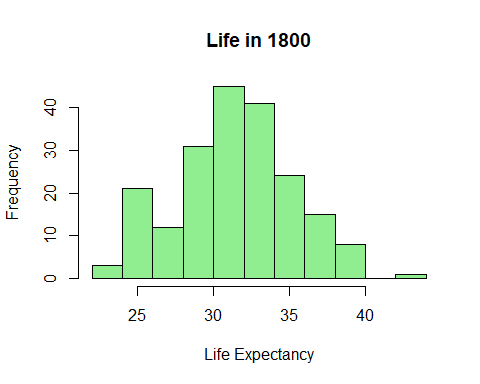
LifeExp

colnames(life)[1] <- 'country'

#Life expecancy in 1800  
summary(life$'1800')

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 23.39 29.00 31.80 31.49 33.90 42.85 798

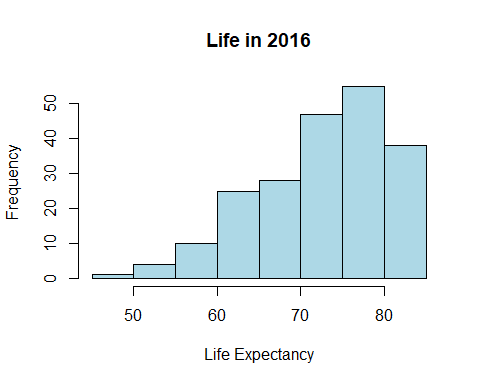
hist(life$'1800', xlab='Life Expectancy', main='Life in 1800', col='lightgreen')



#Life expectancy in 2016  
summary(life$'2016')

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 48.86 67.17 74.50 72.56 78.65 84.80 791

hist(life$'2016', xlab='Life Expectancy', main='Life in 2016', col='lightblue')



library(tidyr)  
lifel <- gather(life, 'year', 'life\_exp', -'country')

#checked and saw lots of missing values that needed to be cleaned  
sum(is.na(lifel))

## [1] 172926

#ommiting the blank rows  
lifel <- na.omit(lifel)

sum(is.na(lifel))

## [1] 0

ggplot(lifel, aes(x=year, y=life\_exp)) +geom\_boxplot(colour = "coral2")+  
 scale\_x\_discrete('Year', limits=c('1800', '2016'))

## Warning: Removed 43448 rows containing non-finite values (stat\_boxplot).



#We can see that life expecatncy starts to change for most of the world in the 1920's  
ggplot(lifel, aes(x=year, y=life\_exp)) +geom\_boxplot(alpha=1/100, color='chartreuse2') +  
 scale\_x\_discrete(breaks = seq('1800','2016', by=25))

