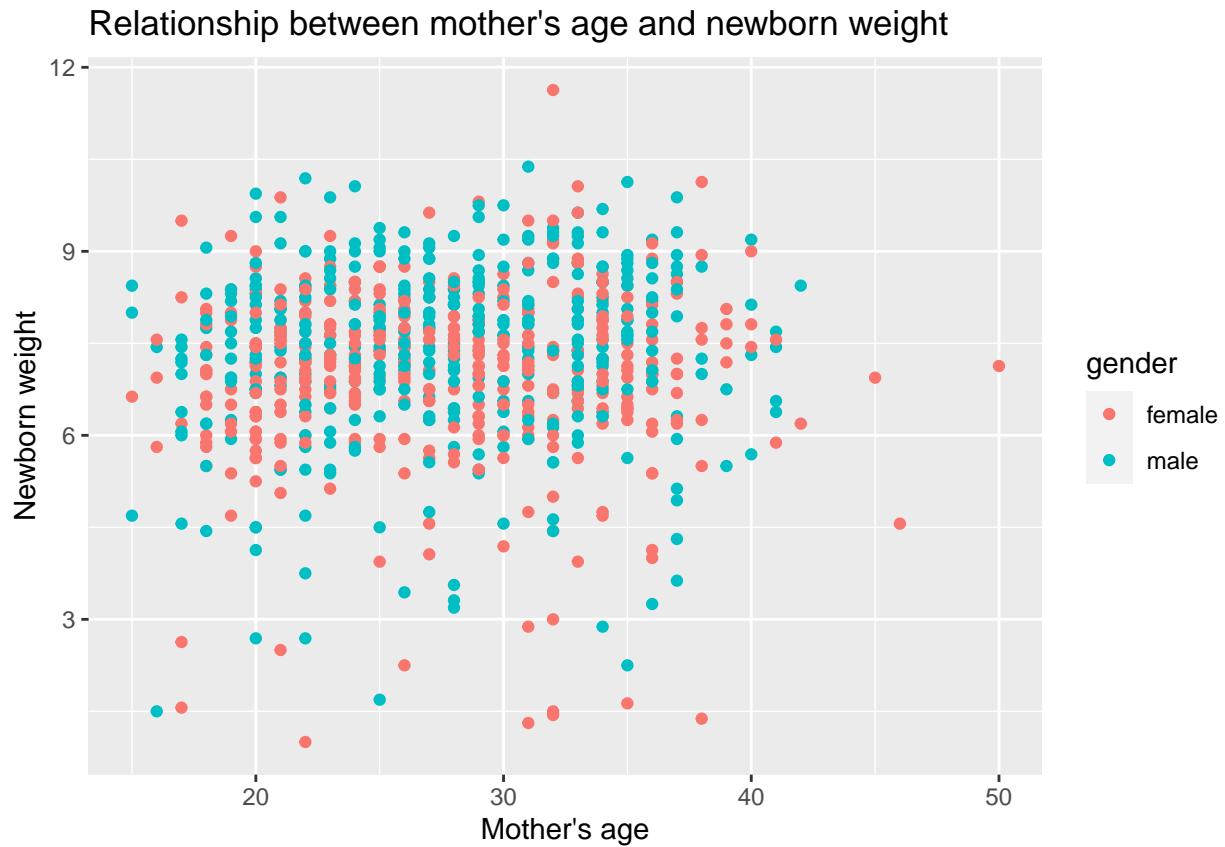


1. A. The color statement classified or grouped cases by a categorical variable, `premie`.

B. 3

C. ~36 weeks

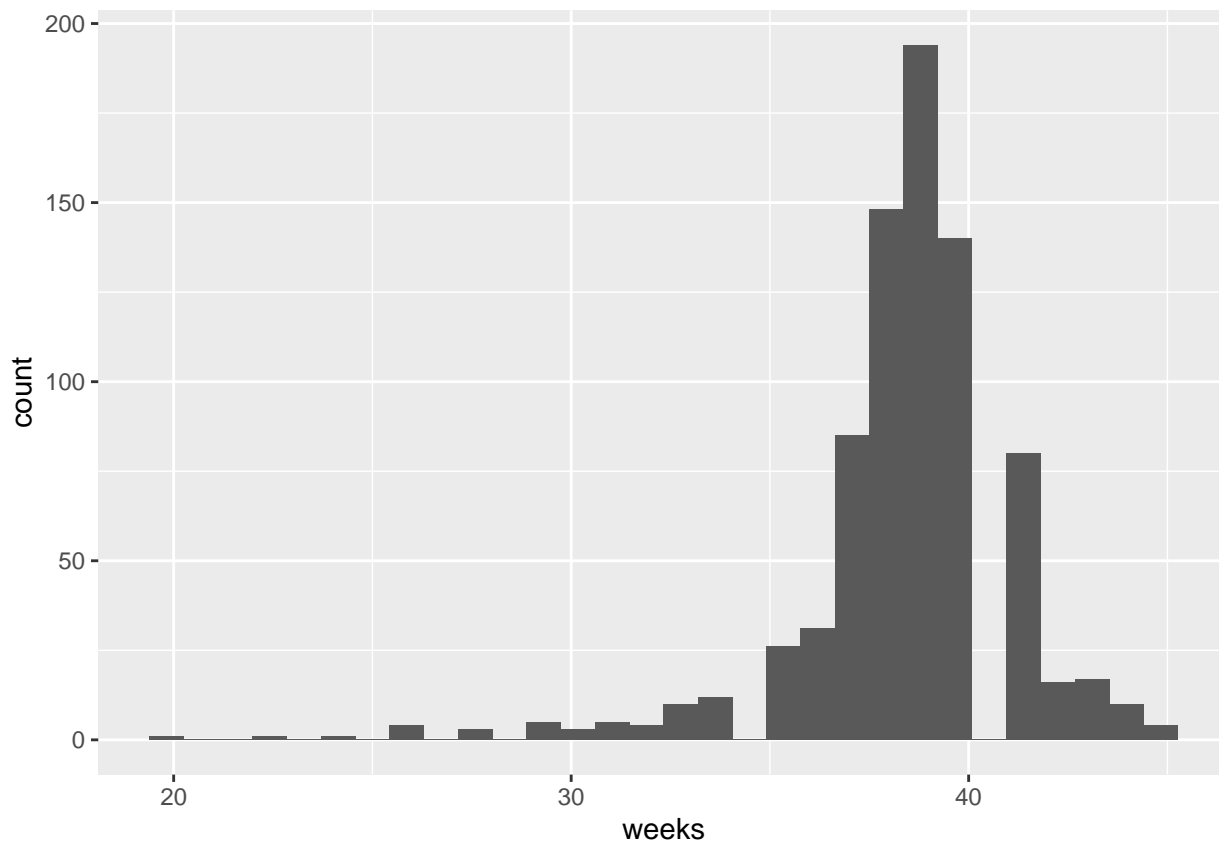
```
ggplot(data = nc, aes(x = mage, y = weight, color = gender)) +
  geom_point() +
  labs(x = "Mother's age", y = "Newborn weight",
       title = "Relationship between mother's age and newborn weight")
```



1. There appears to be no relationship between `mage` and `weight` as the distribution of `weight` remains similar as `mage` increases.

Histograms

```
ggplot(data = nc, aes(x = weeks))+  
  geom_histogram()
```



1.

A. **count** refers to the number of cases which fell within the binned values, in this case, the number of babies born at each interval of **weeks** of gestation.

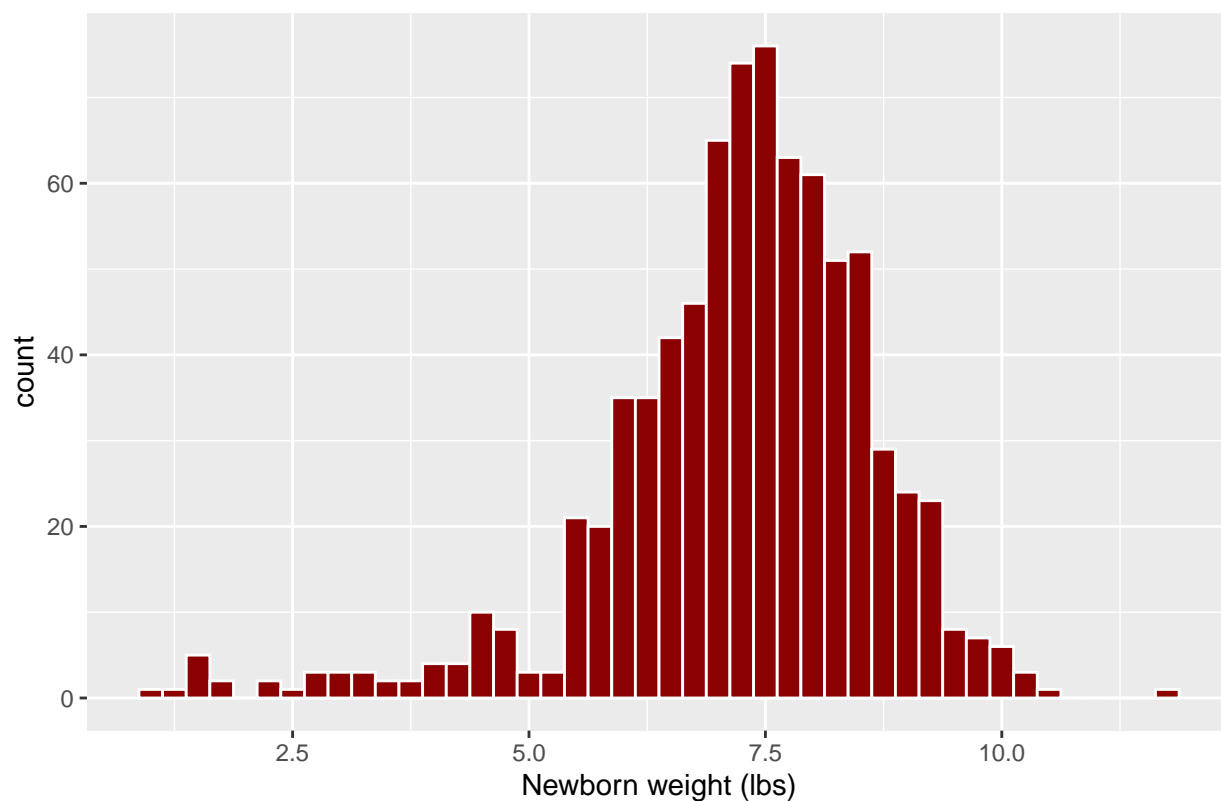
B. ~38 weeks

C. There would be one bin and it would include all cases, ~800, because every birth would fall within the set parameter whether it was at 15 **weeks** or 48 **weeks**.

2.

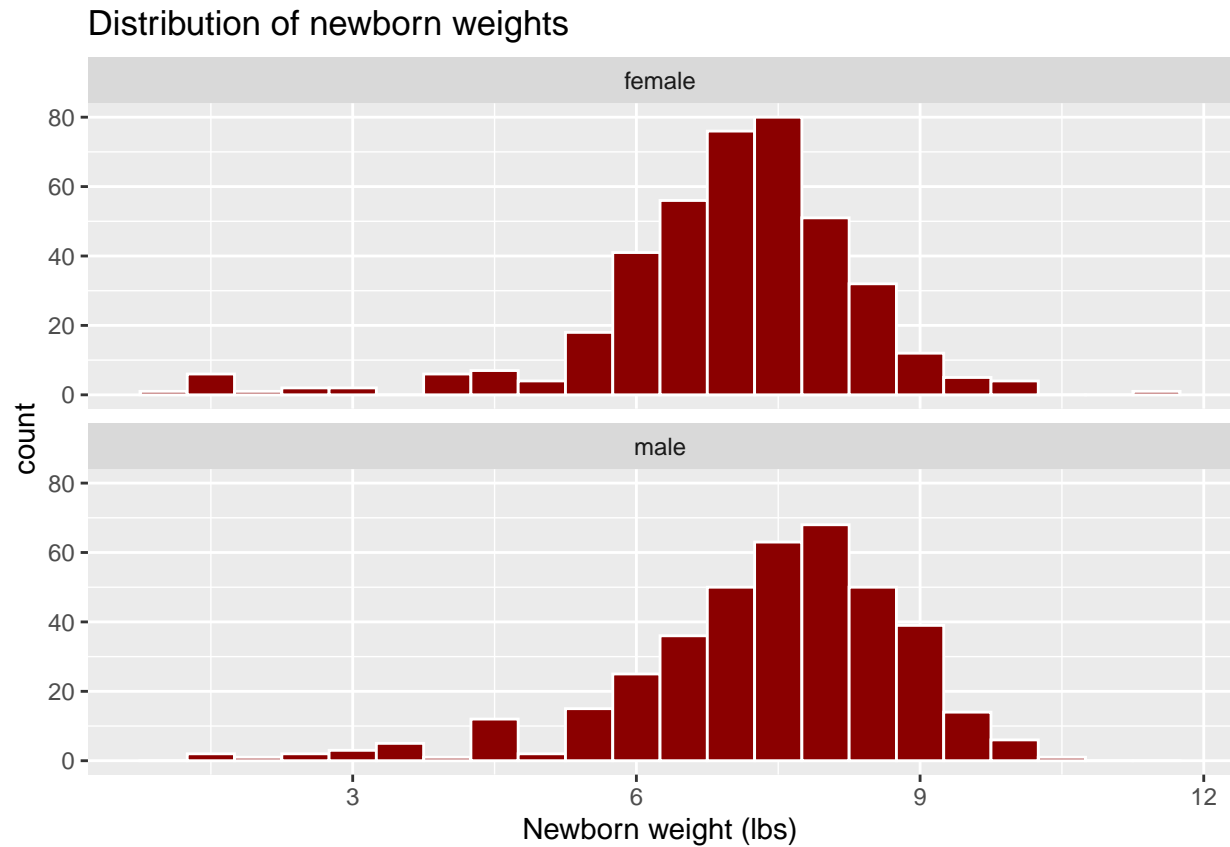
```
ggplot(data = nc, aes(x = weight))+
  geom_histogram(binwidth = 0.25, color = "white", fill = "dark red") +
  labs(x = "Newborn weight (lbs)",
       title = "Distribution of newborn weights")
```

Distribution of newborn weights



Faceting

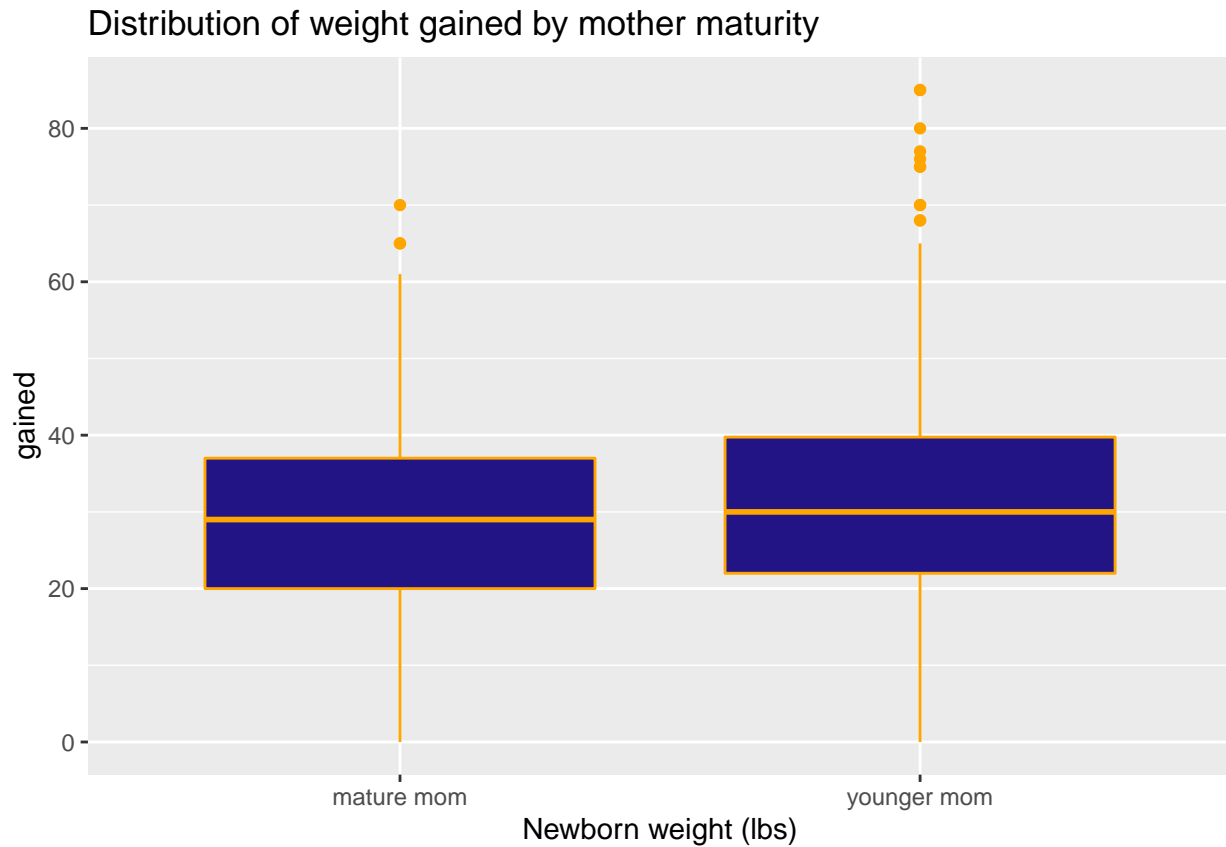
```
ggplot(data = nc, aes(x = weight))+  
  geom_histogram(binwidth = 0.5, color = "white", fill = "dark red") +  
  facet_wrap(~ gender, ncol = 1) +  
  labs(x = "Newborn weight (lbs)",  
       title = "Distribution of newborn weights")
```



1. Males appear to have a slightly larger average birth weight.

Boxplots

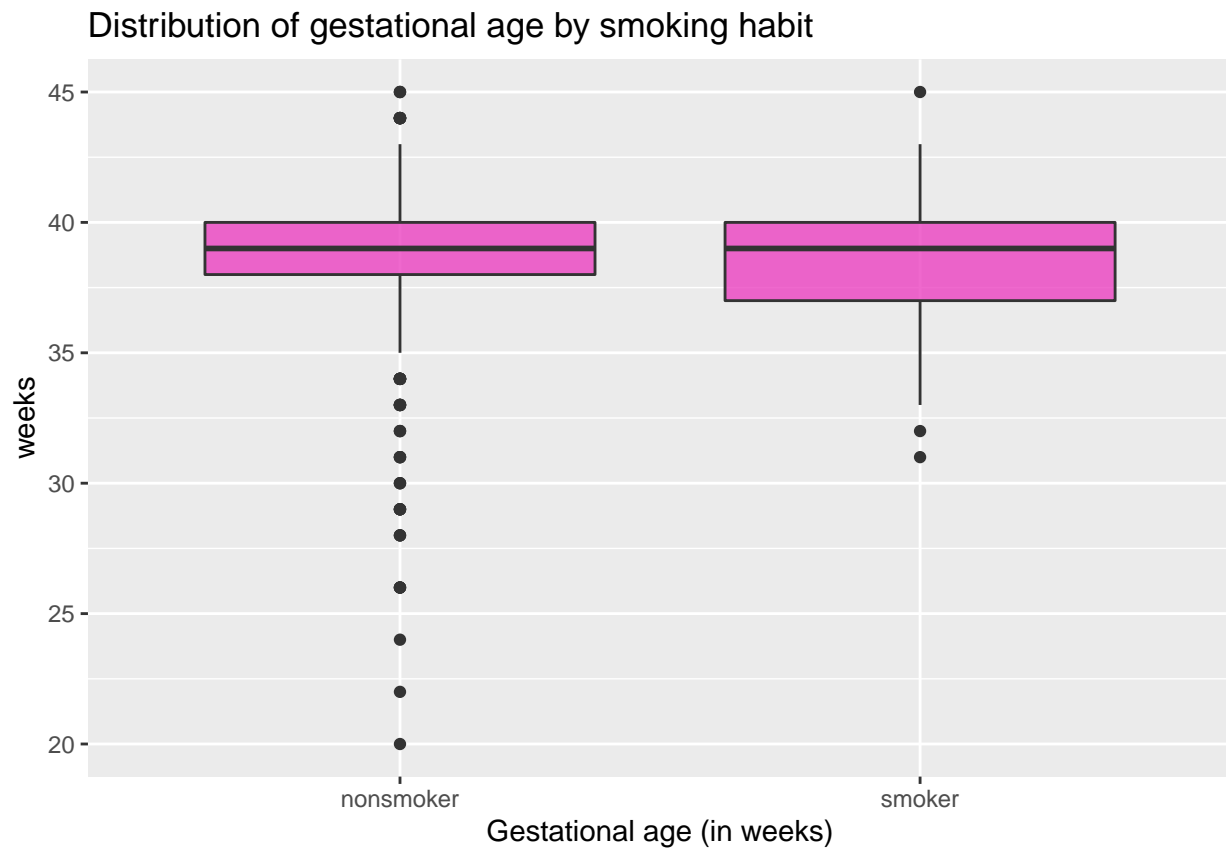
```
ggplot(data = nc, aes(x = mature, y = gained)) +  
  geom_boxplot(color = "orange", fill = "#231485") +  
  labs(x = "Newborn weight (lbs)",  
       title = "Distribution of weight gained by mother maturity")
```



8. It appears the median weight gained for younger moms is slightly higher than mature moms.

9. Smokers appear to have greater variation in gestational age, though far more outliers are present amongst non-smokers.

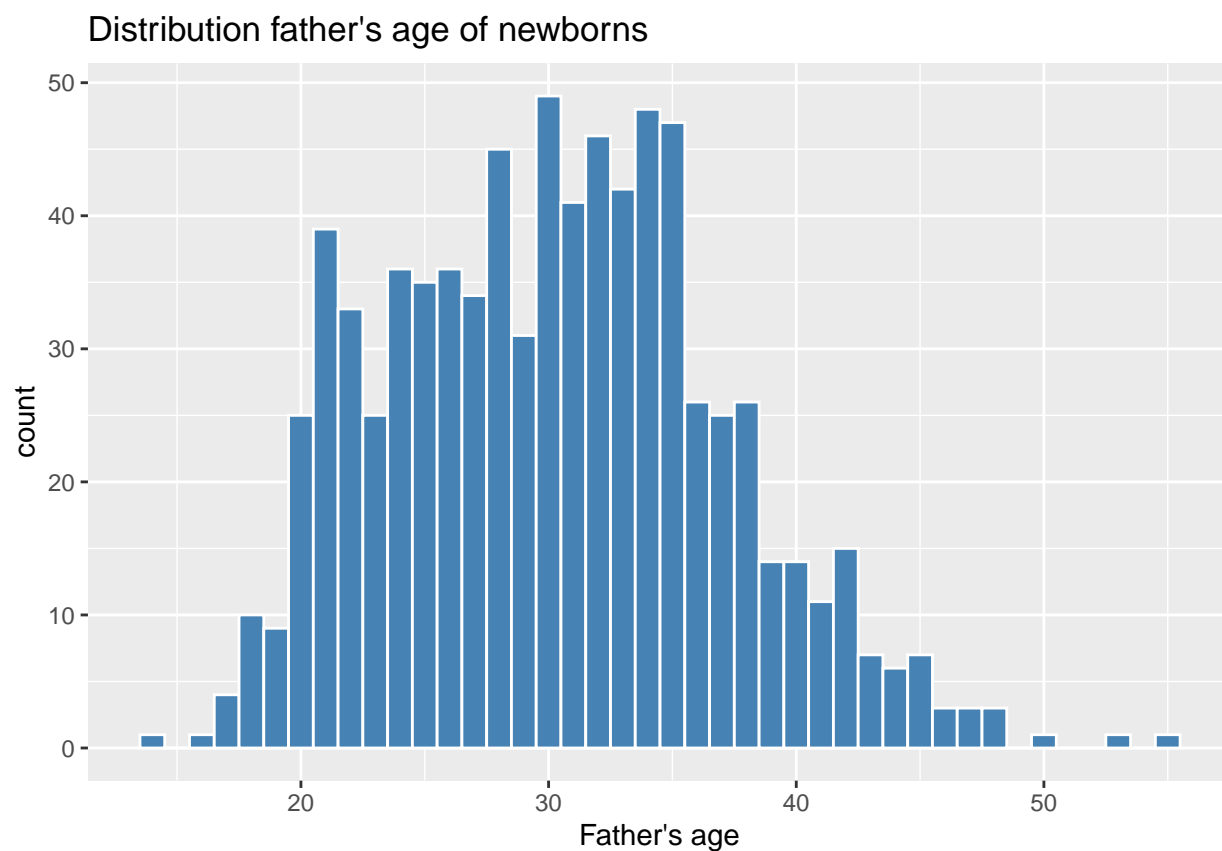
```
ggplot(data = nc, aes(x = habit, y = weeks)) +  
  geom_boxplot(fill = "#EB3BBFC4") +  
  labs(x = "Gestational age (in weeks)",  
       title = "Distribution of gestational age by smoking habit")
```



Independent Practice

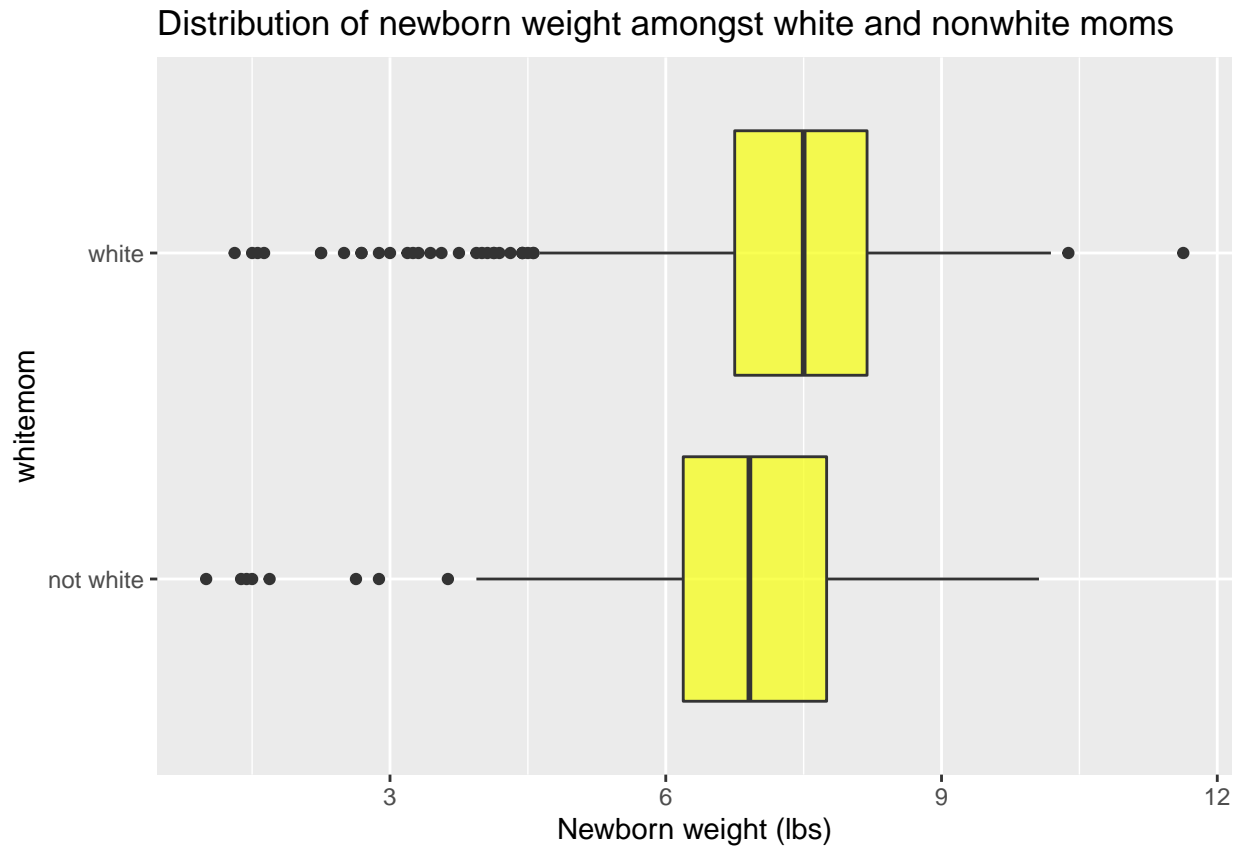
1. The distribution appears to slightly skew right, but is mostly symmetrical.

```
ggplot(data = nc, aes(x = fage))+
  geom_histogram(binwidth = 1, color = "white", fill = "steelblue") +
  labs(x = "Father's age",
       title = "Distribution father's age of newborns")
```

2. Median weight of white newborns is greater than nonwhite.

```
ggplot(data = nc, aes(x = weight, y = whitemom)) +  
  geom_boxplot(fill = "#F7FF12B8") +  
  labs(x = "Newborn weight (lbs)",  
       title = "Distribution of newborn weight amongst white and nonwhite moms")
```



3. No. There appears to be no relationship between the mother's age and the gestational age.

```
ggplot(data = nc, aes(x = mage, y = weeks)) +
  geom_point(color = "blue") +
  labs(x = "Mother's age", y = "Gestational age (in weeks)",
       title = "Relationship between mother's age and gestational age")
```

Relationship between mother's age and gestational age

