1 - Summary

This week, we explored flexible methods for modeling nonlinear relationships using splines, smoothing techniques, and generalized additive models (GAMs). These approaches help preserve the interpretability of linear models while improving fit and reducing overfitting in complex datasets.

2 - Concepts

- **Step functions**: Divide the input space into intervals and fit constant values in each, but lack smoothness and flexibility.
- **Splines**: Piecewise polynomials joined smoothly at knots, with different types:
    - **Linear splines**: Continuous, piecewise linear functions.
    - **Cubic splines**: Ensure continuity of first and second derivatives across knots.
    - **Natural splines**: Add constraints at the boundary to prevent erratic behavior.
- **Knot selection**: Knots can be placed at quantiles of the data or spaced evenly; the number and location of knots affect model flexibility.
- **Generalized additive models (GAMs)**: Combine multiple smooth functions additively, allowing nonlinear relationships for each predictor while keeping the model interpretable.
- **Model fitting in R**: Use `bs()` for B-splines, `ns()` for natural splines, `smooth.spline()` for smoothing splines, and `gam()` from the `mgcv` package for GAMs.

3 - Uncertainties

One thing I'm still figuring out is how to choose the best number and placement of knots when using splines. The lecture mentioned options like using quantiles, but it's unclear when one method is better than another. I also found the concept of smoothing splines a bit abstract, especially how the roughness penalty balances bias and variance. Finally, with GAMs, I'm still a little unclear on how to interpret the fitted functions when they are nonlinear and how to determine which variables should be modeled linearly versus nonlinearly in practice.