Analysis for steak restaurants based on Yelp review data

STAT 628 Module3

Huitong Kou, Zihang Wang, Peibin Rui

UNIVERSITY OF WISCONSIN–MADISON

Steak Restaurants on Yelp

Data Cleaning

- **Reviews for restaurants in US:** Madison, Cleveland, Pittsburgh, Urbana-Champaign, etc

- Mainly focus on **steak restaurants**.

- **Natural Language Processing (NLP)** using **NTLK** package in Python.

- Processing example:

Raw review: The food was very good. The waitress was very inventive. The only thing is my I felt so sick after eating it. I'm not sure if it's just my stomach not being able to handle it or what but I felt terrible. The staff at the front was very friendly and talkative.
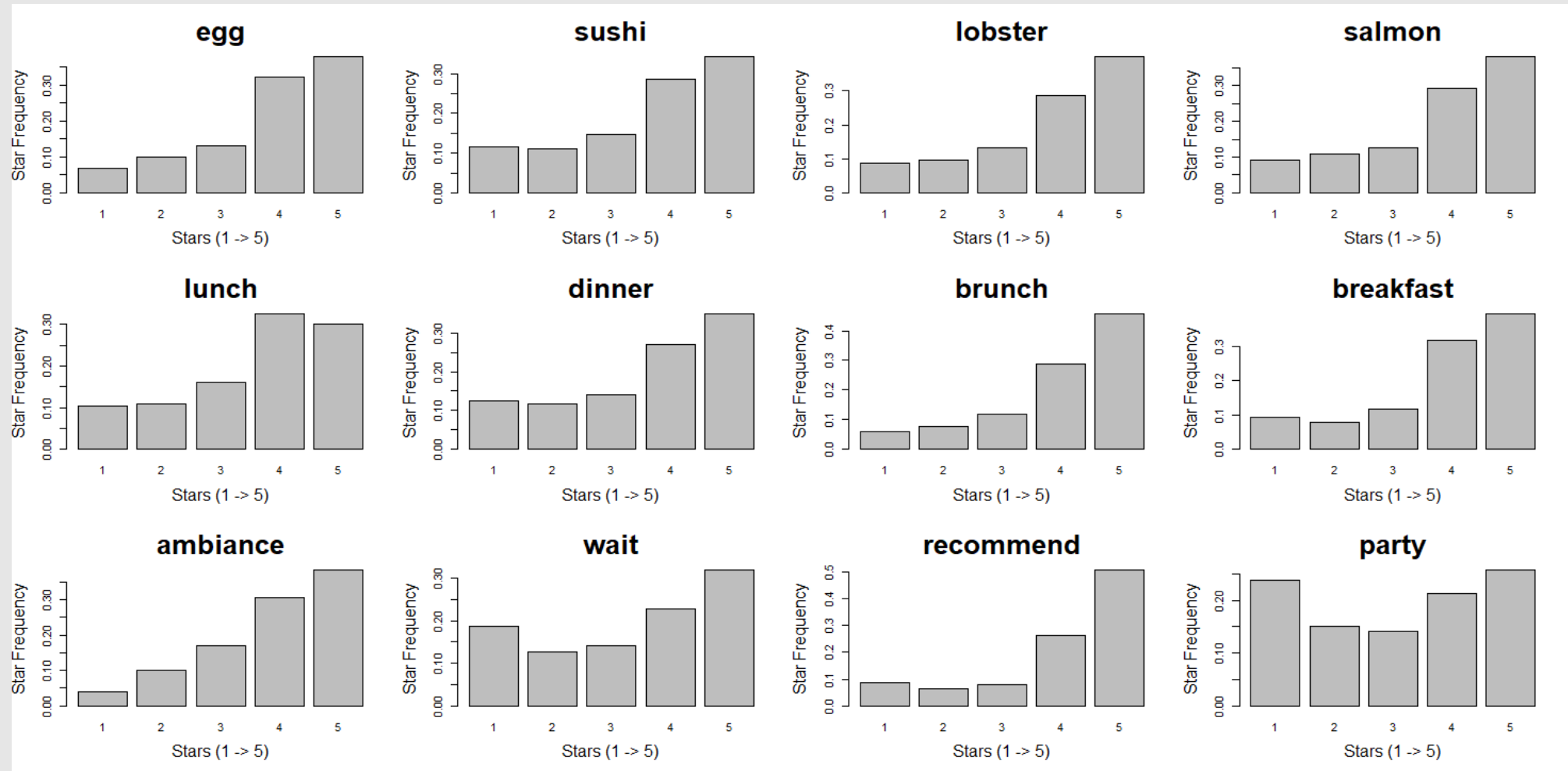
Proccessed review: 'food', 'good', 'waitress', 'inventive', 'thing', 'felt', 'sick', 'eat', 'sure', 'stomach', 'able', 'handle', 'felt', 'terrible', 'staff', 'front', 'friendly', 'talkative'.
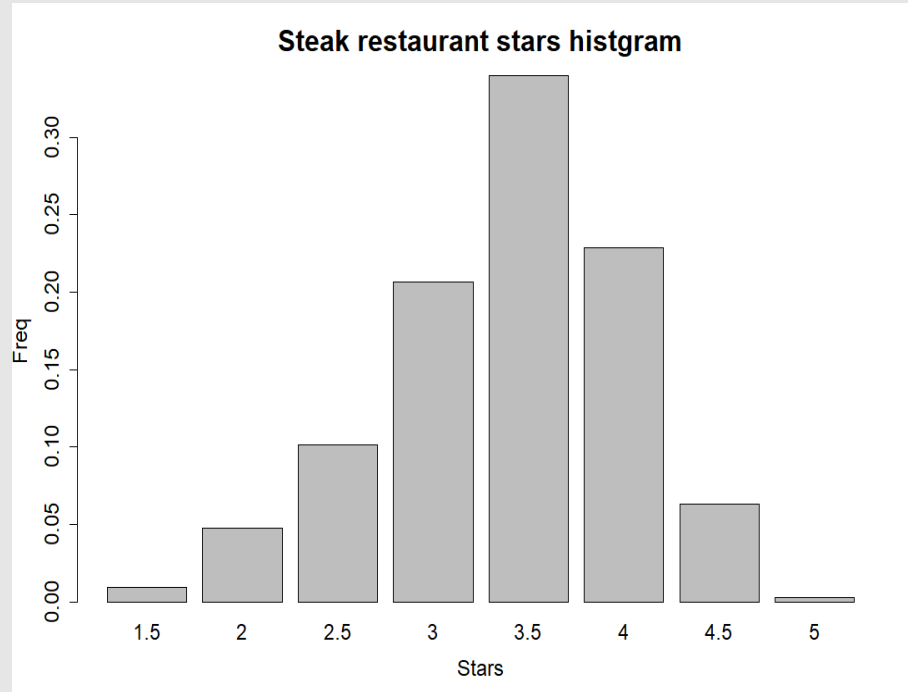
Steak Types
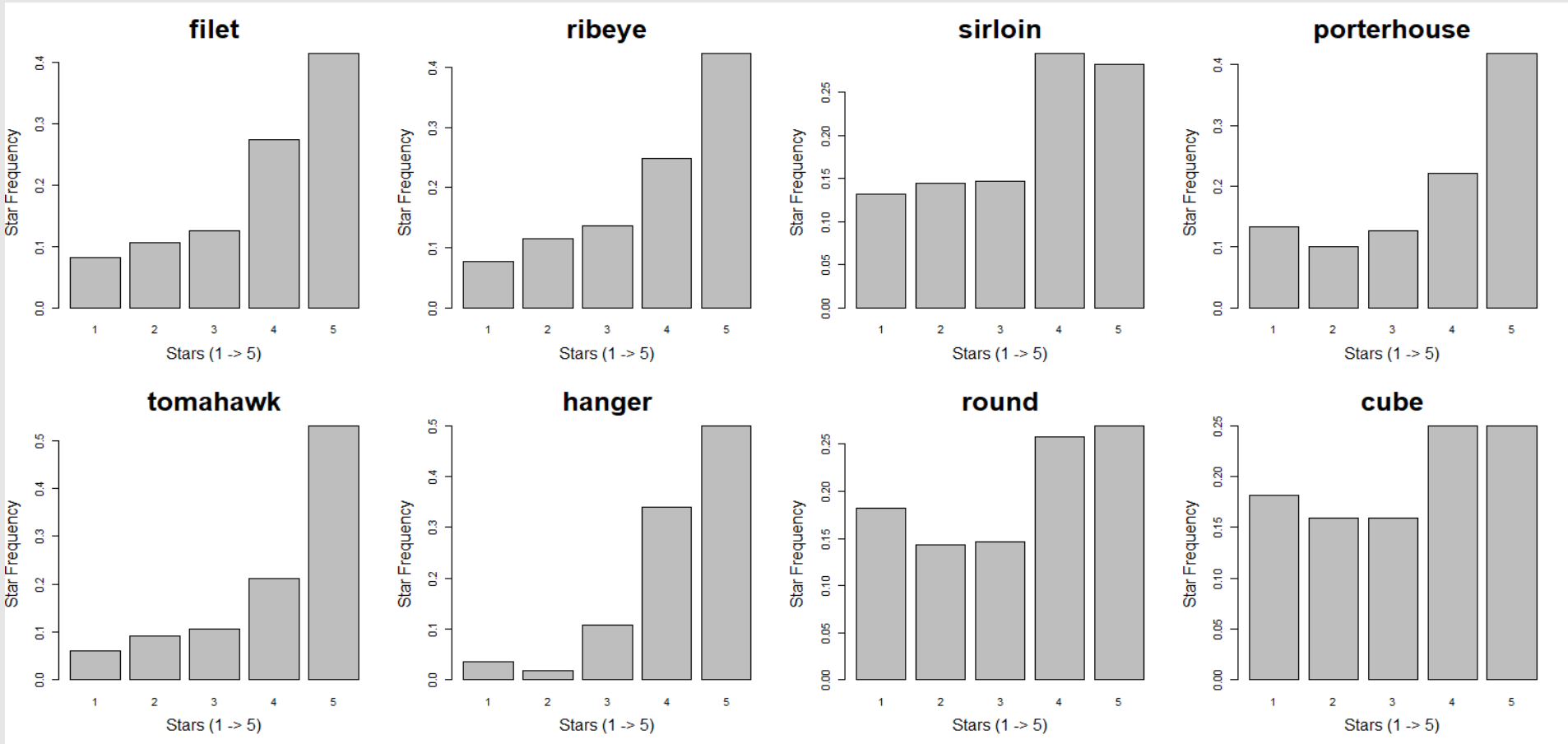
Factors Other Than Steak

T-test on Business Attributes

Steak restaurant stars histgram

attribute	p.value
RestaurantsReservations	0.040536672387
NoiseLevel	0.115368255456
RestaurantsAttire	0.000007760033
WiFi	0.799610824593
OutdoorSeating	0.044275411757
RestaurantsDelivery	0.012348344280
RestaurantsGoodForGroups	0.454369986827

➢ **Significant:** Restaurants Reservations (True), Restaurants Attire (Dressy), Outdoor Seating (True) and Restaurants Delivery (False).

➢ **Non-significant:** The Noise Level (Quiet), WiFi (Free), Restaurants Good For Groups (True).

Multiple Linear Regression

➢ Regression statistics and QQ-plot

Coefficients	Estimate	Lower Bound	Upper Bound	Standard Error	P-value
Intercept	3.3798	3.3567	3.4028	0.0118	$<2 \times 10^{-16}$
filet	0.0667	0.0312	0.1023	0.0182	2.374×10^{-4}
ribeye	0.0713	0.0231	0.1195	0.0246	3.732×10^{-3}
brunch	0.0626	0.0183	0.1069	0.0226	5.602×10^{-3}
ambiance	0.1319	0.0618	0.2020	0.0358	2.379×10^{-4}
recommend	0.2894	0.2488	0.3300	0.0207	3.313×10^{-44}
party	-0.0685	-0.1076	-0.0295	0.0199	5.850×10^{-4}
$R^2 = 0.3365$	$Adjusted\ R^2 = 0.3347$		$p-value: < 2.2 \times 10^{-16}$		
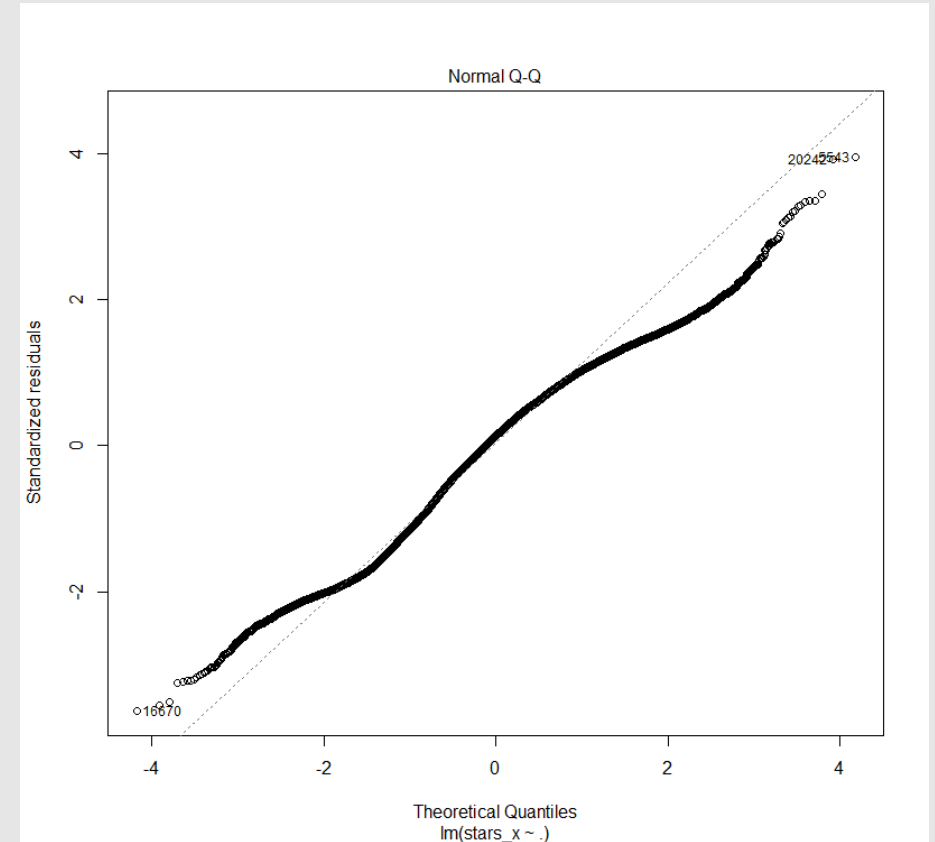
➢ **Steak Types:**
- Make improvement: filet, ribeye, tomahawk, hanger and porterhouse steaks
- Be careful: sirloin, round and cube steaks

➢ **Other foods:**
- Special: lobster and salmon
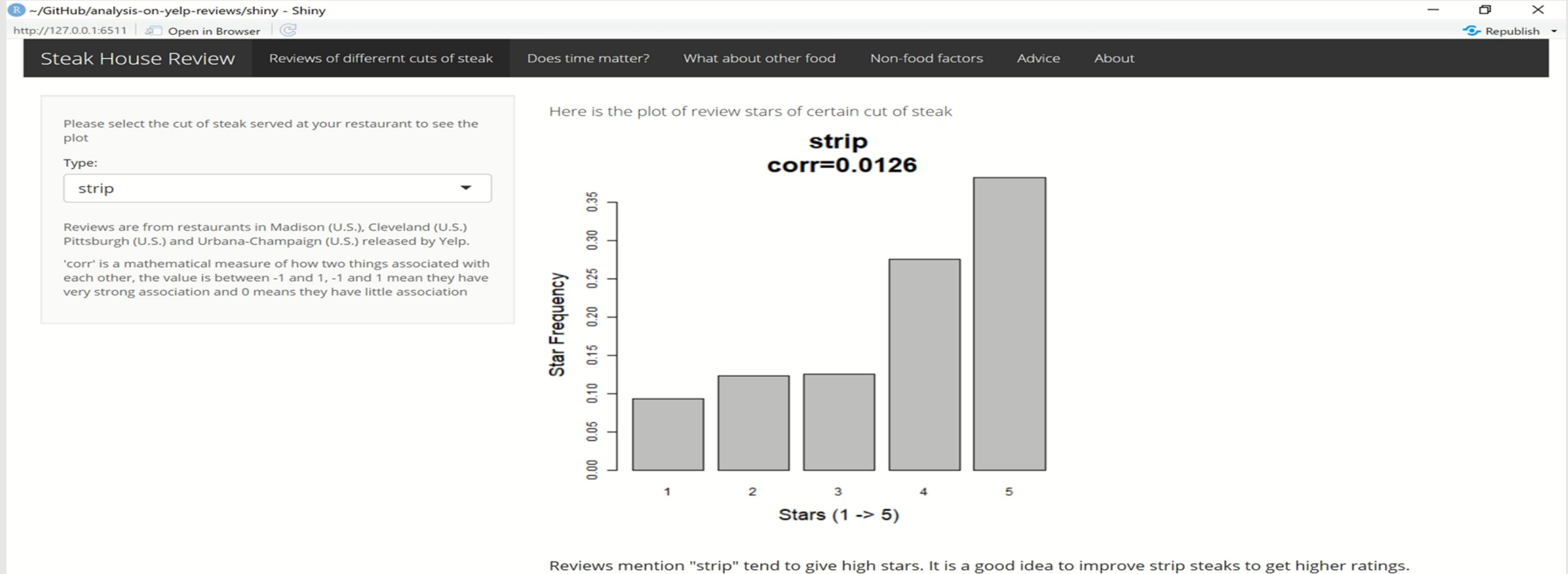- Side orders: desserts, salad, cheese, sushi
- Meal time: brunch

➢ **Environmental factors:**
- Comfortable ambiance
- No noisy activities such as parties
- Encourage recommendations

Shiny app

Conclusion and discussion

> **Summary:**
- Cleaned review data from Yelp
- Exploratory data analysis
- Multiple linear regression model and tests
- Recommendations and plans

> **Limitations:**
- Data may not be completely cleaned up
- Lack of marketing knowledge

THANK YOU