R Exercises

October 27, 2019

Exercise 1

Q1. Which of the following will NOT return TRUE?

A. FALSE == FALSE
B. 10-5 == sqrt(25)
C. TRUE > FALSE
D. 'a' > 'b'

Q2. What is the output when we execute the following code? x <- 3 y <- 2 y <- 17.4 x+y

A. [1] 3 2 17.4
B. [1] 22.4
C. [1] 20.4
D. [1] 5

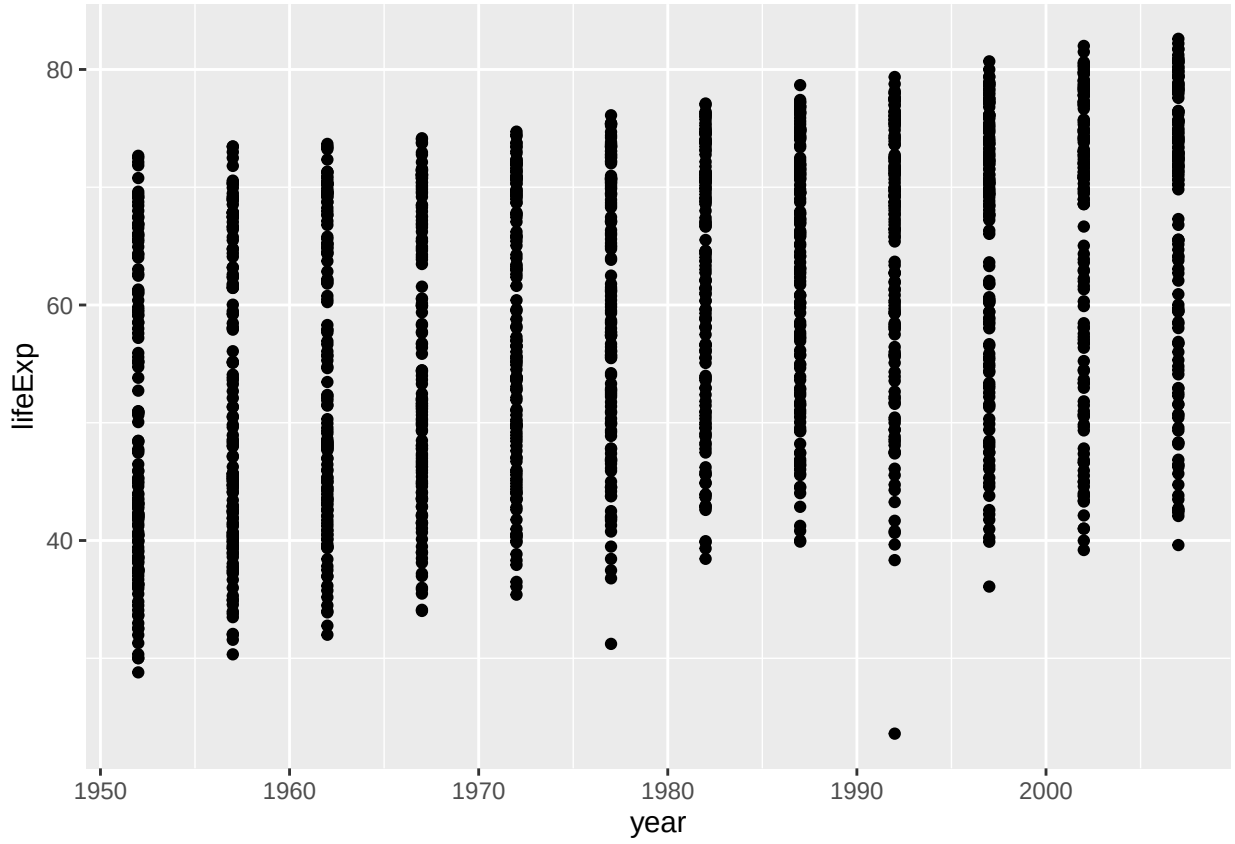
Q3. Use `str()` to look at the structure of the dataframe and `summary()` to get information about the variables.

What are its columns?
How many rows and columns are there?
What is the earliest year in the `year` column?
What is the average life expectancy?
What is the largest population?

Exercise 2 : ggplot

3a. How would we look at life expectancy overtime with ggplot?

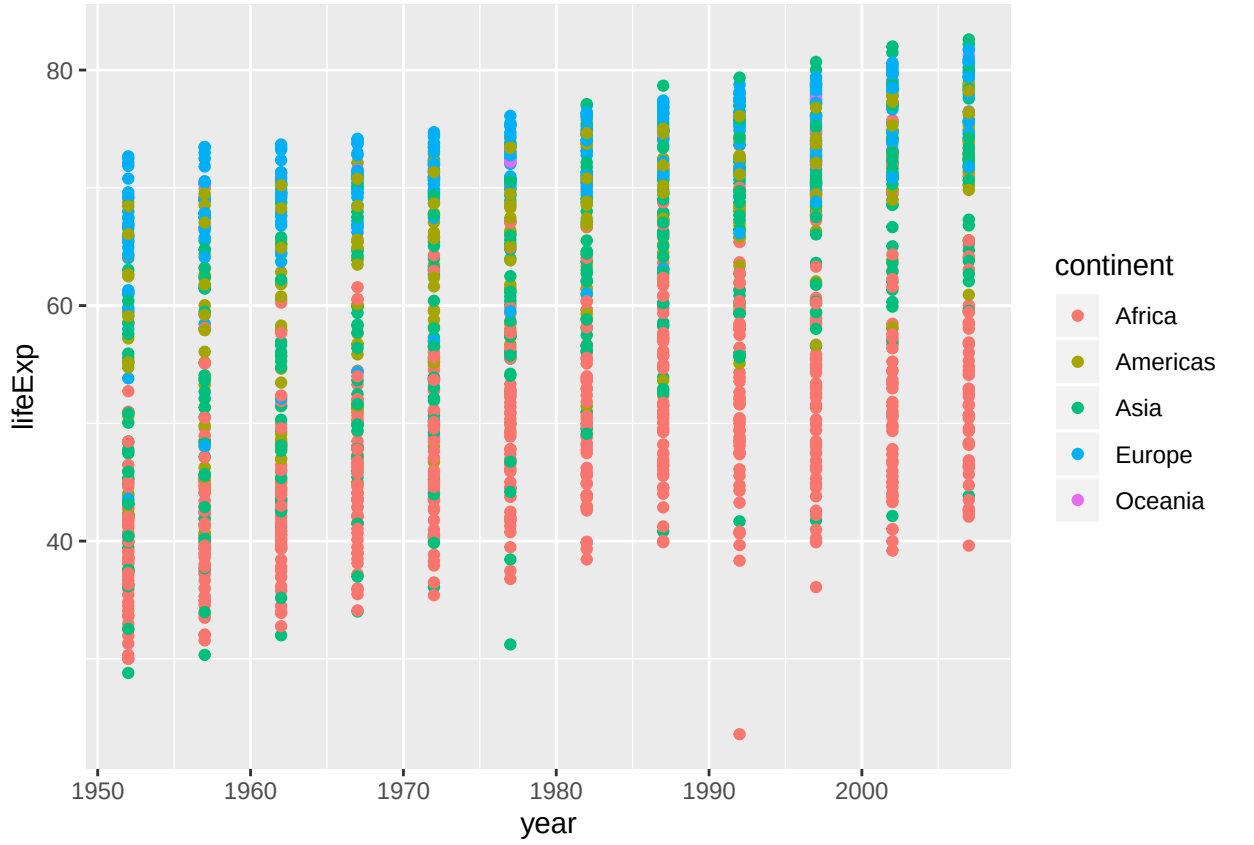
```
ggplot(data = gapminder, mapping = aes(x = year, y = lifeExp)) +
  geom_point()
```



3b. Try to add color.

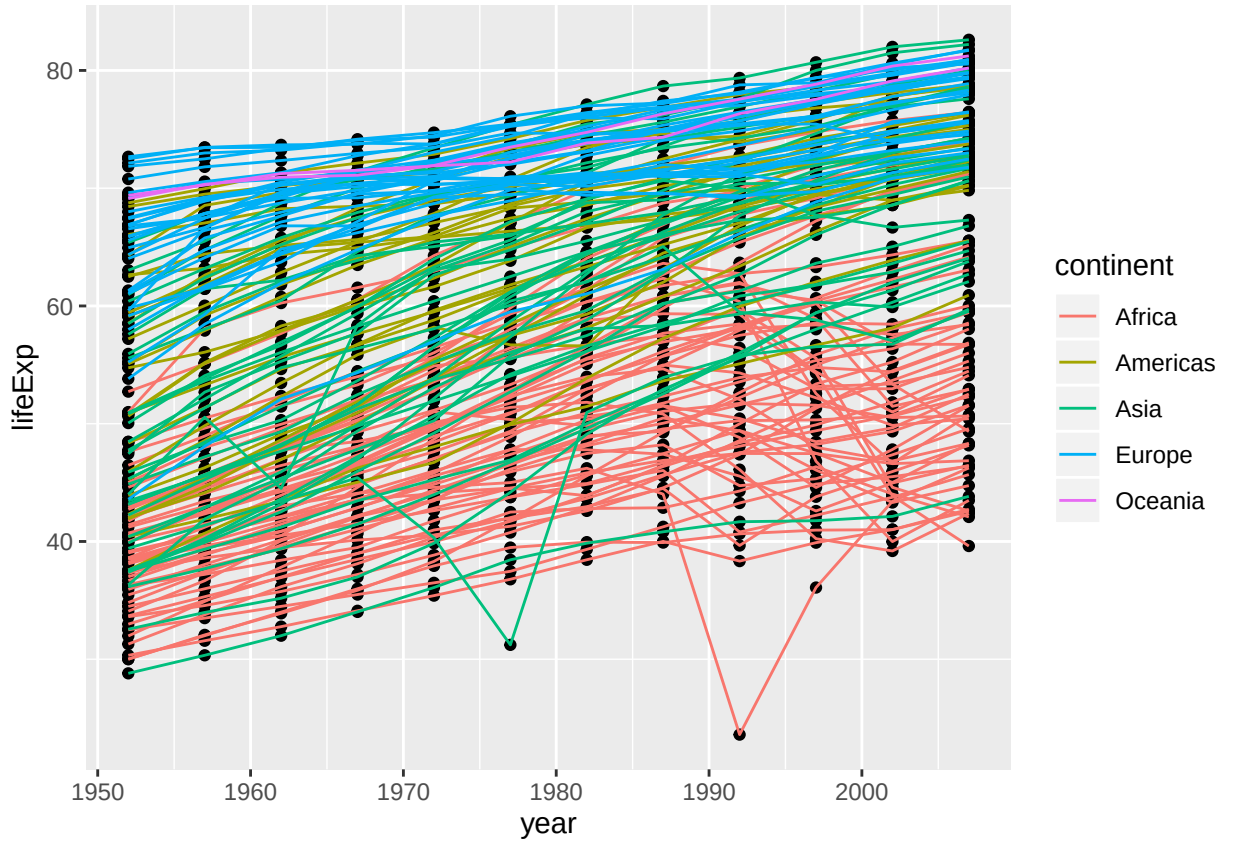
There is a color argument for the aes function
You can can tell ggplot to color by continent

```
ggplot(data = gapminder, mapping = aes(x = year, y = lifeExp, color = continent)) +
  geom_point()
```



3c. Switch the order of the geom_ functions. What happened?

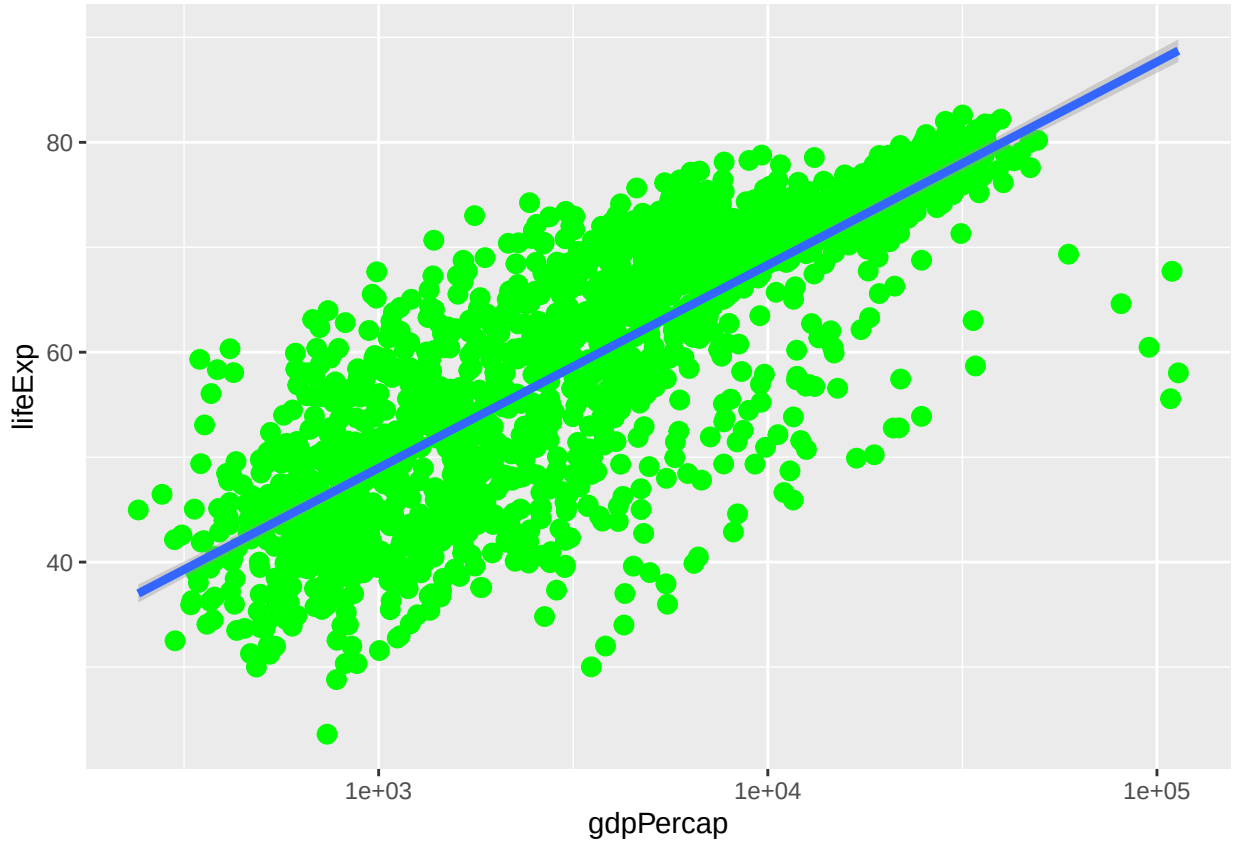
```
ggplot(data = gapminder, mapping = aes(x=year, y=lifeExp, by=country)) +
  geom_point() + geom_line(mapping = aes(color=continent))
```



3d. Change the color and size of the points outside of aes.

Note: ?geom_point

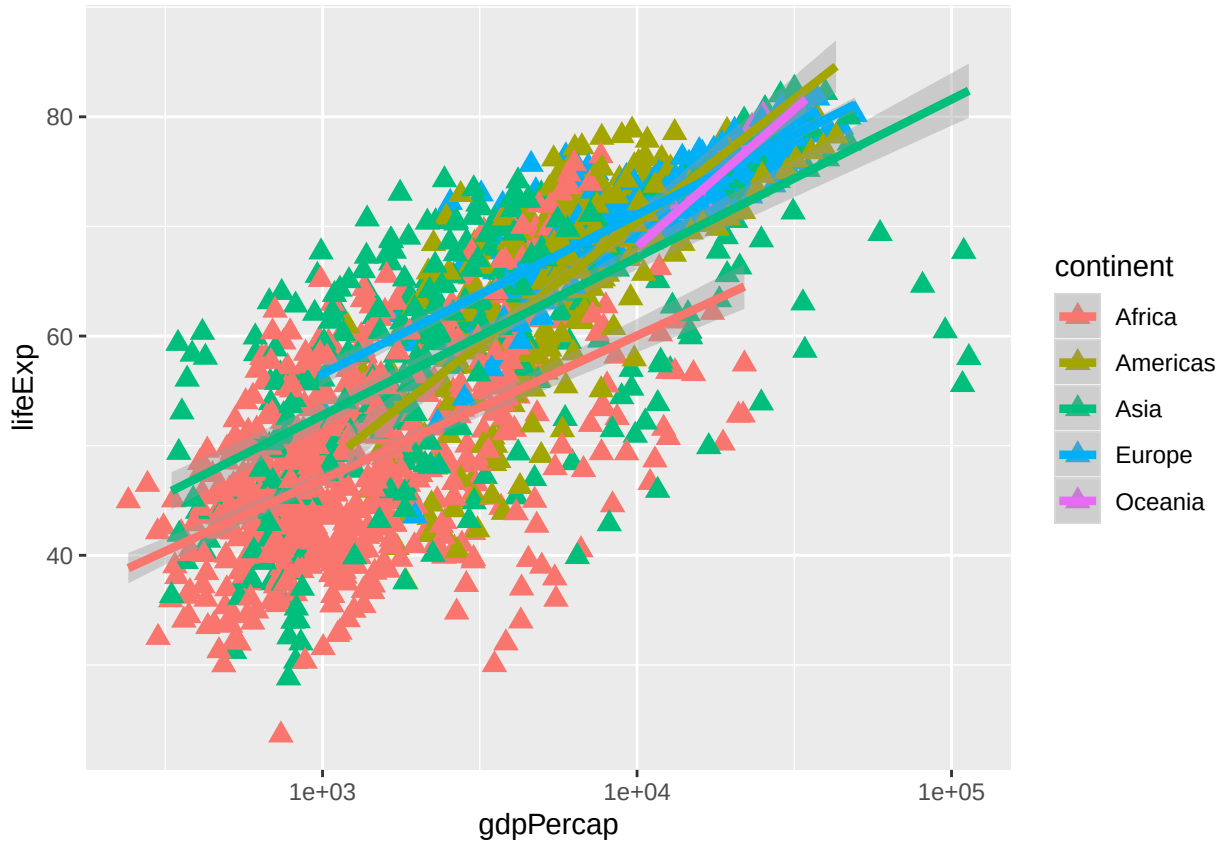
```
ggplot(data = gapminder, mapping = aes(x = gdpPercap, y = lifeExp)) +
  geom_point(size=3, color="green") + scale_x_log10() +
  geom_smooth(method="lm", size=1.5)
```



3e. Now change the shape of the points and color the points by the continent

Note: You'll get multiple linear model fits
Note: ?geom_point

```
ggplot(data = gapminder, mapping = aes(x = gdpPercap, y = lifeExp, color = continent)) +
  geom_point(size=3, shape=17) + scale_x_log10() +
  geom_smooth(method="lm", size=1.5)
```

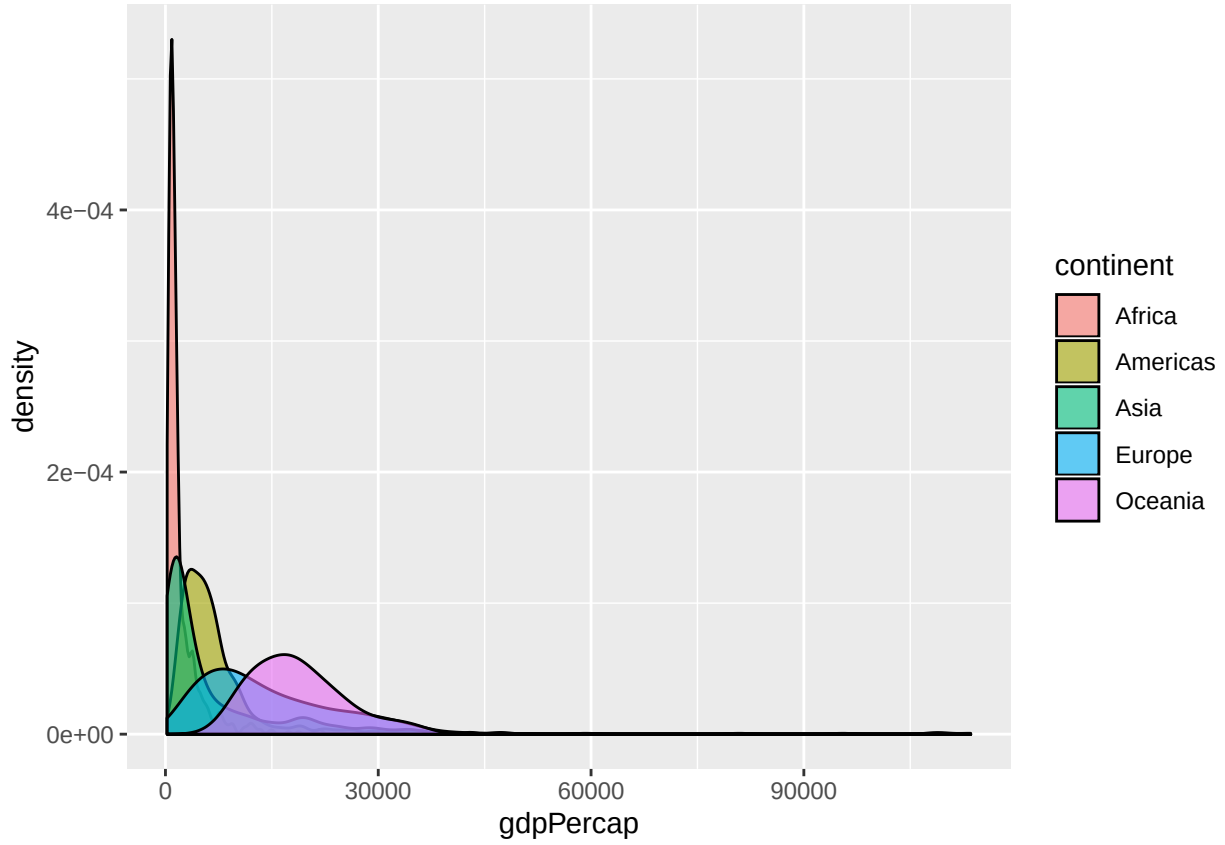


3f. Create a density plot of GDP per capita, filled by continent.

Advanced: Transform the x axis to better visualise the data spread.
Add a facet layer to panel the density plots by year.

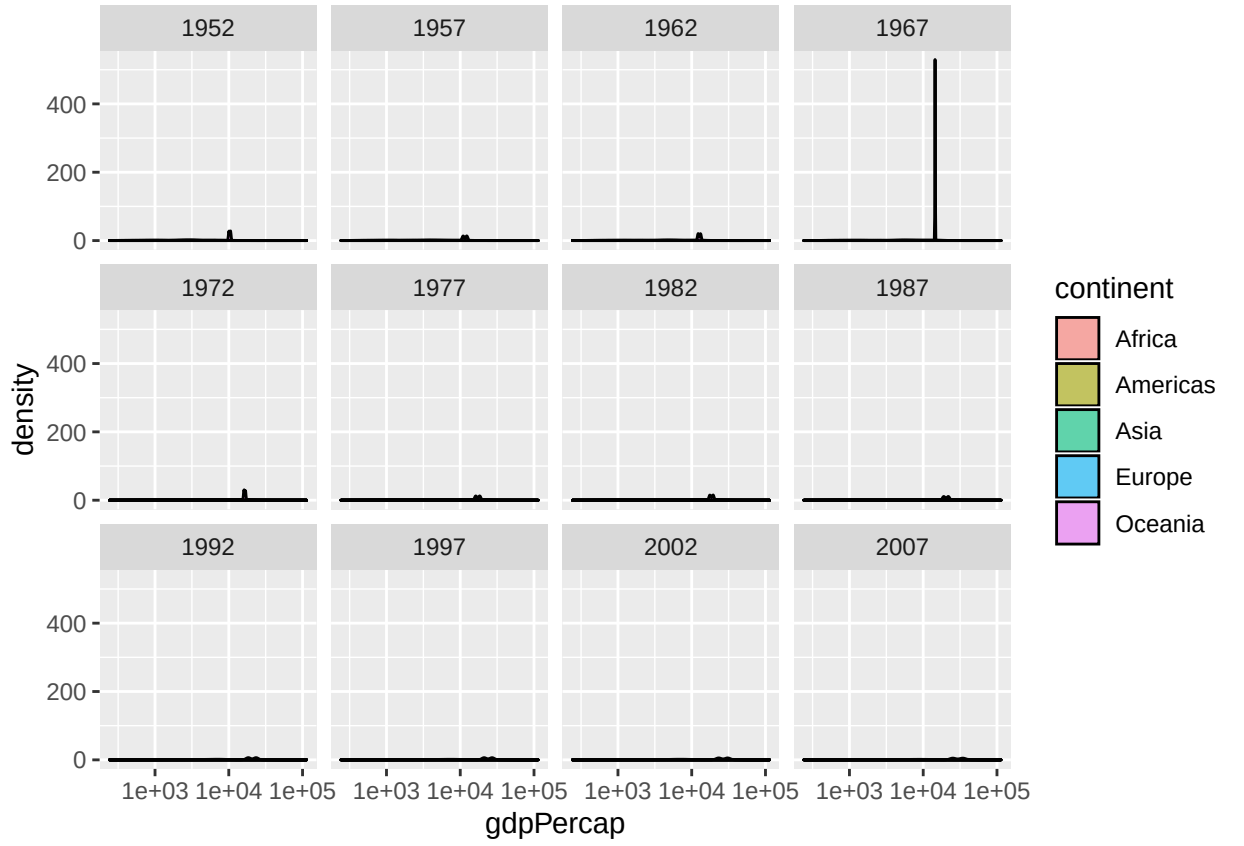
Answer:

```
ggplot(data = gapminder, mapping = aes(x = gdpPercap, fill=continent)) +
  geom_density(alpha=0.6)# + facet_wrap( ~ year) + scale_x_log10()
```



Advanced answer:

```
ggplot(data = gapminder, mapping = aes(x = gdpPercap, fill=continent)) +
  geom_density(alpha=0.6) + facet_wrap( ~ year) + scale_x_log10()
```



Exercise 3: loops

Write a script that loops through the gapminder data by continent and prints out whether the mean life expectancy is smaller or larger than 50 years.

Answer

```
meanLifeExp <- mean(gapminder$lifeExp)

for(cont in unique(gapminder$continent)){
  if(mean(gapminder[gapminder$continent == cont, 'lifeExp']) > meanLifeExp){
    print(paste(cont, 'lifeExp is above average'))
  }
  else if (mean(gapminder[gapminder$continent == cont, 'lifeExp']) < meanLifeExp){
    print(paste(cont, 'lifeExp if below average'))
  }
  else{
    print(paste(cont, 'lifeExp is average'))
  }
}
```

```
## [1] "Asia lifeExp is above average"
## [1] "Europe lifeExp is above average"
## [1] "Africa lifeExp if below average"
## [1] "Americas lifeExp is above average"
## [1] "Oceania lifeExp is above average"
```

Exercise 4: Dplyr

4a. Write a single command (which can span multiple lines and includes pipes) that will produce a dataframe that has the African values for lifeExp, country and year, but not for other Continents. How many rows does your dataframe have and why?

```
year_country_lifeExp_Africa <- gapminder %>%
  filter(continent == "Africa") %>%
  select(year, country, lifeExp)
head(year_country_lifeExp_Africa)
```

```
##   year country lifeExp
## 1 1952 Algeria  43.077
## 2 1957 Algeria  45.685
## 3 1962 Algeria  48.303
## 4 1967 Algeria  51.407
## 5 1972 Algeria  54.518
## 6 1977 Algeria  58.014
```

4b. Create a new dataframe that contains the minimum (MinExp) and maximum (MaxExp) life expectancies for each country.

```
MinMaxExp <- gapminder %>%
  group_by(country) %>%
  summarize(MinExp = min(lifeExp),
            MaxExp = max(lifeExp))

head(MinMaxExp)
```

```
## # A tibble: 6 x 3
##   country     MinExp MaxExp
##   <chr>        <dbl>  <dbl>
## 1 Afghanistan   28.8   43.8
## 2 Albania       55.2   76.4
## 3 Algeria       43.1   72.3
## 4 Angola        30.0   42.7
## 5 Argentina     62.5   75.3
## 6 Australia     69.1   81.2
```

4c. Calculate the average life expectancy per country. Which has the longest average life expectancy and which has the shortest average life expectancy?

```
gapminder %>%
  group_by(country) %>%
  summarize(mean_lifeExp = mean(lifeExp)) %>%
  filter(mean_lifeExp == min(mean_lifeExp) | mean_lifeExp == max(mean_lifeExp))
```

```
## # A tibble: 2 x 2
##   country      mean_lifeExp
##   <chr>               <dbl>
## 1 Iceland              76.5
## 2 Sierra Leone         36.8
```