

Class05: Data Viz with GGPLOT

Qingyun Zheng (A16338254)

Today we are gonna play with plotting and graphics in R.

There are lots of ways to make cool figures in R.

The most popular package is `ggplot2`, which is part of the `tidyverse` collection of packages.
—copilot

There is “base” R graphics (`plot()`, `hist()`, `boxplot()`, etc.).

There is also add-on packages like `ggplot`

To insert any code, click three ’back-tick’s to mark the start of the code, or just click the green button above, that is to the left of the button Run.

```
head(cars, 3)
```

```
speed dist
1      4    2
2      4   10
3      7    4
```

Let’s plot this with “base” R: *option + command + I* would directly insert a code block.

```
plot(cars)
```

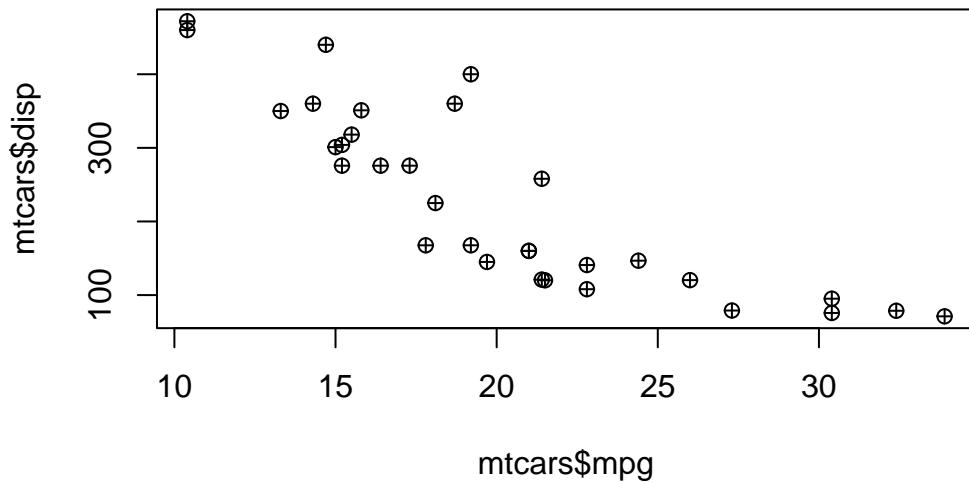


```
head(mtcars)
```

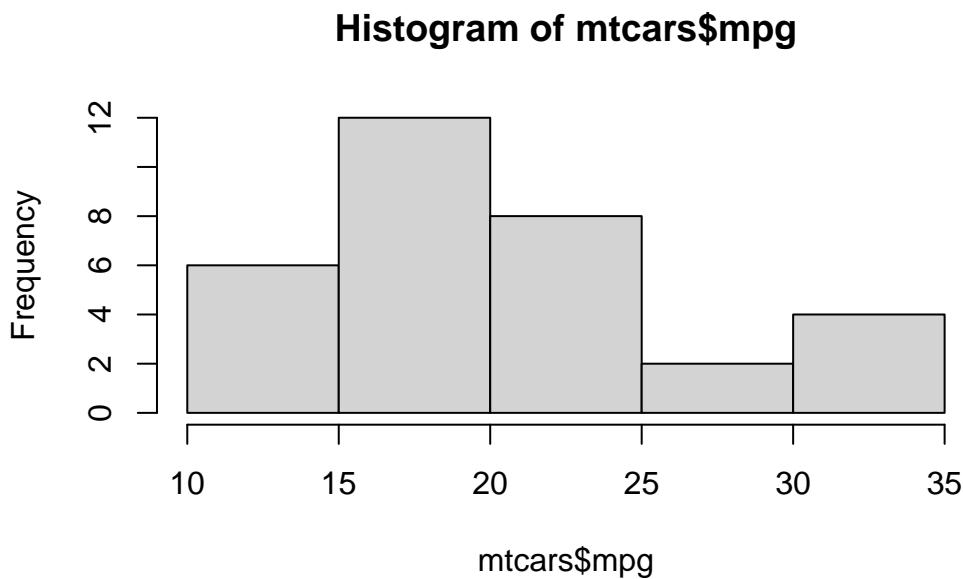
	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

Let's plot mpg vs disp: pch is plotting character, each representing a symbol.

```
plot(mtcars$mpg, mtcars$disp, pch=10)
```



```
hist(mtcars$mpg)
```



GGPLOT (the pkg itself is called `ggplot2`)

The main function in the `ggplot2` package is `ggplot()`. First, I need to install the `ggplot2` package. To do this, I can install any package with the function `install.packages("ggplot2")`.

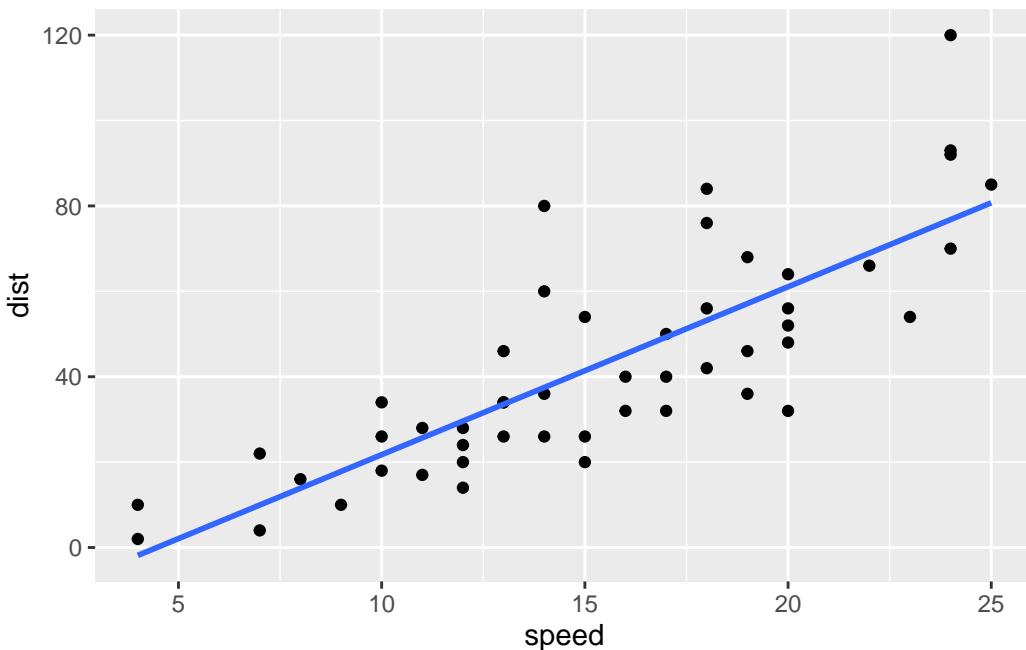
Then I need to load the package with `library(ggplot2)`.

N.B I never want to run `install.packages()` in my quarto source document!! If I do that, every time I render the document, it will try to install the package again. I only need to install a package once on my computer.

`library()` loads the package into my current R session.

```
library(ggplot2)
ggplot(cars)+aes(x=speed, y = dist) + geom_point() + geom_smooth(method = "lm", se = FALSE)

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



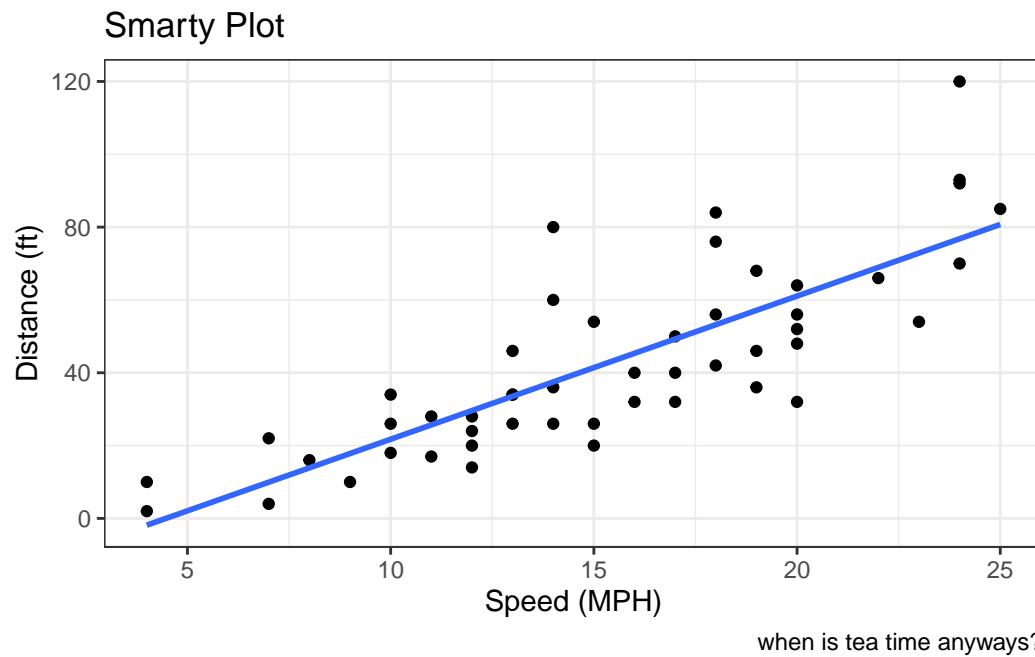
Adding more layers

let's add a line and a title, subtitle an caption as well as custom axis labels

```

library(ggplot2)
ggplot(cars)+aes(x=speed, y = dist) + geom_point() + geom_smooth(method = "lm", se = FALSE) +
`geom_smooth()` ` using formula = 'y ~ x'

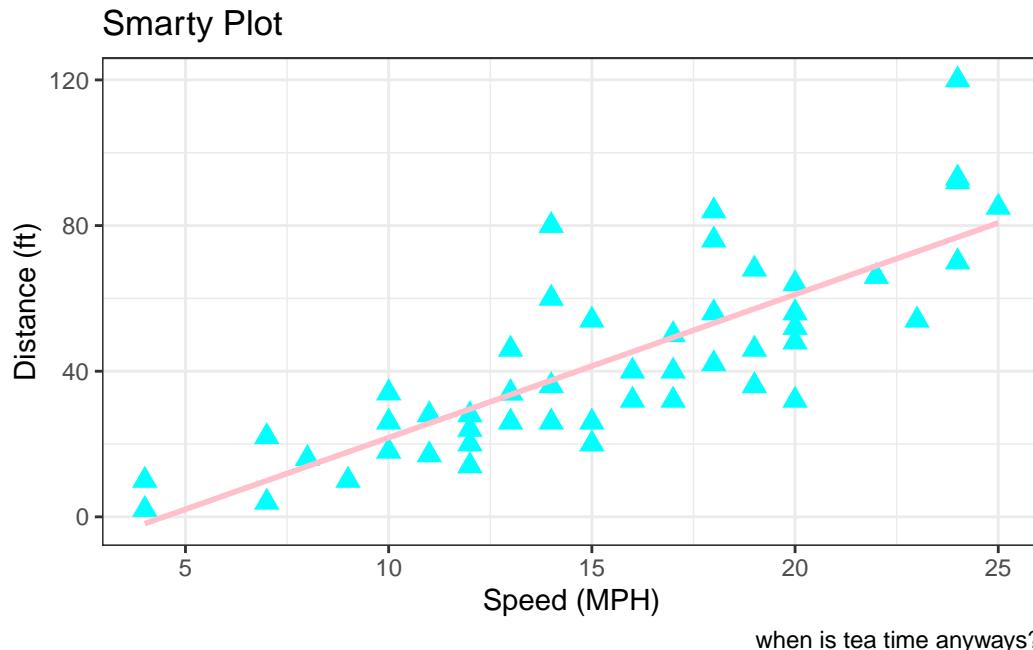
```



```

ggplot(cars)+aes(x=speed, y = dist) + geom_point(color = "cyan", size = 3, shape = 17)+ geom_
`geom_smooth()` ` using formula = 'y ~ x'

```



when is tea time anyways?

Read a file from the web

```
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 wee dataset?

There are 5196 genes in this wee dataset

Q2. How many “up” regulated genes are there?

```
sum(genes$State == "up")
```

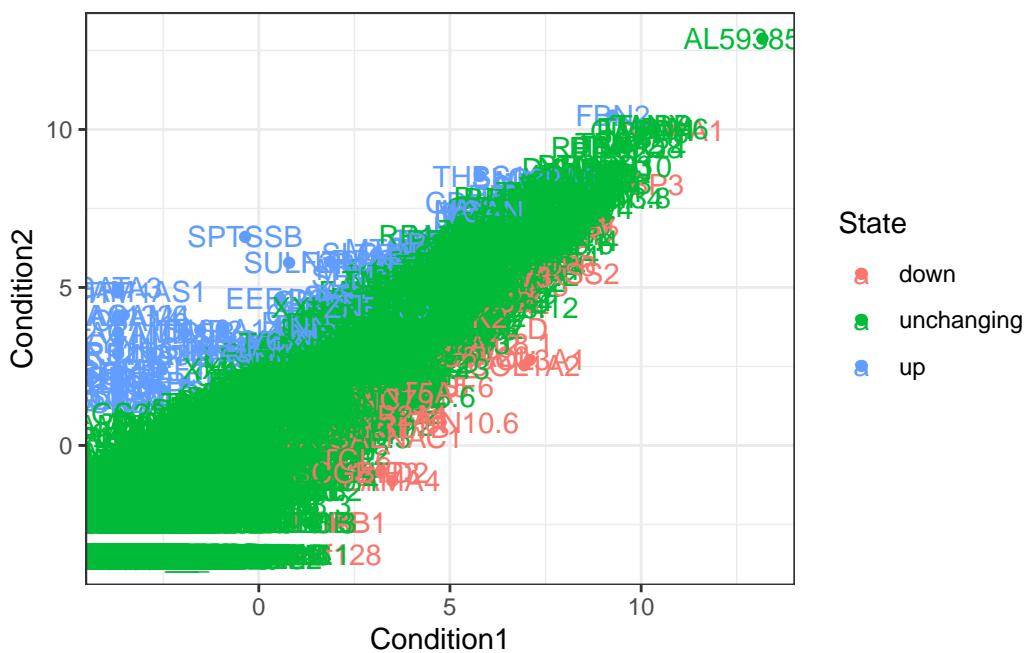
[1] 127

There are 127 upregulated genes.

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

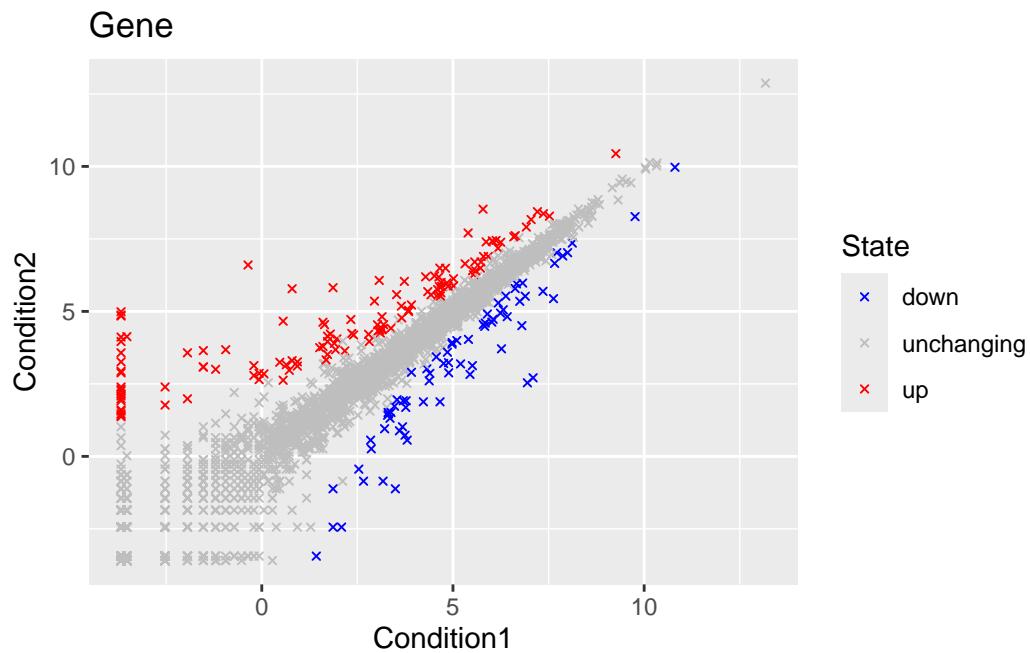
down	unchanging	up
72	4997	127

```
ggplot(genes) +  
  aes(x=Condition1,  
       y= Condition2,  
       col = State,  
       label = Gene) +  
  geom_point() +  
  geom_text() +  
  theme_bw()
```



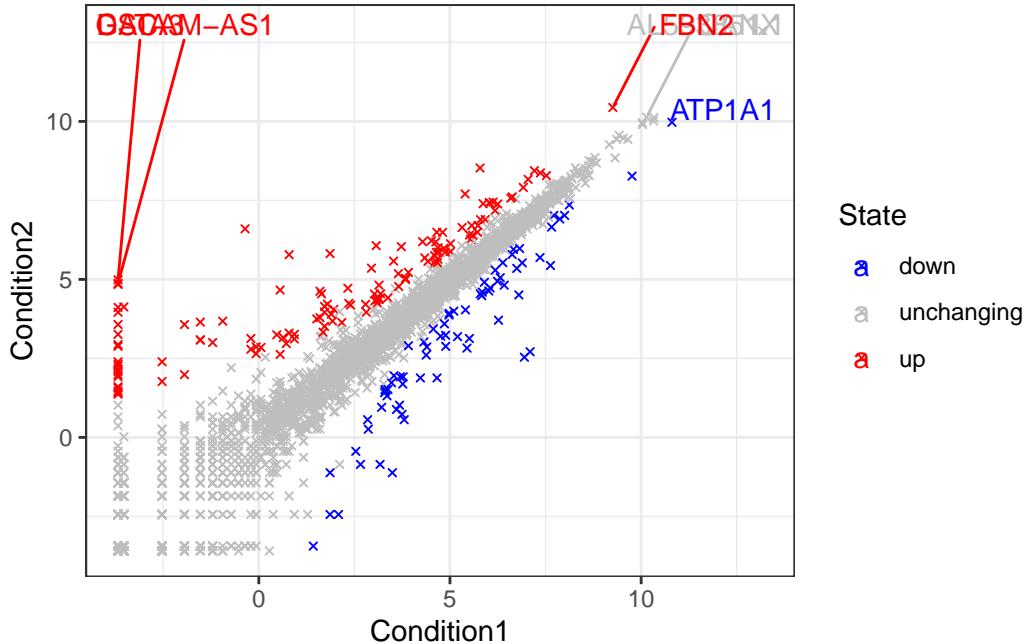
```
p <- ggplot(genes) +  
  aes(x=Condition1, y= Condition2, col = State) +  
  geom_point(pch = 4, size = 1) +  
  scale_color_manual(values = c("blue", "gray", "red"))
```

```
p + labs(title = "Gene")
```



```
library(ggrepel)
ggplot(genes) +
  aes(x=Condition1, y= Condition2, col = State,
      label = Gene) +
  geom_point(pch = 4, size = 1) +
  scale_color_manual(values = c("blue", "gray", "red")) +
  geom_text_repel(max.overlaps = 60) +
  theme_bw()
```

Warning: ggrepel: 5190 unlabeled data points (too many overlaps). Consider increasing max.overlaps



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

```
head(gapminder)
```

	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
5	Afghanistan	Asia	1972	36.088	13079460	739.9811
6	Afghanistan	Asia	1977	38.438	14880372	786.1134

```
tail(gapminder)
```

	country	continent	year	lifeExp	pop	gdpPerCap
1699	Zimbabwe	Africa	1982	60.363	7636524	788.8550
1700	Zimbabwe	Africa	1987	62.351	9216418	706.1573
1701	Zimbabwe	Africa	1992	60.377	10704340	693.4208
1702	Zimbabwe	Africa	1997	46.809	11404948	792.4500

```

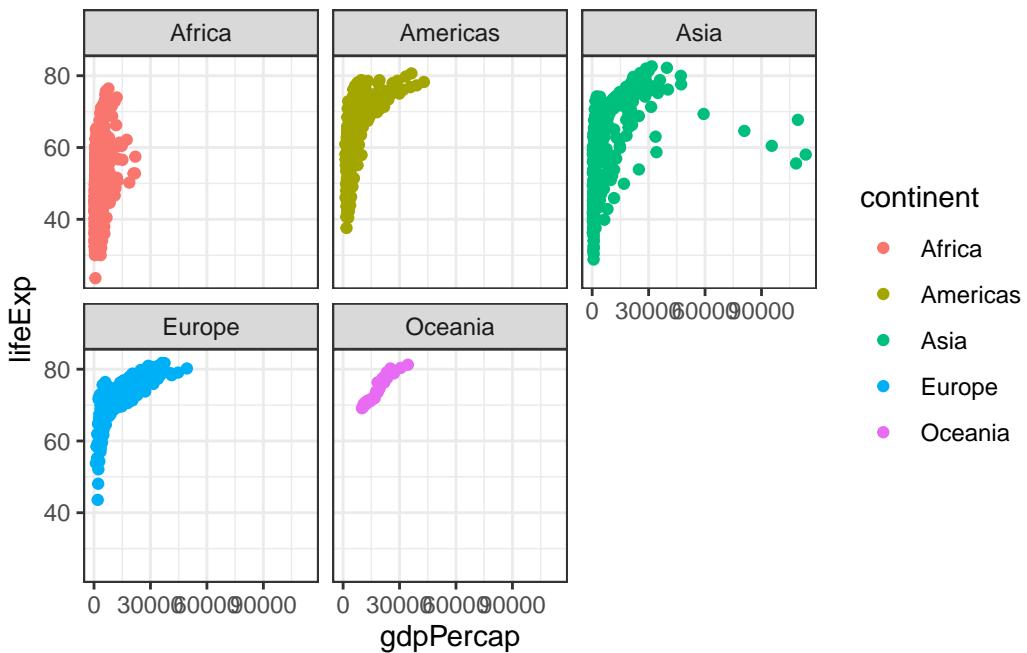
1703 Zimbabwe      Africa 2002  39.989 11926563  672.0386
1704 Zimbabwe      Africa 2007  43.487 12311143  469.7093

```

```

p1 <- ggplot(gapminder) +
  aes(gdpPercap, lifeExp, col = continent) +
  geom_point() +
  facet_wrap(~continent) +
  theme_bw()
p1

```



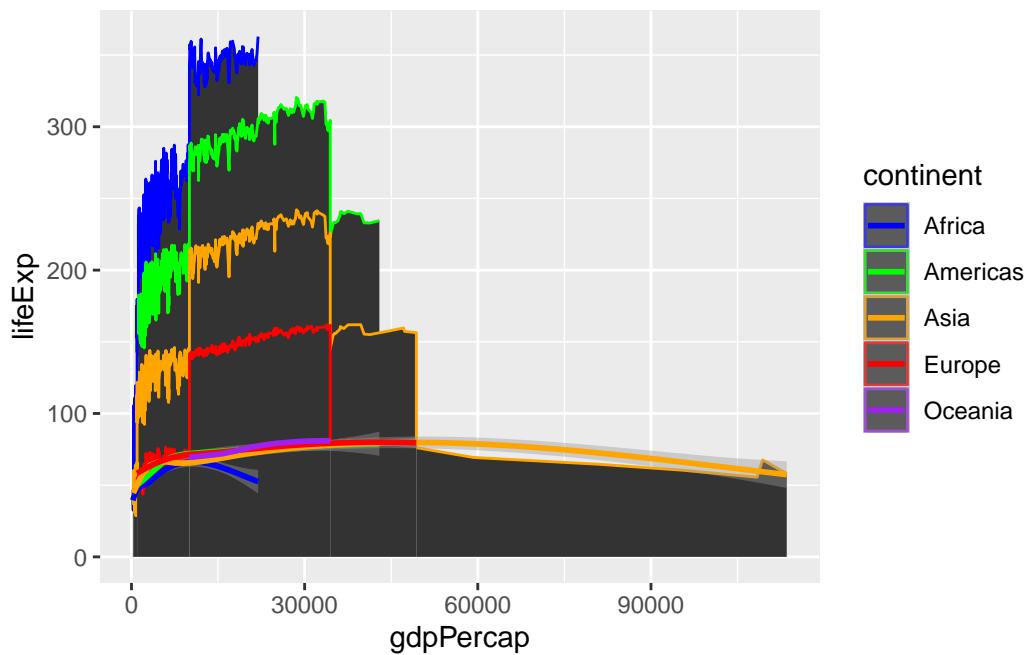
```

p2 <- ggplot(gapminder) +
  aes(gdpPercap, lifeExp, col = continent) +
  geom_area() + geom_smooth() +
  scale_color_manual(values = c("blue", "green", "orange", "red", "purple"))

```

```
p2
```

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



```
library(patchwork)
{p1 / p2}
```

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

