

# DS 5500 Homework 1 - Due Oct. 7

## Instructions

Create a new public Github repository for this homework assignment. The repository should include all of the code necessary to reproduce your submitted solutions. Do not include the raw data in the repository.

*Use the README.md of the Github repository to present your solutions. Answer all questions completely for full credit, including figures and tables where appropriate. For each problem, either provide relevant inline code snippets, or cite the source file where the relevant code lives (with line numbers if appropriate).*

Describe any data processing steps (transformation, filtering, etc.) you perform when solving each problem, providing reasoning where appropriate. You may need to be creative when deciding how to approach each problem, as there may not be a single “correct” solution.

Your solutions should be posted as a **public note** on Piazza in the *hw1* folder with the title “[hw1] - your name name”. Include in the body of the note a link to the Github repository with your solutions.

## Overview

This homework asks you to visualize data from the Gapminder Foundation (<https://www.gapminder.org>).

Download the full Gapminder dataset from the Open Numbers Github repository:

- [https://github.com/open-numbers/ddf-gapminder-systema\\_globalis](https://github.com/open-numbers/ddf-gapminder-systema_globalis)

## Problem 1

Take the Gapminder Test: <http://forms.gapminder.org/s3/test-2018>.

What score did you receive? Did any of the answers surprise you?

Choose a question from the test, re-state it, and answer it using visualization and summarization. Provide a figure and any relevant output with your answer.

## Problem 2

Visualize the distribution of income (GDP / capita) across countries and continents, and how the distribution of income changes over time.

Interpret the visualization and what you notice. Are there any notable trends and/or deviations from that trend?

What caveats apply to your conclusions?

## Problem 3

Use visualization to investigate the relationship between income (GDP / capita), life expectancy, and child mortality over time. How does each measure change over time within each continent?

Interpret your visualizations, noting any trends and/or outliers.

#### **Problem 4**

Choose two variables you have not investigated yet, and visualize their distributions, their relationship with each other, and how these change over time.

Interpret your visualizations, noting any trends and/or outliers.

#### **Problem 5**

Did you use static or interactive plots to answer the previous problems?

Explore the data using the interactive visualization tools at <https://www.gapminder.org/tools>, and watch the TED talk “The best stats you’ve ever seen” at <https://www.youtube.com/watch?v=hVimVzgtD6w>.

Discuss the advantages, disadvantages, and relative usefulness of using interactive/dynamic visualizations versus static visualizations.