Data Visualisation

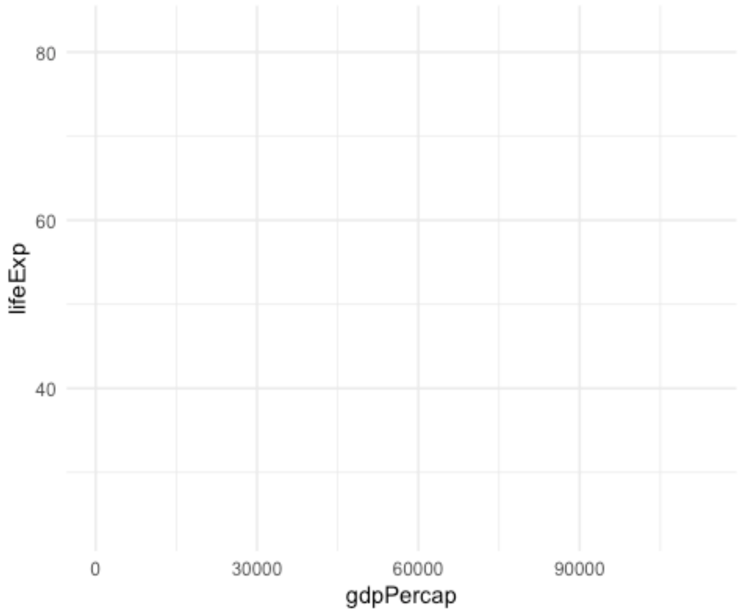
Hans

9/10/2019

```
p_load(tidyverse)
p_load(gapminder)
theme_set(theme_minimal())

glimpse(gapminder)

## Observations: 1,704
## Variables: 6
## $ country   <fct> Afghanistan, Afghanistan, Afghanistan, Afghanistan, Af…
## $ continent <fct> Asia, Asia, Asia, Asia, Asia, Asia, Asia, Asia, Asia, …
## $ year      <int> 1952, 1957, 1962, 1967, 1972, 1977, 1982, 1987, 1992, …
## $ lifeExp   <dbl> 28.801, 30.332, 31.997, 34.020, 36.088, 38.438, 39.854…
## $ pop       <int> 8425333, 9240934, 10267083, 11537966, 13079460, 148803…
## $ gdpPercap <dbl> 779.4453, 820.8530, 853.1007, 836.1971, 739.9811, 786.…
```

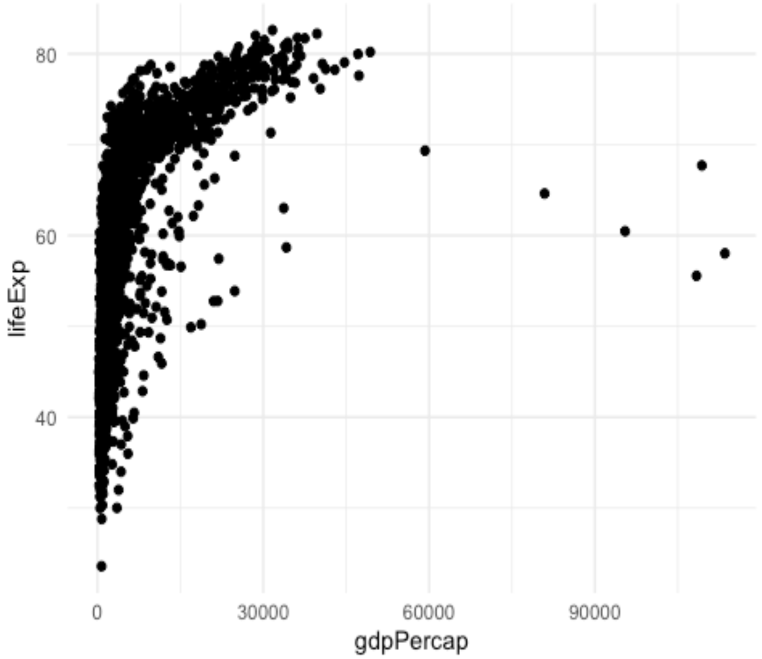
First set up the frame in which to plot

```
p <- ggplot(data=gapminder, mapping = aes(x = gdpPercap, y = lifeExp))
p
```



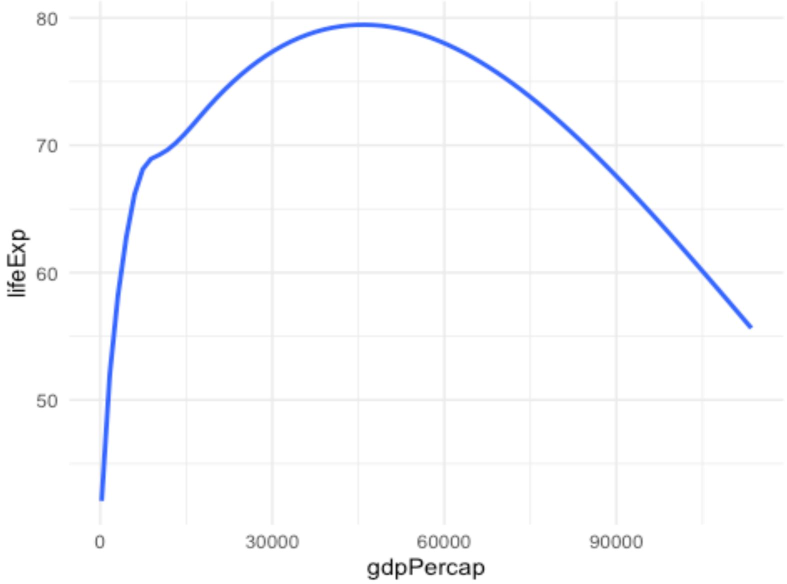
Then add the code to create a dot plot

```
p + geom_point()
```



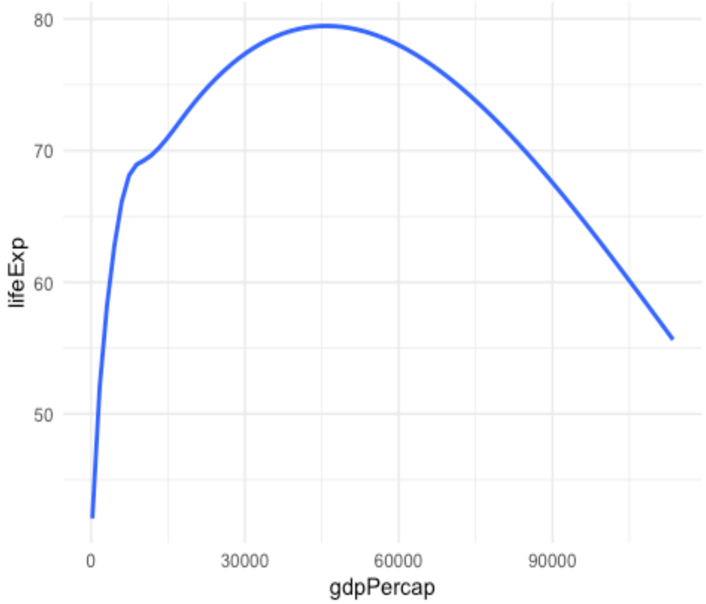
Or, rather than a dot plot, get a smooth line

```
p <- ggplot(data=gapminder, mapping = aes(x = gdpPercap, y = lifeExp))
p + geom_smooth(se = F)

## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



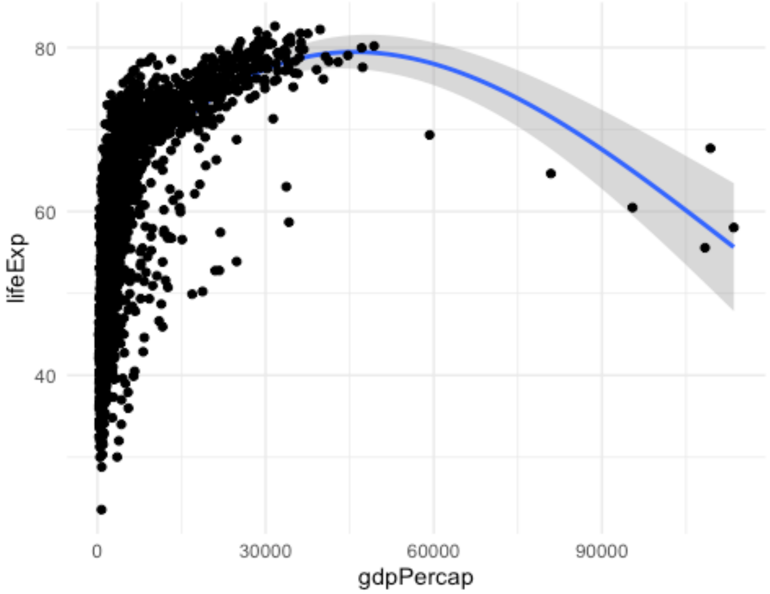
If you like, add a confidence interval around the line

```
p <- ggplot(data=gapminder, mapping = aes(x = gdpPercap, y = lifeExp))
p + geom_smooth(se = F)

## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



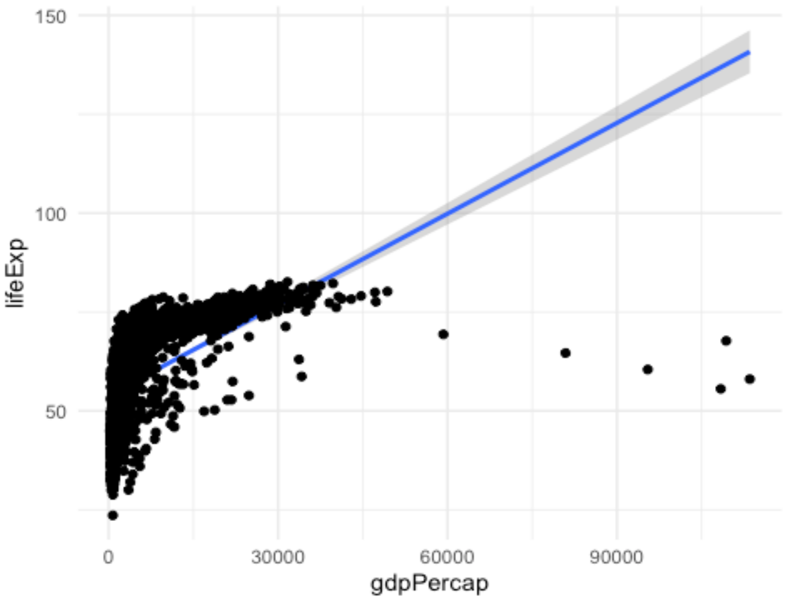
Or have both the points and the line!

```
p <- ggplot(data=gapminder, mapping = aes(x = gdpPercap, y = lifeExp))
p + geom_smooth() +
  geom_point()

## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



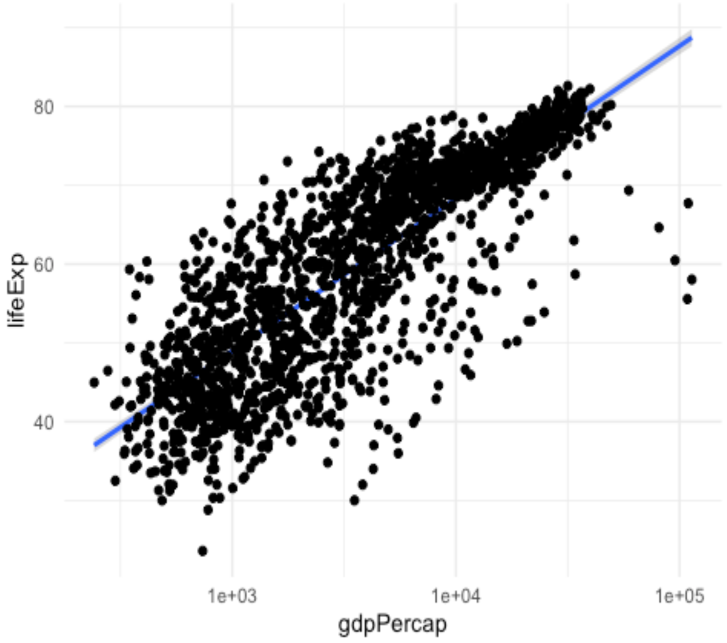
Rather than a smooth line, you can also draw a regression line

```
p <- ggplot(data=gapminder, mapping = aes(x = gdpPercap, y = lifeExp))
p + geom_smooth(method = "lm") +
  geom_point()
```



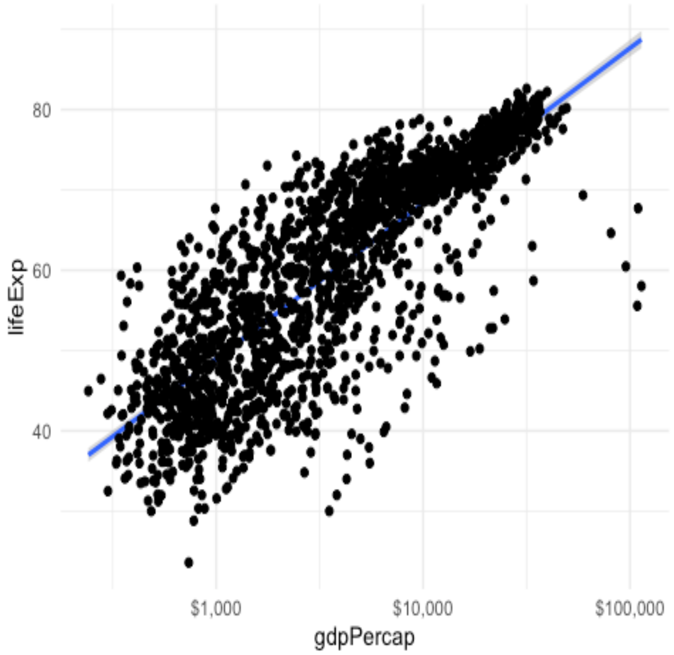
Make the x-axis logarithmic

```
p <- ggplot(data=gapminder, mapping = aes(x = gdpPercap, y = lifeExp))
p + geom_smooth(method = "lm") +
  geom_point() +
  scale_x_log10()
```



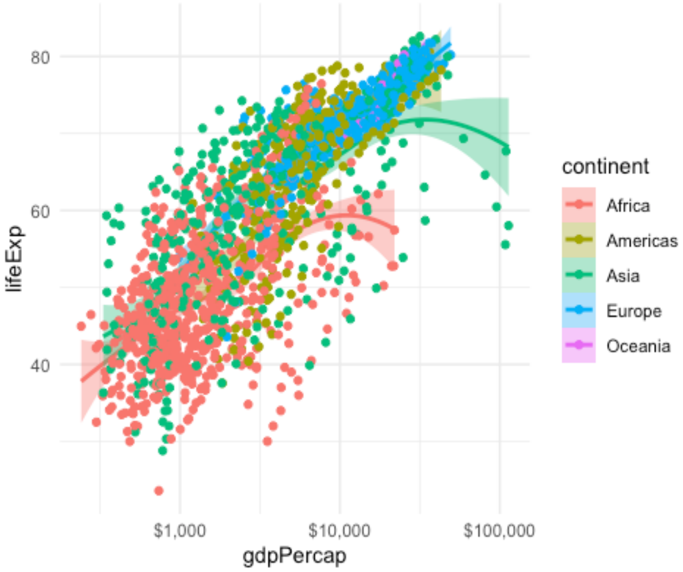
Format the numbers on the x-axis as dollars

```
p <- ggplot(data=gapminder, mapping = aes(x = gdpPercap, y = lifeExp))
p + geom_smooth(method = "lm") +
  geom_point() +
  scale_x_log10(labels=scales::dollar)
```



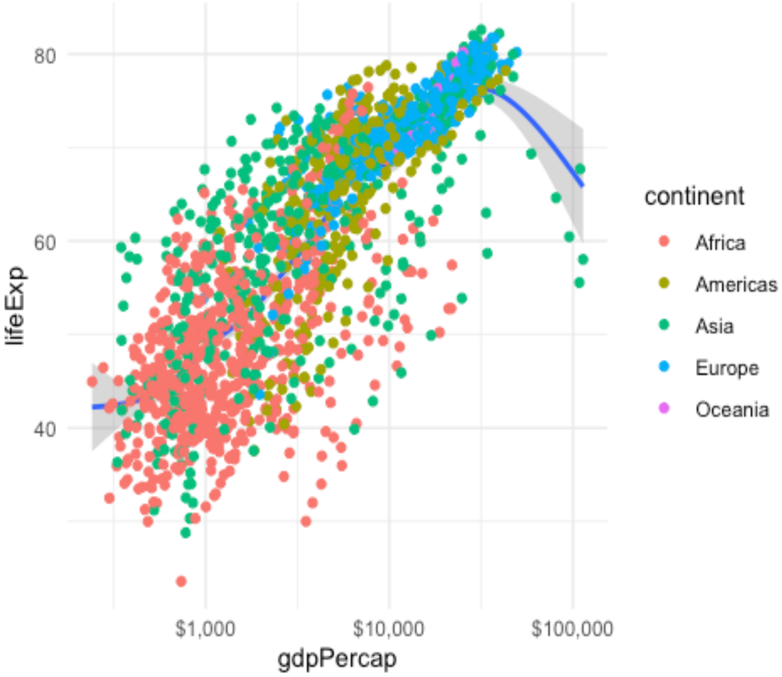
Now use color to visually separate the continents

```
p <- ggplot(data=gapminder, mapping = aes(x = gdpPercap, y = lifeExp, color =
continent, fill = continent))
p + geom_smooth() +
  geom_point() +
  scale_x_log10(labels=scales::dollar)

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



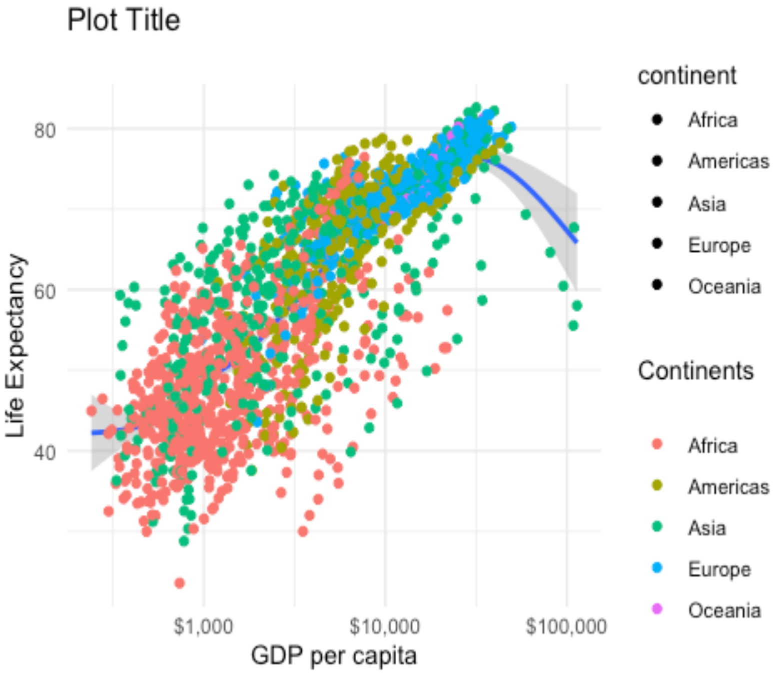
You can apply te color only to the points and not tonthe smooth line

```
p <- ggplot(data = gapminder,
            mapping = aes(x = gdpPercap, y = lifeExp))
p + geom_smooth() +
  geom_point(mapping = aes(color = continent, fill = continent)) +
  scale_x_log10(labels=scales::dollar)

## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



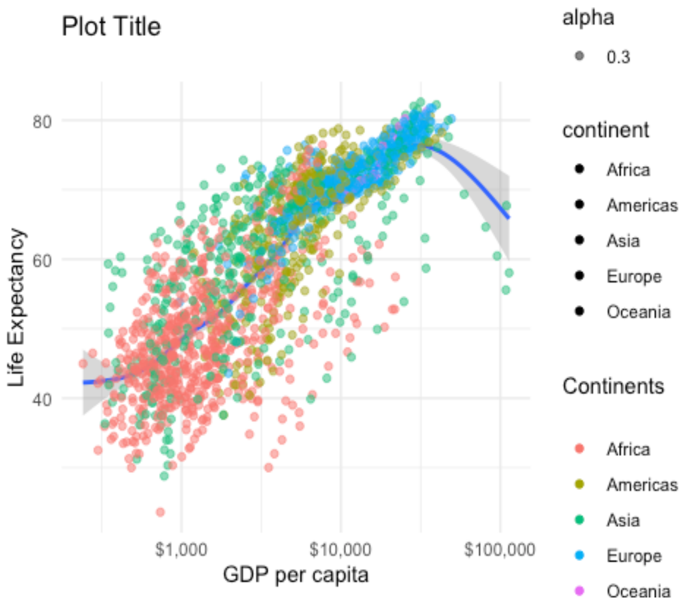
Add some decent looking labels

```
p <- ggplot(data = gapminder,
            mapping = aes(x = gdpPercap, y = lifeExp))
p + geom_smooth() +
  geom_point(mapping = aes(color = continent, fill = continent)) +
  scale_x_log10(labels=scales::dollar) +
  labs(title = "Plot Title\n", x = "GDP per capita", y = "Life Expectancy",
colour = "Continents\n")

## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



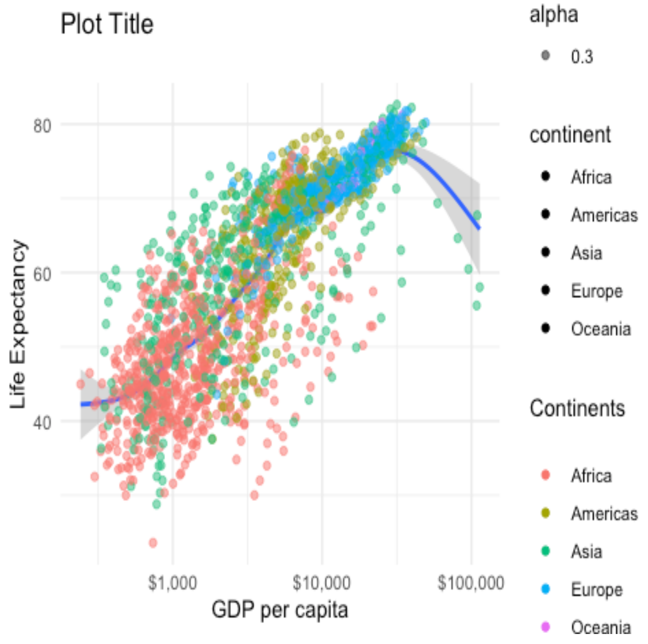
What if you want to make the points less dominant? Use Alpha

```
p <- ggplot(data = gapminder,
            mapping = aes(x = gdpPercap, y = lifeExp))
p + geom_smooth() +
  geom_point(mapping = aes(color = continent, fill = continent, alpha = 0.3))
+
  scale_x_log10(labels=scales::dollar) +
  labs(title = "Plot Title\n", x = "GDP per capita", y = "Life Expectancy",
colour = "Continents\n")

## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



It can be simplified; you do not need to ie the variable names 'data' and 'mapping'

```
p <- ggplot(gapminder,
            aes(x = gdpPercap, y = lifeExp))
p + geom_smooth() +
  geom_point(aes(color = continent, fill = continent, alpha = 0.3)) +
  scale_x_log10(labels=scales::dollar) +
  labs(title = "Plot Title\n", x = "GDP per capita", y = "Life Expectancy",
colour = "Continents\n")

## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```


Plot Title
alpha
continent
Africa
Americas
Asia
Europe
Oceania
Continents
Africa
Americas
Asia
Europe
Oceania
Life Expectancy
$1,000
$10,000
$100,000
GDP per capita