

Lab 5.1: Heart Rate

JAKE JAMES

February 26, 2021

Introduction

Consider the dataset `heart_disease` from the `funModeling` package.

```
##   age gender chest_pain resting_blood_pressure serum_cholesterol
## 1  63   male         1                145         233
## 2  67   male         4                160         286
## 3  67   male         4                120         229
## 4  37   male         3                130         250
## 5  41  female         2                130         204
##   fasting_blood_sugar resting_electro max_heart_rate exer_angina oldpeak slope
## 1             1         2             150         0         2.3     3
## 2             0         2             108         1         1.5     2
## 3             0         2             129         1         2.6     2
## 4             0         0             187         0         3.5     3
## 5             0         2             172         0         1.4     1
##   num_vessels_flour thal heart_disease_severity exer_angina has_heart_disease
## 1             0     6                0         0         no
## 2             3     3                2         1         yes
## 3             2     7                1         1         yes
## 4             0     3                0         0         no
## 5             0     3                0         0         no
```

There are variables related to patient clinic trial. `heart_disease` is a data frame with 303 rows and 16 variables. We'll focus on the following variables in the analysis:

- `age`: age in years (numerical)
- `max_heart_rate`: max heart rate per minute (numerical)
- `thal`: A blood disorder called thalassemia (categorical: 3 = normal; 6 = fixed defect; 7 = reversible defect)
- `has_heart_disease`: Heart disease (categorical: no, yes)
- `gender`: gender of patient (categorical: male, female)

The purpose of this lab is to practice the creative process in exploratory data analysis of asking questions and then investigating those questions using visuals and statistical summaries. It is your job to apply your detective skills to the information hidden in this data. For future use, utilize the modified dataset `heart` according to the R code below:

```
heart=as_tibble(heart_disease) %>%
  select(age, max_heart_rate, thal, has_heart_disease, gender)
head(heart)
```

```
## # A tibble: 6 x 5
##   age max_heart_rate thal has_heart_disease gender
##   <int>         <int> <fct>   <fct>      <fct>
## 1   63           150 6     no         male
## 2   67           108 3     yes        male
## 3   67           129 7     yes        male
## 4   37           187 3     no         male
## 5   41           172 3     no         female
## 6   56           178 3     no         male
```

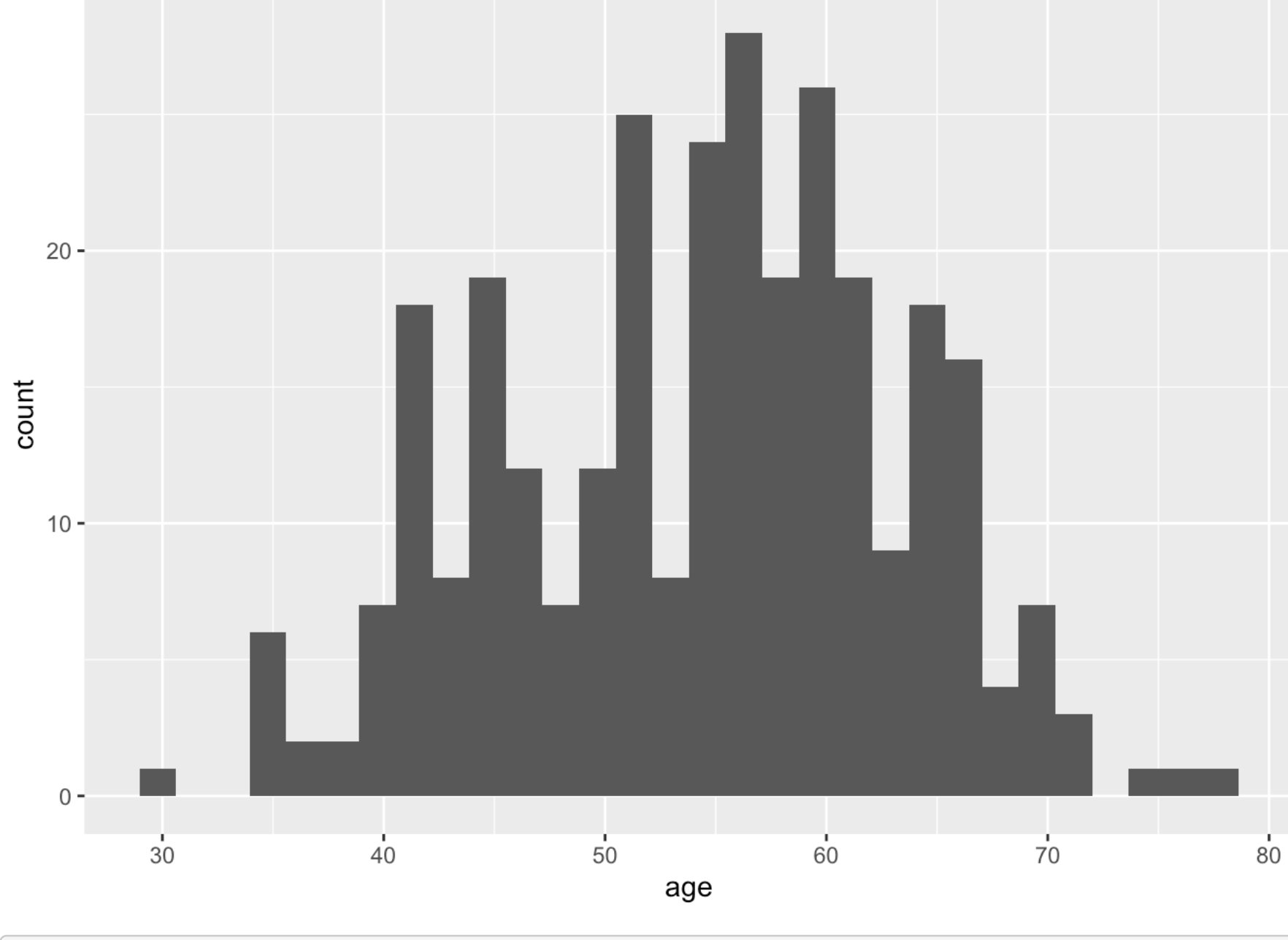
Part 1: Questions About Variation

Question 1: What is the most common age found in the data?

```
## Shows age distribution in the dataset.
```

```
ggplot(heart) +
  geom_histogram(aes(x=age))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
heart %>%
  group_by(age) %>%
  summarise(n=n(), .groups='drop') %>%
  arrange(desc(n)) %>%
  summarise(common.exp=first(age), common.n=first(n), .groups='drop')
```

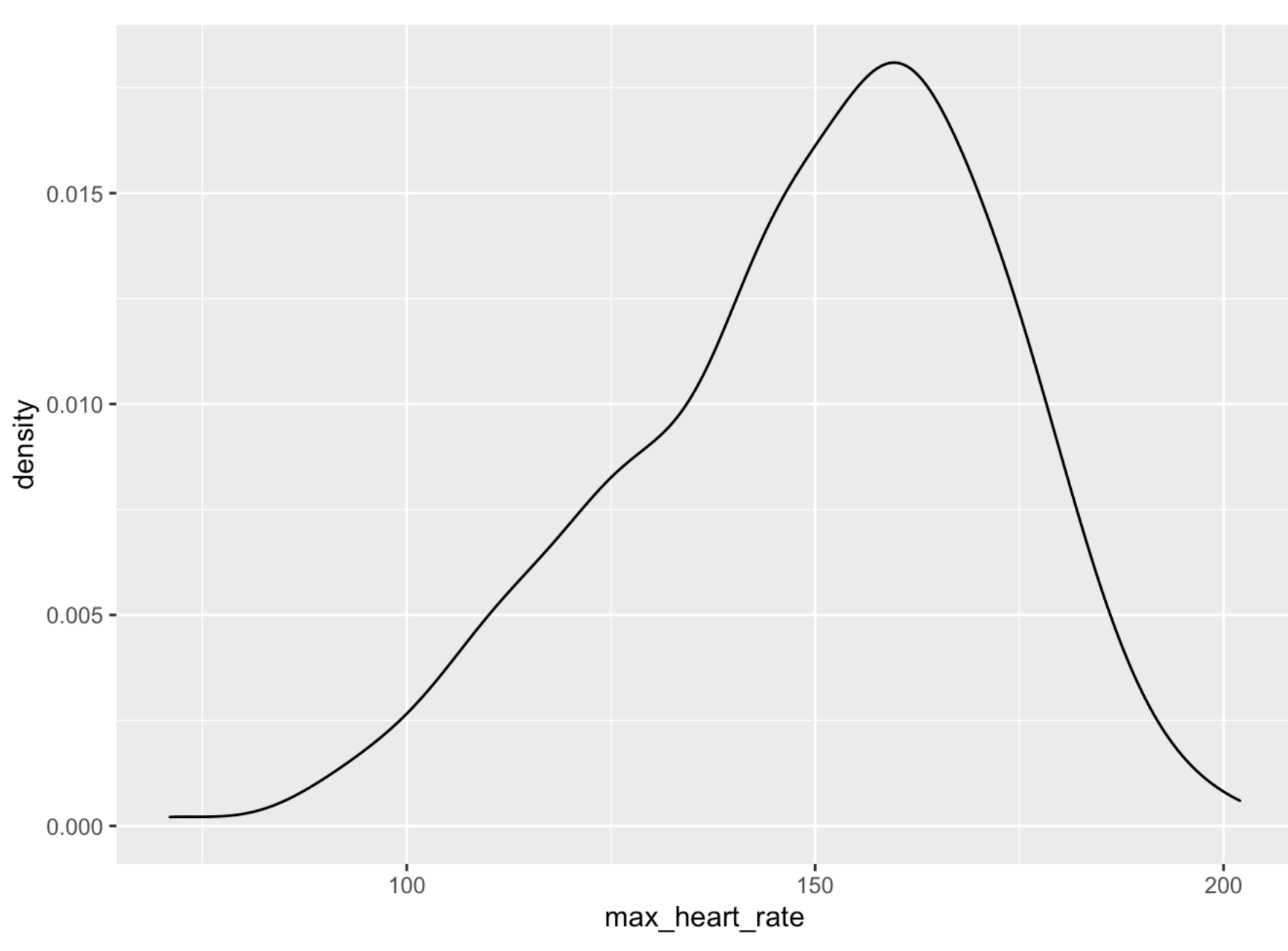
```
## # A tibble: 1 x 2
##   common.exp common.n
##   <int>     <int>
## 1     58         19
```

The most common value for age is 58 and the number of patients of the age is 19.

Question 2: What is the maximum value of max heart rate found in the data?

```
## Shows max_heart_rate distribution in the dataset.
```

```
ggplot(heart) +
  geom_density(aes(x=max_heart_rate))
```



```
heart %>%
  group_by(max_heart_rate) %>%
  summarise(n=n(), .groups='drop') %>%
  arrange(desc(max_heart_rate)) %>%
  summarise(max.max_heart_rate=first(max_heart_rate),
            max.n=first(n), .groups='drop')
```

```
## # A tibble: 1 x 2
##   max.max_heart_rate max.n
##   <int>     <int>
## 1         202         1
```

The maximum `max_heart_rate` was 202 which occurred 1 time in our sample.

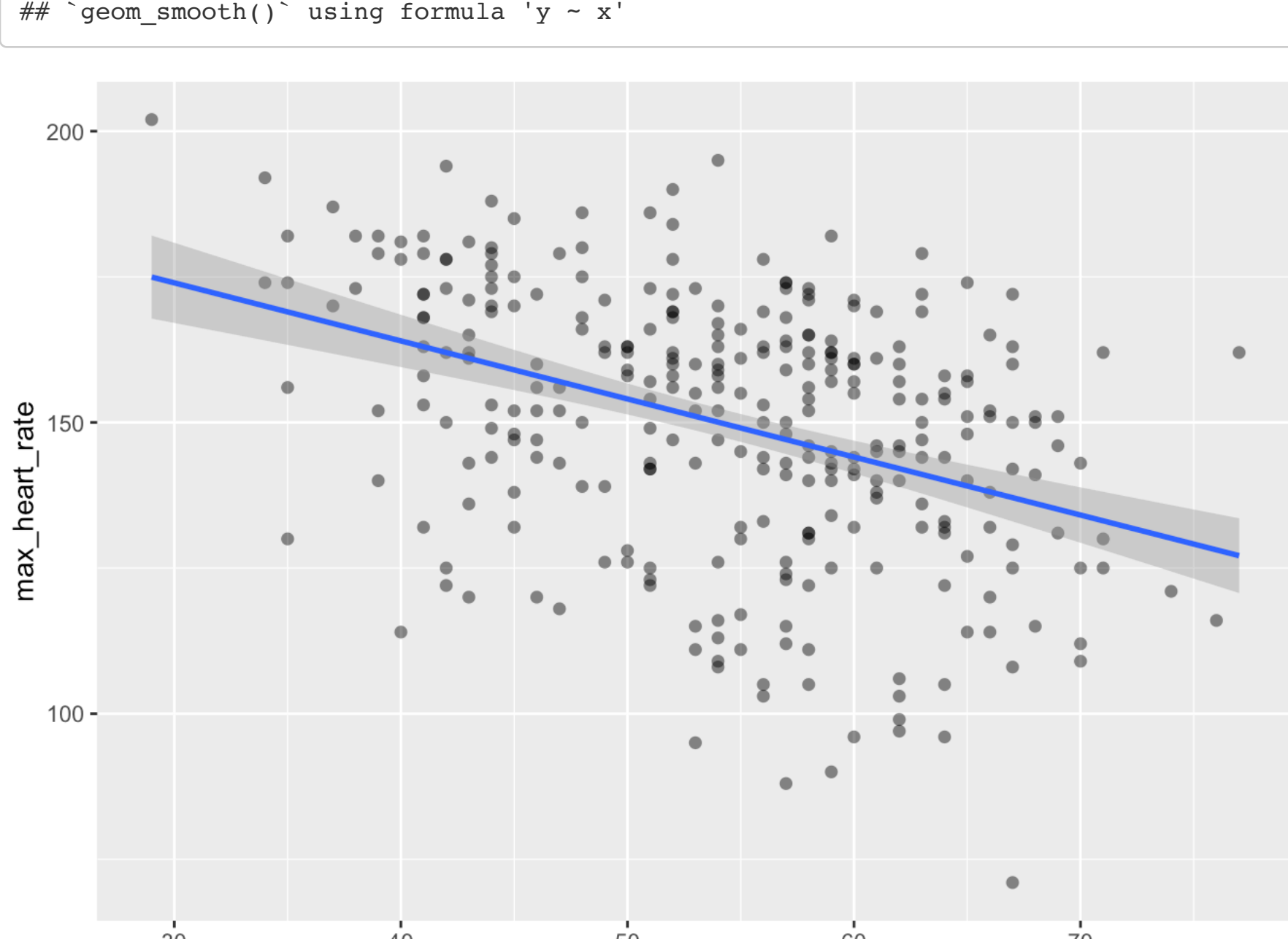
Part 2: Questions about Covariation

Question 3: Is there a relationship between age and max heart rate?

```
## Displays a linear regression line to show the relationship between age and max heart rate.
```

```
ggplot(heart) +
  geom_point(aes(x=age, y=max_heart_rate),
            alpha=0.5, shape=16, size=2) +
  geom_smooth(aes(x=age, y=max_heart_rate), method='lm')
```

```
## `geom_smooth()` using formula 'y ~ x'
```



The max heart rate seems to decrease as the age of patients increases.

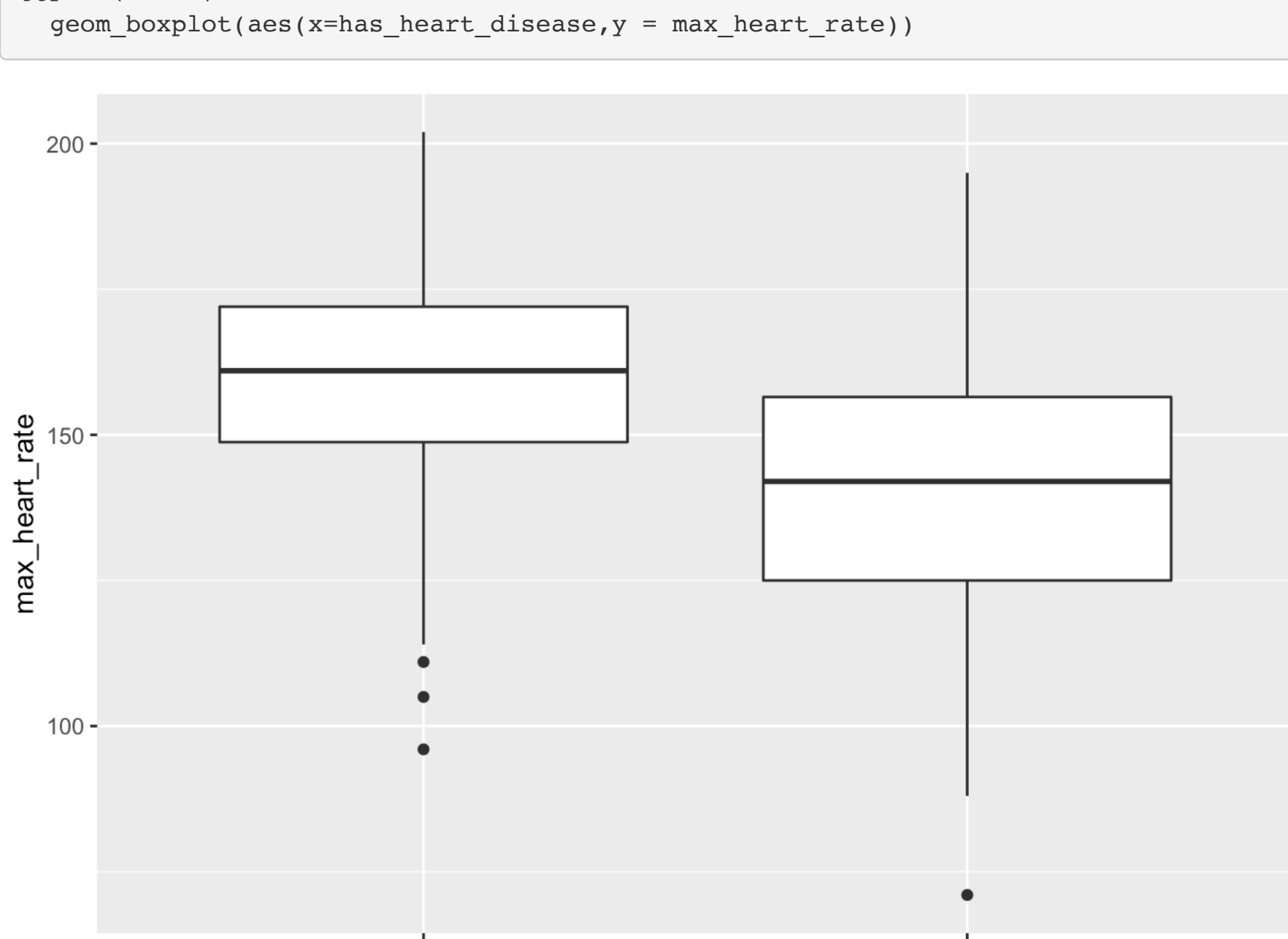
- g. Practically, what reasons do you hypothesize for this observed relationship?

As patients get older, their heart rate deteriorates is worsening conditions.

Question 4: How does max heart rate differ between have heart disease and not have?

```
## Uses `a` boxplot to compare the distribution of max heart rate of patients who have heart disease to the distribution of max heart rate of patients who do not have heart disease.
```

```
ggplot(heart)+
  geom_boxplot(aes(x=has_heart_disease, y = max_heart_rate))
```



```
## Creates a confidence interval comparing the max heart rate of those who have heart disease and those who don't
```

```
heart %>%
  group_by(has_heart_disease) %>%
  summarise(n=n(), mean=mean(max_heart_rate), se=sd(max_heart_rate)/sqrt(n),
            lb=mean-2*se, ub=mean+2*se, .groups='drop')
```

```
## # A tibble: 2 x 6
##   has_heart_disease n mean se lb ub
##   <fct>         <int> <dbl> <dbl> <dbl> <dbl>
## 1 no             164 158. 1.50 155. 161.
## 2 yes           139 139. 1.92 135. 143.
```

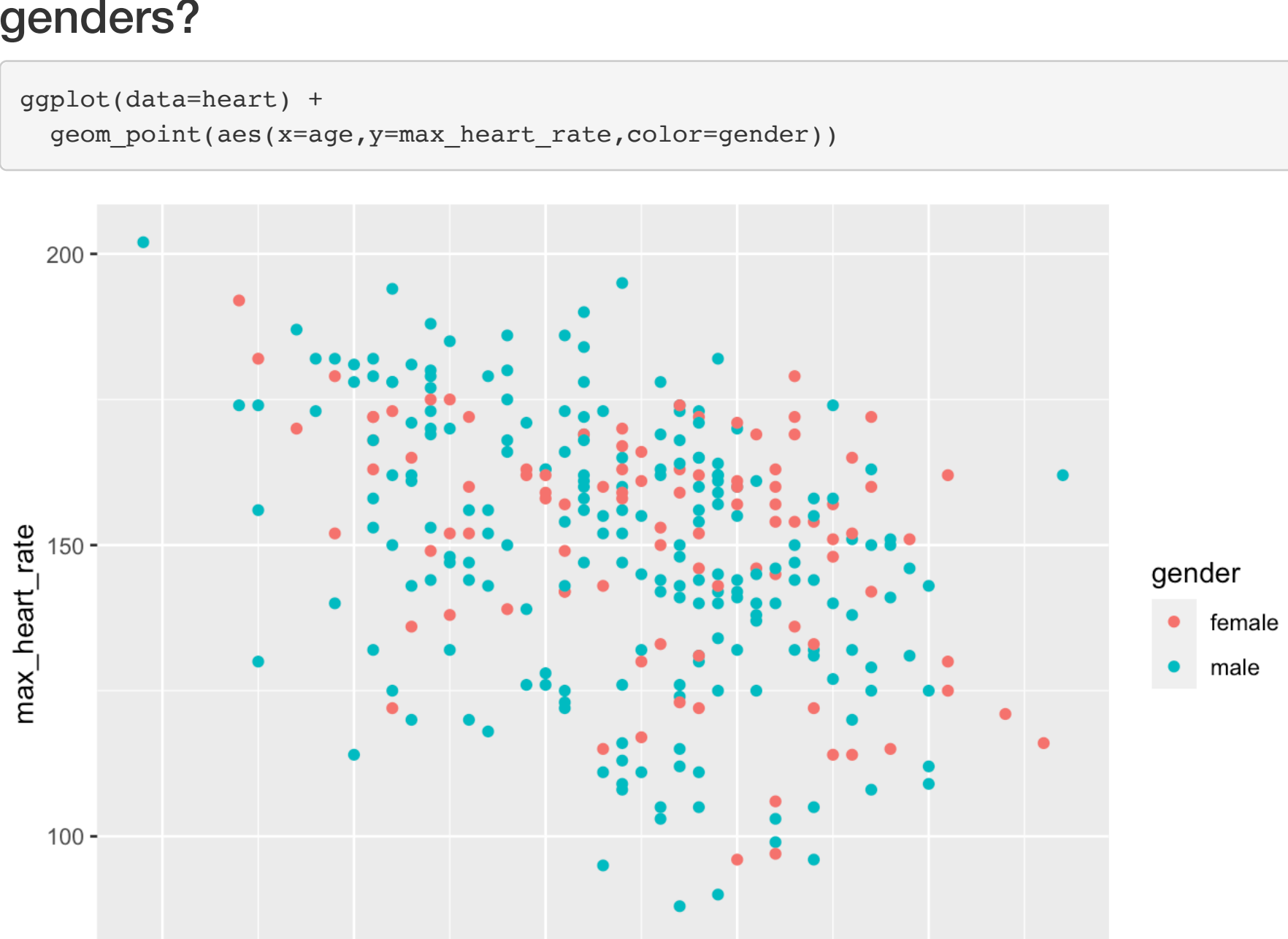
Based on the confidence limits, we have statistical evidence to say that the average `max_heart_rate` for patients who do not have heart disease was larger than the average `max_heart_rate` for patients who have heart disease. yes (yes/no).

- k. How would you explain your answer in terms of the confidence intervals that are constructed above?

We are 95% confident that we have statistically significant evidence to say that the average maximum heart rate of a patient with heart disease will be lower than a patient without heart disease.

Question 5: Does the relationship between age and max heart rate differ between genders?

```
ggplot(data=heart) +
  geom_point(aes(x=age, y=max_heart_rate, color=gender))
```



No.