Time Series HW 3

Mark Greenwood

## HW 3 Key

Graded out of 54.5 points but scored out of 60 to make the math easier for me.

5% bonus if you find someone completely new to work with - that you did not work with on first two assignments.

I mentioned de-seasonalizing of time series, where the seasonal variation is removed from the series to highlight variation at either higher or lower frequencies. There are a variety of techniques for doing this but the simplest is to just subtract the mean for each month from the observations. And the easiest way to find that is using lm(y~month,data=...).

1. For the Bozeman temperature data from HW 1 and 2, estimate a model with month only, subtract its fitted values from the responses (or just extract residuals). Plot the residuals vs the fractional Year variable and compare the plot of this result to the plot of the original time series.

rawd1<-read.csv("https://dl.dropboxusercontent.com/u/77307195/rawbozemandata.csv",header=T)  
  
rawd1$Year<-(floor(rawd1$DATE/100))  
rawd1$Month<-round((rawd1$DATE/100-rawd1$Year)\*100,1)  
rawd1$Yearfrac<-rawd1$Year+(rawd1$Month-1)/12  
rawd1$Monthf<-factor(rawd1$Month)  
  
m\_month<-lm(MMXT~Monthf,data=rawd1)  
source("https://dl.dropboxusercontent.com/u/77307195/concise\_lm.R") #Function to clean up lm model summaries  
print(summary(m\_month), concise = T)

##   
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 31.8357 0.4581 69.490 <1e-04 \*\*\*  
## Monthf2 3.7316 0.6465 5.772 <1e-04 \*\*\*  
## Monthf3 11.2238 0.6465 17.361 <1e-04 \*\*\*  
## Monthf4 22.0161 0.6465 34.054 <1e-04 \*\*\*  
## Monthf5 31.2704 0.6479 48.264 <1e-04 \*\*\*  
## Monthf6 39.7591 0.6479 61.366 <1e-04 \*\*\*  
## Monthf7 49.6484 0.6508 76.293 <1e-04 \*\*\*  
## Monthf8 48.5486 0.6493 74.769 <1e-04 \*\*\*  
## Monthf9 37.6803 0.6508 57.902 <1e-04 \*\*\*  
## Monthf10 25.8731 0.6493 39.847 <1e-04 \*\*\*  
## Monthf11 10.4135 0.6493 16.038 <1e-04 \*\*\*  
## Monthf12 1.8095 0.6508 2.781 0.0055 \*\*   
##   
## Residual standard error: 4.913 on 1362 degrees of freedom  
## Multiple R-squared: 0.9256, Adjusted R-squared: 0.925   
## F-statistic: 1540 on 11 and 1362 DF, p-value: < 1e-04

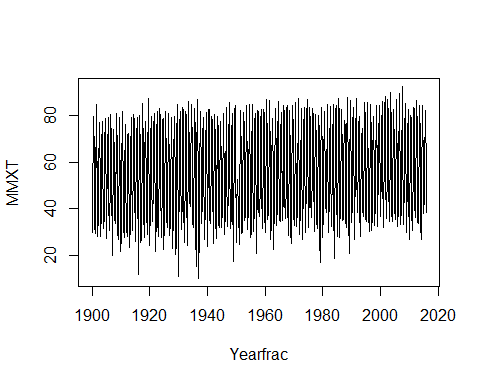
#Or using pander:  
require(pander)  
  
pander(summary(m\_month))

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| **Monthf2** | 3.732 | 0.6465 | 5.772 | 9.689e-09 |
| **Monthf3** | 11.22 | 0.6465 | 17.36 | 3.802e-61 |
| **Monthf4** | 22.02 | 0.6465 | 34.05 | 2.117e-184 |
| **Monthf5** | 31.27 | 0.6479 | 48.26 | 3.531e-297 |
| **Monthf6** | 39.76 | 0.6479 | 61.37 | 0 |
| **Monthf7** | 49.65 | 0.6508 | 76.29 | 0 |
| **Monthf8** | 48.55 | 0.6493 | 74.77 | 0 |
| **Monthf9** | 37.68 | 0.6508 | 57.9 | 0 |
| **Monthf10** | 25.87 | 0.6493 | 39.85 | 8.311e-231 |
| **Monthf11** | 10.41 | 0.6493 | 16.04 | 3.741e-53 |
| **Monthf12** | 1.809 | 0.6508 | 2.781 | 0.005501 |
| **(Intercept)** | 31.84 | 0.4581 | 69.49 | 0 |

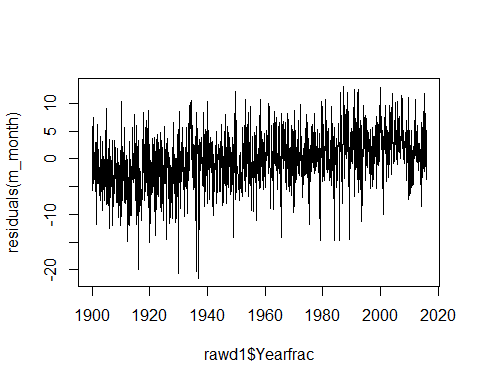
Fitting linear model: MMXT ~ Monthf

|  |  |  |  |
| --- | --- | --- | --- |
| Observations | Residual Std. Error |  | Adjusted |
| 1374 | 4.913 | 0.9256 | 0.925 |

plot(MMXT~Yearfrac,data=rawd1,type="l")



plot(residuals(m\_month)~rawd1$Yearfrac,type="l")



* The first plot has the clear seasonality present and a wide range on the y-axis that makes seeing the long term change in the mean hard although there is a hint of increasing mean temperatures in the plot.
* Once the series is de-seasonalized, the trend is a little more obvious as are unusual months in the series, especially colder than average ones. It also appears that there could be some decrease in variability in the final decades of the series as compared to the earliest years. It is unclear if this is a data quality change (digital vs analog) or an actual change in the process.
* With the seasonal variation removed, we can focus on the trend more clearly since no fixed variation at the seasonal frequency (every twelve months) can be present. We might still see some variation at this frequency in subsets of the series if it changed amplitude or phase slightly over time.

1. In the de-seasonalized Bozeman temperature data set, re-assess evidence for the linear trend. Compare the result (test statistic, degrees of freedom and size of p-value) of just fitting a linear time trend in these de-seasonalized responses to the result from our original model with a linear year component and a month adjustment (not the quadratic trend model).

m2<-lm(residuals(m\_month)~rawd1$Yearfrac)  
print(summary(m2), concise = T)

##   
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.012e+02 7.229e+00 -13.99 <1e-04 \*\*\*  
## rawd1$Yearfrac 5.166e-02 3.691e-03 14.00 <1e-04 \*\*\*  
##   
## Residual standard error: 4.579 on 1372 degrees of freedom  
## Multiple R-squared: 0.1249, Adjusted R-squared: 0.1243   
## F-statistic: 195.9 on 1 and 1372 DF, p-value: < 1e-04

m3<-lm(MMXT~Yearfrac+Monthf,data=rawd1)  
print(summary(m3), concise = T)

##   
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -69.290162 7.266158 -9.536 <1e-04 \*\*\*  
## Yearfrac 0.051674 0.003706 13.942 <1e-04 \*\*\*  
## Monthf2 3.701446 0.604984 6.118 <1e-04 \*\*\*  
## Monthf3 11.189381 0.604985 18.495 <1e-04 \*\*\*  
## Monthf4 21.977317 0.604986 36.327 <1e-04 \*\*\*  
## Monthf5 31.224003 0.606297 51.500 <1e-04 \*\*\*  
## Monthf6 39.708393 0.606299 65.493 <1e-04 \*\*\*  
## Monthf7 49.564963 0.608994 81.388 <1e-04 \*\*\*  
## Monthf8 48.465835 0.607645 79.760 <1e-04 \*\*\*  
## Monthf9 37.629365 0.608976 61.791 <1e-04 \*\*\*  
## Monthf10 25.822126 0.607627 42.497 <1e-04 \*\*\*  
## Monthf11 10.315109 0.607657 16.975 <1e-04 \*\*\*  
## Monthf12 1.732846 0.608989 2.845 0.0045 \*\*   
##   
## Residual standard error: 4.597 on 1361 degrees of freedom  
## Multiple R-squared: 0.9349, Adjusted R-squared: 0.9343   
## F-statistic: 1628 on 12 and 1361 DF, p-value: < 1e-04

* With t(1372)=14 and a p-value<0.0001, there is strong evidence of a linear trend after we remove any seasonality in the time series. This is very similar to the original result of t(1361)=13.94 from question 1 in HW 2. The degrees of freedom in the earlier model were slightly lower to more correctly reflect that we accounted for (and thus needed to estimate) the month to month variation within that model. This version is not accounting for that adjustment even though we did use some of the original information in the data set to estimate those monthly means.

1. I briefly discussed the HADCRUT data set in class. We will consider the HADCRUT4 series of temperature anomalies for the Nothern Hemisphere. The fully up to date data set is available at: <http://www.metoffice.gov.uk/hadobs/hadcrut4/data/current/time_series/HadCRUT.4.4.0.0.monthly_nh.txt>

Download the ensemble median monthly northern hemisphere temperature data. We will use the entire time series that is currently available (January 1850 to July 2016). You might want to read <http://www.metoffice.gov.uk/hadobs/hadcrut4/data/current/series_format.html> for more information on the columns in the data set.

Because the time series is complete over the time frame under consideration, you can use ts() to create a time variable instead of messing around with their Year/Month variable.

Make a plot versus time of the ensemble medians and use that as your response variable in the following questions. Discuss trend, seasonality, outliers, and variability.

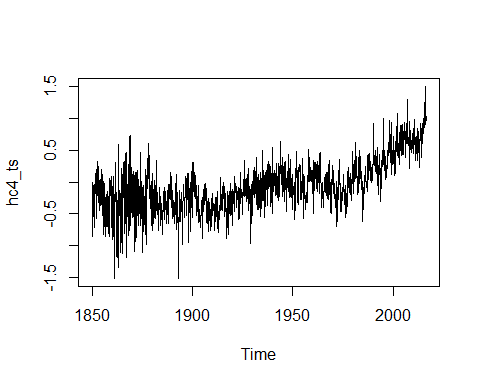
data\_ensemble<-read.table("http://www.metoffice.gov.uk/hadobs/hadcrut4/data/4.4.0.0/time\_series/HadCRUT.4.4.0.0.monthly\_nh.txt")  
head(data\_ensemble)

## V1 V2 V3 V4 V5 V6 V7 V8 V9 V10  
## 1 1850/01 -0.845 -0.921 -0.748 -1.037 -0.652 -1.379 -0.311 -1.052 -0.635  
## 2 1850/02 -0.046 -0.116 0.043 -0.248 0.155 -0.617 0.524 -0.265 0.173  
## 3 1850/03 -0.699 -0.795 -0.633 -0.892 -0.507 -1.124 -0.274 -0.911 -0.496  
## 4 1850/04 -0.601 -0.669 -0.526 -0.800 -0.402 -0.955 -0.247 -0.812 -0.387  
## 5 1850/05 -0.413 -0.491 -0.366 -0.610 -0.216 -0.669 -0.157 -0.627 -0.213  
## 6 1850/06 0.000 -0.068 0.054 -0.185 0.186 -0.217 0.218 -0.197 0.194  
## V11 V12  
## 1 -1.417 -0.271  
## 2 -0.657 0.564  
## 3 -1.175 -0.230  
## 4 -1.012 -0.186  
## 5 -0.749 -0.091  
## 6 -0.294 0.291

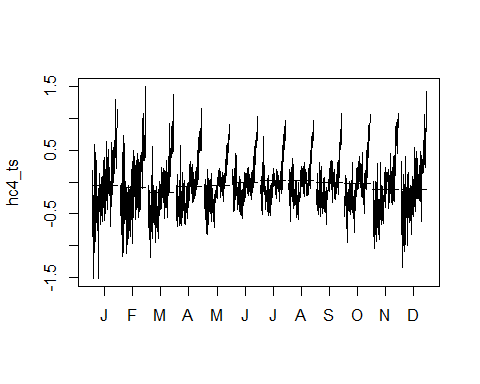
data\_hc4<-data\_ensemble[,c(1,2)]  
names(data\_hc4)<-c("Year","TempAnomaly\_C")  
head(data\_hc4)

## Year TempAnomaly\_C  
## 1 1850/01 -0.845  
## 2 1850/02 -0.046  
## 3 1850/03 -0.699  
## 4 1850/04 -0.601  
## 5 1850/05 -0.413  
## 6 1850/06 0.000

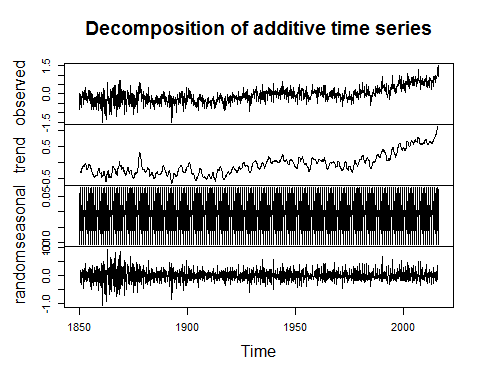
hc4\_ts<-ts(data\_hc4$TempAnomaly\_C,start=c(1850,1),freq=12)  
require(TSA) #To be able to use season  
hc42<-data.frame(TempAnomaly\_C=data\_hc4$TempAnomaly\_C,Year=time(hc4\_ts),Month=season(hc4\_ts))  
plot(hc4\_ts)



monthplot(hc4\_ts)



plot(decompose(hc4\_ts)) #See CM 1.5.3 for more details



* There is an increasing mean temperature anomaly in the last 50 or so years. The increases do not appear to be linear over this long time frame.
* The variability seems to have decreased after 1900 - maybe due to more weather stations or a change in the process?
* There is little or no apparent seasonality that is obvious in the plot of the entire time series. Certainly not the wide swings we would expect from Winter to Summer.
* There are a few outliers, with the most extreme in the late 1880s with a very cold temperature anomaly. Similar cold values earlier in the series are not outliers since there are neighboring similar values.

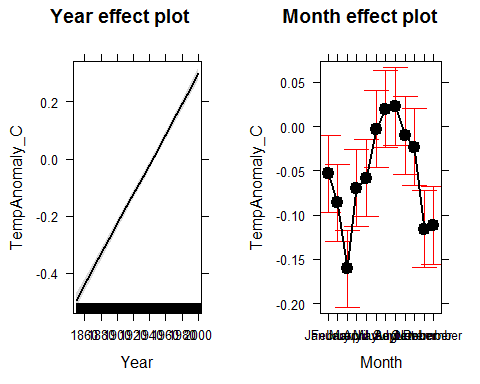
1. Our main focus with these data will be on estimating the long-term trend, starting with polynomial trend models. But first, check for seasonality in a model that accounts for a linear time trend. Use our previous fractional year for the trend. Report an effects plot and a test for the month model component.

Note: when you use time() to generate the Year variable from a time series object it retains some time series object information that can cause conflicts later on. Create a new variable in your data.frame that uses something like as.vector(time(tsdataname)).

m1<-lm(TempAnomaly\_C~Year+Month,data=hc42)  
require(car)  
Anova(m1)

## Anova Table (Type II tests)  
##   
## Response: TempAnomaly\_C  
## Sum Sq Df F value Pr(>F)   
## Year 129.090 1 1567.241 < 2.2e-16 \*\*\*  
## Month 6.104 11 6.737 3.389e-11 \*\*\*  
## Residuals 163.583 1986   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

require(effects)  
plot(allEffects(m1))



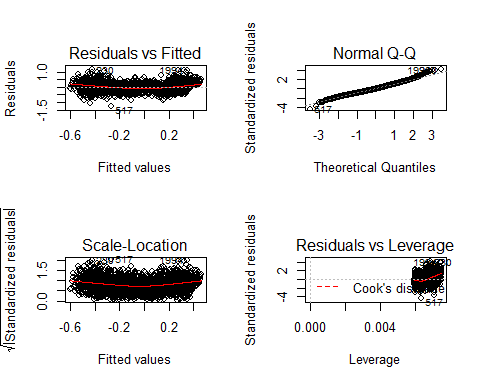
print(summary(m1),concise=T)

##   
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.027e+01 2.590e-01 -39.654 < 1e-04 \*\*\*  
## Year 5.285e-03 1.335e-04 39.588 < 1e-04 \*\*\*  
## MonthFebruary -3.281e-02 3.141e-02 -1.045 0.296288   
## MonthMarch -1.072e-01 3.141e-02 -3.413 0.000656 \*\*\*  
## MonthApril -1.621e-02 3.141e-02 -0.516 0.605890   
## MonthMay -5.031e-03 3.141e-02 -0.160 0.872754   
## MonthJune 5.062e-02 3.141e-02 1.612 0.107154   
## MonthJuly 7.306e-02 3.141e-02 2.326 0.020102 \*   
## MonthAugust 7.620e-02 3.145e-02 2.423 0.015502 \*   
## MonthSeptember 4.339e-02 3.145e-02 1.379 0.167953   
## MonthOctober 3.006e-02 3.145e-02 0.956 0.339359   
## MonthNovember -6.260e-02 3.145e-02 -1.990 0.046701 \*   
## MonthDecember -5.818e-02 3.145e-02 -1.850 0.064507 .   
##   
## Residual standard error: 0.287 on 1986 degrees of freedom  
## Multiple R-squared: 0.4525, Adjusted R-squared: 0.4492   
## F-statistic: 136.8 on 12 and 1986 DF, p-value: < 1e-04

* I didn't ask for any specific discussion so did not grade your sentences other than looking for you to extract the test result fully from the output but a sentence for the test would be useful.
* There is strong evidence of a seasonal variation (F(11,1986)=6.74, p-value<0.0001) in the HADCRUT median series *after accounting for a linear trend in the responses*.
* The seasonal pattern is a little odd with coldest month being march. July and August being warmest makes some sense. Remember that this is the seasonality left after removing the monthly means from 1951-1990. So there is some missed seasonality in the responses given the way they are generated. It is fairly minor at 0.20 degrees C of maximum difference in months compared to about 0.8 degrees C of change in the linear trend over this timeframe.

1. Check the residuals versus fitted values for any evidence of nonlinearity in the residuals vs fitted that was missed by the model with a linear trend and month component. Also note any potential issues with the constant variance assumption.

par(mfrow=c(2,2))  
plot(m1)



* The variability looks to be relatively consistent but there is somewhat larger variability at the smallest fitted values. The original plot vs time made this more clear (or a plot of residuals vs time).
* There appears to be a fairly clear curve in the residuals vs fitted which suggests that some nonlinearity was missed in the first model.

1. You can add higher order polynomial terms to models using x1+I(x1^2)+I(x1^3)... or using the poly function, such as poly(x1,3,raw=T) for a cubic polynomial that includes the linear and quadratic components (we want this!). The raw=T keeps the variables in their raw or input format. Estimate the same model but now using polynomial trends that steps up from linear (poly(time,1,raw=T)) and stop when you get a failure to estimate a part of the model. Briefly discuss what happened.

hc43<-data.frame(TempAnomaly\_C=data\_hc4$TempAnomaly\_C,Year=as.vector(time(hc4\_ts)),Month=season(hc4\_ts))  
m2<-lm(TempAnomaly\_C~poly(Year,2,raw=TRUE)+Month,data=hc43)  
m2

##   
## Call:  
## lm(formula = TempAnomaly\_C ~ poly(Year, 2, raw = TRUE) + Month,   
## data = hc43)  
##   
## Coefficients:  
## (Intercept) poly(Year, 2, raw = TRUE)1   
## 2.223e+02 -2.354e-01   
## poly(Year, 2, raw = TRUE)2 MonthFebruary   
## 6.225e-05 -3.281e-02   
## MonthMarch MonthApril   
## -1.072e-01 -1.620e-02   
## MonthMay MonthJune   
## -5.027e-03 5.063e-02   
## MonthJuly MonthAugust   
## 7.306e-02 7.793e-02   
## MonthSeptember MonthOctober   
## 4.512e-02 3.179e-02   
## MonthNovember MonthDecember   
## -6.087e-02 -5.645e-02

m3<-lm(TempAnomaly\_C~poly(Year,3,raw=TRUE)+Month,data=hc43)  
m3

##   
## Call:  
## lm(formula = TempAnomaly\_C ~ poly(Year, 3, raw = TRUE) + Month,   
## data = hc43)  
##   
## Coefficients:  
## (Intercept) poly(Year, 3, raw = TRUE)1   
## -2.893e+03 4.602e+00   
## poly(Year, 3, raw = TRUE)2 poly(Year, 3, raw = TRUE)3   
## -2.441e-03 4.316e-07   
## MonthFebruary MonthMarch   
## -3.291e-02 -1.074e-01   
## MonthApril MonthMay   
## -1.651e-02 -5.432e-03   
## MonthJune MonthJuly   
## 5.012e-02 7.246e-02   
## MonthAugust MonthSeptember   
## 7.782e-02 4.491e-02   
## MonthOctober MonthNovember   
## 3.149e-02 -6.127e-02   
## MonthDecember   
## -5.695e-02

m4<-lm(TempAnomaly\_C~poly(Year,4,raw=TRUE)+Month,data=hc43)  
m4

##   
## Call:  
## lm(formula = TempAnomaly\_C ~ poly(Year, 4, raw = TRUE) + Month,   
## data = hc43)  
##   
## Coefficients:  
## (Intercept) poly(Year, 4, raw = TRUE)1   
## 2.146e+05 -4.458e+02   
## poly(Year, 4, raw = TRUE)2 poly(Year, 4, raw = TRUE)3   
## 3.472e-01 -1.202e-04   
## poly(Year, 4, raw = TRUE)4 MonthFebruary   
## 1.560e-08 -3.291e-02   
## MonthMarch MonthApril   
## -1.074e-01 -1.650e-02   
## MonthMay MonthJune   
## -5.425e-03 5.013e-02   
## MonthJuly MonthAugust   
## 7.246e-02 7.885e-02   
## MonthSeptember MonthOctober   
## 4.594e-02 3.252e-02   
## MonthNovember MonthDecember   
## -6.024e-02 -5.592e-02

m5<-lm(TempAnomaly\_C~poly(Year,5,raw=TRUE)+Month,data=hc43)  
m5 #Fails at 5th order polynomial with failing to estimate one of the polynomial coefficients

##   
## Call:  
## lm(formula = TempAnomaly\_C ~ poly(Year, 5, raw = TRUE) + Month,   
## data = hc43)  
##   
## Coefficients:  
## (Intercept) poly(Year, 5, raw = TRUE)1   
## 2.146e+05 -4.458e+02   
## poly(Year, 5, raw = TRUE)2 poly(Year, 5, raw = TRUE)3   
## 3.472e-01 -1.202e-04   
## poly(Year, 5, raw = TRUE)4 poly(Year, 5, raw = TRUE)5   
## 1.560e-08 NA   
## MonthFebruary MonthMarch   
## -3.291e-02 -1.074e-01   
## MonthApril MonthMay   
## -1.650e-02 -5.425e-03   
## MonthJune MonthJuly   
## 5.013e-02 7.246e-02   
## MonthAugust MonthSeptember   
## 7.885e-02 4.594e-02   
## MonthOctober MonthNovember   
## 3.252e-02 -6.024e-02   
## MonthDecember   
## -5.592e-02

* The correct answer here is that the multicollinearity grew so extreme that the model was not estimable as specified. You might also note that 2000^5 is a really large number and we are bordering on having values that are hard to compute and store. That is an issue but not the main issue. To show this, consider rescaling of the Year variable by dividing by 2000:

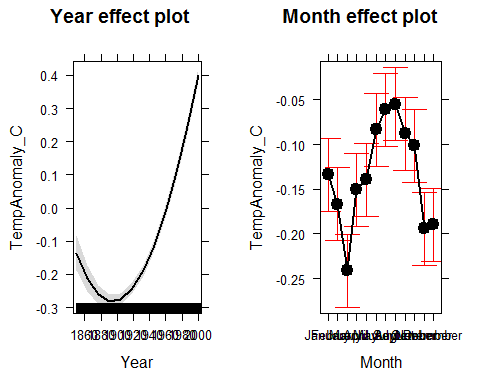
m5<-lm(TempAnomaly\_C~poly(Year/2000,5,raw=TRUE)+Month,data=hc43)  
m5

##   
## Call:  
## lm(formula = TempAnomaly\_C ~ poly(Year/2000, 5, raw = TRUE) +   
## Month, data = hc43)  
##   
## Coefficients:  
## (Intercept) poly(Year/2000, 5, raw = TRUE)1   
## 2.146e+05 -8.917e+05   
## poly(Year/2000, 5, raw = TRUE)2 poly(Year/2000, 5, raw = TRUE)3   
## 1.389e+06 -9.615e+05   
## poly(Year/2000, 5, raw = TRUE)4 poly(Year/2000, 5, raw = TRUE)5   
## 2.496e+05 NA   
## MonthFebruary MonthMarch   
## -3.291e-02 -1.074e-01   
## MonthApril MonthMay   
## -1.650e-02 -5.425e-03   
## MonthJune MonthJuly   
## 5.013e-02 7.246e-02   
## MonthAugust MonthSeptember   
## 7.885e-02 4.594e-02   
## MonthOctober MonthNovember   
## 3.252e-02 -6.024e-02   
## MonthDecember   
## -5.592e-02

* It fails at the same polynomial because the failure is about multi-collinearity and that the inverse of is not computable (the matrix is not full rank). There is nothing magical about the 5th order polynomial here, it just is where it happened in this situation. I've had situations where cubics caused issues and others where you could go a little higher.

1. If we center or, even better, make the polynomial functions orthogonal to one another, we can avoid the issue in the previous question. Use poly(x1,?,raw=F) and step up the polynomial order for time until the p-value for the last coefficient (use summary()) is "large", reporting the single test result for each step in the building process. Then drop back one order and re-fit the model. Report the effects plot of the resulting model and describe the estimated trend. Note: aside from access to orthogonal polynomials the poly function interfaces with Anova and the effects package really nicely.

plot(allEffects(m4)) #Plotting last estimable model from the raw variable version



m2o<-lm(TempAnomaly\_C~poly(Year,2,raw=FALSE)+Month,data=hc43)  
summary(m2o)$coefficients[1:3,]

## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -0.05408131 0.0198368 -2.726312 6.460937e-03  
## poly(Year, 2, raw = FALSE)1 11.36184772 0.2563485 44.321876 7.438213e-299  
## poly(Year, 2, raw = FALSE)2 5.75690413 0.2563529 22.456948 1.039852e-99

m3o<-lm(TempAnomaly\_C~poly(Year,3,raw=FALSE)+Month,data=hc43)  
summary(m3o)$coefficients[c(1,4),]

## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -0.05377813 0.01962449 -2.740358 6.192213e-03  
## poly(Year, 3, raw = FALSE)3 1.68591689 0.25360583 6.647784 3.833515e-11

m4o<-lm(TempAnomaly\_C~poly(Year,4,raw=FALSE)+Month,data=hc43)  
summary(m4o)$coefficients[c(1,5),]

## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -0.05420912 0.01911989 -2.835221 4.625830e-03  
## poly(Year, 4, raw = FALSE)4 2.55726933 0.24709130 10.349492 1.744968e-24

m5o<-lm(TempAnomaly\_C~poly(Year,5,raw=FALSE)+Month,data=hc43)  
summary(m5o)$coefficients[c(1,6),]

## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -0.05386487 0.01894766 -2.842824 4.517378e-03  
## poly(Year, 5, raw = FALSE)5 1.49407377 0.24485925 6.101766 1.258292e-09

m6o<-lm(TempAnomaly\_C~poly(Year,6,raw=FALSE)+Month,data=hc43)  
summary(m6o)$coefficients[c(1,7),]

## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -0.05354585 0.01875655 -2.854781 4.351465e-03  
## poly(Year, 6, raw = FALSE)6 -1.56359500 0.24239794 -6.450529 1.396915e-10

m7o<-lm(TempAnomaly\_C~poly(Year,7,raw=FALSE)+Month,data=hc43)  
summary(m7o)$coefficients[c(1,8),]

## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -0.05419699 0.01831197 -2.959648 3.116474e-03  
## poly(Year, 7, raw = FALSE)7 -2.34724589 0.23664357 -9.918908 1.152128e-22

m8o<-lm(TempAnomaly\_C~poly(Year,8,raw=FALSE)+Month,data=hc43)  
summary(m8o)$coefficients[c(1,9),]

## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -0.05449431 0.01818469 -2.996714 2.763001e-03  
## poly(Year, 8, raw = FALSE)8 1.26189838 0.23500883 5.369579 8.819812e-08

m9o<-lm(TempAnomaly\_C~poly(Year,9,raw=FALSE)+Month,data=hc43)  
summary(m9o)$coefficients[c(1,10),]

## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -0.05404531 0.01802931 -2.997636 2.754713e-03  
## poly(Year, 9, raw = FALSE)9 1.38413571 0.23298955 5.940763 3.343646e-09

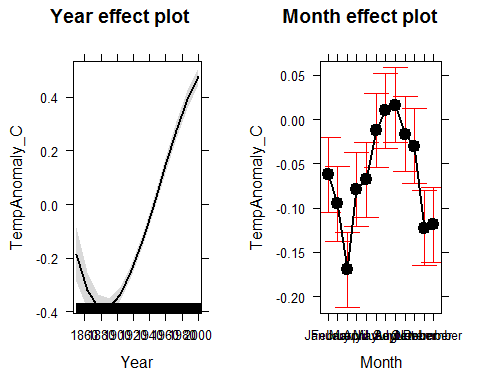
m10o<-lm(TempAnomaly\_C~poly(Year,10,raw=FALSE)+Month,data=hc43)  
summary(m10o)$coefficients[c(1,11),]

## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -0.05412977 0.01802545 -3.002963 0.002707181  
## poly(Year, 10, raw = FALSE)10 0.31861274 0.23295112 1.367724 0.171554074

print(summary(m9o),concise=T)

##   
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.054045 0.018029 -2.998 0.00275 \*\*   
## poly(Year, 9, raw = FALSE)1 11.361883 0.232983 48.767 < 1e-04 \*\*\*  
## poly(Year, 9, raw = FALSE)2 5.757045 0.232987 24.710 < 1e-04 \*\*\*  
## poly(Year, 9, raw = FALSE)3 1.685934 0.232984 7.236 < 1e-04 \*\*\*  
## poly(Year, 9, raw = FALSE)4 2.557257 0.232991 10.976 < 1e-04 \*\*\*  
## poly(Year, 9, raw = FALSE)5 1.494041 0.232986 6.413 < 1e-04 \*\*\*  
## poly(Year, 9, raw = FALSE)6 -1.563434 0.232995 -6.710 < 1e-04 \*\*\*  
## poly(Year, 9, raw = FALSE)7 -2.347166 0.232988 -10.074 < 1e-04 \*\*\*  
## poly(Year, 9, raw = FALSE)8 1.261898 0.232999 5.416 < 1e-04 \*\*\*  
## poly(Year, 9, raw = FALSE)9 1.384136 0.232990 5.941 < 1e-04 \*\*\*  
## MonthFebruary -0.032950 0.025496 -1.292 0.19640   
## MonthMarch -0.107472 0.025497 -4.215 < 1e-04 \*\*\*  
## MonthApril -0.016633 0.025497 -0.652 0.51423   
## MonthMay -0.005608 0.025497 -0.220 0.82595   
## MonthJune 0.049893 0.025497 1.957 0.05051 .   
## MonthJuly 0.072174 0.025498 2.831 0.00469 \*\*   
## MonthAugust 0.078722 0.025536 3.083 0.00208 \*\*   
## MonthSeptember 0.045781 0.025536 1.793 0.07316 .   
## MonthOctober 0.032323 0.025536 1.266 0.20573   
## MonthNovember -0.060475 0.025536 -2.368 0.01797 \*   
## MonthDecember -0.056194 0.025536 -2.201 0.02788 \*   
##   
## Residual standard error: 0.233 on 1978 degrees of freedom  
## Multiple R-squared: 0.6407, Adjusted R-squared: 0.637   
## F-statistic: 176.3 on 20 and 1978 DF, p-value: < 1e-04

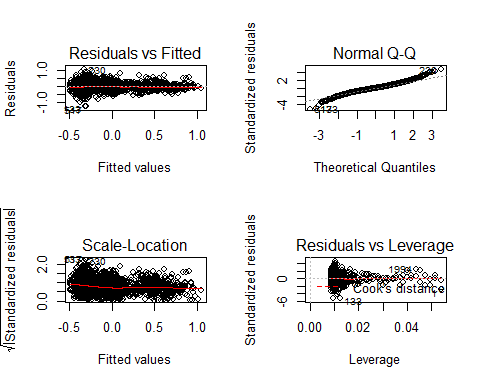
plot(allEffects(m9o))



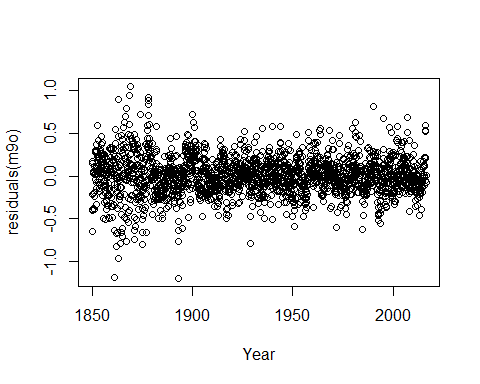
* Squared component in quadratic: t(1985)=22.5, p-value<0.0001
* Cubed component in cubic: t(1984)=6.65, p-value<0.0001
* fourth order in quartic: t(1983)=10.35, p-value<0.0001
* fifth order in quintic: t(1982)=6.1, p-value<0.0001
* sixth order: t(1981)=-6.451, p-value<0.0001
* seventh order: t(1980)=-9.919, p-val<0.0001
* eigth order: t(1979)=5.37, p-value<0.0001
* ninth order: t(1978)=5.9, p-val<0.0001
* 10th order: t(1977)=1.37, p-val=0.172
* So we would step back to the 9th order result and summarize and explore it. The estimated trend when you combine all the polynomial pieces suggests a steep decline in mean temperatures from 1860 to 1890, then a relatively linear and relatively steep increase until around 1980 and then a bit of decrease in the rate of increase in the final years of the data set.
* Note that we could extract approximate slopes for the linear sections with a drop from -0.2 to -0.4 in the first 30 years providing a drop of about 0.07 degrees C per decade. This is followed by an increase from -0.4 to 0.3 between 1890 and 1980 or a change of about 0.08 degrees C per decade over this span. The last part of the time range is not displayed in the effect plot but we could do some sort of similar estimate pulling values from the model using predict() if needed.

1. Check the diagnostic plots from your final model. Does anything improve from the first version. Also plot the residuals vs time and compare that plot to residuals vs fitted.

par(mfrow=c(2,2))  
plot(m9o)



par(mfrow=c(1,1))  
plot(residuals(m9o)~Year,data=hc43)



* The linearity issue seems to be resolved (we don't have any apparent missed nonlinearity).
* The variability may not longer be constant and the normality assumption looks more tenuous as it might be heavy tailed now. These are both related to the higher amounts of variability in the residuals from the lower fitted values.
* The residuals plotted over time show a slight decrease in variation over time but it is possibly more clear with residuals vs fitted that there is changing variance. It makes more sense that the variability in temperatures might be related to the mean although having more variability in low temperatures isn't what I exactly expected.

1. Run the following code so I can see what version of R you are now using:

### Documenting R version

getRversion()

## [1] '3.3.1'