STA 303H1S / 1002 HS -Winter 2018 Assignment # 2
What factors affect baby birth weight?

Due: In Crowdmark via Blackboard by 10pm on Tuesday, February 13, 2018.
Late assignments will be subjected to a penalty of 5% per hour late.

Grading: The grand total for this assignment is 100 marks.

Instructions:

- Use R (or R Studio) to do the analysis for the following questions.

- Use a benchmark significant level of 5%.

- Compile your solution as a PDF document (Word, LaTeX or Rmarkdown can be your base).

- Presentation of solutions is very important. Your assignment should have two main sections- Solutions and Appendix. Include relevant plots and quote relevant numbers from your R output for your solutions. In the Appendix, include your R code and other output. A maximum of 10 marks will be awarded for excellent presentation.

- Write and submit **your own work**. For instance, personalized your code as much as possible, using your first name. **All plots produced must be given a title with the last 4 digits of your student number**.

- Where appropriate, your answers are expected to be written in plain English.

The Data
The source of these data is *Stat Labs* by Nolan and Speed. For the purposes of this assignment a subset of the full data is used and can be found in a file named "bbw.csv" on Blackboard.

This data was used to compare the birth weight of babies born to mothers who were smokers to those whose mothers were nonsmokers in order to determine whether they corroborate the Surgeon General's warning that: "Smoking by pregnant women may result in fetal injury, premature birth, and low birth weight." Our data consists of measurements from 409 babies. Many other questions were investigated as part of this study, however we will focus on the following two questions:
1. Does smoking by mother during pregnancy affect baby birth weight?
2. Does baby birth weight change with gestation maturity?

The variables in the dataset are:

- `bwt`- baby's weight at birth in ounces

- `gestation`- number of days spent in the womb

- `smoke`- an indicator variable which is 1 if the baby's mother smoked and 0 if she did not smoke during pregnancy.

1. (15 marks) Create two new variables: (1) `maturity`- by converting gestational age to a factor with 3 levels; 1 if the baby was preterm and spent less than 259 days in the womb, 3 if gestational age was beyond 293 and 2 otherwise, and (2) `MatSmoke`- a variable that combines maturity level and maternal smoking status. You can use the following R code to do this:

```
maturity=array(0,length(gestation))
MatSmoke=array(0,length(smoke))
for (i in 1:length(gestation))
{
if (gestation[i]<259)
   {maturity[i]=1}
else if (gestation[i]>293)
   {maturity[i]=3}
else {maturity[i]=2}
}
for (i in 1:length(smoke))
{
   if (maturity[i]==1 & smoke[i]==1)
   {MatSmoke[i]="PreSmoke"}
   else if (maturity[i]==1 & smoke[i]==0)
   {MatSmoke[i]="PreNoSmoke"}
   else if (maturity[i]==2 & smoke[i]==1)
   {MatSmoke[i]="NorSmoke"}
   else if (maturity[i]==2 & smoke[i]==0)
   {MatSmoke[i]="NorNoSmoke"}
   else if (maturity[i]==3 & smoke[i]==1)
   {MatSmoke[i]="PostSmoke"}
   else {MatSmoke[i]="PostNoSmoke"}
}
```

Construct three sets of side-by-side boxplots: 1. to compare birth weight between mothers who smoked and those who did not smoke during pregnancy, 2. to compare birth weight among the three maturity levels, and 3. to compare birth weight among the 6 categories of babies grouped by the combination of their maturity level and maternal smoking status. Do there appear to be any differences?

2. (10 marks) Using the R t.test procedure, investigate whether or not there is a difference in the mean birth weight between babies born to mothers who were smokers and babies born to mothers who were nonsmokers.

3. (15 marks) Investigate whether or not there is a difference in mean birth weight among babies classified by gestational maturity, using a one-way analysis of variance. If there is a difference among the levels of maturity, carry out an appropriate analysis to see which levels of maturity differ.

4. (15 marks) Use one-way analysis of variance to investigate whether or not there is a difference in mean birth weight among the six categories of babies classified by the combination of their maturity level and mother's smoking status. If there is evidence of differences among the six categories of babies, carry out an appropriate analysis to see which differ.

5. (10 marks) Do you trust the results of the statistical tests carried out in question 4? Assess whether the necessary assumptions of the model hold.

question of assumption

6. (10 marks) Instead of the one-way classification model used in question 4, a two-way analysis of variance model could have been used with maternal smoking status, maturity level and their interaction. WITHOUT fitting this model, answer the following questions.

   (a) Would the number of predictor variables be the same as in the model used in question 4? Why or why not?

   (b) Would the F-test for the presence of interaction between maturity level and smoking status be statistically significant? How do you know from your results of question 4?

7. (5 marks) Should we be concerned that the data contained different numbers of babies in the three maturity levels? Why or why not?   question of assumption

8. (5 marks) Discuss the use of gestation as a quantitative explanatory variable rather than as a factor in an additive linear model for mean birth weight. Include mathematical equations to describe the difference in models for mean birth weight.

9. (5 marks) Name two additional potential factors of baby birth weight and briefly describe their levels.