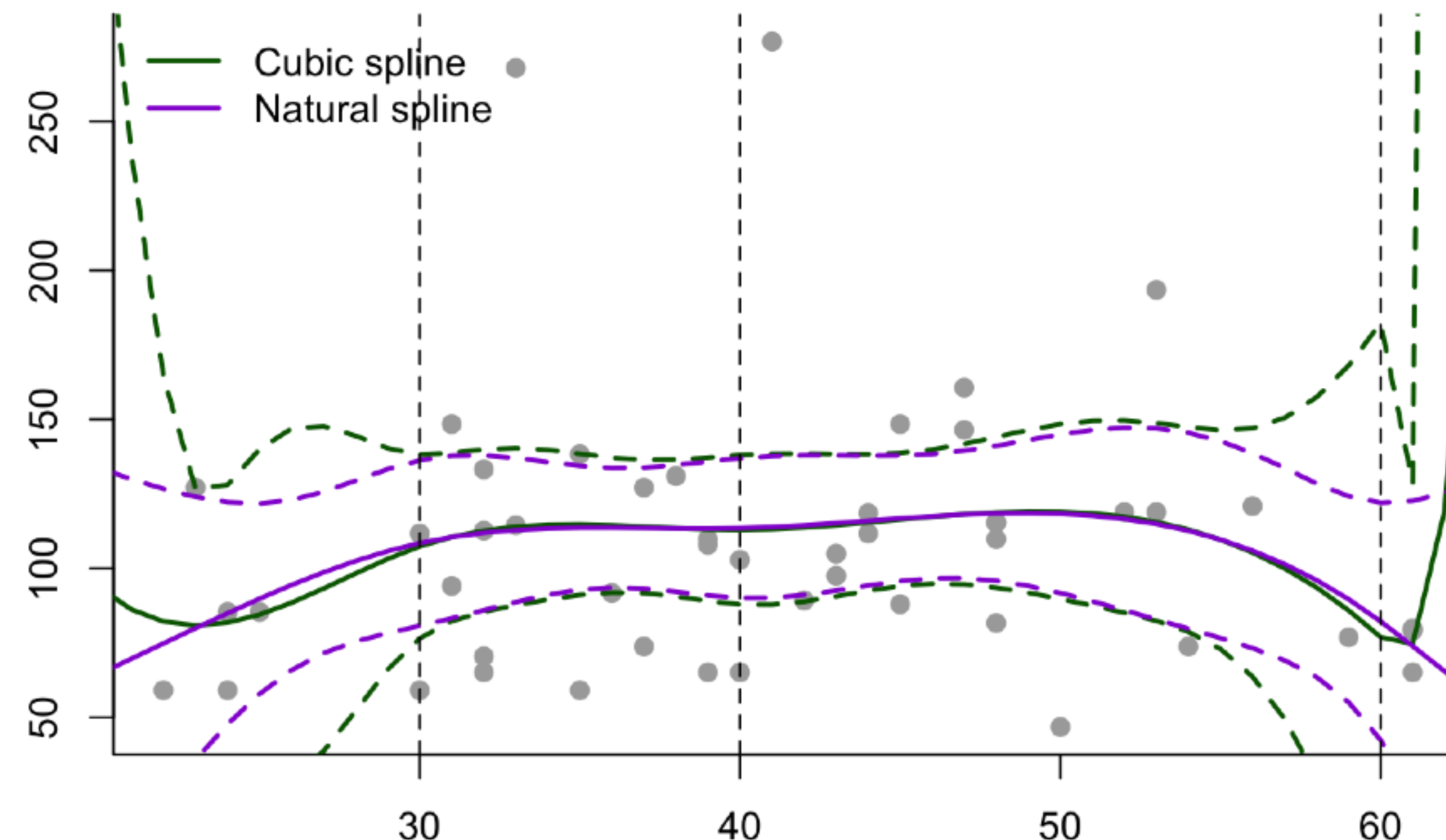# Natural Splines

- Regression splines have high variance at the outer range of the predictor (the tails)
- The confidence intervals at the tails can be wiggly (especially for small samples)

Natural splines are extensions of regression splines which remedy these problems

Two additional constraints at each boundary region:

1. The spline function is constrained to be close to linear when $X <$ smallest knot
2. The spline function is constrained to be close to linear when $X >$ largest knot

How Many Knots?

- Provided there is evidence from the data we can do it empirically:
  - ▸ Place knots where it is clearly obvious there is a distributional shift in direction
  - ▸ Place more knots on regions where we see more variability
  - ▸ Place fewer knots in places which look more stable

- Alternatively, we can place knots in a uniform fashion (25th, 50th, 75th percentiles)