Data Science Principles and Practice
CS x415.1, Fall 2017
Assignment #3
Due: **Tuesday, April 4, 2017 11:59 PM** (no late submissions accepted)
Submit on Canvas / onlinelearning.berkeley.edu (no email submissions accepted)

Problem 1

The U.S. Government Department of Labor, Bureau of Labor Statistics (BLS) publishes a Local Area Unemployent Report, overview available here:

https://www.bls.gov/lau/lauov.htm

Data format description here:

https://www.bls.gov/help/hlpforma.htm#LA

A file named **sf_oak_metro.csv** in the Canvas Files, Assignment 3 folder has local area unemployment statistics for the San Francisco-Oakland-Hayward California Metropolitan Statistical Area (MSA) for the years 1990 - 2016.

Assignment:

From the dataset above, use tidyverse packages to generate an HTML file from RMarkdown with **four separate** ggplot line plots (including points) of the following:

1. Employment (numbers)
2. Labor Force (numbers)
3. Unemploment Rate (rate)
4. Unemployment (numbers)

with:

- The above values on the Y axis (may be labeled simply "value")

- Month abbreviations (Jan, Feb, ... Dec) on the X axis

- Showing the values for the years 1990 - 2016 on each plot indicated by color (include a legend showing color for each year)

- Each plot identified with a title bar or title caption *similar* to the above wording, for example, "Employment: San Francisco-Oakland-Hayward CA Metropolitical Statistical Area"

- A brief description of the problem from the above so that a reader can understand your report.

Problem 2

The file **assignment2_problem2_data_files.zip** contains two files:

gz_2010_us_050_00_20m.json: a geojson spatial polygons (sp) file file of USA counties

california_counties_monthly_employment_2016.tsv: a tsv file with monthly employment statistics for all California counties for 2016

MVP (Minimum Viable Product, for a pass credit grade)

1. A [leaflet choropleth map](#) showing the unemployment rates (only, with legend) for all California counties for December 2016 in your HTML report, generated from RMarkdown
2. A Shiny app that allows the user to choose the month (in a dropdown) for which to display the above choropleth map

FP (Full Product)

1. The same HTML generated from RMarkdown as (1) above
2. Add the following additional features to the above Shiny app :
   1. Use Shiny modules in your solution
   2. The user should be able to choose which month and employment statistic to display: labor force, employed, unemployed, and unemployment rate
   3. County polygons should diplay the label area title: statistic when the cursor hovers over any county
   4. Other map/Shiny app enhancements? For example: use tabs to switch between plot and table presentation of results
   5. Use this application to demonstrate your mastery of the topics we've covered in class!

Note: indicate whether you have opted for MVP or Full Product - **if you submit the FP above, do NOT submit the MVP**. **One** HTML and **one** Shiny app only

Mandtory Implementation Notes

1. Deploy your Shiny app (one only) on shinyapps.io - you will need to sign up to create an account if you do not alread have one (free accounts are available)
2. **Be sure to include and clearly indicate the URL of your Shiny app in your HTML file**
3. Assume the data files are on the same directory as your .Rmd file (don't use setwd() in your code). Your HTML file results should be 100% reproducible: I should be able to Knit it to HTML without modification (of course once the working directory is set by menu option choices within RStudio)

4. Include in your results the as much problem description text as needed for the reader to understand what you are doing

5. In the YAML section of your .Rmd file, include title "Programming Assignment 3," author (your name and email on the same line), and date. Of course, output should be HTML

6. Use the following packages in your solutions: tidyverse, leaflet, maps, geojsonio, spdplyr

7. This is primarily a *learning exercise*. Use the concepts, topics, tools, and approaches discussed in class only in your solution. Clever solutions using other packages, methods, etc., will receive **no** credit (ask if you have any questions about this)

8. Get an early start, no late or email assignments accepted

9. Use fix_column_names from Assignment 2 to convert all data frame column names to snake case (but don't submit this), follow the style guidelines described by Hadley Wickham in this R style guide

Hand In

Hand in three files total (both problems 1 and 2 above in the same file):

1. Your RMarkdown file: asn3_firstname_lastname.Rmd
2. Knitted HTML with source: asn3_firstname_lastname_with_source.html (with source code, i.e. chunk option echo = TRUE)

3. Knitted HTML with no source: asn3_firstname_lastname.html (with no source code, i.e. chunk option echo = FALSE)

4. Embed your shiny app in your RMarkdown

Grading

400 points total. You will be graded on:

- Completeness

- Follows specifications (especially the mandatory specificcations above!)

- Implementation quality (code quality, adherence to style guidelines)

Remember: produce a knitted HTML product that you could hand to your Data Science Manager with the confidence that she would understand, and could reproduce, your report!