EDA

Loading the required libraries

```
library(survey)
```

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

```
## Loading required package: Matrix
```

```
## Loading required package: survival
```

```
##
## Attaching package: 'survey'
```

```
## The following object is masked from 'package:graphics':
##
##     dotchart
```

```
library(sampling)
```

```
##
## Attaching package: 'sampling'
```

```
## The following objects are masked from 'package:survival':
##
##     cluster, strata
```

```
library('ISLR')
library('ggplot2')
library(MASS)
library(tree)

library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following object is masked from 'package:MASS':
##
##     select
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

Loading the dataset

```
# load the dataset
yyc_survey <- read.csv("Citizen_Satisfaction_Survey.csv")

#column names of the dataset
names(yyc_survey)
```

```
##   [1] "Mweight0" "qwave"    "s4qt"     "market2"  "q39"      "q34"
##   [7] "q37"      "q38"      "q30"      "q32x"     "q40"      "sexfix"
##  [13] "q29x"     "q2a"      "q3"       "q24bx_1"  "q24bx_2"  "q24bx_3"
##  [19] "q24bx_4"  "q24bx_5"  "q24bx_6"  "q24bx_7"  "q24cx"    "q10"
##  [25] "q19_1"    "q19_2"    "q19_3"    "q19_4"    "q19_5"    "q19_6"
##  [31] "q19_7"    "q19_8"    "q11a"     "q12"      "q8_1"     "q8_2"
##  [37] "q8_3"     "q8_4"     "q8_5"     "q8_6"     "q8_7"     "q8_8"
##  [43] "q8_9"     "q8_10"    "q8_11"    "q8_12"    "q8_13"    "q8_14"
##  [49] "q8_15"    "q8_16"    "q8_17"    "q8_18"    "q8_19"    "q8_20"
##  [55] "q8_21"    "q8_22"    "q8_23"    "q8_24"    "q8_25"    "q8_26"
##  [61] "q8_27"    "q8_28"    "q8_29"    "q8_30"    "q8_31"    "q8_32"
##  [67] "q8_33"    "q8_34"    "q8_35"    "q9_1_1"   "q9_1_2"   "q9_1_3"
##  [73] "q9_1_4"   "q9_1_5"   "q9_1_6"   "q9_1_7"   "q9_1_8"   "q9_1_9"
##  [79] "q9_1_10"  "q9_1_11"  "q9_1_12"  "q9_1_13"  "q9_1_14"  "q9_1_15"
##  [85] "q9_1_16"  "q9_1_17"  "q9_1_18"  "q9_1_19"  "q9_1_20"  "q9_1_21"
##  [91] "q9_1_22"  "q9_1_23"  "q9_1_24"  "q9_1_25"  "q9_1_26"  "q9_1_27"
##  [97] "q9_1_28"  "q9_1_29"  "q9_1_30"  "q9_1_31"  "q9_1_32"  "q9_1_33"
## [103] "q9_1_34"  "q9_1_35"  "q9_2_1"   "q9_2_2"   "q9_2_3"   "q9_2_4"
## [109] "q9_2_5"   "q9_2_6"   "q9_2_7"   "q9_2_8"   "q9_2_9"   "q9_2_10"
## [115] "q9_2_11"  "q9_2_12"  "q9_2_13"  "q9_2_14"  "q9_2_15"  "q9_2_16"
## [121] "q9_2_17"  "q9_2_18"  "q9_2_19"  "q9_2_20"  "q9_2_21"  "q9_2_22"
## [127] "q9_2_23"  "q9_2_24"  "q9_2_25"  "q9_2_26"  "q9_2_27"  "q9_2_28"
## [133] "q9_2_29"  "q9_2_30"  "q9_2_31"  "q9_2_32"  "q9_2_33"  "q9_2_34"
## [139] "q9_2_35"
```

```
#dim of the dataset
dim(yyc_survey)
```

```
## [1] 10002   139
```

Inspecting the dataset

```
head(yyc_survey)
```

```
##   Mweight0     qwave s4qt market2 q39 q34 q37 q38 q30 q32x q40 sexfix q29x
q2a
## 1     0.51 Year-2021    4      10   4   1   9   3   6    2   2      2    1
6
## 2     0.60 Year-2021    1       8   1   1   9   2   6    2   2      1    2
6
## 3     0.70 Year-2021    2      12   7   1   8   2   6    2   2      2    1
8
## 4     0.49 Year-2021    3       6   9   1   8   3   6    2   3      2    1
```

```
7
## 5     0.77 Year-2021    1      11   4   1  10   2   5    2   2      2    1
8
## 6     0.67 Year-2021    1       6   1   2   1   3   6    2   1      1    2
7
##   q3 q24bx_1 q24bx_2 q24bx_3 q24bx_4 q24bx_5 q24bx_6 q24bx_7 q24cx q10
q19_1
## 1  2       5       5       7       6       1       4       3     3   2
NA
## 2  2       6       7       7       7       6       6       6     4   2
NA
## 3  3       5       3       5       5       1       5       8     3   4
NA
## 4  3       6       1      10      10       1       4       9     3   1
NA
## 5  1       7       3       7       8       7       5       8     3   4
NA
## 6  2       6       7       8       8       6       6       7     2   2
NA
##   q19_2 q19_3 q19_4 q19_5 q19_6 q19_7 q19_8 q11a q12 q8_1 q8_2 q8_3 q8_4
q8_5
## 1     1     4     1     1    NA     1     1    5   4    5    5    5    5
5
## 2     3     4     4     4    NA     4     4    7   2    5    5    5    5
5
## 3     2     3     5     3    NA     3     3    7   2    4    4    4    4
3
## 4     5     5     1     3    NA     3     2    8   4    5    5    5    5
5
## 5     4     4     4     2    NA     1     1    8   2    5    5    5    5
5
## 6     1     3     1     3    NA     5     3    6   5    5    5    5    5
5
##   q8_6 q8_7 q8_8 q8_9 q8_10 q8_11 q8_12 q8_13 q8_14 q8_15 q8_16 q8_17
q8_18
## 1    5    5    5    5     5     5     5     5     5     5     5     5     5
5
## 2    5    5    5    5     5     5     5     5     5     5     5     5     5
5
## 3    2    4    2    4     4     2     4     4     4     4     4     3     3
2
## 4    5    5    5    5     5     5     5     5     5     5     5     5     5
5
## 5    5    5    5    5     5     5     5     5     5     5     5     5     5
5
## 6    5    5    5    5     5     5     5     5     5     5     5     5     5
5
##   q8_19 q8_20 q8_21 q8_22 q8_23 q8_24 q8_25 q8_26 q8_27 q8_28 q8_29 q8_30
q8_31
## 1     2     1     4     3     4     4     1     1     1     3     4     4
```

```
3
## 2     4     4     4     4     3     3     2     4     3     4     4     4
4
## 3     5     5     5     5     5     5     5     5     5     5     5     5
5
## 4     4     4     4     4     4     4     4     3     4     5     4     5
4
## 5     4     2     4     4     4     3     4     4     4     4     4     3
4
## 6     4     4     4     4     4     4     4     4     4     4     4     4
4
##   q8_32 q8_33 q8_34 q8_35 q9_1_1 q9_1_2 q9_1_3 q9_1_4 q9_1_5 q9_1_6 q9_1_7
## 1     1     4     2     4      5      5      5      5      5      5      5
## 2     4     4     1     4      5      5      5      5      5      5      5
## 3     5     5     5     5      4      4      3      3      3      3      4
## 4     4     4     4     4      5      5      5      5      5      5      5
## 5     4     3     3     2      5      5      5      5      5      5      5
## 6     4     4     3     3      5      5      5      5      5      5      5
##   q9_1_8 q9_1_9 q9_1_10 q9_1_11 q9_1_12 q9_1_13 q9_1_14 q9_1_15 q9_1_16
q9_1_17
## 1      5      5       5       5       5       5       5       5       5
5
## 2      5      5       5       5       5       5       5       5       5
5
## 3      3      4       3       3       3       3       3       3       3
3
## 4      5      5       5       5       5       5       5       5       5
5
## 5      5      5       5       5       5       5       5       5       5
5
## 6      5      5       5       5       5       5       5       5       5
5
##   q9_1_18 q9_1_19 q9_1_20 q9_1_21 q9_1_22 q9_1_23 q9_1_24 q9_1_25 q9_1_26
## 1       5       2       2       1       3       1       3       1       1
## 2       5       3       4       4       3       3       3       4       4
## 3       3       5       5       5       5       5       5       5       5
## 4       5       2       4       3       3       4       4       3       3
## 5       5       4       3       3       2       4       3       3       2
## 6       5       3       3       2       2       3       3       3       2
##   q9_1_27 q9_1_28 q9_1_29 q9_1_30 q9_1_31 q9_1_32 q9_1_33 q9_1_34 q9_1_35
## 1       1       3       2       5       1       1       1       1       2
## 2       3       3       3       3       3       3       3       5       4
## 3       5       5       5       5       5       5       5       5       5
## 4       4       5       3       5       2       3       1       1       1
## 5       3       2       3       3       2       3       2       3       3
## 6       3       2       3       3       2       5       2       3       5
##   q9_2_1 q9_2_2 q9_2_3 q9_2_4 q9_2_5 q9_2_6 q9_2_7 q9_2_8 q9_2_9 q9_2_10
## 1      4      4      4      4      4      4      4      4      4       4
## 2      4      4      4      4      4      4      4      4      4       4
## 3      1      1      3      1      3      3      3      3      3       1
```

```
## 4      4      4      4      4      4      4      4      4      4       4
## 5      4      4      4      4      4      4      4      4      4       4
## 6      4      4      4      4      4      4      4      4      4       4
##   q9_2_11 q9_2_12 q9_2_13 q9_2_14 q9_2_15 q9_2_16 q9_2_17 q9_2_18 q9_2_19
## 1       4       4       4       4       4       4       4       4       2
## 2       4       4       4       4       4       4       4       4       1
## 3       1       3       3       1       1       1       2       3       4
## 4       4       4       4       4       4       4       4       4       1
## 5       4       4       4       4       4       4       4       4       3
## 6       4       4       4       4       4       4       4       4       3
##   q9_2_20 q9_2_21 q9_2_22 q9_2_23 q9_2_24 q9_2_25 q9_2_26 q9_2_27 q9_2_28
## 1       2       1       3       1       3       2       2       2       2
## 2       1       1       1       1       1       1       1       1       1
## 3       4       4       4       4       4       4       4       4       4
## 4       3       1       1       3       3       1       2       3       4
## 5       3       3       1       3       1       1       1       1       1
## 6       1       3       1       1       1       1       1       1       3
##   q9_2_29 q9_2_30 q9_2_31 q9_2_32 q9_2_33 q9_2_34 q9_2_35
## 1       2       3       2       2       2       2       1
## 2       1       1       1       1       1       1       1
## 3       4       4       4       4       4       4       4
## 4       3       4       3       1       1       3       1
## 5       3       3       1       1       2       1       2
## 6       1       1       1       3       3       3       3
```

```r
table(yyc_survey$qwave)
```

```
##
## Year-2018 Year-2019 Year-2020 Year-2021
##      2500      2502      2500      2500
```

For our current study we only want to use the 2021 survey data:

```r
library(dplyr)


# filter to include only the survey responses from year 2021
filtered_df <- yyc_survey %>% filter(qwave == 'Year-2021')

dim(filtered_df)
```

```
## [1] 2500  139
```

```r
unique(filtered_df$qwave)
```

```
## [1] "Year-2021"
```

Select the columns to include only demographic features and response variable

```
filtered_df <- filtered_df %>% select(s4qt, q39, q34, q37, q38, q32x, q40,
q29x, q30, q2a)

dim(filtered_df)

## [1] 2500   10
```

Rename the column names for ease.

```
renamed_df <- filtered_df %>% rename("Quadrant" = s4qt, "Income" = q39,
"Tenancy" = q34, "Years_in_yyc" = q37, "Education" = q38, "Children" = q32x,
"Minority" = q40, "Gender" = q29x, "Age" = q30, "Satisfaction_level" = q2a )
names(renamed_df)

##  [1] "Quadrant"           "Income"             "Tenancy"
##  [4] "Years_in_yyc"       "Education"          "Children"
##  [7] "Minority"           "Gender"             "Age"
## [10] "Satisfaction_level"
```

Satisfaction_level is our response variable. Quadrant, Income, Tenancy, Years_in_yyc, Education, Children, Minority, Gender, Age are our predictor variables. ## Inspect the structure of our df

```
str(renamed_df)

## 'data.frame':    2500 obs. of  10 variables:
##  $ Quadrant          : int  4 1 2 3 1 1 1 2 1 1 ...
##  $ Income            : int  4 1 7 9 4 1 4 7 9 2 ...
##  $ Tenancy           : int  1 1 1 1 1 2 1 1 1 1 ...
##  $ Years_in_yyc      : int  9 9 8 8 10 1 11 8 3 7 ...
##  $ Education         : int  3 2 2 3 2 3 3 2 2 3 ...
##  $ Children          : int  2 2 2 2 2 2 2 2 1 2 ...
##  $ Minority          : int  2 2 2 3 2 1 2 2 2 2 ...
##  $ Gender            : int  1 2 1 1 1 2 1 1 1 1 ...
##  $ Age               : int  6 6 6 6 5 6 6 5 4 6 ...
##  $ Satisfaction_level: int  6 6 8 7 8 7 7 10 8 5 ...

df <- renamed_df
str(df)

## 'data.frame':    2500 obs. of  10 variables:
##  $ Quadrant          : int  4 1 2 3 1 1 1 2 1 1 ...
##  $ Income            : int  4 1 7 9 4 1 4 7 9 2 ...
##  $ Tenancy           : int  1 1 1 1 1 2 1 1 1 1 ...
##  $ Years_in_yyc      : int  9 9 8 8 10 1 11 8 3 7 ...
##  $ Education         : int  3 2 2 3 2 3 3 2 2 3 ...
##  $ Children          : int  2 2 2 2 2 2 2 2 1 2 ...
##  $ Minority          : int  2 2 2 3 2 1 2 2 2 2 ...
```

```
##  $ Gender            : int  1 2 1 1 1 2 1 1 1 1 ...
##  $ Age               : int  6 6 6 6 5 6 6 5 4 6 ...
##  $ Satisfaction_level: int  6 6 8 7 8 7 7 10 8 5 ...
```

Our predictors and response variables are categorical but are coded as int. Let us take a closer look at each variable

variable “Quadrant”

```
table(df$Quadrant)

##
##   1   2   3   4
## 659 603 721 517
```

Converting the Quadrant column to nominal column

```
df <- df %>%
  mutate(Quadrant = factor(Quadrant, levels = c(1, 2, 3, 4),
                           labels = c("SW", "SE", "NW", "NE")))

table(df$Quadrant)

##
##  SW  SE  NW  NE
## 659 603 721 517
```

Variable “Income”

```
table(df$Income)

##
##   1   2   3   4   5   6   7   8   9
## 153 179 241 179 190 254 203 240 861

df <- df %>%
  mutate(Income = factor(Income, levels = c(1, 2, 3, 4, 5, 6, 7, 8, 9),
                           labels = c("<30k", "30k-45k", "45k-60k", "60k-
75k", "75k-90k", "90k-105", "105k-120k", ">120k", "Don't know")))

table(df$Income)

##
##       <30k    30k-45k    45k-60k    60k-75k    75k-90k    90k-105  105k-
120k
##        153        179        241        179        190        254
203
##      >120k Don't know
##        240        861
```

We wanted to treat Income level as an ordinal variable. However, since there are a large number of records(861 out of 2500) that have Income as “Dont know”, we decided to treat it as nominal variable.

Variable Tenancy

```
table(df$Tenancy)

##
##    1    2    3    4    5
## 2012  430   12   38    8
```

Converting the Tenancy column to nominal column

```
df <- df %>%
  mutate(Tenancy = factor(Tenancy, levels = c(1, 2, 3, 4, 5),
                          labels = c("Own", "Rent", "Other", "Neither",
"Don't know")))

table(df$Tenancy)

##
##        Own       Rent      Other    Neither Don't know
##       2012        430         12         38          8
```

Variable Years_in_yyc

```
table(df$Years_in_yyc)

##
##   1   2   3   4   5   6   7   8   9  10  11  12
## 112 146 195 228 302 202 230 182 250 165 479   9

#df <- df %>% mutate(Years_in_yyc = ifelse(Years_in_yyc == 12,
11,Years_in_yyc))

#table(df$Years_in_yyc)

df <- df %>%
  mutate(Years_in_yyc = factor(Years_in_yyc,
                               levels = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11,
12),
                               labels = c("< 5 years", "5 to <10 years", "10
to <15 years",
                                          "15 to <20 years", "20 to <25
years", "25 to <30 years",
                                          "30 to <35 years", "35 to <40
years", "40 to <45 years",
                                          "45 to <50 years", "50 years or
more", "Prefer not to answer"),
                               ordered = TRUE))
```

```r
table(df$Years_in_yyc)
```

```
##
##            < 5 years       5 to <10 years      10 to <15 years
##                  112                  146                  195
##      15 to <20 years      20 to <25 years      25 to <30 years
##                  228                  302                  202
##      30 to <35 years      35 to <40 years      40 to <45 years
##                  230                  182                  250
##      45 to <50 years     50 years or more Prefer not to answer
##                  165                  479                    9
```

```r
mode_value <- sort(table(df$Years_in_yyc), decreasing = TRUE)[1]
mode_value
```

```
## 50 years or more
##              479
```

Variable Education

```r
table(df$Education)
```

```
##
##    1    2    3    4
##  378  879 1160   83
```

```r
df <- df %>% mutate(Education = factor(Education,
                                       levels = c(1,2,3,4),
                                       labels = c("High School or less", "
Post Secondary or College Diploma", "University or PG degree", "Don't
know")))
```

level 4 = “Dont know” and there are 83 such records. As somebody with a university degree or a pg degree would not be answering “Dont know” we decided against making this variable ordinal, choosing nominal instead.

```r
table(df$Education)
```

```
##
##               High School or less  Post Secondary or College Diploma
##                               378                                879
##           University or PG degree                          Don't know
##                              1160                                 83
```

Variable Children

```r
table(df$Children)
```

```
##
##    1    2    3
##  672 1821    7
```

```r
df <- df %>% mutate(Children = factor(Children,
                                      levels = c(1,2,3),
                                      labels = c("Yes", "No",
"Don't know")))

table(df$Children)

##
##         Yes          No Don't know
##         672        1821           7
```

Variable Minority

```r
table(df$Minority)

##
##    1    2    3
##  477 1947   76

df <- df %>% mutate(Minority = factor(Minority,
                                      levels = c(1,2,3),
                                      labels = c("Yes", "No", "Don't know")))

table(df$Minority)

##
##         Yes          No Don't know
##         477        1947          76
```

Variable Gender

```r
table(df$Gender)

##
##    1    2    3    7
## 1314 1170    8    8

df <- df %>% mutate(Gender = factor(Gender,
                                    levels = c(1,2,3, 7),
                                    labels = c("Yes", "No", "Other", "Prefer
not to answer")))

table(df$Gender)

##
##                  Yes                   No                Other
##                 1314                 1170                    8
## Prefer not to answer
##                    8
```

Variable Age

```r
table(df$Age)
```

```
##
##   1   2   3   4   5   6   7
## 148 225 381 416 496 784  50
```

```
df <- df %>%
  mutate(Age = factor(Age,
                      levels = c(1, 2, 3, 4, 5, 6, 7),
                      labels = c("18-24 years", "25-34 years", "35-44 years",
                                 "45-54 years", "55-64years", "65+ years",
                                 "Prefer not to answer")))

table(df$Age)
```

```
##
##          18-24 years          25-34 years          35-44 years
##                  148                  225                  381
##          45-54 years          55-64years            65+ years
##                  416                  496                  784
## Prefer not to answer
##                   50
```

```
str(df)
```

```
## 'data.frame':    2500 obs. of  10 variables:
##  $ Quadrant          : Factor w/ 4 levels "SW","SE","NW",..: 4 1 2 3 1 1 1
2 1 1 ...
##  $ Income            : Factor w/ 9 levels "<30k","30k-45k",..: 4 1 7 9 4 1
4 7 9 2 ...
##  $ Tenancy           : Factor w/ 5 levels "Own","Rent","Other",..: 1 1 1 1
1 2 1 1 1 1 ...
##  $ Years_in_yyc      : Ord.factor w/ 12 levels "< 5 years"<"5 to <10
years"<..: 9 9 8 8 10 1 11 8 3 7 ...
##  $ Education         : Factor w/ 4 levels "High School or less",..: 3 2 2
3 2 3 3 2 2 3 ...
##  $ Children          : Factor w/ 3 levels "Yes","No","Don't know": 2 2 2 2
2 2 2 2 1 2 ...
##  $ Minority          : Factor w/ 3 levels "Yes","No","Don't know": 2 2 2 3
2 1 2 2 2 2 ...
##  $ Gender            : Factor w/ 4 levels "Yes","No","Other",..: 1 2 1 1 1
2 1 1 1 1 ...
##  $ Age               : Factor w/ 7 levels "18-24 years",..: 6 6 6 6 5 6 6
5 4 6 ...
##  $ Satisfaction_level: int  6 6 8 7 8 7 7 10 8 5 ...
```

```
df <- df %>%
  mutate(Satisfaction = factor(ifelse(Satisfaction_level > 5, "Yes", "No"),
                                      levels = c("No", "Yes")))

table(df$Satisfaction)
```

```
##
##   No  Yes
##  319 2181
```

```
sum(table(df$Satisfaction))
```

```
## [1] 2500
```

```
Age_Satis <- table(df$Age, df$Satisfaction)
Age_Satis
```

```
##
##                          No Yes
##   18-24 years             8 140
##   25-34 years            14 211
##   35-44 years            35 346
##   45-54 years            54 362
##   55-64years             72 424
##   65+ years             122 662
##   Prefer not to answer   14  36
```

```
chisq.test(Age_Satis)
```

```
##
##  Pearson's Chi-squared test
##
## data:  Age_Satis
## X-squared = 37.553, df = 6, p-value = 1.374e-06
```

P value is less than 0.05, we reject the null hypothesis and conclude that Age and Satisfaction are dependent

```
Income_Satis <- table(df$Income, df$Satisfaction)
Income_Satis
```

```
##
##                No Yes
##   <30k         37 116
##   30k-45k      26 153
##   45k-60k      27 214
##   60k-75k      22 157
##   75k-90k      21 169
##   90k-105      26 228
##   105k-120k    14 189
##   >120k        25 215
##   Don't know  121 740
```

```
chisq.test(Income_Satis)
```

```
##
##  Pearson's Chi-squared test
##
```

```
## data:  Income_Satis
## X-squared = 29.694, df = 8, p-value = 0.0002394
```

```
Gender_Satis <- table(df$Gender, df$Satisfaction)
Gender_Satis
```

```
##
##                            No  Yes
##   Yes                     193 1121
##   No                      122 1048
##   Other                     2    6
##   Prefer not to answer      2    6
```

```
chisq.test(Gender_Satis)
```

```
## Warning in chisq.test(Gender_Satis): Chi-squared approximation may be
incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  Gender_Satis
## X-squared = 12.26, df = 3, p-value = 0.006544
```

```
Years_Satis <- table(df$Years_in_yyc, df$Satisfaction)
Years_Satis
```

```
##
##                           No Yes
##   < 5 years                6 106
##   5 to <10 years          10 136
##   10 to <15 years         21 174
##   15 to <20 years         22 206
##   20 to <25 years         29 273
##   25 to <30 years         29 173
##   30 to <35 years         27 203
##   35 to <40 years         22 160
##   40 to <45 years         44 206
##   45 to <50 years         24 141
##   50 years or more        84 395
##   Prefer not to answer     1   8
```

```
chisq.test(Years_Satis)
```

```
##
##  Pearson's Chi-squared test
##
## data:  Years_Satis
## X-squared = 31.801, df = 11, p-value = 0.0008208
```

```
Education_Satis <- table(df$Education, df$Satisfaction)
Education_Satis
```

```
##
##                                         No  Yes
##   High School or less                   60  318
##    Post Secondary or College Diploma   134  745
##   University or PG degree              102 1058
##   Don't know                            23   60
```

```
chisq.test(Education_Satis)
```

```
##
##  Pearson's Chi-squared test
##
## data:  Education_Satis
## X-squared = 41.23, df = 3, p-value = 5.845e-09
```

```
Quadrant_Satis <- table(df$Quadrant, df$Satisfaction)
Quadrant_Satis
```

```
##
##       No Yes
##   SW  80 579
##   SE  84 519
##   NW  77 644
##   NE  78 439
```

```
chisq.test(Quadrant_Satis)
```

```
##
##  Pearson's Chi-squared test
##
## data:  Quadrant_Satis
## X-squared = 6.288, df = 3, p-value = 0.09841
```

```
Children_Satis <- table(df$Children, df$Satisfaction)
Children_Satis
```

```
##
##                No  Yes
##   Yes          66  606
##   No          248 1573
##   Don't know    5    2
```

```
chisq.test(Children_Satis)
```

```
## Warning in chisq.test(Children_Satis): Chi-squared approximation may be
## incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  Children_Satis
## X-squared = 28.064, df = 2, p-value = 8.054e-07
```

```r
Minority_Satis <- table(df$Minority, df$Satisfaction)
Minority_Satis
```

```
##
##                No  Yes
##   Yes          47  430
##   No          253 1694
##   Don't know   19   57
```

```r
chisq.test(Minority_Satis)
```

```
##
##  Pearson's Chi-squared test
##
## data:  Minority_Satis
## X-squared = 13.945, df = 2, p-value = 0.0009373
```

```r
Tenancy_Satis <- table(df$Tenancy, df$Satisfaction)
Tenancy_Satis
```

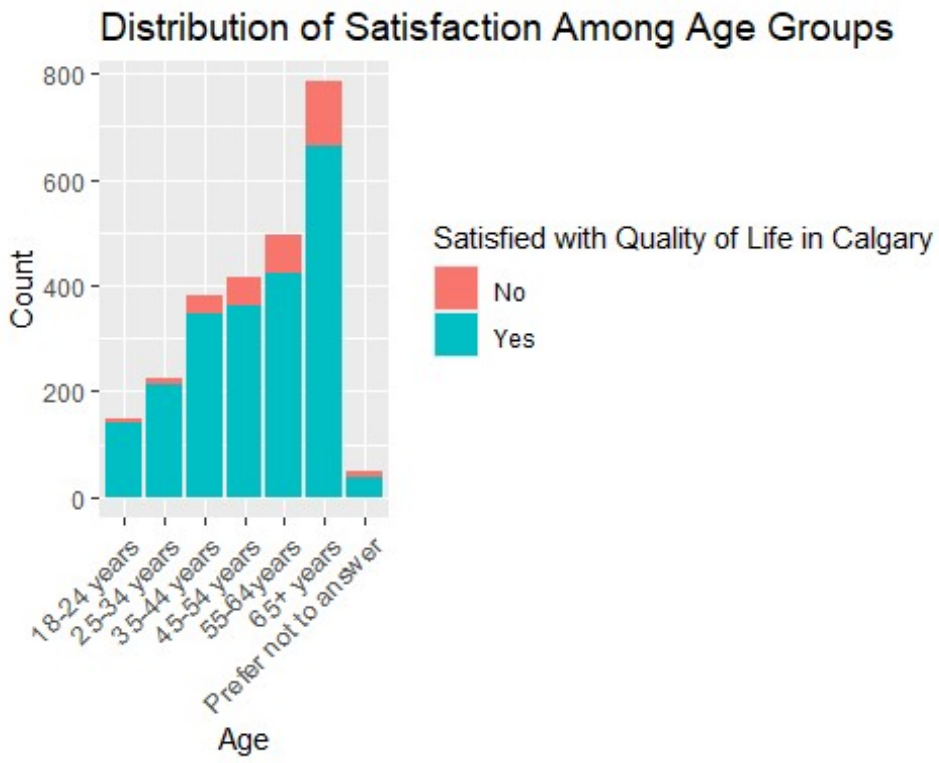
```
##
##                No  Yes
##   Own         255 1757
##   Rent         54  376
##   Other         1   11
##   Neither       5   33
##   Don't know    4    4
```

```r
chisq.test(Tenancy_Satis)
```

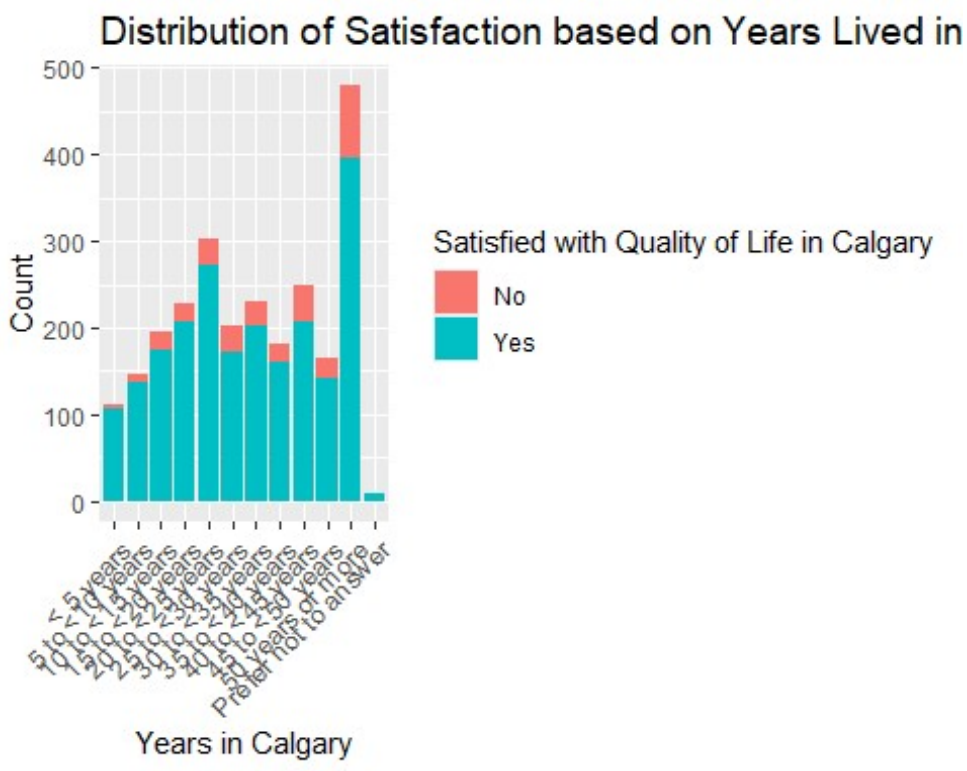
```
## Warning in chisq.test(Tenancy_Satis): Chi-squared approximation may be
## incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  Tenancy_Satis
## X-squared = 10.212, df = 4, p-value = 0.037
```

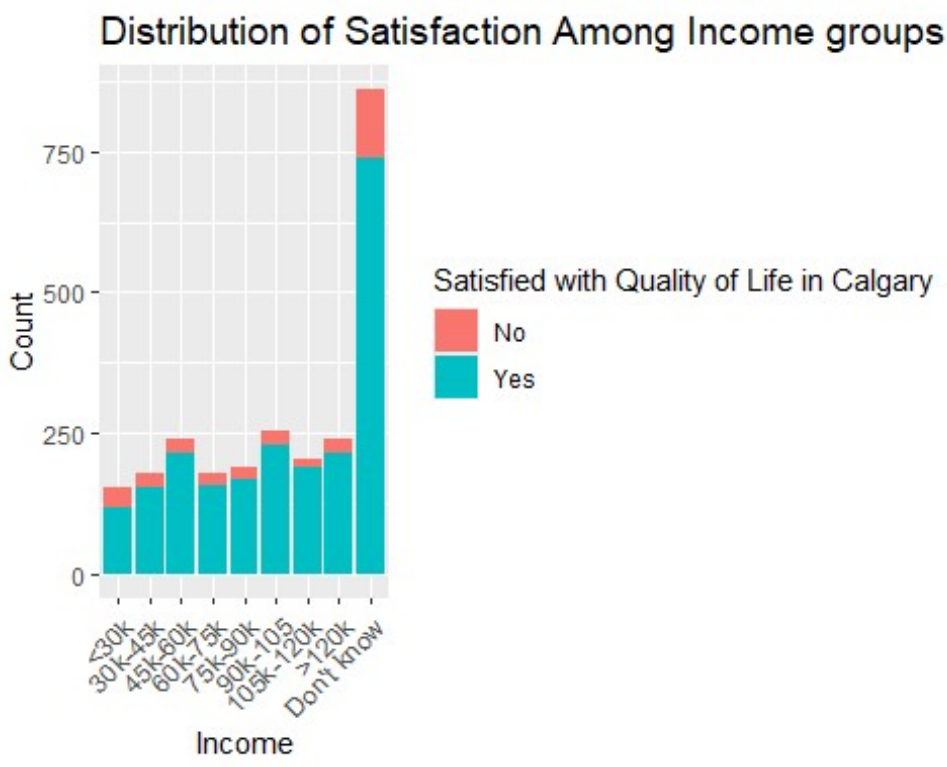
```r
ggplot(df, aes(x=Age, fill=Satisfaction)) +
geom_bar() +
xlab("Age") +
ylab("Count") +
ggtitle("Distribution of Satisfaction Among Age Groups") +
scale_fill_discrete(name = "Satisfied with Quality of Life in Calgary",
labels = c("No", "Yes"))+
    theme(plot.title = element_text(size = 14),
        axis.text.x = element_text(angle = 45, hjust = 1, size = 10))
```



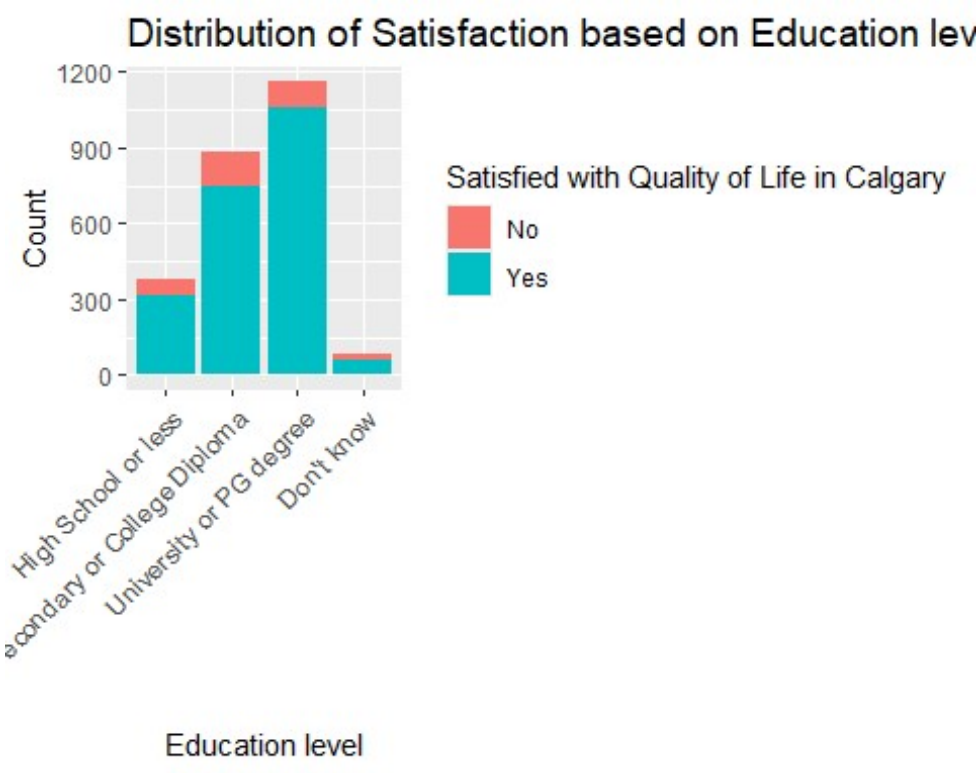

```r
ggplot(df, aes(x=Years_in_yyc, fill=Satisfaction)) +
geom_bar() +
xlab("Years in Calgary") +
ylab("Count") +
ggtitle("Distribution of Satisfaction based on Years Lived in Calgary") +
scale_fill_discrete(name = "Satisfied with Quality of Life in Calgary",
labels = c("No", "Yes"))+
    theme(plot.title = element_text(size = 14),
        axis.text.x = element_text(angle = 45, hjust = 1, size = 10))
```



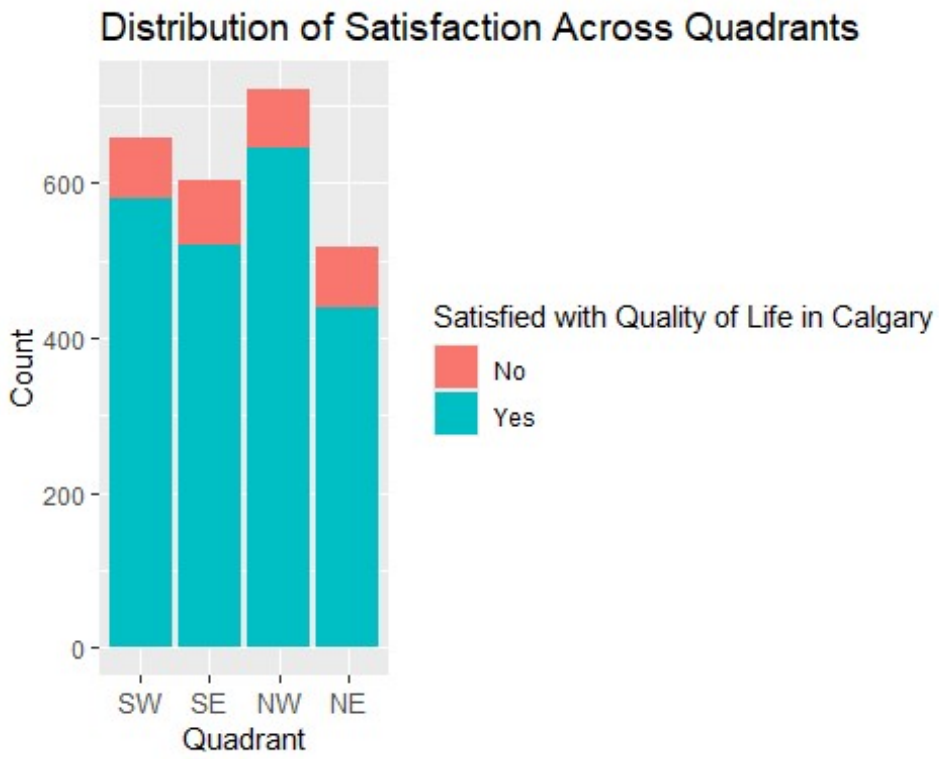

```r
ggplot(df, aes(x=Income, fill=Satisfaction)) +
geom_bar() +
xlab("Income") +
ylab("Count") +
ggtitle("Distribution of Satisfaction Among Income groups") +
scale_fill_discrete(name = "Satisfied with Quality of Life in Calgary",
labels = c("No", "Yes"))+
    theme(plot.title = element_text(size = 14),
        axis.text.x = element_text(angle = 45, hjust = 1, size = 10))
```



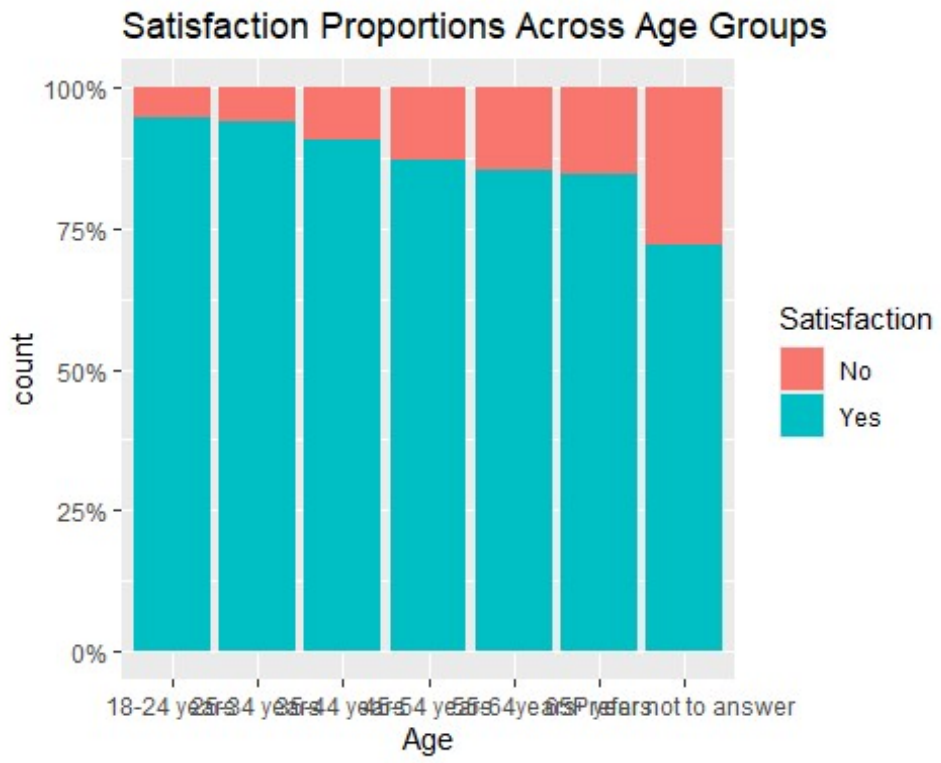

```r
ggplot(df, aes(x=Education, fill=Satisfaction)) +
geom_bar() +
xlab("Education level") +
ylab("Count") +
ggtitle("Distribution of Satisfaction based on Education level") +
scale_fill_discrete(name = "Satisfied with Quality of Life in Calgary",
labels = c("No", "Yes"))+
    theme(plot.title = element_text(size = 14),
        axis.text.x = element_text(angle = 45, hjust = 1, size = 10))
```



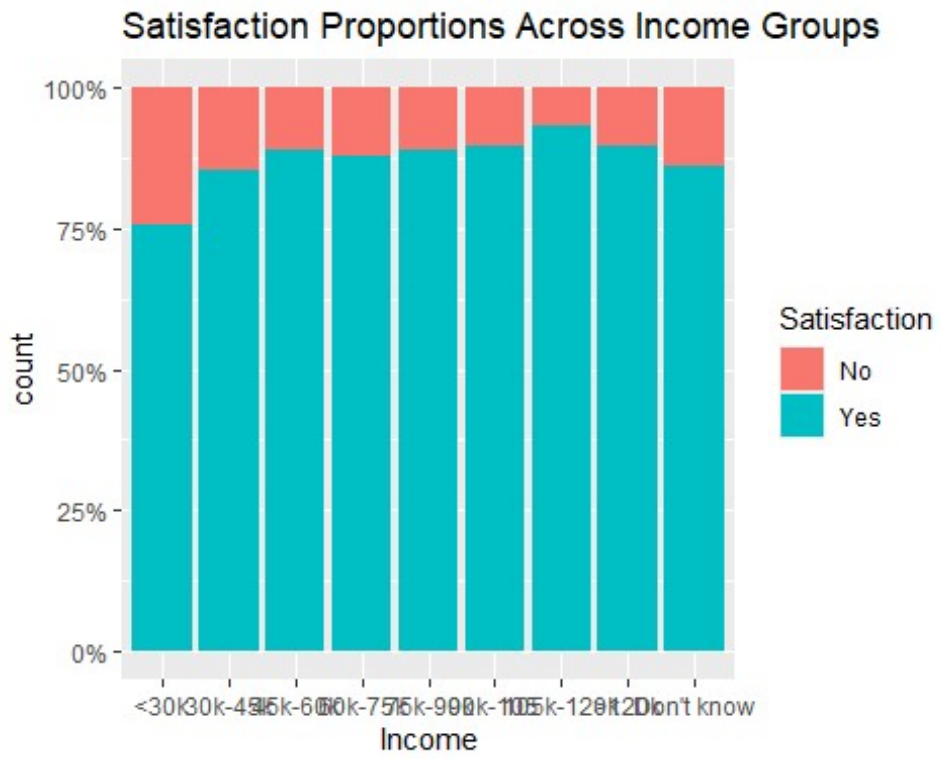

```r
ggplot(df, aes(x=Quadrant, fill=Satisfaction)) +
geom_bar() +
xlab("Quadrant") +
ylab("Count") +
ggtitle("Distribution of Satisfaction Across Quadrants") +
scale_fill_discrete(name = "Satisfied with Quality of Life in Calgary",
labels = c("No", "Yes"))+
    theme(plot.title = element_text(size = 14),
        axis.text.x = element_text(angle = 0, size = 10))
```



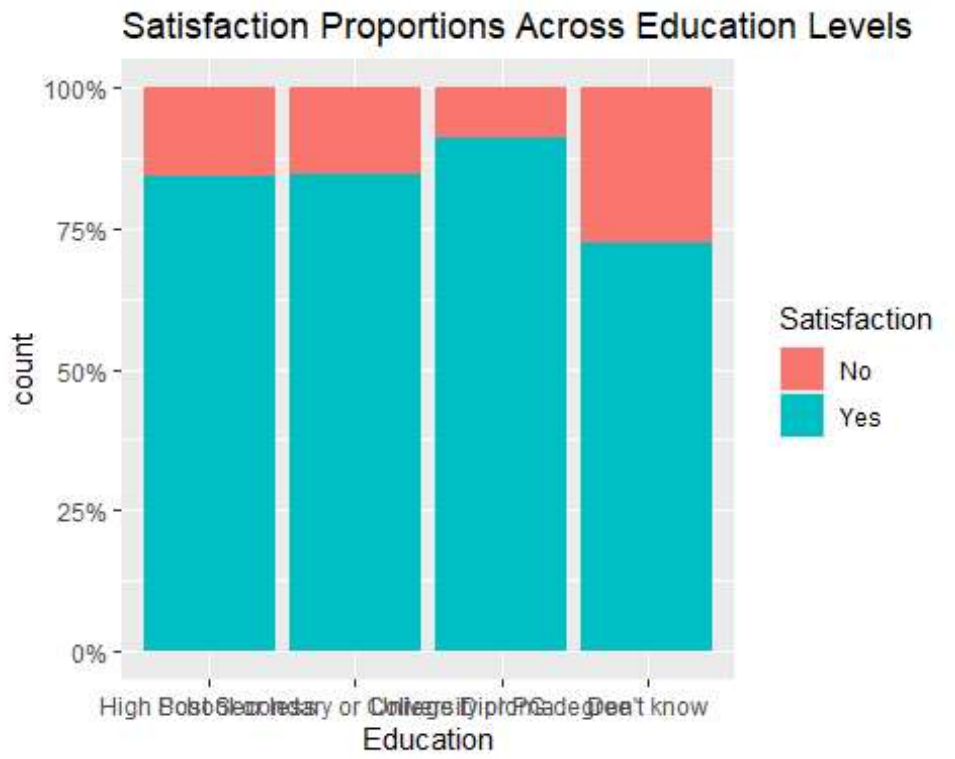

```r
ggplot(df, aes(x = Age, fill = Satisfaction)) +
  geom_bar(position = "fill") +
  scale_y_continuous(labels = scales::percent_format()) +
  ggtitle("Satisfaction Proportions Across Age Groups")
```



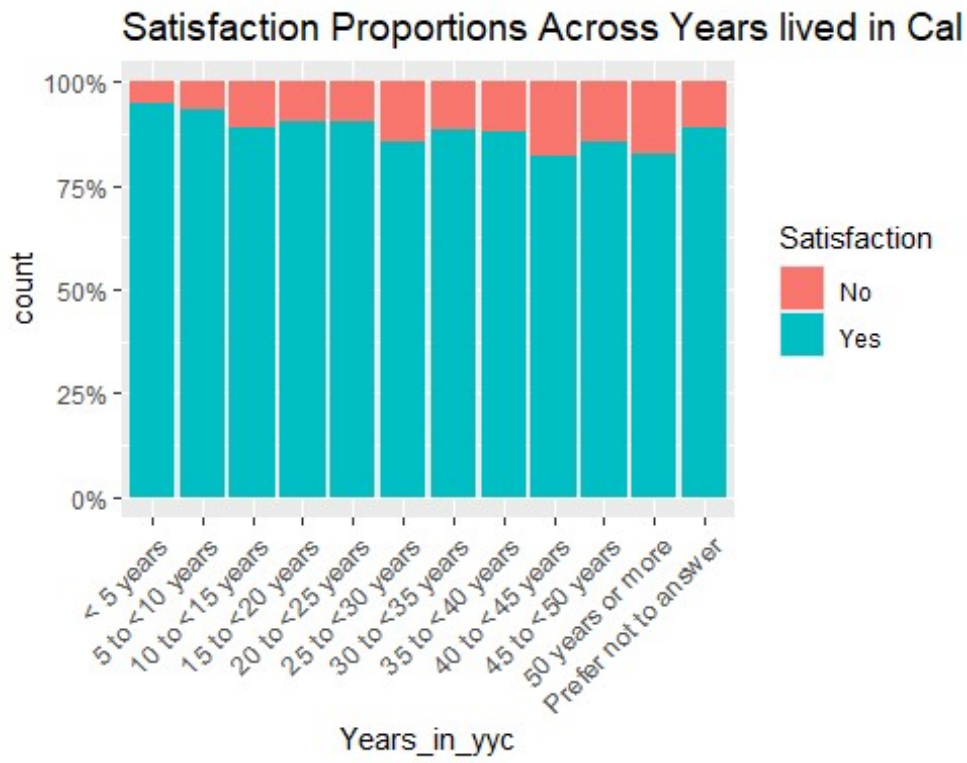

```r
ggplot(df, aes(x = Income, fill = Satisfaction)) +
  geom_bar(position = "fill") +
  scale_y_continuous(labels = scales::percent_format()) +
  ggtitle("Satisfaction Proportions Across Income Groups")
```




```r
ggplot(df, aes(x = Education, fill = Satisfaction)) +
  geom_bar(position = "fill") +
  scale_y_continuous(labels = scales::percent_format()) +
  ggtitle("Satisfaction Proportions Across Education Levels")
```




```
ggplot(df, aes(x = Years_in_yyc, fill = Satisfaction)) +
  geom_bar(position = "fill") +
  scale_y_continuous(labels = scales::percent_format()) +
  ggtitle("Satisfaction Proportions Across Years lived in Calgary")+
      theme(plot.title = element_text(size = 14),
        axis.text.x = element_text(angle = 45, hjust = 1, size = 10))
```


Satisfaction Proportions Across Years lived in Cal
100%
75%
50%
25%
0%
count
< 5 years
5 to <10 years
10 to <15 years
15 to <20 years
20 to <25 years
25 to <30 years
30 to <35 years
35 to <40 years
40 to <45 years
45 to <50 years
50 years or more
Prefer not to answer
Years_in_yyc
Satisfaction
No
Yes