class 5: Data visualization with ggplot

Yang Liu (A16629623)

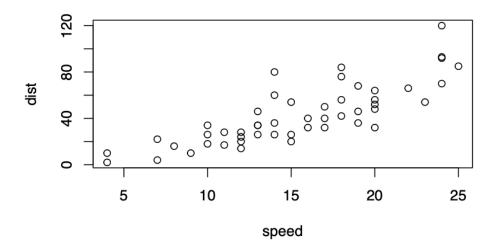
Table of contents

A more complicated scatter plot															4
Exploring their gapmider dataset															17

Today we will have our first play with the **ggplot2** package - one of the most popular graphics packages on the planet.

There are many plotting systems in R. These include so called "base" plotting/graphics.

plot(cars)



Base plot is generally rather short code and somewhat dull plots - but it is always there for you and is fast for big datasets.

If i want to use ggplot2 it takes some more work

```
# ggplot(cars)
```

I need to install the package first. To do this I can use the function 'install.packages("ggplot2")' Every time I want to use a package I need to load it up with a 'libarary()' call.

```
library(ggplot2)
```

Now finally I can use ggplot

```
ggplot(cars)
```

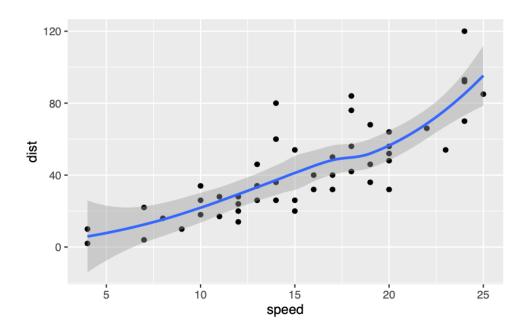
Every ggplot has at least 3 things:

- data (the data.frame witht the data you want to plot)
- aes (the asthetic mapping of the data to the plot)
- geom (how do you want the plot to look, points, lines, columns)

head(cars)

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

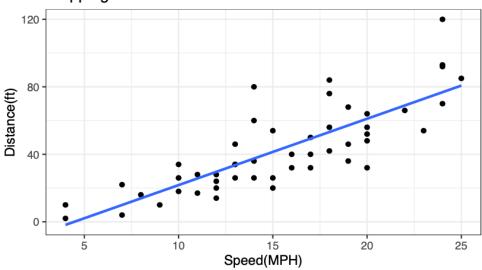
 $\ensuremath{\text{`geom_smooth()`}}\ \ensuremath{\text{using method}}\ = \ensuremath{\text{'loess'}}\ \ensuremath{\text{and formula}}\ = \ensuremath{\text{'y}}\ \sim \ensuremath{\text{x'}}$



I want a linear model and no standard error bounds shown on my plot.

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

Stopping Distnace of Old Cars



From the 'cars' dataset

A more complicated scatter plot

Here we make a plot of gene expression data:

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

```
Gene Condition1 Condition2
                                                                     A4GNT -3.6808610 -3.4401355 unchanging
2
                                                                             AAAS 4.5479580 4.3864126 unchanging
3
                                                                     AASDH 3.7190695 3.4787276 unchanging
                                                                               AATF 5.0784720 5.0151916 unchanging
                                                                             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$State)
                                                           down unchanging
                                                                                                                                                                                                                                                                                                      up
                                                                                                                                                                      4997
                                                                               72
                                                                                                                                                                                                                                                                                             127
                         genes$State == "up"
                               [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE TRUE FALSE
                      [13] FALSE F
                      [25] FALSE F
                      [37] FALSE FALSE
                      [49] FALSE F
                      [61] FALSE F
                      [73] FALSE F
                      [85] FALSE FALSE
                      [97] FALSE FALSE
```

```
[109] FALSE FALSE
 [121] FALSE FALSE
[133] FALSE 
[145] FALSE 
 [157] FALSE FALSE
[169] FALSE 
[181] FALSE 
[193] FALSE 
[205] FALSE 
[217] FALSE 
[229] FALSE 
[241] FALSE 
[253] FALSE 
[265] FALSE 
[277] FALSE 
 [289] FALSE 
[301] FALSE 
 [313] FALSE FALSE
[325] FALSE 
 [337] FALSE FALSE
 [349] FALSE FALSE
 [361] FALSE FALSE
 [373] FALSE FALSE
 [385] FALSE FALSE
 [397] FALSE FALSE
 [409] FALSE 
[421] TRUE FALSE F
[433] FALSE 
[445] FALSE 
[457] FALSE 
[469] FALSE 
 [481] FALSE 
 [493] FALSE FALSE FALSE TRUE FALSE FALSE FALSE TRUE FALSE
 [505] FALSE FALSE
 [517] FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE
 [529] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
 [541] FALSE FALSE
 [553] FALSE FALSE
 [565] FALSE FALSE
 [577] FALSE FALSE
 [589] FALSE FALSE
 [601] FALSE FALSE
 [613] FALSE FALSE
```

```
[625] FALSE FALSE
           [637] FALSE FALSE
           [649] FALSE FALSE
           [661] FALSE FALSE
           [673] FALSE FALSE
           [685] FALSE FALSE
           [697] FALSE FALSE
           [709] FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE FALSE
           [721] FALSE FALSE
           [733] TRUE FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
           [745] FALSE FALSE
           [757] FALSE 
           [769] FALSE FALSE
           [781] FALSE FALSE
           [793] FALSE FALSE
           [805] FALSE 
           [817] FALSE FALSE
           [829] FALSE FALSE
           [841] FALSE FALSE
           [853] FALSE FALSE
           [865] FALSE FALSE
           [877] FALSE FALSE
           [889] FALSE FALSE
           [901] FALSE 
          [913] FALSE 
           [925] FALSE FALSE
          [937] FALSE 
          [949] TRUE FALSE F
          [961] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE
          [973] FALSE 
          [985] FALSE 
          [997] FALSE 
[1009] FALSE FALSE
[1021] FALSE FALSE
[1033] FALSE FALSE
[1045] FALSE FALSE
[1057] TRUE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE TRUE TRUE
[1069] FALSE FALSE
[1081] FALSE FALSE
[1093] FALSE FALSE
[1105] FALSE FALSE
[1117] FALSE FALSE
[1129] FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE FALSE
```

```
[1141] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE
[1153] FALSE FALSE
[1165] FALSE FALSE FALSE TRUE FALSE 
[1177] FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE TRUE FALSE FALSE
[1189] FALSE FALSE
[1201] FALSE FALSE
[1213] FALSE FALSE
[1225] FALSE FALSE
[1237] FALSE FALSE
[1249] FALSE FALSE
[1261] FALSE FALSE
[1273] FALSE TRUE FALSE 
[1285] FALSE FALSE
[1297] FALSE FALSE
[1309] FALSE FALSE
[1321] FALSE FALSE
[1333] TRUE FALSE 
[1345] FALSE FALSE
[1357] FALSE TRUE FALSE 
[1369] FALSE FALSE
[1381] FALSE FALSE
[1393] FALSE FALSE
[1405] FALSE FALSE
[1417] FALSE FALSE
[1429] FALSE FALSE
[1441] FALSE FALSE
[1453] FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE TRUE FALSE FALSE
[1465] FALSE FALSE
[1477] FALSE FALSE
[1489] FALSE FALSE
[1501] FALSE FALSE
[1513] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE
[1525] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE
[1537] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE
[1549] FALSE FALSE
[1561] FALSE FALSE
[1573] FALSE FALSE
[1585] FALSE FALSE
[1597] FALSE FALSE
[1609] FALSE FALSE
[1621] FALSE FALSE
[1633] FALSE FALSE
[1645] FALSE FALSE
```

[1657] FALSE FALSE TRUE FALSE TRUE FALSE F [1669] FALSE [1681] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE [1693] FALSE [1705] TRUE FALSE [1717] FALSE [1729] FALSE FALSE TRUE FALSE [1741] FALSE [1753] FALSE TRUE FALSE [1765] FALSE [1777] FALSE [1789] FALSE [1801] FALSE [1813] FALSE FALSE FALSE FALSE FALSE TRUE FALSE TRUE TRUE FALSE FALSE [1825] FALSE [1837] FALSE [1849] FALSE [1861] FALSE [1873] FALSE FALSE TRUE FALSE [1885] FALSE [1897] FALSE [1909] FALSE [1921] FALSE TRUE FALSE FALSE TRUE TRUE FALSE FALSE FALSE FALSE FALSE FALSE [1933] FALSE FALSE FALSE TRUE FALSE [1945] FALSE [1957] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE [1969] FALSE [1981] TRUE FALSE [1993] FALSE FALSE FALSE TRUE TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [2005] FALSE [2017] FALSE [2029] FALSE [2041] FALSE [2053] FALSE [2065] FALSE [2077] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE [2089] FALSE [2101] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE [2113] FALSE [2125] FALSE [2137] FALSE [2149] FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE FALSE [2161] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE TRUE

```
[2173] FALSE FALSE
[2185] FALSE FALSE
[2197] FALSE FALSE
[2209] FALSE FALSE
[2221] FALSE FALSE
[2233] FALSE FALSE
[2245] FALSE FALSE FALSE FALSE FALSE TRUE FALSE TRUE FALSE FALSE FALSE
[2257] FALSE FALSE
[2269] FALSE FALSE
[2281] FALSE FALSE
[2293] FALSE FALSE
[2305] FALSE FALSE
[2317] FALSE FALSE
[2329] FALSE FALSE TRUE TRUE FALSE F
[2341] FALSE FALSE
[2353] FALSE FALSE
[2365] FALSE FALSE
[2377] FALSE FALSE
[2389] FALSE FALSE TRUE FALSE 
[2401] FALSE FALSE
[2413] FALSE FALSE
[2425] FALSE FALSE
[2437] FALSE FALSE
[2449] FALSE FALSE
[2461] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE TRUE
[2473] FALSE FALSE TRUE FALSE 
[2485] TRUE FALSE 
[2497] FALSE FALSE
[2509] FALSE FALSE
[2521] FALSE FALSE
[2533] FALSE FALSE
[2545] FALSE FALSE
[2557] FALSE FALSE
[2569] FALSE FALSE
[2581] FALSE FALSE
[2593] FALSE FALSE
[2605] TRUE FALSE 
[2617] FALSE TRUE FALSE 
[2629] FALSE FALSE
[2641] FALSE FALSE
[2653] FALSE FALSE
[2665] FALSE FALSE
[2677] FALSE FALSE
```

```
[2689] FALSE FALSE
[2701] FALSE FALSE
[2713] FALSE FALSE
[2725] FALSE FALSE
[2737] FALSE FALSE
[2749] FALSE FALSE
[2761] FALSE FALSE
[2773] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE
[2785] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE
[2797] FALSE FALSE
[2809] TRUE FALSE 
[2821] FALSE FALSE
[2833] FALSE FALSE
[2845] FALSE FALSE
[2857] FALSE FALSE TRUE FALSE 
[2869] FALSE FALSE
[2881] FALSE FALSE
[2893] FALSE FALSE
[2905] FALSE FALSE
[2917] FALSE FALSE
[2929] FALSE FALSE
[2941] FALSE FALSE
[2953] FALSE FALSE
[2965] FALSE FALSE
[2977] FALSE FALSE
[2989] FALSE FALSE
[3001] FALSE FALSE
[3013] FALSE FALSE
[3025] FALSE FALSE
[3037] FALSE FALSE TRUE FALSE 
[3049] FALSE FALSE
[3061] FALSE FALSE
[3073] FALSE FALSE
[3085] FALSE FALSE
[3097] FALSE FALSE
[3109] TRUE FALSE 
[3121] FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[3133] FALSE FALSE
[3145] FALSE FALSE
[3157] FALSE FALSE
[3169] FALSE FALSE
[3181] FALSE FALSE
[3193] FALSE FALSE
```

```
[3205] FALSE FALSE
[3217] FALSE FALSE
[3229] FALSE FALSE
[3241] FALSE FALSE
[3253] FALSE FALSE
[3265] FALSE FALSE
[3277] FALSE FALSE
[3289] FALSE FALSE
[3301] FALSE FALSE
[3313] FALSE FALSE
[3325] FALSE FALSE
[3337] FALSE FALSE
[3349] FALSE FALSE
[3361] FALSE FALSE
[3373] FALSE FALSE
[3385] FALSE FALSE
[3397] FALSE FALSE
[3409] FALSE FALSE
[3421] FALSE FALSE
[3433] FALSE FALSE
[3445] FALSE FALSE
[3457] FALSE FALSE
[3469] FALSE FALSE
[3481] FALSE FALSE
[3493] FALSE FALSE
[3505] FALSE FALSE
[3517] FALSE FALSE
[3529] FALSE FALSE
[3541] FALSE FALSE
[3553] FALSE FALSE
[3565] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE
[3577] FALSE FALSE
[3589] FALSE FALSE
[3601] FALSE FALSE
[3613] FALSE FALSE
[3625] FALSE FALSE
[3637] FALSE FALSE
[3649] FALSE FALSE
[3661] FALSE FALSE
[3673] FALSE FALSE
[3685] FALSE FALSE
[3697] FALSE FALSE
[3709] FALSE FALSE
```

```
[3721] FALSE FALSE
[3733] FALSE FALSE
[3745] FALSE FALSE
[3757] FALSE FALSE
[3769] FALSE FALSE
[3781] FALSE FALSE
[3793] FALSE FALSE
[3805] FALSE FALSE
[3817] FALSE FALSE
[3829] FALSE FALSE
[3841] FALSE FALSE
[3853] FALSE FALSE
[3865] FALSE FALSE
[3877] FALSE FALSE
[3889] FALSE FALSE
[3901] FALSE FALSE
[3913] FALSE FALSE
[3925] FALSE FALSE
[3937] FALSE FALSE
[3949] FALSE FALSE
[3961] FALSE FALSE
[3973] FALSE FALSE
[3985] FALSE FALSE
[3997] FALSE FALSE
[4009] FALSE FALSE
[4021] FALSE FALSE
[4033] FALSE TRUE FALSE 
[4045] FALSE FALSE
[4057] FALSE FALSE
[4069] FALSE FALSE
[4081] FALSE FALSE
[4093] FALSE FALSE
[4105] FALSE FALSE
[4117] FALSE FALSE
[4129] FALSE FALSE
[4141] FALSE FALSE
[4153] FALSE FALSE
[4165] FALSE FALSE
[4177] FALSE FALSE
[4189] FALSE FALSE
[4201] FALSE FALSE
[4213] FALSE FALSE
[4225] FALSE FALSE
```

```
[4237] FALSE FALSE
[4249] FALSE FALSE
[4261] FALSE FALSE
[4273] FALSE FALSE
[4285] FALSE FALSE
[4297] FALSE FALSE
[4309] FALSE FALSE
[4321] FALSE FALSE
[4333] FALSE FALSE FALSE FALSE FALSE TRUE TRUE FALSE FALSE TRUE
[4345] FALSE FALSE
[4357] FALSE FALSE
[4369] FALSE FALSE
[4381] FALSE FALSE
[4393] FALSE FALSE
[4405] FALSE FALSE
[4417] FALSE FALSE
[4429] FALSE TRUE FALSE 
[4441] FALSE FALSE
[4453] FALSE FALSE
[4465] FALSE FALSE
[4477] FALSE FALSE
[4489] FALSE TRUE FALSE 
[4501] FALSE FALSE
[4513] FALSE FALSE
[4525] FALSE FALSE
[4537] FALSE FALSE
[4549] FALSE FALSE
[4561] FALSE FALSE
[4573] FALSE FALSE
[4585] FALSE FALSE
[4597] FALSE FALSE
[4609] FALSE FALSE
[4621] FALSE FALSE FALSE TRUE TRUE FALSE F
[4633] FALSE FALSE
[4645] FALSE FALSE
[4657] FALSE FALSE
[4669] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[4681] FALSE FALSE
[4693] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[4705] FALSE FALSE
[4717] FALSE FALSE TRUE FALSE 
[4729] FALSE FALSE
[4741] FALSE TRUE FALSE FALSE
```

```
[4753] FALSE FALSE
[4765] FALSE FALSE
[4777] FALSE FALSE
[4789] FALSE FALSE
[4801] TRUE FALSE 
[4813] FALSE FALSE
[4825] FALSE FALSE
[4837] FALSE FALSE
[4849] FALSE FALSE
[4861] FALSE FALSE
[4873] FALSE FALSE
[4885] FALSE FALSE
[4897] FALSE FALSE
[4909] FALSE FALSE
[4921] FALSE FALSE
[4933] FALSE FALSE
[4945] FALSE FALSE
[4957] FALSE FALSE FALSE TRUE TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[4969] FALSE FALSE
[4981] FALSE FALSE
[4993] FALSE FALSE
[5005] FALSE FALSE
[5017] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE
[5029] FALSE FALSE
[5041] FALSE FALSE
[5053] FALSE FALSE
[5065] FALSE FALSE
[5077] FALSE FALSE
[5089] FALSE FALSE FALSE TRUE FALSE 
[5101] FALSE FALSE
[5113] FALSE FALSE
[5125] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE
[5137] FALSE FALSE
[5149] FALSE FALSE
[5161] FALSE FALSE
[5173] FALSE FALSE
[5185] FALSE FALSE
```

```
round( sum(genes$State == "up") / nrow(genes) * 100, 2)
```

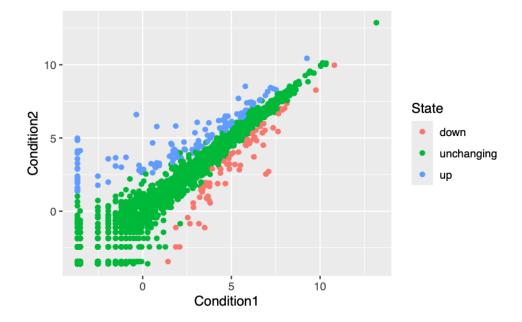
[1] 2.44

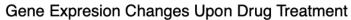
Another way to calculate percentage of total genes is up-regulated in this dataset?

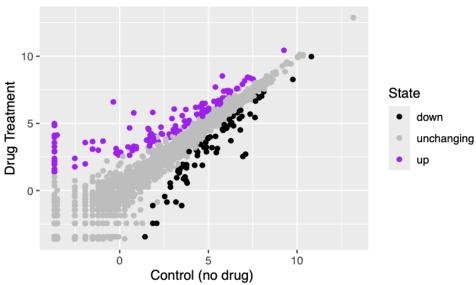
```
n.gene <- nrow(genes)
n.up <- sum(genes$State == "up")
up.present <- n.up/n.gene * 100
round (up.present, 2)</pre>
```

[1] 2.44

```
bp <- ggplot(genes) +
    aes(x=Condition1, y=Condition2, col=State) +
    geom_point()
bp</pre>
```







Exploring their gapmider dataset

Here we will load up the gaomider dataset to get practive with different aes mappings.

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

How many rows and columns can be found in this dataframe?

```
row <- nrow(gapminder)
column <- ncol(gapminder)
row</pre>
```

[1] 1704

column

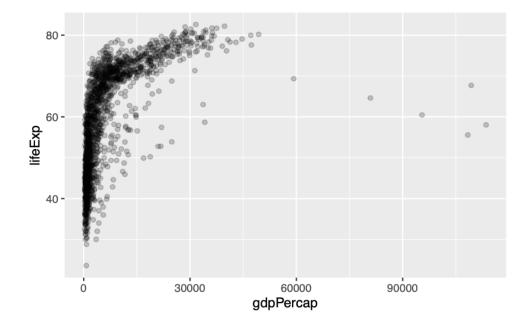
[1] 6

head(gapminder)

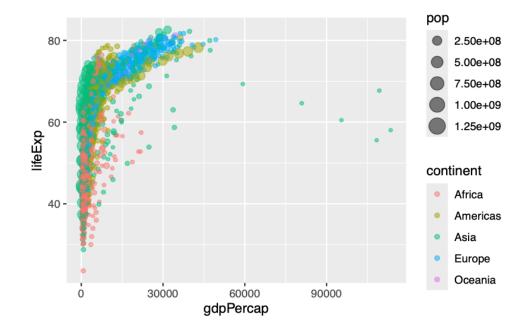
[1] 142

```
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
                  Asia 1972
5 Afghanistan
                            36.088 13079460 739.9811
6 Afghanistan
                  Asia 1977 38.438 14880372 786.1134
Q. How many countries are there in this data set?
  table(gapminder$year)
1952 1957 1962 1967 1972 1977 1982 1987 1992 1997 2002 2007
 table(gapminder$continent)
  Africa Americas
                     Asia
                           Europe
                                  Oceania
                              360
     624
             300
                     396
                                        24
I could use the 'unique()' function...
  unique(gapminder$continent)
[1] "Asia"
                        "Africa"
              "Europe"
                                   "Americas" "Oceania"
  length (unique(gapminder$continent))
[1] 5
  length (unique(gapminder$country))
```

```
ggplot(gapminder) +
  aes(x=gdpPercap, y=lifeExp) +
  geom_point(alpha=0.2)
```



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



library(dplyr)

Attaching package: 'dplyr'

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

filter, lag

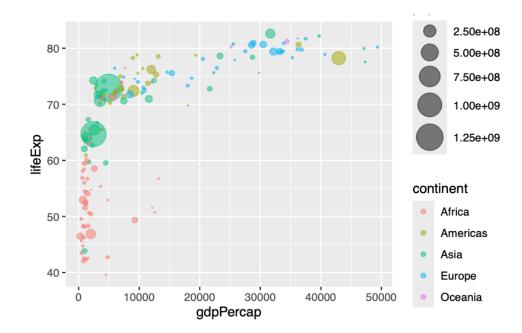
The following objects are masked from 'package:base':

intersect, setdiff, setequal, union

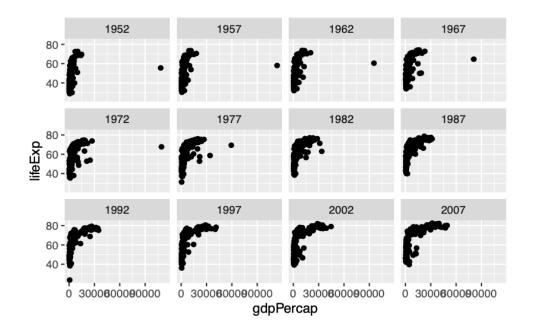
```
gapminder_2007 <- gapminder %>% filter(year==2007)
head(gapminder_2007)
```

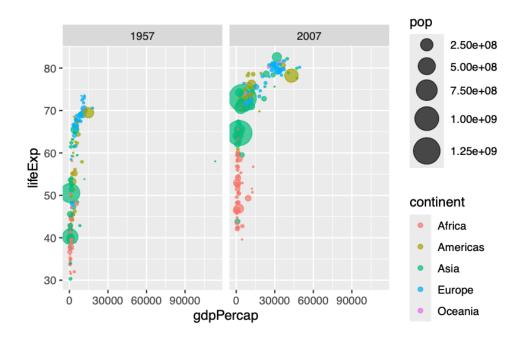
country continent year lifeExp pop gdpPercap
1 Afghanistan Asia 2007 43.828 31889923 974.5803
2 Albania Europe 2007 76.423 3600523 5937.0295

```
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) +
  aes(x = gdpPercap, y = lifeExp) +
  geom_point() +
  facet_wrap(~year)
```





sessionInfo()

R version 4.3.3 (2024-02-29)

Platform: x86_64-apple-darwin20 (64-bit)

Running under: macOS Sonoma 14.0

Matrix products: default

BLAS: /Library/Frameworks/R.framework/Versions/4.3-x86_64/Resources/lib/libRblas.0.dylib LAPACK: /Library/Frameworks/R.framework/Versions/4.3-x86_64/Resources/lib/libRlapack.dylib;

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_3.5.0

loaded via a namespace (and not attached):

[1]	vctrs_0.6.4	nlme_3.1-164	cli_3.6.1	knitr_1.45
[5]	rlang_1.1.2	xfun_0.41	generics_0.1.3	jsonlite_1.8.7
[9]	labeling_0.4.3	glue_1.6.2	colorspace_2.1-0	htmltools_0.5.7
[13]	scales_1.3.0	fansi_1.0.5	rmarkdown_2.25	grid_4.3.3
[17]	evaluate_0.23	munsell_0.5.0	tibble_3.2.1	fastmap_1.1.1
[21]	yam1_2.3.7	lifecycle_1.0.4	compiler_4.3.3	pkgconfig_2.0.3
[25]	mgcv_1.9-1	rstudioapi_0.15.0	farver_2.1.1	lattice_0.22-5
[29]	digest_0.6.33	R6_2.5.1	tidyselect_1.2.0	utf8_1.2.4
[33]	splines_4.3.3	pillar_1.9.0	magrittr_2.0.3	Matrix_1.6-5
[37]	withr_2.5.2	tools_4.3.3	gtable_0.3.4	