**The Geography of Radio**

Title: *Real-time Radio Station Audio Sampling, Processing, and Logging as a Proxy for Spatiotemporal Popular Music Trends*

Authors: Michael Felzan, Bryan Runck

Date: Nov 11th 2021

**Project Repository:**<https://github.com/fezfelzan/Geography-of-Radio.git>

**Abstract**

This project develops an ETL process for iteratively sampling radio station streams to generate a database describing spatiotemporal music trends throughout the US. We present a Python-based process for iteratively sampling MP3 clips of from a collection of online radio streams, identifying the songs playing in those sampled audio clips using a music recognition API, retrieving detailed genre tags of associated with those songs from a music database API, retrieving technical details about each radio tower associated with the URL stream (ERP, HAAT, distance given a specific dBu, geographic coordinates), and organizing all of this data into spreadsheets which are interoperable with a GIS software like ArcGIS Pro. We additionally present methods for performing spatial analysis on this dataset, which include the creation of a statistical surface (kriging) showing the normalized prevalence of certain genres linked to the geographic locations they are broadcasted in, and the times at which they are played. We postulate that the broadcast signal reach (distance from radio tower which would yield 60 dBμ, what is often deemed the ‘edge’ of local coverage/reception for a station) may be used as a proxy for describing cultural aspects—or more specifically, what kinds of music are valued—of a given geographic region, based on the assumption that the station’s broadcast reach and content choices are a product of local funding. Working under this assumption, we present a novel case study which analyzes the intersection of geography and radio – by way of real-time audio data-scaping of FM station’s web broadcasts, and analysis spatiotemporal cultural trends across the US, based on the types of music represented in each sampled radio stream. This report concludes with a discussion of the implications, drawbacks, and opportunities for future work with the material presented in this project.

**Introduction**

Personal computers and smart devices now are completely integrated in our everyday lives, serving as our primary mediums through which we communicate, consume media, organize travel routes, and now due to the COVID-19 Pandemic, it is becoming ever more prevalent. It is therefore no surprise that a large majority of FM radio stations in the US have already set up web-access to their live broadcasts. A team of programmers from Amsterdam capitalized on the now widespread digital availability of radio streams through designing ‘radio.garden,’ a website which allows users to tune into thousands of live radio streams by clicking on glowing points on a world map corresponding to the actual locations of those station’s radio towers. The ability to access FM radio broadcasts—which continue to be hyper-localized avenues for information and music consumption by those who tune in—via the web offers vast opportunities associated with programmatic data-scraping, and the ability to use radio as a proxy for a geographic area’s culture. Radio, as it relates to the fields of geography and GIS, seems to be untapped territory; Wikle et al. (2010) in their report on the ‘new geography’ of religious radio confirm that the existing “geographic research focused on radio is limited in scope.”

This report presents a novel ETL methodology which programmatically samples audio from a variety of radio stations in the US, in order to compile a database which details the types of music played by specific radio stations at given dates and times. This project assumes that the songs which receive airplay by radio stations may be representative of music preferences and associated cultures of the locations which receive each respective station’s broadcast. The main intent of this project is to assess the feasibility of using radio broadcasts as a data source for analyzing cultural geography at a given time.

**Methods**

* [**Title:**](https://arxiv.org/help/prep#title) *required*
* [**Authors:**](https://arxiv.org/help/prep#author) Michael Felzan, Bryan Runck (?)
* [**Abstract:**](https://arxiv.org/help/prep#abstracts)
* Introduction
* Input Data?
* Methods?
* Results
* Discussion
* Conclusion (?)
* [**Comments:**](https://arxiv.org/help/prep#comments) optional, but recommended
* [**Report-no:**](https://arxiv.org/help/prep#report) *required (when supplied by author's institution)*
* [**Category:**](https://arxiv.org/help/prep#subj) depends on archive
* [**Journal-ref:**](https://arxiv.org/help/prep#journal) reserved for publication info
* [**DOI:**](https://arxiv.org/help/prep#doi) reserved publication DOI
* [**MSC-class:**](https://arxiv.org/help/prep#msc) math archives only
* [**ACM-class:**](https://arxiv.org/help/prep#acm) cs archives only

Machine Learning Engineering by Burkov

(tensor flow?0

Building machine learning pipelines

<https://arxiv.org/>

-figure out personal hardware limitations, also generally how much computing power <all stations in US> connections would require

-figure out, given cores

- io connection maximums

-api limits

-5,622 stations in US

-email MSI

<https://www.msi.umn.edu/content/hpc>

🡪 write

<https://docs.mongodb.com/guides/server/introduction/>

VMs on Google Cloud

<https://cloud.google.com/compute/docs/network-bandwidth>

<https://developer.apple.com/forums/thread/38674>

<https://stackoverflow.com/questions/410616/increasing-the-maximum-number-of-tcp-ip-connections-in-linux>