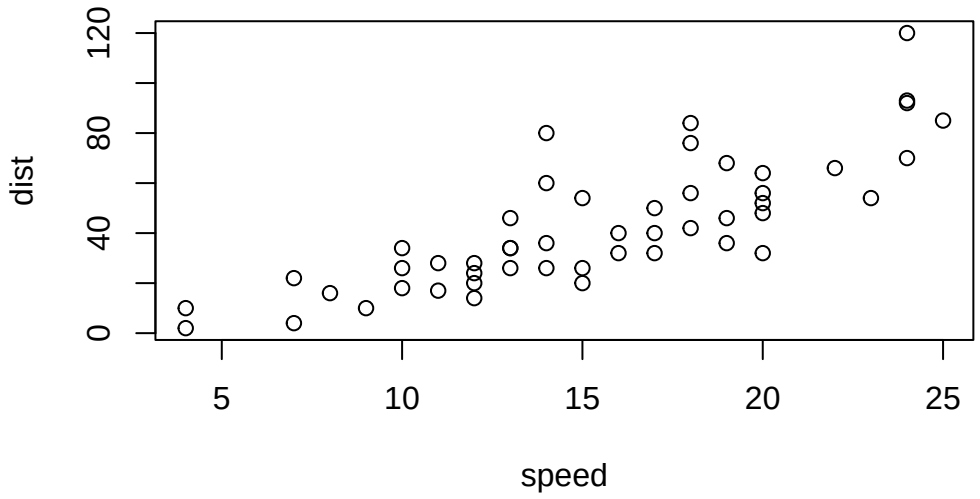
class05

Rita

Plotting in R

R has the multiple plotting and graphics systems. The most popular of which is **ggplot2**.

```
plot(cars)
```

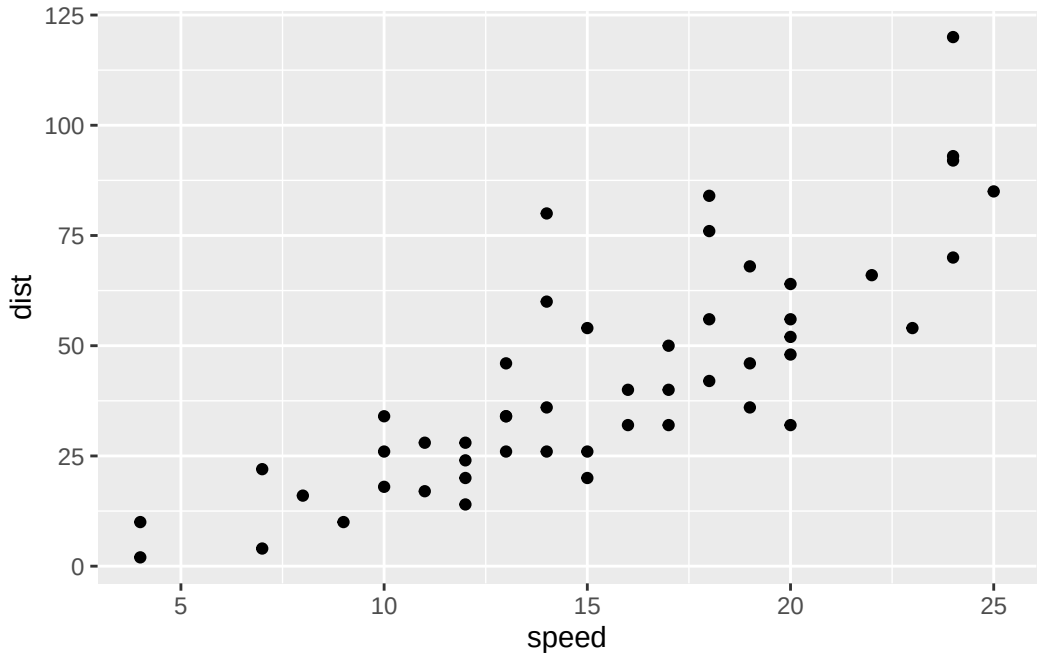


ggplot is much more verbose than plot. Need to install the package.

```
# install.packages("ggplot2")
library(ggplot2)
```

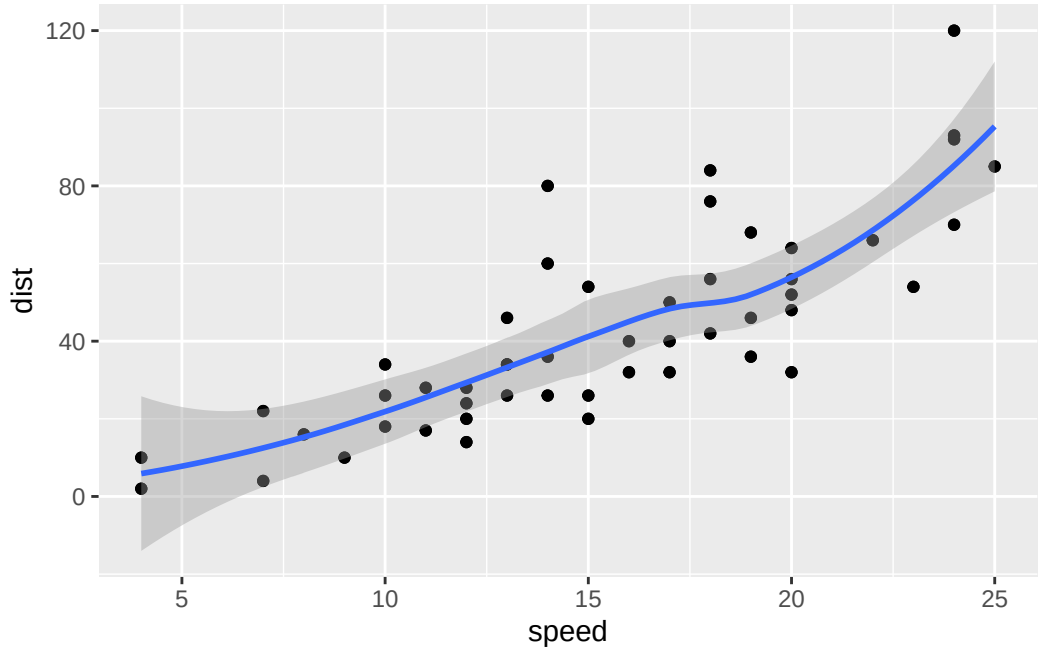
```
bb <- ggplot(cars, aes(speed, dist)) + geom_point()

bb
```




```
bb +
  aes(x=speed, y=dist) +
  geom_point() +
  geom_smooth()
```

```
`geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

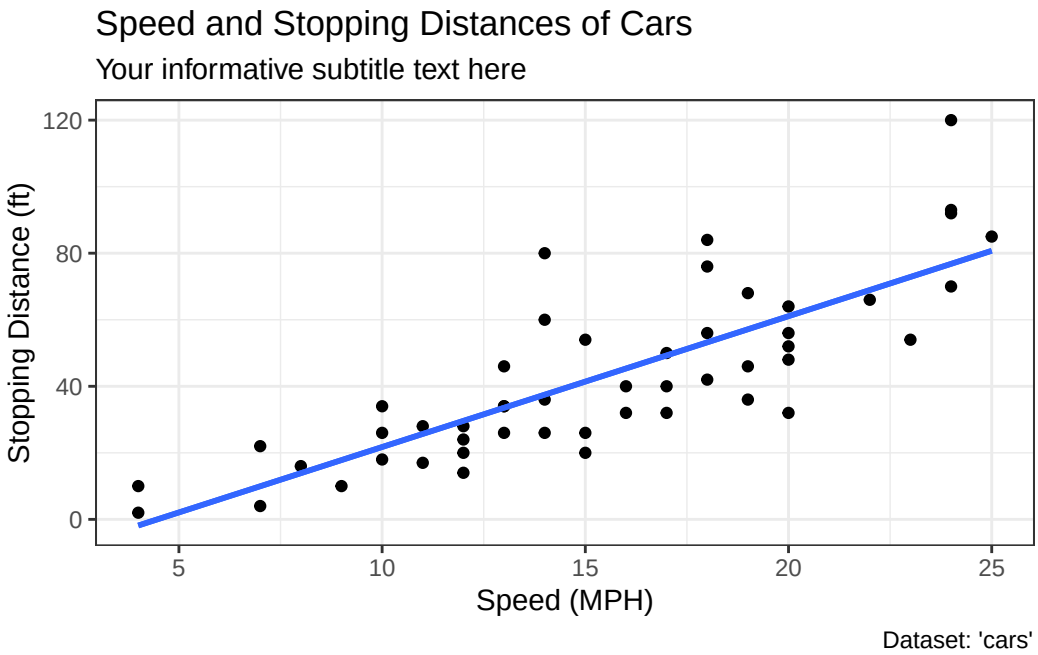


ggplot needs three things (data, aes(x,y)) + geom

```
cc<-bb + geom_smooth(method='lm', se=FALSE)
```

```
cc+ labs(title="Speed and Stopping Distances of Cars",
       x="Speed (MPH)",
       y="Stopping Distance (ft)",
       subtitle = "Your informative subtitle text here",
       caption="Dataset: 'cars'") +
  geom_smooth(method="lm", se=FALSE) +
  theme_bw()
```

```
`geom_smooth()` using formula = 'y ~ x'
`geom_smooth()` using formula = 'y ~ x'
```



look at the top rows of the genes dataframe.

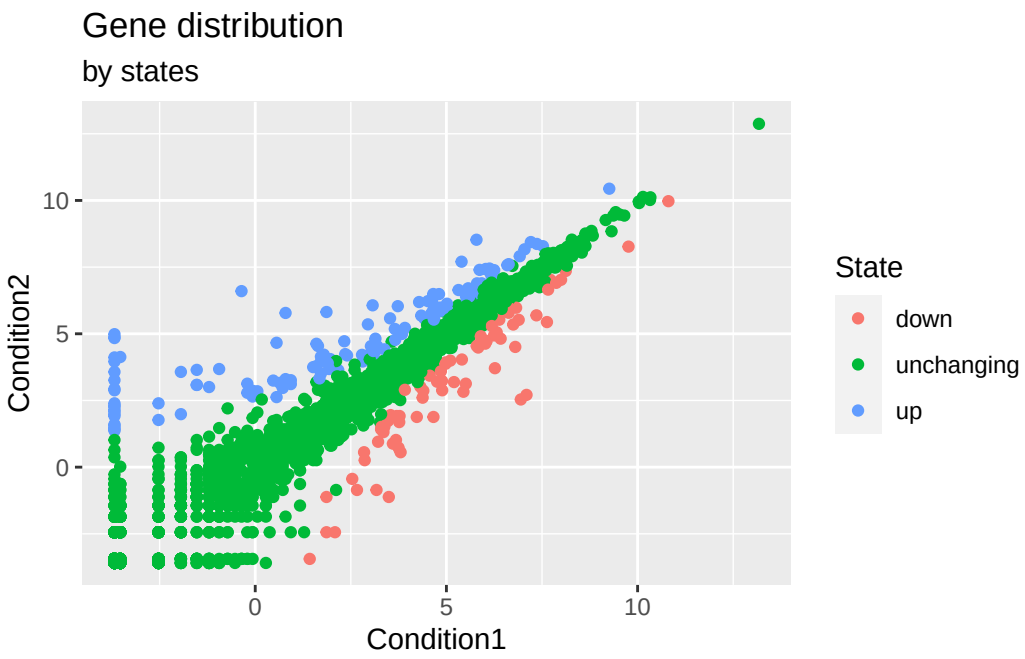
```
url<-"https://bioboot.github.io/bimm143_W23/class-material/up_down_expression.txt"
genes<-read.delim(url)

head(genes)
```

```
        Gene Condition1 Condition2      State
1      A4GNT -3.6808610 -3.4401355 unchanging
2       AAAS  4.5479580  4.3864126 unchanging
3      AASDH  3.7190695  3.4787276 unchanging
4       AATF  5.0784720  5.0151916 unchanging
5       AATK  0.4711421  0.5598642 unchanging
6 AB015752.4 -3.6808610 -3.5921390 unchanging
```

plot the data

```
ggplot(genes) +
    aes(x=Condition1, y=Condition2,color=State) +
    geom_point() +
    labs(title="Gene distribution", subtitle="by states")
```



check the number rows and columns

```
nrow(genes)
```

```
[1] 5196
```

```
ncol(genes)
```

```
[1] 4
```

```
table(genes$State)
```

```
      down unchanging         up
        72       4997        127
```

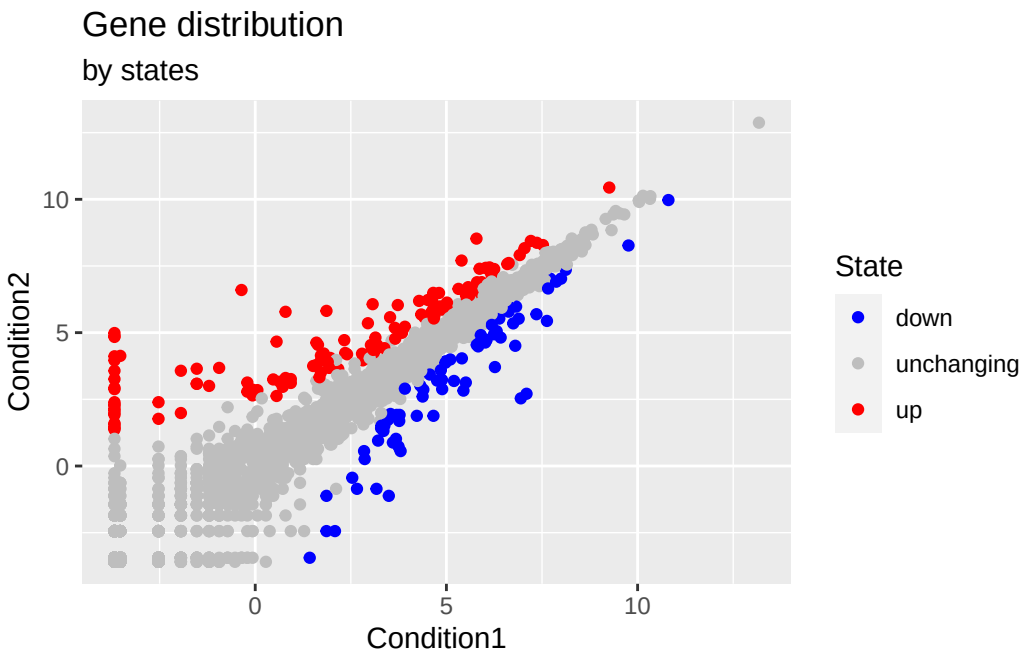
```
round( table(genes$State)/nrow(genes) * 100, 2 )
```

```
      down unchanging         up
      1.39      96.17       2.44
```

This is the percentage of each categories

```
p<-ggplot(genes) +
    aes(x=Condition1, y=Condition2,color=State) +
    geom_point() +
    labs(title="Gene distribution", subtitle="by states")

p+scale_colour_manual( values=c("blue","gray","red") )
```



Extra practice: The gap minder dataset

```
url <- "https://raw.githubusercontent.com/jennybc/gapminder/master/inst/extdata/gapminder.

gapminder <- read.delim(url)
```

```
library(dplyr)
```

```
Attaching package: 'dplyr'
```

```
The following objects are masked from 'package:stats':
```

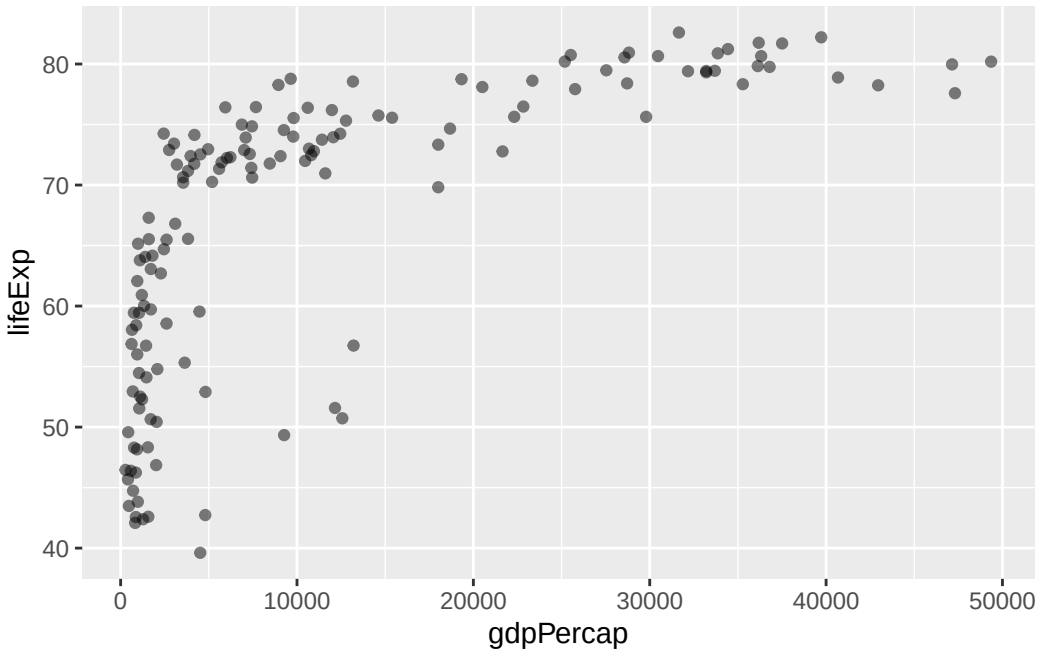
```
    filter, lag
```

```
The following objects are masked from 'package:base':
```

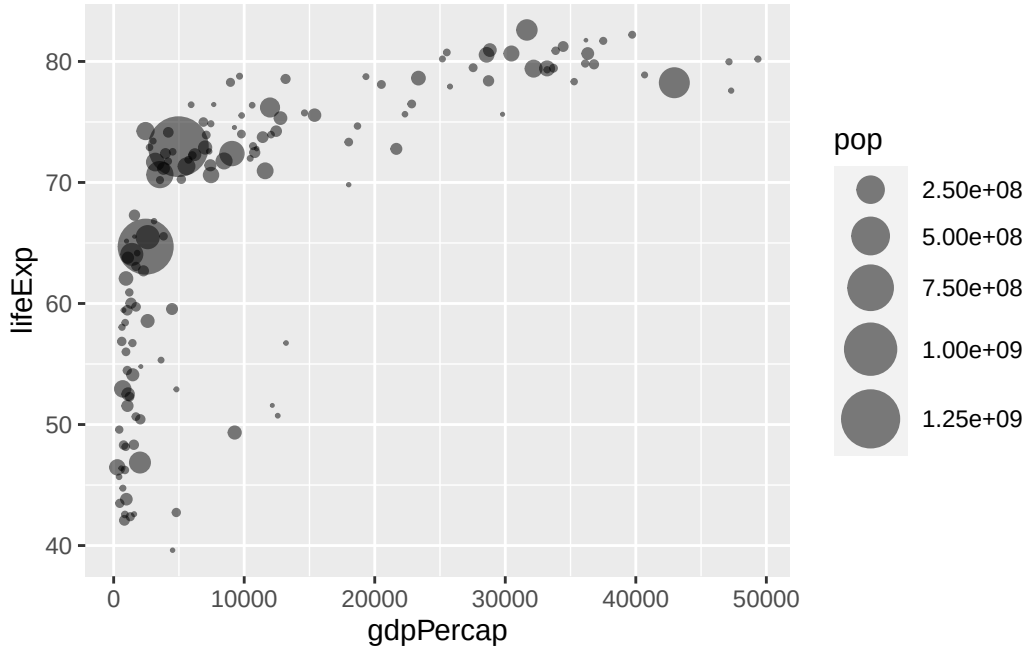
```
    intersect, setdiff, setequal, union
```

```
gapminder_2007 <- gapminder %>% filter(year==2007)
```

```
ggplot(gapminder_2007) +
  aes(x=gdpPercap, y=lifeExp) +
  geom_point(alpha=0.5)
```

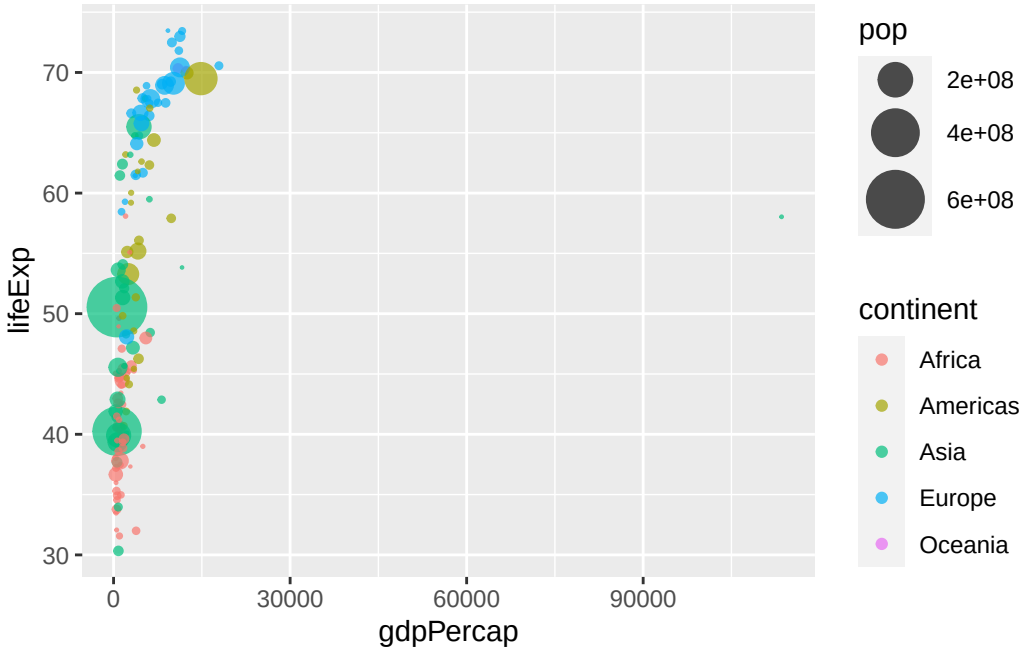


```
ggplot(gapminder_2007) +
  geom_point(aes(x = gdpPercap, y = lifeExp,
                 size = pop), alpha=0.5) +
  scale_size_area(max_size = 10)
```



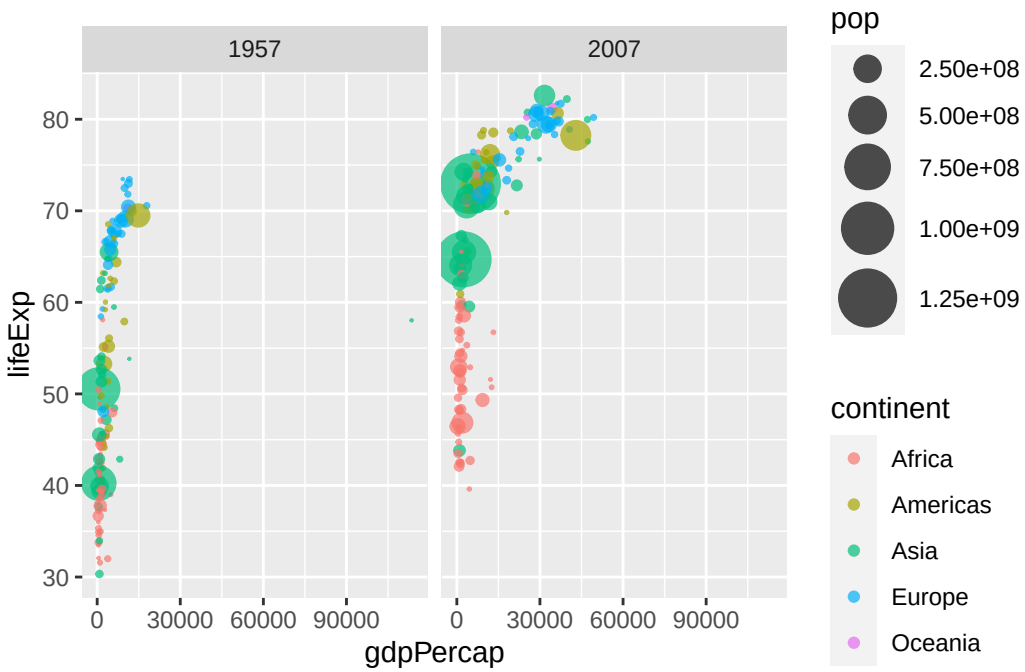

```
gapminder_1957 <- gapminder %>% filter(year==1957)

ggplot(gapminder_1957) +
  aes(x = gdpPercap, y = lifeExp, color=continent,
                 size = pop) +
  geom_point(alpha=0.7) +
  scale_size_area(max_size = 10)
```




```
gapminder_1957 <- gapminder %>% filter(year==1957 | year==2007)

ggplot(gapminder_1957) +
  geom_point(aes(x = gdpPercap, y = lifeExp, color=continent,
                 size = pop), alpha=0.7) +
  scale_size_area(max_size = 10) +
  facet_wrap(~year)
```


1957
2007
lifeExp
80
70
60
50
40
30
0
30000
60000
90000
gdpPercap
pop
2.50e+08
5.00e+08
7.50e+08
1.00e+09
1.25e+09
continent
Africa
Americas
Asia
Europe
Oceania