Back to Week 5 | Lessons | Prev | Next

Identify a research question similar to questions we’ve talked about in this course. Use the Behavioral Risk Factor Surveillance System (BRFSS) dataset (provided below).

All analysis must be completed using the R programming language via RStudio, and your write up must be an R Markdown document. To help you get started we provide a template Rmd file below (see Rmd template in the Required files section below). Download this file, and fill in each section.

IMPORTANT: Analyses completed using software other than R, or not written up using R Markdown, will receive a 0 on the project regardless of their content.

Required files

- Data - Save this file in the same directory as the Rmd template (provided below).

NOTE: If you are using Chrome as your browser you might need to change the .gz at the end of the extension to .Rdata in the file you downloaded.

brfss2013.RData

- Codebook - Review this file to find out what each column in the data represents.

brfss_codebook.html

- Rmd template - You must use this template to write up your project. Save the data and this file in the same directory.

intro_data_prob_project.Rmd

- Assessment rubric - You might want to review the assessment rubric while working on your project so that you have some idea of how your peers will evaluate your work.

intro_data_prob_project_rubric.html

Instructions

Your project will consist of 3 parts:

1. **Data**: (3 points) Describe how the observations in the sample are collected, and the implications of this data collection method on the scope of inference (generalizability / causality). Note that you might will need to look into documentation on the BRFSS to answer this question. See http://www.cdc.gov/brfss/ as well as the "More information on the data" section below.
2. **Research questions**: (11 points) Come up with at least three research questions that you want to answer using these data. You should phrase your research questions in a way that matches up with the scope of inference your dataset allows for. Make sure that at least two of these questions involve at least three variables. You are welcomed to create new variables based on existing ones. With each question include a brief discussion (1-2 sentences) as to why this question is of interest to you and/or your audience.
3. **EDA**: (30 points) Perform exploratory data analysis (EDA) that addresses each of the three research questions you outlined above. Your EDA should contain numerical summaries and visualizations. Each R output and plot should be accompanied by a brief interpretation.

In addition to these parts, there are also 6 points allocated to format, overall organization, and readability of your project. Total points add up to 50 points. See the assessment rubric (provided above) for more details on how your peers will evaluate your work.

You can begin working on the project immediately. Please save your work as you go along. When you're ready to submit your work for evaluation, remember to click the "Submit" button.

After you submit your project, please provide feedback to others on their projects. Please assess at least 3 projects. This peer assessment will not only provide you with experience with a data set and research questions but also prepare for your projects in the future courses in the Specialization.

More information on the data

The Behavioral Risk Factor Surveillance System (BRFSS) is a collaborative project between all of the states in the United States (US) and participating US territories and the Centers for Disease Control and Prevention (CDC). The BRFSS is administered and supported by CDC's Population Health Surveillance Branch, under the Division of Population Health at the National Center for Chronic Disease Prevention and Health Promotion. BRFSS is an ongoing surveillance system designed to measure behavioral risk factors for the non-institutionalized adult population (18 years of age and older) residing in the US. The BRFSS was initiated in 1984, with 15 states collecting surveillance data on risk behaviors through monthly telephone interviews. Over time, the number of states participating in the survey increased; by 2001, 50 states, the

District of Columbia, Puerto Rico, Guam, and the US Virgin Islands were participating in the BRFSS. Today, all 50 states, the District of Columbia, Puerto Rico, and Guam collect data annually and American Samoa, Federated States of Micronesia, and Palau collect survey data over a limited point- in-time (usually one to three months). In this document, the term “state” is used to refer to all areas participating in BRFSS, including the District of Columbia, Guam, and the Commonwealth of Puerto Rico.

The BRFSS objective is to collect uniform, state-specific data on preventive health practices and risk behaviors that are linked to chronic diseases, injuries, and preventable infectious diseases that affect the adult population. Factors assessed by the BRFSS in 2013 include tobacco use, HIV/AIDS knowledge and prevention, exercise, immunization, health status, healthy days — health-related quality of life, health care access, inadequate sleep, hypertension awareness, cholesterol awareness, chronic health conditions, alcohol consumption, fruits and vegetables consumption, arthritis burden, and seatbelt use. Since 2011, BRFSS conducts both landline telephone- and cellular telephone-based surveys. In conducting the BRFSS landline telephone survey, interviewers collect data from a randomly selected adult in a household. In conducting the cellular telephone version of the BRFSS questionnaire, interviewers collect data from an adult who participates by using a cellular telephone and resides in a private residence or college housing.

Health characteristics estimated from the BRFSS pertain to the non-institutionalized adult population, aged 18 years or older, who reside in the US. In 2013, additional question sets were included as optional modules to provide a measure for several childhood health and wellness indicators, including asthma prevalence for people aged 17 years or younger.

Source: Duke University Data and Visualization Services

Frequently Asked Questions

Do I have to use R for my project? Yes. While there are other statistical packages and/or programming languages that may be perfectly appropriate for your project, since one of the goals of this course is to learn R, all analysis **must** be completed in R and using the Rmd template provided above. Projects completed using other statistical packages and/or programming languages will receive a 0 on the project.

Where can I find a list of R commands that might be useful for the project? Refer to the previous labs and see the RStudio cheatsheets for dplyr, ggplot2, and RMarkdown.

Who am I writing for? Write as if you are explaining your results to whomever would be interested in your research question, whether this is another scholar in your field or peers sharing your interest in the topic. This audience may not have taken a statistics course. You must be statistically accurate and use correct statistical terminology, but must also explain your conclusions in a way that anyone can understand.

Does my project have to be written in English? Yes, your project must be written in English; this is the only way to ensure that the students who are assigned to review your project can understand it.

What is a peer assessment? Peer Assessment is when students in a course evaluate a fellow student's work. First, each student submits an assignment. Then, the students who have submitted an assignment are given other students' assignments to evaluate, according to provided criteria. Finally, each student receives a grade that is based on the other students' evaluations.

Can I use a paper I've worked on for another course or purpose? No. Please create a unique project for this course. Do not use your master's thesis, work you have published elsewhere or work you have submitted for another course. In the past, students who have submitted work they used elsewhere were reported as submitting plagiarized work.

What if I think the project I am assessing has been plagiarized?

1. Assess the project according to the Evaluation/Feedback directions.
2. Report plagiarism to Courserahttps://learner.coursera.help/hc/en-us/articles/209818863-Coursera-Honor-Code

How do I avoid plagiarism? "In an instructional setting, plagiarism occurs when a writer deliberately uses someone else’s language, ideas, or other original (not common-knowledge) material without acknowledging its source." - The Council of Writing Program Administrators Therefore, please give credit for all of the sources you have used. Copying and pasting from a site without giving the source is plagiarism, and will be reported. For more information, see this tutorial on avoiding plagiarism. In your own project, give credit to all sources you used, even if you have paraphrased them or if they are your own work but published elsewhere.

Mark as completed