ESM 206 Fall 2020 Assignment 6

Task 1: Reading & discussion
Task 2: Chi-square
Task 3: Two-sample t-test
Task 4: Data visualization

Due Tuesday 2020-12-08 by 5pm PST

Follow instructions carefully.
You will complete Tasks 2, 3 & 4 each in a separate .Rmd, but all **within in a single private version-controlled R project,** then add Jessica (couture322) as a collaborator to your private GitHub repo for grading.

TASK 1: Read and respond to Mah (2017) *Environmental justice in the age of big data*

Post your response to the course Slack channel by **5pm PST on Friday 12/4** to give your classmates the opportunity to read and respond to your post before the Tuesday due date.

For Task 1, read and respond to Dr. Alice Mah's 2017 paper *Environmental justice in the age of big data: challenging toxic blind spots of voice, speed, and expertise*. (...if this sounds familiar to some of you, it was discussed at a previous TidyTuesday). **In 5 - 6 sentences,** share your response to Dr. Mah's paper in the ESM 206 #ethics-and-bias channel on Slack. You can base your response on the questions below, generally discuss insights you gained from the paper, consider the challenges in the context of a real example in your community, or come up with your own topics & questions. Then **respond to at least 2 of your classmates' posts in 3 - 4 sentences each by the assignment due date.**

Here are some questions to consider (you're also encouraged to come up with your own questions / prompts to discuss):

- Which challenges described in the paper are specific to "big data", and which do you think apply to *all* environmental data?
- Dr. Mah states: "...most toxic exposure cases are slow-burning and unspectacular, beyond the scales and frames of analysis of fast-moving media or science." Brainstorm real-world examples of these "slow-burning" environmental injustices. What are some examples of "slow-burning" environmental injustices near where you live? Who is impacted?
- What opportunities and challenges do you expect exist for early career environmental scientists trying to address or overcome these "toxic blind spots" in environmental data justice?

TASK 2: Survey responses - political and environmentalist self-identification

Data for Task 2 are from a survey conducted by Chelsea Batavia and Andrews Forest LTER to explore "Influences on charitable giving for conservation." The survey was conducted in August 2017 and yielded 1,331 individual respondents.

Data citation: Andrews Forest LTER Site and C. Batavia. 2019. Influences on charitable giving for conservation: Online survey data of 1,331 respondents across the US, August 2017 ver 3. Environmental Data Initiative. https://doi.org/10.6073/pasta/926e6270e324a1322a900da14d38b96c

Metadata, including survey questions and possible responses, are here: https://portal.edirepository.org/nis/metadataviewer?packageid=knb-lter-and.5444.3

For Task 2, you are asked: To describe the association between respondents' political affiliation (binned "conservative" or "liberal" by self-identification) and if they consider themself an environmentalist. The variables in the dataset (conservation_survey.csv) you will use are:

- **ENV_ID:** Response to survey statement "I consider myself an environmentalist," with outcomes 1 = Yes, 2 = No, 3 = Unsure
- **POL_OR:** Response to survey question "How do you self-identify politically?" Response values 1 - 3 are "conservative" (strongly to slightly), and values 5 - 7 are "liberal" (slightly to strongly)

a. Create a new private repo on GitHub, cloned to create a new version-controlled R project (title the repo & project ***assignment6_firstname_lastname***, e.g. *assignment6_allison_horst*)

b. Create a data/ folder within the project, and copy the **conservation_survey.csv** file (download from GauchoSpace) into it

c. Create a src/ folder in the project. Create a new RMarkdown document, save in src/ with file name **a6_task2_firstname_lastname.Rmd** (e.g. a6_task2_allison_horst.Rmd)

d. In your R Markdown document for Task 2, write organized code to:

   i. Create a simplified version of the survey data to only keep the variables you need for this task, `ENV_ID` and `POL_OR`

   ii. To your simplified data frame, add a new column that bins the political party into a binary response of either "conservative" (if POL_OR is 1 - 3) or "liberal" (if POL_OR 5 - 7). All other political affiliations (4, 8, 9) should be excluded for this task. Hint: you might want to use dplyr::mutate() + dplyr::case_when() here.

iii. Create a finalized table that shows the **counts** and **proportions** of "liberal" and "conservative" survey respondents who responded "Yes", "No" and "Unsure" to the statement "I consider myself an environmentalist." Hint: see the Week 9 Coding Lesson key for hints on how to do this. **This table should appear in the knitted document, with a table caption above it.**

Basically the overall (unfinished) table structure will be something like the table below, but finalized and containing the correct counts in each cell:

	Yes	**No**	**Unsure**
Conservative			
Liberal			

iv. Then, write code to convert the counts table to a contingency table (hint: column_to_rownames()) and perform a chi-square test for independence to answer: is there an association between political affiliation (conservative or liberal-leaning) and if a person considers themself an environmentalist?

v. In 2-3 sentences, using in-line referencing to the chi-square test results, describe the results of the statistical test *and* briefly highlight major takeaways that your reader should learn from your analysis (i.e., you should refer to the actual proportions here).

What should show up in your knitted HTML for Task 2?

- A brief descriptive title up top
- Your name
- Add a 2 - 3 sentence mini-introduction describing the survey data, the question you're answering, and the method you'll use
- Your finalized table of counts and proportions (with a table caption above)
- Your 2 - 3 sentence summary of the hypothesis test results and major takeaways
- Add the data citation at the end
- **THAT'S IT (no code, warnings, messages, additional info)**

e. Make sure you stage, commit, pull & push your .Rmd and html to your repo for safekeeping. Move on to Task 3…

TASK 3: Lizard lengths

For Task 3, you will be using the Jornada Basin LTER lizards data (lizards.csv, on GauchoSpace) that was used in the chi-square coding lesson in Week 9.

Data source: Lightfoot, D. 2017. Lizard pitfall trap data (LTER-II, LTER-III) ver 36. Environmental Data Initiative. https://doi.org/10.6073/pasta/ff37b4cace16a9943575f3fd7067064e

Metadata is here:
https://portal.edirepository.org/nis/metadataviewer?packageid=knb-lter-jrn.210007001.36

a. Copy the lizards.csv file into the data/ folder of your existing Assignment 6 project
b. Create a new RMarkdown document in the project, and save in src/ with file name **a6_task3_firstname_lastname.Rmd** (e.g. a6_task3_allison_horst.Rmd)
c. For Task 3, you are asked to compare total lengths (variable = **total_length**, in millimeters) for **female lizards** (sex == "F") of two species (variable **spp**): CNUN (*Cnemidophorus uniparens*, the Desert Grassland lizard) and CNTE (*Cnemidophorus tessalatus*, the Colorado Checkered lizard).

   Write code to:

   i. Create and store a simplified version of the data frame that only includes variables ***spp, sex,*** and ***total_length***
   ii. Filter to only include observations for female lizards of species CNUN and CNTE
   iii. Create a finalized visualization comparing total lengths of female lizards for the two species. Include a figure caption below it.
   iv. Prepare a finalized summary table that contains the mean and standard deviation for total length, and sample size, for each group (female CNUN and CNTE lizards)
   v. Answer: is there a significant difference in total length difference in mean total length between female lizards for the two species? Note: you **should do exploratory analyses (e.g. histograms, QQ-plots) to determine if assumptions of a hypothesis test you decide on are appropriate, and include code used to do that, even though it won't show up in your knitted document.**
   vi. Find the effect size (Cohen's *d*) of the difference between means for the two species.

   What should show up in your knitted HTML for Task 3?

   - A brief descriptive title up top
   - Your name
   - Add a 2 - 3 sentence mini-introduction describing the data, the question you're answering, and the method you'll use
   - Your finalized figure, with a figure caption below
   - Your finalized summary table, with a table caption above
   - A 1 - 2 sentence summary statement describing your findings, that at least includes: the results of your hypothesis test, the effect size, and the actual means difference between the two groups
   - Add the data citation at the end

- **THAT'S IT (no code, warnings, messages, additional info)**

f. Make sure you stage, commit, pull & push your .Rmd and html to your repo for safekeeping. Move on to Task 4…

TASK 4: Beautiful data visualization, Mono Lake Levels

The only thing you'll do for this task is create the most perfect, useful finalized data visualization you can. It should be highly customized.

Some background: Mono Lake is a terminal saline lake in Lee Vining, California, northeast of Yosemite National Park in the Pamidu Toiyabe (Sierra Nevada), and is home to the Kootzaduka'a (or Kutzadika'a) Tribe - learn more about the Kootzaduka'a Paiute here.

Read about the recent history of Mono Lake here (lots of other useful information on the Mono Lake Committee website), including major events and legal decisions that have impacted the lake levels. The '**mono_lake_annual_levels.csv'** file on GauchoSpace contains data for the measured lake level (feet above sea level), volume, surface area, and annual change from the Mono Basin Clearinghouse.

For Task 4:

a. Copy the mono_lake_annual_levels.csv file into the data/ folder of your existing Assignment 6 project
b. Create a new RMarkdown document in the project, and save in src/ with file name **a6_task4_firstname_lastname.Rmd** (e.g. a6_task4_allison_horst.Rmd)
c. Read in the data, and simplify if needed (remember you can *skip* rows when using read_csv…)
d. Make a clear, professional, beautiful data visualization showing the change in Mono Lake levels (feet above sea level) for all years included in the data. Add a figure caption below it.

You should carefully consider and finalize every aspect of the graph. Would it be useful to add annotation for important events in Mono Lake's recent history? Or critical lake levels? Carefully think about and update every detail of your graph to correctly, clearly, responsibly, and beautifully communicate the recent history of Mono Lake levels.

Some important water levels you might consider indicating on your data viz:

- At **6,377 feet above sea level**, land bridges emerge between the mainland and islands, allowing predators (e.g. coyotes) to access major bird nesting sites
- From *The Mono Basin Ecosystem: Effects of Changing Lake Level*, published in 1987 by the Mono Basin Ecosystem Study Committee, Board on Environmental Studies and Toxicology (NRC): "The critical food resources for aquatic birds using Mono Lake are

brine shrimp and brine flies. If the lake fell to levels at which the birds' food sources were adversely affected, the bird populations would be reduced. The decrease in availability of brine shrimp for food would begin to affect those birds relying on them – eared grebes and California gulls – at a salinity of 120 g/L (lake level of **6360 ft**). The impacts would be acute at salinities above 150 g/L (**6350 ft**)."

The only things that should show up in your knitted HTML for Task 4 are:

- A brief descriptive title
- Your name
- Your beautiful data visualization
- A professional figure caption
- **THAT'S IT**

ASSIGNMENT 6 NOTES & GRADING:

Tasks 2, 3, and 4 each be in an individual R Markdown document, and should all be contained in a **single private version controlled R Project**. You will add Jessica (couture322) as a collaborator to your repo so she can clone & run your code to check.

- The assignment is worth 30 points total. Each Task (2, 3 & 4) is worth 10 points.
- You will be graded on the specific components requested in the knitted HTML for each task. For each, we will ask:
    - Are results correct?
    - Are the requested results (graphs, tables, text statements) professionally and clearly presented when knitted to HTML?
    - Is the code well-organized and reproducible? (i.e., will Jessica be able to clone and run your .Rmds without issue?)

If you can answer **yes** to the above for each task, you should feel good about your submissions.

DONE? CONGRATULATIONS, YOU'RE DONE WITH YOUR LAST 206 ASSIGNMENT.