```r
library(ggplot2)

# read in data
bb = read.table('babies.txt', header=1)
bb$smoke[bb$smoke == 0] <- "Non-smoker"
bb$smoke[bb$smoke == 1] <- "Smoker"
bb <- bb[bb$smoke!=9,]

# remove extreme outliers
bb <- bb[bb$smoke!=9,]
bb <- bb[bb$weight<750,]
bb <- bb[bb$height<75,]
bb <- bb[bb$age<50,]
bb <- bb[bb$gestation<500,]

# boxplot data
bb = rbind(ns, s)
p = ggplot(bb, aes(x=smoke, y=bwt, group=smoke)) + geom_boxplot()
p + labs(title="Baby Weights in Smoking vs Non-smoking Mothers",
         x="Mother's Smoking Status",
         y="Baby Weight (oz)")

# print mean and sd for smokers and non-smokers
cat("non-smoker\n")
cat("mean", mean(ns$bwt), '\n')
cat("sd", sd(ns$bwt), '\n')
cat("smoker\n")
cat("mean", mean(s$bwt), '\n')
cat("sd", sd(s$bwt))

# Q-Q Plot
qqnorm(s$bwt, pch = 1, frame=FALSE, main="Smoker")
qqline(s$bwt, col="steelblue", lwd=2)

# split data into non-smokers and smokers
ns<-bb[bb$smoke=="Non-smoker",]
s<-bb[bb$smoke=="Smoker",]

# filter smokers and non-smokers by 'box whisker' method
Q1 = summary(ns$bwt)['1st Qu.']
Q3 = summary(ns$bwt)['3rd Qu.']
IQR = Q3-Q1
min_cutoff = Q1 - (1.5*IQR)
max_cutoff = Q3 + (1.5*IQR)
ns<-ns[ns$bwt > min_cutoff, ]
ns<-ns[ns$bwt < max_cutoff, ]
```

```r
ns<-ns[ns$bwt < max_cutoff, ]

Q1 = summary(s$bwt)['1st Qu.']
Q3 = summary(s$bwt)['3rd Qu.']
IQR = Q3-Q1
min_cutoff = Q1 - (1.5*IQR)
max_cutoff = Q3 + (1.5*IQR)
s<-s[s$weight > min_cutoff, ]
s<-s[s$weight < max_cutoff, ]

# create histogram
p = ggplot(bb, aes(bwt, fill=smoke)) + geom_histogram(alpha=.5, aes(y=..density..), position='identity')
p + labs(title="Density of Baby Weights in Smoking vs Non-smoking Mothers",
         x="Baby Weight (oz)",
         y="Density")

# generate gestational periods table
means <- list()
sds <- list()
for (i in 32:46) {
    week_lower = i
    week_upper = i + 1
    day_lower = week_lower * 7
    day_upper = week_upper * 7
    t <- ns[ns$gestation < day_upper, ]
    mean <- sum(t$bwt) / nrow(t)
    means[i] <- mean
    sd <- sd(t$bwt)
    sds[i] <- sd
    cat("week", i, ":", mean, "sd:", sd, "\n")
}

```