

# example-R.R

Sam Do

2022-02-01

```
#Cars dataset  
cars
```

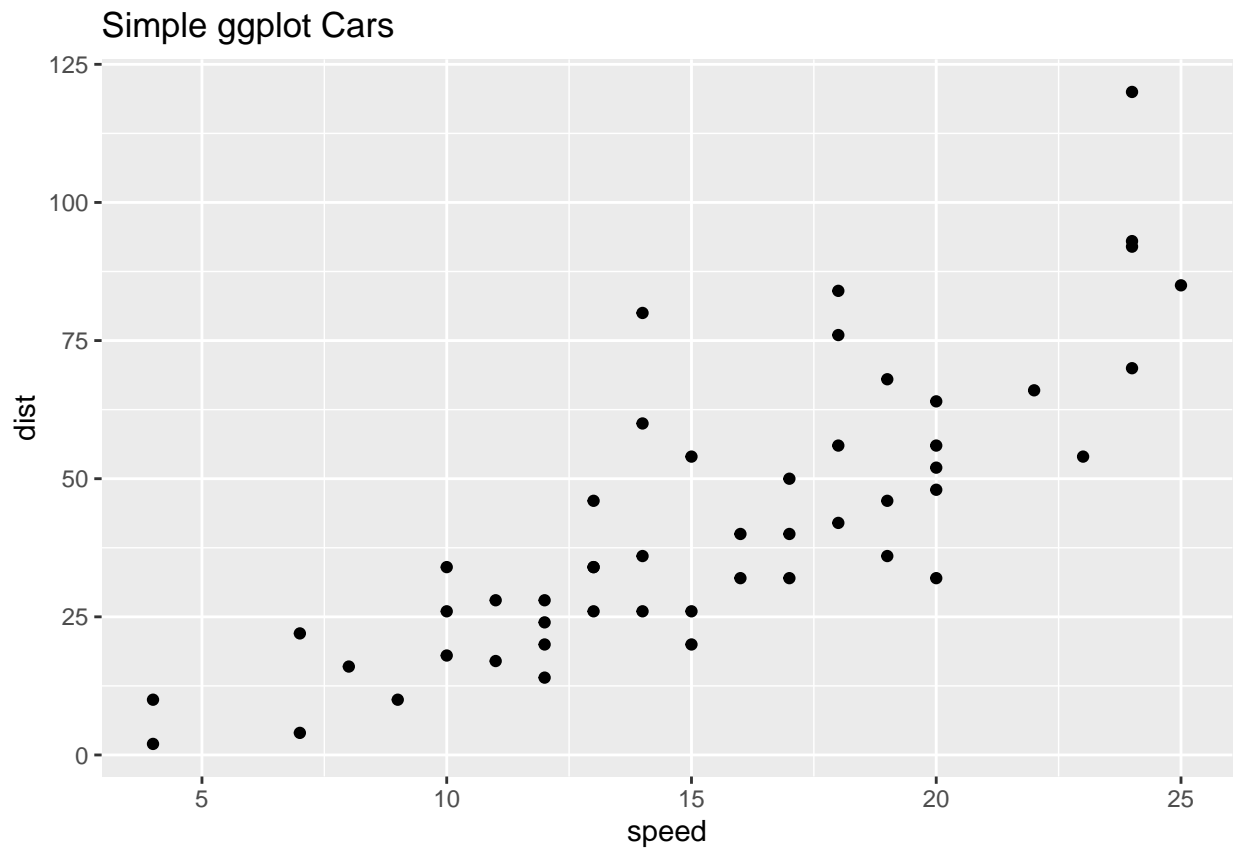
```
##      speed dist  
## 1         4    2  
## 2         4   10  
## 3         7    4  
## 4         7   22  
## 5         8   16  
## 6         9   10  
## 7        10   18  
## 8        10   26  
## 9        10   34  
## 10       11   17  
## 11       11   28  
## 12       12   14  
## 13       12   20  
## 14       12   24  
## 15       12   28  
## 16       13   26  
## 17       13   34  
## 18       13   34  
## 19       13   46  
## 20       14   26  
## 21       14   36  
## 22       14   60  
## 23       14   80  
## 24       15   20  
## 25       15   26  
## 26       15   54  
## 27       16   32  
## 28       16   40  
## 29       17   32  
## 30       17   40  
## 31       17   50  
## 32       18   42  
## 33       18   56  
## 34       18   76  
## 35       18   84  
## 36       19   36  
## 37       19   46  
## 38       19   68
```

```
## 39    20    32
## 40    20    48
## 41    20    52
## 42    20    56
## 43    20    64
## 44    22    66
## 45    23    54
## 46    24    70
## 47    24    92
## 48    24    93
## 49    24   120
## 50    25    85
```

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

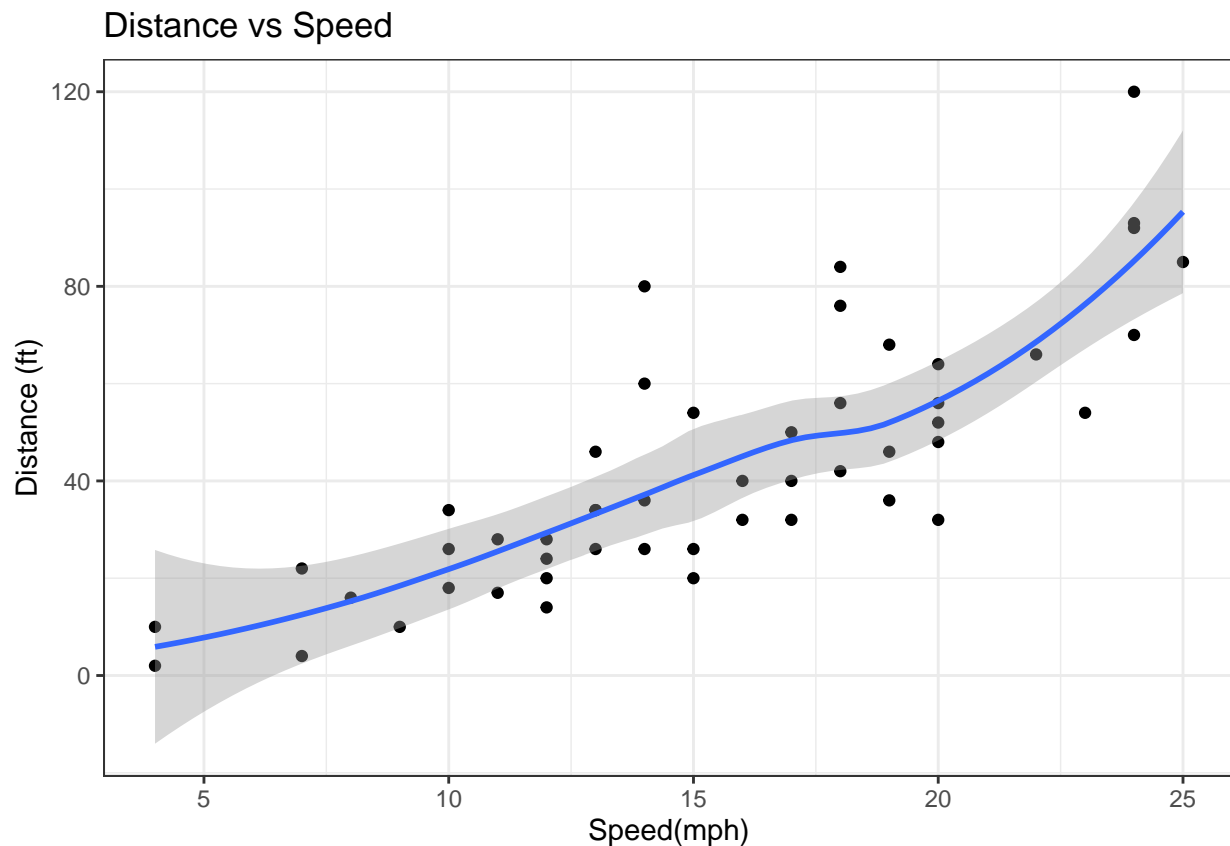
```
## Warning in register(): Can't find generic 'scale_type' in package ggplot2 to
## register S3 method.
```

```
ggplot(cars) + aes(speed, dist, main= "ggplot cars") +
  geom_point() +
  labs(title= "Simple ggplot Cars")
```



```
graph1 <- ggplot(cars) +
  aes(speed, dist, main="ggplot cars") +
  geom_point() +
  geom_smooth() +
  labs(title = "Distance vs Speed") +
  theme_bw() +
  xlab("Speed(mph)") +
  ylab("Distance (ft)")
graph1
```

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



```
#RNASeq experiment dataset
```

```
#Read data into R
```

```
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  unchanged
## 6 AB015752.4 -3.6808610 -3.5921390  unchanged
```

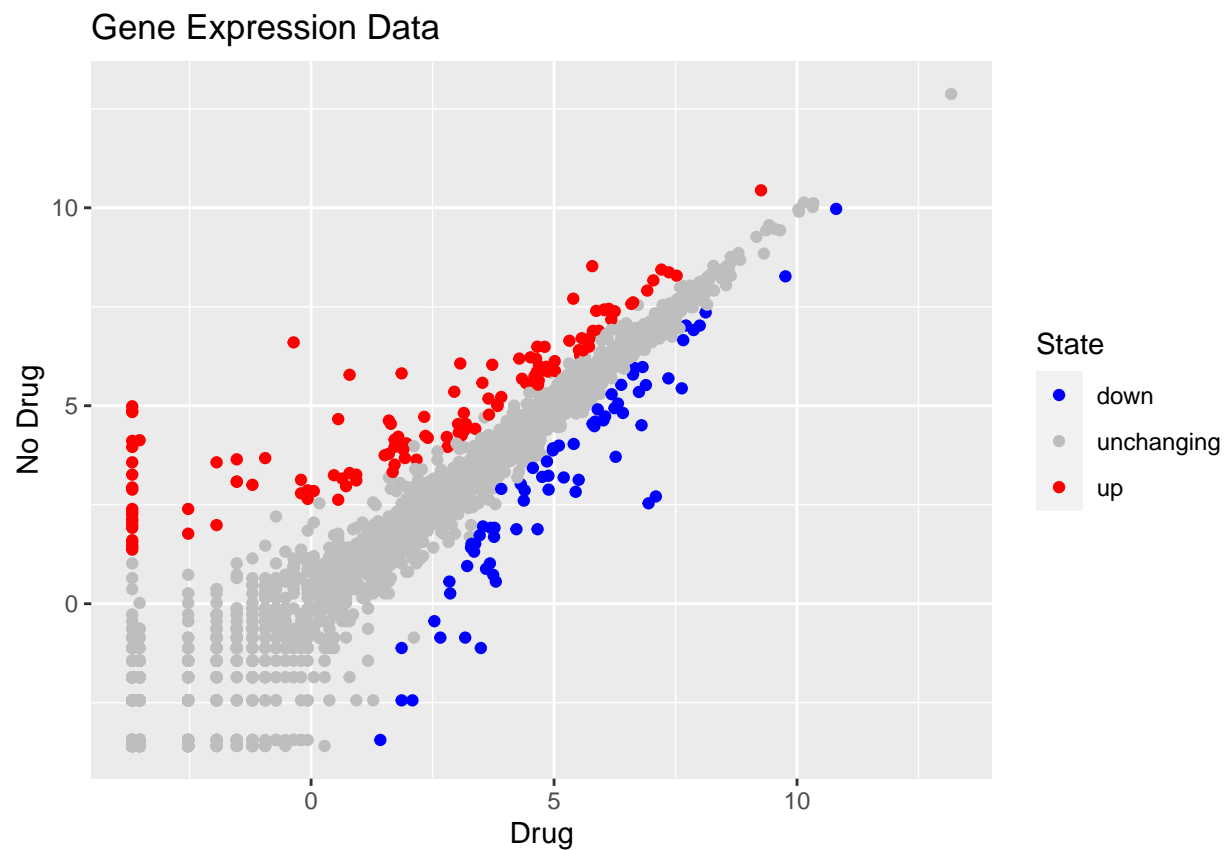
```
table(genes$State)
```

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

```
round(table(genes["State"])/nrow(genes)*100,2)
```

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

```
#Plot dataset
library(ggplot2)
ggplot(genes) + aes(Condition1, Condition2, color=State) +
  geom_point() +
  labs(title="Gene Expression Data") +
  xlab("Drug") +
  ylab("No Drug") +
  scale_colour_manual( values=c("blue","gray","red") )
```



```
#Population Dataset
#install.packages("gapminder")
#install.packages("dplyr")
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
```

```
gapminder_top5 <- gapminder %>% filter( year == 2007) %>%
  arrange(desc(pop)) %>% top_n(5, pop)
gapminder_top5
```

```
## # A tibble: 5 x 6
##   country      continent  year lifeExp      pop gdpPercap
##   <fct>        <fct>    <int>  <dbl>    <int>    <dbl>
## 1 China        Asia      2007   73.0 1318683096  4959.
## 2 India        Asia      2007   64.7 1110396331  2452.
## 3 United States Americas  2007   78.2 301139947   42952.
## 4 Indonesia    Asia      2007   70.6 223547000   3541.
## 5 Brazil        Americas  2007   72.4 190010647   9066.
```

```
library(ggplot2)
ggplot(gapminder_top5) +
  geom_col(aes(x = country, y = pop, fill=continent)) +
  labs(title="Top 5 Countries With Highest Population 2007") +
  xlab("Country") + ylab("Population")
```

Top 5 Countries With Highest Population 2007

