Pick four - comparing trends in population over time

Purpose

The purpose of this report is to compare the population trends for four countries of your choosing.

Required Libraries

```
library(ggplot2)
```

Data

- Gapminder data available here. Gapminder data is licensed CC-BY 3.0.
- Processed data via [@jennybc](https://github.com/jennybc), R package available here. The data-raw sub-directory reveals the journey from Gapminder.org's Excel workbooks to increasingly clean and tidy data.

Read in data: To read in the data, make sure this file is in the same directory/folder as the `gapminderDataFiveYear.txv` file. To set the proper working directory go to Session > Set Working Directory > To Source File Location.

```
gapMinder <- read.delim("../../gapminderDataFiveYear.tsv")

##Check data
head(gapMinder) #First 10 lines of dataset
```

```
##       country year      pop continent lifeExp gdpPercap
## 1 Afghanistan 1952  8425333      Asia   28.80     779.4
## 2 Afghanistan 1957  9240934      Asia   30.33     820.9
## 3 Afghanistan 1962 10267083      Asia   32.00     853.1
## 4 Afghanistan 1967 11537966      Asia   34.02     836.2
## 5 Afghanistan 1972 13079460      Asia   36.09     740.0
## 6 Afghanistan 1977 14880372      Asia   38.44     786.1
```

```
dim(gapMinder) #number of rows and columns in data set
```

```
## [1] 1704    6
```

You can see what countries are available by looking at the how many unique categories are in the country column of the gapMinder dataset.

```
levels(gapMinder$country)
```

Pick Four Countries

Now pick four countries that you are intrested in. Just replace with the countries name below.

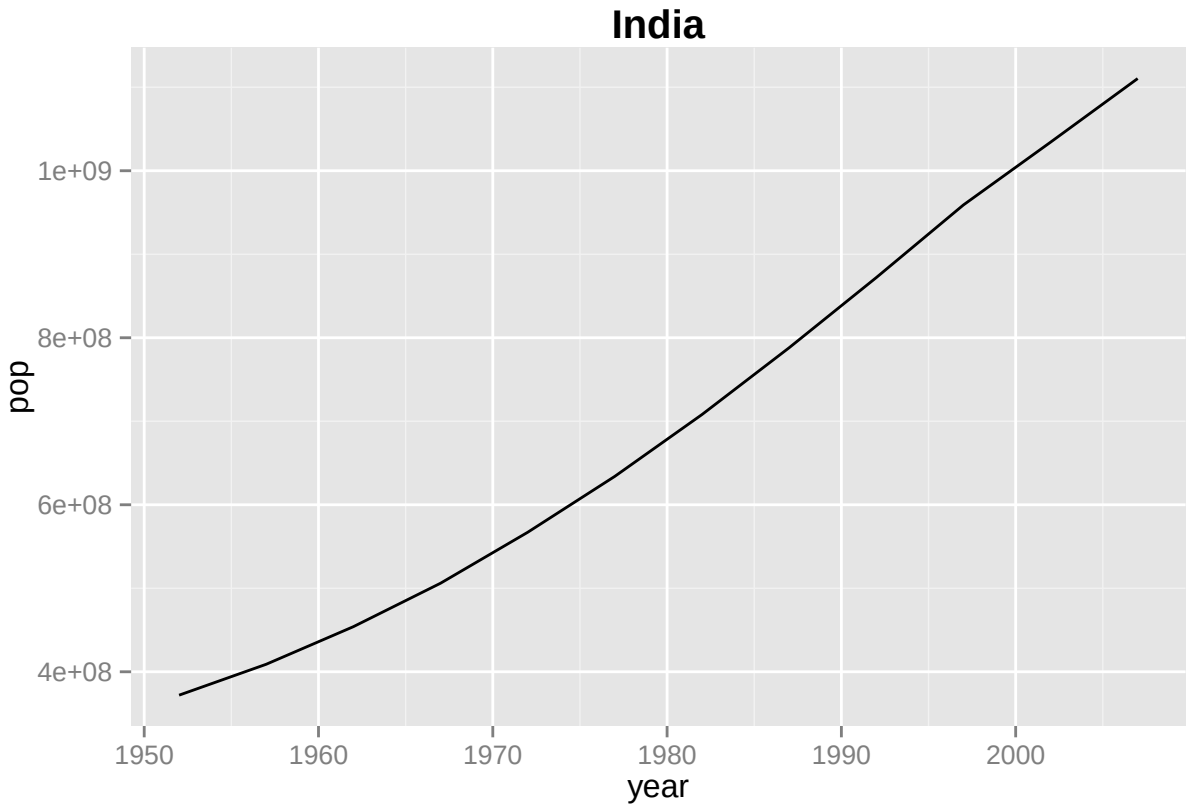
```
countryName1 <- "India"
countryName2 <- "United States"
countryName3 <- "Nigeria"
countryName4 <- "Germany"
```

Country One

We want to look at how population changes over time for the first country.

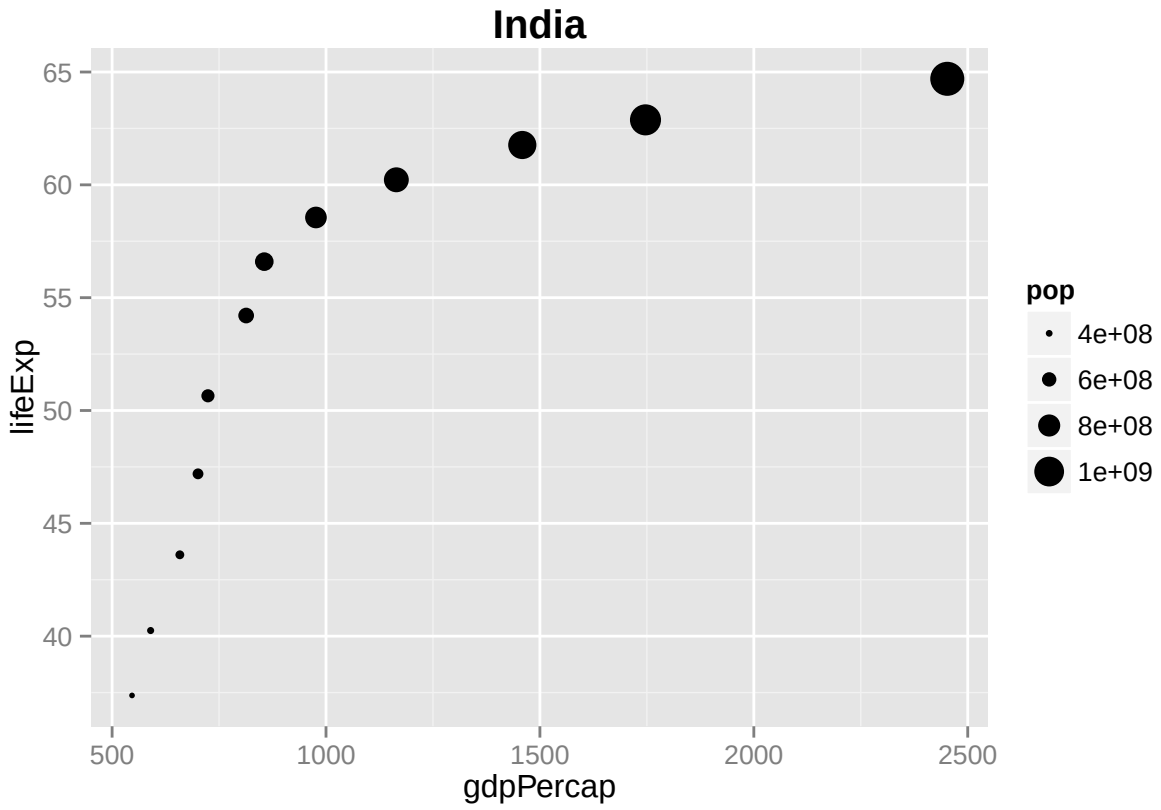
```
country1 <- subset(gapMinder, country == countryName1)

ggplot(country1, aes(year, pop, label = country)) +
  geom_path() +
  ggtitle(countryName1) +
  theme(plot.title = element_text(size = 15, face = "bold"))
```



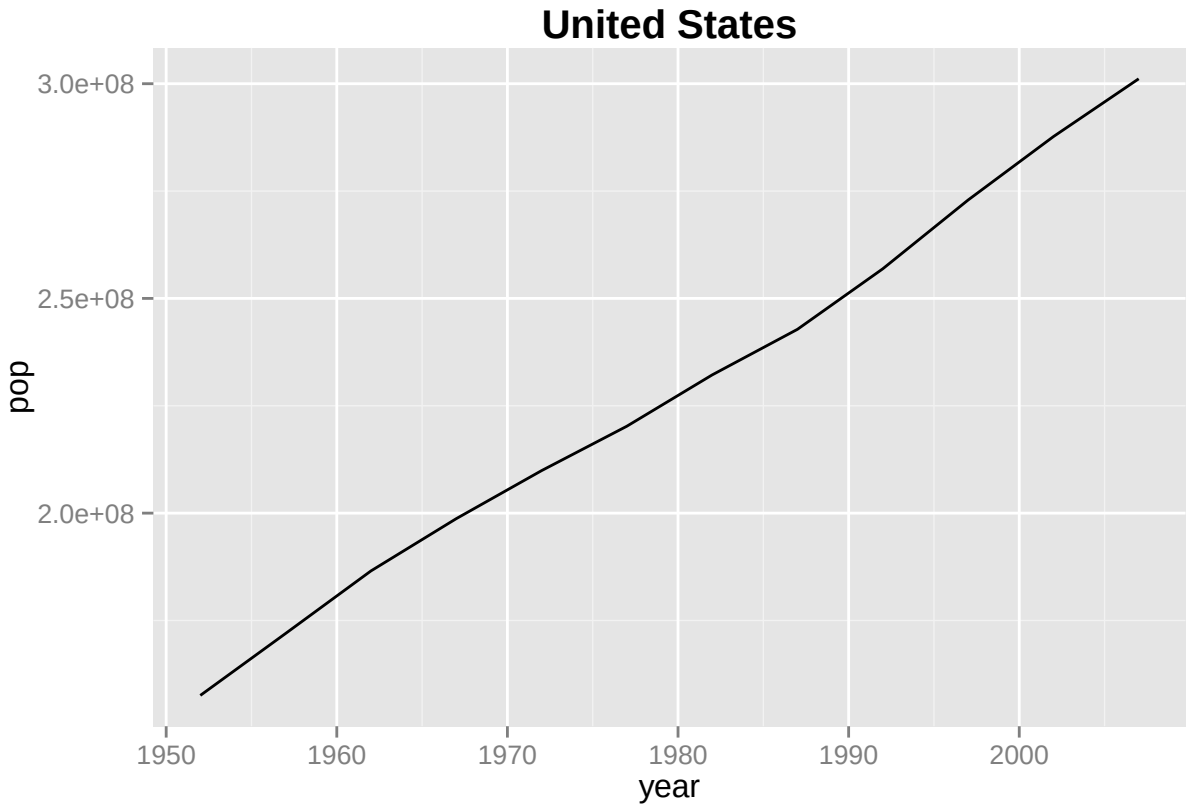
This second graph is looking at the correlation between life expectency (lifeExp) and income per person (gdpPercap). The size of the points on the plot represent.

```
ggplot(country1, aes(gdpPercap, lifeExp, label = country, size = pop)) +
  geom_point() +
  ggtitle(countryName1) +
  theme(plot.title = element_text(size = 15, face = "bold"))
```

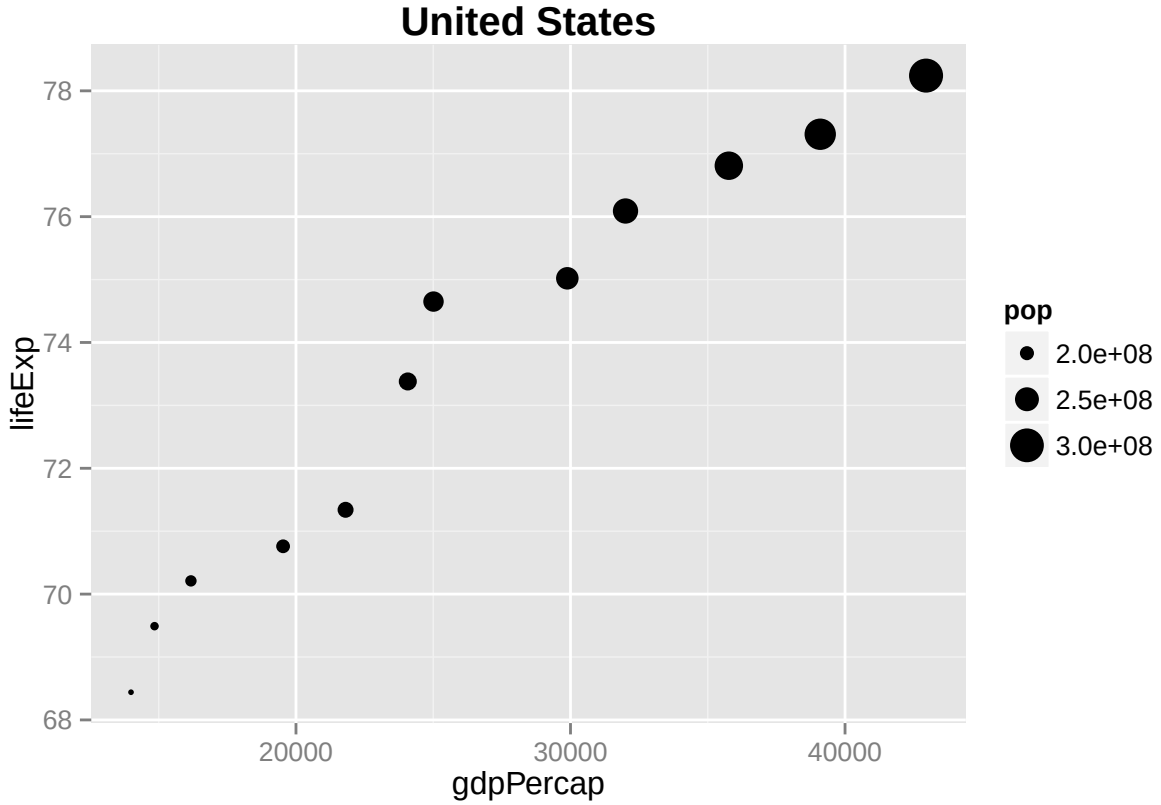


Country 2

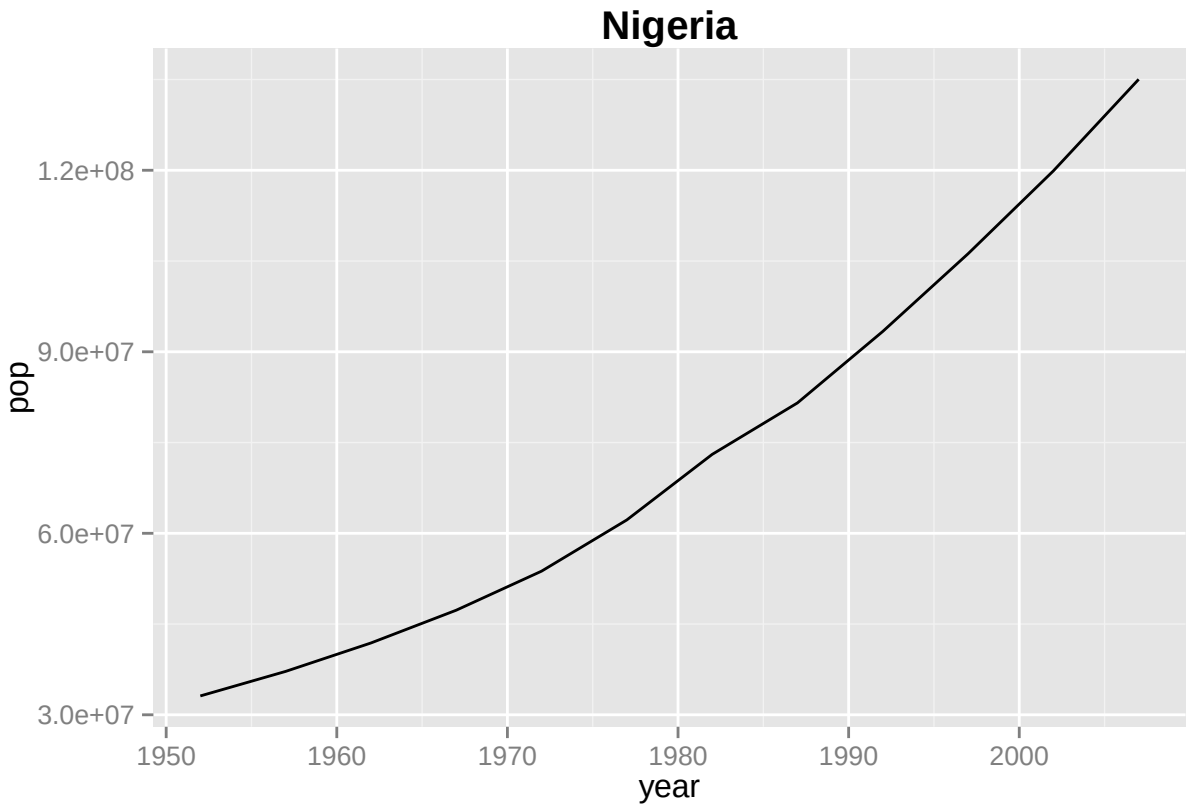
We will do this for each country. Since the code is very similar, we will omit viewing it below.



Notes: In a real report you can add in information about the results of the analysis you are performing. That way your code, analysis, questions, and results are all in one place.

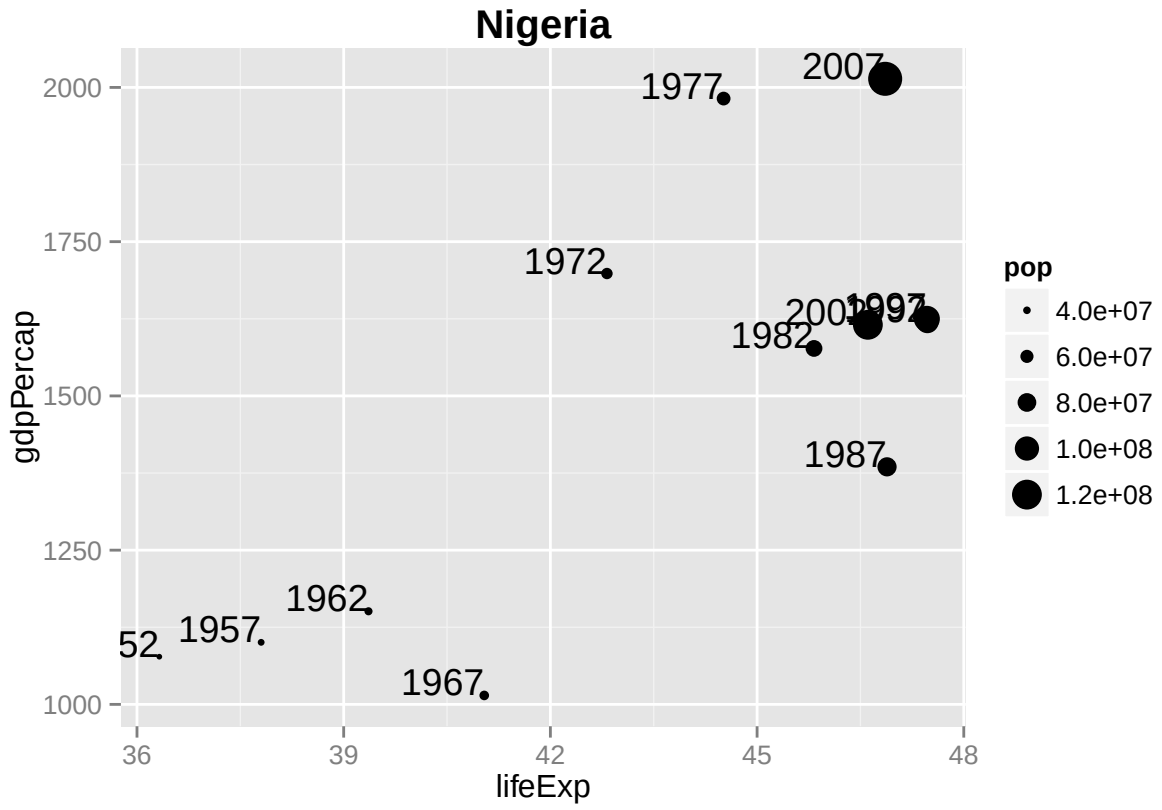


Country 3

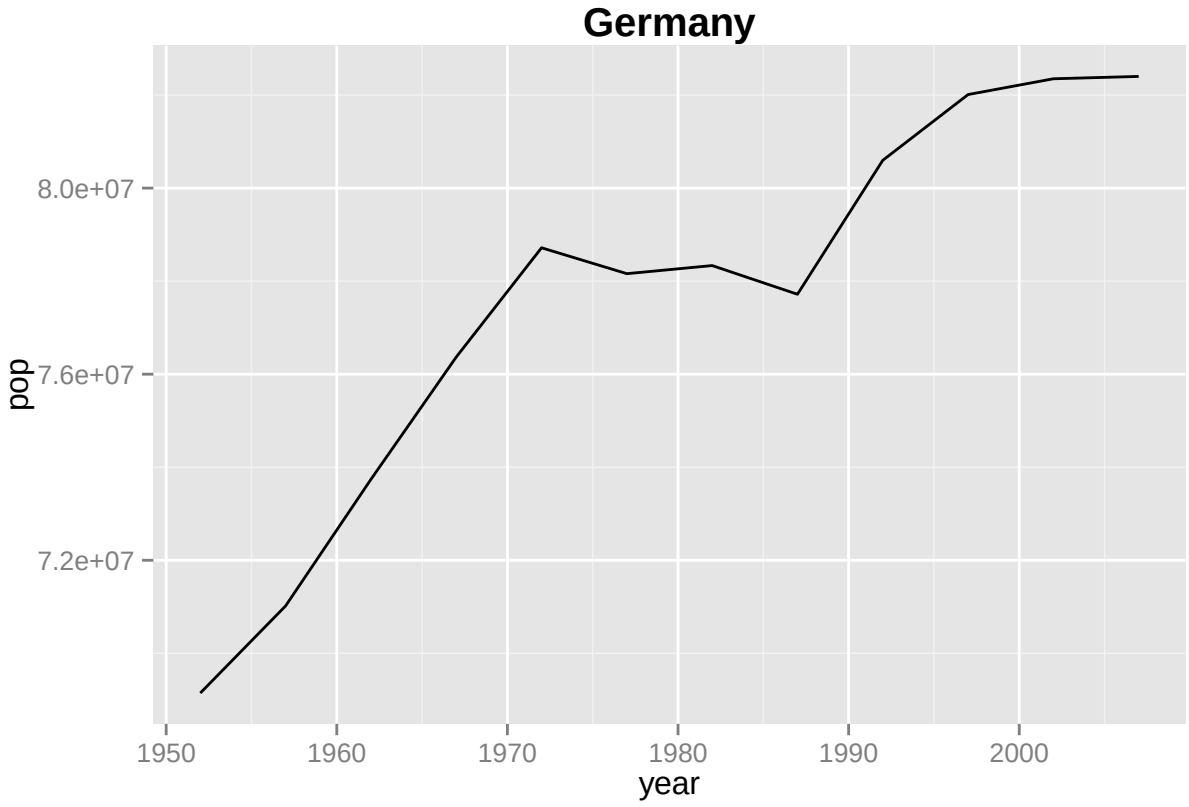


Notes Maybe a country has an unusual distribution and we want to label the graph with the year. So we add a bit of code and label with year.

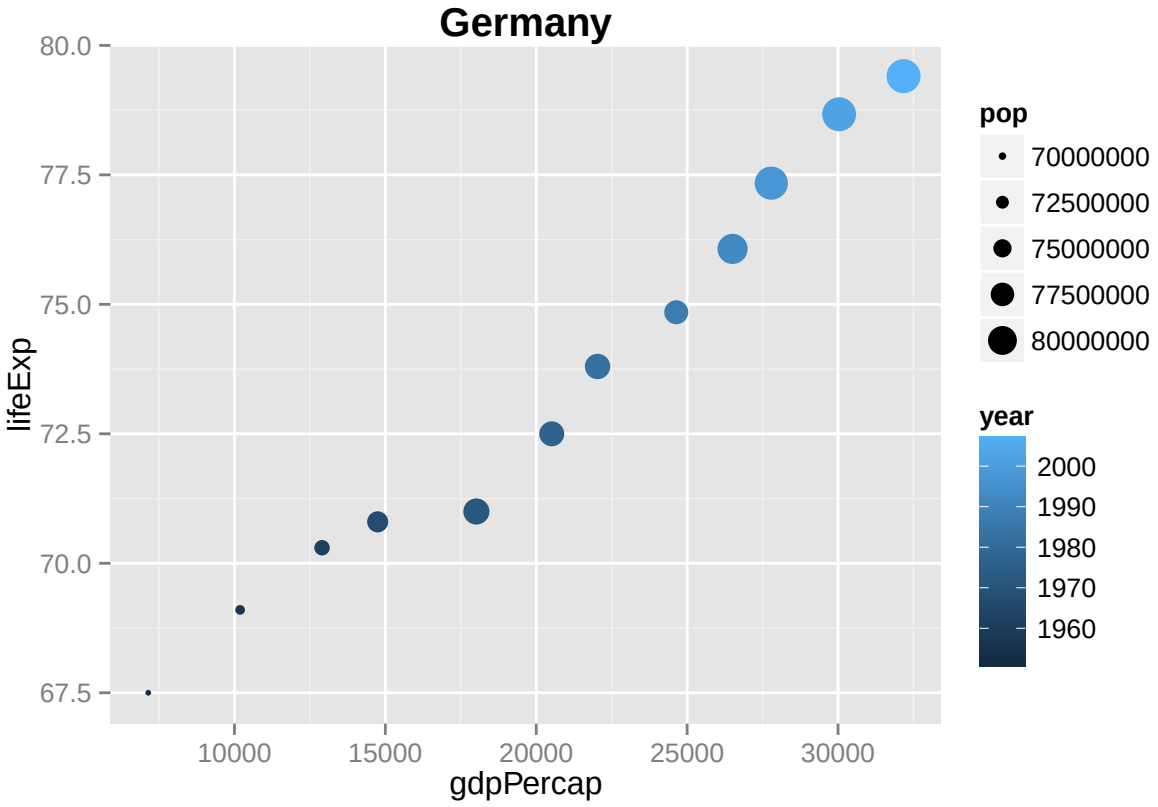
```
ggplot(country3, aes(lifeExp, gdpPercap, size = pop, label = year)) +
  geom_point() +
  geom_text(hjust = 1, vjust = 0, size = 5) +
  ggtitle(countryName3) +
  theme(plot.title = element_text(size = 15, face = "bold"))
```


Nigeria
2000
1750
1500
1250
1000
gdpPercap
36
39
42
45
48
lifeExp
2007
1977
1972
1982
1987
1962
1957
52
1967
pop
4.0e+07
6.0e+07
8.0e+07
1.0e+08
1.2e+08

Country 4


Germany
8.0e+07
7.6e+07
7.2e+07
pop
1950
1960
1970
1980
1990
2000
year

Notes: Or maybe try out labeling the year by adding color.

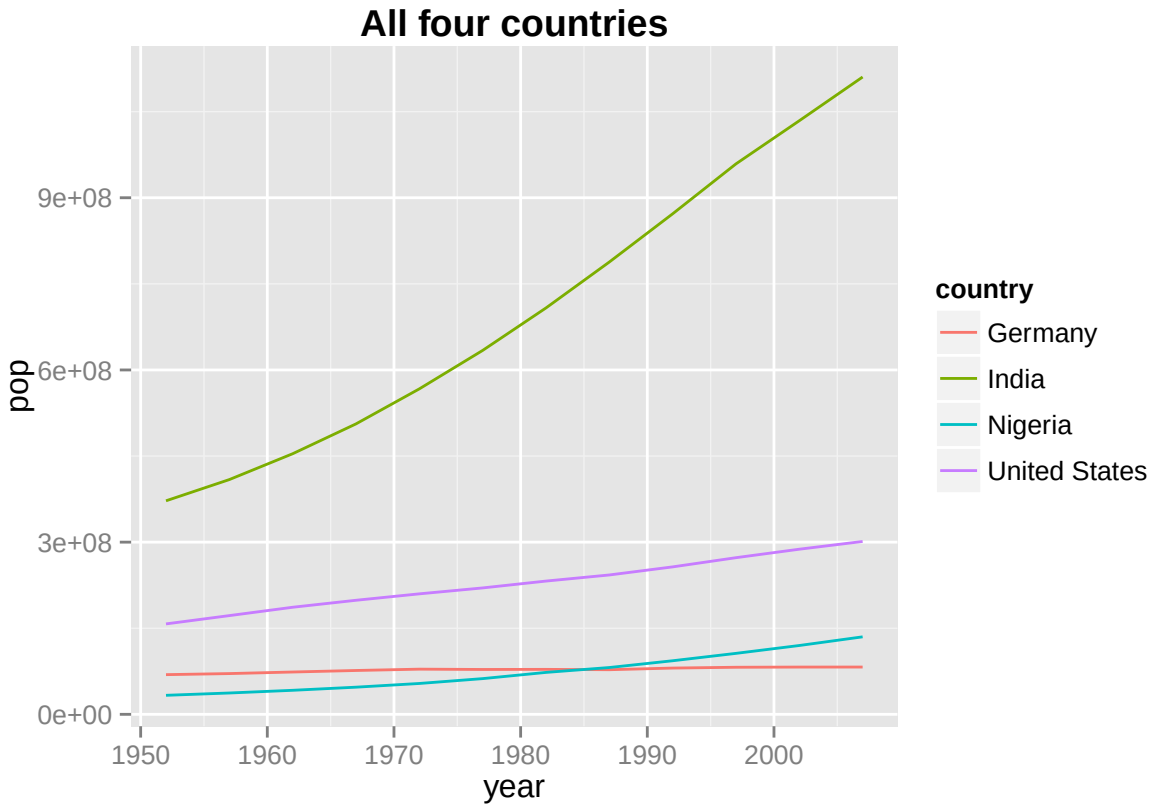


All four countries

Let's add all four countries together and to see how they compare.

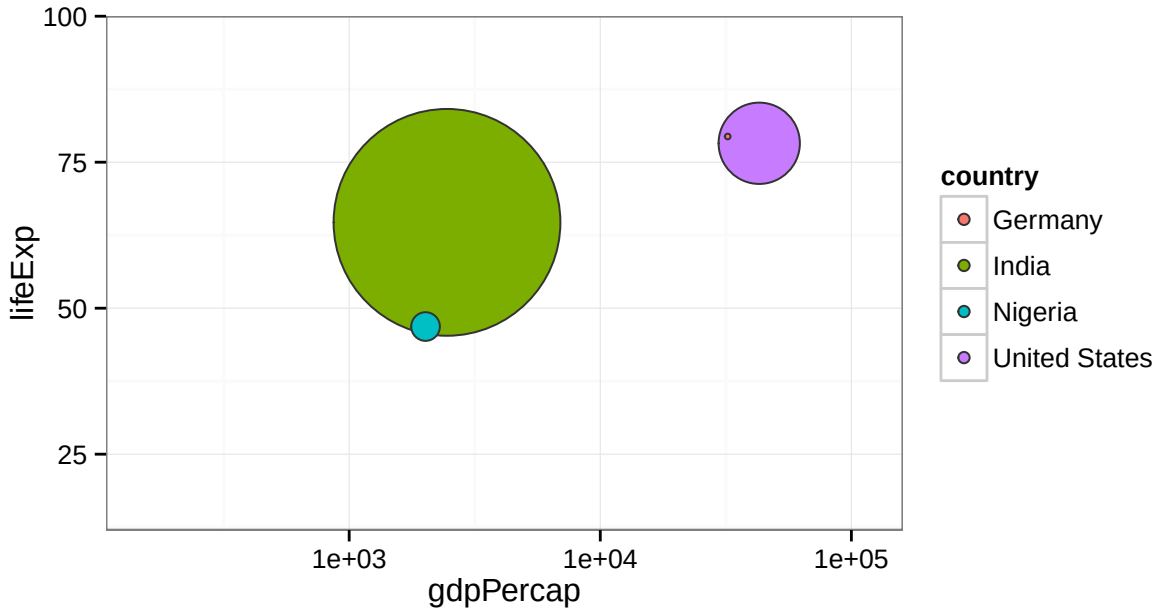
```
#Add subsetted data together
allCountries <- rbind(country1, country2, country3, country4)

#Notice the code for this is similar to when
#we are just looking at one country just
#with the added the color option
ggplot(allCountries, aes(year, pop, color=country)) +
  geom_path() +
  ggtitle("All four countries") +
  theme(plot.title = element_text(lineheight=.8, face = "bold"))
```



What about what is occuring in a particular year? You can change the year by changing the code in the `year == 2007` section. To look at what years are possible use `allCountries$year`.

```
ggplot(subset(allCountries, year == 2007),
       aes(x = gdpPercap, y = lifeExp, fill = country, size=sqrt(pop/pi))) +
       scale_x_log10(limits = c(150, 115000)) +
        ylim(c(16, 96)) +
  geom_point(pch = 21, color = 'grey20') +
  scale_size_continuous(range=c(1,40)) +
  coord_fixed(ratio = 1/43) +
  theme_bw() +
  guides(size=FALSE) +
  theme(strip.text = element_text(size = rel(1.1)))
```



You can plot all the years at once also!

```
ggplot(allCountries,
       aes(x = gdpPercap, y = lifeExp, fill = country, size = sqrt(pop/pi))) +
       scale_x_log10(limits = c(150, 115000)) +
        ylim(c(16, 96)) +
  geom_point(pch = 21, color =   'grey20', alpha = 0.8) +
  scale_size_continuous(range = c(1,40)) +
  coord_fixed(ratio = 1/43) +
  theme_bw() +
  guides(size = FALSE) +
  theme(strip.text = element_text(size = rel(1.1)))
```