Main Question

How is mountain snow affected by air temperature?

Data Source

Air Temperature Data

- Data Source : Western juniper-dominated Experimental Catchments in south western Idaho, USA
- Metadata URL : https://data.gov/
- Data URL : https://ndownloader.figshare.com/files/44334659
- Content : Air temperature data of six places at south wester Idaho, USA in celsius, water_year, month, day calendar_year from 2007-2013.

	water_year	month	day	calendar_year	hour	sme2_ta_C	smf1_ta_C	smg1_ta_C	smg2_ta_C	smm1_ta_C	smm2_ta_C
1	2008	10	01	2007	00	9.0	9.4	9.7	9.2	8.8	8.7
2	2008	10	01	2007	01	8.1	8.7	9.0	8.4	8.6	8.1
3	2008	10	01	2007	02	7.2	8.4	8.8	8.0	8.4	7.6
4	2008	10	01	2007	03	6.6	7.8	8.3	7.5	7.8	7.1
5	2008	10	01	2007	04	6.2	7.7	7.9	7.1	7.5	6.7

Snow Depth Data

- Data Source : Western juniper-dominated Experimental Catchments in south western Idaho, USA
- Metadata URL : https://data.gov/
- Data URL : https://ndownloader.figshare.com/files/44334722
- Content : Snow depth data of six places at south wester Idaho, USA in millimeter, water_year, month, day calendar_year from 2007-2013.

	water_year	month	day	calendar_year	hour	sme2_sd_mm	smf1_sd_mm	smg1_sd_mm	smg2_sd_mm	smm1_sd_mm	smm2_sd_mm
1	2008	10	01	2007	00	0	0	0	0	0	0
2	2008	10	01	2007	01	0	0	0	0	0	0
3	2008	10	01	2007	02	0	0	0	0	0	0
4	2008	10	01	2007	03	0	0	0	0	0	0
5	2008	10	01	2007	04	0	0	0	0	0	0

Data Structure and Quality

These datasets are collected from authentic sources. Which is reflect the real world data and accurate. All dataset information is complete and very few missing values.

Data Source Licenses

All data are received from US government open sources and available for public use. The licence is us-pd and has no copyright restrictions.

Data Pipeline

The data pipeline, includes the following stages:

Technology Used

The data pipeline is implemented by Python using libraries like SQLite for data storage, Pandas for data manipulation andm os for interacting with the operating system.

In [ ]:
```
import pandas as pd
import sqlite3
import os
```

Transformation

The transformation stage is critical in preparing the data for analysis. Key steps in this process include:

1. Date Table Creation: Two comprehensive date tables are generated to ensure consistency in date-related data across both datasets. This function, create_date_table, creates a DataFrame with a range of dates, including various components like year, month, day, and quarter. This standardization aids in accurate temporal analysis and comparison. This table is joined with the covid and crime datasets to perform analysis.

2. Reanaming and deleting columns: Removes the unnecessary columns from the datasets, like water_year. and marge month, day, calendar_year colunm into a single column name date. it also reanames some columns for efficient use during analysis.

3. Data Type Conversions: The script includes conversions of data types, ensuring that each column in the datasets is of the appropriate data type for analysis. For instance, converting strings to dates.

The above steps ensure that the data is clean, consistent, and ready for analysis. This detailed process is crucial for obtaining reliable and accurate insights from the data.

In [ ]:
```
# URLs of the datasets
air_temperature_url = 'https://ndownloader.figshare.com/files/44334659'
snow_depth_url = 'https://ndownloader.figshare.com/files/44334722'
```

In [ ]:
```
# SQLite database file path
data_dir = './data'
db_path = os.path.join(data_dir, 'climate_data.sqlite')
```

In [ ]:
```
def run(self):
    # Ensure the data directory exists
    if not os.path.exists(os.path.dirname(self.db_path)):
    os.makedirs(os.path.dirname(self.db_path))

    # Connect to the SQLite database (it will be created if it doesn't exist)
    conn = sqlite3.connect(self.db_path)

    # Transform and save the air temperature dataset to SQLite
    self.save_to_sqlite(self.air_temperature_url, 'air_temperature', conn)

    # Transform and save the snow depth dataset to SQLite
    self.save_to_sqlite(self.snow_depth_url, 'snow_depth', conn)

    # Close the database connection
    conn.close()
```

In [ ]:
```
# Load the dataset
df = pd.read_csv(url)
```

In [ ]:
```
# Remove the "water_year" column if it exists
if 'water_year' in df.columns:
    df.drop(columns=['water_year'], inplace=True)
```

In [ ]:
```
# Rename air temperature column name
if 'sme2_ta_C' in df.columns:
    df.rename(columns={'sme2_ta_C': 'sme2 temperature in celsius'}, inplace=True)
if 'smf1_ta_C' in df.columns:
    df.rename(columns={'smf1_ta_C': 'smf1 temperature in celsius'}, inplace=True)
if 'smg1_ta_C' in df.columns:
    df.rename(columns={'smg1_ta_C': 'smg1 temperature in celsius'}, inplace=True)
if 'smg2_ta_C' in df.columns:
    df.rename(columns={'smg2_ta_C': 'smg2 temperature in celsius'}, inplace=True)
if 'smm1_ta_C' in df.columns:
    df.rename(columns={'smm1_ta_C': 'smm1 temperature in celsius'}, inplace=True)
if 'smm2_ta_C' in df.columns:
    df.rename(columns={'smm2_ta_C': 'smm2 temperature in celsius'}, inplace=True)
```

Challenges Encountered

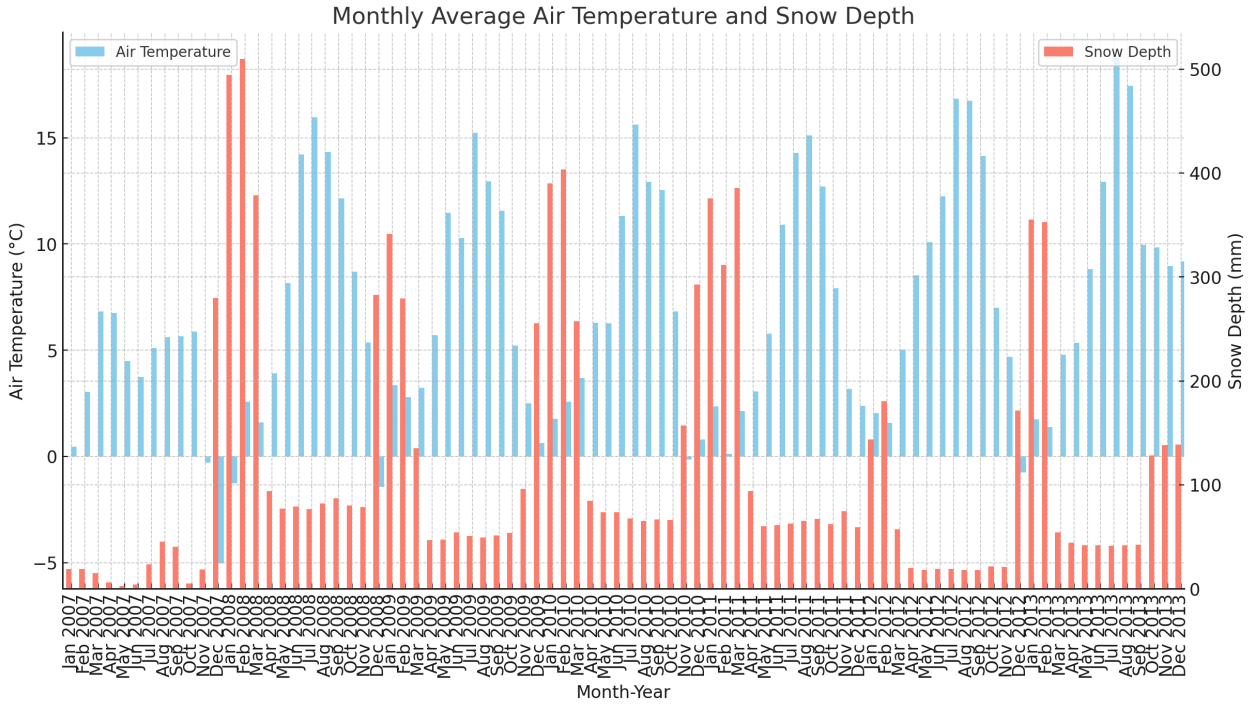
1. Incorrect data in the dataset, needed careful analysis to resolve the issue.
2. Handling missing data presented a significant challenge, requiring careful consideration to avoid wrong results.

3. Inconsistent data formats across sources, required standardization.

Results

The data is presented in the form of plots and bargraphs. These representations aim to depict trends and patterns without imparting any interpretation. Key visualizations include:

1. Two tables in the SQLite database represent the output of the data pipeline during the time of November 2007 to September 2013 air temperature increases, snow depth tends to decrease.
2. This inverse relationship is consistent with climate science, where warmer temperatures lead to snowmelt and reduced snow accumulation.
3. Warmer temperatures cause more rapid snowmelt, exposing bare ground.
4. Rising air temperatures lead to shifts in ecosystems and landscapes. Glaciers recede, permafrost thaws, and plant zones migrate, altering the face of our planet.



Limitations

1. Long tearm data required to get accurate result.
2. The study period may not be sufficient to establish long-term trends.