Tipping Behavior Analysis

Byunghyun Hwang

Data Summary

This dataset contains 244 restaurant transactions from a U.S. restaurant, where each record includes the total bill, tip amount, customer demographic information, and dining information (Badole 2024).

Variables

- **total_bill**: total cost of the meal ($)
- **tip**: tip amount ($)
- **sex**: gender of the customer (Male of Female)
- **smoker**: smoker or non-smoker (Yes or No)
- **day**: day of visit (Thursday to Sunday)
- **time**: time to visit the restaurant (Lunch or Dinner)
- **size**: number of people at table

Research Purpose

In the United States, tipping in restaurants has been a cultural practice to appreciate good service. Nowadays, however, some consumers have expressed frustration arguing that tipping has become pressured by staff. This can be especially confusing for travelers who are unfamiliar with the tip culture since they do not know how much they are supposed to tip.

For restaurant employees, tipping remains a significant part of their income. This results in a natural tension between customers and service staff members.

Therefore, this study explores two key research questions:

what is a reasonable tip amount, and what types of customers tend to tip more?

For this analysis, three R packages are used: ggplot2, dplyr, and scales (Wickham 2023; Wickham et al. 2023; Wickham, Seidel, and Dunnington 2023).

Research Questions

Main Questions

- What is a reasonable tip amount?
- What types of customers tend to tip more?

Sub-Questions

- Do larger groups leave higher tips?
- Do dinner customers tip more than lunch customers?
- Do tipping rates differ by day?
- Do tipping rates differ by day and meal time?
- Does spending on more on a meal lead to higher tip rates?

EDA1

1. Bill Distribution

```
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
3.07   13.35   17.80   19.79   24.13   50.81
```


Distribution of Total Bill
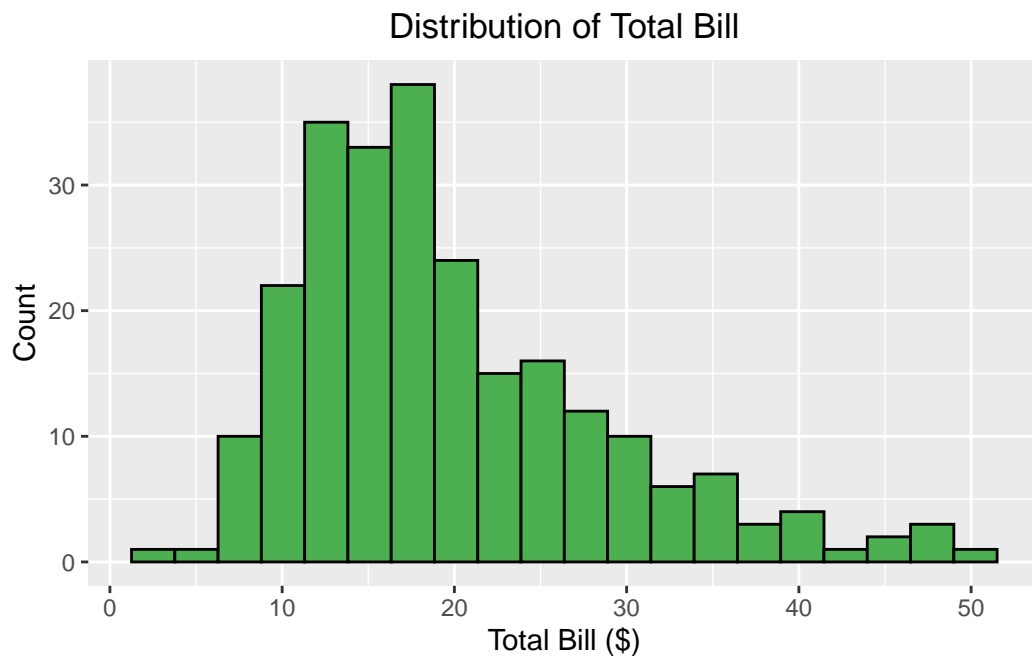
The average and median total bills are approximately $20 and $18, respectively, with values ranging from approximately $3 to $51. The mean is slightly higher than the median due to extreme values. Most customers spend between roughly $13 and $24. This indicates that this restaurant is a moderately priced restaurant.

2. Tip Amount

```
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
1.000   2.000   2.900   2.998   3.562  10.000
```

Distribution of Tips
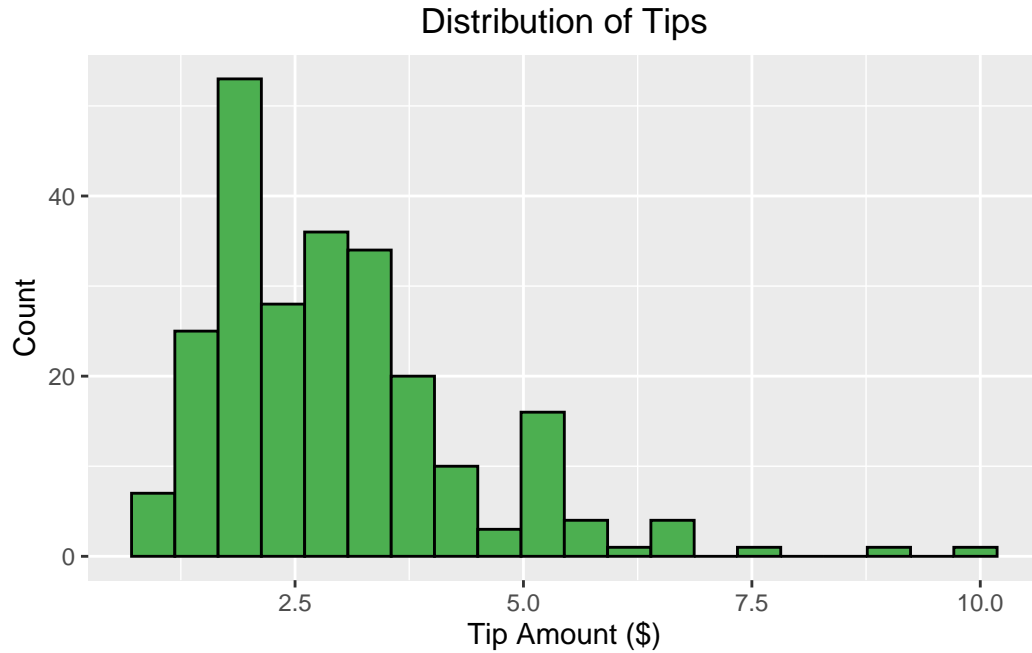
The average and median tip amounts are approximately $3, with tips ranging from $1 to $10. Most customers leave tips between roughly $2 and $3.6. This indicates that most customers leave modest tip amounts.

3. Tip Rates

```
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
3.564  12.913  15.477  16.080  19.148  71.034
```

Distribution of Tip Rates
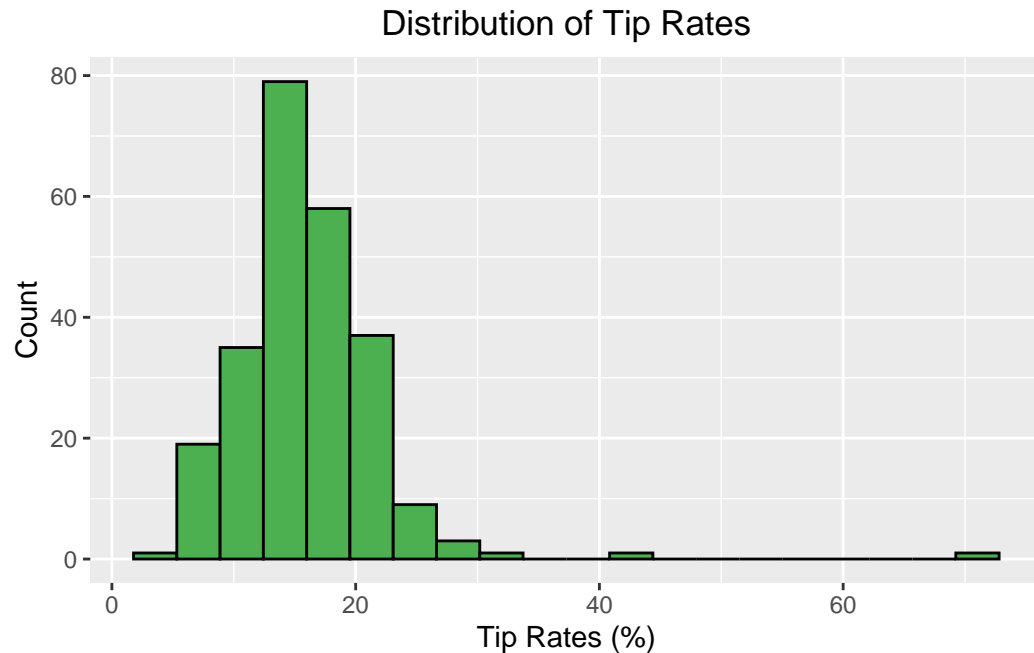
The average and median tip rates are approximately 16% and 15.5%, respectively. The mean is slightly higher than the median due to extreme values. Most customers leave tips between roughly 13% and 19%. Although tip rates range from 3.6% to 71%, these extreme cases are outliers and do not represent typical tipping behavior.

EDA2

1. Remove Outliers

While conducting the initial exploratory data analysis for the first research question, it was found that outliers influenced the statistical results. To ensure a more accurate analysis, outliers were removed based on abnormal tip rates.

2. Tip Rates by Group Size

We often assume that large groups at restaurants are associated with celebrations. It seems reasonable to expect that they might tip more due to a fest atmosphere. But, is this actually true?

```
# A tibble: 2 x 4
  group_type   count tip_rates_mean tip_rates_median
  <fct>        <int>          <dbl>            <dbl>
1 Small (1-2)    156           0.16            0.155
2 Large (3+)      84          0.149            0.15
```

Tip Rates by Group Size
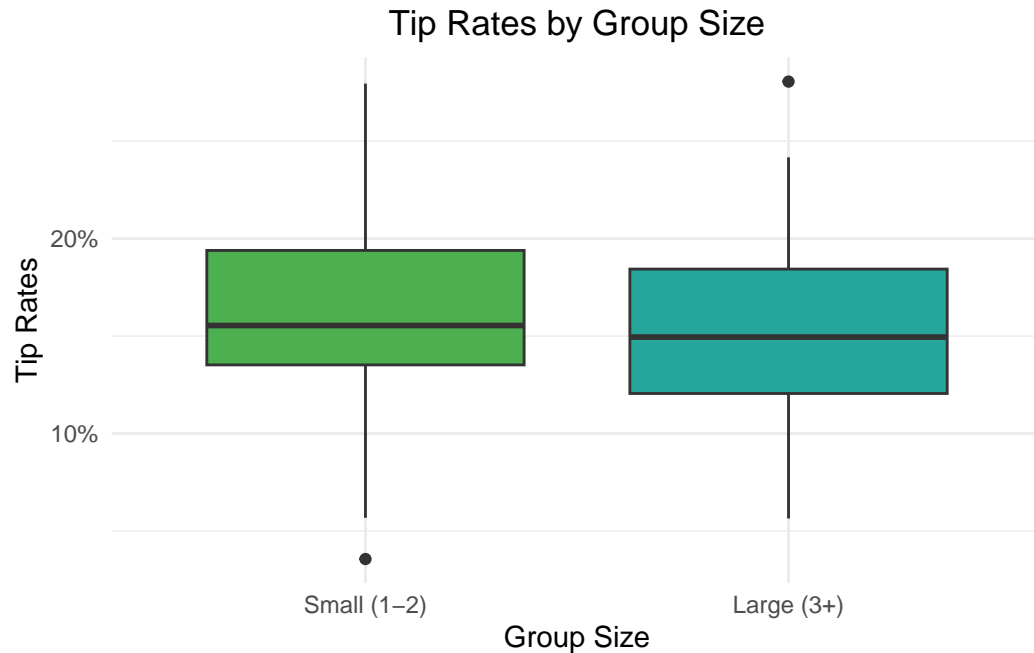

```
        Wilcoxon rank sum test with continuity correction

data:  tip_rate by group_type
W = 7440, p-value = 0.08363
alternative hypothesis: true location shift is not equal to 0
```

Since the original group sizes were unevenly distributed, so customers were re-categorized into two groups (1-2 people vs 3+ people) to ensure a more balanced comparison.

The box plot suggests that smaller groups tend to leave slightly higher tip rates than larger groups. To examine whether this difference is statistically meaningful, a Wilcoxon rank-sum test was conducted. The result indicates no statistically significant difference in tip rate (p-value > 0.05). In other words, while smaller groups appear to tip slightly more, this difference is not statistically significant in this dataset.

3. Tip Rates by Meal Time

Do you think meal time (lunch or dinner) influences tipping behavior? It is not obvious, but dinner customers might leave higher tips since dinner is typically a more relaxed and social occasion compared to lunch.

```
# A tibble: 2 x 4
  time    count tip_rate_mean tip_rate_median
  <fct>   <int>         <dbl>           <dbl>
1 Lunch      68         0.164           0.154
2 Dinner    172         0.153           0.152
```

Tip Rates by Meal Time
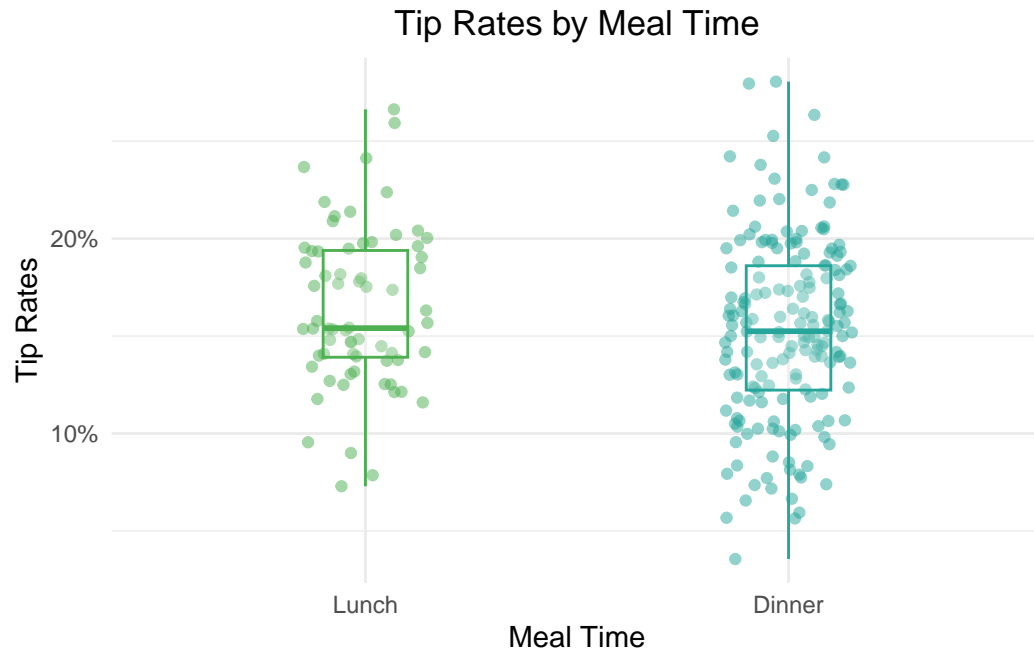


```
    Wilcoxon rank sum test with continuity correction

data:  tip_rate by time
W = 6577, p-value = 0.1328
alternative hypothesis: true location shift is not equal to 0
```

The box plot suggests that lunch customers tend to leave slightly higher tip rates than dinner customers. To examine whether this difference is statistically meaningful, a Wilcoxon rank-sum test was conducted. The result indicates no statistically significant difference in tip rate (p-value > 0.05). In other words, although lunch customers appear to tip slightly more, this difference is not statistically significant in this dataset.

4. Tip Rates by Day

For most people, Friday might be the happiest day of the week since it is the beginning of the weekend. It is reasonable to expect that customers might leave higher tips on Fridays since the atmosphere is more relaxed and social compared to other days. Does the data support this expectation?

```
# A tibble: 4 x 4
  day   count mean_tip_rate median_tip_rate
  <fct> <int>         <dbl>           <dbl>
1 Thur     62          16.1            15.4
2 Fri      19          17.0            15.6
3 Sat      85          15.0            15.0
4 Sun      74          15.6            16.0
```

Average Tip Rates by Day

Median Tip Rates by Day
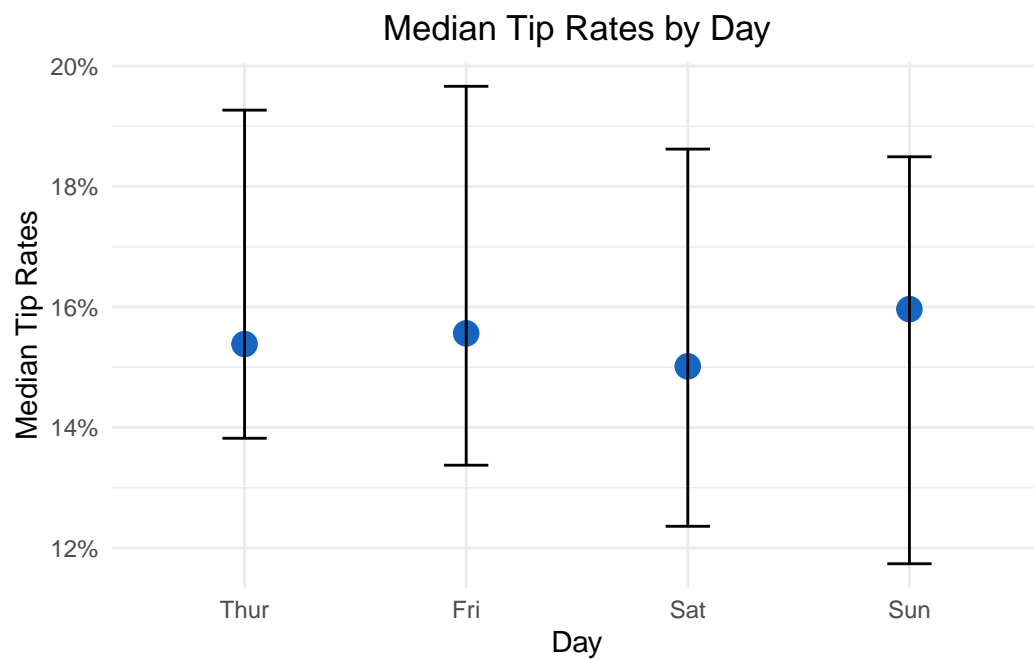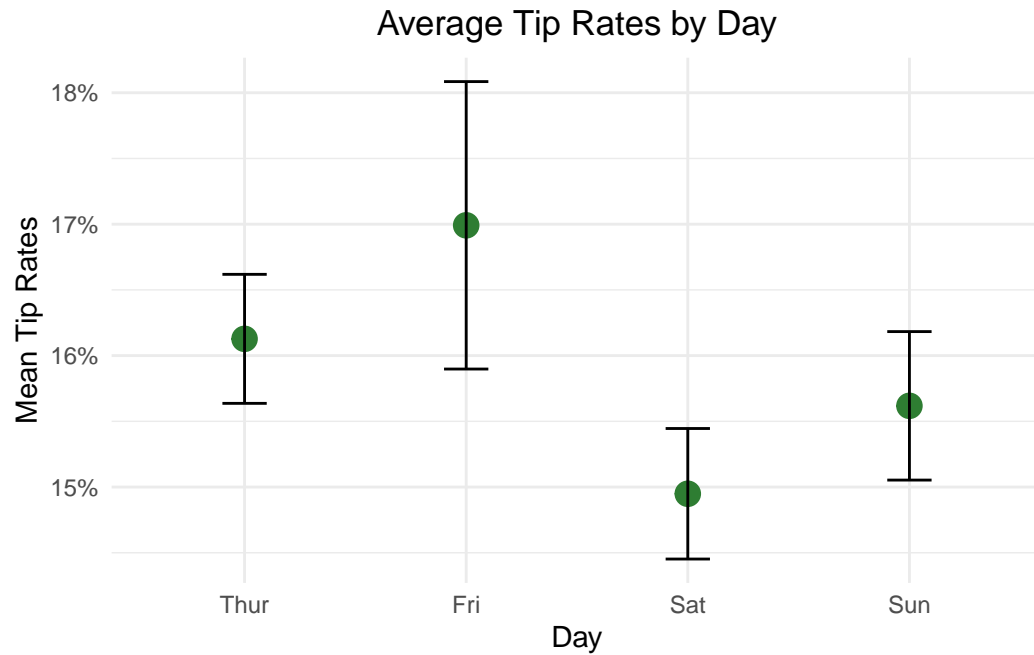

```
        Kruskal-Wallis rank sum test

data:  tip_rate by day
Kruskal-Wallis chi-squared = 2.622, df = 3, p-value = 0.4536
```

The graphs show that customers tend to leave higher tip rates on average, and there are occasional big tippers on Fridays. In contrast, Sundays show relatively stable and consistent tipping behavior

with a higher median tip rate. Saturday shows the lowest average and median tip rates among all days. To examine whether these differences are statistically meaningful, a Kruskal-Wallis rank sum test was conducted. The result indicates no statistically significant difference in tip rate across days (p-value > 0.05). In other words, while customers appear to leave higher tip rates on Friday and Sunday, the differences are not statistically significant in this dataset.

5. Tip Rates by Day and Meal Time

Even on the same day, tip rates may differ depending on the meal time. For example, it is reasonable to expect that customers might leave higher tip rates during Friday dinners since it is the start of the weekend and people tend to be in a more relaxed mood.

Tip Rates by Day and Meal Time
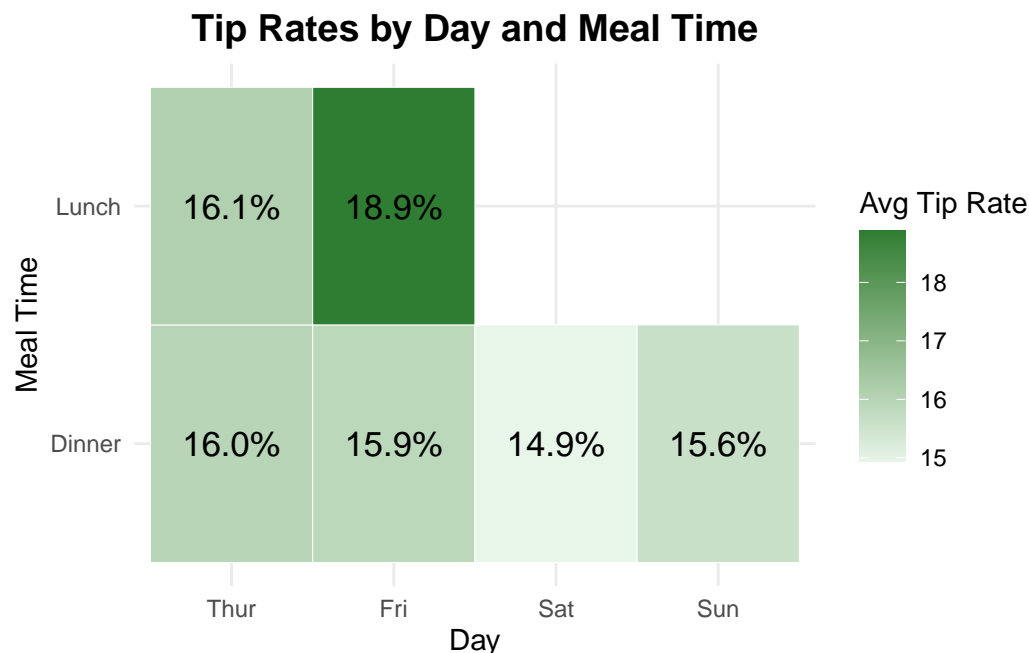
```
        Wilcoxon rank sum exact test

data:  tip_rate by time
W = 26, p-value = 0.1956
alternative hypothesis: true location shift is not equal to 0
```

In the previous analysis, Fridays showed a tendency for higher tip rates on average, with occasional big tippers. The heat map shows that Friday lunch customers tip slightly more than Friday dinner customers. To examine whether this difference is statistically meaningful, a Wilcoxon rank sum exact test was conducted. The result indicates no statistically significant difference in tip rate between Friday lunch and dinner (p-value > 0.05). In other words, although Friday lunch customers appear to tip more for Friday dinner customers, this difference is not statistically significant in this dataset.

6. Tip Rates and Total Bill

Many people might assume that customers who spend more on their meals tend to leave higher tip rates. It seems reasonable since they might be more generous tippers. Let's look at the data to find out whether this expectation holds true.

```
# A tibble: 2 x 4
  bill_group count mean_tip_rate median_tip_rate
  <chr>      <int>         <dbl>           <dbl>
1 High Bill    120          14.6            14.7
2 Low Bill     120          16.7            16.0
```

Tip Amount vs Total Bill
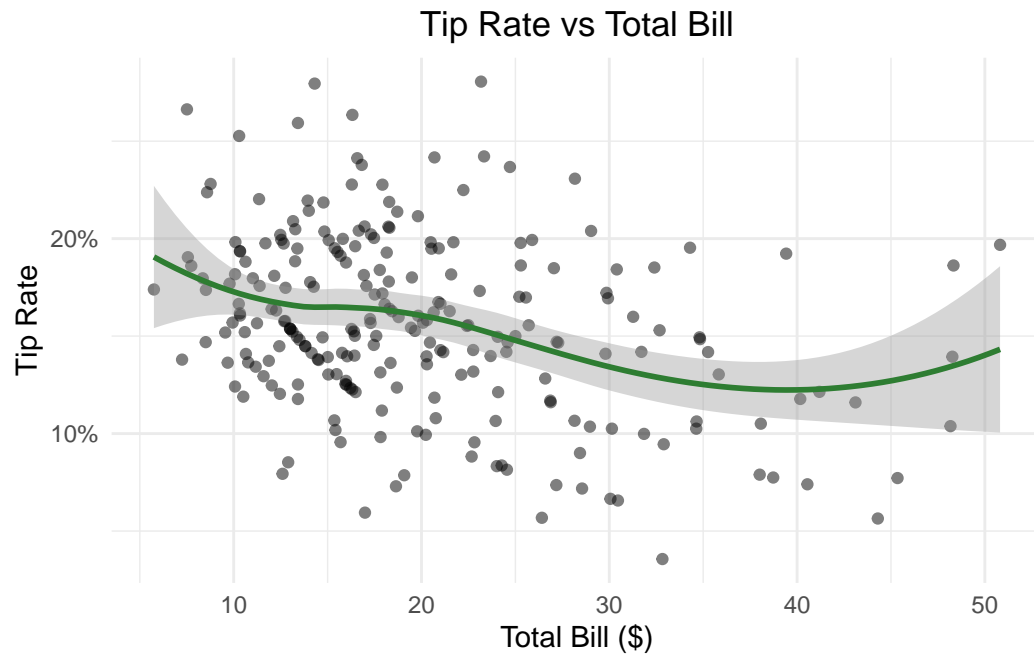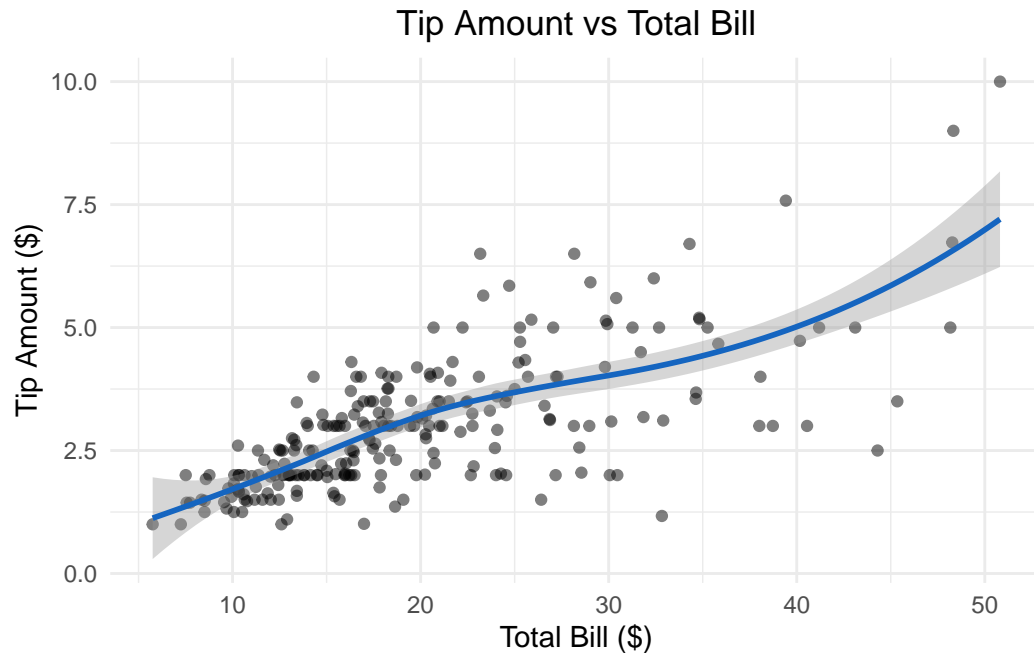

```
    Wilcoxon rank sum test with continuity correction

data:  tip_rate by bill_group
W = 5403, p-value = 0.0008359
alternative hypothesis: true location shift is not equal to 0
```

The tip amount vs total bill graph shows that tip amounts increase as total bills increase. However, The tip rate vs total bill graph shows a different pattern. Tip rates do not increase with spending and instead show a non-positive relationship. To investigate the pattern further, customers are divided into two groups, low bill and high bill, based on the median bill amount. Low bill customers tend to leave higher and more consistent tip rates than high bill customers. A Wilcoxon rank sum exact test confirms that the difference is statistically significant (p-value < 0.05). In other words, customer who spend less on their meals tend to leave higher tip rates consistently than customers who spend more.

Conclusion

Overall, the analysis suggests that typical tip rates in this moderately priced restaurant fall between 13% and 19%. Customers are generally expected to leave tip rates within the range.

Regarding what types of customers tend to tip more, group size, meal time and day did not show statistically significant associations with tip rates. However, this does not imply that these factors have no influence on tipping behavior; rather, the dataset may not have enough sample size or variation to detect strong effects. Prior research on tipping behavior in Europe suggests that tipping behavior is influenced more by customer characteristics such as income, than by situational or service factors (Fernandez et al. 2024). This supports findings of this study that although

smaller groups, lunch customers, and Friday/Sunday customers tended to leave slightly higher tip rates, the differences were not statistically significant.

However, it was statistically significant that customers who spent less on their meals tended to leave higher tip rates than those who spent more. This outcome aligns with previous research showing that tipping behavior is not determined solely by spending level and may reflect social norms and status-seeking motivations (Lynn 2025).

In summary, customer characteristics and tipping norms appear to have a stronger influence on tipping behavior, while situational and price-level factors provide some insight but have a more limited role.

References

Badole, Saurabh. 2024. "Restaurant Tips Dataset." https://www.kaggle.com/datasets/saurabhbadole/restaurant-tips-dataset.

Fernandez, Sebastien, Stefan Gossling, Carlos Martin-Rios, Valerie Fointiat, Susana Pasamar, Rami Isaac, and Merete Lunde. 2024. "To Tip or Not to Tip? Explaining Tipping Behavior in Restaurants with Service-Inclusive Pricing." *International Journal of Hospitality Management* 117: 103640. https://doi.org/10.1016/j.ijhm.2023.103640.

Lynn, Michael. 2025. "How Have u.s. Restaurant Tips Changed over Time?" *International Journal of Hospitality Management* 124: 103969. https://doi.org/10.1016/j.ijhm.2024.103969.

Wickham, Hadley. 2023. *Ggplot2: Elegant Graphics for Data Analysis.* https://CRAN.R-project.org/package=ggplot2.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2023. *Dplyr: A Grammar of Data Manipulation.* https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Dana Seidel, and Dewey Dunnington. 2023. *Scales: Scale Functions for Visualization.* https://CRAN.R-project.org/package=scales.