STAT 605
Final Project Proposal

SAMACH-SHENG-KYUNGJIN-SHRAVAN-MIDHUN

Agenda

Background

Dataset

Variables

Statistical method

Computational steps

Background

Global warming is the long-term crucial issue throughout many decades.

We would like to study the effect of the **climate factors** to the global warming signal.

We scope down the analysis to a significantly huge location, **the United States**.

> *"Do climate factors such as precipitation, snowfall, snow depth significantly affect the probability of global warming?"*

Dataset

Dataset contains daily climate records measured at different stations across the globe.

Comprised by the National Oceanic and Atmospheric Administration.

(Database : Global Historical Climatology Network (GHCN))
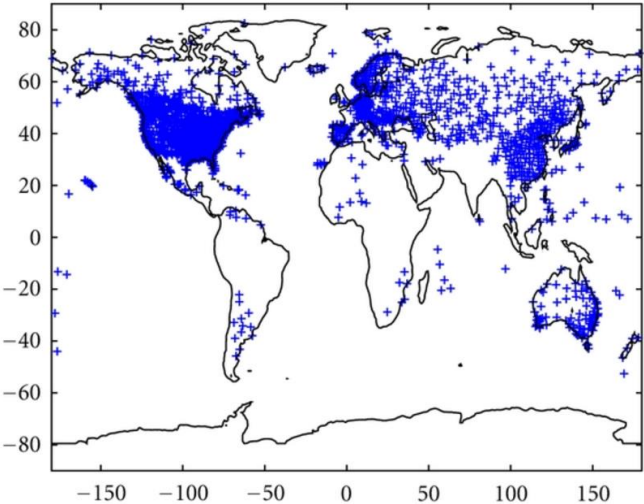
V1	V2	V3	V4	V5	V6	V7	V8
USC00419361	20190101	TMAX	156			7	800
USC00419361	20190101	TMIN	39			7	800
USC00419361	20190101	TOBS	39			7	800
USC00419361	20190101	PRCP	0			7	800
USC00419361	20190101	SNOW	0			7	NA
USC00419361	20190101	SNWD	0			7	NA

Variables

V1	V2	V3	V4	V5	V6	V7	V8
USC00419361	20190101	TMAX	156			7	800
USC00419361	20190101	TMIN	39			7	800
USC00419361	20190101	TOBS	39			7	800
USC00419361	20190101	PRCP	0			7	800
USC00419361	20190101	SNOW	0			7	NA
USC00419361	20190101	SNWD	0			7	NA

Station_id

Date_id

Measure

Value of measure
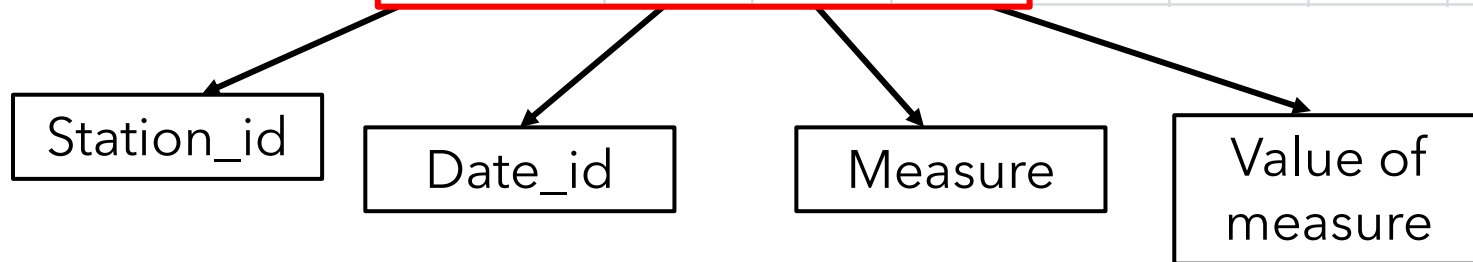
Measures

TMAX = Max temperature
TMIN = Min temperature
TOBS = Temp at a time observed
PRCP = Precipitation
SNOW = Snowfall prcp
SNWD = Snow depth
And others …

Statistical method

> **"Do climate factors such as precipitation, snowfall, snow depth significantly affect the probability of global warming?"**
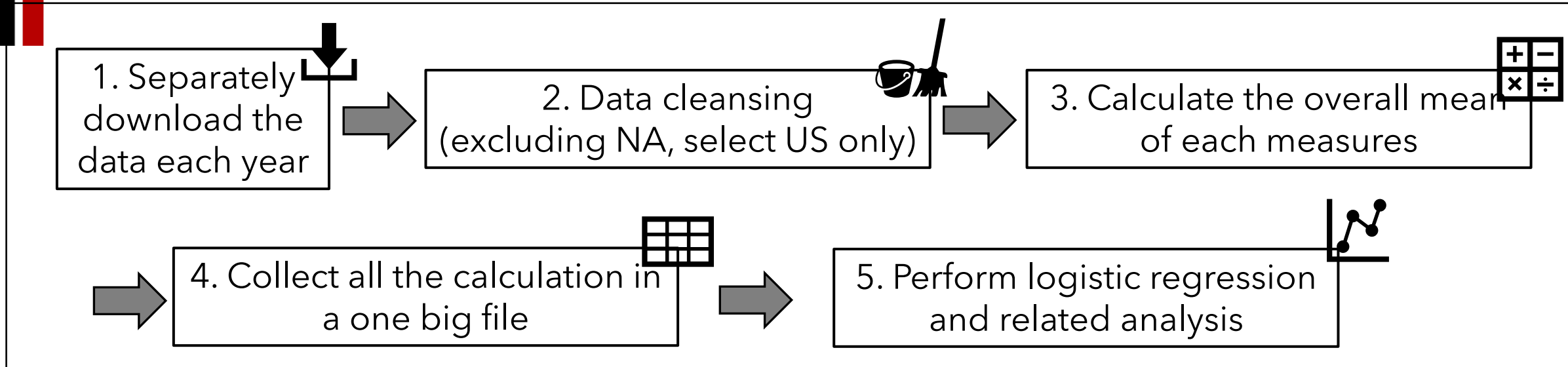
Perform **binary logistic regression** to evaluate the probability of global warming using **R**

We create data column **"is_sign"** based on the average temperature each year and label
- **1 as the temperature is greater than the average (the sign of overheat)**
- **0 as the temperature is lower than the average (no sign of overheat)**

We then model the probability of global warming (is_sign) based on each climate factors

Computational steps

1. Separately download the data each year → 2. Data cleansing (excluding NA, select US only) → 3. Calculate the overall mean of each measures

→ 4. Collect all the calculation in a one big file → 5. Perform logistic regression and related analysis

Date_id	Temperature	Precipitation	Snowfall	Snow depth
19730101	xxx	xxx	xxx	xxx
19730102	xxx	xxx	xxx	xxx
...				
20190112	xxx	xxx	xxx	xxx