

Class05 gene ggplot

Tianru Zhang, PID: A15432834

Oct 12, 2021

```
#class05 Data Visualization
```

```
# scatter plot
```

```
library(ggplot2)
```

```
#every ggplot has three layers: data, aesthetics, geometry
```

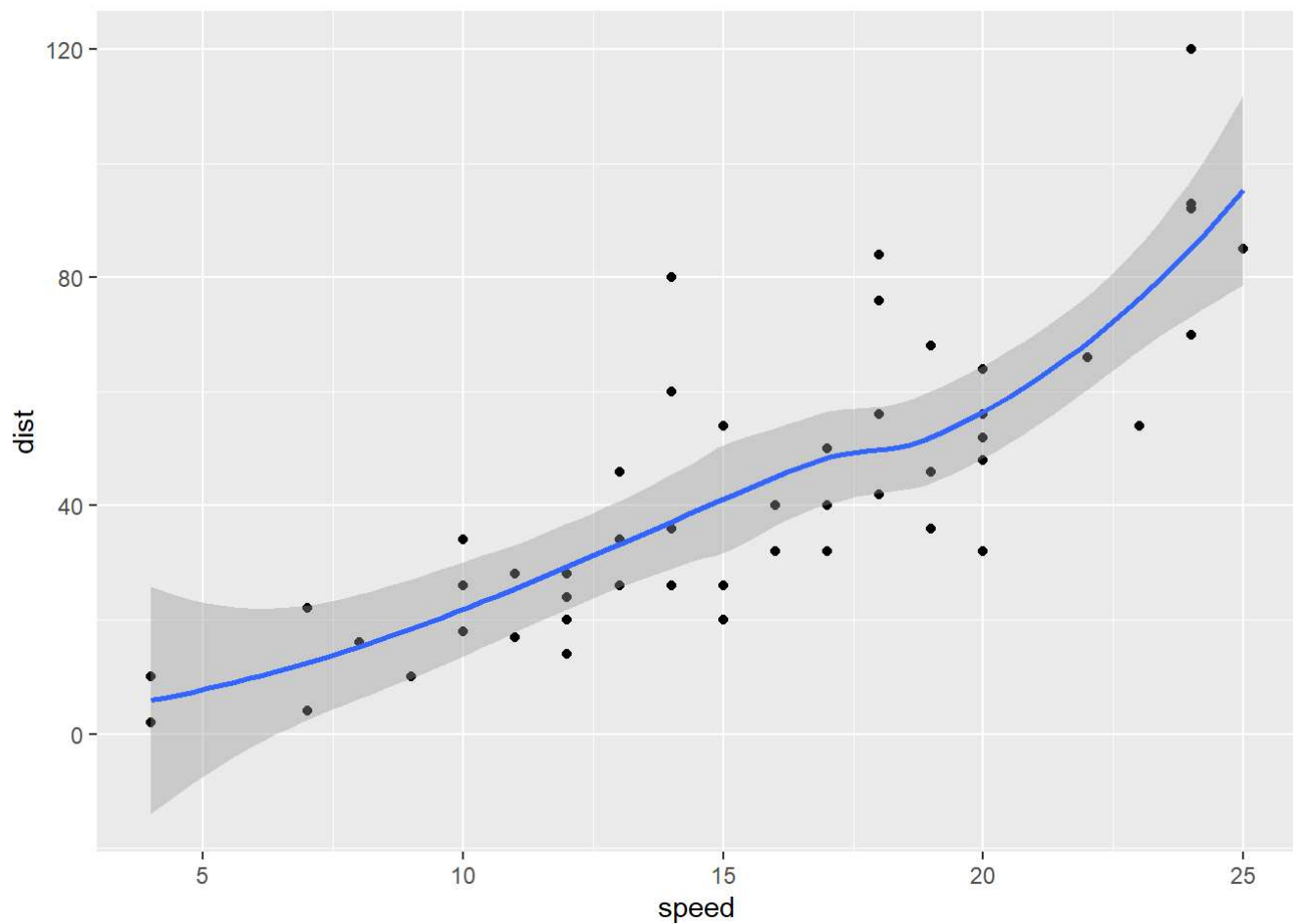
```
ggplot(data=cars) +
```

```
  aes(x=speed, y=dist) +
```

```
  geom_point()+
```

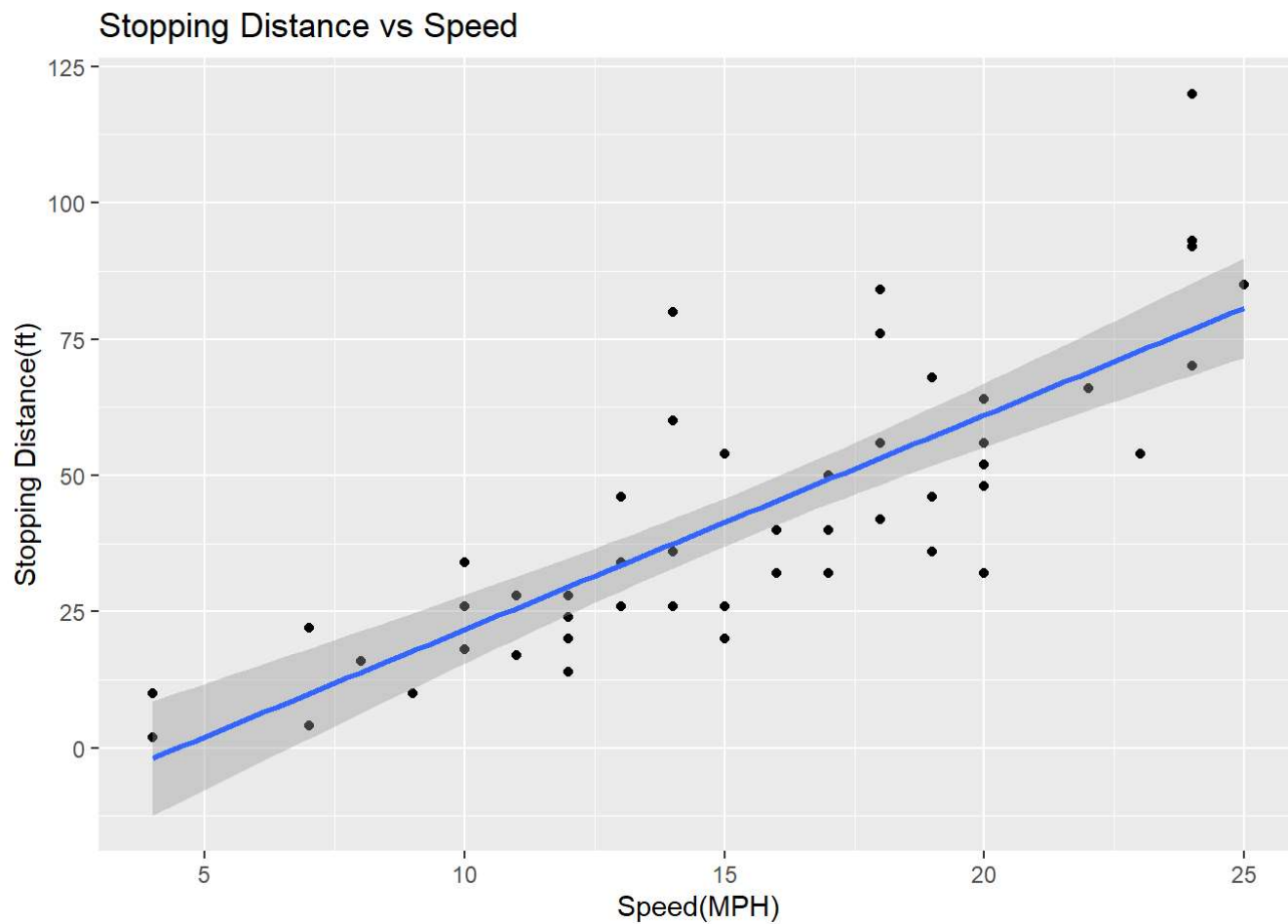
```
  geom_smooth()
```

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

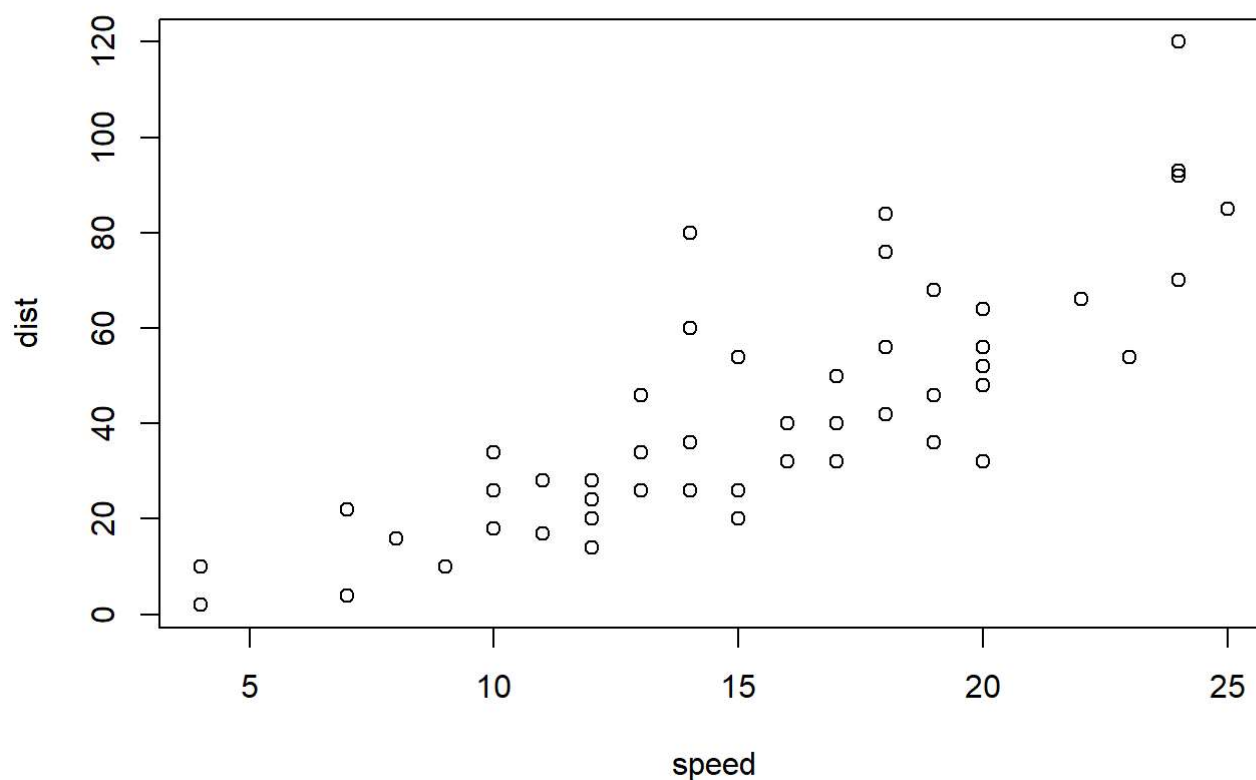


```
#change to linear model
p <- ggplot(data=cars) +
  aes(x=speed, y=dist) +
  geom_point()+
  geom_smooth(method="lm")
p + labs(title="Stopping Distance vs Speed",
  x="Speed(MPH)",
  y="Stopping Distance(ft)")
```

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



```
#Base graphic is shorter
plot(cars)
```



```
#read the dataset first
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
```

```
#Q How many genes
nrow(genes)
```

```
## [1] 5196
```

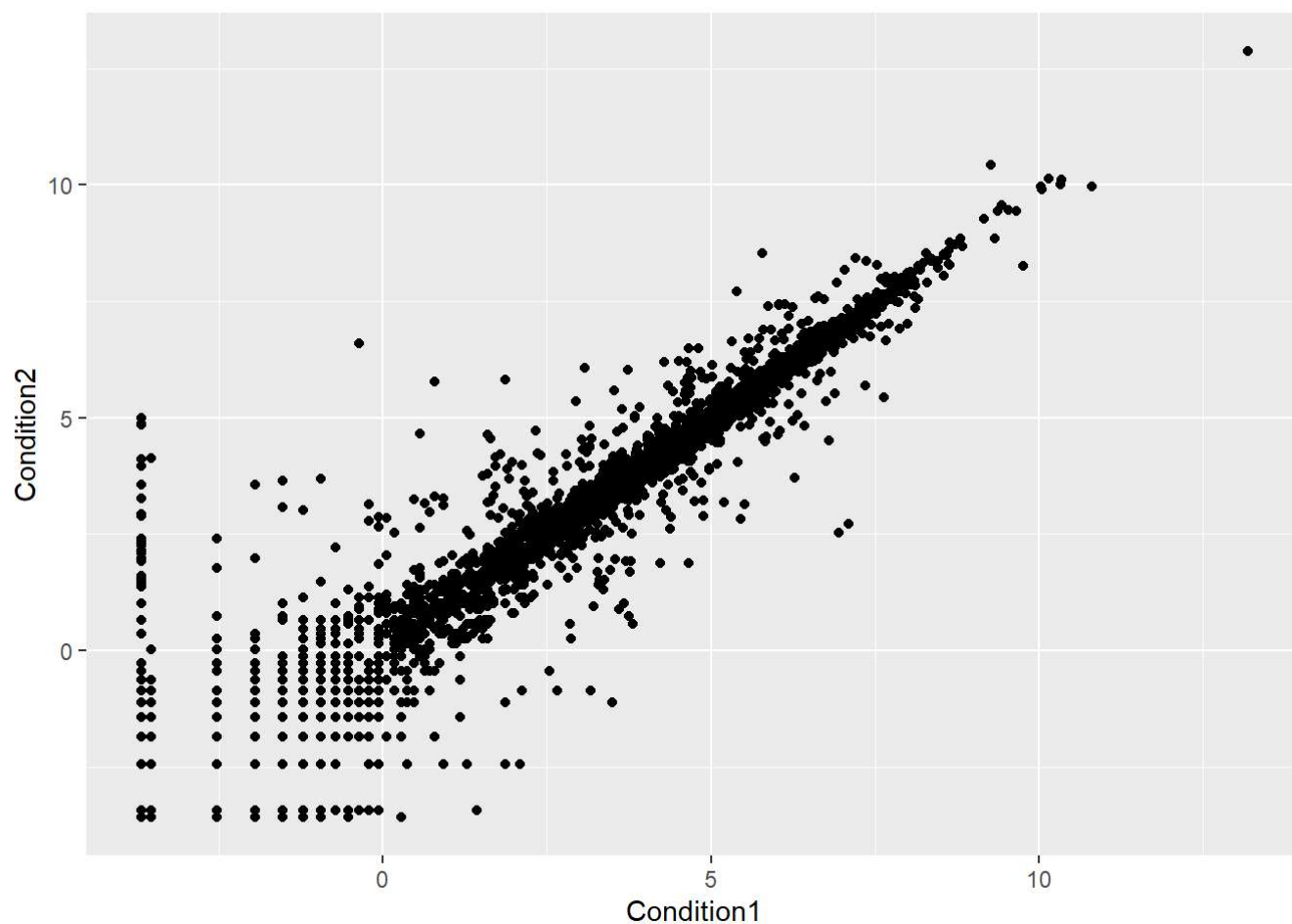
```
#Q how to access state column
table(genes$State)
```

```
##
##      down  unchanged      up
##      72      4997      127
```

```
#Q How many % genes turning up/down
prec<- table(genes$State) /nrow(genes) *100
round(prec, 2)
```

```
##
##      down  unchanged      up
##      1.39      96.17      2.44
```

```
#Generate a plot that shows the gene expression
ggplot(genes) +
  aes(x=Condition1, y=Condition2) +
  geom_point()
```

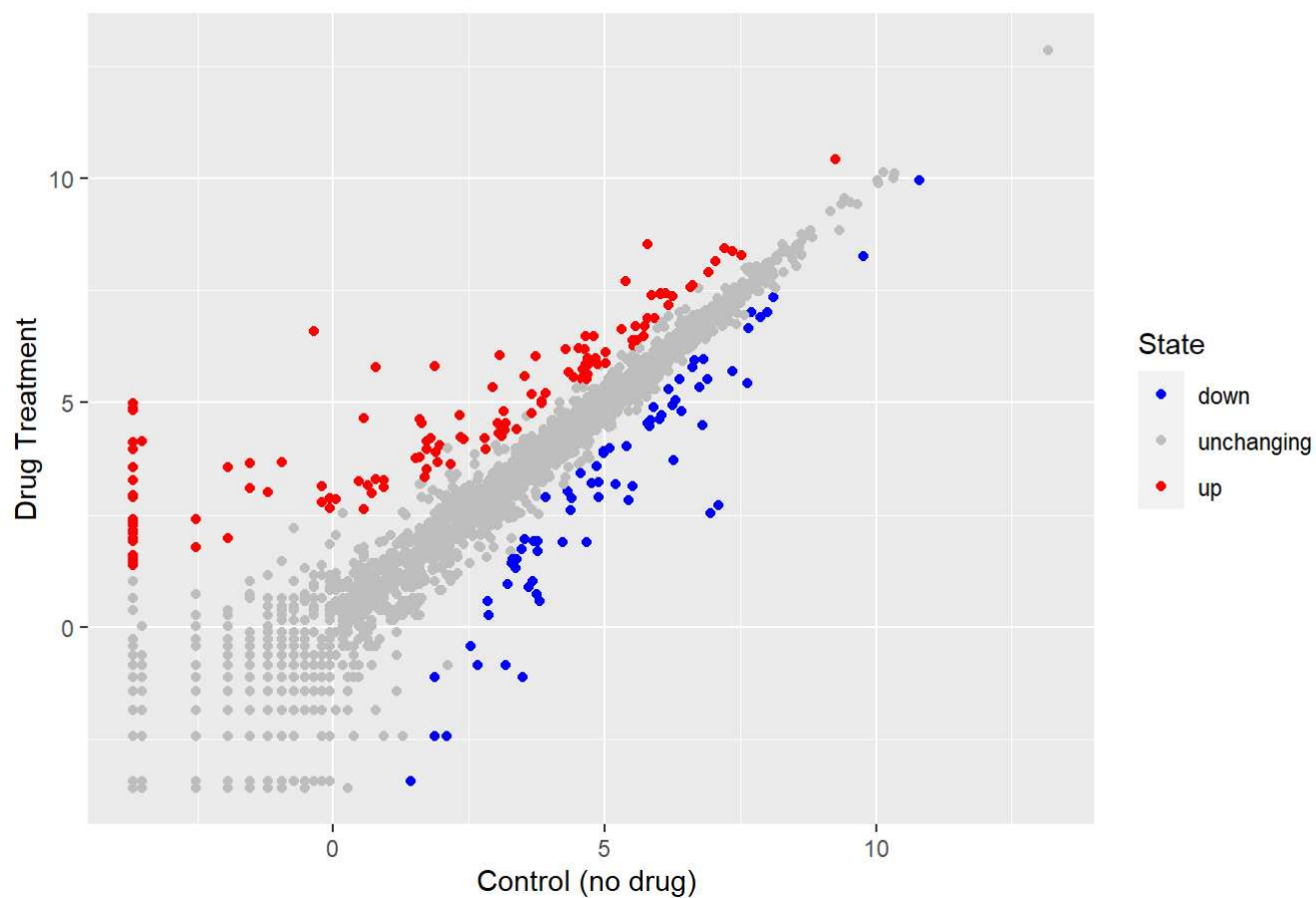


```
# distinguish the genes by their states under drug treatment
g <- ggplot(genes) +
  aes(x=Condition1, y=Condition2, col=State) +
  geom_point()

#change color scales
g <- g + scale_colour_manual( values=c("blue","gray","red") )

#add legends and titles
g+labs(title="Gene Expression Changes Upon Drug Treatment",
       x="Control (no drug) ",
       y="Drug Treatment")
```

Gene Expression Changes Upon Drug Treatment



```
#Focus on gapminder dataset
#install.packages("gapminder")
library(gapminder)

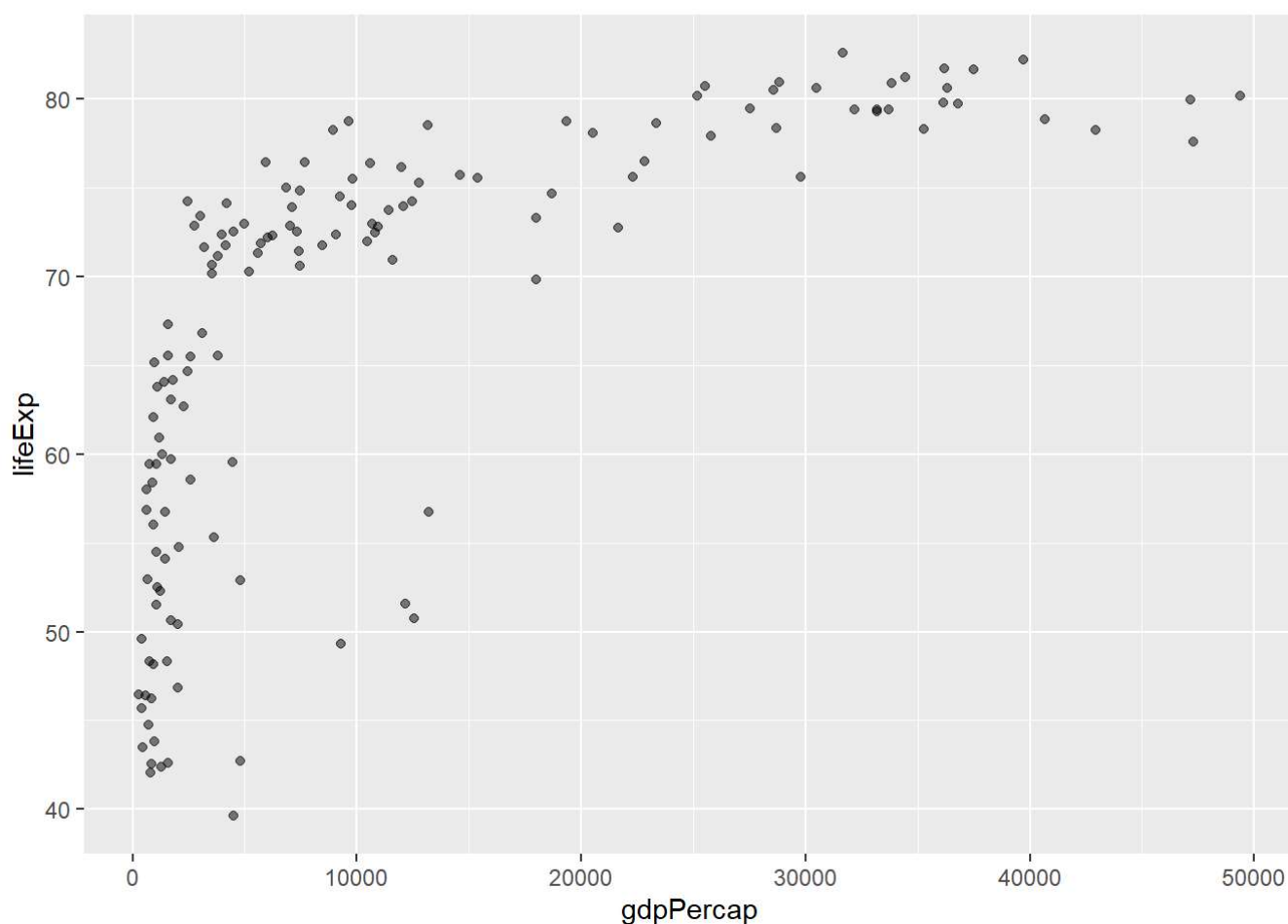
#install.packages("dplyr")
library(dplyr)
```

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

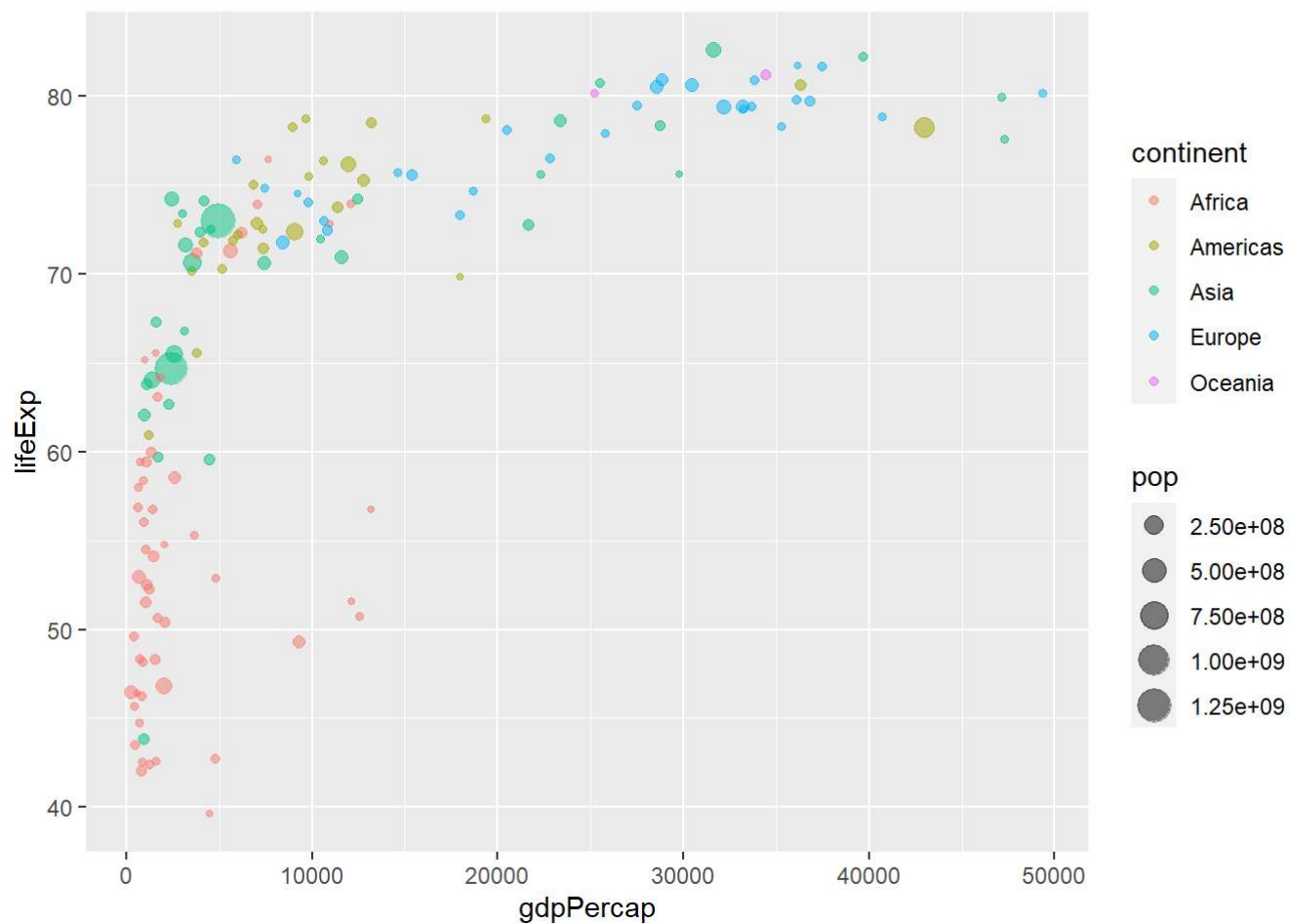
```
## The following objects are masked from 'package:stats':
##
##   filter, lag
```

```
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

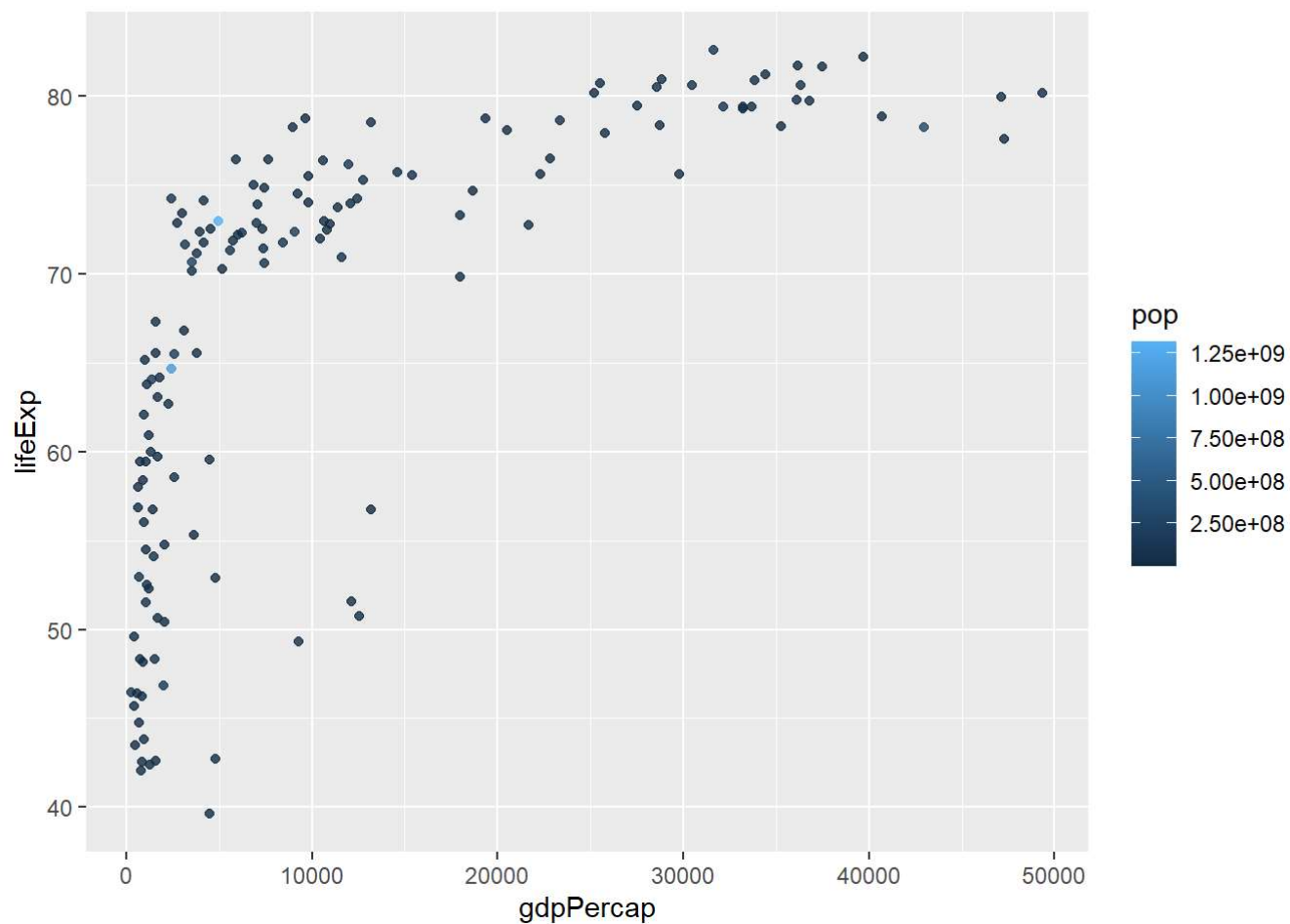
```
gapminder_2007 <- gapminder %>% filter(year==2007)
ggplot(gapminder_2007) +
  aes(x=gdpPercap, y=lifeExp) +
  geom_point(alpha=0.5)
```



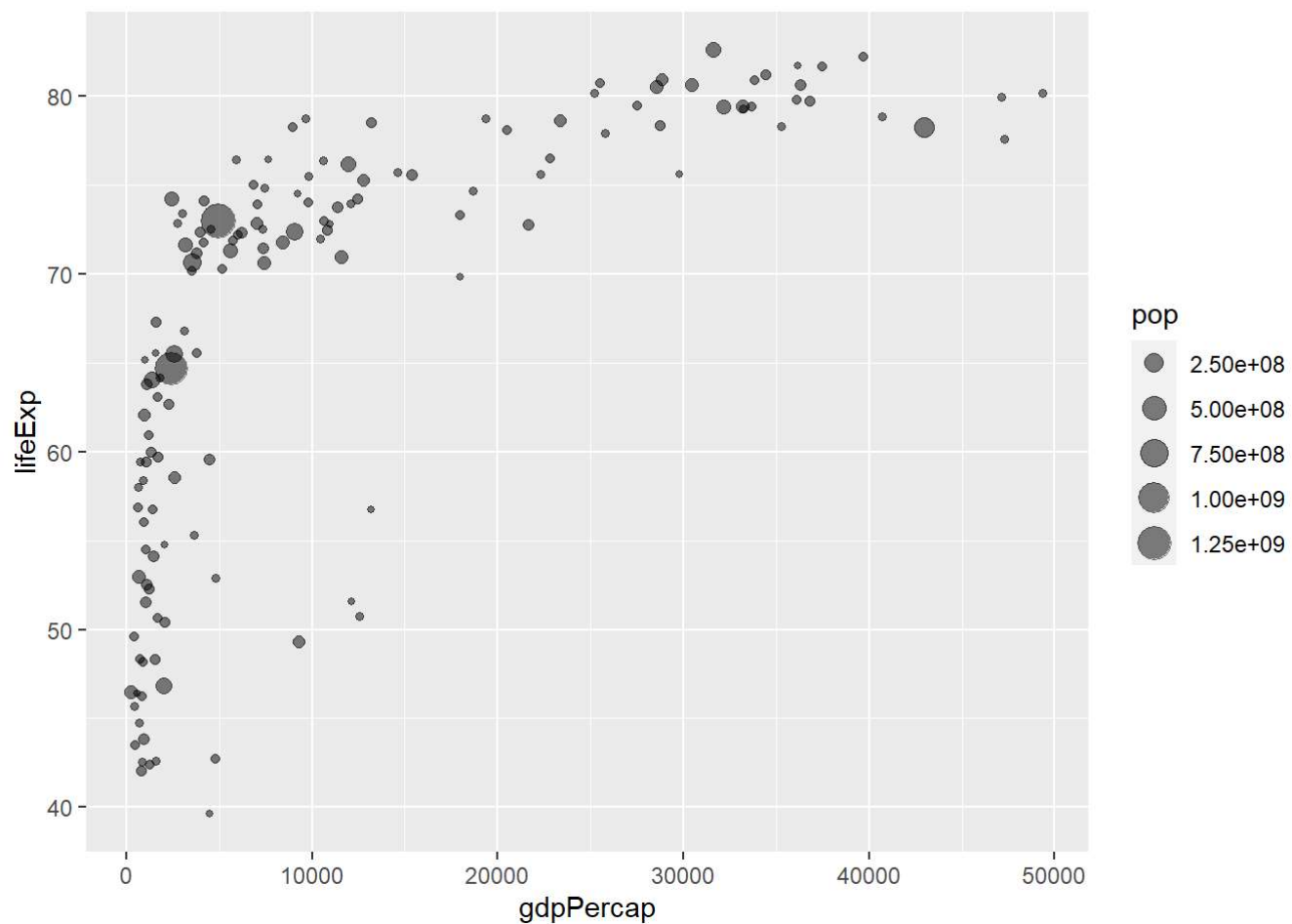
```
# create another plot for the Life expectancy vs GDP: diff col for continents
ggplot(gapminder_2007) +
  aes(x=gdpPercap, y=lifeExp, color=continent, size=pop) +
  geom_point(alpha=0.5)
```



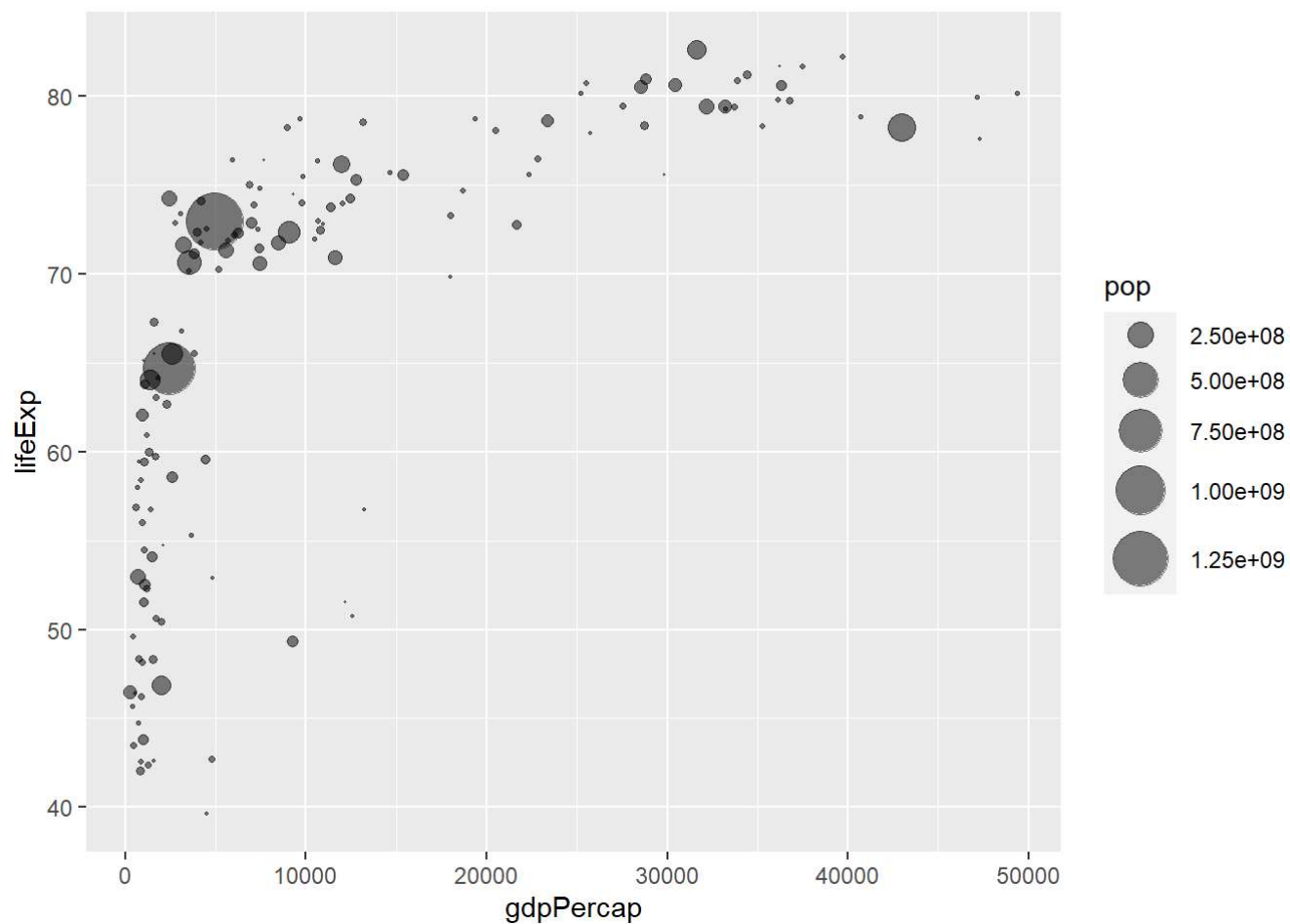
```
# create a plot for life expectancy vs GDP:
#color the points by the numeric variable population pop
ggplot(gapminder_2007) +
  aes(x = gdpPercap, y = lifeExp, color = pop) +
  geom_point(alpha=0.8)
```



```
#set the point size based on the country
ggplot(gapminder_2007) +
  aes(x = gdpPercap, y = lifeExp, size = pop) +
  geom_point(alpha=0.5)
```

```
#reflect the actual population differences  
ggplot(gapminder_2007) +  
  geom_point(aes(x = gdpPercap, y = lifeExp,  
                 size = pop), alpha=0.5) +  
  scale_size_area(max_size = 10)
```



```
# produce a contrast with year 1957
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)
```

