Caring Analysis

Jonathan Luu

Exploratory analysis
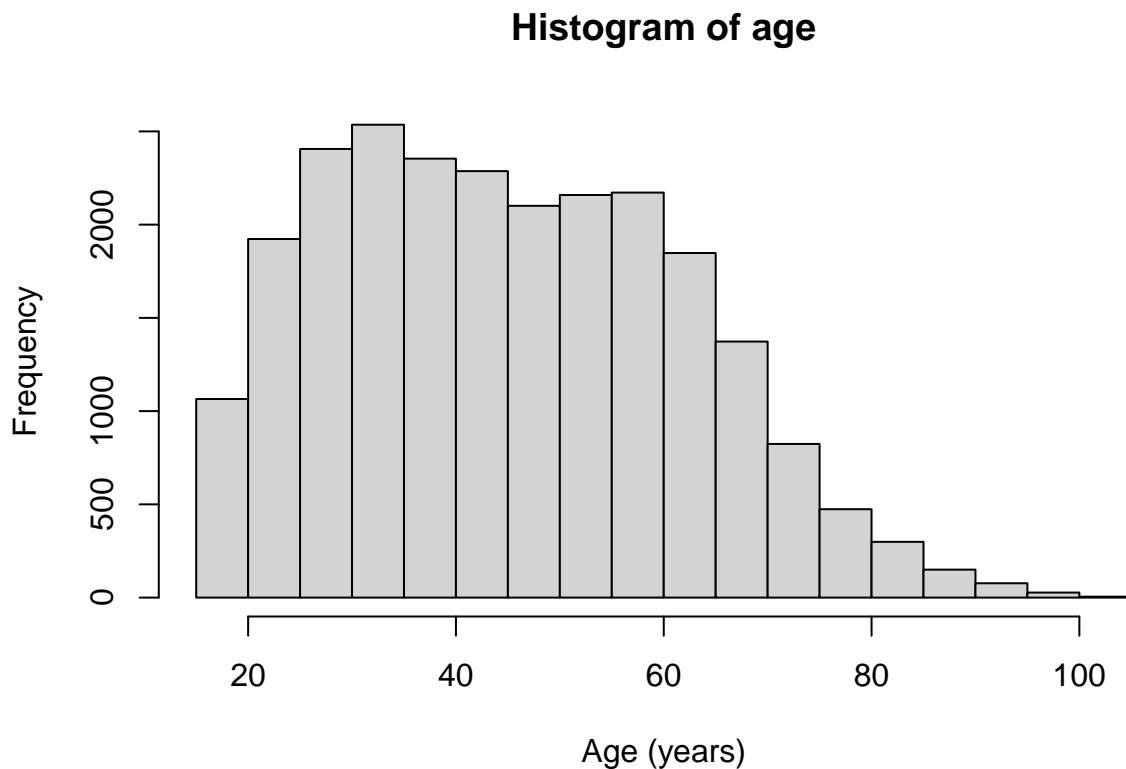
```
# Number of unique entries
dat.distinct %>% nrow()
```

```
[1] 24080
```

```
# Age distribution
dat.distinct$Age %>% summary()
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  18.00   32.00   44.00   45.64   58.00  105.00
```

```
dat.distinct$Age %>% hist(main="Histogram of age", xlab="Age (years)")
```

Histogram of age

```r
# Language
dat.distinct$Language %>% unique() %>% length() # 64 different languages spoken
```

```
[1] 64
```

```r
dat.distinct$Language %>% table() %>% sort(decreasing = T) %>%
  head(n=15) %>% as_tibble() %>% rename(Language=".") # Most popular languages
```

```
# A tibble: 15 x 2
   Language          n
   <chr>         <int>
 1 "English"     13693
 2 "Spanish"      5622
 3 "Russian"      1117
 4 ""              962
 5 "Nepali"        750
 6 "Vietnamese"    651
 7 "Arabic"        433
 8 "Swahili"       210
 9 "Somali"        123
10 "Burmese"        54
11 "Turkish"        37
12 "Portuguese"     36
13 "Other"          33
14 "Kinyarwanda"    31
15 "Dari"           27
```
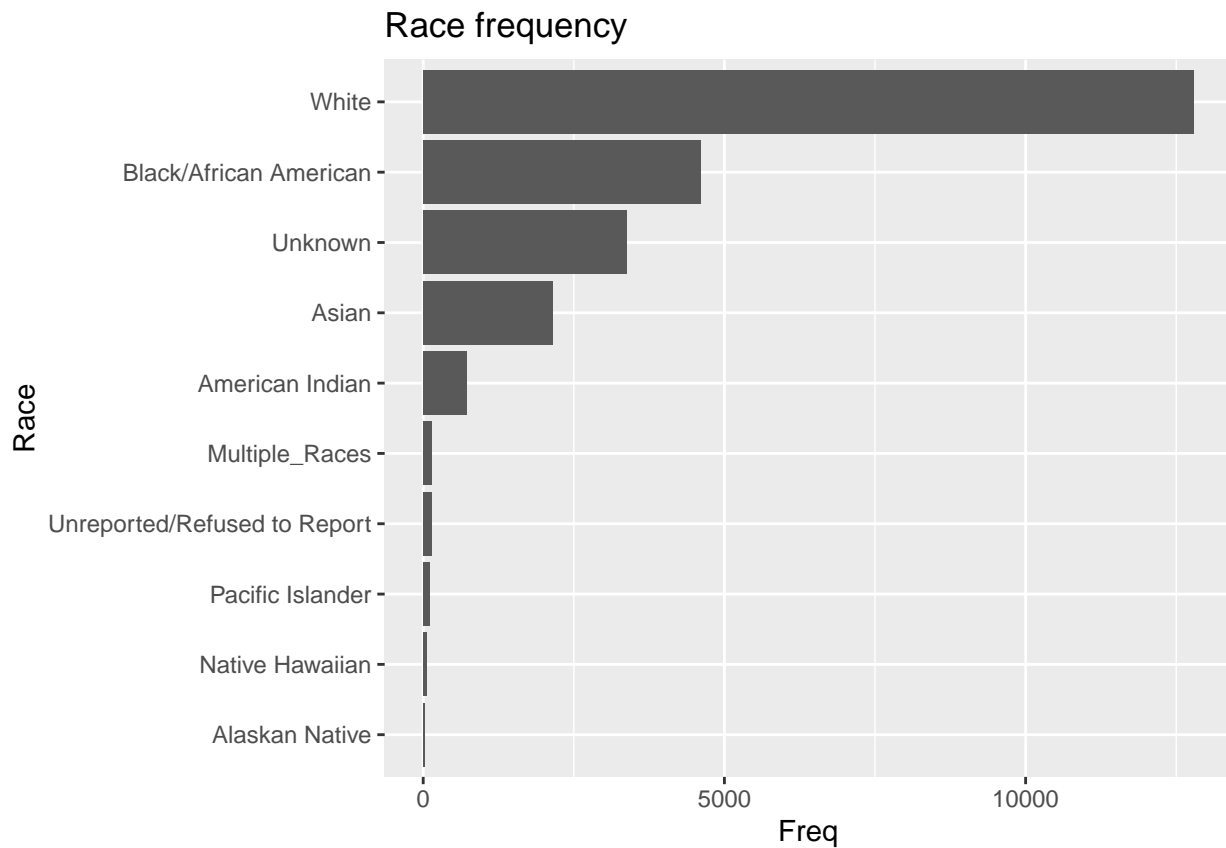
```
# Race
race.table <- dat.distinct$Race %>% table() %>% sort() %>%
  as.data.frame() %>% rename(Race=".")
race.table %>% as_tibble()
```

```
# A tibble: 10 x 2
   Race                     Freq
   <fct>                   <int>
 1 Alaskan Native             19
 2 Native Hawaiian            53
 3 Pacific Islander          104
 4 Unreported/Refused to Report  133
 5 Multiple_Races            139
 6 American Indian           718
 7 Asian                    2151
 8 Unknown                  3370
 9 Black/African American   4605
10 White                   12788
```
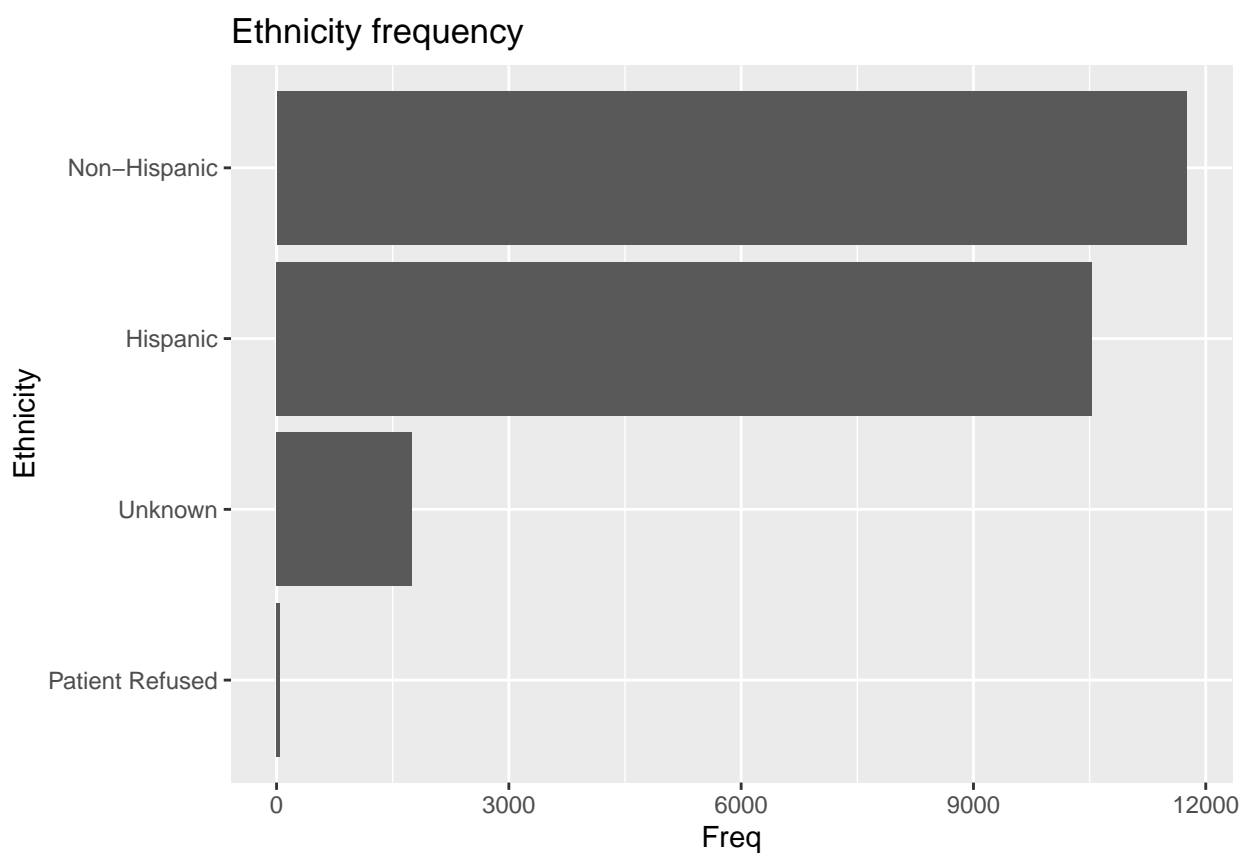
```
# Race barplot
race.table %>% ggplot(aes(x=Freq, y=Race)) + geom_bar(stat="identity") +
  theme(axis.text.y=element_text(angle=0,hjust=1,vjust=0.5)) +
  ggtitle("Race frequency")
```


Race frequency

```
# Ethnicity
eth.table <- dat.distinct$Ethnicity %>% table() %>% sort() %>%
  as.data.frame() %>% rename(Ethnicity=".")
eth.table %>% as_tibble()
```

```
# A tibble: 4 x 2
  Ethnicity        Freq
  <fct>           <int>
1 Patient Refused    45
2 Unknown          1754
3 Hispanic        10529
4 Non-Hispanic    11752
```

```
# Ethnicity barplot
eth.table %>% ggplot(aes(x=Freq, y=Ethnicity)) + geom_bar(stat="identity") +
  theme(axis.text.y=element_text(angle=0,hjust=1,vjust=0.5)) +
  ggtitle("Ethnicity frequency")
```

```r
# Number of visits
dat.distinct$Num_Visits %>% summary()
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000   2.000   5.000   9.819  12.000 268.000
```

```r
# Most common number of visits
dat.distinct$Num_Visits %>% table() %>% sort(decreasing=T) %>%
  head(n=10) %>% as_tibble() %>% rename("Num Visit"=".")
```

```
# A tibble: 10 x 2
   `Num Visit`     n
   <chr>       <int>
 1 2            4604
 2 1            3430
 3 3            2202
 4 4            1539
 5 5            1298
 6 6            1129
 7 7             923
 8 8             852
 9 9             697
10 10            658
```
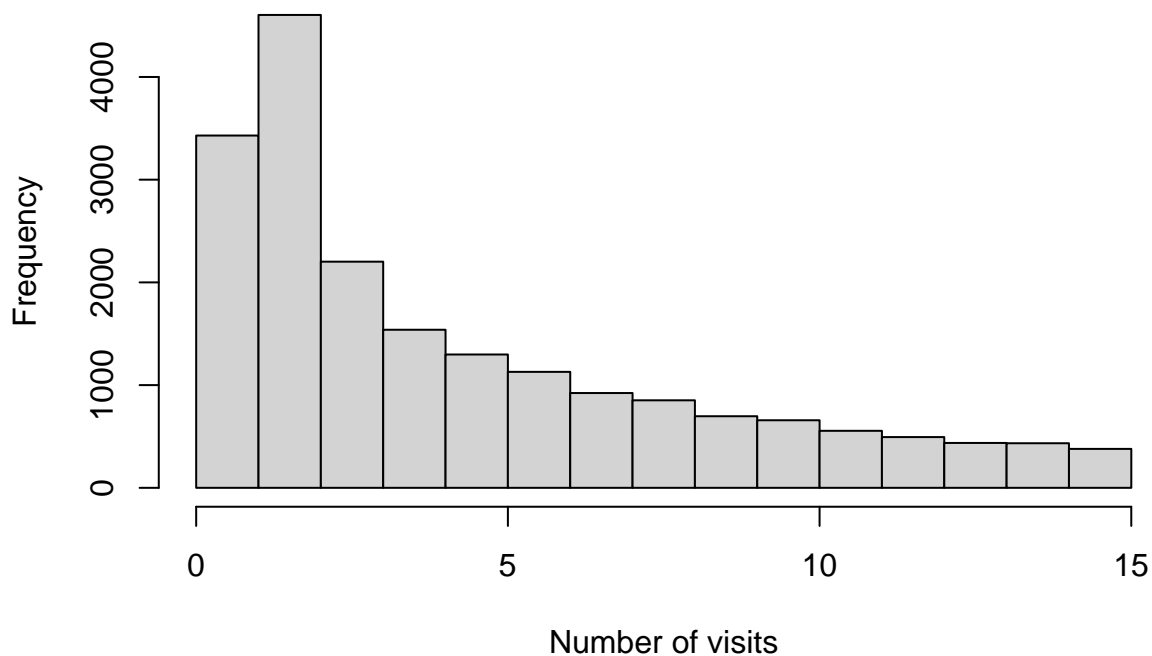
```r
dat.distinct %>% filter(Num_Visits <= 15) %>% pull(Num_Visits) %>%
  hist(breaks=0:15, main="Histogram of number of visits", xlab="Number of visits")
```

Histogram of number of visits

Survey question descriptives

```
# Response types
dat$MEAS_DISP_NAME %>% unique() %>% length() # 45 different response types
```

```
[1] 45
```

```
options(pillar.print_max = 50, pillar.print_min = 50)
dat$MEAS_DISP_NAME %>% table() %>% sort(decreasing=T) %>%
  as_tibble() %>% rename(Question=".")
```

```
# A tibble: 45 x 2
   Question                                                               n
   <chr>                                                              <int>
 1 ""                                                                 21611
 2 "How often does anyone, including family and friends, threaten you wit~  2570
 3 "How often does anyone, including family and friends, physically hurt ~  2568
 4 "How often does anyone, including family and friends, scream or curse ~  2559
 5 "How often does anyone, including family and friends, insult or talk d~  2555
 6 "Relationship Safety Total Score: >11 is abnormal"                  2550
 7 "Would you like assistance with any of the above items?"            1740
 8 "How do you learn best?"                                            1496
 9 "In the past 12 months, has lack of transportation kept you from medic~  1241
10 "Number of positive responses to housing questions"                 1164
11 "What is your living situation today?"                              1157
12 "How hard is it for you to pay for the very basics like food, housing,~  1120
13 "Number of positive responses to food security questions"           1091
14 "Do you feel these kinds of stress these days?"                     1090
15 "Within the past 12 months, you worried that your food would run out b~  1089
16 "What do you want help with?"                                       1076
17 "Within the past 12 months, the food you bought just didn't last and y~  1075
18 "What is the highest grade or year of school you completed?"        1052
19 "How often do you see or talk to people that you care about and feel c~   998
20 "Are you currently employed?"                                        960
21 "How often do you feel lonely or isolated from those around you?"     938
22 "Do you have someone you could call if you needed help?"             916
23 "Think about the place you live. Do you have problems with any of the ~   899
24 "In the past 12 months has the electric, gas, oil, or water company th~   884
25 "On average, how many minutes do you engage in exercise at this level?"  760
26 "On average, how many days per week do you engage in moderate to stren~   753
27 "Weekly Physical Activity"                                           746
28 "Would you like help finding a job?"                                 628
29 "Are you seeking work?"                                              625
30 "Would you like information about language classes or other educationa~   621
31 "Would you like to be connected to job training resources?"          616
32 "Hard to pay for: Utilities"                                         562
33 "Hard to pay for: Food"                                              556
34 "Hard to pay for: Rent/Mortgage payment"                            550
35 "Hard to pay for: Transportation"                                    549
36 "Hard to pay for: Phone"                                             548
37 "Hard to pay for: Clothing"                                          546
38 "Hard to pay for: Medicine or medical care"                         543
```

```
39 "Hard to pay for: Child care"                                      542
40 "Hard to pay for: Health insurance"                                534
41 "Hard to pay for: Other"                                           495
42 "Type of assistance"                                               408
43 "Are you having any problems with your boss?"                      257
44 "What is your current work situation?"                             249
45 "Hard to pay for: Specify Other"                                    46
```

```
# Response categories
dat$DOMAIN_GRP %>% table() %>% sort(decreasing=T) %>%
  as_tibble() %>% rename(Response=".") # 13 categories of responses
```

```
# A tibble: 13 x 2
   Response                        n
   <chr>                        <int>
 1 ""                           21611
 2 "Relationship Safety"        12802
 3 "Help Desired"                5089
 4 "Financial/Other"             4924
 5 "Stress and Social Isolation" 3942
 6 "Food Insecurity"             3811
 7 "Housing"                     3220
 8 "Physical Activity"           2259
 9 "Employment"                  2091
10 "Transportation"              1790
11 "Health Literacy"             1496
12 "Utilities"                   1446
13 "Education"                   1052
```

```
# Individuals with the most # of responses
dat$PAT_ID %>% table() %>% sort(decreasing = T) %>% head(n=10) %>%
  as_tibble() %>% rename(ID=".")
```

```
# A tibble: 10 x 2
   ID            n
   <chr>     <int>
 1 Z3347793    135
 2 Z8131701    105
 3 Z5873535    100
 4 Z8215632     97
 5 Z7376860     93
 6 Z2817799     92
 7 Z3362887     92
 8 Z7227493     90
 9 Z6210794     88
10 Z3906641     87
```

```
# Filter by response category
dat.resp <- dat %>% select(PAT_ID, FLO_MEAS_ID, DOMAIN_GRP) %>%
  count(PAT_ID, DOMAIN_GRP) %>% mutate_all(na_if,"") %>%
  pivot_wider(names_from = DOMAIN_GRP, values_from = n)
dat.resp %>% select(-PAT_ID) %>% summary()
```

```
         NA         Food Insecurity   Help Desired    Relationship Safety
 Min.   :1        Min.   : 1.000   Min.   : 1.000   Min.   : 1.000
 1st Qu.:1        1st Qu.: 3.000   1st Qu.: 1.000   1st Qu.: 5.000
 Median :1        Median : 3.000   Median : 2.000   Median : 5.000
 Mean   :1        Mean   : 3.477   Mean   : 3.223   Mean   : 5.772
 3rd Qu.:1        3rd Qu.: 4.000   3rd Qu.: 4.000   3rd Qu.: 5.000
 Max.   :1        Max.   :11.000   Max.   :24.000   Max.   :20.000
 NA's   :2469     NA's   :22984    NA's   :22501    NA's   :21862
   Education        Employment      Financial/Other  Health Literacy
 Min.   :1.000    Min.   :1.000    Min.   : 1.000   Min.   : 1.000
 1st Qu.:1.000    1st Qu.:2.000    1st Qu.: 1.000   1st Qu.: 1.000
 Median :1.000    Median :2.000    Median : 5.000   Median : 1.000
 Mean   :1.087    Mean   :2.352    Mean   : 4.757   Mean   : 1.943
 3rd Qu.:1.000    3rd Qu.:3.000    3rd Qu.: 8.000   3rd Qu.: 3.000
 Max.   :4.000    Max.   :8.000    Max.   :29.000   Max.   :14.000
 NA's   :23112    NA's   :23191    NA's   :23045    NA's   :23310
    Housing      Physical Activity Stress and Social Isolation Transportation
 Min.   : 1.000  Min.   :1.000    Min.   : 1.00             Min.   :1.000
 1st Qu.: 2.000  1st Qu.:3.000    1st Qu.: 3.00             1st Qu.:1.000
 Median : 3.000  Median :3.000    Median : 4.00             Median :2.000
 Mean   : 2.995  Mean   :3.112    Mean   : 3.67             Mean   :1.736
 3rd Qu.: 3.000  3rd Qu.:3.000    3rd Qu.: 4.00             3rd Qu.:2.000
 Max.   :11.000  Max.   :9.000    Max.   :12.00             Max.   :9.000
 NA's   :23005   NA's   :23354    NA's   :23006             NA's   :23049
   Utilities
 Min.   :1.000
 1st Qu.:1.000
 Median :1.000
 Mean   :1.552
 3rd Qu.:2.000
 Max.   :5.000
 NA's   :23148
```

```r
# Ranges of screening
dat.resp[is.na(dat.resp)] <- 0
dat.resp <- dat.resp %>% select(-"NA") %>% mutate(sum = rowSums(across(where(is.numeric))))

# Number of questions answered
dat.resp %>% filter(sum==0) %>% nrow() #21611 people did not have a survey answer
```

```
[1] 21611
```

```r
dat.resp %>% filter(sum > 0) %>% select(sum) %>% summary() # 2469 individuals who did answer
```

```
      sum
 Min.   :  1.00
 1st Qu.:  5.00
 Median :  8.00
 Mean   : 17.79
 3rd Qu.: 30.00
 Max.   :135.00
```

Questions

Which are the narrowed-down 7 questions?