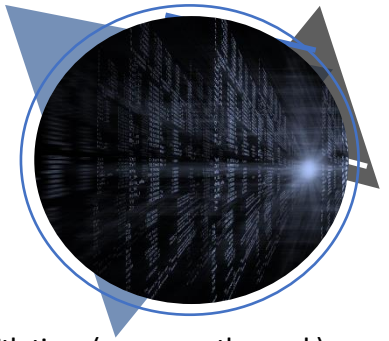




BDHSC Research Data Repositories

The University of South Carolina Big Data Health Science Center (BDHSC) is pleased to announce that it will help facilitate access for USC faculty to a wide range of BDHSC research data repositories (RDR) for big data health science research. The current BDHSC RDR include the following datasets:

- 1. Cellphone-based Population Flow Data** were obtained from SafeGraph and processed by the [Geoinformation and Big Data Research Lab](#) at the [Center for GIScience and Geospatial Big Data](#) in collaboration with BDHSC for academic research purposes. This data contains the monthly and weekly visitations flows originating from over 230,000 Census Block Groups (CBGs) to over 5 million Points of Interest (POIs) in the US from 01/01/2018 to 08/30/2022. In total, this dataset has 9.5 billion ODT flows and can be requested in two formats: 1) individual Origin-Destination-Time (ODT) flows filtered with time (year, month, week) and geographic location.
- 2. Twitter Data** were collected again by the [Geoinformation and Big Data Research Lab](#) at the [Center for GIScience and Geospatial Big Data](#) for academic research purposes. This is a live dataset that contains worldwide tweets covering over 10 years from 2012 to present (real-time tweets are being collected around the clock). The total number of tweets as of December 2022 is around 18.6 billion. There are two types of Twitter data in the database: geotagged tweets and randomly sampled tweets. The geotagged tweets are continuously collected using the official Twitter Streaming API with geo filters. The randomly sampled tweets were downloaded from the Internet Archive.



For information on the data repositories and data request, please visit the BDHSC website at: <https://bigdata.sc.edu/research-data-repositories/>.

For questions regarding the BDHSC or RDR initiative, please contact Dilek Akgun at akgun@mailbox.sc.edu.