

Sequía en Chile

CC5212-1 Procesamiento masivo de datos

Profesor: Aidan Hogan

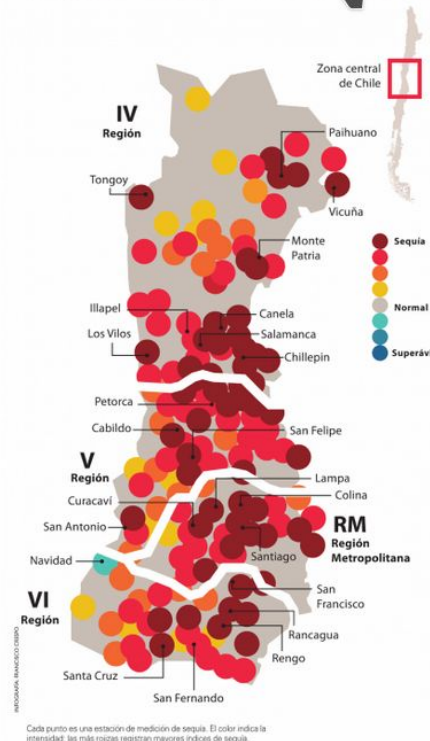
Integrantes: Cristian Ahumada
Felipe Flores
Francisco Munoz

Grupo: 17

Motivación

- Extensa sequía en Chile (casi 10 años)
- valles centrales fuertemente afectados
- numerosos estudios realizados
- registros históricos disponibles

Mapa: los lugares más afectados con la peor sequía de los últimos cien años



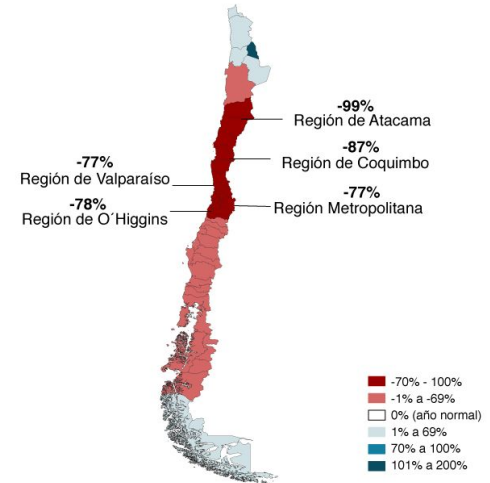
EL MEDIO AMBIENTE

Triste, pero cierto: las lluvias ya no están bien dibujadas en invierno. Ni siquiera asoman como borbotones. Los prados tampoco tienen ese alboroto a tierra recién regada. Y los árboles frutales, peor aún, ya no se reventan en primavera. Es una situación preocupante, aterradora a decir verdad, pero es la pura y santa verdad. Según los científicos del Centro del Clima y la

ma mediterránea (lluvias concentradas en invierno) de estas regiones se parecen cada vez más, debido a este desplazamiento, al clima semi-árido del sector sur del Monte Chico".

Las zonas más afectadas, detalla Garreaud, quien tiene un magister en Ciencias de la Geofísica en la Universidad de Chile, corresponde al trazo de tierra que va desde la Quinta a la Sexta Región. "El fenómeno es muy exten-

Mapa de déficit/superávit de precipitaciones acumuladas a agosto 2019*



*Comparación con el promedio histórico a la misma fecha entre los años 1981-2010

Objetivos

- Analizar cambios en lluvias registradas a lo largo del país durante los últimos 100 años y relacionar su comportamiento con la actual sequía.
- Preguntas del proyecto:
 - ¿Cómo varía la precipitación a lo largo de los años y a lo largo del país?
 - ¿Cuales son las zonas con mayor reducción en sus precipitaciones?
 - ¿Han ocurrido sequías de esta magnitud en el pasado?
 - ¿Cuando empezaron a disminuir las precipitaciones?

Datos

- Dataset creado por CR2 (Center for Climate and Resilience Research).
- Agrupa registros de lluvias de tres instituciones: DGA, DMC y GHCN.
- Contiene registros de **precipitación acumulada diaria** para **1249 estaciones** de Chile, la Antártica y otros países de Latinoamérica.
- Los registros inician en **Enero de 1900** y cubren hasta **Marzo de 2018**.
- Contiene **53.794.430** de registros.

Datos

El dataset contiene:

- Información correspondiente a cada estación como: código, institución, latitud, longitud, fecha de inicio de funcionamiento, etc
- registros diarios de precipitación acumulada medida en milímetros.
- días no registrados son identificados como -9999

Metodología

Las tareas realizadas fueron:

1. Preprocesar los datos: reordenar tabla y filtrar datos.
2. Definición de zonas geográficas.
3. Cálculo de precipitación acumulada anual.
4. Cálculo de índice de precipitación estándar (SPI) anual por zonas.
5. Cálculo de precipitación anual promedio para cada estación.

Tareas 1, 4 y 5 fueron realizadas con Apache Spark

Tareas 2 y 3 fueron realizadas con Apache Pig

Preprocesamiento

- Rotación de tabla
- reducir número de columnas
- filtrar datos fuera de chile continental
- eliminar registros malo (-9999)

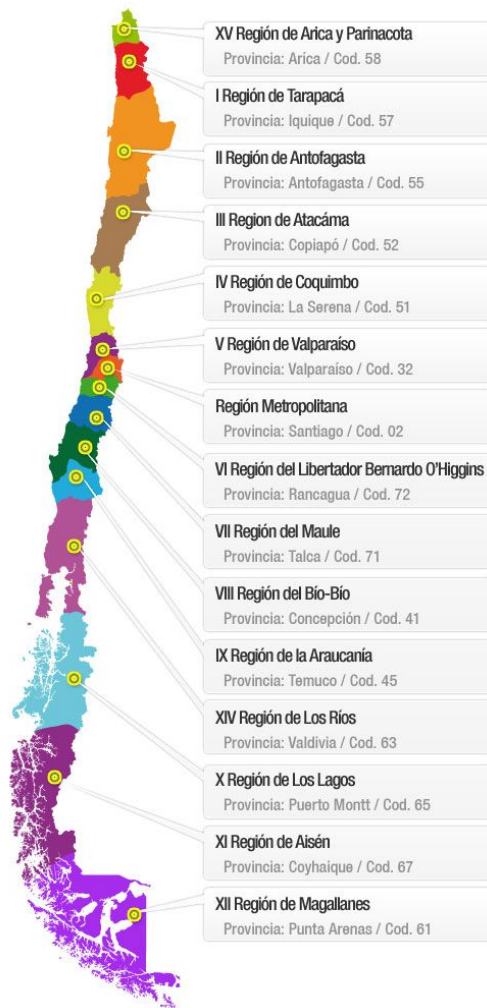
estación	12123	53535	24244	3334
latitud	-31.23	-27.0	-34.5	-23.3
...
1900-01-01	0.0	0.7	5.0	-9999
1900-01-02	30.0	-9999	3.0	-9999



Estación	Año	Mes	Dia	Precip
12123	1900	01	01	0.0
12123	1900	01	02	30.0
53535	1900	01	01	0.7
24244	1900	01	01	5.0
24244	1900	01	02	3.0

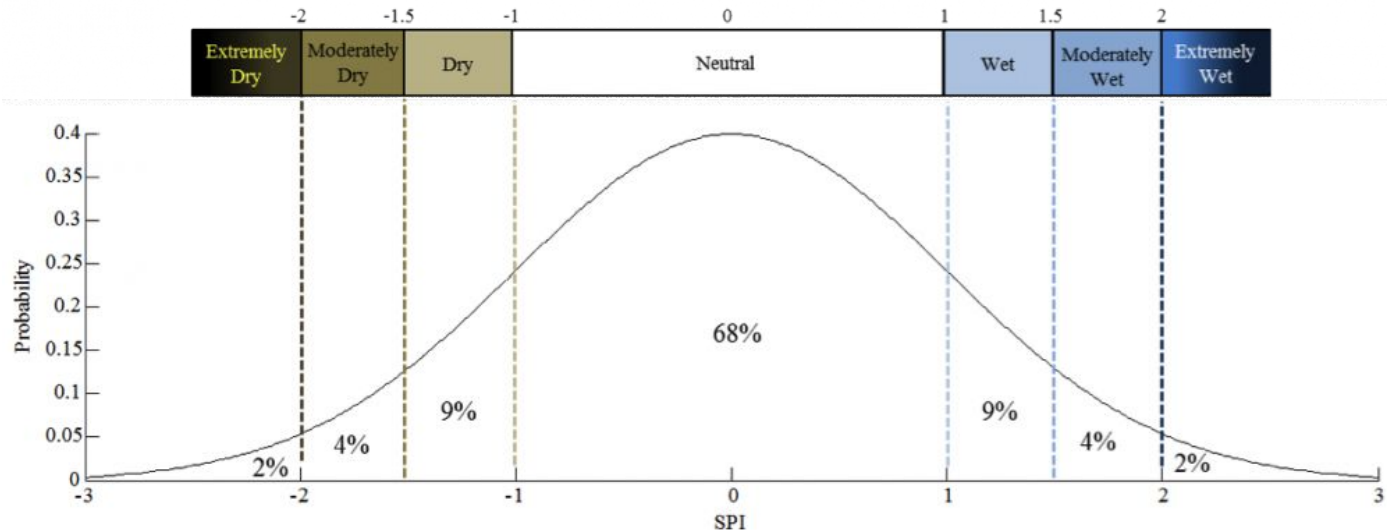
Zonas Geográficas

- división por regiones administrativas
- división por regiones naturales

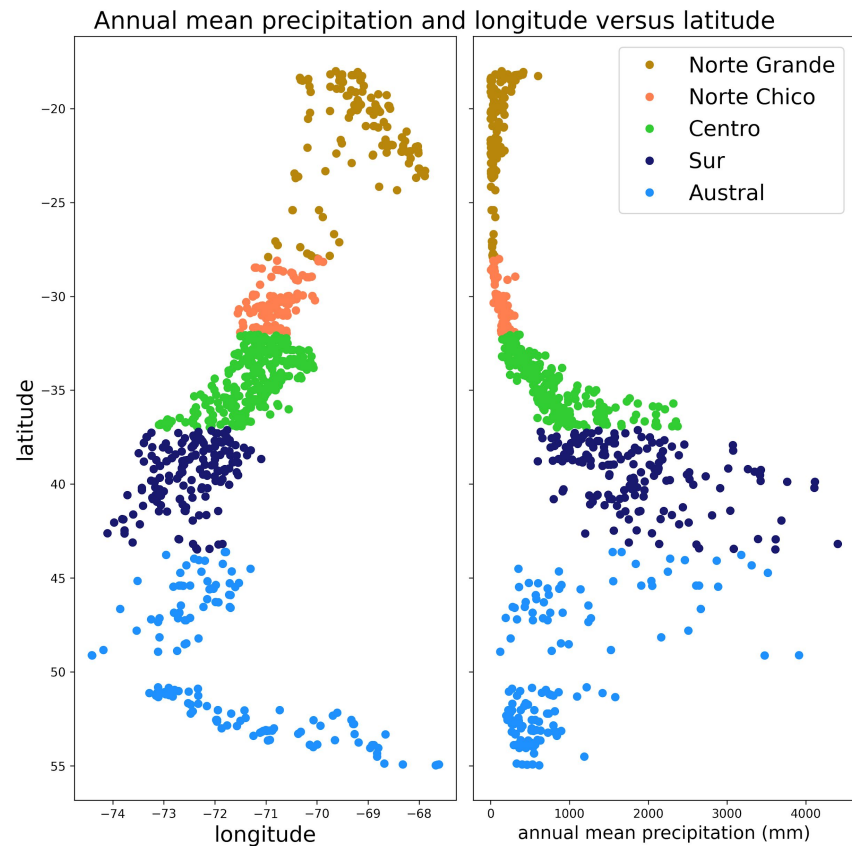
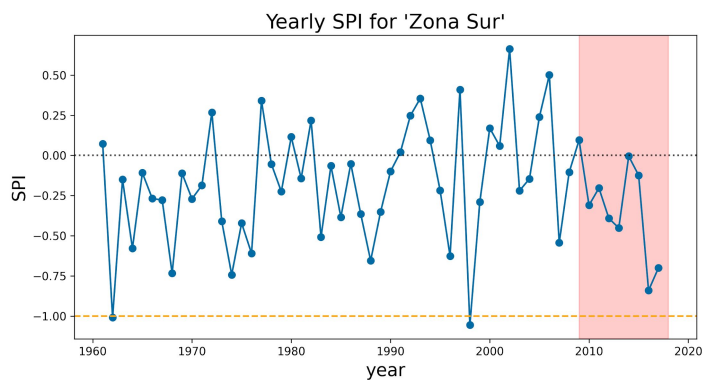
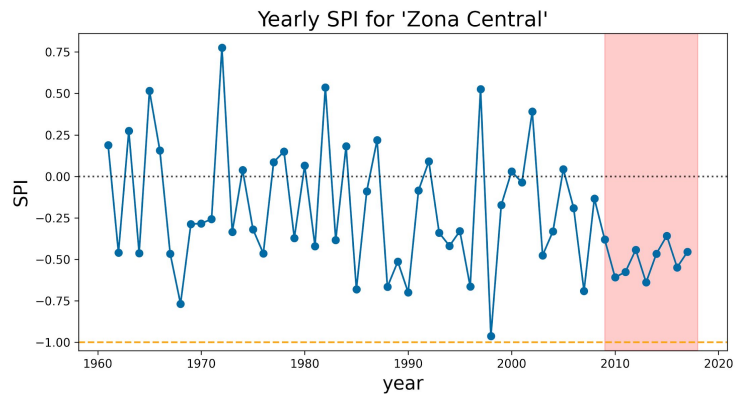


Índice de precipitación estándar (SPI)

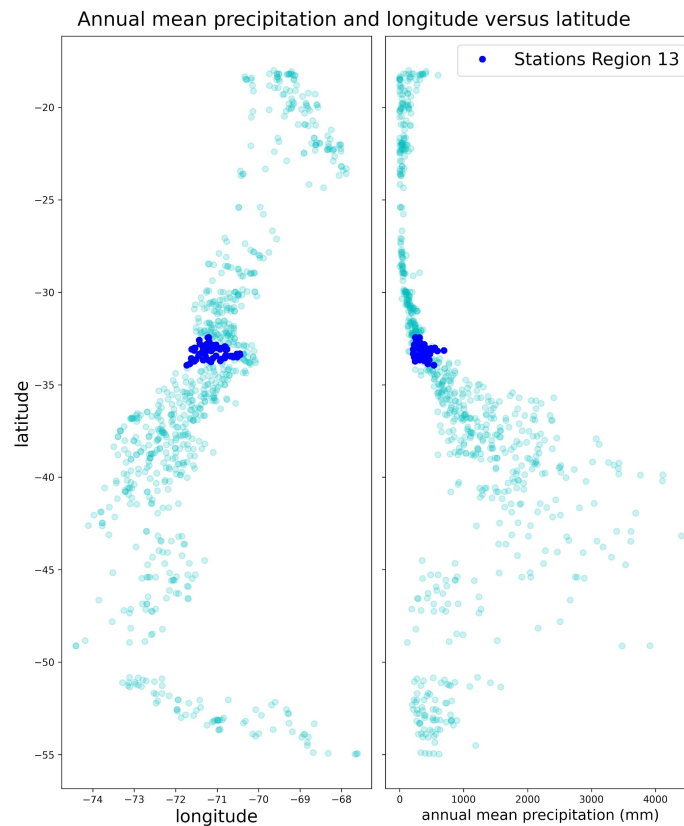
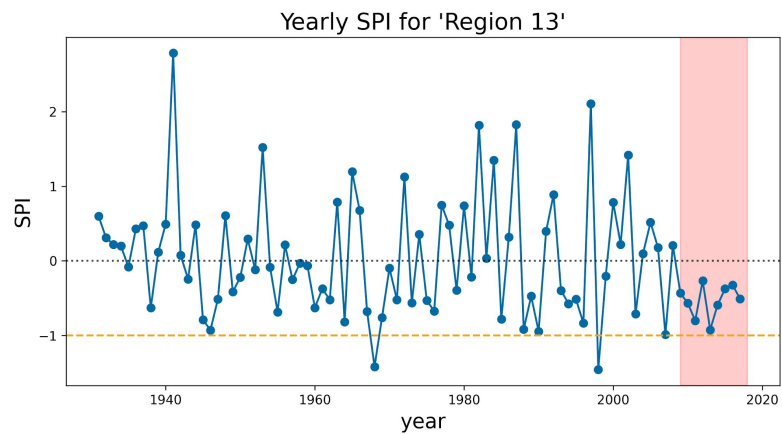
- Se procesan los datos para obtener un valor normalizado.
- Normalización en base una distribución normal.
- La distancia del promedio define sequía o no sequía.



Resultados



Resultados



Conclusiones

- Preprocesamiento complejo de datos crudos
 - no es directo rotar una tabla y luego colapsar sus columnas
- Definición de zonas geográficas desafiante
 - elegir los límites correctos y construir la tabla correspondiente
 - Los límites políticos no necesariamente se condicen con los límites geográficos o naturales.

Conclusiones

- Precipitación acumulada anual fácil de calcular
- Cálculo de SPI difícil de realizar
 - requirió un procesamiento de varios pasos
- Apache Pig y Apache Spark fáciles de manejar