This project uses the data from the Nature paper [Momentous Sprint at the 2156 Olympics](http://www.nature.com/nature/journal/v431/n7008/full/431525a.html)?, reproduces the results of that paper, and reproduces the r*eductio ad absurdum* argument by Kenneth Rice, highlighted in the Calling Bull [Gender Gap Case Study](https://www.callingbullshit.org/case_studies/case_study_gender_gap_running.html).

## Learning Objectives

In this case study, together we will:

* Reproduce scientific results, but debunk the conclusions with a *reductio ad absurdum* argument and some programming/math.
* Practice prior R syntax around plotting, functions, math, and data.frames.
* Begin “free coding,” creating your own plan for a code and implementing it (great practice for the final project!).

**Revisiting linear fits, the programming process, and math functions**

In past case studies, we’ve practiced a few different skills.

* In Bill Gates Goes to the Moon, we practiced getting into R Studio, arithmetic, Fermi estimation, and “back of the envelope calculations.”
* In [Caffeine-Free in R](https://docs.google.com/document/d/1v90mPh2a-G-oH3IBZuGbN9ECOLWOuIvJJjjGHEsvx-s/edit), we practiced writing and saving, scripts, assigning variables, and professional practices around code checks and commenting.
* In [Storks and Babies in R](https://docs.google.com/document/d/1SZjkhepznXkET9bIWzoQrfGzLp7cwsDhHnPI1_ewOVQ/edit?usp=sharing), we practiced programming and model design, data.frames, plotting, legends, and linear fitting.
* In [Mammograms](https://docs.google.com/document/d/1n8_63626nlwK8OV1L0eEKbOnOecn2L4W8DGfmw9hvko/edit?usp=sharing), we practiced critiquing visualizations, the programming and modeling process, visual implementation design, using formulas in spreadsheets, and verbal interpretation of probabilities.
* In our “R skills” refresher, we practiced for loops, and utilizing the power of indices and vector structures.
* In [Track and Field in R](https://docs.google.com/document/d/1g_sgZdFzJvYmwZjGlFd_oZyStu1po2vAk14QlSn39Ww/edit?usp=sharing), we practiced drawing and saving random variables, plotting, installing and loading packages. We emphasized self-guided exploration of the design process and of working with new functions by using Google, ? and args().

In this project, we will use all of the above skills, but you will be coming up with the plan, and executing it.

## MODELING/PROGRAMMING

**Modeling** and **programming** are broad terms that encompass a process including elements such as planning, implementation, testing, revision, feedback, etc., for some purpose. In some past exercises, I have told you which modeling process you will use. In this project, you will practice computational thinking by considering the case study at hand, and think through the steps needed in order to accomplish a reproduction of the results.

**Goal: Abstracted idea in context**

The original *Nature* paper was probably a “tongue-in-cheek” warning about extrapolating too far beyond the dataset. But it is fun to figure out why it can be so dangerous to do so through finding clever ways to refute their bullshit. The argument I’d like you to use is r*eductio ad absurdum* argument by Kenneth Rice, highlighted in the Calling Bull Case Study:

“*Sir — A. J. Tatem and colleagues calculate that women may out-sprint men by the middle of the twenty-second century (Nature* ***431****,525; 2004). They omit to mention, however, that (according to their analysis) a far more interesting race should occur in about 2636, when times of less than zero seconds will be recorded.*

*In the intervening 600 years, the authors may wish to address the obvious challenges raised for both time-keeping and the teaching of basic statistics.”*

Bergstrom and West of Calling Bull replicate the results and fit the linear fit curves to the data. They even find the latest data, add that to the fit and recalculate!

For the following project, you will use R/RStudio to:

1. Import the data as a data.frame.
2. Find out where the linear fit lines intersect (thus reproducing the results from the Nature paper).
3. Find out where the linear fit lines intersect 0, thus reproducing the r*eductio ad absurdum* argument.
4. Re-create the Calling Bull graph. For now, just worry about the data that is used by the Nature paper. However, make sure that the graph also includes the visual representation of #2 & #3.

You need not do these in the order above. In fact, you may choose to plot before finding the values of the intersections so that you can check your results visually. There are also steps implied by the above outcomes, but are not explicit. Your goals will be to make a plan to achieve the outcome above and then execute it.

Helpful hints:

* To get ahold of the data:
	+ Go to the original paper (swim upstream!!),
	+ Save the original data in the supplement as an Excel file in your working directory,
	+ Clean the data up and resave,
	+ Import the data into RStudio. Importing using RStudio is easy - go to File>Import dataset>From Excel.
	+ Copy and paste the code provided into the first line of your script.
* Here’s a [quick reference](http://www.countbio.com/web_pages/left_object/R_for_biology/R_fundamentals/multiple_curves_R.html) to plotting that may be helpful.
* See this link of calculating the necessary intersection points ([link](https://docs.google.com/document/d/1eolCdH01vB-IIt6qUelDwswhyo5bnE5gmnIO4jKxRyQ/edit?usp=sharing))

**IMPORTANT:** For this case study, please go to the Lyceum Assignment for this Case Study which will create your own version to type responses to the discussion questions, type your plan in, and insert your graph. **Each person should submit this sheet, their data file, and their R file in Lyceum by Sunday. You may collaborate on the ideas, but your submission should be in your own words, including comments on your code in your own words.**

## PLAN

**\* Coding professional practice - Intentionally plan your code before you start typing.**

**Team member names:**

**Discussion 1**

Create a plan to achieve the four outcomes above. Discuss the plan with your group. Once you agree, write it down in YOUR OWN WORDS in your copy of the Case Study. Add more steps if necessary.

1. Prepare workspace by...

1.
2.
3.

**Discussion Question 2**

**\*Coding professional practice - sketch out your plan in psuedocode before implementing. This may also prompt you to identify some commands you may have to look up. Remember to Google, use ?command, and args().**

Sketch out some psuedocode for the plan above ([example here](https://www.unf.edu/~broggio/cop2221/2221pseu.htm)). Compare and contrast approaches with your group members. What commands will you need to know in order to carry out your plan? Will this require any particular use of structures? What packages will you need to install to use these commands? Write a sketch of your code below in the psuedocode section.

PSUEDOCODE:

## IMPLEMENT

I only have an empty skeleton below - you fill in the rest. Use # to add explanations to your code

### Step 1: Set-up your workspace

Start a new R-script (in RStudio, go to **File > New File > R-script**).

Add the file information to the top of your script.

# Diaz Eaton, DCS 105 A Calling Bull (should be your name, not mine)

# Gender Gap Case Study, March 6, 2019

# Team member names

# Clear out any old variable names with the first code line

rm(list=ls())

Save the script into a working directory. Save the Excel files with the data you need into the same working directory.

#Install and load any packages needed.

### Step 2:

**\*Coding professional practice - name the parameters close to what they mean. Use # (aka “comment out”) before descriptors that explain what you are doing and/or the units, or where the numbers you are using come from.**

# Assign any variables needed

### Step 3:

### Step 4:

**\*Coding professional practice - build in ways to check your code to make sure it is functioning as intended.**

Look at your environment window in RStudio. Are the variables storing as expected?

### Step 5:

**Discussion Question 3. How can you write a self-check into your script code?** Implement a self-check into your code. What result did you get? Same as Calling Bull? Different? Is the code working properly?

## ANALYZE

**Insert the generated graph here**

**Discussion Question 4**

What does it mean/what story does it tell? What did you learn?

**(Optional) If time - Discuss with at least one other group**

What were the challenges of this project? How did your approach differ from others? Strengths/Weaknesses of the differing approaches?

**CHALLENGE (Optional)**

✨ Find the more recent track data - re-do the analysis. Any change? When will the momentus race take place?

Try plotting with ggplot2: reference [link 1](https://community.rstudio.com/t/multiple-linear-regression-lines-in-a-graph-with-ggplot2/9328), [link 2](https://aosmith.rbind.io/2018/11/16/plot-fitted-lines/) Hint: this may require re-thinking/reshaping your dataset into gender as a category.

## SUBMIT

This project is worth 20 points. This is not a “check-off.” This is an individually submitted project assignment and so **each person in your group needs to submit separately and in their own words**.

* Did you professionally comment all code in your own words?
* Is it clear who is submitting the code as well as who your team members are who also contributed intellectually to the final product?
* Did you insert the final graph in this report?
* Did you complete all goals, fill in all the blanks, and answer all discussion questions in your own words? (See the rubric below for a detailed checklist)
* Did you attach the Excel file for the cleaned data you used?
* Did you attach the final, well commented code from your R script?

|  |  |
| --- | --- |
| **Rubric for Gender Gap Case Study** | **Pts** |
| Assignment includes all items “Each person should submit **this sheet, their data file, and their R file in Lyceum by Sunday.”**  | \_\_\_/3 |
| Professional practice: Code shows commenting & Collaborators acknowledged. “You may collaborate on the ideas, but your submission should be **in your own words, including comments on your code in your own words.”** | \_\_\_/4 |
| Planning: Discussion question 1&2 (1 pt each: plan discussed & filled in, breakdown from plan to psuedocode complete and filled in), and steps in implementation section are filled in (2pts) | \_\_\_/4 |
| Project Tasks completed - For the following project, use R to:(1 pt) Find out where the linear fit lines intersect (thus reproducing the results from the Nature paper).(1 pt) Find out where the linear fit lines intersect 0, thus reproducing the reductio ad absurdum argument. (2 pts) Re-create the graph with the data used by the Nature paper, illustrating the two intersections above to show consistent agreement with the intersection answers above. (2 pts) Shows professional practice for making a good data viz - Data, Fit line, Color or filled pts w/Legend or explanation, Axis labels, Axis limits. | \_\_\_/61 pt1 pt2 pts2 pts |
| Discussion Question 3&4: Reflection on code - self-checks (1 pt) and sense-making (2 pts) | \_\_\_/2 |
| (Optional Challenges) Adding current years and re-run. Make sure you include second R file and Excel file |  |
| **Total** | **\_\_/20** |