I. Research Question

This analysis seeks to answer the question if states west of the continental divide receive a larger proportion of their annual snowfall during the period of January and February than states east of the continental divide. The hypothesis is that States west of the divide, defined here as Utah and Idaho, receive the majority of their snowfall in the early winter months, January and February, whereas states east of the divide, defined here as Colorado and Wyoming, receive the majority of their snowfall in the late winter months, March and April. Due to the higher elevation of states west of the continental divide, one would assume that they receive more snowfall during the entire winter. However, this fails to take into account the geography preceding Colorado and Wyoming. The Great Basin, a basin formed by the borders of a number of different mountain ranges creates the meteorological conditions for Utah and Idaho to receive more snowfall in the early winter months. There has been some study of this question, Zhang et. al. published a paper that sought to answer if snowfall is diverging between mountains and plains.[1] But, many of these studies have not been updated with contemporary weather data from the 21st century, indeed Zhang et al. is the only paper that incorporates data from the 2010s. Given the increasing interest and impacts of global warming, there has been little study to determine the impact of the decreases in snowfall. This is particularly salient because states in the Intermountain West rely heavily on winter snowfall to feed their reservoirs for the coming year until the snows return in the fall. Additionally, there has been some research comparing

[1] Jiyang Zhang, Haochi Zheng, Xiaodong Zhang, and Jeffrey VanLooy. 2020. "Changes in Regional Snowfall in Central North America (1961–2017): Mountain Versus Plains." *Geosciences* 10 (157): 157. doi:10.3390/geosciences10050157.

snowfall and rainfall trends across all states west of the Rockies, with the authors concluding that finding that snowfall decreased and rainfall increased in January and March.[2]

II. Data Collection and Cleaning

The data for this analysis was pulled from NOAA's National Centers for Environmental Information. The data came in a proprietary format that needed to be converted to a machine-readable format (CSV in this case). The goal of this analysis is to determine if different areas receive more snow during different parts of winter. One issue that arose early on was whether temperature should have a place in this analysis, it was decided that temperature would not be a variable due to the impact of windchill as thermometers often read at or above 32°F, yet snow would be falling and accumulating. One issue with the data from National Centers for Environmental Information (NCEI) is there are a significant number of missing observations where no reading was taken and additionally there are a couple of codes with M indicating a missing observation and T indicating a trace observation. Trace observations are an effort to acknowledge that there was some precipitation that fell, yet the precipitation did not amount to anything measurable.

After the station and location data was downloaded from NOAA, it was imported and cleaned down to a list of stations in Colorado, Idaho, Utah, and Wyoming. Using the list of stations the weather data was collected using a for loop to download the files and then convert them from the '.dly' format to a '.csv' format. From here, it was filtered down to only the months present in the study and observations containing snow measurements. The weather data

[2] Knowles, Noah, Michael D. Dettinger, and Daniel R. Cayan. 2006. "Trends in Snowfall versus Rainfall in the Western United States." *Journal of Climate* 19 (18): 4545–59. https://search-ebscohost-com.ezproxy.pratt.edu/login.aspx?direct=true&db=edsjsr&AN=edsjsr.26259253&site=eds-live&scope=site.

was then combined into a single file and transposed from wide form to long form using openrefine. It was at this point that an effort was made to input some data to account for the missing observations. However, due to the nature of this analysis, a relatively small number of weather stations spread out over a large geographic area, the variogram was unable to offer predictions for the entire grid area. This was because for many areas, there was no data to create a model off. Before the kriging process could begin it was necessary to filter down the weather data removing NA values

III. Methods

After the kriging was unsuccessful, the analysis went ahead by beginning with subsetting the data into four separate groups, January and February for Utah and Idaho, January and February for Colorado and Wyoming, March and April for Utah and Idaho, and March and April for Colorado and Wyoming. After the data was separated some tests to ascertain the distribution of the data were ran. All four subsets failed, so a MannWhitneyU test was used to test the hypothesis.

IV. Results

The first part of the hypothesis, Utah and Idaho receive a higher proportion (>50%) of total snowfall in January and February than Colorado and Wyoming do in January and February, resulted in a p-value of 0.34, thus failing to reject the null hypothesis that the distribution between the two was equal. The second part of the hypothesis, Colorado and Wyoming receive a higher proportion (>50%) of total snowfall in March and April than Utah and Idaho do in March and April, resulted in a different outcome than the first. With a p-value of $2.44 * 10^{-43}$, the null

hypothesis was rejected and it was shown that Colorado and Wyoming receive a higher proportion snowfall in March and April than Utah and Idaho do in the same period.

V. Discussion

While part of the hypothesis was proven using MannWhitneyU, there are more factors present, most notably the differences in elevation between the two sides of the continental divide, with much of Colorado being high alpine climate whereas much of Utah is high desert, which warms up faster than the higher elevations. Additional analysis would perhaps utilize a spatially-integrated regression model to incorporate more factors to more fully understand weather patterns. A longer time horizon would also offer a more complete picture of snowfall distributions.