ESM 244 Assignment 3

Due 2pm on Tuesday 3/5. Prepare your individual efforts for Tasks 1, 2, 3 and 4 in a single well-formatted Rmarkdown document. Submit the **final knitted HTML and your .Rmd** to the GauchoSpace link. **Include all code (but not messages/warnings/etc.) in your knitted HTML.** We want to see the code and the outputs requested, but nothing else.

Task 1: Open science perspectives

Open science practices are emerging, evolving and imperfect.

For Task 1, read and think critically about two publications re: open science and open practices:

1. McKiernan, E.C. et al. (2016). *How open science helps researchers succeed*. eLife. 2016; 5: e16800. doi: 10.7554/eLife.16800. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4973366/

2. Bahlai, C. et al (2018). *Open Science Isn't Always Open to All Scientists.* American Scientist. https://www.americanscientist.org/article/open-science-isnt-always-open-to-all-scientists

In ~ 600 - 800 words, prepare a brief, well-written statement:

- What are some important goals of "open science" and how do they benefit researchers and environmental scientists?
- What are some problems with how open science is currently approached regarding inclusiveness, equity and accessibility?
- Describe the tools, skills and approaches you are using to make your work more reproducible and open (at a level that makes sense for you - e.g. within research groups, with your adviser, sharing with the public, etc.)
- Describe limitations of your work or workflow regarding openness, inclusiveness and accessibility. What are some ways that you could overcome those limitations moving forward?

Task 2: Truckee River flow (2000 – 2016)

For Task 2, use monthly averaged discharge (cfs) data for the Truckee River (measured near Truckee, CA) for **January 2000 – September 2016**, accessed from the USGS Water Data for the Nation website and compiled into the file 'truckee_flow.xls' on GauchoSpace.

Explore and simplify the dataset before loading into your .Rmd.

For Part 2, you are asked to explore trends and forecasting in the time-series data.

a. Create a graph with the **decomposed** time series information (original, trend, seasonal, and residual). You do not need to further finalize this graph. You do not need to add a figure caption.

In text below the decomposed graphs, generally describe the time series data in 2 - 3 sentences. For example, does it appear additive or multiplicative? Stationary or non-stationary? Are there outliers? Is there seasonality? Does there appear to be a cyclical trend *beyond* the normal seasonality? Etc.

b. Use either Holt-Winters **or** ARIMA (your choice – or you can do both and compare) to forecast the Truckee River for 5 years after the final observation in the dataset. Visualize the forecasted data in a graph (you don't need to finalize this in ggplot) with updated axis labels and a figure caption.

c. For whichever model (Holt-Winters or ARIMA) you chose in (b), visually assess the model residuals.

For Task 2, your knitted html should include:

- All code (exclude any messages/warnings) for Task 2
- The plot of the decomposed time series (unfinalized)
- Below the decomposed graph, a 2 – 3 sentence description of the time series data/patterns/trends
- The unfinalized graph with 5-year forecasted Truckee discharge based on your Holt-Winters or ARIMA model (no figure caption needed)
- The graphs you used to visualize model residuals (no figure caption needed)

Task 3: Mapping California's National Parks

Using shapefile data for California counties (from Lab 6, or pull your own) and National Parks shapefile data (nps_boundary) provided on GauchoSpace, create a beautiful finalized map of **California's National Parks**. Add text labels for the parks. You will need to learn how to add map labels on your own, but here's a hint for one approach (of many): ggrepel functions for text labels want x/y coordinates. See sf::st_centroid and sf:: st_coordinates.

For Task 3, your knitted html should include:

- Any code to read, manipulate and map the spatial data
- Your final map with California's National Parks shown and labeled (no caption required)

Task 4: Lizards in the Northern Chihuahuan Desert – Data Wrangling and Analysis

Download the dataset 'lter_lizard_pitfall.csv' from GauchoSpace. For more information and metadata, visit: https://portal.lternet.edu/nis/metadataviewer?packageid=knb-lter-jrn.210007001.36.

Study summary (from the authors): *"In conjunction with net primary production studies, consumer and faunal studies are conducted at or near NPP sites using pitfall traps. We use live traps, not employing ethylene glycol or other killing/preservative agents, with traps checked once a week at the minimum. Sampling-with-replacement is used with the lizards. Variables measured include species, sex, recapture status, snout-vent length, total length, weight, and whether tail is broken or whole."*

Citation: Lightfoot D. 1991. Lizard pitfall trap data (LTER-II, LTER-III). Environmental Data Initiative. http://dx.doi.org/10.6073/pasta/ff37b4cace16a9943575f3fd7067064e. Dataset accessed 2/26/2018.

For Task 4, you are asked to answer the following:

1. For all lizards trapped at site 'CALI', do weights of male and female adult lizards differ significantly?
2. For lizards trapped at the 'CALI' site, is there a significant difference in the proportion of adult male and female lizards with broken tails?

For Task 4, your knitted html should include:

- Any code to read, manipulate and analyze the lizard data
- A final 1-sentence statement answering each question, suitable for publication