5.)

(a) As $\lambda \to \infty$, $\hat{g}_2$ will have the smaller training RSS, as it allows for a potentially more complex fit with more parms, and thus will make training RSS smaller. ✓

(b.) This answer depends on the underlying structure/properties of the true data generating function of $\{y_i\}$. For example, if $\{y_i\}$ generated from a 3rd degree polynomial in x_i, ✓ I would expect $\hat{g}_2$ to have smaller test set RSS. If $\{y_i\}$ generated from 2nd degree polynomial, $\hat{g}_1$ test set RSS likely smaller.

(c) when $\lambda = 0 \Rightarrow$ no penalty.

Thus again, $\hat{g}_2$ will have train set RSS at least as small, and test set RSS will be smaller for ($\hat{g}_1$, $\hat{g}_2$) if DGP closer to (2nd, 3rd) degree polynomial.

Thus here, $\hat{g}_1 = \hat{g}_2$, as both will try to minimize SSE (and likely both perfectly interpolate $\{y_i\}$, thus their train and test RSS will be equal!