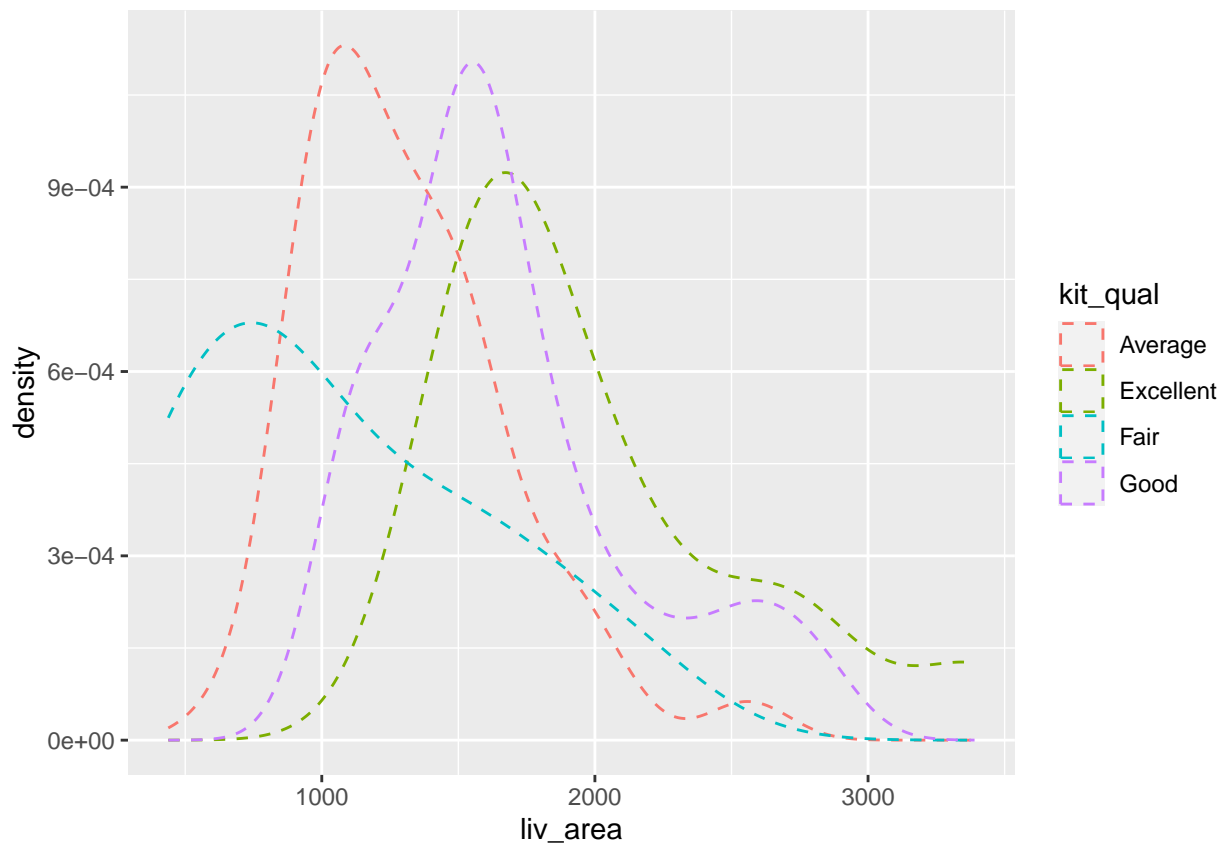# HW5

Andrew Shao

2024-10-09

5.10.2 Q1

```r
library(r02pro)
```

```
## Warning: package 'r02pro' was built under R version 4.3.3
```
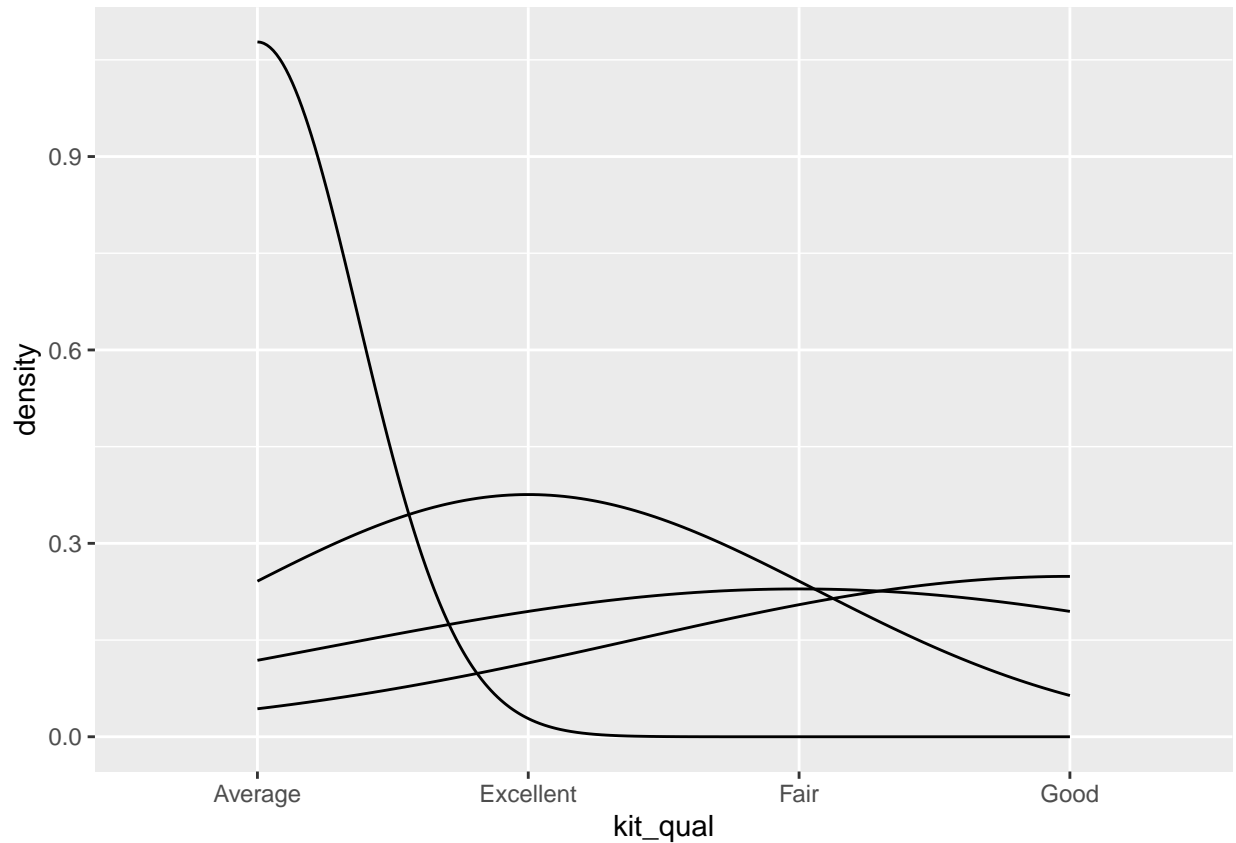
```r
library(ggplot2)

ggplot(data = sahp) +
  geom_density(aes(x = liv_area,
                   color = kit_qual),
               linetype = 'dashed')
```

Living area is skewed right for all the kitchen qualities. As kitchen quality improves the center of the living area distribution shifts to the right.
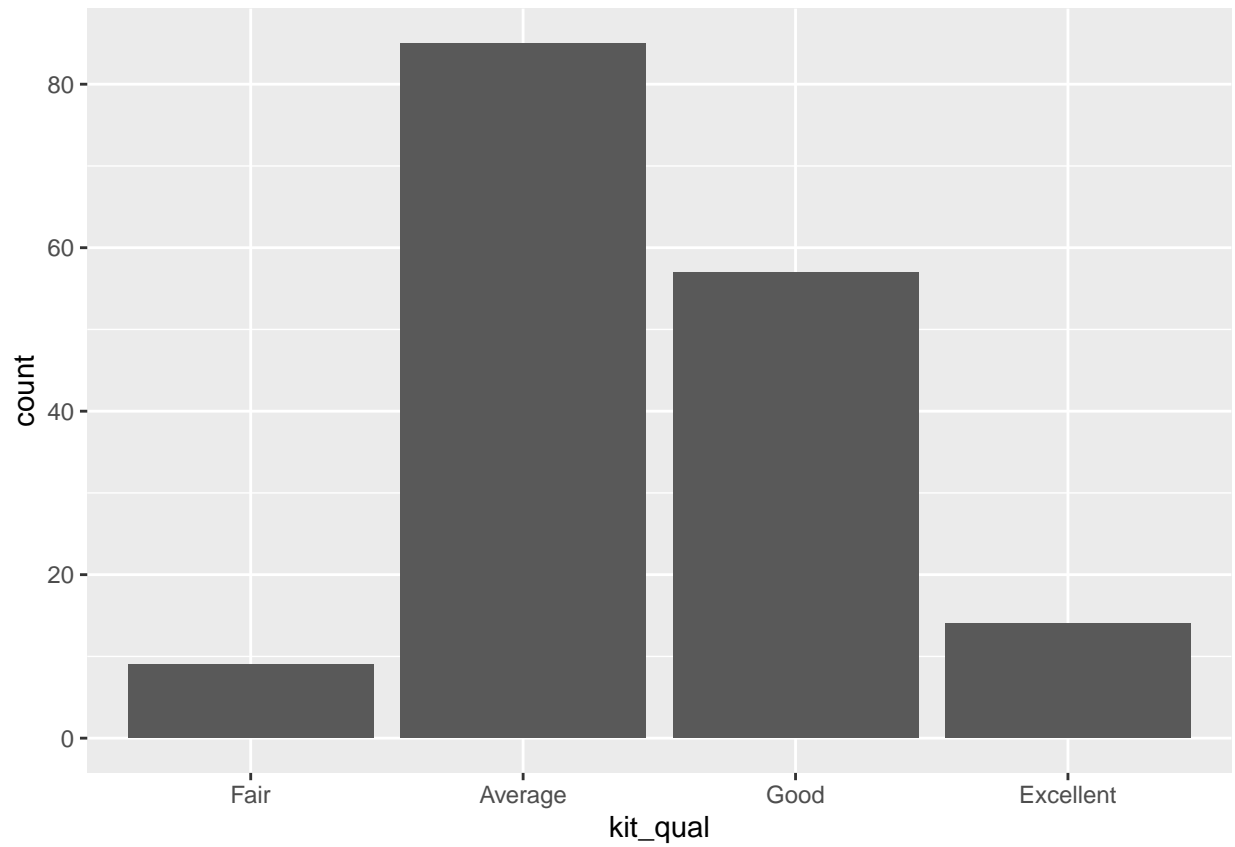
5.10.2 Q2

```r
ggplot(data = sahp) +
  geom_density(aes(x = kit_qual))
```


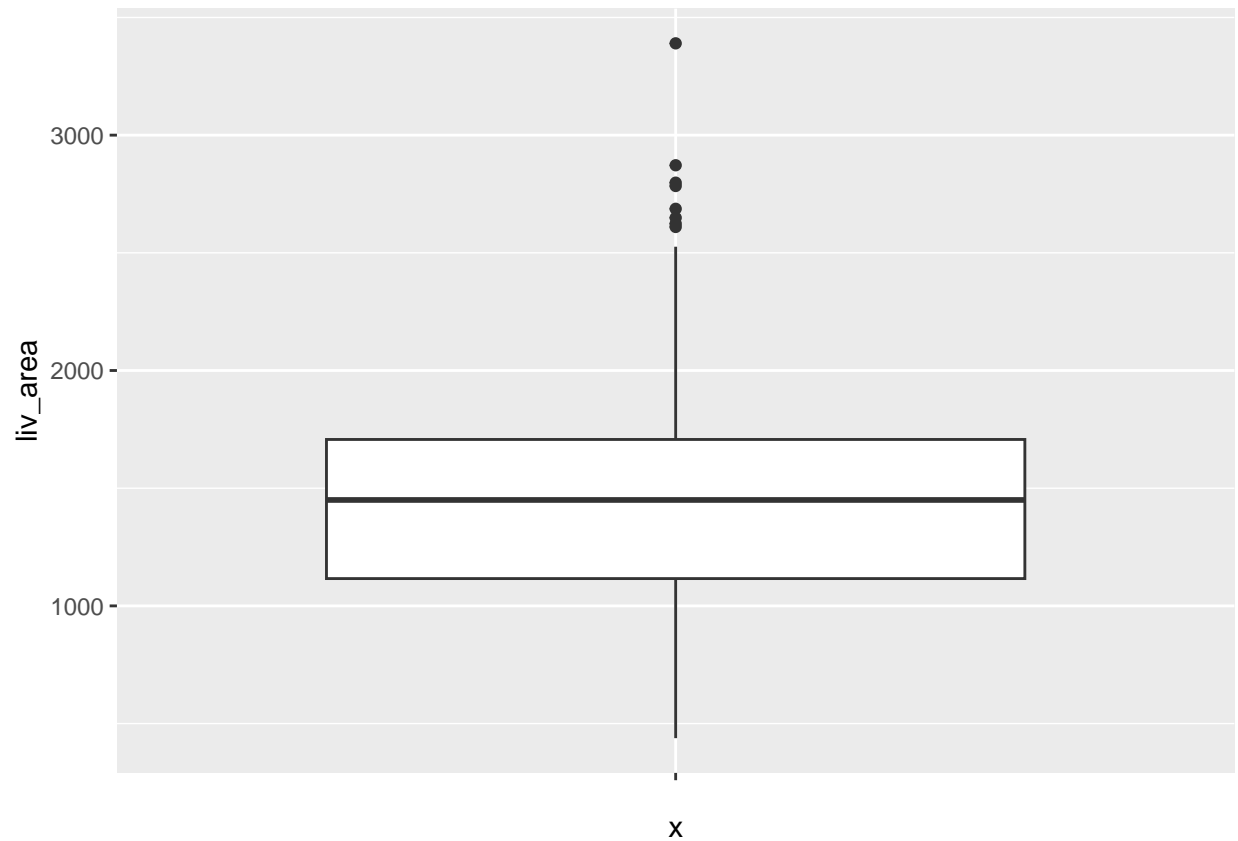
This plot is not informative since kit_qual is a categorical variable it doesn't make sense to use a density plot.

```r
library(forcats)
ggplot(data = sahp) +
  geom_bar(aes(x = fct_relevel(kit_qual,
                              c('Fair',
                                'Average',
                                'Good',
                                'Excellent')
                              ))) +
  xlab('kit_qual')
```

5.11.6 Q1

```
ggplot(data = sahp) +
  geom_boxplot(aes(x = '', y = liv_area))
```

Line in the middle:

```r
median(sahp$liv_area)
```

```
## [1] 1450
```

Lower hinge:

```r
quantile(sahp$liv_area, 0.25)
```

```
##   25%
## 1116
```

Upper hinge:

```r
quantile(sahp$liv_area, 0.75)
```

```
##   75%
## 1707
```

Lower whisker:

```
quantile(sahp$liv_area, 0.25) - 1.5 * IQR(sahp$liv_area)
```
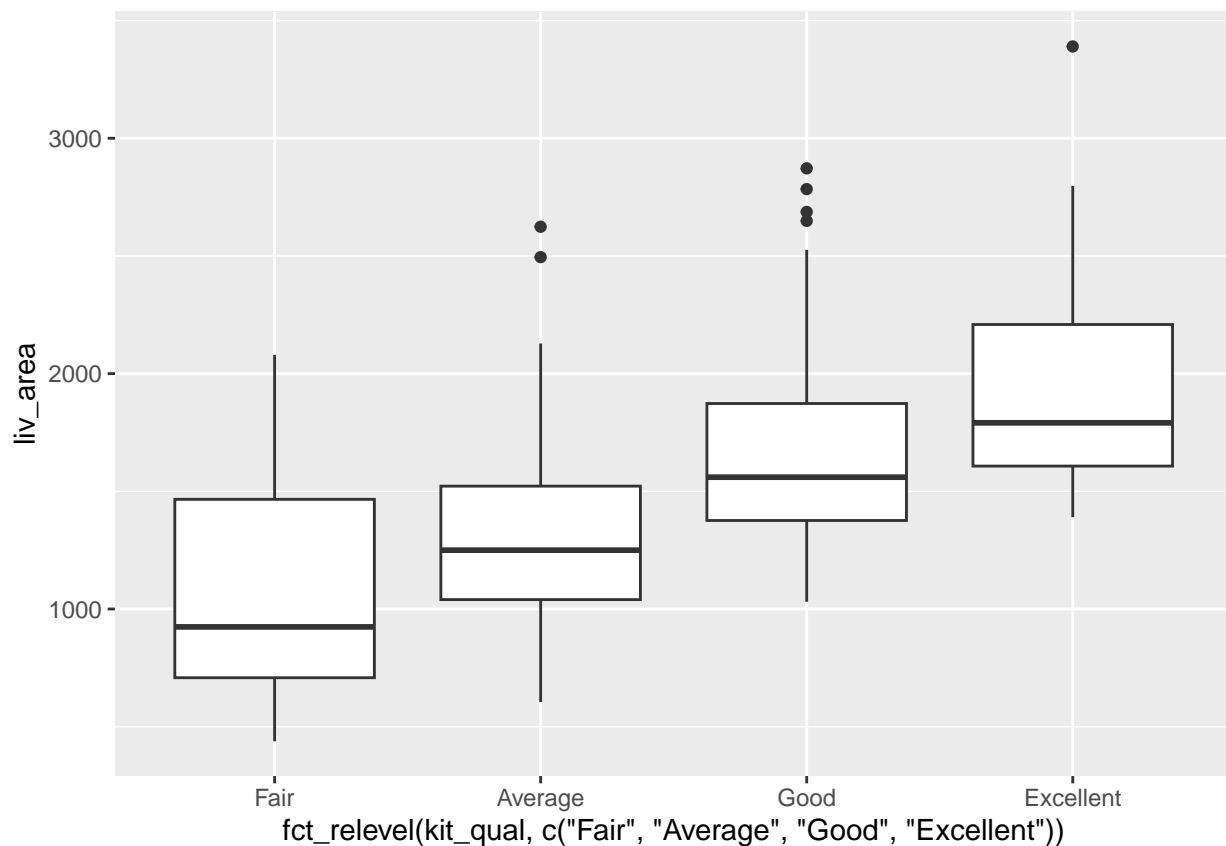
```
##    25%
## 229.5
```

Upper whisker:

```
quantile(sahp$liv_area, 0.75) + 1.5 * IQR(sahp$liv_area)
```

```
##     75%
## 2593.5
```

5.11.6 Q2

```
ggplot(data = sahp) +
  geom_boxplot(aes(x = fct_relevel(kit_qual,
                          c('Fair',
                            'Average',
                            'Good',
                            'Excellent')
                          ),
                   y = liv_area))
```
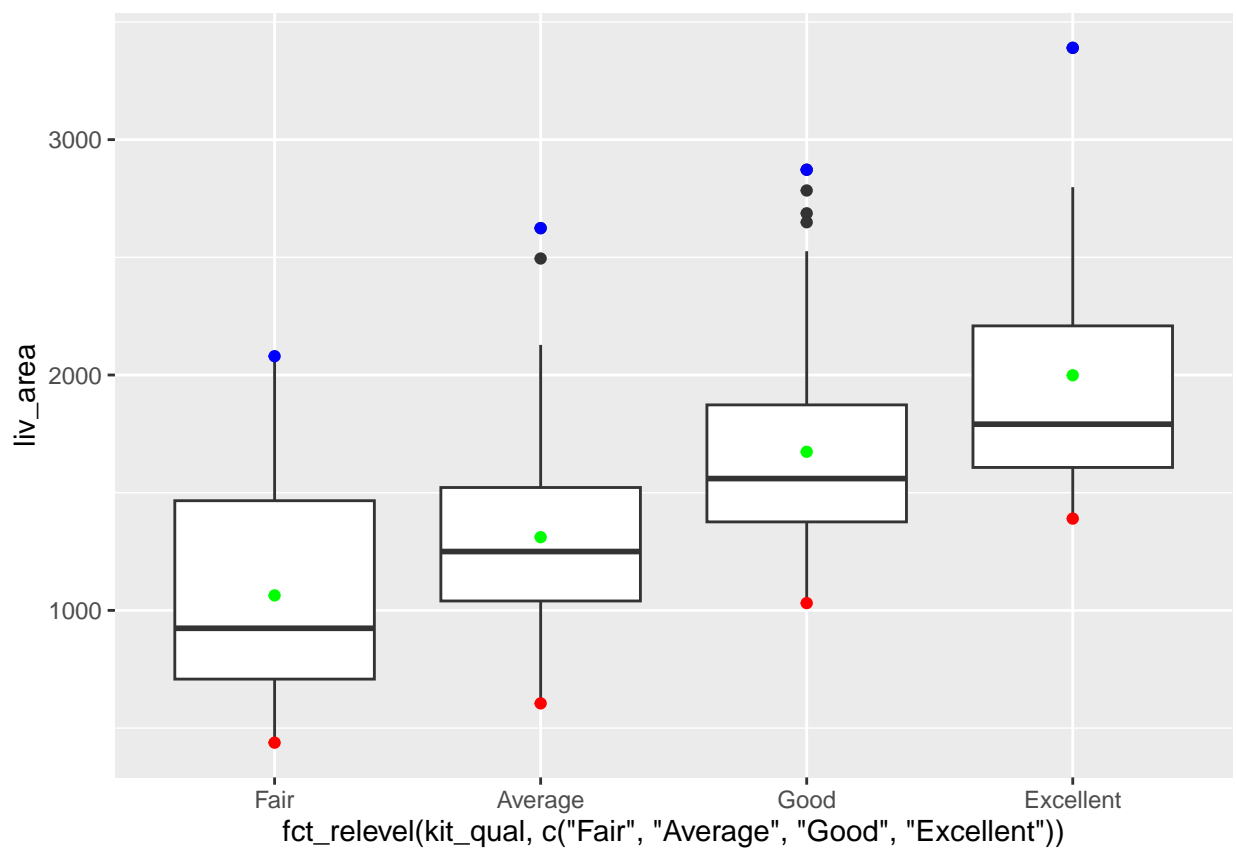

As kitchen quality improves the distribution of living area shifts in the positive direction.

5.11.6 Q3

```
ggplot(data = sahp,
       aes(x = fct_relevel(kit_qual,
                           c('Fair',
                             'Average',
                             'Good',
                             'Excellent')
                          ),
           y = liv_area)) +
  geom_boxplot() +
  geom_point(stat = 'summary',
             fun = 'min',
             color = 'red') +
  geom_point(stat = 'summary',
             fun = 'max',
             color = 'blue') +
  geom_point(stat = 'summary',
             fun = 'mean',
             color = 'green')
```
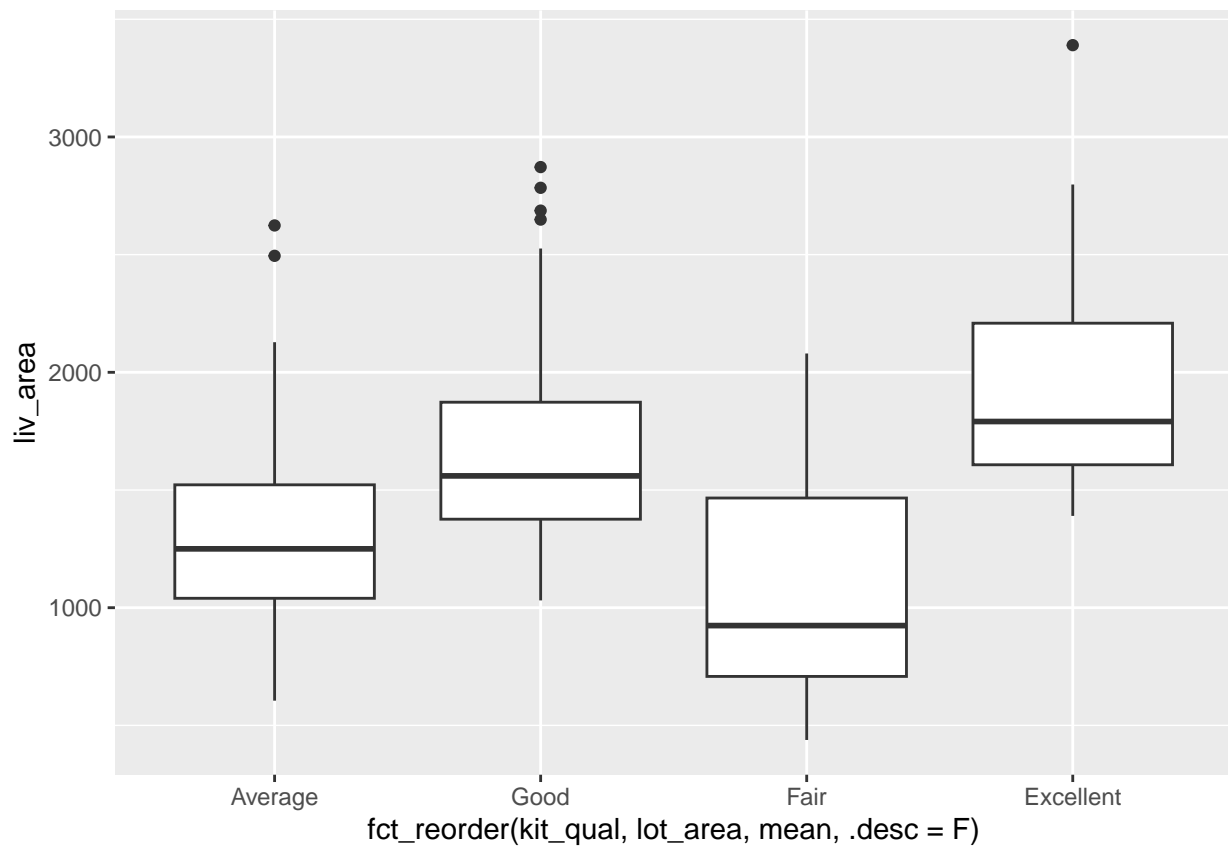
5.11.6 Q4

```r
ggplot(data = sahp) +
  geom_boxplot(aes(x = fct_reorder(kit_qual,
                                   lot_area,
                                   mean,
                                   .desc = F
                                   ),
                   y = liv_area))
```
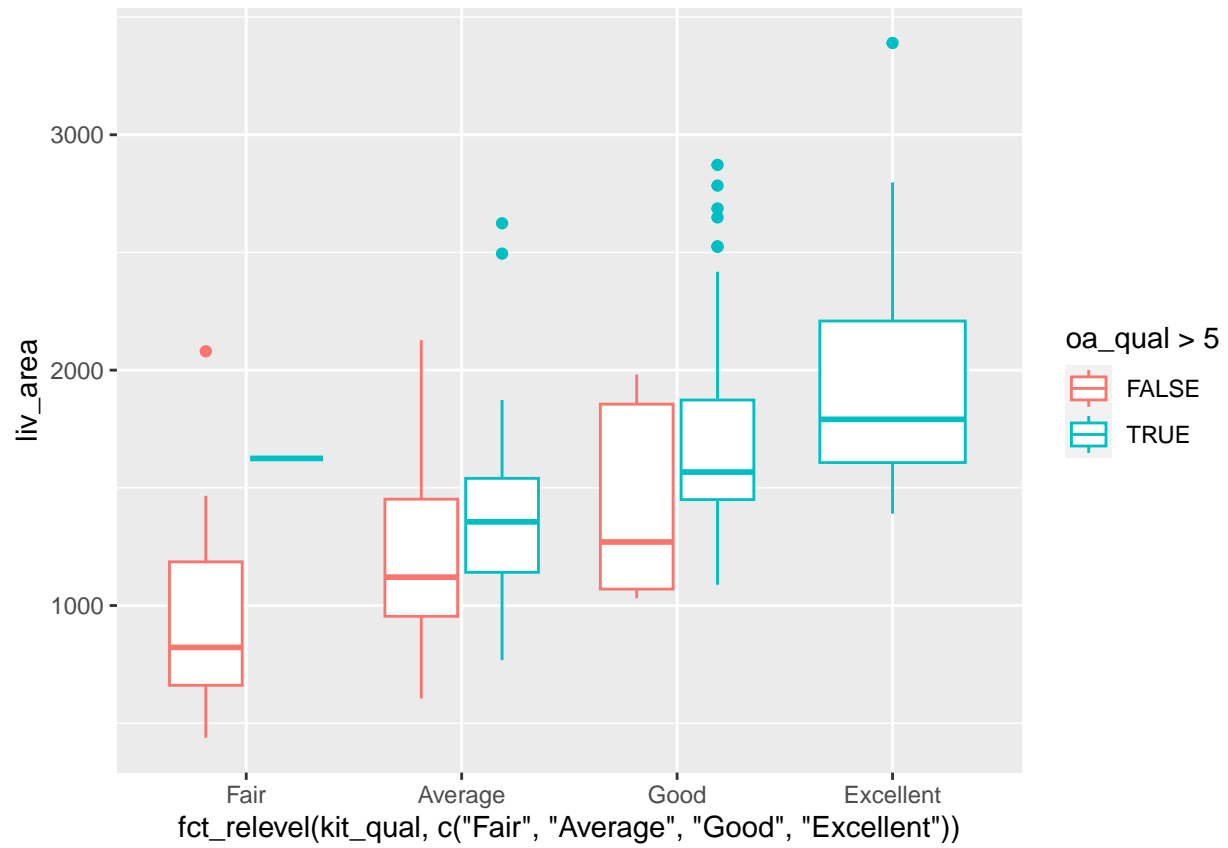

5.11.6 Q5

```r
ggplot(data = remove_missing(sahp, vars = 'oa_qual')) +
  geom_boxplot(aes(x = fct_relevel(kit_qual,
                                   c('Fair',
                                     'Average',
                                     'Good',
                                     'Excellent')
                                   ),
                   y = liv_area,
                   colour = oa_qual > 5))
```
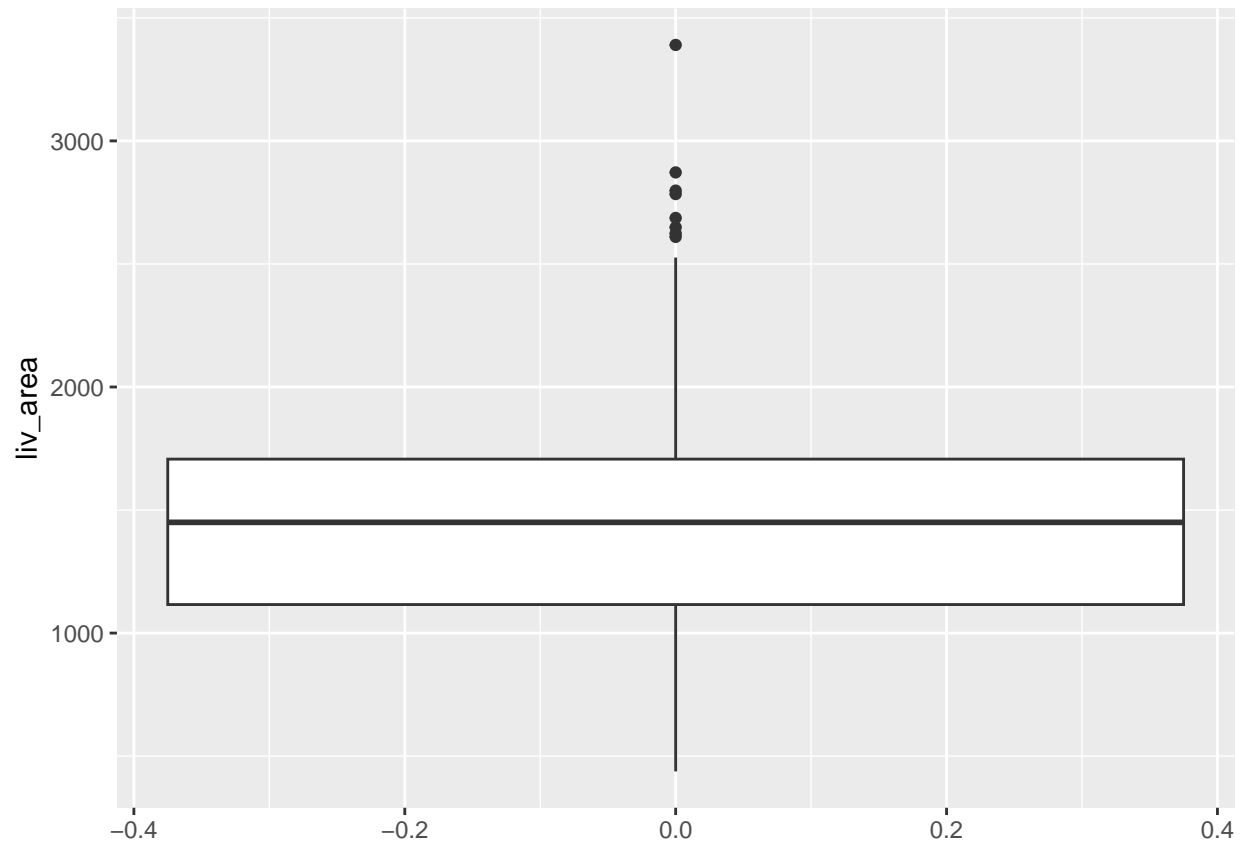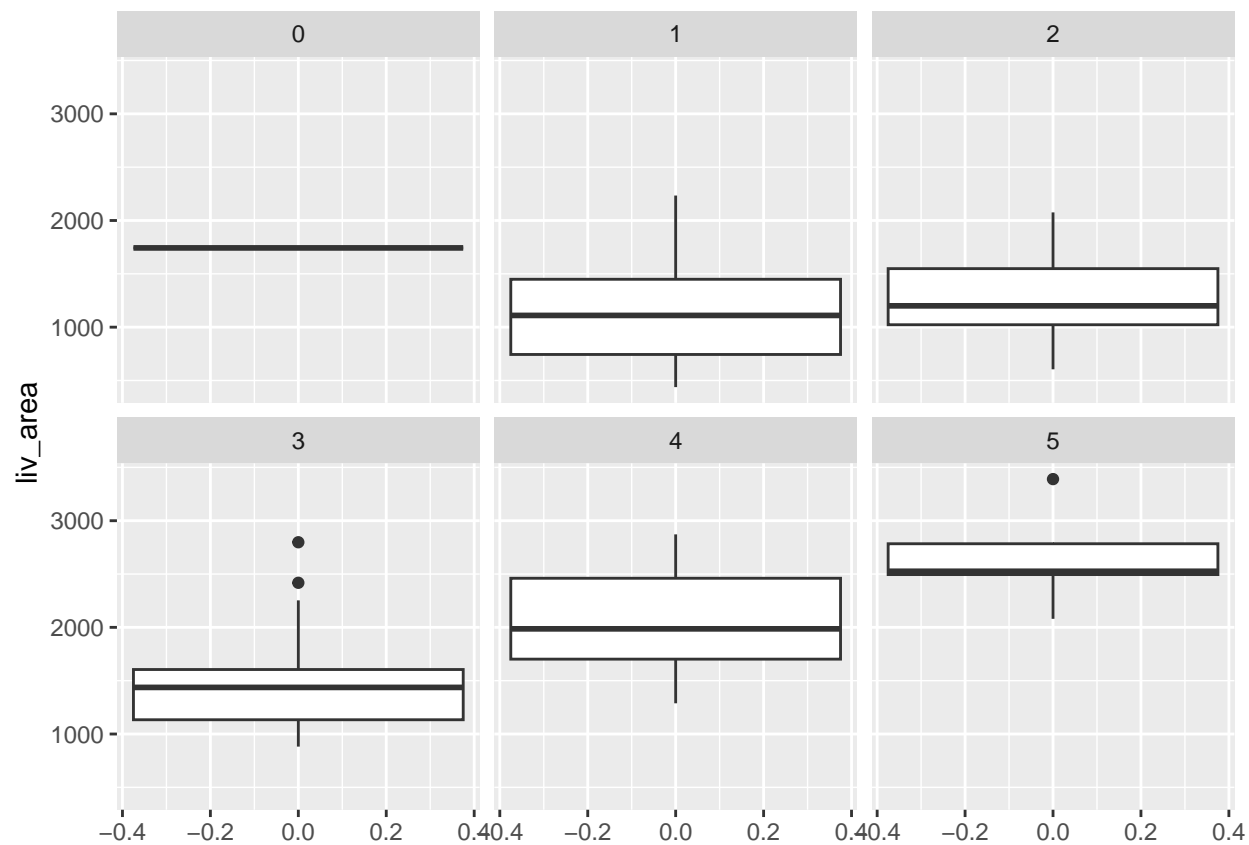
```
## Warning: Removed 1 rows containing missing values.
```


5.12.3 Q1

```r
my_boxplot <- ggplot(data = sahp) +
  geom_boxplot(aes(y = liv_area))

my_boxplot
```
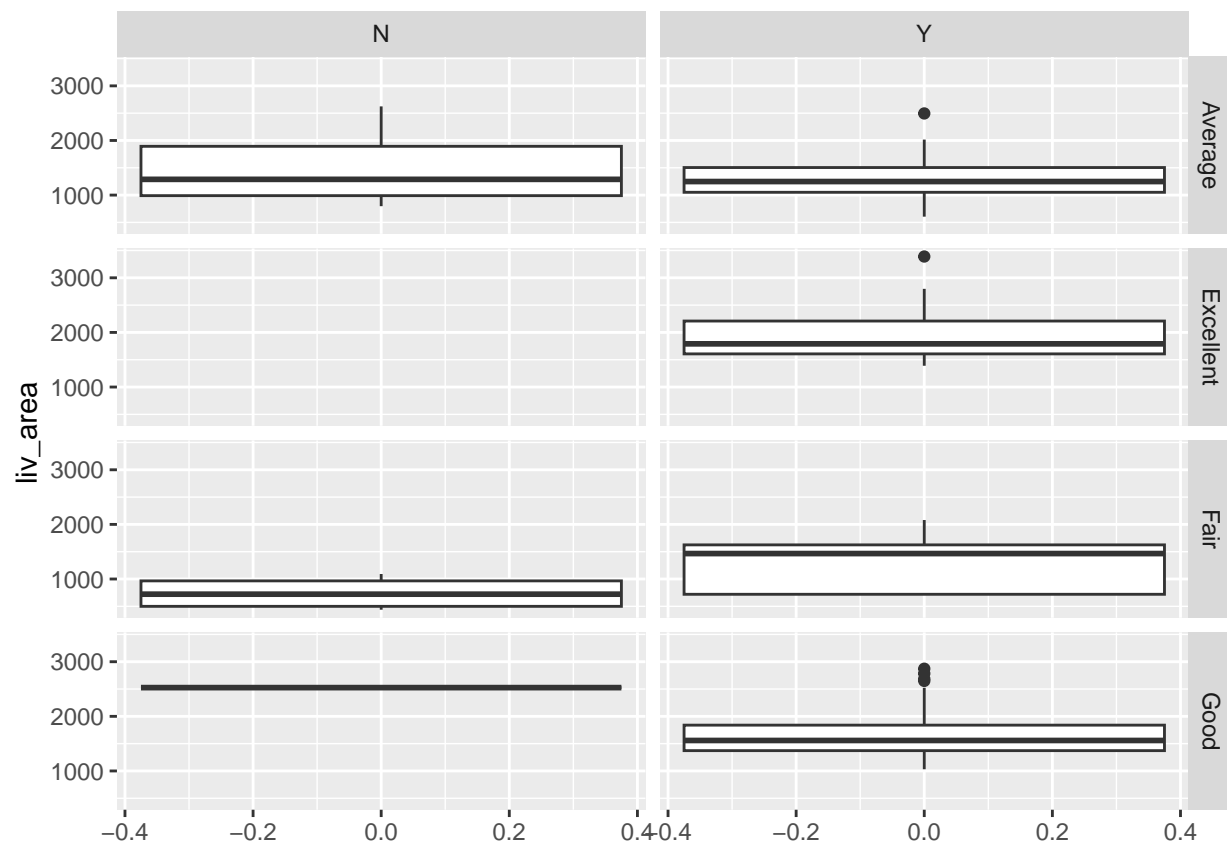
5.12.3 Q2

```
my_boxplot + facet_wrap('bedroom')
```

As bedroom count increases so does the distribution of living area... which makes sense.

5.12.3 Q3

```
my_boxplot + facet_grid(rows = vars(kit_qual),
                        cols = vars(central_air))
```

There is an empty plot for `kit_qual` value of `"Excellent"` and `central_air` value of `"N"`, this is because there is no observations within the dataset with that combination of values.