**Basic Info**

Project Title: MultiStream

We are doing a technique study based on the MultiStream technique for visualizing hierarchical time-series data using multi-resolution streamgraphs. The paper detailing this technique can be found at: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8267086.

Team Members: Ryan Capps and Ellie Nguyen

Project Repo: https://github.com/ryjo44/IntroToVisProject

**Background and Motivation**

The main source of inspiration for choosing this project came from assignment 2 in this course. We both really enjoyed working with hierarchical time-series data: specifically, we both found the user interaction component interesting. Going into project selection we know we wanted to find a technique focusing on a data vis for this type of data, and found the MultiStream technique.

**Objectives**

We have two main objectives for this project. Firstly, we both want to learn more about visualizing hierarchical data and how to create effective and interesting user interactions with it. Secondly, we wanted to work on understanding technical writing in the vis field. In the course of the class we weren’t exposed to too many formal studies and reports on visualization techniques, and being able to work from these kinds of reports is a very valuable skill to have moving forward.

**Data**

There are a few different data sources we plan on using for this project. The most important one for us is the dataset provided in Assignment 2. This dataset is of the type that this visualization technique is designed for, and we already know that this data is clean and in a usable form. This dataset will be the primary one that we use during our implementation of the MultiStream technique in order to make sure everything is functioning properly.

Once we have implemented this technique to our satisfaction, we plan on using the following two data sets to investigate its effectiveness.

International Visitation to the US: <https://data.world/zika-virus-data/international-visitation-to-us?fbclid=IwAR1Un5b9NK2S8I2MoCNE1q0vns0EAo0rWK2xXEePYWCXzdKbVGhVsjwC0Ms>

US Permanent Visa Applications:

<https://www.kaggle.com/jboysen/us-perm-visas?fbclid=IwAR2hrGC7JHY2MI9F1UKo-fo4Ysv10Sw6j6Bp5Z0oATNsI6bwxUmLRMuxKx4>

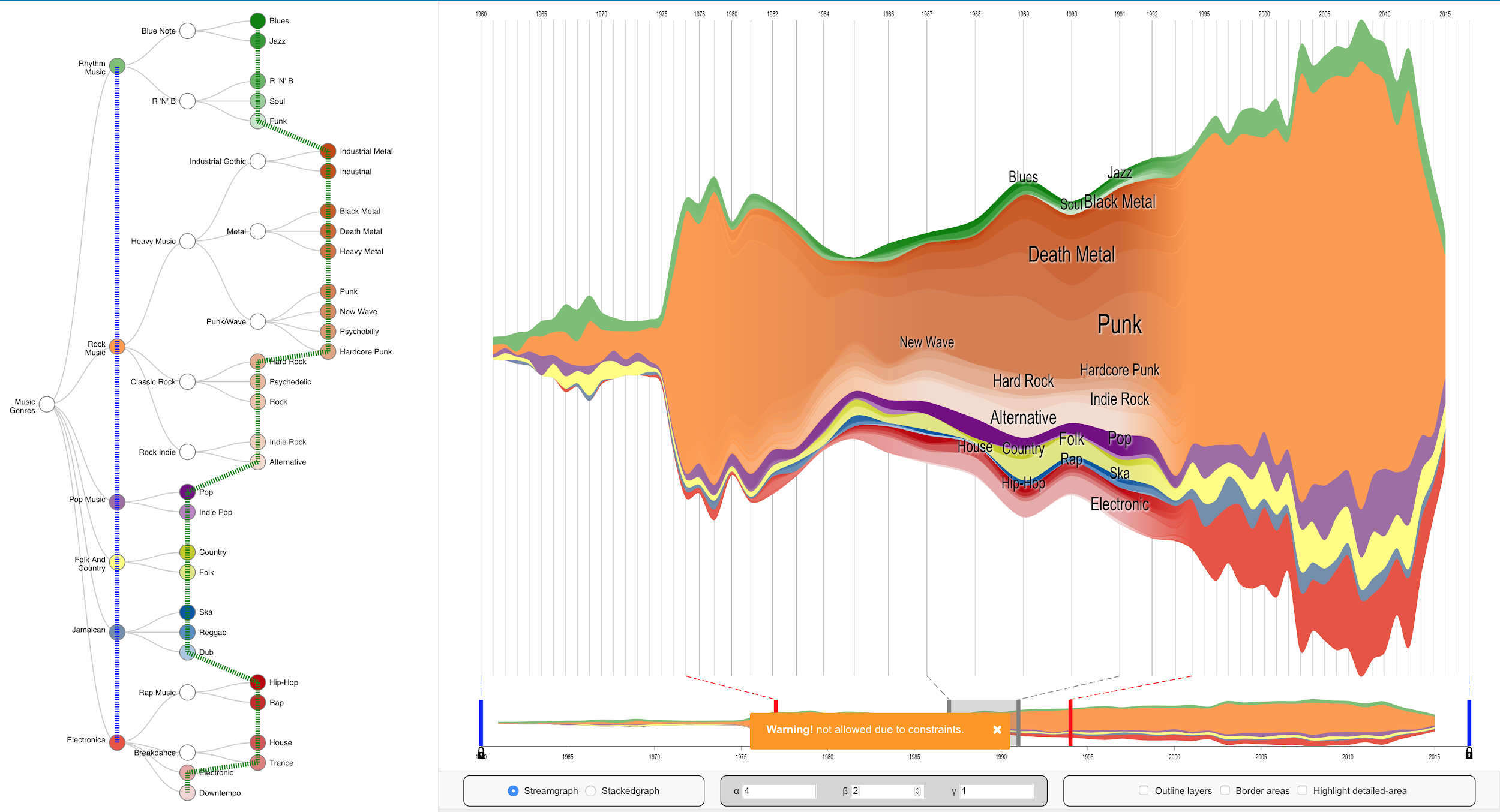
**Data Processing**

We plan on devoting less than a week to data processing. Upon initial inspection, the datasets that we have selected look to be pretty clean and well organized. The main change we will have to make to them is adding in more hierarchical structure. The two datasets both focus heavily on location, but only really define location in one way. For example, the US Visa Application dataset lists only the country of origin of the applicant. From this, we can add in more general categories (like continent) which will allow us to add more hierarchical structure while still staying true to the data.

We plan on using R for the bulk of our data processing, and then using d3 to aggregate it into the correct hierarchical structure. All of the code we need for the d3 section is available through our previous assignments.

**Must-Have Features:**

There will be three main views in the visualization: the overall view which shows the full time series with time-view controller, the zoom view showing the zoomed part of time series and the hierarchy manager.



Ex 1: A screenshot of MultiStream implementation

The main visualization techniques used in MultiStream are stacked graph and streamgraph visualization. The techniques will visualize temporal patterns over the time series to reveal patterns and trends. For example, we could see what the time period that has the highest numbers of visa applications using a steamgraph. We would support changing back and forth between a steamgraph and a stacked graph.

We would also support viewing hierarchical structure in time series. A problem arises with multiple time series is the view might be cluttered with too many streams and it is hard for the users to decipher the trend of each series. Therefore, using colors and layer order, the relationship between individual time series in a hierarchy would be shown.A hierarchy manager would be implemented to allow the users to specify the levels of hierarchy that they want to visualize. For example, we would group the visa applications and the international visitation into a hierarchy of continents and countries. The user could change the view of the hierarchical structure of time series at different levels of details. For example, the user could view the visualization by continents or choose to see each country in a specific continent.

To control the view of the time dimension, we will implement the controller described in the paper, which could change the multiresolution view over the time dimension, divide it into detailed-area, context-areas and transition areas. The user could change the detail area to move along the time axel or change the width of each area

**Project Schedule:**

Week 1:

* + Completed proposal
    - Ryan: Basic Info, Background and Motivation, Objectives, Data, and Data Processing
    - Ellie: Visualization Technique, Must-Have Features, Optional Features, Project Schedule, Analysis
* Week 2:
  + Ryan:
    - In charge of data collection and data cleaning
    - Set up data structures
    - Make a prototype of the steamgraph and stackgraph without all the interaction
  + Ellie:
    - Implement the hierarchical manager to update the data structures
    - Implement the time series view
* Week 3:
  + Ryan:
    - Complete the analysis’s draft of the technique with multiple datasets
  + Ellie:
    - Implement the zoom and detail for each time series
    - Make sure the interaction are in place
* Week 4: Prepare the presentation
* Week 5: Final project submission and peer assessment completed