**Flood and Global Warming: Visualization and Prediction**

Pengyu Wang & Kyoun Huh

**Introduction**

As a common type of disaster, flood claimed lives and damaged properties throughout history. Notwithstanding the frequent occurrences of floods, our society is still underprepared while facing these disasters. To study the features and patterns of floods could help save lives and minimize economic losses.

With the power of big data technologies, this project aims to study the main floods all over the world since 1985. We present their locations and severities, examine the common causes of floods and how these causes changed through decades. Given the context of global warming, we also present how the temperatures have been changed in different areas of the world. Then, we attempt to link the temperature changes to flood occurrences. In addition to exploring the correlation between temperature and flood, a machine learning model is built to predict probabilities of floods based on historical temperature records. The final results of this project are presented by a visualization dashboard and a report.

**Products**

1. Interactive Visualization Dashboard

2. Report

3. Prediction model

**Data Sources**

Flood Data: Global Active Archive of Large Flood Events, Dartmouth Flood Observatory, University of Colorado.

Temperature Data: Earth Surface Temperature, Berkeley Earth.

**Technologies**

Data Storage: AWS S3 + AWS RDS (PostgreSQL)

ETL: PySpark on AWS EMR

Machine Learning: Spark ML

Visualization: Plotly + Heroku

Version control: GitHub