AirlineAnalysis.R

puj83

Mon Jul 29 20:11:13 2019

```
# Airline X

# Guest Satisfaction Survey Data

# Business Problem

# Given data set of 8,000 guest satisfaction survey records to determine the
key drivers for likelihood to recommend Airline X.
# The dependent variable in the data set is Q1.
# key drivers you found, how you derived them (data used, manipulations/trans
formations if you used any, models run, sampling, etc.),
# and what information/data, if necessary, you'd like to make a recommendatio
n on key drivers.

GuestData<-read.csv(file = "C:/Users/puj83/OneDrive/CV/Cases/AirlineX/Guest_S
atisfaction_Survey_Data.csv", header = T, sep = ",")
# saveRDS(GuestData, file = "GuestData.rds")

# GuestData<-load(file = "GuestData.rds")

names(GuestData)
```

```
##  [1] "Q1"                  "Q132"                "Q94"
##  [4] "Q91_1"               "Q91_2"               "Q91_3"
##  [7] "Q135"                "Q115"                "Q95"
## [10] "Q53_1"               "Q53_2"               "Q53_3"
## [13] "Q53_4"               "Q55"                 "Q119"
## [16] "Q116"                "Q131"                "Q109"
## [19] "Q96"                 "Q100"                "Q102_1"
## [22] "Q102_2"              "Q102_3"              "Q102_4"
## [25] "Q104"                "Q134"                "Q138"
## [28] "Q130"                "Q106_1"              "Q106_2"
## [31] "Q106_3"              "Q126"                "Q108"
## [34] "Q110"                "Q112"                "Q118"
## [37] "Q141"                "Q129"                "Q114_1"
## [40] "Q114_2"              "Q114_3"              "Q116.1"
## [43] "Q117"                "Q137"                "Q120"
## [46] "Q35"                 "Q37_1"               "Q37_2"
## [49] "Q37_3"               "Q37_4"               "Q37_5"
## [52] "Q111"                "Q113"                "Q112.1"
## [55] "Q8"                  "Q9"                  "Q11"
## [58] "Q12"                 "Q107"                "Q105"
```

```
## [61] "Q106"                "Q40"                 "Q15"
## [64] "Q18"                 "Q20"                 "Q22"
## [67] "Q24"                 "Q28_1"               "Q34"
## [70] "Q36_1"               "Q36_2"               "Q3"
## [73] "Q5"                  "Q6_1"                "Q6_2"
## [76] "Q6_3"                "Q6_4"                "Q7"
## [79] "Q60"                 "Q114"                "Q107.1"
## [82] "Q108.1"              "Q48"                 "Q50"
## [85] "Q52"                 "Q54_1"               "Q54_2"
## [88] "Q54_3"               "Q54_4"               "Q54_5"
## [91] "Q145"                "Q99"                 "Q127"
## [94] "Q100.1"              "Booking.Channel"     "Flight.Date.Time"
## [97] "Pax.per.PNR"         "Segment1.Destination" "Segment1.Origin"
```

```
library(dplyr)
```

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

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

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

```
# install.packages("tibble")
# install.packages("pdflatex")

library(rlang)
```

```
## Warning: package 'rlang' was built under R version 3.5.3
```

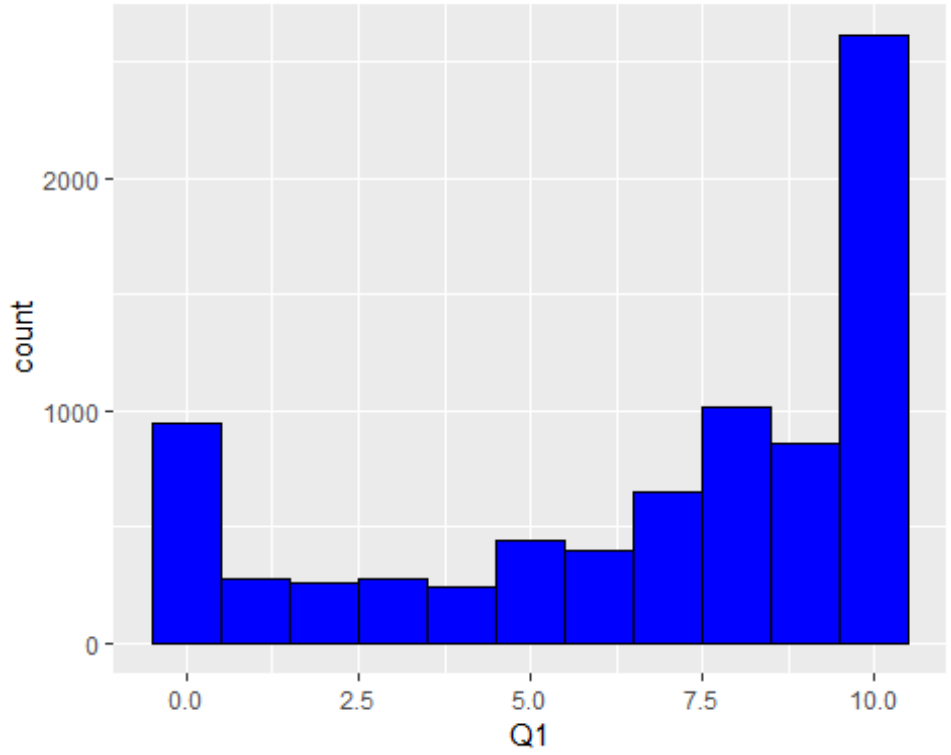
```
library(tibble)
```

```
## Warning: package 'tibble' was built under R version 3.5.3
```

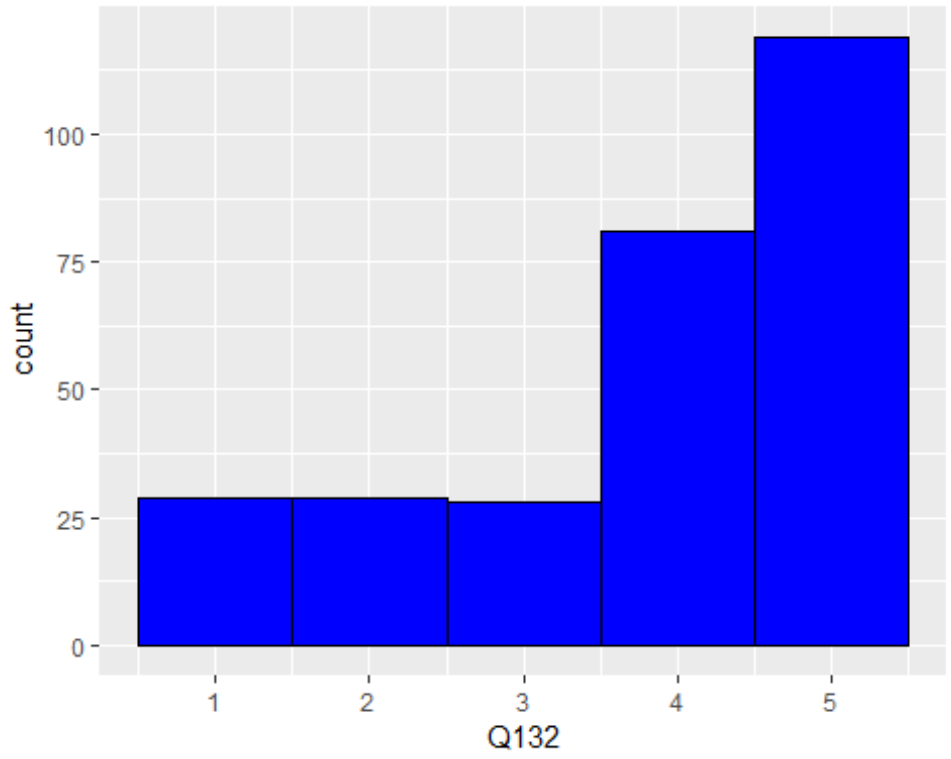
```
library(ggplot2)
# library(pdflatex)

GuestData<-GuestData %>% mutate_all(as.factor)

# Convert all columns to factors

# Transform all 'ordinal' columns into 'numerical columns'

# Treating missing values:

# Missing observations will be ignored and analysis is done on the variables
present.
```

```r
# Look at all levels of the variables.

GuestData2<-GuestData %>% sapply(levels)

library(plyr)

## ------------------------------------------------------------------------

## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, th
en dplyr:
## library(plyr); library(dplyr)

## ------------------------------------------------------------------------

##
## Attaching package: 'plyr'

## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize

# [1] "Q1"                  "Q132"                "Q94"                 "Q
91_1"                "Q91_2"               "Q91_3"               "Q135"
"Q115"                "Q95"
# [10] "Q53_1"               "Q53_2"               "Q53_3"               "
Q53_4"               "Q55"                 "Q119"                "Q116"
"Q131"                "Q109"
# [19] "Q96"                 "Q100"                "Q102_1"              "
Q102_2"              "Q102_3"              "Q102_4"              "Q104"
"Q134"                "Q138"
# [28] "Q130"                "Q106_1"              "Q106_2"              "
Q106_3"              "Q126"                "Q108"                "Q110"
"Q112"                "Q118"
# [37] "Q141"                "Q129"                "Q114_1"              "
Q114_2"              "Q114_3"              "Q116.1"              "Q117"
"Q137"                "Q120"
# [46] "Q35"                 "Q37_1"               "Q37_2"               "
Q37_3"               "Q37_4"               "Q37_5"               "Q111"
"Q113"                "Q112.1"
# [55] "Q8"                  "Q9"                  "Q11"                 "
Q12"                 "Q107"                "Q105"                "Q106"
"Q40"                 "Q15"
# [64] "Q18"                 "Q20"                 "Q22"                 "
Q24"                 "Q28_1"               "Q34"                 "Q36_1"
"Q36_2"               "Q3"
# [73] "Q5"                  "Q6_1"                "Q6_2"                "
Q6_3"                "Q6_4"                "Q7"                  "Q60"
"Q114"                "Q107.1"
```

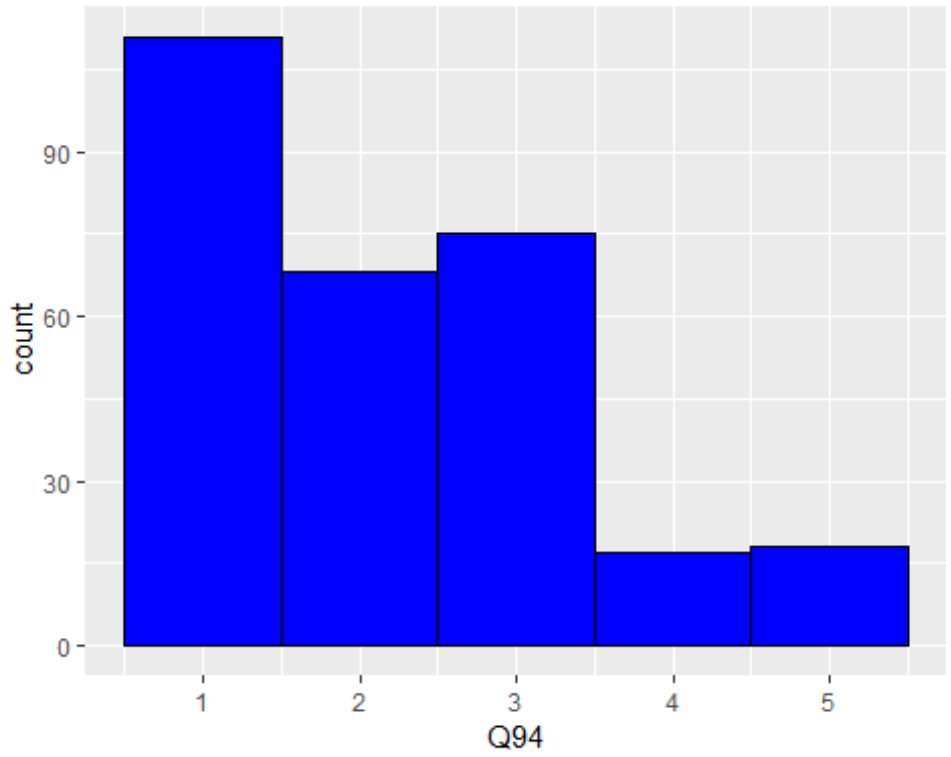
```
# [82] "Q108.1"                "Q48"                    "Q50"                    "
Q52"                   "Q54_1"                 "Q54_2"                 "Q54_3"
"Q54_4"                 "Q54_5"
# [91] "Q145"                  "Q99"                    "Q127"                   "
Q100.1"                "Booking.Channel"      "Flight.Date.Time"     "Pax.per.
PNR"           "Segment1.Destination" "Segment1.Origin"

# Response Variable (Factors that explain the performance of this variable):
# $Q1
# ""
# "0"
# "1"
# "10"
# "2"
# "3"
# "4"
# "5"
# "6"
# "7"
# "8"
# "9"
# "On\na scale from 0-10, how likely are you to recommend AirlineX Airlines t
o a friend\nor colleague?"

GuestData$Q1<-as.numeric(as.character(GuestData$Q1))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q1)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7 rows containing non-finite values (stat_bin).
```



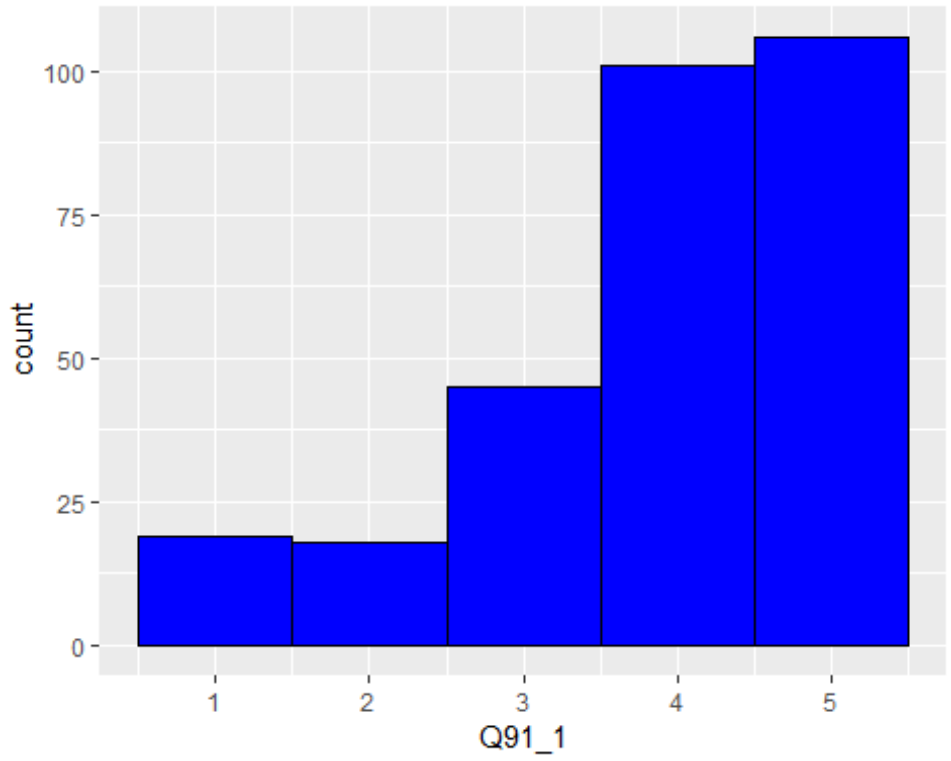

```
# $Q132
#""
#"Extremely dissatisfied"
#"Extremely satisfied"
#"How satisfied were you with your kiosk experience?"
#"Neither satisfied nor dissatisfied"
#"Somewhat dissatisfied"
#"Somewhat satisfied"

GuestData$Q132<-revalue(GuestData$Q132, c("Extremely dissatisfied" = "1", "So
mewhat dissatisfied" = "2", "Neither satisfied nor dissatisfied" = "3", "Some
what satisfied" = "4", "Extremely satisfied" = "5"))
GuestData$Q132<-as.numeric(as.character(GuestData$Q132))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q132)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7715 rows containing non-finite values (stat_bin).
```



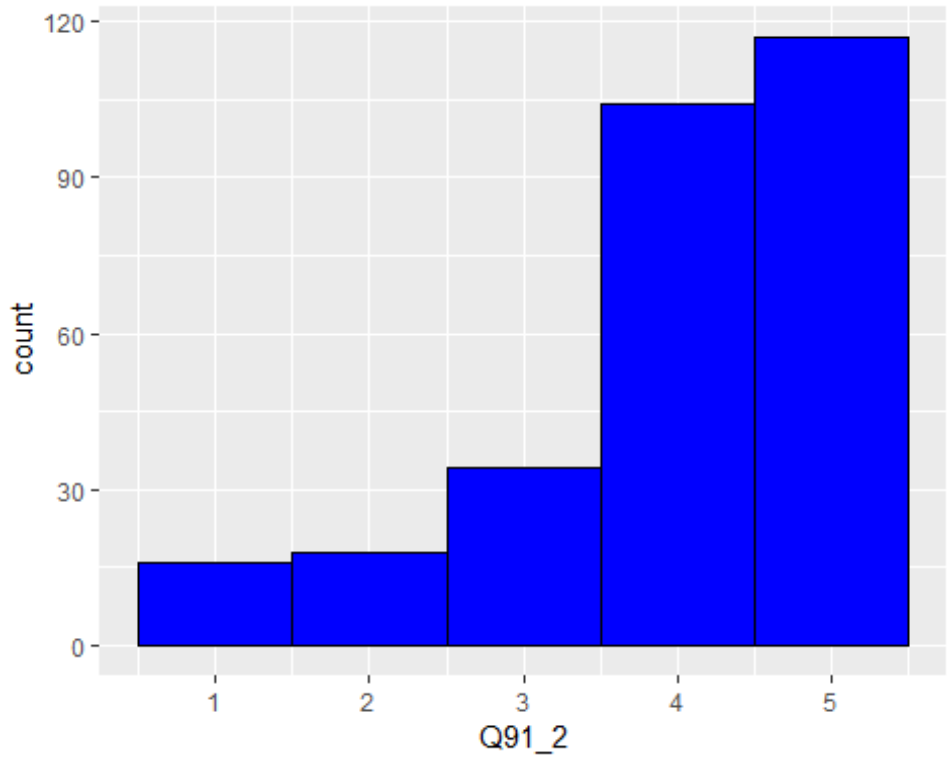

```
#$Q94
#""
#"How long were the lines to use the kiosks?"
#"Long"
#"Normal"
#"Short"
#"Very long"
#"Very short"

GuestData$Q94<-revalue(GuestData$Q94, c("Very short"="1", "Short"="2", "Norma
l"="3", "Long"="4", "Very long"="5"))
GuestData$Q94<-as.numeric(as.character(GuestData$Q94))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q94)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7712 rows containing non-finite values (stat_bin).
```



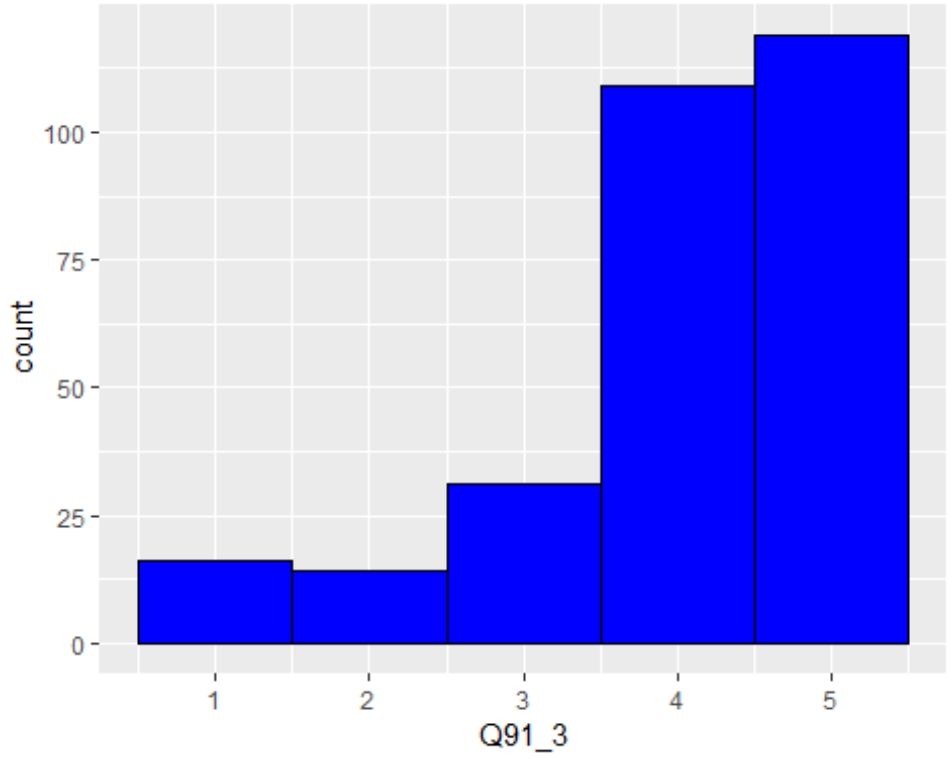

```
#$Q91_1
#""
#"Agree"
#"Disagree"
#"Neither agree nor disagree"
#"Please rate the following statements about your check-in kiosk interaction: \nThe check-in kiosk was... - Responsive"
#"Strongly agree"
#"Strongly Disagree"

GuestData$Q91_1<-revalue(GuestData$Q91_1, c("Strongly Disagree"="1", "Disagree"="2", "Neither agree nor disagree" = "3", "Agree" = "4", "Strongly agree" = "5"))
GuestData$Q91_1<-as.numeric(as.character(GuestData$Q91_1))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q91_1)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7712 rows containing non-finite values (stat_bin).
```



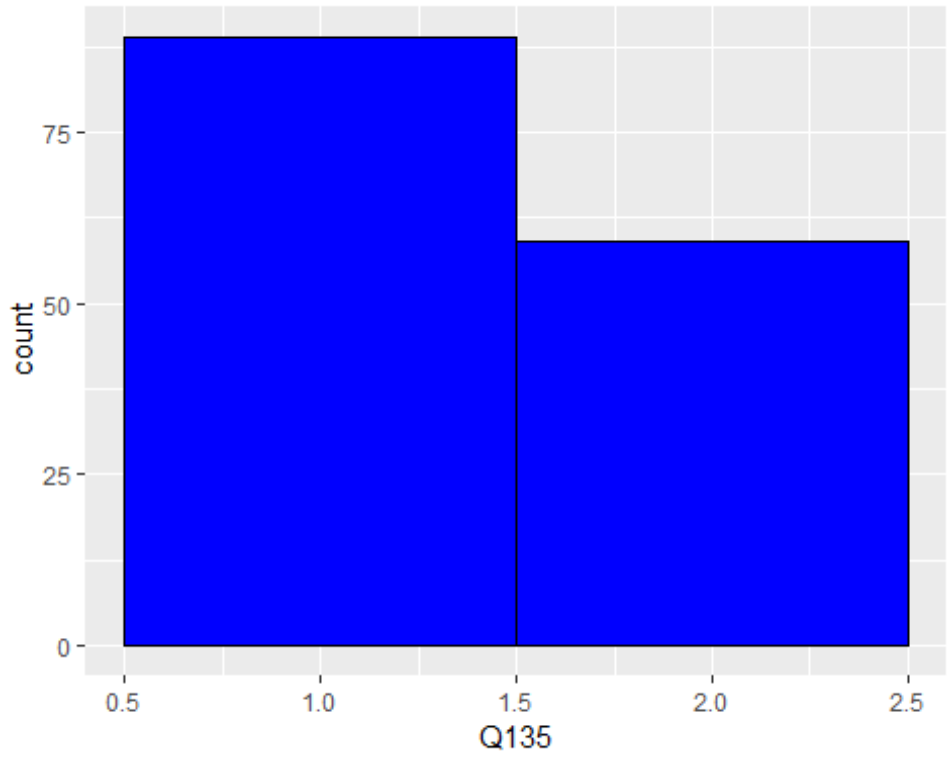

```
#$Q91_2
#""
#"Agree"
#"Disagree"
#"Neither agree nor disagree"
#"Please rate the following statements about your check-in kiosk interaction:
\nThe check-in kiosk was... - Easy-to-use"
#"Strongly agree"
#"Strongly Disagree"

GuestData$Q91_2<-revalue(GuestData$Q91_2, c("Strongly Disagree"="1", "Disagre
e"="2", "Neither agree nor disagree" = "3", "Agree" = "4", "Strongly agree" =
"5"))
GuestData$Q91_2<-as.numeric(as.character(GuestData$Q91_2))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q91_2)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7712 rows containing non-finite values (stat_bin).
```



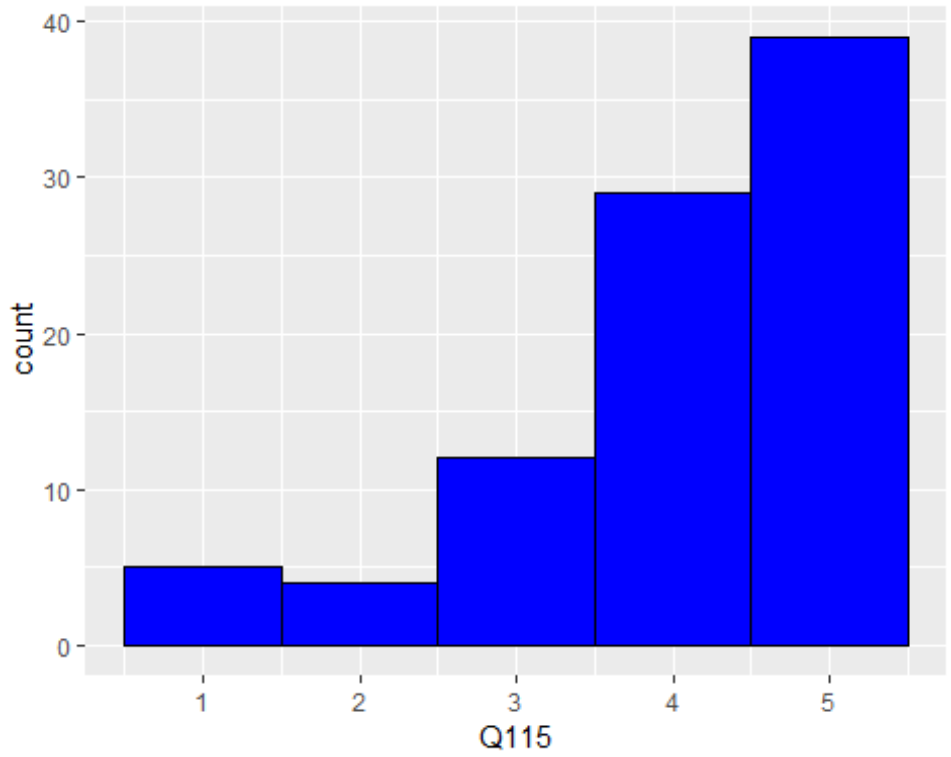

```
#$Q91_3
#""
#"Agree"
#"Disagree"
#"Neither agree nor disagree"
#"Please rate the following statements about your check-in kiosk interaction:
\nThe check-in kiosk was... - Available"
#"Strongly agree"
#"Strongly Disagree"

GuestData$Q91_3<-revalue(GuestData$Q91_3, c("Strongly Disagree"="1", "Disagre
e"="2", "Neither agree nor disagree" = "3", "Agree" = "4", "Strongly agree" =
"5"))
GuestData$Q91_3<-as.numeric(as.character(GuestData$Q91_3))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q91_3)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7712 rows containing non-finite values (stat_bin).
```



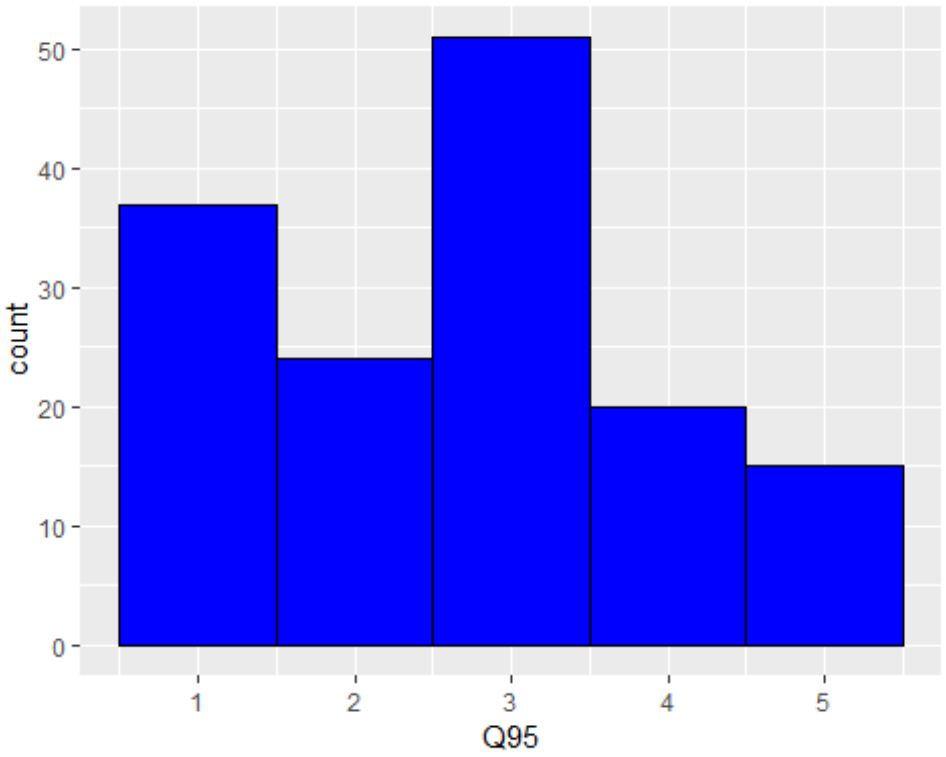

```
#$Q135
#""
#"Did you use the self-bag tags printed from the kiosk?"
#"No"
#"Yes"

GuestData$Q135<-revalue(GuestData$Q135, c("Yes"="1", "No"="2"))
GuestData$Q135<-as.numeric(as.character(GuestData$Q135))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q135)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7853 rows containing non-finite values (stat_bin).
```



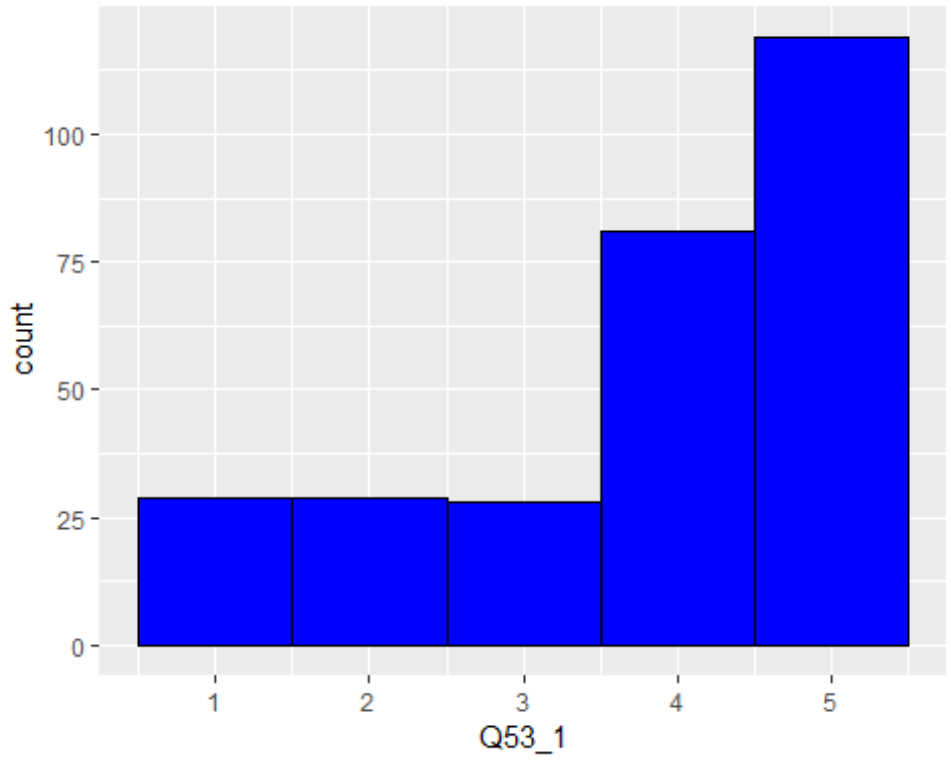

```
#$Q115
#""
#"Neither satisfied nor dissatisfied"
#"Not satisfied at all"
#"Overall, how satisfied were you with the self-bag tag process?"
#"Somewhat dissatisfied"
#"Somewhat satisfied"
#"Very satisfied"

GuestData$Q115<-revalue(GuestData$Q115, c("Not satisfied at all" = "1", "Some
what dissatisfied" = "2", "Neither satisfied nor dissatisfied" = "3", "Somewh
at satisfied" = "4", "Very satisfied" = "5"))
GuestData$Q115<-as.numeric(as.character(GuestData$Q115))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q115)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7912 rows containing non-finite values (stat_bin).
```



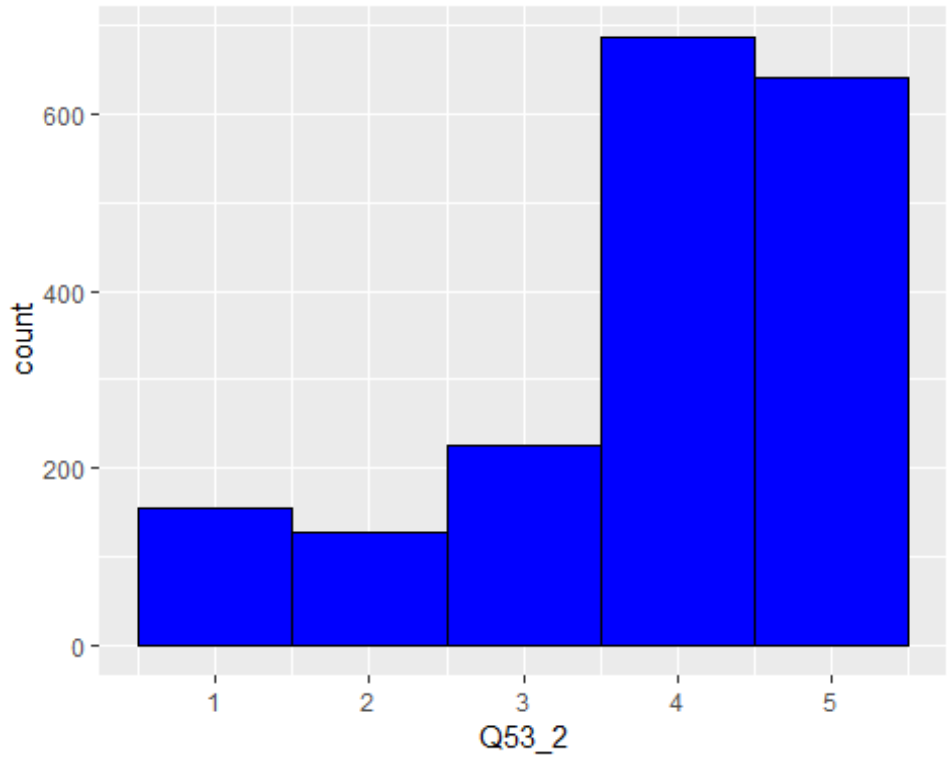

```
#$Q95
#""
#"How long were the lines to check your bags?"
#"Long"
#"Normal"
#"Short"
#"Very long"
#"Very short"

GuestData$Q95<-revalue(GuestData$Q95, c("Very short"="1", "Short"="2", "Norma
l"="3", "Long"="4", "Very long"="5"))
GuestData$Q95<-as.numeric(as.character(GuestData$Q95))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q95)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7854 rows containing non-finite values (stat_bin).
```



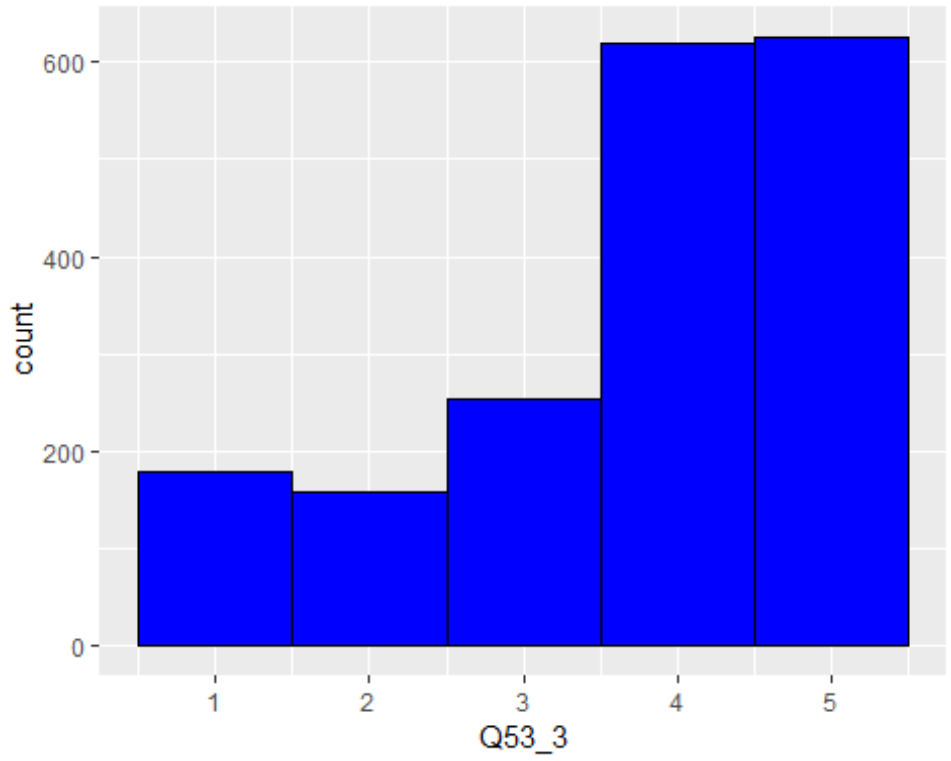

```
# $Q53_1
#""
#"Agree"
#"Disagree"
#"Neither agree nor disagree"
#"Please rate the following statements about your check-in experience:\n\nThe terminal lobby was... - Clean" "Strongly agree"
#"Strongly disagree"

GuestData$Q53_1<-revalue(GuestData$Q53_1, c("Strongly disagree"="1", "Disagree"="2", "Neither agree nor disagree" = "3", "Agree" = "4", "Strongly agree" = "5"))
GuestData$Q53_1<-as.numeric(as.character(GuestData$Q132))

ggplot(GuestData, aes(x=Q53_1)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7715 rows containing non-finite values (stat_bin).
```



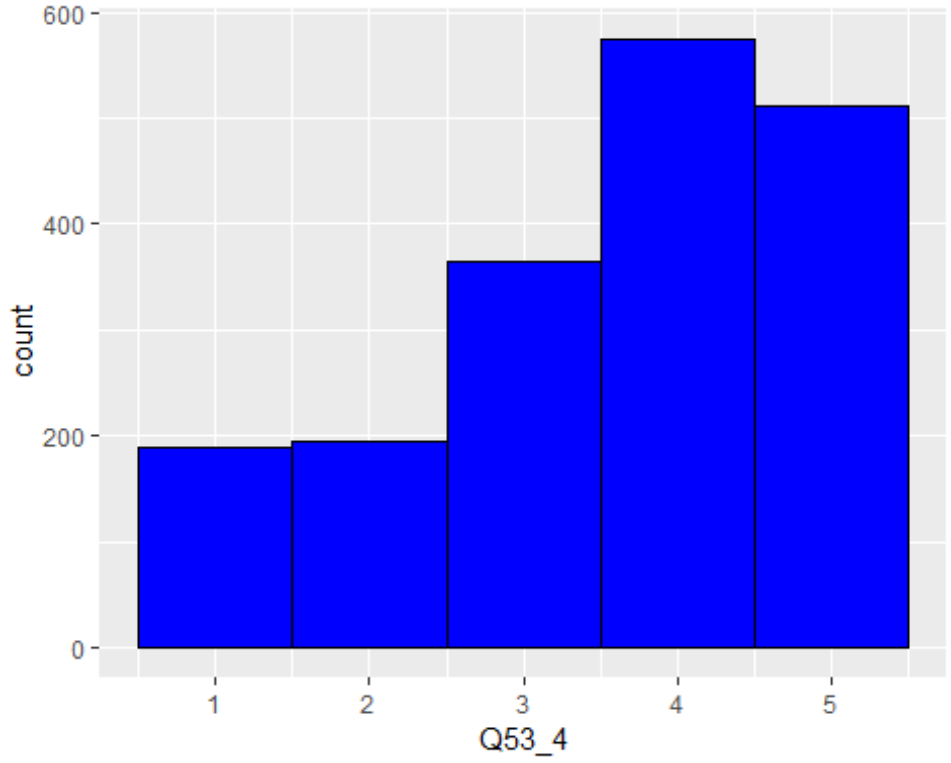

```
#$Q53_2
#""
#"Agree"
#"Disagree"
#"Neither agree nor disagree"
#"Please rate the following statements about your check-in experience:\n\nThe
terminal lobby was... - Organized"
#"Strongly agree"
#"Strongly disagree"

GuestData$Q53_2<-revalue(GuestData$Q53_2, c("Strongly disagree"="1", "Disagre
e"="2", "Neither agree nor disagree" = "3", "Agree" = "4", "Strongly agree" =
"5"))
GuestData$Q53_2<-as.numeric(as.character(GuestData$Q53_2))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q53_2)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 6166 rows containing non-finite values (stat_bin).
```



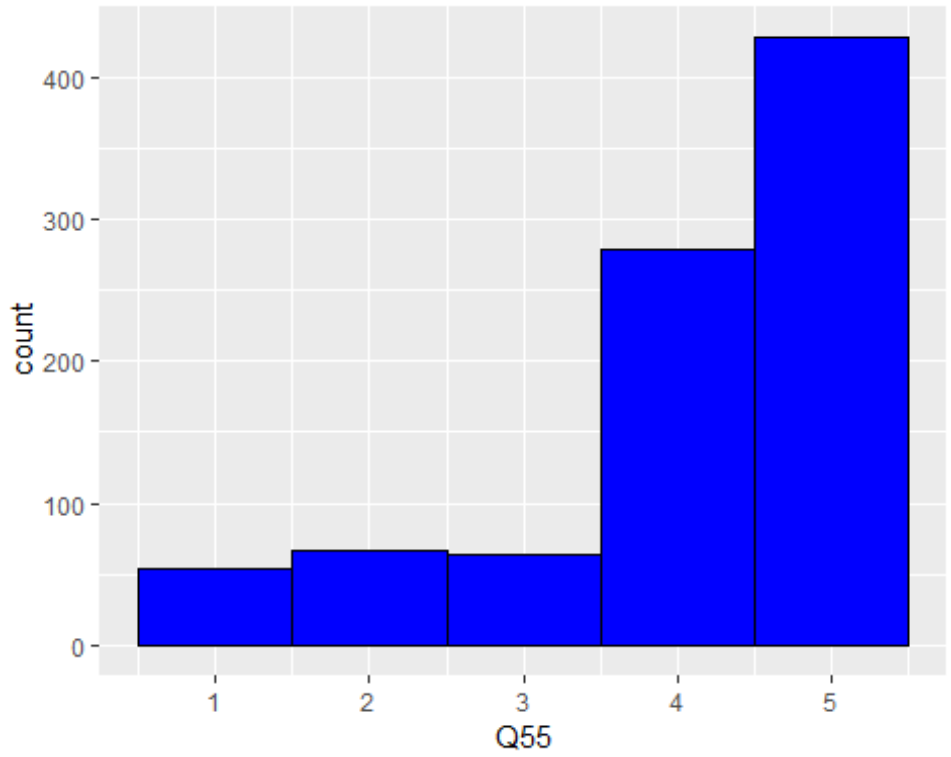

```
#$Q53_3
#""
#"Agree"
#"Disagree"
#"Neither agree nor disagree"
#"Please rate the following statements about your check-in experience:\n\nThe terminal lobby was... - Fast" "Strongly agree"
#"Strongly disagree"

GuestData$Q53_3<-revalue(GuestData$Q53_3, c("Strongly disagree"="1", "Disagree"="2", "Neither agree nor disagree" = "3", "Agree" = "4", "Strongly agree" = "5"))
GuestData$Q53_3<-as.numeric(as.character(GuestData$Q53_3))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q53_3)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 6166 rows containing non-finite values (stat_bin).
```



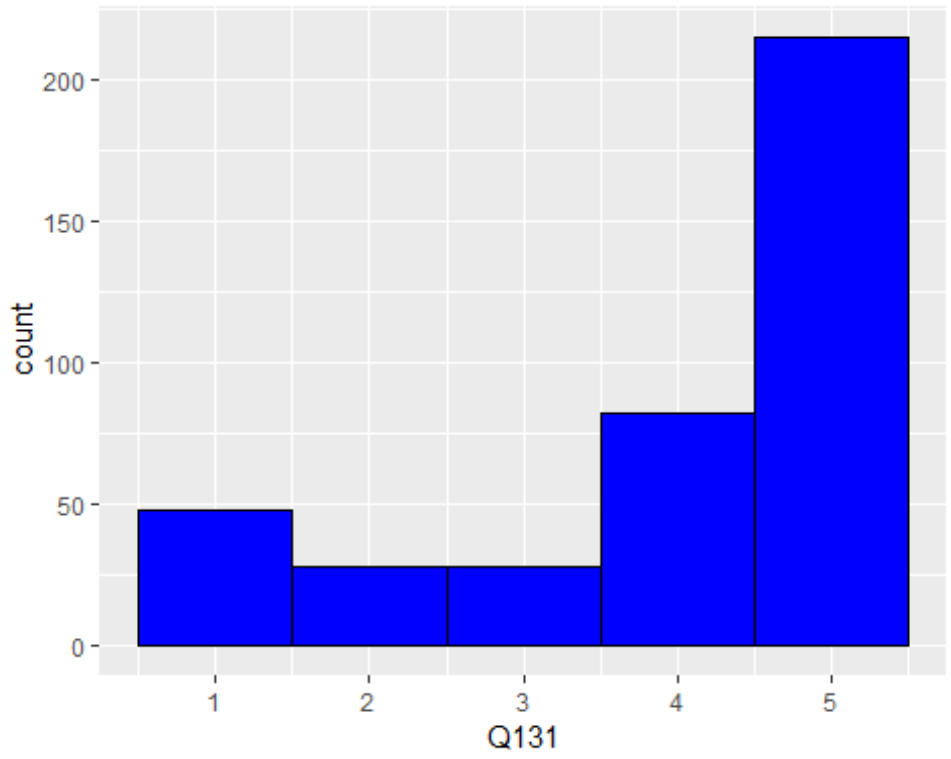

```
#$Q53_4
#""
#"Agree"
#"Disagree"
#"Neither agree nor disagree"
#"Please rate the following statements about your check-in experience:\n\nThe terminal lobby was... - Spacious" "Strongly agree"
#"Strongly disagree"

GuestData$Q53_4<-revalue(GuestData$Q53_4, c("Strongly disagree"="1", "Disagree"="2", "Neither agree nor disagree" = "3", "Agree" = "4", "Strongly agree" = "5"))
GuestData$Q53_4<-as.numeric(as.character(GuestData$Q53_4))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q53_4)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 6166 rows containing non-finite values (stat_bin).
```



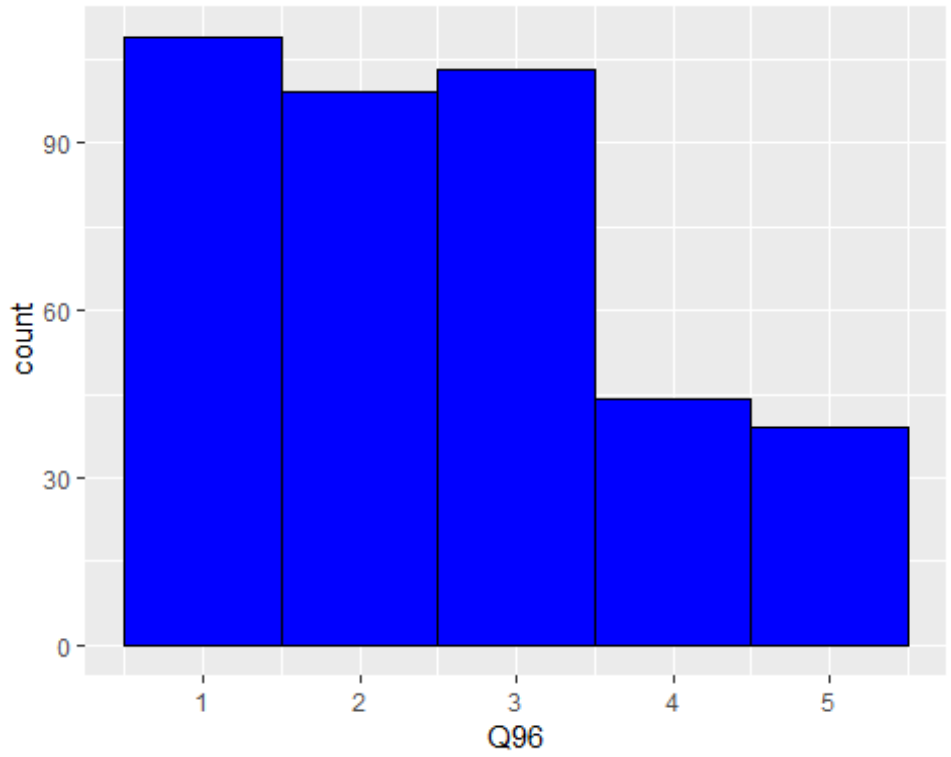

```
#$Q55
#""
#"Extremely clear"
#"Extremely unclear"
#"How clear were directions to check-in and bag drop after arriving to the te
rminal?"
#"Neither clear nor unclear"
#"Somewhat clear"
#"Somewhat unclear"

GuestData$Q55<-revalue(GuestData$Q55, c("Extremely unclear"="1", "Somewhat un
clear"="2", "Neither clear nor unclear" = "3", "Somewhat clear" = "4", "Extre
mely clear" = "5"))
GuestData$Q55<-as.numeric(as.character(GuestData$Q55))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q55)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7109 rows containing non-finite values (stat_bin).
```



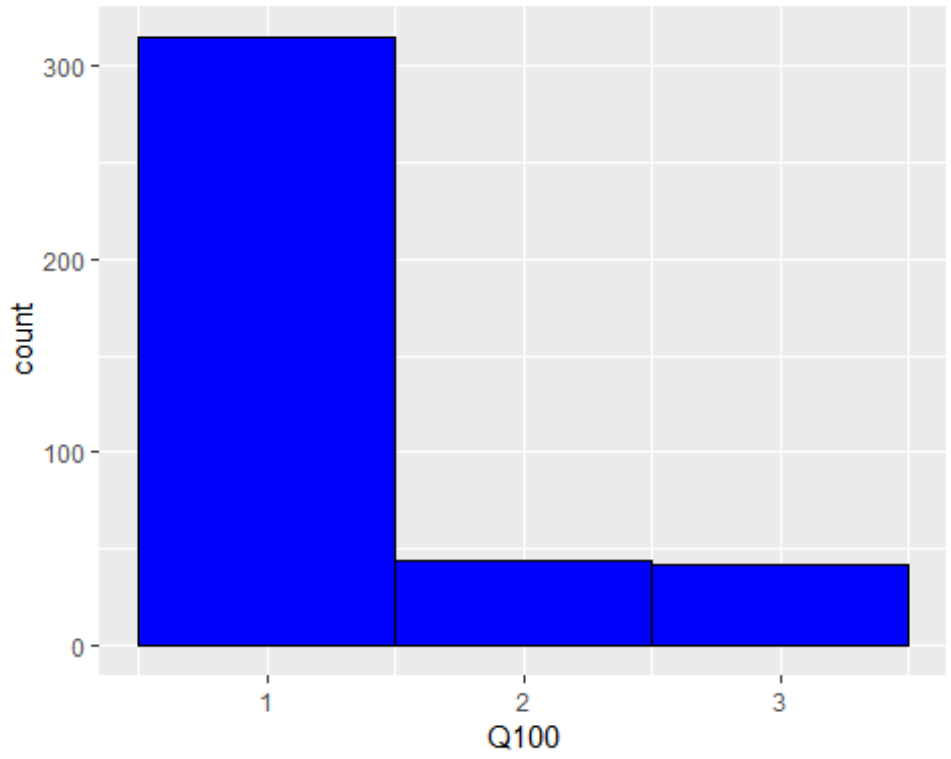

```
#$Q131
#""
#"Extremely dissatisfied"
#"Extremely satisfied"
#"How satisfied were you with your check-in or bag-drop counter experience?"
#"Neither satisfied nor dissatisfied"
#"Somewhat dissatisfied"
#"Somewhat satisfied"

GuestData$Q131<-revalue(GuestData$Q131, c("Extremely dissatisfied" = "1", "So
mewhat dissatisfied" = "2", "Neither satisfied nor dissatisfied" = "3", "Some
what satisfied" = "4", "Extremely satisfied" = "5"))
GuestData$Q131<-as.numeric(as.character(GuestData$Q131))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q131)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7600 rows containing non-finite values (stat_bin).
```



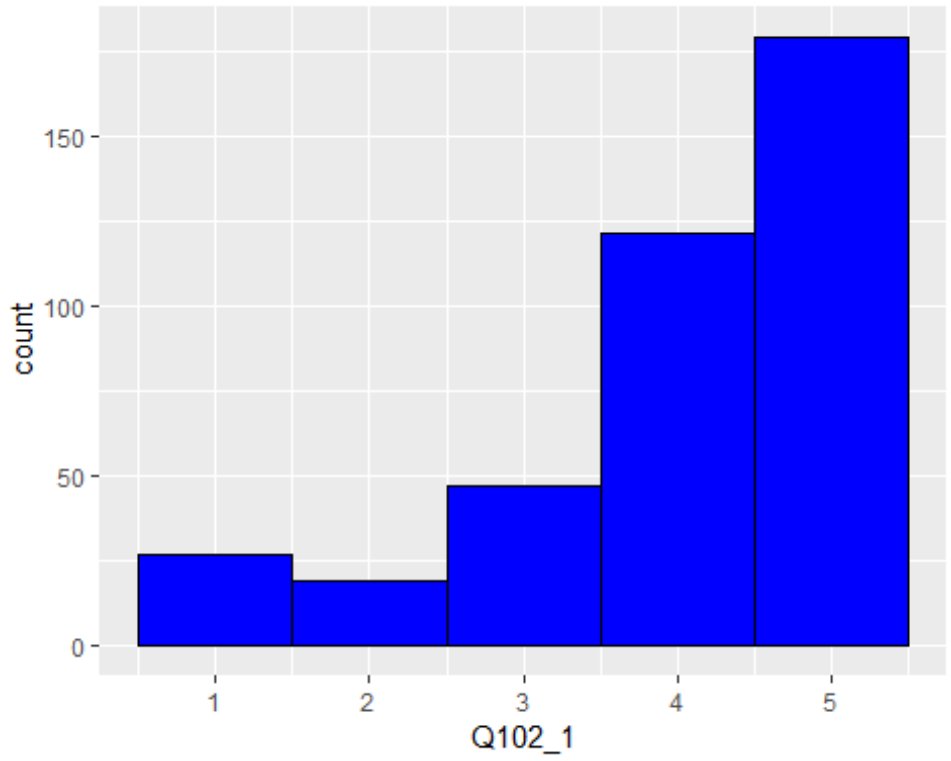

```
#$Q96
#""
#"How long were the lines to check-in at the counter?"
#"Long"
#"Normal"
#"Short"
#"Very long"
#"Very short"

GuestData$Q96<-revalue(GuestData$Q96, c("Very short"="1", "Short"="2", "Norma
l"="3", "Long"="4", "Very long"="5"))
GuestData$Q96<-as.numeric(as.character(GuestData$Q96))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q96)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7607 rows containing non-finite values (stat_bin).
```



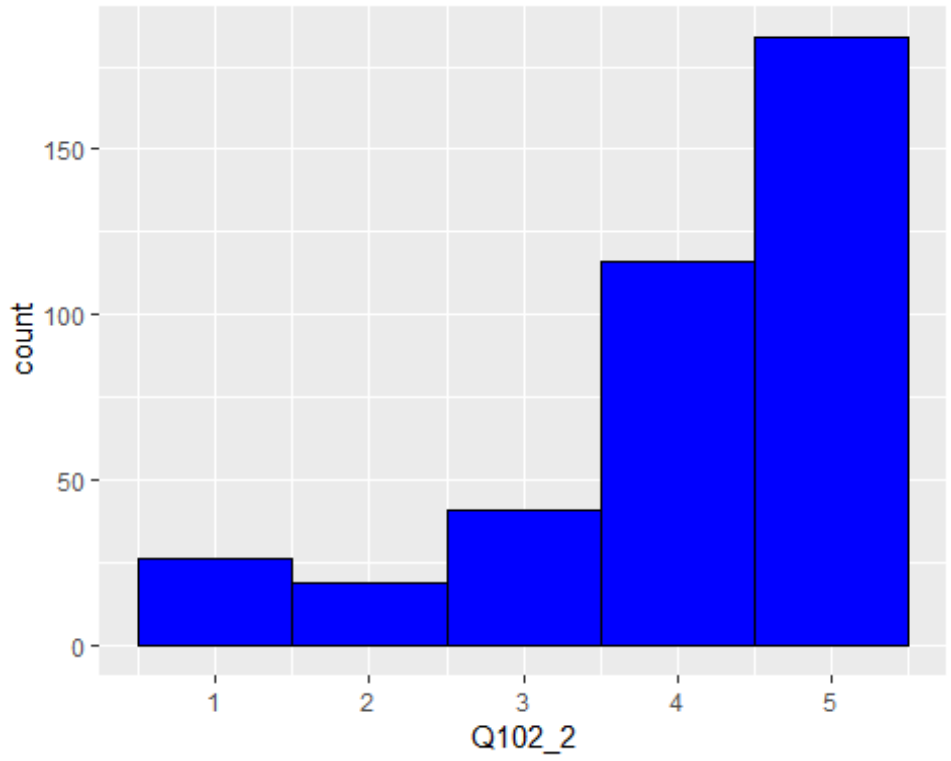

```
#$Q100
#""
#"After reaching the counter, were you greeted by the counter agent?"
#"I don't remember"
#"No"
#"Yes"

GuestData$Q100<-revalue(GuestData$Q100, c("Yes"="1", "I don't remember" = "2"
, "No"="3"))
GuestData$Q100<-as.numeric(as.character(GuestData$Q100))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q100)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7600 rows containing non-finite values (stat_bin).
```



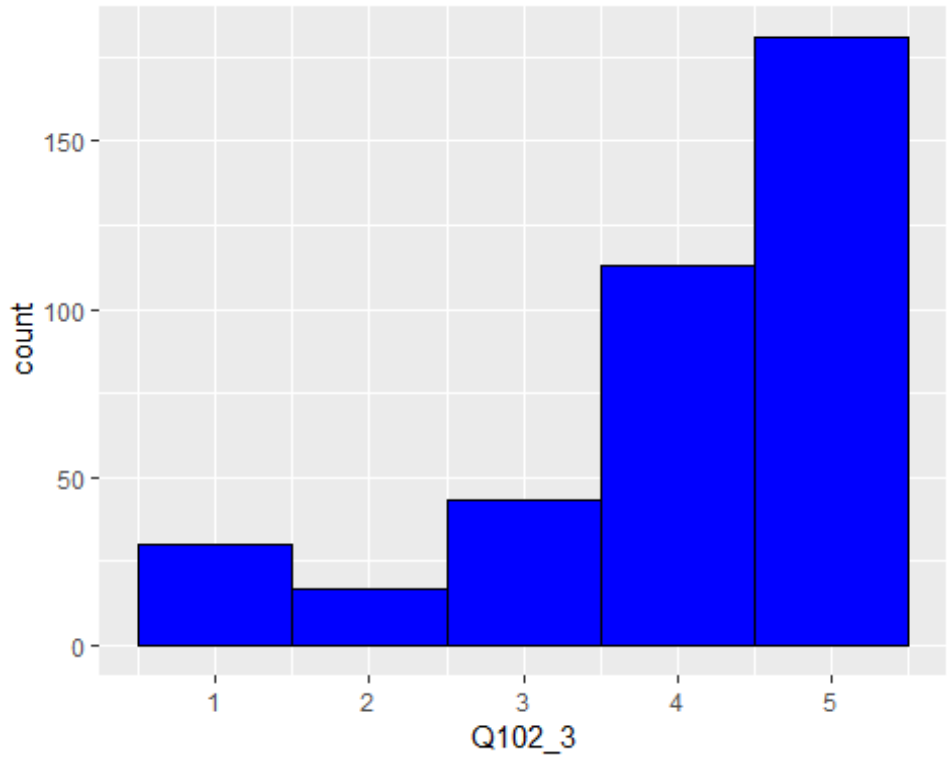

```
#$Q102_1
#""
#"Agree"
#"Disagree"
#"Neither agree nor disagree"
#"Please rate the following\nstatements:\n\n \n\nThe\ncustomer service agents
at the check-in/bag-drop counter on your flight from [Field-Journey%20Origin%
20Name] were: - Attentive"
#"Strongly agree"
#"Strongly disagree"

GuestData$Q102_1<-revalue(GuestData$Q102_1, c("Strongly disagree"="1", "Disag
ree"="2", "Neither agree nor disagree" = "3", "Agree" = "4", "Strongly agree"
= "5"))
GuestData$Q102_1<-as.numeric(as.character(GuestData$Q102_1))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q102_1)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7608 rows containing non-finite values (stat_bin).
```



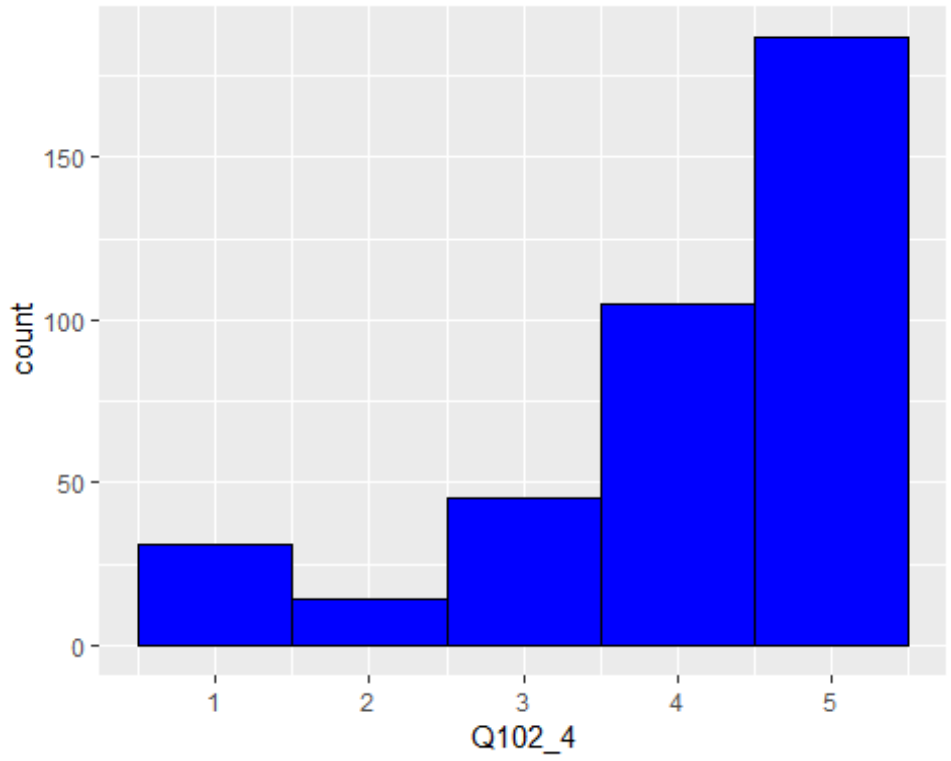

```
#$Q102_2
#""
#"Agree"
#"Disagree"
#"Neither agree nor disagree"
#"Please rate the following\nstatements:\n\n \n\nThe\ncustomer service agents
at the check-in/bag-drop counter on your flight from [Field-Journey%20Origin%
20Name] were: - Professional"
#"Strongly agree"
#"Strongly disagree"

GuestData$Q102_2<-revalue(GuestData$Q102_2, c("Strongly disagree"="1", "Disag
ree"="2", "Neither agree nor disagree" = "3", "Agree" = "4", "Strongly agree"
= "5"))
GuestData$Q102_2<-as.numeric(as.character(GuestData$Q102_2))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q102_2)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7615 rows containing non-finite values (stat_bin).
```



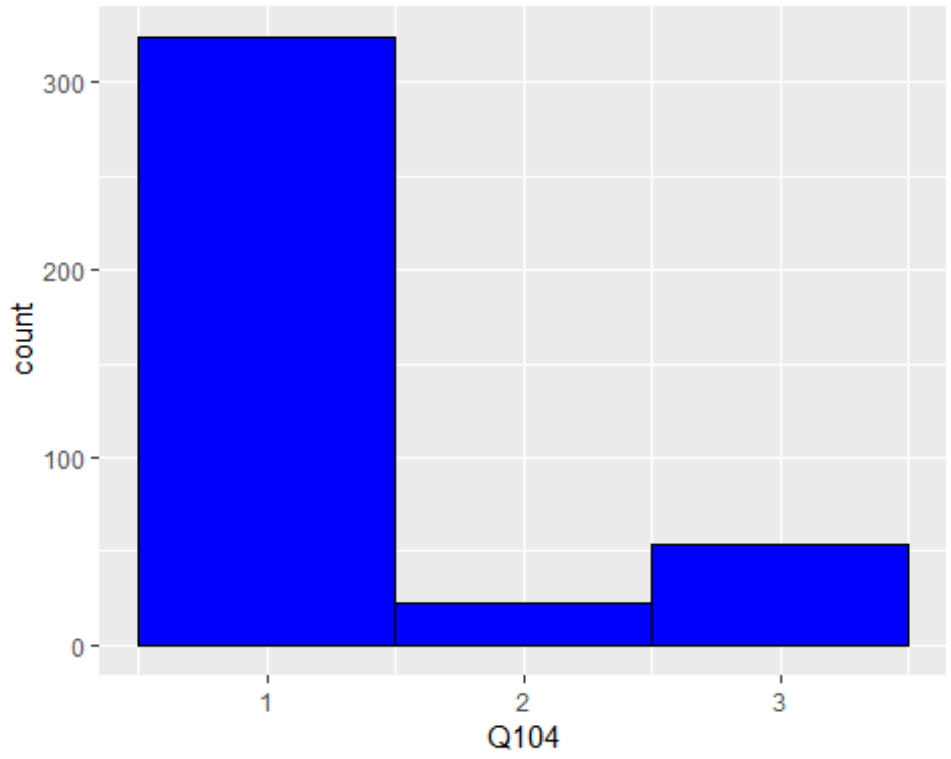

```
#$Q102_3
#""
#"Agree"
#"Disagree"
#"Neither agree nor disagree"
#"Please rate the following\nstatements:\n\n \n\nThe\ncustomer service agents
at the check-in/bag-drop counter on your flight from [Field-Journey%20Origin%
20Name] were: - Friendly"
#"Strongly agree"
#"Strongly disagree"

GuestData$Q102_3<-revalue(GuestData$Q102_3, c("Strongly disagree"="1", "Disag
ree"="2", "Neither agree nor disagree" = "3", "Agree" = "4", "Strongly agree"
= "5"))
GuestData$Q102_3<-as.numeric(as.character(GuestData$Q102_3))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q102_3)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7617 rows containing non-finite values (stat_bin).
```



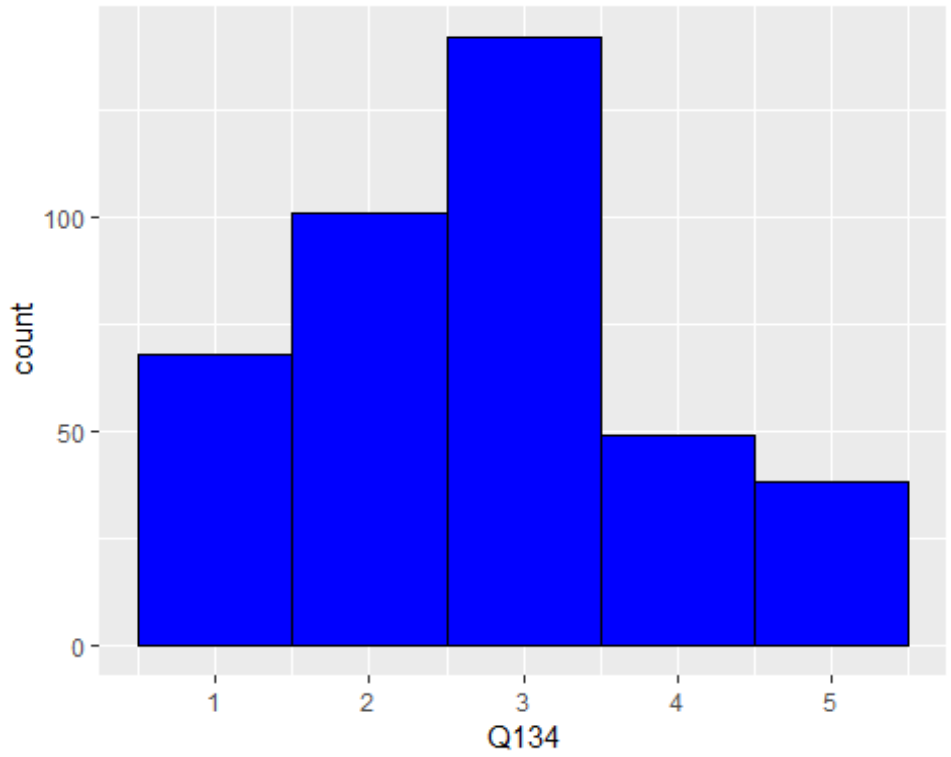

```
#$Q102_4
#""
#"Agree"
#"Disagree"
#"Neither agree nor disagree"
#"Please rate the following\nstatements:\n\n \n\nThe\ncustomer service agents
at the check-in/bag-drop counter on your flight from [Field-Journey%20Origin%
20Name] were: - Helpful"
#"Strongly agree"
#"Strongly disagree"

GuestData$Q102_4<-revalue(GuestData$Q102_4, c("Strongly disagree"="1", "Disag
ree"="2", "Neither agree nor disagree" = "3", "Agree" = "4", "Strongly agree"
= "5"))
GuestData$Q102_4<-as.numeric(as.character(GuestData$Q102_4))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q102_4)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7619 rows containing non-finite values (stat_bin).
```



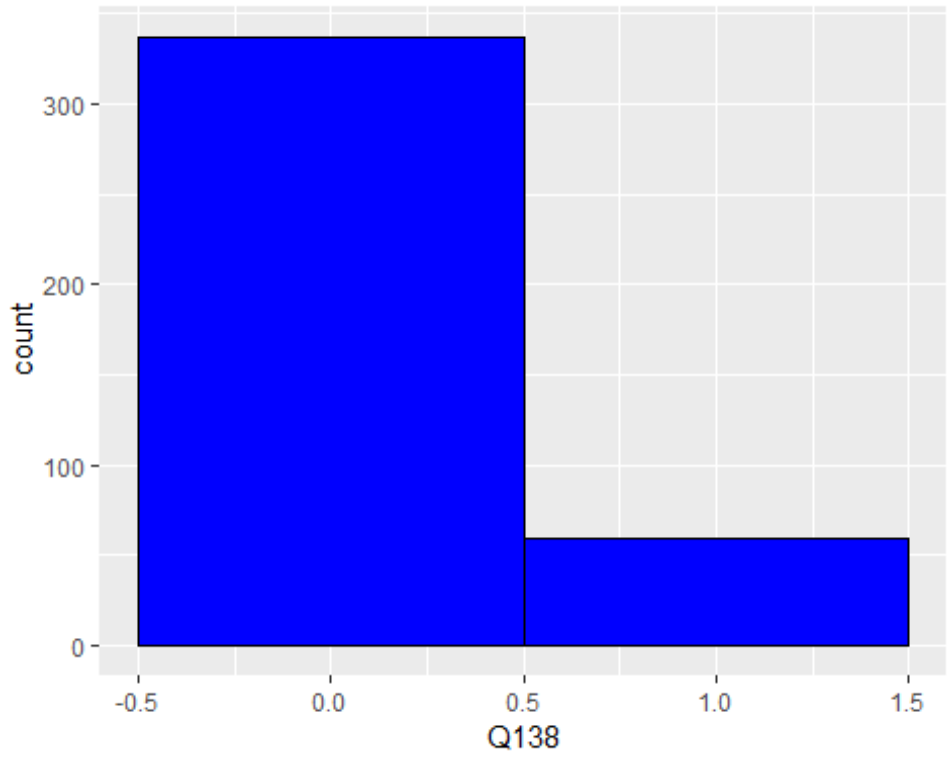

```
#$Q104
#""
#"At the end of your ticket counter interaction, did the agent provide you wi
th the current gate information and boarding time?"
#"I don't remember"
#"No"
#"Yes"

GuestData$Q104<-revalue(GuestData$Q104, c("Yes"="1", "I don't remember" = "2"
, "No"="3"))
GuestData$Q104<-as.numeric(as.character(GuestData$Q104))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q104)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7600 rows containing non-finite values (stat_bin).
```



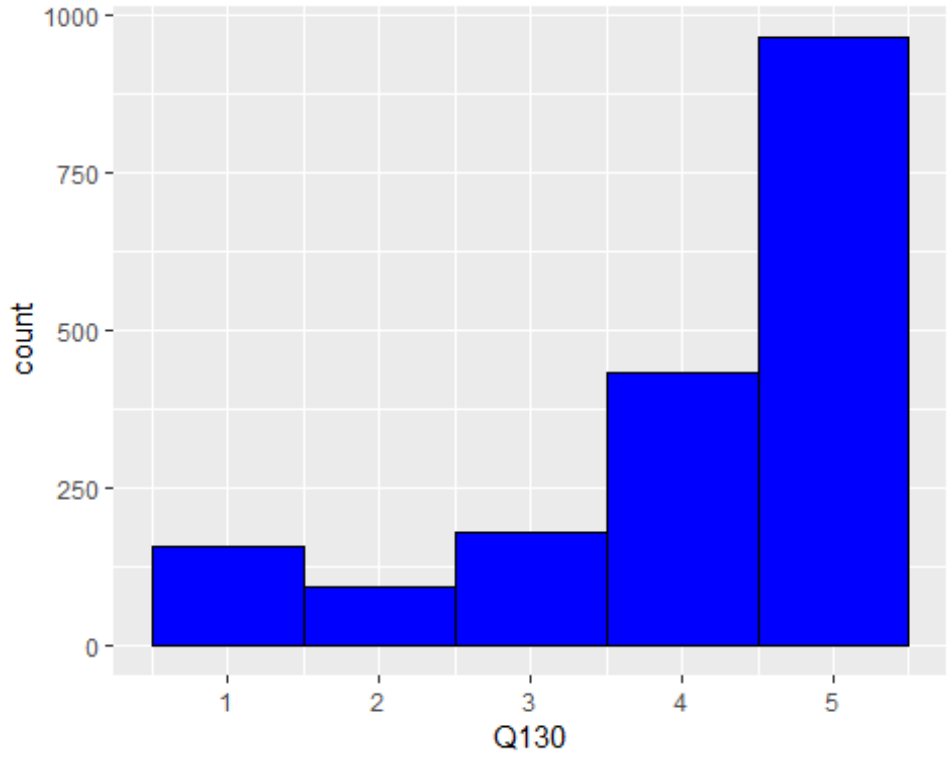

```
#$Q134
#""
#"How long were the lines to get through the TSA security line?"
#"Long"
#"Normal"
#"Short"
#"Very long"
#"Very short"

GuestData$Q134<-revalue(GuestData$Q134, c("Very short"="1", "Short"="2", "Nor
mal"="3", "Long"="4", "Very long"="5"))
GuestData$Q134<-as.numeric(as.character(GuestData$Q134))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q134)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7603 rows containing non-finite values (stat_bin).
```



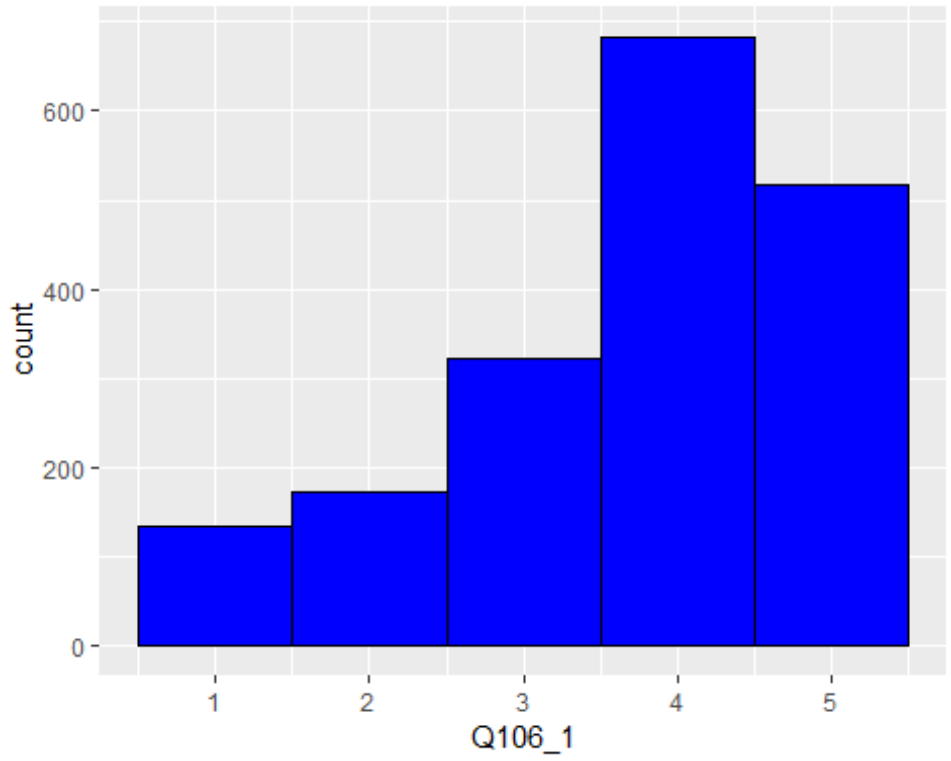

```
#$Q138
#""
#"Did you experience any issues during check-in?"
#"No"
#"Yes"

GuestData$Q138<-revalue(GuestData$Q138, c("Yes"="1", "No"="0"))
GuestData$Q138<-as.numeric(as.character(GuestData$Q138))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q138)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7605 rows containing non-finite values (stat_bin).
```



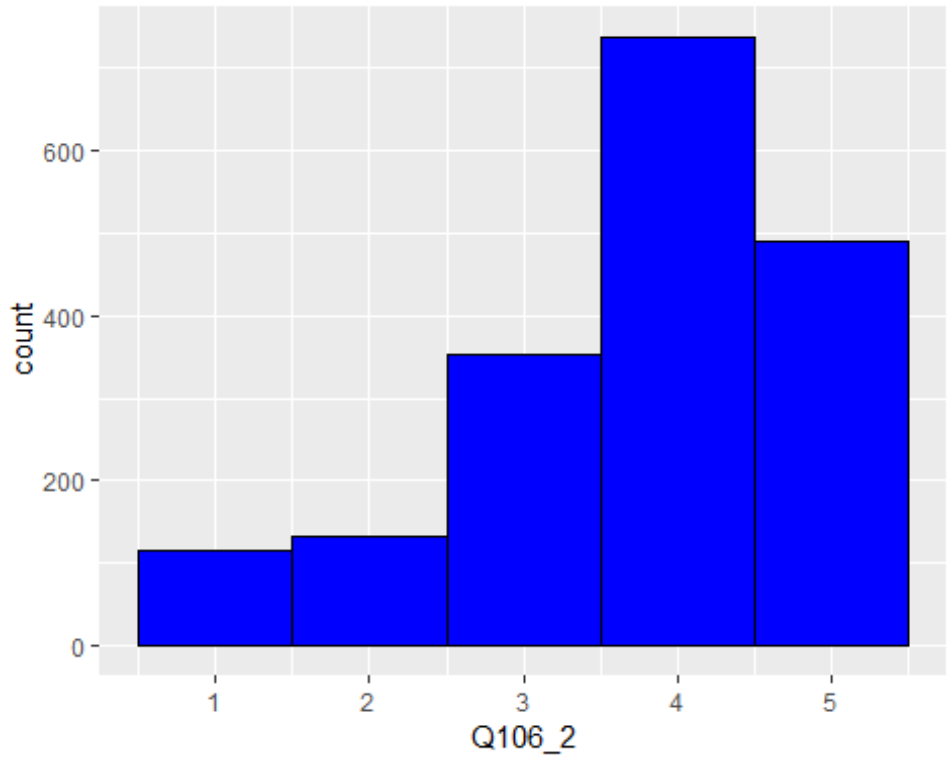

```
#$Q130
#""
#"Extremely dissatisfied"
#"Extremely satisfied"
#"How satisfied were you with your gate experience?"
#"Neither satisfied nor dissatisfied"
#"Somewhat dissatisfied"
#"Somewhat satisfied"

GuestData$Q130<-revalue(GuestData$Q130, c("Extremely dissatisfied" = "1", "So
mewhat dissatisfied" = "2", "Neither satisfied nor dissatisfied" = "3", "Some
what satisfied" = "4", "Extremely satisfied" = "5"))
GuestData$Q130<-as.numeric(as.character(GuestData$Q130))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q130)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 6177 rows containing non-finite values (stat_bin).
```



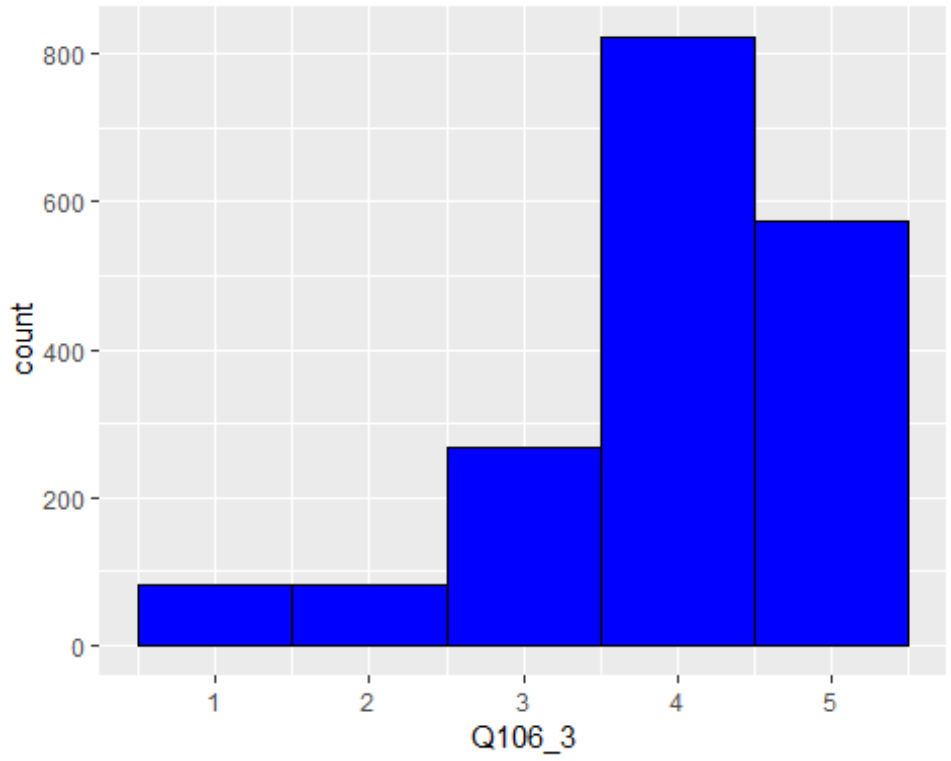

```
#$Q106_1
#""
#"Agree"
#"Disagree"
#"Neither agree nor disagree"
#"Please rate the following statements about the gate and boarding area:\n\nT
he gate area was... - Spacious"
#"Strongly agree"
#"Strongly disagree"

GuestData$Q106_1<-revalue(GuestData$Q106_1, c("Strongly disagree"="1", "Disag
ree"="2", "Neither agree nor disagree" = "3", "Agree" = "4", "Strongly agree"
= "5"))
GuestData$Q106_1<-as.numeric(as.character(GuestData$Q106_1))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(Q106_1)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 6174 rows containing non-finite values (stat_bin).
```



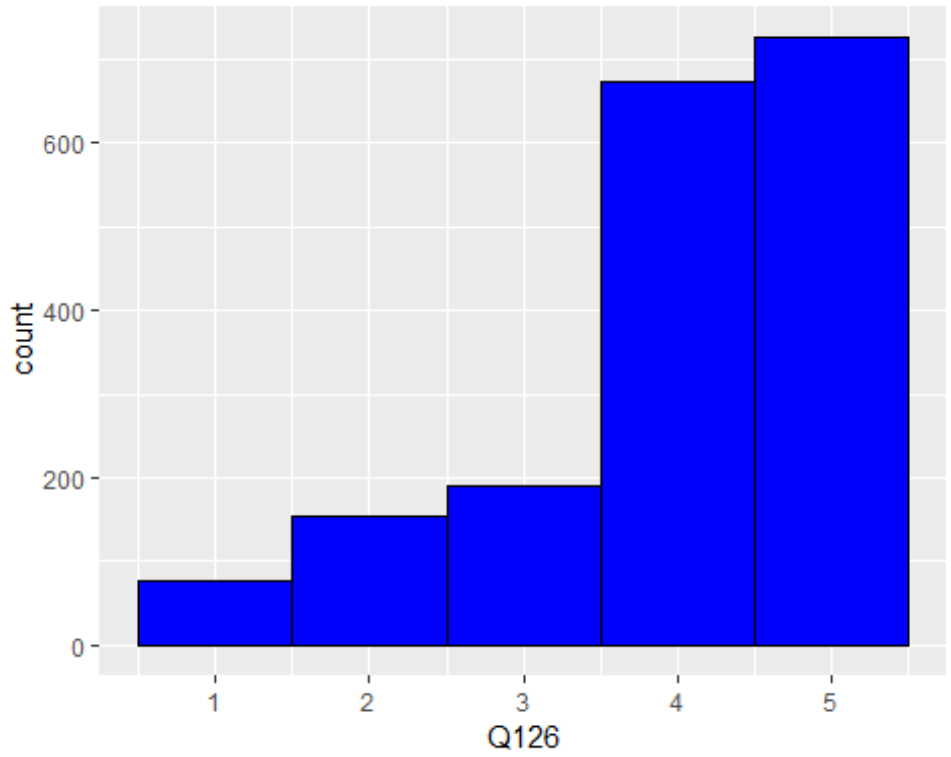

```
#$Q106_2
#""
#"Agree"
#"Disagree"
#"Neither agree nor disagree"
#"Please rate the following statements about the gate and boarding area:\n\nT
he gate area was... - Comfortable"
#"Strongly agree"
#"Strongly disagree"

GuestData$Q106_2<-revalue(GuestData$Q106_2, c("Strongly disagree"="1", "Disag
ree"="2", "Neither agree nor disagree" = "3", "Agree" = "4", "Strongly agree"
= "5"))
GuestData$Q106_2<-as.numeric(as.character(GuestData$Q106_2))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q106_2)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 6174 rows containing non-finite values (stat_bin).
```



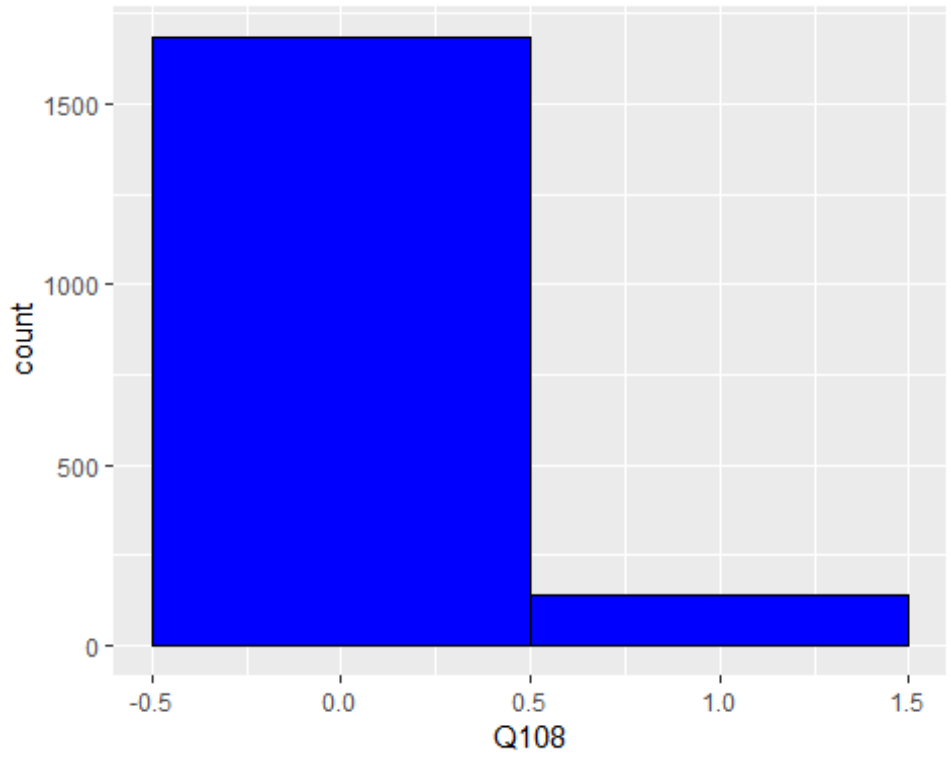

```
#$Q106_3
#""
#"Agree"
#"Disagree"
#"Neither agree nor disagree"
#"Please rate the following statements about the gate and boarding area:\n\nT
he gate area was... - Clean"
#"Strongly agree"
#"Strongly disagree"

GuestData$Q106_3<-revalue(GuestData$Q106_3, c("Strongly disagree"="1", "Disag
ree"="2", "Neither agree nor disagree" = "3", "Agree" = "4", "Strongly agree"
= "5"))
GuestData$Q106_3<-as.numeric(as.character(GuestData$Q106_3))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q106_3)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 6174 rows containing non-finite values (stat_bin).
```



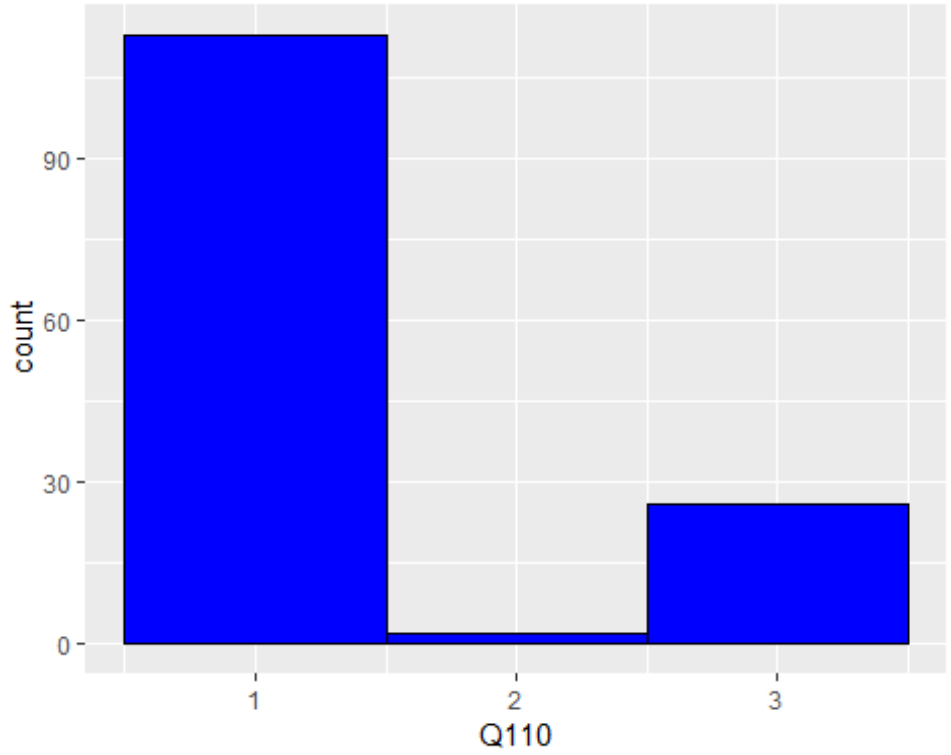

```
#$Q126
#""
#"Clear"
#"How clear were announcements made at the gate?"
#"Neither clear nor poor"
#"Poor"
#"Very clear"
#"Very poor"

GuestData$Q126<-revalue(GuestData$Q126, c("Very poor"="1", "Poor"="2", "Neith
er clear nor poor" = "3", "Clear" = "4", "Very clear" = "5"))
GuestData$Q126<-as.numeric(as.character(GuestData$Q126))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q126)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 6180 rows containing non-finite values (stat_bin).
```



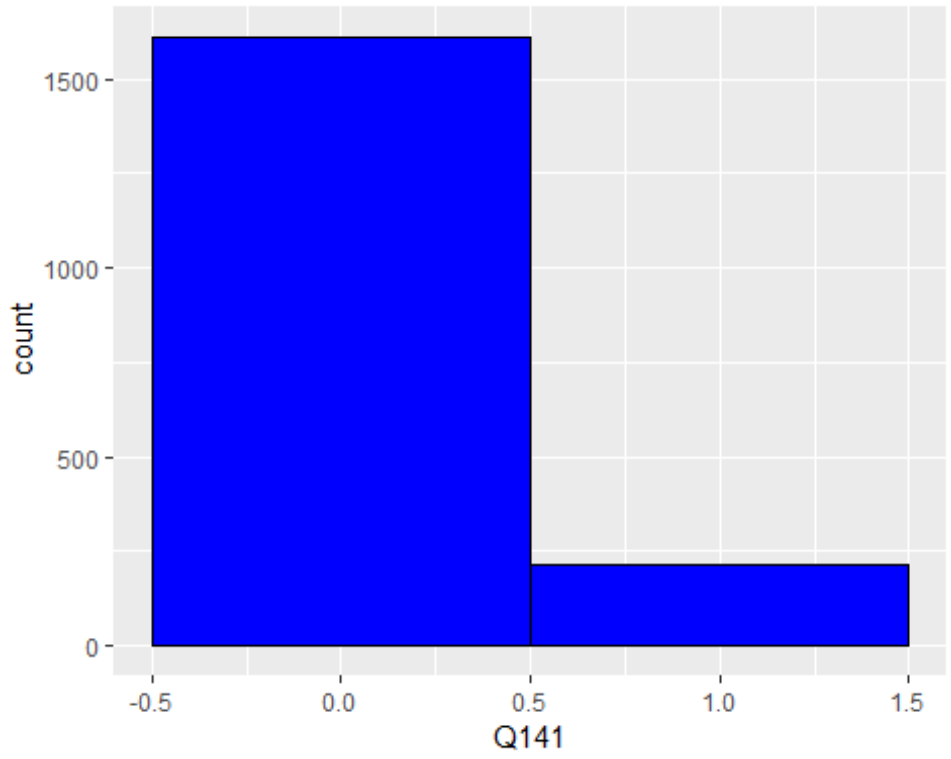

```
#$Q108
#""
#"Did you require any assistance from the gate agent after going through secu
rity?"
#"No"
#"Yes"

GuestData$Q108<-revalue(GuestData$Q108, c("Yes"="1", "No"="0"))
GuestData$Q108<-as.numeric(as.character(GuestData$Q108))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q108)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 6174 rows containing non-finite values (stat_bin).
```



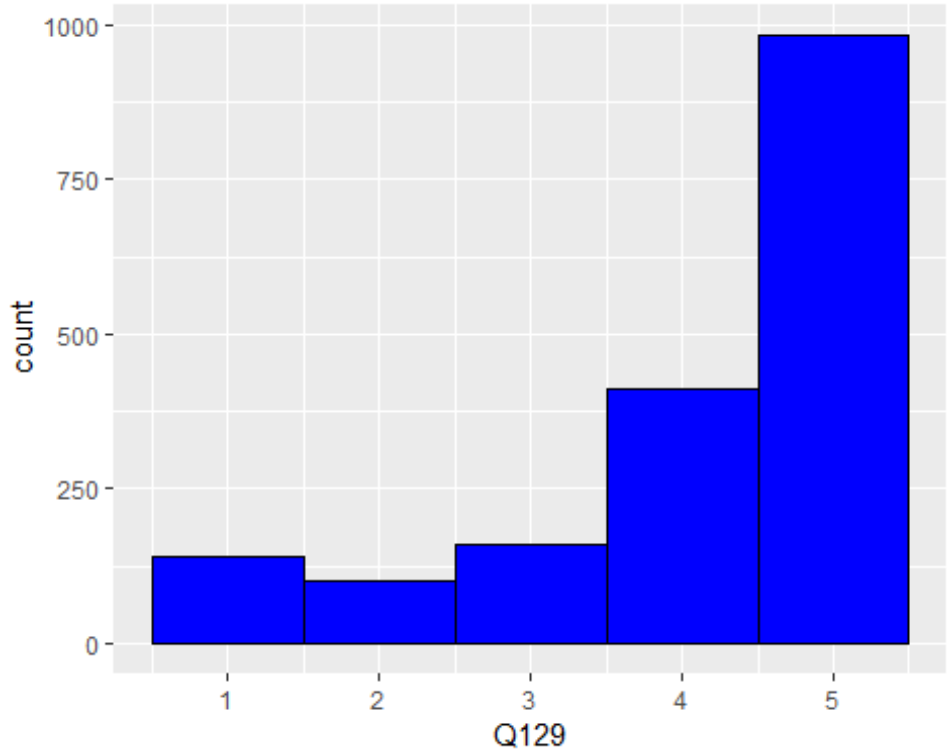

```
#$Q110
#""
#"I don't remember"
#"No"
#"Were AirlineX team members available at the gate counter to assist you?"
#"Yes"

GuestData$Q110<-revalue(GuestData$Q110, c("Yes"="1", "I don't remember" = "2"
, "No"="3"))
GuestData$Q110<-as.numeric(as.character(GuestData$Q110))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q110)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7860 rows containing non-finite values (stat_bin).
```



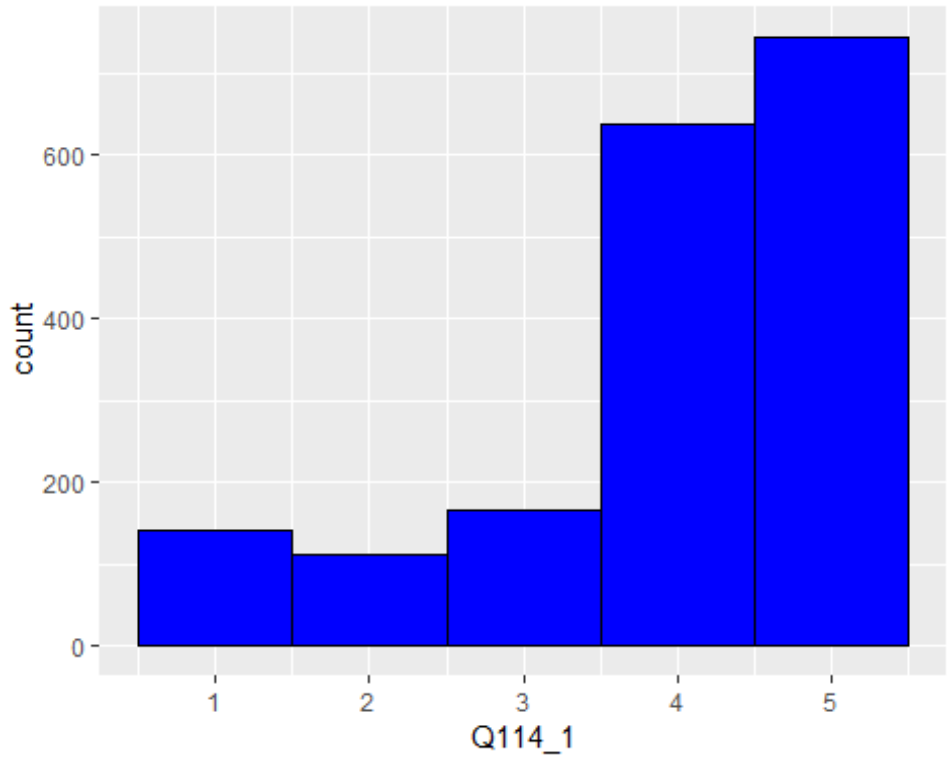

```
#$Q141
#""
#"Did you experience any issues at the gate?"
#"No"
#"Yes"

GuestData$Q141<-revalue(GuestData$Q141, c("Yes"="1", "No"="0"))
GuestData$Q141<-as.numeric(as.character(GuestData$Q141))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q141)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 6176 rows containing non-finite values (stat_bin).
```



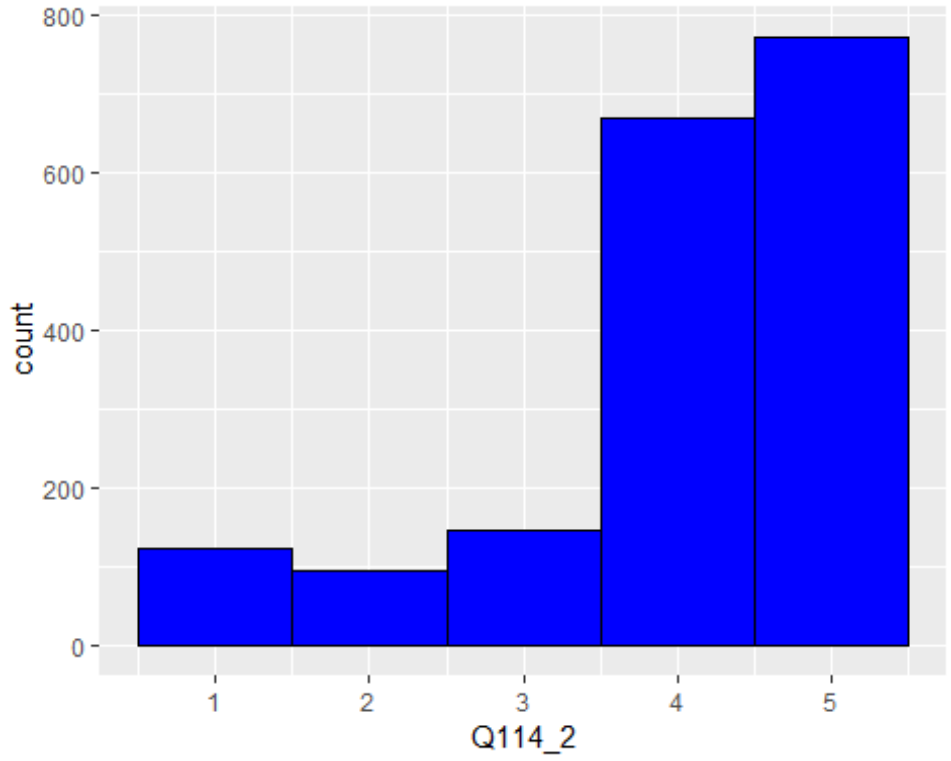

```
#$Q129
#""
#"Extremely dissatisfied"
#"Extremely satisfied"
#"How satisfied were you with the boarding process?"
#"Neither satisfied nor dissatisfied"
#"Somewhat dissatisfied"
#"Somewhat satisfied"

GuestData$Q129<-revalue(GuestData$Q129, c("Extremely dissatisfied" = "1", "So
mewhat dissatisfied" = "2", "Neither satisfied nor dissatisfied" = "3", "Some
what satisfied" = "4", "Extremely satisfied" = "5"))
GuestData$Q129<-as.numeric(as.character(GuestData$Q129))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q129)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 6205 rows containing non-finite values (stat_bin).
```



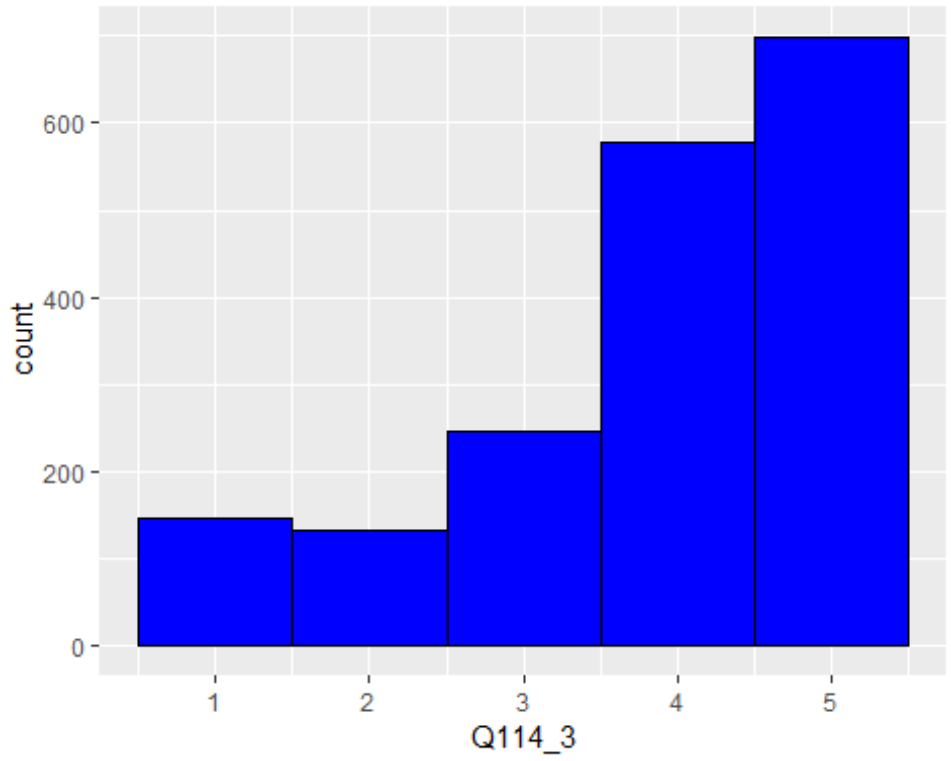

```
#$Q114_1
#""
#"Agree"
#"Disagree"
#"Neither agree nor disagree"
#"Please rate the following statements about the boarding process:\n\nThe boa
rding process was... - Organized"
#"Strongly agree"
#"Strongly disagree"

GuestData$Q114_1<-revalue(GuestData$Q114_1, c("Strongly disagree"="1", "Disag
ree"="2", "Neither agree nor disagree" = "3", "Agree" = "4", "Strongly agree"
= "5"))
GuestData$Q114_1<-as.numeric(as.character(GuestData$Q114_1))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q114_1)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 6200 rows containing non-finite values (stat_bin).
```



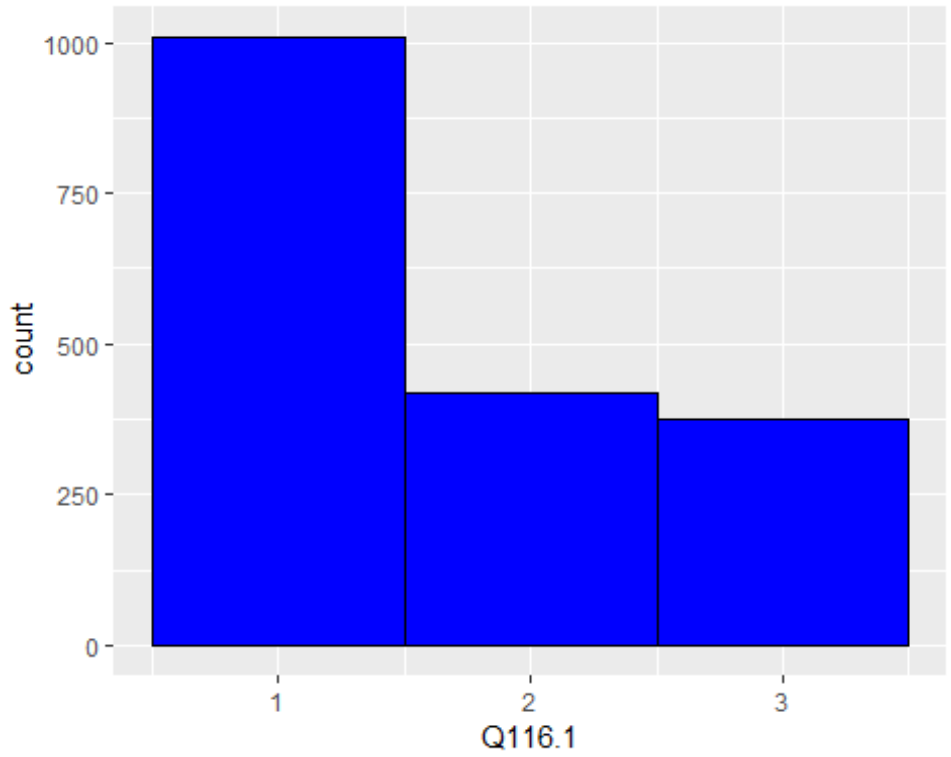

```
#$Q114_2
#""
#"Agree"
#"Disagree"
#"Neither agree nor disagree"
#"Please rate the following statements about the boarding process:\n\nThe boa
rding process was... - Easy to understand"
#"Strongly agree"
#"Strongly disagree"

GuestData$Q114_2<-revalue(GuestData$Q114_2, c("Strongly disagree"="1", "Disag
ree"="2", "Neither agree nor disagree" = "3", "Agree" = "4", "Strongly agree"
= "5"))
GuestData$Q114_2<-as.numeric(as.character(GuestData$Q114_2))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q114_2)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 6200 rows containing non-finite values (stat_bin).
```




```
#$Q114_3
#""
#"Agree"
#"Disagree"
#"Neither agree nor disagree"
#"Please rate the following statements about the boarding process:\n\nThe boa
rding process was... - Fast"
#"Strongly agree"
#"Strongly disagree"

GuestData$Q114_3<-revalue(GuestData$Q114_3, c("Strongly disagree"="1", "Disag
ree"="2", "Neither agree nor disagree" = "3", "Agree" = "4", "Strongly agree"
= "5"))
GuestData$Q114_3<-as.numeric(as.character(GuestData$Q114_3))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q114_3)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 6200 rows containing non-finite values (stat_bin).
```



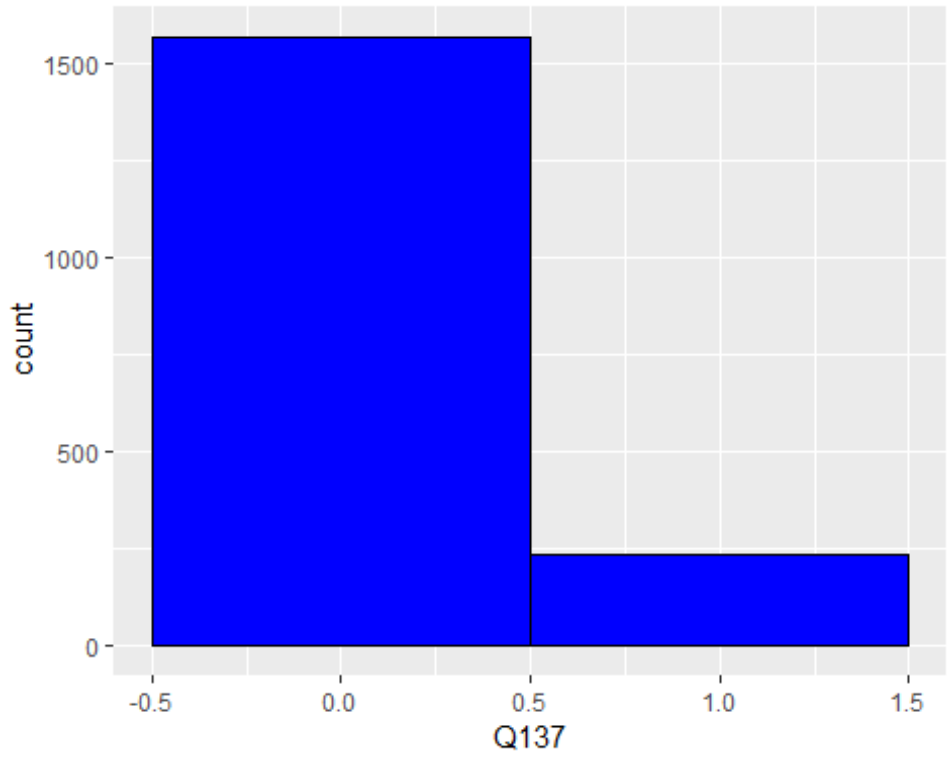

```
#$Q116.1
#""
#"Did the gate agent thank you for flying AirlineX?"
#"I don't remember"
#"No"
#"Yes"

GuestData$Q116.1<-revalue(GuestData$Q116.1, c("Yes"="1", "I don't remember" =
"2", "No"="3"))
GuestData$Q116.1<-as.numeric(as.character(GuestData$Q116.1))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q116.1)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 6200 rows containing non-finite values (stat_bin).
```



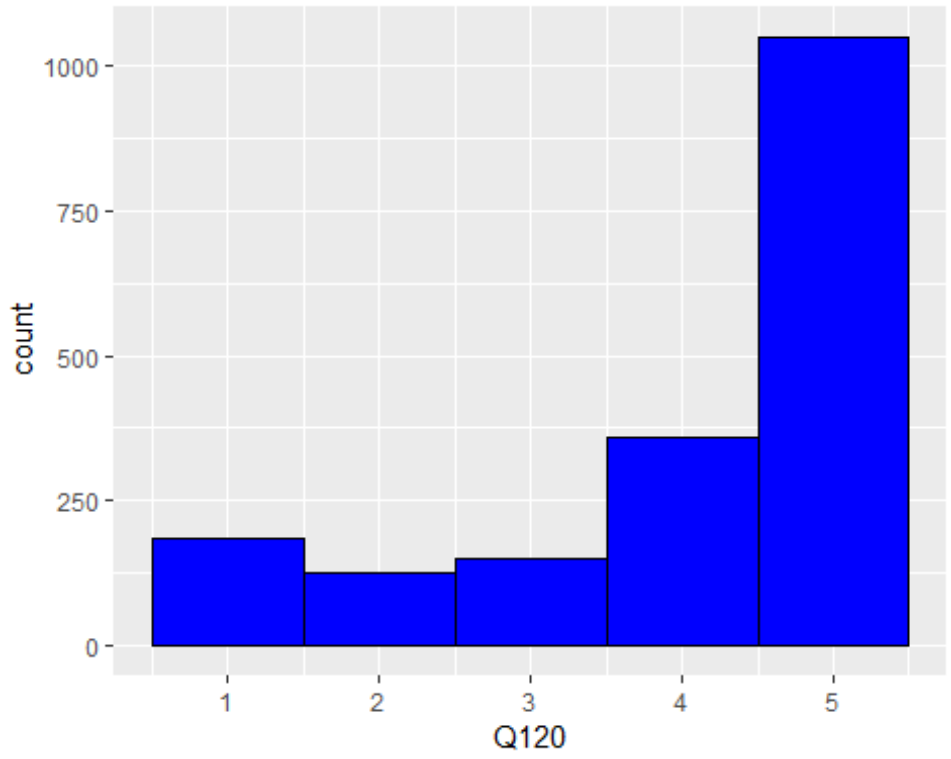

```
#$Q137
#""
#"Did you experience any issues during the boarding process?"
#"No"
```

Yes’

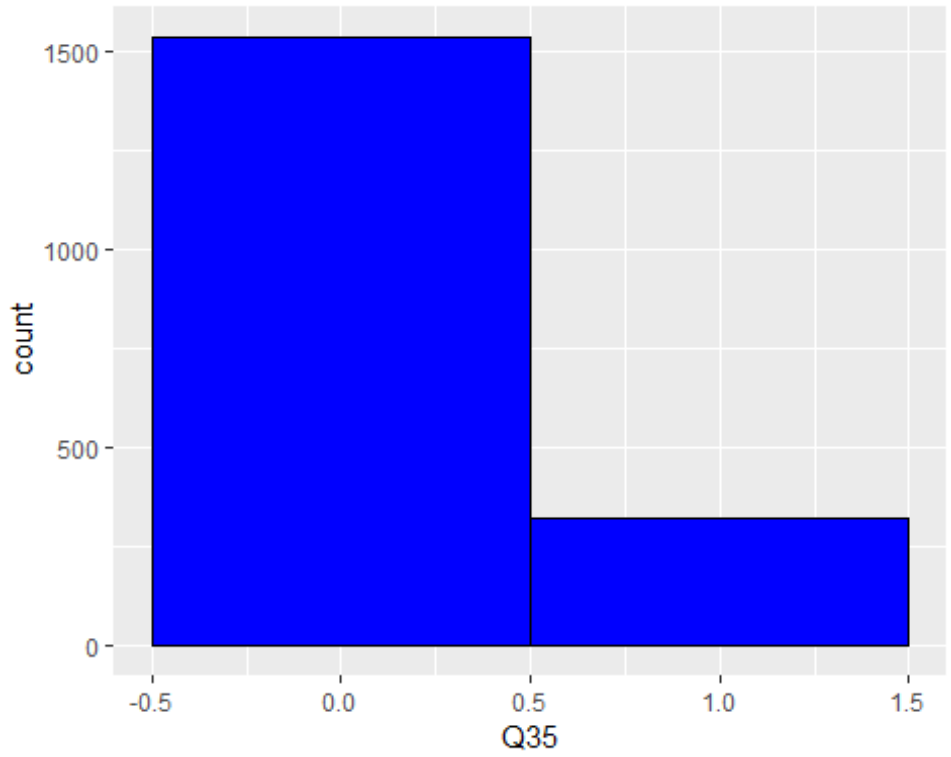
```
GuestData$Q137<-revalue(GuestData$Q137, c("Yes"="1", "No"="0"))
GuestData$Q137<-as.numeric(as.character(GuestData$Q137))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q137)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 6202 rows containing non-finite values (stat_bin).
```



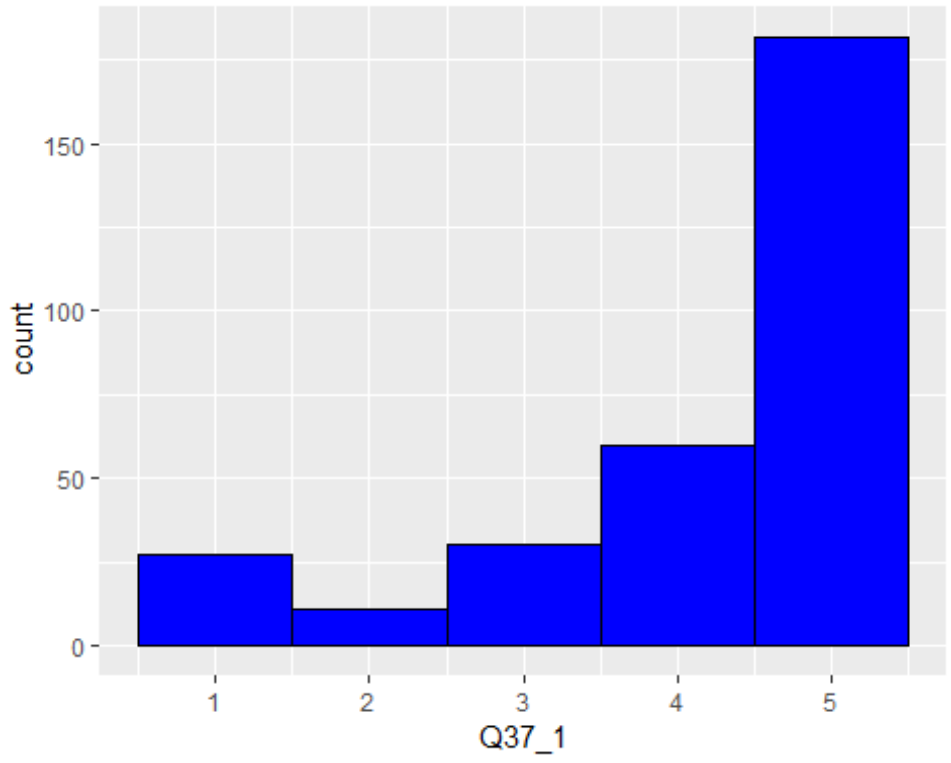

```
#$Q120
#""
#"Extremely dissatisfied"
#"Extremely satisfied"
#"How satisfied were you with your arrival experience?"
#"Neither satisfied nor dissatisfied"
#"Somewhat dissatisfied"
#"Somewhat satisfied"

GuestData$Q120<-revalue(GuestData$Q120, c("Extremely dissatisfied" = "1", "So
mewhat dissatisfied" = "2", "Neither satisfied nor dissatisfied" = "3", "Some
what satisfied" = "4", "Extremely satisfied" = "5"))
GuestData$Q120<-as.numeric(as.character(GuestData$Q120))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q120)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 6138 rows containing non-finite values (stat_bin).
```



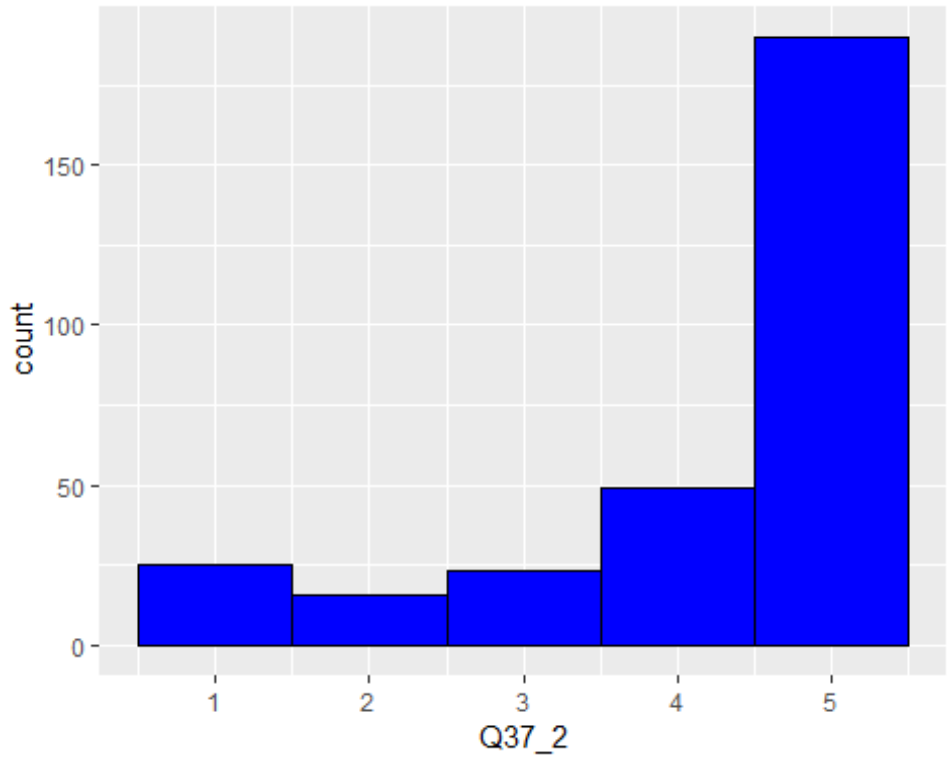

```
#$Q35
#""
#"Did you interact with\na gate agent after deplaning?"
#"No"
#"Yes"

GuestData$Q35<-revalue(GuestData$Q35, c("Yes"="1", "No"="0"))
GuestData$Q35<-as.numeric(as.character(GuestData$Q35))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q35)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 6141 rows containing non-finite values (stat_bin).
```



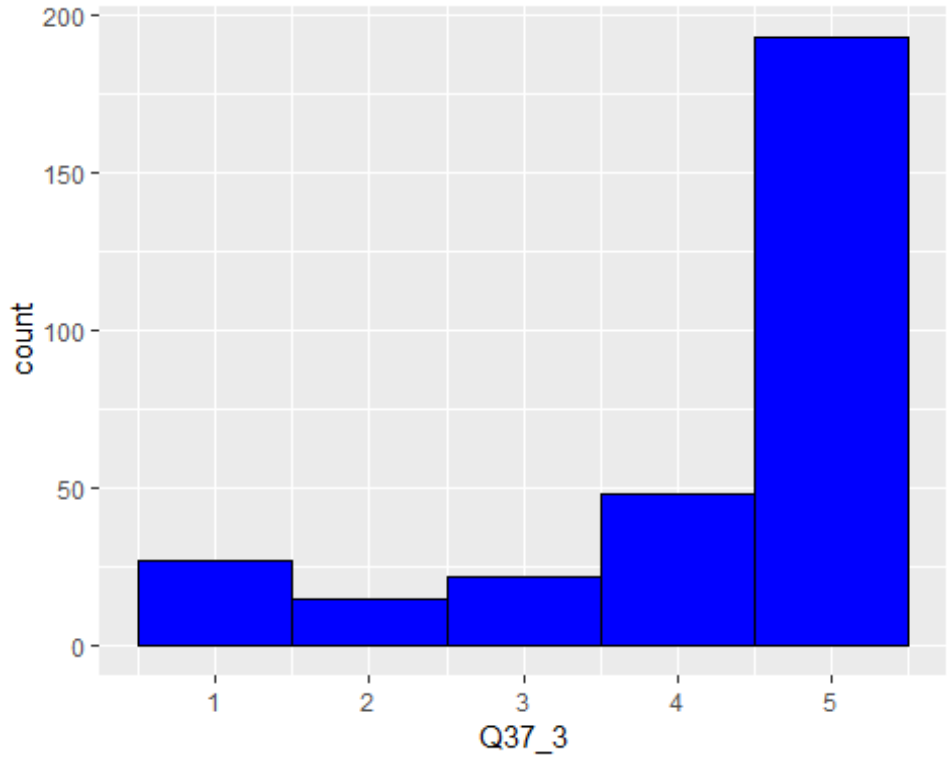

```
#$Q37_1
#""
#"Neither agree nor disagree"
#"Somewhat agree"
#"Somewhat disagree"
#"Strongly agree"
#"Strongly disagree"
#"The customer service\nagent at your arrival gate was... - Available"

GuestData$Q37_1<-revalue(GuestData$Q37_1, c("Strongly disagree"="1", "Somewha
t disagree"="2", "Neither agree nor disagree" = "3", "Somewhat agree" = "4",
"Strongly agree" = "5"))
GuestData$Q37_1<-as.numeric(as.character(GuestData$Q37_1))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q37_1)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7691 rows containing non-finite values (stat_bin).
```



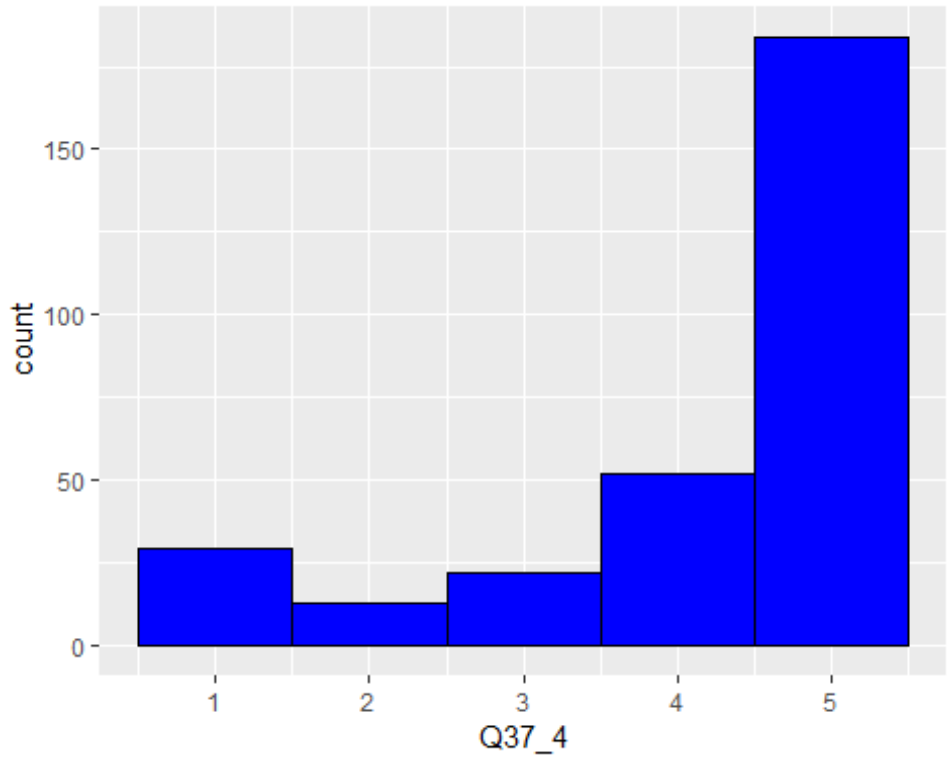

```
#$Q37_2
#""
#"Neither agree nor disagree"
#"Somewhat agree"
#"Somewhat disagree"
#"Strongly agree"
#"Strongly disagree"
#"The customer service\nagent at your arrival gate was... - Professional"

GuestData$Q37_2<-revalue(GuestData$Q37_2, c("Strongly disagree"="1", "Somewha
t disagree"="2", "Neither agree nor disagree" = "3", "Somewhat agree" = "4",
"Strongly agree" = "5"))
GuestData$Q37_2<-as.numeric(as.character(GuestData$Q37_2))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q37_2)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7698 rows containing non-finite values (stat_bin).
```



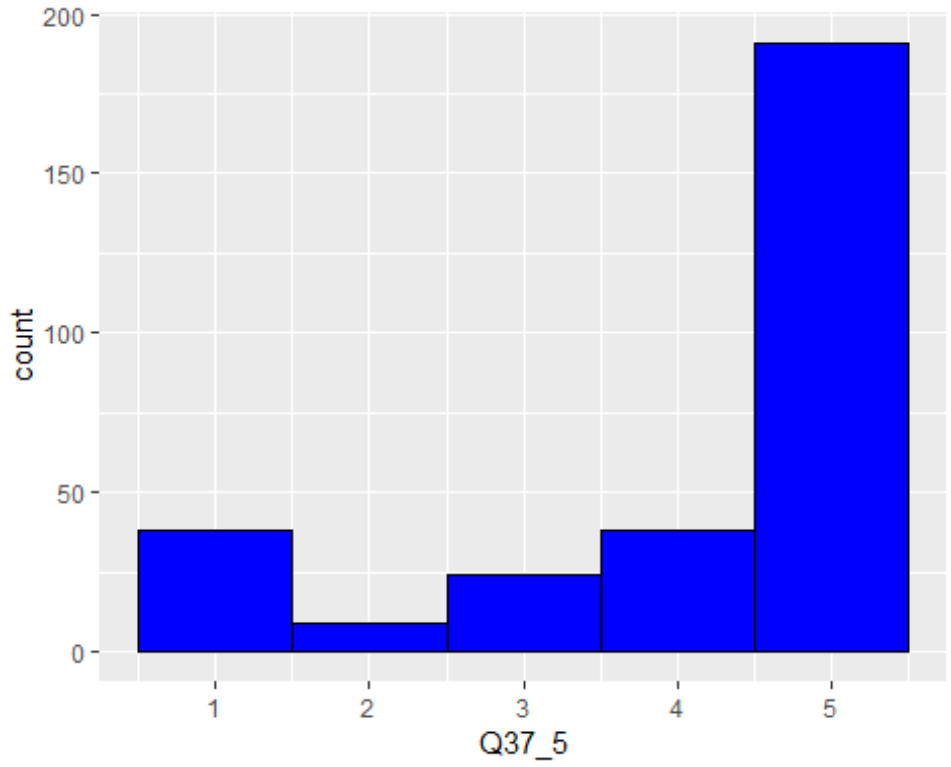

```
#$Q37_3
#""
#"Neither agree nor disagree"
#"Somewhat agree"
#"Somewhat disagree"
#"Strongly agree"
#"Strongly disagree"
#"The customer service\nagent at your arrival gate was... - Friendly"

GuestData$Q37_3<-revalue(GuestData$Q37_3, c("Strongly disagree"="1", "Somewha
t disagree"="2", "Neither agree nor disagree" = "3", "Somewhat agree" = "4",
"Strongly agree" = "5"))
GuestData$Q37_3<-as.numeric(as.character(GuestData$Q37_3))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q37_3)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7696 rows containing non-finite values (stat_bin).
```



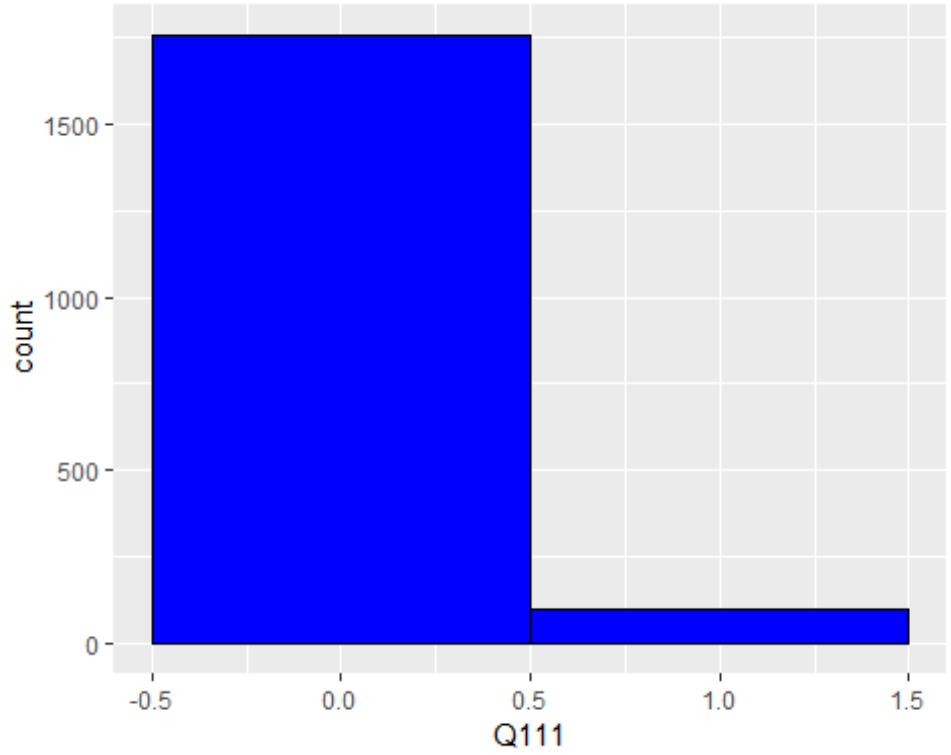

```
#$Q37_4
#""
#"Neither agree nor disagree"
#"Somewhat agree"
#"Somewhat disagree"
#"Strongly agree"
#"Strongly disagree"
#"The customer service\nagent at your arrival gate was... - Attentive"

GuestData$Q37_4<-revalue(GuestData$Q37_4, c("Strongly disagree"="1", "Somewha
t disagree"="2", "Neither agree nor disagree" = "3", "Somewhat agree" = "4",
"Strongly agree" = "5"))
GuestData$Q37_4<-as.numeric(as.character(GuestData$Q37_4))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q37_4)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7701 rows containing non-finite values (stat_bin).
```



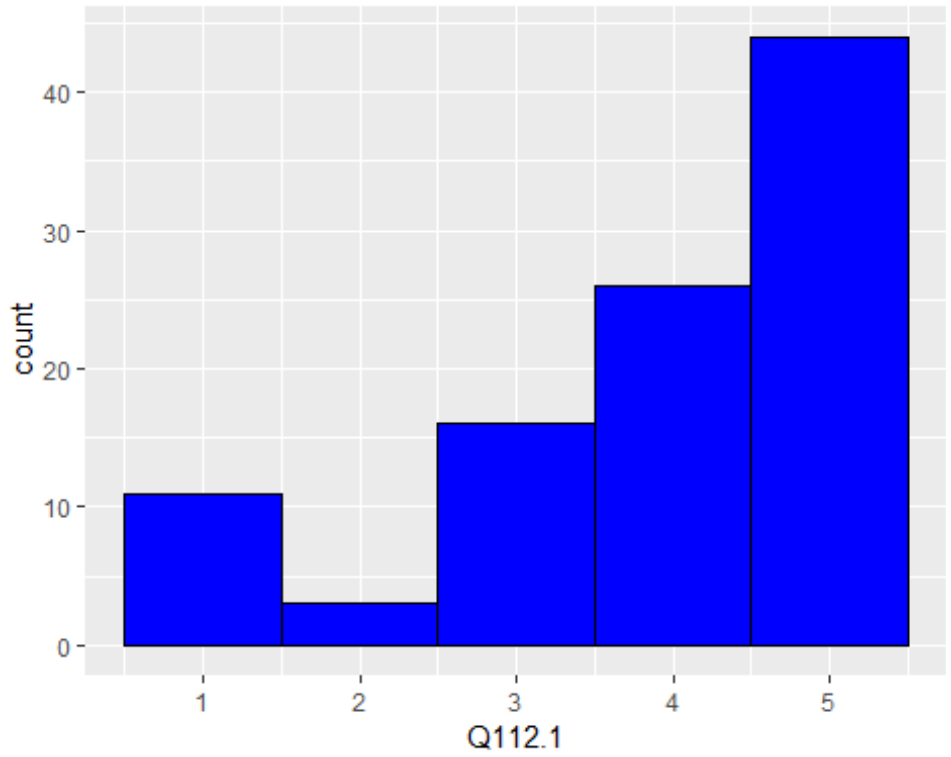

```
#$Q37_5
#""
#"Neither agree nor disagree"
#"Somewhat agree"
#"Somewhat disagree"
#"Strongly agree"
#"Strongly disagree"
#"The customer service\nagent at your arrival gate was... - Helpful"

GuestData$Q37_5<-revalue(GuestData$Q37_5, c("Strongly disagree"="1", "Somewha
t disagree"="2", "Neither agree nor disagree" = "3", "Somewhat agree" = "4",
"Strongly agree" = "5"))
GuestData$Q37_5<-as.numeric(as.character(GuestData$Q37_5))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q37_5)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7701 rows containing non-finite values (stat_bin).
```



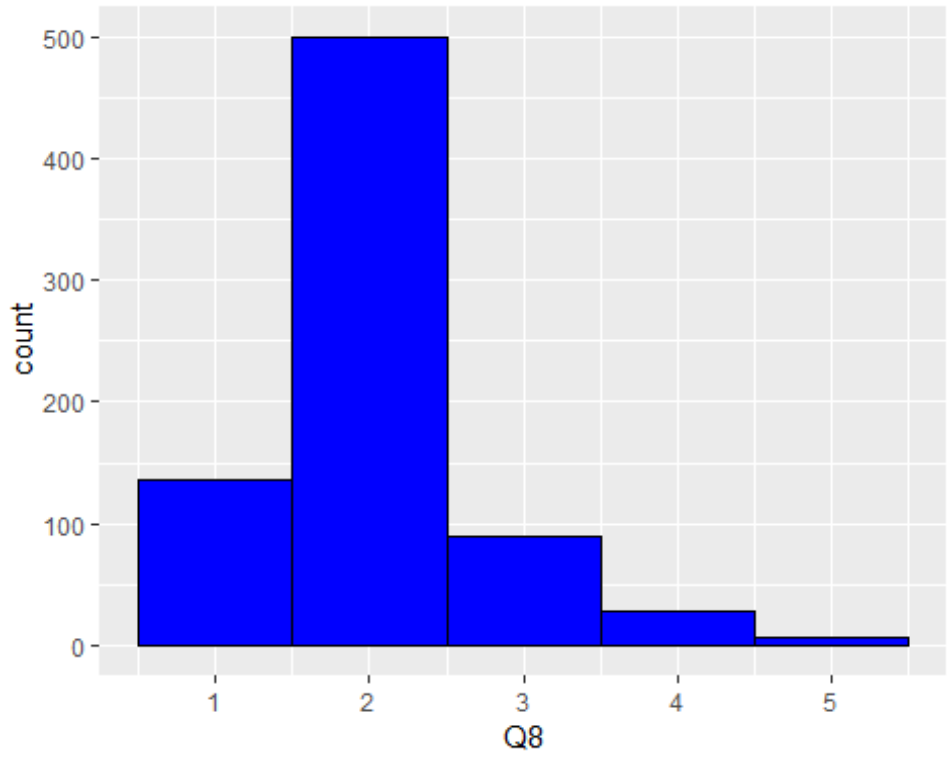

```
#$Q111
#""
#"Did you gate-check a stroller or wheelchair?"
#"No"
#"Yes"

GuestData$Q111<-revalue(GuestData$Q111, c("Yes"="1", "No"="0"))
GuestData$Q111<-as.numeric(as.character(GuestData$Q111))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q111)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 6143 rows containing non-finite values (stat_bin).
```



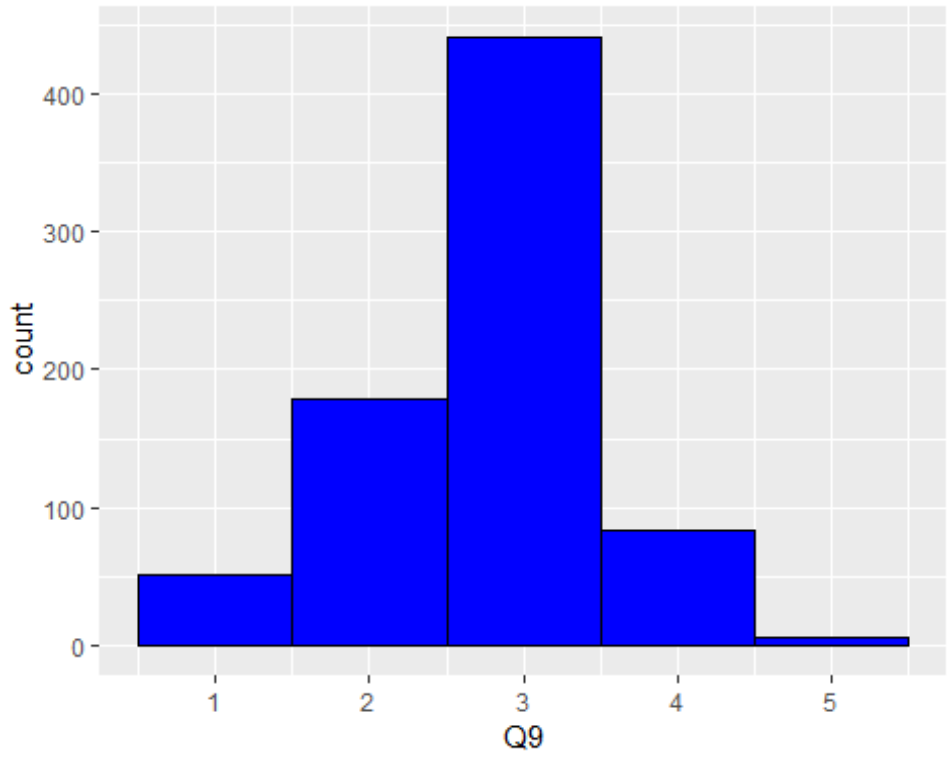

```
# $Q112.1
# ""
# "How quickly was your gate-checked stroller or wheelchair returned to you u
pon arrival to your destination?"
# "Neither quickly nor slowly"
# "Quickly"
# "Slowly"
# "Very quickly"
# "Very slowly"

GuestData$Q112.1<-revalue(GuestData$Q112.1, c("Very slowly"="1", "Slowly" = "
2", "Neither quickly nor slowly"="3", "Quickly" = "4", "Very quickly"="5"))
GuestData$Q112.1<-as.numeric(as.character(GuestData$Q112.1))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q112.1)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7901 rows containing non-finite values (stat_bin).
```



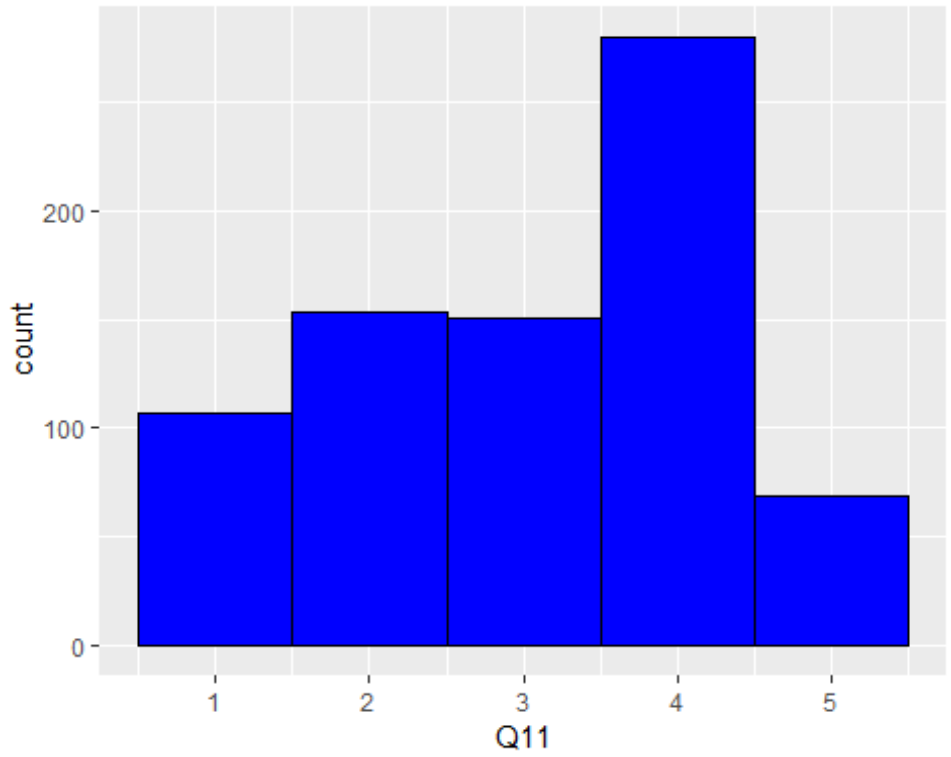

```
# $Q8
# ""
# "Clean"
# "Dirty"
# "Filthy"
# "How\nclean was the aircraft cabin upon boarding?"
# "Somewhat between clean and dirty"
# "Spotless"

GuestData$Q8<-revalue(GuestData$Q8, c("Spotless"="1", "Clean" = "2", "Somewha
t between clean and dirty"="3", "Dirty" = "4", "Filthy"="5"))
GuestData$Q8<-as.numeric(as.character(GuestData$Q8))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q8)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7240 rows containing non-finite values (stat_bin).
```



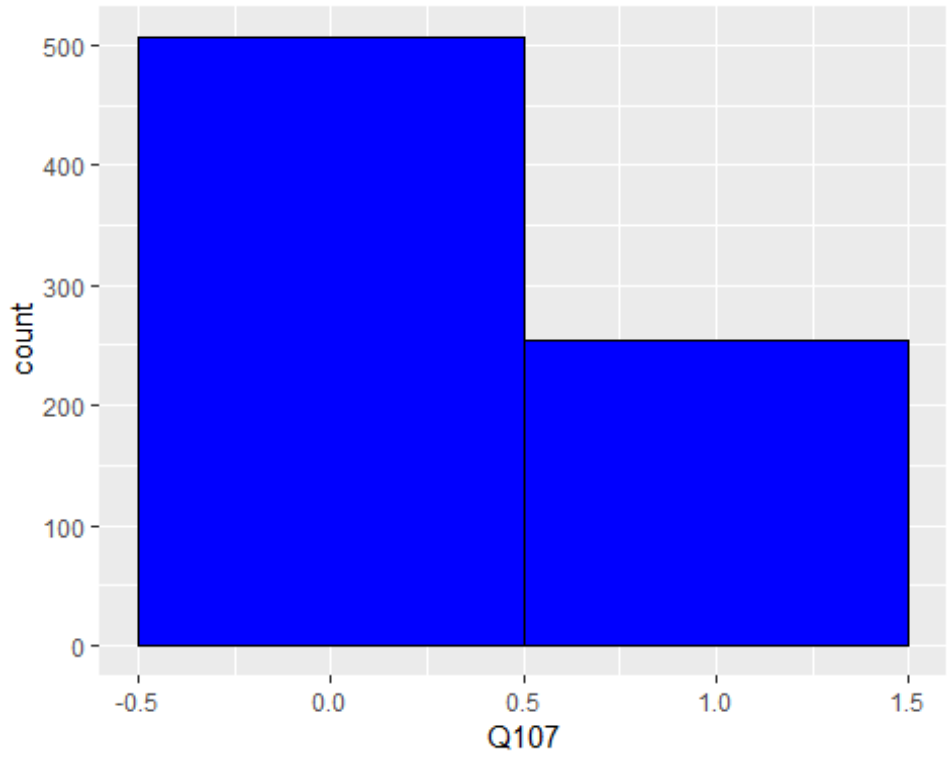

```
# $Q9
# ""
# "Chilly"
# "Extremely Cold"
# "Extremely Hot"
# "How\nwas the cabin temperature on the aircraft?"
# "Perfect"
# "Warm"

GuestData$Q9<-revalue(GuestData$Q9, c("Extremely Cold"="1", "Chilly" = "2", "
Perfect"="3", "Warm" = "4", "Extremely Hot"="5"))
GuestData$Q9<-as.numeric(as.character(GuestData$Q9))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q9)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7242 rows containing non-finite values (stat_bin).
```



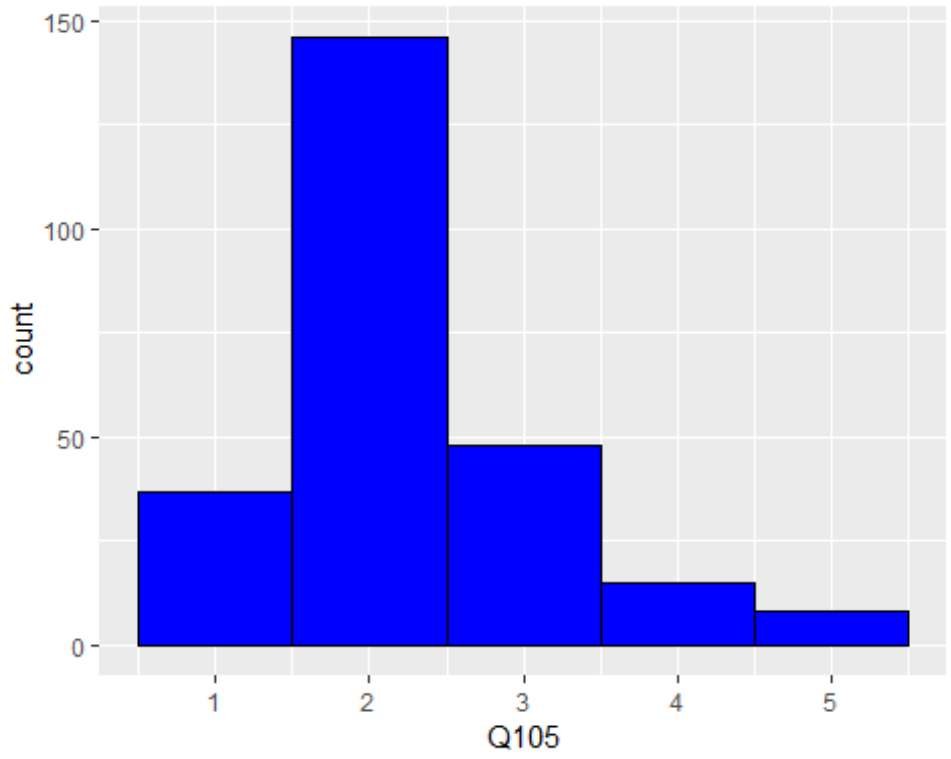

```
# $Q11
# ""
# "Extremely comfortable"
# "Extremely uncomfortable"
# "How comfortable were the\nseats?"
# "Neither comfortable nor uncomfortable"
# "Somewhat comfortable"
# "Somewhat uncomfortable"

GuestData$Q11<-revalue(GuestData$Q11, c("Extremely uncomfortable" ="1", "Some
what uncomfortable" = "2", "Neither comfortable nor uncomfortable"="3",
                                        "Somewhat comfortable" = "4", "Extrem
ely comfortable"="5"))

GuestData$Q11<-as.numeric(as.character(GuestData$Q11))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q11)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7241 rows containing non-finite values (stat_bin).
```



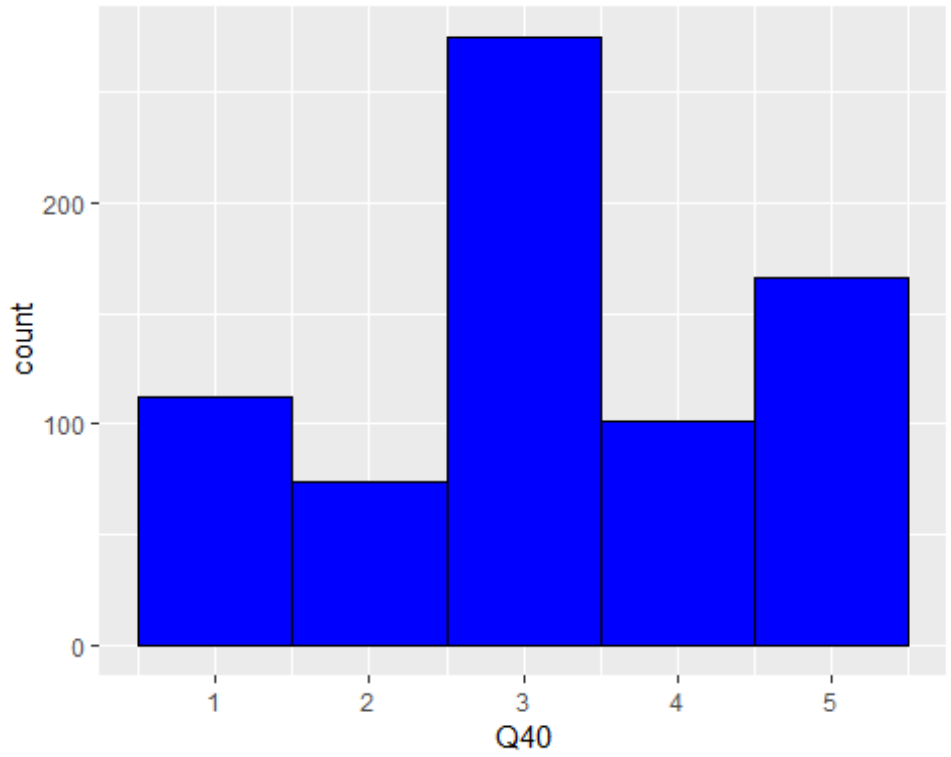

```
# $Q107
# ""
# "Did you use the onboard lavatory?"
# "No"
# "Yes"

GuestData$Q107<-revalue(GuestData$Q107, c("Yes"="1", "No"="0"))
GuestData$Q107<-as.numeric(as.character(GuestData$Q107))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q107)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7240 rows containing non-finite values (stat_bin).
```



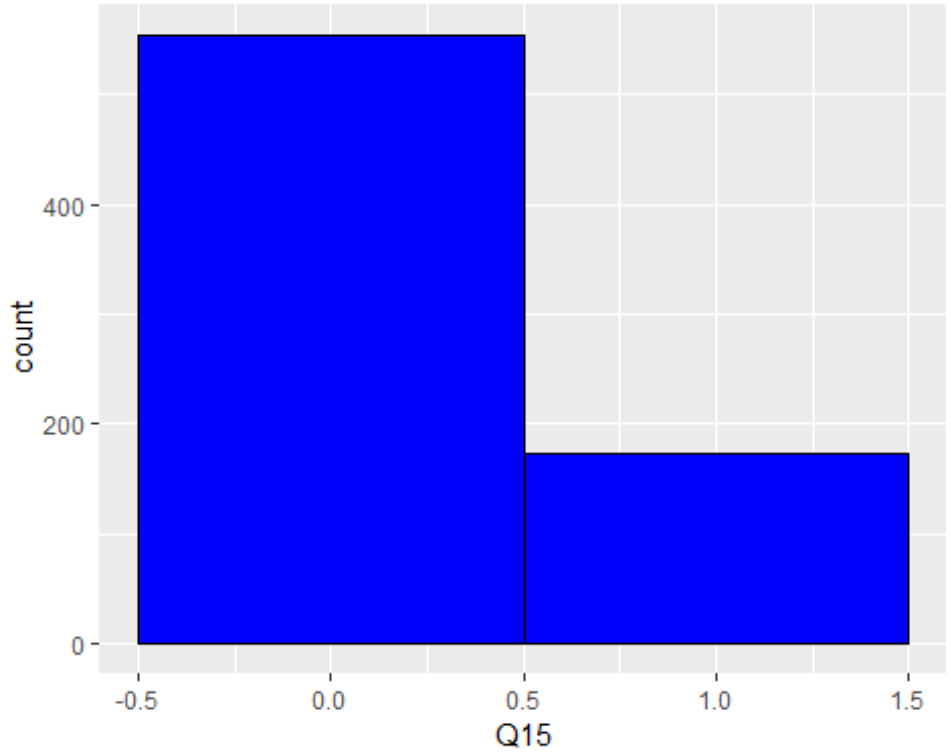

```
# $Q105
# ""
# "Clean"
# "Dirty"
# "Filthy"
# "How clean was the lavatory onboard your flight?"
# "Somewhat between clean and dirty"
# "Spotless"

GuestData$Q105<-revalue(GuestData$Q105, c("Spotless"="1", "Clean" = "2", "Som
ewhat between clean and dirty"="3", "Dirty" = "4", "Filthy"="5"))
GuestData$Q105<-as.numeric(as.character(GuestData$Q105))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q105)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7747 rows containing non-finite values (stat_bin).
```



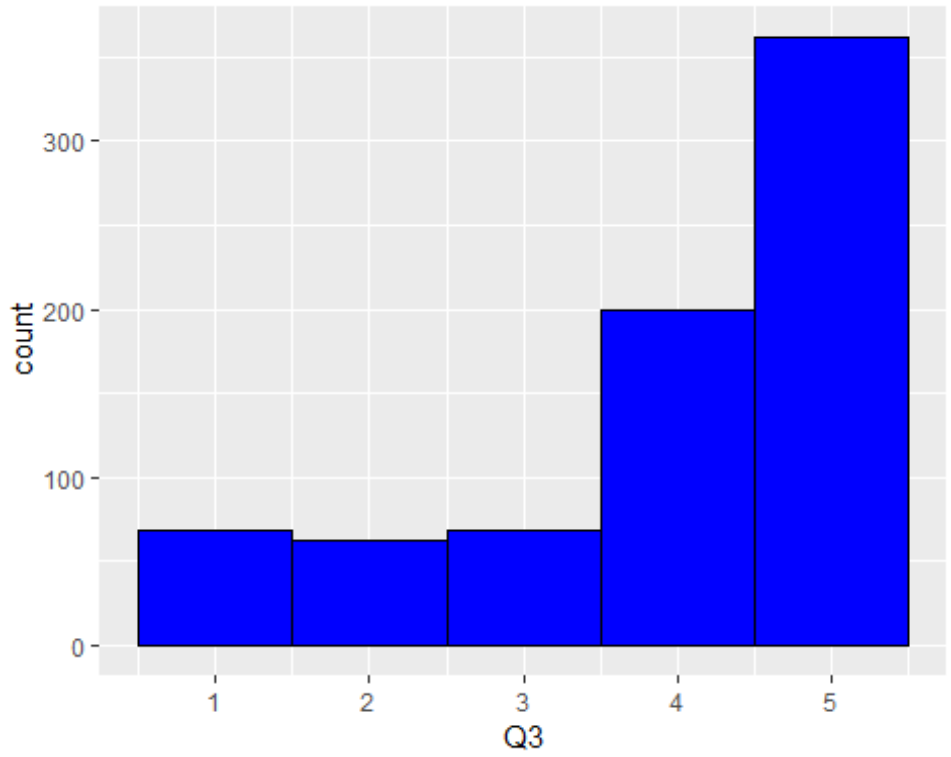

```
# $Q40
# ""
# "Extremely dissatisfied"
# "Extremely satisfied"
# "Neither satisfied nor dissatisfied"
# "Overall, how satisfied are you with AirlineX's onboard food and beverage e
xperience?"
# "Somewhat dissatisfied"
# "Somewhat satisfied"

GuestData$Q40<-revalue(GuestData$Q40, c("Extremely dissatisfied" = "1", "Some
what dissatisfied" = "2", "Neither satisfied nor dissatisfied" = "3", "Somewh
at satisfied" = "4", "Extremely satisfied" = "5"))
GuestData$Q40<-as.numeric(as.character(GuestData$Q40))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q40)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7273 rows containing non-finite values (stat_bin).
```



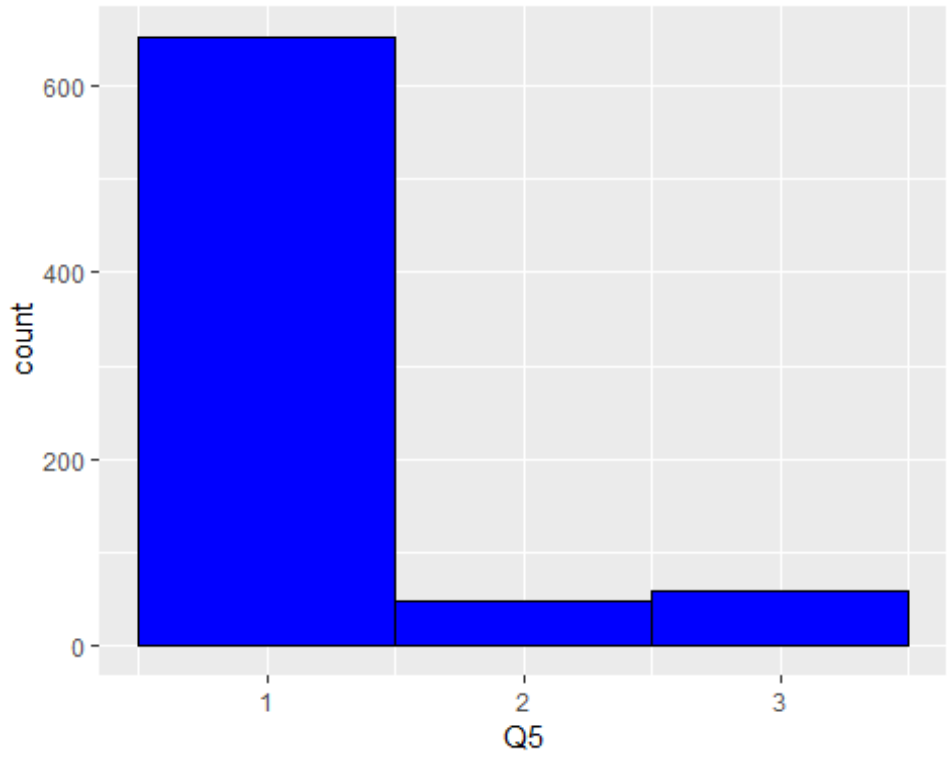

```
# $Q15
# ""
# "Did you make a food or beverage purchase onboard during this flight?"
# "No"
# "Yes"

GuestData$Q15<-revalue(GuestData$Q15, c("Yes"="1", "No"="0"))
GuestData$Q15<-as.numeric(as.character(GuestData$Q15))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q15)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7274 rows containing non-finite values (stat_bin).
```



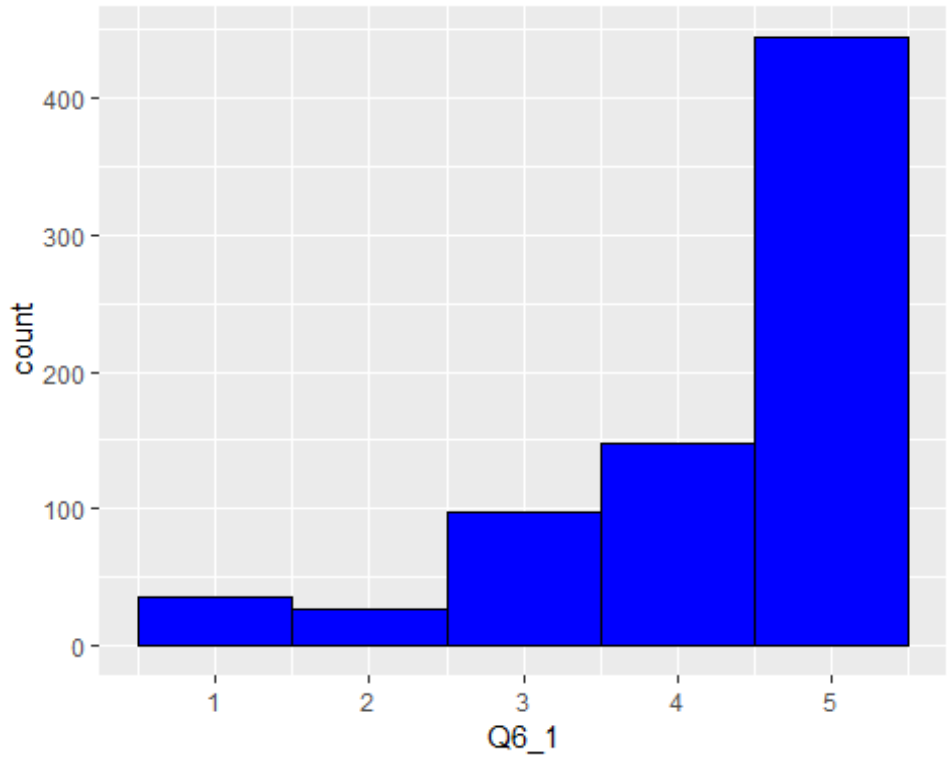

```
# $Q3
# ""
# "Extremely dissatisfied"
# "Extremely satisfied"
# "How satisfied were you with your in-flight experience?"
# "Neither satisfied nor dissatisfied"
# "Somewhat dissatisfied"
# "Somewhat satisfied"

GuestData$Q3<-revalue(GuestData$Q3, c("Extremely dissatisfied" = "1", "Somewh
at dissatisfied" = "2", "Neither satisfied nor dissatisfied" = "3", "Somewhat
satisfied" = "4", "Extremely satisfied" = "5"))
GuestData$Q3<-as.numeric(as.character(GuestData$Q3))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q3)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7241 rows containing non-finite values (stat_bin).
```



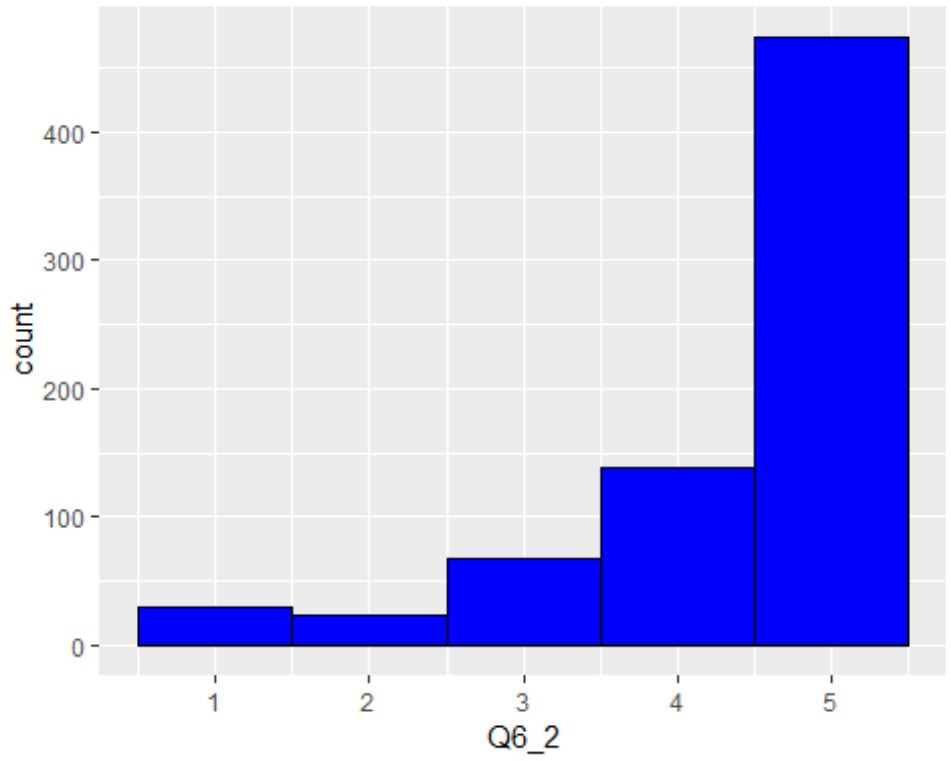

```
# $Q5
# ""
# "I don't remember"
# "No"
# "Were you greeted by the\nflight attendants as you boarded the aircraft?"
# "Yes"

GuestData$Q5<-revalue(GuestData$Q5, c("Yes"="1", "I don't remember" = "2", "N
o"="3"))
GuestData$Q5<-as.numeric(as.character(GuestData$Q5))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q5)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7243 rows containing non-finite values (stat_bin).
```



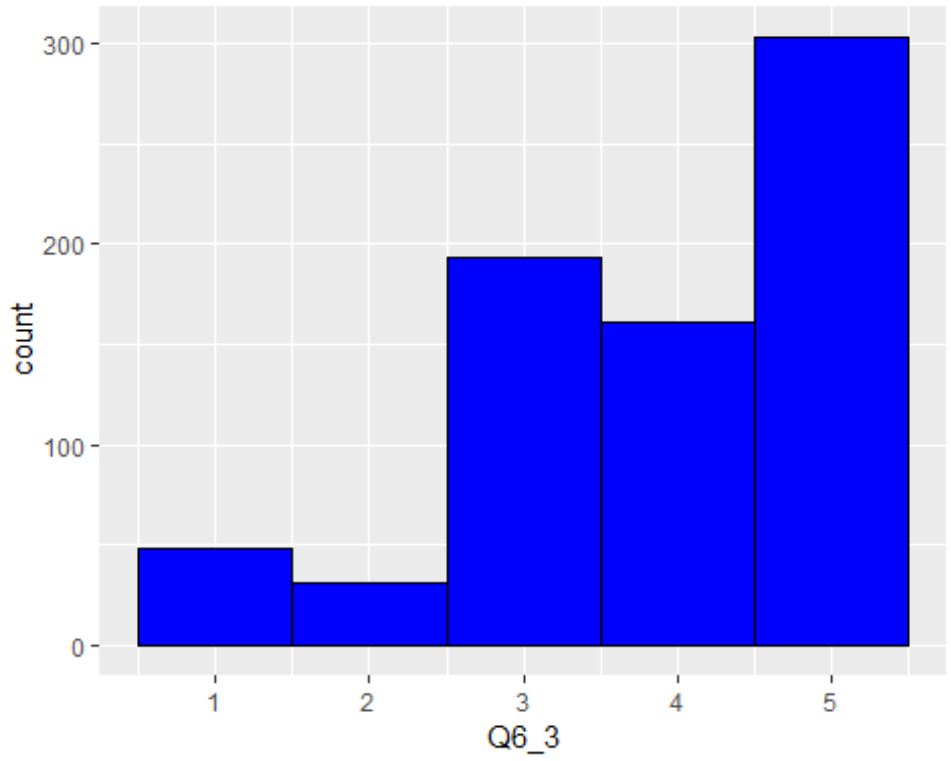

```
# $Q6_1
# ""
# "Neither agree nor disagree"
# "Please\nrate the following statements: \n\n\nThe flight attendants on your
flight from [Field-Journey%20Origin%20Name] to\n[Field-Journey%20Destination%
20Name] were... - Attentive"
# "Somewhat agree"
# "Somewhat disagree"
# "Strongly agree"
# "Strongly disagree"

GuestData$Q6_1<-revalue(GuestData$Q6_1, c("Strongly disagree"="1", "Somewhat
disagree"="2", "Neither agree nor disagree" = "3", "Somewhat agree" = "4", "S
trongly agree" = "5"))
GuestData$Q6_1<-as.numeric(as.character(GuestData$Q6_1))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q6_1)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7249 rows containing non-finite values (stat_bin).
```



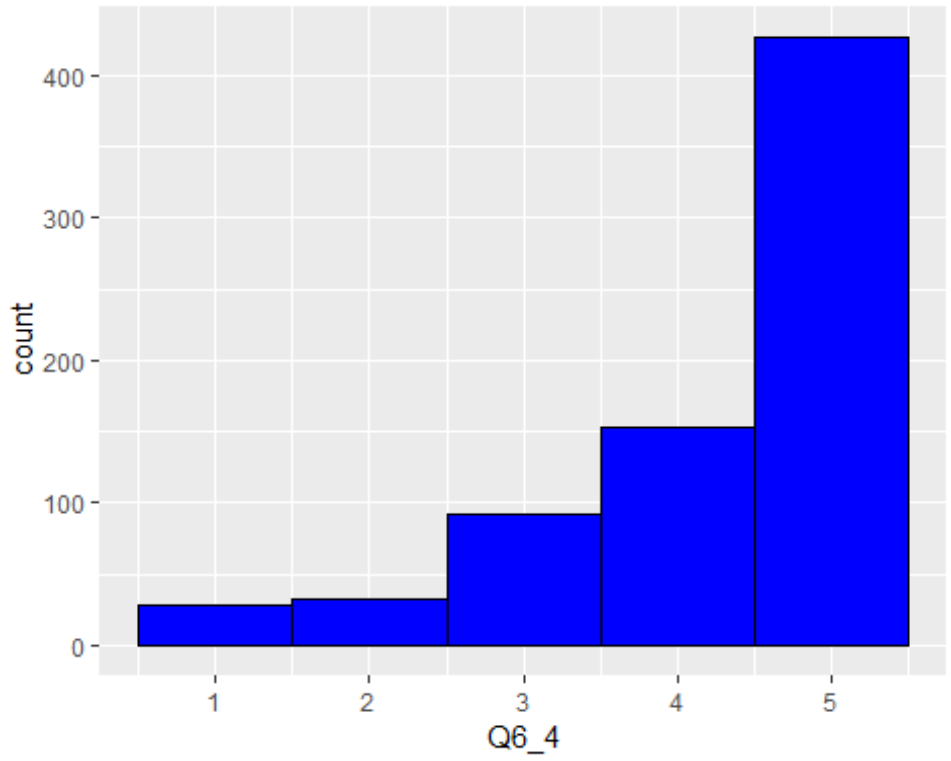

```
# $Q6_2
# ""
# "Neither agree nor disagree"
# "Please\nrate the following statements: \n\n\nThe flight attendants on your
flight from [Field-Journey%20Origin%20Name] to\n[Field-Journey%20Destination%
20Name] were... - Professional"
# "Somewhat agree"
# "Somewhat disagree"
# "Strongly agree"
# "Strongly disagree"

GuestData$Q6_2<-revalue(GuestData$Q6_2, c("Strongly disagree"="1", "Somewhat
disagree"="2", "Neither agree nor disagree" = "3", "Somewhat agree" = "4", "S
trongly agree" = "5"))
GuestData$Q6_2<-as.numeric(as.character(GuestData$Q6_2))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q6_2)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7268 rows containing non-finite values (stat_bin).
```



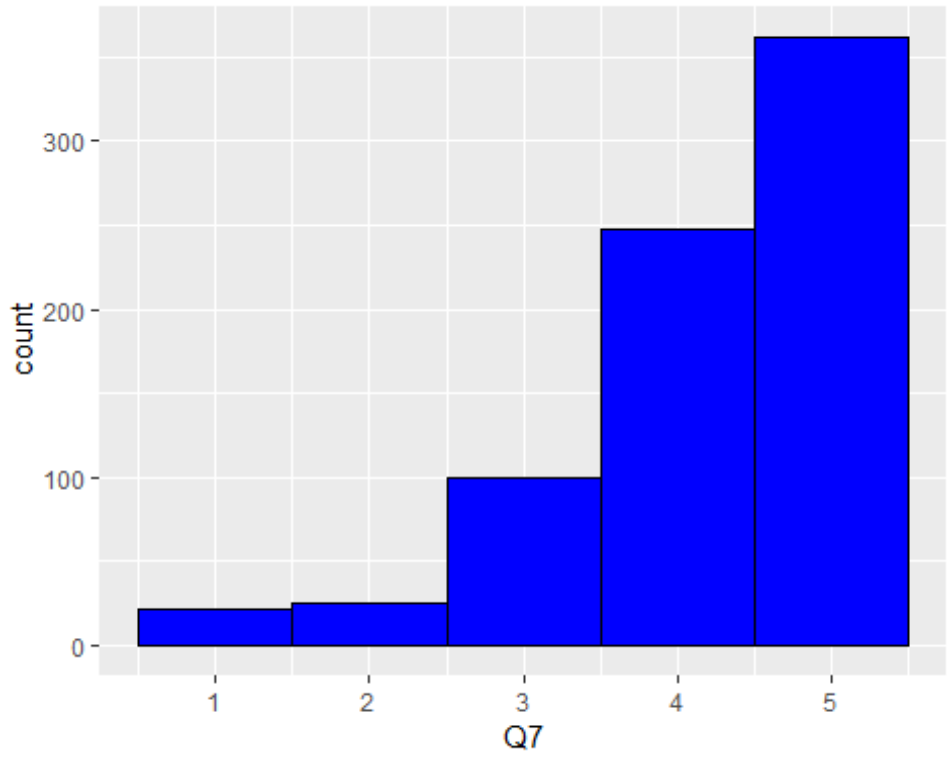

```
# $Q6_3
# ""
# "Neither agree nor disagree"
# "Please\nrate the following statements: \n\n\nThe flight attendants on your
flight from [Field-Journey%20Origin%20Name] to\n[Field-Journey%20Destination%
20Name] were... - Fun"
# "Somewhat agree"
# "Somewhat disagree"
# "Strongly agree"
# "Strongly disagree"

GuestData$Q6_3<-revalue(GuestData$Q6_3, c("Strongly disagree"="1", "Somewhat
disagree"="2", "Neither agree nor disagree" = "3", "Somewhat agree" = "4", "S
trongly agree" = "5"))
GuestData$Q6_3<-as.numeric(as.character(GuestData$Q6_3))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q6_3)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7265 rows containing non-finite values (stat_bin).
```



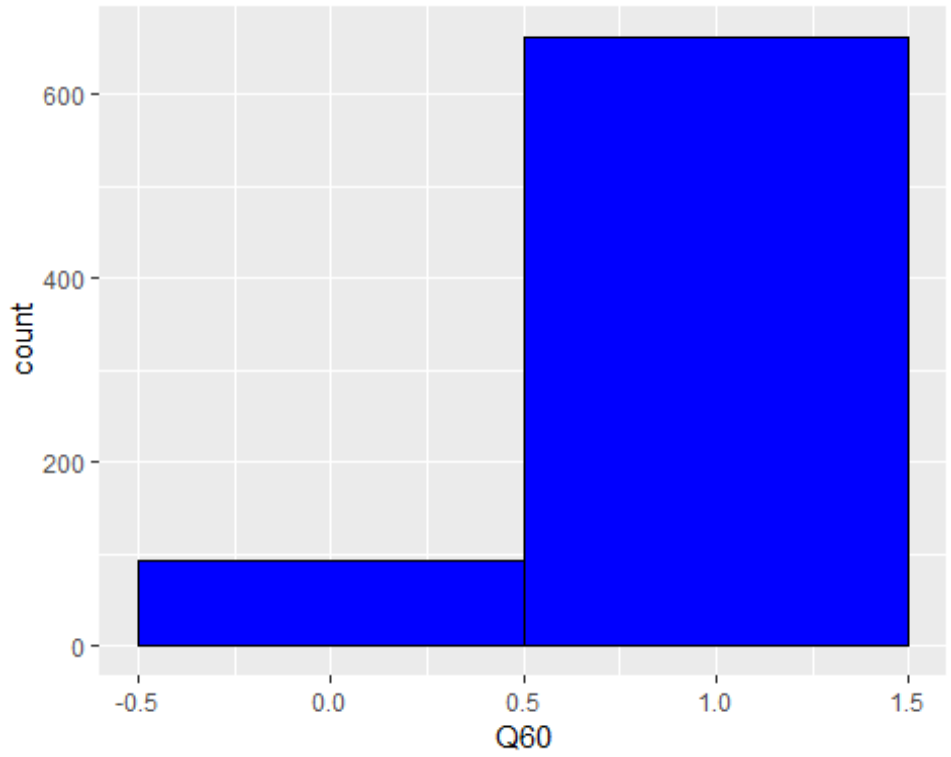

```
# $Q6_4
# ""
# "Neither agree nor disagree"
# "Please\nrate the following statements: \n\n\nThe flight attendants on your
flight from [Field-Journey%20Origin%20Name] to\n[Field-Journey%20Destination%
20Name] were... - Friendly"
# "Somewhat agree"
# "Somewhat disagree"
# "Strongly agree"
# "Strongly disagree"

GuestData$Q6_4<-revalue(GuestData$Q6_4, c("Strongly disagree"="1", "Somewhat
disagree"="2", "Neither agree nor disagree" = "3", "Somewhat agree" = "4", "S
trongly agree" = "5"))
GuestData$Q6_4<-as.numeric(as.character(GuestData$Q6_4))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q6_4)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7268 rows containing non-finite values (stat_bin).
```



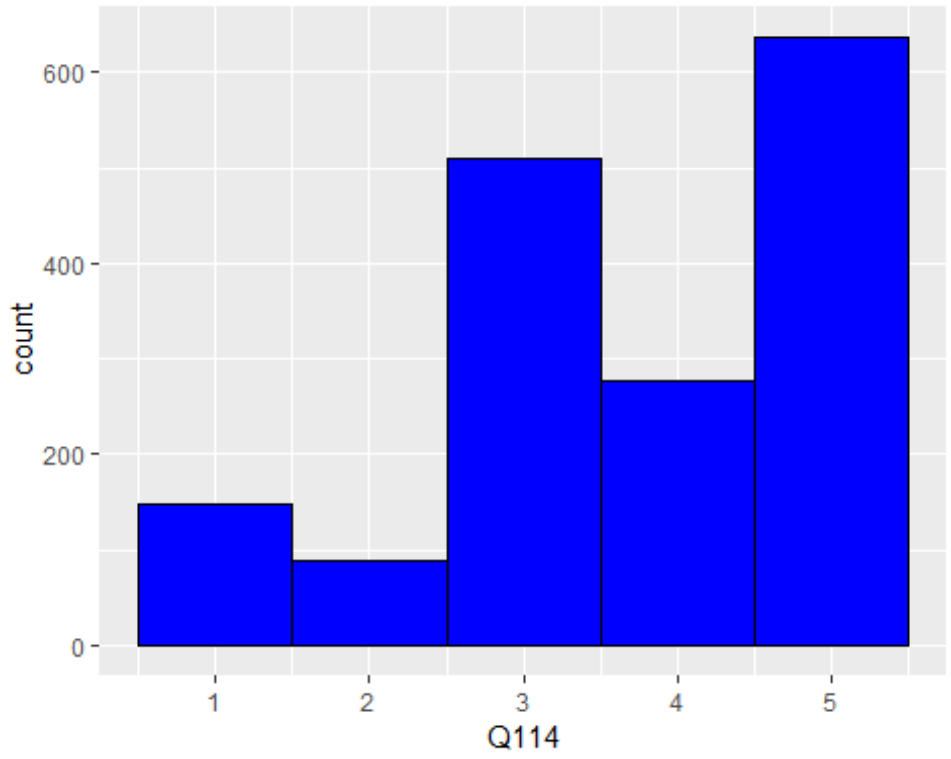

```
# $Q7
# ""
# "Extremely well"
# "How well were you kept\nup-to-date with information updates from the pilot
s and flight attendants?"
# "Moderately well"
# "Not well at all"
# "Slightly well"
# "Very well"

GuestData$Q7<-revalue(GuestData$Q7, c("Not well at all" ="1", "Slightly well"
="2", "Moderately well"="3", "Very well" ="4", "Extremely well" ="5"))
GuestData$Q7<-as.numeric(as.character(GuestData$Q7))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q7)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7245 rows containing non-finite values (stat_bin).
```



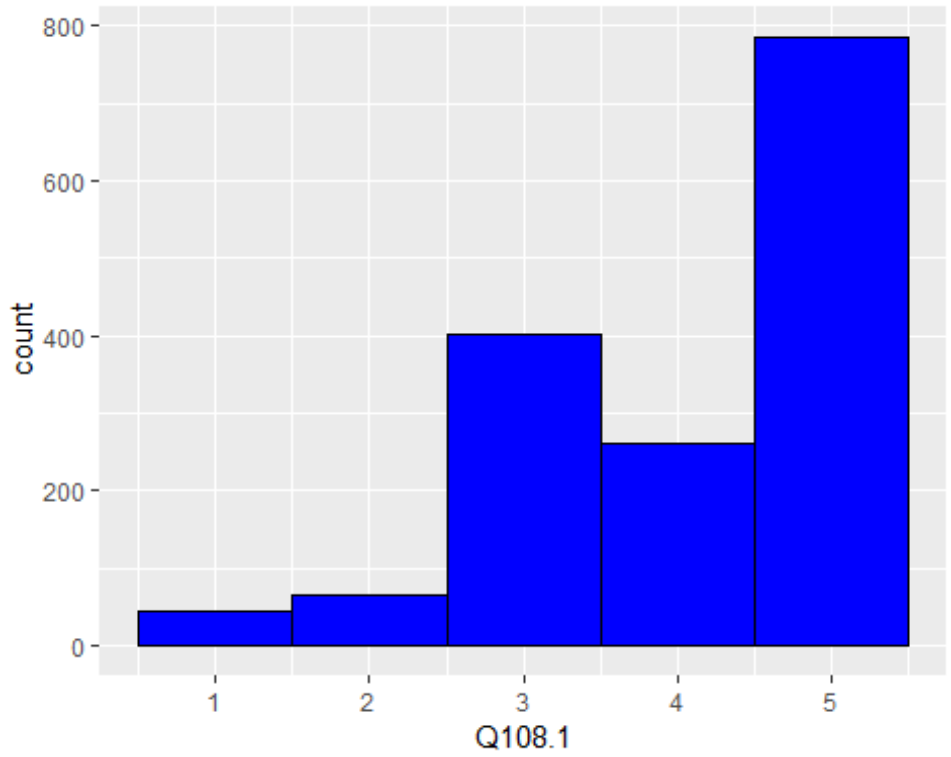

```
# $Q60
# ""
# "Did\nthe flight crew thank you for flying AirlineX as you deplaned?"
# "No"
# "Yes"

GuestData$Q60<-revalue(GuestData$Q60, c("Yes"="1", "No"="0"))
GuestData$Q60<-as.numeric(as.character(GuestData$Q60))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q60)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 7245 rows containing non-finite values (stat_bin).
```



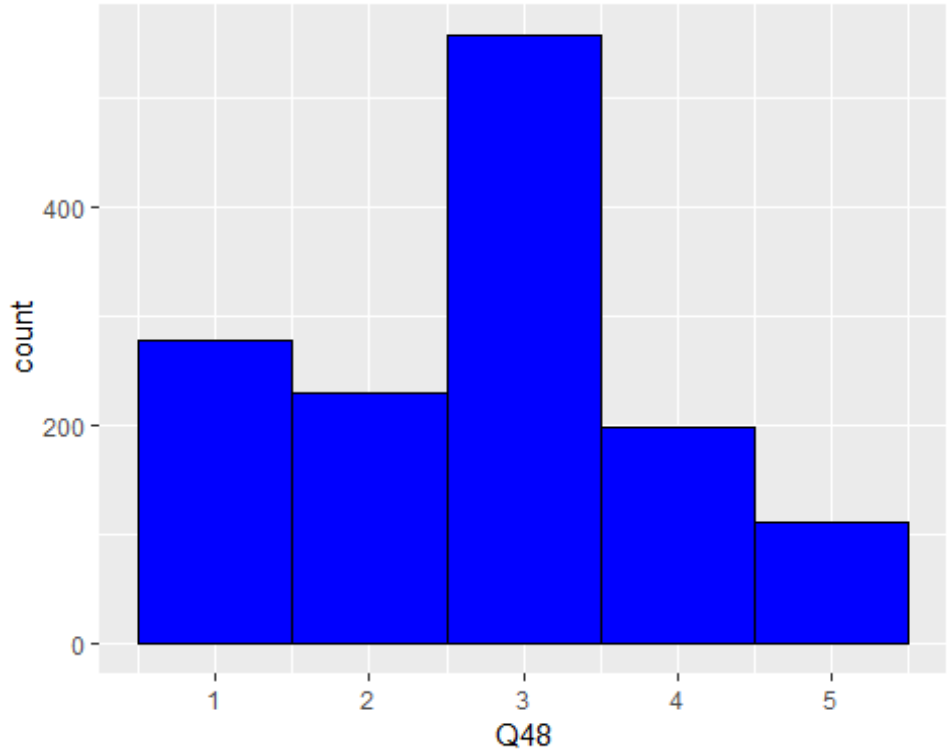

```
# $Q114
# ""
# "Extremely dissatisfied"
# "Extremely satisfied"
# "How satisfied were you with your baggage claim experience?"
# "Neither satisfied nor dissatisfied"
# "Somewhat dissatisfied"
# "Somewhat satisfied"

GuestData$Q114<-revalue(GuestData$Q114, c("Extremely dissatisfied" = "1", "So
mewhat dissatisfied" = "2", "Neither satisfied nor dissatisfied" = "3", "Some
what satisfied" = "4", "Extremely satisfied" = "5"))
GuestData$Q114<-as.numeric(as.character(GuestData$Q114))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q114)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 6341 rows containing non-finite values (stat_bin).
```



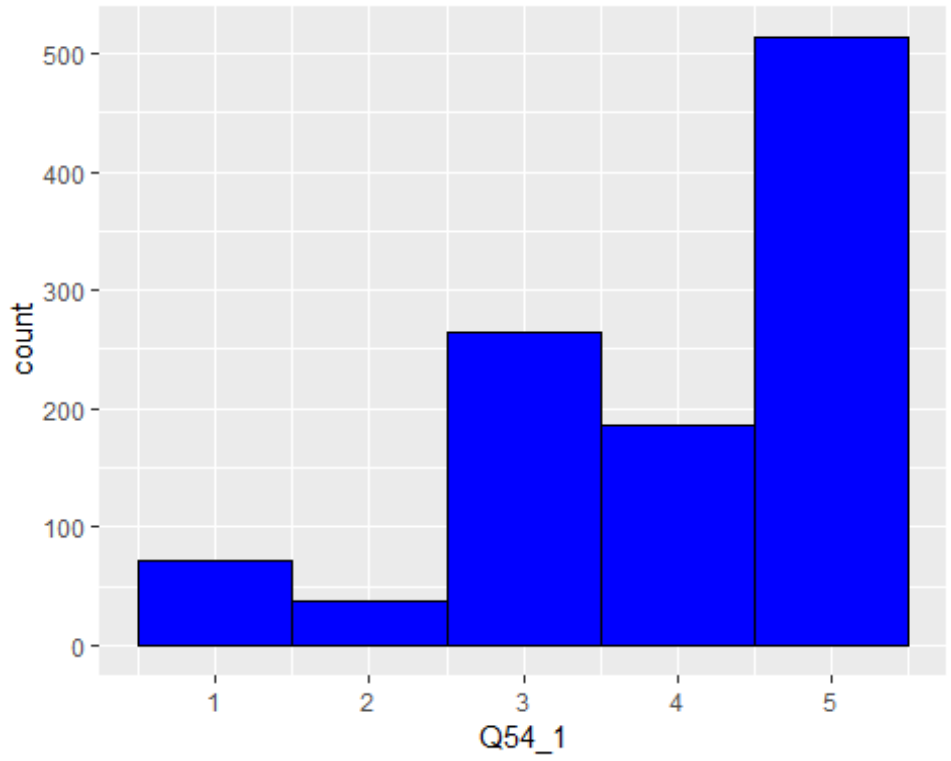

```
# $Q108.1
# ""
# "Extremely difficult"
# "Extremely easy"
# "How easy was it to obtain baggage claim information?"
# "Neither easy nor difficult"
# "Somewhat difficult"
# "Somewhat easy"

GuestData$Q108.1<-revalue(GuestData$Q108.1, c("Extremely difficult"="1", "Som
ewhat difficult"="2", "Neither easy nor difficult"="3",
                                              "Somewhat easy"="4", "Extremely
easy"="5"))
GuestData$Q108.1<-as.numeric(as.character(GuestData$Q108.1))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q108.1)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 6444 rows containing non-finite values (stat_bin).
```



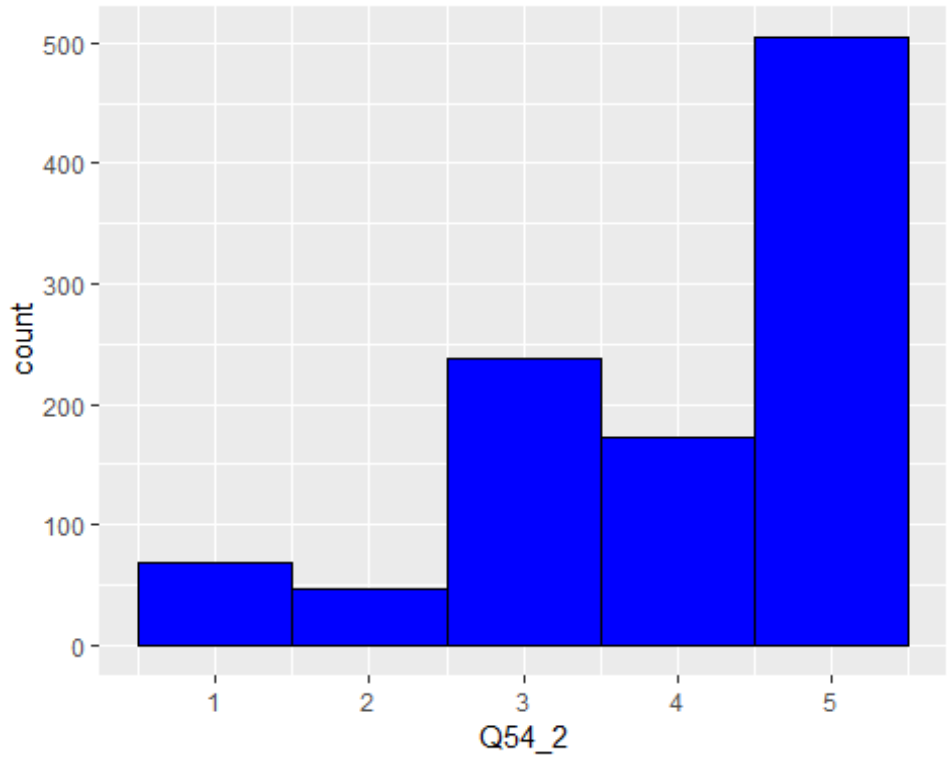

```
# $Q48
# ""
# "How long was the wait\nto retrieve your checked baggage?"
# "Long"
# "Normal"
# "Short"
# "Very long"
# "Very short"

GuestData$Q48<-revalue(GuestData$Q48, c("Very short"="1", "Short"="2", "Norma
l"="3", "Long"="4", "Very long"="5"))
GuestData$Q48<-as.numeric(as.character(GuestData$Q48))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q48)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 6627 rows containing non-finite values (stat_bin).
```



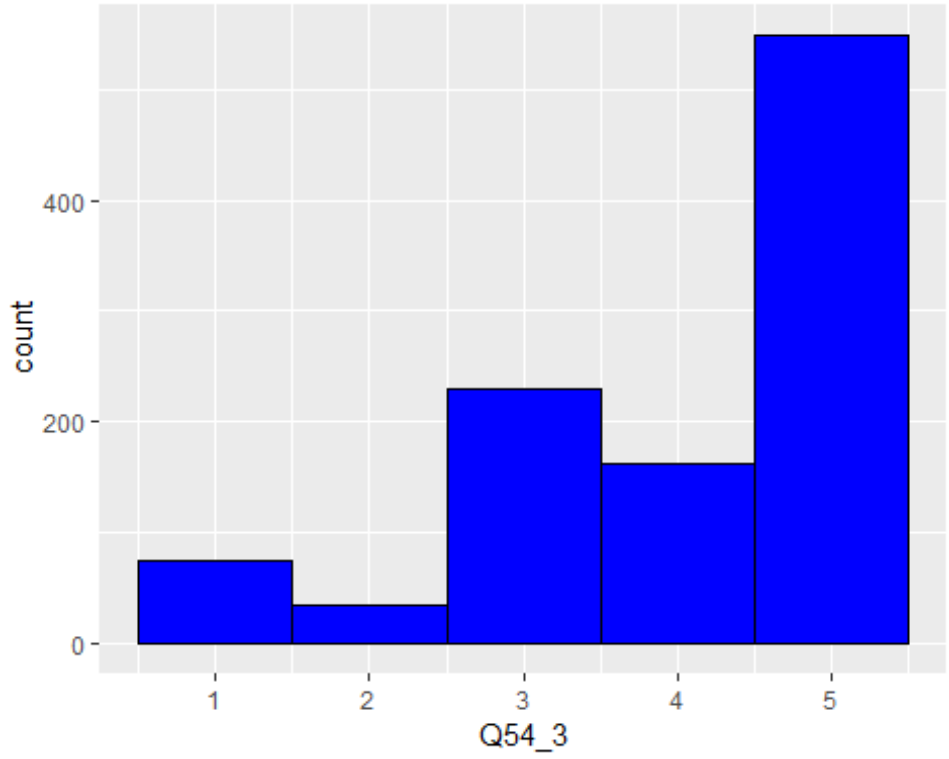

```
# $Q54_1
# ""
# "Neither agree nor disagree"
# "Somewhat agree"
# "Somewhat disagree"
# "Strongly agree"
# "Strongly disagree"
# "The customer service agent\nyou spoke with was: - Available"

GuestData$Q54_1<-revalue(GuestData$Q54_1, c("Strongly disagree"="1", "Somewha
t disagree"="2", "Neither agree nor disagree" = "3", "Somewhat agree" = "4",
"Strongly agree" = "5"))
GuestData$Q54_1<-as.numeric(as.character(GuestData$Q54_1))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q54_1)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 6927 rows containing non-finite values (stat_bin).
```



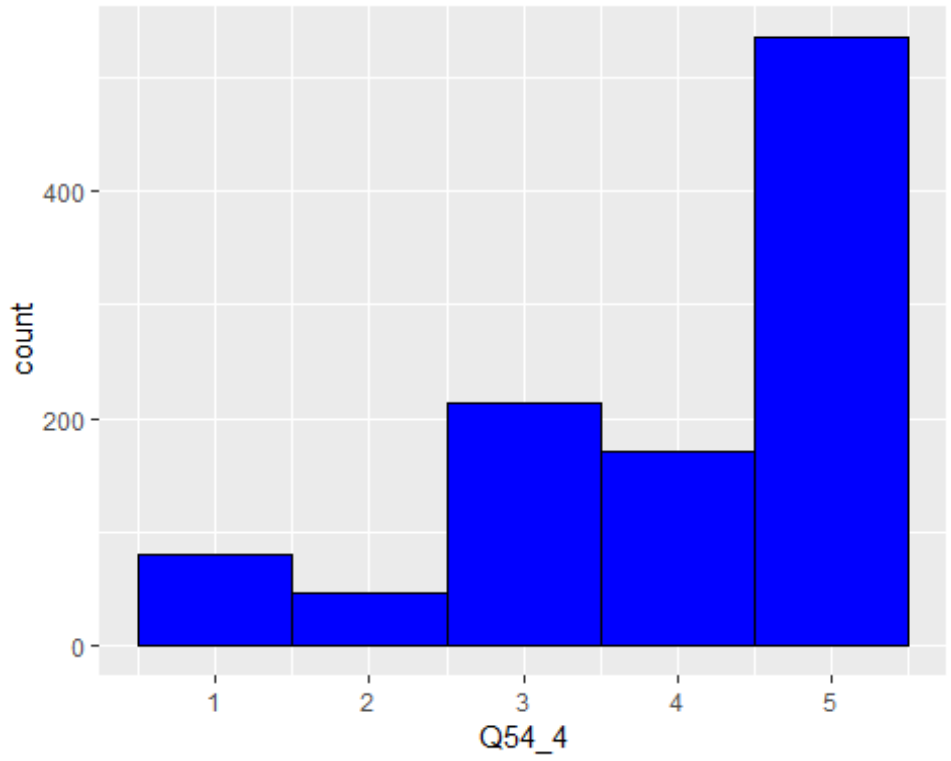

```
# $Q54_2
# ""
# "Neither agree nor disagree"
# "Somewhat agree"
# "Somewhat disagree"
# "Strongly agree"
# "Strongly disagree"
# "The customer service agent\nyou spoke with was: - Attentive"

GuestData$Q54_2<-revalue(GuestData$Q54_2, c("Strongly disagree"="1", "Somewha
t disagree"="2", "Neither agree nor disagree" = "3", "Somewhat agree" = "4",
"Strongly agree" = "5"))
GuestData$Q54_2<-as.numeric(as.character(GuestData$Q54_2))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q54_2)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 6969 rows containing non-finite values (stat_bin).
```



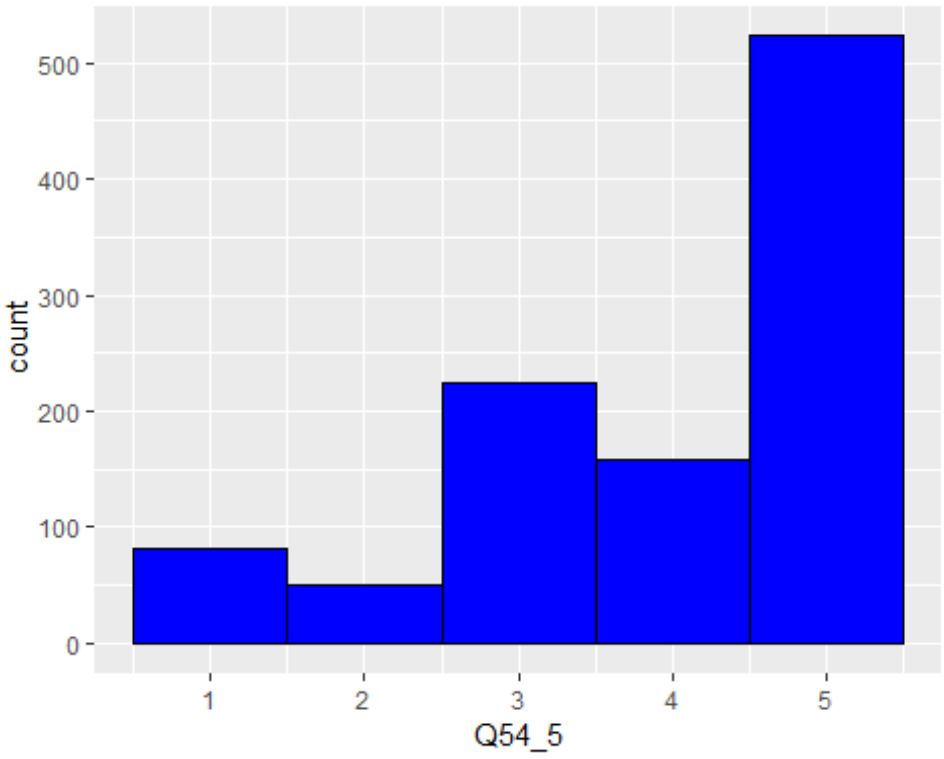

```
# $Q54_3
# ""
# "Neither agree nor disagree"
# "Somewhat agree"
# "Somewhat disagree"
# "Strongly agree"
# "Strongly disagree"
# "The customer service agent\nyou spoke with was: - Professional"

GuestData$Q54_3<-revalue(GuestData$Q54_3, c("Strongly disagree"="1", "Somewha
t disagree"="2", "Neither agree nor disagree" = "3", "Somewhat agree" = "4",
"Strongly agree" = "5"))
GuestData$Q54_3<-as.numeric(as.character(GuestData$Q54_3))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q54_3)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 6951 rows containing non-finite values (stat_bin).
```



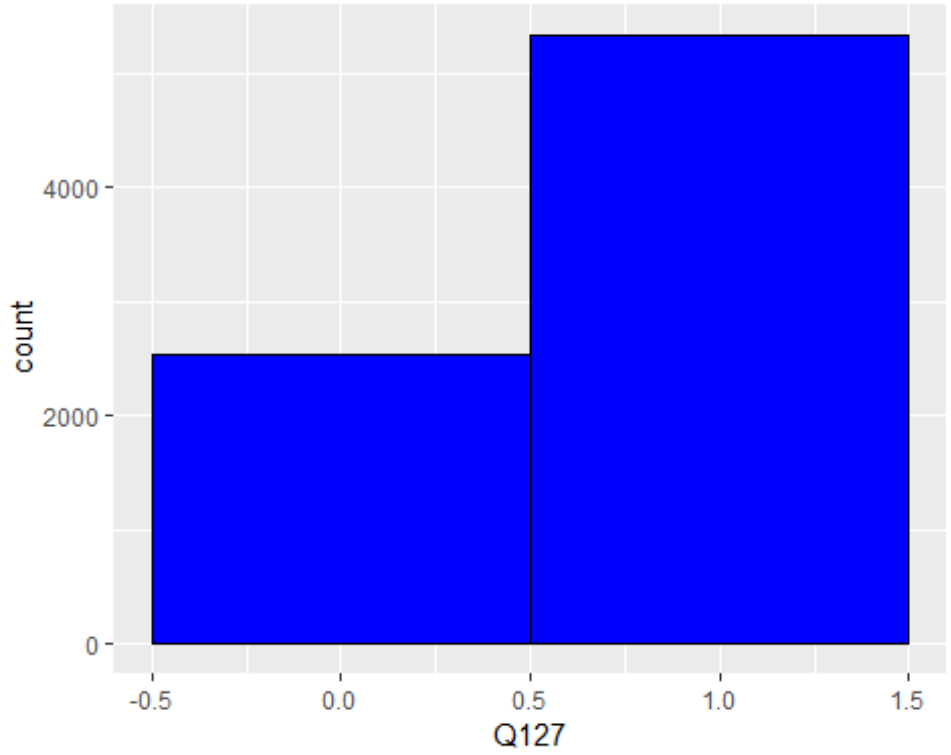

```
# $Q54_4
# ""
# "Neither agree nor disagree"
# "Somewhat agree"
# "Somewhat disagree"
# "Strongly agree"
# "Strongly disagree"
# "The customer service agent\nyou spoke with was: - Friendly"

GuestData$Q54_4<-revalue(GuestData$Q54_4, c("Strongly disagree"="1", "Somewha
t disagree"="2", "Neither agree nor disagree" = "3", "Somewhat agree" = "4",
"Strongly agree" = "5"))
GuestData$Q54_4<-as.numeric(as.character(GuestData$Q54_4))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q54_4)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 6954 rows containing non-finite values (stat_bin).
```



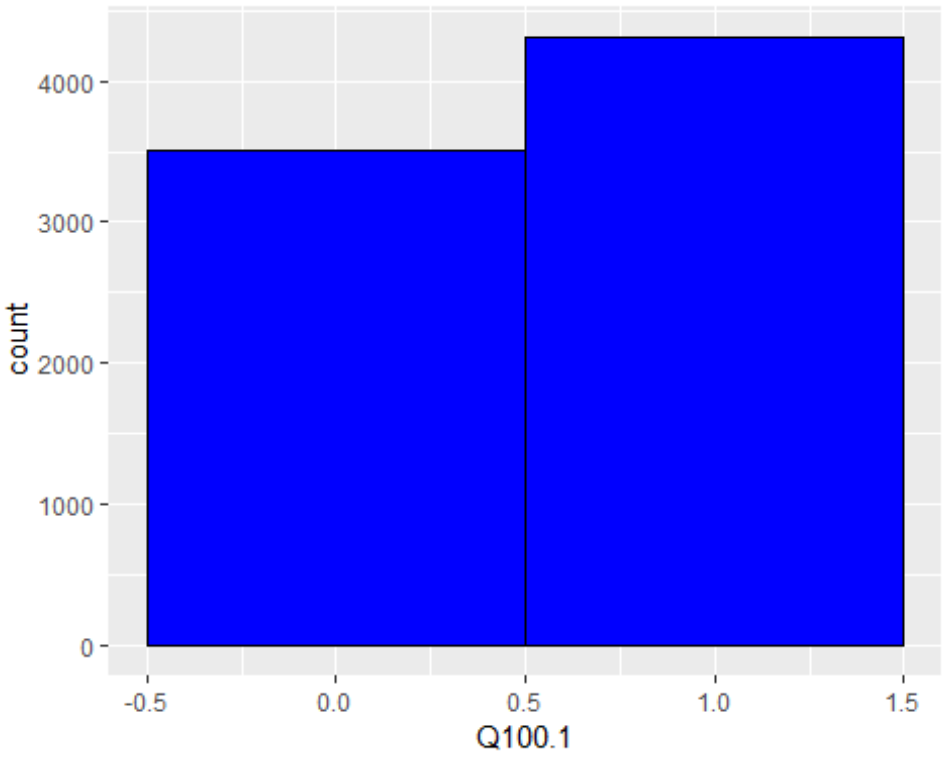

```
# $Q54_5
# ""
# "Neither agree nor disagree"
# "Somewhat agree"
# "Somewhat disagree"
# "Strongly agree"
# "Strongly disagree"
# "The customer service agent\nyou spoke with was: - Helpful"

GuestData$Q54_5<-revalue(GuestData$Q54_5, c("Strongly disagree"="1", "Somewha
t disagree"="2", "Neither agree nor disagree" = "3", "Somewhat agree" = "4",
"Strongly agree" = "5"))
GuestData$Q54_5<-as.numeric(as.character(GuestData$Q54_5))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q54_5)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 6961 rows containing non-finite values (stat_bin).
```




```
# $Q127
# ""
# "Do you plan on flying with AirlineX within the next year?"
# "No"
# "Yes"

GuestData$Q127<-revalue(GuestData$Q127, c("Yes"="1", "No"="0"))
GuestData$Q127<-as.numeric(as.character(GuestData$Q127))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q127)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 123 rows containing non-finite values (stat_bin).
```




```
# $Q100.1
# ""
# "Did any of our team members help make your experience on AirlineX Airlines
a positive one?"
# "No"
# "Yes"

GuestData$Q100.1<-revalue(GuestData$Q100.1, c("Yes"="1", "No"="0"))
GuestData$Q100.1<-as.numeric(as.character(GuestData$Q100.1))

## Warning: NAs introduced by coercion

ggplot(GuestData, aes(x=Q100.1)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")

## Warning: Removed 181 rows containing non-finite values (stat_bin).
```

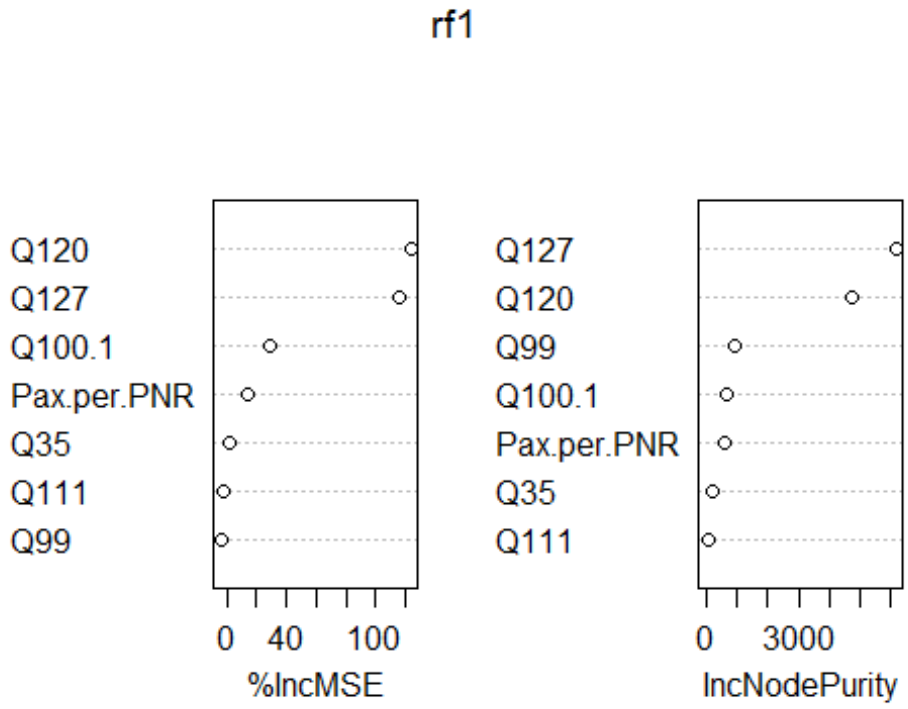



```
GuestData1<-GuestData
# colnames(GuestData1) <- as.character(unlist(GuestData1[1,]))
# GuestData1 = GuestData1[-1, ]

GuestData1[ GuestData1 == "Pax per PNR" ] <- NA
GuestData1[ GuestData1 == "Including your flight on [Field-Flight%20Date], ho
w many flights have you taken in the past twelve months?" ] <- NA

# names(GuestData1)

GuestData1$Q116<-NULL
GuestData1$Q119<-NULL
GuestData1$Q109<-NULL
GuestData1$Q112<-NULL
GuestData1$Q118<-NULL
GuestData1$Q117<-NULL
GuestData1$Q113<-NULL
GuestData1$Q12<-NULL
GuestData1$Q106<-NULL
GuestData1$Q18<-NULL
GuestData1$Q20<-NULL
GuestData1$Q22<-NULL
GuestData1$Q24<-NULL
GuestData1$Q28_1<-NULL
GuestData1$Q34<-NULL
GuestData1$Q36_1<-NULL
```

```
GuestData1$Q36_2<-NULL
GuestData1$Q107.1<-NULL
GuestData1$Q50<-NULL
GuestData1$Q52<-NULL
GuestData1$Q145<-NULL
GuestData1$Booking.Channel<-NULL
GuestData1$Flight.Date.Time<-NULL
GuestData1$Segment1.Destination<-NULL
GuestData1$Segment1.Origin<-NULL

GuestData1 = as.data.frame(sapply(GuestData1, as.numeric))

# Random Forest Model

# install.packages("sfsmisc")
# install.packages("tibble")
# install.packages("rsample")
# install.packages("randomForest")
# install.packages("ranger")
# install.packages("caret")
# install.packages("h2o")
# install.packages("AmesHousing")

library(sfsmisc)

## Warning: package 'sfsmisc' was built under R version 3.5.3

##
## Attaching package: 'sfsmisc'

## The following object is masked from 'package:dplyr':
##
##     last

library(tibble)
library(rsample)

## Warning: package 'rsample' was built under R version 3.5.3

## Loading required package: tidyr

library(randomForest)

## Warning: package 'randomForest' was built under R version 3.5.3

## randomForest 4.6-14

## Type rfNews() to see new features/changes/bug fixes.

##
## Attaching package: 'randomForest'
```

```
## The following object is masked from 'package:ggplot2':
##
##     margin

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

library(ranger)

## Warning: package 'ranger' was built under R version 3.5.3

##
## Attaching package: 'ranger'

## The following object is masked from 'package:randomForest':
##
##     importance

library(caret)

## Warning: package 'caret' was built under R version 3.5.3

## Loading required package: lattice

library(h2o)

## Warning: package 'h2o' was built under R version 3.5.3

##
## ----------------------------------------------------------------------
##
## Your next step is to start H2O:
##     > h2o.init()
##
## For H2O package documentation, ask for help:
##     > ??h2o
##
## After starting H2O, you can use the Web UI at http://localhost:54321
## For more information visit http://docs.h2o.ai
##
## ----------------------------------------------------------------------

##
## Attaching package: 'h2o'

## The following objects are masked from 'package:stats':
##
##     cor, sd, var

## The following objects are masked from 'package:base':
##
##     %*%, %in%, &&, ||, apply, as.factor, as.numeric, colnames,
```

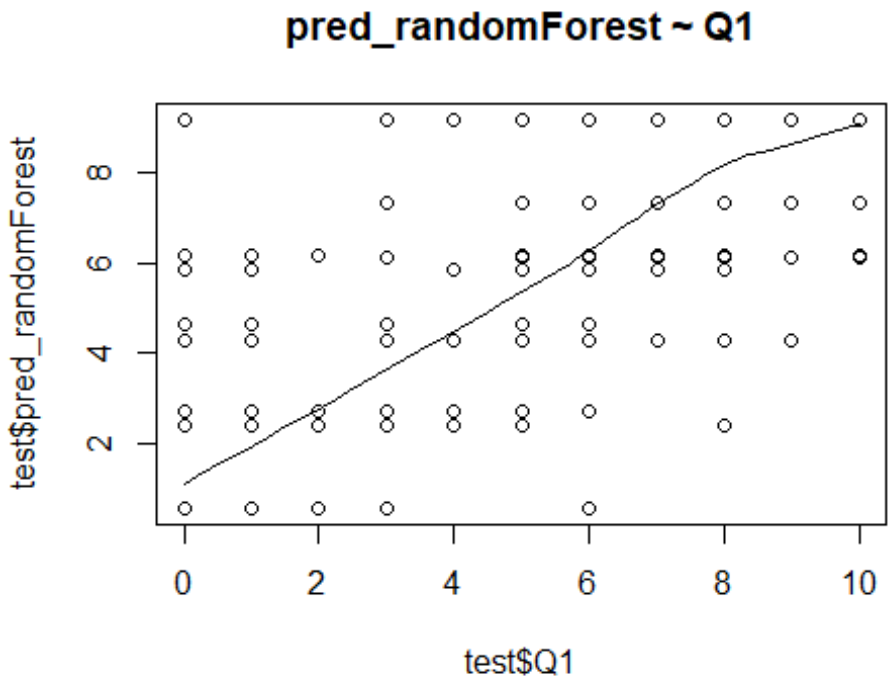
```
##     colnames<-, ifelse, is.character, is.factor, is.numeric, log,
##     log10, log1p, log2, round, signif, trunc

library(AmesHousing)

## Warning: package 'AmesHousing' was built under R version 3.5.3

set.seed(123)

# Treating missing values

GuestData1<-GuestData1[, -which(colMeans(is.na(GuestData1)) > 0.77)]

## 75% of the sample size
smp_size <- floor(0.75 * nrow(GuestData1))

## set the seed to make your partition reproducible
set.seed(123)
train_ind <- sample(seq_len(nrow(GuestData1)), size = smp_size)

train <- GuestData1[train_ind, ]
test <- GuestData1[-train_ind, ]

# for reproducibility

set.seed(123)

train$Q1<-as.numeric(as.character(train$Q1))

rf1<-randomForest(Q1 ~.,
                  data = train, ntree = 500,
                  mtry = 4, importance = TRUE, na.action = na.omit)

print(rf1)

##
## Call:
##  randomForest(formula = Q1 ~ ., data = train, ntree = 500, mtry = 4,
importance = TRUE, na.action = na.omit)
##                Type of random forest: regression
##                      Number of trees: 500
## No. of variables tried at each split: 4
##
##           Mean of squared residuals: 4.279212
##                     % Var explained: 64.93

varImpPlot(rf1)
```




```
rf2<-randomForest(Q1 ~ Q127 + Q120,
                  data = train, ntree = 500,
                  mtry = 2, importance = TRUE, na.action = na.omit)

print(rf2)

##
## Call:
##  randomForest(formula = Q1 ~ Q127 + Q120, data = train, ntree = 500,
mtry = 2, importance = TRUE, na.action = na.omit)
##                Type of random forest: regression
##                      Number of trees: 500
## No. of variables tried at each split: 2
##
##           Mean of squared residuals: 3.926502
##                     % Var explained: 67.68

test$pred_randomForest<-predict(rf2, test)

simple.fit = lm(pred_randomForest~Q1, data = test)
summary(simple.fit)

##
## Call:
## lm(formula = pred_randomForest ~ Q1, data = test)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -5.7046 -1.0413  0.1379  0.8406  7.1647
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.03210    0.16541   12.29   <2e-16 ***
## Q1           0.70268    0.02181   32.23   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.66 on 466 degrees of freedom
##   (1533 observations deleted due to missingness)
## Multiple R-squared:  0.6903, Adjusted R-squared:  0.6896
## F-statistic:  1038 on 1 and 466 DF,  p-value: < 2.2e-16
```

```
scatter.smooth(x=test$Q1, y=test$pred_randomForest, main = "pred_randomForest
~ Q1") # scatterplot
```

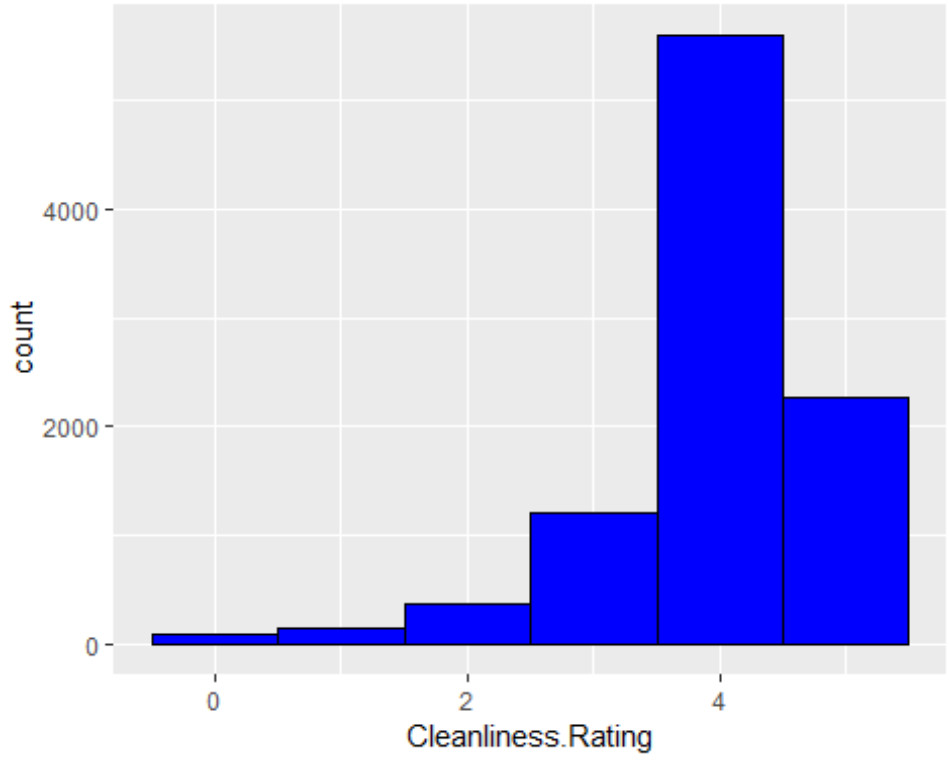



```
# Excel Test Data

# Business Problem:

# The attached workbook contains guest survey data for aircraft cabin cleanli
ness ratings.
# Each aircraft has a "tail number" that uniquely identifies the aircraft.
# Every aircraft stays overnight at a particular airport where the cleaning o
peration takes place.
```

```r
# Every morning, each aircraft departs, flies to different airports serving m
any guests and again stays overnight at some airport at the end of the day.
# Each guest's cleanliness rating shows the effectiveness of the cleaning ope
ration at the overnight airport.
# For simplicity, you can use a 0 to 5 scale to represent the cleanliness rat
ing where 0 = filthy and 5 = spotless.

# Questions to answer:
#  Which airports are the best at cleaning aircraft?
#  Do certain airports impact the aircraft cleanliness ratings more than othe
rs?
#  What other factors could drive the cleanliness ratings to be low?

Surveys<-read.csv(file = "C:/Users/puj83/OneDrive/CV/Cases/AirlineX/Surveys.c
sv", header = T, sep = ",")
Tail_Number_Data<-read.csv(file = "C:/Users/puj83/OneDrive/CV/Cases/AirlineX/
Tail_Number_Data.csv", header = T, sep = ",")
Overnight_Station_Data<-read.csv(file = "C:/Users/puj83/OneDrive/CV/Cases/Air
lineX/Overnight_Station_Data.csv", header = T, sep = ",")
Airport_Delivery_Dates<-read.csv(file = "C:/Users/puj83/OneDrive/CV/Cases/Air
lineX/Airport_Delivery_Dates.csv", header = T, sep = ",")

names(Surveys)

## [1] "PNR"                "Origin"             "Destination"
## [4] "Flight.Date.Time"   "Survey.Date"        "Cleanliness.Rating"

# [1] "PNR"                "Origin"             "Destination"        "Flight.
Date.Time"   "Survey.Date"        "Cleanliness.Rating"

names(Tail_Number_Data)

## [1] "PNR"                  "Departure.Date"       "Tail.Number"
## [4] "Airport.Code.Deprt.." "Airport.Code.Arrvl.."

# [1] "PNR"                  "Departure.Date"       "Tail.Number"          "A
irport.Code.Deprt.." "Airport.Code.Arrvl.."

names(Overnight_Station_Data)

## [1] "Airport.Code.Deprt.." "Tail.Number"          "Departure.Date"

# "Airport.Code.Deprt.." "Tail.Number"          "Departure.Date"

names(Airport_Delivery_Dates)

## [1] "Tail.Number"            "Aircraft.Delivery.Date"

# [1] "Tail.Number"            "Aircraft.Delivery.Date"

Surveys$PNR<-as.character(Surveys$PNR)
```

```r
Tail_Number_Data$PNR<-as.character(Tail_Number_Data$PNR)

dim(Surveys)

## [1] 9399    6

dim(Tail_Number_Data)

## [1] 8606    5

# Left outer join Tail_Number_Data so that the specified Tail Numbers would h
ave a cabin cleanliness rating.

Test<-left_join(Surveys, Tail_Number_Data, by = "PNR")
dim(Test)

## [1] 9676   10

names(Test)

##  [1] "PNR"                  "Origin"               "Destination"
##  [4] "Flight.Date.Time"     "Survey.Date"          "Cleanliness.Rating"
##  [7] "Departure.Date"       "Tail.Number"          "Airport.Code.Deprt.."
## [10] "Airport.Code.Arrvl.."

# Concatenate Tail.Number and Departure.Date in Overnight_Station_Data for a
unique ID.
# Concatenate Tail.Number and Departure.Date in Test so you can join Overnigh
t_Station_Data on Test Data.

Test$x <- paste(Test$Tail.Number, "-", Test$Departure.Date)
Overnight_Station_Data$x <- paste(Overnight_Station_Data$Tail.Number, "-", Ov
ernight_Station_Data$Departure.Date)

Test$x<-as.character(Test$x)
Overnight_Station_Data$x<-as.character(Overnight_Station_Data$x)

#Get rid of Tail.Number, Departure.Date in Overnight_Station_Data to get rid
of duplicate columns (when they join)
#Change the name of "Airport.Code.Deprt.." to "Overnight_Airport"in Overnight
_Station_Data

Overnight_Station_Data$Tail.Number<-NULL
Overnight_Station_Data$Departure.Date<-NULL
colnames(Overnight_Station_Data)[colnames(Overnight_Station_Data)=="Airport.C
ode.Deprt.."] <- "Overnight_Airport"

# Left outer join Overnight_Station_Data on Test to get Overnight_Aiport matc
hed up with cleanliness rating in original Test Data.

Test<-left_join(Test, Overnight_Station_Data, by = "x")
```

```
# Left outer join Airport_Delivery_Dates on Test to get Aircraft Delivery Dat
es matched up with cleanliness rating in original Test Data.

Test<-left_join(Test, Airport_Delivery_Dates, by = "Tail.Number")

Test$Cleanliness.Rating<-as.factor(Test$Cleanliness.Rating)
levels(Test$Cleanliness.Rating)

## [1] "Clean"          "Dirty"          "Filthy"         "Somewhat Clean"
## [5] "Somewhat Dirty" "Spotless"

# Convert cleanliness rating where 0 = filthy and 5 = spotless.

Test$Cleanliness.Rating<-revalue(Test$Cleanliness.Rating, c("Filthy"="0", "Di
rty"="1", "Somewhat Dirty"="2", "Somewhat Clean"="3", "Clean"="4", "Spotless"
="5"))
Test$Cleanliness.Rating<-as.numeric(as.character(Test$Cleanliness.Rating))

# Plot histogram to for data exploration

ggplot(Test, aes(x=Cleanliness.Rating)) +
  geom_histogram(binwidth=1, colour="black", fill="blue")
```




```
# Write to csv and do further analysis in PowerBI.

# write.csv(Test, "C:/Users/puj83/OneDrive/CV/Cases/AirlineX/Test.csv")
```