gl05_3_hoermann

Paul Hörmann

11/30/2019

Aufgabe 3: Datensatz "Marketing"

```
##
## 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
```

```
dmarketing = data.table(read.csv("./Marketing.csv", sep = ";"))
head(dmarketing)
```

```
##    MarketSize LocationID AgeOfStore Promotion Week Sales
## 1:        373          3         12         1    1 44540
## 2:        373          3         12         1    2 37940
## 3:        373          3         12         1    3 45490
## 4:        373          3         12         1    4 46050
## 5:        253          7         15         1    1 42920
## 6:        253          7         15         1    2 42160
```

a

```
dmarketingEncoded <- mutate(dmarketing,
  Size = ifelse(MarketSize > 500, "Supermarkt",
                ifelse(MarketSize <= 301, "Geschaeft", "Markt")))
head(dmarketingEncoded)
```

```
##   MarketSize LocationID AgeOfStore Promotion Week Sales      Size
## 1        373          3         12         1    1 44540     Markt
## 2        373          3         12         1    2 37940     Markt
## 3        373          3         12         1    3 45490     Markt
## 4        373          3         12         1    4 46050     Markt
## 5        253          7         15         1    1 42920 Geschaeft
## 6        253          7         15         1    2 42160 Geschaeft
```

```
dSize = dmarketingEncoded %>% group_by(Size) %>% summarize(
  nr = length(LocationID)
```

```
)
dSize
```

```
## # A tibble: 3 x 2
##   Size          nr
##   <chr>      <int>
## 1 Geschaeft    208
## 2 Markt        196
## 3 Supermarkt   144
```

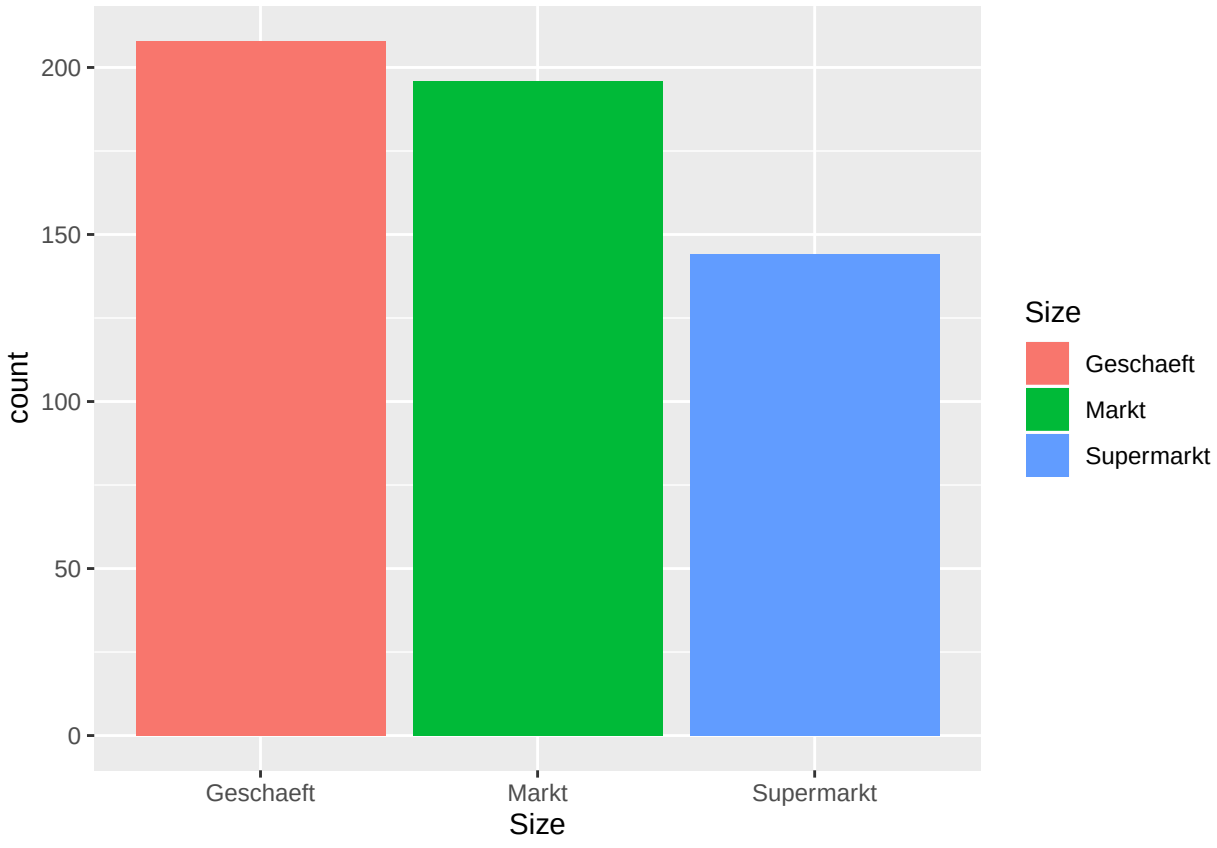
```
dmarketingEncoded %>% group_by(Promotion) %>% summarize(
  nr = length(LocationID)
)
```

```
## # A tibble: 3 x 2
##   Promotion    nr
##       <int> <int>
## 1         1   172
## 2         2   188
## 3         3   188
```

b

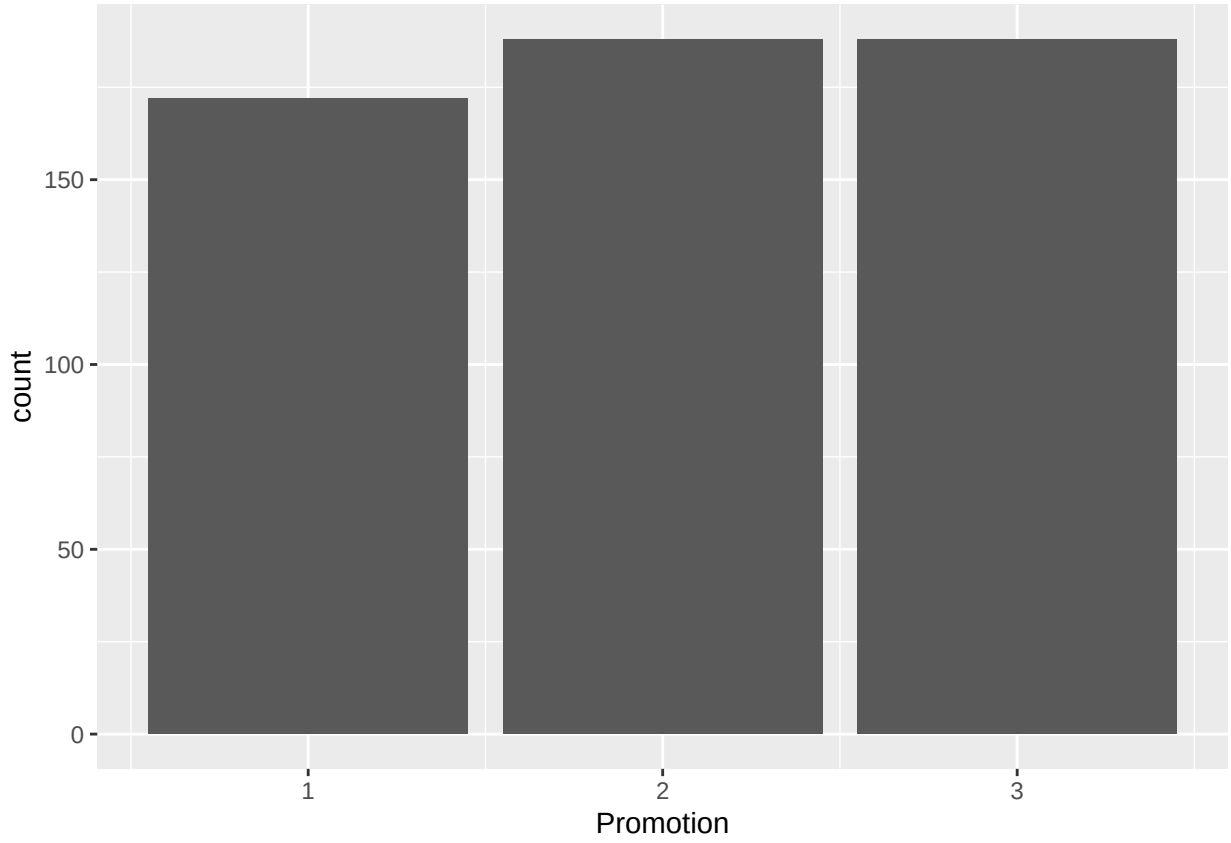
```
ggplot(dmarketingEncoded, aes(x=Size, fill=Size)) + geom_histogram(position = "dodge",stat = "count")
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```



```r
ggplot(dmarketingEncoded, aes(x=Promotion), fill=Psdfromotion) + geom_histogram(position = "dodge",stat
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```



c

```r
dmarketingByLoc = data.table(dmarketingEncoded %>% group_by(LocationID))

calculateSales = function (data, wstart, wend) {
  currentCompany = 0
  dLength = length(data$LocationID)
  data[, Diff := 0]
  index = 4 * currentCompany + 1
  while (index < dLength) {
    index = currentCompany * 4 + 1
    for (week in wstart:wend) {
      index = index + 1
      data[index]$Diff = data[index,]$Sales - data[index - 1,]$Sales
    }
    currentCompany = currentCompany + 1
  }
  data
}
dmarketingDiff = calculateSales(dmarketingByLoc, 2, 4)
```

Mittelwert Verkäufe erste Woche

```
msw1 = colMeans(dmarketingByLoc %>% filter(Week == 1) %>% select(Sales))
msw1
```

```
##    Sales
## 53790.58
```

Mittelwert +- SD

```
sdsw1 = dmarketingByLoc %>% filter(Week == 1) %>% summarize(SD = sd(Sales))

paste(c(round(msw1), " +- ", round(sdsw1)), collapse = '')
```

```
## [1] "53791 +- 16969"
```

Mittelwert & SD am Ende der Studie

```
msw4 = colMeans(dmarketingByLoc %>% filter(Week == 4) %>% select(Sales))
sdsw4 = dmarketingByLoc %>% filter(Week == 4) %>% summarize(SD = sd(Sales))

paste(c(round(msw4), " +- ", round(sdsw4)), collapse = '')
```

```
## [1] "55581 +- 16942"
```

Median, IQR & Spannweite zum Beginn der Studie

```
statw1 = dmarketingByLoc %>% filter(Week == 1) %>% summarize(Median = median(Sales), IQR = IQR(Sales), r
statw1
```

```
##   Median   IQR         range
## 1  50200 15630 24750 - 99650
```

MW aller numerischen Werte

```
mwnum = colMeans(dmarketingByLoc %>% select_if(is.numeric))
mwnum
```

```
##   MarketSize   LocationID   AgeOfStore    Promotion         Week
##   403.635036   479.656934     8.503650     2.029197     2.500000
##        Sales         Diff
## 54288.065693     0.000000
```

MW Verkäufe gruppiert nach Geschäftsgröße

```
statgsize = dmarketingByLoc %>% group_by(Size) %>% summarize(Mean = mean(Sales))
statgsize
```

```
## # A tibble: 3 x 2
##   Size         Mean
##   <chr>       <dbl>
## 1 Geschaeft  48716.
## 2 Markt      47992.
## 3 Supermarkt 70906.
```

e i

```
bySize = dmarketingDiff %>% group_by(Size)

ei = data.table(W1 = bySize %>% filter(Week == 1) %>% summarize(Sales = mean(Sales)),
                W1 = (bySize %>% filter(Week == 2) %>% summarize(Diff = mean(Diff)))[,2],
```

```
                W4 = (bySize %>% filter(Week == 4) %>% summarize(Diff = mean(Diff)))[,2])
ei
```

```
##       W1.Size W1.Sales   W1.Diff   W4.Diff
## 1:  Geschaeft 48005.77  873.2692 2318.0769
## 2:      Markt 47872.65 -745.3061 1858.9796
## 3: Supermarkt 70201.39 -590.0000 -168.8889
```

e ii

```
byPromo = dmarketingDiff %>% group_by(Promotion)

eii = data.table(W1 = byPromo %>% filter(Week == 1) %>% summarize(Sales = mean(Sales)),
                 W1 = (byPromo %>% filter(Week == 2) %>% summarize(Diff = mean(Diff)))[,2],
                 W4 = (byPromo %>% filter(Week == 4) %>% summarize(Diff = mean(Diff)))[,2])
eii
```

```
##    W1.Promotion W1.Sales   W1.Diff   W4.Diff
## 1:            1 58244.42 -966.0465  616.0465
## 2:            2 47730.21 -190.2128  535.1064
## 3:            3 55776.17  811.2766 3274.6809
```

e iii

```
byPromoAndSize = dmarketingDiff %>% group_by(Promotion, Size)

eiii = data.table(W1 = byPromoAndSize %>% filter(Week == 1) %>% summarize(Sales = mean(Sales)),
                 W1 = (byPromoAndSize %>% filter(Week == 2) %>% summarize(Diff = mean(Diff)))[,3],
                 W4 = (byPromoAndSize %>% filter(Week == 4) %>% summarize(Diff = mean(Diff)))[,3])
eiii
```

```
##    W1.Promotion    W1.Size W1.Sales    W1.Diff    W4.Diff
## 1:            1  Geschaeft 51692.86  -895.7143  -106.1905
## 2:            1      Markt 53445.45 -1581.8182  1505.4545
## 3:            1 Supermarkt 75550.91  -484.5455  1105.4545
## 4:            2  Geschaeft 41019.33  2124.0000  1495.3333
## 5:            2      Markt 44466.11 -1983.8889   977.7778
## 6:            2 Supermarkt 59117.14  -363.5714 -1062.8571
## 7:            3  Geschaeft 49716.25  2022.5000  6271.2500
## 8:            3      Markt 47873.50   829.5000  2846.5000
## 9:            3 Supermarkt 78959.09  -983.6364  -305.4545
```

d

```
t.test(ei[,2:4], eii[,2:4])
```

```
##
##  Welch Two Sample t-test
##
## data:  ei[, 2:4] and eii[, 2:4]
## t = 0.032552, df = 15.961, p-value = 0.9744
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -27039.57  27882.74
```

```
## sample estimates:
## mean of x mean of y
##  18847.33  18425.74
```

P Wert beinahe 1, daher akzeptieren wir H1, welche lautet dass die Werte nicht auf der gleichen Verteilung beruhen.