Assignment 4 - ESM 244 (Winter 2021)

Tasks 1, 2 & 3 due Monday, March 8th, 5pm PST (submit on GauchoSpace).
Complete your personal website by March 15th.
Complete and prepare a presentation for your Shiny app by your sign-up session.

Task 1: Agglomerative hierarchical clustering (submit html on GS)

For Task 1, use hierarchical clustering by complete linkage to create a dendrogram showing multivariate clustering for water chemistry by ***site***, saved as **lastname_firstname_a4_task1.html**.

To perform hierarchical clustering by **site**, you'll want to make a data frame that has a single summary row per site (e.g. based on means from all observations at that site), *then* calculate the euclidean distance before performing complete linkage agglomerative hierarchical clustering.

Link to the data:
https://drive.google.com/file/d/16rYLBi-CgvjcWhOsX1SLdD9HHUMP9m2l/view?usp=sharing

DATA & METADATA SOURCE:
SBC LTER: Stream chemistry in the Santa Barbara Coastal drainage area, ongoing since 2000
Creators: Santa Barbara Coastal LTER, & Melack, John M
Citation: Santa Barbara Coastal LTER and J. Melack. 2019. SBC LTER: Land: Stream chemistry in the Santa Barbara Coastal drainage area, ongoing since 2000 ver 16. Environmental Data Initiative. https://doi.org/10.6073/pasta/67a558a24ceed9a0a5bf5e46ab841174.

NOTES: Make sure you convert -999 values to NA. If you have a dataset that you are *more interested in exploring with hierarchical clustering*, you are welcome to use an alternative dataset of your choosing instead.

Task 2: Parameter Estimation – Wild Fish Catch (submit html on GS)

Source: Global wild fish catch and aquaculture production, compiled by Earth Policy Institute with 1950-2010 from U.N. Food and Agriculture Organization (FAO), *Global Capture Production* and *Global Aquaculture Production*, electronic databases, at www.fao.org/fishery/topic/16140/en.

Get the data: **fish_catch.csv**

For Task 2, you will find an equation with parameters estimated by nonlinear least squares for the increase in global **wild fish catch** from 1950 – 2012. ***Hint: You will want to set 1950 = 0 (i.e., create a new column with years starting at 0, instead of value 1950, and use those values for your model…).***

For Task 2:

Prepare a knitted .html saved as **lastname_firstname_a4_task2.html** that contains:

a) An exploratory graph of wild catch over time (does not need to be finalized). Include the exploratory graph in your knitted HTML.
b) In text below the exploratory graph: What type of relationship describes the trend? What does that look like mathematically (include an equation, possibly using LaTeX)? What are your initial estimates for the parameters in the model?
c) Use nonlinear least squares to find parameters for your model describing wild catch. Report the parameter outcomes (with units) in text or a table.
d) Prepare a **finalized** (publication quality) graph showing both the original data *and* your model output. No figure caption required. The code and graphs should appear in your knitted HTML (OK to use code folding).

Task 3: Bootstrapped Confidence Interval for Proportions (submit html on GS)

The following data are from the 2014 UCSB Campus Climate Project Final Report (prepared by Rankin & Associates Consulting, available at *http://campusclimate.ucop.edu/_common/files/pdf-climate/ucsb-full-report.pdf)*.

In the study, 22 out of 36 surveyed UCSB community members (61%) identifying as nonbinary/genderqueer responded that they had personally experienced “exclusionary, offensive, hostile or intimidating conduct” (compared to 26% and 19% for those identifying as women and men, respectively).

Your goal is to find a confidence interval for the proportion of nonbinary/genderqueer students experiencing exclusionary, hostile or intimidating experience at UCSB using bootstrapping. Create a vector reflecting the collected survey data (n = 36), then find the bootstrapped 95% confidence interval for the *proportion* of genderqueer individuals experiencing exclusionary conduct based on the 2014 UCSB survey data.

Hint: *this task will require you to recreate the data* ***and*** *create a function that calculates the proportion of a specified outcome for each bootstrap sample. The proportions for each bootstrap sample are what will be plotted in your histogram below.*

For Task 3:
In your nicely organized .html saved as **lastname_firstname_a4_task3.html**:

a) Include any code you wrote to make the original sample vector, create the “proportions function,” and find bootstrap samples (use at least 10,000 bootstrap samples here).
b) A histogram of bootstrapped proportions (does not need to be finalized). The histogram should show up in your knitted HTML.
c) A final statement (in text) describing the bootstrapped CI in the context of the data (suitable for publication).

Task 4: Keep finalizing your personal website (by Monday 3/15)

We’re excited to see your personalized website. We will start checking your links to your sites on Monday 3/15, so make sure to finalize by then.

Task 5: Prepare to present your Shiny app (by your session)

Here is the rubric for your Shiny app term project. Each group should prepare a ~10 minute presentation on your app that you will give during the window that you’ve signed up for, leaving several minutes for Q&A from the group. You are expected to share your screen to demo your app.

In your presentation, you should:

- Describe the purpose of the app
- Walk through each tab
- Demonstrate the widgets and reactive outputs
- Describe a challenge you had in building the app
- Describe remaining problems/what you’d like to include but couldn’t
- Each member should share what you’re most proud of