

```
## Error: <text>:4:0: unexpected end of input
## 2: library(xtable)
## 3: opts_chunk$set(fig.path='figure/', fig.height=4)
## ^
```

# Lista 1

Mikołaj Langner, Marcin Kostrzewa

31.3.2021

## 1 Etap I

```
df <- read.csv('churn.txt')
dim(df)

## [1] 3333   21

head(df)

##   State Account.Length Area.Code   Phone Int.l.Plan VMail.Plan VMail.Message
## 1    KS          128     415 382-4657       no      yes        25
## 2    OH          107     415 371-7191       no      yes        26
## 3    NJ          137     415 358-1921       no       no         0
## 4    OH           84     408 375-9999      yes       no         0
## 5    OK           75     415 330-6626      yes       no         0
## 6    AL          118     510 391-8027      yes       no         0
##   Day.Mins Day.Calls Day.Charge Eve.Mins Eve.Calls Eve.Charge Night.Mins
## 1    265.1     110    45.07    197.4      99    16.78    244.7
## 2    161.6     123    27.47    195.5     103    16.62    254.4
## 3    243.4     114    41.38    121.2     110    10.30    162.6
## 4    299.4      71    50.90     61.9      88     5.26    196.9
## 5    166.7     113    28.34    148.3     122    12.61    186.9
## 6    223.4      98    37.98    220.6     101    18.75    203.9
##   Night.Calls Night.Charge Intl.Mins Intl.Calls Intl.Charge CustServ.Calls
## 1         91     11.01     10.0        3     2.70        1
## 2        103     11.45     13.7        3     3.70        1
## 3        104      7.32     12.2        5     3.29        0
## 4         89      8.86      6.6        7     1.78        2
## 5        121      8.41     10.1        3     2.73        3
## 6        118      9.18      6.3        6     1.70        0
##   Churn.
## 1 False.
## 2 False.
## 3 False.
## 4 False.
## 5 False.
## 6 False.

str(df)

## 'data.frame': 3333 obs. of  21 variables:
## $ State      : Factor w/ 51 levels "AK","AL","AR",...: 17 36 32 36 37 2 20 25 19 50 ...
## $ Account.Length: int  128 107 137 84 75 118 121 147 117 141 ...
```

```

## $ Area.Code      : int 415 415 415 408 415 510 510 415 408 415 ...
## $ Phone         : Factor w/ 3333 levels "327-1058","327-1319",...: 1927 1576 1118 1708 111 2254 1048 81
## $ Int.l.Plan    : Factor w/ 2 levels "no","yes": 1 1 1 2 2 2 1 2 1 2 ...
## $ VMail.Plan    : Factor w/ 2 levels "no","yes": 2 2 1 1 1 1 2 1 1 2 ...
## $ VMail.Message : int 25 26 0 0 0 0 24 0 0 37 ...
## $ Day.Mins      : num 265 162 243 299 167 ...
## $ Day.Calls     : int 110 123 114 71 113 98 88 79 97 84 ...
## $ Day.Charge    : num 45.1 27.5 41.4 50.9 28.3 ...
## $ Eve.Mins      : num 197.4 195.5 121.2 61.9 148.3 ...
## $ Eve.Calls     : int 99 103 110 88 122 101 108 94 80 111 ...
## $ Eve.Charge    : num 16.78 16.62 10.3 5.26 12.61 ...
## $ Night.Mins    : num 245 254 163 197 187 ...
## $ Night.Calls   : int 91 103 104 89 121 118 118 96 90 97 ...
## $ Night.Charge  : num 11.01 11.45 7.32 8.86 8.41 ...
## $ Intl.Mins     : num 10 13.7 12.2 6.6 10.1 6.3 7.5 7.1 8.7 11.2 ...
## $ Intl.Calls    : int 3 3 5 7 3 6 7 6 4 5 ...
## $ Intl.Charge   : num 2.7 3.7 3.29 1.78 2.73 1.7 2.03 1.92 2.35 3.02 ...
## $ CustServ.Calls: int 1 1 0 2 3 0 3 0 1 0 ...
## $ Churn.        : Factor w/ 2 levels "False.","True.": 1 1 1 1 1 1 1 1 1 1 ...

```

```

df$Area.Code <- as.factor(df$Area.Code)
df$Churn. <- as.factor(df$Churn.)
df$Int.l.Plan <- as.factor(df$Int.l.Plan)
df$VMail.Plan <- as.factor(df$VMail.Plan)
df$State <- as.factor(df$State)

```

```

sapply(df, function(x) sum(is.na(x)))

##           State Account.Length       Area.Code       Phone       Int.l.Plan
##             0            0            0            0            0
## VMail.Plan VMail.Message       Day.Mins       Day.Calls       Day.Charge
##             0            0            0            0            0
##       Eve.Mins     Eve.Calls     Eve.Charge     Night.Mins     Night.Calls
##             0            0            0            0            0
## Night.Charge     Intl.Mins     Intl.Calls     Intl.Charge CustServ.Calls
##             0            0            0            0            0
##       Churn.
##             0

```

```
df <- subset(df, select=-Phone)
```

```

sapply(df[, sapply(df, is.factor)], levels)

## $State
## [1] "AK" "AL" "AR" "AZ" "CA" "CO" "CT" "DC" "DE" "FL" "GA" "HI" "IA" "ID" "IL"
## [16] "IN" "KS" "KY" "LA" "MA" "MD" "ME" "MI" "MN" "MO" "MS" "MT" "NC" "ND" "NE"
## [31] "NH" "NJ" "NM" "NV" "NY" "OH" "OK" "OR" "PA" "RI" "SC" "SD" "TN" "TX" "UT"
## [46] "VA" "VT" "WA" "WI" "WV" "WY"
##
## $Area.Code
## [1] "408" "415" "510"

```

```

## 
## $Int.l.Plan
## [1] "no"  "yes"
##
## $VMail.Plan
## [1] "no"  "yes"
##
## $Churn.
## [1] "False." "True."

```

**summary(df)**

```

##      State    Account.Length  Area.Code Int.l.Plan VMail.Plan
##  WV     : 106   Min.    : 1.0    408: 838   no :3010   no :2411
##  MN     :  84   1st Qu.: 74.0   415:1655 yes: 323   yes: 922
##  NY     :  83   Median  :101.0   510: 840
##  AL     :  80   Mean    :101.1
##  OH     :  78   3rd Qu.:127.0
##  OR     :  78   Max.    :243.0
##  (Other):2824
##      VMail.Message      Day.Mins      Day.Calls      Day.Charge
##  Min.   : 0.000   Min.   : 0.0   Min.   : 0.0   Min.   : 0.00
##  1st Qu.: 0.000   1st Qu.:143.7  1st Qu.: 87.0  1st Qu.:24.43
##  Median : 0.000   Median :179.4   Median :101.0   Median :30.50
##  Mean   : 8.099   Mean   :179.8   Mean   :100.4   Mean   :30.56
##  3rd Qu.:20.000  3rd Qu.:216.4  3rd Qu.:114.0  3rd Qu.:36.79
##  Max.   :51.000   Max.   :350.8   Max.   :165.0   Max.   :59.64
##
##      Eve.Mins      Eve.Calls      Eve.Charge      Night.Mins
##  Min.   : 0.0   Min.   : 0.0   Min.   : 0.00   Min.   : 23.2
##  1st Qu.:166.6  1st Qu.: 87.0  1st Qu.:14.16  1st Qu.:167.0
##  Median :201.4  Median :100.0   Median :17.12  Median :201.2
##  Mean   :201.0  Mean   :100.1   Mean   :17.08  Mean   :200.9
##  3rd Qu.:235.3  3rd Qu.:114.0  3rd Qu.:20.00  3rd Qu.:235.3
##  Max.   :363.7  Max.   :170.0   Max.   :30.91  Max.   :395.0
##
##      Night.Calls      Night.Charge      Intl.Mins      Intl.Calls
##  Min.   :33.0   Min.   : 1.040   Min.   : 0.00   Min.   : 0.000
##  1st Qu.:87.0  1st Qu.: 7.520   1st Qu.: 8.50   1st Qu.: 3.000
##  Median :100.0  Median : 9.050   Median :10.30   Median : 4.000
##  Mean   :100.1  Mean   : 9.039   Mean   :10.24   Mean   : 4.479
##  3rd Qu.:113.0  3rd Qu.:10.590  3rd Qu.:12.10  3rd Qu.: 6.000
##  Max.   :175.0  Max.   :17.770   Max.   :20.00  Max.   :20.000
##
##      Intl.Charge      CustServ.Calls      Churn.
##  Min.   :0.000   Min.   :0.000   False.:2850
##  1st Qu.:2.300  1st Qu.:1.000   True. : 483
##  Median :2.780  Median :1.000
##  Mean   :2.765  Mean   :1.563
##  3rd Qu.:3.270  3rd Qu.:2.000
##  Max.   :5.400  Max.   :9.000
##
```

## 2 Etap II

```
library(ggplot2)
library(GGally)
library(tidyr)
library(dplyr)
library(EnvStats)
library(DescTools)
```

```
factors <- subset(df, select=sapply(df, is.factor))
numerics <- subset(df, select=sapply(df, function(x) !is.factor(x)))
```

a)

```
extended.summary <- function(X) {
  c(mean = mean(X), trimmed.mean = mean(X, trim=.0025), min = min(X), Q1 = quantile(X, .25, names=F)
}

sapply(numerics, extended.summary)

##          Account.Length VMail.Message Day.Mins Day.Calls Day.Charge
## mean           101.06481     8.09901 179.77510 100.43564 30.562307
## trimmed.mean   101.00392     8.02050 179.80413 100.46850 30.567241
## min            1.00000     0.00000  0.00000  0.00000  0.000000
## Q1             74.00000     0.00000 143.70000  87.00000 24.430000
## median         101.00000     0.00000 179.40000 101.00000 30.500000
## Q3             127.00000    20.00000 216.40000 114.00000 36.790000
## max            243.00000    51.00000 350.80000 165.00000 59.640000
## range1         1.00000     0.00000  0.00000  0.00000  0.000000
## range2         243.00000    51.00000 350.80000 165.00000 59.640000
## sd              39.82211    13.68837  54.46739  20.06908  9.259435
## IQR            53.00000    20.00000  72.70000  27.00000 12.360000
##          Eve.Mins Eve.Calls Eve.Charge Night.Mins Night.Calls Night.Charge
## mean           200.98035 100.11431 17.083540 200.87204 100.10771  9.039325
## trimmed.mean   201.00362 100.13205 17.085520 200.83841 100.10280  9.037814
## min            0.00000  0.00000  0.000000 23.20000  33.00000  1.040000
## Q1             166.60000 87.00000 14.160000 167.00000  87.00000  7.520000
## median         201.40000 100.00000 17.120000 201.20000 100.00000  9.050000
## Q3             235.30000 114.00000 20.000000 235.30000 113.00000 10.590000
## max            363.70000 170.00000 30.910000 395.00000 175.00000 17.770000
## range1         0.00000  0.00000  0.000000 23.20000  33.00000  1.040000
## range2         363.70000 170.00000 30.910000 395.00000 175.00000 17.770000
## sd              50.71384 19.92263  4.310668  50.57385  19.56861  2.275873
## IQR            68.70000 27.00000  5.840000  68.30000  26.00000  3.070000
##          Intl.Mins Intl.Calls Intl.Charge CustServ.Calls
## mean           10.23729  4.479448  2.7645815      1.562856
## trimmed.mean  10.24206  4.458245  2.7658728      1.551703
## min            0.00000  0.000000  0.0000000     0.000000
## Q1             8.50000  3.000000  2.3000000     1.000000
## median         10.30000  4.000000  2.7800000     1.000000
## Q3             12.10000  6.000000  3.2700000     2.000000
```

```

## max           20.00000 20.000000 5.4000000 9.0000000
## range1        0.00000 0.000000 0.0000000 0.0000000
## range2        20.00000 20.000000 5.4000000 9.0000000
## sd            2.79184 2.461214 0.7537726 1.315491
## IQR           3.60000 3.000000 0.9700000 1.0000000

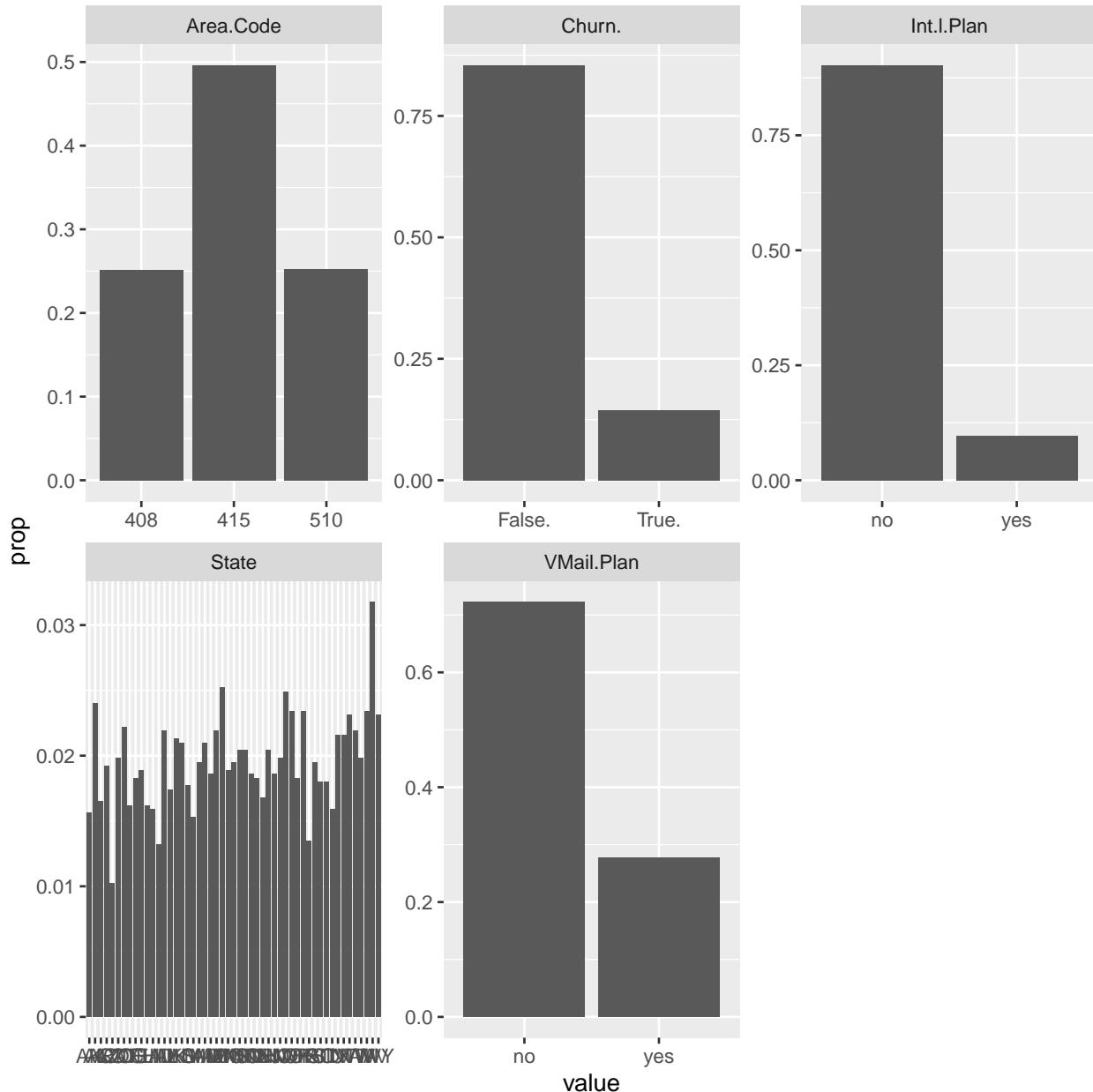
```

b)

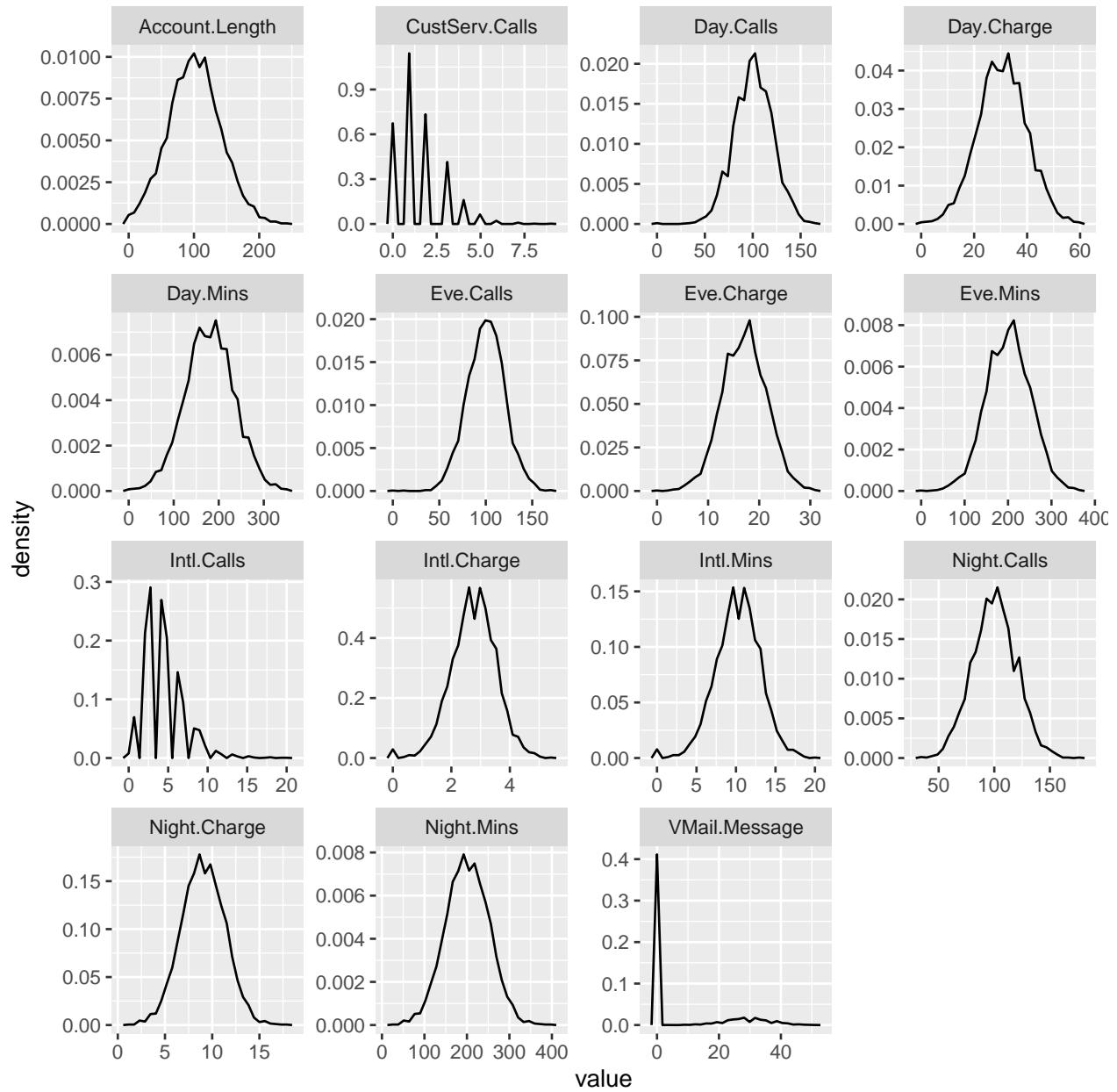
```

ggplot(gather(factors), aes(value)) +
  geom_bar(aes(y=..prop.., group=1)) +
  facet_wrap(~key, scales='free')

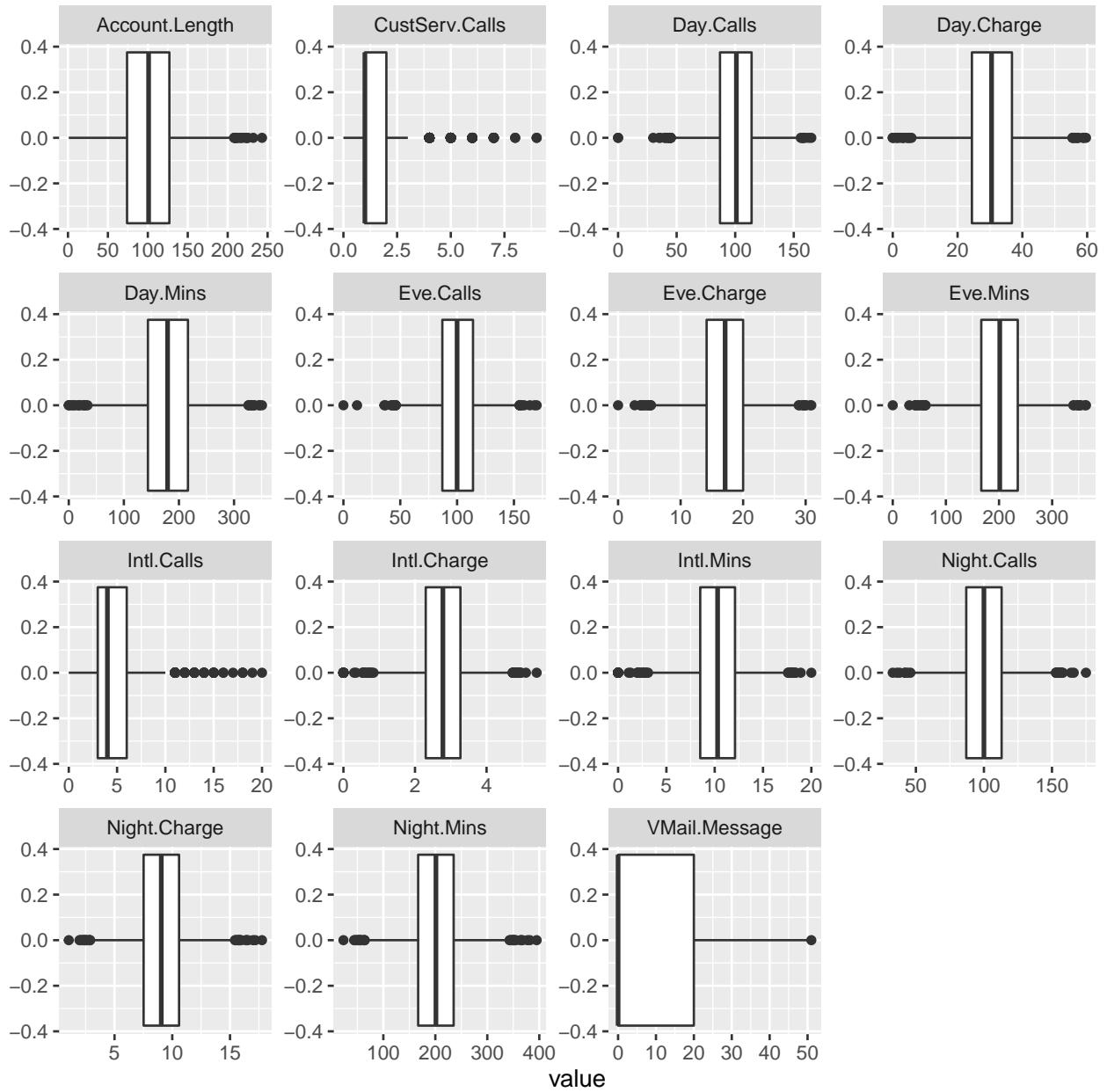
```



```
ggplot(gather(numerics), aes(value)) +
  geom_freqpoly(aes(y=..density..)) +
  facet_wrap(~key, scales='free')
```

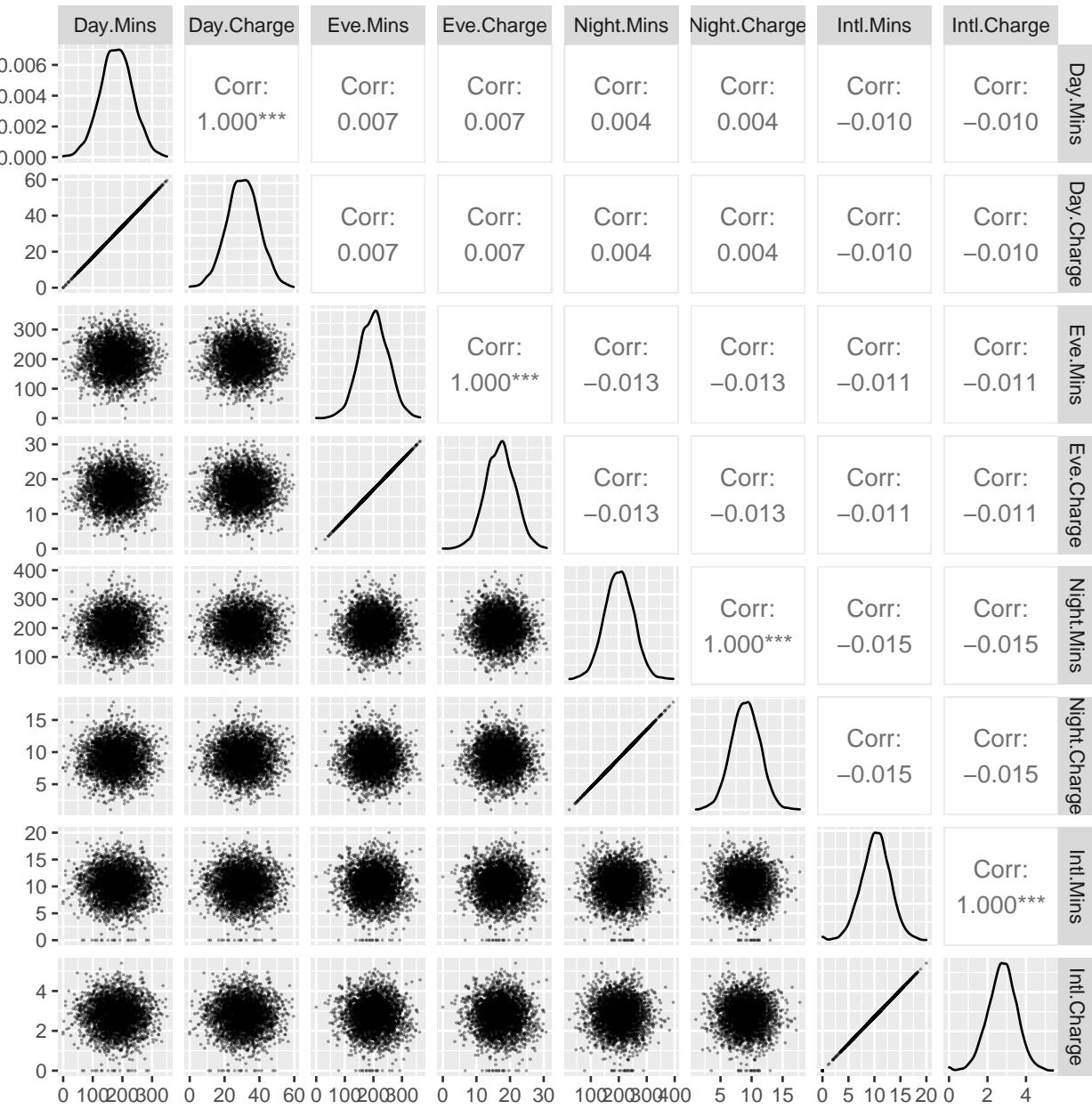


```
ggplot(gather(numerics), aes(value)) +
  geom_boxplot(aes(x=value)) +
  facet_wrap(~key, scales='free')
```



c)

```
continuous <- subset(numerics, select=sapply(numerics, function(x) !is.integer(x)))
gpairs(continuous,
       lower=list(continuous=wrap("points", alpha=.4, size=.01)))
```



d)

```
sapply(numerics, Range)
```

```
## Account.Length  VMail.Message      Day.Mins      Day.Calls      Day.Charge
##      242.00          51.00       350.80      165.00       59.64
##      Eve.Mins        Eve.Calls     Eve.Charge    Night.Mins    Night.Calls
##      363.70          170.00       30.91      371.80       142.00
##  Night.Charge      Intl.Mins     Intl.Calls  Intl.Charge CustServ.Calls
##      16.73           20.00        20.00        5.40          9.00
```

```

sapply(factors, levels)

## $State
## [1] "AK" "AL" "AR" "AZ" "CA" "CO" "CT" "DC" "DE" "FL" "GA" "HI" "IA" "ID" "IL"
## [16] "IN" "KS" "KY" "LA" "MA" "MD" "ME" "MI" "MN" "MO" "MS" "MT" "NC" "ND" "NE"
## [31] "NH" "NJ" "NM" "NV" "NY" "OH" "OK" "OR" "PA" "RI" "SC" "SD" "TN" "TX" "UT"
## [46] "VA" "VT" "WA" "WI" "WV" "WY"
##
## $Area.Code
## [1] "408" "415" "510"
##
## $Int.l.Plan
## [1] "no" "yes"
##
## $VMail.Plan
## [1] "no" "yes"
##
## $Churn.
## [1] "False." "True."

```

```
sapply(numerics, skewness)
```

	Account.Length	VMail.Message	Day.Mins	Day.Calls	Day.Charge
##	0.096606294	1.264823634	-0.029077067	-0.111786639	-0.029083268
##	Eve.Mins	Eve.Calls	Eve.Charge	Night.Mins	Night.Calls
##	-0.023877456	-0.055563139	-0.023857989	0.008921291	0.032499570
##	Night.Charge	Intl.Mins	Intl.Calls	Intl.Charge	CustServ.Calls
##	0.008886237	-0.245135939	1.321478166	-0.245286508	1.091359482

```
sapply(numerics, cv)
```

	Account.Length	VMail.Message	Day.Mins	Day.Calls	Day.Charge
##	0.3940255	1.6901282	0.3029752	0.1998203	0.3029691
##	Eve.Mins	Eve.Calls	Eve.Charge	Night.Mins	Night.Calls
##	0.2523324	0.1989988	0.2523287	0.2517715	0.1954755
##	Night.Charge	Intl.Mins	Intl.Calls	Intl.Charge	CustServ.Calls
##	0.2517746	0.2727127	0.5494459	0.2726534	0.8417223

### 3 Etap III

a)

```

numerics <- data.frame(numerics, Churn. = df$Churn.)

aggregate(. ~ Churn., numerics, extended.summary)

##   Churn. Account.Length.mean Account.Length.trimmed.mean Account.Length.min
## 1 False.          100.79368                  100.73202           1.00000
## 2 True.           102.66460                  102.62162           1.00000
##   Account.Length.Q1 Account.Length.median Account.Length.Q3 Account.Length.max

```

```

## 1      73.00000      100.00000      127.00000      243.00000
## 2      76.00000      103.00000      127.00000      225.00000
##   Account.Length.range1 Account.Length.range2 Account.Length.sd
## 1          1.00000      243.00000      39.88235
## 2          1.00000      225.00000      39.46782
##   Account.Length.IQR VMail.Message.mean VMail.Message.trimmed.mean
## 1          54.00000      8.604561      8.526446
## 2          51.00000      5.115942      5.037422
##   VMail.Message.min VMail.Message.Q1 VMail.Message.median VMail.Message.Q3
## 1          0.000000      0.000000      0.000000      22.000000
## 2          0.000000      0.000000      0.000000      0.000000
##   VMail.Message.max VMail.Message.range1 VMail.Message.range2 VMail.Message.sd
## 1          51.000000      0.000000      51.000000      13.913125
## 2          48.000000      0.000000      48.000000      11.860138
##   VMail.Message.IQR Day.Mins.mean Day.Mins.trimmed.mean Day.Mins.min
## 1          22.000000      175.17575      175.25127      0.00000
## 2          0.000000      206.91408      207.04511      0.00000
##   Day.Mins.Q1 Day.Mins.median Day.Mins.Q3 Day.Mins.max Day.Mins.range1
## 1          142.82500      177.20000      210.30000      315.60000      0.00000
## 2          153.25000      217.60000      265.95000      350.80000      0.00000
##   Day.Mins.range2 Day.Mins.sd Day.Mins.IQR Day.Calls.mean
## 1          315.60000      50.18166      67.47500      100.28316
## 2          350.80000      68.99779      112.70000      101.33540
##   Day.Calls.trimmed.mean Day.Calls.min Day.Calls.Q1 Day.Calls.median
## 1          100.30959      0.00000      87.00000      100.00000
## 2          101.41372      0.00000      87.50000      103.00000
##   Day.Calls.Q3 Day.Calls.max Day.Calls.range1 Day.Calls.range2 Day.Calls.sd
## 1          114.00000      163.00000      0.00000      163.00000      19.80116
## 2          116.50000      165.00000      0.00000      165.00000      21.58231
##   Day.Calls.IQR Day.Charge.mean Day.Charge.trimmed.mean Day.Charge.min
## 1          27.00000      29.780421      29.793262      0.000000
## 2          29.00000      35.175921      35.198191      0.000000
##   Day.Charge.Q1 Day.Charge.median Day.Charge.Q3 Day.Charge.max
## 1          24.282500      30.120000      35.750000      53.650000
## 2          26.055000      36.990000      45.210000      59.640000
##   Day.Charge.range1 Day.Charge.range2 Day.Charge.sd Day.Charge.IQR
## 1          0.000000      53.650000      8.530835      11.467500
## 2          0.000000      59.640000      11.729710      19.155000
##   Eve.Mins.mean Eve.Mins.trimmed.mean Eve.Mins.min Eve.Mins.Q1 Eve.Mins.median
## 1          199.04330      199.07831      0.00000      164.50000      199.60000
## 2          212.41014      212.38981      70.90000      177.10000      211.30000
##   Eve.Mins.Q3 Eve.Mins.max Eve.Mins.range1 Eve.Mins.range2 Eve.Mins.sd
## 1          233.20000      361.80000      0.00000      361.80000      50.29217
## 2          249.45000      363.70000      70.90000      363.70000      51.72891
##   Eve.Mins.IQR Eve.Calls.mean Eve.Calls.trimmed.mean Eve.Calls.min Eve.Calls.Q1
## 1          68.70000      100.03860      100.06488      0.00000      87.00000
## 2          72.35000      100.56108      100.53015      48.00000      87.00000
##   Eve.Calls.median Eve.Calls.Q3 Eve.Calls.max Eve.Calls.range1 Eve.Calls.range2
## 1          100.00000      114.00000      170.00000      0.00000      170.00000
## 2          101.00000      114.00000      168.00000      48.00000      168.00000
##   Eve.Calls.sd Eve.Calls.IQR Eve.Charge.mean Eve.Charge.trimmed.mean
## 1          19.95841      27.00000      16.918909      16.921890
## 2          19.72471      27.00000      18.054969      18.053243
##   Eve.Charge.min Eve.Charge.Q1 Eve.Charge.median Eve.Charge.Q3 Eve.Charge.max

```

```

## 1      0.000000  13.980000   16.970000  19.820000 30.750000
## 2      6.030000  15.055000   17.960000  21.205000 30.910000
##   Eve.Charge.range1 Eve.Charge.range2 Eve.Charge.sd Eve.Charge.IQR
## 1      0.000000  30.750000   4.274863   5.840000
## 2      6.030000  30.910000   4.396762   6.150000
##   Night.Mins.mean Night.Mins.trimmed.mean Night.Mins.min Night.Mins.Q1
## 1      200.13319 200.09041   23.20000  165.90000
## 2      205.23168 205.24865   47.40000  171.25000
##   Night.Mins.median Night.Mins.Q3 Night.Mins.max Night.Mins.range1
## 1      200.25000 234.90000  395.00000  23.20000
## 2      204.80000 239.85000  354.90000  47.40000
##   Night.Mins.range2 Night.Mins.sd Night.Mins.IQR Night.Calls.mean
## 1      395.00000 51.10503   69.00000  100.05825
## 2      354.90000 47.13282   68.60000  100.39959
##   Night.Calls.trimmed.mean Night.Calls.min Night.Calls.Q1 Night.Calls.median
## 1      100.05430 33.00000  87.00000  100.00000
## 2      100.38669 49.00000  85.00000  100.00000
##   Night.Calls.Q3 Night.Calls.max Night.Calls.range1 Night.Calls.range2
## 1      113.00000 175.00000  33.00000  175.00000
## 2      115.00000 158.00000  49.00000  158.00000
##   Night.Calls.sd Night.Calls.IQR Night.Charge.mean Night.Charge.trimmed.mean
## 1      19.50625 26.00000  9.006074  9.004150
## 2      19.95066 30.00000  9.235528  9.236299
##   Night.Charge.min Night.Charge.Q1 Night.Charge.median Night.Charge.Q3
## 1      1.040000 7.470000  9.010000  10.570000
## 2      2.130000 7.705000  9.220000  10.795000
##   Night.Charge.max Night.Charge.range1 Night.Charge.range2 Night.Charge.sd
## 1      17.770000 1.040000  17.770000  2.299768
## 2      15.970000 2.130000  15.970000  2.121081
##   Night.Charge.IQR Intl.Mins.mean Intl.Mins.trimmed.mean Intl.Mins.min
## 1      3.100000 10.158877 10.163999  0.000000
## 2      3.090000 10.700000 10.698753  2.000000
##   Intl.Mins.Q1 Intl.Mins.median Intl.Mins.Q3 Intl.Mins.max Intl.Mins.range1
## 1      8.400000 10.200000 12.000000 18.900000  0.000000
## 2      8.800000 10.600000 12.800000 20.000000  2.000000
##   Intl.Mins.range2 Intl.Mins.sd Intl.Mins.IQR Intl.Calls.mean
## 1      18.900000 2.784489 3.600000  4.532982
## 2      20.000000 2.793190 4.000000  4.163561
##   Intl.Calls.trimmed.mean Intl.Calls.min Intl.Calls.Q1 Intl.Calls.median
## 1      4.512341 0.000000 3.000000  4.000000
## 2      4.137214 1.000000 2.000000  4.000000
##   Intl.Calls.Q3 Intl.Calls.max Intl.Calls.range1 Intl.Calls.range2
## 1      6.000000 19.000000 0.000000 19.000000
## 2      5.000000 20.000000 1.000000 20.000000
##   Intl.Calls.sd Intl.Calls.IQR Intl.Charge.mean Intl.Charge.trimmed.mean
## 1      2.441984 3.000000 2.7434035 2.7447920
## 2      2.551575 3.000000 2.8895445 2.8892100
##   Intl.Charge.min Intl.Charge.Q1 Intl.Charge.median Intl.Charge.Q3
## 1      0.0000000 2.2700000 2.7500000 3.2400000
## 2      0.5400000 2.3800000 2.8600000 3.4600000
##   Intl.Charge.max Intl.Charge.range1 Intl.Charge.range2 Intl.Charge.sd
## 1      5.1000000 0.0000000 5.1000000 0.7517843
## 2      5.4000000 0.5400000 5.4000000 0.7541521
##   Intl.Charge.IQR CustServ.Calls.mean CustServ.Calls.trimmed.mean

```

```

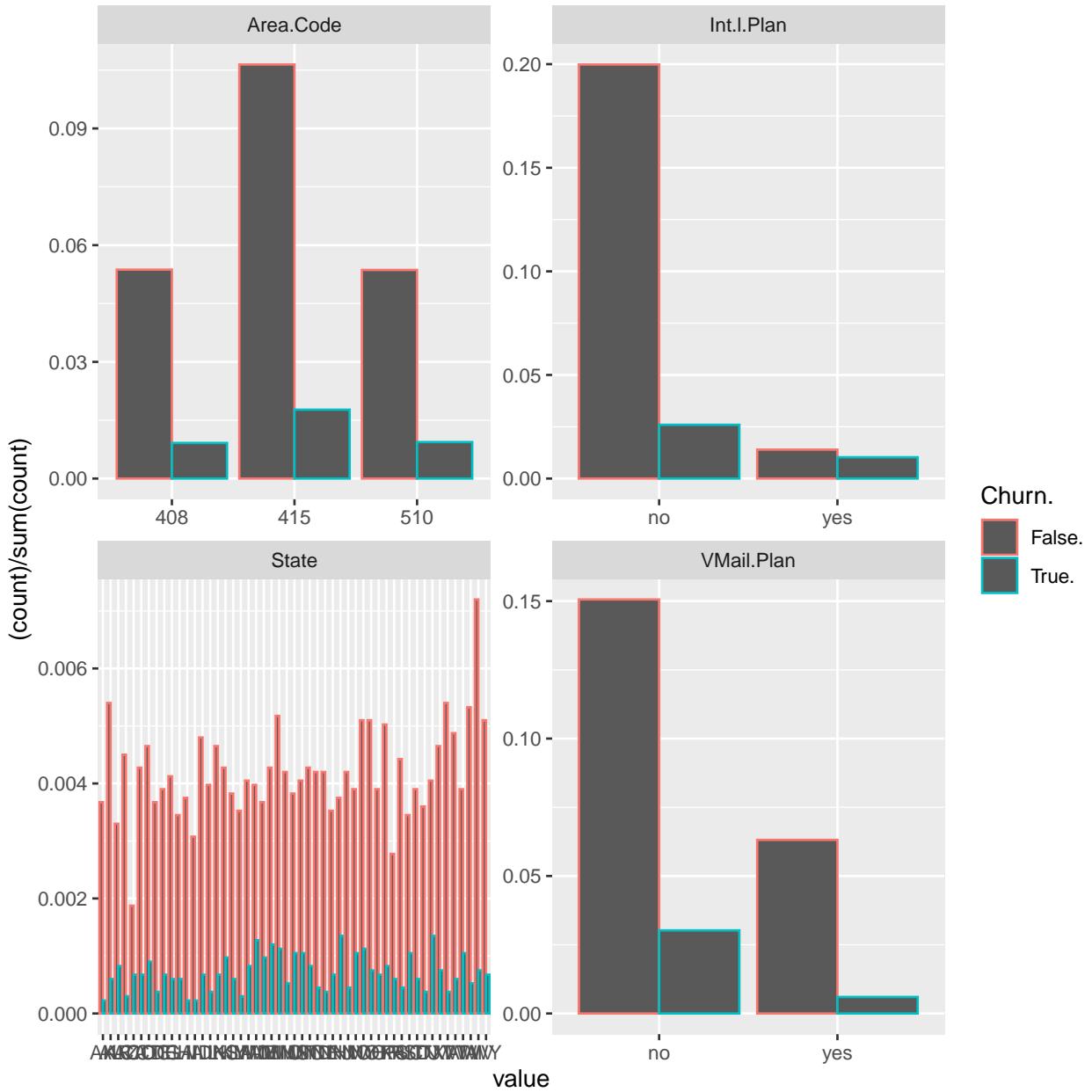
## 1      0.9700000      1.449825      1.440056
## 2      1.0800000      2.229814      2.220374
##   CustServ.Calls.min CustServ.Calls.Q1 CustServ.Calls.median CustServ.Calls.Q3
## 1      0.000000      1.000000      1.000000      2.000000
## 2      0.000000      1.000000      2.000000      4.000000
##   CustServ.Calls.max CustServ.Calls.range1 CustServ.Calls.range2
## 1      8.000000      0.000000      8.000000
## 2      9.000000      0.000000      9.000000
##   CustServ.Calls.sd CustServ.Calls.IQR
## 1      1.163883      1.000000
## 2      1.853275      3.000000

```

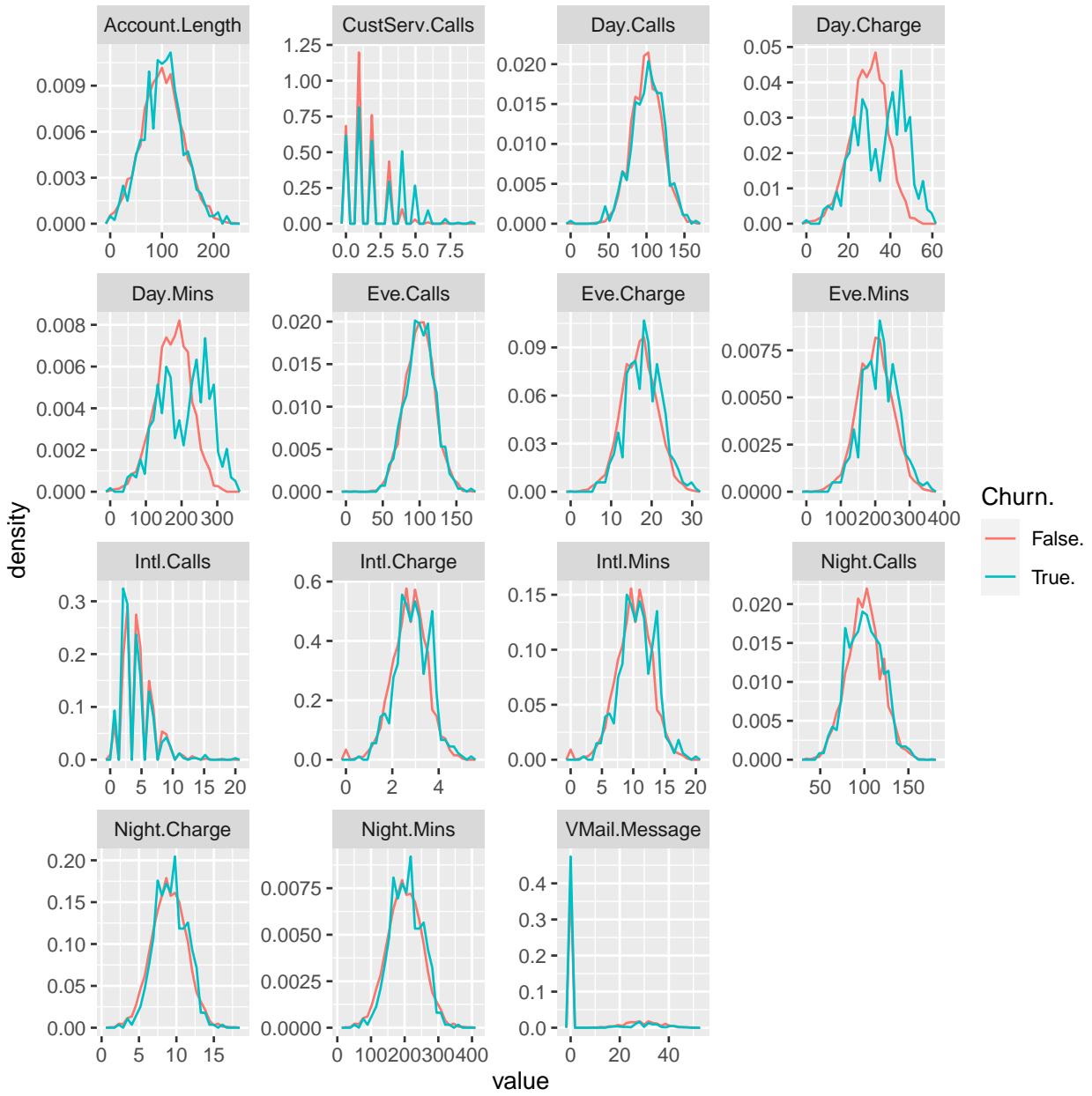
```

ggplot(gather(factors, "key", "value", -Churn.), aes(value, color=Churn.)) +
  geom_bar(position='dodge', aes(y=..count../sum(..count..))) +
  facet_wrap(~key, scales='free')

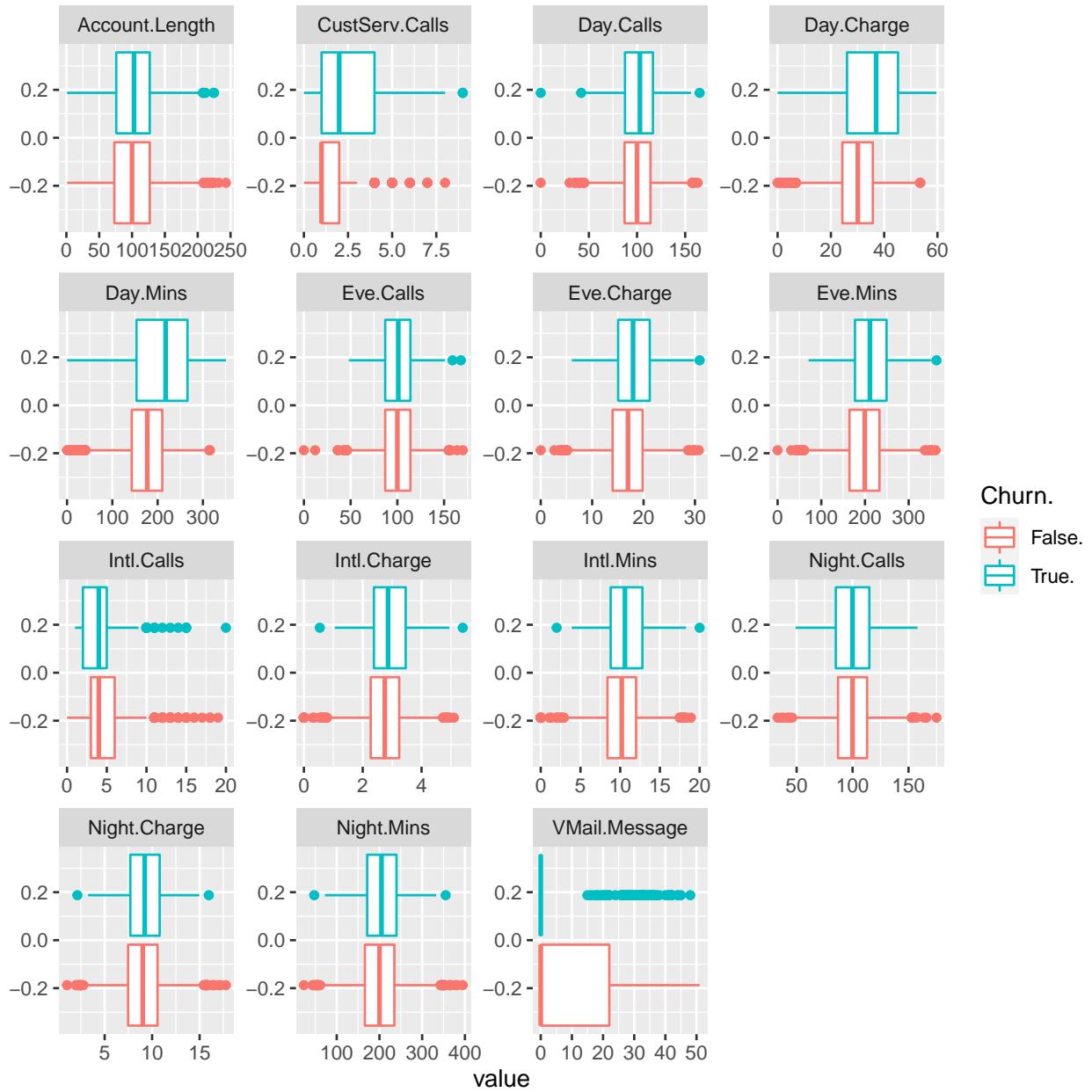
```



```
ggplot(gather(numerics, "key", "value", -Churn.), aes(value, color=Churn.)) +
  geom_freqpoly(aes(y=..density...)) +
  facet_wrap(~key, scales='free')
```



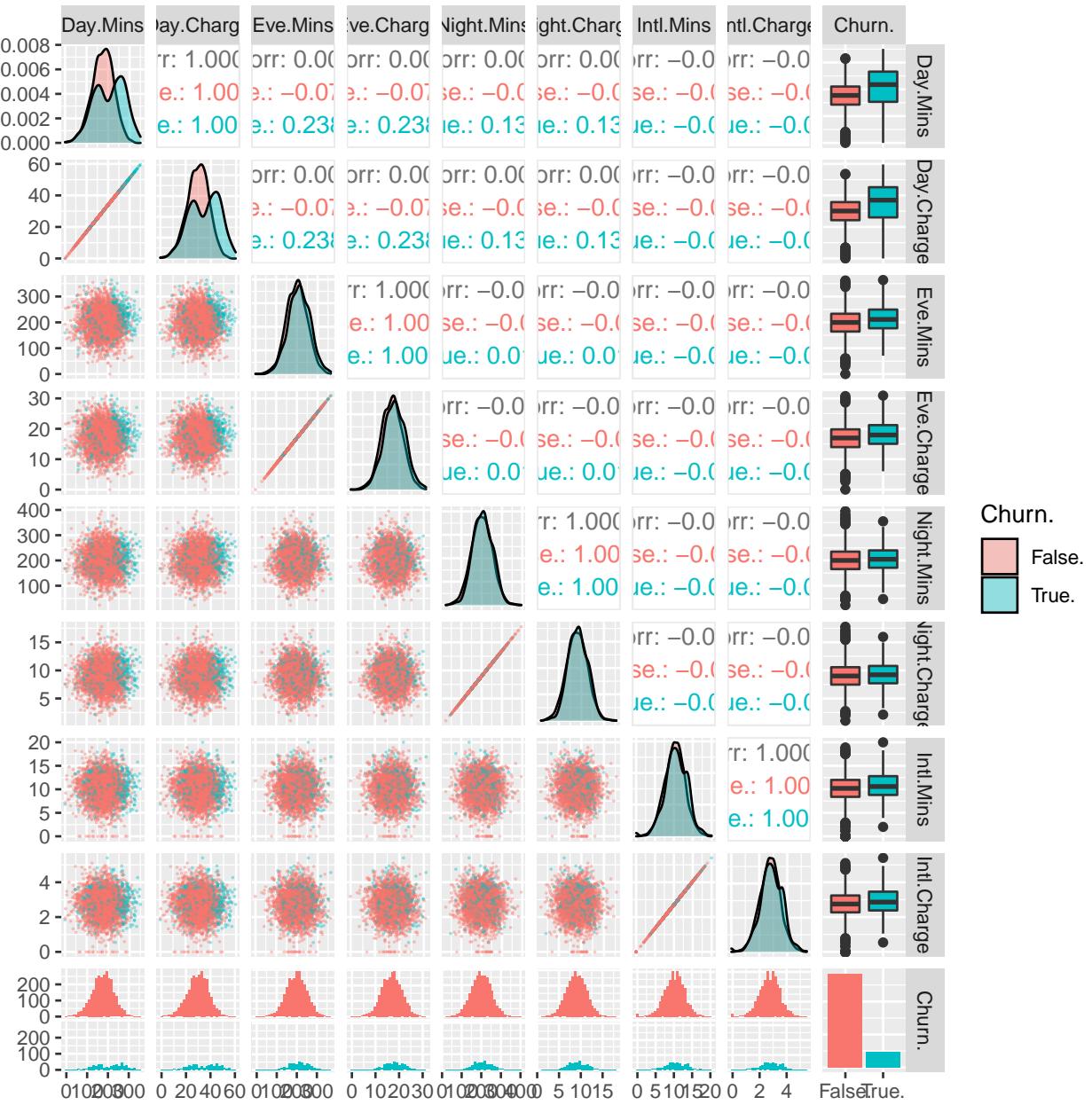
```
ggplot(gather(numerics, "key", "value", -Churn.), aes(value, color=Churn.)) +
  geom_boxplot(aes(x=value)) +
  facet_wrap(~key, scales='free')
```



```

continuous <- subset(numerics, select=sapply(numerics, function(x) !is.integer(x)))
integers <- subset(numerics, select=sapply(numerics, is.integer))
integers <- data.frame(integers, Churn. = continuous$Churn.)
continuous %>% ggpairs(..,
  mapping = ggplot2::aes(color=Churn.),
  legend=1,
  lower=list(continuous=wrap("points", alpha=.4, size=.01)),
  diag=list(continuous=wrap("densityDiag", alpha=.4)))

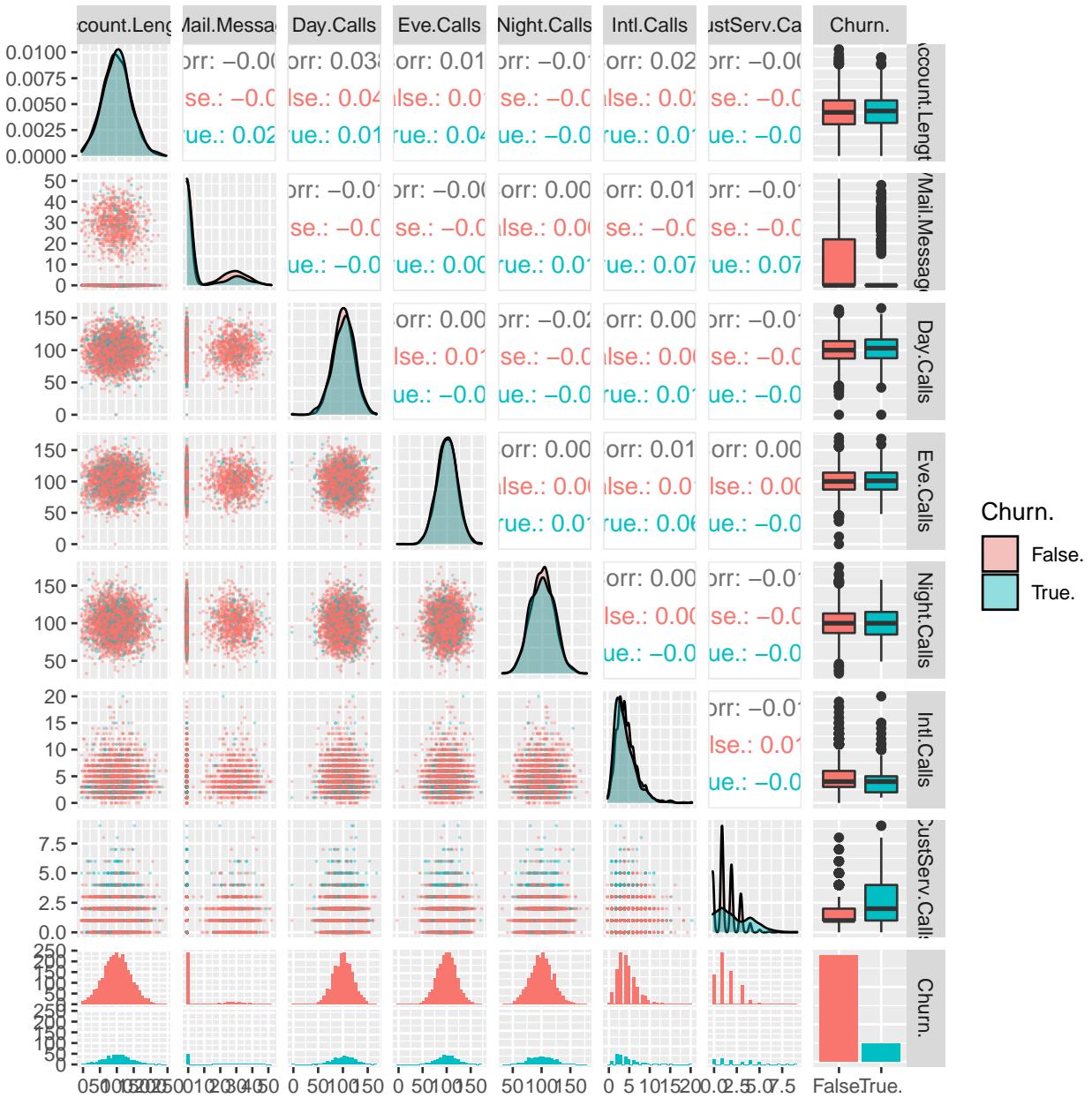
```



```

integers %>% ggpairs(., 
  mapping = ggplot2::aes(color=Churn.),
  legend=1,
  lower=list(continuous=wrap("points", alpha=.4, size=.01)),
  diag=list(continuous=wrap("densityDiag", alpha=.4)))

```



```

aggregate(. ~ Churn., numerics, Range)

##   Churn. Account.Length VMail.Message Day.Mins Day.Calls Day.Charge Eve.Mins
## 1 False.          242           51    315.6     163      53.65   361.8
## 2 True.           224           48    350.8     165      59.64   292.8
##   Eve.Calls Eve.Charge Night.Mins Night.Calls Night.Charge Intl.Mins Intl.Calls
## 1     170     30.75    371.8      142     16.73     18.9       19
## 2     120     24.88    307.5      109     13.84     18.0       19
##   Intl.Charge CustServ.Calls
## 1      5.10            8
## 2      4.86            9

```

```
aggregate(. ~ Churn., factors, levels)
```

```
##   Churn. State Area.Code Int.l.Plan VMail.Plan
## 1 False.    NULL      NULL      NULL      NULL
## 2 True.     NULL      NULL      NULL      NULL
```

```
aggregate(. ~ Churn., numerics, skewness)
```

```
##   Churn. Account.Length VMail.Message Day.Mins Day.Calls Day.Charge
## 1 False.      0.09052327     1.167435 -0.2263781 -0.0632955 -0.2264046
## 2 True.       0.13713205     2.040412 -0.1992228 -0.3533924 -0.1992087
##   Eve.Mins Eve.Calls Eve.Charge Night.Mins Night.Calls Night.Charge
## 1 -0.04313957 -0.06921711 -0.04310337  0.017229862  0.02694851  0.017240153
## 2  0.03312938  0.02985619  0.03314373 -0.005073416  0.06149375 -0.005463696
##   Intl.Mins Intl.Calls Intl.Charge CustServ.Calls
## 1 -0.29377989    1.259044 -0.29390182     0.8868008
## 2  0.02139163    1.714231  0.02100866     0.7036041
```

```
aggregate(. ~ Churn., numerics, cv)
```

```
##   Churn. Account.Length VMail.Message Day.Mins Day.Calls Day.Charge Eve.Mins
## 1 False.      0.3956830     1.616948 0.2864646 0.1974525 0.2864579 0.2526695
## 2 True.       0.3844346     2.318271 0.3334611 0.2129789 0.3334585 0.2435331
##   Eve.Calls Eve.Charge Night.Mins Night.Calls Night.Charge Intl.Mins Intl.Calls
## 1 0.1995071  0.2526678  0.2553551  0.1949489  0.2553574 0.2740942 0.5387146
## 2 0.1961466  0.2435209  0.2296567  0.1987126  0.2296653 0.2610458 0.6128349
##   Intl.Charge CustServ.Calls
## 1  0.2740335     0.8027753
## 2  0.2609934     0.8311344
```

b)

## 4 Etap IV