

HW 3

Frank Edwards

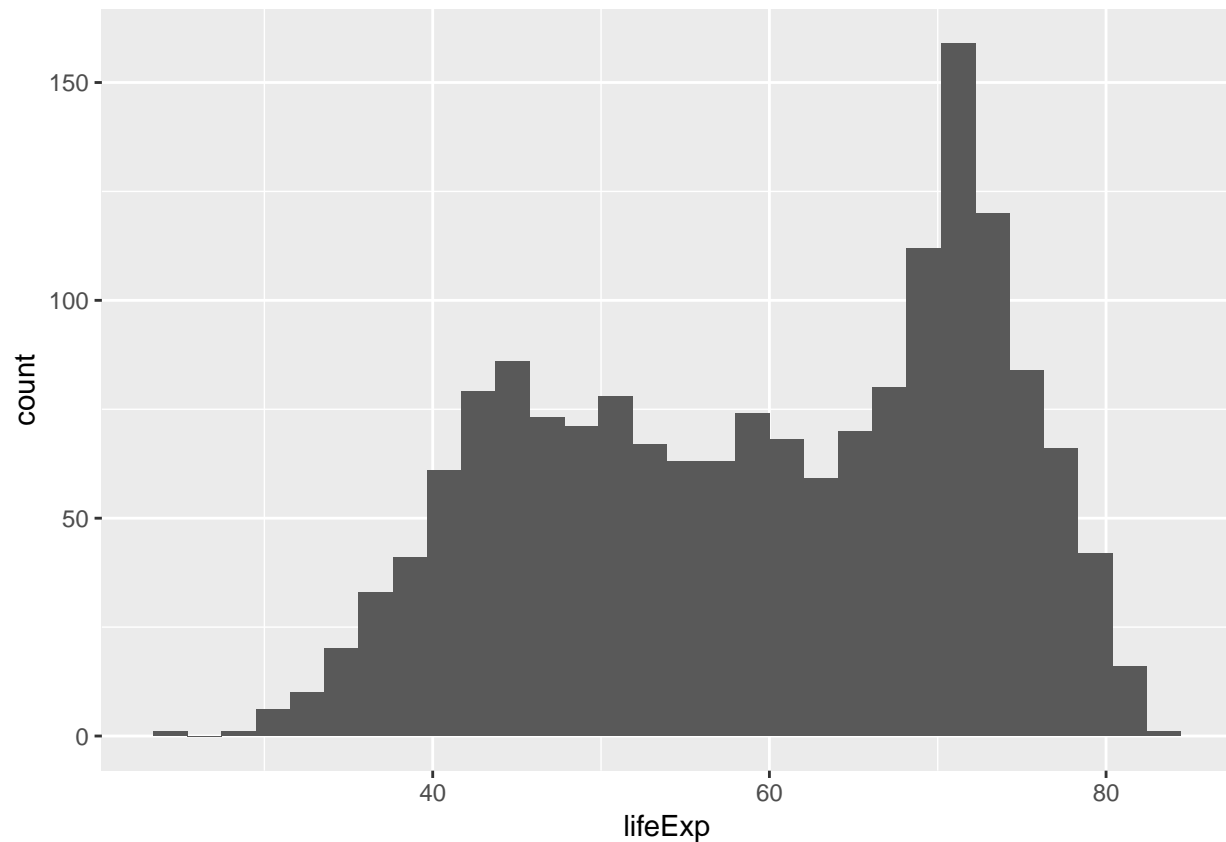
2/15/2019

Question 1

1. Provide a histogram of life expectancy (lifeExp). Interpret this histogram.

```
ggplot(gapminder, aes(x = lifeExp)) +  
  geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

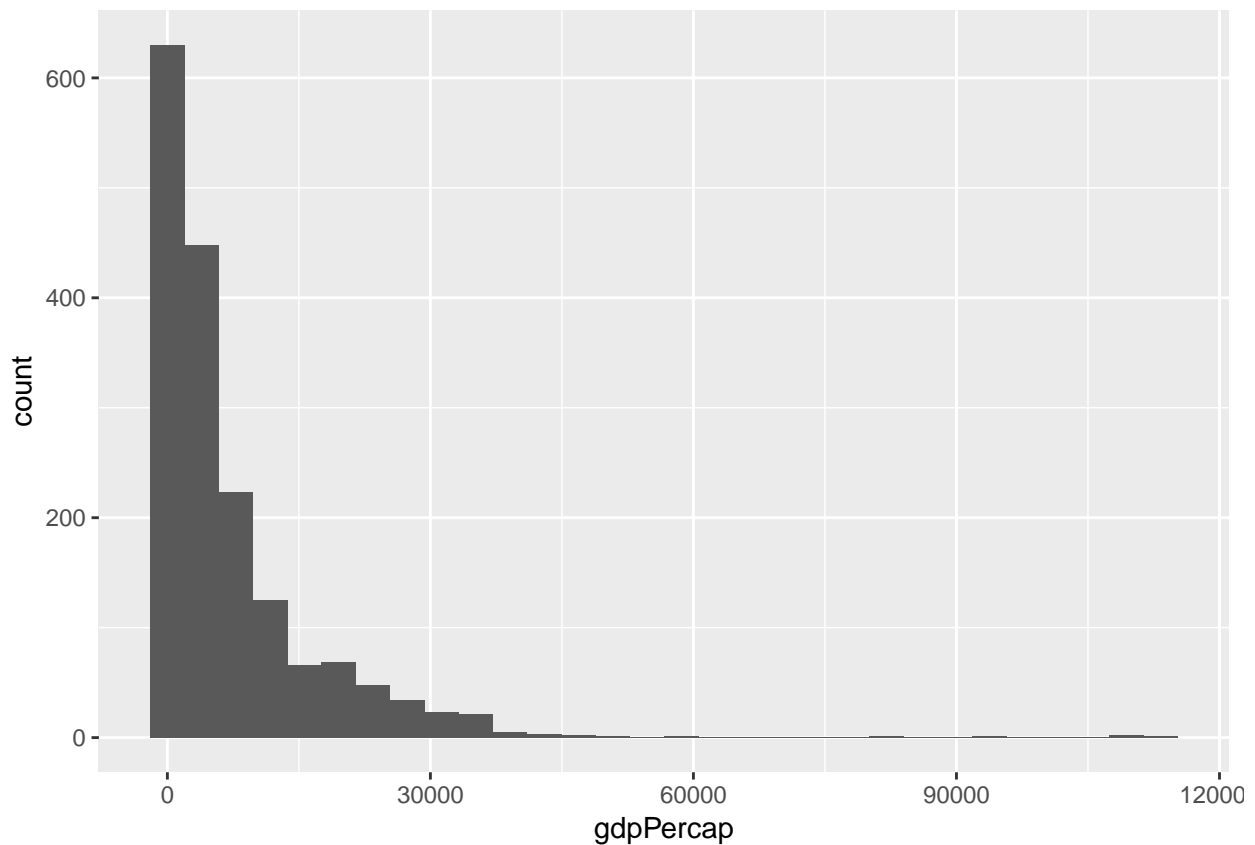


This histogram shows life expectancy across the countries in the Gapminder data. It shows two distinct clusters of countries, indicated by two relative modal values. One group of countries has relatively low life expectancy. The histogram has a peak at around a life expectancy of 45. Another group has relatively high life expectancy, with a peak at about 75 years.

2. Provide a histogram of GDP per capita (gdpPercap). Interpret this histogram.

```
ggplot(gapminder,  
  aes(x = gdpPercap)) +  
  geom_histogram()
```

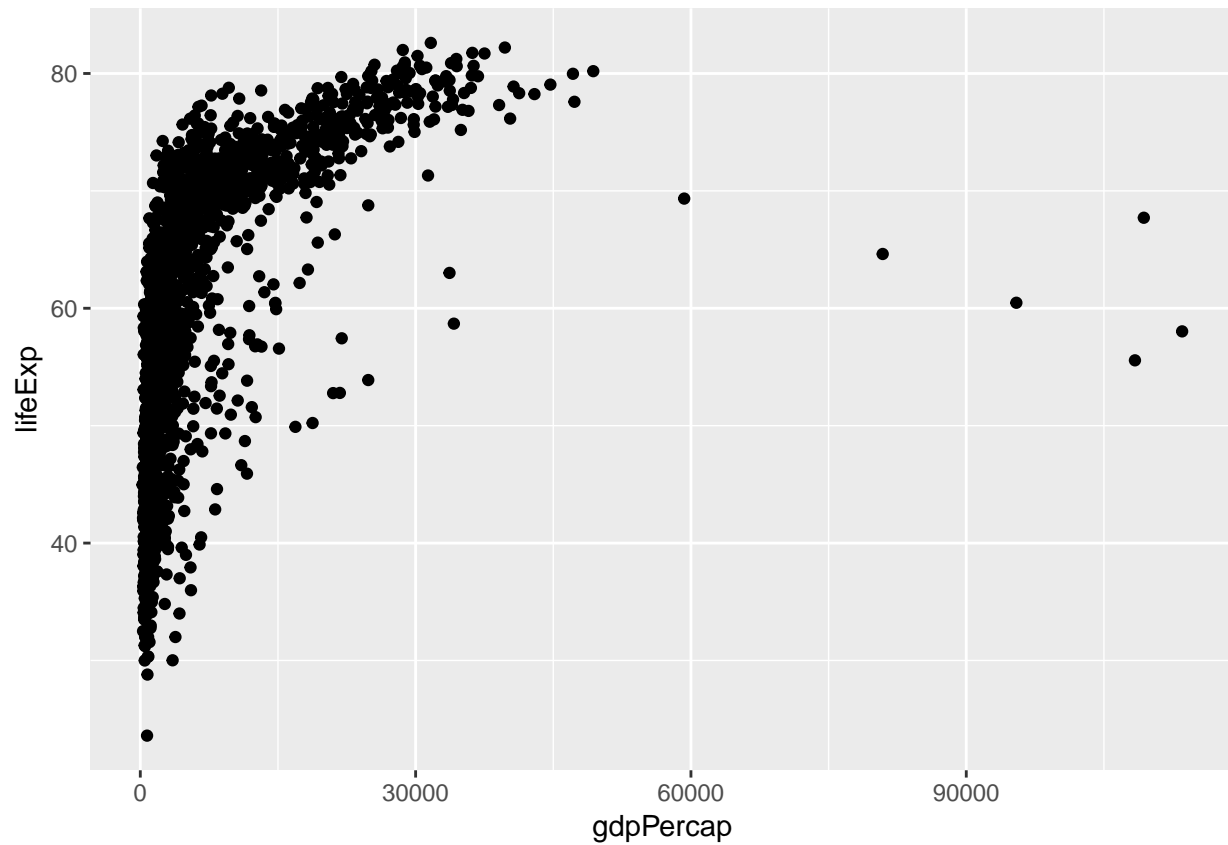
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



This histogram shows the distribution of GDP per capita across the countries in the gapminder data. It shows that the bulk of countries have relatively low GDP per capita, but that a handful of countries have extremely high GDP per capita. It suggests a high level of inequality in GDP per capita across countries.

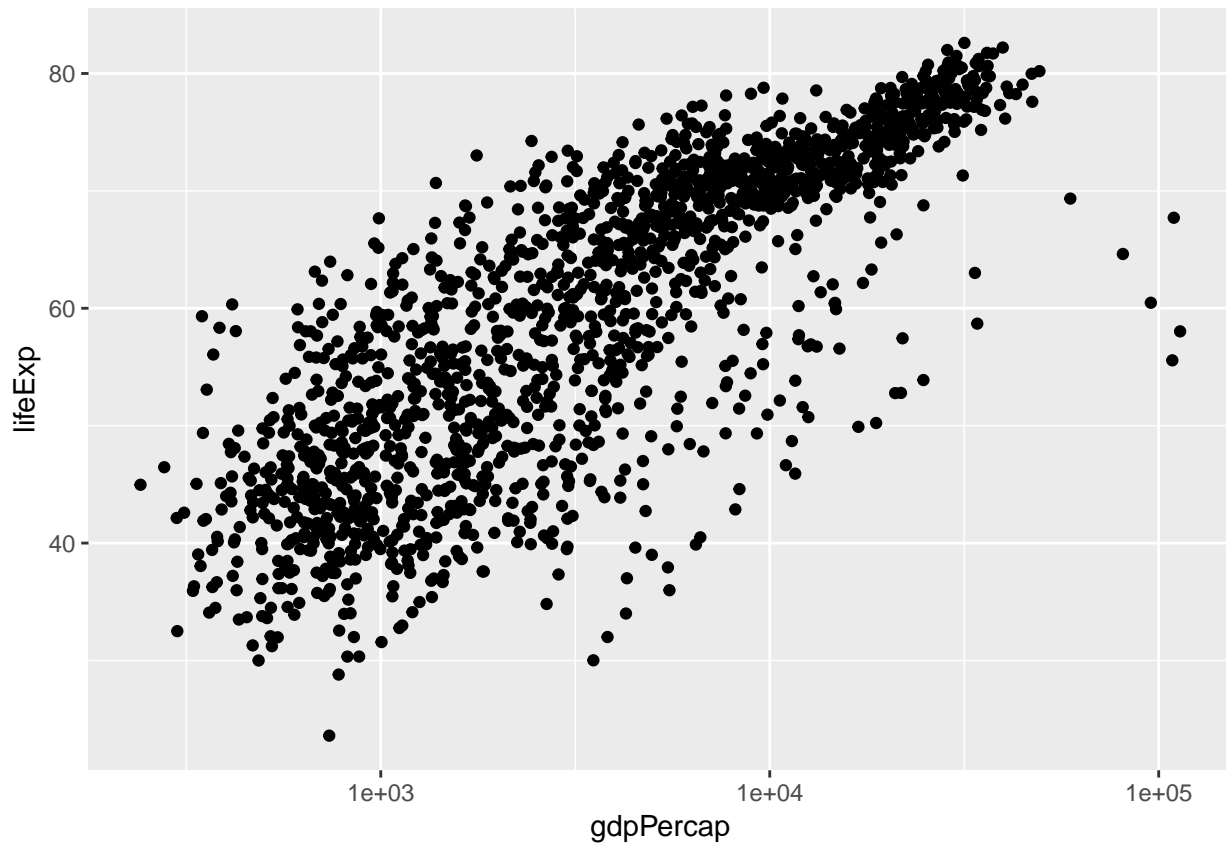
3. Provide a scatterplot of life expectancy by GDP per capita

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



4. Transform GDP per capita using a log scale either manually or using `scale_x_log10()`

```
ggplot(gapminder,  
  aes(x = gdpPercap, y = lifeExp)) +  
  geom_point() +  
  scale_x_log10()
```



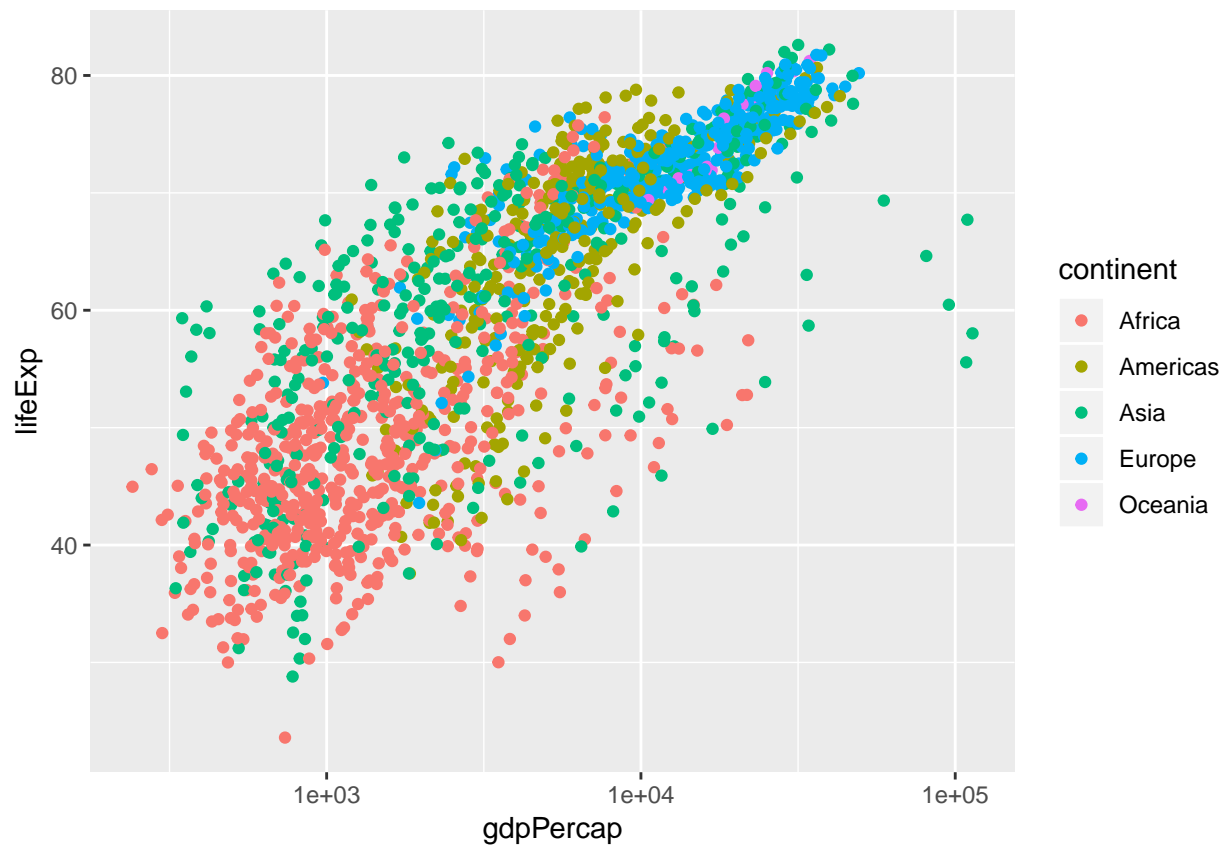
5. Interpret the relationship between life expectancy and gdp per capita

The relationship between life expectancy and $\log(\text{GDP per capita})$ appears to be linear. The scatterplot indicates that countries with higher levels of GDP tend to have higher life expectancies. Countries with lower levels of GDP, $\text{GDP}_{percap} < 1000$ all have life expectancies of less than 70 years, with most countries in this range falling below a life expectancy of 60 years. We could further quantify this relationship by estimating a linear regression model:

Question 2

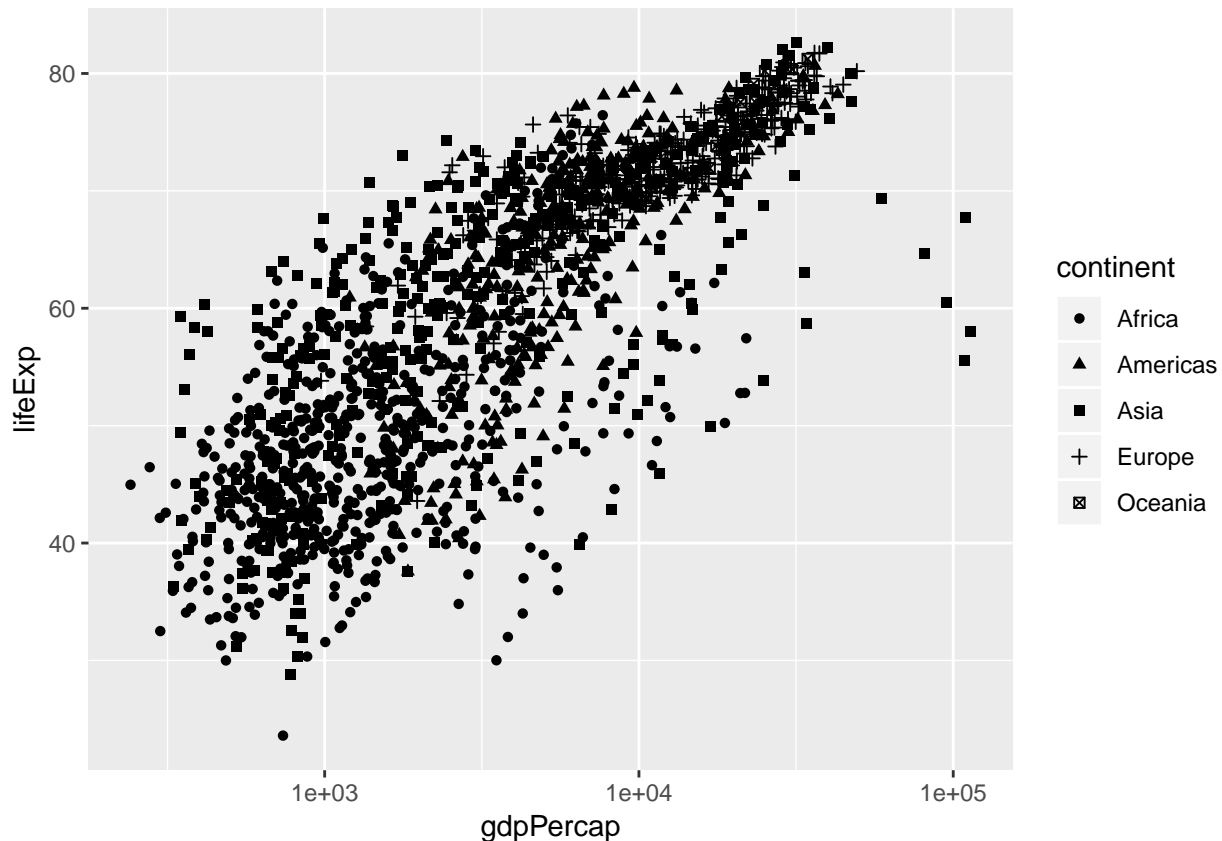
1. Using your scatterplot from 1.5., add a color aesthetic to your plot for continent.

```
ggplot(gapminder,
       aes(x = gdpPercap, y = lifeExp, color = continent)) +
  geom_point() +
  scale_x_log10()
```



2. Now, replace your color aesthetic with a shape aesthetic.

```
ggplot(gapminder,  
  aes(x = gdpPerCap, y = lifeExp, shape = continent)) +  
  geom_point() +  
  scale_x_log10()
```



3. Which plot is easier to interpret, 2.1 or 2.2?

2.1 is much easier to interpret. Color makes the levels clearly distinguishable. Shapes in scatterplots blur together when clustered.

Question 3

1. Estimate a regression equation for the relationship between life expectancy, GDP per capita, and continent.

```
m1<-lm(lifeExp~log(gdpPercap) * continent,
      data = gapminder)
```

I estimate a linear regression of the relationship between life expectancy, GDP per capita, and continent. I interact the terms for log GDP per capita and continent to allow both the estimated intercepts and slopes of regression lines to vary by continent.

2. Provide a display of the estimated model using broom::tidy()

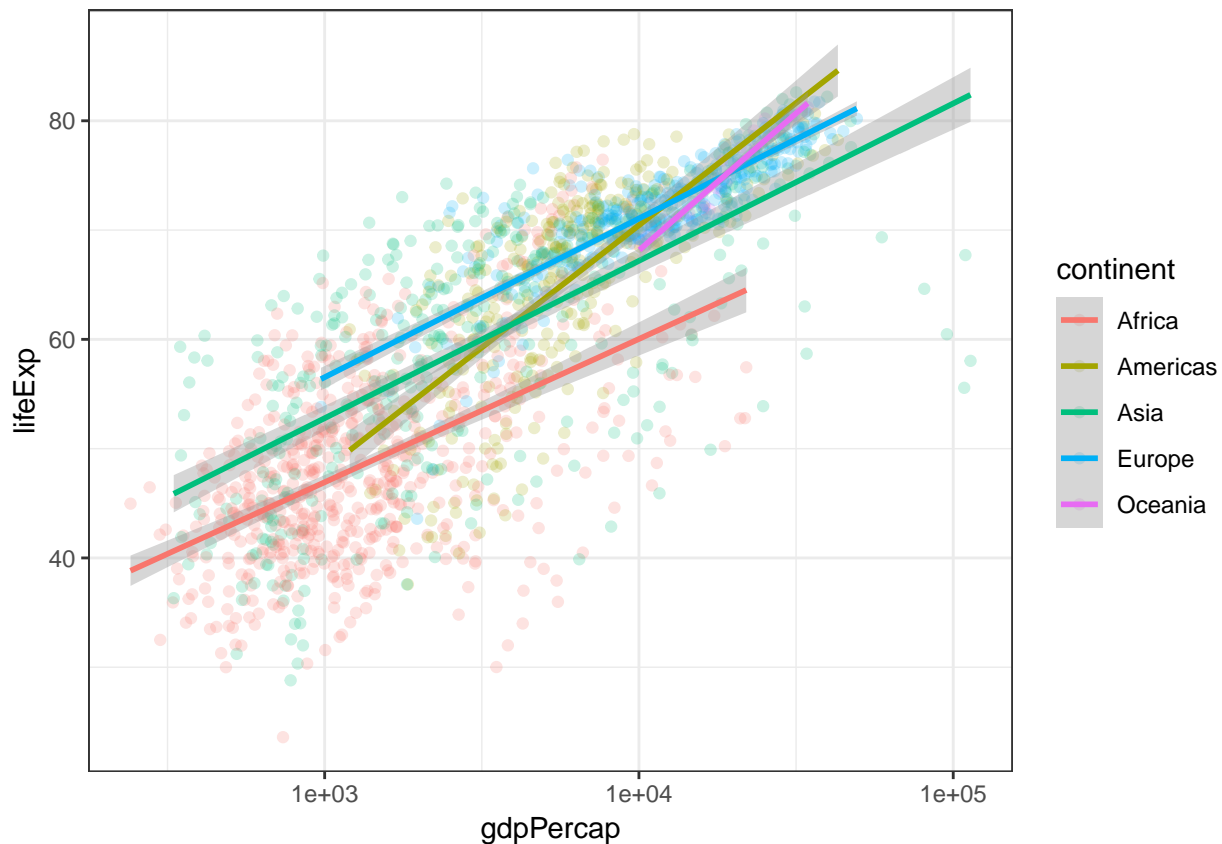
```
tidy(m1)
```

```
## # A tibble: 10 x 5
##   term                                estimate std.error statistic  p.value
##   <chr>                                <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept)                        7.60      2.37      3.21 1.37e- 3
## 2 log(gdpPercap)                     5.69      0.325    17.5 2.48e-63
## 3 continentAmericas                  -26.7      5.63     -4.74 2.31e- 6
## 4 continentAsia                      1.98      3.23      0.614 5.39e- 1
## 5 continentEurope                    5.37      5.26      1.02 3.08e- 1
## 6 continentOceania                   -39.8     42.6     -0.934 3.50e- 1
```

```
## 7 log(gdpPercap):continentAmericas 4.03 0.674 5.97 2.80e- 9
## 8 log(gdpPercap):continentAsia 0.561 0.421 1.33 1.83e- 1
## 9 log(gdpPercap):continentEurope 0.614 0.597 1.03 3.04e- 1
## 10 log(gdpPercap):continentOceania 5.20 4.36 1.19 2.33e- 1
```

3. Provide a visual that summarizes your estimated relationship

```
ggplot(gapminder,
  aes(x = gdpPercap, y = lifeExp, color = continent)) +
  geom_point(alpha = 0.2) +
  scale_x_log10() +
  geom_smooth(method = "lm") +
  theme_bw()
```



The plotted regression lines show the range of the observed data and differences in level and slope by continent. Africa has the lowest average life expectancy, but there is substantial variation across the continent over time. The steepest slopes are in Oceania and the Americas, indicating a larger difference between the lowest life expectancies and the highest life expectancies on these continents for the country-years in the data.