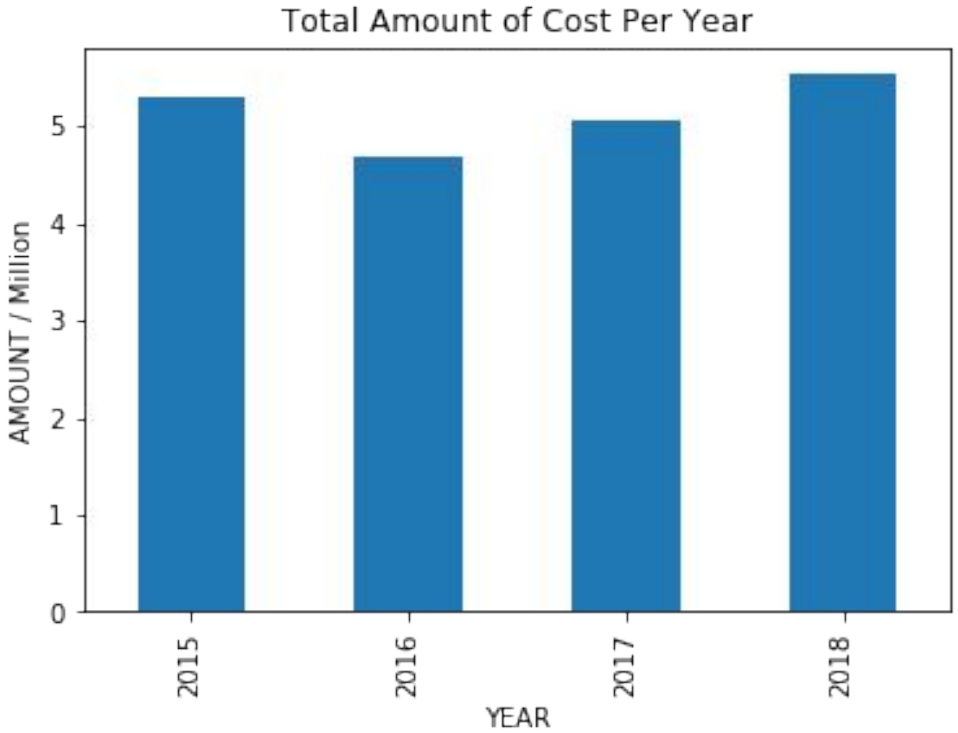
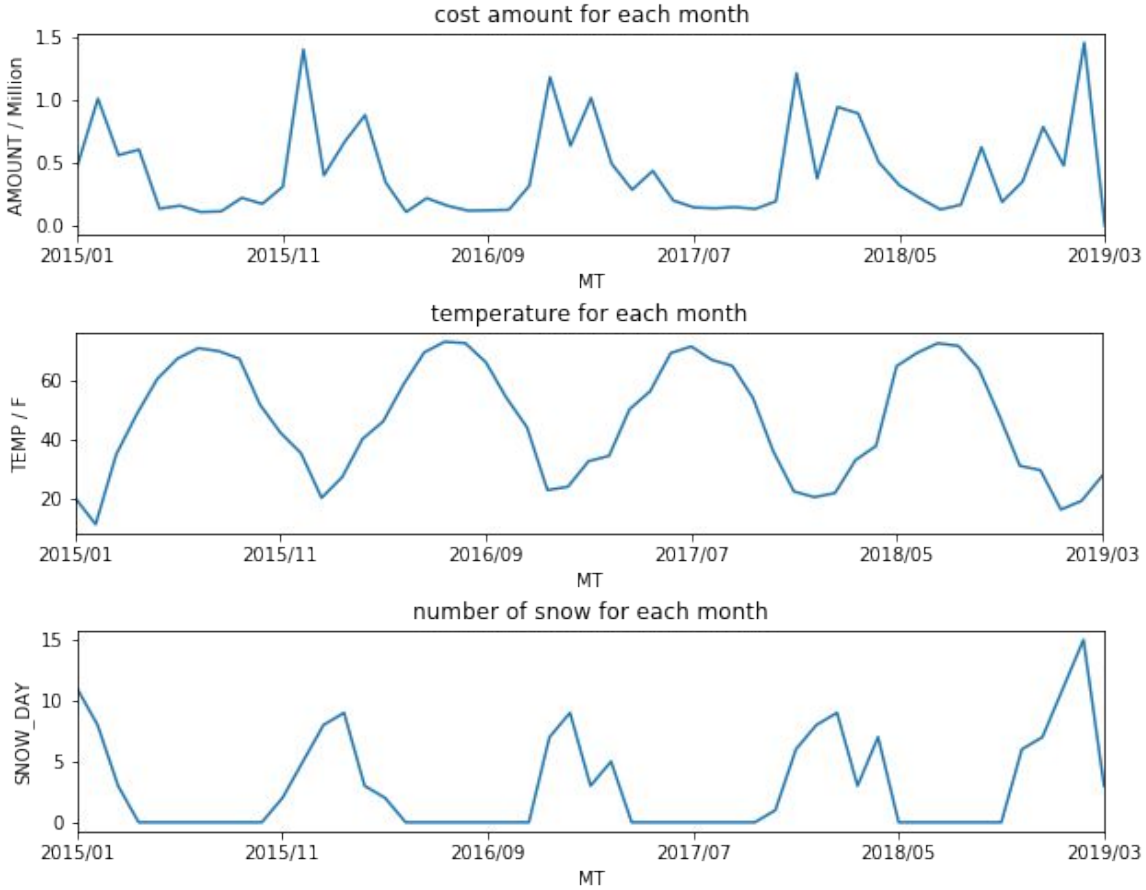
Relationship Between Spending and Weather in Madison

Swaraj Rao
Whitney Long

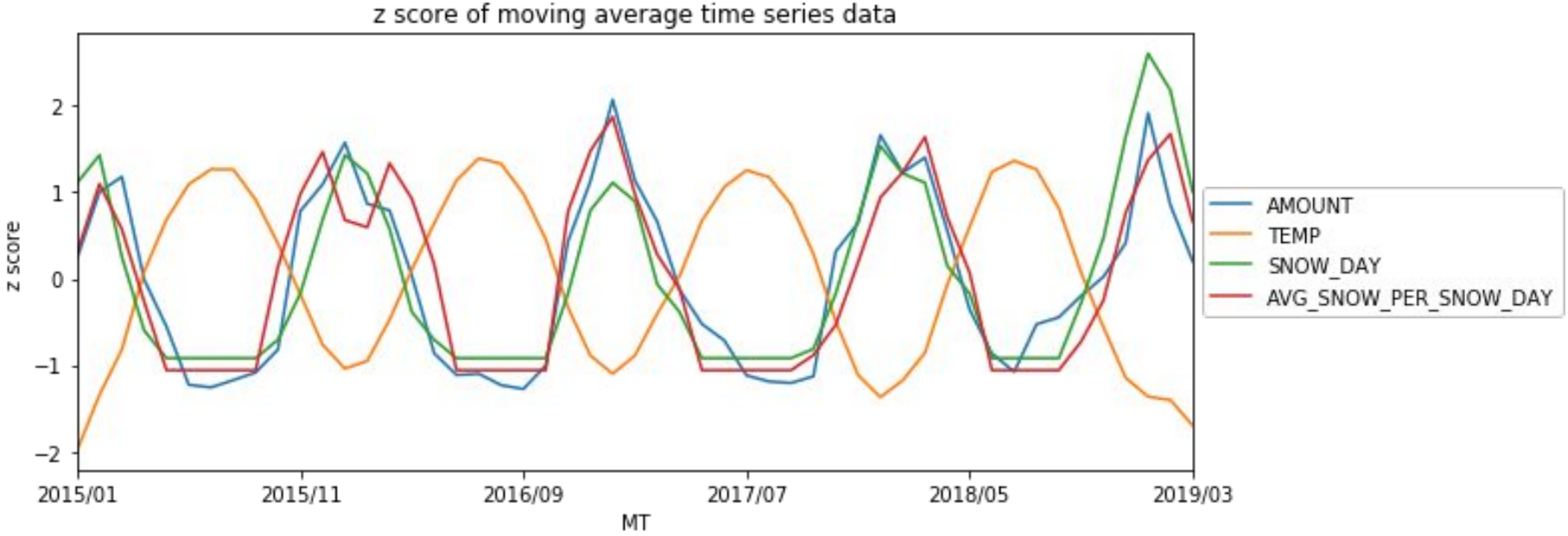
Total Expense in Different Years



Overview of historical cost, temperature, and snow

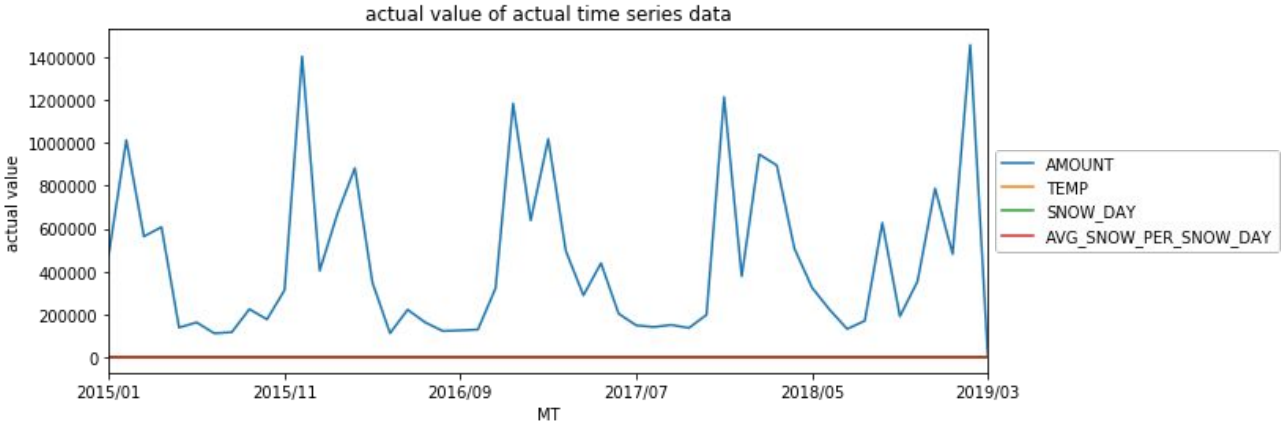


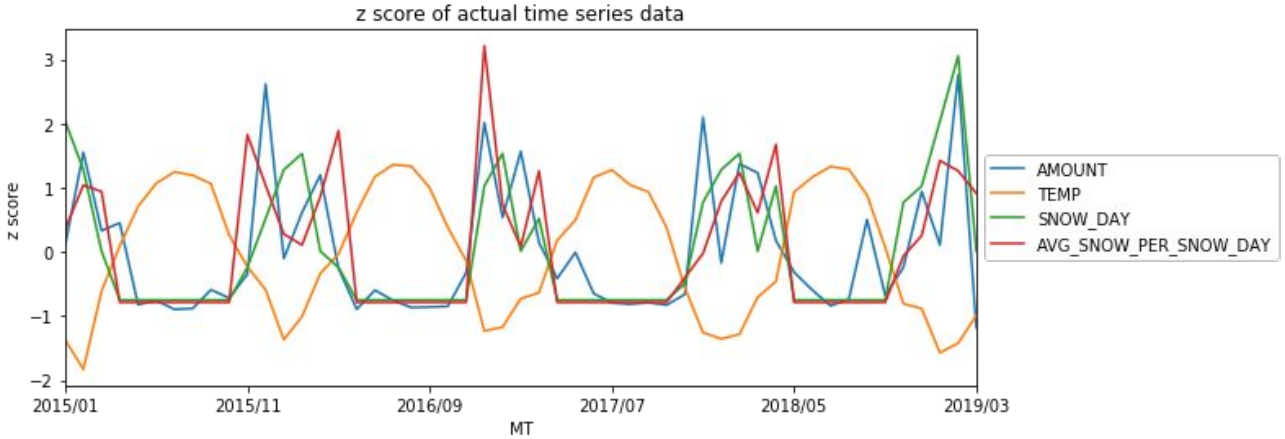
Data Visualization of Four Examined Variables

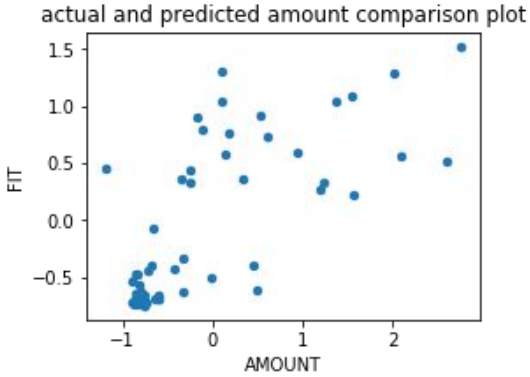


- Instead of using the actual value for each variable, z score, which is computed by minus the mean and divided by standard deviation, is used so that the values are in the same scale.
- Considering the possibility of prepaid and postpaid payments, the real cost amount of one month is estimated by averaging the recorded amount of two adjacent months and itself.

Why Z-score? Why Moving-average?

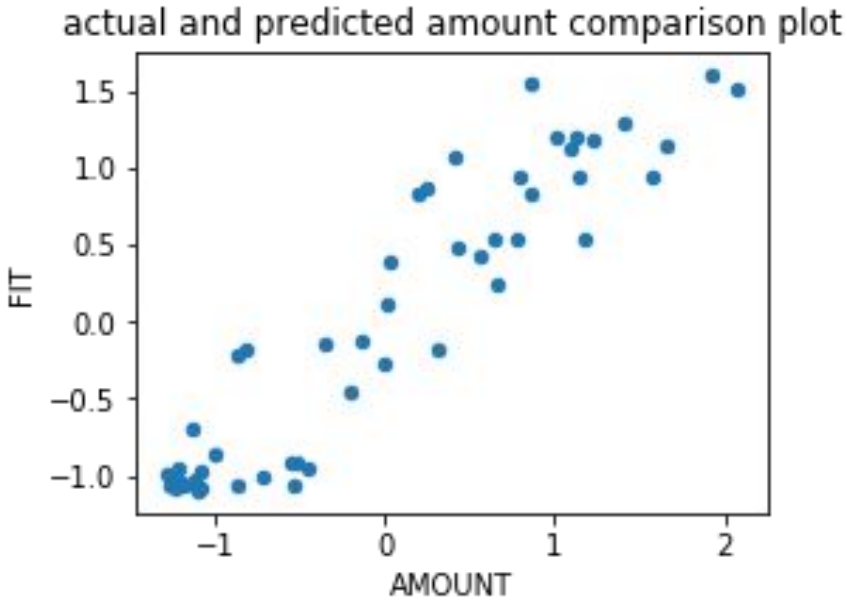






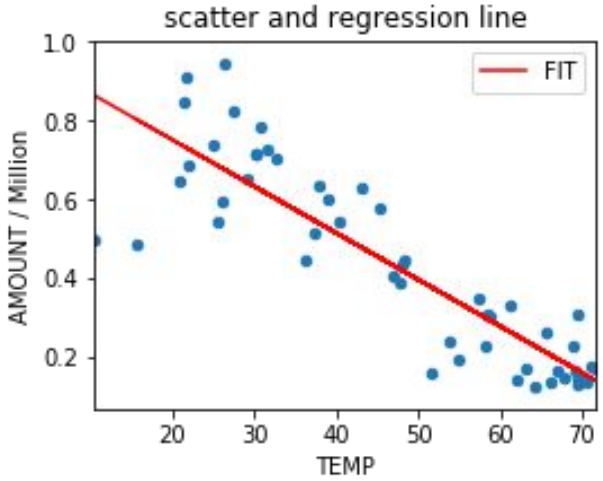
Standard Deviation of the difference between actual z score and predicted z score is 0.7097
Regression without moving average

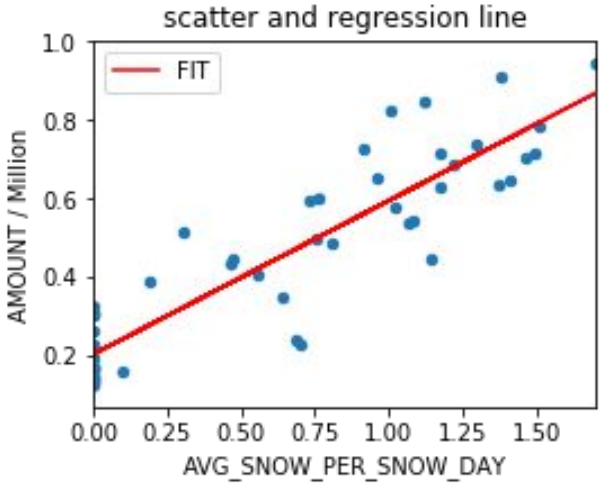
Linear Regression of Four Variables

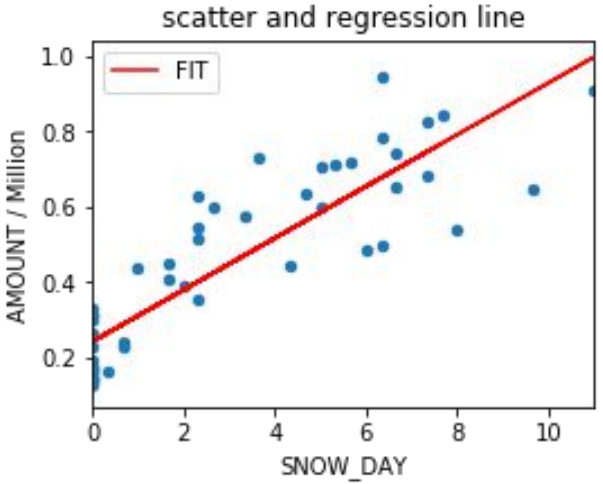


AMOUNT = -0.2483 * TEMP + 0.2127 * SNOW_DAY + 0.5186 * AVG_SNOW_PER_SNOW_DAY
Standard Deviation of the difference between actual z score and predicted z score is 0.3582

Linear Regression of Each Variable:
Temperature, Average Snow Per Snow Day, and the Total Number of Snow Day







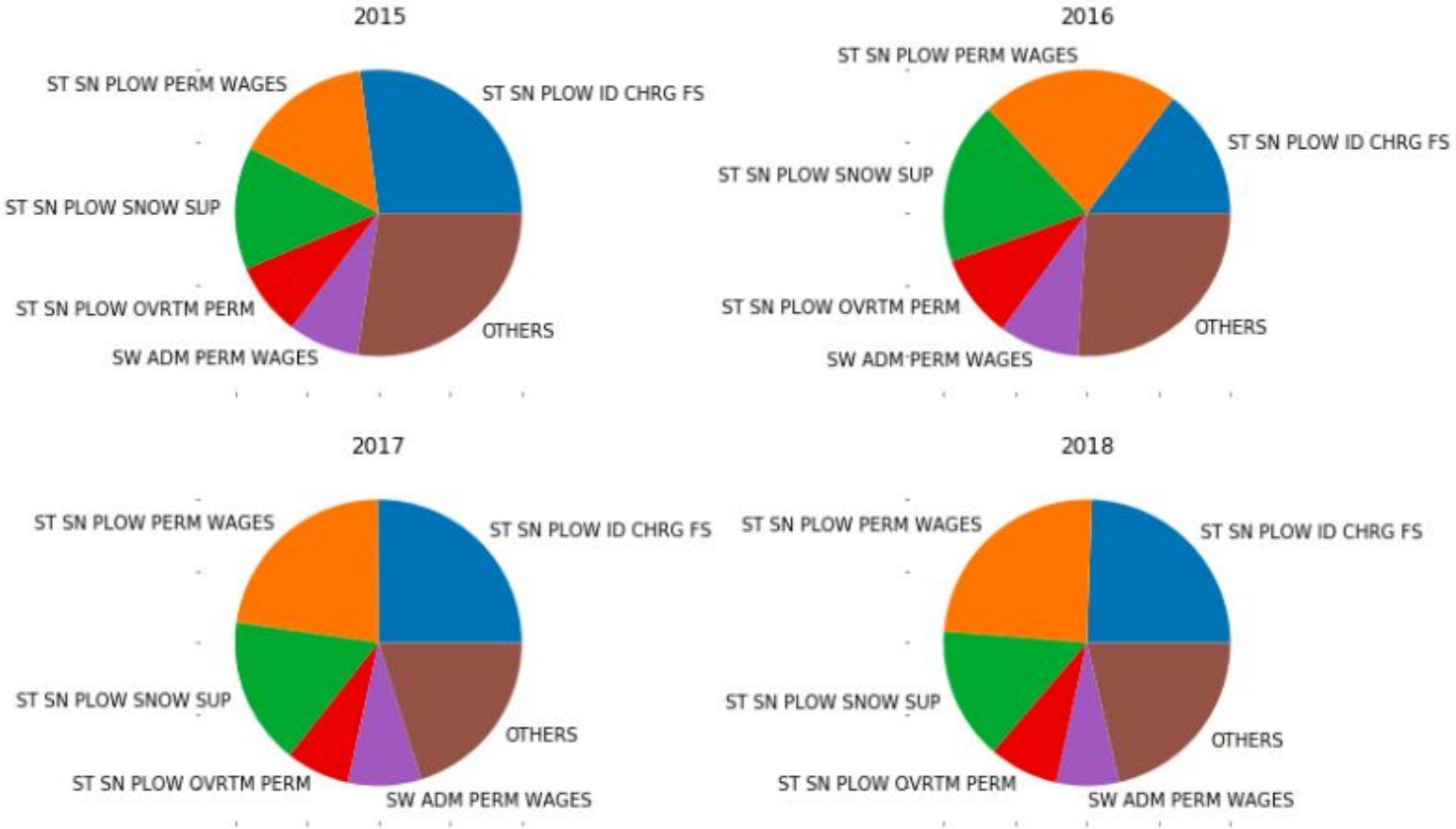
Std Err: 0.1193

Std Err: 0.1054

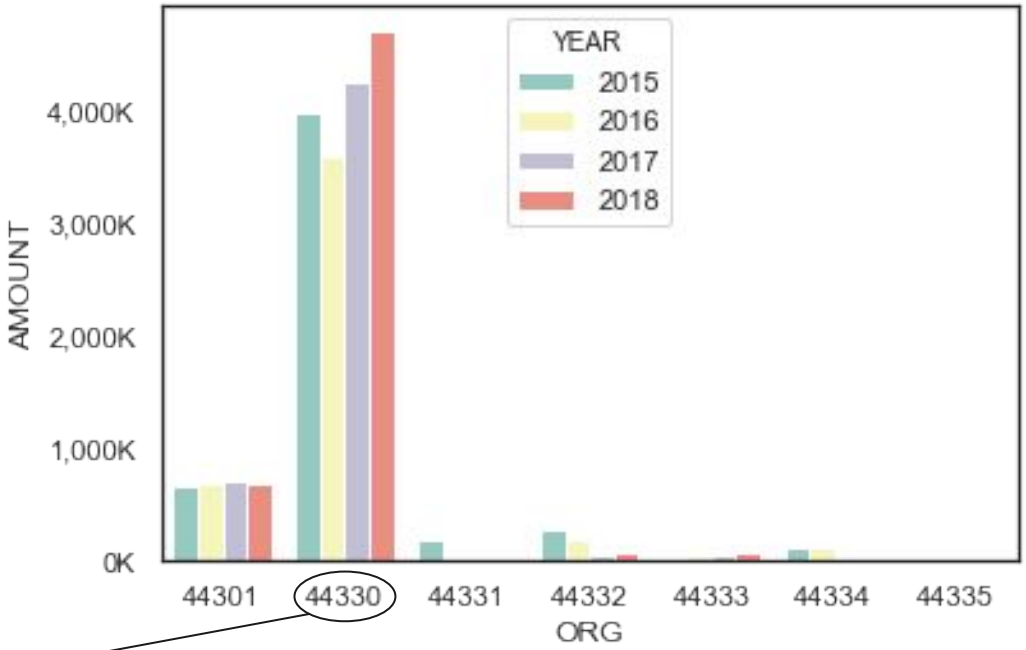
Std Err: 0.1186

AMOUNT = -0.0118 * TEMP + 0.9868
AMOUNT = 0.3910 * AVG_SNOW_PER_SNOW_DAY + 0.2032
AMOUNT = 0.0689 * SNOW_DAY + 0.2407

Proportion of Cost in Different Categories

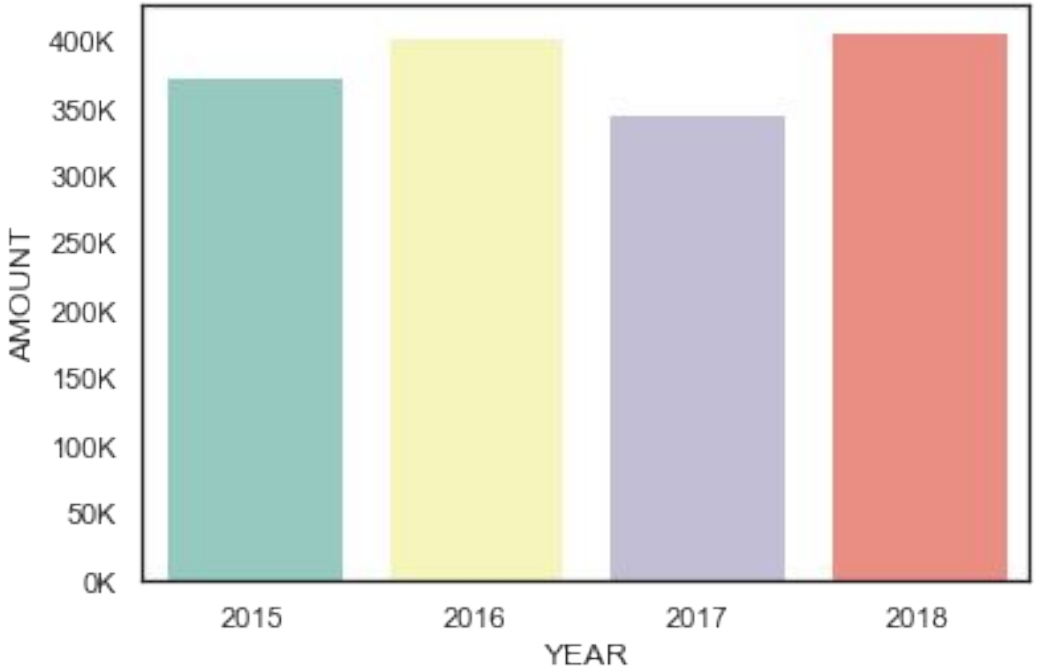


City spending data by org over that timespan



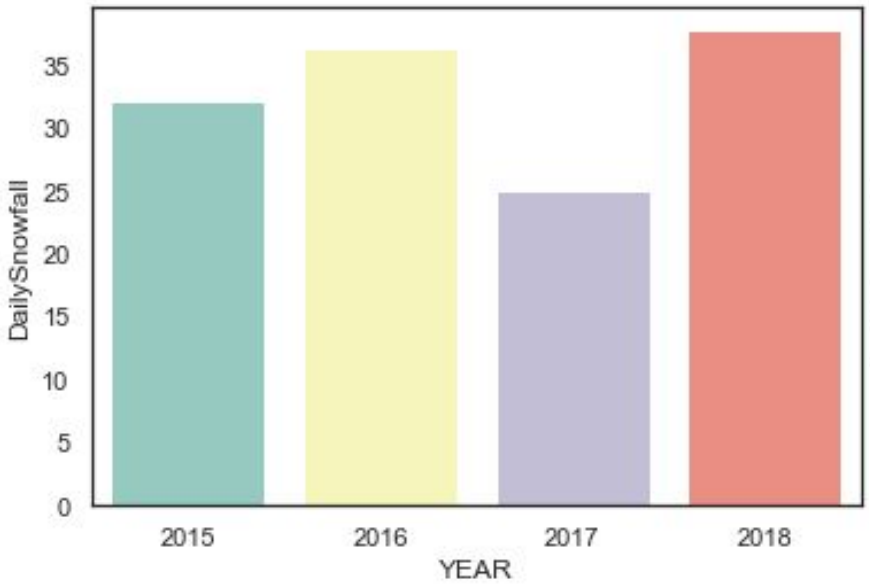
Snow Plowing organization

Snow plowing overtime pay over the years

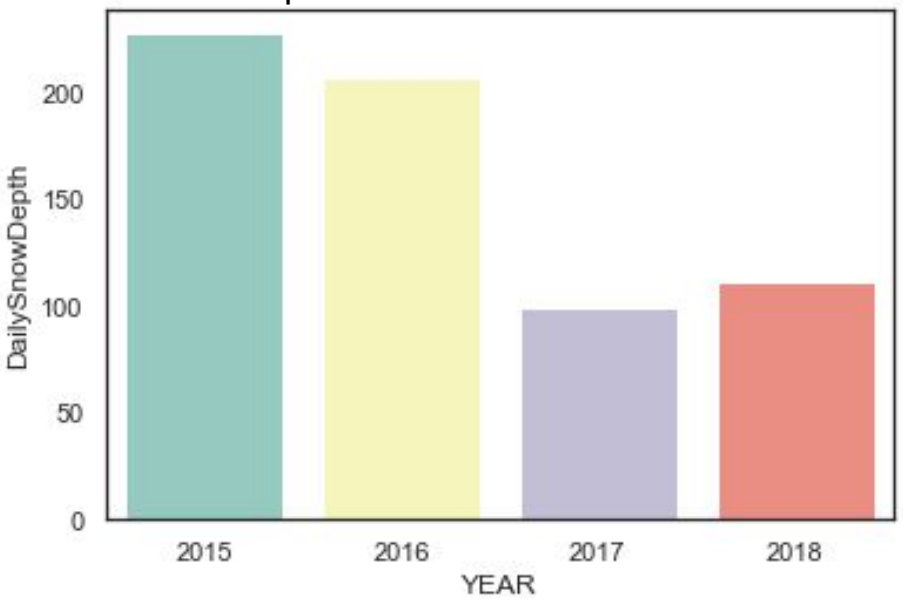


Cumulative daily snowfall and snow depth

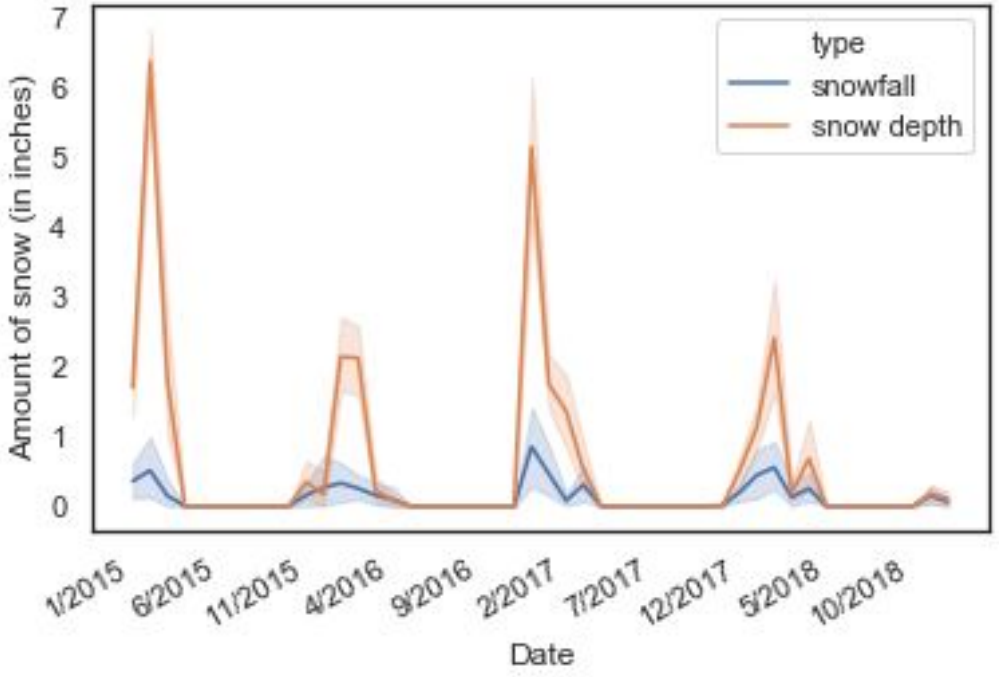
Snowfall



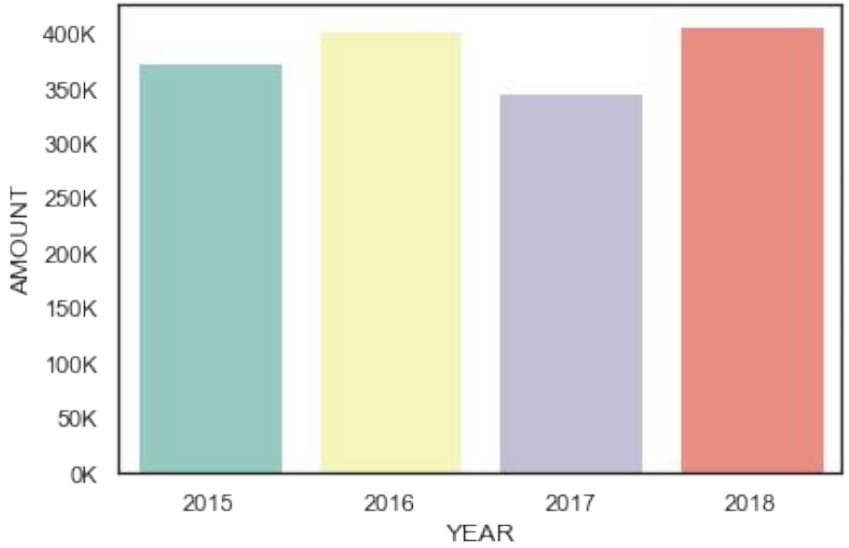
Snow depth



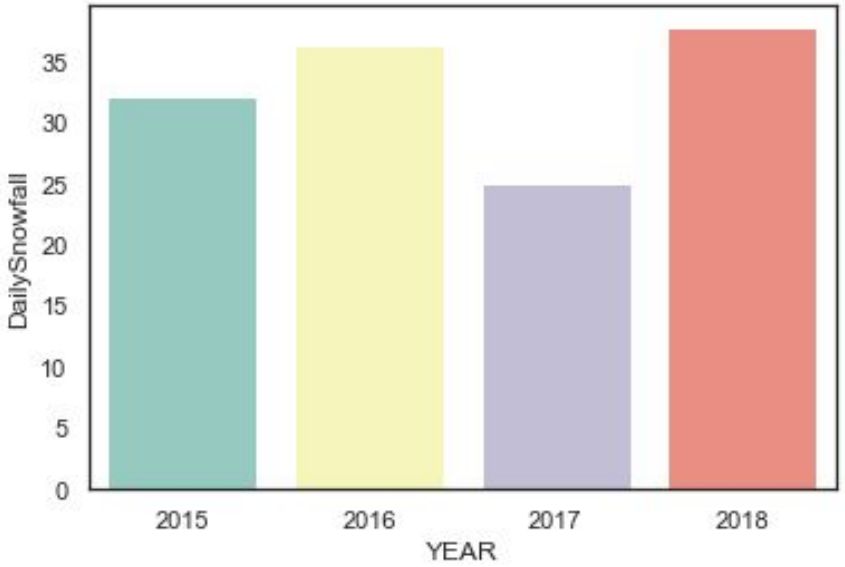
Comparison of snowfall and snow depth



Similarities to daily snowfall over the years

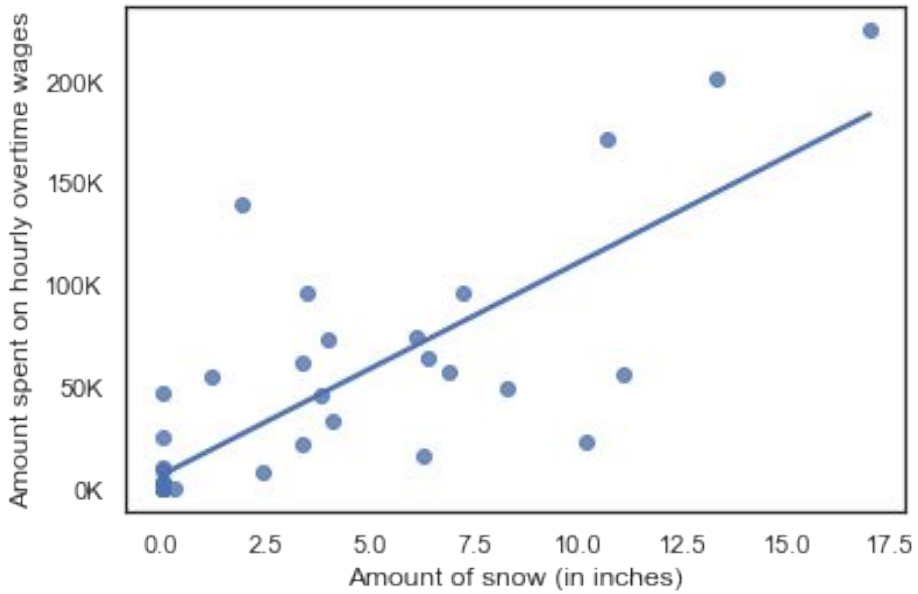


Yearly spending on overtime wages



Yearly cumulative snowfall

Regression model between monthly spending and snowfall



Simple linear regression model to predict overtime hourly wages from predicted snowfall

Sum of X = 131.5 Mean X = 2.7396 (where **X = monthly snowfall**)

Sum of Y = 1702749.35 Mean Y = 35473.9448 (where **Y = monthly spending on overtime wages**)

Sum of squares (SSX) = 827.0948

Sum of products (SP) = 8607676.0319

Regression Equation = $\hat{y} = bX + a$

b = SP/SSX = 8607676.03/827.09 = 10407.12155

a = MY - bMX = 35473.94 - (10407.12*2.74) = 6962.76804

$\hat{y} = 10407.12155X + 6962.76804$

Correlation coefficient of the model

$$r = \frac{1}{n-1}\left(\frac{\sum_x \sum_y (x-\bar{x})(y-\bar{y})}{s_x s_y}\right)$$

ws:

X = monthly cumulative snowfall

Y = monthly cumulative spending on overtime wages

Using this equation our r = 0.8071

0.7 < |r| <= 1 represents a strong correlation between variables

Predicted spending on overtime wages vs actual spending

	Predicted	Actual	Error
2019	453106.07	347048.18	30%

Using snowfall and spending data from January and February 2019.

Future

- At the end of the year, check if regression model has a lower error % when predicting spending on overtime wages

- Possibly use more objects in analysis
  - Snow supplies object
  - Snow plowing object

- Include more variables in regression analysis to decrease standard deviation and find more accurate models

Questions?