# moderndive: statistical inference using the tidyverse

## Quick info

* Workshop [webpage](https://www.causeweb.org/cause/uscots/uscots19/workshop/4)
* Location: Penn Stater Room 106
* Times:
  + Installation help: Wednesday May 15th 12:30pm to 1:00pm
  + Day 1: Wednesday, May 15th at 1:00pm to 4:30pm
  + Day 2: Thursday, May 16th at 8:30am to 12:00 pm

## Preparation

In preparation for the workshop, please complete the following. We’ll be in Room 106 30 minutes early on Wednesday in case you have installation questions.

* Make sure you’ve completed the following [background survey](https://docs.google.com/forms/d/e/1FAIpQLSftYCfOHObBEs9pnf1BzuBBJYb0aanagVHdYagsQIrrEFf5_Q/viewform).
* Bring colored pens/pencils
* Have the following software installed on your computer.
  + RStudio version 1.2.792 or greater
  + R version 3.5.2 (2018-12-20) or greater. Type sessionInfo() in R to get the version number.
  + The following packages. sessionInfo()will once again provide this info.
    - tidyverse version 1.2.1 or greater
    - devtools version 2.0.2 or greater
    - usethis version 1.5.0 or greater
    - moderndive version 0.2.0 or greater
    - fivethirtyeight version 0.4.0 or greater

## Day 1 Schedule

Please have the following links open in your browser:

* Development version of moderndive <https://moderndive.netlify.com/>
* This Google Doc accessible at bit.ly/USCOTS2019: I will post notes here.
* [Google Drive](https://drive.google.com/drive/u/0/folders/1aYRV3ZoukIEk-rFHa31LsgwzgqzxVphQ): I will share files here.

### Chapter 2: Exploring data in R & RStudio

1. Chalk talk, pen/paper, or tactile exercise:
   1. What would flight information look like?
2. Replicating exercise on computer:
   1. See USCOTS\_tidyverse\_day\_1.Rmd
3. Exercise:
   1. See [Chapter 2](https://moderndive.netlify.com/2-getting-started.html).
   2. Run the following code to load the CSV from Google Sheets and save it in a data frame called africa  
      library(readr)

africa <- read\_csv("https://docs.google.com/spreadsheets/d/e/2PACX-1vSttoZ6MRXOTqw6yMUJY5rUxgFuoqLq-VQIQsccSsxONSwJ935evaIrBhZv3oR6CIo4wLuTbvdlErTT/pub?gid=0&single=true&output=csv")

1. Discussion:

### Chapter 3: Data visualization

1. Chalk talk, pen/paper, or tactile exercise:
   1. Grammar of graphics
2. Replicating exercise on computer.
   1. See USCOTS\_tidyverse\_day\_1.Rmd
3. Exercise:
   1. Create a scatterplot of
      1. y = number of countries guesses
      2. x = height
   2. Create boxplot with
      1. y = number of countries
      2. x = num\_countries\_range\_guess
4. Discussion:

### Chapter 4: Data wrangling

1. Chalk talk, pen/paper, or tactile exercise:
   1. Fruit basket
2. Replicating exercise on computer.
   1. See USCOTS\_tidyverse\_day\_1.Rmd
3. Exercise:
   1. Compute the mean number of countries guess SPLIT by priming number
   2. Mutate the africa data frame to convert height in inches to cm, and make sure this variable is saved in africa. 2.54 conversion factor
4. Discussion:

### Chapter 5: Tidy data & data importing

1. Chalk talk, pen/paper, or tactile exercise:
   1. “tidy” data
2. Replicating exercise on computer.
3. Exercise:
   1. Practice [publishing](https://twitter.com/rudeboybert/status/1055821833512071168) a Google Sheet to a web-readable .csv.
4. Discussion:

### Chapter 6: Basic regression

1. Chalk talk, pen/paper, or tactile exercise:
   1. Example of SLR to illustrate residuals
   2. Example of regression with categorical variable name (Bert, Jenny, Miles)
2. Replicating exercise on computer.
3. Exercise 1: Numerical x
   1. Create a viz of regression of
      * y = num countries guesses
      * x = height
   2. Get regression table
4. Exercise 2: Categorical x
   1. Get regression table of
      * y = num countries guessed
      * x = priming number (14 vs 94)
5. Discussion:

## Day 2 Schedule

Please have the following links open in your browser:

* Development version of moderndive <https://moderndive.netlify.com/>
* This Google Doc accessible at bit.ly/USCOTS2019: I will post notes here.
* [Google Drive](https://drive.google.com/drive/u/0/folders/1aYRV3ZoukIEk-rFHa31LsgwzgqzxVphQ): I will share files here.

### Chapter 7: Multiple regression

1. Chalk talk, pen/paper, or tactile exercise:
   1. Two contrasting 1 num x + 1 cat x multiple regression
2. Replicating exercise on computer.
   1. See USCOTS\_tidyverse\_day\_2.Rmd
3. Exercise:
   1. Visualize africa data with an interaction model
      1. y = num countries guessed
      2. x1 = height
      3. x2 = categorical priming number (14 vs 94)
   2. Get regression table
4. Discussion:

### Chapter 8: Sampling

1. Chalk talk, pen/paper, or tactile exercise: sampling bowl
2. Replicating exercise on computer.
   1. See USCOTS\_tidyverse\_day\_2.Rmd
3. Exercise:
   1. Take 33 replicates of samples of size 50 from virtual bowl
   2. For each of 33 replicates, summarize them with  
      prop\_red = sum(color == “red”)/50
   3. Visualize these 33 prop\_red with a histogram
4. Discussion:
   1. These ideas are not new i.e. the “what” has been done before
      1. Statkey
      2. Mosaic package: do() operator for repeated tasks
   2. What’s different is the “how”, via the tidyverse

### Chapter 9: Confidence intervals via bootstrap resampling

1. Chalk talk, pen/paper, or tactile exercise: resampling slips of paper representing pennies
2. Replicating exercise on computer: resampling virtual pennies
3. Exercise:
   1. Construct bootstrap REsampling distribution using code from Ch8 exercise based on SINGLE sample of n=50 pennies
   2. Identify the middle 95% of values of x\_bar AKA mu\_hat
4. Discussion:
   1. See [MD 8.2.4](https://moderndive.netlify.com/8-sampling.html#different-shovels) Fig 8.11 for comparing effect of sample size as justification for showing formulas, for SE example that has n in denominator
   2. MD 9.4.4 for computing bootstrap CI’s numerically
   3. [Tweet](https://twitter.com/ModernDive/status/1110982189460713472) illustrating process for tactile resampling exercise.

### Chapter 10: Hypothesis testing via permutation resampling

1. Chalk talk, pen/paper, or tactile exercise: shuffling deck of cards to randomly reassign (binary) gender per null hypothesis
2. Replicating exercise on computer: NA
3. Exercise:
   1. [Video](https://www.youtube.com/watch?v=2pHhjx9hyM4) on gender discrimination in promotions
   2. [Worksheet](https://docs.google.com/spreadsheets/d/1MK_HBenQw9RXfTesRnqFth5ZDMR84pny_wDyYr6dLPs/edit#gid=585143573) on gender discrimination in promotions
4. Discussion:

### Chapter 11: Inference for regression

1. Chalk talk, pen/paper, or tactile exercise: None:
2. Replicating exercise on computer. See USCOTS\_tidyverse\_day\_2.Rmd
3. Exercise: Take infer package out for a spin. Resampling done using tidyverse

library(infer)

# Null distribution of test statistic

null\_distribution <- promotions %>%

specify(formula = decision ~ gender, success = "promoted") %>%

hypothesize(null = "independence") %>%

generate(reps = 1000, type = "permute") %>%

calculate(stat = "diff in props", order = c("male", "female"))

# Observed test statistic

obs\_test\_stat <- promotions %>%

specify(decision ~ gender, success = "promoted") %>%

calculate(stat = "diff in props", order = c("male", "female"))

obs\_test\_stat

# Plot!

ggplot(null\_distribution, aes(x=stat)) +

geom\_histogram(bins = 10) +

geom\_vline(xintercept = obs\_test\_stat$stat, col="red")

1. Discussion:
   1. infer.netlify.com is the package homepage -> Articles tab
   2. Comment about having single page *cheat sheet* with all arguments to infer verbs to do:
      1. One sample prop & mean
      2. Two sample prop & mean
      3. ANOVA
      4. Chi-sq
      5. Simple linear regression