

UCLA Extension Data Science Intensive

Instructor: William Yu

Project 1

Updated 9/28/2018

- Submit your results (including R script and any output you got) through Canvas.
[Create a new R script in RStudio: File -> New File -> R Script](#)

A. Using R to Write a Formula to Calculate the Mortgage Payment

- Taylor wanted to buy a house with a price \$582,000. She could only afford the down payment 20% of the price, which is \$116,400. Therefore she went to the bank to ask for a 30-years mortgage loan for \$465,600. The banker told her the current (annual) fixed mortgage rate is 4.5%. The amortized mortgages have the following formula to show Taylor's future monthly repayment for this mortgage.
 - x is the total amount of loan (in the case, x=465,600)
 - z is the mortgage interest rate (in the case, z=0.045)
 - y is the years of the loan (in the case, y=30)
 - c is the monthly mortgage rate (in the case, c=z/12)
 - n is the total months of the loan (in the case, n=y*12)
 - p is the monthly repayment for Taylor

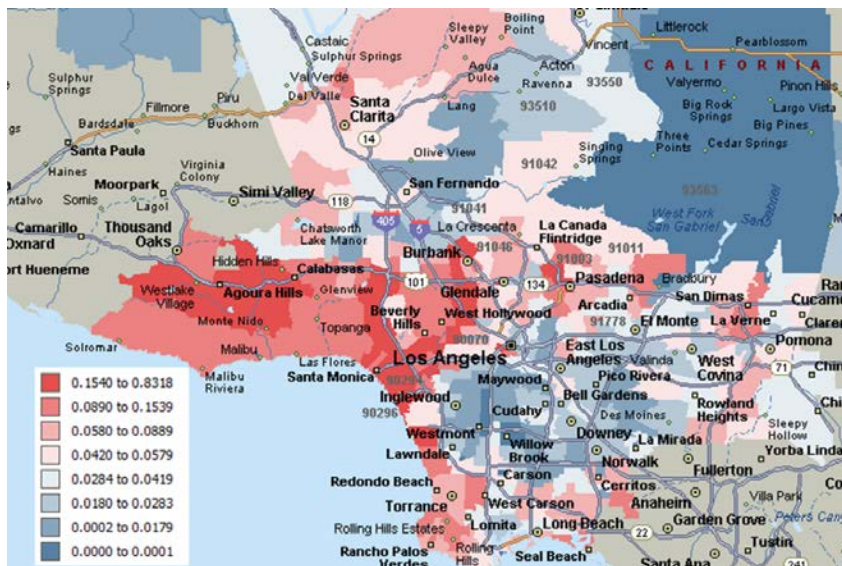
$$p = \frac{x \cdot [c \cdot (1 + c)^n]}{(1 + c)^n - 1}$$

- How much will be the monthly payment for Taylor?
- Use RStudio script to write a simple code to show the result for Taylor.
- [Note: R can only read data x=465600 instead of 465,600.]

B. Calculate the Density of Tech Job by Zip Code in Los Angeles

- Read my report: "The Tech Industry in California and Los Angeles," for Anderson Forecast Q3 Economic Outlook as well as my presentation slides.
- We are going to calculate the data I use for Figure 7 (as shown below) in the report or Slides page #9. We will learn to plot the data in the map for the future project in coming weeks.
- Download the data (P01_LA zipcode payroll.xlsx) into your computer.
- You need to produce the data output in which to show the payroll employment for the total industry, the information sector, and the professional, scientific, & technical skills sector aligned by zip code in 2017.
- And then calculate the percentage of tech job= (information jobs + professional scientific technical jobs) / total jobs.

- The expected output (laz17tech.csv) is attached in the Project 1 folder.
- Reminder and Hint:
 - There might be a lot of data cleaning/management to do this project. For instance:
 - Replacing NA with some value.
 - Remove “Total” in Zip Code column.
 - Replace ***** with 0
 - Covert Column 5 and 6 from character to Numeric.
 - E.g. `laz2017[,c(1,5:6)]=sapply(laz2017[,c(1,5:6)], as.numeric)`
 - Use “?sapply” to check what it means.
 - Install “dplyr” package and use some function such as `left_join` in order to line up the zip code.
 - You may need to use functions such as “`subset`” and “`gsub`”.
 - If it doesn’t work in the beginning, try to convert the data you imported to data.frame. Such as:
`laz2017 <- data.frame(read_excel("P01_LA zipcode payroll.xlsx", sheet="2017"))`
 - Don’t get frustrated if it doesn’t work! Almost many things will not work for everyone.
 - Of course you can do some data cleaning and management in Excel first. But I hope you can do these preprocessing tasks directly in R.



C. Bonus project

If you can modify D01b_hs R Script to make the Excel/CSV output more appealing and closer to a desirable one (such as W01d_hs1.all.xlsx), you will get a bonus.