

# Homework 2

*Your Name*

*Due: October 12, 2019 at 11:59pm*

## Homework Policies:

*You are encouraged to discuss problem sets with your fellow students (and with the Course Instructor of course), but you must write your own final answers, in your own words. Solutions prepared “in committee” or by copying someone else’s paper are not acceptable. This violates the Brown standards of plagiarism, and you will not have the benefit of having thought about and worked the problem when you take the examinations.*

*All answers must be in complete sentences and all graphs must be properly labeled.*

*In this homework you will be required to use .Rmd to do it., you can then knit to a word document of PDF to turn it in.*

***For the PDF Version of this assignment: PDF***

***For the R Markdown Version of this assignment: RMarkdown***

## Turning the Homework in:

*Please turn the homework in through canvas. You may use a pdf, html or word doc file to turn the assignment in.*

PHP 1510 Assignment Link

PHP 2510 Assignment Link

## The Data

We will work with the dataset called gapminder, this is a cleaned up version from Gapminder Data. Gapminder contains a lot of great data on all of the nations of the world. We first need to install the gapminder package in R.

```
install.packages("gapminder")  
  
library(dplyr)  
library(gapminder)  
gapminder
```

## Problems for Everyone

Use **dplyr** functions to address the following questions. For some you can just use arrange to print your solutions to the top.

1. How many unique countries are represented per continent?
2. Which European nation had the lowest GDP per capita in 1997?
3. According to the data available, what was the average life expectancy across each continent in the 1980s?
4. What 5 countries have the highest total GDP over all years combined?
5. What countries and years had life expectancies of *at least* 80 years? *Only output the columns of interest: country, life expectancy and year (in that order).*
6. Which three countries have had the most consistent population estimates (i.e. lowest standard deviation) across the years of available data?
7. Follow the steps below to create a plot about life expectancy.

- a. Create a plot of life expectancy over time where each country has its own line (`group=country`).
  - b. Add a layer to this plot where you use `geom_smooth(method="lm")`
  - c. Add a layer to this plot where you use `geom_smooth(method="lm")` but it is colored by continent.
8. Interpret the graph you created in 7.
  - a. What types of patterns are you seeing?
  - b. Do all countries follow this pattern?
  - c. etc...
9. Create boxplots of life expectancy by continent. Add a layer using `geom_jitter()` to see how the points fall in these boxplots. **\*\*Hint: using `alpha=0.04` inside the jitter will lighten the points\***
10. Interpret the plot you made in 9.

## PHP 2560 Only

11. Which combinations of continent (besides Asia) and year have the highest average population across all countries? *your output should include all results sorted by highest average population.* With what you already know, this one may stump you. See this Q&A for how to **ungroup** before **arranging**.
12. Consider the function below

```
hourly_delay <- filter(
  summarise(
    group_by(
      filter(
        flights,
        !is.na(dep_delay)
      ),
      month, day, year, hour
    ),
    delay=mean(dep_delay),
    n=n()
  ),
  n>10
)
```

- a. What are some problems with this function?
- b. How easy is it to follow the logic of this?
- c. Rewrite this using piping to make it more understandable.
- d. Does your rewritten command give the same results?