

Visualizaciones precipitaciones y temperatura media.



Integrantes:

- Benjamín Bustos A.
- Felipe Sanhueza C.
- Maximiliano Aguilar S.
- Samuel Sanchez P.

Base de datos

El set de datos utilizado para el proyecto consiste en cinco tablas obtenidas desde el Centro de Ciencia del Clima y Resiliencia (CR)².

1. Datos de precipitación: Cr2 PrDaily 2019 Ghcn Zip
[<http://www.cr2.cl/datos-de-precipitacion/>]
2. Datos de temperaturas medias: Cr2 TasDaily 2019 Ghcn Zip
[<http://www.cr2.cl/datos-de-temperatura/>]
3. Datos de temperaturas minimas: Cr2 TasminDaily 2019 Ghcn Zip
[<http://www.cr2.cl/datos-de-temperaturas-extremas/>]
4. Datos de temperaturas máximas: CR2 TasmxDaily 2019 Ghcn Zip
[<http://www.cr2.cl/datos-de-temperaturas-extremas/>]
5. Datos de caudales: cr2_qlfxDaily_2018.zip
[<http://www.cr2.cl/datos-de-caudales/>]



Descripción de los datos

Las 5 bases de datos contienen una estructura similar de tablas:

1. **Descripción.**

Este archivo .txt no es una tabla, es un documento de texto el cual describe la fuente de donde se adquirieron los datos, las variables contenidas en los datos y notas sobre estos.

2. **Estaciones.**

Esta tabla contiene el listado de estaciones utilizadas, con los metadatos indicados en la tabla de datos.

3. **Mediciones.**

Esta tabla contiene toda la información respecto a las mediciones realizadas por las estaciones y, también incluye la información para caracterizar estas.

Descripción de los datos

1. **Datos de precipitación**
 - a. Cantidad de registros: 43934
 - b. Cantidad de estaciones: 1255
2. **Datos de temperaturas medias**
 - a. Cantidad de registros: 43934
 - b. Cantidad de estaciones: 901
3. **Datos de temperaturas minimas**
 - a. Cantidad de registros: 43935
 - b. Cantidad de estaciones: 926
4. **Datos de temperaturas máximas**
 - a. Cantidad de registros: 43934
 - b. Cantidad de estaciones: 920
5. **Datos de caudales**
 - a. Cantidad de registros: 43181
 - b. Cantidad de estaciones: 810



Center for Climate
and Resilience Research

Descripción de los datos

- En conjunto se consideran alrededor de 2100 estaciones.
 - Sobre estas se toman las diferentes mediciones.
 - Se incluyen mediciones de Sudamérica y algunas estaciones australes.
- Las mediciones son de una resolución temporal diaria y su periodo de medición varían de acuerdo a su tipo:
 - Precipitaciones: Enero de 1900 - Marzo de 2020.
 - Temperaturas medias: Marzo de 1941 - Marzo de 2020.
 - Temperaturas minimas: Enero de 1900 - Abril de 2020.
 - Temperaturas maximas: Enero de 1900 - Abril de 2020.
 - Caudales: Febrero de 1913 - Marzo de 2018.
- En caso de no existir mediciones un dia, el valor ingresado es "-9999".

Tipos de datos - Estaciones.

Cada una de las 5 tablas de estaciones tiene los siguientes campos:

- **Código estación** (Nominal categórico)
- **Institución** (Nominal categórico)
- **Fuente** (Nominal categórico)
- **Nombre estación** (Nominal arbitrario)
- **Altitud** (Ordinal intervalo)
- **Latitud** (Ordinal cero absoluto)
- **Longitud** (Ordinal cero absoluto)
- **Código de cuenca** (Nominal categórico)
- **Nombre de cuenca** (Nominal categórico)
- **Código de subcuenca** (Nominal categórico)
- **Nombre de subcuenca** (Nominal categórico)
- **Fecha comienzo obs.** (Ordinal intervalo)
- **Fecha fin obs.** (Ordinal intervalo)
- **Cantidad obs.** (Ordinal cero absoluto)
- **Inicio obs. en estación automática** (Ordinal intervalo)

Tipos de datos - Mediciones.

En las cinco tablas de mediciones las primeras filas están destinadas a describir los valores ya representados en sus tablas de estaciones respectivas y así funcionar como encabezados.

Las filas que restan contienen mediciones individuales realizadas por cada estación y contienen la **fecha** (Ordinal intervalo) junto con el valor de la medición realizada, cuyo tipo de dato para cada tabla es el siguiente:

- Precipitación: (mm - Ordinal cero absoluto)
- Temperatura media (°C - Ordinal intervalo)
- Temperatura minima (°C - Ordinal intervalo)
- Temperatura maxima (°C - Ordinal intervalo)
- Caudal (m^3/s - Ordinal cero absoluto)



Fase exploratoria



Problemas iniciales.

Análisis inicial de los datos:

- Tablas de medición presentan cada estación como una columna.
- Muchos registros en tablas de medición con valores faltantes en fechas más pretéritas.
- Valores faltantes en registros representados con “-9999”.
- Diferentes tipos de datos entre los valores de una misma columna.
- Valores incorrectos de acuerdo al contexto del problema

Preprocesamiento inicial.

Se utiliza Python con la librería Pandas para gestionar la depuración de datos de forma eficaz e intuitiva.

- Remoción de tuplas desde 1900 a 1969.
- Remoción de encabezados (1ras 14 tuplas)
- Detección de nulos y duplicados.
- Conversión de formato de los valores en la tabla de medición a uno estándar.
- Reemplazo de datos con valor “-9999” por “null”.
- Agrupación de registros de mediciones por mes.
- Creación de tablas transpuestas de medición.

Preprocesamiento inicial.

- Remoción de datos de observaciones de las tablas de estación y unión de estas en una común.
- Geolocalización de direcciones según latitud y longitud (Librería GeoPy)
- Agregación de país y región como columnas a partir de direccion
- Redimensionamiento de las tablas transpuestas e intersección con la de estaciones, identificando cada una con un nuevo campo: "tipo_medicion"
- Unión de todas las tablas del paso anterior en una nueva gran tabla: big_data

Descripción de los datos preprocesados.

1. Datos de precipitación

- a. Cantidad de registros: 18353 -> 603
- b. Cantidad de estaciones: 1255

2. Datos de temperaturas medias

- a. Cantidad de registros: 18353 -> 603
- b. Cantidad de estaciones: 901

3. Datos de temperaturas minimas

- a. Cantidad de registros: 18353 -> 603
- b. Cantidad de estaciones: 926

4. Datos de temperaturas máximas

- a. Cantidad de registros: 18353 -> 603
- b. Cantidad de estaciones: 920

5. Datos de caudales

- a. Cantidad de registros: 17600 -> 579
- b. Cantidad de estaciones: 810

Datos big_data

- a. Cantidad de registros: 1675414
- b. Cantidad de estaciones: 16



Center for Climate
and Resilience Research



Hipótesis y datos

Hipótesis

1. Las temperaturas incrementarán a través de los años lo cual, reducirá las precipitaciones y a su vez, causará una reducción en los caudales.
2. ¿Cómo se comparan las mediciones en un rango de 1 a 10 años previos y, posteriores a desastres naturales en Chile?
3. ¿Existen diferencias significativas entre las mediciones de temperatura en diferentes coordenadas de latitud y longitud en un determinado tiempo?
4. ¿Donde ocurrieron la mayor cantidad de precipitaciones en Chile? ¿Existe alguna relación con algún evento en particular?
5. En base al nivel de precipitación y la temperatura de una zona, ¿es posible identificar distintos climas?

Relevancia de los datos en relación a las hipótesis

Respecto a los datos y su relación con la hipótesis podemos notar lo siguiente:

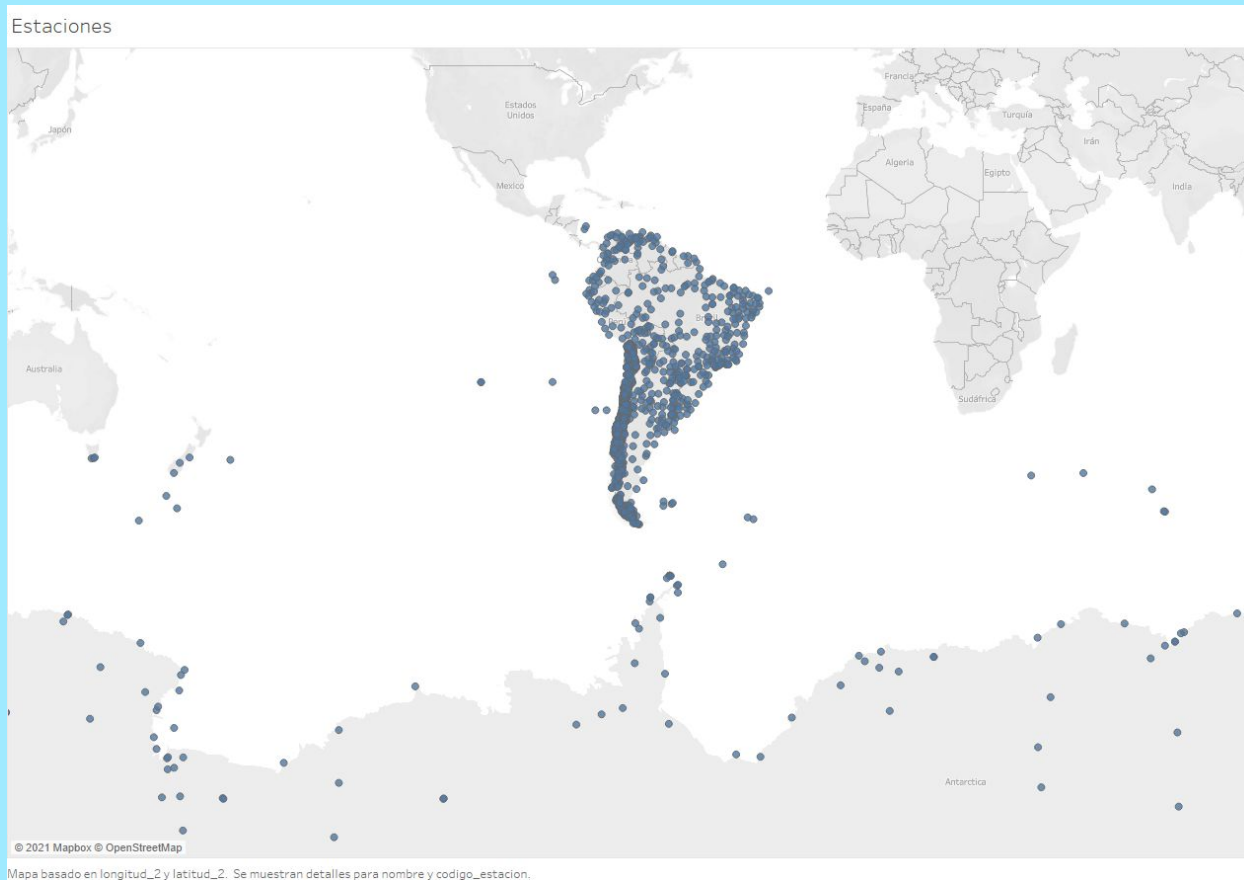
- Contamos con una base de datos con estaciones posicionadas estratégicamente con el fin de otorgar cobertura geográfica a través del hemisferio sur con énfasis en américa del sur.
- Estas mismas estaciones recolectan diferentes mediciones de características climáticas de interés, tales como temperatura, precipitaciones y caudales.
- Tenemos un periodo de 50 años con datos relevantes por lo cual, podemos observar cambios climáticos ocurridos en el tiempo.
- Las mediciones están geolocalizadas por país y región lo que permite poder agrupar sobre estas variables e identificar patrones por sectores.

...

Visualizaciones preliminares

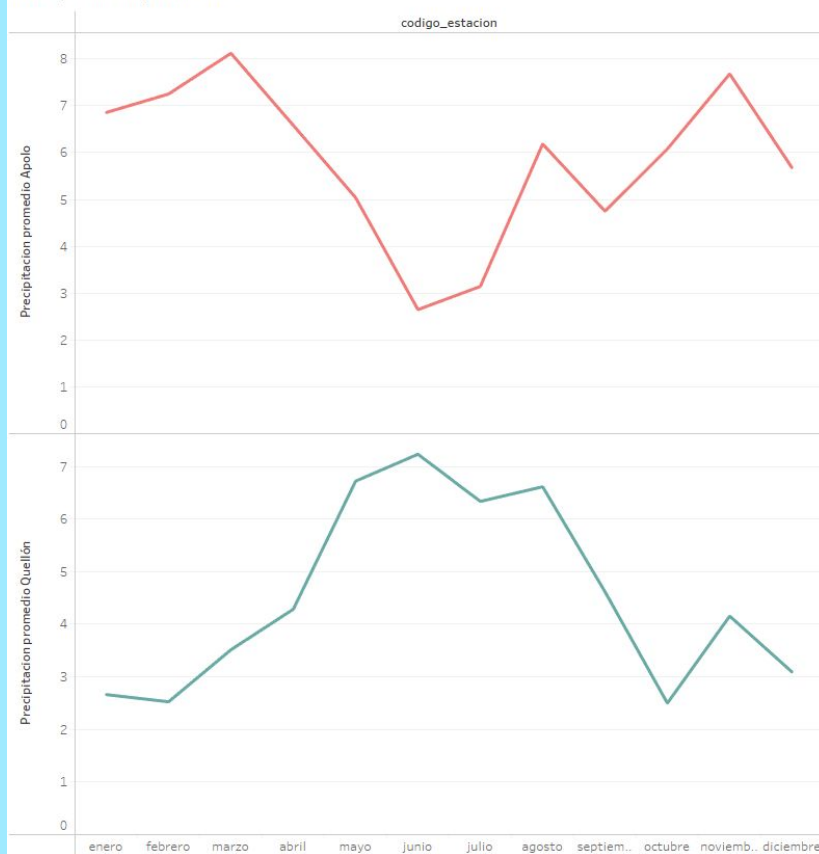


Ubicación geográfica de estaciones



Precipitaciones Apolo vs Quellón

Precipitación promedio

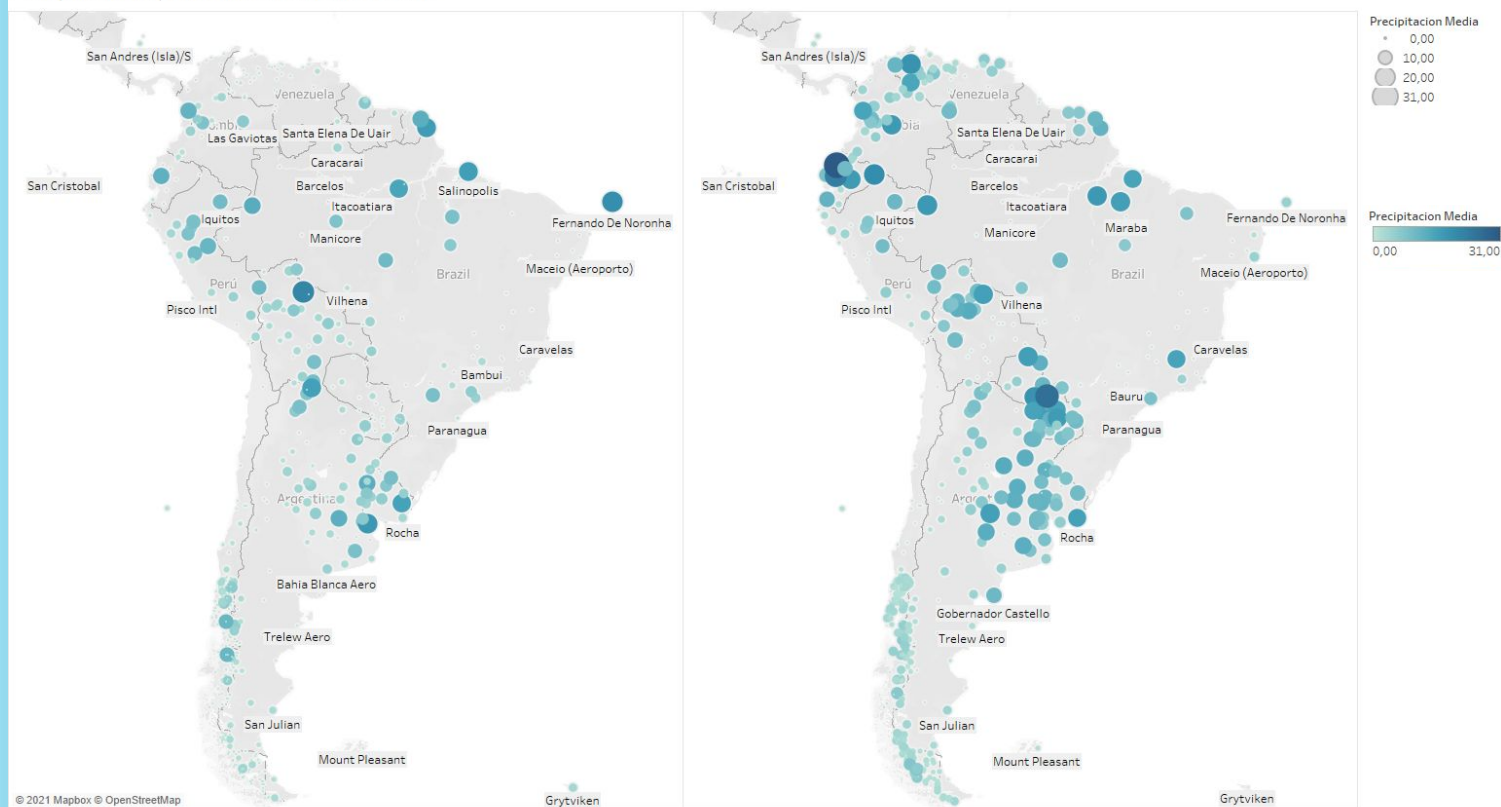


Apolo vs Quellón



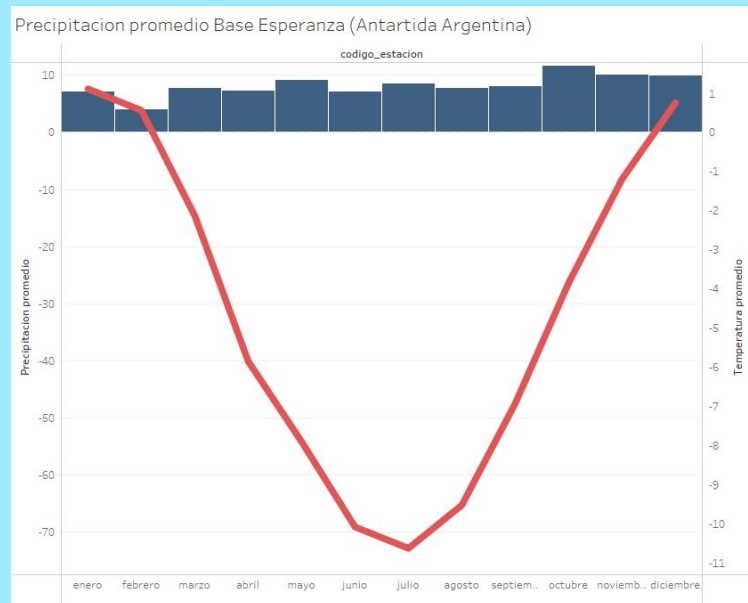
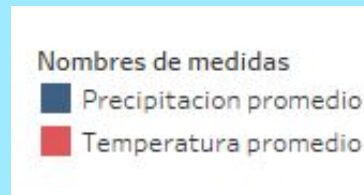
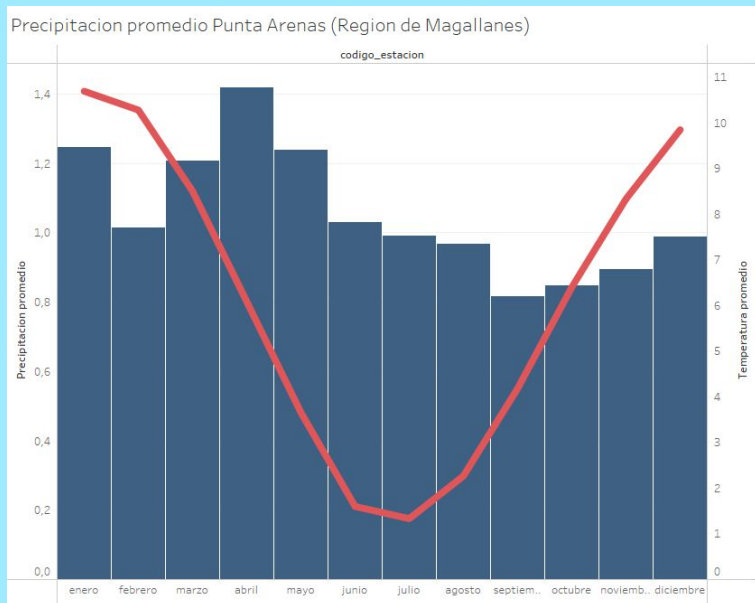
Comparación precipitaciones promedio 1980 - 2012

Precipitaciones promedio marzo 1980 vs 2012



Mapa basado en longitud_2 y longitud_2 y latitud_2. Las marcas se etiquetan por nombre. Para el panel Longitud_2: El color muestra detalles acerca de fecha_1980. El tamaño muestra detalles acerca de fecha_1980. Para el panel Longitud_2 (2): El color muestra detalles acerca de fecha_2012. El tamaño muestra detalles acerca de fecha_2012.

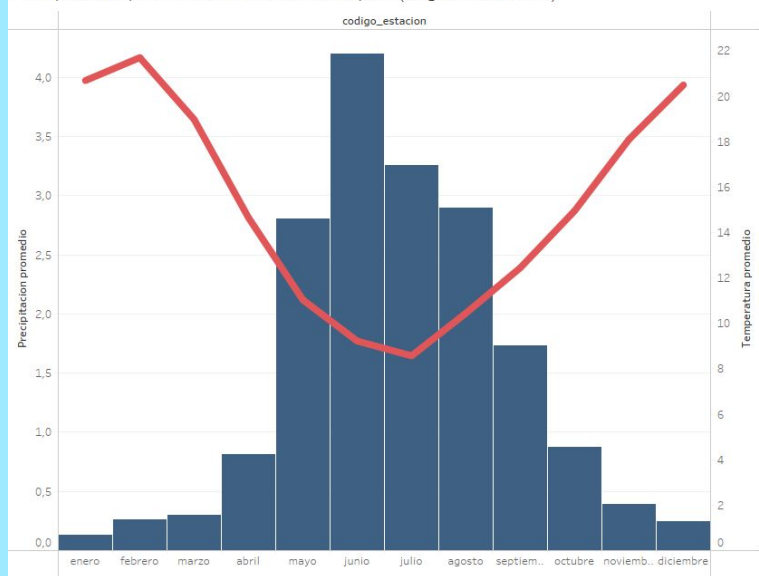
Climogramas



- Variables gráficas:
 - Color
 - Forma
 - Posicion

Climogramas

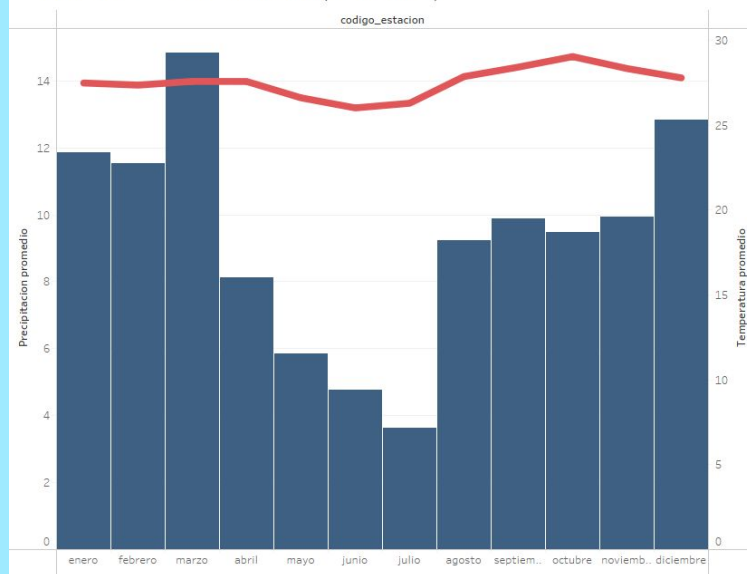
Precipitación promedio Río Claro en Rauquen (Region del Maule)



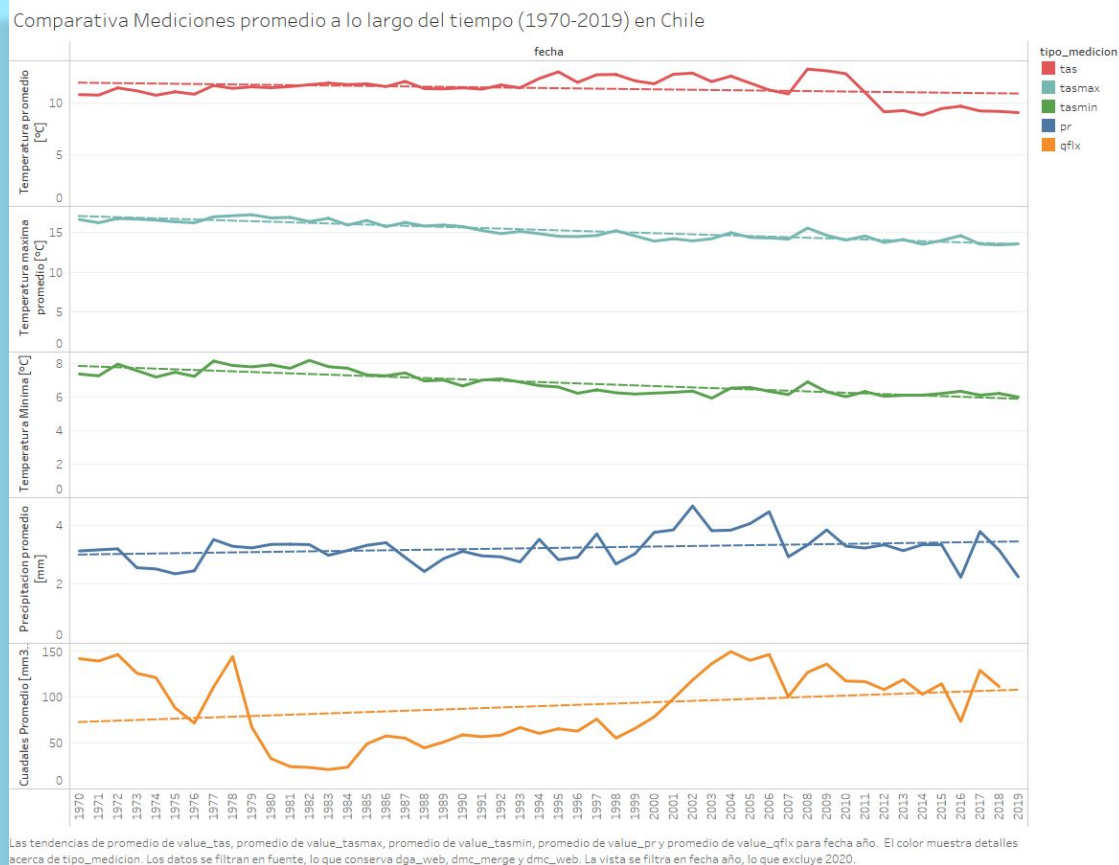
Nombres de medidas

- Precipitación promedio
- Temperatura promedio

Precipitación promedio San Joaquín (Beni, Bolivia)

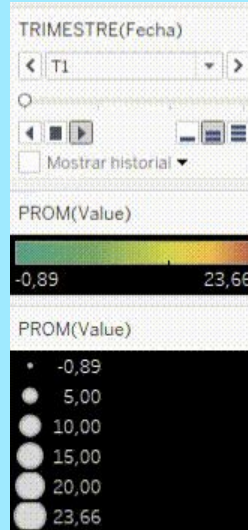


Comparación Mediciones promedio Chile



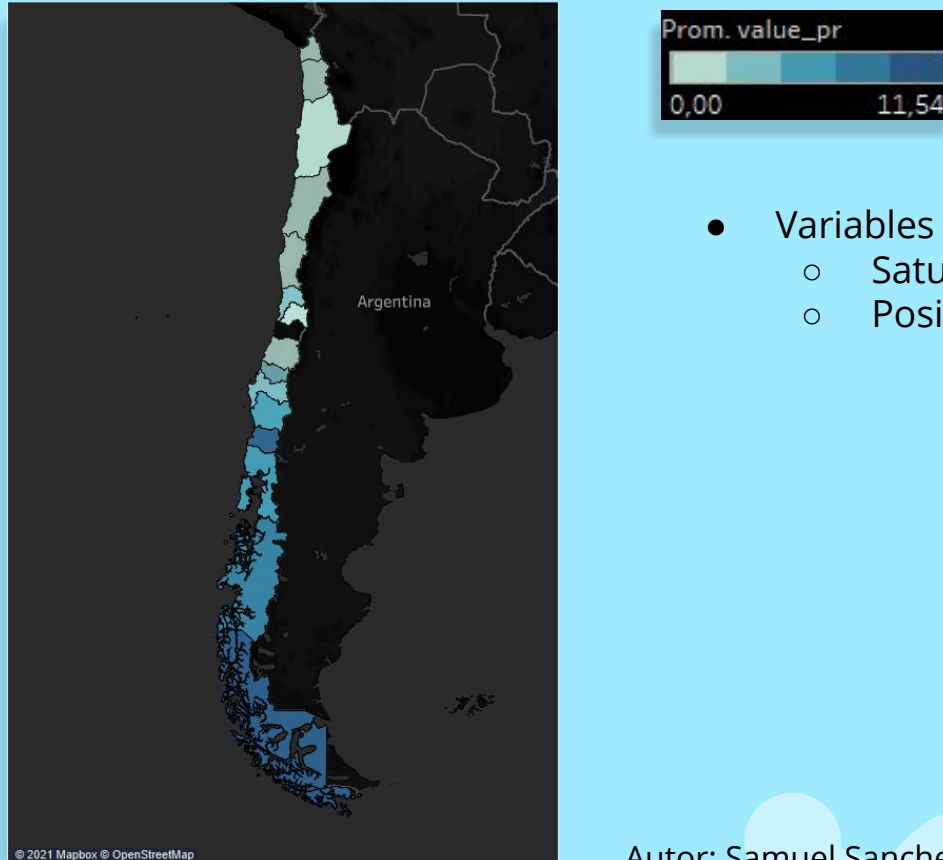
- Variables gráficas:
 - Color.
 - Posición.

Temperatura media a través del tiempo en Chile



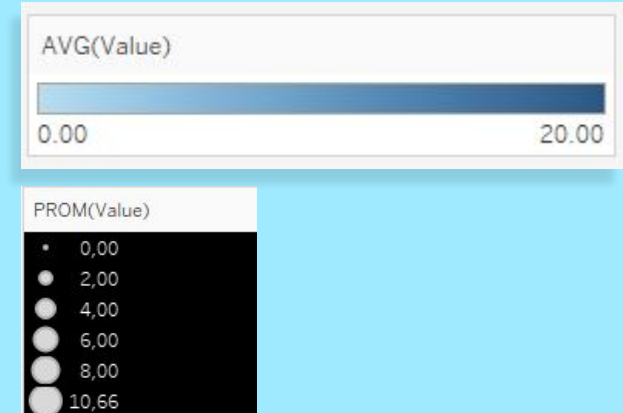
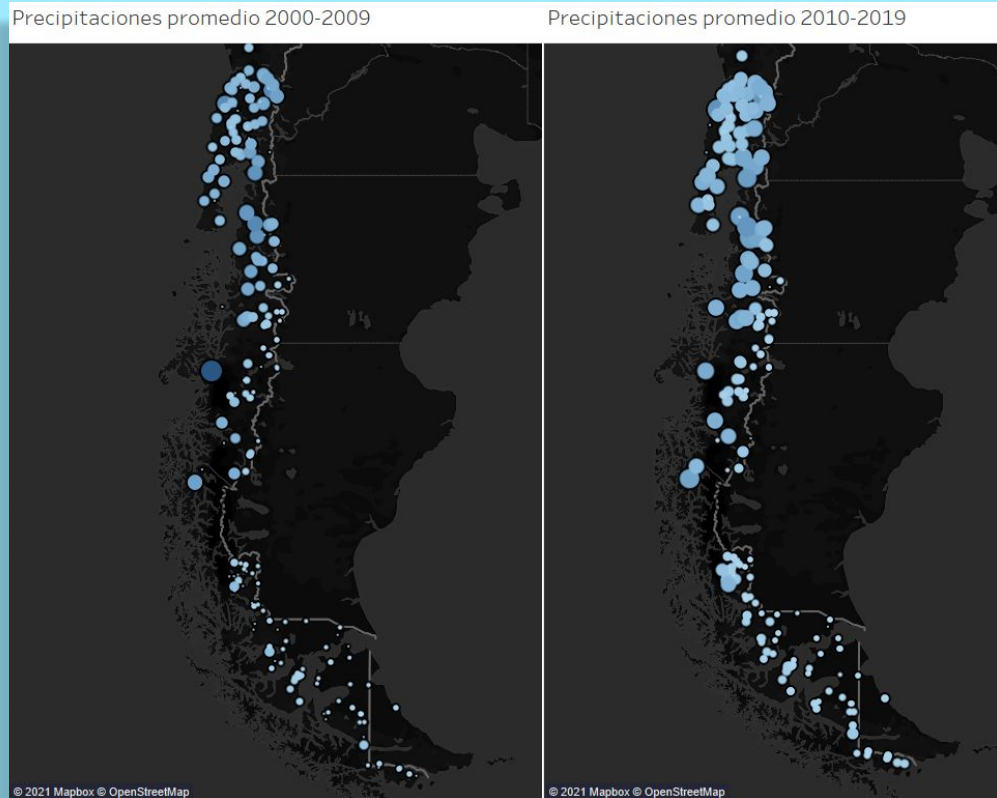
- Variables gráficas:
 - Color.
 - Posición.
 - Tamaño.
 - Tiempo

Precipitación Promedio Chile por región



- Variables gráficas:
 - Saturación.
 - Posición.

Precipitación promedio 2000-2009 vs 2010-2019



- Variables gráficas:
 - Saturación.
 - Tamaño.
 - Posición.

Autor: Felipe Sanhueza C.

Temperatura promedio 2000-2009 vs 2010-2019

Temperaturas promedio 2000-2009



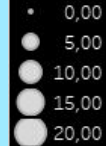
Temperaturas promedio 2010-2019



AVG(Value)

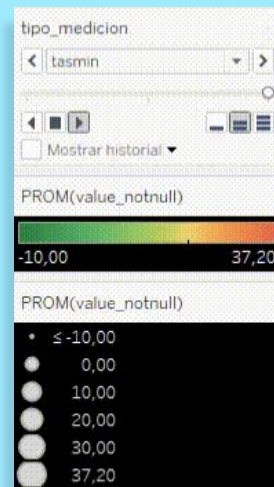
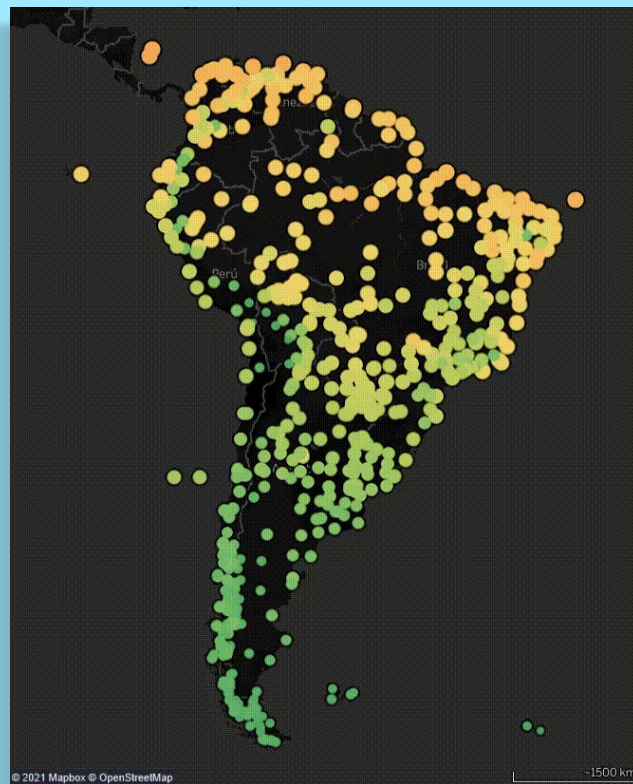


PROM(Value)



- Variables gráficas:
 - Tonalidad.
 - Tamaño.
 - Posición.

Comparación Temperaturas mínimas, promedio y máximas



- Variables gráficas:
 - Color
 - Tamaño
 - Posicion
 - Tiempo

Autores: Samuel Sanchez P. y Benjamín Bustos A.

Tareas a realizar

1. Continuar preprocesamiento hasta obtener un estado lo suficientemente pulcro en los datos.
2. Explorar comportamiento de los datos en momentos de anomalías climáticas o desastres naturales.
3. Explorar las hipótesis y ver si existe algún resultado más relevante ajeno a estas.
4. Confeccionar animación interactiva a partir de los datos para modelar el comportamiento del clima a través de los años.
5. Usar algoritmos de clasificación para describir futuras mediciones posibles.

Justificación

- Se comienza con estas visualizaciones sobre los datos ya preprocesados las cuales, dan una intuición inicial respecto a lo que nos gustaría obtener finalmente.
- Al comparar temporalmente 2 estaciones particulares, podemos observar fenómenos climáticos a través del tiempo.
- Por otro lado, en el mapa geográfico podemos ver como al agrupar las estaciones y sus mediciones, la visualización permite observar más claramente las condiciones climáticas de un momento temporal particular.
- Debido a la relación de la hipótesis, esperamos que al responder una de estas, obtengamos información relevante para complementar el análisis de las demás hipótesis.
- Finalmente, vale la pena mencionar que al explorar las hipótesis, se espera que se encuentren comportamientos o visualizaciones interesantes por lo que, estas podrían cambiar para la fase final.

Trabajo realizado.

- Analisis inicial: Trabajo conjunto en paralelo por todos los integrantes del grupo.
- Preprocesamiento: Trabajo individual con comunicación grupal de todos los integrantes.
 - 1ra Parte: Principalmente Benjamin y Felipe.
 - 2da Parte: Principalmente Maximiliano y Samuel.
- Hipotesis: Discusión y trabajo grupal.
- Visualizaciones: Mencionados anteriormente en cada visualización.
- Presentacion: Trabajo conjunto en paralelo y de forma individual por todos los integrantes.

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, infographics & images by **Freepik** and illustrations by **Stories**

Please keep this slide for attribution

Thanks!



Do you have any questions?

youremail@freepik.com

+91 620 421 838

yourcompany.com



Muchas gracias

