

Class 05: Data Visualization with GGPLOT

Karen Guerrero

Table of contents

Our first plot	1
----------------	---

Our first plot

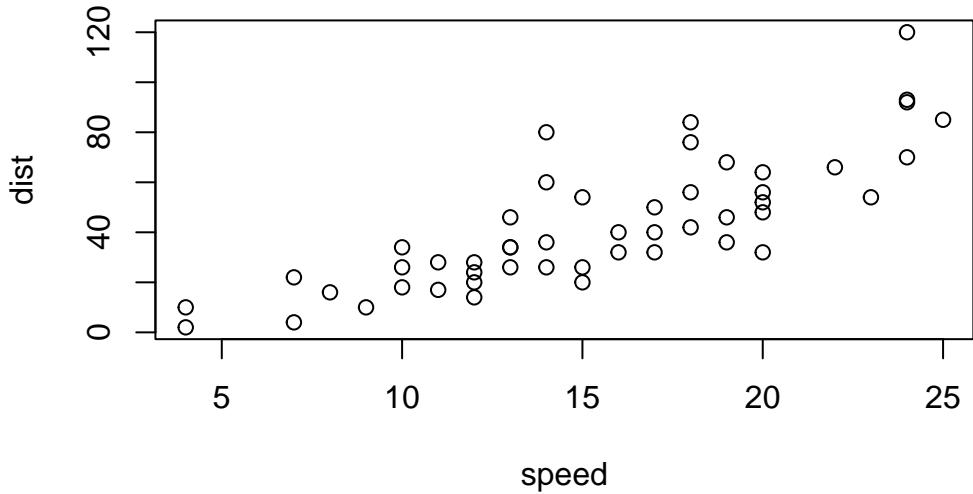
R has base graphics

```
| cars
```

```
  speed dist
1      4     2
2      4    10
3      7     4
4      7    22
5      8    16
6      9    10
7     10    18
8     10    26
9     10    34
10    11    17
11    11    28
12    12    14
13    12    20
14    12    24
15    12    28
16    13    26
17    13    34
18    13    34
19    13    46
```

```
20    14    26
21    14    36
22    14    60
23    14    80
24    15    20
25    15    26
26    15    54
27    16    32
28    16    40
29    17    32
30    17    40
31    17    50
32    18    42
33    18    56
34    18    76
35    18    84
36    19    36
37    19    46
38    19    68
39    20    32
40    20    48
41    20    52
42    20    56
43    20    64
44    22    66
45    23    54
46    24    70
47    24    92
48    24    93
49    24   120
50    25    85
```

```
| plot(cars)
```



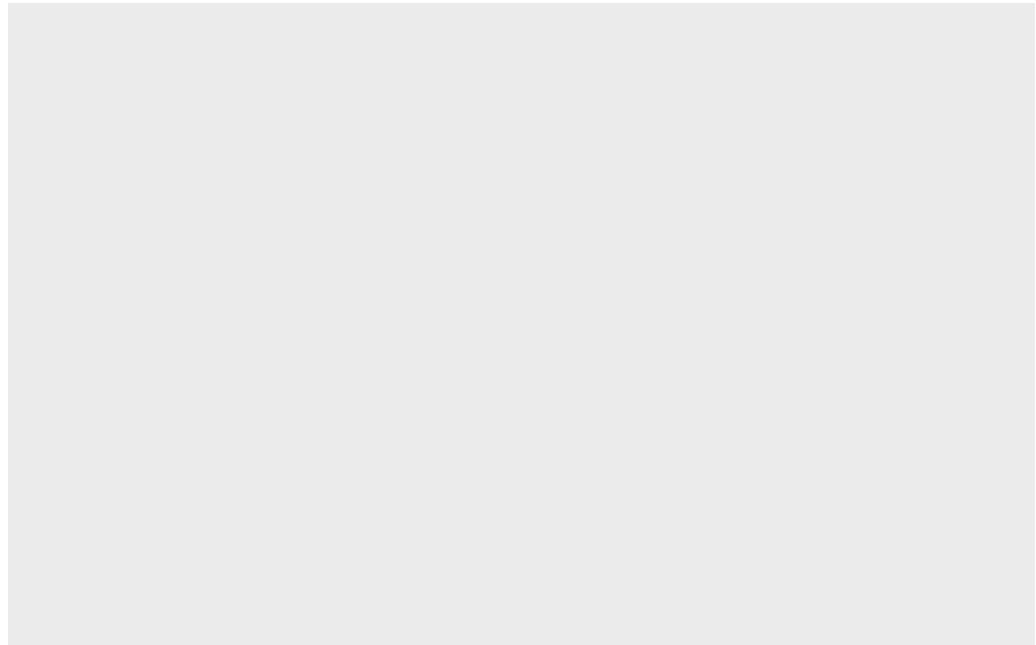
```
head(cars)
```

	speed	dist
1	4	2
2	4	10
3	7	4
4	7	22
5	8	16
6	9	10

```
library(ggplot2)
```

How would I plot this with 'ggplot2'?

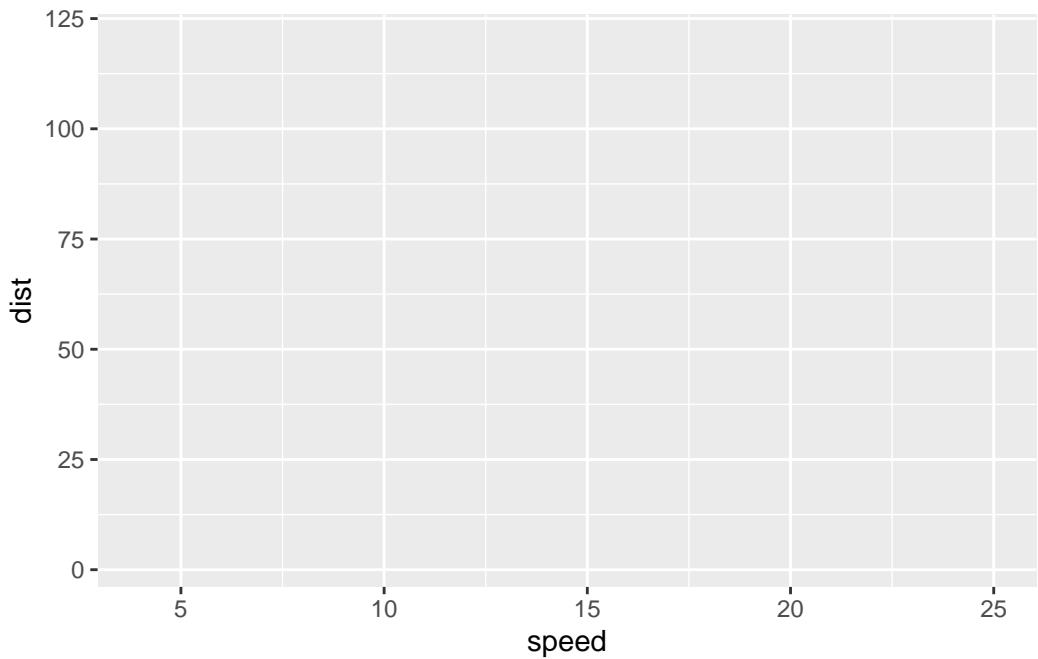
```
ggplot(cars)
```



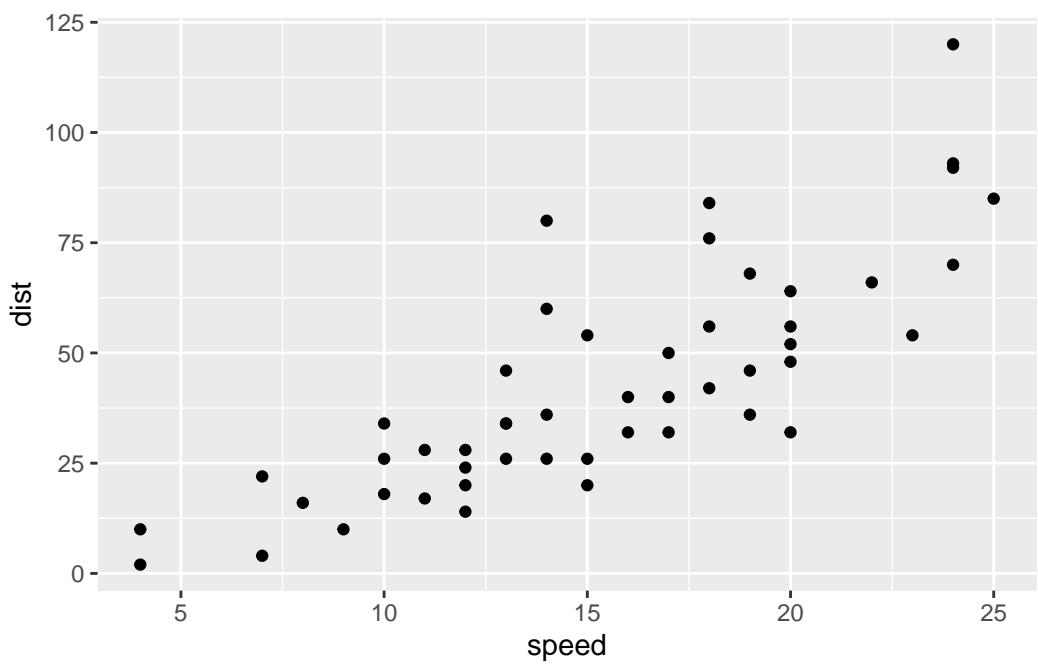
Every ggplot needs at least 3 layers:

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

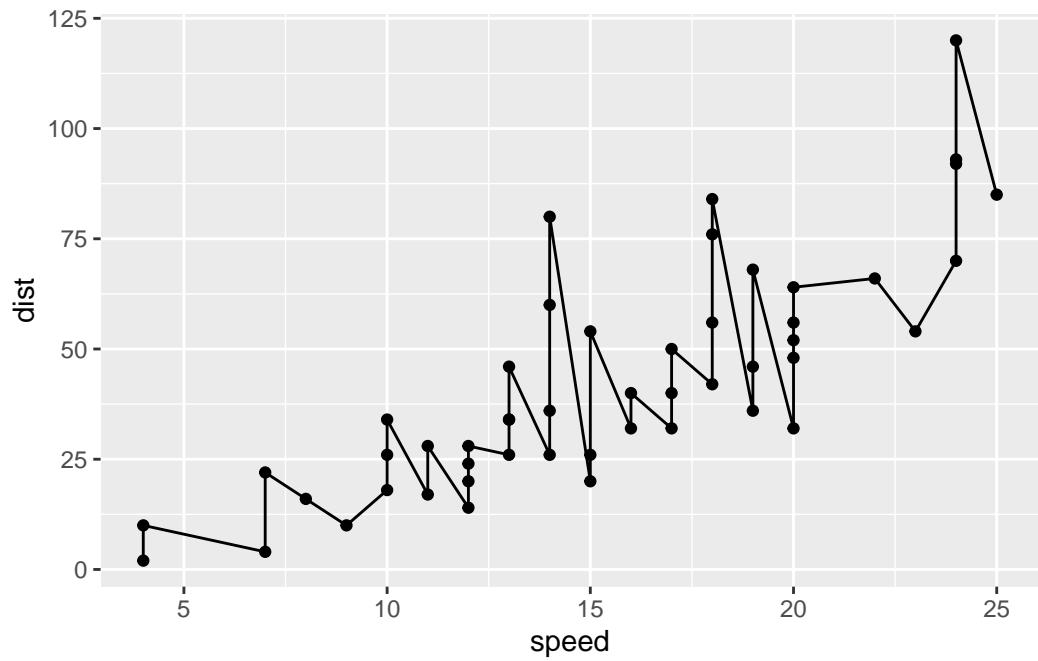
```
ggplot(cars) + aes(x=speed, y=dist)
```



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

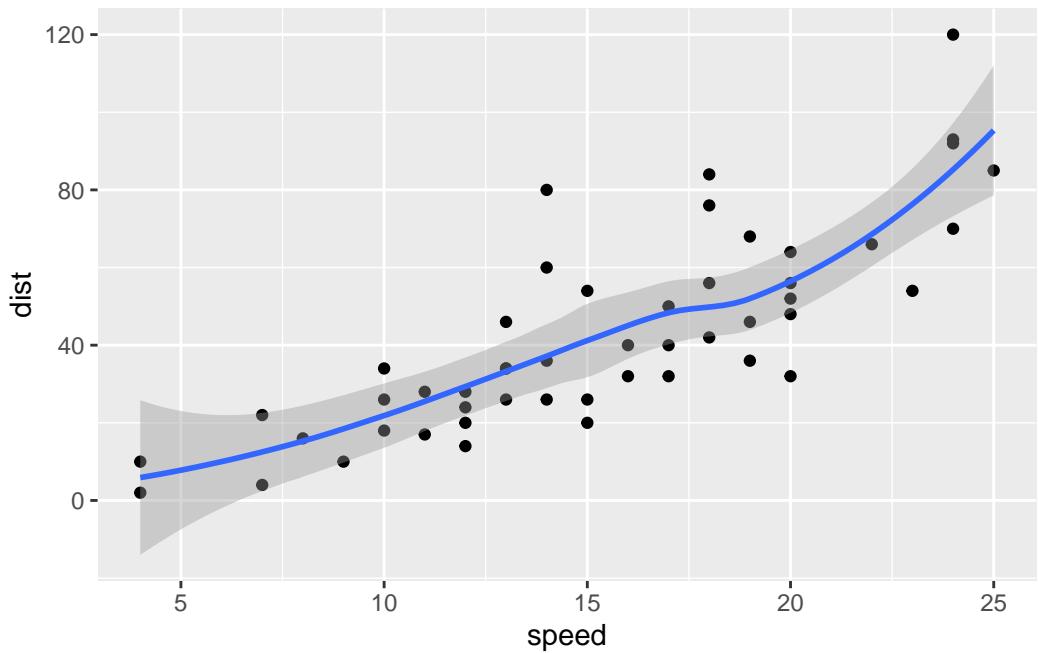


```
ggplot(cars) + aes(x=speed, y=dist) + geom_point() + geom_line()
```

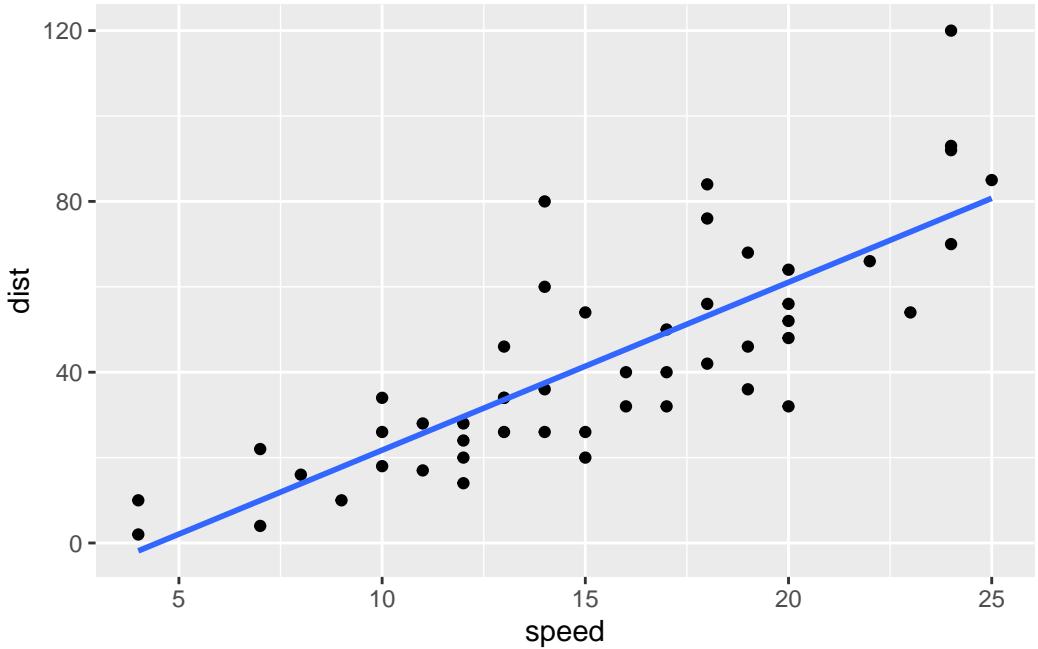


```
ggplot(cars) + aes(x=speed, y=dist) + geom_point() + geom_smooth()
```

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



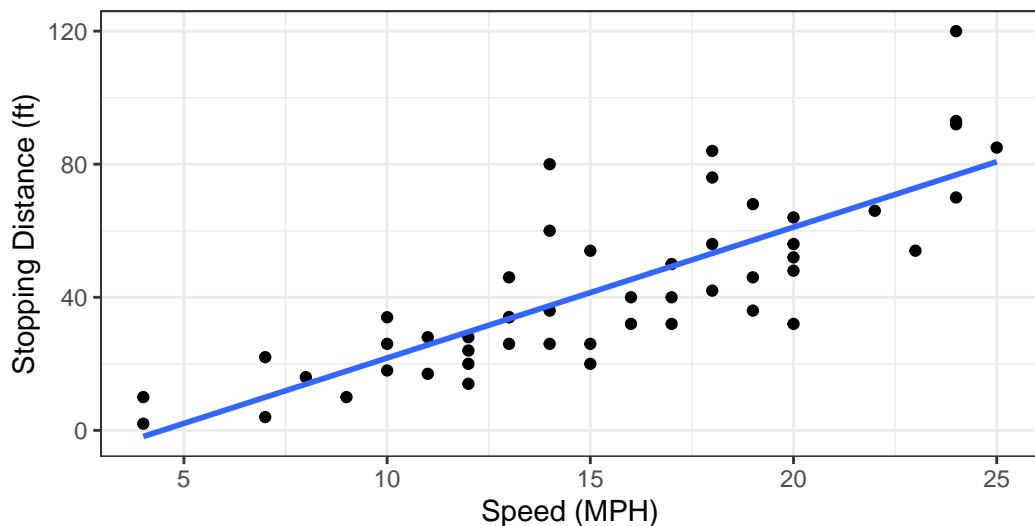
```
ggplot(cars) + aes(x=speed, y=dist) + geom_point() + geom_smooth(method="lm", se=FALSE)  
`geom_smooth()` using formula 'y ~ x'
```



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

Speed and Stopping Distances of Cars

Your informative subtitle text here



Dataset: 'cars'

```
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
```

There are 5196 genes in this data set.

```
colnames(genes)
```

```
[1] "Gene"      "Condition1" "Condition2" "State"
```

```
  ncol(genes)
```

```
[1] 4
```

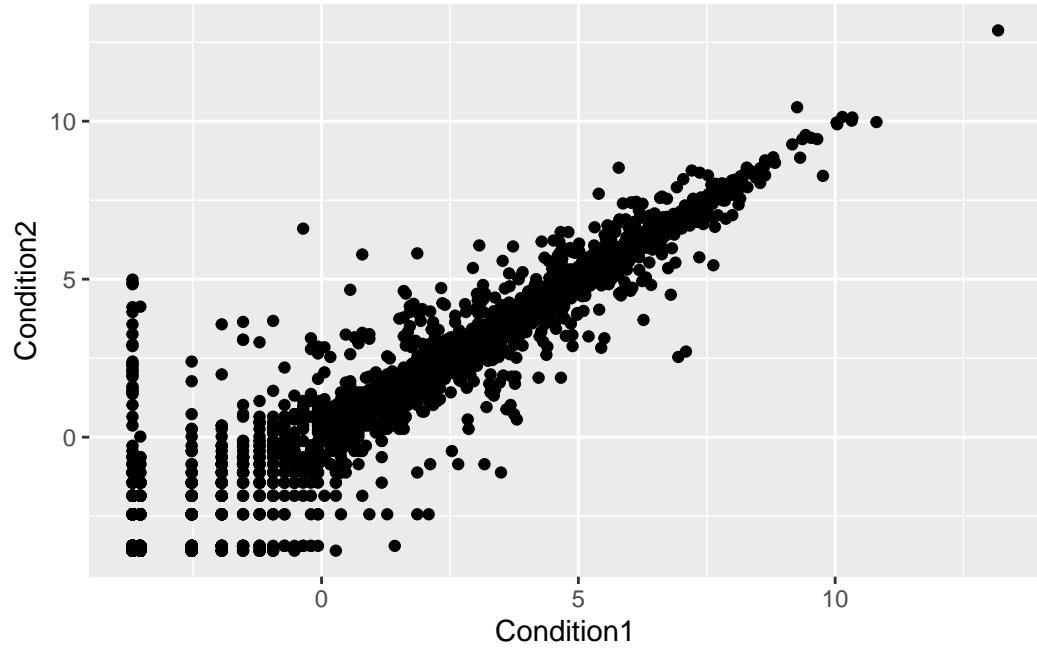
```
  table(genes$State)
```

	down	unchanging	up
	72	4997	127

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

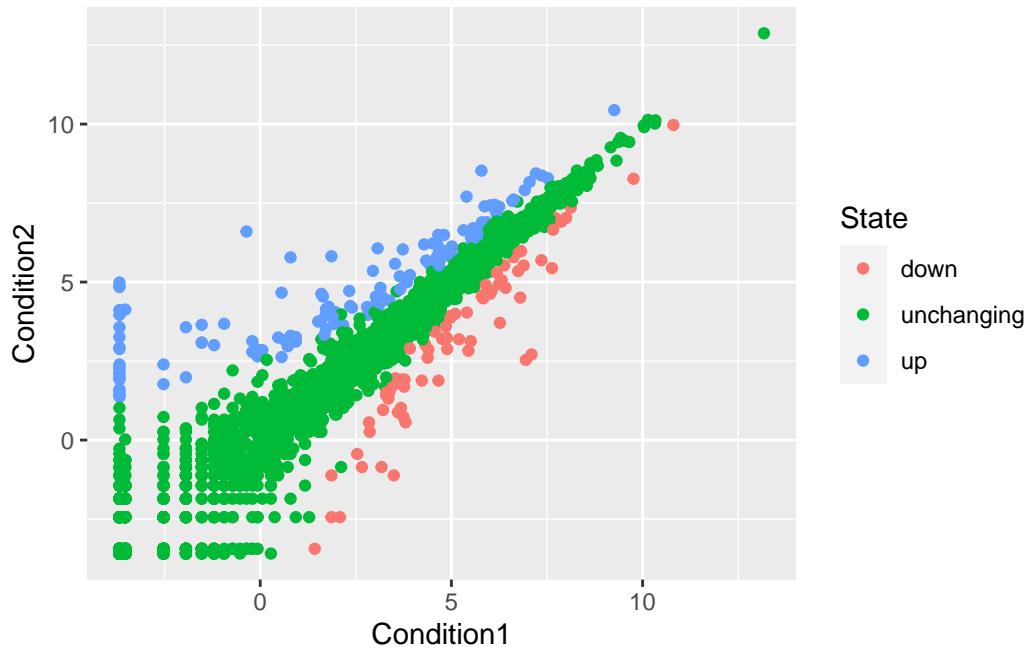
	down	unchanging	up
	1.39	96.17	2.44

```
  ggplot(genes) + aes(x=Condition1, y=Condition2) + geom_point()
```

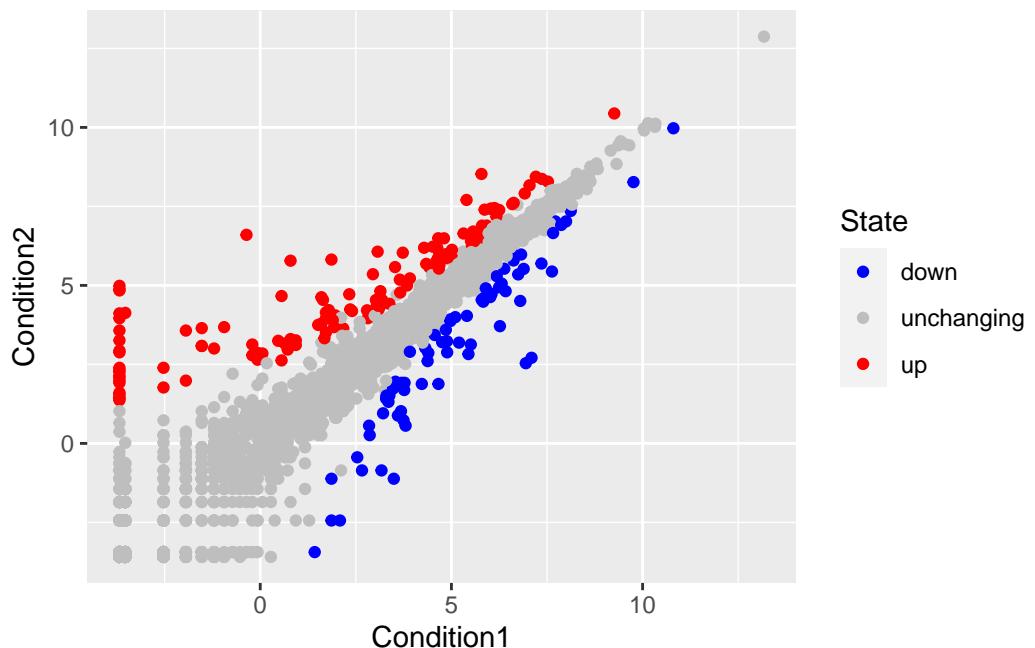


```
p <- ggplot(genes) + aes(x=Condition1, y=Condition2, col=State) + geom_point()
```

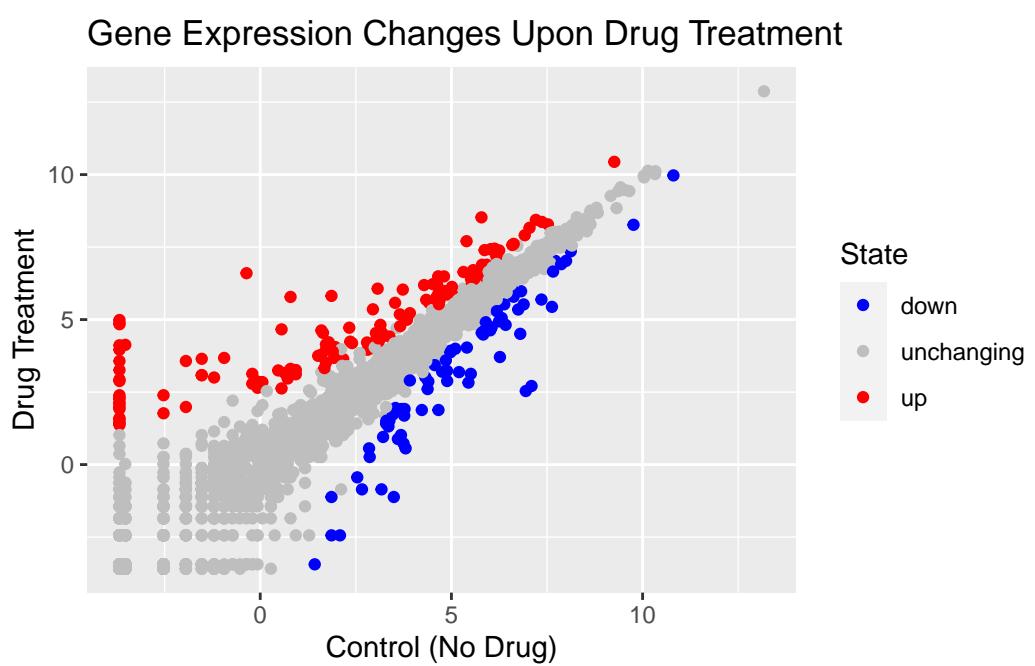
```
p
```



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



```
p + scale_colour_manual( values=c("blue","gray","red") ) + labs(title="Gene Expression Change Upon Drug Treatment")
```



```

#install.packages("gapminder")
library(gapminder)

#install.packages("dplyr")

library(dplyr)

Attaching package: 'dplyr'

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

filter, lag

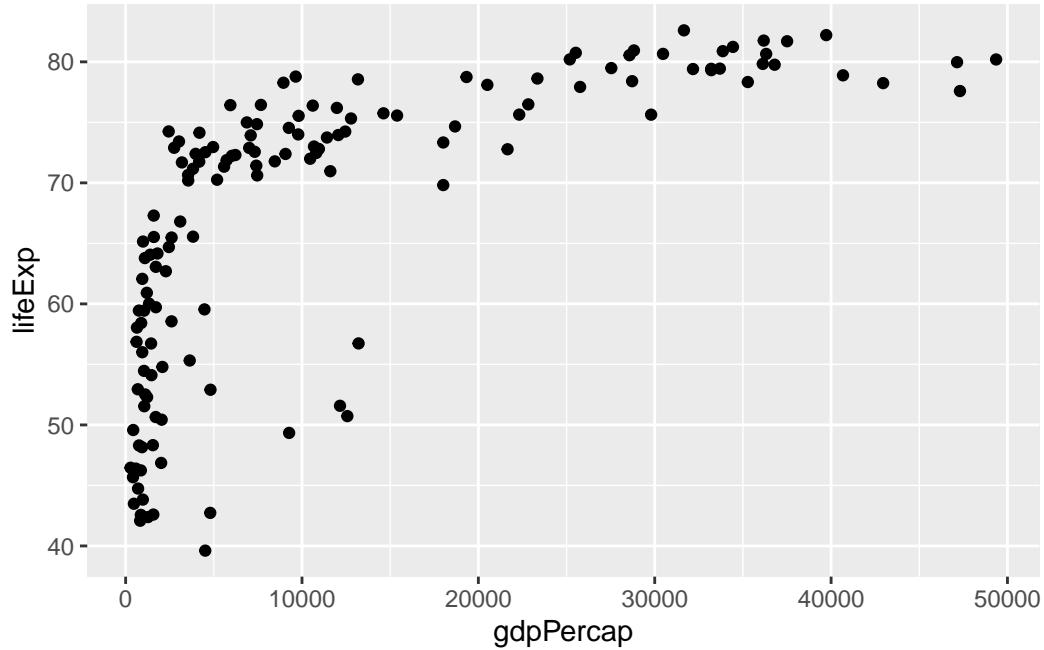
The following objects are masked from 'package:base':

intersect, setdiff, setequal, union

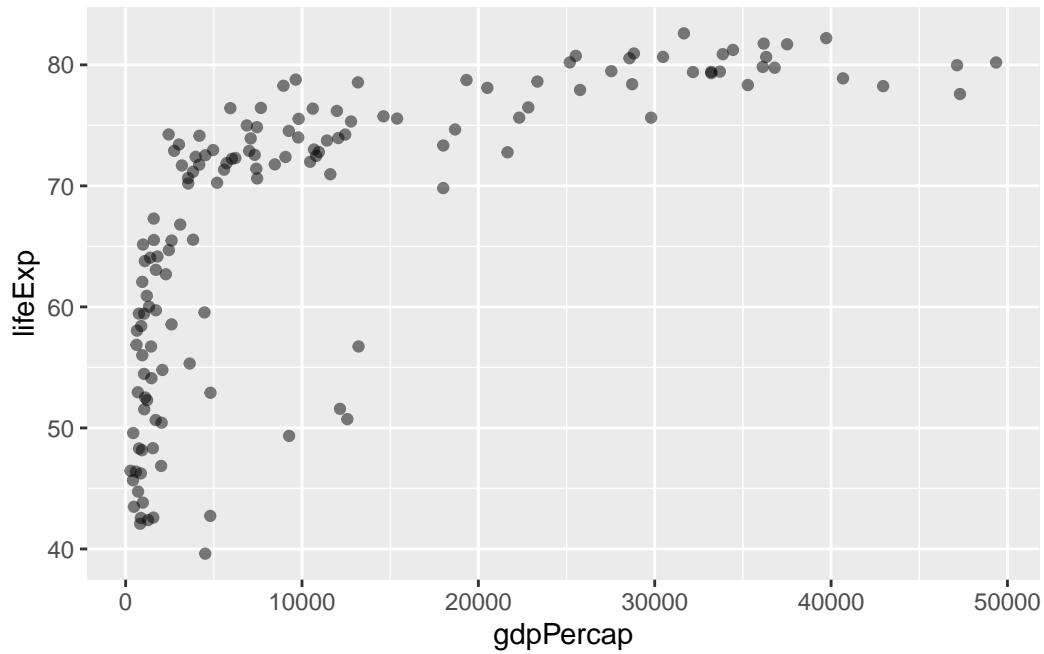
gapminder_2007 <- gapminder %>% filter(year==2007)

ggplot(gapminder_2007) + aes(x=gdpPercap, y=lifeExp) + geom_point()

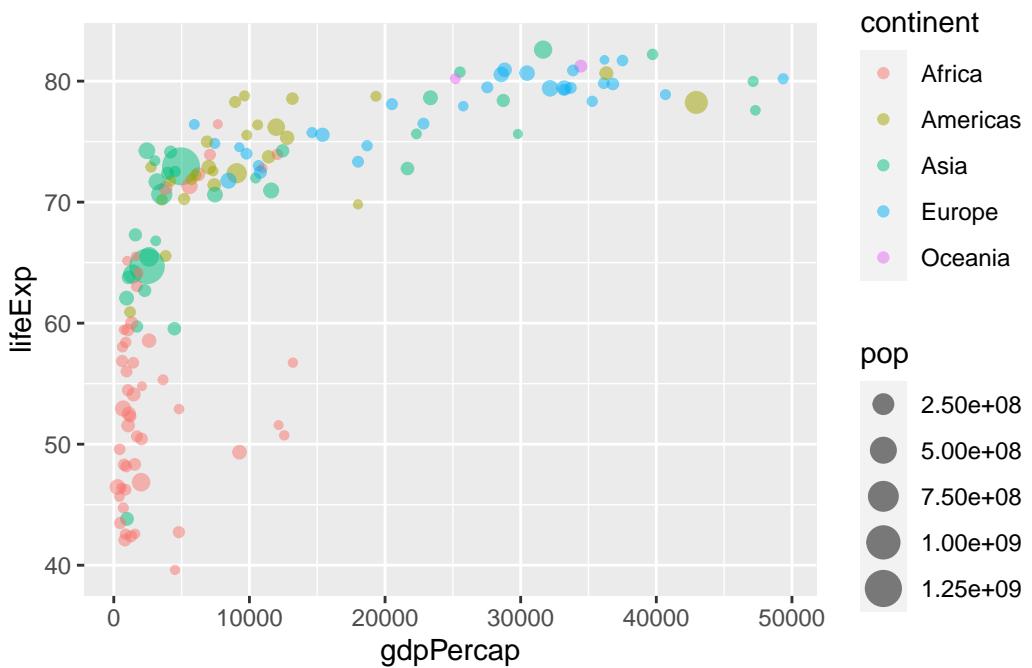
```



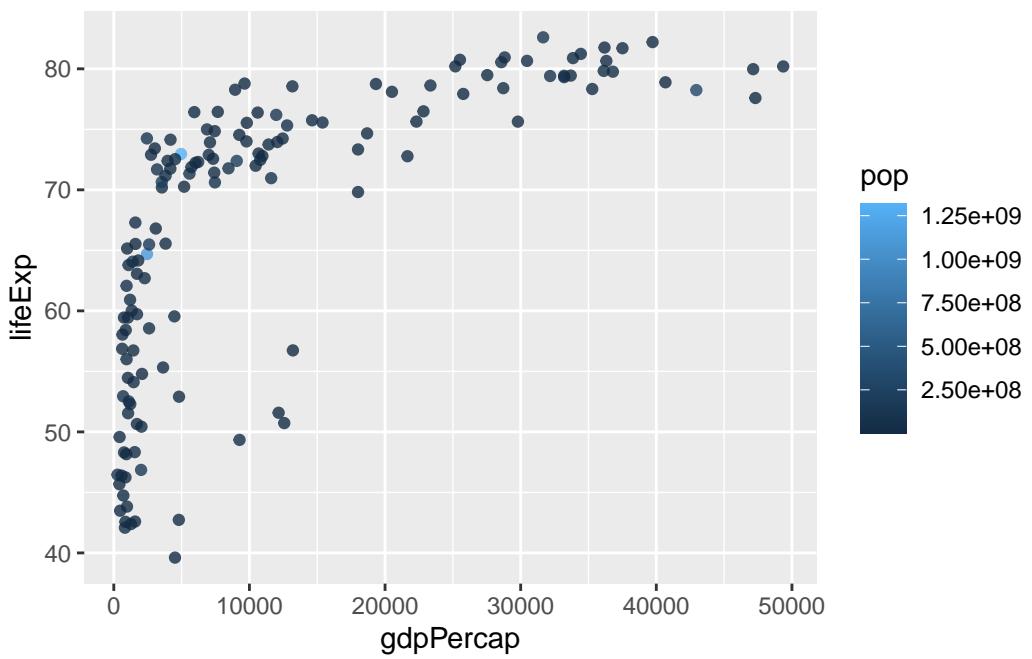
```
ggplot(gapminder_2007) + aes(x=gdpPercap, y=lifeExp) + geom_point(alpha=0.5)
```



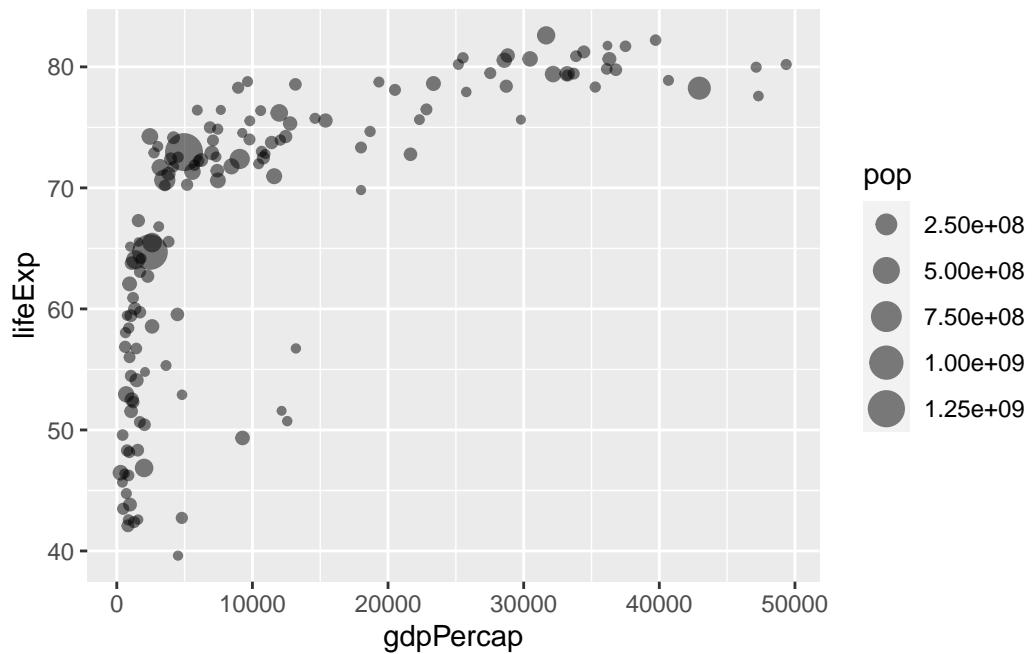
```
ggplot(gapminder_2007) + aes(x=gdpPercap, y=lifeExp, color=continent, size=pop) + geom_point()
```



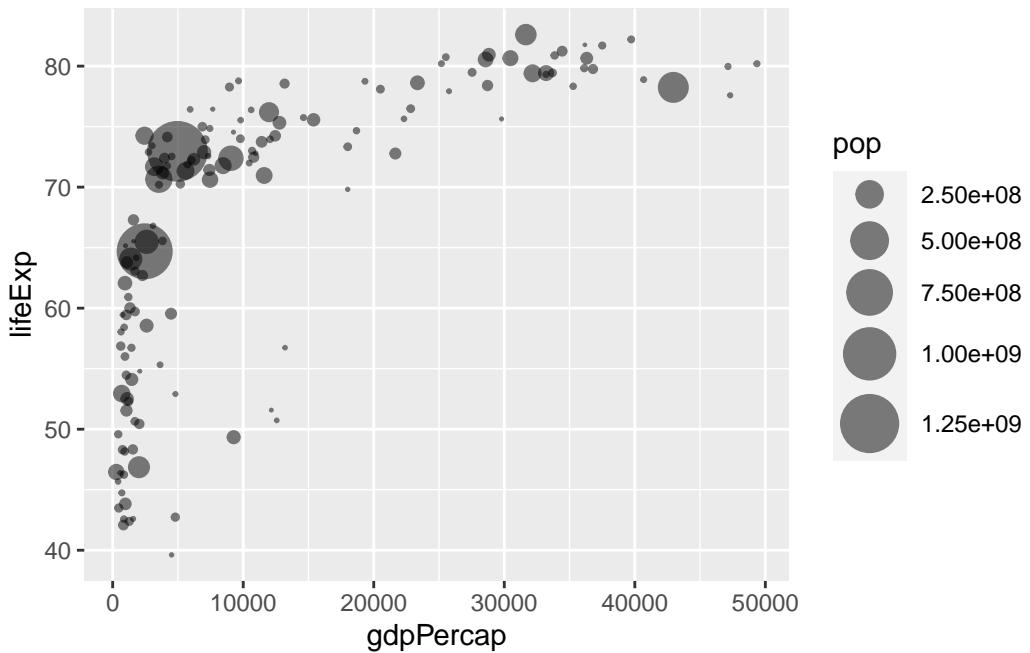
```
ggplot(gapminder_2007) + aes(x=gdpPercap, y=lifeExp, color=pop) + geom_point(alpha=0.8)
```



```
ggplot(gapminder_2007) + aes(x=gdpPercap, y=lifeExp, size=pop) + geom_point(alpha=0.5)
```

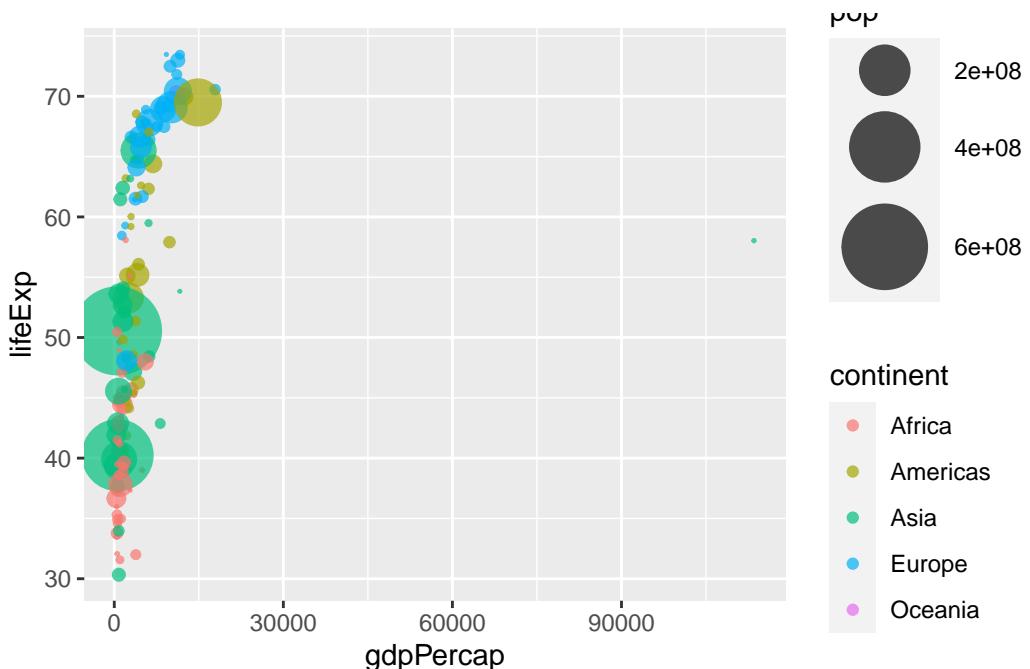


```
ggplot(gapminder_2007) + geom_point(aes(x = gdpPercap, y = lifeExp, size = pop), alpha=0.5)
```



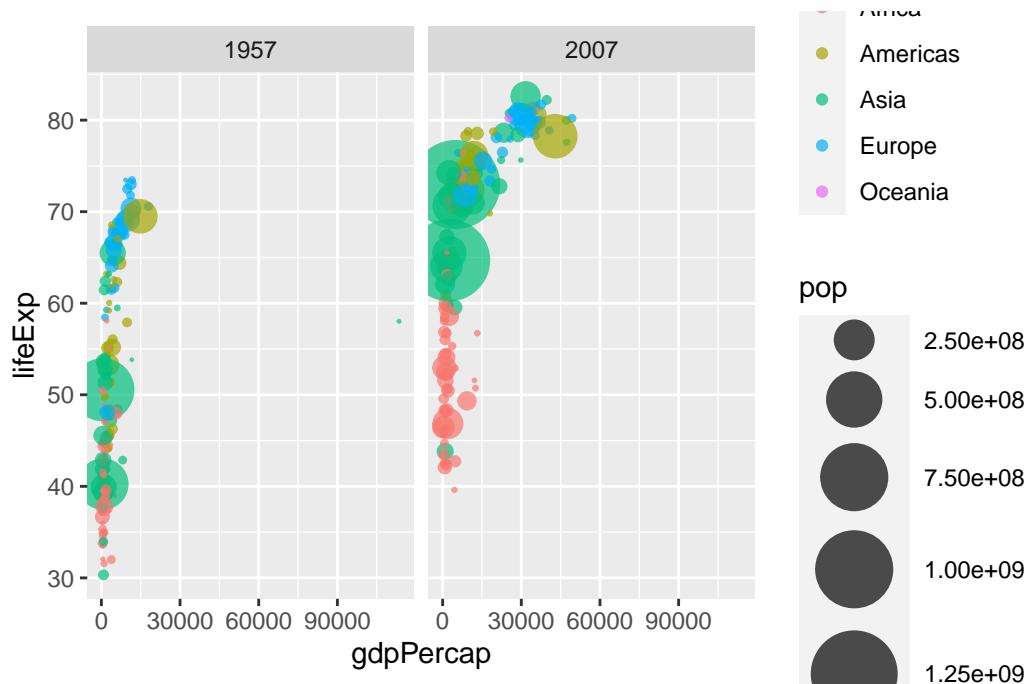
```
gapminder_1957 <- gapminder %>% filter(year==1957)
```

```
ggplot(gapminder_1957) + aes(x=gdpPercap, y=lifeExp, color=continent, size=pop) + geom_poi
```



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

```
ggplot(gapminder_1957) + aes(x=gdpPercap, y=lifeExp, color=continent, size=pop) + geom_poin
```

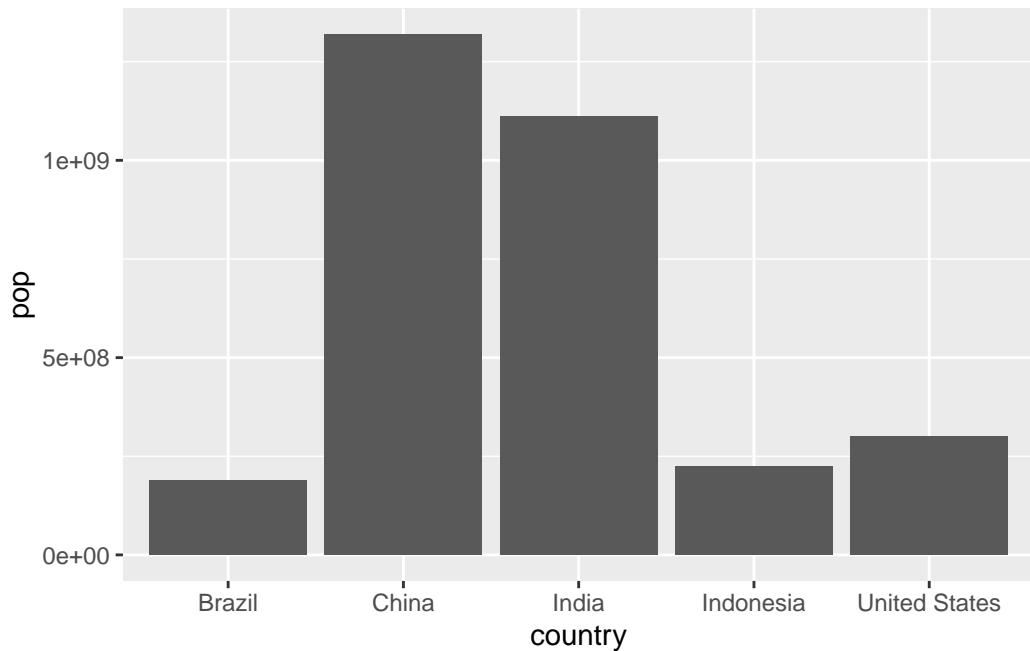


```
gapminder_top5 <- gapminder %>% filter(year==2007) %>% arrange(desc(pop)) %>% top_n(5, pop)
```

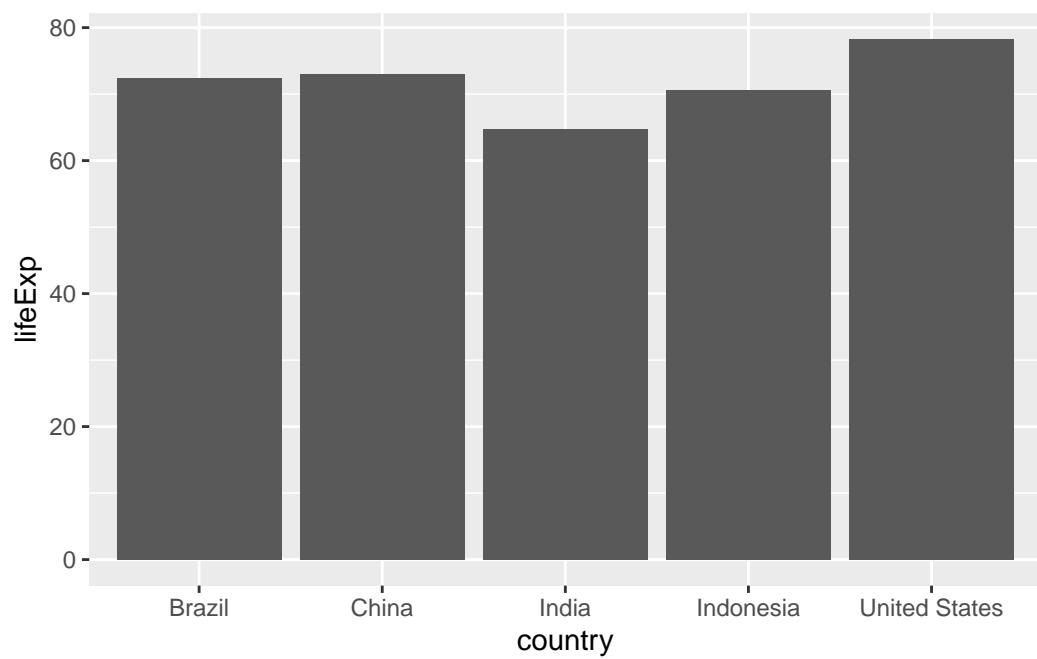
```
# A tibble: 5 x 6
```

	country	continent	year	lifeExp	pop	gdpPercap
	<fct>	<fct>	<int>	<dbl>	<int>	<dbl>
1	China	Asia	2007	73.0	1318683096	4959.
2	India	Asia	2007	64.7	1110396331	2452.
3	United States	Americas	2007	78.2	301139947	42952.
4	Indonesia	Asia	2007	70.6	223547000	3541.
5	Brazil	Americas	2007	72.4	190010647	9066.

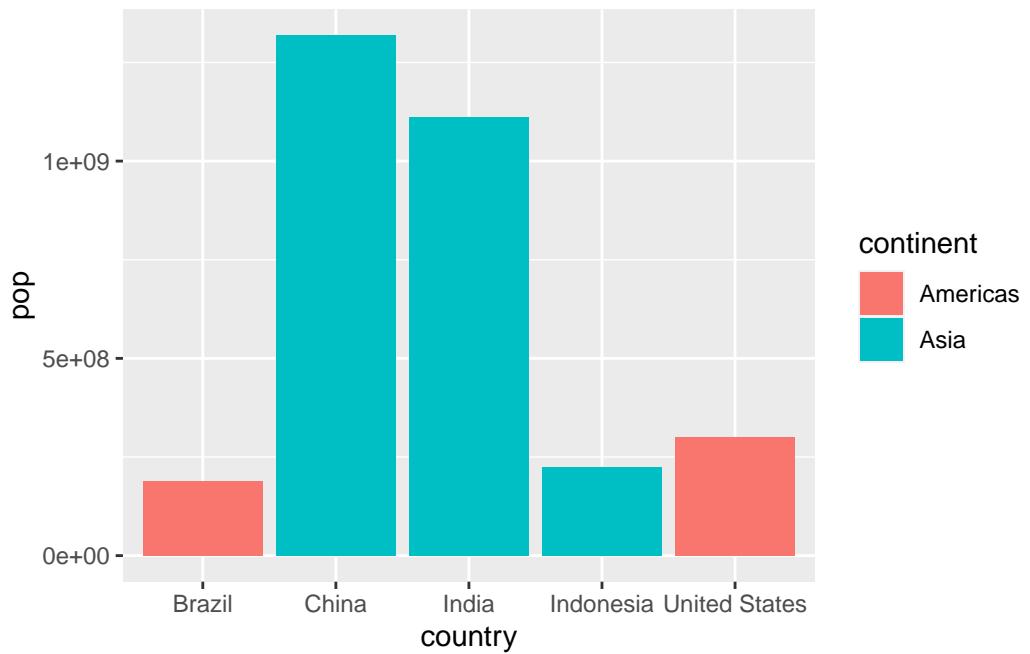
```
ggplot(gapminder_top5) + geom_col(aes(x = country, y = pop))
```



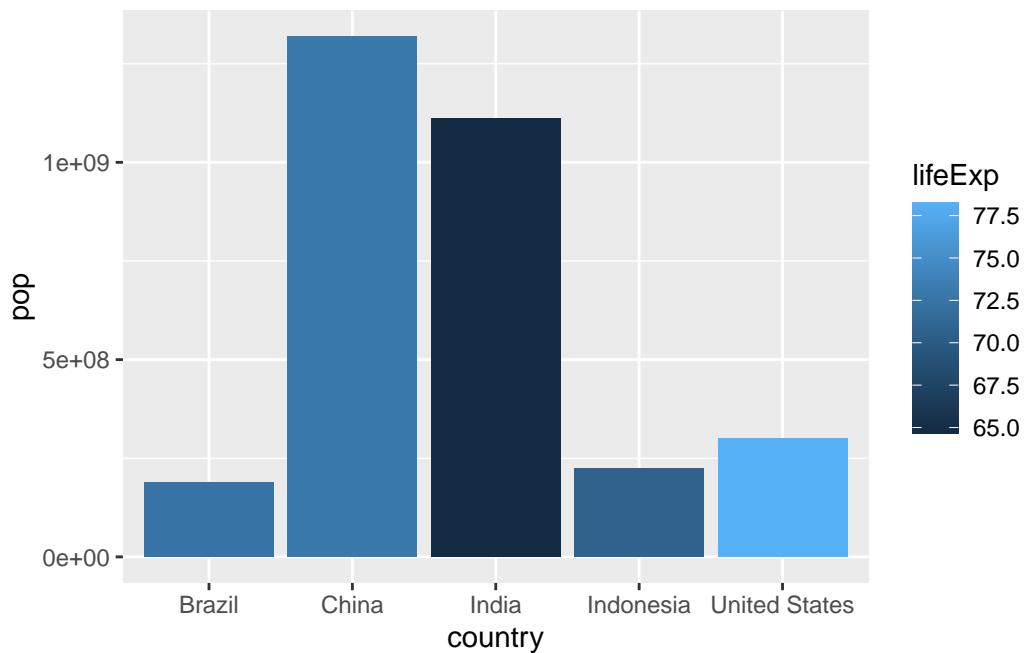
```
ggplot(gapminder_top5) + geom_col(aes(x = country, y = lifeExp))
```



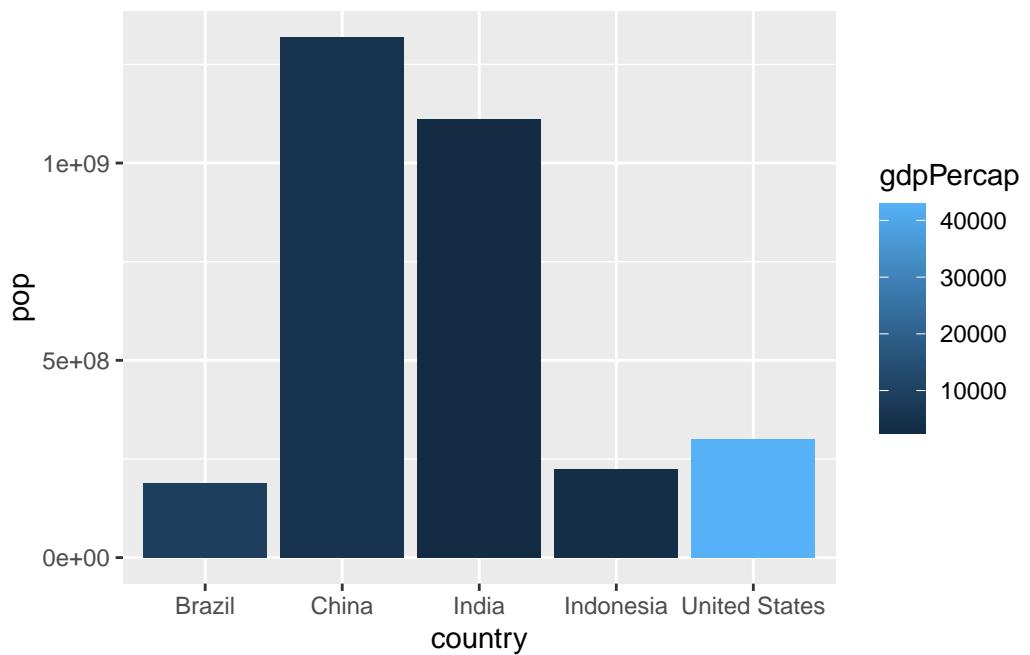
```
ggplot(gapminder_top5) + geom_col(aes(x = country, y = pop, fill=continent))
```



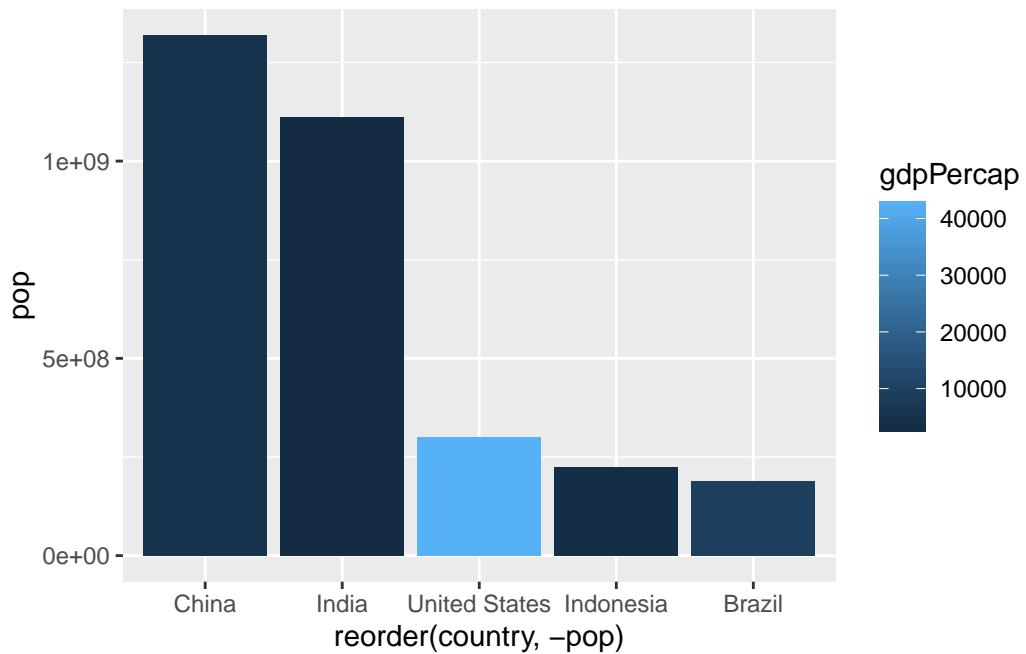
```
ggplot(gapminder_top5) + geom_col(aes(x = country, y = pop, fill=lifeExp))
```



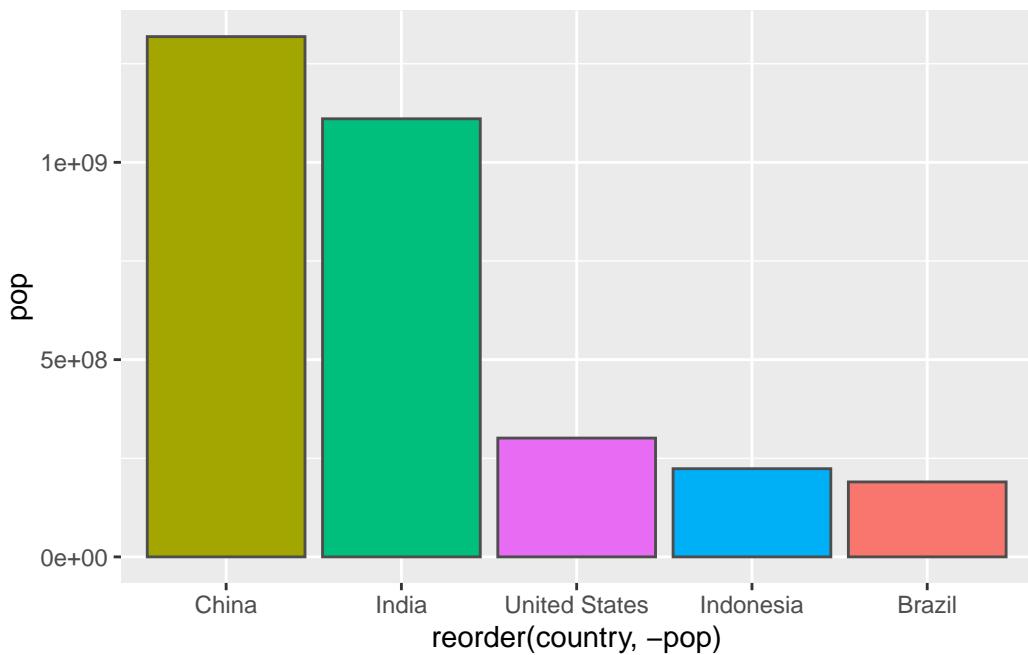
```
ggplot(gapminder_top5) + geom_col(aes(x = country, y = pop, fill=gdpPercap))
```



```
ggplot(gapminder_top5) + geom_col(aes(x = reorder(country, -pop), y = pop, fill=gdpPercap))
```



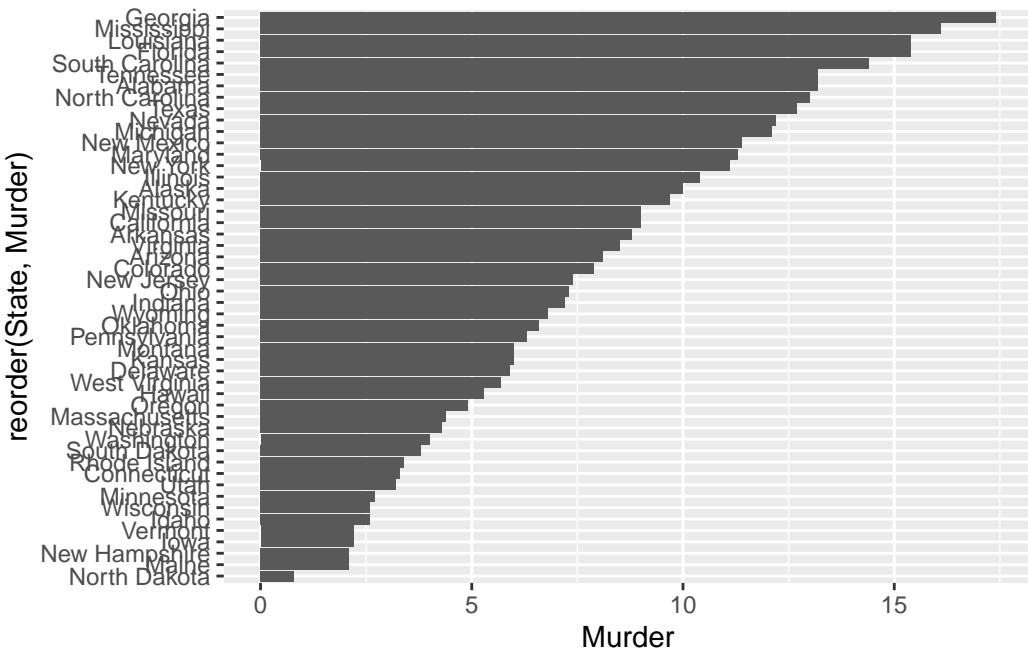
```
ggplot(gapminder_top5) + aes(x = reorder(country, -pop), y = pop, fill=country) + geom_col()
```



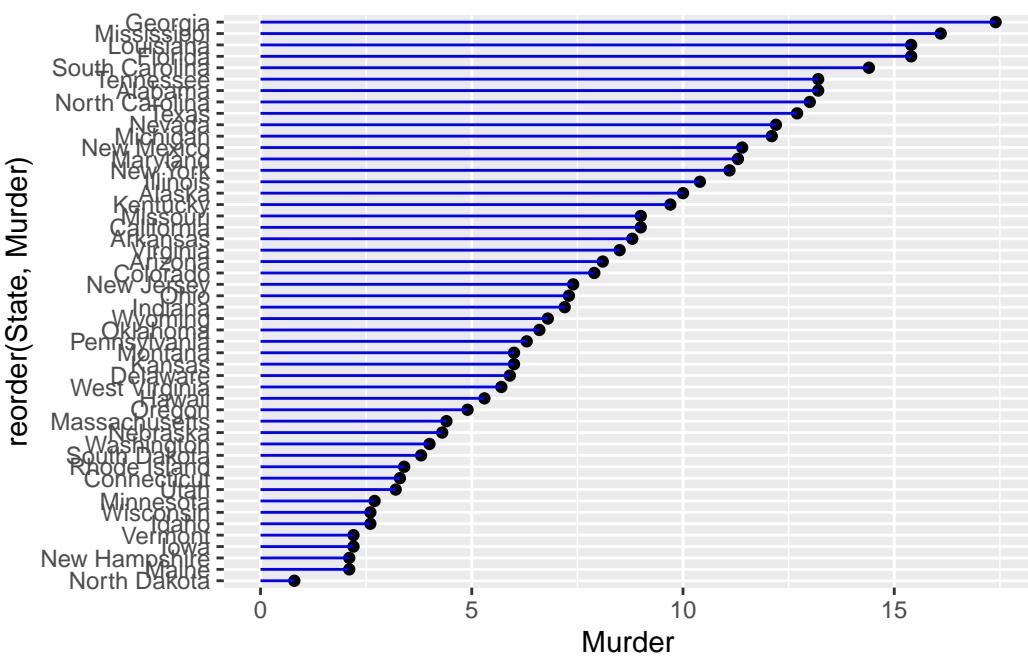
```
head(USArrests)
```

	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()
```



```
ggplot(USArrests) + aes(x=reorder(State,Murder), y=Murder) + geom_point() + geom_segment(a
```

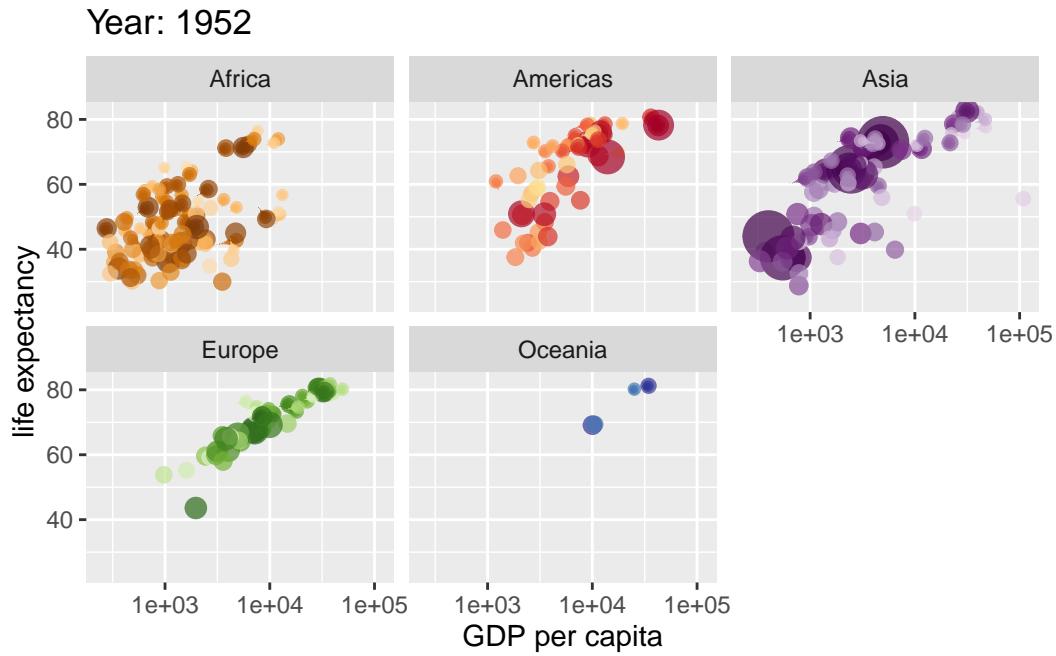


```

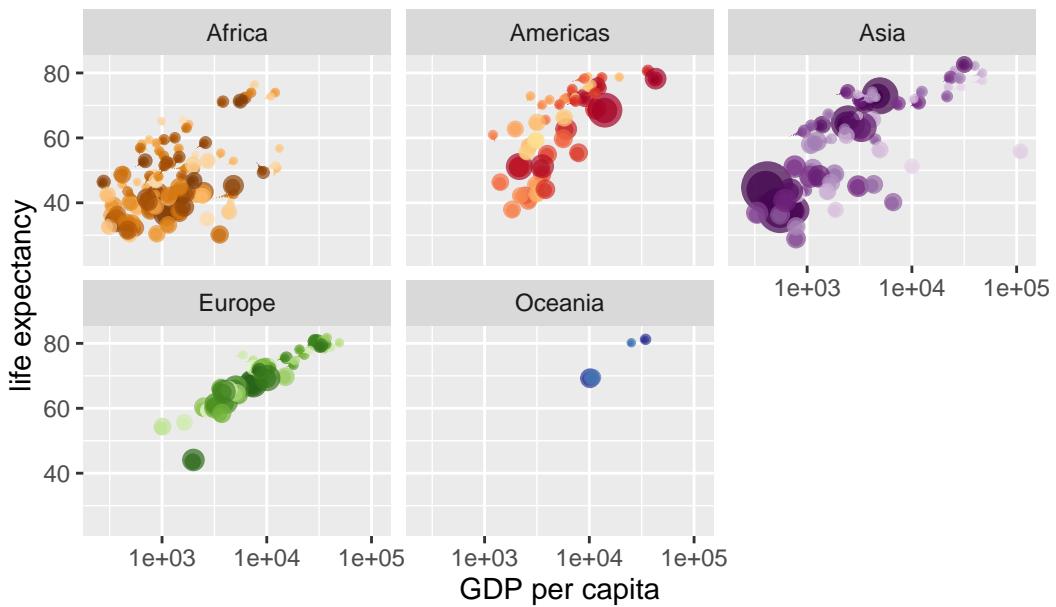
library(gapminder)
library(gganimate)

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)

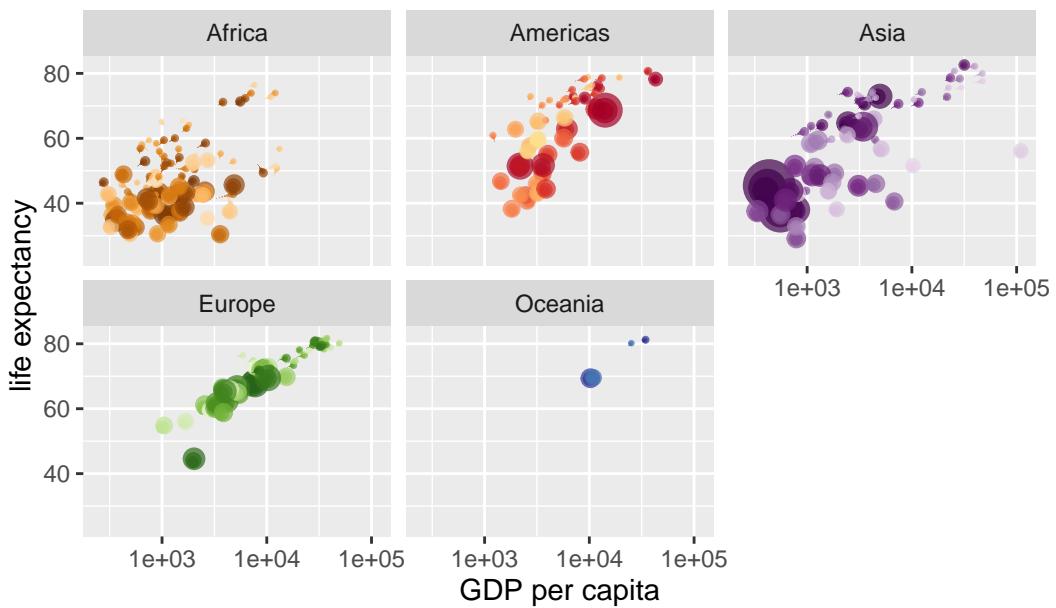
```



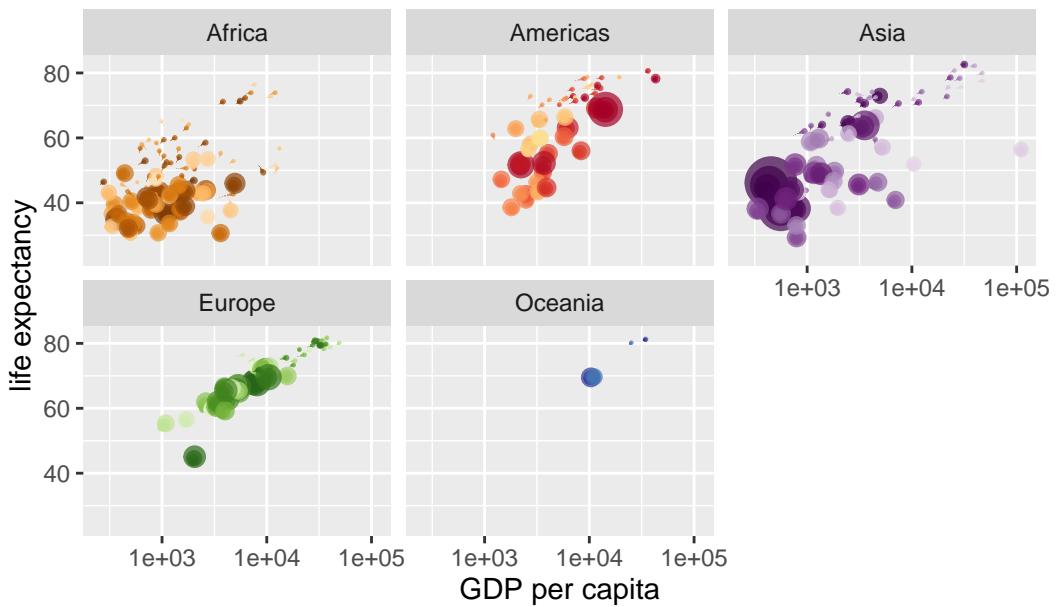
Year: 1953



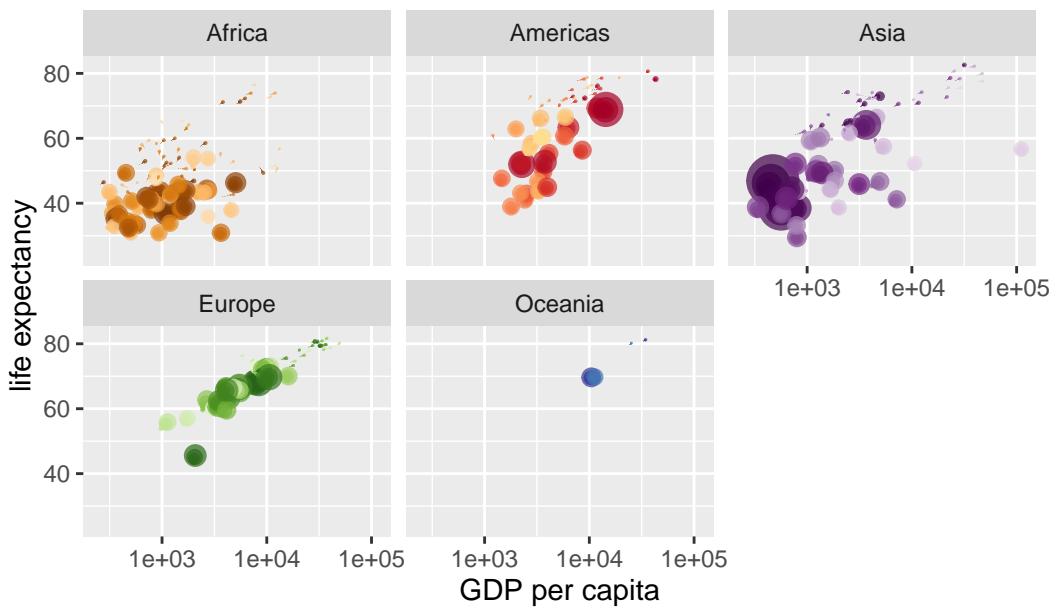
Year: 1953



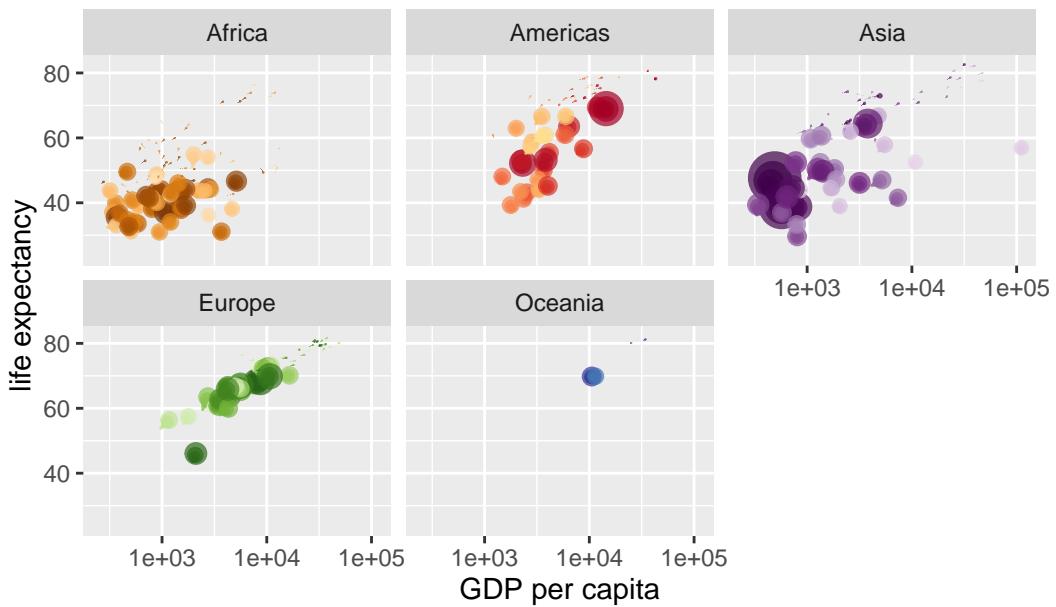
Year: 1954



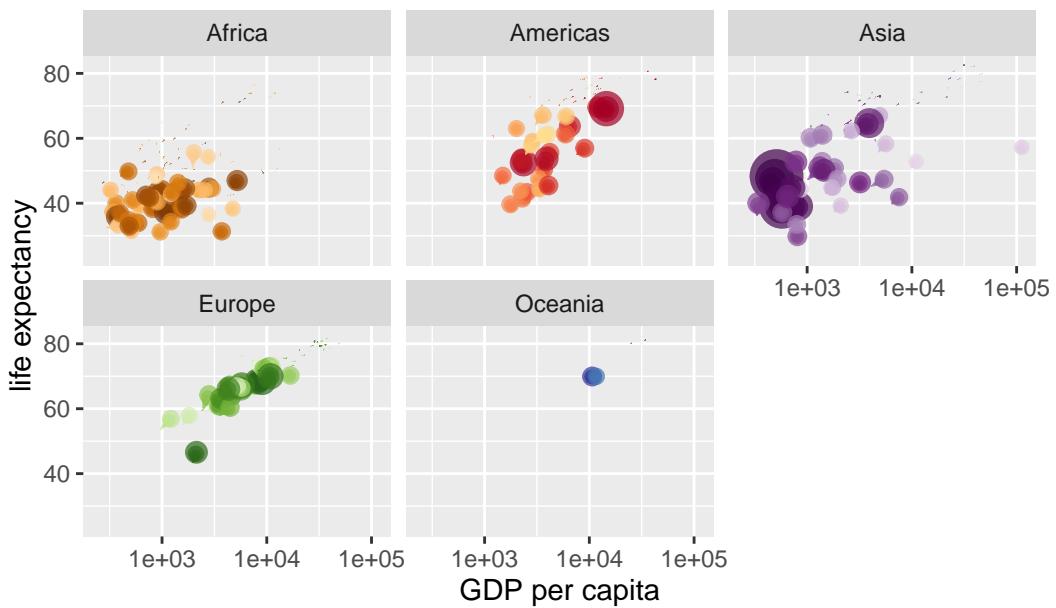
Year: 1954



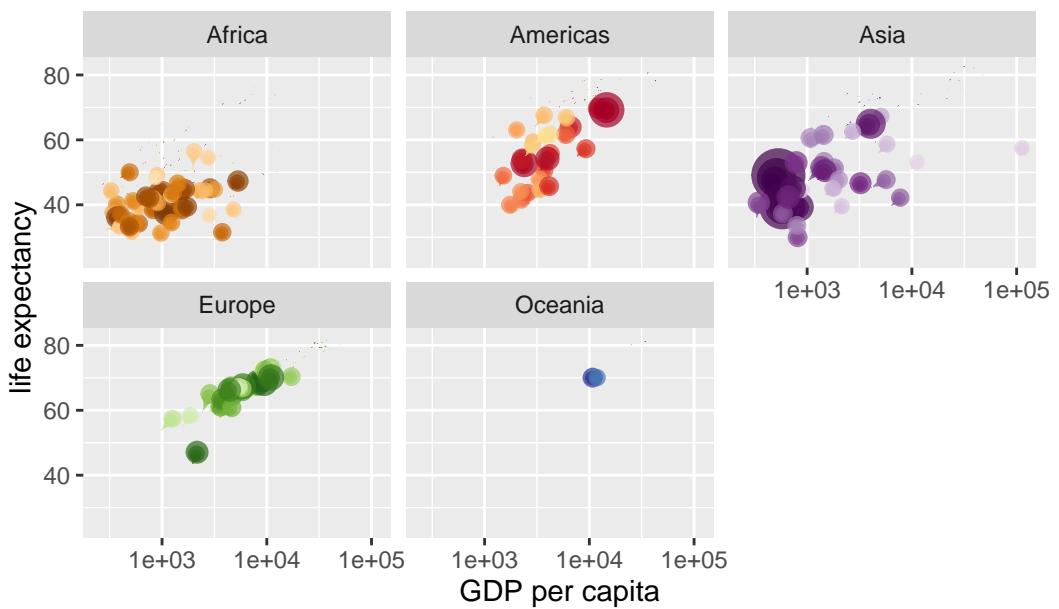
Year: 1955



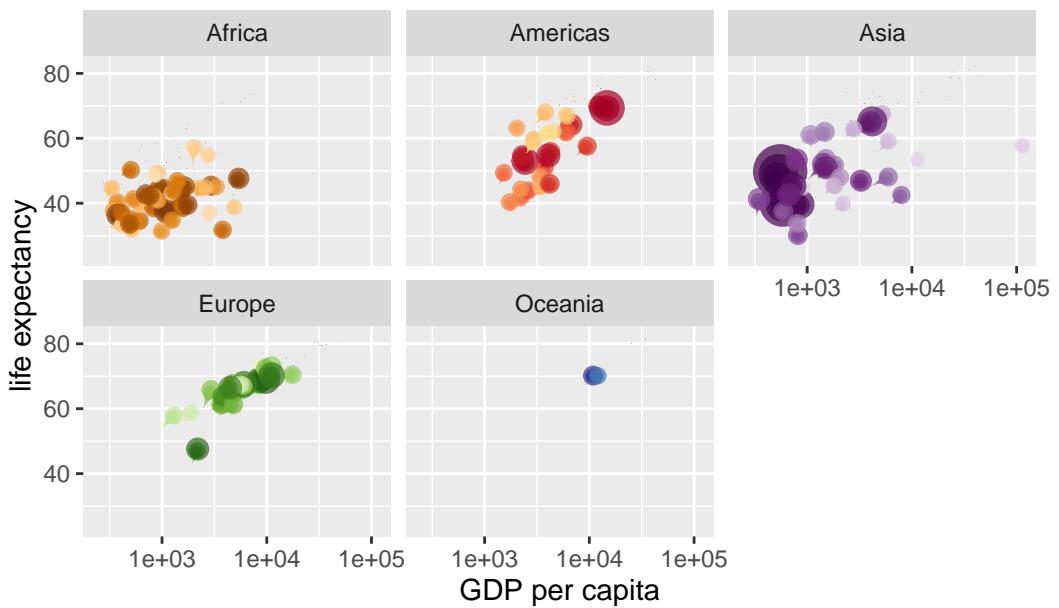
Year: 1955



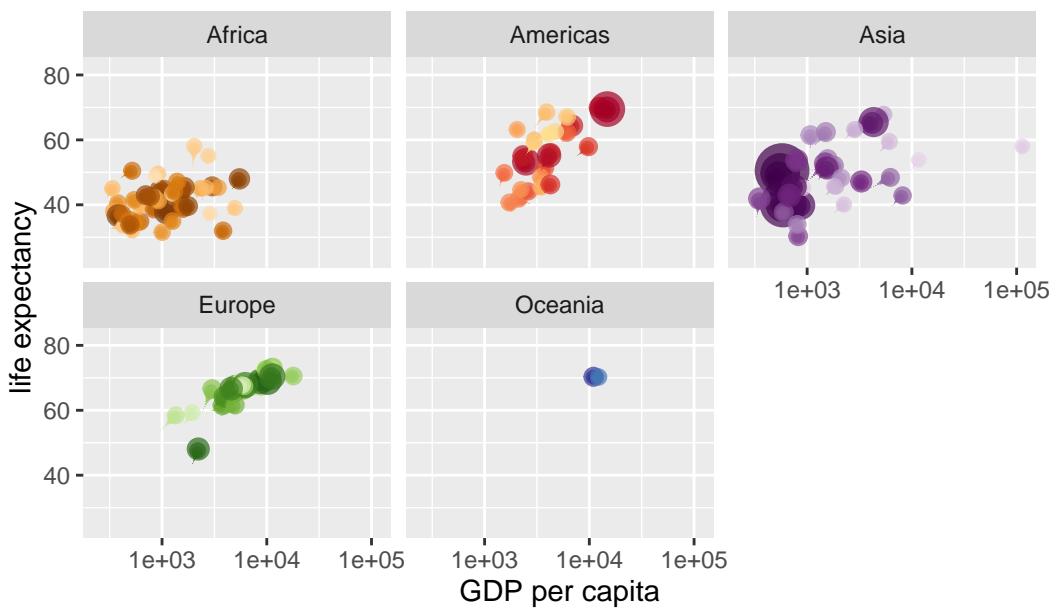
Year: 1956



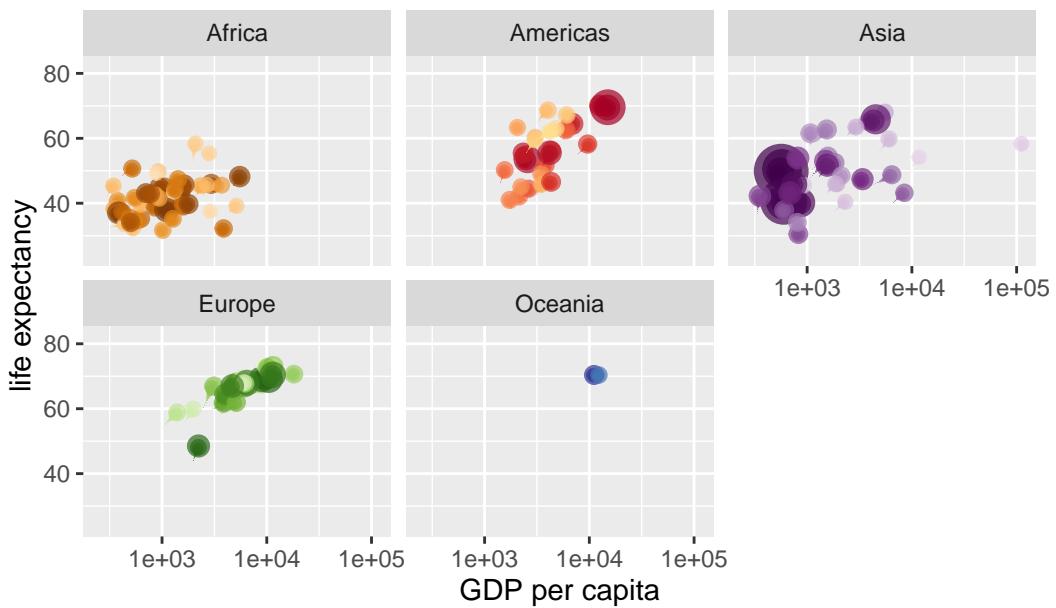
Year: 1956



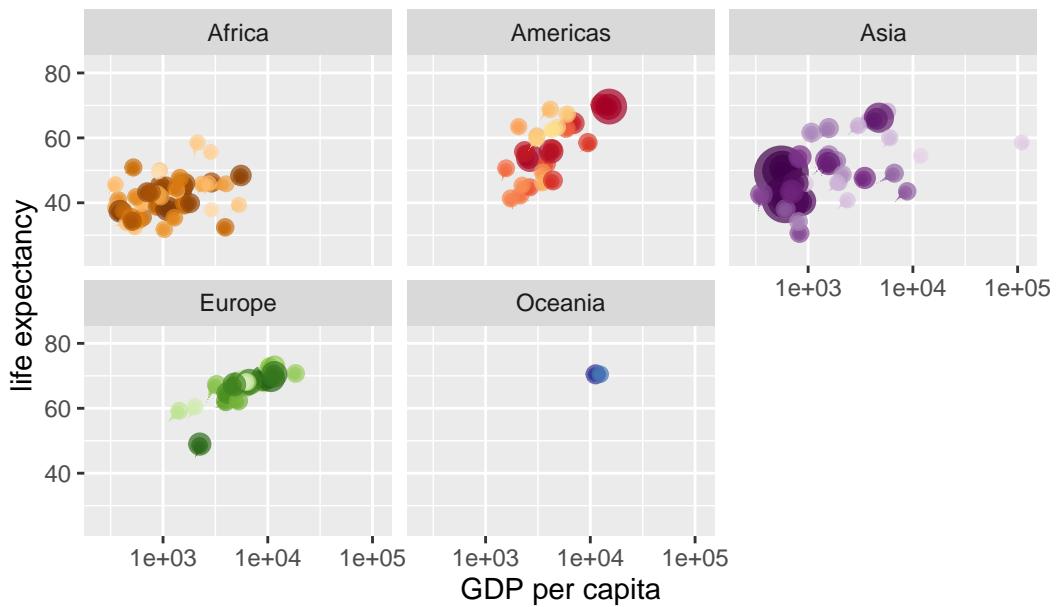
Year: 1957



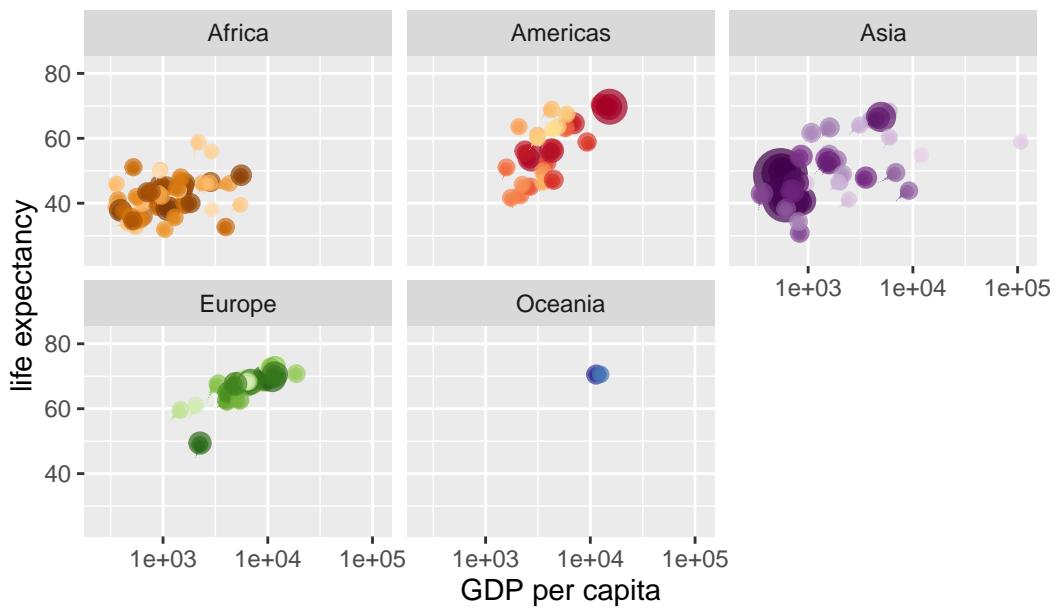
Year: 1958



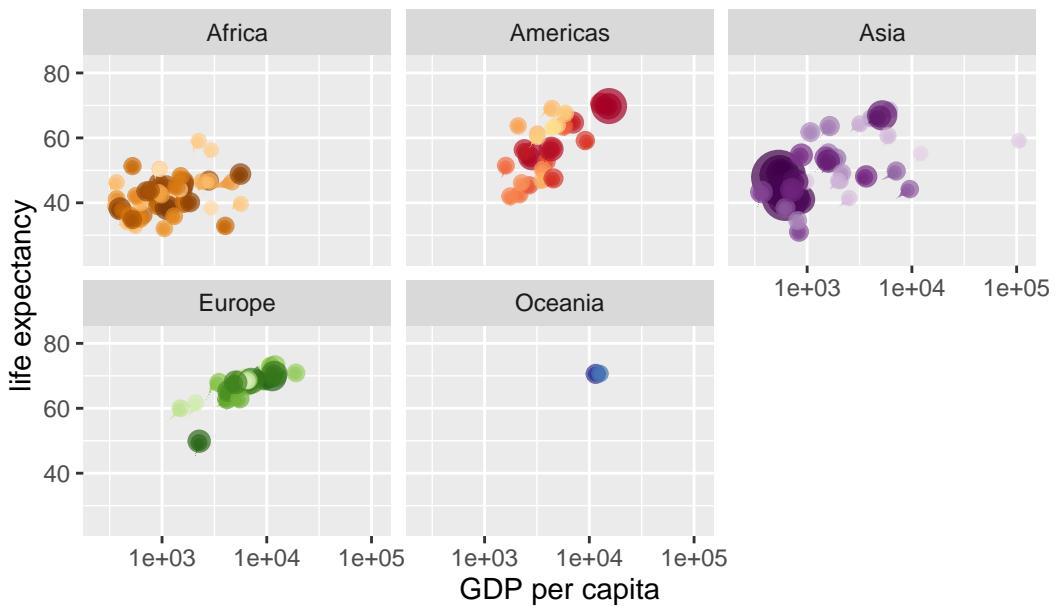
Year: 1958



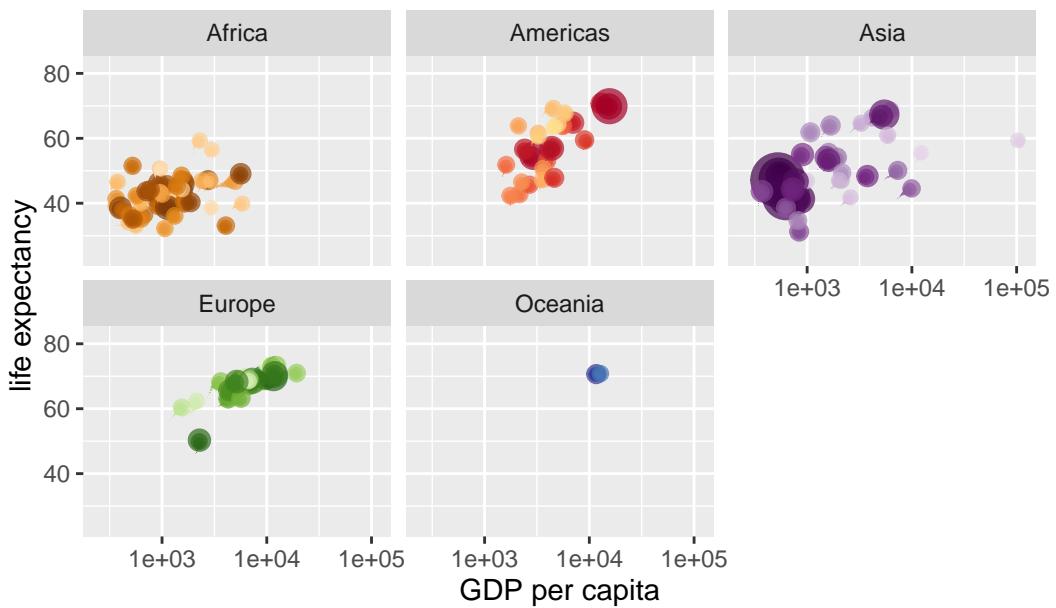
Year: 1959



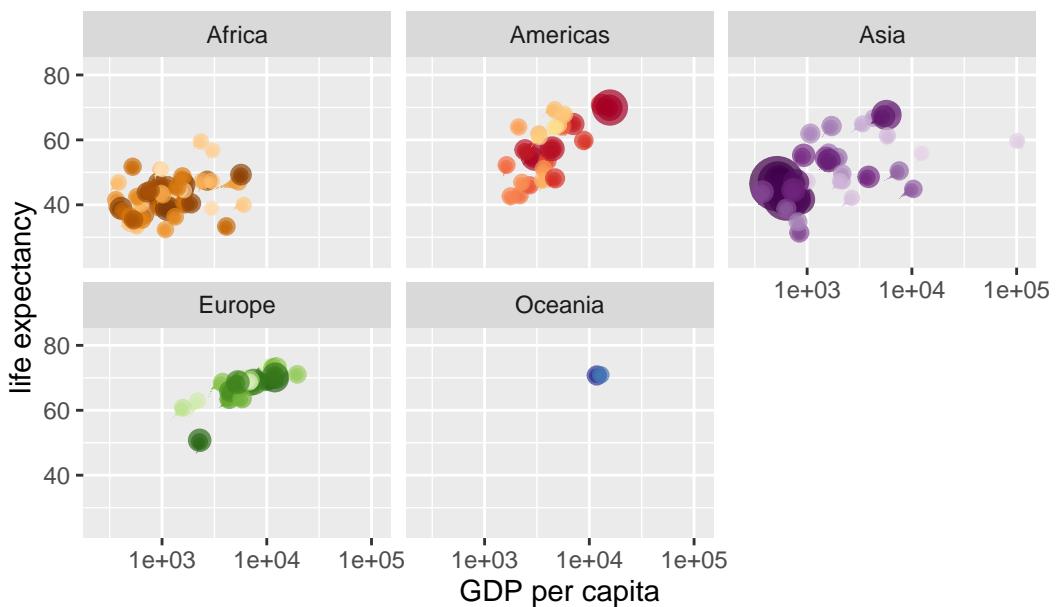
Year: 1959



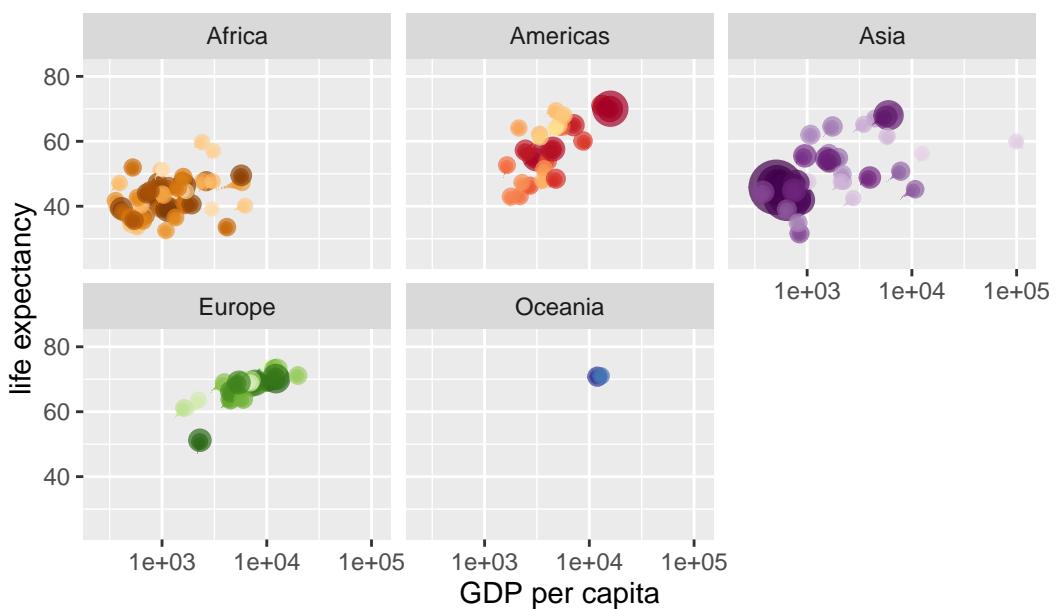
Year: 1960



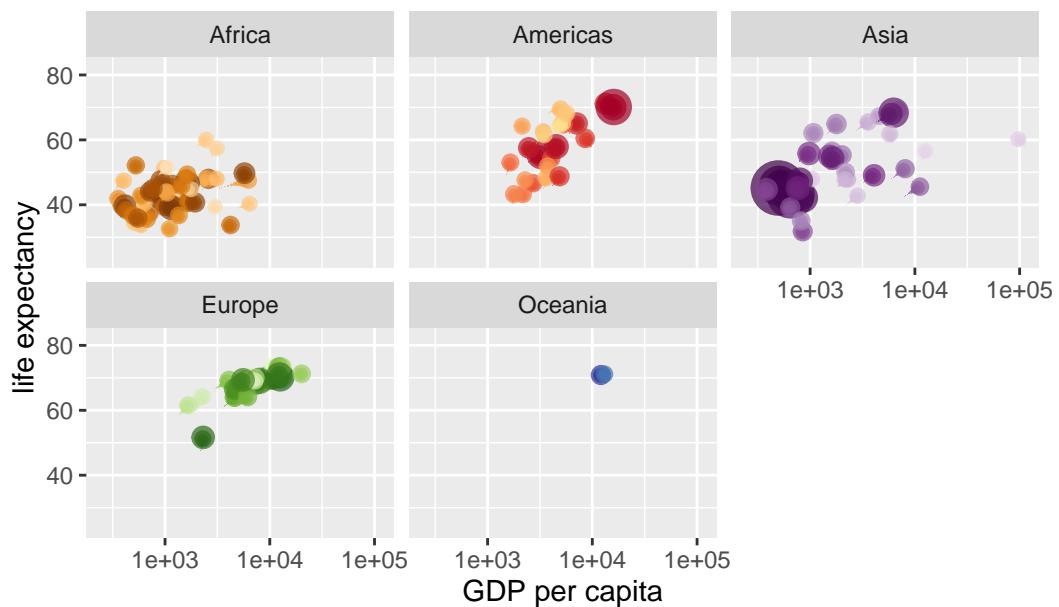
Year: 1960



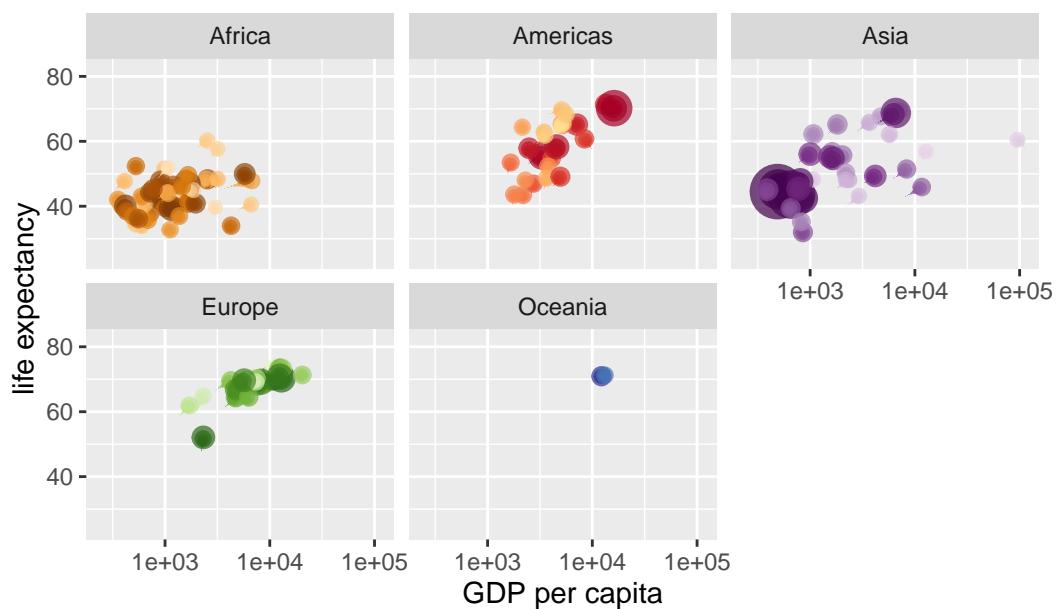
Year: 1961



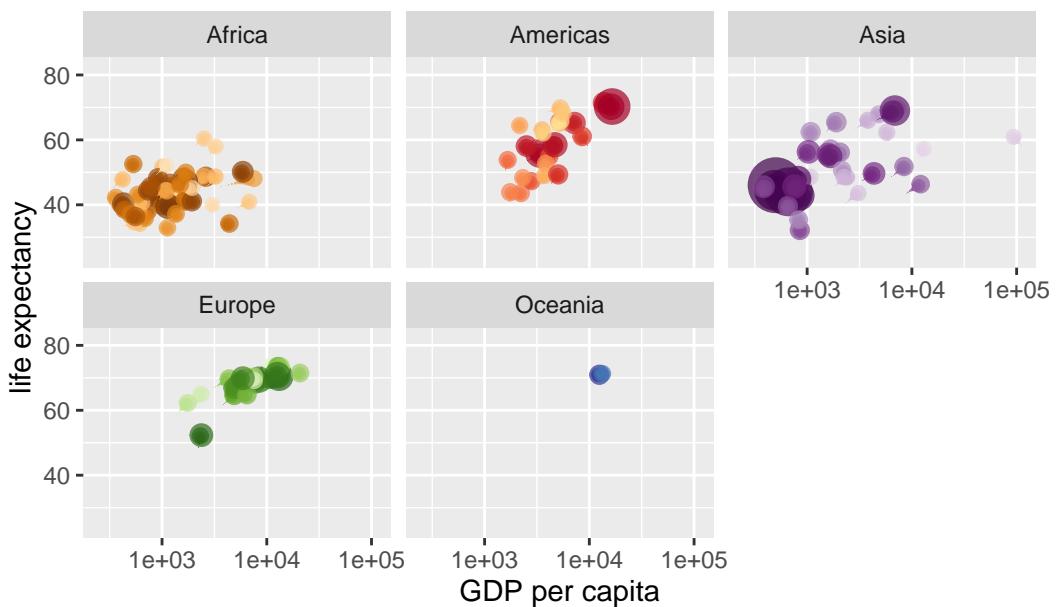
Year: 1961



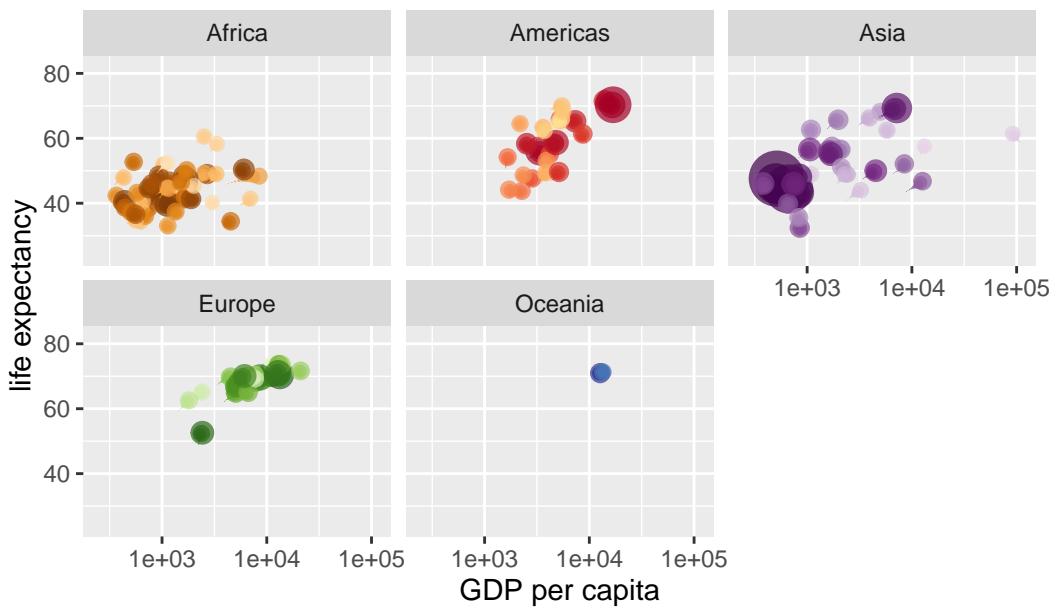
Year: 1962



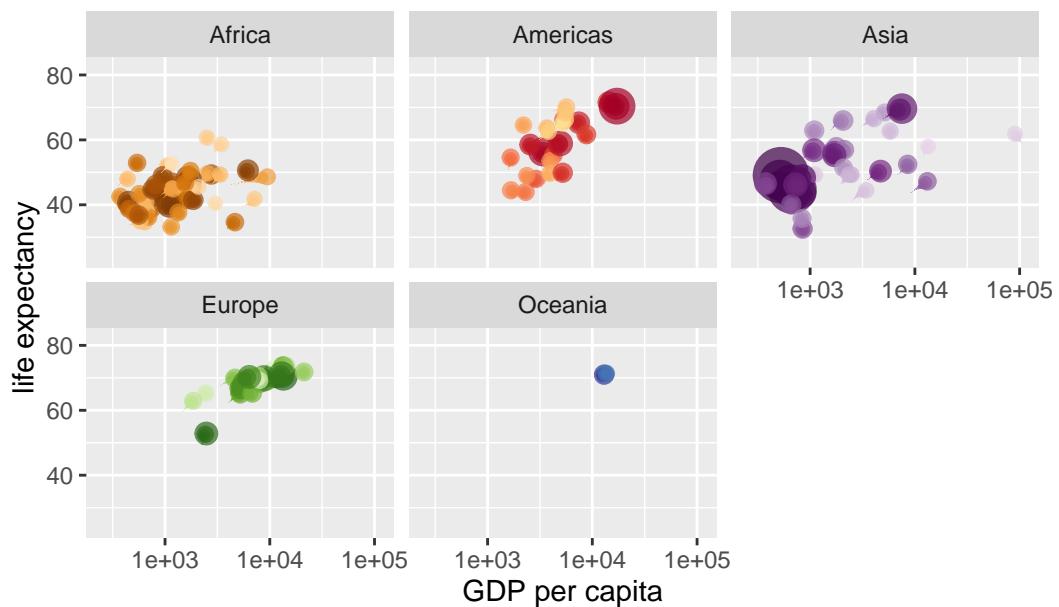
Year: 1963



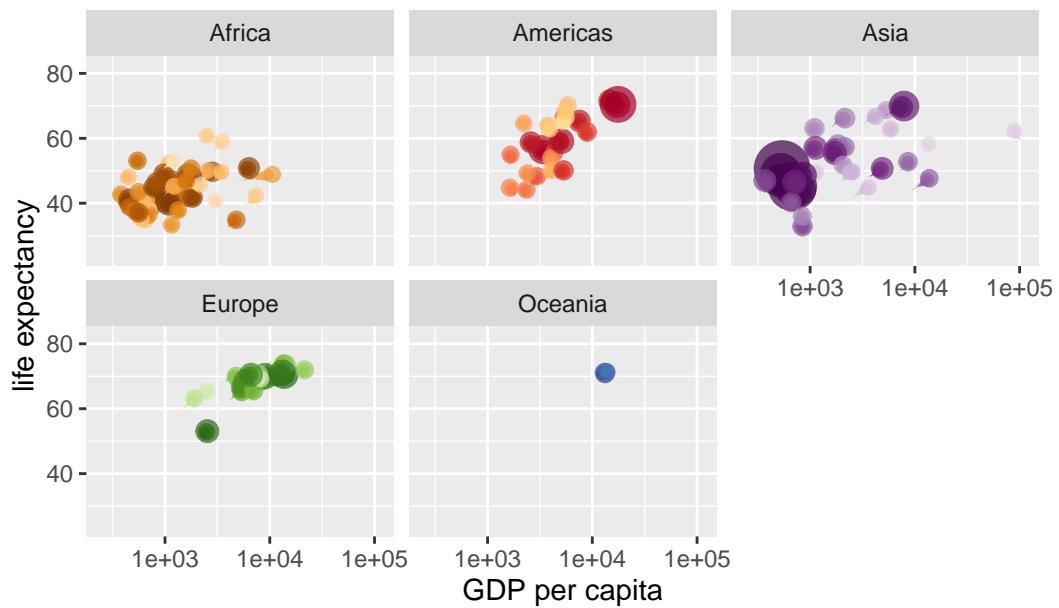
Year: 1963



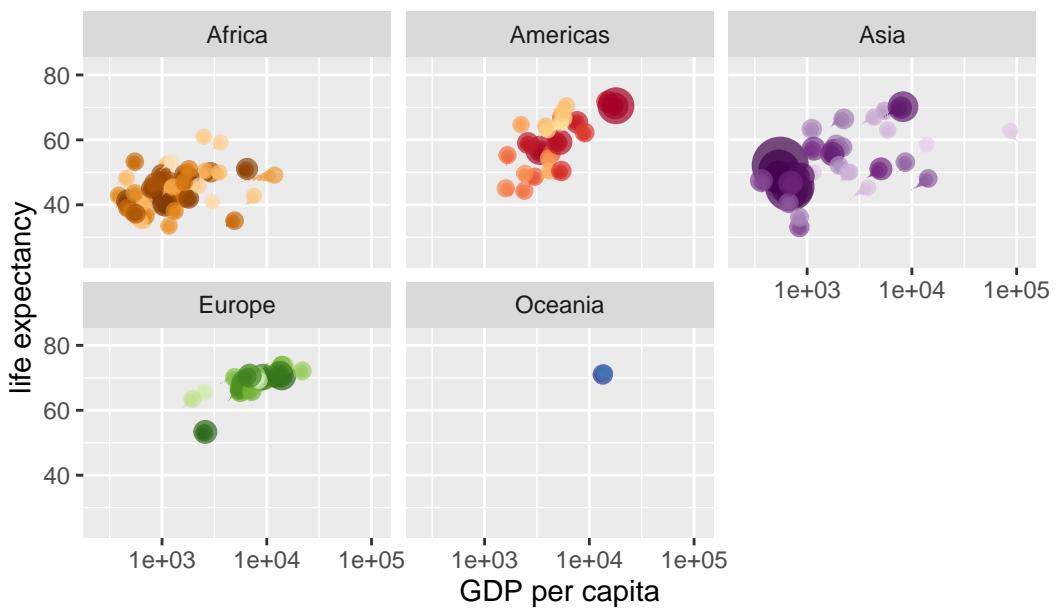
Year: 1964



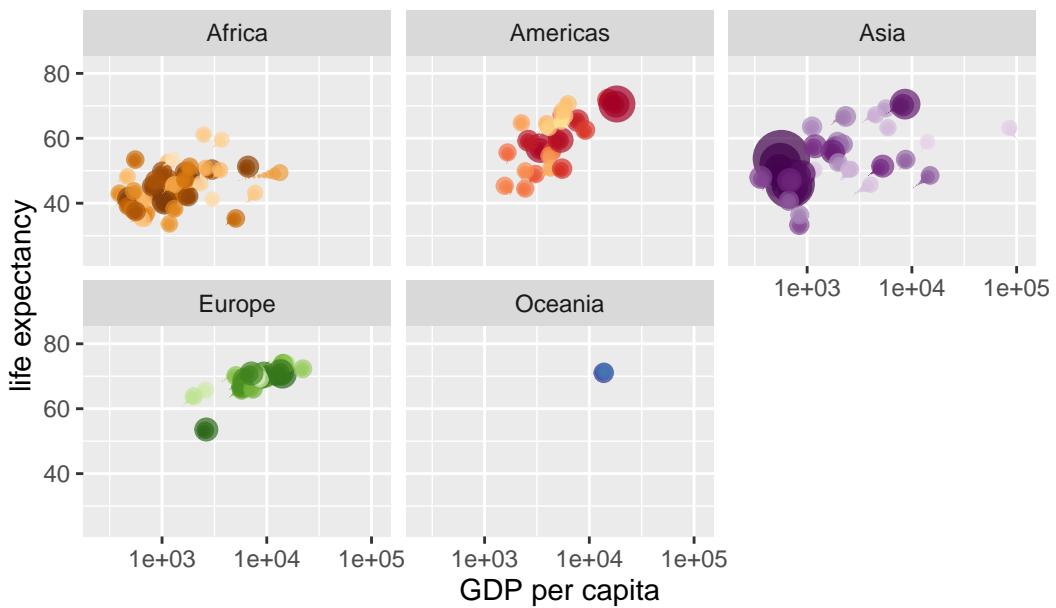
Year: 1964



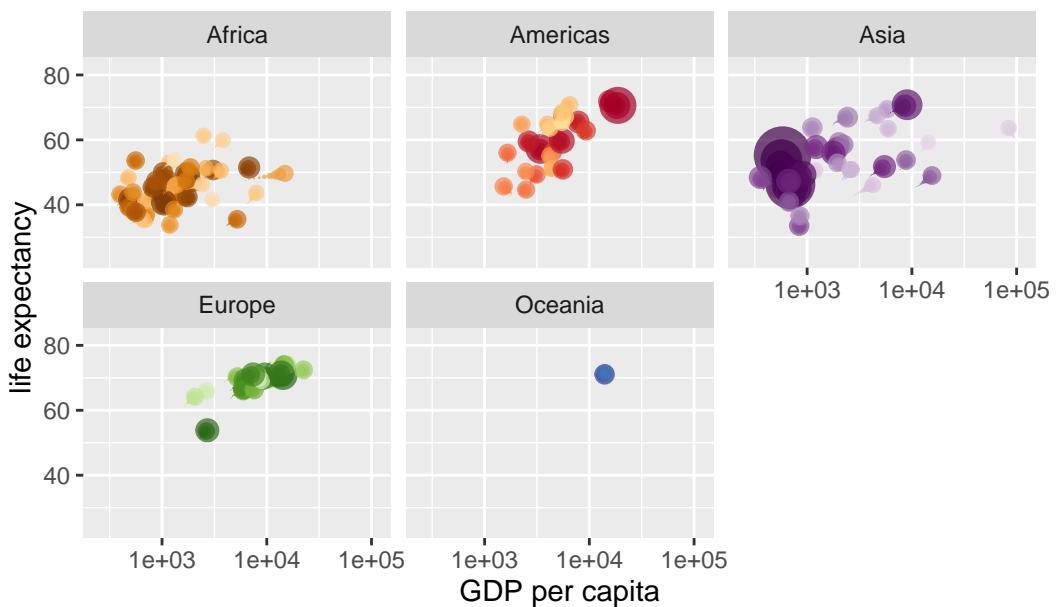
Year: 1965



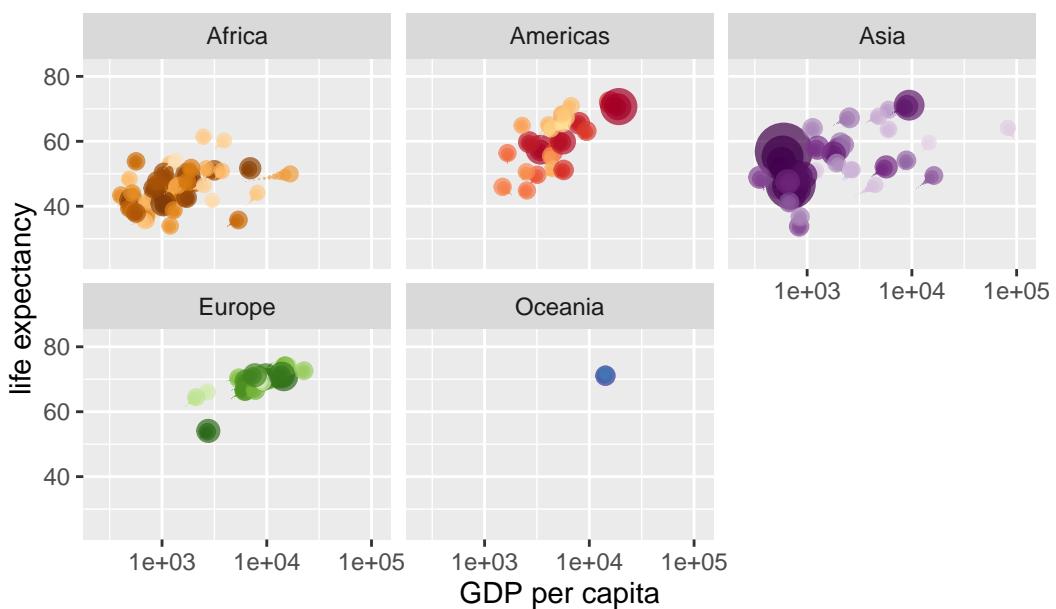
Year: 1965



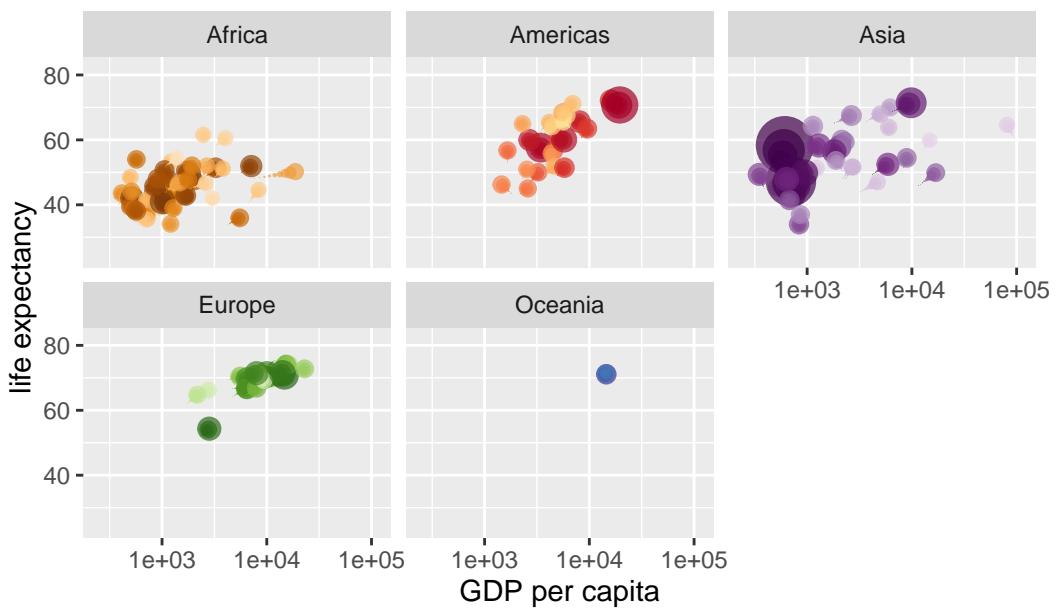
Year: 1966



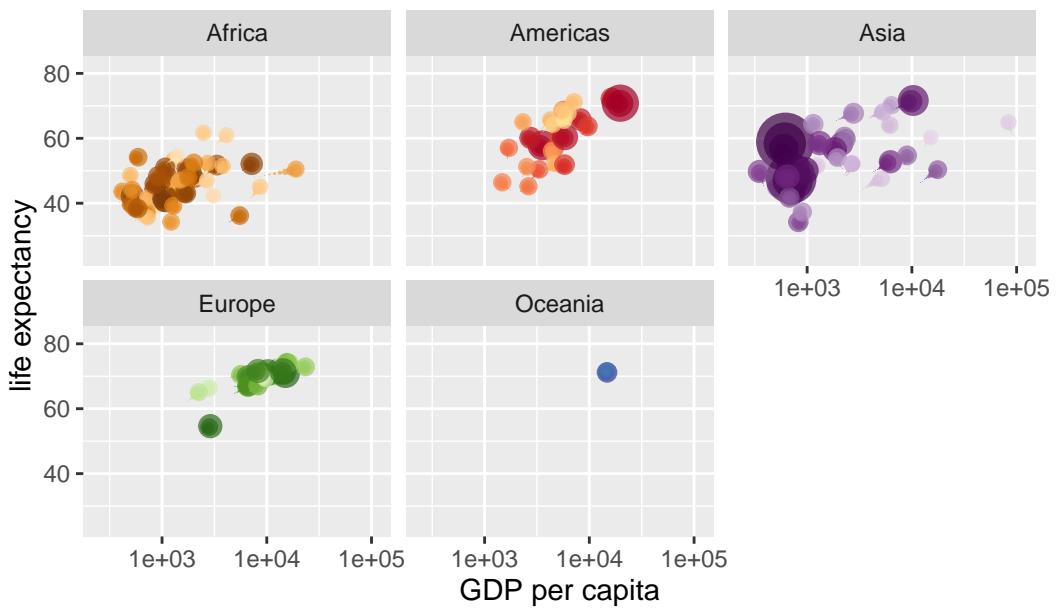
Year: 1966



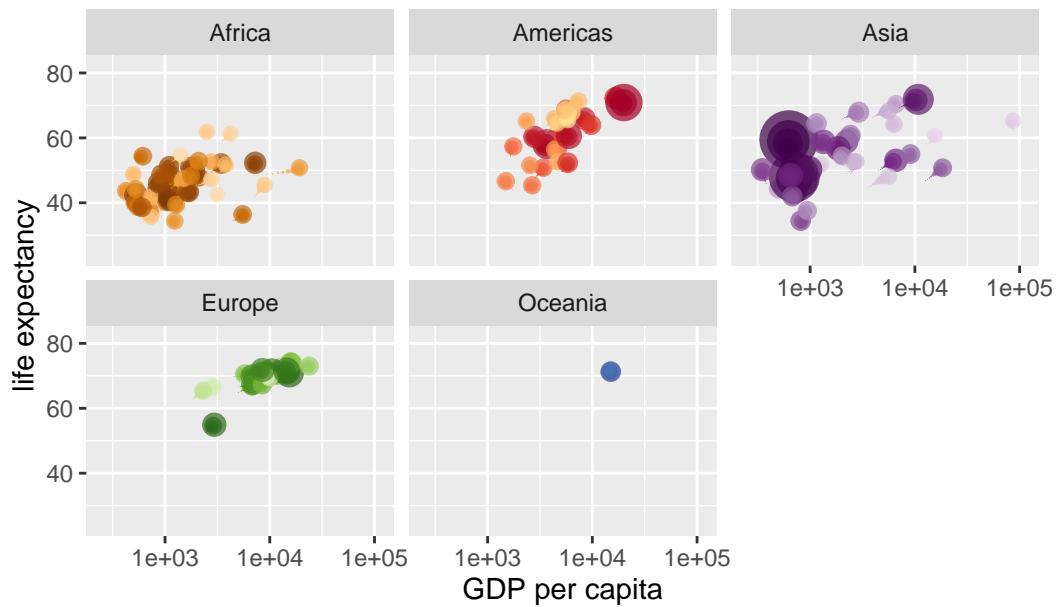
Year: 1967



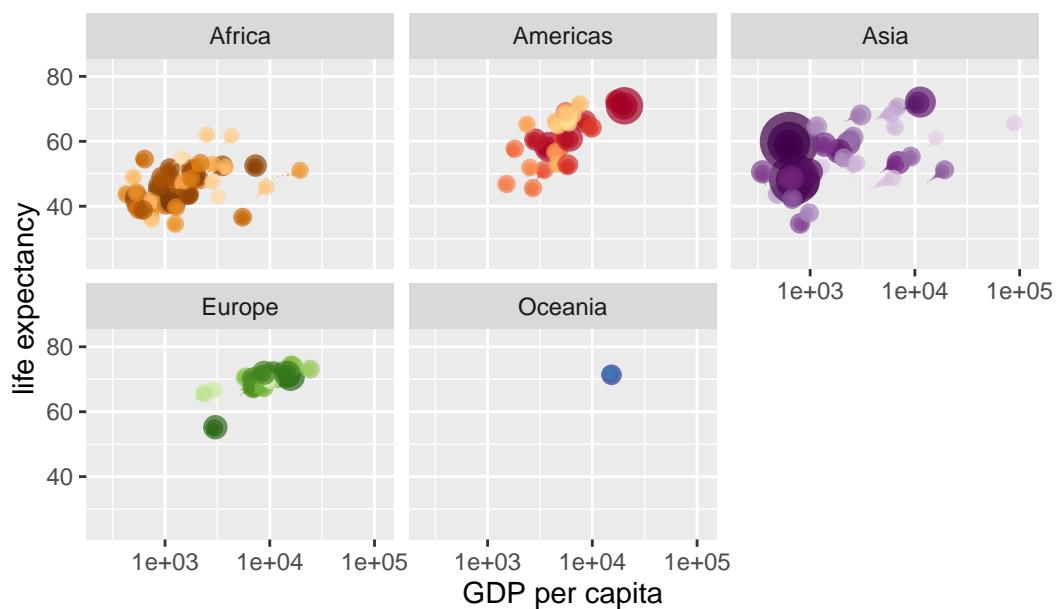
Year: 1968



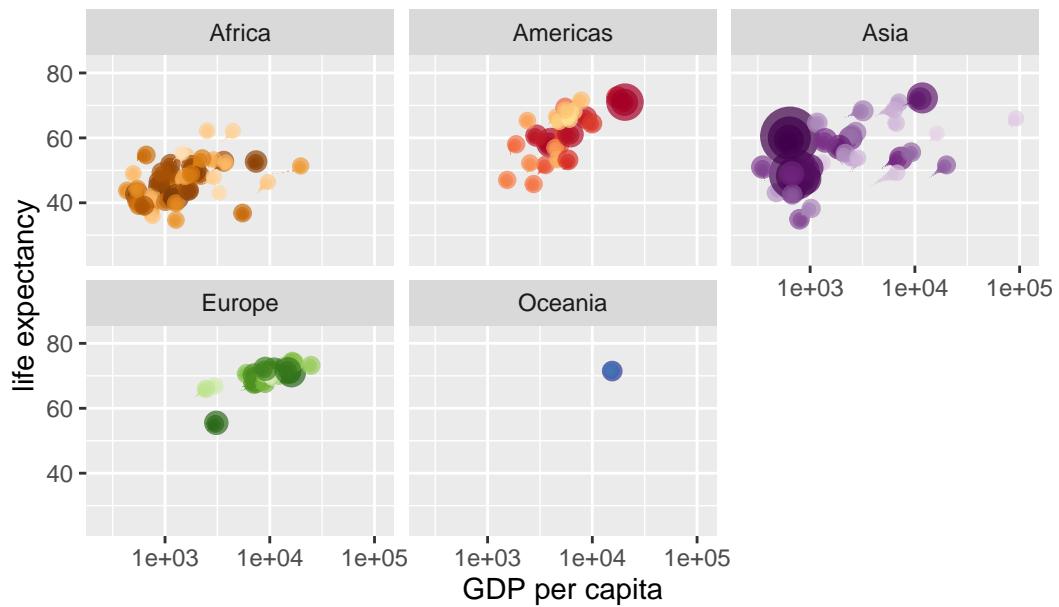
Year: 1968



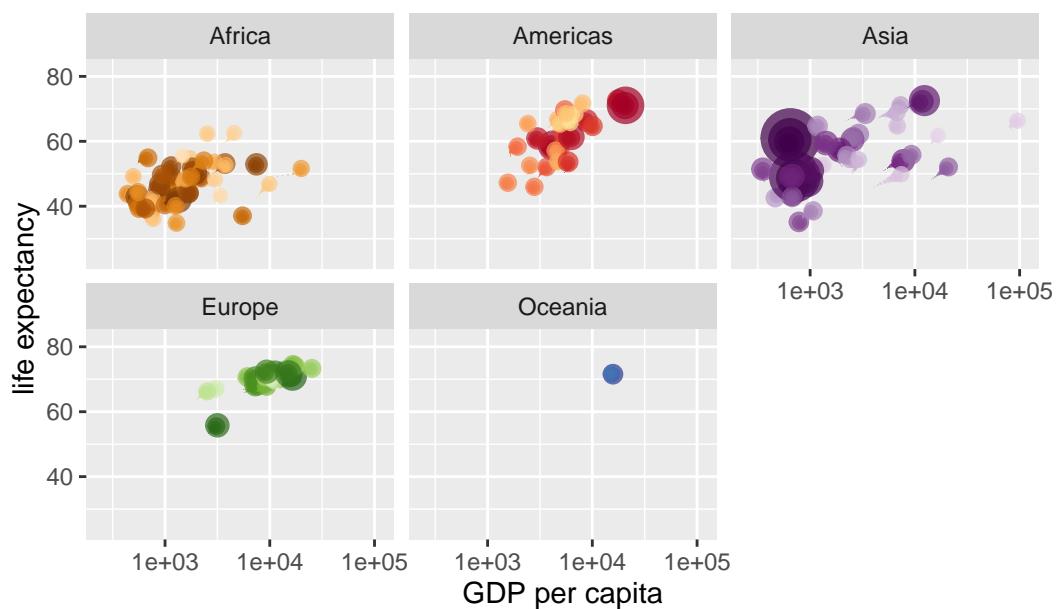
Year: 1969



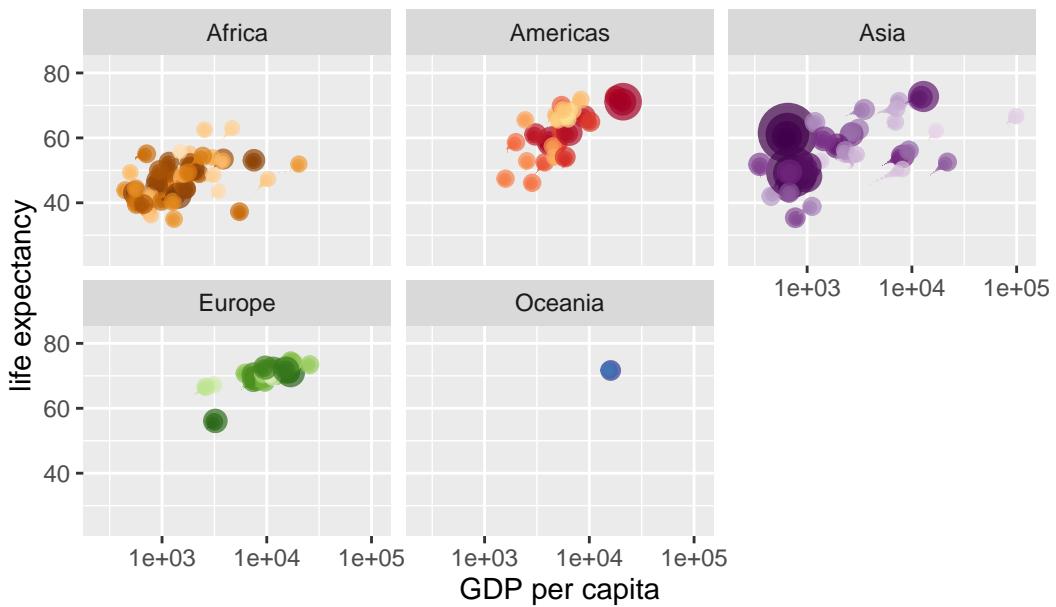
Year: 1969



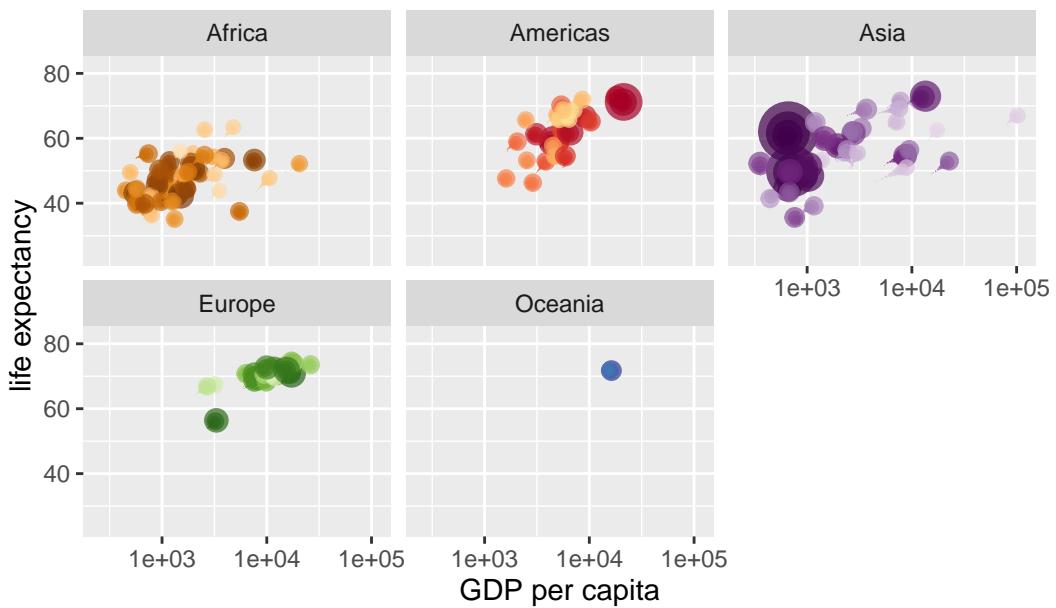
Year: 1970



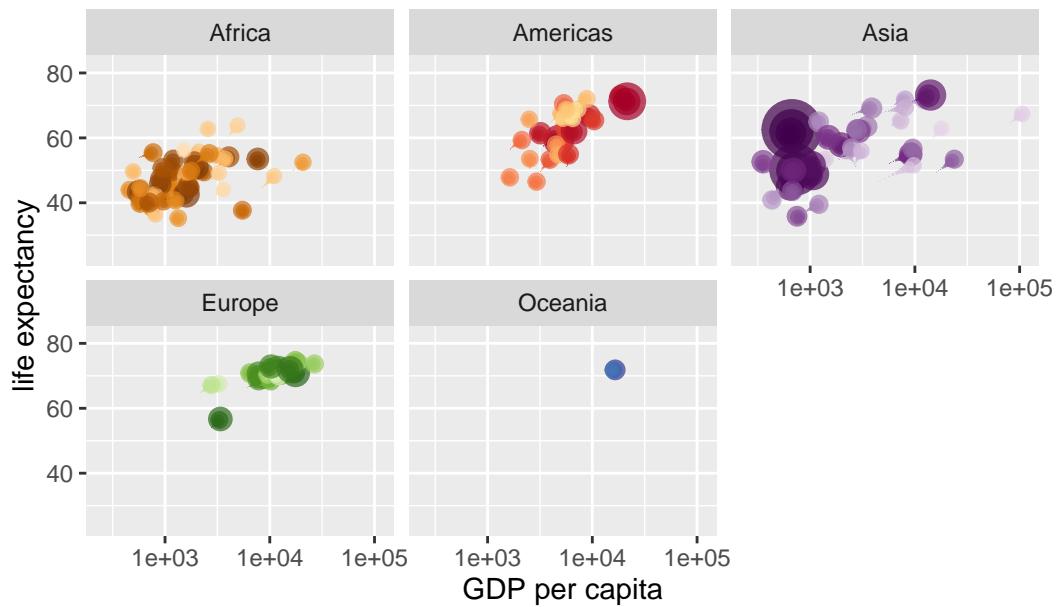
Year: 1970



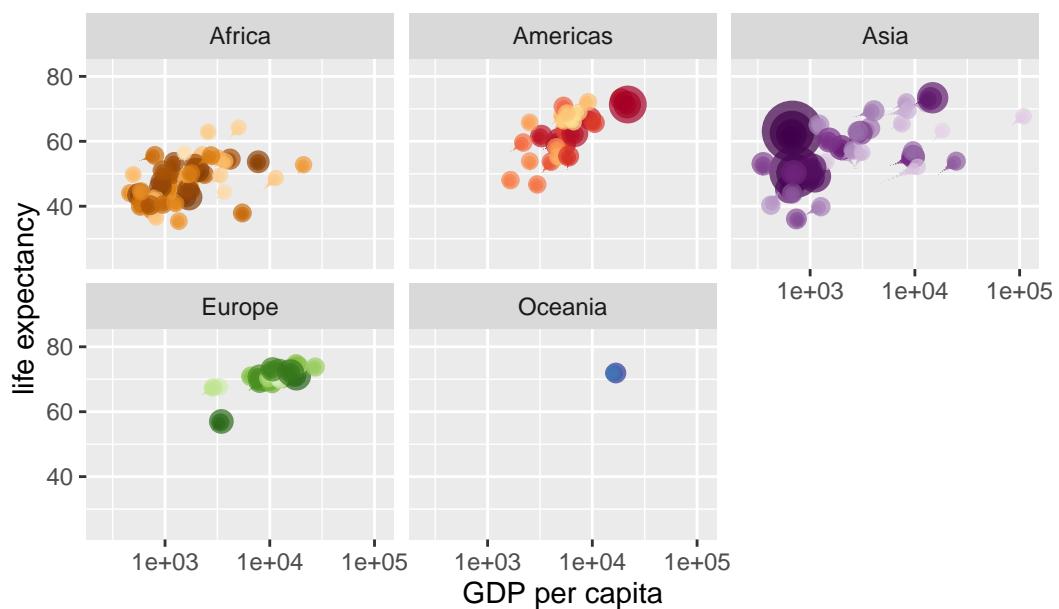
Year: 1971



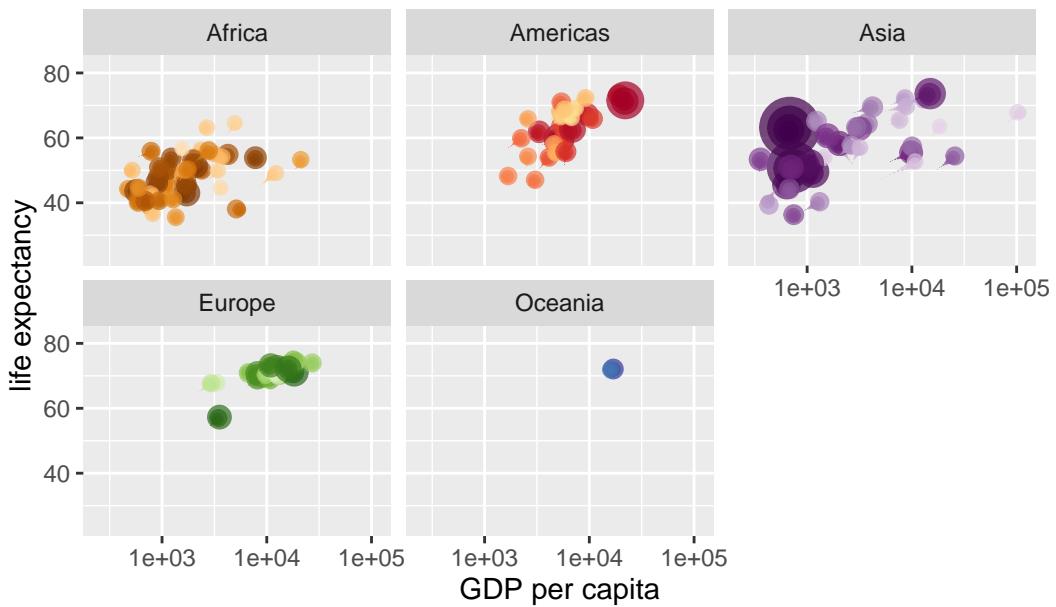
Year: 1971



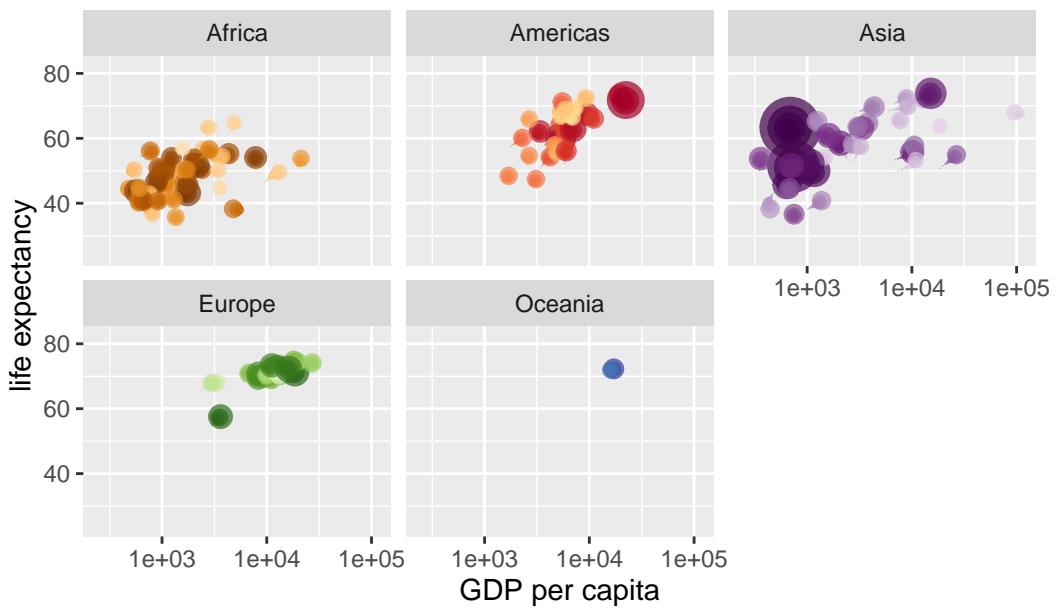
Year: 1972



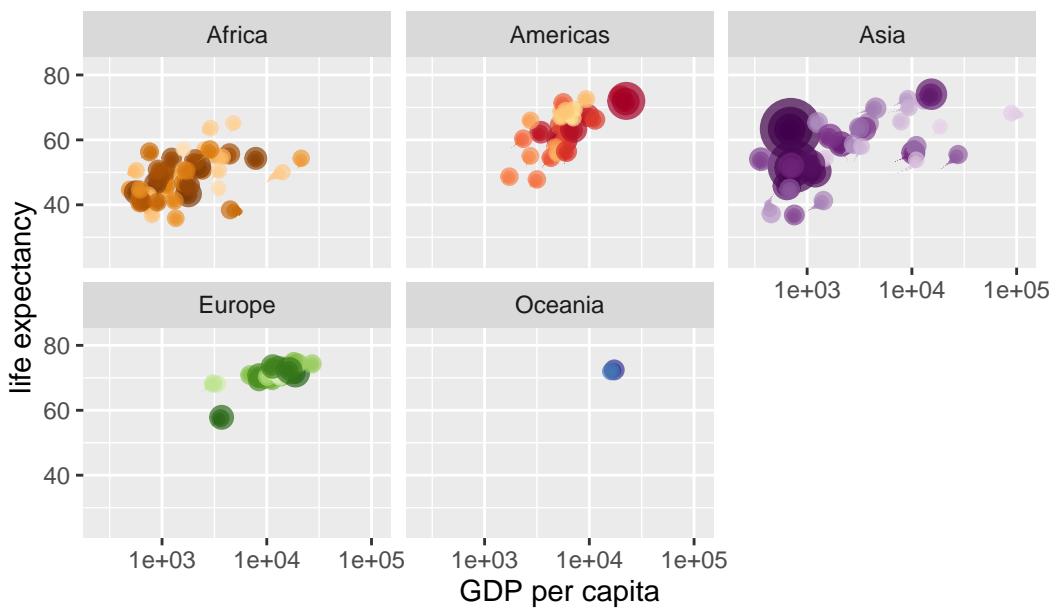
Year: 1973



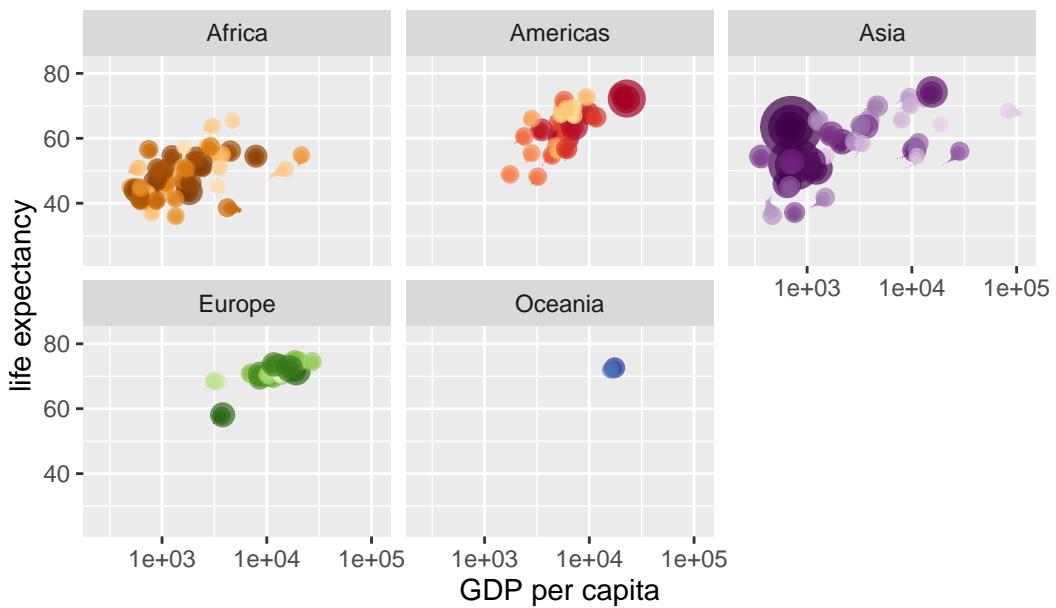
Year: 1973



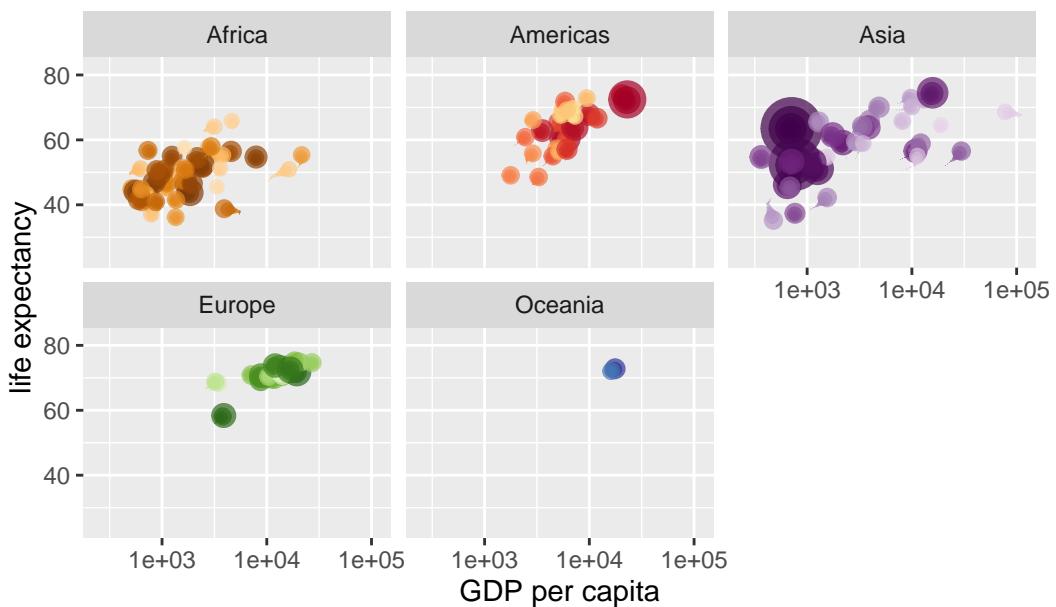
Year: 1974



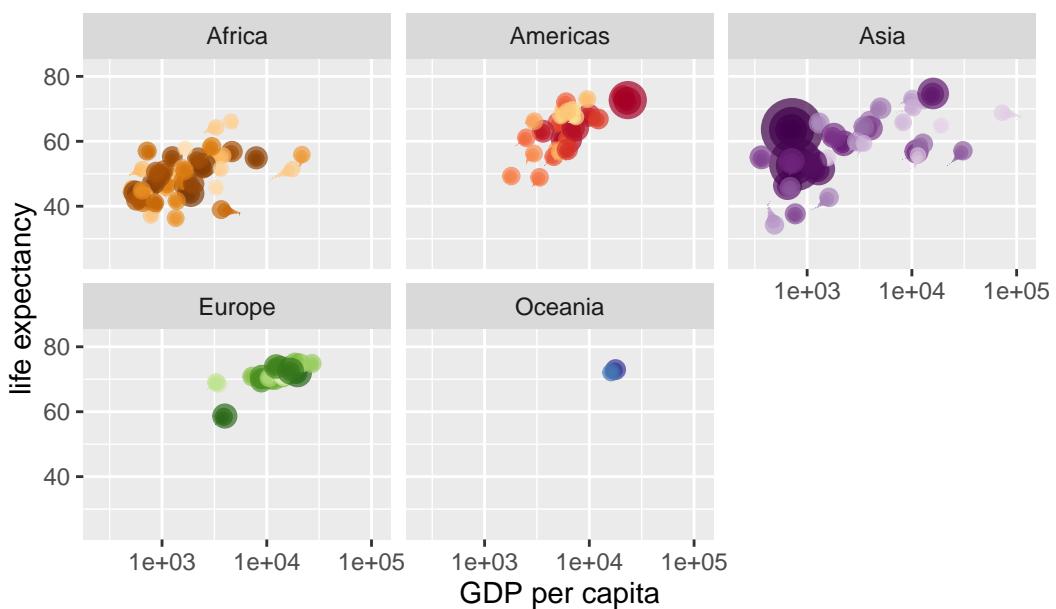
Year: 1974



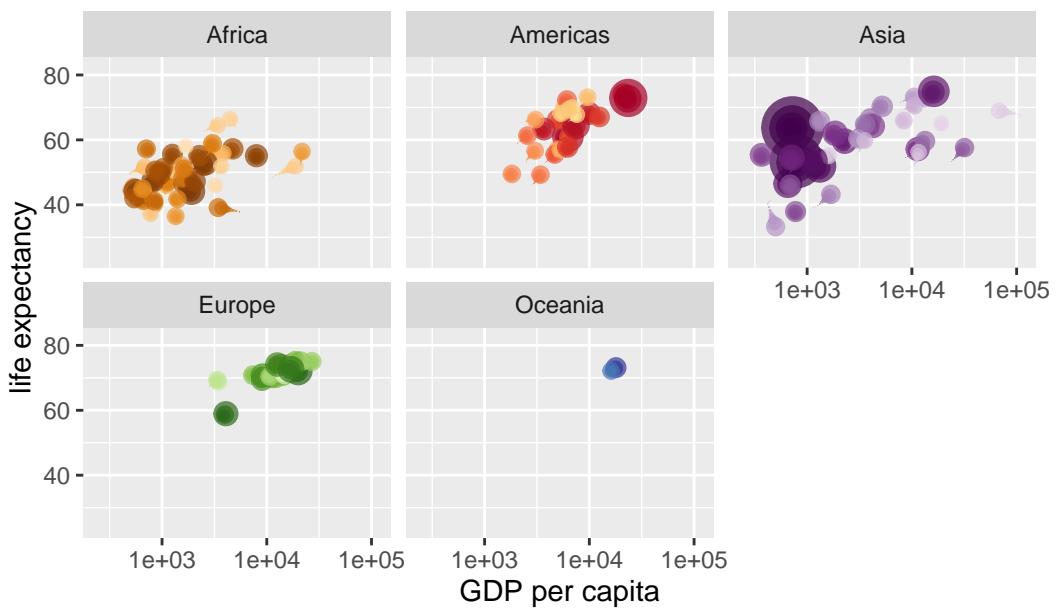
Year: 1975



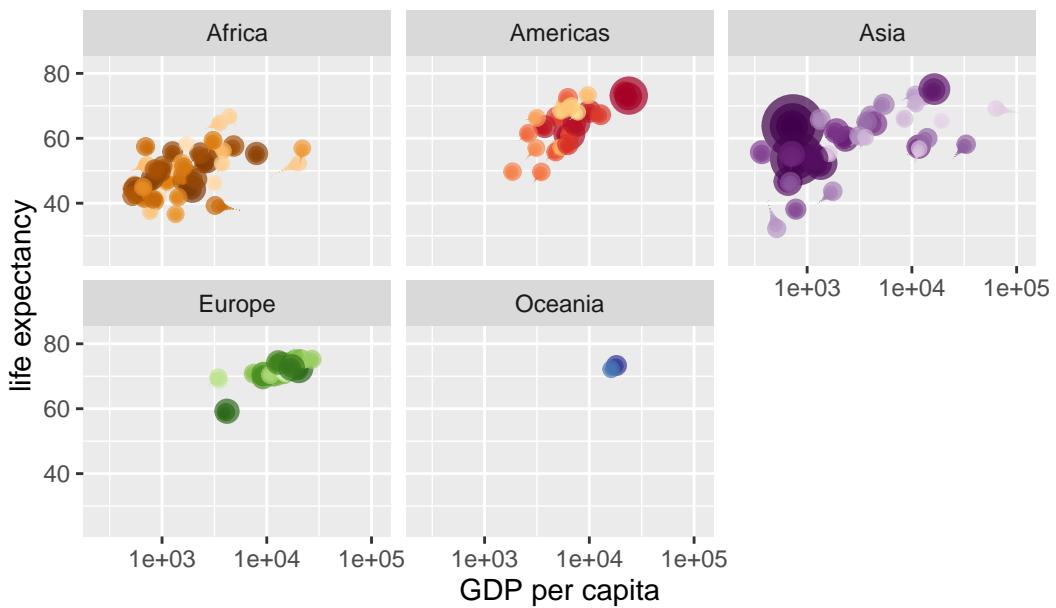
Year: 1975



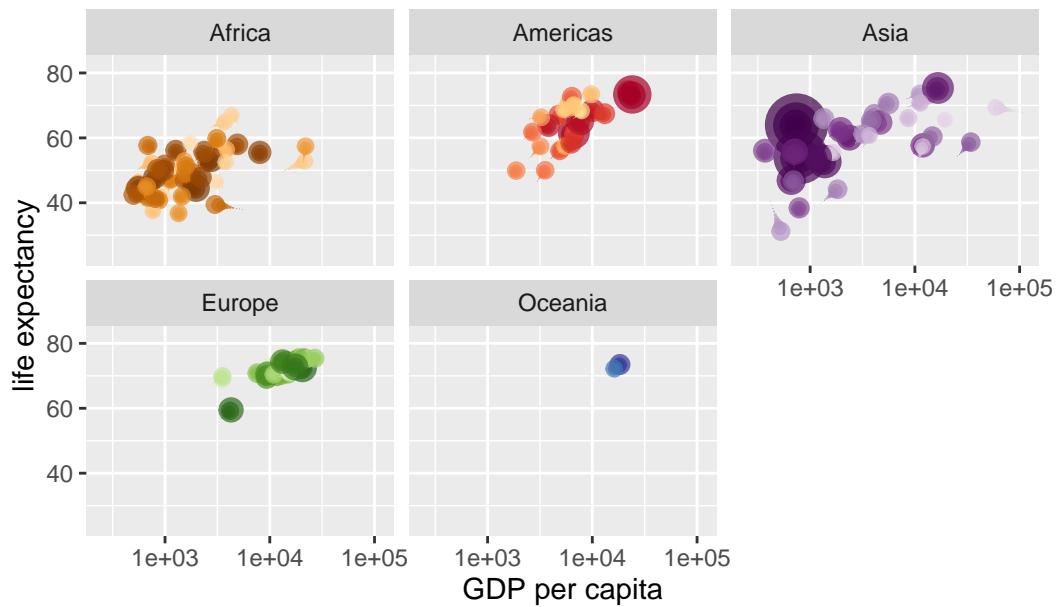
Year: 1976



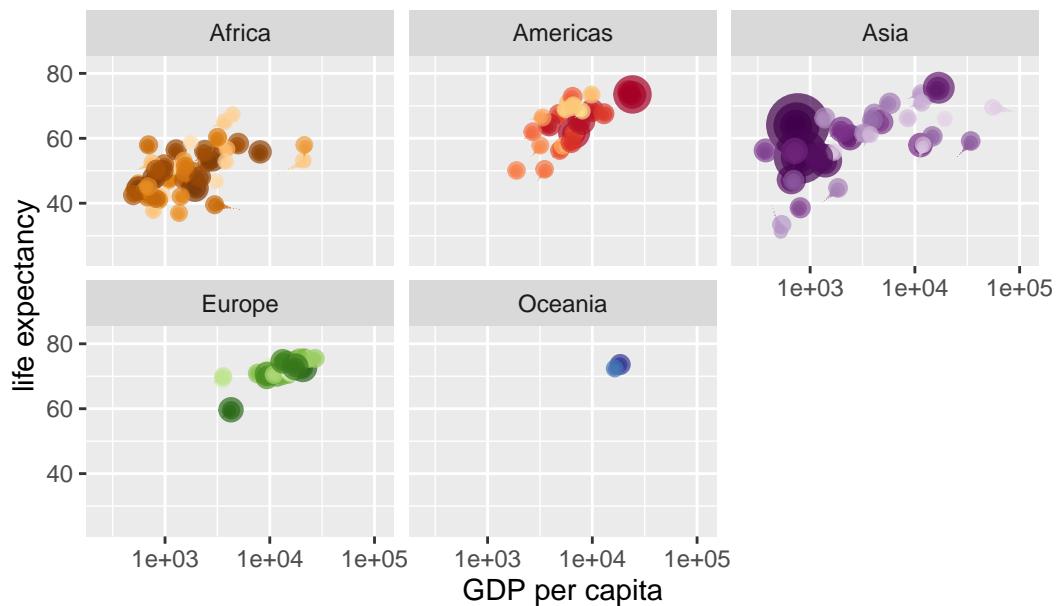
Year: 1976



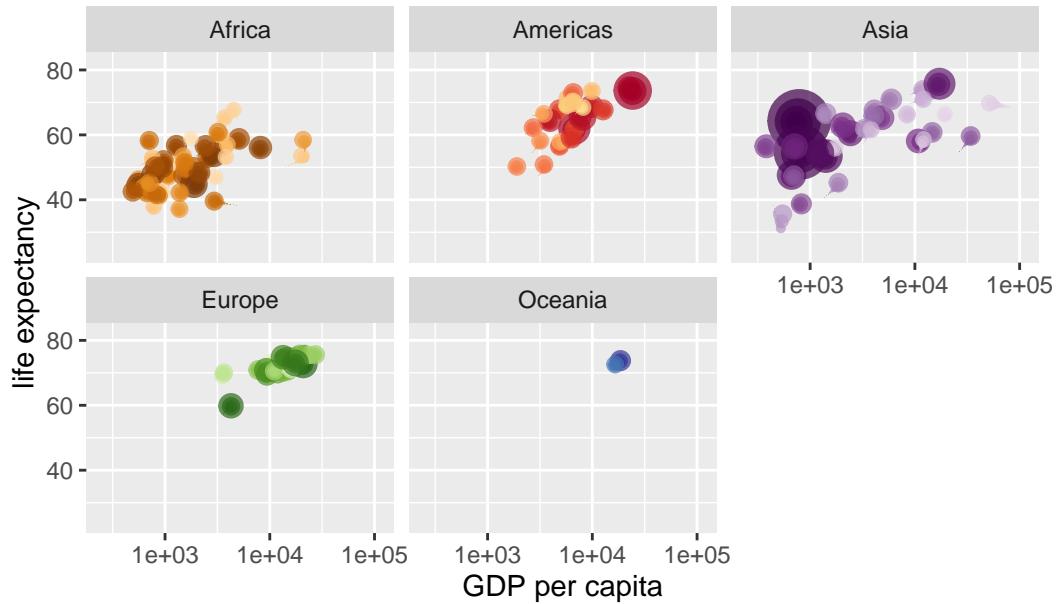
Year: 1977



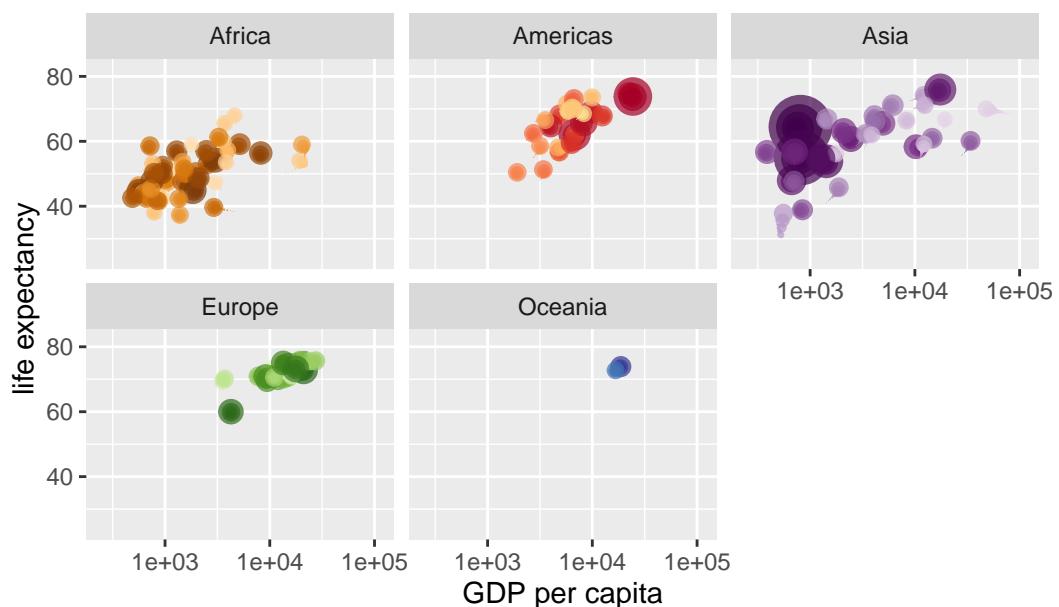
Year: 1978



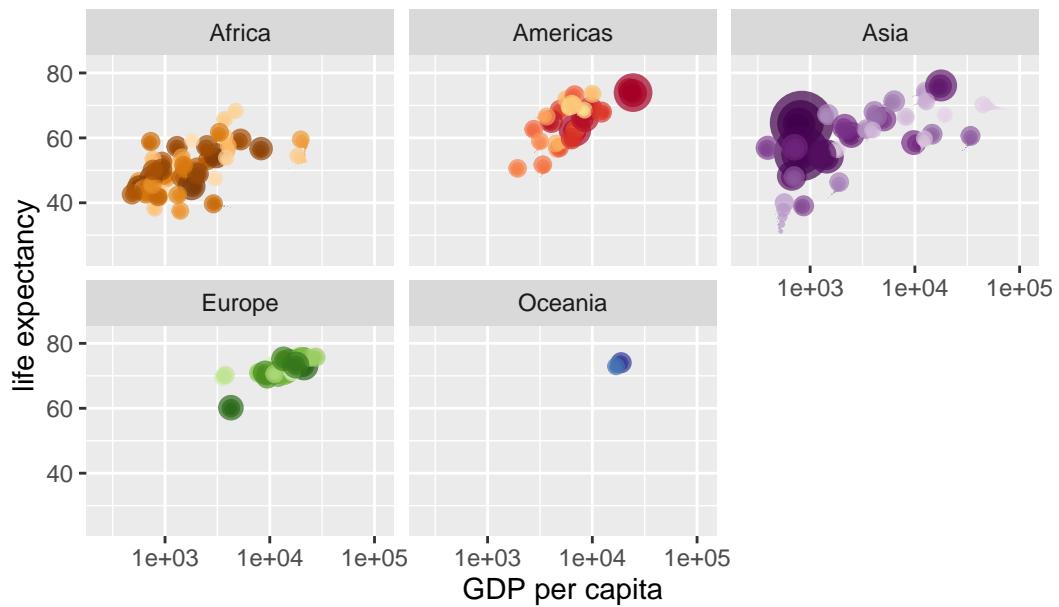
Year: 1978



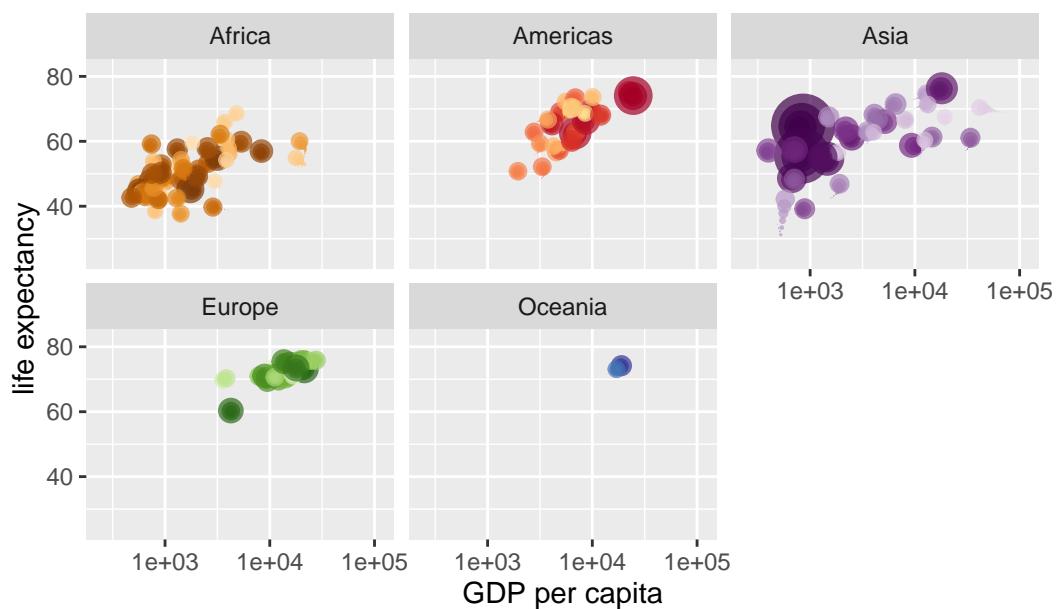
Year: 1979



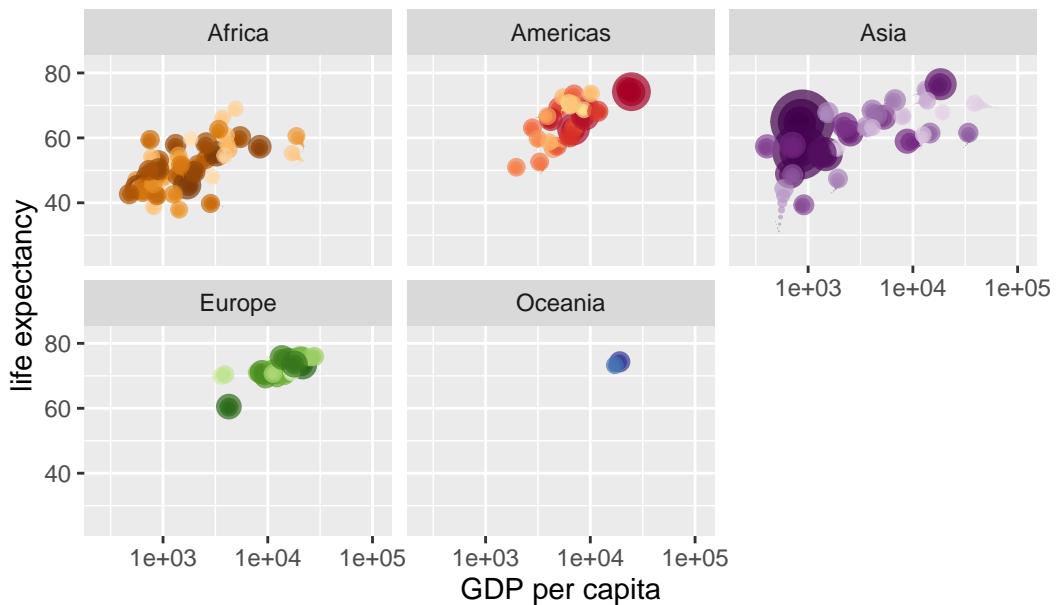
Year: 1979



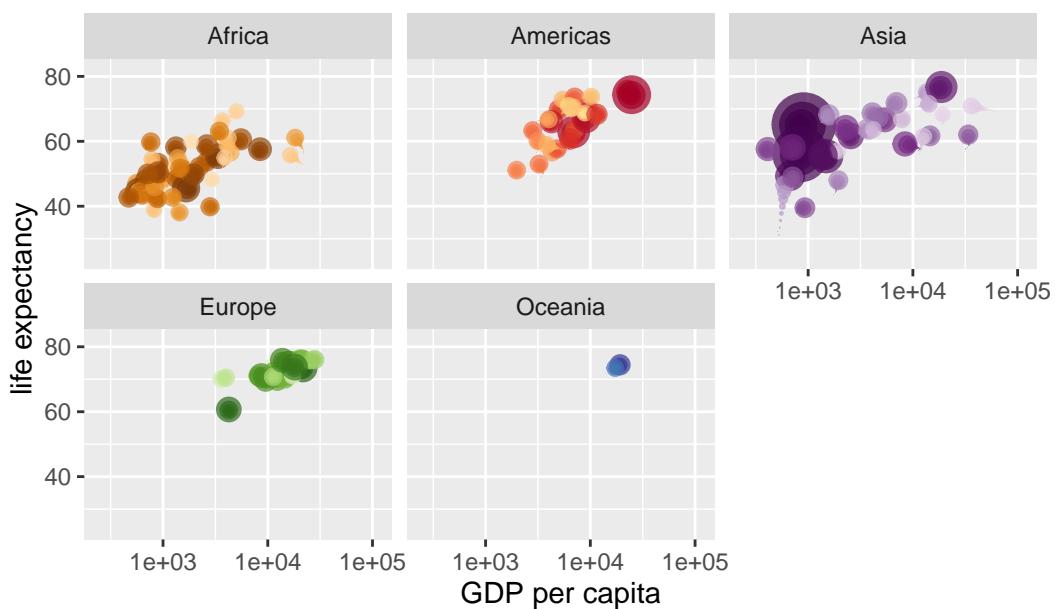
Year: 1980



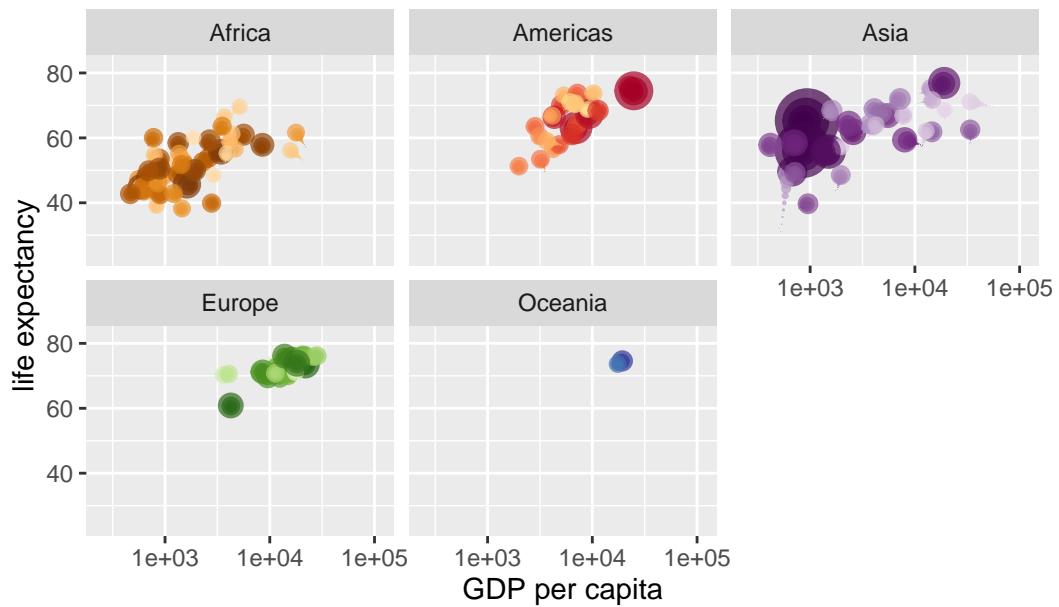
Year: 1980



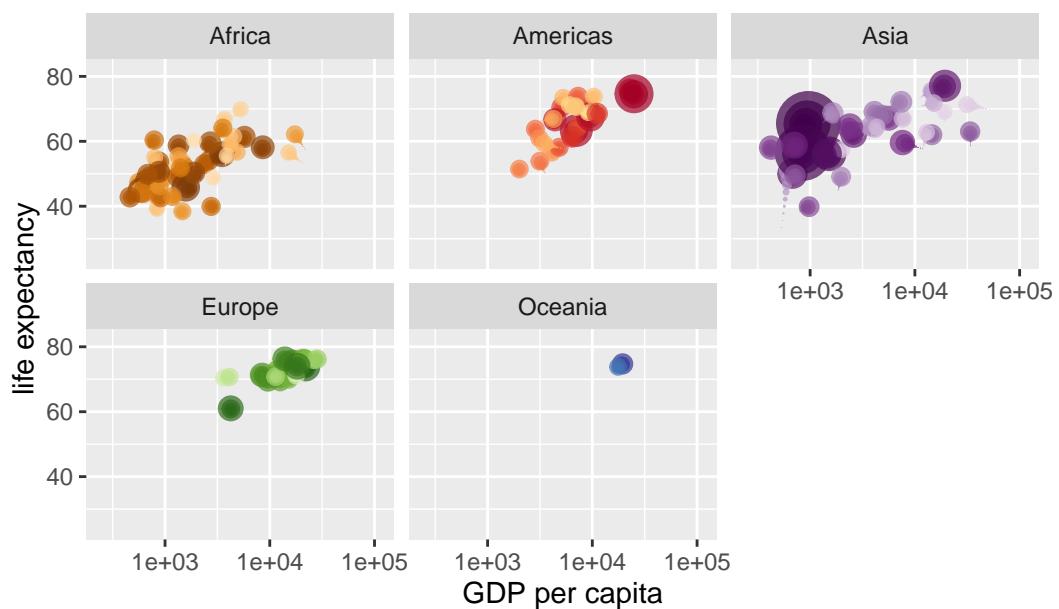
Year: 1981



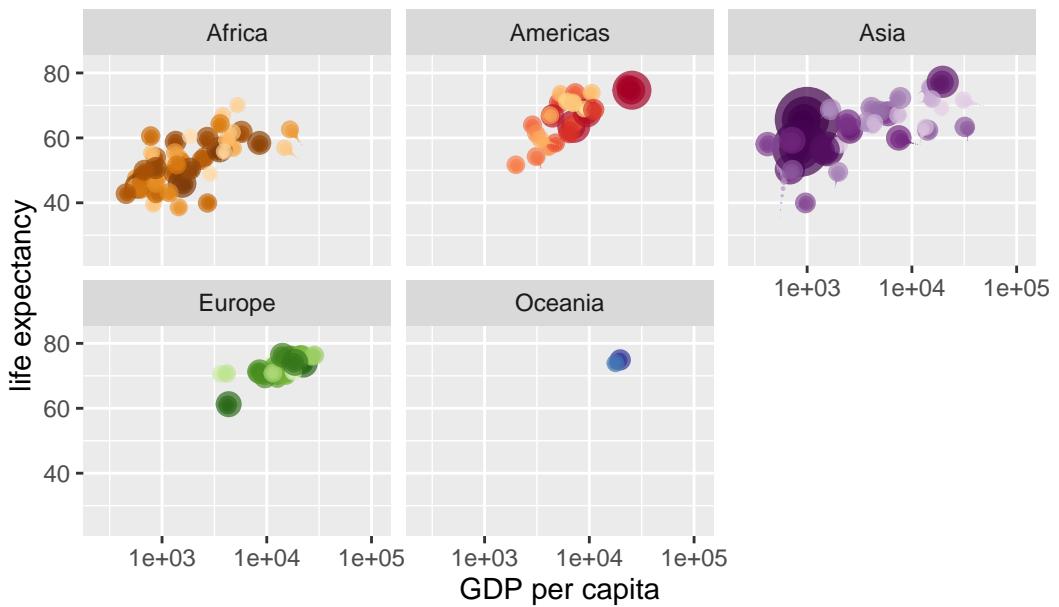
Year: 1981



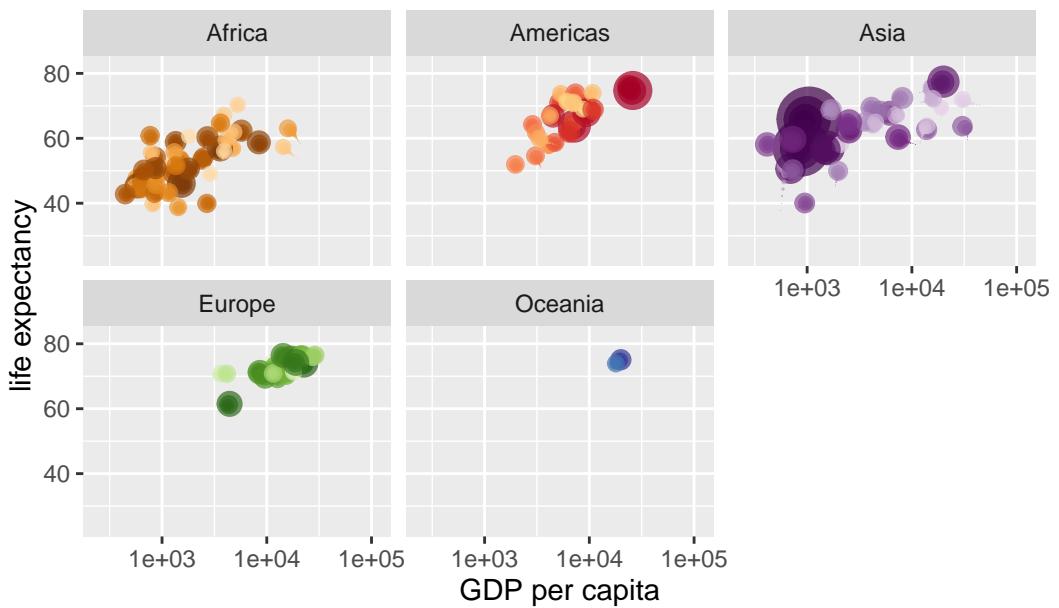
Year: 1982



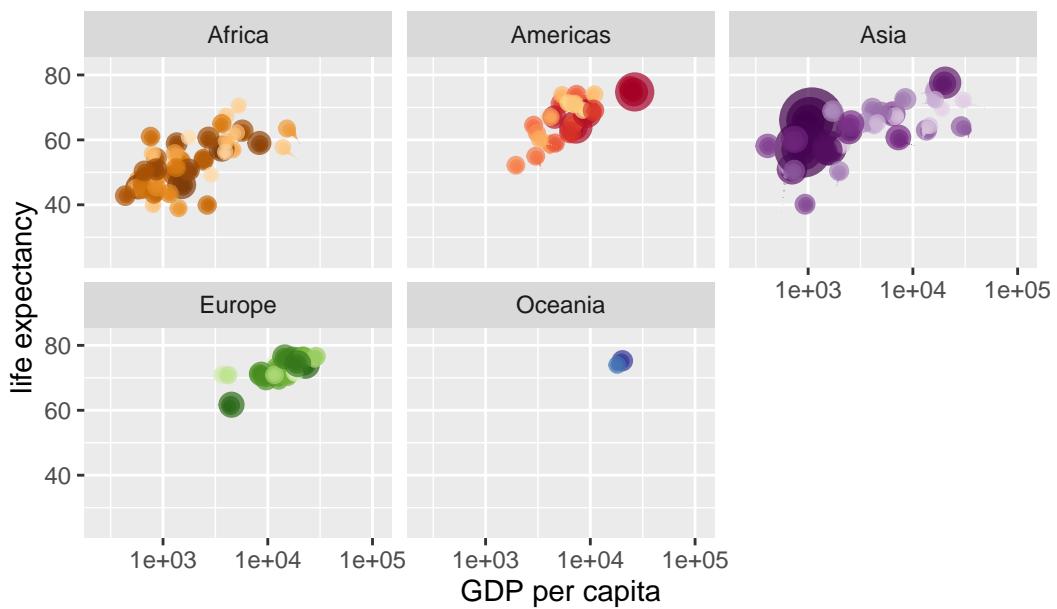
Year: 1983



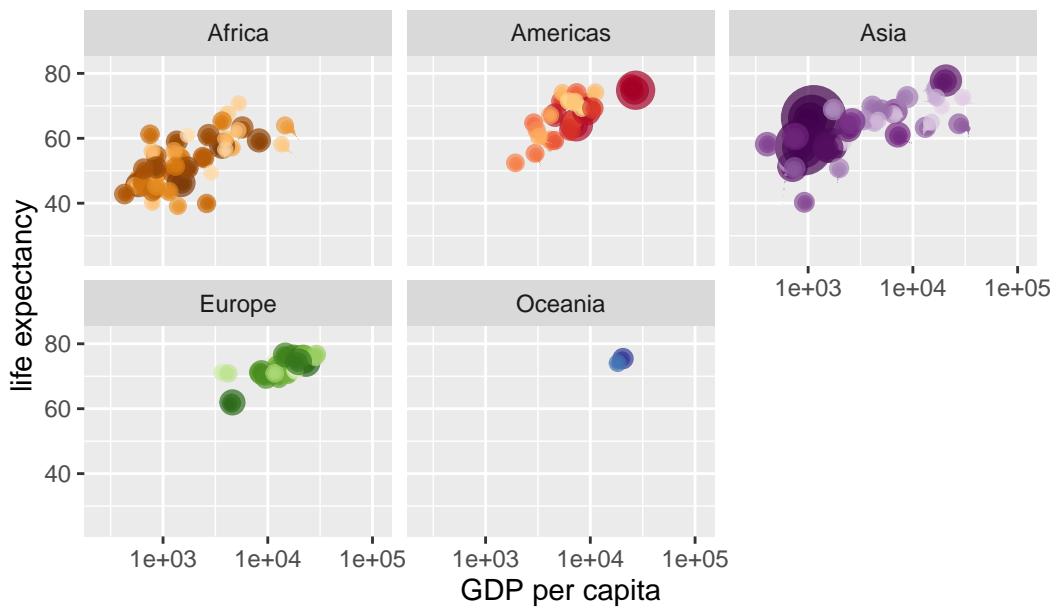
Year: 1983



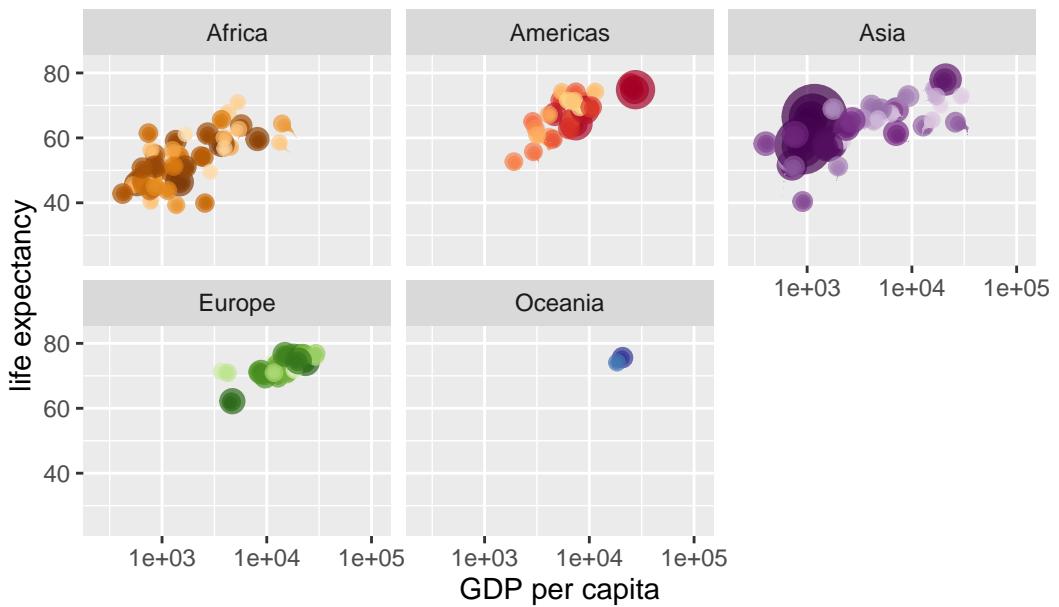
Year: 1984



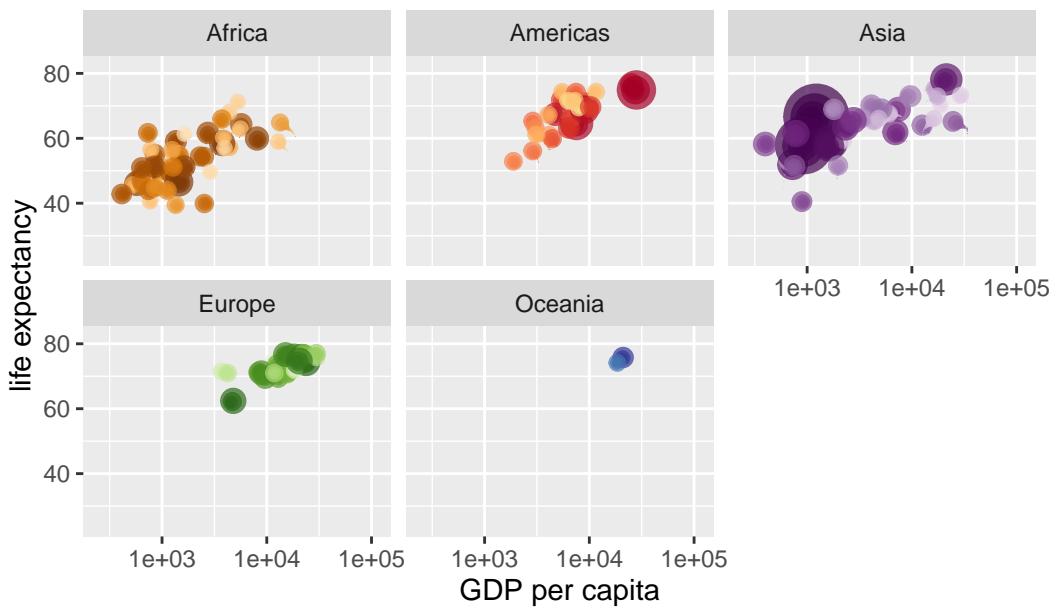
Year: 1984



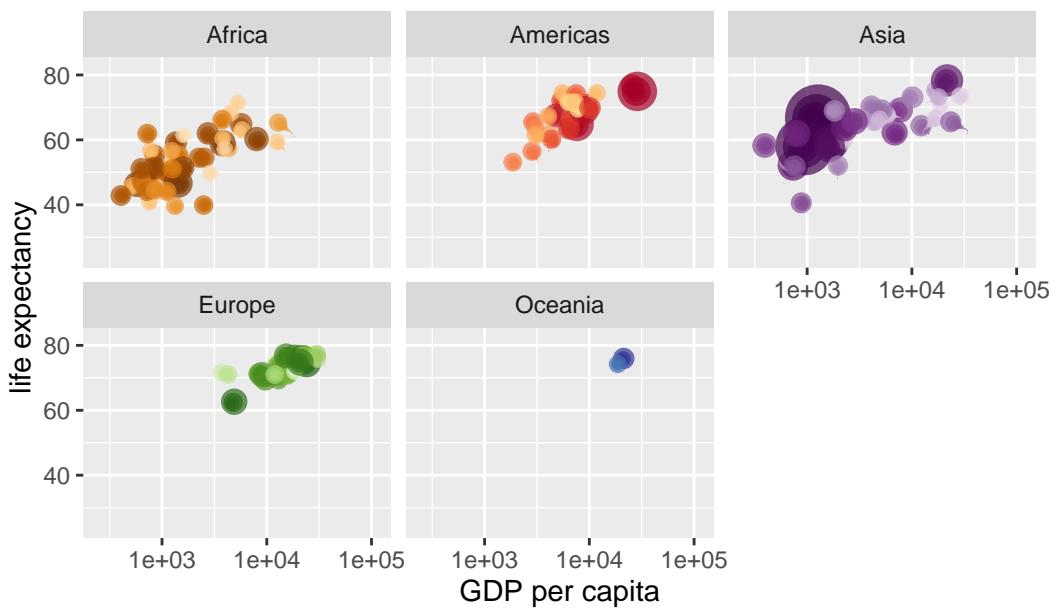
Year: 1985



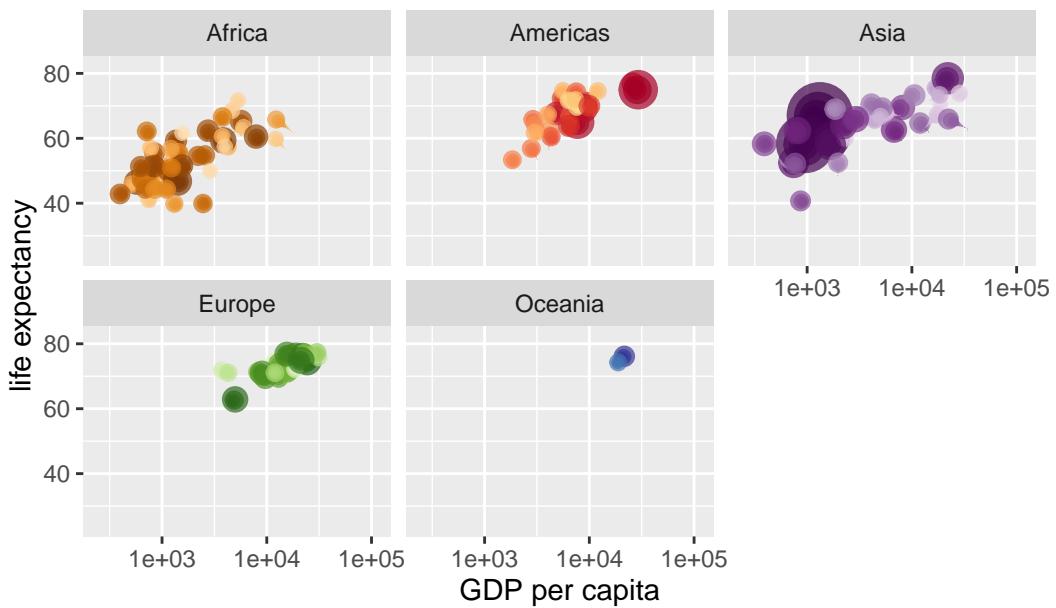
Year: 1985



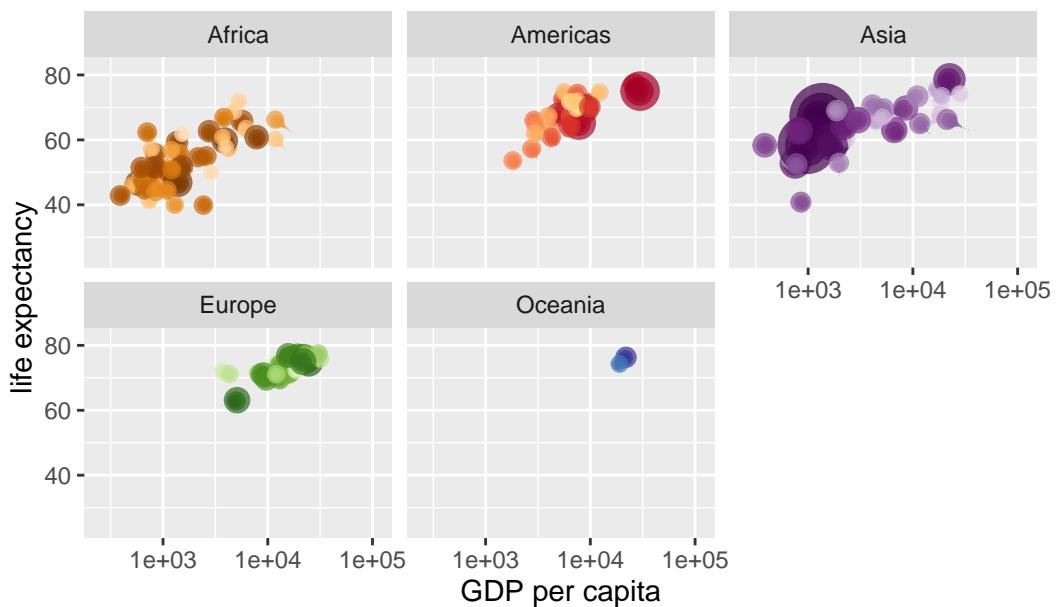
Year: 1986



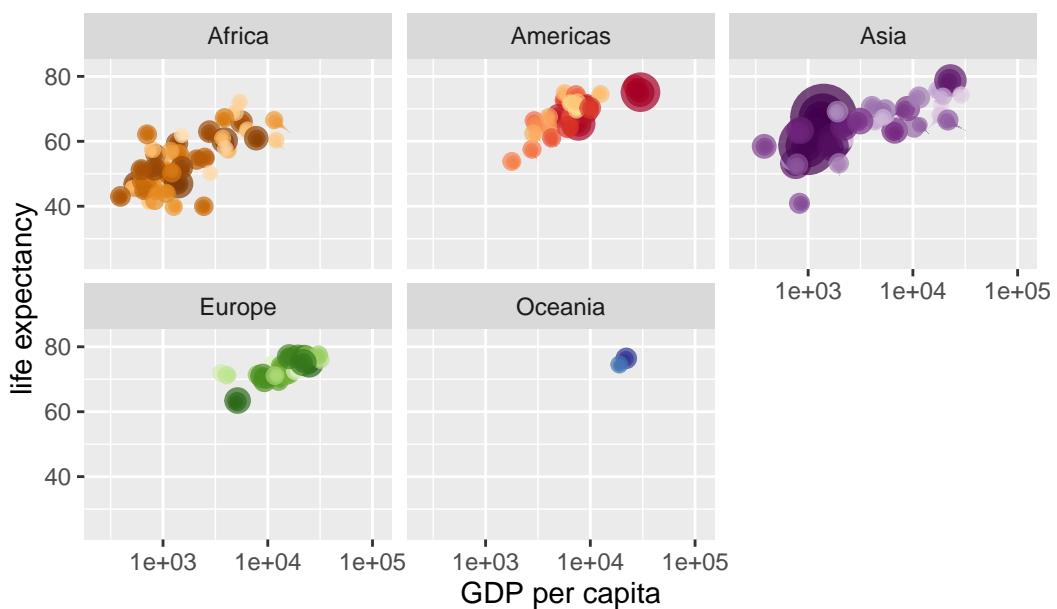
Year: 1986



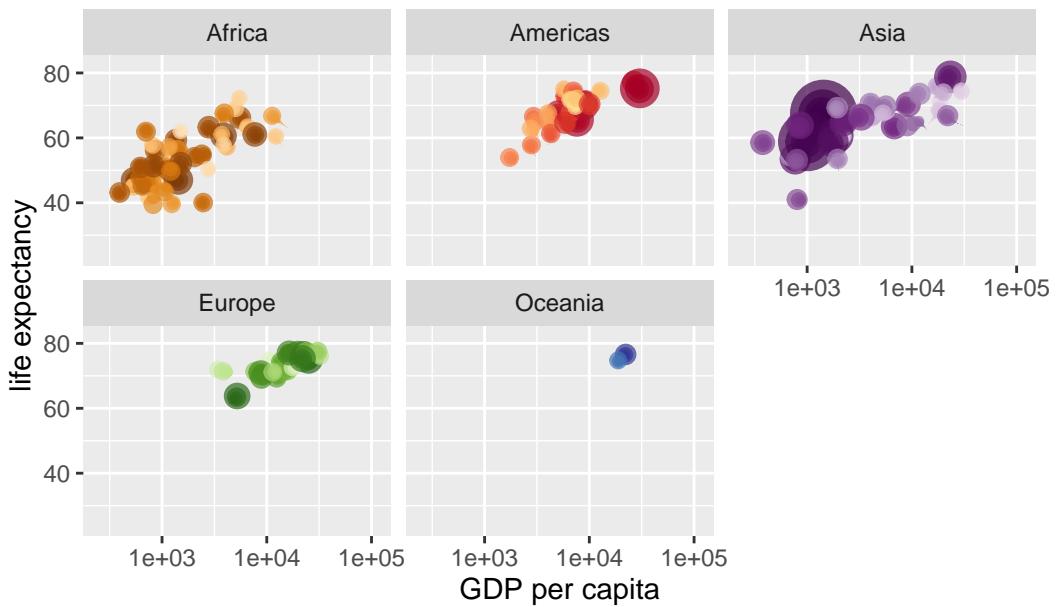
Year: 1987



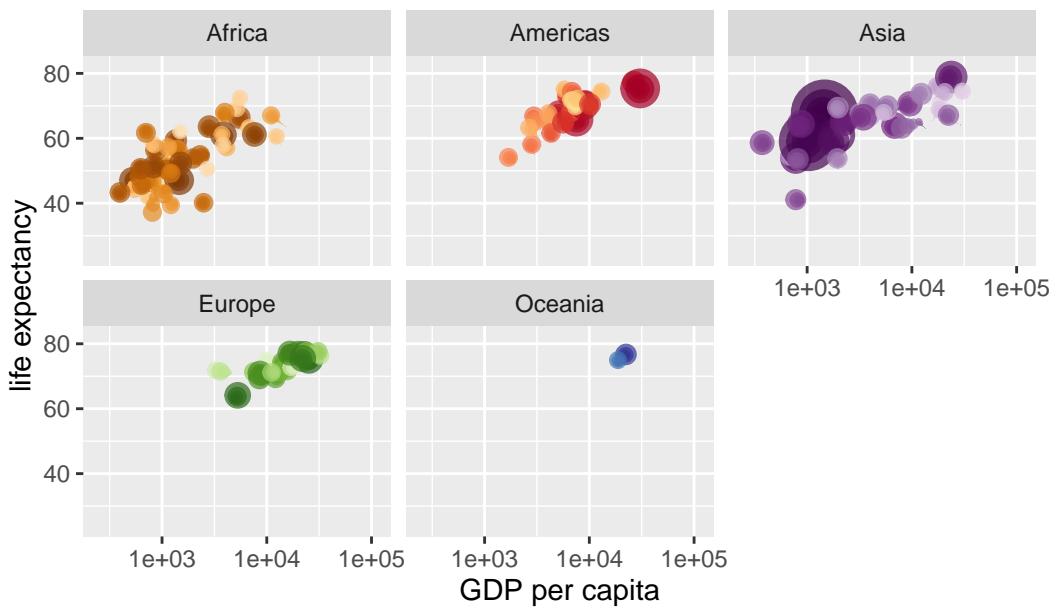
Year: 1988



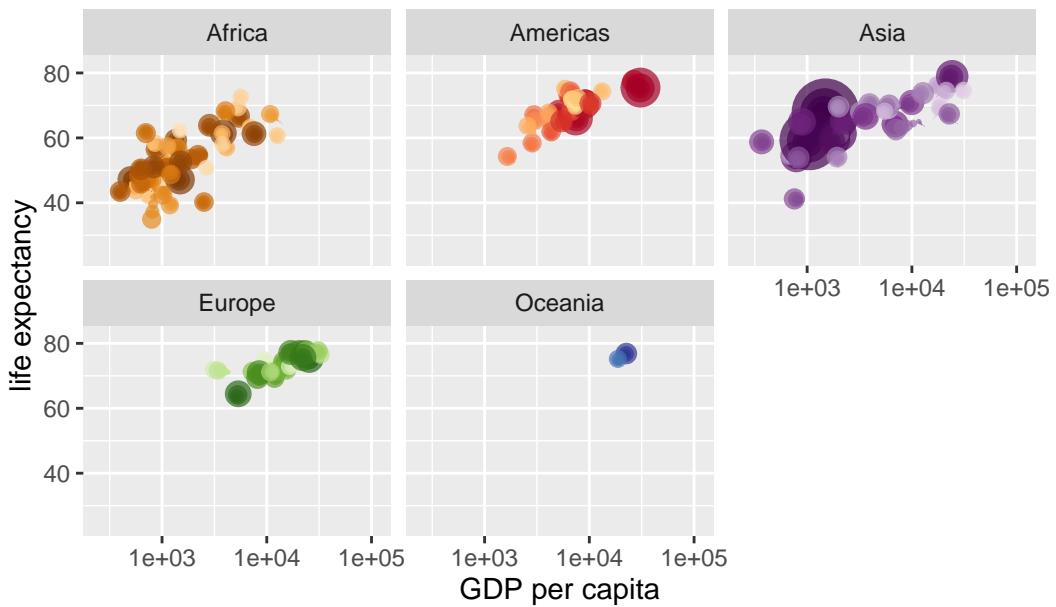
Year: 1988



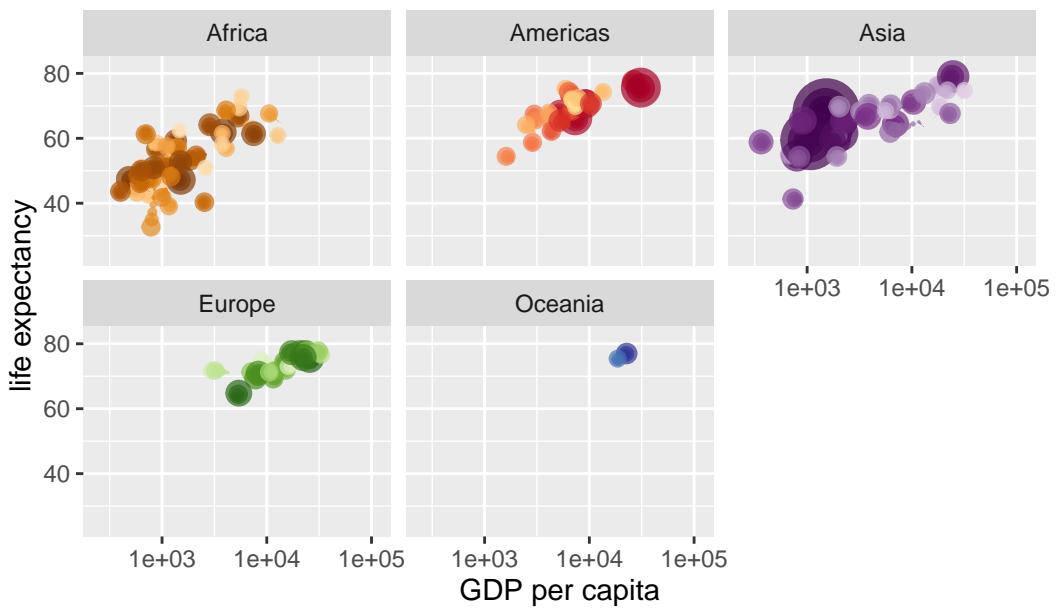
Year: 1989



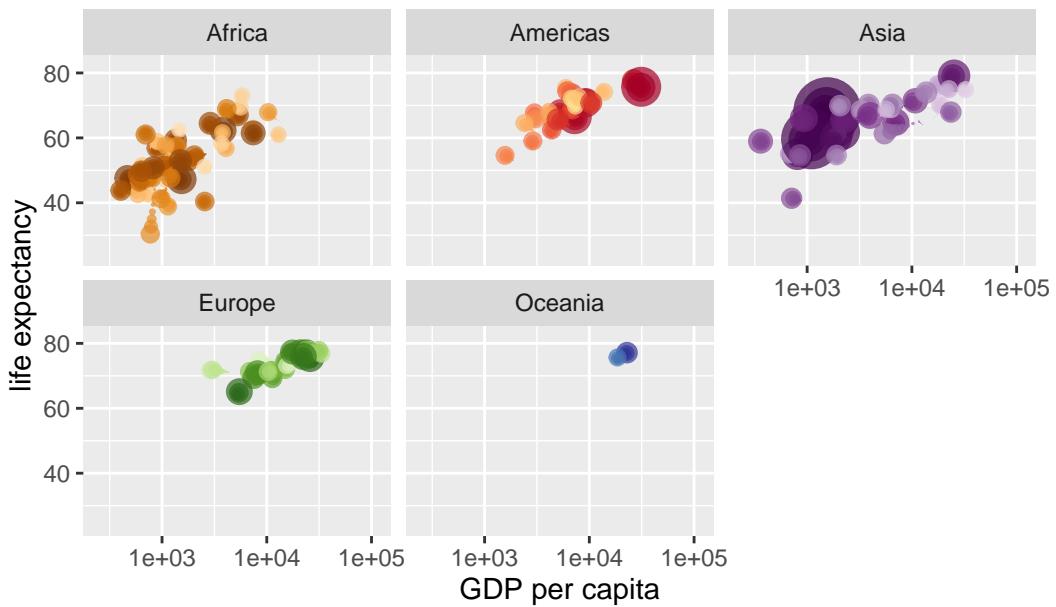
Year: 1989



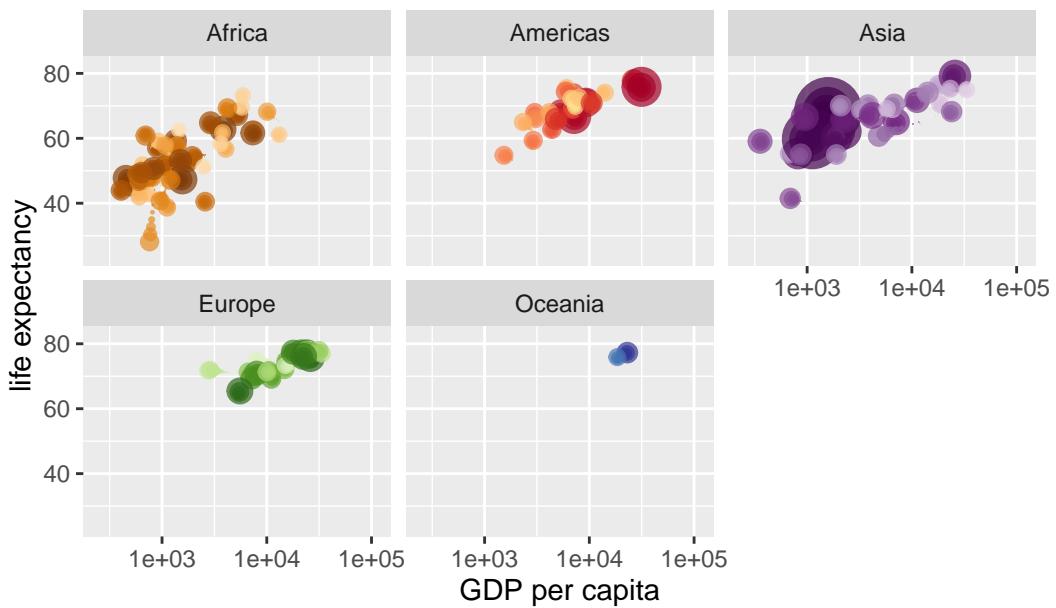
Year: 1990



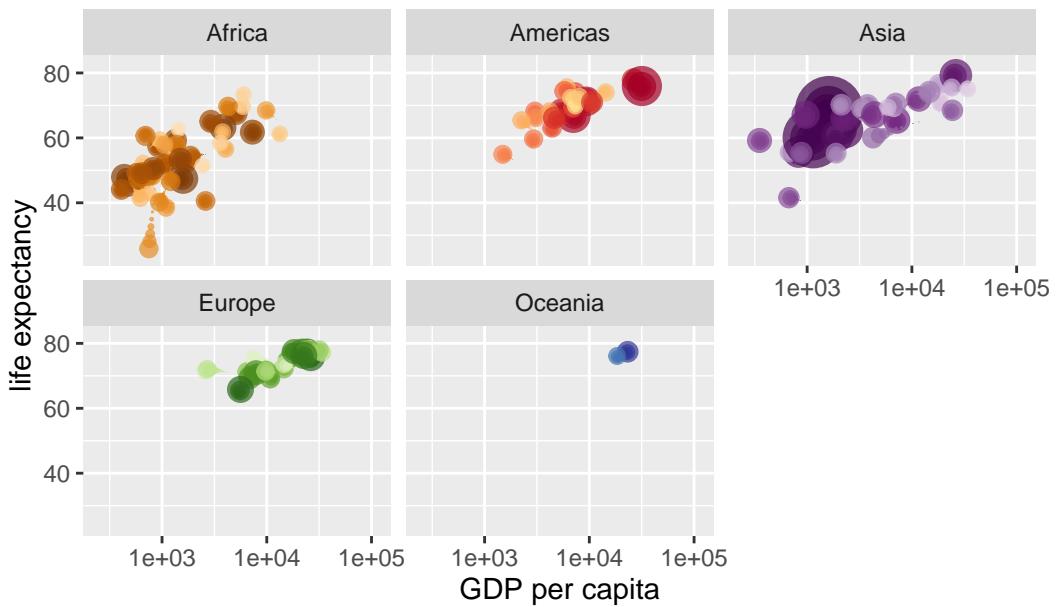
Year: 1990



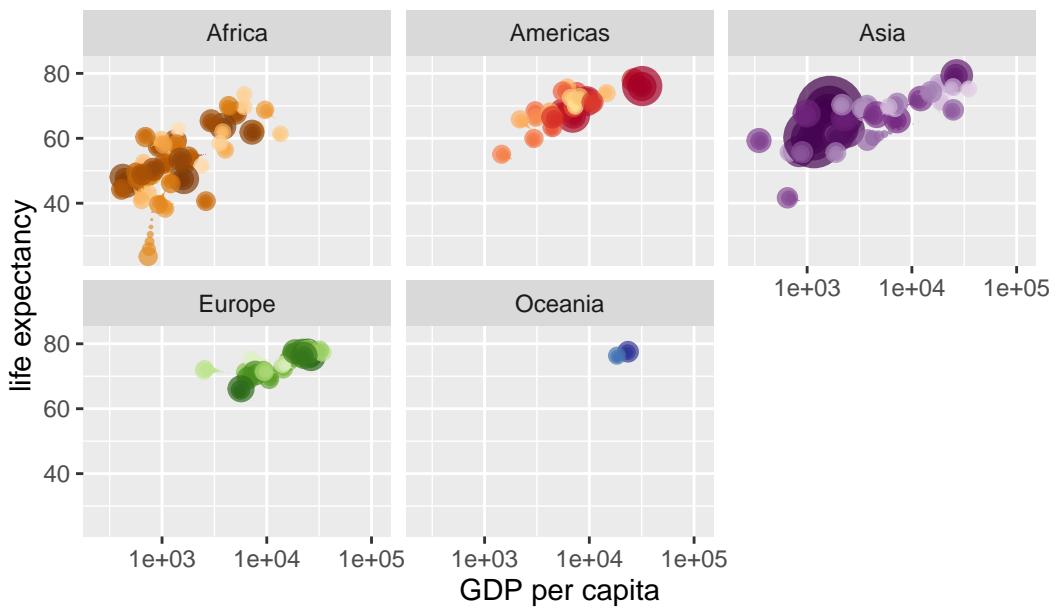
Year: 1991



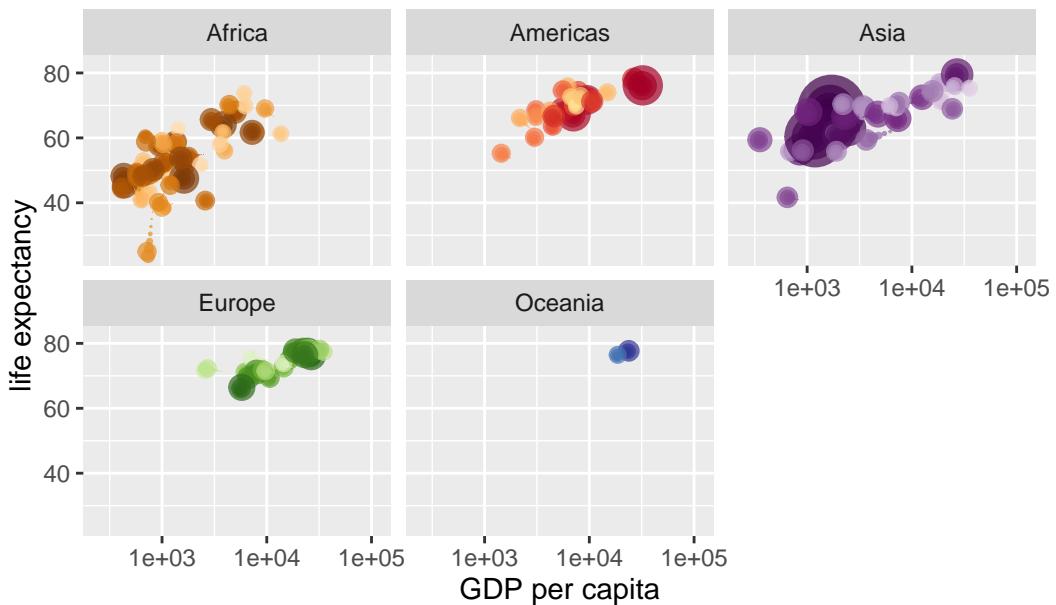
Year: 1991



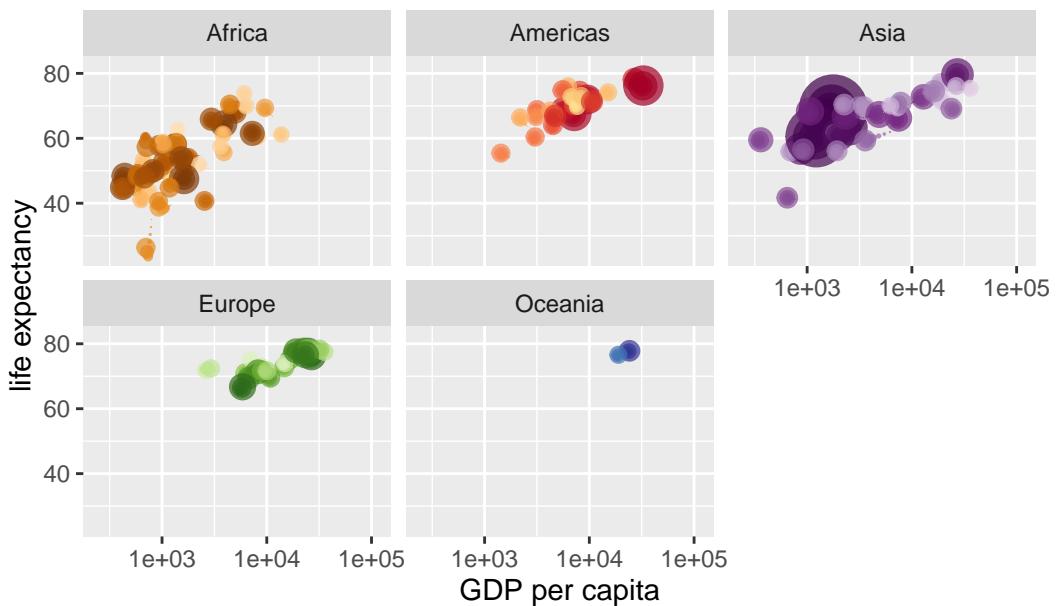
Year: 1992



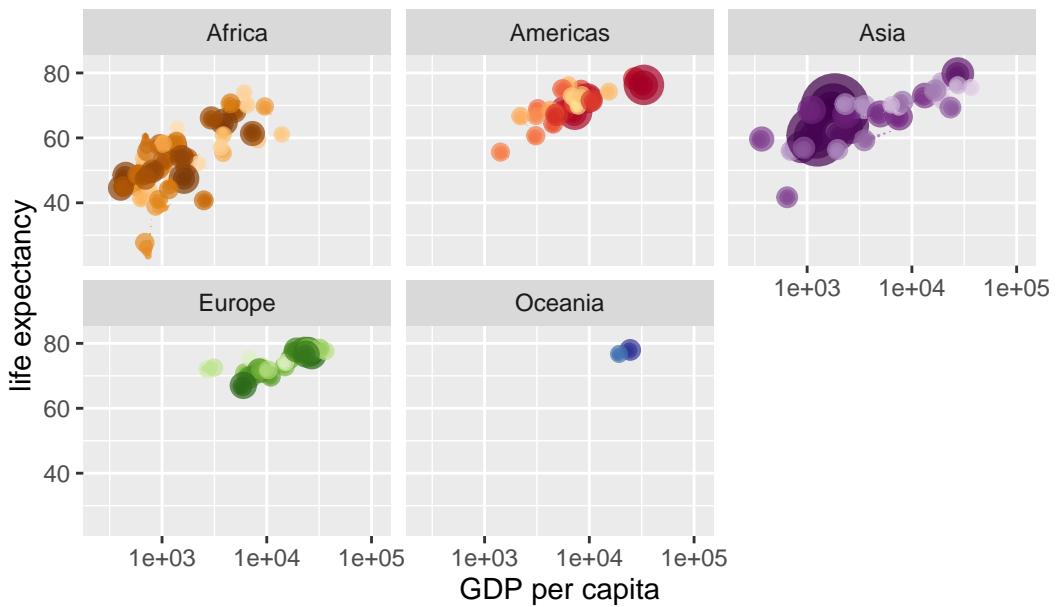
Year: 1993



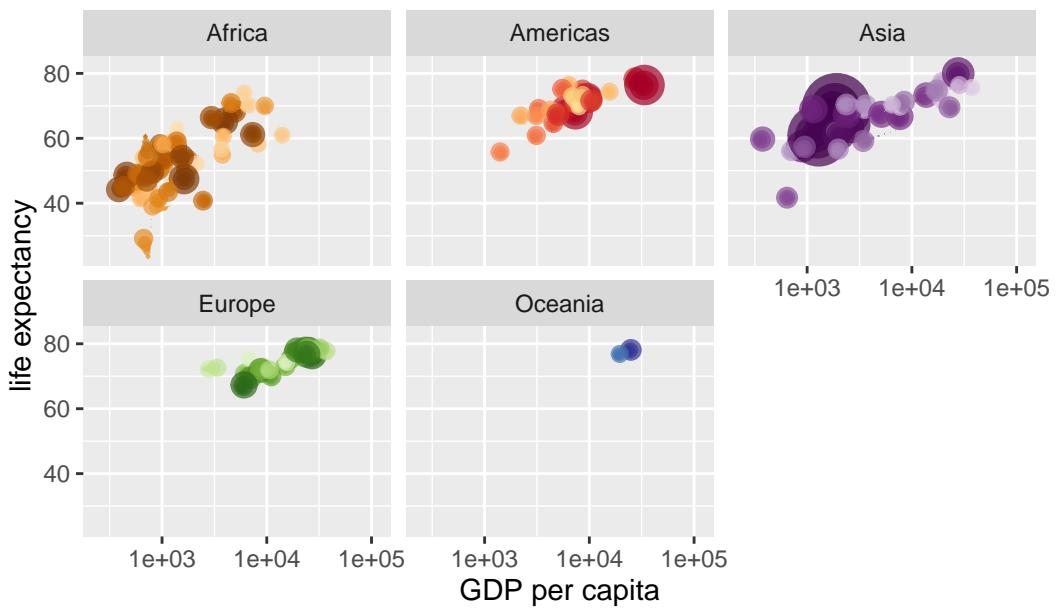
Year: 1993



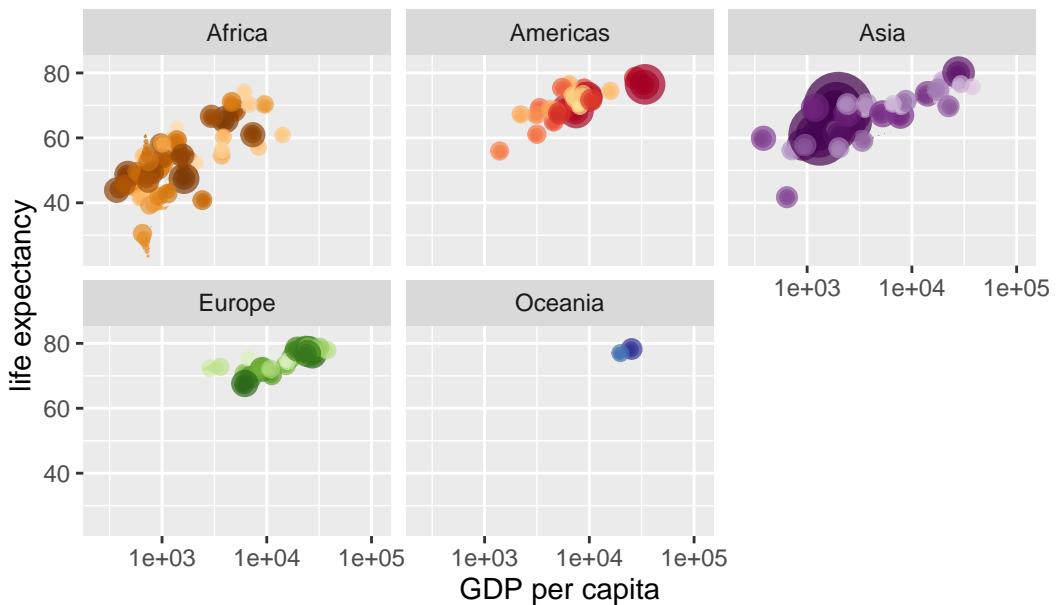
Year: 1994



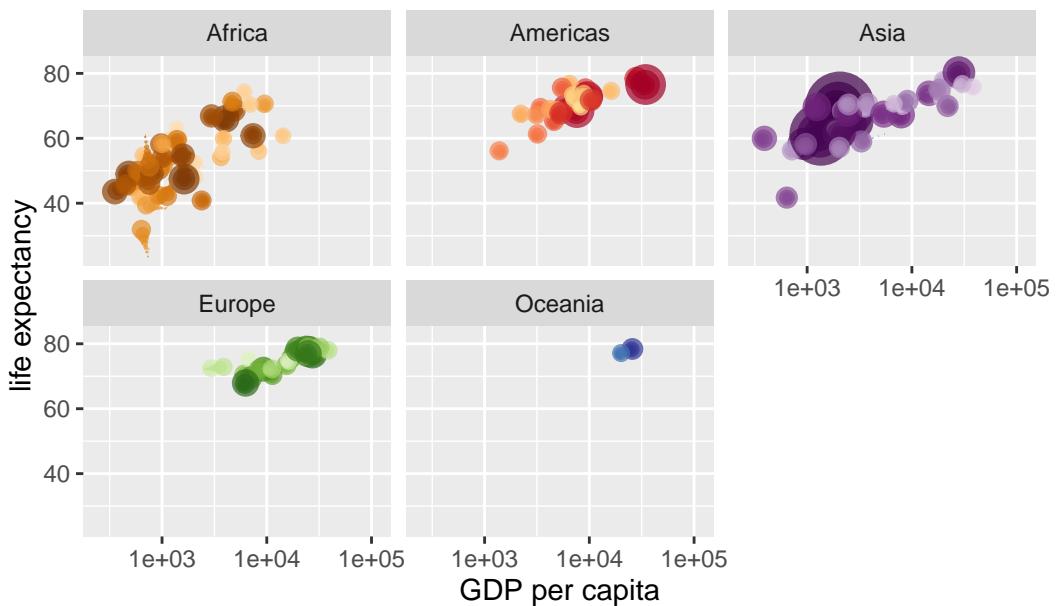
Year: 1994



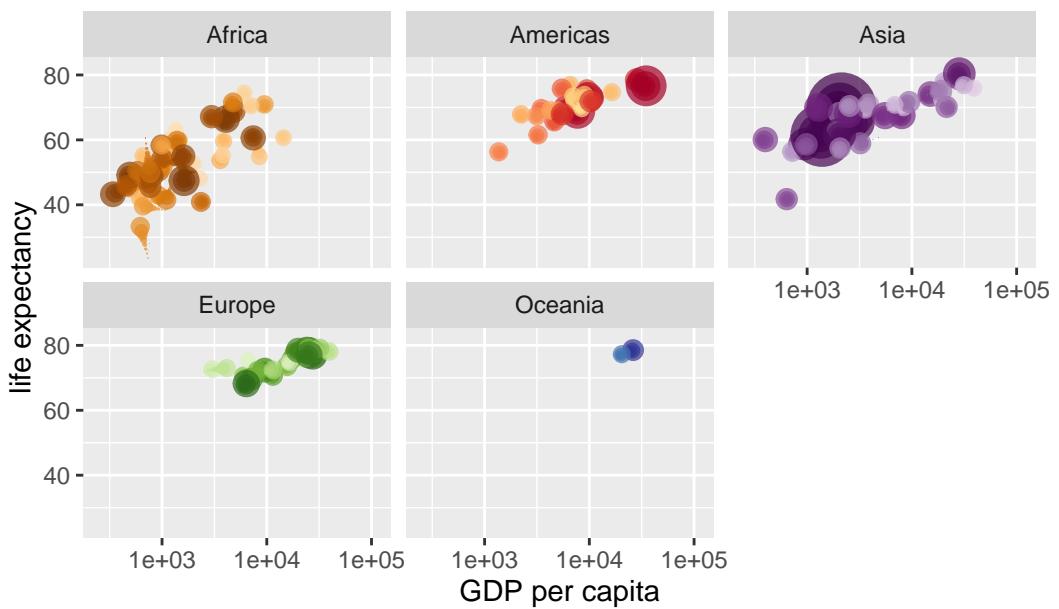
Year: 1995



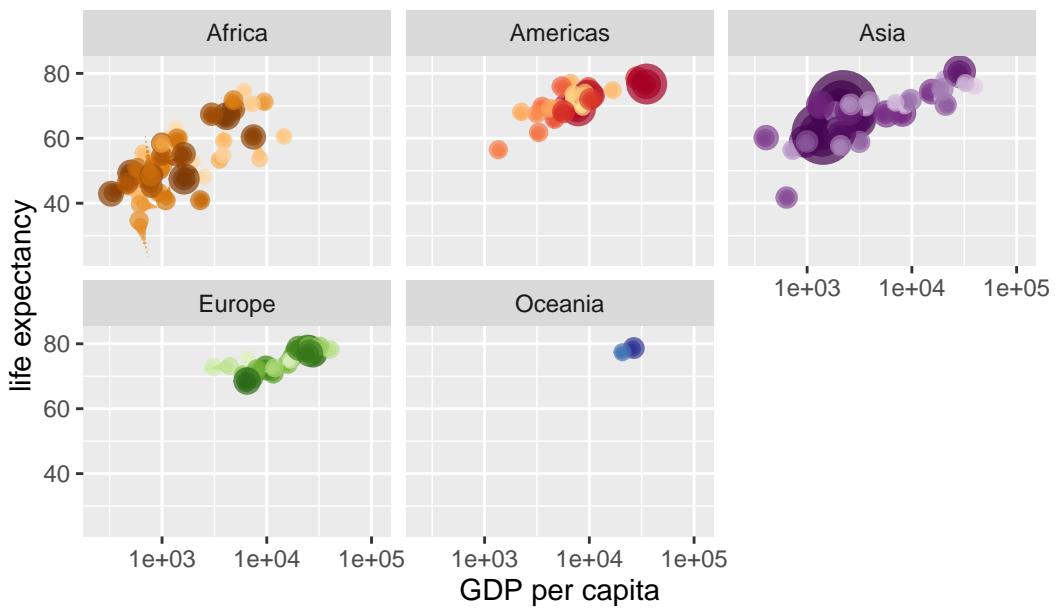
Year: 1995



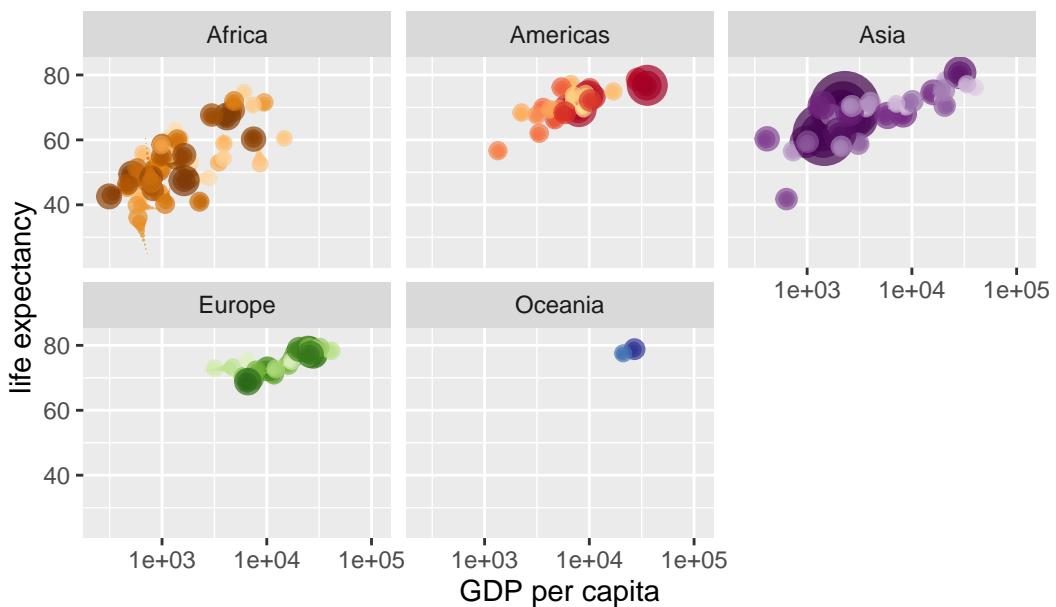
Year: 1996



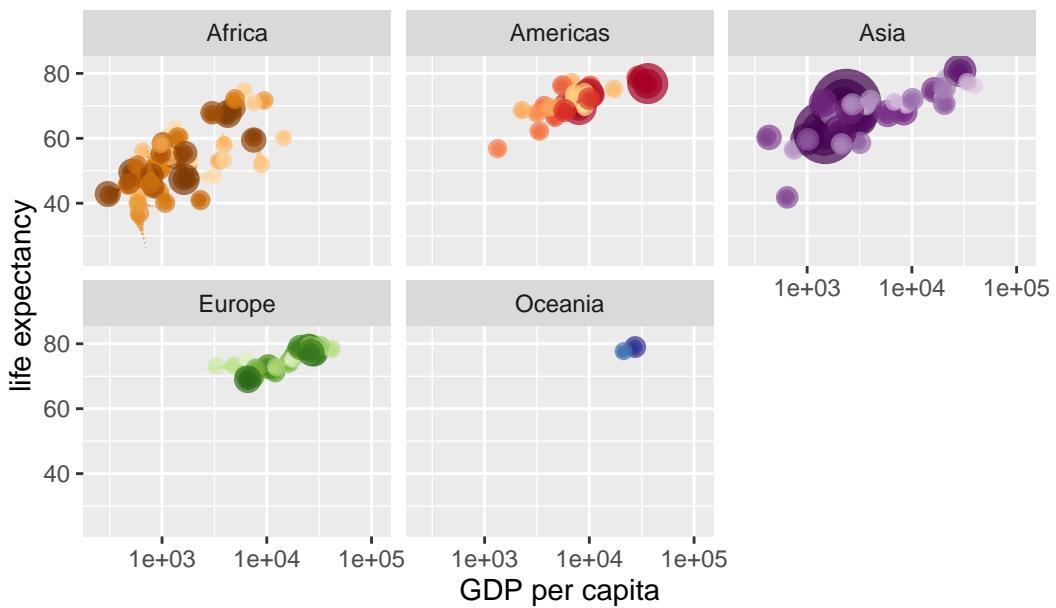
Year: 1996



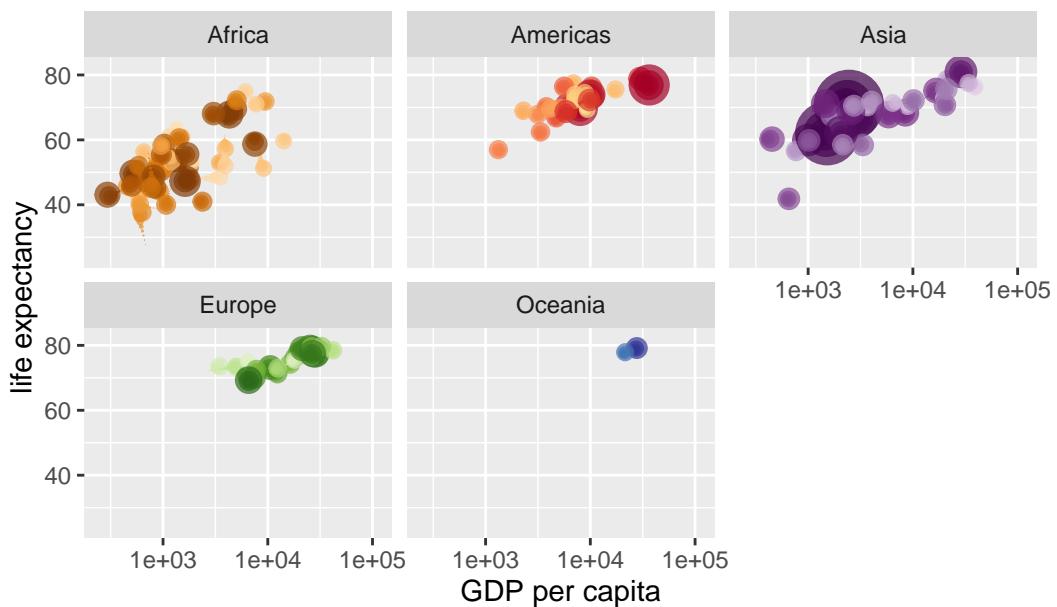
Year: 1997



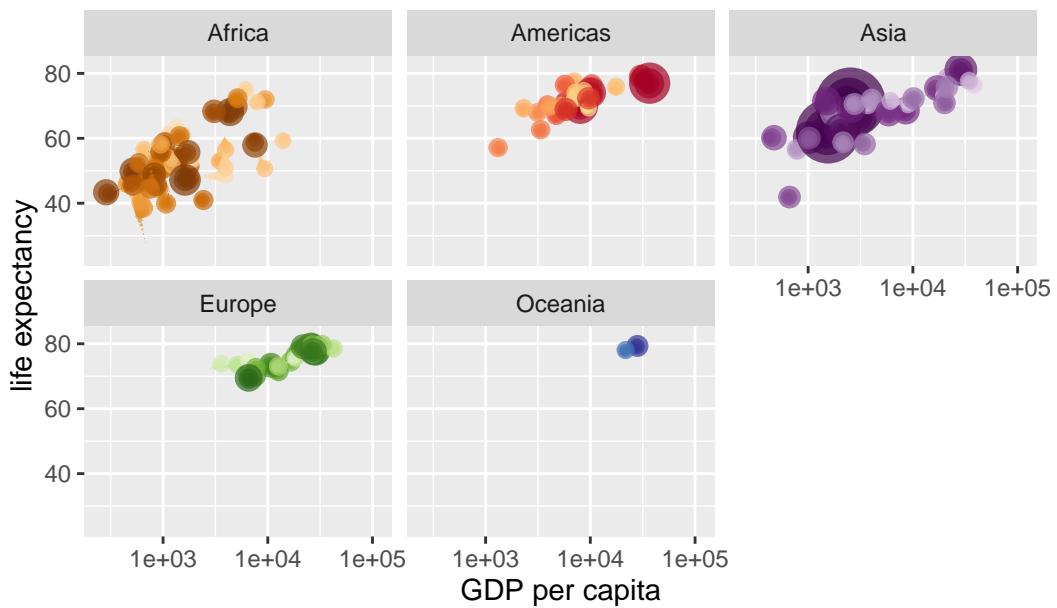
Year: 1998



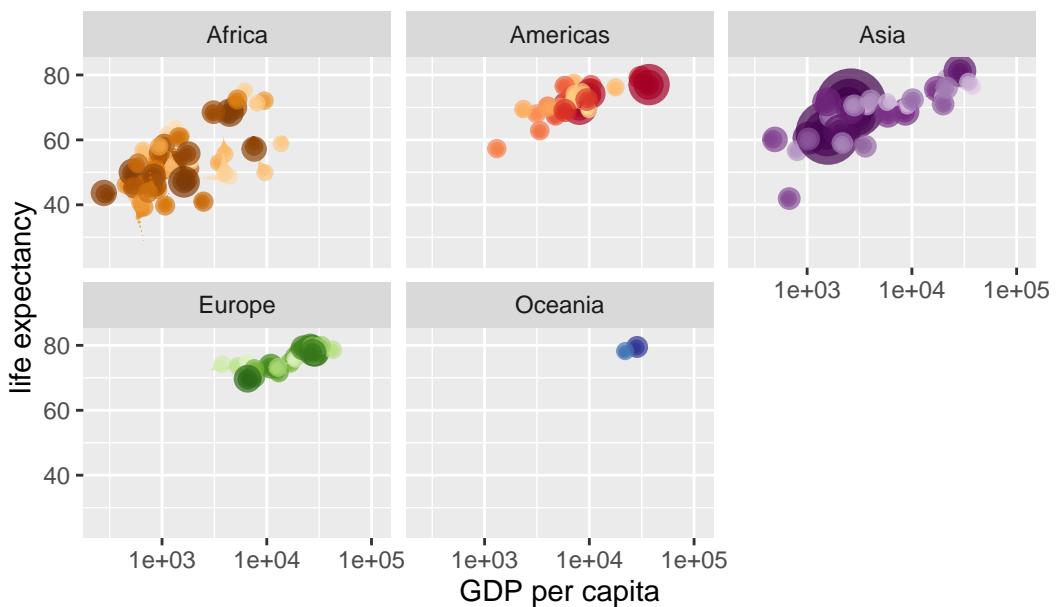
Year: 1998



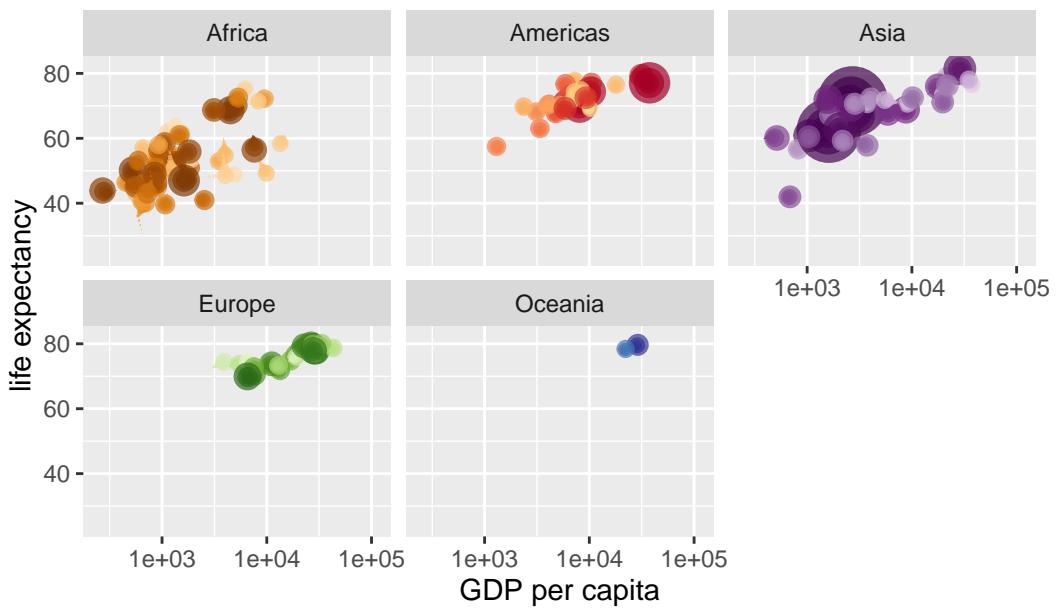
Year: 1999



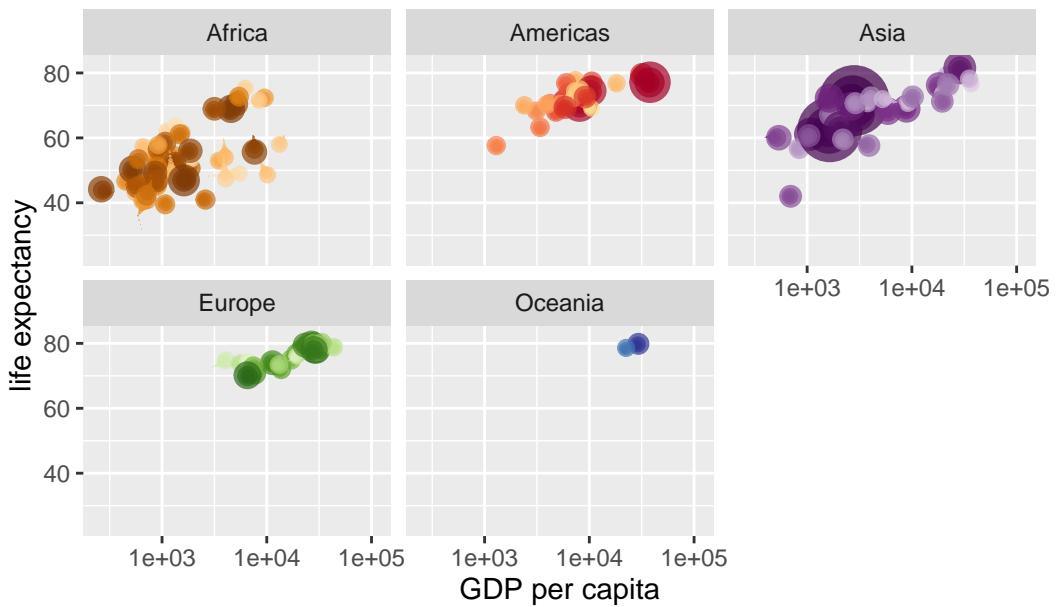
Year: 1999



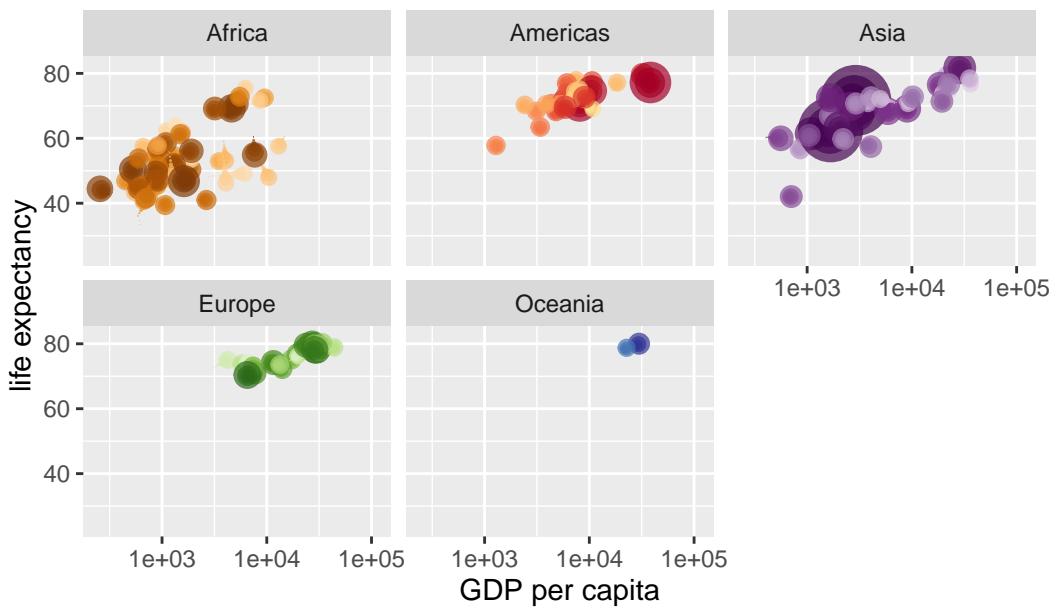
Year: 2000



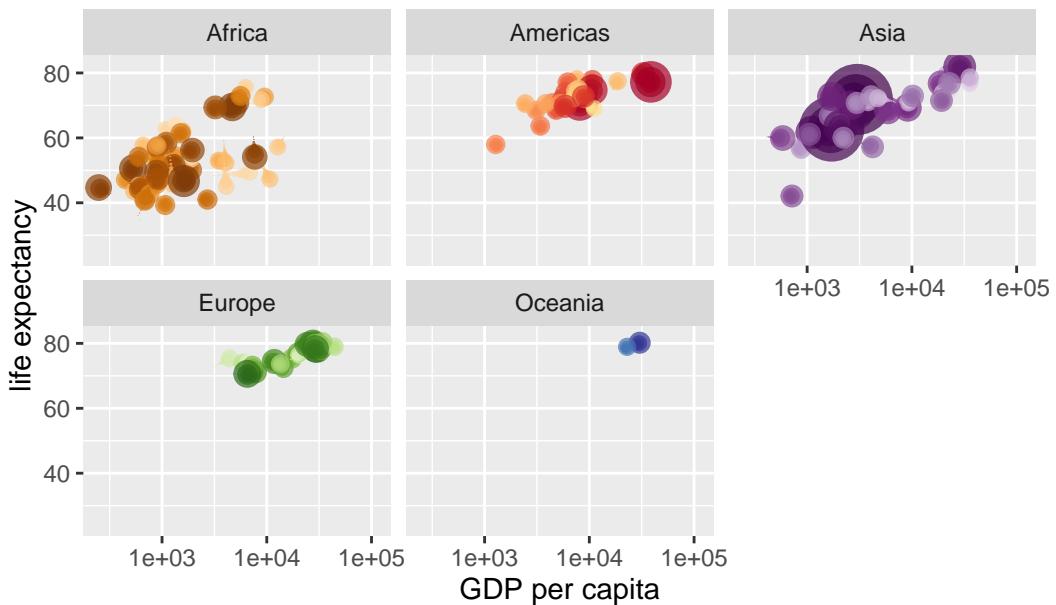
Year: 2000



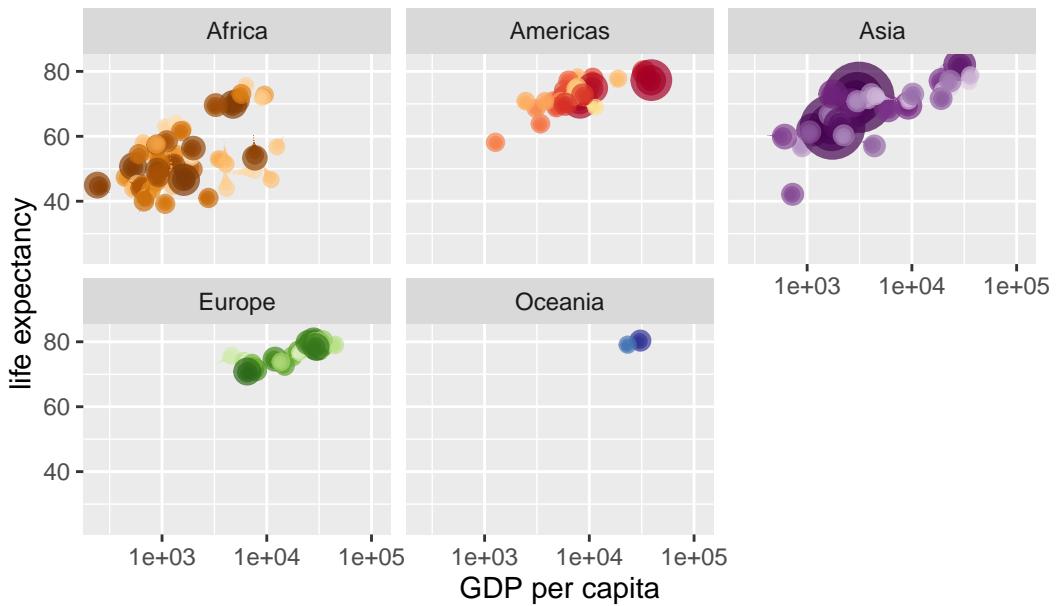
Year: 2001



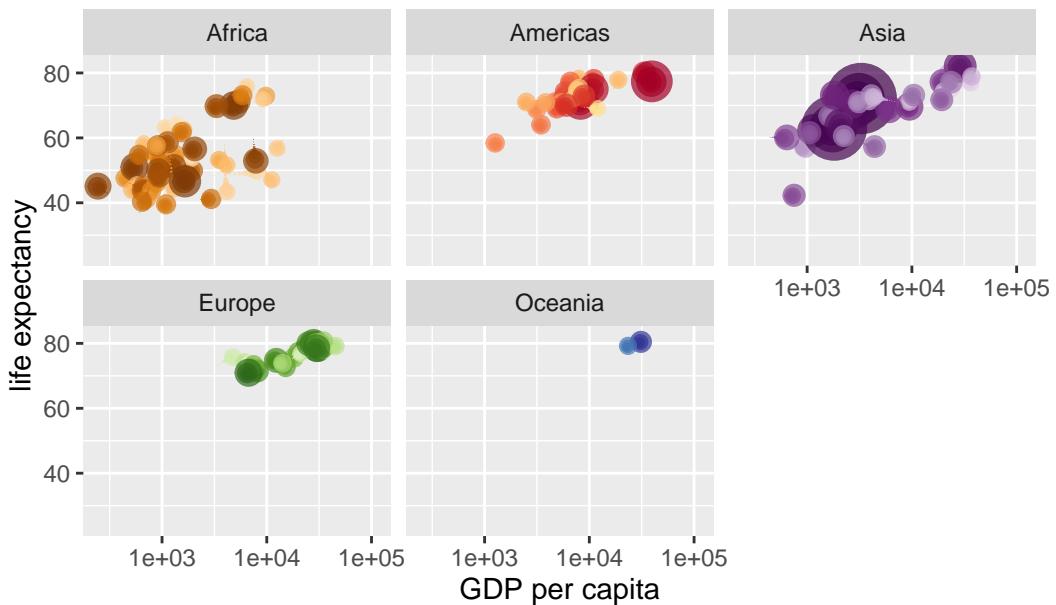
Year: 2001



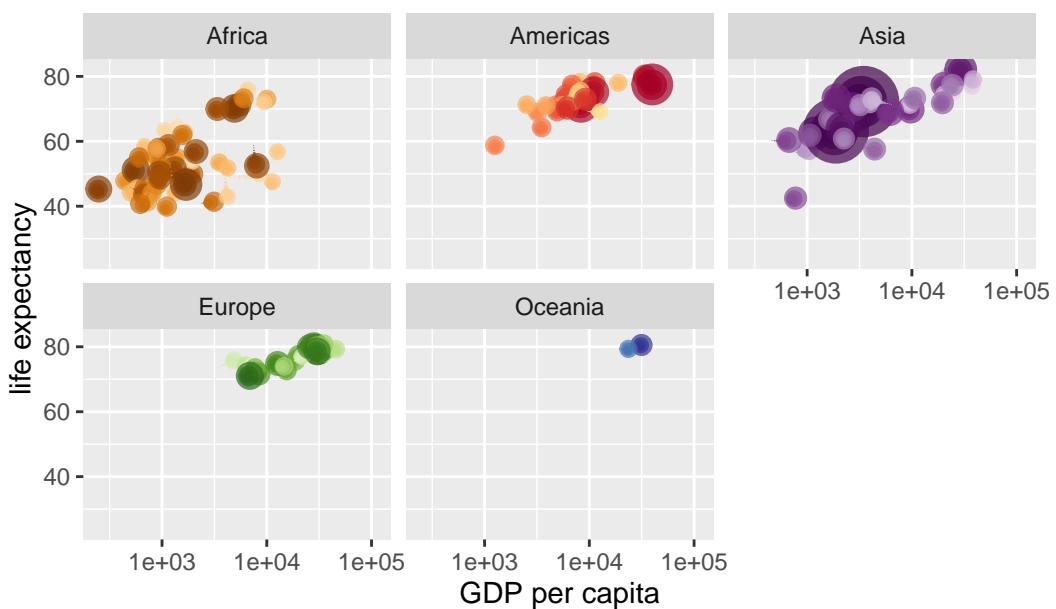
Year: 2002



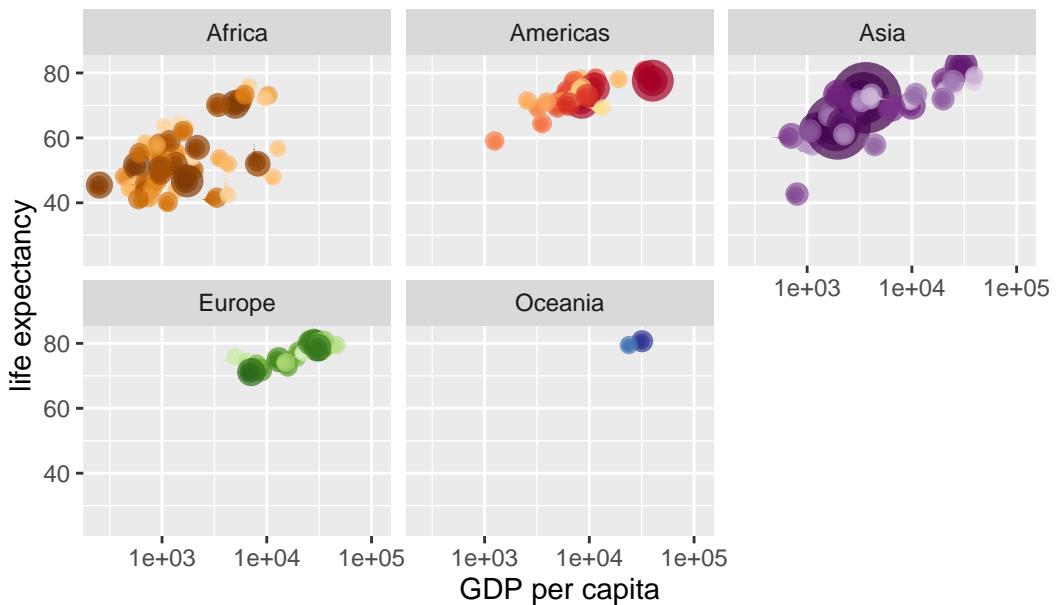
Year: 2003



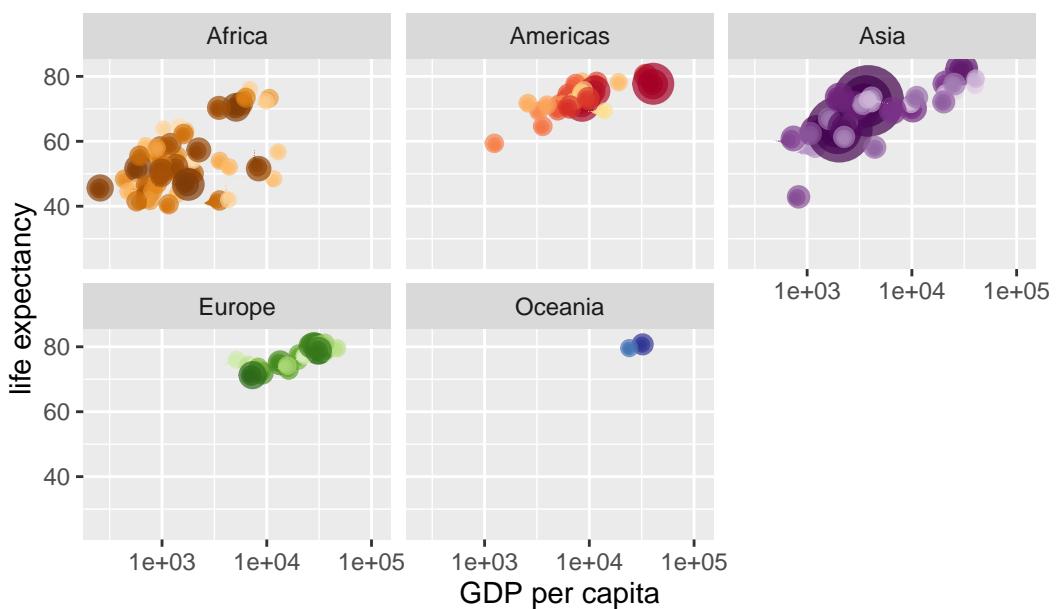
Year: 2003



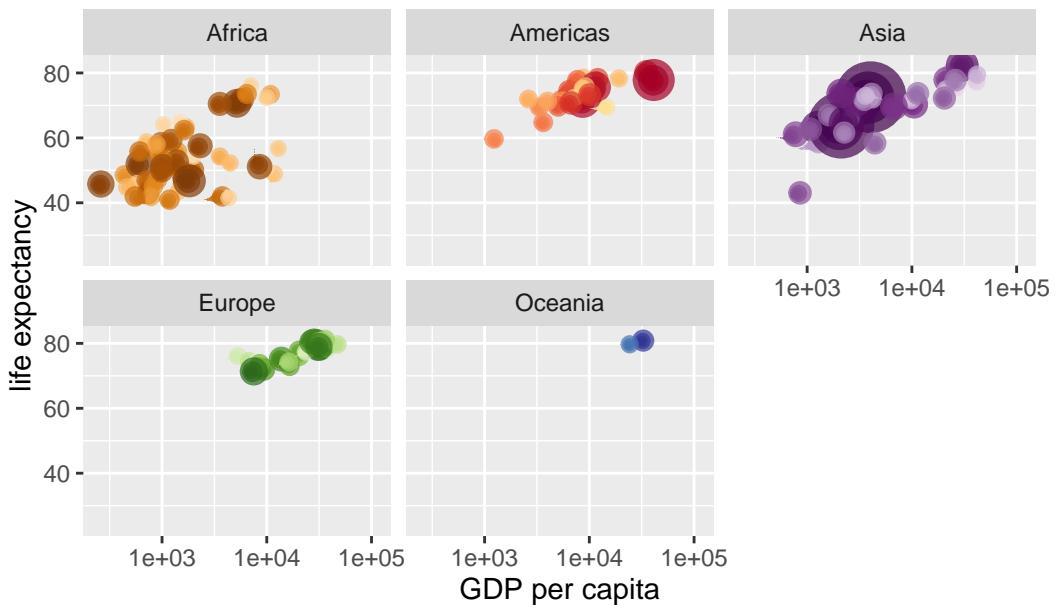
Year: 2004



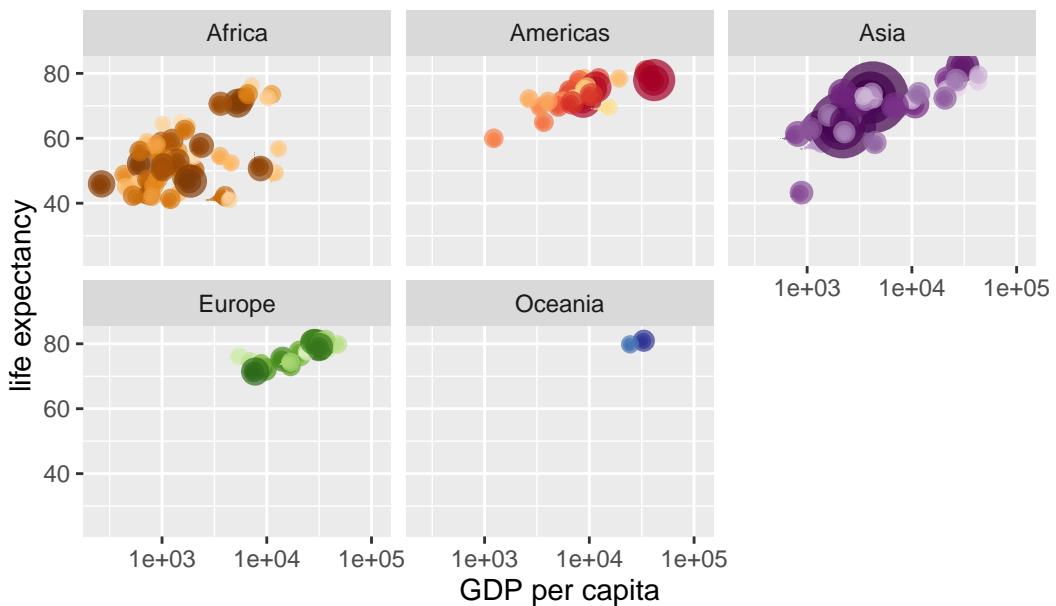
Year: 2004



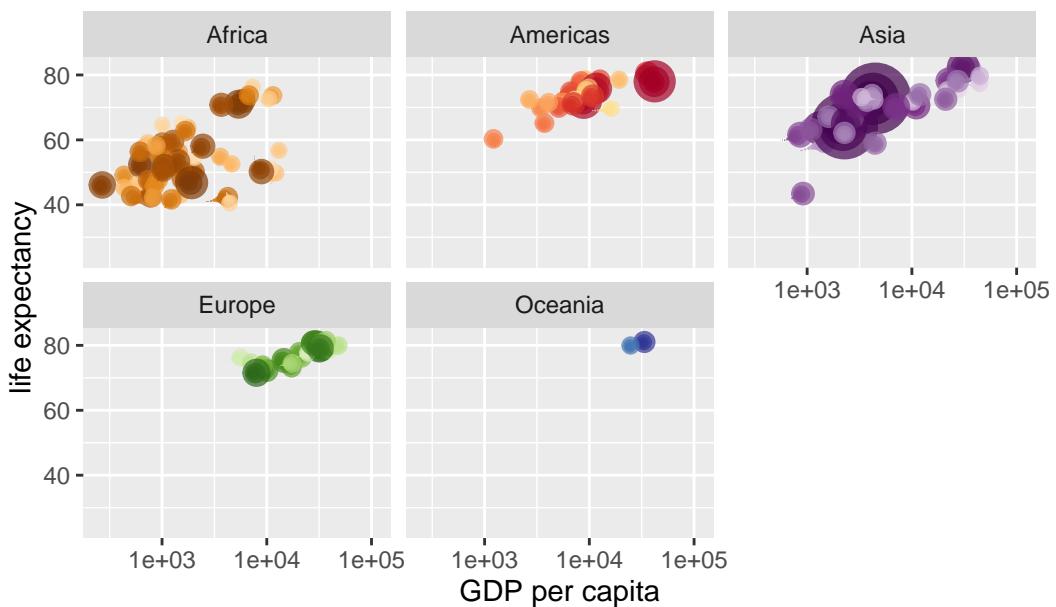
Year: 2005



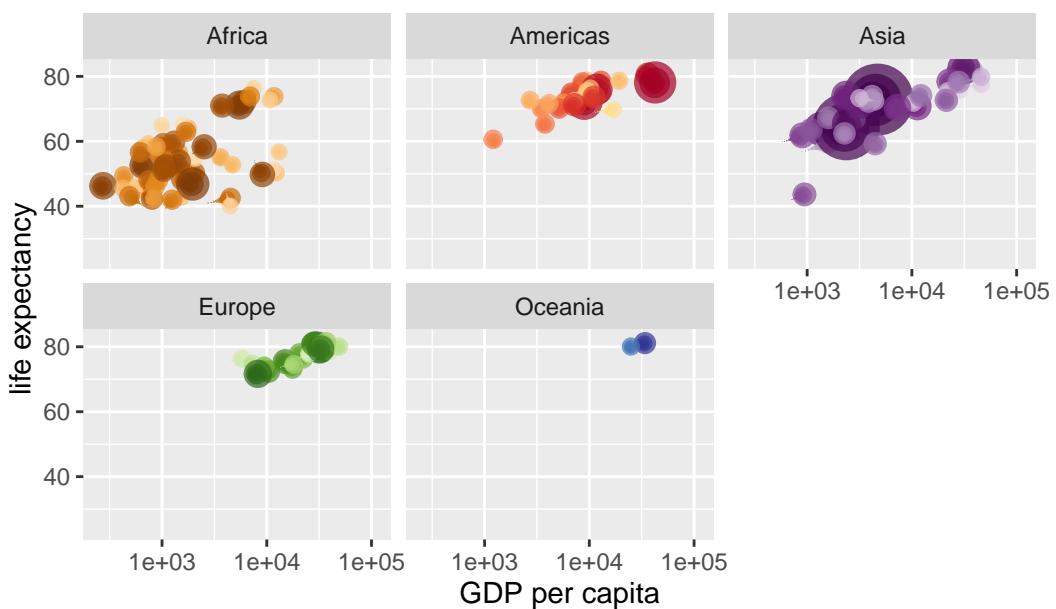
Year: 2005



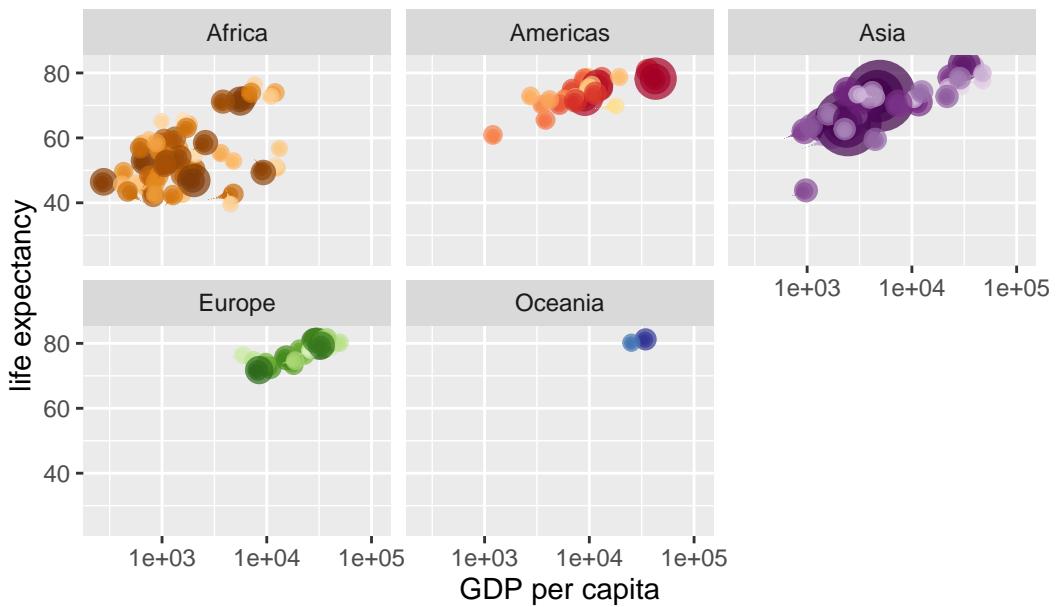
Year: 2006



Year: 2006



Year: 2007



```
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(disp, qsec))
p4 <- ggplot(mtcars) + geom_bar(aes(carb))
```

```
(p1 | p2 | p3) / p4
```

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

