Project 2: Overview

Background

The goal of this project was to extract information from three untidy datasets and prepare these datasets for downstream analysis. The untidy data were selected from options and/or formats identified/shared by DATA607 students.

Our project team (Daniel Moscoe, Ethan Haley, and Sean Connin) identified the following untidy datasets for cleaning and analyses, each pertaining to aspects of hurricane activity in the Atlantic and related impacts on New Orleans, LA over time: i

1. Wikipedia HTML Tables – measures of hurricane activity, intensity, and damages since 1900
2. Excel Spreadsheets – demographic changes in New Orleans attributed to Hurricane Katrina (2005)
3. Google Maps – spatial data (GeoJASON) for surviving landmarks and homes of jazz musicians in New Orleans.

The datasets identified by our team afforded us opportunity to construct a body-of-work with thematic coherence while challenging us to acquire and prepare data from different sources and in different forms.

Methods

Our team met via. videoconference four times over two weeks to ensure agreement on project goals development, and results. Regular asynchronous exchanges via. Zoom also enabled us to provide each other assistance and updates during interim periods.

Data and associated graphics were prepared in the RStudio work environment - with raw data, R scripts, and related materials hosted on a shared Github repository. This repository can be accessed at https://github.com/sconnin/New\_Orleans\_Demographics.

The repository includes the following files:

Proj2\_Untidy\_Wikipedia\_Tables.txt – raw data extracted from multiple HTML tables

Proj2\_Wikipedia.rmd – R markdown script used to clean, graph, and analyze the Wikipedia data

Proj2\_ Atlantc\_Hurricanes.csv – tidy dataset extracted from Wikipedia html tables and saved as .cvs

This report is organized into sections for each of the three datasets. Each section includes an Overview, R Script, and Summary. Documentation related to data source(s) and description are included in the section overviews. Results, references, and recommendations are discussed in the section summaries.

Our project assessment (observations, lessons learned, etc) are included as final Conclusions.