Stat 21 Homework 4

Problem 10 Solution

```
voter_turnout <- tibble(law_type = c(rep("No law",18),
                                     rep("Non-strict non-photo", 14),
                                     rep("Strict non-photo",2),
                                     rep("Non-strict photo",8),
                                     rep("Strict photo",8)),
                        turnout_rate = c(52.8,56.5,61.6,68.4,70.7,66.4,67.2,74.1,62.8,57.4,64.4,
                                         54.5,56.8,64.5,66.2,63.6,63.7,59.5,50.2,64.7,56.8,
                                         52.3,60.8,71.4,61.8,62.2,58.6,64.2,63.7,70,61,54.9,56.7,
                                         62.9,51.4,58.6,59.1,64.7,60,59.2,42.3,64.5,
                                         69.5,66.1,51.1,55.2,57.7,56.4,59.1,58.8))
head(voter_turnout)
```

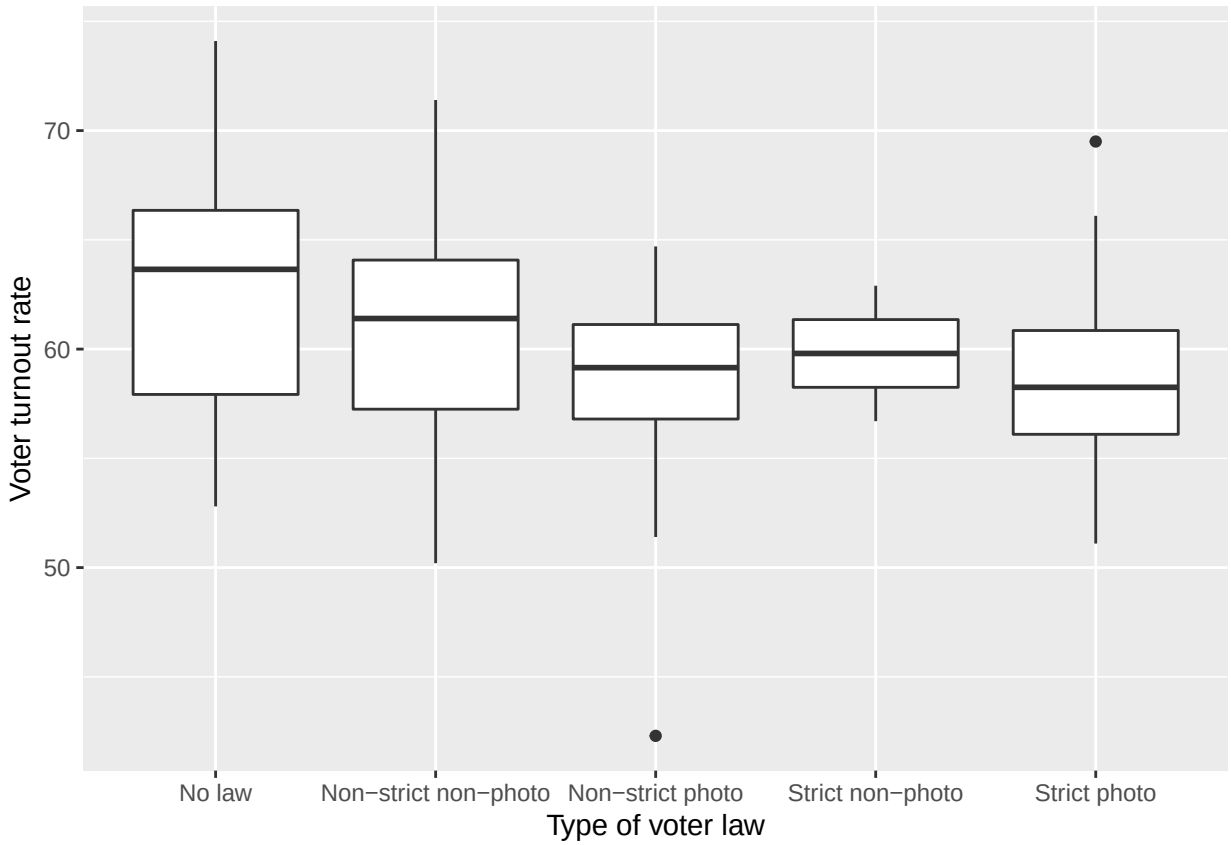
```
## # A tibble: 6 x 2
##   law_type turnout_rate
##   <chr>           <dbl>
## 1 No law           52.8
## 2 No law           56.5
## 3 No law           61.6
## 4 No law           68.4
## 5 No law           70.7
## 6 No law           66.4
```

Problem 10

Plot the data with box plots for voter turnout. Does the equal variance assumption seem reasonable? What does this mean about the estimate in Problem 9?

Solution Problem 10:

```
ggplot(voter_turnout, aes(x=factor(law_type), y=turnout_rate)) +
  geom_boxplot() +
  xlab("Type of voter law") +
  ylab("Voter turnout rate")
```



Based on on the box plots, the equal variance assumption seems reasonable for every group except those states with strict non-photo ID laws, which seems to have a much smaller spread in turnout rates than the other groups of states. This seems to indicate that the estimate for σ^2 from Problem 9 is not going to be reliable. However, there are only two data points within this group of states with strict non-photo ID laws. Since this data represents different states in the US, there is no way we can collect more data for this group which would normally be the solution here. A more appropriate solution for this example is to instead consider combining this group with another level of the categorical variable that it may be similar to. Perhaps it makes most sense to combine states with strict non-photo laws and those with strict photo laws into a single level of states with "strict laws - either photo or non-photo". This would give us a more reliable estimate for σ^2 as all groups would have at least a few observations and the spread of each of the box plots would not be too different.

For this problem I am looking for two things: a plot and exposition that interprets the homogeneity of variances assumption within the context of the box plots and the data itself.

Plot part - Full credit for clearly labeled box plots. Partial credit for correct plots but not clearly/neatly labeled.

Exposition part

- Full credit if noted that the equal variances assumption may be appropriate but we cannot tell because there are only two data points in the "strict non-photo" group and therefore the estimate from Problem 9 may or may not be reliable, we have no way to tell without further adjustment.
- Partial credit if noted that equal variances assumption does not seem appropriate because of the "strict non-photo group" and therefore the estimate for σ^2 from Problem 9 is not reliable.