

포브스 선정 Top2000 기업 데이터

```

> getwd()
[1] "/Users/baektaehyun/Desktop/전산통계학/실습파일"
> library(xlsx)
> ds <- read.xlsx("Forbes-2021.xlsx",header = T, sheetIndex = 1)
> head(ds)
  Rank      Name      Country Sales Profit
1    1      ICBC      China $190.5 B $45.8 B
2    2 JPMorgan Chase United States $136.2 B $40.4 B
3    3 Berkshire Hathaway United States $245.5 B $42.5 B
4    4 China Construction Bank      China $173.5 B $39.3 B
5    5 Saudi Arabian Oil Company (Saudi Aramco) Saudi Arabia $229.7 B $49.3 B
6    6 Apple United States      $294 B $63.9 B

  Assets Market.Value Year
1 $4,914.7 B      $249.5 B 2021
2 $3,689.3 B      $464.8 B 2021
3  $873.7 B      $624.4 B 2021
4 $4,301.7 B      $210.4 B 2021
5  $510.3 B      $1,897.2 B 2021
6  $354.1 B      $2,252.3 B 2021
> reStr <- function(x) gsub("^\\s+|\\s+$|[$BM,]", "", x)
> for(i in 4:7){
+   ds[,i] <- as.numeric(reStr(ds[,i]))
+ }
> head(ds)
  Rank      Name      Country Sales Profit Assets
1    1      ICBC      China 190.5   45.8 4914.7
2    2 JPMorgan Chase United States 136.2   40.4 3689.3
3    3 Berkshire Hathaway United States 245.5   42.5 873.7
4    4 China Construction Bank      China 173.5   39.3 4301.7
5    5 Saudi Arabian Oil Company (Saudi Aramco) Saudi Arabia 229.7   49.3 510.3
6    6 Apple United States 294.0   63.9 354.1

  Market.Value Year
1      249.5 2021
2      464.8 2021
3      624.4 2021
4      210.4 2021
5     1897.2 2021
6     2252.3 2021
> ds[!complete.cases(ds),] #결측값 확인
  Rank      Name Country Sales Profit Assets Market.Value Year
1933 1824 Bank of Greece Greece    1.3  942.8    NA      366 2021
2033 1918 Bank of Greece Greece    1.3  942.8    NA      366 2021
> table(ds$Country)
  Argentina      Australia      Austria      Bahrain
           2             34             9             1
Bangladesh      Belgium      Bermuda      Brazil
           1             9             6            22
Canada          Chile      China      Colombia
          58             5      307             3
Cyprus          Czech Republic      Denmark      Egypt
           1             1            12             1
Finland          France      Germany      Greece
           9            56            58             6
Hong Kong      Hungary      India      Indonesia
          64             2            55             6
Ireland        Israel      Italy      Japan
          18            13            24            229
Kazakhstan      Kenya      Kuwait      Luxembourg
           2             1             3             7
Malaysia        Mexico      Monaco      Morocco
           9            11            1             4
Netherlands     Nigeria      Norway      Oman
          20             3             8             1
Peru            Philippines      Poland      Portugal
           1             7             7             4
Qatar           Russia      Saudi Arabia      Singapore
           6            25            13             9
South Africa    South Korea      Spain      Sweden
          16            62            21            34
Switzerland     Taiwan      Thailand      Turkey
          46            46            15            10
United Arab Emirates United Kingdom      United States      Venezuela
           9            69            611             1
Vietnam
           6

```

```
> tmp <- sort(table(ds$Country), decreasing = T)
> tmp
```

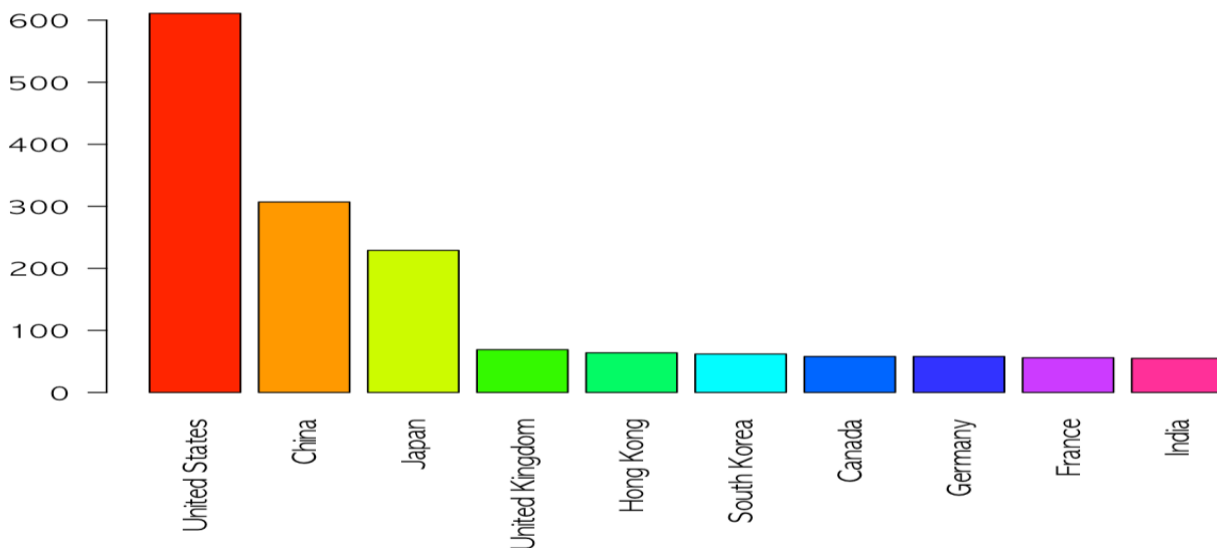
United States	China	Japan	United Kingdom
611	307	229	69
Hong Kong	South Korea	Canada	Germany
64	62	58	58
France	India	Switzerland	Taiwan
56	55	46	46
Australia	Sweden	Russia	Italy
34	34	25	24
Brazil	Spain	Netherlands	Ireland
22	21	20	18
South Africa	Thailand	Israel	Saudi Arabia
16	15	13	13
Denmark	Mexico	Turkey	Austria
12	11	10	9
Belgium	Finland	Malaysia	Singapore
9	9	9	9
United Arab Emirates	Norway	Luxembourg	Philippines
9	8	7	7
Poland	Bermuda	Greece	Indonesia
7	6	6	6
Qatar	Vietnam	Chile	Morocco
6	6	5	4
Portugal	Colombia	Kuwait	Nigeria
4	3	3	3
Argentina	Hungary	Kazakhstan	Bahrain
2	2	2	1
Bangladesh	Cyprus	Czech Republic	Egypt
1	1	1	1
Kenya	Monaco	Oman	Peru
1	1	1	1
Venezuela			
1			

```
> top.10.contry <- tmp[1:10]
> top.10.contry
```

United States	China	Japan	United Kingdom	Hong Kong
611	307	229	69	64
South Korea	Canada	Germany	France	India
62	58	58	56	55

```
> par(mar=c(8,4,4,2))
> barplot(top.10.contry,
+         main = '기업수 상위 10개국',
+         col=rainbow(10),
+         las=2)
```

기업수 상위 10개국



```

> tmp <- ds[order(ds$Market.Value, decreasing = T),]
> top.10.Market_value <- tmp[1:10,c('Name', 'Market.Value')]
> top.10.Market_value #2021년 기준 시가총액 상위 10개 기업

```

	Name	Market.Value
6	Apple	2252.3
15	Microsoft	1966.6
5	Saudi Arabian Oil Company (Saudi Aramco)	1897.2
10	Amazon	1711.8
13	Alphabet	1538.9
1932	Cattolica Assicurazioni	975.0
2032	Cattolica Assicurazioni	975.0
1714	Rite Aid	974.0
1639	Nishi-nippon Financial Holdings	967.0
1718	77 Bank	966.0

```

> korea <- subset(ds, Country == 'South Korea')
> korea[,c('Name', 'Country', 'Market.Value')] #포브스 2000에 속하는 한국기업

```

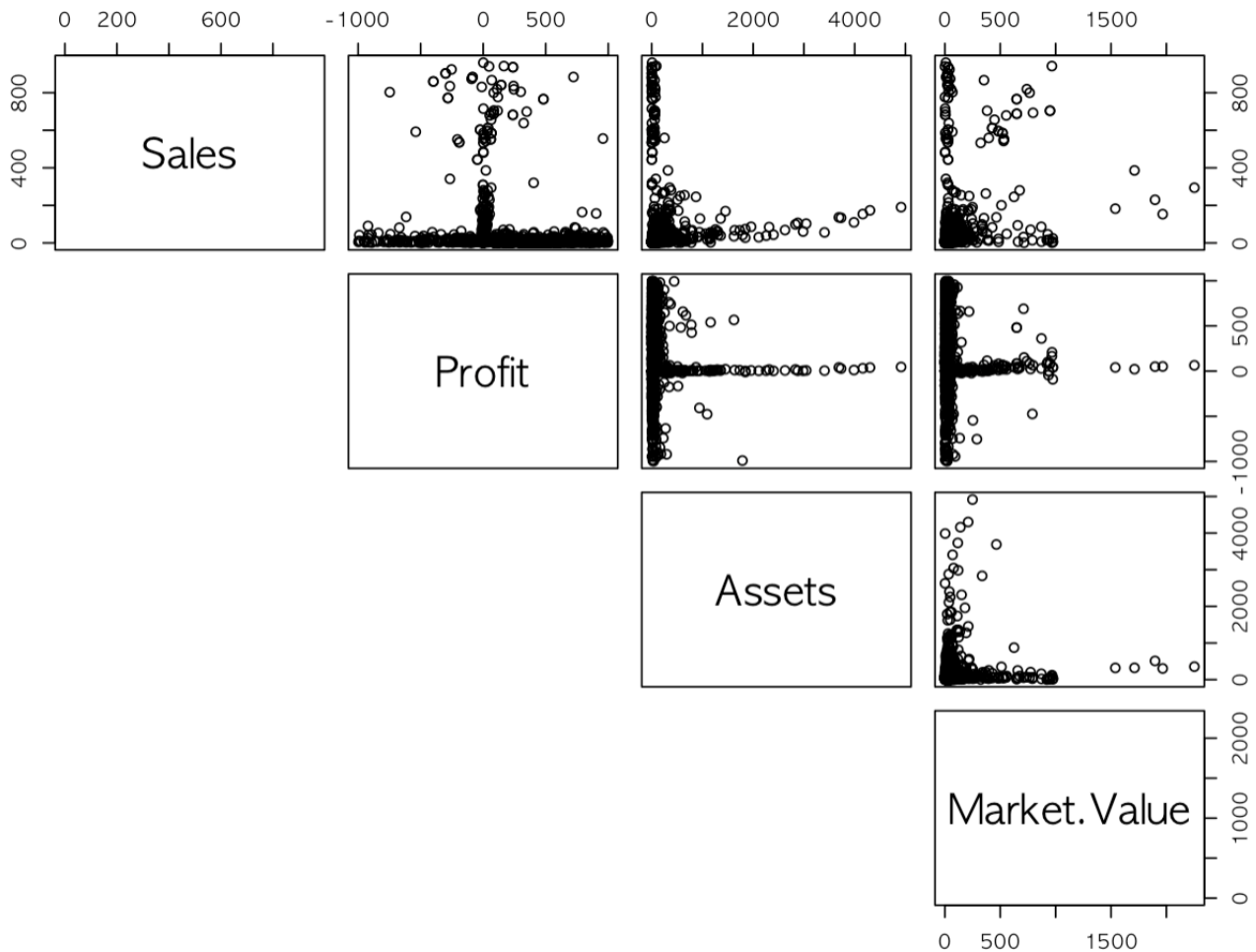
	Name	Country	Market.Value
11	Samsung Electronics	South Korea	510.5
155	Hyundai Motor	South Korea	54.0
173	SK Hynix	South Korea	84.2
229	KB Financial Group	South Korea	18.4
248	Shinhan Financial Group	South Korea	17.6
266	KIA	South Korea	31.2
269	Posco	South Korea	23.4
279	LG Electronics	South Korea	27.2
292	Korea Electric Power	South Korea	13.8
353	Hyundai Mobis	South Korea	25.3
390	Samsung Life Insurance	South Korea	13.0
427	Samsung C&T	South Korea	20.5
446	Hana Financial Group	South Korea	10.7
474	SK Telecom	South Korea	19.3
480	LG Chem	South Korea	62.5
659	SK Holdings	South Korea	13.4
662	Woori Financial Group	South Korea	6.6
672	Industrial Bank of Korea	South Korea	6.1
728	Samsung Fire & Marine	South Korea	7.3
760	SK Innovation	South Korea	20.7
773	Samsung SDI	South Korea	41.6
832	Hanwha	South Korea	2.0
863	Naver	South Korea	51.7
899	KT	South Korea	6.2
911	Db Insurance	South Korea	2.4
954	CJ Cheiljedang	South Korea	6.0
980	LG	South Korea	17.3
1032	Coupang	South Korea	78.8
1053	Meritz Financial Group	South Korea	2.0
1102	LG Household & Health Care	South Korea	23.4
1115	CJ Corporation	South Korea	2.6
1131	Mirae Asset Daewoo	South Korea	5.7
1172	Korea Investment Holdings	South Korea	5.5
1202	LG Display	South Korea	7.9
1233	Korea Gas	South Korea	2.6
1258	Hyundai Marine & Fire	South Korea	1.7
1288	Hyundai Steel	South Korea	5.8
1361	Lotte Shopping	South Korea	3.1
1371	Doosan	South Korea	790.0
1378	BNK Financial Group	South Korea	1.9
1395	E-mart	South Korea	4.1
1397	Hyundai Heavy Industries Holdings	South Korea	3.7
1408	Celltrion	South Korea	37.7
1417	NH Investment & Securities	South Korea	3.2
1426	Hyundai Glovis	South Korea	6.2
1455	Korea Shipbuilding & Offshore Engineering	South Korea	8.7
1461	Samsung Electro-Mechanics	South Korea	13.1
1464	Samsung SDS	South Korea	13.3
1468	Kiwoom Securities	South Korea	3.3

```

> tmp <- ds[,4:7] #sales, profit, assets,market.Value 추출
> tmp <- tmp[complete.cases(tmp),]
> plot(tmp, lower.panel=NULL) #산점도
> cor(tmp) #상관계수 -> 기업의 자산과 자산의 어느정도 서로 영향을 끼침

```

	Sales	Profit	Assets	Market.Value
Sales	1.00000000	-0.1510503	0.05272886	0.3048553
Profit	-0.15105026	1.00000000	-0.13956449	-0.1285085
Assets	0.05272886	-0.1395645	1.00000000	0.1028012
Market.Value	0.30485529	-0.1285085	0.10280122	1.00000000



대기오염 데이터

```
> getwd()
[1] "/Users/baektaehyun/Desktop/전산통계학/실습파일"
> setwd("/Users/baektaehyun/desktop/전산통계학/실습파일")
> library(xlsx)
> files <- c("2021년 1월.xlsx", "2021년 2월.xlsx", "2021년 3월.xlsx", "2021년 4월.xlsx", "2021년 5월.xlsx", "2021년 6월.xlsx")
> ds<-NULL
> for(f in files){
+   tmp <- read_xlsx(paste("/Users/baektaehyun/desktop/전산통계학/실습파일/", f, sep = ""))
+   ds <- rbind(ds, tmp)
+   print(f)
+ }
[1] "2021년 1월.xlsx"
[1] "2021년 2월.xlsx"
[1] "2021년 3월.xlsx"
[1] "2021년 4월.xlsx"
[1] "2021년 5월.xlsx"
[1] "2021년 6월.xlsx"
> colnames(ds)
[1] "지역"          "망"          "측정소코드" "측정소명"    "측정일시"    "SO2"
[7] "CO"           "O3"          "NO2"         "PM10"        "PM25"        "주소"
> head(ds)
# A tibble: 6 × 12
  지역 망 측정소...¹ 측정소명 측정...² SO2 CO O3 NO2 PM10 PM25 주소
<chr> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <chr>
1 서울 중구 도시대기 111121 중구 2.02e9 0.003 0.4 0.02 0.018 27 14 서울...
2 서울 중구 도시대기 111121 중구 2.02e9 0.003 0.5 0.017 0.021 26 14 서울...
3 서울 중구 도시대기 111121 중구 2.02e9 0.003 0.5 0.014 0.026 25 16 서울...
4 서울 중구 도시대기 111121 중구 2.02e9 0.003 0.4 0.021 0.018 22 13 서울...
5 서울 중구 도시대기 111121 중구 2.02e9 0.003 0.5 0.01 0.029 25 13 서울...
6 서울 중구 도시대기 111121 중구 2.02e9 0.003 0.6 0.005 0.036 27 16 서울...
# ... with abbreviated variable names ¹측정소코드, ²측정일시
```

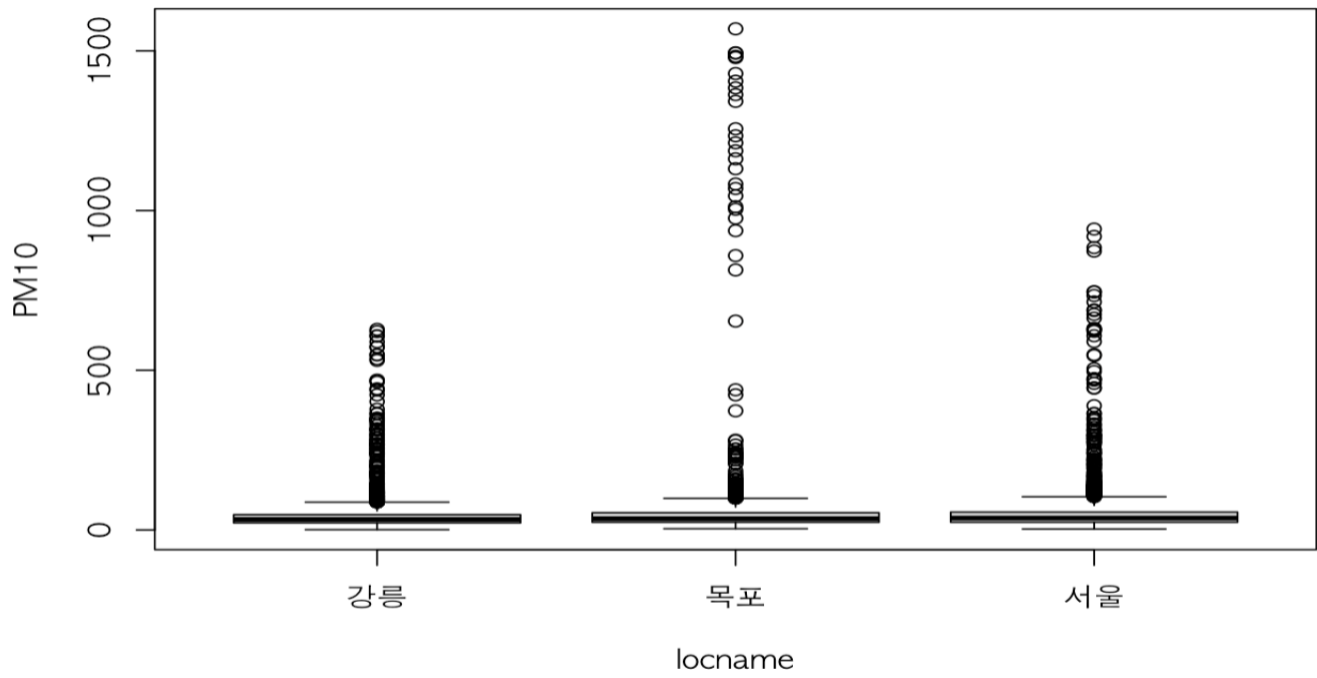
```
> range(ds$측정일시) # 측정기간 20210101 ~ 2021063024
[1] 2021010101 2021063024
> ds_new <- cbind(ds[,3], ds[,5], ds[,6:11]) #사용할 데이터들을 따로 추출하여 데이터 처리
> colnames(ds_new) <- c("loc", "mdate", colnames(ds_new)[3:8])
> unique(ds_new$loc)
[1] "111121" "111122" "111123" "111124" "111125" "111131" "111141" "111142" "111143"
[10] "111151" "111152" "111154" "111161" "111162" "111171" "111181" "111191" "111201"
[19] "111202" "111212" "111213" "111231" "111232" "111241" "111242" "111251"
[28] "111261" "111262" "111263" "111264" "111273" "111274" "111275" "111281" "111282"
[37] "111291" "111301" "111311" "111312" "131111" "131112" "131113" "131114" "131115"
[46] "131116" "131117" "131118" "131120" "131121" "131123" "131124" "131125" "131126"
[55] "131128" "131129" "131131" "131132" "131133" "131141" "131142" "131144" "131145"
[64] "131161" "131163" "131191" "131192" "131193" "131194" "131195" "131196" "131197"
[73] "131198" "131201" "131202" "131211" "131212" "131222" "131223" "131231" "131232"
[82] "131233" "131234" "131235" "131236" "131241" "131242" "131243" "131244" "131245"
[91] "131246" "131247" "131341" "131342" "131343" "131344" "131345" "131346" "131371"
[100] "131372" "131373" "131374" "131381" "131382" "131383" "131384" "131385" "131392"
[109] "131394" "131395" "131411" "131412" "131413" "131414" "131415" "131416" "131417"
[118] "131441" "131442" "131443" "131444" "131451" "131453" "131454" "131471" "131472"
[127] "131473" "131474" "131475" "131501" "131502" "131531" "131532" "131541" "131542"
[136] "131551" "131552" "131553" "131554" "131555" "131556" "131557" "131558" "131561"
[145] "131562" "131571" "131571" "131581" "131583" "131584" "131591" "131592" "131593"
[154] "131611" "131612" "131621" "131622" "131991" "132112" "132113" "132118" "132119"
[163] "132401" "132402" "132901" "132902" "132991" "132992" "132993" "132994" "221112"
[172] "221131" "221141" "221142" "221152" "221162" "221163" "221172" "221181" "221182"
[181] "221183" "221191" "221192" "221193" "221202" "221211" "221212" "221213" "221221"
[190] "221231" "221233" "221241" "221251" "221253" "221271" "221281" "221282" "221283"
[199] "221284" "221901" "221902" "221903" "238111" "238112" "238113" "238120" "238121"
[208] "238122" "238123" "238124" "238125" "238126" "238127" "238128" "238129" "238130"
[217] "238131" "238132" "238133" "238134" "238141" "238142" "238143" "238144" "238145"
[226] "238146" "238151" "238152" "238161" "238181" "238182" "238183" "238184" "238191"
[235] "238201" "238203" "238211" "238212" "238241" "238361" "238362" "238363" "238371"
[244] "238373" "238374" "238375" "238376" "238377" "238378" "238379" "238381" "238401"
[253] "238411" "238421" "238431" "238441" "238451" "238461" "238471" "238481" "238491"
[262] "238501" "238511" "238901" "238902" "324115" "324121" "324123" "324124" "324125"
[271] "324133" "324134" "324135" "324136" "324142" "324143" "324148" "324155" "335115"
[280] "336111" "336112" "336121" "336124" "336125" "336127" "336128" "336131" "336132"
[289] "336133" "336134" "336135" "336141" "336151" "336161" "336221" "336222" "336223"
[298] "336352" "336353" "336354" "336355" "336361" "336441" "336442" "336451" "336452"
[307] "336461" "336462" "336471" "336481" "336491" "336501" "336511" "336521" "336522"
[316] "336523" "336531" "336541" "336551" "336561" "336571" "336572" "336901" "336902"
[325] "336903" "339111" "339112" "339113" "339114" "339121" "339122" "339123" "339131"
[334] "339211" "339312" "422114" "422115" "422121" "422132" "422133" "422141" "422154"
[343] "422155" "422161" "422171" "422201" "422202" "422203" "422204" "422205" "422206"
[352] "437112" "437113" "437114" "437115" "437116" "437117" "437118" "437119" "437122"
[361] "437123" "437124" "437125" "437131" "437132" "437133" "437141" "437151" "437152"
[370] "437153" "437154" "437161" "437171" "437181" "437191" "437201" "437202" "437203"
[379] "437211" "437221" "437371" "437401" "437402" "437411" "437412" "437421" "437431"
[388] "437541" "437542" "437551" "437561" "437571" "437581" "437591" "437901" "525111"
[397] "525112" "525121" "525141" "525142" "525143" "525144" "525151" "525161" "525162"
[406] "525171" "525172" "525173" "533112" "533113" "533114" "533115" "533116" "534111"
[415] "534112" "534114" "534115" "534116" "534341" "534342" "534343" "534411" "534421"
[424] "534422" "534423" "534424" "534433" "534434" "534441" "534442" "534443" "534444"
```

```
> unique(ds$지역)
[1] "서울 종로구" "서울 용산구" "서울 중로구" "서울 광진구" "서울 성동구"
[6] "서울 중랑구" "서울 동대문구" "서울 성북구" "서울 도봉구" "서울 은평구"
[11] "서울 서대문구" "서울 마포구" "서울 강서구" "서울 구로구" "서울 영등포구"
[16] "서울 동작구" "서울 관악구" "서울 강남구" "서울 서초구" "서울 송파구"
[21] "서울 강동구" "서울 금천구" "서울 강북구" "서울 양천구" "서울 노원구"
[26] "경기 수원시" "경기 성남시" "경기 의정부시" "경기 안양시" "경기 광명시"
[31] "경기 안산시" "경기 과천시" "경기 구리시" "경기 의왕시" "경기 시흥시"
[36] "경기 남양주시" "경기 평택시" "경기 파주시" "경기 고양시" "경기 군포시"
[41] "경기 용인시" "경기 이천시" "경기 포천시" "경기 김포시" "경기 군포시"
[46] "경기 오산시" "경기 하남시" "경기 화성시" "경기 양주시" "경기 동두천시"
[51] "경기 안성시" "경기 여주시" "경기 연천군" "경기 가평군" "경기 양평군"
[56] "강원 춘천시" "강원 양구군" "강원 동해시" "강원 철원군" "강원 화천군"
[61] "강원 인제군" "강원 고성군" "부산 중구" "부산 동구" "부산 영도구"
[66] "부산 부산진구" "부산 동래구" "부산 남구" "부산 사상구" "부산 북구"
[71] "부산 금정구" "부산 해운대구" "부산 사하구" "부산 강서구" "부산 연제구"
[76] "부산 기장군" "부산 수영구" "부산 서구" "경남 창원시" "울산 동구"
[81] "울산 중구" "울산 남구" "울산 울주군" "울산 북구" "경남 진주시"
[86] "경남 하동군" "경남 김해시" "경남 거제시" "경남 사천시" "경남 양산시"
[91] "경남 밀양시" "경남 통영시" "경남 고성군" "경남 거창군" "경남 함안군"
[96] "경남 함양군" "경남 남해군" "경남 산청군" "경남 의령군" "경남 창녕군"
[101] "경남 합천군" "광주 동구" "광주 서구" "광주 남구" "광주 북구"
[106] "광주 광산구" "전북 전주시" "전남 목포시" "전남 여수시" "전남 순천시"
[111] "전남 나주시" "전남 담양군" "전남 장성군" "전남 광양시" "전남 해남군"
[116] "전남 영암군" "전남 화순군" "전남 영광군" "전남 장흥군" "전남 진도군"
[121] "전남 완도군" "전남 함평군" "전남 고흥군" "전남 신안군" "전남 무안군"
[126] "전남 강진군" "전남 곡성군" "전남 구례군" "전남 보성군" "제주 제주시"
[131] "제주 서귀포시" "대구 중구" "대구 수성구" "대구 동구" "대구 서구"
[136] "대구 남구" "대구 북구" "대구 달서구" "대구 달성군" "경북 포항시"
[141] "경북 경주시" "경북 김천시" "경북 안동시" "경북 구미시" "경북 영주시"
[146] "경북 경산시" "경북 상주시" "경북 칠곡군" "경북 영덕군" "경북 문경시"
[151] "경북 성주군" "경북 영천시" "경북 의성군" "경북 울진군" "경북 봉화군"
[156] "경북 울릉군" "경북 고령군" "경북 군위군" "경북 영양군" "경북 예천군"
[161] "경북 청도군" "대전 대덕구" "대전 중구" "대전 유성구" "대전 동구"
[166] "대전 서구" "충북 청주시" "충남 천안시" "충남 공주시" "충남 부여군"
[171] "충남 서산시" "충남 당진시" "충남 아산시" "충남 논산시" "충남 태안군"
[176] "충남 예산군" "충남 보령시" "충남 홍성군" "충남 금산군" "충남 청양군"
[181] "충남 계룡시" "충남 서천군" "세종 세종시" "강원 원주시" "강원 강릉시"
[186] "강원 삼척시" "강원 평창군" "강원 정선군" "강원 횡성군" "강원 양양군"
[191] "강원 속초시" "강원 홍천군" "강원 영월군" "강원 태백시" "충북 충주시"
[196] "충북 제천시" "충북 단양군" "충북 괴산군" "충북 진천군" "충북 음성군"
[201] "충북 영동군" "충북 증평군" "충북 보은군" "충북 옥천군" "전북 군산시"
[206] "전북 익산시" "전북 정읍시" "전북 남원시" "전북 고창군" "전북 부안군"
[211] "전북 김제시" "전북 완주군" "전북 진안군" "전북 임실군" "전북 무주군"
[216] "전북 순창군" "전북 장수군" "경북 청송군" "인천 중구" "인천 동구"
[221] "인천 남동구" "인천 미추홀구" "인천 부평구" "인천 서구" "인천 계양구"
[226] "인천 강화군" "인천 연수구" "경기 부천시" "인천 옹진군"
```

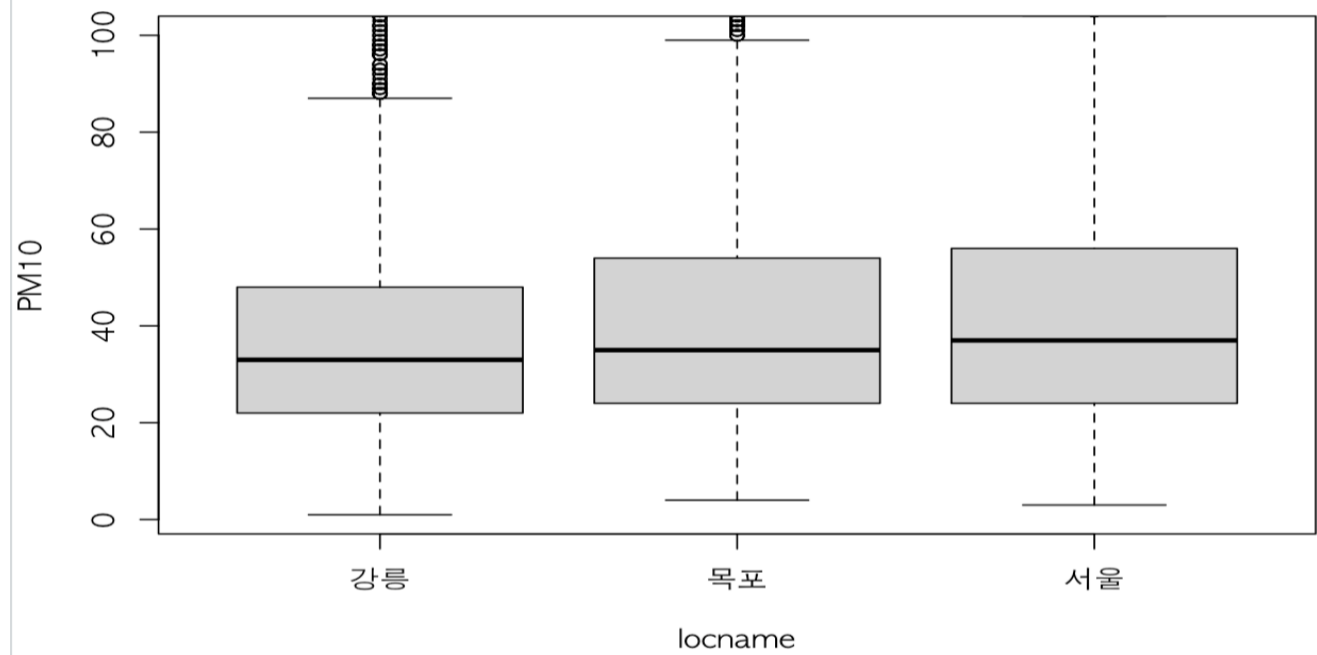
```
> ds_new <- ds_new[complete.cases(ds_new),]
> mdate <- as.character(ds_new$mdate)
> ds_new$year <- as.numeric(substr(mdate, 1, 4))
> ds_new$month <- as.numeric(substr(mdate, 5, 6))
> ds_new$hour <- as.numeric(substr(mdate, 9, 10))
> ds_new$locname <- NA
> ds_new$locname[ds_new$loc == 111123] <- "서울"
> ds_new$locname[ds_new$loc == 336111] <- "목포"
> ds_new$locname[ds_new$loc == 632132] <- "강릉"
> ds_new <- ds_new[complete.cases(ds_new),] #ds_new의 결측값 제거
> str(ds_new)
'data.frame': 10864 obs. of 12 variables:
 $ loc : chr "111123" "111123" "111123" "111123" ...
 $ mdate : num 2.02e+09 2.02e+09 2.02e+09 2.02e+09 2.02e+09 ...
 $ S02 : num 0.002 0.002 0.002 0.003 0.003 0.002 0.003 0.003 0.003 0.004 ...
 $ CO : num 0.5 0.6 0.6 0.6 0.6 0.6 0.7 0.7 0.8 0.8 0.9 ...
 $ O3 : num 0.022 0.018 0.013 0.011 0.008 0.003 0.002 0.002 0.004 0.007 ...
 $ NO2 : num 0.016 0.02 0.025 0.027 0.032 0.037 0.039 0.041 0.04 0.039 ...
 $ PM10 : num 24 25 27 23 24 26 27 30 33 35 ...
 $ PM25 : num 14 14 16 13 14 16 18 18 19 19 ...
 $ year : num 2021 2021 2021 2021 2021 2021 ...
 $ month : num 1 1 1 1 1 1 1 1 1 1 ...
 $ hour : num 1 2 3 4 5 6 7 8 9 10 ...
 $ locname: chr "서울" "서울" "서울" "서울" ...
> head(ds_new)
      loc      mdate      S02      CO      O3      NO2      PM10      PM25      year      month      hour      locname
1489 111123 2021010101 0.002 0.5 0.022 0.016 24 14 2021 1 1 서울
1490 111123 2021010102 0.002 0.6 0.018 0.020 25 14 2021 1 2 서울
1491 111123 2021010103 0.002 0.6 0.013 0.025 27 16 2021 1 3 서울
1492 111123 2021010104 0.003 0.6 0.011 0.027 23 13 2021 1 4 서울
1493 111123 2021010105 0.003 0.6 0.008 0.032 24 14 2021 1 5 서울
1494 111123 2021010106 0.002 0.7 0.003 0.037 26 16 2021 1 6 서울
```

```
30 ds_new
31 boxplot(PM10~locname, data=ds_new,
32         main='미세먼지 농도 분포') # PM10 미세먼지의 장소별 상자 농도
33 boxplot(PM10~locname, data=ds_new, #미세먼지의 농도 범위를 100으로 제한하여 다시 그림
34         main='미세먼지 농도 분포',
35         ylim=c(1,100))
36
```

미세먼지 농도 분포



미세먼지 농도 분포




```

> tmp.month <- aggregate(ds_new[,7],
+                         by = list(month = ds_new$month,
+                         loc = ds_new$locname), FUN = "mean") #PM 10 미세먼지의 월별 평균
> tmp.month
  month  loc      x
1     1 강릉 31.62857
2     2 강릉 38.62021
3     3 강릉 54.94598
4     4 강릉 44.81135
5     5 강릉 53.72452
6     6 강릉 27.76969
7     1 목포 45.29585
8     2 목포 44.63595
9     3 목포 94.92143
10    5 목포 42.32614
11    6 목포 30.95571
12    1 서울 41.50342
13    2 서울 50.25988
14    3 서울 69.53940
15    4 서울 42.57283
16    5 서울 63.27891
17    6 서울 31.91854

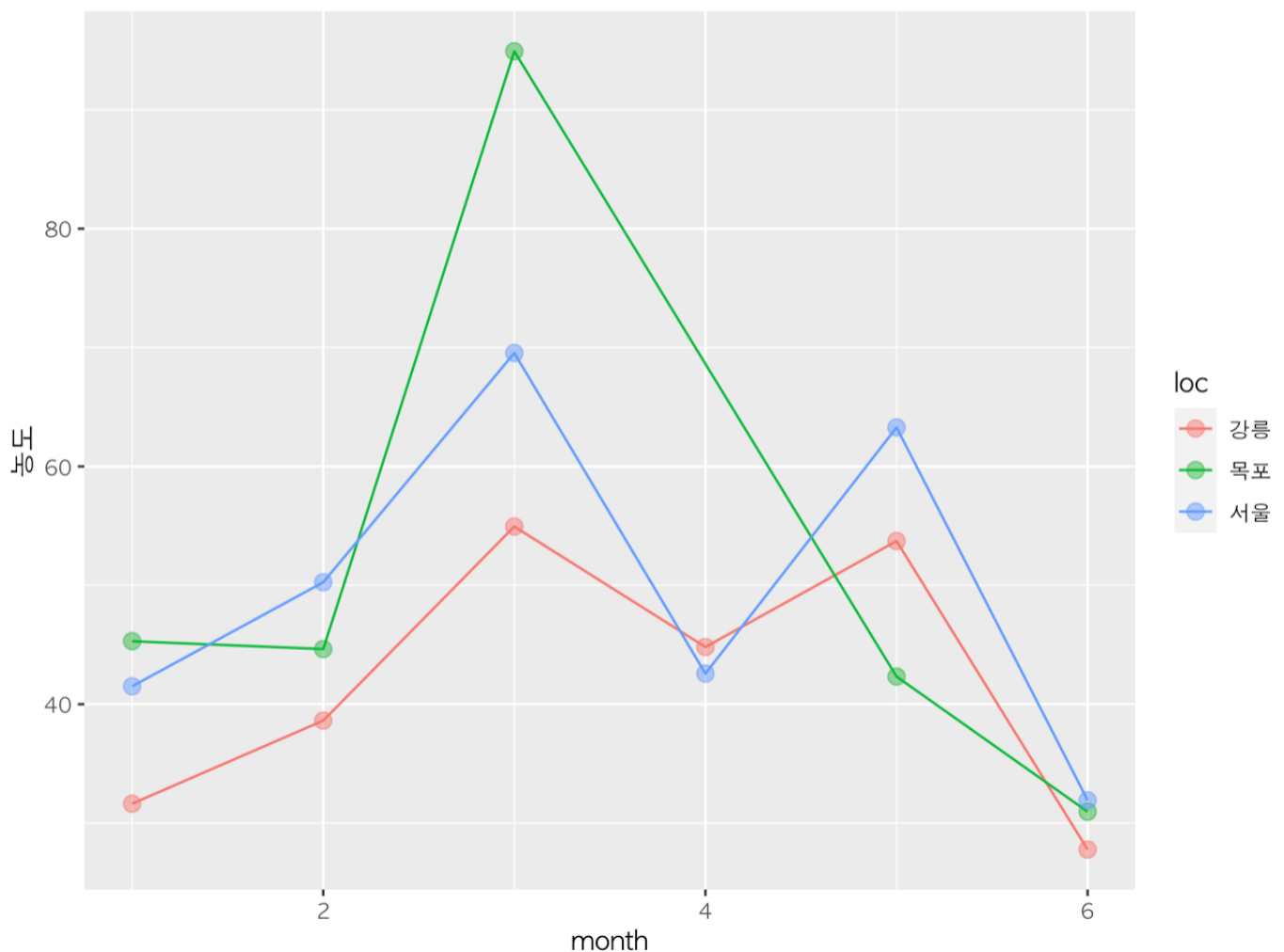
```

```

> ggplot(tmp.month,
+        aes(x=month,
+            y = x,
+            colour = loc,
+            group = loc))+
+   geom_line() +
+   geom_point(size = 3, shape = 19, alpha = 0.5) +
+   ggtitle("월별 PM10 농도 변화")+
+   ylab("농도")

```

월별 PM10 농도 변화

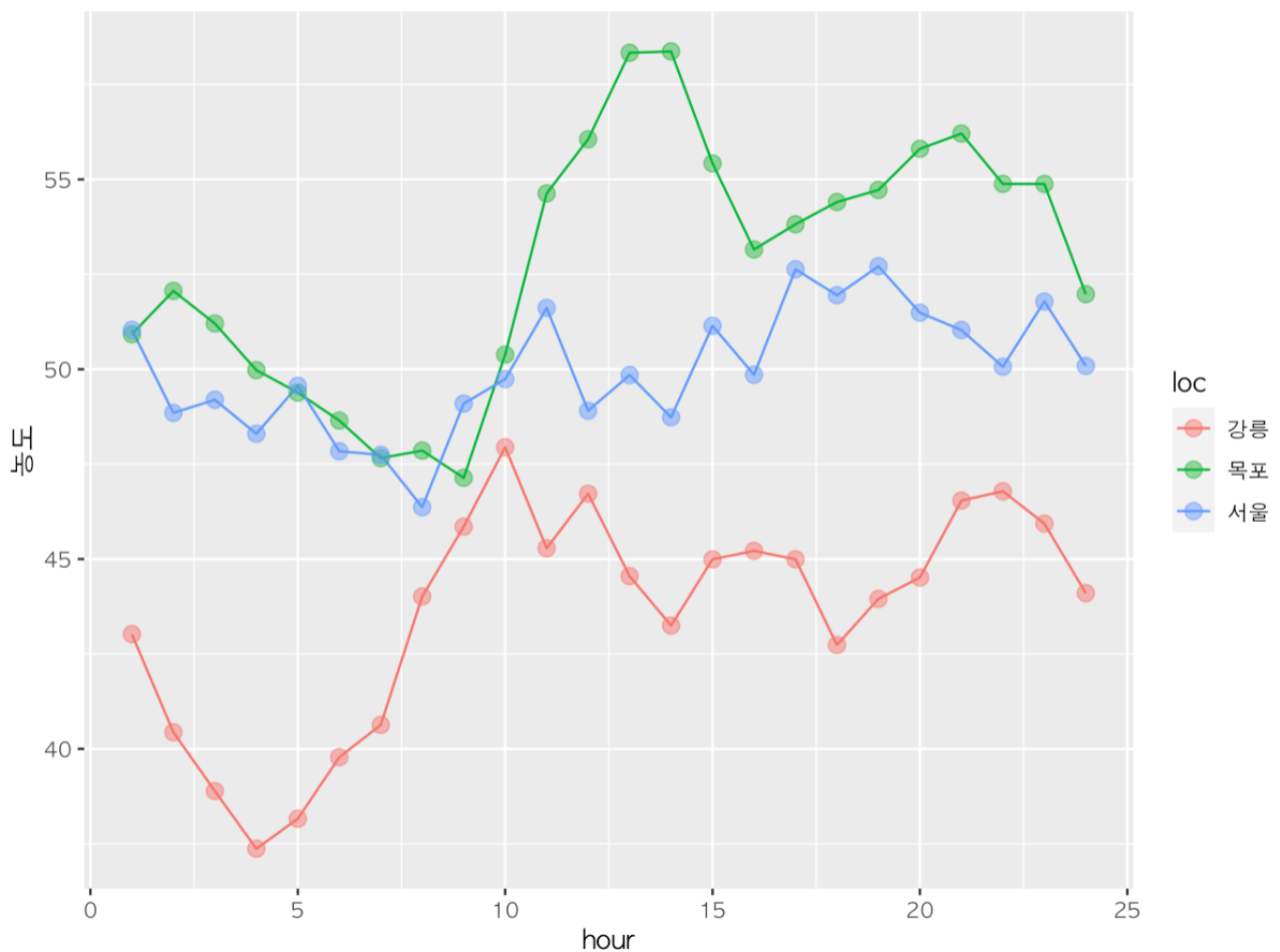



```

> tmp.hour <- aggregate(ds_new[,7],
+                       by=list(hour=ds_new$hour,
+                               loc=ds_new$locname), FUN='mean') #PM10 미세먼지의 시간대별 평균
> tmp.hour$loc = as.factor(tmp.hour$loc)
> head(tmp.hour)
  hour loc      x
1    1 강릉 43.02055
2    2 강릉 40.43624
3    3 강릉 38.89041
4    4 강릉 37.36913
5    5 강릉 38.15894
6    6 강릉 39.78000
> ggplot(tmp.hour, aes(x=hour,y=x, colour=loc, group=loc))+
+   geom_line( )+
+   geom_point(size=3, shape=19, alpha=0.5)+
+   ggtitle('시간별 PM10 농도 변화')+ylab('농도')

```

시간별 PM10 농도 변화



```

> set.seed(100)
> plot(ds_new[sample(nrow(ds_new),5000),3:7], lower.panel=NULL)
> cor(ds_new[,3:7])
      S02      CO      O3      NO2      PM10
S02  1.0000000  0.3574590 -0.120743280  0.46576647  0.058865706
CO   0.35745898  1.0000000 -0.478062707  0.77890229  0.146717661
O3  -0.12074328 -0.4780627  1.000000000 -0.63347444  0.005997464
NO2  0.46576647  0.7789023 -0.633474443  1.000000000  0.052656834
PM10 0.05886571  0.1467177  0.005997464  0.05265683  1.000000000
> idx_max <- which(tmp.month$x == max(tmp.month$x)) #가장 미세먼지가 많았던 날
> tmp.month[idx_max,]
  month loc      x
9      3 목포 94.92143
> idx_min <- which(tmp.month$x == min(tmp.month$x)) # 가장 미세먼지가 적었던날
> tmp.month[idx_min,]
  month loc      x
6      6 강릉 27.76969
>

```

