Econ 448 First Week Exercise

Check Canvas for Due Date

Alex Lin

10/05/2025

```r
library(ggplot2)
library(gapminder)
library(dplyr)
```

Getting Started with R

You will familiarize yourself with the R computing language, R Studio environment, and R Markdown language. If you are already familiar with these, then lucky you, you are ahead of me! If not, read on.

There are two ways to complete this project: Downloading and installing R on your own machine (or using an already installed version of R on a campus computer), or running this notebook in the cloud version of R.

Installing R on Your Personal Machine

- Download R and RStudio (recommended over using R alone)
- (Windows) Install Rtools. Note that there are instructions on this page for creating a text file called .Renviron that identifies a path for R to use to find Rtools. It seemse complicated, but is fairly straightforward in practice.

Using R Studio in the Cloud - Preferred

- I have set up a Jupyter Notebook for use in this class to make it really easy for you to run these programs in a cloud version of R Studio. The link is in the class navigation pane.

R Markdown and Notebooks

This is the html document produced by an R Markdown Notebook. You can download the code to run in R above (Code>Download Rmd). When you execute code within the notebook, the results appear beneath the code. You can download this file, experiment with it, and change it to suit your needs. When you are ready to finalize your results, make sure that you have run all of the relevant chunks of code and then click preview (do not knit!) above to create a html document of your own with code and results included.

The grey box below is called a chunk. Try executing this chunk by clicking the green arrow on the far left, the *Run* button within the chunk, or by placing your cursor inside it and pressing *Ctrl+Shift+Enter*.

```
## This is a comment and the # tells R not to read it.  Run the command below to print out the message
print ("Hello World")
```

```
## [1] "Hello World"
```

Once you've installed RStudio, you have to install your packages (but only do this once). Open a script (File > New File > R Script) and type the commands below (or run them in the notebook).

If you use the class jupyterhub, you may not need to install any packages.

```
## Run this code if you haven't run it before, but only needs to be run once after you have installed R
##if (!require("pacman")) install.packages("pacman") #pacman is the package that installs packages nice
##pacman::p_load(rmarkdown, dplyr, gapminder, tidyverse, xml2, ggplot2, dplyr, lifecycle) #these are th
```

After installing your packages, you need to load the ones you will be using in your program every time.

```
## Only needs to be run once per session.
library(ggplot2)
library(gapminder)
library(dplyr)
```

Hint: You can also include the commands above at the start of a script if you decide to create and run your own code.

Introduction to Data Work in R

```
#Below, you can practice a calculation in R.  If you run the code below, R will output the sum.
3+1
```

```
## [1] 4
```

```
#But maybe you want to perform an operation and save the results in a variable so that you can use them
#This code shows you how to assign a value to a variable x using a calculation, and then prints the val
x<-3+1
x
```

```
## [1] 4
```

```
#In fact, R thinks in vectors and matrices works best when you assign a bunch of values to a variable.
x1<-c(1,2,3,4) #This makes a 1 dimensional vector
x2<-matrix(c(1,2,3,4), nrow=2, ncol=2) #This makes a 2x2 matrix
x1
```

```
## [1] 1 2 3 4
```

```
x2
```

```
##      [,1] [,2]
## [1,]    1    3
## [2,]    2    4
```

You can try modifying the numbers above yourself and making different size matrices, etc.

Data Frames

Data frames are a useful way that R organizes data. These are flexible versions of matrices that have special properties that make them easy to work with. Later, you will import some actual data from the internet into a data frame, but for now we will use the gapminder data that is included in the package you loaded earlier.

Gapminder (gapminder.org) data is data on life expectancy and GDP by country over time. It can be used to calculate Preston Curves, as discussed in class. For now, let's just perform a few operations on it.

```
#Some quick summary statistics
summary(gapminder)
```
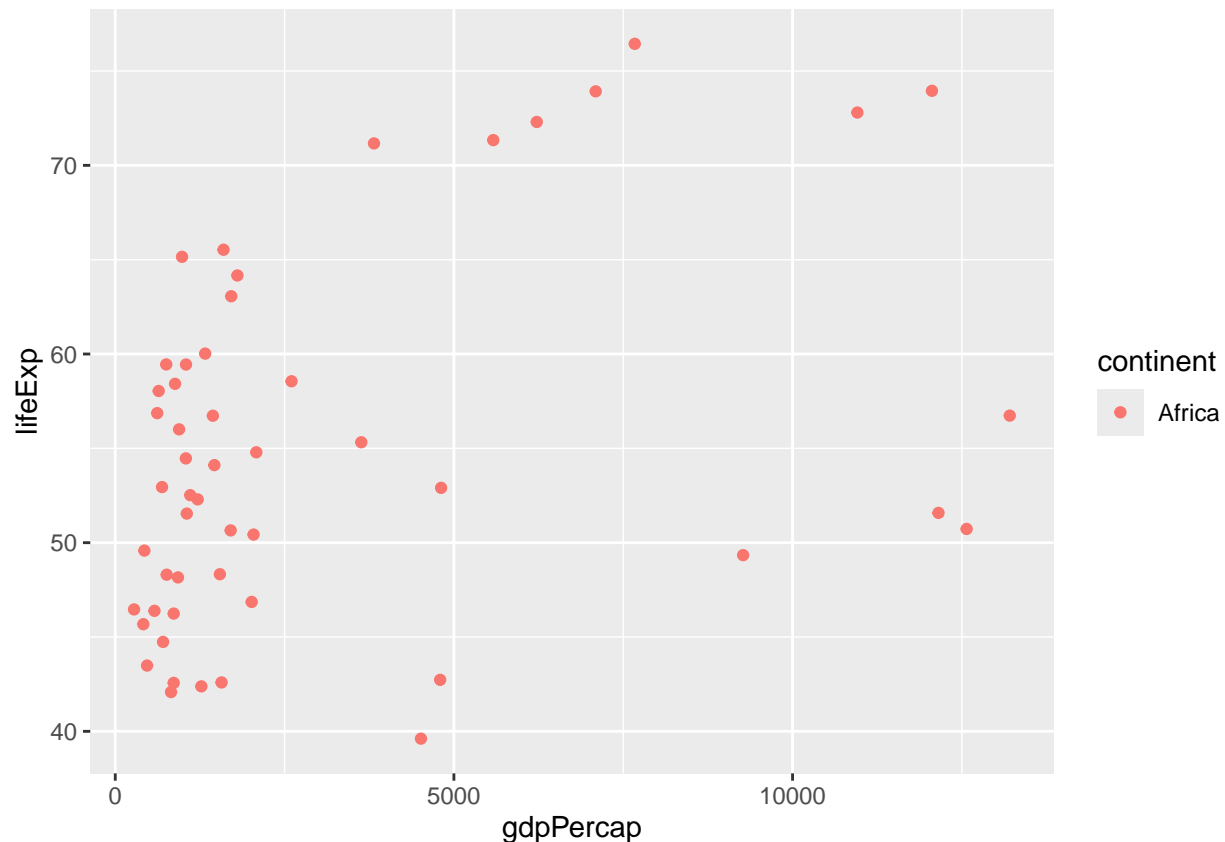
```
##      country          continent       year         lifeExp
##  Afghanistan:  12   Africa  :624   Min.   :1952   Min.   :23.60
##  Albania    :  12   Americas:300   1st Qu.:1966   1st Qu.:48.20
##  Algeria    :  12   Asia    :396   Median :1980   Median :60.71
##  Angola     :  12   Europe  :360   Mean   :1980   Mean   :59.47
##  Argentina  :  12   Oceania : 24   3rd Qu.:1993   3rd Qu.:70.85
##  Australia  :  12                  Max.   :2007   Max.   :82.60
##  (Other)    :1632
##       pop               gdpPercap
##  Min.   :6.001e+04   Min.   :    241.2
##  1st Qu.:2.794e+06   1st Qu.:   1202.1
##  Median :7.024e+06   Median :   3531.8
##  Mean   :2.960e+07   Mean   :   7215.3
##  3rd Qu.:1.959e+07   3rd Qu.:   9325.5
##  Max.   :1.319e+09   Max.   :113523.1
##
```

```
# Load relevant data into data frame DF_2007 for the year 2007 and the continent of Africa.
## What is the median life expectancy in the continent in 2007?
DF<-data.frame(gapminder)
DF_2007<-filter(DF, year==2007 & continent=="Africa")
summary(DF_2007)
```

```
##          country        continent      year         lifeExp
##  Algeria     : 1   Africa  :52   Min.   :2007   Min.   :39.61
##  Angola      : 1   Americas: 0   1st Qu.:2007   1st Qu.:47.83
##  Benin       : 1   Asia    : 0   Median :2007   Median :52.93
##  Botswana    : 1   Europe  : 0   Mean   :2007   Mean   :54.81
##  Burkina Faso: 1   Oceania : 0   3rd Qu.:2007   3rd Qu.:59.44
##  Burundi     : 1                 Max.   :2007   Max.   :76.44
##  (Other)     :46
##       pop              gdpPercap
##  Min.   :   199579   Min.   :  277.6
##  1st Qu.:  2909226   1st Qu.:  863.0
##  Median : 10093310   Median : 1452.3
##  Mean   : 17875763   Mean   : 3089.0
##  3rd Qu.: 19363654   3rd Qu.: 3993.5
##  Max.   :135031164   Max.   :13206.5
##
```

Plot the relationship below. The plot represents the variation in life expectancy (within a single continent) in a single period in time. The level of health technology should be (approximately) the same across all points and so the plotted relationship is like a snapshot of the variation in life expectancy with income at that particular time and state of the world.

```
## Create a scatter plot of life expectancy and GDP per capita in 2007 for the continent of Africa.
ggplot(DF_2007, aes(x=gdpPercap, y=lifeExp, color=continent))+geom_point()
```


Now let's look at life expectancy as a function of income for a single country at different periods of time.
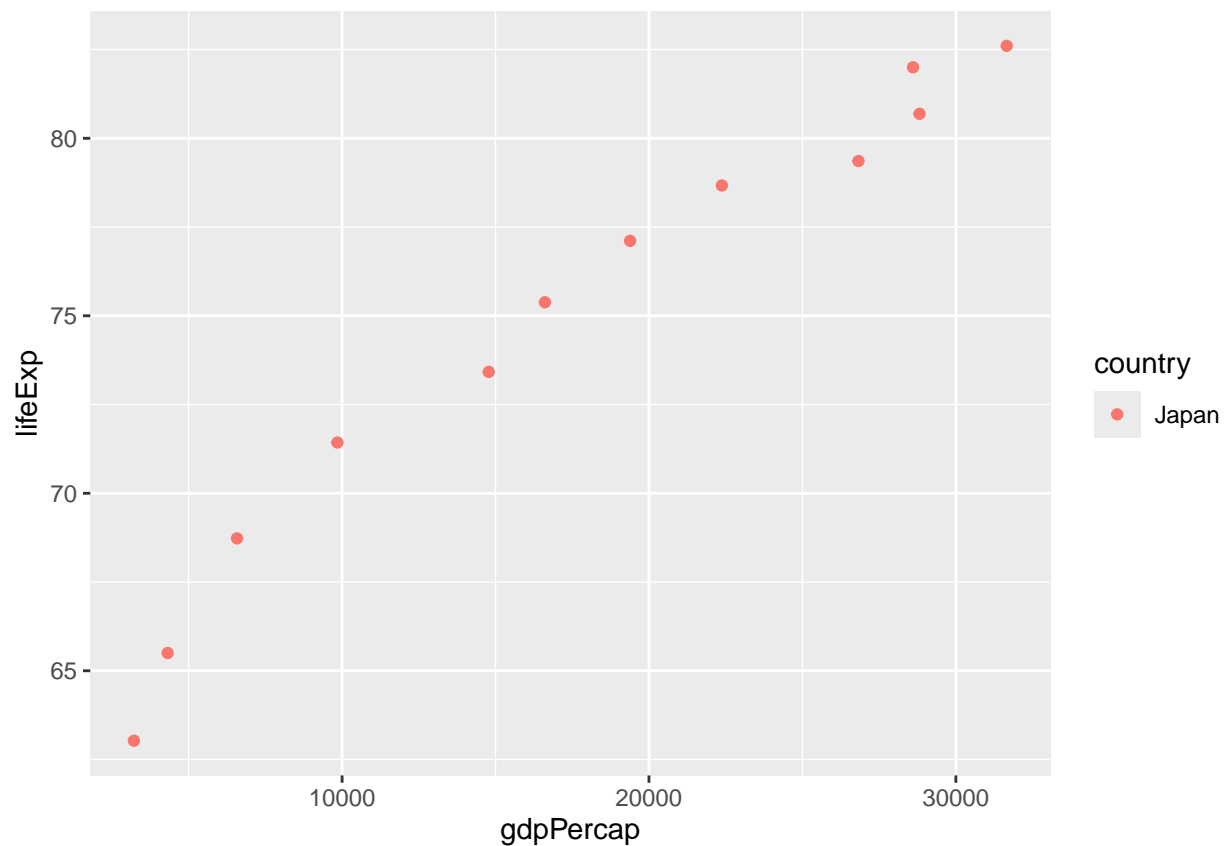
```
# Load relevant data into data frame DF_country for a country of your choosing for all years of data.
DF_country<-filter(gapminder, country=="Japan")
summary(DF_country)
```

```
##          country       continent       year         lifeExp
##   Japan       :12   Africa  : 0   Min.   :1952   Min.   :63.03
##   Afghanistan: 0   Americas: 0   1st Qu.:1966   1st Qu.:70.75
##   Albania    : 0   Asia    :12   Median :1980   Median :76.25
##   Algeria    : 0   Europe  : 0   Mean   :1980   Mean   :74.83
##   Angola     : 0   Oceania : 0   3rd Qu.:1993   3rd Qu.:79.69
##   Argentina  : 0                 Max.   :2007   Max.   :82.60
##   (Other)    : 0
##        pop               gdpPercap
##   Min.   : 86459025   Min.   : 3217
##   1st Qu.: 99576898   1st Qu.: 9030
##   Median :116163724   Median :17997
```

```
##  Mean   :111758808   Mean   :17751
##  3rd Qu.:124736076   3rd Qu.:27270
##  Max.   :127467972   Max.   :31656
##
```

Plot the relationship below. The plot represents the evolution of life expectancy as the economy grows for one country, incorporating the social, political, and health contexts of that particular country, as well as changing health technology.

```
## Create a scatter plot of life expectancy and GDP per capita for the country of Japan.
ggplot(DF_country, aes(x=gdpPercap, y=lifeExp, color=country))+geom_point()
```


**** ## To Do To complete this assignment, answer the following questions. You can submit these results however you choose, but a simple option would be to modify this notebook to answer the questions above, execute the code, and hit "preview" in R Studio to create a html document that can be submitted with code and results.

1. What is the median life expectancy in the world in 1952? In 2007?

```
global_medians <-
  gapminder %>%
  filter(year %in% c(1952, 2007)) %>%
  group_by(year) %>%
  summarise(median_lifeExp = median(lifeExp), .groups = "drop")

global_medians
```

```
## # A tibble: 2 x 2
##    year median_lifeExp
##   <int>          <dbl>
## 1  1952           45.1
## 2  2007           71.9
```

2. Choose one continent other than Africa. What is the median life expectancy in that continent in 1952 and 2007? Plot the life expectancy against GDP per capita for every country in that continent for the year 2007.
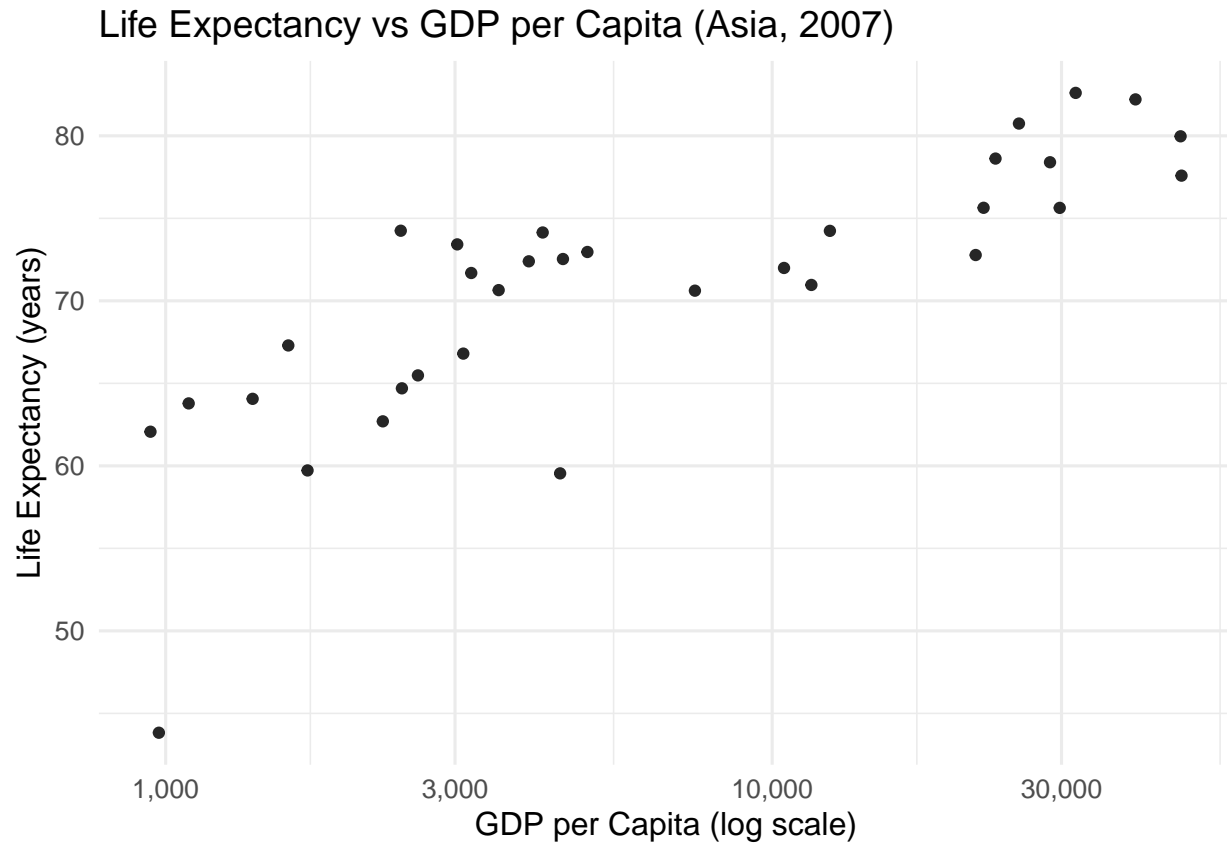
```
continent_medians <-
  gapminder %>%
  filter(continent == "Asia", year %in% c(1952, 2007)) %>%
  group_by(year) %>%
  summarise(median_lifeExp = median(lifeExp), .groups = "drop")

continent_medians
```

```
## # A tibble: 2 x 2
##    year median_lifeExp
##   <int>          <dbl>
## 1  1952           44.9
## 2  2007           72.4
```

```
df_cont_2007 <-
  gapminder %>%
  dplyr::filter(continent == "Asia", year == 2007)

ggplot(df_cont_2007, aes(x = gdpPercap, y = lifeExp)) +
  geom_point(alpha = 0.85) +
  scale_x_log10(labels = scales::label_comma()) +
  labs(
    title = "Life Expectancy vs GDP per Capita (Asia, 2007)",
    x = "GDP per Capita (log scale)",
    y = "Life Expectancy (years)"
  ) +
  theme_minimal(base_size = 12)
```
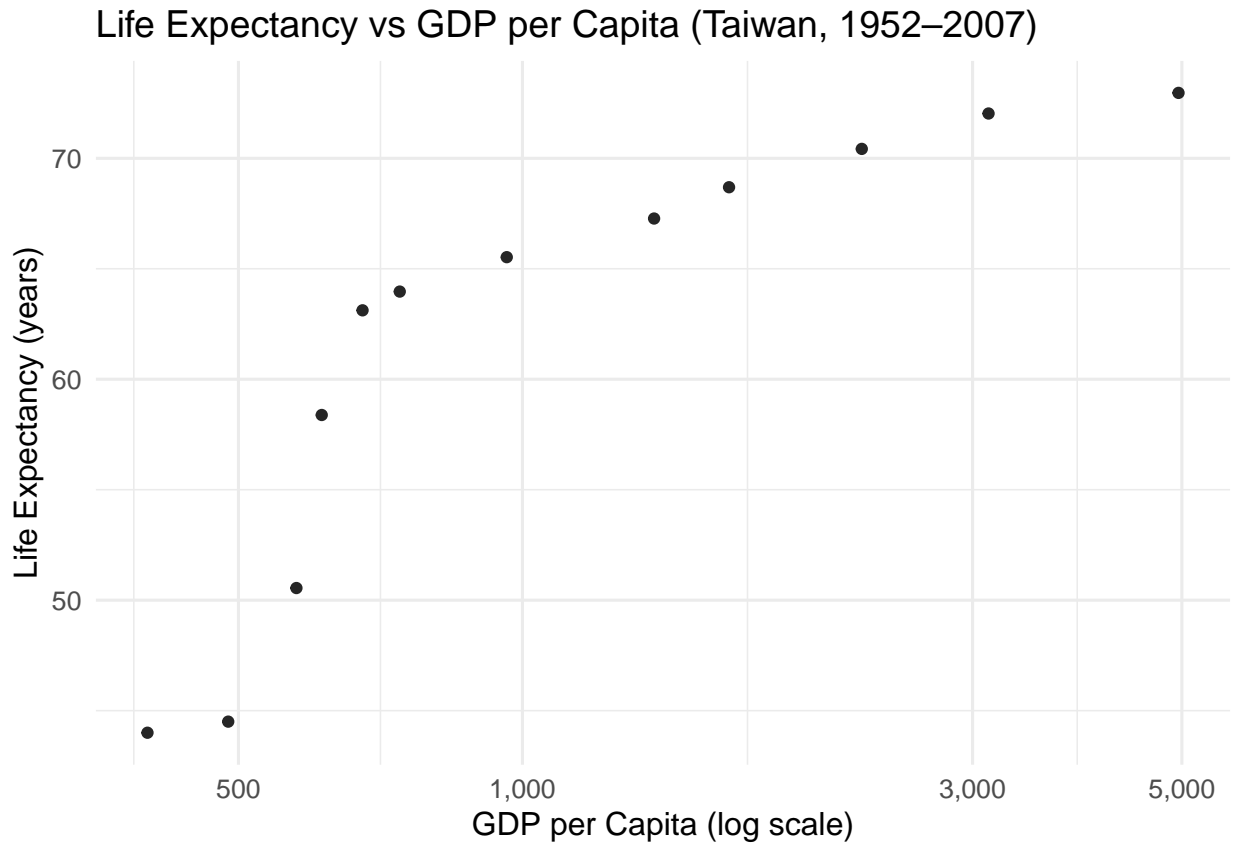
Life Expectancy vs GDP per Capita (Asia, 2007)

3. Choose a country other than Japan. Plot the life expectancy against GDP per capita from 1952 to 2007.

```r
df_country <-
  gapminder %>%
  dplyr::filter(country == "China")

ggplot(df_country, aes(x = gdpPercap, y = lifeExp)) +
  geom_point(alpha = 0.85) +
  scale_x_log10(labels = scales::label_comma()) +
  labs(
    title = paste0("Life Expectancy vs GDP per Capita (Taiwan, 1952-2007)"),
    x = "GDP per Capita (log scale)",
    y = "Life Expectancy (years)"
  ) +
  theme_minimal(base_size = 12)
```

Life Expectancy vs GDP per Capita (Taiwan, 1952–2007)

4. What is your experience with R? If you haven't used R before, do you have experience with any other programming languages or statistical software (SPSS, SAS, Stata, etc.) or Excel?

A: I have used the R language in several classes, including Economics, Applied Math, and Informatics. ****