

# Class 5: Data Visualization with ggplot

Deborah Sigaud Kutner (PID: A69027018)

2024-01-24

## Graphic systems in R

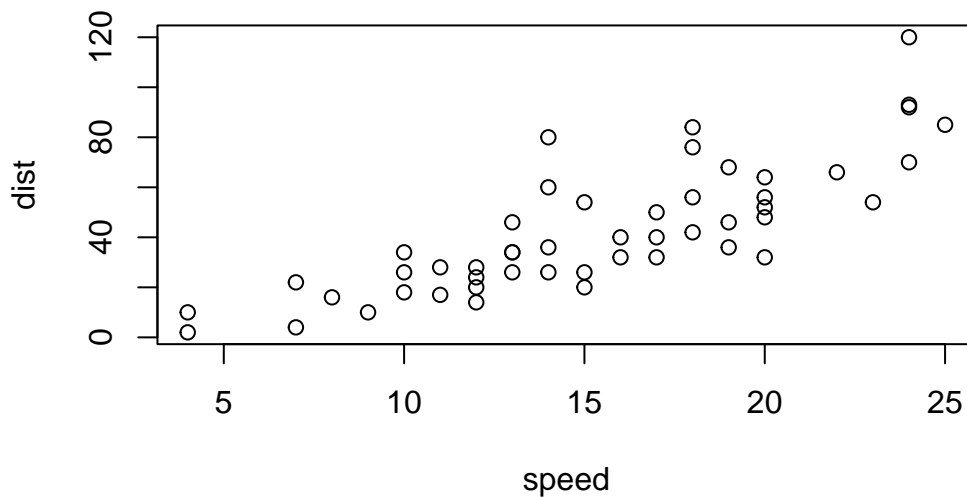
There are many graphic systems for R, including the so-called *base R* and those in add-on packages like 'ggplot2'.

### Dataset: cars

```
head(cars)
```

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

```
plot(cars) # Base r plot
```



Same plot with ggplot2. You need to specify:

- **data** (data.frame)
- **aesthetics** (how the data map to the plot)
- **geometry** (type of plot - e.g. line, dots, boxplot, etc.)

```
# install.packages("ggplot2") #install package
library(ggplot2) #load up the package
p1 <- ggplot(data=cars, aes(x= speed, y=dist)) +
  geom_point() +
  geom_smooth()+
  xlab("Speed") + ylab("Stopping Distance (ft)") + ggtitle("Stopping Distance of Old Cars")
  theme_bw()
```

Same plot with a linear regression and no confidence interval.

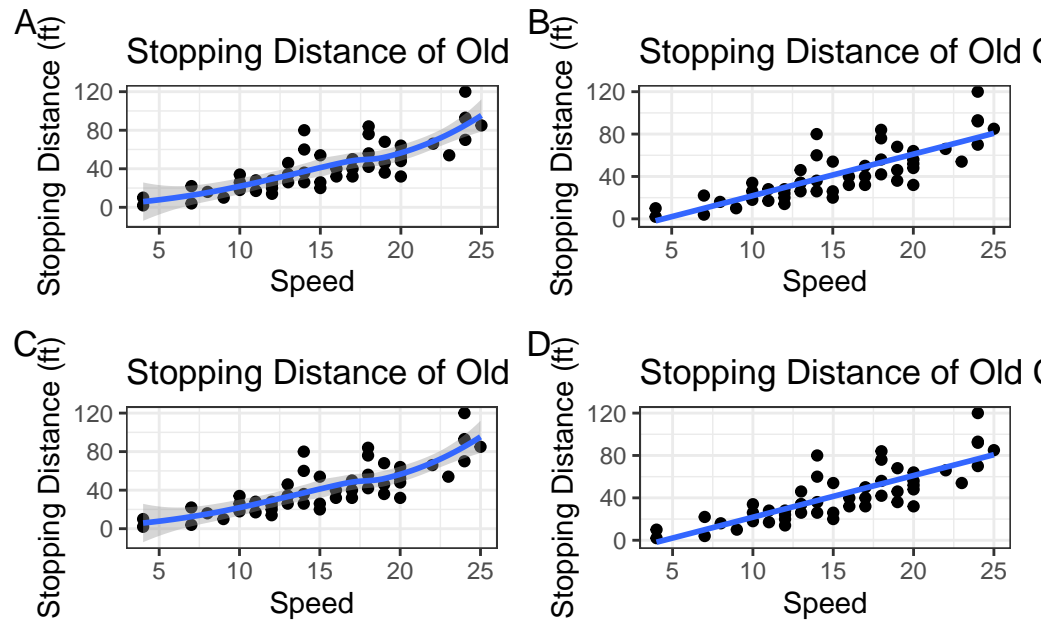
```
p2 <- ggplot(data=cars, aes(x= speed, y=dist)) +
  geom_point() +
  geom_smooth(method="lm", se=F)+
  xlab("Speed") + ylab("Stopping Distance (ft)") + ggtitle("Stopping Distance of Old Cars")
  theme_bw()
```

Combining multiple plots in one figure:

```
library("patchwork")
```

Warning: package 'patchwork' was built under R version 4.3.2

```
(p1 | p2) / (p1 | p2) + plot_annotation(tag_levels = 'A')
```



## Dataset: Genes

```
url <- "https://bioboot.github.io/bimm143_S20/class-material/up_down_expression.txt"
genes <- read.delim(url)
head(genes, 5)
```

	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

```
# How many genes are in this dataset?  
nrow(genes)
```

```
[1] 5196
```

```
# How many columns did you find?  
ncol(genes)
```

```
[1] 4
```

```
colnames(genes)
```

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

```
# How many 'up' regulated genes are there?  
table(genes$State) # There are 127 "up"
```

down	unchanging	up
72	4997	127

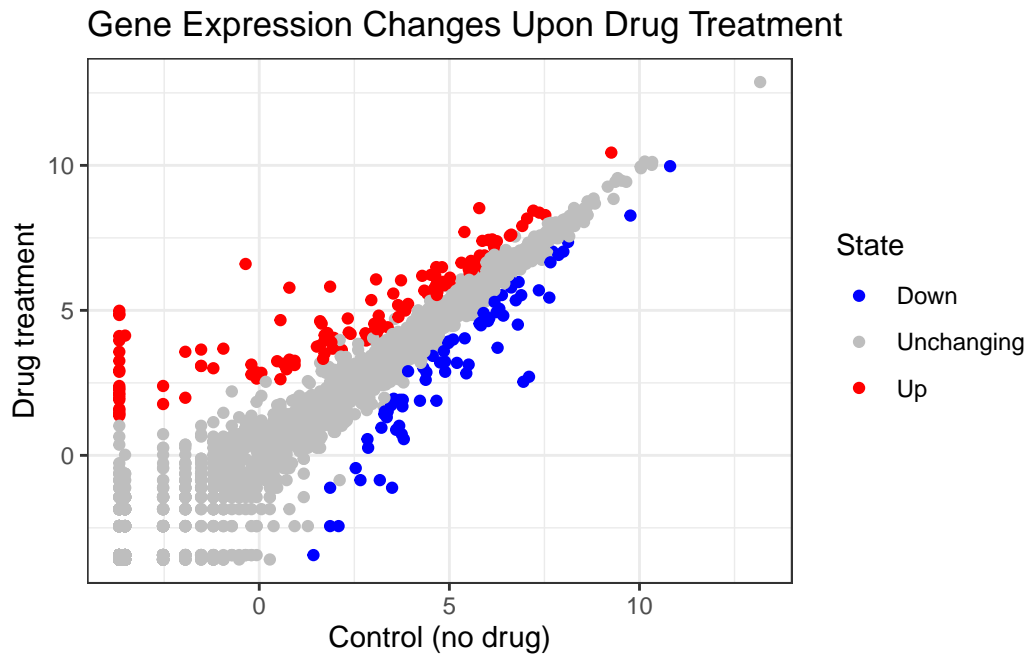
```
# What fraction of total genes is up-regulated in this dataset?  
table(genes$State == "up")/nrow(genes)*100 # 2.44%
```

FALSE	TRUE
97.555812	2.444188

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

down	unchanging	up
1.39	96.17	2.44

```
ggplot(data=genes, aes(x=Condition1, y=Condition2, color = State)) +
  geom_point() +
  scale_colour_manual(labels=c("Down", "Unchanging", "Up"), values=c("blue","gray","red"))
xlab("Control (no drug)") + ylab("Drug treatment") + ggtitle("Gene Expression Changes Upon Drug Treatment")
theme_bw()
```



```
library("plotly") # Won't work in a PDF
```

Warning: package 'plotly' was built under R version 4.3.2

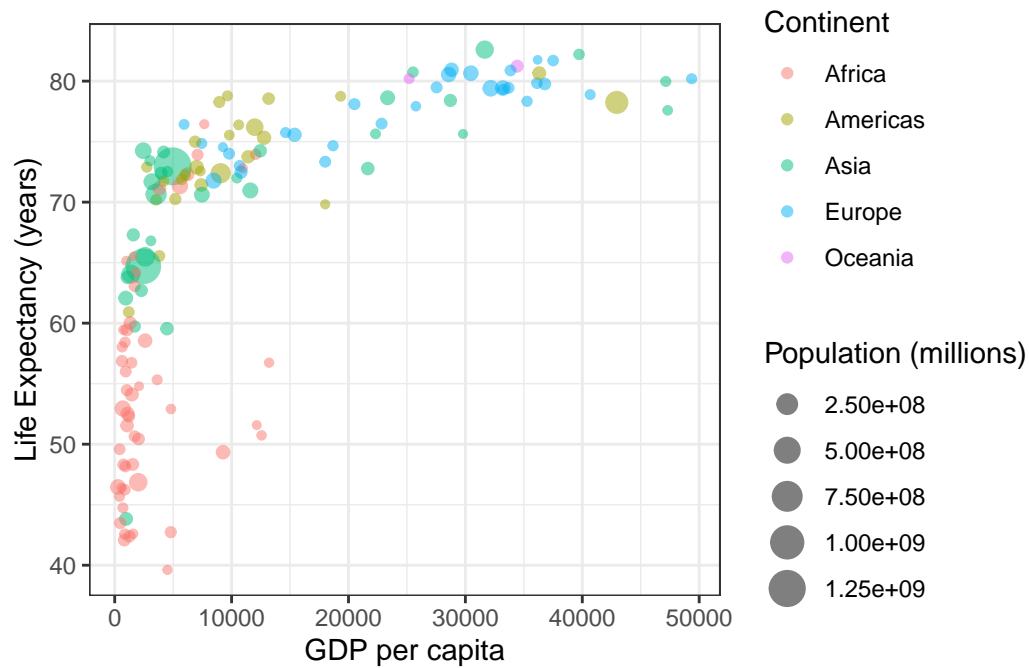
```
# ggplotly(ggplot(data=genes, aes(x=Condition1, y=Condition2, color = State, name = Gene))
#   geom_point() +
#   scale_colour_manual(labels=c("Down", "Unchanging", "Up"), values=c("blue","gray","red"))
#   xlab("Control (no drug)") + ylab("Drug treatment") + ggtitle("Gene Expression Changes Upon Drug Treatment")
#   theme_bw())
```

## Dataset: gapminder

```
url <- "https://raw.githubusercontent.com/jennybc/gapminder/master/inst/extdata/gapminder.  
gapminder <- read.delim(url)  
head(gapminder)
```

	country	continent	year	lifeExp	pop	gdpPercap
1	Afghanistan	Asia	1952	28.801	8425333	779.4453
2	Afghanistan	Asia	1957	30.332	9240934	820.8530
3	Afghanistan	Asia	1962	31.997	10267083	853.1007
4	Afghanistan	Asia	1967	34.020	11537966	836.1971
5	Afghanistan	Asia	1972	36.088	13079460	739.9811
6	Afghanistan	Asia	1977	38.438	14880372	786.1134

```
library(dplyr)  
  
gapminder_2007 <- gapminder %>% filter(year==2007)  
  
ggplot(data=gapminder_2007) +  
  aes(x= gdpPercap, y=lifeExp, color=continent, size=pop) +  
  geom_point(alpha = 0.5) + #make the dots slightly transparent  
  labs(size = "Population (millions)", color = "Continent", x="GDP per capita", y="Life Ex  
  theme_bw()
```



## Dataset: USArrests

```
head(USArrests)
```

	Murder	Assault	UrbanPop	Rape
Alabama	13.2	236	58	21.2
Alaska	10.0	263	48	44.5
Arizona	8.1	294	80	31.0
Arkansas	8.8	190	50	19.5
California	9.0	276	91	40.6
Colorado	7.9	204	78	38.7

```
USArrests$State <- rownames(USArrests)

ggplot(USArrests) +
  aes(x=reorder(State,Murder), y=Murder) +
  geom_point() +
  geom_segment(aes(x=State,
                  xend=State,
                  y=0,
```

```

    yend=Murder), color="blue") +
coord_flip() +
xlab("State") + ylab("Number of murders") +
theme_bw()

```

