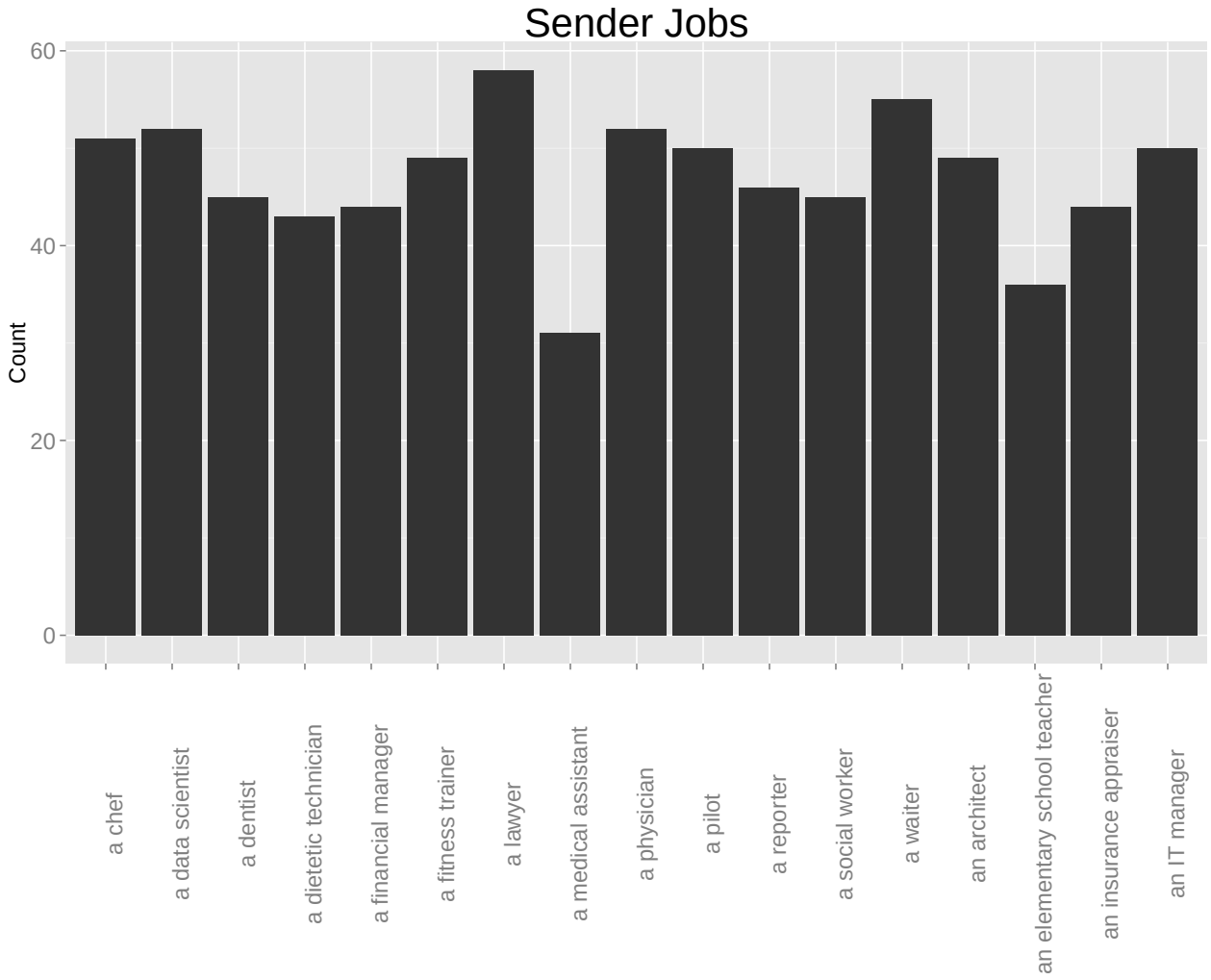
Housing Discrimination

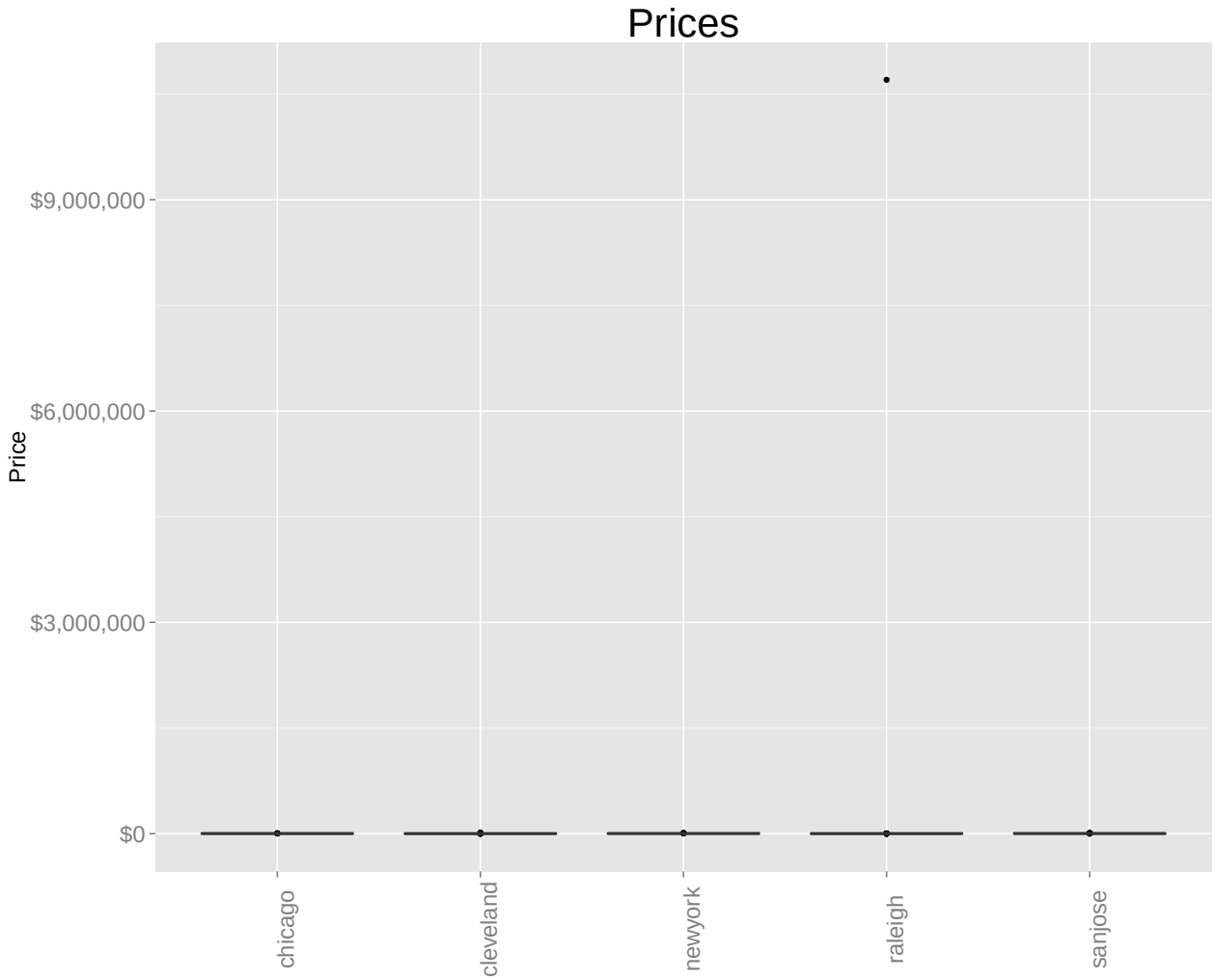
```
library(ggplot2)
library(scales)
library(reshape2)
library(dplyr)
setwd('/Users/nathanielblack/Downloads/')
cl.data <- read.csv('w241_final_project.csv',header = T)
#head(cl.data)

cast.price <- function(text)
  {
  text2 = substr(text, 2, 15)
  text2 = gsub(",", "", text2)
  return(as.numeric(text2))
  }

cl.data['price_clean'] <- lapply(cl.data['price'], FUN = cast.price)
summary(cl.data[c('sender_race','sender_social_status','sender_job','price_clean','outcome')])
```

```
##  sender_race sender_social_status            sender_job
##  black:400   high:400             a lawyer        : 58
##  white:400   low :400             a waiter        : 55
##                                   a data scientist: 52
##                                   a physician     : 52
##                                   a chef          : 51
##                                   a pilot         : 50
##                                   (Other)         :482
##   price_clean          outcome
##  Min.   :     249   Min.   :0.000
##  1st Qu.:     970   1st Qu.:0.000
##  Median :    1700   Median :0.000
##  Mean   :   15670   Mean   :0.385
##  3rd Qu.:    2450   3rd Qu.:1.000
##  Max.   :10701155   Max.   :1.000
##  NA's   :23
```


Sender Jobs
Count
60
40
20
0
a chef
a data scientist
a dentist
a dietetic technician
a financial manager
a fitness trainer
a lawyer
a medical assistant
a physician
a pilot
a reporter
a social worker
a waiter
an architect
an elementary school teacher
an insurance appraiser
an IT manager

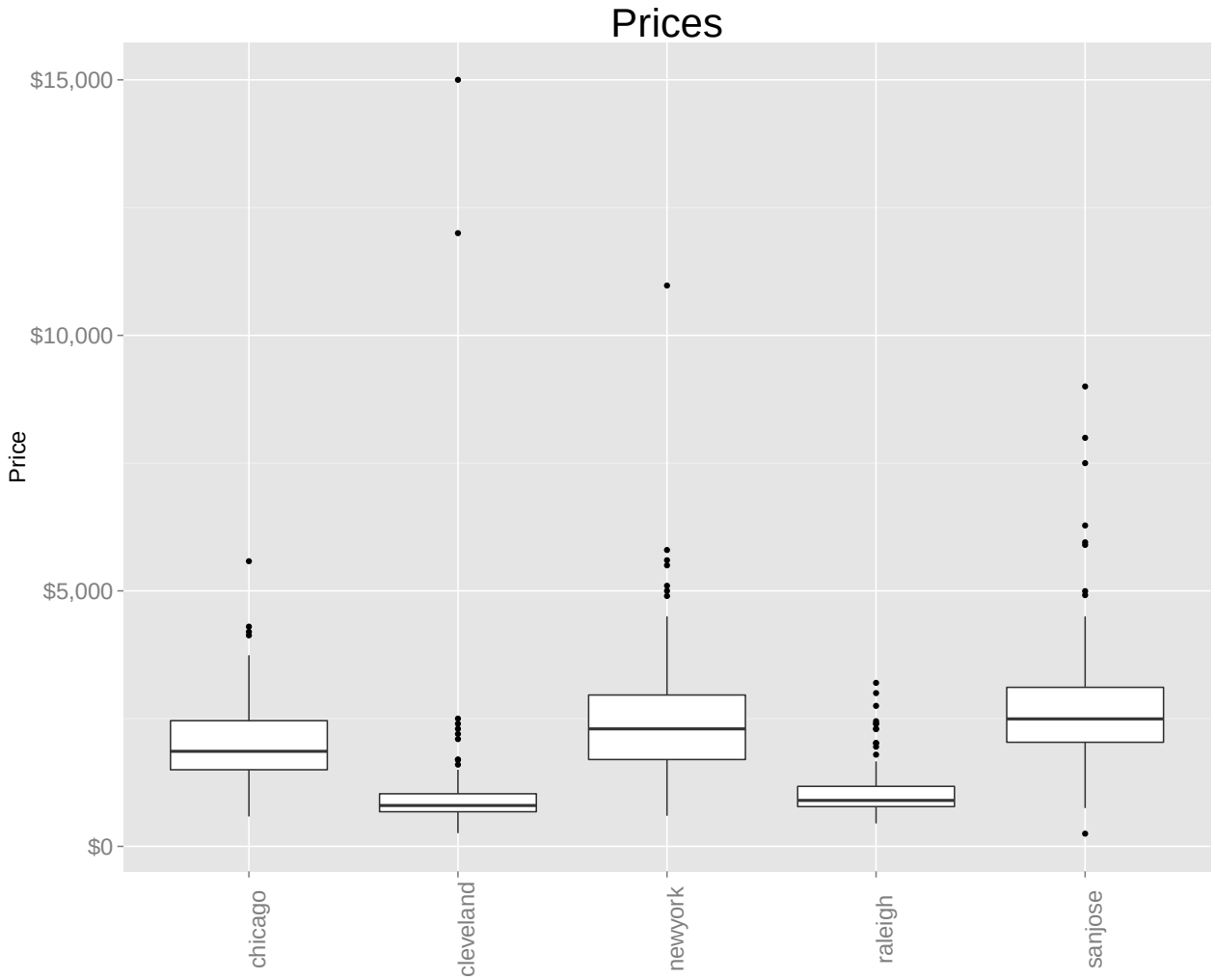


There is an outlier in Raleigh...

```
outlier <- filter(cl.data,cl.data['price_clean'] > 1000000)
outlier[c('id','price','city')]
```

```
##           id       price    city
## 1 4974054862 $10,701,155 raleigh
```

Re-plot without outlier



Response Rates

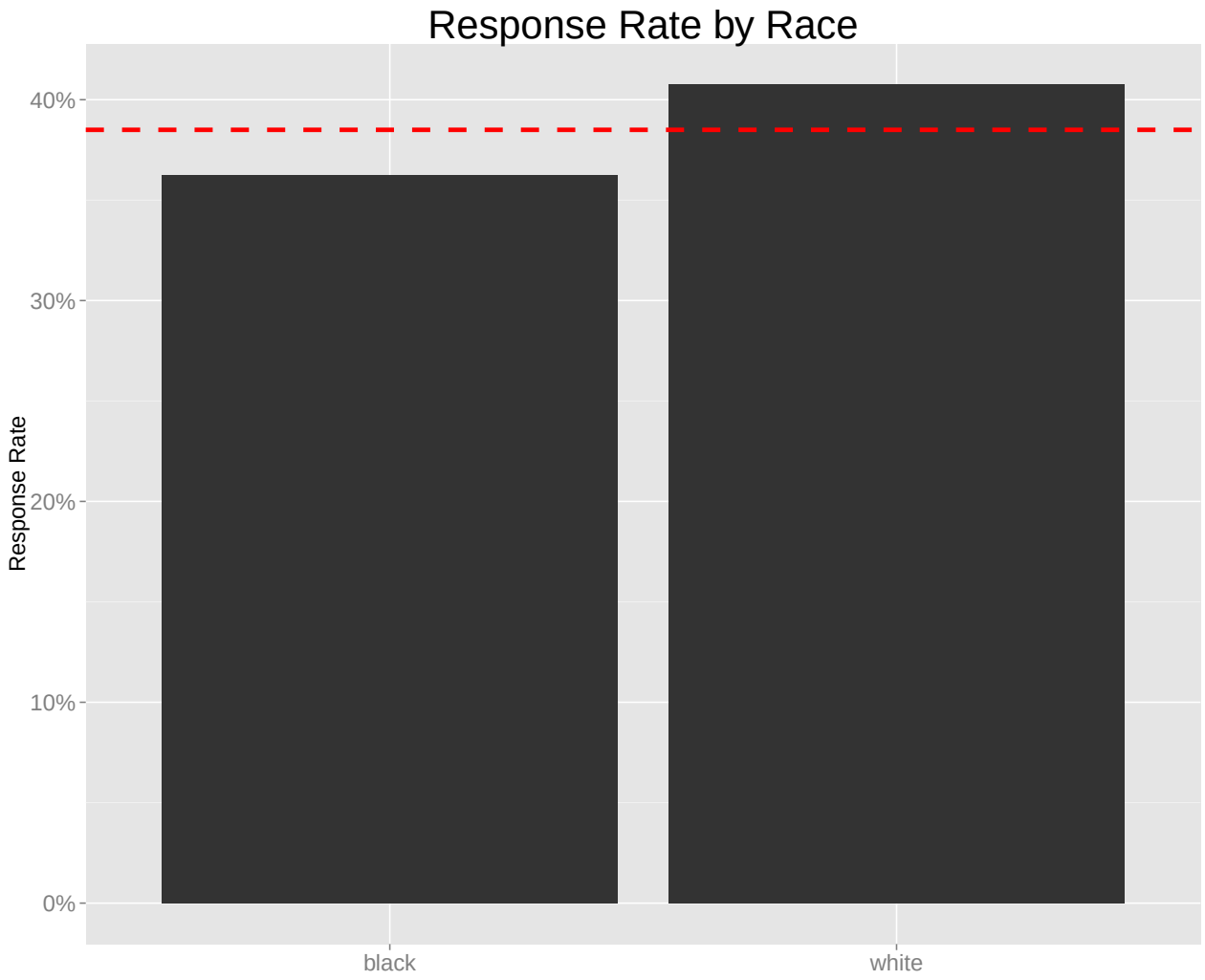
The overall response rate was 38.5%

```
overall.response <- mean(cl.data$outcome)
overall.response
```

```
## [1] 0.385
```

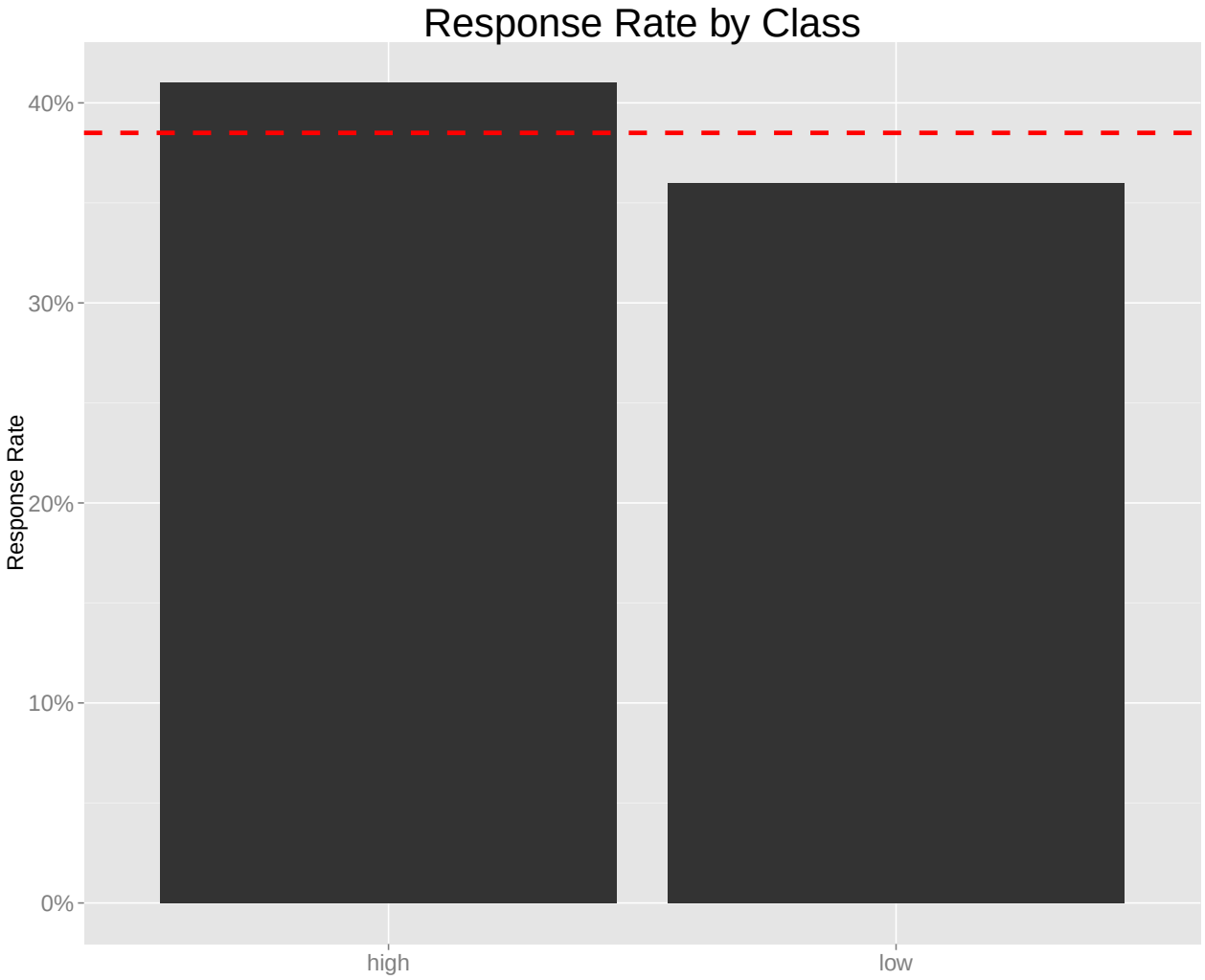
Race

```
## Source: local data frame [2 x 2]
##
##   sender_race response_rate
## 1       black        0.3625
## 2       white        0.4075
```



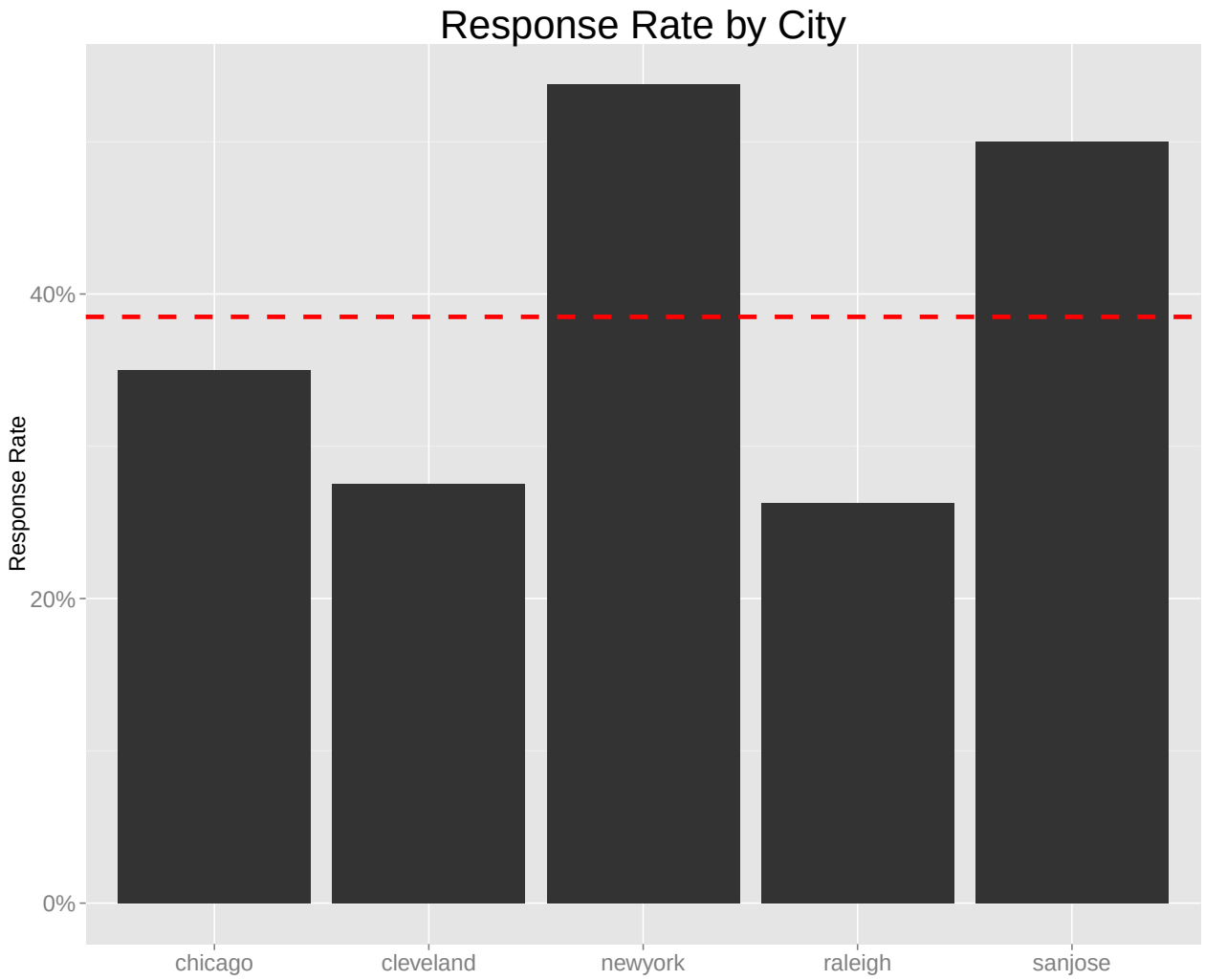
Class

```
## Source: local data frame [2 x 2]
##
##   sender_social_status response_rate
## 1                 high          0.41
## 2                  low          0.36
```



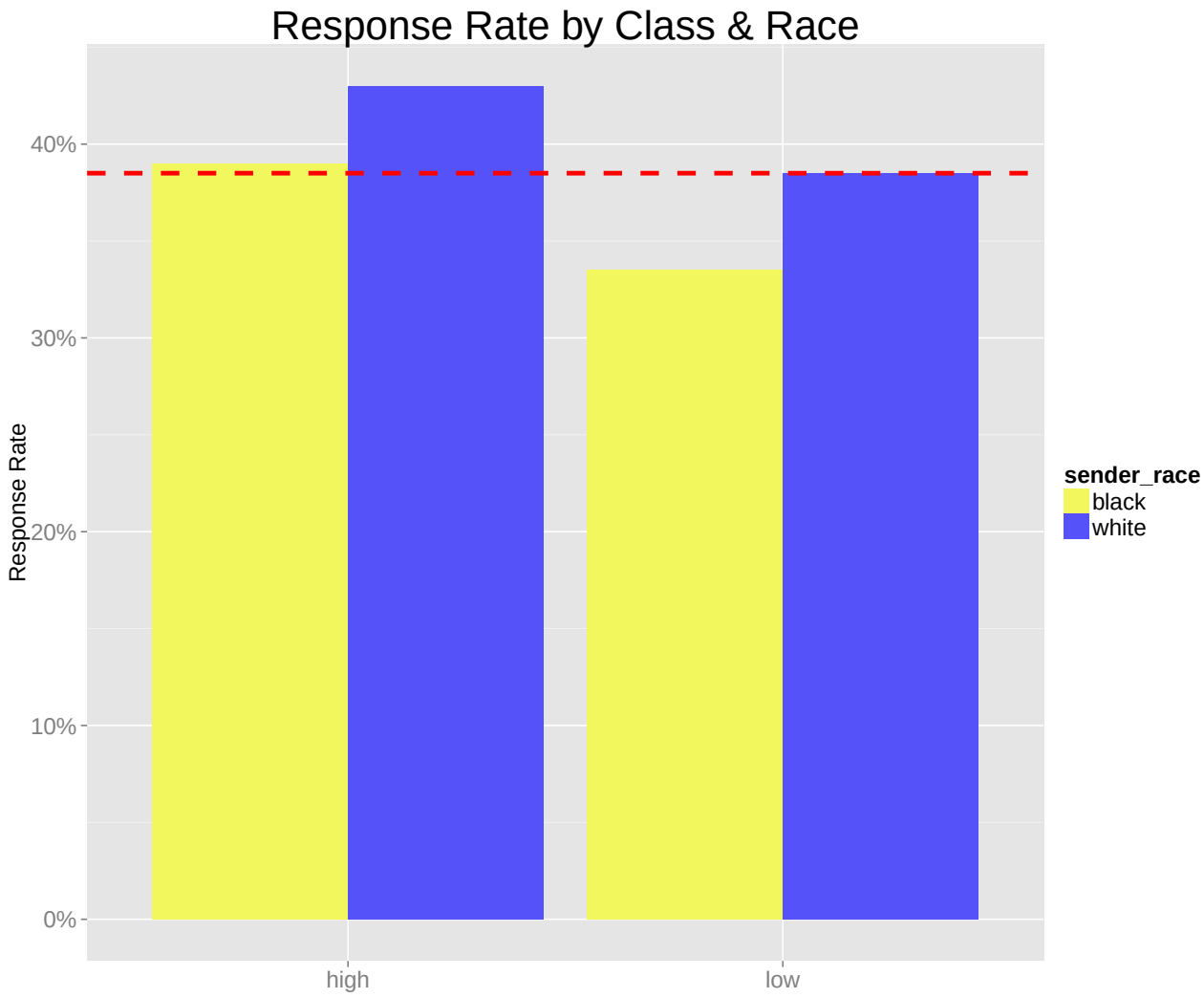
Response by City

```
## Source: local data frame [5 x 2]
##
##        city response_rate
## 1   chicago        0.3500
## 2 cleveland        0.2750
## 3   newyork        0.5375
## 4   raleigh        0.2625
## 5   sanjose        0.5000
```



Race and Class

```
## Source: local data frame [4 x 3]
## Groups: sender_social_status
##
##   sender_social_status sender_race response_rate
## 1                 high       black         0.390
## 2                 high       white         0.430
## 3                  low       black         0.335
## 4                  low       white         0.385
```



Race and Class by City

```
## Source: local data frame [20 x 4]
## Groups: sender_social_status, sender_race
##
##    sender_social_status sender_race      city response_rate
## 1                  high       black   chicago         0.375
## 2                  high       black cleveland         0.225
## 3                  high       black   newyork         0.500
## 4                  high       black   raleigh         0.275
## 5                  high       black   sanjose         0.575
## 6                  high       white   chicago         0.425
## 7                  high       white cleveland         0.325
## 8                  high       white   newyork         0.675
## 9                  high       white   raleigh         0.250
## 10                 high       white   sanjose         0.475
## 11                  low       black   chicago         0.250
## 12                  low       black cleveland         0.275
## 13                  low       black   newyork         0.350
## 14                  low       black   raleigh         0.275
```

```
## 15                 low       black   sanjose         0.525
## 16                 low       white   chicago         0.350
## 17                 low       white cleveland         0.275
## 18                 low       white   newyork         0.625
## 19                 low       white   raleigh         0.250
## 20                 low       white   sanjose         0.425
```

Response by Job

```
## Source: local data frame [17 x 3]
## Groups: sender_job
##
##                      sender_job sender_social_status response_rate
## 1                        a chef                  low     0.3333333
## 2             a data scientist                 high     0.4423077
## 3                     a dentist                 high     0.3111111
## 4         a dietetic technician                  low     0.3023256
## 5           a financial manager                 high     0.4545455
## 6             a fitness trainer                  low     0.4081633
```

```
## 7                       a lawyer                high      0.3620690
## 8            a medical assistant                 low      0.4516129
## 9                     a physician                high      0.4615385
## 10                        a pilot                high      0.5200000
## 11                     a reporter                 low      0.2608696
## 12                a social worker                 low      0.3555556
## 13                       a waiter                 low      0.2727273
## 14                   an architect                high      0.3673469
## 15 an elementary school teacher                 low      0.5000000
## 16        an insurance appraiser                 low      0.4318182
## 17                  an IT manager                high      0.3600000
```