

# class05: Data visualization with ggplot

Toheeb-Balogun

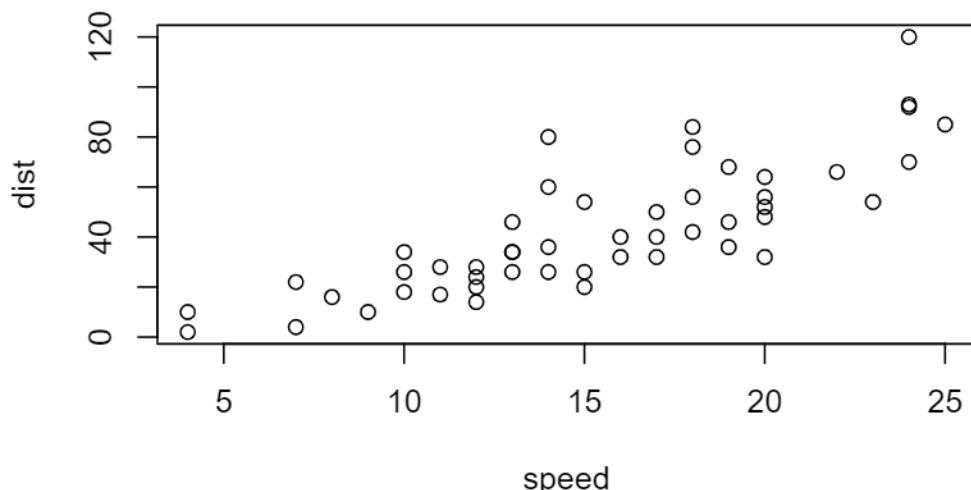
## Table of contents

|                                     |          |
|-------------------------------------|----------|
| <b>Our first plot</b>               | <b>1</b> |
| Add more interesting plot . . . . . | 4        |

## Our first plot

R has base graphics

```
plot(cars)
```



How would I plot this with ggplot2? NO!

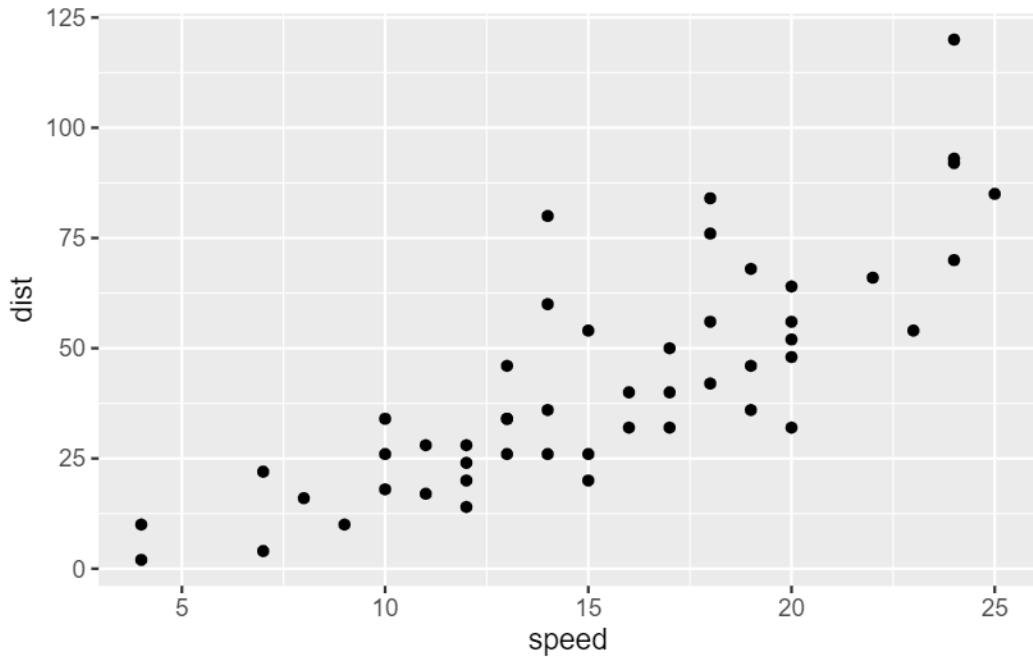
We need to install and load the ggplot2 package first. To install any package in R we use the `install.packages`

Before I can use this package, I need to load it with a `library()` call.

Every ggplot need at least 3 layers

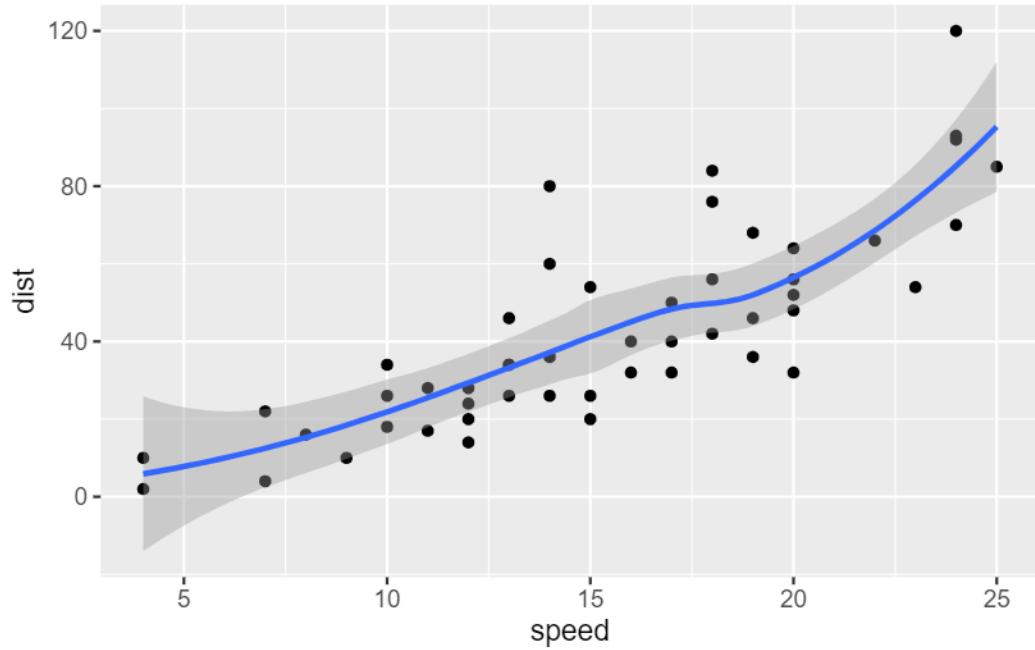
- **Data** (i.e the `data.frame` we have),
- **Aes** (the aesthetic mapping of our data to do what we want to plot)
- **Geoms** (How we want to plot this stuff!)

```
library(ggplot2)
ggplot(data=cars) +
  aes(x=speed, y=dist) +
  geom_point()
```



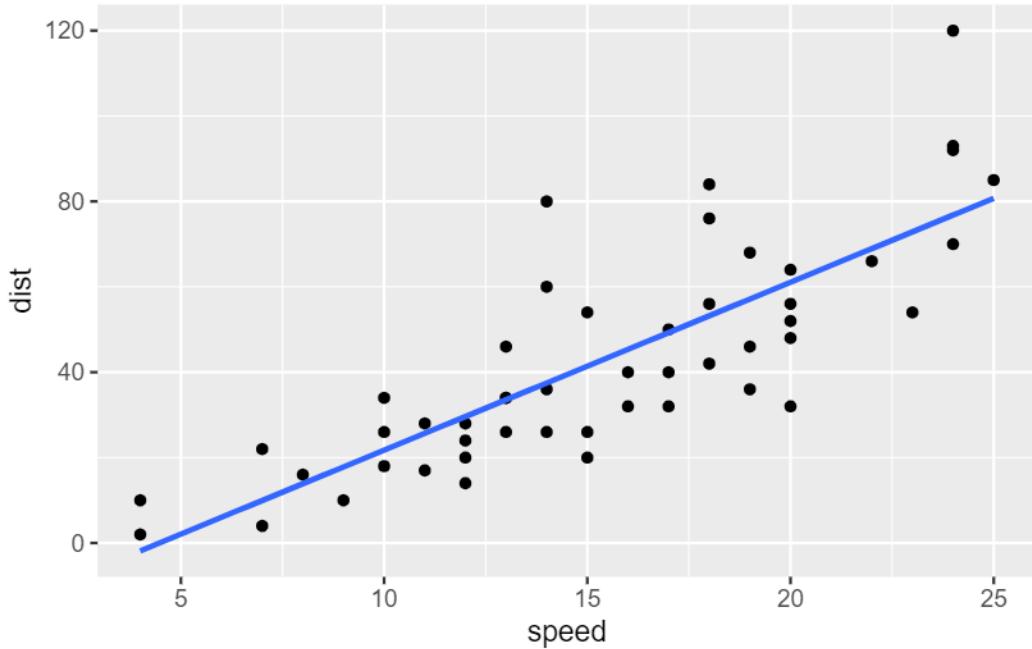
```
library(ggplot2)
ggplot(data=cars)+ 
  aes(x=speed, y=dist) +
  geom_point()+
  geom_smooth()
```

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



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

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



## Add more interesting plot

#First read data from online

```
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 |

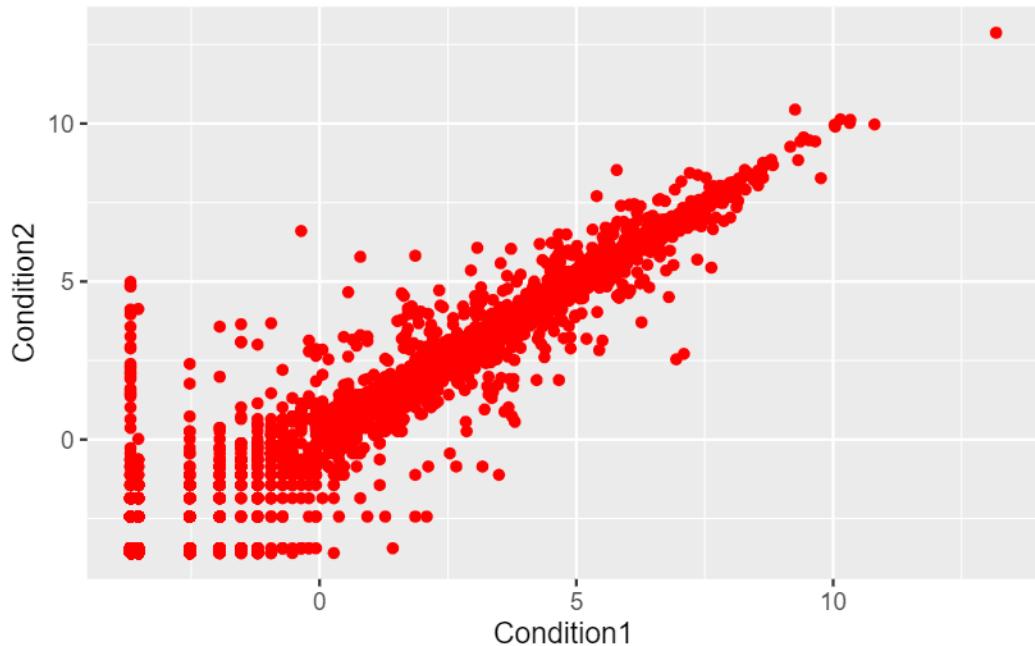
Q1. How many genes are in this dataset

```
nrow(genes)
```

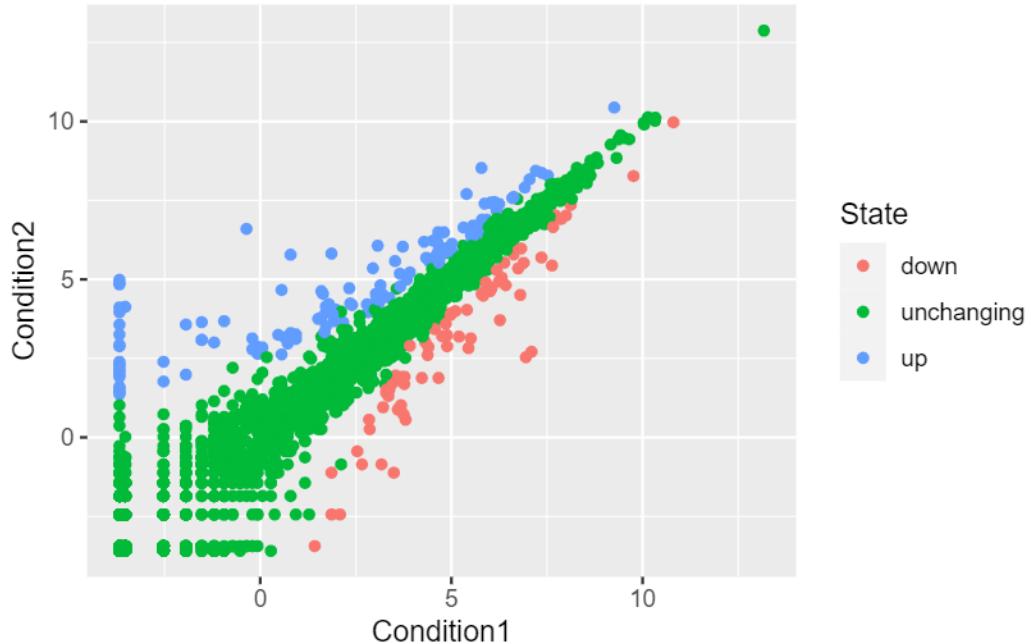
```
[1] 5196
```

```
#There are 5196 genes in this data set
```

```
ggplot(genes) +  
  aes(x=Condition1, y=Condition2, col=State) +  
  geom_point(col="red")
```



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



#Q2. print the columnnames and the number of columns

```
colnames(genes)
```

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

```
ncol(genes)
```

```
[1] 4
```

```
table(genes$State)[3]
```

```
up  
127
```

```
#the code below does the same thing!  
#table(genes[,4])[3]  
round( table(genes$State)/nrow(genes) * 100, 2 )
```

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

There are 127 'up' regulated genes

#Plotly is a package where you can hover a point and see the gene #>%>%: this called foward pipe opeartor.This operator will forward a value, or the result of an expression, into the next function call/expression

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

Attaching package: 'dplyr'

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

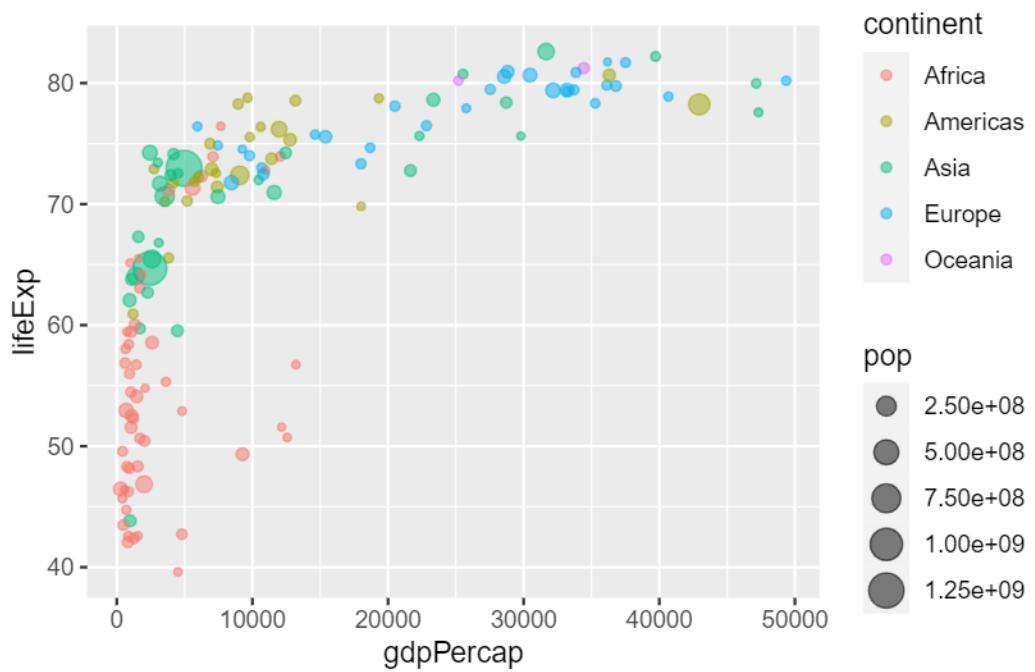
filter, lag

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

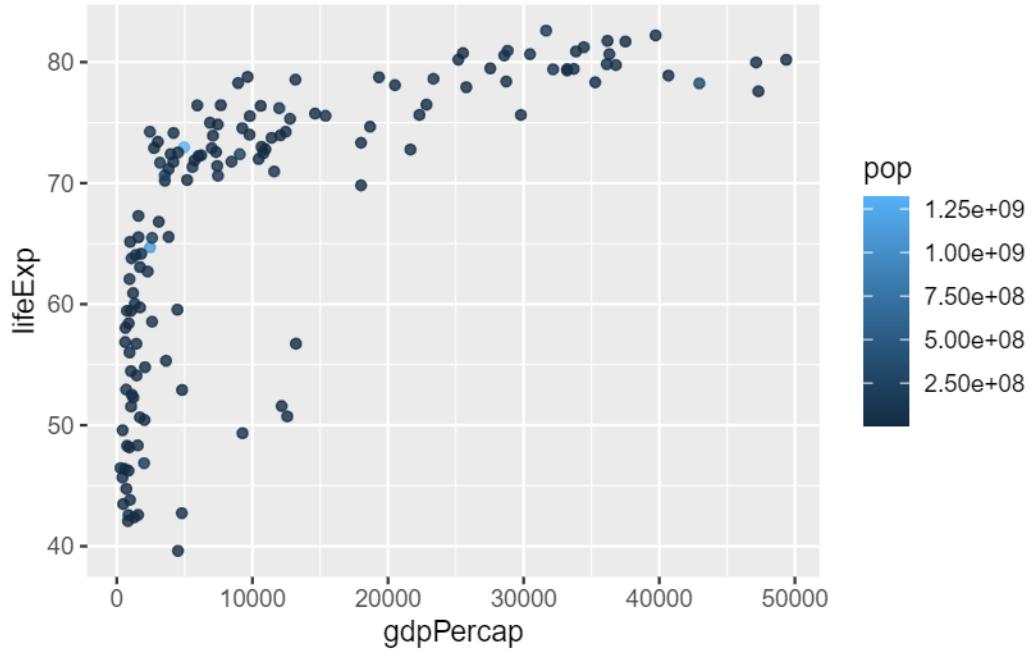
intersect, setdiff, setequal, union

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

library(ggplot2)
ggplot(gapminder_2007) +
  aes(x=gdpPercap, y=lifeExp, color=continent, size=pop) +
  geom_point(alpha=0.5)
```

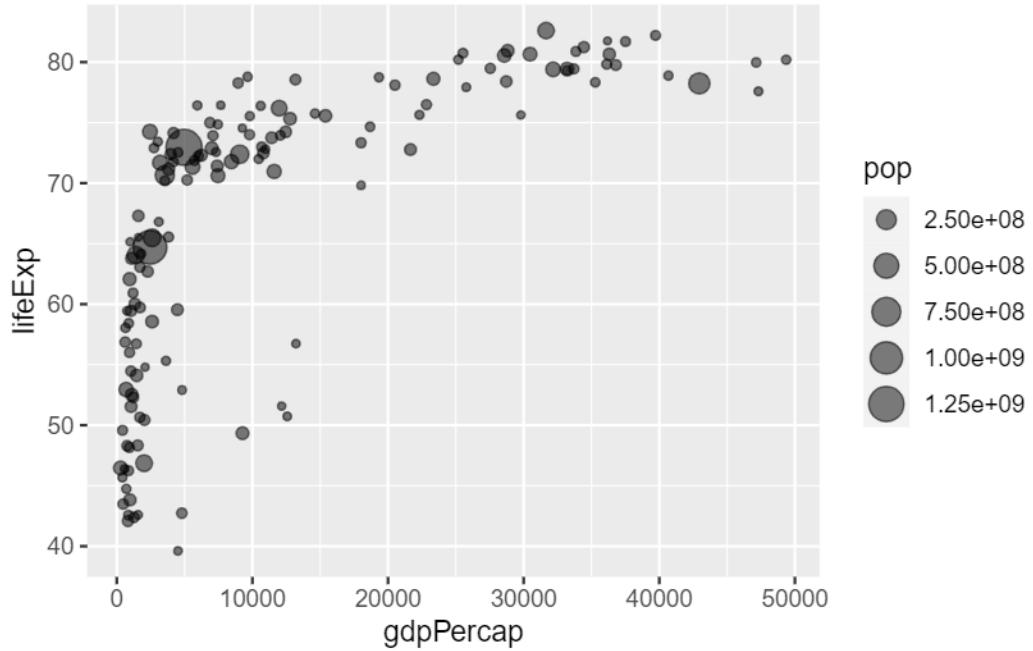


```
ggplot(gapminder_2007) +  
  aes(x=gdpPercap, y=lifeExp, color=pop) +  
  geom_point(alpha=0.8)
```

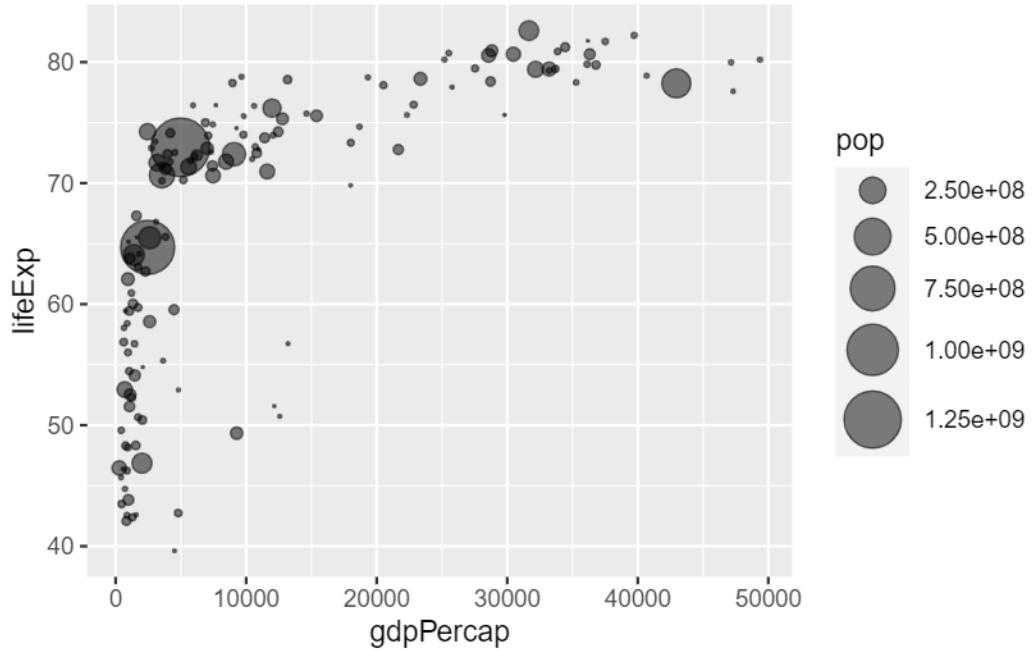


#Adjusting point size

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



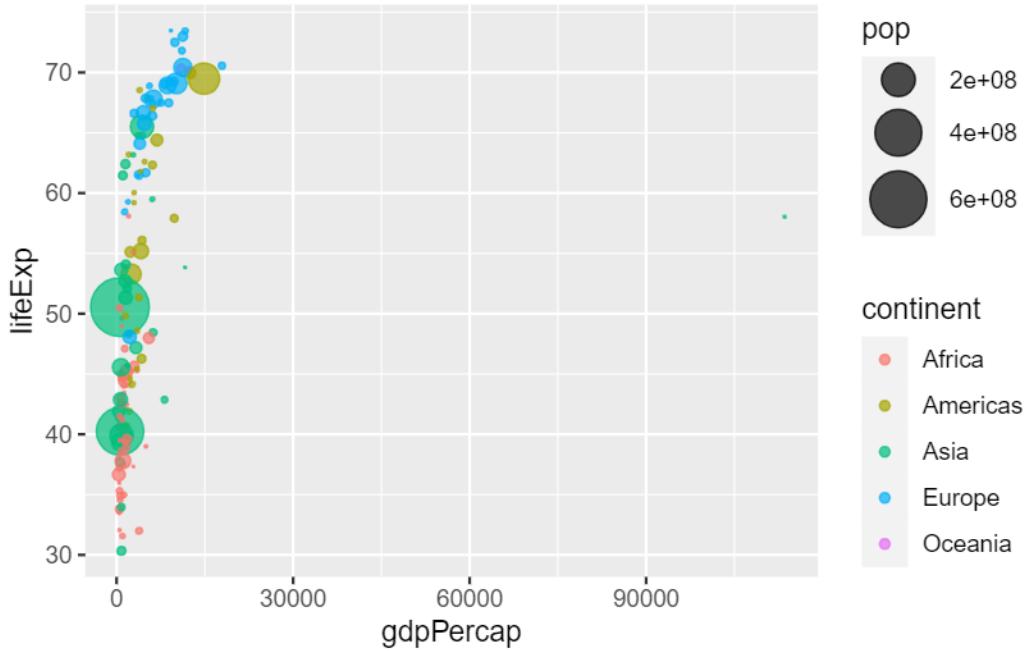
```
ggplot(gapminder_2007) +  
  geom_point(aes(x = gdpPerCap, y = lifeExp,  
                 size = pop), alpha=0.5) +  
  scale_size_area(max_size = 10)
```



#steps to produce 1957 plots

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

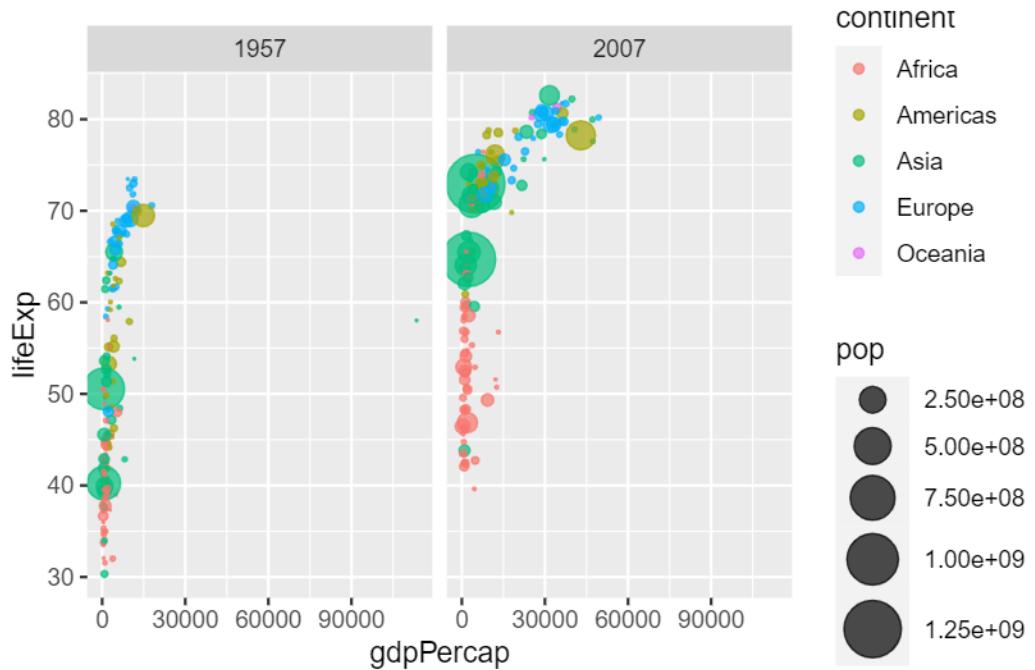
ggplot(gapminder_1957) +
  aes(x = gdpPercap, y = lifeExp, color=continent,
      size = pop) +
  geom_point(alpha=0.7) +
  scale_size_area(max_size = 10)
```



#plots for 1957 and 2007

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

ggplot(gapminder_1957) +
  geom_point(aes(x = gdpPercap, y = lifeExp, color=continent,
                 size = pop), alpha=0.7) +
  scale_size_area(max_size = 10) +
  facet_wrap(~year)
```



#Bar charts #desc() is used to sort a variable in descending order.

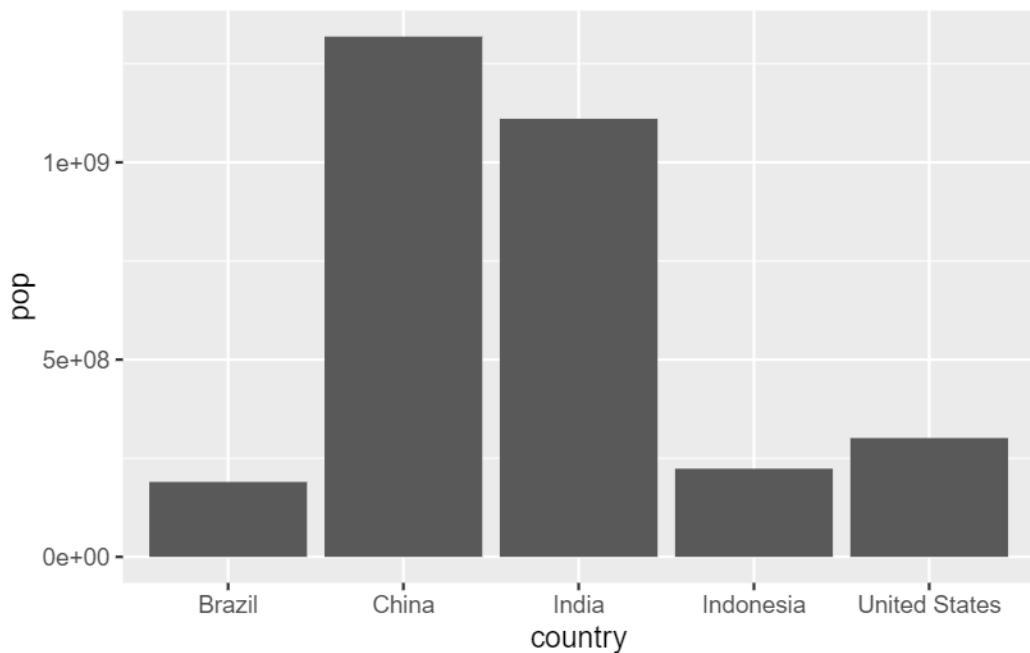
```
gapminder_top5 <- gapminder %>%
  filter(year==2007) %>%
  arrange(desc(pop)) %>%
  top_n(5, pop)

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  |

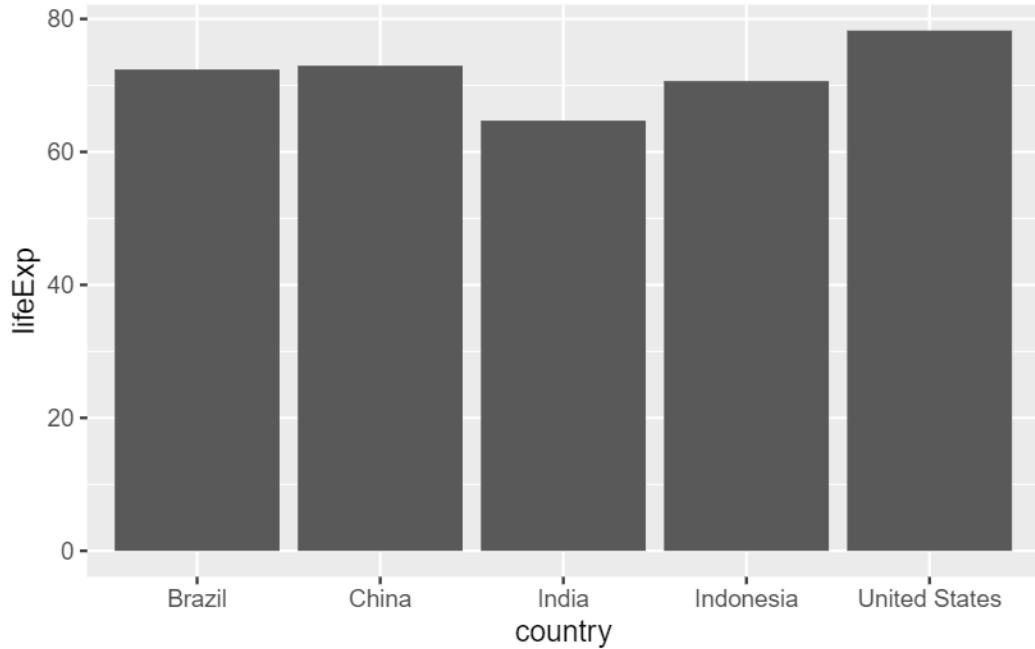
#Creating a simple bar chart

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



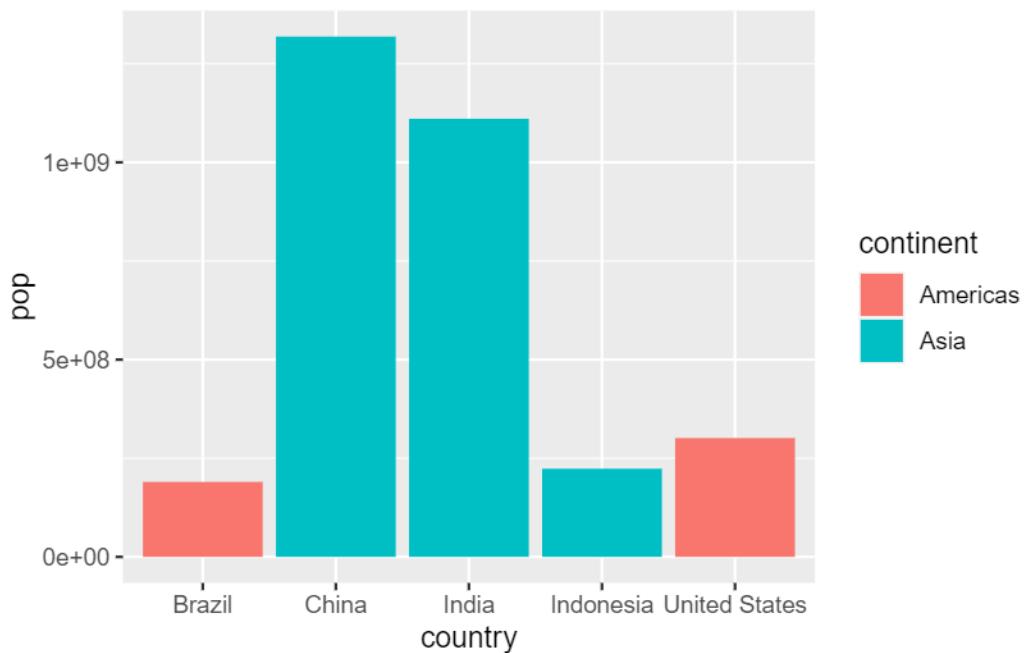
#bar chart showing the life expectancy of the five biggest countries by population in 2007

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

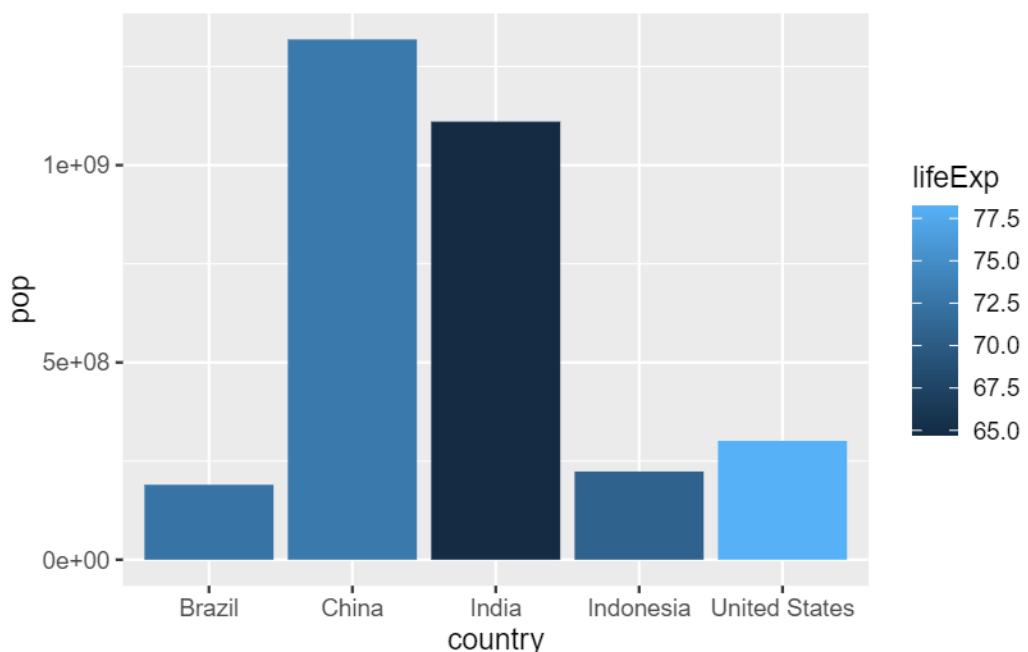


#Filling bars with color

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

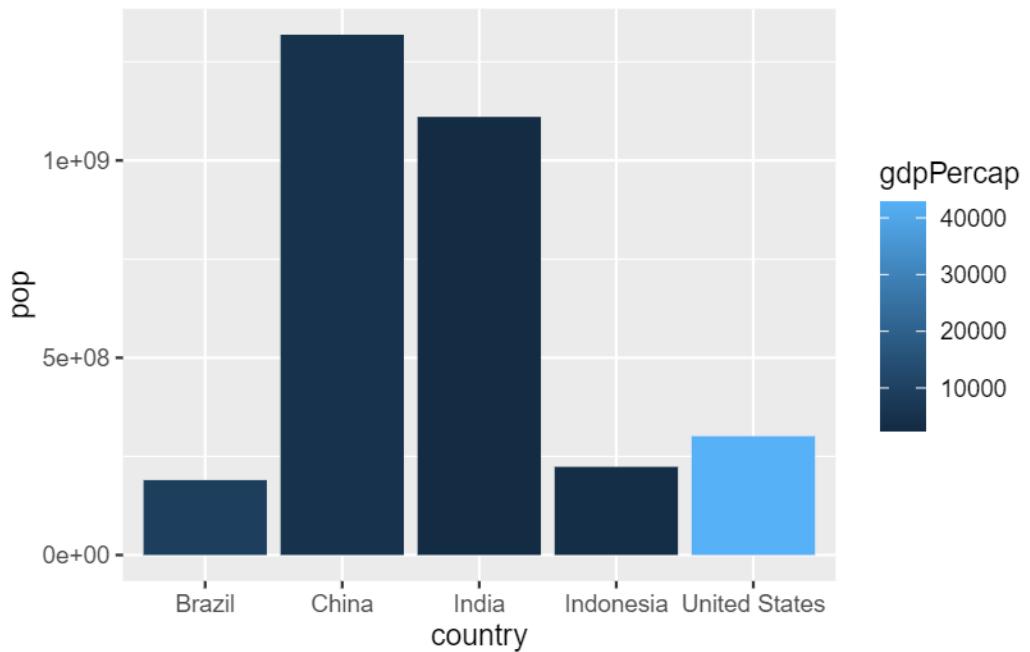


```
ggplot(gapminder_top5) +  
  geom_col(aes(x=country, y=pop, fill= lifeExp))
```



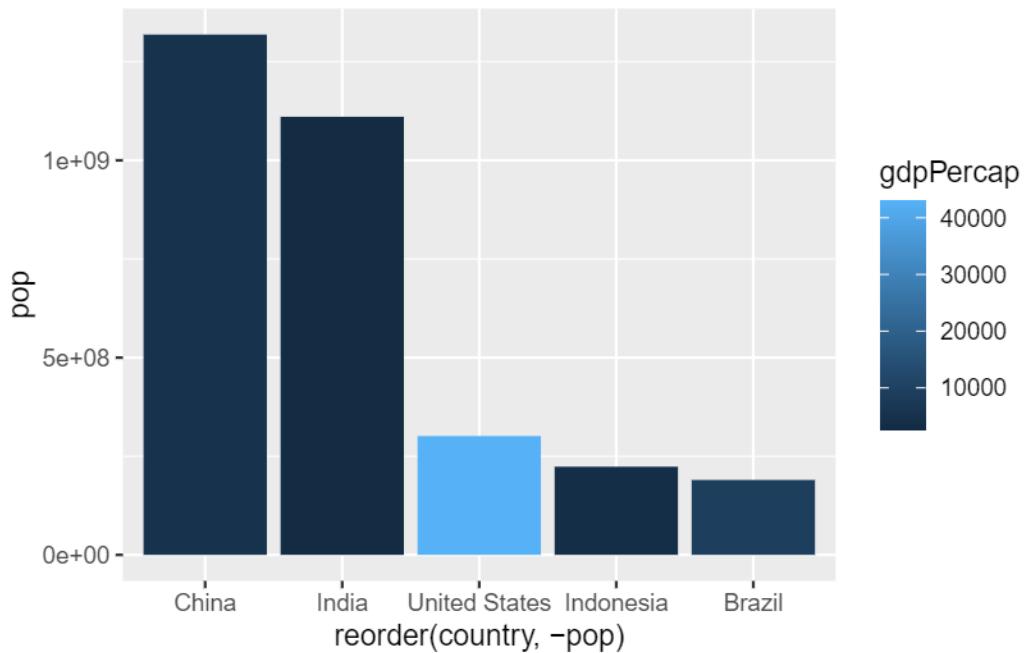
```
#plot population by country
```

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

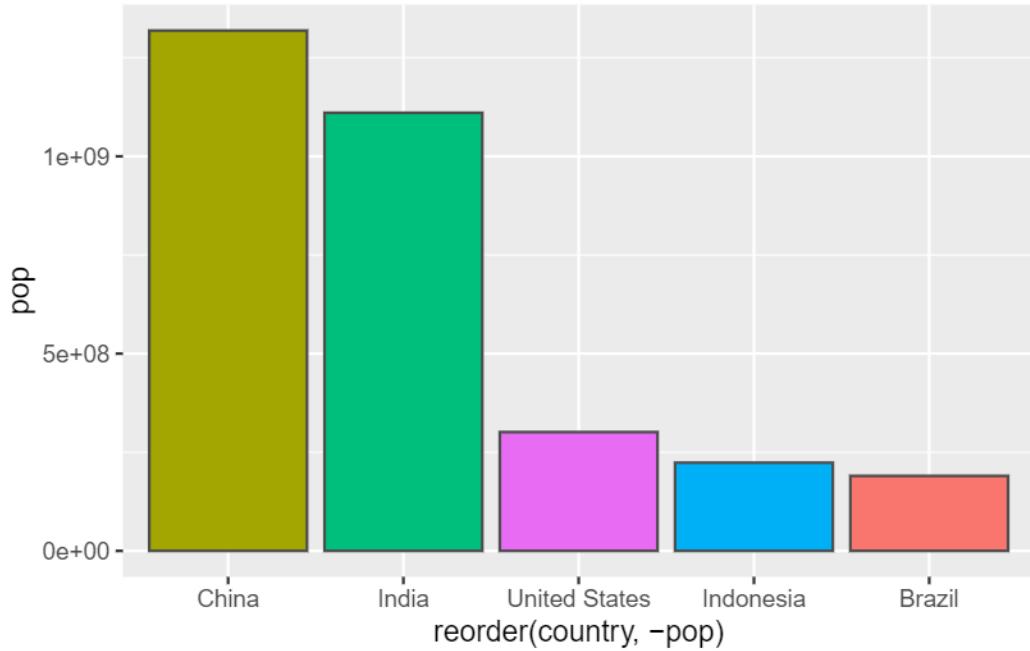


```
#Changing the order of bars
```

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

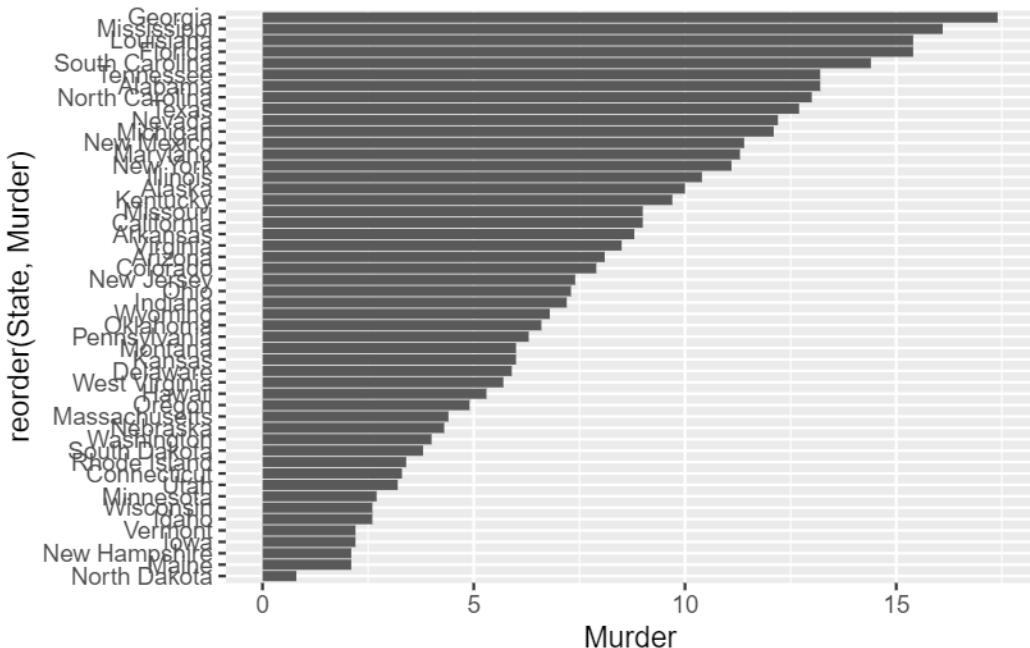


```
ggplot(gapminder_top5) +  
  aes(x=reorder(country, -pop), y=pop, fill=country) +  
  geom_col(col="gray30") +  
  guides(fill="none")
```

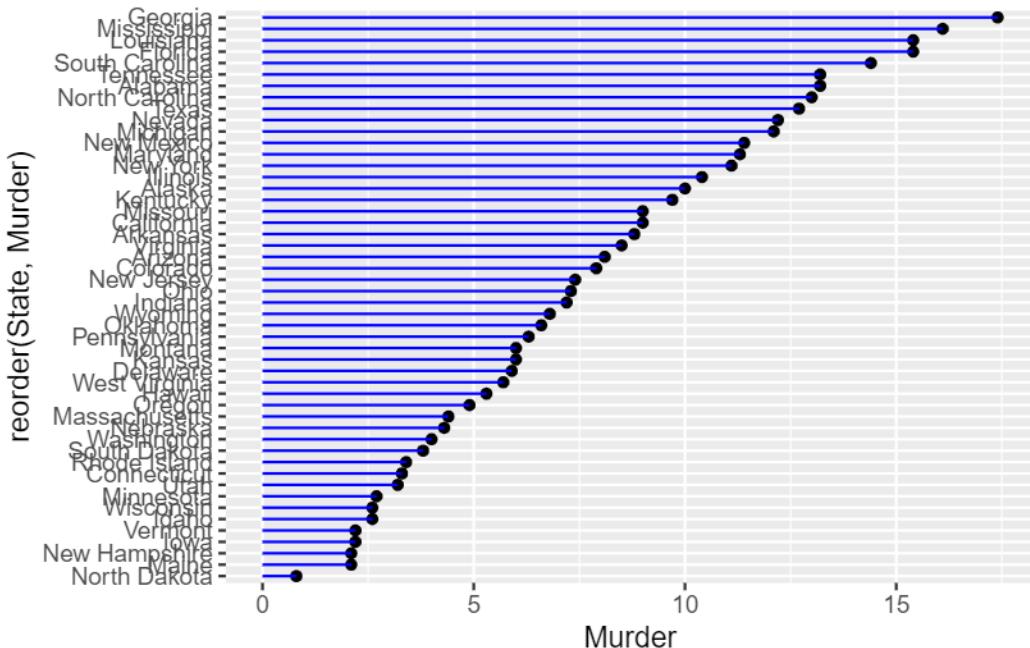


#Flipping bar charts

```
#head(USArrests)
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(aes(x=State,
                   xend=State,
                   y=0,
                   yend=Murder), color="blue") +
  coord_flip()
```



```
#Extensions: Animation #Yah!!! my animation worked after installing suggested packages by R. Hurray!!
```

```
library(gapminder)
```

```
Attaching package: 'gapminder'
```

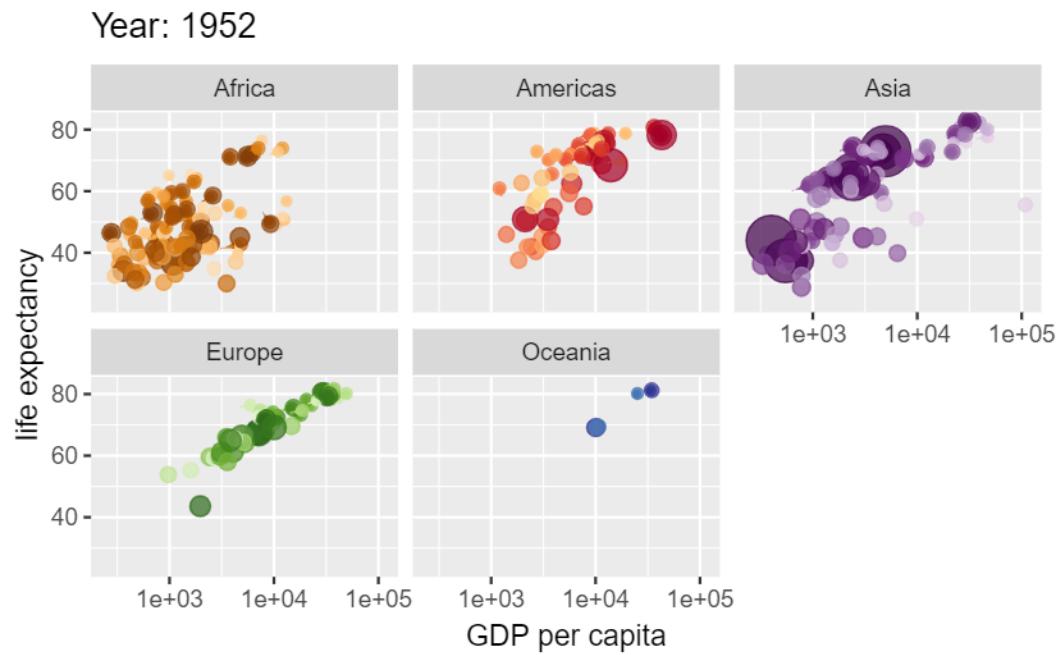
```
The following object is masked _by_ '.GlobalEnv':
```

```
gapminder
```

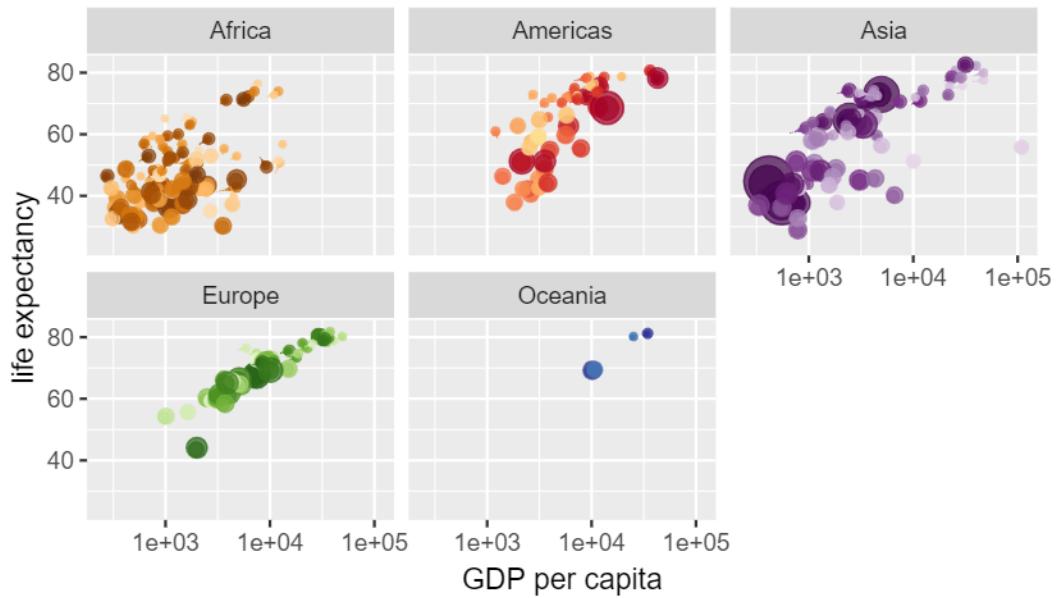
```
library(gganimate)
```

```
# Setup nice regular ggplot of the gapminder data
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) +
```

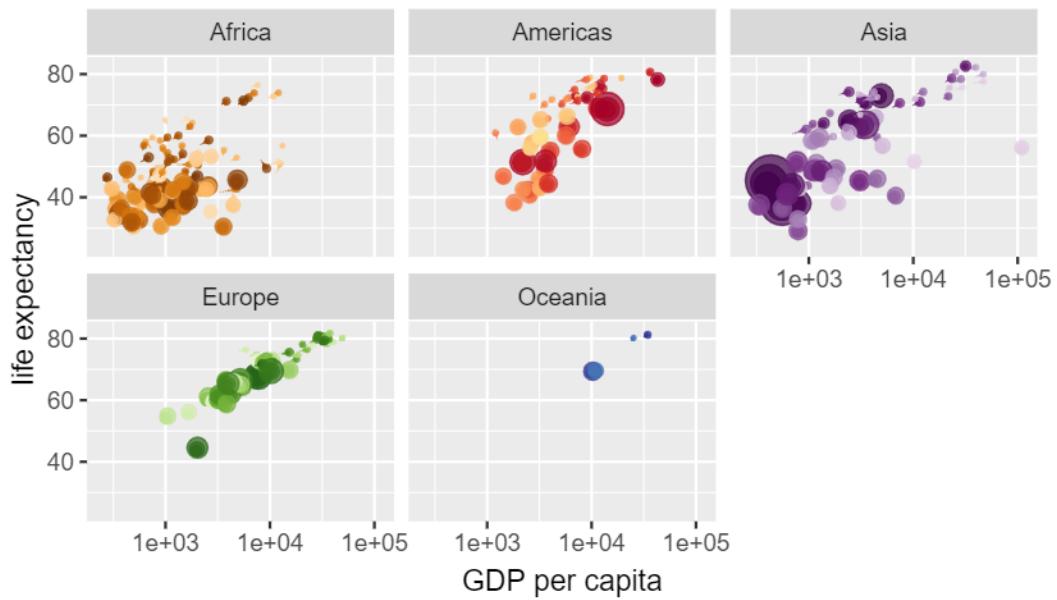
```
# Here comes the gganimate specific bits
labs(title = 'Year: {frame_time}', x = 'GDP per capita', y = 'life expectancy') +
transition_time(year) +
shadow_wake(wake_length = 0.1, alpha = FALSE)
```



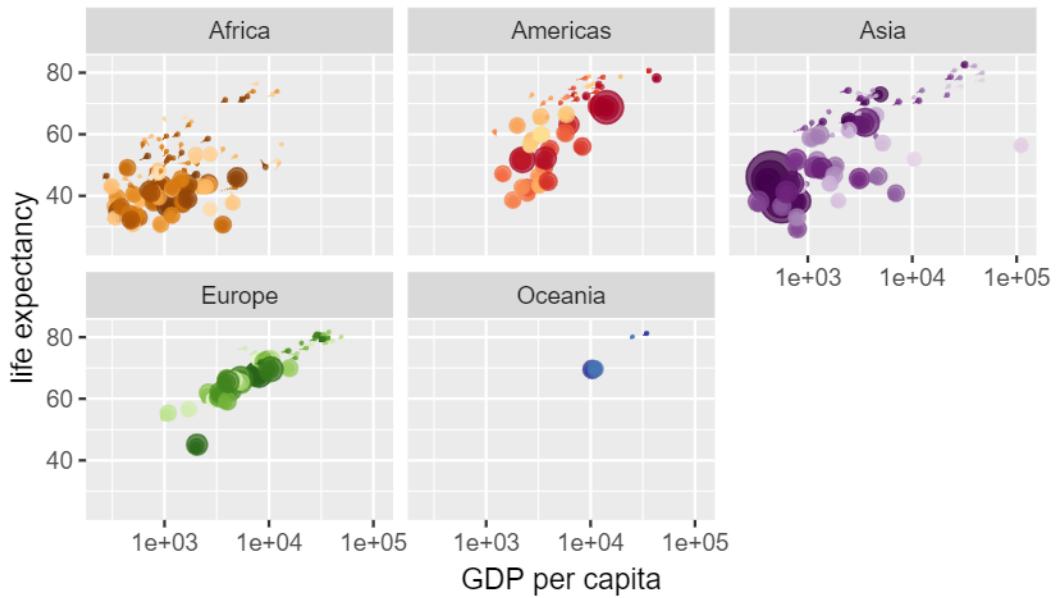
Year: 1953



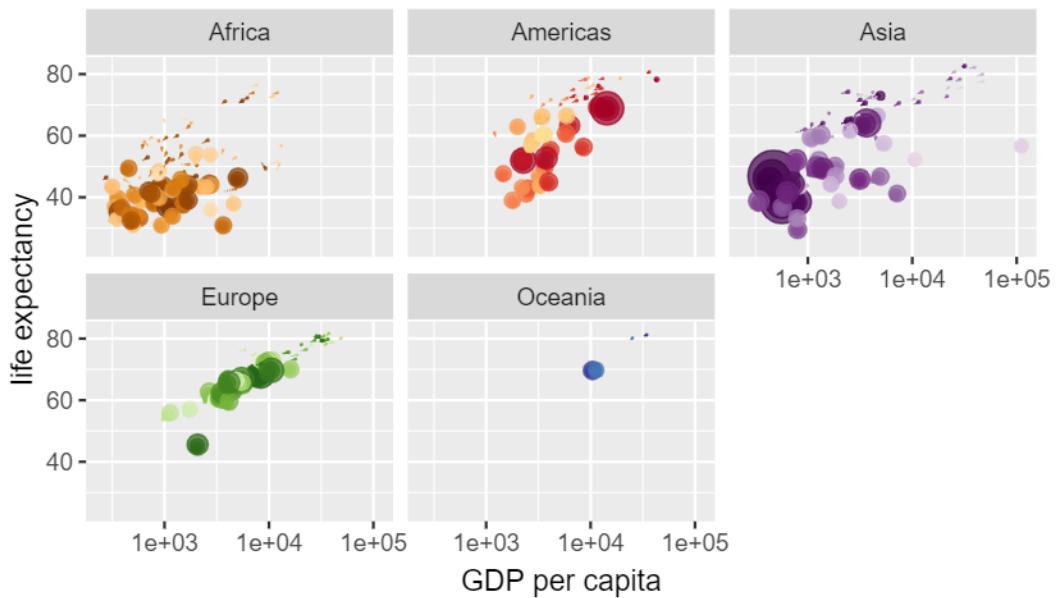
Year: 1953



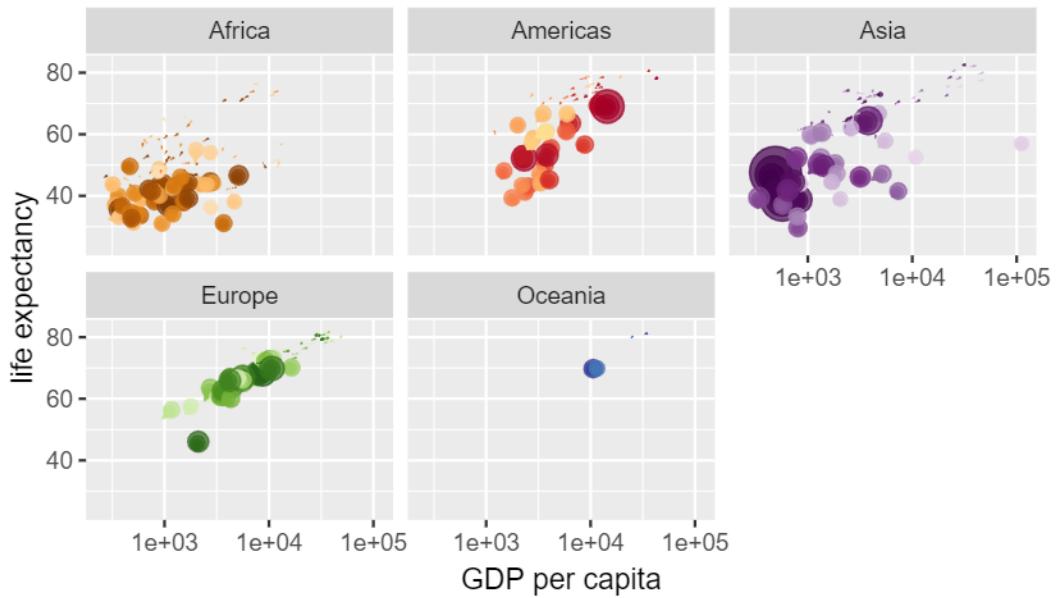
Year: 1954



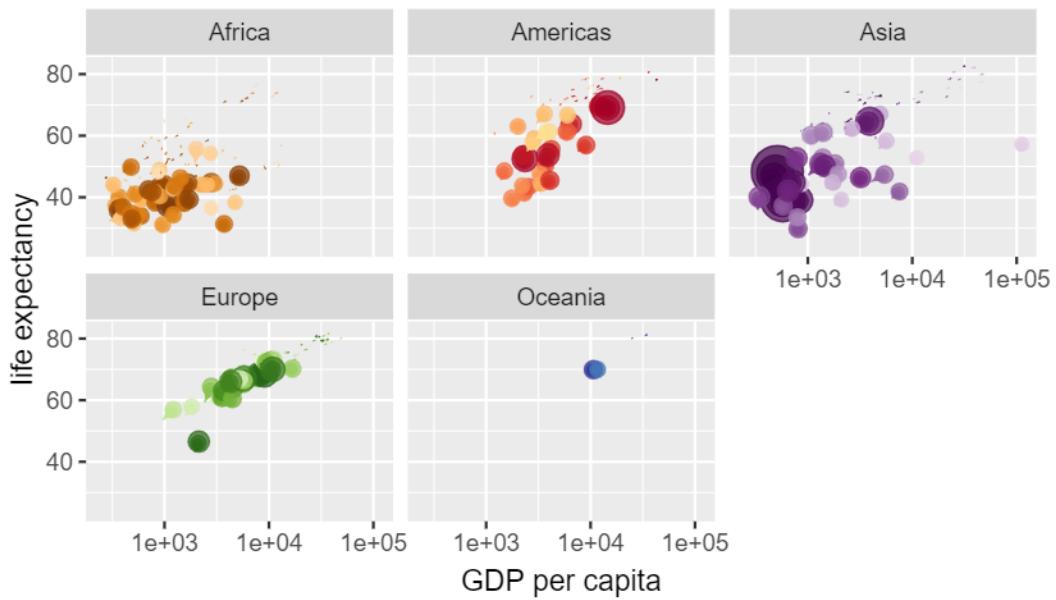
Year: 1954



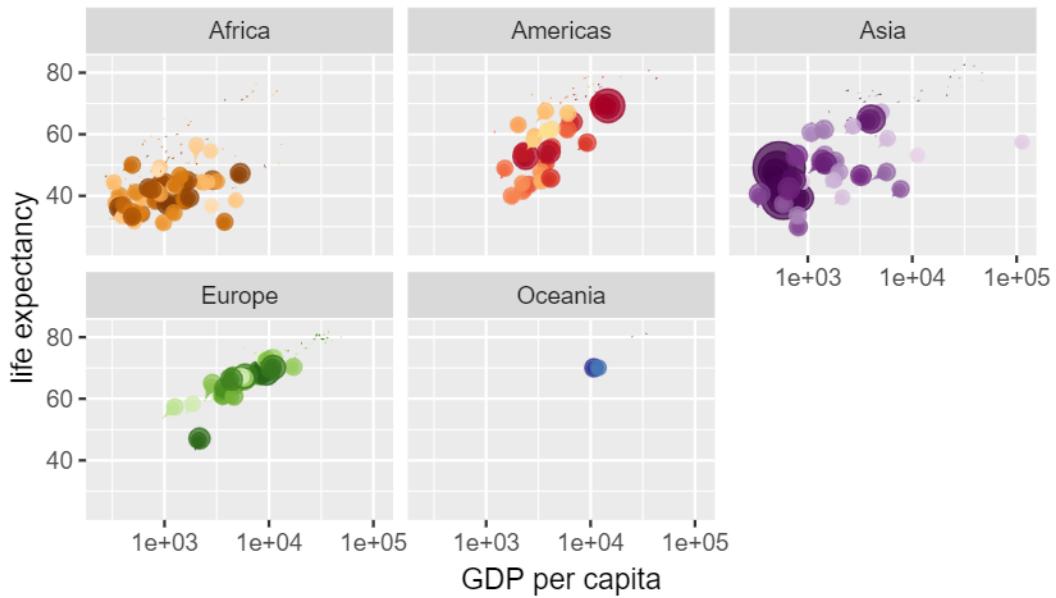
Year: 1955



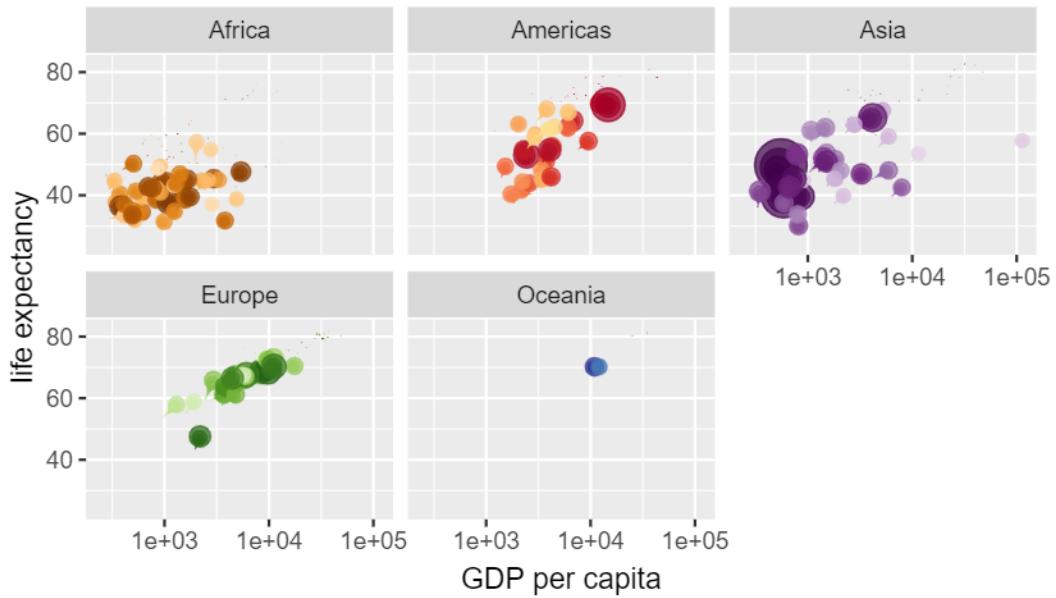
Year: 1955



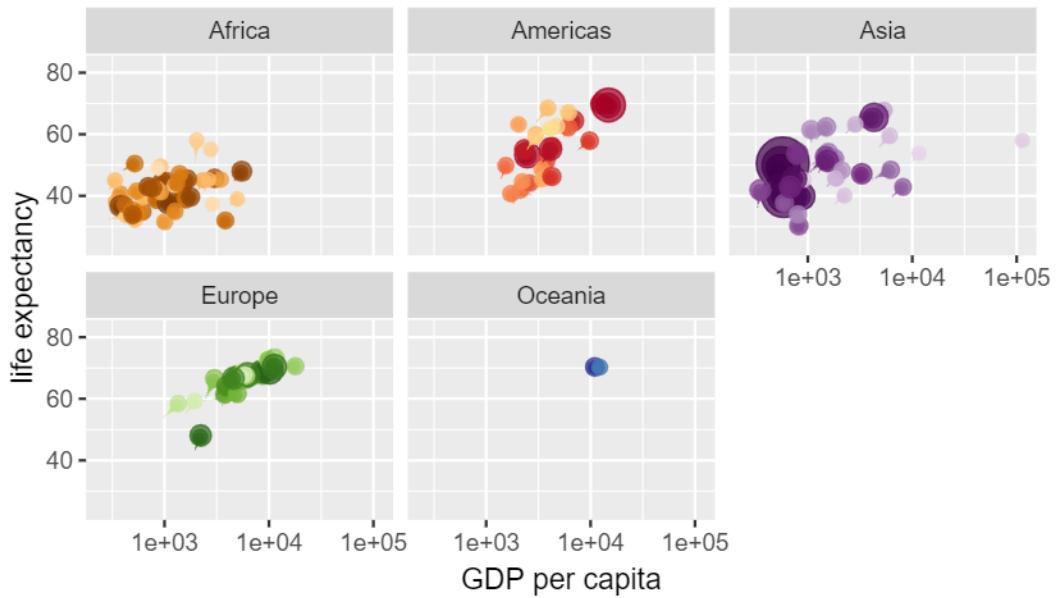
Year: 1956



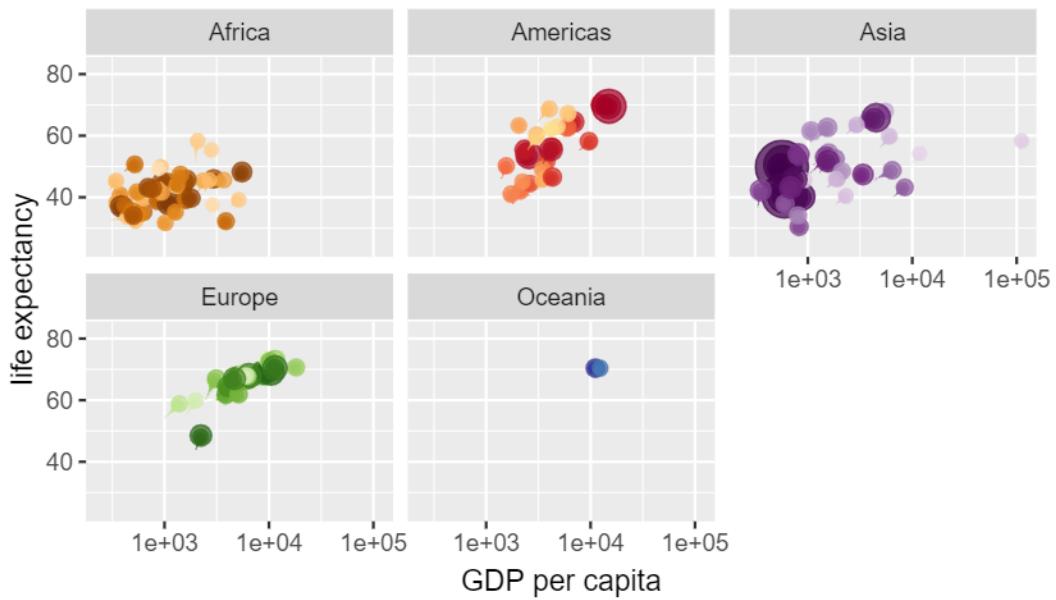
Year: 1956



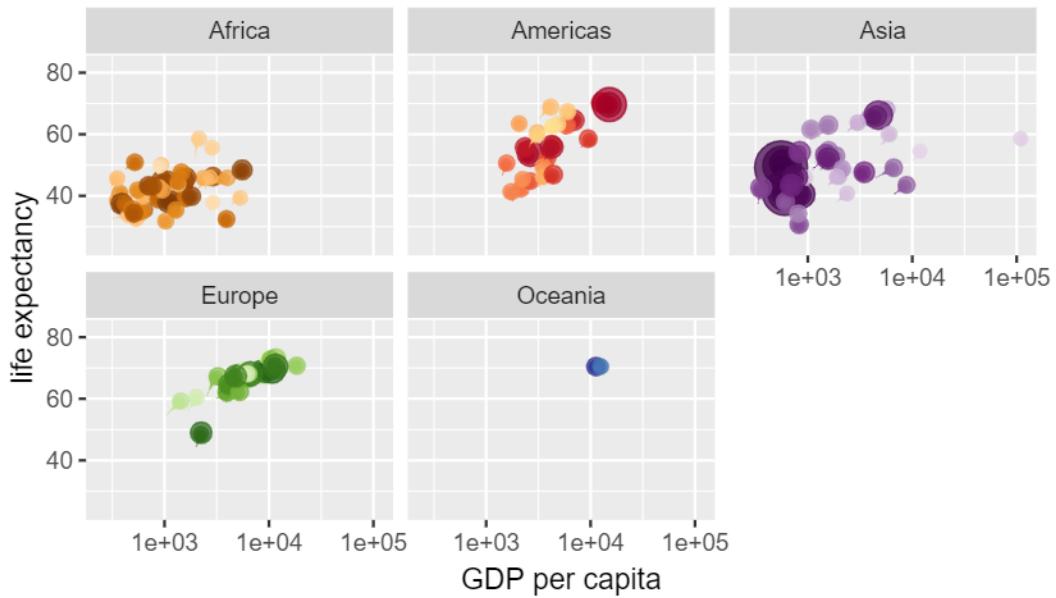
Year: 1957



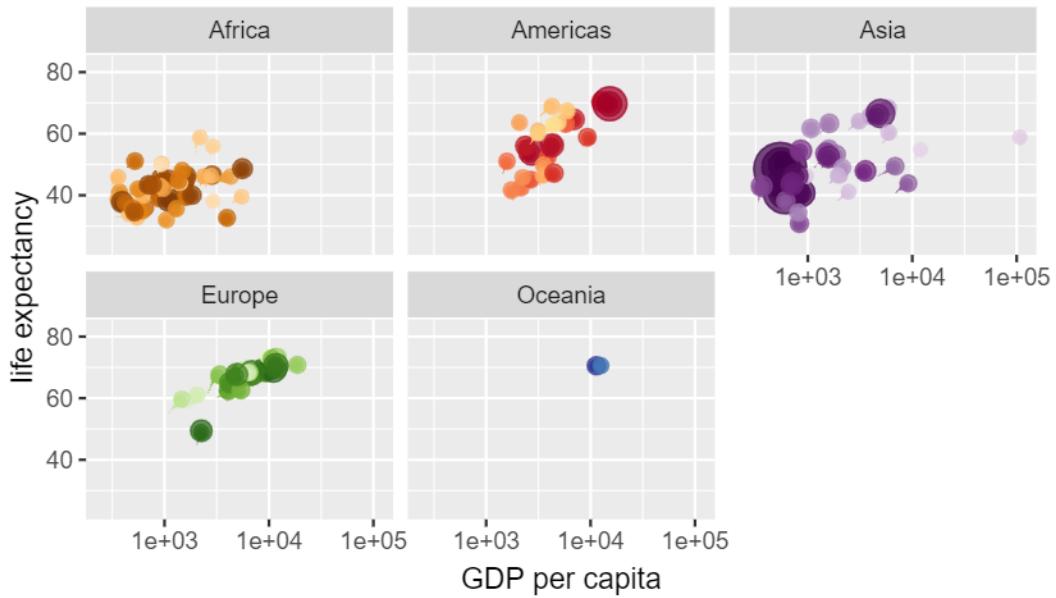
Year: 1958



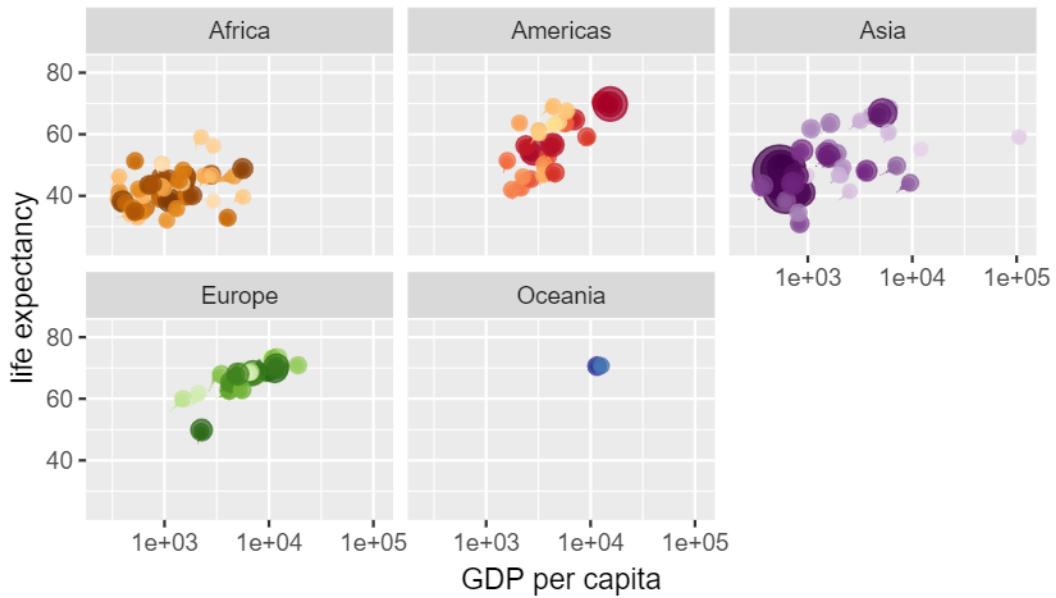
Year: 1958



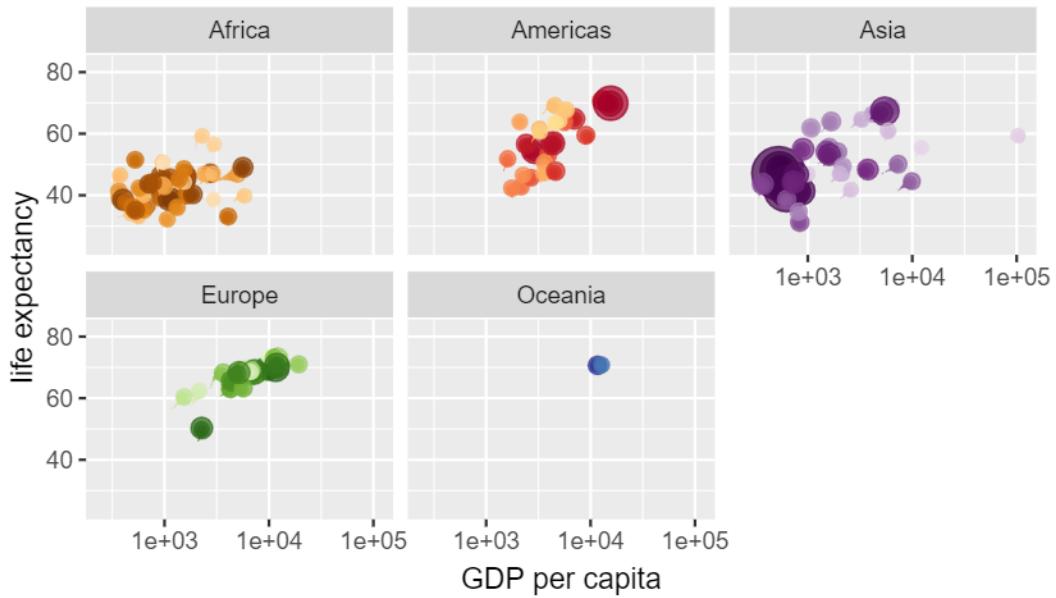
Year: 1959



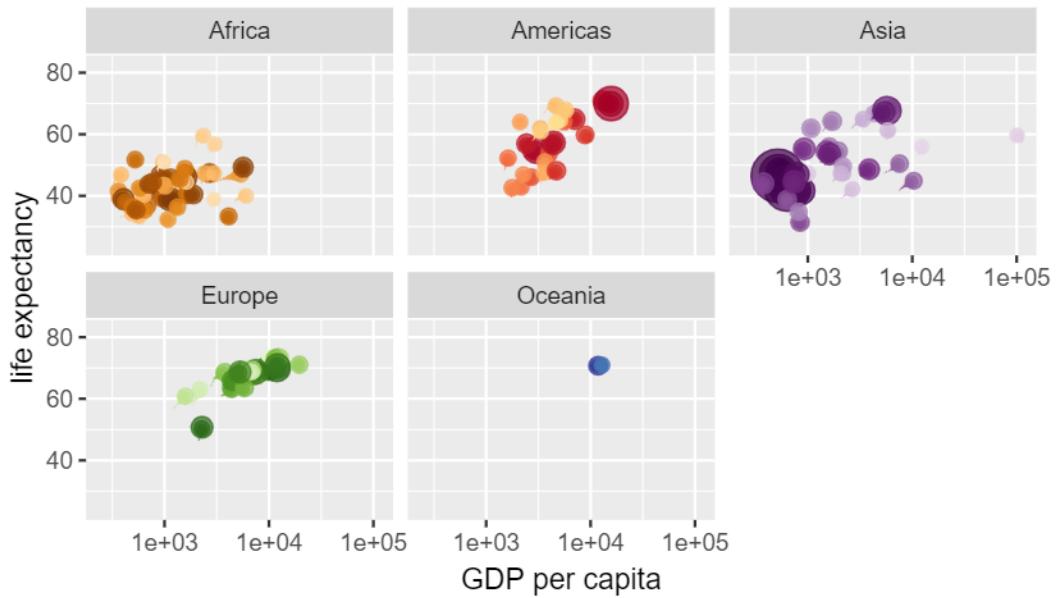
Year: 1959



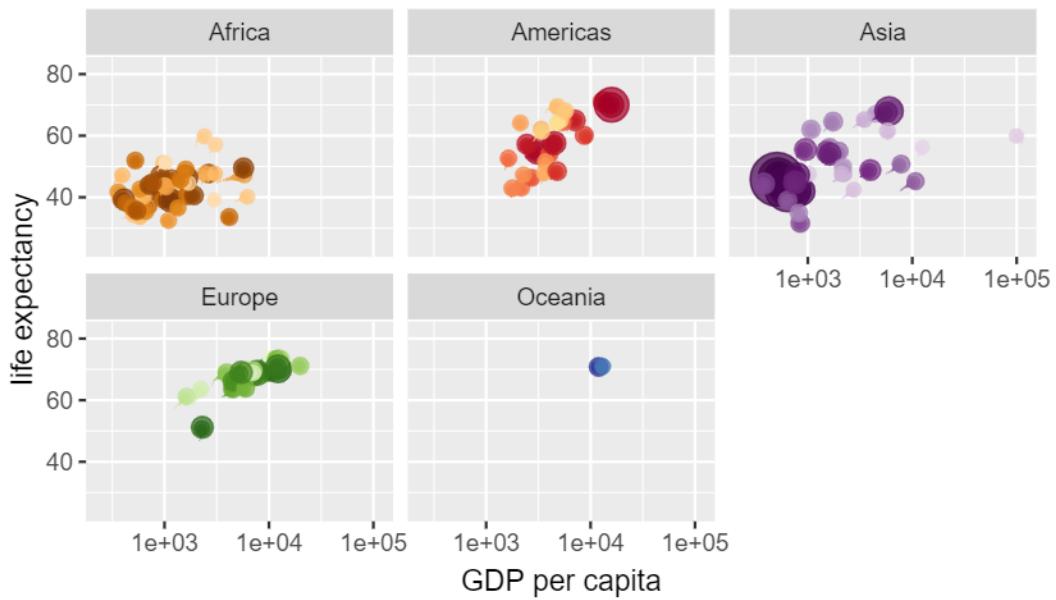
Year: 1960



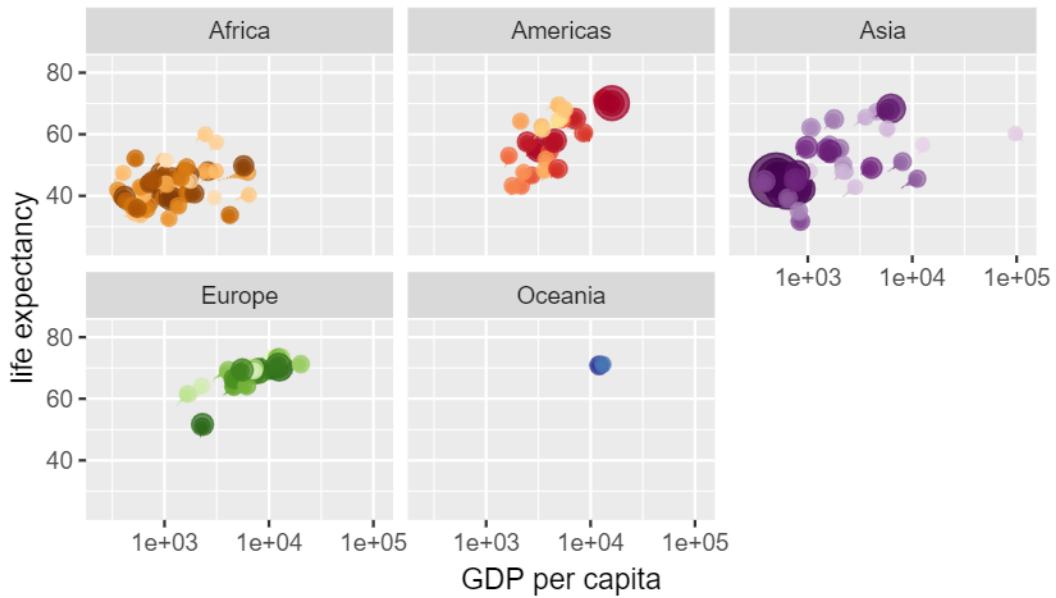
Year: 1960



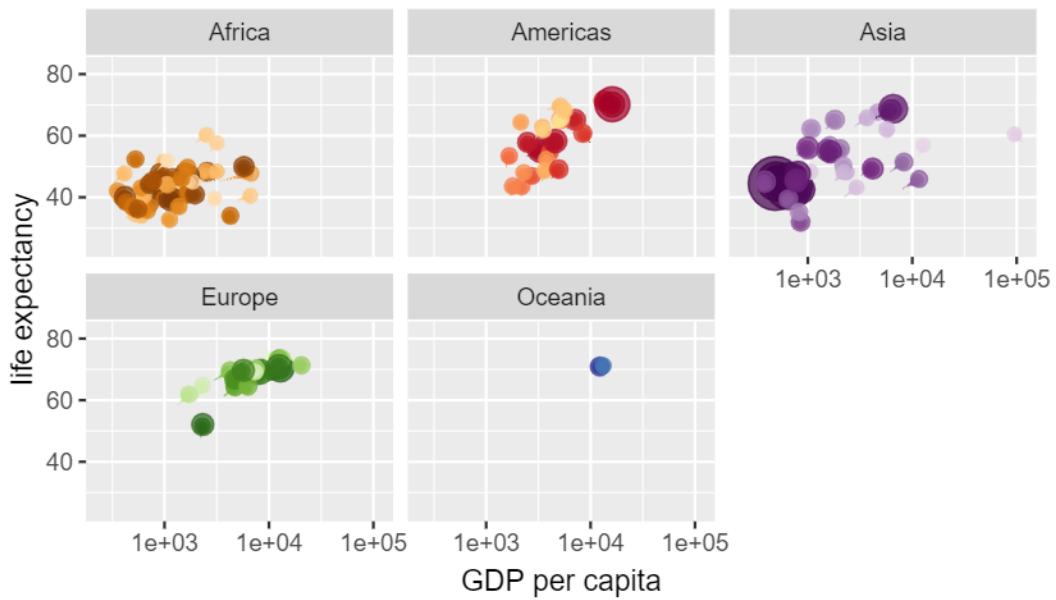
Year: 1961



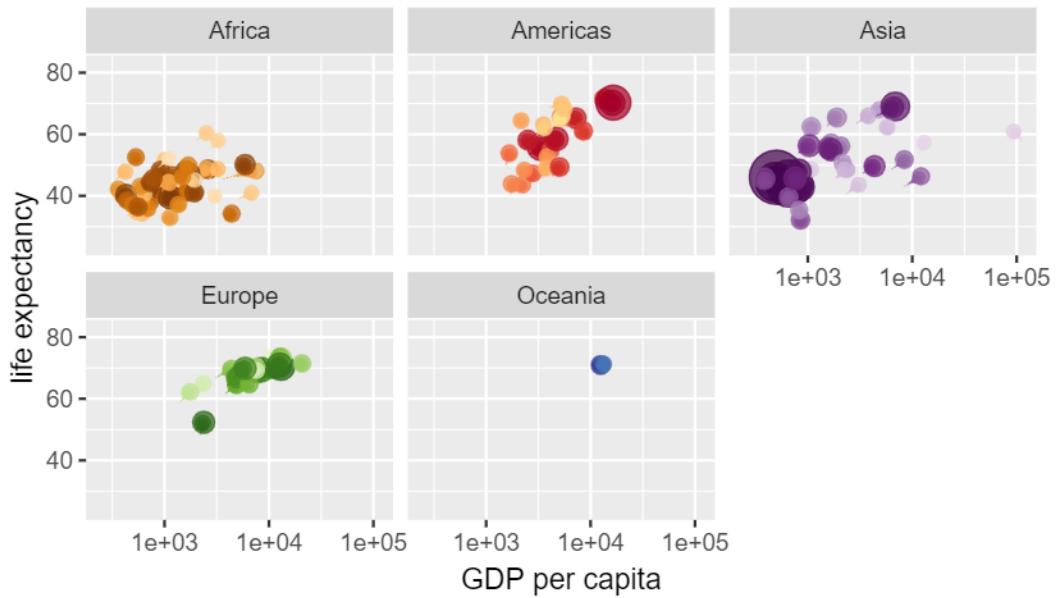
Year: 1961



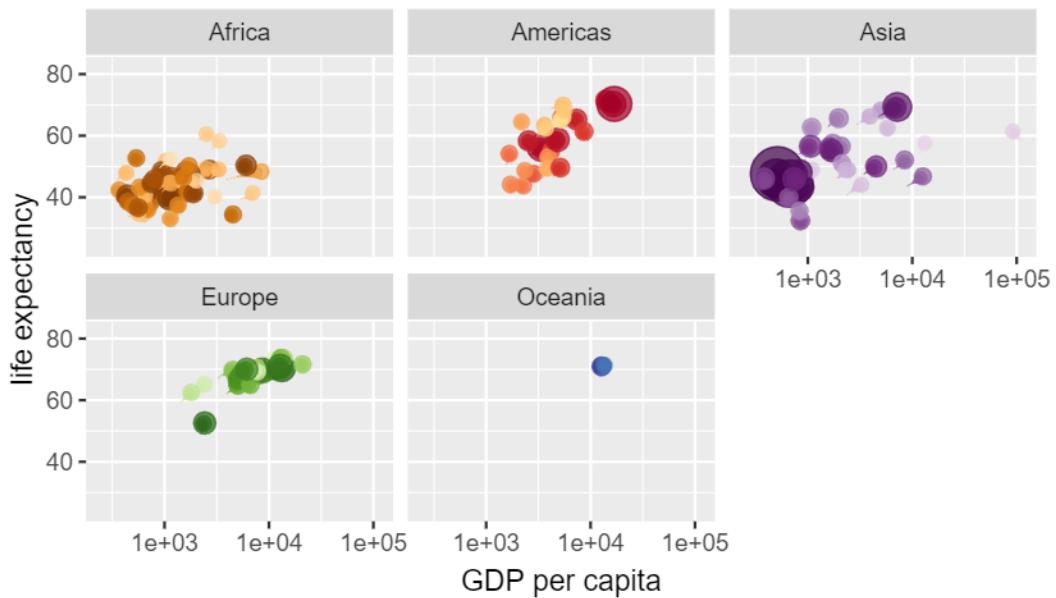
Year: 1962



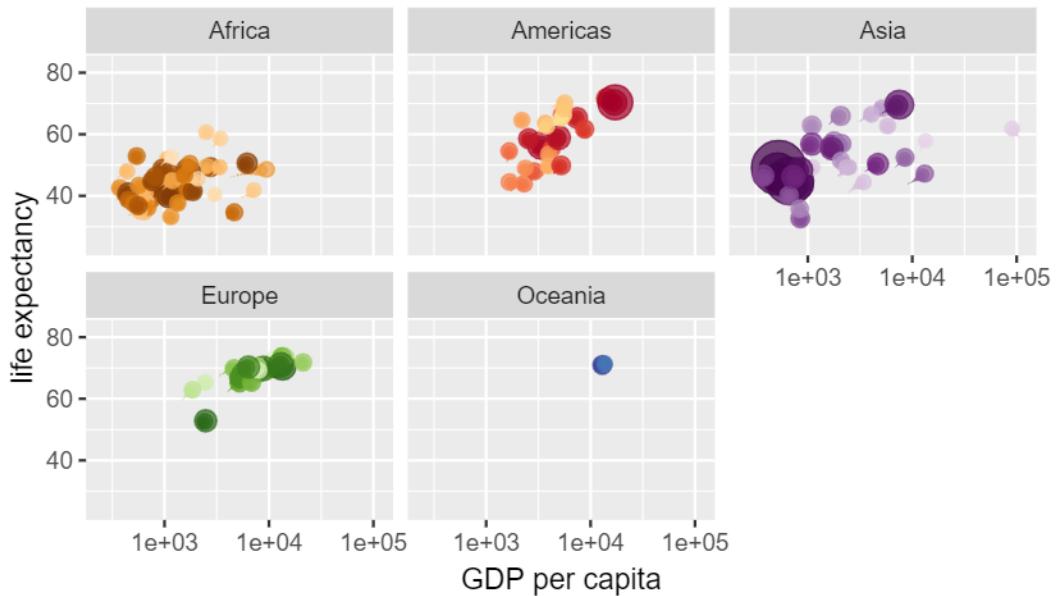
Year: 1963



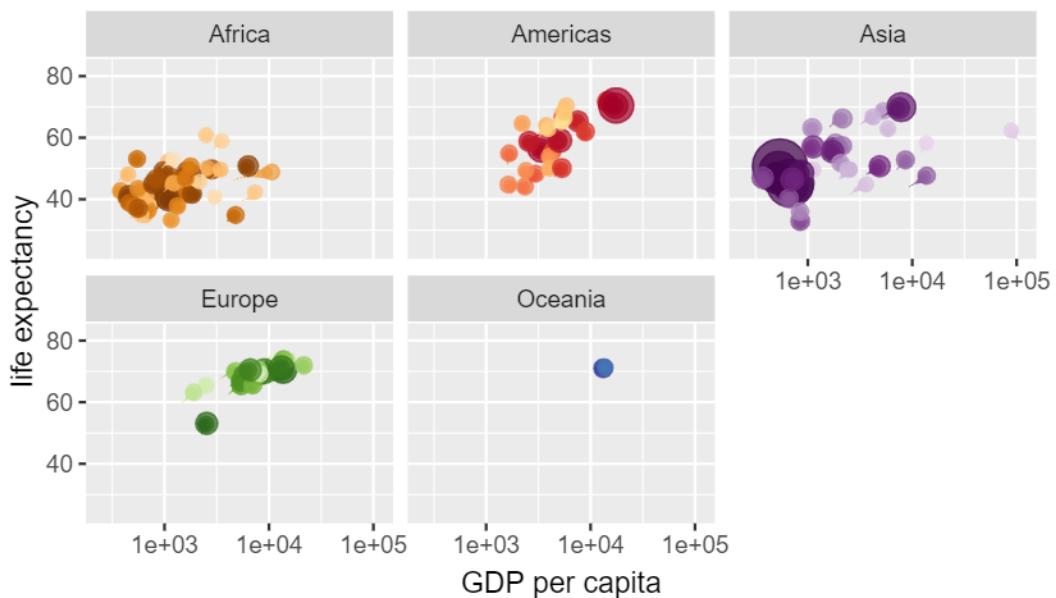
Year: 1963



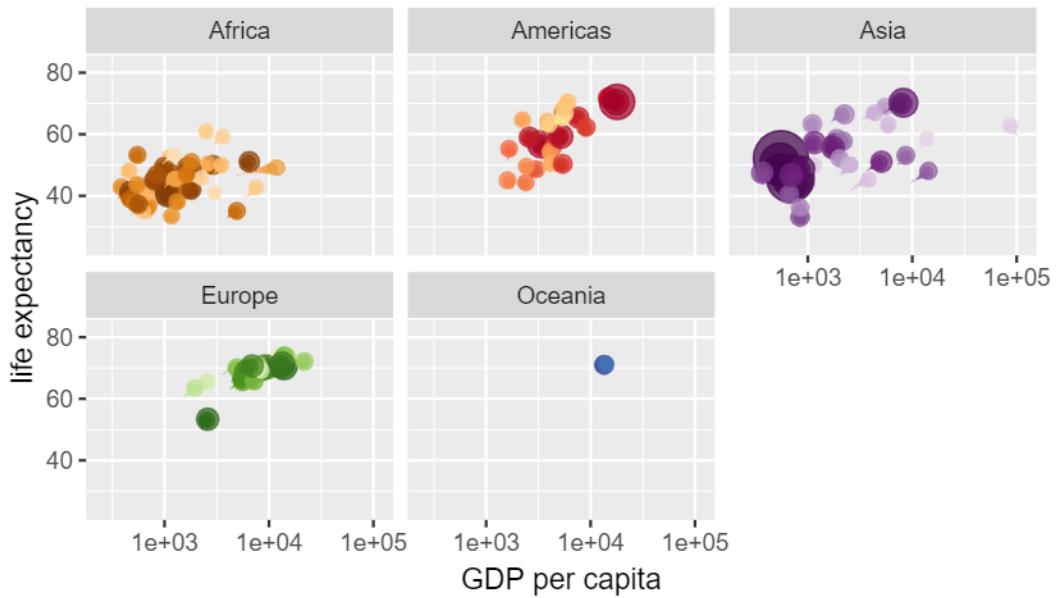
Year: 1964



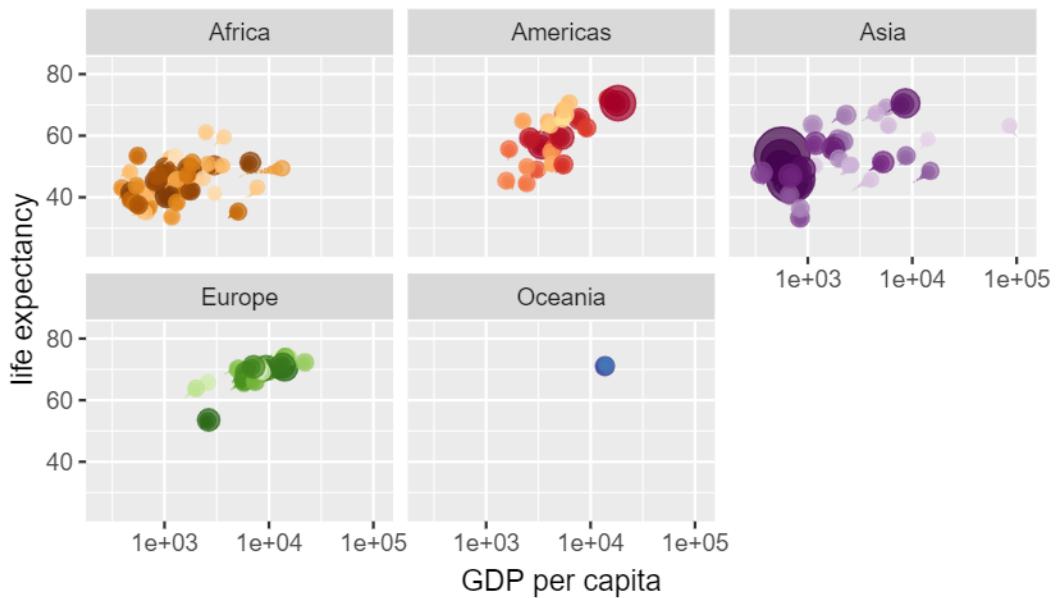
Year: 1964



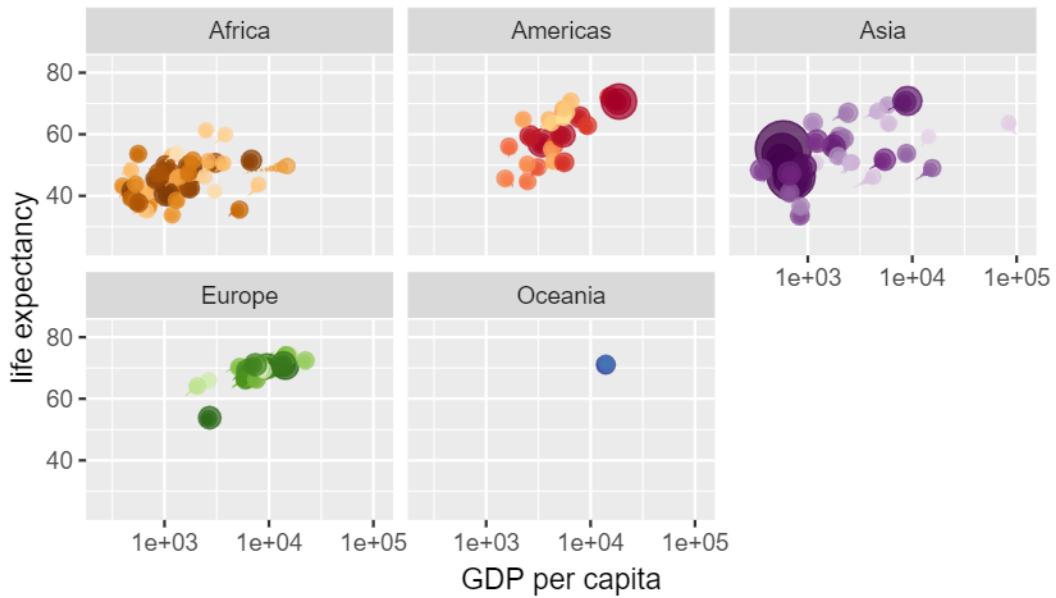
Year: 1965



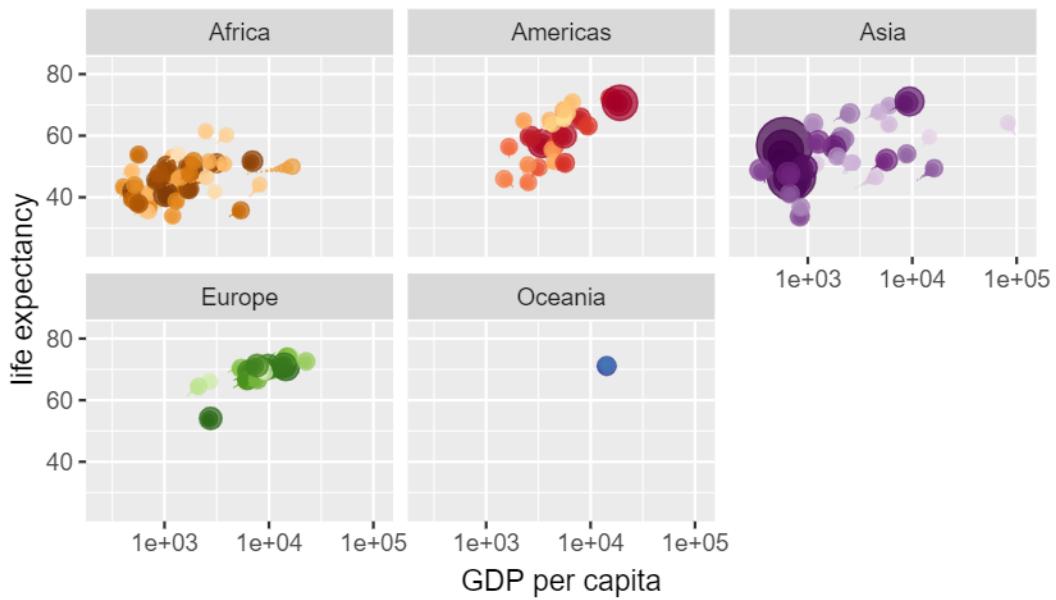
Year: 1965



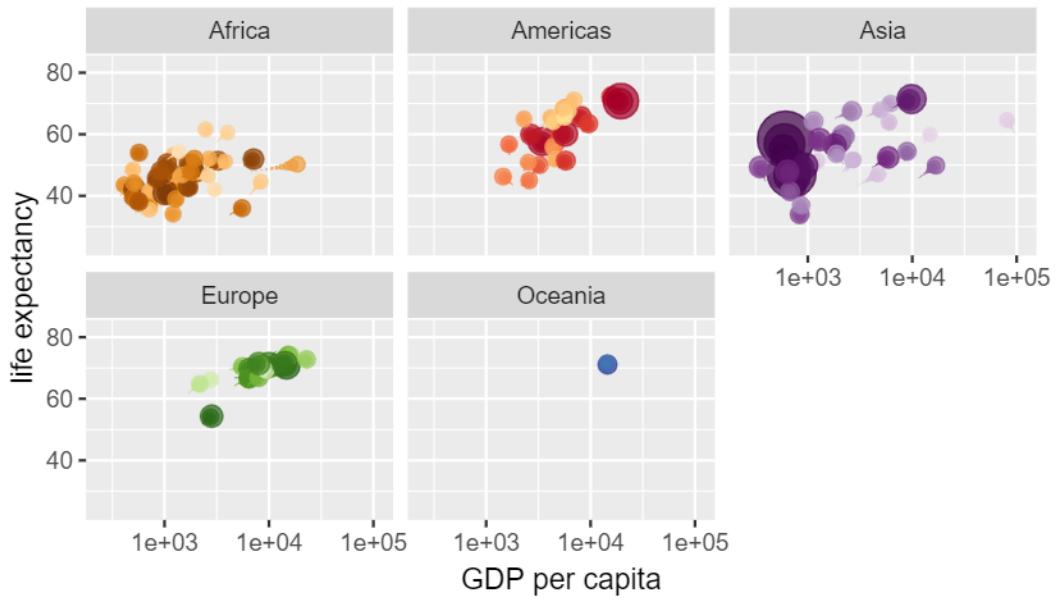
Year: 1966



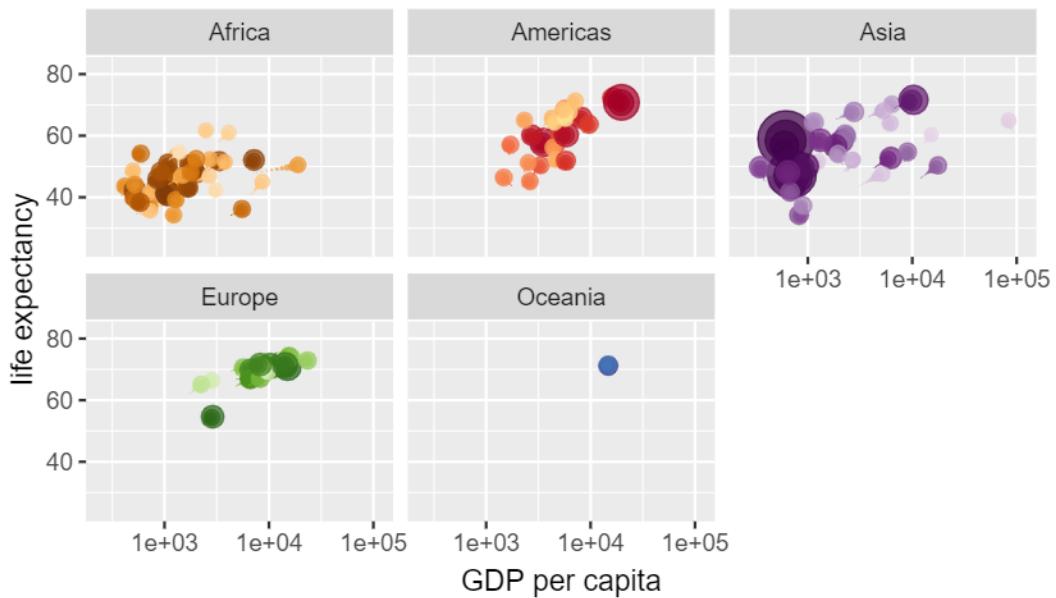
Year: 1966



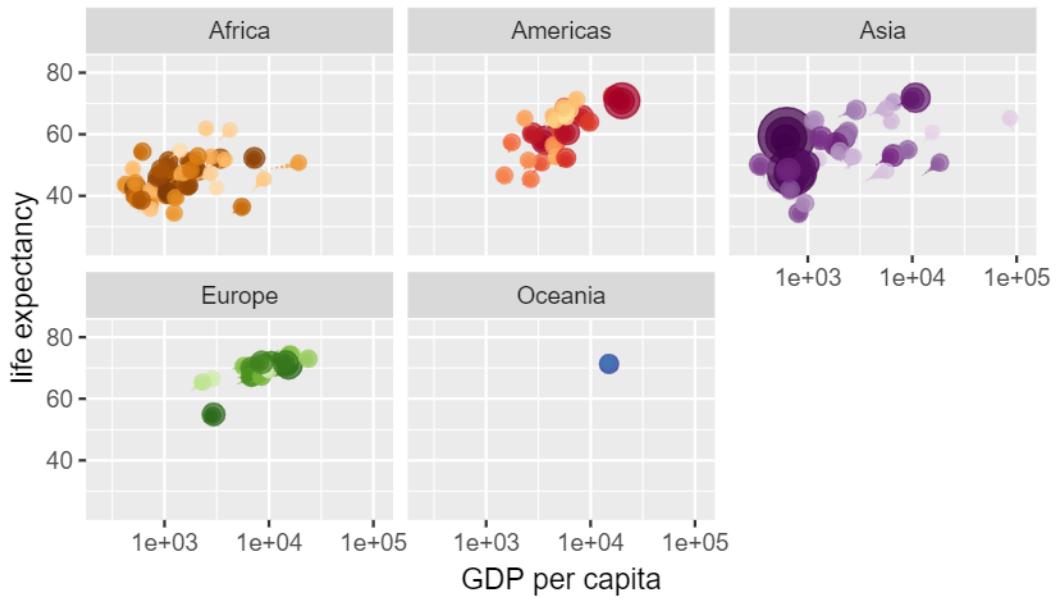
Year: 1967



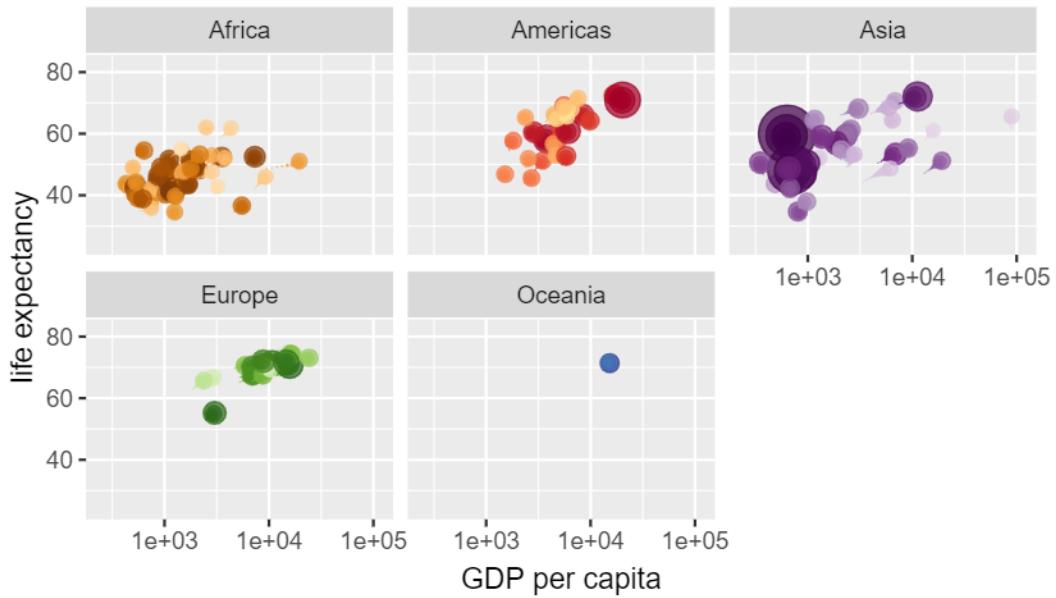
Year: 1968



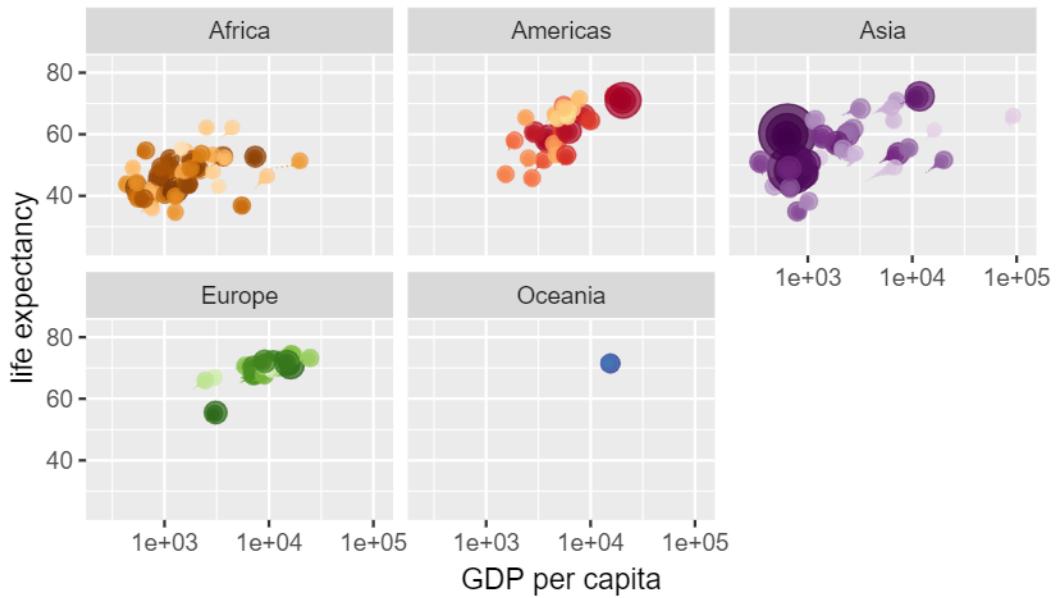
Year: 1968



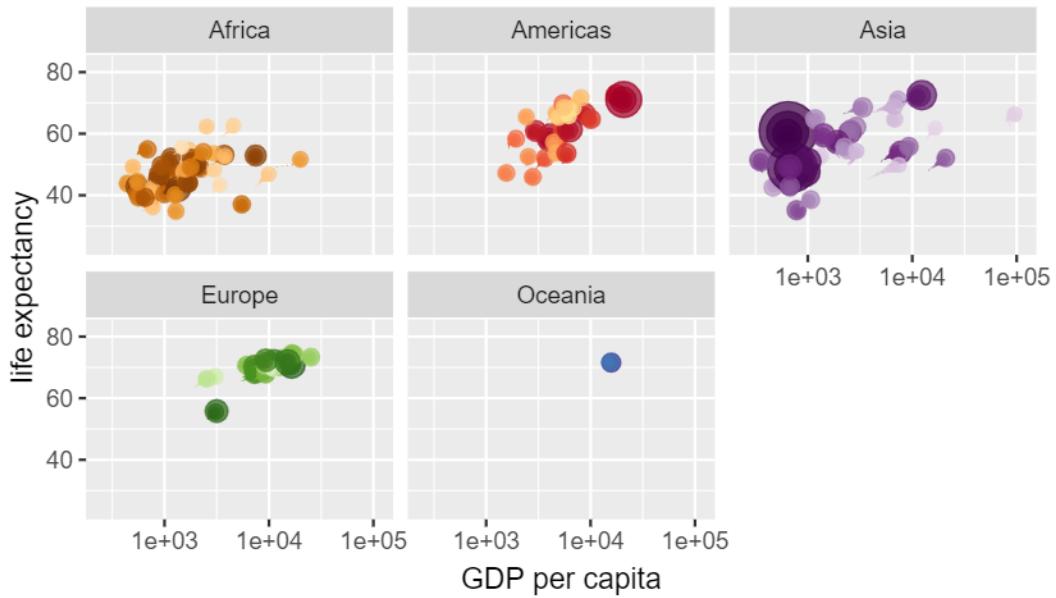
Year: 1969



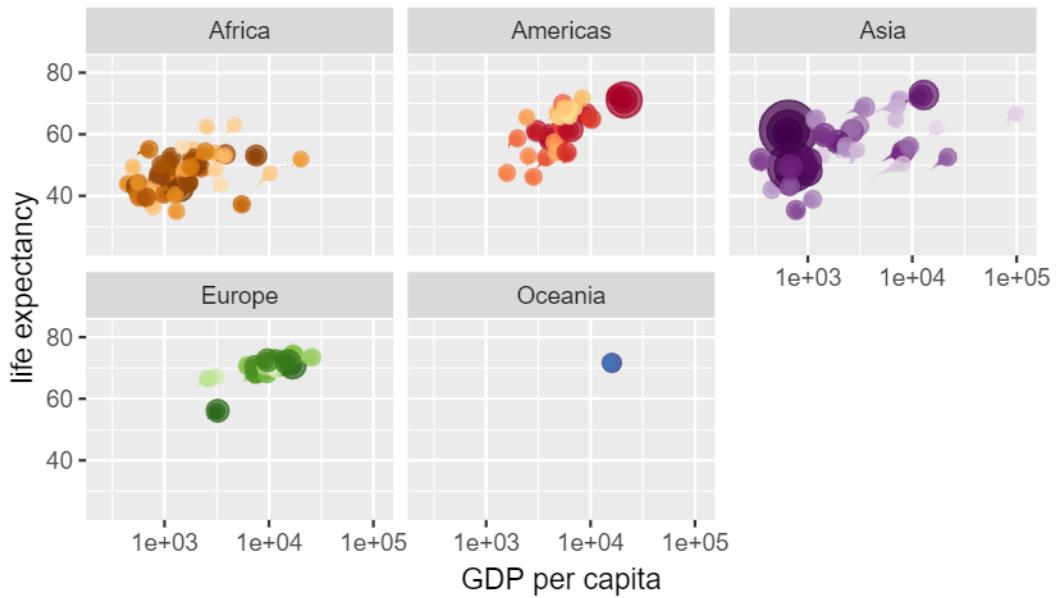
Year: 1969



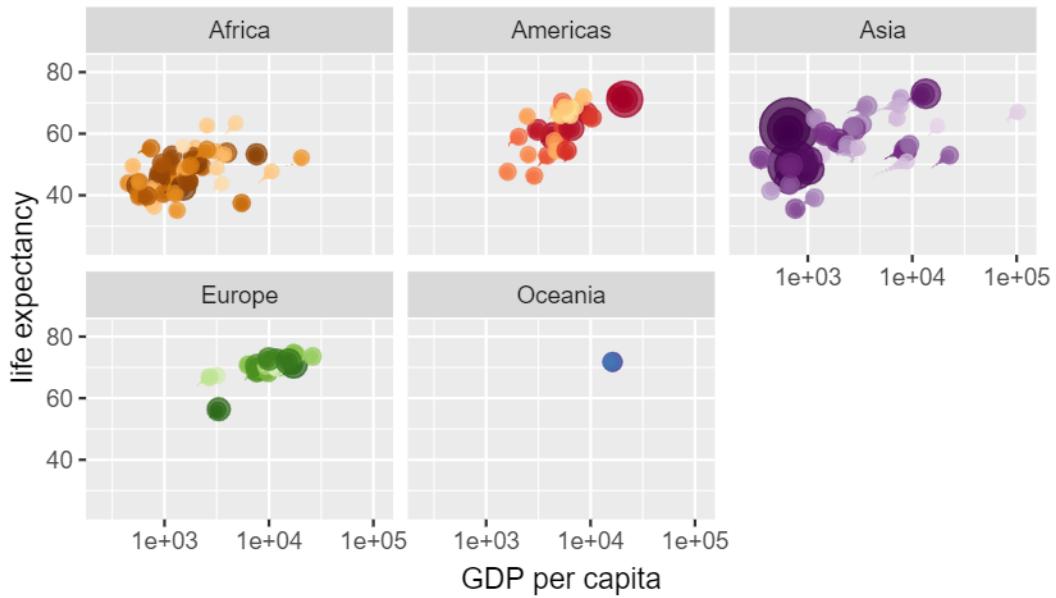
Year: 1970



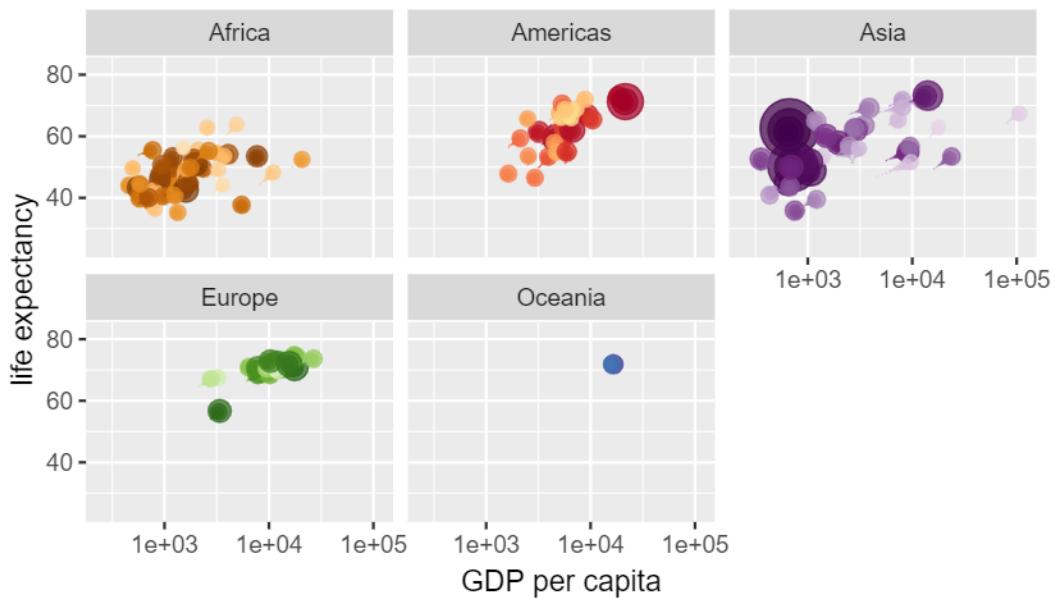
Year: 1970



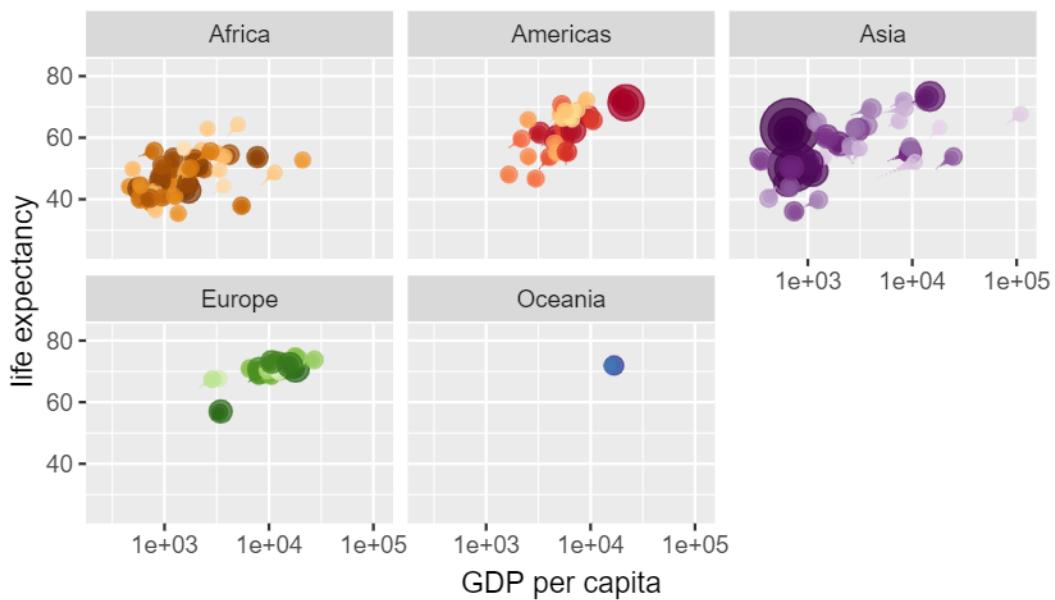
Year: 1971



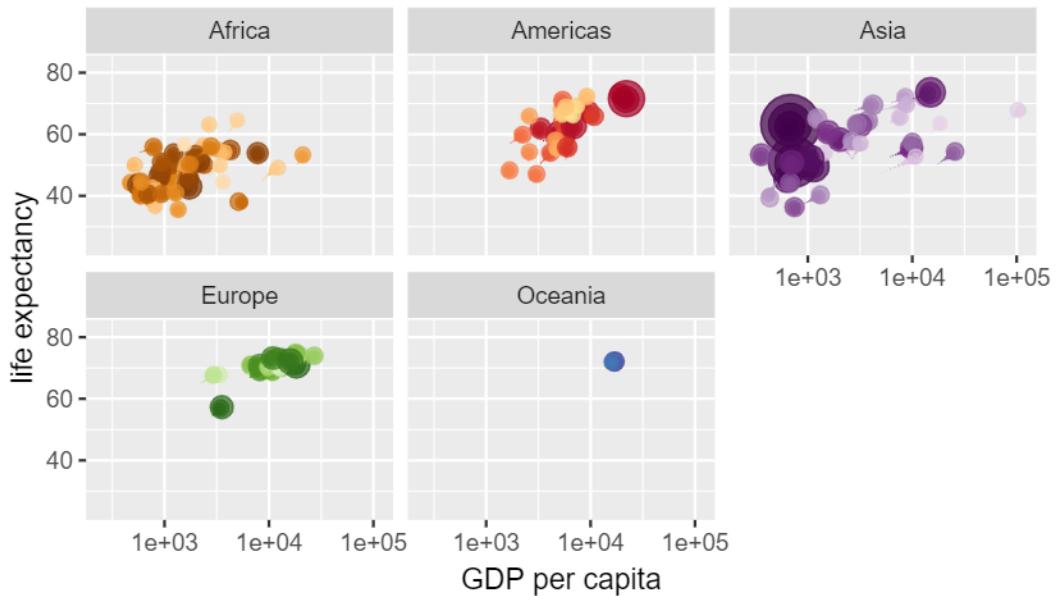
Year: 1971



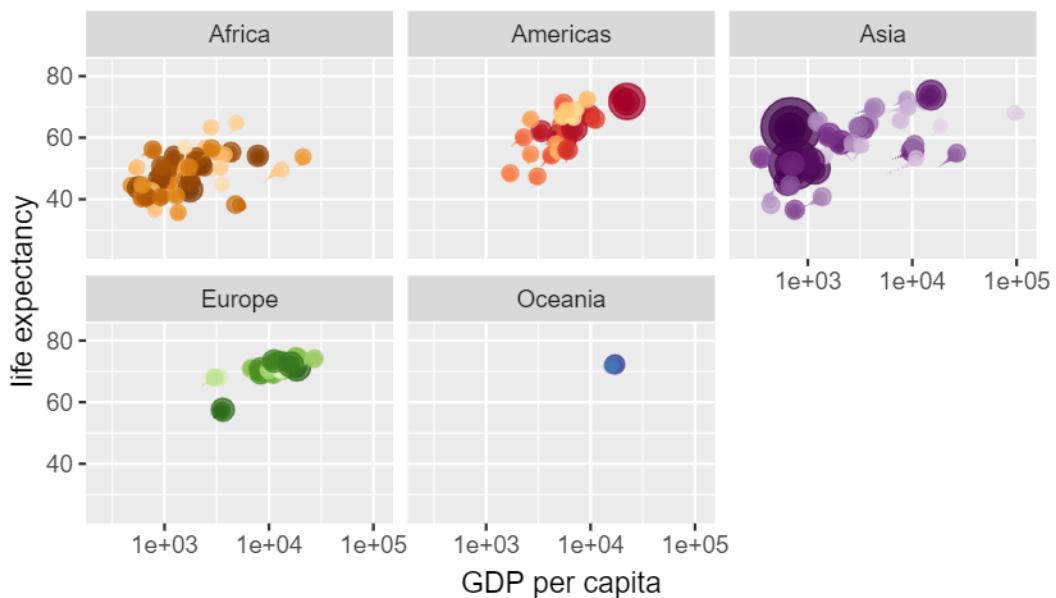
Year: 1972



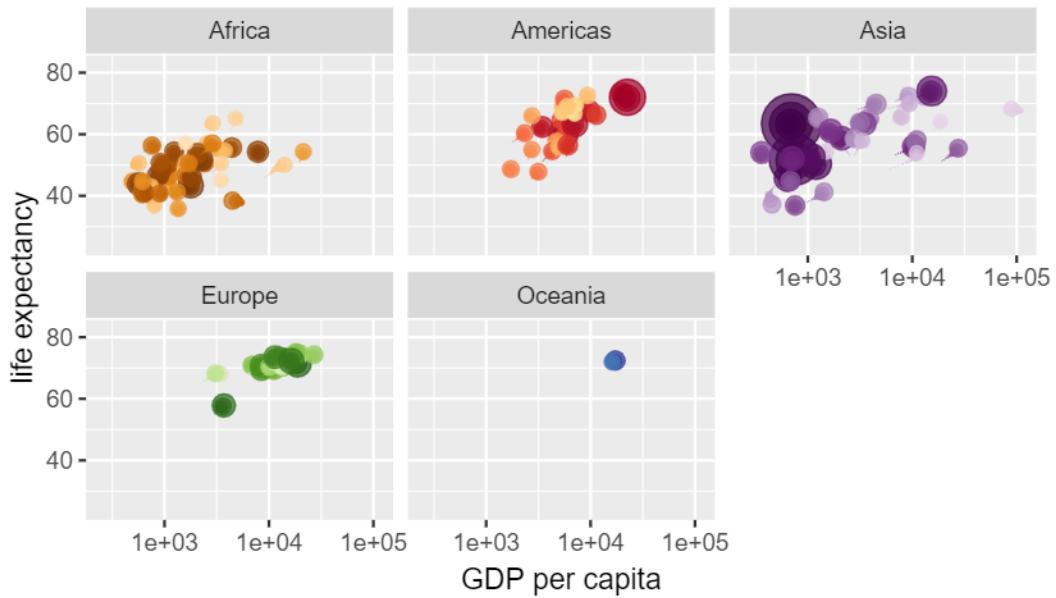
Year: 1973



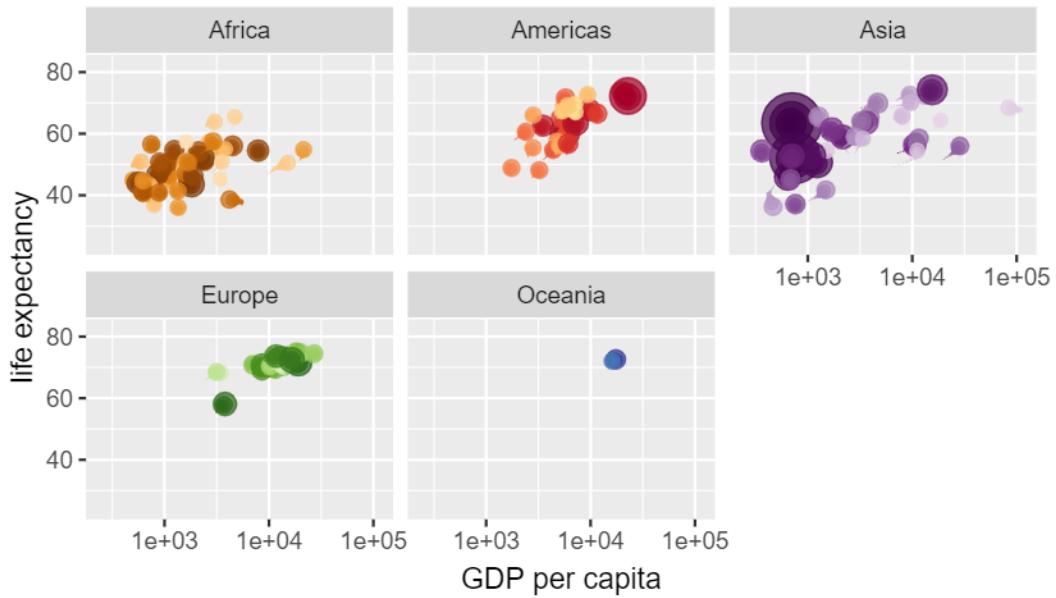
Year: 1973



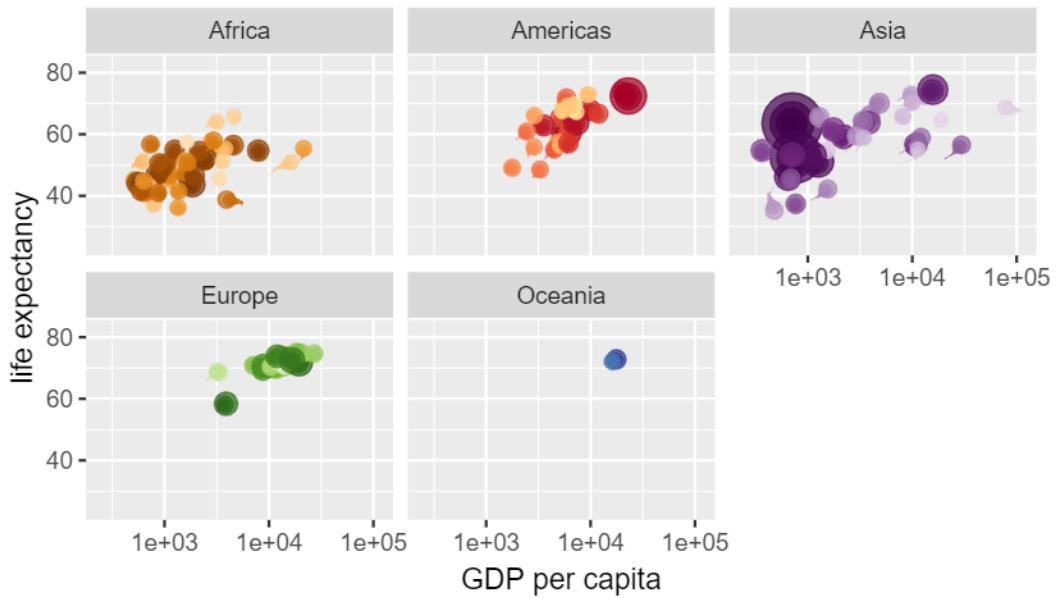
Year: 1974



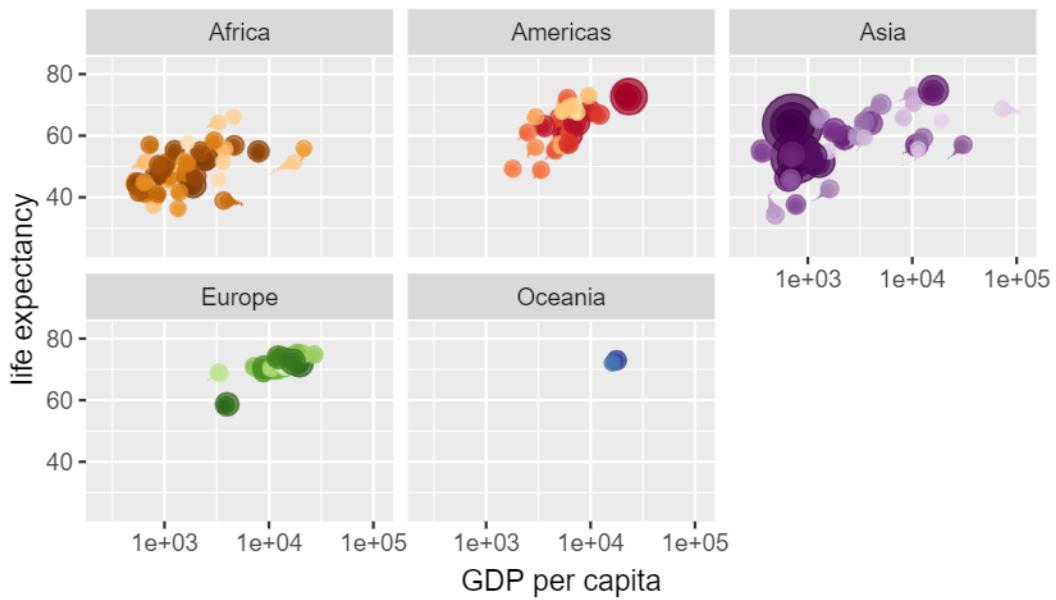
Year: 1974



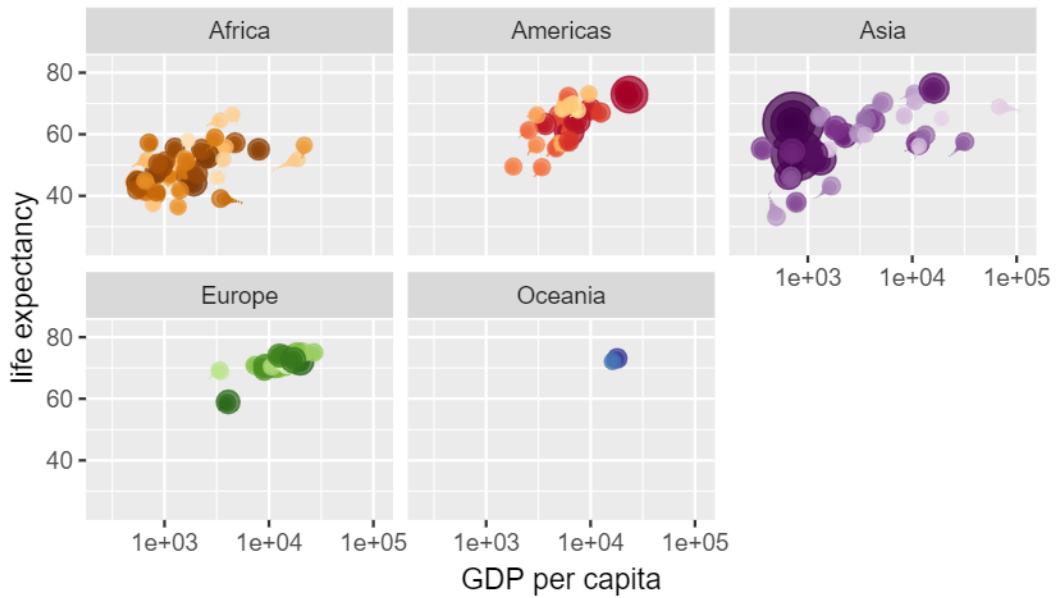
Year: 1975



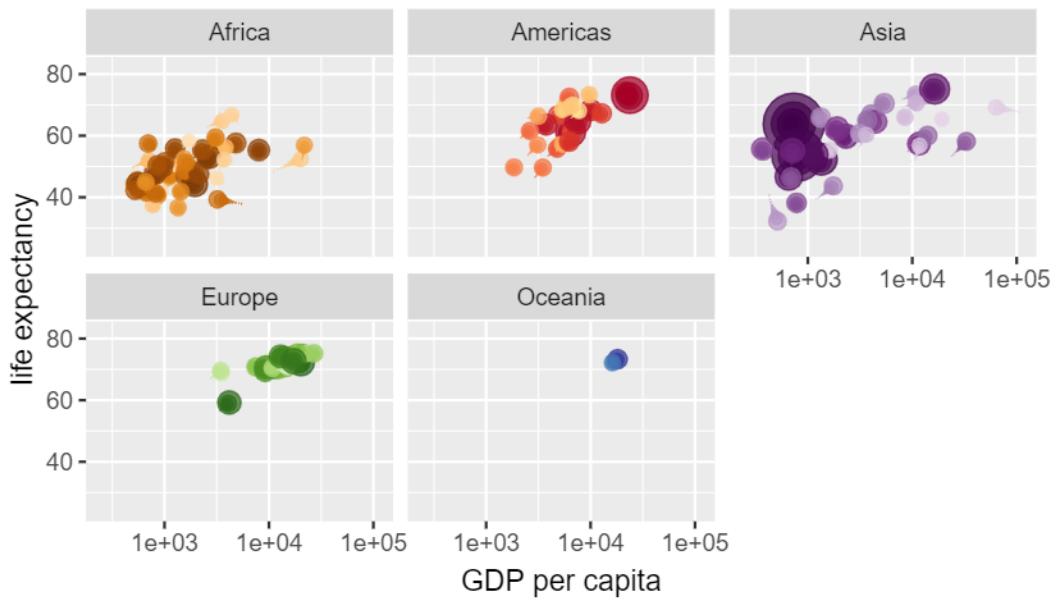
Year: 1975



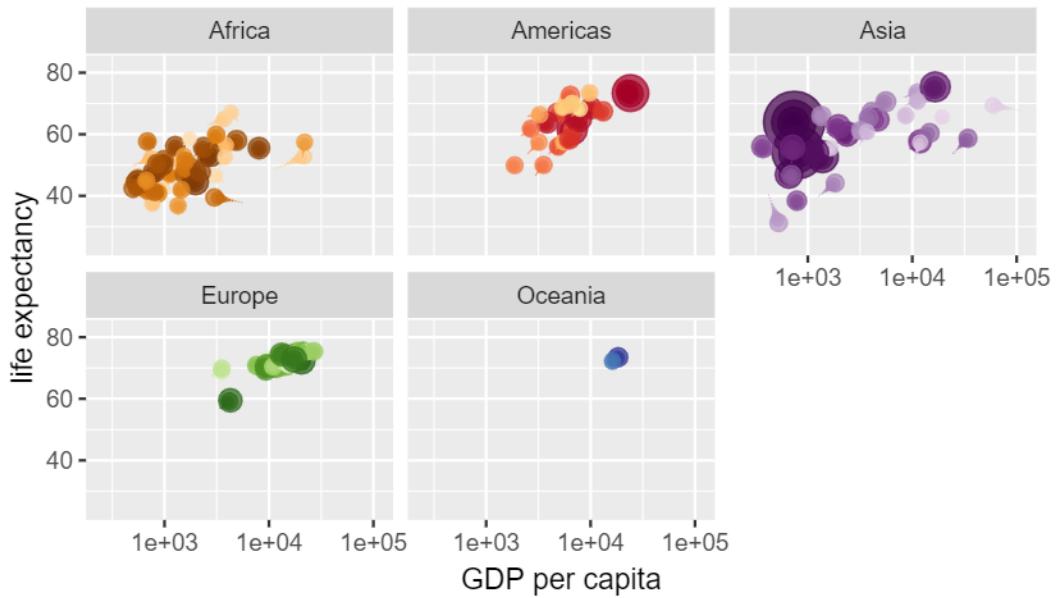
Year: 1976



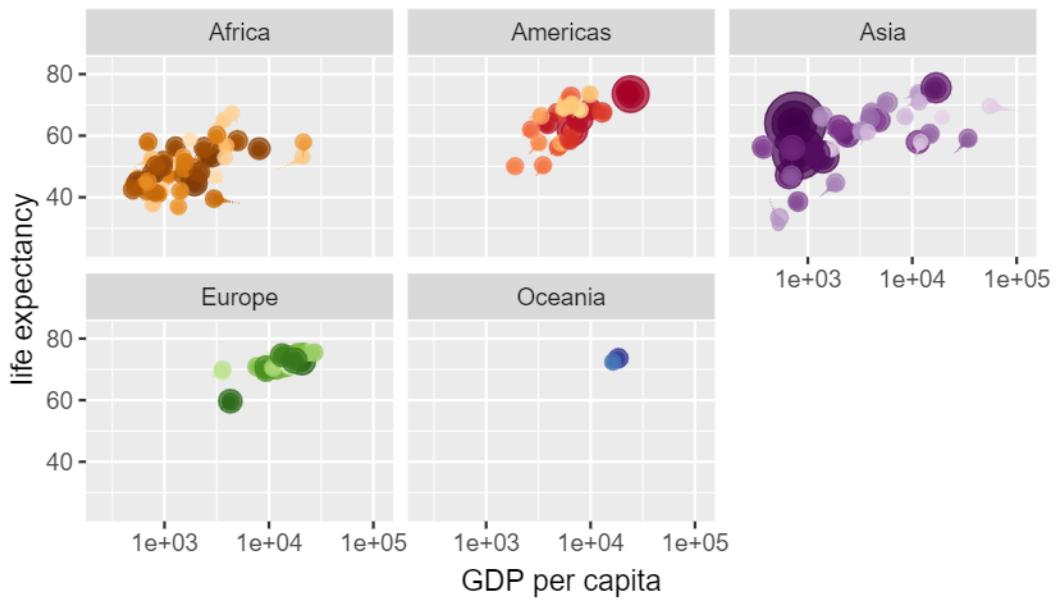
Year: 1976



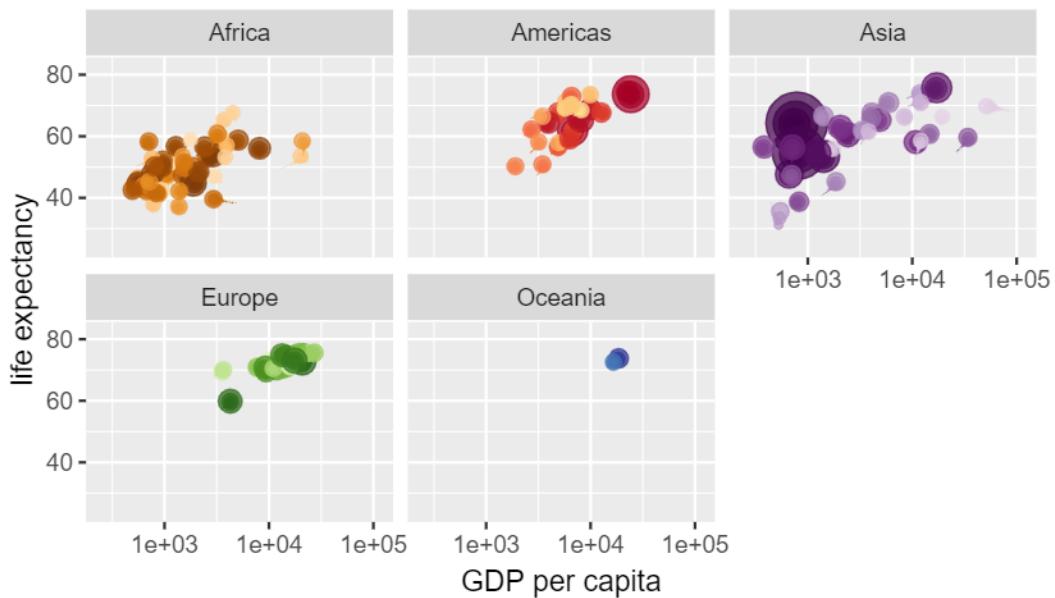
Year: 1977



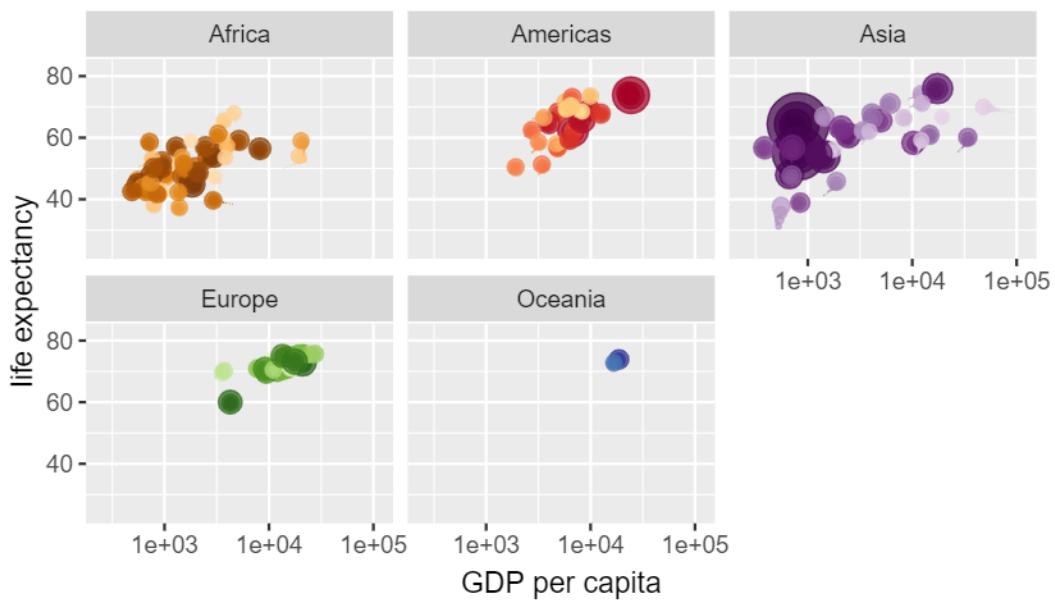
Year: 1978



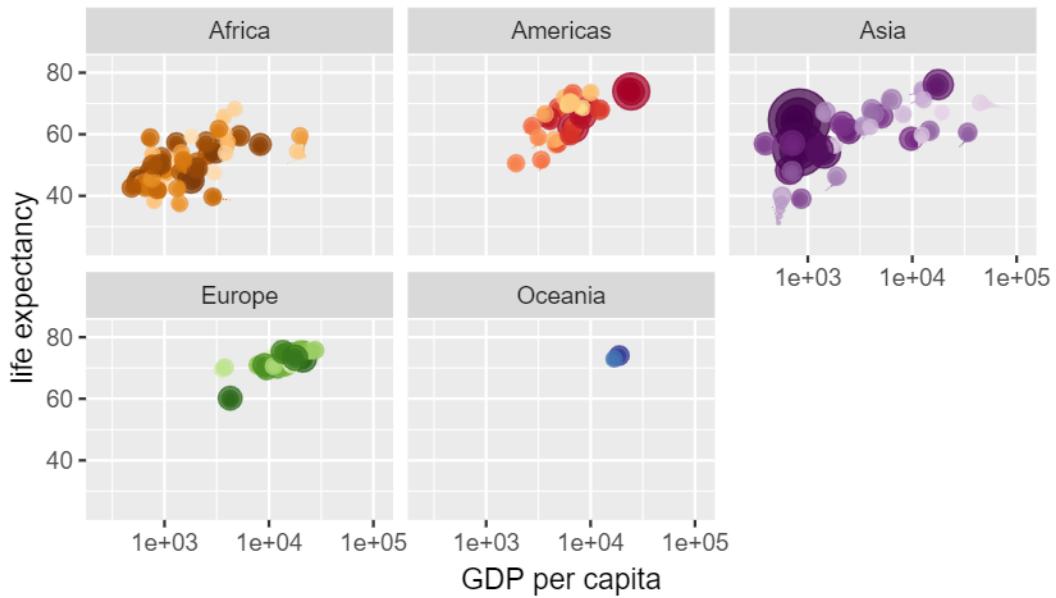
Year: 1978



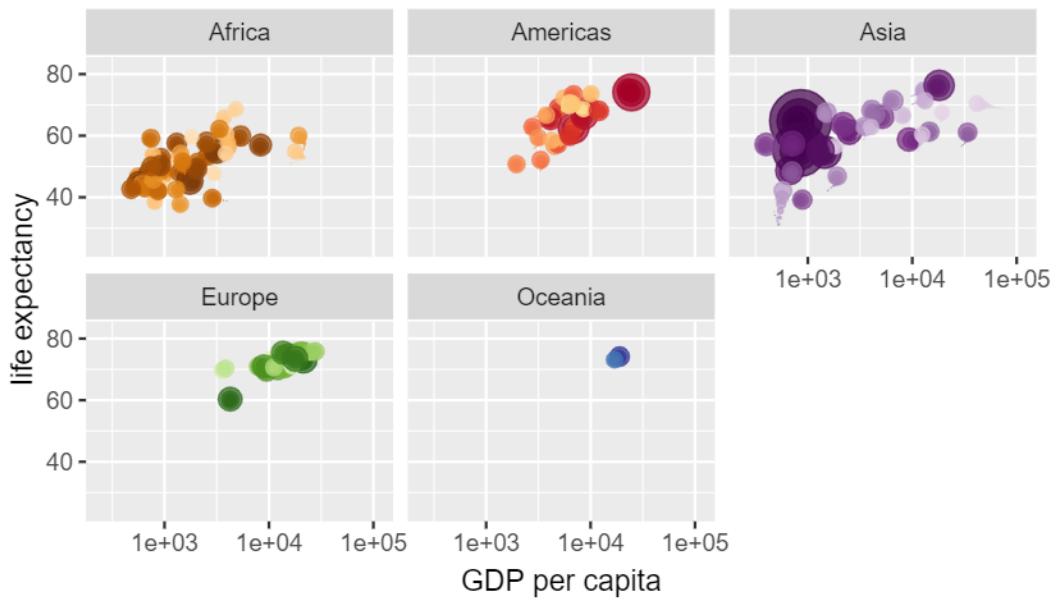
Year: 1979



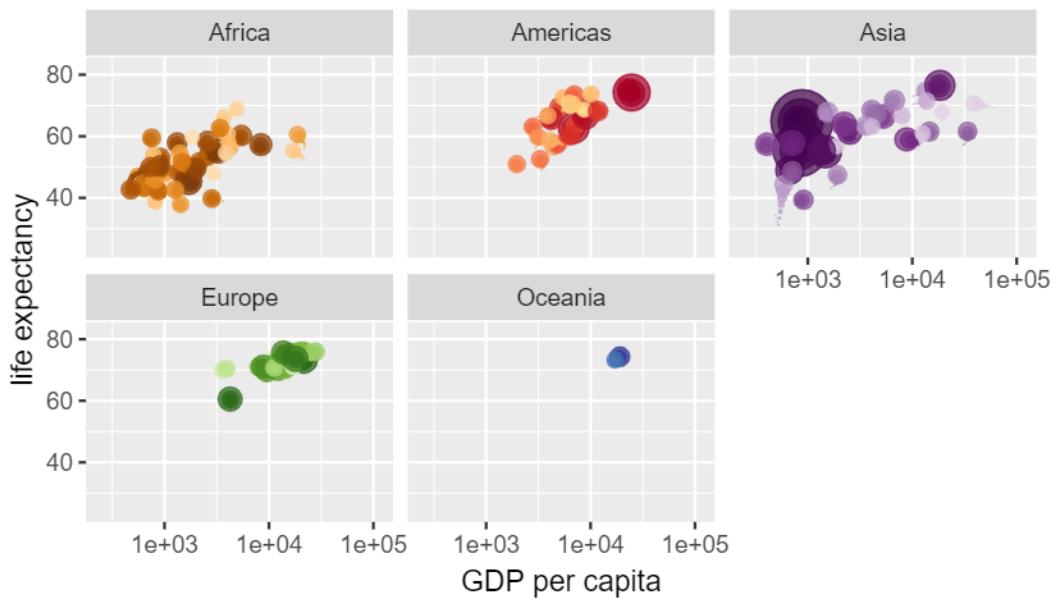
Year: 1979



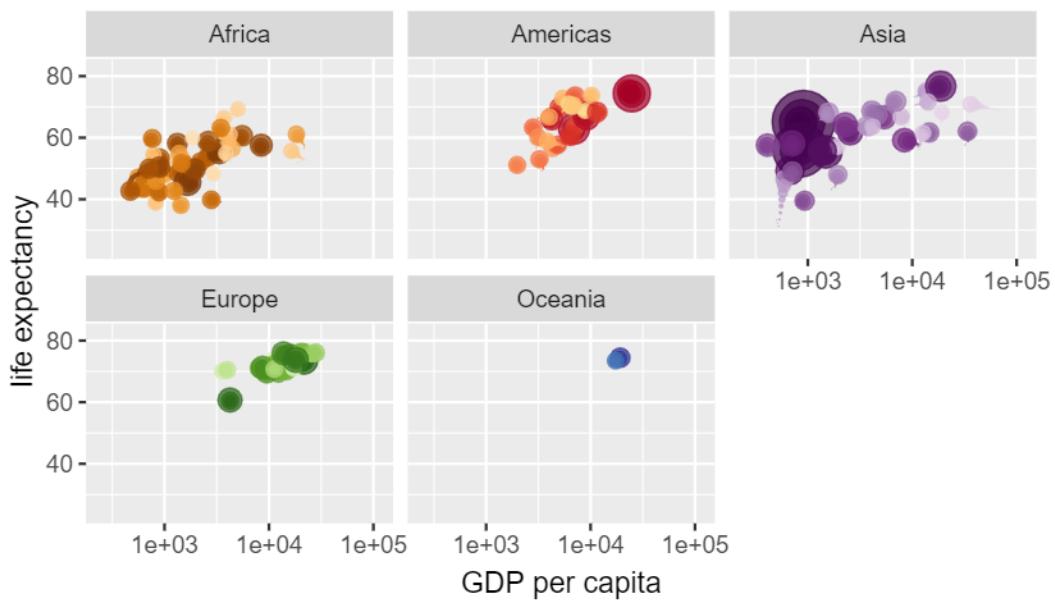
Year: 1980



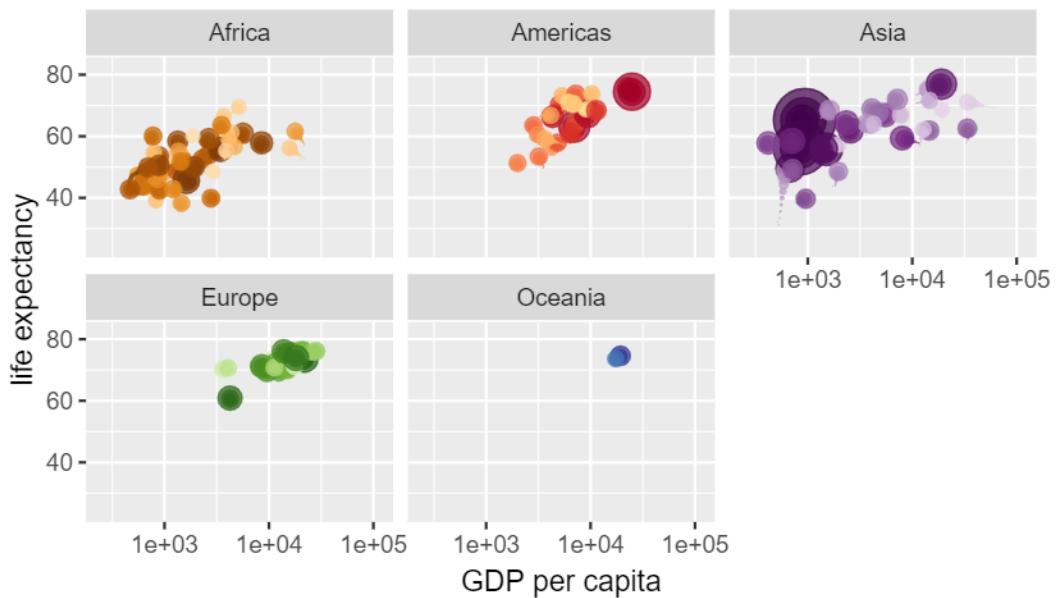
Year: 1980



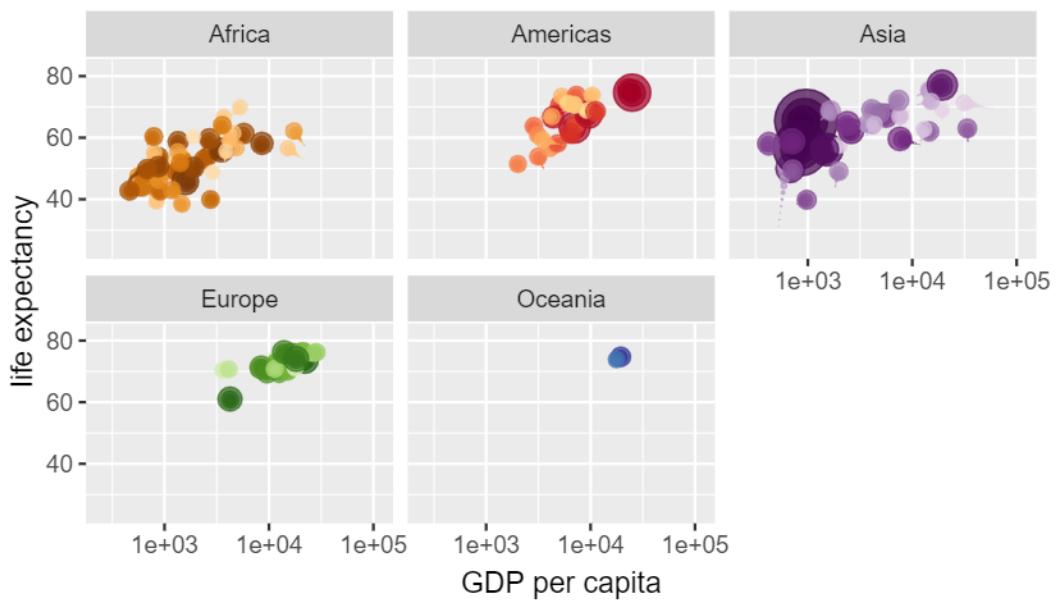
Year: 1981



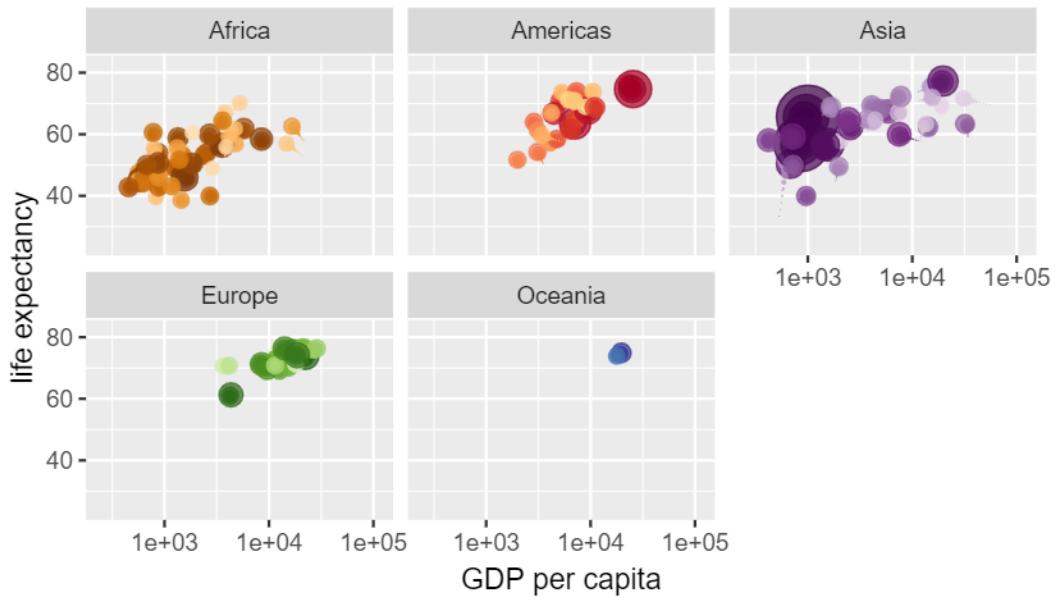
Year: 1981



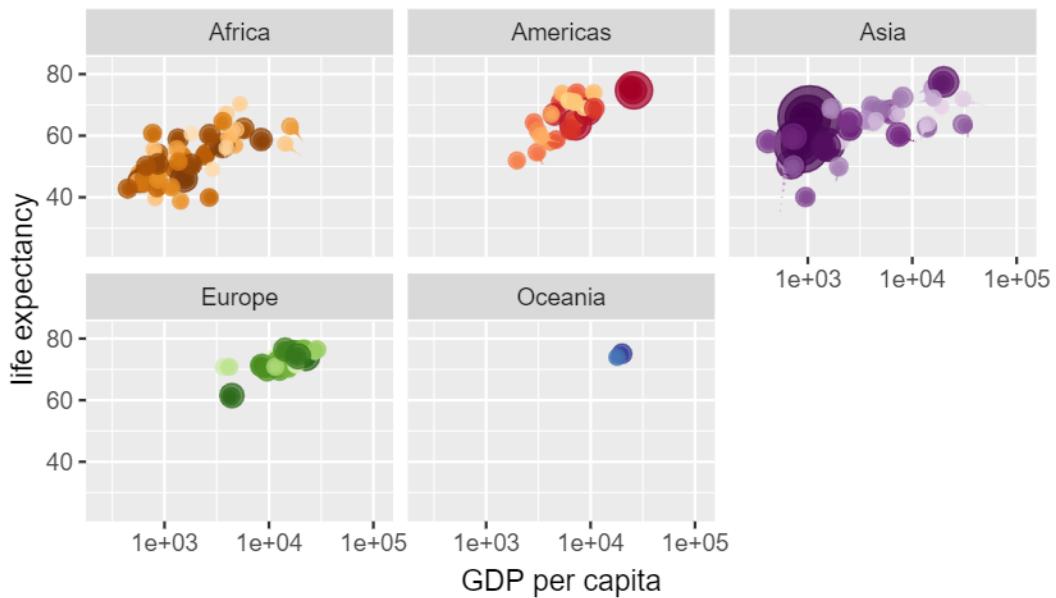
Year: 1982



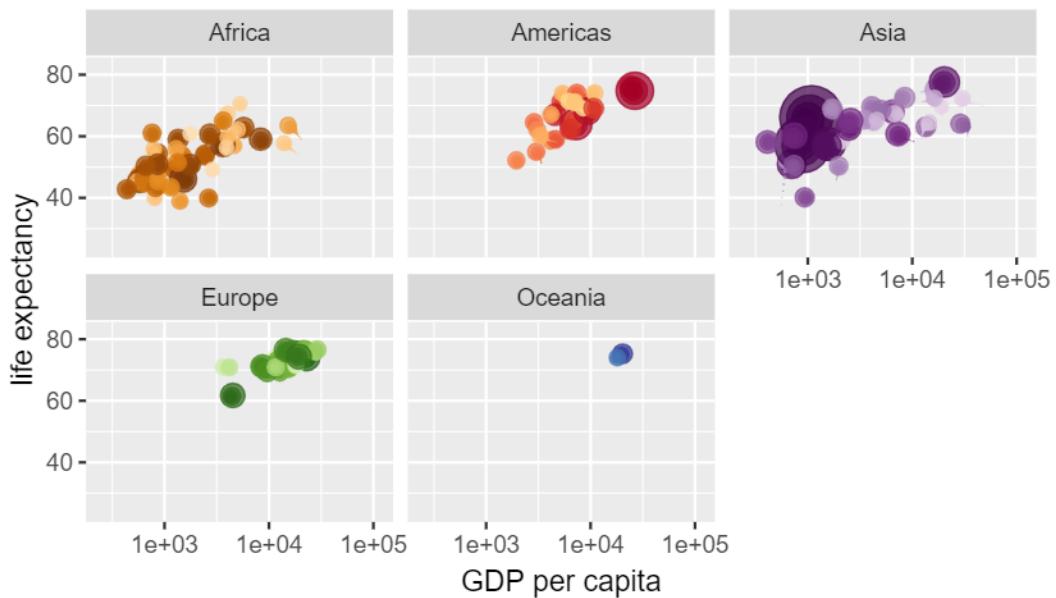
Year: 1983



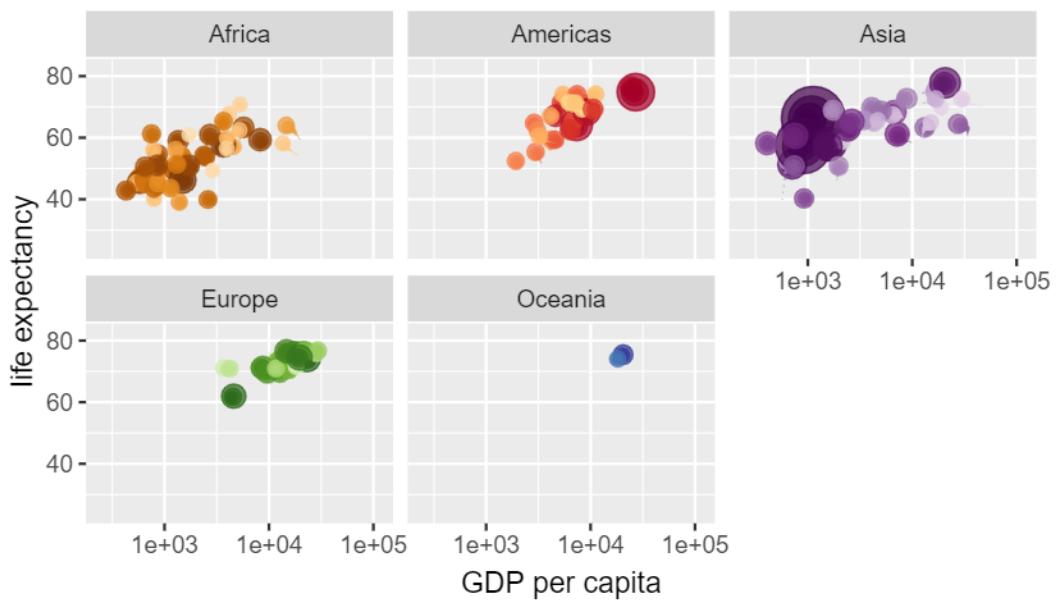
Year: 1983



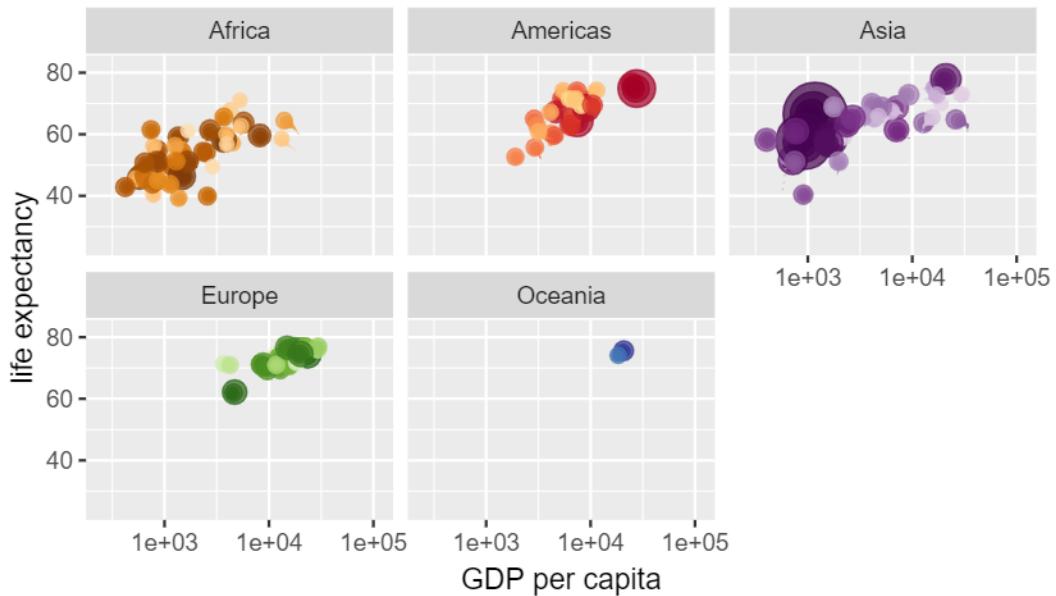
Year: 1984



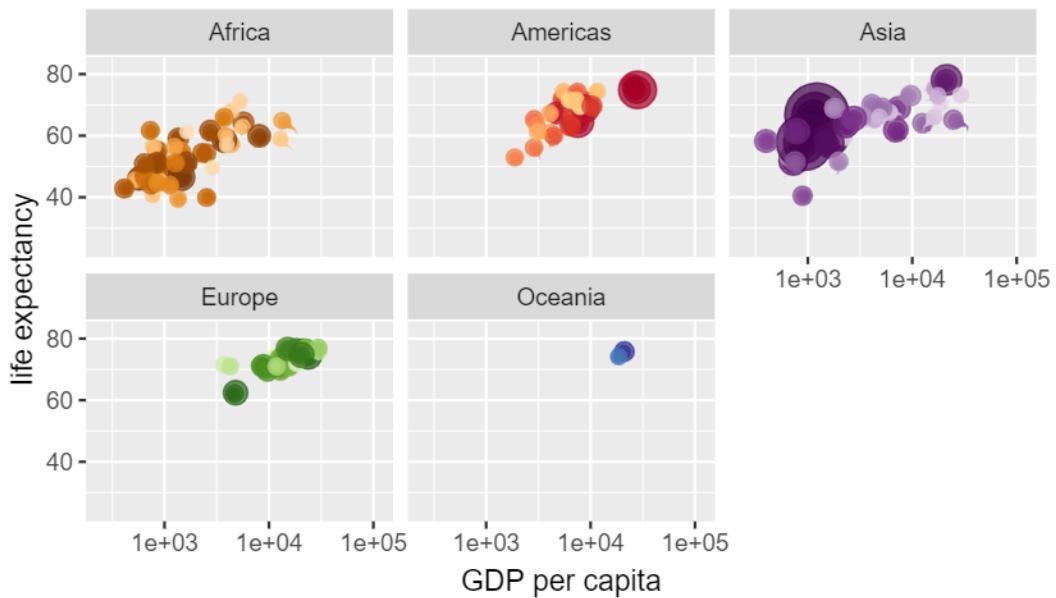
Year: 1984



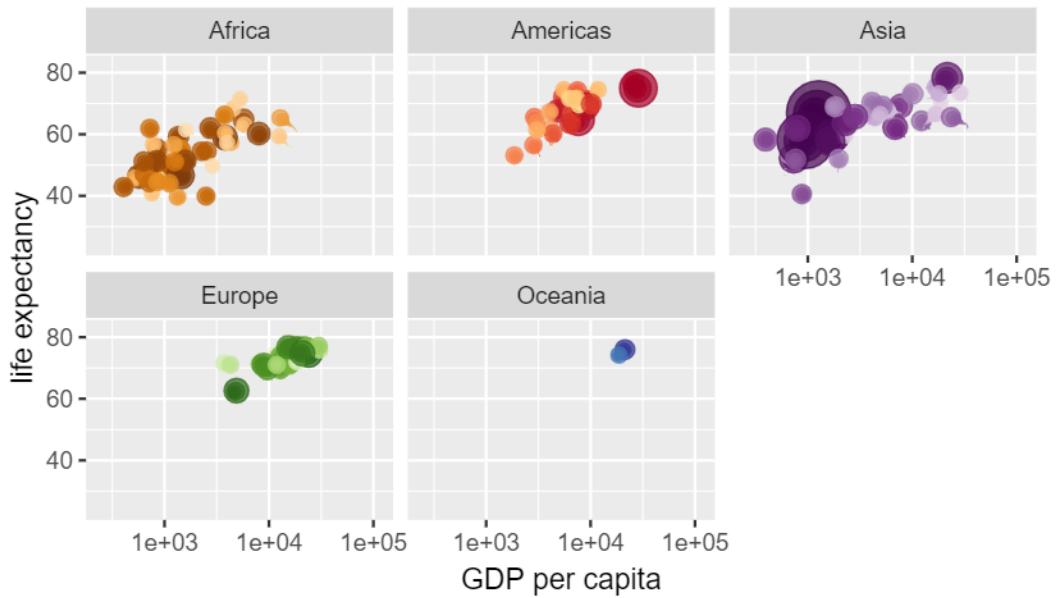
Year: 1985



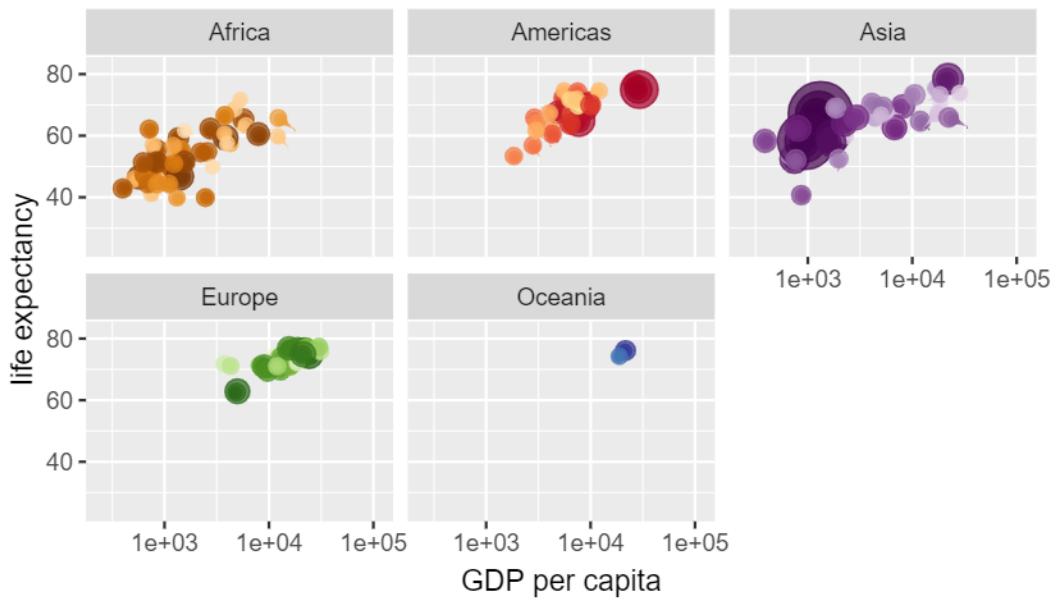
Year: 1985



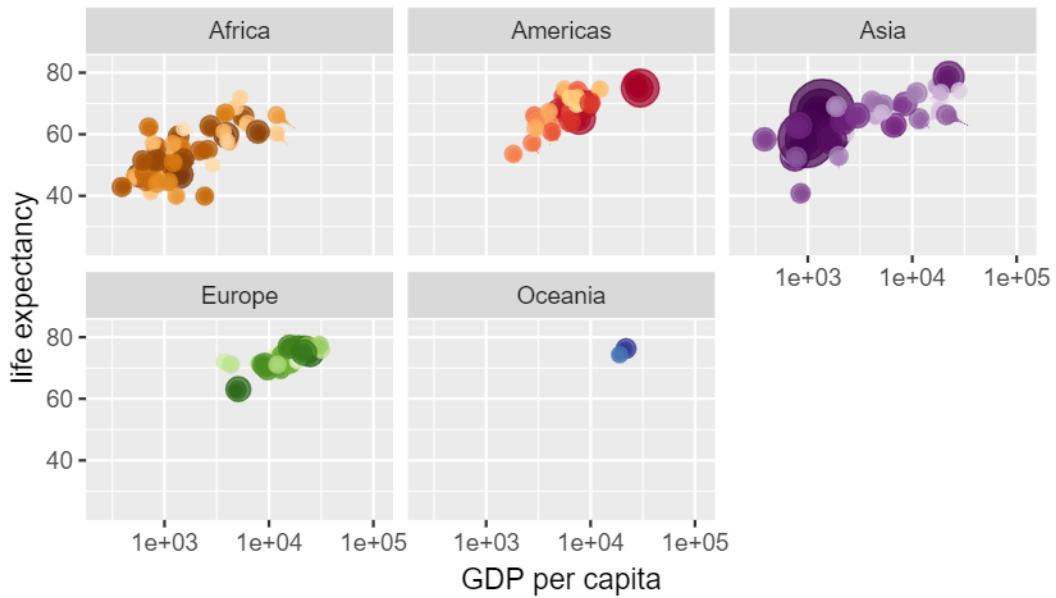
Year: 1986



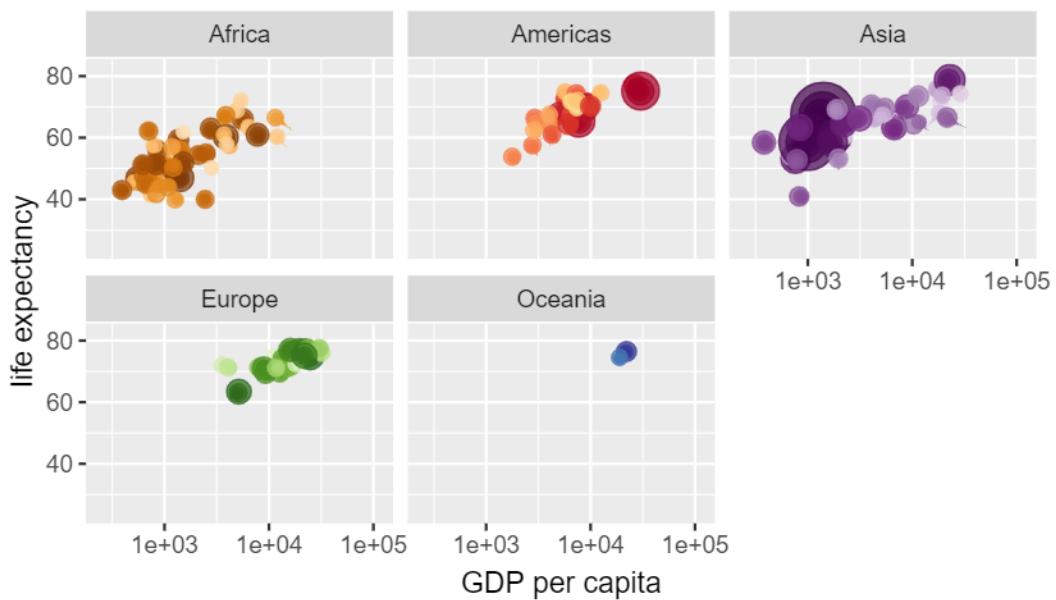
Year: 1986



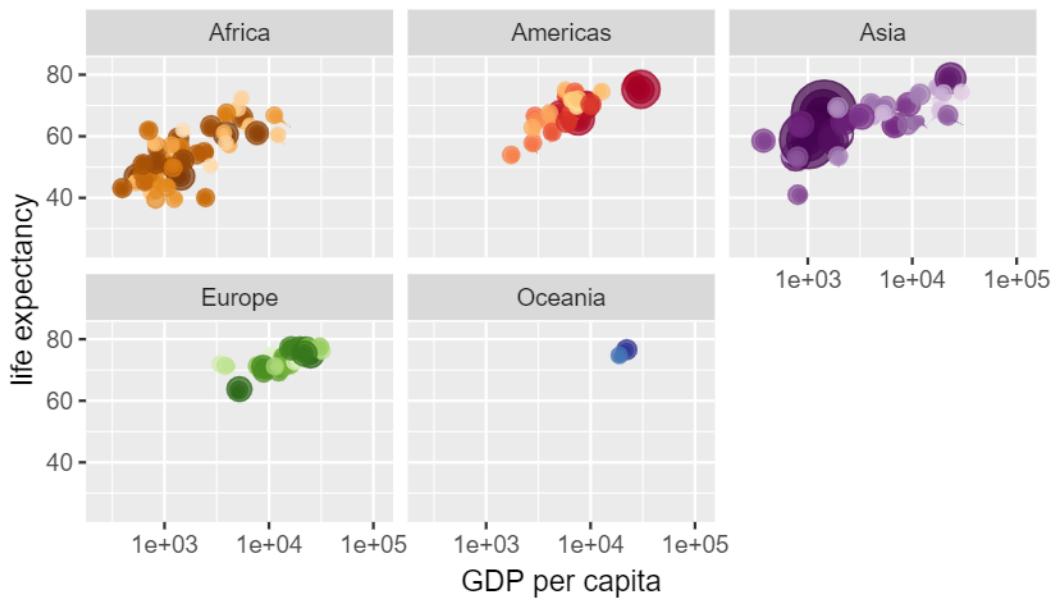
Year: 1987



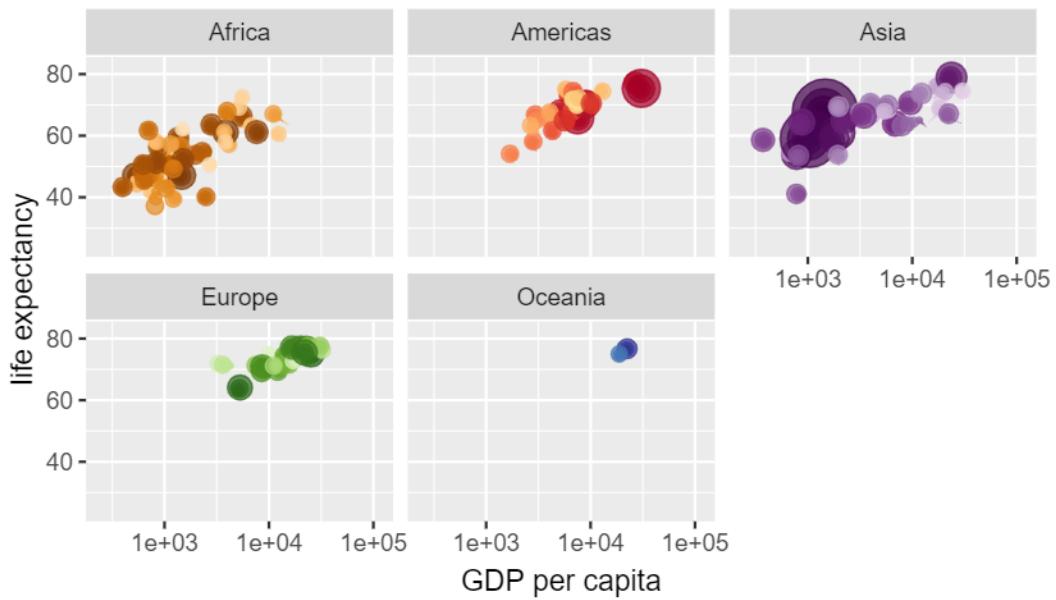
Year: 1988



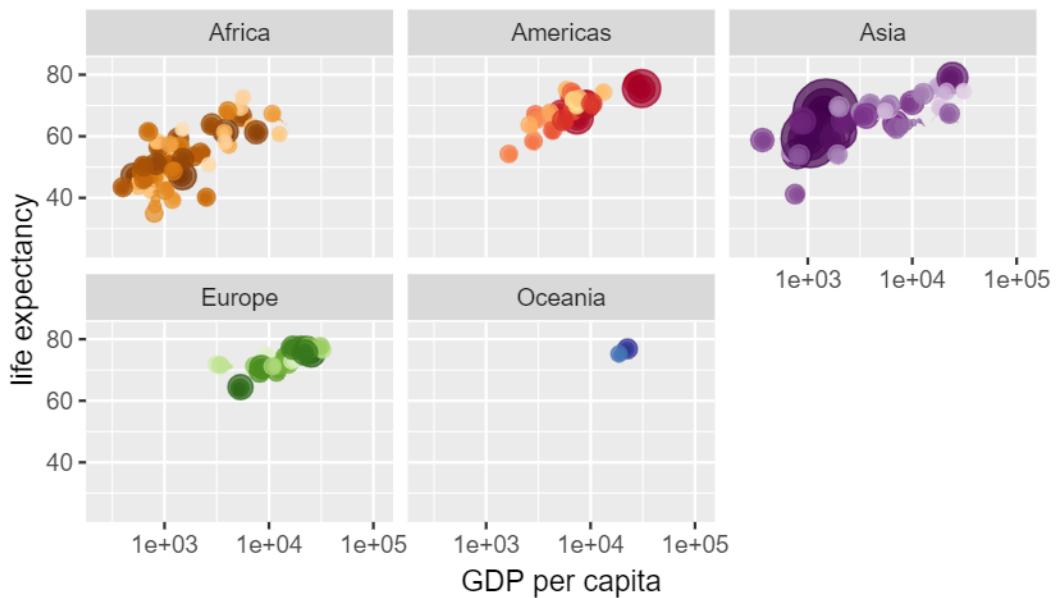
Year: 1988



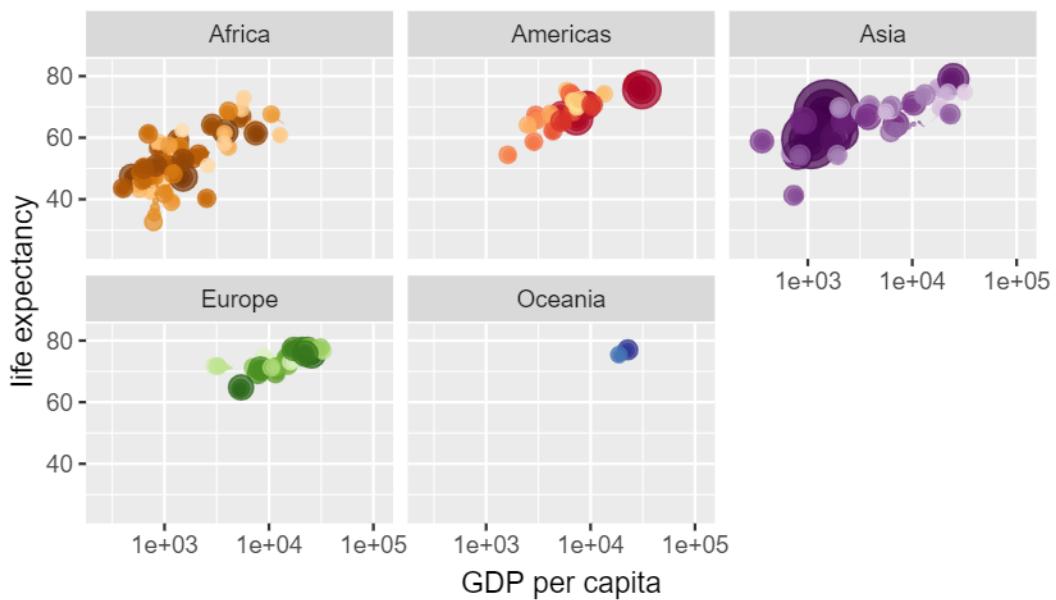
Year: 1989



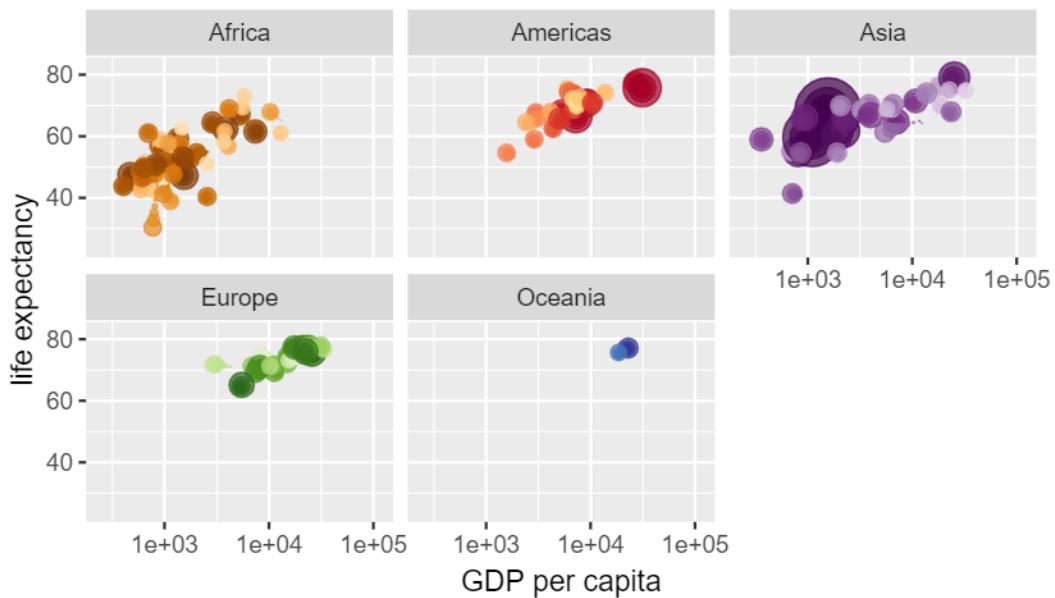
Year: 1989



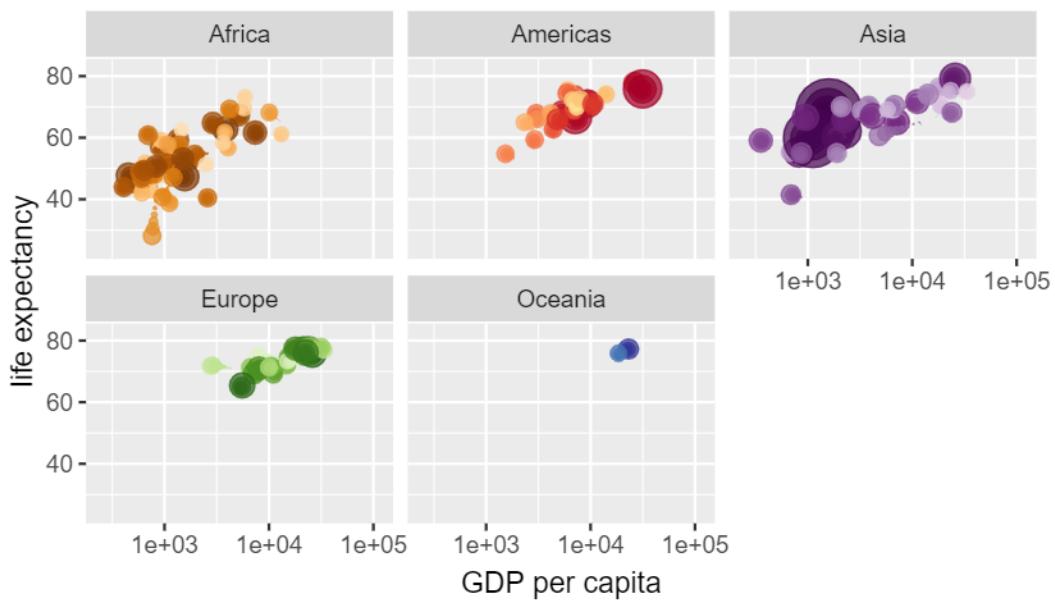
Year: 1990



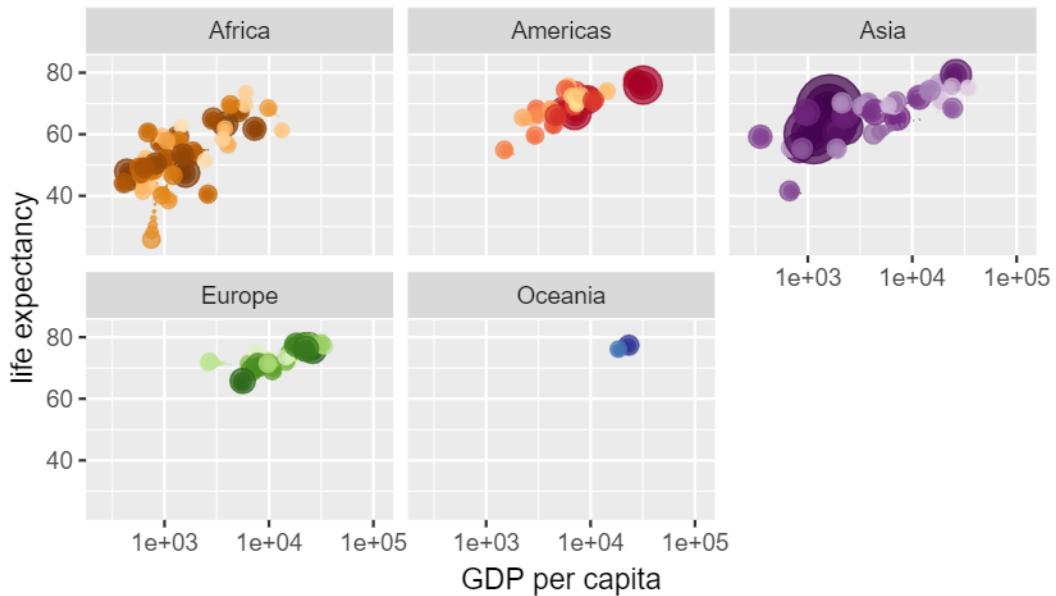
Year: 1990



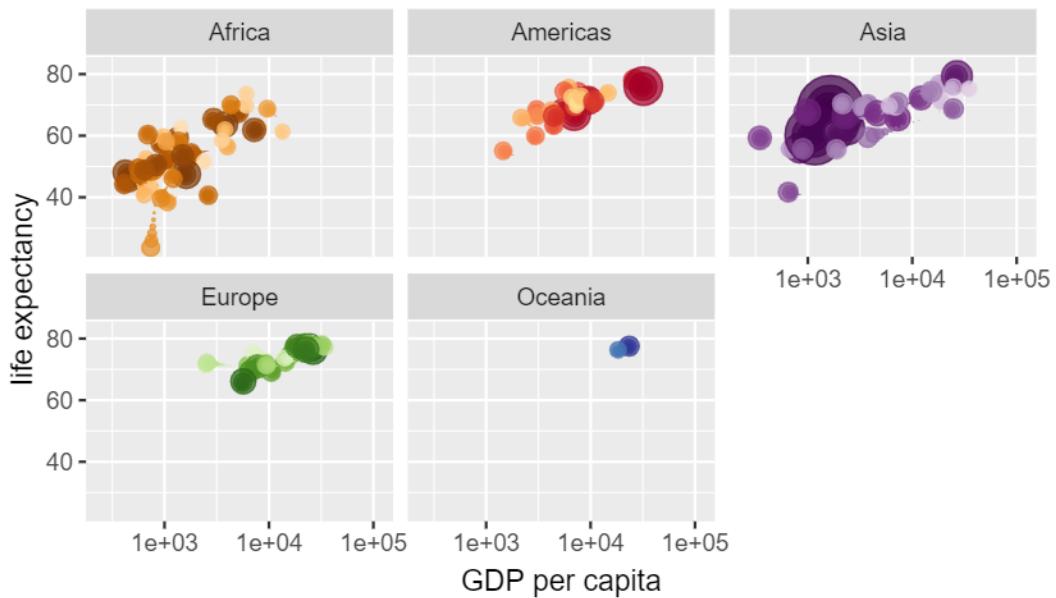
Year: 1991



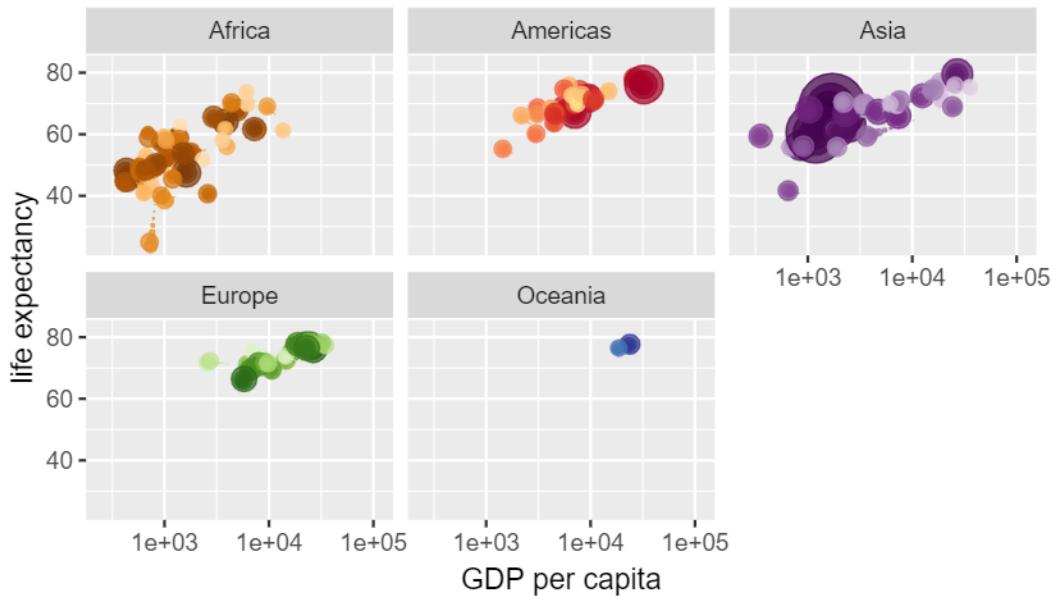
Year: 1991



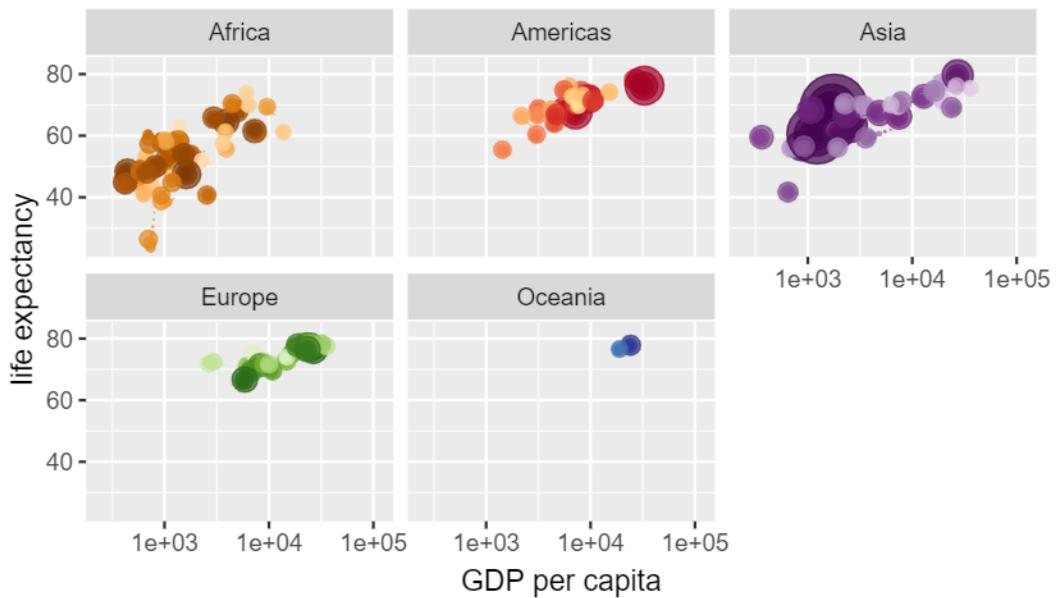
Year: 1992



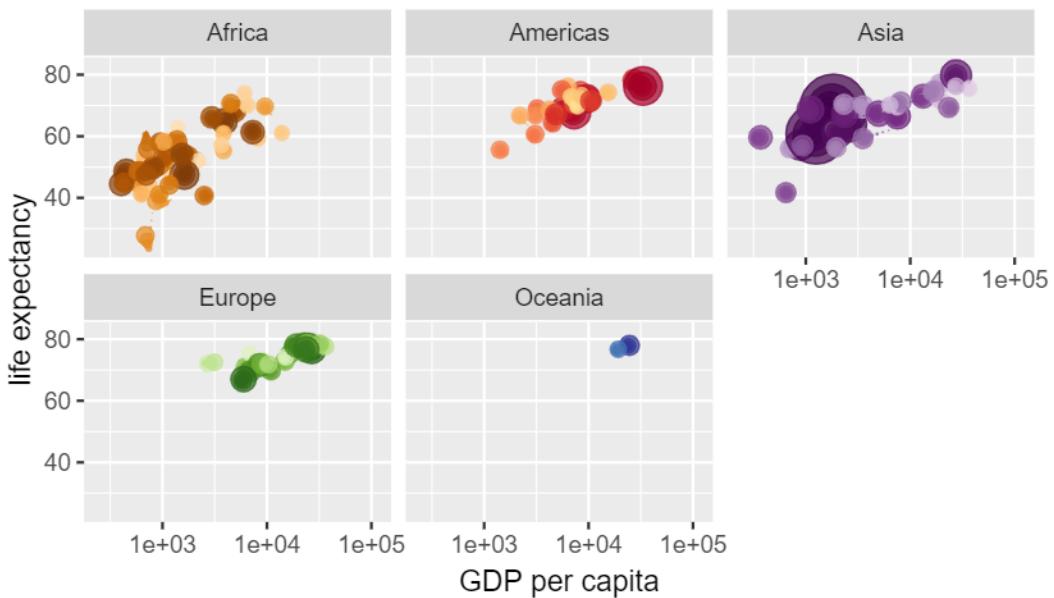
Year: 1993



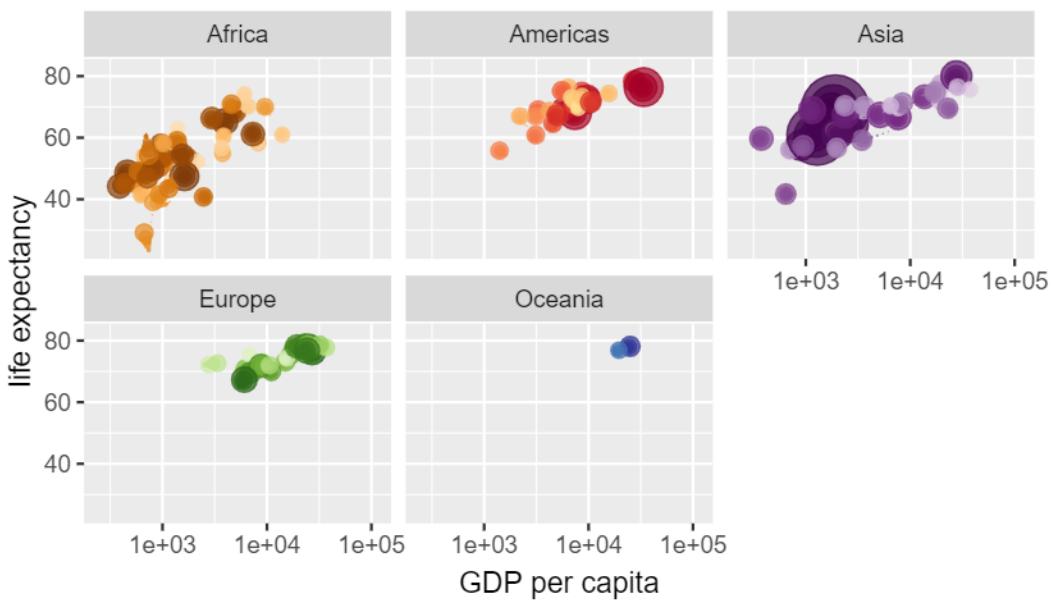
Year: 1993



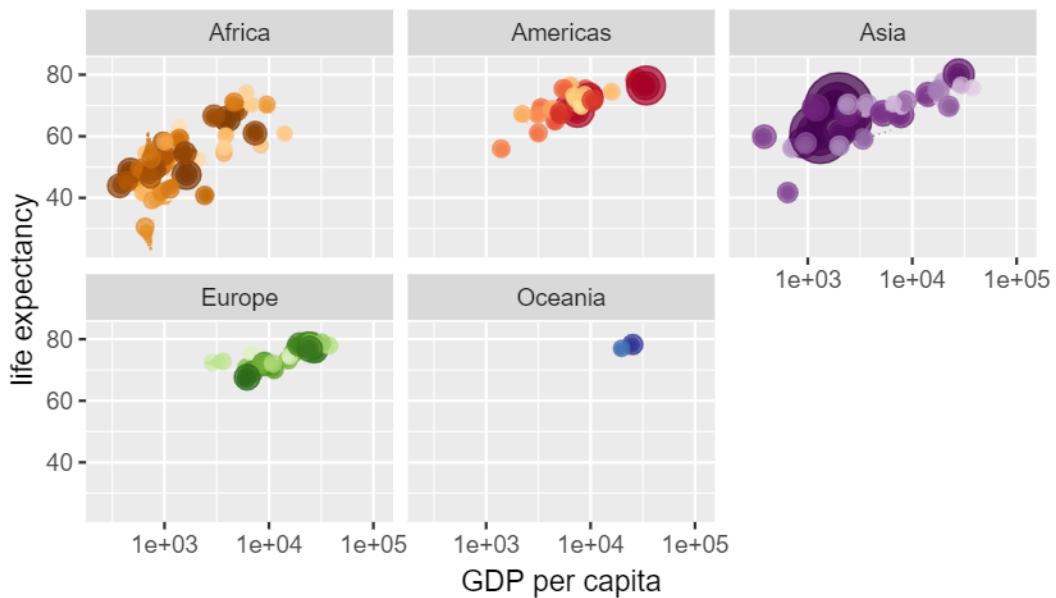
Year: 1994



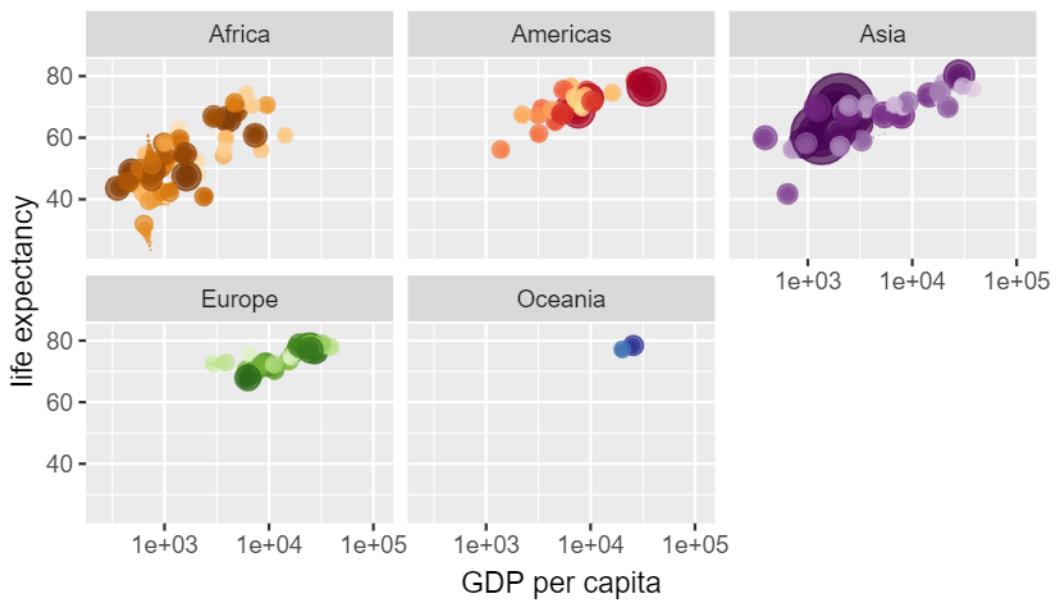
Year: 1994



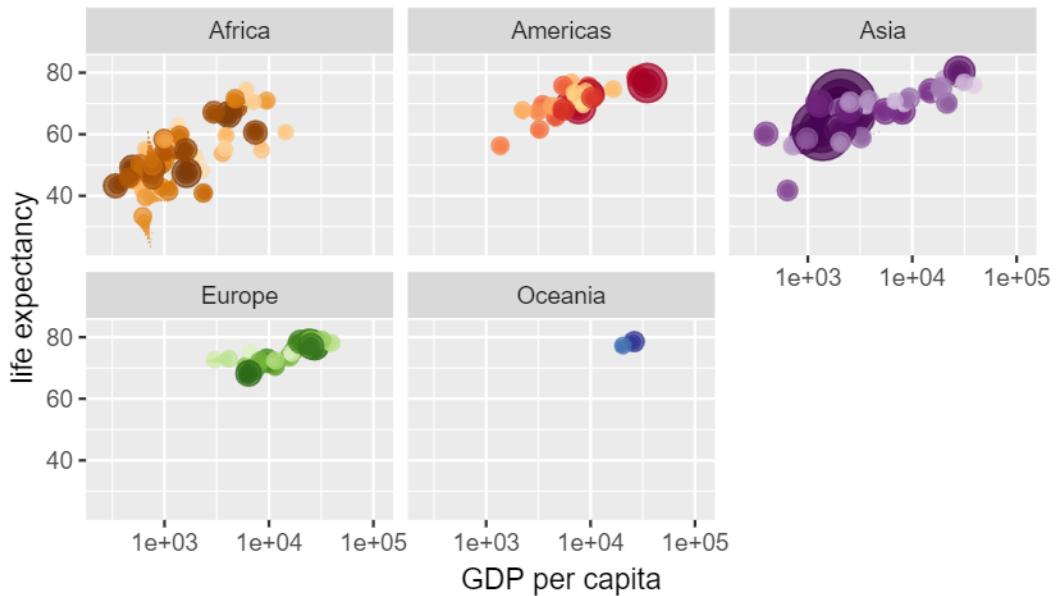
Year: 1995



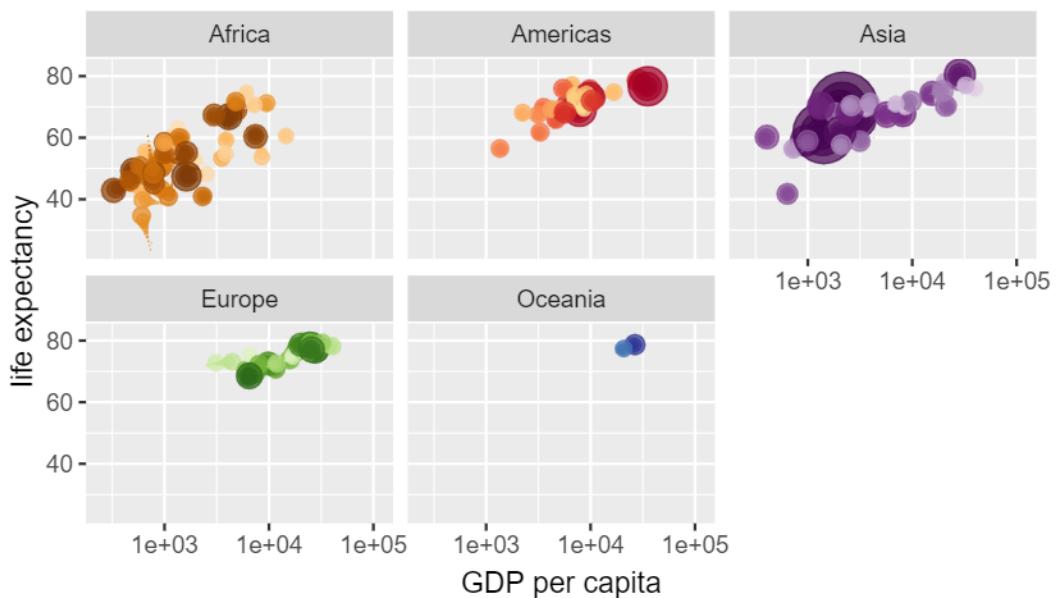
Year: 1995



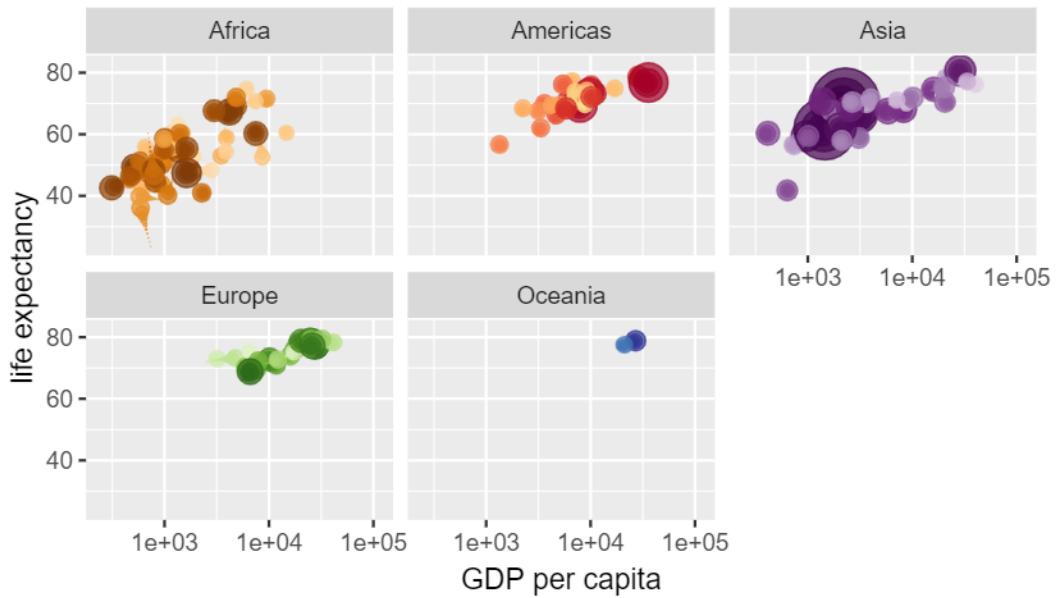
Year: 1996



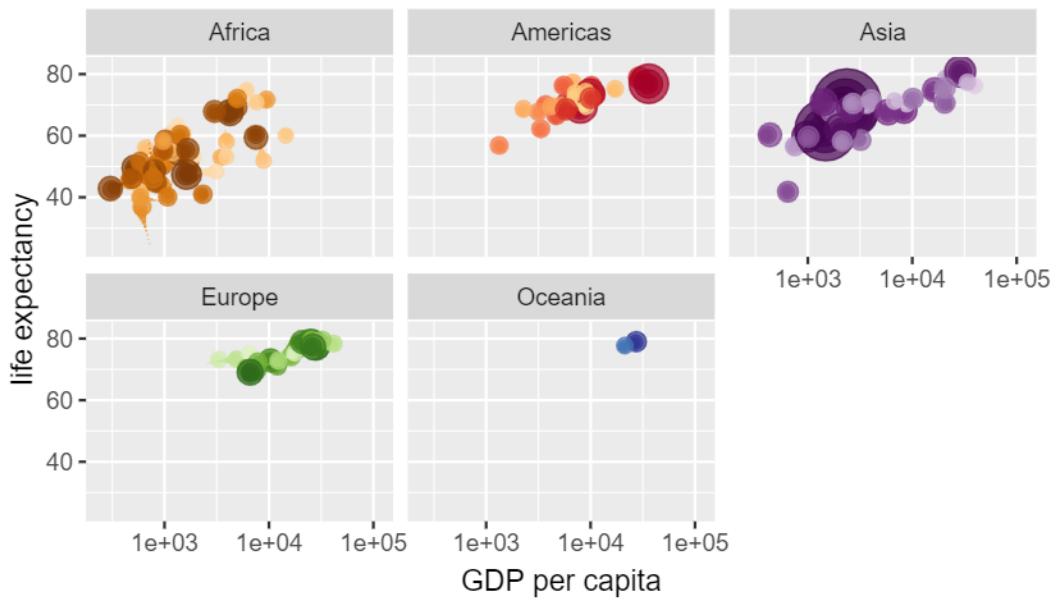
Year: 1996



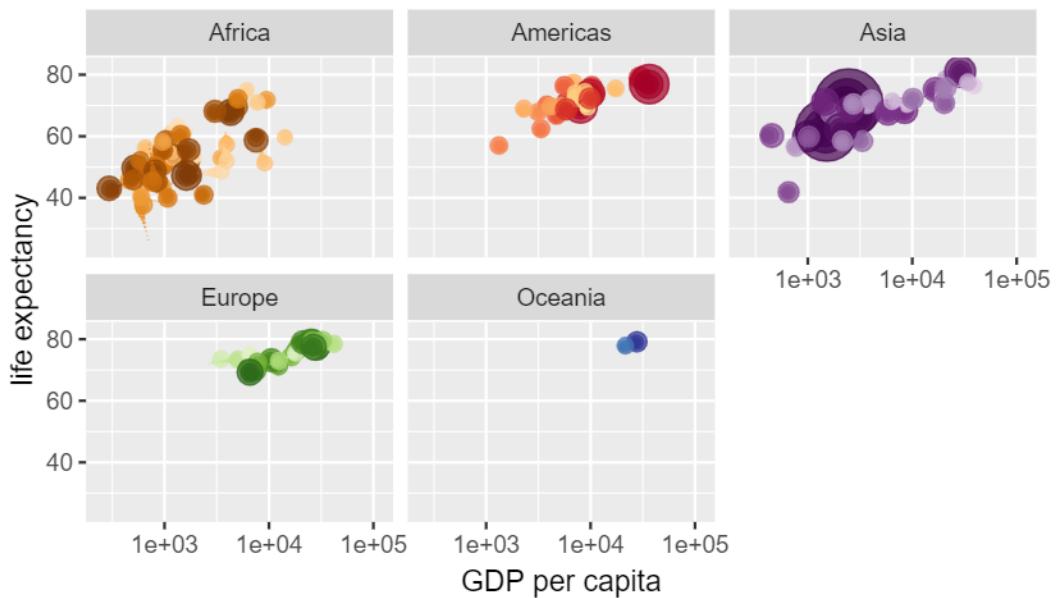
Year: 1997



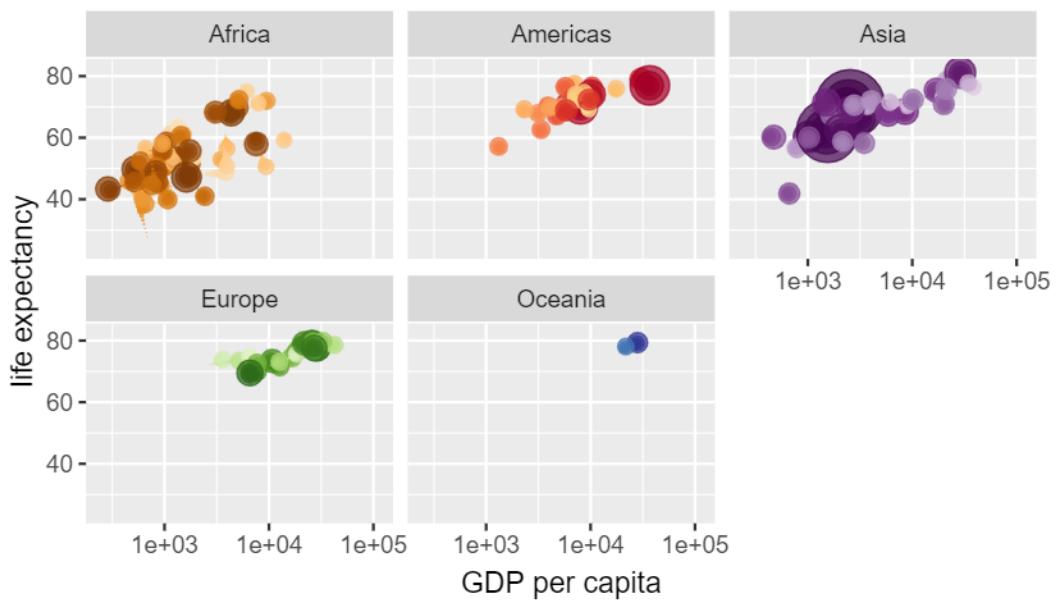
Year: 1998



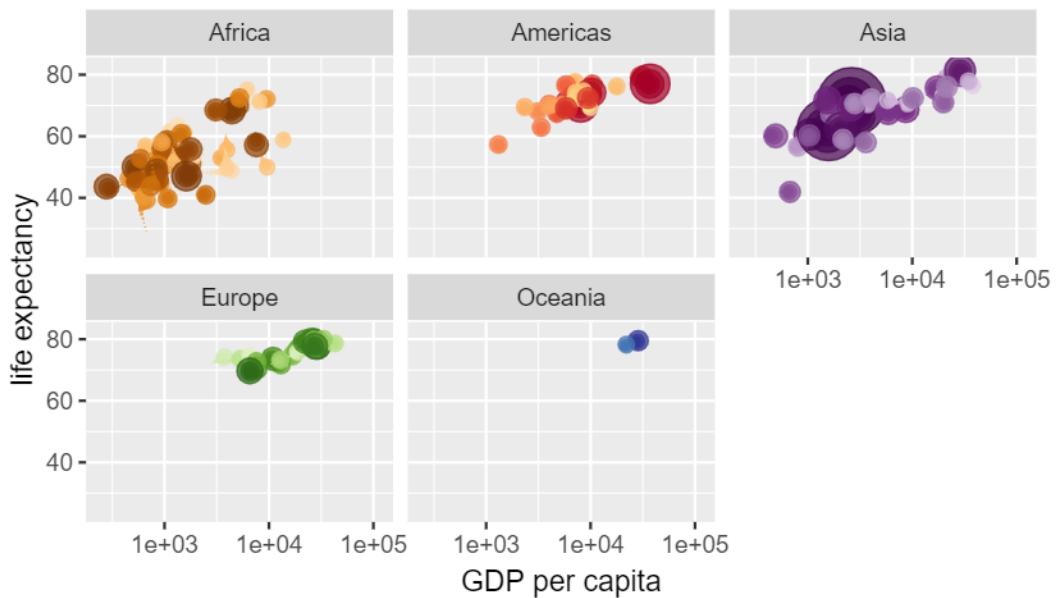
Year: 1998



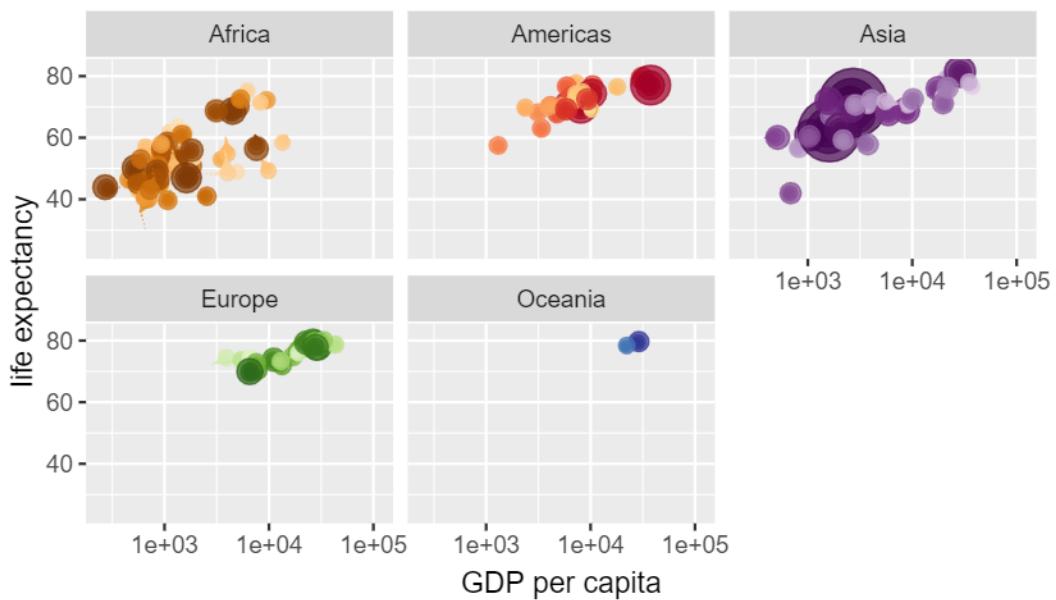
Year: 1999



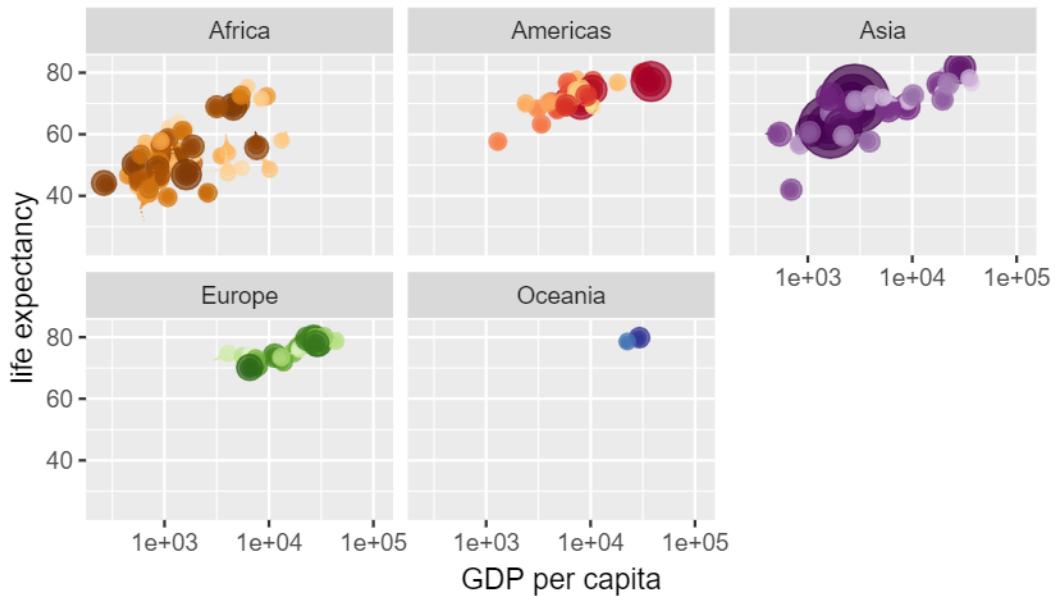
Year: 1999



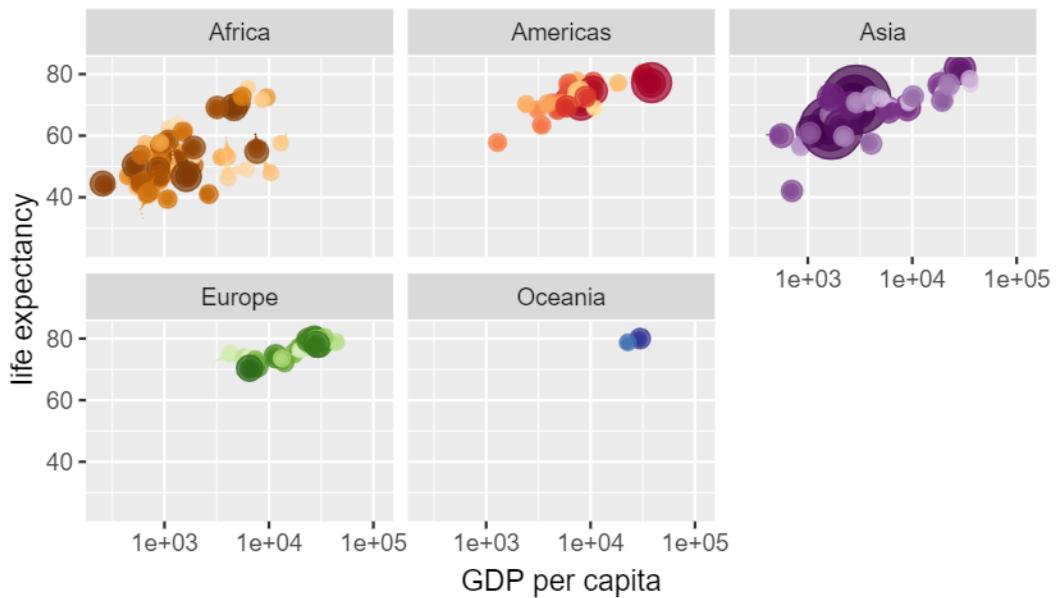
Year: 2000



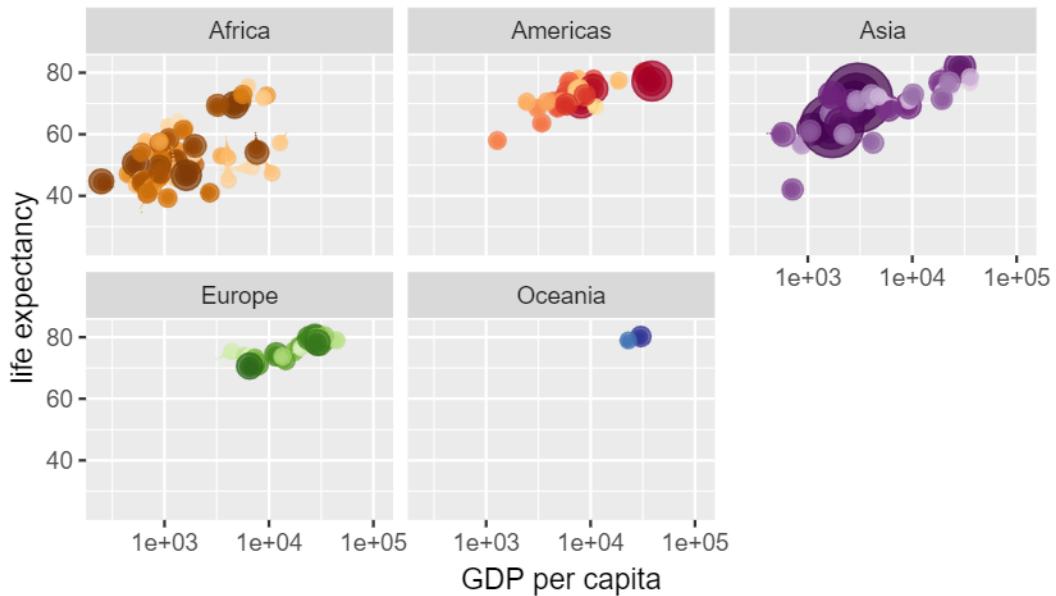
Year: 2000



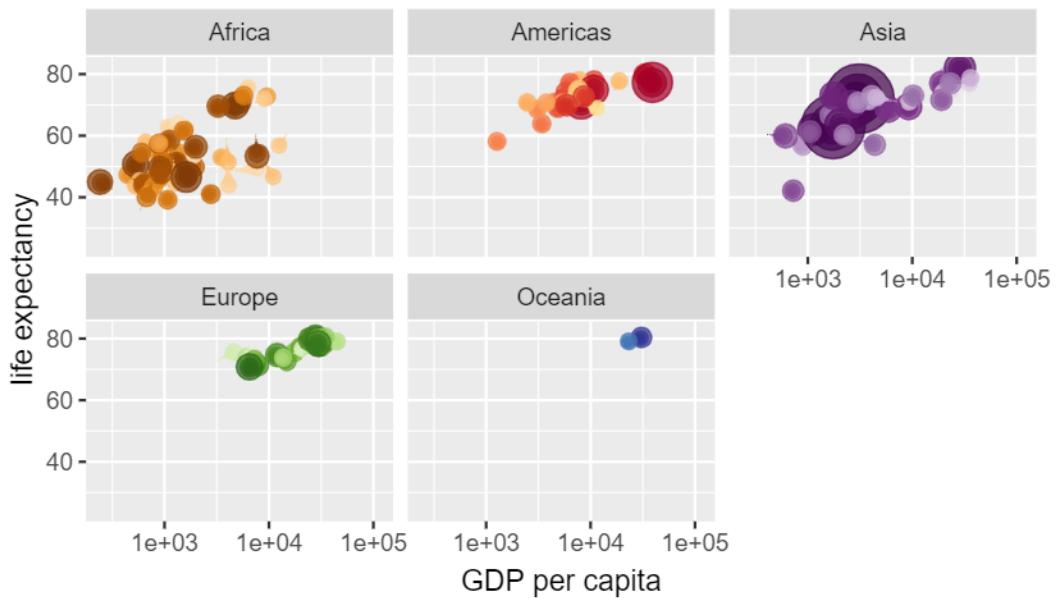
Year: 2001



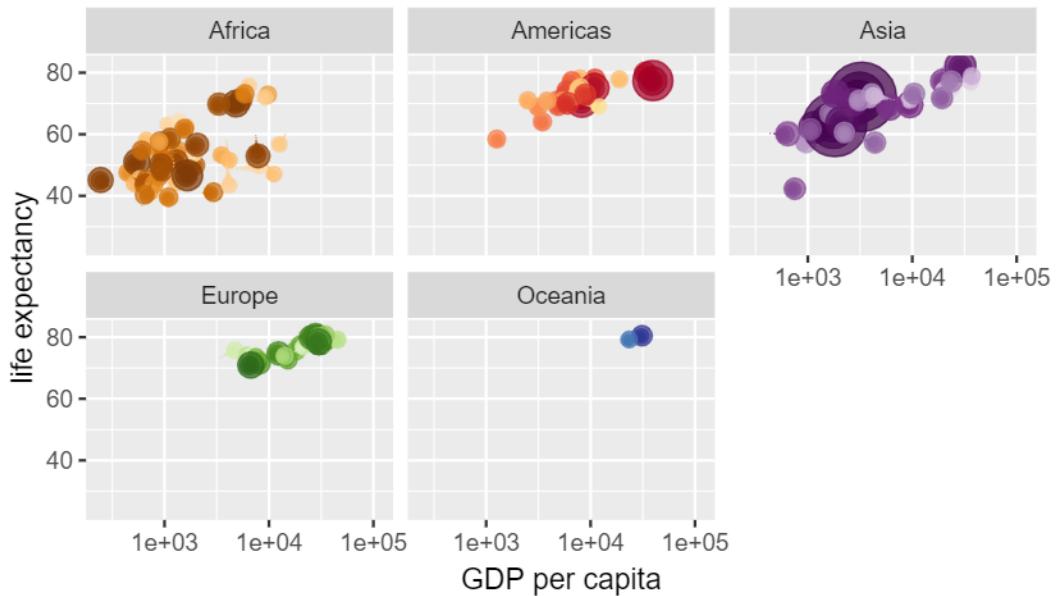
Year: 2001



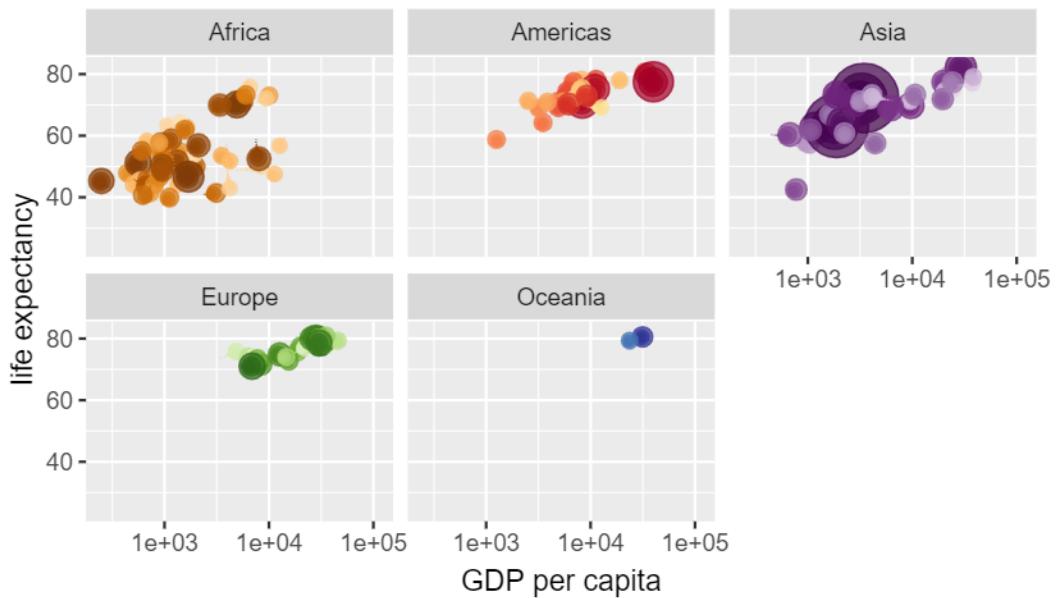
Year: 2002



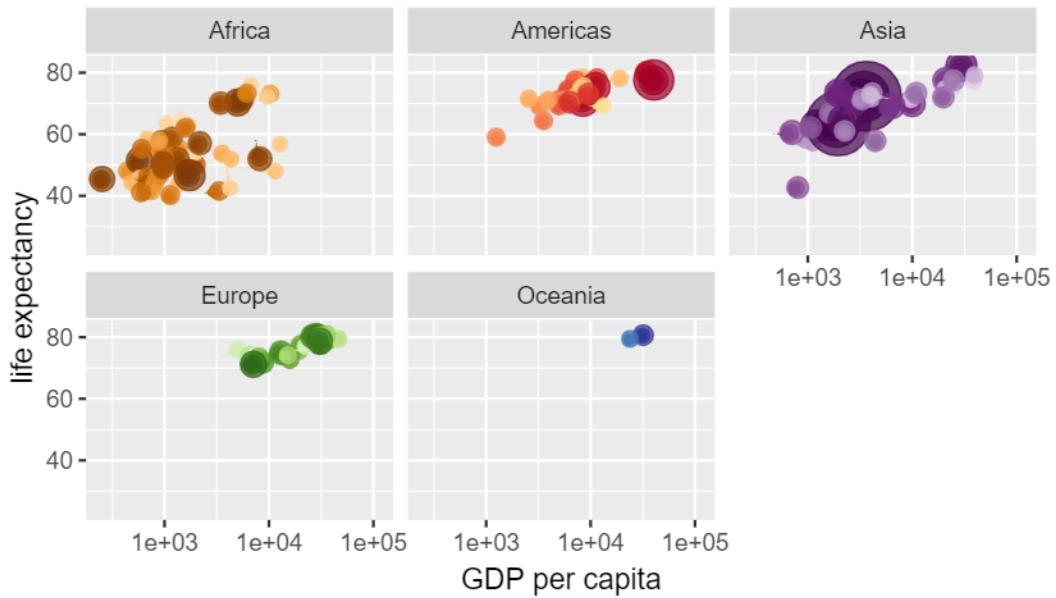
Year: 2003



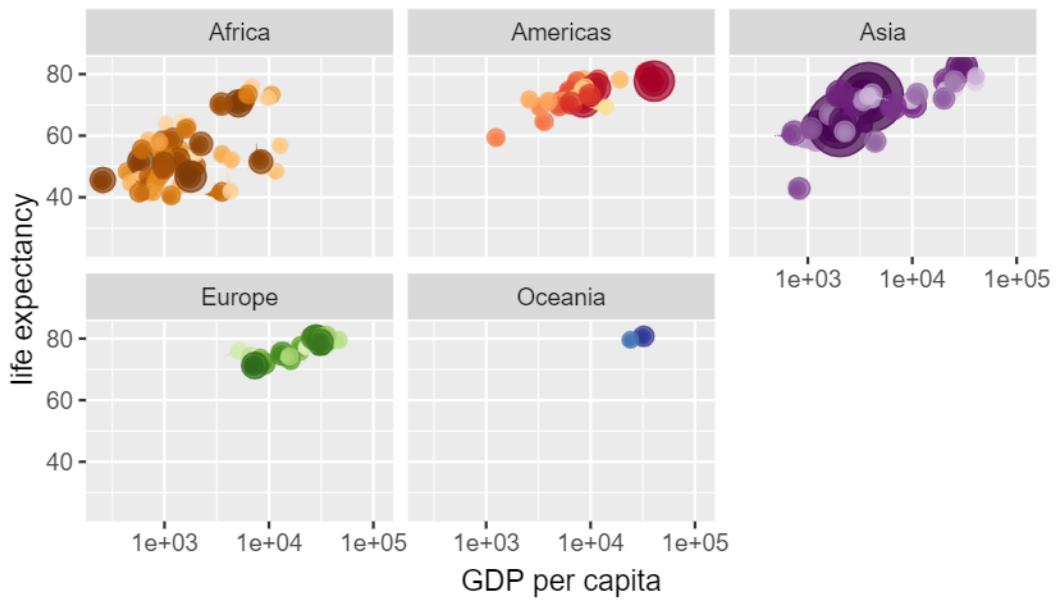
Year: 2003



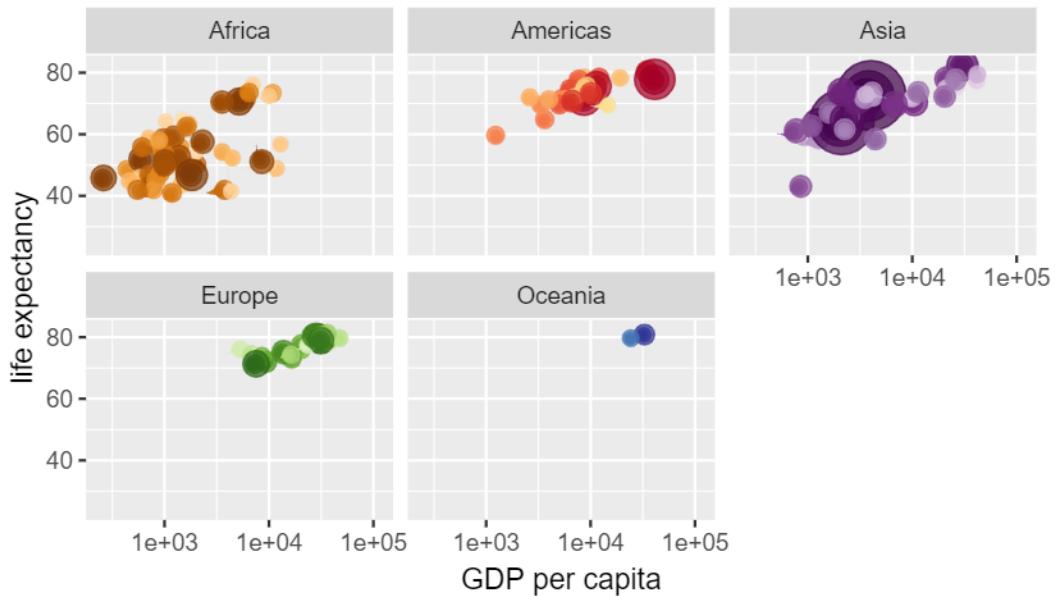
Year: 2004



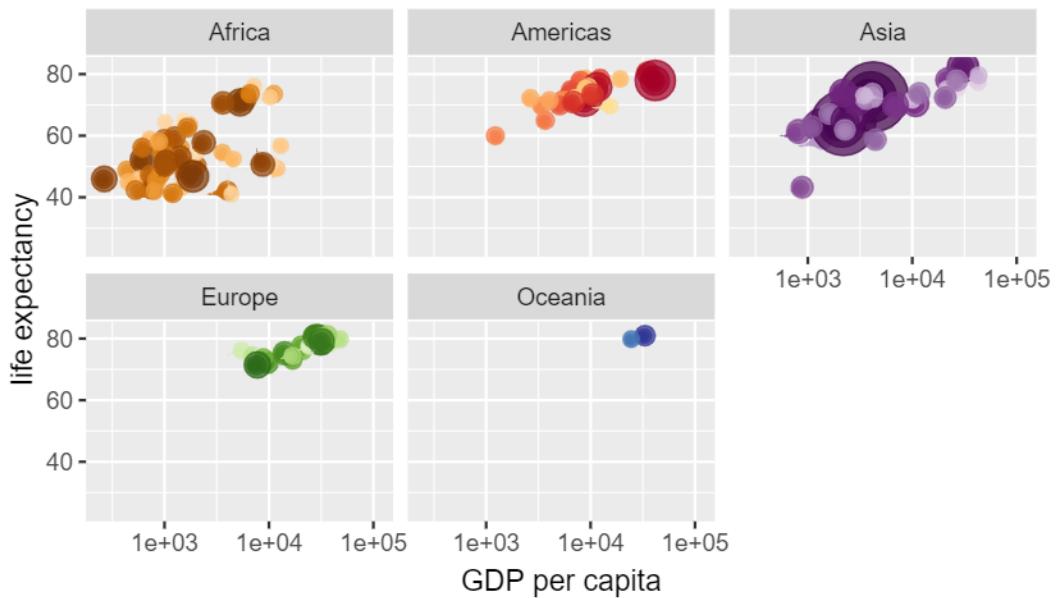
Year: 2004



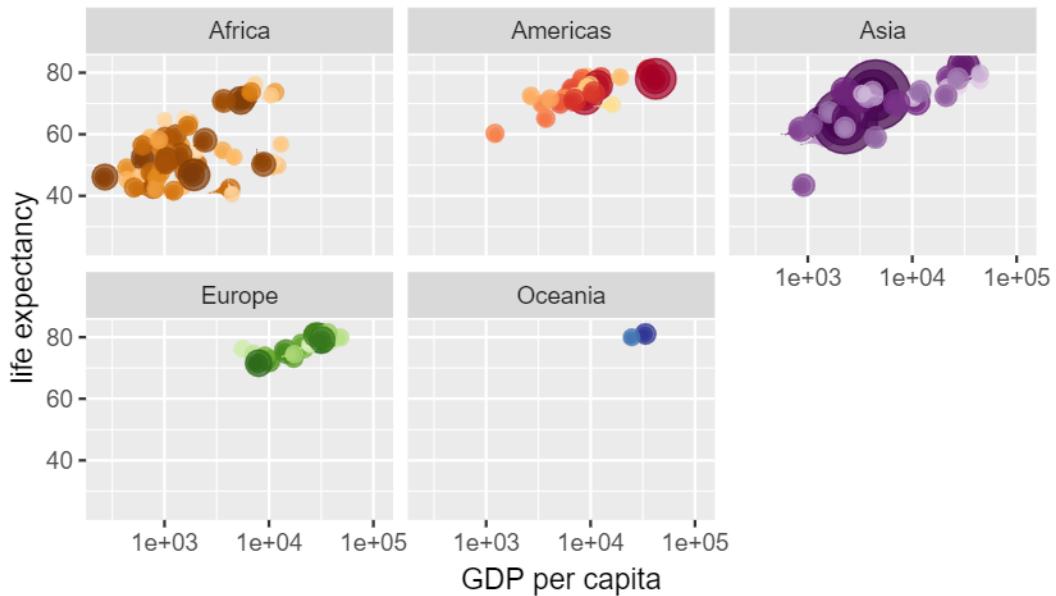
Year: 2005



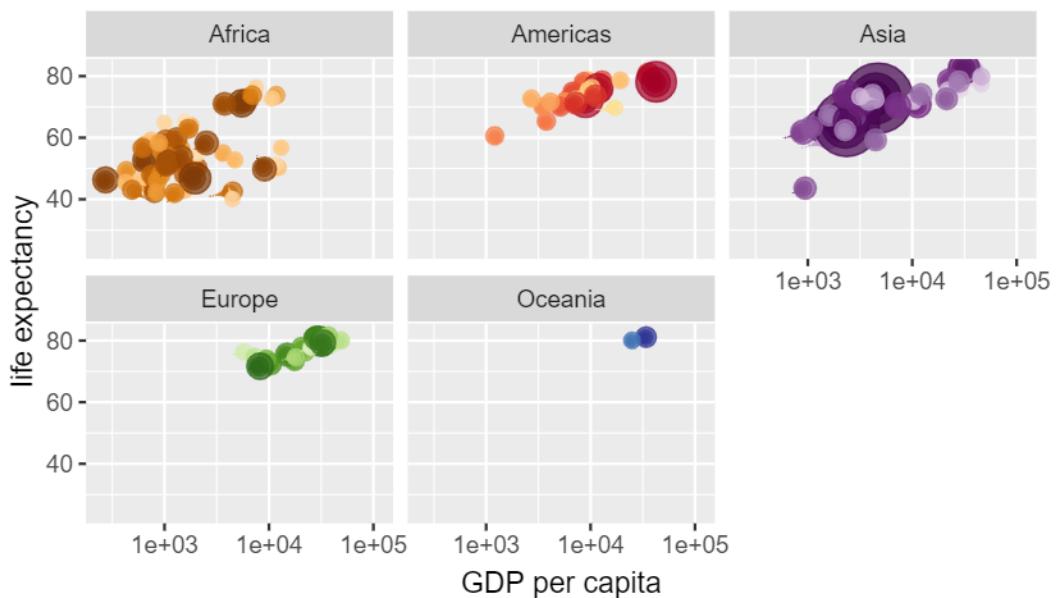
Year: 2005



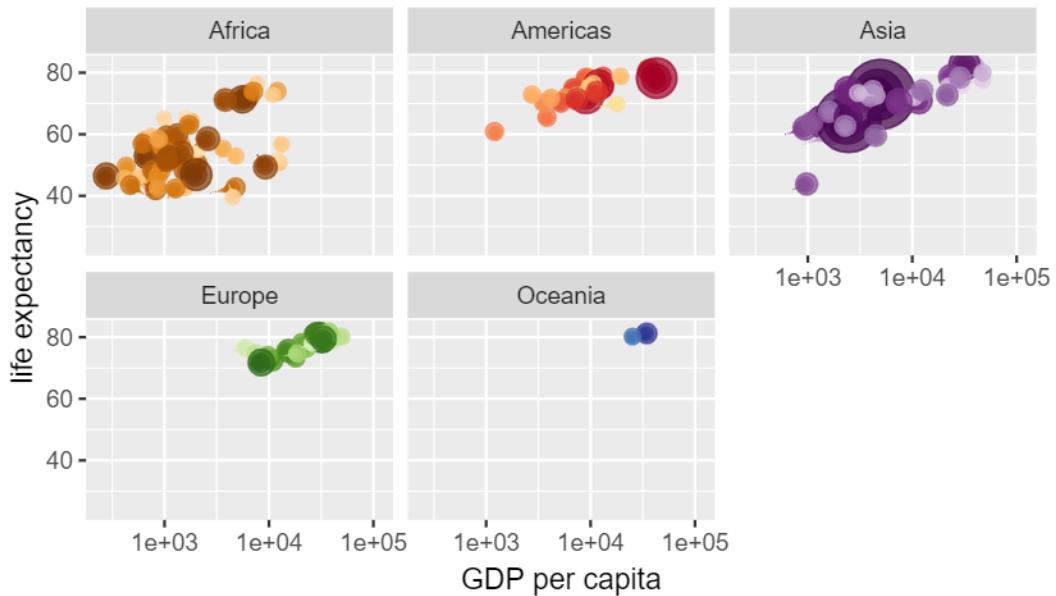
Year: 2006



Year: 2006



Year: 2007



#Combining plots

```
library(patchwork)

# Setup some example plots
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))

# Use patchwork to combine them here:
(p1 | p2 | p3) /
  p4

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

