Stat 557 - Homework 1

Carson Sievert

September 2, 2012

Problem 1.1:

```
data(UCBAdmissions)
odds.ratio <- function(x) {
    return(x[1] * x[4]/(x[2] * x[3]))
}
UCB <- as.data.frame(UCBAdmissions)
marginal <- ddply(UCB, .(Admit, Gender), summarize, Freq = sum(Freq))
odds.ratio(marginal$Freq)

## [1] 1.841
```

The odds ratio indicates that males have nearly twice the odds of women at being admitted.

Problem 1.2:

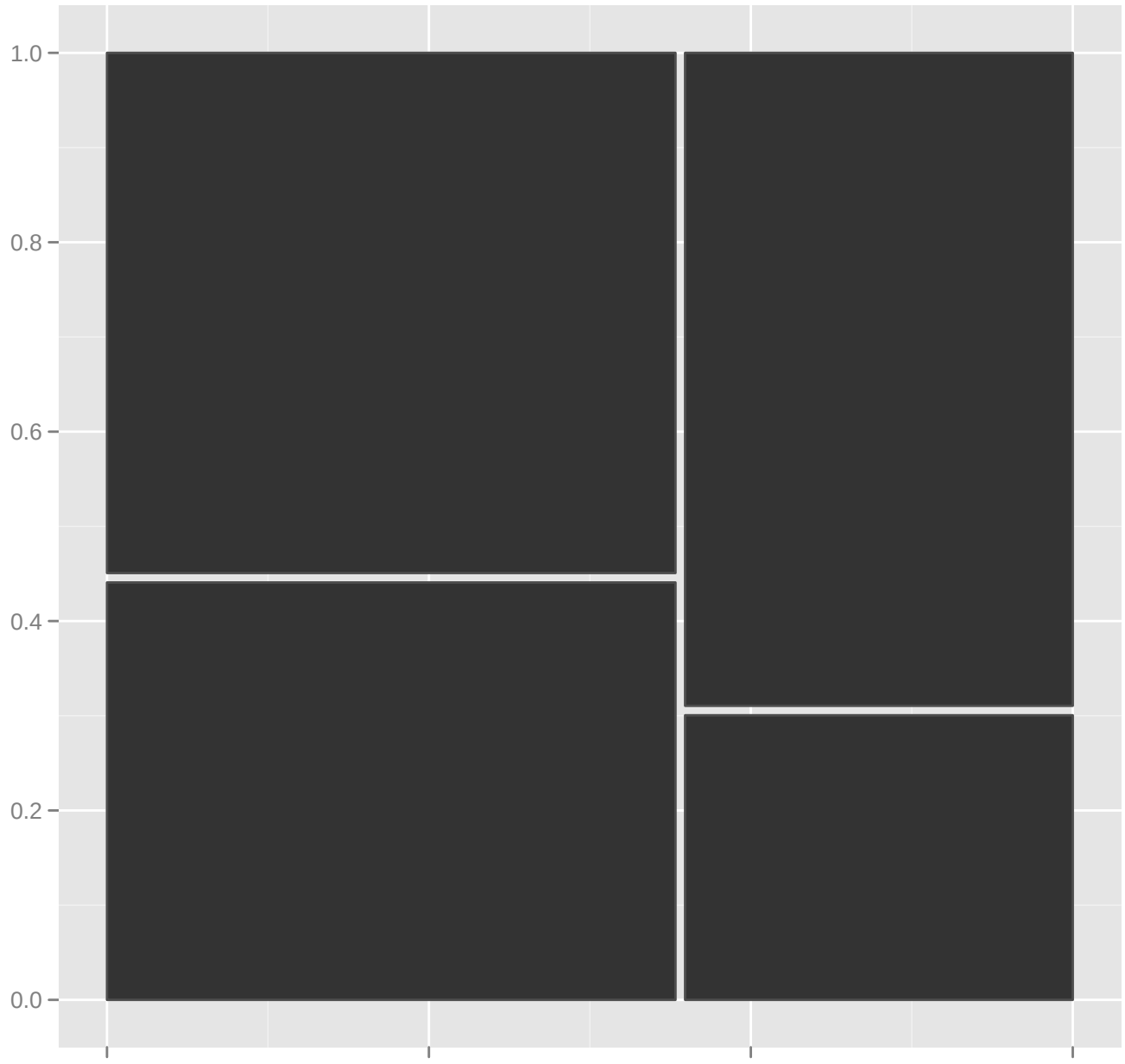
```
ddply(UCB, .(Dept), summarize, odds.ratio = odds.ratio(Freq))

##   Dept odds.ratio
## 1    A     0.3492
## 2    B     0.8025
## 3    C     1.1331
## 4    D     0.9213
## 5    E     1.2216
## 6    F     0.8279
```

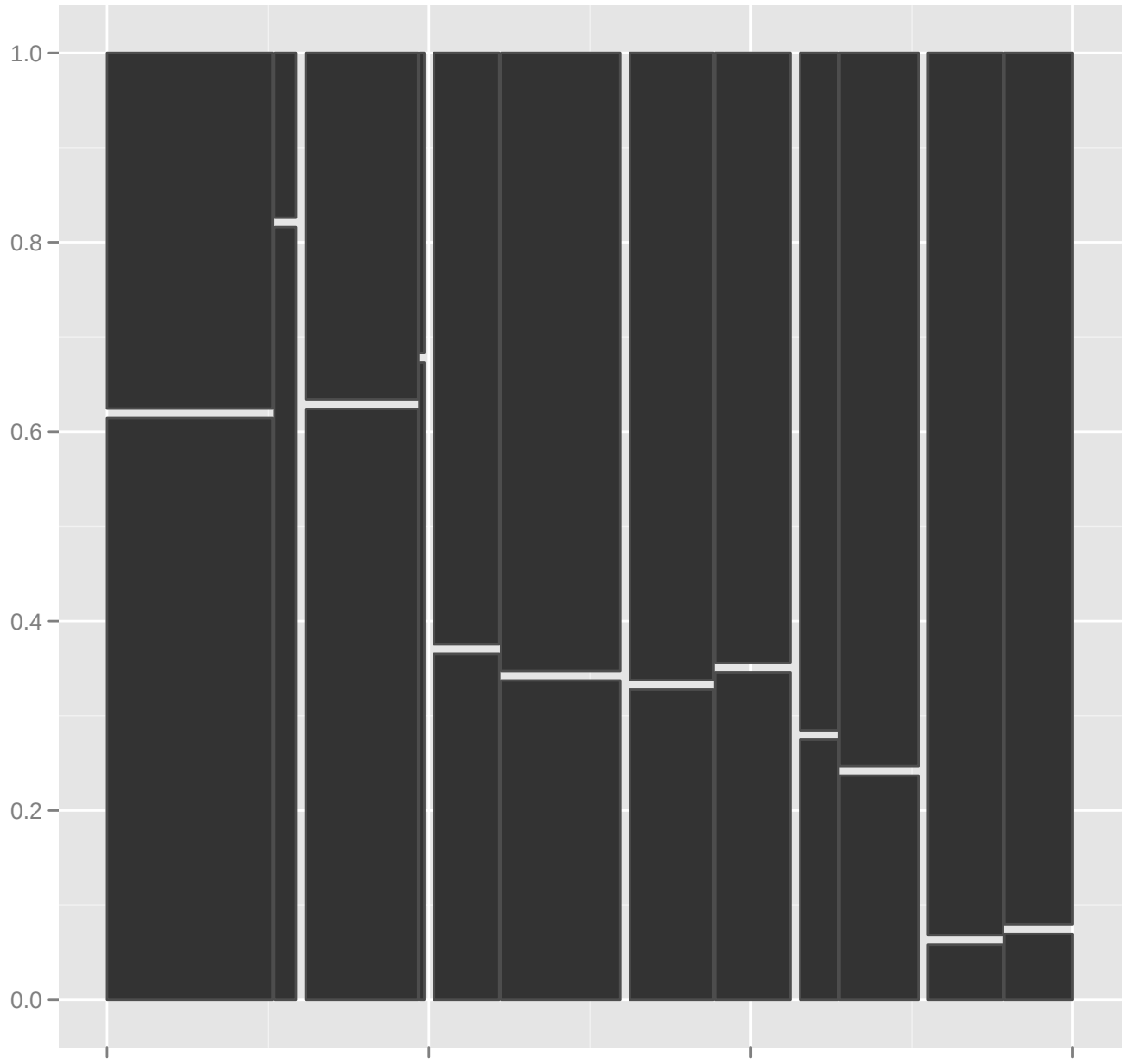
This portrays an example of Simpson's paradox. That is, the most of the conditonal associations of gender and admit rates (given department) indicate that women have better odds than men at being admitted.

Problem 1.3:

```
# Mosaic plot of marginal association
prodplot(marginal, Freq ~ Admit + Gender, c("vspine", "hspine"), subset = level ==
    2)
```



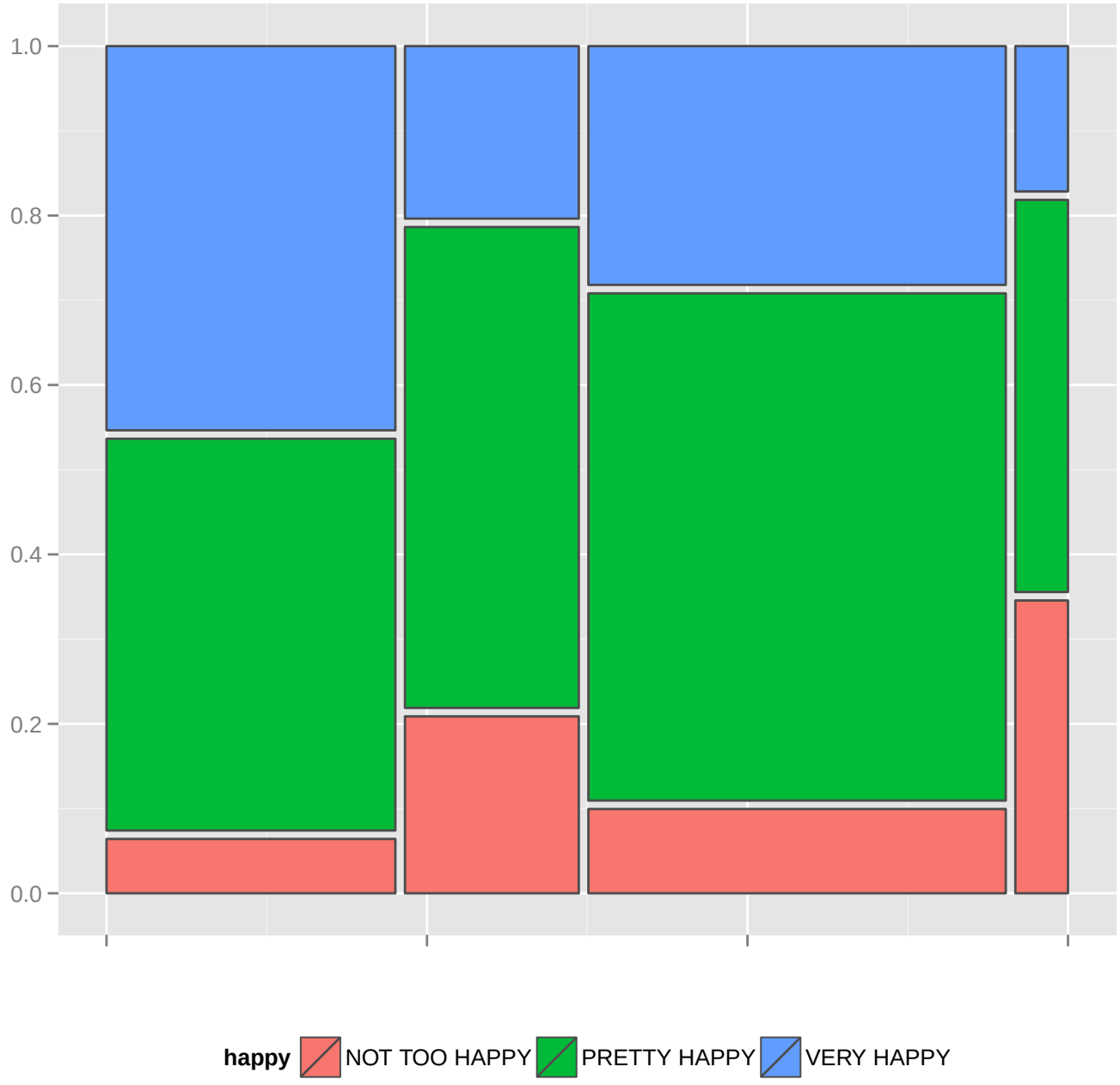

```
# Mosaic plot of conditional associations
prodplot(UCB, Freq ~ Admit + Gender + Dept, c("vspine", "hspine", "hspine"), subset =
level ==
    3)
```



Problem 2:

```
setwd("~/Dropbox/ISU/Classes/557/HW1/data/")
p <- read.csv("politics.csv")
p2 <- p[complete.cases(p), ]
p3 <- subset(as.data.frame(xtabs(~happy + sex + race + health, data = p2)), Freq >
    0)
```

```
prodplot(p3, Freq ~ happy + health, c("vspine", "hspine")) + aes(fill = happy) +
    opts(legend.position = "bottom", legend.direction = "horizontal")
```

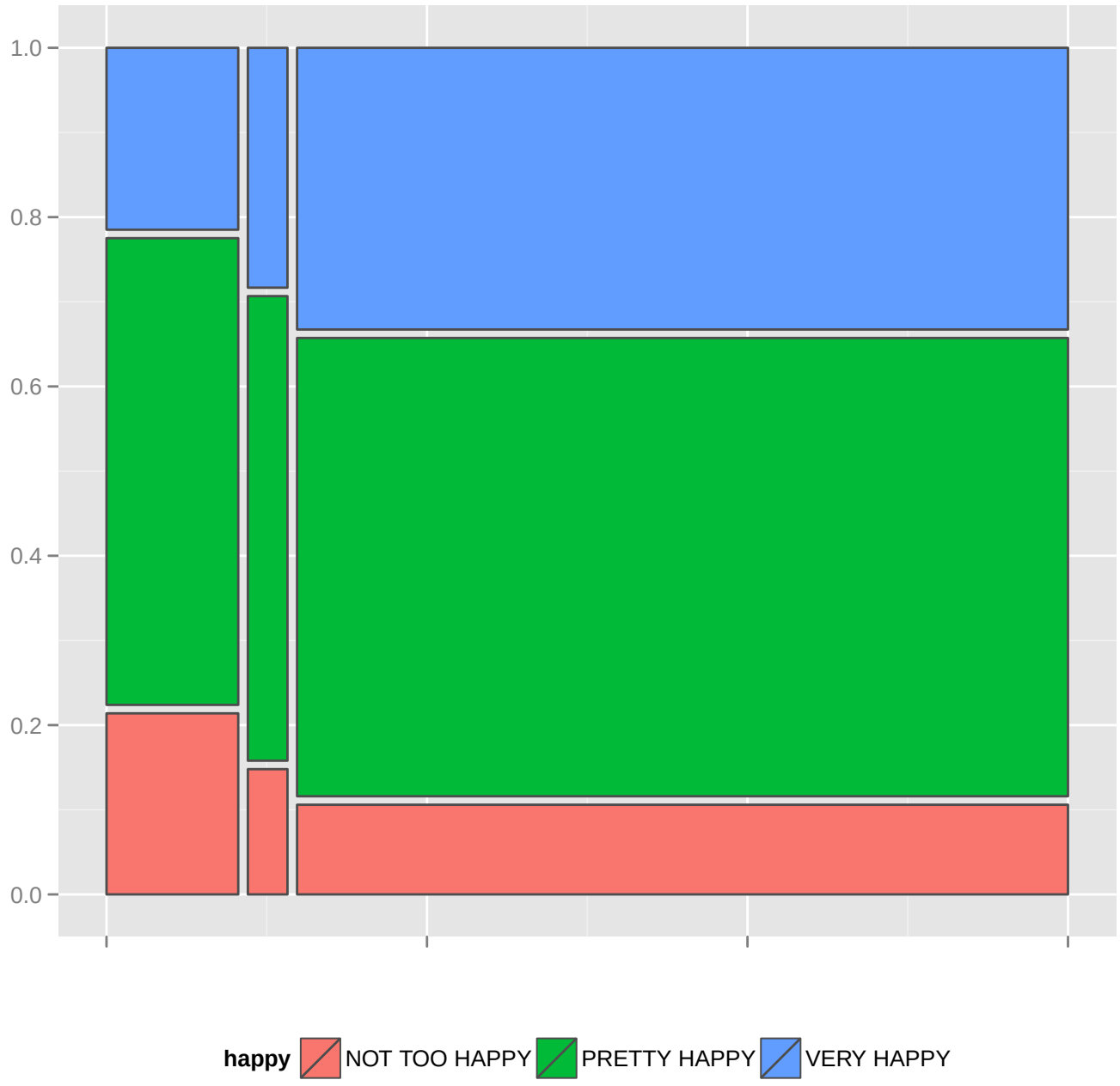



```
unique(p3$health)

## [1] EXCELLENT FAIR      GOOD      POOR
## Levels: EXCELLENT FAIR GOOD POOR
```

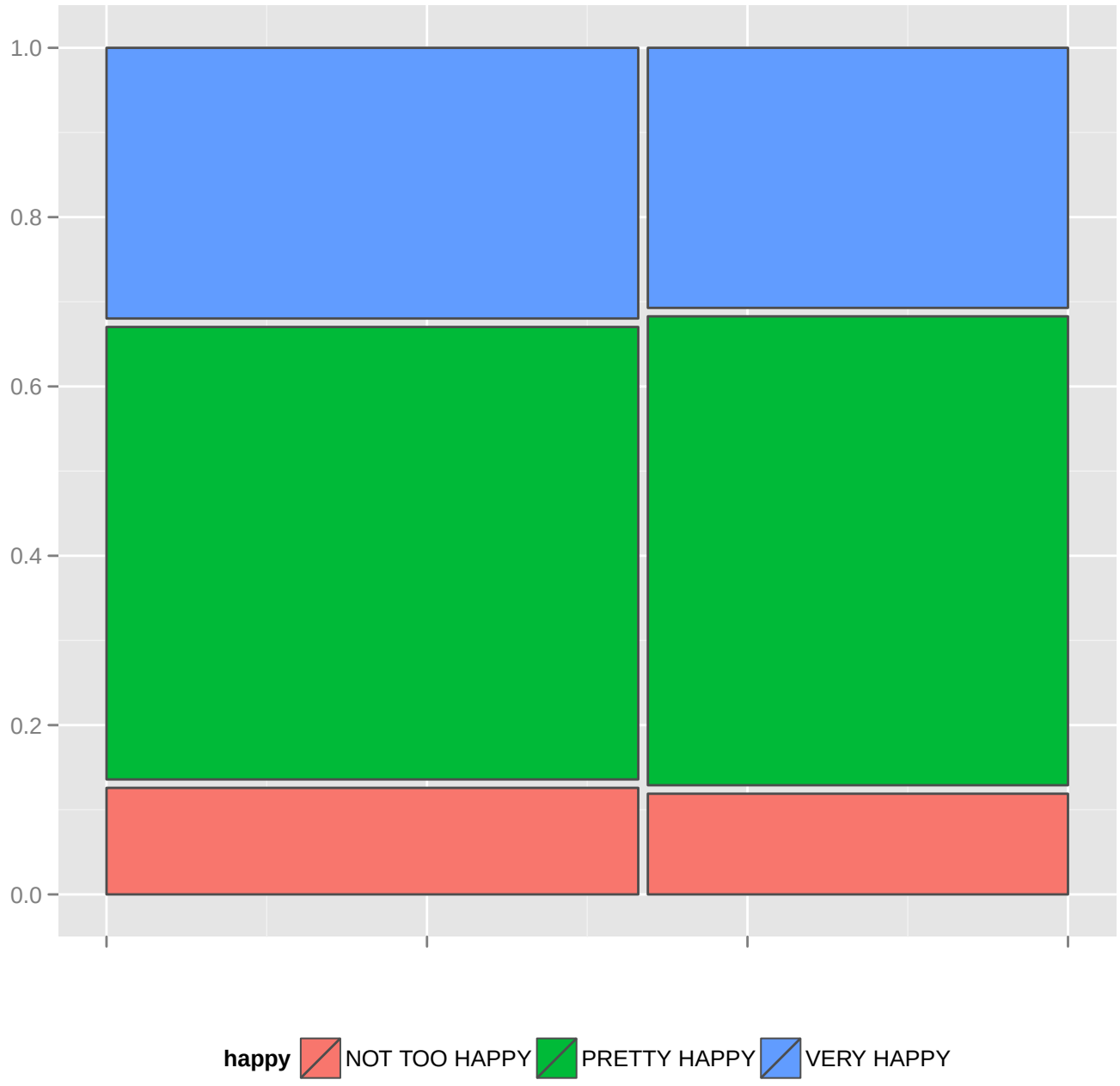
As expected, happiness decreases as health deteriorates.

```
prodplot(p3, Freq ~ happy + race, c("vspine", "hspine")) + aes(fill = happy) +
opts(legend.position = "bottom",
    legend.direction = "horizontal")
```



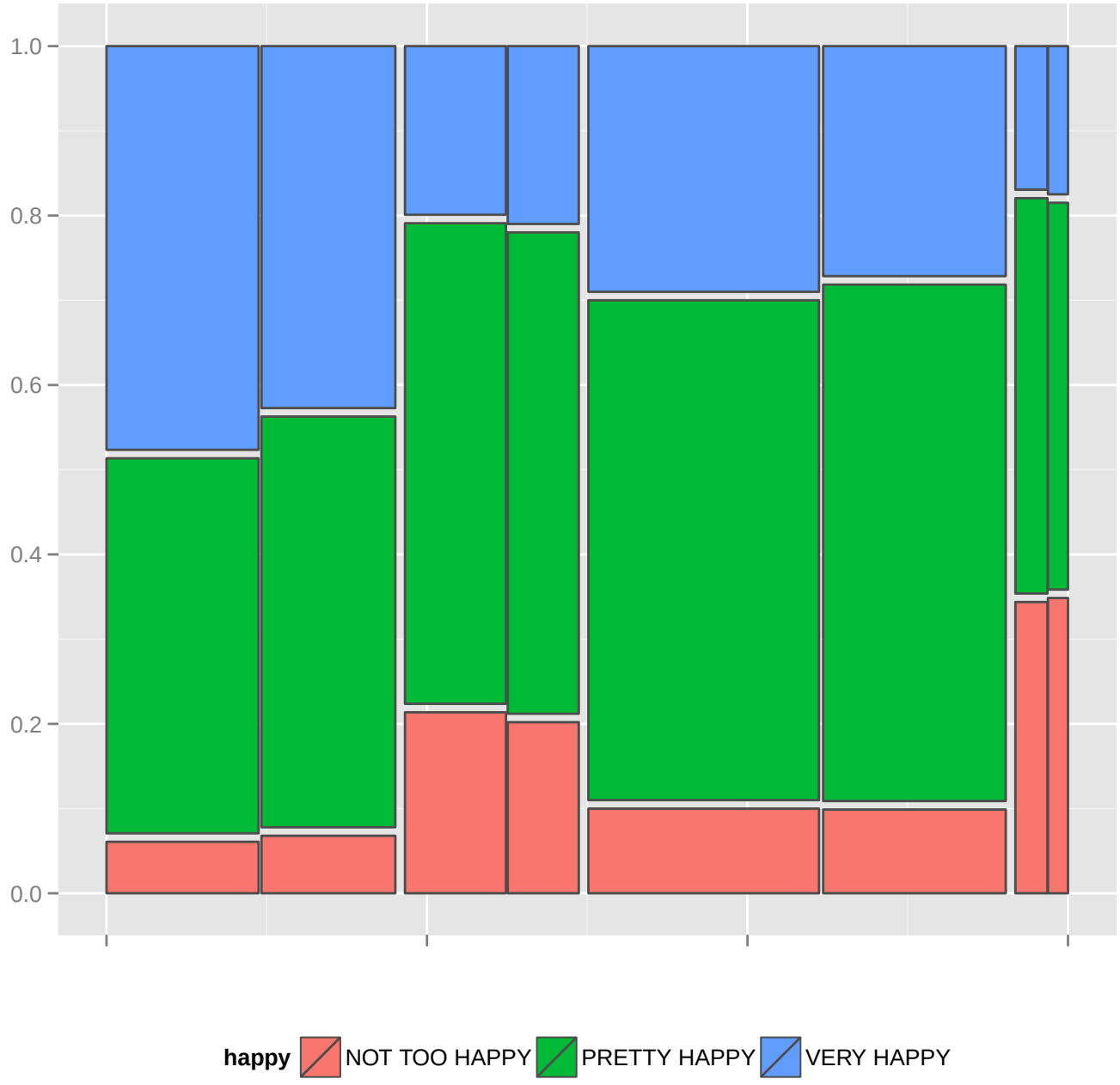
"whites" are generally the happiest race, "other" are second happiest and "blacks" are least happy.

```
prodplot(p3, Freq ~ happy + sex, c("vspine", "hspine")) + aes(fill = happy) +
opts(legend.position = "bottom",
    legend.direction = "horizontal")
```



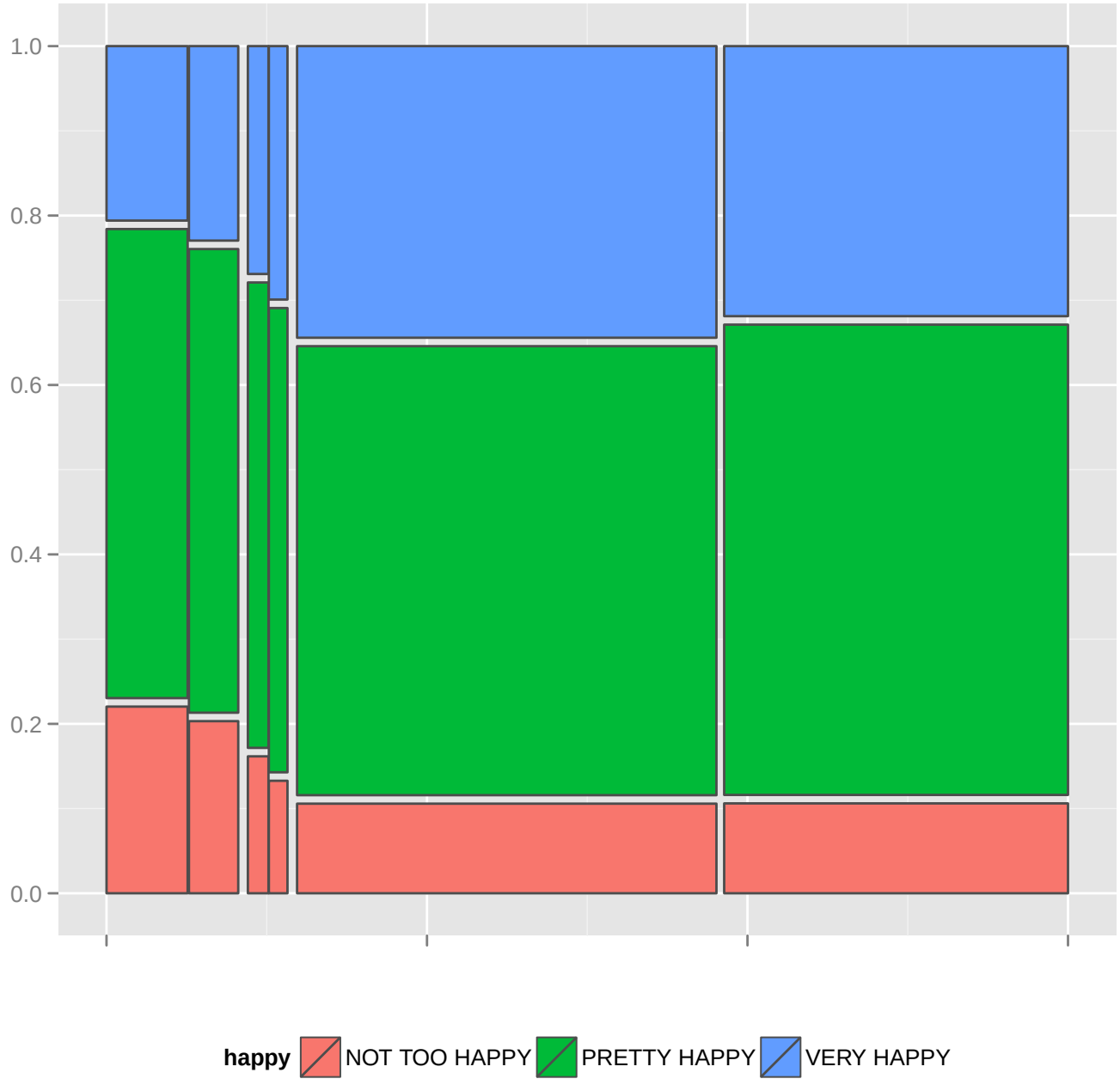
Men and women seem to be about equally happy. However, men are more likely to be on the extreme ends of the scale.

```
prodplot(p3, Freq ~ happy + sex + health, c("vspine", "hspine", "hspine")) + aes(fill =
happy) +
    opts(legend.position = "bottom", legend.direction = "horizontal")
```



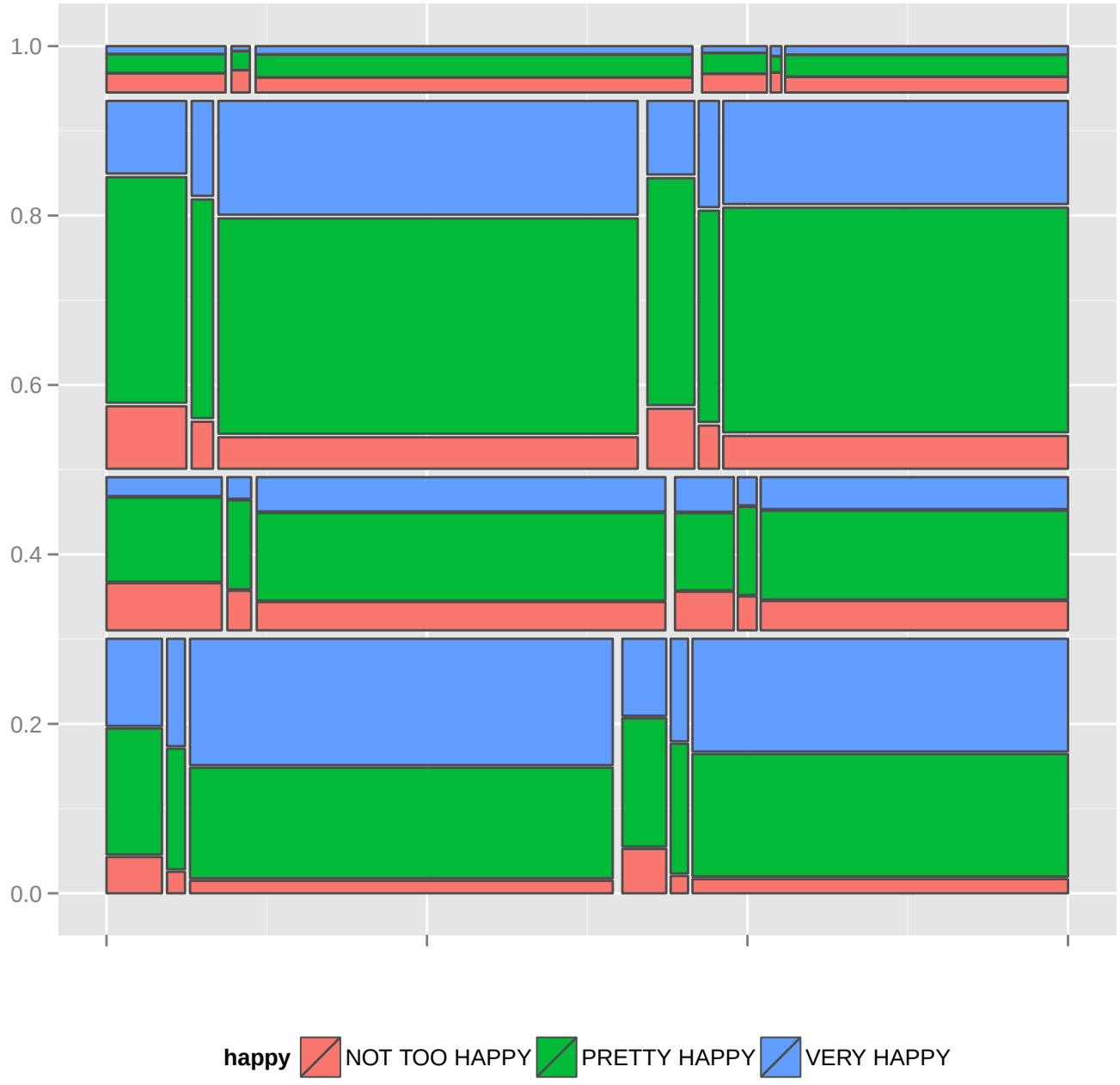
Again we see men and women equally happy at each level of health.

```
prodplot(p3, Freq ~ happy + sex + race, c("vspine", "hspine", "hspine")) + aes(fill =
happy) +
    opts(legend.position = "bottom", legend.direction = "horizontal")
```



Now we begin to see some differences between men and women within race. "Black" and "other" males are less happy than their female counterparts, while "white" males are slightly happier than their female counterparts.

```
prodplot(p3, Freq ~ happy + race + sex + health, c("vspine", "hspine", "hspine",
    "vspine")) + aes(fill = happy) + opts(legend.position = "bottom", legend.direction =
"horizontal")
```

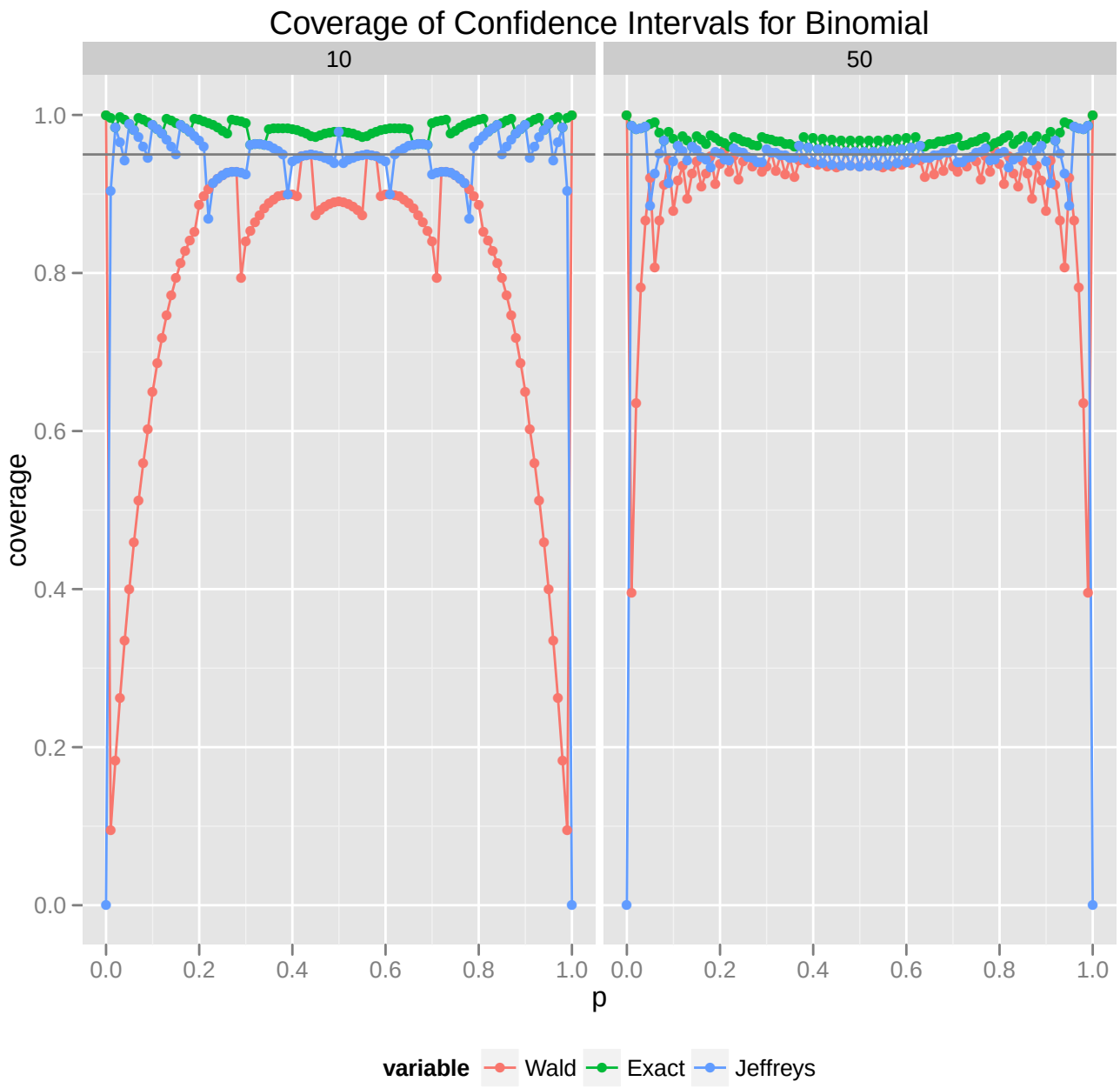


At nearly every level of health, "whites" are generally the happiest race, "other" are second happiest and "blacks" are least happy. However, those differences become smaller as health deteriorates.

Problem 3:

```
Wald <- function(y, n, alpha = 0.05, ...) {
    p <- y/n
    li <- qnorm(1 - alpha/2) * sqrt(p * (1 - p)/n)
    return(c(p - li, p + li))
```

```
}
exact <- function(y, n, alpha = 0.05, ...) {
    pL <- 1/(1 + (n - y + 1)/(y * qf(alpha/2, 2 * y, 2 * (n - y + 1))))
    pU <- 1/(1 + (n - y)/((y + 1) * qf(1 - alpha/2, 2 * (y + 1), 2 * (n - y))))
    if (is.nan(pL))
        pL <- 0
    if (is.nan(pU))
        pU <- 1
    return(c(pL, pU))
}
Jeffreys <- function(y, n, alpha = 0.05, ...) {
    pL <- qbeta(alpha/2, y + 1/2, n - y + 1/2)
    pU <- qbeta(1 - alpha/2, y + 1/2, n - y + 1/2)
    return(c(pL, pU))
}
coverage <- function(p, ci.method, n = 10, ...) {
    I <- ldply(0:n, function(x) {
        ci <- ci.method(y = x[1], n = n, ...)
        if ((p >= ci[1]) & (p <= ci[2]))
            return(1) else return(0)
    })
    bi <- dbinom(0:n, size = n, prob = p)
    return(sum(I * bi))
}
setup <- data.frame(expand.grid(p = seq(0, 1, by = 0.01), n = c(10, 50)))
res <- ddply(setup, .(n, p), summarise, Wald = coverage(p, Wald, n), Exact = coverage(p,
    exact, n), Jeffreys = coverage(p, Jeffreys, n))
res.melt <- melt(res, id.vars = c("n", "p"), variable_name = "Method")
qplot(p, value, colour = variable, data = res.melt, geom = c("line", "point"), group =
variable,
    main = "Coverage of Confidence Intervals for Binomial", facets = . ~ n, ylab =
"coverage") +
    geom_hline(yintercept = 0.95, color = "grey50") + opts(legend.position = "bottom",
legend.direction = "horizontal")
```



For $p \in (0, 0.2] \cup [0.8, 1)$, the Jeffreys interval is a huge improvement (in terms of coverage) compared to Wald and a slight improvement from the exact method. In fact, among the three methods it performs the best for nearly every possible value of p.