Assignment 2 Emily Chu

Code

This is an R Markdown (http://rmarkdown.rstudio.com) Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Cmd+Shift+Enter*.

r r plot(cars)

Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Cmd+Option+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Cmd+Shift+K* to preview the HTML file).

Data frame for this assignment: lowbwt

1. List the level of measurement for each of the following variables:

a. LOW: Ordinal-level
b. SMOKE: Ordinal-level
c. BWT: Interval-Ratio

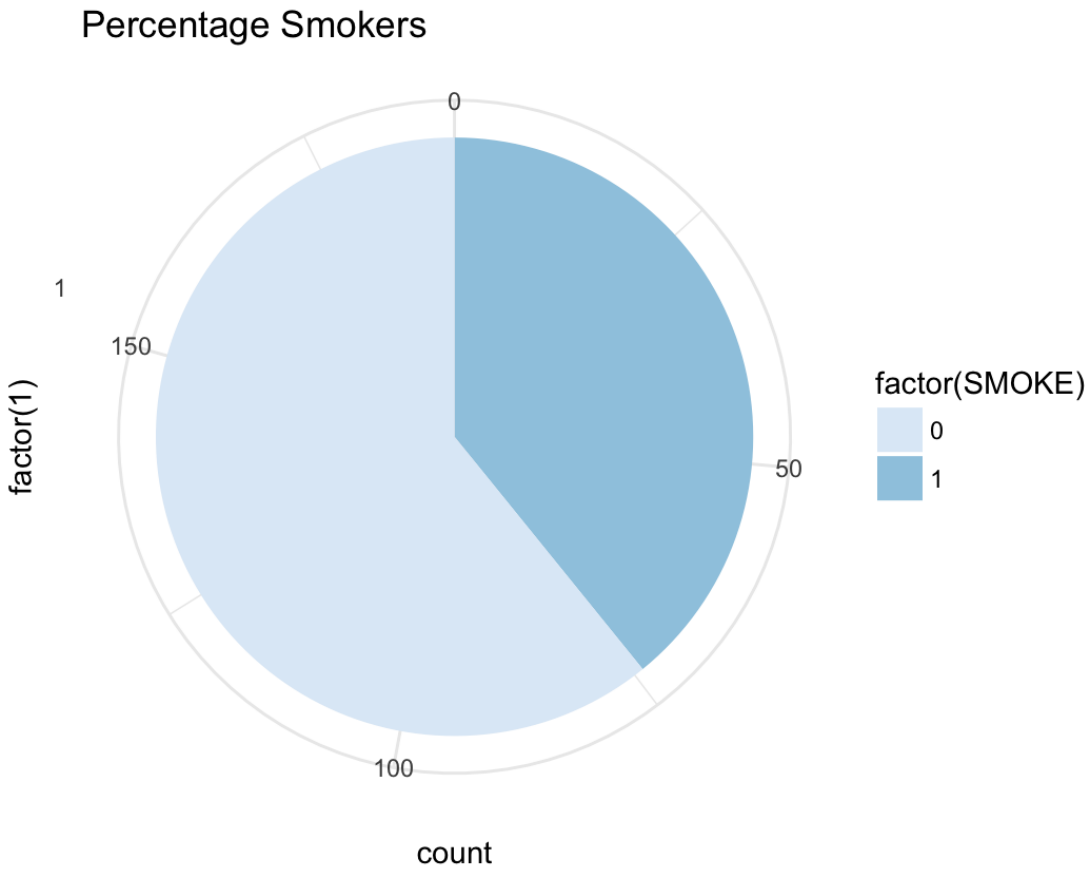
2. Show a frequency table and an appropriate graph for the variable Smoking Status During Pregnancy.

r r table(lowbwt$SMOKE)

```
  0   1
115  74
```

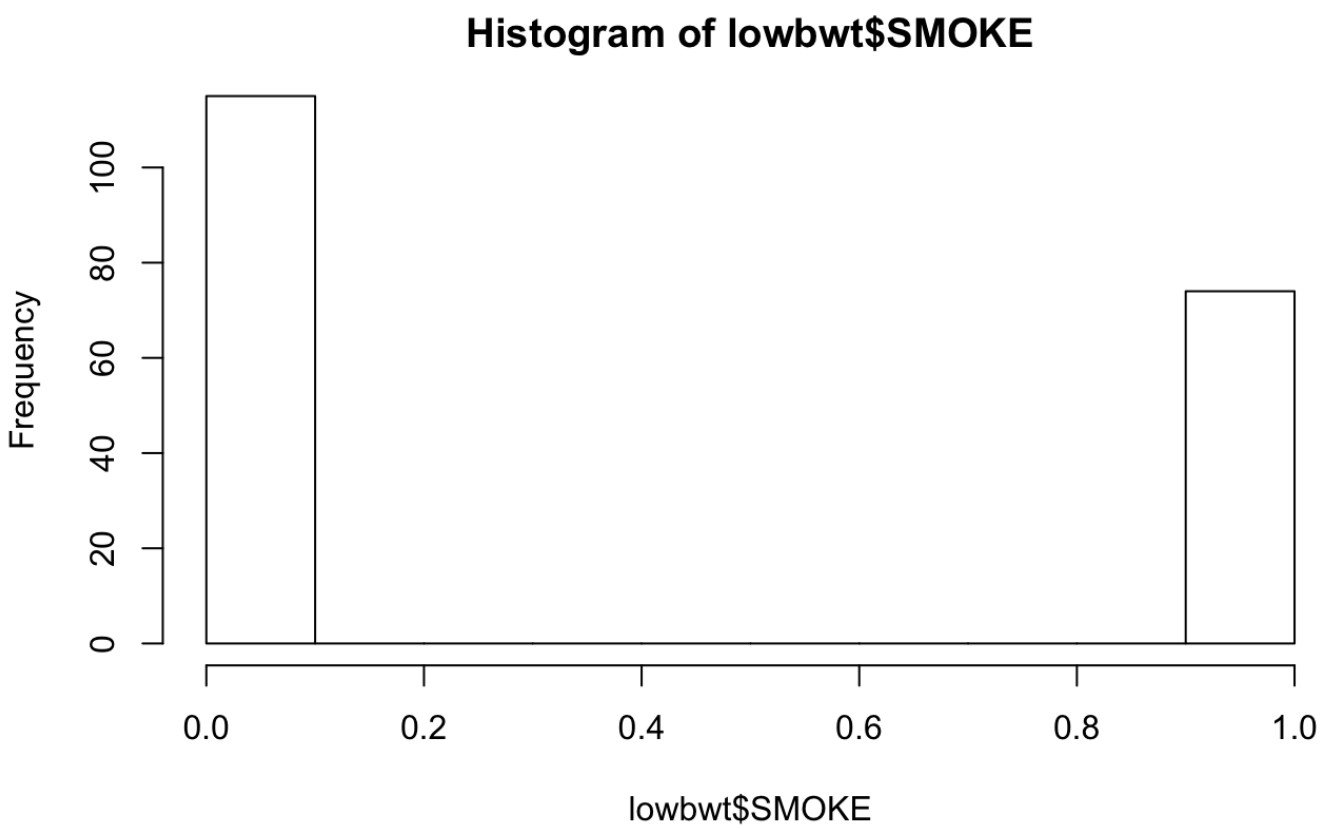
r r prop.table(table(lowbwt$SMOKE))

```
        0         1
0.6084656 0.3915344
```



3. Describe the distribution of Smoking Status During Pregnancy.

Smoking Status During Pregnancy is a ordinal-level variable. Thus, mean and median are not calculated. However, other descriptive statistics can be used for this variable. Around 60% of mothers in the dataset are non-smokers while 40% are smokers. The mode is non-smoking mothers with a frequency of 115. If a histogram was created, it would look like the below.



Hide

```
names(sort(-table(lowbwt$SMOKE)))[1]
```

```
[1] "0"
```

4. Show a frequency table and an appropriate graph for the variable Low Birth Weight.

Hide

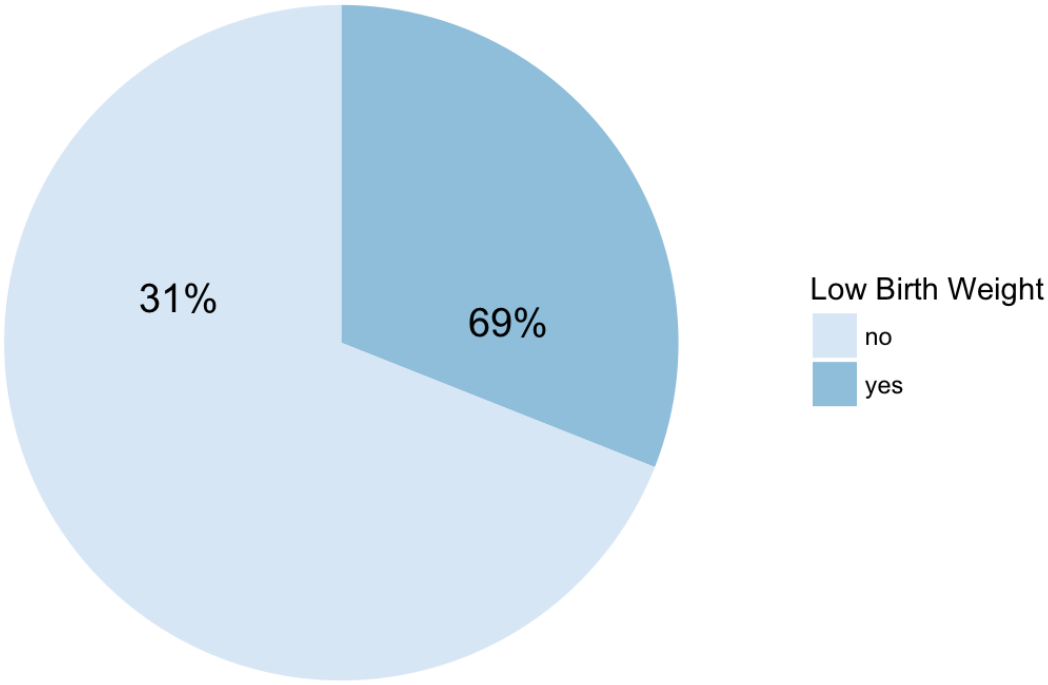
```
table(lowbwt$LOW)
```

```
  0   1
130  59
```

Hide

```
prop.table(table(lowbwt$LOW))
```

```
        0         1
0.6878307 0.3121693
```

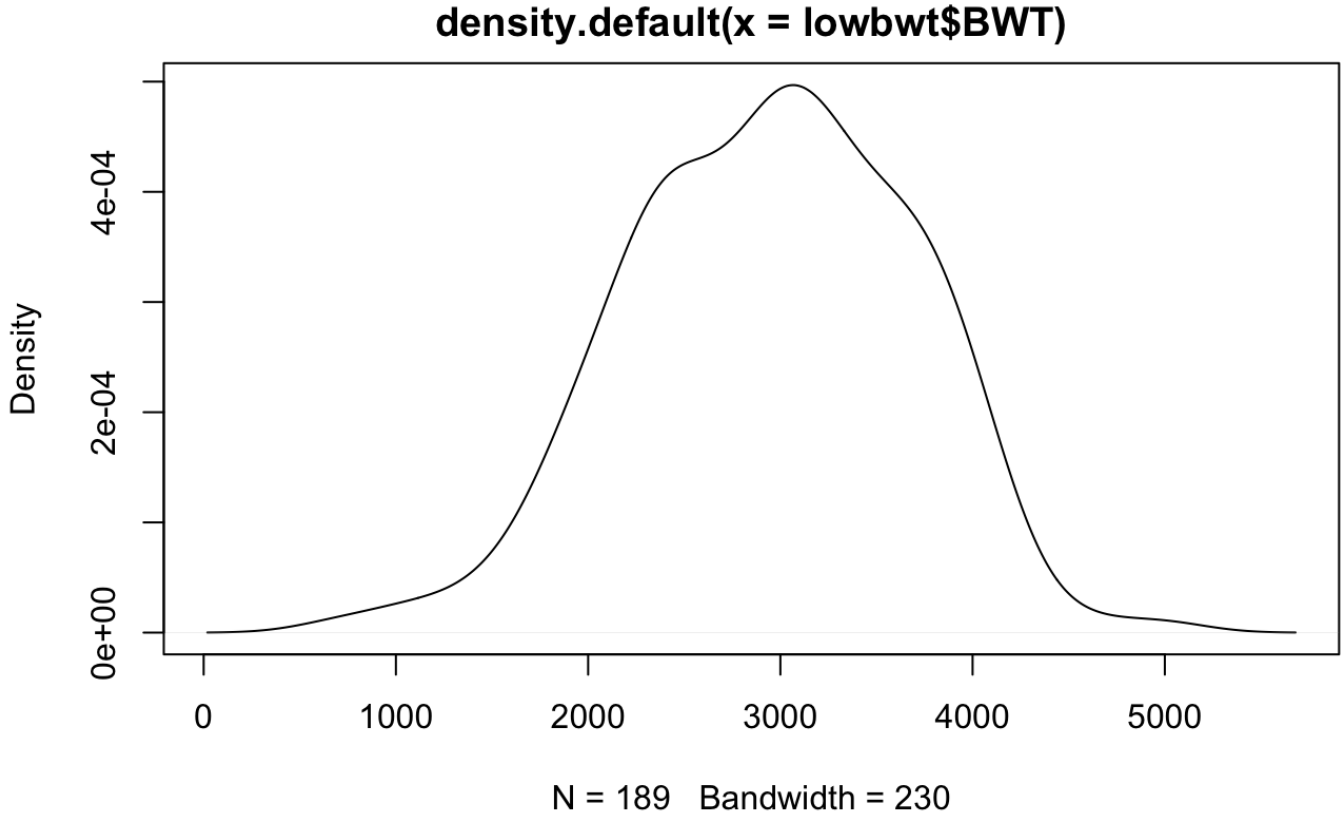


5. Show summary statistics and an appropriate graph for Birth Weight in Grams(BWT).

Hide

```
summary(lowbwt$BWT)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
    709    2414    2977    2945    3475    4990 
```

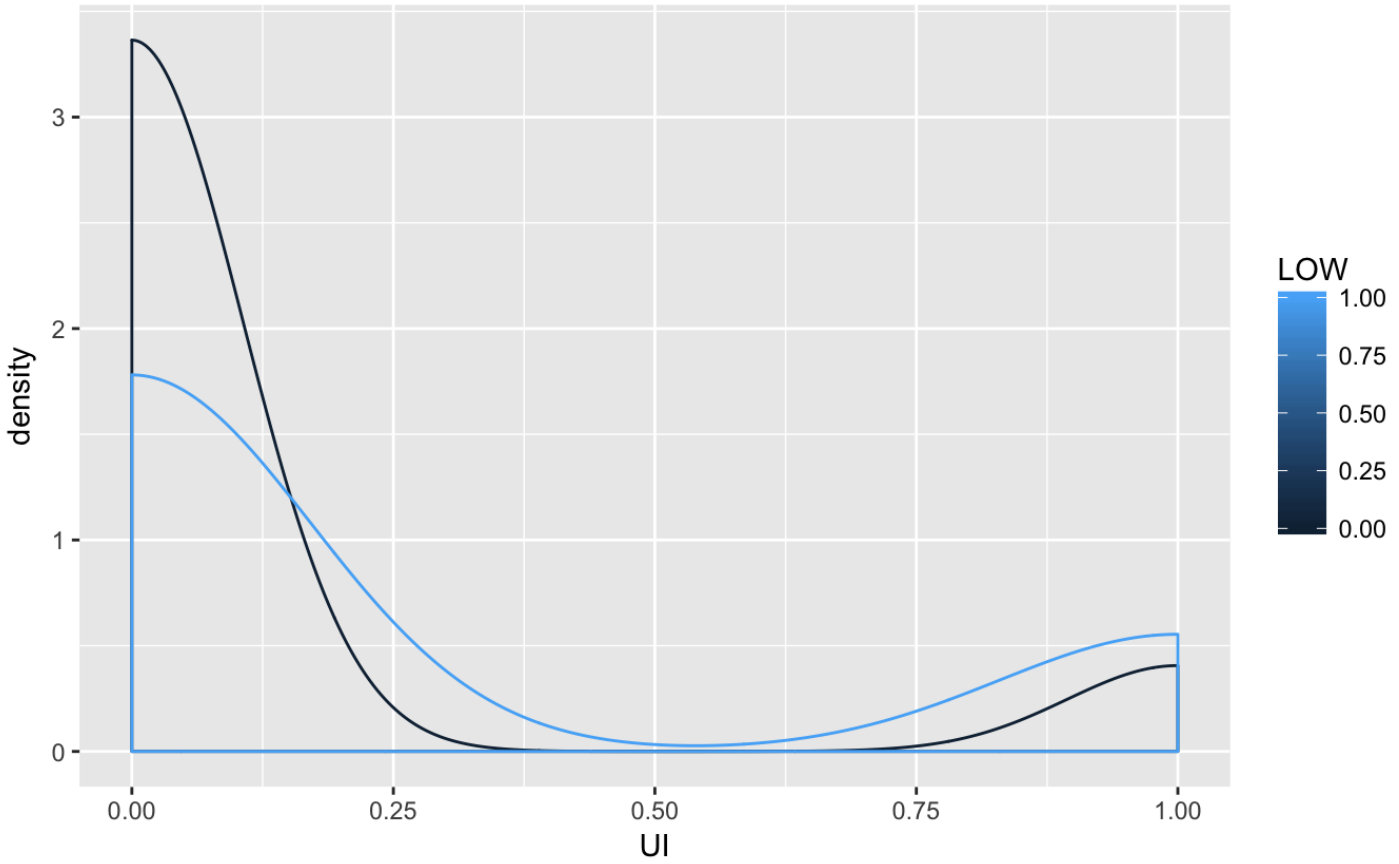


6. Describe the distribution of Birth Weight in Grams(BWT).

The distribution of BWT is almost normal, only very slightly skewed negatively to the left, with a mean of 2945g.

7. Create a kernal density plot to describe the distribution of Birth Weight in Grams (BWT) for those who suffer from uterine irritability and then for those who do not.

Try to use the code to overlay these plots on top of each other. Describe the pattern. The uterine irritablity variable is called UI. Remember you need install and library packages ggplot2.packages before creating this kernal density plot.



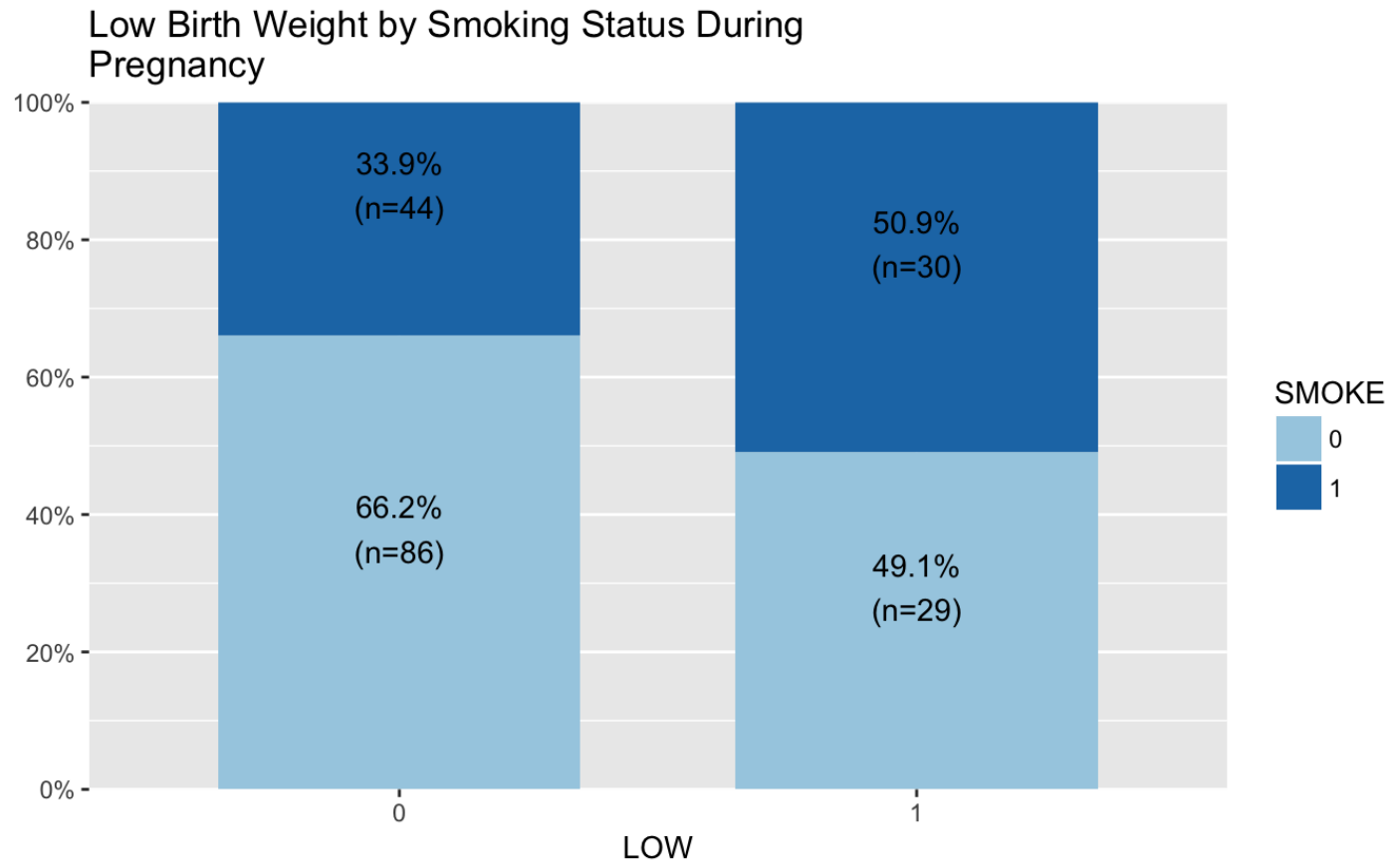
Hide

```
summary(lowbwt$UI)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
 0.0000  0.0000  0.0000  0.1481  0.0000  1.0000 
```

8. Show a crosstab of Low Birth Weight BY Smoking Status During Pregnancy.

Be sure to show and interpret the percentages (interpreting the column percenages is sufficient)



9. Does there appear to be a relationship between Low Birth Weight and Smoking Status During Pregnancy? Uterine Irritability? Explain.

There does appear to be a relationship between Low Birth Weight and Smoking Status during Pregnancy: out of the non-low weight babies, 33.9% of them came from smokers. Out of the low weight babies, on the other hand, 50.9% of them came from smokers.

There also seems to be a relationship between Low Birth Weight and Uterine Irritability. In the kernal density graph, there is a higher density of low weight babies (LOW=1) if there is presence of urine irritability, and there is a higher density of non-low weight babies if there is no presence of urine irritability.

Submit to Canvas: your notebook. The notebook should include your code, the output and (importantly!) your interpretation of the findings.

Advanced material

Here's an example of a density plot overlay which also shows the mean of two groups. Here I use the drug treatment dataset which we worked with last time. The example looks at the beck depression scale by site. Study participants were not randomized by site. In this example we show the mean on the plot itself.

	SITE <fctr>	**grp.mean** <dbl>
1	0	17.90705
2	1	16.13400

2 rows

Hide

```
library(plyr)
library(ggplot2)
drugtreatment$SITE <- factor(drugtreatment$SITE) # set site as factor rather than interv
al ratio variable
becsite <- ddply(drugtreatment, "SITE", summarise, grp.mean=mean(BECK))
head(becsite)

c<-ggplot(drugtreatment, aes(x=BECK, group=SITE, col=SITE)) +
  geom_density(size=1)+
  geom_vline(data=becsite, aes(xintercept=grp.mean, color=SITE),
             linetype="dashed", size=1)
c
```