

예제 데이터 1

http://wolpack.hnu.ac.kr/iBOOKs/about_R/power_demand.csv

2011년 1월 1일부터 13년 12월 31일, 3년간 전력수요

“PEAK_HOUR”-최대수요시각 “REAL_LOAD”-전력수요 “AVG_TMPR”-평균온도
 “MIN_TMPR”-최저기온 “MAX_TMPR”-최고기온 “MAX_WND_SPD” -최대풍속
 “AVG_WND_SPD”-평균풍속 “RAIN_AMT”-강수량 “AVG_HMDT” -평균습도
 “CLD_AMT”-구름량 “SUN_HOUR”-일조량 “HOLIDAY_YN”-휴일여부

PEAK_HO	REAL_LO	AVG_TMPR	MIN_TMPR	MAX_TMPR	MAX_WNI	AVG_WNI	RAIN_AM	AVG_HMDT	CLD_AMT	SUN_HOU	HOLIDAY_YN
1	59796	-4.89	-9.174	-0.05	4.958	2.276	0	56.902	2.457	7.701	Y
24	59173	-2.986	-7.223	2.674	4.292	1.969	0	60.416	0.103	8.803	N
12	71768	-2.682	-7.027	1.591	4.528	2.17	0	65.038	4.689	5.652	N
12	71697	-1.927	-5.512	1.149	4.954	2.249	0	65.993	5.111	3.068	N
11	71845	-1.263	-4.408	1.866	6.696	3.222	0	58.845	6.06	4.227	N
11	73012	-4.524	-6.952	-1.337	5.379	2.752	0	53.598	1.55	7.898	N
11	73676	-4.878	-9.102	0.068	5.211	2.638	0	46.323	0.143	9.072	N
11	65100	-1.797	-7.387	3.229	5.171	2.329	0	57.975	4.416	5.251	N

예제 데이터 2

http://wolpack.hnu.ac.kr/iBOOKs/about_R/population_vital.csv

국내 인구동태 월별 (2000년 1월 ~ 2015년 12월)

- 출생(명) birth
- 사망(명) death
- 결혼(건) marriage
- 이혼(건) divorce
- 경제활동인구(명)

date_mom	birth	death	marriage	divrce	eco_pop
Jan.00	61229	24269	30120	9045	21442
Feb.00	56305	20138	26187	9392	21458
Mar.00	59448	21404	30418	10878	21810
Apr.00	52672	20253	28179	8640	22104
May.00	52058	19668	35330	10210	22337
Jun.00	46338	18589	29029	10324	22346
Jul.00	47992	19110	20488	10083	22482
Aug.00	50163	19599	18780	11216	22256
Sep.00	52430	19963	15844	9445	22432
Oct.00	53882	21510	26823	10626	22534
Nov.00	51886	20836	33343	9887	22395

시간도표 그리기

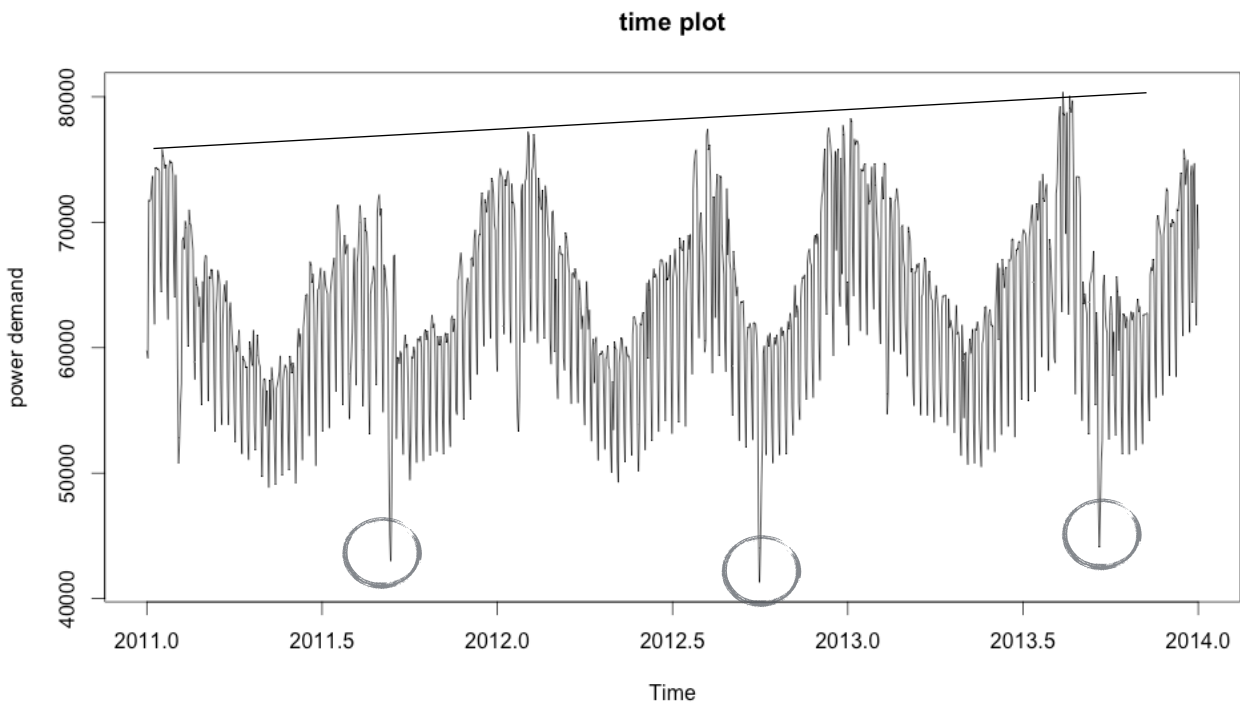
```
ds.pd=read.csv("power_demand.csv") #csv 외부데이터 R 데이터 읽기
names(ds.pd) #데이터 내 변수보기
yt.pd=ts(ds.pd$REAL_LOAD,frequency=365,start=c(2011,1,1)) #1일주기,2011년1월1일
plot.ts(yt.pd,main="time plot",ylab="power demand") #제목, y-축제목, 시간도표
```

함수 ts() : 시계열 데이터 만들기

- frequency 옵션 : 년=1, 월=12, 분기=4, 일=365
- start() 옵션 : 시계열 시각 시작 및 끝 날짜 지정, 월 c(2011,1)

함수 plot.ts() : 시간도표 그리기

- main 옵션 : 그래프 제목 지정, ylab=y-축 제목 지정



- 일별 자료이므로 당연히 7일(계절성) 주기 있음-전력수요이므로 월별주기(계절성) 있음
- 추석연휴 - 동그라, 전력수요가 이상적으로 낮음
- 직선적 증가 경향을 (직선) 보이고 있음 - 전력수요 연도별 증가 경향 보임

```

ds.pv=read.csv("population_vital.csv") #csv 외부데이터 R 데이터 읽기
names(ds.pv) #데이터 내 변수보기
yt.bt=ts(ds.pv$birth,frequency=12,start=c(2000,1)) #12일 주기,2000년 1월
yt.dt=ts(ds.pv$death,frequency=12,start=c(2000,1)) #12일 주기,2000년 1월
yt.bt.dt=data.frame(cbind(yt.bt,yt.dt))

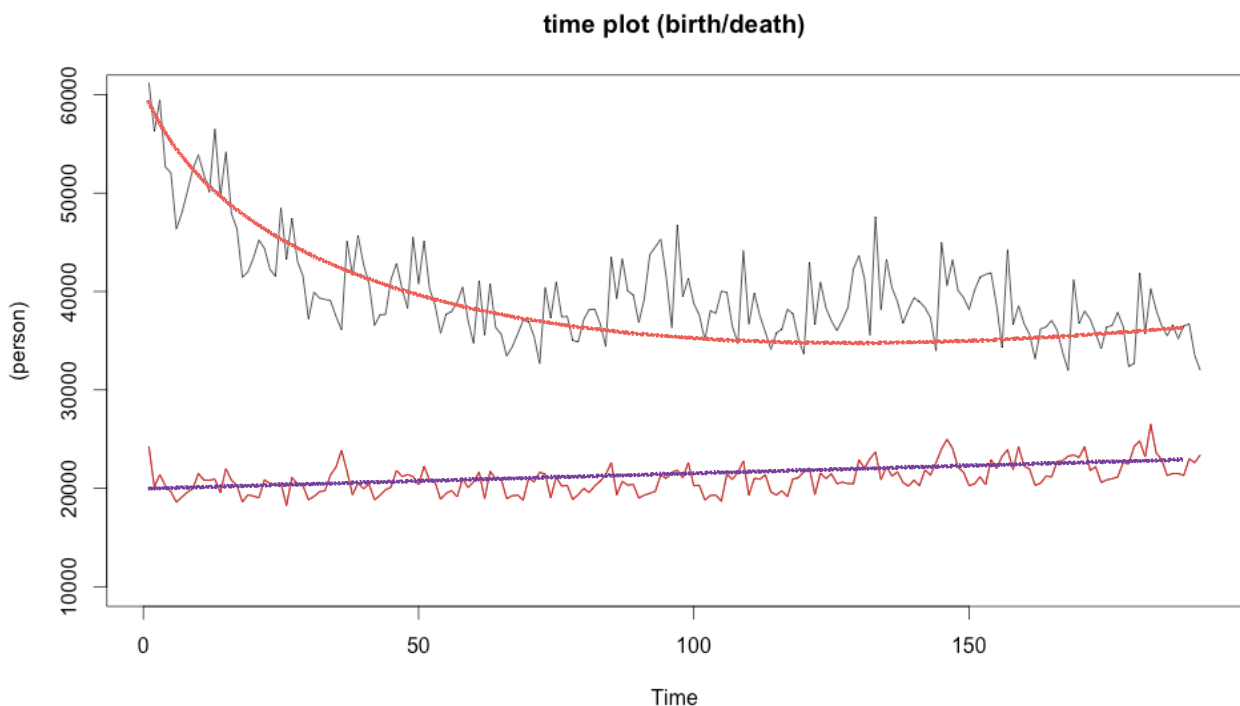
plot.ts(yt.bt.dt$yt.bt,main="time plot (birth/death)",ylab="(person)",ylim=c(10000,60000))
#제목, y-축제목, 시간도표

lines(yt.bt.dt$yt.dt,main="time plot (birth/death)",ylab="(person)",col="red") #제목, y-축
제목, 시간도표

```

함수 lines() : 라인 산점도 겹쳐 그리기

- col="red" 옵션 : 선을 붉은 색으로 표현



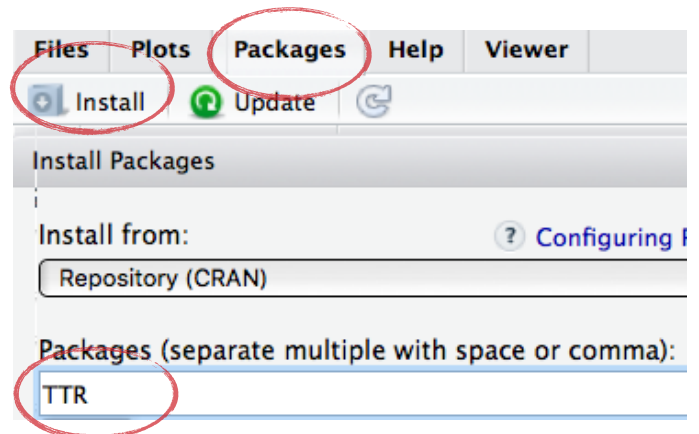
- 지수적 감소하는 경향을 보임 (출생자 수)
- 사망자 수는 직선적 증가
- 월별 자료이므로 12주기 계절성 있음

이동 평균법 Moving Average

```
library(TTR) #Smoothing 함수 불러오기
fit.yt=SMA(yt.pd,n=7) #주기(m)=7 이동평균법
fit2.yt=SMA(yt.pd,n=28) #주기(m)=28 이동평균법
fit3.yt=filter(yt.pd,rep(1/7,7),sides=1) #가중치1/7, 7개관측치, 1=이동평균, 2=중심이동평균
ds.pd2=data.frame(cbind(yt,fit.yt,fit2.yt))
plot.ts(ds.pd2$yt,main="Time Plot(raw, MA7, MA28)",ylab="power demand",col="black")
#월데이터
lines(ds.pd2$fit.yt,main="Time Plot(raw, MA7, MA28)",ylab="power
demand",col="red",lty="dashed") # m=7 MA
lines(ds.pd2$fit2.yt,main="Time Plot(raw, MA7, MA28)",ylab="power
demand",col="blue",lty="dashed") # m=28 MA
```

함수 library() : 함수 패키지

- 예측방법 함수 SMA는 TTR에 있음 - 처음 사용 시에는 옆 그림처럼 Packages -> Install 에서 TTR을 설치



함수 SMA() : 이동평균 구하기

- n 옵션 : 주기(m) 옵션임

함수 filter() : 가중 이동평균 구하기

- rep() 옵션 : 가중치 부여, (0.2,5) - 5개 관측치에 0.2 가중값 - 합은 반드시 1
- sides 옵션 : 관측치 사용 지정 1=이동평균과 동일(가장 최근 m개)-(..., t-1,t), 2=중심 기준으로 양쪽, 본 예제의 경우 m=7인 경우 (t-2, t-1, t, t+1, t+2) 관측치 사용

함수 data.frame() : 오브젝트를 데이터로 변환

함수 cbind() : 오브젝트들 열 합치기

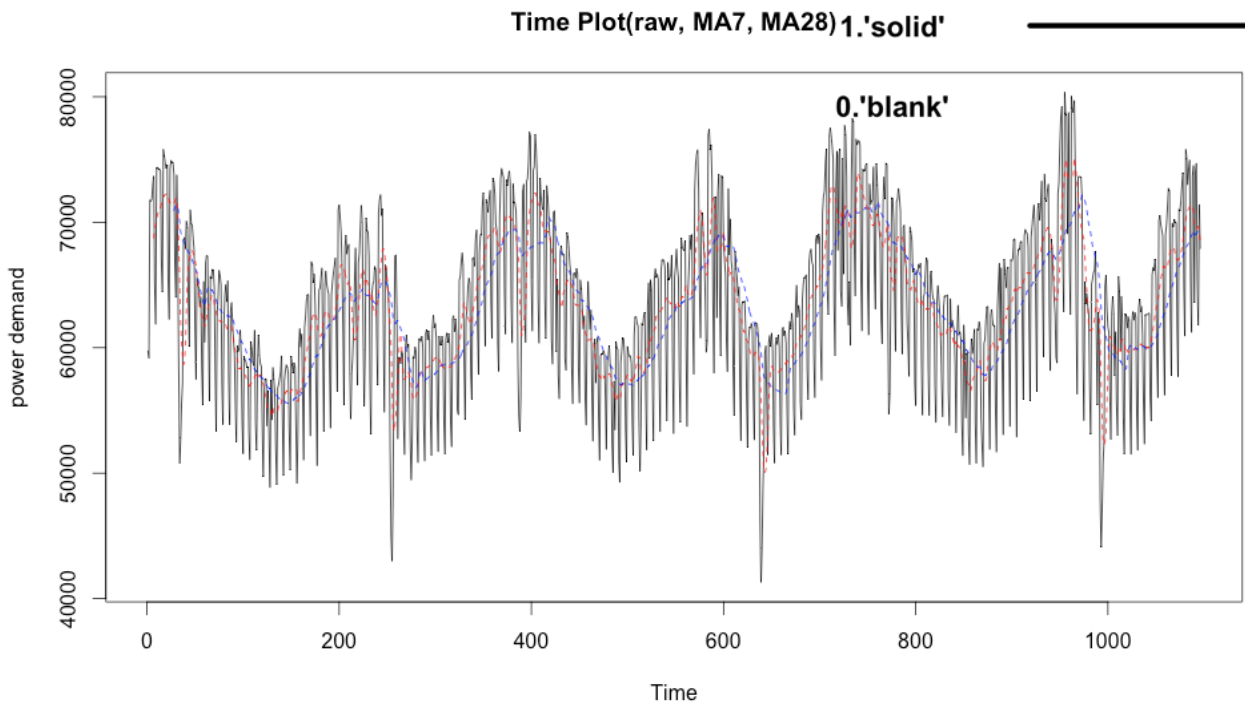
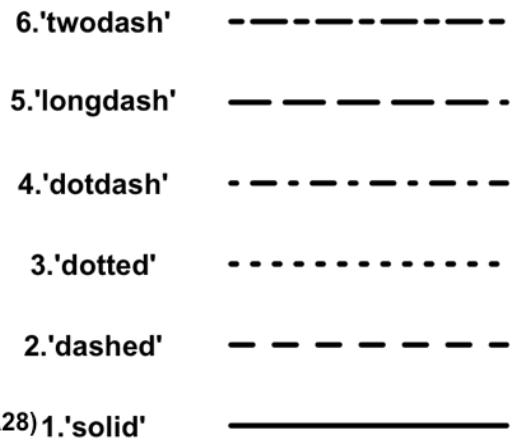
함수 plot.ts() : plot.ts

- ts() 함수에 의해 시계열 관측치에 날짜 부여하였으나 data.frame() 함수로 인하여 날짜가 사라졌으므로 더 이상 x-축은 날짜가 나타나지 않는다.

함수 lines() : 라인 산점도 겹쳐 그리기

- col="blue" 옵션 : 선을 블루 색으로 표현

- lty="dashed" : 선의 타입을 대쉬로 함,



- 단기, 중장기적으로 증가 추세 - 붉은 선, 파란 선 모두 증가

```

library(TTR) #Smoothing 함수 불러오기
fit.yt.bt=SMA(yt.bt,n=12) #주기(m)=12 이동평균법
fit2.yt.bt=SMA(yt.bt,n=60) #주기(m)=60 이동평균법

ds.pv2=data.frame(cbind(yt.bt,fit.yt.bt,fit2.yt.bt))

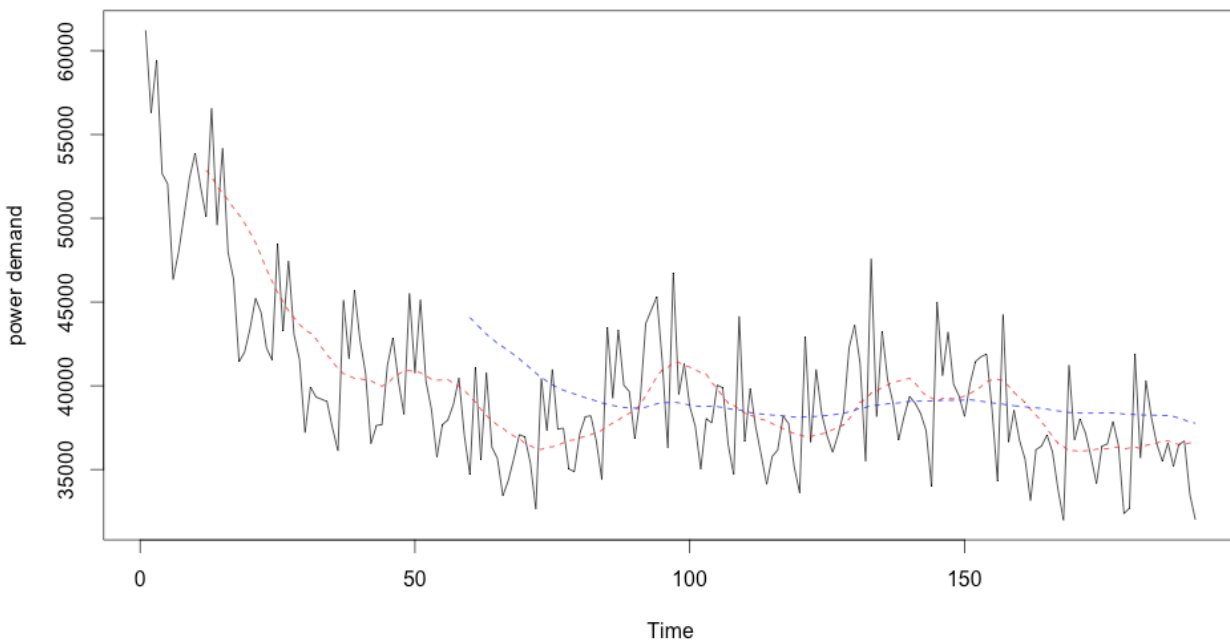
plot.ts(ds.pv2$yt.bt,main="Time Plot(raw, MA12, MA60)",ylab="power
demand",col="black") #원데이터

lines(ds.pv2$fit.yt.bt,main="Time Plot(raw, MA12, MA60)",ylab="power
demand",col="red",lty="dashed") # m=12 MA

lines(ds.pv2$fit2.yt.bt,main="Time Plot(raw, MA12, MA60)",ylab="power
demand",col="blue",lty="dashed") # m=60 MA

```

Time Plot(raw, MA12, MA60)



- 출생율 : 단기적으로는(붉은 선) 현상 유지하거나 반등이 있겠지만 중장기적(파란색)으로는 낮아질 것으로 예측됨
- $m=12$, 1년 이동평균보다 5년 이동평균 직선에 가까움