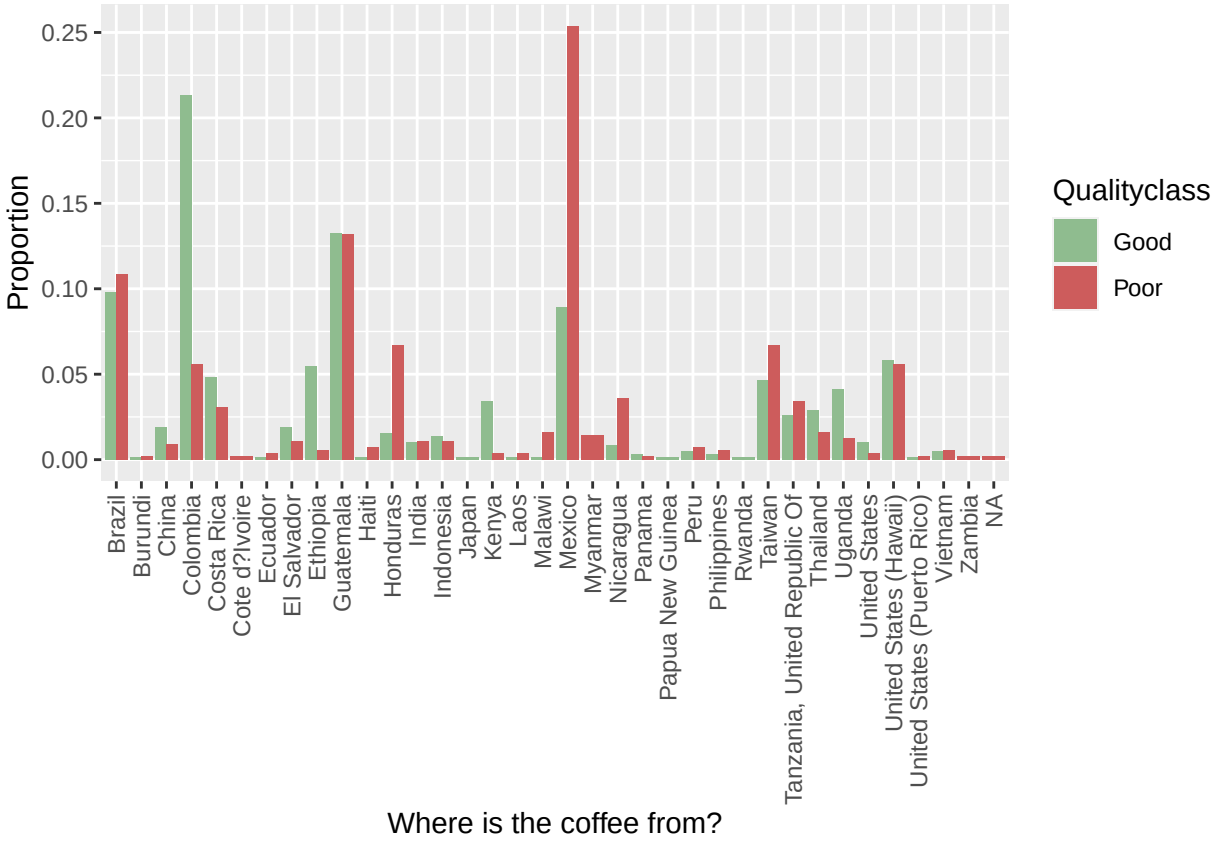
Group_14_Analysis

```
summary(coffee)
```

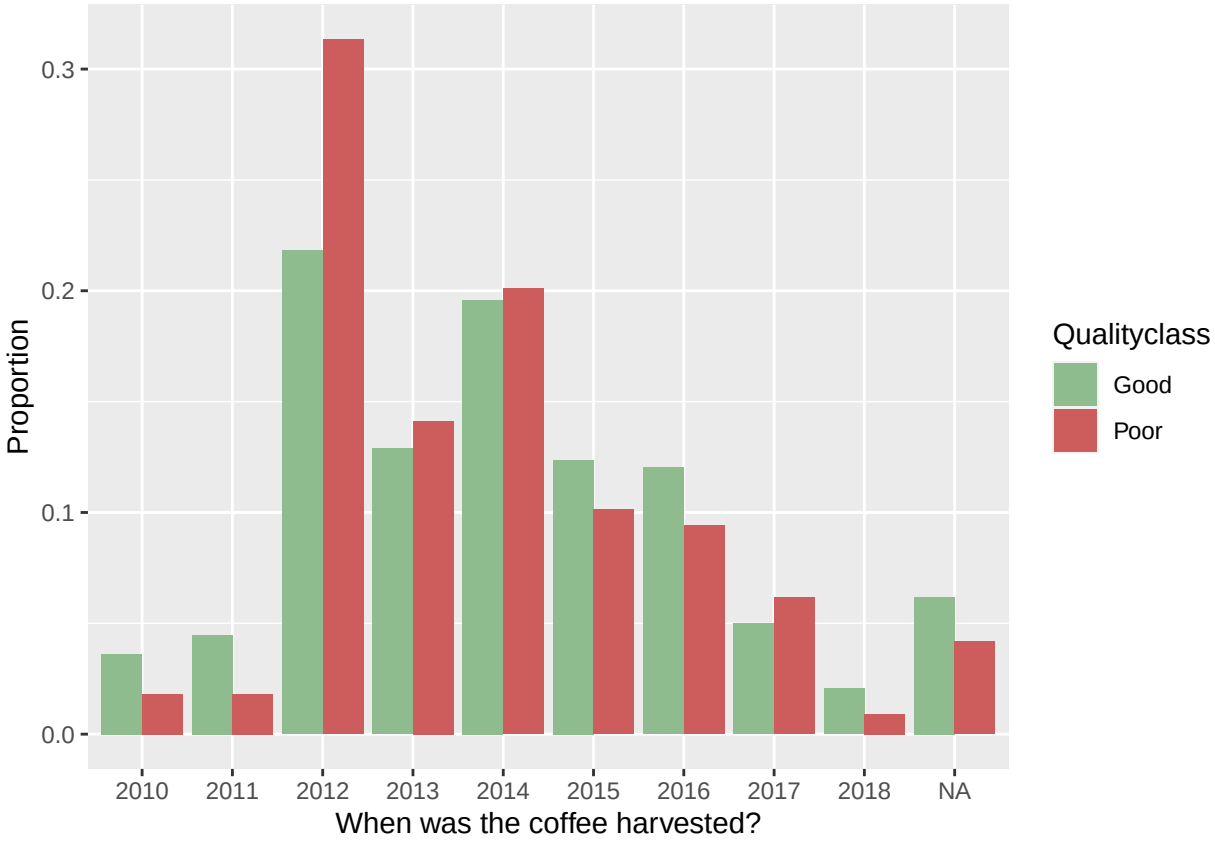
```
##  country_of_origin      aroma           flavor         acidity
##  Length:1134        Min.   :0.000   Min.   :0.000   Min.   :0.000
##  Class :character   1st Qu.:7.420   1st Qu.:7.330   1st Qu.:7.330
##  Mode  :character   Median :7.580   Median :7.580   Median :7.580
##                     Mean   :7.568   Mean   :7.523   Mean   :7.538
##                     3rd Qu.:7.750   3rd Qu.:7.750   3rd Qu.:7.750
##                     Max.   :8.750   Max.   :8.830   Max.   :8.750
##
##  category_two_defects altitude_mean_meters   harvested    Qualityclass
##  Min.   : 0.000       Min.   :     1       Min.   :2010   Length:1134
##  1st Qu.: 0.000       1st Qu.:  1100       1st Qu.:2012   Class :character
##  Median : 2.000       Median :  1311       Median :2014   Mode  :character
##  Mean   : 3.501       Mean   :  1653       Mean   :2014
##  3rd Qu.: 4.000       3rd Qu.:  1600       3rd Qu.:2015
##  Max.   :55.000       Max.   :190164       Max.   :2018
##                       NA's   :199          NA's   :59
```

```
ggplot(data = coffee, aes(x=country_of_origin, group = Qualityclass)) +
  geom_bar(aes(y = ..prop.., fill= Qualityclass), stat="count", position="dodge") +
   scale_fill_manual(values = c("darkseagreen", "indianred")) +
  labs(x= "Where is the coffee from?", y="Proportion") +
  ggpubr::rotate_x_text() # The first plot is a bar chart showing the proportion of coffee quality clas
```

```
## Warning: The dot-dot notation (`..prop..`) was deprecated in ggplot2 3.4.0.
## i Please use `after_stat(prop)` instead.
```



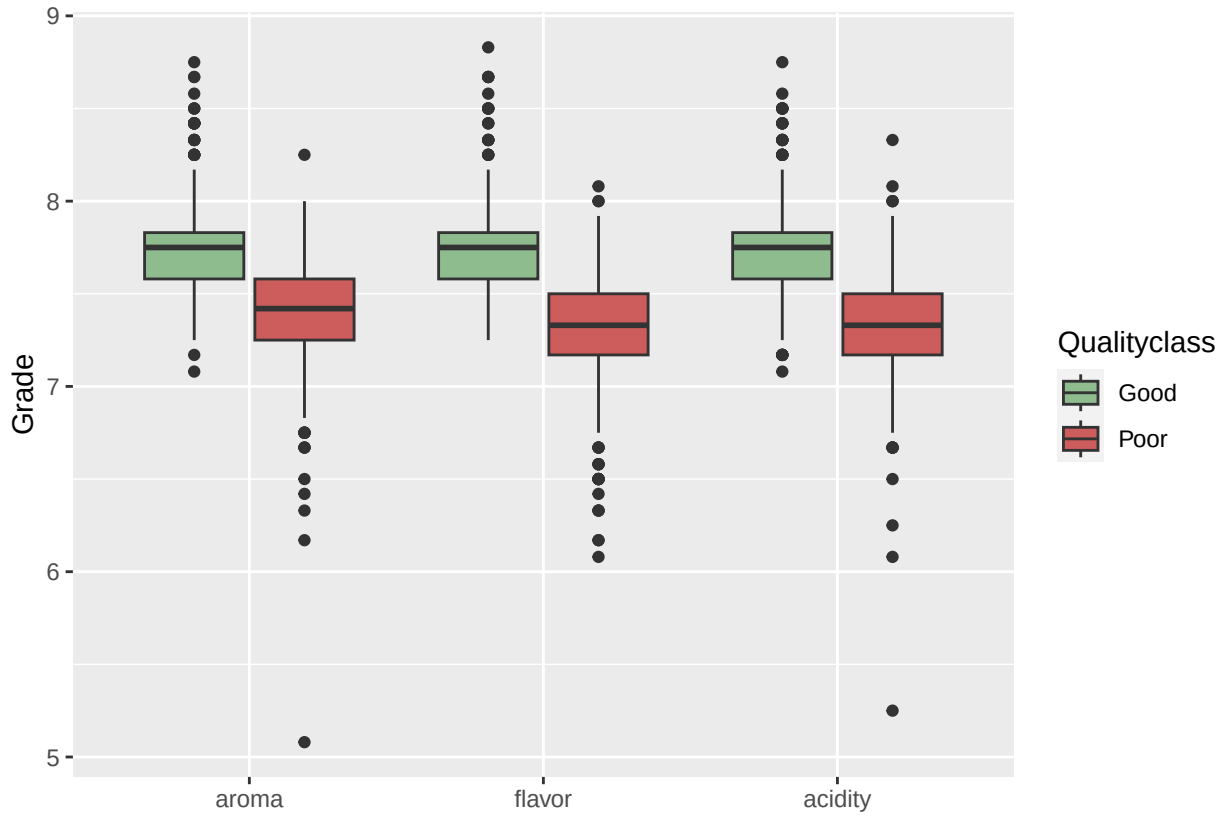

```
coffee$harvested <- as.factor(coffee$harvested)
ggplot(data = coffee, aes(x=harvested, group = Qualityclass)) +
  geom_bar(aes(y = ..prop.., fill= Qualityclass), stat="count", position="dodge") +
  scale_fill_manual(values = c("darkseagreen", "indianred")) +
  labs(x= "When was the coffee harvested?", y="Proportion") #The second plot is a bar chart showing the
```



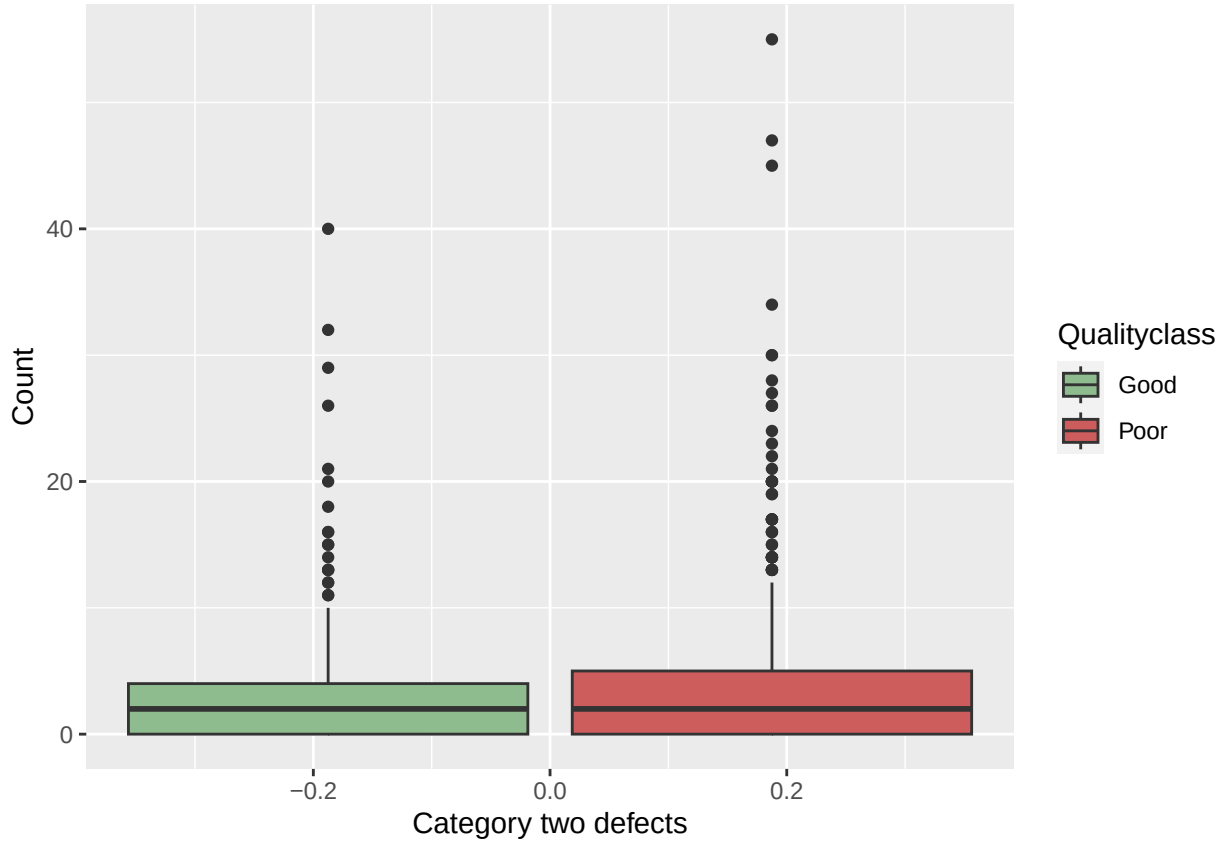

```
#Some data wrangling to get a nice boxplot with aroma, flavor and acidity
coffee_long <- melt(coffee, id="Qualityclass")
```

```
## Warning: attributes are not identical across measure variables; they will be
## dropped
```

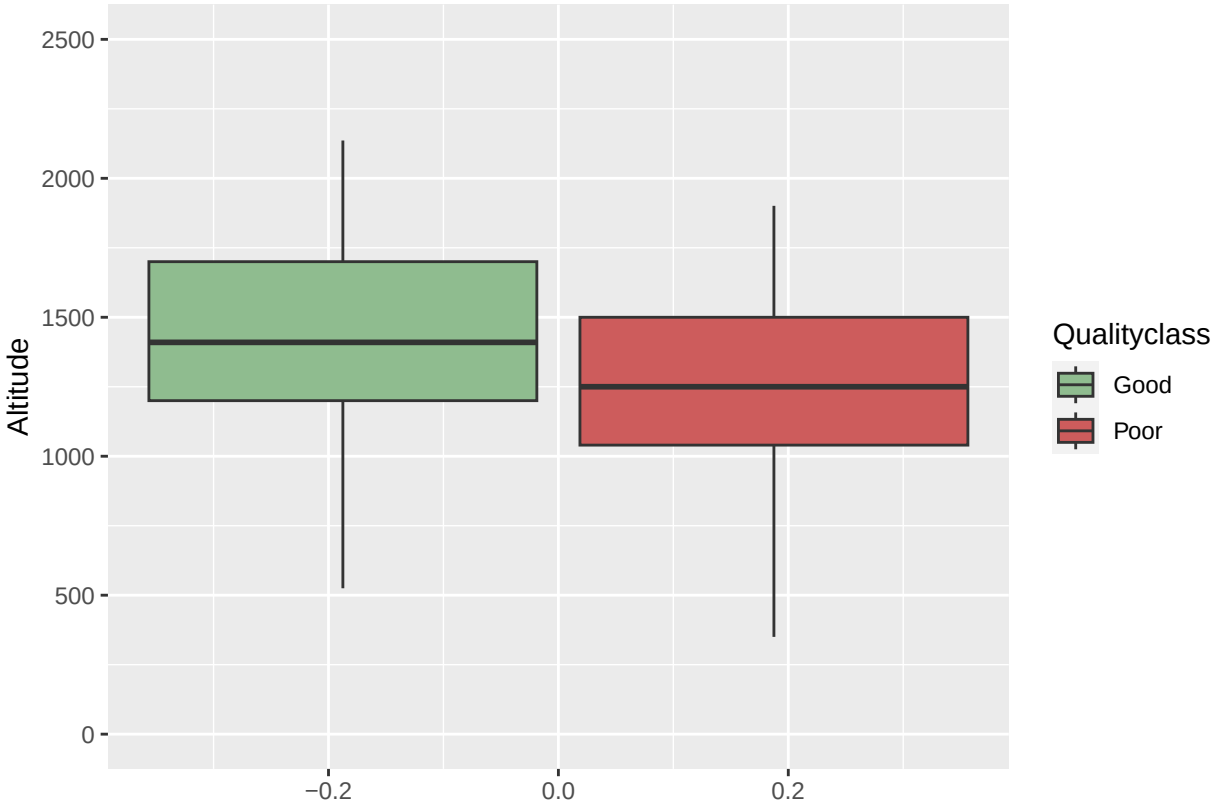
```
coffee_qualities <- coffee_long %>% filter(!(variable %in% c("country_of_origin", "category_two_defects
coffee_qualities$value <- as.numeric(coffee_qualities$value)
ggplot(coffee_qualities, aes(x = variable, y=value)) +
  geom_boxplot(aes(fill= Qualityclass)) +
  scale_fill_manual(values = c("darkseagreen", "indianred")) +
  labs(x="", y="Grade")#The third plot is a boxplot comparing the grades of coffee based on different q
```




```
ggplot(coffee, aes(y=category_two_defects)) +
  geom_boxplot(aes(fill=Qualityclass)) +
  scale_fill_manual(values = c("darkseagreen", "indianred")) +
  labs(x="Category two defects", y="Count")#The fourth plot is a boxplot comparing the number of catego
```




```
coffee_alt <- coffee %>% drop_na(altitude_mean_meters)
ggplot(coffee_alt, aes(y=altitude_mean_meters)) +
  geom_boxplot(aes(fill=Qualityclass), outlier.shape = NA) +
  coord_cartesian(ylim =  c(0, 2500)) +
  scale_fill_manual(values = c("darkseagreen", "indianred")) +
  labs(x="", y="Altitude")#The fifth plot is a boxplot comparing the altitude of coffee farms by qualit
```


2500
2000
1500
1000
500
0
Altitude
−0.2
0.0
0.2
Qualityclass
Good
Poor