

Class 5: Data Visualization with GGLOT

Julia Ainsworth

Table of contents

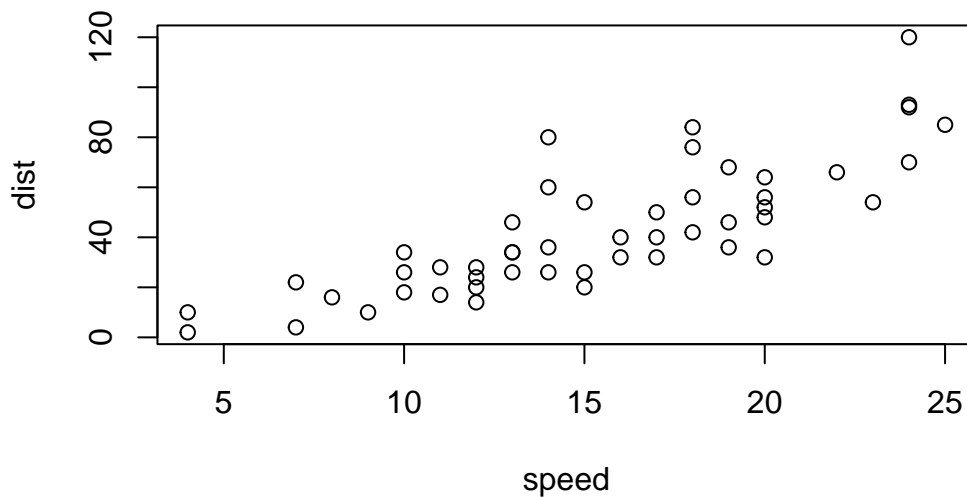
Our first plot	1
New graph for genes upregulated/downregulated with drug treatment	4
gapminder extension	7
Bar charts	10
Animation:	15
Combining plots for 10:	16

Our first plot

R has base graphics

Keyboard shortcut to insert a chunk of code: opt + cmd + i

```
plot(cars)
```



How would I plot this with ggplot2?

If I hadn't installed ggplot2 yet, this would be the first step:

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

But before I use it, need to re-load (called a “library call”) in every session. This is the case for every package ‘library(ggplot2)’

```
# library(ggplot2)
library(ggplot2)
```

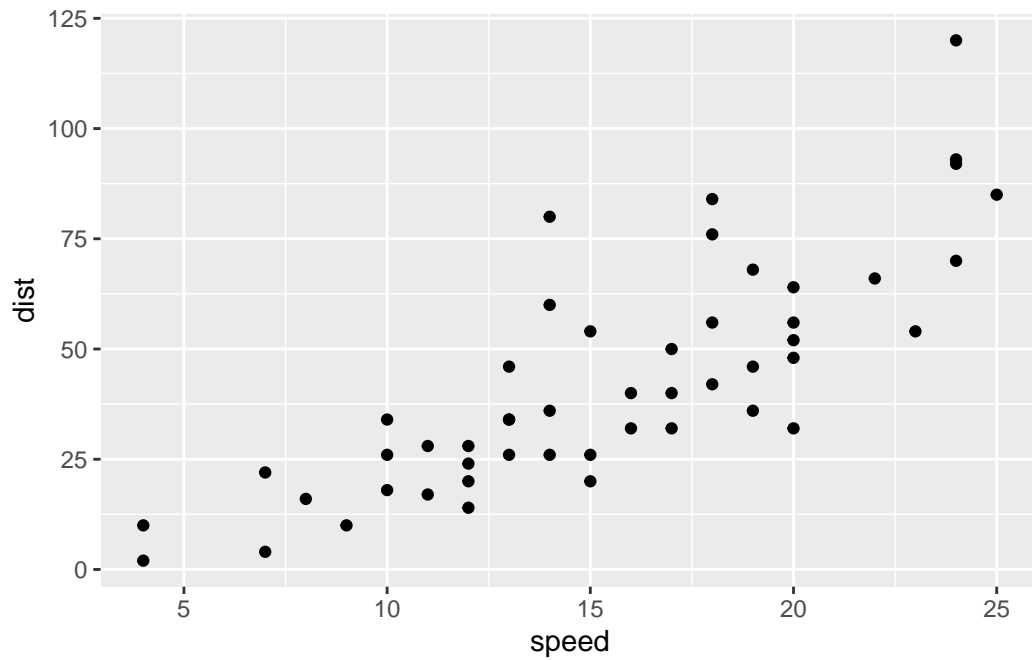
Every ggplot needs at least 3 layers:

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

Simplest version of a ggplot:

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

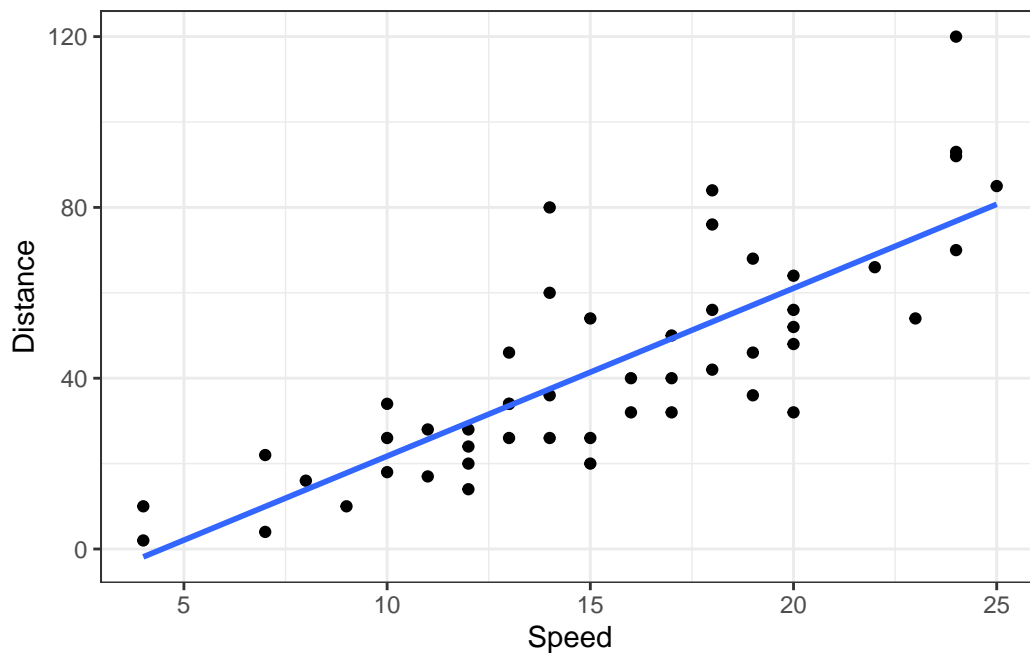
```
geom_point()
```



Adding a trend line to our data:

```
ggplot(data=cars) +  
  aes(x=speed, y=dist) +  
  geom_point() +  
  geom_smooth(method = lm, se = FALSE)+  
  xlab("Speed") +  
  ylab("Distance") +  
  theme_bw()
```

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



New graph for genes upregulated/downregulated with drug treatment

```
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

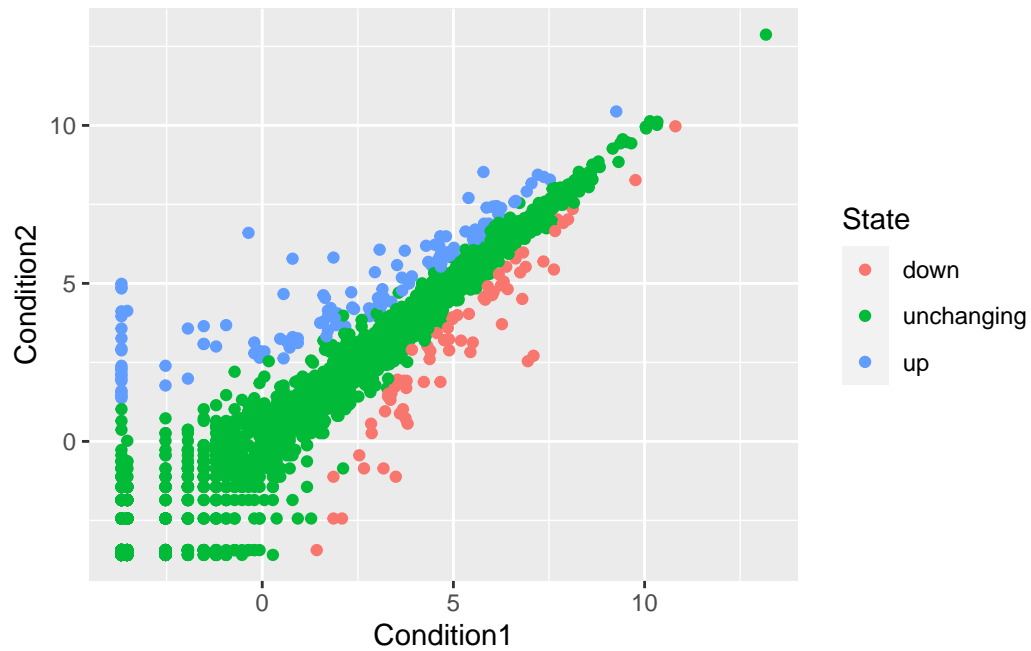
```
round( table(genes$State)/nrow(genes) * 100, 2 )
```

down	unchanging	up
1.39	96.17	2.44

Answers to questions: Q: There are 5196 genes in this data set.

Q: There are `ncol(genes)` columns in the genes data set. Graph of the genes up and down-regulated:

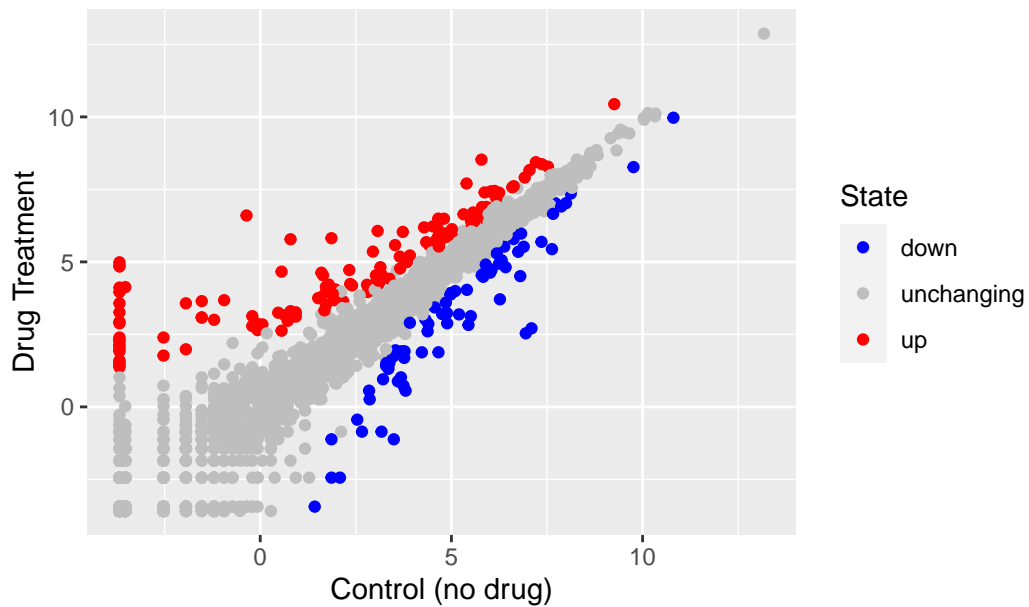
```
p <- ggplot(genes) +  
  aes(Condition1, Condition2, col=State) +  
  geom_point()  
p
```



Graph showing the color changing: does p carry over? Answer: yes!

```
p +
  scale_color_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



gapminder extension

```
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':

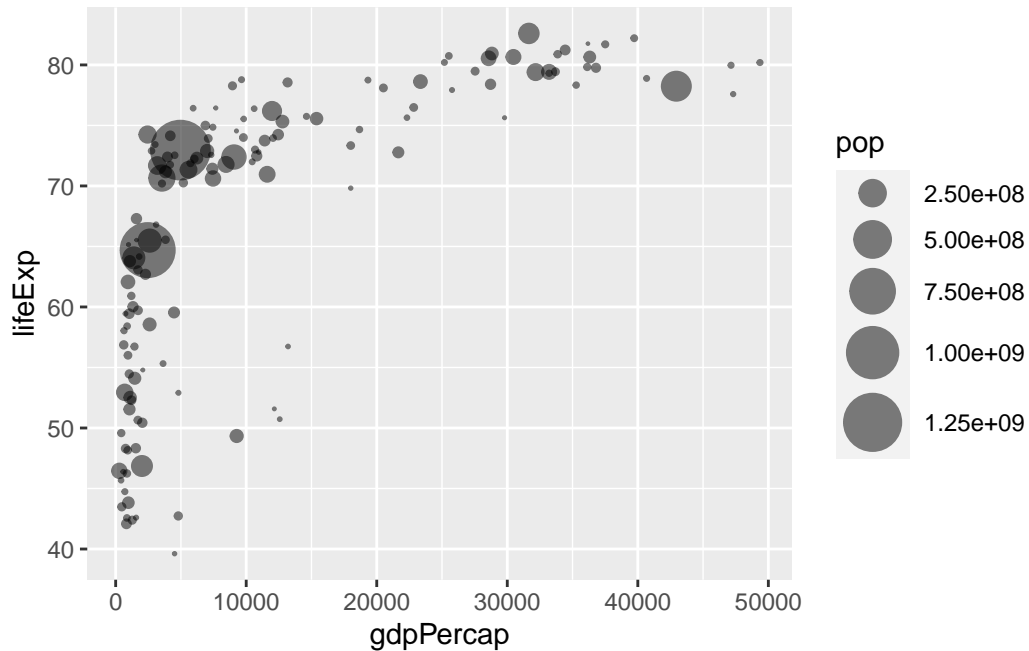
intersect, setdiff, setequal, union

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

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

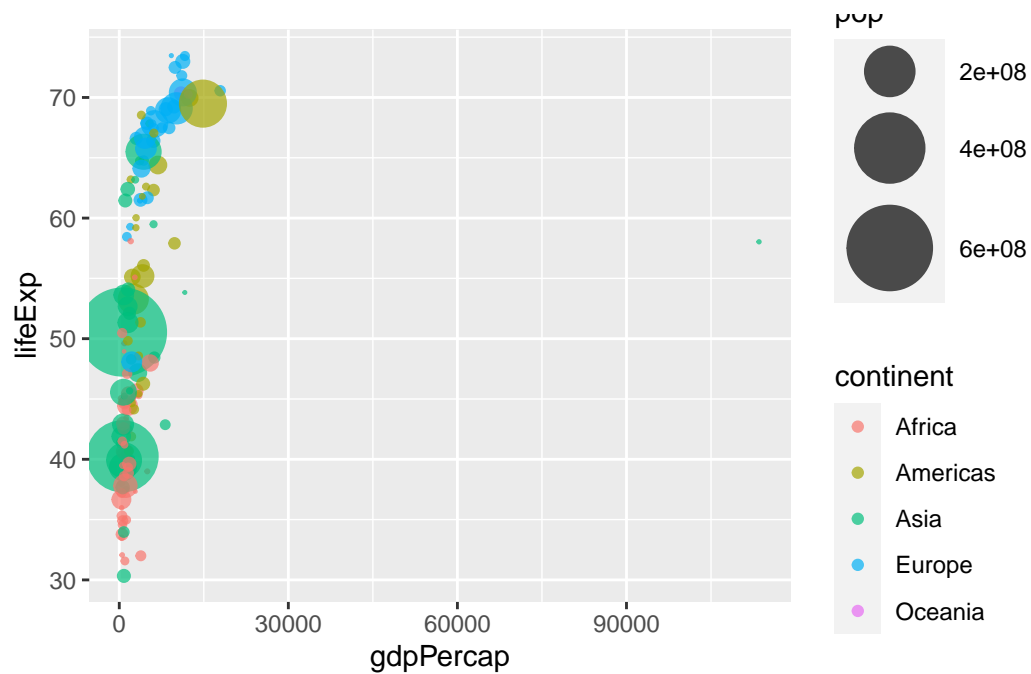
First basic scatterplot of gapminder

```
ggplot(gapminder_2007) +  
  aes(gdpPercap, lifeExp, size= pop) +  
  geom_point(alpha = 0.5) +  
  scale_size_area(max_size = 10)
```



Redoing it for 1957

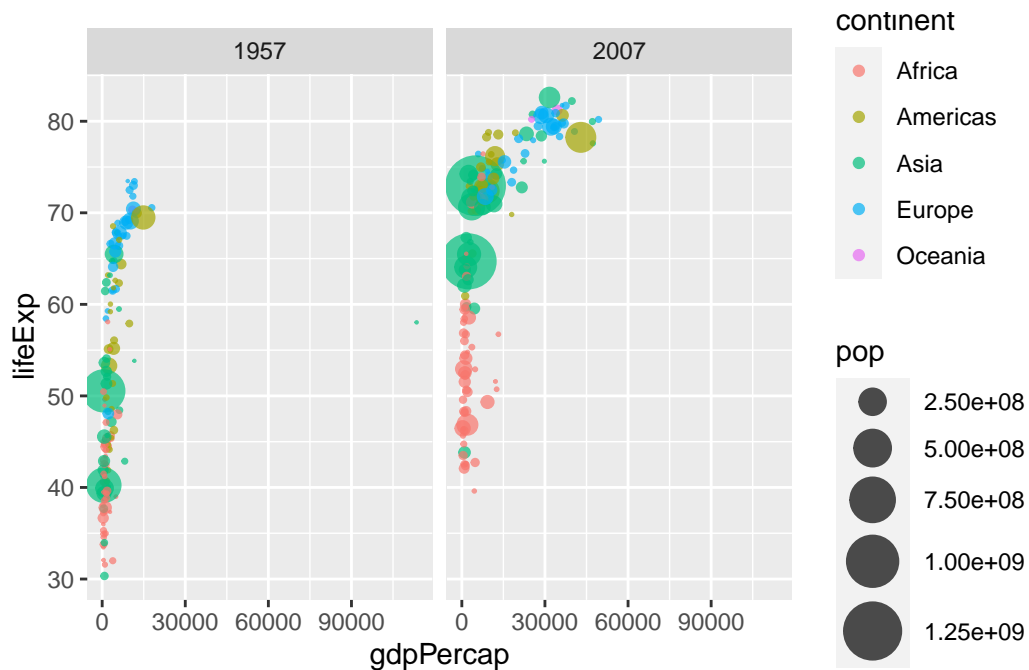
```
gapminder_1957 <- gapminder %>% filter(year==1957)  
  
ggplot(gapminder_1957) +  
  aes(gdpPercap, lifeExp, col = continent, size = pop) +  
  geom_point(alpha = 0.7) +  
  scale_size_area(max_size = 15)
```

Facet wrapping: getting two graphs next to each other

```
gapminder_1957 <- gapminder %>% filter(year==1957 | year==2007)

ggplot(gapminder_1957) +
  aes(gdpPercap, lifeExp, color = continent, size = pop)+
  geom_point(alpha = 0.7) +
  scale_size_area(max_size = 10) +
  facet_wrap(~year)
```



Bar charts

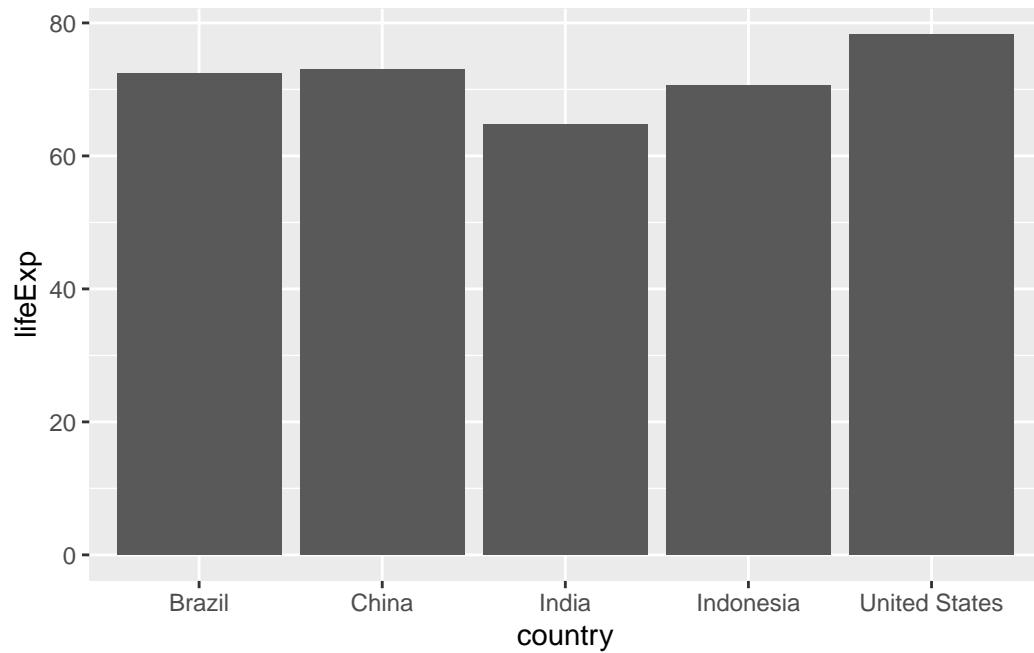
```
gapminder_top5 <- gapminder %>%
  filter(year==2007) %>%
  arrange(desc(lifeExp)) %>%
  top_n(5, pop)
```

```
gapminder_top5
```

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

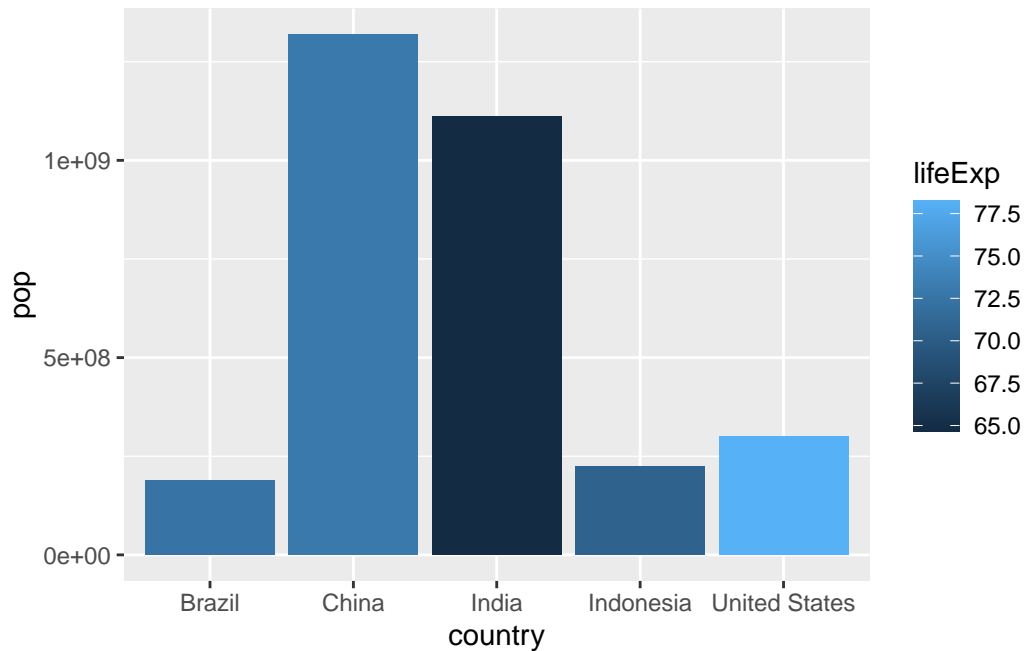
Bar chart:

```
ggplot(gapminder_top5) +
  geom_col(aes(x = country, y = lifeExp))
```



Putting colors in:

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



Plotting population size by country:

```
gapminder_top5pop <- gapminder %>%
  filter(year==2007) %>%
  arrange(desc(pop)) %>%
  top_n(5, pop)

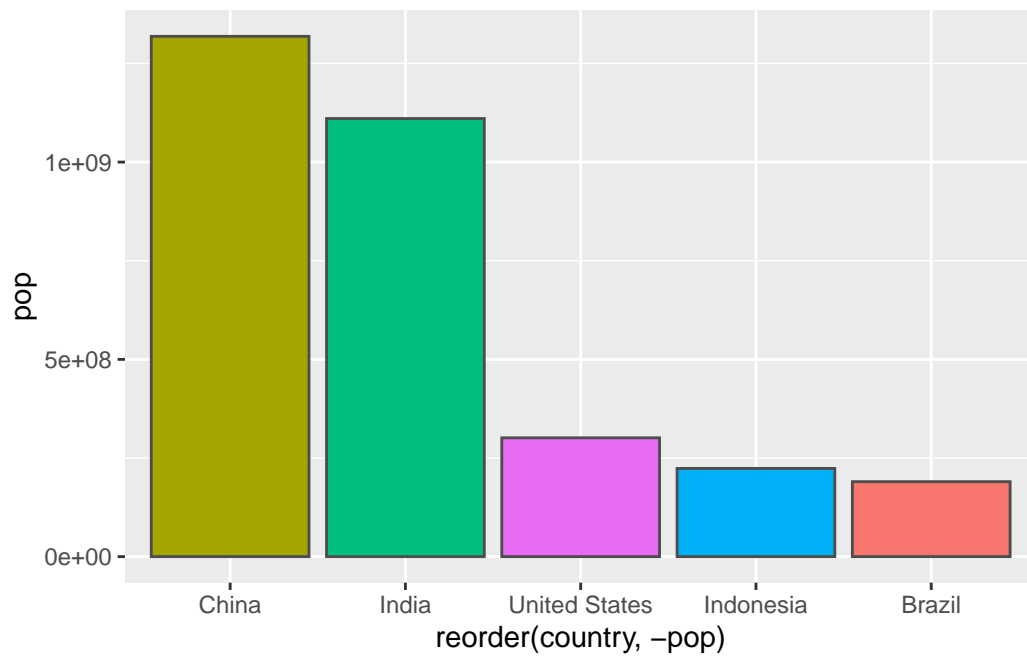
gapminder_top5pop
```

	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

Barplot:

```
ggplot(gapminder_top5pop) +
  aes(x=reorder(country, -pop), y=pop, fill = country)+
  geom_col(col="gray30") +
```

```
guides(fill="none")
```

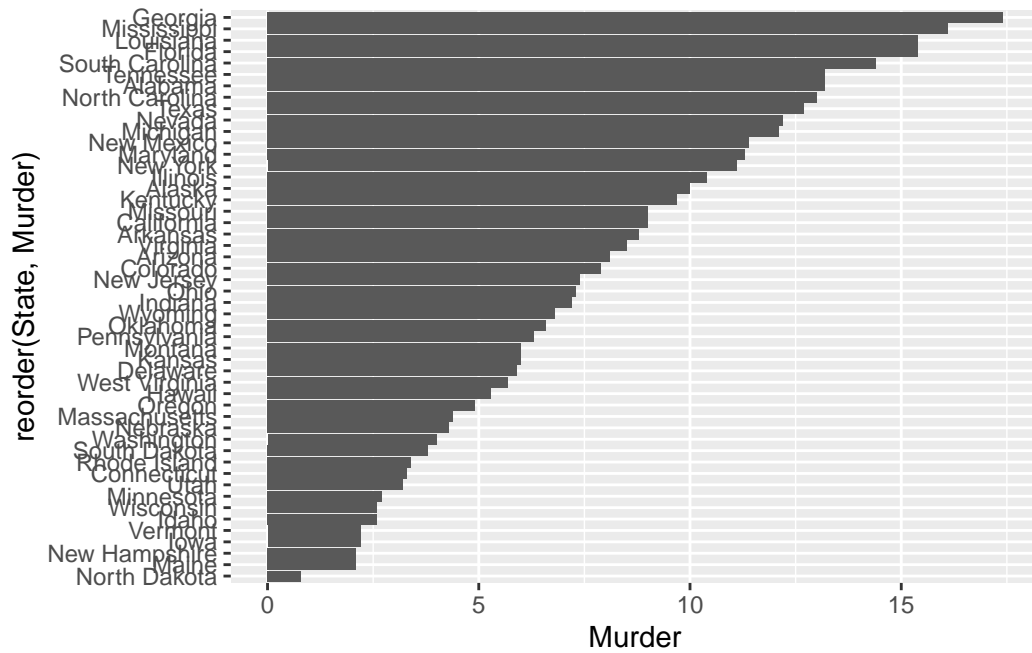


Flipping bar charts:

```
head(USArrests)
```

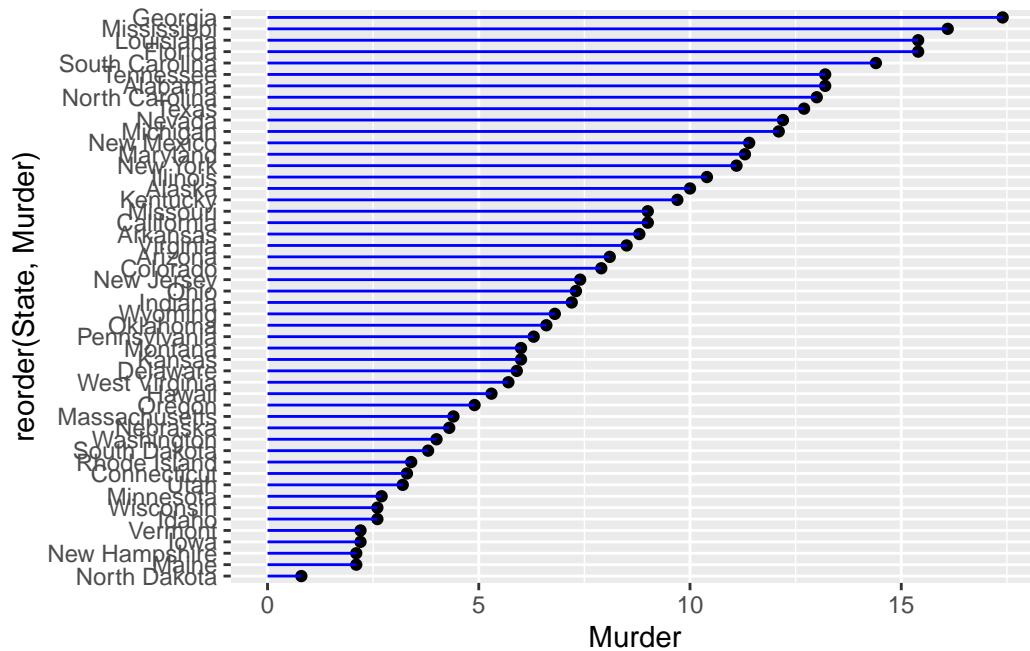
	Murder	Assault	UrbanPop	Rape
Alabama	13.2	236	58	21.2
Alaska	10.0	263	48	44.5
Arizona	8.1	294	80	31.0
Arkansas	8.8	190	50	19.5
California	9.0	276	91	40.6
Colorado	7.9	204	78	38.7

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



Segmented bar plot:

```
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()
```



Animation:

```
library(gapminder)
library(gganimate)

# Setup nice regular ggplot of the gapminder data
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 by continent
  facet_wrap(~continent) +
  # Here comes the gganimate specific bits
  labs(title = 'Year: {frame_time}', x = 'GDP per capita', y = 'life expectancy') +
  transition_time(year) +
  shadow_wake(wake_length = 0.1, alpha = FALSE)
```

Combining plots for 10:

```
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'

