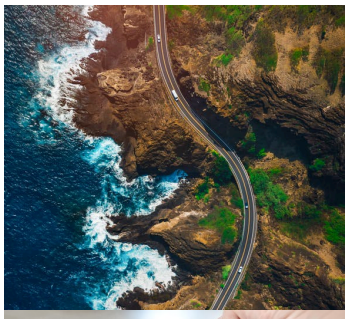

Unit 10: Advanced Linear Regression Modeling and ANOVA

Case Studies:

- We will examine the relationship between melanoma rate, global latitude, whether you living on the coast, and the *interaction between global latitude and whether you living on the coast:*
  - to introduce the concept of a **linear regression model *with interaction effects***
  - to examine how to **conduct inference on the slopes of ALL variables in a regression analysis at the same time** (rather than one at a time like we did in the last chapter)
  - to introduce the concept of **nested linear regression models.**

- We will examine the relationship political affiliation (dem/ind/no pref/other/rep) and age:
  - to introduce the concept of **a one-way ANOVA**

- We will examine the relationship between animal brain weight and body weight:
  - To explore the relationship between the **t-test and the F-test on linear regression coefficients.**

Summary of Concepts:

- **Research questions we will be able to answer**
  - **Numerical Response + Some Explanatory Variables**
    - Is there a linear association between the interaction of two explanatory variables and a response variable?
    - Is there evidence to suggest that that there is a linear association between *at least* one of our explanatory variables and the response variable?
    - Is there evidence to suggest that a linear regression model with *k* additional explanatory variables is more adequate in describing the data than the linear regression model without these additional k explanatory variables?
  - **Numerical Response and a Categorical Explanatory Variable (with >2 levels)**
    - Is there evidence to suggest that that there is an association between a categorical variable (with >2 levels) and a numerical variable?
- **Definitions/Properties**
  - Interaction effect term
  - False positive rate of a hypothesis test
  - F-distribution
- **Descriptive Analytics**
  - How to visualize the relationship between two numerical variables and a categorical variable.

- **Modeling**
  - **<u>Fit a multiple linear regression model with interaction variables</u>**
    - Interpretation of interaction variable slopes.
- **Inference**
  - **<u>Numerical Response + Some Explanatory Variables</u>**
    - Is there evidence to suggest that *at least one* of the population slopes $\beta_1, \beta_2, \ldots, \beta_p$ in a multiple linear regression model is non-zero?
    - Is there evidence to suggest that *at least one* of the *additional k* population slopes $\beta_{p+1}, \beta_{p+2}, \ldots, \beta_{p+k}$ in a multiple linear regression model is non-zero?
  - **<u>Numerical Response + Categorical Explanatory Variable (with 2 levels)</u>**
    - If you have <u>*p>2* population means</u> $(\mu_1, \mu_2, \ldots \mu_p)$ Is there evidence to suggest that *at least one* pair of these population means is different from eachother?