EDA

Lewei Er

2024-02-21

```
ExpData(data)
```

```
##                                          Descriptions       Value
## 1                                   Sample size (nrow)      978582
## 2                              No. of variables (ncol)          16
## 3                      No. of numeric/interger variables           7
## 4                                No. of factor variables           0
## 5                                  No. of text variables           9
## 6                               No. of logical variables           0
## 7                            No. of identifier variables           1
## 8                                  No. of date variables           0
## 9               No. of zero variance variables (uniform)           1
## 10                 %. of variables having complete cases 68.75% (11)
## 11   %. of variables having >0% and <50% missing cases      0% (0)
## 12 %. of variables having >=50% and <90% missing cases   12.5% (2)
## 13          %. of variables having >=90% missing cases  18.75% (3)
```

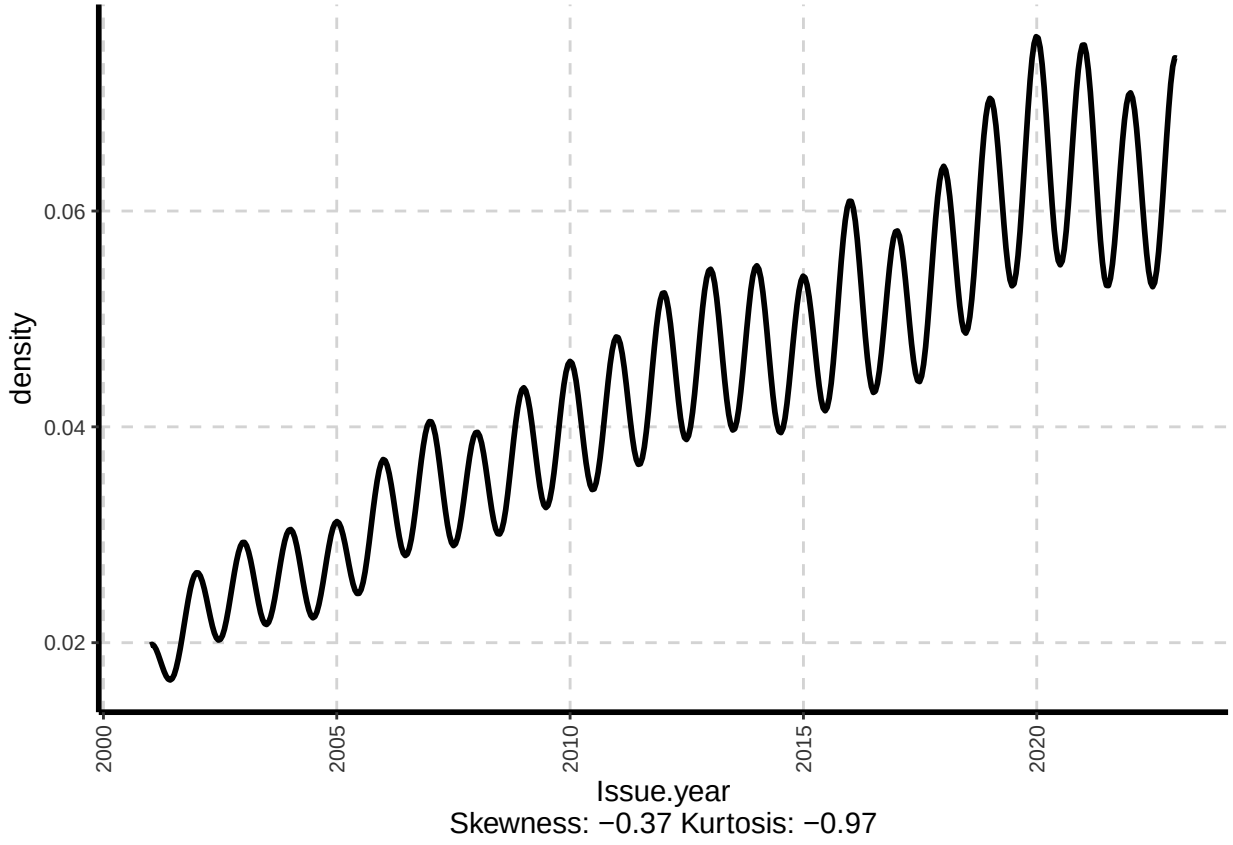
```
ExpData(data, type = 2)
```

```
##    Index        Variable_Name Variable_Type Sample_n Missing_Count
## 1      1        Policy.number     character   978582             0
## 2      2          Policy.type     character   978582             0
## 3      3           Issue.year       integer   978582             0
## 4      4            Issue.age       integer   978582             0
## 5      5                  Sex     character   978582             0
## 6      6          Face.amount       integer   978582             0
## 7      7        Smoker.Status     character   978582             0
## 8      8   Underwriting.Class     character   978582             0
## 9      9       Urban.vs.Rural     character   978582             0
## 10    10               Region       integer   978582             0
## 11    11 Distribution.Channel     character   978582             0
## 12    12      Death.indicator       integer    40376        938206
## 13    13        Year.of.Death       integer    40376        938206
## 14    14      Lapse.Indicator     character   110889        867693
## 15    15        Year.of.Lapse       integer   110889        867693
## 16    16       Cause.of.Death     character    39540        939042
##    Per_of_Missing No_of_distinct_values
## 1           0.000                978582
## 2           0.000                     2
## 3           0.000                    23
## 4           0.000                    40
## 5           0.000                     2
## 6           0.000                     6
```

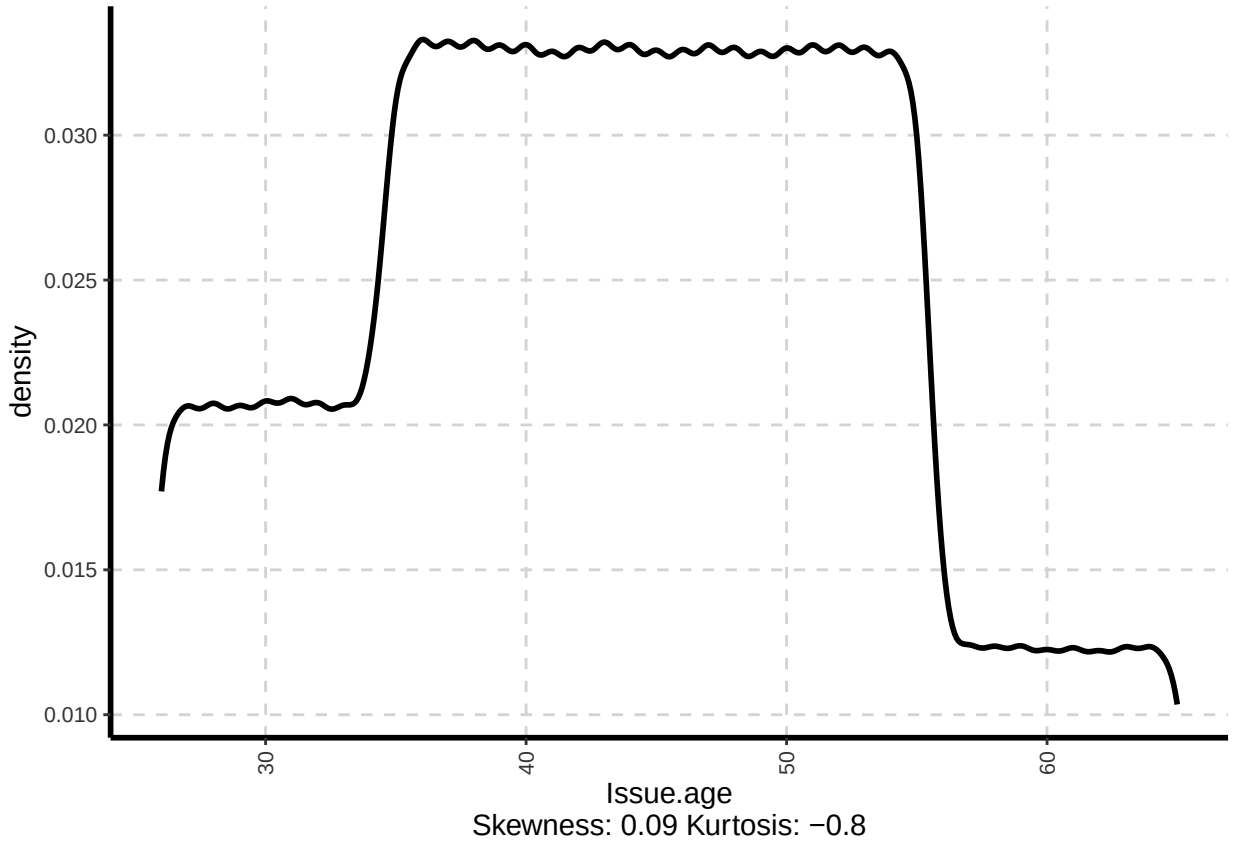
```
## 7           0.000                   2
## 8           0.000                   4
## 9           0.000                   2
## 10          0.000                   6
## 11          0.000                   3
## 12          0.959                   1
## 13          0.959                  23
## 14          0.887                   3
## 15          0.887                  23
## 16          0.960                  17
```

```
ExpNumViz(data)
```

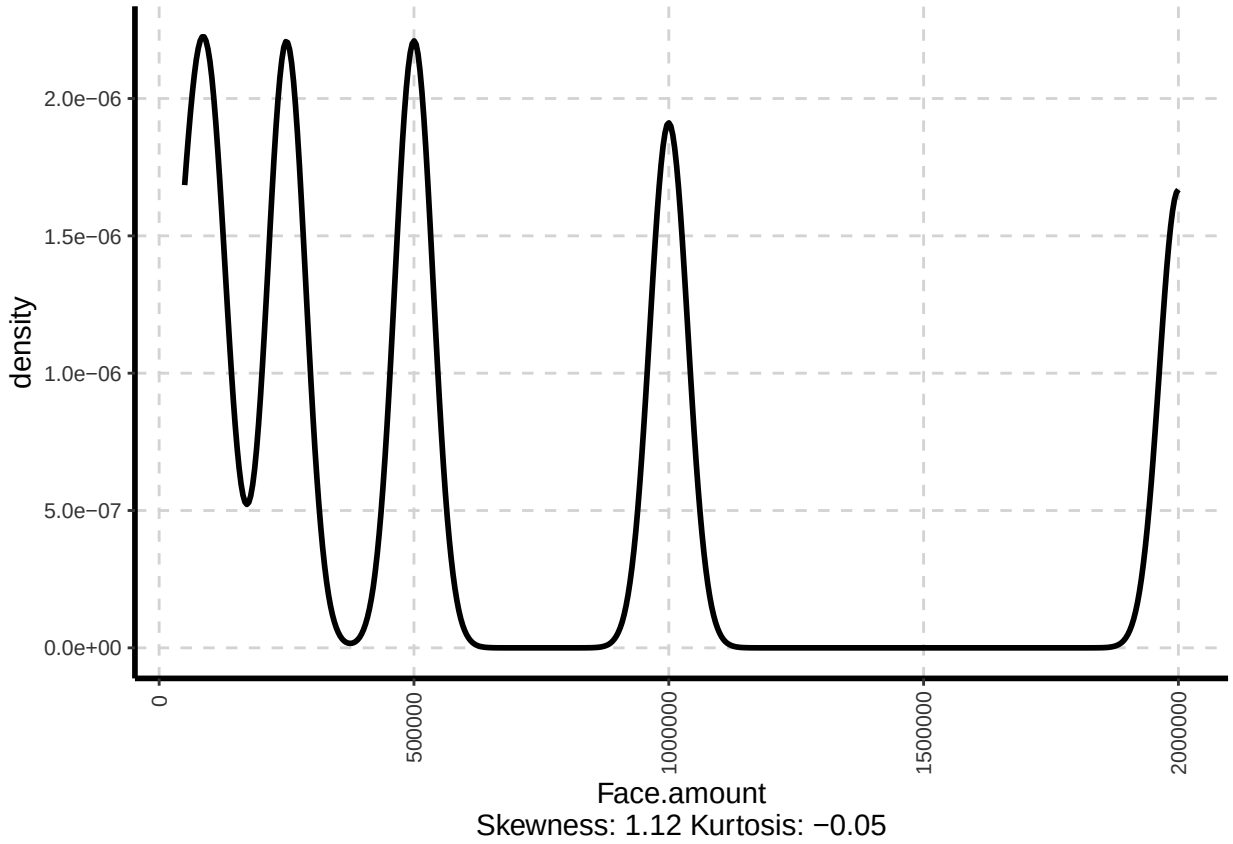
```
## [[1]]
```



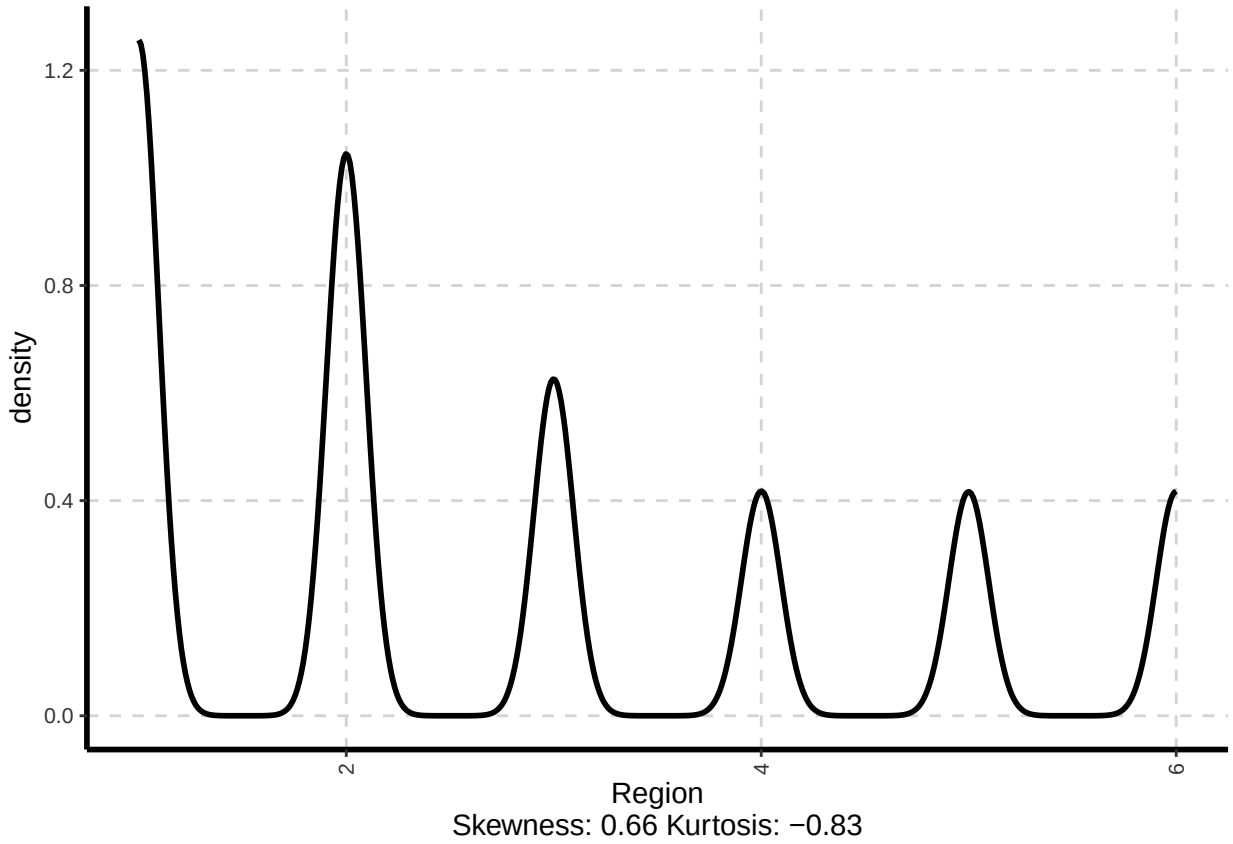

```
##
## [[2]]
```


0.030
0.025
0.020
0.015
0.010
density
30
40
50
60
Issue.age
Skewness: 0.09 Kurtosis: −0.8

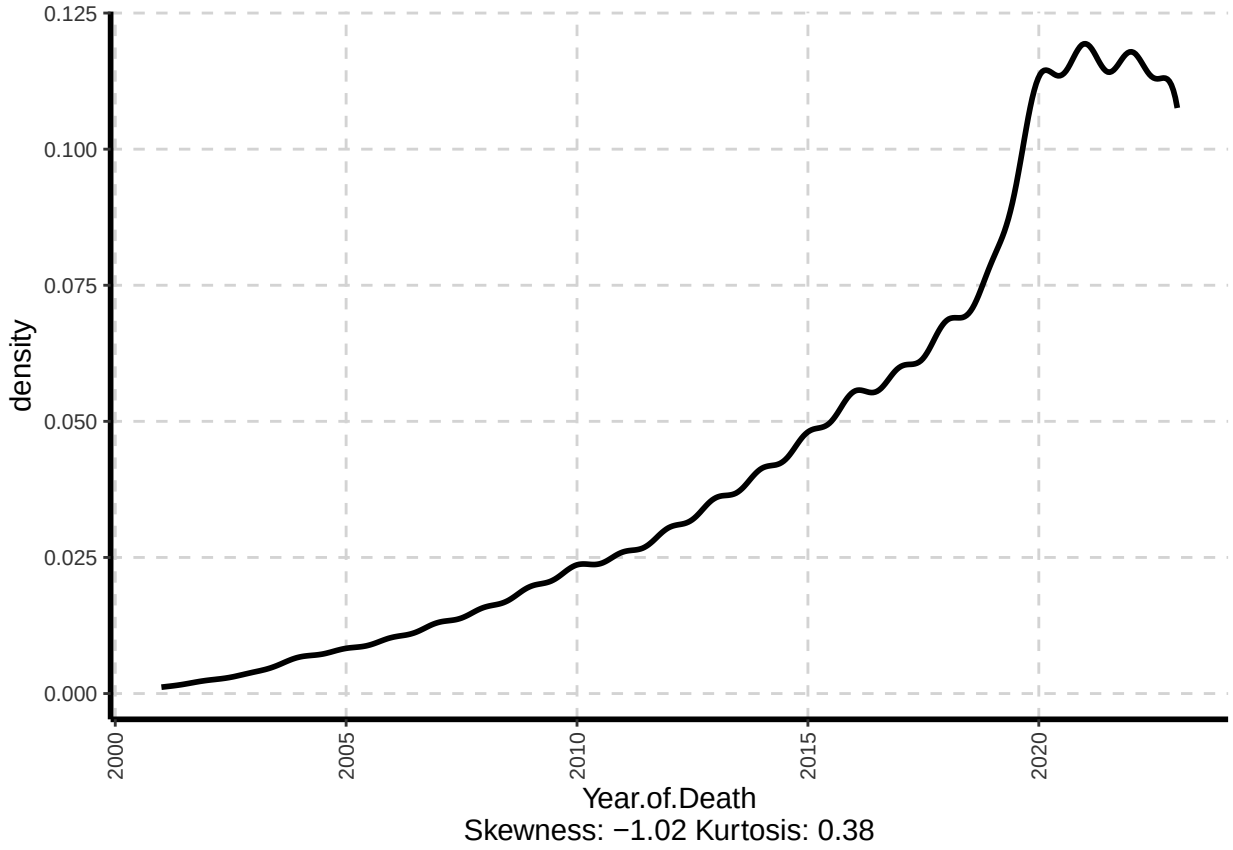
```
##
## [[3]]
```


2.0e−06
1.5e−06
1.0e−06
5.0e−07
0.0e+00
density
0
500000
1000000
1500000
2000000
Face.amount
Skewness: 1.12 Kurtosis: −0.05

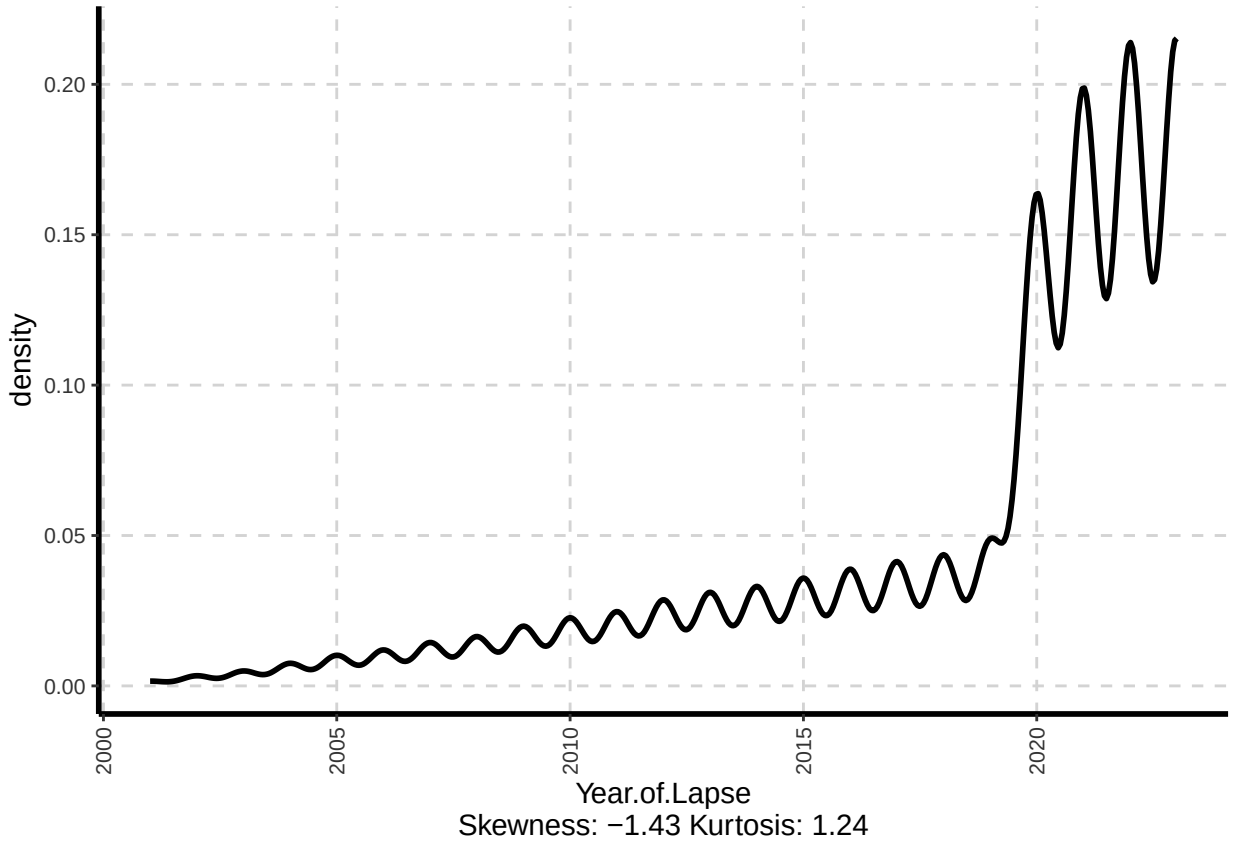
```
##
## [[4]]
```




```
##
## [[5]]
```

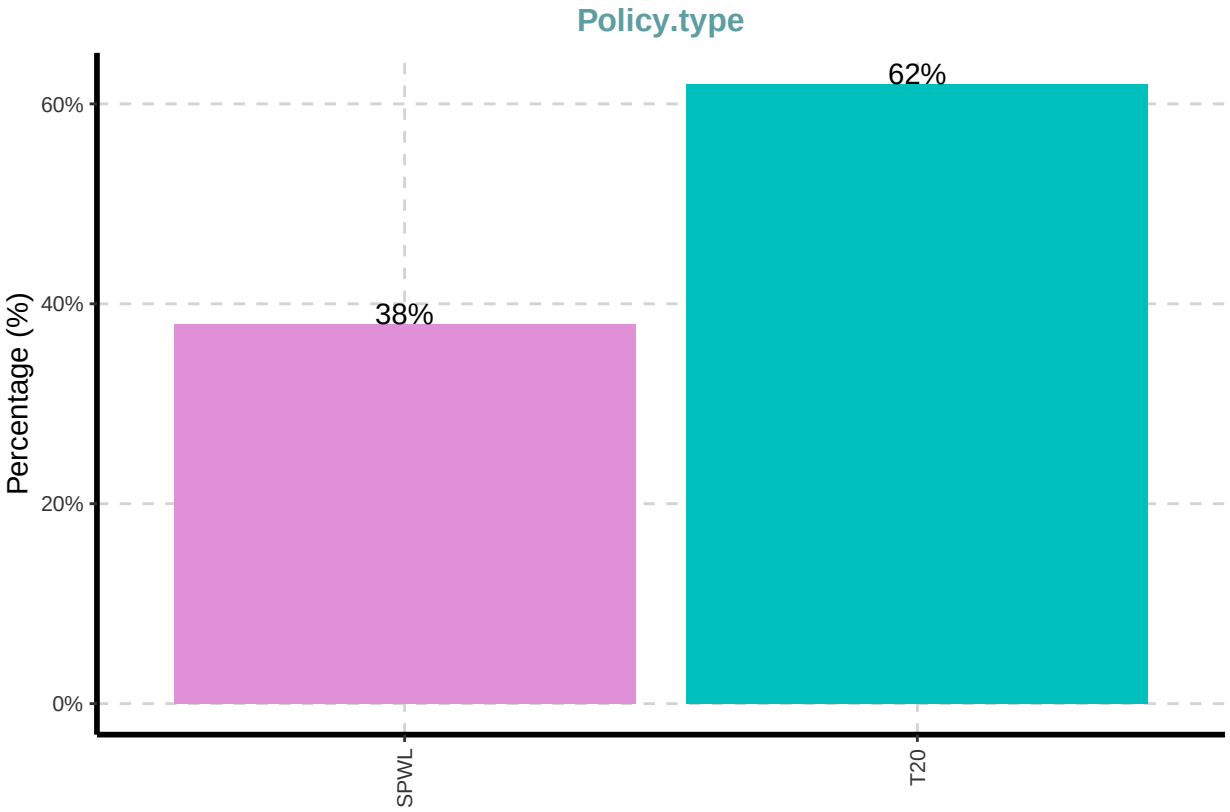

0.125
0.100
0.075
0.050
0.025
0.000
density
2000
2005
2010
2015
2020
Year.of.Death
Skewness: −1.02 Kurtosis: 0.38

```
##
## [[6]]
```

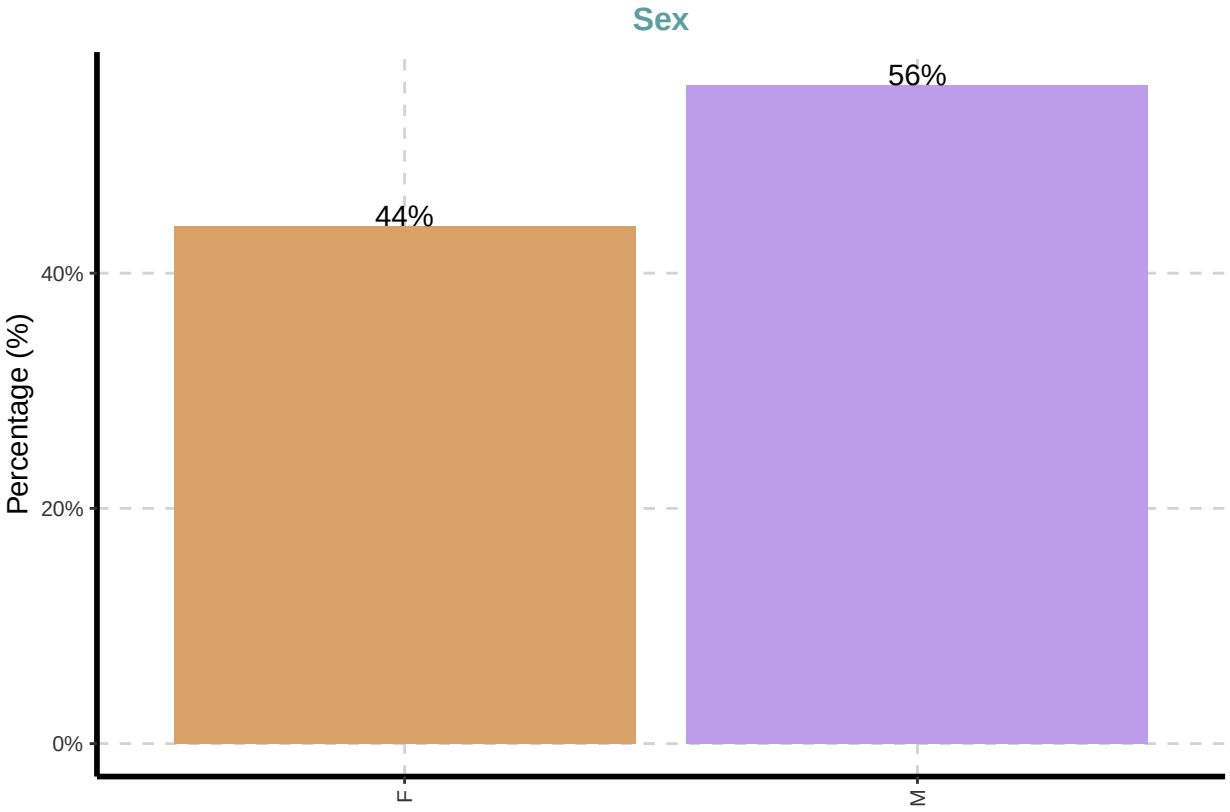



```
ExpCatViz(data)
```

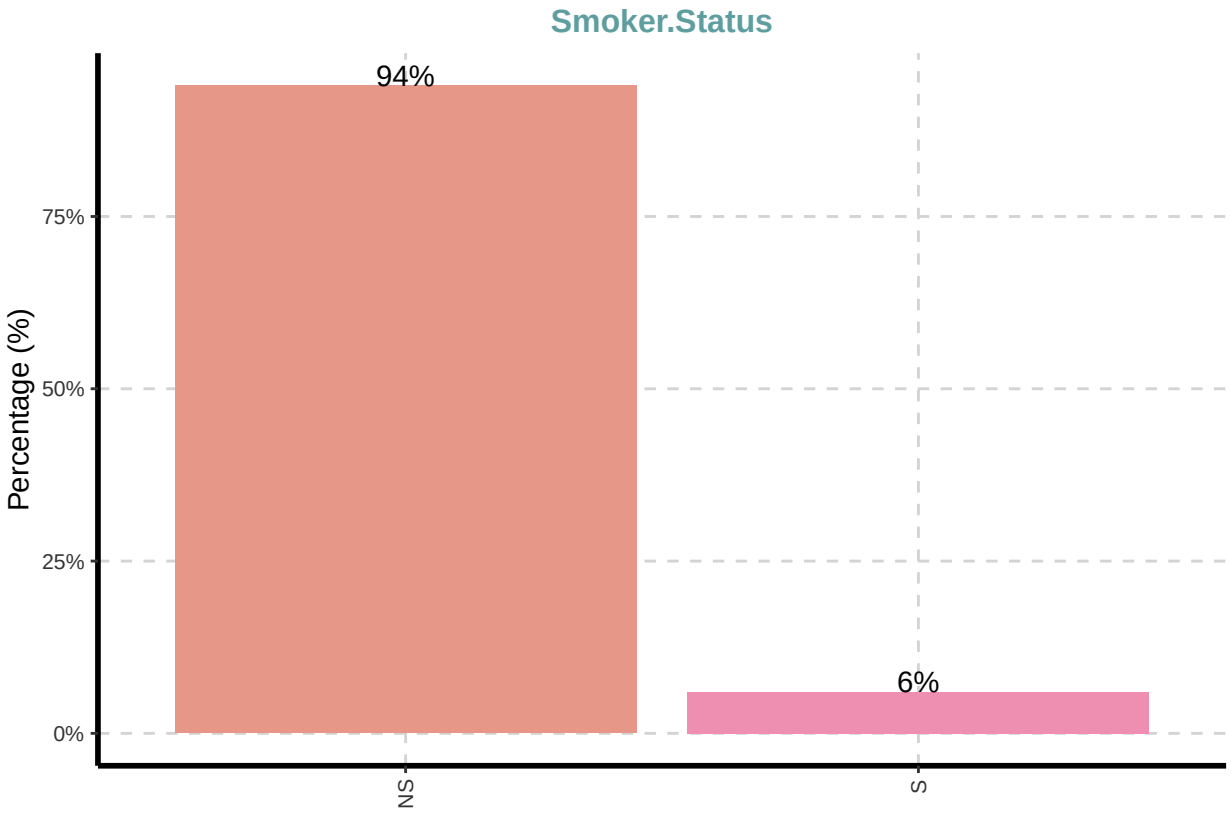
```
## [[1]]
```


Policy.type
62%
38%
60%
40%
20%
0%
Percentage (%)
SPWL
T20

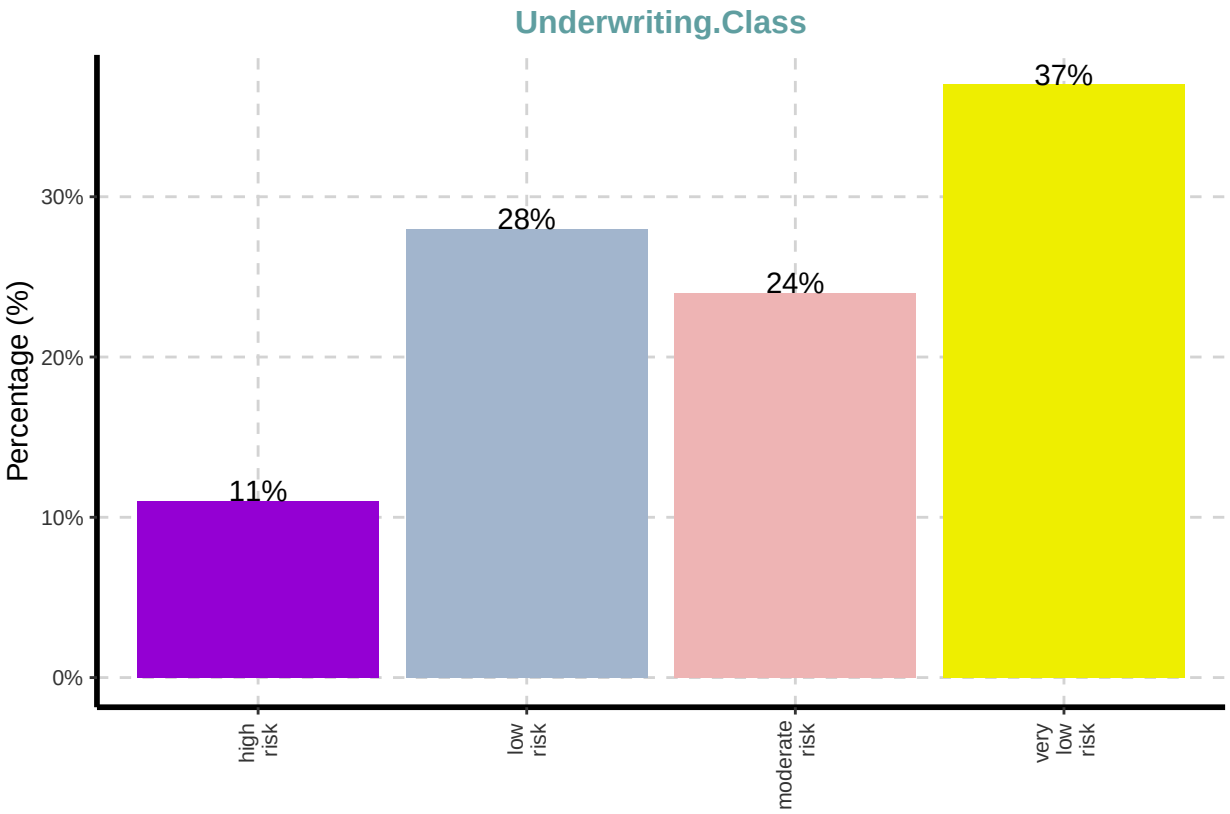
```
##
## [[2]]
```


Sex
56%
44%
40%
20%
0%
Percentage (%)
F
M

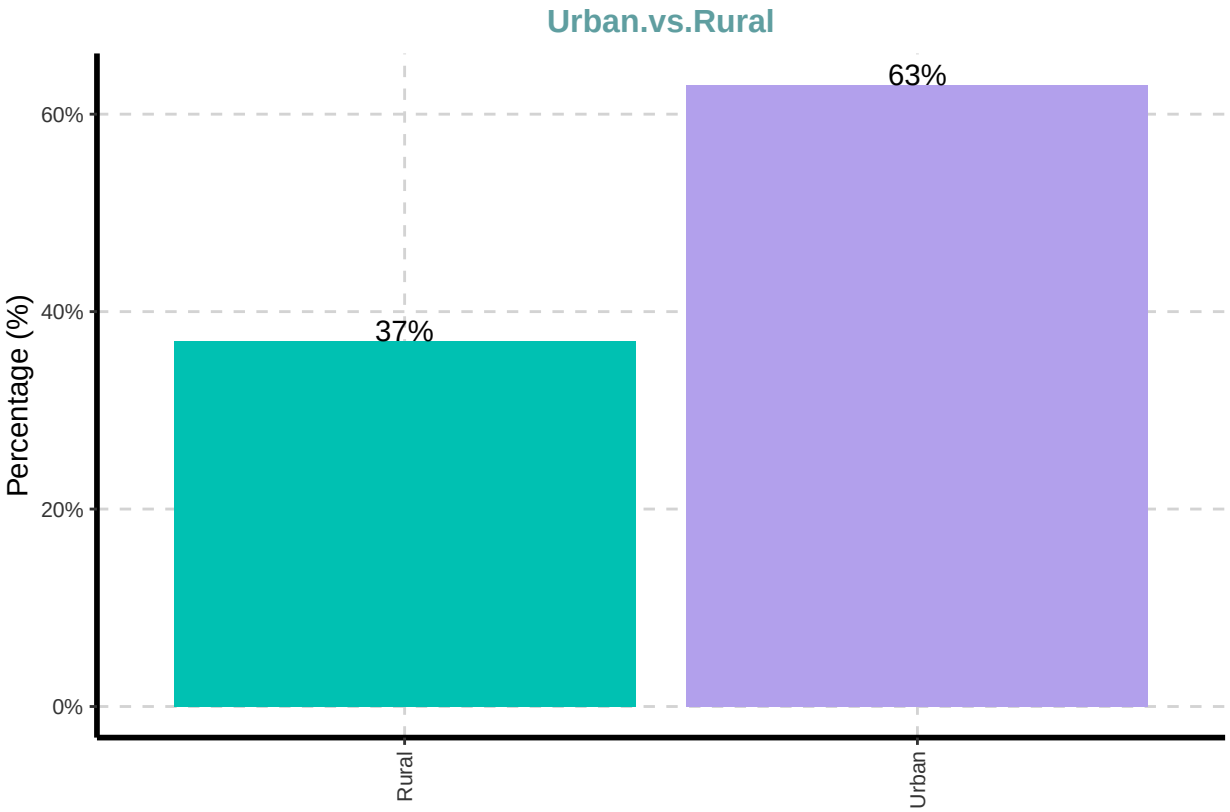
```
##
## [[3]]
```



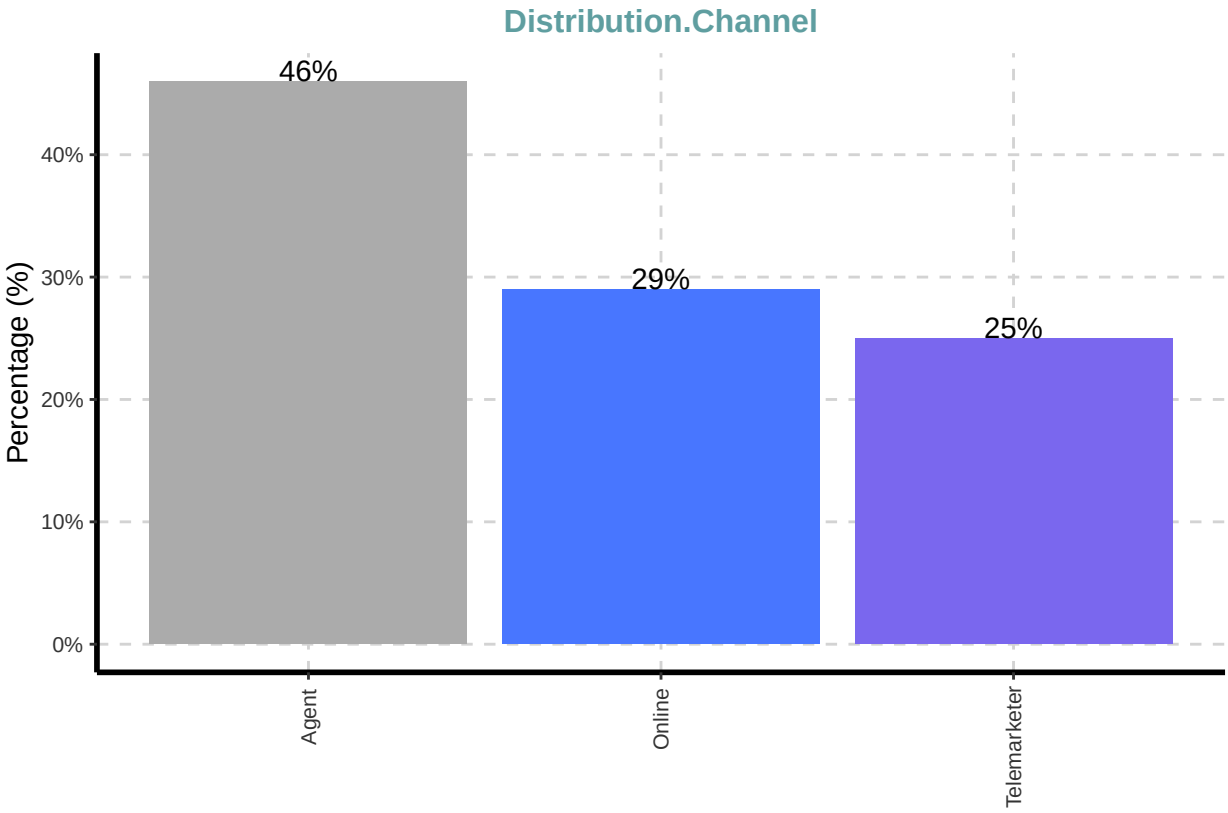

```
##
## [[4]]
```


Underwriting.Class
Percentage (%)
30%
20%
10%
0%
11%
28%
24%
37%
high risk
low risk
moderate risk
very low risk

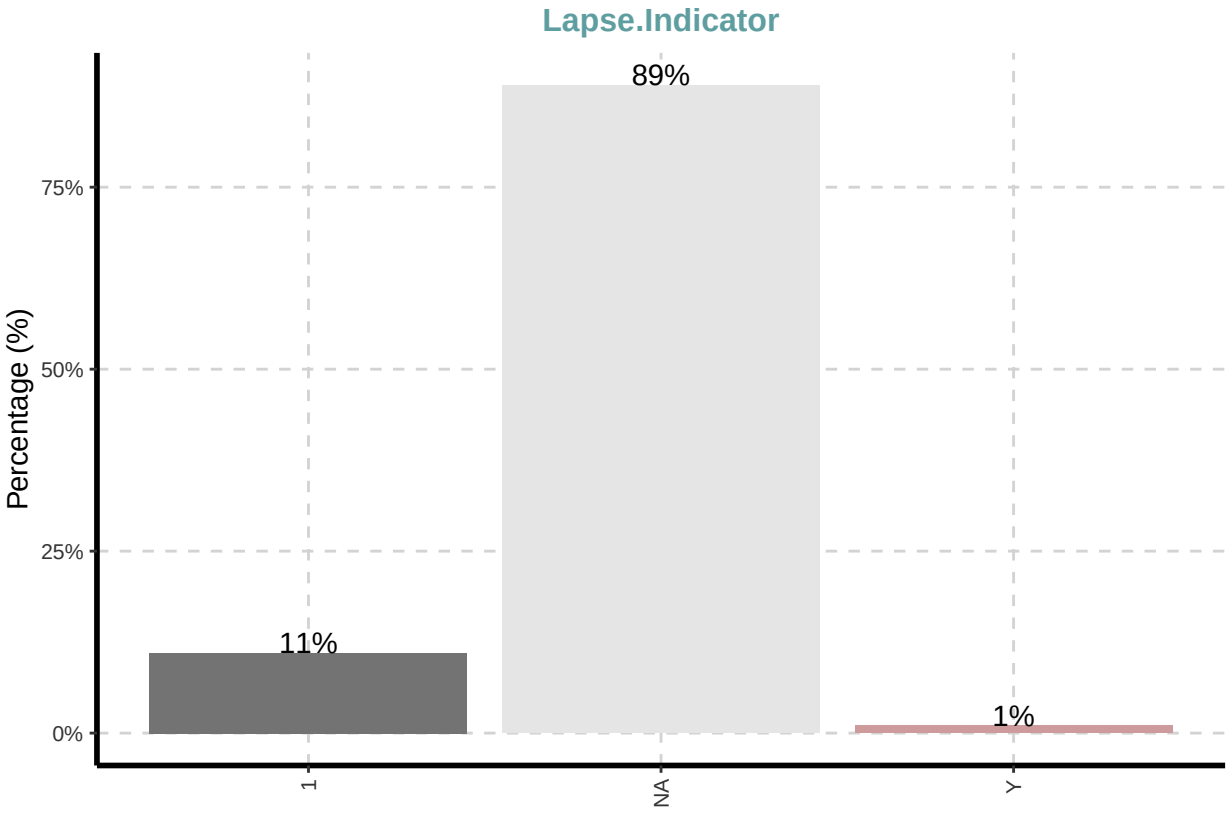
```
##
## [[5]]
```


Urban.vs.Rural
Percentage (%)
60%
40%
20%
0%
37%
63%
Rural
Urban

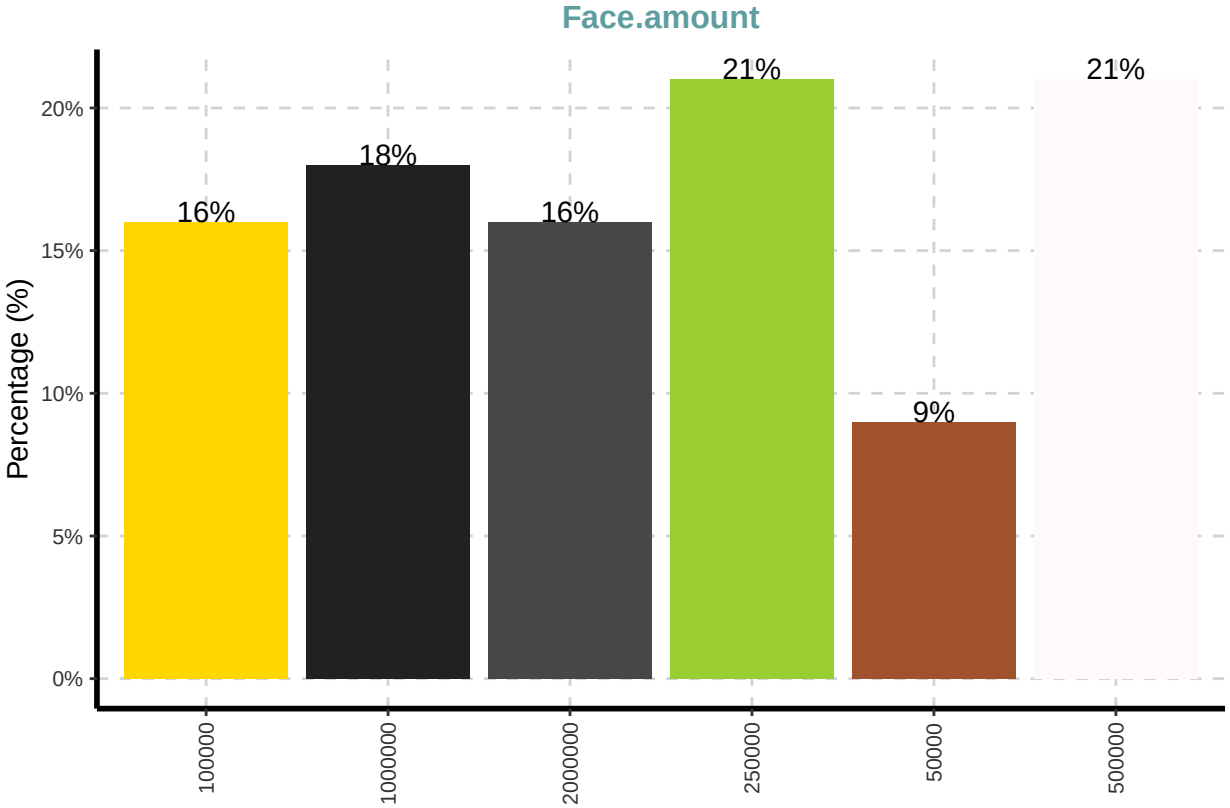
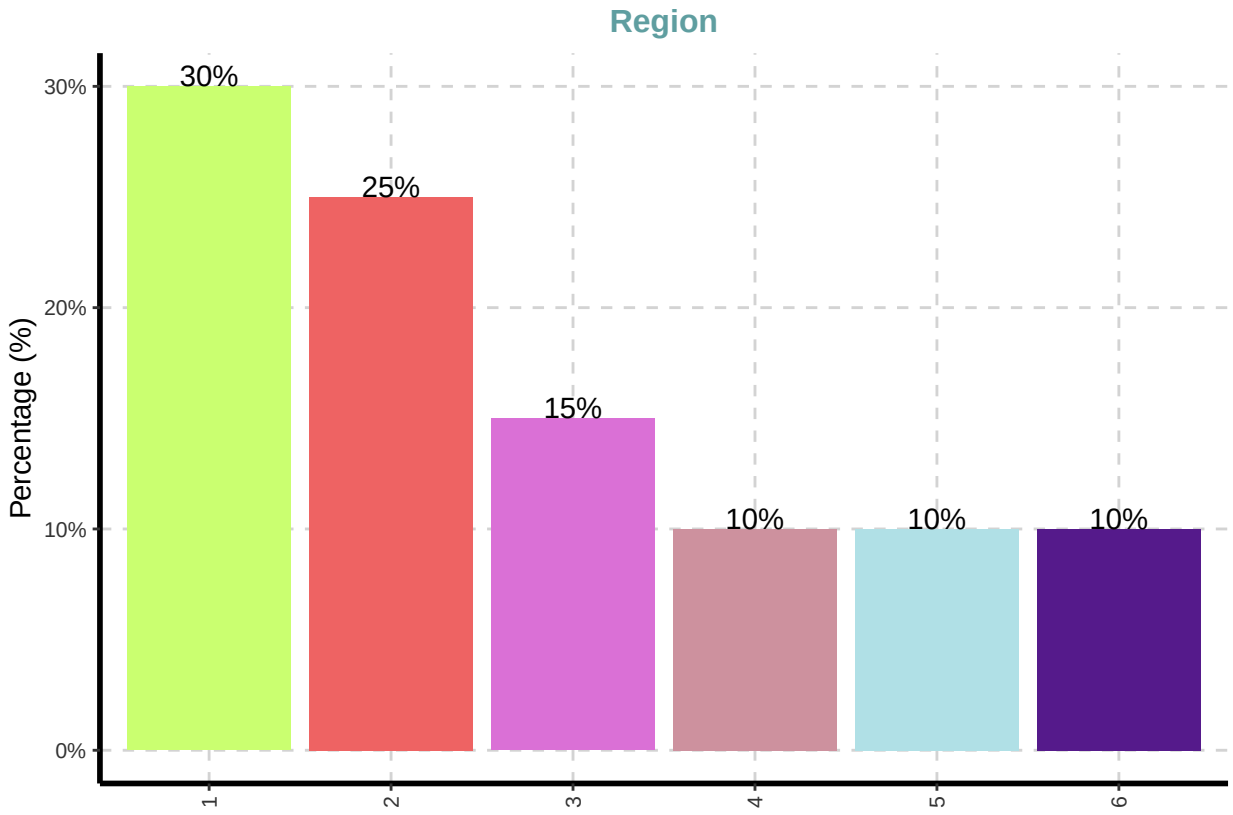
```
##
## [[6]]
```


Distribution.Channel
Percentage (%)
0%
10%
20%
30%
40%
46%
29%
25%
Agent
Online
Telemarketer

```
##
## [[7]]
```


Lapse.Indicator
Percentage (%)
75%
50%
25%
0%
89%
11%
1%
1
NA
Y

```
##
## [[8]]
```


Face.amount
Percentage (%)
20%
15%
10%
5%
0%
16%
18%
16%
21%
9%
21%
100000
1000000
2000000
250000
50000
500000

```
##
## [[9]]
```