

# Analyse der Tipping Daten (Draft)

## 1. Hintergrund

In einem Restaurant in einer Shopping Mall, notiert eine Bedienung für 244 Rechnungen die folgenden Informationen:

1. Rechnungsbetrag (USD)
2. Trinkgeld (USD)
3. Geschlecht (der bezahlenden Person)
4. Raucher (ist ein Raucher in der Gruppe der Gäste)
5. Wochentag (Donnerstag bis Sonntag)
6. Tageszeit (Tag oder Nacht)
7. Anzahl der Personen in der Gruppe (1-6)

## 2. Ziel der Studie

Unter welchen Umständen kann die Bedienung hohe Trinkgelder erwarten?

## 3. Datenqualität und Transformation

Es existieren keine fehlende Werte, noch sind andere Fehler in den Daten auffällig. Als abgeleitete Variable wird die Variable "Trinkgeld %" erstellt aus Trinkgeld/Rechnungsbetrag.

## 4. Analyse

### 4.1 Allgemein

#### 4.1.1 Univariat

Für die kategoriellen Variablen ergeben sich folgende Tabellen:

```
> lapply(tips[,3:7], table)
$Sex
```

```
Female  Male
      87   157
```

```
$Smoker
```

```
  No  Yes
151   93
```

\$Day

Friday	Saturday	Sunday	Thursday
19	87	76	62

\$Time

Day	Night
68	176

\$Size.of.Party

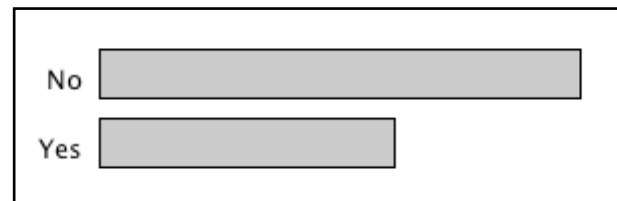
1	2	3	4	5	6
4	156	38	37	5	4

Mit den korrespondierenden Graphiken:

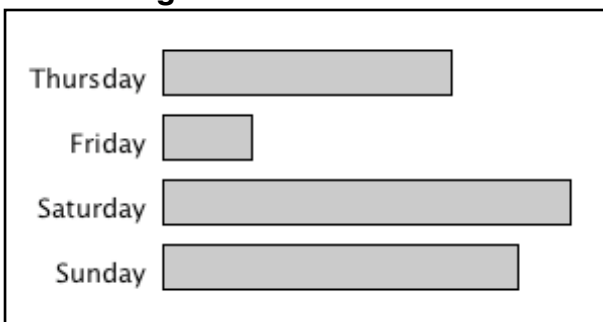
**Geschlecht:**



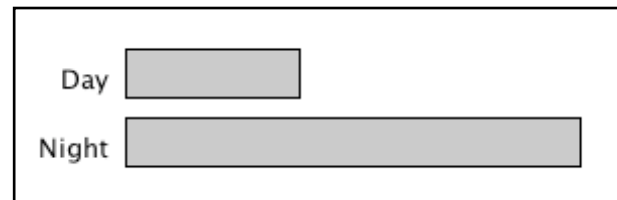
**Raucher:**



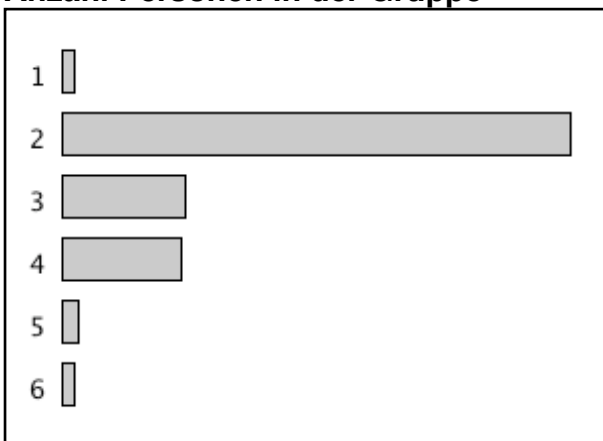
**Wochentag:**



**Tageszeit:**



**Anzahl Personen in der Gruppe**



Für die stetigen Variablen erhalten wir:

```
> lapply(tips[,-(3:7)], summary)
```

```
$Total.Bill.in...incl..Tax
```

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

```
$Tip.in..
```

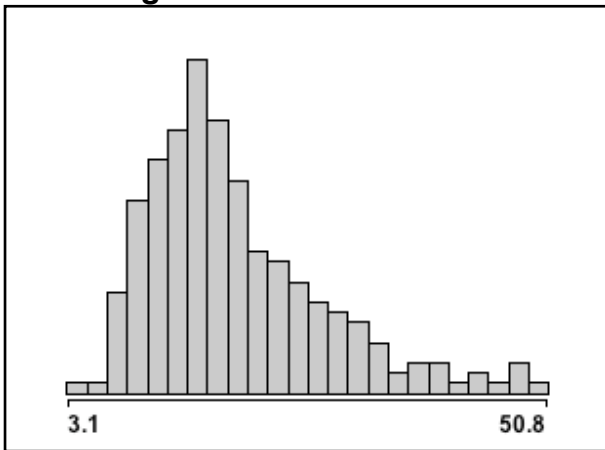
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.000	2.000	2.900	2.998	3.562	10.000

```
$Tip.in...of.Bill
```

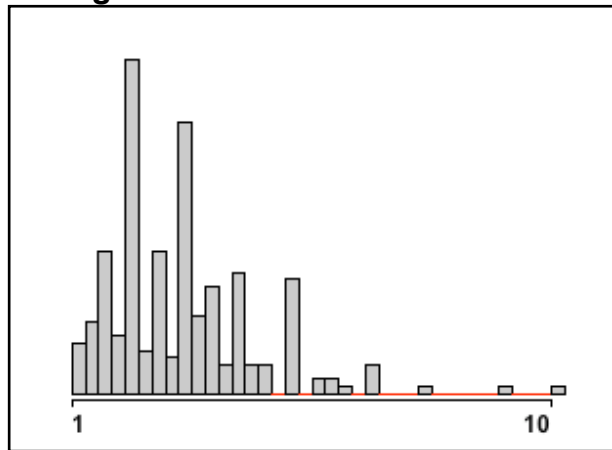
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.03564	0.12910	0.15480	0.16080	0.19150	0.71030

mit den korrespondierenden Graphiken:

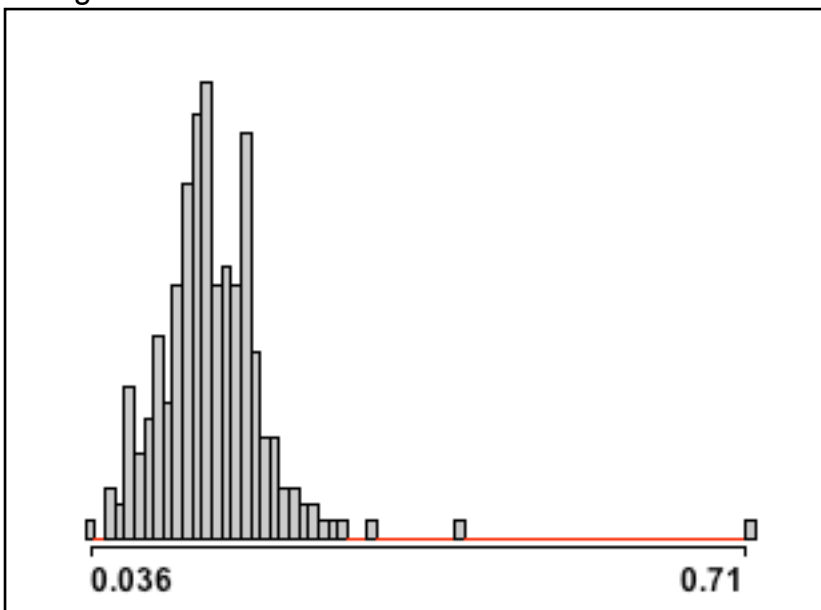
**Rechnung:**



**Trinkgeld:**



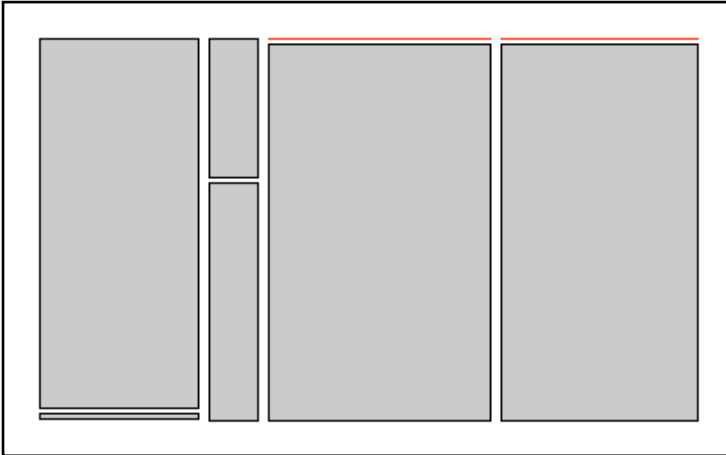
**Trinkgeldrate:**



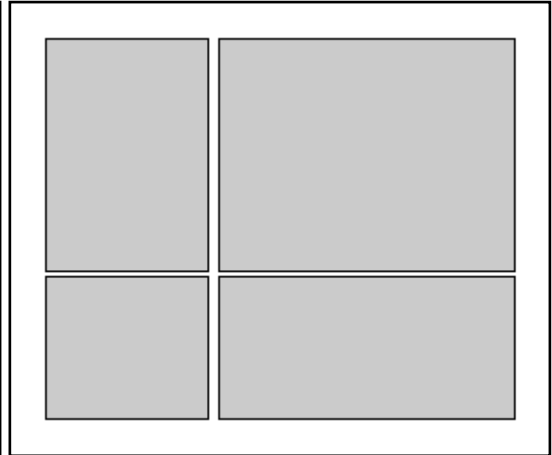
#### 4.1.2 Bivariat

Als relevante bivariate Interaktionen ergeben sich

##### Wochentag x Tageszeit

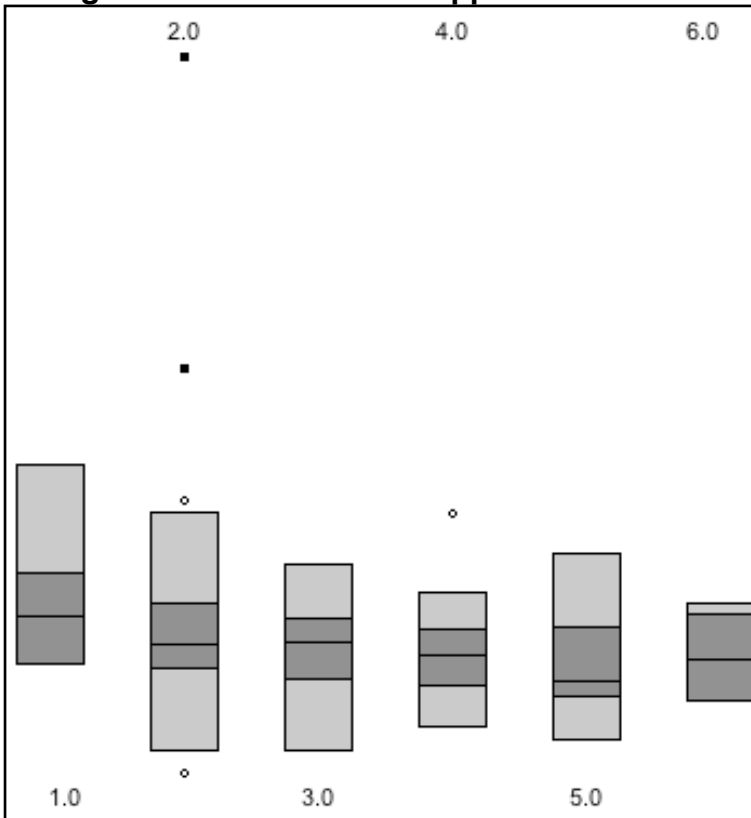


##### Geschlecht x Raucher



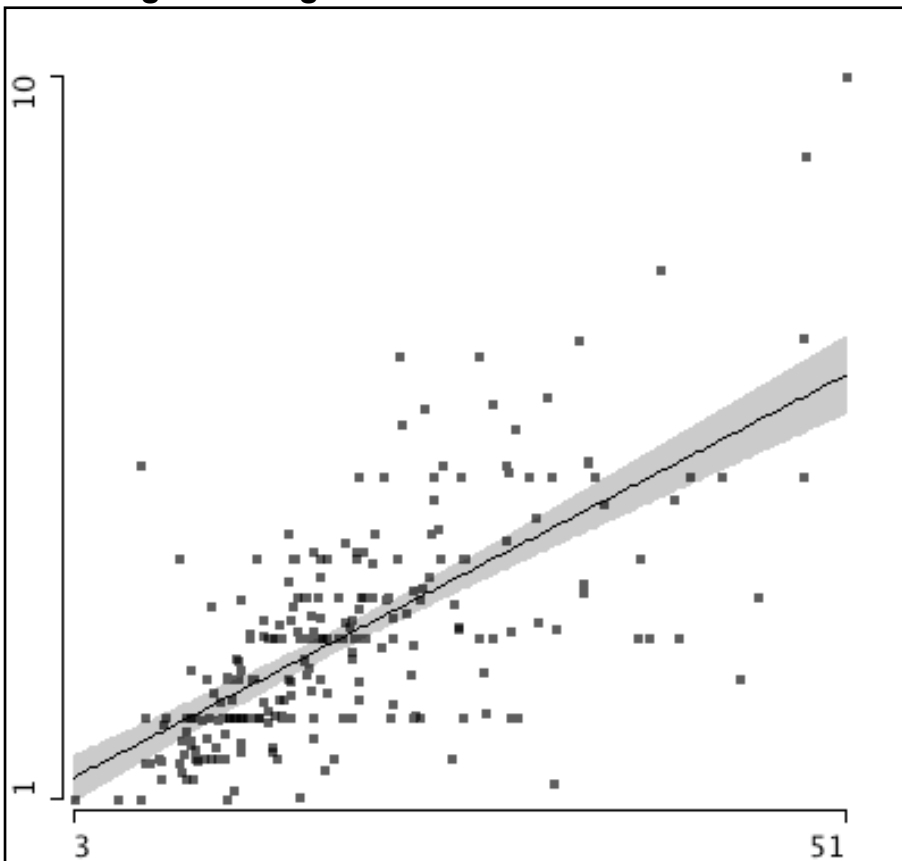
sowie

##### Trinkgeldrate | Größe der Gruppe



Die Betrachtung des Zusammenhangs zwischen Trinkgeld und Rechnungsbetrag liefert

## Rechnung vs. Trinkgeld



Als lineares Modell kann man die Trinkgeldrate mit

```
> lm(Tip.in..~Total.Bill.in...incl..Tax, data=tips)
```

Call:

```
lm(formula = Tip.in.. ~ Total.Bill.in...incl..Tax, data = tips)
```

Coefficients:

```
      (Intercept)  Total.Bill.in...incl..Tax  
            0.9203                0.1050
```

ohne Konstante schätzen.

```
> lm(Tip.in..~Total.Bill.in...incl..Tax -1, data=tips)
```

Call:

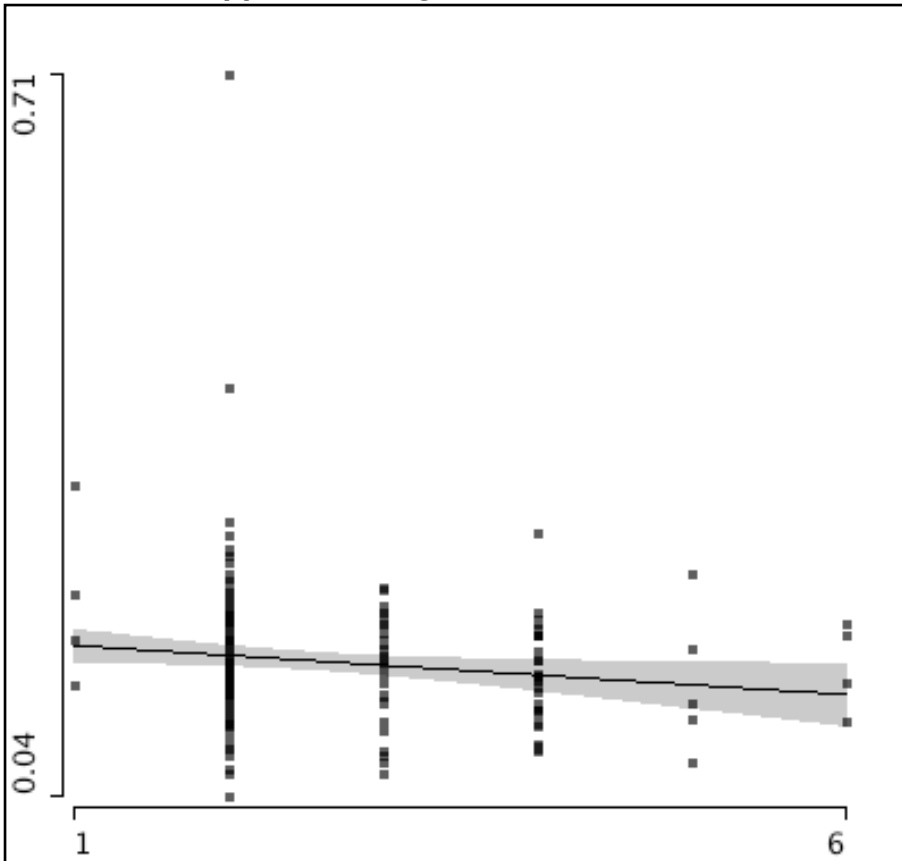
```
lm(formula = Tip.in.. ~ Total.Bill.in...incl..Tax - 1, data =  
tips)
```

Coefficients:

```
Total.Bill.in...incl..Tax  
            0.1437
```

Eine Betrachtung der Trinkgeldrate in Abhängigkeit von der Größe der Gruppe ergibt:

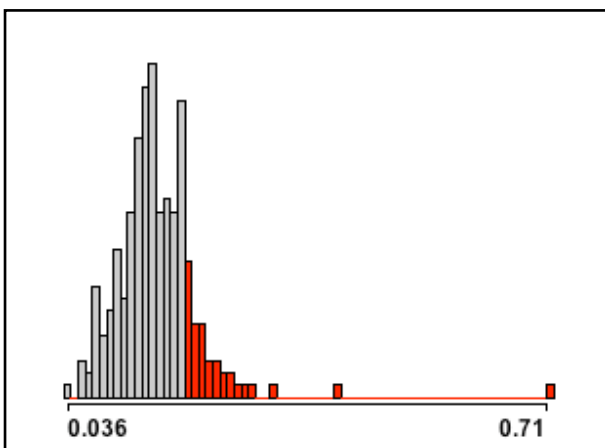
## Größe der Gruppe vs. Trinkgeldrate



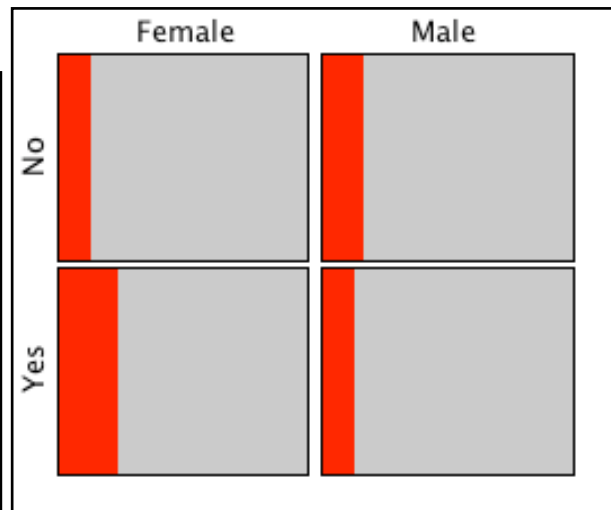
## 4.2 Optimierung der Trinkgeldes

Die stärkste und konsistenteste Abhängigkeit des Trinkgeldes ergibt sich auf den Kombinationen von Rauchverhalten und Geschlecht:

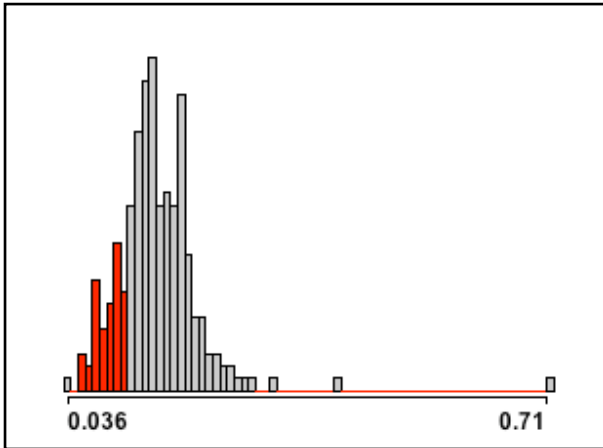
### Trinkgeldrate



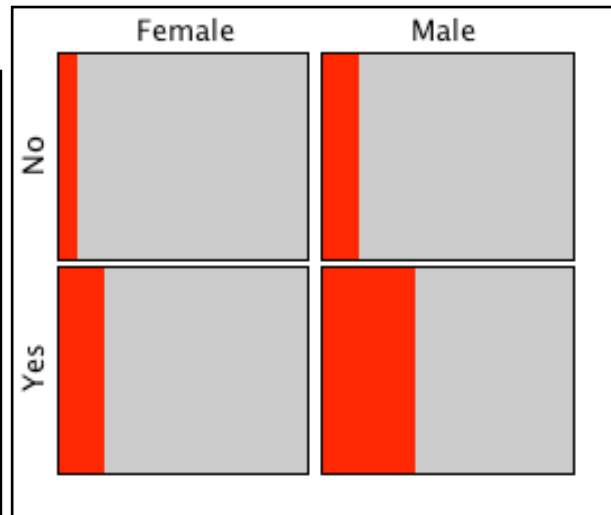
### Geschlecht x Raucher



## Trinkgeldrate



## Geschlecht x Raucher



Die Gruppengröße ist nicht sehr aussagekräftig, genau wie die zeitliche Information, da dort die Untergruppen zu klein sind für sinnvolle Aussagen.

## 5. Wesentliche Ergebnisse

### 5.1 Allgemein

Es zeigt sich, dass

1. sehr wenige Gruppen der Größe 1, 5 und 6 notiert wurden, sowie relativ wenige am Freitag
2. Rauchverhalten und Geschlecht unabhängig sind
3. Tageszeit und Wochentag extrem unbalanciert sind.
4. sehr hohe Trinkgeldraten mit kleinen Rechnungen assoziiert sind
5. die häufigste Trinkgeldrate bei 15% liegt
6. ...

### 5.2 Trinkgeldoptimierung

Annahme: Es ist sinnvoller die Trinkgeldrate als das Trinkgeld selber zu optimieren.

Hohe Trinkgeldraten ergeben sich für weibliche Nichtraucher, während extrem kleine Trinkgeldraten bei männlichen Rauchern beobachtet werden können.

## 6. Zusammenfassung

Die Daten können eine übliche Trinkgeldrate von 15%-20% bestätigen. Eine Aussage über hohe bzw. niedrige Trinkgeldraten ist nur bzgl. Rauchverhalten in Verbindung mit Geschlecht möglich.

...