Investigating the Factors Affecting Birthweight

Andrew Hayes, id = 21321503

13 October, 2022

Question of Interest

Are the factors of smoking, previous history of hypertension or urinary irritability associated with whether babies were born with low birthweight (less than 2,500 grams)?

Load the Libraries and Data Needed

Load the required libraries so you can use them, and then make the birthweight data available ('lowbwt') as follows:

```
library(tidyverse)
library(aplore3)

data(lowbwt)
```

The low birthweight data is from the "Applied Logistic Regression" textbook by Hosmer and Lemeshow. The following is a description of the variables in this dataset.

Name	Description
subject	identification code
low	low birthweight ("< 2500 g" or ">= 2500 g")
age	age of mother
lwt	weight at last menstrual period (pounds)
race	race (Black, White, Other)
smoke	smoked during pregnancy (Yes, No)
ptl	premature labour history (None, One, Two, etc.)
ht	history of hypertension (Yes, No)
ui	uterine irritability (Yes, No)
ftv	number of visits to physician during 1st trimester (None, One, Two, etc.)
bwt	birthweight (in grams)

Subjective Impressions

The key variable of interest is `low` which represents whether a baby is born with low birthweight, defined as a birthweight below 2,500 grams.

```
lowbwt %>% select(low) %>% table()
```

```
## low
## >= 2500 g  < 2500 g
##      130        59
```

Let's explore the association between history of hypertension and low birthweight by tabulating the data.

```
lowbwt %>% select(low, ht) %>% table()
```

```
##            ht
## low         No Yes
##    >= 2500 g 125   5
##    < 2500 g   52   7
```

It seems there were not many mothers with hypertension, but the proportions of low weight babies is very much higher for mothers suffering from hypertension status than those that were not.

```
lowbwt %>% select(low, ht) %>% table() %>% prop.table(margin = 2)
```

```
##            ht
## low               No       Yes
##    >= 2500 g 0.7062147 0.4166667
##    < 2500 g  0.2937853 0.5833333
```

Task: In the following R chunk explore the association between uterine irritability and whether the babies were born with low birthweight, using both the counts and appropriate percentages. Explain the results in words.

```
lowbwt %>% select(low, ui) %>% table()
```

```
##            ui
## low         No Yes
##    >= 2500 g 116  14
##    < 2500 g   45  14
```

```
lowbwt %>% select(low, ui) %>% table() %>% prop.table(margin = 2)
```

```
##            ui
## low               No       Yes
##    >= 2500 g 0.7204969 0.5000000
##    < 2500 g  0.2795031 0.5000000
```

It seems that there were relatively few mothers with uterine irritability in the dataset, only 28. Of those 28, precisely half the babies had low weight, and the rest did not. When we compare these proportions to the rest of the dataset, we see that uterine irritability is definitely associated with low birth rate, although to a lesser degree than hypertension. Babies were almost 1.8 times more likely to be born with low weight if their mother suffered from uterine irritability than babies whose mothers did not.

Task: In the following R chunk explore the association between smoking status and whether the babies were born with low birthweight, using both the counts and appropriate percentages. Explain the results in words.

```
lowbwt %>% select(low, smoke) %>% table()
```

```
##            smoke
## low          No Yes
##    >= 2500 g 86  44
##    < 2500 g  29  30
```

```
lowbwt %>% select(low, smoke) %>% table() %>% prop.table(margin = 2)
```

```
##            smoke
## low              No       Yes
##    >= 2500 g 0.7478261 0.5945946
##    < 2500 g  0.2521739 0.4054054
```

Smoking definitely has an association with low birth weight, although to a lesser degree than uterine irritability. Babies of smokers were about 1.6 times more likely than the babies of non-smokers to have low birth rate. However, babies of smokers were more likely to be non-low weight than low weight. There are also a lot more mothers who are smokers than have uterine irritability, so the smoker data should be less sensitive to outliers.

Now we will create some barcharts.
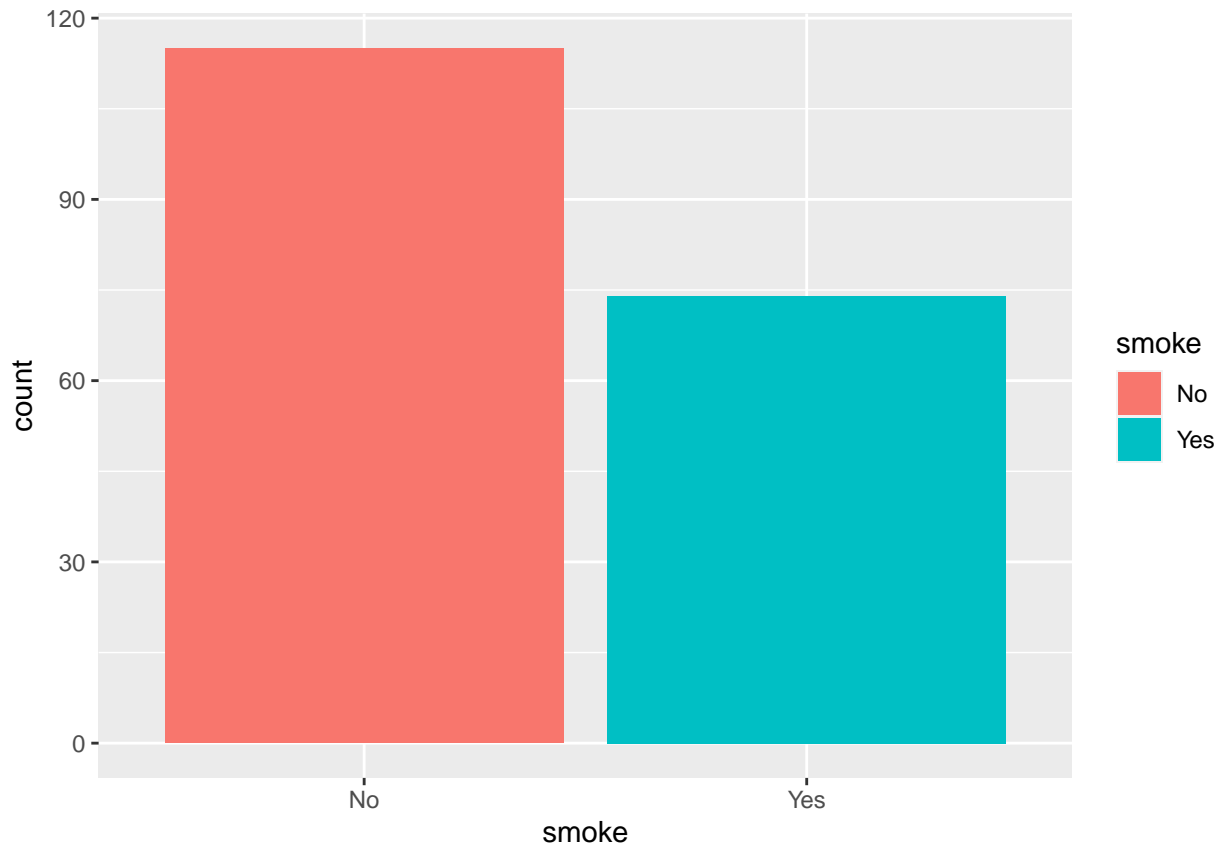
Barchart of Low Birthweight

The following is a frequency plot of the low birthweight status.

```
ggplot(lowbwt, aes(x = low, fill = low)) +
  geom_bar()
```
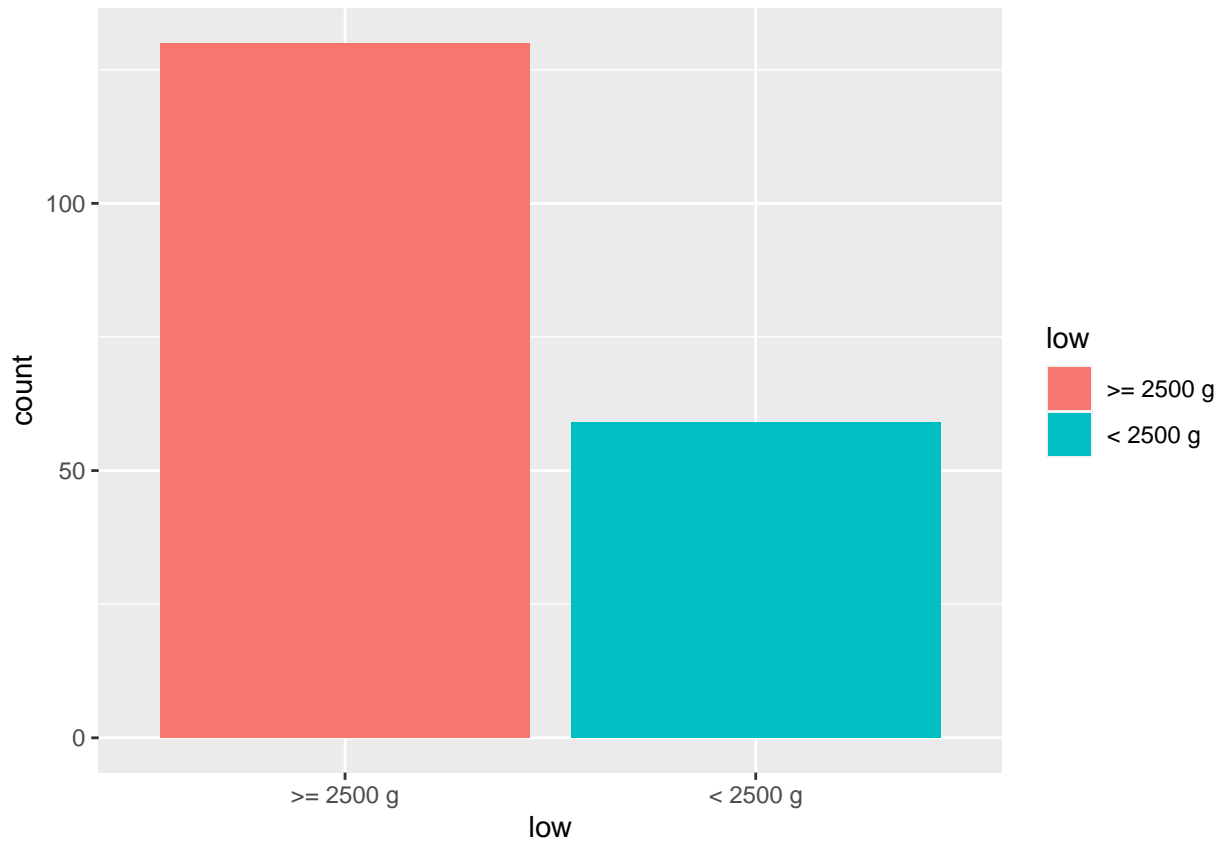
Task: In the following `R` chunk create a frequency plot of the smoking status.

```
ggplot(lowbwt, aes(x = smoke, fill = smoke)) +
  geom_bar()
```

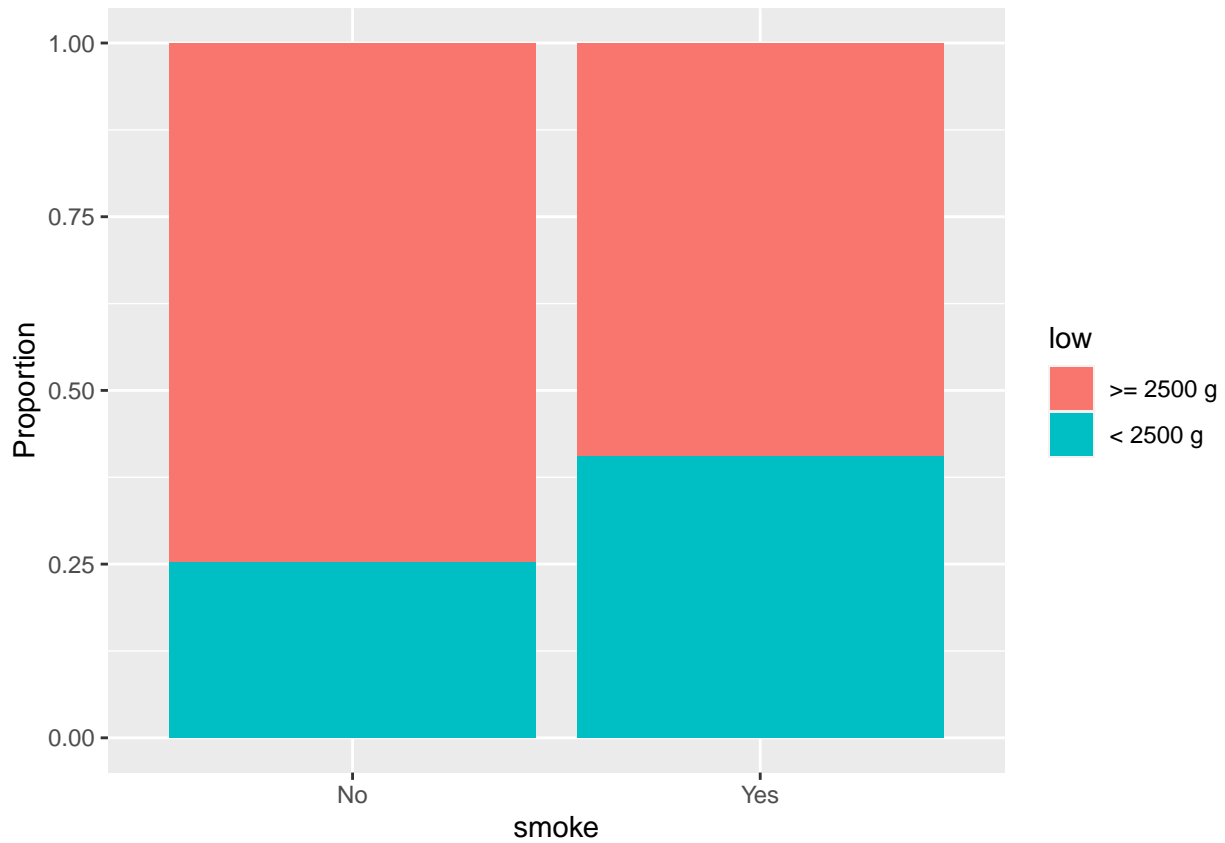## Stacked Barchart of Low Birthweight by Hypertension Status

Below is a relative frequency plot of the low birthweight of the babies against the hypertension status of the mothers using a stacked barchart.

```
ggplot(lowbwt, aes(x = ht)) +
    geom_bar(aes(fill = low), position = "fill") +
    ylab("Proportion")
```
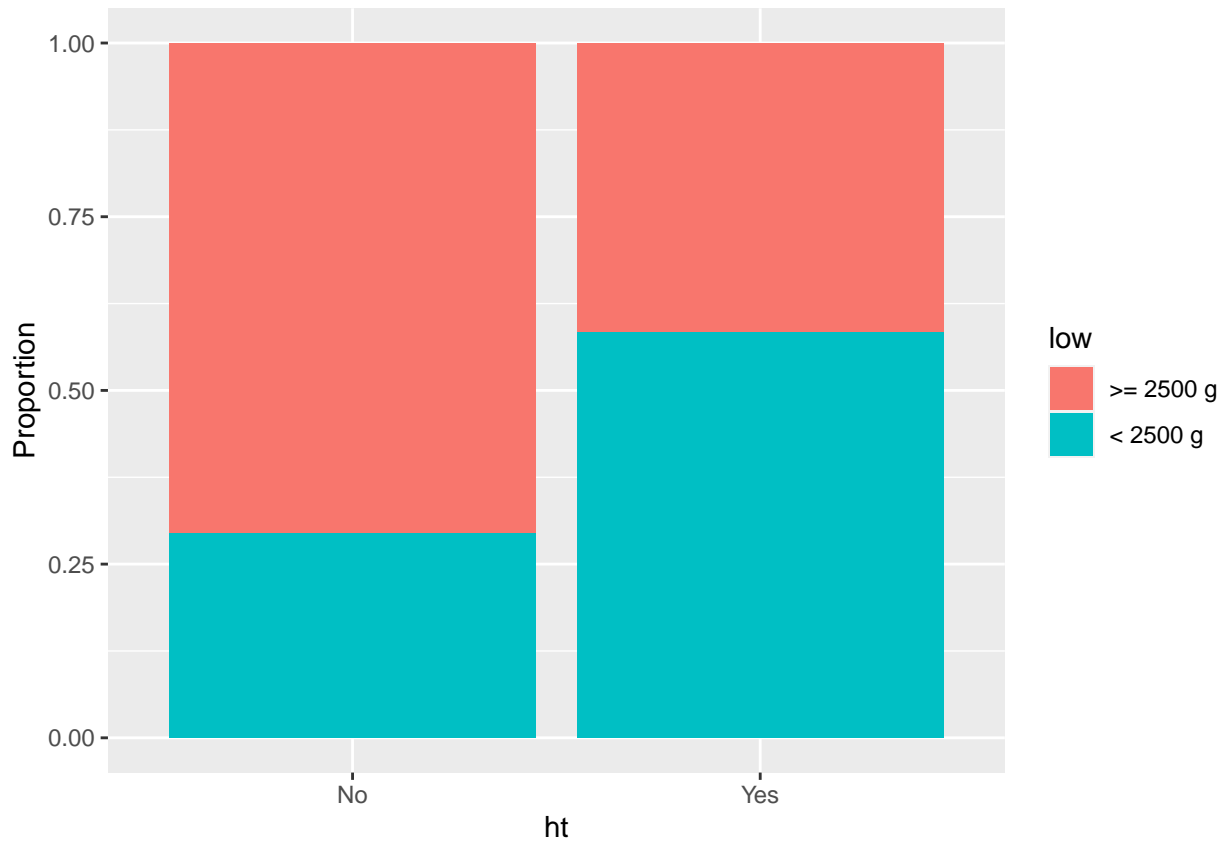
Task: Create a stacked barchart of low birthweight by smoking status by inserting an `R` chunk and relevant code below.

```
ggplot(lowbwt, aes(x = smoke)) +
    geom_bar(aes(fill = low), position = "fill") +
    ylab("Proportion")
```

Task: Create a stacked barchart of low birthweight by uterine irritability by inserting an R chunk and relevant code below.

```
ggplot(lowbwt, aes(x = ui)) +
    geom_bar(aes(fill = low), position = "fill") +
    ylab("Proportion")
```
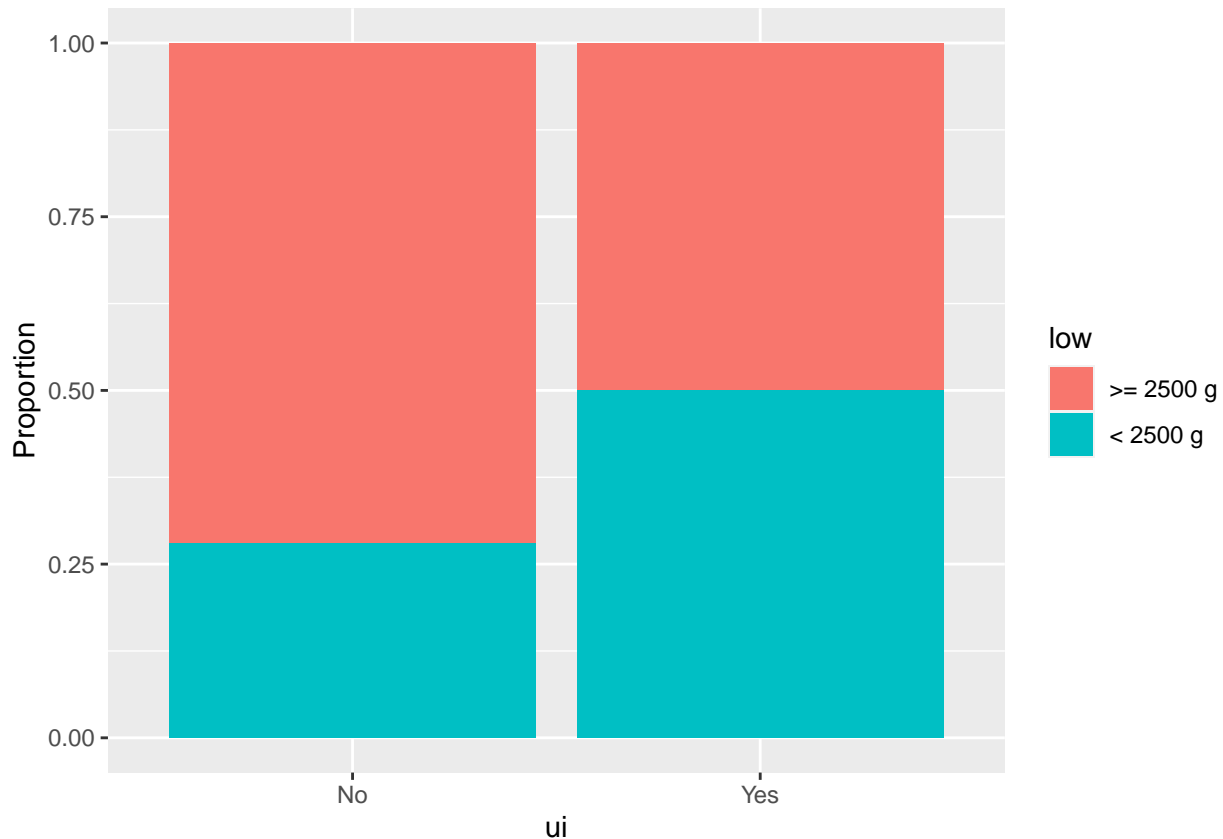
Task: Once you have created the plots, explain your interpretation of which factors are associated with low birthweight based on the three barcharts. State which factor you think is most associated with birthweight.

Based off of the 3 barcharts, I would say that all 3 factors are associated with low birthweight. It appears to me that hypertension is most associated with low birthweight (0.5833333), followed by uterine irritability (0.5000000), followed by smoke (0.4054054). However, it may be relevant to note that the sample size of each of those factors is rather small, and therefore sensitive to outliers.

The following `R` chunk produces a boxplot of the birthweight distribution.

```
lowbwt %>% ggplot(aes(y = bwt)) +
  geom_boxplot() +
  labs(y = "Birthweight (in grams)")
```
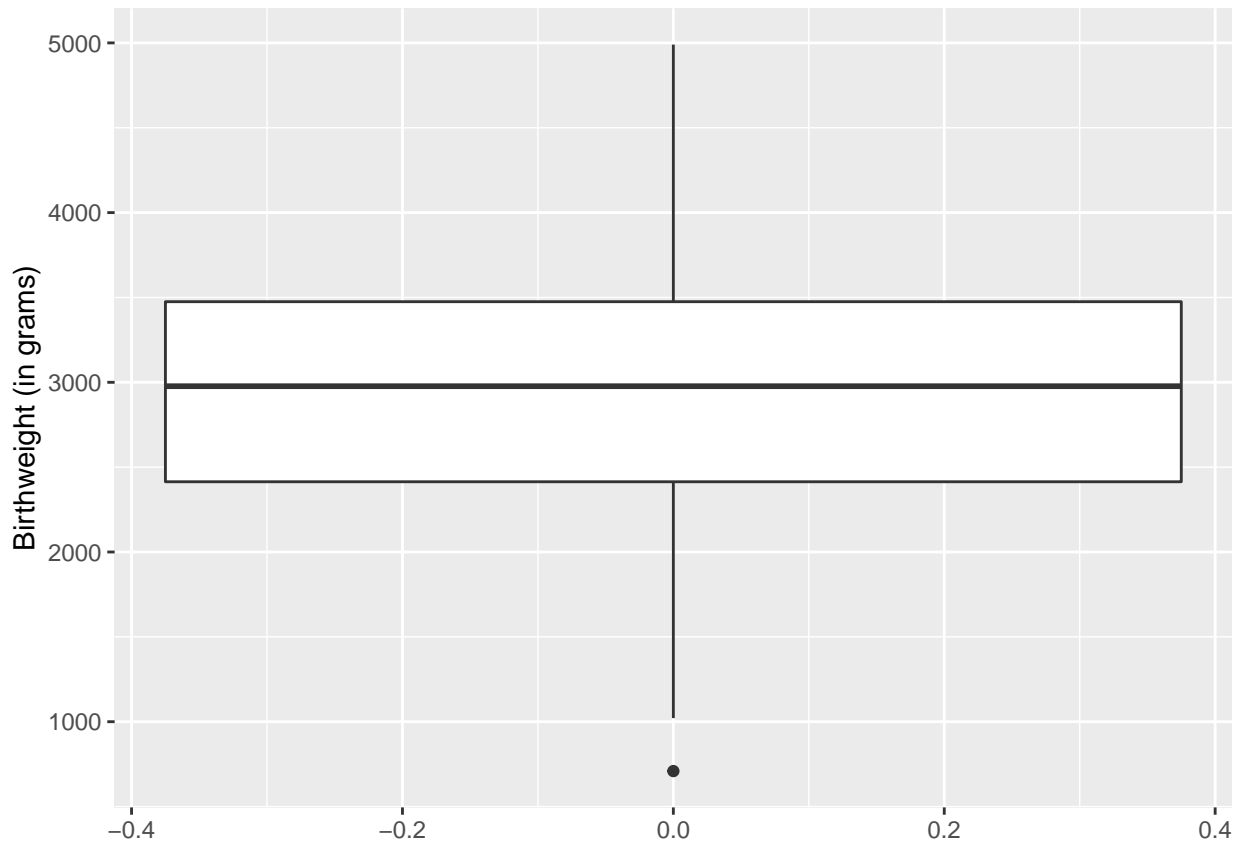
Task: In the previous task you stated which factor you believe was most associated with birthweight, so you can explore the impact on the distribution in more detail. Create a graph of side-by-side boxplots comparing the birthweight distribution for each level of that factor (e.g. comparing mothers who had uterine irritability and those who did not), by inserting an R chunk and relevant code below.

[Hint: we used side-by-side boxplots in the week 4 lab and in the Exploratory Data Analysis worksheet]

```
lowbwt %>% ggplot(aes(y = bwt, x = factor(ht))) +
  geom_boxplot() +
  labs(x = "ht")
```
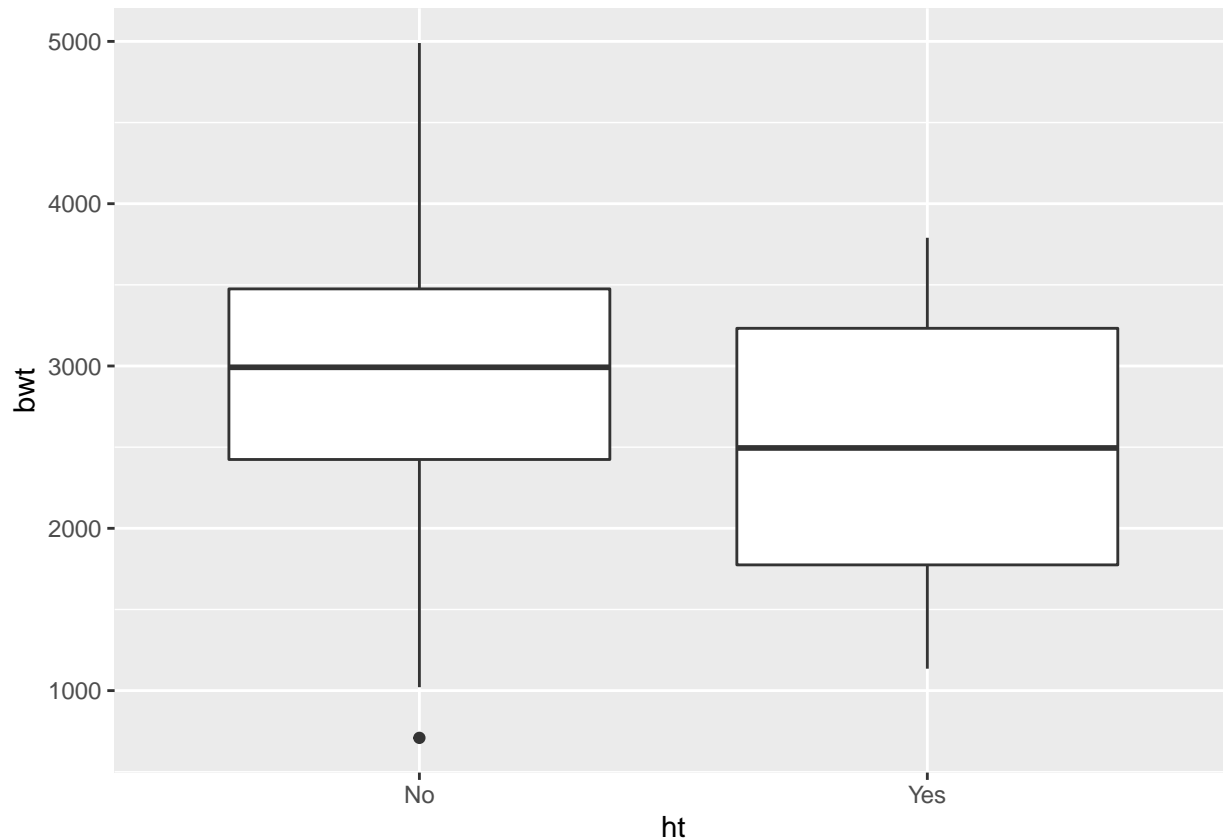
Conclusion

Task: Write a short conclusion of whether you think low birthweight of babies can be predicted based on whether the mother smoked, has hypertension or uterine irritability.

It's clear that there is some association between low birthweight of babies based off whether the mother smoked, had hyptertension, or uterine irritability, However, it's more difficult to say whether low birthweight can be predicted by any of these factors.

I would say that low birthweight can be predicted by hypertension, as the majority of babies born to mothers with hypertension have low birthrate (0.5833333).

However, uterine irritability is a predictor of an increased chance of low birthweight, but not low birthweight itself. It appears from the data that it is 50-50 (0.5000000) whether a baby born to a mother with uterine irritability will have low birthrate, so you couldn't make an accurate prediction about the baby of a mother with uterine irritability. However, it most certainly does predict increased chances of low birthweight.

Finally, while smoking is a predictor of increased chance of low birthweight (0.4054054), the majority of babies born to smokers do not have low birthweight, which means that the only reasonable prediction you could make about a baby born to a smoker is that it *will not* have a low birthweight, although there is a significant chance that it will, which is significantly higher than the chance of a non-smoker having a baby with low birthweight.

Final Task: "knit" the file as a Word of PDF document and submit it via the relevant link on Blackboard before the deadline.