Statistics 5350/7110
Forecasting

Lecture 9
Scatterplot Smoothing

Professor Robert Stine

Admin Issues

- Questions
  - TA office hours
- Assignments
  - Questions regarding Assignment 2
- Quick review
  - Detrending
  - Differencing vs regression
  - Trade-offs: Possible costs vs possible benefits

	Difference	**Fit Regression**
Stochastic trend **Random walk**	✔	Still nonstationary Inflated precision
Deterministic **Linear trend**	Loss of efficiency	✔

Today's Topics

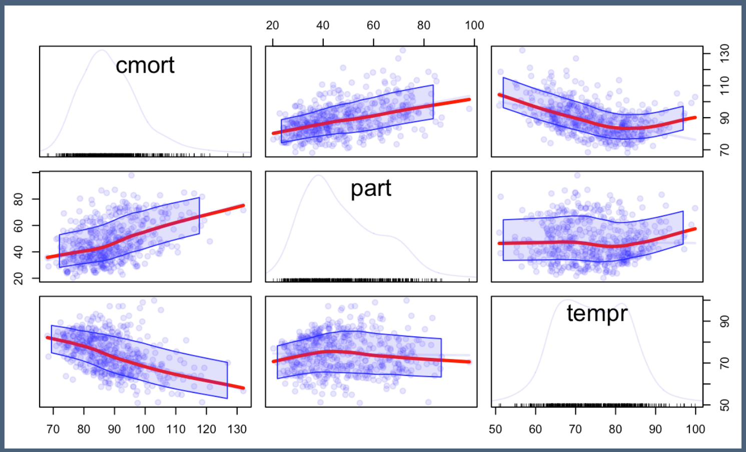
Text, §3.3

- Scatterplot smoothing
  - Common graphical procedure typically used with cross-sectional data
  - Dependence leads to special issues with time series (more next time)
- Motivation
  - Regression fit is a weighted sum of the response, but the weights might surprise you
  - Local averages
- Methods
  - Bias-variance trade-off
  - Local averaging with a kernel
  - Smoothing splines
  - Loess

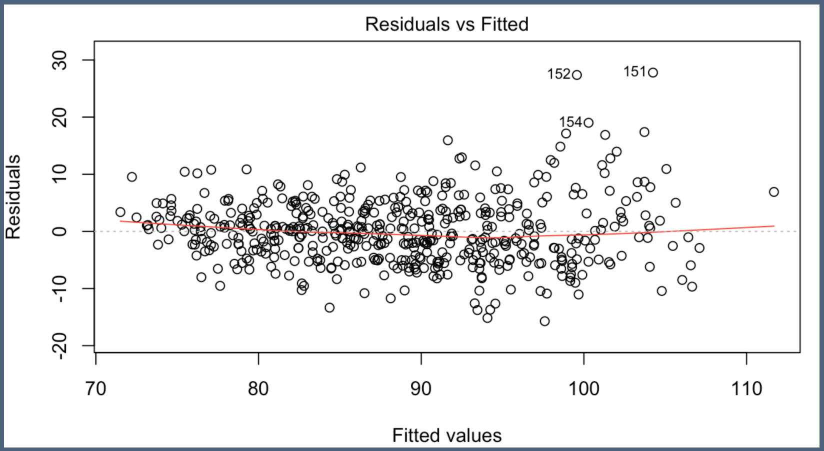
Examples

- Visual benefit
  - Distinguish imagined pattern from something real
  - Frequent in R diagnostic plots
- Goal here
  - Understand methods well-enough to modify defaults

Scatterplot Matrix



Calibration Plot



Smoothing

- Regression is smoothing
  - Fit in simple regression illustrates general expression: Fit is a weighted sum of response

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i = \sum_{j=1}^{n} h_{ij} Y_j, \quad h_{ij} = \frac{1}{n} + \frac{(x_i - \overline{X})(x_j - \overline{X})}{SS_X}$$

origin of name "hat values" for R function

leverage = h_{ii}

  - Values of response with most weight are not necessarily those close to x_i
  - Weights chosen to reduce the mean squared error (MSE) of the fit
  - Example: Boundary points have the most effect when estimating a linear trend
- Scatterplot smoothing
  - Smoothed value is a weighted average of "nearby" observations

$$\hat{Y} = \frac{\sum w_i Y_i}{\sum w_i}$$

  - What does it mean to be nearby?
  - A moving average makes sense or a sequence of equally spaced values, ... But how many to average?

Dangers of cross-validation with time series illustrated in the Red file.

Bias-Variance Trade-off

- Working model
  - View data as independent random deviations around smooth function

$$y_i = f(x_i) + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma_\epsilon^2)$$

  - Squared error of an estimator

$$E\left(y_0 - \hat{f}(x_0)\right)^2 = \sigma_\epsilon^2 + \underbrace{\left(f(x_0) - E\,\hat{f}(x_0)\right)^2}_{\text{squared bias}} + \underbrace{\text{Var}\,\hat{f}(x_0)}_{\text{variance}}$$

- Trade-off
  - Smooth, nearly constant estimator has large bias, but small variance
  - Rough, highly responsive estimator has large variance, but small bias
- Making a choice?
  - Prior knowledge of the unknown function $f(x)$
  - Data-driven methods such as cross-validation not well-suited to time series (not independent errors) e.g. positively autocorrelated leads to under-smoothing (Example simulation in Lecture_9.Rmd)

Moving Average

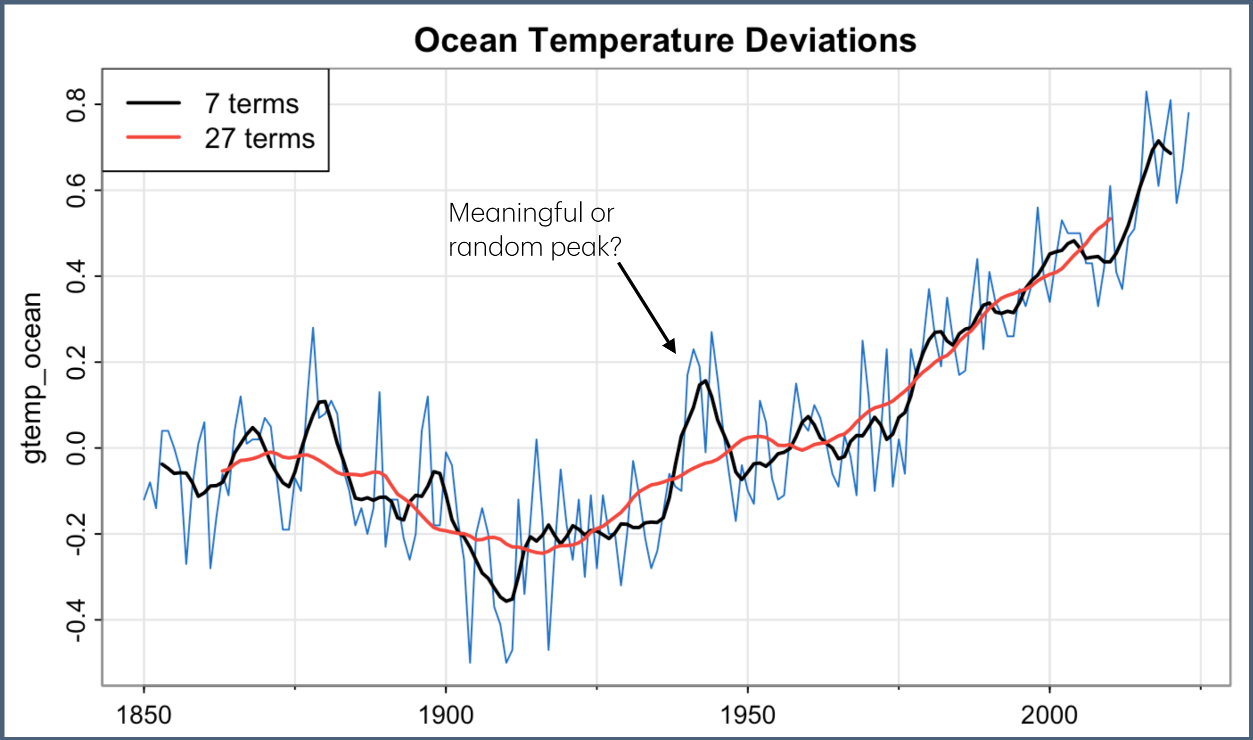
Example 3.16

- Average adjacent values
  - Simple to define when data are equally spaced

$$\hat{Y}_i = \sum_{j=-k}^{k} a_j X_{t-j}$$

with $a_j = 1/(2k+1)$

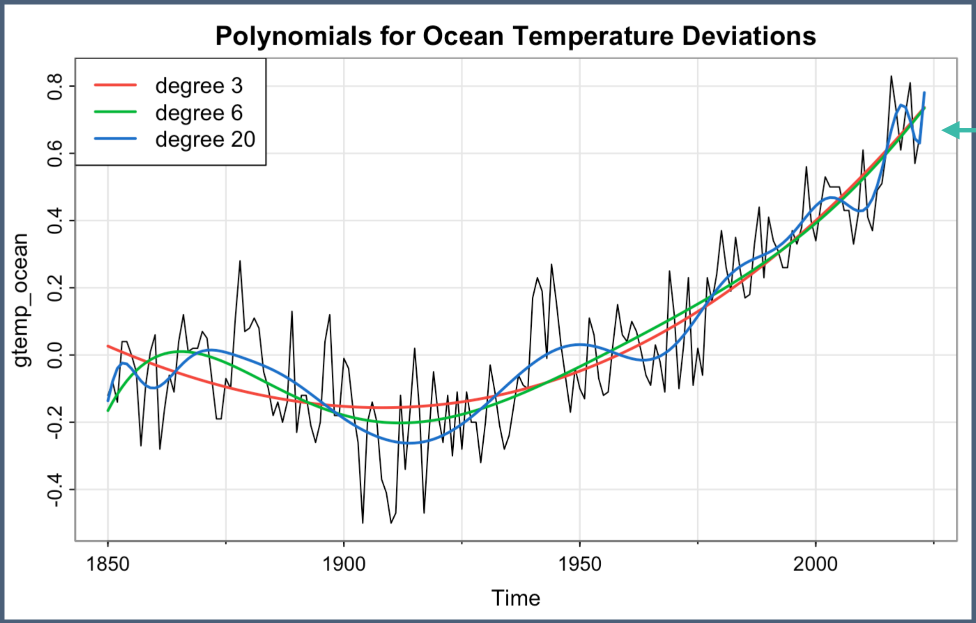
- Issues
  - Number to average
  - Choice of weights
  - Behavior at boundaries
- Conceptual issue
  - What are we estimating?
  - What's $f(x)$ in this context?



Code in this example uses slightly different weights

Polynomials

- Polynomials
  - With high-enough degree can approximate any smooth curve
  - Issues at the boundaries (Gibbs effect, must diverge to ±∞)
- Degree and smoothness
  - Higher degree allows more curvature, greater "flexibility"
  - Global smoothness
  - Choice of degree?



Boundary "wiggle"

Fit with orthogonal polynomials in LS regression

Smoothing Splines

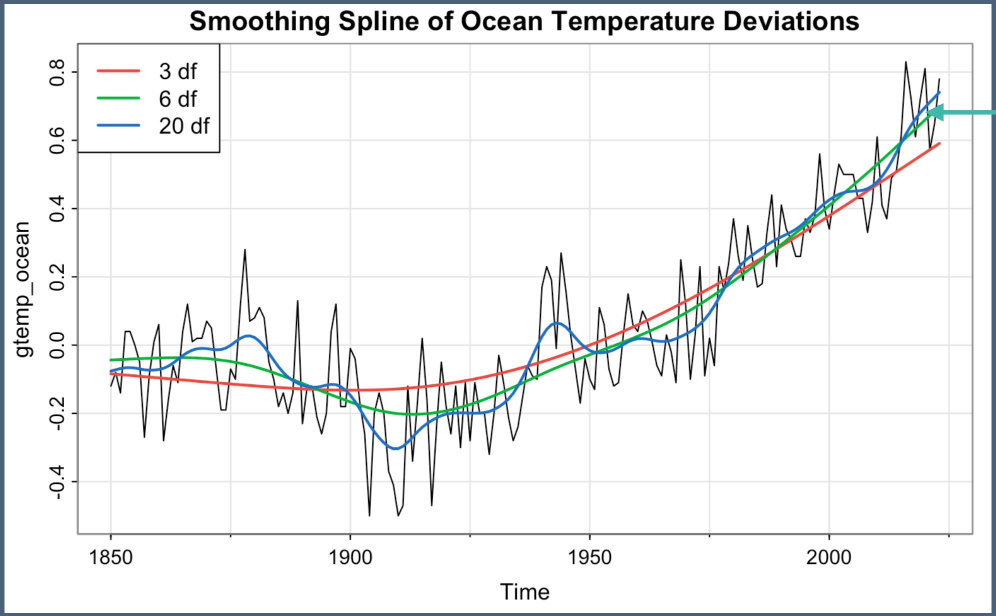
- Polynomials
  - Closer approximation requires high-degree of curvature
  - Issues at the boundaries (Gibbs effect)
- Cubic spline
  - Introduces local fit, analogous to moving average
  - Interpolates data with piecewise cubic polynomial
  - Continuous first and second derivatives
- Smoothing spline
  - Doesn't interpolate
  - Trades fidelity to data with a smoothing penalty

$$\hat{f} = \arg\min_{f} \sum_{i} (y_i - f(x_i))^2 + \lambda \int f''(x)^2 dx$$

  - Choice of λ equivalent to an equivalent degrees of freedom

Smoothing Spline

- Improved behavior over polynomials
  - Allows local change without high degree
  - Improved behavior near boundaries
- Choices
  - Have to choose degree of smoothness
  - Global smoothness



Eliminates boundary "wiggle"

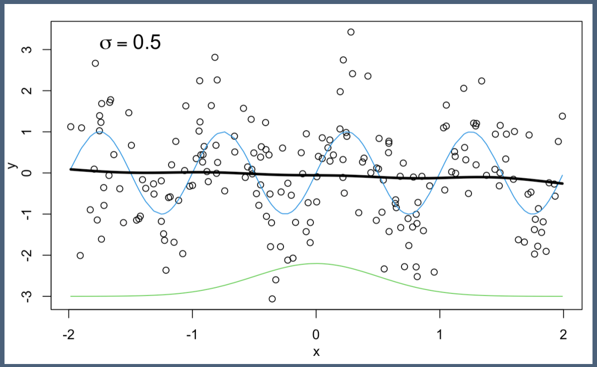
Kernel Smoothing

Example 3.17

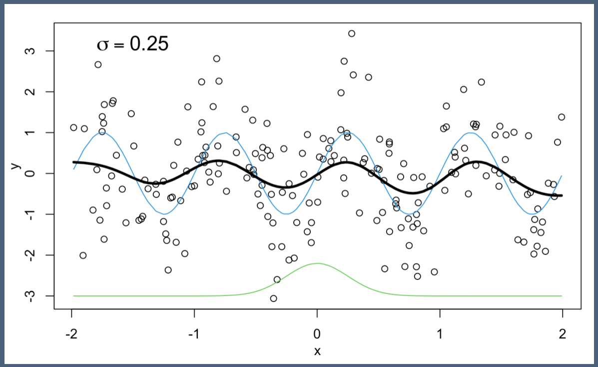
- Generalize moving average
  - What happens if the Y values are not equally spaced?
- Kernel smoothing
  - Use a function (the kernel) to set the weights. Handles the boundary problem.

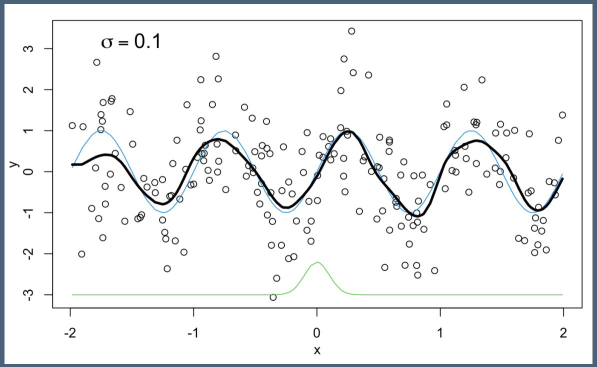
$$\hat{Y}_i = \frac{\sum_{j=1}^{n} K(x_i, x_j)\, Y_j}{\sum K(x_i, x_j)}$$

  - Example: Let K be the normal probability density function, using the SD to control the smoothness
  - Question: How to choose the bandwidth, the bias/variance trade-off



Low variance, but high bias





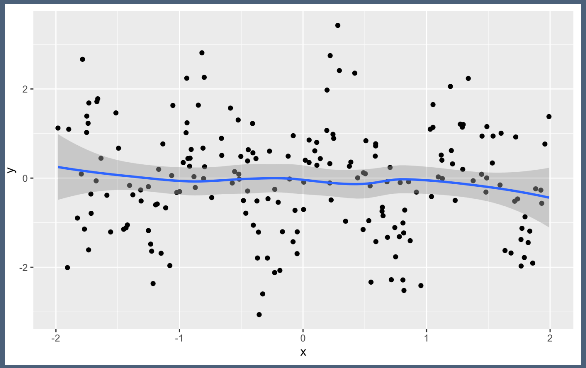
Higher variance, but less bias

Lowess and Loess

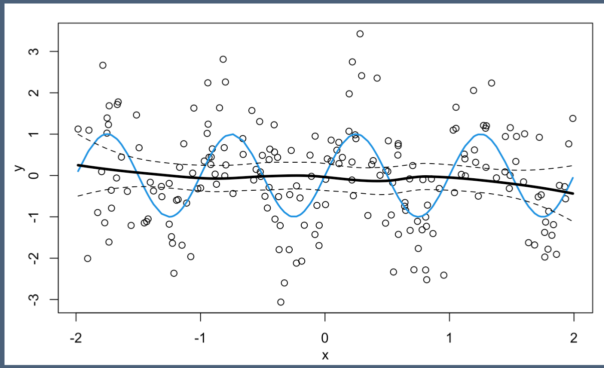
Example 3.18

- Modern scatterplot smoothers
  - Locally weighted polynomial regression (as in a kernel) estimates smooth values
  - Theory and implementation provide point-wise confidence intervals
  - Loess is a revision of predecessor lowess with more theory
  - Modern implementation found in many software packages (e.g. ggplot)
- Examples
  - Issue remains: How much smoothing is desirable. Default span is often too smooth.

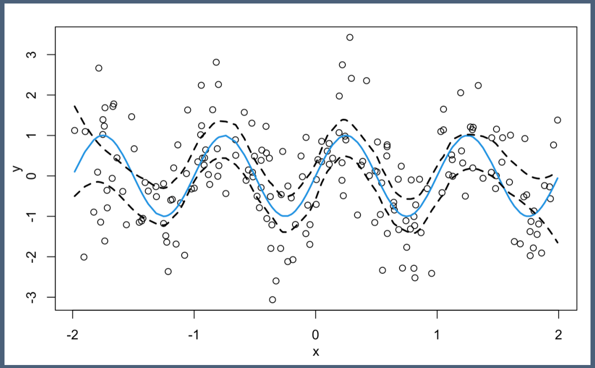
Default weights computed from Gaussian density



ggplot version, default span



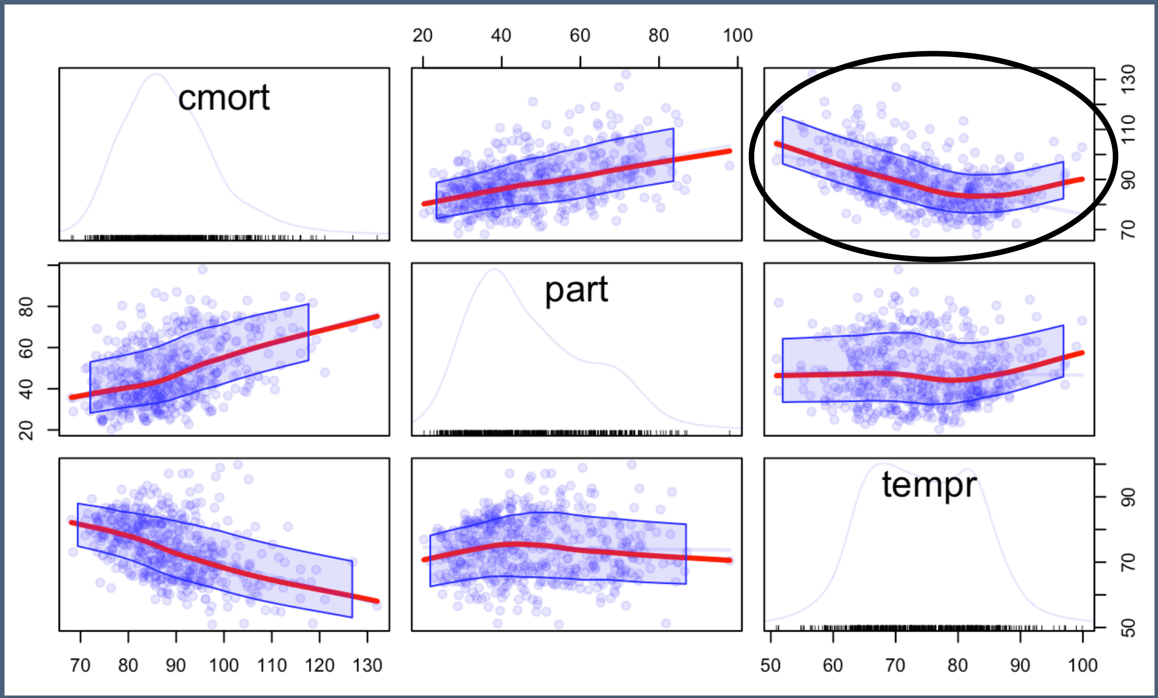
R version, default span



Smaller span

Mortality Data Revisited

- Scatterplot matrix
  - Data are mortality in LA county with particulates and temperature
  - Scatterplot matrix includes smooth curves to check for nonlinearity

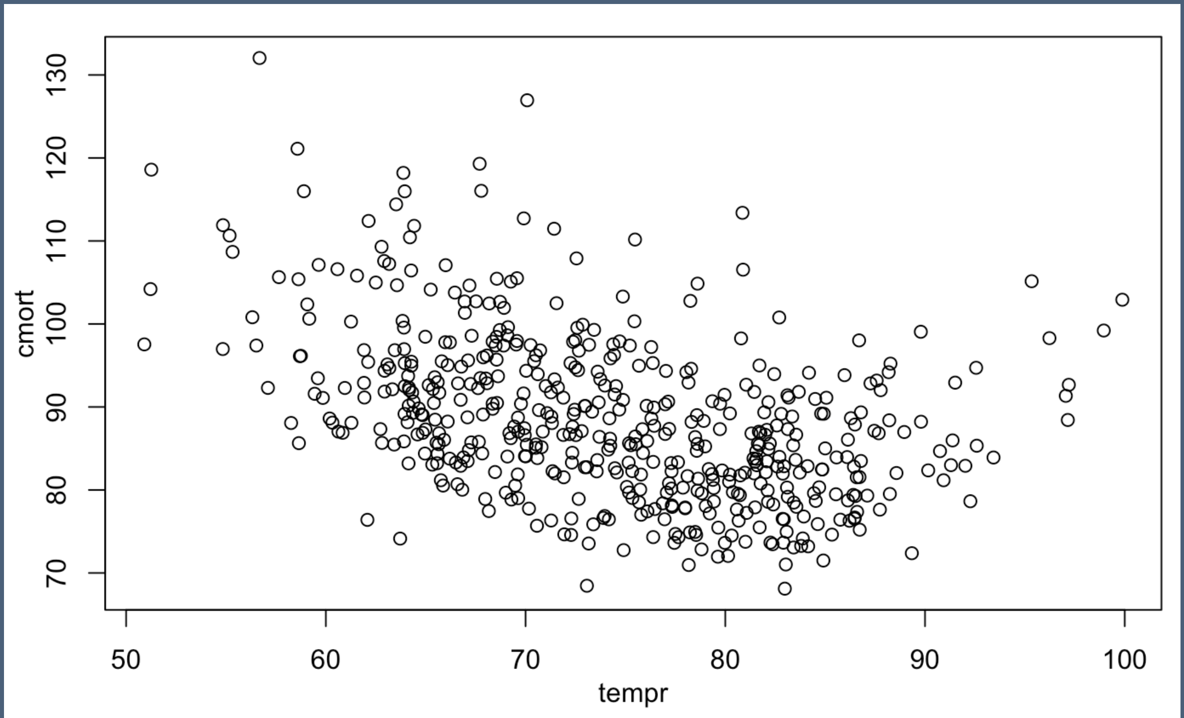


Transparent colors are helpful when data overlap

Mortality Data Revisited

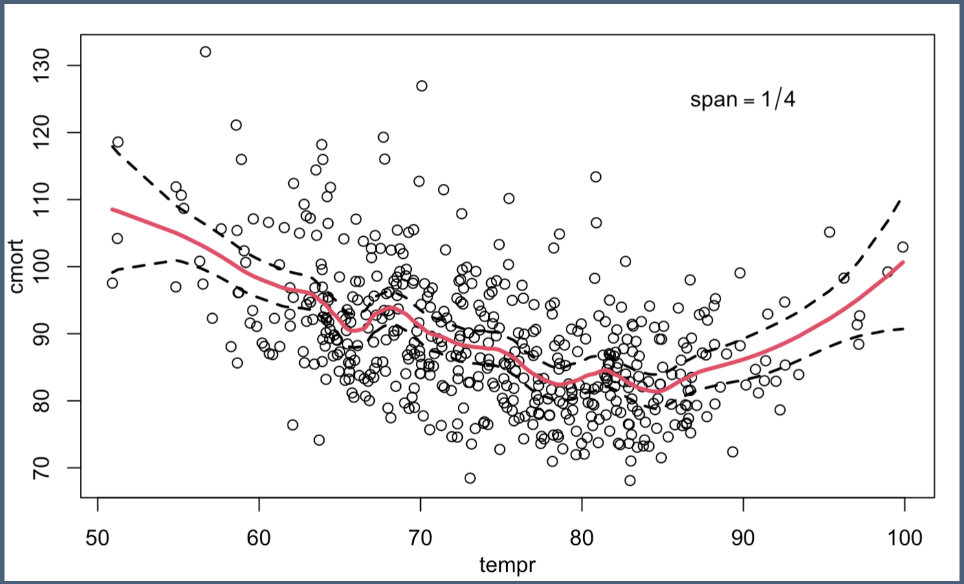
Example 3.19

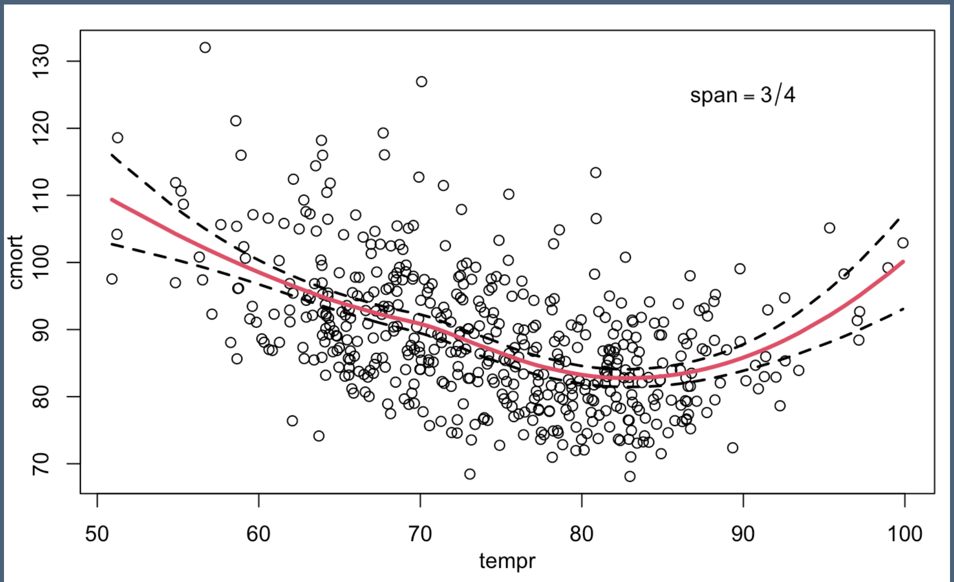
- Mortality appears higher away from temperatures near 75-80 degrees
  - Note that this plot does not show the time dimension of these series



Mortality Data Revisited

- Effects of span size
  - Smaller span captures local fluctuations, but more variable (wider intervals since less data)
  - Wider span often misses local features, but offers narrower confidence intervals.
  - Recommend: Explore a variety of spans

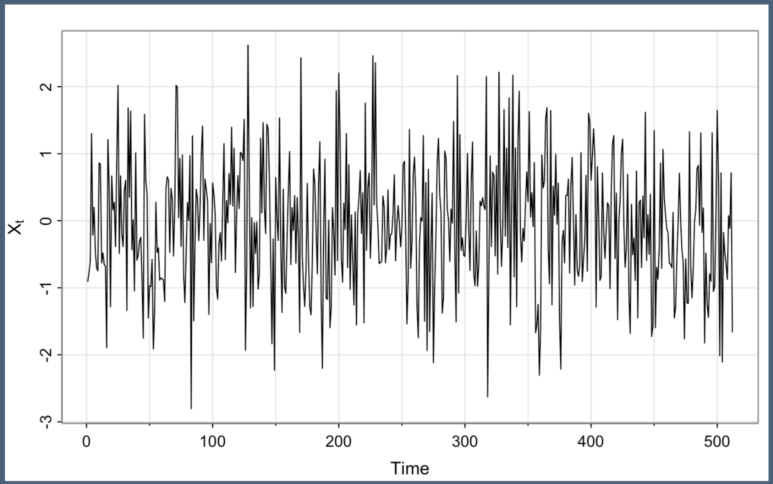


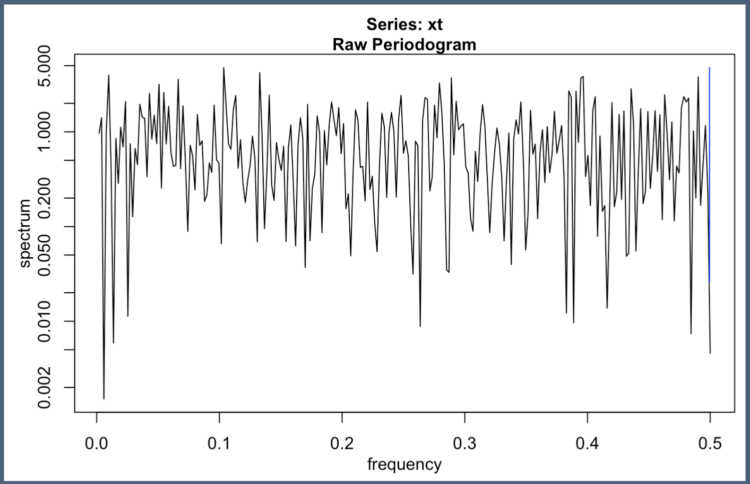


Note: these plots do not show the time dimension of these series

Smoothing Time Series

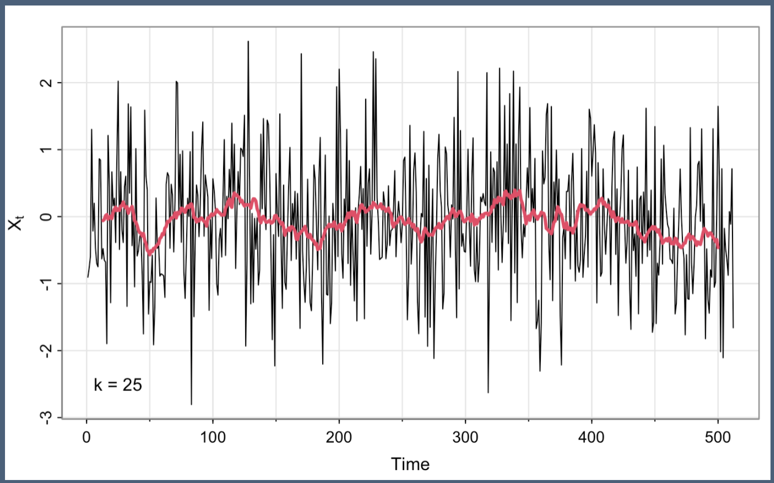
- A note of caution
  - Smoothing a time series has subtle effects
- Example
  - Start with a random, iid normal series, the canonical "white noise"
- Periodogram
  - Variance explained by sinusoids of the form cos(2 π f_k t) on grid of equally-spaced frequencies
  - Equal energy at every frequency … flat … the color spectrum of white light

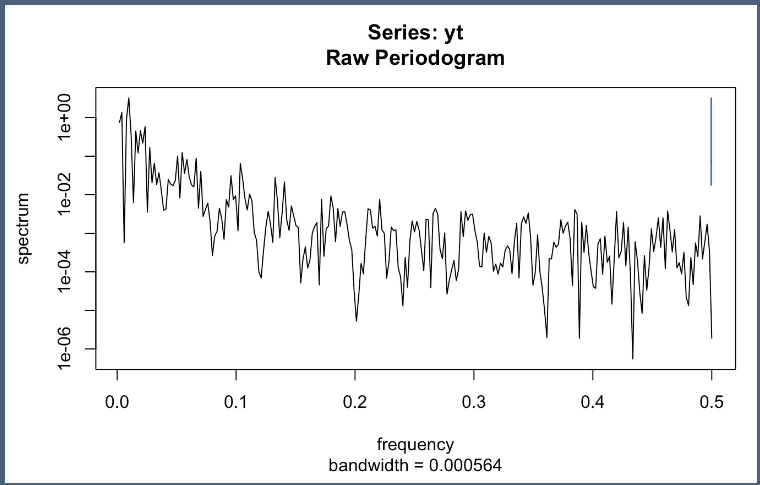




Smoothing Time Series

- What happens when we smooth the time series?
  - We change how variability in the data spreads out over frequencies
- Example
  - Smoother: Equally weighted moving average
  - Result:
    More variation at lower frequencies: the series has less high-frequency variation: it's smoother!





What's next?

- Decomposing a time series
  - Time series == Trend + Seasonal + Noise

- Common to find "seasonally adjusted" economic data