What numerical or categorical features were in the data?

Numerical features included the # of resorts per state, skiable area, days open, terrain parks, nightskiing, resorts per 100k population and resorts per 100k miles. Categorical features included state, quartile and resort features (e.g. fast Quads, snow making equipment).

Was there any pattern suggested of a relationship between state and ticket price? What did this lead us to decide regarding which features to use in subsequent modeling?

In the first two components, there is a spread of states across the first component. It looks like Vermont and New Hampshire might be off on their own a little in the second dimension, although they're really no more extreme than New York and Colorado are in the first dimension. It looks like resorts\_per\_100kcapita and resorts\_per\_100ksq\_mile might count for quite a lot, in a positive sense.

What aspects of the data (e.g. relationships between features) should you remain wary of when you come to perform feature selection for modeling?

You need to be wary of missing values and these should be accounted for in an appropriate way (e.g. removed, estimated, etc.). Also, it’s important to determine if there is an obvious correlation between features.

Two key points that must be addressed are the choice of target feature for your modelling and how, if at all, you're going to handle the states labels in the data.

By exploring the state summary data, exploration of the resort-level data can be performed in more detail. Which can help guide how (or whether) to use the state labels in the data. For example, engineering a resort's share of the supply for a given state.

Having merged the state summary features into the ski resort data, add "state resort competition" features:

* ratio of resort skiable area to total state skiable area
* ratio of resort days open to total state days open
* ratio of resort terrain park count to total state terrain park count
* ratio of resort night skiing area to total state night skiing area

Once these features have been derived within the context of its state, the state columns can be dropped. Their main purpose was to understand what share of states' skiing "assets" is accounted for by each resort.