

# Yelp Data Analysis

Yanqi Huang  
Huiyu Jiang  
Xinjie Ye

University of Wisconsin-Madison

March 7, 2018

# Data Description

Keys	Values
star	1
name	McDonald's
text	Seriously cannot stand this McDonald's. They NEVER get my order right. Food almost always sucks! ...
date	2014-12-29
city	Glendale
longitude	-112.205020
latitude	33.509597
categories	['Burgers', 'Fast Food', 'Restaurants']

# Text Process

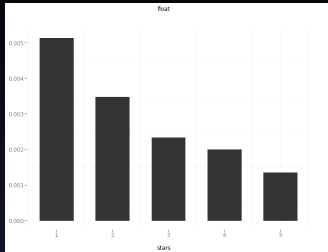
## **Aim1:** Postive/Negative words

- Stratified sample 100000 reviews
- Clean data: *n't* → *not*, *delete* :?!"(), *upper* → *lower*
- Count frequencies of all words
- Select useful words

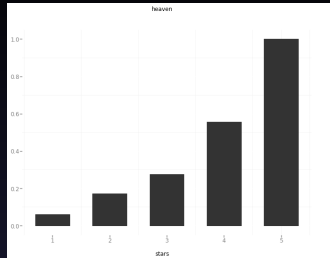
## **Aim2:** Special Expressions :) :( etc

- Delete all English letters: a-zA-Z
- Count frequencies of special expressions

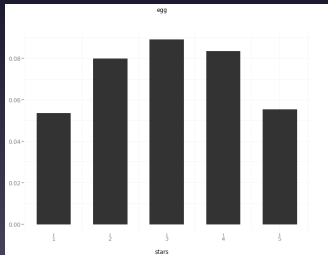
# Word Selection



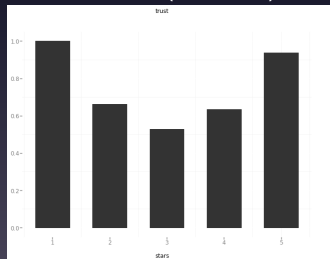
Negative(*float*)



Positive(*heaven*)

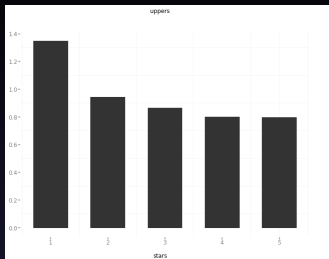


Neutral(*egg*)

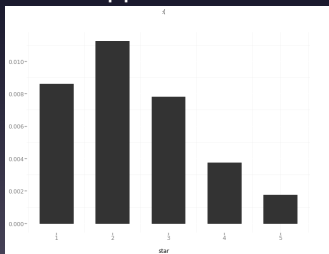


Intensive(*trust*)

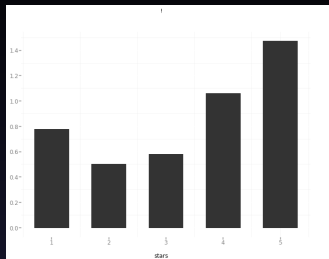
# Symbol Selection



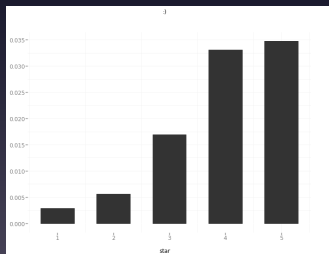
Upper words



:(



!



:)

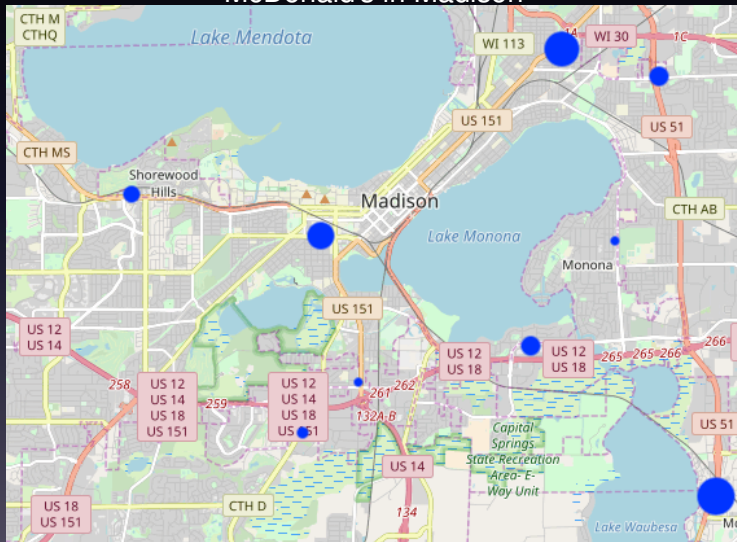
# Data Description

Keys	Values
star	1
name	McDonald's
text	Seriously cannot stand this McDonald's. They NEVER get my order right. Food almost always sucks! ...
date	2014-12-29
city	Glendale
longitude	-112.205020
latitude	33.509597
categories	['Burgers', 'Fast Food', 'Restaurants']

# Data Visualization

shiny app link: <https://carolxxxshiny.shinyapps.io/Module2>

## McDonald's in Madison



# Next Step

- Sentiment Analysis of new texts:  
*Based on frequencies of useful words and symbols*
- Consider seasonal difference for restaurants
- Apply linear model and Machine Learning methods to train models
- Adjust model parameters to make better predictions

To Be Continued...