Final Project

Ingrid J. Lu 260773949;

28/10/2021

```r
# Set-up
df_fh <- readr::read_csv(here::here("fetal_health.csv")) %>%
  select(!starts_with("histogram"))
```

Background

Your sister and her partner are expecting a child soon, and she just went to her obstetrician for her routine check. Because she in her second trimester, her obstetrician asks her to do a fetal cardiogram. The results will not get back to her until a week later. Your sister is a bit of a hypochondriac, so she is afraid that there is something wrong with her child.

So, her partner turns to you, a Masters student who is studying inferential statistics for health, hoping to find some results to calm your sister down.

Luckily, you have identified a publicly available dataset that includes thousands of fetal cardiogram results, and the classification of these babies' health status. You need to convince your sister that she and her child will be safe.

Note: the objective of this exercise is to consolidate all the important concepts covered in EPIB607. When answer each question, be sure to include any units and assumptions and define all parameters, when appropriate. The following questions are based on the publicly available dataset "Fetal Classification", please find all attribute information of the data from the link.

Question 1 Data Visualization and Summary Statistics

a)

Is this data tidy? If no, transform it into an untidy data. If yes, provide an explanation.

b)

Looking at the 3 different classification of fetal health status and each fetus' baseline heart rate provide an appropriate graphic summarizing the distribution of each of baseline heart rates for each class. Be sure to provide the correct title and label for the plot.

```r
# Question 1

# 1.b
df_fh$fetal_health <- ifelse(df_fh$fetal_health == 1, "Normal", ifelse(df_fh$fetal_health == 2, "Suspec
```

```
df_fh %>% group_by(fetal_health) %>%
  ggplot(aes(x = fetal_health, y = `baseline value`)) +
  geom_boxplot()
```
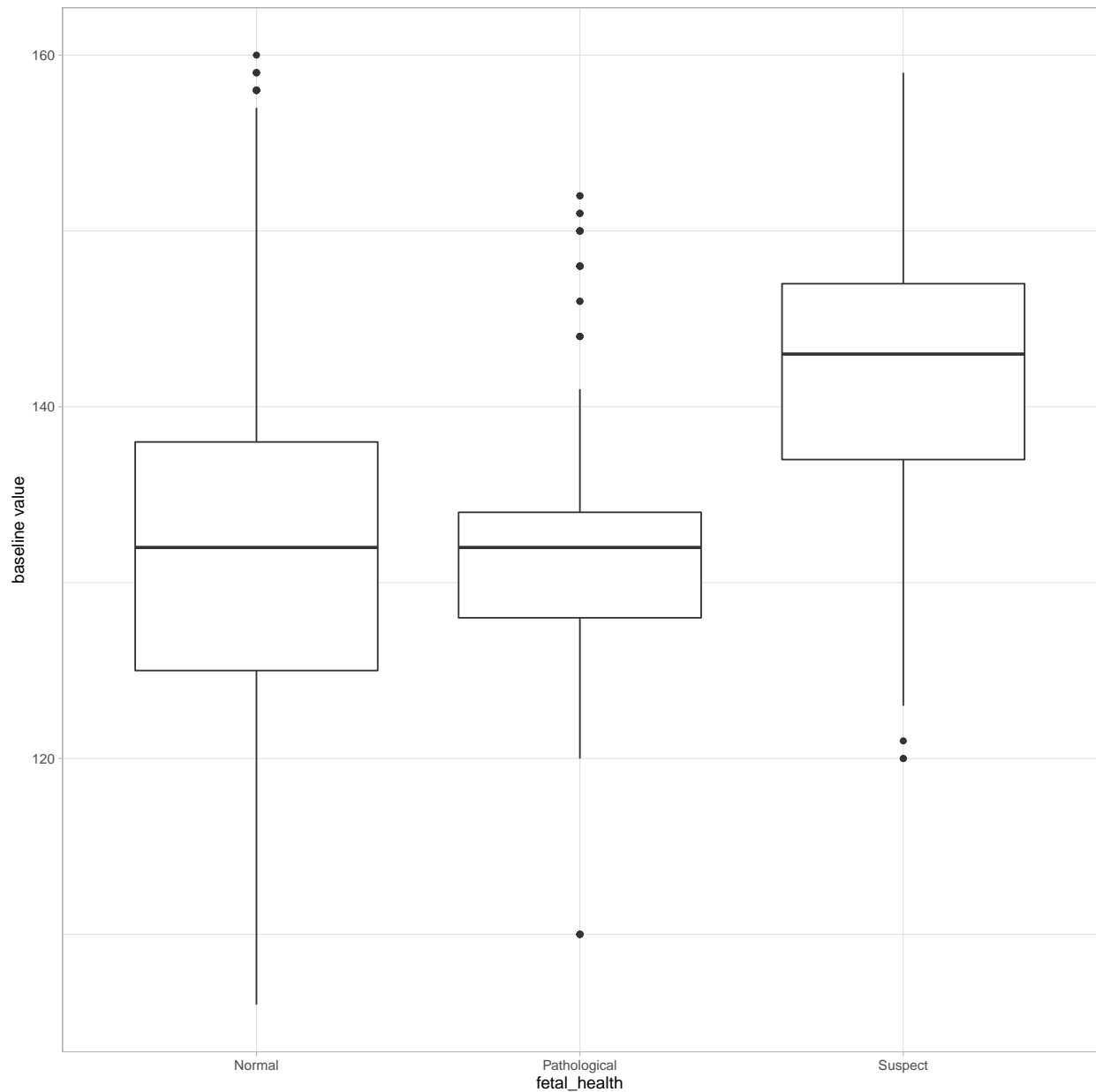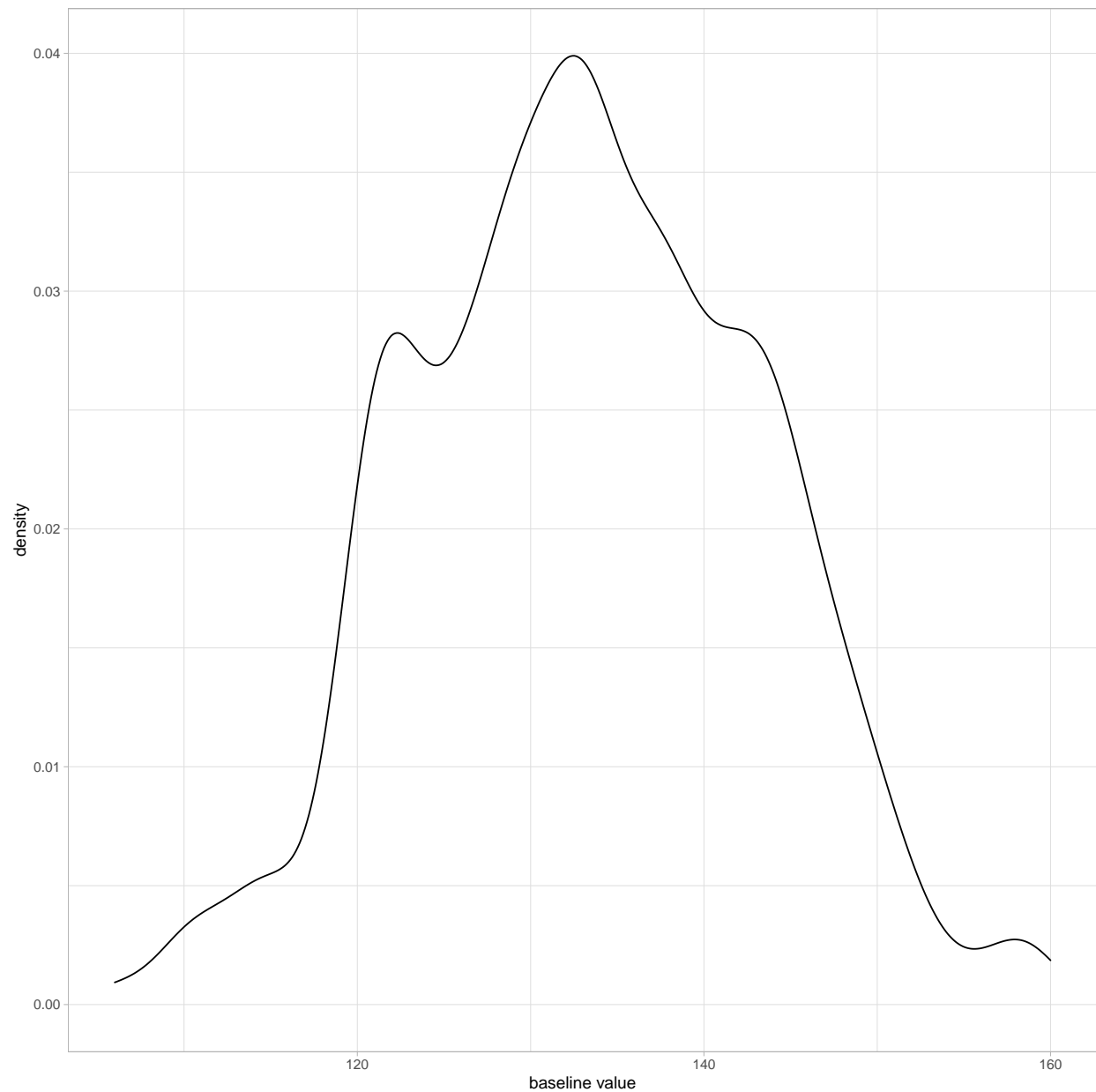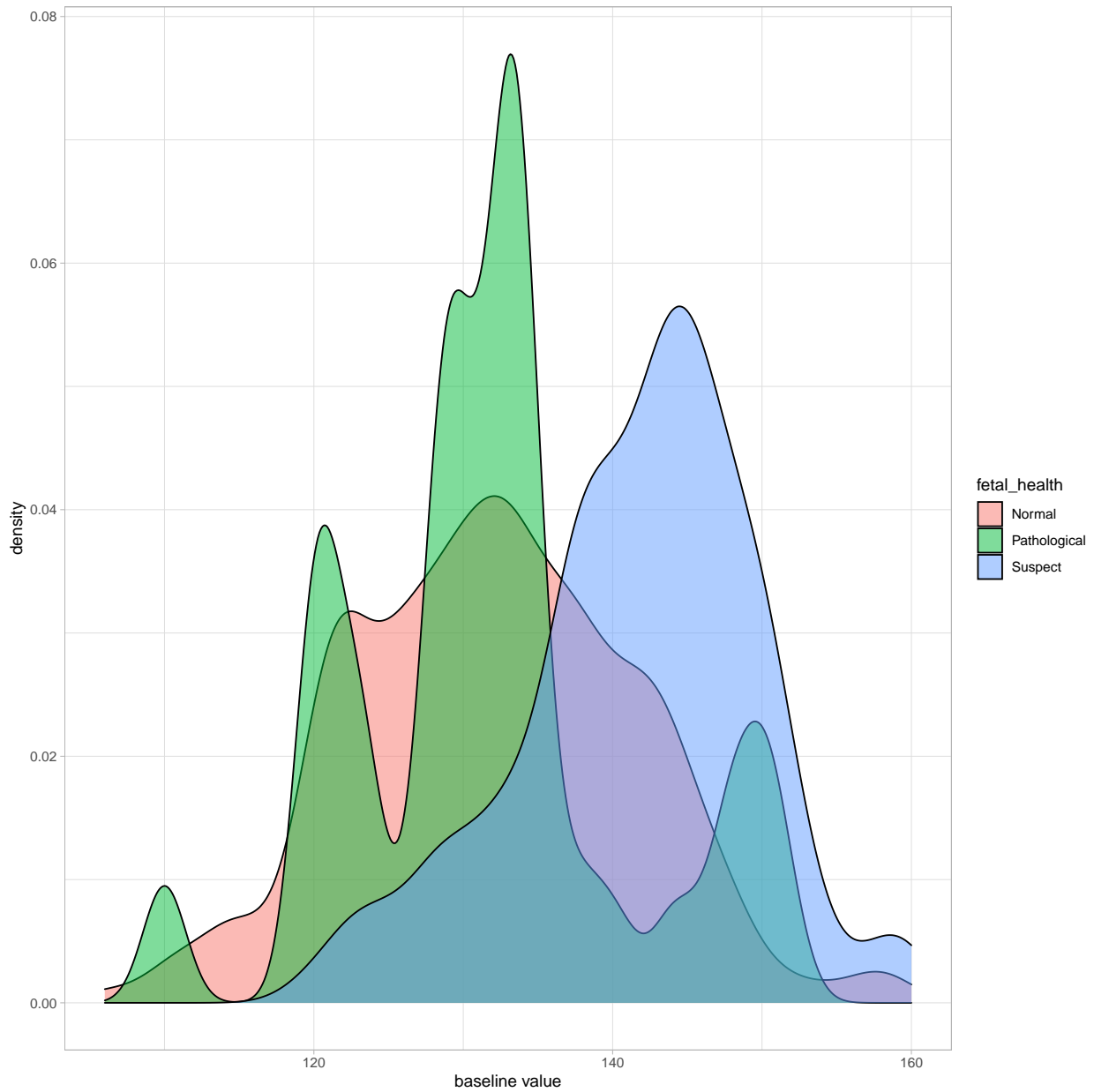

c)

Describe the distribution of the baseline heart rate for all participants in this sample, is the baseline heart rate normally distributed, comment on any skewness. What about each class? Use an appropriate graph to answer this question.

```
## 1.c
```

```r
df_fh %>%
  ggplot(aes(x = `baseline value`))+
  geom_density()
```



```r
df_fh %>%
  ggplot(aes(x = `baseline value`, fill = fetal_health))+
  geom_density(alpha = 0.5)
```

Question 2

b)

What is the mean baseline heart rate for each class?

c)

Since we have a small sample size for those who are suspected to be pathological and those who are truly pathological, what is the bootstrapped mean baseline heart rate for each class?

```
# Question 2
```

```
# Question 3
```

```
# Question 4
```

```
# Question 5
```

```
# Question 6
```