Assignment 1 - ESM 244 (Winter 2020)

All parts due by 2:00pm on Thursday 2019-02-06

Data wrangling & viz, principal components analysis, and outlining your Shiny app

- **Tasks 1 & 2:** Submit **individual** knitted HTML for Tasks 1 & 2 through GauchoSpace
- **Task 3:** Add your Shiny app information here (**one entry per group**)
- **Task 4:** Create a Shiny app visual outline and submit **one hard copy per group**

Task 1 (individual): Data wrangling & visualization (showshoe hares)

For Task 1 you will wrangle, then create a final visualization and summary table, of physical characteristics for snowshoe hares observed in Bonanza Creek Experimental Forest.

Data: download showshoe_data.csv here
Information & metadata: click here
Study citation: Kielland K., F. S. Chapin, R. W. Ruess. 2017. Snowshoe hare physical data in Bonanza Creek Experimental Forest: 1999-Present. Environmental Data Initiative. https://doi.org/10.6073/pasta/03dce4856d79b91557d8e6ce2cbcdc14.

You get to decide what variables and summary statistics to visualize. Here's what you should prepare: ***A single polished HTML (knitted from .Rmd), planning that this might be a post/project you'd include on your personal blogdown site, that includes at least:***

1. A useful descriptive introductory summary (3 - 4 sentences) of what's contained in the project
2. Image(s), with captions and necessary citation(s), of showshoe hares and/or a map of the study area
3. All of your organized and well-annotated code (with warnings/messages hidden) used to create **at least**:

   - One finalized graph about the Bonanza Creek snowshoe hare population (you pick which variables, how you want to wrangle it beforehand, and which type of visual to create - but make sure it is beautifully finalized)
   - One finalized HTML table (probably created using kable & kableExtra) containing summary statistics about the snowshoe hares (again, you pick which variables, and how you want to group/summarize them)

4. Make sure that both your figure and table appear in your final knitted document, each with a useful caption. Include text associated with each to help the audience understand and interpret the results.

Task 2 (individual): Principal components analysis

For this task, I'll provide a couple of datasets that you *can* use for PCA exploration, but you are also welcome to find/choose a **different** dataset to use. You only need to use **one** dataset to practice running PCA, creating a biplot, and interpreting the results. Whichever dataset you choose, create a finalized HTML (knitted from .Rmd) that can be included as a project on your blogdown site including:

1. A useful descriptive introductory summary (3 - 4 sentences) that helps the audience understand the data (include a citation as necessary) and what you'll be exploring
2. All of your organized and well-annotated code (with warnings/messages hidden) used to wrangle then run PCA, then create a **finalized** PCA biplot that appears (with a caption) in your knitted HTML
3. A summary (bulletpoints are fine) pointing out some major takeaways from your PCA that can be gleaned from the biplot
4. You are also welcome to include a photo in the HTML to add some visual interest (not required for Task 2).

Here are two datasets that you **can** work with, either in whole or in part (i.e. you are welcome to wrangle in order to limit observations and/or variables, just make sure to describe how you are limiting things in your project), or feel free to choose your own:

- *Miscellaneous environmental and climatic variables (country-level)*
  - The file: world_env_vars.csv
  - Compiled and provided by @zander_venter on Kaggle, described as: "This data is all acquired through Google Earth Engine (https://earthengine.google.com/) where publicly available remote sensing datasets have been uploaded...Most of the data is derived by calculating the mean for each country at a reduction scale of about 10km."
- *Food nutrient information for raw fruits and veggies from USDA (National Nutrient Database, now FoodData Central):*
  - The file: usda_nutrients.csv
  - Note: If you use this dataset, you'll probably want to narrow the scope of your PCA (e.g. by limiting the food types and/or nutrients explored)

Task 3: Each group, fill out this Shiny app description spreadsheet

The guidelines and grading rubric for your Shiny app are HERE. For Task 1, each group should add their information to this spreadsheet by 2:00pm on Tuesday 2/4/2020.

Task 4: Shiny app outline (complete/submit one per group)

The goal of Task 4 is to help you get started with your final project - a Shiny app in R. Please see the final requirements here. Your app design will change over the course of creating it. That's OK. This is a starting point to get you rolling. I strongly recommend checking out RStudio's Shiny gallery: https://shiny.rstudio.com/gallery/ for inspiration and ideas.
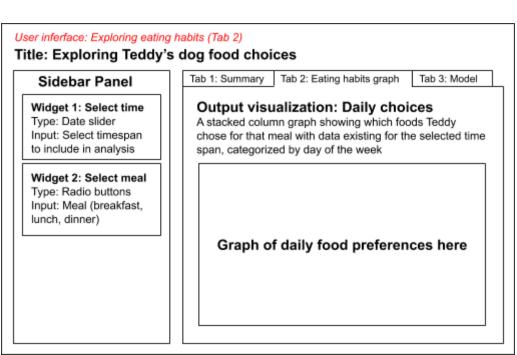
For Task 4, you will submit **one hard copy per group on Tuesday 2/4/2019** of a Shiny app outline document that contains the following (bulletpoints encouraged):

a. **App summary:** A 2 - 3 sentence "big picture" summary of what you want your app to do
b. **Data:** A ¼ - ½ page detailed description of the data you'll use in your app, including citations as appropriate. Some things to consider: What variables exist? How many observations? Are the data currently in tidy format?
c. **Widgets:** Briefly describe (½ page) the 3 (or more) widgets that you will include in your app. What type of widgets will you include (e.g. slider, radio buttons, etc. - see the Widgets gallery at the link above for possibilities)? What will be the inputs for each widget?
d. **Outputs:** Briefly describe (½ page) the 3 (or more) final outputs (tables, graphs, maps, figures, etc.) that will be produced based on widget selections.
e. **Sketch your user interface.** Produce a simple illustration (digital recommended) of your app's user interface. Each tab should have its own "slide" in your illustration. An example is included below (yours will be different - this is just to give you an idea of the type of thing I'd expect).

Note: the more detailed and thoughtful you are in preparing this app outline, the easier your life will be when actually making it.

Example app sketch (for Task 4):

User inferface: front page (Tab 1)

Title: Exploring Teddy's dog food choices

| Tab 1: Summary | Tab 2: Eating habits graph | Tab 3: Model |

HERE: A Teddy picture

In this tab (Tab 1): A summary of the app, what it does, how to use it, and a description of the data (including citations as necessary)

User inferface: Exploring eating habits (Tab 2)

Title: Exploring Teddy's dog food choices

Sidebar Panel

| Tab 1: Summary | Tab 2: Eating habits graph | Tab 3: Model |

Widget 1: Select time
Type: Date slider
Input: Select timespan to include in analysis

Widget 2: Select meal
Type: Radio buttons
Input: Meal (breakfast, lunch, dinner)

Output visualization: Daily choices
A stacked column graph showing which foods Teddy chose for that meal with data existing for the selected time span, categorized by day of the week

Graph of daily food preferences here

Title: Exploring Teddy's dog food choices

| | Tab 1: Summary | Tab 2: Eating habits graph | Tab 3: Model |

Sidebar Panel

Text description of model (inputs, outputs, assumptions, type of model, etc.)

Output: Graph of model predictions + table

A graph of predicted choices by multinomial logistic regression, and a finalized table of model outputs

Graph of predicted probabilities for food choices

Table of model outputs