# Table of Content

# Exploring Data

## Summarizing Data

- Examining the "age" and "balance" averages of each "Marital Status" group

| **Marital** | **Age Average** | **Balance Average** |
| --- | --- | --- |
| **divorced** | 45.475 | 1122.390 |
| **married** | 43.454 | 1463.196 |
| **single** | 33.927 | 1460.415 |

- Examining the "age" and "balance" averages of each "Job" who subscribed in this campaign

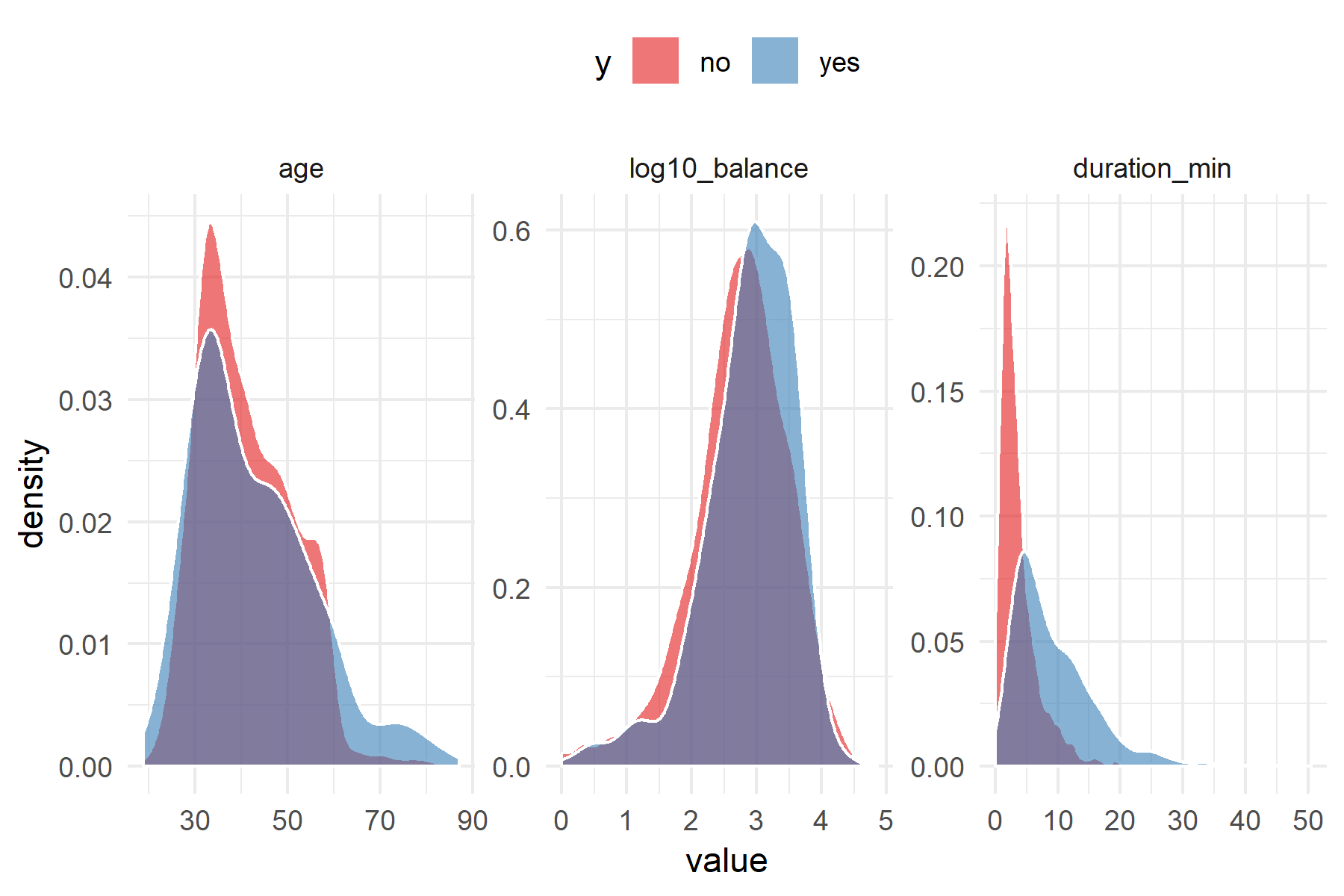
| **Job** | **Y** | **Age Average** | **Balance Average** |
| --- | --- | --- | --- |
| **housemaid** | yes | 45.714 | 3900.357 |
| **retired** | yes | 67.259 | 2480.333 |
| **management** | yes | 39.115 | 1733.550 |
| **self-employed** | yes | 45.550 | 1546.700 |
| **technician** | yes | 40.398 | 1425.795 |
| **unknown** | yes | 52.000 | 1349.857 |
| **unemployed** | yes | 40.308 | 1336.077 |
| **admin.** | yes | 41.207 | 1331.638 |
| **student** | yes | 24.211 | 1198.263 |
| **services** | yes | 37.868 | 1112.658 |
| **blue-collar** | yes | 38.667 | 1026.493 |
| **entrepreneur** | yes | 42.200 | 943.133 |

- Examining the "duration" of each "Contact Method"

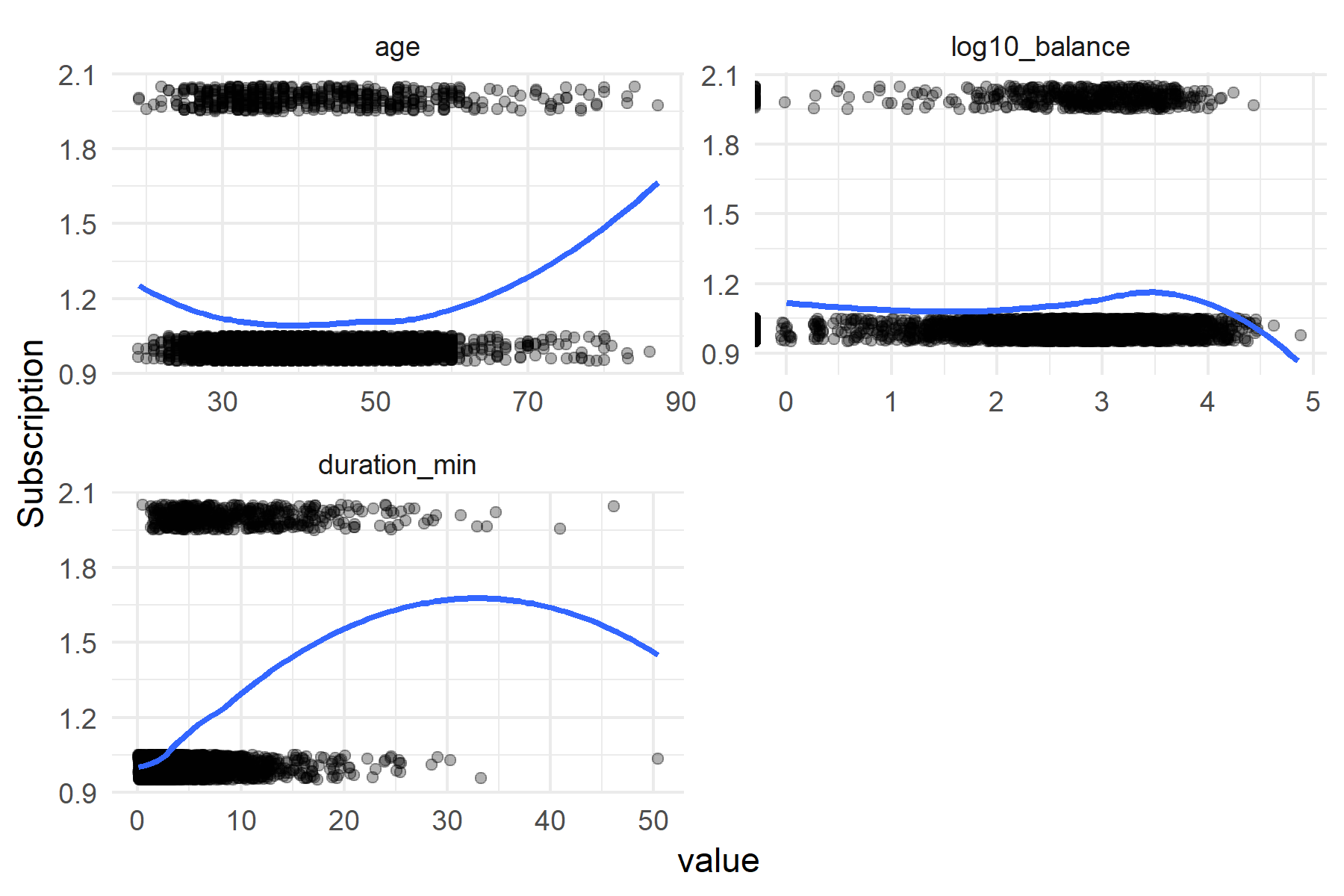
| **Contact** | **Duration Average Min** |
| --- | --- |
| **cellular** | 4.452 |
| **telephone** | 4.056 |
| **unknown** | 4.363 |

## Visualizing Data

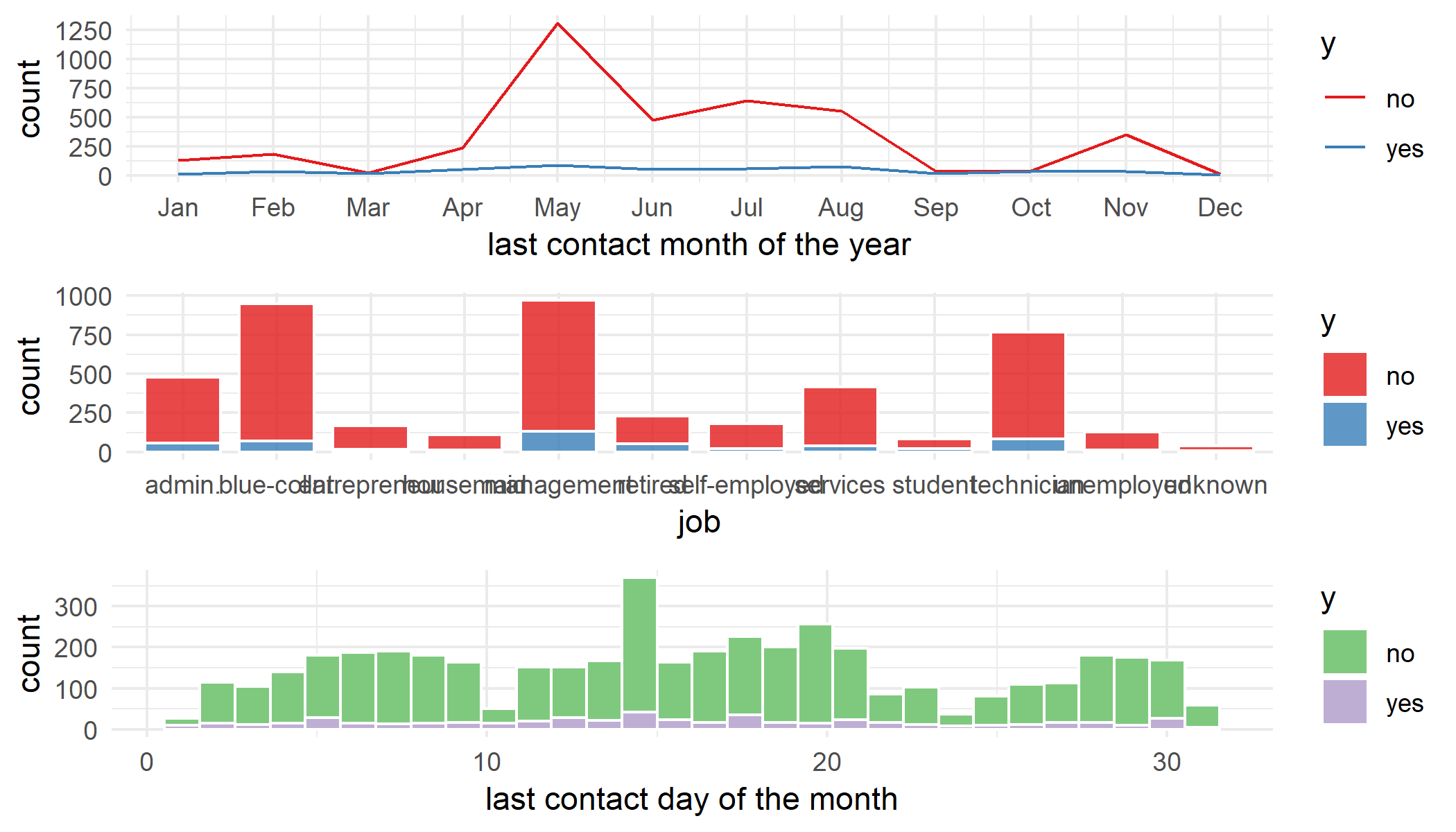
- Visualizing distribution of numeric columns:



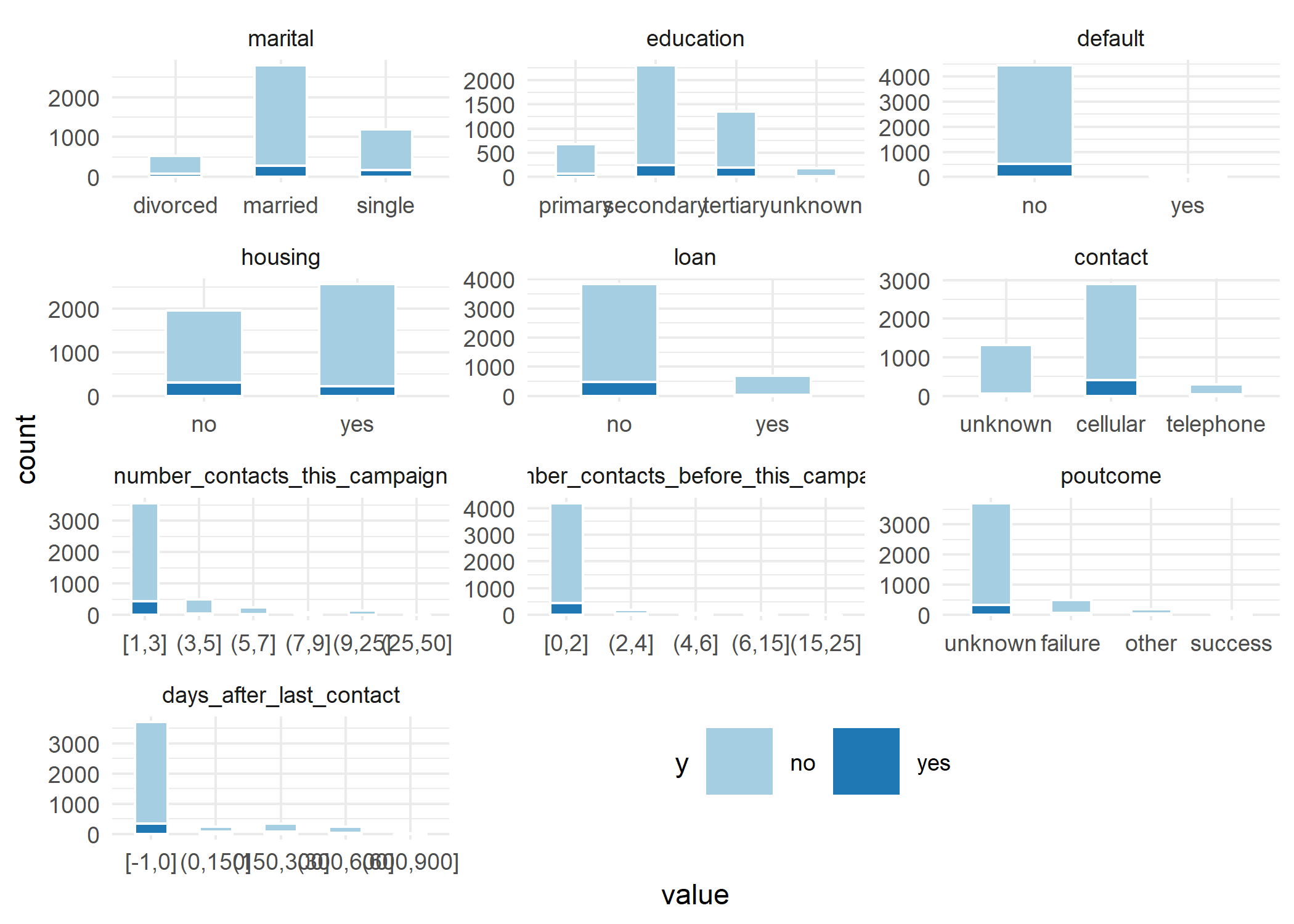
- Visualizing relationship between numeric columns and y:



- spotting days, months and jobs in which users tend to subscribe



- Visualizing frequency of categorical variables and their relationship with y:



# Modeling Data

## Preparing data for modeling

- one hot encoding of categorical variables and normalizing numeric ones

dimenstions of original data: 4521 , 17

dimenstions of new encoded data: 4521 , 95

## Train/Test Splitting

- splitting data to train and test datasets (80/20)

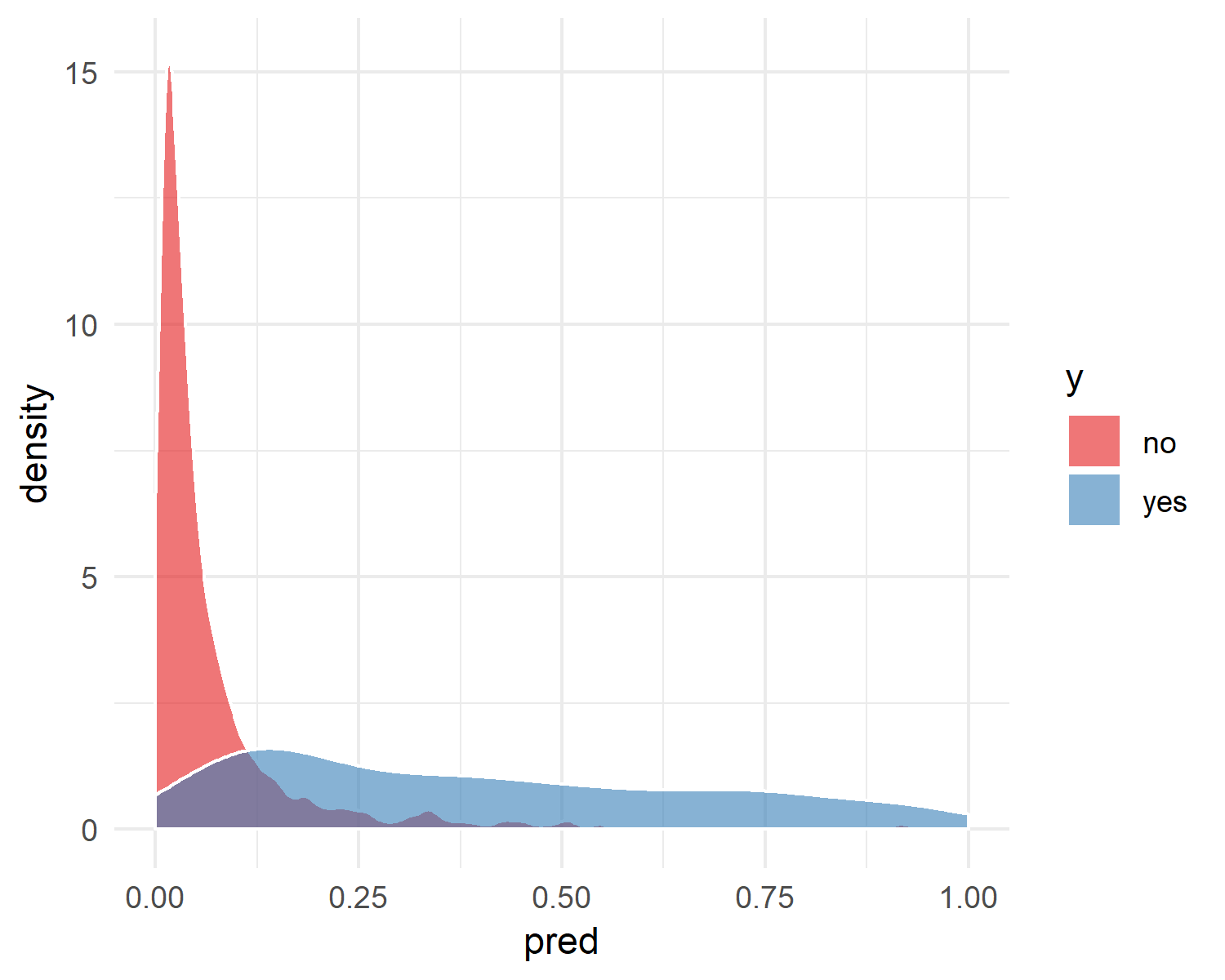
nrows of training dataset: 3616

nrows of test dataset: 905

## Fitting Classification Model

- Applying Logistic Regression algorithm to data:

Spotting Classification Threshold Probability Value:



The model seems to classify well at probablity: 0.15

## Evaluating the Model

- Measuring Model Accuracy:

|  |  |
| --- | --- |
| **train accuracy** | 0.86 |
| **test accuracy** | **0.87** |