

Day 2, Digital Trace Data

Lesson Plan: Summer Institutes in Computational Social Science 2021

Prepared by SICSS-Rutgers, Adapted from SICSS-Duke

Summary

An open-ended group exercise to gain practice with collecting digital trace data, formulating related research questions, and extending upon the data in research design and/or presentation.

Learning Objectives

- Participants gain practice collecting digital trace data via screen-scraping or working with APIs, as well as formulating a hybrid research design and/or building apps and bots for social science research.
- Participants may generate more potential ideas for group projects in the second week.

Activity

- Split into small groups.
- Complete small group popcorn prompt
- Phase One: Groups should select one person to take notes and report group process
- Complete the prompts described below for Phase One
- Phase Two: Groups should select one person to take notes and report group results
- Complete the prompts describe below for Phase Two
- Come back together as a large group and discuss projects at the end of the day.

Rough schedule

Phase One: 1:45pm - 2:30pm

- 15 minutes: introduction of activity and breaking into small groups
- 30 minutes: read the activity and go through the Phase One question prompts

Phase Two: 2:30-5:00pm

- 120 minutes: complete Phase Two question prompts
- 30 minutes: group discussion of projects

Prompt

Phase One:

1. For 15 minutes, work together to identify a topic that you believe digital trace data (e.g., data collected through screen-scraping or working with APIs) can help describe or be used to address a specific research question.
2. Identify a relevant sampling frame. For example, if your research topic is about politics, your sampling frame might be a list of elected officials on Twitter.
3. Discuss the strengths/weaknesses of the data for addressing your research topic.

Phase Two:

4. Collect the data!
5. Improve Upon the Digital Trace Data. Choose one of the below:
 - Build a Shiny App to present your data in a user-friendly way for exploring or describing the topic (also allowed to use another program for app creation)
 - Develop a hybrid research design that allows you to combine the digital trace data with some other type of data that will generate more information and address some of the weaknesses you identified with the data (e.g., creating a bot to interact with subjects).
6. Discuss the strengths/limitations of what you have completed.

Resources

(See also folder with code examples for screen scraping, using rtweet, tuber (YouTube API package), and shiny.)

APIS:

- R packages for APIs: RgoogleMaps, googlelanguageR, rOpenSci, WDI, rOpenGov, rtimes, rtweet
- <https://docs.google.com/spreadsheets/d/1ZEr3okdlb0zctmX0MZKo-gZKPsq5WGn1nJOxPV7al-Q/edit?usp=sharing>

More Screen Scraping Examples:

- <https://www.analyticsvidhya.com/blog/2017/03/beginners-guide-on-web-scraping-in-r-using-rvest-with-hands-on-knowledge/>

Using R Shiny

- [Example code for building keyword search shiny app from SICSS-Rutgers](#)
- <https://shiny.rstudio.com/articles/basics.html>
- <https://mastering-shiny.org/index.html>
- Data storage: <https://shiny.rstudio.com/articles/persistent-data-storage.html>

Google's Flutter

- <https://flutter.dev/docs/codelabs>