firstdraft

2025-10-23

Priya Mantraratnam

```r
options(repos = c(CRAN = "https://cran.rstudio.com/"))

install.packages("ggfortify")
```

```
## Installing package into 'C:/Users/harip/AppData/Local/R/win-library/4.5'
## (as 'lib' is unspecified)
```

```
## package 'ggfortify' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\harip\AppData\Local\Temp\RtmpAFhqWC\downloaded_packages
```

```r
install.packages("mvnormtest")
```

```
## Installing package into 'C:/Users/harip/AppData/Local/R/win-library/4.5'
## (as 'lib' is unspecified)
```

```
## package 'mvnormtest' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\harip\AppData\Local\Temp\RtmpAFhqWC\downloaded_packages
```

```r
install.packages("datarium")
```

```
## Installing package into 'C:/Users/harip/AppData/Local/R/win-library/4.5'
## (as 'lib' is unspecified)
```

```
## package 'datarium' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\harip\AppData\Local\Temp\RtmpAFhqWC\downloaded_packages
```

```r
install.packages("ggplot2")
```

```
## Installing package into 'C:/Users/harip/AppData/Local/R/win-library/4.5'
## (as 'lib' is unspecified)
## package ’ggplot2’ successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\harip\AppData\Local\Temp\RtmpAFhqWC\downloaded_packages
```

```r
install.packages("caret")
```

```
## Installing package into ’C:/Users/harip/AppData/Local/R/win-library/4.5’
## (as ’lib’ is unspecified)
```

```
## package ’caret’ successfully unpacked and MD5 sums checked
```

```
## Warning: cannot remove prior installation of package ’caret’
```

```
## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying
## C:\Users\harip\AppData\Local\R\win-library\4.5\00LOCK\caret\libs\x64\caret.dll
## to C:\Users\harip\AppData\Local\R\win-library\4.5\caret\libs\x64\caret.dll:
## Permission denied
```

```
## Warning: restored ’caret’
```

```
##
## The downloaded binary packages are in
##  C:\Users\harip\AppData\Local\Temp\RtmpAFhqWC\downloaded_packages
```

```r
install.packages("mvtnorm")
```

```
## Installing package into ’C:/Users/harip/AppData/Local/R/win-library/4.5’
## (as ’lib’ is unspecified)
```

```
## package ’mvtnorm’ successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\harip\AppData\Local\Temp\RtmpAFhqWC\downloaded_packages
```

```r
install.packages("pROC")
```

```
## Installing package into ’C:/Users/harip/AppData/Local/R/win-library/4.5’
## (as ’lib’ is unspecified)
```

```
## package ’pROC’ successfully unpacked and MD5 sums checked
```

```
## Warning: cannot remove prior installation of package ’pROC’
```

```
## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying
## C:\Users\harip\AppData\Local\R\win-library\4.5\00LOCK\pROC\libs\x64\pROC.dll to
## C:\Users\harip\AppData\Local\R\win-library\4.5\pROC\libs\x64\pROC.dll:
## Permission denied
## Warning: restored ’pROC’
```

```
##
## The downloaded binary packages are in
##  C:\Users\harip\AppData\Local\Temp\RtmpAFhqWC\downloaded_packages
```

```r
install.packages("tinytex")
```

```
## Installing package into ’C:/Users/harip/AppData/Local/R/win-library/4.5’
## (as ’lib’ is unspecified)
```

```
## package ’tinytex’ successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\harip\AppData\Local\Temp\RtmpAFhqWC\downloaded_packages
```

```r
install.packages("scales")
```

```
## Installing package into ’C:/Users/harip/AppData/Local/R/win-library/4.5’
## (as ’lib’ is unspecified)
```

```
## package ’scales’ successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\harip\AppData\Local\Temp\RtmpAFhqWC\downloaded_packages
```

```r
install.packages("janitor")
```

```
## Installing package into ’C:/Users/harip/AppData/Local/R/win-library/4.5’
## (as ’lib’ is unspecified)
```

```
## package ’janitor’ successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\harip\AppData\Local\Temp\RtmpAFhqWC\downloaded_packages
```

```r
library(MASS)
library(datarium)
library(ggplot2)
library(broom)
library(ggfortify)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.1     v stringr   1.5.2
## v lubridate 1.9.4     v tibble    3.3.0
## v purrr     1.1.0     v tidyr     1.3.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## x dplyr::select() masks MASS::select()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(mvnormtest)
library(data.table)
```

```
##
## Attaching package: 'data.table'
##
## The following objects are masked from 'package:lubridate':
##
##     hour, isoweek, mday, minute, month, quarter, second, wday, week,
##     yday, year
##
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
##
## The following object is masked from 'package:purrr':
##
##     transpose
```

```
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'
##
## The following object is masked from 'package:dplyr':
##
##     combine
```

```
library(dplyr)
library(tinytex)
library(ggplot2)
library(tidyr)
library(dplyr)
library(scales)
```

```
##
## Attaching package: 'scales'
##
## The following object is masked from 'package:purrr':
##
##     discard
##
## The following object is masked from 'package:readr':
##
##     col_factor
```

Histograms for higher education dataset: https://archive.ics.uci.edu/dataset/697/predict+students+dropout+and+academic+success

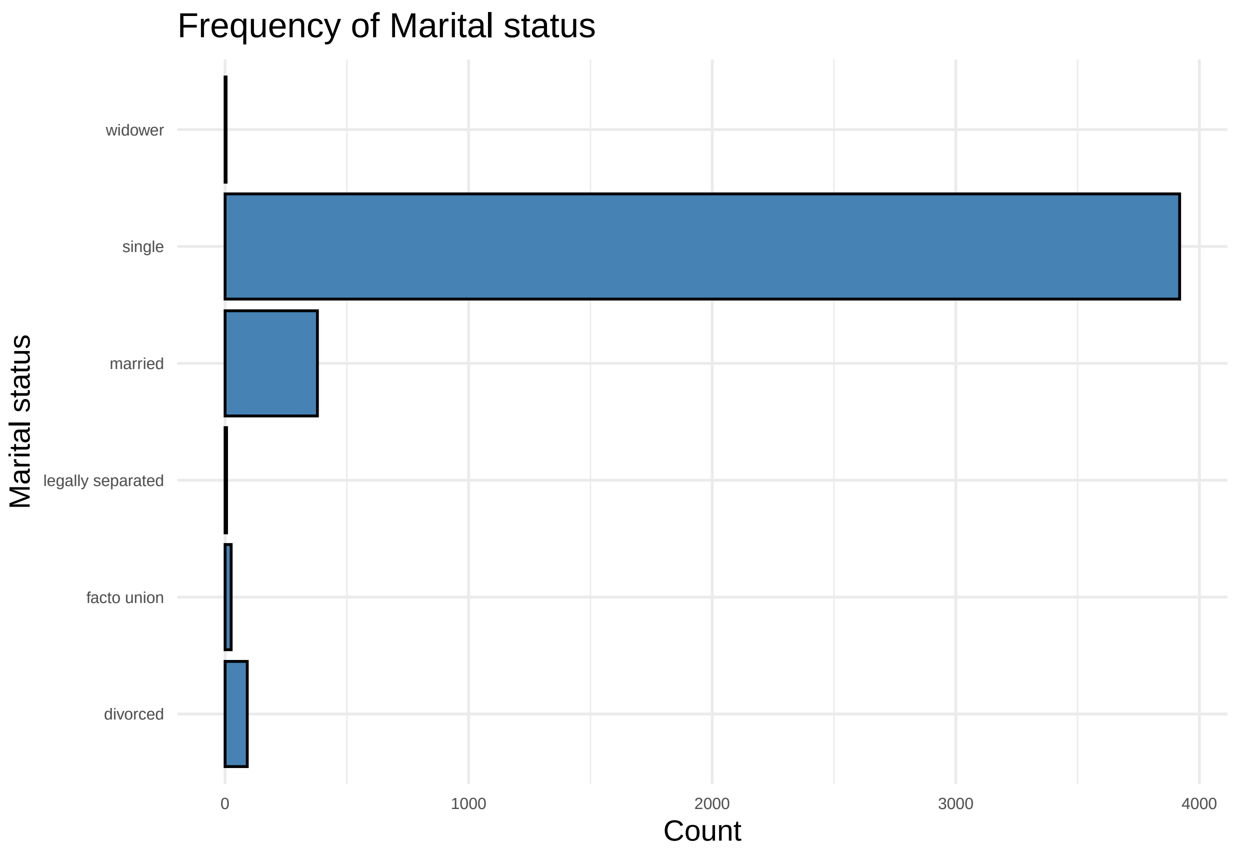
```
data1 <- read.csv("C:/Users/harip/Downloads/predict+students+dropout+and+academic+success/data.csv", sep
head(data1)
```

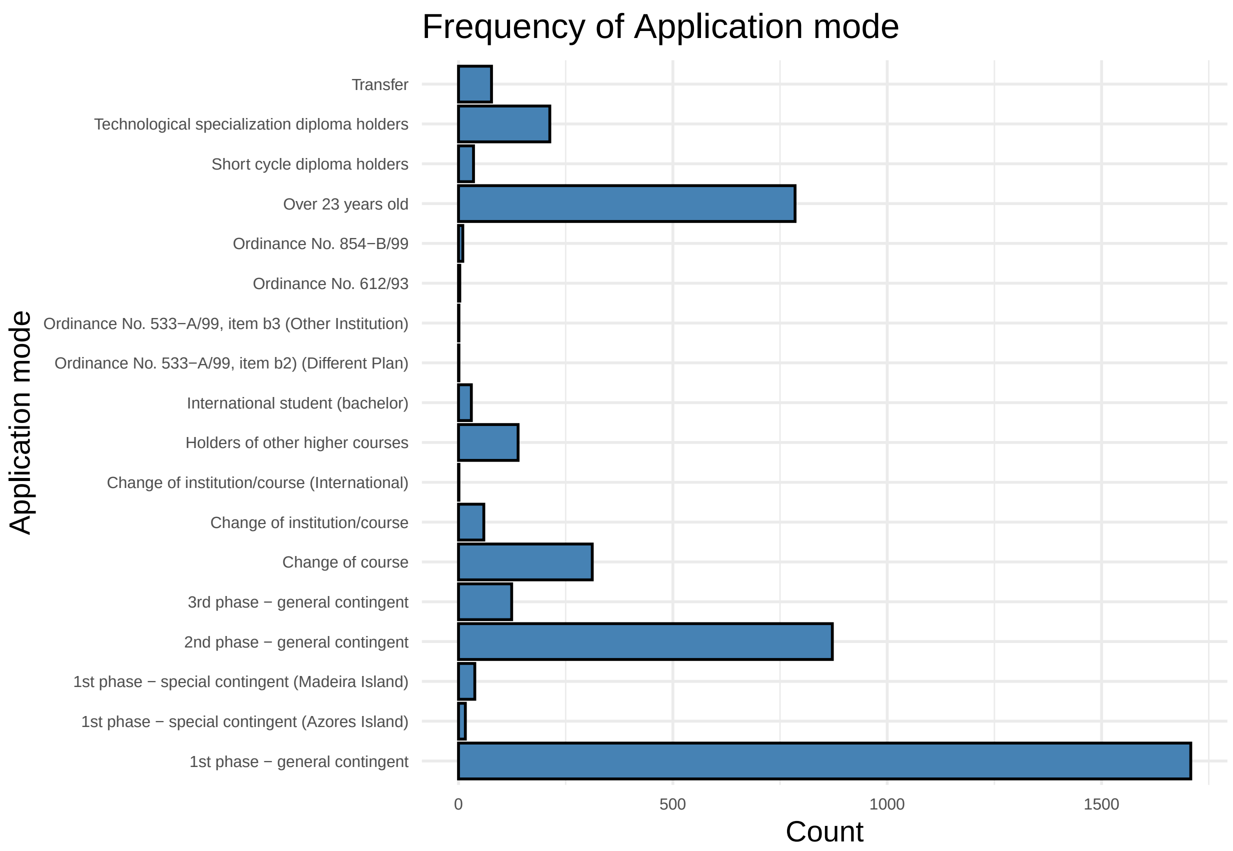
```
##   Marital.status Application.mode Application.order Course
## 1              1               17                 5    171
## 2              1               15                 1   9254
## 3              1                1                 5   9070
## 4              1               17                 2   9773
## 5              2               39                 1   8014
## 6              2               39                 1   9991
##   Daytime.evening.attendance. Previous.qualification
## 1                           1                      1
## 2                           1                      1
## 3                           1                      1
## 4                           1                      1
## 5                           0                      1
## 6                           0                     19
##   Previous.qualification..grade. Nacionality Mother.s.qualification
## 1                          122.0           1                     19
## 2                          160.0           1                      1
## 3                          122.0           1                     37
## 4                          122.0           1                     38
## 5                          100.0           1                     37
## 6                          133.1           1                     37
##   Father.s.qualification Mother.s.occupation Father.s.occupation
## 1                     12                   5                   9
## 2                      3                   3                   3
## 3                     37                   9                   9
## 4                     37                   5                   3
## 5                     38                   9                   9
## 6                     37                   9                   7
##   Admission.grade Displaced Educational.special.needs Debtor
## 1           127.3         1                         0      0
## 2           142.5         1                         0      0
## 3           124.8         1                         0      0
## 4           119.6         1                         0      0
## 5           141.5         0                         0      0
## 6           114.8         0                         0      1
##   Tuition.fees.up.to.date Gender Scholarship.holder Age.at.enrollment
## 1                       1      1                  0                20
## 2                       0      1                  0                19
## 3                       0      1                  0                19
## 4                       1      0                  0                20
## 5                       1      0                  0                45
## 6                       1      1                  0                50
##   International Curricular.units.1st.sem..credited.
## 1             0                                   0
## 2             0                                   0
## 3             0                                   0
## 4             0                                   0
## 5             0                                   0
## 6             0                                   0
##   Curricular.units.1st.sem..enrolled. Curricular.units.1st.sem..evaluations.
## 1                                   0                                      0
## 2                                   6                                      6
## 3                                   6                                      0
## 4                                   6                                      8
## 5                                   6                                      9
## 6                                   5                                     10
##   Curricular.units.1st.sem..approved. Curricular.units.1st.sem..grade.
## 1                                   0                          0.00000
## 2                                   6                         14.00000
## 3                                   0                          0.00000
## 4                                   6                         13.42857
## 5                                   5                         12.33333
## 6                                   5                         11.85714
##   Curricular.units.1st.sem..without.evaluations.
## 1                                              0
## 2                                              0
## 3                                              0
## 4                                              0
## 5                                              0
## 6                                              0
##   Curricular.units.2nd.sem..credited. Curricular.units.2nd.sem..enrolled.
## 1                                   0                                   0
## 2                                   0                                   6
## 3                                   0                                   6
## 4                                   0                                   6
## 5                                   0                                   6
## 6                                   0                                   5
##   Curricular.units.2nd.sem..evaluations. Curricular.units.2nd.sem..approved.
## 1                                      0                                   0
## 2                                      6                                   6
## 3                                      0                                   0
## 4                                     10                                   5
## 5                                      6                                   6
## 6                                     17                                   5
##   Curricular.units.2nd.sem..grade.
## 1                          0.00000
## 2                         13.66667
## 3                          0.00000
## 4                         12.40000
## 5                         13.00000
## 6                         11.50000
##   Curricular.units.2nd.sem..without.evaluations. Unemployment.rate
## 1                                              0              10.8
## 2                                              0              13.9
## 3                                              0              10.8
## 4                                              0               9.4
## 5                                              0              13.9
## 6                                              5              16.2
##   Inflation.rate   GDP   Target
## 1            1.4  1.74  Dropout
## 2           -0.3  0.79 Graduate
## 3            1.4  1.74  Dropout
## 4           -0.8 -3.12 Graduate
## 5           -0.3  0.79 Graduate
## 6            0.3 -0.92 Graduate
```

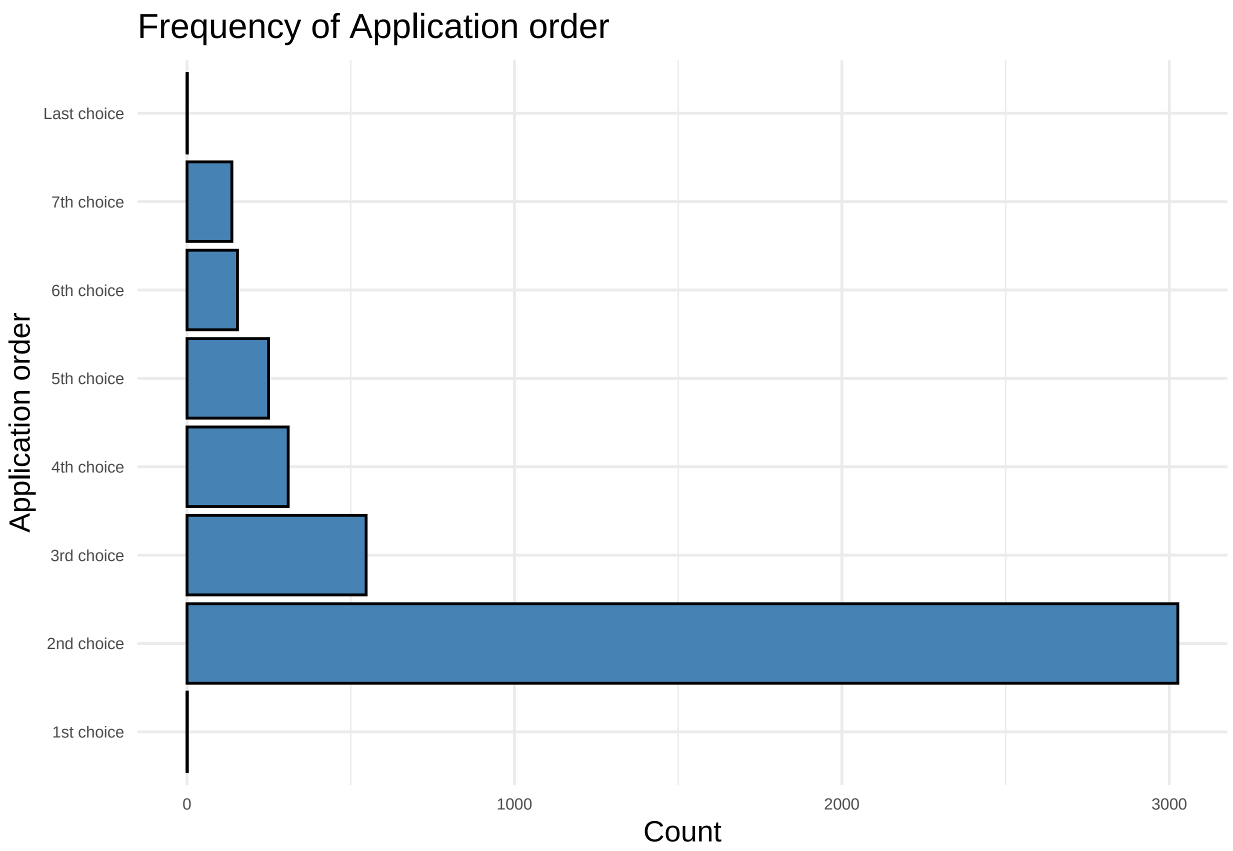
```
colnames(data1)
```

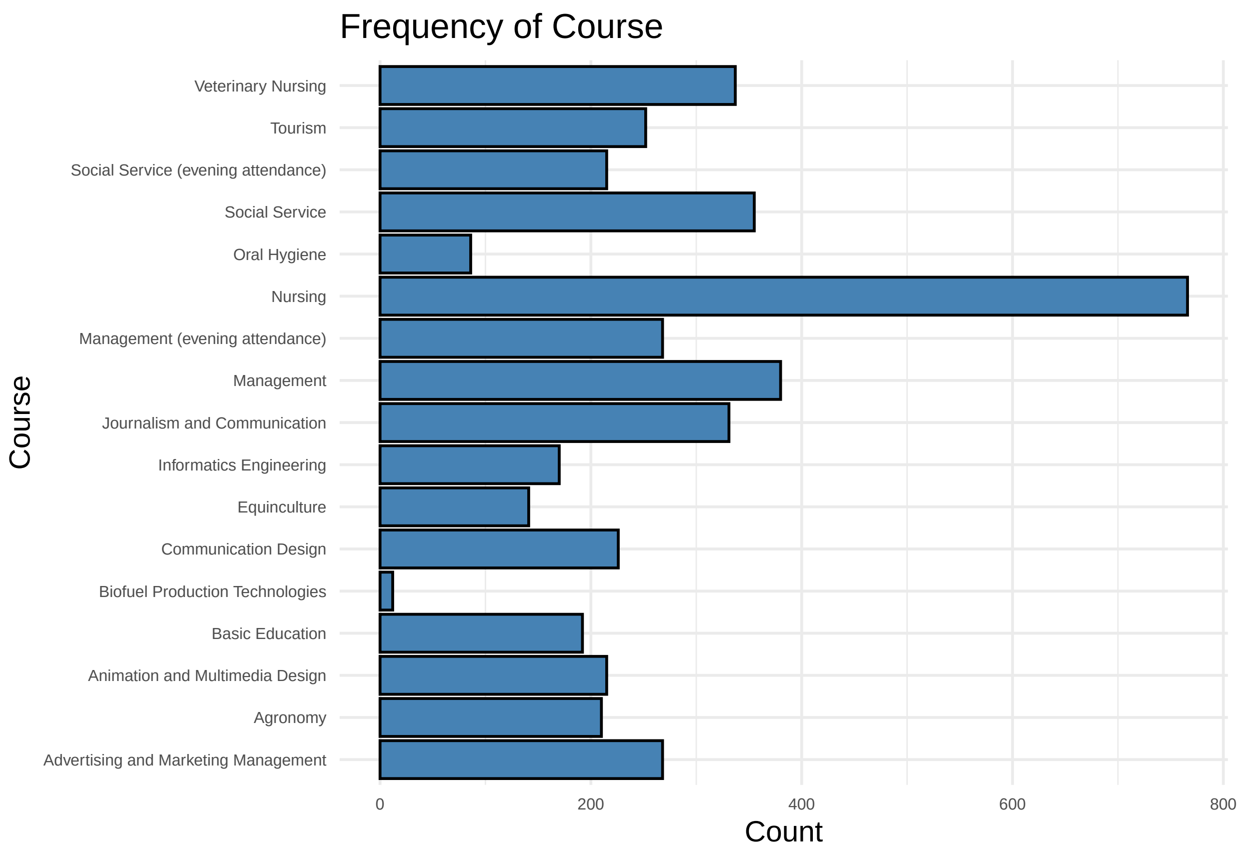
```
##  [1] "Marital.status"
##  [2] "Application.mode"
##  [3] "Application.order"
##  [4] "Course"
##  [5] "Daytime.evening.attendance."
##  [6] "Previous.qualification"
##  [7] "Previous.qualification..grade."
##  [8] "Nacionality"
##  [9] "Mother.s.qualification"
## [10] "Father.s.qualification"
## [11] "Mother.s.occupation"
## [12] "Father.s.occupation"
## [13] "Admission.grade"
## [14] "Displaced"
## [15] "Educational.special.needs"
## [16] "Debtor"
## [17] "Tuition.fees.up.to.date"
## [18] "Gender"
## [19] "Scholarship.holder"
## [20] "Age.at.enrollment"
## [21] "International"
## [22] "Curricular.units.1st.sem..credited."
## [23] "Curricular.units.1st.sem..enrolled."
## [24] "Curricular.units.1st.sem..evaluations."
## [25] "Curricular.units.1st.sem..approved."
## [26] "Curricular.units.1st.sem..grade."
## [27] "Curricular.units.1st.sem..without.evaluations."
## [28] "Curricular.units.2nd.sem..credited."
## [29] "Curricular.units.2nd.sem..enrolled."
## [30] "Curricular.units.2nd.sem..evaluations."
## [31] "Curricular.units.2nd.sem..approved."
## [32] "Curricular.units.2nd.sem..grade."
## [33] "Curricular.units.2nd.sem..without.evaluations."
## [34] "Unemployment.rate"
## [35] "Inflation.rate"
## [36] "GDP"
## [37] "Target"
```

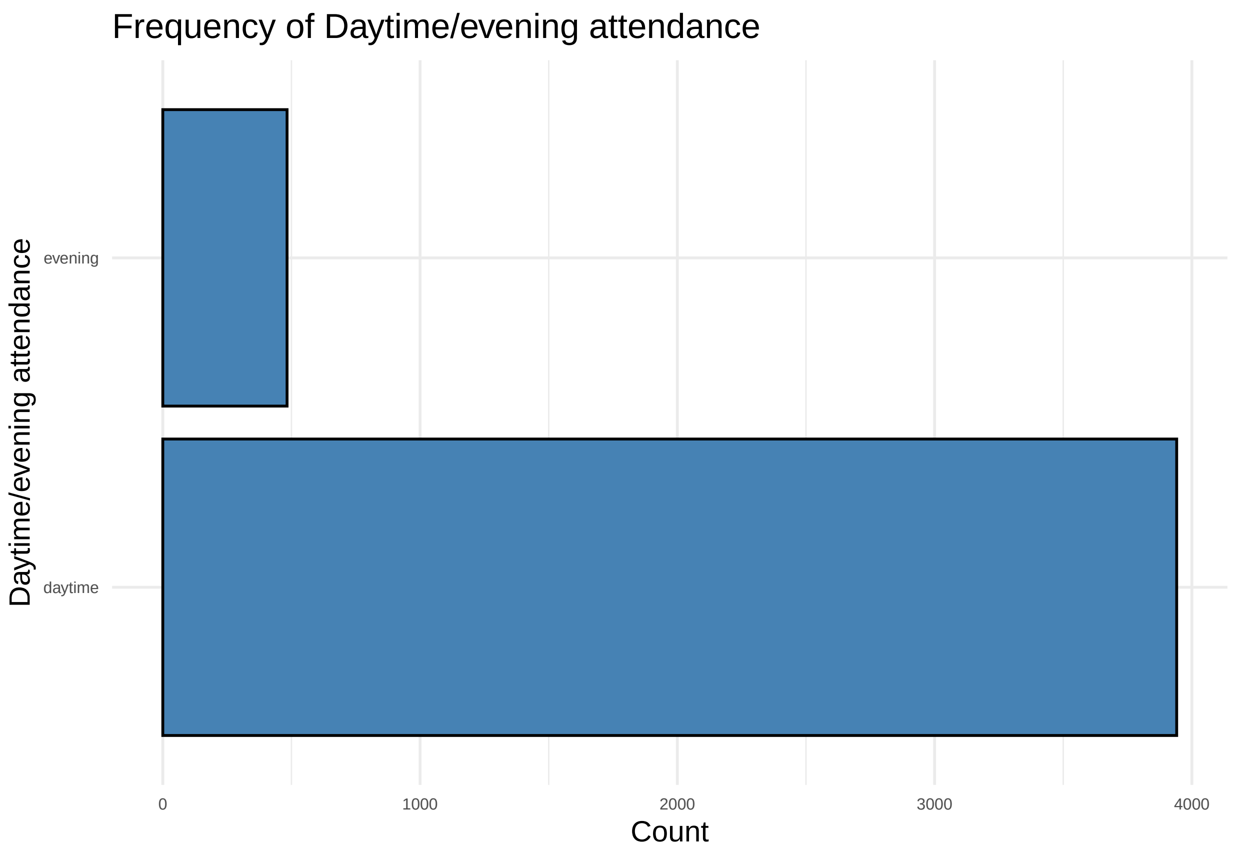
```
names(data1) <- c("Marital status",
                  "Application mode",
                  "Application order",
                  "Course",
                  "Daytime/evening attendance",
                  "Previous qualification",
                  "Previous qualification grade",
                  "Nationality",
                  "Mother's qualification",
                  "Father's qualification",
                  "Mother's occupation",
                  "Father's occupation",
                  "Admission grade",
                  "Displaced",
                  "Educational special needs",
                  "Debtor",
                  "Tuition fees up to date",
                  "Gender",
                  "Scholarship holder",
                  "Age at enrollment",
                  "International",
                  "Semester 1 credited units",
                  "Semester 1 enrolled units",
                  "Semester 1 evaluations",
                  "Semester 1 approved units",
                  "Semester 1 grade",
                  "Semester 1 units without evaluations",
                  "Semester 2 credited units",
                  "Semester 2 enrolled units",
                  "Semester 2 evaluations",
                  "Semester 2 approved units",
                  "Semester 2 grade",
                  "Semester 2 units without evaluations",
                  "Unemployment rate",
                  "Inflation rate",
                  "GDP",
                  "Target")

 data1$"Marital status" <- ifelse(data1$"Marital status" == 1, "single",
                           ifelse(data1$"Marital status" == 2, "married",
                           ifelse(data1$"Marital status" == 3, "widower",
                           ifelse(data1$"Marital status" == 4, "divorced",
                           ifelse(data1$"Marital status" == 5, "facto union",
                           ifelse(data1$"Marital status" == 6, "legally separated", NA))))))

 data1$"Application mode" <- ifelse(data1$"Application mode" == 1, "1st phase - general contingent",
                             ifelse(data1$"Application mode" == 2, "Ordinance No. 612/93",
                             ifelse(data1$"Application mode" == 5, "1st phase - special contingent (Azore
                             ifelse(data1$"Application mode" == 7, "Holders of other higher courses",
                             ifelse(data1$"Application mode" == 10, "Ordinance No. 854-B/99",
                             ifelse(data1$"Application mode" == 15, "International student (bachelor)",
                             ifelse(data1$"Application mode" == 16, "1st phase - special contingent (Made
                             ifelse(data1$"Application mode" == 17, "2nd phase - general contingent",
                             ifelse(data1$"Application mode" == 18, "3rd phase - general contingent",
                             ifelse(data1$"Application mode" == 26, "Ordinance No. 533-A/99, item b2) (Di
                             ifelse(data1$"Application mode" == 27, "Ordinance No. 533-A/99, item b3 (Oth
                             ifelse(data1$"Application mode" == 39, "Over 23 years old",
                             ifelse(data1$"Application mode" == 42, "Transfer",
                             ifelse(data1$"Application mode" == 43, "Change of course",
                             ifelse(data1$"Application mode" == 44, "Technological specialization diploma
                             ifelse(data1$"Application mode" == 51, "Change of institution/course",
                             ifelse(data1$"Application mode" == 53, "Short cycle diploma holders",
                             ifelse(data1$"Application mode" == 57, "Change of institution/course (Inter

 data1$"Application order" <- ifelse(data1$"Application order" == 0, "1st choice",
                              ifelse(data1$"Application order" == 1, "2nd choice",
                              ifelse(data1$"Application order" == 2, "3rd choice",
                              ifelse(data1$"Application order" == 3, "4th choice",
                              ifelse(data1$"Application order" == 4, "5th choice",
                              ifelse(data1$"Application order" == 5, "6th choice",
                              ifelse(data1$"Application order" == 6, "7th choice",
                              ifelse(data1$"Application order" == 7, "8th choice",
                              ifelse(data1$"Application order" == 8, "9th choice choice",
                              ifelse(data1$"Application order" == 9, "Last choice", NA))))))))))

 data1$"Course" <- ifelse(data1$"Course" == 33, "Biofuel Production Technologies",
                   ifelse(data1$"Course" == 171, "Animation and Multimedia Design",
                   ifelse(data1$"Course" == 8014, "Social Service (evening attendance)",
                   ifelse(data1$"Course" == 9003, "Agronomy",
                   ifelse(data1$"Course" == 9070, "Communication Design",
                   ifelse(data1$"Course" == 9085, "Veterinary Nursing",
                   ifelse(data1$"Course" == 9119, "Informatics Engineering",
                   ifelse(data1$"Course" == 9130, "Equinculture",
                   ifelse(data1$"Course" == 9147, "Management",
                   ifelse(data1$"Course" == 9238, "Social Service",
                   ifelse(data1$"Course" == 9254, "Tourism",
                   ifelse(data1$"Course" == 9500, "Nursing",
                   ifelse(data1$"Course" == 9556, "Oral Hygiene",
                   ifelse(data1$"Course" == 9670, "Advertising and Marketing Management",
                   ifelse(data1$"Course" == 9773, "Journalism and Communication",
                   ifelse(data1$"Course" == 9853, "Basic Education",
                   ifelse(data1$"Course" == 9991, "Management (evening attendance)", NA)))))))))))))))))

 data1$"Daytime/evening attendance" <- ifelse(data1$"Daytime/evening attendance" == 1, "daytime", "evenin

 data1$"Previous qualification" <- ifelse(data1$"Previous qualification" == 1, "Secondary education",
                                   ifelse(data1$"Previous qualification" == 2, "Higher education - bachel
                                   ifelse(data1$"Previous qualification" == 3, "Higher education - degre
                                   ifelse(data1$"Previous qualification" == 4, "Higher education - maste
                                   ifelse(data1$"Previous qualification" == 5, "Higher education - docto
                                   ifelse(data1$"Previous qualification" == 6, "Frequency of higher educa
                                   ifelse(data1$"Previous qualification" == 9, "12th year of schooling -
                                   ifelse(data1$"Previous qualification" == 10, "11th year of schooling
                                   ifelse(data1$"Previous qualification" == 12, "Other - 11th year of sc
                                   ifelse(data1$"Previous qualification" == 14, "10th year of schooling"
                                   ifelse(data1$"Previous qualification" == 15, "10th year of schooling
                                   ifelse(data1$"Previous qualification" == 19, "Basic education 3rd cycl
                                   ifelse(data1$"Previous qualification" == 38, "Basic education 2nd cycl
                                   ifelse(data1$"Previous qualification" == 39, "Technological speciali
                                   ifelse(data1$"Previous qualification" == 40, "Higher education - degre
                                   ifelse(data1$"Previous qualification" == 42, "Professional higher tec
                                   ifelse(data1$"Previous qualification" == 43, "Higher education - mast

 data1$"Nationality" <- ifelse(data1$"Nationality" == 1, "Portuguese",
                        ifelse(data1$"Nationality" == 2, "German",
                        ifelse(data1$"Nationality" == 6, "Spanish",
                        ifelse(data1$"Nationality" == 11, "Italian",
                        ifelse(data1$"Nationality" == 13, "Dutch",
                        ifelse(data1$"Nationality" == 14, "English",
                        ifelse(data1$"Nationality" == 17, "Lithuanian",
                        ifelse(data1$"Nationality" == 21, "Angolan",
                        ifelse(data1$"Nationality" == 22, "Cape Verdean",
                        ifelse(data1$"Nationality" == 24, "Guinean",
                        ifelse(data1$"Nationality" == 25, "Mozambican",
                        ifelse(data1$"Nationality" == 26, "Santomean",
                        ifelse(data1$"Nationality" == 32, "Turkish",
                        ifelse(data1$"Nationality" == 41, "Brazilian",
                        ifelse(data1$"Nationality" == 62, "Romanian",
                        ifelse(data1$"Nationality" == 100, "Moldova (Republic of)",
                        ifelse(data1$"Nationality" == 101, "Mexican",
                        ifelse(data1$"Nationality" == 103, "Ukrainian",
                        ifelse(data1$"Nationality" == 105, "Russian",
                        ifelse(data1$"Nationality" == 108, "Cuban",
                        ifelse(data1$"Nationality" == 109, "Columbian", NA)))))))))))))))))))))

 data1$"Mother's qualification" <- ifelse(data1$"Mother's qualification" == 1, "Secondary Education - 12t
                                   ifelse(data1$"Mother's qualification" == 2, "Higher Education - Bachel
                                   ifelse(data1$"Mother's qualification" == 3, "Higher Education - Degre
                                   ifelse(data1$"Mother's qualification" == 4, "Higher Education - Maste
                                   ifelse(data1$"Mother's qualification" == 5, "Higher Education - Docto
                                   ifelse(data1$"Mother's qualification" == 6, "Frequency of Higher Educa
                                   ifelse(data1$"Mother's qualification" == 9, "12th Year of Schooling -
                                   ifelse(data1$"Mother's qualification" == 10, "11th Year of Schooling
                                   ifelse(data1$"Mother's qualification" == 11, "7th Year (Old)",
                                   ifelse(data1$"Mother's qualification" == 12, "Other - 11th Year of Sc
                                   ifelse(data1$"Mother's qualification" == 14, "10th Year of Schooling"
                                   ifelse(data1$"Mother's qualification" == 18, "General commerce course
                                   ifelse(data1$"Mother's qualification" == 19, "Basic Education 3rd Cyc
                                   ifelse(data1$"Mother's qualification" == 22, "Technical-professional
                                   ifelse(data1$"Mother's qualification" == 26, "7th year of schooling",
                                   ifelse(data1$"Mother's qualification" == 27, "2nd cycle of the general
                                   ifelse(data1$"Mother's qualification" == 29, "9th Year of Schooling -
                                   ifelse(data1$"Mother's qualification" == 30, "8th year of schooling",
                                   ifelse(data1$"Mother's qualification" == 34, "Unknown",
                                   ifelse(data1$"Mother's qualification" == 35, "Can't read or write",
                                   ifelse(data1$"Mother's qualification" == 36, "Can read without having
                                   ifelse(data1$"Mother's qualification" == 37, "Basic education 1st cyc
                                   ifelse(data1$"Mother's qualification" == 38, "Basic Education 2nd Cyc
                                   ifelse(data1$"Mother's qualification" == 39, "Technological specializ
                                   ifelse(data1$"Mother's qualification" == 40, "Higher education - degr
                                   ifelse(data1$"Mother's qualification" == 41, "Specialized higher stud
                                   ifelse(data1$"Mother's qualification" == 42, "Professional higher tec
                                   ifelse(data1$"Mother's qualification" == 43, "Higher Education - Mast
                                   ifelse(data1$"Mother's qualification" == 44, "Higher Education - Doct

 data1$"Father's qualification" <- ifelse(data1$"Father's qualification" == 1, "Secondary Education - 12t
                                   ifelse(data1$"Father's qualification" == 2, "Higher Education - Bachel
                                   ifelse(data1$"Father's qualification" == 3, "Higher Education - Degre
                                   ifelse(data1$"Father's qualification" == 4, "Higher Education - Maste
                                   ifelse(data1$"Father's qualification" == 5, "Higher Education - Docto
                                   ifelse(data1$"Father's qualification" == 6, "Frequency of Higher Educa
                                   ifelse(data1$"Father's qualification" == 9, "12th Year of Schooling -
                                   ifelse(data1$"Father's qualification" == 10, "11th Year of Schooling
                                   ifelse(data1$"Father's qualification" == 11, "7th Year (Old)",
                                   ifelse(data1$"Father's qualification" == 12, "Other - 11th Year of Sc
                                   ifelse(data1$"Father's qualification" == 13, "2nd year complementary
                                   ifelse(data1$"Father's qualification" == 14, "10th Year of Schooling"
                                   ifelse(data1$"Father's qualification" == 18, "General commerce course
                                   ifelse(data1$"Father's qualification" == 19, "Basic Education 3rd Cyc
                                   ifelse(data1$"Father's qualification" == 20, "Complementary High Scho
                                   ifelse(data1$"Father's qualification" == 22, "Technical-professional
                                   ifelse(data1$"Father's qualification" == 25, "Complementary High Scho
                                   ifelse(data1$"Father's qualification" == 26, "7th year of schooling",
                                   ifelse(data1$"Father's qualification" == 27, "2nd cycle of the genera
                                   ifelse(data1$"Father's qualification" == 29, "9th Year of Schooling -
                                   ifelse(data1$"Father's qualification" == 30, "8th year of schooling",
                                   ifelse(data1$"Father's qualification" == 31, "General Course of Admin
                                   ifelse(data1$"Father's qualification" == 33, "Supplementary Accountin
                                   ifelse(data1$"Father's qualification" == 34, "Unknown",
                                   ifelse(data1$"Father's qualification" == 35, "Can't read or write",
                                   ifelse(data1$"Father's qualification" == 36, "Can read without having
                                   ifelse(data1$"Father's qualification" == 37, "Basic education 1st cyc
                                   ifelse(data1$"Father's qualification" == 38, "Basic Education 2nd Cyc
                                   ifelse(data1$"Father's qualification" == 39, "Technological specializ
                                   ifelse(data1$"Father's qualification" == 40, "Higher education - degr
                                   ifelse(data1$"Father's qualification" == 41, "Specialized higher stud
                                   ifelse(data1$"Father's qualification" == 42, "Professional higher tec
                                   ifelse(data1$"Father's qualification" == 43, "Higher Education - Mast
                                   ifelse(data1$"Mother's qualification" == 44, "Higher Education - Doct

 data1$"Mother's occupation" <- ifelse(data1$"Mother's occupation" == 0, "Student",
                                ifelse(data1$"Mother's occupation" == 1, "Representatives of the Legisla
                                ifelse(data1$"Mother's occupation" == 2, "Specialists in Intellectual an
                                ifelse(data1$"Mother's occupation" == 3, "Intermediate Level Technicians
                                ifelse(data1$"Mother's occupation" == 4, "Administrative staff",
                                ifelse(data1$"Mother's occupation" == 5, "Personal Services, Security an
                                ifelse(data1$"Mother's occupation" == 6, "Farmers and Skilled Workers in
                                ifelse(data1$"Mother's occupation" == 7, "Skilled Workers in Industry, C
                                ifelse(data1$"Mother's occupation" == 8, "Installation and Machine Opera
                                ifelse(data1$"Mother's occupation" == 9, "Unskilled Workers",
                                ifelse(data1$"Mother's occupation" == 10, "Armed Forces Professions 90 -
                                ifelse(data1$"Mother's occupation" == 99, "(Blank)",
                                ifelse(data1$"Mother's occupation" == 122, "Health professionals",
                                ifelse(data1$"Mother's occupation" == 123, "Teachers",
                                ifelse(data1$"Mother's occupation" == 125, "Specialists in information a
                                ifelse(data1$"Mother's occupation" == 131, "Intermediate level science a
                                ifelse(data1$"Mother's occupation" == 132, "Technicians and professional
                                ifelse(data1$"Mother's occupation" == 134, "Intermediate level technicia
                                ifelse(data1$"Mother's occupation" == 141, "Office workers, secretaries
                                ifelse(data1$"Mother's occupation" == 143, "Data, accounting, statistica
                                ifelse(data1$"Mother's occupation" == 144, "Other administrative support
                                ifelse(data1$"Mother's occupation" == 151, "Personal service workers",
                                ifelse(data1$"Mother's occupation" == 152, "Sellers",
                                ifelse(data1$"Mother's occupation" == 153, "Personal care workers and the
                                ifelse(data1$"Mother's occupation" == 171, "Skilled construction workers
                                ifelse(data1$"Mother's occupation" == 173, "Skilled workers in printing,
                                ifelse(data1$"Mother's occupation" == 175, "Workers in food processing,
                                ifelse(data1$"Mother's occupation" == 191, "Cleaning workers",
                                ifelse(data1$"Mother's occupation" == 192, "Unskilled workers in agricul
                                ifelse(data1$"Mother's occupation" == 193, "Unskilled workers in extract
                                ifelse(data1$"Mother's occupation" == 194, "Meal preparation assistants"

 data1$"Father's occupation" <- ifelse(data1$"Father's occupation" == 0, "Student",
                                ifelse(data1$"Father's occupation" == 1, "Representatives of the Legisla
                                ifelse(data1$"Father's occupation" == 2, "Specialists in Intellectual and
                                ifelse(data1$"Father's occupation" == 3, "Intermediate Level Technicians
                                ifelse(data1$"Father's occupation" == 4, "Administrative staff",
                                ifelse(data1$"Father's occupation" == 5, "Personal Services, Security and
                                ifelse(data1$"Father's occupation" == 6, "Farmers and Skilled Workers in
                                ifelse(data1$"Father's occupation" == 7, "Skilled Workers in Industry, C
                                ifelse(data1$"Father's occupation" == 8, "Installation and Machine Opera
                                ifelse(data1$"Father's occupation" == 9, "Unskilled Workers",
                                ifelse(data1$"Father's occupation" == 10, "Armed Forces Professions 90 -
                                ifelse(data1$"Father's occupation" == 99, "(Blank)",
                                ifelse(data1$"Father's occupation" == 101, "Armed Forces Officers",
                                ifelse(data1$"Father's occupation" == 102, "Armed Forces Sergeants",
                                ifelse(data1$"Father's occupation" == 103, "Other Armed Forces personnel
                                ifelse(data1$"Father's occupation" == 112, "Directors of administrative
                                ifelse(data1$"Father's occupation" == 114, "Hotel, catering, trade and o
                                ifelse(data1$"Father's occupation" == 121, "Specialists in the physical
                                ifelse(data1$"Father's occupation" == 122, "Health professionals",
                                ifelse(data1$"Father's occupation" == 123, "Teachers",
                                ifelse(data1$"Father's occupation" == 124, "Specialists in finance, acco
                                ifelse(data1$"Father's occupation" == 131, "Intermediate level science a
                                ifelse(data1$"Father's occupation" == 132, "Technicians and professional
                                ifelse(data1$"Father's occupation" == 134, "Intermediate level technicia
                                ifelse(data1$"Father's occupation" == 135, "Information and communicatio
                                ifelse(data1$"Father's occupation" == 141, "Office workers, secretaries
                                ifelse(data1$"Father's occupation" == 143, "Data, accounting, statistica
                                ifelse(data1$"Father's occupation" == 144, "Other administrative support
                                ifelse(data1$"Father's occupation" == 151, "Personal service workers",
                                ifelse(data1$"Father's occupation" == 152, "Sellers",
                                ifelse(data1$"Father's occupation" == 153, "Personal care workers and the
                                ifelse(data1$"Father's occupation" == 154, "Protection and security serv
                                ifelse(data1$"Father's occupation" == 161, "Market-oriented farmers and
                                ifelse(data1$"Father's occupation" == 163, "Farmers, livestock keepers,
                                ifelse(data1$"Father's occupation" == 171, "Skilled construction workers
                                ifelse(data1$"Father's occupation" == 172, "Skilled workers in metallurg
                                ifelse(data1$"Father's occupation" == 174, "Skilled workers in electrici
                                ifelse(data1$"Father's occupation" == 175, "Workers in food processing,
                                ifelse(data1$"Father's occupation" == 181, "Fixed plant and machine oper
                                ifelse(data1$"Father's occupation" == 182, "Assembly workers",
                                ifelse(data1$"Father's occupation" == 183, "Vehicle drivers and mobile e
                                ifelse(data1$"Father's occupation" == 192, "Unskilled workers in agricul
                                ifelse(data1$"Father's occupation" == 193, "Unskilled workers in extract
                                ifelse(data1$"Father's occupation" == 194, "Meal preparation assistants"
                                ifelse(data1$"Father's occupation" == 195, "Street vendors (except food)

 data1$"Displaced" <- ifelse(data1$"Displaced" == 1, "yes", "no")

 data1$"Educational special needs" <- ifelse(data1$"Educational special needs" == 1, "yes", "no")

 data1$"Debtor" <- ifelse(data1$"Debtor" == 1, "yes", "no")

 data1$"Tuition fees up to date" <- ifelse(data1$"Tuition fees up to date" == 1, "yes", "no")

 data1$"Gender" <- ifelse(data1$"Gender" == 1, "Male", "Female")

 data1$"Scholarship holder" <- ifelse(data1$"Scholarship holder" == 1, "yes", "no")

 data1$"International" <- ifelse(data1$"International" == 1, "yes", "no")

 categorical_cols <- c("Marital status",
                       "Application mode",
                       "Application order",
                       "Course",
                       "Daytime/evening attendance",
                       "Previous qualification",
                       "Nationality",
                       "Mother's qualification",
                       "Father's qualification",
                       "Mother's occupation",
                       "Father's occupation",
                       "Displaced",
                       "Educational special needs",
                       "Debtor",
                       "Tuition fees up to date",
                       "Gender",
                       "Scholarship holder",
                       "International")

 library(ggplot2)
 library(scales)

 for (i in categorical_cols) {
   plot <- ggplot(data1, aes(x = .data[[i]])) +
     geom_bar(fill = "steelblue", color = "black") +
     labs(
       title = paste("Frequency of", i),
       x = i,
       y = "Count"
     ) +
     theme_minimal() +
     scale_x_discrete(
       labels = label_wrap(150)
     ) +
     coord_flip() + theme(
       axis.text = element_text(size = 6)
     )
   print(plot)
 }
```



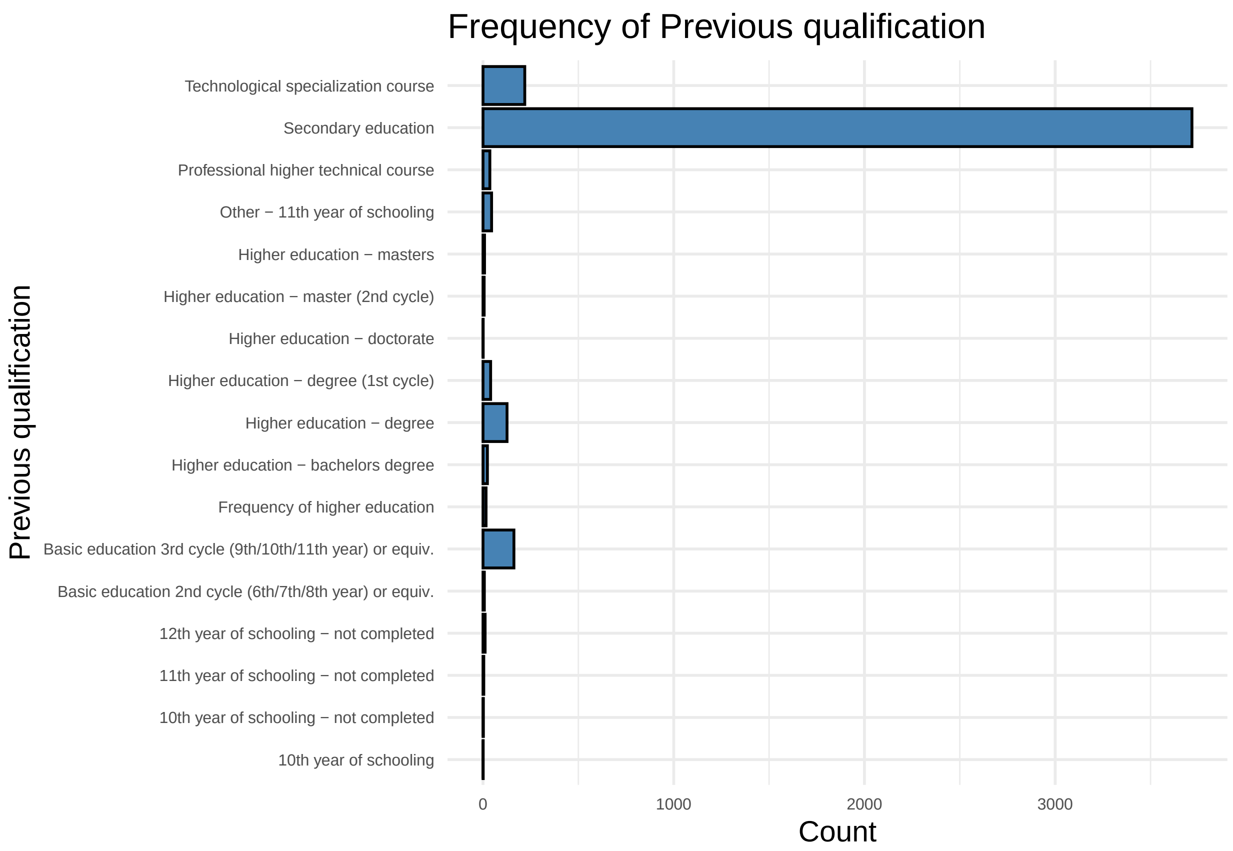

Frequency of Application mode
Application mode
Transfer
Technological specialization diploma holders
Short cycle diploma holders
Over 23 years old
Ordinance No. 854−B/99
Ordinance No. 612/93
Ordinance No. 533−A/99, item b3 (Other Institution)
Ordinance No. 533−A/99, item b2) (Different Plan)
International student (bachelor)
Holders of other higher courses
Change of institution/course (International)
Change of institution/course
Change of course
3rd phase − general contingent
2nd phase − general contingent
1st phase − special contingent (Madeira Island)
1st phase − special contingent (Azores Island)
1st phase − general contingent
0
500
1000
1500
Count

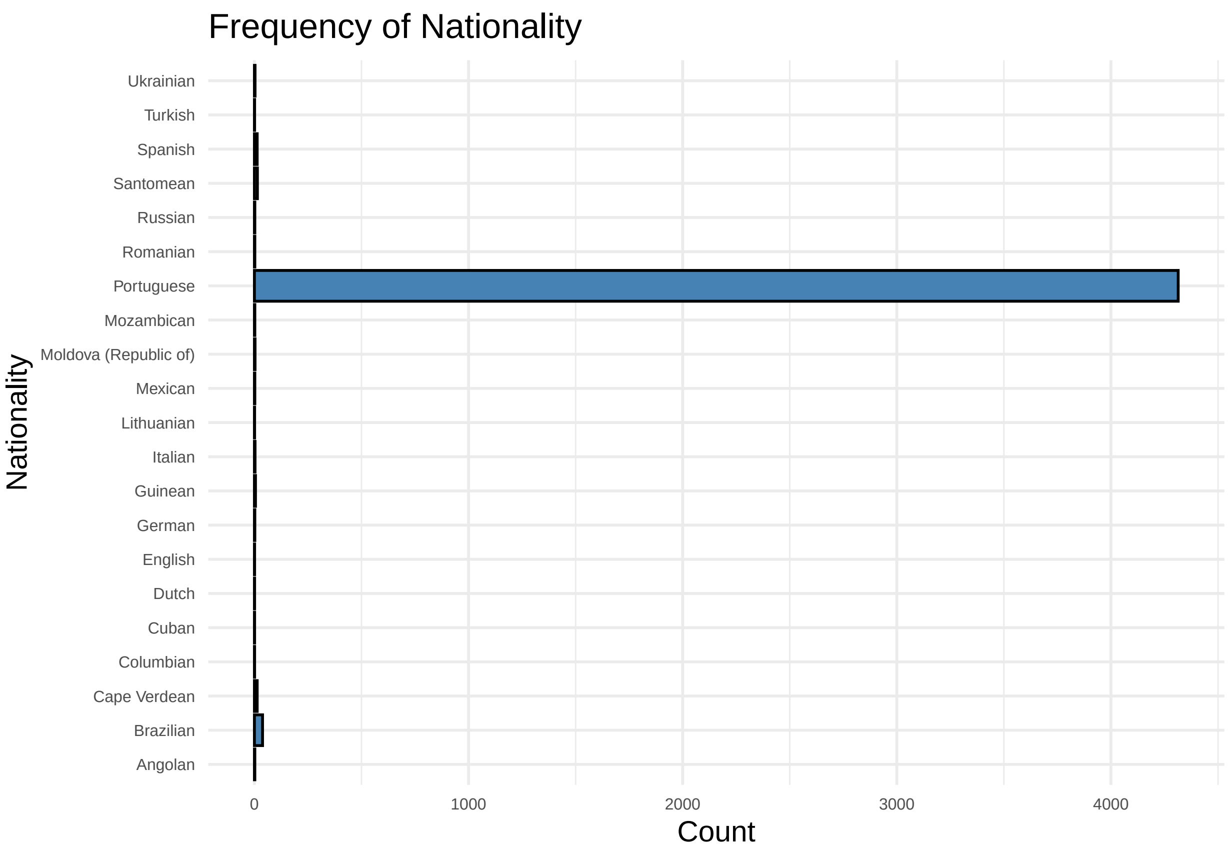

Frequency of Application order
Application order
Last choice
7th choice
6th choice
5th choice
4th choice
3rd choice
2nd choice
1st choice
0
1000
2000
3000
Count

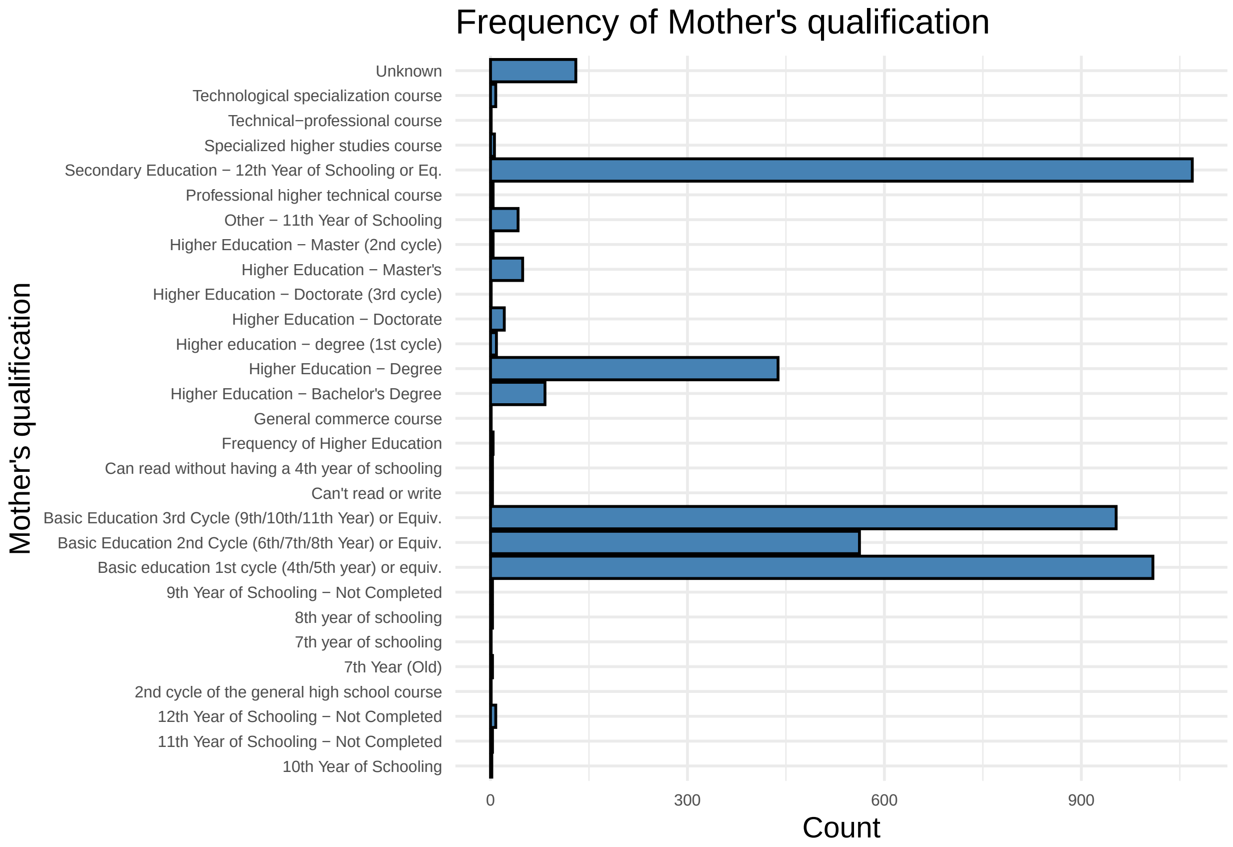

Frequency of Course
Course
Count
Veterinary Nursing
Tourism
Social Service (evening attendance)
Social Service
Oral Hygiene
Nursing
Management (evening attendance)
Management
Journalism and Communication
Informatics Engineering
Equinculture
Communication Design
Biofuel Production Technologies
Basic Education
Animation and Multimedia Design
Agronomy
Advertising and Marketing Management
0
200
400
600
800

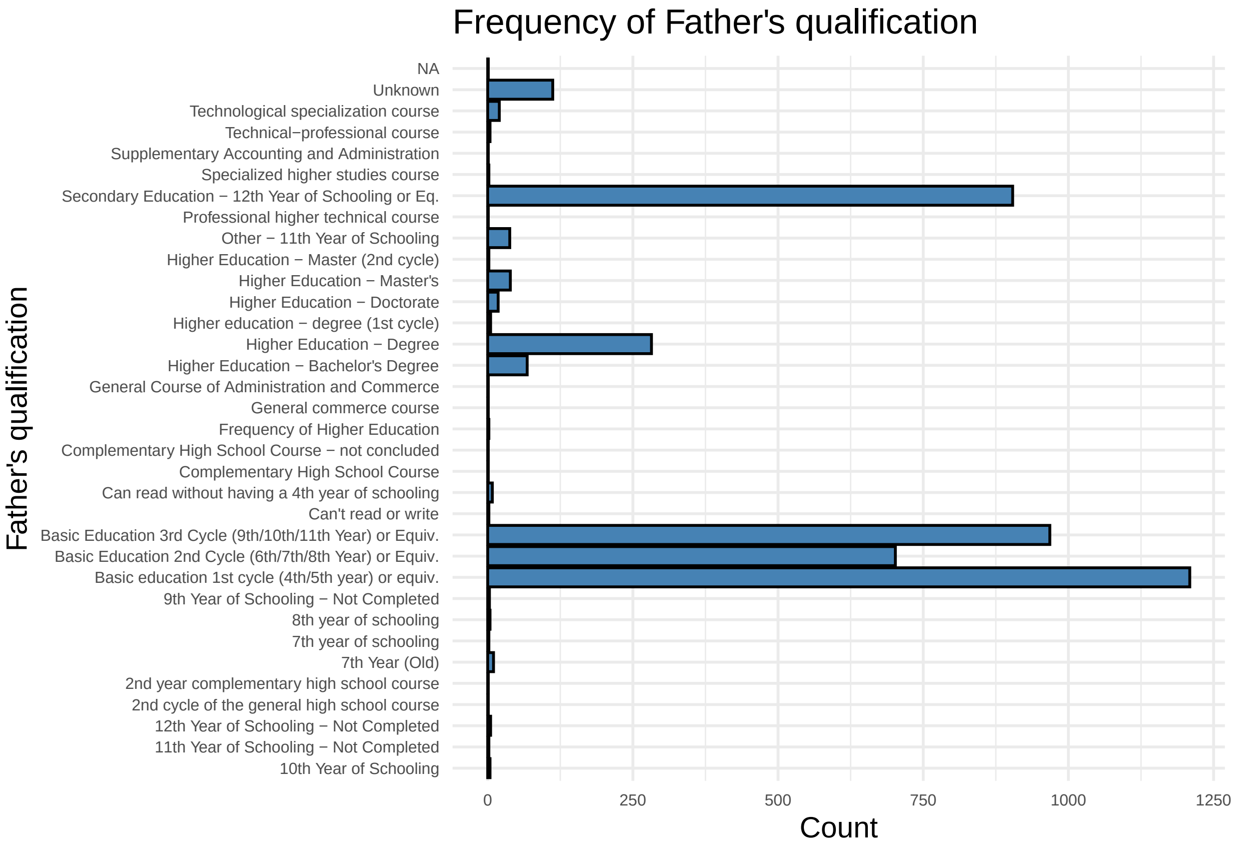


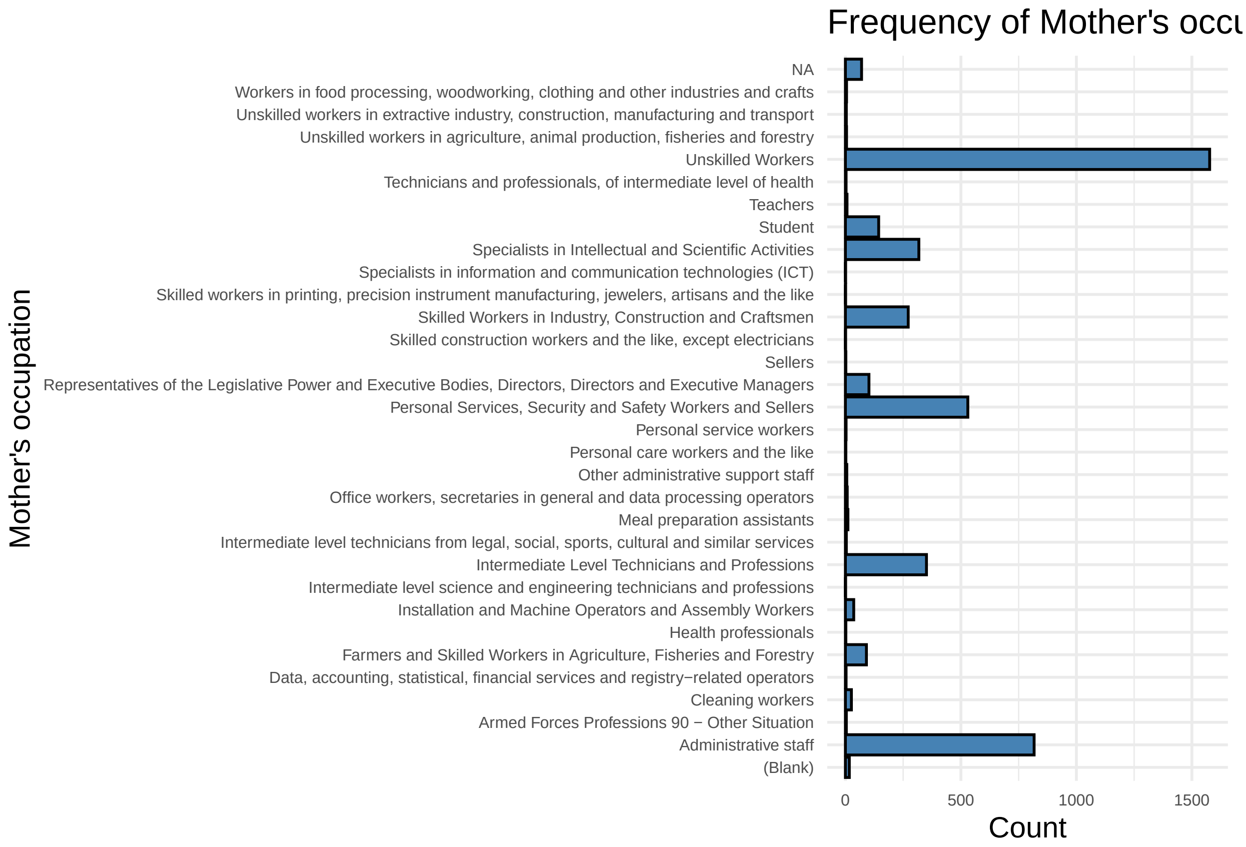
Frequency of Daytime/evening attendance
Daytime/evening attendance
evening
daytime
0
1000
2000
3000
4000
Count

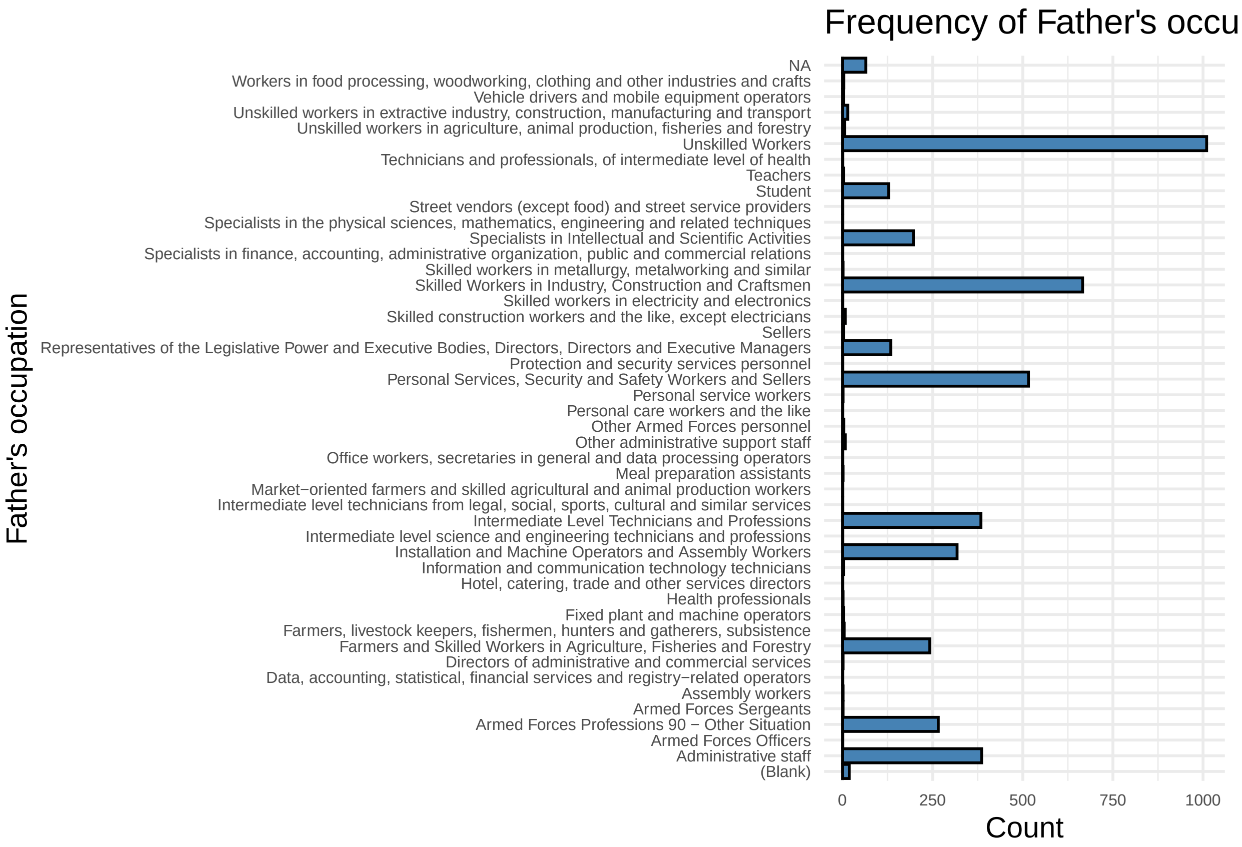

Frequency of Previous qualification
Previous qualification
Technological specialization course
Secondary education
Professional higher technical course
Other − 11th year of schooling
Higher education − masters
Higher education − master (2nd cycle)
Higher education − doctorate
Higher education − degree (1st cycle)
Higher education − degree
Higher education − bachelors degree
Frequency of higher education
Basic education 3rd cycle (9th/10th/11th year) or equiv.
Basic education 2nd cycle (6th/7th/8th year) or equiv.
12th year of schooling − not completed
11th year of schooling − not completed
10th year of schooling − not completed
10th year of schooling
0
1000
2000
3000
Count

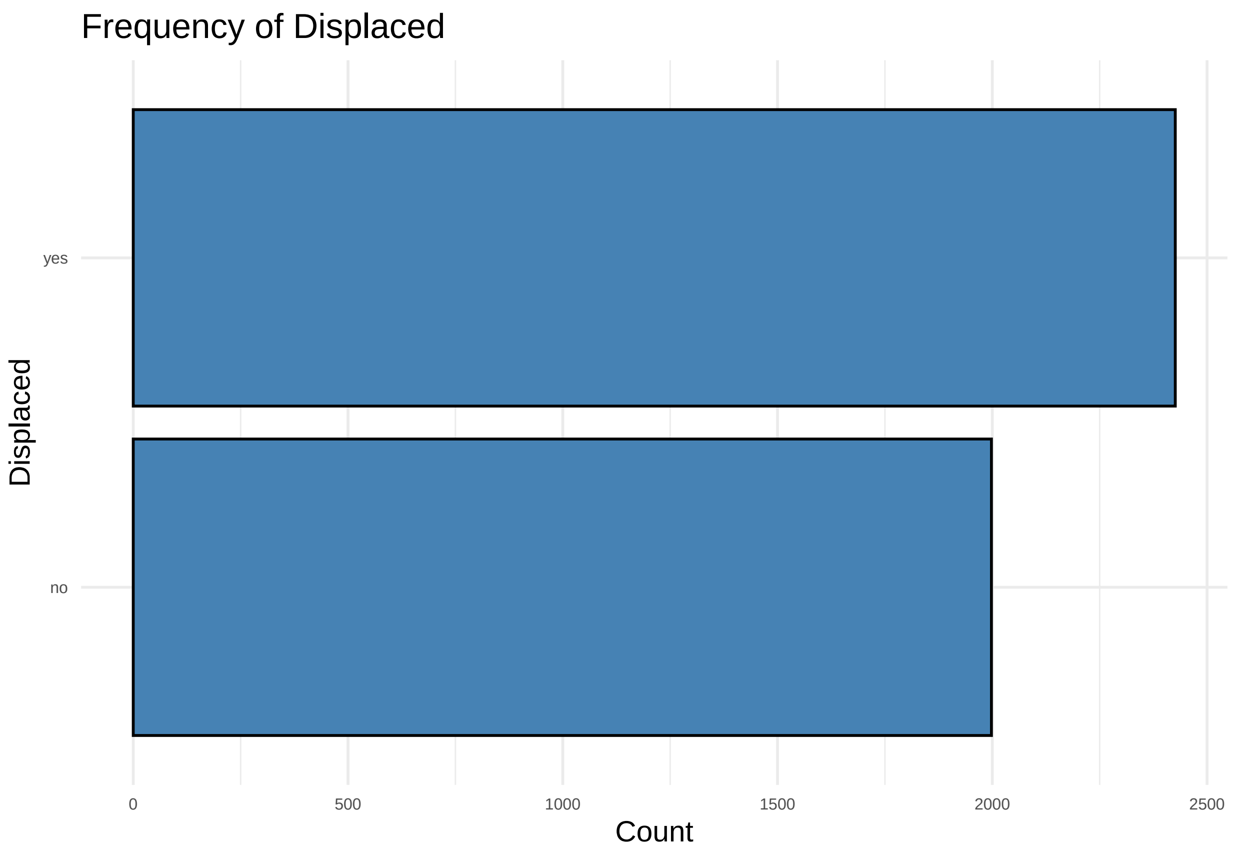

Frequency of Nationality
Nationality
Ukrainian
Turkish
Spanish
Santomean
Russian
Romanian
Portuguese
Mozambican
Moldova (Republic of)
Mexican
Lithuanian
Italian
Guinean
German
English
Dutch
Cuban
Columbian
Cape Verdean
Brazilian
Angolan
0
1000
2000
3000
4000
Count

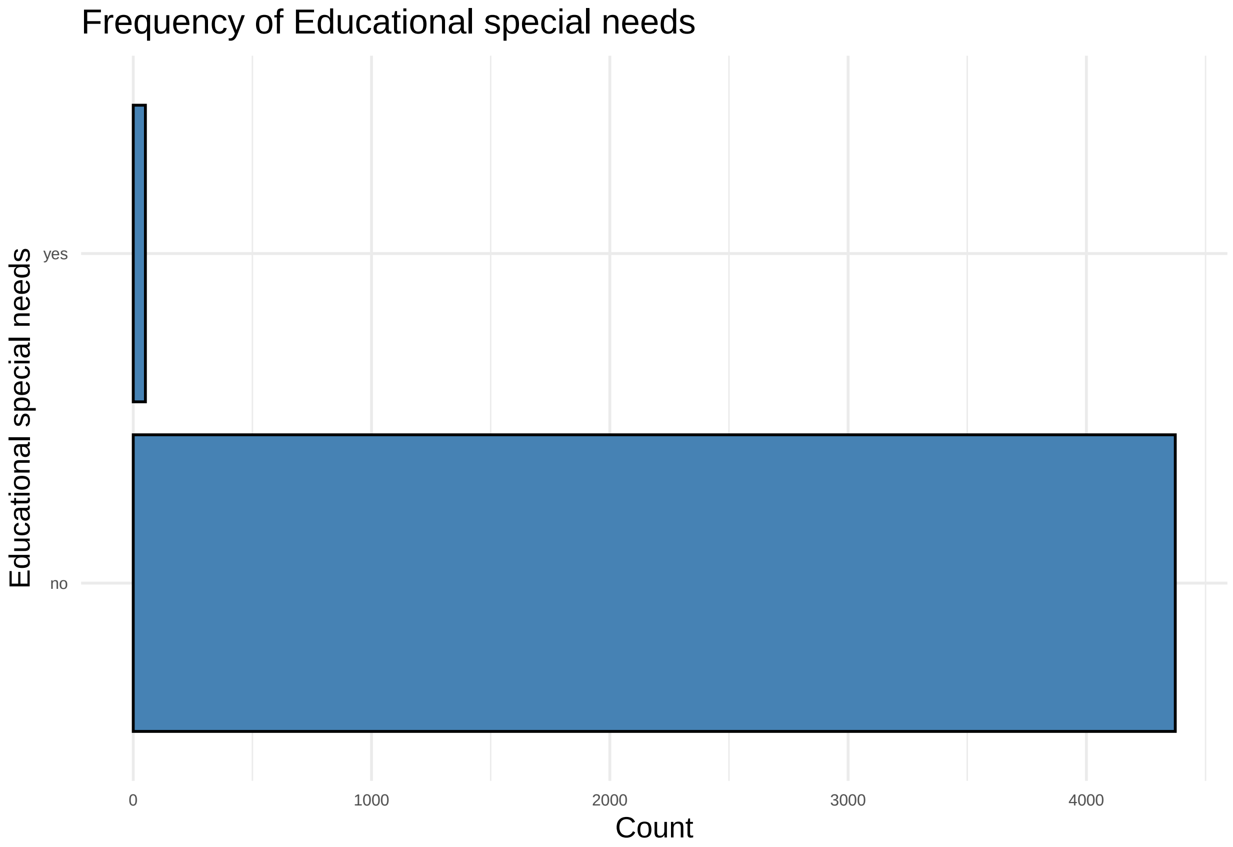

Frequency of Mother's qualification
Mother's qualification
Unknown
Technological specialization course
Technical−professional course
Specialized higher studies course
Secondary Education − 12th Year of Schooling or Eq.
Professional higher technical course
Other − 11th Year of Schooling
Higher Education − Master (2nd cycle)
Higher Education − Master's
Higher Education − Doctorate (3rd cycle)
Higher Education − Doctorate
Higher education − degree (1st cycle)
Higher Education − Degree
Higher Education − Bachelor's Degree
General commerce course
Frequency of Higher Education
Can read without having a 4th year of schooling
Can't read or write
Basic Education 3rd Cycle (9th/10th/11th Year) or Equiv.
Basic Education 2nd Cycle (6th/7th/8th Year) or Equiv.
Basic education 1st cycle (4th/5th year) or equiv.
9th Year of Schooling − Not Completed
8th year of schooling
7th year of schooling
7th Year (Old)
2nd cycle of the general high school course
12th Year of Schooling − Not Completed
11th Year of Schooling − Not Completed
10th Year of Schooling
0
300
600
900
Count

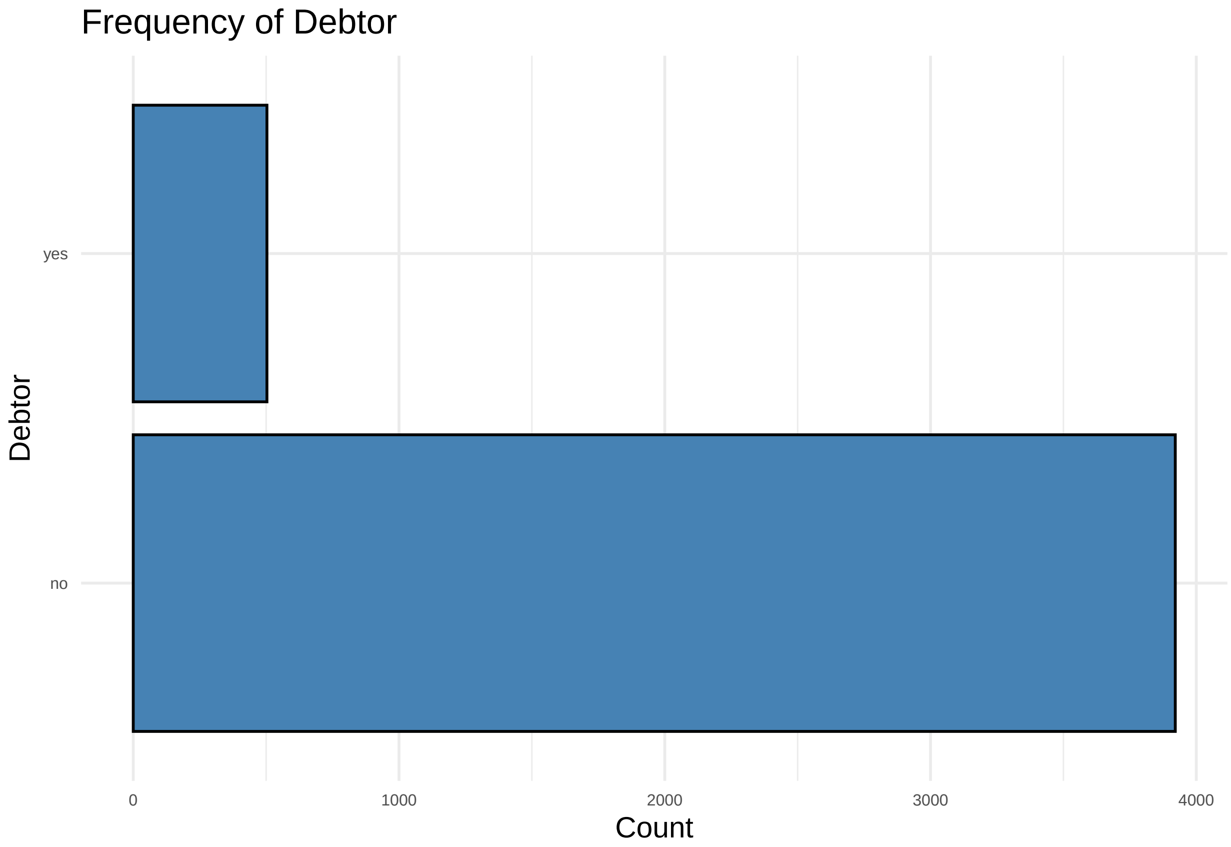

Frequency of Father's qualification
Father's qualification
NA
Unknown
Technological specialization course
Technical−professional course
Supplementary Accounting and Administration
Specialized higher studies course
Secondary Education − 12th Year of Schooling or Eq.
Professional higher technical course
Other − 11th Year of Schooling
Higher Education − Master (2nd cycle)
Higher Education − Master's
Higher Education − Doctorate
Higher education − degree (1st cycle)
Higher Education − Degree
Higher Education − Bachelor's Degree
General Course of Administration and Commerce
General commerce course
Frequency of Higher Education
Complementary High School Course − not concluded
Complementary High School Course
Can read without having a 4th year of schooling
Can't read or write
Basic Education 3rd Cycle (9th/10th/11th Year) or Equiv.
Basic Education 2nd Cycle (6th/7th/8th Year) or Equiv.
Basic education 1st cycle (4th/5th year) or equiv.
9th Year of Schooling − Not Completed
8th year of schooling
7th year of schooling
7th Year (Old)
2nd year complementary high school course
2nd cycle of the general high school course
12th Year of Schooling − Not Completed
11th Year of Schooling − Not Completed
10th Year of Schooling
0
250
500
750
1000
1250
Count

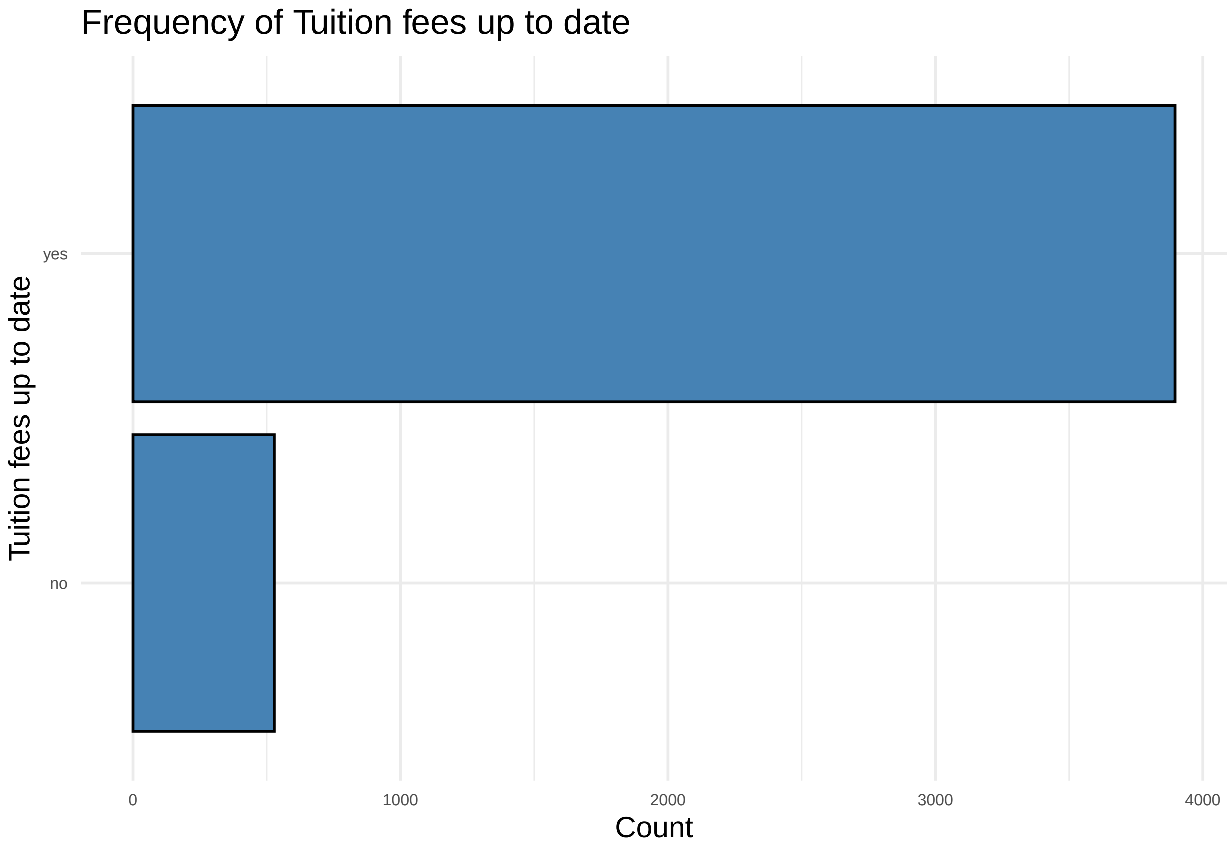

Frequency of Mother's occu
Mother's occupation
NA
Workers in food processing, woodworking, clothing and other industries and crafts
Unskilled workers in extractive industry, construction, manufacturing and transport
Unskilled workers in agriculture, animal production, fisheries and forestry
Unskilled Workers
Technicians and professionals, of intermediate level of health
Teachers
Student
Specialists in Intellectual and Scientific Activities
Specialists in information and communication technologies (ICT)
Skilled workers in printing, precision instrument manufacturing, jewelers, artisans and the like
Skilled Workers in Industry, Construction and Craftsmen
Skilled construction workers and the like, except electricians
Sellers
Representatives of the Legislative Power and Executive Bodies, Directors, Directors and Executive Managers
Personal Services, Security and Safety Workers and Sellers
Personal service workers
Personal care workers and the like
Other administrative support staff
Office workers, secretaries in general and data processing operators
Meal preparation assistants
Intermediate level technicians from legal, social, sports, cultural and similar services
Intermediate Level Technicians and Professions
Intermediate level science and engineering technicians and professions
Installation and Machine Operators and Assembly Workers
Health professionals
Farmers and Skilled Workers in Agriculture, Fisheries and Forestry
Data, accounting, statistical, financial services and registry−related operators
Cleaning workers
Armed Forces Professions 90 − Other Situation
Administrative staff
(Blank)
0
500
1000
1500
Count

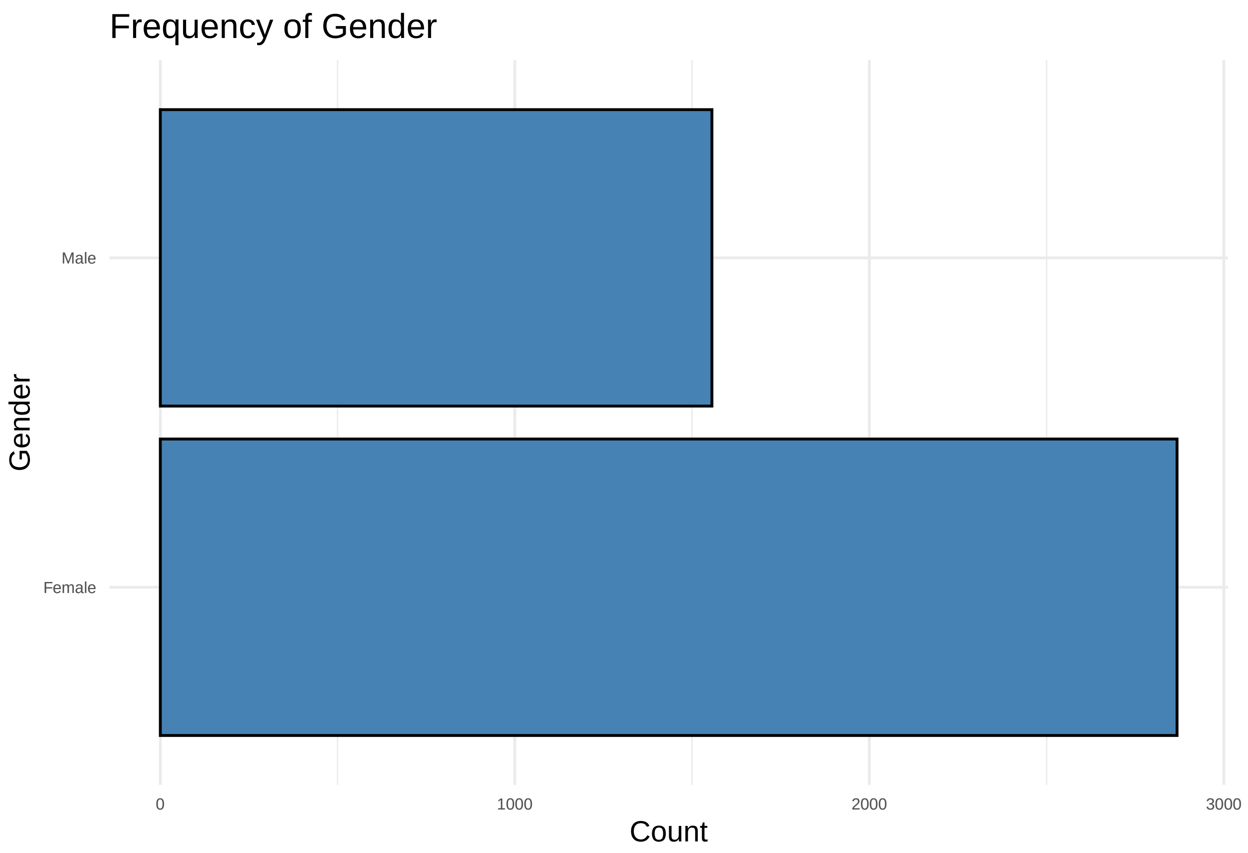

Frequency of Father's occu
Father's occupation
NA
Workers in food processing, woodworking, clothing and other industries and crafts
Vehicle drivers and mobile equipment operators
Unskilled workers in extractive industry, construction, manufacturing and transport
Unskilled workers in agriculture, animal production, fisheries and forestry
Unskilled Workers
Technicians and professionals, of intermediate level of health
Teachers
Student
Street vendors (except food) and street service providers
Specialists in the physical sciences, mathematics, engineering and related techniques
Specialists in Intellectual and Scientific Activities
Specialists in finance, accounting, administrative organization, public and commercial relations
Skilled workers in metallurgy, metalworking and similar
Skilled Workers in Industry, Construction and Craftsmen
Skilled workers in electricity and electronics
Skilled construction workers and the like, except electricians
Sellers
Representatives of the Legislative Power and Executive Bodies, Directors, Directors and Executive Managers
Protection and security services personnel
Personal Services, Security and Safety Workers and Sellers
Personal service workers
Personal care workers and the like
Other Armed Forces personnel
Other administrative support staff
Office workers, secretaries in general and data processing operators
Meal preparation assistants
Market−oriented farmers and skilled agricultural and animal production workers
Intermediate level technicians from legal, social, sports, cultural and similar services
Intermediate Level Technicians and Professions
Intermediate level science and engineering technicians and professions
Installation and Machine Operators and Assembly Workers
Information and communication technology technicians
Hotel, catering, trade and other services directors
Health professionals
Fixed plant and machine operators
Farmers, livestock keepers, fishermen, hunters and gatherers, subsistence
Farmers and Skilled Workers in Agriculture, Fisheries and Forestry
Directors of administrative and commercial services
Data, accounting, statistical, financial services and registry−related operators
Assembly workers
Armed Forces Sergeants
Armed Forces Professions 90 − Other Situation
Armed Forces Officers
Administrative staff
(Blank)
0
250
500
750
1000
Count

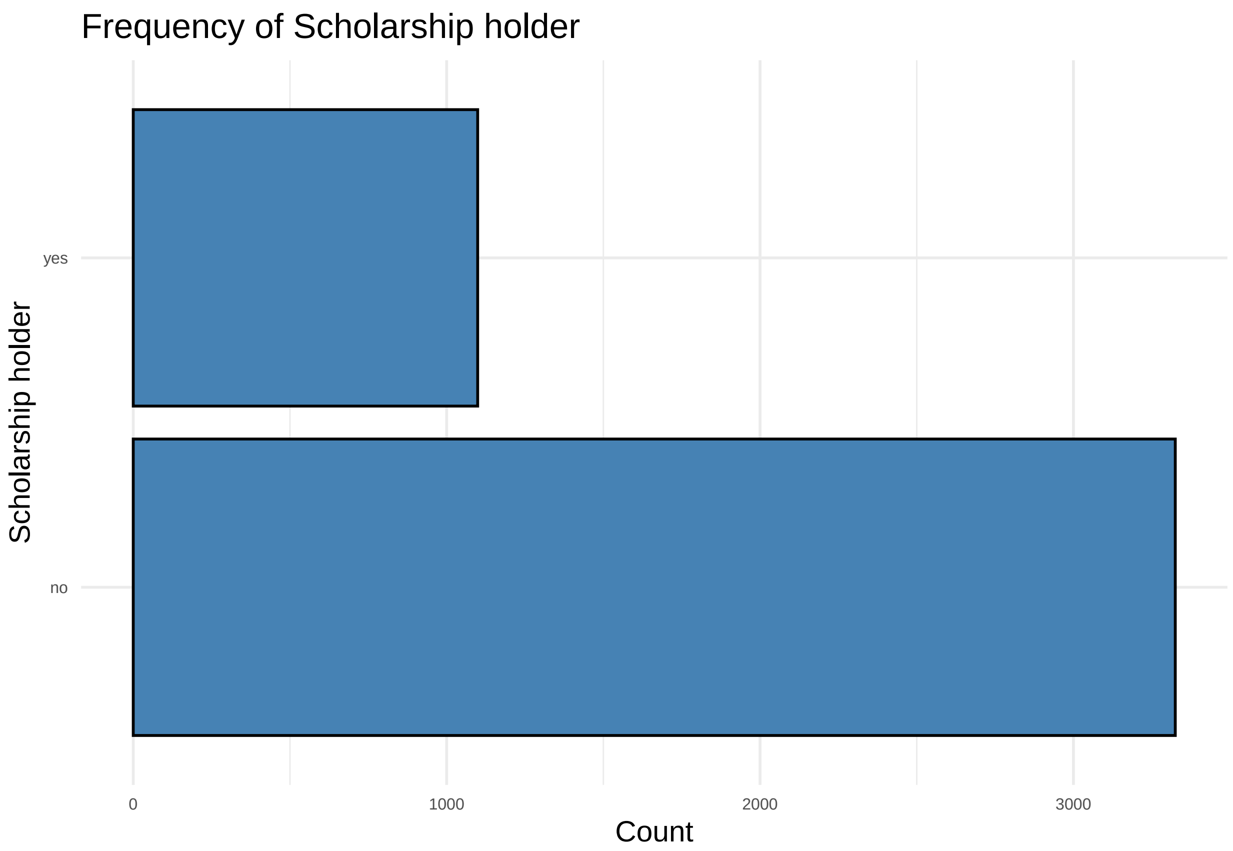

Frequency of Displaced
Displaced
yes
no
Count
0
500
1000
1500
2000
2500

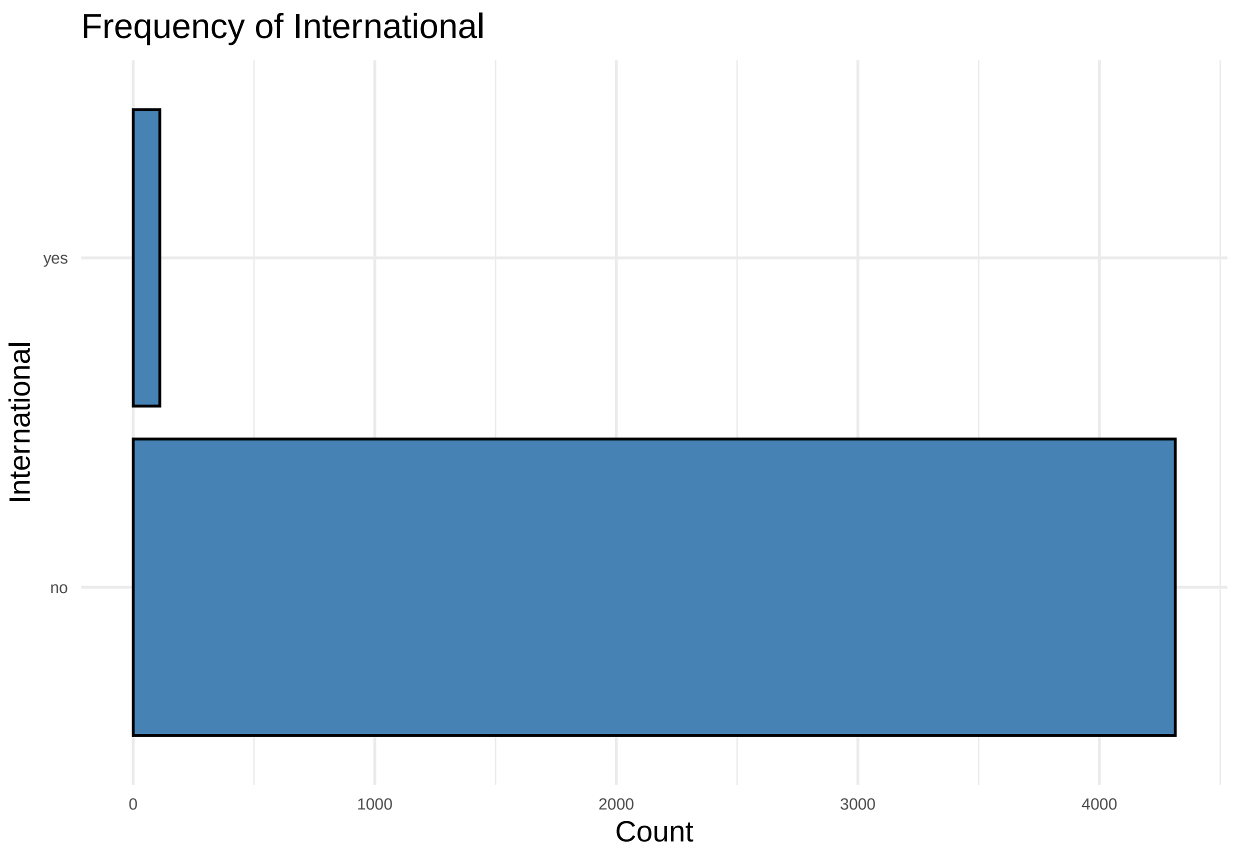

Frequency of Educational special needs
Educational special needs
yes
no
0
1000
2000
3000
4000
Count


Frequency of Debtor
Debtor
yes
no
0
1000
2000
3000
4000
Count

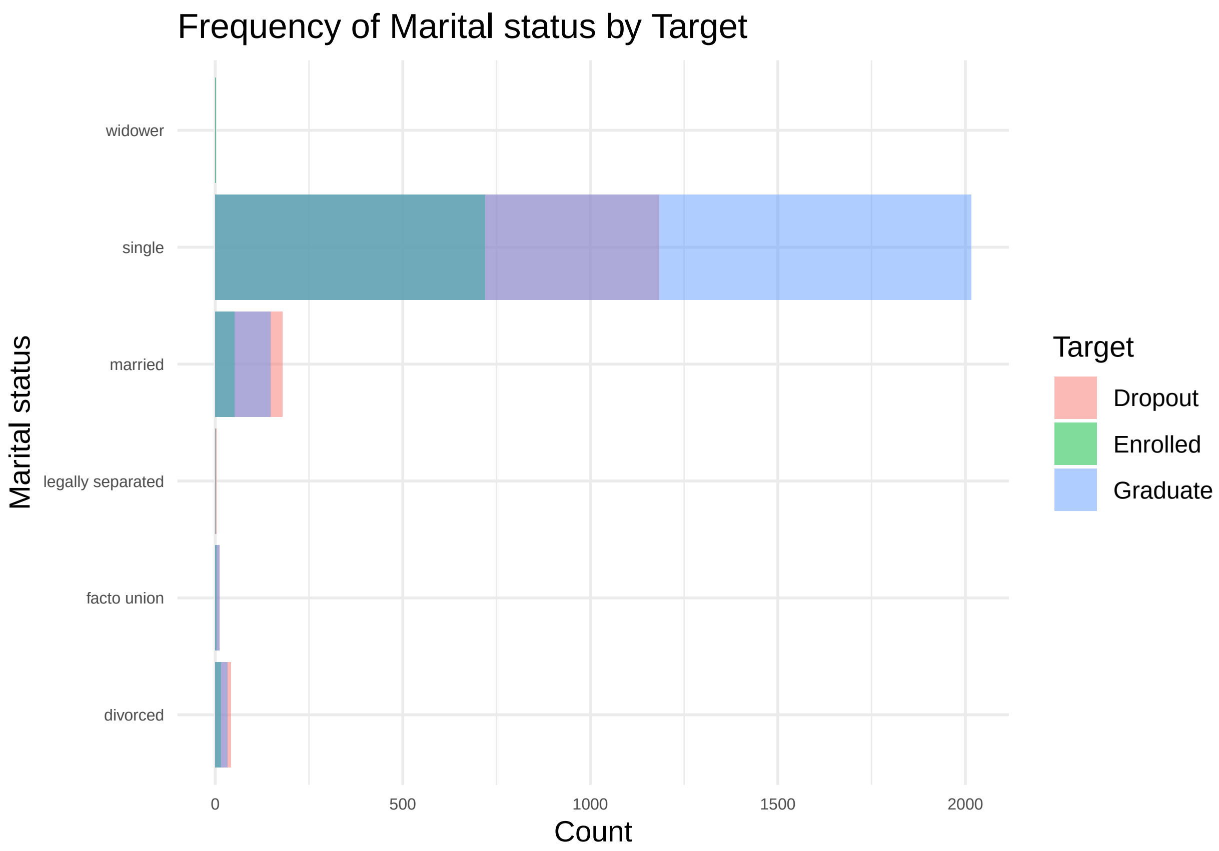

Frequency of Tuition fees up to date
Tuition fees up to date
yes
no
0
1000
2000
3000
4000
Count

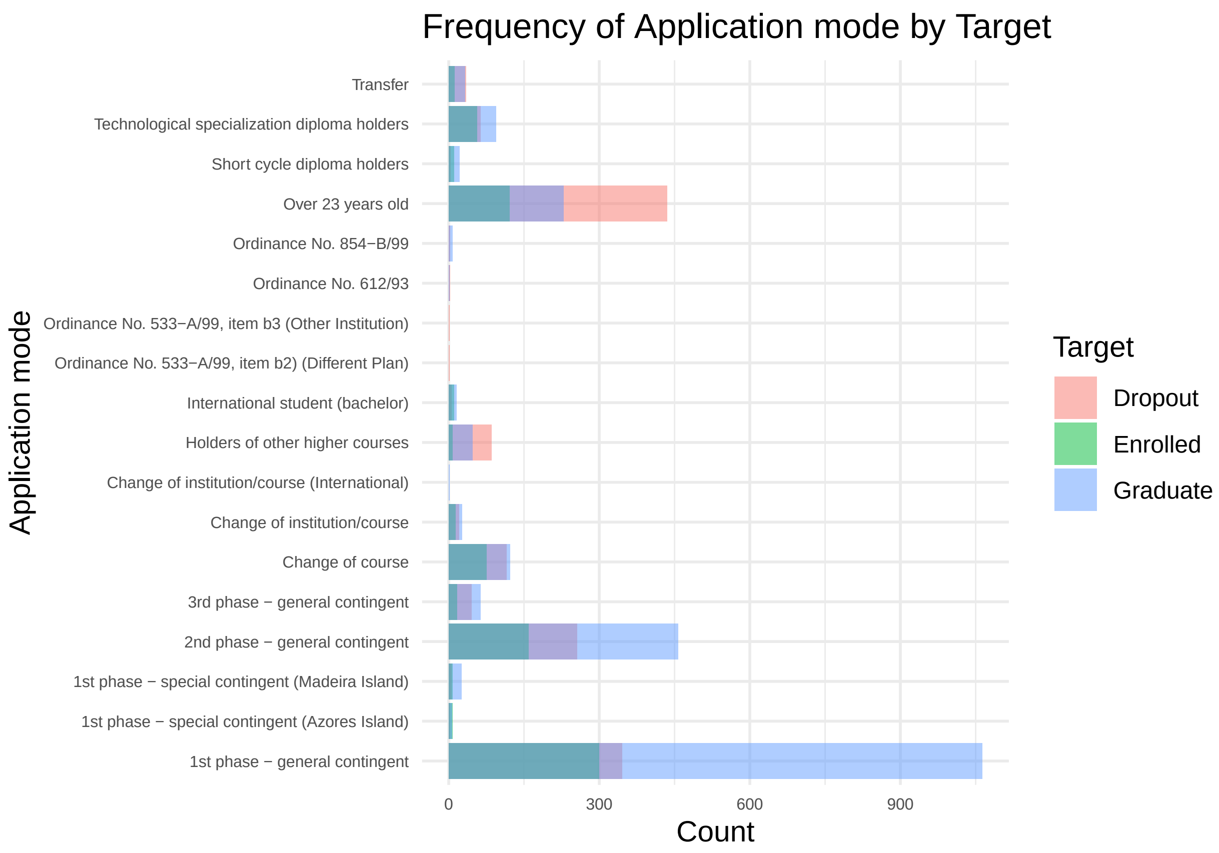

Frequency of Gender
Gender
Male
Female
Count
0
1000
2000
3000

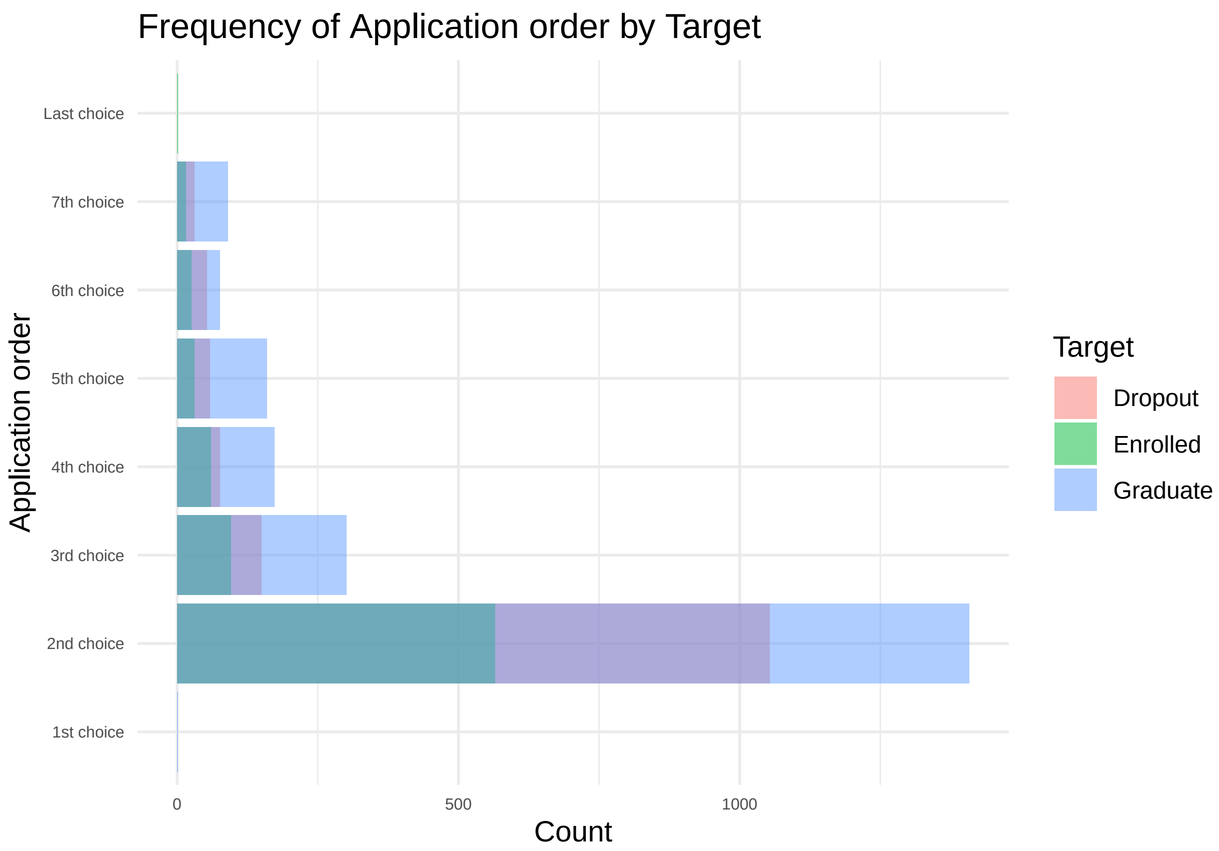

Frequency of Scholarship holder
Scholarship holder
yes
no
0
1000
2000
3000
Count

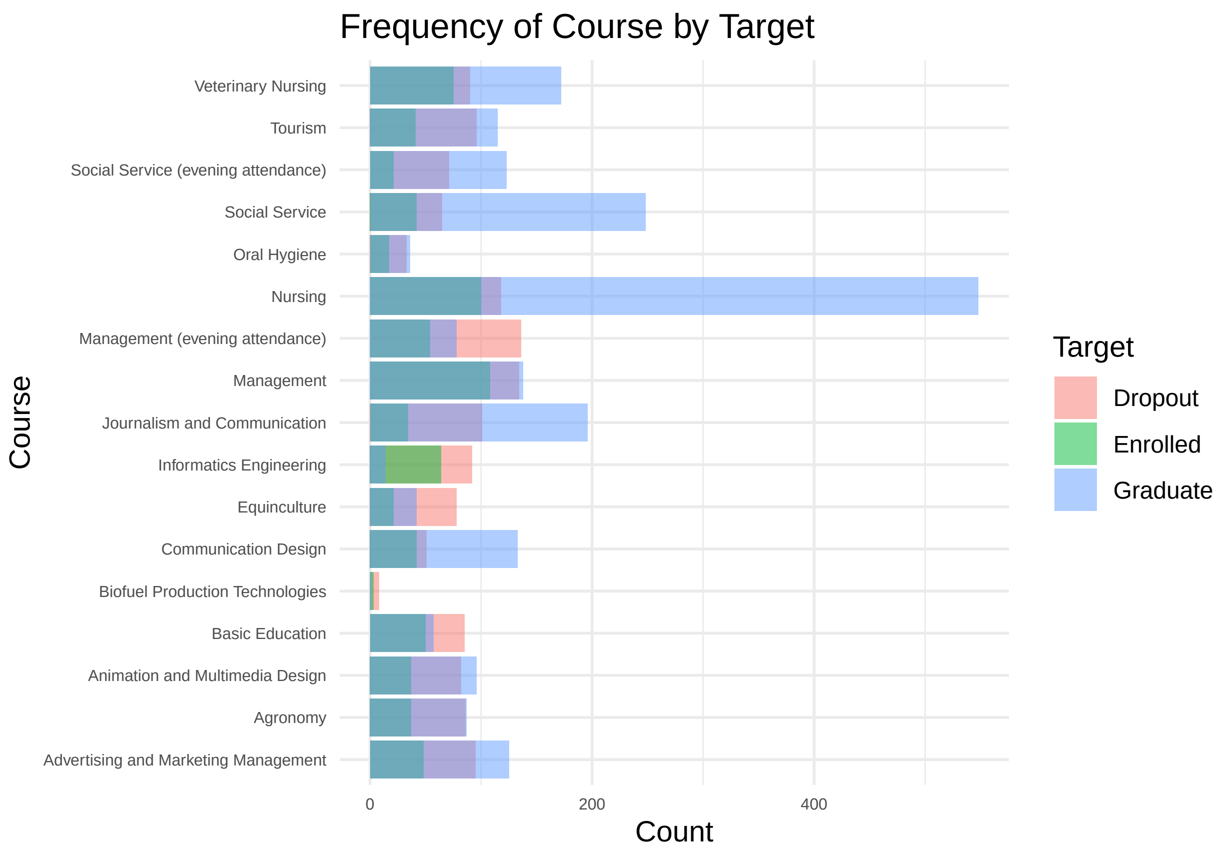


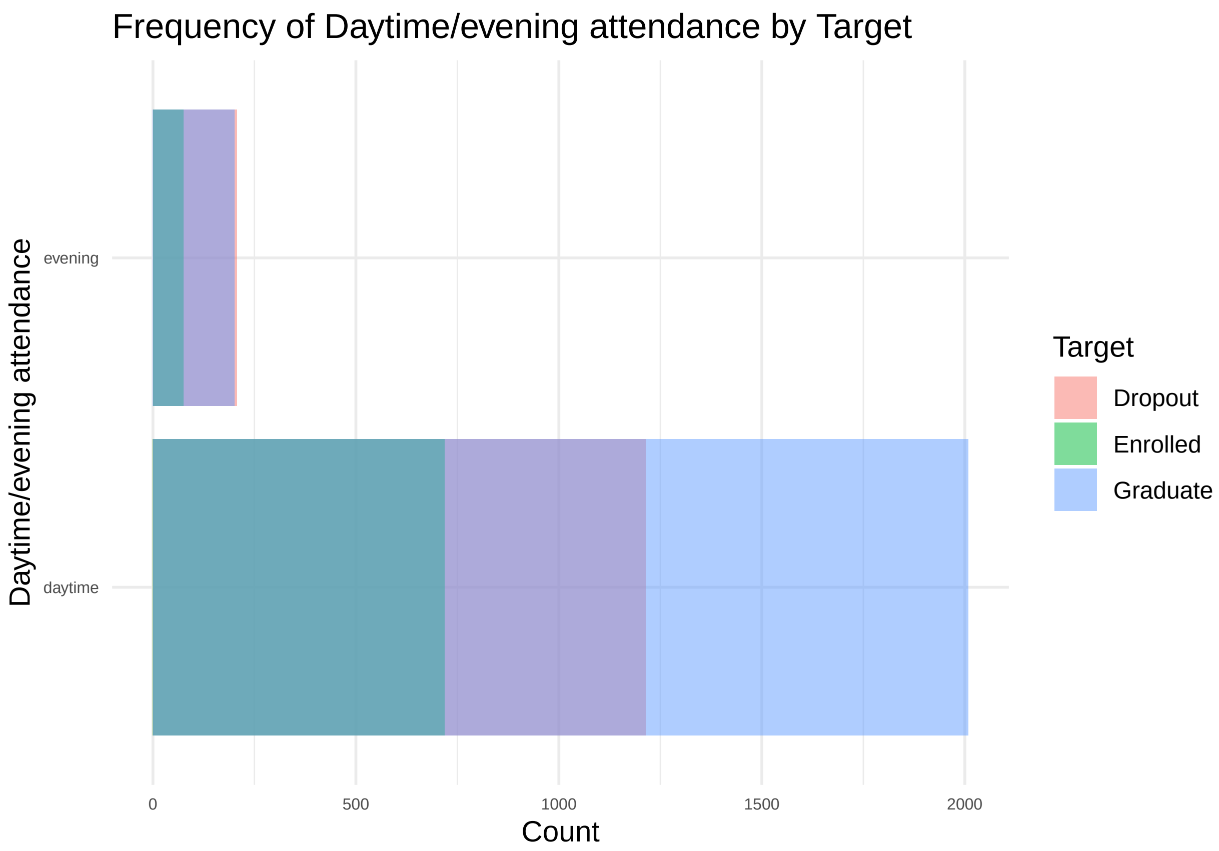
```r
for (i in categorical_cols) {
  plot <- ggplot(data1, aes(x = .data[[i]], fill = Target)) +
    geom_bar(position = "identity", alpha=0.5) +
    labs(
      title = paste("Frequency of", i, "by Target"),
      x = i,
      y = "Count"
    ) +
    theme_minimal() +
    scale_x_discrete(
      labels = label_wrap(150)
    ) +
    coord_flip() + theme(
      axis.text = element_text(size = 6)
    )
  print(plot)
}
```

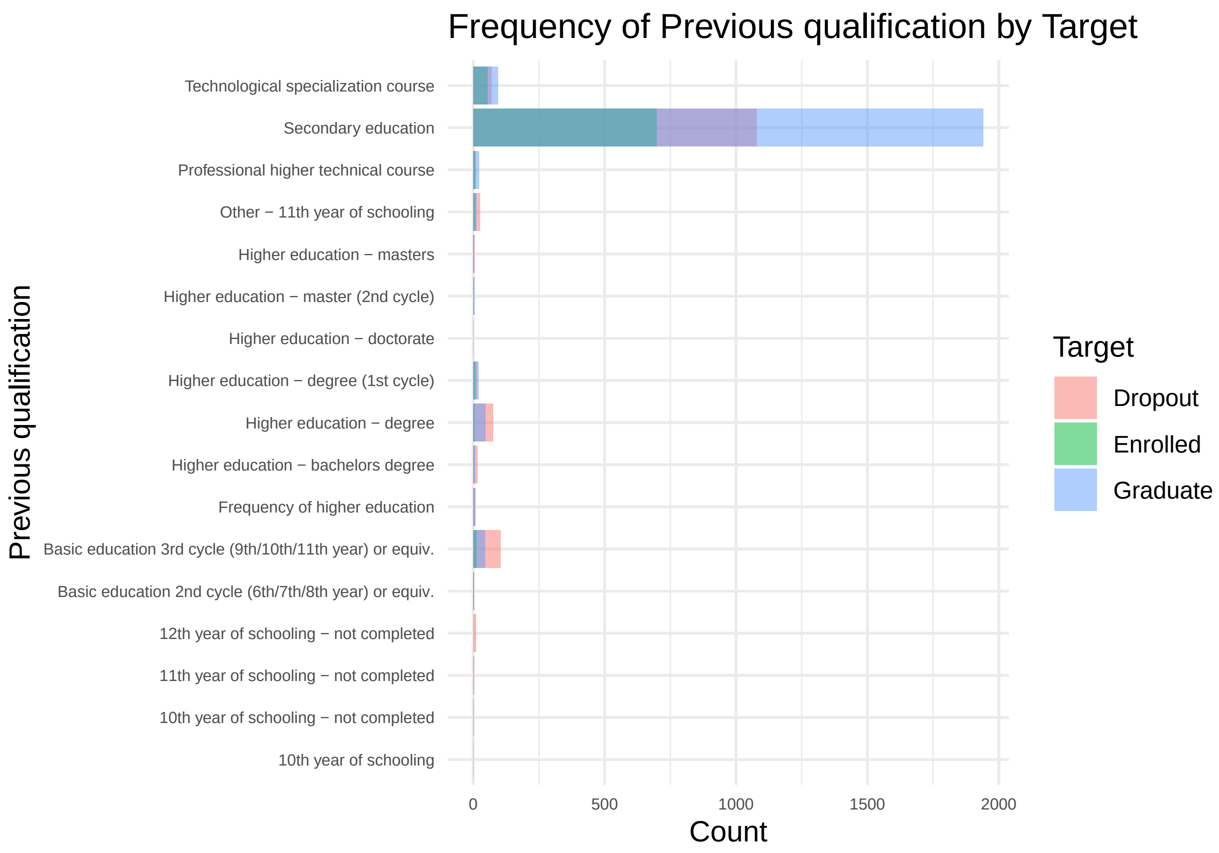

Frequency of Marital status by Target
Marital status
widower
single
married
legally separated
facto union
divorced
0
500
1000
1500
2000
Count
Target
Dropout
Enrolled
Graduate

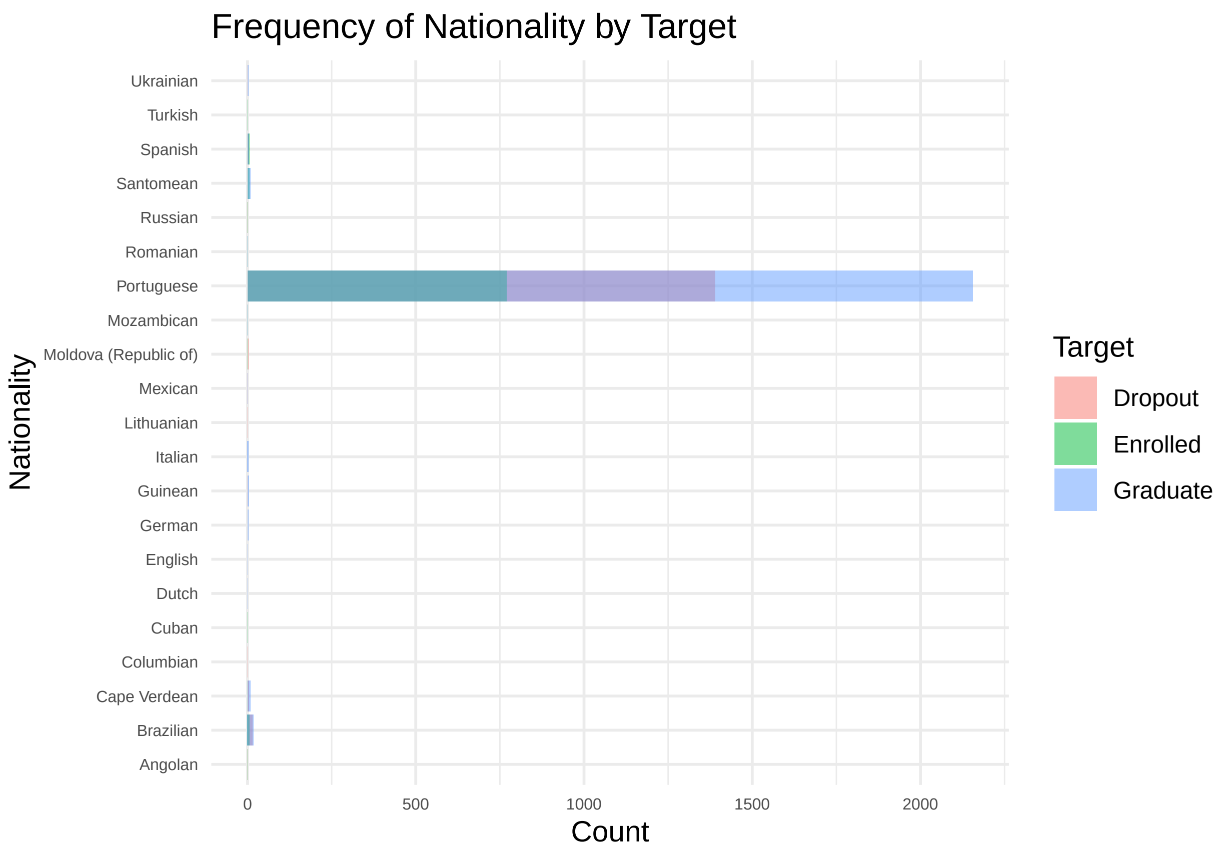

Frequency of Application mode by Target
Application mode
Transfer
Technological specialization diploma holders
Short cycle diploma holders
Over 23 years old
Ordinance No. 854−B/99
Ordinance No. 612/93
Ordinance No. 533−A/99, item b3 (Other Institution)
Ordinance No. 533−A/99, item b2) (Different Plan)
International student (bachelor)
Holders of other higher courses
Change of institution/course (International)
Change of institution/course
Change of course
3rd phase − general contingent
2nd phase − general contingent
1st phase − special contingent (Madeira Island)
1st phase − special contingent (Azores Island)
1st phase − general contingent
0
300
600
900
Count
Target
Dropout
Enrolled
Graduate

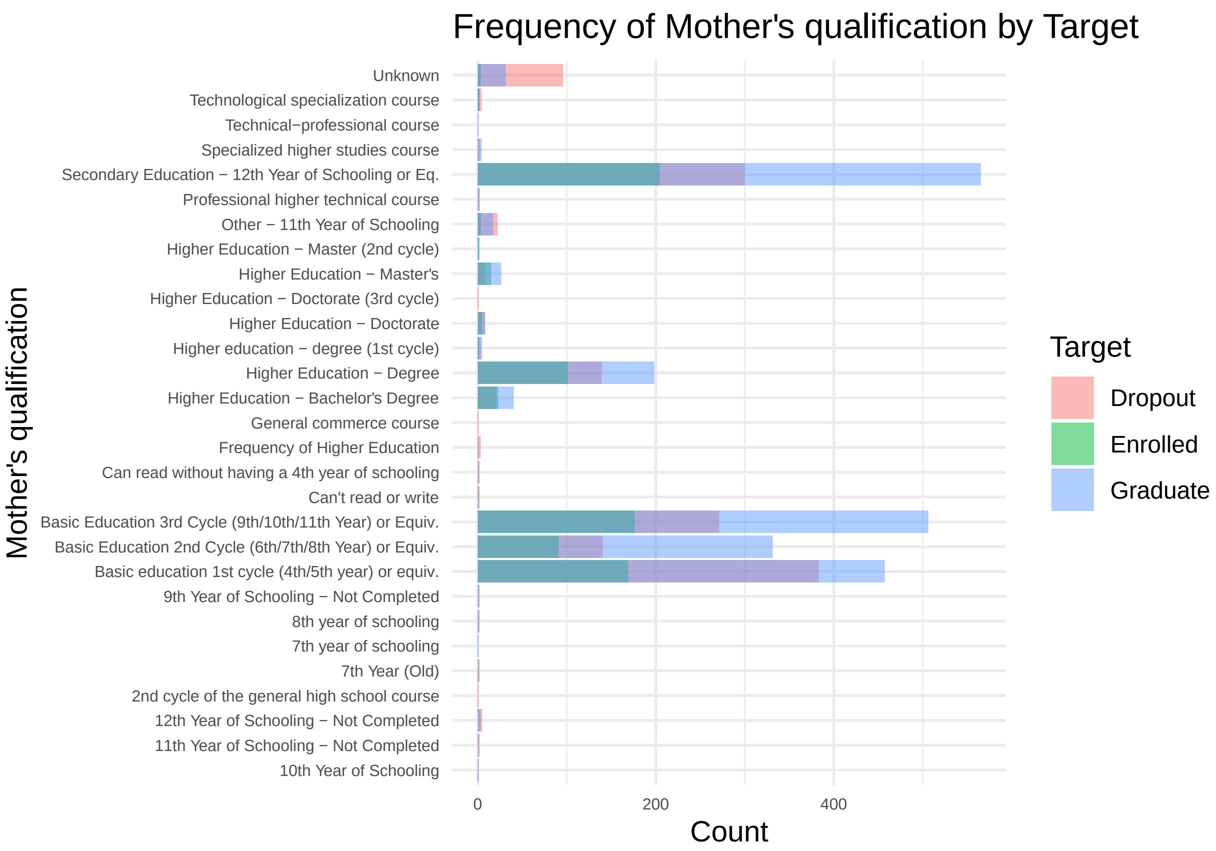

Frequency of Application order by Target
Application order
Last choice
7th choice
6th choice
5th choice
4th choice
3rd choice
2nd choice
1st choice
0
500
1000
Count
Target
Dropout
Enrolled
Graduate

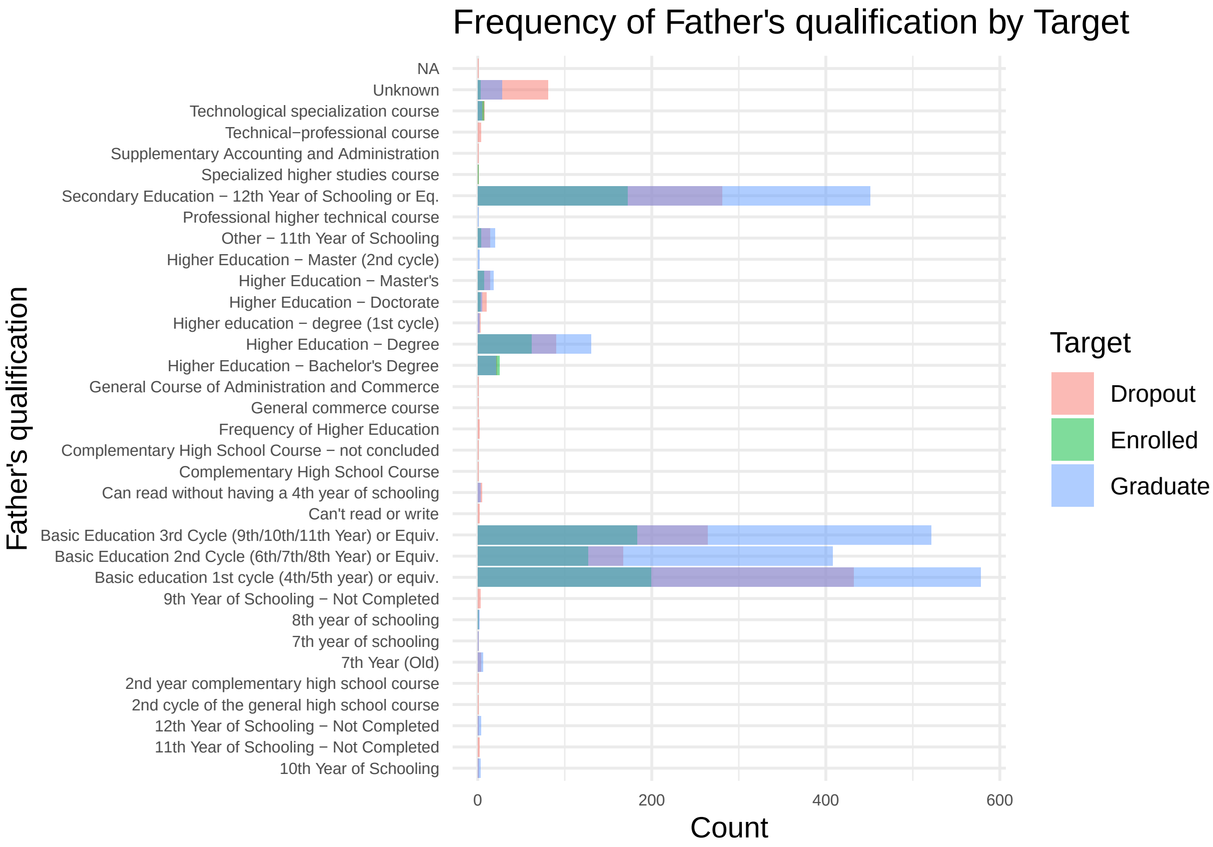

Frequency of Course by Target
Course
Veterinary Nursing
Tourism
Social Service (evening attendance)
Social Service
Oral Hygiene
Nursing
Management (evening attendance)
Management
Journalism and Communication
Informatics Engineering
Equinculture
Communication Design
Biofuel Production Technologies
Basic Education
Animation and Multimedia Design
Agronomy
Advertising and Marketing Management
0
200
400
Count
Target
Dropout
Enrolled
Graduate

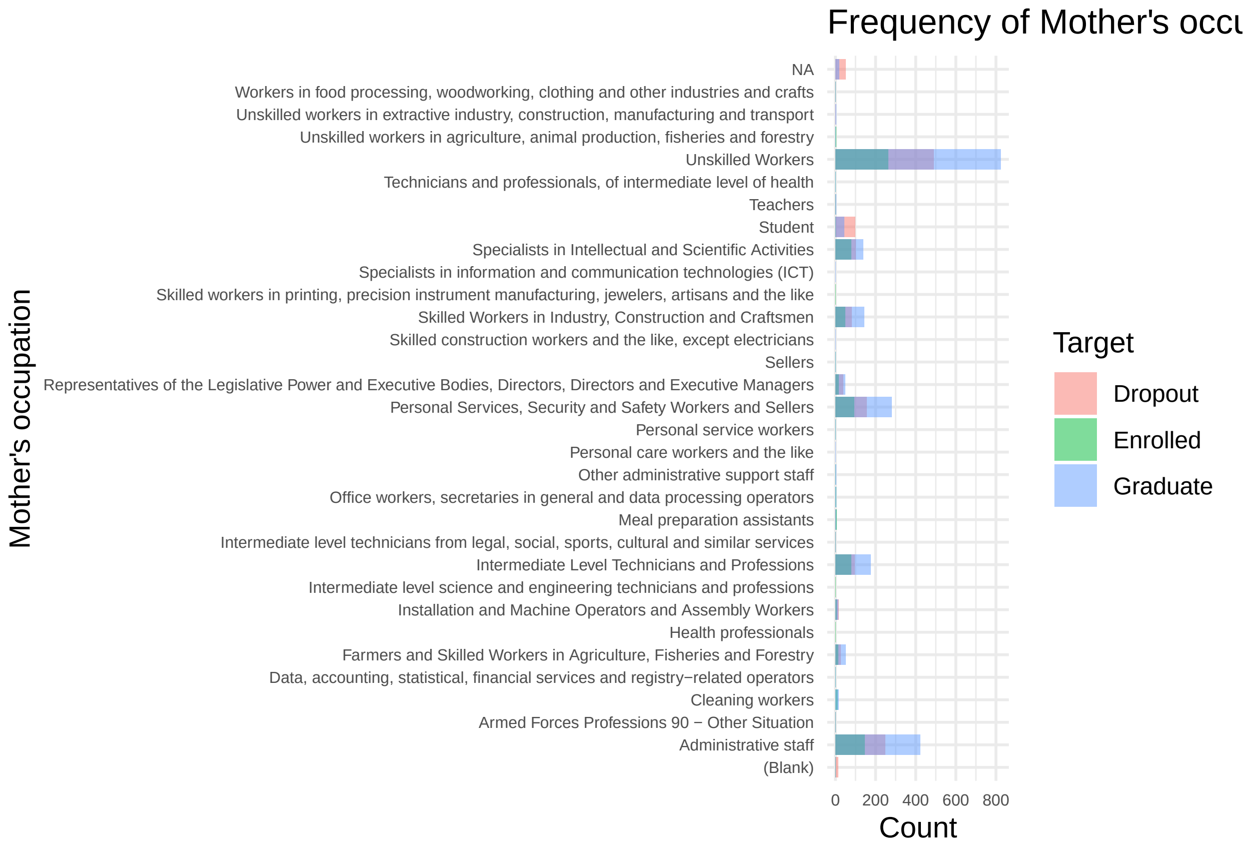

Frequency of Daytime/evening attendance by Target
Daytime/evening attendance
evening
daytime
0
500
1000
1500
2000
Count
Target
Dropout
Enrolled
Graduate

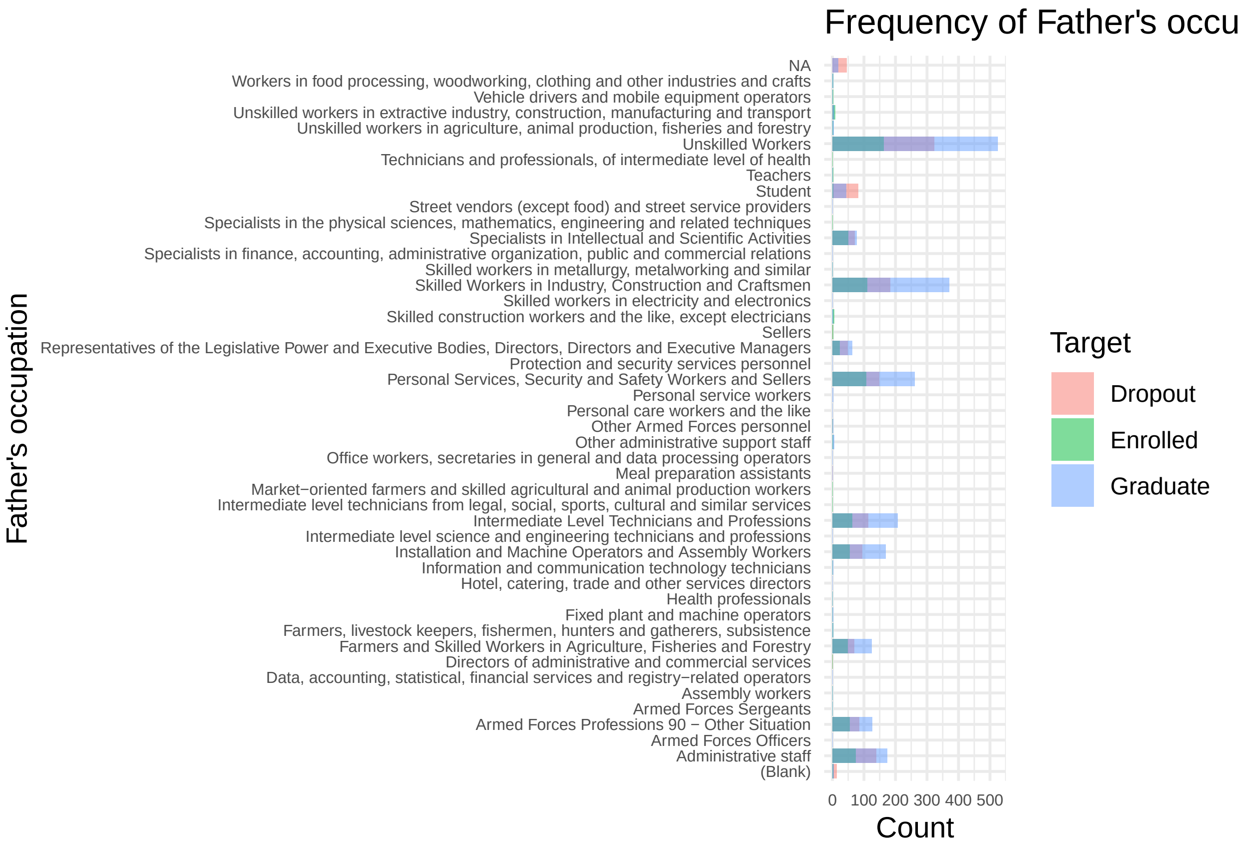

Frequency of Previous qualification by Target
Previous qualification
Technological specialization course
Secondary education
Professional higher technical course
Other − 11th year of schooling
Higher education − masters
Higher education − master (2nd cycle)
Higher education − doctorate
Higher education − degree (1st cycle)
Higher education − degree
Higher education − bachelors degree
Frequency of higher education
Basic education 3rd cycle (9th/10th/11th year) or equiv.
Basic education 2nd cycle (6th/7th/8th year) or equiv.
12th year of schooling − not completed
11th year of schooling − not completed
10th year of schooling − not completed
10th year of schooling
0
500
1000
1500
2000
Count
Target
Dropout
Enrolled
Graduate

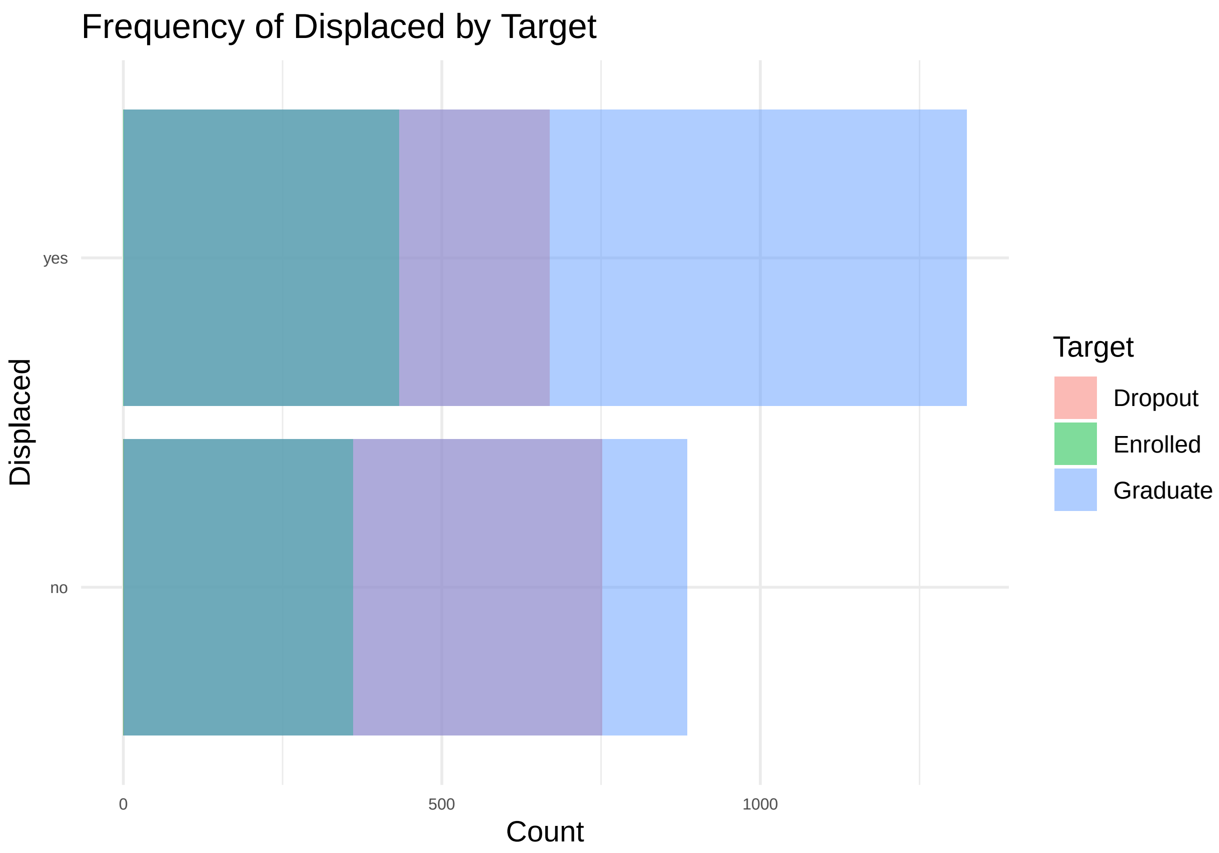

Frequency of Nationality by Target
Ukrainian
Turkish
Spanish
Santomean
Russian
Romanian
Portuguese
Mozambican
Moldova (Republic of)
Mexican
Lithuanian
Italian
Guinean
German
English
Dutch
Cuban
Columbian
Cape Verdean
Brazilian
Angolan
Nationality
0
500
1000
1500
2000
Count
Target
Dropout
Enrolled
Graduate

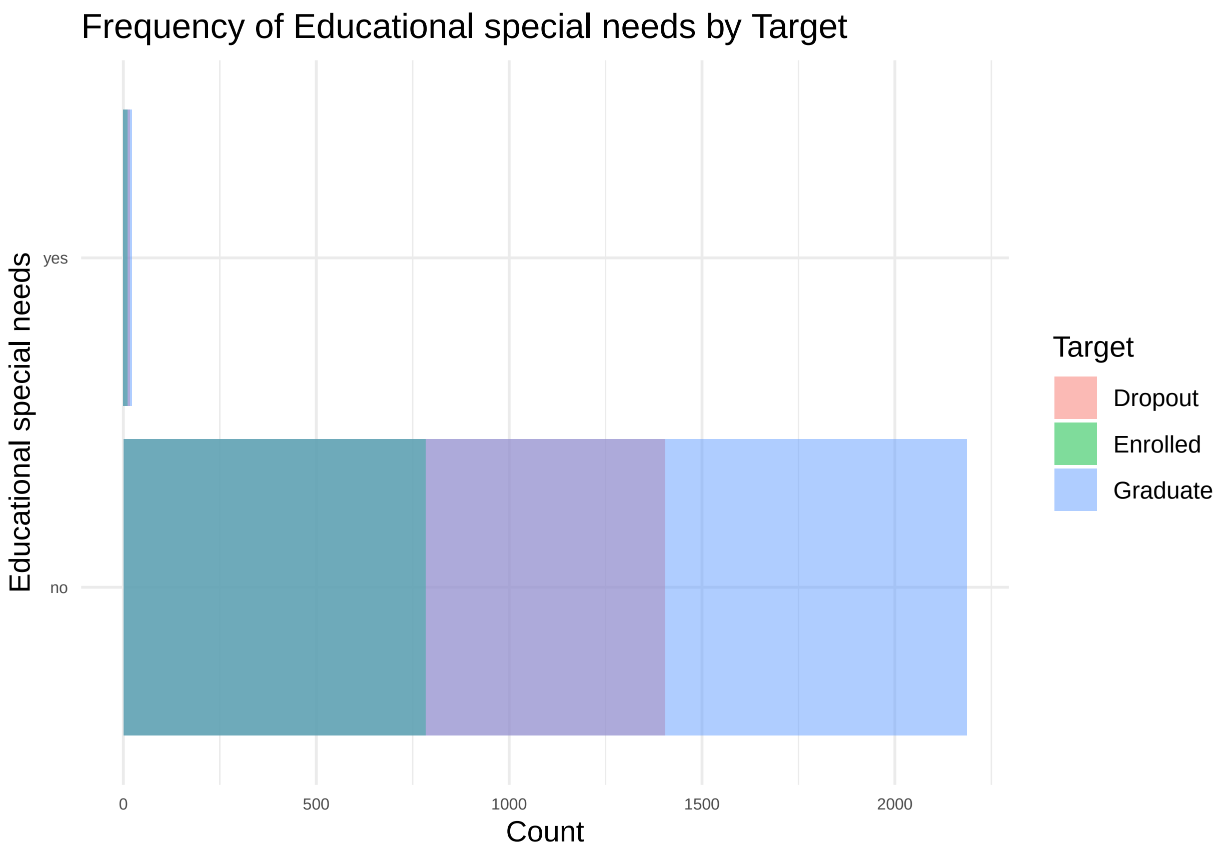

Frequency of Mother's qualification by Target
Mother's qualification
Unknown
Technological specialization course
Technical−professional course
Specialized higher studies course
Secondary Education − 12th Year of Schooling or Eq.
Professional higher technical course
Other − 11th Year of Schooling
Higher Education − Master (2nd cycle)
Higher Education − Master's
Higher Education − Doctorate (3rd cycle)
Higher Education − Doctorate
Higher education − degree (1st cycle)
Higher Education − Degree
Higher Education − Bachelor's Degree
General commerce course
Frequency of Higher Education
Can read without having a 4th year of schooling
Can't read or write
Basic Education 3rd Cycle (9th/10th/11th Year) or Equiv.
Basic Education 2nd Cycle (6th/7th/8th Year) or Equiv.
Basic education 1st cycle (4th/5th year) or equiv.
9th Year of Schooling − Not Completed
8th year of schooling
7th year of schooling
7th Year (Old)
2nd cycle of the general high school course
12th Year of Schooling − Not Completed
11th Year of Schooling − Not Completed
10th Year of Schooling
0
200
400
Count
Target
Dropout
Enrolled
Graduate

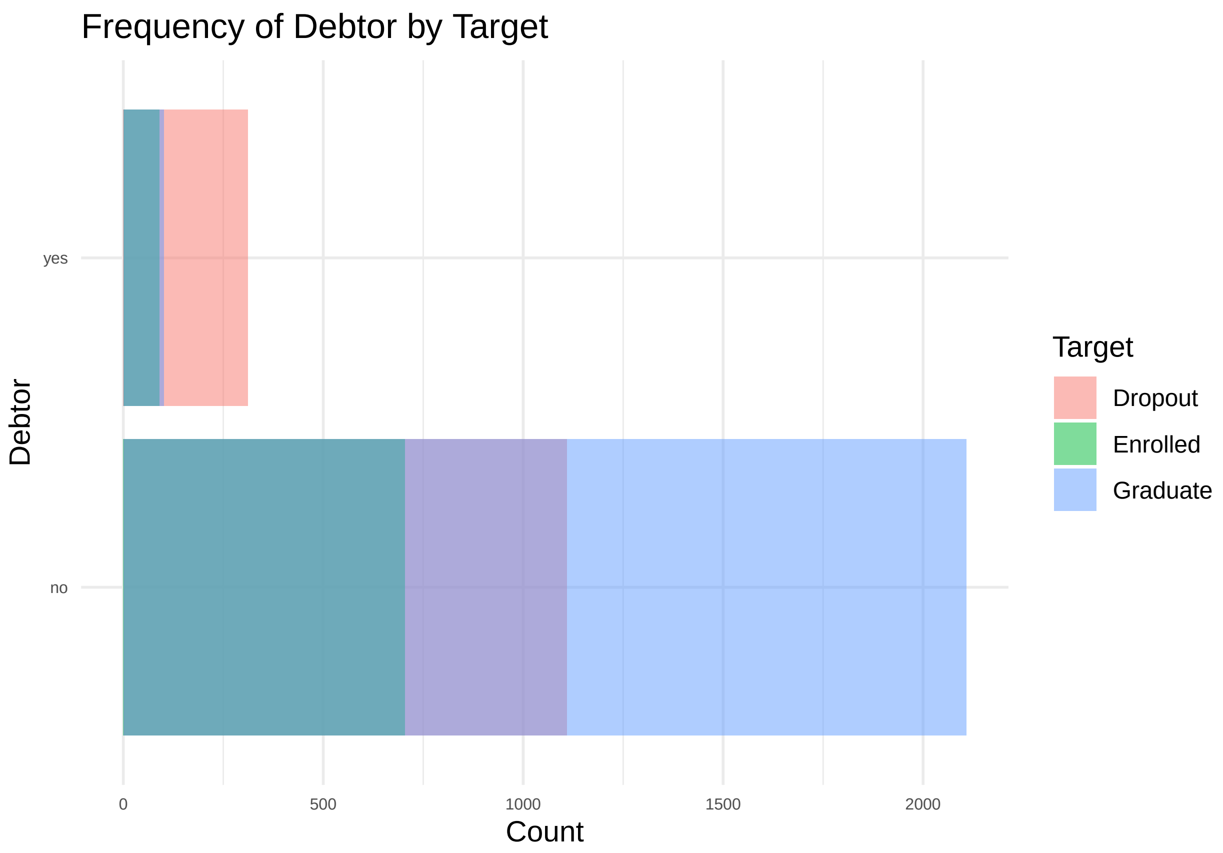

Frequency of Father's qualification by Target
Father's qualification
NA
Unknown
Technological specialization course
Technical−professional course
Supplementary Accounting and Administration
Specialized higher studies course
Secondary Education − 12th Year of Schooling or Eq.
Professional higher technical course
Other − 11th Year of Schooling
Higher Education − Master (2nd cycle)
Higher Education − Master's
Higher Education − Doctorate
Higher education − degree (1st cycle)
Higher Education − Degree
Higher Education − Bachelor's Degree
General Course of Administration and Commerce
General commerce course
Frequency of Higher Education
Complementary High School Course − not concluded
Complementary High School Course
Can read without having a 4th year of schooling
Can't read or write
Basic Education 3rd Cycle (9th/10th/11th Year) or Equiv.
Basic Education 2nd Cycle (6th/7th/8th Year) or Equiv.
Basic education 1st cycle (4th/5th year) or equiv.
9th Year of Schooling − Not Completed
8th year of schooling
7th year of schooling
7th Year (Old)
2nd year complementary high school course
2nd cycle of the general high school course
12th Year of Schooling − Not Completed
11th Year of Schooling − Not Completed
10th Year of Schooling
0
200
400
600
Count
Target
Dropout
Enrolled
Graduate

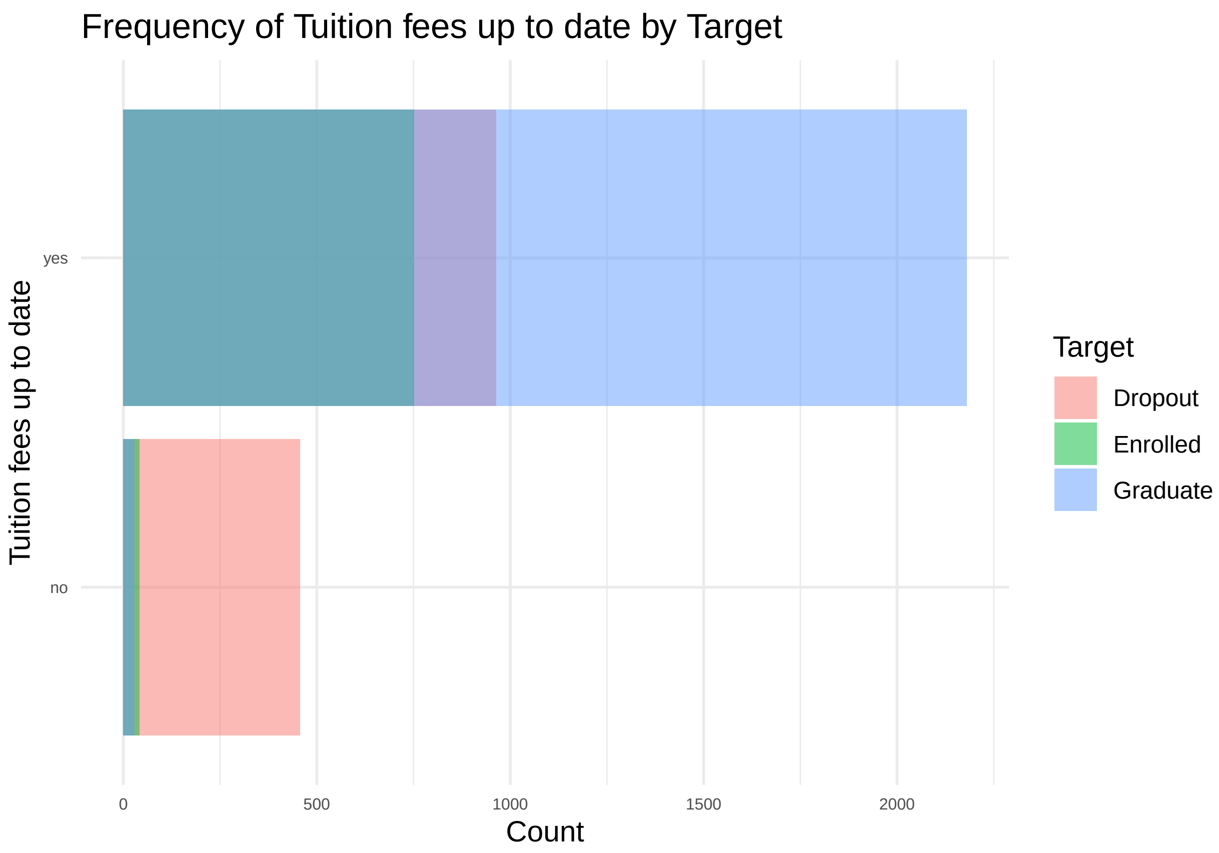

Frequency of Mother's occu
Mother's occupation
NA
Workers in food processing, woodworking, clothing and other industries and crafts
Unskilled workers in extractive industry, construction, manufacturing and transport
Unskilled workers in agriculture, animal production, fisheries and forestry
Unskilled Workers
Technicians and professionals, of intermediate level of health
Teachers
Student
Specialists in Intellectual and Scientific Activities
Specialists in information and communication technologies (ICT)
Skilled workers in printing, precision instrument manufacturing, jewelers, artisans and the like
Skilled Workers in Industry, Construction and Craftsmen
Skilled construction workers and the like, except electricians
Sellers
Representatives of the Legislative Power and Executive Bodies, Directors, Directors and Executive Managers
Personal Services, Security and Safety Workers and Sellers
Personal service workers
Personal care workers and the like
Other administrative support staff
Office workers, secretaries in general and data processing operators
Meal preparation assistants
Intermediate level technicians from legal, social, sports, cultural and similar services
Intermediate Level Technicians and Professions
Intermediate level science and engineering technicians and professions
Installation and Machine Operators and Assembly Workers
Health professionals
Farmers and Skilled Workers in Agriculture, Fisheries and Forestry
Data, accounting, statistical, financial services and registry−related operators
Cleaning workers
Armed Forces Professions 90 − Other Situation
Administrative staff
(Blank)
0
200
400
600
800
Count
Target
Dropout
Enrolled
Graduate

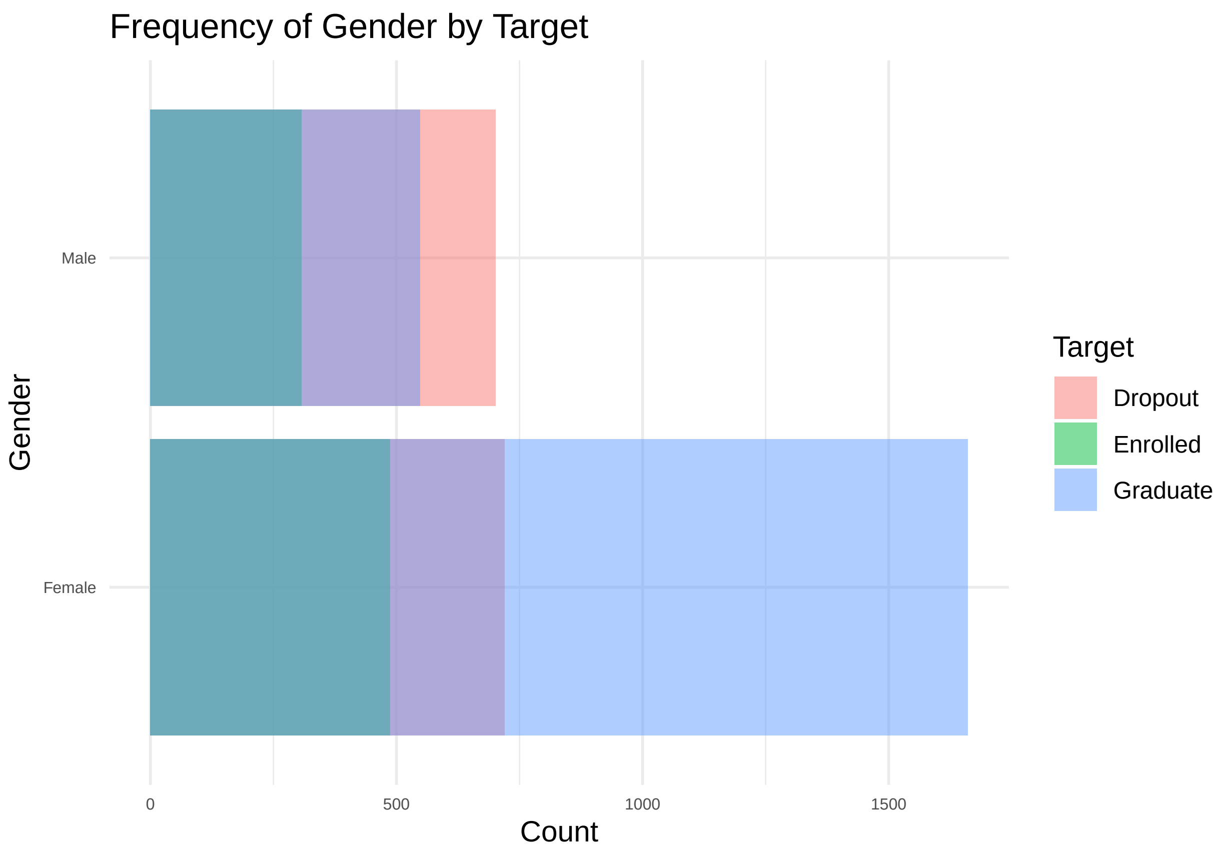

Frequency of Father's occu
Father's occupation
NA
Workers in food processing, woodworking, clothing and other industries and crafts
Vehicle drivers and mobile equipment operators
Unskilled workers in extractive industry, construction, manufacturing and transport
Unskilled workers in agriculture, animal production, fisheries and forestry
Unskilled Workers
Technicians and professionals, of intermediate level of health
Teachers
Student
Street vendors (except food) and street service providers
Specialists in the physical sciences, mathematics, engineering and related techniques
Specialists in Intellectual and Scientific Activities
Specialists in finance, accounting, administrative organization, public and commercial relations
Skilled workers in metallurgy, metalworking and similar
Skilled Workers in Industry, Construction and Craftsmen
Skilled workers in electricity and electronics
Skilled construction workers and the like, except electricians
Sellers
Representatives of the Legislative Power and Executive Bodies, Directors, Directors and Executive Managers
Protection and security services personnel
Personal Services, Security and Safety Workers and Sellers
Personal service workers
Personal care workers and the like
Other Armed Forces personnel
Other administrative support staff
Office workers, secretaries in general and data processing operators
Meal preparation assistants
Market−oriented farmers and skilled agricultural and animal production workers
Intermediate level technicians from legal, social, sports, cultural and similar services
Intermediate Level Technicians and Professions
Intermediate level science and engineering technicians and professions
Installation and Machine Operators and Assembly Workers
Information and communication technology technicians
Hotel, catering, trade and other services directors
Health professionals
Fixed plant and machine operators
Farmers, livestock keepers, fishermen, hunters and gatherers, subsistence
Farmers and Skilled Workers in Agriculture, Fisheries and Forestry
Directors of administrative and commercial services
Data, accounting, statistical, financial services and registry−related operators
Assembly workers
Armed Forces Sergeants
Armed Forces Professions 90 − Other Situation
Armed Forces Officers
Administrative staff
(Blank)
0
100
200
300
400
500
Count
Target
Dropout
Enrolled
Graduate

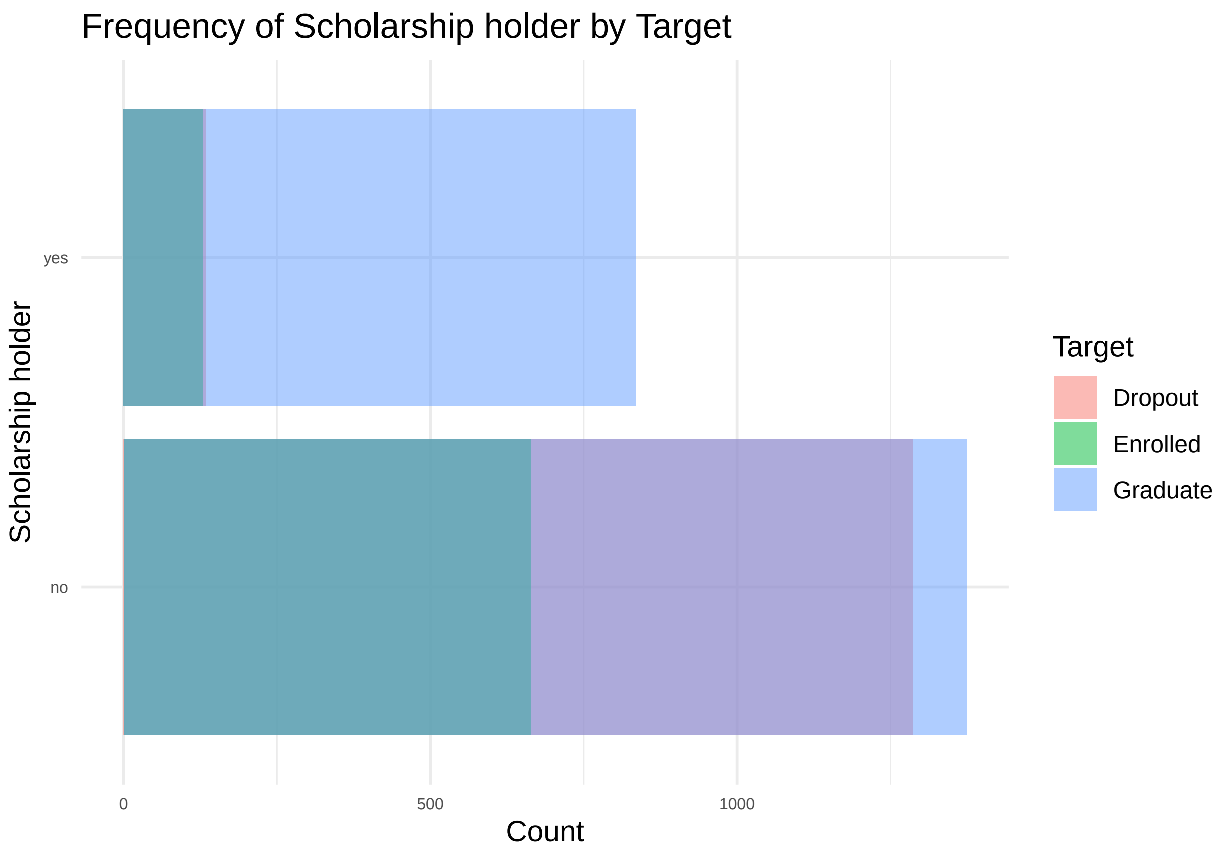

Frequency of Displaced by Target
Displaced
yes
no
Count
0
500
1000
Target
Dropout
Enrolled
Graduate

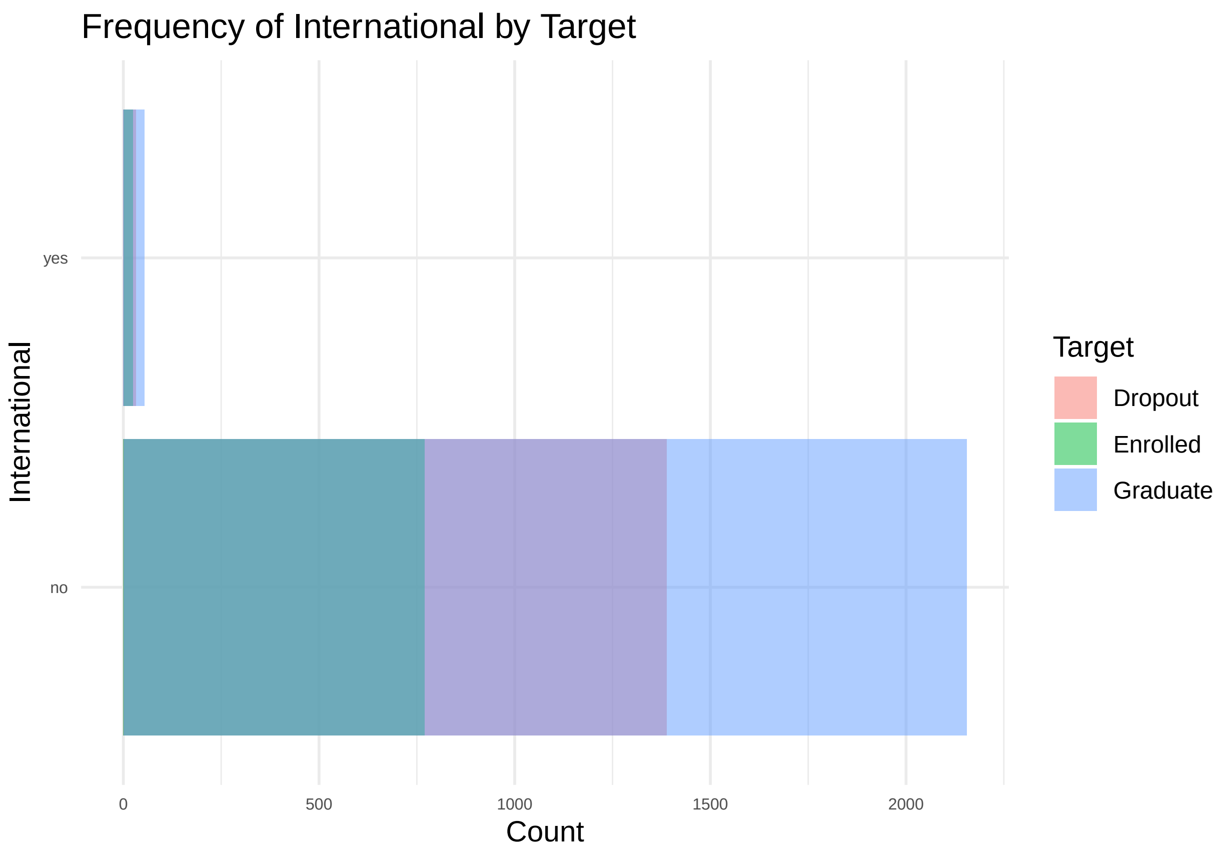

Frequency of Educational special needs by Target
Educational special needs
yes
no
0
500
1000
1500
2000
Count
Target
Dropout
Enrolled
Graduate


Frequency of Debtor by Target
Debtor
yes
no
0
500
1000
1500
2000
Count
Target
Dropout
Enrolled
Graduate

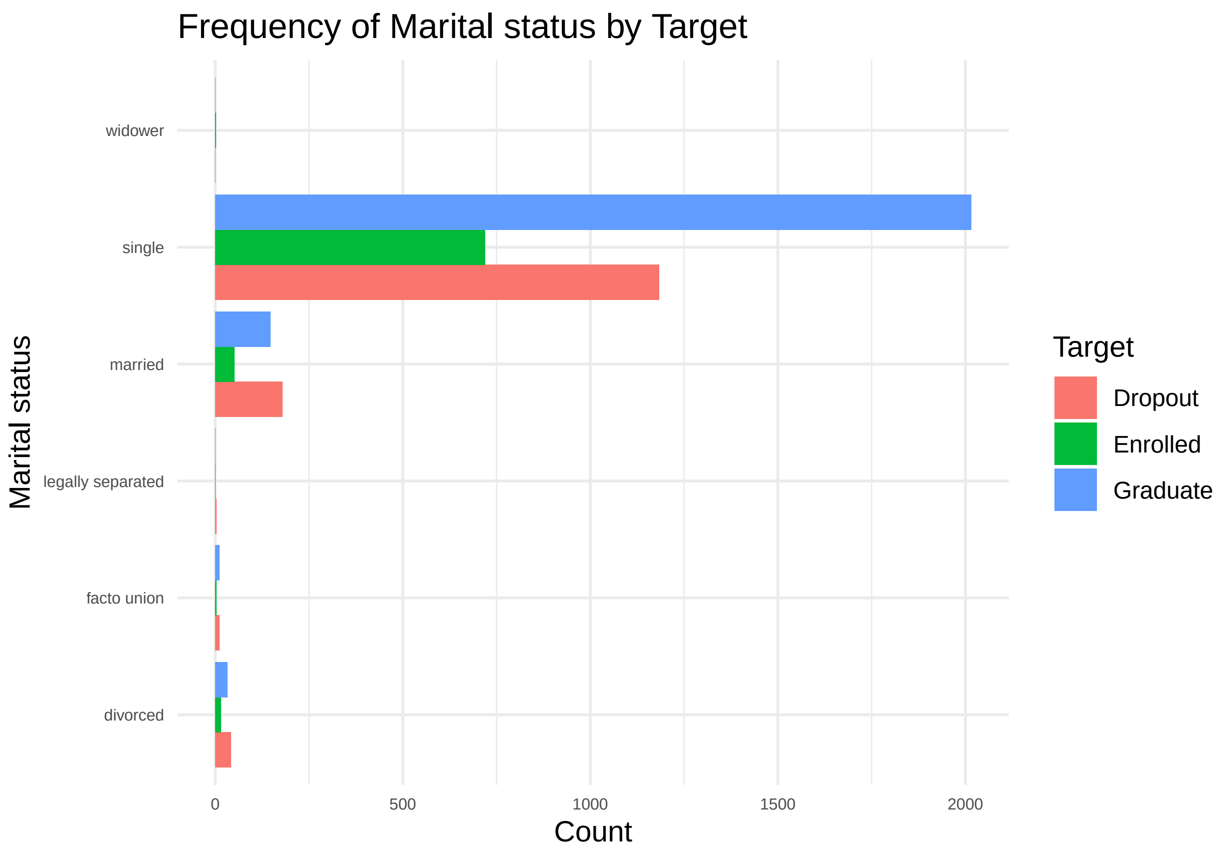

Frequency of Tuition fees up to date by Target
Tuition fees up to date
yes
no
0
500
1000
1500
2000
Count
Target
Dropout
Enrolled
Graduate

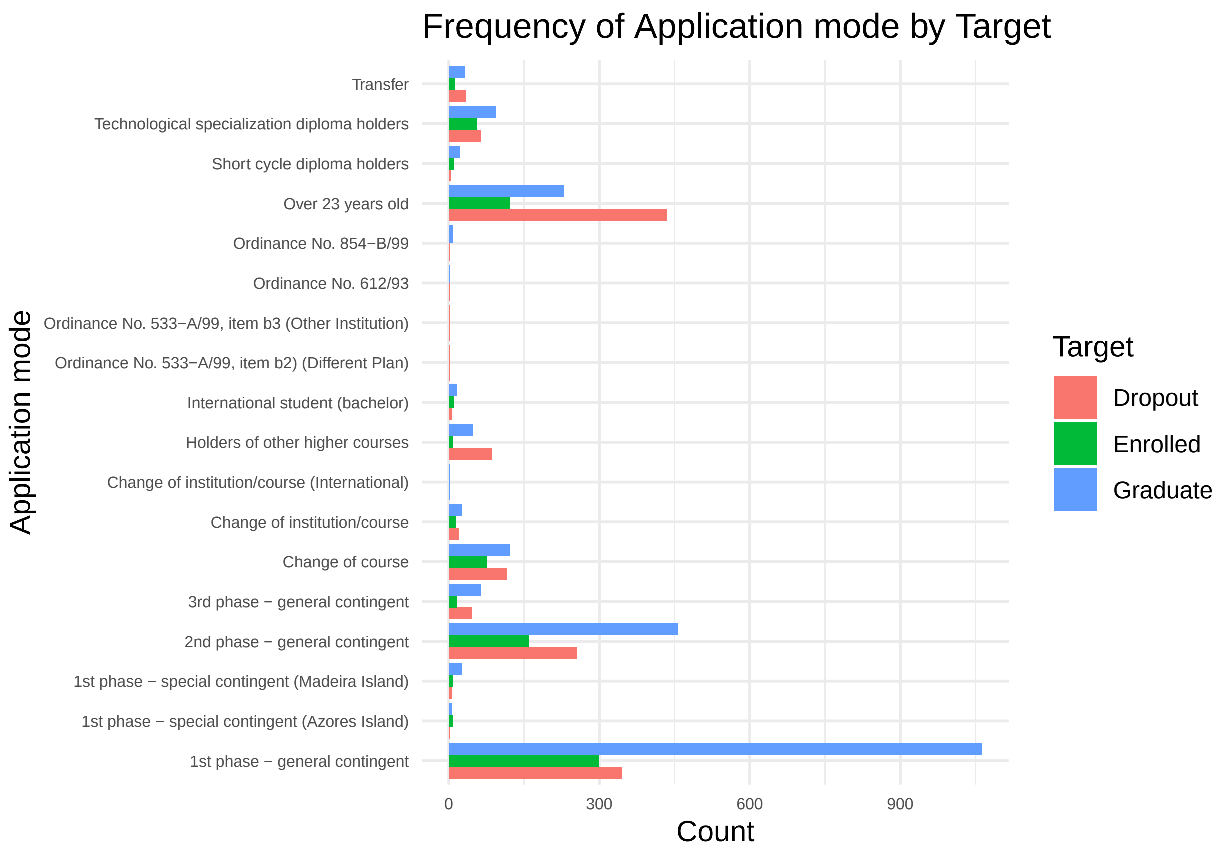

Frequency of Gender by Target
Gender
Male
Female
0
500
1000
1500
Count
Target
Dropout
Enrolled
Graduate

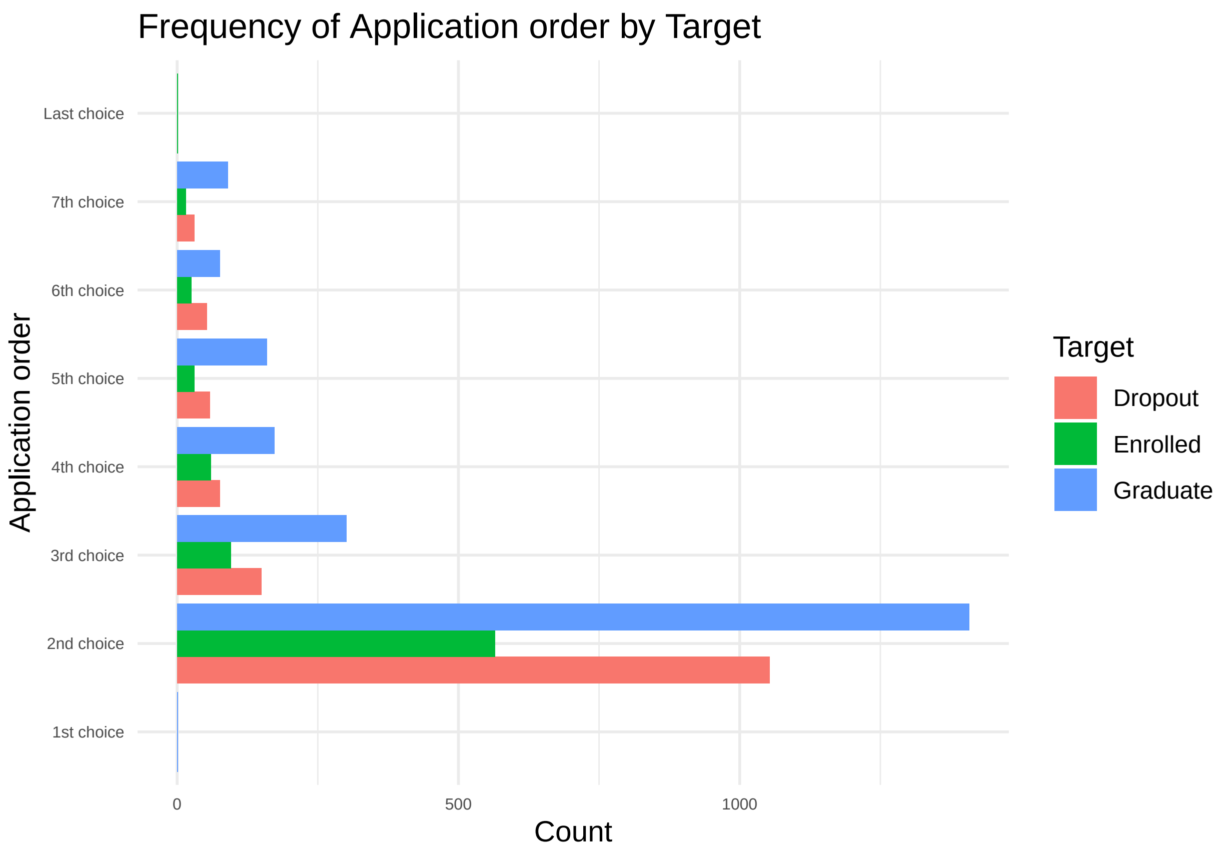

Frequency of Scholarship holder by Target
Scholarship holder
yes
no
0
500
1000
Count
Target
Dropout
Enrolled
Graduate

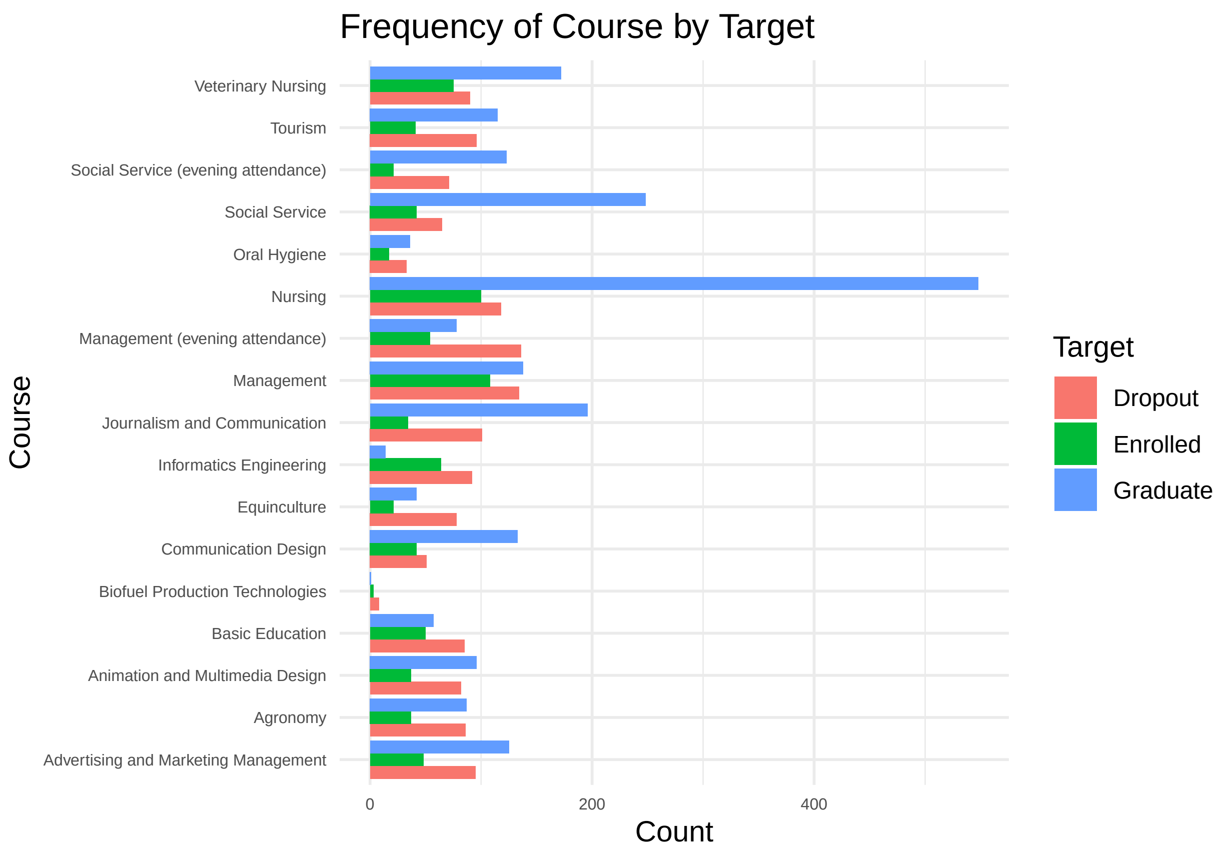


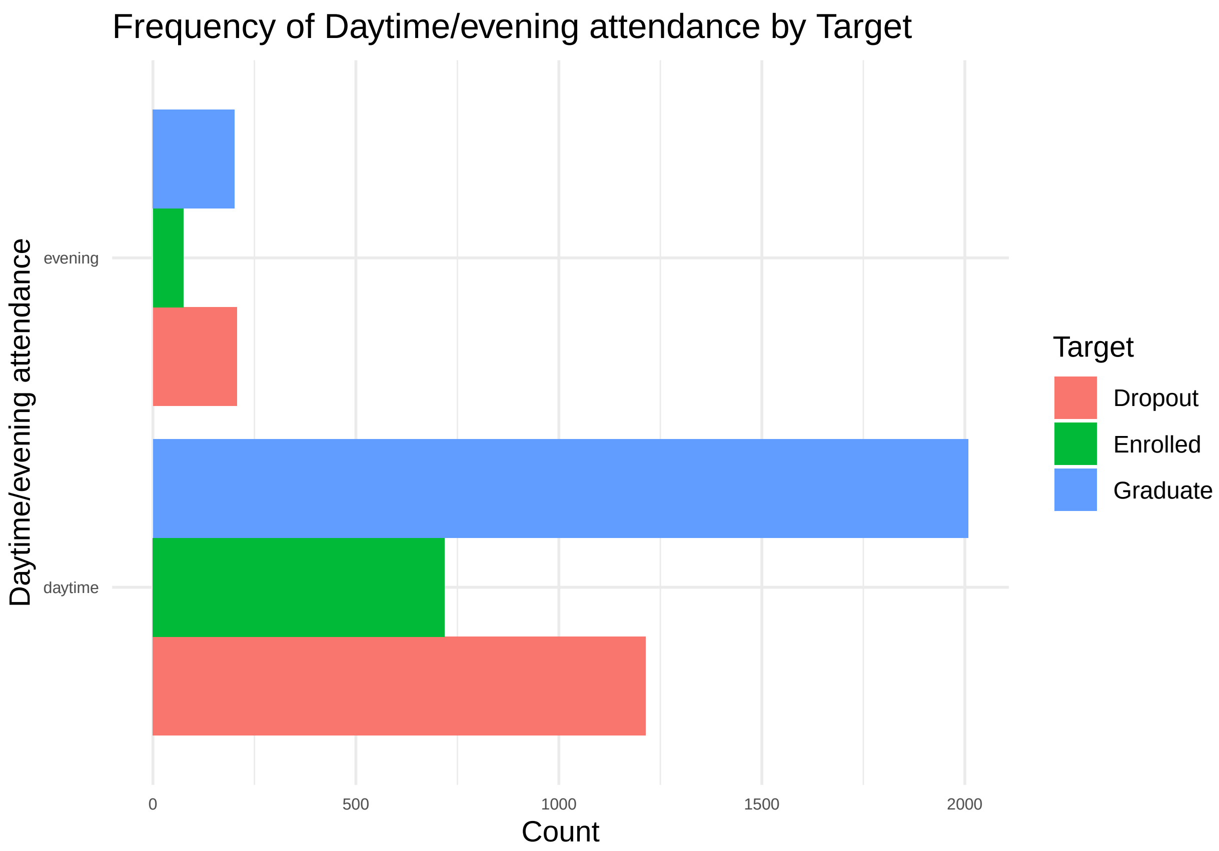
```
for (i in categorical_cols) {
  plot <- ggplot(data1, aes(x = .data[[i]], fill = Target)) +
  geom_bar(position = "dodge") +
  labs(
    title = paste("Frequency of", i, "by Target"),
    x = i,
    y = "Count"
    ) +
    theme_minimal() +
    scale_x_discrete(
      labels = label_wrap(150)
    ) +
    coord_flip() + theme(
      axis.text = element_text(size = 6)
    )
  print(plot)
}
```

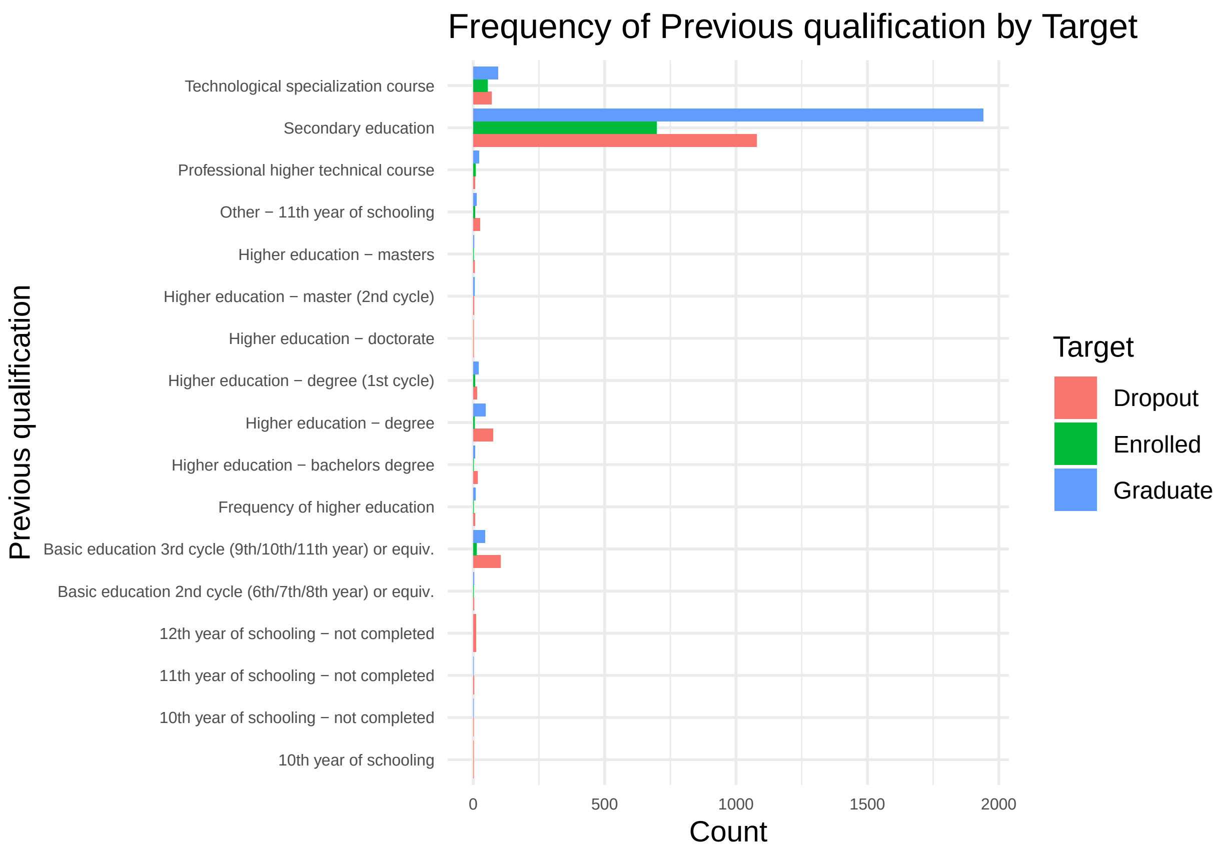

Frequency of Marital status by Target
Marital status
widower
single
married
legally separated
facto union
divorced
0
500
1000
1500
2000
Count
Target
Dropout
Enrolled
Graduate

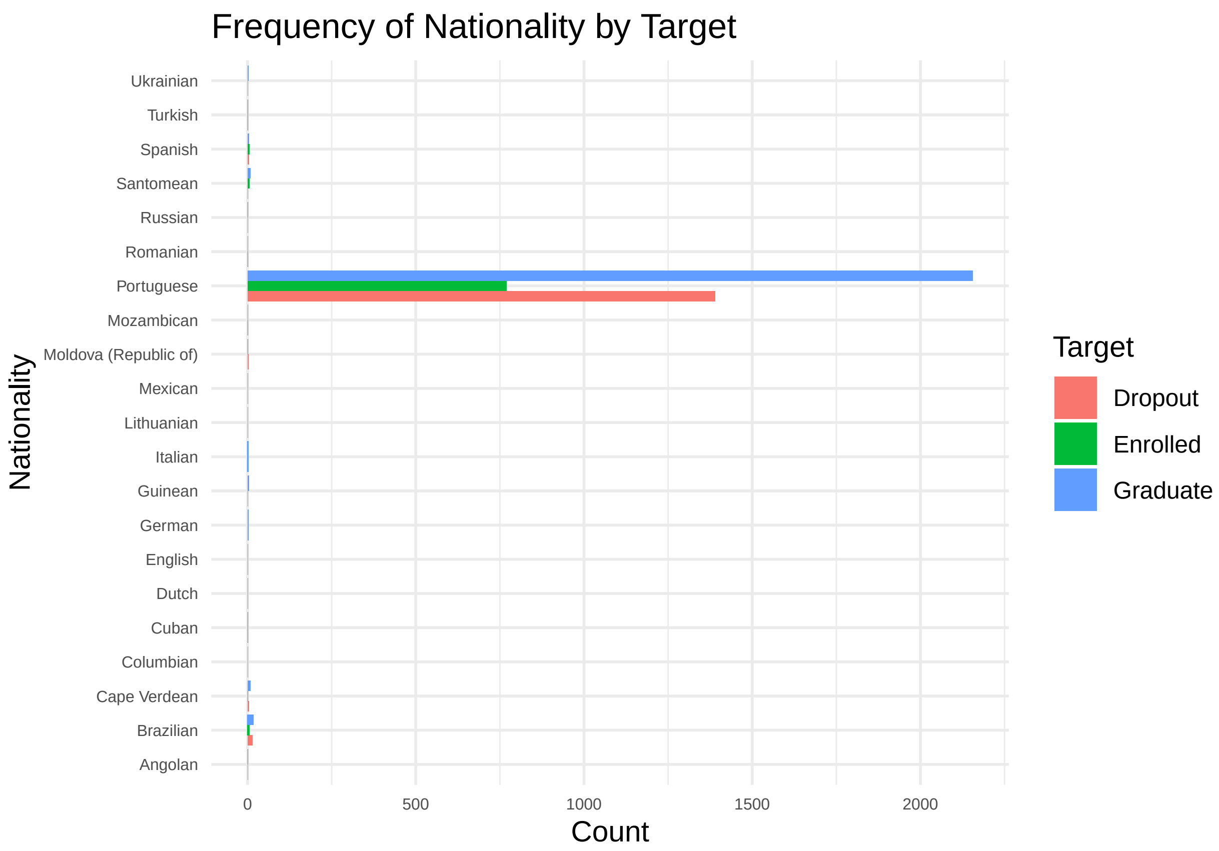

Frequency of Application mode by Target
Application mode
Transfer
Technological specialization diploma holders
Short cycle diploma holders
Over 23 years old
Ordinance No. 854−B/99
Ordinance No. 612/93
Ordinance No. 533−A/99, item b3 (Other Institution)
Ordinance No. 533−A/99, item b2) (Different Plan)
International student (bachelor)
Holders of other higher courses
Change of institution/course (International)
Change of institution/course
Change of course
3rd phase − general contingent
2nd phase − general contingent
1st phase − special contingent (Madeira Island)
1st phase − special contingent (Azores Island)
1st phase − general contingent
0
300
600
900
Count
Target
Dropout
Enrolled
Graduate

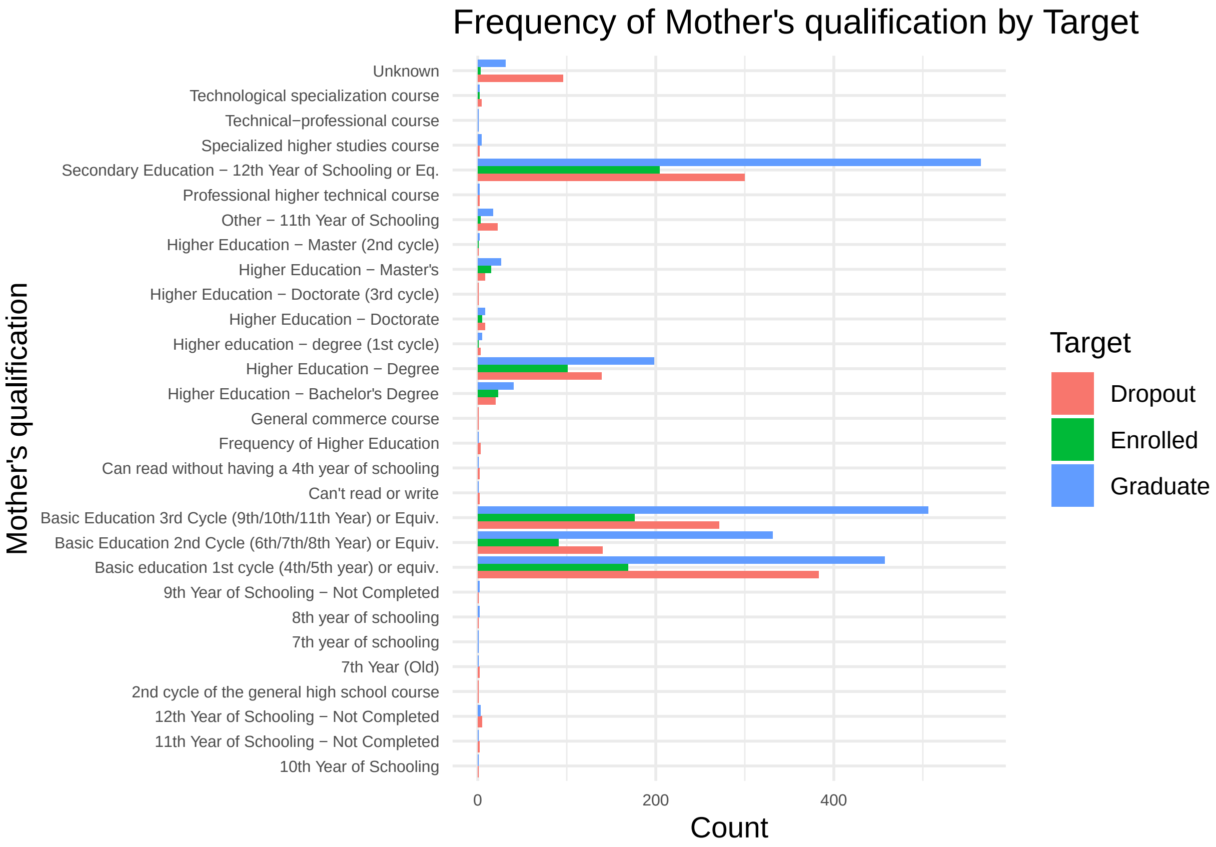

Frequency of Application order by Target
Application order
Last choice
7th choice
6th choice
5th choice
4th choice
3rd choice
2nd choice
1st choice
0
500
1000
Count
Target
Dropout
Enrolled
Graduate

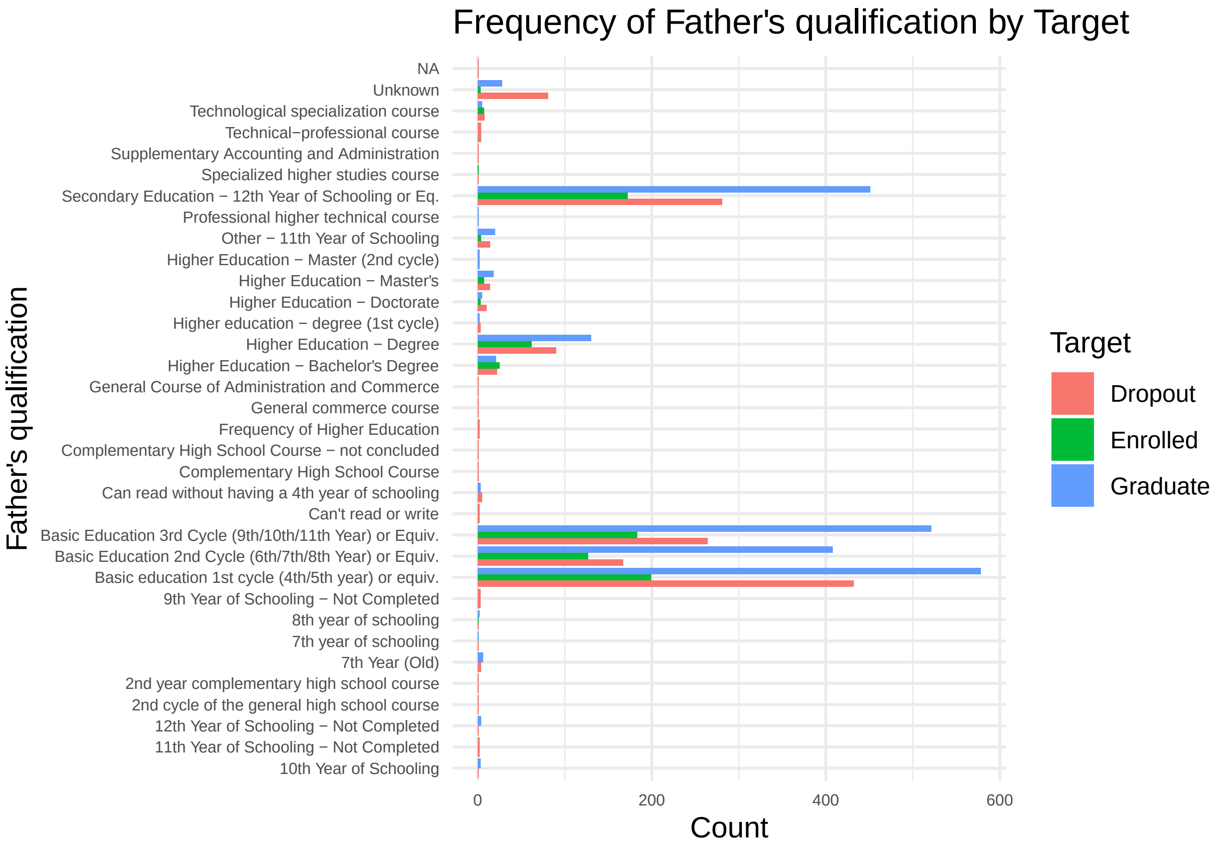

Frequency of Course by Target
Course
Veterinary Nursing
Tourism
Social Service (evening attendance)
Social Service
Oral Hygiene
Nursing
Management (evening attendance)
Management
Journalism and Communication
Informatics Engineering
Equinculture
Communication Design
Biofuel Production Technologies
Basic Education
Animation and Multimedia Design
Agronomy
Advertising and Marketing Management
0
200
400
Count
Target
Dropout
Enrolled
Graduate

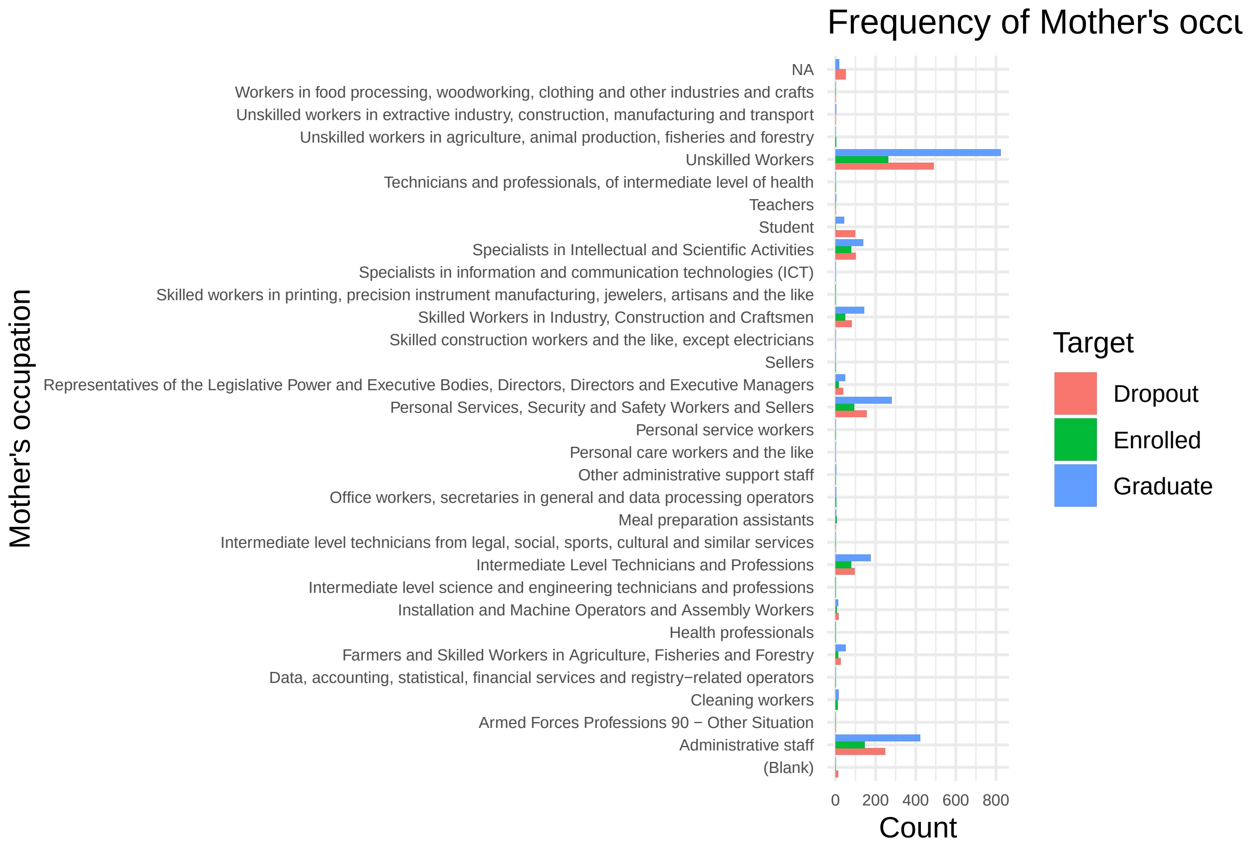

Frequency of Daytime/evening attendance by Target
Daytime/evening attendance
evening
daytime
0
500
1000
1500
2000
Count
Target
Dropout
Enrolled
Graduate


Frequency of Previous qualification by Target
Previous qualification
Technological specialization course
Secondary education
Professional higher technical course
Other − 11th year of schooling
Higher education − masters
Higher education − master (2nd cycle)
Higher education − doctorate
Higher education − degree (1st cycle)
Higher education − degree
Higher education − bachelors degree
Frequency of higher education
Basic education 3rd cycle (9th/10th/11th year) or equiv.
Basic education 2nd cycle (6th/7th/8th year) or equiv.
12th year of schooling − not completed
11th year of schooling − not completed
10th year of schooling − not completed
10th year of schooling
0
500
1000
1500
2000
Count
Target
Dropout
Enrolled
Graduate

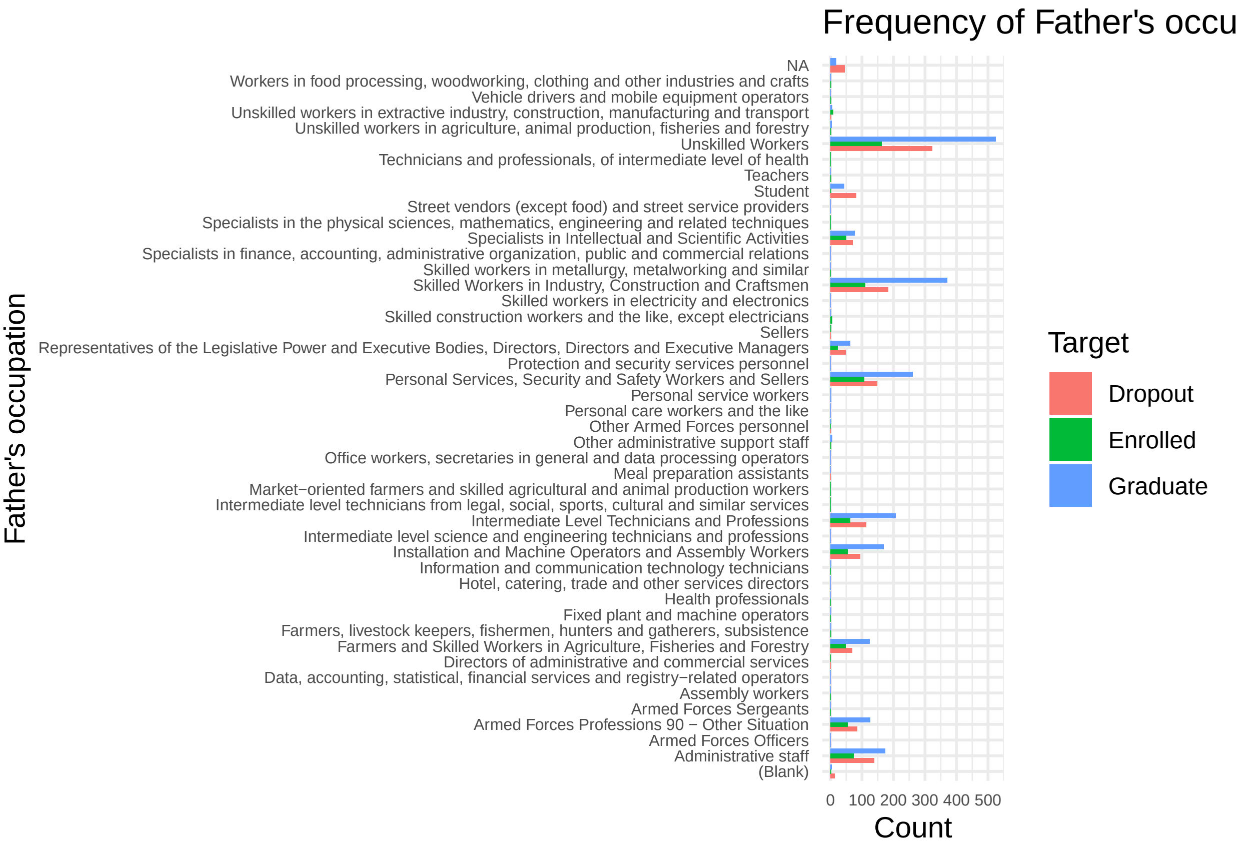

Frequency of Nationality by Target
Nationality
Ukrainian
Turkish
Spanish
Santomean
Russian
Romanian
Portuguese
Mozambican
Moldova (Republic of)
Mexican
Lithuanian
Italian
Guinean
German
English
Dutch
Cuban
Columbian
Cape Verdean
Brazilian
Angolan
0
500
1000
1500
2000
Count
Target
Dropout
Enrolled
Graduate

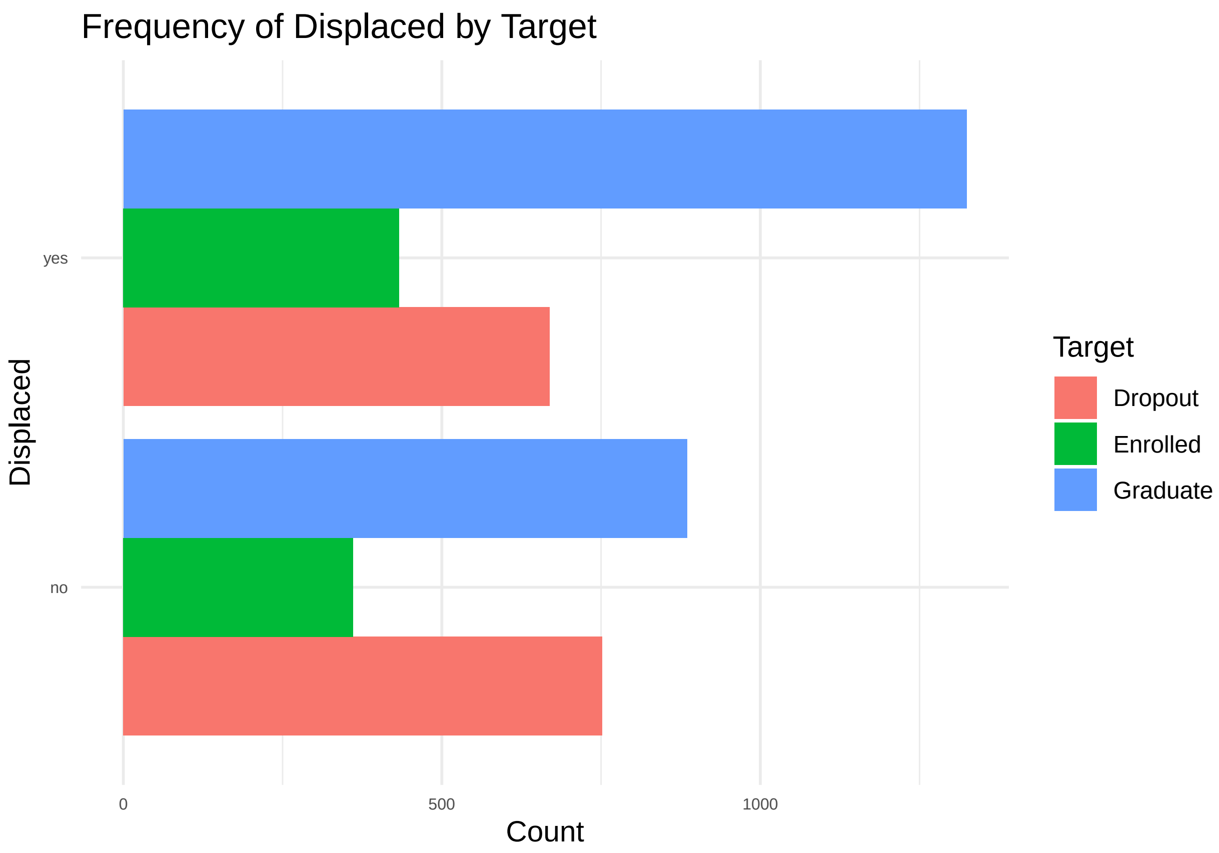

Frequency of Mother's qualification by Target
Mother's qualification
Unknown
Technological specialization course
Technical−professional course
Specialized higher studies course
Secondary Education − 12th Year of Schooling or Eq.
Professional higher technical course
Other − 11th Year of Schooling
Higher Education − Master (2nd cycle)
Higher Education − Master's
Higher Education − Doctorate (3rd cycle)
Higher Education − Doctorate
Higher education − degree (1st cycle)
Higher Education − Degree
Higher Education − Bachelor's Degree
General commerce course
Frequency of Higher Education
Can read without having a 4th year of schooling
Can't read or write
Basic Education 3rd Cycle (9th/10th/11th Year) or Equiv.
Basic Education 2nd Cycle (6th/7th/8th Year) or Equiv.
Basic education 1st cycle (4th/5th year) or equiv.
9th Year of Schooling − Not Completed
8th year of schooling
7th year of schooling
7th Year (Old)
2nd cycle of the general high school course
12th Year of Schooling − Not Completed
11th Year of Schooling − Not Completed
10th Year of Schooling
0
200
400
Count
Target
Dropout
Enrolled
Graduate

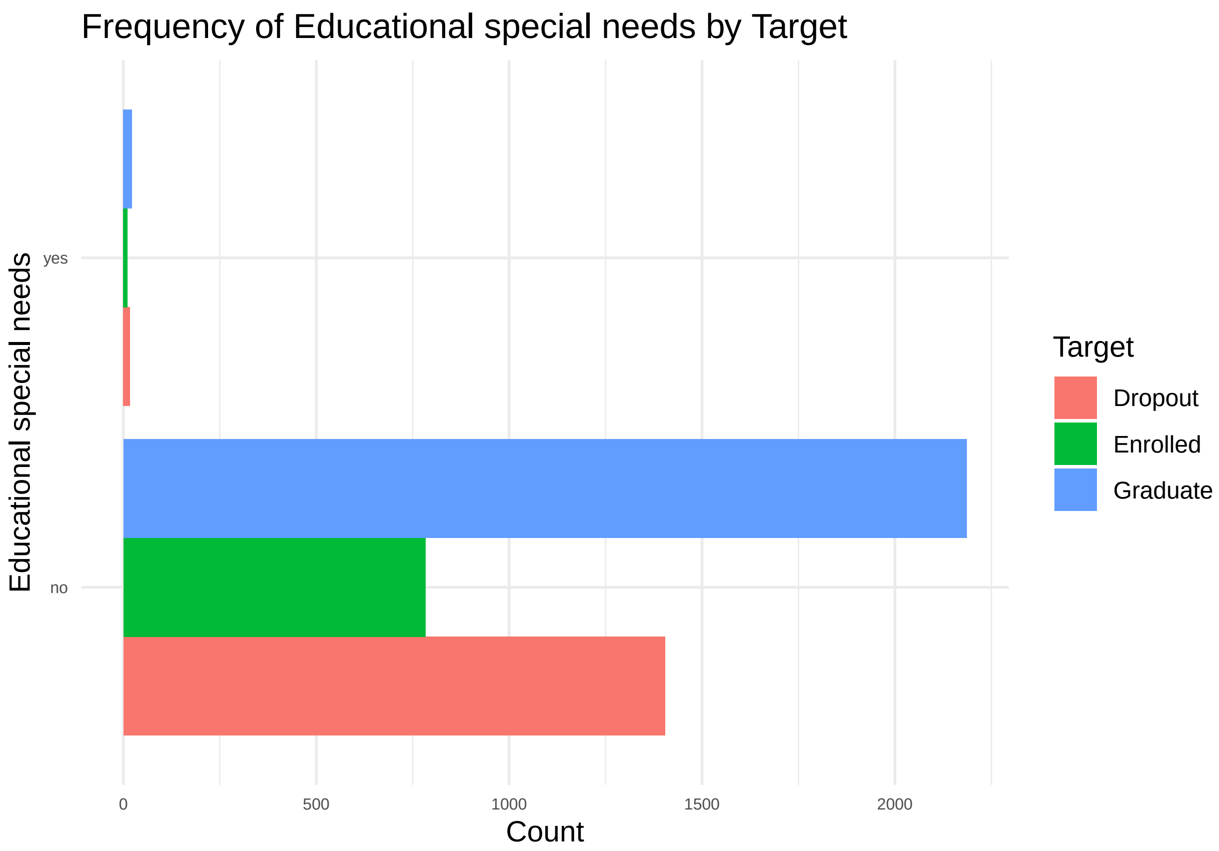

Frequency of Father's qualification by Target
Father's qualification
NA
Unknown
Technological specialization course
Technical−professional course
Supplementary Accounting and Administration
Specialized higher studies course
Secondary Education − 12th Year of Schooling or Eq.
Professional higher technical course
Other − 11th Year of Schooling
Higher Education − Master (2nd cycle)
Higher Education − Master's
Higher Education − Doctorate
Higher education − degree (1st cycle)
Higher Education − Degree
Higher Education − Bachelor's Degree
General Course of Administration and Commerce
General commerce course
Frequency of Higher Education
Complementary High School Course − not concluded
Complementary High School Course
Can read without having a 4th year of schooling
Can't read or write
Basic Education 3rd Cycle (9th/10th/11th Year) or Equiv.
Basic Education 2nd Cycle (6th/7th/8th Year) or Equiv.
Basic education 1st cycle (4th/5th year) or equiv.
9th Year of Schooling − Not Completed
8th year of schooling
7th year of schooling
7th Year (Old)
2nd year complementary high school course
2nd cycle of the general high school course
12th Year of Schooling − Not Completed
11th Year of Schooling − Not Completed
10th Year of Schooling
0
200
400
600
Count
Target
Dropout
Enrolled
Graduate

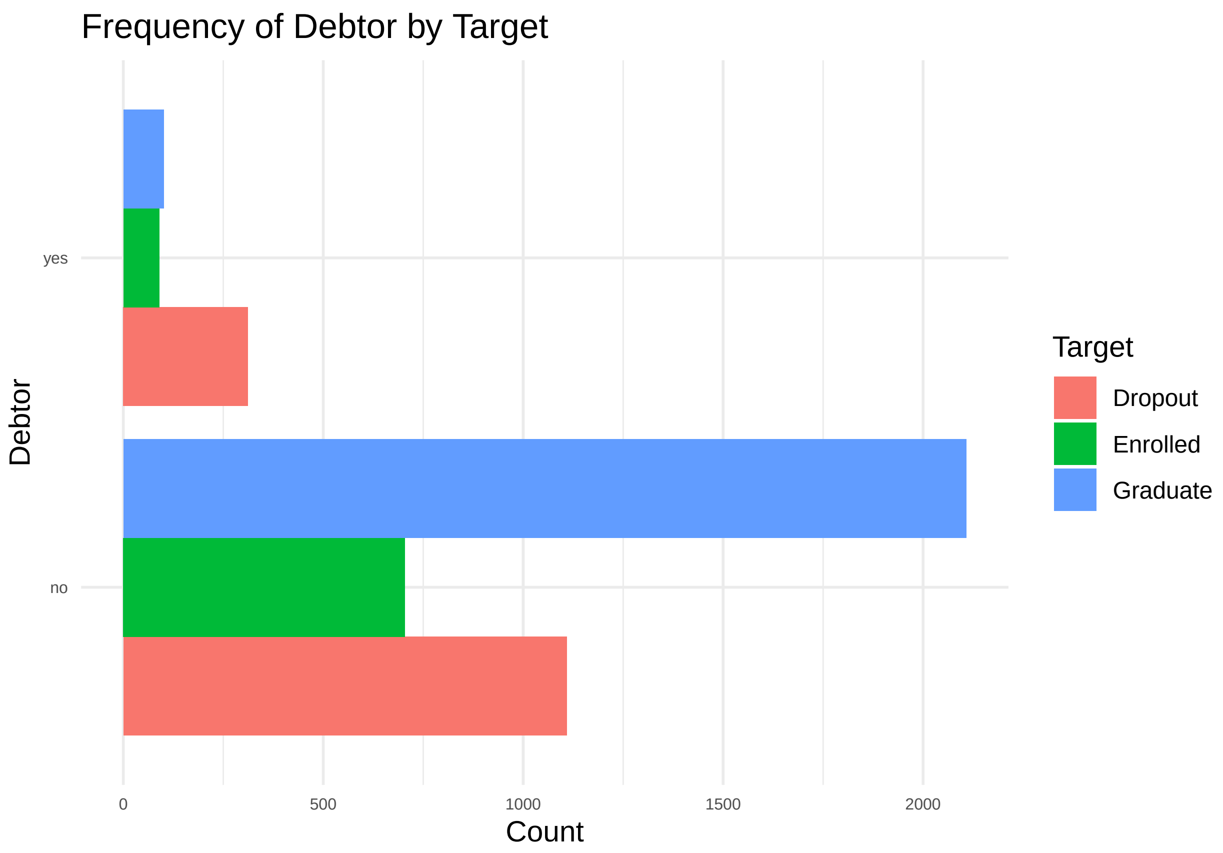


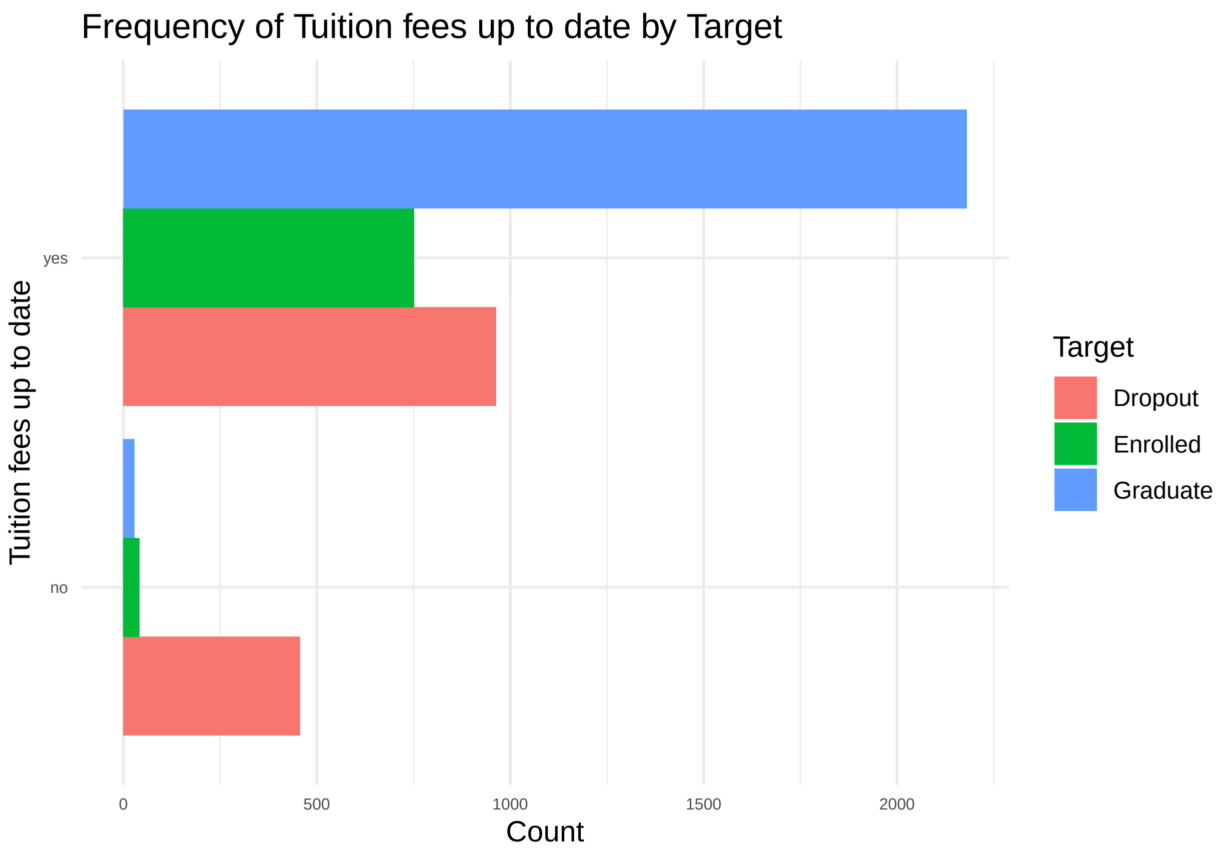
Frequency of Mother's occu
Mother's occupation
NA
Workers in food processing, woodworking, clothing and other industries and crafts
Unskilled workers in extractive industry, construction, manufacturing and transport
Unskilled workers in agriculture, animal production, fisheries and forestry
Unskilled Workers
Technicians and professionals, of intermediate level of health
Teachers
Student
Specialists in Intellectual and Scientific Activities
Specialists in information and communication technologies (ICT)
Skilled workers in printing, precision instrument manufacturing, jewelers, artisans and the like
Skilled Workers in Industry, Construction and Craftsmen
Skilled construction workers and the like, except electricians
Sellers
Representatives of the Legislative Power and Executive Bodies, Directors, Directors and Executive Managers
Personal Services, Security and Safety Workers and Sellers
Personal service workers
Personal care workers and the like
Other administrative support staff
Office workers, secretaries in general and data processing operators
Meal preparation assistants
Intermediate level technicians from legal, social, sports, cultural and similar services
Intermediate Level Technicians and Professions
Intermediate level science and engineering technicians and professions
Installation and Machine Operators and Assembly Workers
Health professionals
Farmers and Skilled Workers in Agriculture, Fisheries and Forestry
Data, accounting, statistical, financial services and registry−related operators
Cleaning workers
Armed Forces Professions 90 − Other Situation
Administrative staff
(Blank)
0
200
400
600
800
Count
Target
Dropout
Enrolled
Graduate

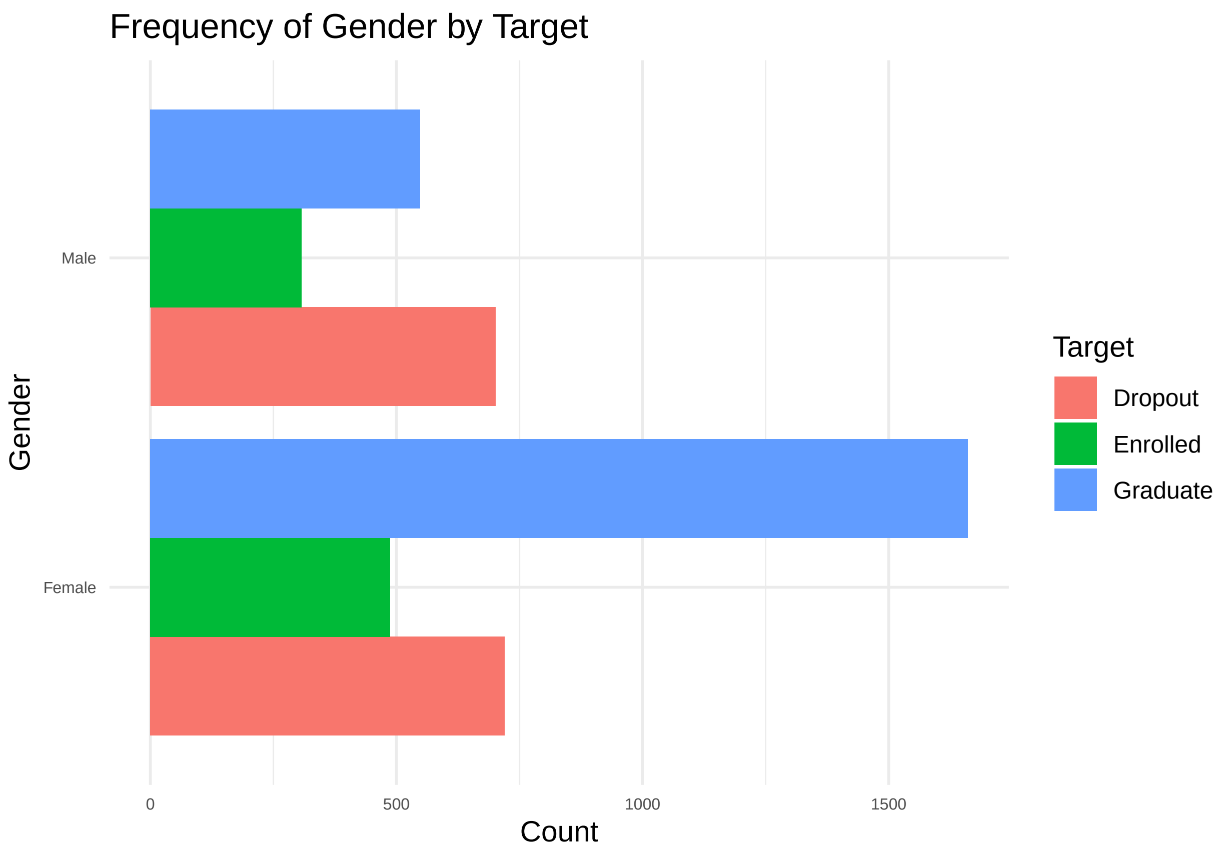

Frequency of Father's occu
Father's occupation
NA
Workers in food processing, woodworking, clothing and other industries and crafts
Vehicle drivers and mobile equipment operators
Unskilled workers in extractive industry, construction, manufacturing and transport
Unskilled workers in agriculture, animal production, fisheries and forestry
Unskilled Workers
Technicians and professionals, of intermediate level of health
Teachers
Student
Street vendors (except food) and street service providers
Specialists in the physical sciences, mathematics, engineering and related techniques
Specialists in Intellectual and Scientific Activities
Specialists in finance, accounting, administrative organization, public and commercial relations
Skilled workers in metallurgy, metalworking and similar
Skilled Workers in Industry, Construction and Craftsmen
Skilled workers in electricity and electronics
Skilled construction workers and the like, except electricians
Sellers
Representatives of the Legislative Power and Executive Bodies, Directors, Directors and Executive Managers
Protection and security services personnel
Personal Services, Security and Safety Workers and Sellers
Personal service workers
Personal care workers and the like
Other Armed Forces personnel
Other administrative support staff
Office workers, secretaries in general and data processing operators
Meal preparation assistants
Market−oriented farmers and skilled agricultural and animal production workers
Intermediate level technicians from legal, social, sports, cultural and similar services
Intermediate Level Technicians and Professions
Intermediate level science and engineering technicians and professions
Installation and Machine Operators and Assembly Workers
Information and communication technology technicians
Hotel, catering, trade and other services directors
Health professionals
Fixed plant and machine operators
Farmers, livestock keepers, fishermen, hunters and gatherers, subsistence
Farmers and Skilled Workers in Agriculture, Fisheries and Forestry
Directors of administrative and commercial services
Data, accounting, statistical, financial services and registry−related operators
Assembly workers
Armed Forces Sergeants
Armed Forces Professions 90 − Other Situation
Armed Forces Officers
Administrative staff
(Blank)
0
100
200
300
400
500
Count
Target
Dropout
Enrolled
Graduate

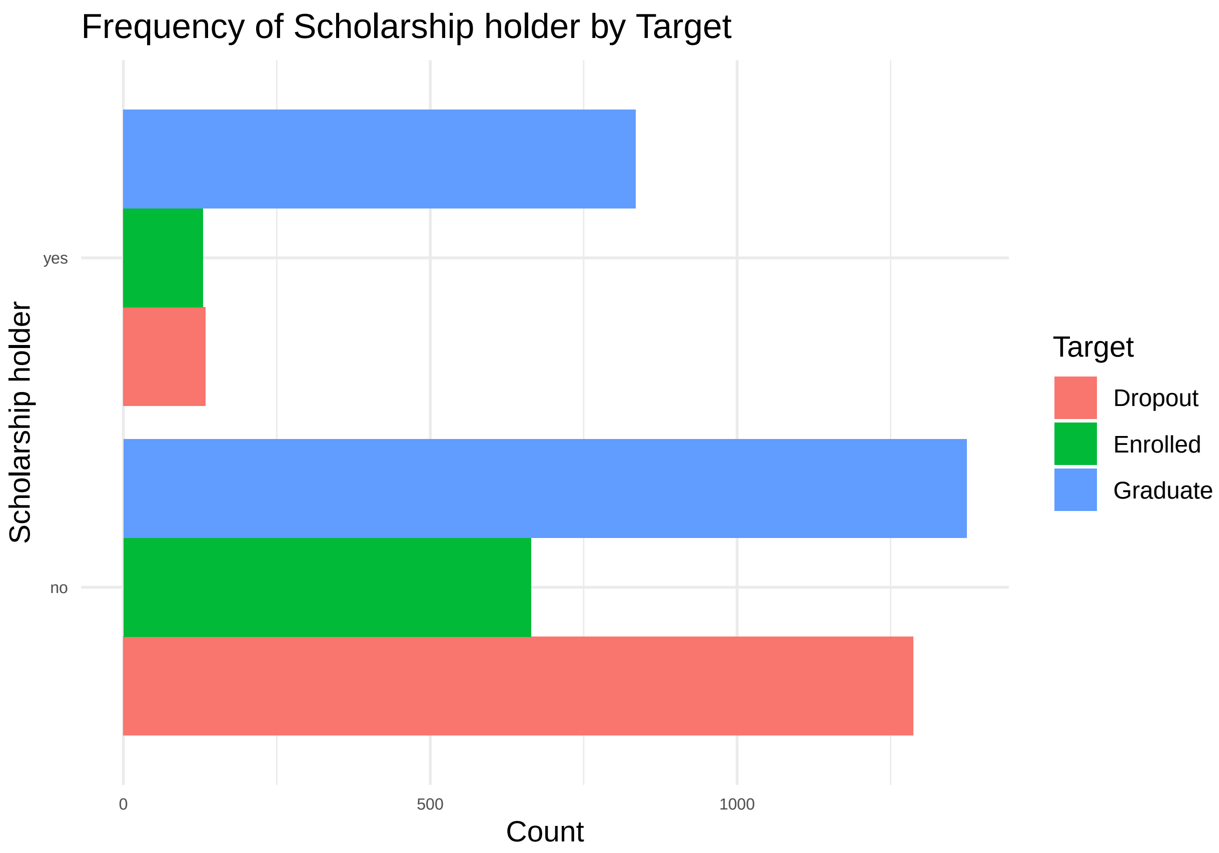

Frequency of Displaced by Target
Displaced
yes
no
0
500
1000
Count
Target
Dropout
Enrolled
Graduate

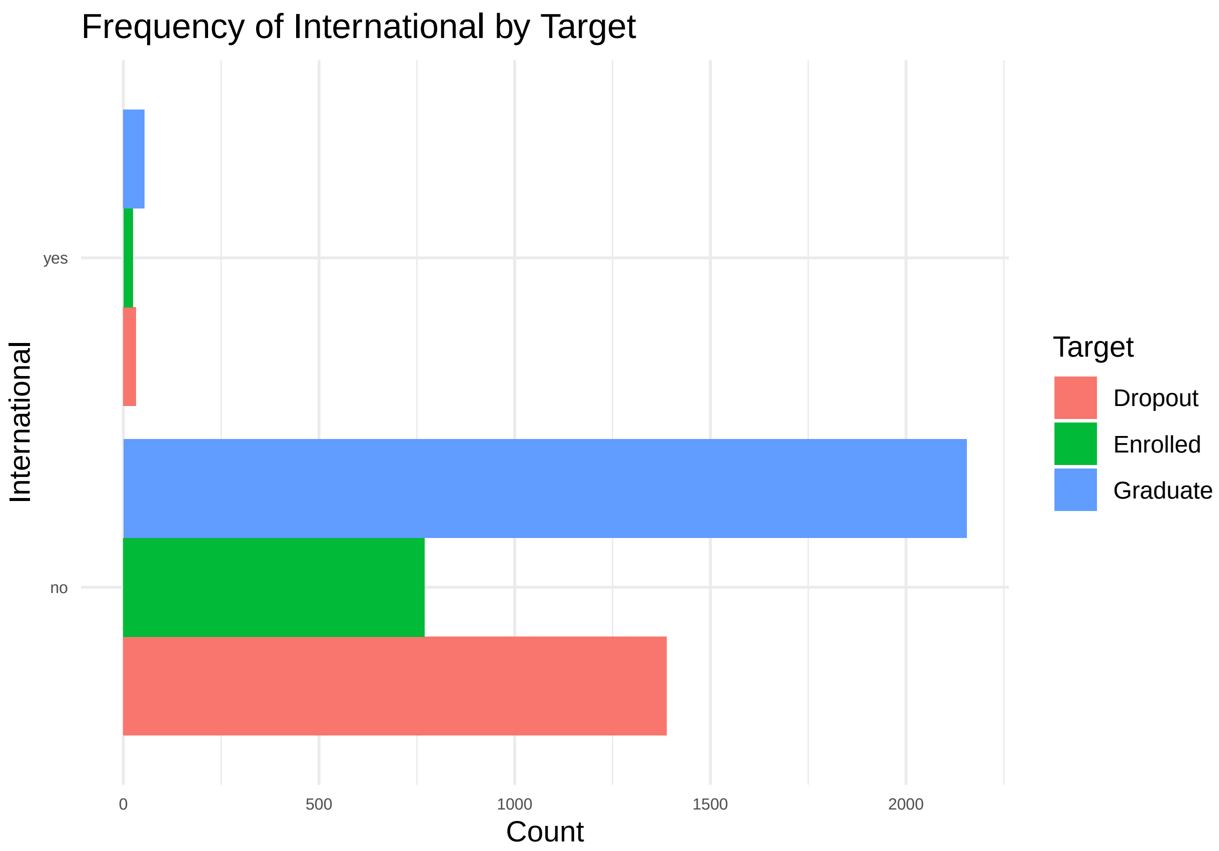

Frequency of Educational special needs by Target
Educational special needs
yes
no
0
500
1000
1500
2000
Count
Target
Dropout
Enrolled
Graduate


Frequency of Debtor by Target
Debtor
yes
no
0
500
1000
1500
2000
Count
Target
Dropout
Enrolled
Graduate

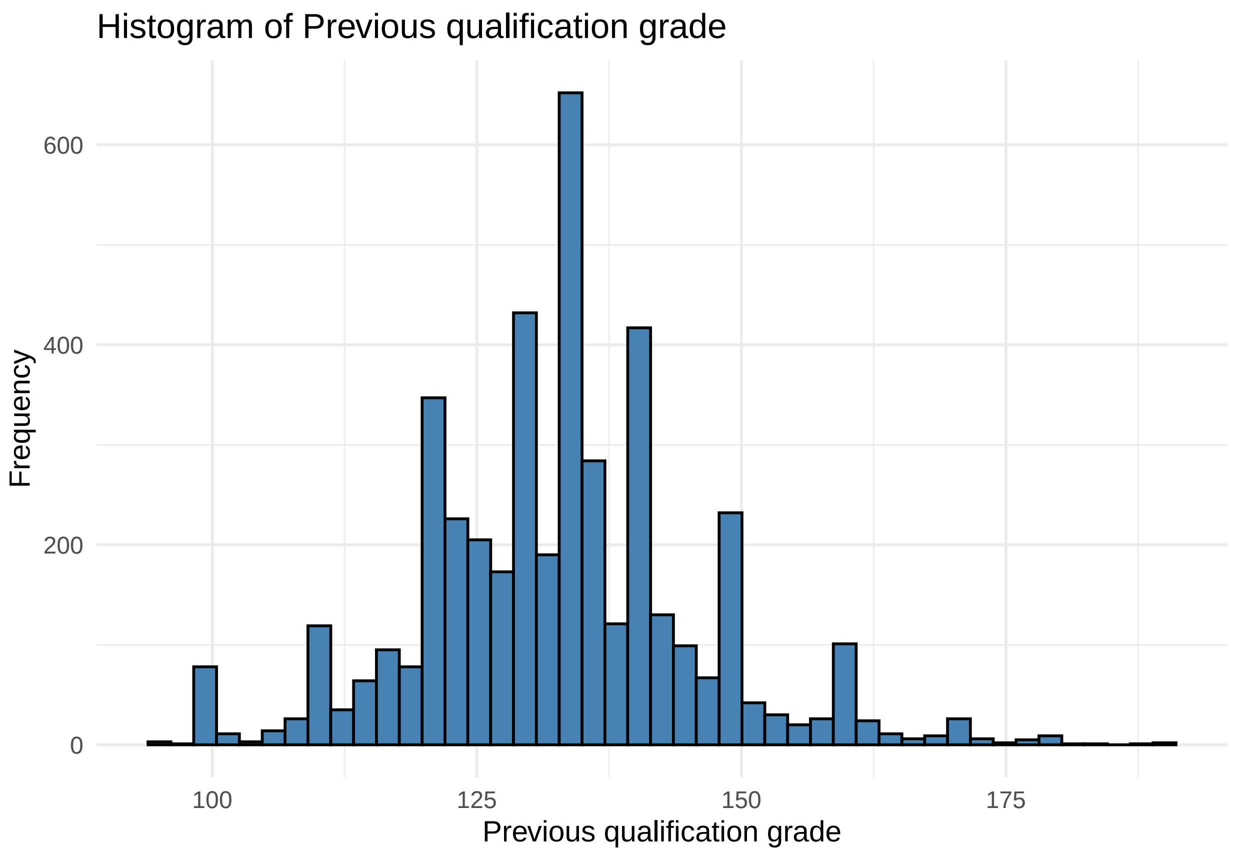

Frequency of Tuition fees up to date by Target
Tuition fees up to date
yes
no
0
500
1000
1500
2000
Count
Target
Dropout
Enrolled
Graduate

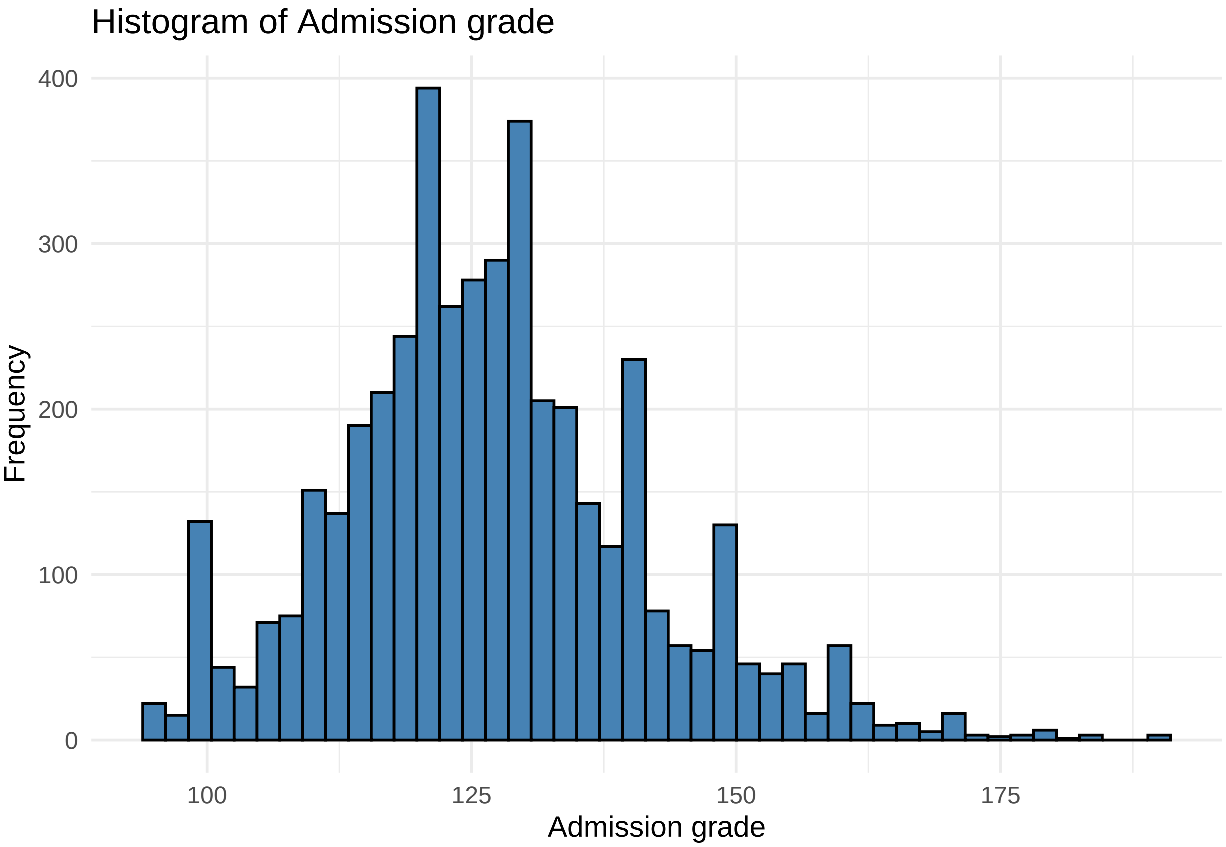

Frequency of Gender by Target
Gender
Male
Female
Count
0
500
1000
1500
Target
Dropout
Enrolled
Graduate

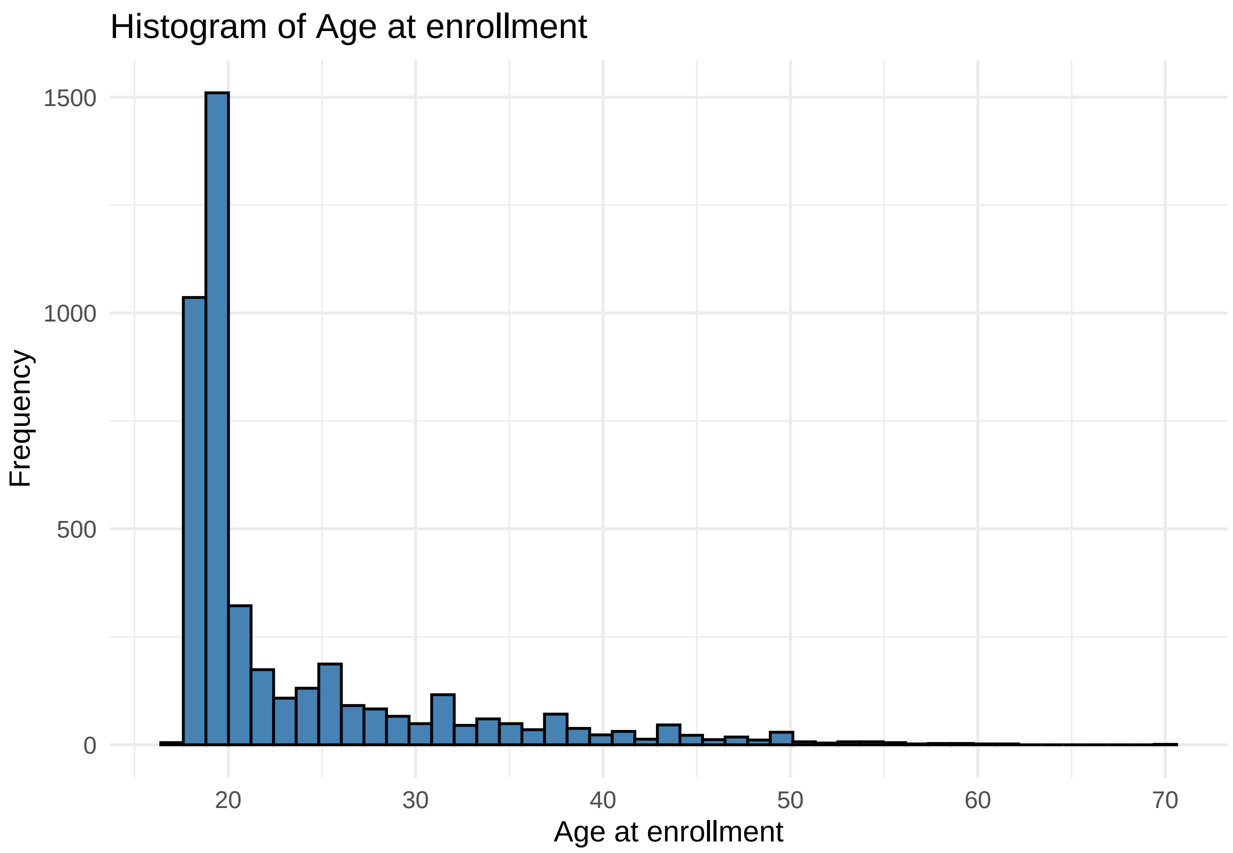

Frequency of Scholarship holder by Target
Scholarship holder
yes
no
0
500
1000
Count
Target
Dropout
Enrolled
Graduate

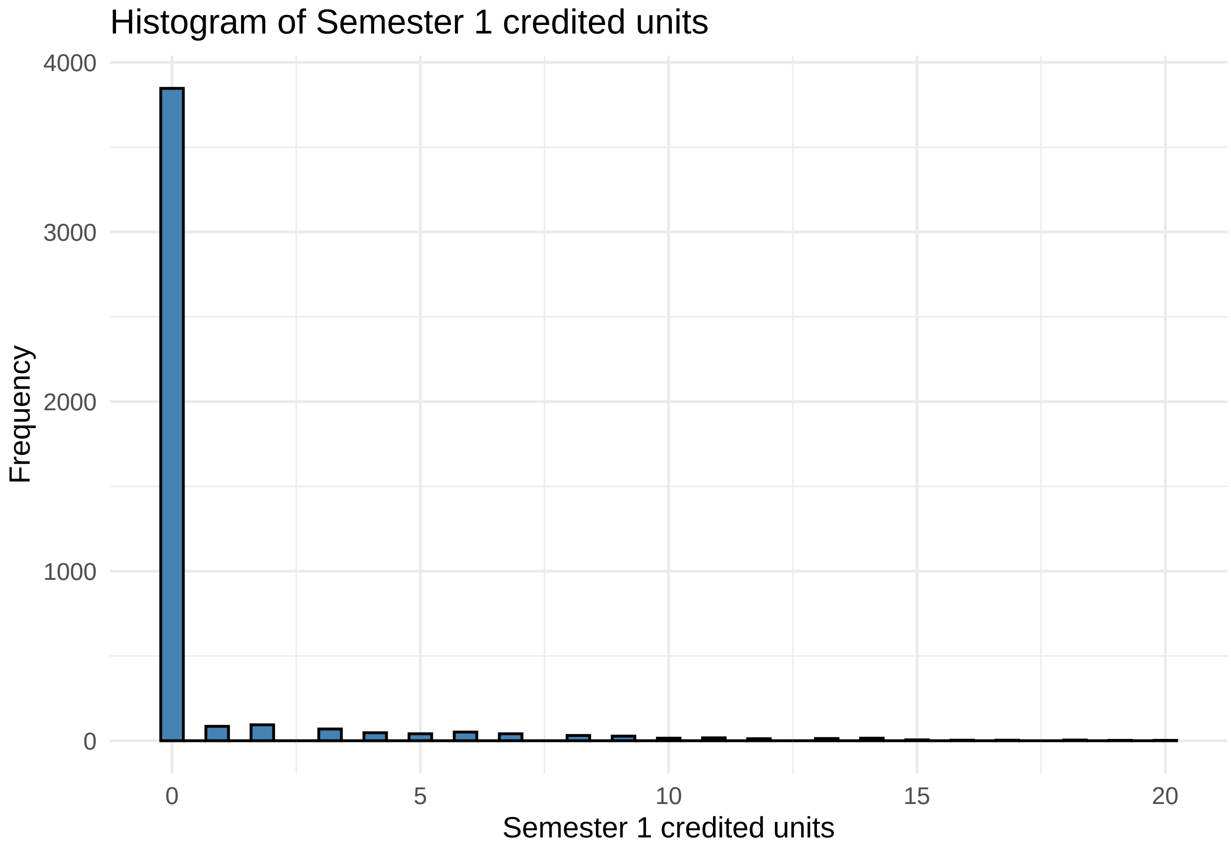


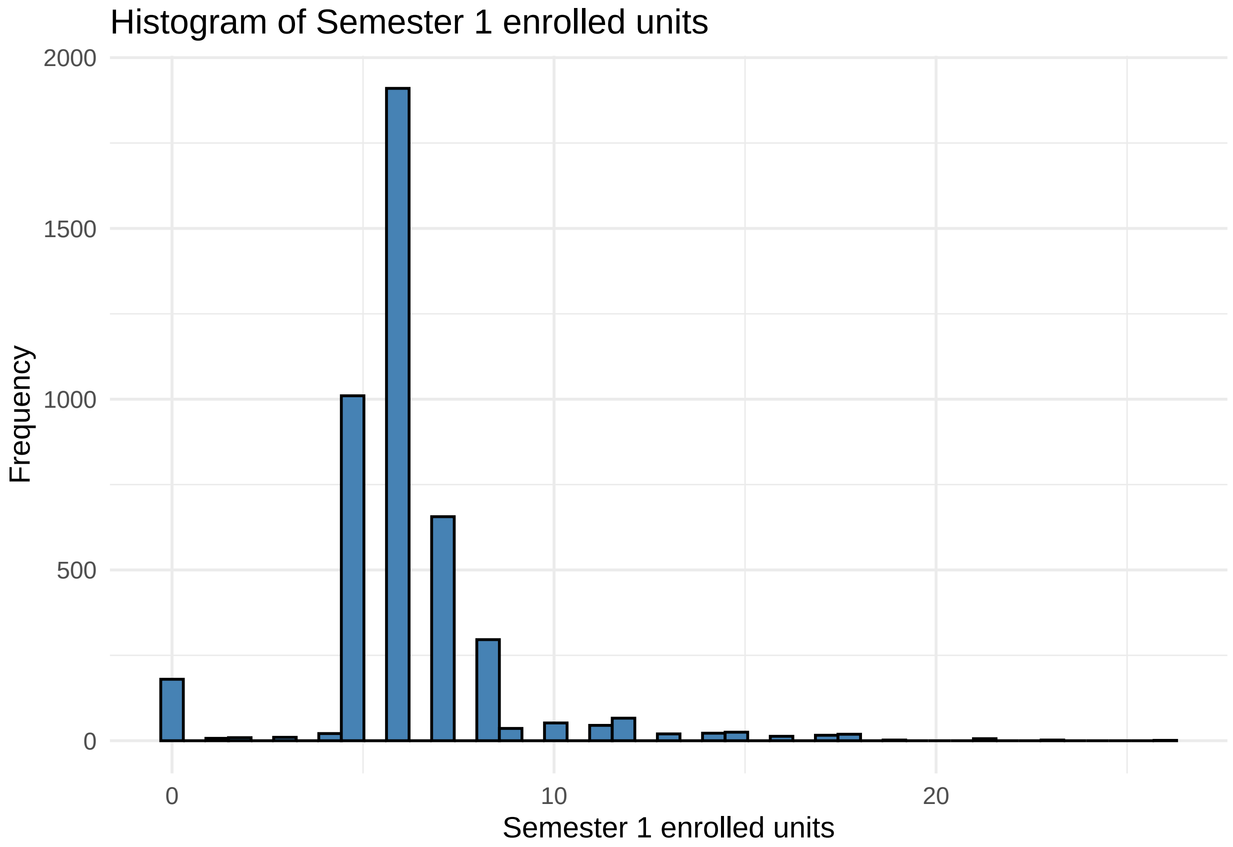
```r
numeric_cols <- names(data1)[sapply(data1, is.numeric)]

for (i in numeric_cols) {
  plot <- ggplot(data1, aes(x = .data[[i]])) +
    geom_histogram(bins = 45, fill = "steelblue", color = "black") +
    labs(title = paste("Histogram of", i), x = i, y = "Frequency") +
    theme_minimal()
  print(plot)
}
```

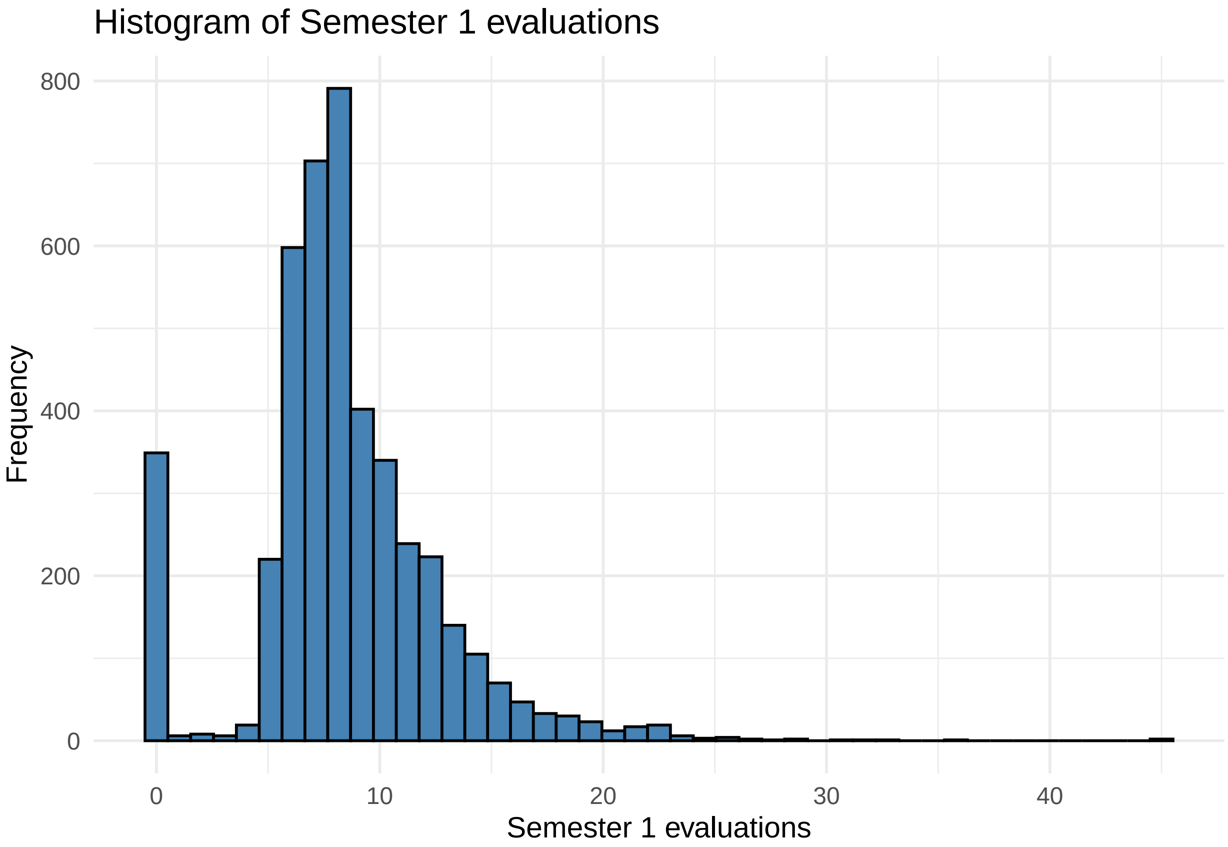

Histogram of Previous qualification grade
Frequency
600
400
200
0
100
125
150
175
Previous qualification grade

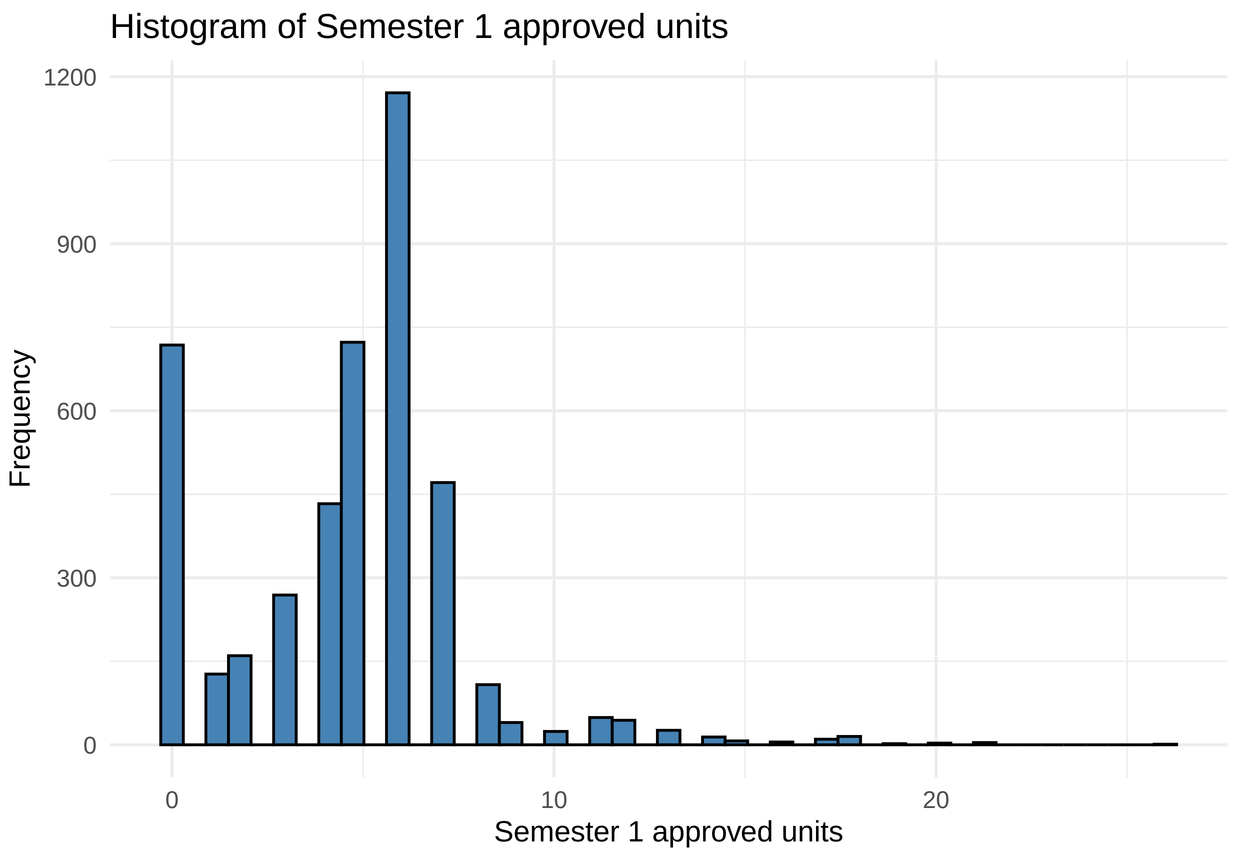

Histogram of Admission grade
Frequency
400
300
200
100
0
100
125
150
175
Admission grade


Histogram of Age at enrollment
Frequency
1500
1000
500
0
20
30
40
50
60
70
Age at enrollment

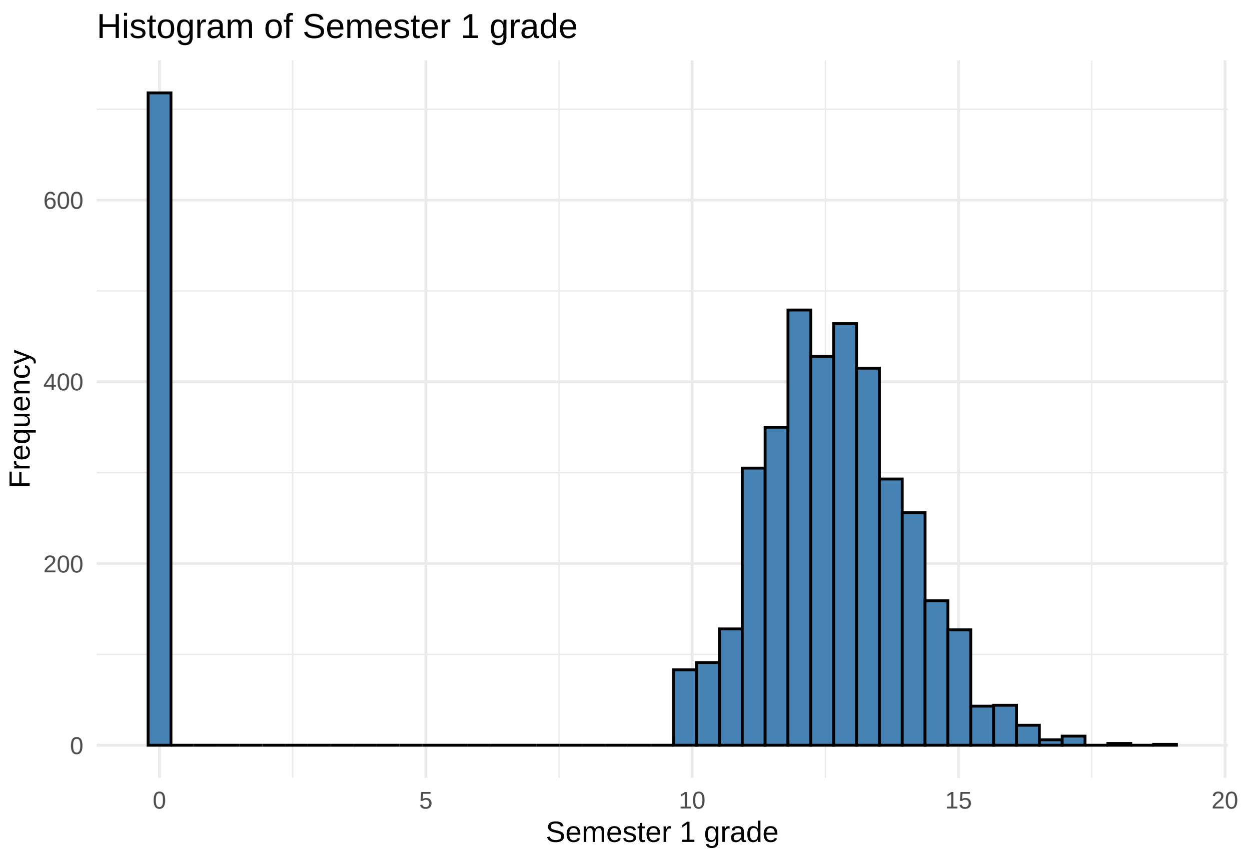

Histogram of Semester 1 credited units
Frequency
4000
3000
2000
1000
0
0
5
10
15
20
Semester 1 credited units

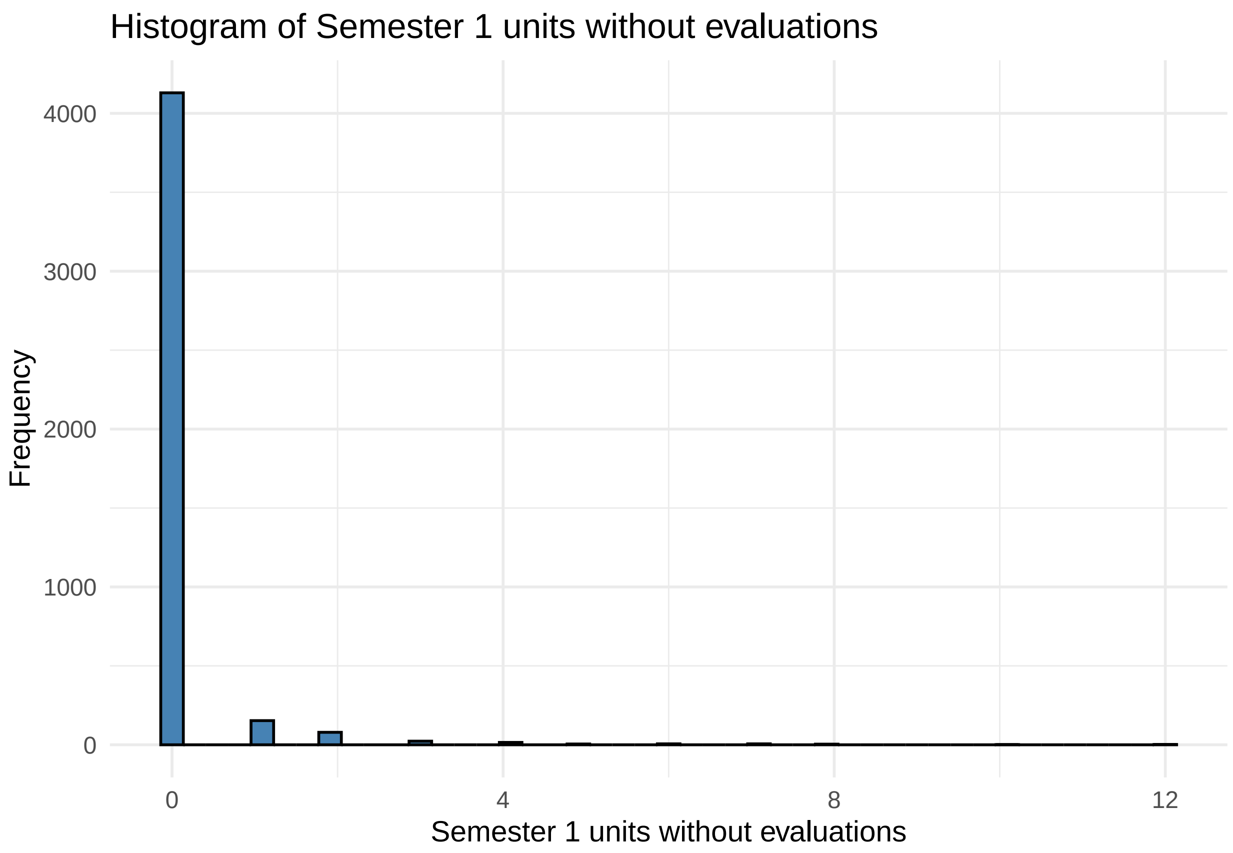

Histogram of Semester 1 enrolled units
Frequency
2000
1500
1000
500
0
0
10
20
Semester 1 enrolled units

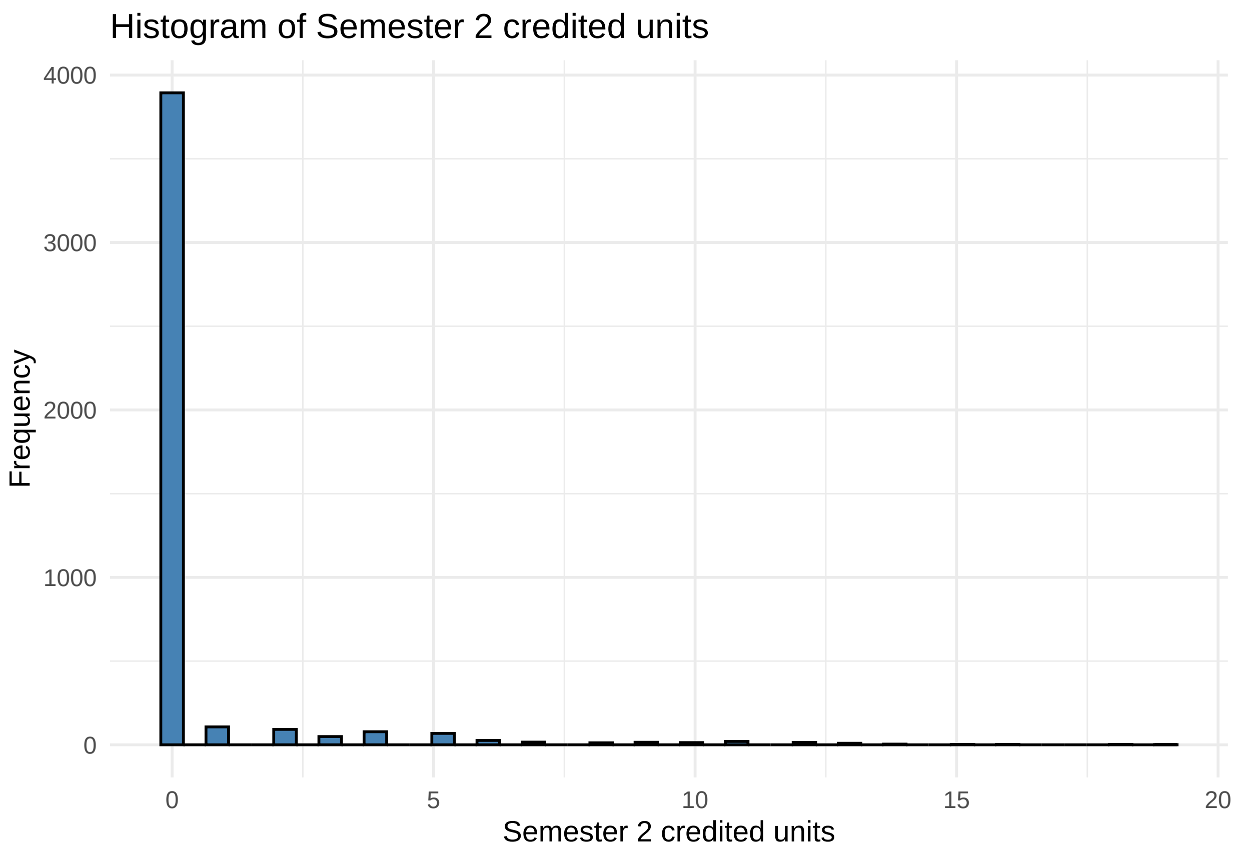

Histogram of Semester 1 evaluations
Frequency
800
600
400
200
0
0
10
20
30
40
Semester 1 evaluations

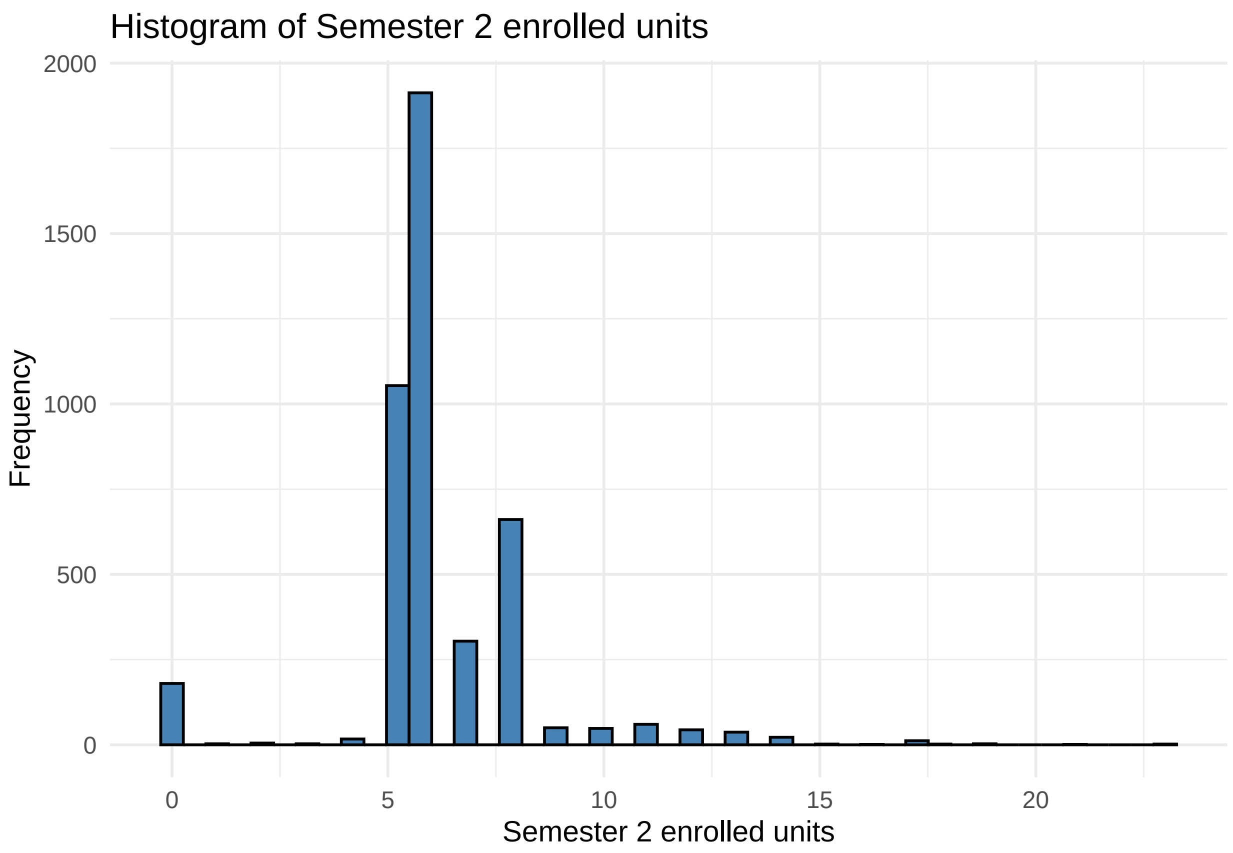


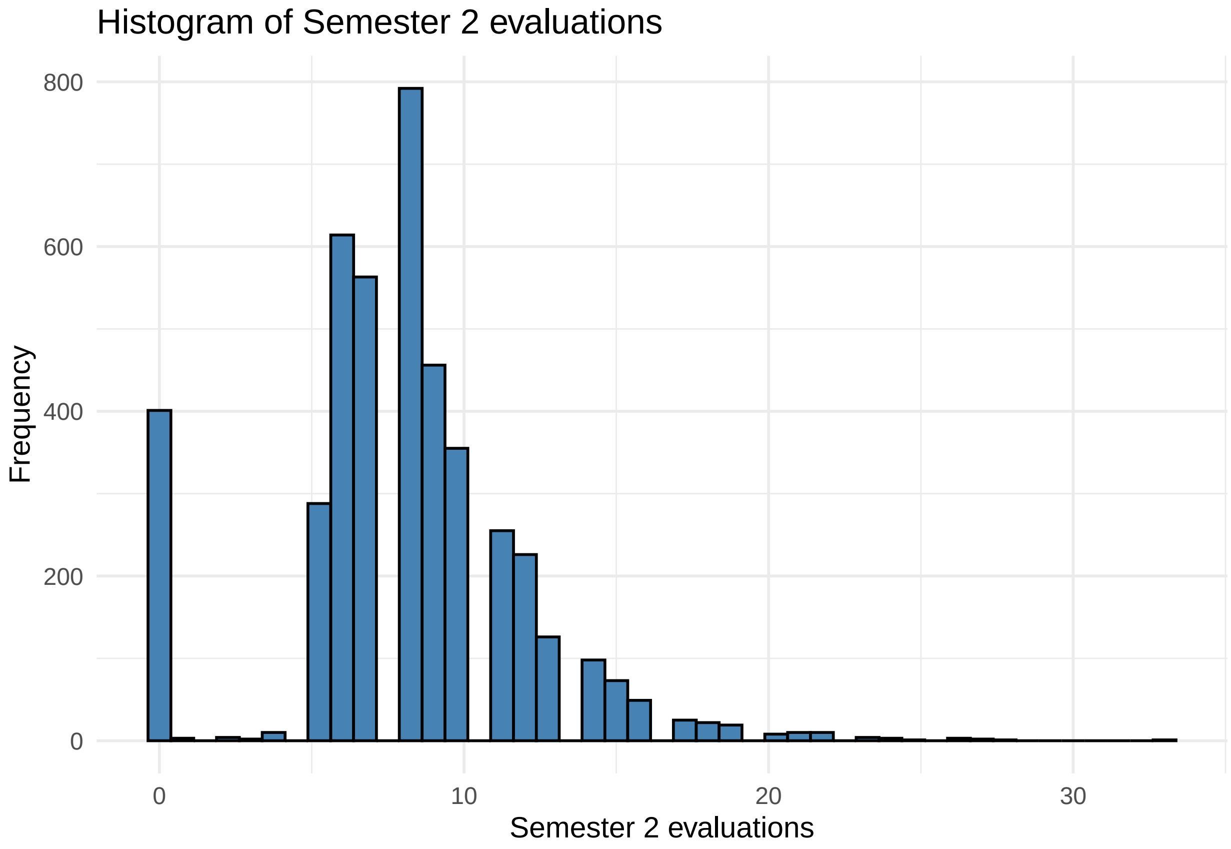
Histogram of Semester 1 approved units
Frequency
1200
900
600
300
0
0
10
20
Semester 1 approved units

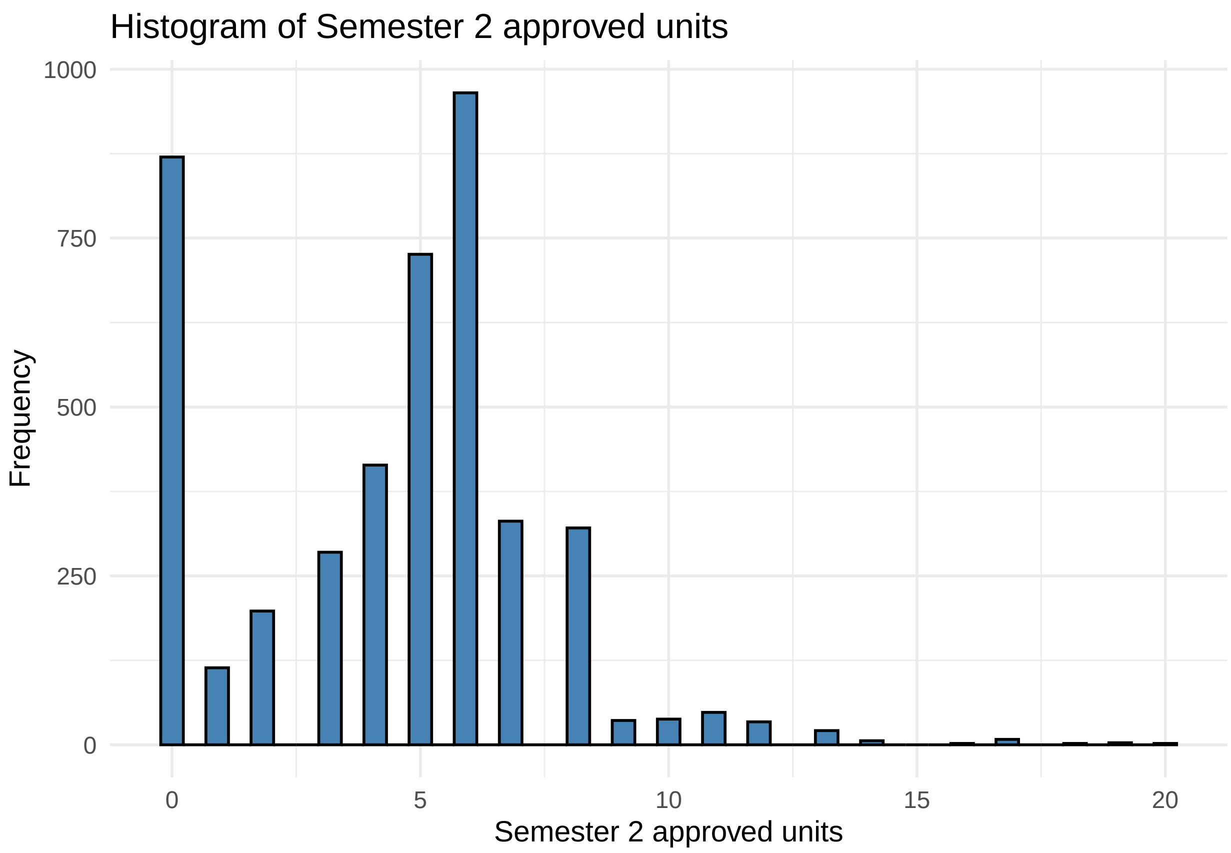

Histogram of Semester 1 grade
Frequency
600
400
200
0
0
5
10
15
20
Semester 1 grade

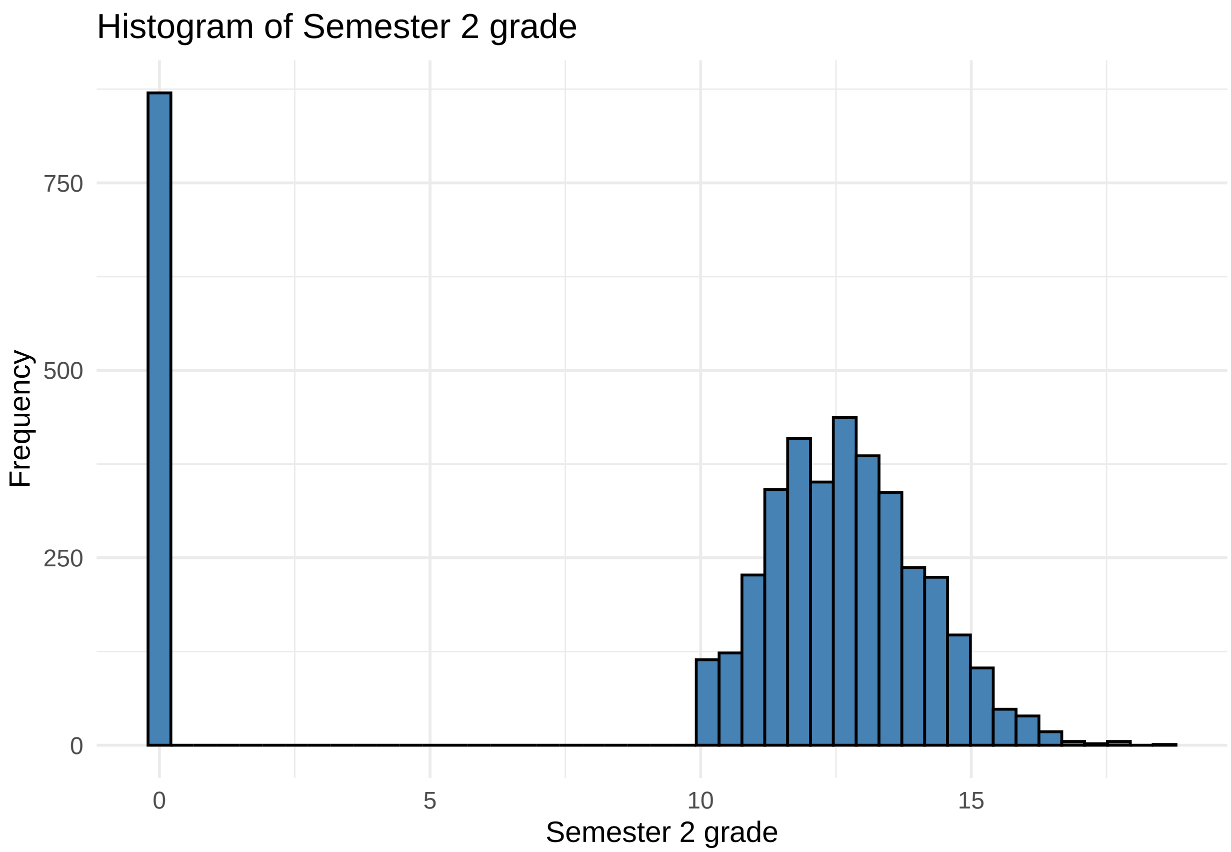

Histogram of Semester 1 units without evaluations
Frequency
4000
3000
2000
1000
0
0
4
8
12
Semester 1 units without evaluations

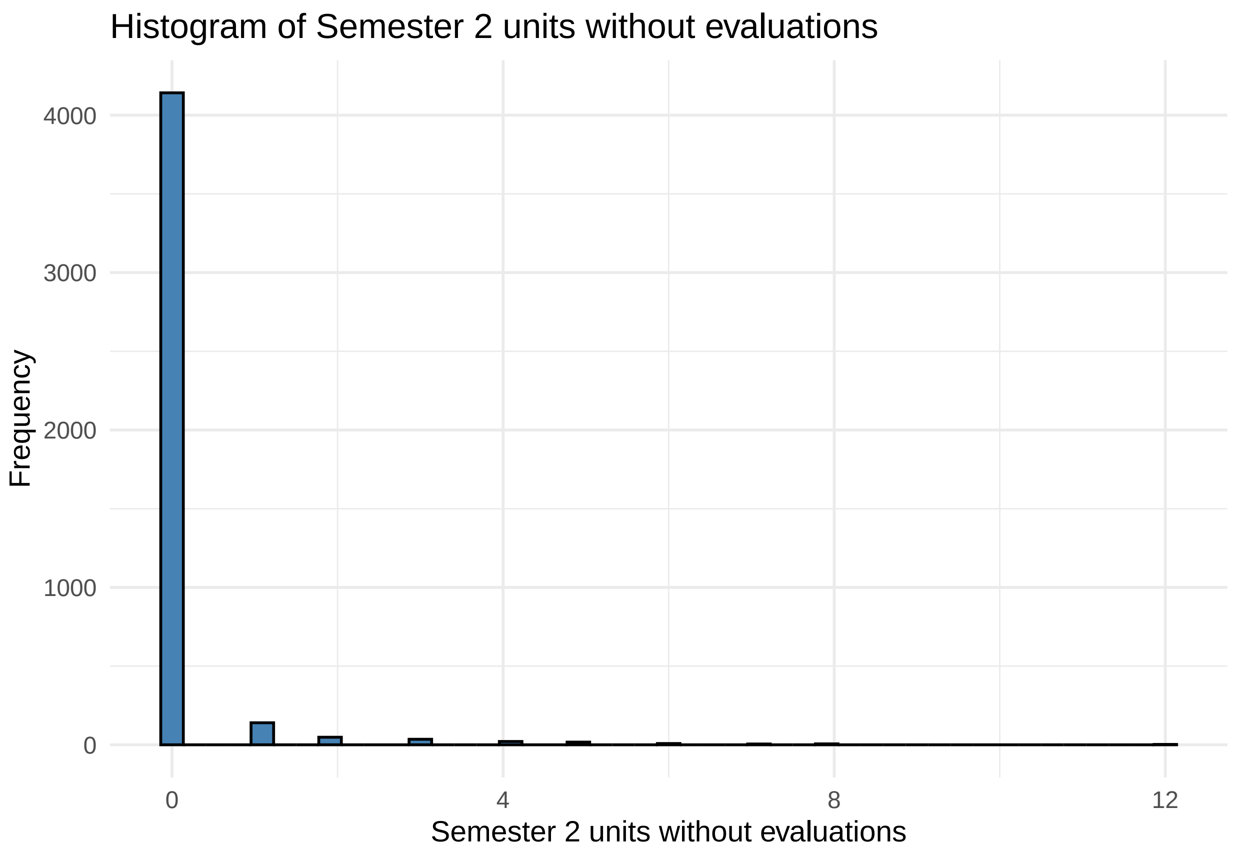

Histogram of Semester 2 credited units
Frequency
4000
3000
2000
1000
0
0
5
10
15
20
Semester 2 credited units

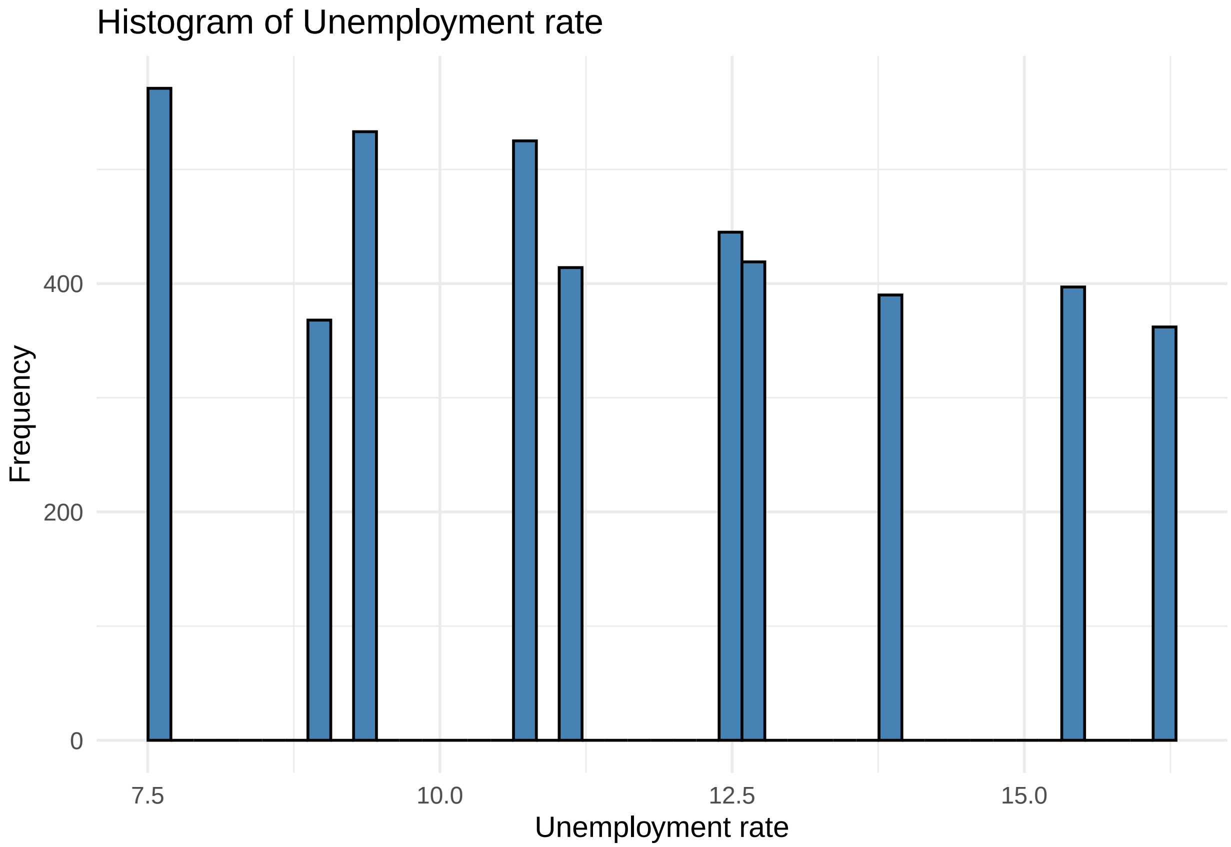

Histogram of Semester 2 enrolled units
Frequency
2000
1500
1000
500
0
0
5
10
15
20
Semester 2 enrolled units

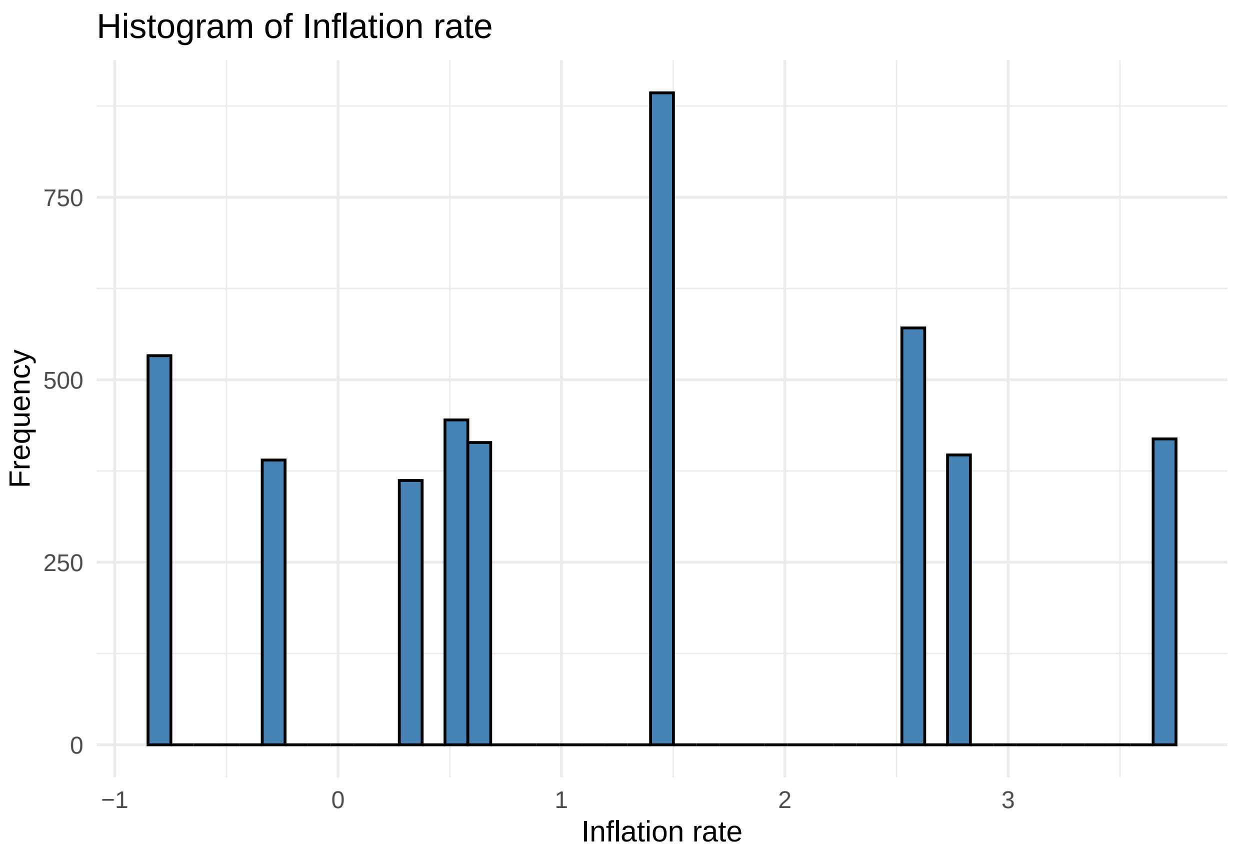

Histogram of Semester 2 evaluations
Frequency
800
600
400
200
0
0
10
20
30
Semester 2 evaluations

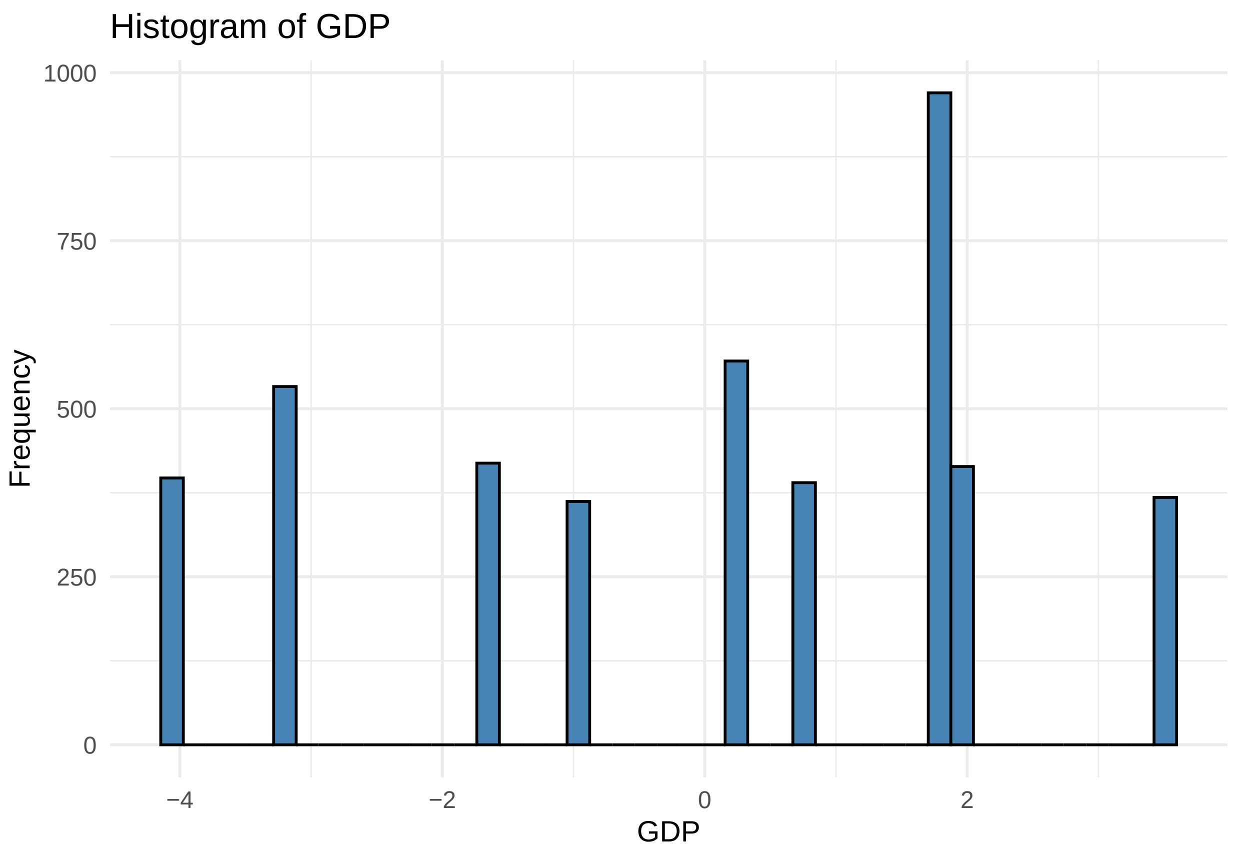

Histogram of Semester 2 approved units
Frequency
1000
750
500
250
0
0
5
10
15
20
Semester 2 approved units


Histogram of Semester 2 grade
Frequency
750
500
250
0
0
5
10
15
Semester 2 grade

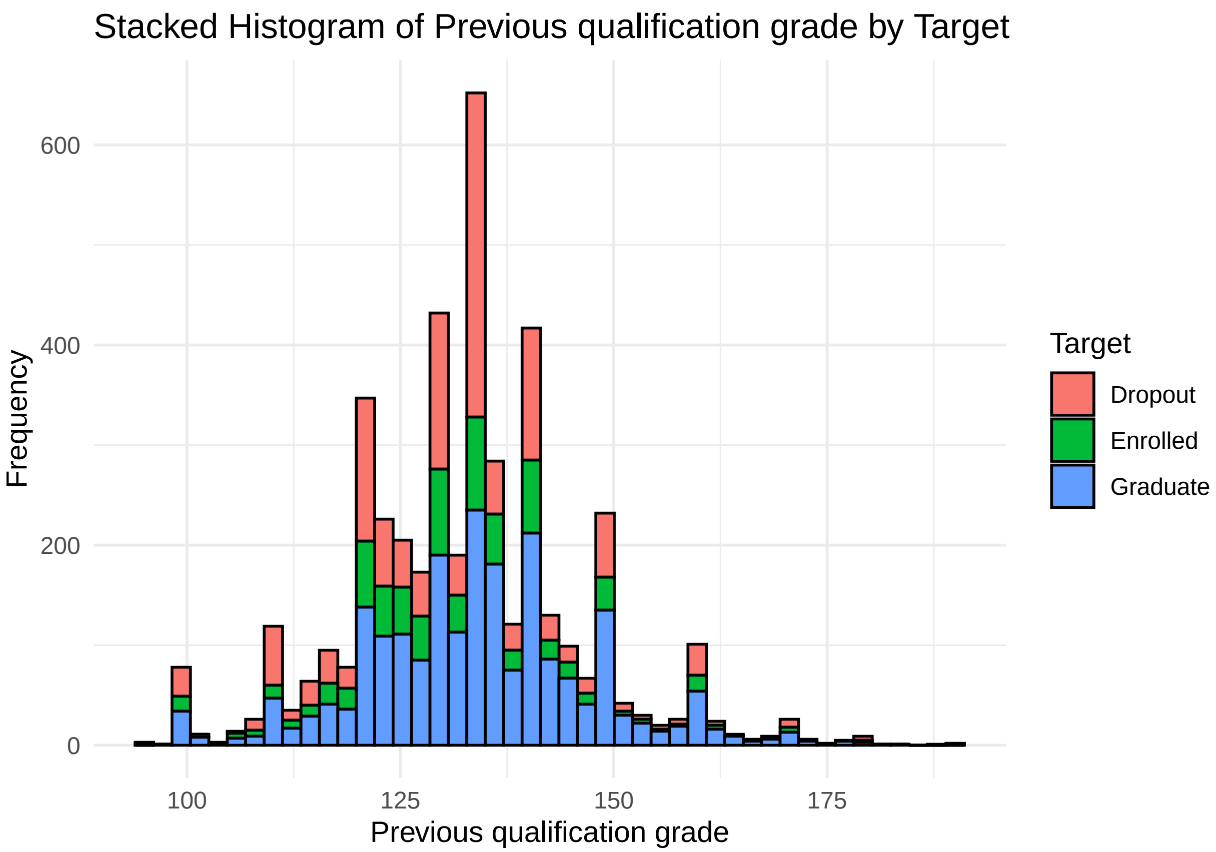

Histogram of Semester 2 units without evaluations
Frequency
4000
3000
2000
1000
0
0
4
8
12
Semester 2 units without evaluations

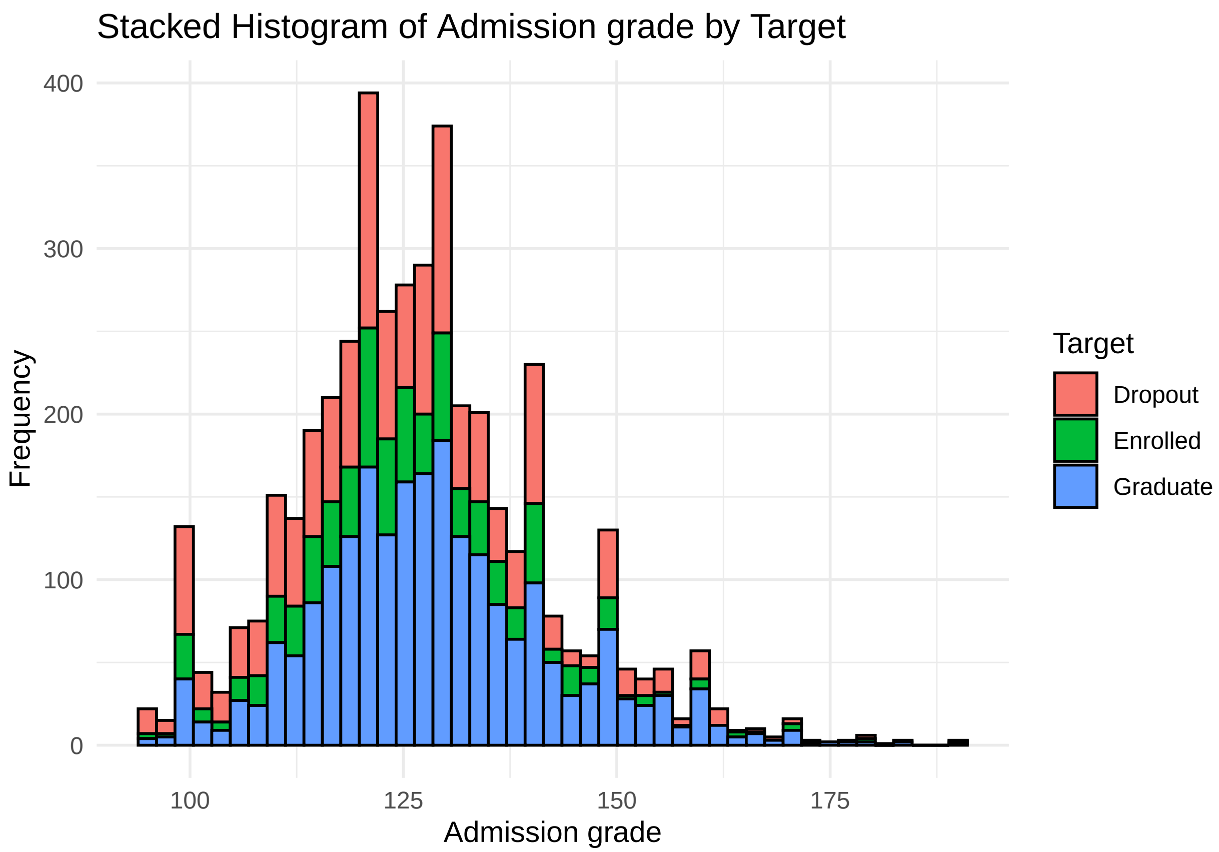

Histogram of Unemployment rate
Frequency
0
200
400
7.5
10.0
12.5
15.0
Unemployment rate

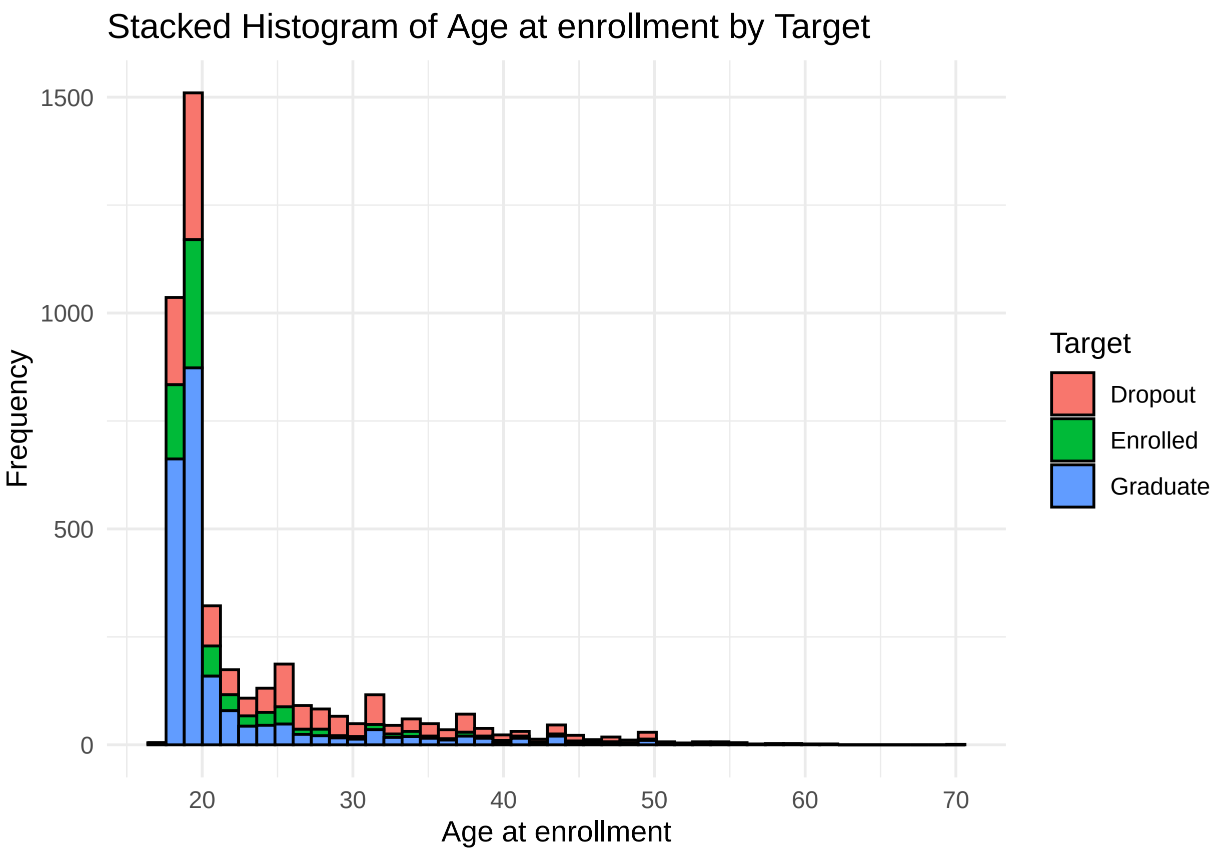

Histogram of Inflation rate
Frequency
750
500
250
0
−1
0
1
2
3
Inflation rate



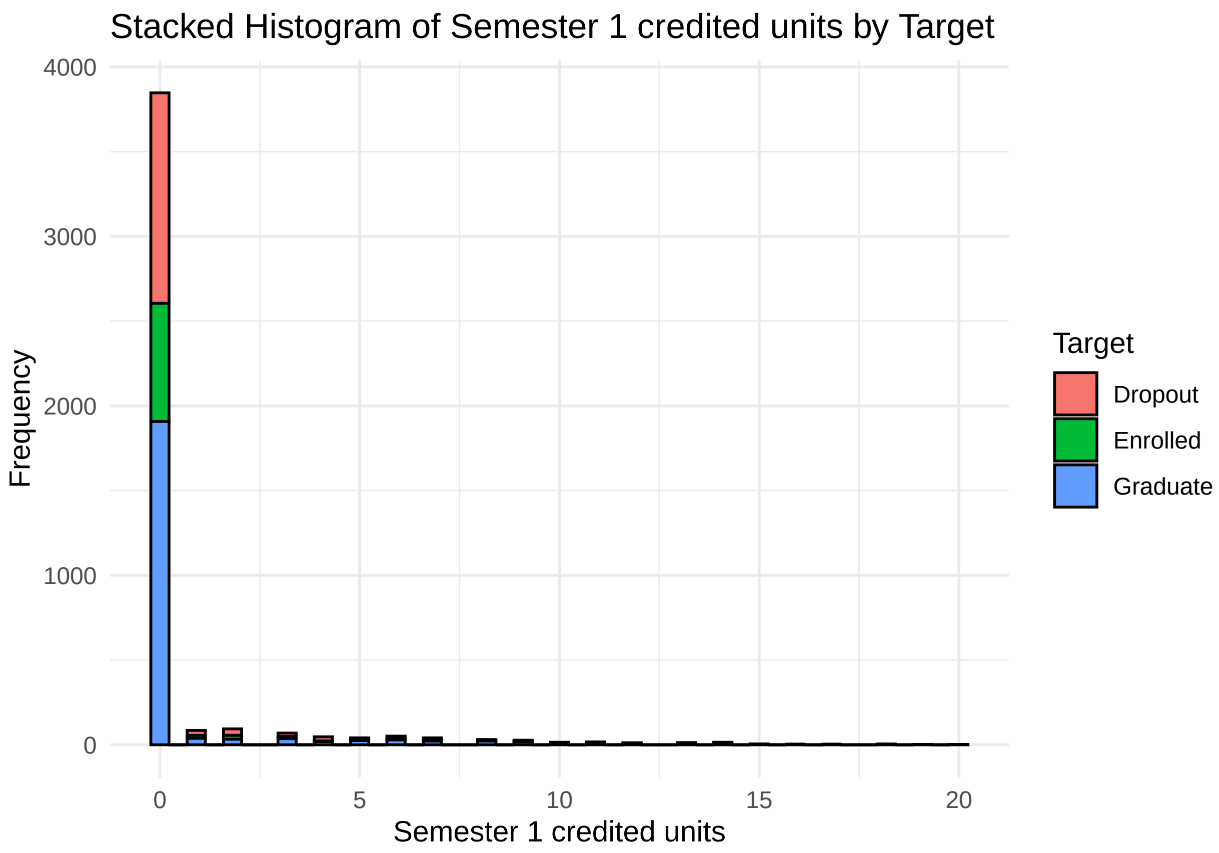
```
for (i in numeric_cols) {
  plot <- ggplot(data1, aes(x = .data[[i]], fill = Target)) +
    geom_histogram(bins = 45, color = "black") +
    labs(title = paste("Stacked Histogram of", i, "by Target"), x = i, y = "Frequency") +
    theme_minimal()
  print(plot)
}
```

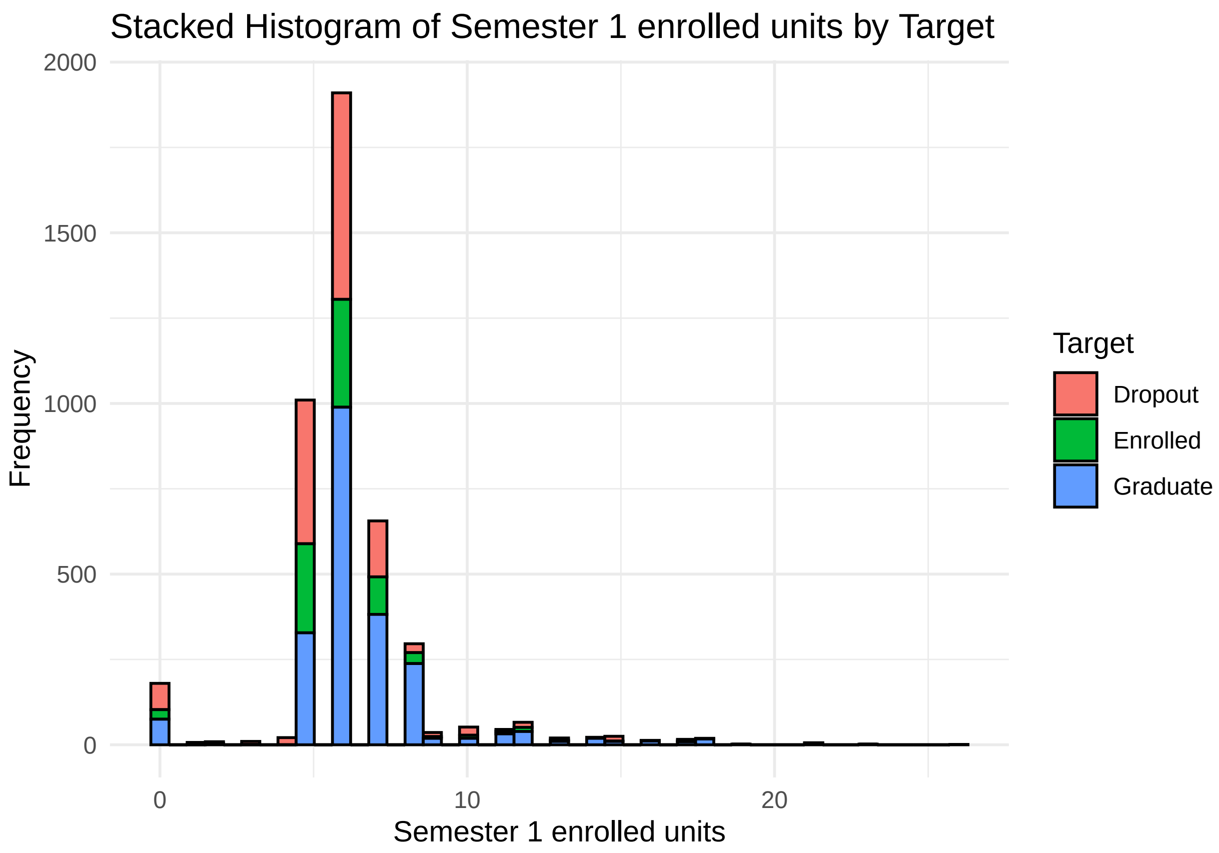

Stacked Histogram of Previous qualification grade by Target
Frequency
600
400
200
0
100
125
150
175
Previous qualification grade
Target
Dropout
Enrolled
Graduate

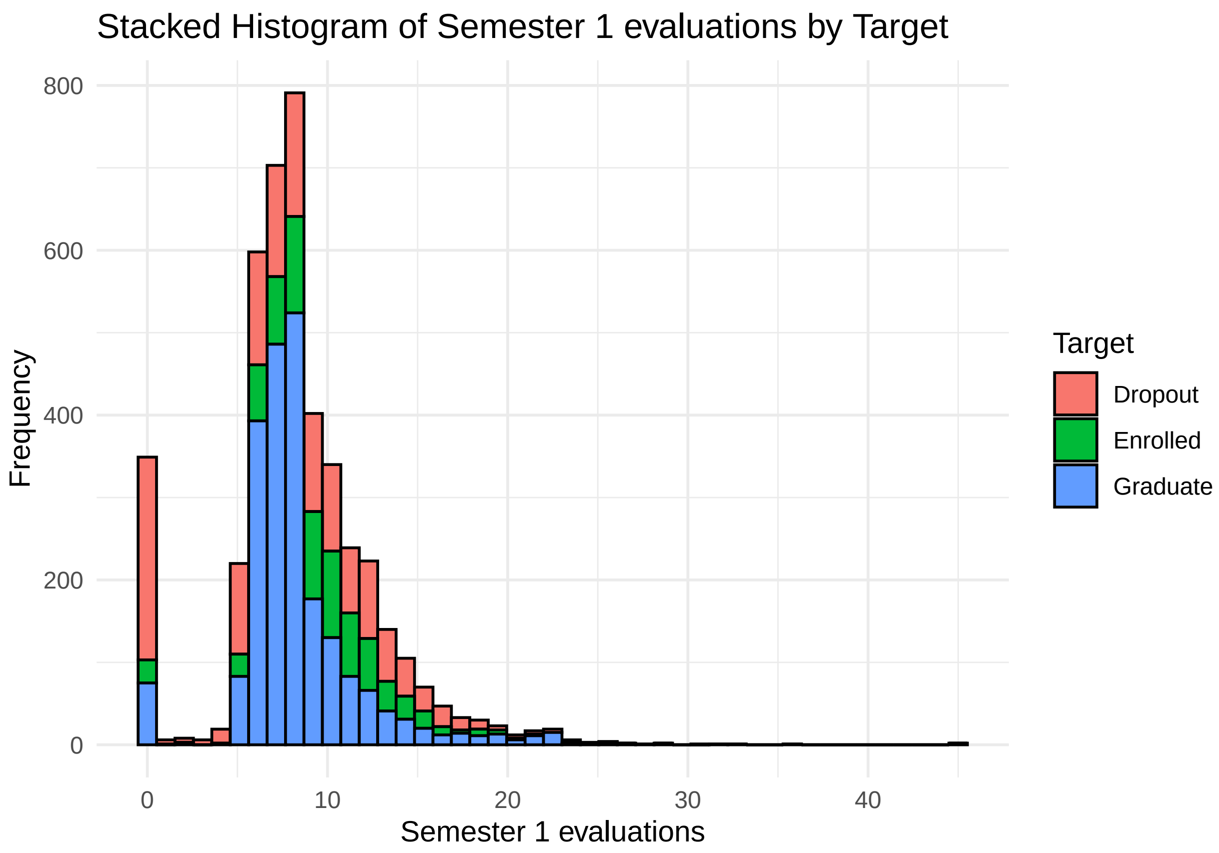

Stacked Histogram of Admission grade by Target
Frequency
400
300
200
100
0
100
125
150
175
Admission grade
Target
Dropout
Enrolled
Graduate

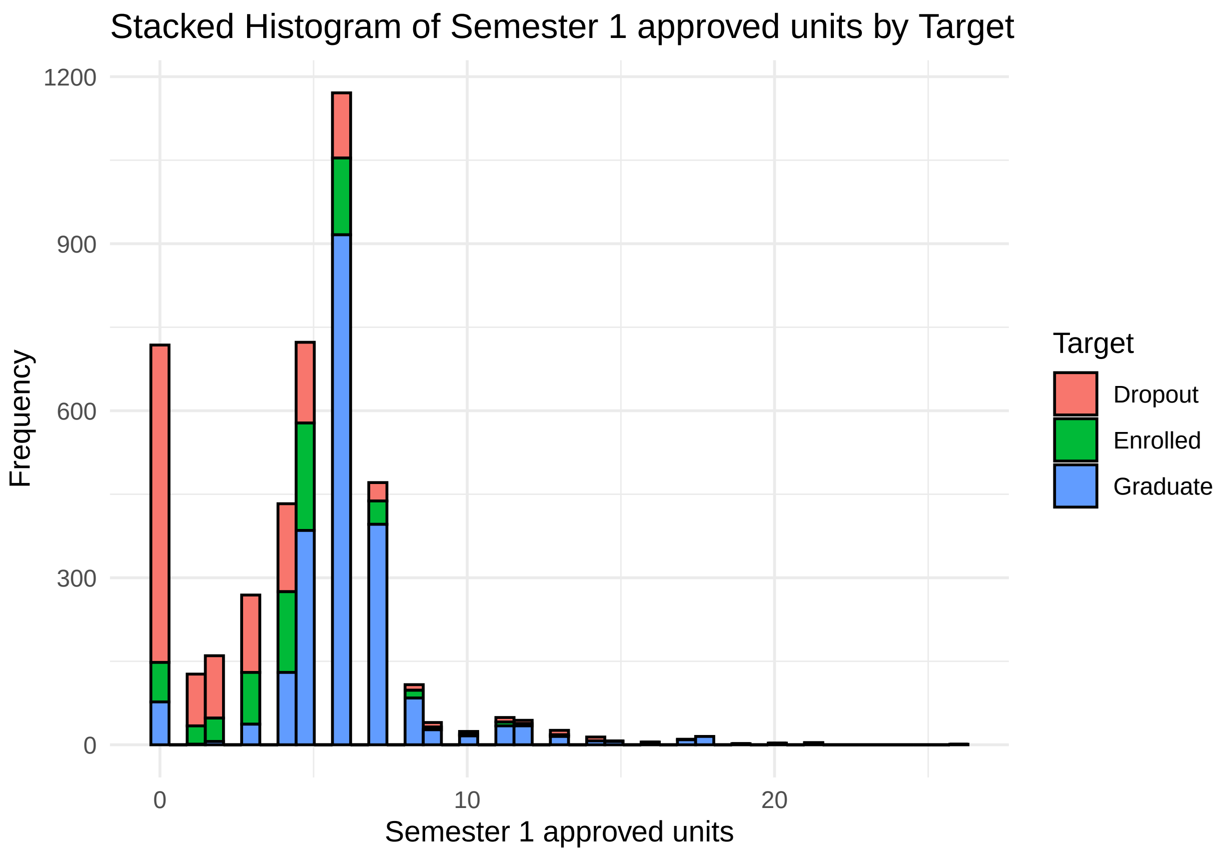


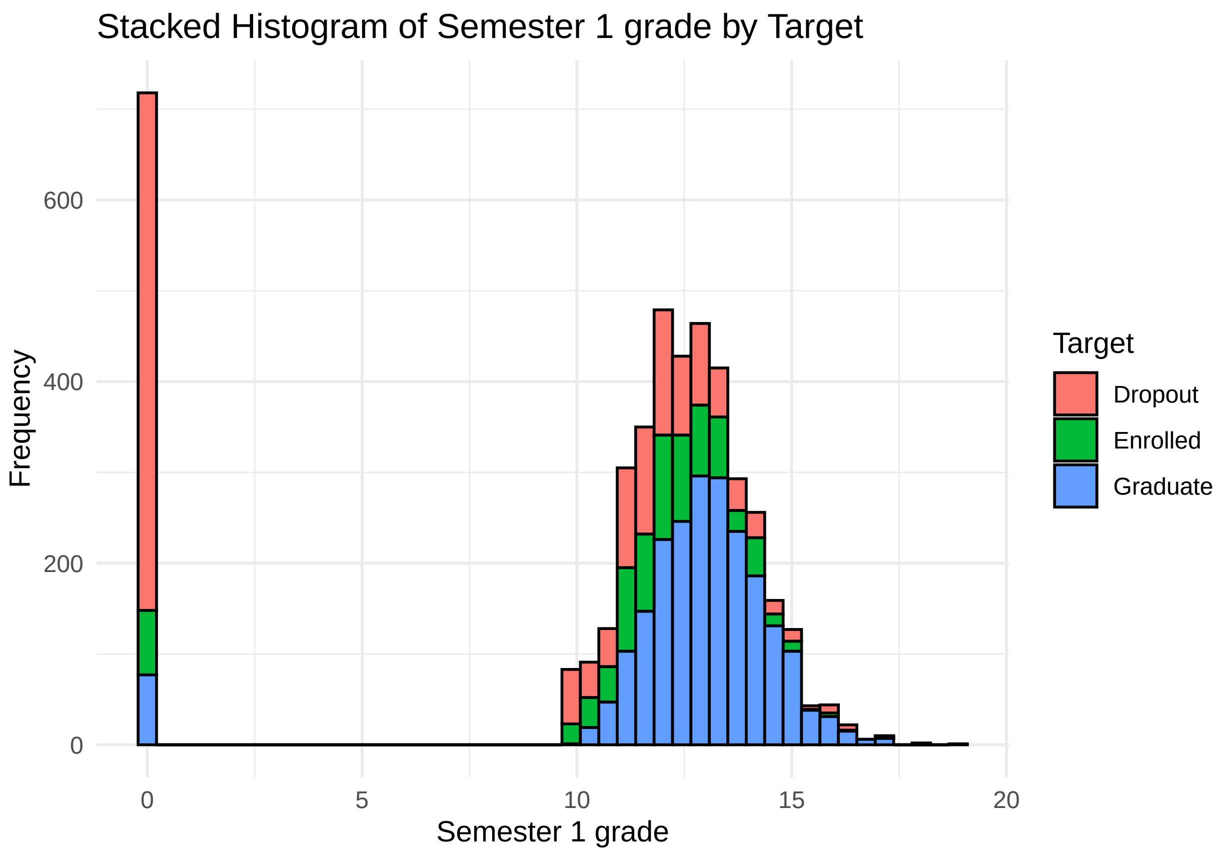
Stacked Histogram of Age at enrollment by Target
Frequency
1500
1000
500
0
20
30
40
50
60
70
Age at enrollment
Target
Dropout
Enrolled
Graduate

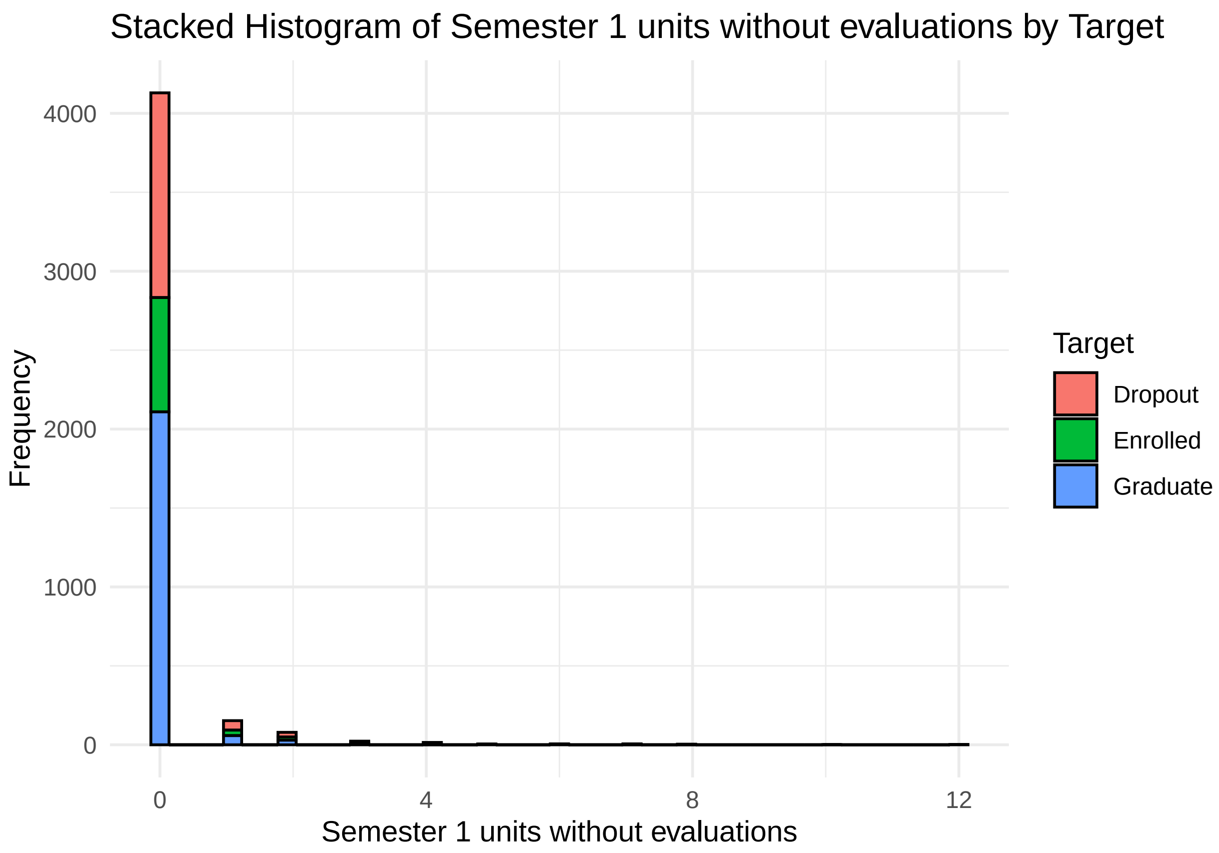

Stacked Histogram of Semester 1 credited units by Target
Frequency
4000
3000
2000
1000
0
0
5
10
15
20
Semester 1 credited units
Target
Dropout
Enrolled
Graduate

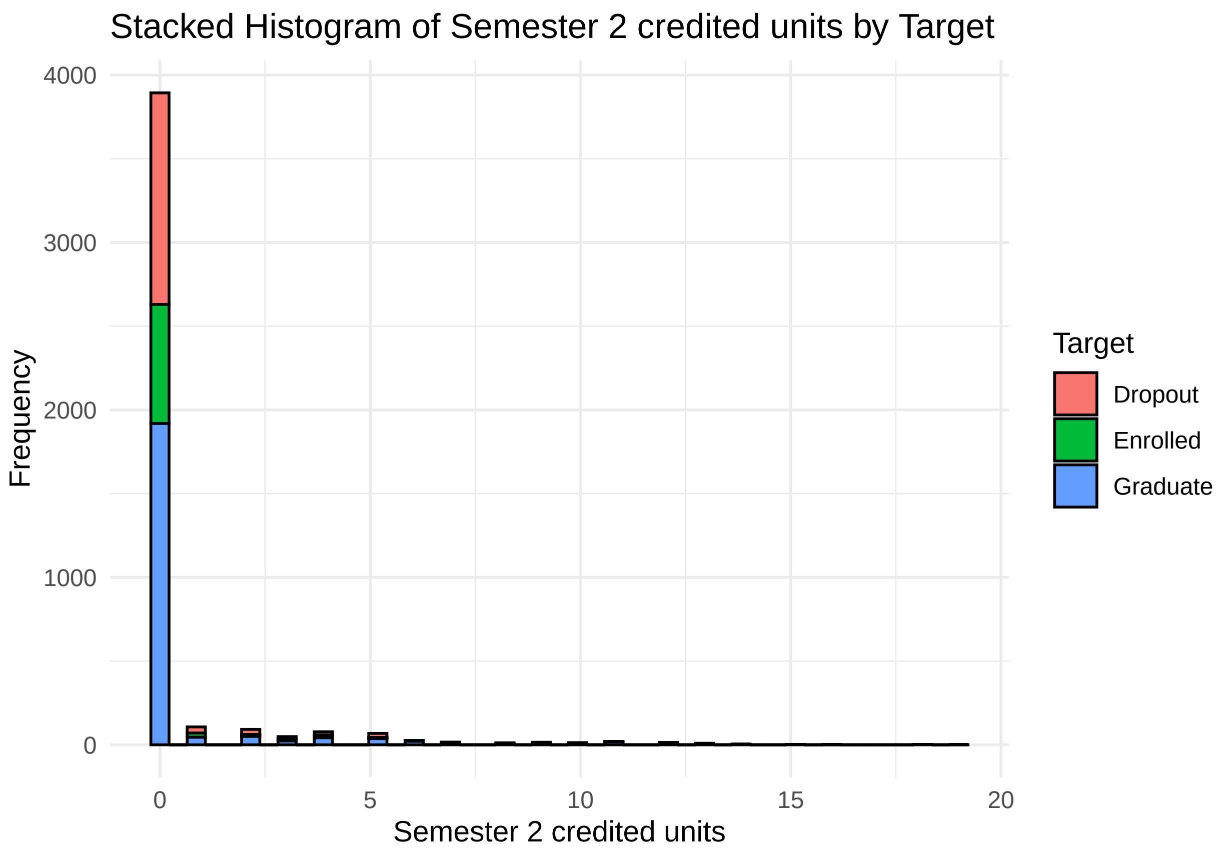

Stacked Histogram of Semester 1 enrolled units by Target
Frequency
2000
1500
1000
500
0
0
10
20
Semester 1 enrolled units
Target
Dropout
Enrolled
Graduate

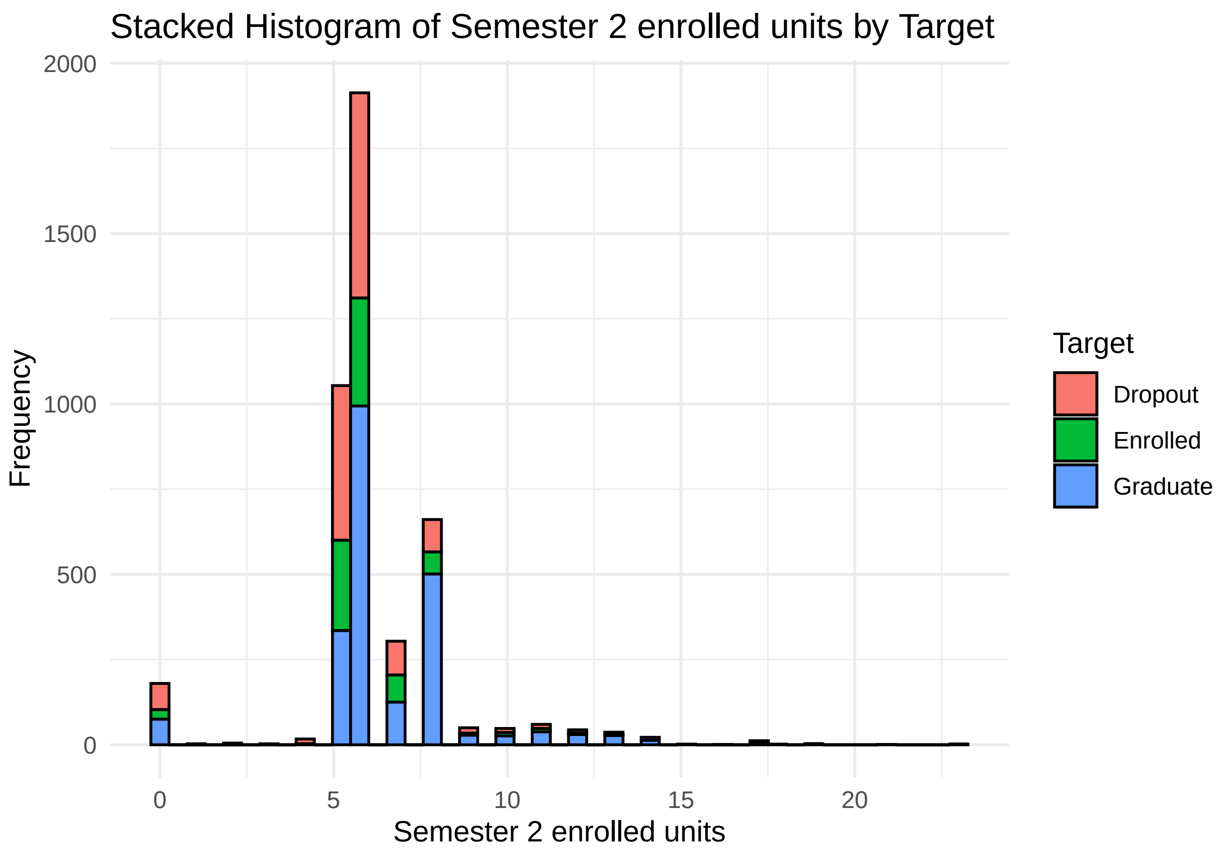

Stacked Histogram of Semester 1 evaluations by Target
Frequency
800
600
400
200
0
0
10
20
30
40
Semester 1 evaluations
Target
Dropout
Enrolled
Graduate

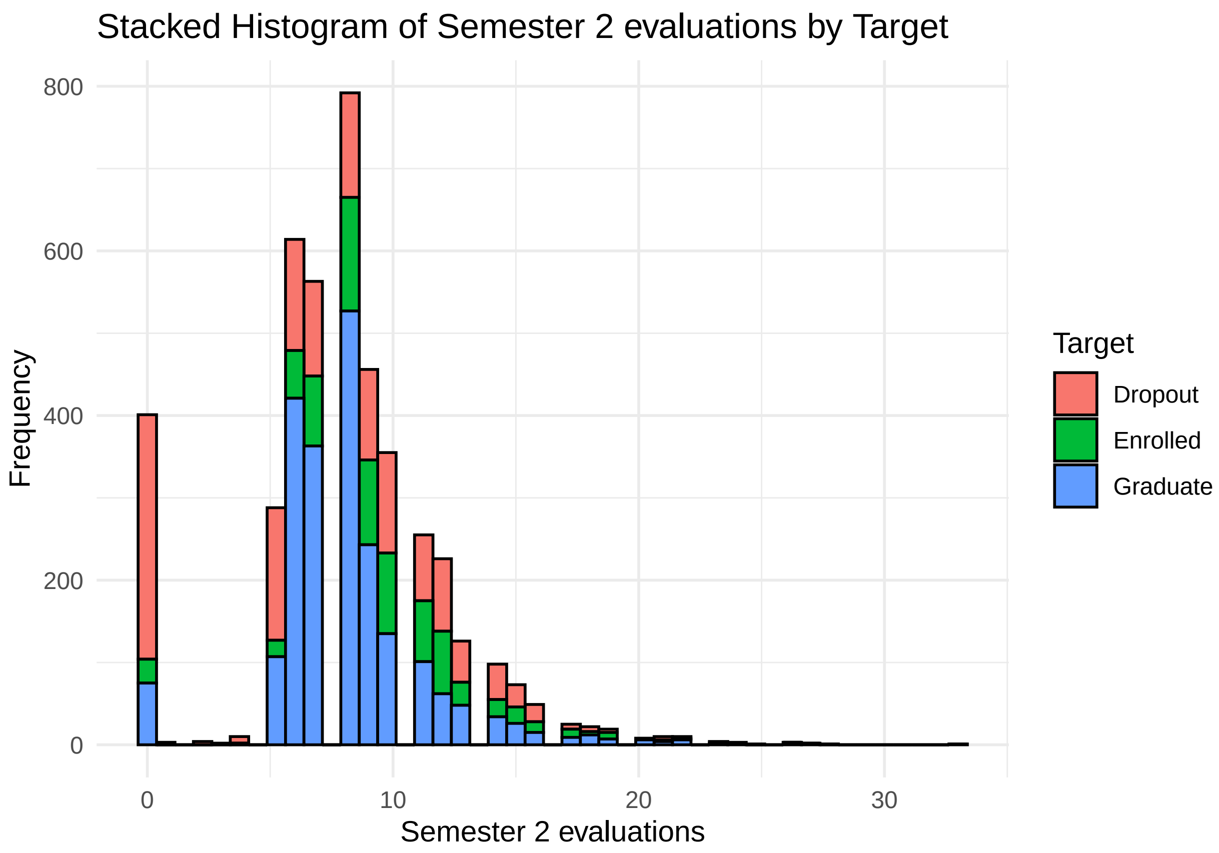

Stacked Histogram of Semester 1 approved units by Target
Frequency
1200
900
600
300
0
0
10
20
Semester 1 approved units
Target
Dropout
Enrolled
Graduate

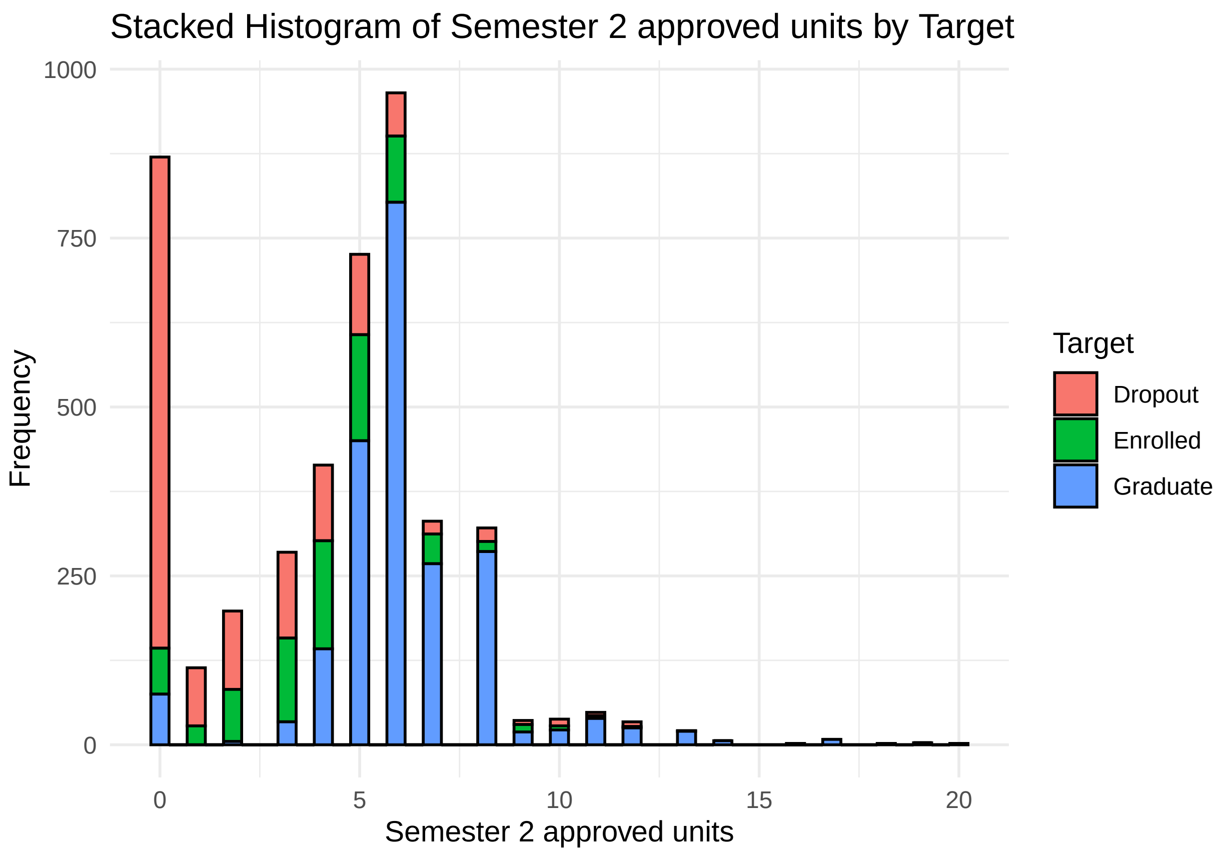

Stacked Histogram of Semester 1 grade by Target
Frequency
600
400
200
0
0
5
10
15
20
Semester 1 grade
Target
Dropout
Enrolled
Graduate

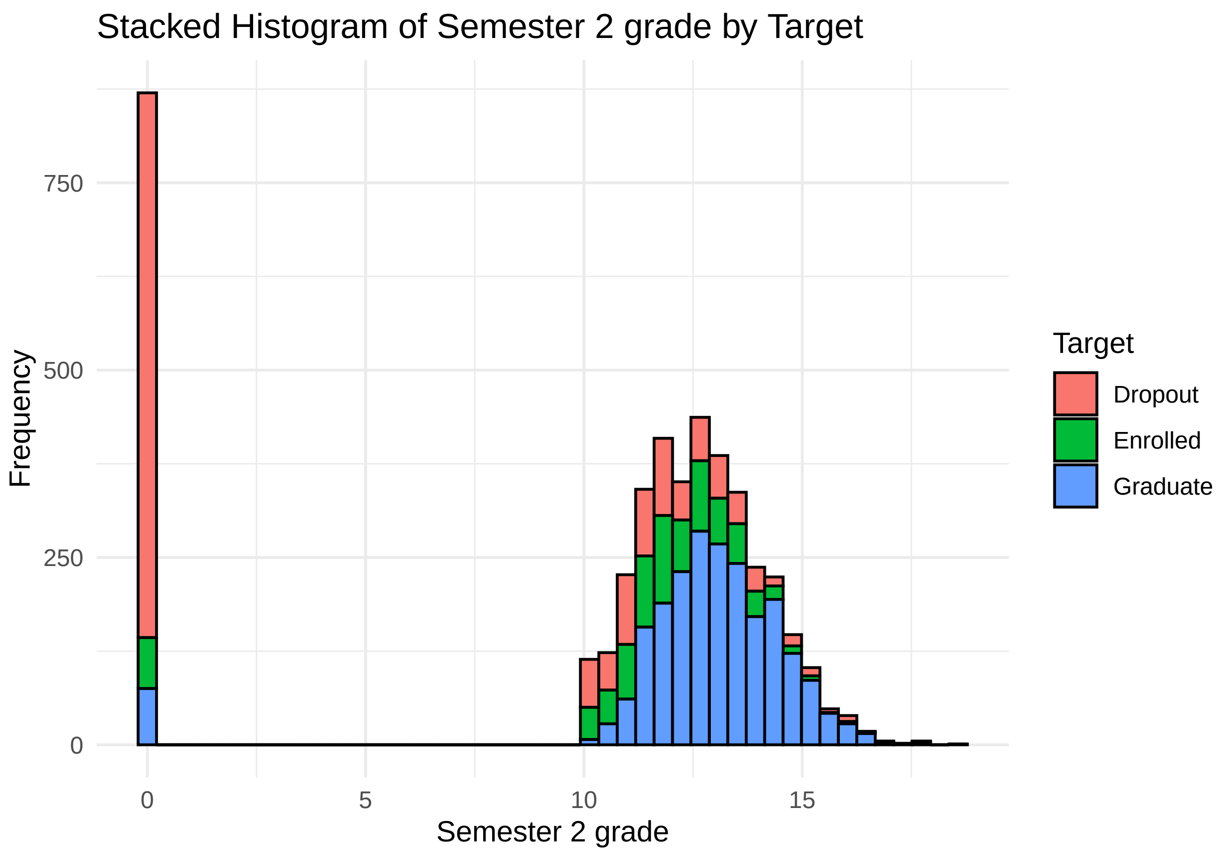

Stacked Histogram of Semester 1 units without evaluations by Target
Frequency
4000
3000
2000
1000
0
0
4
8
12
Semester 1 units without evaluations
Target
Dropout
Enrolled
Graduate

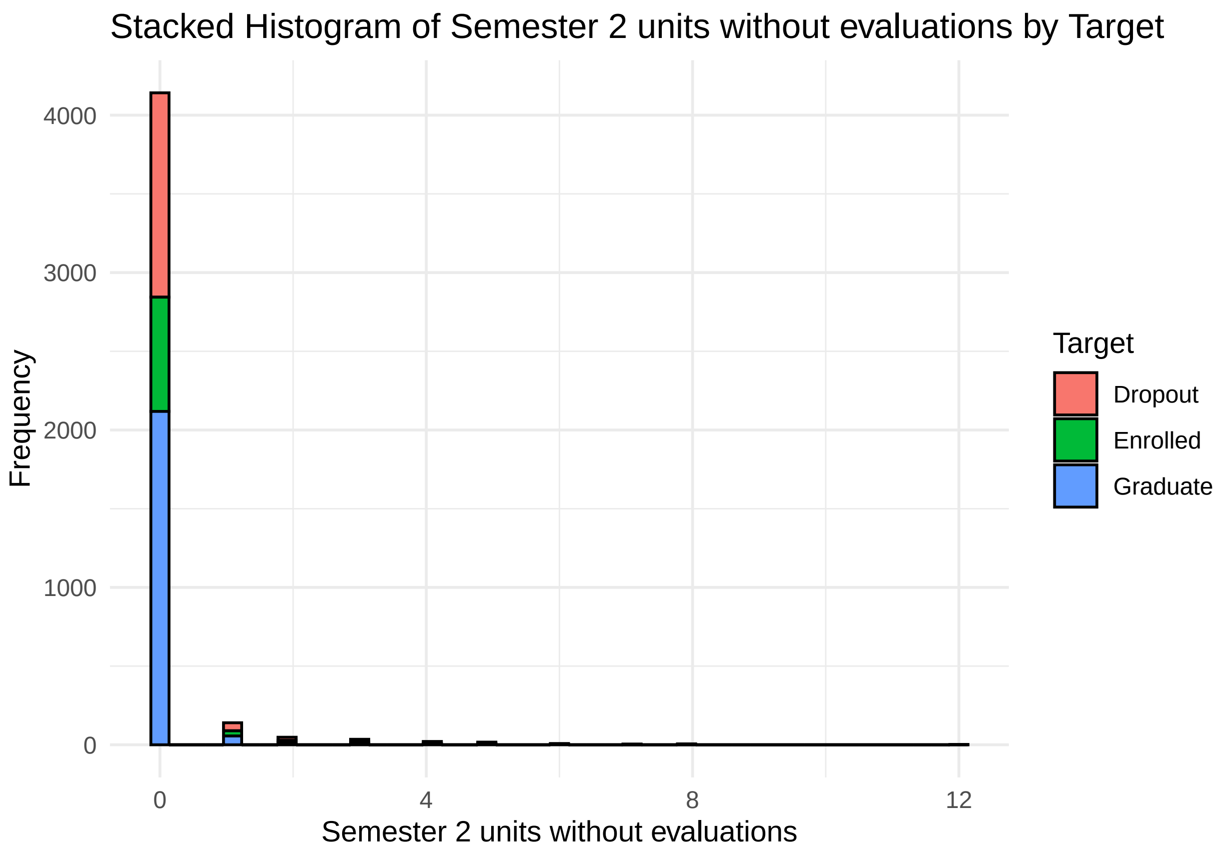

Stacked Histogram of Semester 2 credited units by Target
Frequency
4000
3000
2000
1000
0
0
5
10
15
20
Semester 2 credited units
Target
Dropout
Enrolled
Graduate

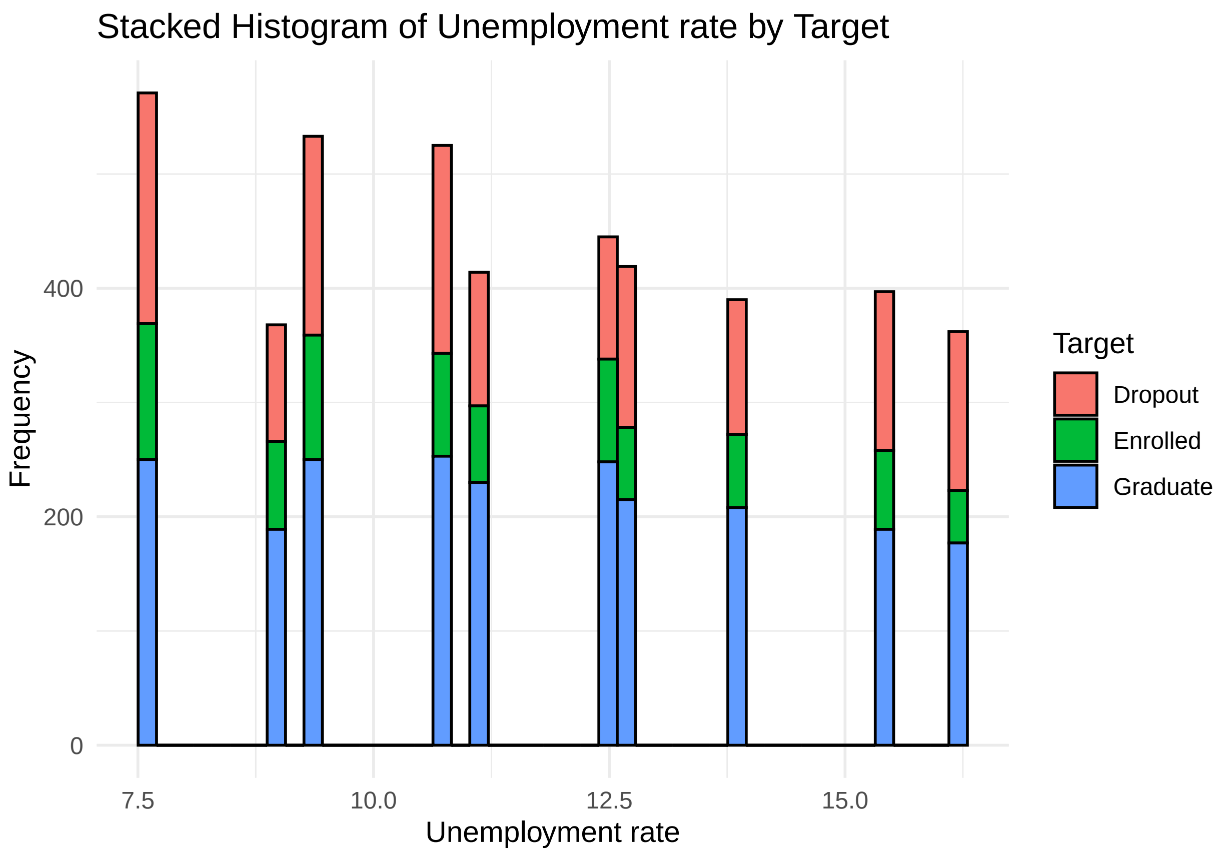

Stacked Histogram of Semester 2 enrolled units by Target
Frequency
2000
1500
1000
500
0
0
5
10
15
20
Semester 2 enrolled units
Target
Dropout
Enrolled
Graduate

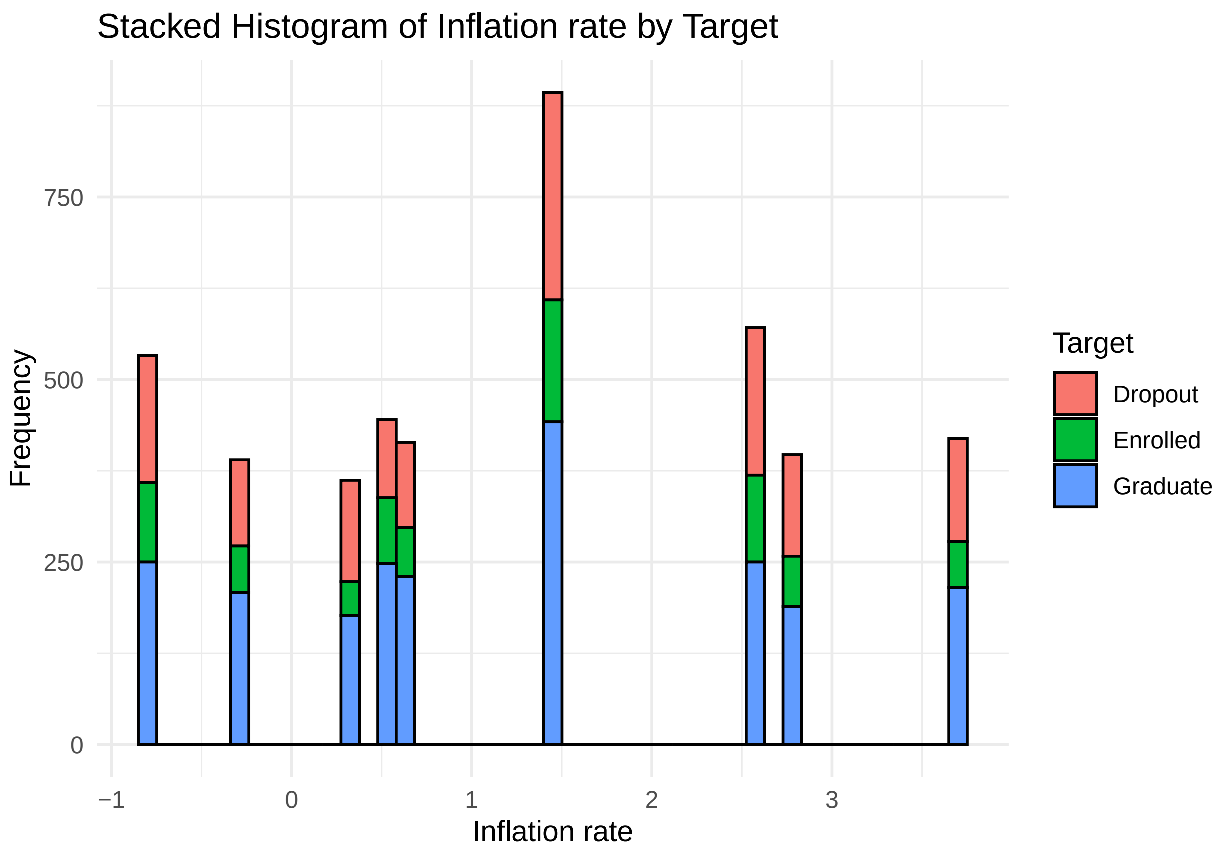

Stacked Histogram of Semester 2 evaluations by Target
Frequency
800
600
400
200
0
0
10
20
30
Semester 2 evaluations
Target
Dropout
Enrolled
Graduate

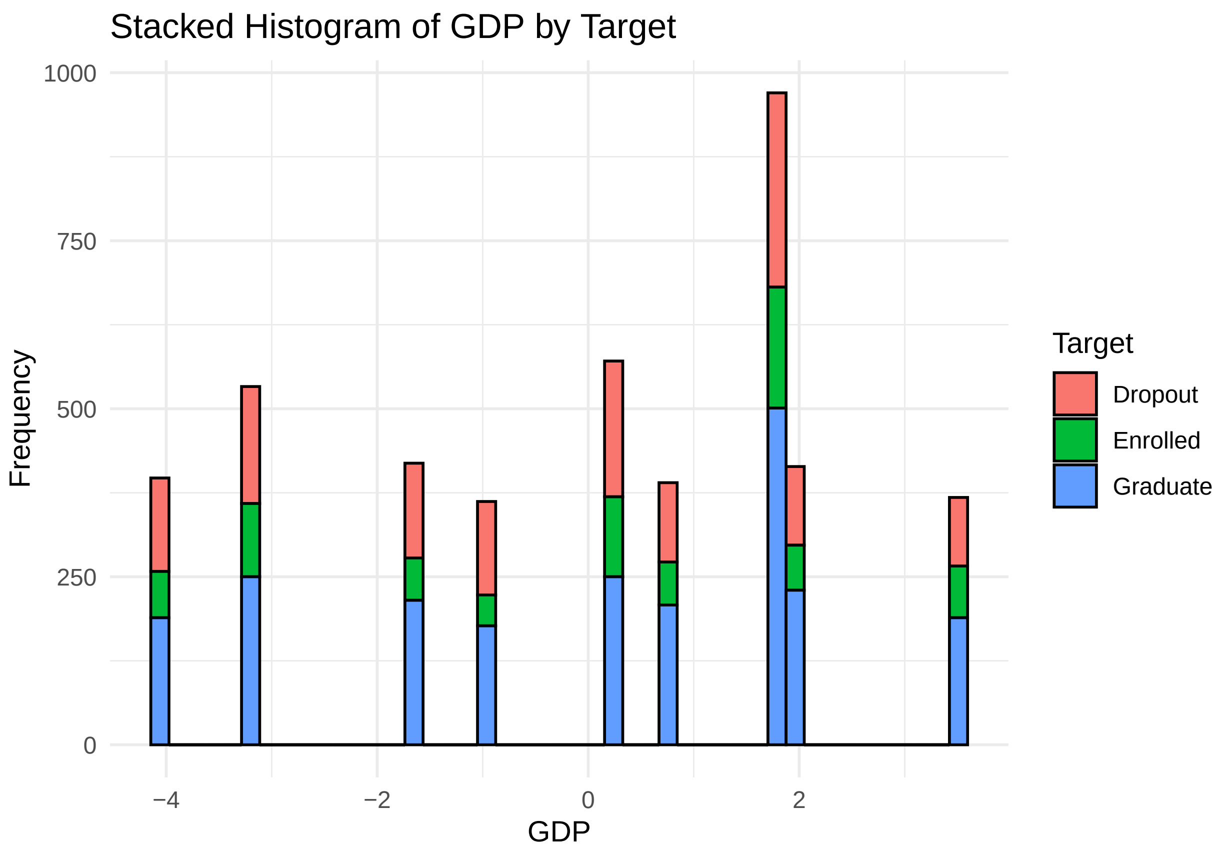

Stacked Histogram of Semester 2 approved units by Target
Frequency
1000
750
500
250
0
0
5
10
15
20
Semester 2 approved units
Target
Dropout
Enrolled
Graduate


Stacked Histogram of Semester 2 grade by Target
Frequency
750
500
250
0
0
5
10
15
Semester 2 grade
Target
Dropout
Enrolled
Graduate

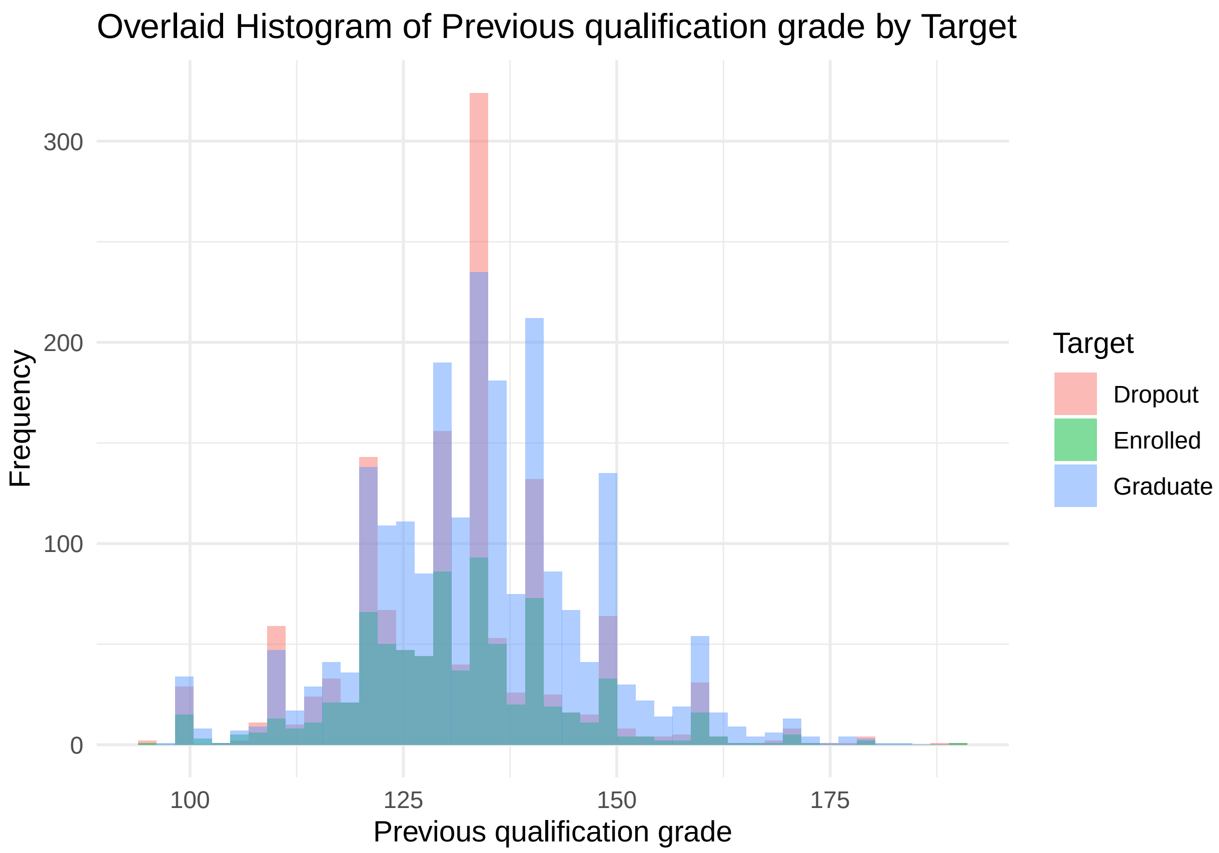

Stacked Histogram of Semester 2 units without evaluations by Target
Frequency
4000
3000
2000
1000
0
0
4
8
12
Semester 2 units without evaluations
Target
Dropout
Enrolled
Graduate

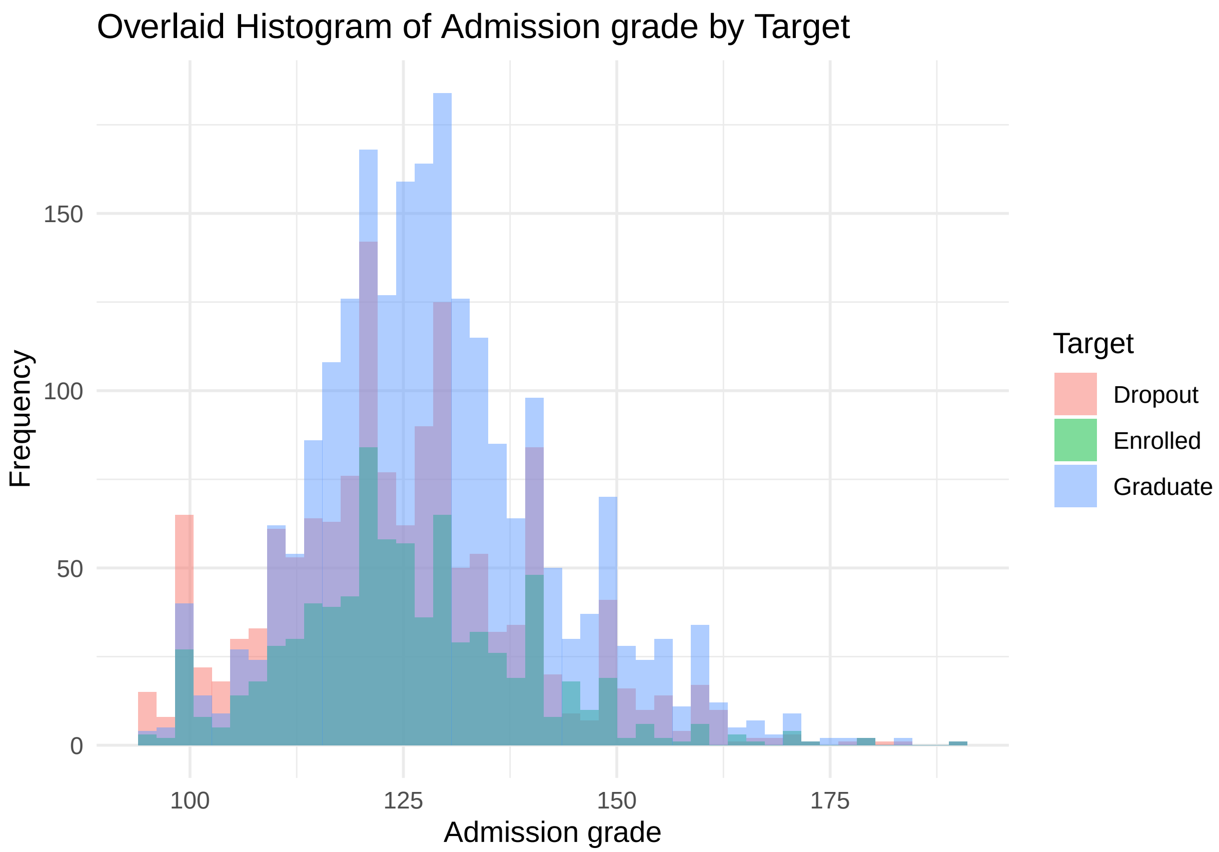

Stacked Histogram of Unemployment rate by Target
Frequency
0
200
400
7.5
10.0
12.5
15.0
Unemployment rate
Target
Dropout
Enrolled
Graduate

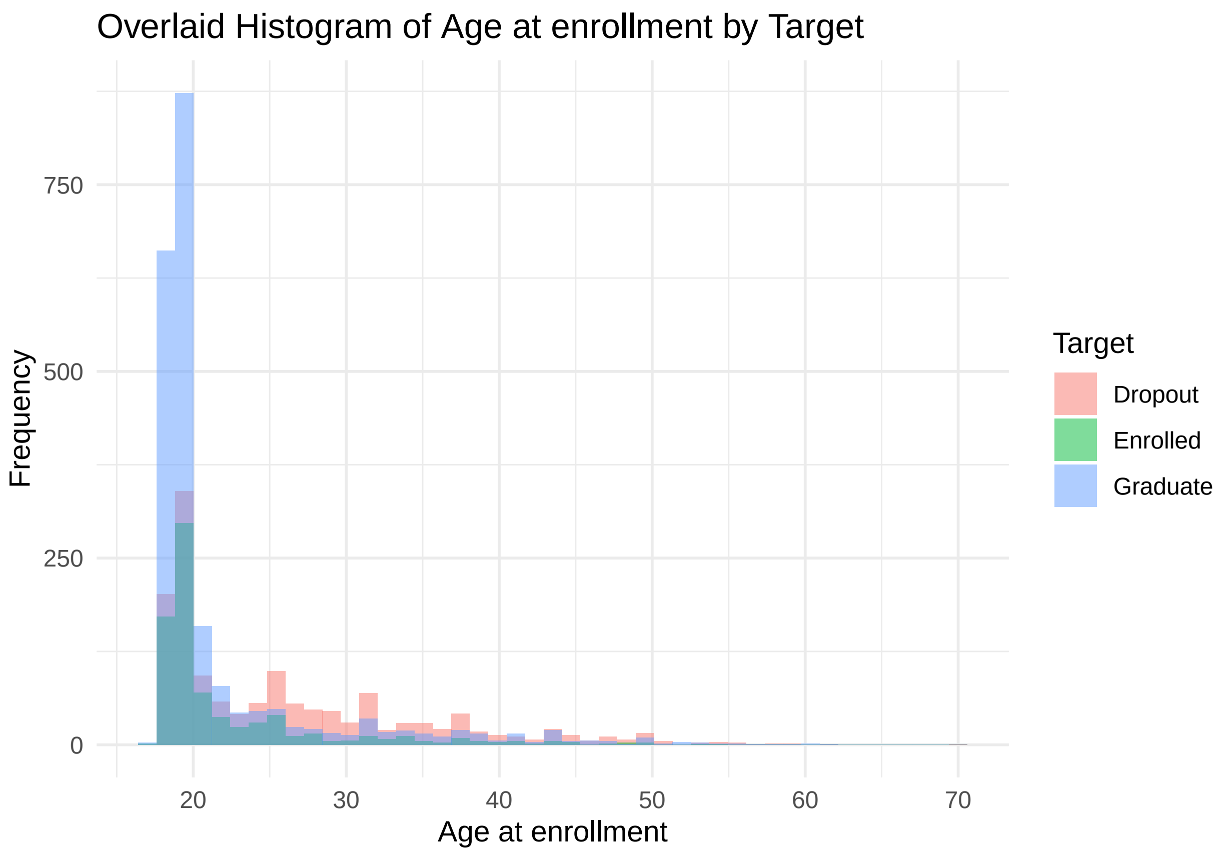

Stacked Histogram of Inflation rate by Target
Frequency
750
500
250
0
−1
0
1
2
3
Inflation rate
Target
Dropout
Enrolled
Graduate

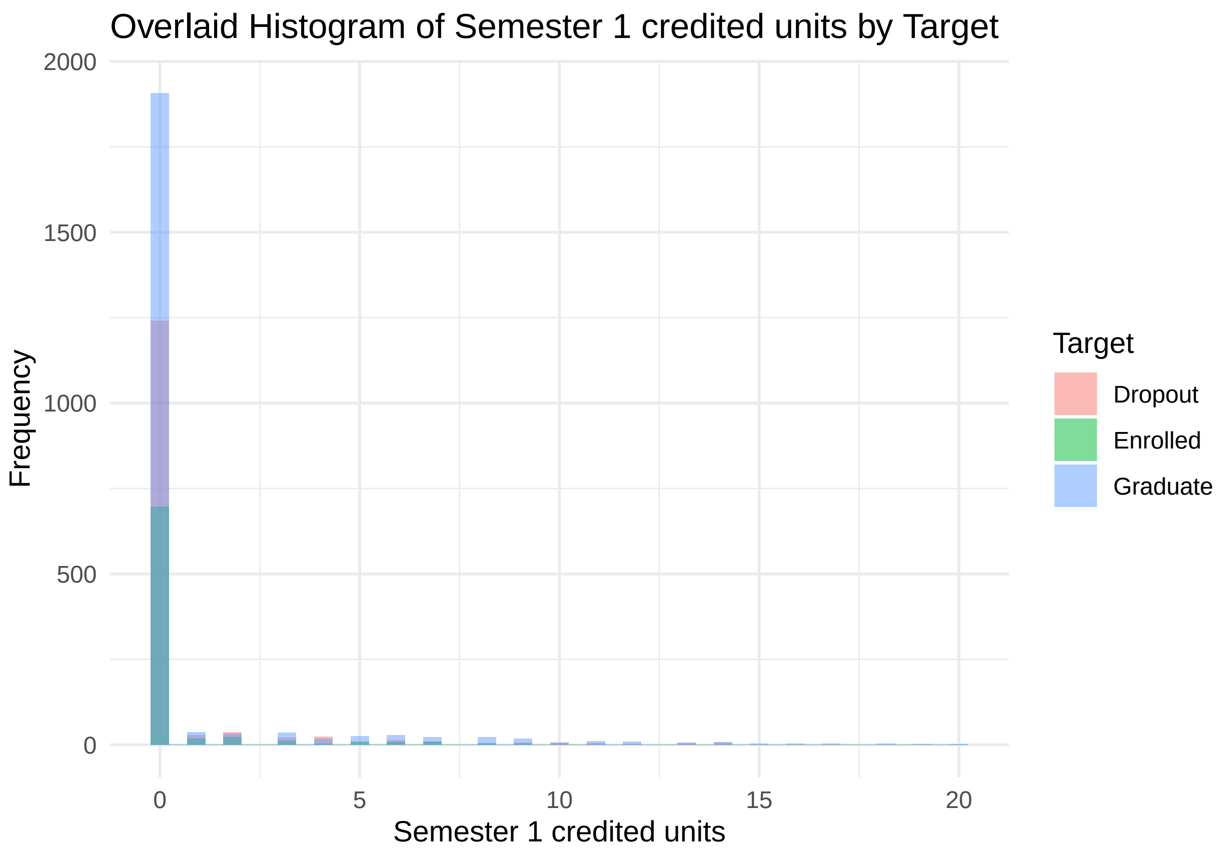


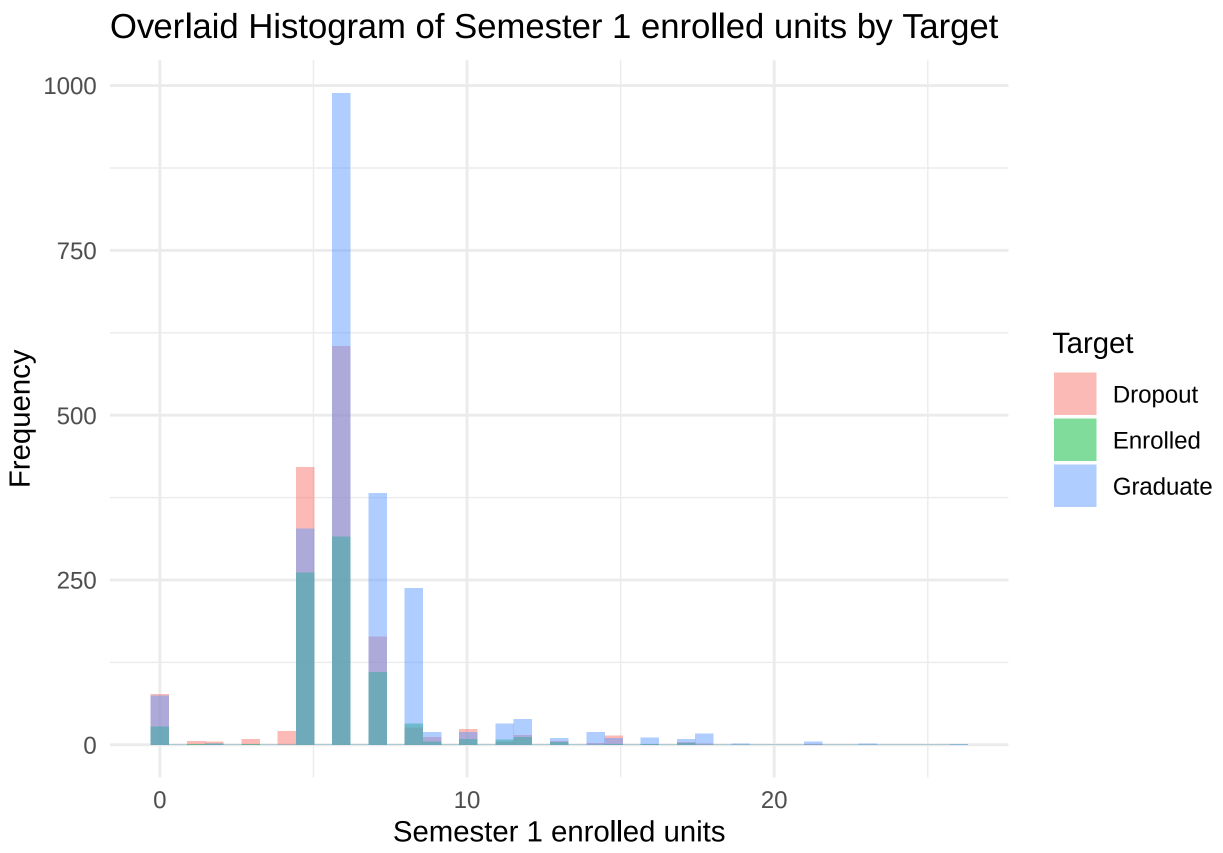
```
for (i in numeric_cols) {
  plot <- ggplot(data1, aes(x = .data[[i]], fill = Target)) +
    geom_histogram(bins = 45, position = "identity", alpha = 0.5) +
    labs(title = paste("Overlaid Histogram of", i, "by Target"), x = i, y = "Frequency") +
    theme_minimal()
  print(plot)
}
```

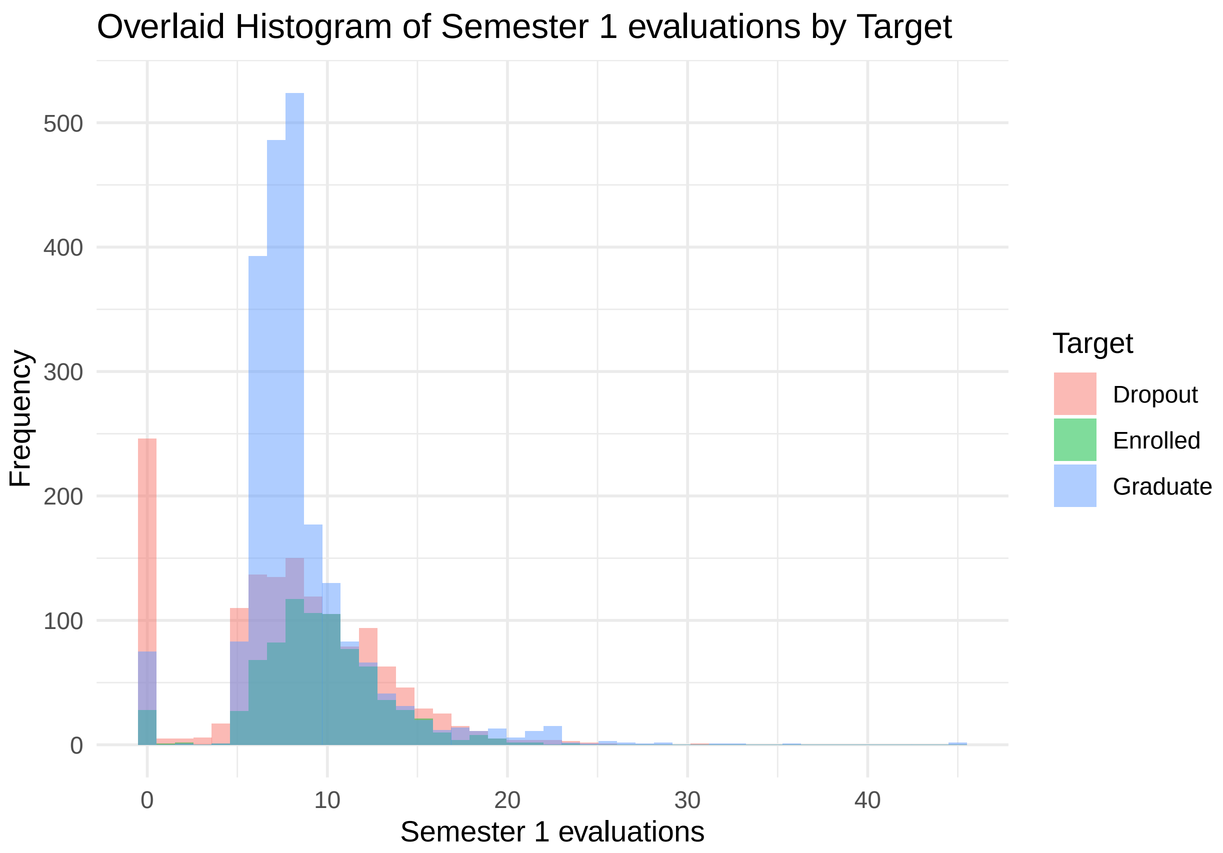

Overlaid Histogram of Previous qualification grade by Target
Frequency
300
200
100
0
100
125
150
175
Previous qualification grade
Target
Dropout
Enrolled
Graduate

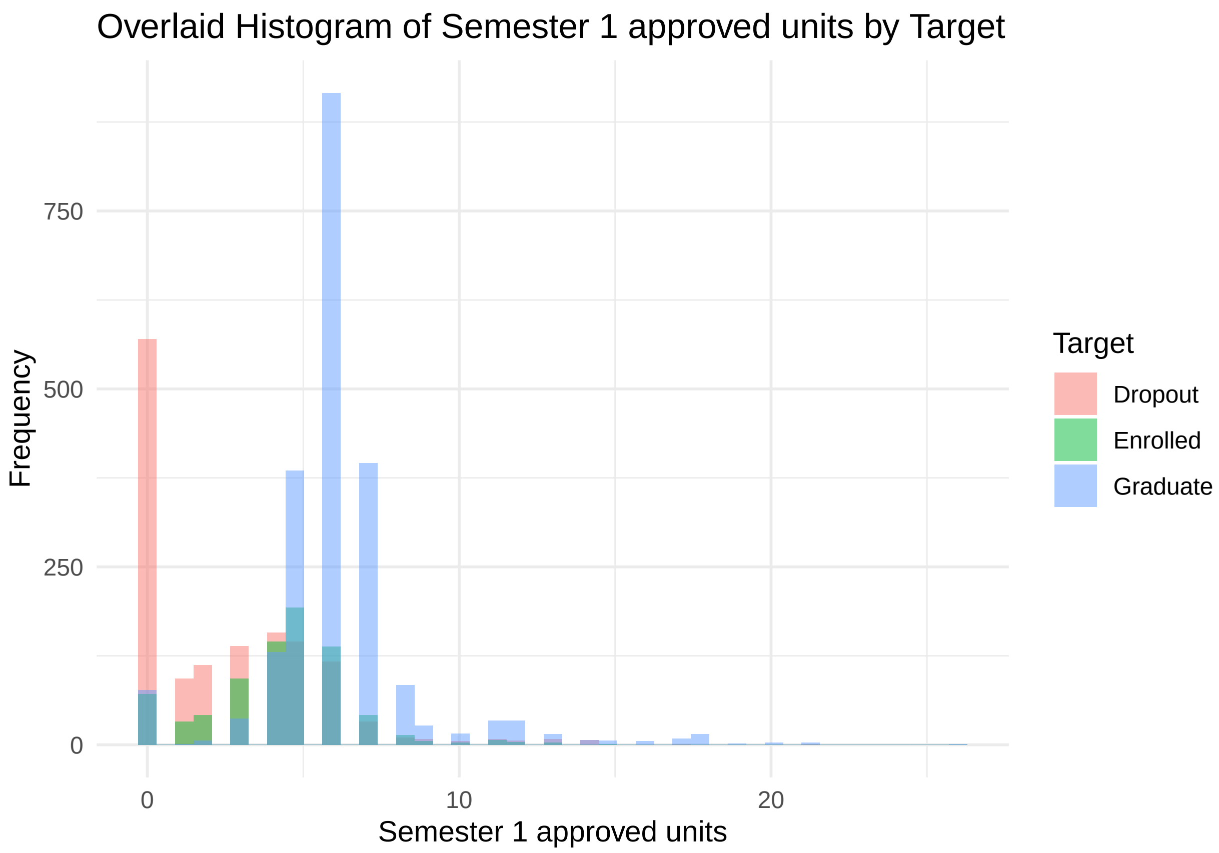

Overlaid Histogram of Admission grade by Target
Frequency
0
50
100
150
100
125
150
175
Admission grade
Target
Dropout
Enrolled
Graduate

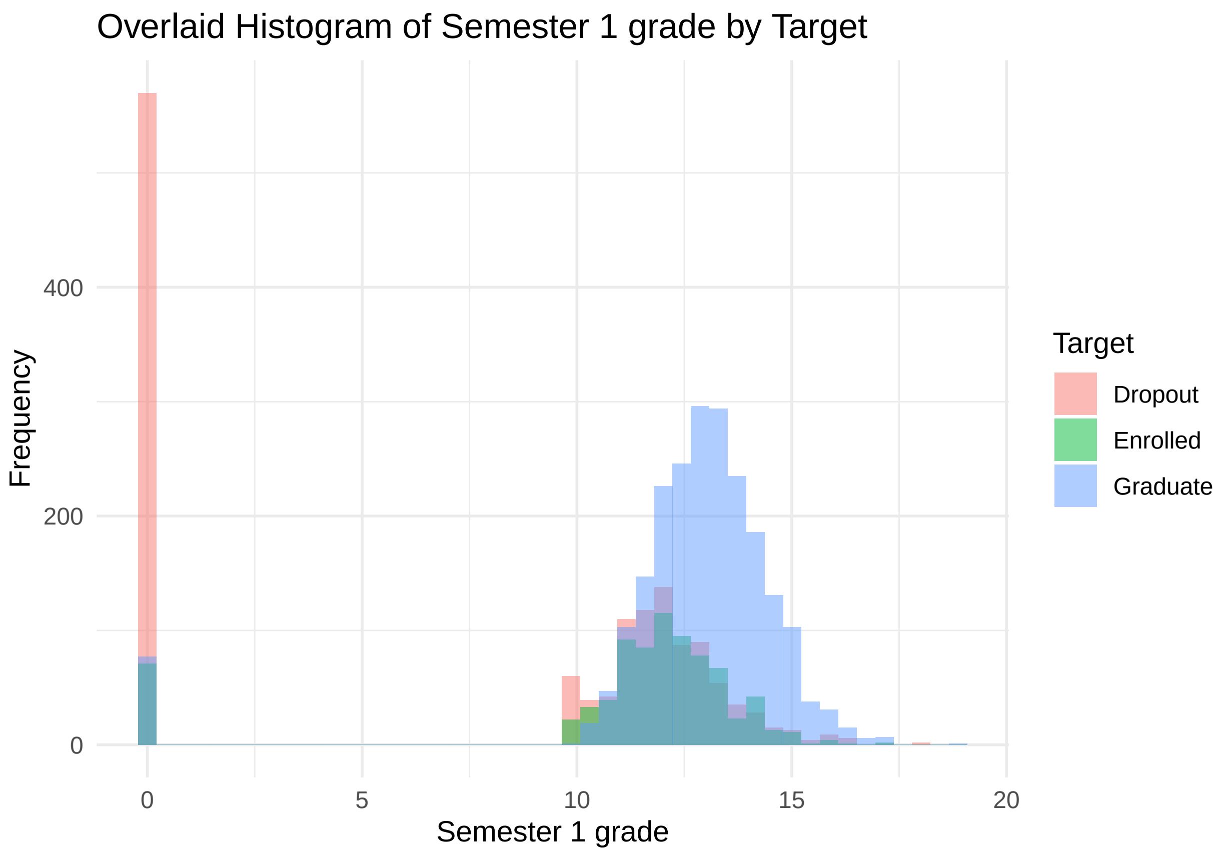

Overlaid Histogram of Age at enrollment by Target
Frequency
750
500
250
0
20
30
40
50
60
70
Age at enrollment
Target
Dropout
Enrolled
Graduate

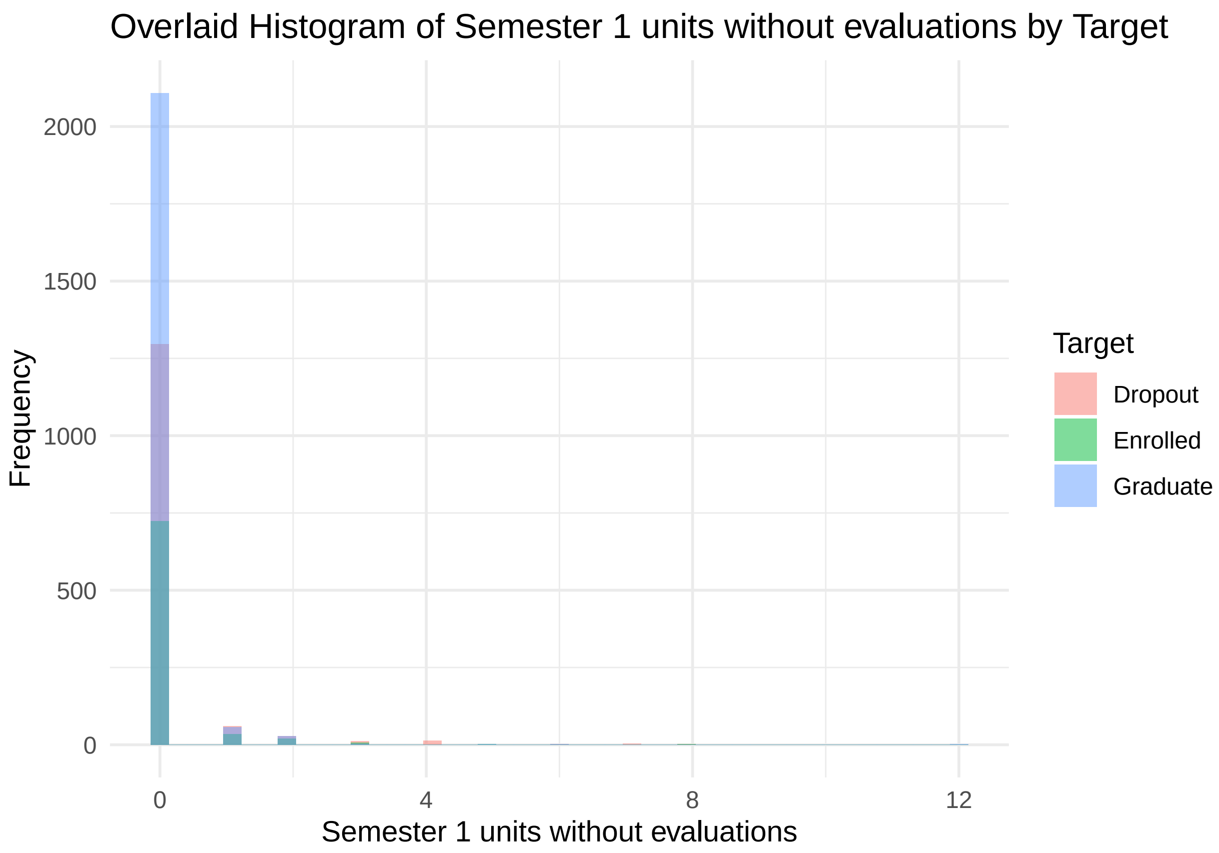

Overlaid Histogram of Semester 1 credited units by Target
Frequency
2000
1500
1000
500
0
0
5
10
15
20
Semester 1 credited units
Target
Dropout
Enrolled
Graduate

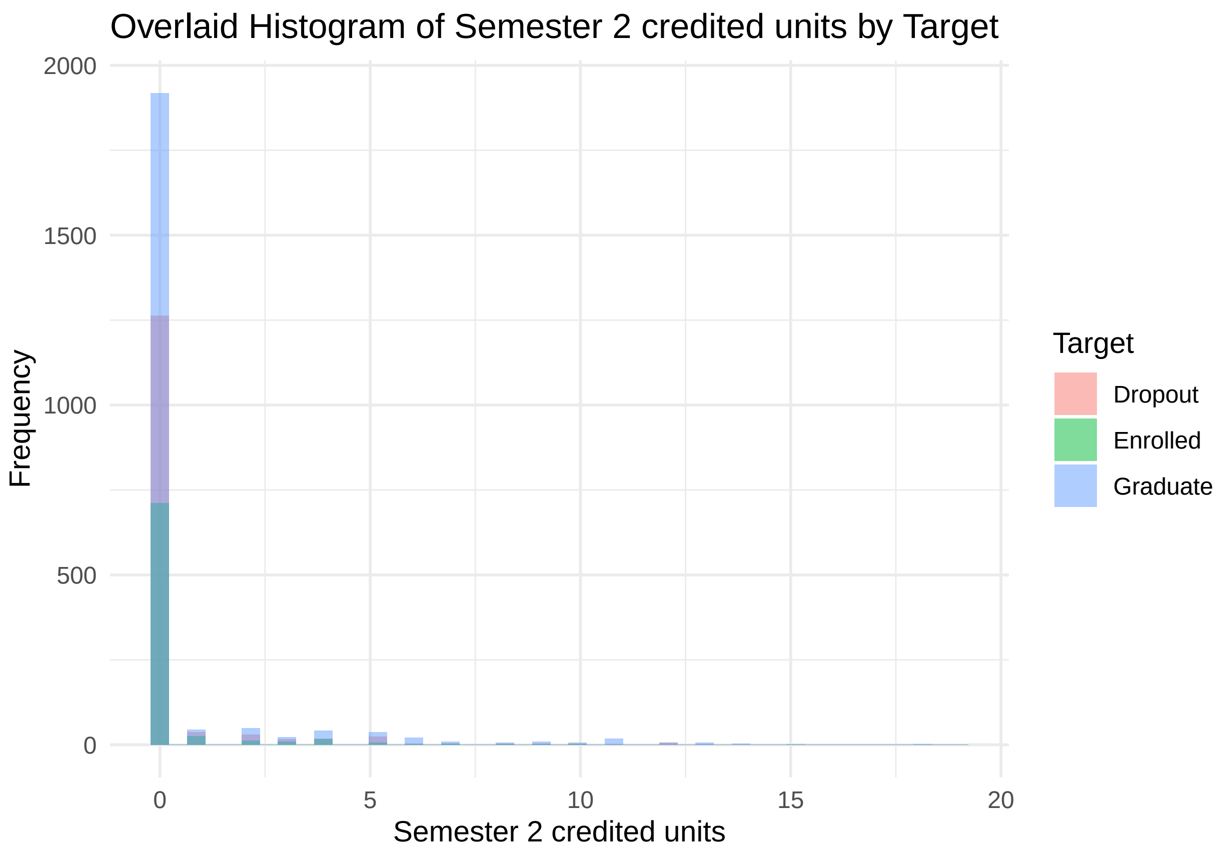

Overlaid Histogram of Semester 1 enrolled units by Target
Frequency
1000
750
500
250
0
0
10
20
Semester 1 enrolled units
Target
Dropout
Enrolled
Graduate

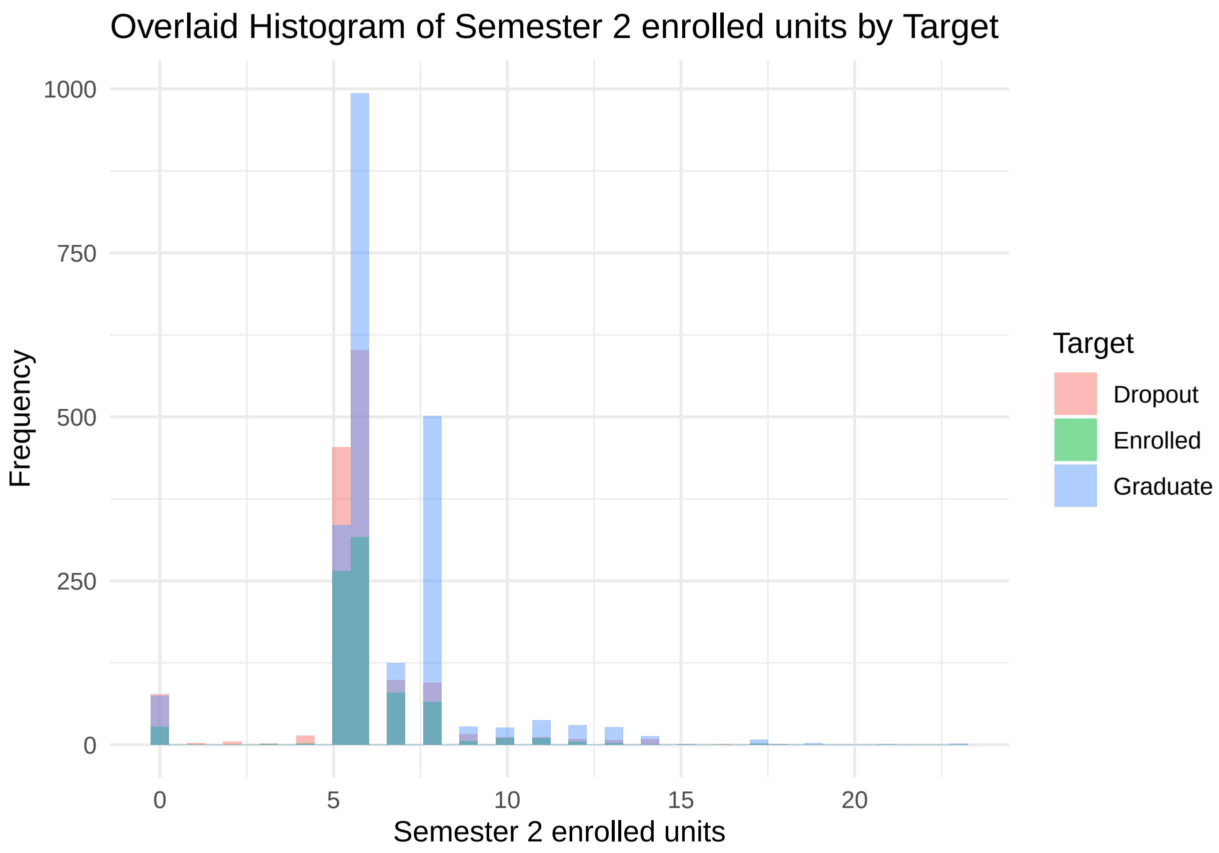

Overlaid Histogram of Semester 1 evaluations by Target
Frequency
500
400
300
200
100
0
0
10
20
30
40
Semester 1 evaluations
Target
Dropout
Enrolled
Graduate

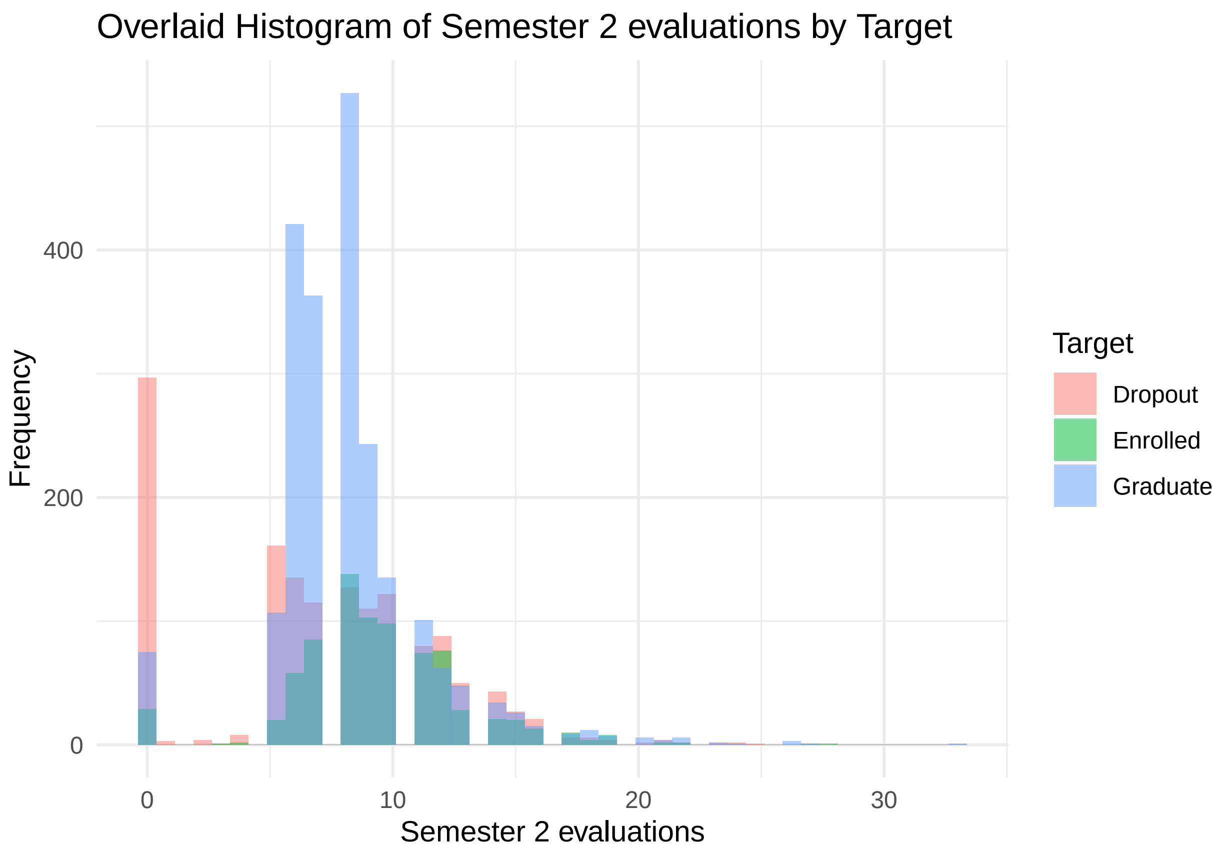

Overlaid Histogram of Semester 1 approved units by Target
Frequency
750
500
250
0
0
10
20
Semester 1 approved units
Target
Dropout
Enrolled
Graduate

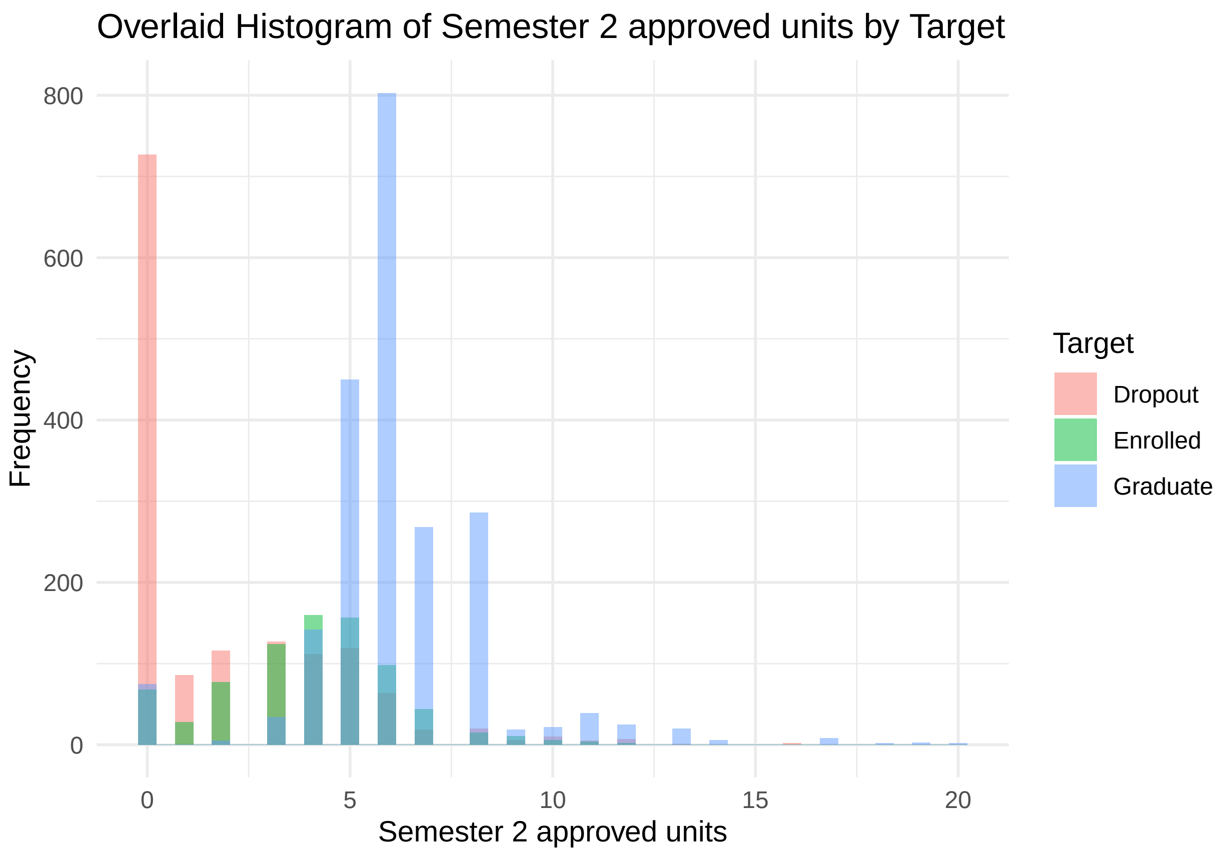

Overlaid Histogram of Semester 1 grade by Target
Frequency
400
200
0
0
5
10
15
20
Semester 1 grade
Target
Dropout
Enrolled
Graduate

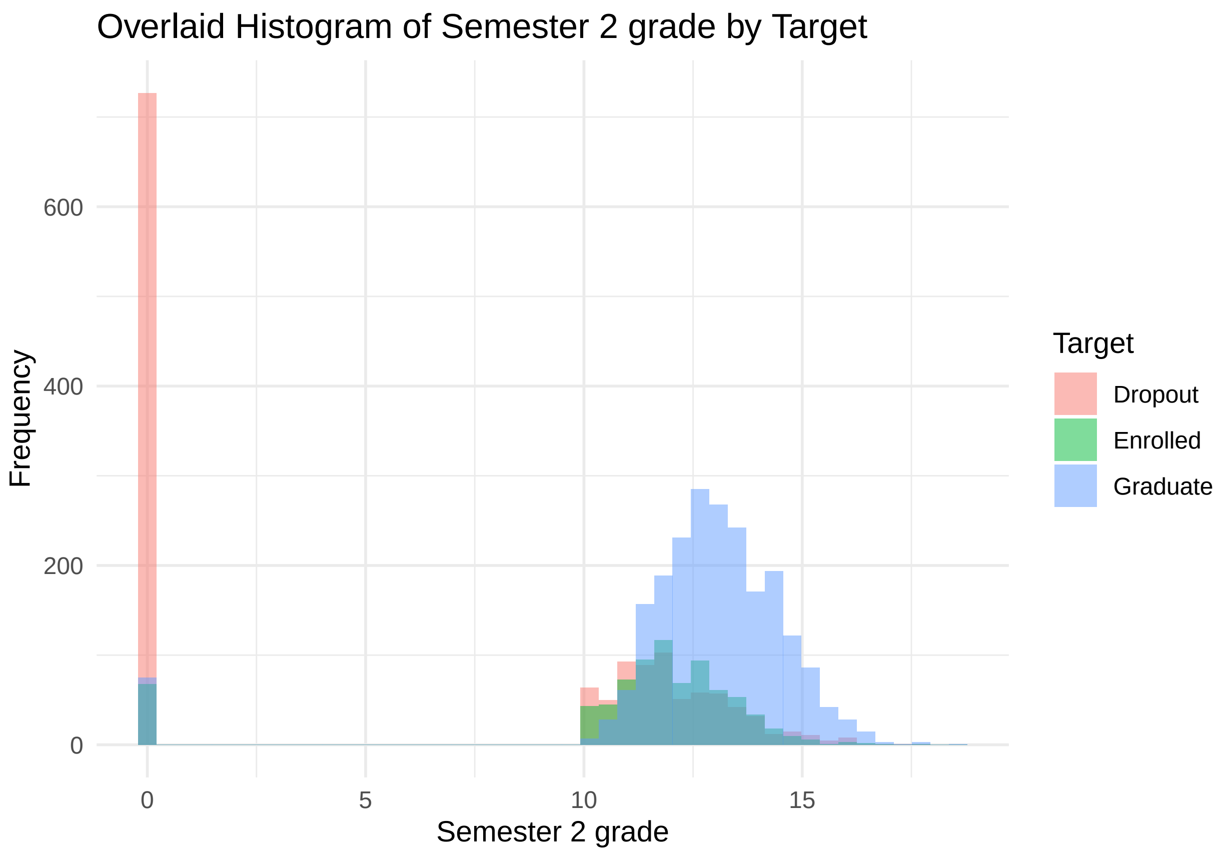

Overlaid Histogram of Semester 1 units without evaluations by Target
Frequency
2000
1500
1000
500
0
0
4
8
12
Semester 1 units without evaluations
Target
Dropout
Enrolled
Graduate

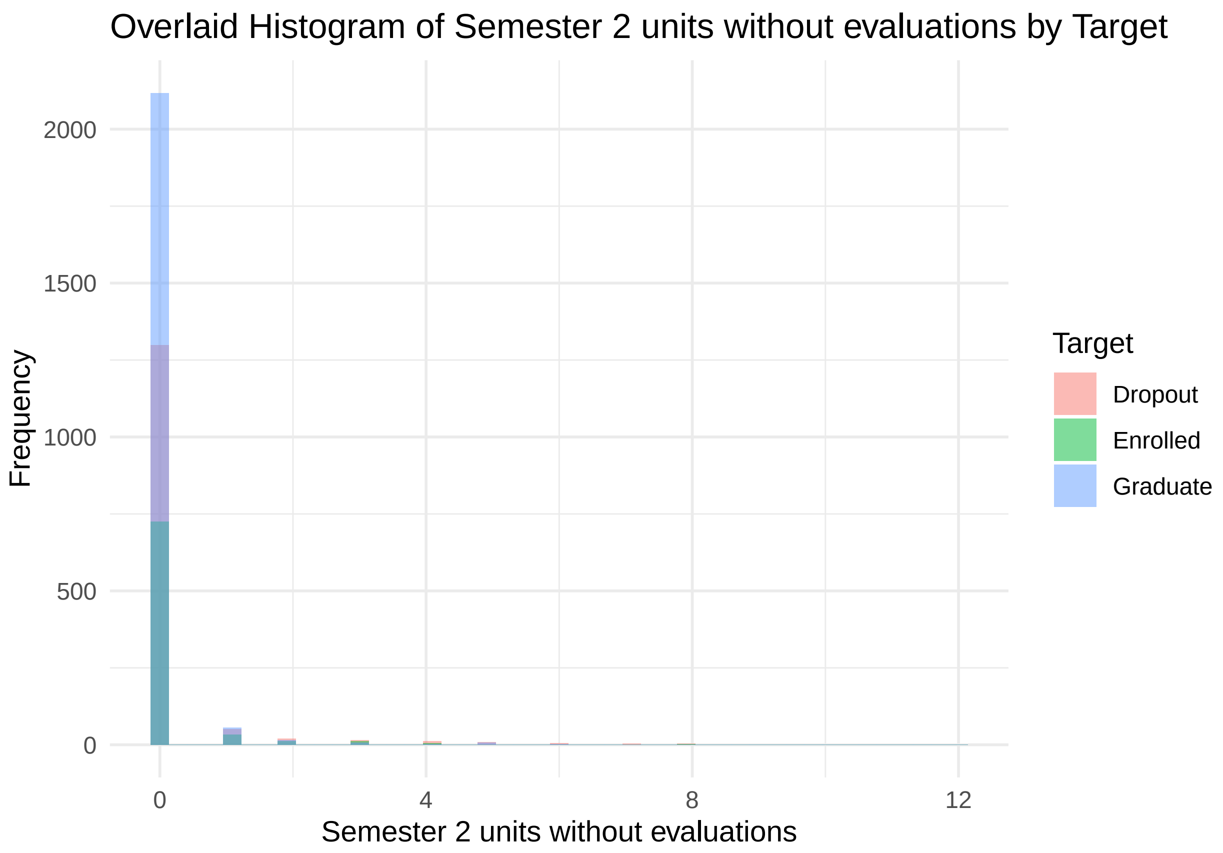

Overlaid Histogram of Semester 2 credited units by Target
Frequency
2000
1500
1000
500
0
0
5
10
15
20
Semester 2 credited units
Target
Dropout
Enrolled
Graduate

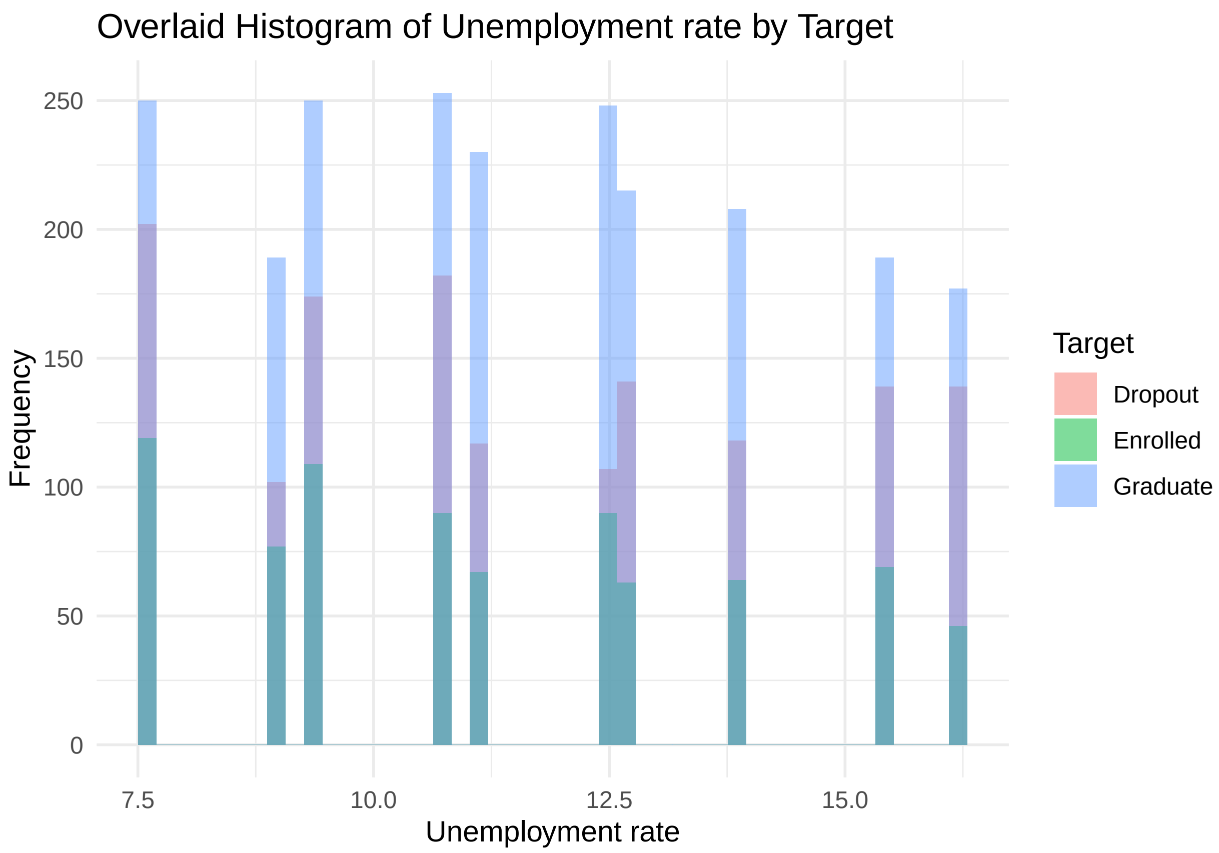

Overlaid Histogram of Semester 2 enrolled units by Target
Frequency
1000
750
500
250
0
0
5
10
15
20
Semester 2 enrolled units
Target
Dropout
Enrolled
Graduate

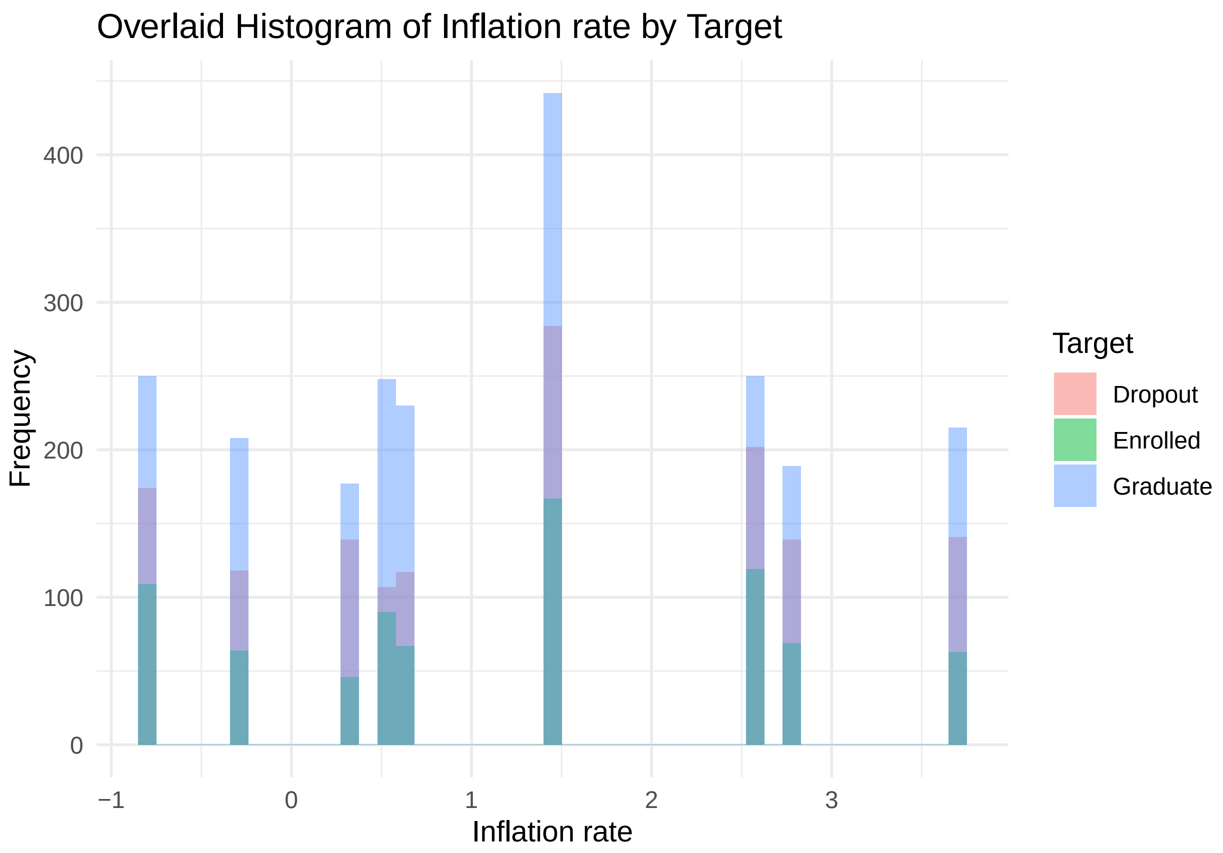

Overlaid Histogram of Semester 2 evaluations by Target
Frequency
400
200
0
0
10
20
30
Semester 2 evaluations
Target
Dropout
Enrolled
Graduate

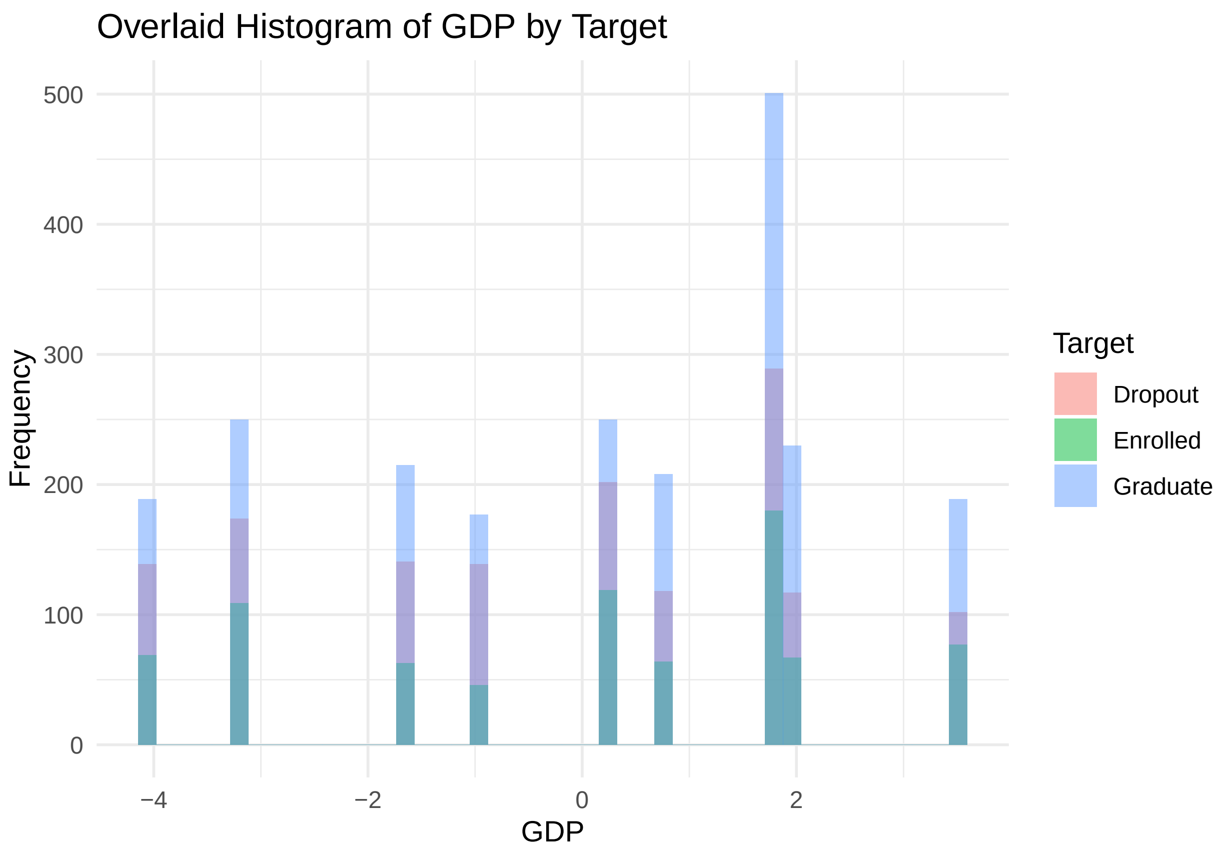

Overlaid Histogram of Semester 2 approved units by Target
Frequency
0
200
400
600
800
0
5
10
15
20
Semester 2 approved units
Target
Dropout
Enrolled
Graduate


Overlaid Histogram of Semester 2 grade by Target
Frequency
600
400
200
0
0
5
10
15
Semester 2 grade
Target
Dropout
Enrolled
Graduate

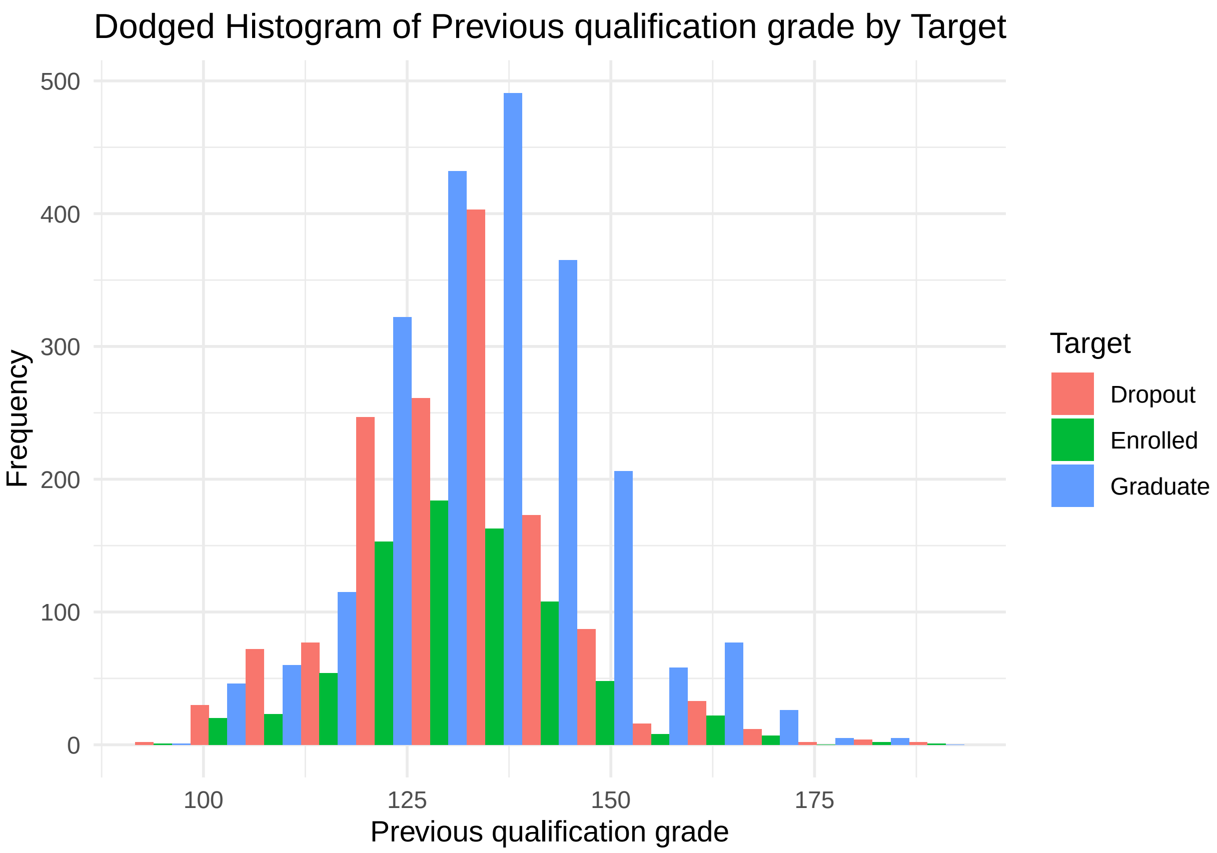

Overlaid Histogram of Semester 2 units without evaluations by Target
Frequency
2000
1500
1000
500
0
0
4
8
12
Semester 2 units without evaluations
Target
Dropout
Enrolled
Graduate

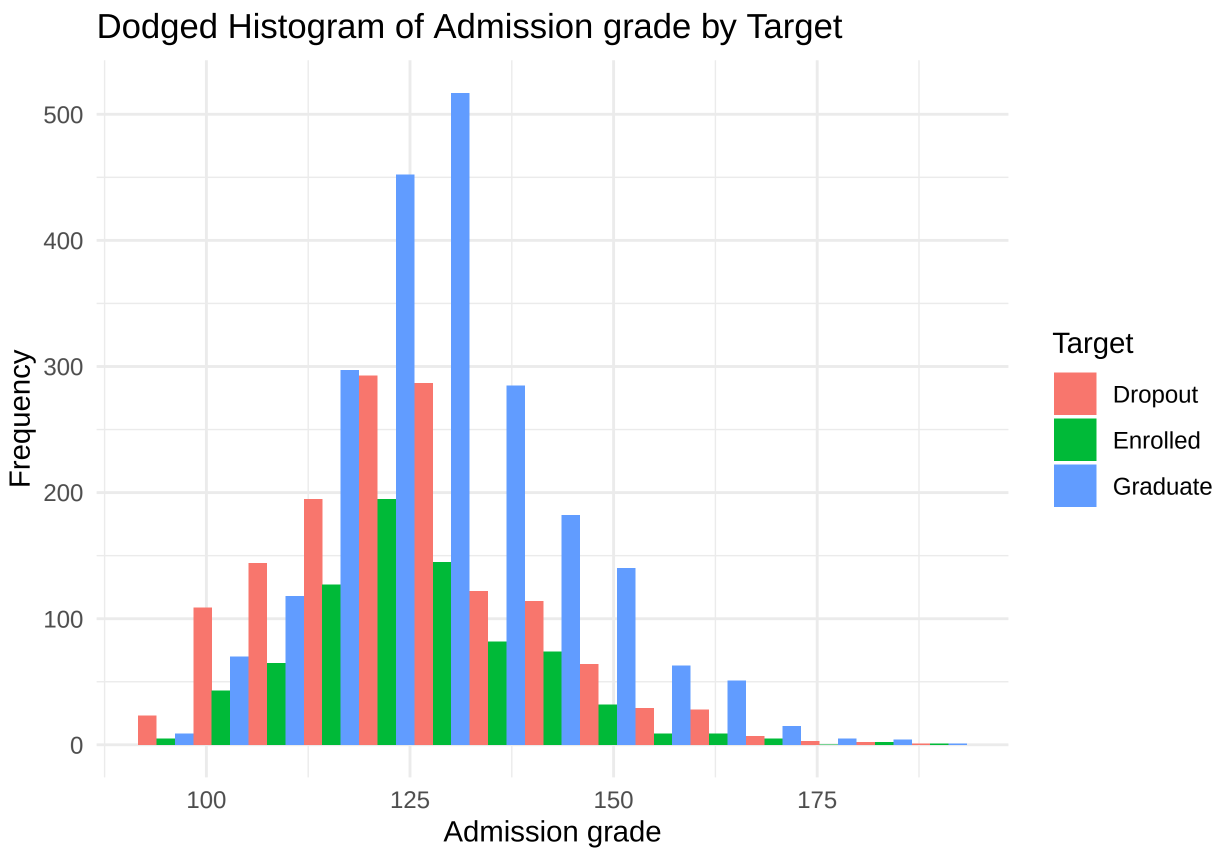

Overlaid Histogram of Unemployment rate by Target
Frequency
0
50
100
150
200
250
7.5
10.0
12.5
15.0
Unemployment rate
Target
Dropout
Enrolled
Graduate

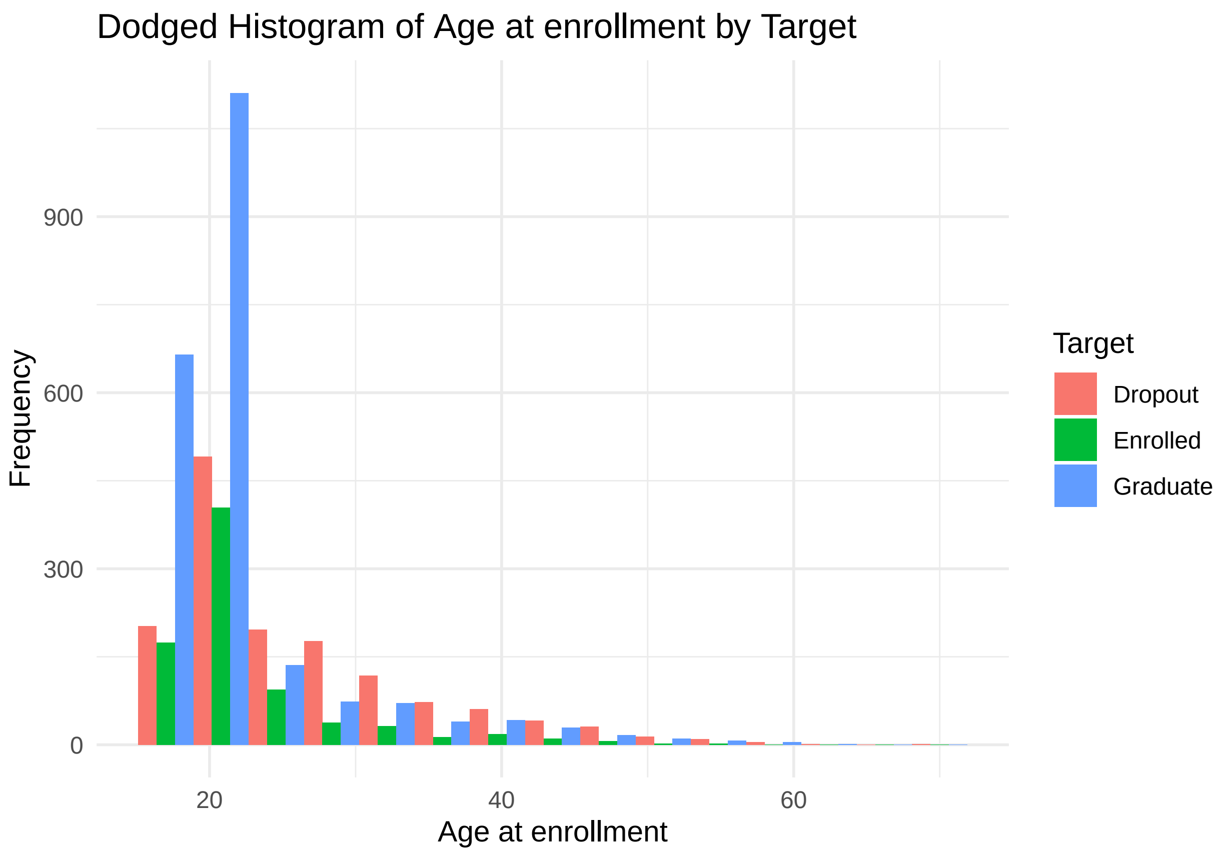

Overlaid Histogram of Inflation rate by Target
Frequency
400
300
200
100
0
−1
0
1
2
3
Inflation rate
Target
Dropout
Enrolled
Graduate

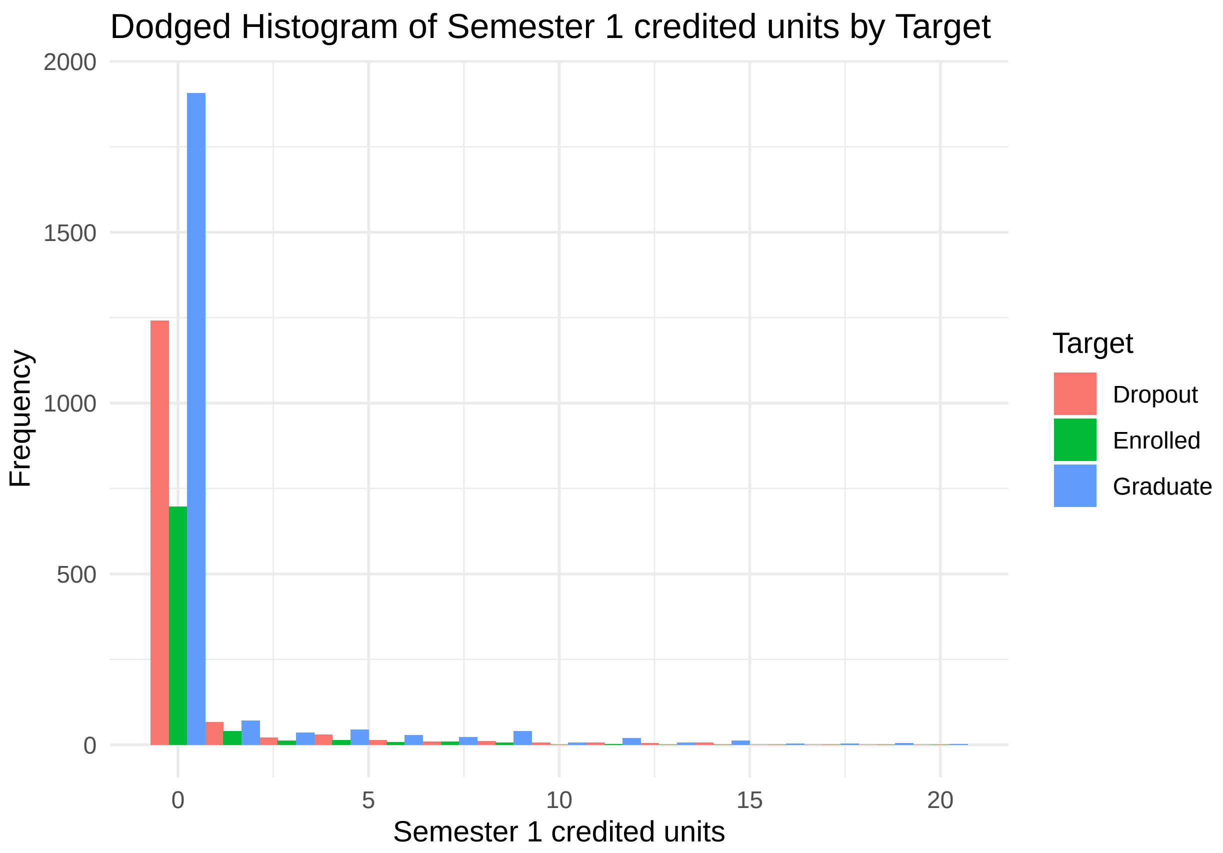


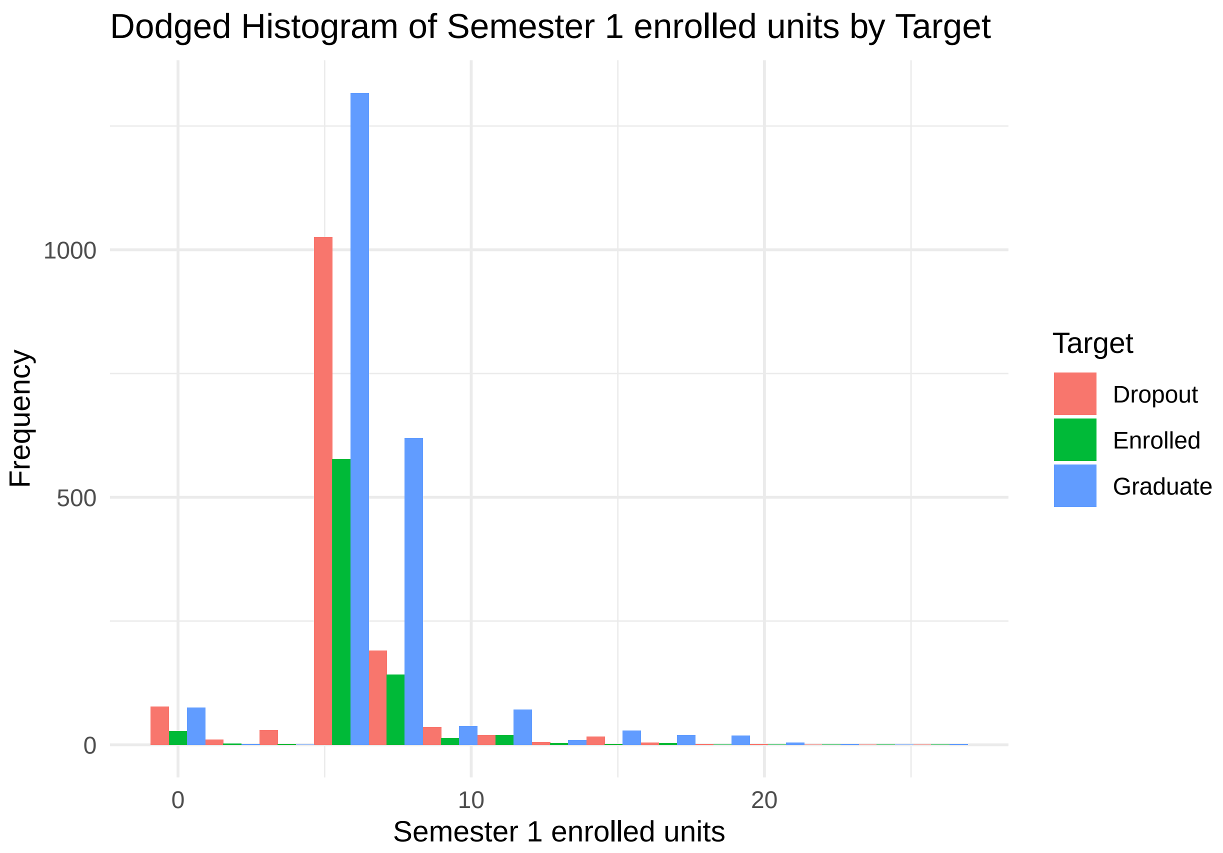
```r
for (i in numeric_cols) {
  plot <- ggplot(data1, aes(x = .data[[i]], fill = Target)) +
    geom_histogram(bins = 15, position = "dodge") +
    labs(title = paste("Dodged Histogram of", i, "by Target"), x = i, y = "Frequency") +
    theme_minimal()
  print(plot)
}
```

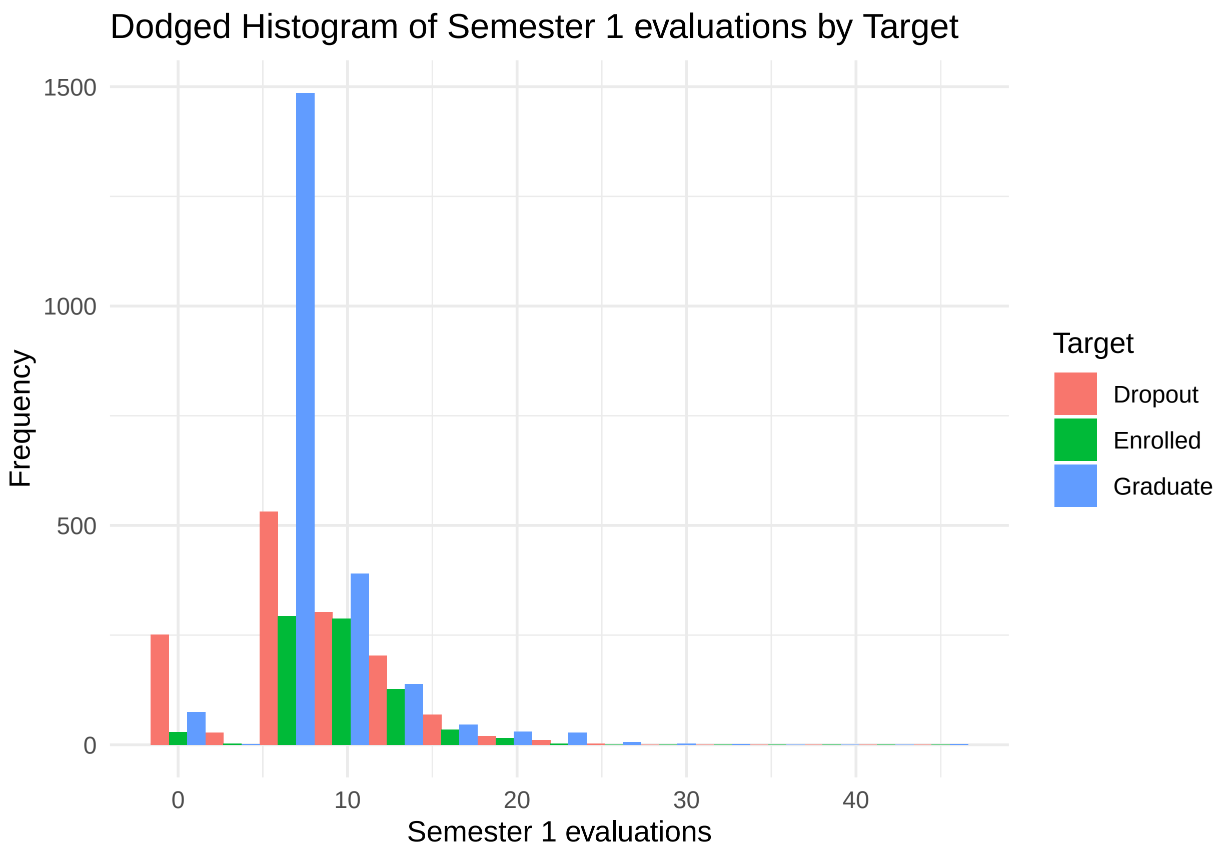

Dodged Histogram of Previous qualification grade by Target
Frequency
500
400
300
200
100
0
100
125
150
175
Previous qualification grade
Target
Dropout
Enrolled
Graduate

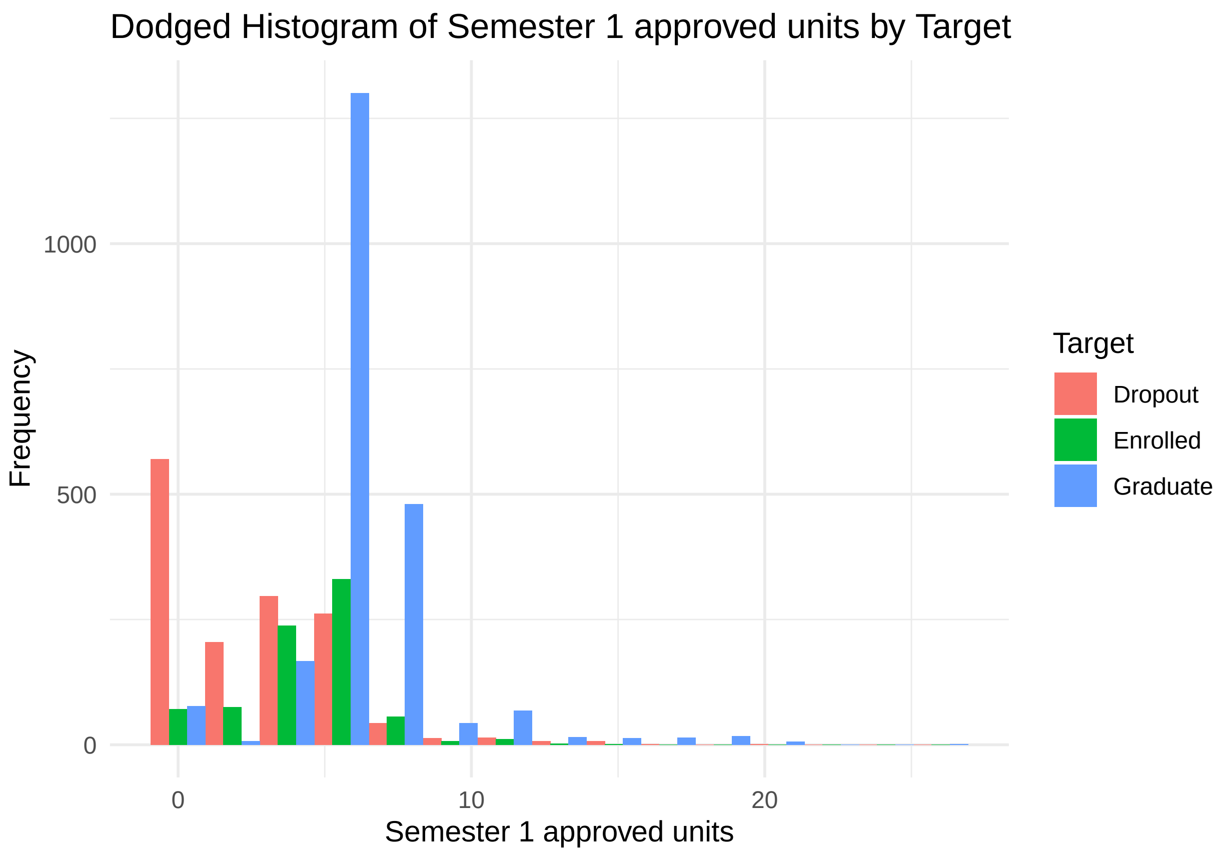

Dodged Histogram of Admission grade by Target
Frequency
500
400
300
200
100
0
100
125
150
175
Admission grade
Target
Dropout
Enrolled
Graduate

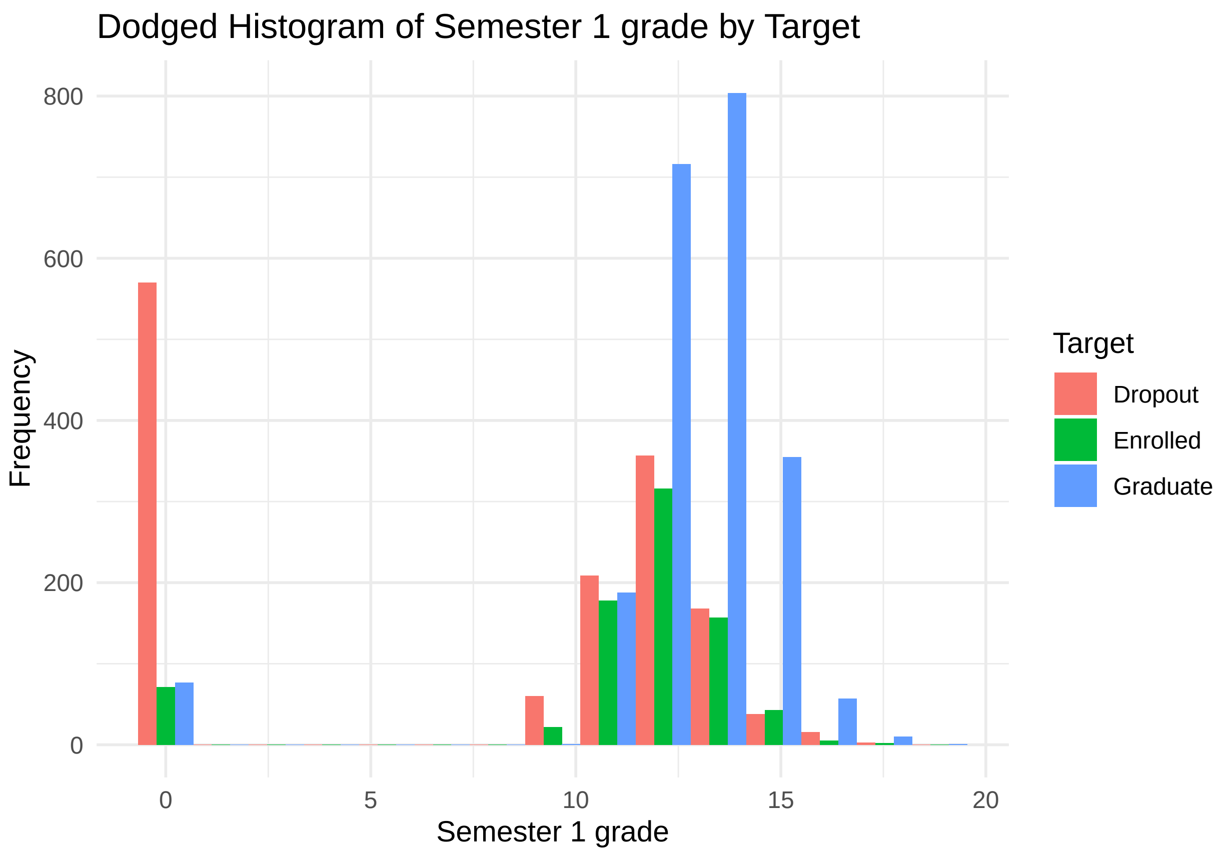

Dodged Histogram of Age at enrollment by Target
Frequency
900
600
300
0
20
40
60
Age at enrollment
Target
Dropout
Enrolled
Graduate

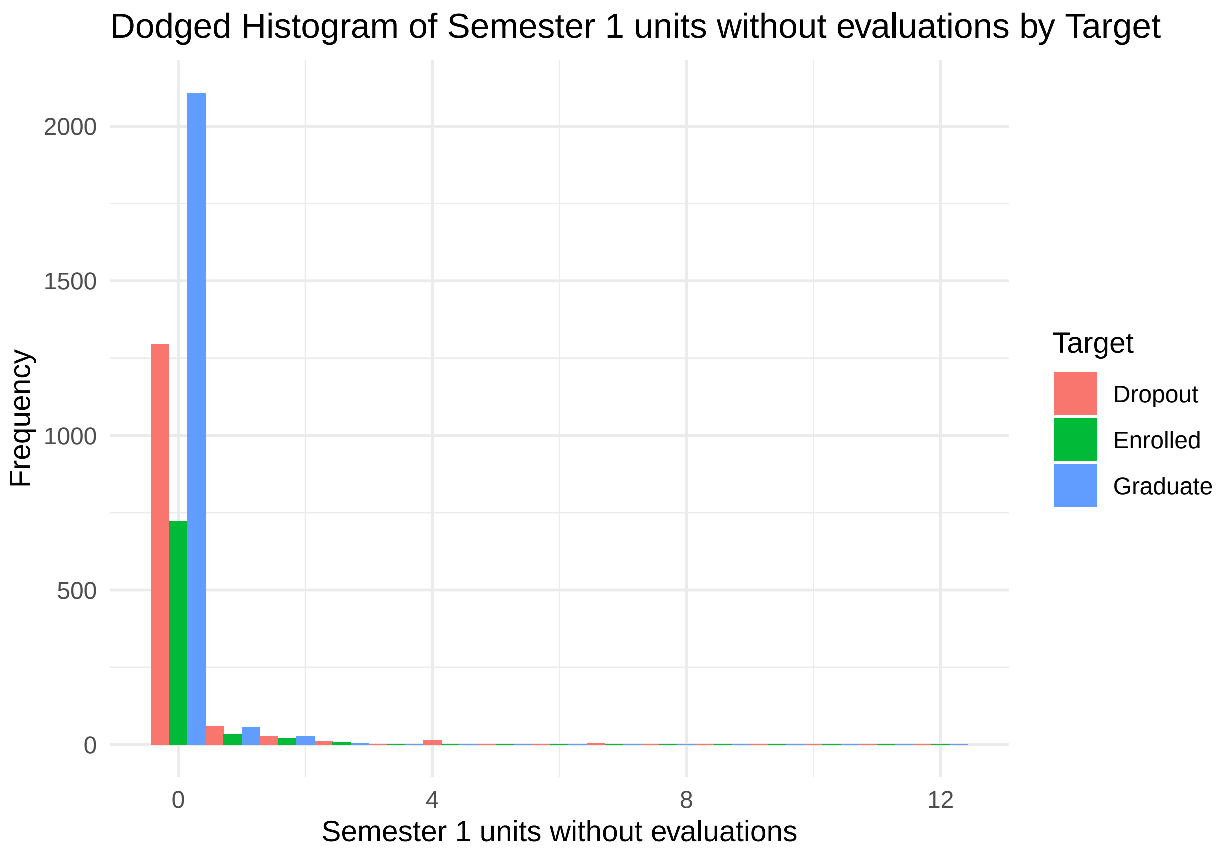

Dodged Histogram of Semester 1 credited units by Target
Frequency
2000
1500
1000
500
0
0
5
10
15
20
Semester 1 credited units
Target
Dropout
Enrolled
Graduate

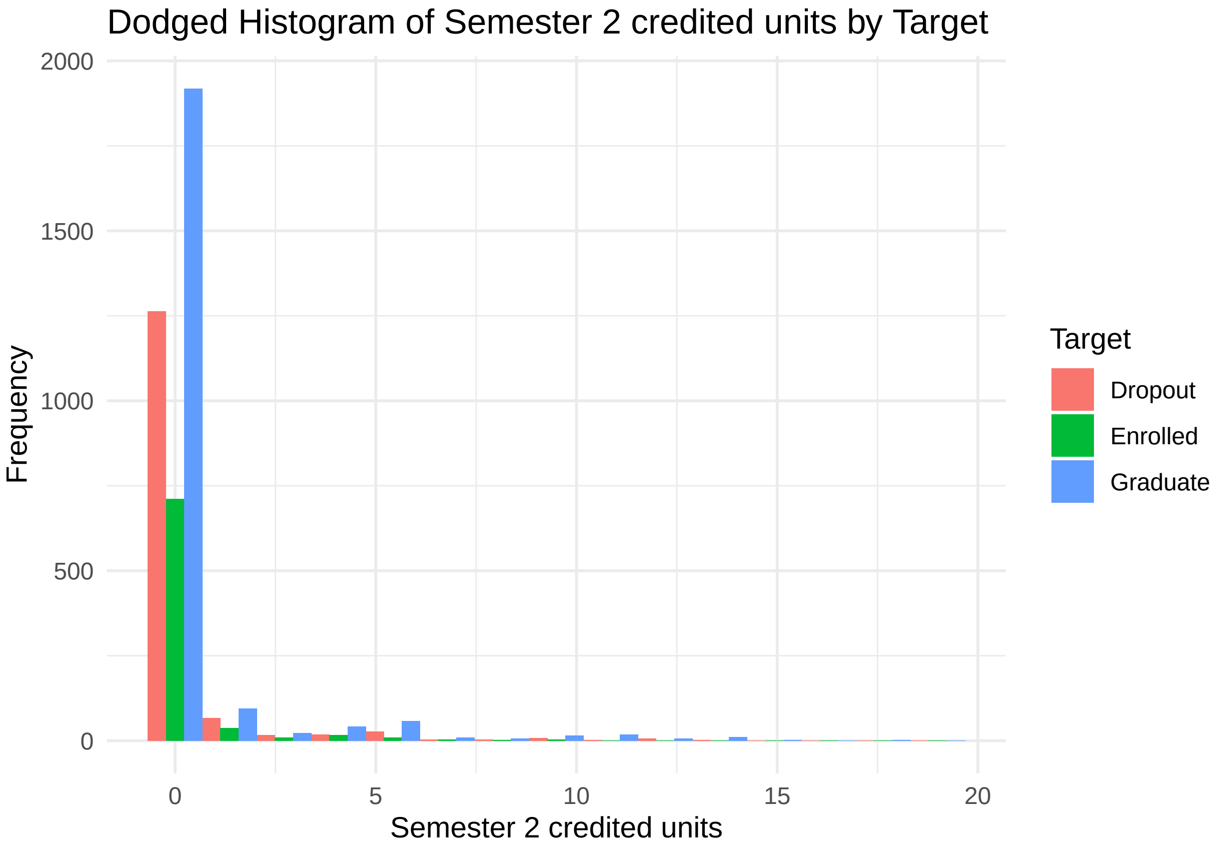

Dodged Histogram of Semester 1 enrolled units by Target
Frequency
1000
500
0
0
10
20
Semester 1 enrolled units
Target
Dropout
Enrolled
Graduate

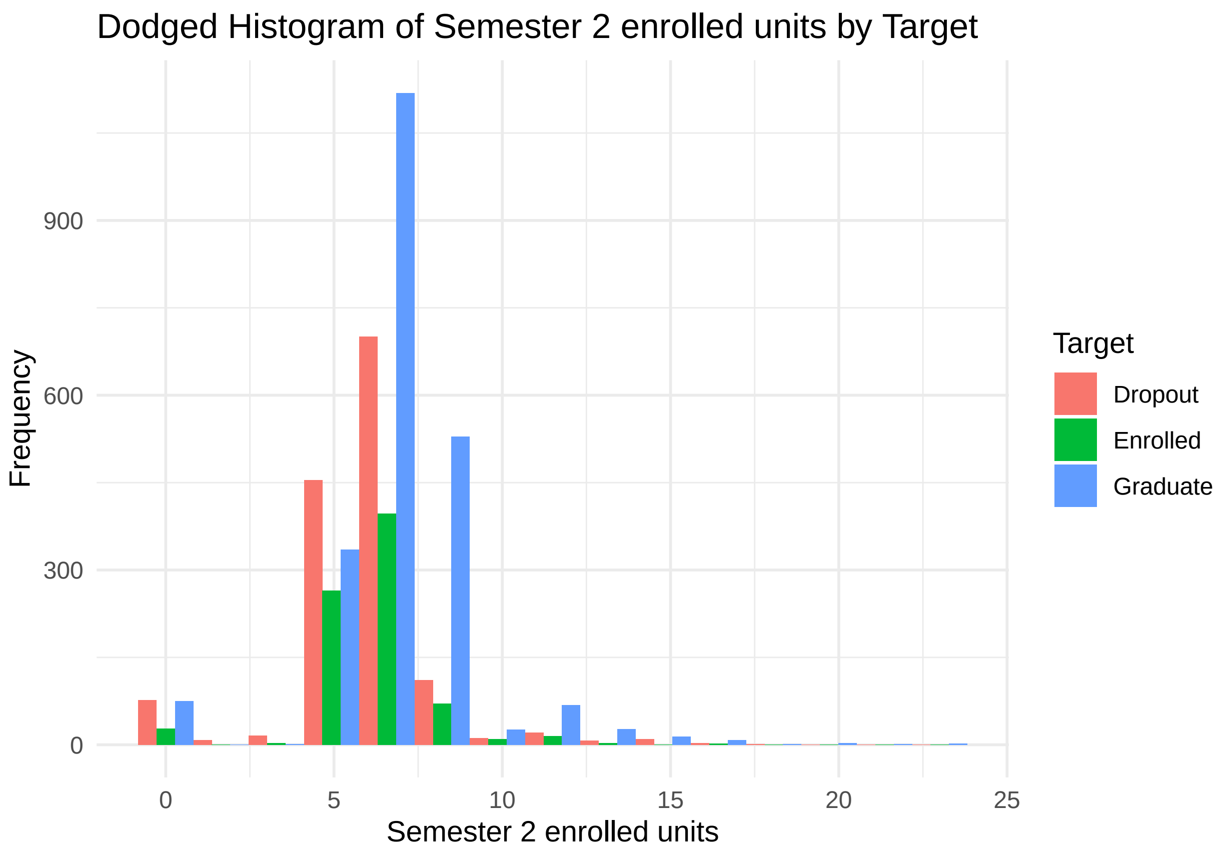

Dodged Histogram of Semester 1 evaluations by Target
1500
1000
500
0
Frequency
0
10
20
30
40
Semester 1 evaluations
Target
Dropout
Enrolled
Graduate

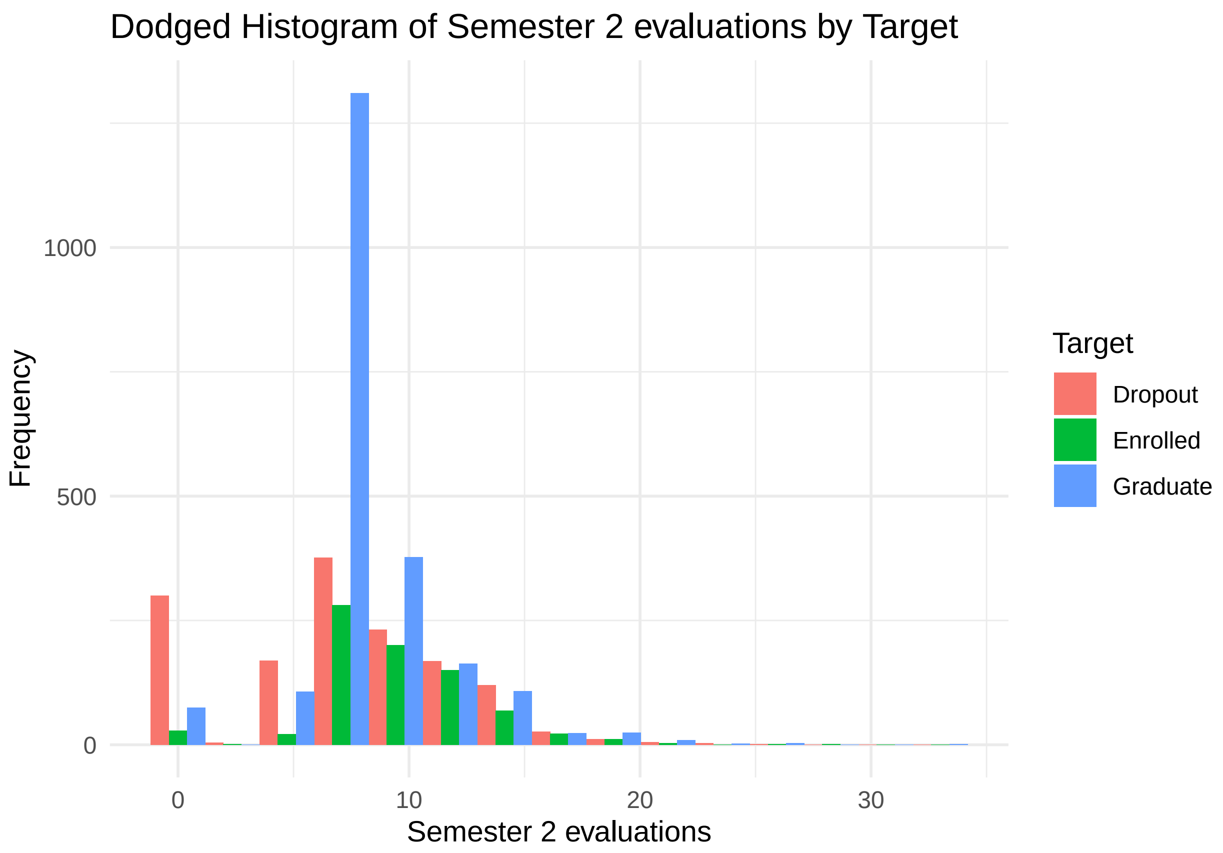

Dodged Histogram of Semester 1 approved units by Target
Frequency
1000
500
0
0
10
20
Semester 1 approved units
Target
Dropout
Enrolled
Graduate

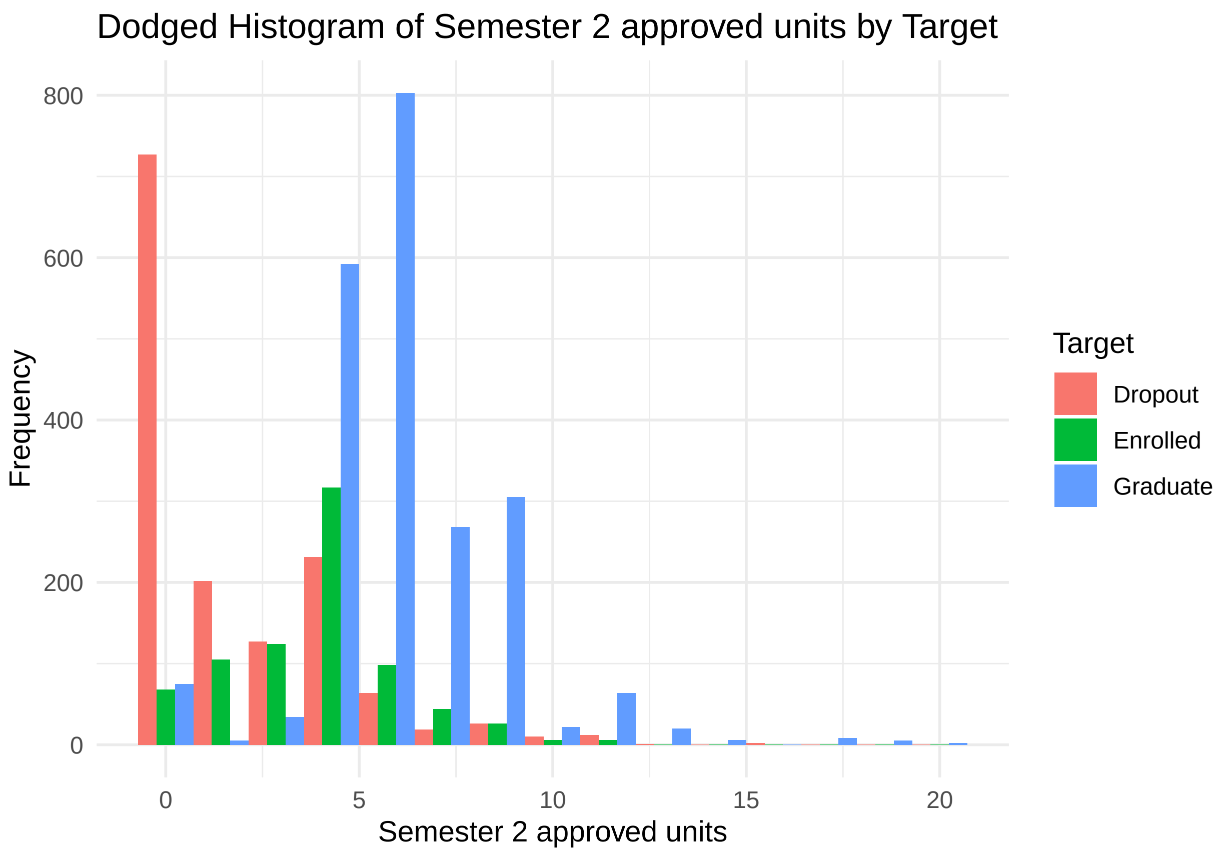

Dodged Histogram of Semester 1 grade by Target
Frequency
800
600
400
200
0
0
5
10
15
20
Semester 1 grade
Target
Dropout
Enrolled
Graduate

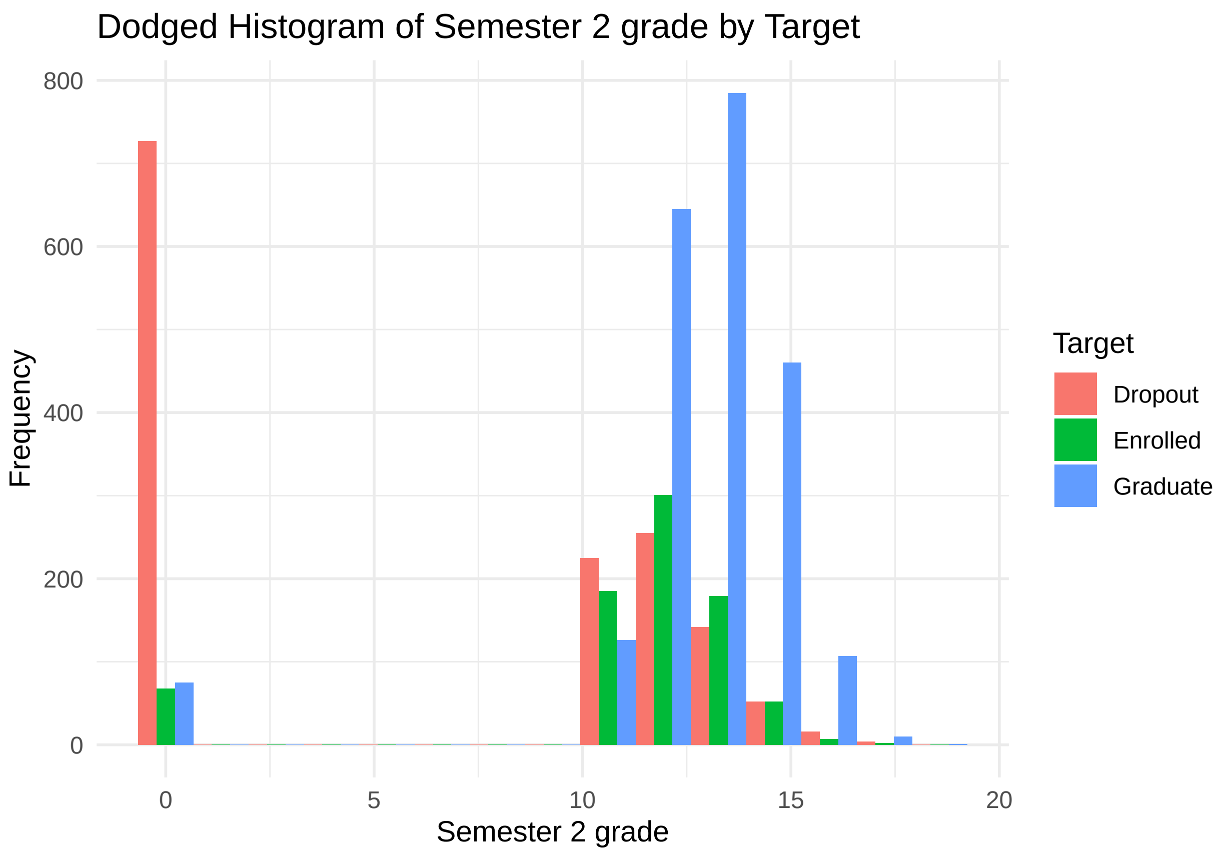

Dodged Histogram of Semester 1 units without evaluations by Target
Frequency
2000
1500
1000
500
0
0
4
8
12
Semester 1 units without evaluations
Target
Dropout
Enrolled
Graduate

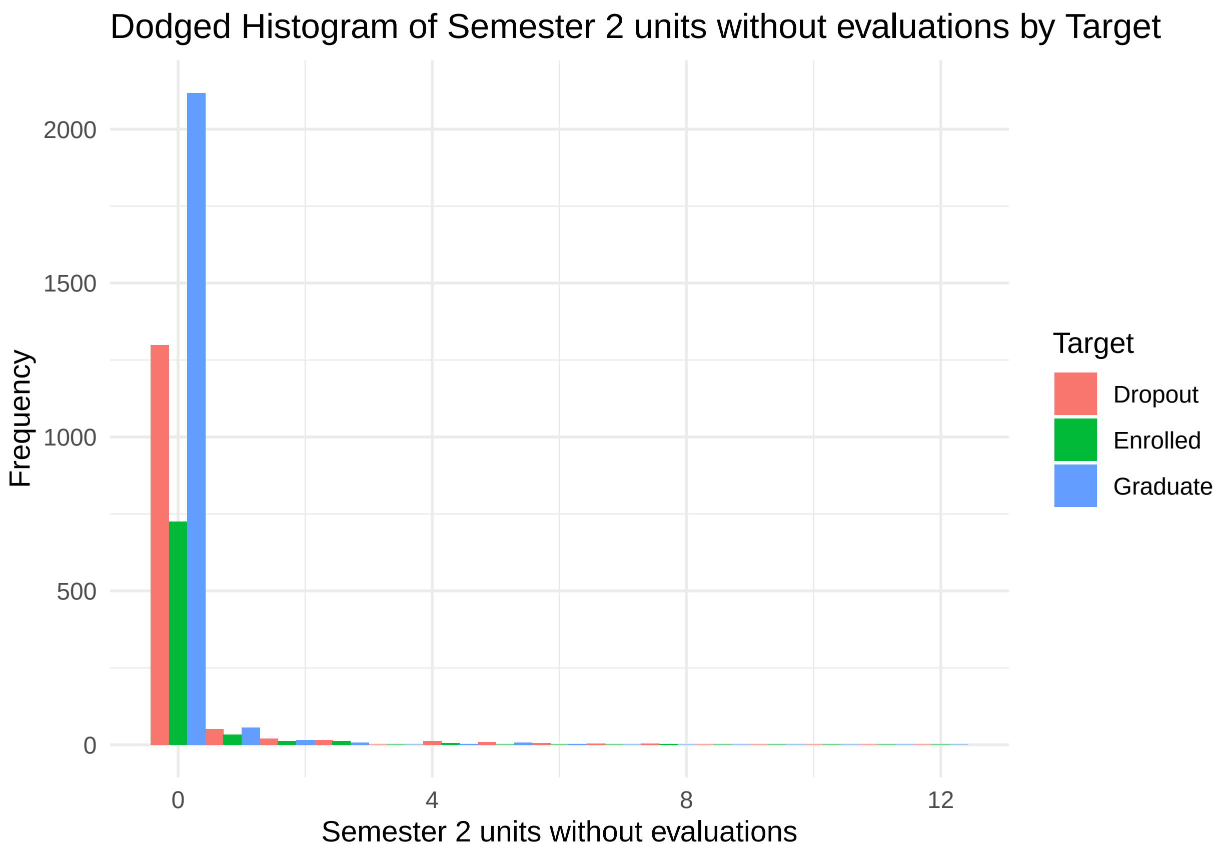

Dodged Histogram of Semester 2 credited units by Target
Frequency
2000
1500
1000
500
0
0
5
10
15
20
Semester 2 credited units
Target
Dropout
Enrolled
Graduate

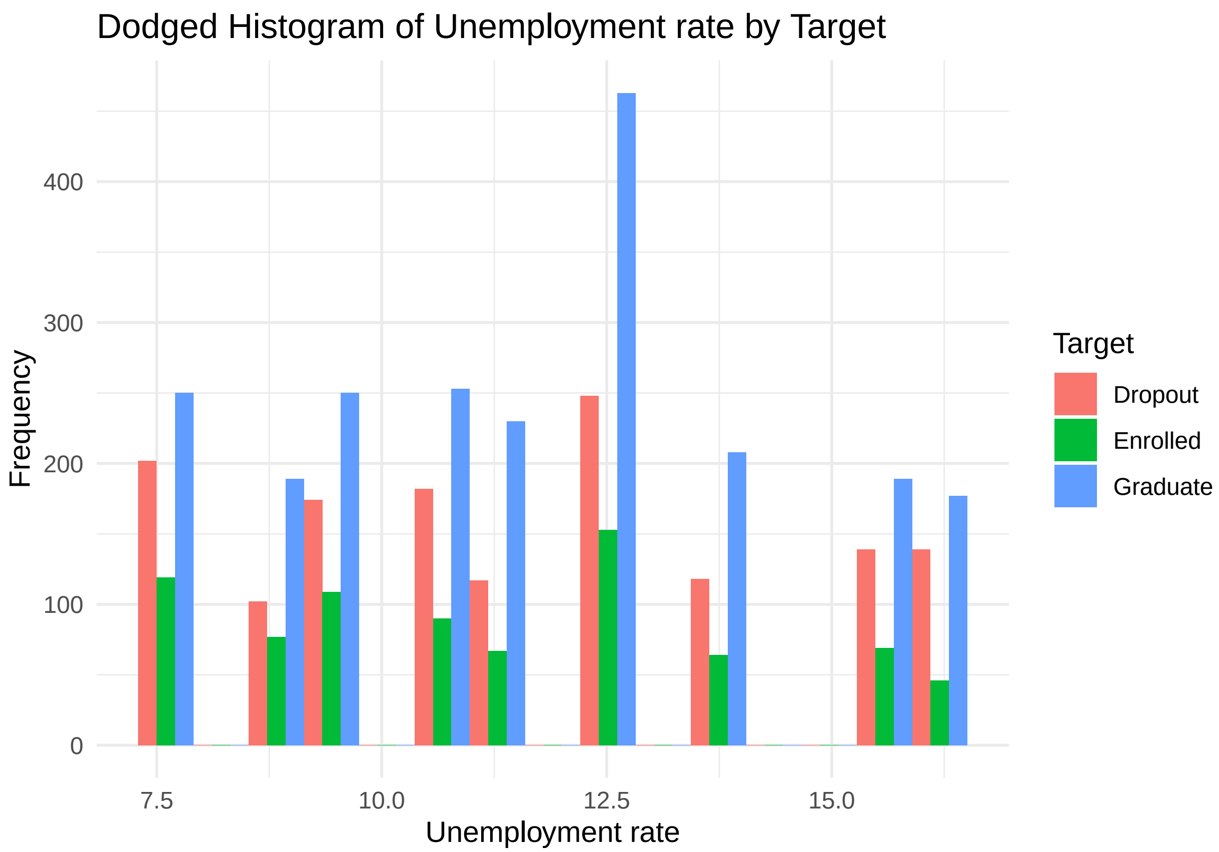

Dodged Histogram of Semester 2 enrolled units by Target
Frequency
900
600
300
0
0
5
10
15
20
25
Semester 2 enrolled units
Target
Dropout
Enrolled
Graduate

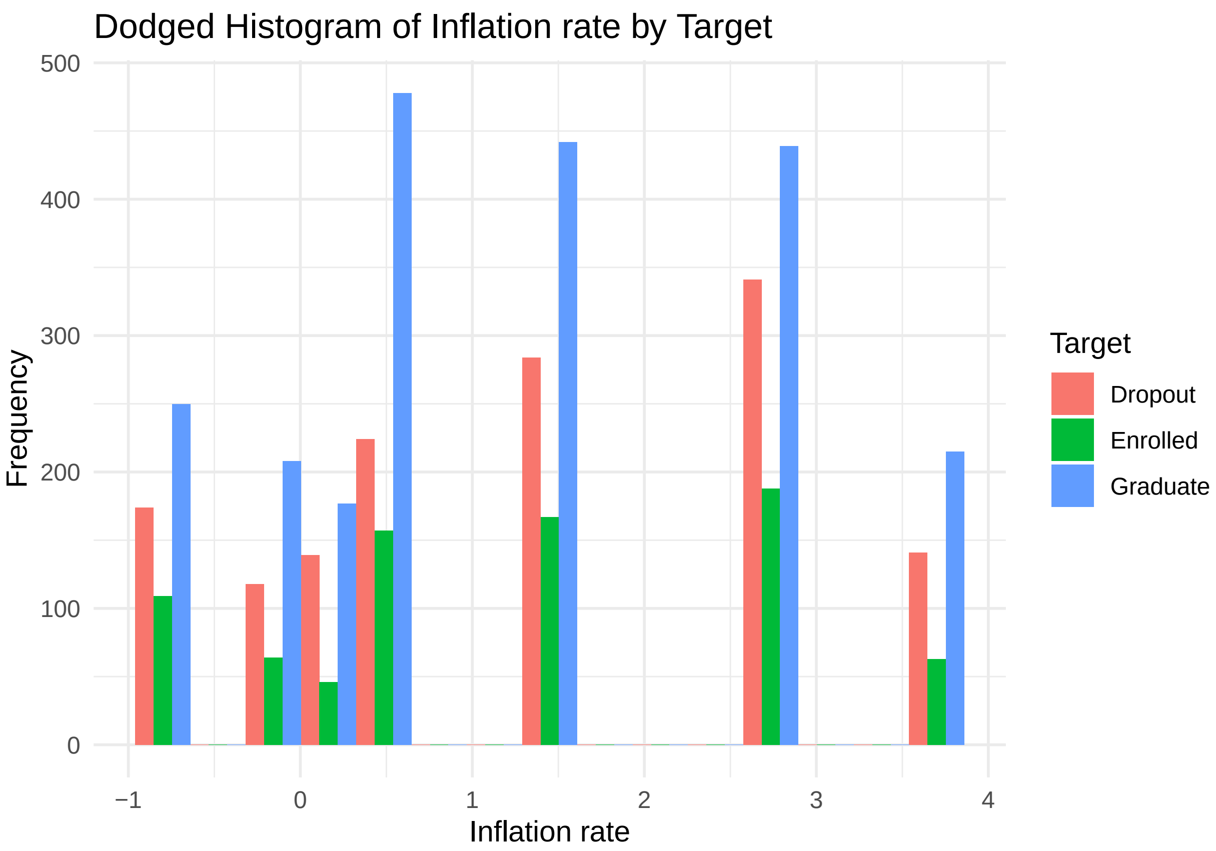

Dodged Histogram of Semester 2 evaluations by Target
Frequency
1000
500
0
0
10
20
30
Semester 2 evaluations
Target
Dropout
Enrolled
Graduate

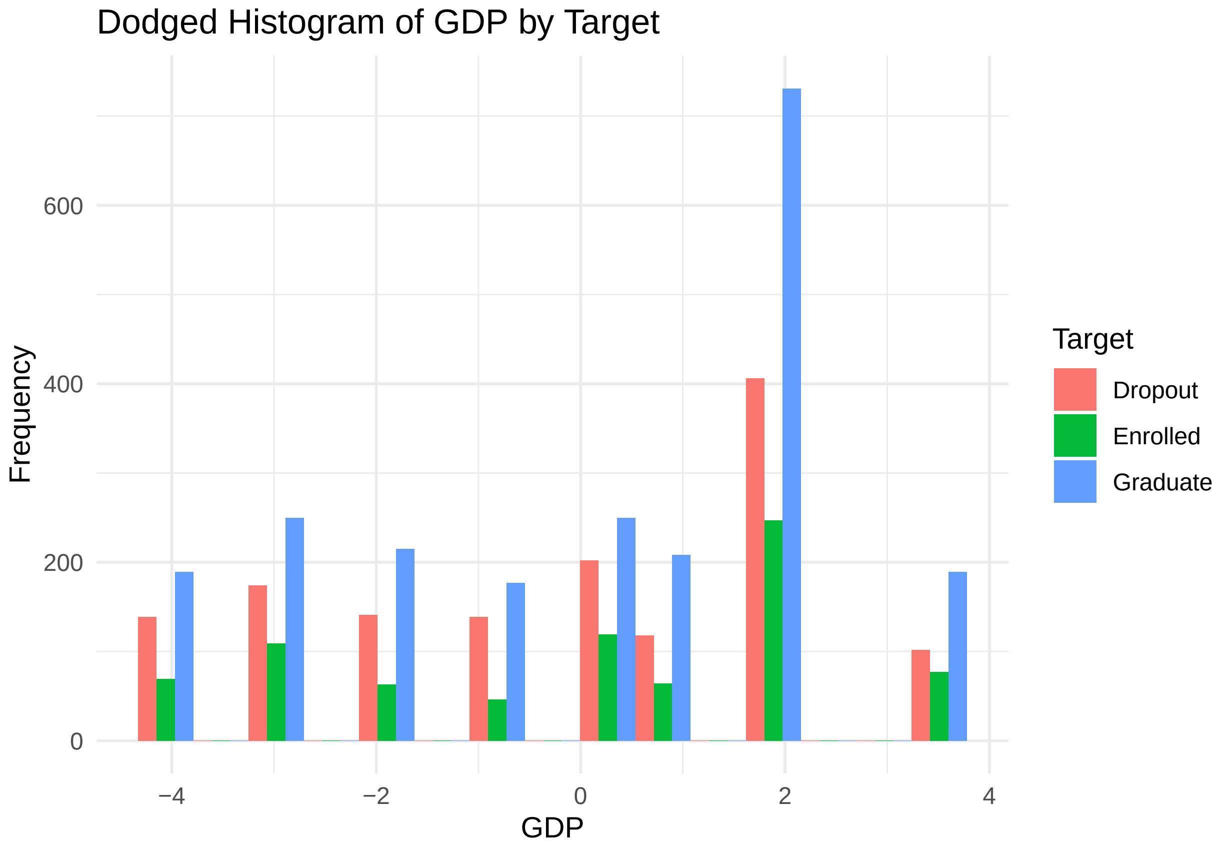

Dodged Histogram of Semester 2 approved units by Target
Frequency
800
600
400
200
0
0
5
10
15
20
Semester 2 approved units
Target
Dropout
Enrolled
Graduate


Dodged Histogram of Semester 2 grade by Target
Frequency
800
600
400
200
0
0
5
10
15
20
Semester 2 grade
Target
Dropout
Enrolled
Graduate


Dodged Histogram of Semester 2 units without evaluations by Target
Frequency
2000
1500
1000
500
0
0
4
8
12
Semester 2 units without evaluations
Target
Dropout
Enrolled
Graduate


Dodged Histogram of Unemployment rate by Target
Frequency
0
100
200
300
400
7.5
10.0
12.5
15.0
Unemployment rate
Target
Dropout
Enrolled
Graduate


Dodged Histogram of Inflation rate by Target
Frequency
500
400
300
200
100
0
−1
0
1
2
3
4
Inflation rate
Target
Dropout
Enrolled
Graduate



Categorical variables showing little variation: - marital status (single) - application order (2nd) - daytime/evening attendance (daytime) - previous qualification (secondary education) - nationality (Portuguese) - educational special needs (no) - debtor (no) - tuition fees up to date (yes) - international (no)

Other notes: - course with highest frequency is nursing - more displaced than not, but almost half and half - about 2/3 women, 1/3 men - about 3/4 without scholarship, 1/4 with scholarship

Numeric columns with almost all zeros: - Semester 1 credited units - Semester 1 units without evaluations - Semester 2 credited units - Semester 2 units without evaluations

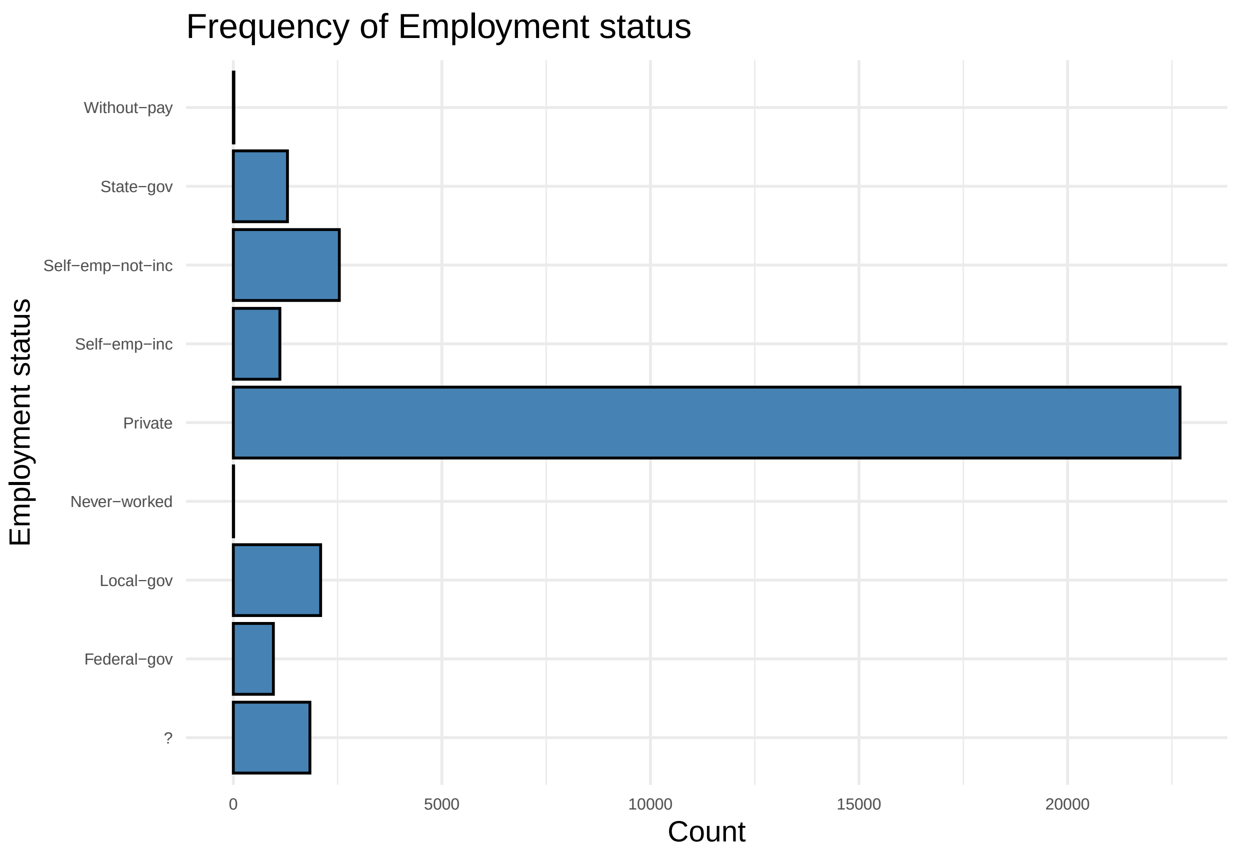
Notable variables: - Debtor (for Dropout) - Tuition fees up to date (for Dropout) - Scholarship holder - Semester 1 approved units - Semester 1 grade - Semester 2 approved units - Semester 2 grade - Semester 1 evaluations - Semester 2 evaluations

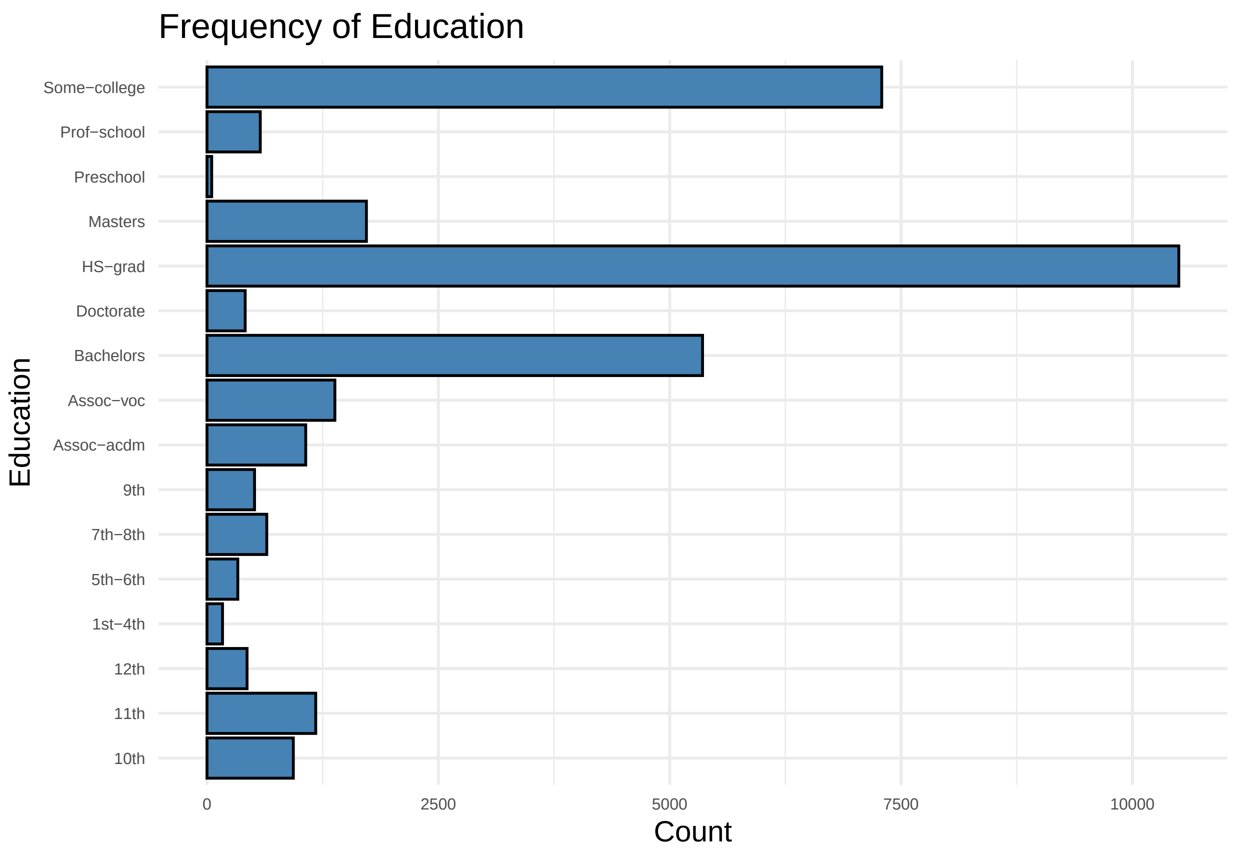
Histograms for income dataset: https://archive.ics.uci.edu/dataset/2/adult https://www.kaggle.com/code/yashhvyass/adult-census-income-logistic-reg-explained-86-2

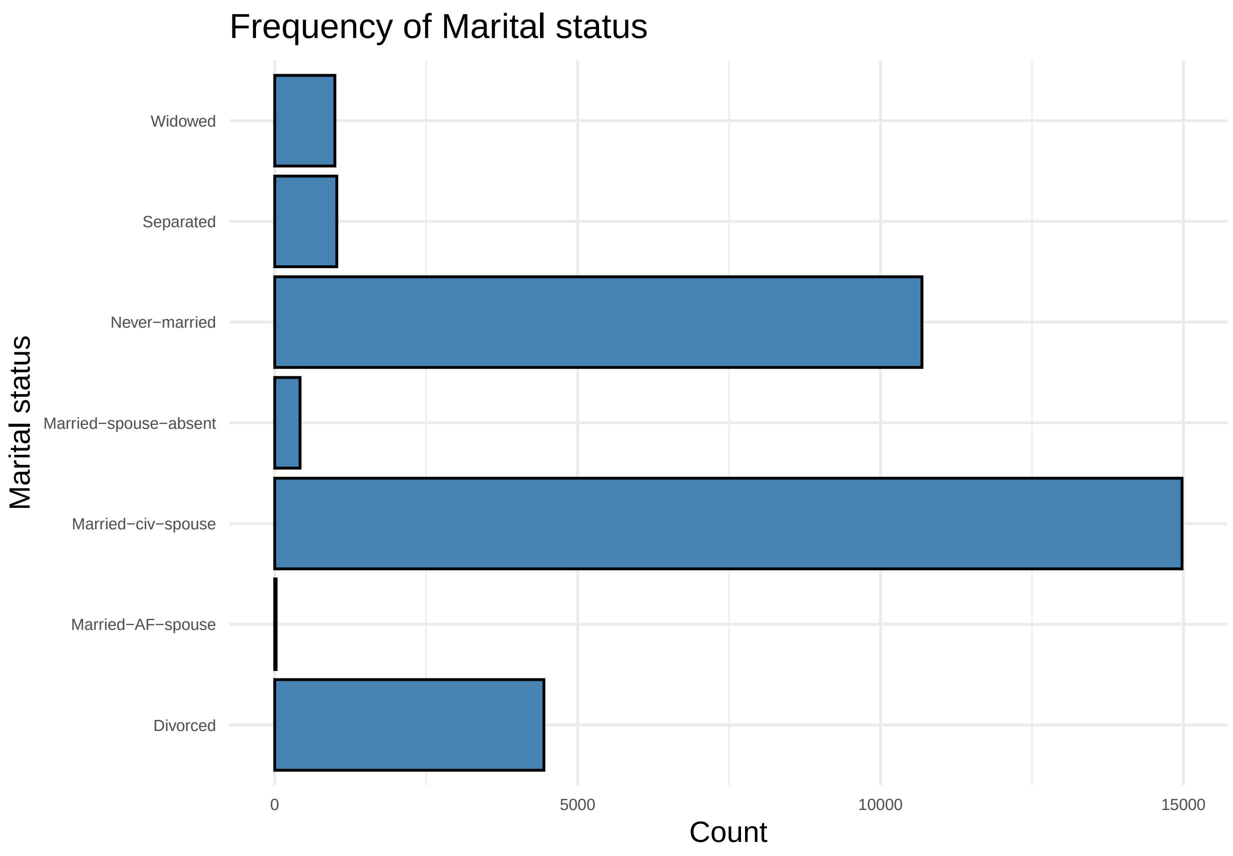
```
data2 <- read.csv("C:/Users/harip/Downloads/adult/adult.data", header=FALSE)
head(data2)
```

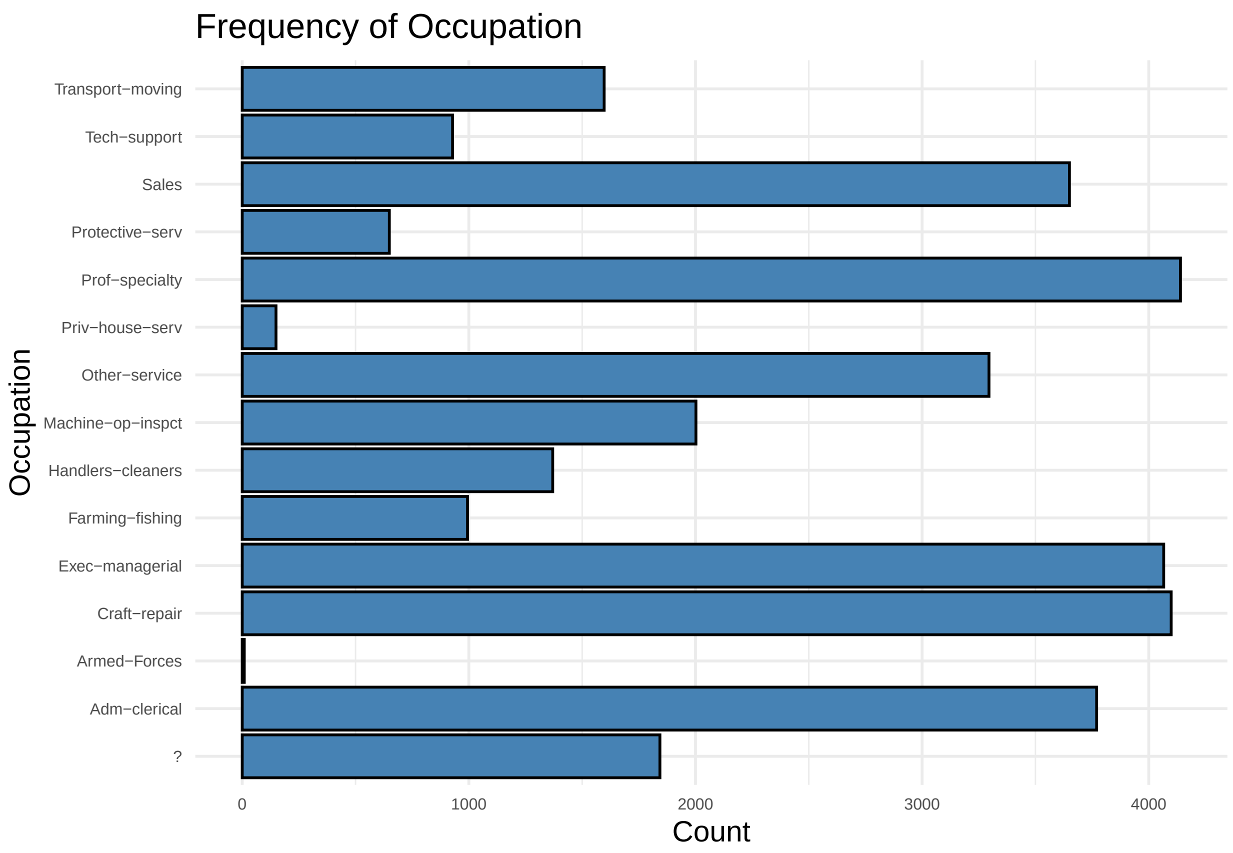
```
##   V1                V2     V3         V4 V5                  V6
## 1 39         State-gov  77516  Bachelors 13       Never-married
## 2 50  Self-emp-not-inc  83311  Bachelors 13  Married-civ-spouse
## 3 38           Private 215646    HS-grad  9            Divorced
## 4 53           Private 234721       11th  7  Married-civ-spouse
## 5 28           Private 338409  Bachelors 13  Married-civ-spouse
## 6 37           Private 284582    Masters 14  Married-civ-spouse
##                     V7             V8      V9     V10   V11  V12 V13            V14
## 1       Adm-clerical  Not-in-family  White    Male 2174   0  40  United-States
## 2    Exec-managerial        Husband  White    Male    0   0  13  United-States
## 3  Handlers-cleaners  Not-in-family  White    Male    0   0  40  United-States
## 4  Handlers-cleaners        Husband  Black    Male    0   0  40  United-States
## 5     Prof-specialty           Wife  Black  Female    0   0  40           Cuba
## 6    Exec-managerial           Wife  White  Female    0   0  40  United-States
##      V15
## 1  <=50K
## 2  <=50K
## 3  <=50K
## 4  <=50K
## 5  <=50K
## 6  <=50K
```

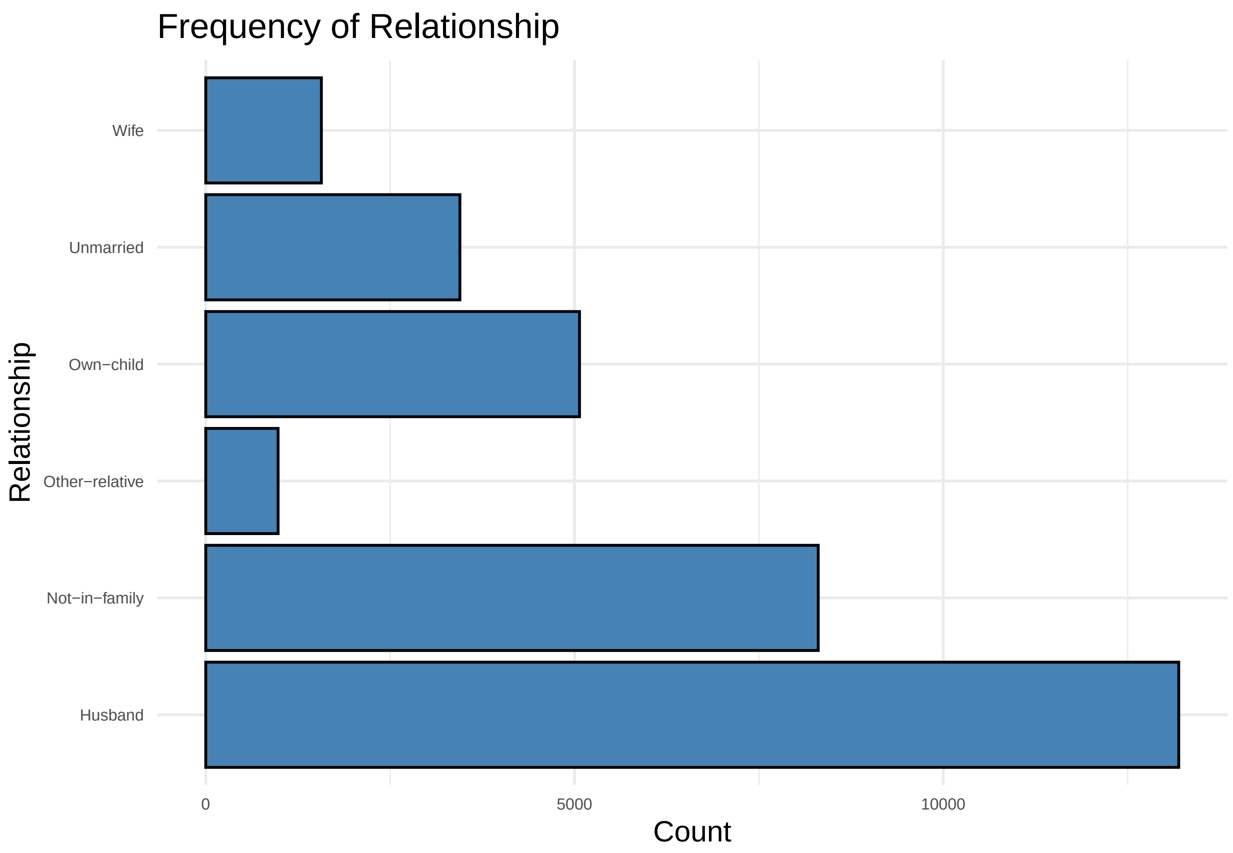
```
# final weight is the number of people the Census believes the entry represents
names(data2) <- c("Age",
                  "Employment status",
                  "Final weight",
                  "Education",
                  "Years of education",
                  "Marital status",
                  "Occupation",
                  "Relationship",
                  "Race",
                  "Sex",
                  "Capital gain",
                  "Capital loss",
                  "Hours worked per week",
                  "Country of origin",
                  "Target")

categorical_cols <- c("Employment status", "Education", "Marital status", "Occupation", "Relationship",

library(ggplot2)
library(scales)

for (i in categorical_cols) {
  plot <- ggplot(data2, aes(x = .data[[i]])) +
    geom_bar(fill = "steelblue", color = "black") +
    labs(
      title = paste("Frequency of", i),
      x = i,
      y = "Count"
    ) +
    theme_minimal() +
    scale_x_discrete(
      labels = label_wrap(150)
    ) +
    coord_flip() + theme(
      axis.text = element_text(size = 6)
    )
  print(plot)
}
```

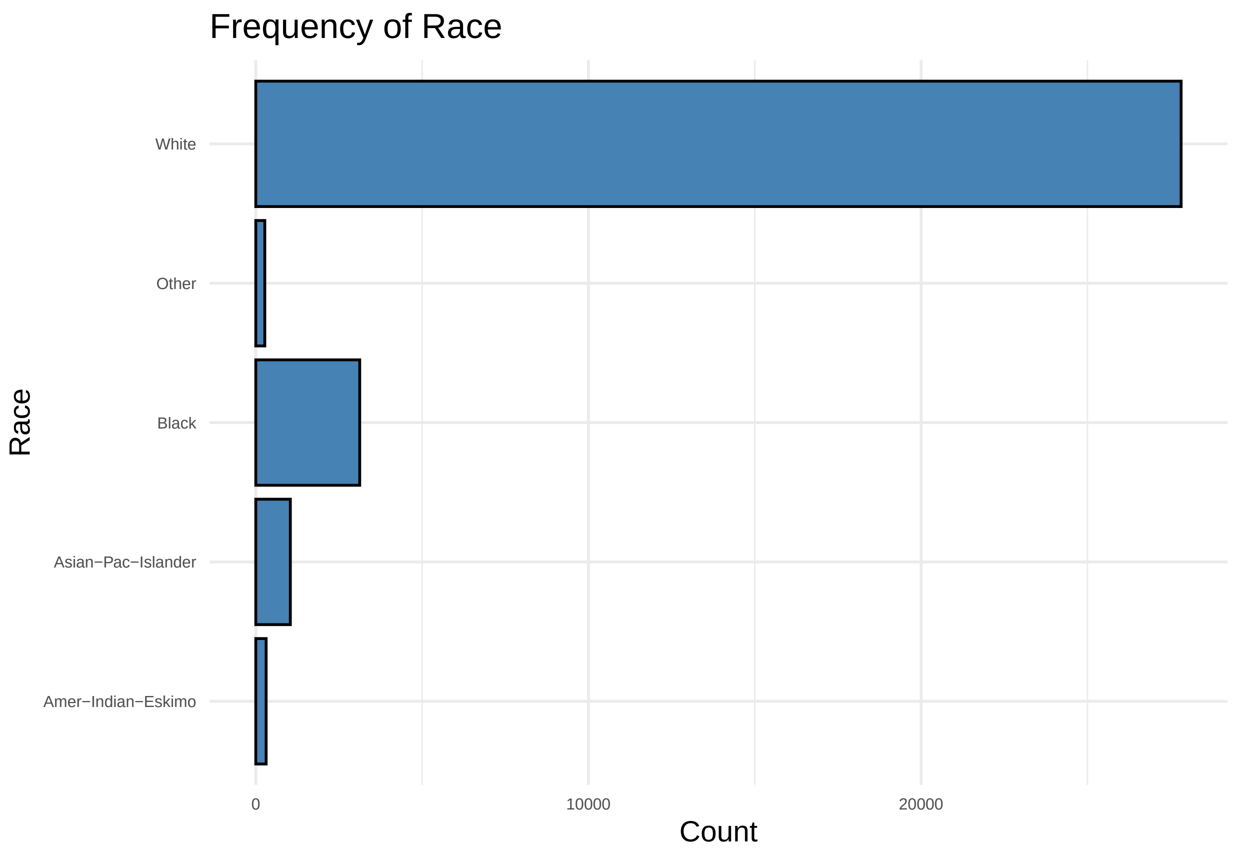

Frequency of Employment status
Employment status
Without−pay
State−gov
Self−emp−not−inc
Self−emp−inc
Private
Never−worked
Local−gov
Federal−gov
?
0
5000
10000
15000
20000
Count

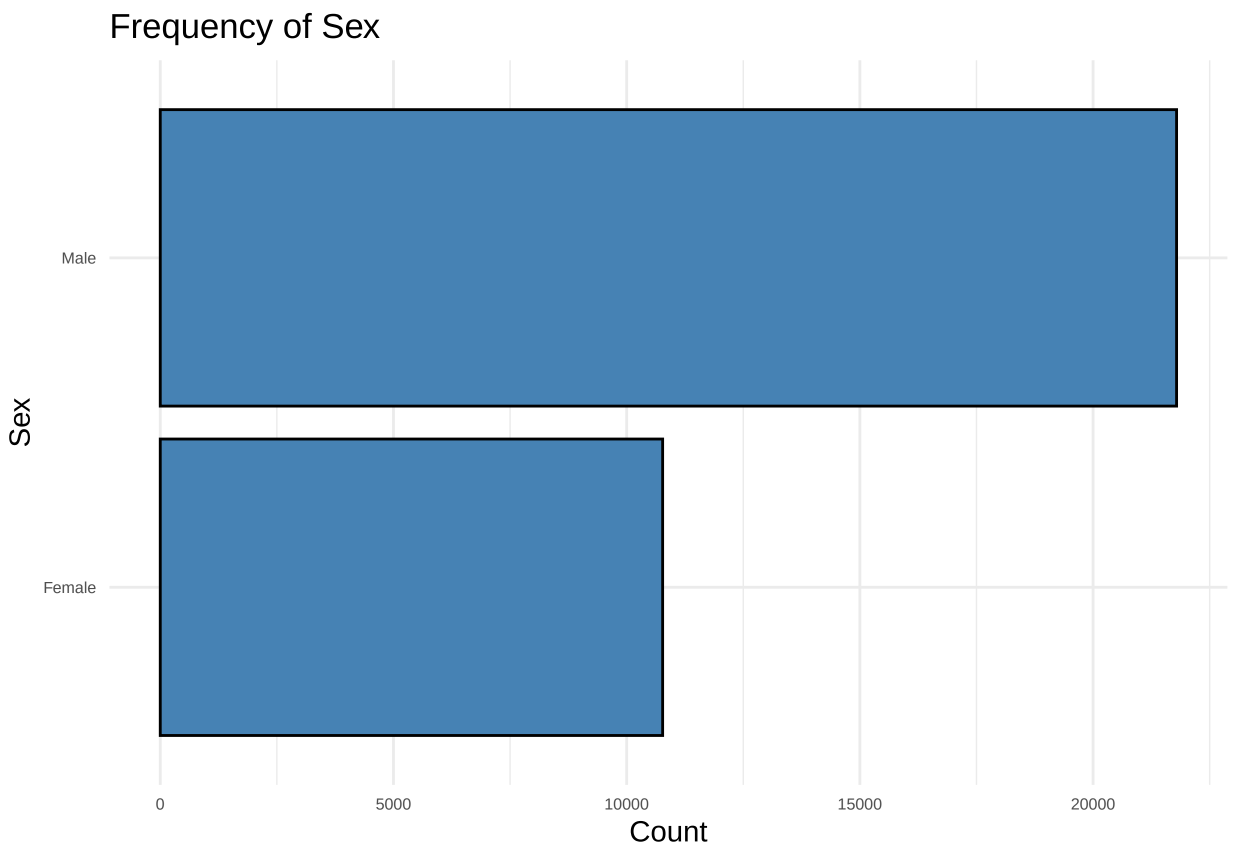

Frequency of Education
Education
Some−college
Prof−school
Preschool
Masters
HS−grad
Doctorate
Bachelors
Assoc−voc
Assoc−acdm
9th
7th−8th
5th−6th
1st−4th
12th
11th
10th
0
2500
5000
7500
10000
Count

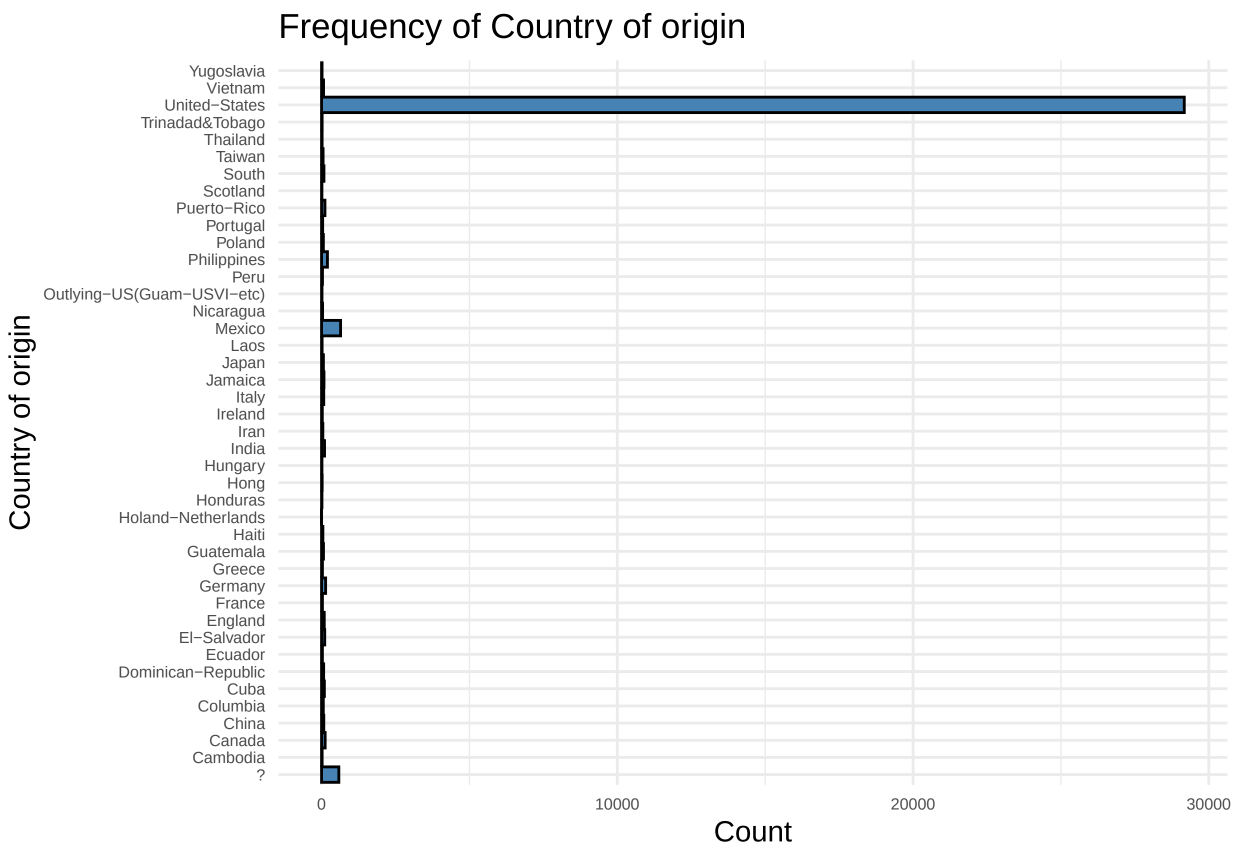

Frequency of Marital status
Marital status
Widowed
Separated
Never−married
Married−spouse−absent
Married−civ−spouse
Married−AF−spouse
Divorced
0
5000
10000
15000
Count


Frequency of Occupation
Occupation
Transport−moving
Tech−support
Sales
Protective−serv
Prof−specialty
Priv−house−serv
Other−service
Machine−op−inspct
Handlers−cleaners
Farming−fishing
Exec−managerial
Craft−repair
Armed−Forces
Adm−clerical
?
0
1000
2000
3000
4000
Count

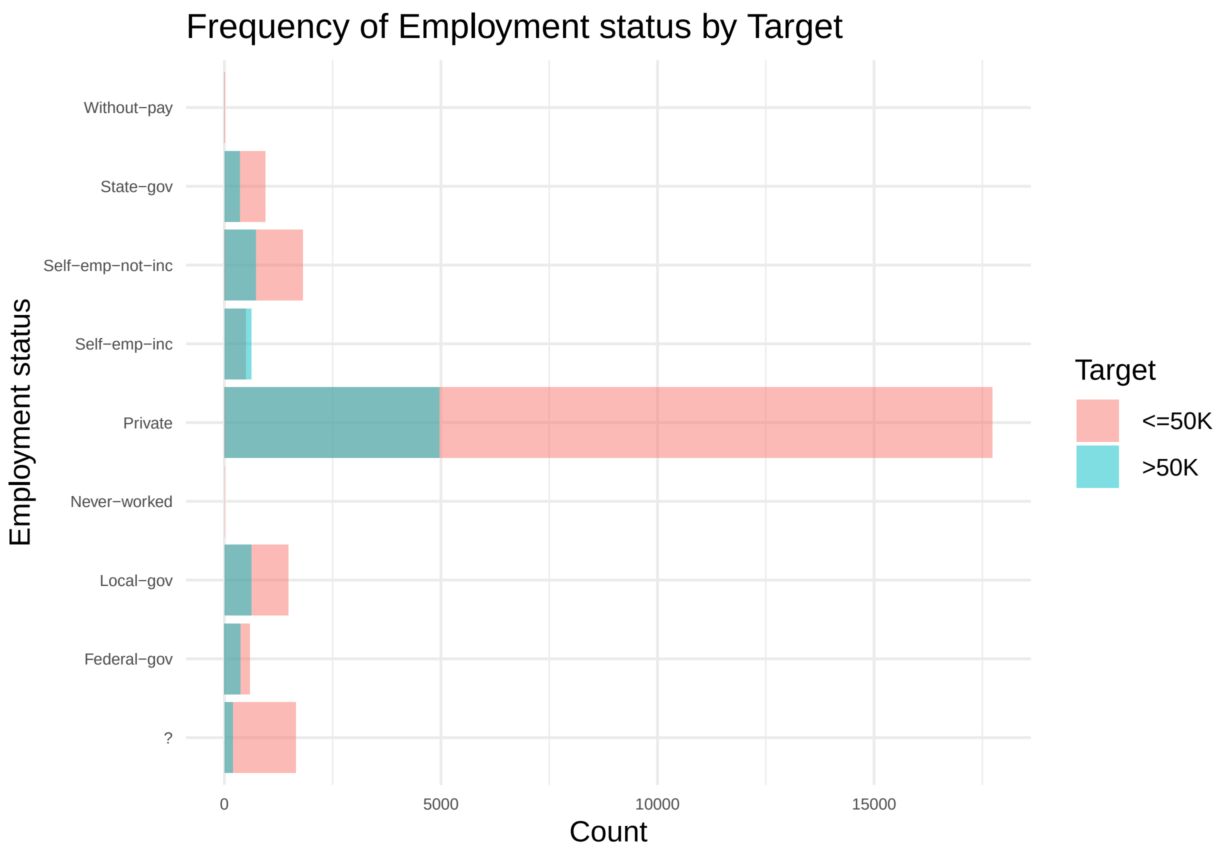

Frequency of Relationship
Relationship
Wife
Unmarried
Own−child
Other−relative
Not−in−family
Husband
0
5000
10000
Count

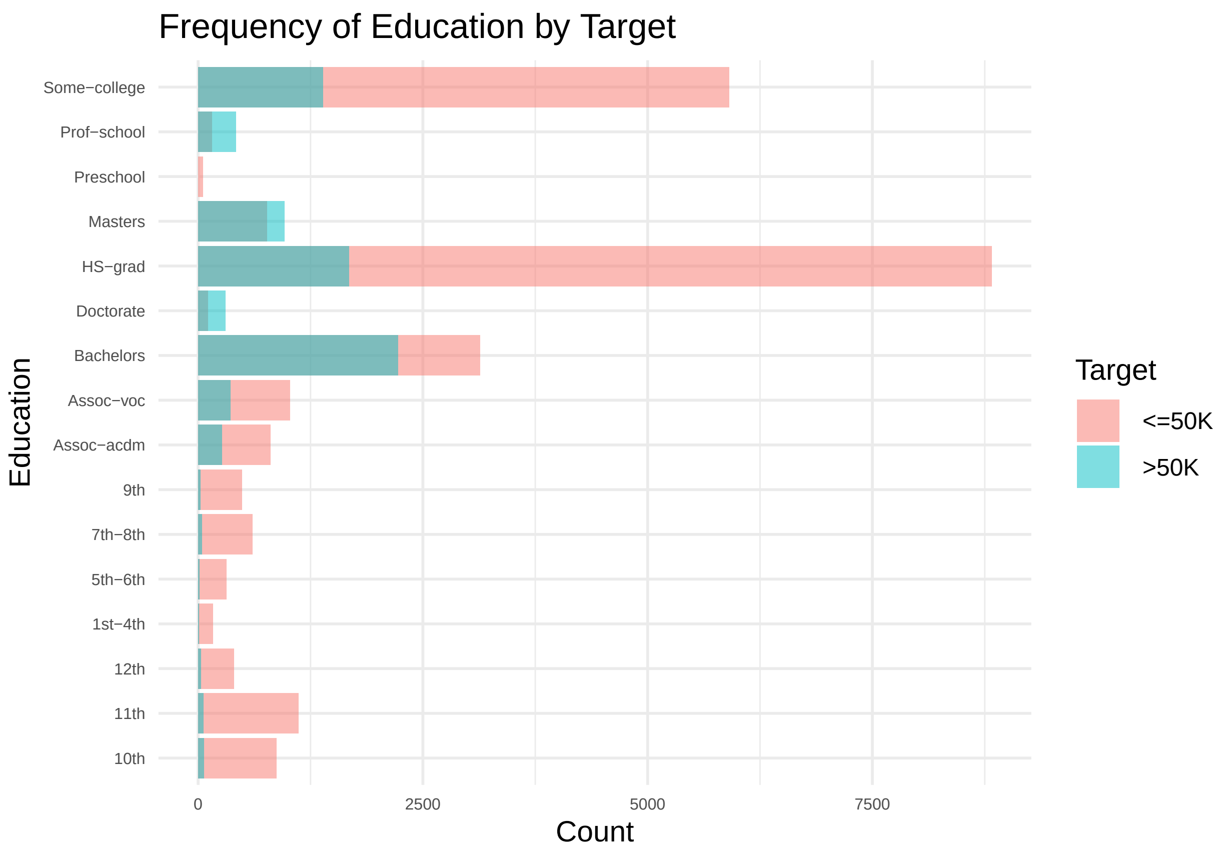

Frequency of Race
White
Other
Black
Asian−Pac−Islander
Amer−Indian−Eskimo
Race
0
10000
20000
Count

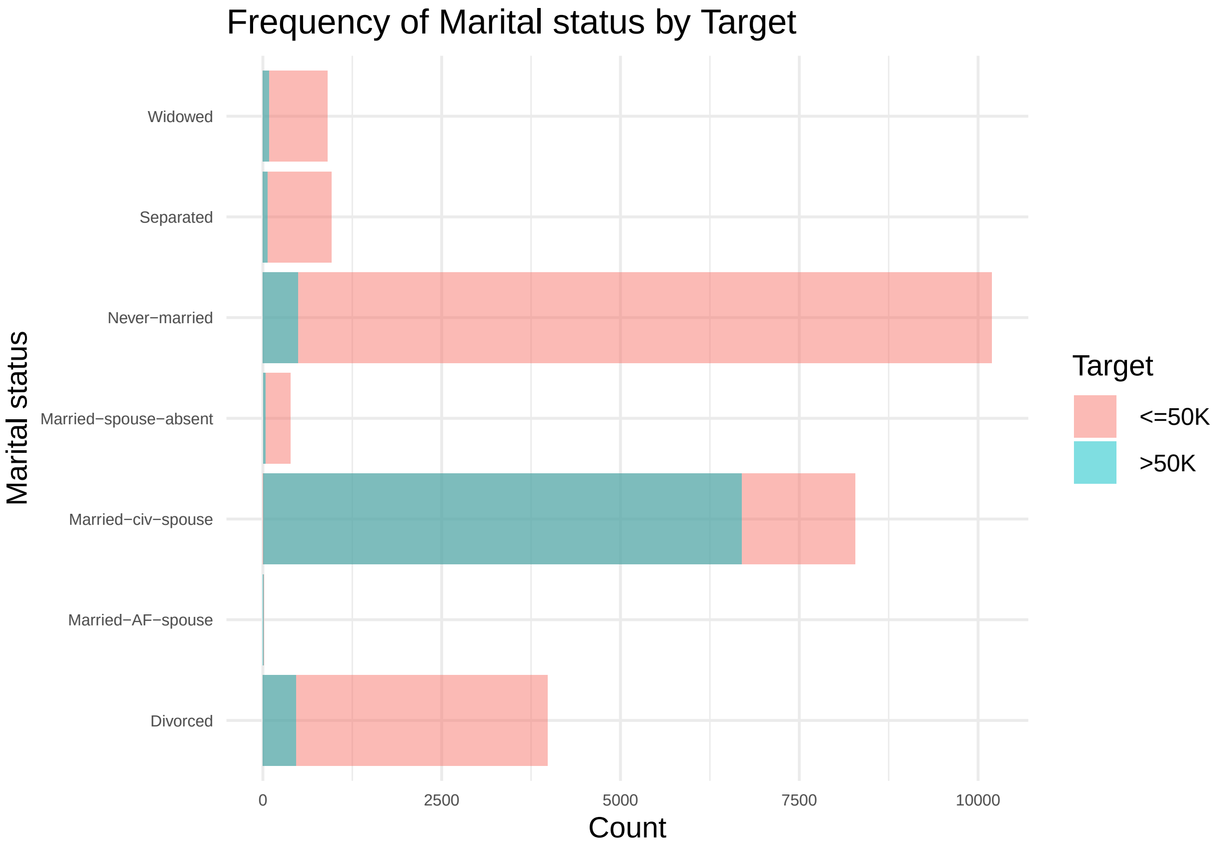

Frequency of Sex
Male
Female
Sex
0
5000
10000
15000
20000
Count

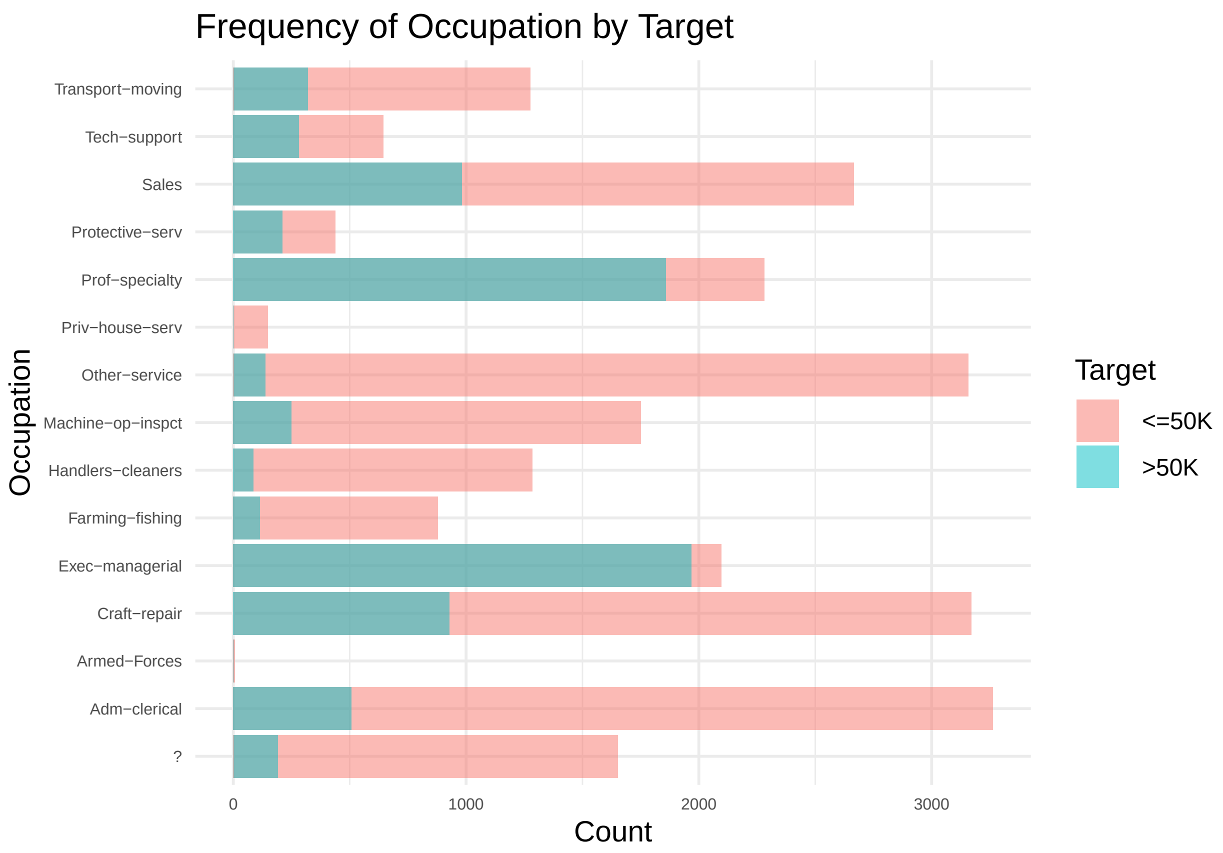


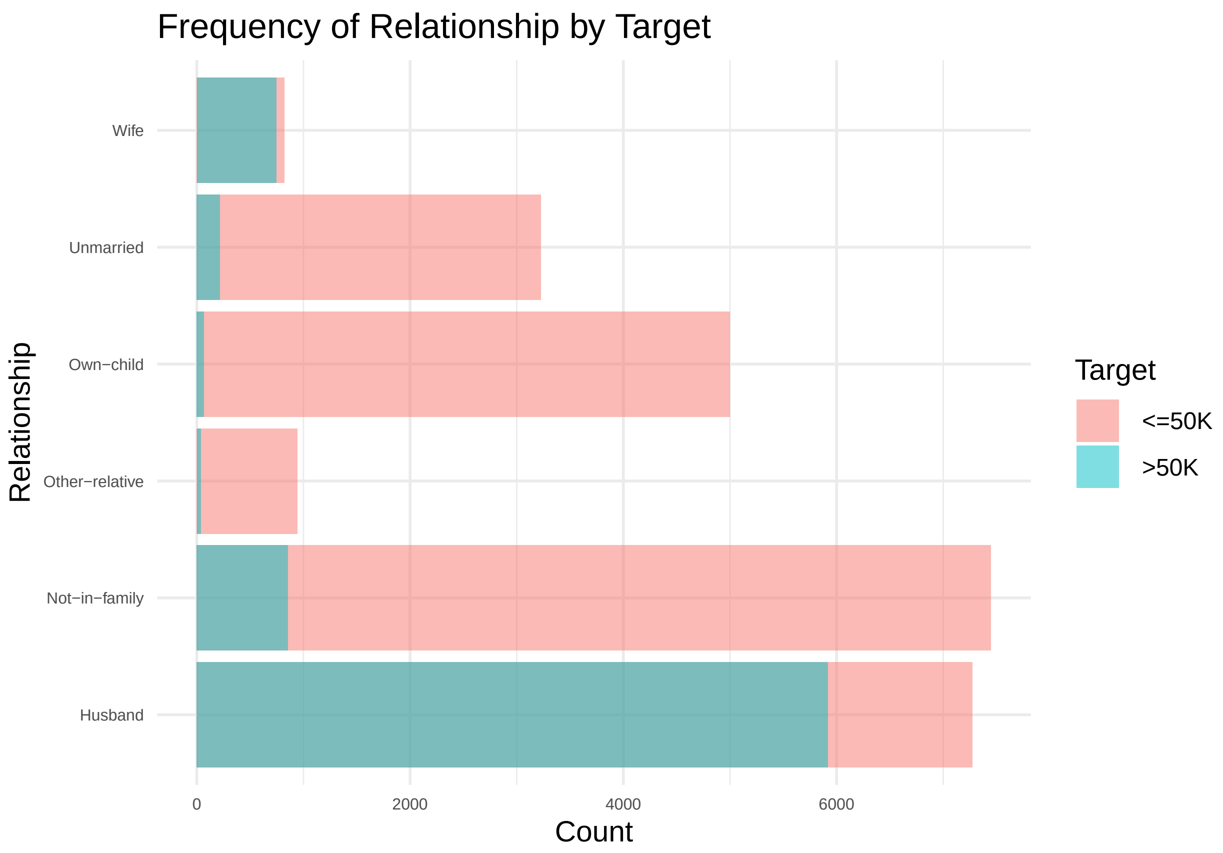
```
for (i in categorical_cols) {
  plot <- ggplot(data2, aes(x = .data[[i]], fill = Target)) +
    geom_bar(position = "identity", alpha=0.5) +
    labs(
      title = paste("Frequency of", i, "by Target"),
      x = i,
      y = "Count"
    ) +
    theme_minimal() +
    scale_x_discrete(
      labels = label_wrap(150)
    ) +
    coord_flip() + theme(
      axis.text = element_text(size = 6)
    )
  print(plot)
}
```

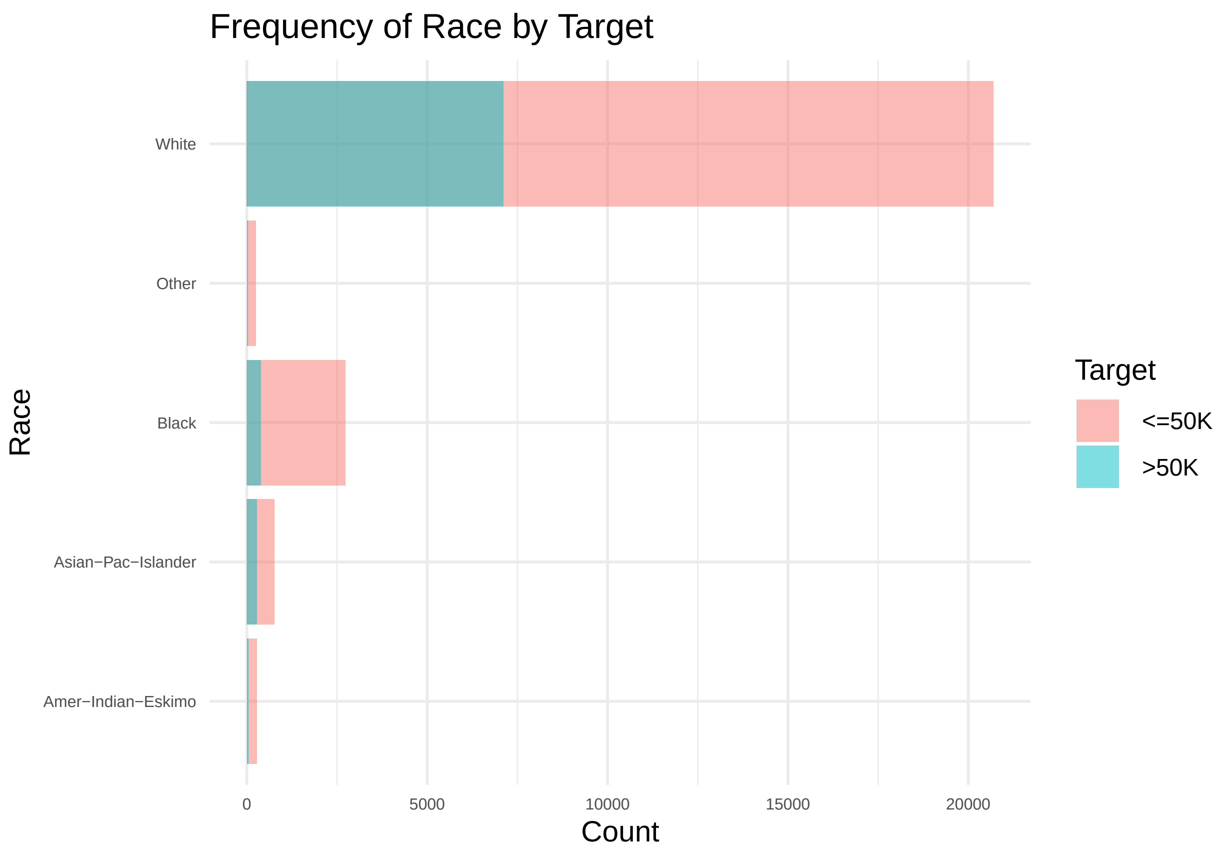

Frequency of Employment status by Target
Employment status
Without−pay
State−gov
Self−emp−not−inc
Self−emp−inc
Private
Never−worked
Local−gov
Federal−gov
?
0
5000
10000
15000
Count
Target
<=50K
>50K

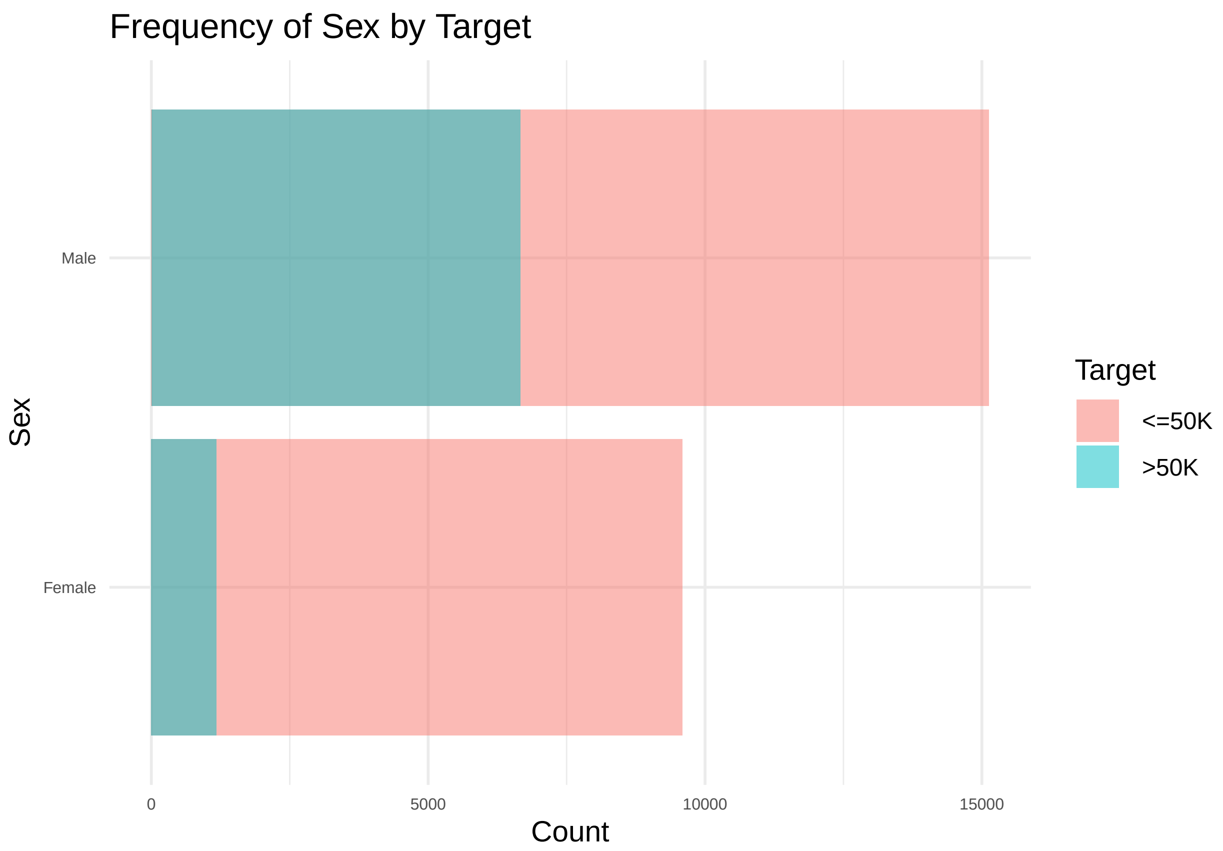

Frequency of Education by Target
Education
Some−college
Prof−school
Preschool
Masters
HS−grad
Doctorate
Bachelors
Assoc−voc
Assoc−acdm
9th
7th−8th
5th−6th
1st−4th
12th
11th
10th
0
2500
5000
7500
Count
Target
<=50K
>50K

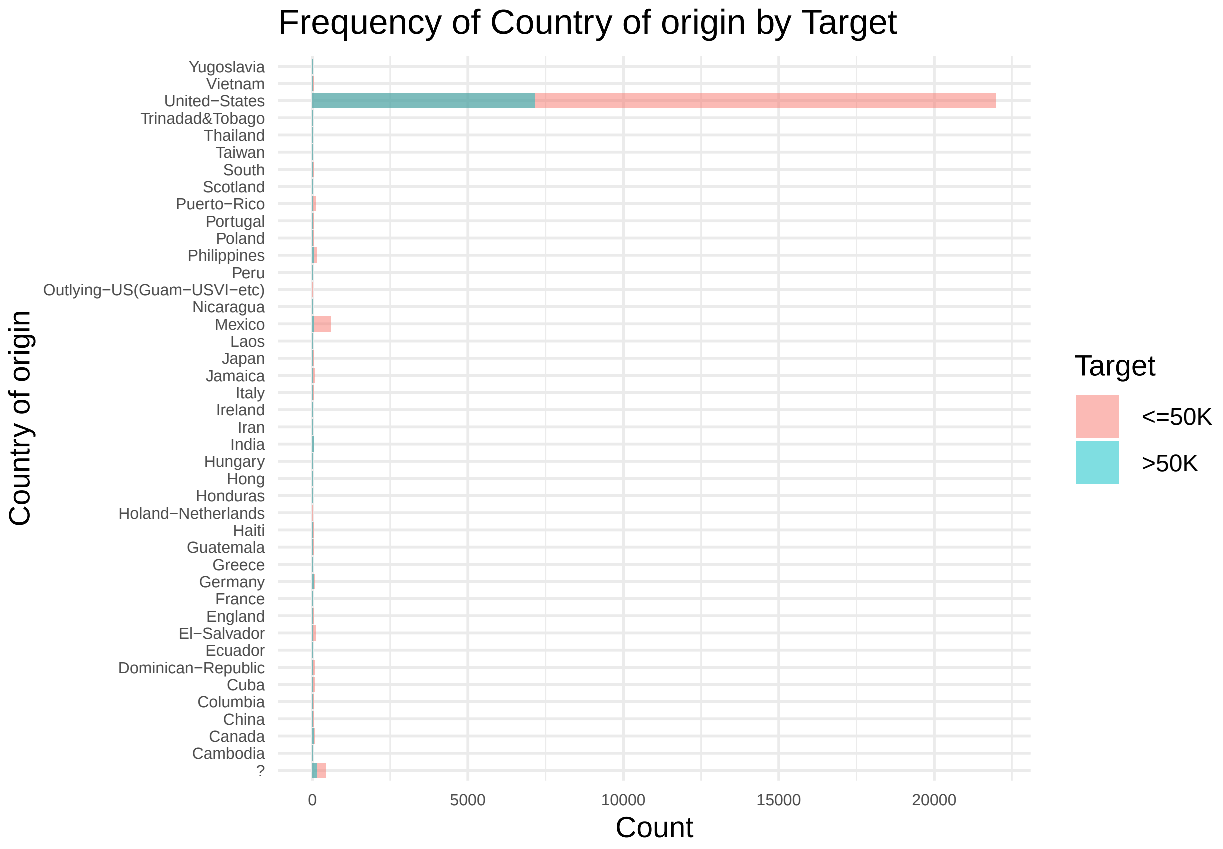

Frequency of Marital status by Target
Marital status
Widowed
Separated
Never−married
Married−spouse−absent
Married−civ−spouse
Married−AF−spouse
Divorced
0
2500
5000
7500
10000
Count
Target
<=50K
>50K


Frequency of Occupation by Target
Occupation
Transport−moving
Tech−support
Sales
Protective−serv
Prof−specialty
Priv−house−serv
Other−service
Machine−op−inspct
Handlers−cleaners
Farming−fishing
Exec−managerial
Craft−repair
Armed−Forces
Adm−clerical
?
0
1000
2000
3000
Count
Target
<=50K
>50K

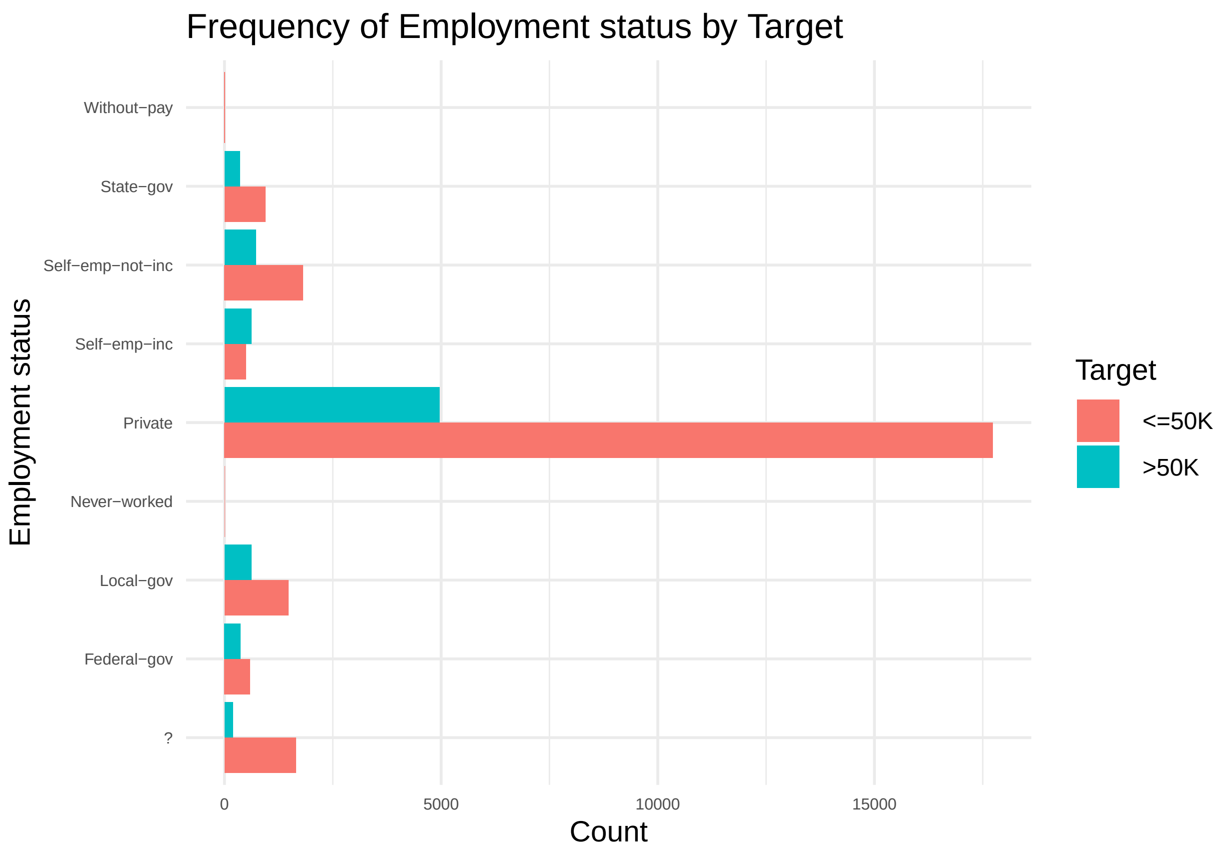

Frequency of Relationship by Target
Relationship
Wife
Unmarried
Own−child
Other−relative
Not−in−family
Husband
0
2000
4000
6000
Count
Target
<=50K
>50K

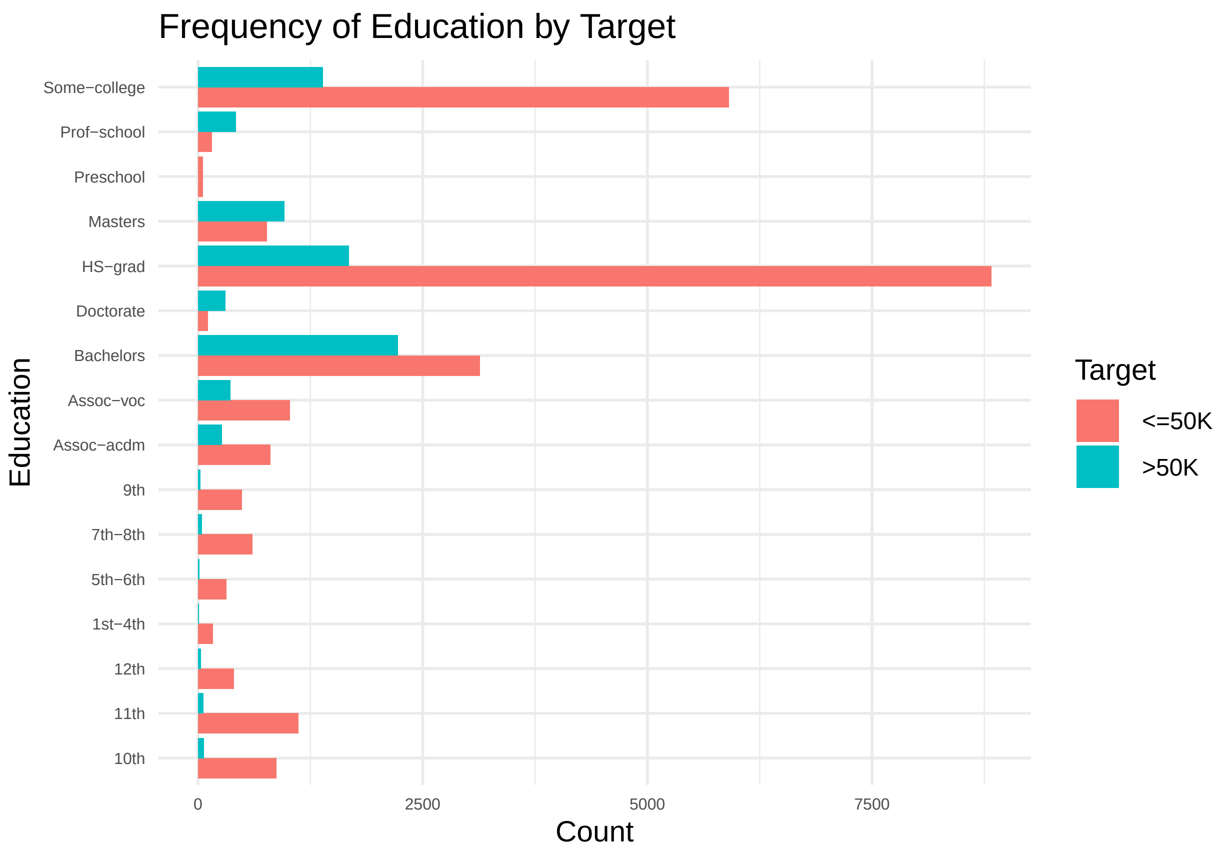

Frequency of Race by Target
Race
White
Other
Black
Asian−Pac−Islander
Amer−Indian−Eskimo
0
5000
10000
15000
20000
Count
Target
<=50K
>50K

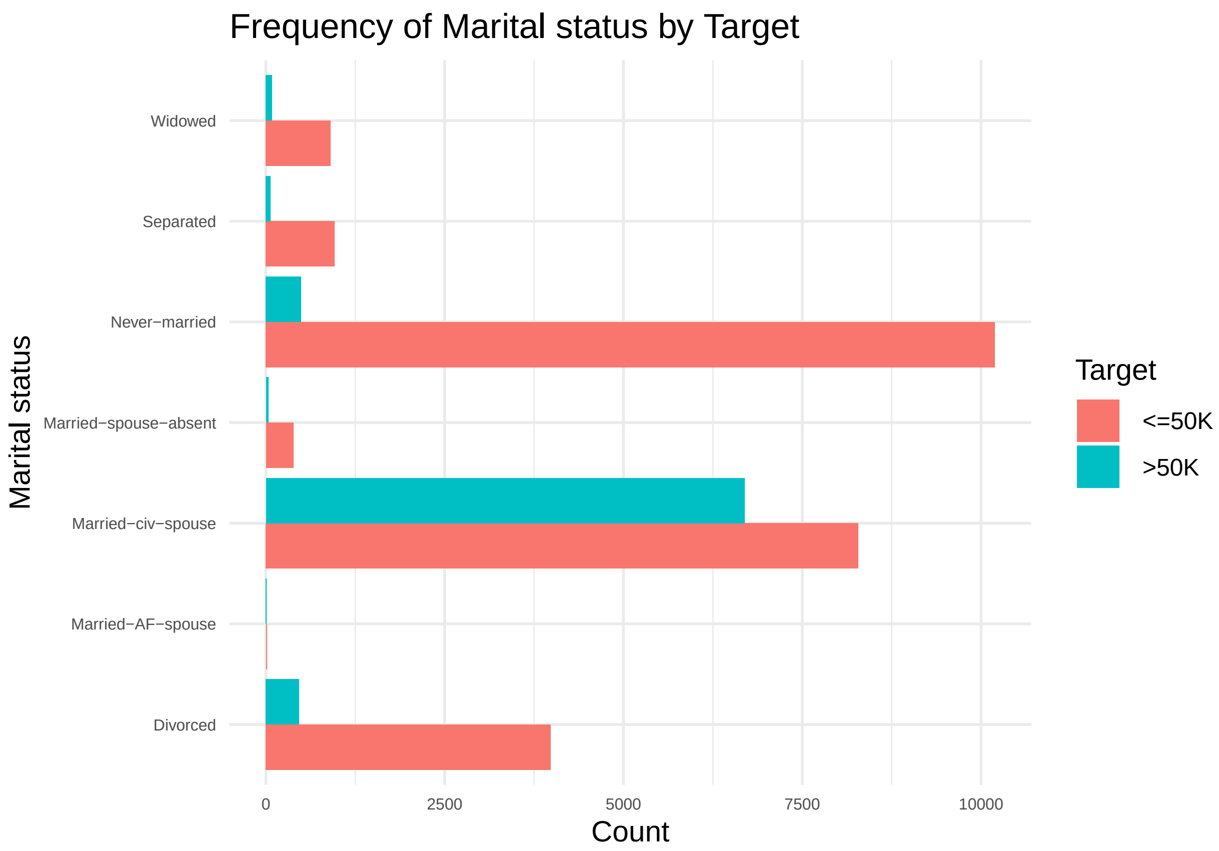

Frequency of Sex by Target
Sex
Male
Female
Count
0
5000
10000
15000
Target
<=50K
>50K

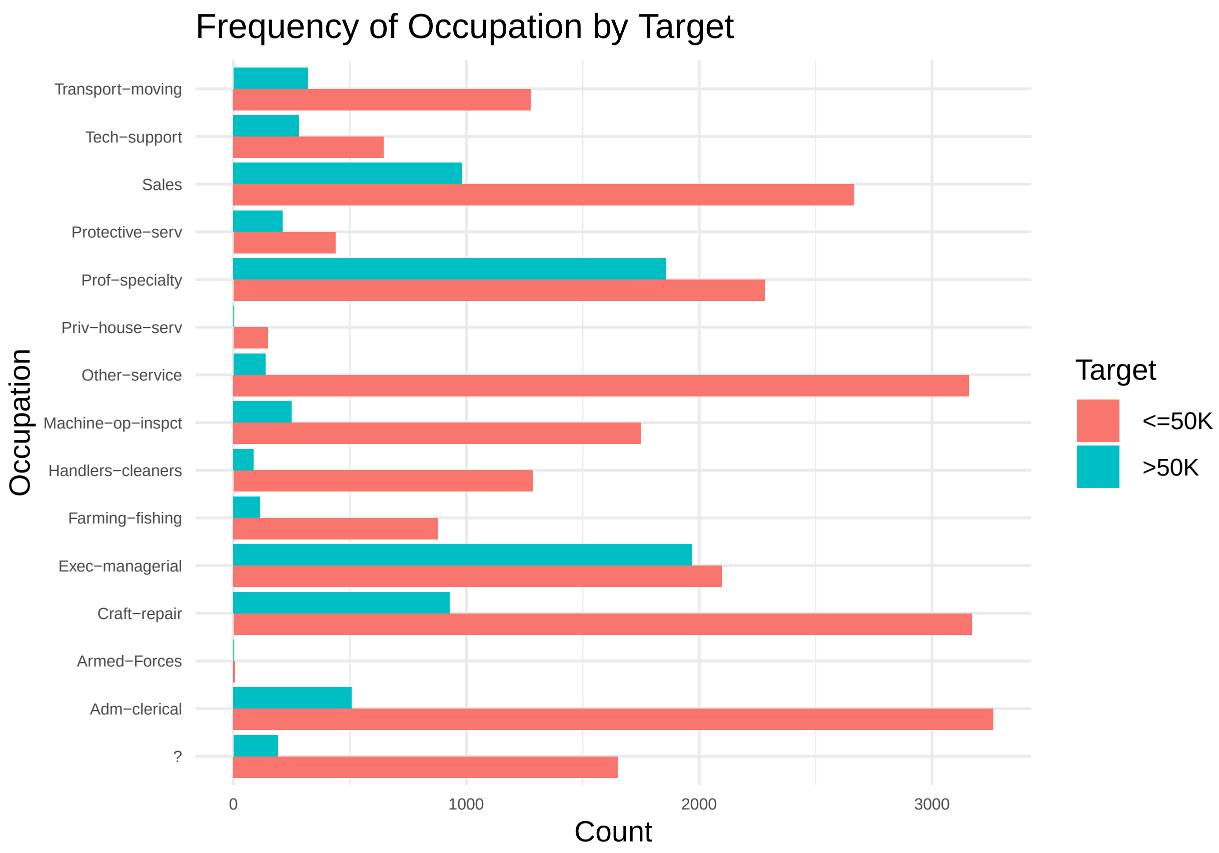


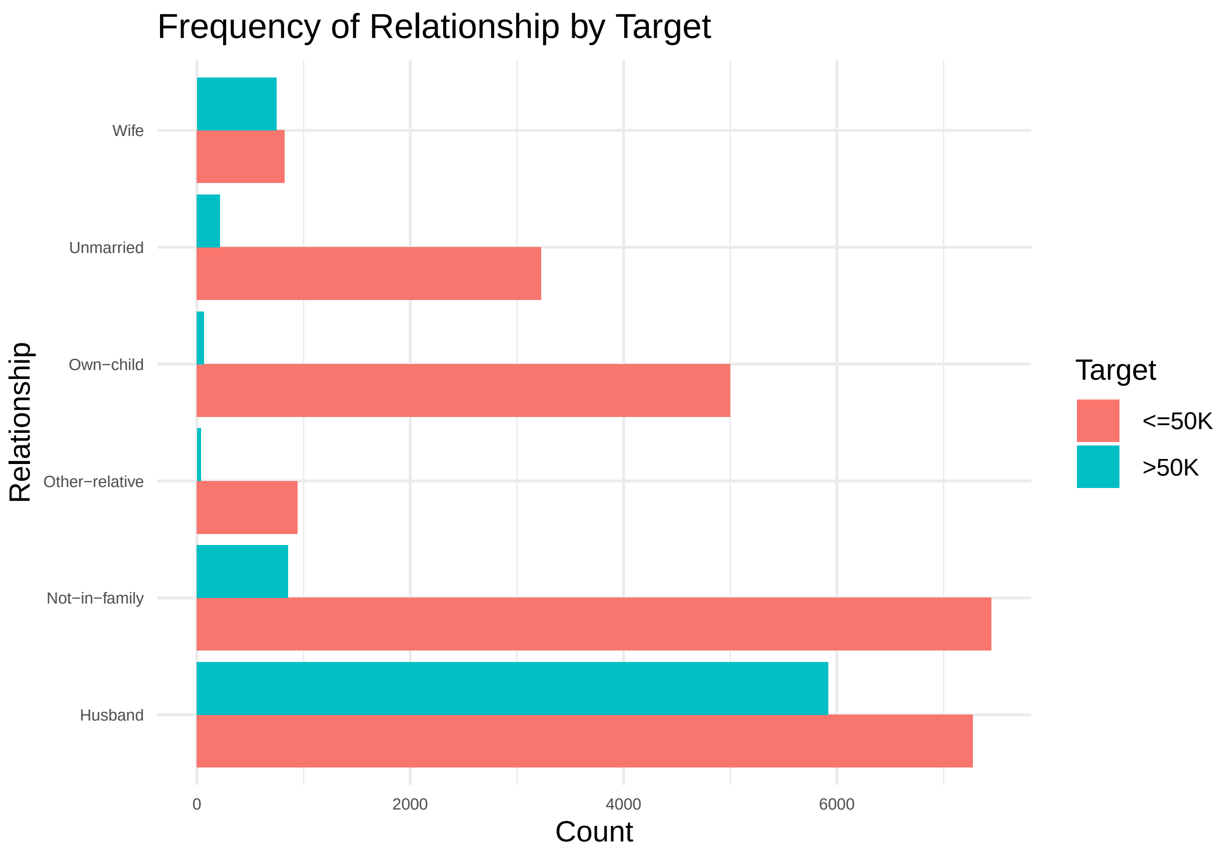
```r
for (i in categorical_cols) {
  plot <- ggplot(data2, aes(x = .data[[i]], fill = Target)) +
    geom_bar(position = "dodge") +
    labs(
      title = paste("Frequency of", i, "by Target"),
      x = i,
      y = "Count"
    ) +
    theme_minimal() +
    scale_x_discrete(
      labels = label_wrap(150)
    ) +
    coord_flip() + theme(
      axis.text = element_text(size = 6)
    )
  print(plot)
}
```

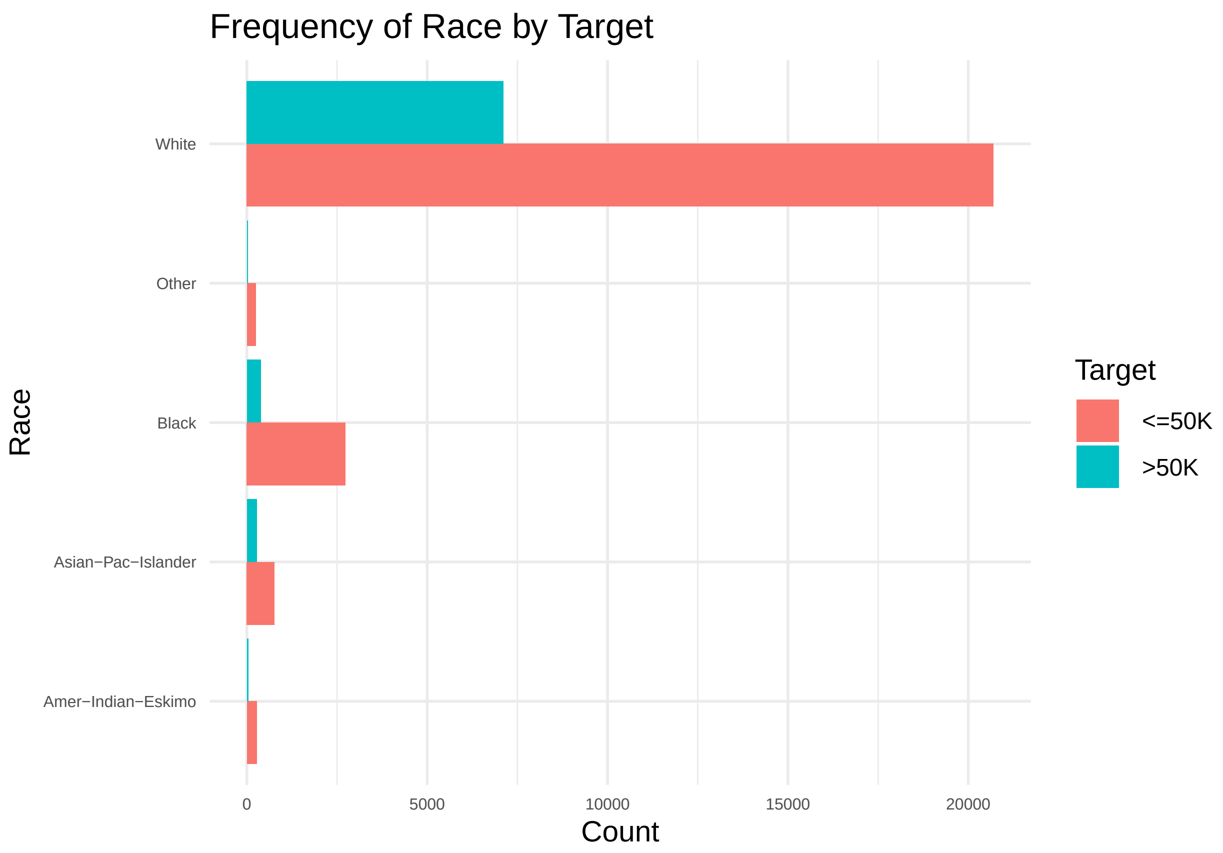

Frequency of Employment status by Target
Employment status
Without−pay
State−gov
Self−emp−not−inc
Self−emp−inc
Private
Never−worked
Local−gov
Federal−gov
?
0
5000
10000
15000
Count
Target
<=50K
>50K

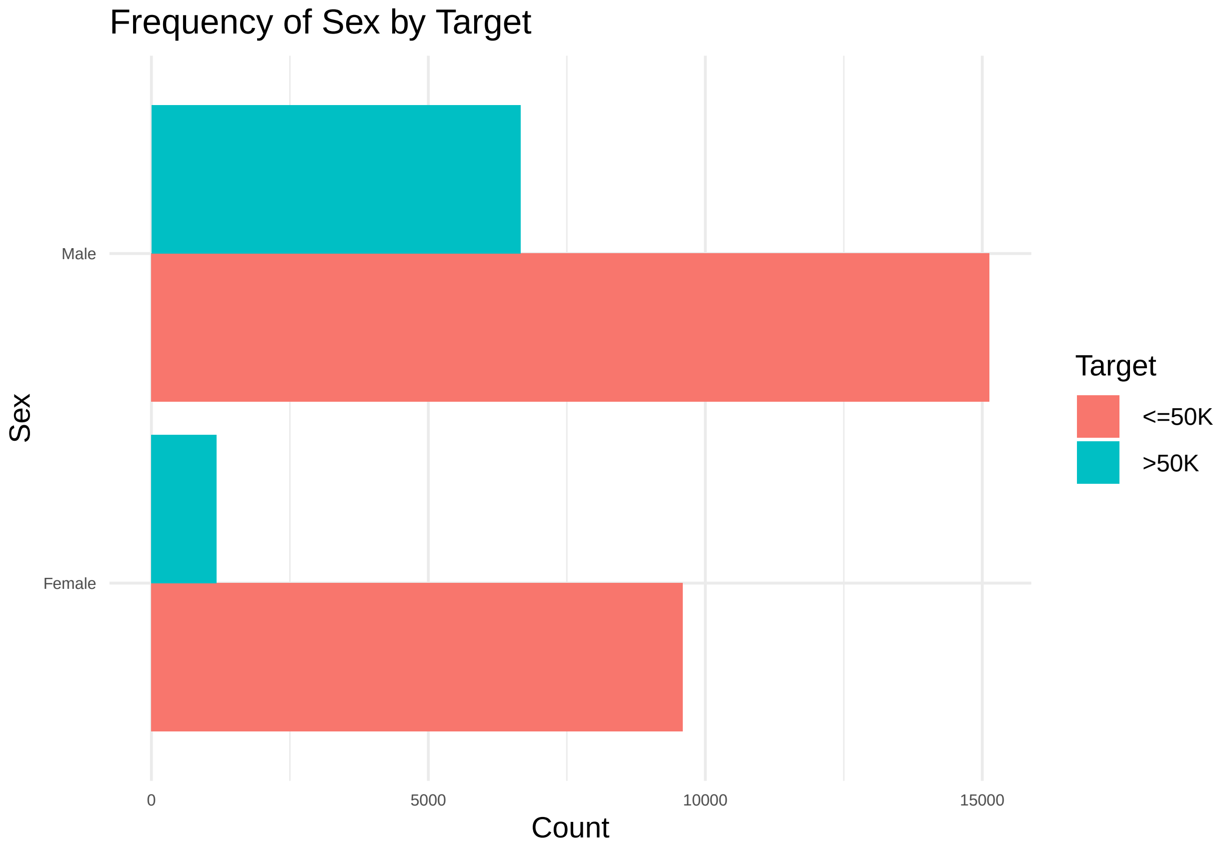

Frequency of Education by Target
Education
Some−college
Prof−school
Preschool
Masters
HS−grad
Doctorate
Bachelors
Assoc−voc
Assoc−acdm
9th
7th−8th
5th−6th
1st−4th
12th
11th
10th
0
2500
5000
7500
Count
Target
<=50K
>50K

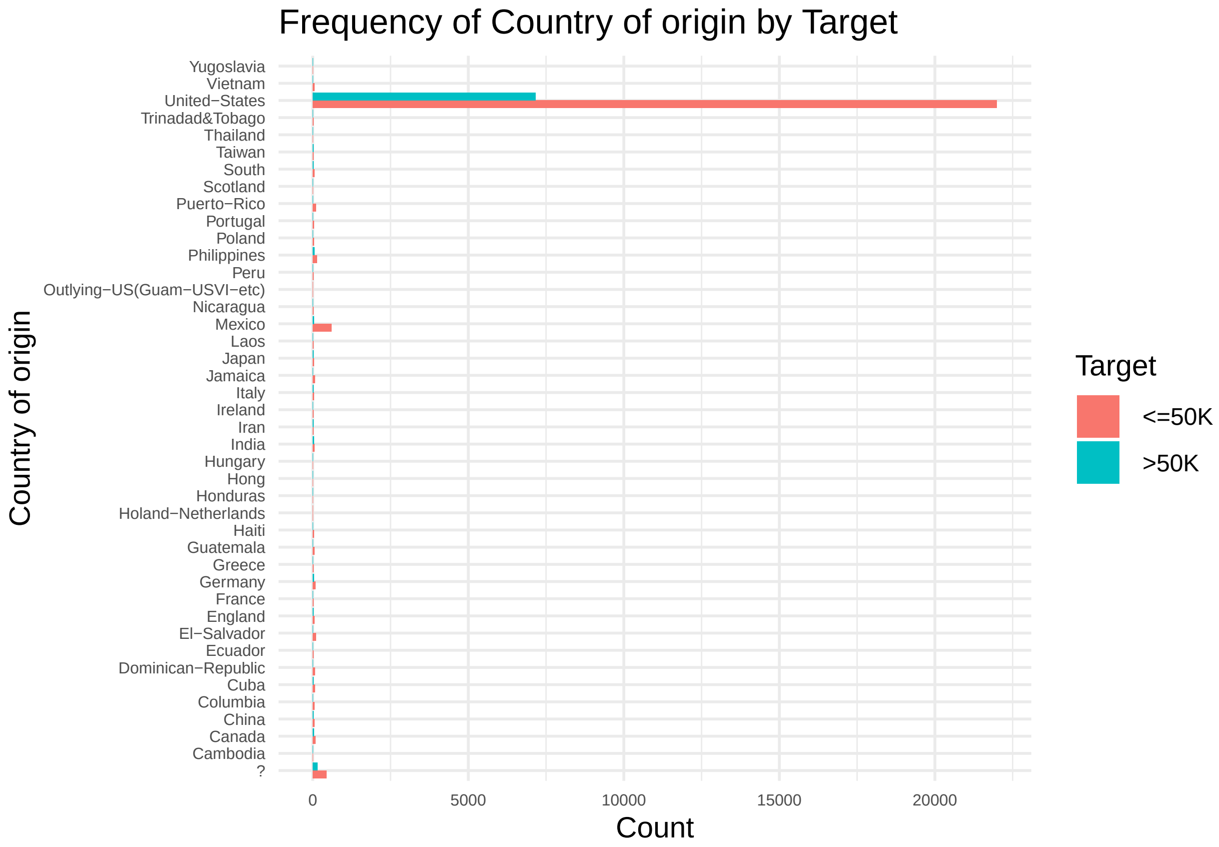

Frequency of Marital status by Target
Marital status
Widowed
Separated
Never−married
Married−spouse−absent
Married−civ−spouse
Married−AF−spouse
Divorced
0
2500
5000
7500
10000
Count
Target
<=50K
>50K


Frequency of Occupation by Target
Occupation
Transport−moving
Tech−support
Sales
Protective−serv
Prof−specialty
Priv−house−serv
Other−service
Machine−op−inspct
Handlers−cleaners
Farming−fishing
Exec−managerial
Craft−repair
Armed−Forces
Adm−clerical
?
0
1000
2000
3000
Count
Target
<=50K
>50K


Frequency of Relationship by Target
Relationship
Wife
Unmarried
Own−child
Other−relative
Not−in−family
Husband
0
2000
4000
6000
Count
Target
<=50K
>50K


Frequency of Race by Target
Race
White
Other
Black
Asian−Pac−Islander
Amer−Indian−Eskimo
0
5000
10000
15000
20000
Count
Target
<=50K
>50K

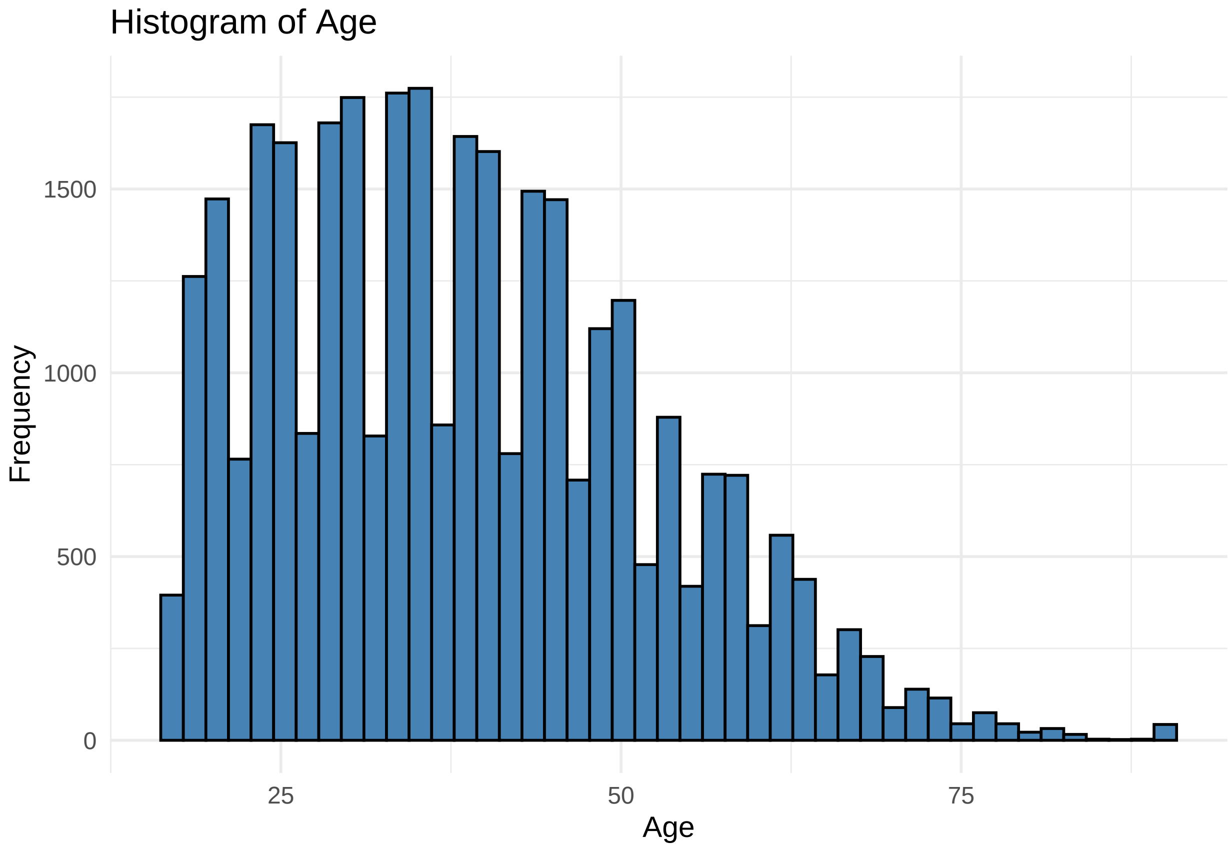

Frequency of Sex by Target
Sex
Male
Female
0
5000
10000
15000
Count
Target
🟥 <=50K
🟦 >50K

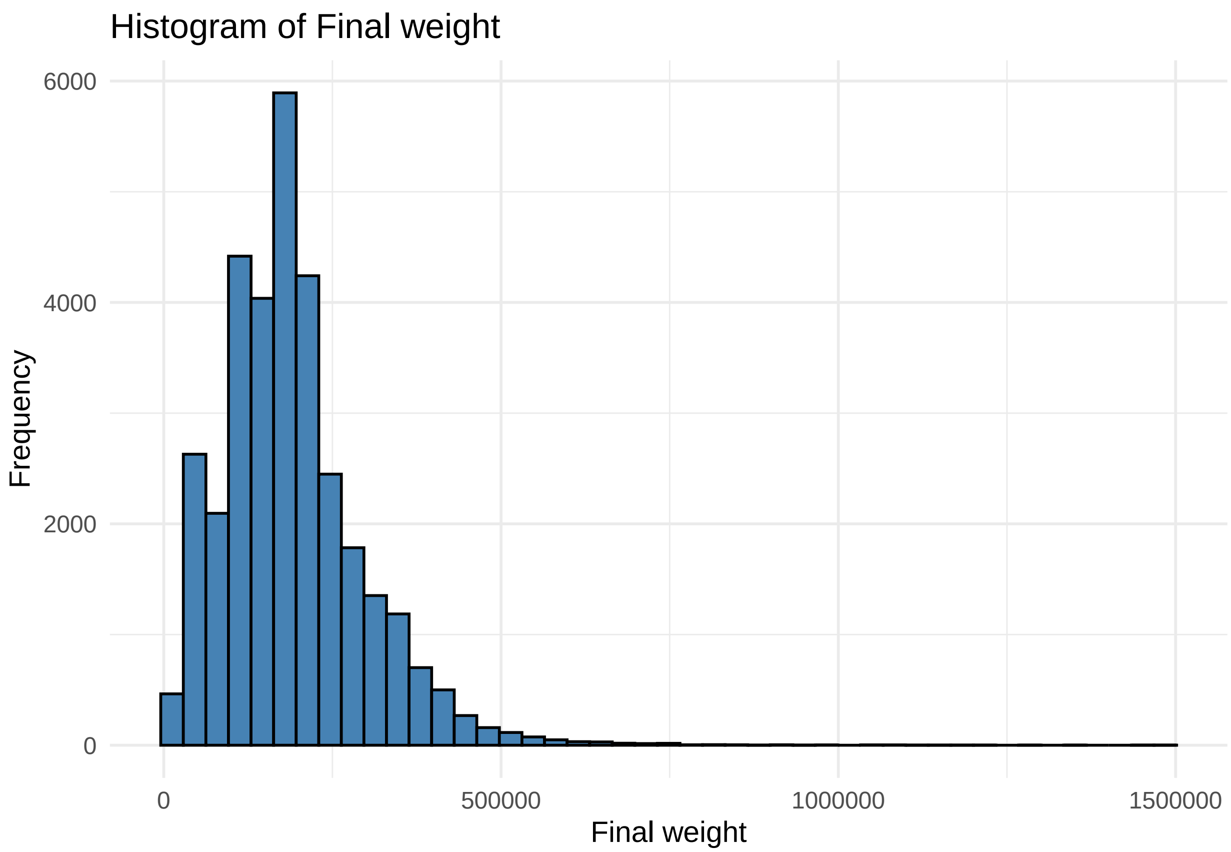


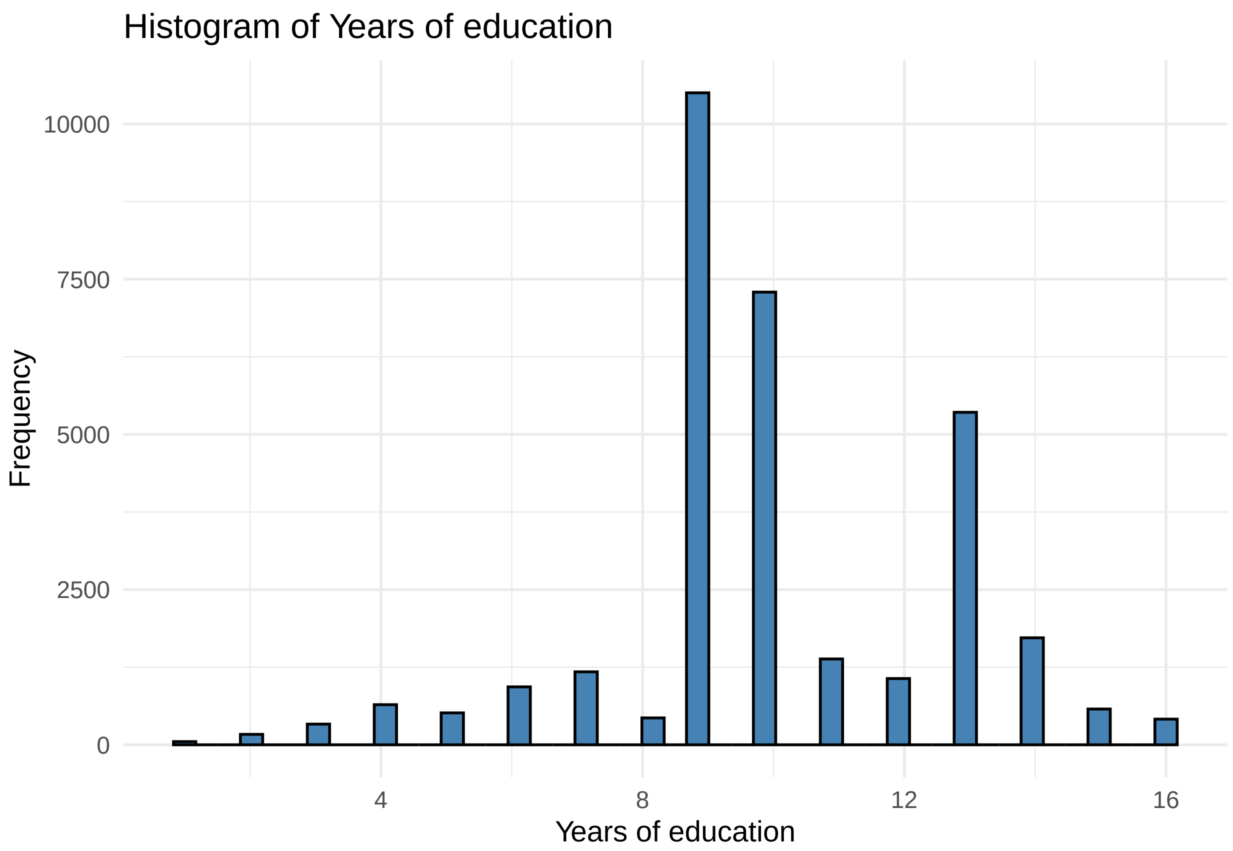
```r
numeric_cols <- names(data2)[sapply(data2, is.numeric)]
numeric_cols
```

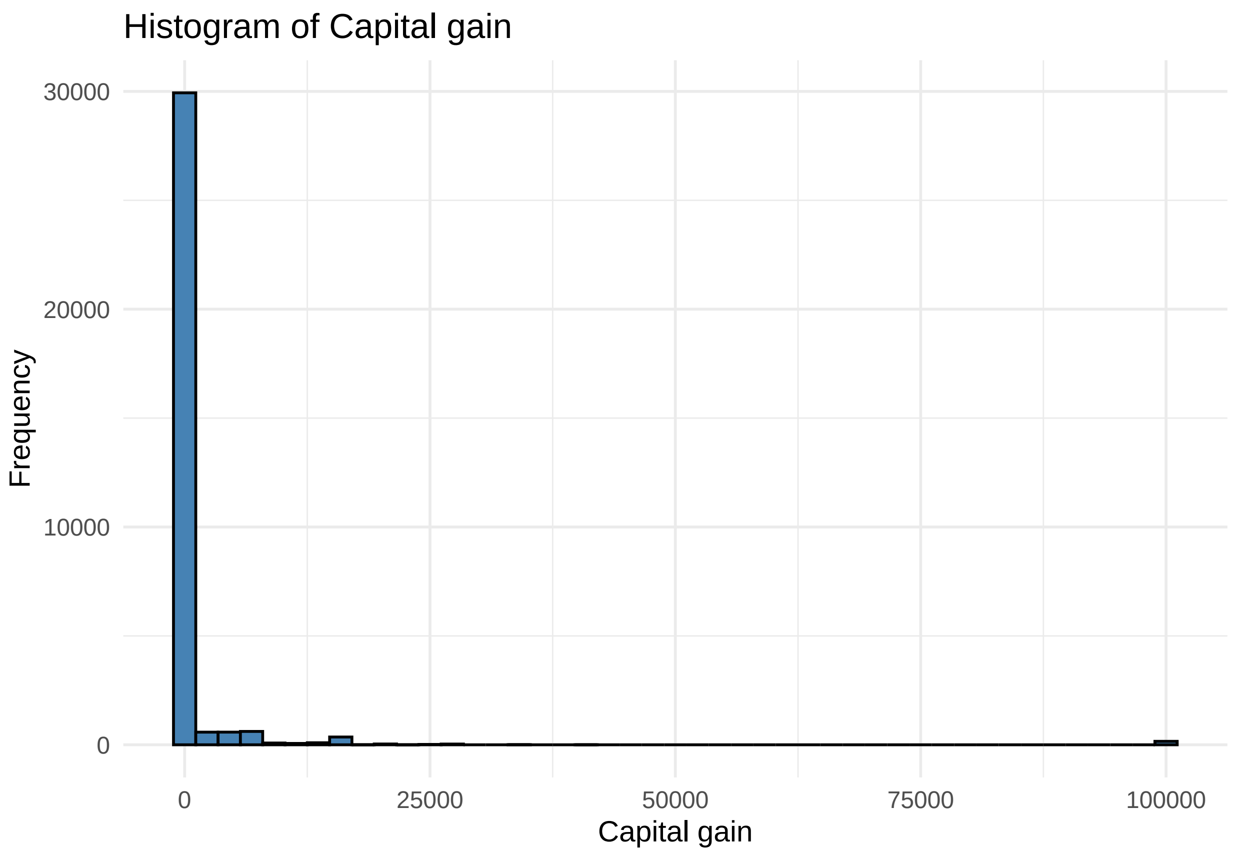
```
## [1] "Age"                "Final weight"       "Years of education"
## [4] "Capital gain"       "Capital loss"       "Hours worked per week"
```

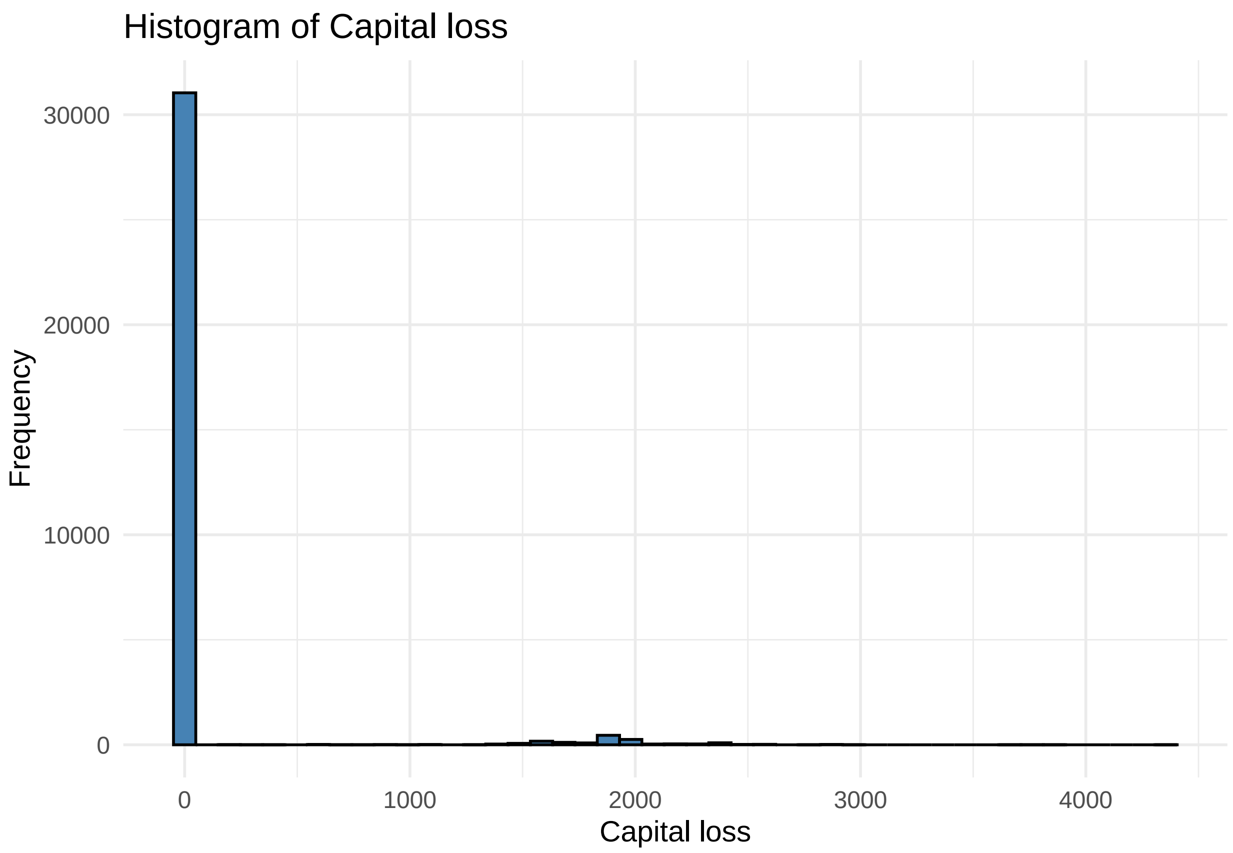
```r
for (i in numeric_cols) {
  plot <- ggplot(data2, aes(x = .data[[i]])) +
    geom_histogram(bins = 45, fill = "steelblue", color = "black") +
    labs(title = paste("Histogram of", i), x = i, y = "Frequency") +
    theme_minimal()
  print(plot)
}
```

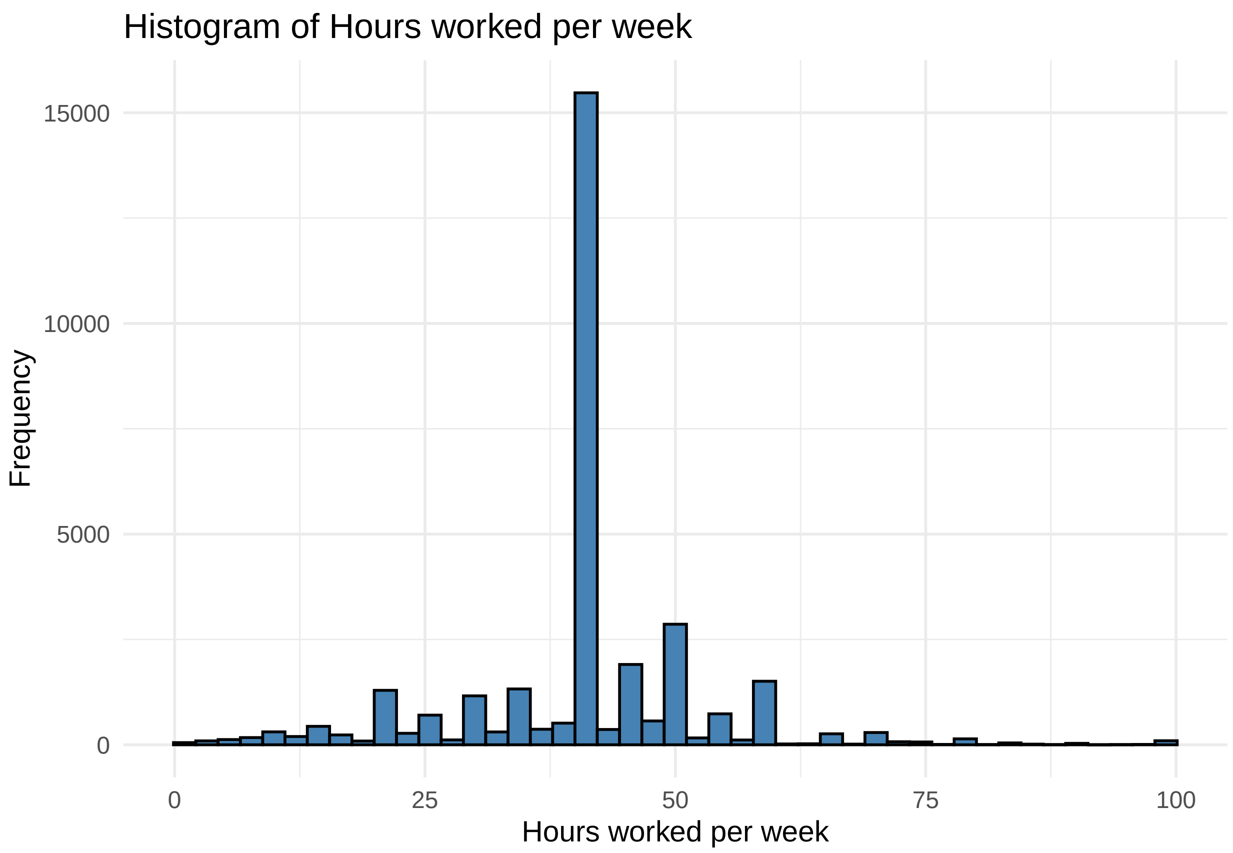

Histogram of Age
Frequency
0
500
1000
1500
25
50
75
Age


Histogram of Final weight
Frequency
6000
4000
2000
0
0
500000
1000000
1500000
Final weight

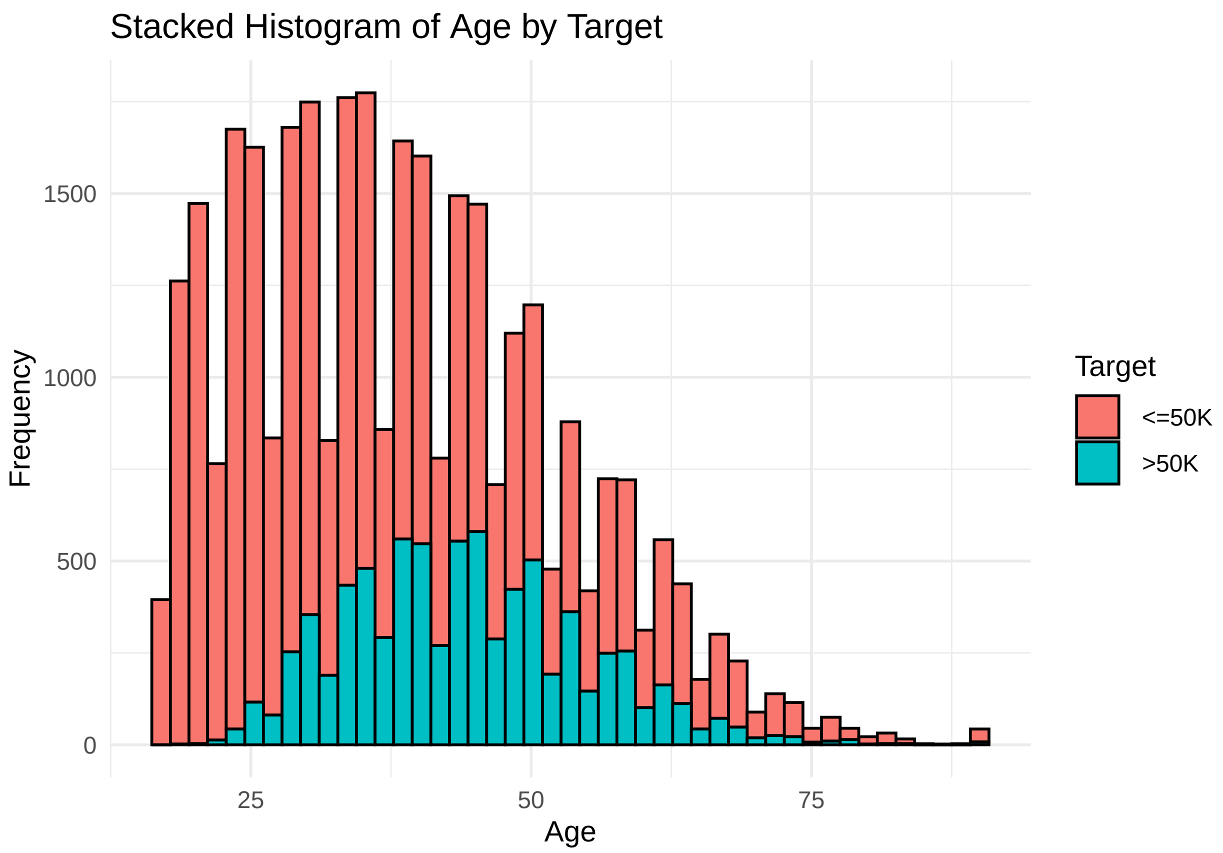

Histogram of Years of education
Frequency
10000
7500
5000
2500
0
4
8
12
16
Years of education

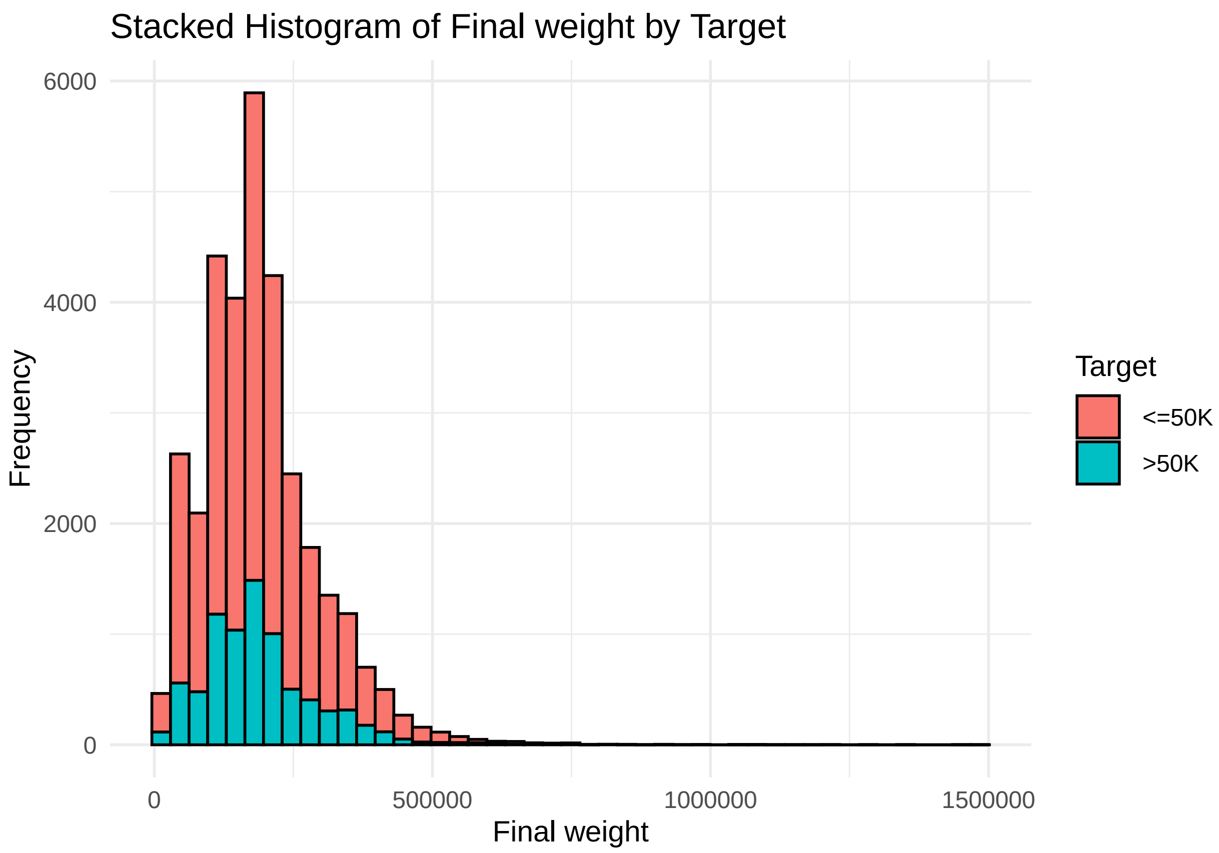

Histogram of Capital gain
Frequency
30000
20000
10000
0
0
25000
50000
75000
100000
Capital gain

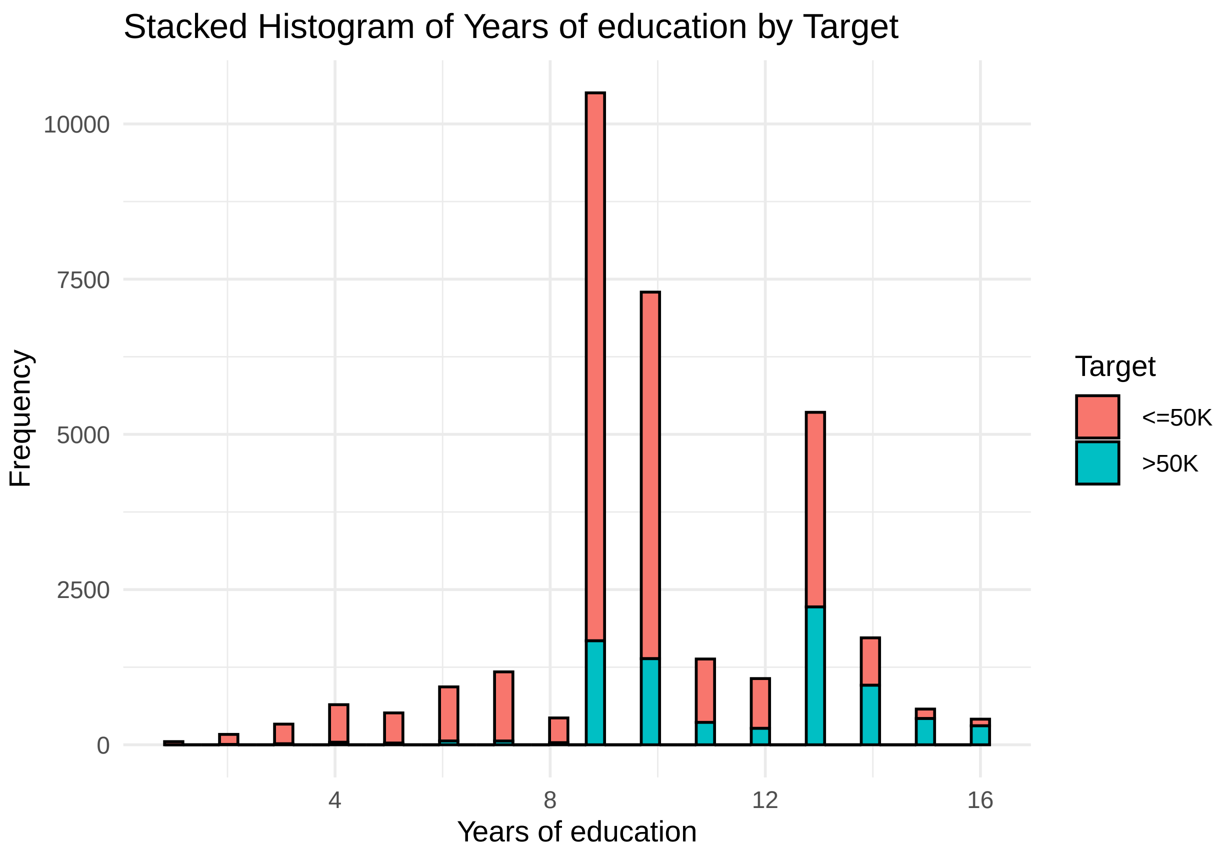

Histogram of Capital loss
30000
20000
10000
0
Frequency
0
1000
2000
3000
4000
Capital loss

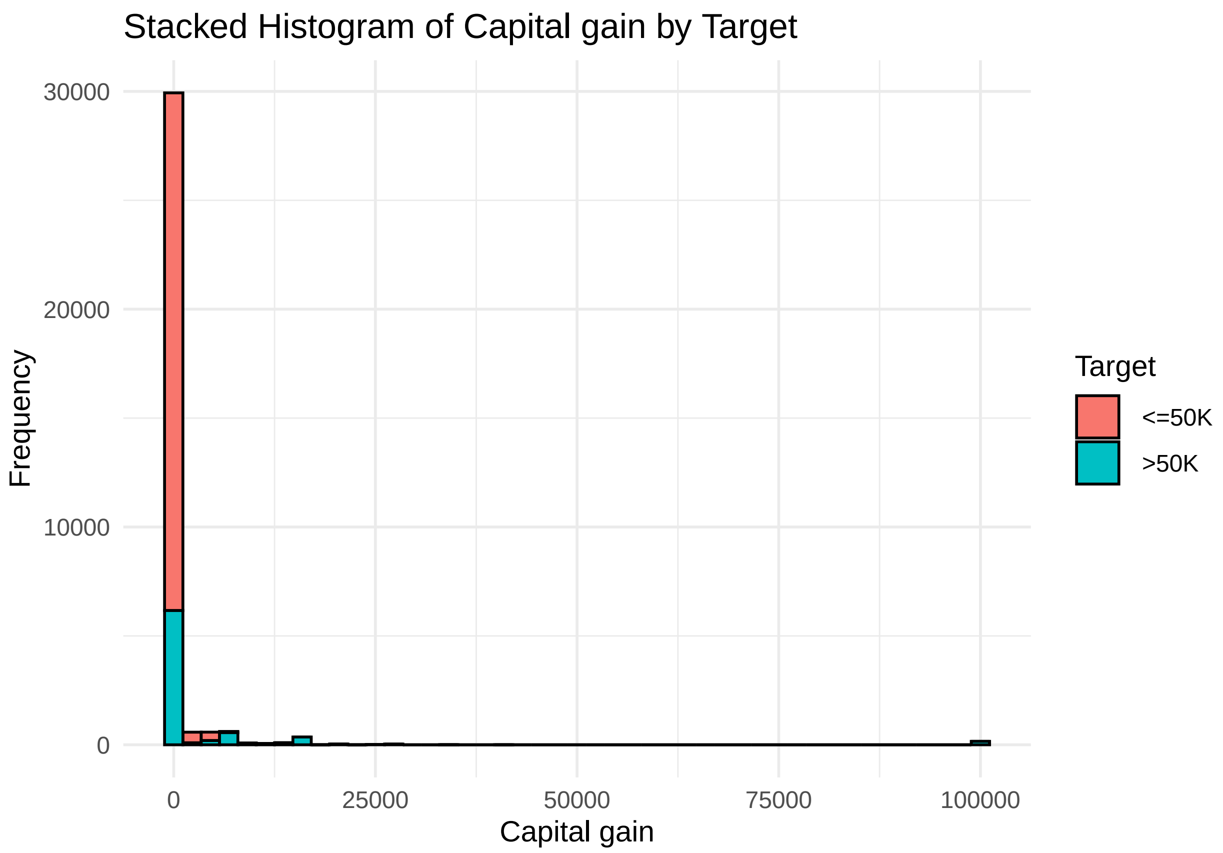


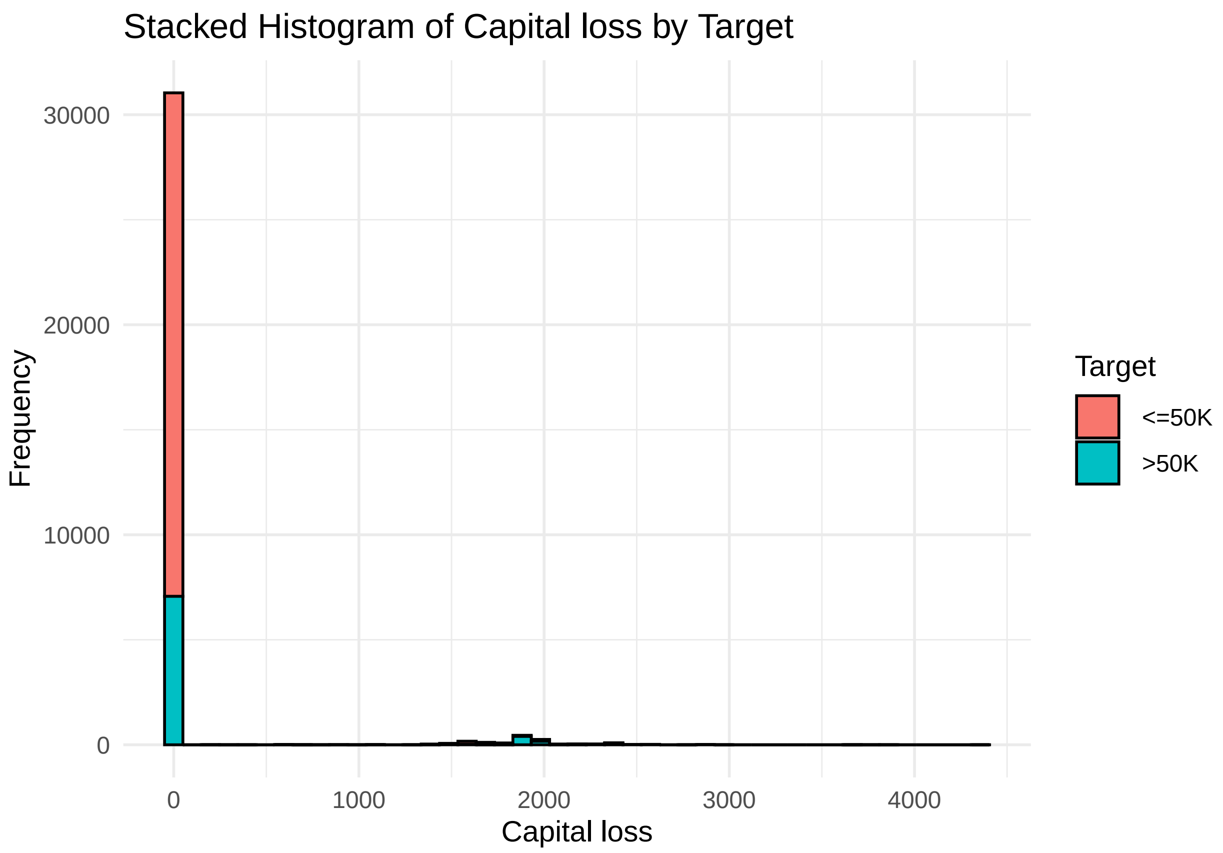
```
for (i in numeric_cols) {
  plot <- ggplot(data2, aes(x = .data[[i]], fill = Target)) +
    geom_histogram(bins = 45, color = "black") +
    labs(title = paste("Stacked Histogram of", i, "by Target"), x = i, y = "Frequency") +
    theme_minimal()
  print(plot)
}
```

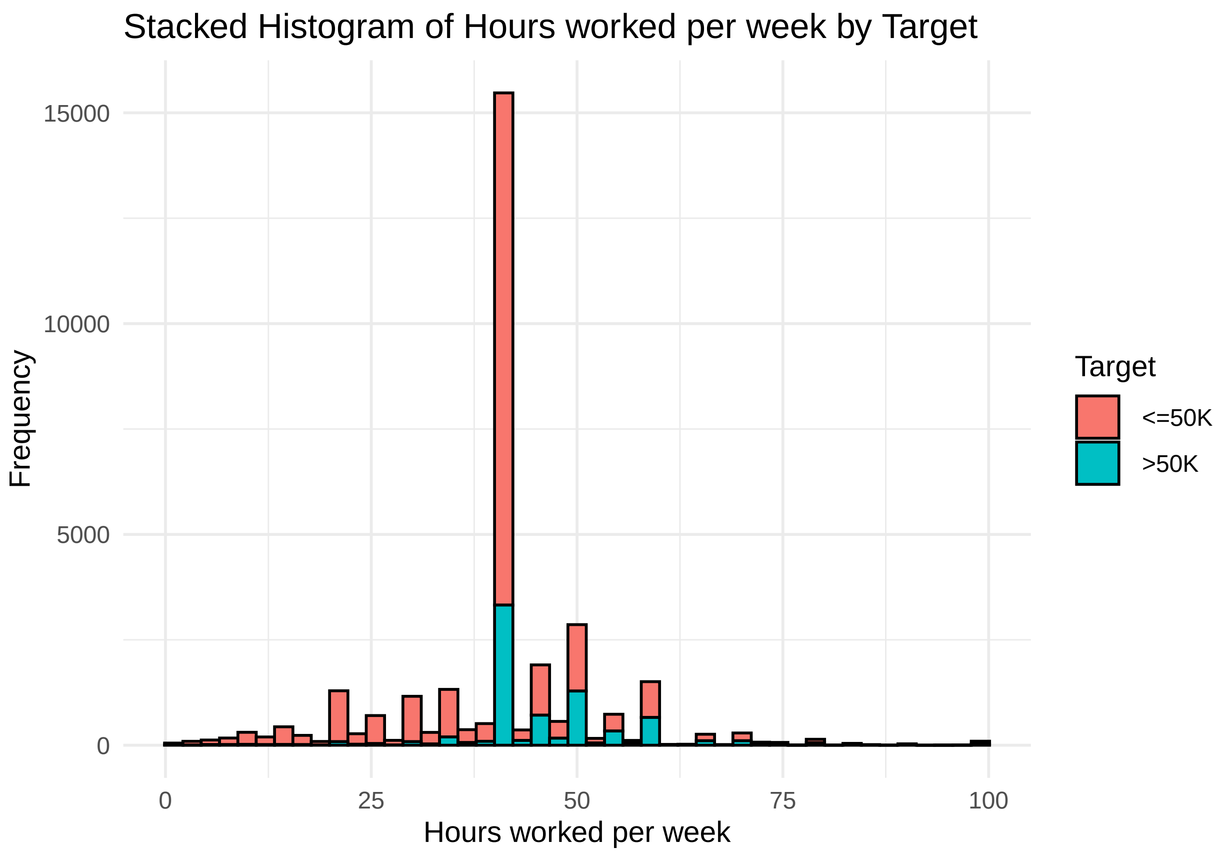

Stacked Histogram of Age by Target
Frequency
0
500
1000
1500
25
50
75
Age
Target
<=50K
>50K


Stacked Histogram of Final weight by Target
6000
4000
2000
0
Frequency
0
500000
1000000
1500000
Final weight
Target
<=50K
>50K

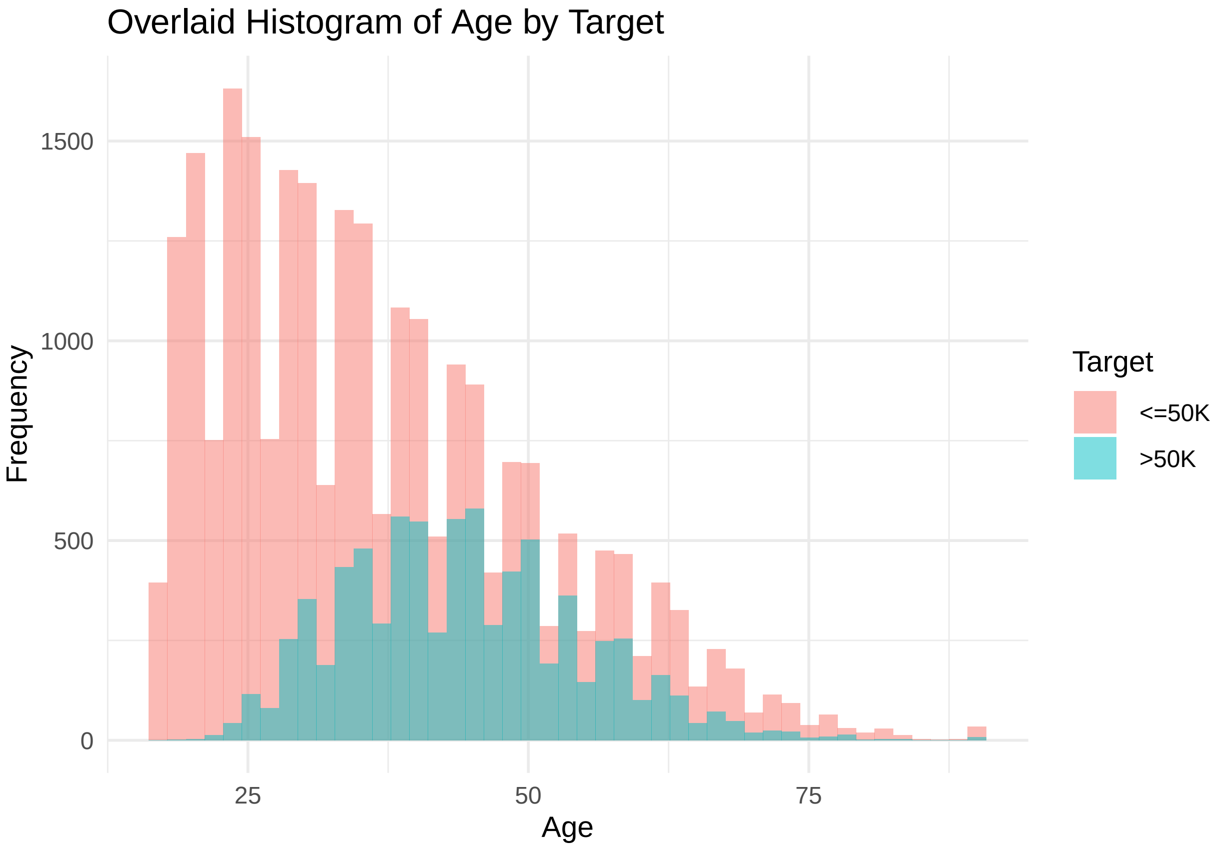

Stacked Histogram of Years of education by Target
Frequency
10000
7500
5000
2500
0
4
8
12
16
Years of education
Target
<=50K
>50K

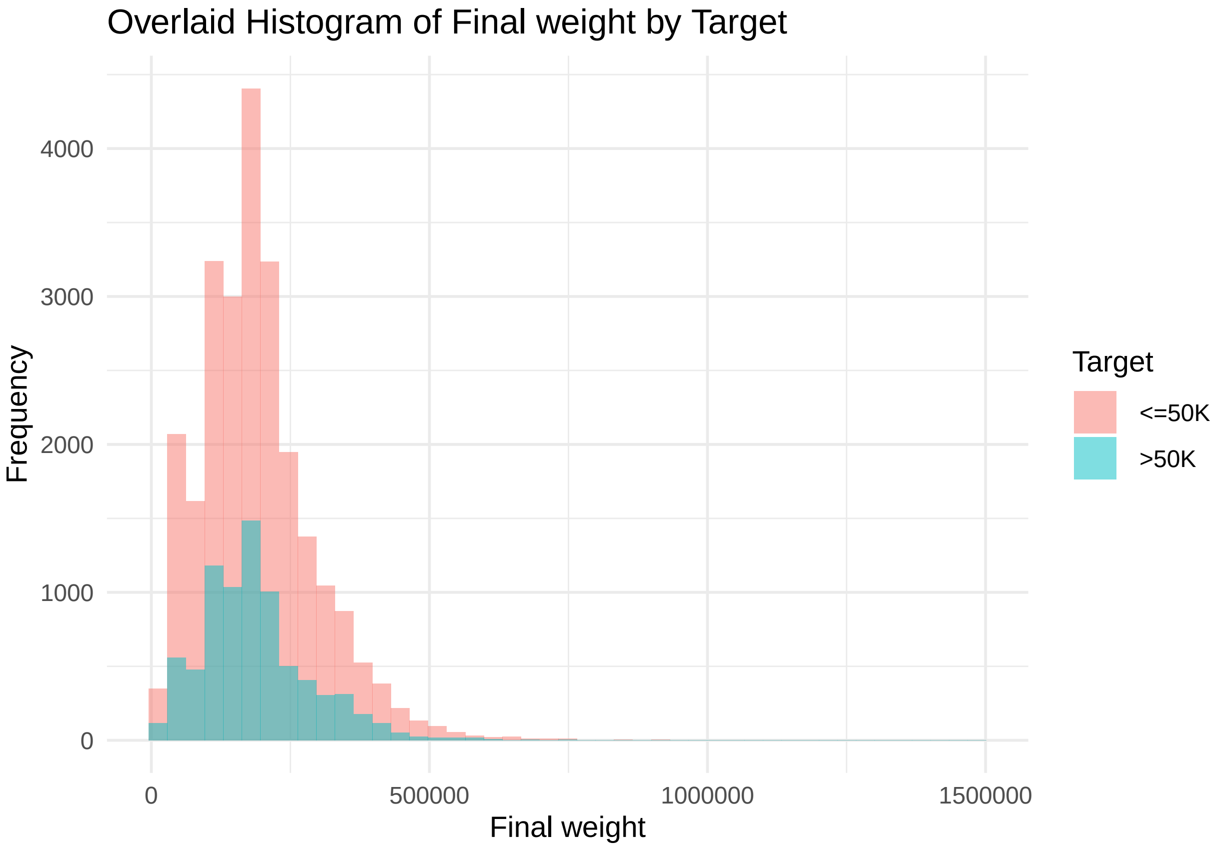

Stacked Histogram of Capital gain by Target
Frequency
30000
20000
10000
0
0
25000
50000
75000
100000
Capital gain
Target
<=50K
>50K

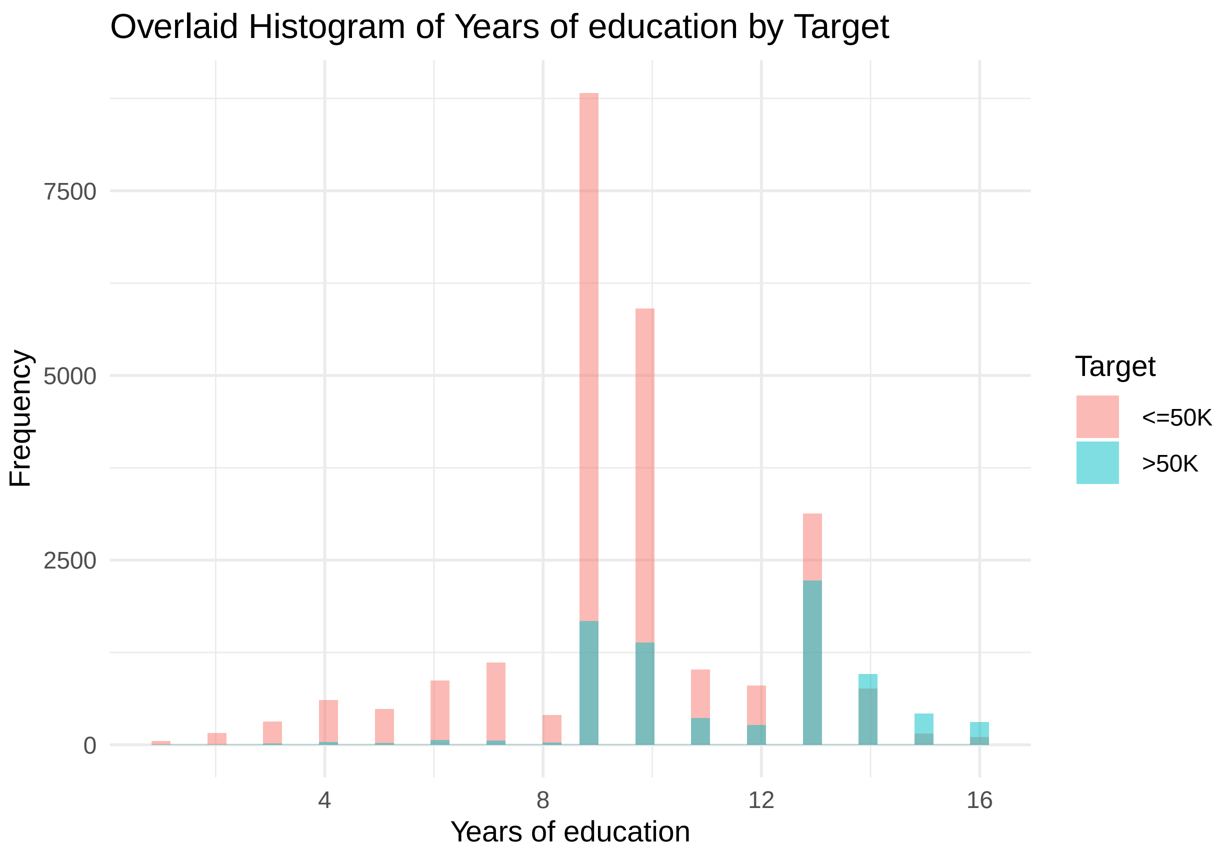

Stacked Histogram of Capital loss by Target
Frequency
30000
20000
10000
0
0
1000
2000
3000
4000
Capital loss
Target
<=50K
>50K

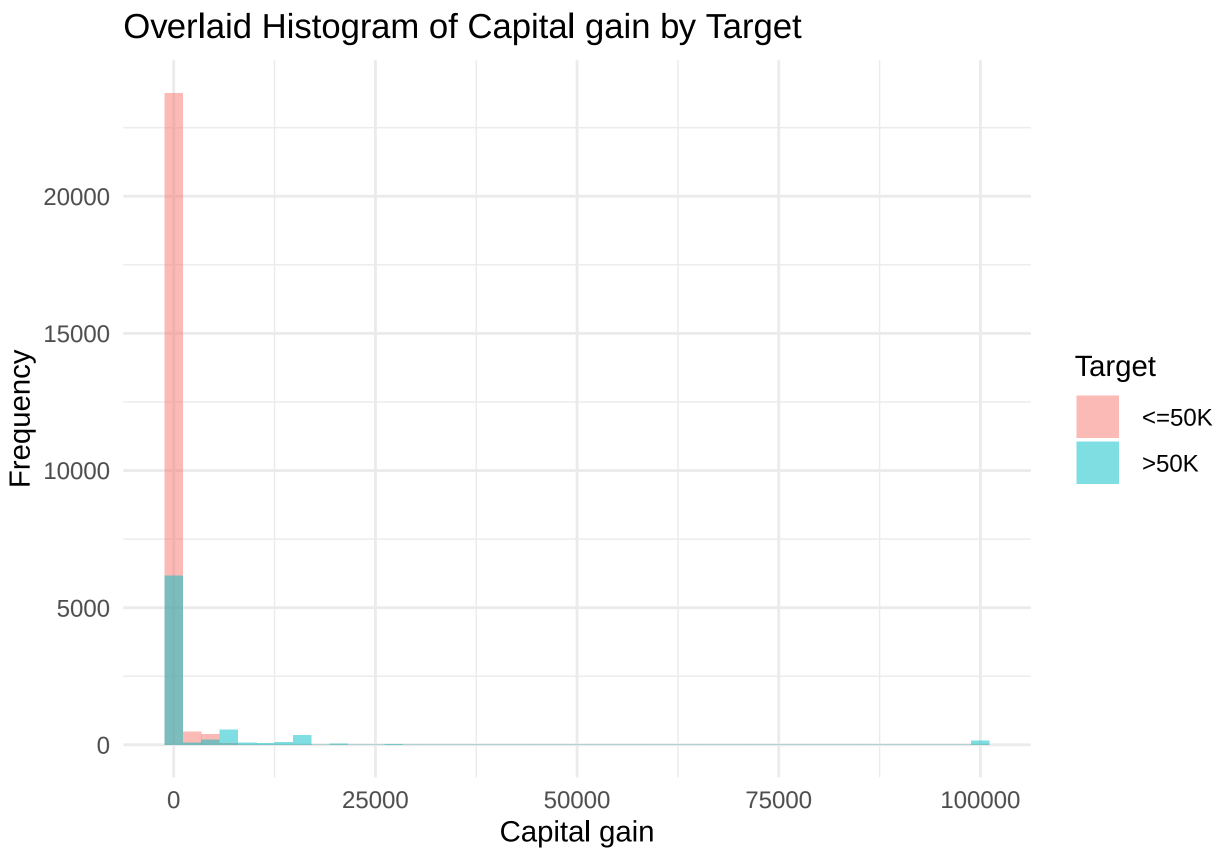


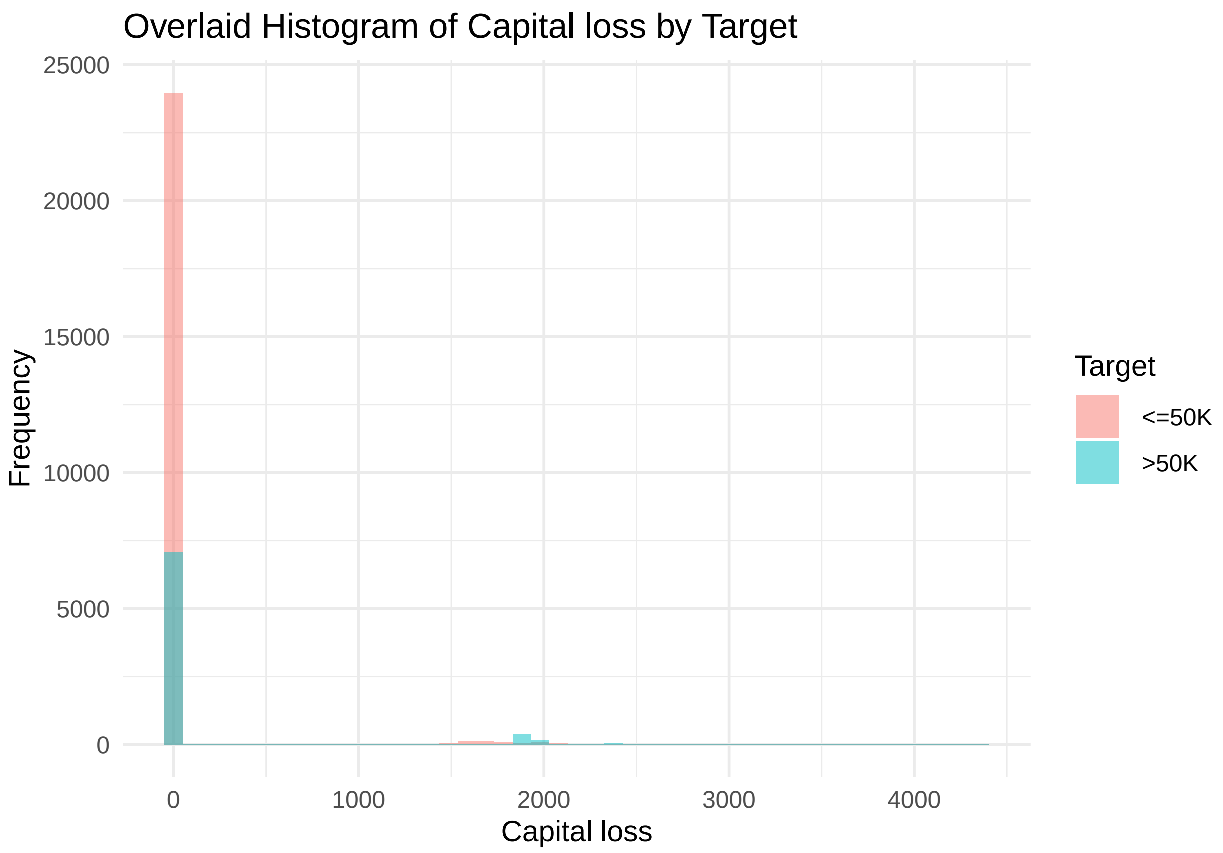
```
for (i in numeric_cols) {
  plot <- ggplot(data2, aes(x = .data[[i]], fill = Target)) +
    geom_histogram(bins = 45, position = "identity", alpha = 0.5) +
    labs(title = paste("Overlaid Histogram of", i, "by Target"), x = i, y = "Frequency") +
    theme_minimal()
  print(plot)
}
```

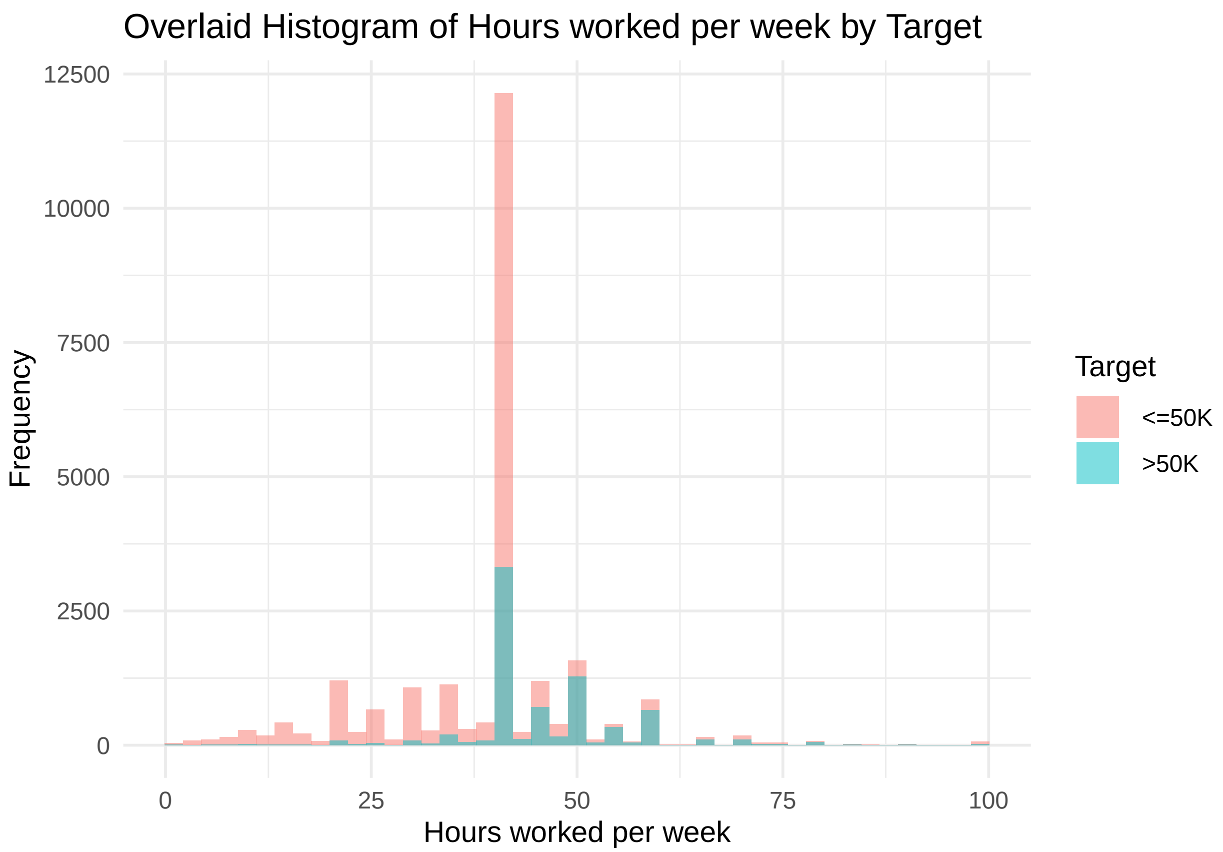

Overlaid Histogram of Age by Target
Frequency
1500
1000
500
0
25
50
75
Age
Target
<=50K
>50K


Overlaid Histogram of Final weight by Target
Frequency
4000
3000
2000
1000
0
0
500000
1000000
1500000
Final weight
Target
<=50K
>50K

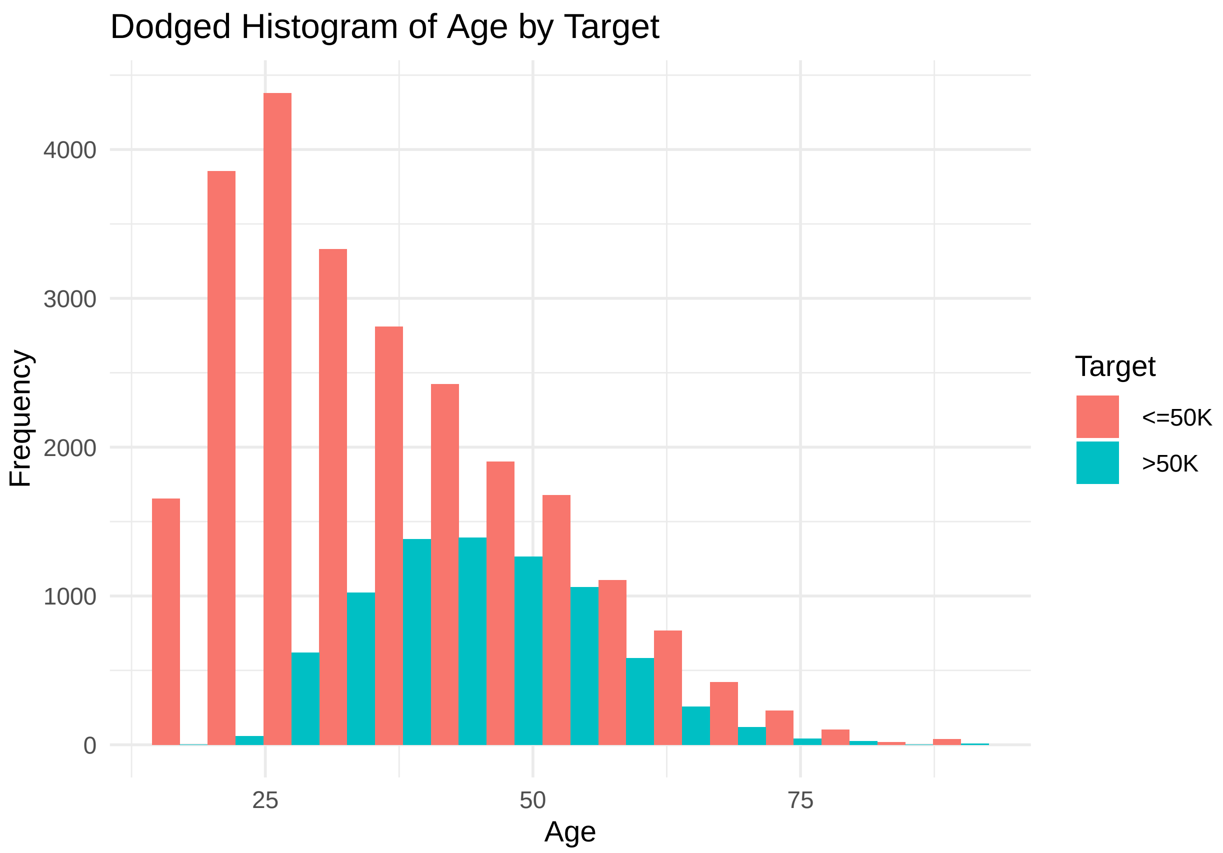

Overlaid Histogram of Years of education by Target
Frequency
7500
5000
2500
0
4
8
12
16
Years of education
Target
<=50K
>50K

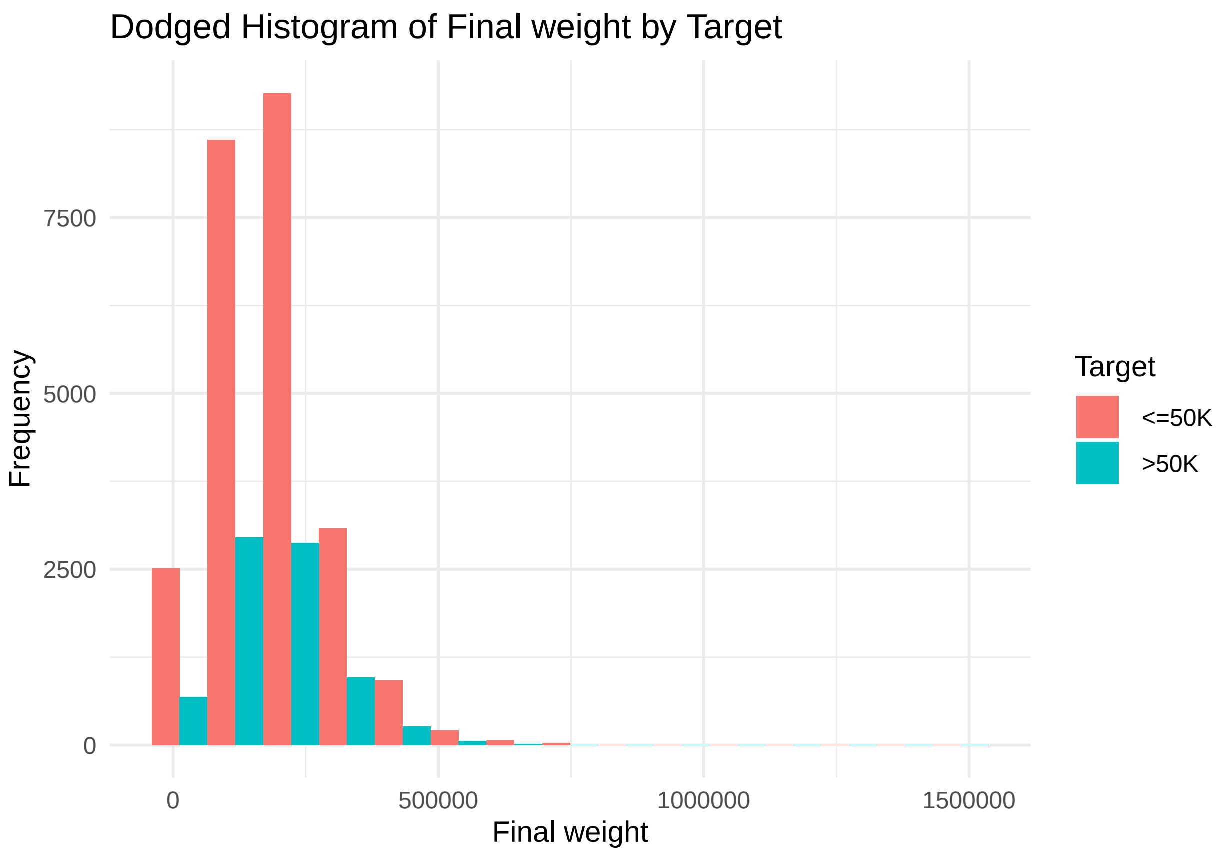

Overlaid Histogram of Capital gain by Target
Frequency
20000
15000
10000
5000
0
0
25000
50000
75000
100000
Capital gain
Target
<=50K
>50K

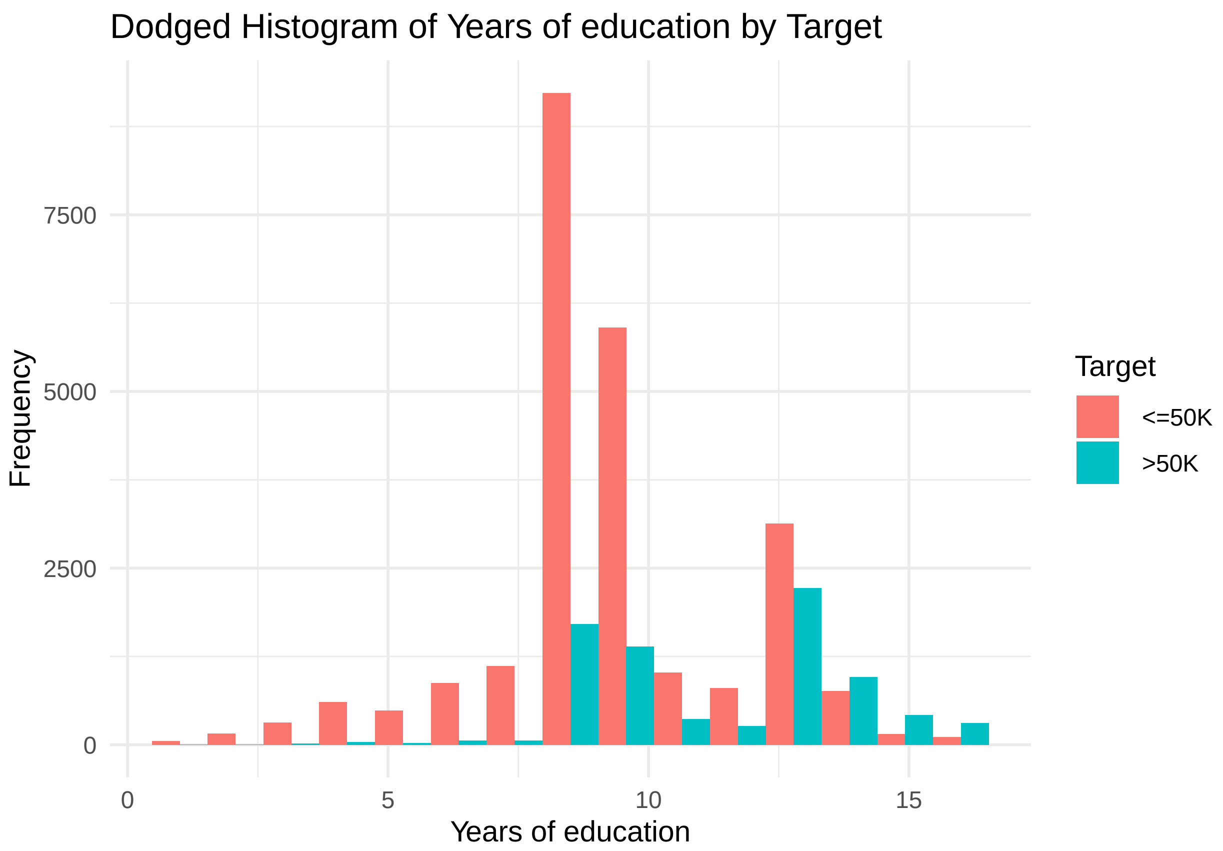

Overlaid Histogram of Capital loss by Target
Frequency
25000
20000
15000
10000
5000
0
0
1000
2000
3000
4000
Capital loss
Target
<=50K
>50K

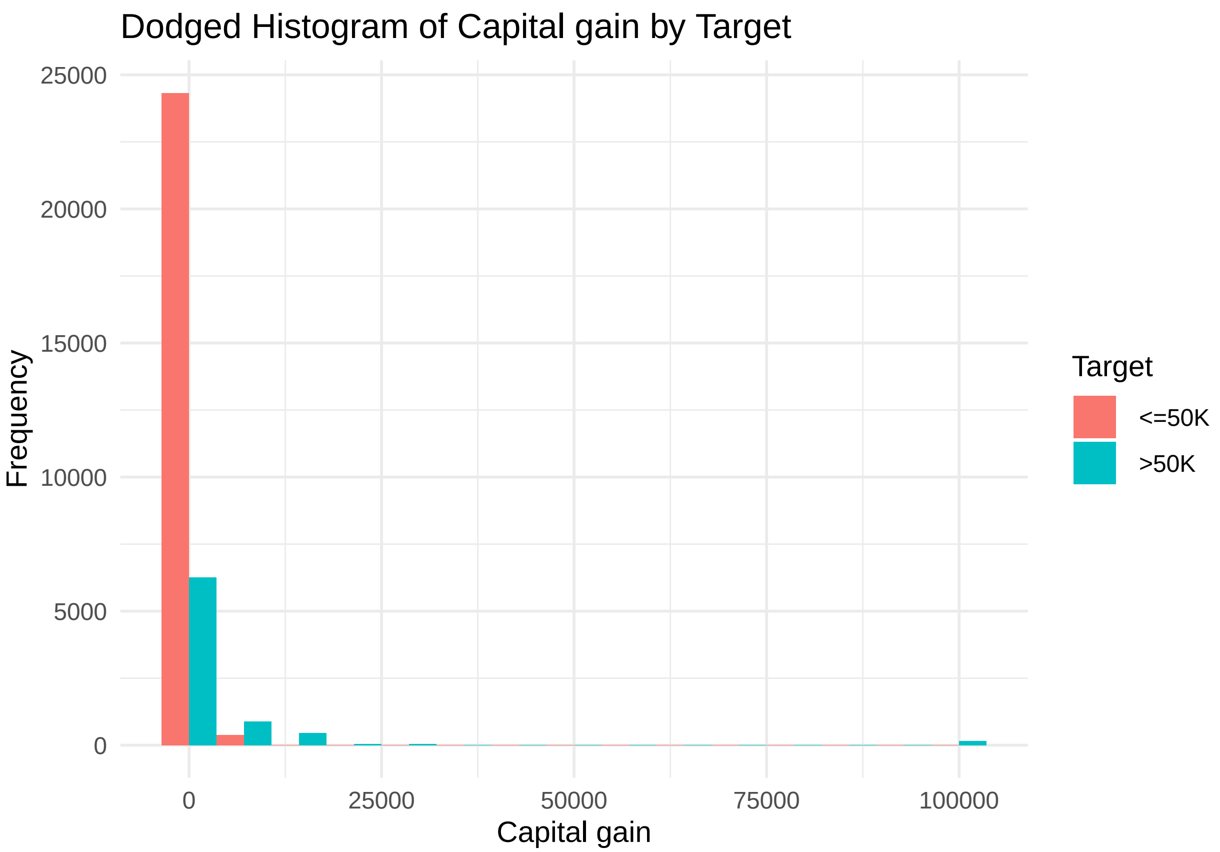


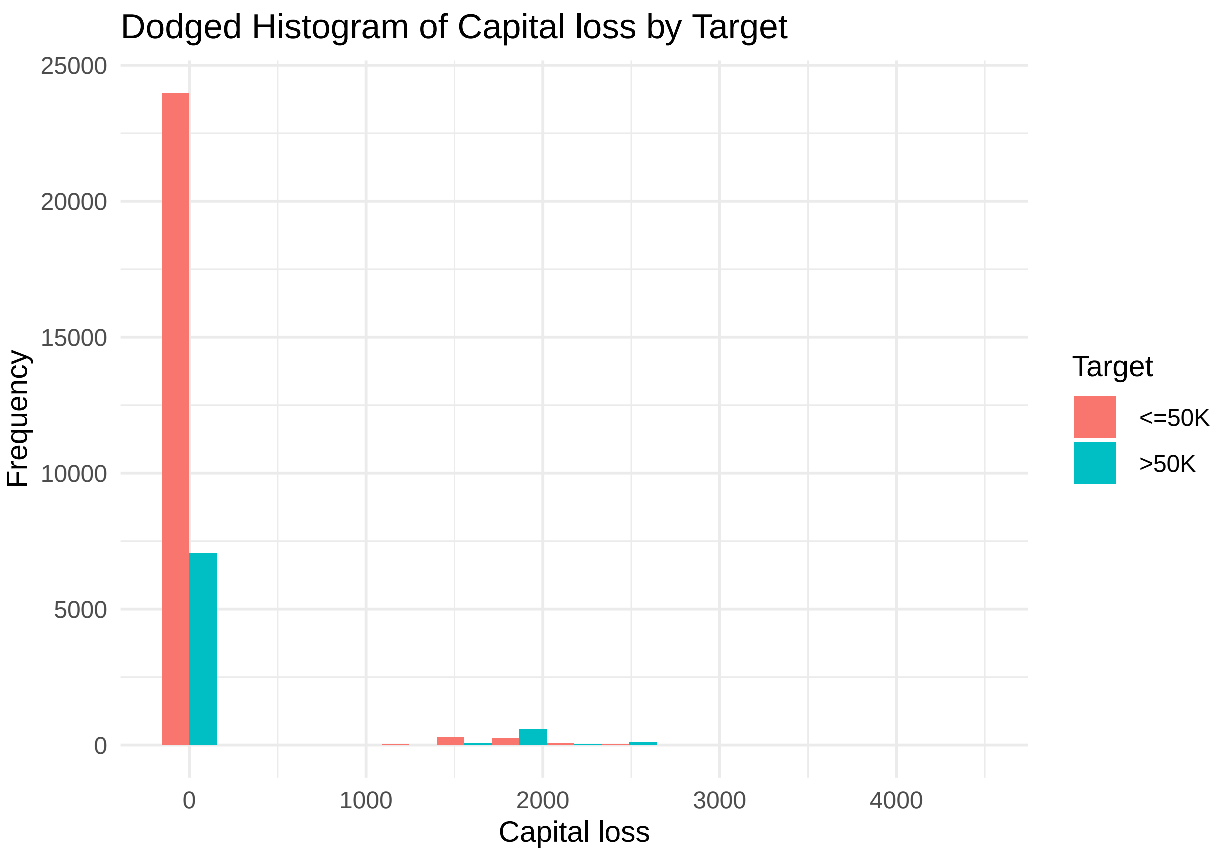
```
for (i in numeric_cols) {
  plot <- ggplot(data2, aes(x = .data[[i]], fill = Target)) +
    geom_histogram(bins = 15, position = "dodge") +
    labs(title = paste("Dodged Histogram of", i, "by Target"), x = i, y = "Frequency") +
    theme_minimal()
  print(plot)
}
```

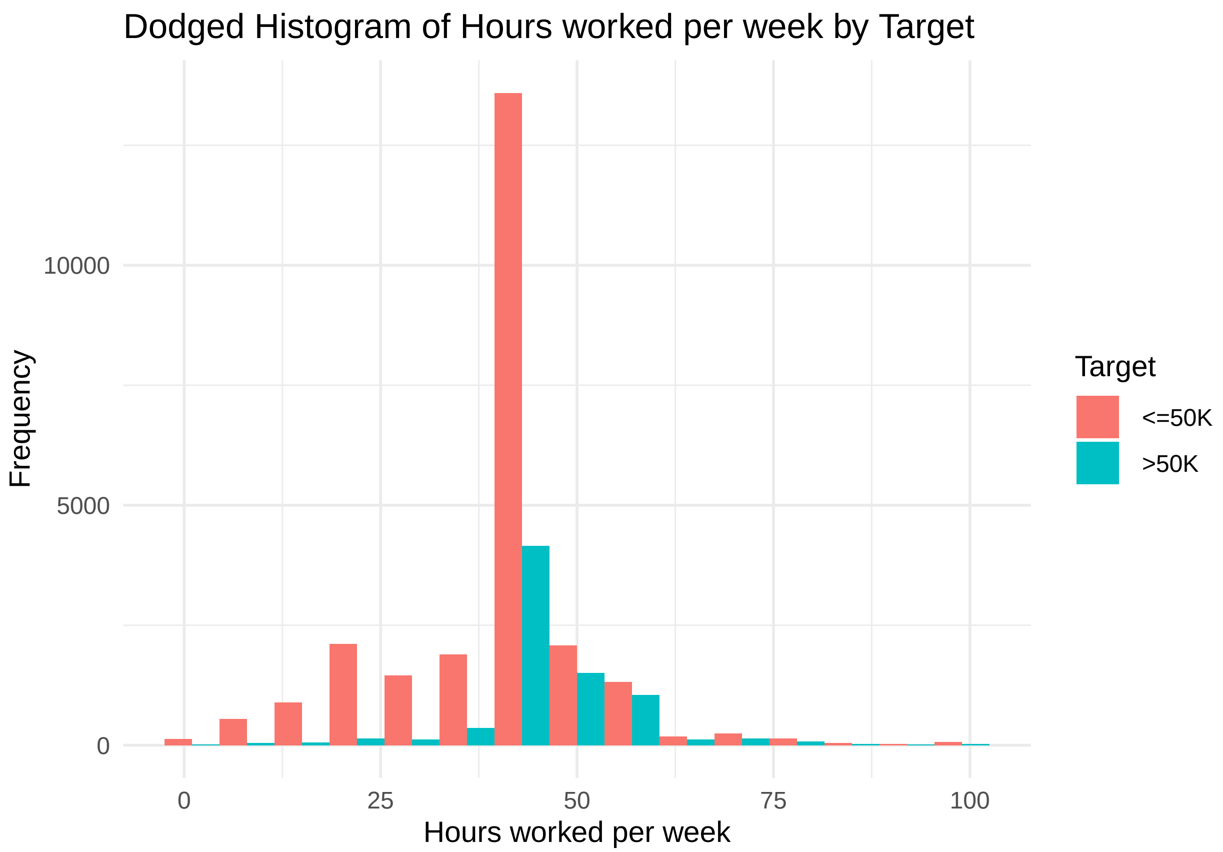

Dodged Histogram of Age by Target
Frequency
4000
3000
2000
1000
0
25
50
75
Age
Target
<=50K
>50K


Dodged Histogram of Final weight by Target
Frequency
7500
5000
2500
0
0
500000
1000000
1500000
Final weight
Target
<=50K
>50K


Dodged Histogram of Years of education by Target
Frequency
7500
5000
2500
0
0
5
10
15
Years of education
Target
<=50K
>50K


Dodged Histogram of Capital gain by Target
Frequency
25000
20000
15000
10000
5000
0
0
25000
50000
75000
100000
Capital gain
Target
<=50K
>50K


Dodged Histogram of Capital loss by Target
Frequency
25000
20000
15000
10000
5000
0
0
1000
2000
3000
4000
Capital loss
Target
<=50K
>50K



Notes: - employment status is mostly private but has variation - race is mostly white but has variation - country of origin is mostly US - 2/3 male, 1/3 female - hours worked per week is mostly 40, with some variation - capital loss and gain are each mostly zero - age may be truncated - age is skewed to the right for <=50K - hours worked per week: <=50K has a larger spread

Notable variables: - relationship - occupation - marital status