**Basic Info.** The project title, your names, e-mail addresses, UIDs.

Title: Utah Snowfall Comparison

Morgan Rasmussen, u1261516@umail.utah.edu, u1261516

Chase Maughan, u0962361@umail.utah.edu, u0962361

Ed Wright, u1378788@umail.utah.edu, u1378788

**Background and Motivation.** Discuss your motivations and reasons for choosing this project, especially any background or research interests that may have influenced your decision.

Salt Lake City was selected in late November by International Olympic Committee leaders as the preferred host for the 2034 Winter Games. This will be the second time the state has held the winter olympics.  The reason to host them in Utah is clear, Utah is globally known for having some of the best snow on earth. But with recent concerns over climate change, are our winters as great as they once were?  Will there even be enough snowfall to host the 2034 olympics? We plan to delve into the climate data to visualize and understand how snowfall in Utah has changed over the last 30 years.

**Project Objectives.**

Questions:

How has climate change affected snowfall in Utah?

Has there been a significant change in snowfall in Utah in the last 30 years, and how does it compare to average temperature in Utah in that area?

Was last winter (2022-2023), which felt much wetter and saw more snow, in line with the trend or an anomaly?

Can we use past and current winter snowfall trends to predict what the snowfall will be in 2034?

Learning objectives:

We want to be able to clearly demonstrate snowfall through graphs and statistical findings how snowfall over time has affected different Utah cities (Logan, Salt Lake City, and St. George)

We want to be able to statistically prove that there has or hasn't been a statistically significant change in temperature and snowfall.

We want to show that winter 2022-2023 was or was not an anomaly in snowfall compared to the previous 30 years.

We want to be able to use the data we obtained to be able to predict the snowfall in Utah in the year 2034.

**Data Description and Acquisition.** What format is your data in? How many items are there? What attributes do those items have? Are there special structures in it (e.g., networks, geographical)? From where and how are you collecting your data? If appropriate, provide a link to your data sources.

- This part should be specific enough that the instructional staff is assured you have or will be able to obtain data.
    - If it's online through direct download, link to the specific page from which you will download it.
    - If you will scrape it from the web, link to the page from which you will scrape it and a statement regarding how you have confirmed you are permitted to scrape it.
    - If you will use an API to access it, link to the documentation of the API and explain how you have access to that API
        - If it requires an account, state you have one.
        - If it doesn't require an account, state that it does not require one.
    - If it is data you have access to through other means, describe in detail what the data is, how you have access to it, and why you have permission to use it.

We obtained our weather data from oikolab.com. We gathered weather data from the past 30 years, with the mean temperature, snowfall, snow depth and total precipitation. It's all in csv files. To download these files we had to create an account at Oiklab and receive an API key. https://weatherdownloader.oikolab.com/downloader

**Ethical Considerations.** Complete a stakeholder analysis for your project.

- Who may be affected by your project and its outcomes? How could you project be used for harm?
    - "There are no ethical considerations" must be *strongly* defended. No one successfully done this before in this class.

If we were to say Utah was losing snow, people may be less likely to go skiing here, impacting the economy. It could also create a sense of urgency that would cause an excess of people to come to Utah. Alternatively, our data might show a change minor enough to not influence people's decisions.

Some stakeholders may include current residents of Utah, those who own or work for ski resorts, and prospective skiers from elsewhere. Depending on the outcome, these groups will have different reactions. None of these parties are necessarily in conflict with each other, but some of their interests may not align fully.

**Data Cleaning and Processing.** Do you expect to do substantial data cleanup or data extraction? What quantities do you plan to derive from your data? How will data processing be implemented?

Currently our working datasets will be pulled from the Oikolab weather API. These datasets are downloaded as three different CSV files corresponding to three different cities in Utah; Logan, Salt Lake City, and St. George. The principal variables in the dataset will be the monthly temperature, snow fall, snow depth, and total precipitation for each city spanning back 30 years. Each dataset will be around 360 rows long. To prepare the data for analysis and for plotting the data sets will need to be combined. Followed by identifying both null data points and possible outliers in the data. Depending on the completeness of the data we will either remove the null data points or replace them with the mean temperature for that given year. For outliers, we will identify them initially visually, but statistical tests may be necessary for further identification (using Z scores or Interquartile range). A majority of this data cleaning and processing can be completed using python scripts in jupyter notebook.

**Exploratory Analysis.** Which methods and visualizations are you planning to use to look at your dataset?

The primary focus of the project is to compare snowfall over time in the three different Utah cities. This can be simply visualized using a Simple Regression Plot, years on the X axis and snowfall on the Y axis. This Simple Regression Plot will also aid in identifying possible outliers in the data. To show a comparison between snowfall in the cities we will use an analysis of variance (ANOVA). Plotting each city against each other to show the difference in snowfall. Further analysis of the other variables, temperature, snow depth, and total precipitation can also be completed using these methods. If a particular year stands out as having more or less snowfall we will use a Chi Squared Test to test for significance, namely the 2022-2023 winter. We also plan on investigating methods to be able to predict snowfall over the next ten years.

The primary method for analyizing existing data is linear regression. We used this to train an autoregression model to predict future snowfall.

**Analysis Methodology.** How are you planning to analyze your data?

- What specific questions do you hope to calculate?
- What methods (from class or otherwise) do you think you will use?

We will be attempting to determine if climate change has a statistically significant effect on snowfall in Utah. We will use this conclusion to predict snowfall in the coming years, especially 2034. I suspect this will involve linear regression as the primary means of creating a model that relates the two. See Exploratory Analysis for more detail. Perhaps the most important part of the project, however, is that we need to present our findings in a clear and non-misleading way.

**Project Schedule.** Make sure that you plan your work so that you can avoid a big rush right before the final project deadline, and delegate different modules and responsibilities among your team members. Write this in terms of weekly deadlines.

Week of March 17: Compile and Clean Data

Week of March 24: Compile and Clean Data

Week of March 31: First Milestone

Week of April 7: Make graphs/other visuals, hypothesis test

Week of April 14: Make the video/finish project


Additional Info:


https://utahavalanchecenter.org/alta-monthly-snowfall

https://wcc.sc.egov.usda.gov/reportGenerator/edit/customMultiTimeSeriesGroupByStationReport/monthly/start_of_period/856:UT:SNTL%7Cid=%22%22%7Cname/POR_BEGIN,POR_END/SNWD::average_1991

https://oikolab.com/

Github Repository

Data Acquisition and Description

We got our data from Oikolab. They get their data from various sources, such as NOAA, GEFS, CHIRPS, and ERA5. The data came in a csv file, with a variety of variables to choose from, all of which we explored in our autoregression file, in which we decided which variables to focus on in our correlation matrix.

Milestones:

Linear regression models

Previously done

1. Preprocessing data
2. Encoded categorical data
3. OLS Summary for each city
4. Mean snowfall for February by year by city graphs

Need to do

1. Add a regression line to plots
2. Change x-axis ticks
3. Add comments to this code

Auto regression models

Previously done

1. Preprocessing data
2. Encoded categorical data
3.  Fit an auto-regression model with the SLC all Param data
4. Get a predicted snowfall for Feb 10 2034 (which day is arbitrary)
5.

Need to do

1. Get an accuracy for regression model/other statistical analysis of the autoregression
2. Find another model to make a comparison
3. Add bullet for prediction for feb 2034 to other plots
4. Possibly run other city datasets through the auto regression model.
5. Clean code

Data

Previously done

1. Reformat data/other preprocessing.
2. Create a correlation matrix to identify which variable correlates most with Snowfall in SLC
3. Find historic mean, upper limit, lower limit  for Feb 1994-2024 in SLC

Need to do

1. Show statistical significance or insignificance of data and plots.
2. Make sure all datasets cover the same years.

Okay we need to upload the data, what's above this in one document, and then the peer review in a different document

Peer Feedback

Auto regressive order 1, t-1, testing residuals, lag it by one year, minimize residual and error term, find data that actually affected snowfall, estimate, given this equation, what is our estimate, if residuals are within epsilon error, high r^2 and residuals are very small

[Autoregressive (AR) Models Python Examples: Time-series Forecasting - Analytics Yogi (vitalflux.com)](#)

[Autoregressions - statsmodels 0.15.0 (+222)](#)

[Autoregression Models for Time Series Forecasting With Python - MachineLearningMastery.com](#)

```
from statsmodels.tsa.ar_model import AutoReg
```

OLS versus auto regressive, AR1, AR2, AR3, AR3 is the best one

Lake effect, temperature

Independent is time/temperature

Gaussi, gamma, fit data to distribution curve, second or third standard deviation away to find outliers

Causality, why below or above that much snowfall, bassian

Climate change use all three, Salt Lake City for Olympics 2034

Look at data to see how accurate it is

Look at 2002, did they have enough snow, this is how much snow vail resorts is getting

Data for park city

Data collection for oikolab

Justify if you have a low r^2

What cities matter for Olympics? Get data for them

Get the data or change the story that we want to tell

Logan and st George, runoff going to the Colorado, feeding Las Vegas and California, see if that works, policy recommendations

Organize data into panel data, 1-slc 2-park city 3-sandy

Heat exchange with C, cloud formation, momentum in the atmosphere

Planting trees, where to be environmentally impactful

Do a literature review, or Bao will rip you apart

Add to a study that's already there

Aggregate supply and demand water to Utah

Another ethical consideration is water rights

Correlation matrix

Less time, with more variables

DEEPNOTE


Josh, Isaac, Bella