SPPH581A: Foundations of Public Health Computing in R

Time: Fridays from 14:00-17:00
Location: Pharmacy 3208

Instructor:	**Sarah Henderson**
Email:	sarah.henderson@bccdc.ca
Phone:	604-910-9144
Office locations:	Environmental Health Services, BCCDC (Tuesday-Thursday) Room 327A, SPPH (Friday)
Office hours:	By appointment only
Teaching assistant:	Angela Yao
Email:	angelayao3@gmail.com

Required equipment:

- To get the most out of the course, bring your own laptop to class. Some tablets running Windows will be adequate, but an iPad will not.
- You must have administrative access to this machine.
- Windows, OSX, and Linux are all fine.
- If you cannot bring a laptop to class, sharing a laptop with one of your classmates will be better than no laptop at all.
- Please ensure that your operating system is up-to-date and that you have wireless internet access.

Recommended textbook:

- The Art of R Programming by Norman Matloff (tinyurl.com/zk4r9kh).
- This textbook is not required, but it is highly recommended as an ongoing resource for anyone who will continue using R for their research after the course is complete.

Overall objective:

- The entire course is designed to lead students through the data gathering, cleaning, analysis, and visualization necessary to reproduce key results for one peer-reviewed journal article recently published in Environmental Health Perspectives. (ehp.niehs.nih.gov/ehp224/). Note that the mortality dataset used in the course will be realistic, but completely falsified to ensure no risk to personal privacy. As such, the results we obtain will not be the same as those reported in the paper.

Course format:

- There will be 12 weeks of lectures. The first eight weeks will cover data preparation and the next four weeks will cover data analysis. This breakdown generally reflects the nature of data analysis, where 2/3 of the time (or more) is dedicated to data preparation and 1/3 of the time is dedicated to analysis and visualization.

- Each lecture will be a stand-alone guided tutorial that we follow together in class, with students working alone, in pairs, or in small groups as they prefer. If we do not complete the tutorial in class, you MUST complete it in your own time prior to the next class.
- Because the tutorials are stand-alone, it is not strictly necessary for you to attend class. However, students who do not attend will miss valuable context and insight provided by the instructor, the TA, and your fellow students. Marks will also be given for class participation.

Grading:

- 30% based on in-class participation, as evaluated by the instructor (10%), the TA (10%), and your fellow students (10%).
- 30% based on in-class midterm, which will be completed in assigned pairs.
- 40% based on take-home final assignment, which will be completed alone.

Class participation:

- Programming is, by nature, a collaborative activity and this course reflects that reality. The highest marks in the class will go to students whose attitudes are collaborative rather than competitive. You can teach each other far more than we can teach you.

Midterm exam:

- You will be given a large, dirty dataset that requires cleaning and organization. You will work with an assigned partner using your notes, the internet, and whatever else you need to prepare the dataset for analysis.
- At the end of the midterm you will submit your R script for evaluation.
- Your code will be graded for correctness, not elegance.

Final assignment:

- You will be responsible for designing a small study of interest to you or to someone else. The only limitation is that you must be able to share the required data with the course instructor and the TA. Secure data storage can be arranged, if necessary.
- You will do the data cleaning and analysis necessary to address your study question. Your code must demonstrate that you can carefully organize a project, that you can use loops, and that you can write your own custom functions.
- You will submit a draft manuscript to be considered as a Brief Article in the American Journal of Public Health. Specifically, the article will present preliminary findings with up to 1200 words in the main text, up to 200 words in a structured abstract, up to 2 tables and/or figures, and no more than 12 references. The article must have separate sections for the Introduction, Methods, Results, Discussion, and Public Health Implications.
- You will also submit the RProject folder associated with your draft manuscript.
- Your project will be marked on its overall quality. The grades will consider your experience with R at the beginning of the course.

Very brief course schedule:

Activity	Objectives
Lecture 1 Sep 9*	- Install R, RStudio, and Git - Load mortality data and start cleaning
Lecture 2 Sep 16	- Continue cleaning mortality data - Convert character variables to date format
Lecture 3 Sep 23	- Finish cleaning mortality data - Save cleaned file in R format
Lecture 4 Sep 30	- Download temperature data from the web using R - Loops and flow control
Lecture 5 Oct 7	- Clean temperature data - Use pipelines to pass objects between functions
Lecture 6 Oct 14*	- Prepare case-crossover dataset with custom functions - Merge with temperature data
In-class midterm Oct 21	
Lecture 7 Oct 28	- Geocode the residential 6-digit postal codes - Apply national Postal Code Conversion File (PCCF)
Lecture 8 Nov 4	- Create a map of the decedent locations - Assign values for the 16 spatial variables at each location
Remembrance Day Nov 11	
Lecture 9 Nov 18*	- Identify the split values for each spatial variable - Conditional logistic regression
Lecture 10 Nov 25	- Run the case-crossover analyses for each of the variables - Gather the results into a format that will allow easy pglotting
Lecture 11 Dec 2	- Learn about different plotting devices - Start plotting paper figures
Lecture 12 Dec 9	- Finish plotting paper figures - More on debugging
Take-home final assignment due Dec 16	

*Indicates a conflict for students in SPPH526-DL. Angela will offer a special lecture for these students sometime earlier in the week.