

UCLA Extension Data Science Intensive

Instructor: William Yu

Project 1

- Submit your results (including R script and a summary excel file) 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.***

p is the total amount of loan (in the case, $x=582,000$)

d is the percentage of down-payment (in the case, $d=0.2$)

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, $n=y*12$)

m is the monthly repayment for Taylor

$$m = \frac{[p \times (1-d)] \times [c \times (1+c)^n]}{(1+c)^n - 1}$$

- How much will be the monthly re-payment for Taylor?
- Use RStudio script to write a simple code to show the result for Taylor.
- Note: [R can only read data x=582000 instead of 582,000](#). Put p , d , z , y at the top 4 lines.

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.
- 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 (per) = (information employment + professional scientific technical employment) / total employment.
- The expected output (laz17tech.csv) is attached in the Project 1 folder.
- Use Tableau Public to plot the pec (and save it into your excel output file). Chart 1 is the plot I got from a different commercial package. Chart 2 is the plot I got through Tableau.
- Reminder and Hint:
 - There are a lot of data cleaning/management to do for this project because the data provider (from CA employment development department) did not make the data so user friendly (**most variables are strings/characters, not numeric!!**). For instance, you will need to do:
 - To make it run smoothly, try to convert the data you imported to data.frame. Such as: `laz2017 <- data.frame(read_excel("P01_LA zipcode payroll.xlsx", sheet="2017"))`
 - Replacing NA with some value so you can use it to filter the data. NA are at those rows in which the industry and NAICS' names are absent.
 - 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 some function such as "merge" in order to line up the zip code.
 - You may need to use functions such as "subset" and "gsub".
 - 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.

Chart 1.

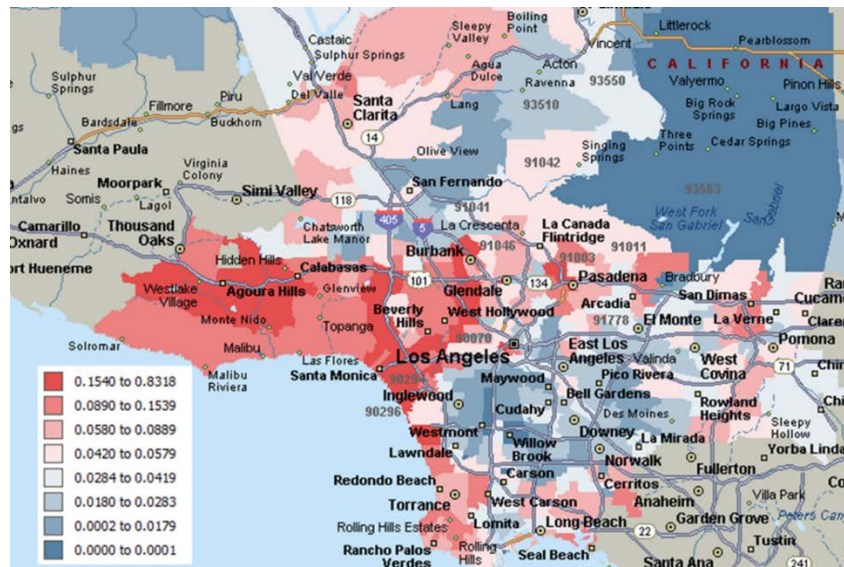


Chart 2.

