

Class 5: Data visualization with ggplot2

Ellice Wang

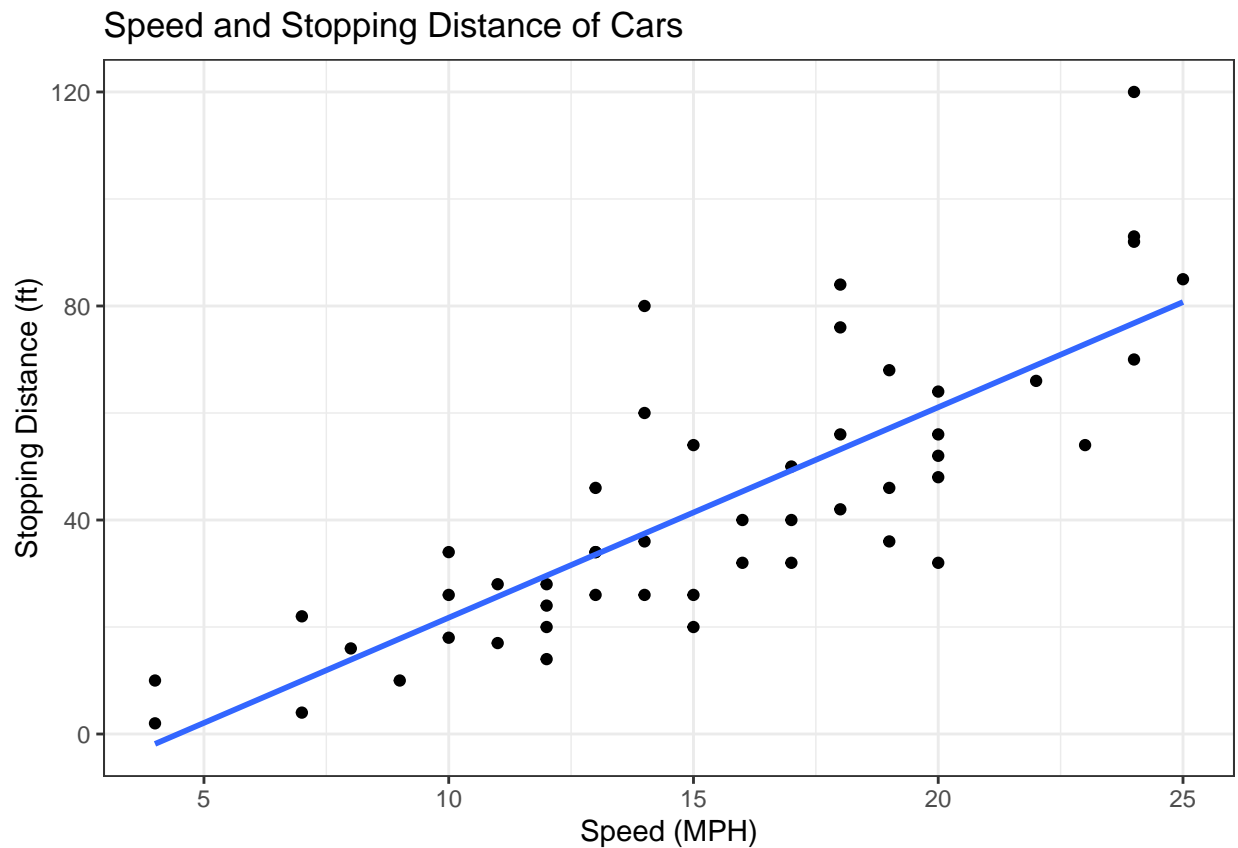
2025-01-15

Class 5: Data visualization with ggplot2

```
#Load in ggplot2
library(ggplot2)

# plot speed and stopping distance of car dataset
ggplot(cars) + aes(x=speed, y=dist) + geom_point() + geom_smooth(method=lm,
                                                                    se=FALSE) +
  labs(title="Speed and Stopping Distance of Cars", x="Speed (MPH)",
       y="Stopping Distance (ft)") + theme_bw()

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



```
# load in dataframe from url
url <- "https://bioboot.github.io/bimm143_S20/class-material/up_down_expression.txt"
genes <- read.delim(url)
table(genes$State)
```

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

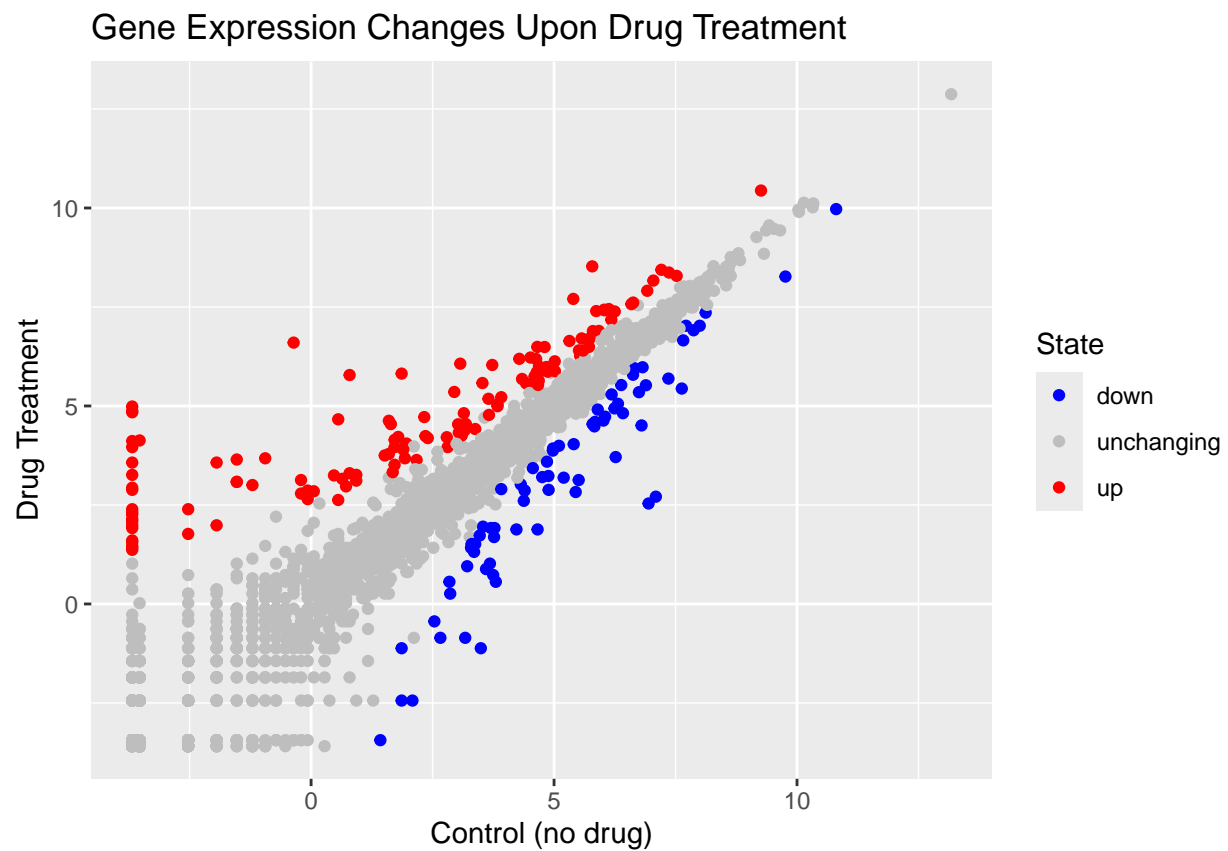
```
# already knew there were 127 up regulated genes based on prev code
127/5196
```

```
## [1] 0.02444188
```

```
# more universal solution to find frequency of all genes
round(table(genes$State)/nrow(genes) * 100, 2)
```

```
##
##      down  unchanging      up
##      1.39      96.17      2.44
```

```
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="Drug Treatment")
p + scale_colour_manual(values=c("blue", "gray", "red"))
```



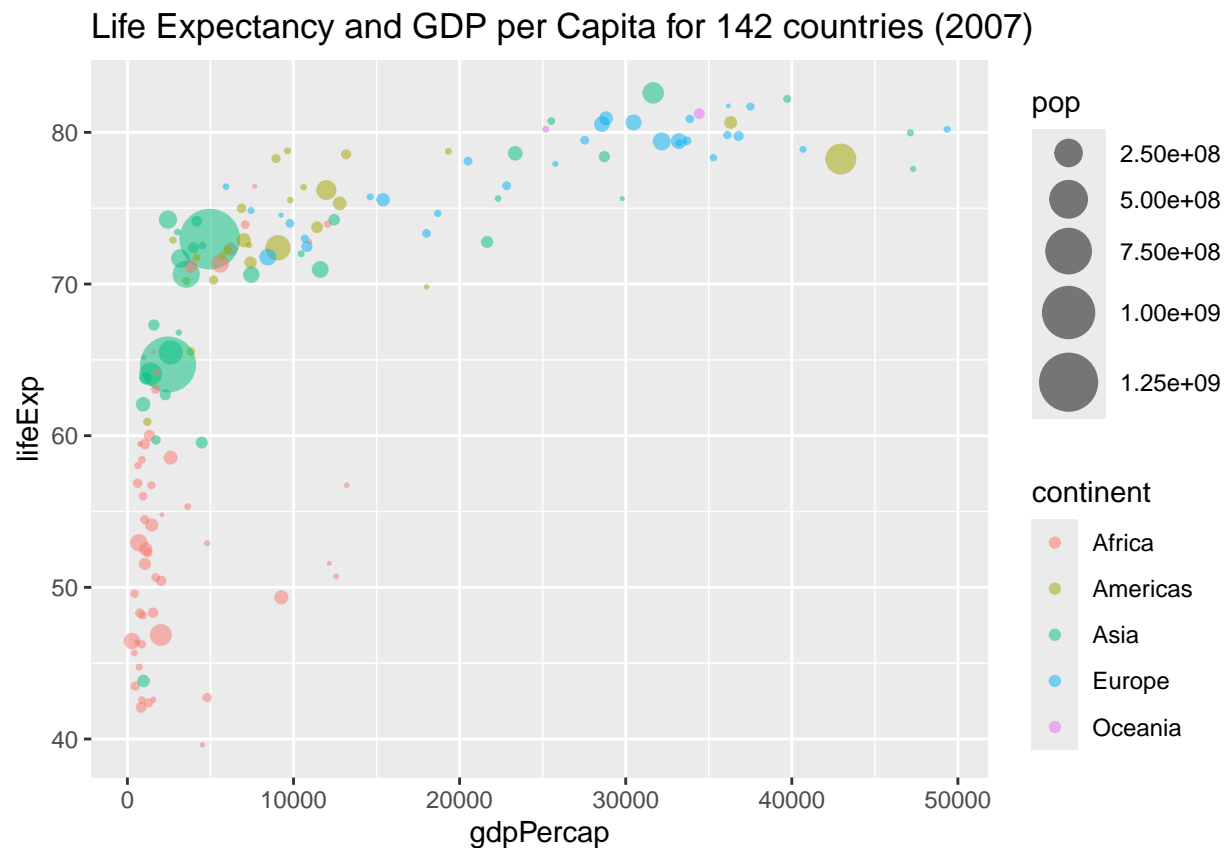
7. Going Further

```
# load in dataset from url
url <- "https://raw.githubusercontent.com/jennybc/gapminder/master/inst/extdata/gapminder.tsv"
gapminder <- read.delim(url)

# load in dplyr package
library(dplyr)

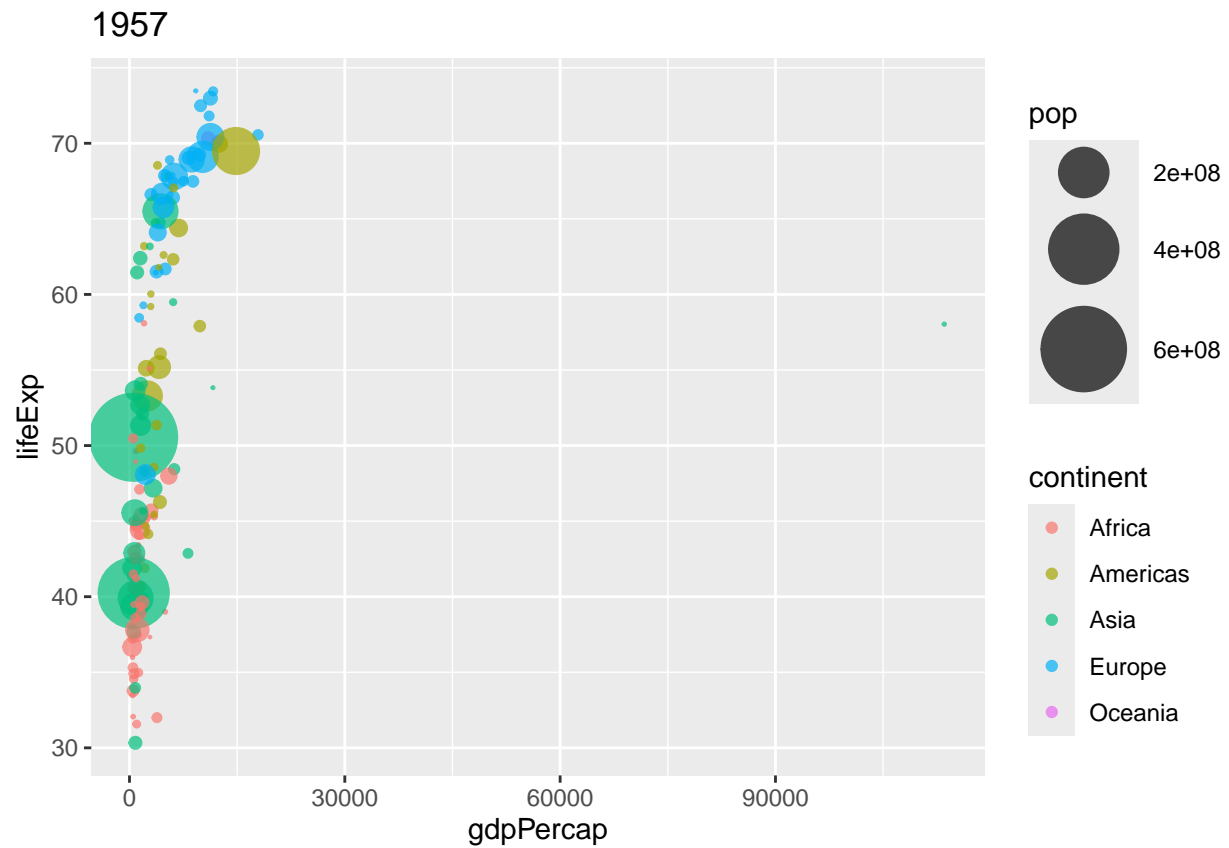
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

gapminder_2007 <- gapminder %>% filter(year==2007)
ggplot(gapminder_2007) + aes(x=gdpPercap, y=lifeExp, color=continent,
                           size=pop) + geom_point(alpha=0.5) +
  labs(title="Life Expectancy and GDP per Capita for 142 countries (2007)") +
  scale_size_area(max_size=10)
```

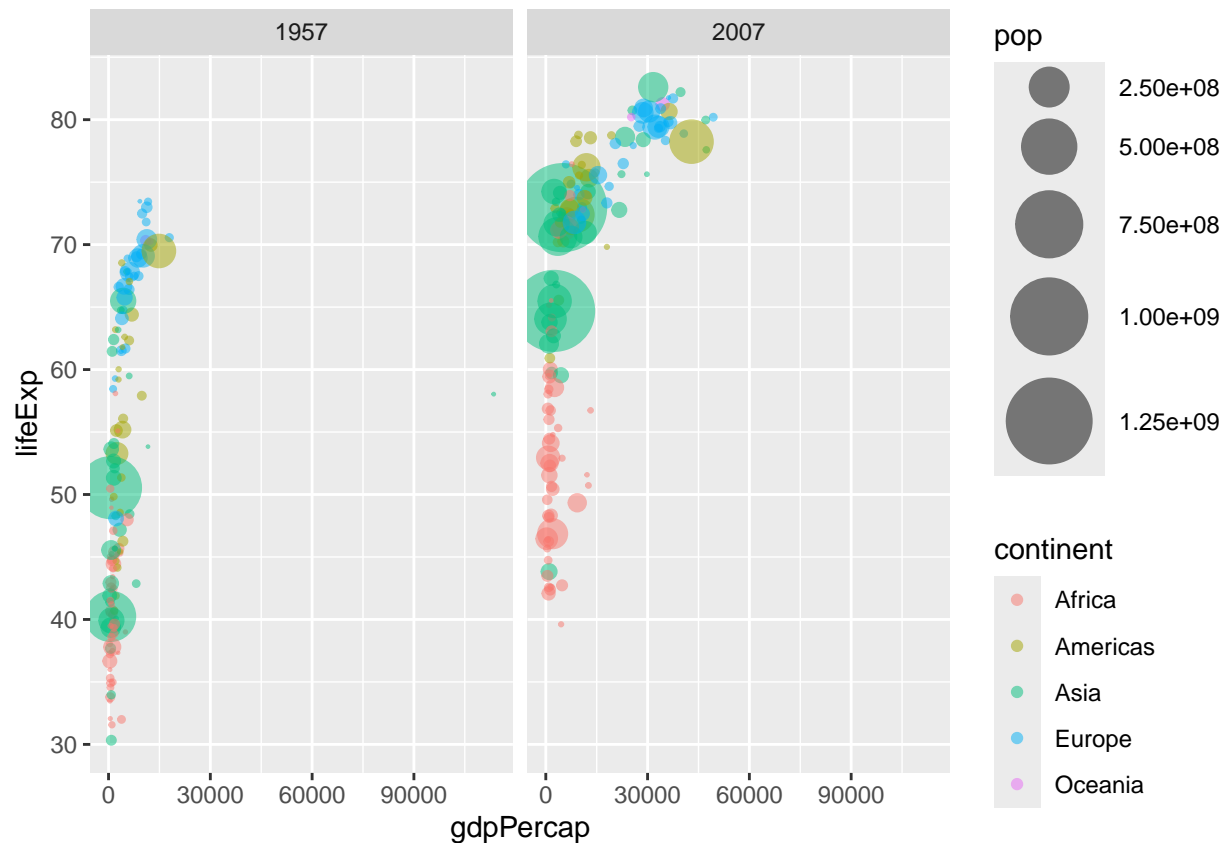


```
gapminder_1957 <- gapminder %>% filter(year==1957)
ggplot(gapminder_1957) + aes(x=gdpPercap, y=lifeExp, color=continent, size=pop)+
```

```
geom_point(alpha=0.7) + scale_size_area(max_size = 15) + labs(title="1957")
```



```
gapminder_combo <- gapminder %>% filter(year == 1957 | year == 2007)
ggplot(gapminder_combo) + aes(x=gdpPercap, y=lifeExp, color=continent, size=pop) +
  geom_point(alpha=0.5) + scale_size_area(max_size = 15) + facet_wrap(~year)
```

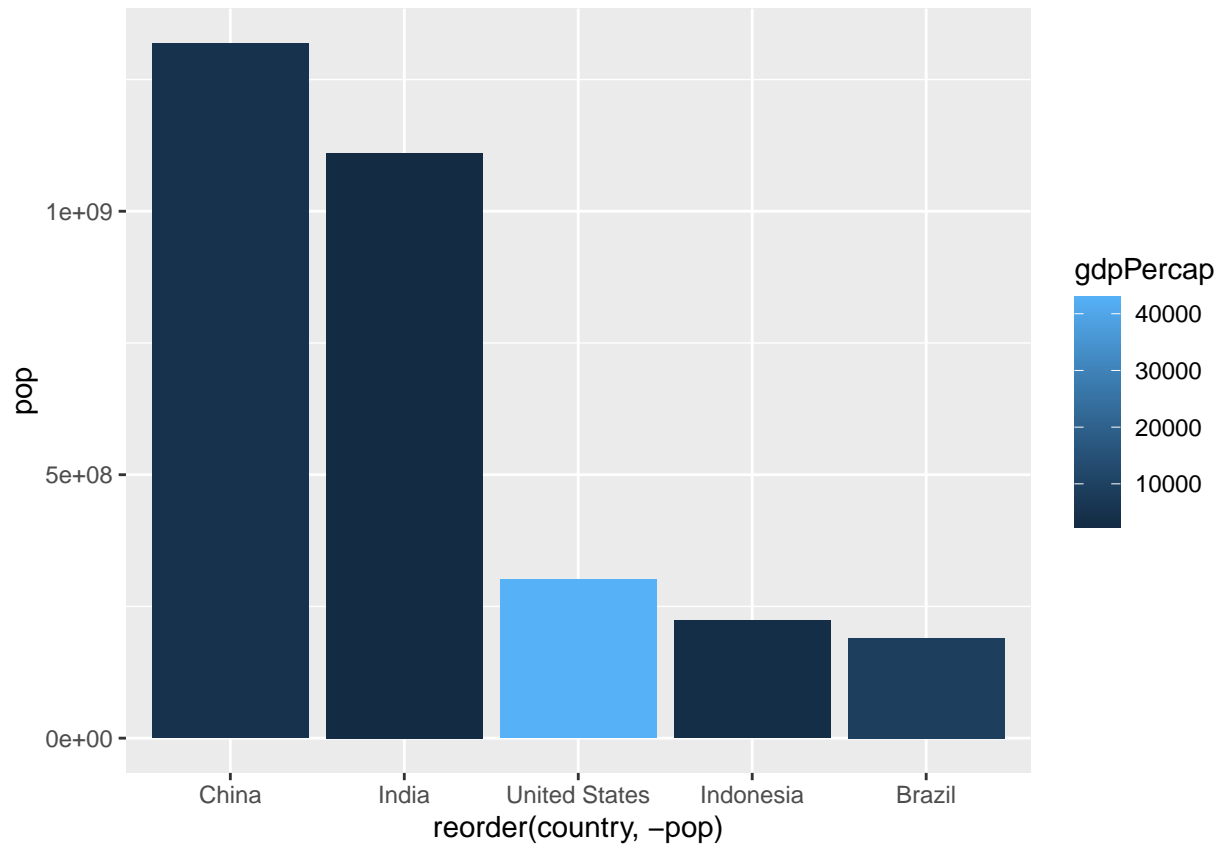


```
gapminder_top5 <- gapminder %>% filter(year==2007) %>% arrange(desc(pop)) %>%
  top_n(5, pop)
```

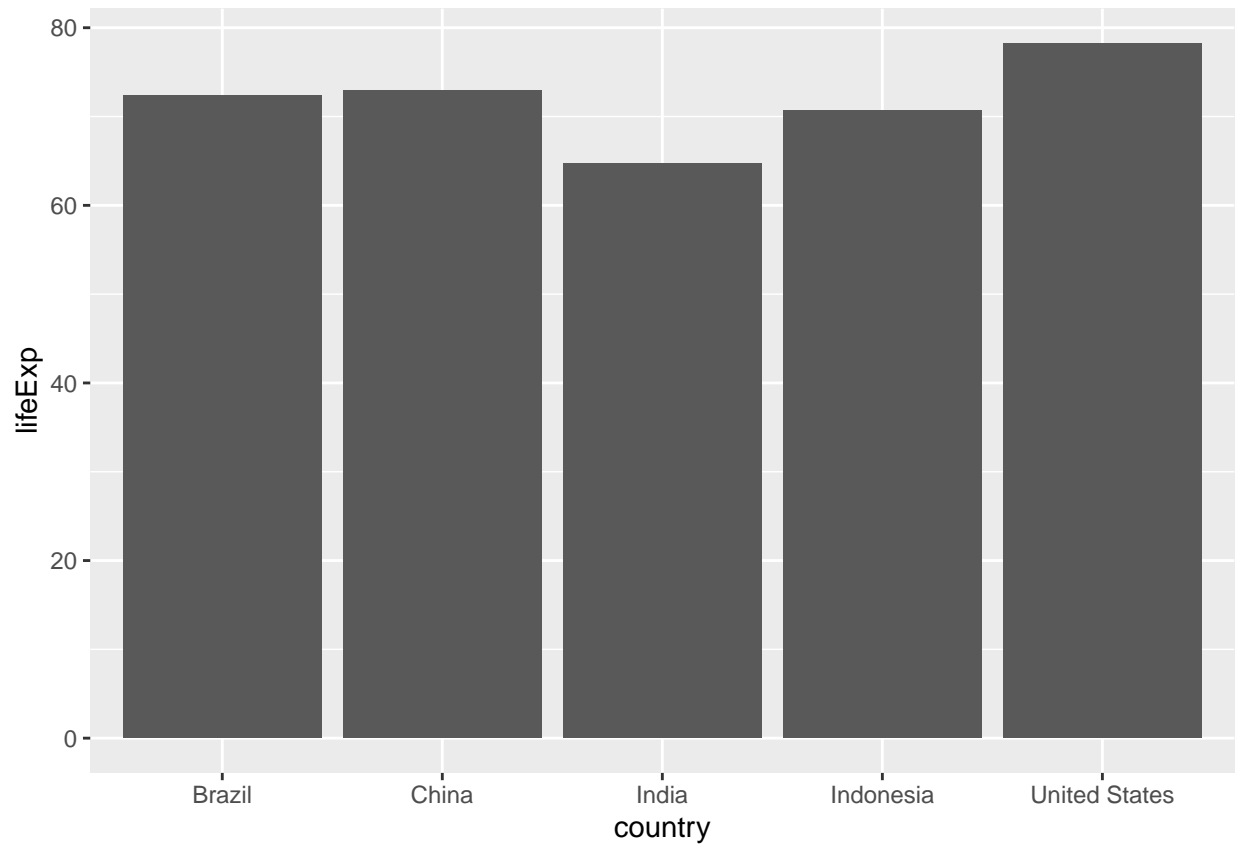
```
gapminder_top5
```

```
##      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=reorder(country,-pop), y=pop,
  fill=gdpPercap))
```



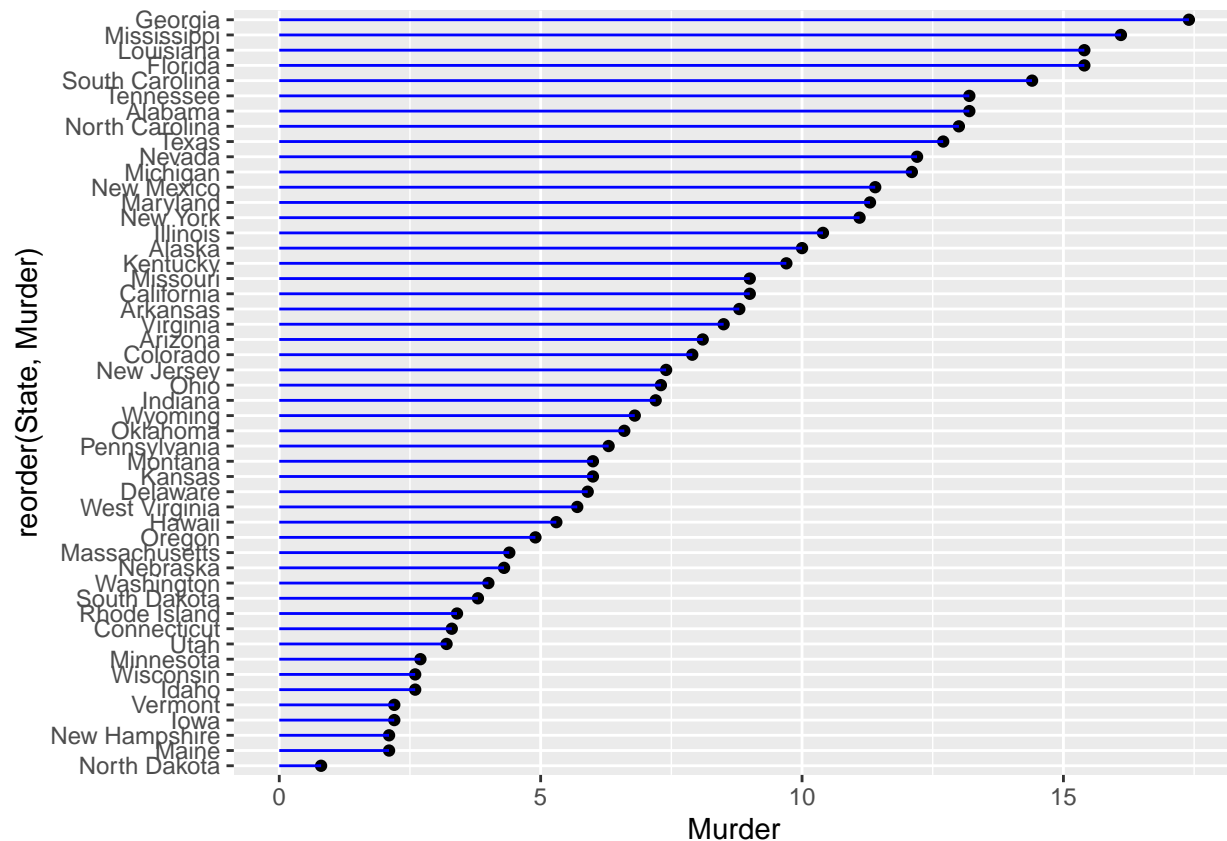
```
ggplot(gapminder_top5) + geom_col(aes(x=country, y=lifeExp))
```



```
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()
```



```
sessionInfo()
```

```
## R version 4.4.0 (2024-04-24)
## Platform: x86_64-apple-darwin20
## Running under: macOS Monterey 12.7.1
##
## Matrix products: default
## BLAS: /Library/Frameworks/R.framework/Versions/4.4-x86_64/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.4-x86_64/Resources/lib/libRlapack.dylib; LAPACK
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## time zone: America/Los_Angeles
## tzcode source: internal
##
## attached base packages:
## [1] stats graphics grDevices utils datasets methods base
##
## other attached packages:
## [1] dplyr_1.1.4 ggplot2_3.5.1
##
## loaded via a namespace (and not attached):
## [1] vctrs_0.6.5 nlme_3.1-166 cli_3.6.3 knitr_1.49
## [5] rlang_1.1.4 xfun_0.50 generics_0.1.3 labeling_0.4.3
## [9] glue_1.8.0 colorspace_2.1-1 htmltools_0.5.8.1 scales_1.3.0
```


## [13]	rmarkdown_2.29	grid_4.4.0	evaluate_1.0.3	munsell_0.5.1
## [17]	tibble_3.2.1	fastmap_1.2.0	yaml_2.3.10	lifecycle_1.0.4
## [21]	compiler_4.4.0	pkgconfig_2.0.3	mgcv_1.9-1	rstudioapi_0.17.1
## [25]	farver_2.1.2	lattice_0.22-6	digest_0.6.37	R6_2.5.1
## [29]	tidyselect_1.2.1	splines_4.4.0	pillar_1.10.1	magrittr_2.0.3
## [33]	Matrix_1.7-1	withr_3.0.2	tools_4.4.0	gtable_0.3.6