Econ 448 First Week Exercise

Code

Melissa Knox

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 (https://cran.r-project.org/) and RStudio (https://www.rstudio.com/products/rstudio/download/) (recommended over using R alone)
- (Windows) Install Rtools (https://cran.r-project.org/bin/windows/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 (http://rmarkdown.rstudio.com) 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*.

Hide

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

Hide

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

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

Hide

```
## 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

Hide

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

```
[1] 4
```

Hide

```
#But maybe you want to perform an operation and save the results in a variable so that y
ou can use them later in other calculations?
#This code shows you how to assign a value to a variable x using a calculation, and then
prints the value of x.
x<-5+1
x
```

```
[1] 6
```

Hide

```
#In fact, R thinks in vectors and matrices works best when you assign a bunch of values
to a variable.
x1<-c(5,10,15,20,25,30,35,40) #This makes a 1 dimensional vector
x2<-matrix(c(5,10,15,20,25,30,35,40), nrow=2, ncol=4) #This makes a 2x2 matrix
x1
```

```
[1]  5 10 15 20 25 30 35 40
```

Hide

```
x2
```

```
     [,1] [,2] [,3] [,4]
[1,]    5   15   25   35
[2,]   10   20   30   40
```

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.

Hide

```
#Some quick summary statistics
summary(gapminder)
```

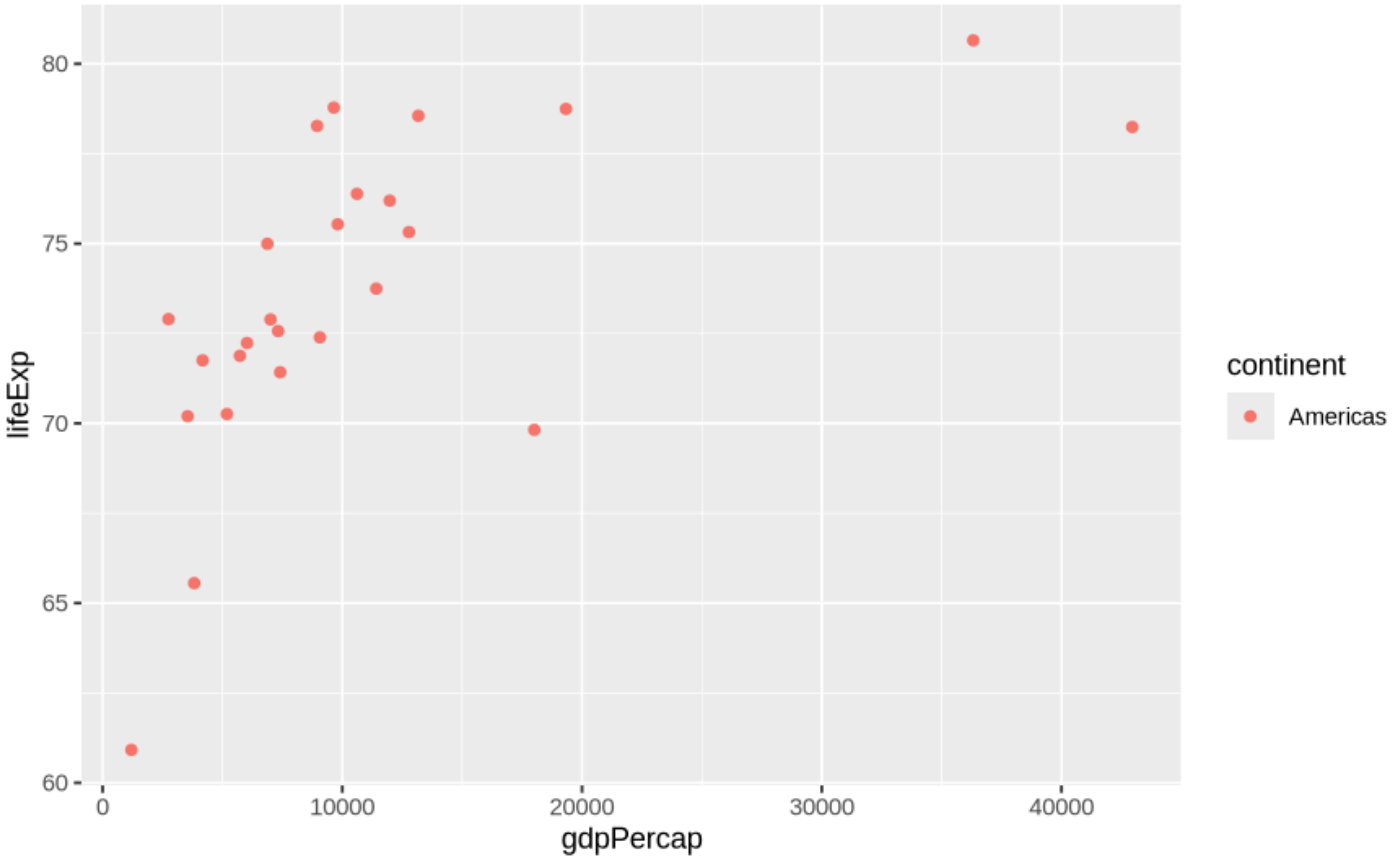
```
        country        continent        year         lifeExp           pop
gdpPercap
 Afghanistan:  12   Africa  :624   Min.   :1952   Min.   :23.60   Min.   :6.001e+04   Mi
n.   :   241.2
 Albania    :  12   Americas:300   1st Qu.:1966   1st Qu.:48.20   1st Qu.:2.794e+06   1s
t Qu.:  1202.1
 Algeria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

Hide

```
# Load relevant data into data frame DF_2007 for the year 2007 and the continent of Afri
ca.
## What is the median life expectancy in the continent in 2007?
DF<-data.frame(gapminder)
DF_1952<-filter(DF, year==1952 & continent=="Americas")
summary(DF_1952)
```

```
      country      continent       year         lifeExp           pop              gdpPe
rcap
 Argentina: 1   Africa  : 0   Min.   :1952   Min.   :37.58   Min.   :   662850   Min.
: 1398
 Bolivia  : 1   Americas:25   1st Qu.:1952   1st Qu.:45.26   1st Qu.:  1555876   1st Q
u.: 2428
 Brazil   : 1   Asia    : 0   Median :1952   Median :54.74   Median :  3146381   Median
: 3048
 Canada   : 1   Europe  : 0   Mean   :1952   Mean   :53.28   Mean   : 13806098   Mean
: 4079
 Chile    : 1   Oceania : 0   3rd Qu.:1952   3rd Qu.:59.42   3rd Qu.:  8025700   3rd Q
u.: 3940
 Colombia : 1                 Max.   :1952   Max.   :68.75   Max.   :157553000   Max.
:13990
 (Other)  :19
```

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.



Now let's look at life expectancy as a function of income for a single country at different periods of time.

Hide

```
# Load relevant data into data frame DF_country for a country of your choosing for all y
ears of data.
DF_country<-filter(gapminder, country=="Japan")
summary(DF_country)
```

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.

Hide

```
## 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? Median life expectancy for 1952 is 45.14 and 72.9 in 2007.

Hide

```
summary(DF_1952)
```

```
       country       continent       year         lifeExp          pop             gd
pPercap
 Afghanistan:  1   Africa  :52   Min.   :1952   Min.   :28.80   Min.   :    60011   Min.
:   298.9
 Albania    :  1   Americas:25   1st Qu.:1952   1st Qu.:39.06   1st Qu.:  1452026   1st
Qu.:   864.8
 Algeria    :  1   Asia    :33   Median :1952   Median :45.14   Median :  3943953   Medi
an :  1968.5
 Angola     :  1   Europe  :30   Mean   :1952   Mean   :49.06   Mean   : 16950402   Mean
:  3725.3
 Argentina  :  1   Oceania : 2   3rd Qu.:1952   3rd Qu.:59.77   3rd Qu.:  9168198   3rd
Qu.:  3913.5
 Australia  :  1                 Max.   :1952   Max.   :72.67   Max.   :556263527   Max.
:108382.4
 (Other)    :136
```

Hide

```
summary(DF_2007)
```

```
      country      continent       year         lifeExp           pop             gdpPe
rcap
 Argentina: 1   Africa  : 0   Min.   :2007   Min.   :60.92   Min.   :  1056608   Min.
: 1202
 Bolivia  : 1   Americas:25   1st Qu.:2007   1st Qu.:71.75   1st Qu.:  5675356   1st Q
u.: 5728
 Brazil   : 1   Asia    : 0   Median :2007   Median :72.90   Median :  9319622   Median
: 8948
 Canada   : 1   Europe  : 0   Mean   :2007   Mean   :73.61   Mean   : 35954847   Mean
:11003
 Chile    : 1   Oceania : 0   3rd Qu.:2007   3rd Qu.:76.38   3rd Qu.: 28674757   3rd Q
u.:11978
 Colombia : 1                 Max.   :2007   Max.   :80.65   Max.   :301139947   Max.
:42952
 (Other)  :19
```

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. I chose the Americas and the life expectancy in 1952 was 54.74 and in 2007,72.90.

Hide

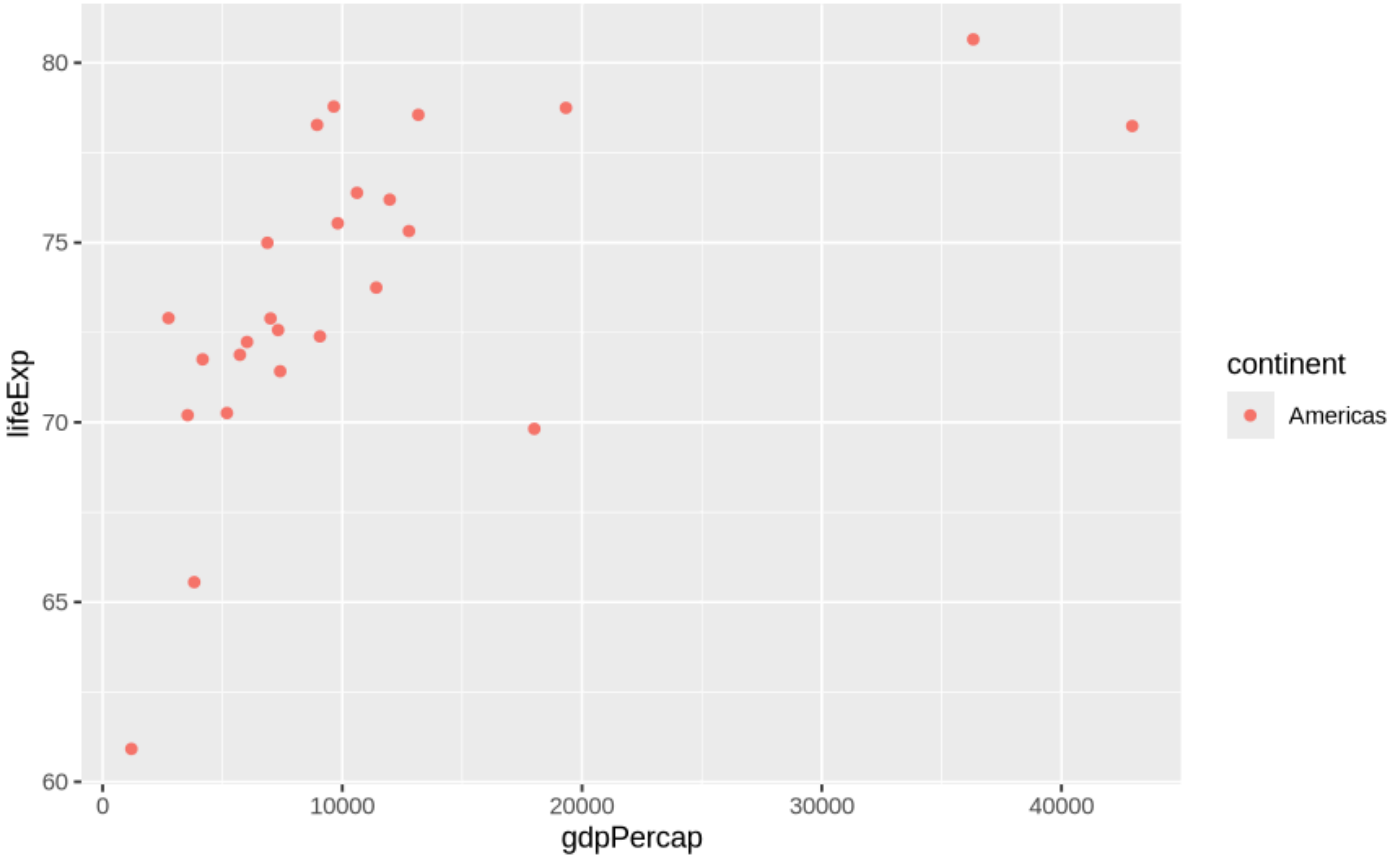
```
summary(DF_1952)
```

```
      country      continent       year         lifeExp           pop             gdpPe
rcap
 Argentina: 1   Africa  : 0   Min.   :1952   Min.   :37.58   Min.   :   662850   Min.
: 1398
 Bolivia  : 1   Americas:25   1st Qu.:1952   1st Qu.:45.26   1st Qu.:  1555876   1st Q
u.: 2428
 Brazil   : 1   Asia    : 0   Median :1952   Median :54.74   Median :  3146381   Median
: 3048
 Canada   : 1   Europe  : 0   Mean   :1952   Mean   :53.28   Mean   : 13806098   Mean
: 4079
 Chile    : 1   Oceania : 0   3rd Qu.:1952   3rd Qu.:59.42   3rd Qu.:  8025700   3rd Q
u.: 3940
 Colombia : 1                 Max.   :1952   Max.   :68.75   Max.   :157553000   Max.
:13990
 (Other)  :19
```

Hide

```
summary(DF_2007)
```

```
      country      continent       year         lifeExp           pop             gdpPe
rcap
 Argentina: 1   Africa  : 0   Min.   :2007   Min.   :60.92   Min.   :  1056608   Min.
: 1202
 Bolivia  : 1   Americas:25   1st Qu.:2007   1st Qu.:71.75   1st Qu.:  5675356   1st Q
u.: 5728
 Brazil   : 1   Asia    : 0   Median :2007   Median :72.90   Median :  9319622   Median
: 8948
 Canada   : 1   Europe  : 0   Mean   :2007   Mean   :73.61   Mean   : 35954847   Mean
:11003
 Chile    : 1   Oceania : 0   3rd Qu.:2007   3rd Qu.:76.38   3rd Qu.: 28674757   3rd Q
u.:11978
 Colombia : 1                 Max.   :2007   Max.   :80.65   Max.   :301139947   Max.
:42952
 (Other)  :19
```

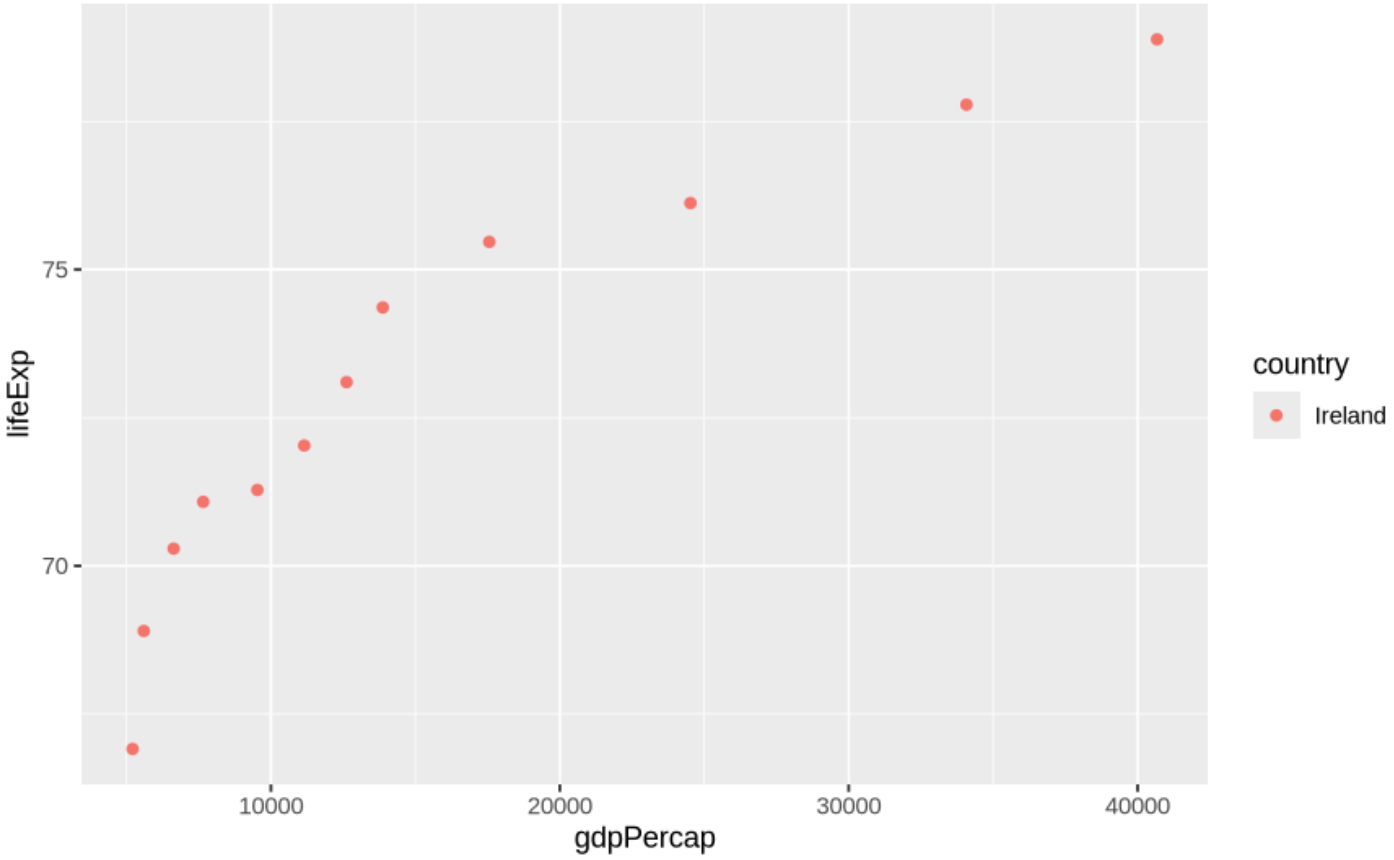


3. Choose a country other than Japan. Plot the life expectancy against GDP per capita from 1952 to 2007.

Hide

```
summary(DF_country)
```

```
        country      continent       year         lifeExp           pop            gdpPe
rcap
 Ireland    :12   Africa  : 0   Min.   :1952   Min.   :66.91   Min.   :2830000   Min.
: 5210
 Afghanistan: 0   Americas: 0   1st Qu.:1966   1st Qu.:70.88   1st Qu.:2939142   1st Q
u.: 7400
 Albania    : 0   Asia    : 0   Median :1980   Median :72.56   Median :3375950   Median
:11885
 Algeria    : 0   Europe  :12   Mean   :1980   Mean   :73.02   Mean   :3340826   Mean
:15759
 Angola     : 0   Oceania : 0   3rd Qu.:1993   3rd Qu.:75.63   3rd Qu.:3585129   3rd Q
u.:19300
 Argentina  : 0                 Max.   :2007   Max.   :78.89   Max.   :4109086   Max.
:40676
 (Other)    : 0
```



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? I used R during my Stats 211 at North Seattle College back in 2019 or 2020. I will be refreshing myself on the intricacies of R again as it has several years since using R and I am not a programmer by trade. I have used Excel to sort data for a class I took in the Evans school for Public Policy that utilized GIS. Two quarters ago, I was in Labor Market Analysis and used Stata to a very small degree.