

# Exploring Data through Visualizations

Here, you will use two independent data sets from the `mdsr` and `gapminder` packages. The `mdsr` package is for the [Modern data science with R. 2nd Edition \(2021\)](#) by Baumer, B. S., Kaplan, D. T., & Horton, N. J.

```
# load the packages
```

```
library(mdsr)
```

```
library(gapminder)
```

```
library(tidyverse)
```

1. Using the `CIACountries` data set from the `mdsr` package, create a bar plot where the x-axis is the population and the y-axis is the top 15 most populous countries. Color each bar according to the country's area. Write a paragraph describing your observations.
2. Using the `gapminder` dataset from the `gapminder` package, create a subset where you take only years 1990-2000. Take the maximum population in each country and average the life expectancy, and make sure that the continent variable is retained for each country. Create a scatterplot where the x-axis is the gdp per capita and the y-axis is the life expectancy. Make the size of each point to be proportional to the population and color each point to be proportional to the continent. Write a paragraph describing your observations.
3. Explore either data sets and make a plot where you incorporate 3-4 variables. Write a paragraph describing your observations.
4. (bonus) Use the `full_join()` function to combine the `gapminder` and `CIACountries` data set using the country variable. Note that the `gapminder` data set has the year variable - so you need to do some extra data wrangling on it. Create a scatter plot comparing life expectancy and education. Color each point to be proportional to the continent variable, and the size proportional to the net users. Write a paragraph describing your observations.