

# Italian restaurants in NYC

MULTIPLE AND LOGISTIC REGRESSION IN R



Ben Baumer  
Instructor



# ZAGAT 2013

## La Masseria

FOOD	DECOR	SERVICE	COST
<u>23</u>	<u>21</u>	<u>22</u>	<u>\$57</u>

Showgoers tout this “quick-pace” Hell’s Kitchen Southern Italian for its “hearty” cooking, “sweet farmhouse” setting and “fast” service that “gets you to the theater on time”; “crowded” conditions and “pricey-but-worth-the-money” tabs complete the overall appealing picture.

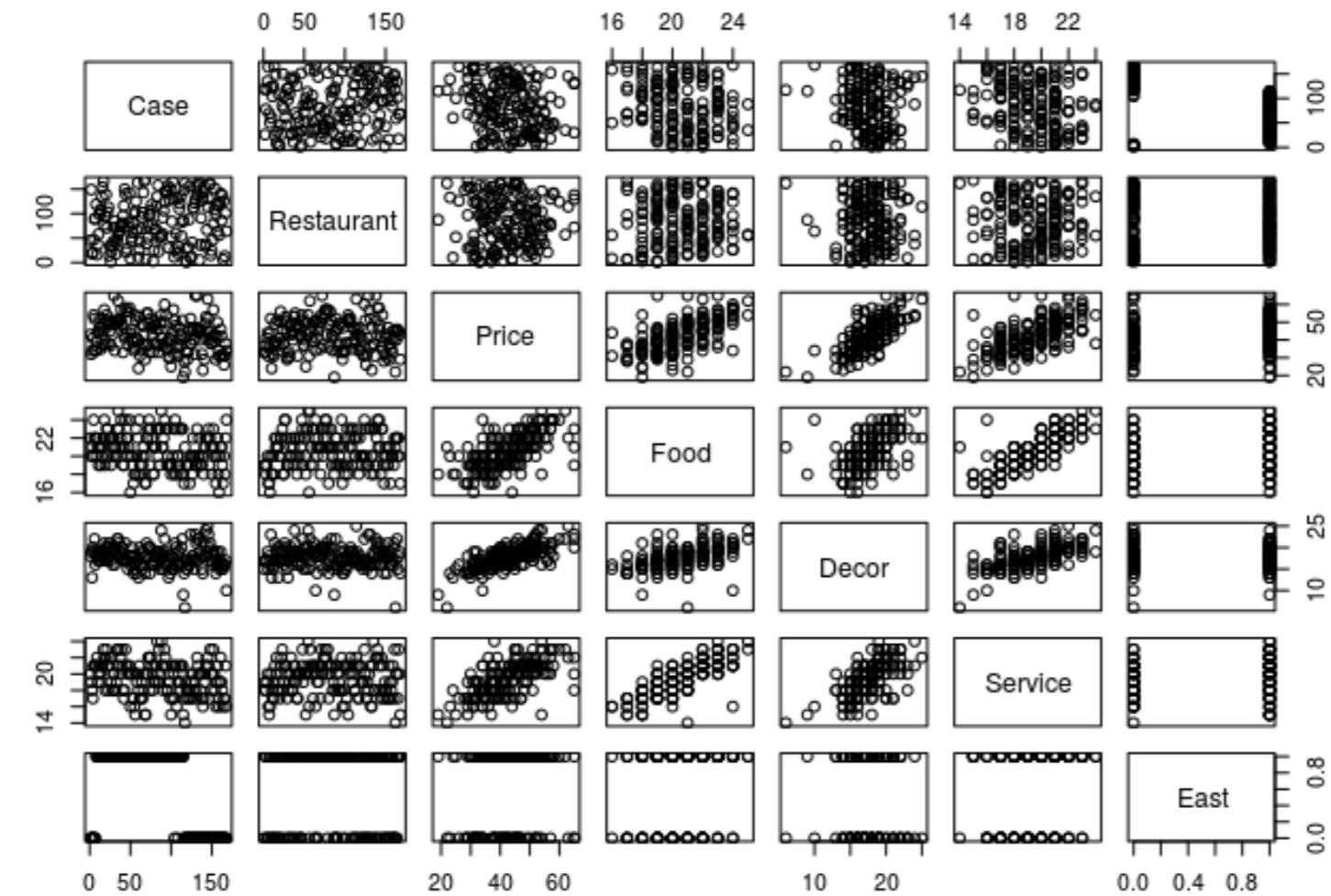
# Exploring the data

```
glimpse(nyc)
```

```
## Observations: 168
## Variables: 7
## $ Case      <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, ...
## $ Restaurant <fctr> Daniella Ristorante, Tello's Ristorante, Biricchin...
## $ Price      <int> 43, 32, 34, 41, 54, 52, 34, 34, 39, 44, 45, 47, 52, ...
## $ Food       <int> 22, 20, 21, 20, 24, 22, 22, 20, 22, 21, 19, 21, 21, ...
## $ Decor      <int> 18, 19, 13, 20, 19, 22, 16, 18, 19, 17, 17, 19, 19, ...
## $ Service    <int> 20, 19, 18, 17, 21, 21, 21, 22, 19, 20, 21, 20, ...
## $ East       <int> 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, ...
```

# EDA

```
pairs(nyc)
```



# Let's practice!

MULTIPLE AND LOGISTIC REGRESSION IN R

# Incorporating another variable

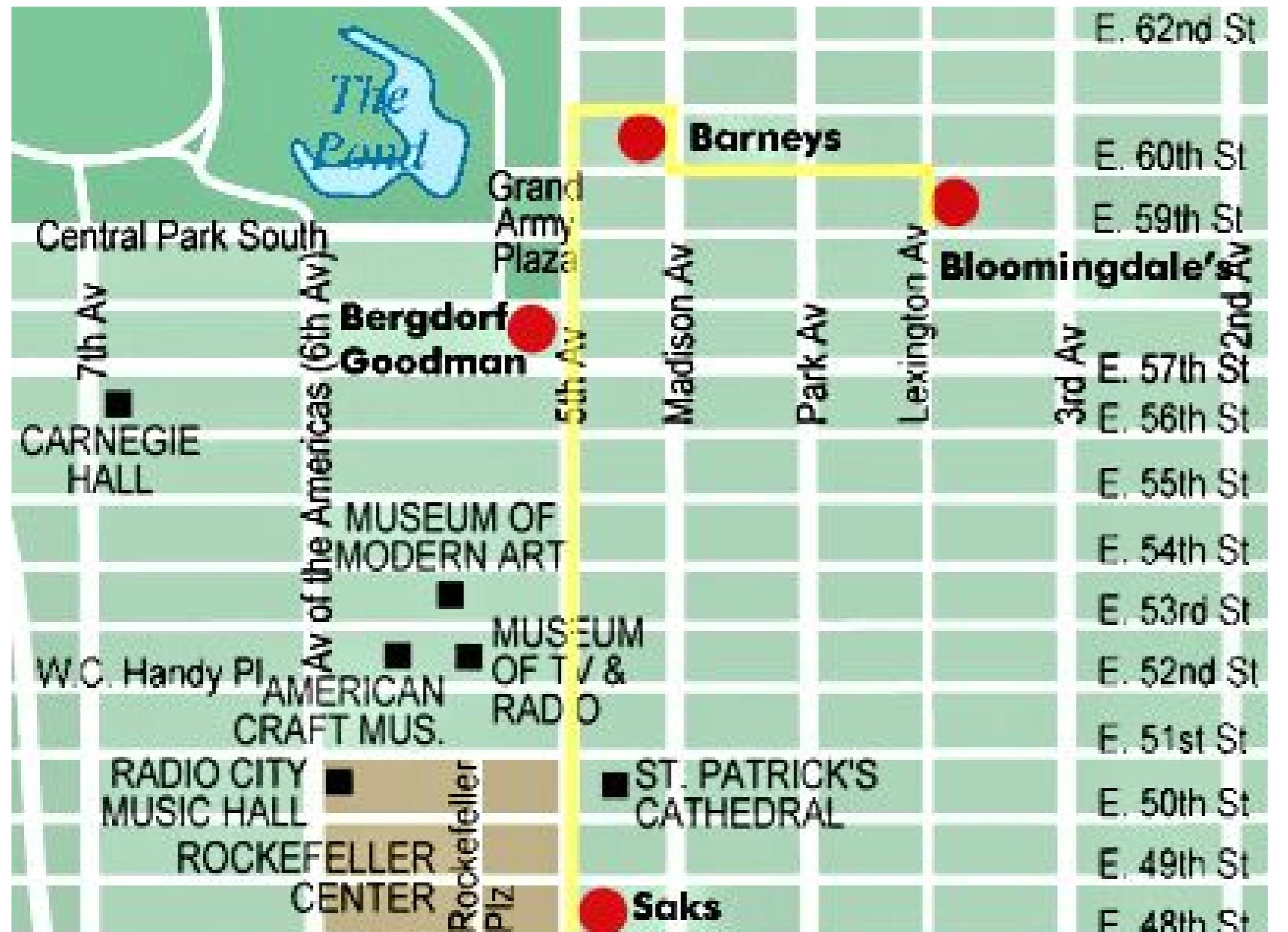
MULTIPLE AND LOGISTIC REGRESSION IN R



Ben Baumer  
Instructor







# The price of location

```
nyc %>%  
  group_by(East) %>%  
  summarize(mean_price = mean(Price))
```

```
# A tibble: 2 x 2  
  East mean_price  
  <int>     <dbl>  
1     0     40.43548  
2     1     44.01887
```



# Let's practice!

MULTIPLE AND LOGISTIC REGRESSION IN R

# Higher dimensions

MULTIPLE AND LOGISTIC REGRESSION IN R



Ben Baumer  
Instructor



# Building a full model

- Response variable:
  - Price
- Explanatory variables
  - Food
  - Service
  - Decor
  - East (categorical)
- Unusable
  - Case
  - Restaurant

# Collinearity

```
nyc %>%  
  mutate(Price_cents = Price / 100) %>%  
  summarize(cor_collinear = cor(Price, Price_cents))
```

```
##     cor_collinear  
## 1
```

# Multicollinearity

- Explanatory variables are highly correlated
- Unstable coefficient estimates
- Doesn't affect  $R^2$
- Be skeptical of surprising results

# Let's practice!

MULTIPLE AND LOGISTIC REGRESSION IN R

# Wrap-up

MULTIPLE AND LOGISTIC REGRESSION IN R



Ben Baumer  
Instructor