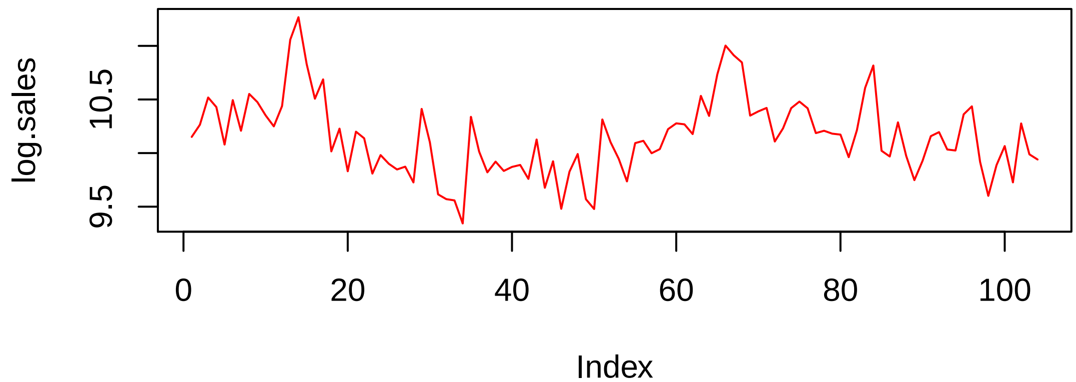
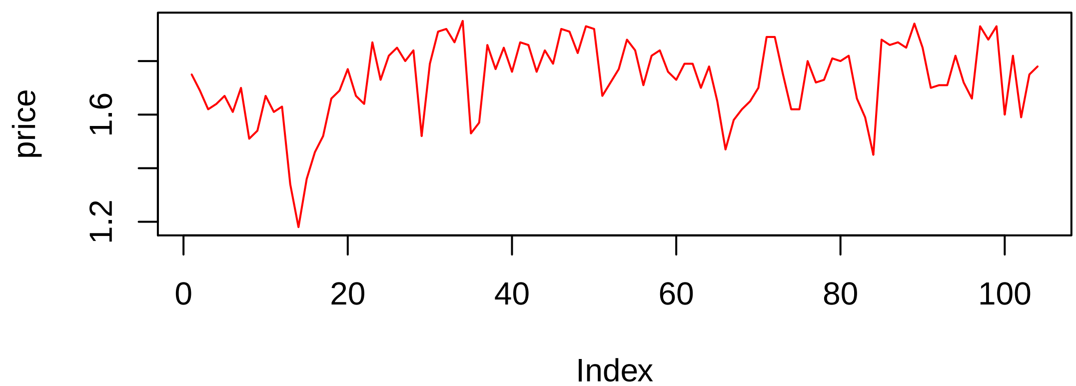
Homework-6

Tadesse Zemicheal

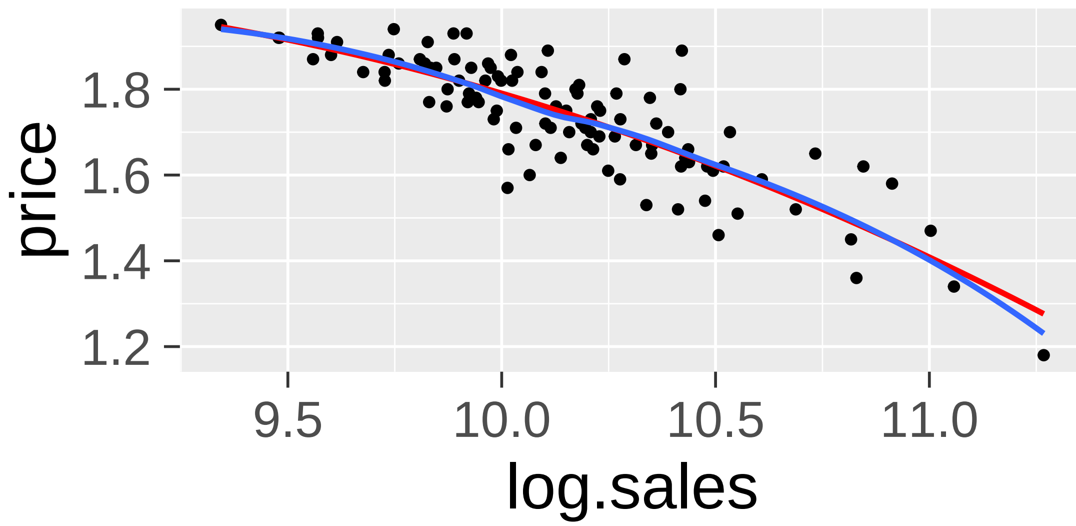
February 24, 2016

How does price relate to sales?

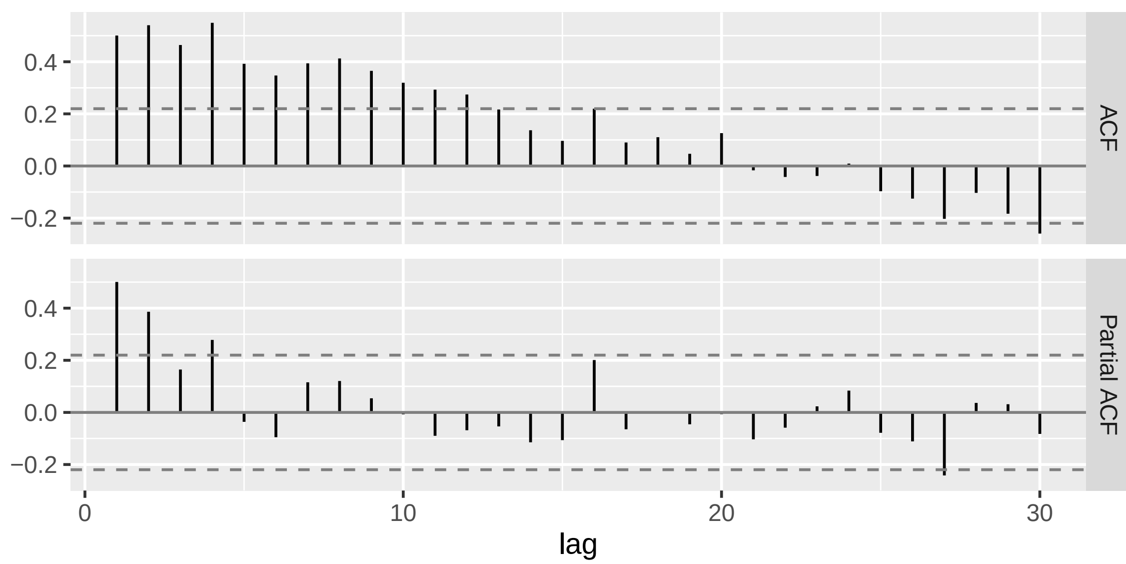




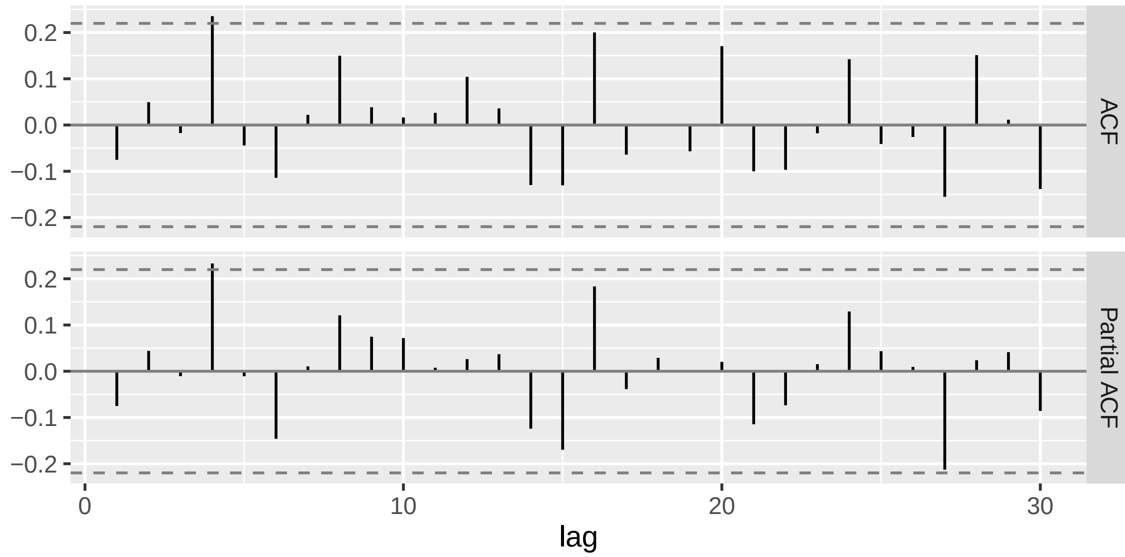
###



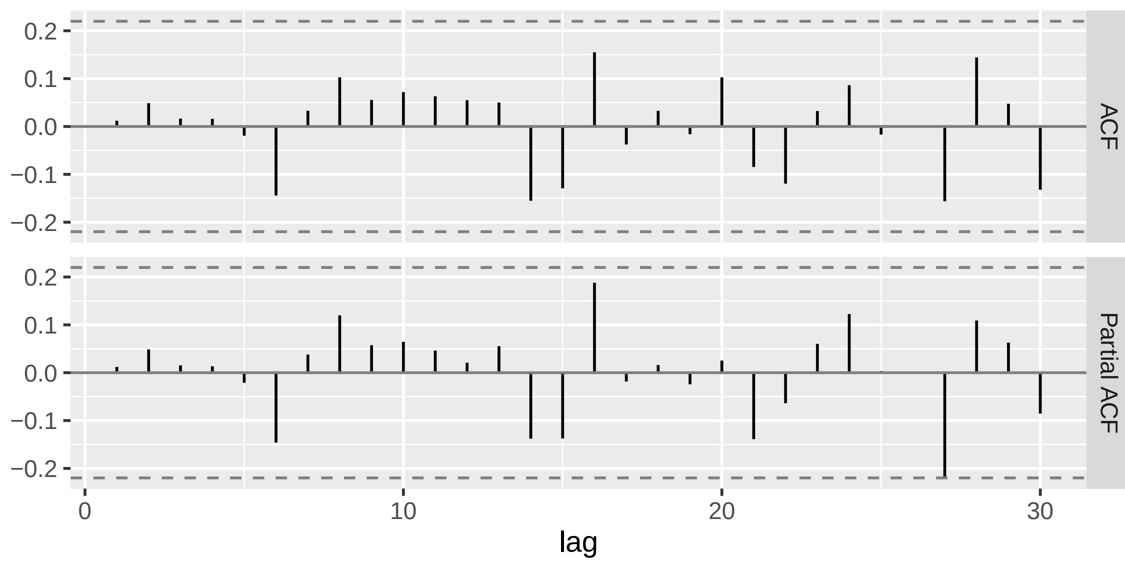
Fit an ols model
Fit model with no correlation

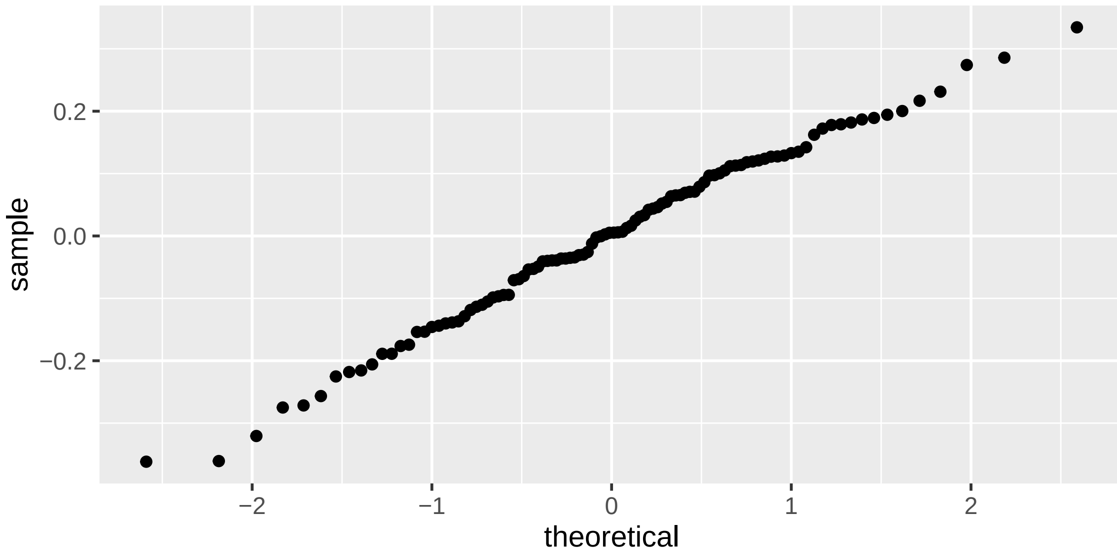


From the ACF and PACF graph the model fits good for arima(1,0,1) and it fits better for AR(4,0) as well. ACF of arima(1,0,1)



ACF of arima(4,0,0)



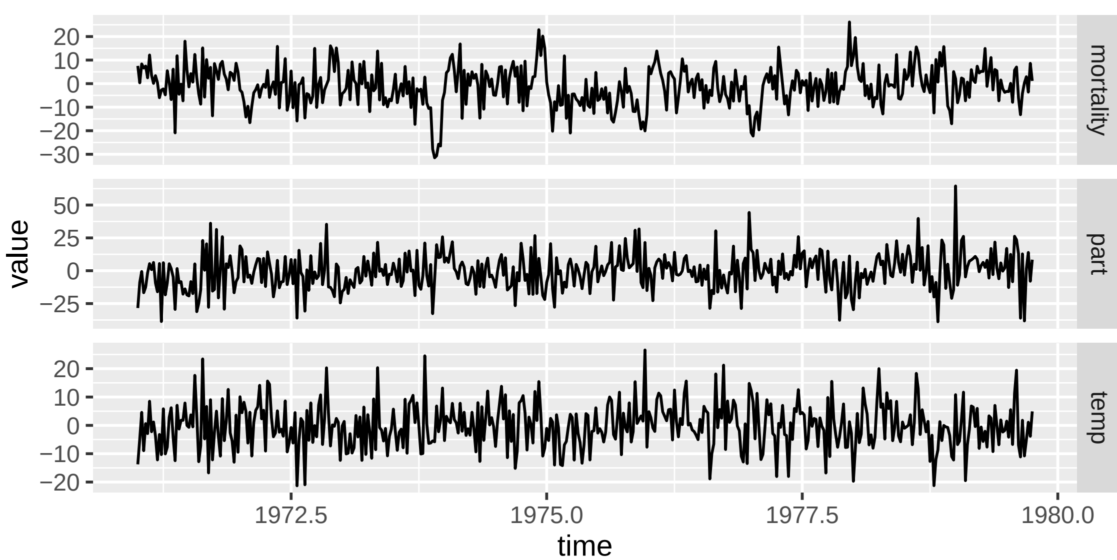


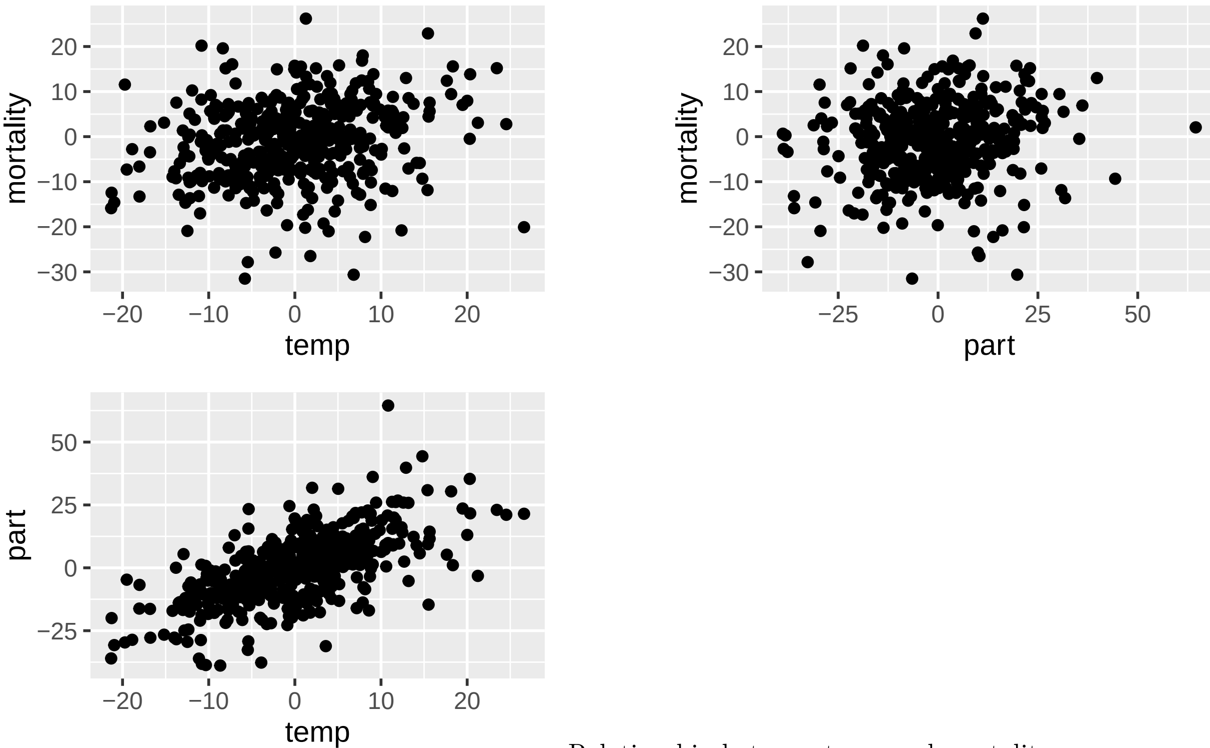
It is estimated that an increase in price of 10 cents is associated with decrease in median sales of 17.42 (95% CI 19.17 to `{r} round(ci[2], 2)`).

Q2 In class we looked a modelling the relationship between mortality, temperature and particulate matter. Repeat the analysis but seasonally difference all three series first. Compare the results.

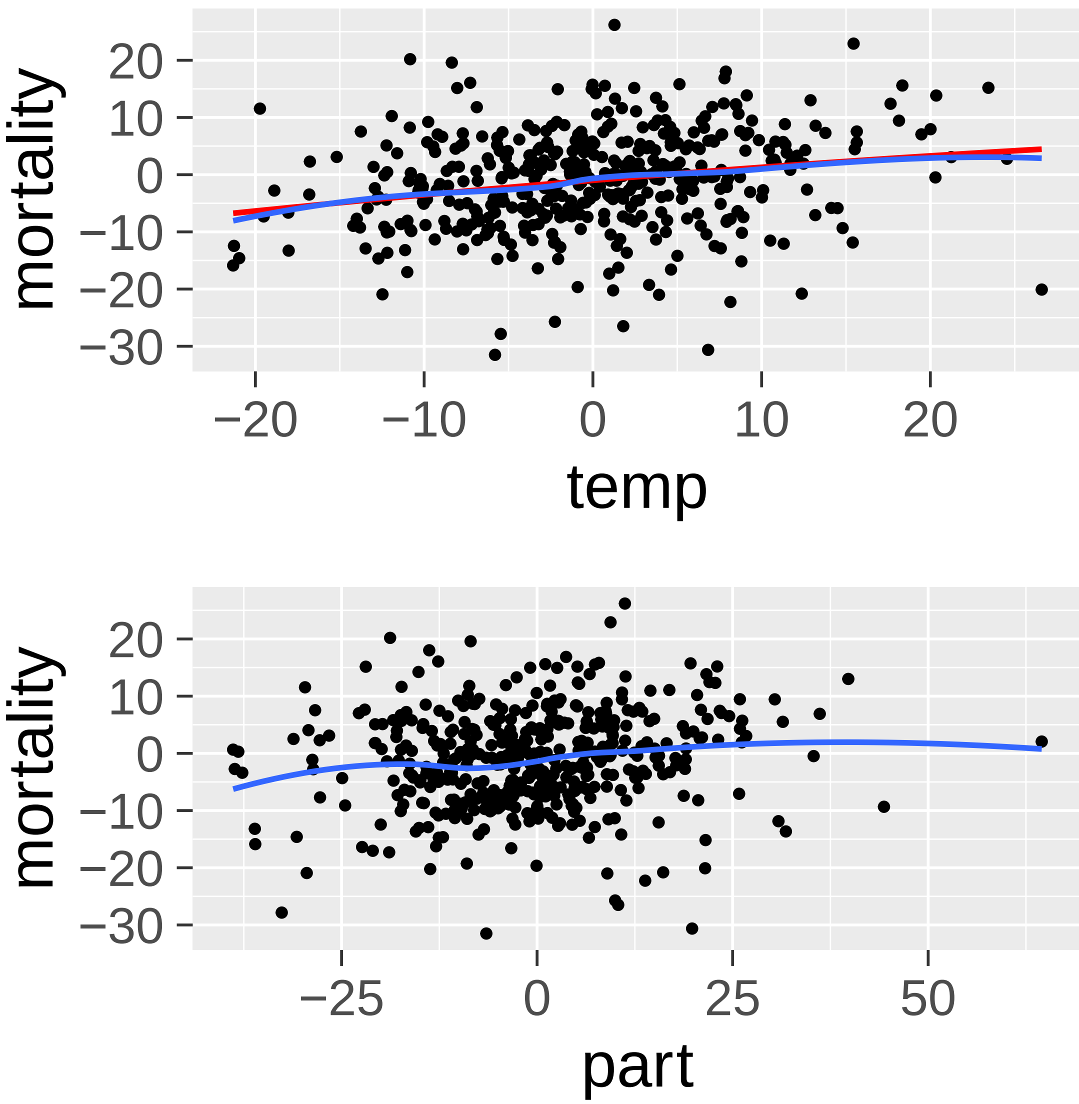
Remove seasonality with S=52

Seasonaliy remove series looks like.

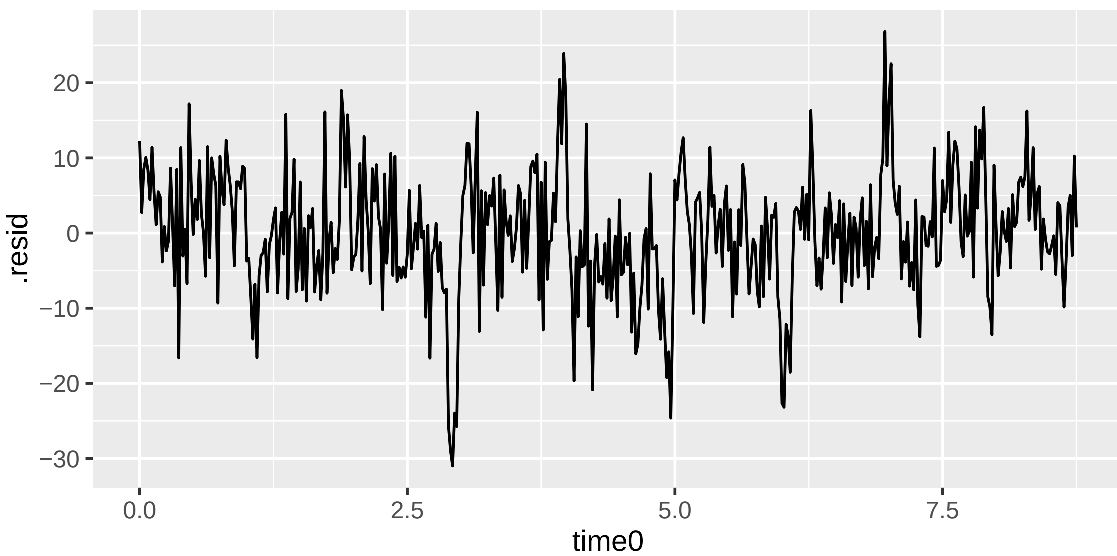




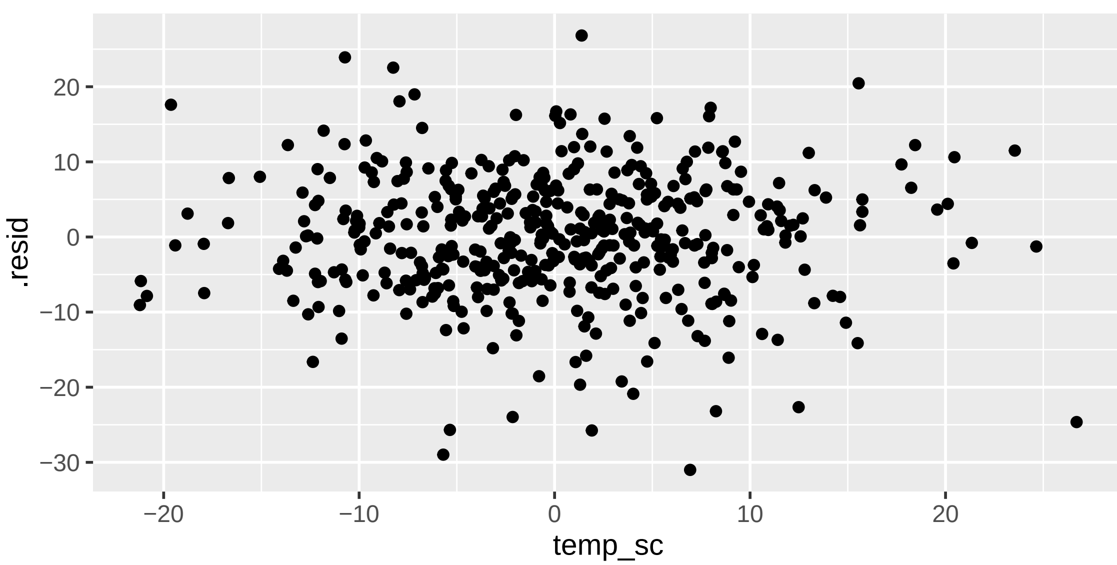
Relationship between temp and mortality



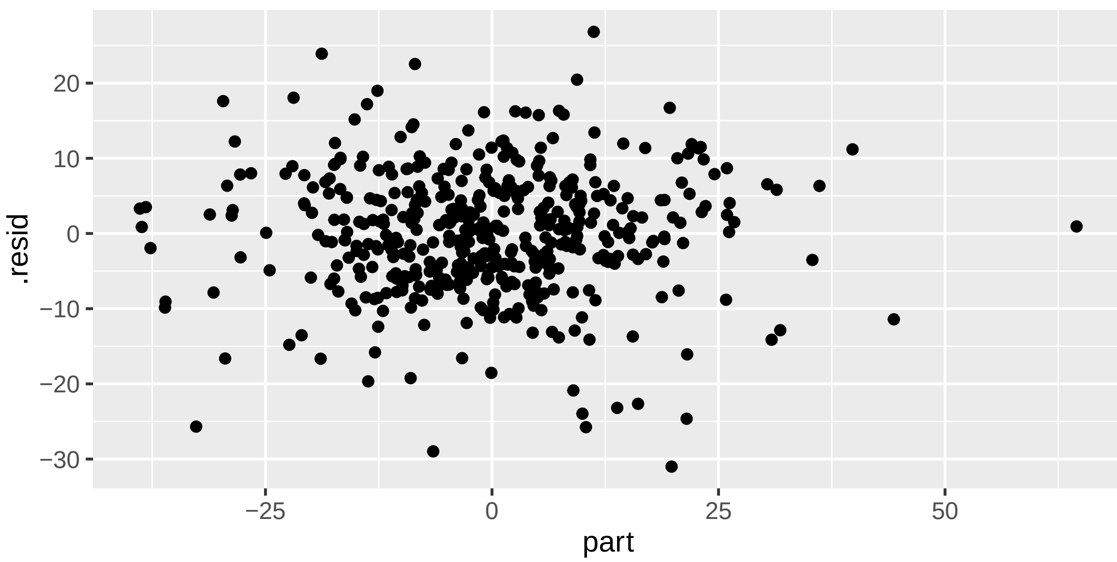
Residuals versus covariates analysis.



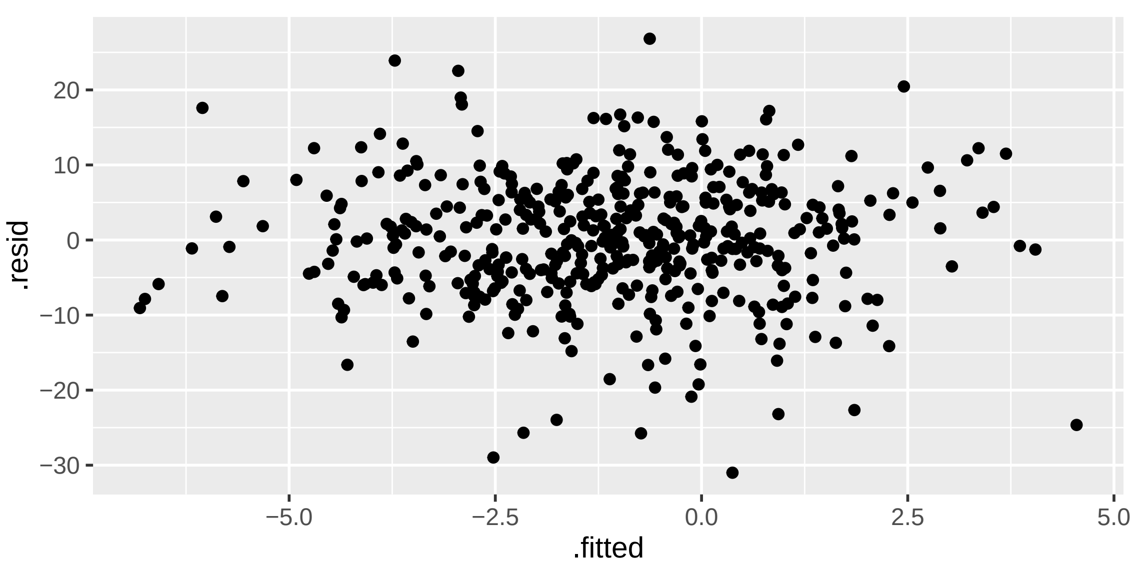
20
10
0
−10
−20
−30
.resid
0.0
2.5
5.0
7.5
time0

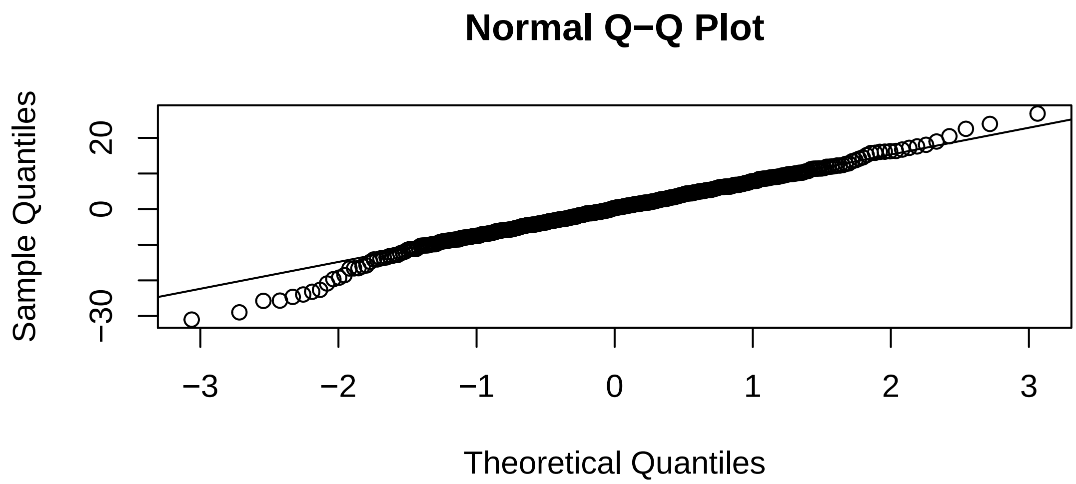


20
10
0
−10
−20
−30
.resid
−20
−10
0
10
20
temp_sc

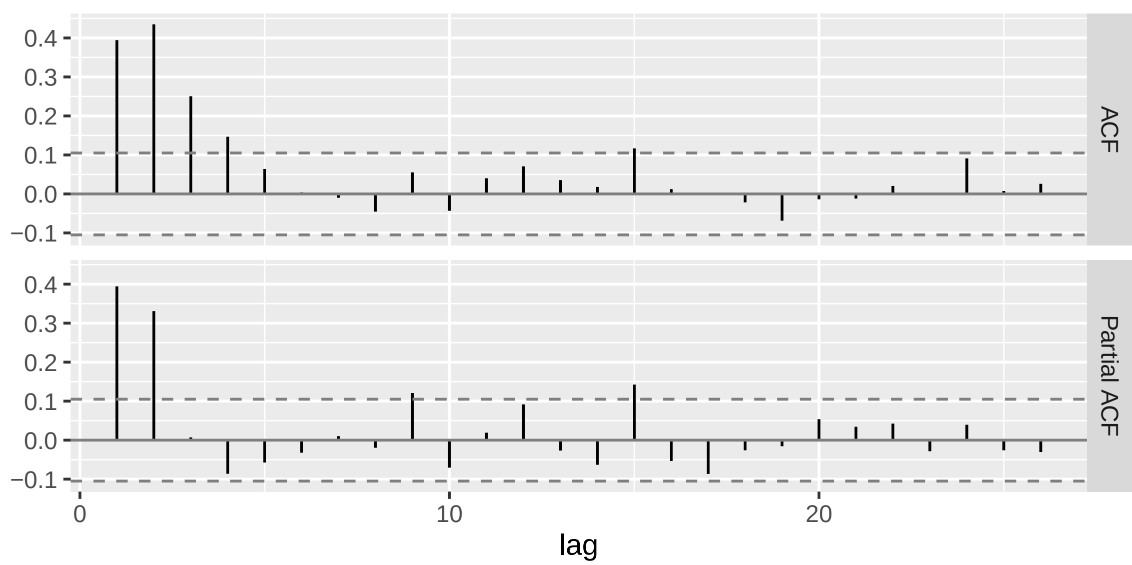


Assumption and Diagnosis

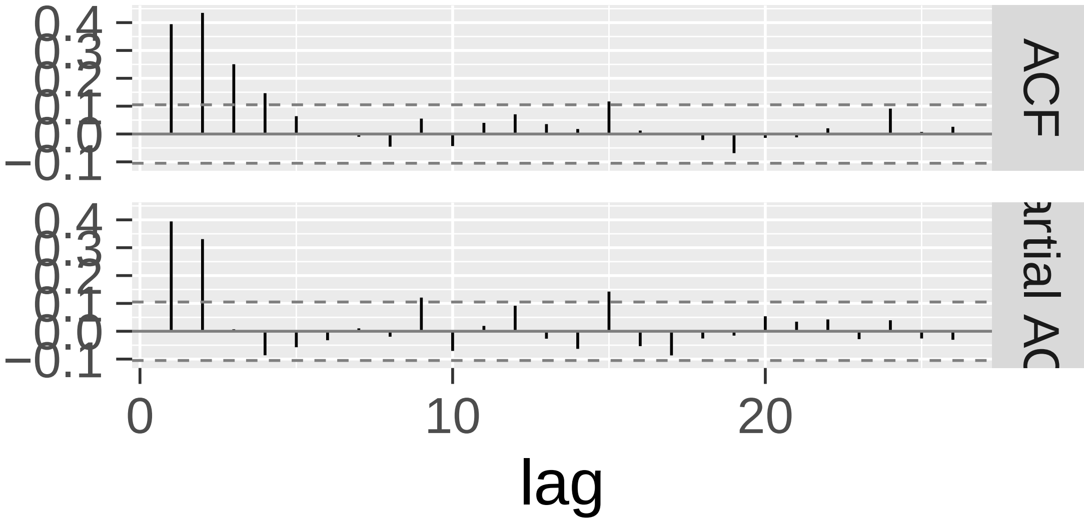




Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−30
0
20
−3
−2
−1
0
1
2
3

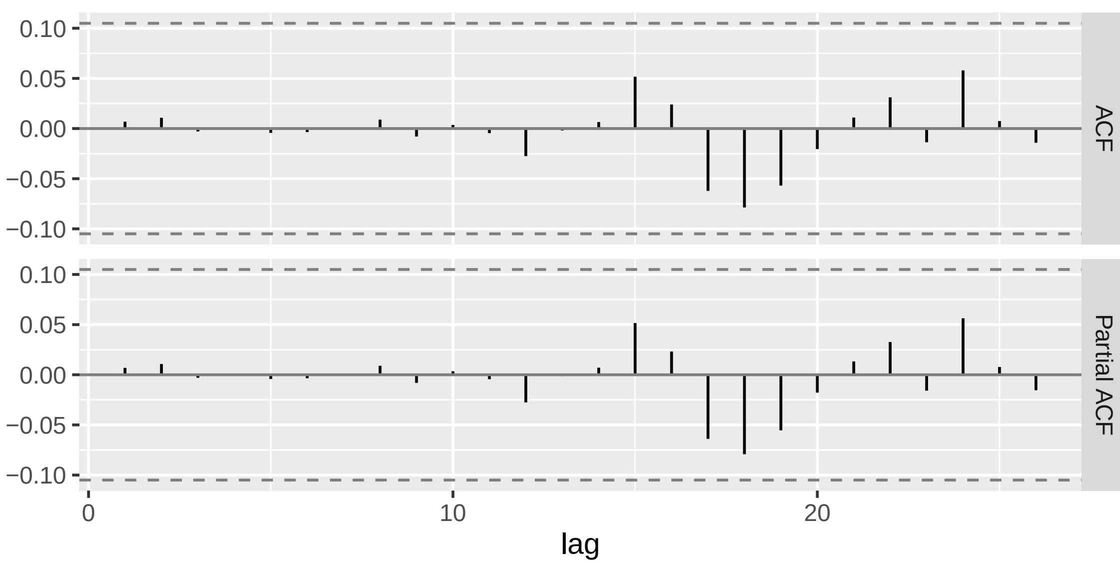


ACF
Partial ACF
lag
0.4
0.3
0.2
0.1
0.0
−0.1
0
10
20

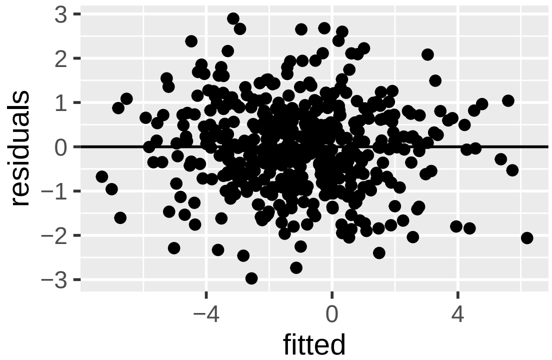


ACF and PACF shows there is a significant in the first 2 and at 15. The best model that fit is AR(15). Only the centered temp is important response variable.

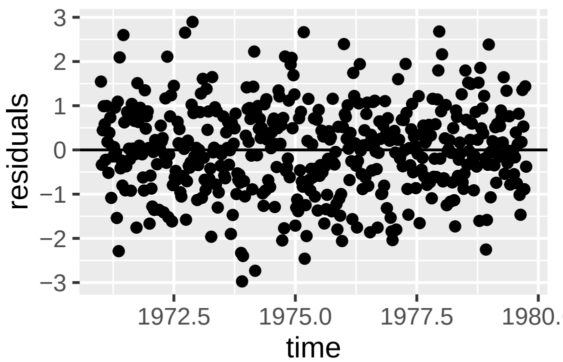
Diagnosis of the residual.

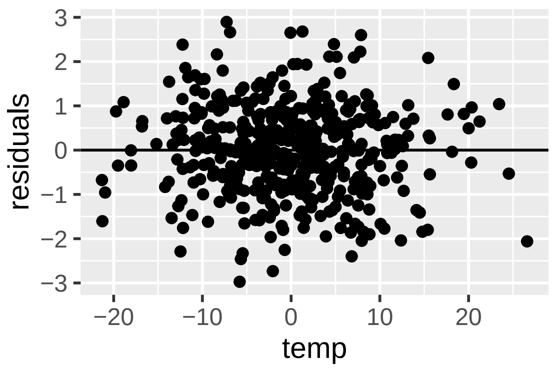


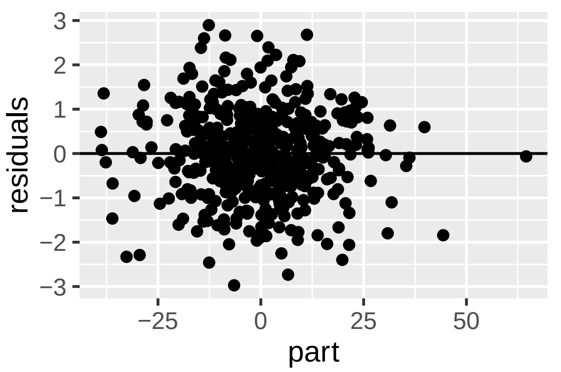
Looks



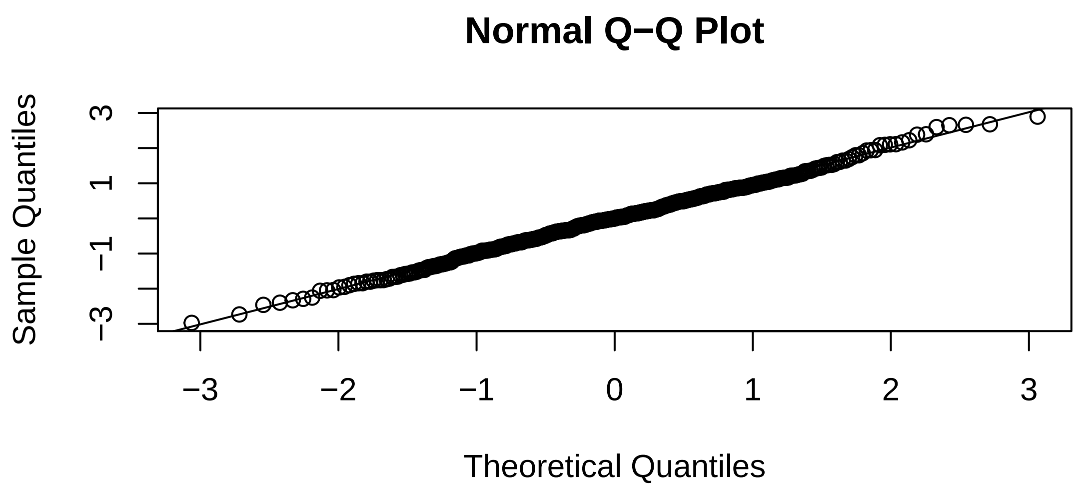
to fit good. The residual are now white noise.

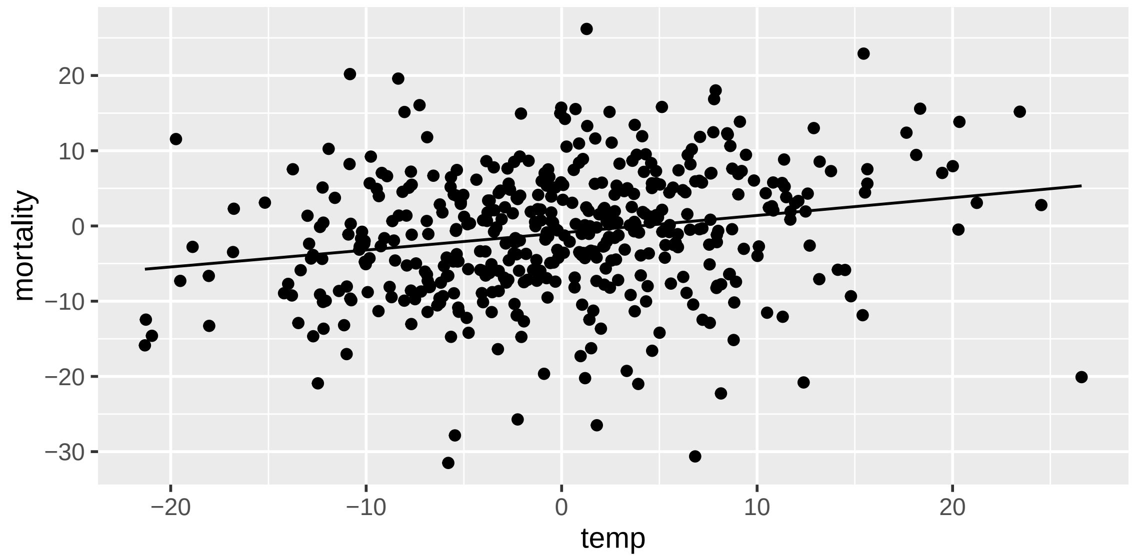






Forecast temp versus mortality





The analysis shows that removing seasonality by differencing with lag 52, results better fit for the model . Differencing get rid of the main correlation in the noise caused by time.

Appendix

```
knitr::opts_chunk$set(echo=FALSE, message = FALSE,
  warning = FALSE, results = "hide", fig.height = 3, fig.width = 6)
library(ggplot2)
library(dplyr)
library(tidyr)
library(TSA)
source(url("http://stat565.cwick.co.nz/code/fortify-ts.r"))
source(url("http://stat565.cwick.co.nz/code/get_acf.R")) # my code for examine_corr
big_font <- theme_grey(base_size =  24)
```

```r
data(bluebirdlite)
log.sales <- bluebirdlite[, 1]
price <- bluebirdlite[, 2]

plot(log.sales,type='l',col='red')
plot(price,type='l',col='red')
fit_lm <- lm(log.sales ~ price, data = bluebirdlite)
summary(fit_lm)
qplot(log.sales, price, data = bluebirdlite) +
  geom_smooth(method = "lm", formula = y ~ poly(x, 2), se = FALSE, colour = "red") +
  geom_smooth(se = FALSE)  +
  big_font
fit <- arima(log.sales, xreg = price, order = c(0,0,0))
examine_corr(residuals(fit),lag.max=30)
#it is okay
fit.2 <- arima(log.sales, xreg = price, order = c(1,0,1))
examine_corr(residuals(fit.2),lag.max=30)
#but AR(4,0) fits good
fit.4 <- arima(log.sales, xreg = price, order = c(4,0,0))
examine_corr(residuals(fit.4),lag.max=30)
qplot(sample = residuals(fit.4))
fit.4
coef(fit.4)["xreg"]
confint(fit.4)["xreg", ]

# backtransform and use % decrease with price increase of 0.1
(est <- 100*(1 - exp(0.1 * coef(fit.4)["xreg"])))
(ci <- 100*(1 - exp(0.1 * confint(fit.4)["xreg", ])))

library(ggplot2)
library(dplyr)
# install.packages("tidyr")
library(tidyr)
load(url("http://www.stat.pitt.edu/stoffer/tsa3/tsa3.rda"))
source(url("http://stat565.cwick.co.nz/code/fortify-ts.r"))
source(url("http://stat565.cwick.co.nz/code/get_acf.R")) # my code for examine_corr
big_font <- theme_grey(base_size =  24)

mort <- diff(cmort, lag = 52)
temp <- diff(tempr, lag = 52)
#Remove seasonality with S=52
part <- diff(part, lag = 52)

mort <- data.frame(mortality = mort, part = part, temp = temp)
mort$time <- fortify(cmort)$time[53:508]

# Seasonal deferenced plot
qplot(time, value, data = gather(mort, variable, value, -time), geom = "line") +
  facet_grid(variable ~ ., scale = "free")
#Checking the model
qplot(temp, mortality, data = mort)
qplot(part, mortality, data = mort)
qplot(temp, part, data = mort)
```

```r
qplot(temp, mortality, data = mort) +
  geom_smooth(method = "lm", formula = y ~ poly(x, 2), se = FALSE, colour = "red") +
  geom_smooth(se = FALSE)  +
  big_font

qplot(part, mortality, data = mort) + geom_smooth(se = FALSE) +
  big_font

mort <- mutate(mort, temp_sc = temp - mean(temp),
               temp_2 =temp_sc^2,
               time0 = time - min(time))

fit_lm <- lm(mortality ~ time0 + temp_sc + temp_2 + part, data = mort, na.action = na.omit)
summary(fit_lm)
mort_lm <- fortify(fit_lm)
qplot(time0, .resid, data = mort_lm, geom= "line")
qplot(temp_sc, .resid, data = mort_lm)
qplot(part, .resid, data = mort_lm)
# residuals versus fitted
qplot(.fitted, .resid, data = mort_lm)

# normality of residuals
qqnorm(mort_lm$.resid)
qqline(mort_lm$.resid)
# correlation of residuals
examine_corr(residuals(fit_lm))
last_plot() + big_font
# AR (2)? violates regression assumptions
# two ways to fit
library(nlme)
gls_fit <- gls(mortality ~ time0 + temp_sc + temp_2 + part, data = mort,
    correlation = corARMA(p = 15), method = "ML")
summary(gls_fit)

# or
# auto.arima(mort$mortality)
mort$residuals <- residuals(gls_fit, type = "normalized")
mort$fitted <- fitted(gls_fit)
examine_corr(mort$residuals)
qplot(fitted, residuals, data = mort) + geom_hline(yintercept = 0)
qplot(time, residuals, data = mort) + geom_hline(yintercept = 0)
qplot(temp, residuals, data = mort) + geom_hline(yintercept = 0)
qplot(part, residuals, data = mort) + geom_hline(yintercept = 0)
qqnorm(mort$residuals)
qqline(mort$residuals)
# plot prediction on temp versus mortality plot
mort$pred <- predict(gls_fit, newdata =
  data.frame(time0 = 0, temp_sc = mort$temp_sc,
             temp_2 = mort$temp_2, part = mean(mort$part)))

qplot(temp, mortality, data = mort) +
  geom_line(aes(y = pred))
```