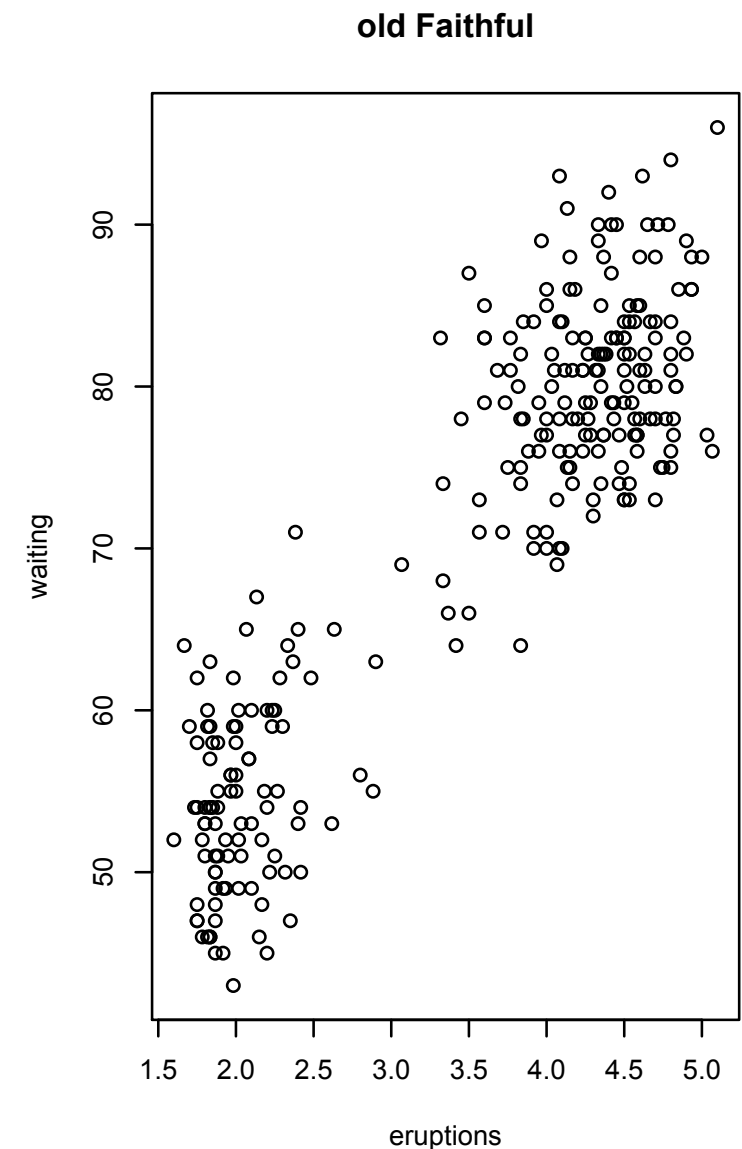
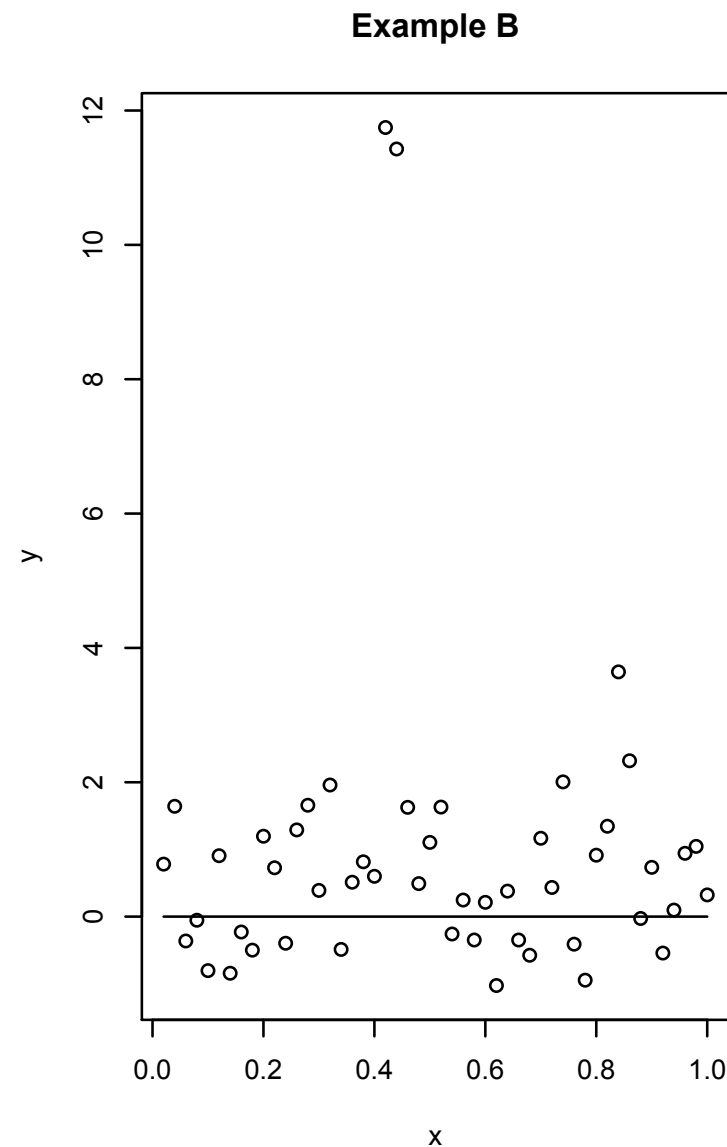
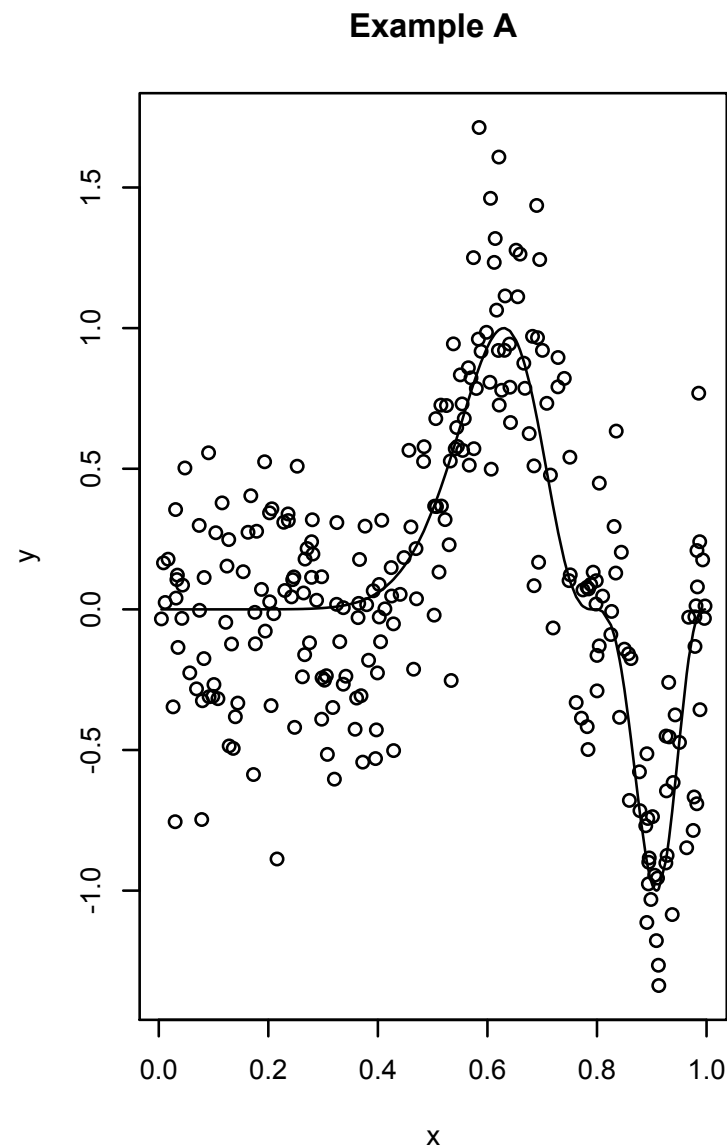


Nonparametric Regression: Examples

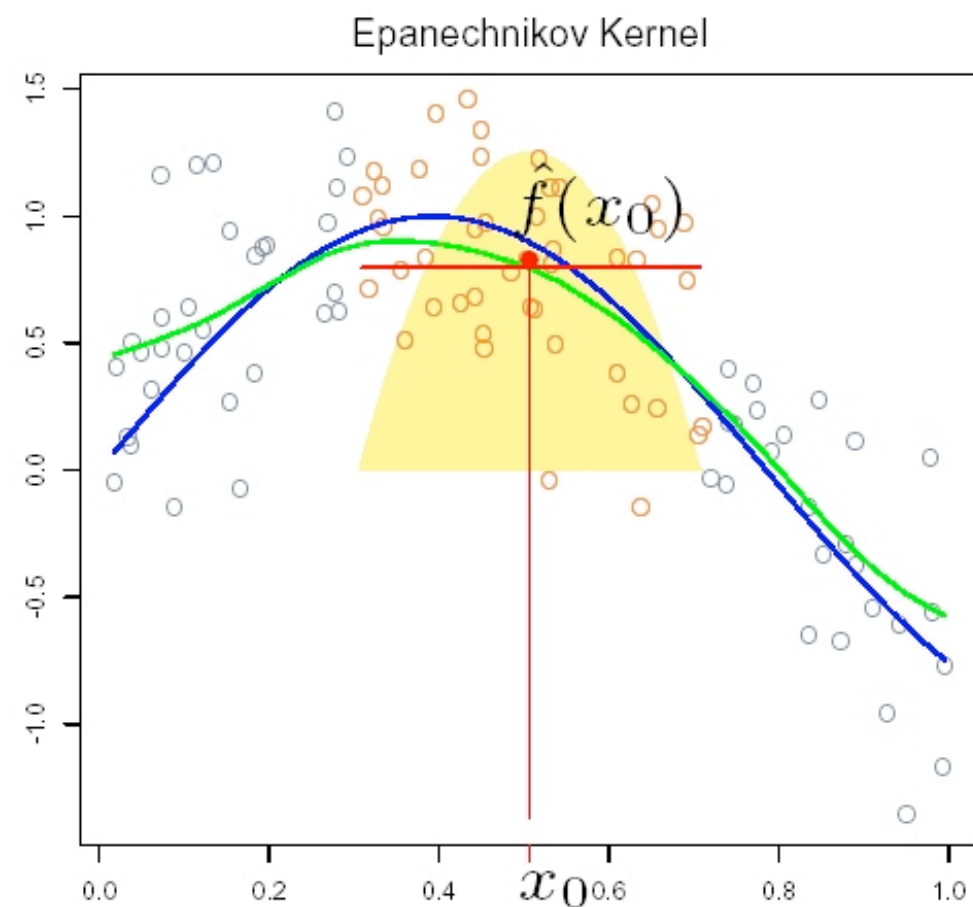
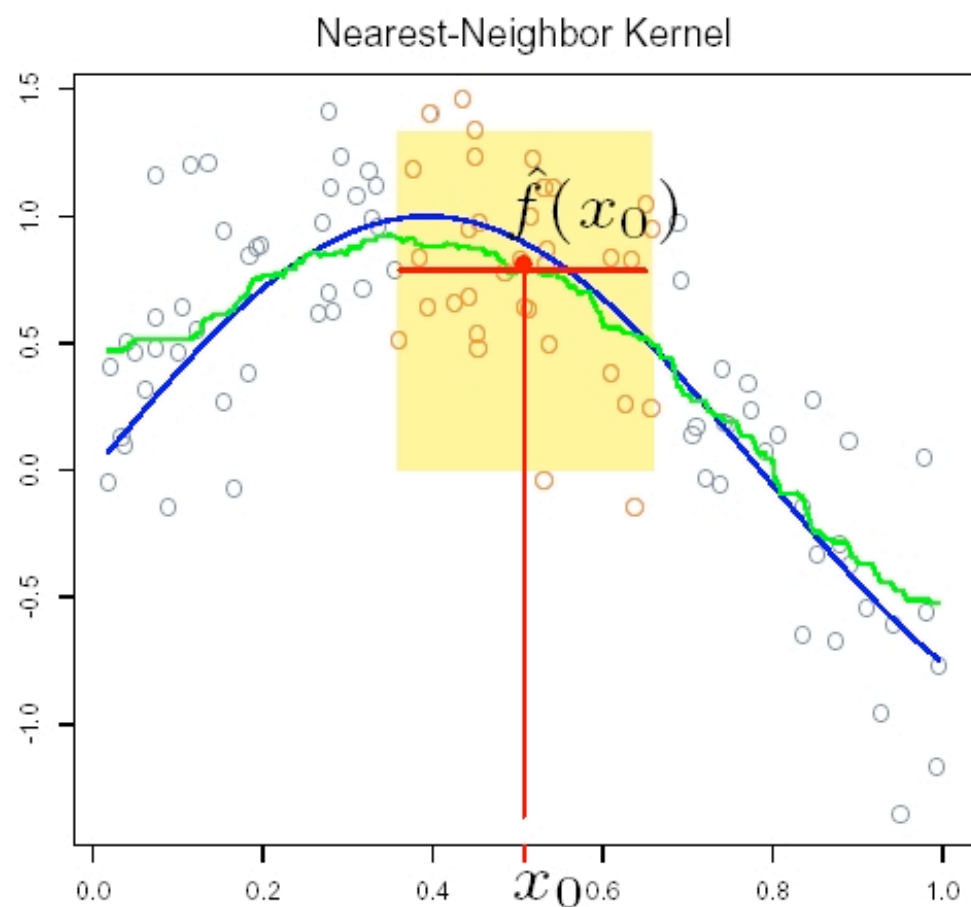


Data examples. **Example A** has varying amounts of curvature, two optima and a point of inflexion. **Example B** has two outliers. The **Old Faithful** provides the challenges of real data.

Kernel Estimators

$$\hat{f}(x) = \text{Ave}(y_i | x_i \in \mathcal{N}_k(x))$$

$$\hat{f}(x_0) = \frac{\sum_{i=1}^N K_\lambda(x_0, x_i) y_i}{\sum_{i=1}^N K_\lambda(x_0, x_i)}$$



$$\hat{f}_\lambda(x_0) = \frac{\sum_{i=1}^n w_i y_i}{\sum_i w_i},$$

where $w_i = K_\lambda(x_0, x_i)$

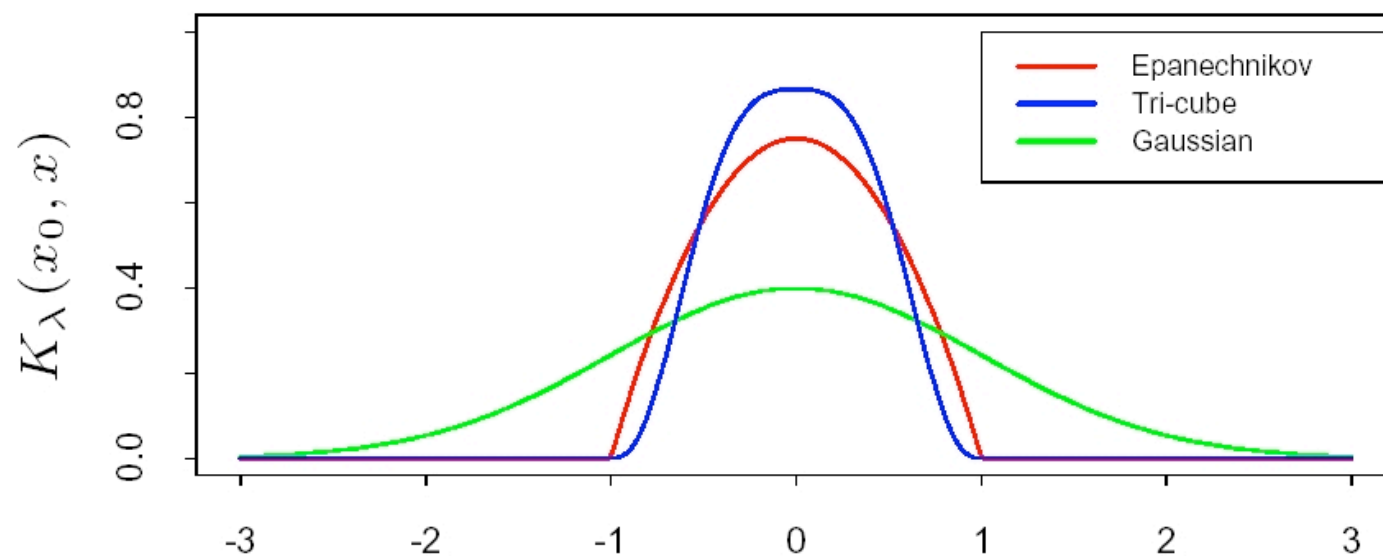
Kernels

$$K_{\lambda}(x_0, x) = D\left(\frac{|x - x_0|}{\lambda}\right)$$

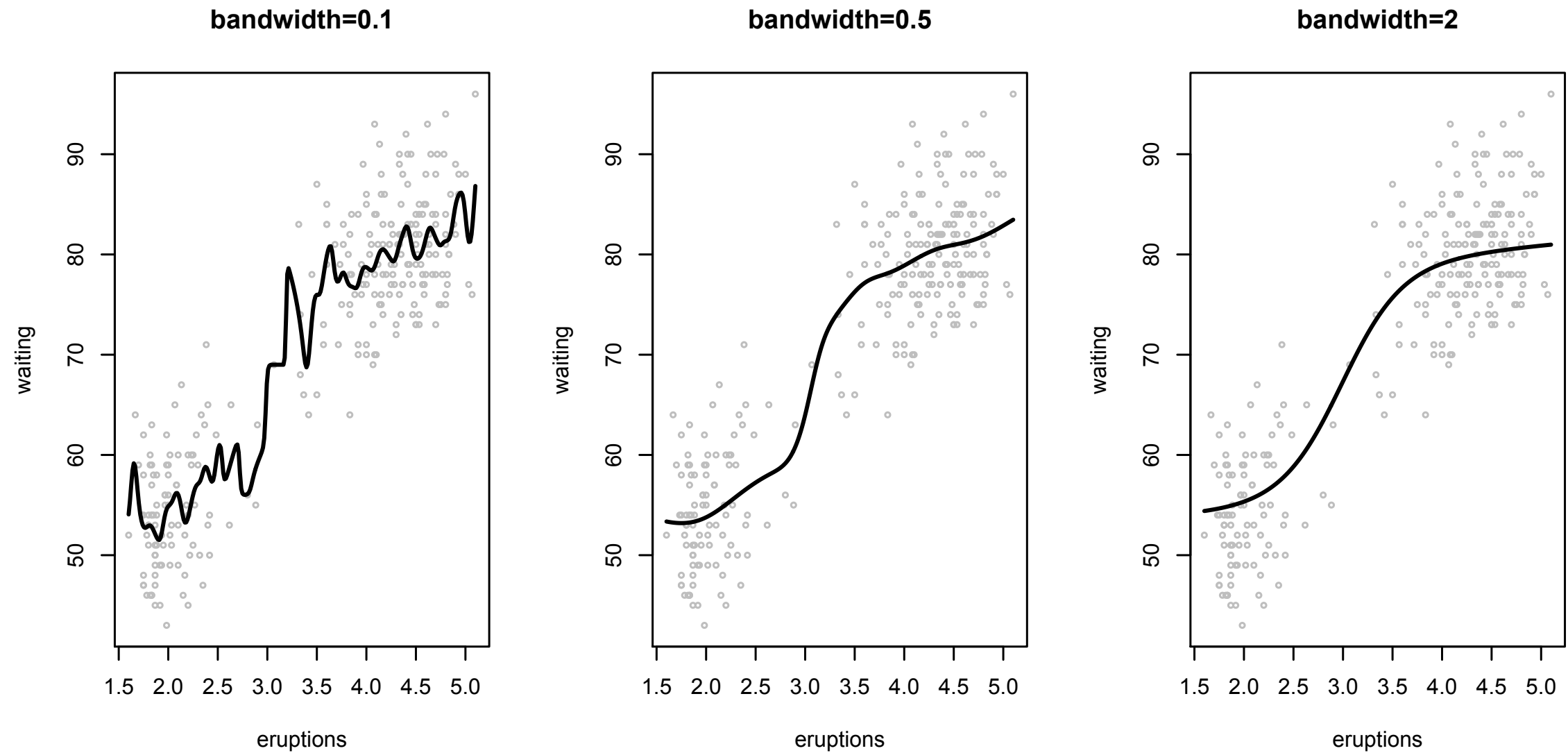
Gaussian : $D(t) = \phi(t) = e^{-\frac{1}{2}t^2}$

Epanechnikov : $D(t) = \begin{cases} \frac{3}{4}(1 - t^2) & \text{if } |t| < 1 \\ 0 & \text{otherwise} \end{cases}$

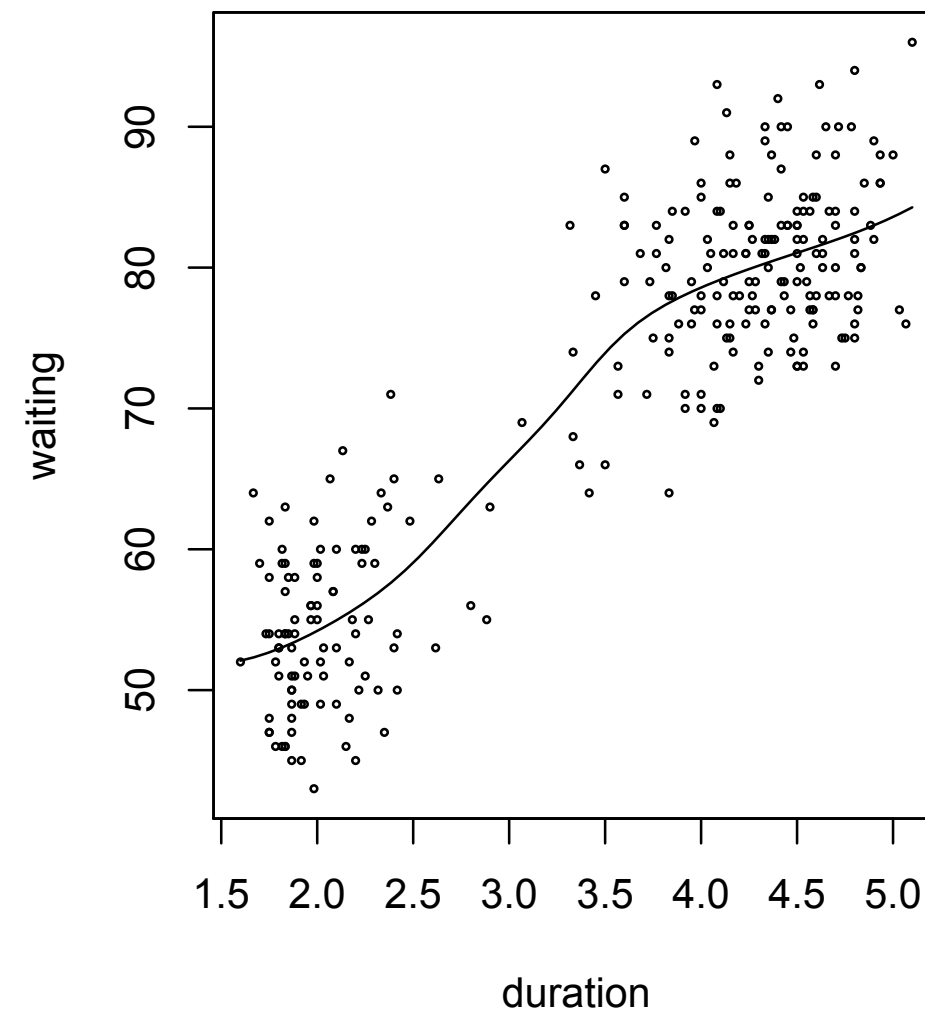
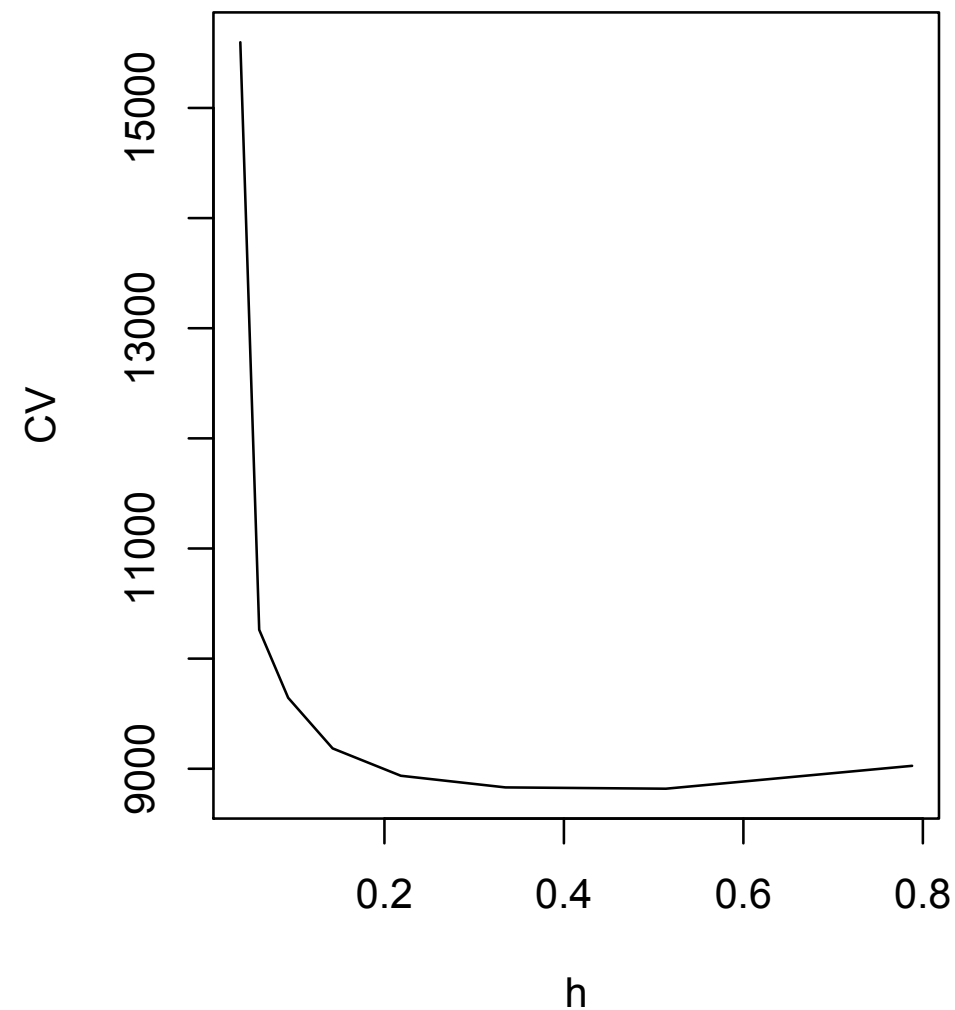
Tri-Cube : $D(t) = \begin{cases} (1 - t^3)^3 & \text{if } |t| < 1 \\ 0 & \text{otherwise} \end{cases}$



Choice of Bandwidth

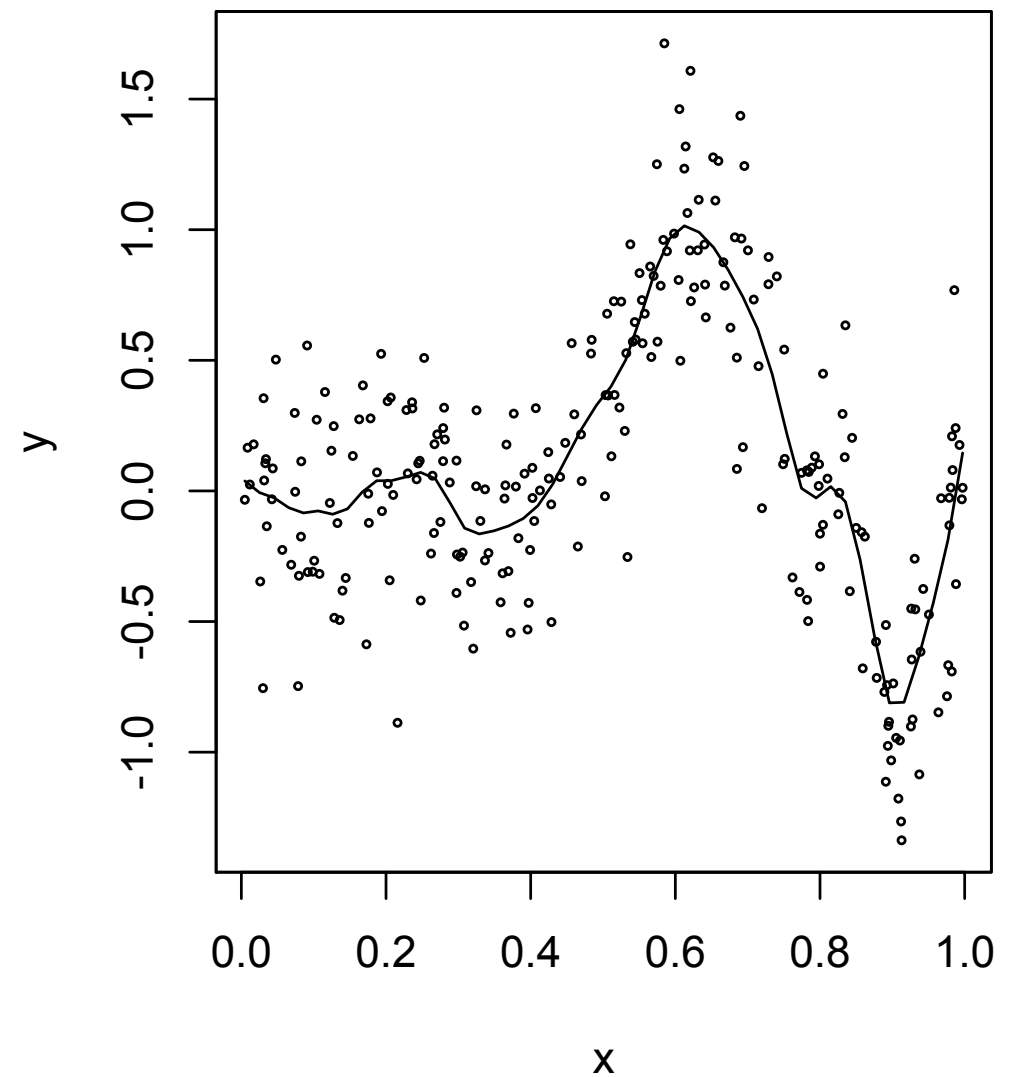
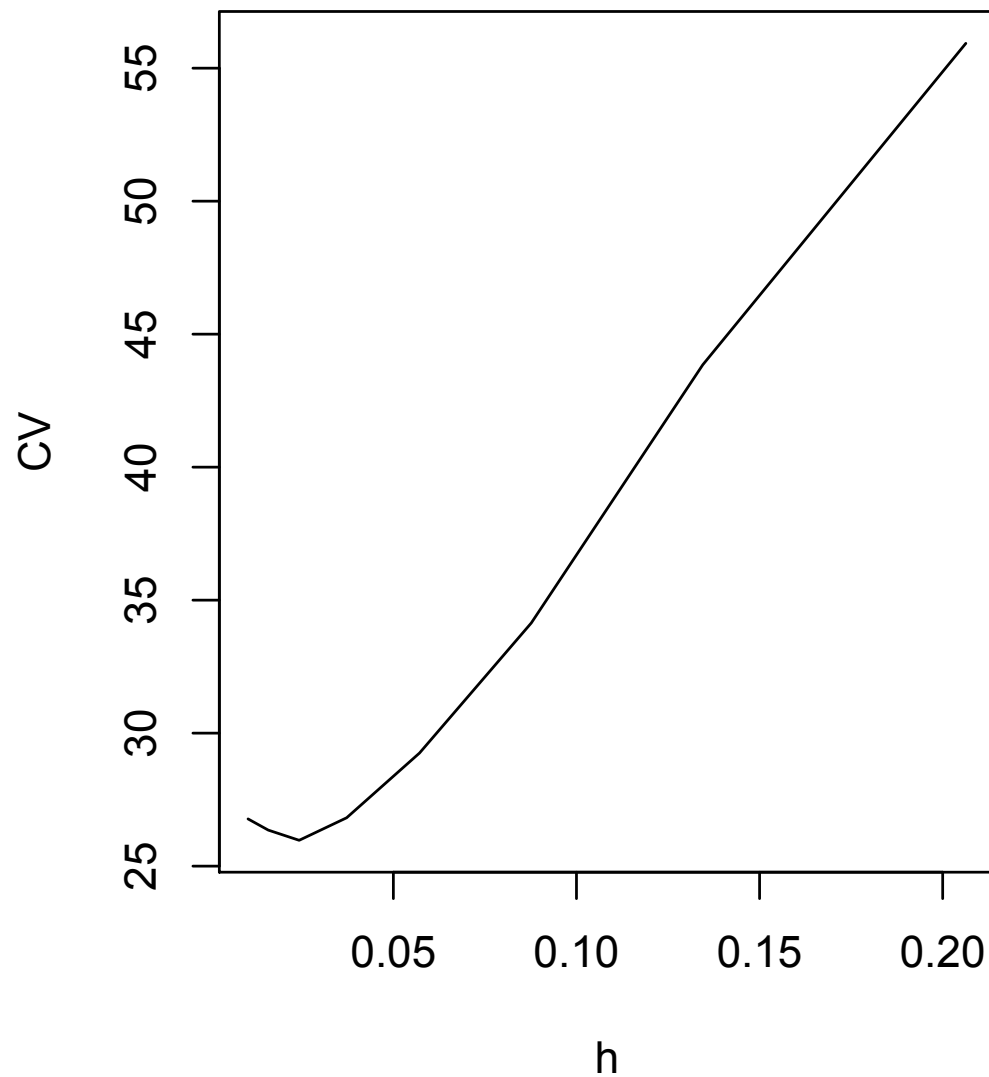


Normal kernel with three different bandwidths on the **Old Faithful data**.

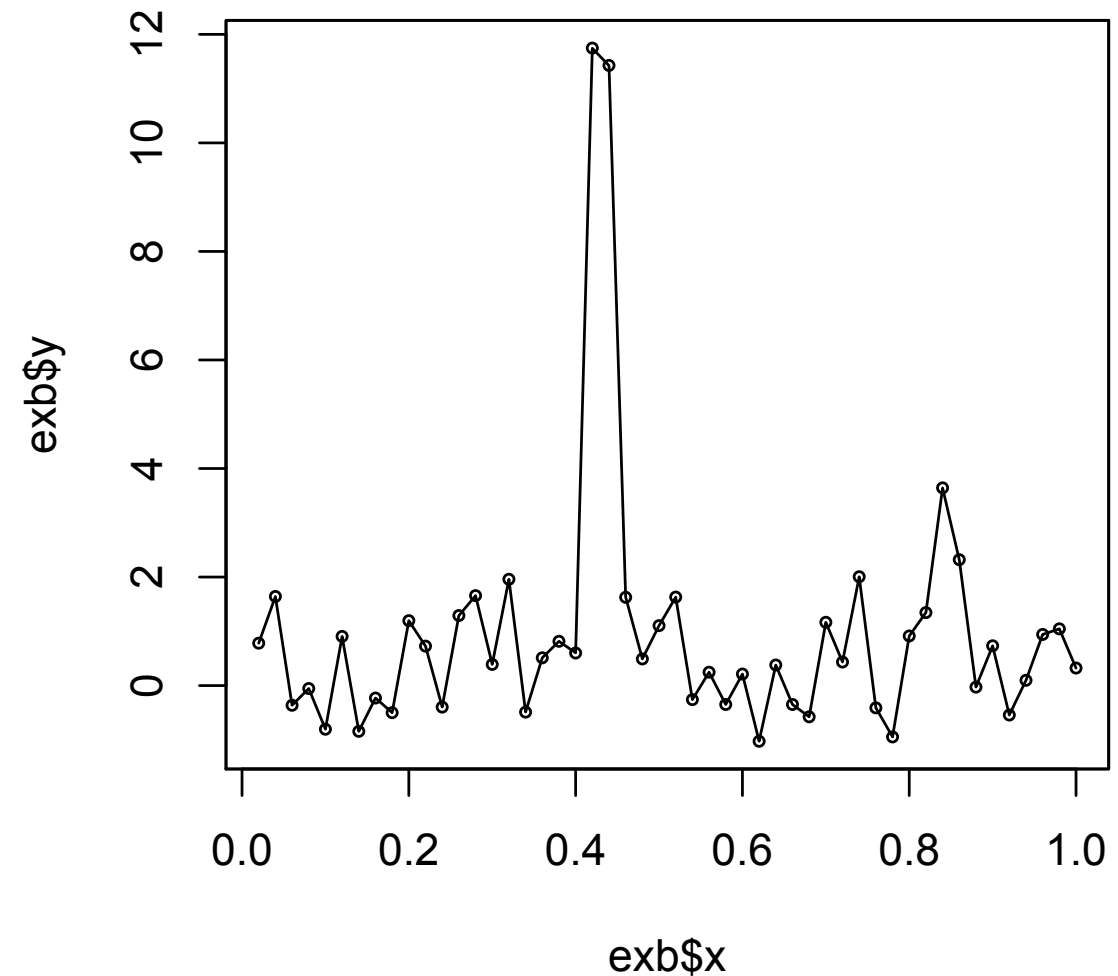
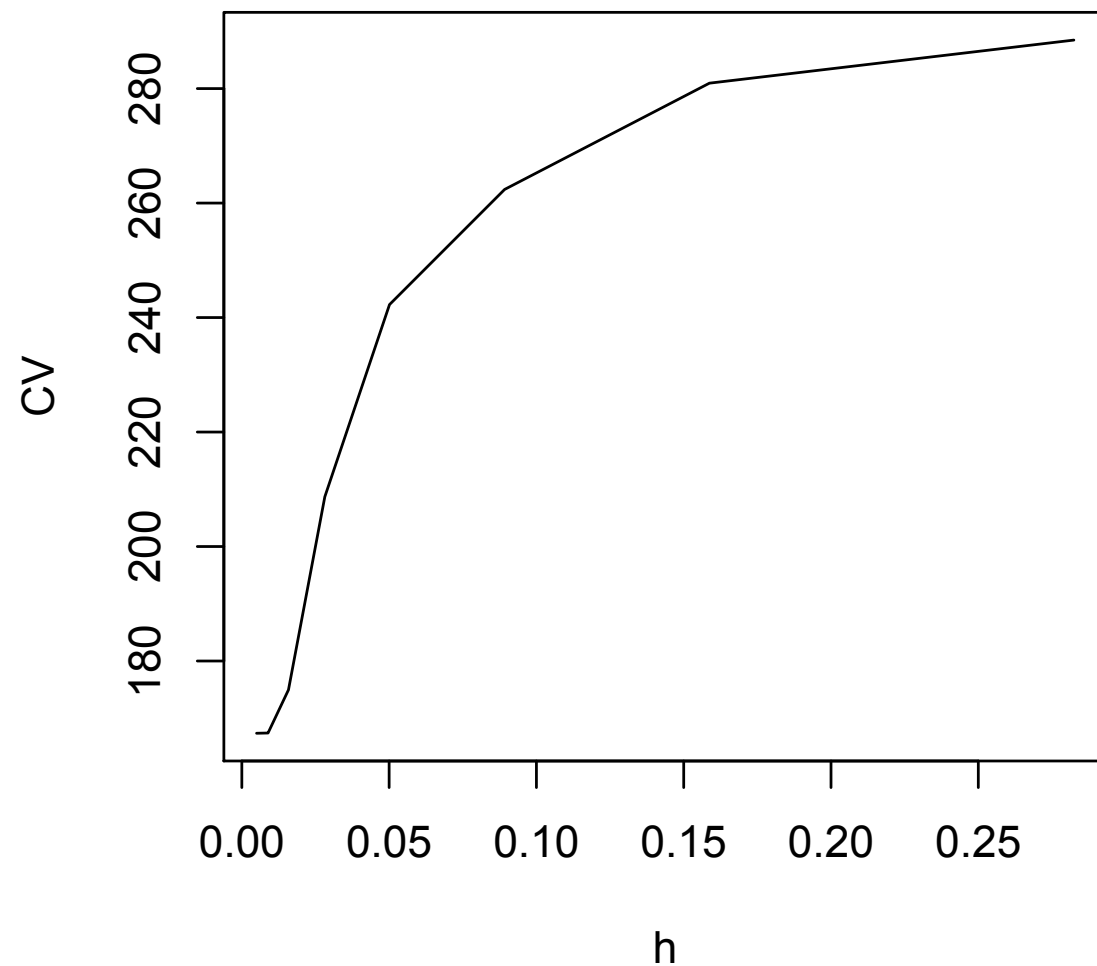


Left: The cross-validation criterion shown as a function of the smoothing parameter. The minimum occurs at a value of 0.424. Notice that the curve is quite flat in the region of the minimum indicating that a wide range of choices will produce acceptable results.

Right: the kernel estimate with this this value of the smoothing parameter.



CV and kernel smoother plots for **Example A**. The minimum of the cross-validation criterion is achieved at $h=0.022$.



CV and kernel smoother plots for **Example B**. the CV choice is at the lower boundary of suggested bandwidths, which is too small and leads to under-smoothing because the regression exactly fits the data.

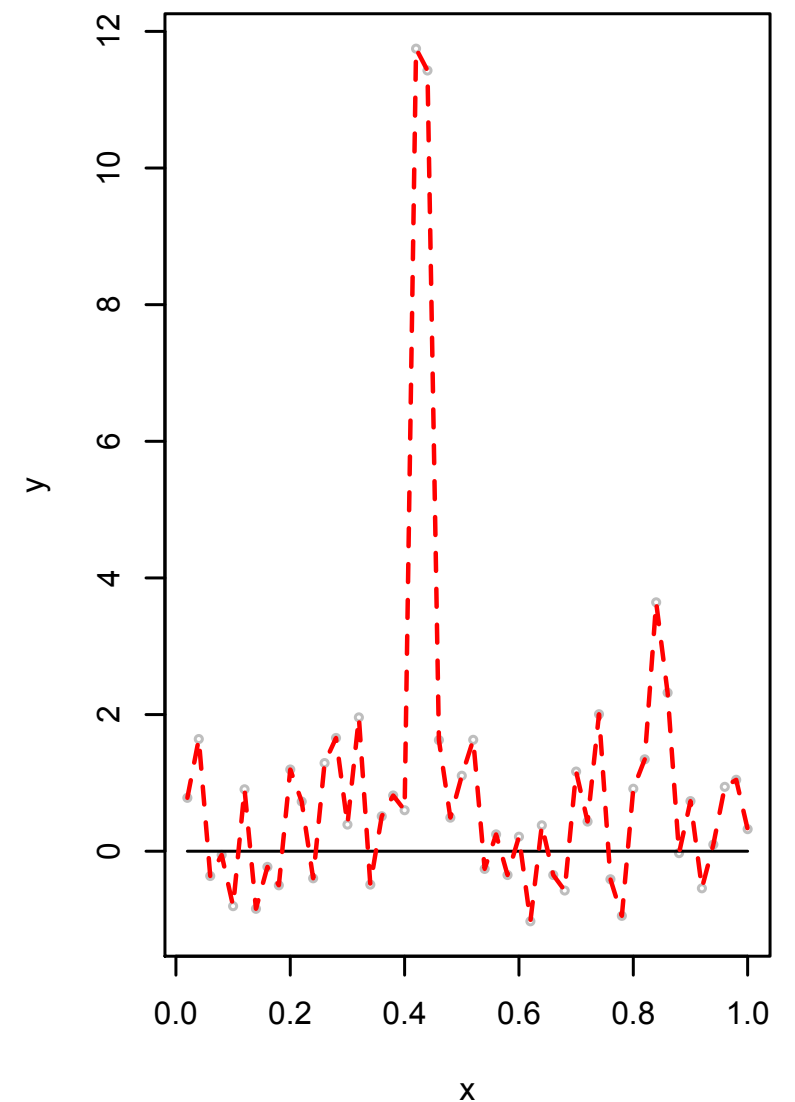
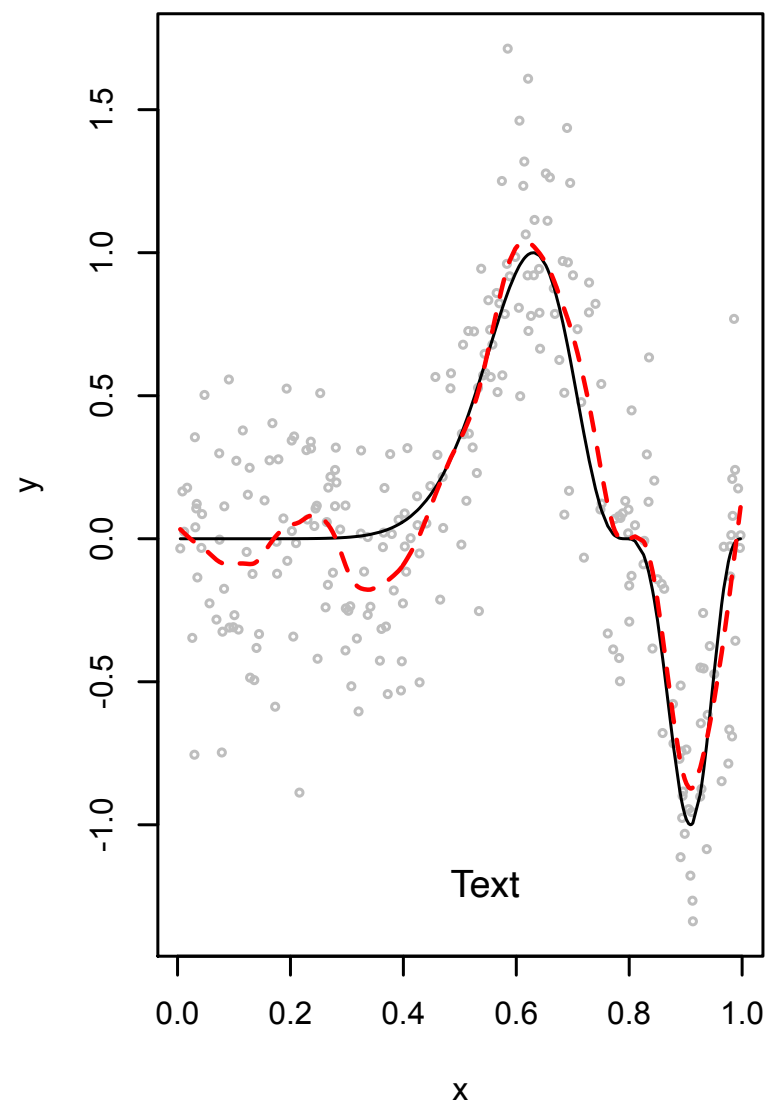
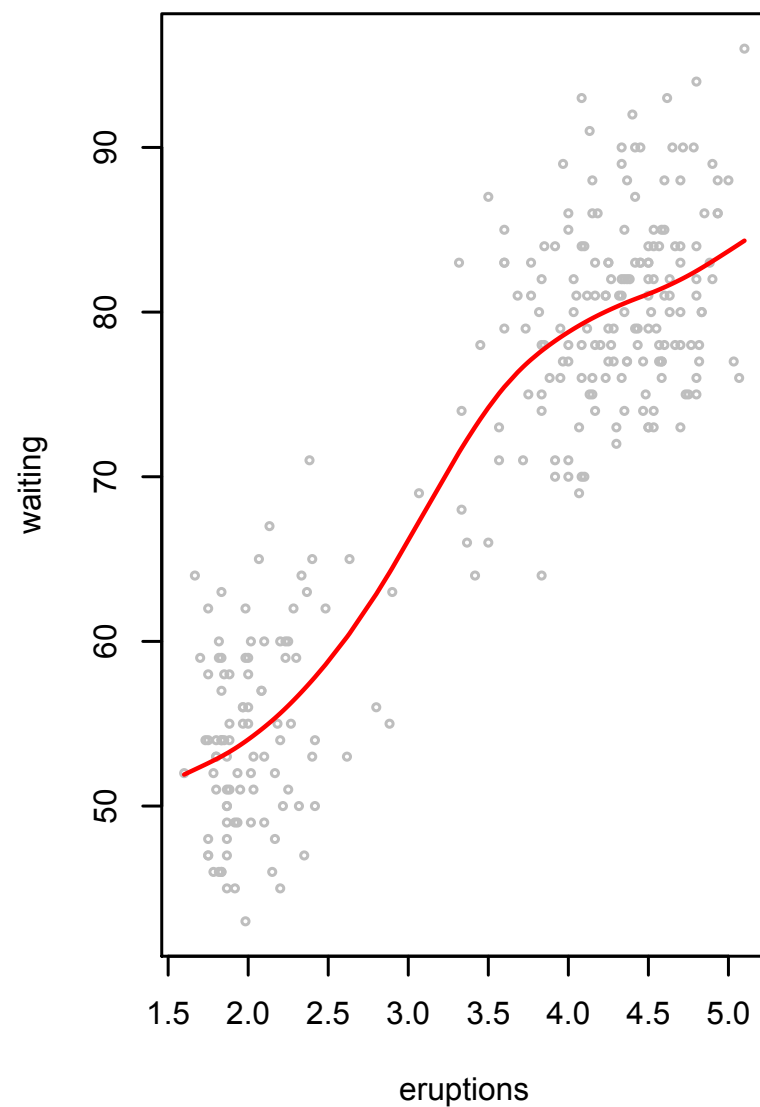
For this example, it'll be ideal to have different bandwidth for different range of the data.

Smoothing Splines

$$\frac{1}{n} \sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \int [f''(x)]^2 dx$$

The objective function = fitting error + roughness penalty.

The minimizer can be shown to be a natural cubic spline, that is, f is continuous up to the 2nd derivative.



Smoothing spline fits. For Examples A and B, the true function is shown as solid and the spline fit as dotted.

In R, CV is used by default to choose the smoothing parameter (λ).

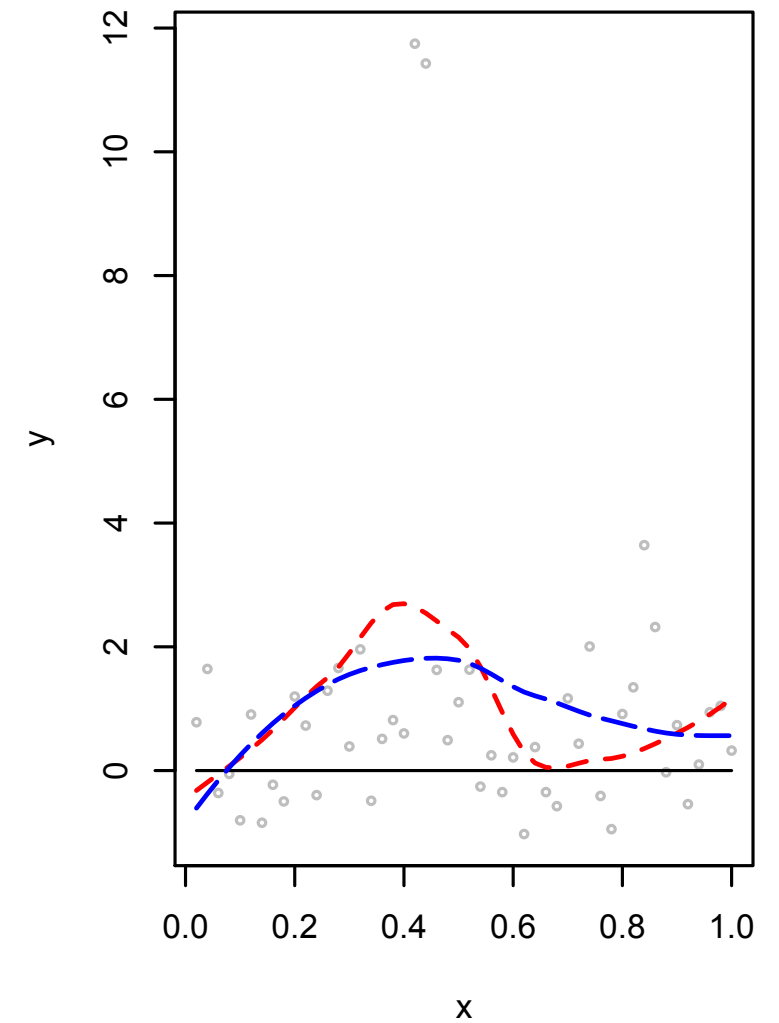
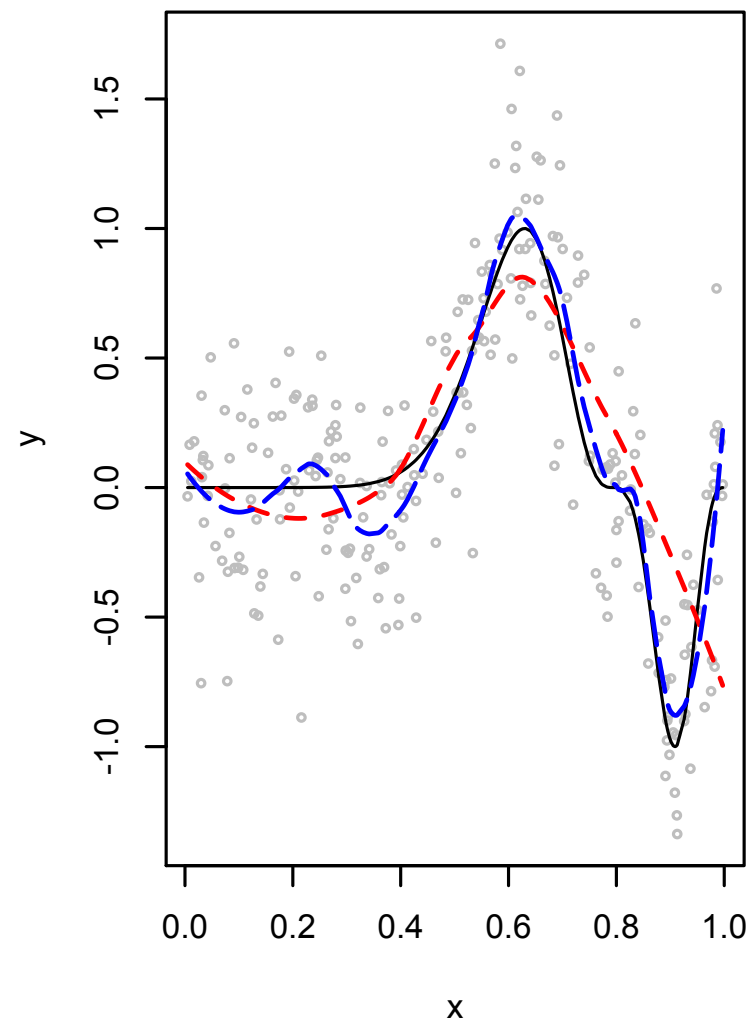
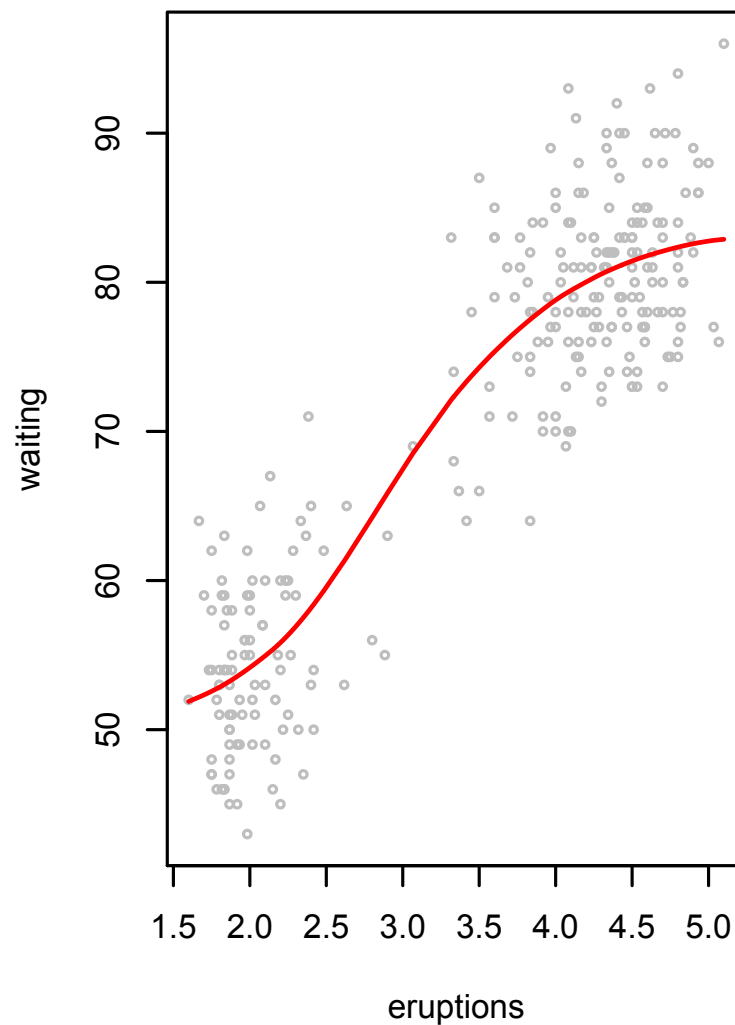
Local Polynomials

Both the kernel and spline methods have been relatively vulnerable to outliers as seen by their performance on Example B.

The **local polynomial** method combines robustness ideas from linear regression and local fitting ideas from kernel methods.

First we select a window. We then fit a polynomial to the data in that window using robust methods. The predicted response at the middle of the window is the fitted value. We then simply slide the window over the range of the data, repeating the fitting process as the window moves.

The most well-known implementation of this type of smoothing is called loess (Cleveland, 1979).



Faithful Data: use the default choice of loess which fits a local quadratic function using 75% of the data (span = 0.75).

Example A: default choice in red, and the optimal one in blue (span = 0.22)

Example B: the optimal choice for span is 1 (i.e., use all the data). We can see that the fitting isn't affected that much by the two outliers.