

NCSU ST 503 Discussion 14

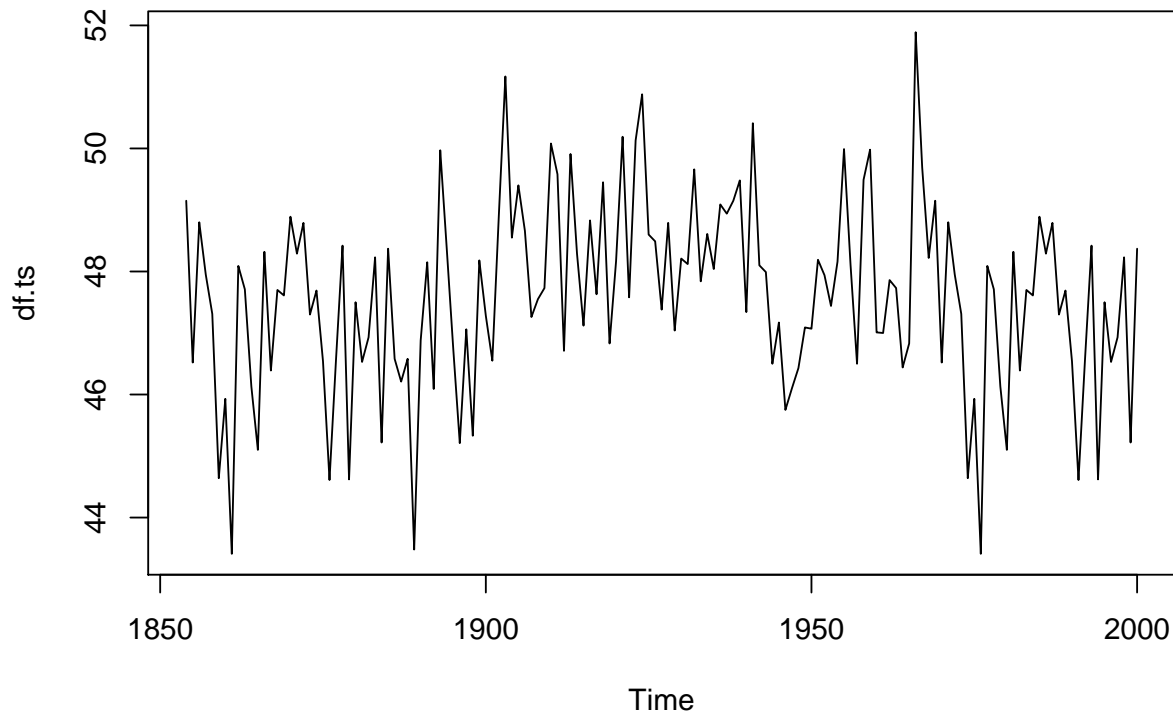
Problem 14.5 Faraway, Julian J. Extending the Linear Model with R:
Generalized Linear, Mixed Effects and Nonparametric Regression Models
CRC Press.

Bruce Campbell

14.5 temp data analysis

The aatemp data comes from the U.S. Historical Climatology network. They are the annual mean temperatures (in degrees Fahrenheit) in Ann Arbor, Michigan, going back about 150 years.

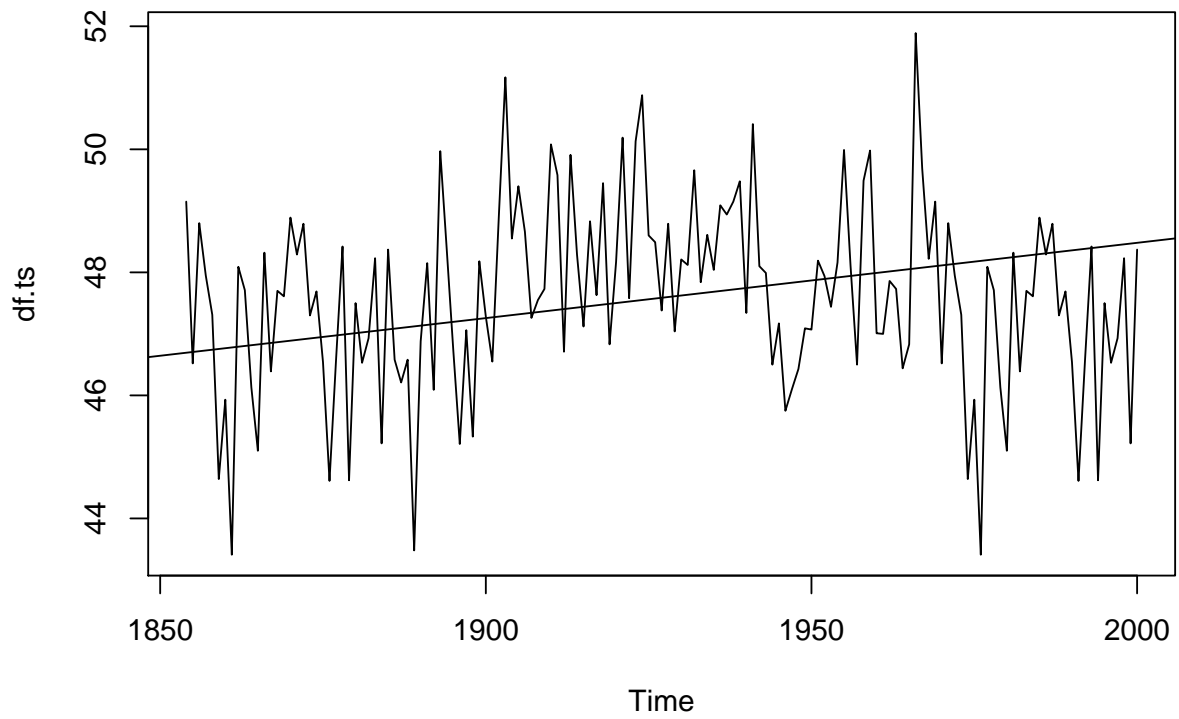
(a) Plot the temperature as a function of time and comment on the underlying trend.



It appears that the overall trend is rising and then falling. There are some higher frequency fluctuations as well.

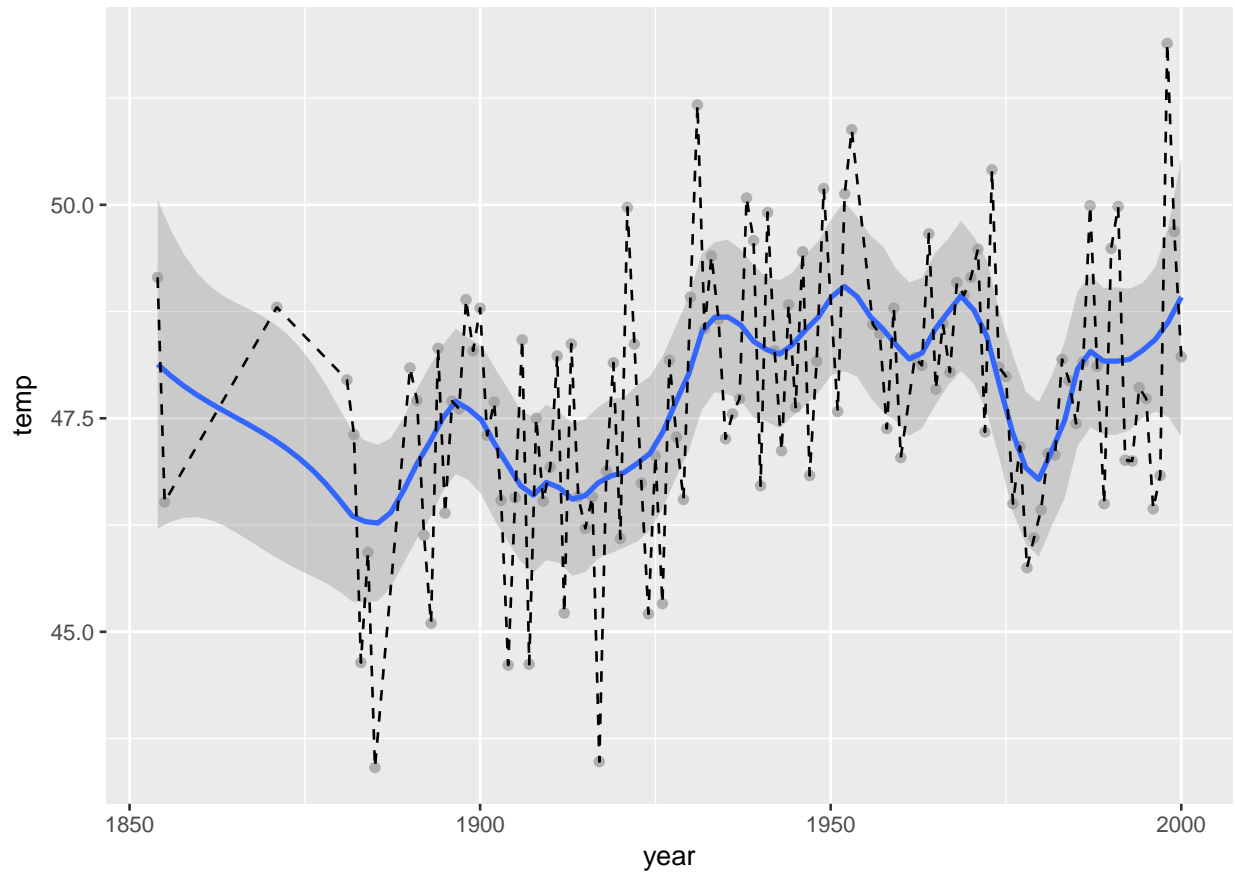
(b) Fit a least squares line to the data and test whether the slope of the line is different from zero. What is the main drawback of this modeling approach?

```
##
## Call:
## lm(formula = temp ~ year, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.9843 -0.9113 -0.0820  0.9946  3.5343
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 24.005510   7.310781   3.284  0.00136 **
## year         0.012237   0.003768   3.247  0.00153 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.466 on 113 degrees of freedom
## Multiple R-squared:  0.08536,    Adjusted R-squared:  0.07727
## F-statistic: 10.55 on 1 and 113 DF,  p-value: 0.001533
```



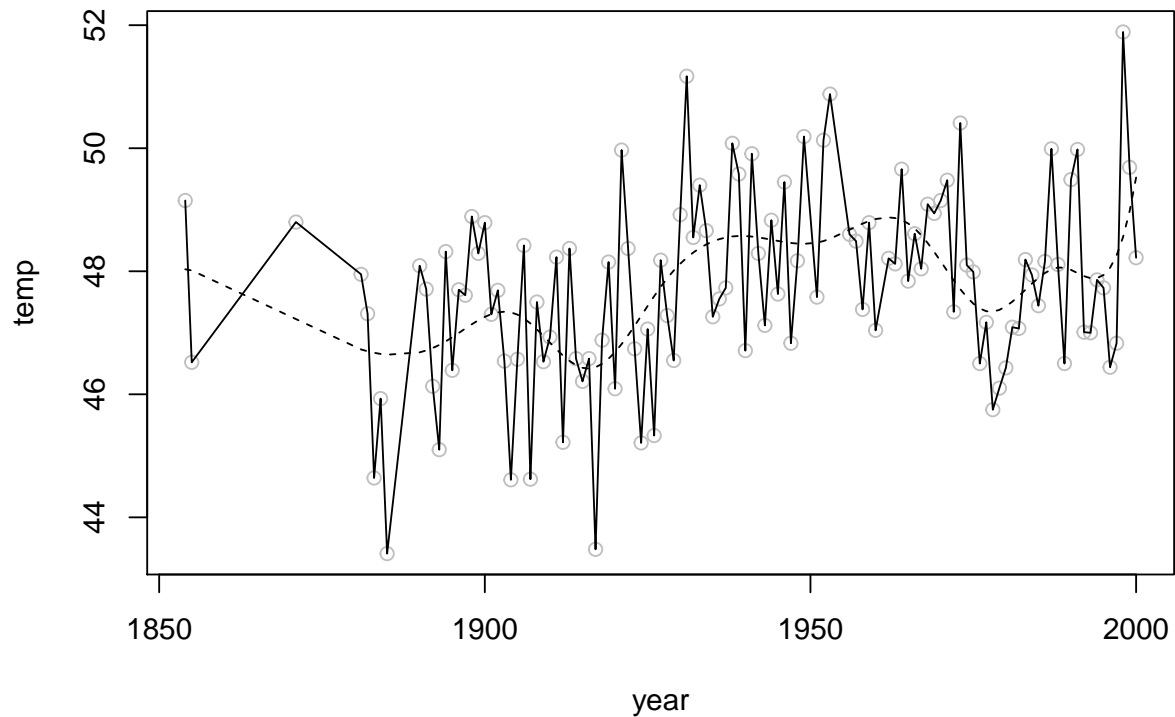
We see that the slope is significant, but none of the nonlinear components are captured in this model.

(c) Fit a Lowess curve to the data using the default amount of smoothing. Display the fit along with a 95% confidence band. What does this say about the underlying trend in the relationship?



This fit captures more of the nonlinearities in the data. The confidence bands are pointwise, the distance between them is determined by the variability of the points in the neighborhood about the band.

(d) Fit a regression spline basis to the data with 12 knots. Display the fit on the data.



(e) Compare this model to the linear fit using an F-test. Which model is preferred? What more needs to be explored with spline fit before drawing conclusions?

```
## Analysis of Variance Table
##
## Model 1: temp ~ year
## Model 2: temp ~ bs(year, 12)
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1     113 242.94
## 2     102 199.67 11    43.275 2.0097 0.03476 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```