

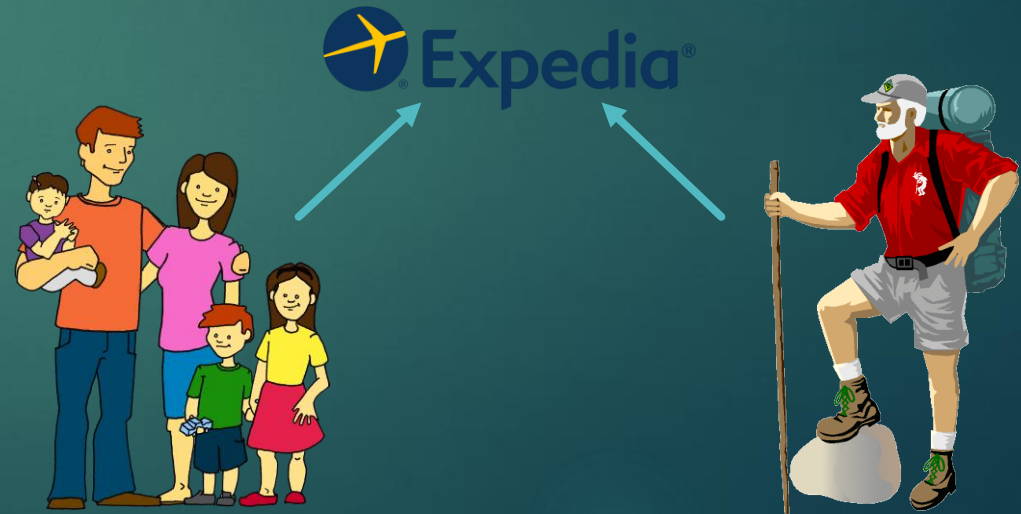
# Hierarchical Clustering and Random Forests to Identify Hotel Rank

KYLE HENKE

MANISH BHATTARAL

KELLIN RUMSEY

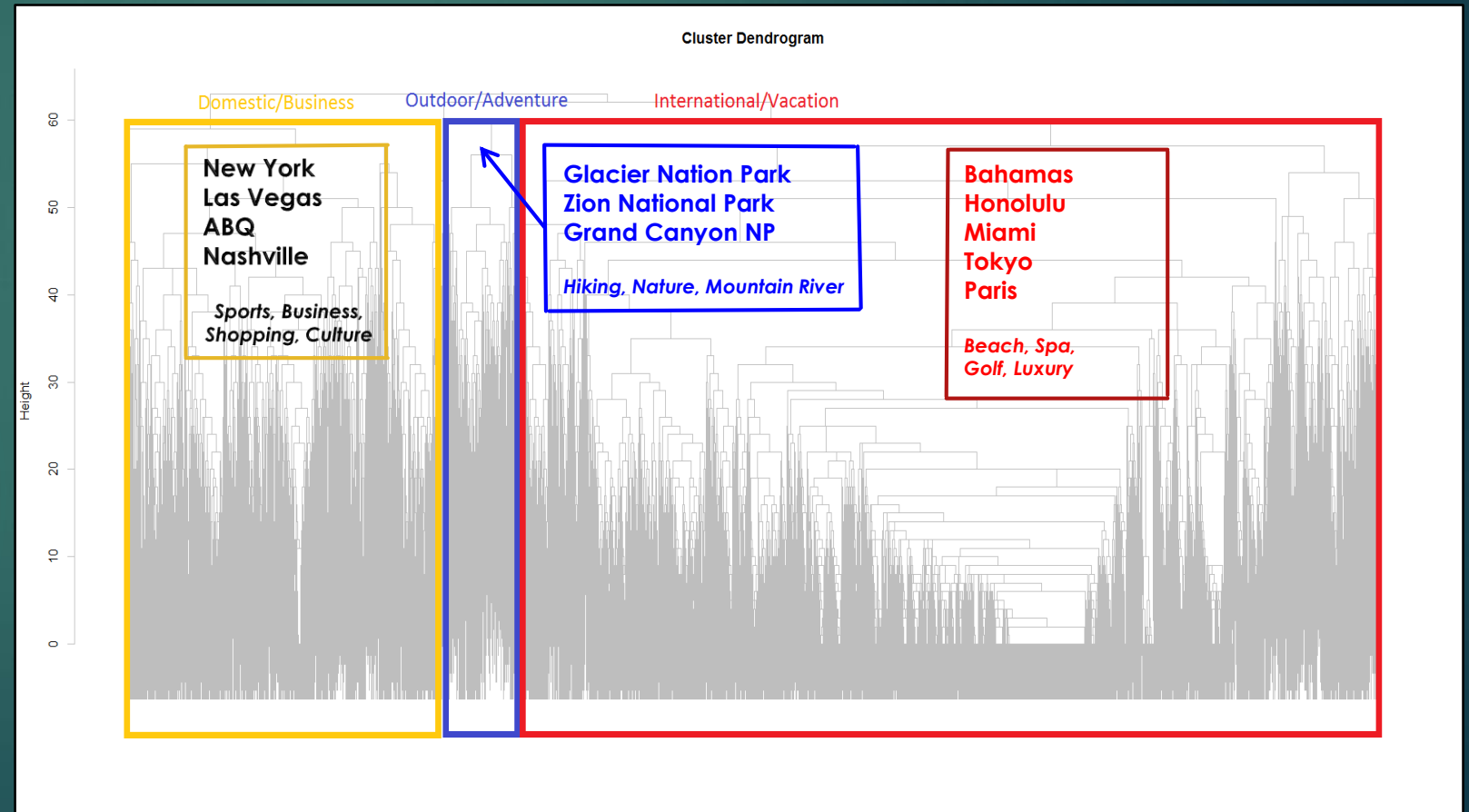
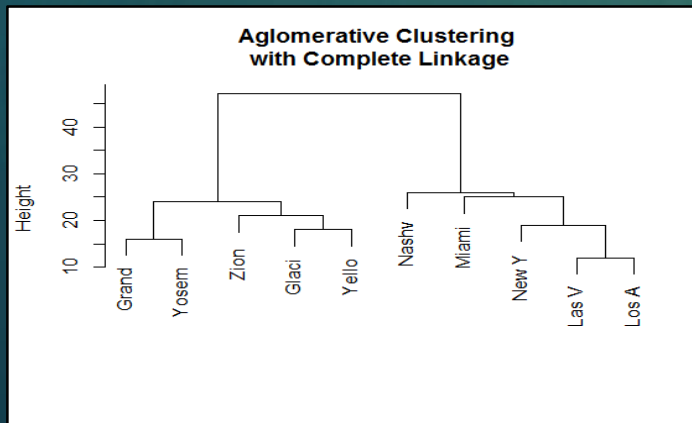
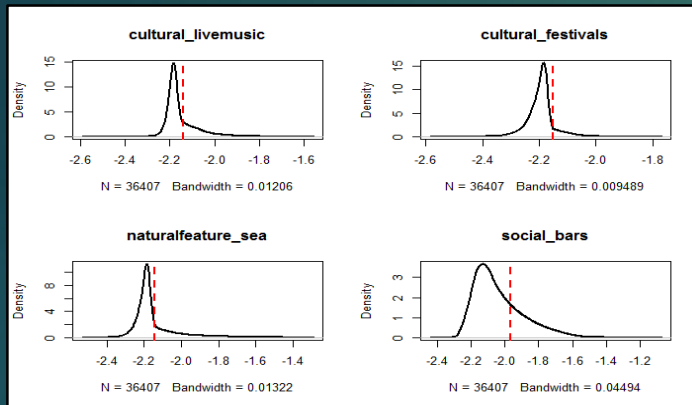
ZACH STUART



# Hierarchical Clustering – Finding Similar Destinations

srch_destination_id	srch_name	srch_type	srch_lat	srch_lon	popular_*
8524	"Albuquerque, NM"	557	35.08684	-106.6726	

Probability of a recommendation on log-scale



# Random Forests – Finding Attractive Hotel Attributes

srch_destination_id	is_booking	hotel_id	prop_is_branded	prop_starrating	distance_band	hist_price_band
8524	1	557	35.08684	-106.6726	VC	L

- ▶ TRAINING THE RANDOM FOREST
  - ▶ Take a random subset
  - ▶ Take a random subset from similar cluster
- ▶ RANKING HOTELS WITHIN A REGION
  - ▶ For each hotel in a destination, use RF to estimate booking probability.
  - ▶ Compare results to actual data.
- ▶ IMPROVING THE METHOD
  - ▶ Filter clustering by search type
  - ▶ Attribute selection in clustering.
  - ▶ SVM

## Zion National Park Predicted Booking Probability (37 hotels)

Hotel Ranking	Random Subset	Cluster Subset	Actual Bookings
1 <sup>st</sup>	11.35%	26.85%	<b>14.56%</b>
2 <sup>nd</sup>	11.35%	<b>21.19%</b>	11.26%
3 <sup>rd</sup>	11.35%	10.83%	9.89%
4 <sup>th</sup>	11.35%	10.36%	7.14%
5 <sup>th</sup>	11.35%	5.49%	6.86%