Investigating the Factors Affecting Birthweight

Smitesh Patil, id = 22223696

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

The dataset you need is available in the `aplore3` package and we will also use the `tidyverse` package. These may not be installed already. If you need to install them then run following code in `R`. You will only need to do this once. After they have been installed you can just delete this chunk and text.

```
install.packages(c("tidyverse", "aplore3"),
                 repos = "https://cran.microsoft.com/", quiet = TRUE)
```

```
## package 'tidyverse' successfully unpacked and MD5 sums checked
## package 'aplore3' successfully unpacked and MD5 sums checked
```

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)

Name	Description
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()
```

```
## .
## >= 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.

Answer: In the dataset, there are not many mothers with uterine irritability. But in the percentage of low weight babies with mothers surrfering from uterin irritablity is much higher.

For low birth weight babies, mothers with uterine irritablity(23.7%) have more that double the rate compared to normal birthweight babies(10%).

Absolute proportion of mothers with uterine irritability is divided equally among low birthweight(50%) and normal birthweight babies(50%).

```
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 = 1)*100
```

```
##            ui
## low               No      Yes
##   >= 2500 g 89.23077 10.76923
##   < 2500 g  76.27119 23.72881
```

```
lowbwt %>% select(low, ui) %>% table() %>% prop.table(margin = 2)*100
```

```
##            ui
## low               No      Yes
##   >= 2500 g 72.04969 50.00000
##   < 2500 g  27.95031 50.00000
```

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.

Answer: Smoking seems to be adversely affecting babies with low birthweight.

Among low weight babies almost 50% of mothers smoked. on the other hand for babies with nomal birthweight only 33% of mothers smoked.

In terms of absolute proportion of mothers with smoking habit, proportion is higher for normal birth weights(59%) and higher for low birth weights(40%)

```
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 = 1)*100
```

```
##            smoke
## low               No      Yes
##   >= 2500 g 66.15385 33.84615
##   < 2500 g  49.15254 50.84746
```

```
lowbwt %>% select(low, smoke) %>% table() %>% prop.table(margin = 2)*100
```

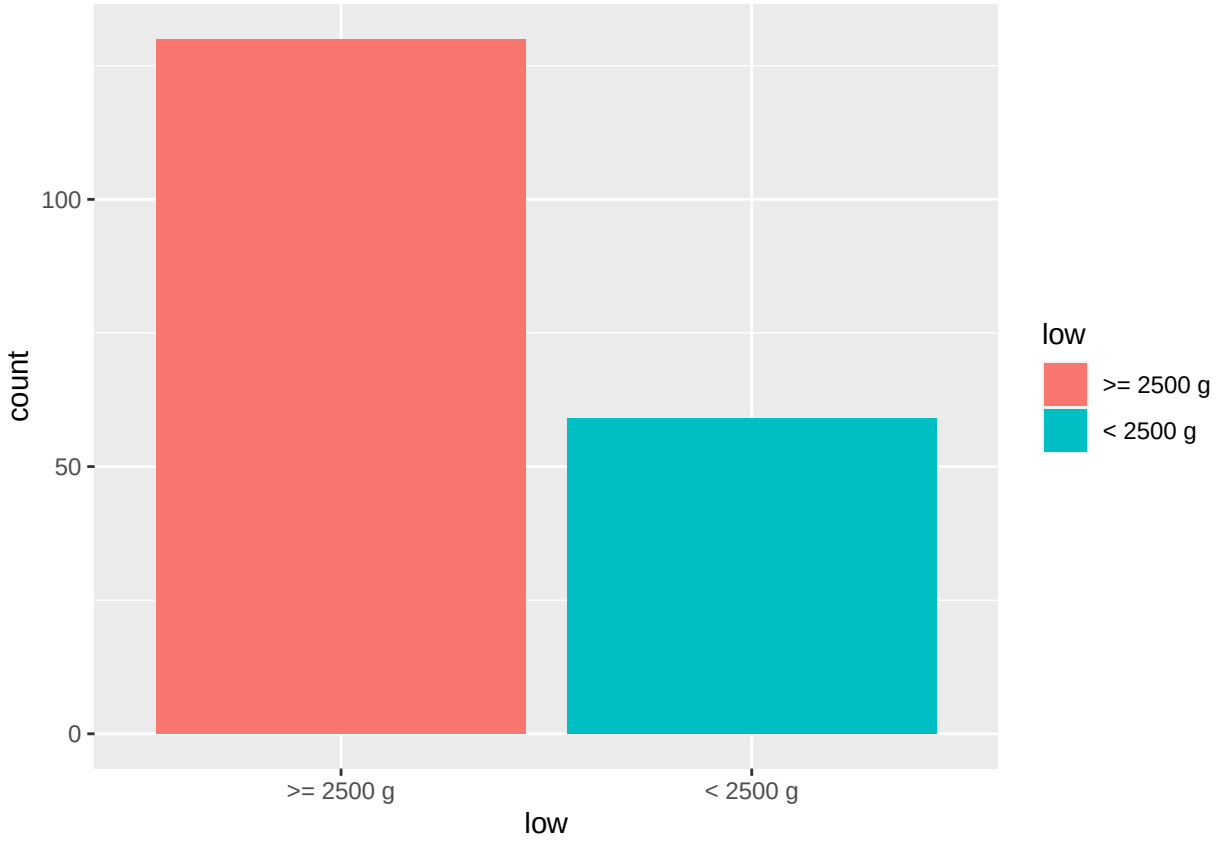
```
##            smoke
## low               No      Yes
##   >= 2500 g 74.78261 59.45946
##   < 2500 g  25.21739 40.54054
```

Now we will create some barcharts.

Barchart of Low Birthweight

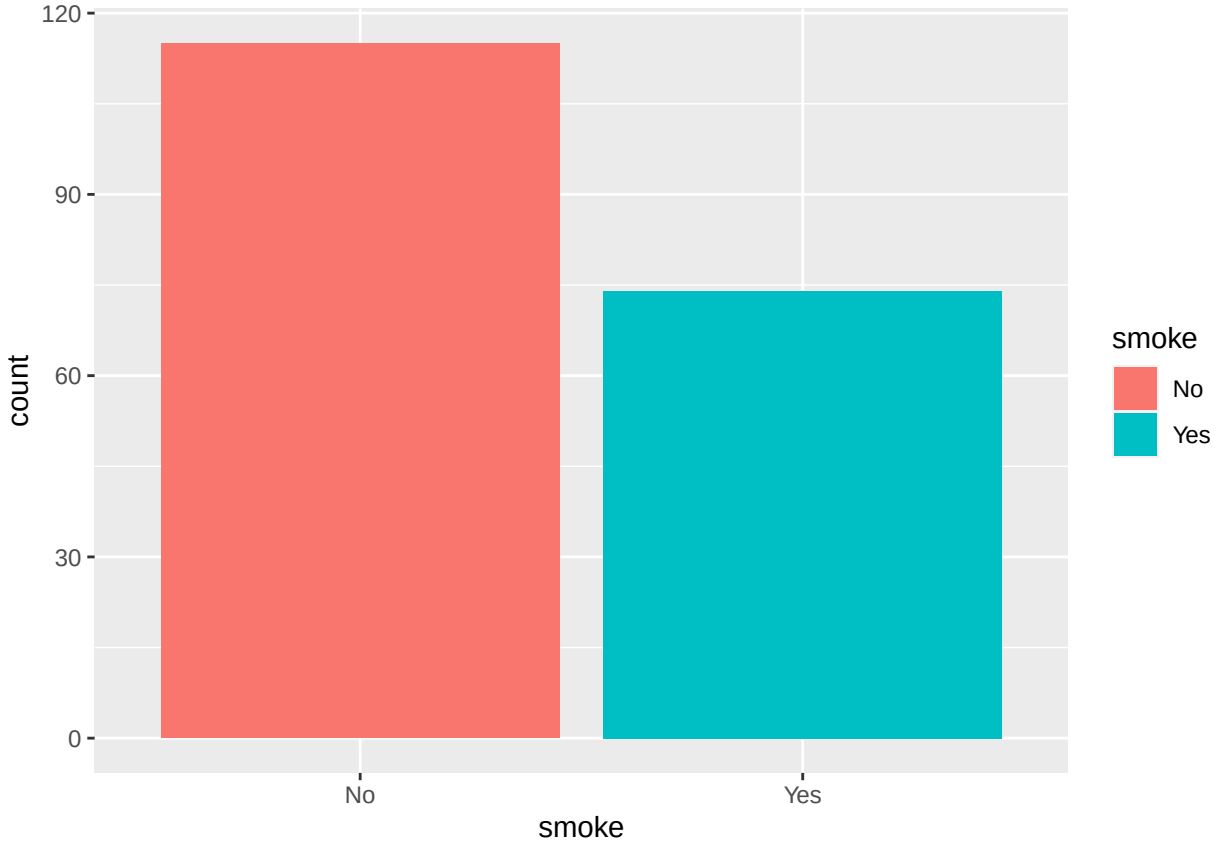
The following is a frequency plot of the low birthweight status.

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



Task: In the following `R` chunk create a frequency plot of the smoking status.

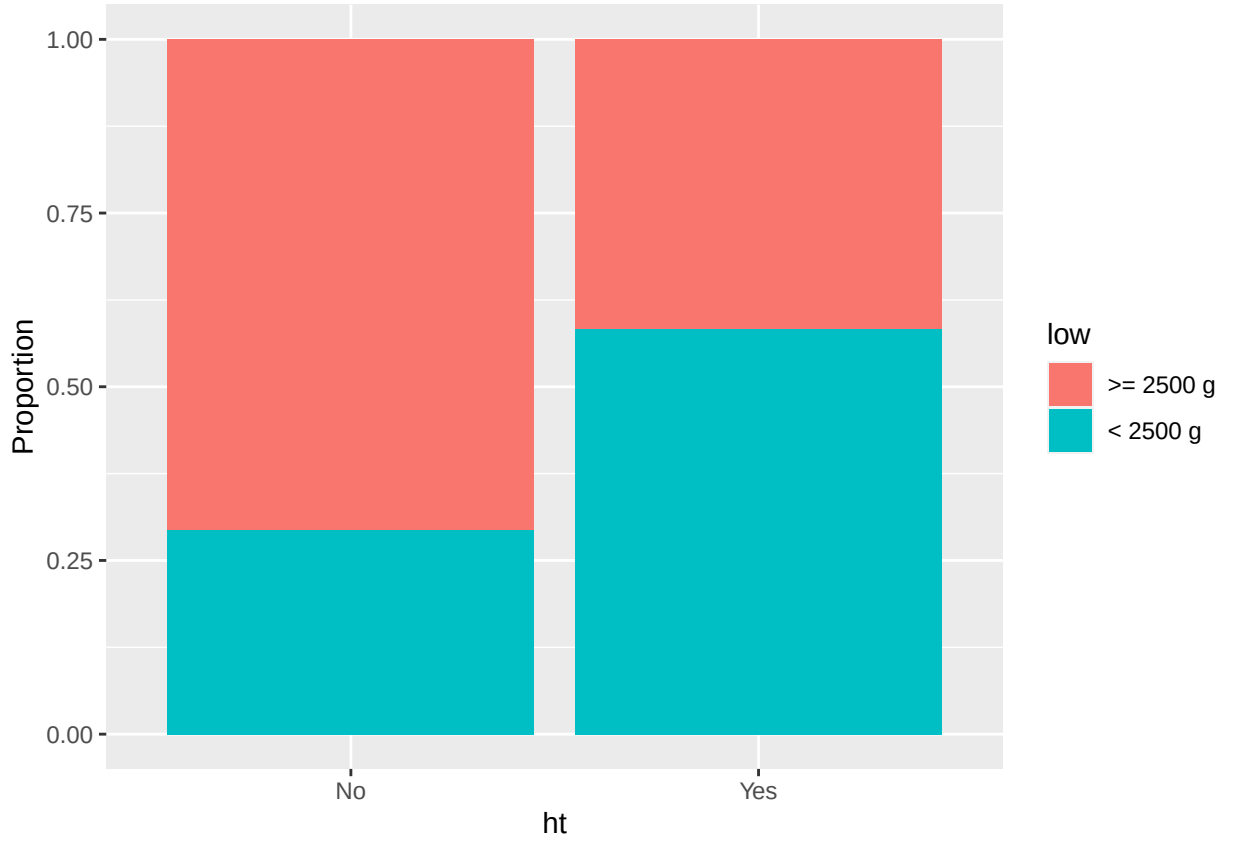
```
ggplot(lowbwt, aes(x = smoke, fill=smoke)) +
  geom_bar()
```



Stacked Barchart of Low Birthweight by Hypertension Status

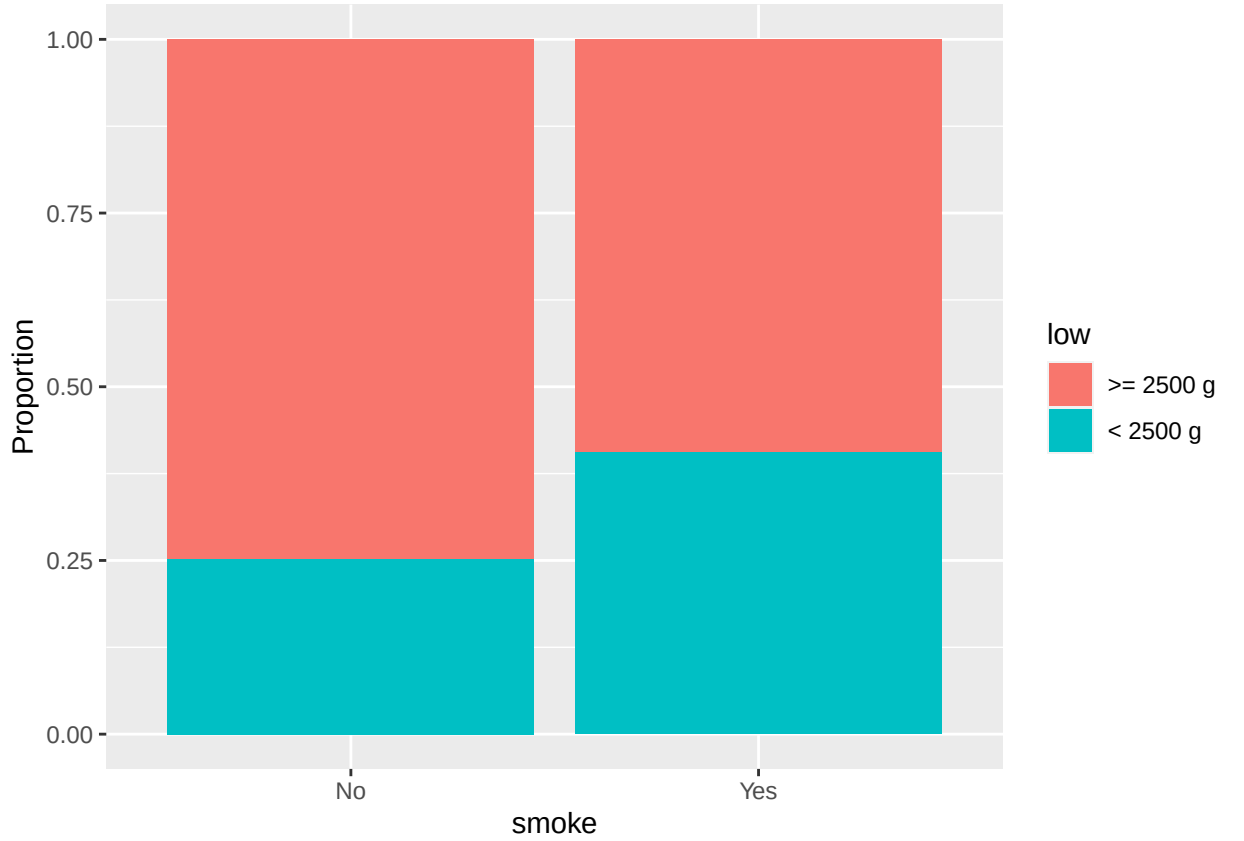
Below is a relative frequency plot of the low birthweight of the babies against the hypertension status of the mothers using a stacked barchart.

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



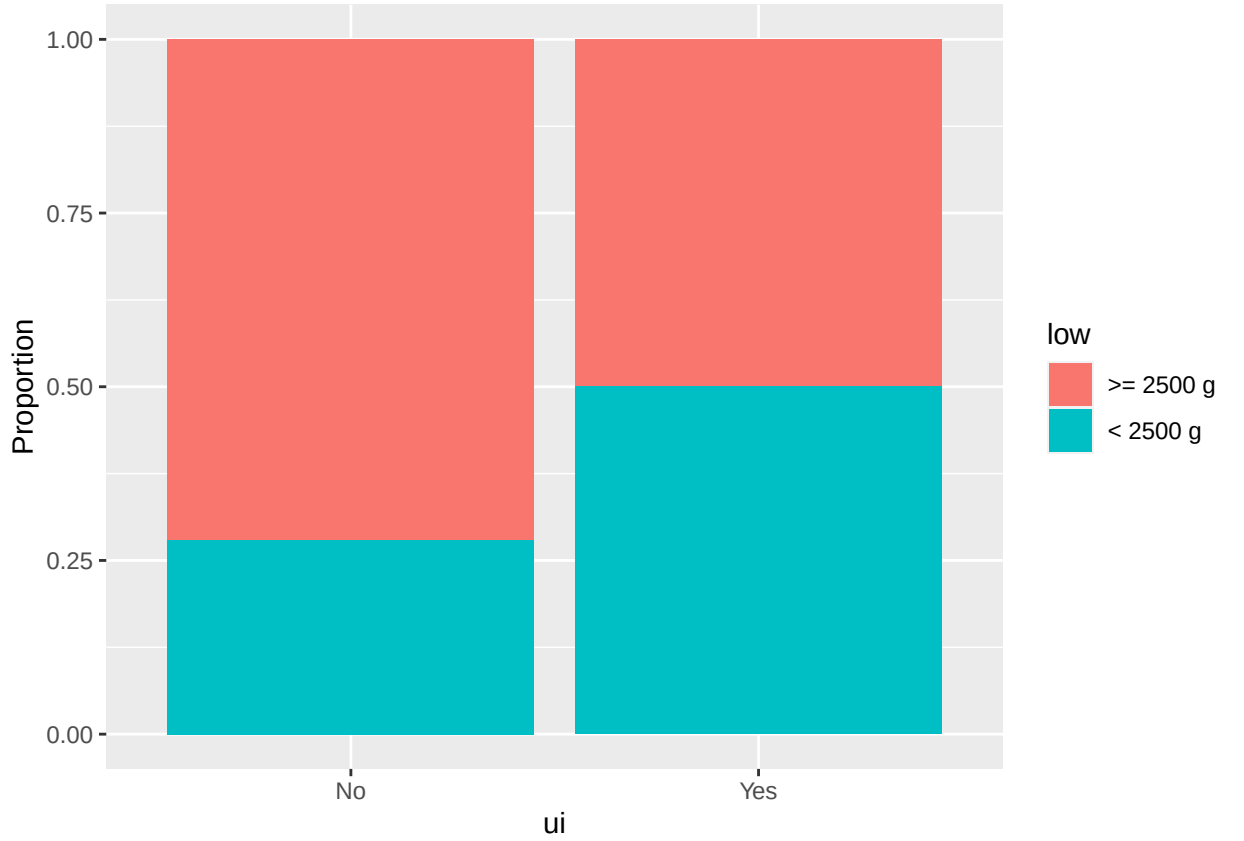
Task: Create a stacked barchart of low birthweight by smoking status by inserting an `R` chunk and relevant code below.

```
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")
```

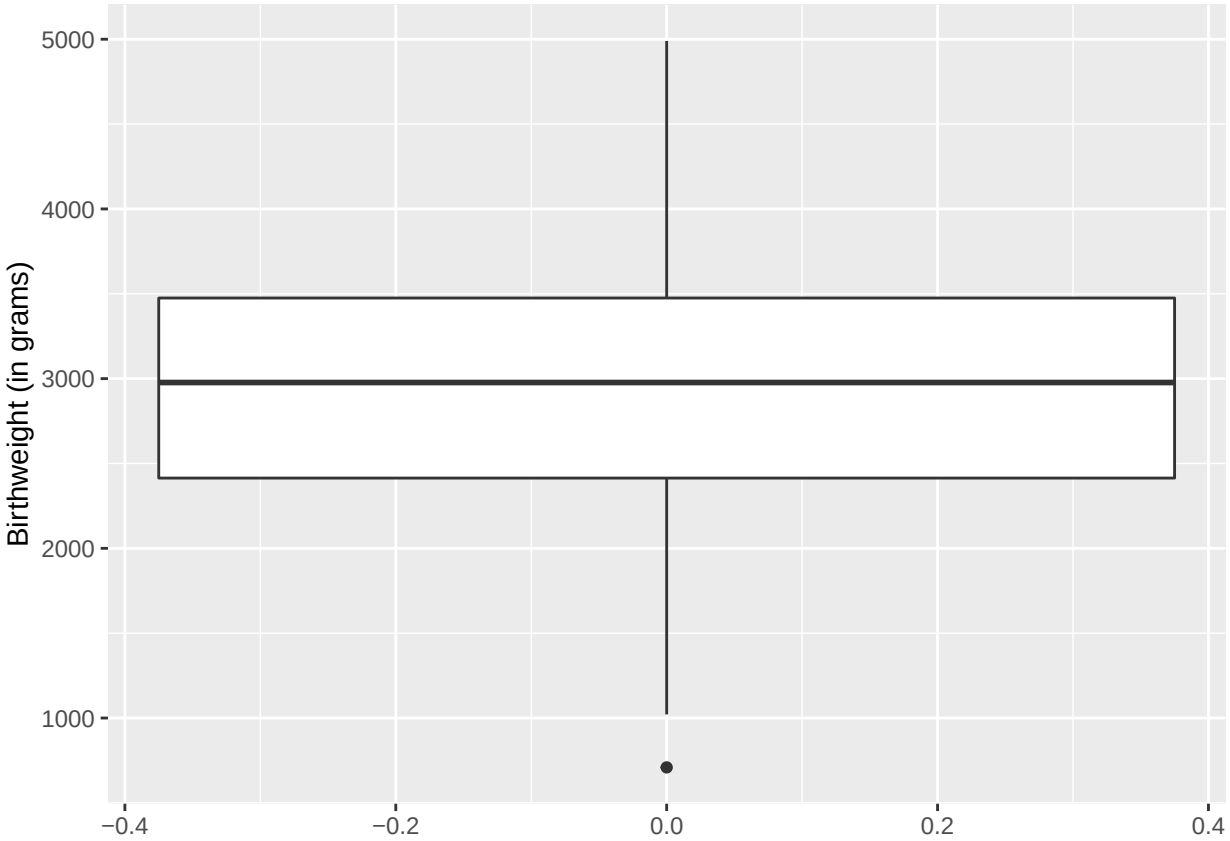


Task: Once you have created the plots, explain your interpretation of which factors are associated with low birthweight based on the three barcharts. State which factor you think is most associated with birthweight.

Answer: Uterine irritability seems to be the factor most associate with birthweght as the proportion mothers of low weight babies seem to higher for uterine irritability(23.7) that women without the condition(10.7).

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

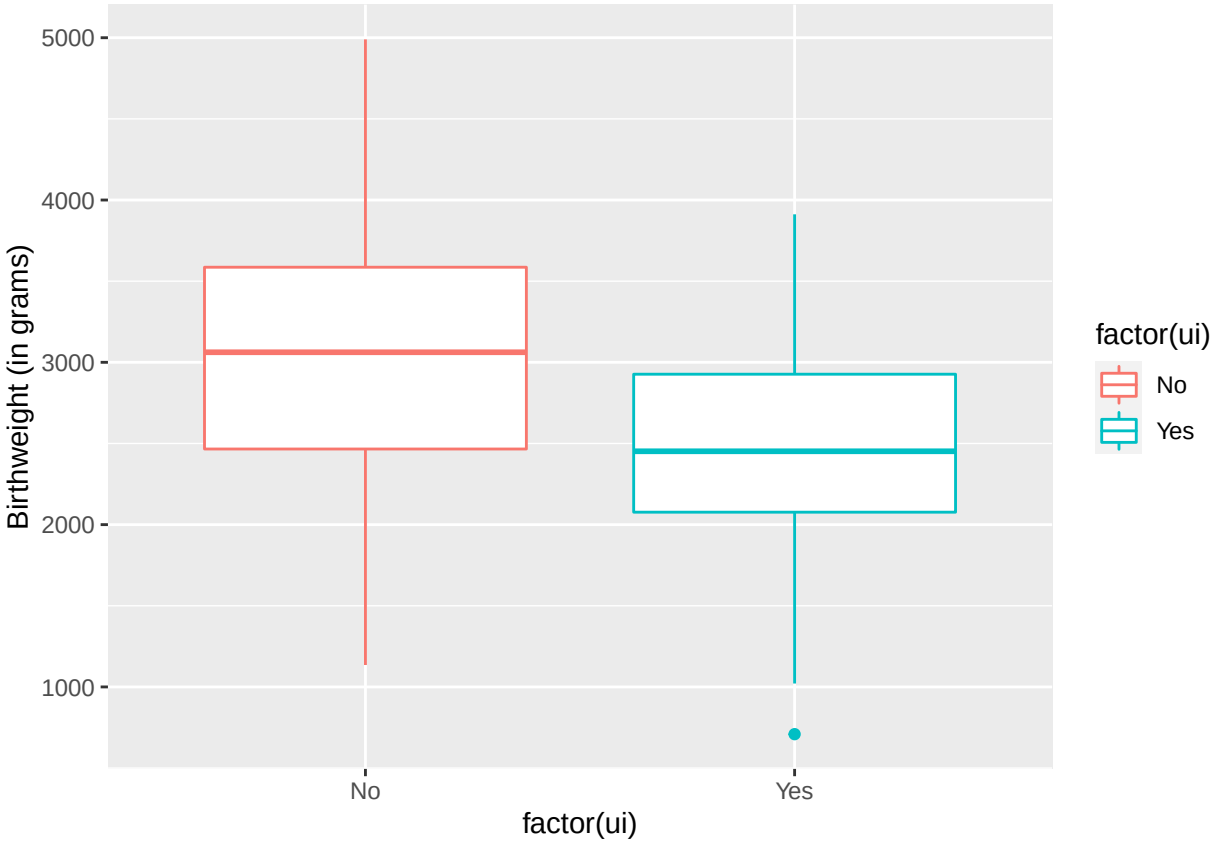
```
lowbwt %>% ggplot(aes(y = bwt)) +
  geom_boxplot() +
  labs(y = "Birthweight (in grams)")
```



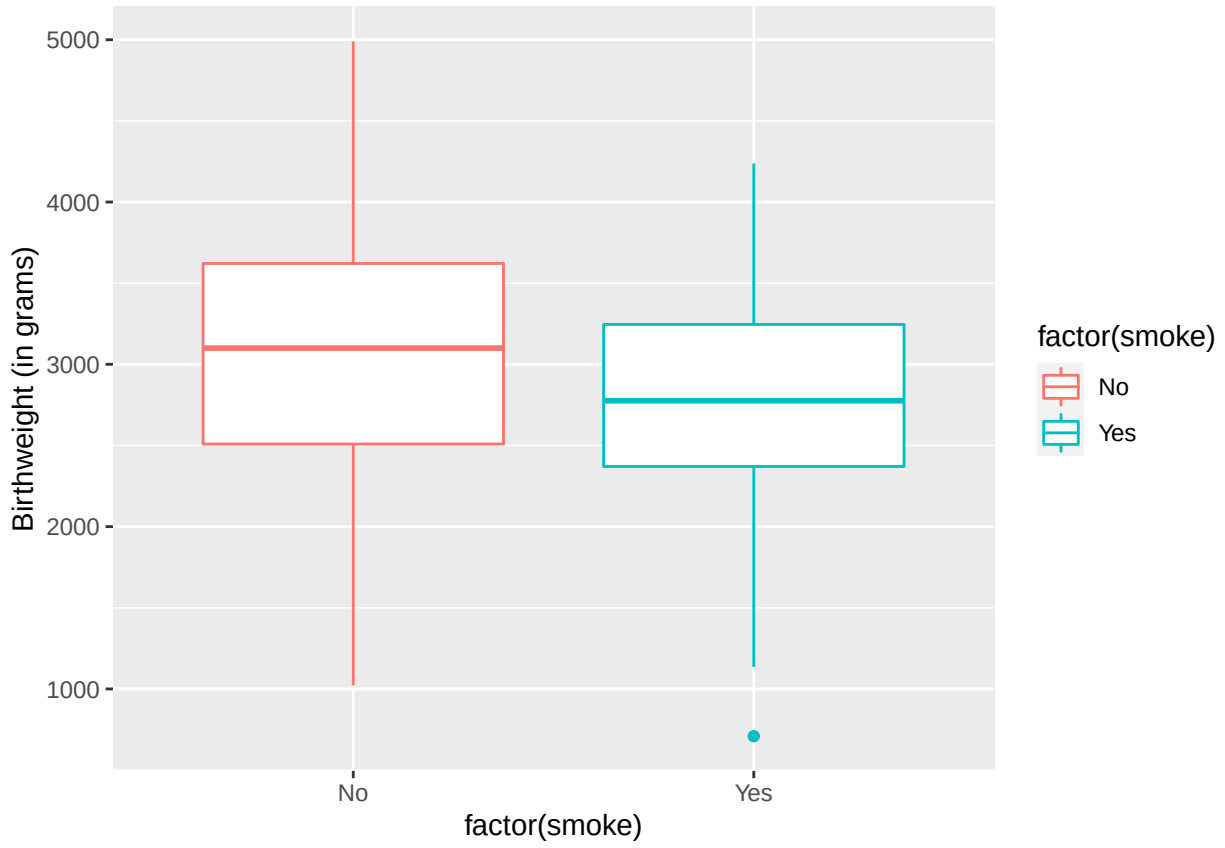
Task: In the previous task you stated which factor you believe was most associated with birthweight, so you can explore the impact on the distribution in more detail. Create a graph of side-by-side boxplots comparing the birthweight distribution for each level of that factor (e.g. comparing mothers who 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]

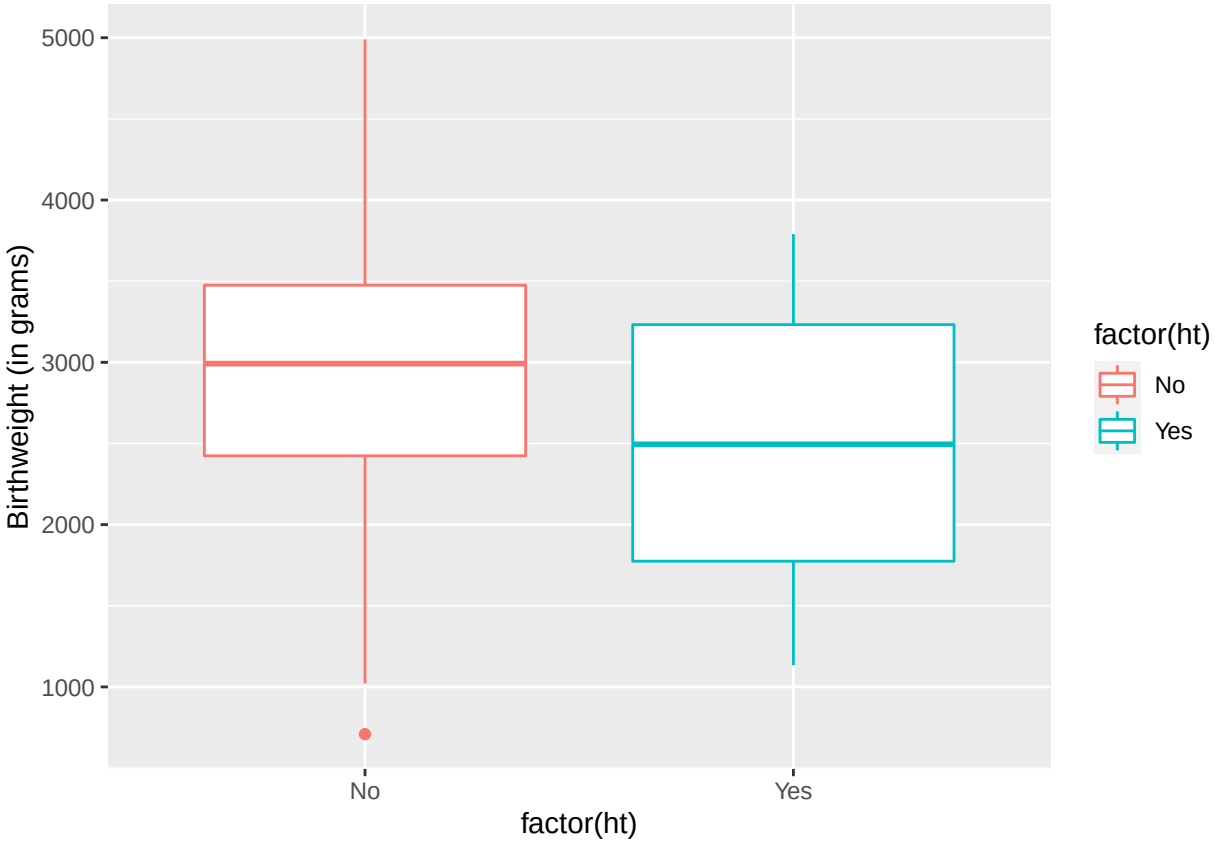
```
lowbwt %>% ggplot(aes(y = bwt, x = factor(ui), color = factor(ui))) +
  geom_boxplot() +
  labs(y = "Birthweight (in grams)")
```




```
lowbwt %>% ggplot(aes(y = bwt, x = factor(smoke), color = factor(smoke))) +
  geom_boxplot() +
  labs(y = "Birthweight (in grams)")
```




```
lowbwt %>% ggplot(aes(y = bwt, x = factor(ht), color = factor(ht))) +
  geom_boxplot() +
  labs(y = "Birthweight (in grams)")
```



‘‘‘

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.

Answer: Yes the factors like smoking, hypertension and uterine irritability seem to be correlated to low birthweight in babies to a varying degree. They can be used to used to predict low weight in babies

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