The Grammar of Graphics

Answers

```
library(tidyverse)
library(magrittr)
library(gapminder)
data(gapminder)
```

1. Slide 38

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

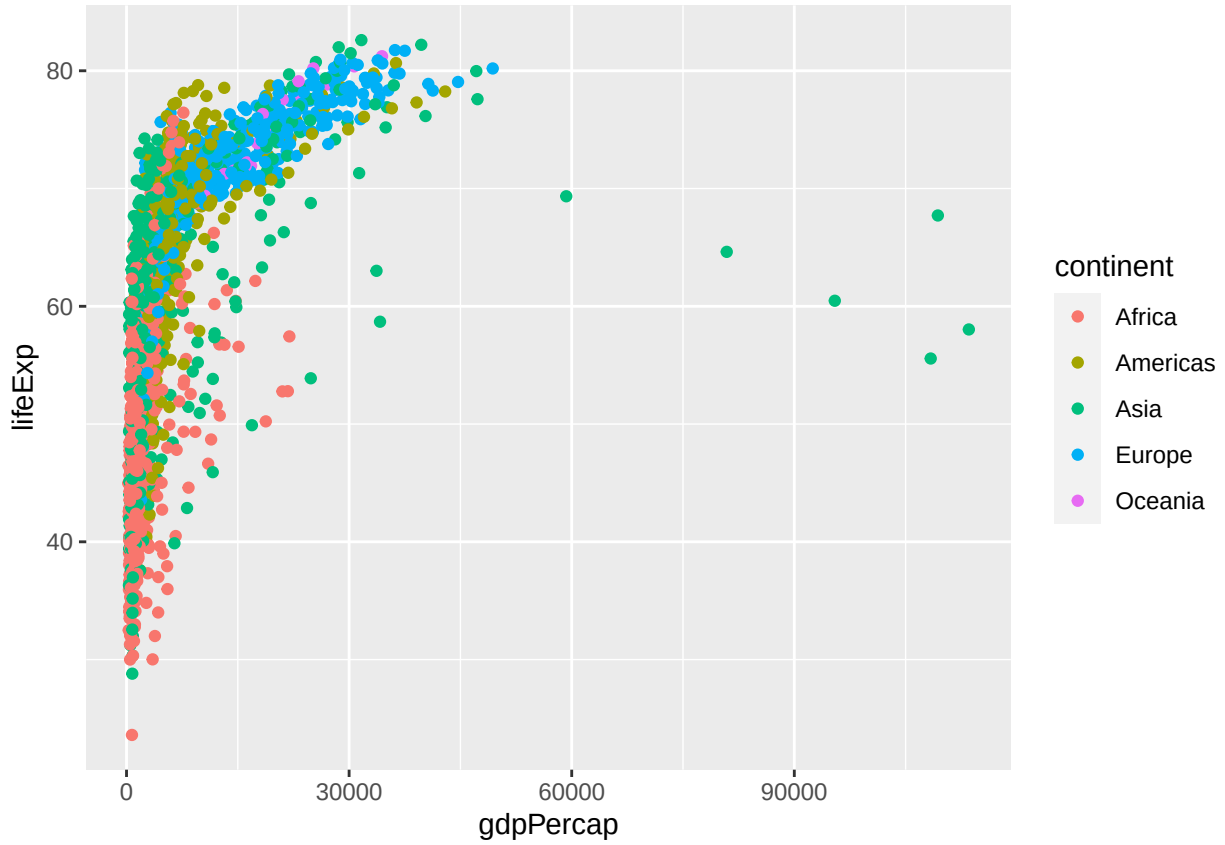


Figure 1: Figure 1

It's impossible to see whats going on in Figure 1! Now lets look at the same graph with both axes scaled by factors of 10 (in other words, on logarithmic axes).

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

Figure 2: Figure 2

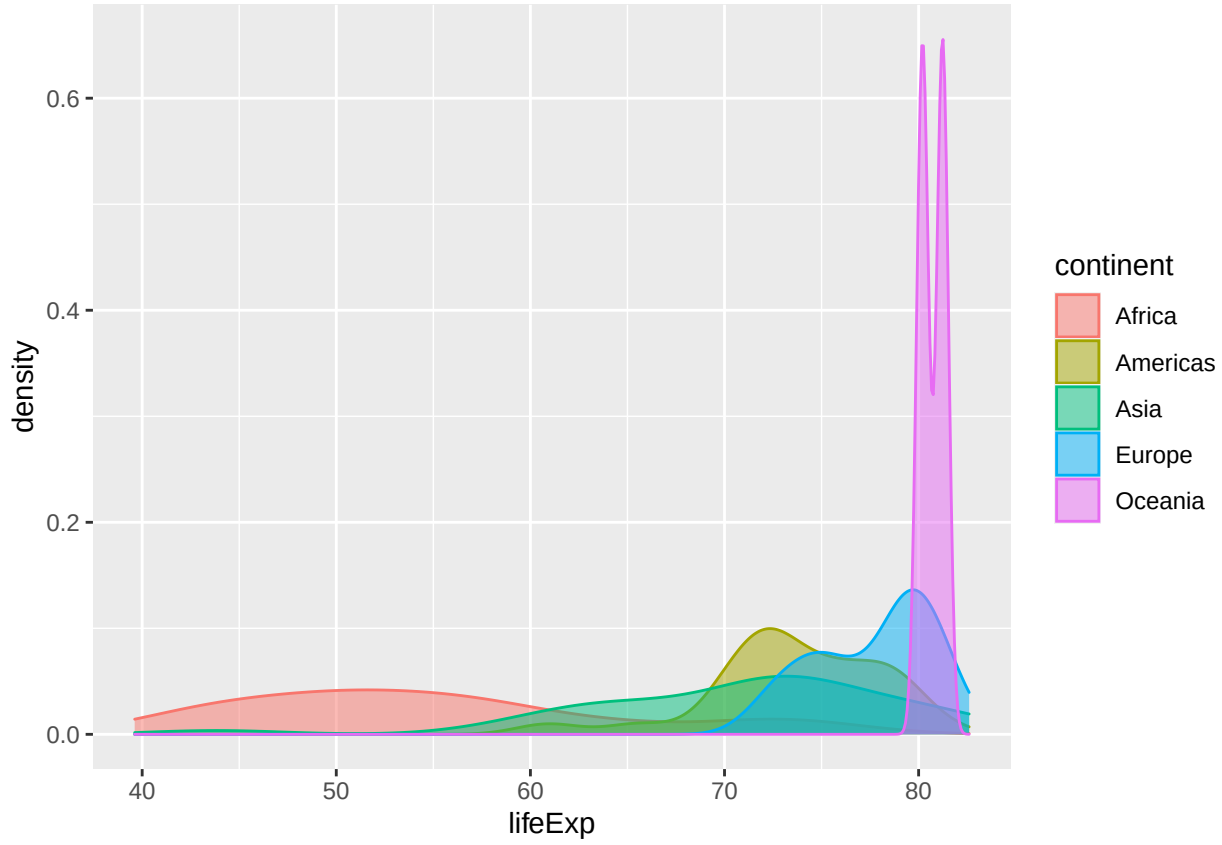
In Figure 2, the points are spread pretty evenly along both axes - so you can really see what's going on.

When does one use the logarithmic scale to visualize data instead of the absolute values?

- When the dataset contains numbers of very different magnitudes, you should consider whether a log transform will enhance the visualization. A log transformation preserves the order of the observations while making outliers less extreme.

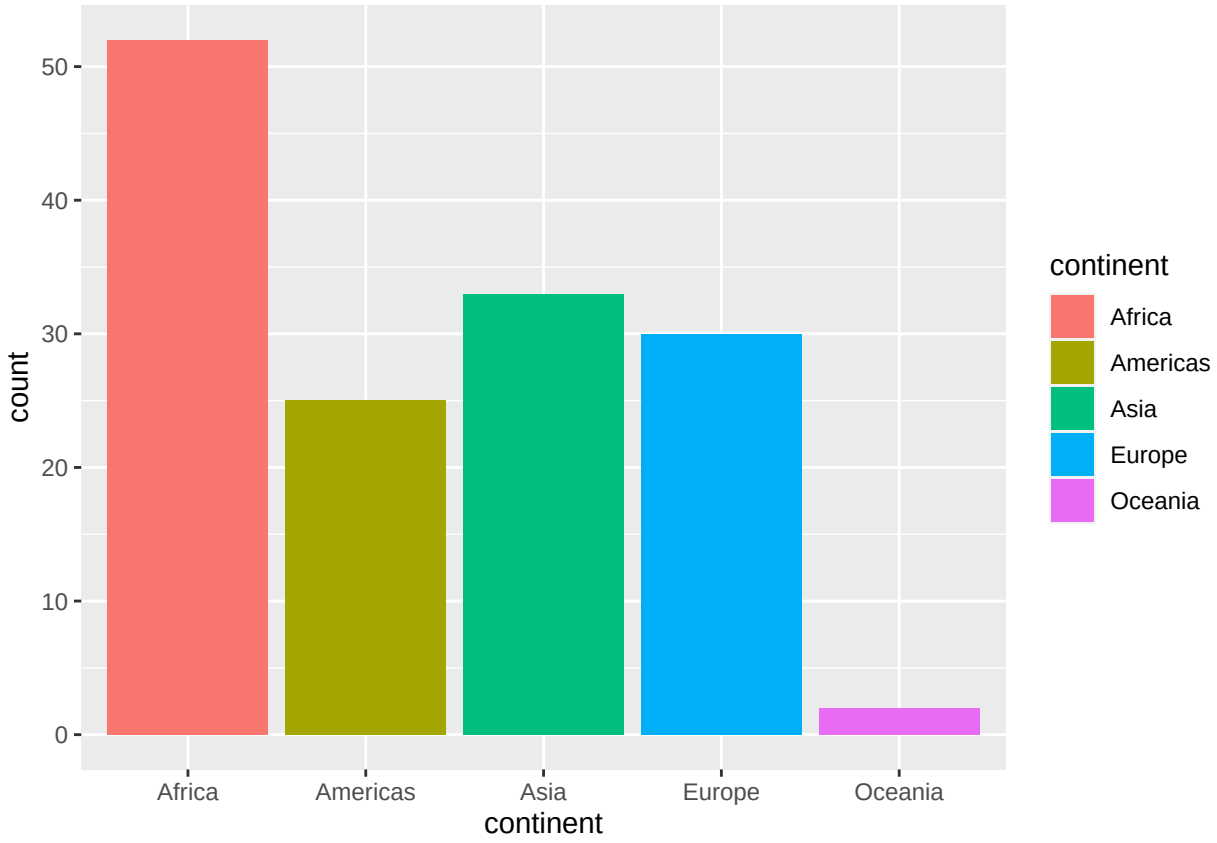
2. Slide 63

```
gapminder2007 <- gapminder %>%
  filter(year == 2007)
ggplot(gapminder2007,
aes(x=lifeExp, col=continent, fill=continent))+
  geom_density(alpha=0.5)
```



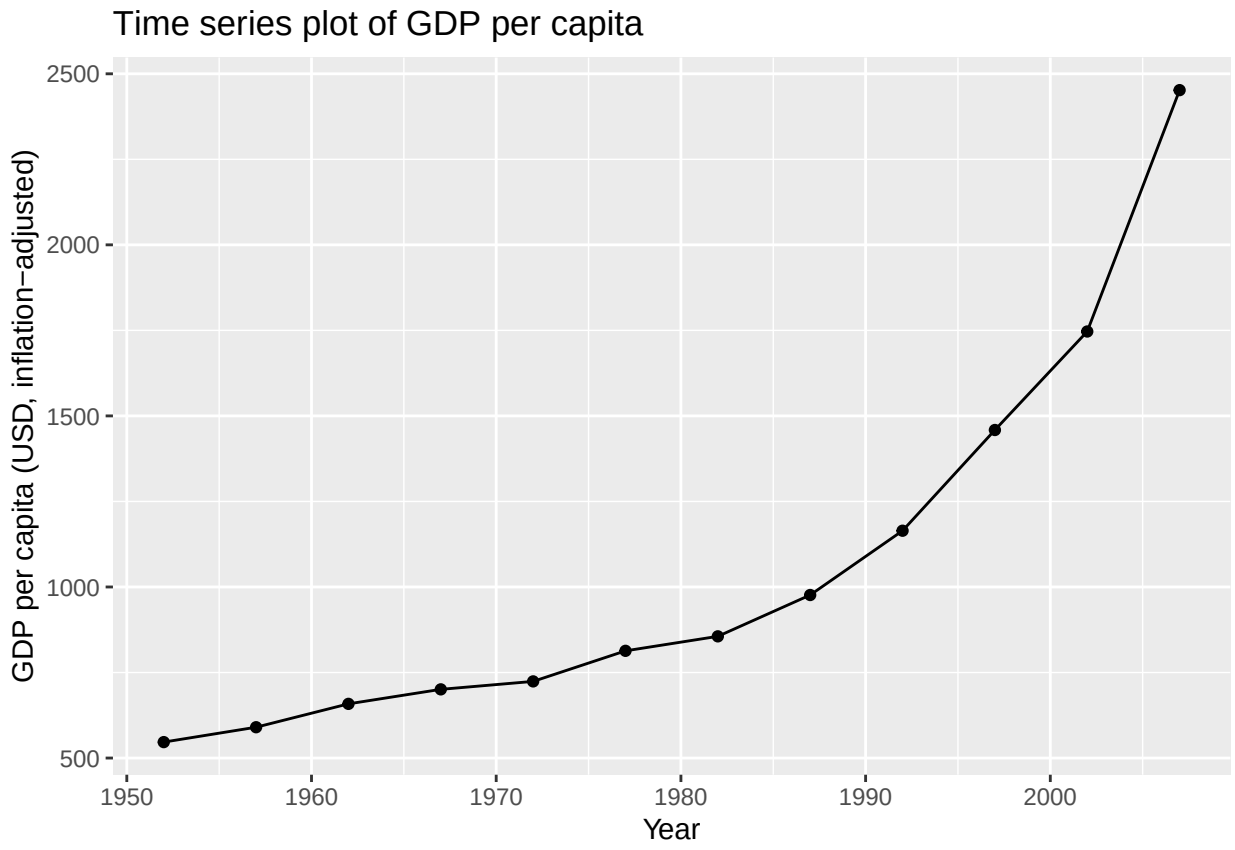
3. Slide 66

```
ggplot(gapminder2007,
aes(x=continent, fill=continent))+
 geom_bar()
```



4. Slide 73

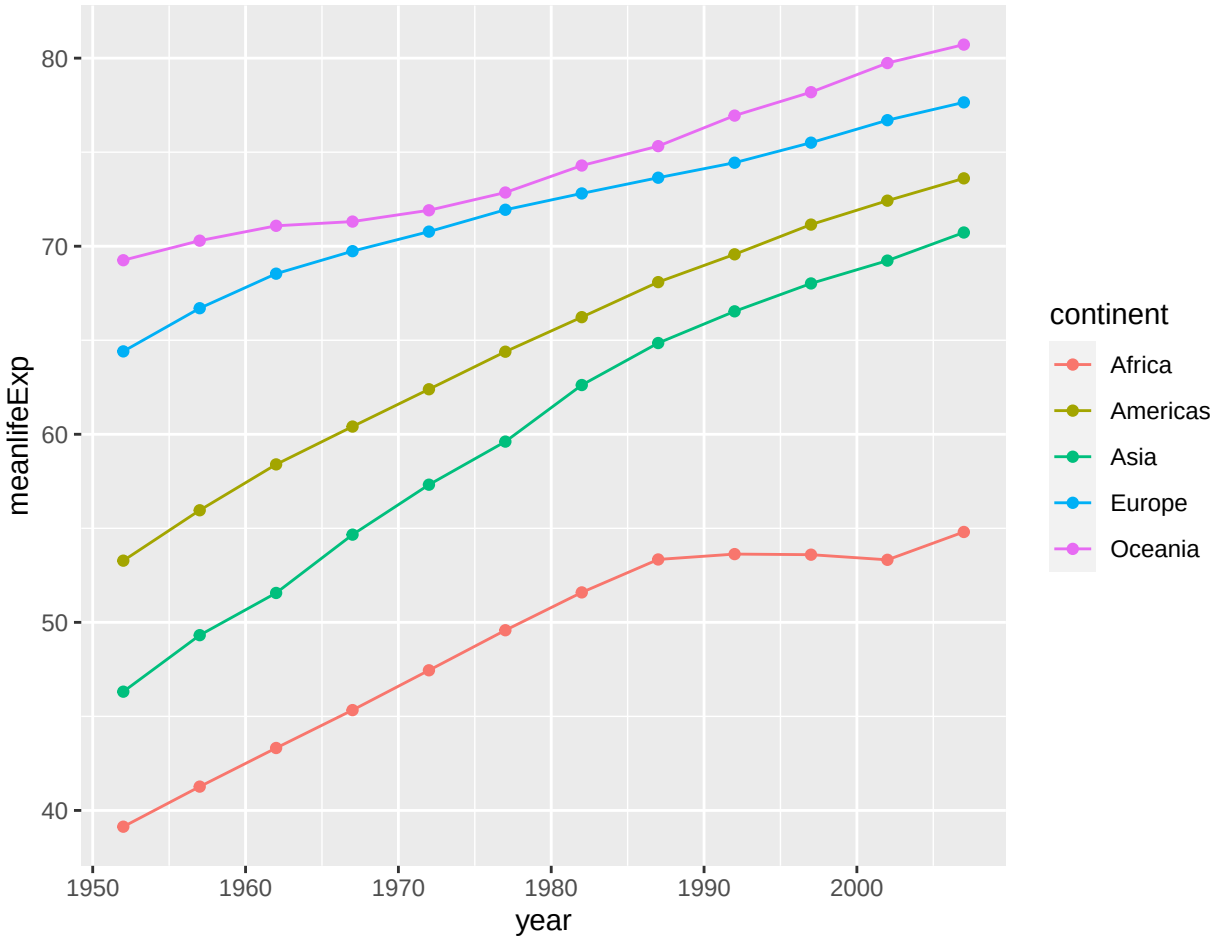
```
gapminder %>%
filter(country == "India") %>%
ggplot(aes(x = year, y = gdpPercap)) +
geom_line() +
  geom_point() +
  labs(title="Time series plot of GDP per capita",
       x="Year",
       y="GDP per capita (USD, inflation-adjusted)")
```



5. Slide 75

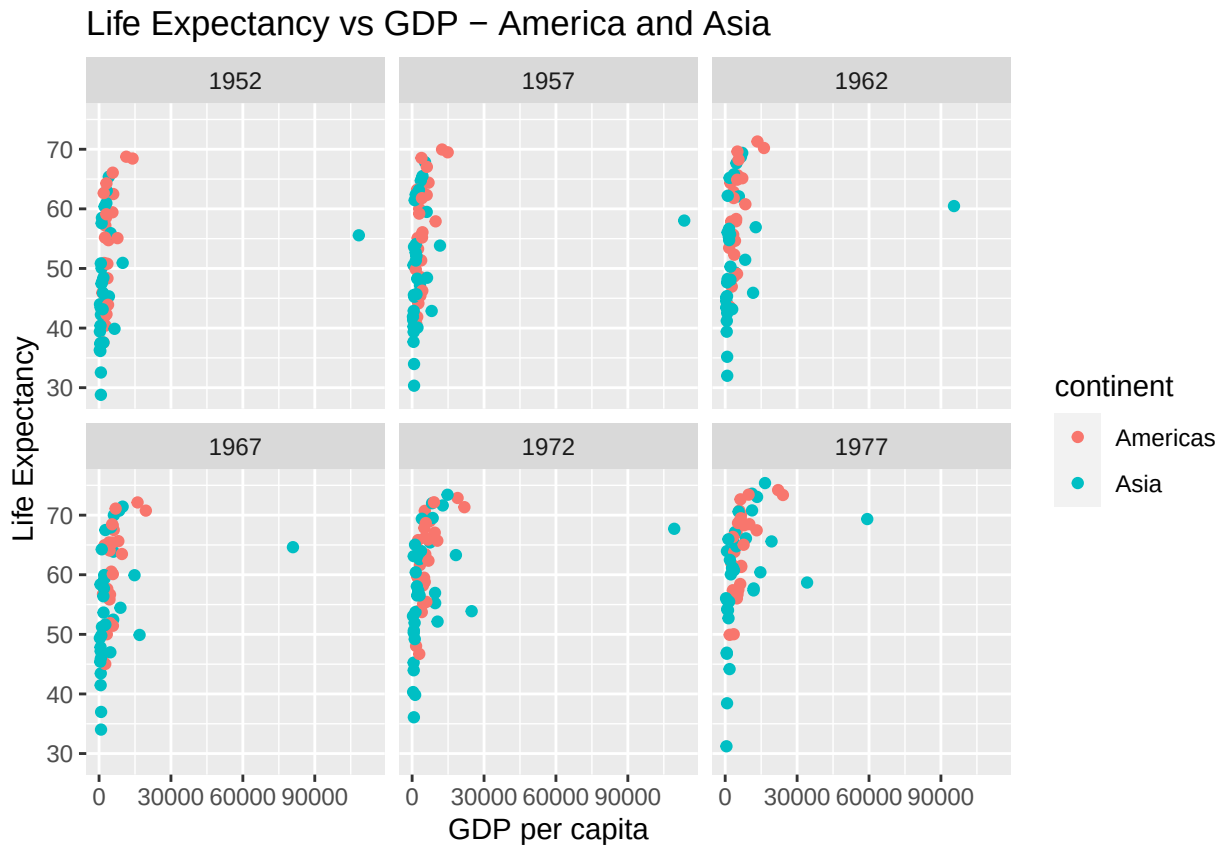
```
avglifeExp <- gapminder %>%
  group_by(continent, year) %>%
  summarise(meanlifeExp=mean(lifeExp))

ggplot(avglifeExp, aes(x=year, y=meanlifeExp, col=continent))+
  geom_line() + geom_point()
```



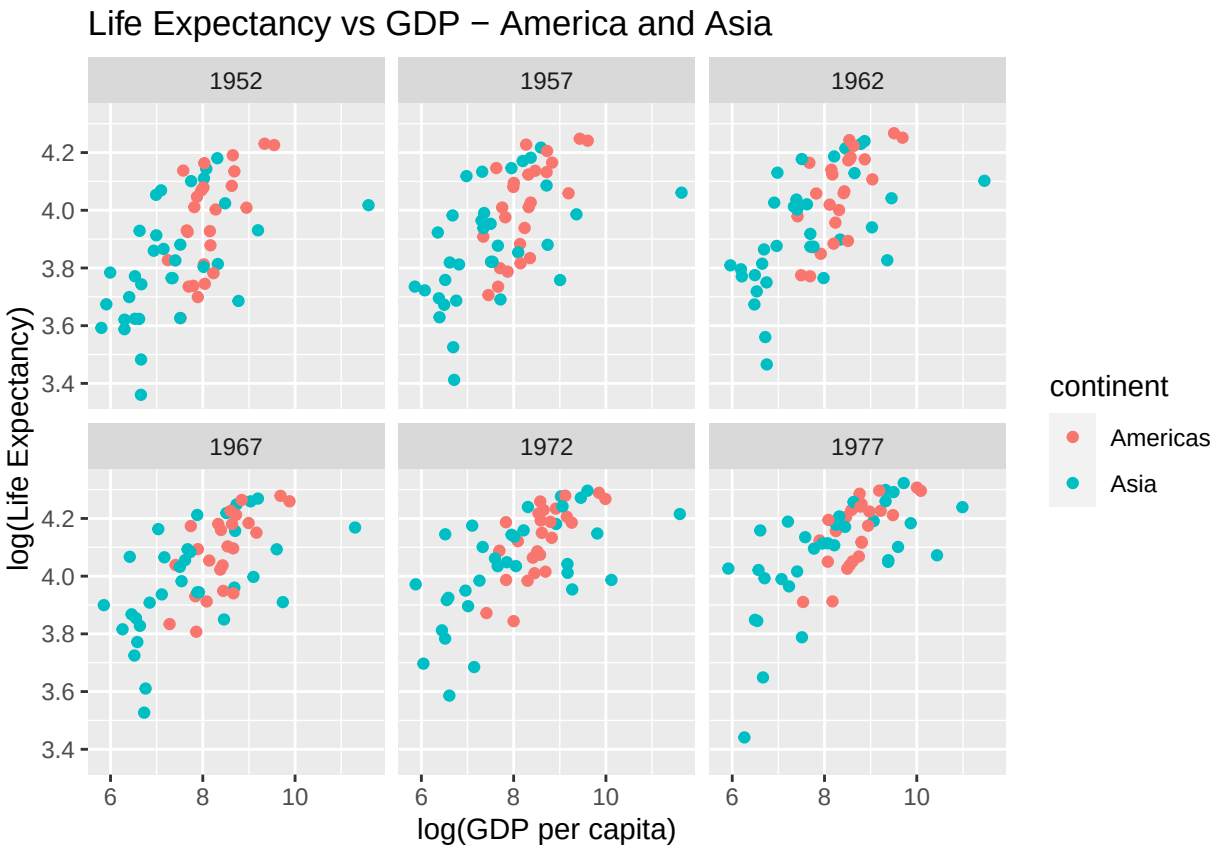
6. Slide 76

```
gapminder %>%
  filter(year %in% c(1952, 1957, 1962, 1967, 1972, 1977)) %>%
  filter(continent %in% c("Asia", "Americas")) %>%
  ggplot(aes(y=lifeExp, x=gdpPercap, color=continent)) +
  geom_point() +
  facet_wrap(~year, ncol=3)+
  labs(title="Life Expectancy vs GDP - America and Asia",
       y = "Life Expectancy",
       x = "GDP per capita")
```



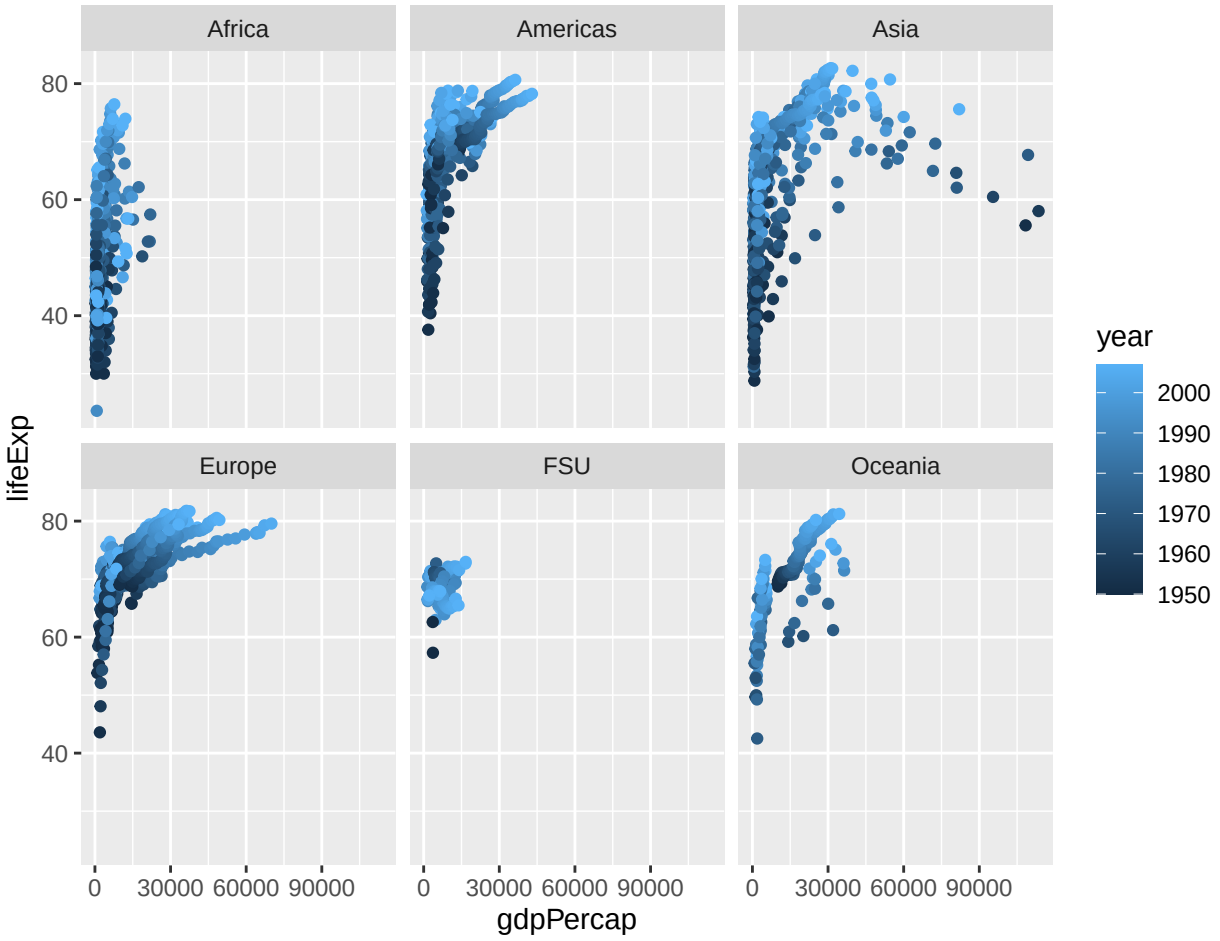
7. Slide 77

```
gapminder %>%
  filter(year %in% c(1952, 1957, 1962, 1967, 1972, 1977)) %>%
  filter(continent %in% c("Asia", "Americas")) %>%
  ggplot(aes(y=log(lifeExp), x=log(gdpPercap), color=continent)) +
  geom_point() +
  facet_wrap(~year, ncol=3)+
  labs(title="Life Expectancy vs GDP - America and Asia",
       y = "log(Life Expectancy)",
       x = "log(GDP per capita)")
```



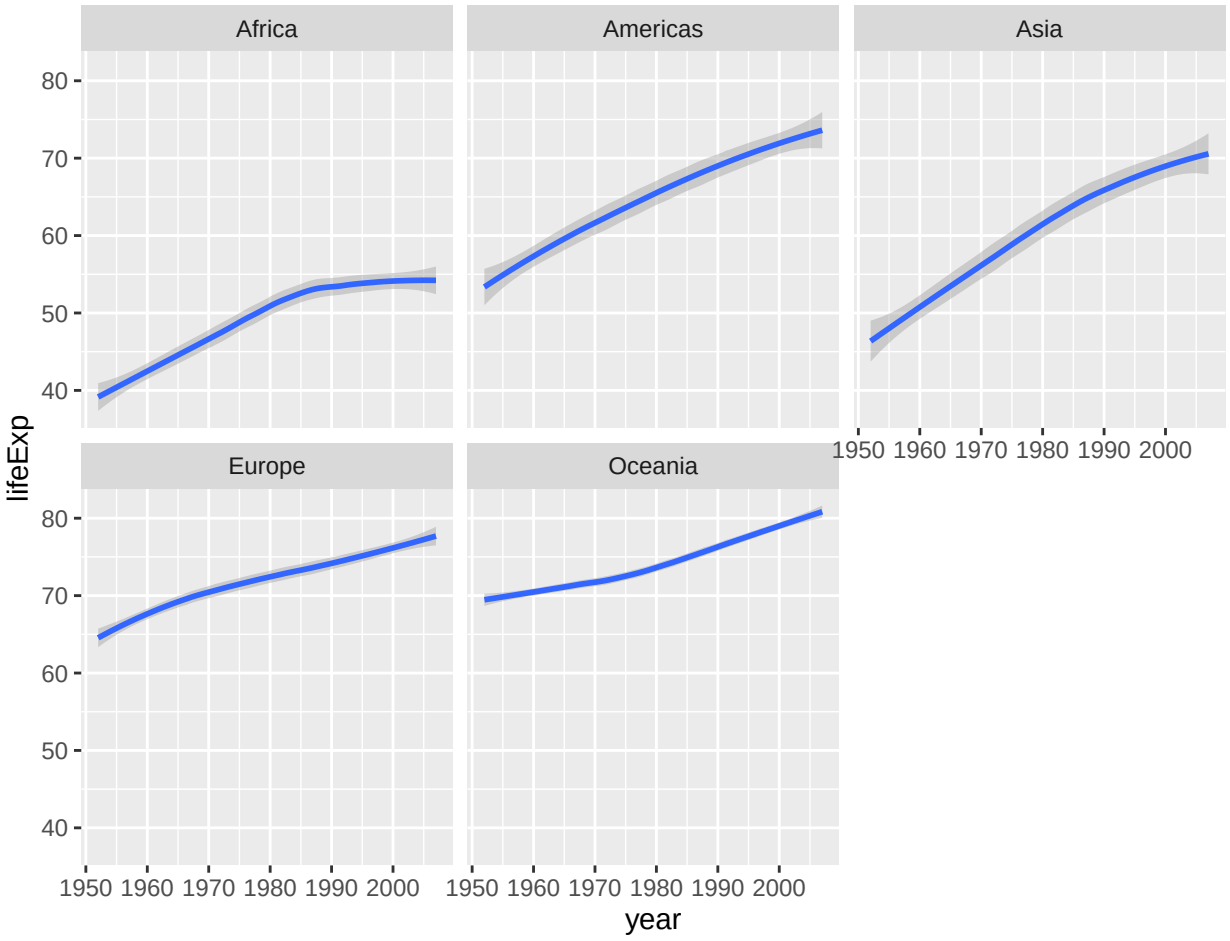
8. Slide 79

```
ggplot(gapminder_unfiltered, aes(gdpPercap, lifeExp, color = year)) +
geom_point() +
facet_wrap(~ continent)
```



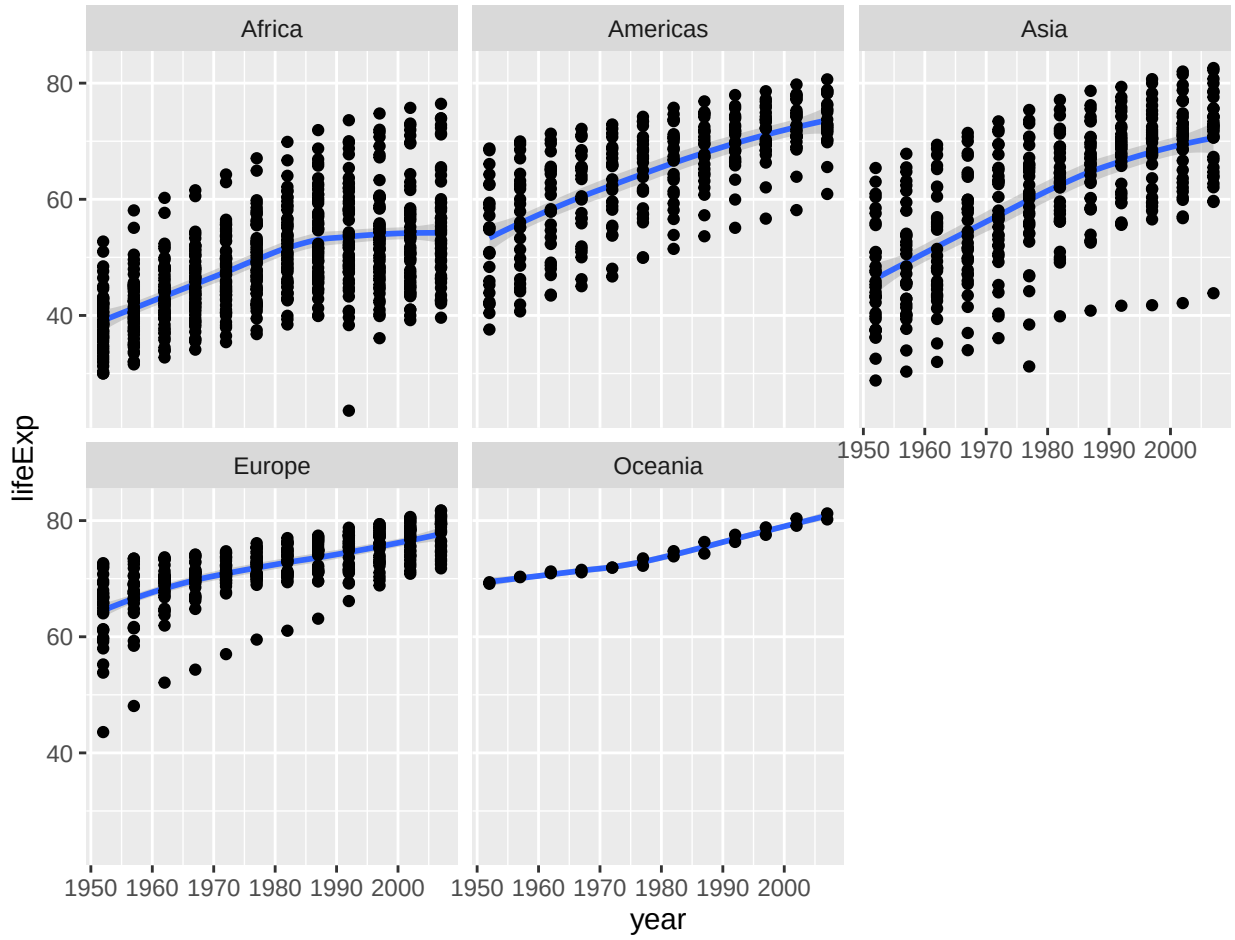
9. Slide 82

```
ggplot(gapminder, aes(y=lifeExp,  x=year)) +
geom_smooth() +
facet_wrap(~ continent)
```



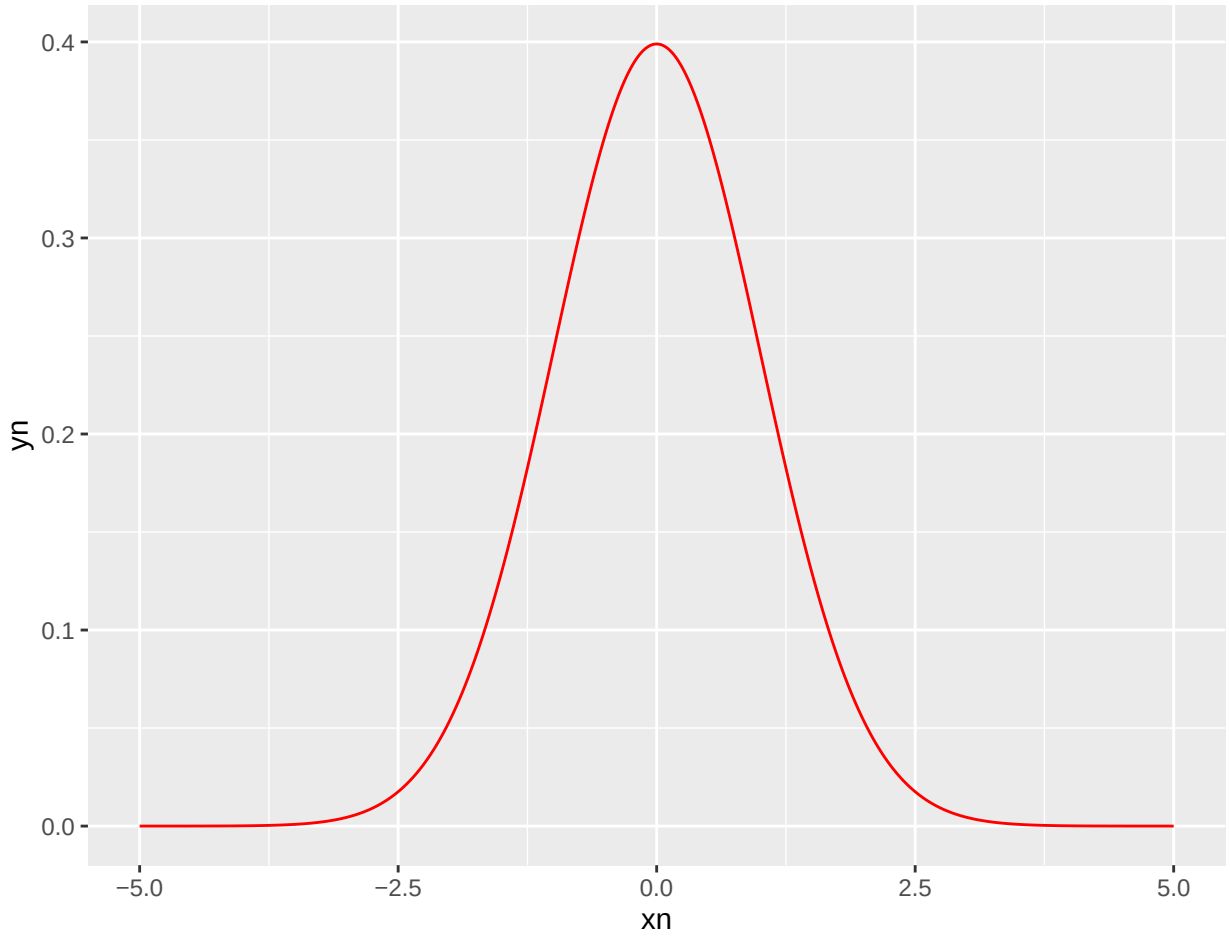
10. Slide 83

```
ggplot(gapminder, aes(y=lifeExp,  x=year)) +
geom_smooth() +
  geom_point() +
facet_wrap(~ continent)
```



11. Slide 84

```
xn <- seq(-5, 5, length=10000)
yn <- dnorm(xn)
df <- data.frame(x=xn, y=yn)
ggplot(df, aes(x=xn, y=yn))+
  geom_line(col="red")
```



Note:

You should include Figure caption for all graphs. It is important to interpret all graphs.