

# gl05\_\_2\_\_hoermann

*Paul Hörmann*

*11/27/2019*

## Aufgabe 2: Tips

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:plyr':
##
##   arrange, count, desc, failwith, id, mutate, rename, summarise,
##   summarize

## The following objects are masked from 'package:data.table':
##
##   between, first, last

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

dtips = data.table(read.csv("./tips.csv"))

## Warning in scan(file = file, what = what, sep = sep, quote = quote, dec =
## dec, : EOF within quoted string

head(dtips)

##      X total_bill  tip    sex smoker day   time size
## 1: 1      16.99 1.01 Female    No  Sun  Dinner    2
## 2: 2      10.34 1.66   Male    No  Sun  Dinner    3
## 3: 3      21.01 3.50   Male    No  Sun  Dinner    3
## 4: 4      23.68 3.31   Male    No  Sun  Dinner    2
## 5: 5      24.59 3.61 Female    No  Sun  Dinner    4
## 6: 6      25.29 4.71   Male    No  Sun  Dinner    4

# ratio function
ratio = function(tip=tip, total_bill=total_bill) {
  round(sum(tip) / sum(total_bill) * 100, digits = 1)
}

# group by and calculate necessary values
aggregate(dtips$tip~sex + day, data = dtips, mean)

##      sex  day dtips$tip
## 1 Female  Fri  2.921667
## 2  Male  Fri  2.816667
## 3 Female  Sat  2.641333
## 4  Male  Sat  3.097500
## 5 Female  Sun  3.234444
## 6  Male  Sun  3.138571
```

```
## 7 Female Thur 2.711538
## 8 Male Thur 3.249091

dmodified = data.table(ddply(.data = dtips, .variables = c("sex", "day", "time"), .fun = summarize,
  n = length(tip),
  tip_mw = round(mean(tip), 2),
  tip_sd = round(sd(tip), 2),
  anteil = ratio(tip, total_bill)))

# order
dmodified$tip_sd = paste("(", dmodified$tip_sd, ")")
dmodified$tip_sd = paste(dmodified$tip_sd, " ", " ")
dmodified$tip = paste(dmodified$tip_mw, dmodified$tip_sd, " ");
dmodified$tip_mw = dmodified$tip_sd = NULL
dmodified$i <- seq.int(nrow(dmodified))
setcolorder(dmodified, c("i", setdiff(names(dmodified), "i")))
dnamed <- plyr::rename(dmodified, c("i" = "", "sex" = "Geschlecht", "day" = "Tag", "time" = "Zeit", "n" =
# add index column
# put index column to the first place
dnamed
```

```
##      Geschlecht Tag   Zeit Anzahl Anteil      Tip MW(SD)
## 1: 1      Female Fri Dinner      5  19.6 2.81 ( 1.21 )
## 2: 2      Female Fri  Lunch      1  25.9 3.48 ( NA )
## 3: 3      Female Sat Dinner     15  13.5 2.64 ( 0.93 )
## 4: 4      Female Sun Dinner      9  15.7 3.23 ( 1.37 )
## 5: 5      Female Thur Lunch     13  16.5 2.71 ( 1.24 )
## 6: 6        Male Fri Dinner      7  12.9 3.03 ( 1.21 )
## 7: 7        Male Fri  Lunch      2  19.9 2.06 ( 0.2 )
## 8: 8        Male Sat Dinner     32  15.8 3.1 ( 1.43 )
## 9: 9        Male Sun Dinner     28  15.6 3.14 ( 1.29 )
## 10: 10       Male Thur Lunch     11  15.8 3.25 ( 1.85 )
```

```
setcolorder(dnamed, c(setdiff(names(dnamed), "Anteil"), "Anteil"))
dnamed = dnamed[order(-Anteil)]
dnamed$Anteil = paste(dnamed$Anteil, "%", " ");
dnamed$Tag = mapvalues(dnamed$Tag, from = c("Thur", "Fri", "Sat", "Sun"), to = c("Do", "Fr", "Sa", "So"));
dnamed$Zeit = mapvalues(dnamed$Zeit, from = c("Lunch", "Dinner"), to = c("Mittag", "Abend"))
dnamed$Geschlecht = mapvalues(dnamed$Geschlecht, from = c("Female", "Male"), to = c("Männlich", "Weiblich"))
dnamed
```

```
##      Geschlecht Tag   Zeit Anzahl      Tip MW(SD) Anteil
## 1: 2    Männlich Fr Mittag      1  3.48 ( NA ) 25.9 %
## 2: 7    Weiblich Fr Mittag      2  2.06 ( 0.2 ) 19.9 %
## 3: 1    Männlich Fr  Abend      5  2.81 ( 1.21 ) 19.6 %
## 4: 5    Männlich Do Mittag     13  2.71 ( 1.24 ) 16.5 %
## 5: 8    Weiblich Sa  Abend     32  3.1 ( 1.43 ) 15.8 %
## 6: 10   Weiblich Do Mittag     11  3.25 ( 1.85 ) 15.8 %
## 7: 4    Männlich So  Abend      9  3.23 ( 1.37 ) 15.7 %
## 8: 9    Weiblich So  Abend     28  3.14 ( 1.29 ) 15.6 %
## 9: 3    Männlich Sa  Abend     15  2.64 ( 0.93 ) 13.5 %
## 10: 6   Weiblich Fr  Abend      7  3.03 ( 1.21 ) 12.9 %
```