Homework 02

DSO 545: Statistical Computing and Data Visualization

Fall 2019

Due Date: Saturday September 14, 2019 (at the end of the day at 11:59pm)

Instructions

- Use Python to answer all questions (Feel free to use any IDE or Jupyter notebook)
- Submit your Python file to blackboard with comments when needed
- USC won't tolerate any kind of cheating
- Good luck

01. Gapminder Data

The mission of the GapMinder foundation is to promote sustainable global development and achievement of the United Nations Millennium Goals by increased use and understanding of statistics and other information about social, economic and environmental development at the local, national and global levels. The dataset (`gapminder.tsv`) that we are using in this homework assignment comes from (https://www.gapminder.org/data/), and it includes six variables:

Variable	Description
country	–
continent	–
year	–
lifeExp	life expectancy at birth
pop	total population
gdpPercap	per-capita GDP

Questions

1. **(1 point)** What is the value of the life expectancy for the 100th row?
2. **(1 point)** Create a subset of the gapminder data that includes the country, year, and gdp per capita for all observations. Save your result in variable `gapminder_subset`.
3. **(1 point)** Which years are presented in this dataset?
4. **(1 point)** What is the average life expectancy for all countries in 2007?
5. **(2 points)** Create a dataframe that has all countries in Europe with GDP per capita less than the average European GDP per capita in 2007? Save your result in a variable `result`.
6. **(2 points)** Using the `result` dataframe from the previous question, create a data dictionary for the countries in Europe with GDP per capita less than the average European GDP per capita in 2007? This dictionary should have two keys: country, and gdpPercap, and the values for each key are represented using a list.

02. Case in Point: Graph Analysis for Consulting and Case Interviews

7. **(2 points)** Read pages 1-10 (excluding Pie Graphs), and create a powerpoint presentation that discusses the main points in these pages (max of 6 slides).