Week 3 SEP 8 - SEP 14 Lab 2 Population Growth

Lysa Vanible

September 14, 2021

##**Introduction**
In this lab we'll explore the population growth rates of the US, as well as their doubling times. The population growth rate tells us how fast a population is growing in percentage terms. The doubling time tells us the number of years it will take for a population to double in size if it continues at a specific growth rate. You will be getting the data for this lab from mortality.org. To do that you will need to create an account and register on the mortality.org website. Go to the website and click on the new user link at the top ofthe left side menu. Input the requested information to get a username and password and make note of those in a secure place. You will need them for multiple labs.

1). Replace "Write your name here" in the YAML header above with your name before you start. MAKE SURE NOT TO MODIFY ANY OTHER PART OF THE HEADER.

2). Again the same general rules we used in the previous labs hold for this lab. Do not delete or modify anything unless it was written by you or the assignment requests it. All R code must be entered in the chunks.

3). To run R code as you work your way through the assignment, highlight the code and click on the "Run" command in the R studio menu and then on "Run Selected Lines" or click on the green button at the top right of a particular chunk to run all the code within it.

4). When you get to the end of the assignment you will knit the document to produce a pdf document using the Knit command in the Rstudio menu.

5). Continue reading from Part A to see what you have to do.

```
knitr::opts_chunk$set(error = TRUE)
```

PART A EXTRACT USA POPULATION DATA FROM 1950 to 2014 We will start by loading the libraries needed for this lab. What the libraries do is give R information on what each command means, sort of like loading a dictionary that R can check commands against so it knows what to do. Run the following code to upload the appropriate packages.

```
library(HMDHFDplus)
```

```
## Warning: package 'HMDHFDplus' was built under R version 4.1.1
```

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.1.1
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.2     v dplyr   1.0.6
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1

## Warning: package 'ggplot2' was built under R version 4.1.1

## Warning: package 'tidyr' was built under R version 4.1.1

## Warning: package 'forcats' was built under R version 4.1.1

## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

In the code below, we import the annual population data for the US from mortality.org and store the information in a variable we call populationUSA. The population comes by specific ages so we aggregate the data for all ages in a year to get the population at the start of the year and the end of the year. We rename the year start population PopJan and the year end population PopDec. After that we extract the data for the years beginning 1950. You'll need to enter in your username and password for mortality.org in the appropriate places to pull the data. Once you've entered in your username and password, you can run the code below. BE VERY CAREFUL NOT TO DELETE OR MODIFY ANYTHING ELSE IN THE CODE.

```
populationUSA <- readHMDweb(CNTRY = 'USA', item ="Population",
username='lysa.vanible@lc.cuny.edu', password='Oceansky2004')
total_pop_by_year_USA <- aggregate(x=populationUSA[4:9], by=list(Year=populationUSA$Year), FUN=sum)
total_pop_by_year_USA <- dplyr::select(total_pop_by_year_USA, Year, Total1, Total2)
total_pop_by_year_USA <- dplyr::rename(total_pop_by_year_USA, PopJan=Total1, PopDec=Total2)
total_pop_by_year_USA<-subset(total_pop_by_year_USA,Year>=1950)
```

PART B CALCULATE POPULATION GROWTH RATES FROM 1950 to 2014

The growth rate describes the percentage increase in population size over a given period. Here, we're looking at annual growth rates since we're looking at the percent change in population from January 1st to December 31st in a given year. To calculate this, we subtract the population in January from the population in December to get the annual absolute change in population, then we divide this by the population in the middle of the year year. We then multiply the result of the calculation by 100 to make it a percentage. You will see that we refer to the variables used in the computation as total_pop_by_year_USA$PopJan and total_pop_by_year_USA$PopDec. The $ sign tells R that the variables, PopJan and PopDec, are stored in the dataset "total_pop_by_year_USA". The new data for the growth rates are stored in "Growth" but in the code we refer to it as "total_pop_by_year_USA$Growth" so that R adds it to the data already contained in "total_pop_by_year_USA". Run the code below to get the annual growth rate of US population since about 1950.

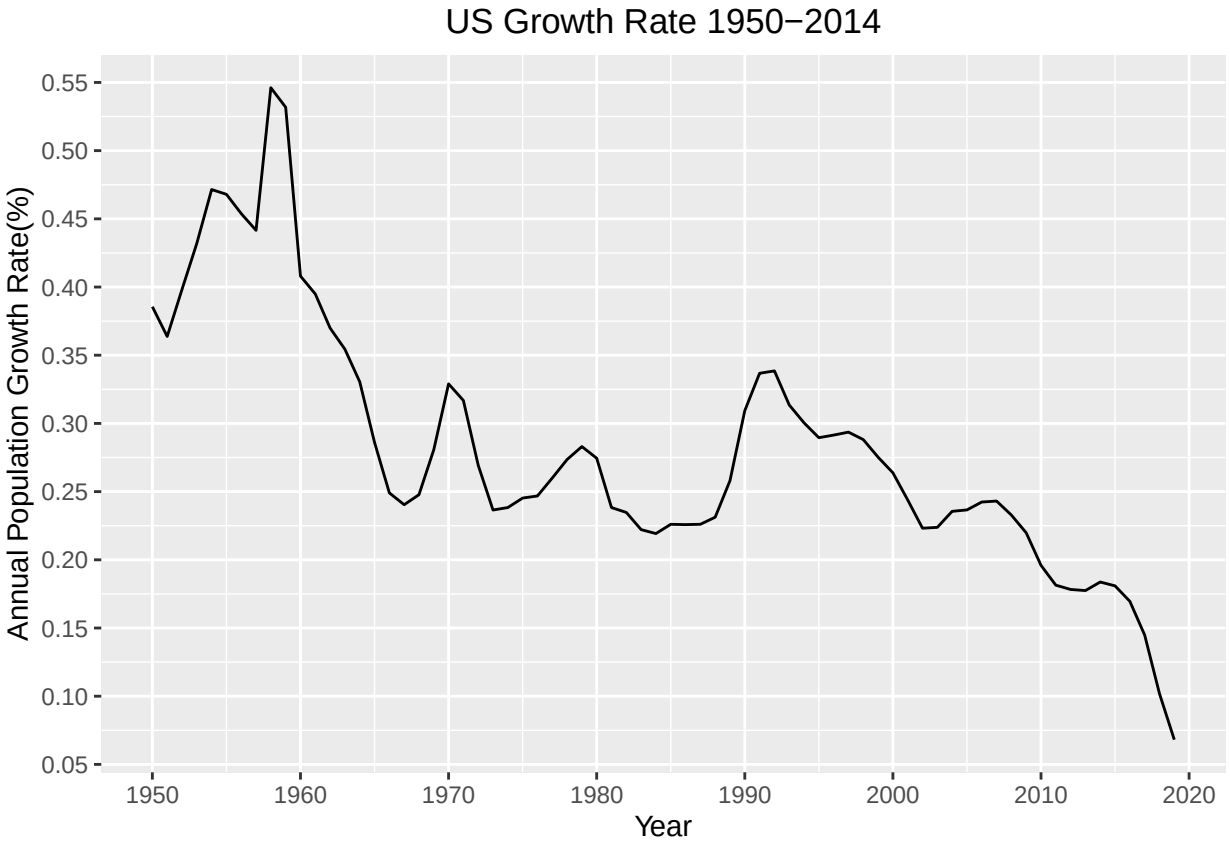
```
#The formula is Growth_Rate = (End Population - Start Population)/(End Population+Start Population)/2*


total_pop_by_year_USA$Growth <- (total_pop_by_year_USA$PopDec
- total_pop_by_year_USA$PopJan) /(total_pop_by_year_USA$PopJan+total_pop_by_year_USA$PopDec)/2
total_pop_by_year_USA$GrowthRate <-total_pop_by_year_USA$Growth*100
```

PART C GRAPH US POPULATION GROWTH RATES FROM 1950 to 2014

Now let's plot the growth rate using ggplot, a very powerful R package used for drawing charts. The code below indicates that ggplot is using data from total_pop_by_year_USA, the data set of US population that was created in the preceeding code chunks. Years will be displayed on the x or horizontal axis, while growth rate will be displayed on the y or vertical axis. Run the code below to see your chart below the chunk.

```
ggplot(total_pop_by_year_USA, aes(x = Year, y = GrowthRate)) +
geom_line() + ggtitle("US Growth Rate 1950-2014") +
scale_x_continuous(breaks = scales::pretty_breaks(n = 10)) +
scale_y_continuous(breaks = scales::pretty_breaks(n = 10))+
ylab(label="Annual Population Growth Rate(%)") +
theme(plot.title = element_text(hjust = 0.5))
```




```
expand_limits(y = 0)
```

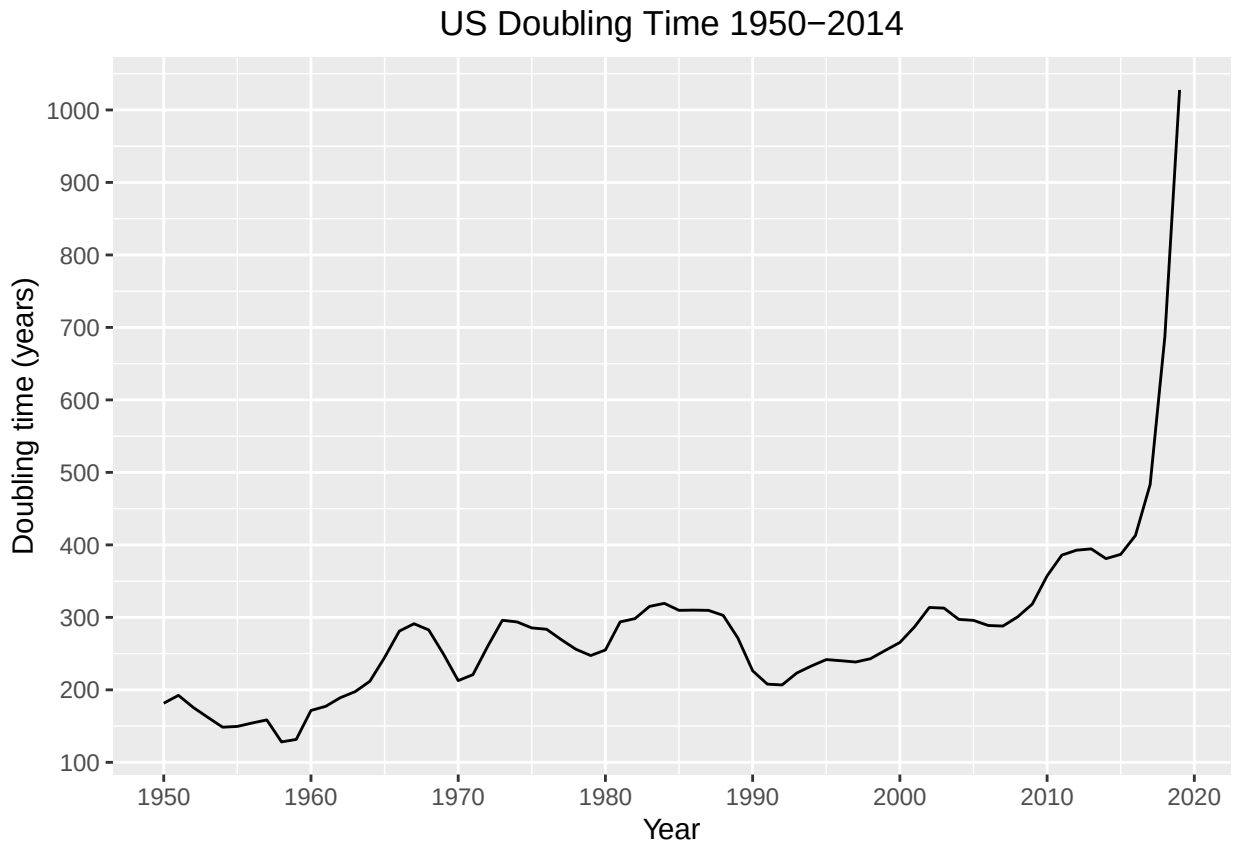
```
## mapping: y = ~y
## geom_blank: na.rm = FALSE
## stat_identity: na.rm = FALSE
## position_identity
```

PART D COMPUTE AND GRAPH US DOUBLING TIME FROM 1950 to 2014 Growth rates are not very intuitive measures, so it can be helpful to study how fast a population is growing by looking at Doubling Time. The formula for computing the doubling time is 70 divided by the growth rate as a percentage. Run the following code to compute the doubling time for the US for the period from 1950 onwards. The information will be placed in the data set "total_pop_by_year_USA" and named DoublingTime.

```
total_pop_by_year_USA$DoublingTime <-70/total_pop_by_year_USA$GrowthRate
```

Now let us graph the doubling time. Run the code below to get ggplot to create the chart below the chunk.

```
ggplot(total_pop_by_year_USA) +
geom_line(mapping=aes(x = Year, y =DoublingTime)) +
ggtitle("US Doubling Time 1950-2014") +
scale_x_continuous(breaks = scales::pretty_breaks(n = 10)) +
scale_y_continuous(breaks = scales::pretty_breaks(n = 10))+
ylab(label="Doubling time (years)") +
theme(plot.title = element_text(hjust = 0.5))
```




```
expand_limits(y = 0)
```

```
## mapping: y = ~y
## geom_blank: na.rm = FALSE
## stat_identity: na.rm = FALSE
## position_identity
```

PART E CONCLUSION

Great work!!!! You are almost at the end of your second lab. You will now prepare to knit the lab. Click on the "Knit" command at the top of this window to turn this document with your code and all the results into a pdf document. When the knit is complete, Rstudio will open the pdf document. Inspect it very carefully to

make sure that you have not missed any part of the assignment. If you have missed any tasks please go back and fix them and knit the document again. Once the pdf document is knitted and you are sure you have completed all the coding assignments in lab 2 prepare to submit it to the week 3 of SOC339's blackboard site. Before submission there are a few more things for you to do.

1). First save your rmarkdown file (the file you have been working on in Rstudio) by clicking on the FILE command in the Rstudio menu and then on SAVE. You will see the rmarkdown document in the FILES folder of the window to the bottom right of your screen. It will be named Lab 2 Population Growth.Rmd.

2). To save the pdf file, locate it in the FILES folder of the window to the bottom right of your screen. It will be named Lab 2 Population Growth.pdf. Click the little box to the left of it. Then click on the MORE tab near the top of the FILES window. Then on EXPORT and then when prompted by RSTUDIO click on DOWNLOAD. RSTUDIO will save the lab 2 Population Growth.pdf to the DOWNLOAD folder of your computer.

3).Go to the blackboard Week 3 folder and follow the instructions to complete lab 2 and upload the Lab 2 Population Growth.pdf.