

2019.01.18 (ZEPPELIN SPARK.R TEST)

2019년 1월 18일 금요일 오전 12:04

spark R 한글 설정

<https://serverfault.com/questions/616790/how-to-add-language-support-on-centos-7-on-docker>

```
[root@dn01 ~]# yum -y -q reinstall glibc-common #설치
```

```
[root@dn01 ~]#vi /etc/yum.conf
```

```
[root@dn01 ~]#locale -a
```

```
[root@dn01 ~]#localectl list-locales | grep ^ko
```

```
[root@dn01 ~]#localectl set-locale LANG=ko_KR.utf8
```

```
%spark.r
```

```
x <- read.df("/user/hadoop/innisfreeOutput1/part-r-00000", source = "csv",inferSchema='true');
```

```
str(x);
```

```
xx <- as.data.frame(x, row.names = NULL, optional = FALSE);
```

```
str(xx);
```

```
names(xx)[1]<-paste("pdtname");
```

```
names(xx)[2]<-paste("month");
```

```
names(xx)[3]<-paste("count");
```

```
str(xx);
```

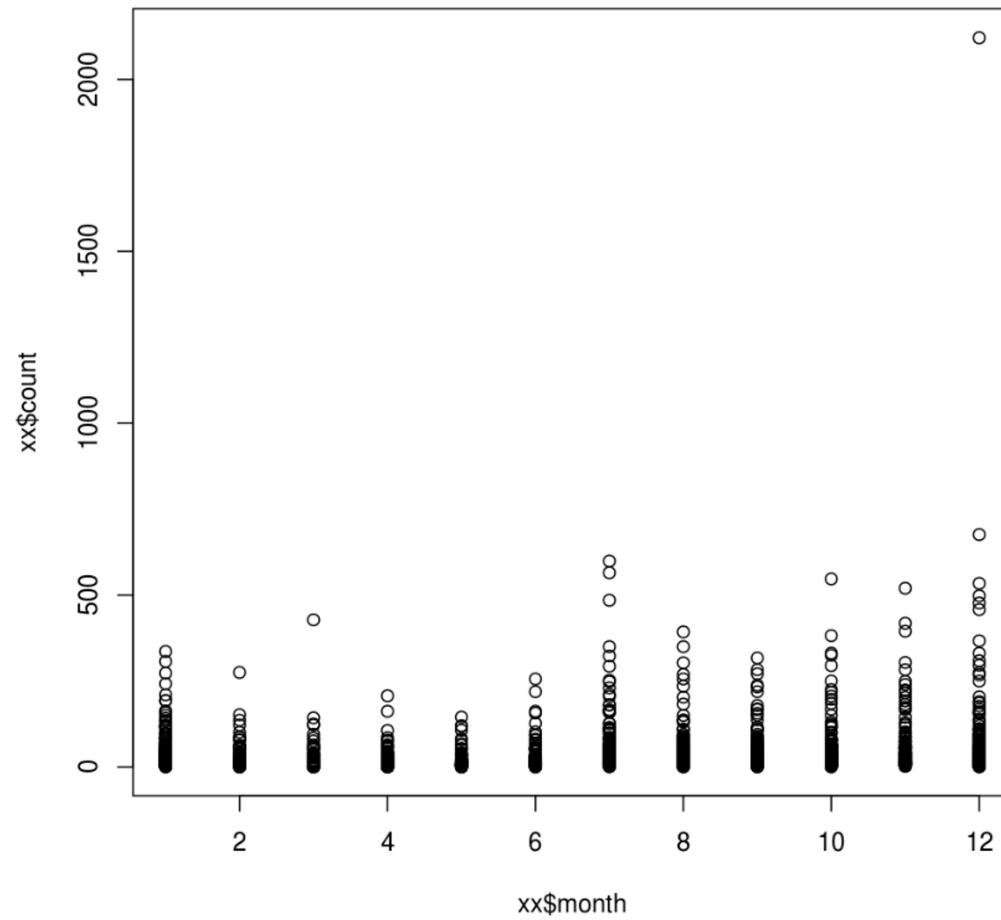
```
library(ggplot2);
```

```
library(plyr);
```

```
library(reshape2);
```

```
plot1 <- plot(xx$month, xx$count);
```

Plot1



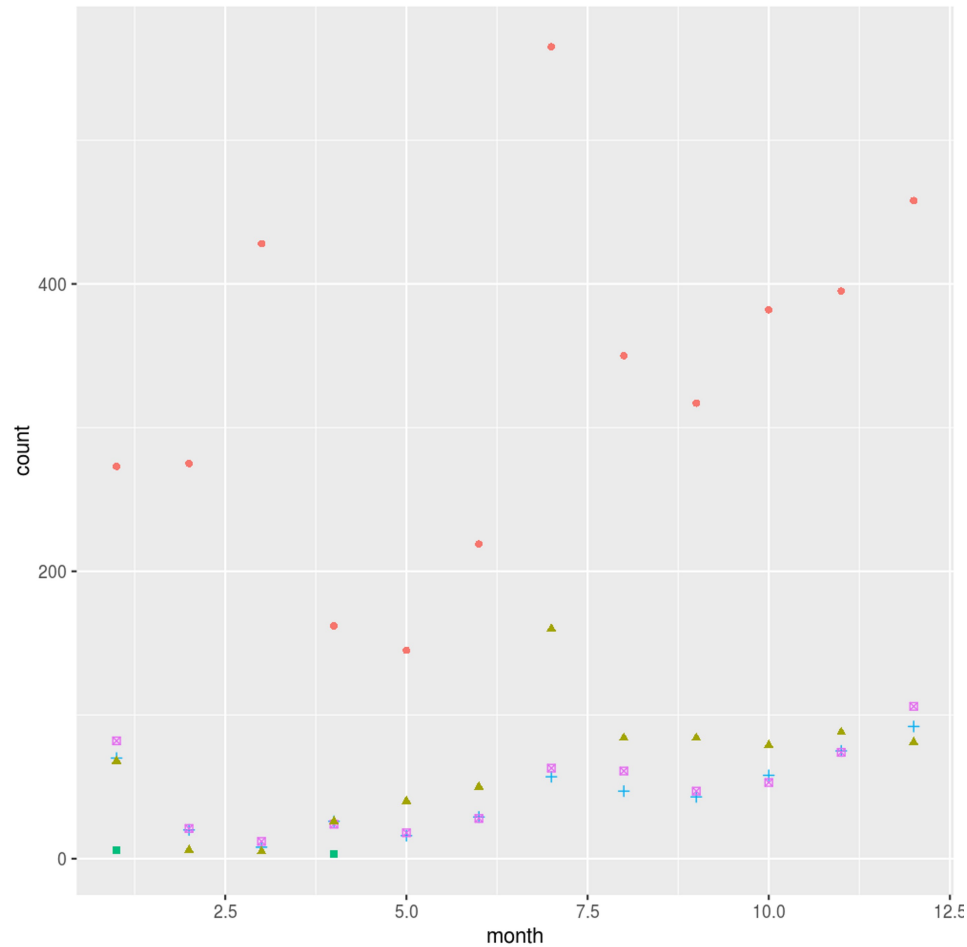
GGPLOT :

- 데이터 프레임(data frame)
- 색상, 크기 같은 외적 요소(aes)
- 점, 선, 모양 같은 기하학적 요소(geoms)
- 통계적 처리 방법(stats)
- aes에서 사용할 스케일(scale)

```
xxx <- xx[1:50,] #사이즈 줄이기
```

```
ggplot1 <- ggplot(xxx, aes(x=month, y=count, col=pdtname, shape=pdtname), size=10) + geom_point() + theme(legend.position = "none") ;
```

Ggplot

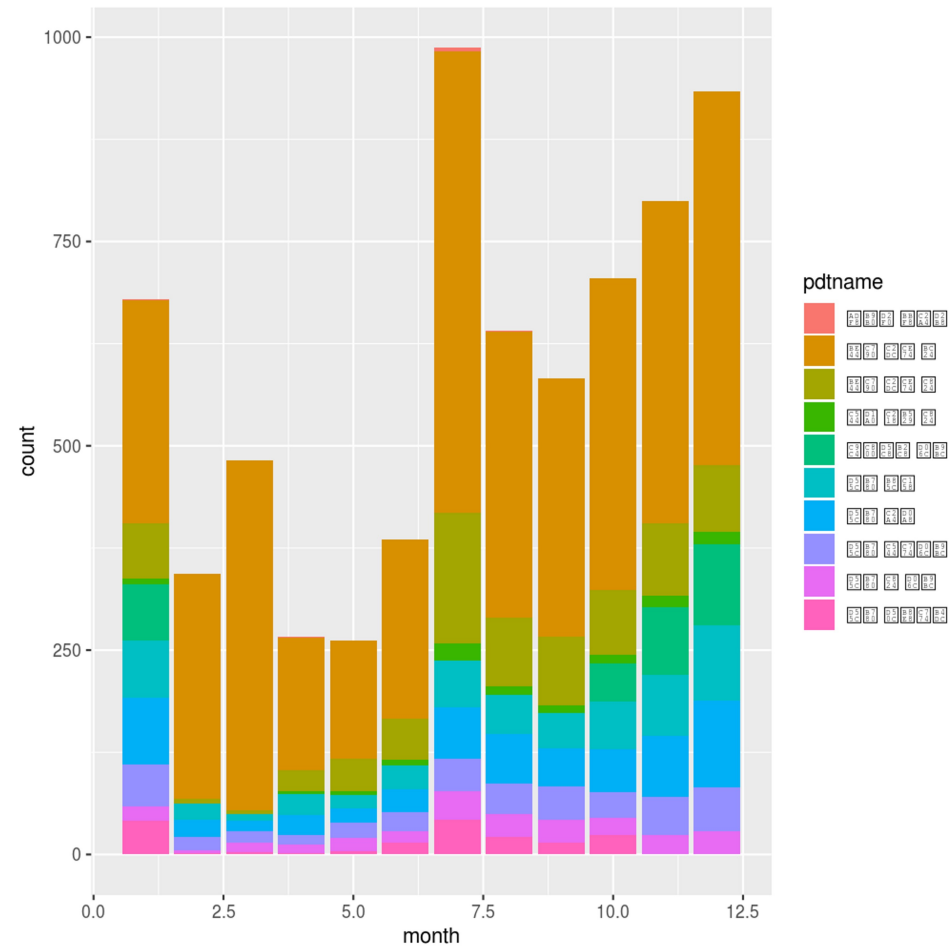


```
xxxx <- xx[1:100,]
```

#“stat” 요소를 “identity” (데이터 프레임의 값을 그대로 사용해서 그래프를 그리는 뜻)

```
g3 <- ggplot(xxxx, aes(month, count, fill=pdtname)) + geom_bar(stat="identity")
```

g3



Col이름 지정 :

```
colnames() = c('pdtname','category','line','volume','price','usergrade','userid','usersex','userage','userdate','reviewpoint','reviewtxt')
```

'SparkDataFrame': 12 variables:

```
$ pdtname      : chr "자연발효 에너지 오일" "자연발효 에너지 오일" "자연발효 에너지 오일" "자연발효 에너지 오일" "자연발효 에너지 오일" "자연발효 에너지 오일" "자연발효 에너지 오일"
$ category     : chr "XD" "XD" "XD" "XD" "XD" "XD"
$ line        : chr "자연발효" "자연발효" "자연발효" "자연발효" "자연발효" "자연발효" "자연발효"
$ volume      : chr "30mL" "30mL" "30mL" "30mL" "30mL" "30mL"
$ price       : chr "28000" "28000" "28000" "28000" "28000" "28000"
$ usergrade   : chr "그린회원" "그린프라임회원" "그린프라임회원" "그린프라임회원" "그린회원" "웰컴회원"
$ userid      : chr " o0i***" " ora**" " sol" " jyy" " tvf**" " als"
$ usersex     : chr "여자" "여자" "여자" "여자" "여자" "여자"
$ usenage     : int 20 30 20 30 20 20
$ userdate    : POSIXct 2019-01-17 2019-01-17 2019-01-16 2019-01-16 2019-01-16 2019-01-15
$ reviewpoint: int 5 5 5 3 5 5
$ reviewtxt   : chr "사워후에 목실에서 바르고 나와요 오일인데 너무 오일리하지도 않으면서 촉촉해요 순하구요 향도 좋아요 강추" "정말 몇 병째인지 모르겠어요!! 세일 할때마다 거의 구입하는 편이에요!"
```