OUTLIER DETECTION IN DEPTH OF SNOW DATA

BY: SCOUT JARMAN

UTAH STATE UNIVERSITY

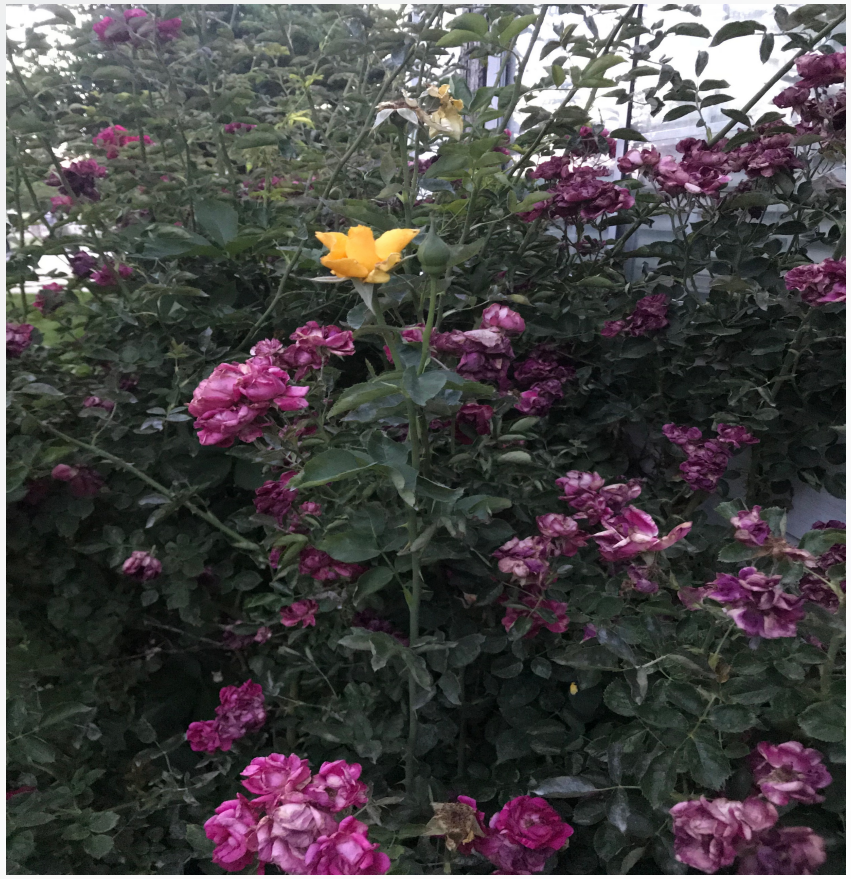
STUDENT RESEARCH SYMPOSIUM 2020

INTRODUCTION: WHAT AND WHY

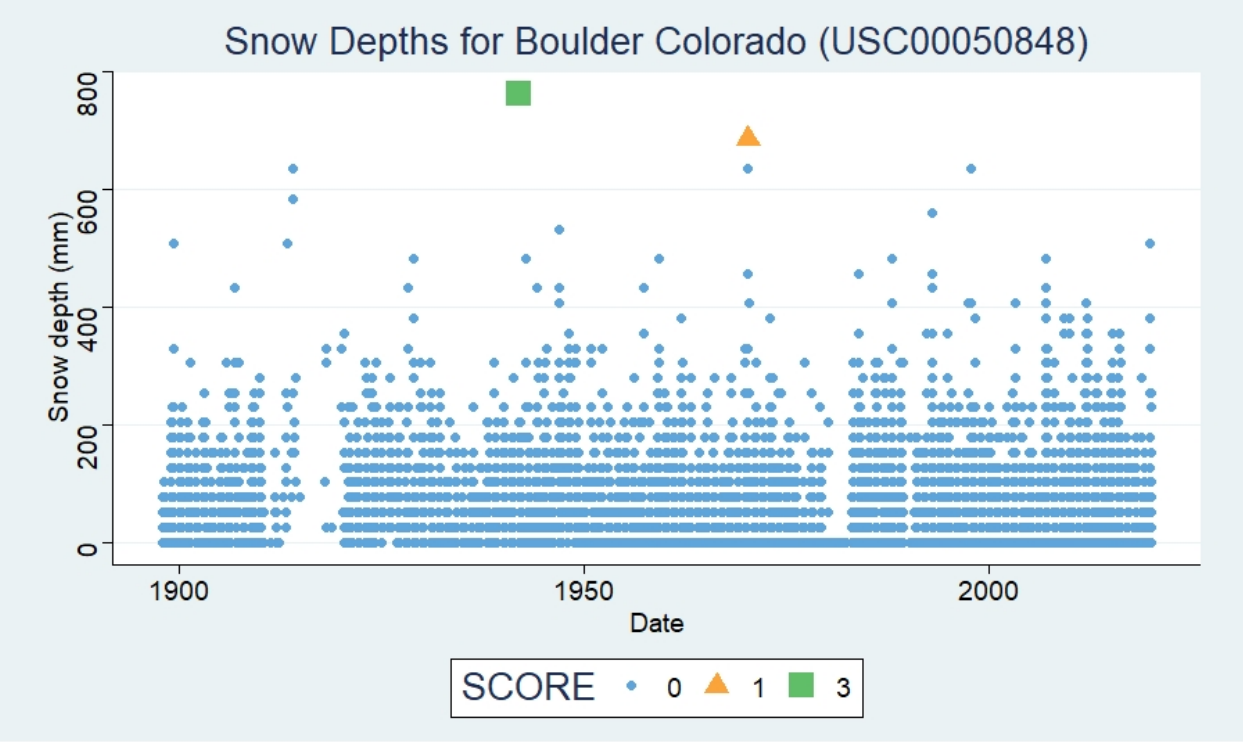
- Buildings need to be built to withstand heavy snow falls.
- A data driven solution is needed to create a national snow map.
- A balance between safety and costs is need, and outliers increase cost.

INTRODUCTION: WHAT AND WHY

- Outliers are points in a data set that are anomalous and are usually a result of a recording error.
- Weather stations across America record climatological data including the depth of snow. These records are prone to transcription error.

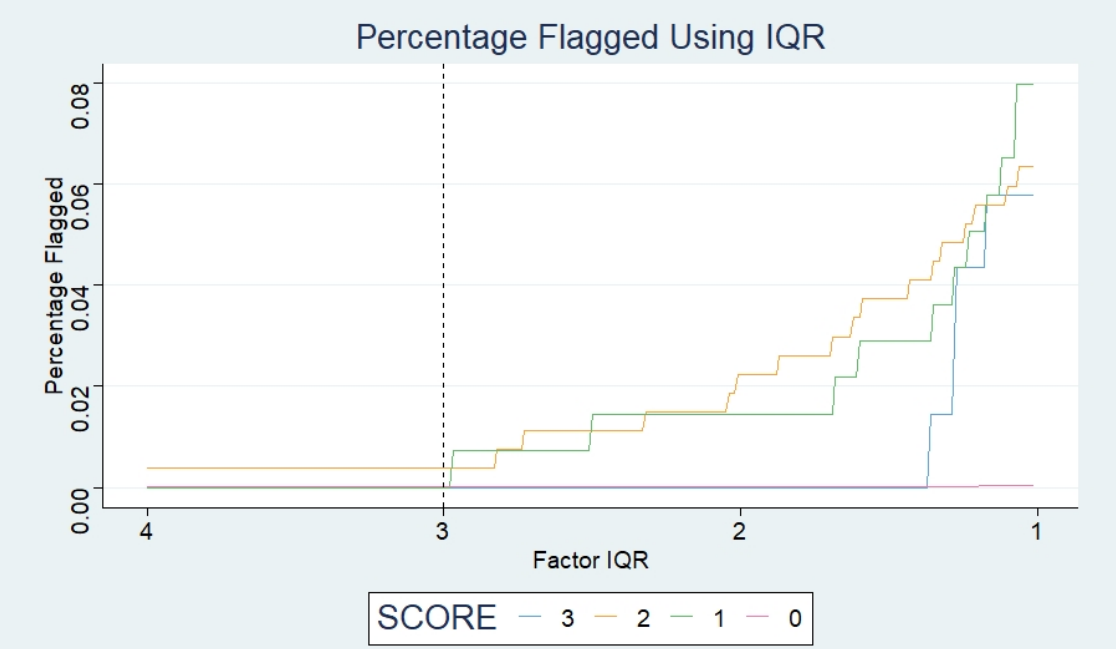


DATA: COLLECTION AND FLAGGING

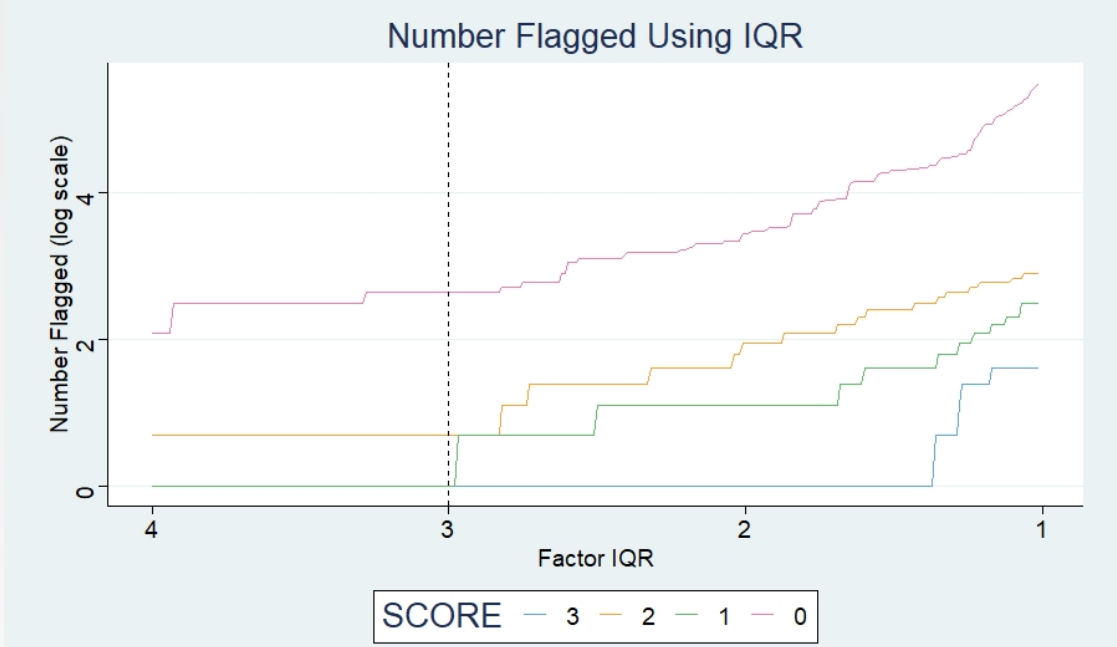


- Stations came from Washington, California, Colorado, Montana, and New Hampshire.
- Outliers were visually determined, and scored from 0 to 3, with 3 being an outlier, and 0 not being an outlier.

INTERQUARTILE RANGE METHOD



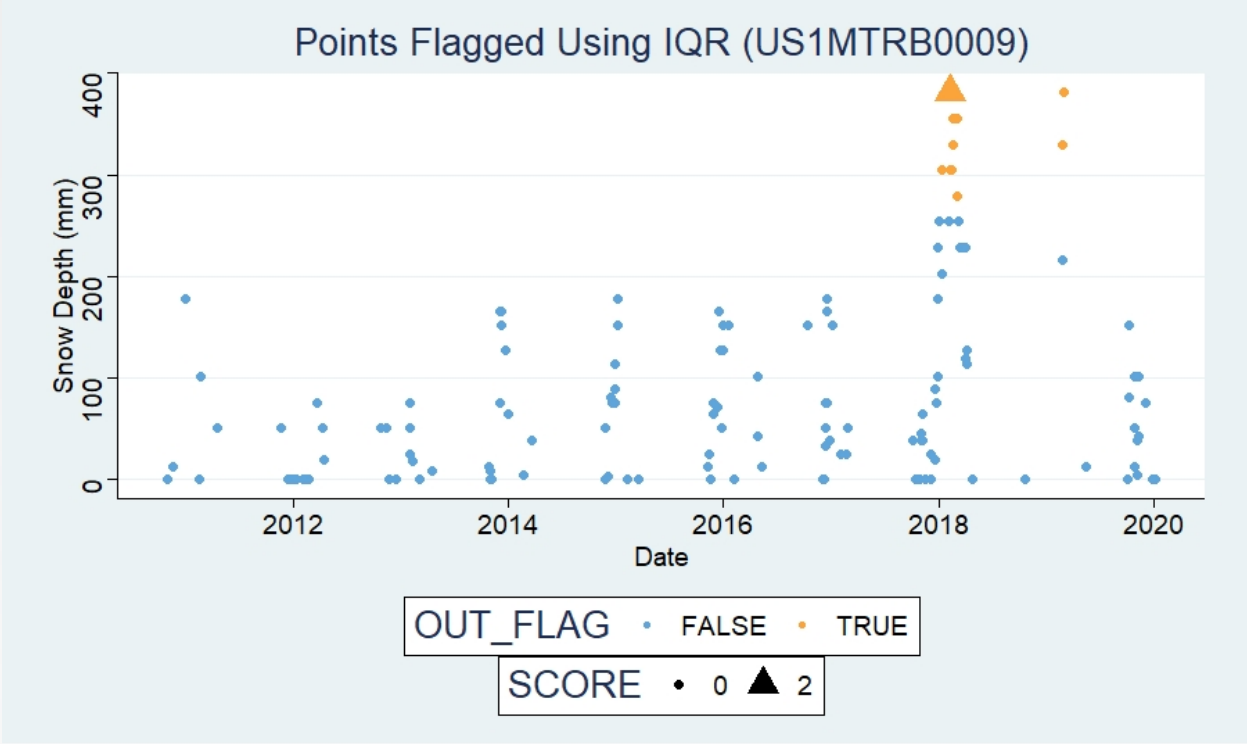
PERCENTAGE OF OBSERVATIONS FLAGGED USING DIFFERENT FACTORS OF THE IQR FOR EVERY SCORE OF OUTLIER. USING A FACTOR OF 3, NO 3'S WERE FLAGGED.



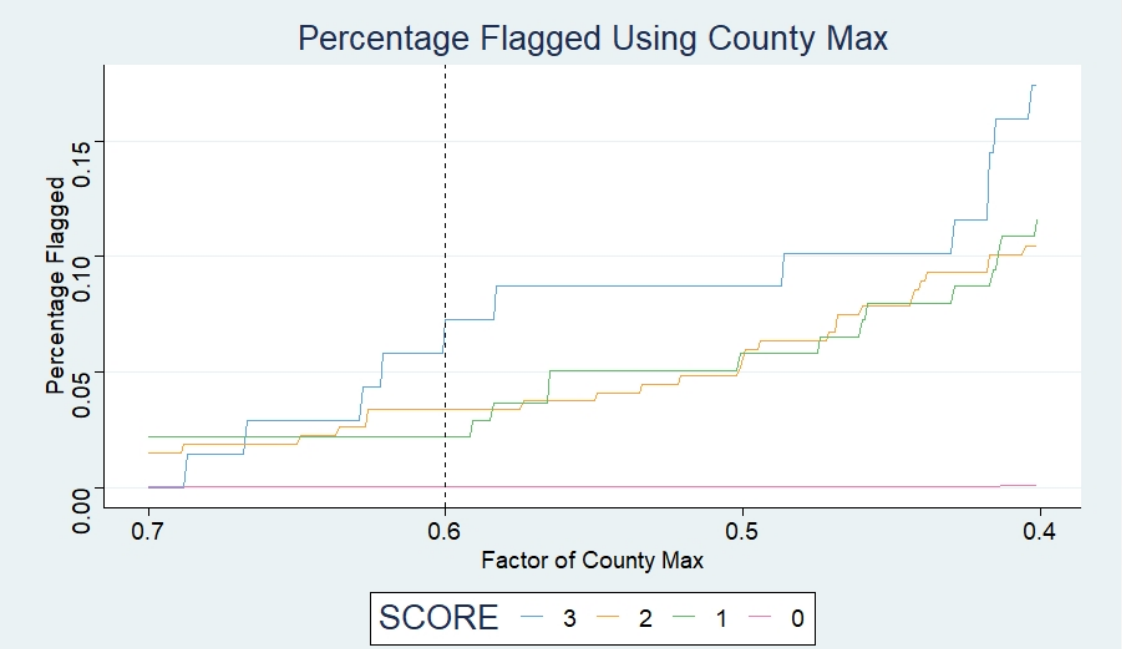
THE LOG NUMBER OF OBSERVATIONS FLAGGED USING DIFFERENT FACTORS OF THE IQR FOR EVERY SCORE OF OUTLIER. USING A FACTOR OF 3 FOR THE IQR, MOSTLY 0'S AND 2'S ARE FLAGGED

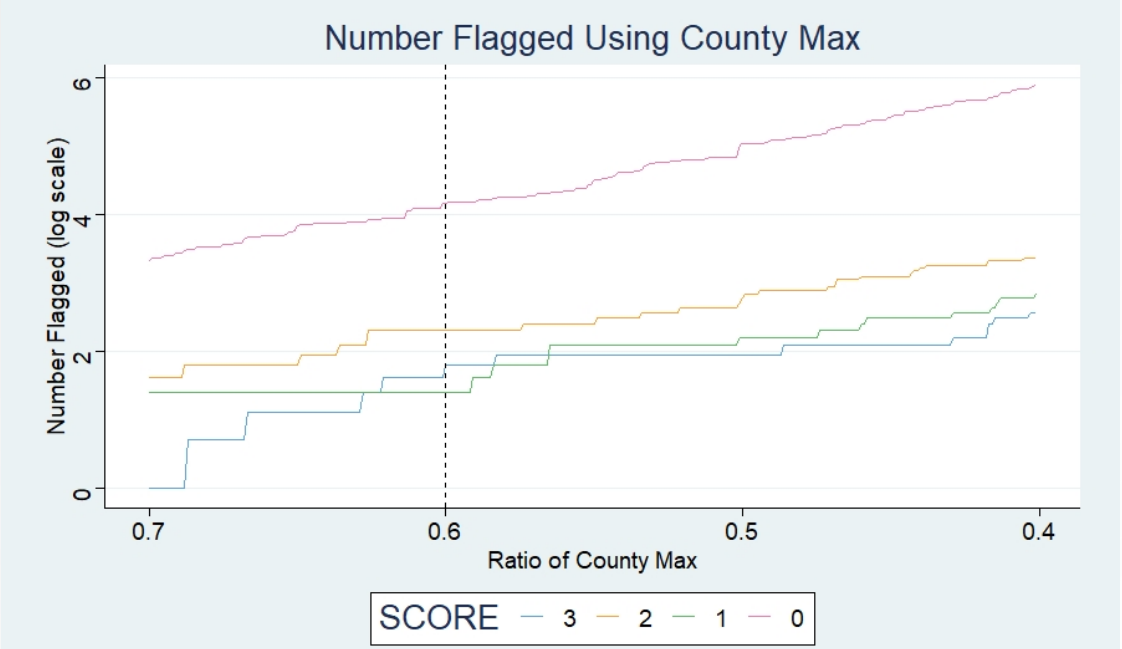
INTERQUARTILE RANGE METHOD

- These were the only points flagged using 3*IQR.
- Pros: simple, self contained, and conservative.
- Cons: Does not catch any 3's, perhaps too conservative.



COUNTY MAX METHOD

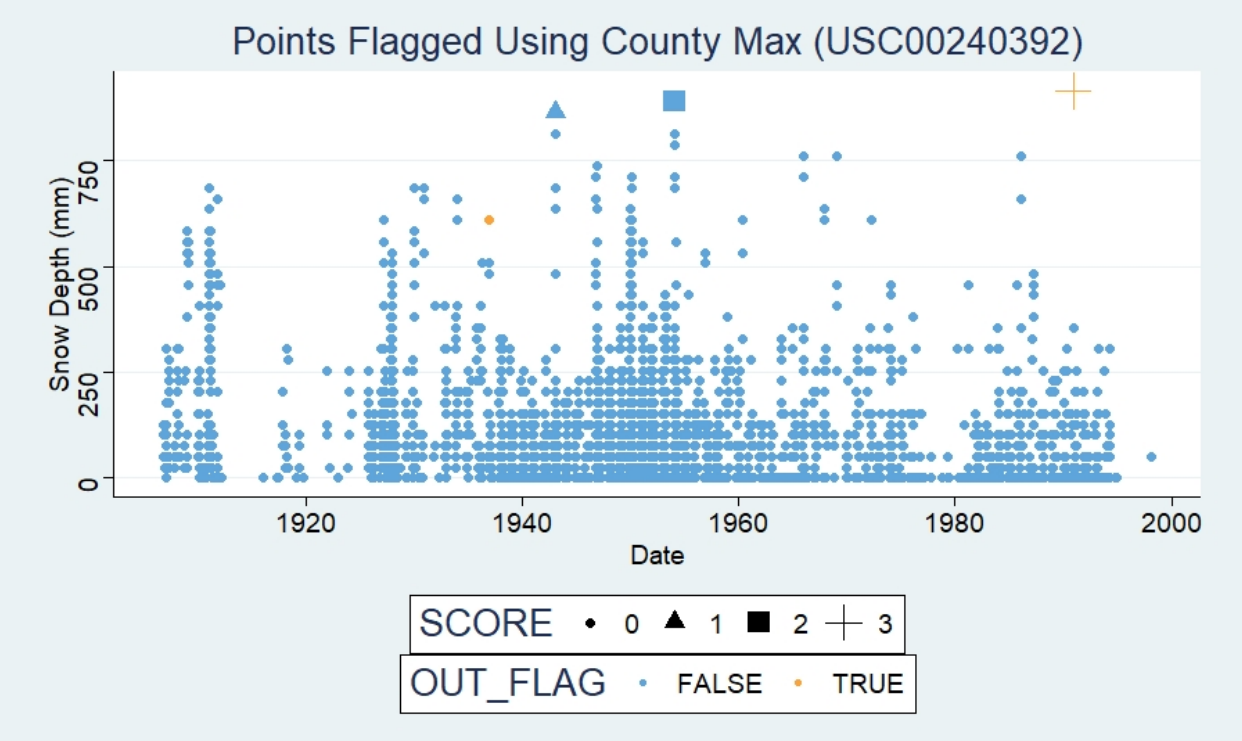




PERCENTAGE OF OBSERVATIONS FLAGGED USING DIFFERENT FACTORS OF THE COUNTY MAX FOR EVERY SCORE OF OUTLIER. USING A FACTOR OF .6, ABOUT 7% OF 3'S ARE FLAGGED.

THE LOG NUMBER OF OBSERVATIONS FLAGGED USING DIFFERENT FACTORS OF THE IQR FOR EVERY SCORE OF OUTLIER. USING A FACTOR OF 3 FOR THE IQR, MOSTLY 0'S AND 2'S ARE FLAGGED

COUNTY MAX METHOD



- Here is an example from Babb Montana.
- Pros: Catches more 3's than IQR
- Cons: Must find county max data, and if county max data if wrong, then this method is incorrect.

FUTURE WORK: PATTERN RECOGNITION

- Four distinct patterns were found; SS, SL, LS, and LL. Where S means large changes in snow depth in a short amount of time, and L means changes in snow depth over a long period of time.
- Most 3's and 2's were SS and SL. Here are 2 examples of SL.