birthwt Dataset

The Dataset

The *birthwt* dataset—stored in the *MASS* library—looks at the variables pertaining to low birth weights.

- *low*: An indicator showing if the child's birth weight was less than 2500 grams.
- *age*: The mother's age in years.
- *lwt*: The mother's weight in pounds at the last menstrual period.
- *race*: An indicator identifying the mother's race (1=white, 2=black, 3=other).
- *smoke*: An indicator recording the mother's smoking status during pregnancy.
- *ptl*: The mother's number of previous premature labors.
- *ht*: An indicator recording if the mother has a history of hypertension.
- *ui*: An indicator recording the presence of uterine irritability
- *ftv*: The mother's number of physician visits during the first trimester.
- *bwt*: The child's birth weight in grams.

Potential Questions

Like many of the best datasets, so many of these variables are intertwined. Even better—in a sense—is that there are a number of potential confounding factors, particularly socioeconomic ones that aren't recorded about our observations that could probably help explain our data better. However, even without that information there are many good questions to ask and try to answer, even if that answer is likely incomplete without information about those confounders.

1. Do white mothers have different rates of uterine irritability than non-white mothers? (2-sample proportion)

2. Do smokers have a history of hypertension? (2-sample proportion)
3. Does having previous premature labors affect the chances of uterine irritability? (2-sample proportion)
4. **Does having previous premature labors affect the chances of low birth weights? (2-sample proportion)**
5. Is the average birth weight of children whose mothers had doctors visits during the first trimester higher than those who didn't? (2-sample mean)
6. Does smoking affect child birth weight? (2-sample mean)
7. **Does hypertension affect child birth weight? (2-sample mean)**
8. Are mother's age and child birth weight associated? (Correlation)
9. **Are mother's weight and child's birth weight associated? (Correlation)**
10. What factors best predict the chances of a low birth weight?*
11. What factors best predict a child's birth weight*

** I would denote this a challenge question. Challenge questions are ones that, due to the open-ended nature of the question or coding that can stretch the student or may not even be in book, challenges students to a certain degree.*

Question 4

It seems reasonable that a history of premature births would be a likely indicator of future premature births. These premature births would likely be indicated by low birth weights, at least with high probability. In this way, we can try to see if having previous premature births leads to a higher chance of low birth weights. As the *ptl* variable only takes on integer values between 0 and 3 in this dataset, we can start a little exploratory analysis with a contingency table.

	0	1	2	3
Non-low birth weight	118	8	3	1
Low Birth Weight	41	16	2	0

In answering this question, we'll combine the *ptl* variable into two groups: those who had no previous premature labors and those who did. This will allow us to look at things using a two-sample test for proportions. This particular contingency table winds out being

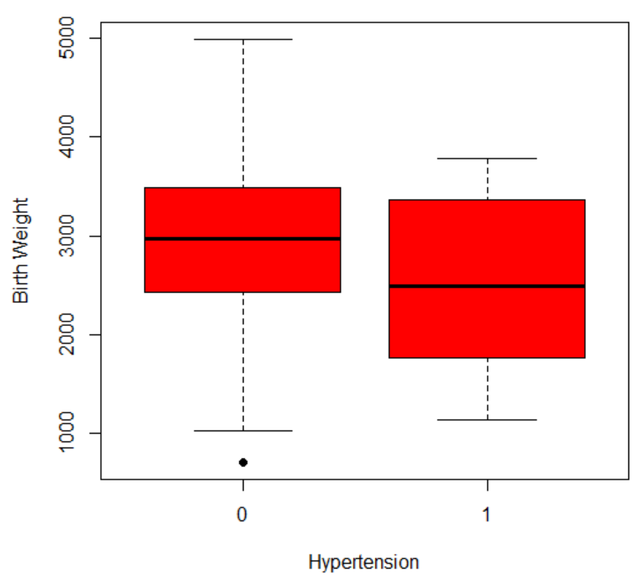
	Previous Premature	No Premature
Non-low birth weight	118	12
Low Birth Weight	41	18

From here, we'll move to testing $H_A : p_{NoPreviuos} - p_{Previous} < 0$, where p is the proportion of births with a low birth weight. We'll test this at the significance level $\alpha = 0.05$. The test returns a test statistic of $t = -\sqrt{13.759} = -3.709$ with a p-value of $P\big(N(0,1) < -3.709\big) = 0.0001$. Thus, we'll reject the null hypothesis and conclude that the population proportion of low birth weights for those with previous premature labors is greater than for those without.

Our sample size assumptions are somewhat questionable here, although technically they're good. Ideally, we'd want to see more data to gain a little more confidence in these assumptions. For a more complete answer of this question, we'd have to look at a logistic regression predicting the probability of low birth weight, or alternatively shift toward predicting the actual birth weight from the number of past premature labors.

```
library(MASS)
data(bwt)
table(bwt$low, bwt$ptl)
prop.test(x=c(41,18),
n=c(41+118,18+12),
alternative="less",correct=F)
-sqrt(13.759)
```

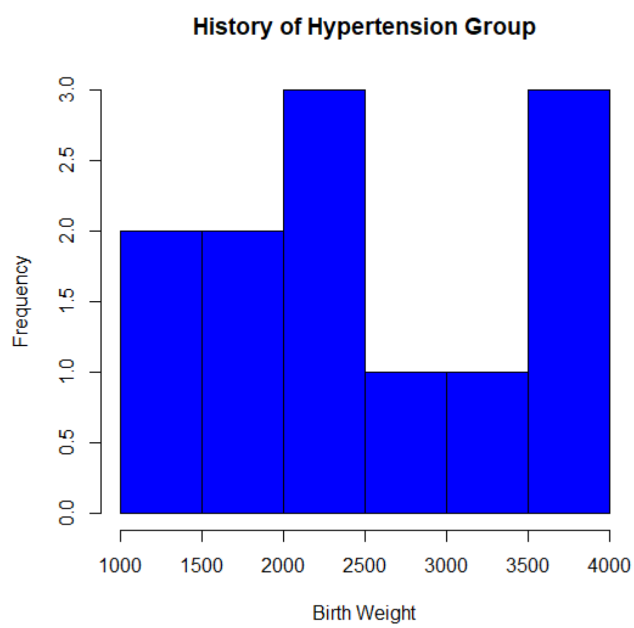
Question 7



Hypertension—or high blood pressure—is one of those things that is related to many things and can affect just as many. With this in mind, it's reasonable to wonder if a mother's hypertension could affect their child's birth weight. A quick look at a boxplot of the data suggests it's possible, even if this affect is due to a confounding factor.

For this test, we'll test $H_A : \mu_{NoHBP} - \mu_{HBP} > 0$, where μ is the population average birth weight, at the $\alpha = 0.01$ level. As always, we check to see if the variances for our two groups are equal, which they indeed are in this question. The test gives us a test statistic of $t = 2.0179$ with a p-value of $P(t_{187} > 2.1079) = 0.02252$. Thus, we'd fail to reject the null hypothesis and conclude that it's plausible and mothers with histories of hypertension and those without have the same population average child birth weight. Although, given how close the p-value is to the significance level, more data might get this p-value to a statistically significant level.

It's important to note that the assumptions tied to this test are definitely in question, specifically for the group with a history of hypertension. The sample size of this group is only $n_{HBP} = 12$, well below our usual standard, and the data for this group looks decidedly non-Normal. Again, more data would likely tip our p-value past the significance level, while also providing us with enough data to meet our assumptions.

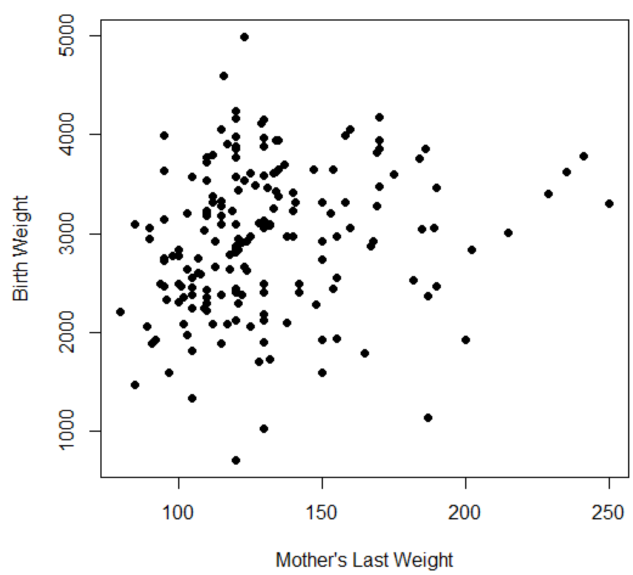



```
library(MASS)
data(birthwt)
boxplot(bwt~ht, data=birthwt,
xlab="Hypertension",ylab="Birth
Weight",col="red",pch=16)
var(birthwt$bwt[birthwt$ht==0])
var(birthwt$bwt[birthwt$ht==1])
t.test(bwt~ht,data=birthwt,
```

```
alternative="greater",var.equal=T)
table(ht)
hist(birthwt$bwt[birthwt$ht==1],
col="blue",xlab="Birth Weight",
main="History of Hypertension Group")
```

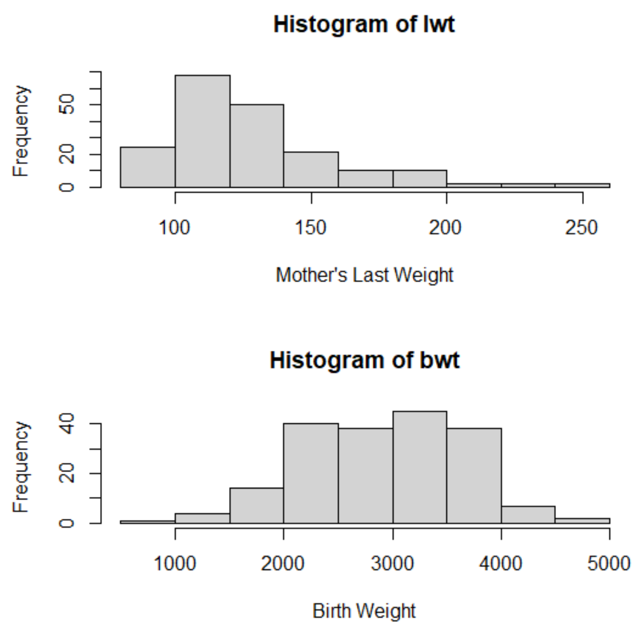
Question 9

In many instances, weight is affected by family history. When that affect starts taking hold, if it is there, is hard to say, so it seems reasonable to ask if mother's weight affects child's birth weight. Now, there is an obvious factor here that is missing: father's weight. However, it wouldn't be outlandish to think that child's weight is correlated with both of these factors. That said, a scatterplot of mother's weight and birth weight seem to suggest a weak correlation.



Since both variables are numeric, we'll use our test for correlation to try and establish a relationship between these two variables. Testing $H_A : \rho \neq 0$ at the $\alpha = 0.05$ level. The test returns a test statistic of $t = 2.5848$ with a p-value of $P\big(t_{187} > 2.5848\big) = 0.0105$, thus we'd reject the null hypothesis and conclude that mother's weight and child's birth are indeed correlated.

Looking at our assumptions, while birth weight is conceivably Normal, the mother's last weight seems slightly—and I do mean ever so slightly—skewed to the right. The main clue to this is the extremely long tail noticeably missing on the left. Formal tests for Normality exist, but we won't go into them. However, though the sample size may be large enough to overcome this slight skew, this assumption's potential violation should be at least noted accordingly.




```
library(MASS)
data(birthwt)
plot(birthwt$lwt,birthwt$bwt,pch=16,
xlab="Mother's Last Weight",ylab="Birth Weight")
cor.test(lwt, bwt, data=birthwt)
par(mfrow=c(2,1))
hist(lwt,xlab="Mother's Last Weight")
hist(bwt,xlab="Birth Weight")
```