Homework5

ShanZhong

April 6, 2019

```
library("dlnm")
```

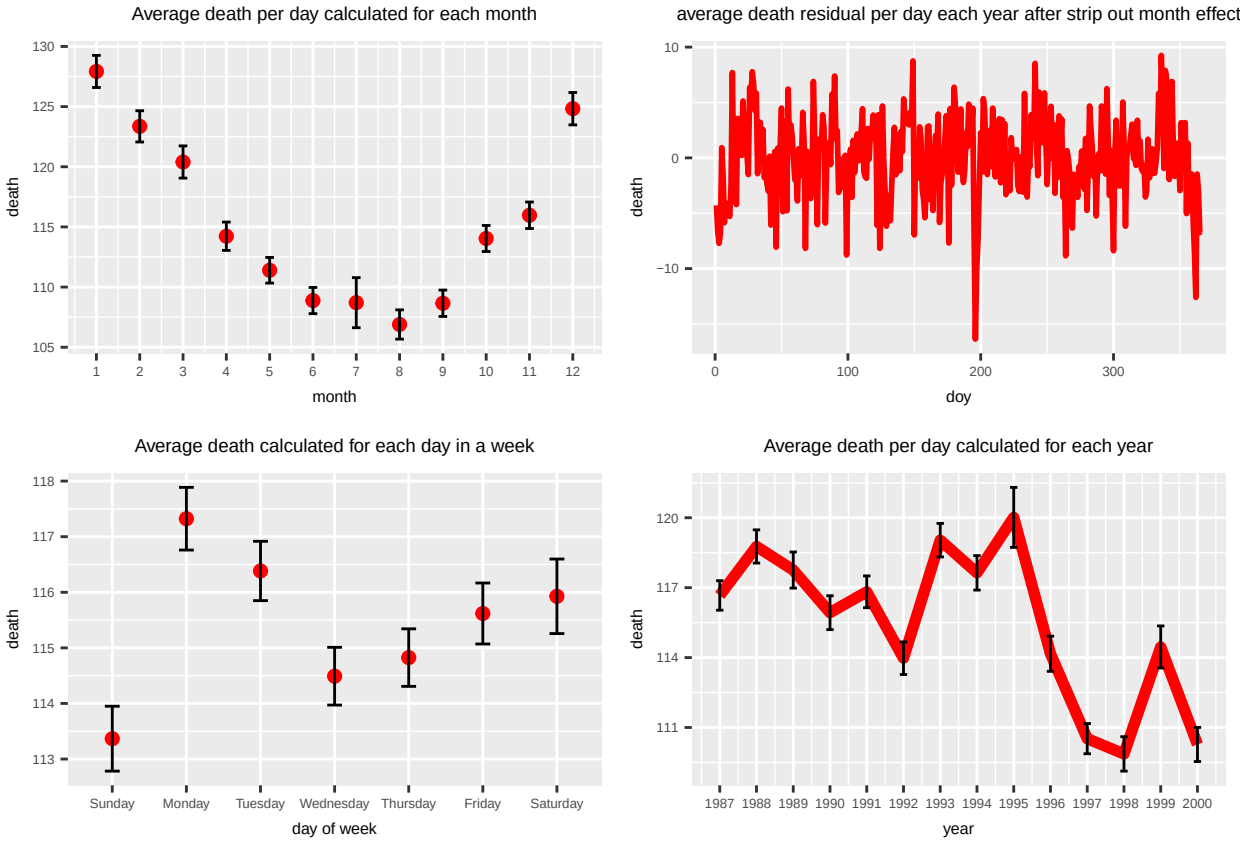
```
## This is dlnm 2.3.9. For details: help(dlnm) and vignette('dlnmOverview').
```

```
library("data.table")
data(chicagoNMMAPS)
dat<-data.table(chicagoNMMAPS)
library("ggplot2")
library("gridExtra")
```

there are strong evidence of trend and seasonality in the death

```
month.death<-dat[,.(death=mean(death),sd=sd(death)/sqrt(.N)),by=c("month")]

p1<-ggplot(data=month.death,aes(x=month,y=death))+geom_point(color="red",size=2)+
  geom_errorbar(aes(ymin=death-sd*1.96,ymax=death+sd*1.96,width=0.2))+
  scale_x_continuous(breaks=c(1:12),labels=paste(1:12))+
  ggtitle("Average death per day calculated for each month")+
  theme(plot.title=element_text(hjust=0.5),text=element_text(size=6))

dat$month<-as.factor(dat$month)
detrend<-lm(death~month,data=dat)
dat$preddeath<-predict(detrend,dat)
dat$residual<-dat$preddeath-dat$death
day.death<-dat[,.(death=mean(residual),sd=sd(residual)/sqrt(.N)),by=c("doy")]

p2<-ggplot(data=day.death,aes(x=doy,y=death))+geom_line(color="red",size=1)+
  ggtitle("average death residual per day each year after strip out month effect")+
  theme(plot.title=element_text(hjust=0.5),text=element_text(size=6))

week.death<-dat[,.(death=mean(death),sd=sd(death)/sqrt(.N)),by=c("dow")]

p3<-ggplot(data=week.death,aes(x=dow,y=death))+geom_point(color="red",size=2)+
  geom_errorbar(aes(ymin=death-sd,ymax=death+sd,width=0.2))+
  ggtitle("Average death calculated for each day in a week")+
  theme(plot.title=element_text(hjust=0.5),text=element_text(size=6))+xlab("day of week")

year.death<-dat[,.(death=mean(death),sd=sd(death)/sqrt(.N)),by=c("year")]

p4<-ggplot(data=year.death,aes(x=year,y=death))+geom_line(color="red",size=2)+
  geom_errorbar(aes(ymin=death-sd,ymax=death+sd,width=0.2))+
  scale_x_continuous(breaks=c(1987:2000),labels=paste(1987:2000))+
  ggtitle("Average death per day calculated for each year")+
  theme(plot.title=element_text(hjust=0.5),text=element_text(size=6))
```

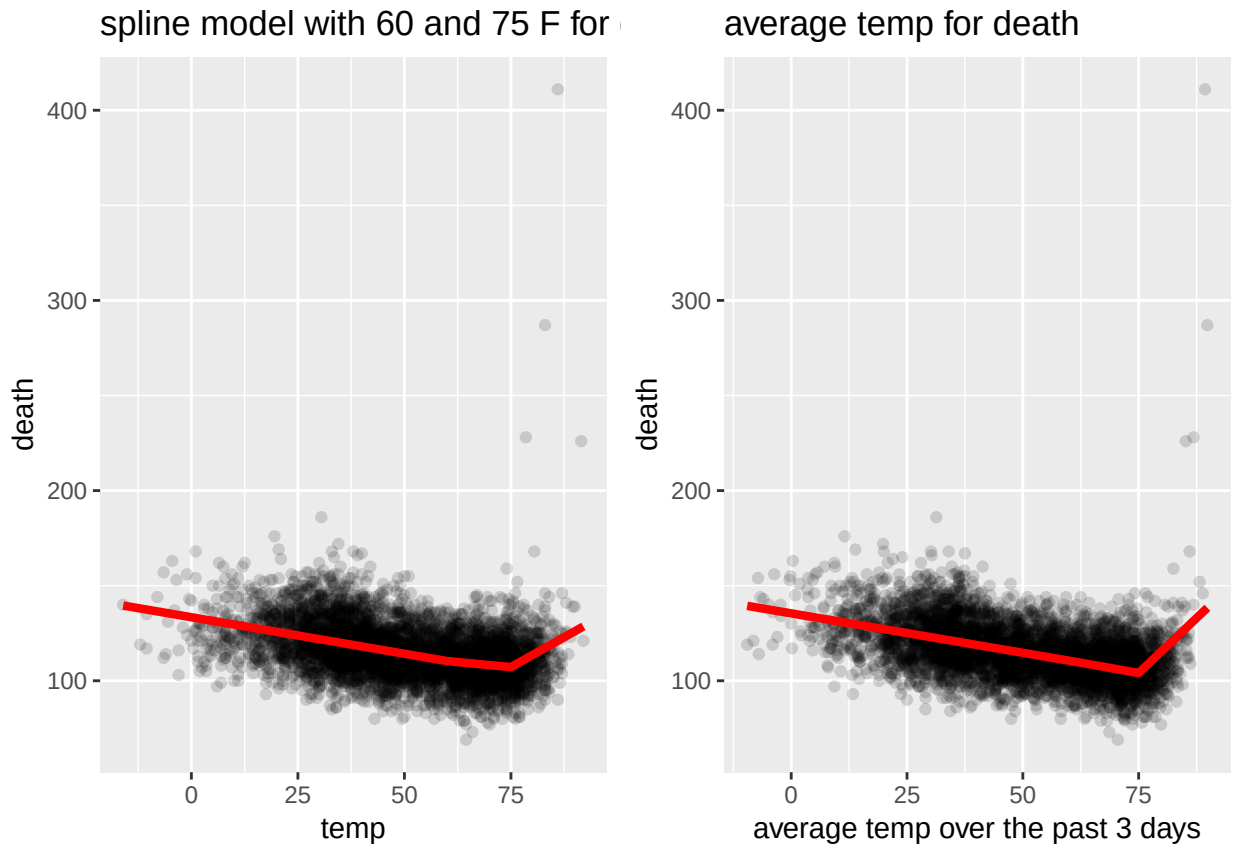
```
grid.arrange(p1,p2,p3,p4,nrow=2)
```



too hot or too cold cause people to die, as expexted

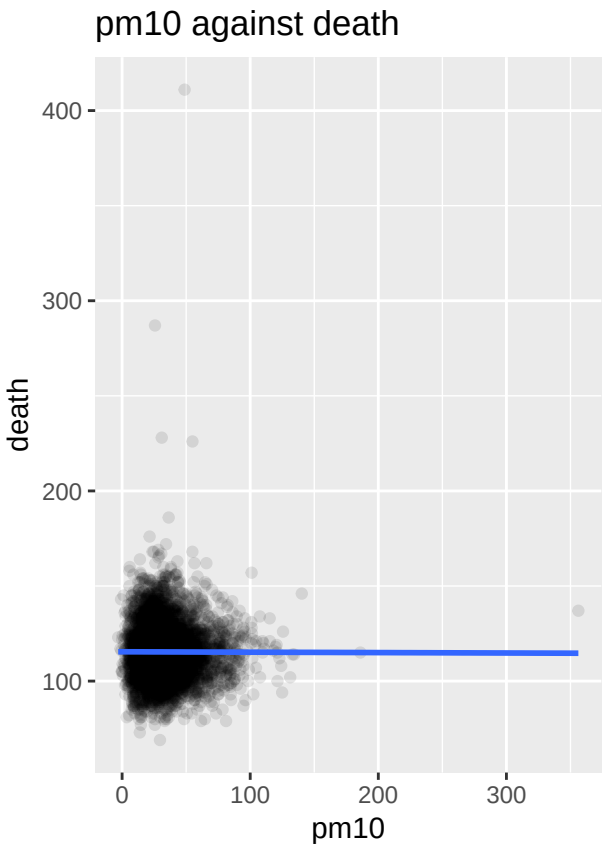
```
dat<-data.table(chicagoNMMAPS)
# C to F
dat$temp<-dat$temp*9/5+32
dat$temp60<-ifelse(dat$temp > 60, dat$temp-60, 0)
dat$temp75<-ifelse(dat$temp > 75, dat$temp-75, 0)

lm_fit<-lm(death~temp+temp60+temp75,data=dat)
predicted_death <- data.table(death_pred = predict(lm_fit, dat), temp=dat$temp)

p1<-ggplot(data=dat,aes(x=temp,y=death))+geom_point(alpha=0.15)+
  geom_line(color="red",size=1.5,data=predicted_death,aes(x=temp,y=death_pred))+
  ggtitle("spline model with 60 and 75 F for death")


store<-shift(dat$temp,1:3,fill=0,type=c("lag"))
dat$modified.temp<-(store[[1]]+store[[2]]+store[[3]])/3
rm(store)
dat$modified.temp60<-ifelse(dat$modified.temp > 60, dat$modified.temp-60, 0)
dat$modified.temp75<-ifelse(dat$modified.temp > 75, dat$modified.temp-75, 0)
```

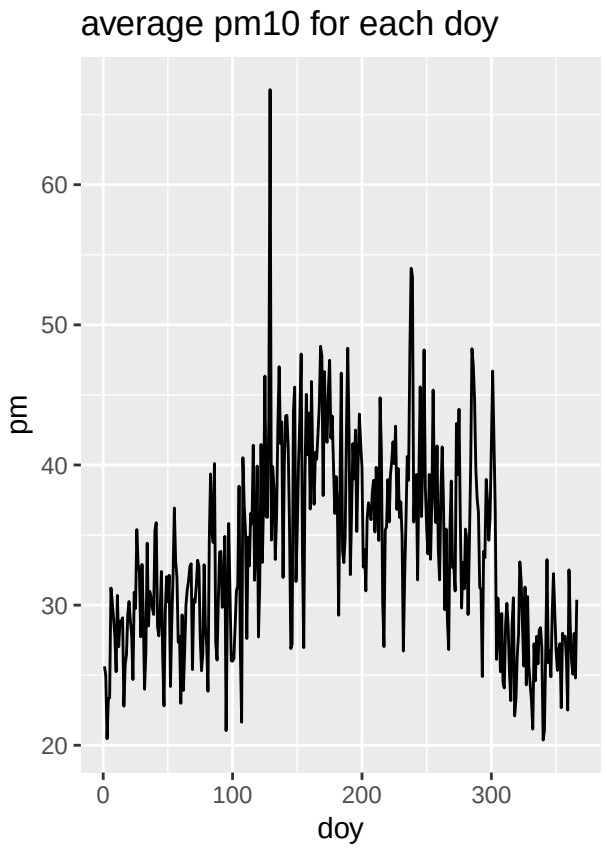
```
lm_fit<-lm(death~modified.temp+modified.temp60+modified.temp75,data=dat)
predicted_death <- data.table(death_pred = predict(lm_fit, dat), modified.temp=dat$modified.temp)

p2<-ggplot(data=dat,aes(x=modified.temp,y=death))+geom_point(alpha=0.15)+
  geom_line(color="red",size=1.5,data=predicted_death,aes(x=modified.temp,y=death_pred))+
  ggtitle("average temp for death")+xlab("average temp over the past 3 days")

grid.arrange(p1,p2,nrow=1)
```



it seems pm10 do not have much impact on death rate

```
dat<-data.table(chicagoNMMAPS)

dat<-dat[complete.cases(dat[ , c("pm10")]),]


p1<-ggplot(data=dat,aes(x=pm10,y=death))+geom_point(alpha=0.1)+geom_smooth(method=lm,se=FALSE)+
  ggtitle("pm10 against death")

a<-dat[,.(pm=mean(pm10)),by=c("doy")]
p2<-ggplot(data=a,aes(x=doy,y=pm))+geom_line()+ggtitle("average pm10 for each doy")
rm(a)

grid.arrange(p1,p2,nrow=1)
```






```
dat<-data.table(chicagoNMMAPS)

dat$cummonth<-((dat$year-1987)*12+dat$month-1)
dat$cumyear<-dat$year-1987
dat$cumday<-dat$date-dat$date[1]
dat$cumseason<-dat$cumyear*4+floor((dat$month-0.01)/3)

# already created
```

```
dat<-data.table(chicagoNMMAPS)
# C to F
dat$temp<-dat$temp*9/5+32
dat$temp60<-ifelse(dat$temp > 60, dat$temp-60, 0)
dat$temp75<-ifelse(dat$temp > 75, dat$temp-75, 0)

store<-shift(dat$temp,1:3,fill=0,type=c("lag"))
dat$modified.temp<-(store[[1]]+store[[2]]+store[[3]])/3
rm(store)
dat$modified.temp60<-ifelse(dat$modified.temp > 60, dat$modified.temp-60, 0)
dat$modified.temp75<-ifelse(dat$modified.temp > 75, dat$modified.temp-75, 0)

dat$year<-as.factor(dat$year)

fit<-glm(death~pm10+year+temp+temp60+temp75+modified.temp+modified.temp60+modified.temp75,data=dat,fami
summary(fit)
```

```
##
```

```
## Call:
## glm(formula = death ~ pm10 + year + temp + temp60 + temp75 +
##     modified.temp + modified.temp60 + modified.temp75, family = poisson,
##     data = dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.7169  -0.7996  -0.0493   0.6998  18.1368
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)      4.918e+00  7.365e-03 667.831  < 2e-16 ***
## pm10             3.087e-04  8.085e-05   3.818 0.000135 ***
## year1988         3.302e-03  7.409e-03   0.446 0.655786
## year1989         8.965e-03  7.479e-03   1.199 0.230648
## year1990        -6.843e-04  7.176e-03  -0.095 0.924027
## year1991         3.958e-03  7.219e-03   0.548 0.583455
## year1992        -2.318e-02  7.184e-03  -3.226 0.001255 **
## year1993         1.771e-02  7.093e-03   2.497 0.012536 *
## year1994         7.415e-03  7.115e-03   1.042 0.297321
## year1995         1.672e-02  7.053e-03   2.371 0.017723 *
## year1996        -2.742e-02  7.138e-03  -3.841 0.000122 ***
## year1997        -5.734e-02  7.193e-03  -7.972 1.56e-15 ***
## year1998        -4.562e-02  7.210e-03  -6.327 2.50e-10 ***
## year1999        -1.980e-02  7.127e-03  -2.778 0.005470 **
## year2000        -5.126e-02  7.231e-03  -7.089 1.36e-12 ***
## temp             1.706e-04  2.107e-04   0.810 0.418174
## temp60           2.342e-03  5.704e-04   4.106 4.03e-05 ***
## temp75           1.334e-03  1.322e-03   1.009 0.312773
## modified.temp   -3.700e-03  2.144e-04 -17.256  < 2e-16 ***
## modified.temp60 -2.805e-03  5.713e-04  -4.909 9.16e-07 ***
## modified.temp75  2.377e-02  1.504e-03  15.799  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 9383.4  on 4862  degrees of freedom
## Residual deviance: 6677.6  on 4842  degrees of freedom
##   (251 observations deleted due to missingness)
## AIC: 38714
##
## Number of Fisher Scoring iterations: 4
```

```
# AIC 40721
# deviance 9873.8
```

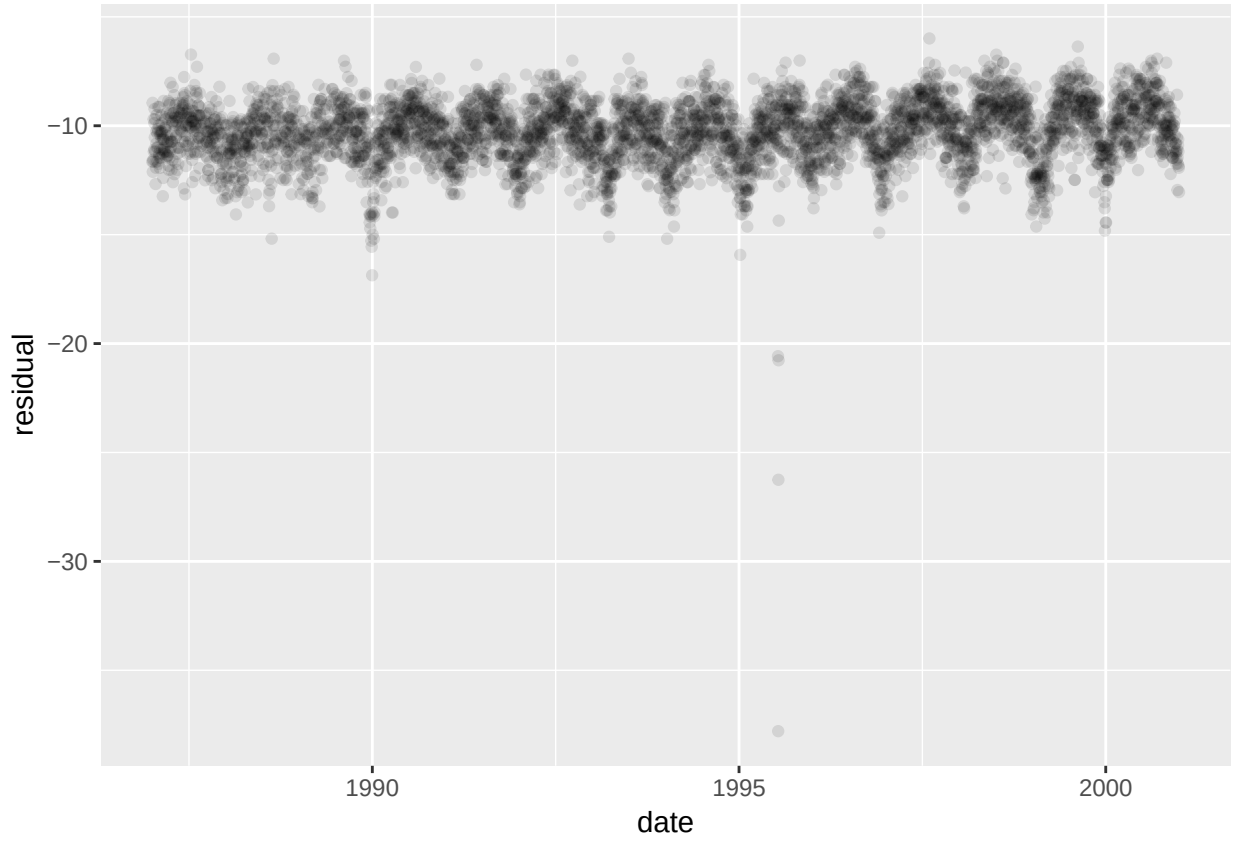
```
dat<-data.table(chicagoNMMAPS)

dat$season<-floor((dat$month-0.01)/3)+1

# C to F
dat$temp<-dat$temp*9/5+32
dat$temp60<-ifelse(dat$temp > 60, dat$temp-60, 0)
dat$temp75<-ifelse(dat$temp > 75, dat$temp-75, 0)
```

```r
store<-shift(dat$temp,1:3,fill=0,type=c("lag"))
dat$modified.temp<-(store[[1]]+store[[2]]+store[[3]])/3
rm(store)
dat$modified.temp60<-ifelse(dat$modified.temp > 60, dat$modified.temp-60, 0)
dat$modified.temp75<-ifelse(dat$modified.temp > 75, dat$modified.temp-75, 0)

fit1<-glm(death~pm10+as.factor(year)+temp+temp60+temp75+modified.temp+modified.temp60+modified.temp75,da
fit2<-glm(death~pm10+as.factor(season)+temp+temp60+temp75+modified.temp+modified.temp60+modified.temp75
fit3<-glm(death~pm10+as.factor(month)+temp+temp60+temp75+modified.temp+modified.temp60+modified.temp75,

table<-matrix(NA,nrow=3,ncol=6)
colnames(table)<-c(" PM10 coef"," Standard error","95% CI","Model df","Deviance","AIC")
rownames(table)<-c("Year","Season","Month")

fit1<-glm(death~pm10+as.factor(year)+temp+temp60+temp75+modified.temp+modified.temp60+modified.temp75,da
fit2<-glm(death~pm10+as.factor(season)+temp+temp60+temp75+modified.temp+modified.temp60+modified.temp75
fit3<-glm(death~pm10+as.factor(month)+temp+temp60+temp75+modified.temp+modified.temp60+modified.temp75,

table[1,1:2]<-summary(fit1)$coefficient[2,1:2]
table[1,4]<-summary(fit1)$df[1]-1
table[1,5]<-summary(fit1)$deviance
table[1,6]<-summary(fit1)$aic

table[2,1:2]<-summary(fit2)$coefficient[2,1:2]
table[2,4]<-summary(fit2)$df[1]-1
table[2,5]<-summary(fit2)$deviance
table[2,6]<-summary(fit2)$aic


table[3,1:2]<-summary(fit3)$coefficient[2,1:2]
table[3,4]<-summary(fit3)$df[1]-1
table[3,5]<-summary(fit3)$deviance
table[3,6]<-summary(fit3)$aic

for( i in 1:3){
  table[i,3]<-paste(as.numeric(table[i,1])-1.96*as.numeric(table[i,2]),
                    as.numeric(table[i,1])+1.96*as.numeric(table[i,2]), collapse="-")
}
table
```

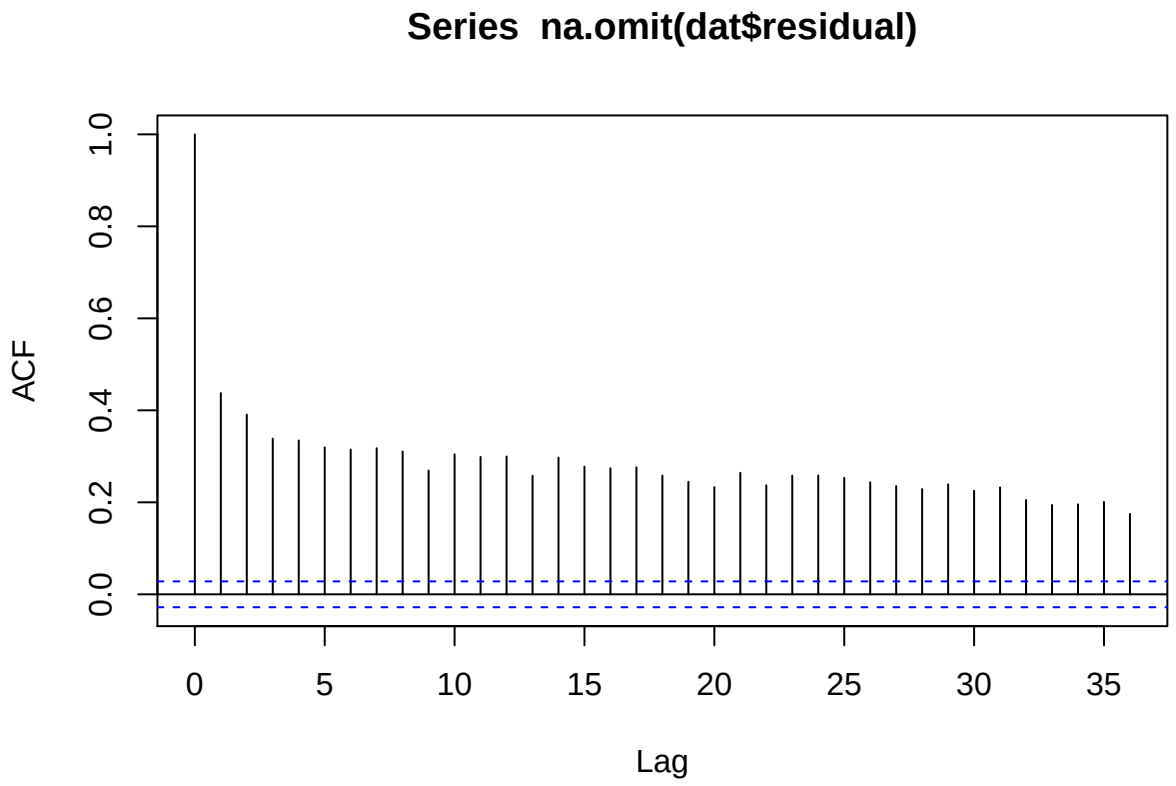
```
##         PM10 coef              Standard error
## Year   "0.000308681708124519" "8.08490594950916e-05"
## Season "0.000331627057560469" "8.08380921562627e-05"
## Month  "0.000316181893219149" "8.18453391078264e-05"
##        95% CI                                        Model df
## Year   "0.000150217551514139 0.000467145864734898" "20"
## Season "0.000173184396934194 0.000490069718186744" "10"
## Month  "0.000155765028567809 0.000476598757870489" "18"
##        Deviance           AIC
## Year   "6677.62026461159" "38713.587324815"
## Season "6941.17393376242" "38957.1409939658"
## Month  "6835.44293623116" "38867.4099964346"
```

```r
# pearson residual
dat$pred<-predict(fit3,dat)
dat$residual<-(dat$pred-dat$death)/sqrt(mean(dat$death))

ggplot(data=dat,aes(x=date,y=residual))+geom_point(alpha=0.1)
```

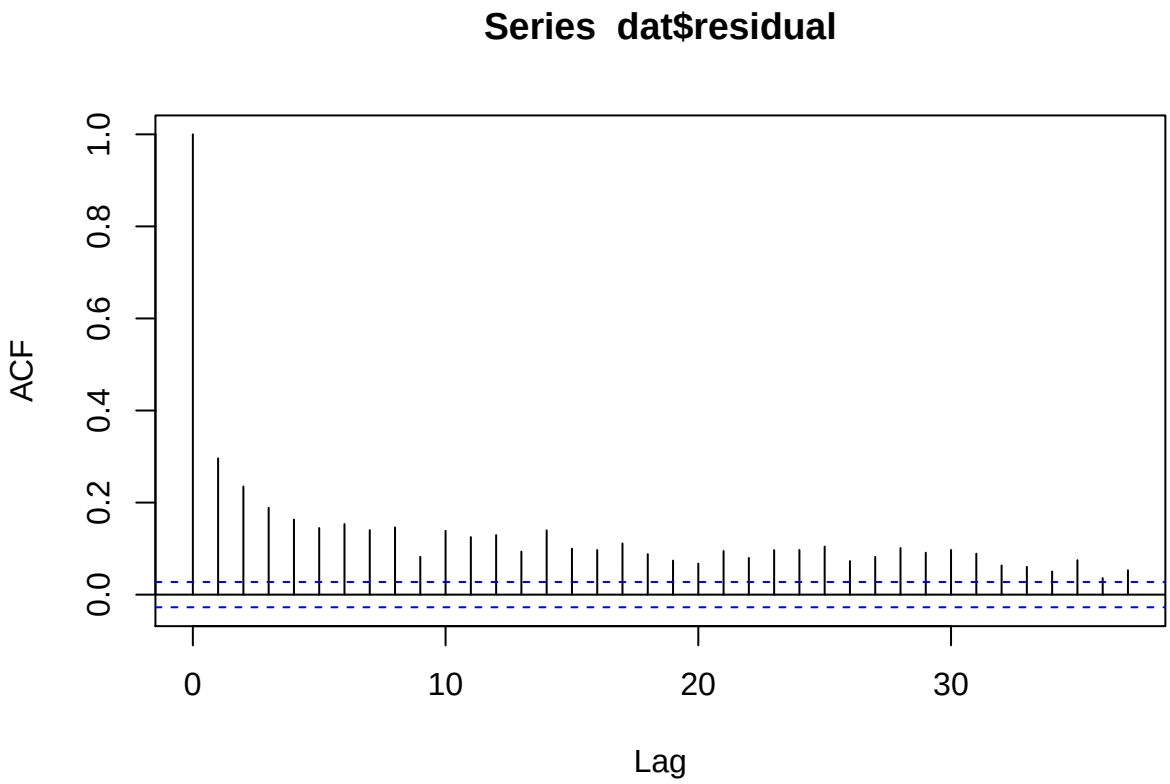
```
## Warning: Removed 251 rows containing missing values (geom_point).
```




```r
acf(na.omit(dat$residual))
```




```r
dat<-data.table(chicagoNMMAPS)

dat$month<-as.factor(dat$month)
detrend<-lm(death~month,data=dat)
dat$preddeath<-predict(detrend,dat)
dat$residual<-dat$preddeath-dat$death

# the trend still lasts if use month model
acf(dat$residual)
```



6 We have great finding! pm10 do affect motarility of humnan being at 99% significance level in our model. The positive coeficient with month, season,year model are all siginificent, suggest a larger pm10 level lead to more death per day. So action is needed.