BDM600 - Lab 7 - Group 8

Ran Arino; Zubeka Dane Dang; Solmaz Heidar Nassab

2024-04-03

Credentials

All members participated in this lab assignment. The file is the merged version.

Start

```
library(vcd)
```

```
## Loading required package: grid
```

```
library(vcdExtra)
```

```
## Loading required package: gnm
```

```
library(logmult)
```

```
## Warning: package 'logmult' was built under R version 4.3.3
```

```
##
## Attaching package: 'logmult'
```

```
## The following object is masked from 'package:gnm':
##
##     se
```

```
## The following object is masked from 'package:vcd':
##
##     assoc
```

```
library(MASS)
library(Lahman)
```

```
## Warning: package 'Lahman' was built under R version 4.3.3
```

5.1

```r
# load data
data("criminal", package = "logmult")
criminal
```

```
##       Age
## Year    15  16  17  18  19
##   1955 141 285 320 441 427
##   1956 144 292 342 441 396
##   1957 196 380 424 462 427
##   1958 212 424 399 442 430
```

(a)

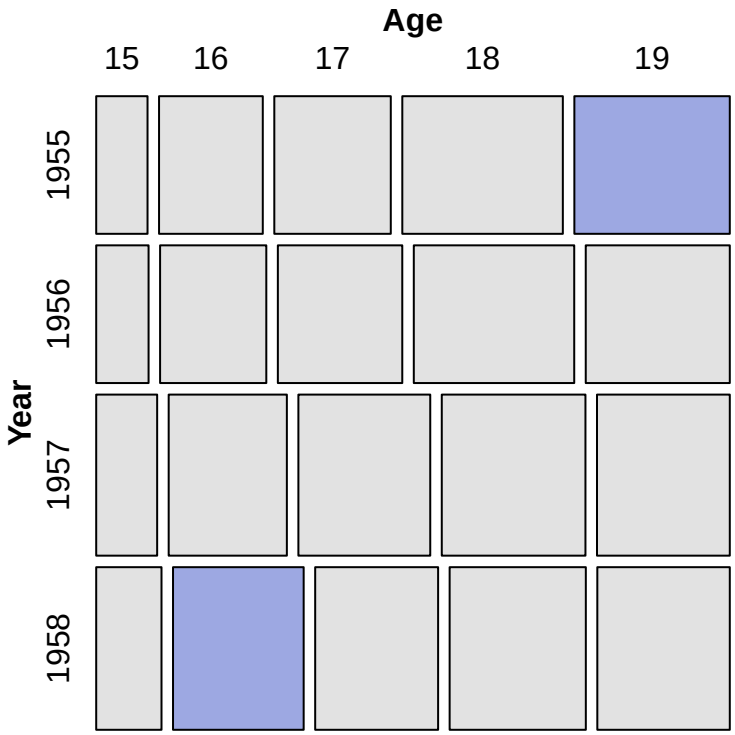
```r
model <- loglm(~ Year + Age, data = criminal)
summary(model)
```

```
## Formula:
## ~Year + Age
## attr(,"variables")
## list(Year, Age)
## attr(,"factors")
##      Year Age
## Year    1   0
## Age     0   1
## attr(,"term.labels")
## [1] "Year" "Age"
## attr(,"order")
## [1] 1 1
## attr(,"intercept")
## [1] 1
## attr(,"response")
## [1] 0
## attr(,".Environment")
## <environment: R_GlobalEnv>
##
## Statistics:
##                       X^2 df     P(> X^2)
## Likelihood Ratio 38.24466 12 0.0001400372
## Pearson          38.41033 12 0.0001315495
```

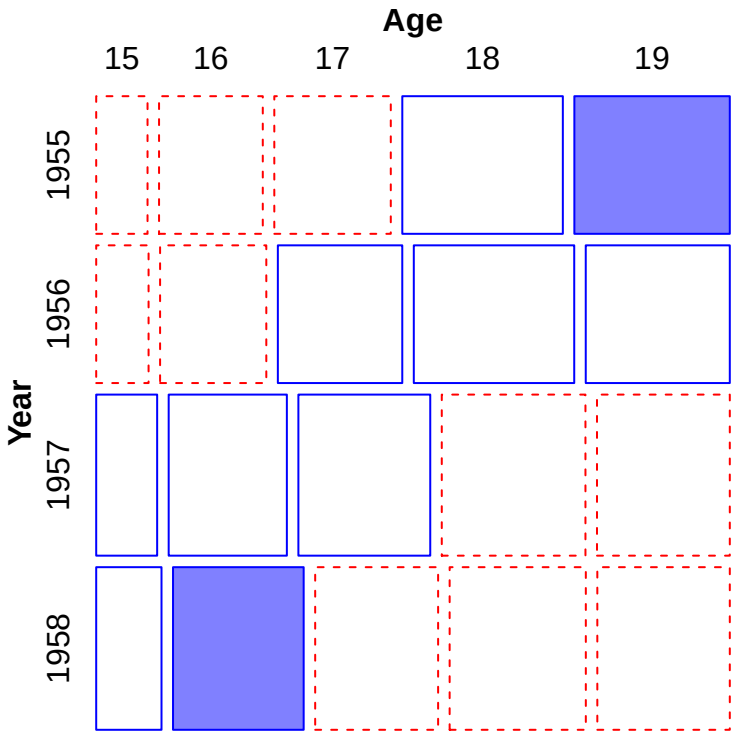
- Both the likelihood ratio and Pearson chi-squared tests have very low p-values (much less than 0.05).
- This result supports that the dropping of charges is associated with age, and this relationship has charged over the years recorded in the dataset.

(b)

```r
# mosaic with the option shade = TRUE
mosaic(criminal, shade = TRUE, legend = FALSE)
```




```
# Friendly shade
mosaic(criminal, shade = TRUE, legend = FALSE, gp = shading_Friendly)
```



- First plot: Darker shades in the combination of a few groups indicate high frequencies, Year 1995 and Age 19, as well as Year 1958 and Age 16. - Second plot: If we see that the cells with solid lines (which indicate higher-than-expected frequencies of dropped charges) are more prevalent in the younger age groups as time progresses from 1955 to 1958. - To wrap up, the mosaic plots, especially with Friendly shade, shows that there is an observable trend where younger men aged 15-19 were more likely to have charges dropped as time went on from 1955 to 1958.

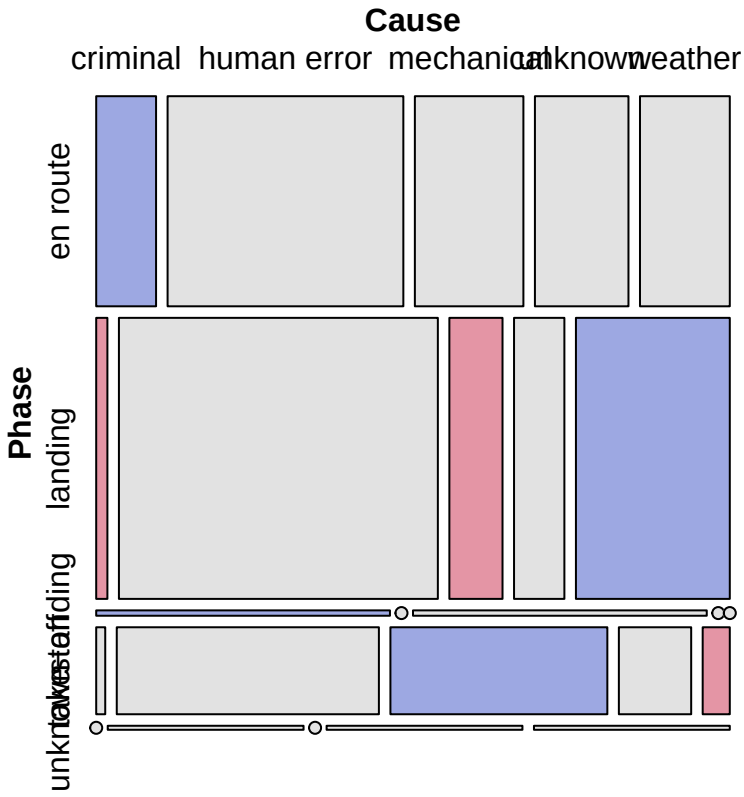
5.2

```
data("AirCrash", package = "vcdExtra")
aircrash_tab <- xtabs(~ Phase + Cause, data = AirCrash)
aircrash_tab
```

```
##           Cause
## Phase      criminal human error mechanical unknown weather
##   en route       16          63         29      25      24
##   landing         4         114         19      18      55
##   standing        2           0          2       0       0
##   take-off        1          29         24       8       3
##   unknown         0           1          0       1       1
```

(a)

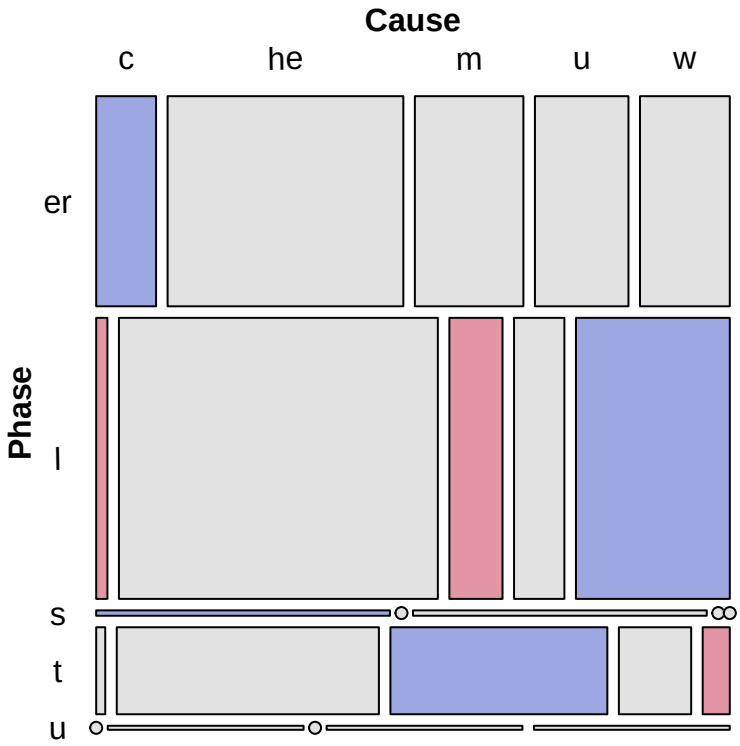
```
mosaic(aircrash_tab, shade = TRUE, legend = FALSE)
```



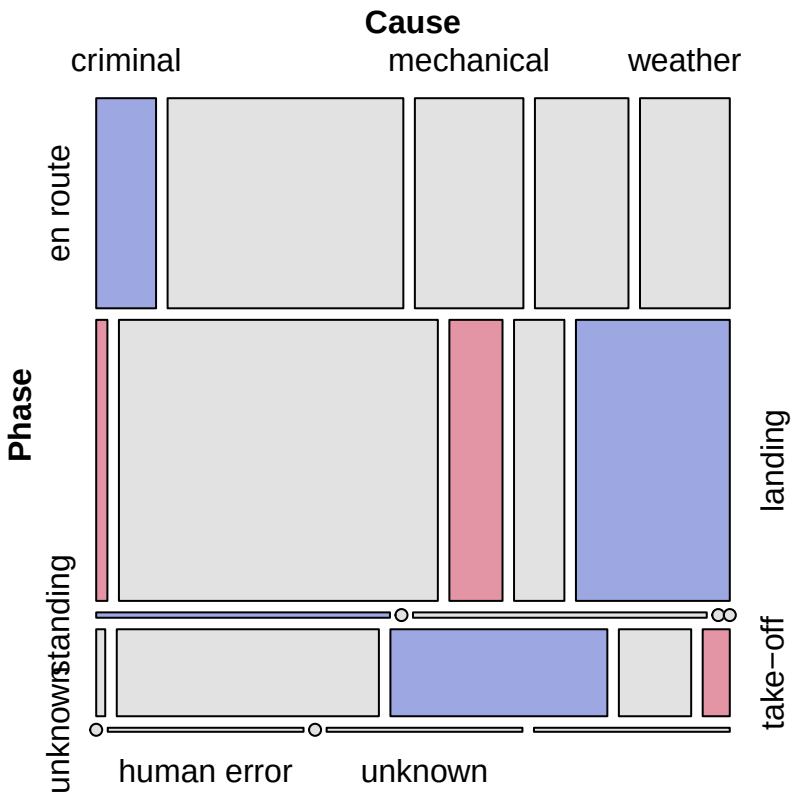
- Here are the combinations that show higher-than-expected frequencies. - En route vs criminal - Landing vs Weather - Take-off vs mechanical - However, interpreting the cells labeled as “unknown” may pose challenges due to insufficient information.

(b)

```
# shortest possible labels
abbrev <- list(abbreviate = TRUE, rot_labels = TRUE, cex = 0.8)
mosaic(aircrash_tab, shade = TRUE, legend = FALSE, labeling_args = abbrev)
```

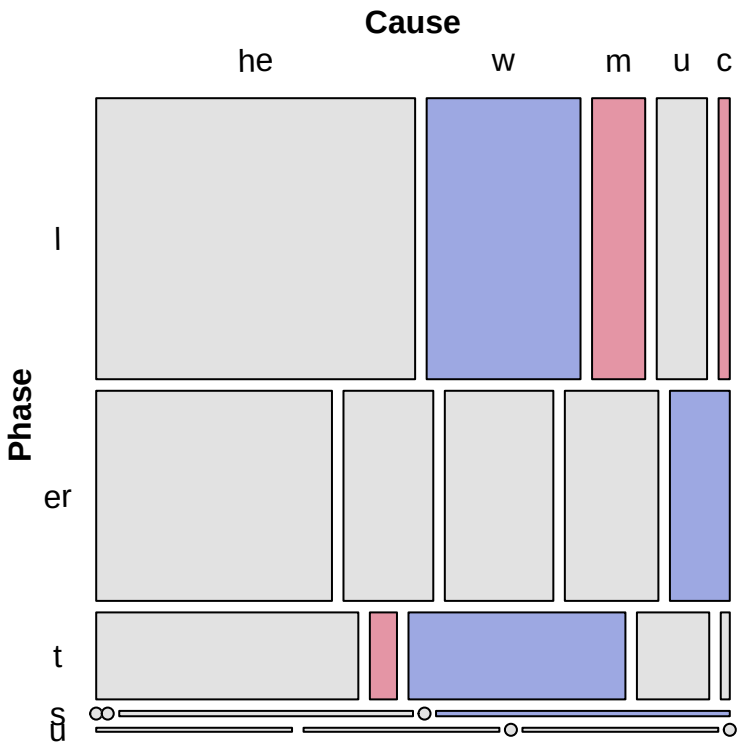



```
# alternate labels
mosaic(aircrash_tab, shade = TRUE, alternate_labels = TRUE, legend = FALSE)
```



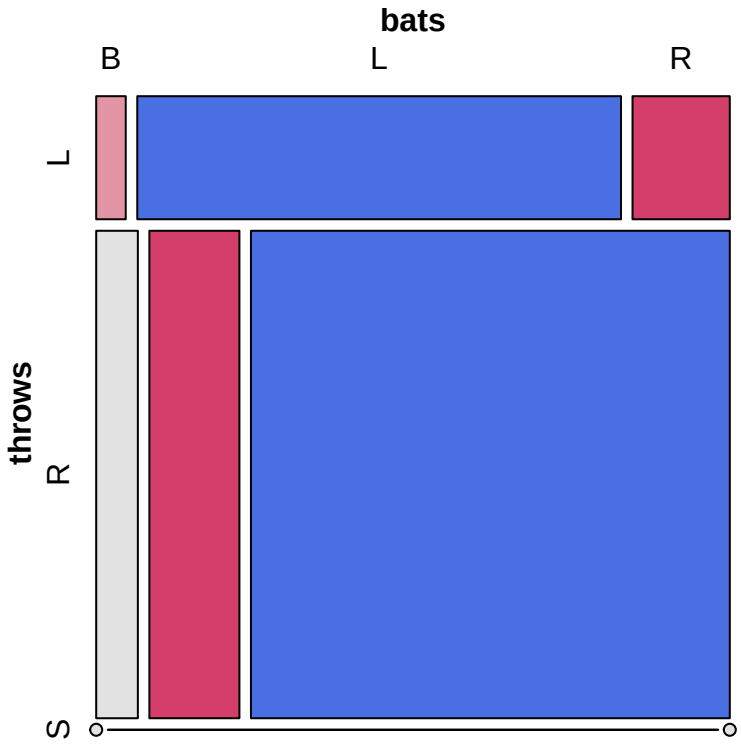
(c)

```r
# Calculate marginal frequencies for each factor
phase_freq <- sort(margin.table(aircrash_tab, 1), decreasing = TRUE)
cause_freq <- sort(margin.table(aircrash_tab, 2), decreasing = TRUE)

# Reorder factors in the data frame
AirCrash$Phase <- factor(AirCrash$Phase, levels = names(phase_freq))
AirCrash$Cause <- factor(AirCrash$Cause, levels = names(cause_freq))

# Now create the mosaic plot with the reordered factors
abbrev <- list(abbreviate = TRUE, rot_labels = TRUE, cex = 0.8)
mosaic(~ Phase + Cause, data = AirCrash, shade = TRUE, legend = FALSE, labeling_args = abbrev)
```



5.3

```
# renamed the file from "Master" to "People"
data("People", package = "Lahman")
basehands <- with(People, table(throws, bats))

mosaic(basehands, shade = TRUE, legend = FALSE)
```



- There are more than expected number of people who throw and swing with the same handedness (left-handed throwers typically bat left-handedly, and right-handed throwers predominantly bat right-handedly).
- On the other hand, there are less than expected number of people who throw and swing with different handedness.

5.7

```
# load data
data("caith", package = "MASS")
caith
```

```
##        fair red medium dark black
## blue    326  38    241  110     3
## light   688 116    584  188     4
## medium  343  84    909  412    26
## dark     98  48    403  681    85
```

(a)

```
# create matrix
caith.mat <- as.matrix(caith)
dimnames(caith.mat)
```

```
## [[1]]
## [1] "blue"   "light"  "medium" "dark"
##
## [[2]]
## [1] "fair"   "red"    "medium" "dark"   "black"
```

```
names(dimnames(caith.mat)) <- c("Eye", "Hair")
caith.mat
```

```
##         Hair
## Eye      fair red medium dark black
##   blue    326  38    241  110     3
##   light   688 116    584  188     4
##   medium  343  84    909  412    26
##   dark     98  48    403  681    85
```

(b)

```
# fit the model using loglm()
caith.model <- loglm(~ Hair + Eye, data = caith.mat, fitted = TRUE)
caith.model
```

```
## Call:
## loglm(formula = ~Hair + Eye, data = caith.mat, fitted = TRUE)
##
## Statistics:
##                       X^2 df P(> X^2)
## Likelihood Ratio 1218.314 12        0
## Pearson          1240.039 12        0
```

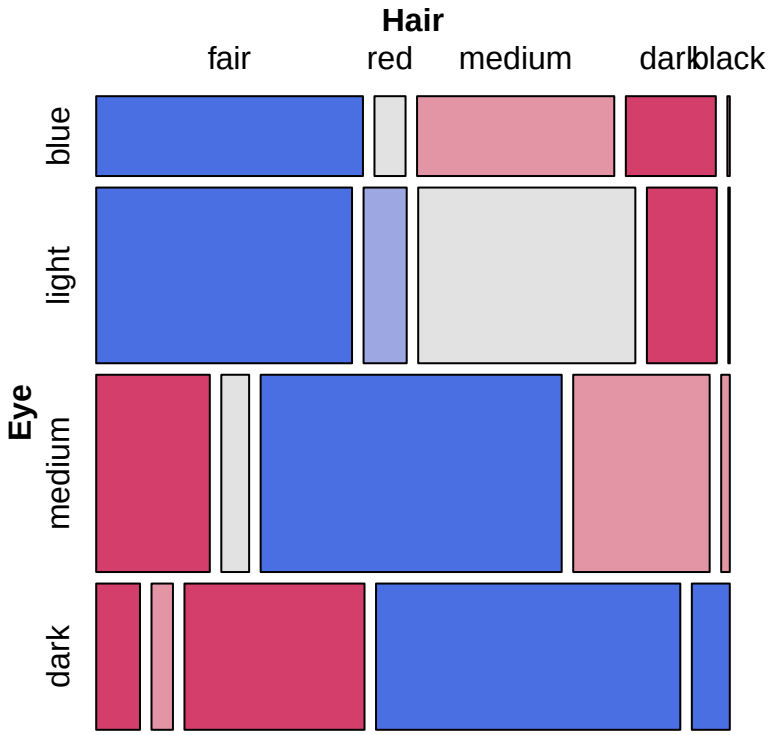
(c)

```
# calculate residuals for this model
residuals(caith.model)
```

```
##         Hair
## Eye            fair         red    medium       dark      black
##   blue     8.632012 -0.01931269 -2.668179  -5.995870 -3.938806
##   light   11.604525  3.31247206 -1.728753 -12.191903 -6.629923
##   medium  -6.555953 -1.06909312  7.400871  -2.190322 -2.195900
##   dark   -16.185525 -2.76828628 -5.413467  16.277849  8.459859
```

(d)

```
# display mosaic
mosaic(caith.mat, shade = TRUE, legend = FALSE)
```


Hair
fair
red
medium
dark
black
Eye
blue
light
medium
dark