MSCS 264: Homework #2

Due Friday, Feb 17 at 11pm

Al Ashir Intisar

This homework covers more of Ch3: Data Visualization.

Refer back to in-class Rmd ch3_2_facet_geoms_labs.Rmd (and possibly ch3_1_ggplot_geom_point_aesthetics.Rmd)

Change the author above to your name. Save a copy of this Rmd to your submit folder.

The code below will add the data from the gapminder package. Notice that the second line creates our dataset, gapminder07 which includes data from only the year 2007! (preview of Ch 5!)

```
library(tidyverse)
```

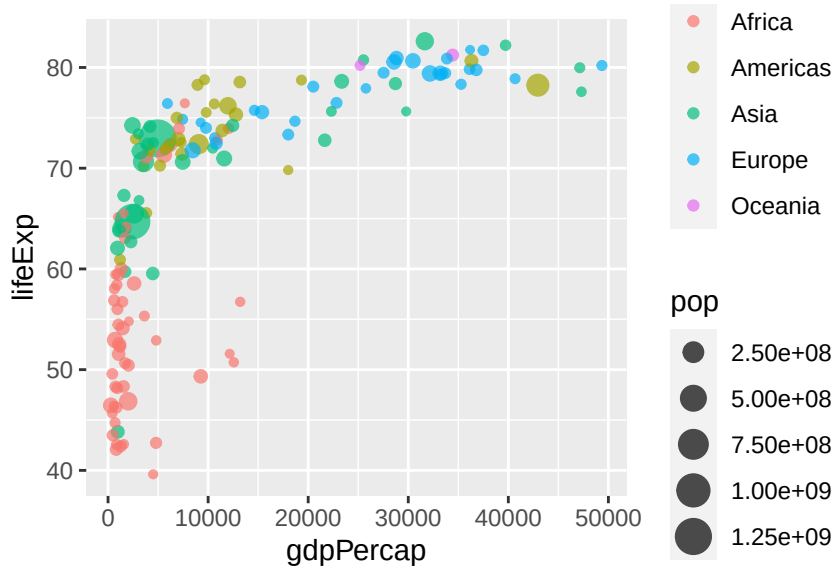
```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.7     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(gapminder)
gapminder07 <- filter(gapminder, year == 2007)
```

Type ?gapminder in your console to see the definitions of the variables.

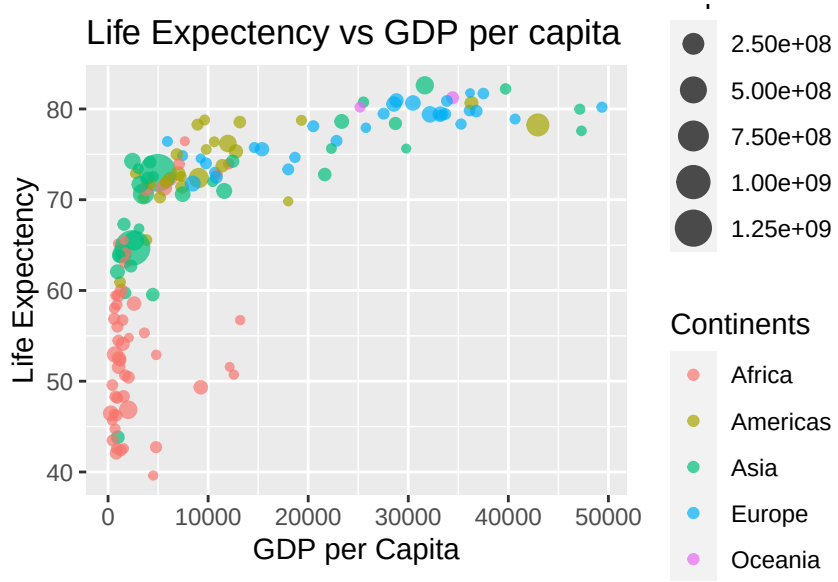
1. In the Homework > images folder, open "hw2_gapminder_plot". Use the gapminder07 dataset to make this plot. (It's a lot, I know! Start by finding example code for a scatterplot; then build up: start with the x and y axis, then add more aesthetics.)

```
ggplot(gapminder07, aes(x = gdpPercap, y = lifeExp, color = continent)) +
  geom_point(aes(size = pop), alpha = .7)
```



2. Use `labs` to modify the labels on the x and y axis, the size and color legends, and add a title to the graph you made in #1.

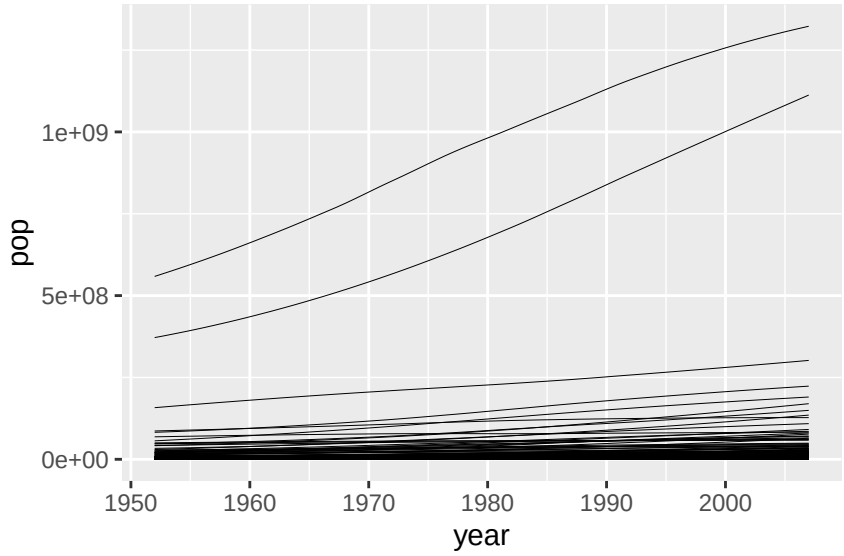
```
ggplot(gapminder07, aes(x = gdpPercap, y = lifeExp, color = continent)) +
  geom_point(aes(size = pop), alpha = .7)+
  labs(title = "Life Expectency vs GDP per capita", color = "Continents", size = "Population")+
  xlab("GDP per Capita")+
  ylab("Life Expectency")
```



3. Modify the code in the graph below so that the lines are thinner, are colored black, and there is no gray uncertainty band behind each line.

```
ggplot(gapminder, aes(x = year, y = pop)) +
  geom_smooth(aes(group = country), color = "black", se = FALSE, size = 0.1)
```

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



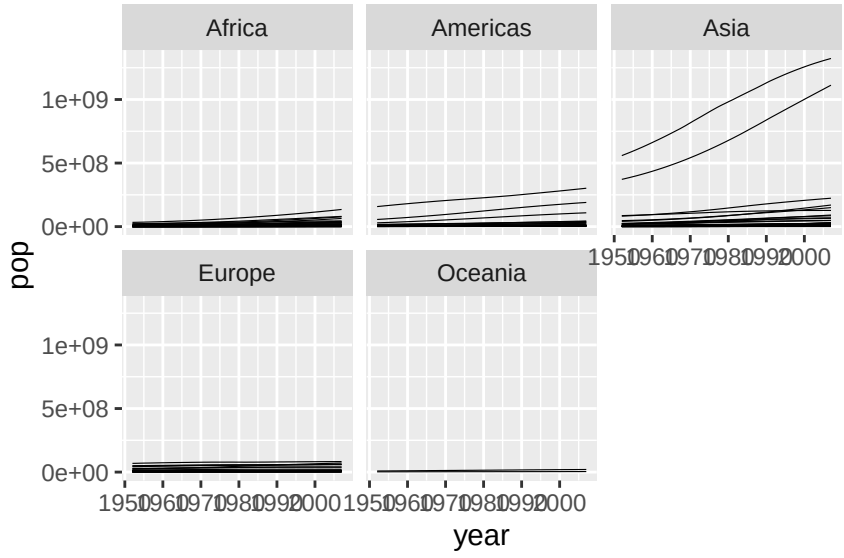
4. Why might we prefer group = country in the code above, rather than color = country?

Ans: In geom smooth for this data if we use color = country then we will only have one line trying to represent a trend in the all country data. But when we use group = country we get one smooth line for each country as a group.

5. Modify your plot above so that each continent is plotted separately.

```
ggplot(gapminder, aes(x = year, y = pop)) +
  geom_smooth(aes(group = country), color = "black", se = FALSE, size = 0.2)+
  facet_wrap(~ continent, nrow = 2)
```

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



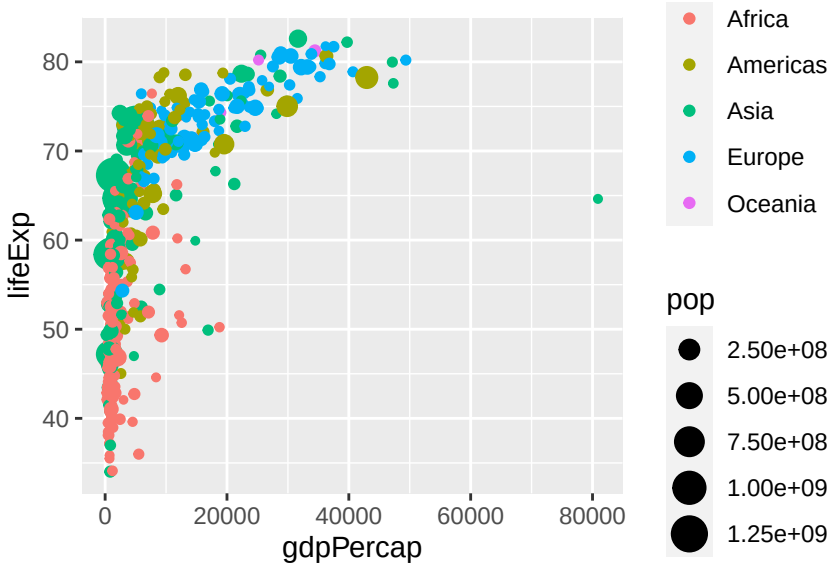
6. Write 2 sentences about what you learn about how population has changed over time from your graph in #5.

Ans: We can see a significant increase in the population in Asia and Americas and a little rise in Africa. But the population is close to stationary in Europe and Oceania.

6. Using the gapminder2 dataset, create a scatterplot that has gdpPercap as the explanatory variable and lifeExp as the response variable.

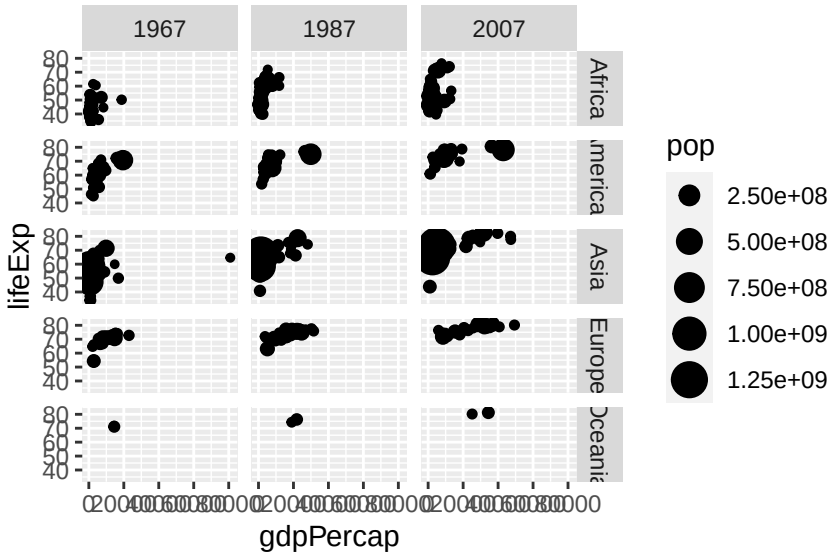
```
gapminder2 <- filter(gapminder, year %in% c(1967, 1987, 2007))
```

```
ggplot(gapminder2, aes(x = gdpPercap, y = lifeExp)) +
  geom_point(aes(size = pop, color = continent))
```



7. Modify your code above to create a grid of scatterplots. Each row of plots should be a continent, and column should indicate year. (note that we have filtered the data to include only the years 1967, 1987 and 2007)

```
ggplot(gapminder2, aes(x = gdpPercap, y = lifeExp)) +
  geom_point(aes(size = pop)) +
  facet_grid(continent ~ year) +
  labs(color = FALSE)
```



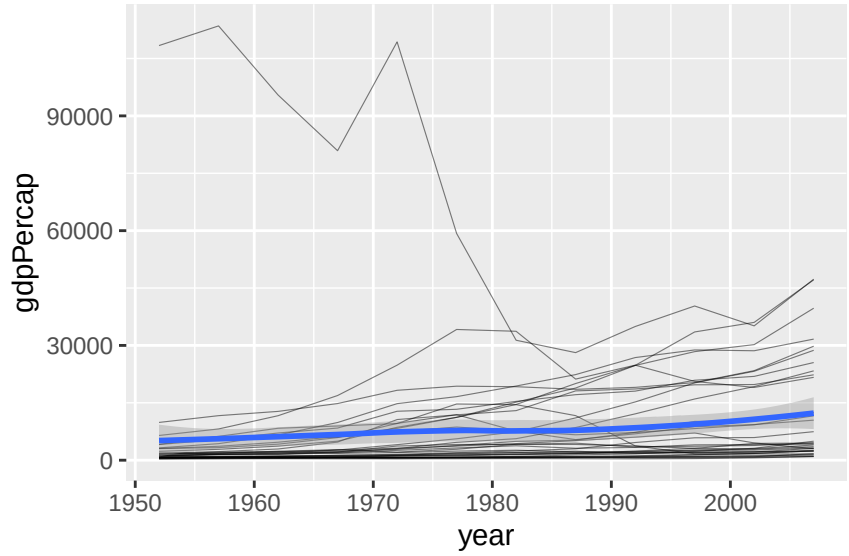
8. Modify the code below so that we have one gray line per country and a single blue smooth overall. See

Homework > images > hw2_gdp_asia.png

```
gapminder_asia <- filter(gapminder, continent == "Asia")

ggplot(gapminder_asia, aes(x = year, y = gdpPercap)) +
  geom_line(aes(group = country), size = 0.2, alpha = 0.5) +
  geom_smooth(aes(x = year, y = gdpPercap))
```

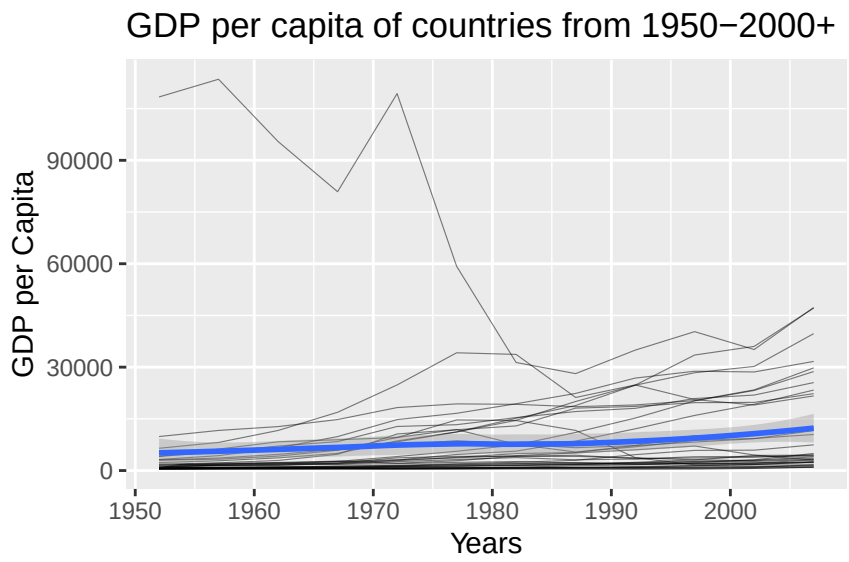
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



9. Add meaningful axis labels and a title to your plot above.

```
ggplot(gapminder_asia, aes(x = year, y = gdpPercap)) +
  geom_line(aes(group = country), size = 0.2, alpha = 0.5) +
  geom_smooth(aes(x = year, y = gdpPercap))+
labs(title = "GDP per capita of countries from 1950-2000+")+
  xlab("Years")+
  ylab("GDP per Capita")
```

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

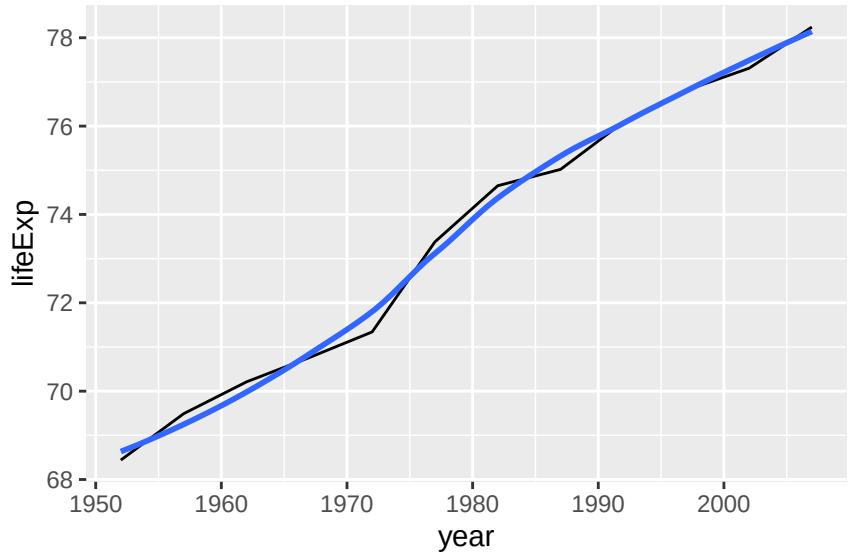


10. How do you think geom_line differs from geom_smooth? Use the example below to help you:

```
gapminder_usa <- filter(gapminder, country == "United States")

ggplot(gapminder_usa, aes(x = year, y = lifeExp)) +
  #geom_point() +
  geom_line() +
  geom_smooth(se = FALSE)
```

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



Ans: Geom_smooth tries to find a trend and makes the line more smooth whereas geom_line just connects the data points and creates lines.