Case: Log transformations of (simulated) data in R

In this exercise you shall simulate data and fit linear regressions to study the effects of log transformations in the four cases:

$$
\begin{aligned}
y_i =& \alpha + \beta x_i + \varepsilon_i \\
\log y_i =& \alpha + \beta x_i + \varepsilon_i \\
\log y_i =& \alpha + \beta \log x_i + \varepsilon_i \\
y_i =& \alpha + \beta \log x_i + \varepsilon_i
\end{aligned}
\qquad i = 1, \ldots, n
$$

where the y_i are all independent and $\varepsilon_i \sim N(0, \sigma^2)$. See also the section about transformations and Table 1-2 in the lecture notes.

Exercise

We already know about the first model (ordinary linear regression).

For each of the three other models:

- Simulate x (or $\log x$) and y (or $\log y$) from the model. Choose values of α, β, σ^2 and n. α and β must be chosen such that the "need" for transformation can clearly be seen in the plot (see below).
- Plot x and y in one plot. Plot (according to the model) the transformed data in another plot.
- Fit a linear regression on the transformed data according to the model. Add the fit to your plots above. Remember to back-transform!
- Perform an ordinary linear regression on x and y (we know this is incorrect). Plot a residual plot and a normal QQ-plot. Comment on these.

Can we always "see" if we should log-transform data. In the case that there is little visual difference btw. log-transform or not, would you then transform your data? Why/why not?

Hint:

```
x <- seq(1.5, 2.5, length = 100) ## make x's
logy <- 3*x + rnorm(100, sd = 0.15) ## make log y_i s
y <- ??(logy) ## transform
plot(x,y)
[...] # Fill in
fit_model <- exp(coef_alpha , coef_beta*x)
lines(x, fit_model)
```