DataViz Midterm I

Your Name Here

General instructions for Midterms:

- Create a new Markdown file
- Change the heading to include your author name
- Save the R Markdown file (named as: [MikeID]-[MidtermI].Rmd – e.g. "mlopez-MidtermI") to somewhere where you'll be able to access it later (zip drive, My Documents, Dropbox, etc)
- Your file should contain the code/commands to answer each question in its own code block, which will also produce plots that will be automatically embedded in the output file
- Each answer must be supported by written statements (unless otherwise specified) as well as any code used
- A printed HTML or PDF copy of this midterm is due in class on Friday at 8:40 AM (October 7). I will not answer any questions on this exam after 3:00 PM on Thursday.
- Each student must abide by Skidmore's honor code. You may use the internet for coding tips only: no solicitating answers or communicating with other students. However, note that class notes and labs are sufficient for finishing this exam.

Part 0 (10 points)

The first 10 points are awarded based on the format, spelling, grammar, and presentation of your HTML/PDF document. This includes eliminating unneeded code and writing (including the preamble above), warnings & messages, and any other text or output that is not part of your answer. Refer to the RMarkdown cheatsheet (given in the first lab) for formatting tips.

Note: You may include the questions themselves in your response, but you don't need to.

Part 1 Understanding graphics (20 points)

This question corresponds to an article and graph provided on the New York Times website (link). The article posits that searches on how to register to vote spiked in markets with a large Hispanic population, and argues that the potential association may be driven by Donald Trump's comments of former Miss Universe Alicia Machado. The graph in question is titled "The Alicia Machado Effect?"

1. Imagine that you got your hands on this particular data set. Using the points provided as a reference, provide what two rows of the data would look like.
2. Referring to the 7 elements of the grammar of graphics, what are this chart's aesthetics, facet, and theme?
3. The writer is debating between a line of best fit and a curve of best fit. Pick which one makes more sense with this data, and support your argument.
4. One of the variables used in the graph is the percent of each market identifying as Hispanic. Estimate it's median.

Part II Data manipulation tools (20 points)

The `gapminder` package contains a data set titled `gapminder` which stores information on life expectency, GDP per capita, and population by country and continent at each five year increment between 1952 and 2007.

You can download the package and take a look at the data using:

```
install.packages("gapminder")
library(gapminder)
gapminder %>% head()
?gapminder
```

Note : *usecodetogetcreditforeachofthefollowingquestions*

1. Which continent shows up most often in this data set?
2. Find the country with the largest life expectency in 1977.
3. Which continent boasts the highest average `gdpPercap` among its countries when looking at year 2007?
4. A social scientiest is interested in comparing life expectency to gdpPercap, and advocates using `lifeExp_per_gdp`, where `lifeExp_per_gdp` = `lifeExp/gdpPercap`. What is the `lifeExp_per_gdp` of the United States in each year?

Part III Graphical Critiques (20 points)

Write clear and concise summaries of the following two graphs:

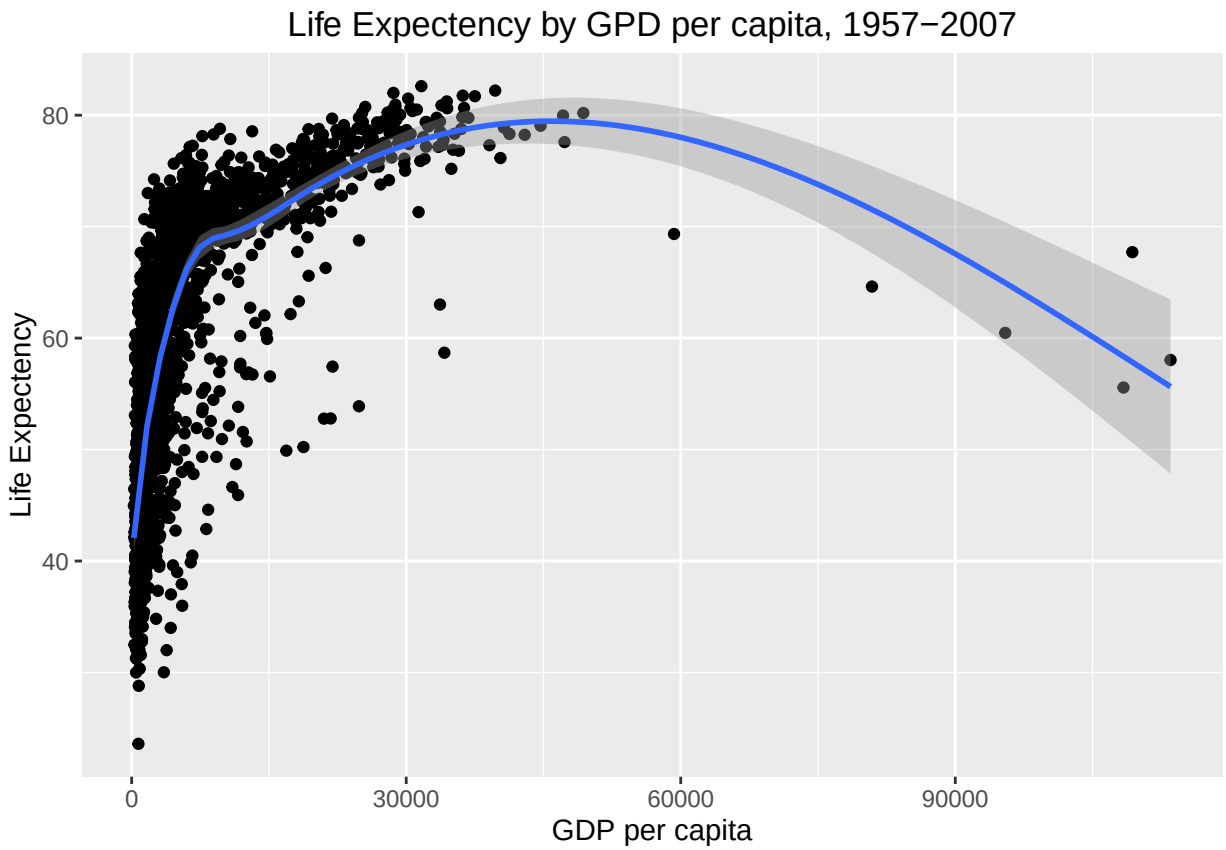
a. State Locations of Joint Statistical Meetings: 1953-2002, shown in this plot (link). The Joint Statistical Meetings is an annual conference of statisticians.

b. Bush and Trump's tax cuts, shown in this plot (link). A more complete article is linked here (link)

Include a brief summary of what the chart is trying to show, comment on problems each chart may have, and propose one idea for making a better chart using the same information.

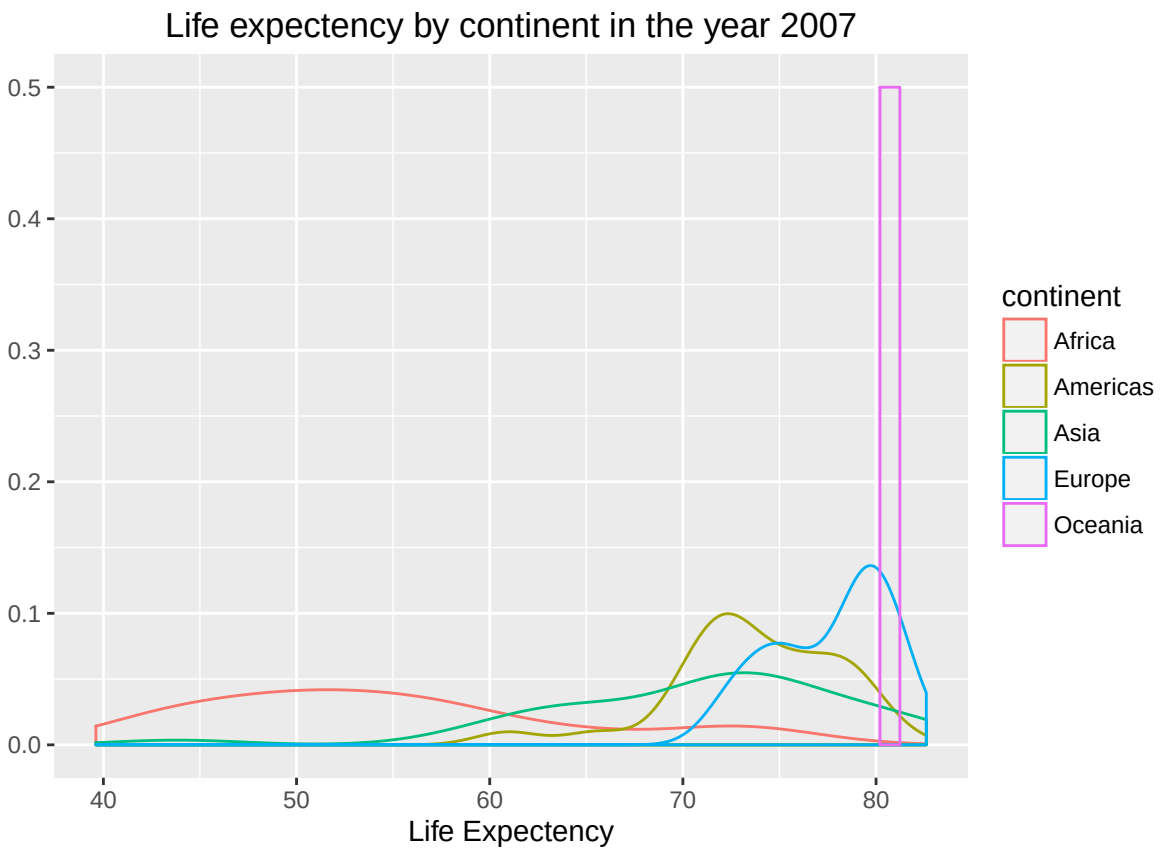
Part IV ggplot, visualization using R (30 points)

1. Make the following chart using the `gapminder` data set.



2. One set of points stands out in the chart above. Using code, identify the country behind those points. What would be one way of making the plot above more visually appealing and intuitive? Note that you should not make a new chart.

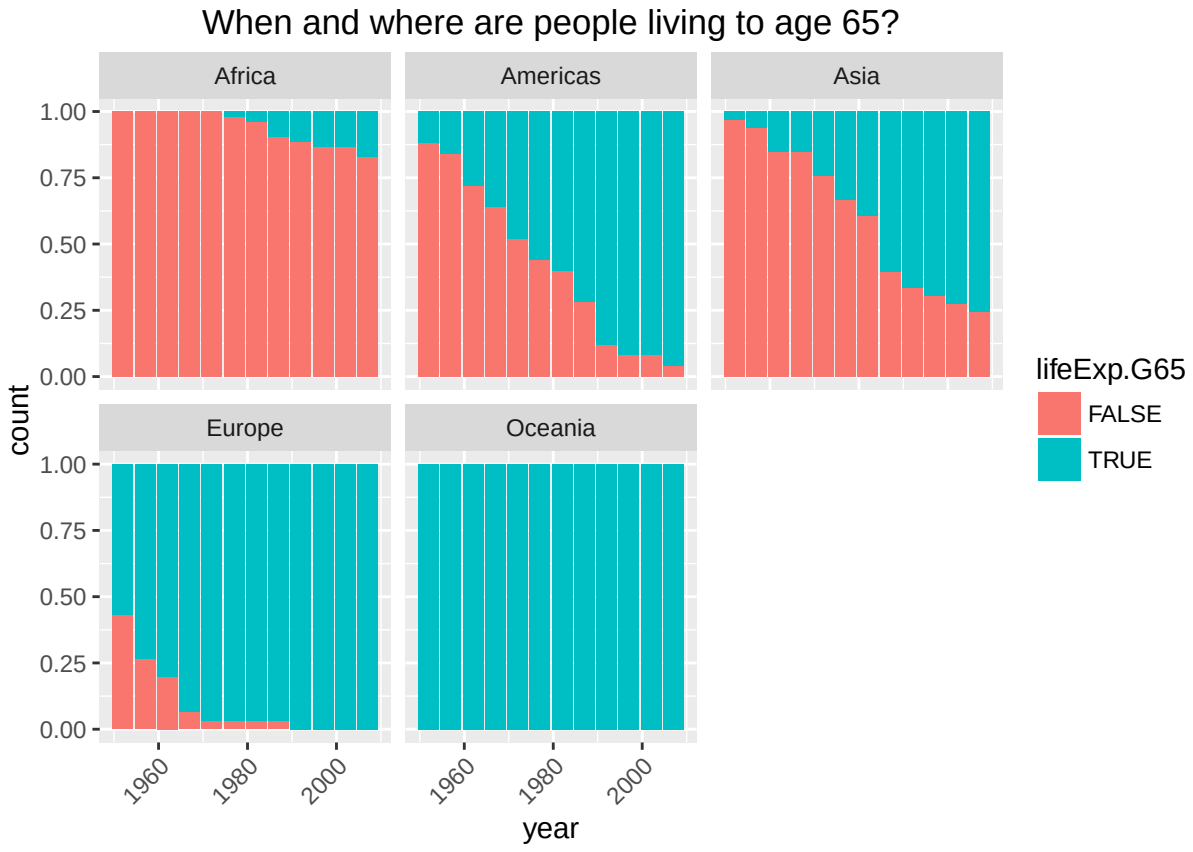
3. Make the following plot using `gapminder`, noting the year in the title.



4. Contrast the shapes of the life expectencies among countries in Asia and Africa in the plot above.

5. Use the following code as a guide to help make the final chart of your exam.

```
gapminder1 <- gapminder %>%
  mutate(lifeExp.G65 = lifeExp > 65)
```



6. What does the graph in question 5 show you? Next, provide 2 ways to improve this graph.