

Investigating the Factors Affecting Birthweight

Chin Zhe Jing, id = 22221970

10 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)
```

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
```

*#There were less cases with mother having uterine irritability, but the proportions of low weight babies is much higher for those mothers having uterine irritability than those that were not.
#Besides, mothers with no uterine irritability are having higher proportions on giving birth to normal weight babies.*

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
```

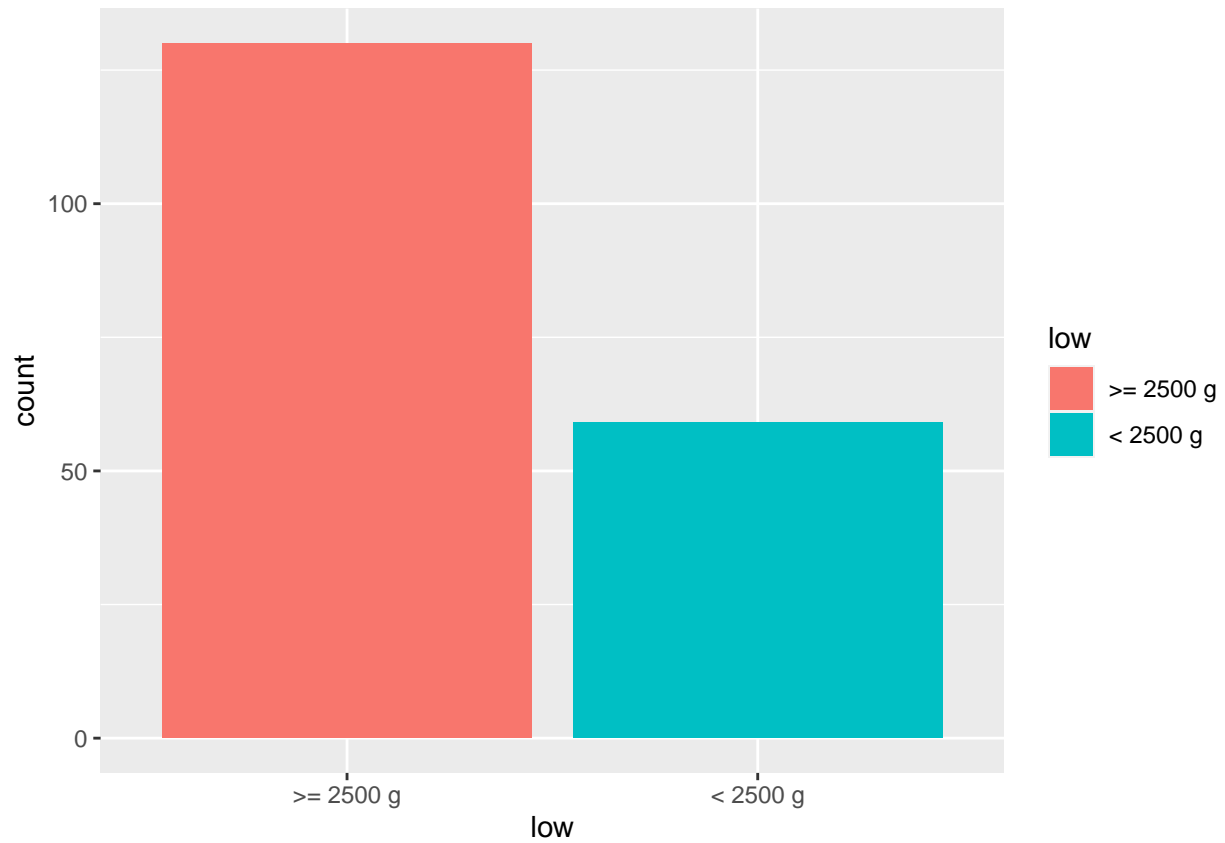
*#The number of smoking mothers is only 41 less than the number of non-smoking mothers.
#The proportions of low weight babies is higher for mothers who smoke than those that were not.*

Now we will create some barcharts.

Barchart of Low Birthweight

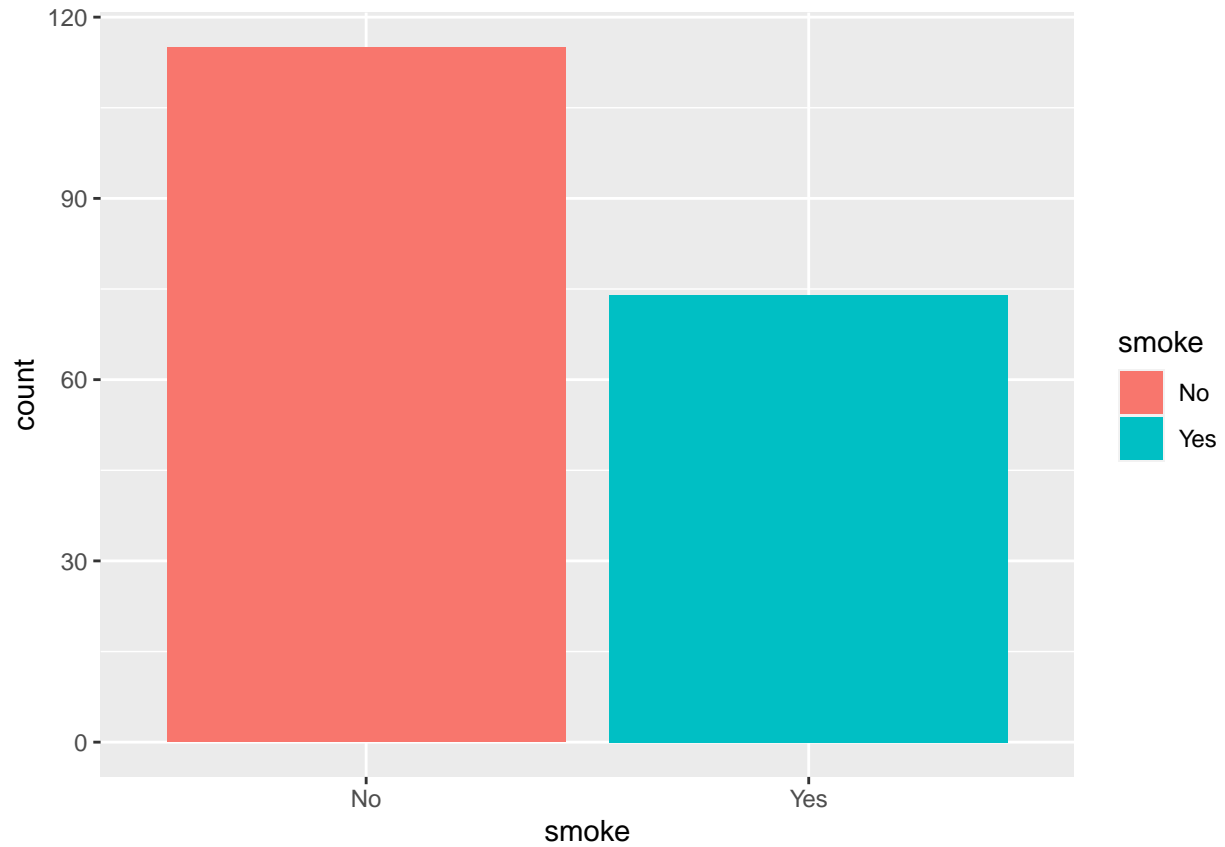
The following is a frequency plot of the low birthweight status.

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



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

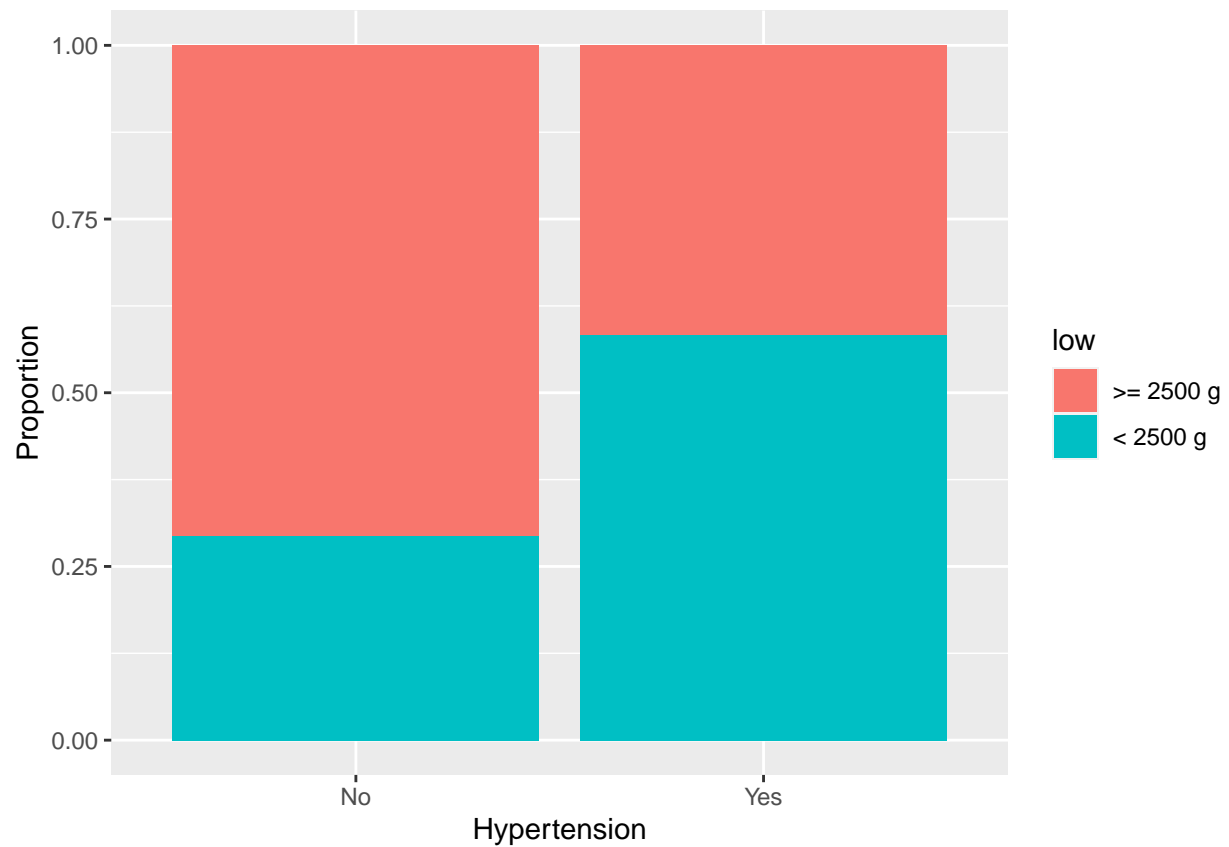
```
ggplot(lowbwt, aes(x = smoke, fill = smoke)) +  
  geom_bar()
```



Stacked Barchart of Low Birthweight by Hypertension Status

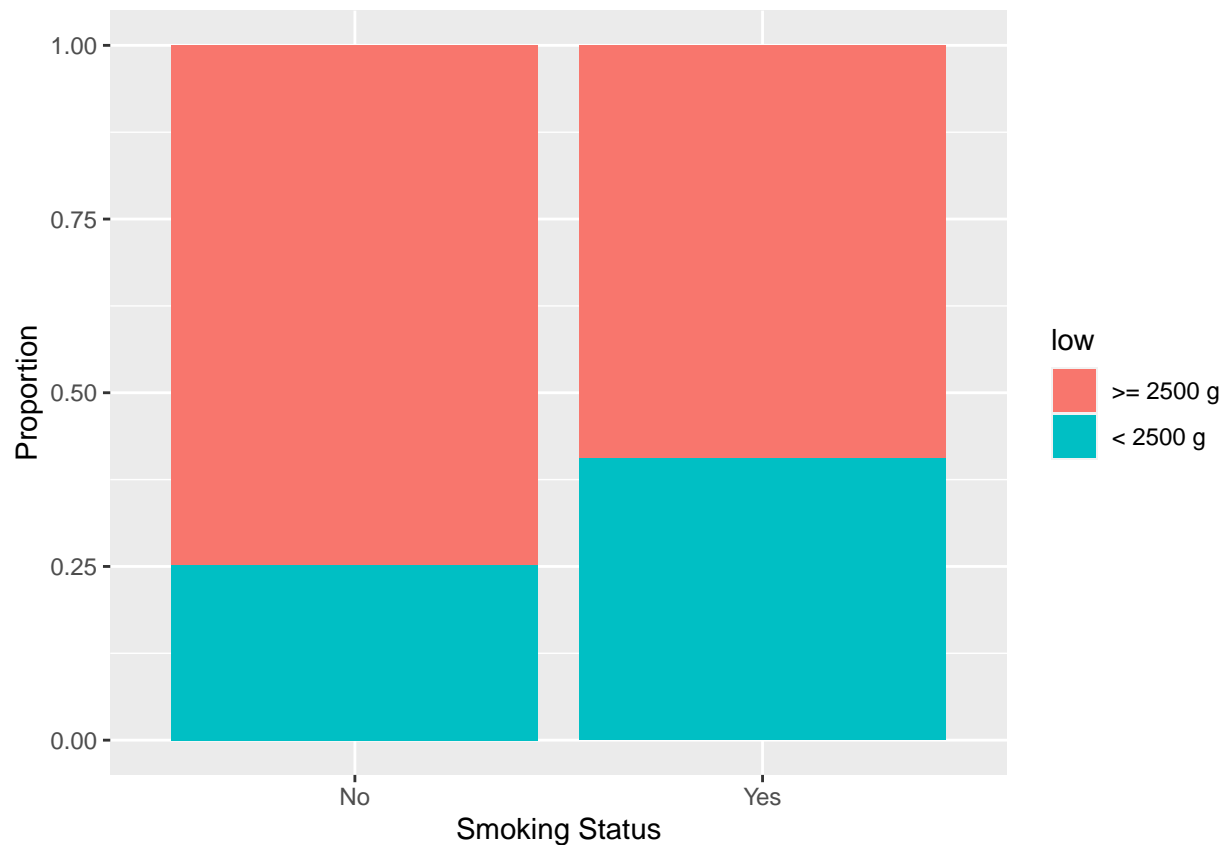
Below is a relative frequency plot of the low birthweight of the babies against the hypertension status of the mothers using a stacked barchart.

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



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

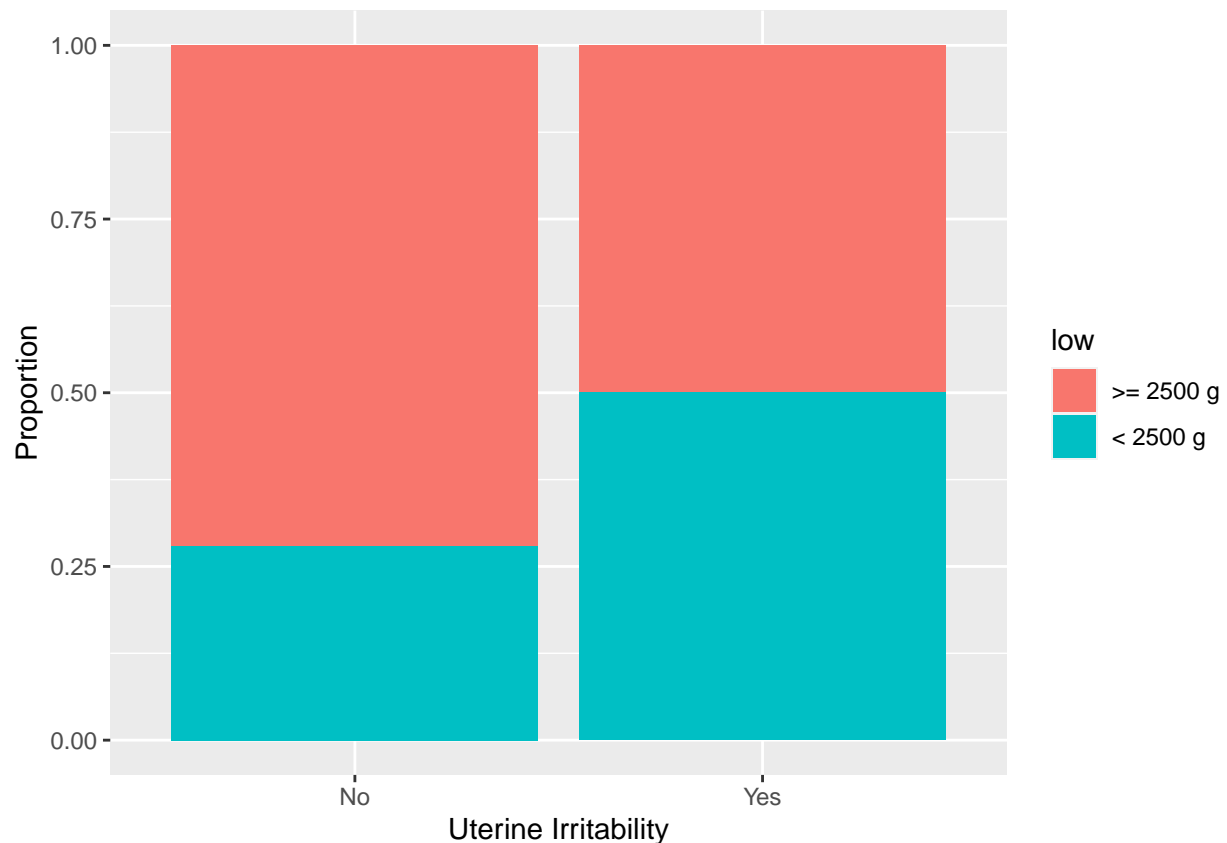
```
ggplot(lowbwt, aes(x = smoke)) +  
  geom_bar(aes(fill = low), position = "fill") +  
  ylab("Proportion") + xlab("Smoking Status")
```



*#From the bar chart, we can clearly see that smoking mothers are having higher proportions
#of giving birth to low birth weight babies than those that
#were not smoking (blue part of yes > blue part of no).
#While non-smoking mothers are having higher proportional to give birth to
#normal weight babies than those that were smoking.*

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") + xlab("Uterine Irritability")
```



*#From the bar chart, we can clearly see that mothers suffering from
 #uterine irritability are having
 #higher proportions of giving birth to low birth weight babies than
 #those that were not (blue part of yes > blue part of no).
 #While mothers without uterine irritability are having higher proportional to give birth to
 #normal weight babies than those that were having uterine irritability.*

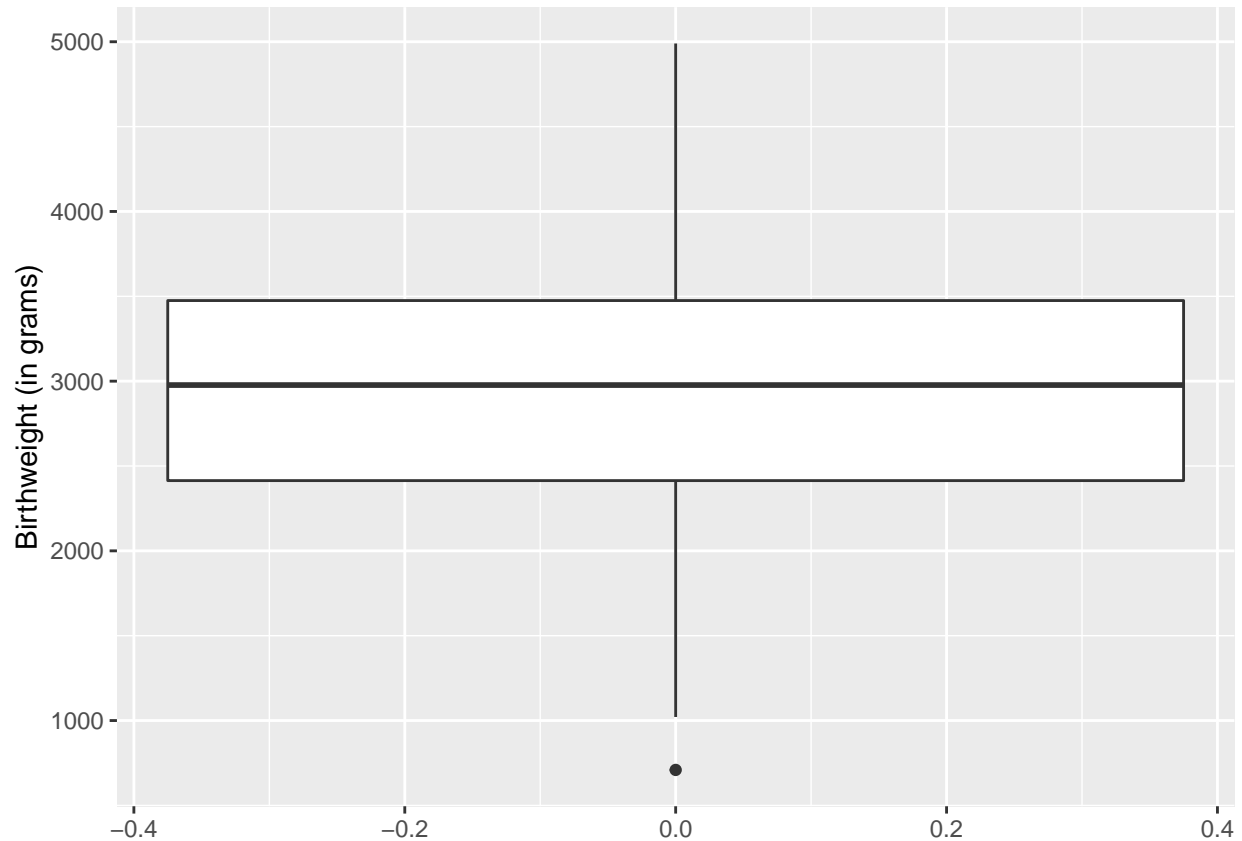
Task: Once you have created the plots, explain your interpretation of which factors are associated with low birthweight based on the three barcharts. State which factor you think is most associated with birthweight.

By comparing the barcharts, the proportion of mothers having hypertension give birth to low weight babies were more than 50%, for smoking mothers were less than 50% and for mothers with uterine irritability were about 50%. Thus, I think that hypertension is most associated with low birthweight.

The proportion of birthweight for mothers without smoking status or hypertension or uterine irritability is more or less similar but having those factors fluctuated the proportions of low weight babies. When compared to hypertension and uterine irritability, the gap between normal weight babies of mothers with or without smoking status is smaller. Thus, I think that the factor that most associated with birthweight should be hypertension.

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

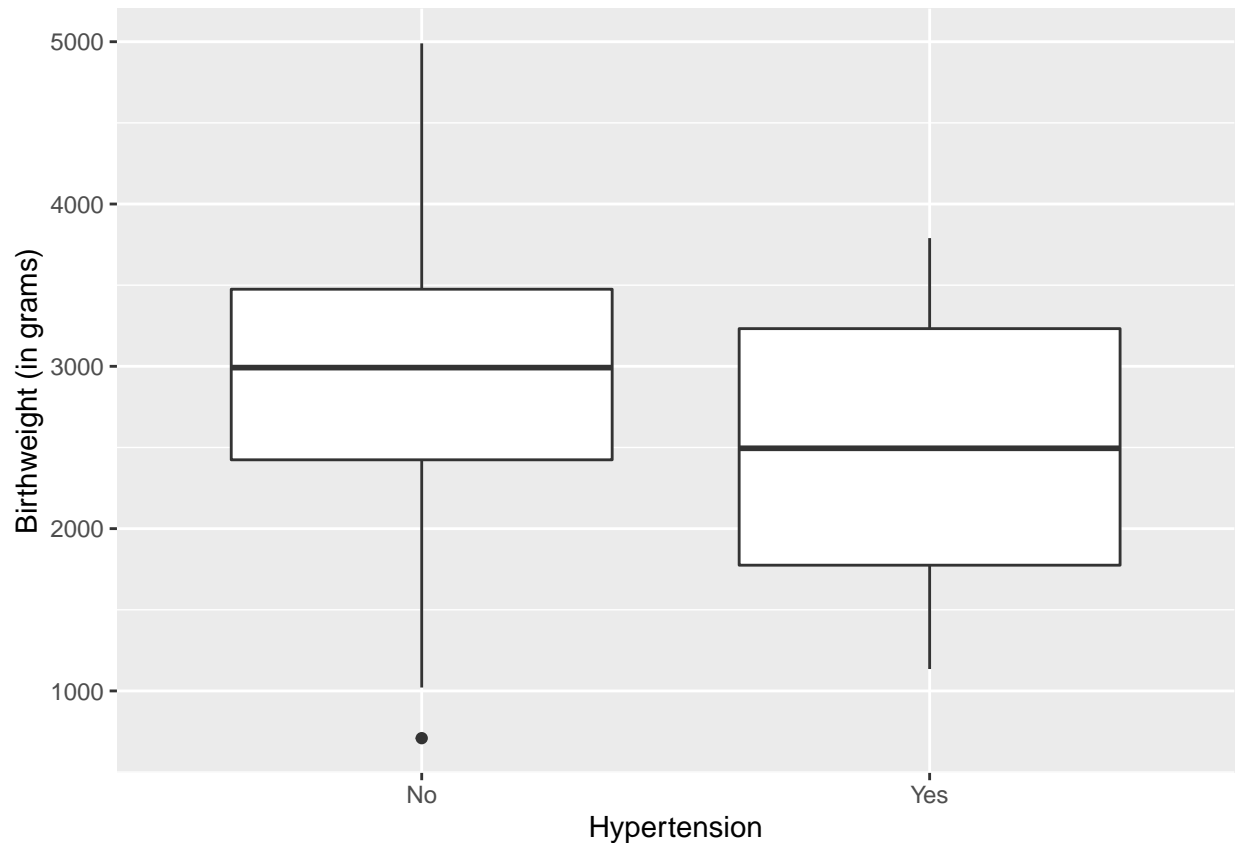

```
lowbwt %>% ggplot(aes(y = bwt)) +
  geom_boxplot() +
  labs(y = "Birthweight (in grams)")
```



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(y = "Birthweight (in grams)", x = "Hypertension")
```



*#From the boxplot, we can see that majority of the samples of "Yes" category is having
 #median at around 2500 birthweight while mothers with no hypertension
 #is having median at around 3000 birthweight.
 #There is an outlier in the "No" category that having baby with extremely
 #low weight (lower than 1000 birthweight) while there is no outlier in the "Yes" category.
 #Both the category seems like a symmetrical distribution.
 #However, the data for mothers with hypertension is extremely
 #small compared to the data of mothers without hypertension.
 #Hence, it is very uncertain as to the shape of the population distribution.
 #Boxplots are poor in small sample size data as they are a very simplified summary.*

*# Furthermore, the "No" category covered a larger range
 #(whisker tip to whisker tip) while the "Yes" category
 #is having larger interquartile range (the box).
 #Both categories data seem dispersed.*

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.

In conclusion, although we have a small dataset on mother with hypertension but the above charts show us that mother has hypertension is most probably to have low birthweight babies compared to mother who smoked or has uterine irritability with more than 50% of proportion in the barchart. Hence, predict low birthweight of babies could possibly be based on hypertension status of mother.

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