

Class 5: Data visualization with ggplot

Jennifer Swindlehurst Chan

Table of contents

Our first plot	1
Cars	1
Gene expression changes	4
Countries	6
US arrests	9

Our first plot

Cars

R has base graphics

```
head(cars)
```

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

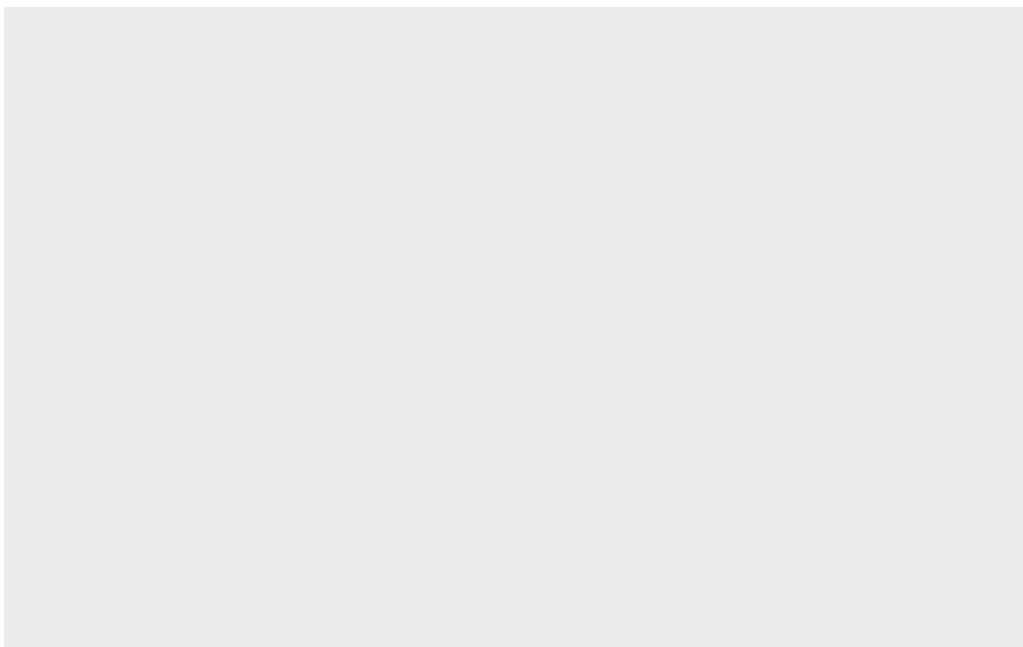
How would I plot this with ggplot2?

We need to install and download the ggplot2 package first. To install any package in R we use the `install.packages()` function.

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

Before I can use this package I need to load it with a `library()` call.

```
library(ggplot2)  
ggplot(cars)
```



Every ggplot needs at least three layers:

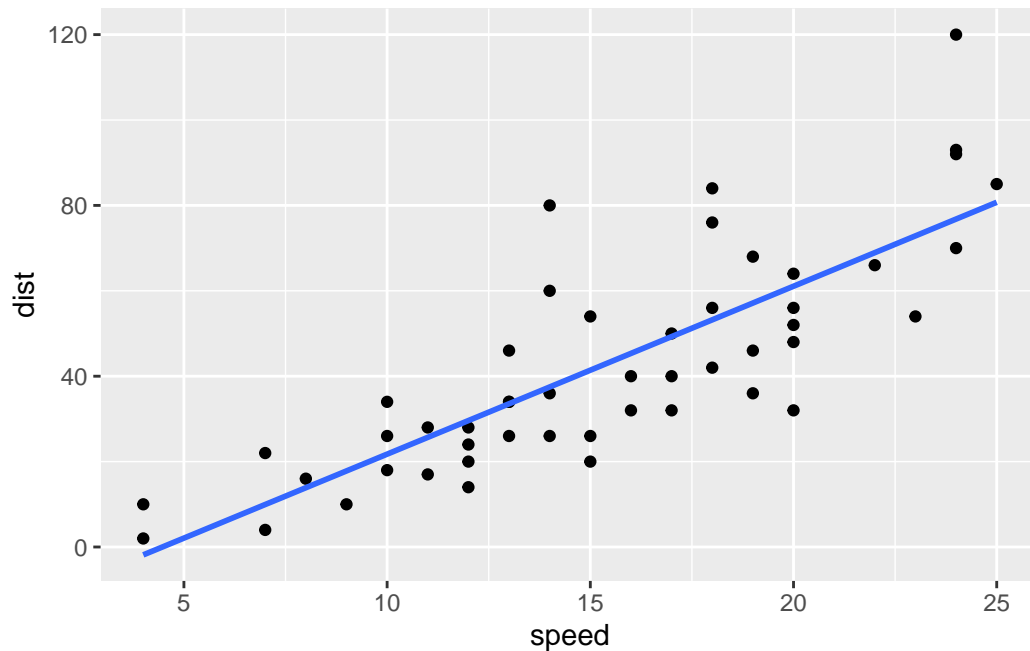
- **Data** (i.e. the data.frame we have)
- **Aes** (the aesthetic mapping of our data to what we want to plot)
- **Geoms** (how we want to plot this stuff!)

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



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

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



Gene expression changes

```
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

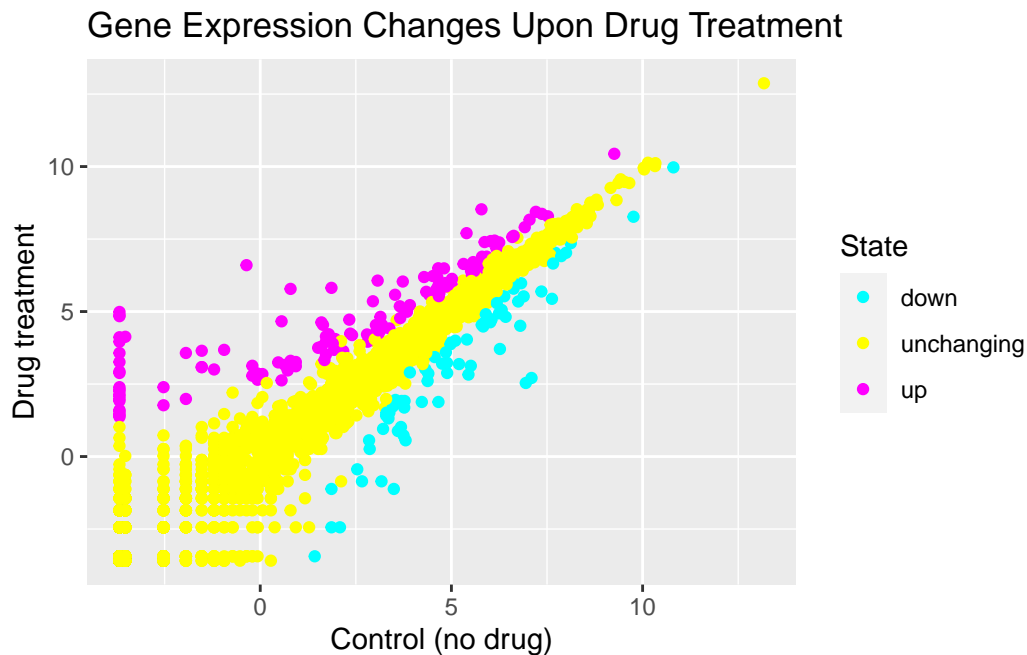
Q1. How many genes are in this data set

```
nrow(genes)
```

```
[1] 5196
```

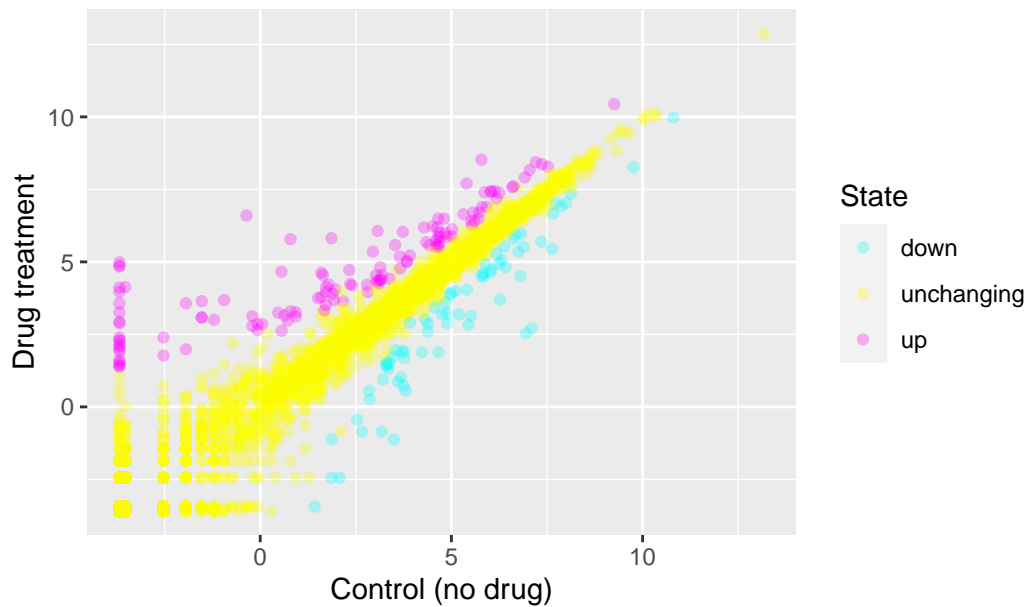
There are 5196 genes in this data set.

```
ggplot(data=genes) +  
  aes(x=Condition1, y=Condition2, col=State) +  
  geom_point() +  
  scale_color_manual(values=c("cyan", "yellow", "magenta")) +  
  ggtitle("Gene Expression Changes Upon Drug Treatment") +  
  xlab("Control (no drug)") +  
  ylab("Drug treatment")
```



```
ggplot(data=genes) +  
  aes(x=Condition1, y=Condition2, col=State) +  
  geom_point(alpha=0.3) +  
  scale_color_manual(values=c("cyan", "yellow", "magenta")) +  
  labs(title = "Gene Expression Changes Upon Drug Treatment", x = "Control (no drug)", y =
```

Gene Expression Changes Upon Drug Treatment



?ggplot

Countries

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

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

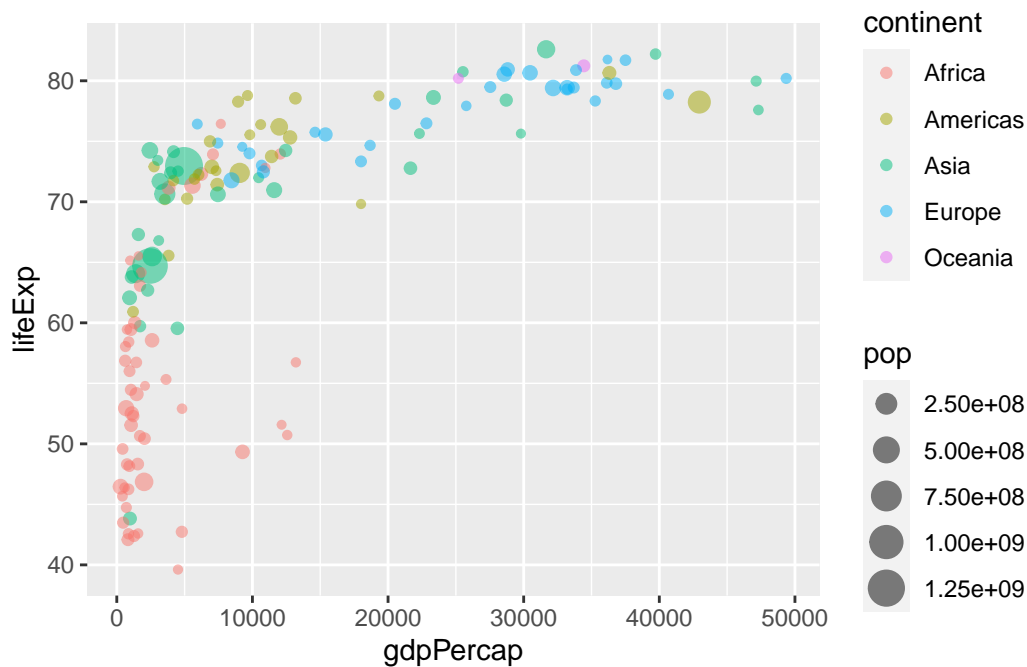
intersect, setdiff, setequal, union

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

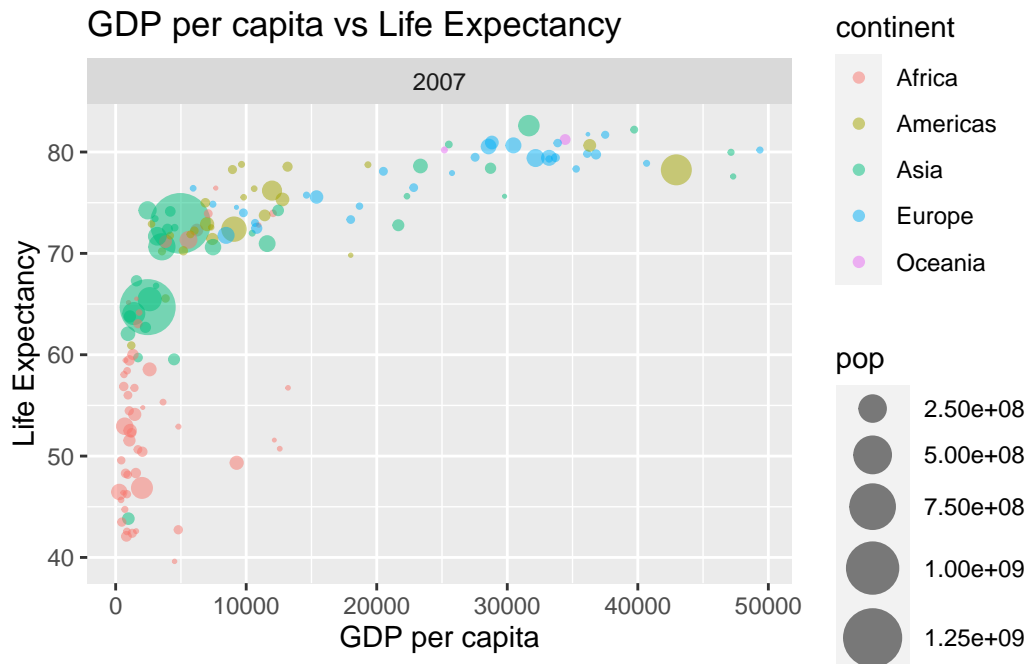
```
head(gapminder_2007)
```

	country	continent	year	lifeExp	pop	gdpPercap
1	Afghanistan	Asia	2007	43.828	31889923	974.5803
2	Albania	Europe	2007	76.423	3600523	5937.0295
3	Algeria	Africa	2007	72.301	33333216	6223.3675
4	Angola	Africa	2007	42.731	12420476	4797.2313
5	Argentina	Americas	2007	75.320	40301927	12779.3796
6	Australia	Oceania	2007	81.235	20434176	34435.3674

```
ggplot(gapminder_2007) +  
  aes(x=gdpPercap, y=lifeExp, color=continent, size=pop) +  
  geom_point(alpha=0.5)
```



```
ggplot(gapminder_2007) +
  geom_point(aes(x=gdpPercap, y=lifeExp, size=pop, color=continent), alpha=0.5) +
  scale_size_area(max_size = 10)+
  facet_wrap(2007)+
  labs(title="GDP per capita vs Life Expectancy", x="GDP per capita", y="Life Expectancy")
```

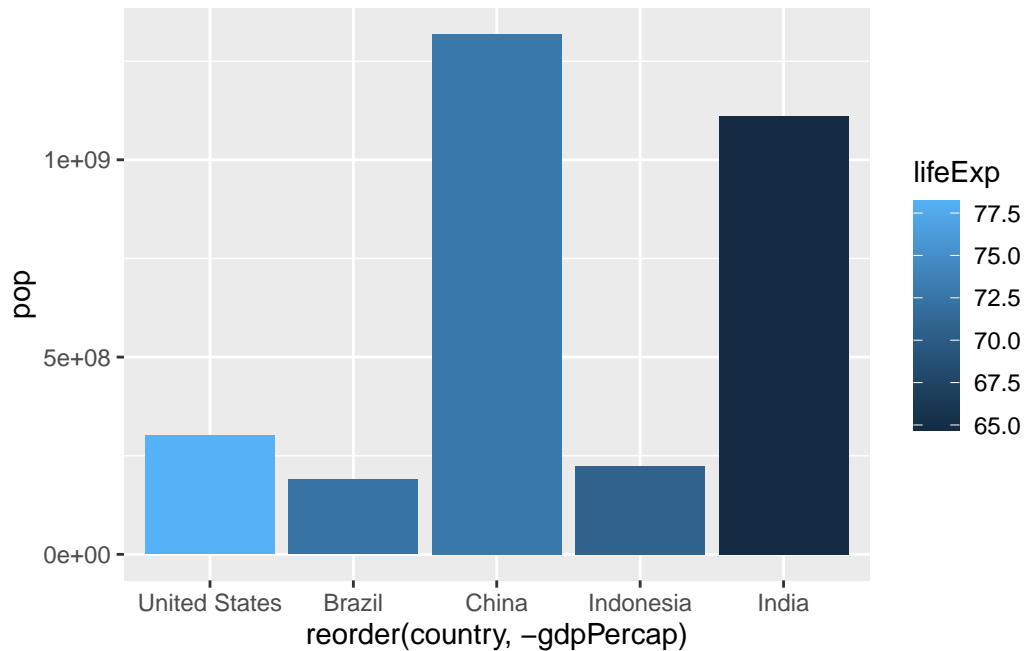


```
gapminder_top5 <- gapminder %>%
  filter(year==2007) %>%
  arrange(desc(pop)) %>%
  top_n(5, pop)
```

```
gapminder_top5
```

	country	continent	year	lifeExp	pop	gdpPercap
1	China	Asia	2007	72.961	1318683096	4959.115
2	India	Asia	2007	64.698	1110396331	2452.210
3	United States	Americas	2007	78.242	301139947	42951.653
4	Indonesia	Asia	2007	70.650	223547000	3540.652
5	Brazil	Americas	2007	72.390	190010647	9065.801


```
ggplot(gapminder_top5) +
  aes(x = reorder(country, -gdpPercap), y = pop, fill = lifeExp)+
  geom_col()
```

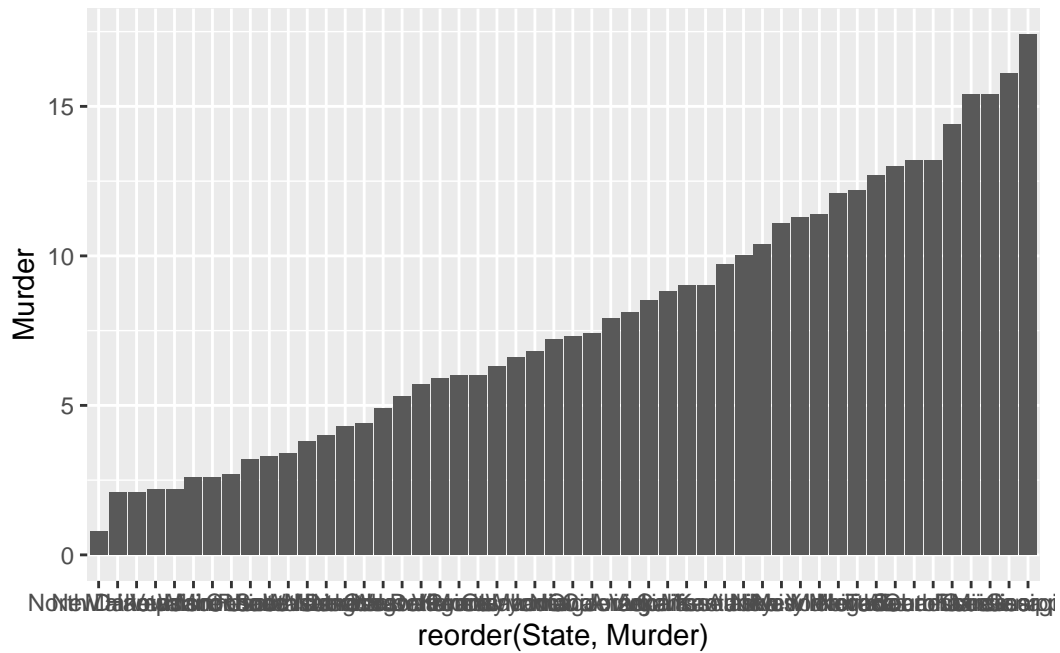


US arrests

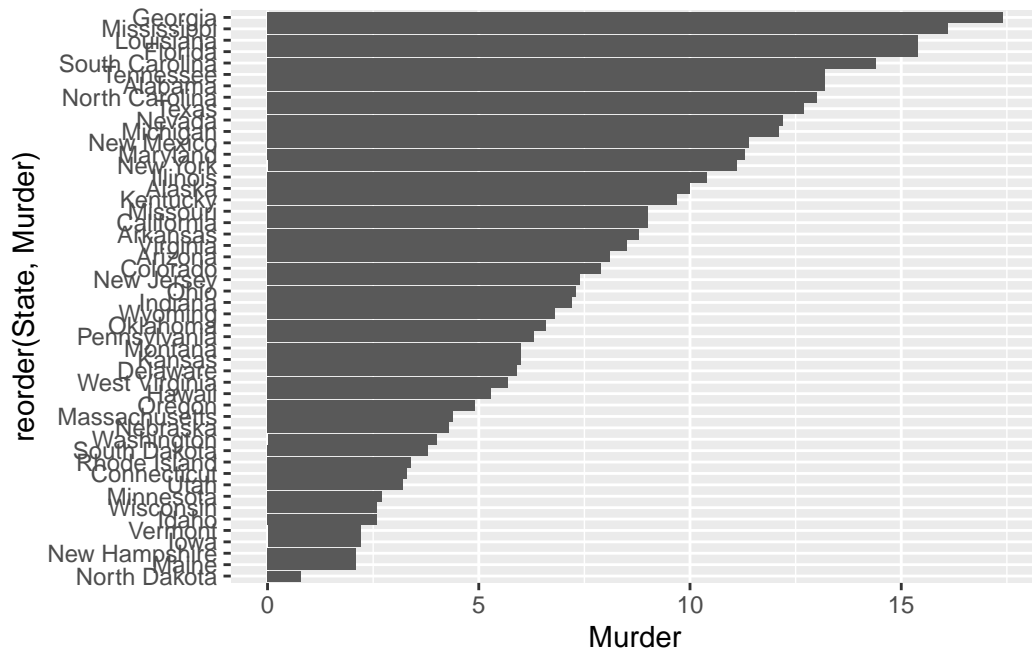
```
head(USArrests)
```

	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

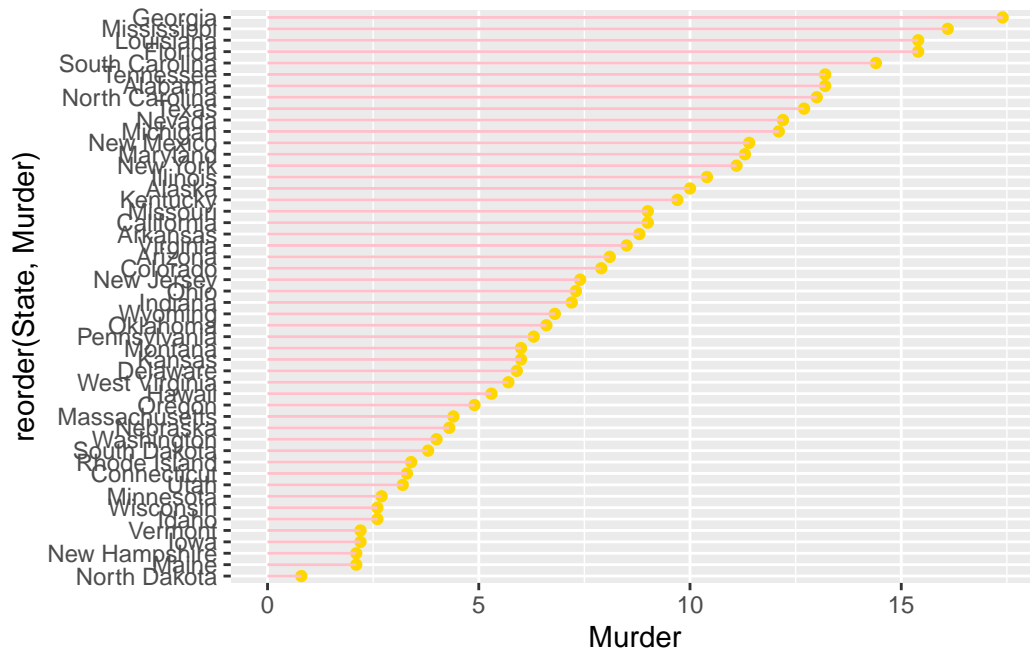
```
USArrests$State <- rownames(USArrests)
ggplot(USArrests) +
  aes(x=reorder(State,Murder), y=Murder) +
  geom_col()
```



```
USArrests$State <- rownames(USArrests)
ggplot(USArrests) +
  aes(x=reorder(State,Murder), y=Murder) +
  geom_col() +
  coord_flip()
```



```
ggplot(USArrests) +
  aes(x=reorder(State,Murder), y=Murder) +
  geom_point(color = "gold") +
  geom_segment(aes(x=State,
                  xend=State,
                  y=0,
                  yend=Murder), color="pink") +
  coord_flip()
```



```
#install.packages("gifski")
#install.packages("gganimate")
```

```
library(gapminder)
library(gganimate)
```

```
#ggplot(gapminder, aes(gdpPercap, lifeExp, size = pop, colour = country)) +
  #geom_point(alpha = 0.7, show.legend = FALSE) +
  #scale_colour_manual(values = country_colors) +
  #scale_size(range = c(2, 12)) +
  #scale_x_log10() +
  #facet_wrap(~continent) +
  #labs(title = 'Year: {frame_time}', x = 'GDP per capita', y = 'life expectancy') +
  #transition_time(year) +
  #shadow_wake(wake_length = 0.1, alpha = FALSE)
```

```
#install.packages("patchwork")
```

```
library(patchwork)
```

```

p1 <- ggplot(mtcars) + geom_point(aes(mpg, disp))
p2 <- ggplot(mtcars) + geom_boxplot(aes(gear, disp, group = gear))
p3 <- ggplot(mtcars) + geom_smooth(aes(dis, qsec))
p4 <- ggplot(mtcars) + geom_bar(aes(carb))

(p1 | p2 | p3) /
  p4

```

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

