Effects of Smoking on Birth Weight of Babies

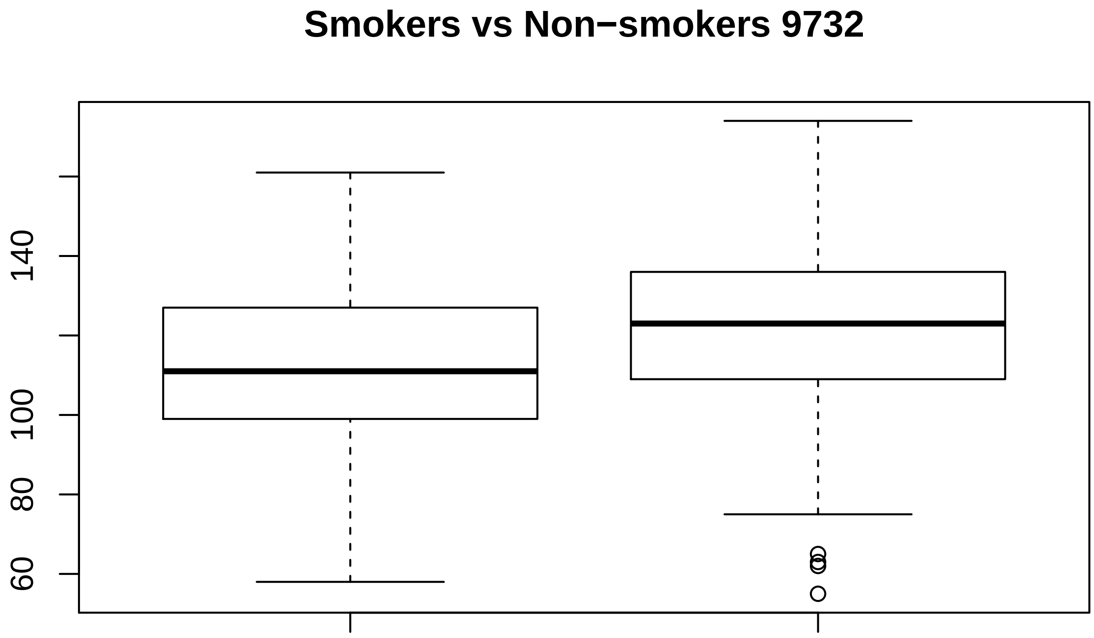
Ted Si Yuan Cheng

2018-02-04

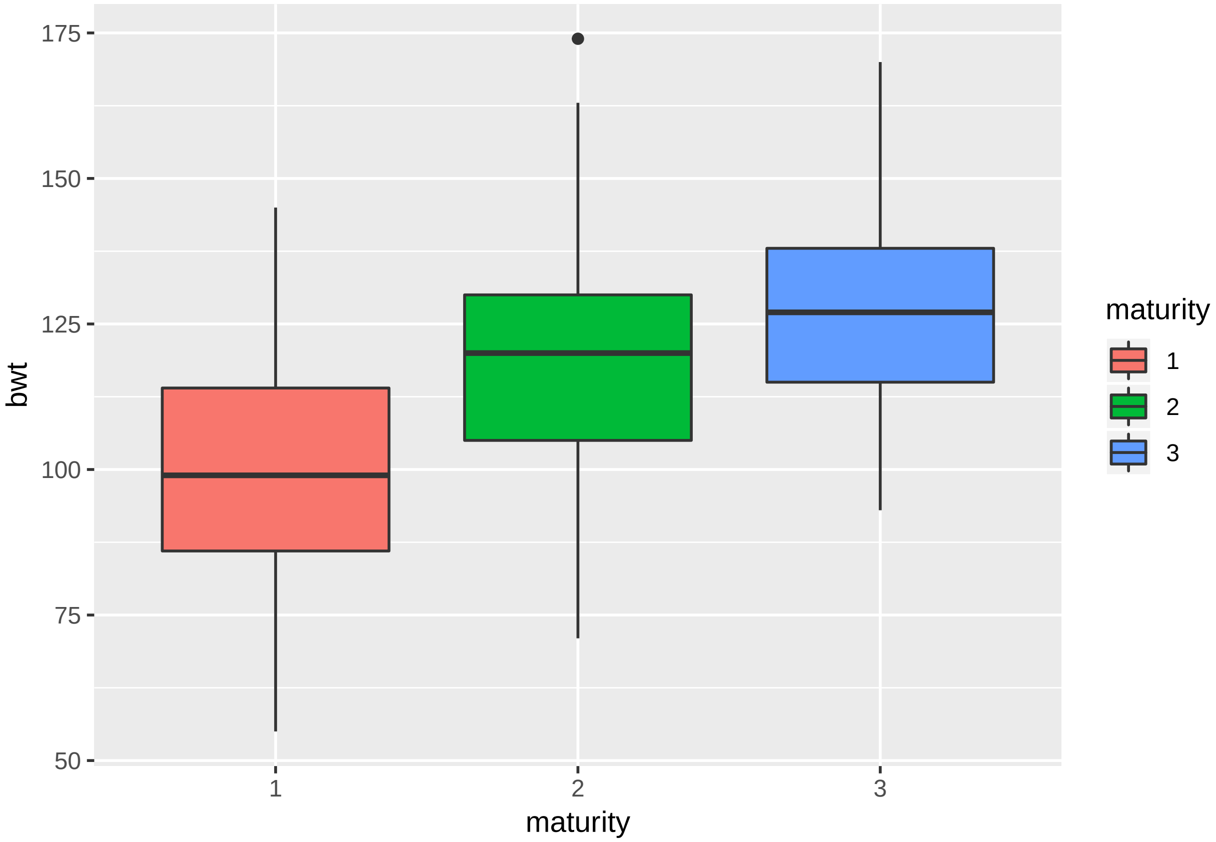
Question 1

Compare birth weight between mothers who smoked and those who did not smoke during pregnancy, 2. compare birth weight among the three maturity levels, and 3.compare birth weight among the 6 categories of babies grouped by the combination of their maturity level and maternal smoking status.

Boxplot of birth weight between mothers who smoked and didn't smoke

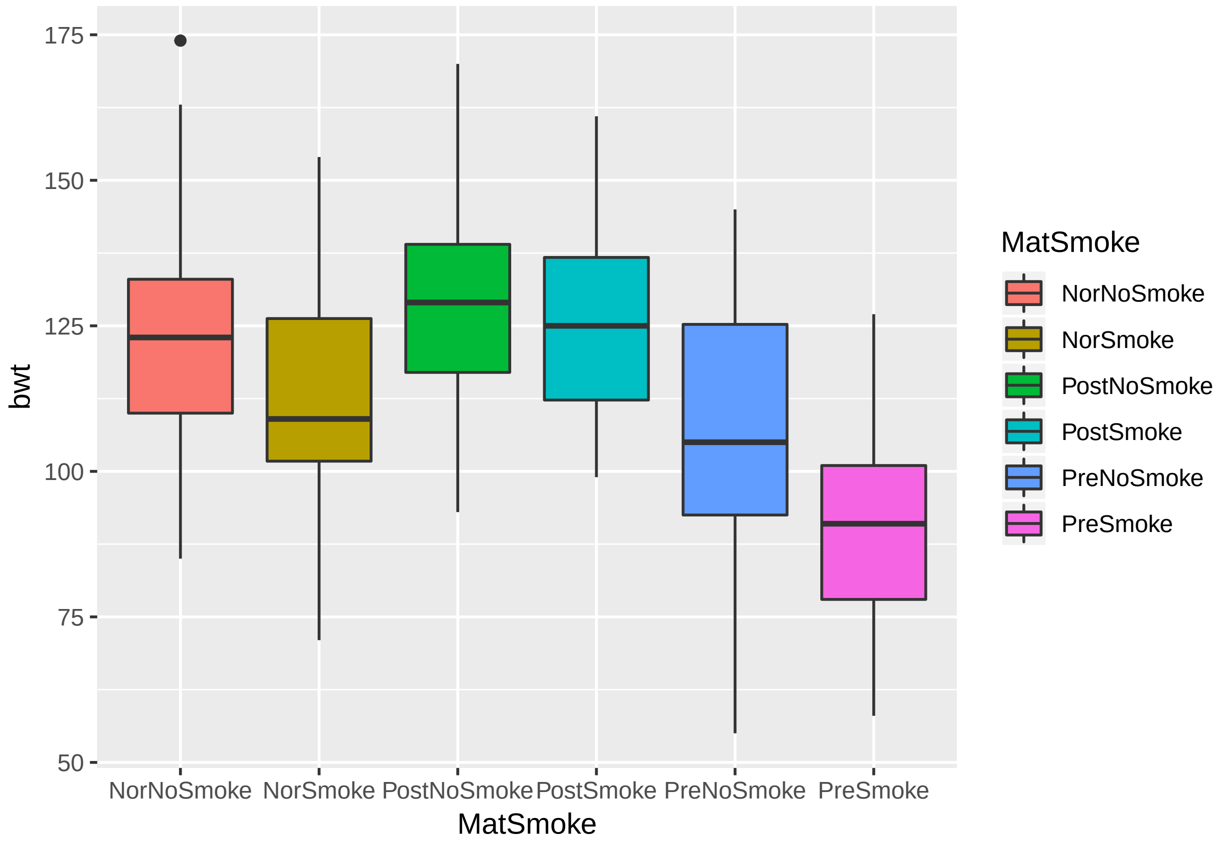


Boxplot of birth weight among three maturity levels 9732



Body weight seems to increase proportionally to maturity levels

Boxplot of birth weight among babies grouped by combination of their maturity level and maternal smoking status 9732



The non-smoker's birth weights appear to higher than their respective group of smokers. Pre-pregnancy smoking seems to cause the lowest birth weight

Question 2

Whether or not there is a diference in the mean birth weight between babies born to mothers who were smokers and babies born to mothers who were nonsmokers with t.test.

The true difference in mean between the smokers and non smokers are not equal to 0 ($p > 0.05$), so we reject $H_0 = 0$ which states there is no difference in the mean birth weight between babies born to smoking versus nonsmoking mothers.

Question 3

Whether or not there is a difference in mean birth weight among babies classified by gestational maturity, using a one-way analysis of variance.

The one-way ANOVA of body weight versus gestation shows that there is a difference as $p < 0.05$.

The post-hoc Bonferroni test shows there is a difference among the mean levels of maturity between all level as all values of $p < 0.05$.

Quesiton 4

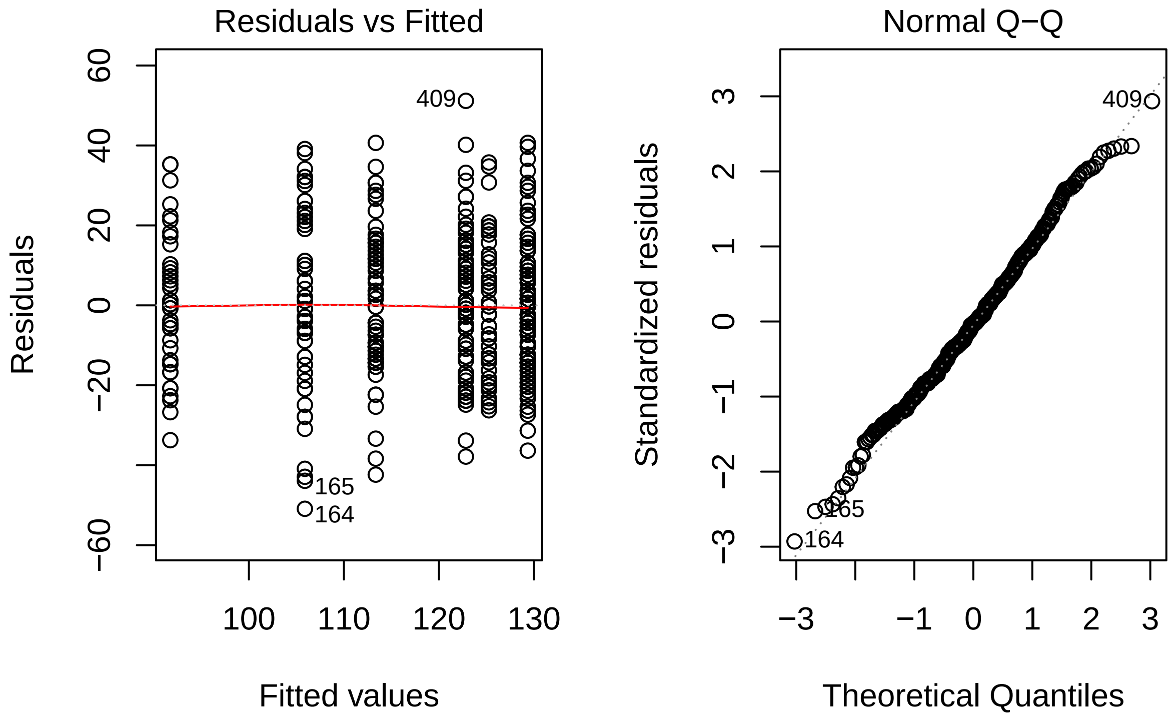
Use one-way analysis of variance to investigate whether or not there is a difference in mean birth weight among the six categories of babies classified by the combination of their maturity level and mother's smoking status.

The one-way ANOVA of body weight versus maturity level and smoking status shows that there is a difference as p < 0.05.

The post-hoc Bonferroni test shows that there is a difference between NorSmoke~NorNoSmoke, PreNoSmoke~NorNoSmoke, PreSmoke~NorNoSmoke, PostNoSmoke~NorSmoke, PostSmoke~NorSmoke, PreNoSmoke~NorSmoke, PreNoSmoke~PostNoSmoke, PreSmoke~PostNoSmoke, PreNoSmoke~PostSmoke, PreSmoke~PostSmoke, PreSmoke~PreNoSmoke, as all p<0.05.

Question 5

Assess whether the necessary assumptions of the model hold.



The assumption of normality for the ANOVA test is met from the linear model qq plot showing normality. The assumption of homogeneous variance also seems to hold, as the spreads are quite close together in the residuals vs fitted graph. However, the test for independency is not as the Chi-Squared test showed p = 0.0004998, rejecting the null hypothesis that the baby weight among the six categories of maturity level and smoking status are independent.

The assumptions are met for the Bonferroni method as it can be used more generally than the Tukey.

Question 6

Would the number of predictor variables be the same as in the model used in question 4? Would the F-test for the presence of interaction between maturity level and smoking status be statistically significant?

a) The number of predictor variables would be more than the model in quesiton 4 as two-way anova will require two predictor variables.

b) Yes it would as from question 4 we see that many of the interactions between smokers and non-smokers at various stages of pregnancy show a significant difference.

Question 7

Should we be concerned that the data contained different numbers of babies in the three maturity levels?

No, as the Bonferroni method can be used with unequal sample sizes as well as equal sample sizes. Outside of the Bonferroni, this may affect our ability to do a t test for the mean difference of one level of bodyweight for the combination of MatSmoke to the other level, likewise for maturity. Doing a t.test with unequal sample sizes could result in inflated errors if the variances are not equal as well.

Question 8

Discuss the use of gestation as a quantitative explanatory variable rather than as a factor in an additive linear model for mean birth weight.

Gestation as a quantitative explanatory variable will yield the equation $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$. y_i denotes the birth weight of the i^{th} baby, x_{i1} is gestation age of the i^{th} baby, and x_{i2} is the categorical variabe. The dummy variable will be 1 if the mother smoked, or 0 if she did not. The mean response function is $\mu_{Y0} = \beta_0 + \beta_1 x_{i1}$ for mothers that did not smoke, and $\mu_{Y1} = (\beta_0 + \beta_2) + \beta_1 x_1$ for mothers that did smoke.

Gestation as a factor will yield the equation $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$. y_i denotes the birth weight of the i^{th} baby, x_{i1} is the maturity level based on the gestation age of the i^{th} baby. The dummy variable x_{i1} is 1 if the gestation age is less than 259 days , 3 if it is >293, and 2 otherwise. x_{i2} is the same categorical variable it was as the previous equation with the same levels. The mean response functions from this model will be in order of maturity level from 1 to 3 and nonsmokers are $\mu_{Y1} = \beta_0 + \beta_1$, $\mu_{Y2} = \beta_0 + 2\beta_1$, $\mu_{Y3} = \beta_0 + 3\beta_1$. For mothers who did smoke in order of maturity level 1 to 3 they will be $\mu_{Y1} = \beta_0 + \beta_1 + \beta_2$, $\mu_{Y2} = (\beta_0 + \beta_2) + 2\beta_1$, $\mu_{Y3} = (\beta_0 + \beta_2) + 3\beta_1$.

Question 9

Name two additional potential factors of baby birth weight and briey describe theirlevels.

One possible factor affecting birth weight could be the mother's nutrition intake during pregnancy. The possible levels could be 1 if the mother undereats, 2 if the mother eats a moderate amount, and 3 if the mother overeats. Hypothetically, a mother who undereats will have infants that will be underweight.

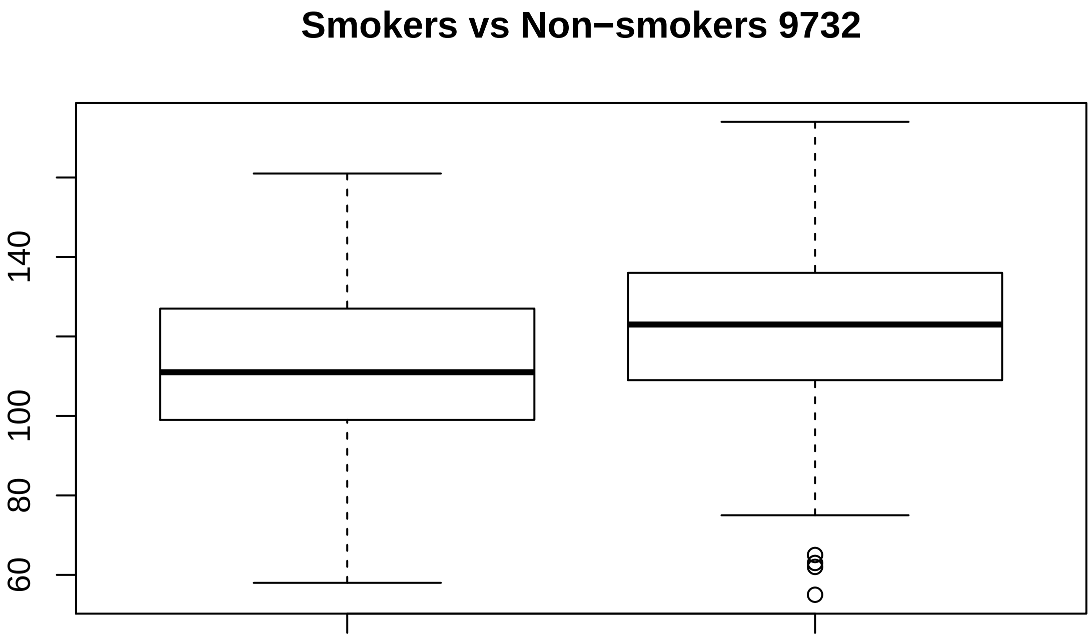
Another factor would be the age of the mother during pregnancy. Studies have shown that mothers who are older than 35 have larger babies, and teens with babies that are underweight. We can therefore have 3 levels as well, 1 being the age of any women/girl under the age of 18, 2 being the range of the age of women from 18 to 35, and 3 being the age of any women 35 and older.

Appendix

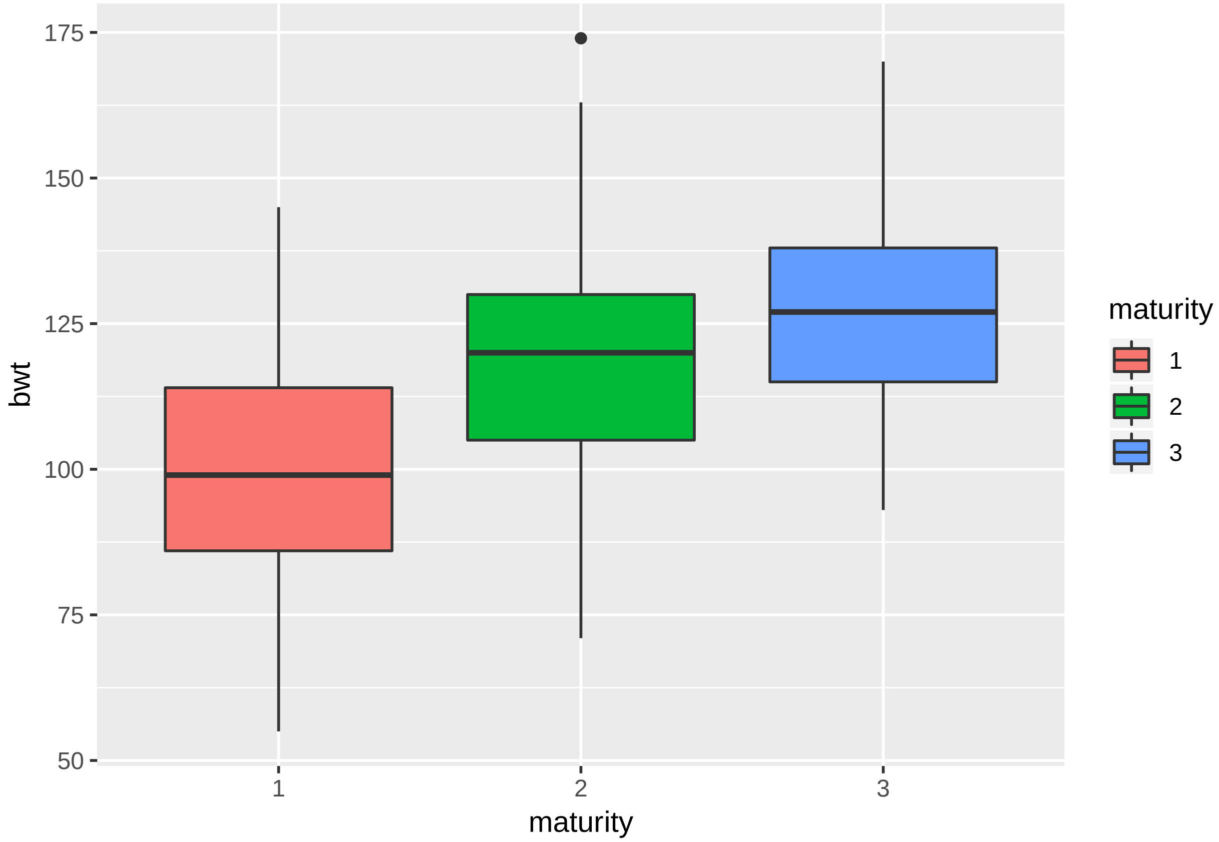
```
knitr::opts_chunk$set(echo = TRUE)
#Import data
bbw <- read.csv("C:\\Users\\Ted\\Documents\\R Projects\\2\\bbw.csv")
attach(bbw)
```

```
## The following objects are masked _by_ .GlobalEnv:
##
##     bwt, gestation, smoke

## The following objects are masked from bbw (pos = 4):
##
##     bwt, gestation, smoke
```

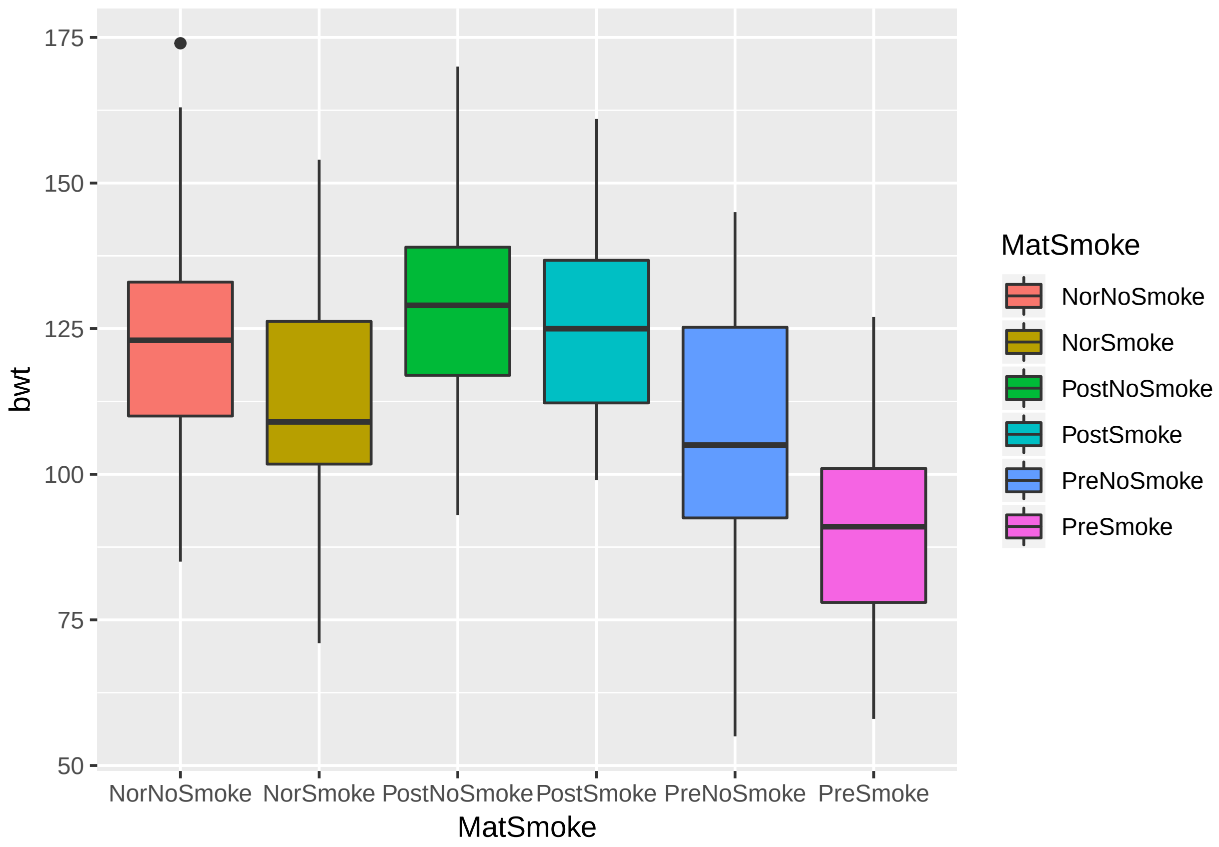
```
#Create variables
smoke <- bbw$smoke
gestation <- bbw$gestation
bwt <- bbw$bwt

#Convert to factors
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"}
}

maturity <- as.factor(maturity)

#Boxplot for smokers vs non-smokers
groupsmoke <- bwt[smoke==1]
groupnosmoke <- bwt[smoke==0]
boxplot(groupsmoke, groupnosmoke, main = 'Smokers vs Non-smokers 9732')
```



```r
#Boxplot for maturity levels
library(ggplot2)
ggplot(bbw, aes(x=maturity, y=bwt, fill=maturity))+geom_boxplot()
```



```
#Boxplot for maturity level and smoking status
ggplot(bbw, aes(x=MatSmoke, y=bwt,fill=MatSmoke))+geom_boxplot()
```



```
#T test for smokers vs nonsmokers
t.test(groupsmoke, groupnosmoke)
```

```
##
##  Welch Two Sample t-test
##
## data:  groupsmoke and groupnosmoke
## t = -4.6409, df = 333.49, p-value = 4.994e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -13.797101  -5.582669
## sample estimates:
## mean of x mean of y
##  111.8589  121.5488
```

```
#ANOVA and Bonferroni test
aov_m <- aov(bwt~maturity)
summary(aov_m)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## maturity      2  46586   23293   71.28 <2e-16 ***
## Residuals   406 132680     327
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
posthocm <- pairwise.t.test(bwt, maturity, p.adjust.method = 'bonf')
posthocm
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  bwt and maturity
##
##   1       2
## 2 1.4e-14 -
## 3 < 2e-16 3.8e-05
##
## P value adjustment method: bonferroni
```

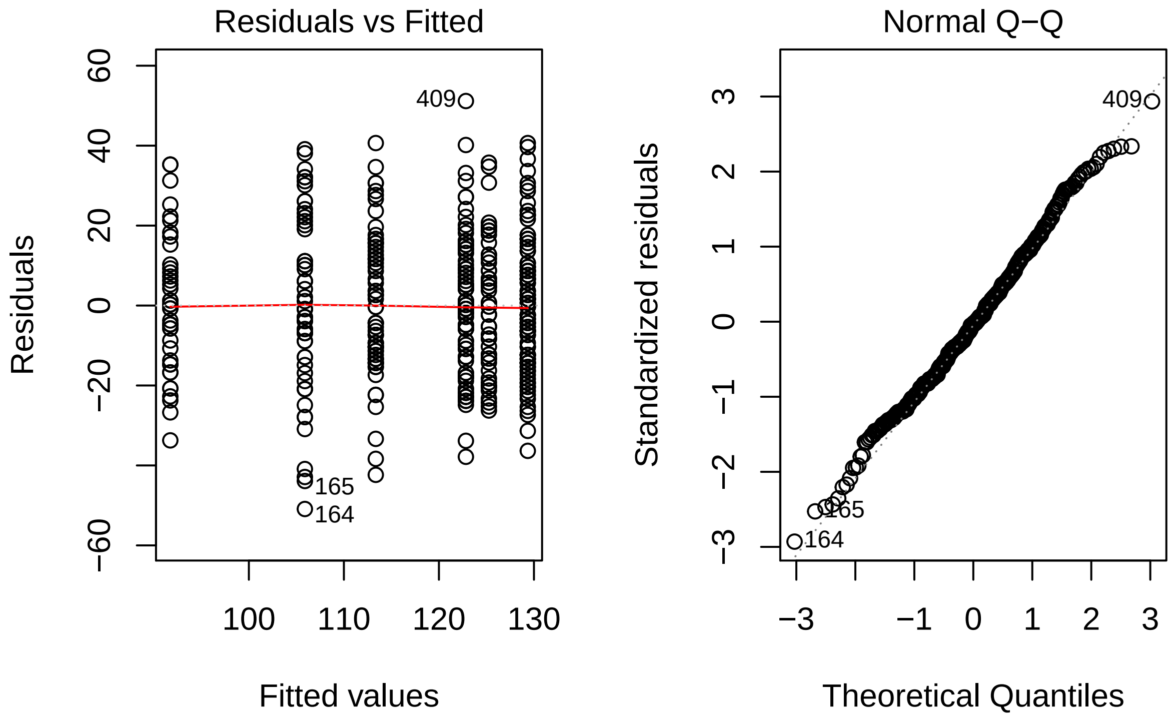
```
#ANOVA and Bonferroni test
aov_ms <- aov(bwt~MatSmoke)
summary(aov_ms)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## MatSmoke      5  55448   11090   36.09 <2e-16 ***
## Residuals   403 123818     307
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
posthocms <- pairwise.t.test(bwt, MatSmoke, p.adjust.method = 'bonf')
posthocms
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  bwt and MatSmoke
##
##             NorNoSmoke NorSmoke PostNoSmoke PostSmoke PreNoSmoke
## NorSmoke    0.0114     -        -           -         -
## PostNoSmoke 0.1625     2.4e-07  -           -         -
## PostSmoke   1.0000     0.0033   1.0000      -         -
## PreNoSmoke  3.2e-07    0.2824   2.4e-13     2.1e-07   -
## PreSmoke    < 2e-16    1.7e-08  < 2e-16     < 2e-16   0.0015
##
## P value adjustment method: bonferroni
```

```
#Check Bonferroni test assumptions
fitms <- lm(bwt~MatSmoke, data=bbw)
par(mfrow=c(1,2))
plot(fitms, which=1:2)
```



```r
chisq.test(bwt, MatSmoke, simulate.p.value = T)
```

```
##
##  Pearson's Chi-squared test with simulated p-value (based on 2000
##  replicates)
##
## data:  bwt and MatSmoke
## X-squared = 558.02, df = NA, p-value = 0.0009995
```