chapter\_2\_R\_intro.R

RYU

Tue Nov 13 09:17:35 2018

library(lubridate)

##   
## Attaching package: 'lubridate'

## The following object is masked from 'package:base':  
##   
## date

library(ggplot2)  
library(astsa)  
library(scales)  
library(car)

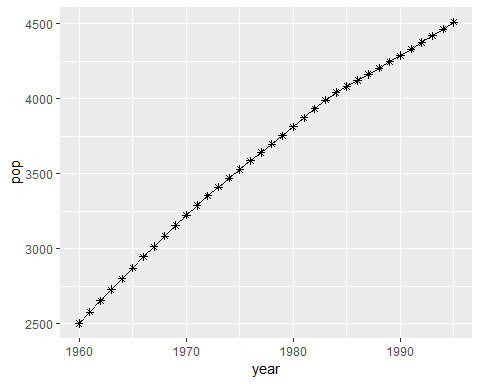
## Loading required package: carData

library(forecast)

##   
## Attaching package: 'forecast'

## The following object is masked from 'package:astsa':  
##   
## gas

# example 2.1  
data <- read.csv('../timedata/population.txt', sep='', header=FALSE)  
pop <- na.omit(c(t(data)))  
pop <- round(pop/10000)  
lnpop <- log(pop)  
t <- 1:length(pop)  
t2 <- t\*\*2  
year <- 1959 + t  
df1 <- data.frame(pop, lnpop, t, t2, year)  
ggplot(data=df1, aes(year, pop)) +   
 geom\_line() +   
 geom\_point(shape=8, color='black')



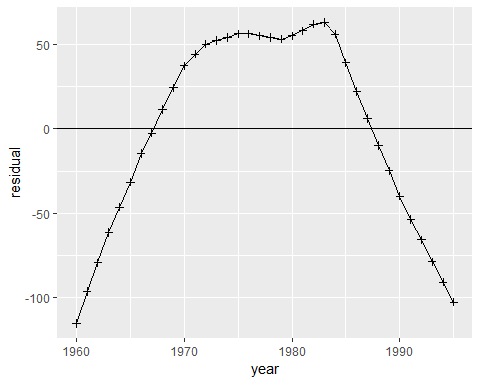
regmodel <- lm(pop~t, data=df1)  
anova(regmodel)

## Analysis of Variance Table  
##   
## Response: pop  
## Df Sum Sq Mean Sq F value Pr(>F)   
## t 1 12628351 12628351 3644.2 < 2.2e-16 \*\*\*  
## Residuals 34 117822 3465   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

summary(regmodel)

##   
## Call:  
## lm(formula = pop ~ t, data = df1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -115.40 -48.30 16.87 54.37 63.29   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2559.3889 20.0385 127.72 <2e-16 \*\*\*  
## t 57.0135 0.9444 60.37 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 58.87 on 34 degrees of freedom  
## Multiple R-squared: 0.9908, Adjusted R-squared: 0.9905   
## F-statistic: 3644 on 1 and 34 DF, p-value: < 2.2e-16

df2 <- data.frame(year, regmodel$fitted.values, regmodel$residuals)  
colnames(df2) <- c("year", "pred", "residual")  
ggplot(data=df2, aes(year, residual)) +   
 geom\_line() +   
 geom\_point(shape=3, color='black') +   
 geom\_hline(yintercept=0)



regmodel <- lm(pop~t+t2, data=df1)  
anova(regmodel)

## Analysis of Variance Table  
##   
## Response: pop  
## Df Sum Sq Mean Sq F value Pr(>F)   
## t 1 12628351 12628351 214681 < 2.2e-16 \*\*\*  
## t2 1 115880 115880 1970 < 2.2e-16 \*\*\*  
## Residuals 33 1941 59   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

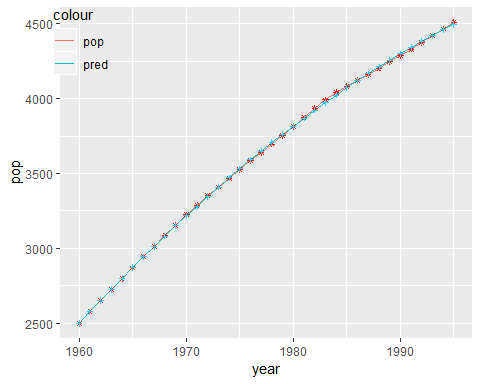
summary(regmodel)

##   
## Call:  
## lm(formula = pop ~ t + t2, data = df1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -11.365 -4.779 -1.049 3.798 17.631   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2421.49090 4.05820 596.69 <2e-16 \*\*\*  
## t 78.78688 0.50576 155.78 <2e-16 \*\*\*  
## t2 -0.58847 0.01326 -44.38 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 7.67 on 33 degrees of freedom  
## Multiple R-squared: 0.9998, Adjusted R-squared: 0.9998   
## F-statistic: 1.083e+05 on 2 and 33 DF, p-value: < 2.2e-16

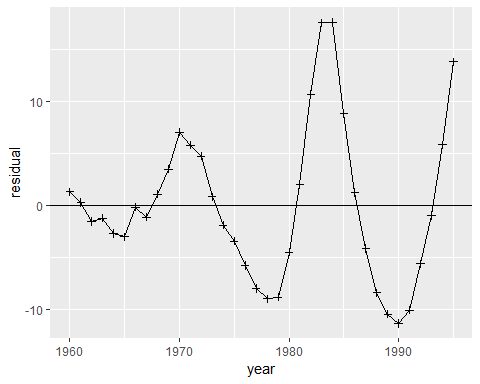
durbinWatsonTest(regmodel)

## lag Autocorrelation D-W Statistic p-value  
## 1 0.7948211 0.3108307 0  
## Alternative hypothesis: rho != 0

df3 <- data.frame(year, pop, regmodel$fitted.values, regmodel$residuals)  
colnames(df3) <- c("year", "pop", "pred", "residual")  
ggplot(data=df3, aes(x=year)) +   
 geom\_line(aes(y=pop, colour='pop')) +   
 geom\_line(aes(y=pred, colour='pred')) +  
 geom\_point(data=df3, aes(y=pop), shape=8, size=1, color='#ff0000') +   
 geom\_point(data=df3, aes(y=pred), shape=3, size=1, color='#33ccff') +   
 theme(legend.position = c(0.05,0.9), legend.background=element\_rect(fill="transparent"))



ggplot(data=df3, aes(year, residual)) +   
 geom\_line() +   
 geom\_point(shape=3, color='black') +   
 geom\_hline(yintercept=0)



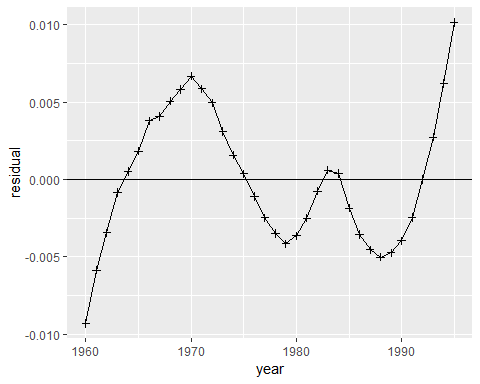
regmodel <- lm(lnpop~t+t2, data=df1)  
anova(regmodel)

## Analysis of Variance Table  
##   
## Response: lnpop  
## Df Sum Sq Mean Sq F value Pr(>F)   
## t 1 1.03011 1.03011 51760.9 < 2.2e-16 \*\*\*  
## t2 1 0.03020 0.03020 1517.5 < 2.2e-16 \*\*\*  
## Residuals 33 0.00066 0.00002   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

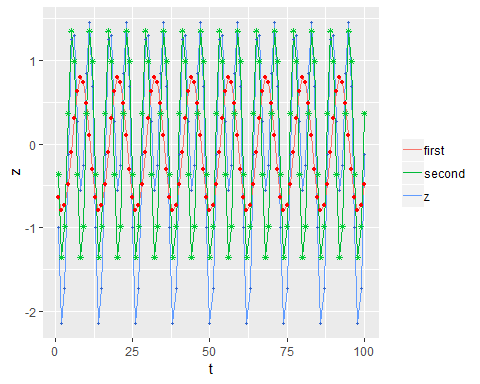
summary(regmodel)

##   
## Call:  
## lm(formula = lnpop ~ t + t2, data = df1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.009306 -0.003520 -0.000374 0.003284 0.010159   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 7.807e+00 2.360e-03 3307.25 <2e-16 \*\*\*  
## t 2.740e-02 2.942e-04 93.14 <2e-16 \*\*\*  
## t2 -3.004e-04 7.712e-06 -38.95 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.004461 on 33 degrees of freedom  
## Multiple R-squared: 0.9994, Adjusted R-squared: 0.9993   
## F-statistic: 2.664e+04 on 2 and 33 DF, p-value: < 2.2e-16

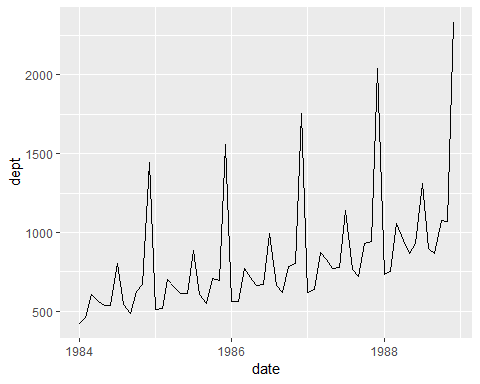
df4 <- data.frame(year, regmodel$residuals)  
colnames(df4) <- c("year", "residual")  
ggplot(data=df4, aes(year, residual)) +   
 geom\_line() +   
 geom\_point(shape=3, color='black') +   
 geom\_hline(yintercept=0)



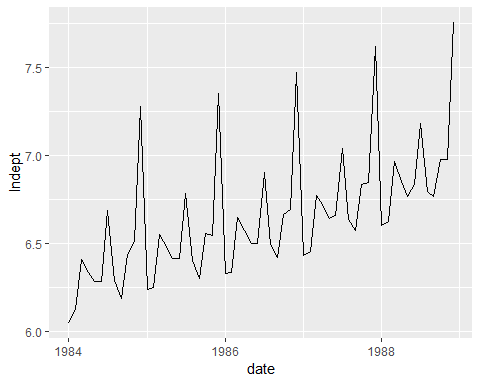
# figure 2.8  
t <- 1:100  
a1 <- -0.8  
a2 <- 1.4  
phi1 <- pi/8  
phi2 <- 3\*pi/4  
first <- a1 \* sin(pi\*t/6 + phi1)  
second <- a2 \* sin(pi\*t/3 + phi2)  
z <- first + second  
df <- data.frame(t, z, first, second)  
ggplot(data=df, aes(t)) +   
 geom\_line(aes(y=z, colour='z')) +   
 geom\_line(aes(y=first, colour='first')) +  
 geom\_line(aes(y=second, colour='second')) +  
 geom\_point(data=df, aes(y=z), shape=20, size=1, color='#3366cc') +   
 geom\_point(data=df, aes(y=first), shape=16, size=1, color='#ff0000') +   
 geom\_point(data=df, aes(y=second), shape=8, size=1, color='#00cc33') +  
 theme(legend.title=element\_blank())



# example 2.2 & 2.4  
data <- read.csv('../timedata/depart.txt', sep='', header=FALSE)  
dept <- na.omit(c(t(data)))  
t <- 1:length(dept)  
lndept <- log(dept)  
date <- ymd("840101") + months(1:length(dept)-1)  
mon <- month(date)  
i1 <- as.integer(mon==1)  
i2 <- as.integer(mon==2)  
i3 <- as.integer(mon==3)  
i4 <- as.integer(mon==4)  
i5 <- as.integer(mon==5)  
i6 <- as.integer(mon==6)  
i7 <- as.integer(mon==7)  
i8 <- as.integer(mon==8)  
i9 <- as.integer(mon==9)  
i10 <- as.integer(mon==10)  
i11 <- as.integer(mon==11)  
i12 <- as.integer(mon==12)  
df1 <- data.frame(date, dept, lndept, i1, i2, i3, i4, i5, i6, i7, i8, i9, i10, i11, i12)  
  
ggplot(data=df1, aes(date, dept)) +  
 geom\_line()



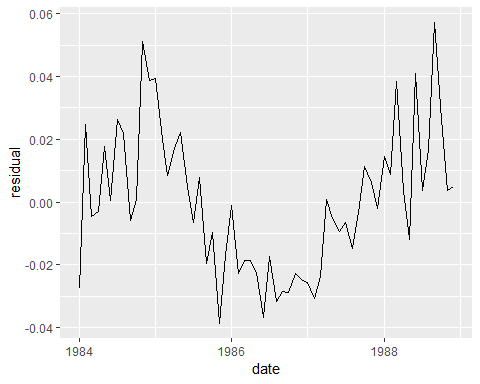
ggplot(data=df1, aes(date, lndept)) +  
 geom\_line()



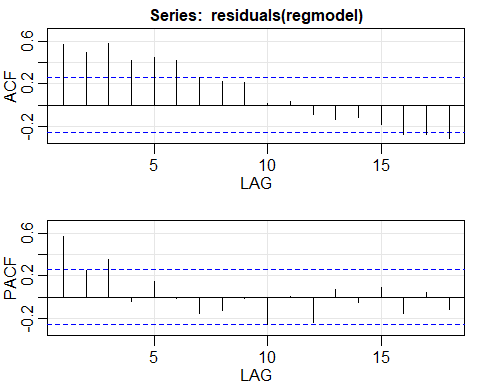
regmodel <- lm(lndept~t+i1+i2+i3+i4+i5+i6+i7+i8+i9+i10+i11+i12+0, data=df1)  
regmodel$coefficients

## t i1 i2 i3 i4 i5   
## 0.01066028 6.06419041 6.08079954 6.38111834 6.29534554 6.21323923   
## i6 i7 i8 i9 i10 i11   
## 6.21977708 6.58850653 6.18428310 6.10011484 6.33345051 6.34171162   
## i12   
## 7.11048162

df2 <- data.frame(date, residual=as.numeric(regmodel$residuals))  
  
ggplot(data=df2, aes(date, residual)) +  
 geom\_line()

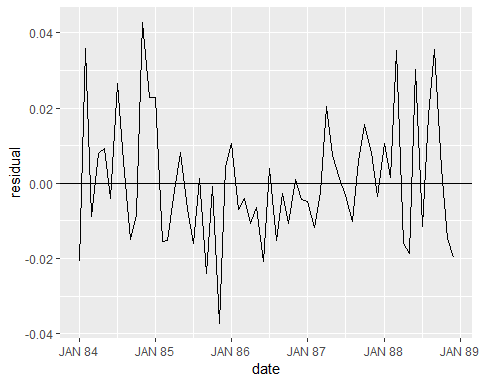


acf2(residuals(regmodel))

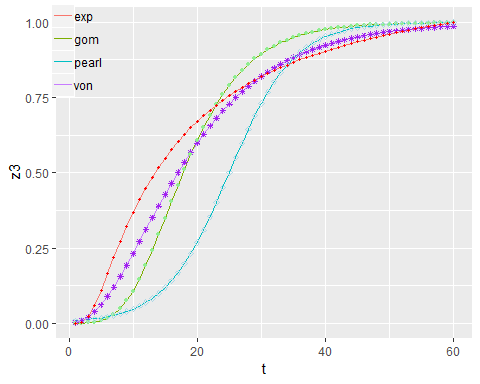


## ACF PACF  
## [1,] 0.57 0.57  
## [2,] 0.50 0.25  
## [3,] 0.58 0.35  
## [4,] 0.42 -0.04  
## [5,] 0.45 0.15  
## [6,] 0.42 -0.01  
## [7,] 0.26 -0.15  
## [8,] 0.22 -0.12  
## [9,] 0.21 -0.01  
## [10,] 0.02 -0.26  
## [11,] 0.04 0.00  
## [12,] -0.09 -0.24  
## [13,] -0.14 0.07  
## [14,] -0.12 -0.05  
## [15,] -0.19 0.09  
## [16,] -0.28 -0.16  
## [17,] -0.28 0.04  
## [18,] -0.32 -0.12

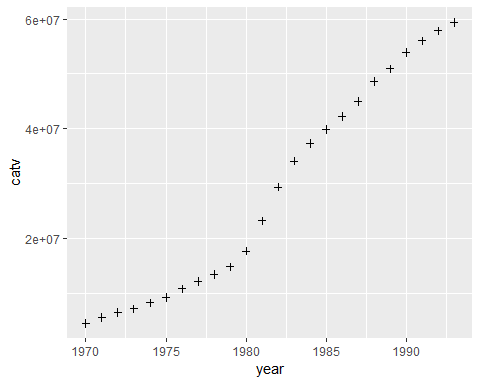
ar3res <- arima(residuals(regmodel),order=c(3,0,0))  
  
df3 <- data.frame(date, residual=as.numeric(ar3res$residuals))  
ggplot(data=df3, aes(date, residual)) +  
 geom\_line() +  
 geom\_hline(yintercept=0) +  
 scale\_x\_date(date\_breaks = "1 year",  
 labels = date\_format("JAN %y"))



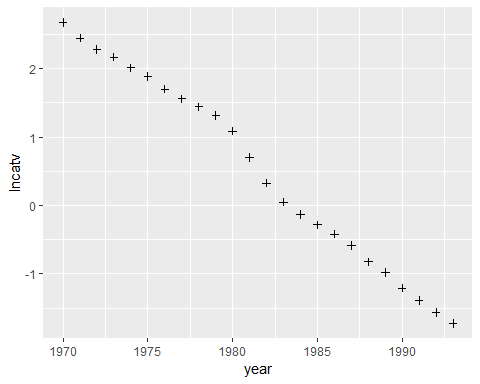
# figure 2.12  
t <- 1:60  
  
b0 <- 0.2  
b1 <- -12  
z1 <- exp(b0 + b1/t)  
  
b0 <- 10  
b1 <- 0.15  
k <- 1  
z2 <- k\*exp(-b0\*exp(-b1\*t))  
  
b0 <- 0.95  
b1 <- 0.09  
z3 <- (1-b0\*exp((-b1)\*t))\*\*3  
  
b0 <- 5  
b1 <- -0.2  
z4 <- k/(1+exp(b0+b1\*t))  
  
df <- data.frame(t, z1, z2, z3, z4)  
  
ggplot(data=df, aes(x=t)) +  
 geom\_line(aes(y=z3, colour='von')) +  
 geom\_line(aes(y=z2, colour='gom')) +  
 geom\_line(aes(y=z4, colour='pearl')) +  
 geom\_line(aes(y=z1, colour='exp')) +  
 geom\_point(data=df, aes(y=z3), shape=8, size=1, color='purple') +   
 geom\_point(data=df, aes(y=z2), shape=16, size=1, color='lightgreen') +   
 geom\_point(data=df, aes(y=z4), shape=5, size=1, color='lightblue') +  
 geom\_point(data=df, aes(y=z1), shape=20, size=1, color='red') +  
 theme(legend.position = c(0.05,0.87), legend.background=element\_rect(fill="transparent"), legend.title=element\_blank())



# example 2.3  
data <- read.csv('../timedata/catv.txt', sep='', header=FALSE)  
catv <- na.omit(c(t(data)))  
t <- 1:length(catv)  
year <- 1969 + t  
k <- 70000000  
lncatv <- log(k/catv-1)  
df1 <- data.frame(year, catv, lncatv)  
  
ggplot(data=df1, aes(year, catv)) +  
 geom\_point(shape=3)



ggplot(data=df1, aes(year, lncatv)) +  
 geom\_point(shape=3)



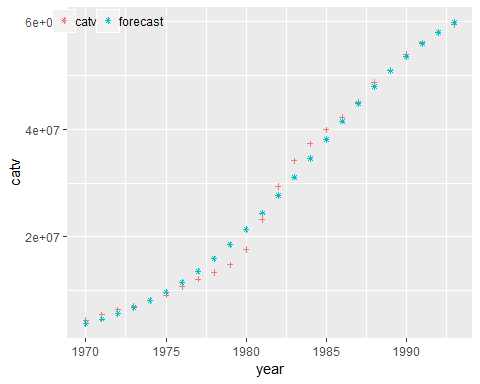
regmodel <- lm(lncatv~year, data=df1)  
anova(regmodel)

## Analysis of Variance Table  
##   
## Response: lncatv  
## Df Sum Sq Mean Sq F value Pr(>F)   
## year 1 45.708 45.708 2563 < 2.2e-16 \*\*\*  
## Residuals 22 0.392 0.018   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

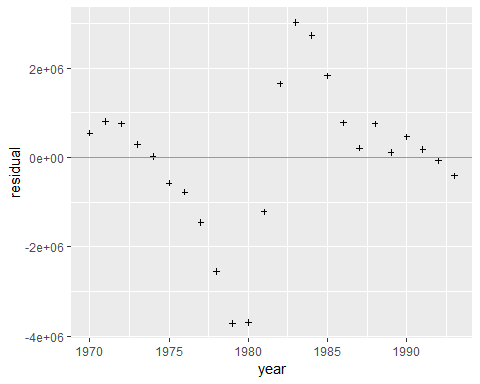
summary(regmodel)

##   
## Call:  
## lm(formula = lncatv ~ year, data = df1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.17388 -0.09974 -0.01448 0.07133 0.29312   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 395.562792 7.803147 50.69 <2e-16 \*\*\*  
## year -0.199364 0.003938 -50.63 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1335 on 22 degrees of freedom  
## Multiple R-squared: 0.9915, Adjusted R-squared: 0.9911   
## F-statistic: 2563 on 1 and 22 DF, p-value: < 2.2e-16

p1 <- k/(exp(regmodel$fitted.values)+1)  
residual <- catv-p1  
df2 <- data.frame(year, regmodel$fitted.values, p1, residual)  
colnames(df2) <- c('year', 'pred', 'p1', 'residual')  
  
ggplot(data=df2, aes(x=year)) +   
 geom\_point(data=df2, aes(y=catv, colour='catv'), shape=3, size=1) +  
 geom\_point(data=df2, aes(y=p1, colour='forecast'), shape=8, size=1) +  
 theme(legend.position = c(0.1, 0.96),   
 legend.background=element\_rect(fill="transparent"),   
 legend.title=element\_blank(),  
 legend.direction = "horizontal")



ggplot(data=df2, aes(x=year, y=residual)) +  
 geom\_point(shape=3, size=1) +  
 geom\_hline(yintercept=0, color="#999999")



# trend model prediction  
z <- c(23,25,27,34,38,47,49,39,57,59,63,64,69,78,73,89,83,84,86,92)  
t <- 1:length(z)  
regmodel <- lm(z~t)  
new\_t <- 1:32  
new <- data.frame(t=new\_t)  
df <- data.frame(predict(lm(z~t), new, interval="prediction", level=0.95))  
df['t'] <- new\_t  
df['z'] <- c(z, rep('NA', 12))  
  
ggplot(data=df, aes(x=t)) +   
 geom\_line(aes(y=fit, colour='prediction')) +   
 geom\_line(aes(y=lwr, colour='lower')) +   
 geom\_line(aes(y=upr, colour='upper')) +   
 geom\_line(aes(x=t,y=as.numeric(z), colour='z')) +  
 geom\_point(aes(y=fit), shape='p') +   
 geom\_point(aes(y=lwr), shape='L') +  
 geom\_point(aes(y=upr), shape='U') +  
 geom\_point(aes(y=as.numeric(z)), shape=20) +  
 geom\_vline(xintercept=21, color='#999999') +  
 theme(legend.position = c(0.2, 0.96),   
 legend.background=element\_rect(fill="transparent"),   
 legend.title=element\_blank(),  
 legend.direction = "horizontal")

## Warning in FUN(X[[i]], ...): 강제형변환에 의해 생성된 NA 입니다  
  
## Warning in FUN(X[[i]], ...): 강제형변환에 의해 생성된 NA 입니다

## Warning: Removed 12 rows containing missing values (geom\_path).

## Warning: Removed 12 rows containing missing values (geom\_point).

