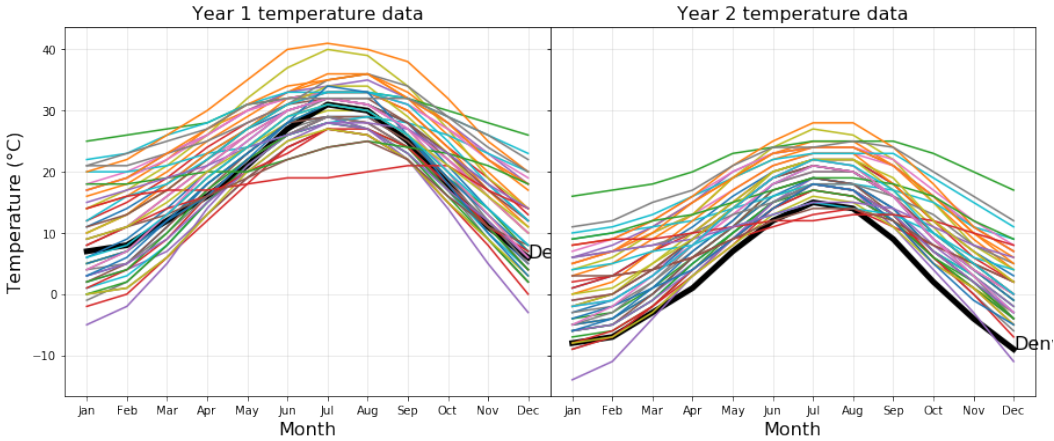
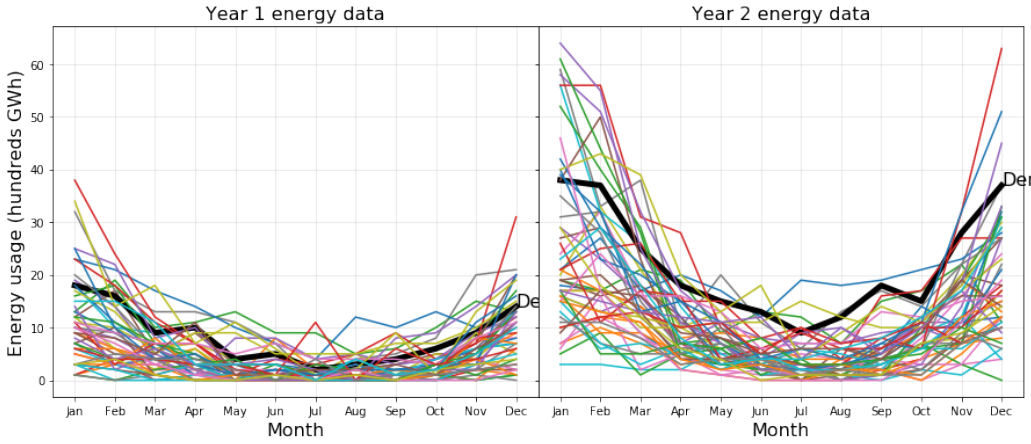
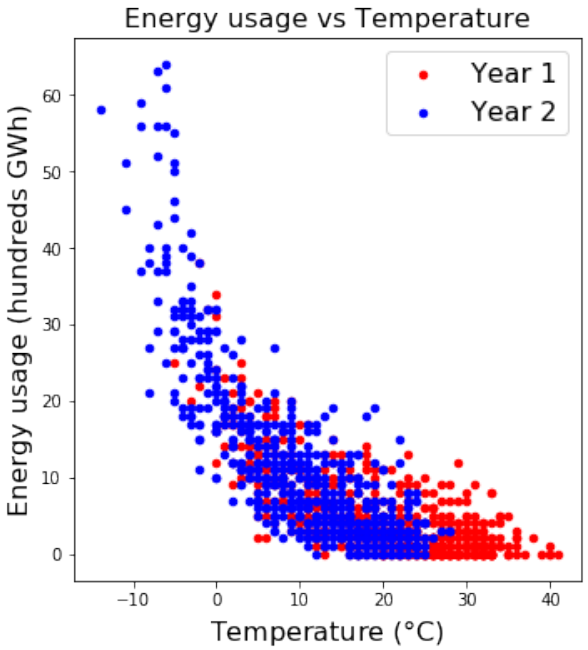
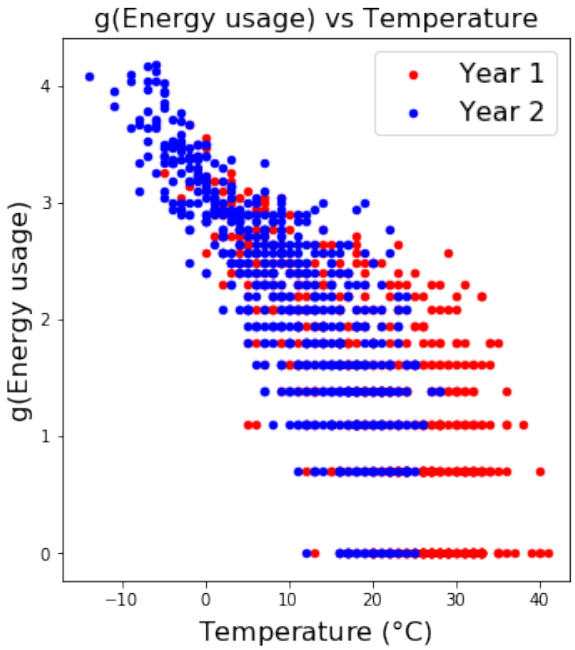
Exploratory Analysis









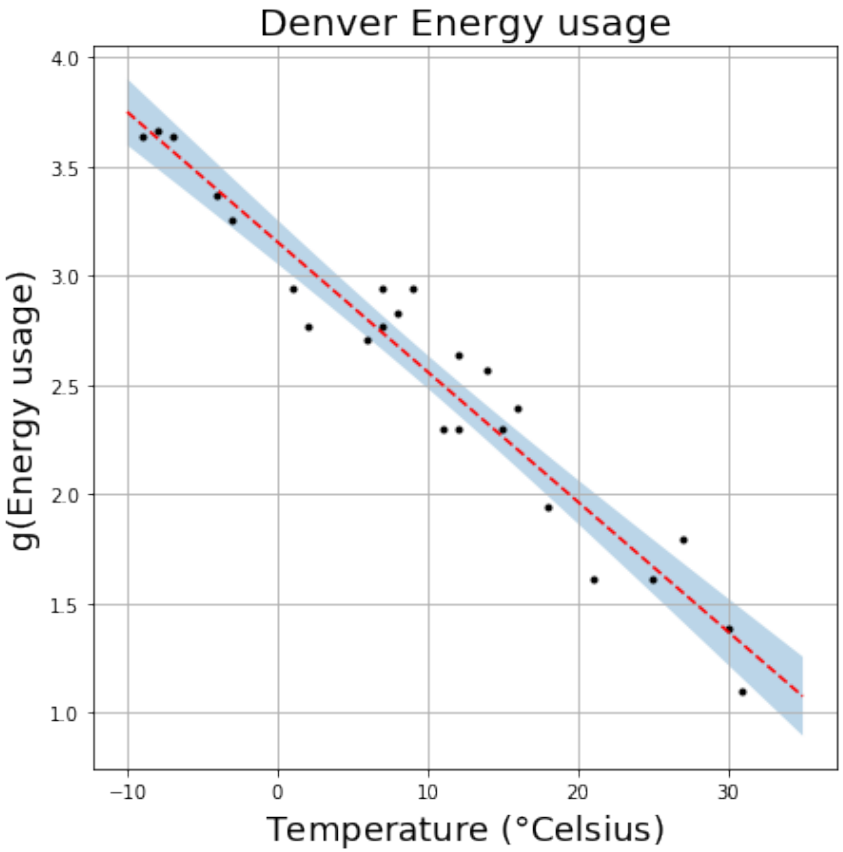
Trend of data between cities not too dissimilar.

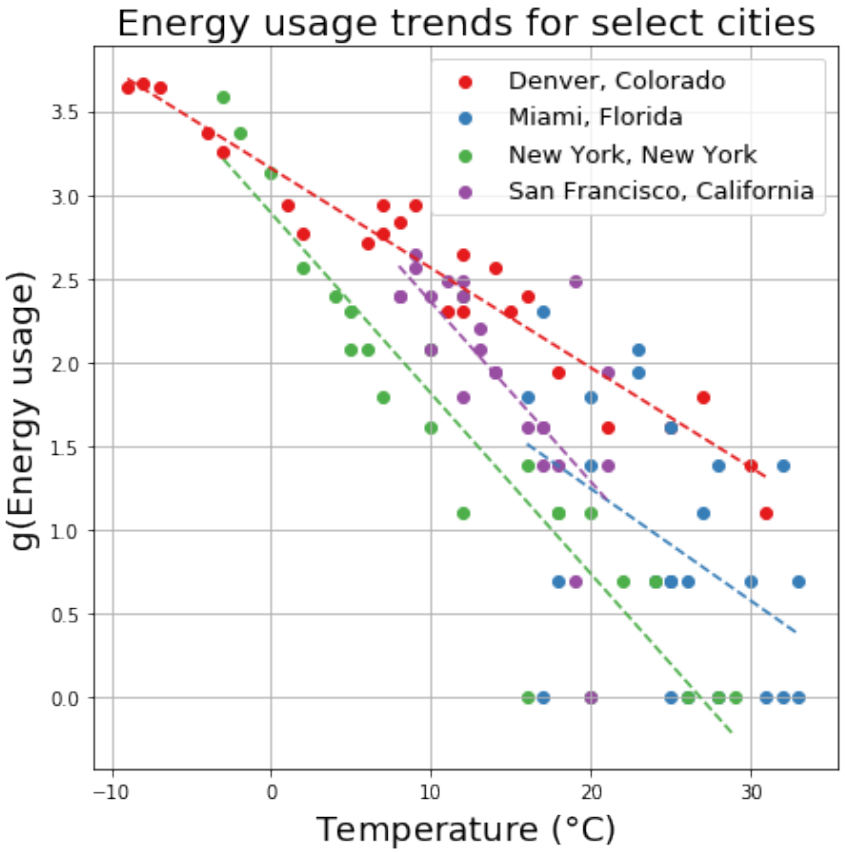
Energy usage scales nonlinearly.

Transformation used:

$$g(y) = ln(y + 1)$$

Model Selection





Considerations:

1) **Log transformation fits Denver data nicely (a case for a simple model).**
2) **Other city data provides additional information for overall structure (don't want to throw out data).**
3) **Other city data different enough to warrant hierarchical model with city-dependent random effects.**

Prediction

```
Linear mixed model fit by REML ['lmerMod']
Formula: LogEnergy ~ Temperature + (1 | City) + (0 + Temperature | City)
   Data: allData

REML criterion at convergence: 1581.5

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.6622 -0.5015  0.1215  0.5997  3.5636

Random effects:
 Groups    Name        Variance Std.Dev.
 City      (Intercept) 0.052672 0.22950
 City.1    Temperature 0.000403 0.02007
 Residual              0.177934 0.42182
Number of obs: 1224, groups:  City, 51

Fixed effects:
            Estimate Std. Error t value
(Intercept)  2.939079   0.039991   73.49
Temperature -0.077786   0.003099  -25.10

Correlation of Fixed Effects:
            (Intr)
Temperature -0.211
```

```
Models:
m0: LogEnergy ~ Temperature
m1: LogEnergy ~ Temperature + (1 | City)
m2: LogEnergy ~ Temperature + (1 | City) + (0 + Temperature | City)
   Df    AIC    BIC   logLik deviance  Chisq Chi Df
m0  3 2214.5 2229.8 -1104.25   2208.5
m1  4 1751.8 1772.3  -871.92   1743.8 464.65      1
m2  5 1577.1 1602.7  -783.55   1567.1 176.75      1
   Pr(>Chisq)
m0
m1  < 2.2e-16 ***
m2  < 2.2e-16 ***
```

Random effects: significant

Residuals vs. Fitted values