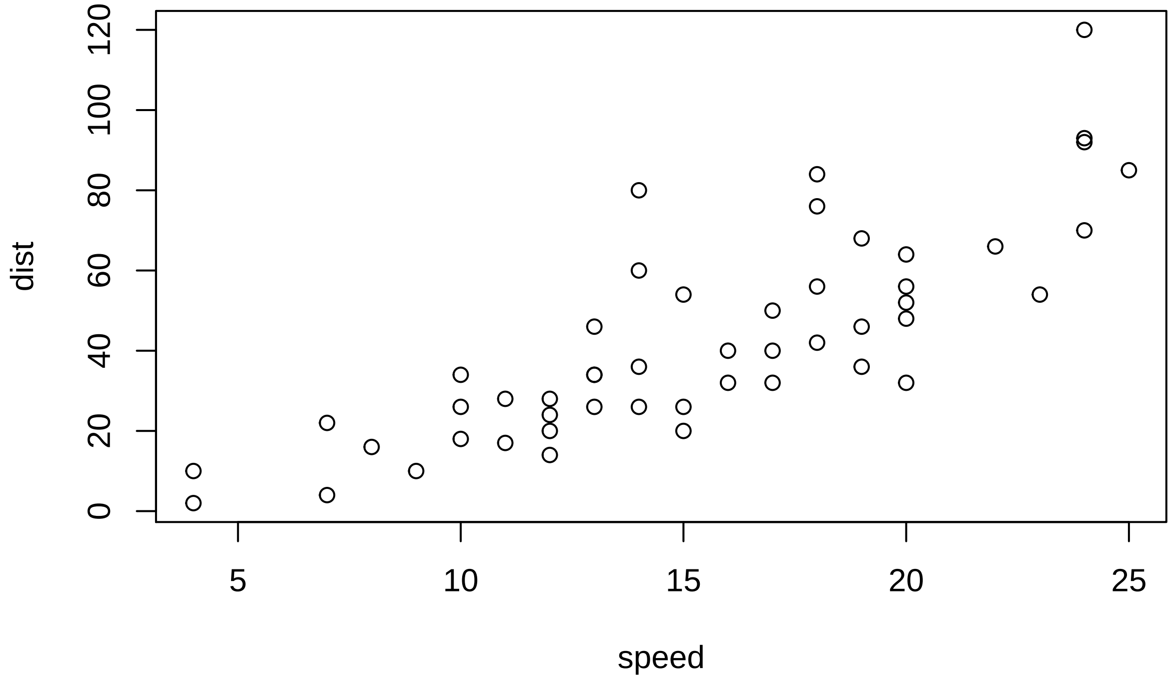
class05.R

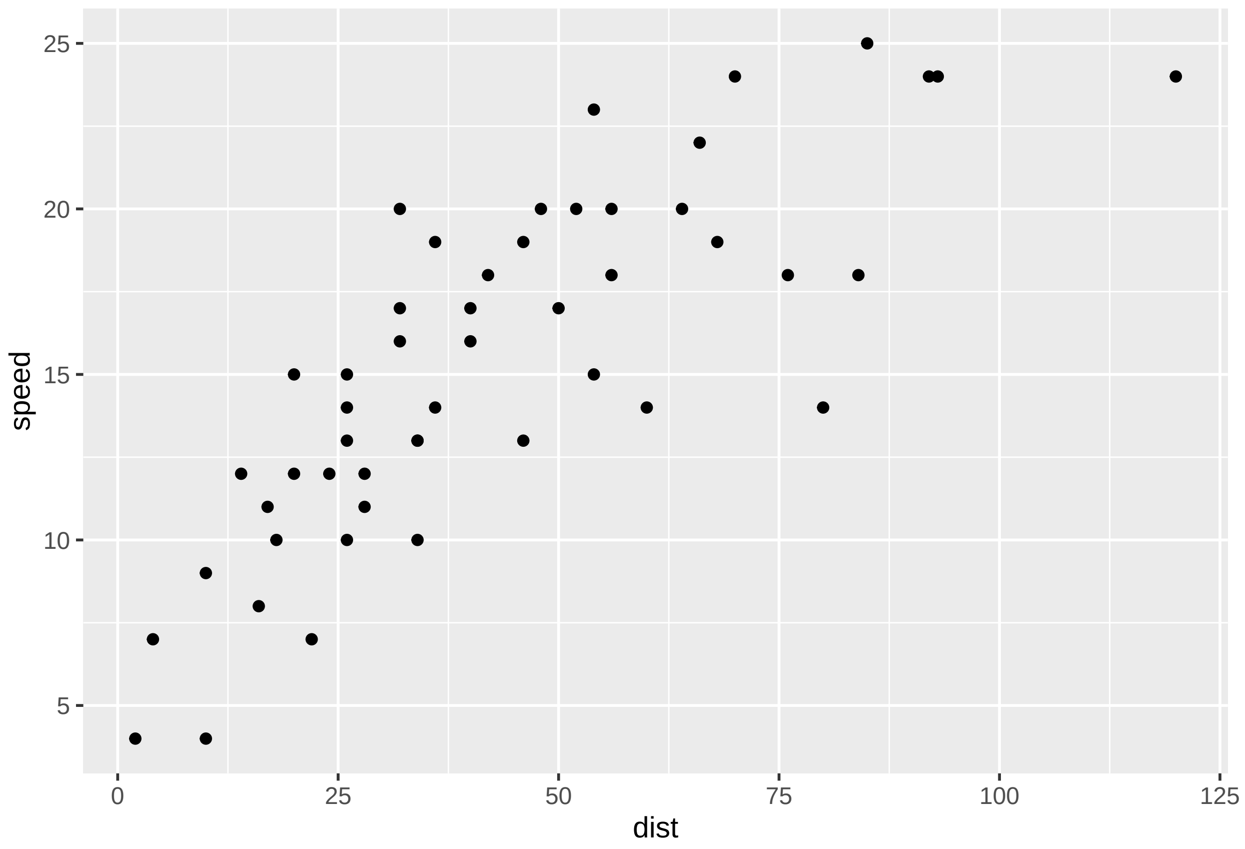
sbhwang

2022-02-05

```r
# Class 05 Data Visualization
# Run shortcut: command + enter
# Source: run everything on your script

# This is the "base" R plot
plot(cars)

# We are going to get a new plotting package called ggplot2
# ggplot(cars) --> cannot because we did not load it yet

# install.packages("ggplot2")
# Once we installed the package, we don't have to install again
# Run it on console rather than script because we don't want to constantly install it

# Now we need to call/load the package
# We have to call library every time we start a new session
library(ggplot2)
```

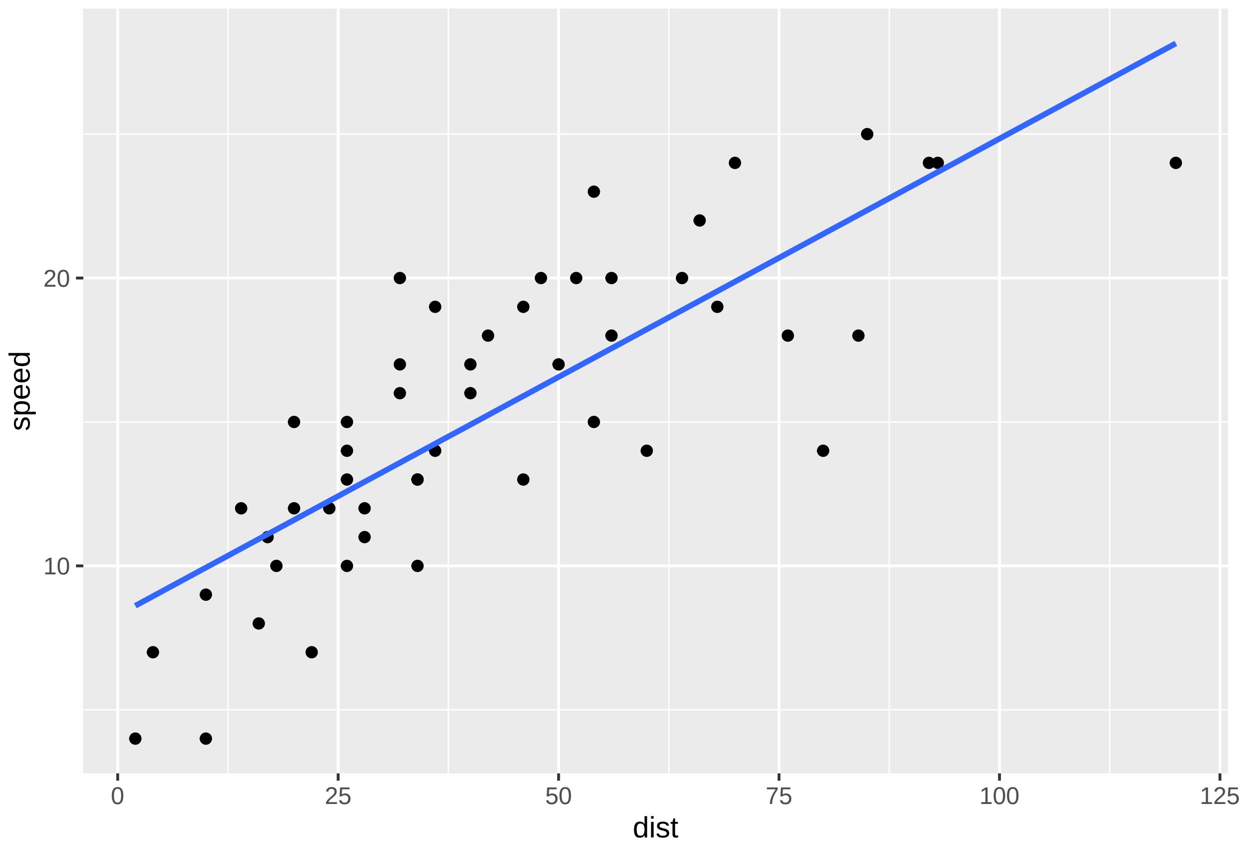


```
# This sets up the plot
#ggplot(cars) --> need aes() and geom() to plot

gg <- ggplot(data=cars) + aes(x=dist, y=speed) + geom_point()
gg
```



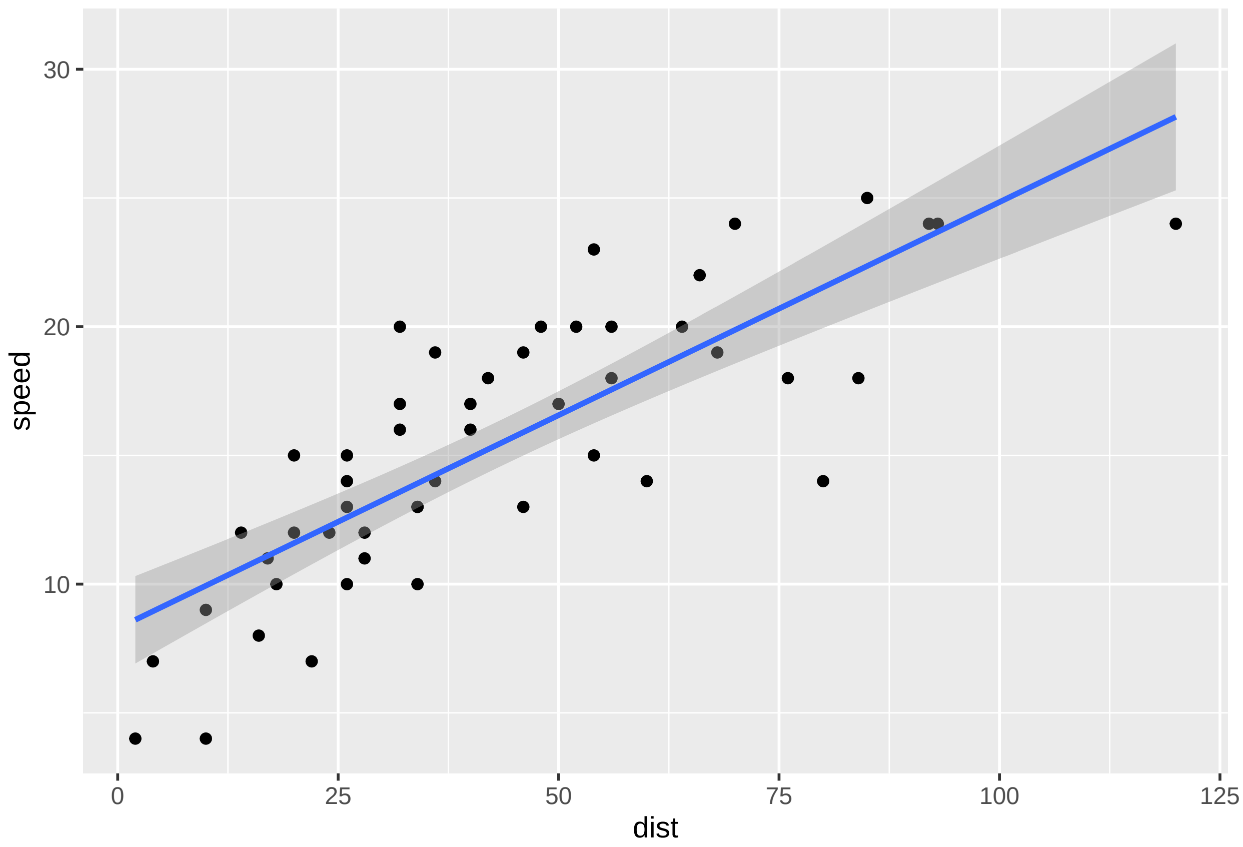
```
# One last thing. lets add a line to the data
gg + geom_smooth(method="lm", se=FALSE)
```

```
## ‘geom_smooth()‘ using formula ’y ~ x’
```



```
ggplot(data=cars) + aes(x=dist, y=speed) + geom_point() +
  geom_smooth(method="lm")
```

```
## ‘geom_smooth()‘ using formula ’y ~ x’
```



```
# genes data frame
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
```

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

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

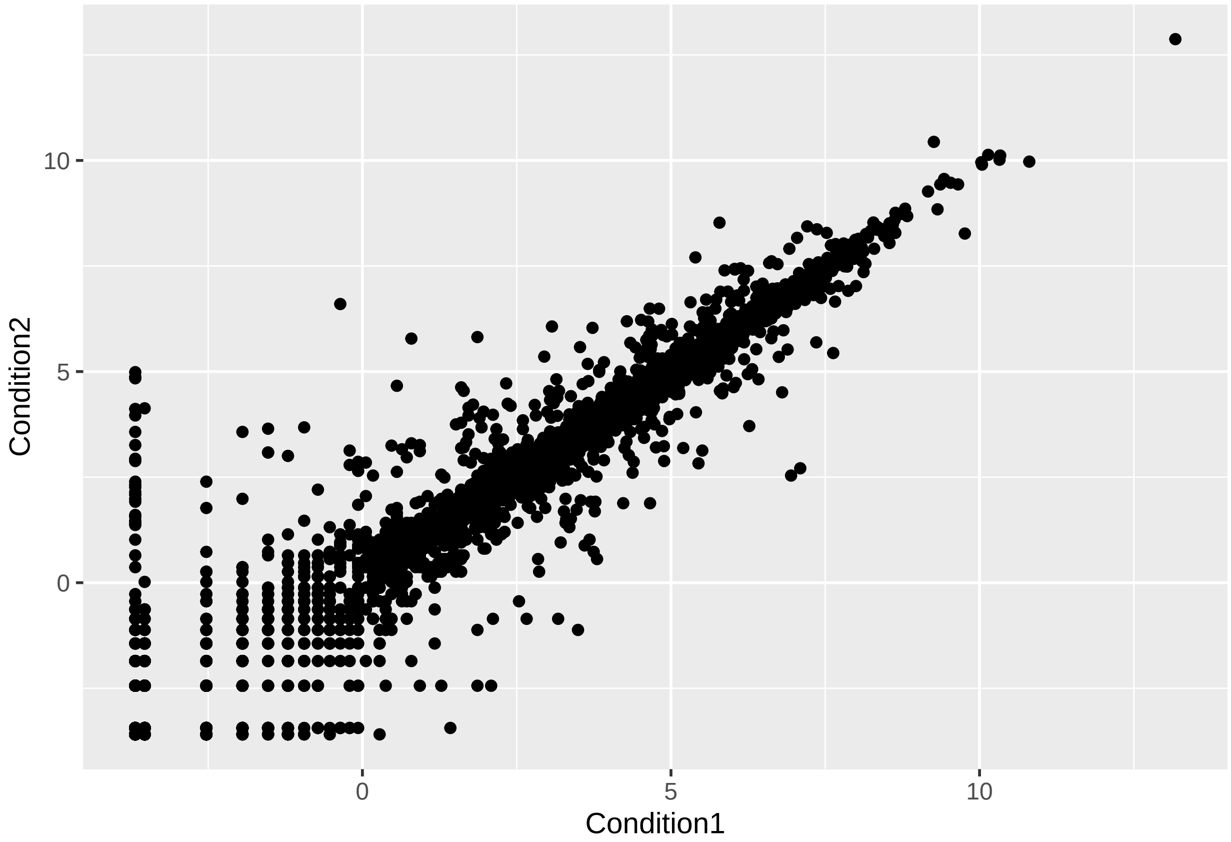
```
127/(72+4997+127)*100
```

```
## [1] 2.444188
```

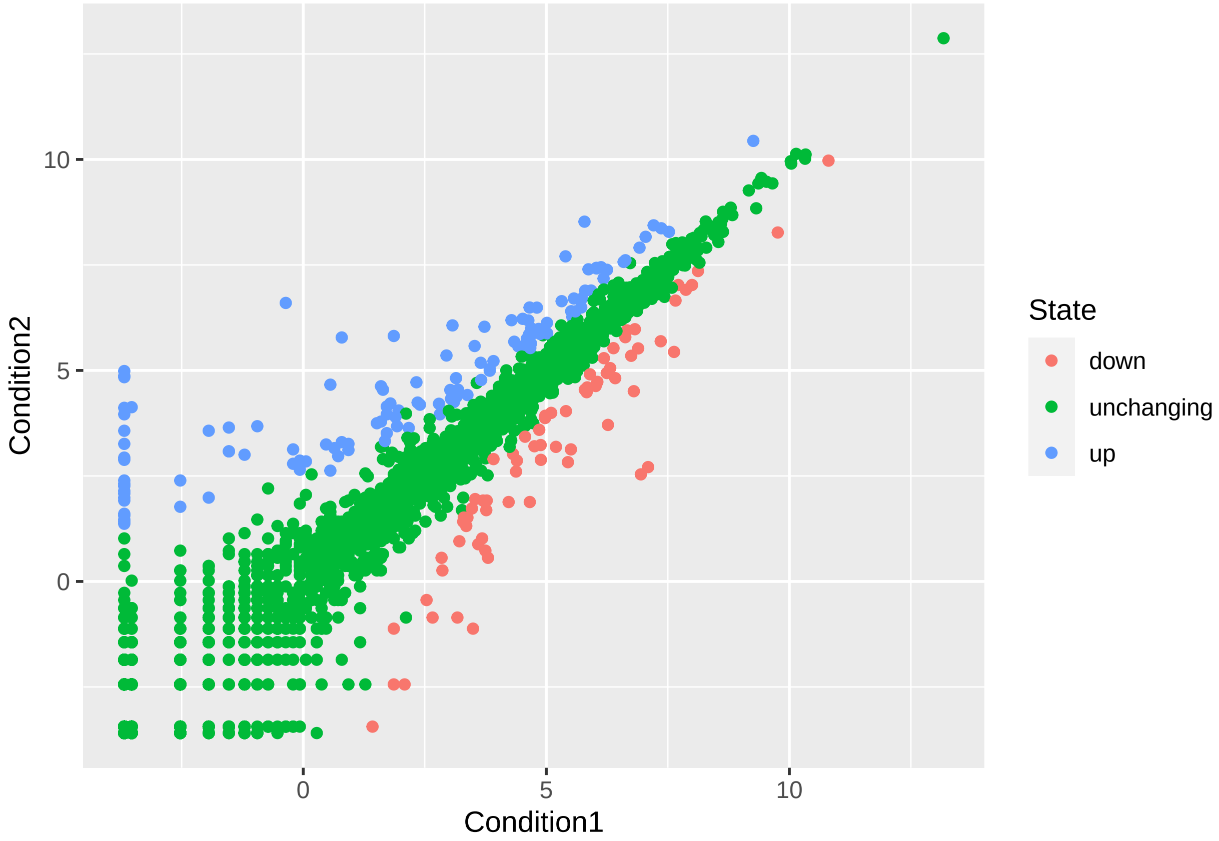
```
# or
round(table(genes$State)/nrow(genes) * 100, 2)
```

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

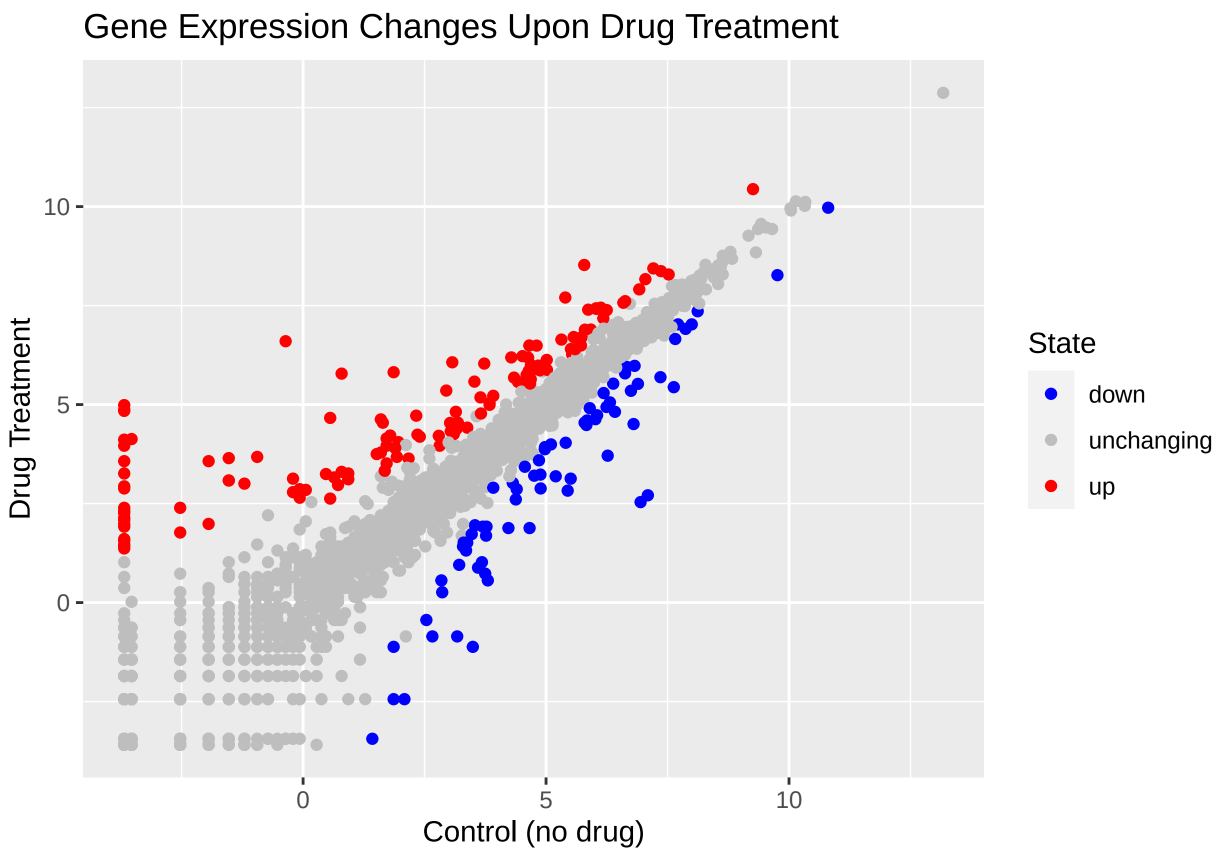
```
# Condition1(no drug) vs. Condition2(drug) plot
ggplot(data=genes) + aes(x=Condition1, y=Condition2) +
  geom_point()
```



```
# Plot that tells whether the difference in expression values between conditions is statistically signi
p <- ggplot(data=genes) + aes(x=Condition1, y=Condition2, color=State) +
  geom_point()
p
```



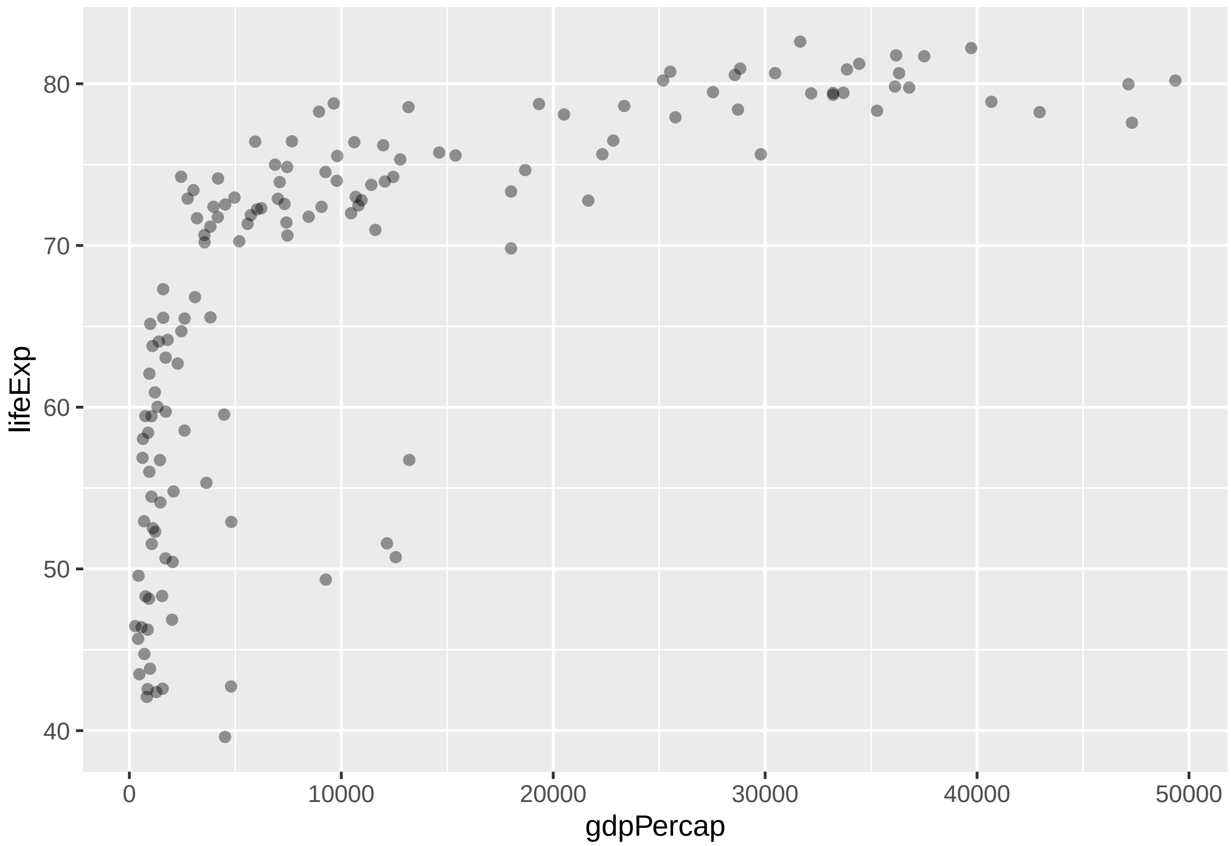
```
p + scale_colour_manual(values = c("blue", "gray", "red")) +
  labs(x="Control (no drug)", y="Drug Treatment",
       title="Gene Expression Changes Upon Drug Treatment")
```



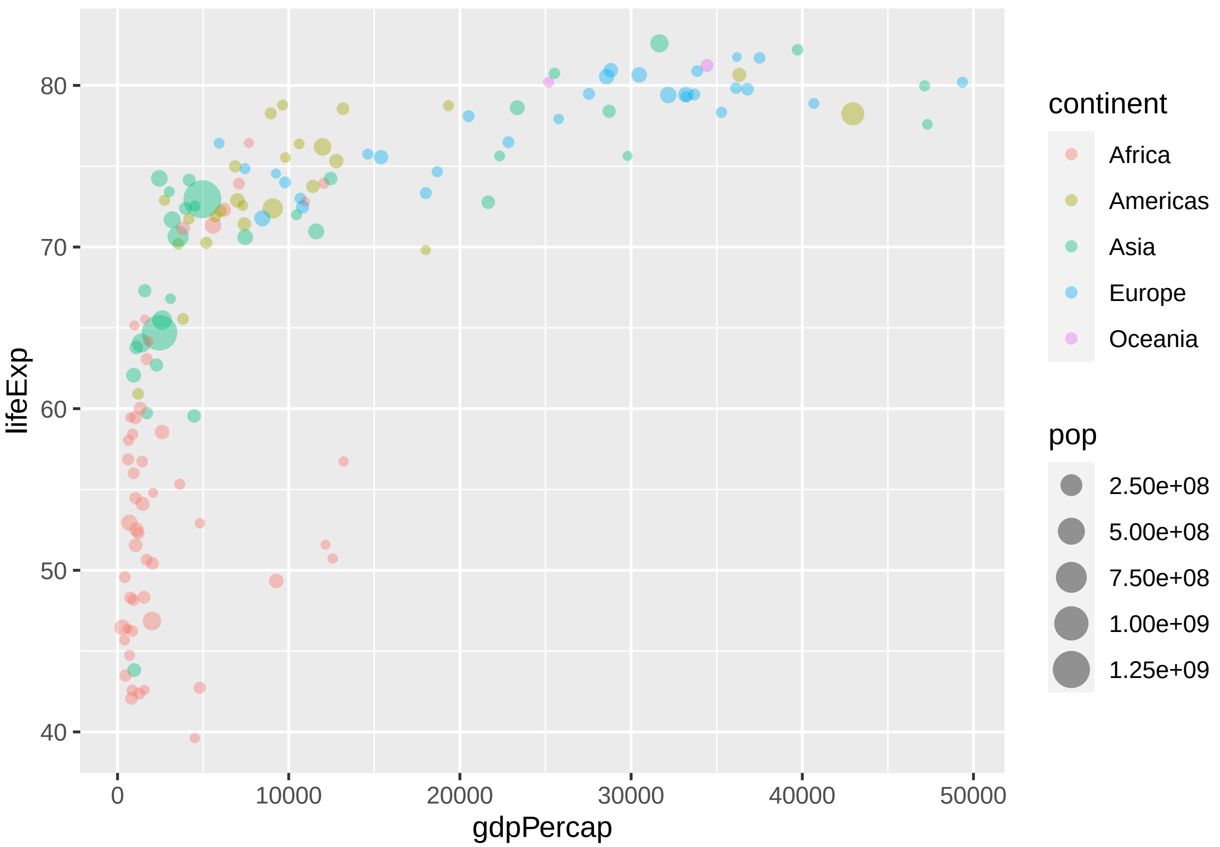
```
# gapminder
# install.packages("gapminder")
library(gapminder)

# File location online
url <- "https://raw.githubusercontent.com/jennybc/gapminder/master/inst/extdata/gapminder.tsv"

# read.delim(): Reads a file in table format and creates a data frame from it, with cases corresponding
gapminder <- read.delim(url)

#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
```

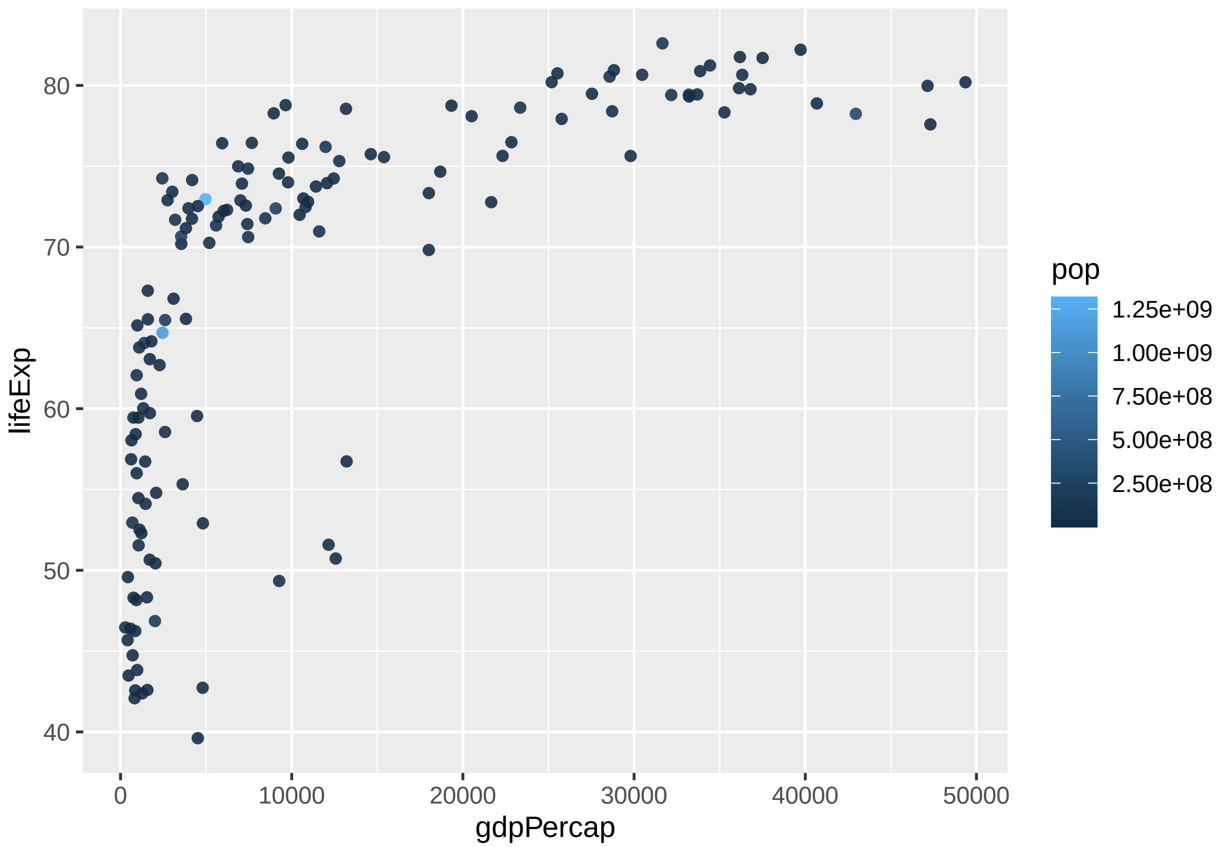
```r
gapminder_2007 <- gapminder %>% filter(year == 2007)

g <- ggplot(gapminder_2007) + aes(x=gdpPercap, y=lifeExp) +
  geom_point(alpha=0.4)
g
```



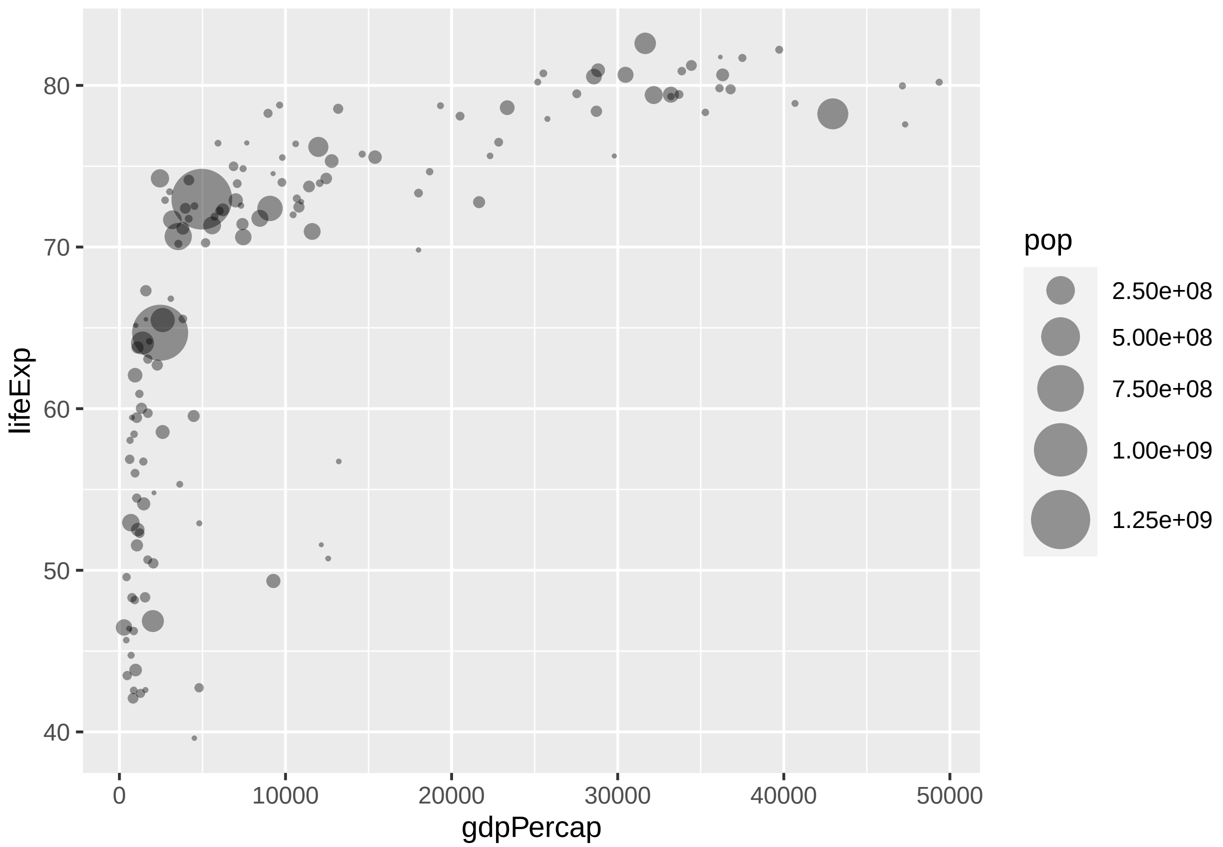
```r
g + aes(col=continent, size=pop)
```



```
g + aes(col=pop) + geom_point(alpha=0.8)
```



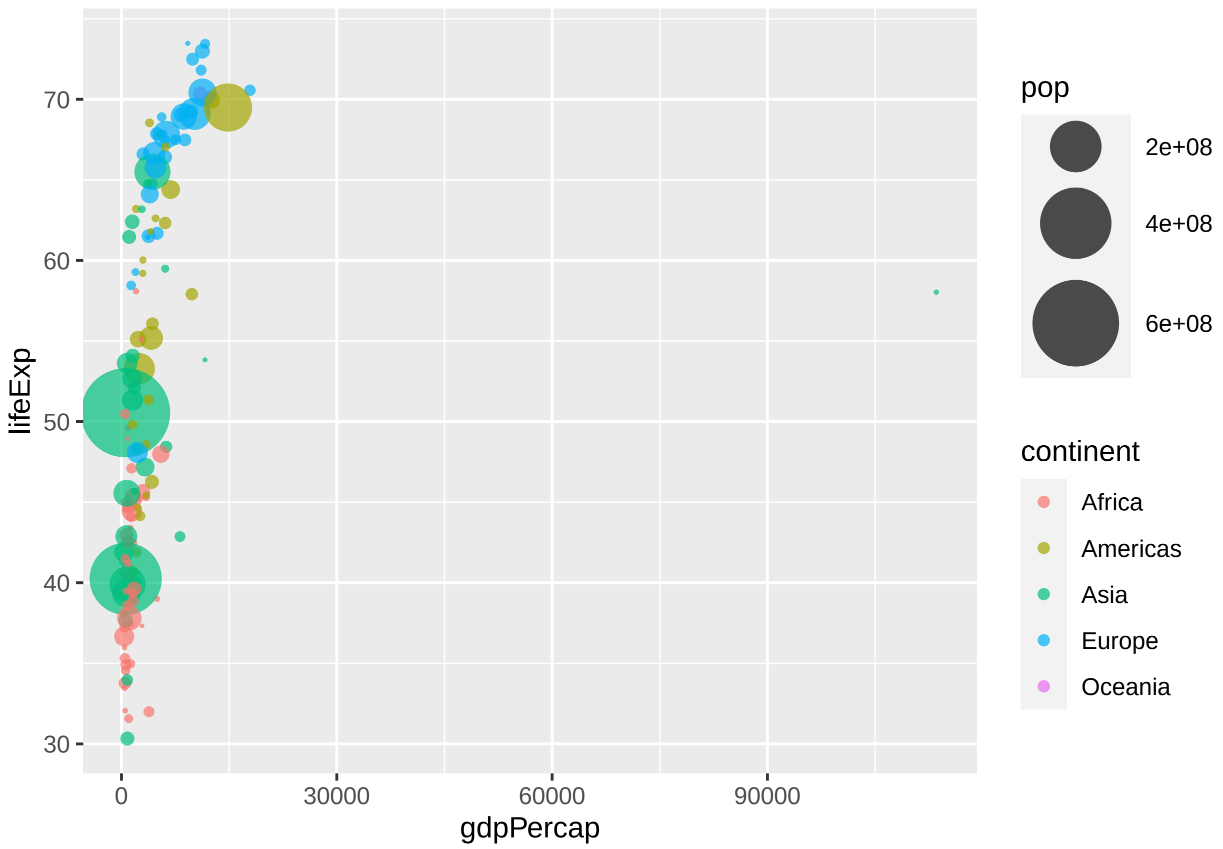
```
# scale_size_area(): scale the area of points to be proportional to the value.
# max_size: size of largest points
g + aes(size=pop) + scale_size_area(max_size=10)
```



```
# gapminder scatterplot for 1957
colnames(gapminder)
```

```
## [1] "country"   "continent" "year"      "lifeExp"   "pop"       "gdpPercap"
```

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

ggplot(gapminder_1957) +
  aes(x=gdpPercap, y=lifeExp, col=continent, size=pop) +
  geom_point(alpha=0.7) + scale_size_area(max_size = 15)
```


70
60
50
40
30
lifeExp
0
30000
60000
90000
gdpPercap
pop
2e+08
4e+08
6e+08
continent
Africa
Americas
Asia
Europe
Oceania

```
#2007
```