

Class05: Data Visualization w ggplot

Nicole Alfonso (A16429176)

Graphics systems in R

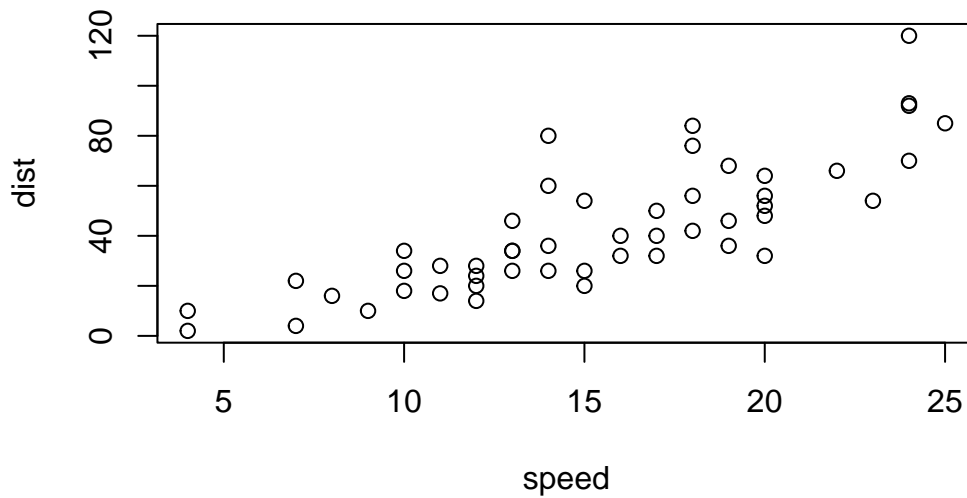
There are many graphics systems in R for making plots and figures.

We have already played a little with “**base R**” graphics and the `plot()` function.

Today we will start learning about a popular graphics package called `ggplot2()`.

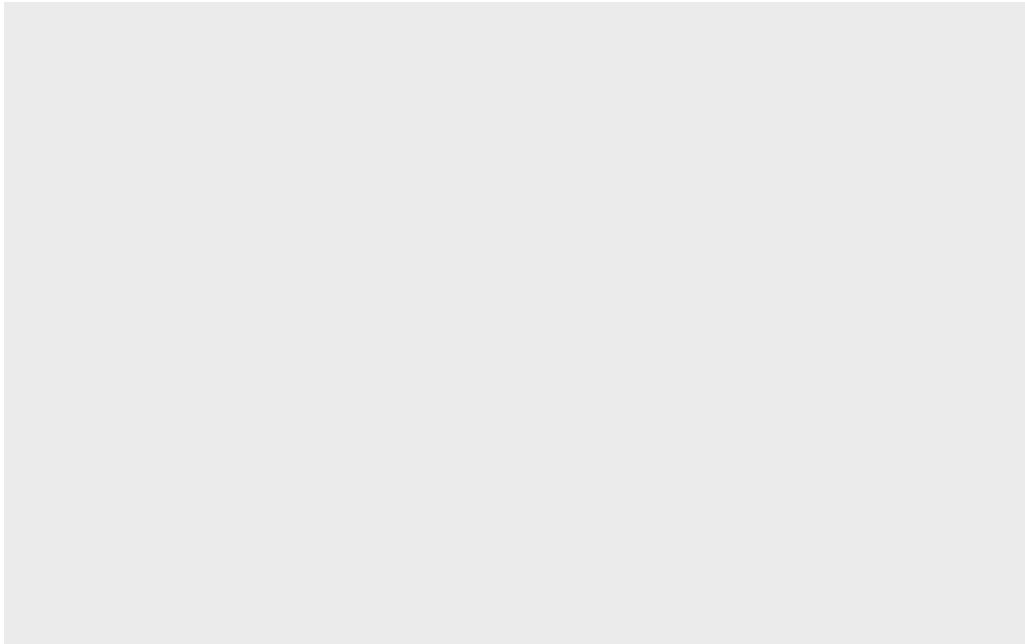
This is an add on package - i.e. we need to install it. I install it like I install any package, with the `install.packages()` function.

```
plot(cars)
```



Before I can use the functions from a package I have to load up the package from my “library”. We use the `library(ggplot2)` command to load it up. You also have to add extra stuff to `ggplot`.

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



Every `ggplot` is made up of at least 3 things: - data (the numbers etc. that will go into your plot) - aesthetics (how the columns of the data map to the plot aesthetics) - geoms (how the plot actually looks, points, bars, lines, etc.)

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

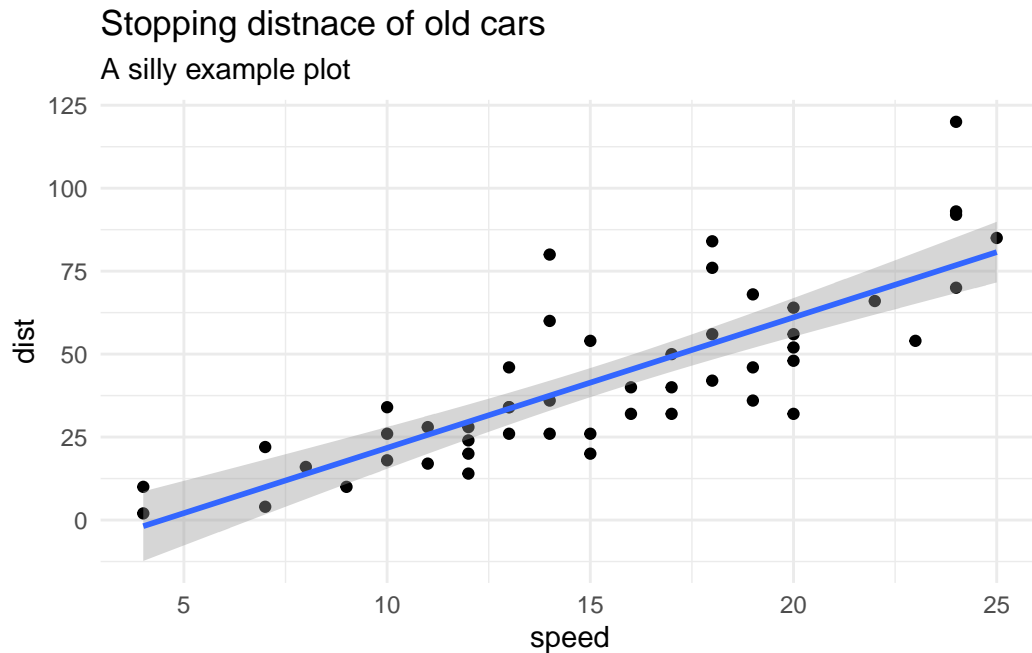


For simple plots, ggplot is more verbose - it takes more code than base R plot.

Add some more layers to our plot

```
ggplot(cars) +  
  aes(x=speed, y=dist)+  
  geom_point() +  
  geom_smooth(method = "lm") +  
  labs(title = "Stopping distnace of old cars",  
        subtitle = "A silly example plot") +  
  theme_minimal()
```

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



```
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
```

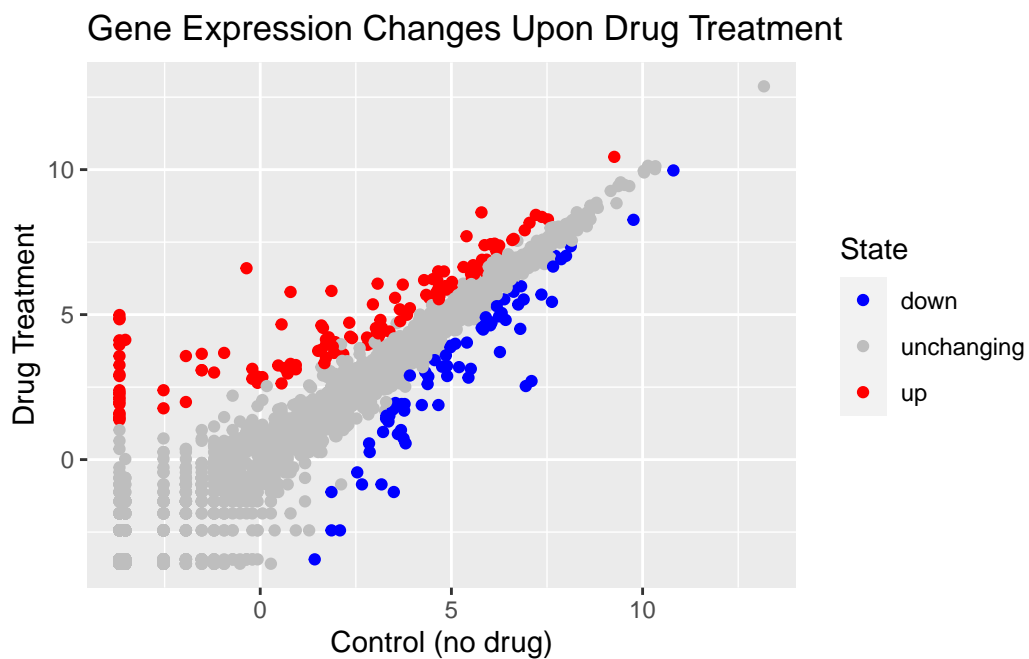
```
table(genes$State)
```

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

```
(127/5196)*100
```

```
[1] 2.444188
```

```
p <- ggplot(genes) +  
  aes(x=Condition1, y=Condition2, col=State) +  
  geom_point() +  
  labs(title = "Gene Expression Changes Upon Drug Treatment", x="Control (no drug)", y="Dr  
p + scale_color_manual(values=c("blue", "gray", "red"))
```



Optional Extension

Question 7

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

Attaching package: 'dplyr'

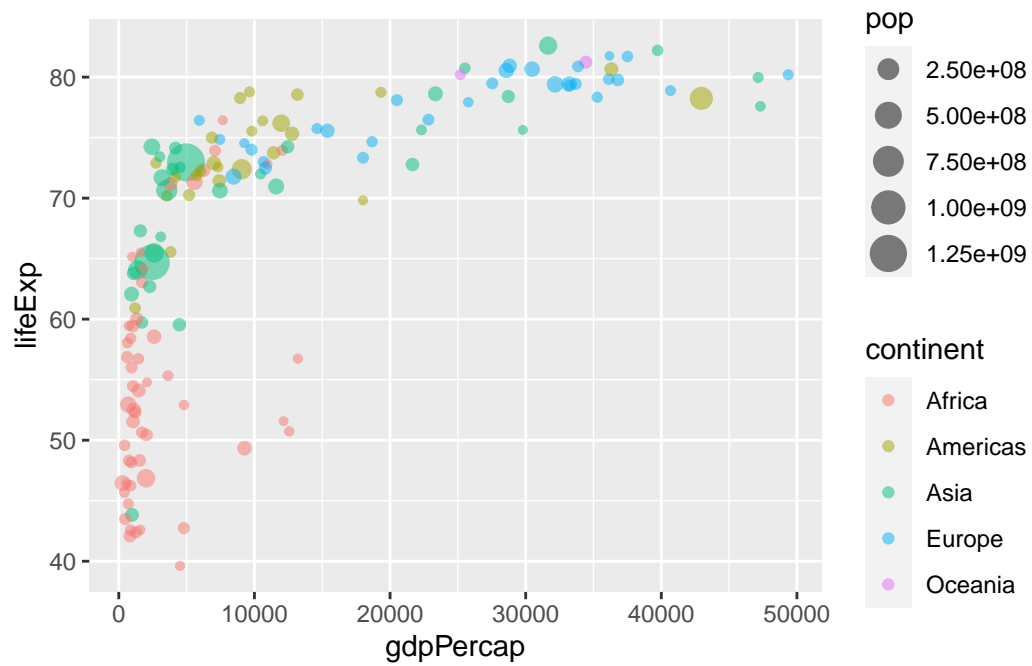
The following objects are masked from 'package:stats':

filter, lag

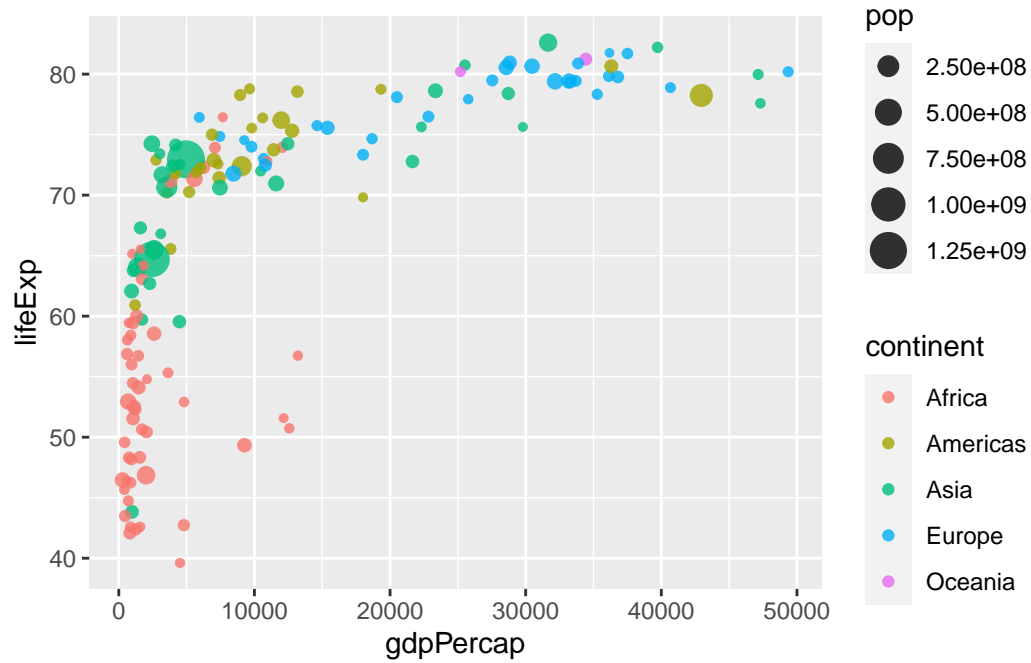
The following objects are masked from 'package:base':

intersect, setdiff, setequal, union

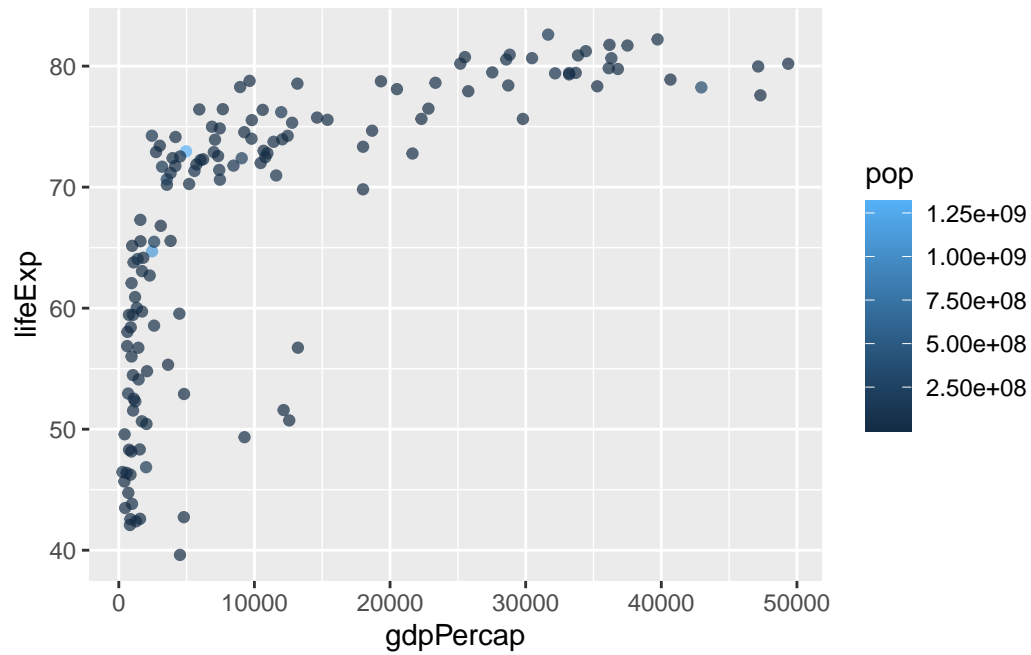
```
gapminder_2007 <- gapminder %>% filter(year==2007)  
  
ggplot(gapminder_2007) +  
  aes(x=gdpPercap, y=lifeExp, color=continent, size=pop) +  
  geom_point(alpha=0.5)
```



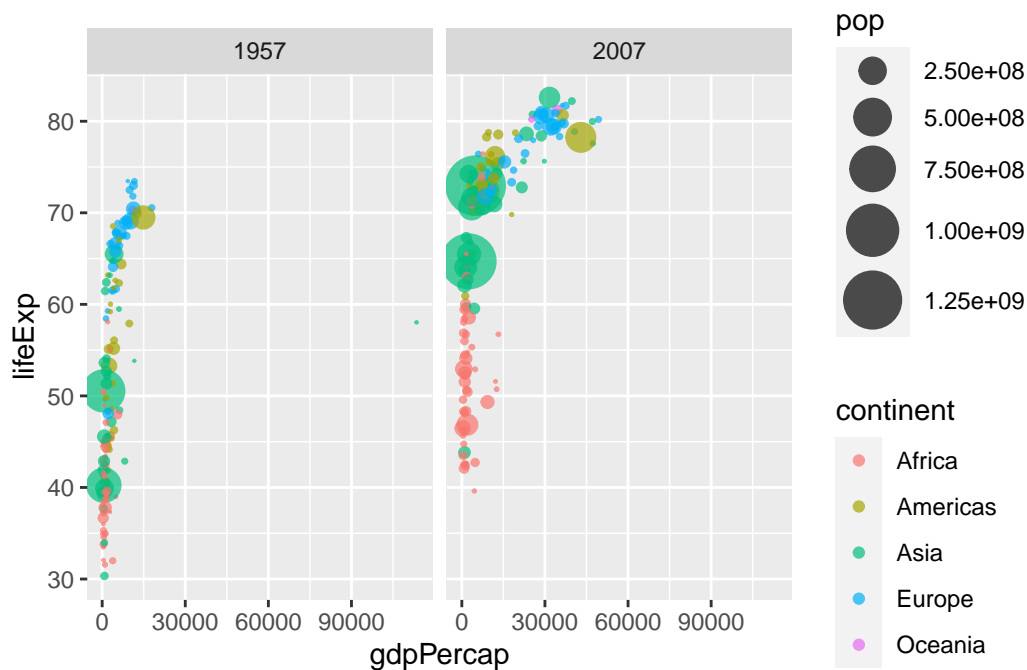
```
ggplot(gapminder_2007) +  
  aes(x=gdpPercap, y=lifeExp, color=continent, size=pop) +  
  geom_point(alpha=0.8)
```



```
ggplot(gapminder_2007) +  
  aes(x=gdpPercap, y=lifeExp, color=pop) +  
  geom_point(alpha=0.7) +  
  scale_size_area(max_size = 10)
```

```
gapminder_1957 <- gapminder %>% filter(year==1957 | year==2007)
ggplot(gapminder_1957) +
  aes(x=gdpPercap, y=lifeExp, color=continent, size=pop) +
  geom_point(alpha=0.7) +
  scale_size_area(max_size = 10) +
  facet_wrap(~year)
```

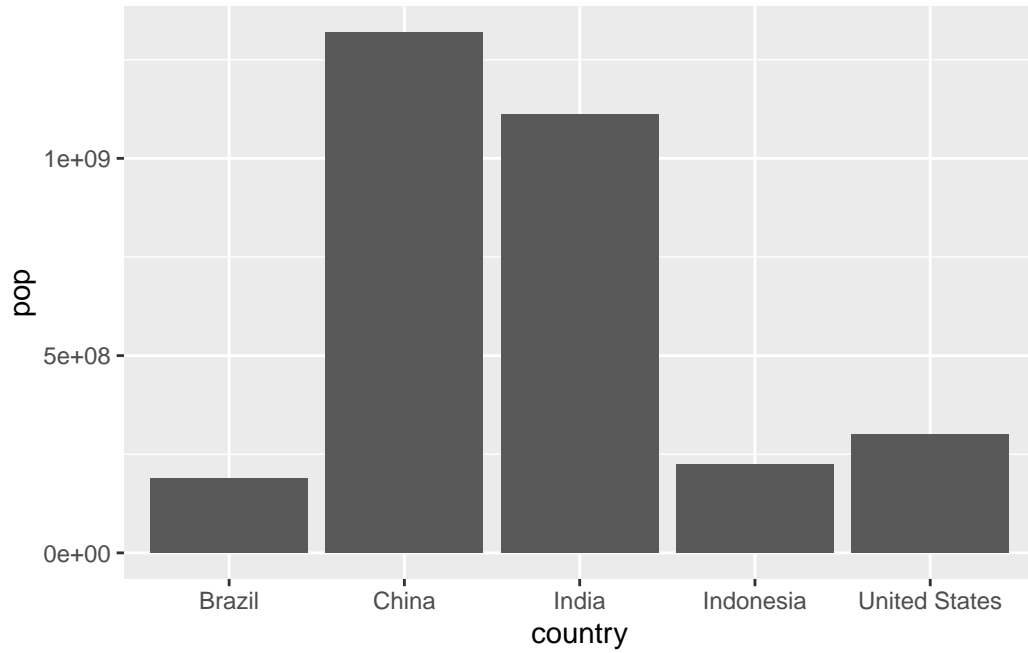


Question 8: Bar Chart

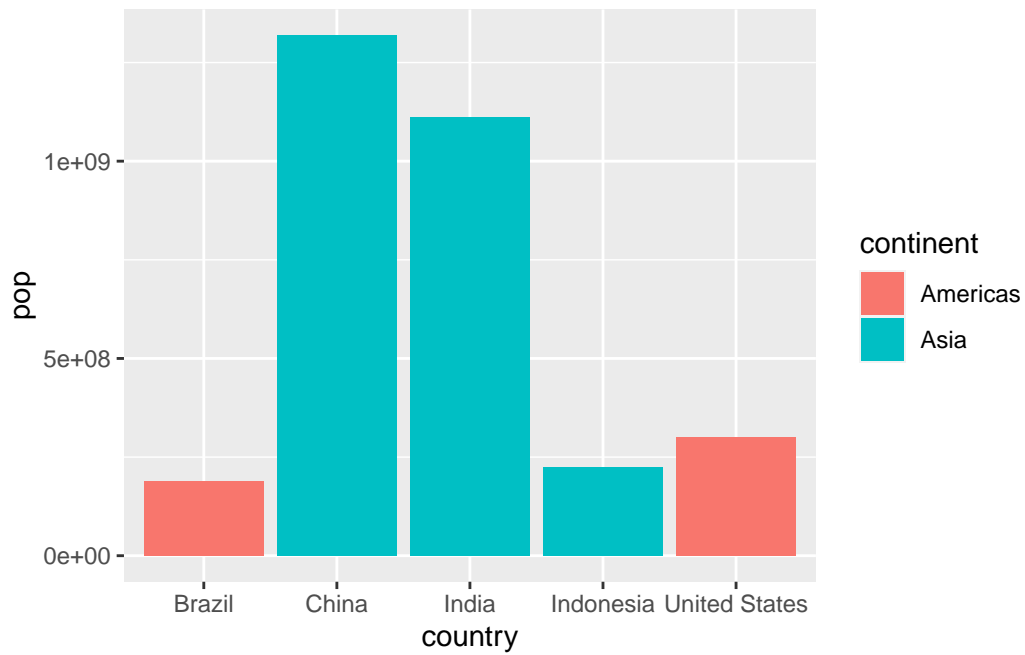
```
gapminder_top5 <- gapminder %>% filter (year==2007) %>% arrange(desc(pop)) %>% top_n(5, pop)
gapminder_top5
```

	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

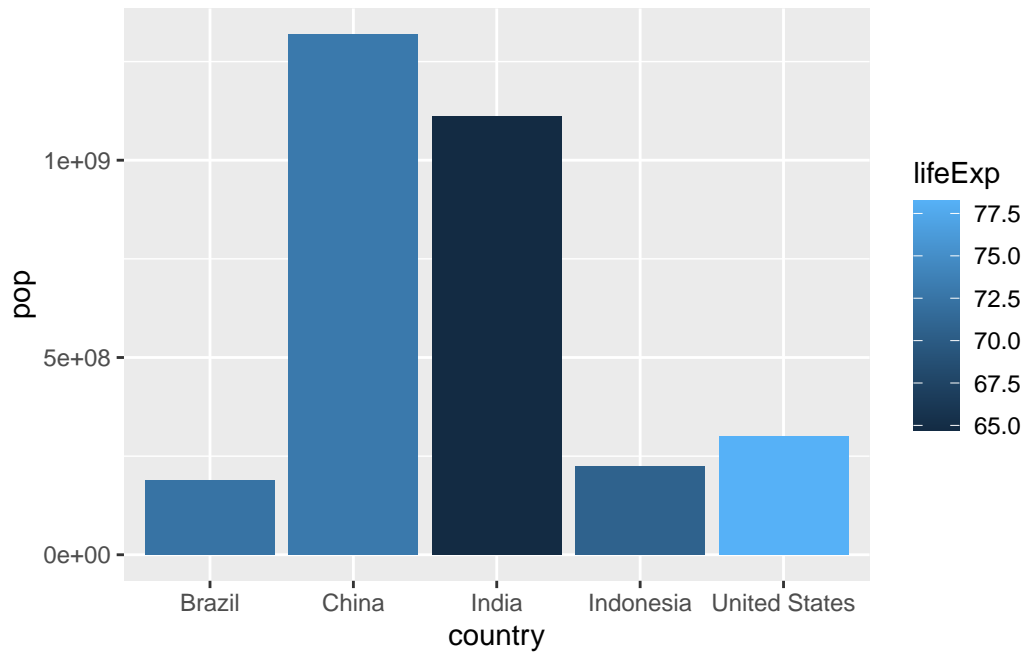
```
ggplot(gapminder_top5) +
  geom_col(aes(x=country, y=pop))
```



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



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



```
ggplot(gapminder_top5) +  
  (aes(x=reorder(country, -pop), y=pop, fill=country)) +  
  geom_col(col="gray30") +  
  guides(fill="none")
```

