UNIVERSITY OF LEEDS

Quantitative Social Research II

Workshop 4: Non-Linear Effects

Jose Pina-Sánchez

Workshop Aims

- Review the linearity assumption
- Learn how to model non-linear relationships within the constraints of the linear model
  - polynomial regression
  - local regression

Workshop Aims: Recap

- Assumptions in the linear regression model ($Y = \alpha + \beta_k X_k + e$):
  - normality: residuals are normally distributed
  - homoskedasticity: the variance of the residuals is constant
  - independence: residuals are independent of each other
  - no multicollinearity
  - perfectly measured variables
  - no missing data (other than missing at random)
  - no unobserved confounders: we control for all common causes of X_1 and Y
  - no reverse causality: Y does not cause X_1
  - **linearity**: the effect of X_1 on Y is the same across the range of X_1

Linearity

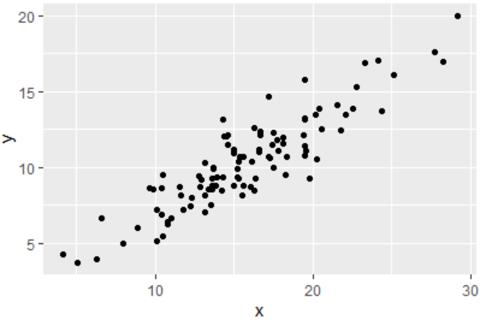
- On average, the change in Y is proportional to the change in X
  - each explanatory variable is multiplied by a coefficient and summed up

    $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_k X_{ki} + e_i$
  - and the average magnitude of e_i is assumed to be the same for all values of X_i

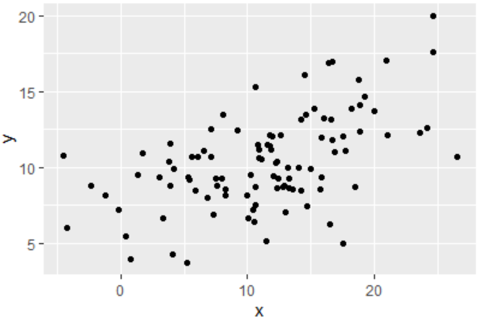
Linearity

- On average, the change in Y is proportional to the change in X
  - each explanatory variable is multiplied by a coefficient and summed up
    $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_k X_{ki} + e_i$
  - and the average magnitude of e_i is assumed to be the same for all values of X_i
- If that is not the case, then the model residuals won't be...
  - exogenous, $Cov(X_i, e_i) = 0$
  - independent, $Cov(e_i, e_j) = 0$,
  - with constant variance, $Var(e_i) = Var(e)$
  - or normally distributed, $N \sim (0, Var(e))$

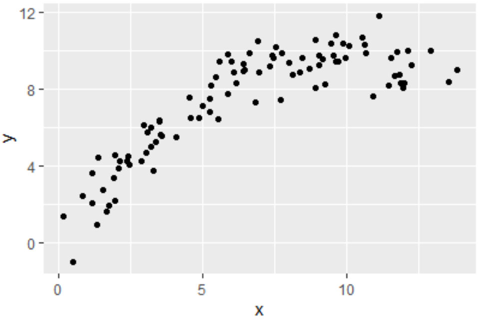
Linear or Non-Linear?



Linear or Non-Linear?



Linear or Non-Linear?



Non-Linear Models

- One way to deal with non-linearity is by transforming the dependent variable
  - we have used log(Y) for right-skewed data
  - and logit(Y) for binary data
  - we can accommodate many other non-normal distributions using 'generalised linear models'
  - which will invoke different parametric assumptions for the residuals
  - Poisson, negative binomial, exponential, gamma, etc.

Non-Linear Models

- One way to deal with non-linearity is by transforming the dependent variable
  - we have used log(Y) for right-skewed data
  - and logit(Y) for binary data
  - we can accommodate many other non-normal distributions using 'generalised linear models'
  - which will invoke different parametric assumptions for the residuals
  - Poisson, negative binomial, exponential, gamma, etc.
- These models however will modify the relationship between Y and all the X_k
  - e.g. if using logit(Y) all β_k are interpreted as changes in the log-odds of Y for a one change in X_k
  - but what if the non-linear relationship only affects one (or a subset) of X_k?
- Then we can use polynomial regression

Polynomial Regression

- We can extend the linear model by raising explanatory variables to a given power
  - let's say we suspect X_1 to have a non-linear relationship with Y
  - we can add a new term X_1^2 in the model to explore a quadratic relationship between X_1 and Y (we allow a point of inflection)

Polynomial Regression

- We can extend the linear model by raising explanatory variables to a given power
  - let's say we suspect X_1 to have a non-linear relationship with Y
  - we can add a new term X_1^2 in the model to explore a quadratic relationship between X_1 and Y (we allow a point of inflection)

$$\overbrace{Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k + e}^{\textit{linear model}}$$

$$\overbrace{Y = \beta_0 + (\beta_{1,1} X_1 + \beta_{1,2} X_1^2) + \beta_2 X_2 + ... + \beta_k X_k + e}^{\textit{linear model with a quadratic effect for } X_1}$$

the effect of X_1 on Y will now be represented by the term within ()

Polynomial Regression

- We can extend the linear model by raising explanatory variables to a given power
  - let's say we suspect X_1 to have a non-linear relationship with Y
  - we can add a new term X_1^2 in the model to explore a quadratic relationship between X_1 and Y (we allow a point of inflection)

$$\overbrace{Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k + e}^{\text{linear model}}$$

$$\overbrace{Y = \beta_0 + (\beta_{1,1} X_1 + \beta_{1,2} X_1^2) + \beta_2 X_2 + ... + \beta_k X_k + e}^{\text{linear model with a quadratic effect for } X_1}$$

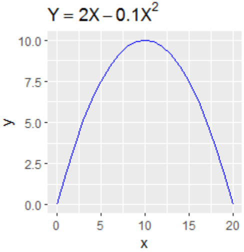
    the effect of X_1 on Y will now be represented by the term within ()

  - we can also explore cubic a relationship (two points of inflection)

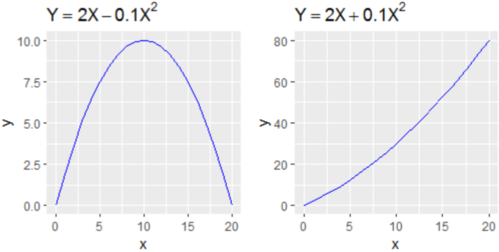
$$\overbrace{Y = \beta_0 + (\beta_{1,1} X_1 + \beta_{1,2} X_1^2 + \beta_{1,3} X_1^3) + \beta_2 X_2 + ... + \beta_k X_k + e}^{\text{linear model with a cubic effect for } X_1}$$

  - we can add more points of inflection but such relationships are rare and we will inevitably incur in problems of multicollinearity

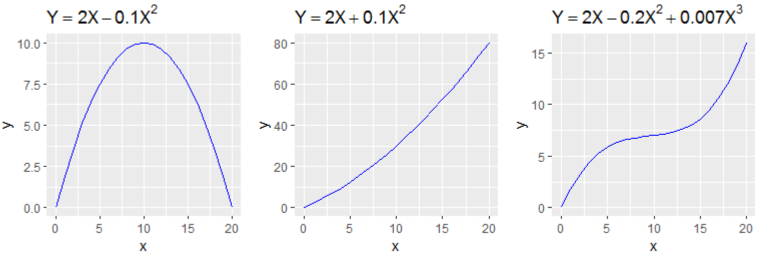
Polynomial Regression: Visually



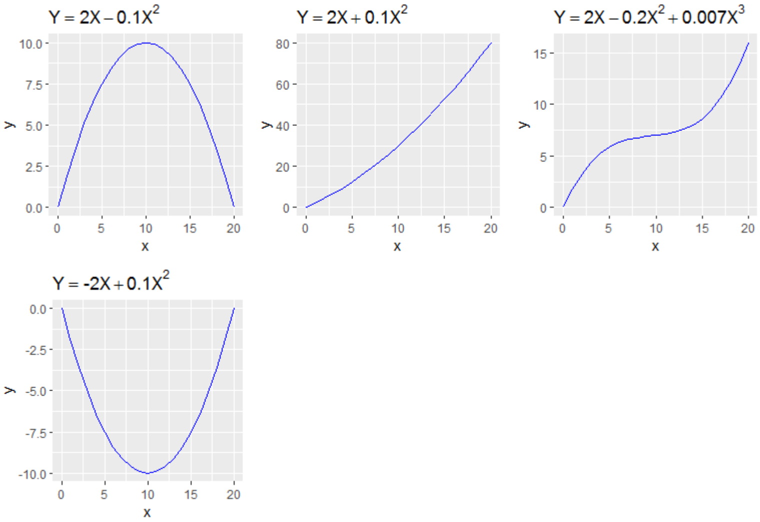
Polynomial Regression: Visually



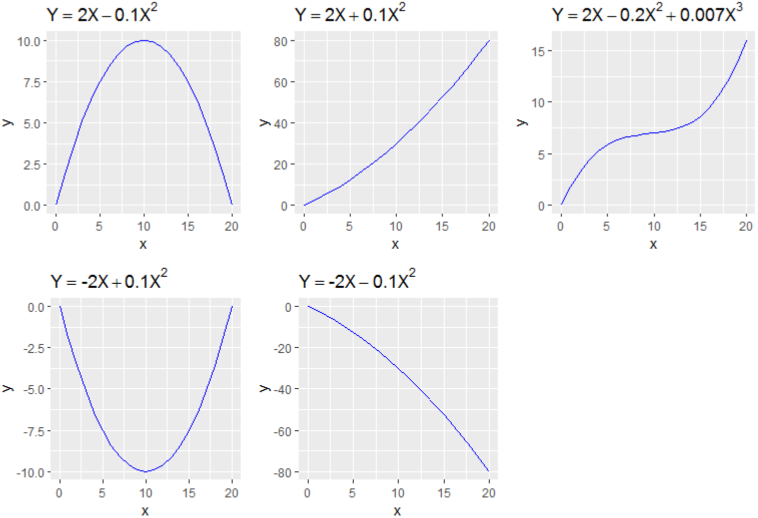
Polynomial Regression: Visually



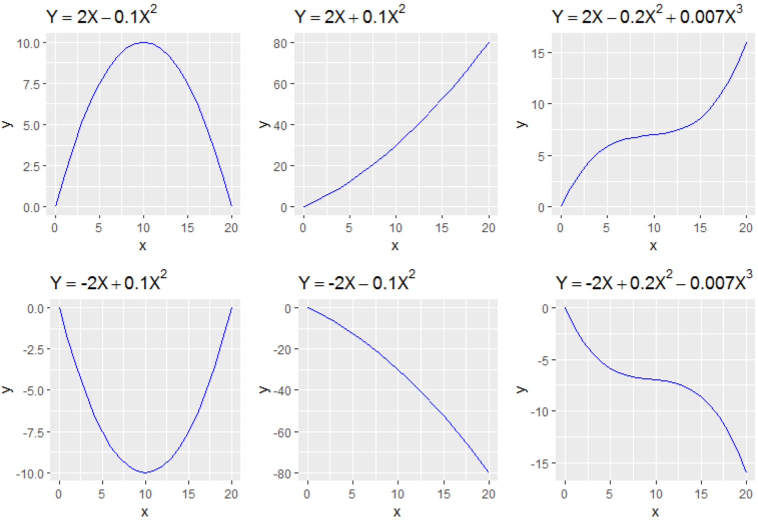
Polynomial Regression: Visually



Polynomial Regression: Visually



Polynomial Regression: Visually



Polynomial Regression

- A simple way to provide a non-linear fit to the relationship between Y and one or more X
  - if we use it in the context of a standard linear model we can keep using OLS
  - we can also use polynomial terms in generalised linear models
- It is still subject to a number of limitations
  - can induce multicollinearity
  - provides additional flexibility but it is still parametric (follows a given function)
  - we need to be extra careful when it comes to extrapolations

Locally Weighted Regression (LOESS)

- LOESS is ‘local’
  - breaks down the sample into different ‘neighborhoods’ around a specific location, X_0
  - the neighborhood is defined as the span (aka bandwidth), i.e. the fraction of the total points used to form neighborhoods
  - e.g. for a span of 0.5 the closest half of the total number of cases is used
  - the same model is estimated in each neighbourhood using different cases
  - obtaining different β across the range of X
- LOESS is ‘weighted’
  - in each of the models estimated cases closer to X_0 are given more weight

Weighted Regression

- For the case of simple linear regression based on OLS
  - we seek to minimise the sum squared error terms, $\Sigma(Y_i - \beta_0 - \beta_1 X_i)^2$
  - solving the system of equations we get the estimators for the constant and the slope as

    $\hat{\beta}_0 = \bar{Y} - \bar{X}\hat{\beta}_1$; and $\hat{\beta}_1 = \dfrac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\Sigma(X_i - \bar{X})^2}$

    where $\bar{Y} = \dfrac{\Sigma Y_i}{N}$; and $\bar{X} = \dfrac{\Sigma X_i}{N}$

Weighted Regression

- For the case of simple linear regression based on OLS
  - we seek to minimise the sum squared error terms, $\Sigma(Y_i - \beta_0 - \beta_1 X_i)^2$
  - solving the system of equations we get the estimators for the constant and the slope as

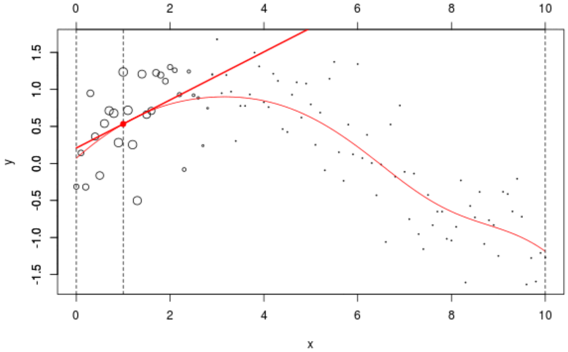
    $\hat{\beta}_0 = \bar{Y} - \bar{X}\hat{\beta}_1$; and $\hat{\beta}_1 = \dfrac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\Sigma(X_i - \bar{X})^2}$

    where $\bar{Y} = \dfrac{\Sigma Y_i}{N}$; and $\bar{X} = \dfrac{\Sigma X_i}{N}$

- For weighted least squares we seek to minimise
  - $\Sigma W_i(Y_i - \beta_0 - \beta 1 X_i)^2$; which leads to
  - $\hat{\beta}_0 = \bar{Y}_W - \bar{X}_W\hat{\beta}_1$; and $\hat{\beta}_1 = \dfrac{\Sigma W_i(X_i - \bar{X}_W)(Y_i - \bar{Y}_W)}{\Sigma W_i(X_i - \bar{X}_W)^2}$

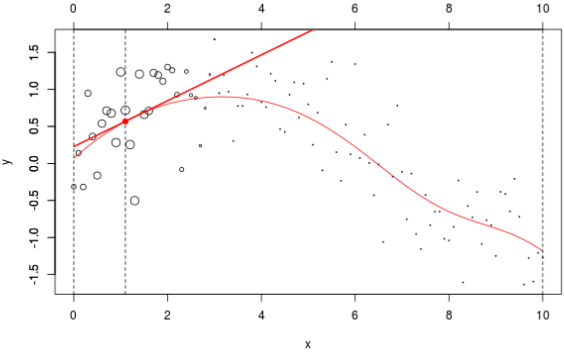
    where $\bar{Y} = \dfrac{\Sigma W_i Y_i}{\Sigma W_i}$; and $\bar{X} = \dfrac{\Sigma W_i X_i}{\Sigma W_i}$

LOESS: Visually



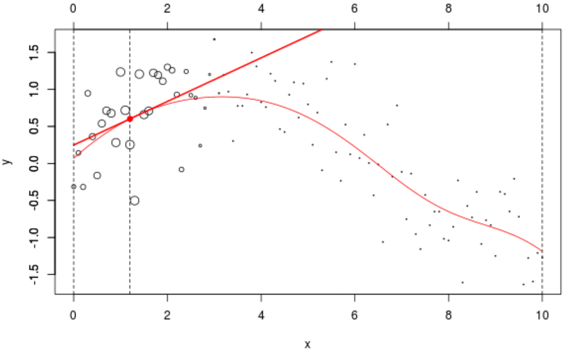
Source: A. Charpentier

LOESS: Visually



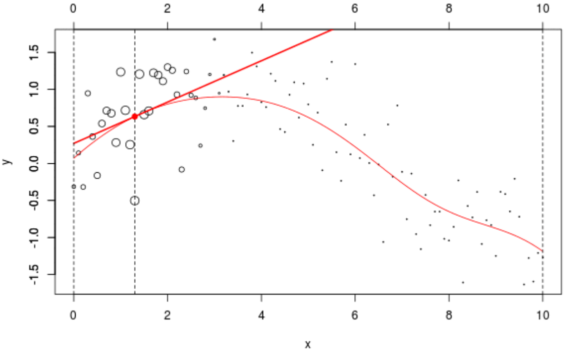
Source: A. Charpentier

LOESS: Visually



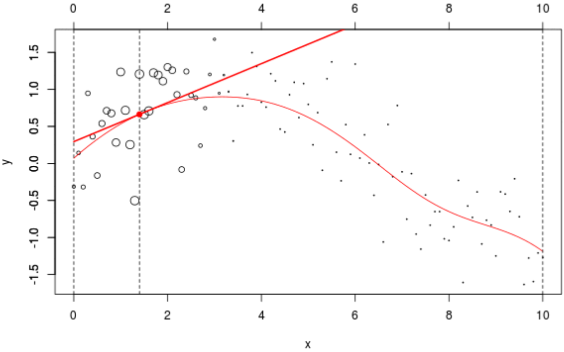
Source: A. Charpentier

LOESS: Visually



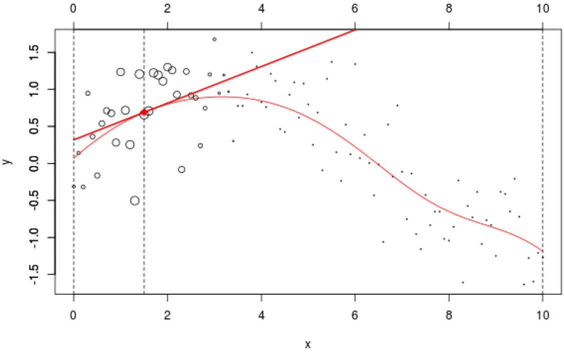
Source: A. Charpentier

LOESS: Visually



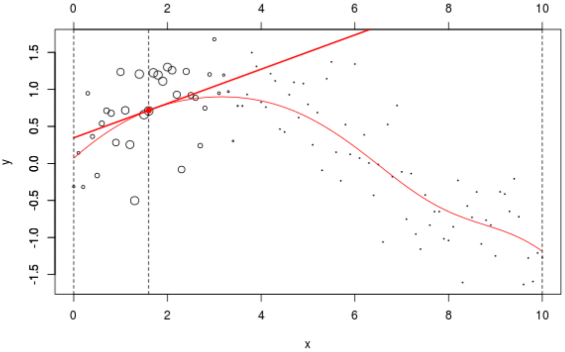
Source: A. Charpentier

LOESS: Visually



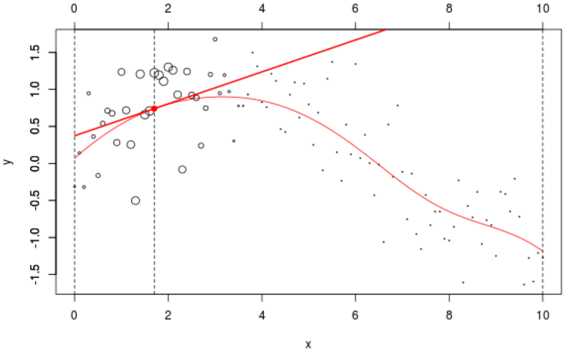
Source: A. Charpentier

LOESS: Visually



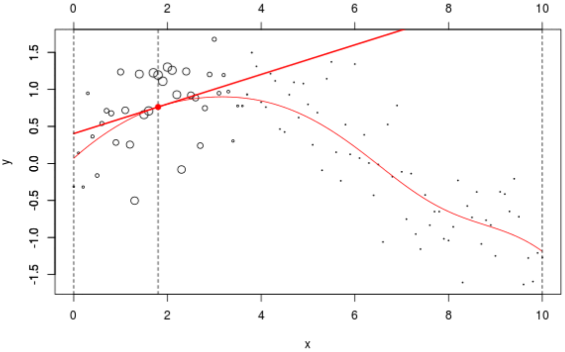
Source: A. Charpentier

LOESS: Visually



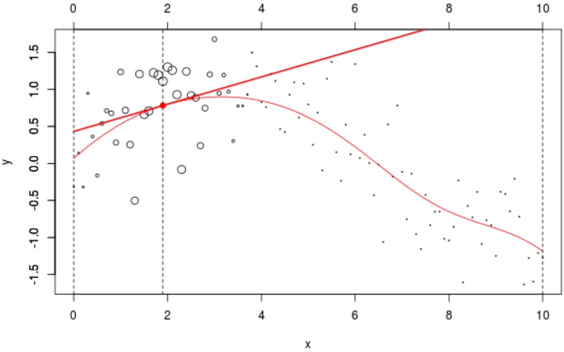
Source: A. Charpentier

LOESS: Visually



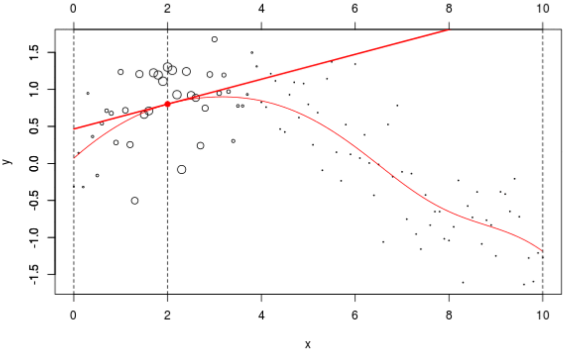
Source: A. Charpentier

LOESS: Visually



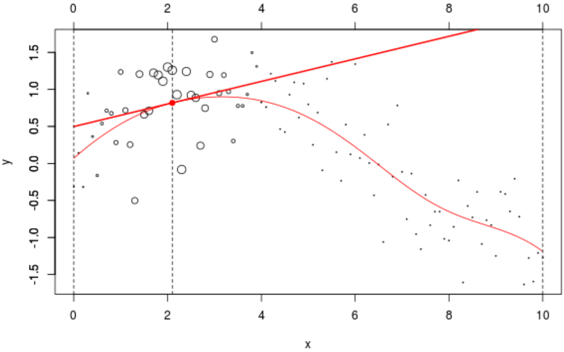
Source: A. Charpentier

LOESS: Visually



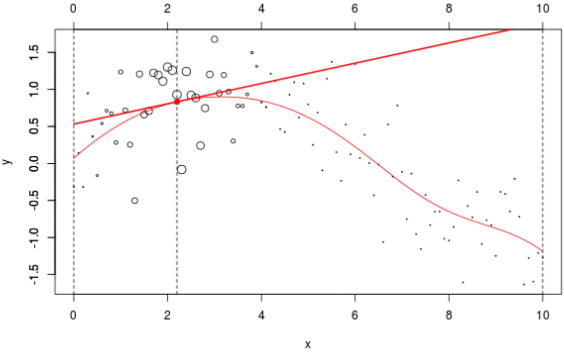
Source: A. Charpentier

LOESS: Visually



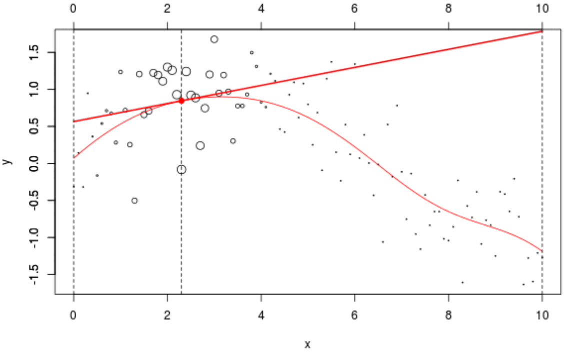
Source: A. Charpentier

LOESS: Visually



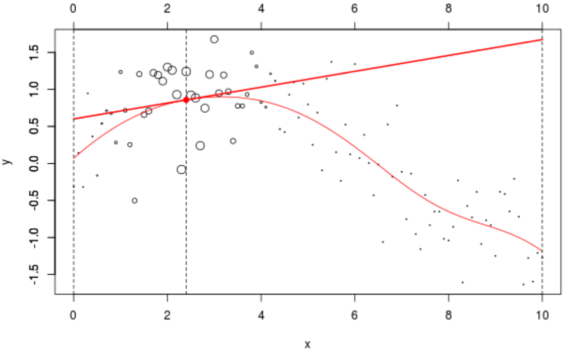
Source: A. Charpentier

LOESS: Visually



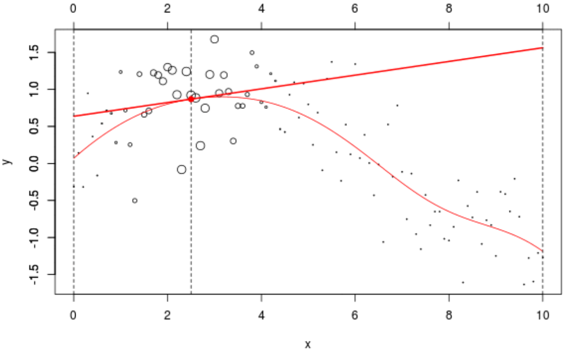
Source: A. Charpentier

LOESS: Visually



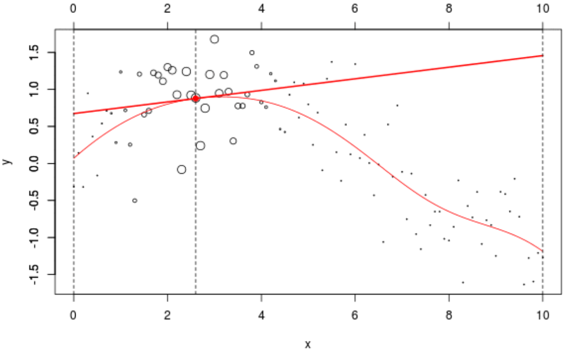
Source: A. Charpentier

LOESS: Visually



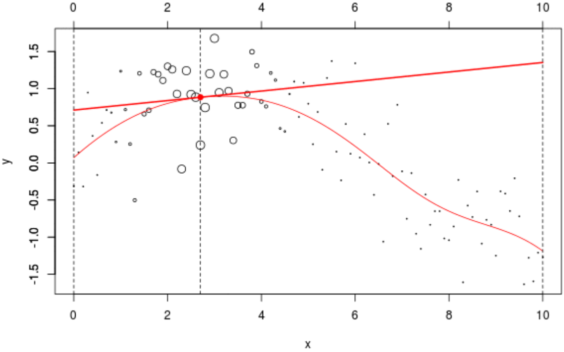
Source: A. Charpentier

LOESS: Visually



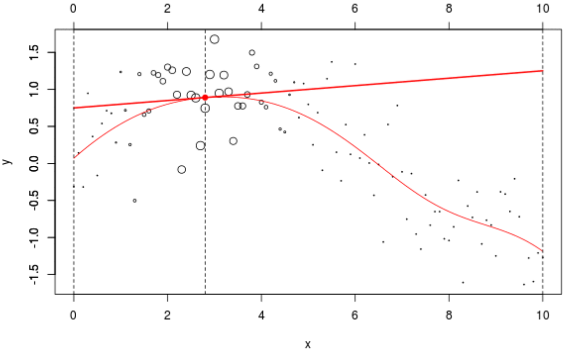
Source: A. Charpentier

LOESS: Visually



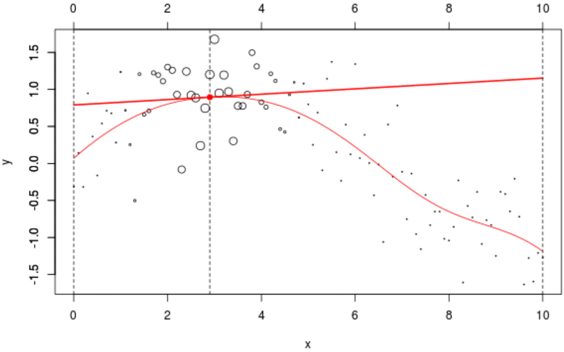
Source: A. Charpentier

LOESS: Visually



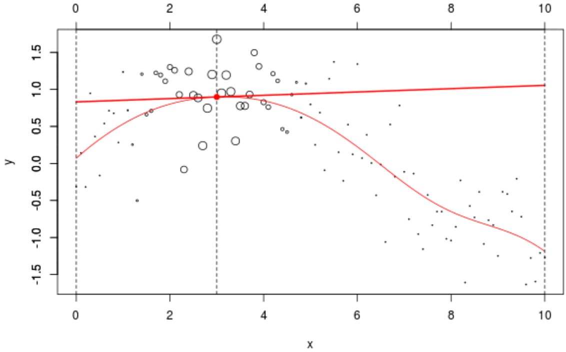
Source: A. Charpentier

LOESS: Visually



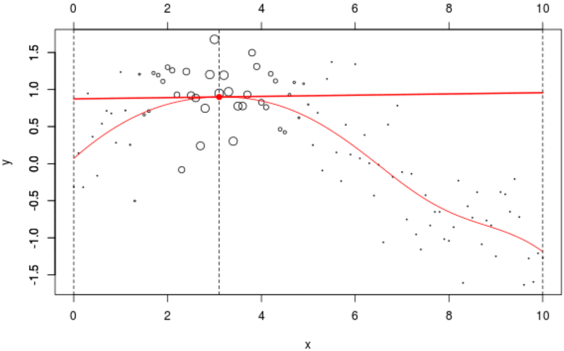
Source: A. Charpentier

LOESS: Visually



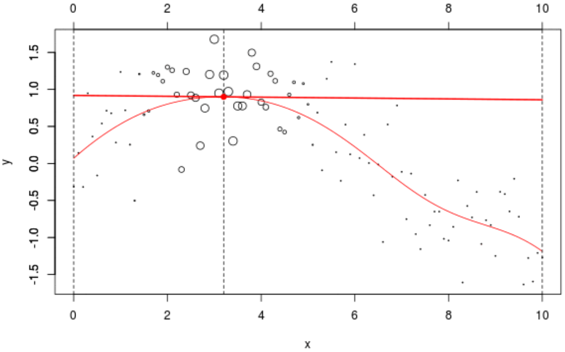
Source: A. Charpentier

LOESS: Visually



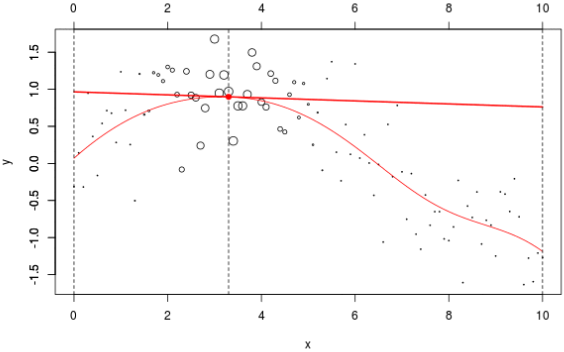
Source: A. Charpentier

LOESS: Visually



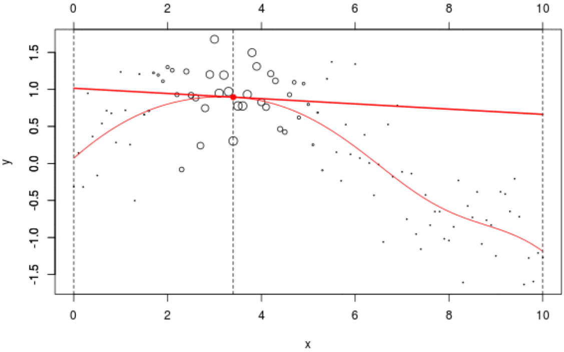
Source: A. Charpentier

LOESS: Visually



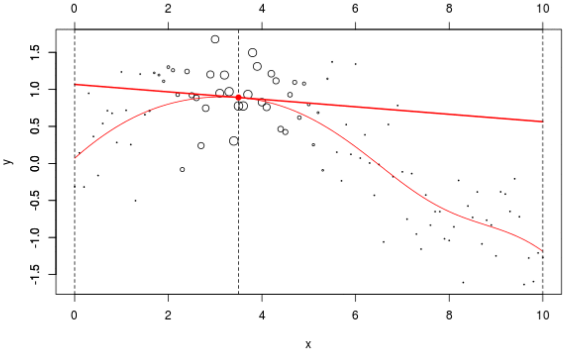
Source: A. Charpentier

LOESS: Visually



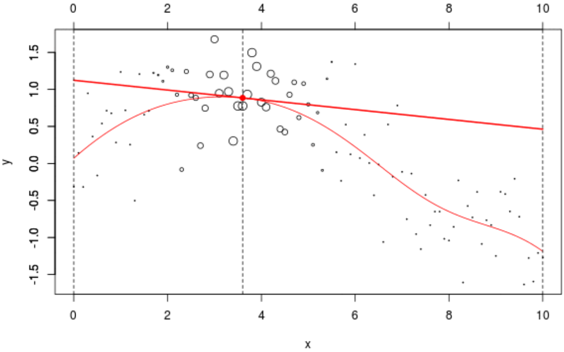
Source: A. Charpentier

LOESS: Visually



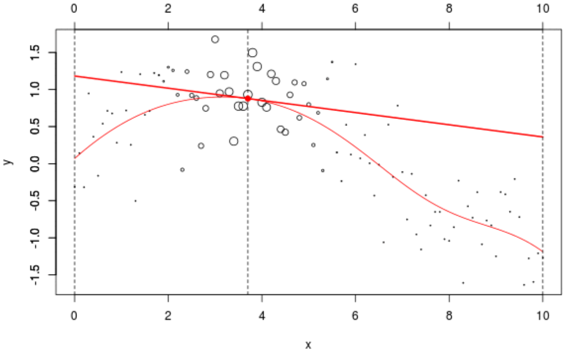
Source: A. Charpentier

LOESS: Visually



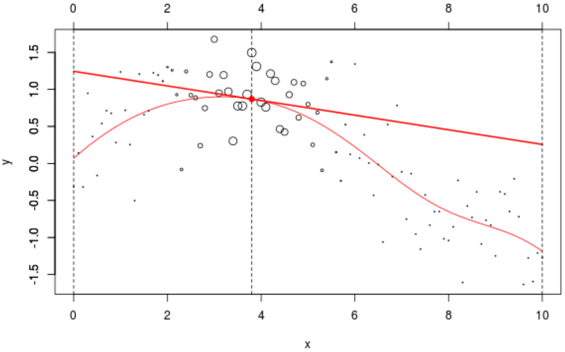
Source: A. Charpentier

LOESS: Visually



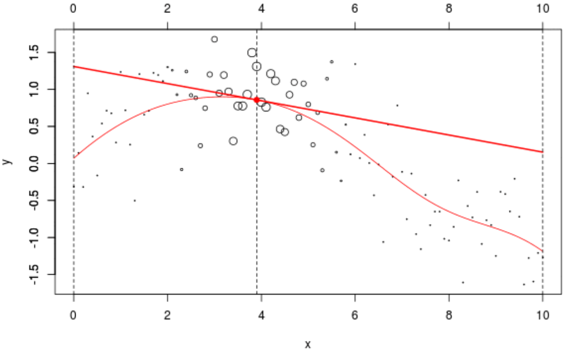
Source: A. Charpentier

LOESS: Visually



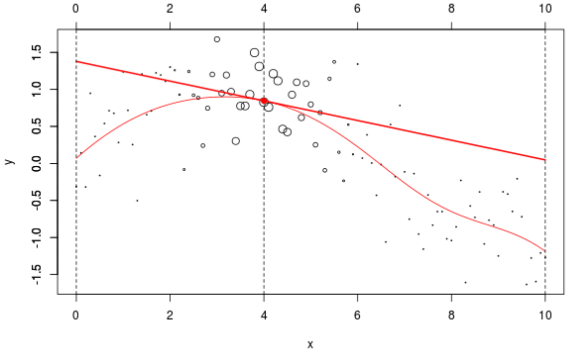
Source: A. Charpentier

LOESS: Visually



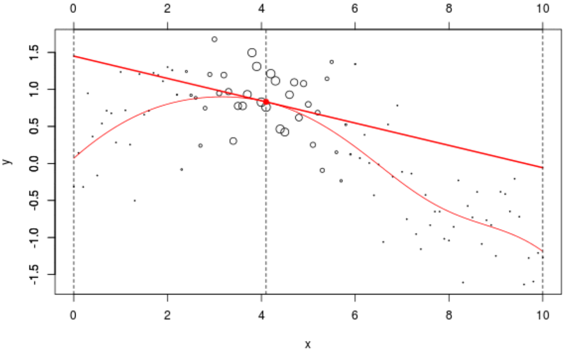
Source: A. Charpentier

LOESS: Visually



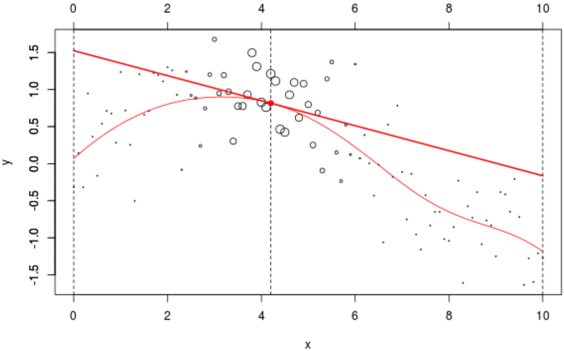
Source: A. Charpentier

LOESS: Visually



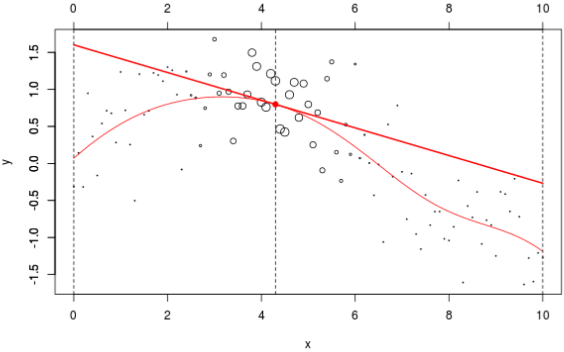
Source: A. Charpentier

LOESS: Visually



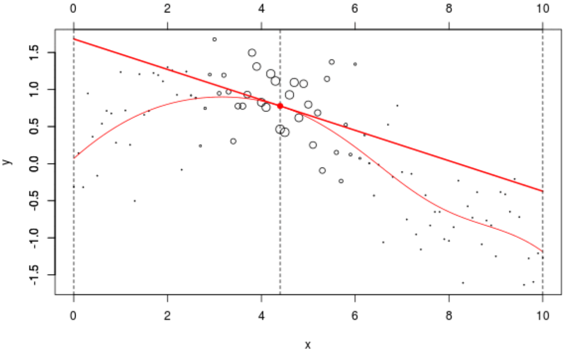
Source: A. Charpentier

LOESS: Visually



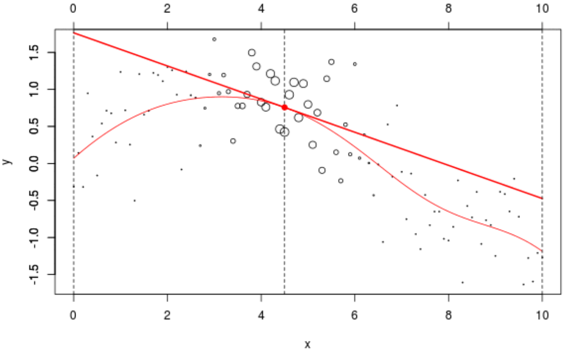
Source: A. Charpentier

LOESS: Visually



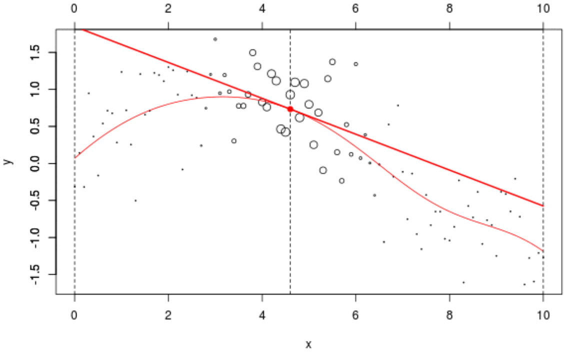
Source: A. Charpentier

LOESS: Visually



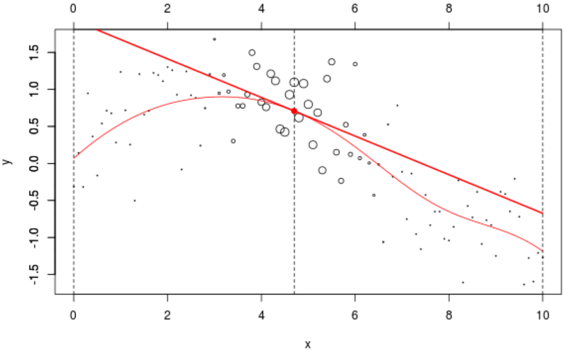
Source: A. Charpentier

LOESS: Visually



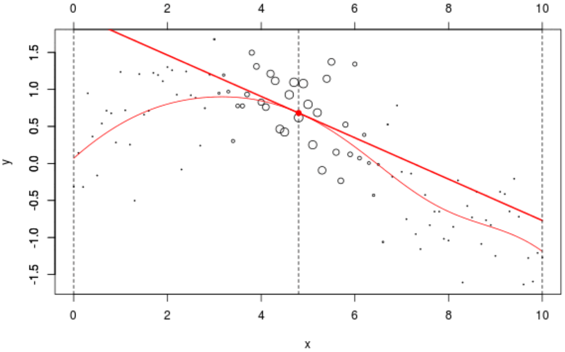
Source: A. Charpentier

LOESS: Visually



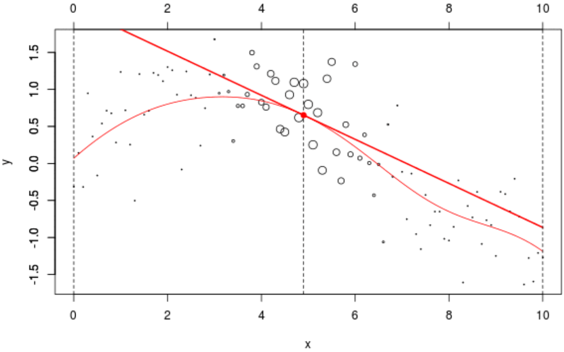
Source: A. Charpentier

LOESS: Visually



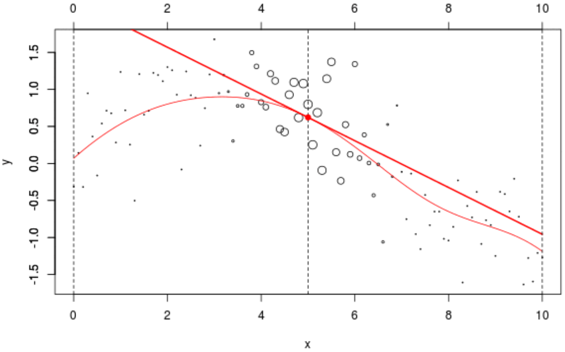
Source: A. Charpentier

LOESS: Visually



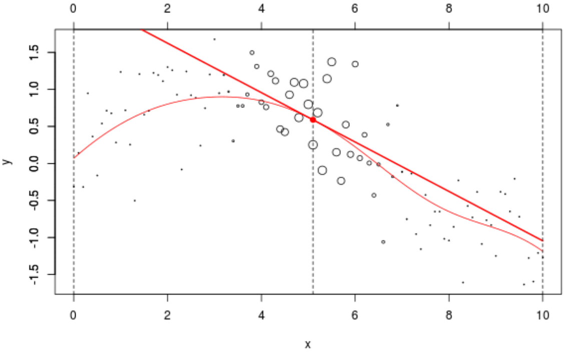
Source: A. Charpentier

LOESS: Visually



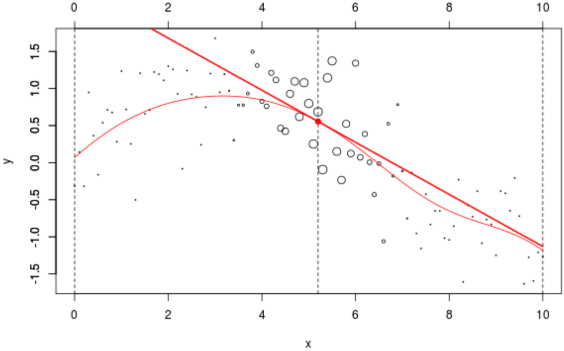
Source: A. Charpentier

LOESS: Visually



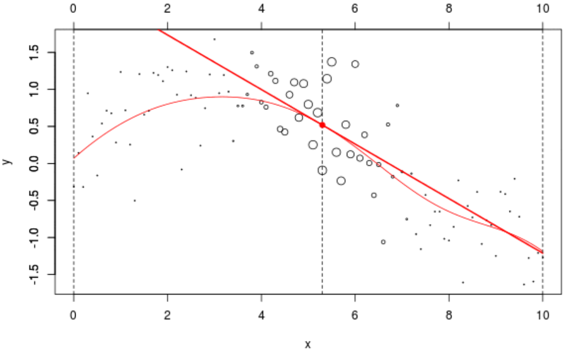
Source: A. Charpentier

LOESS: Visually



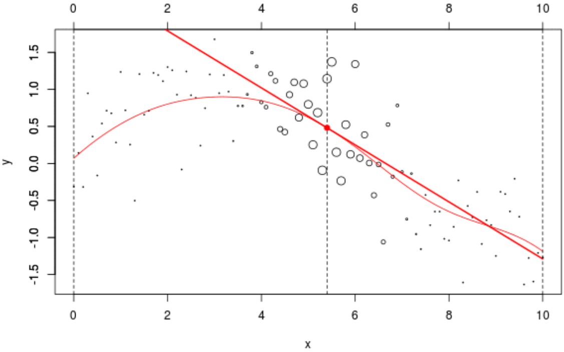
Source: A. Charpentier

LOESS: Visually



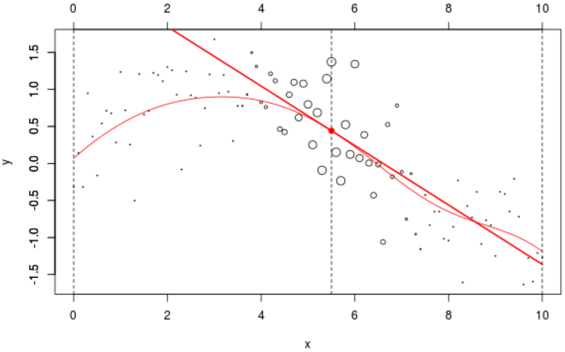
Source: A. Charpentier

LOESS: Visually



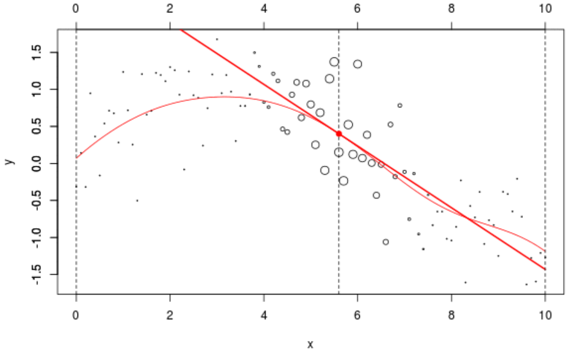
Source: A. Charpentier

LOESS: Visually



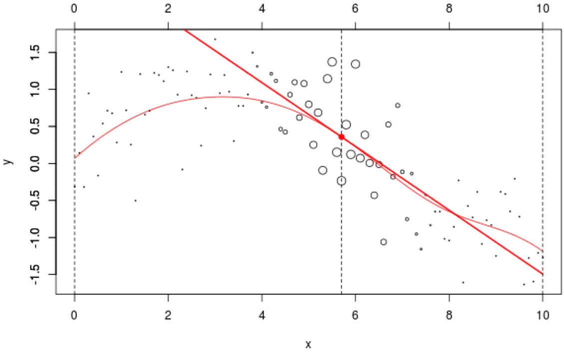
Source: A. Charpentier

LOESS: Visually



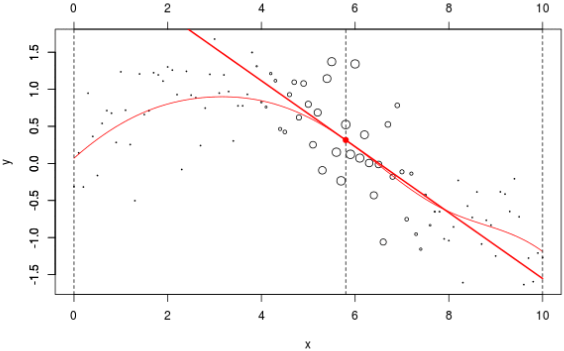
Source: A. Charpentier

LOESS: Visually



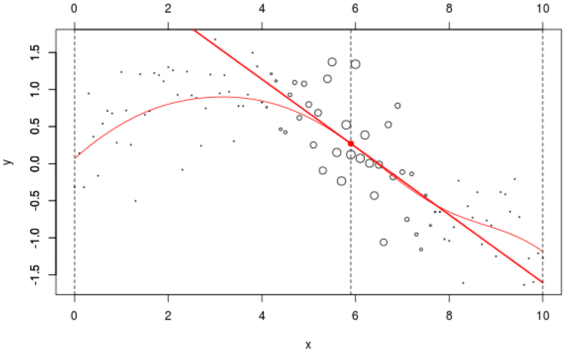
Source: A. Charpentier

LOESS: Visually



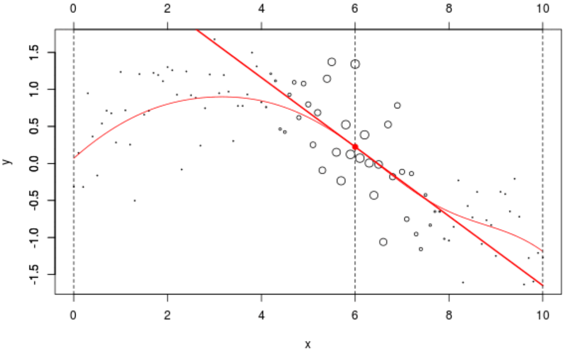
Source: A. Charpentier

LOESS: Visually



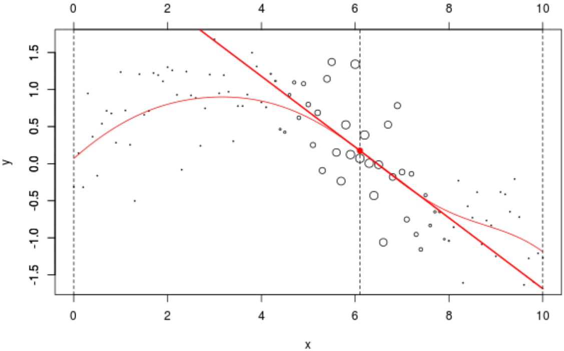
Source: A. Charpentier

LOESS: Visually



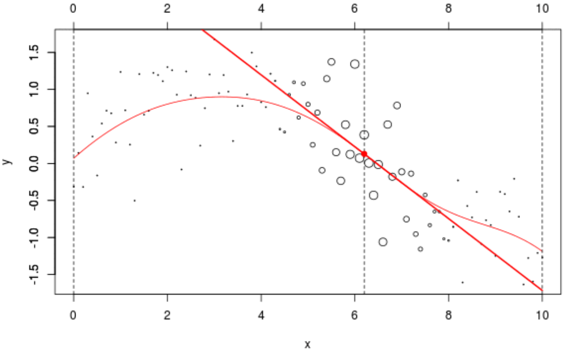
Source: A. Charpentier

LOESS: Visually



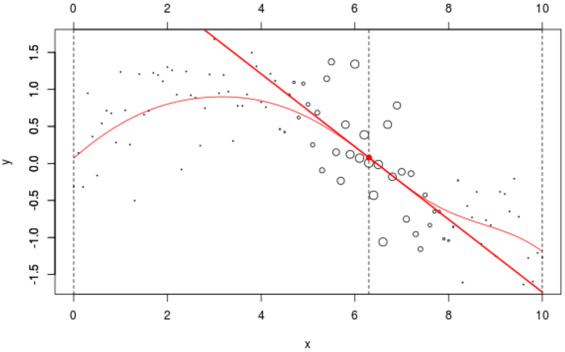
Source: A. Charpentier

LOESS: Visually



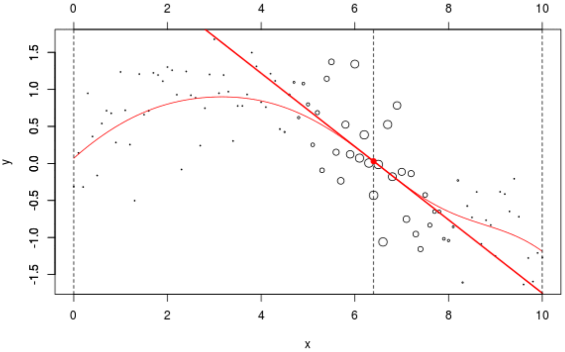
Source: A. Charpentier

LOESS: Visually



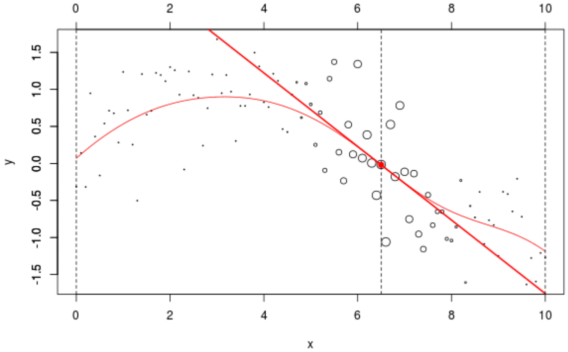
Source: A. Charpentier

LOESS: Visually



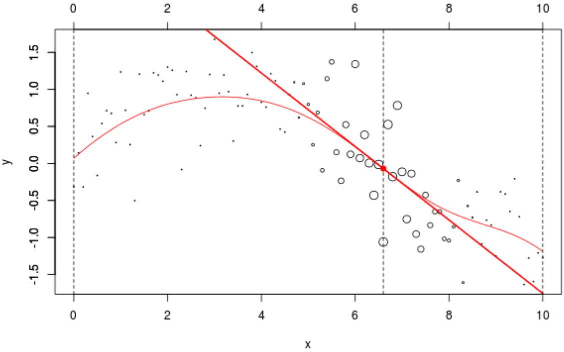
Source: A. Charpentier

LOESS: Visually



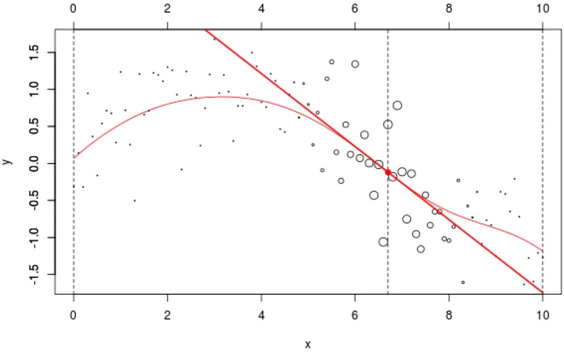
Source: A. Charpentier

LOESS: Visually



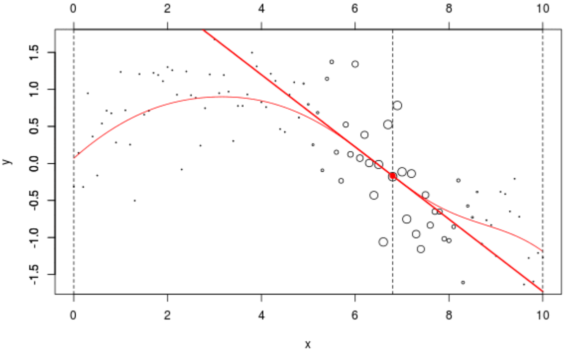
Source: A. Charpentier

LOESS: Visually



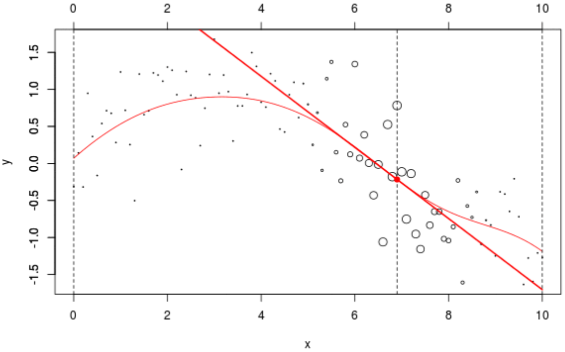
Source: A. Charpentier

LOESS: Visually



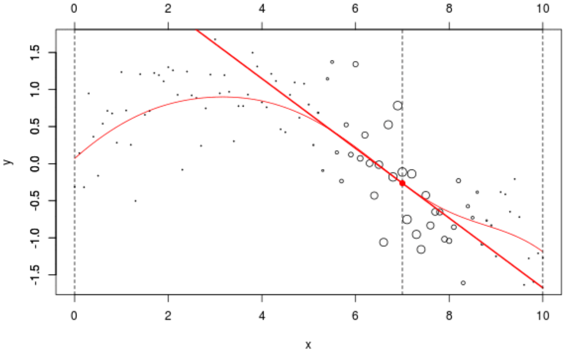
Source: A. Charpentier

LOESS: Visually



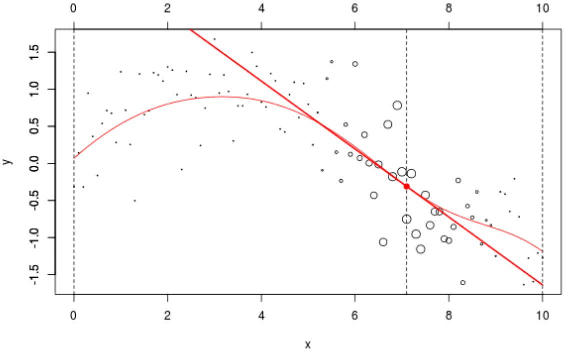
Source: A. Charpentier

LOESS: Visually



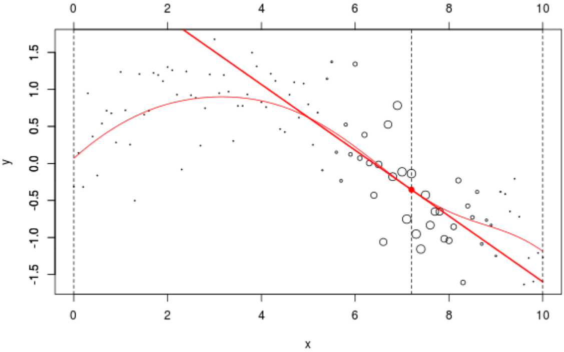
Source: A. Charpentier

LOESS: Visually



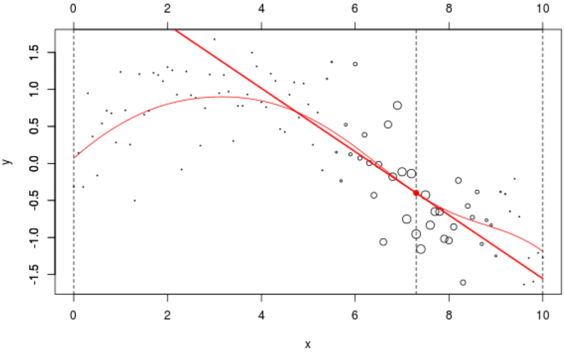
Source: A. Charpentier

LOESS: Visually



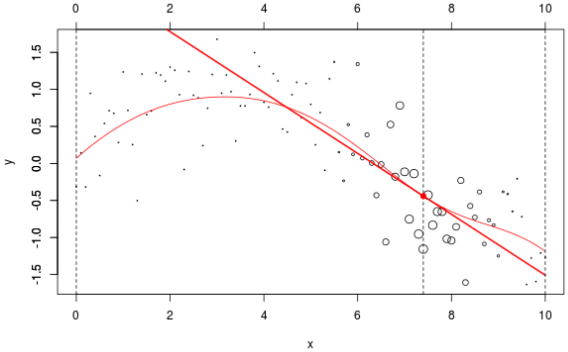
Source: A. Charpentier

LOESS: Visually



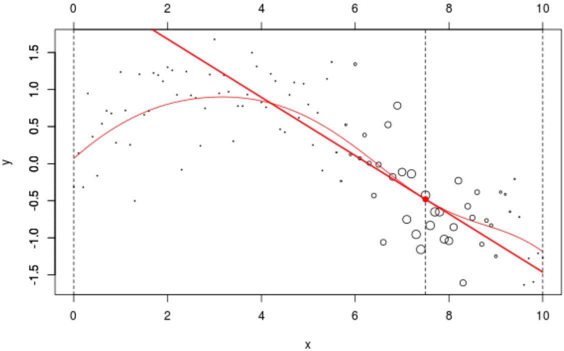
Source: A. Charpentier

LOESS: Visually



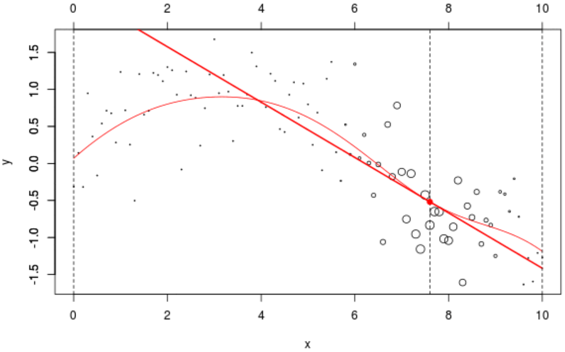
Source: A. Charpentier

LOESS: Visually



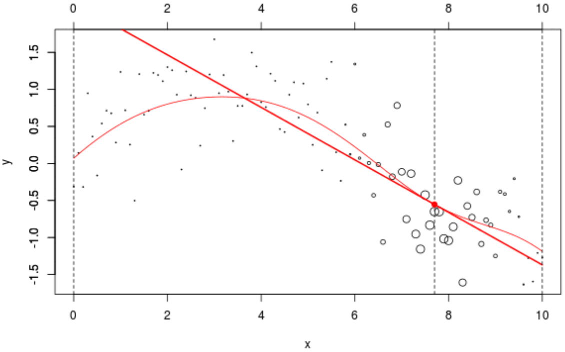
Source: A. Charpentier

LOESS: Visually



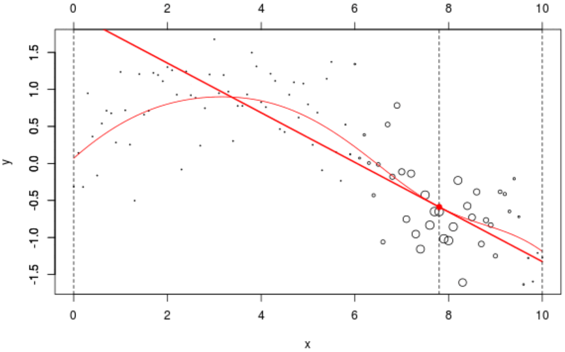
Source: A. Charpentier

LOESS: Visually



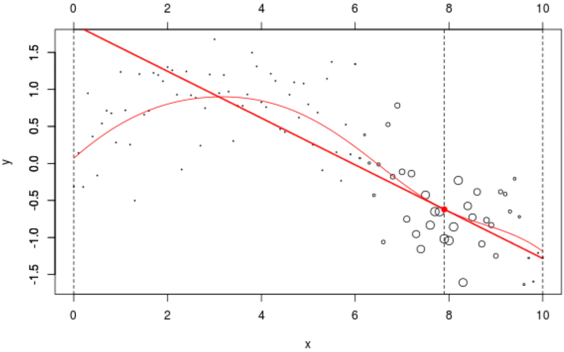
Source: A. Charpentier

LOESS: Visually



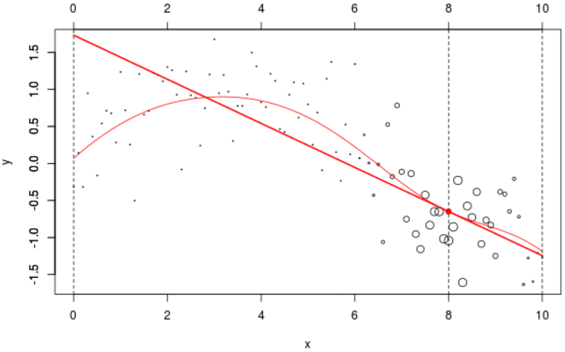
Source: A. Charpentier

LOESS: Visually



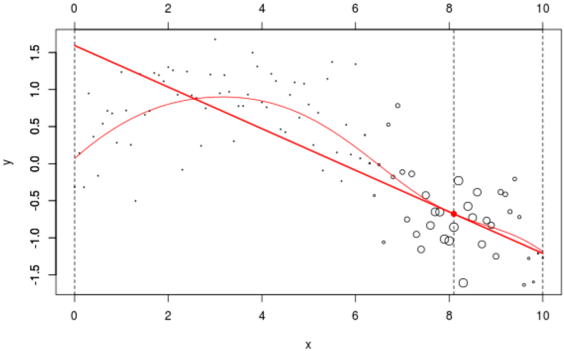
Source: A. Charpentier

LOESS: Visually



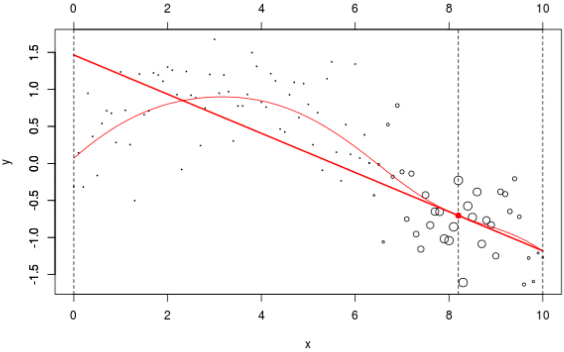
Source: A. Charpentier

LOESS: Visually



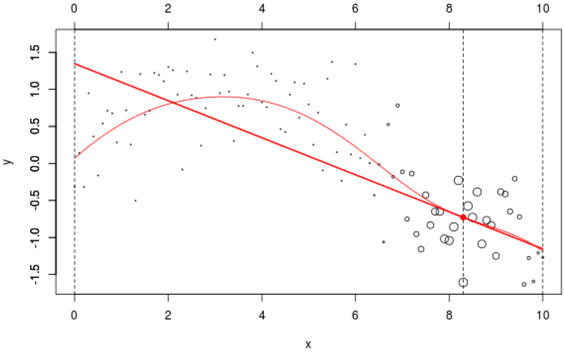
Source: A. Charpentier

LOESS: Visually



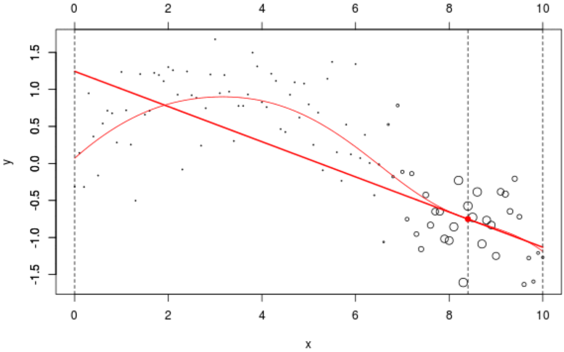
Source: A. Charpentier

LOESS: Visually



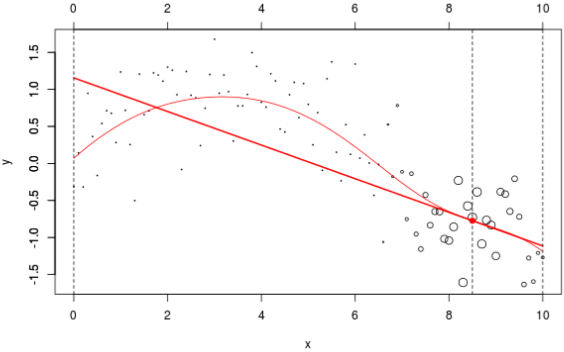
Source: A. Charpentier

LOESS: Visually



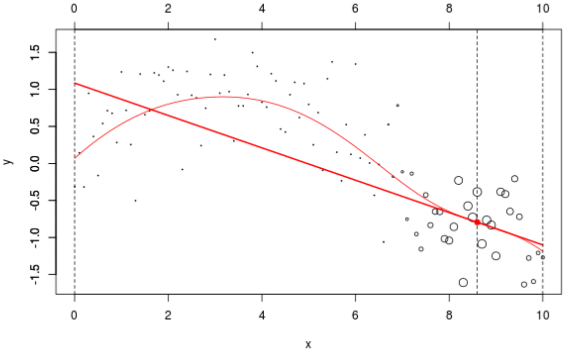
Source: A. Charpentier

LOESS: Visually



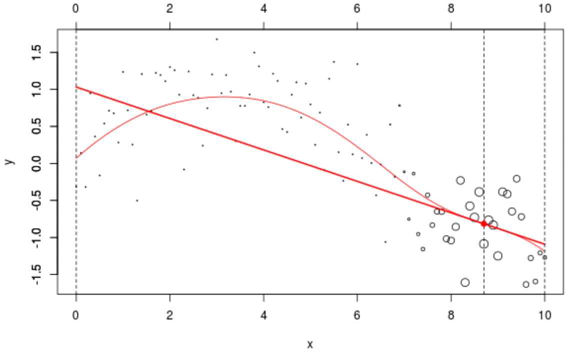
Source: A. Charpentier

LOESS: Visually



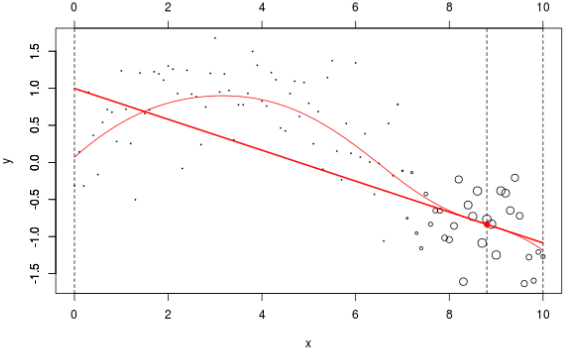
Source: A. Charpentier

LOESS: Visually



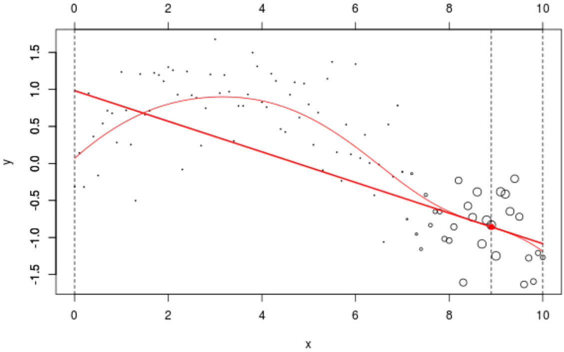
Source: A. Charpentier

LOESS: Visually



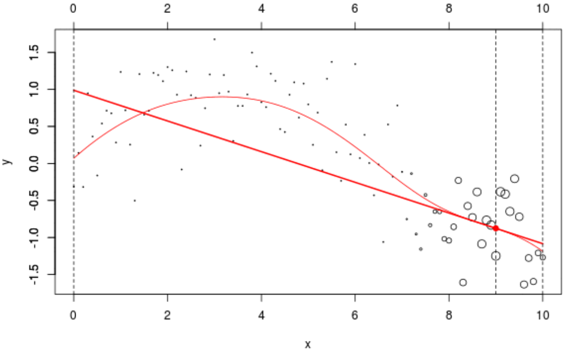
Source: A. Charpentier

LOESS: Visually



Source: A. Charpentier

LOESS: Visually



Source: A. Charpentier

LOESS: Final Considerations

- Parametric models (such as polynomial regression) are normally guided by theory
  - nonparametric models are data-driven
  - good exploratory tools
  - but perhaps they should not be used to test hypotheses

LOESS: Final Considerations

- Parametric models (such as polynomial regression) are normally guided by theory
  - nonparametric models are data-driven
  - good exploratory tools
  - but perhaps they should not be used to test hypotheses
- It is recommended not to use narrow bandwiths
  - if too wide of a bandwith we might miss of the non-linearity
  - the narrower the bandwidth the more noise will be depicted
  - the narrower the bandwidth the higher the SEs
  - there is a trade-off between precision and accuracy

LOESS: Final Considerations

- Parametric models (such as polynomial regression) are normally guided by theory
  - nonparametric models are data-driven
  - good exploratory tools
  - but perhaps they should not be used to test hypotheses
- It is recommended not to use narrow bandwiths
  - if too wide of a bandwith we might miss of the non-linearity
  - the narrower the bandwidth the more noise will be depicted
  - the narrower the bandwidth the higher the SEs
  - there is a trade-off between precision and accuracy
- Its applicability is limited
  - both the outcome and explanatory variable must be continuous
  - can be used for more than one explanatory variable, but it cannot control for other variables
  - to explore non-linear effects non-parametrically while controlling for other variables use generalised additive models (GAM)
  - to explore changes in the effect of discrete explanatory variables across the range of Y we can use quantile regression

Recap

- We can include non-linear effects in linear models easily using polynomial regressions
- If we do not want to impose pre-defined (parametric) functions we can also use local regression
  - we have learnt how LOESS works, probably the most common form of non-parametric regression to deal with non-linear effects
  - similar methods could be explored in more complex settings (GAM and quantile regression)
- Recommended readings
  - on polynomial regression you can read Hanck et al. (2019) Chapter 8 'Nonlinear Regression Functions'
  - on non-parametric models you can read Mahmoud (2019) 'Parametric vs Semi and NonParametric Regression Models'
  - on local regression Irizarry 'Chapter 3. Local Regression'