

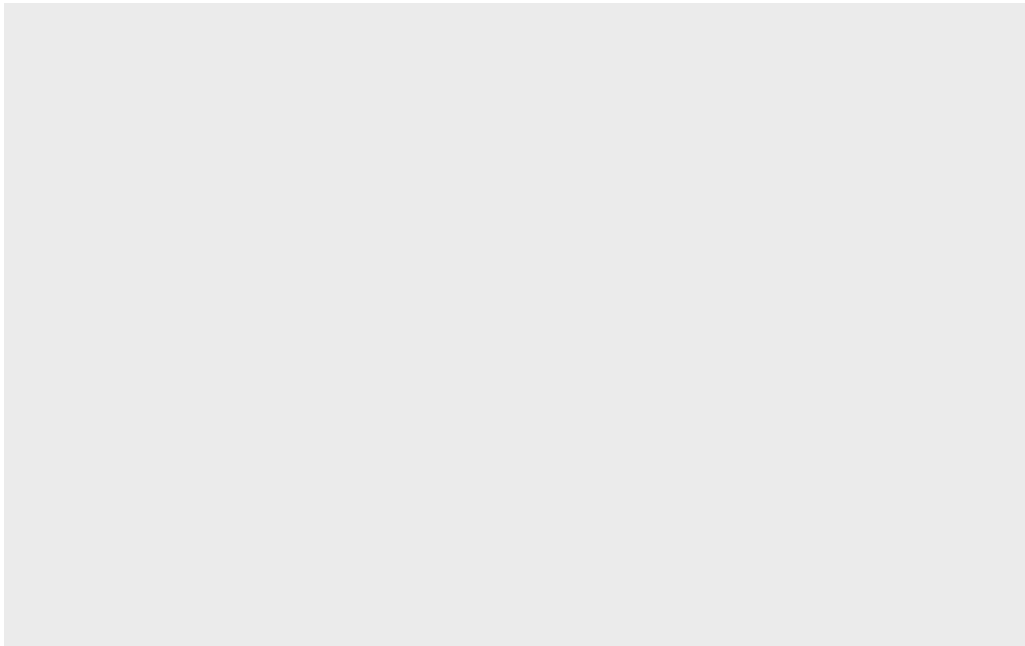
# Class 5: Data Visualization with ggplot

Andy (A12053233)

sections 1-7 ## Using ggplot ## To use ggplot2 we first need to install it on our computers. To do this, we will use the function `install.packages()`.

**Before I use any package functions, I have to load them up with a “library()” call, like so:**

```
library(ggplot2)
ggplot(cars)
```



```
head(cars)
```

	speed	dist
1	4	2
2	4	10
3	7	4
4	7	22
5	8	16
6	9	10

```
tail(cars)
```

	speed	dist
45	23	54
46	24	70
47	24	92
48	24	93
49	24	120
50	25	85

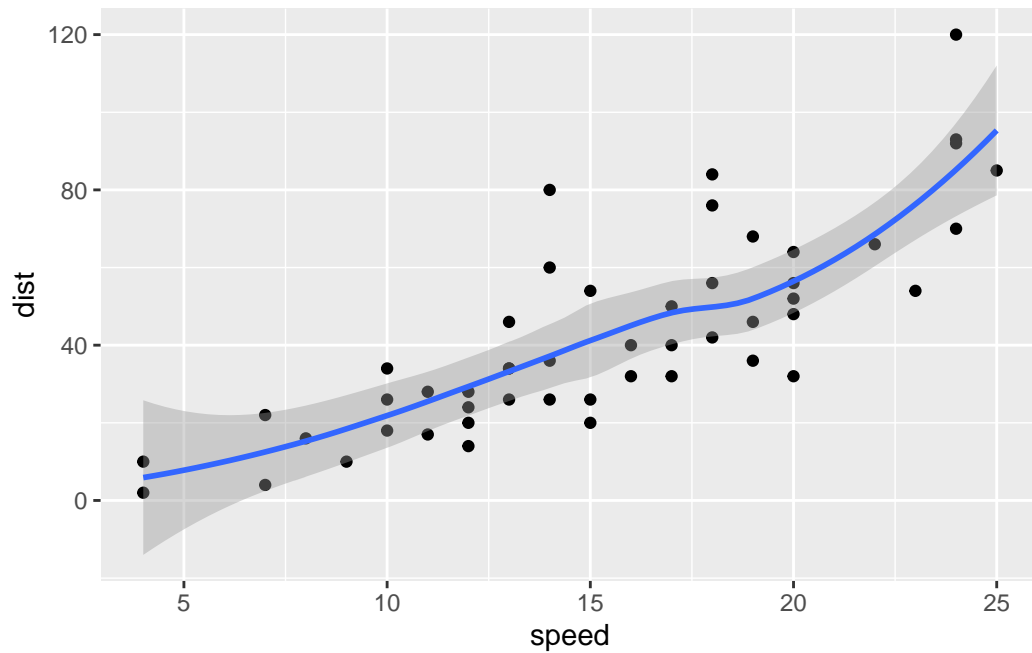
**There is always the “base R” graphics system, i.e. the `plot()` function**

**To use ggplot, I need to spell out at least 3 things:**

- data (the stuff I want to plot, always in the format of a data.frame) -aesthetics (`aes()` values
- how the data map to the plot) -geometries/geoms (how I want things drawn)

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

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

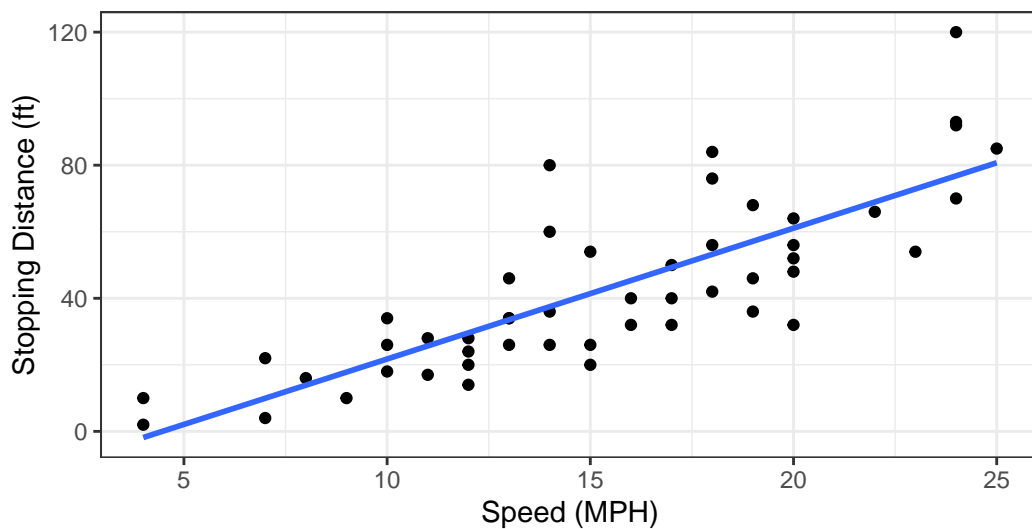


```
ggplot(cars) +
  aes(x=speed, y=dist) +
  geom_point() +
  geom_smooth(method="lm", se=FALSE) +
  labs(title="Speed and Stopping Distance of Cars",
        x="Speed (MPH)",
        y="Stopping Distance (ft)",
        subtitle="Your informative subtitle text",
        caption="Dataset: 'cars'") +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

`geom\_smooth()` using formula = 'y ~ x'

## Speed and Stopping Distance of Cars

Your informative subtitle text



Dataset: 'cars'

```
url <- "https://bioboot.github.io/bimm143_S20/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

```
nrow(genes)
```

```
[1] 5196
```

```
colnames(genes)
```

```
[1] "Gene"      "Condition1" "Condition2" "State"
```

```
ncol(genes)
```

```
[1] 4
```

```
sum(genes$State == "up")
```

```
[1] 127
```

There are 5196 in this dataset There are 127 “Up” regulated genes in this dataset

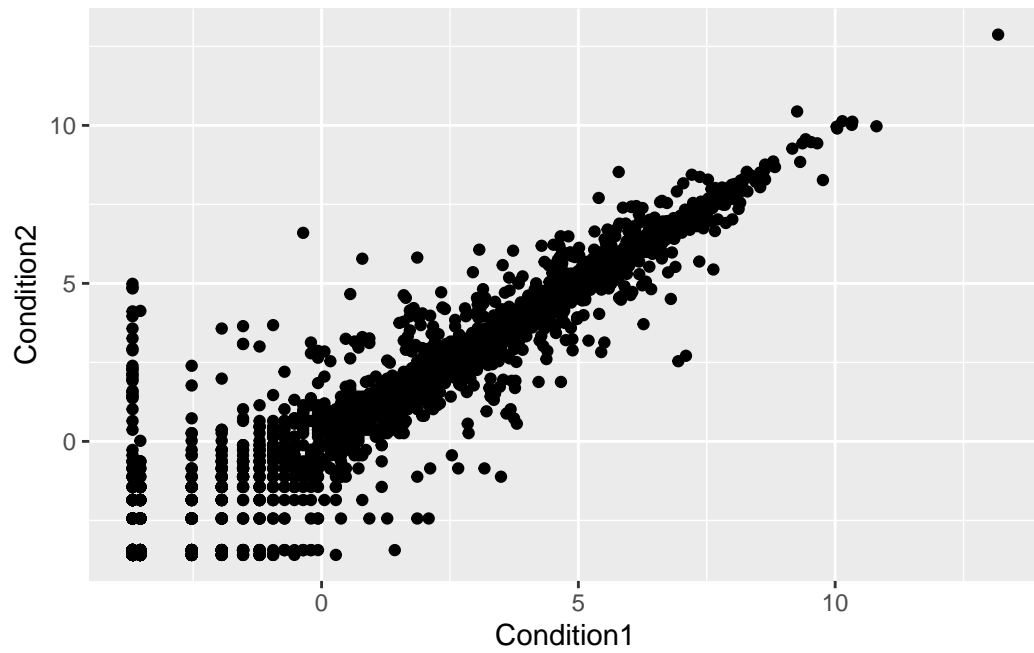
```
table(genes$State)
```

down	unchanging	up
72	4997	127

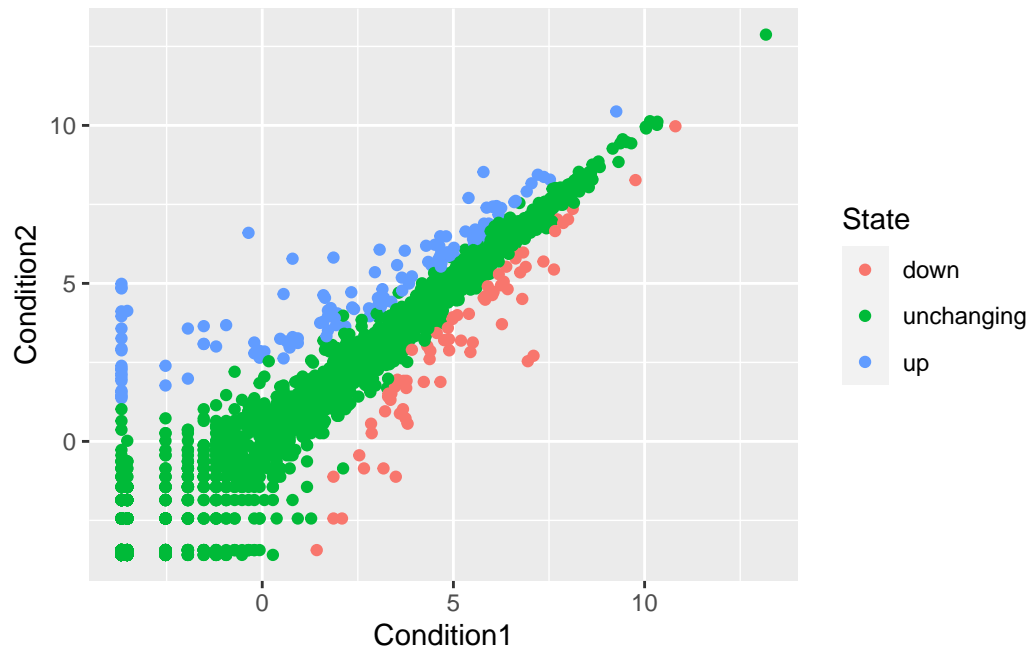
```
round( table(genes$State)/nrow(genes) * 100, 2)
```

down	unchanging	up
1.39	96.17	2.44

```
ggplot(genes) +  
  aes(x=Condition1, y=Condition2) +  
  geom_point()
```

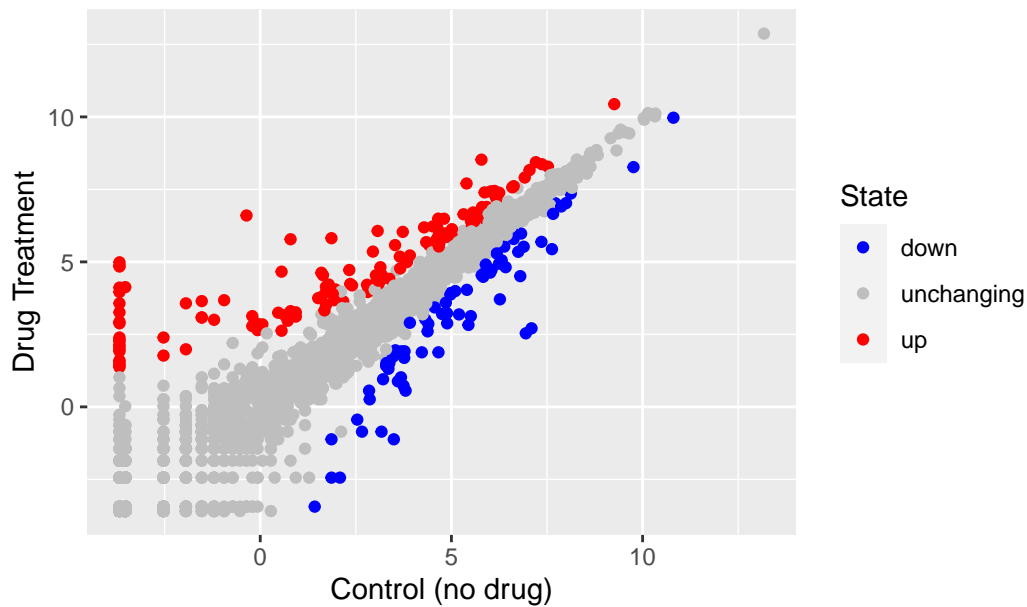


```
p <- ggplot(genes) +  
  aes(x=Condition1, y=Condition2, col=State) +  
  geom_point()  
p
```



```
p + scale_colour_manual( values=c("blue","gray","red")) +
  labs(title="Gene Expression Changes Upon Drug Treatment",
        x="Control (no drug)",
        y="Drug Treatment")
```

## Gene Expression Changes Upon Drug Treatment



```
#install.packages("gapminder")
library(gapminder)
#install.packages("dplyr")
library(dplyr)
```

Attaching package: 'dplyr'

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

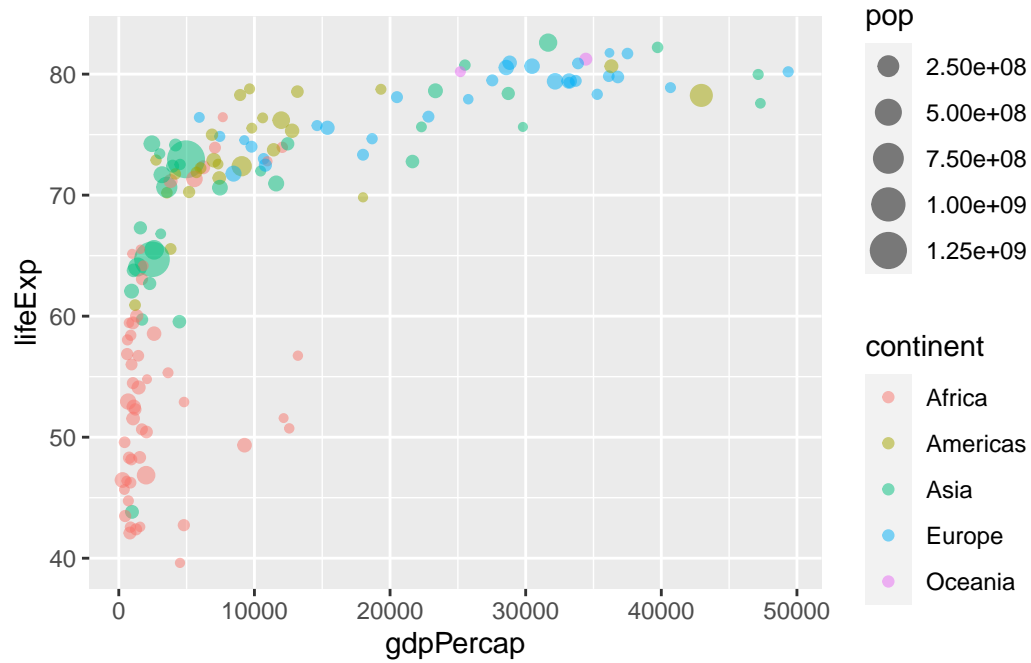
filter, lag

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

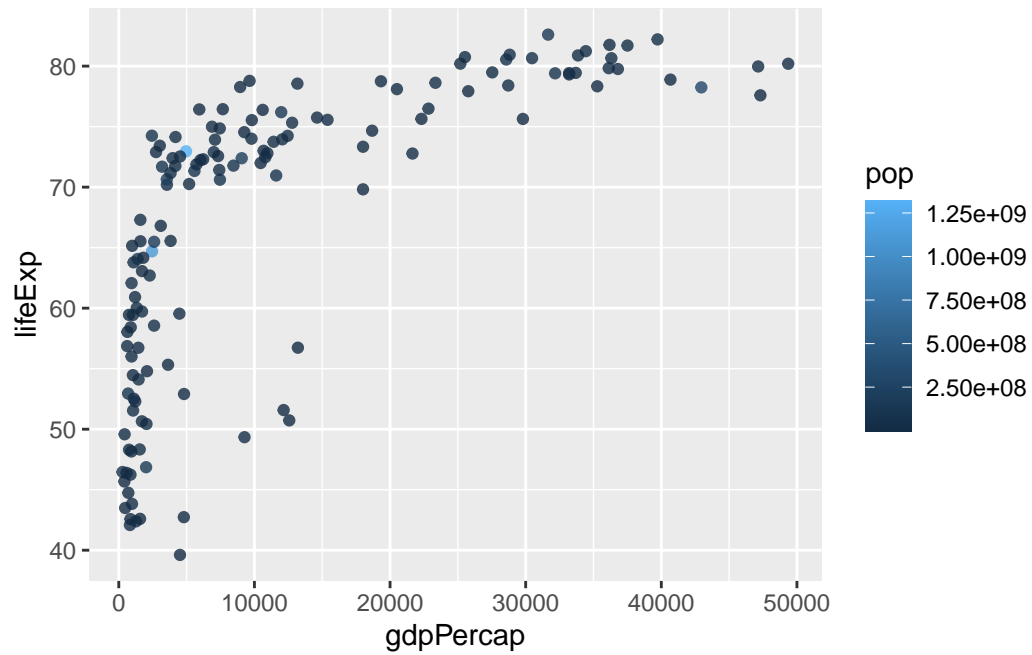
intersect, setdiff, setequal, union

```
gapminder_2007 <- gapminder %>% filter(year==2007)
ggplot(gapminder_2007) +
  aes(x=gdpPercap, y=lifeExp, color=continent, size=pop) +
  geom_point(alpha=0.5)
```

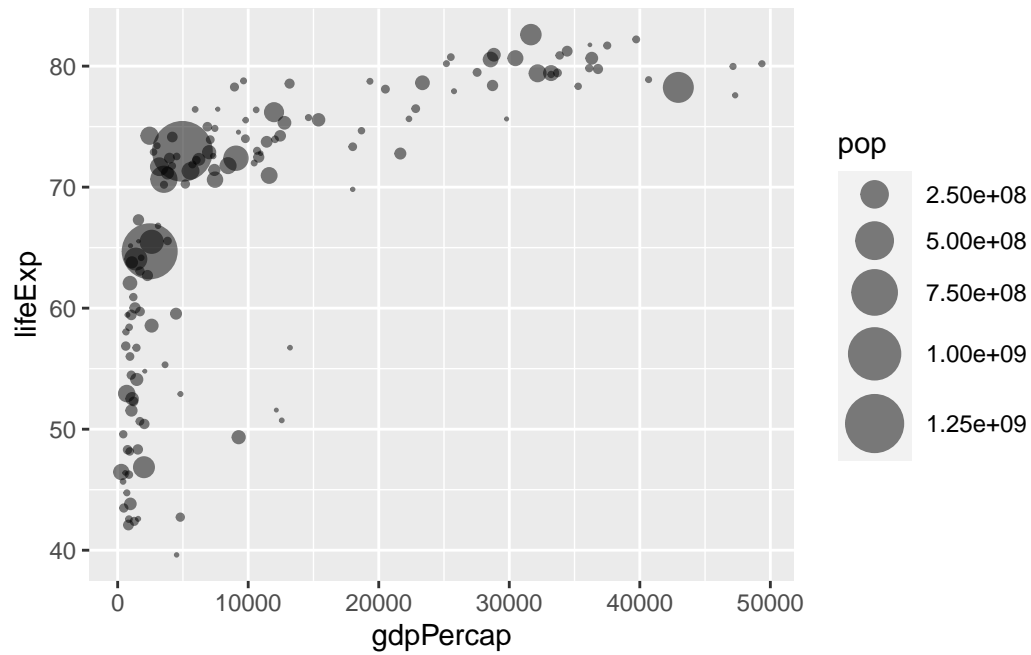




```
ggplot(gapminder_2007) +  
  aes(x=gdpPerCap, y=lifeExp, color=pop) +  
  geom_point(alpha=0.8)
```

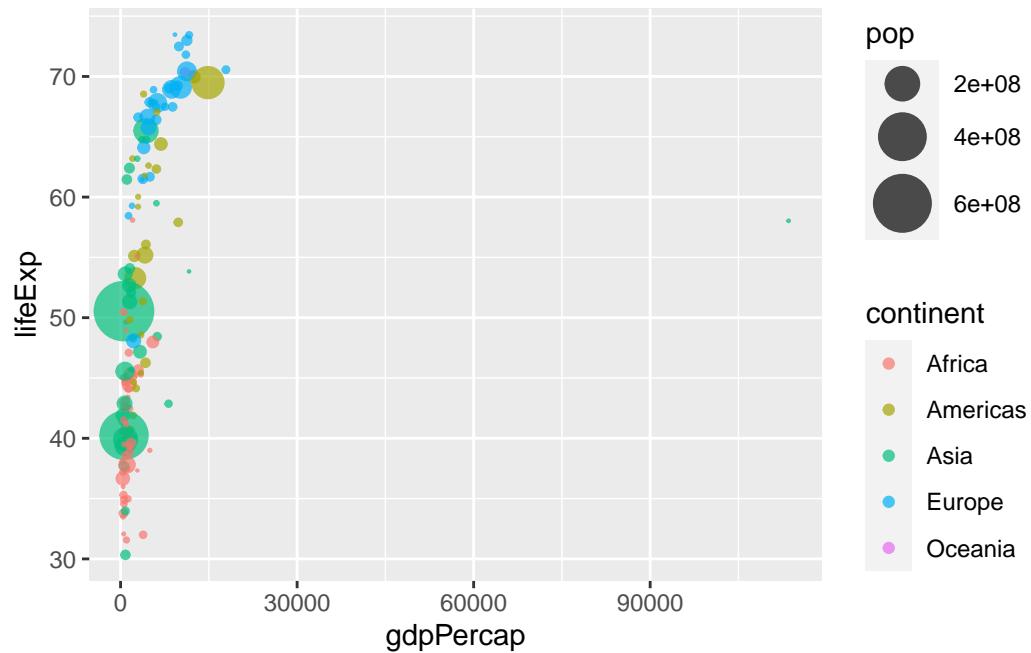


```
ggplot(gapminder_2007) +  
  geom_point(aes(x=gdpPercap, y=lifeExp,  
                 size=pop), alpha=0.5) +  
  scale_size_area(max_size = 10)
```



```
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 = 15) +
  facet_wrap(~year)
```

