

# Week 5 Wednesday - Bioinformatics Class

Mirte Kuijpers

February 2nd, 2022

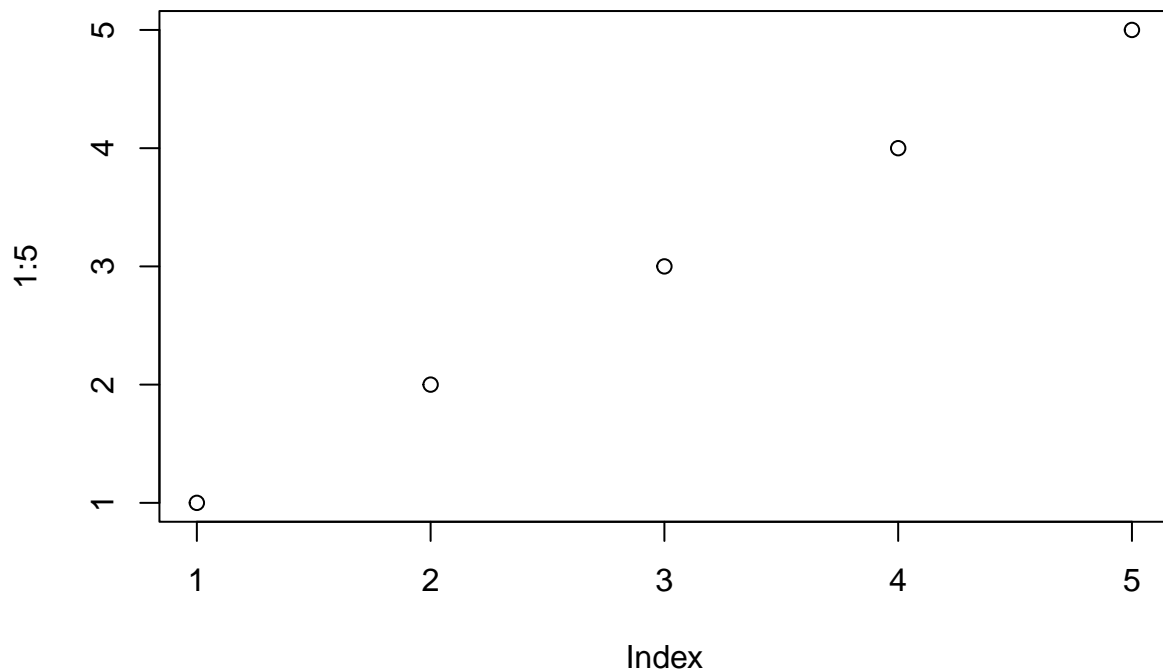
```
library(formatR)
library(knitr)
opts_chunk$set(tidy.opts=list(width.cutoff=80),tidy=TRUE)

#In person class, Wednesday 2nd Feb - Data Visualization

plot(1:5)

#ggplot

##set-up
library("ggplot2")
```



```
## Use cars data
#####
```

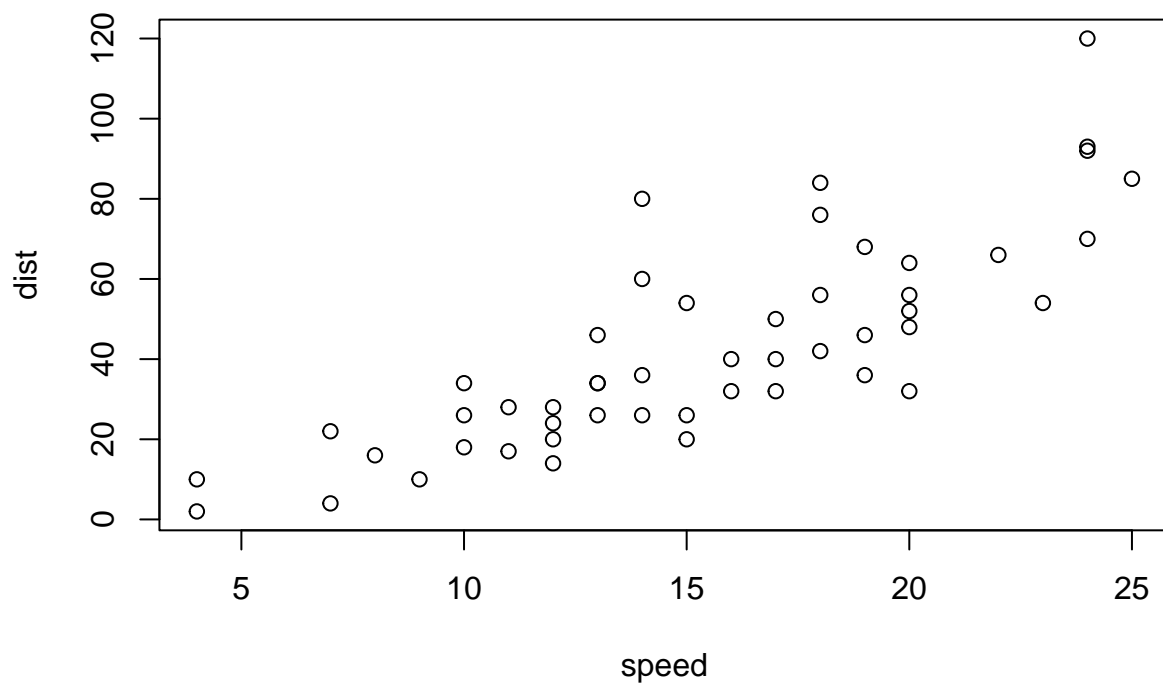
```
###investigate data
head(cars)
```

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

```
str(cars)
```

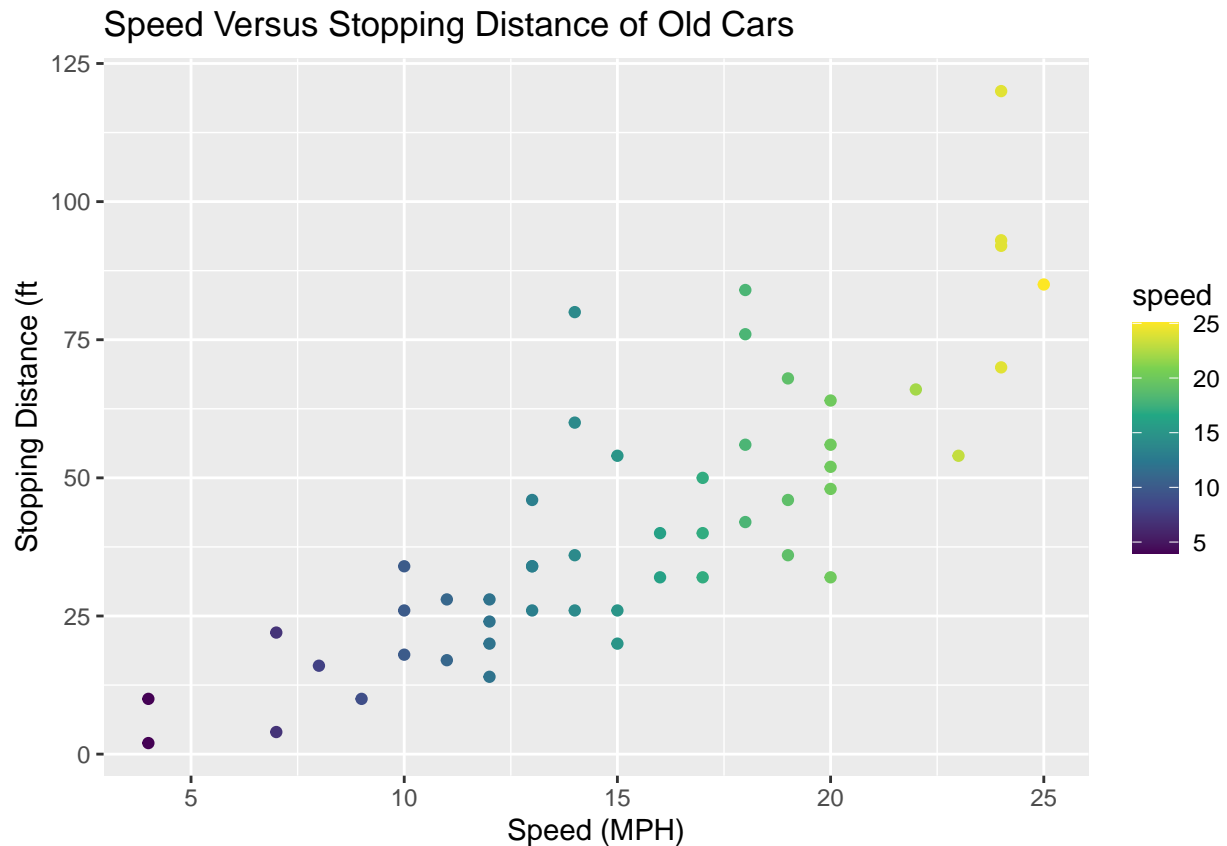
```
## 'data.frame':   50 obs. of  2 variables:
## $ speed: num  4 4 7 7 8 9 10 10 10 11 ...
## $ dist : num  2 10 4 22 16 10 18 26 34 17 ...
```

```
###plot data
plot(cars)
```



```
####first attempt
```

```
ggplot(cars, aes(speed, dist)) +  
  geom_point(aes(colour = speed)) +  
  labs(title="Speed Versus Stopping Distance of Old Cars", x="Speed (MPH)", y="Stopping Distance (ft)") +  
  scale_color_continuous(type = "viridis")
```



```
#best to put what you can into the aes() of ggplot rather than into the geom_'s -> more consistency,
```

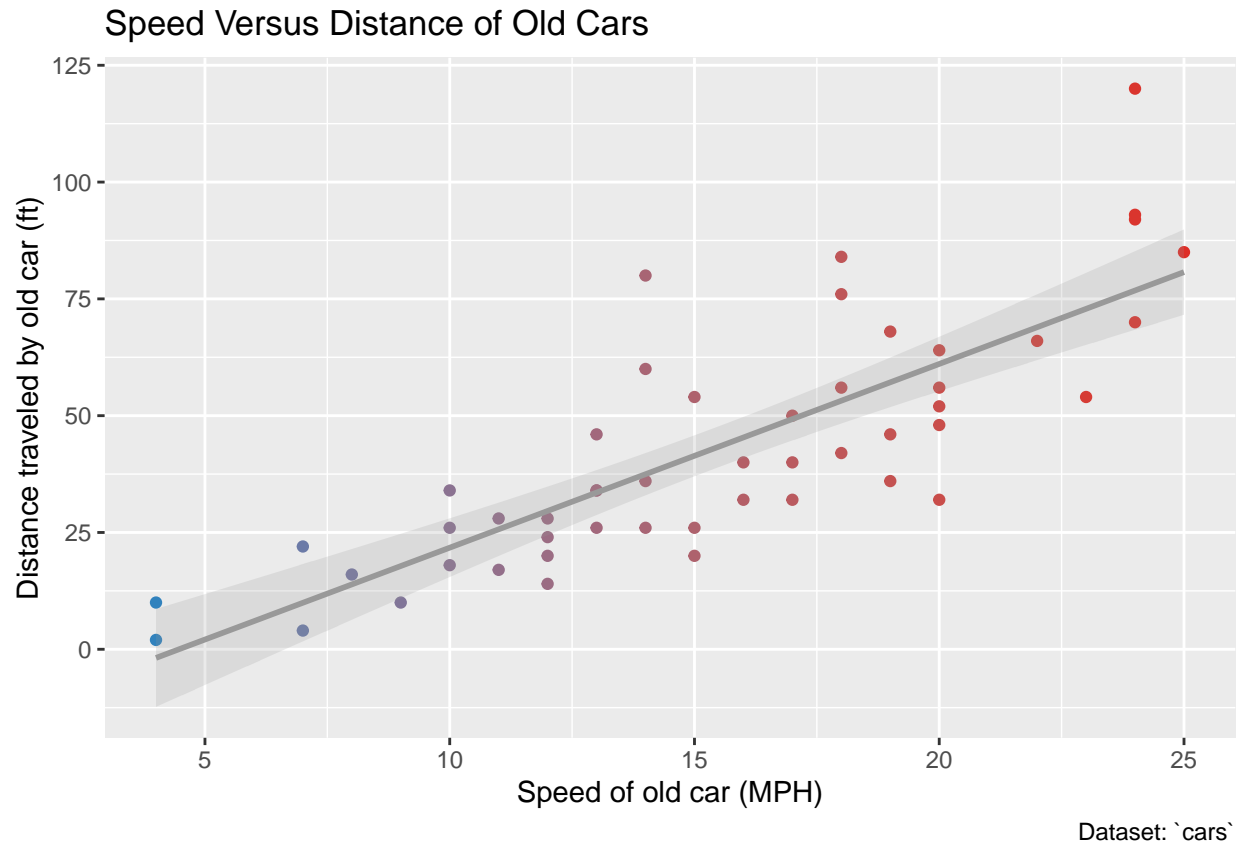
```
####improve colour scale
```

```
basic <- ggplot(cars, aes(speed, dist)) +  
  geom_point(aes(colour = speed), show.legend = FALSE) +  
  labs(title="Speed Versus Distance of Old Cars", x="Speed of old car (MPH)", y="Distance traveled by old car") +  
  scale_colour_gradient(name="Speed", low = "#3182bd", high = "#de2d26", space = "Lab", na.value = "grey")
```

```
####add a trend line
```

```
basic + geom_smooth(method="lm", colour = "grey60", alpha=0.2)
```

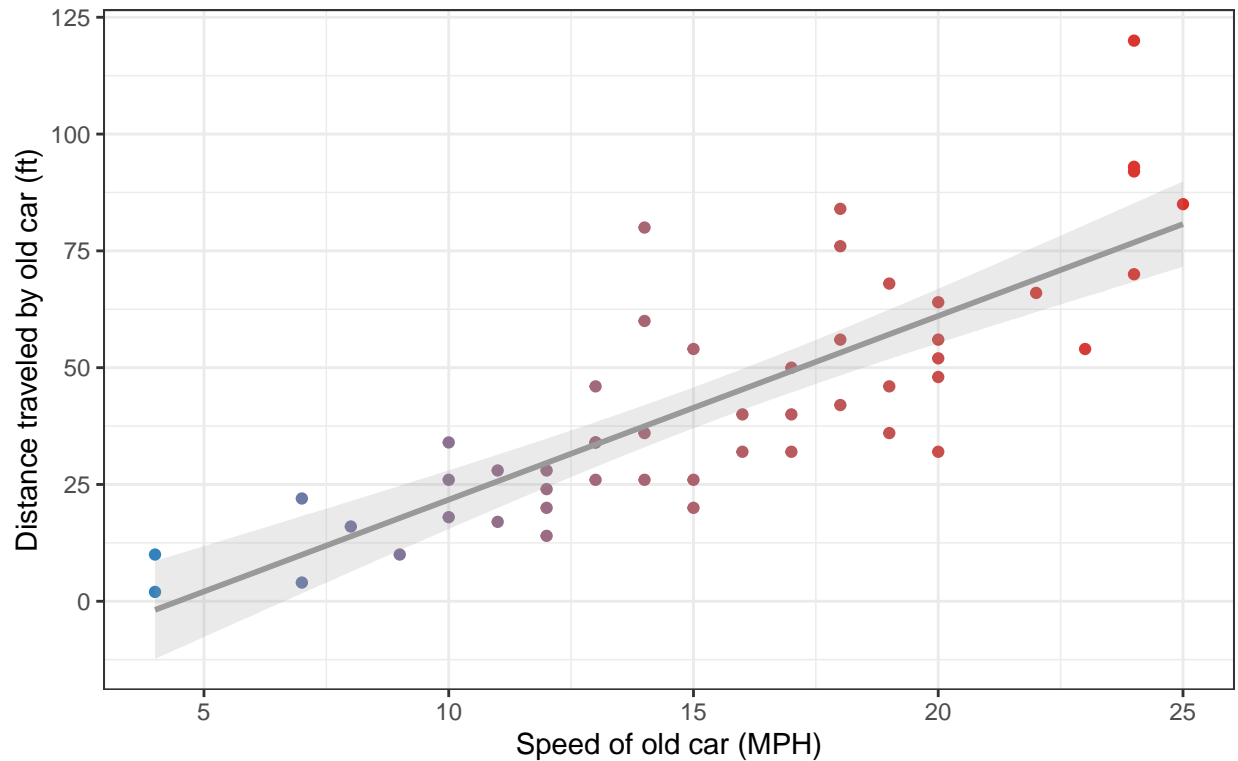
```
## 'geom_smooth()' using formula 'y ~ x'
```



```
####use the black and white theme as suggested in the tutorial
basic + geom_smooth(method="lm", colour = "grey60", alpha=0.2) + theme_bw()
```

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

## Speed Versus Distance of Old Cars



Dataset: `cars`

```
#####
#Gene data (next section)
#####

##get data
url <- "https://bioboot.github.io/bimm143_S20/class-material/up_down_expression.txt"
genes <- read.delim(url)

##observe data
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
```

```
str(genes)
```

```
## 'data.frame':   5196 obs. of  4 variables:
## $ Gene      : chr  "A4GNT" "AAAS" "AASDH" "AATF" ...
## $ Condition1: num  -3.681 4.548 3.719 5.078 0.471 ...
## $ Condition2: num  -3.44 4.39 3.48 5.02 0.56 ...
```

```
## $ State      : chr  "unchanging" "unchanging" "unchanging" "unchanging" ...
```

```
nrow(genes) ###number of genes = 5196
```

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

```
ncol(genes)
```

```
## [1] 4
```

```
colnames(genes)
```

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

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

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

```
### percentage of genes in each state
```

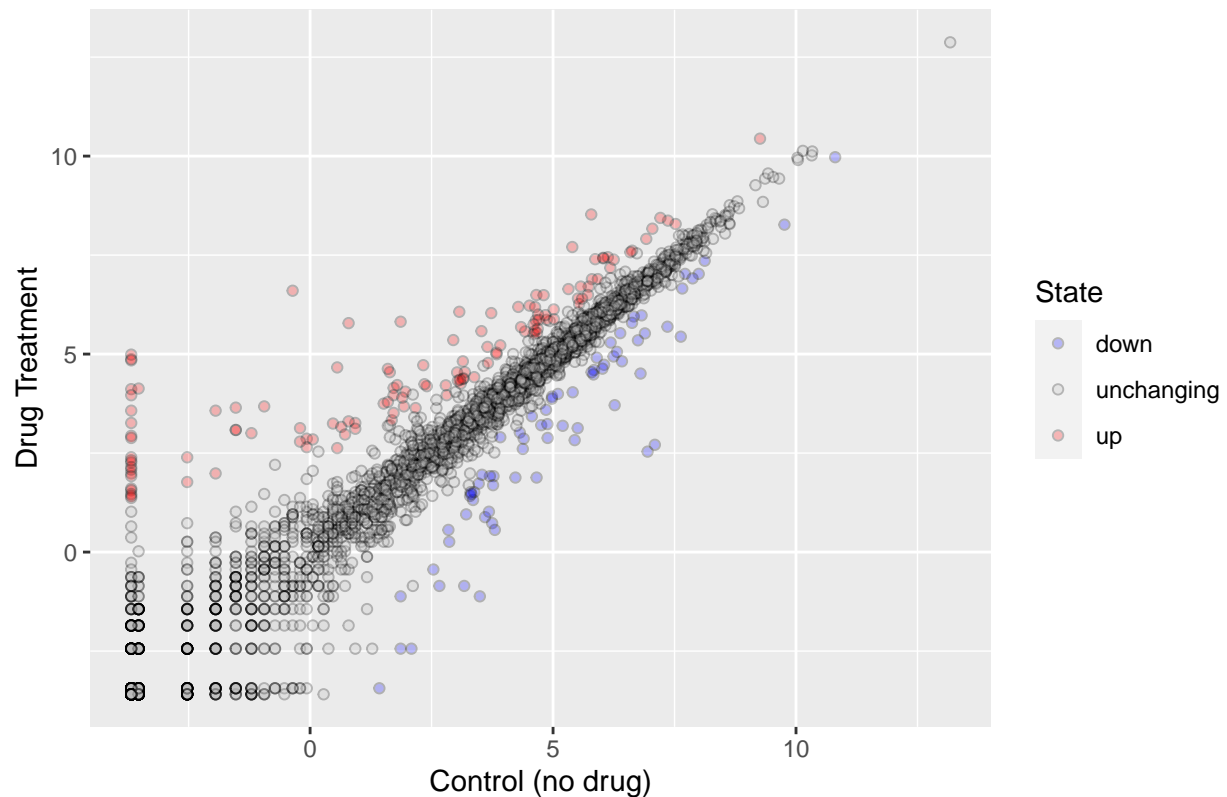
```
round( table(genes$State)/nrow(genes) * 100, 2 )
```

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

```
##plot data
```

```
ggplot(genes, aes(Condition1, Condition2, fill=State)) +  
  geom_point(pch=21, alpha=0.25) +  
  scale_fill_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



*#Section 6 (optional)*

*##further setup for this section, install commented out as only had to be done the first time*  
*#install.packages("gapminder")*

```
library(gapminder)
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
```

```
## Load data
```

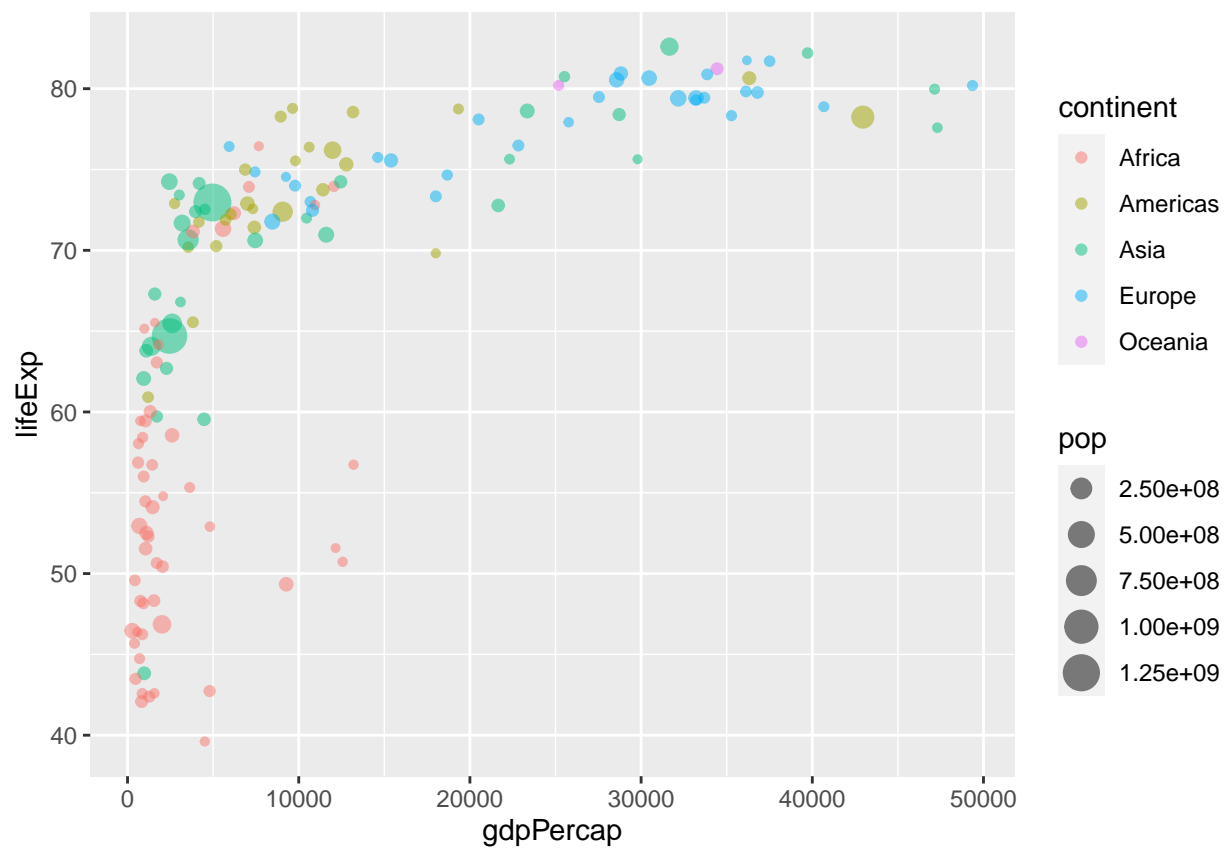
```
url <- "https://raw.githubusercontent.com/jennybc/gapminder/master/inst/extdata/gapminder.tsv"
```

```
gapminder <- read.delim(url) #read.delim is a function which reads a file in table format and creates a
```

```
###get only the data for 2007
```

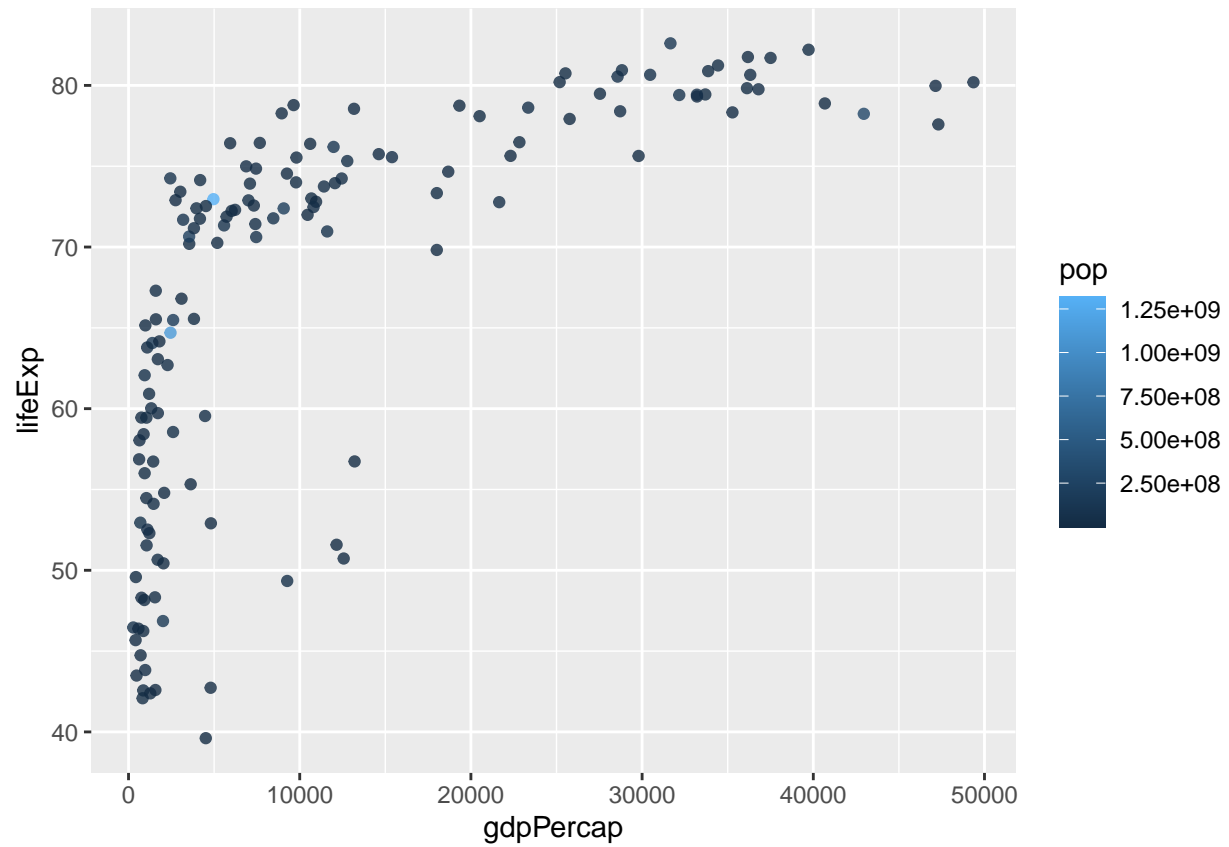
```
gapminder_2007 <- gapminder %>% filter(year==2007) #%>% passes the left hand side of the operator to th

##Begin to plot data
ggplot(gapminder_2007, aes(gdpPercap, lifeExp, color=continent, size=pop)) +
  geom_point(alpha=0.5)
```

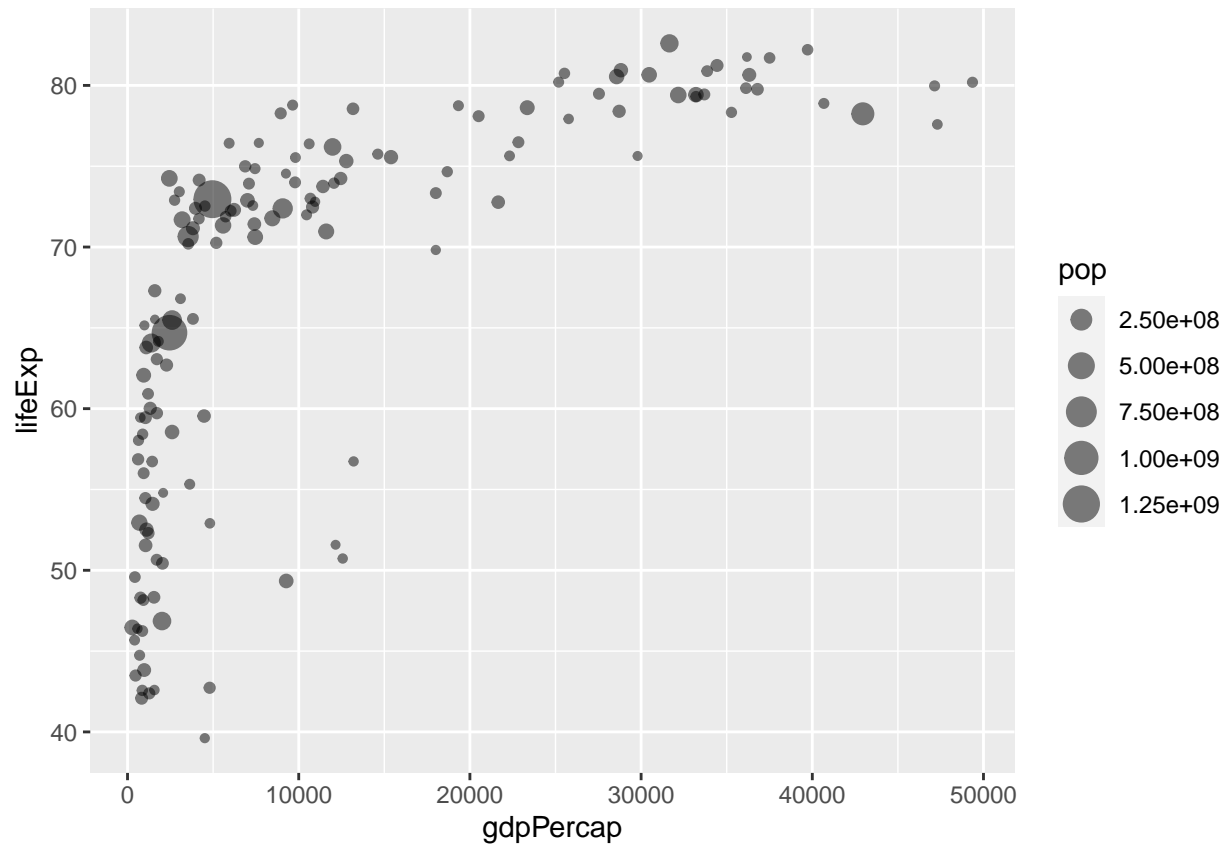


```
###alternative - colour by pop size
ggplot(gapminder_2007) +
  aes(x = gdpPercap, y = lifeExp, color = pop) +
  geom_point(alpha=0.8)
```

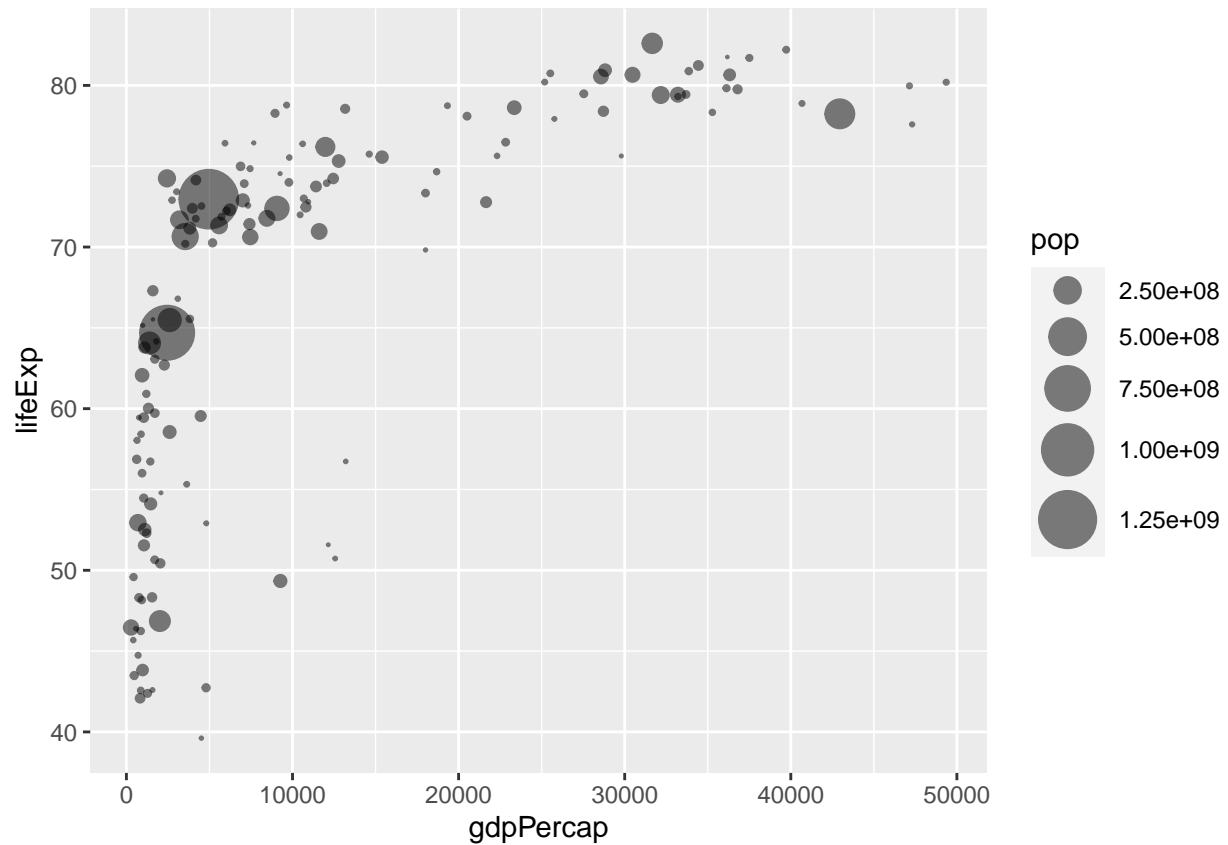




```
###or do size by pop size
ggplot(gapminder_2007) +
  aes(x = gdpPercap, y = lifeExp, size = pop) +
  geom_point(alpha=0.5)
```

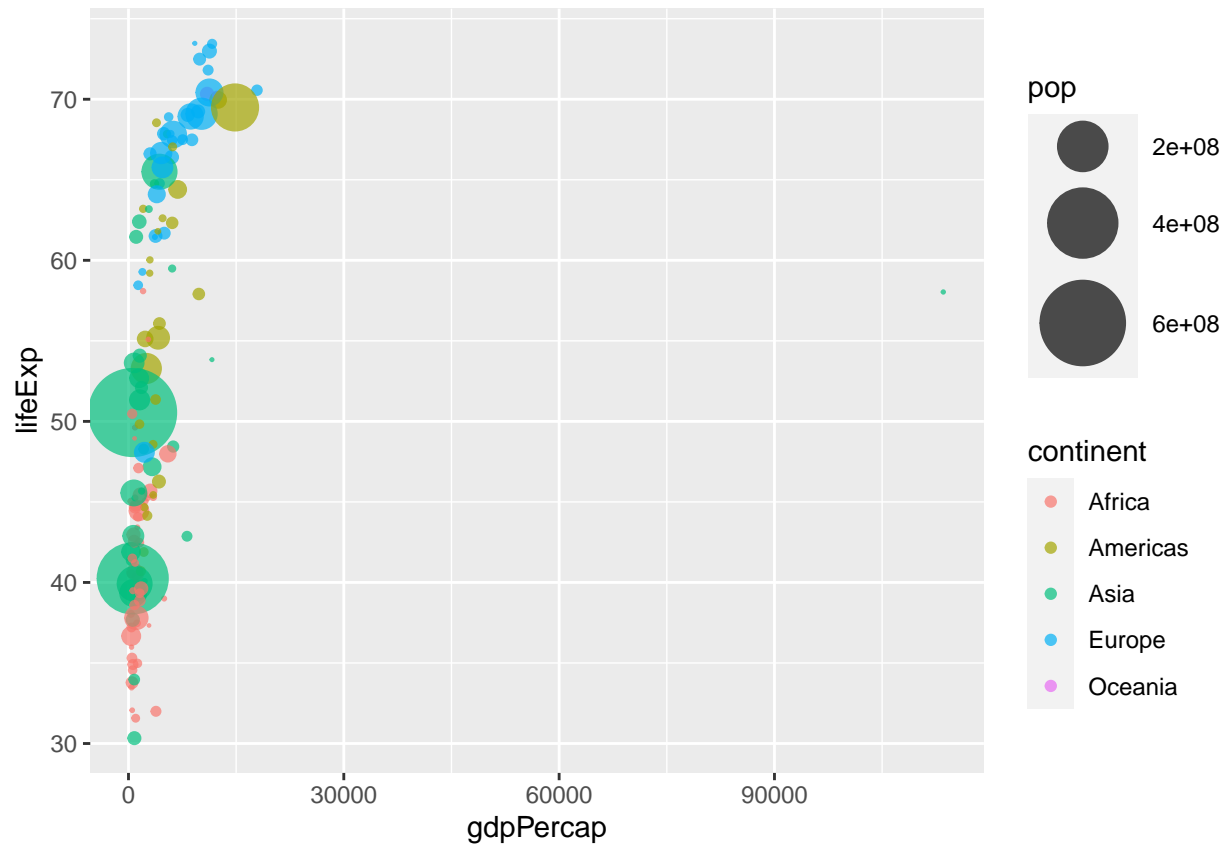


```
####but that doesn't work properly, area doesn't scale as it should so use below instead
ggplot(gapminder_2007) +
  geom_point(aes(x = gdpPercap, y = lifeExp,
                 size = pop), alpha=0.5) +
  scale_size_area(max_size = 10)
```



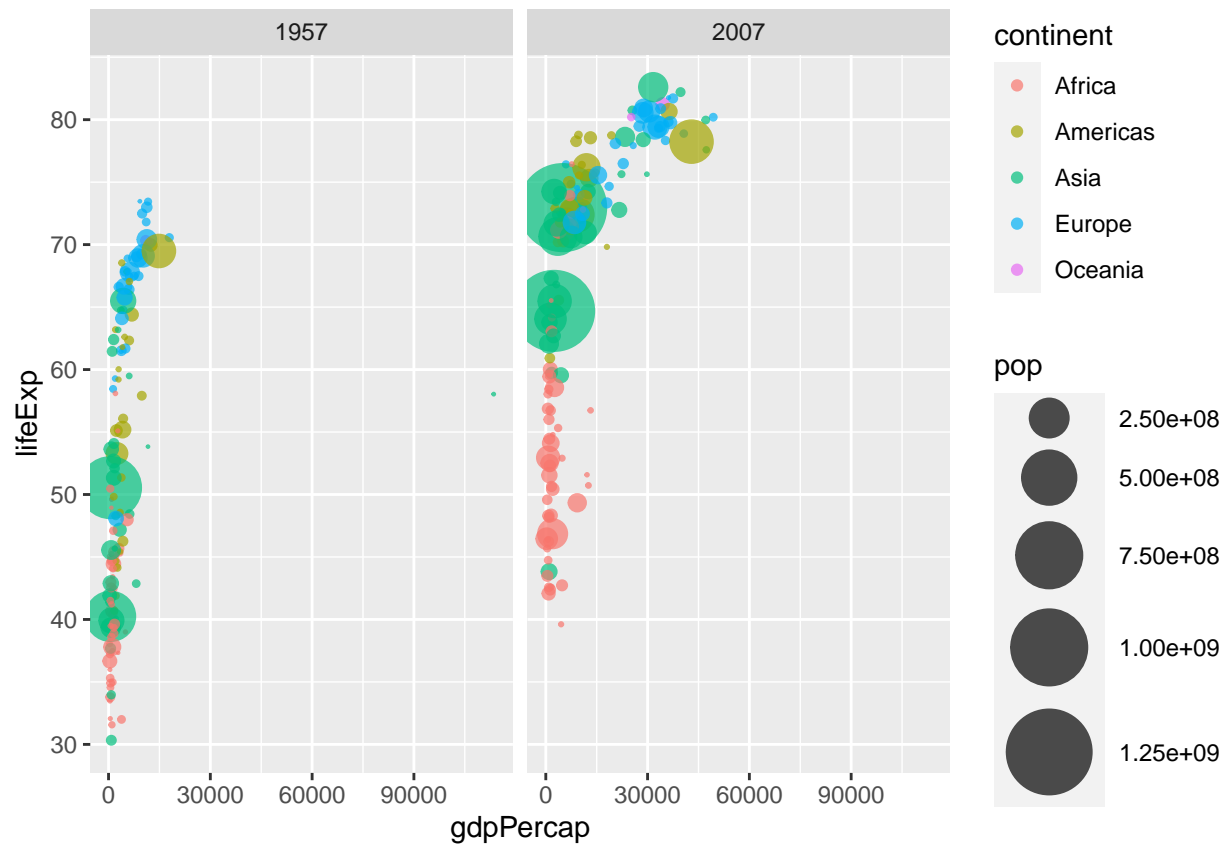
```
##Actual question
###get 1957 as well
gapminder_1957 <- gapminder %>% filter(year==1957)

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



```
##1957 and 2007
gapminder_merge <- rbind(gapminder_1957, gapminder_2007)

ggplot(gapminder_merge) +
  geom_point(aes(x = gdpPercap, y = lifeExp,
                 size = pop, colour=continent), alpha=0.7) +
  scale_size_area(max_size = 15) +
  facet_wrap(~year)
```



```
ggplot(gapminder_merge) +
  geom_point(aes(x = gdpPerCap, y = lifeExp,
                 size = pop, colour=continent), alpha=0.7) +
  scale_size_area(max_size = 15) +
  facet_wrap(~continent)
```

