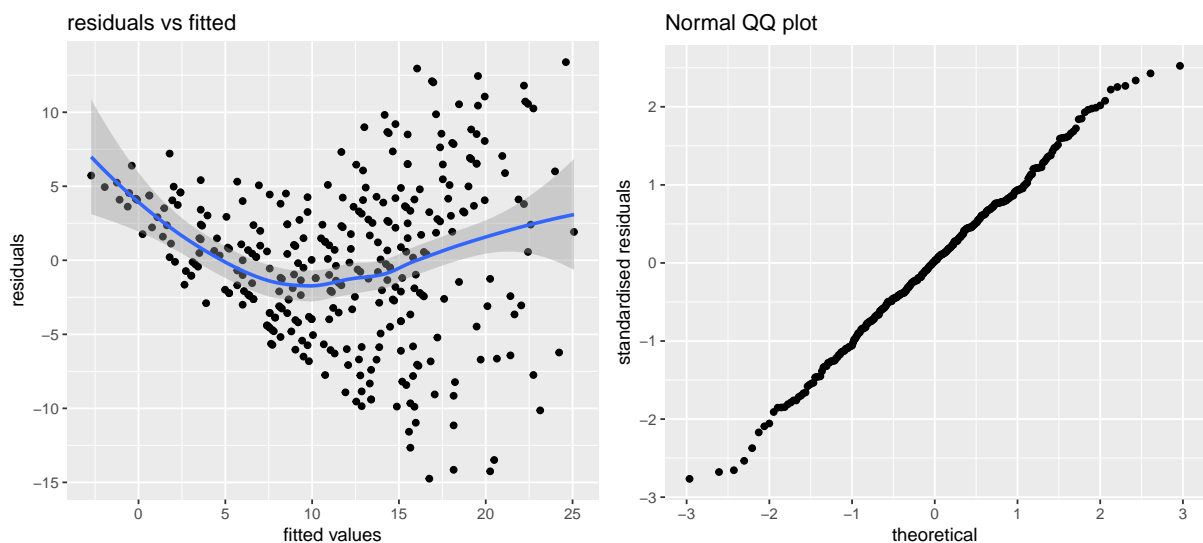


Problem H.2

Consider again the LA ozone data from problem H.1.

- Fit two simple linear regression models to the data using `lm()`. Determine the variable, which has the highest correlation with the response ozone. Use this variable as predictor in the first model. The second model uses the variable `season`, defined in H.1 f) as predictor variable. Interpret the estimate b_1 for both models.
- Perform a graphical residual analysis by plotting the two `lm`-objects from part a). Discuss in each case the first two plots. Which assumptions of the simple linear regression can you check using those two plots? Are these assumptions satisfied for the two models?
- Which model do you prefer? Give reasons for your decision.
- Now write a function `qplot_lm()`, which has as input a `lm`-object and the argument `cat_pred`, which indicates if the predictor variable of the model is categorical. Think about where to use this information. The output of the function is a plot of the form

```
qplot_lm(lm_model, cat_pred = FALSE)
```



Hence, it should produce one graphic showing the two plots side by side. Finally apply `qplot_lm()` to both models from part a).

Hint: Visit <https://www.statmethods.net/management/userfunctions.html> for information on how to write your own function. You might also want to use the functions `rstandard()` and `gridExtra::grid.arrange()` (from the `gridExtra` package).

Remark: The notation `::` allows you to use `grid.arrange()` without loading in advance `gridExtra` with the `library()` command.