

## Exercise 7

1. The map of tann (Oregon Annual Temperature) shows higher annual temperatures along the coast, with lower temperatures in Central and Southern Oregon. The residuals plot shows lower (more negative) residuals in these same areas, with residuals closer to 0 in Western, Northern, and Eastern Oregon.
2. The F-statistic between tann and elevation is 257.9, the p-value is  $2.2e-16$ , the Multiple R-squared value is 0.7413, and the residual standard error is 0.9852. Therefore, there is a significant relationship. The intercept and slope are significantly different from 0 (shown by plotting  $\text{tann} \sim \text{elevation}$ ).
3. There is a discernable pattern in the Residuals vs. Fitted plot, and there is some drift away from the 1:1 line of the Q-Q plot for residuals greater than 1. Thus we should modify this model.
4. There is a positive relationship between our linear model residuals and longitude, and a negative relationship with latitude.
5. For our multiple regression model of elevation and position as predictors, we again have a large F-statistic and a small p-value. Based on t-values, all the predictors are significant. There are still discernable patterns in the residuals plot.
6. There is no discernable pattern in the residuals vs fitted plot, and the normal Q-Q plot is much closer to a straight line, so fitting a local curve improved the fit relative to fitting a straight line.
7. Our second loess regression model fits the tann data better than the multiple regression models. The  $R^2$  value is much higher, and the residual standard error much lower. There is less of a pattern in the residual scatter diagram, although it's worth noting that the Q-Q Normal plot strays from a straight line below 0.
8. Our first loess regression model has a lower  $R^2$  and higher residual standard error than the expanded loess regression model, but its Normal Q-Q plot fits the straight line better. In any case, the loess regression is the optimal model for the tann data.

