A Logistic Regression Approach to CoIL Challenge 2000

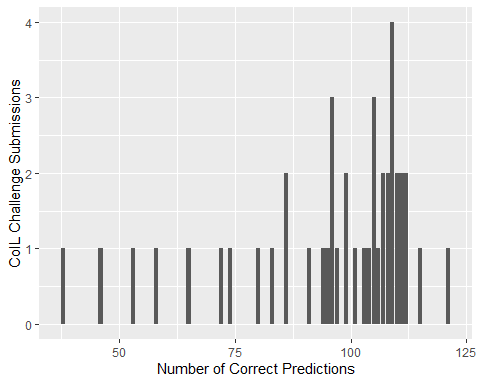
Corey Arnouts, Adam Douglas, Jason Givens-Doyle, Michael Silva

A logistic regression based solution to the CoIL Challenge 2000 is described. The challenge consists of correctly identifying potential customers for an insurance product, and describing their characteristics.

# Introduction

Businesses use data science to extract insights from data. It has many practical business applications. Identifying households to include in a marketing campaign is one application. One example using real world data is the Computational Intelligence and Learning (CoIL) Challenge. The CoIL Challenge competition was held from March 17 to May 8 in 2000. The challenge is to:

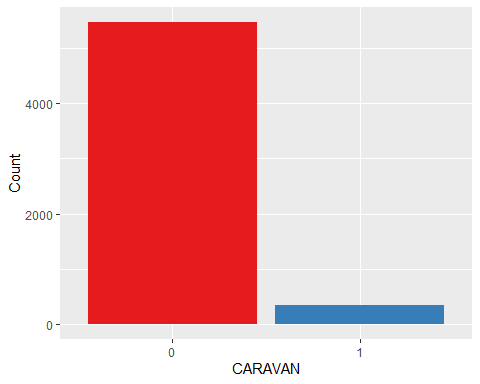
1. Identify potential customers for an insurance policy; and
2. Provide a description of this customer base.



In total 147 participants registered and 43 submitted solutions (Putten, Ruiter, and Someren 2000). The maximum number of policyowners that could be found was 238. The submissions identified 95 policy owners on average. The winning model (Elkan 2000) identified 121 policy owners. Random selection results in identifying 42 policy owners. The standard benchmark tests result in 94 (k-nearest neighbor), 102 (naïve bayes), 105 (neural networks) and 118 (linear) policy owners. (Putten, Ruiter, and Someren 2000). In this paper we set out to complete the first part of the COIL Challenge.

# Exploratory Data Analysis

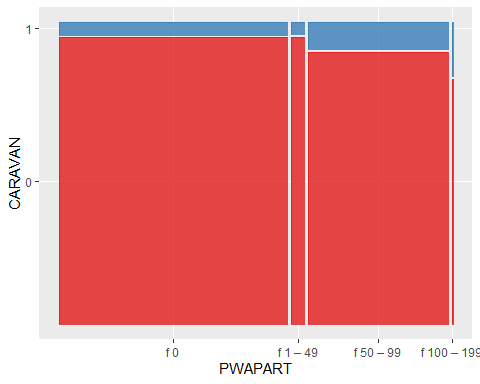
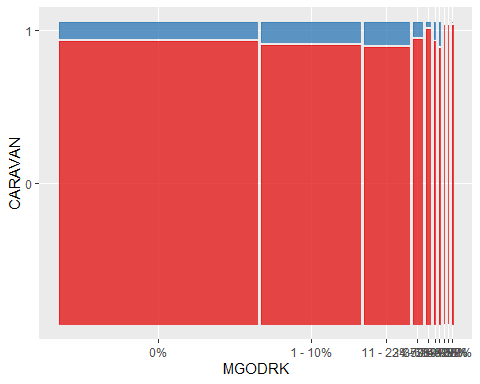
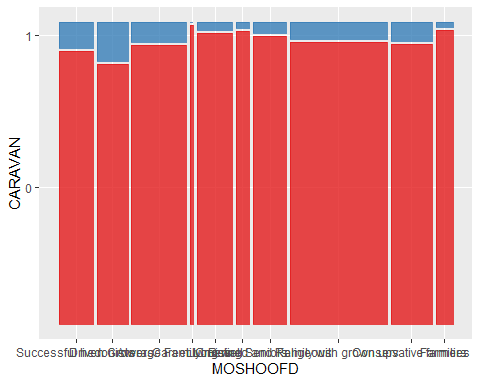
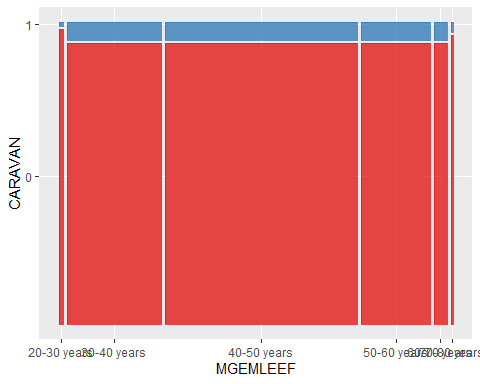
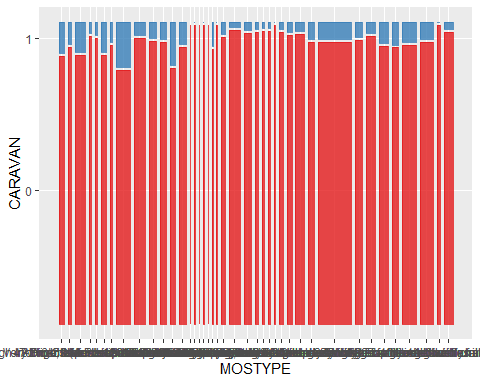
## Response Variable CARAVAN



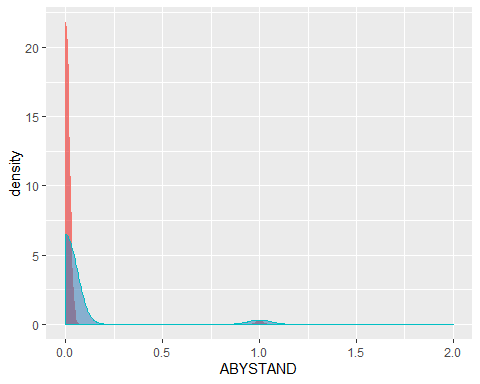
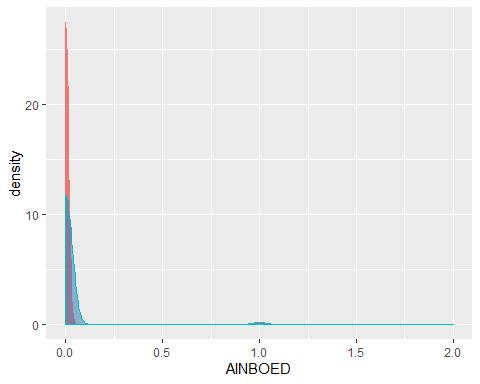
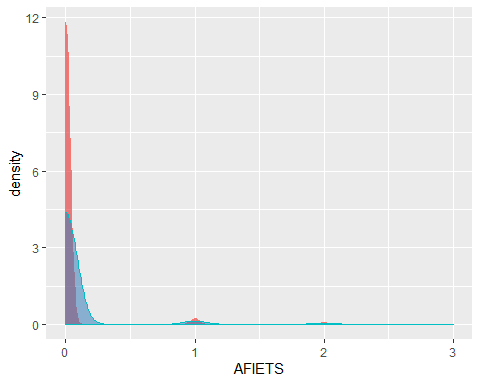
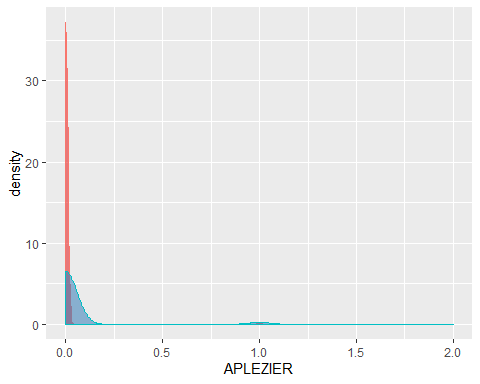
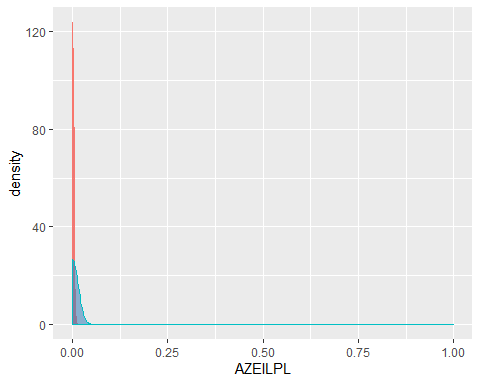
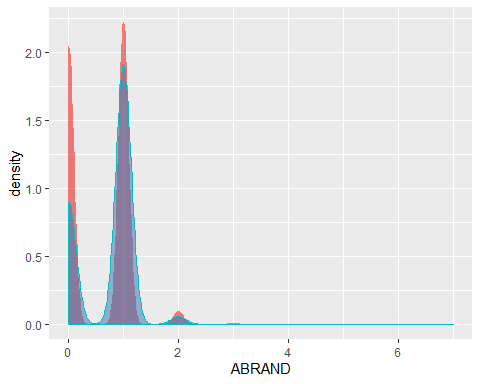
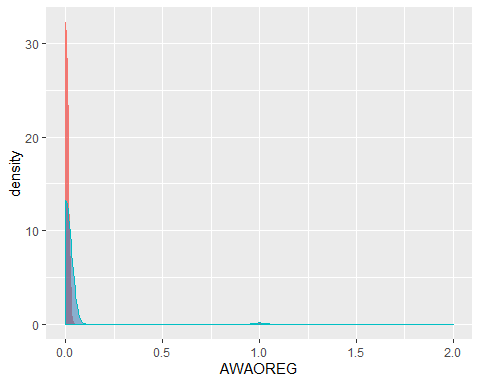
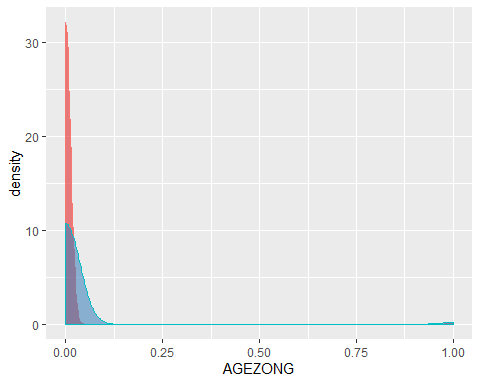
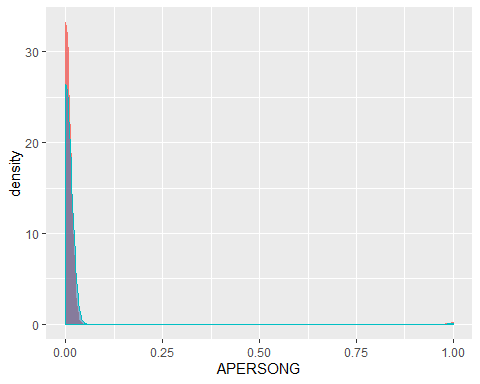
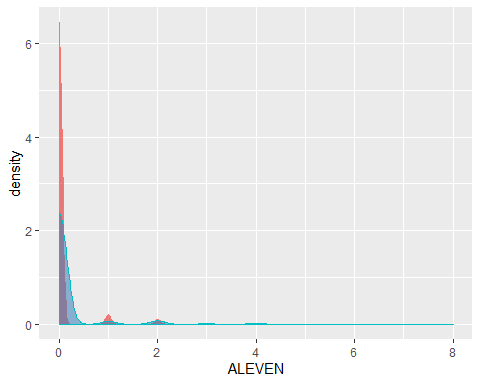
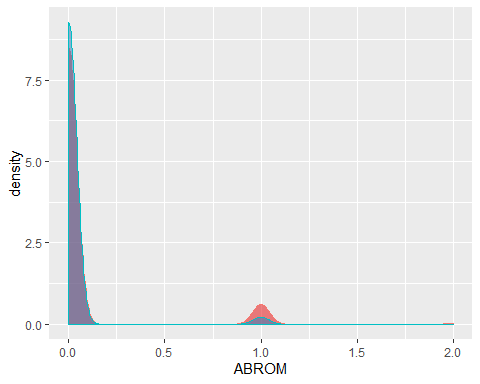
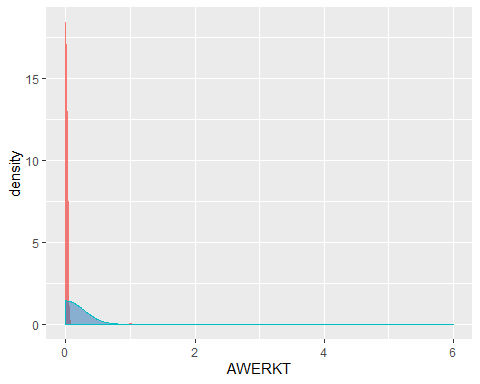
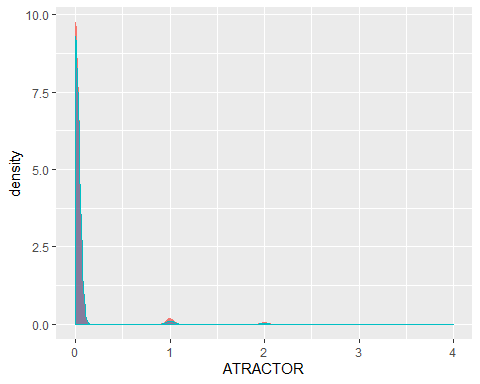
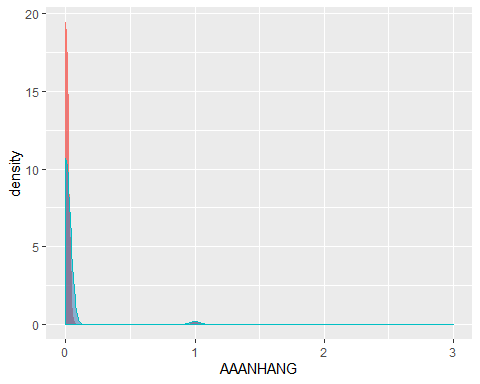
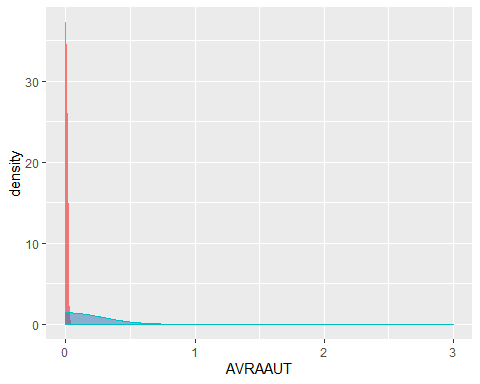
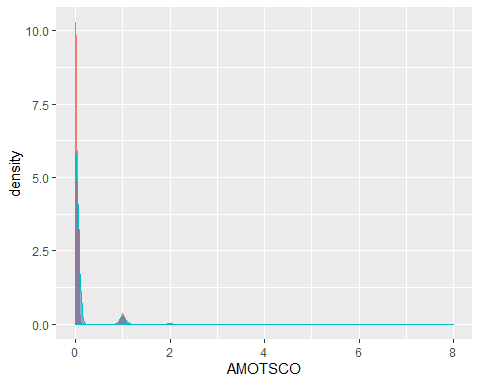
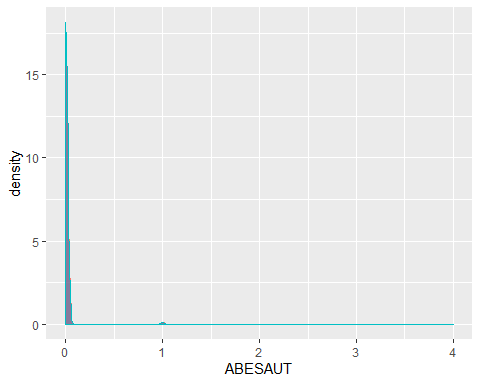
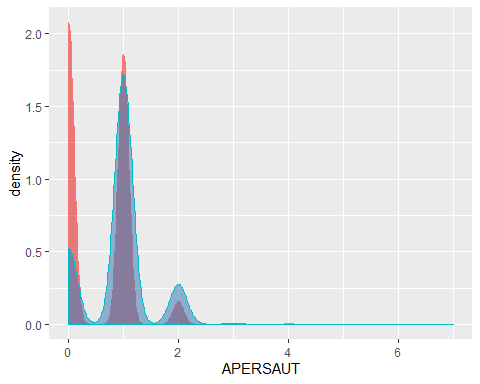
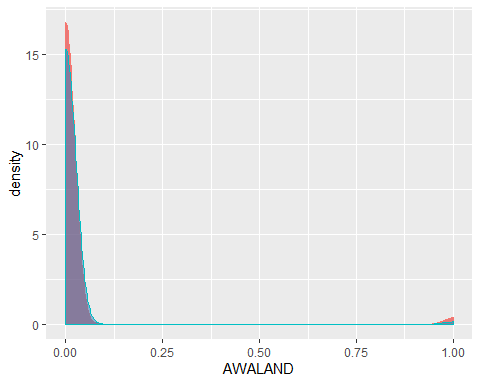
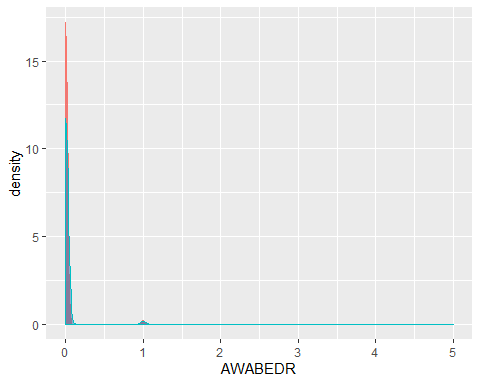
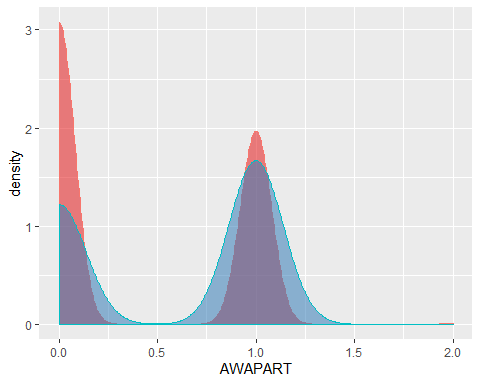
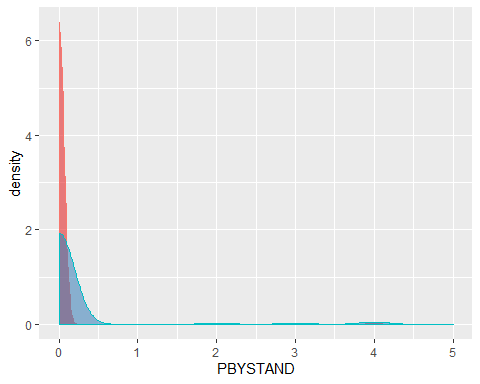
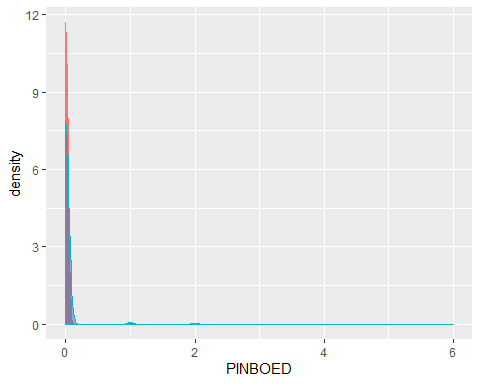
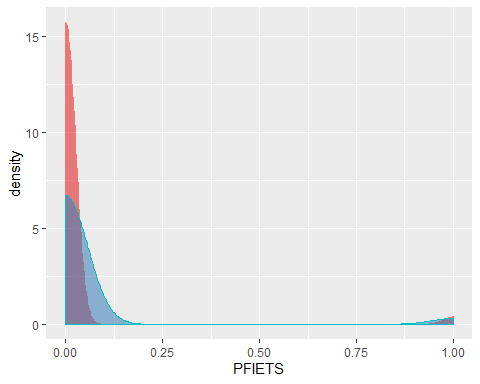
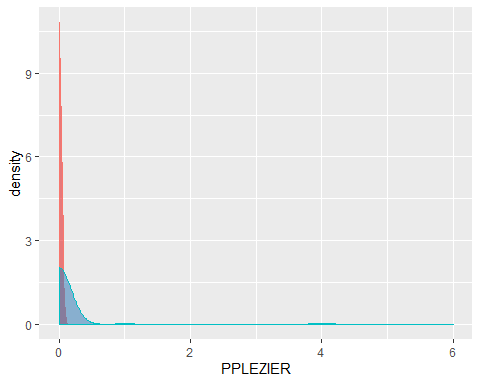
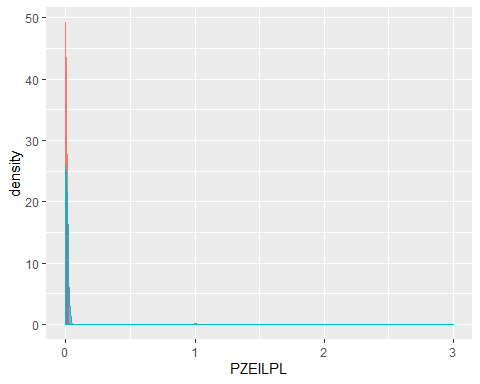
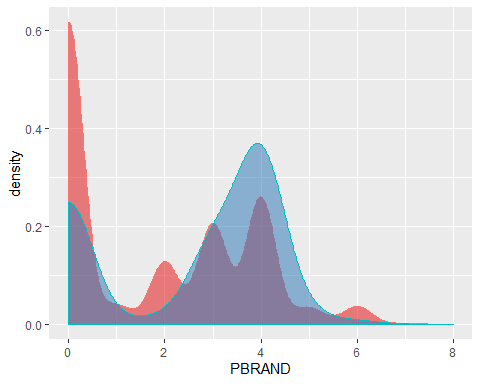
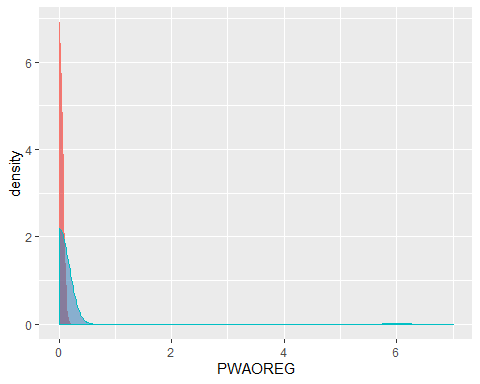
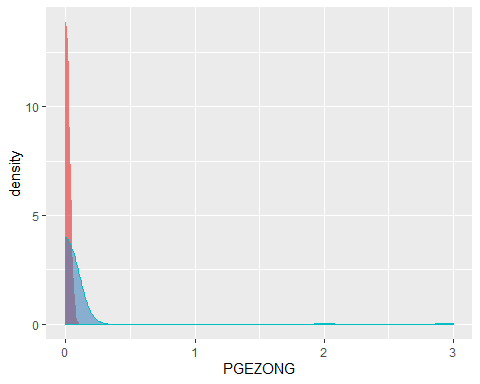
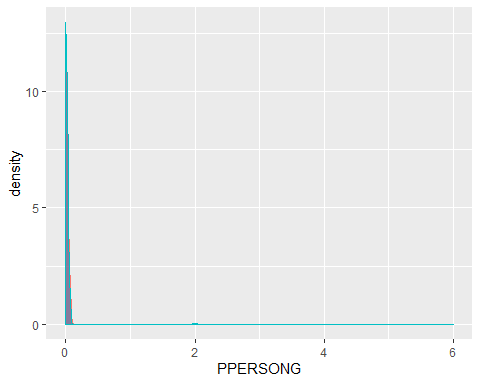
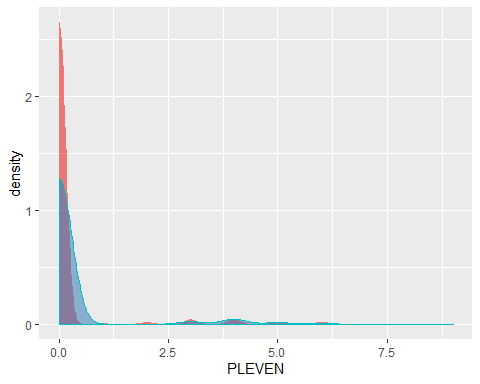
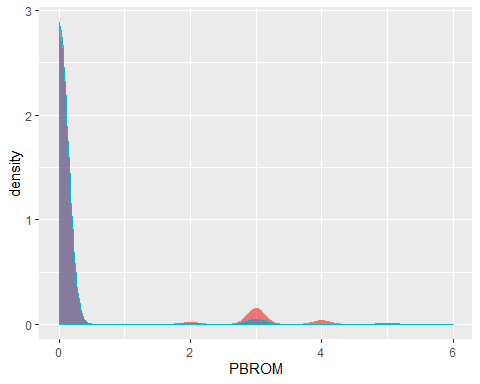
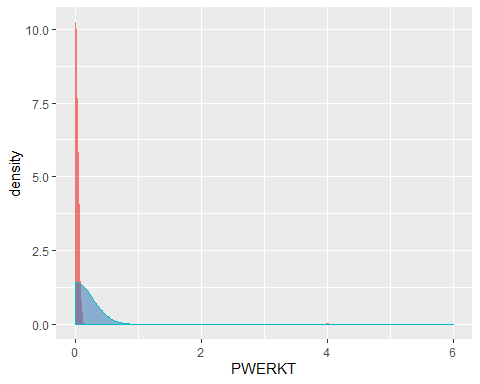
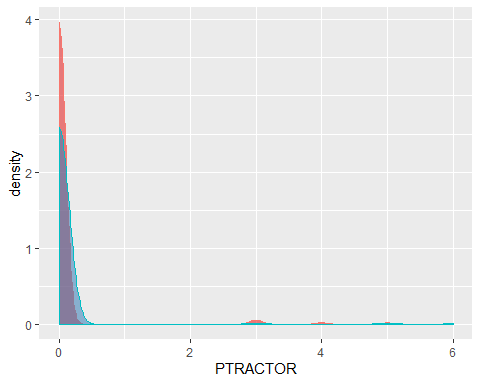
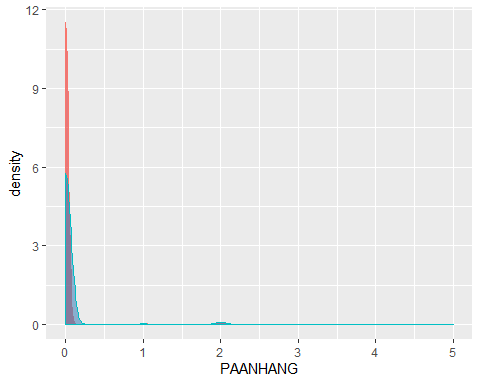
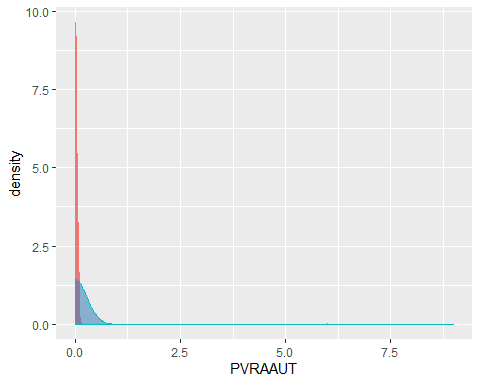
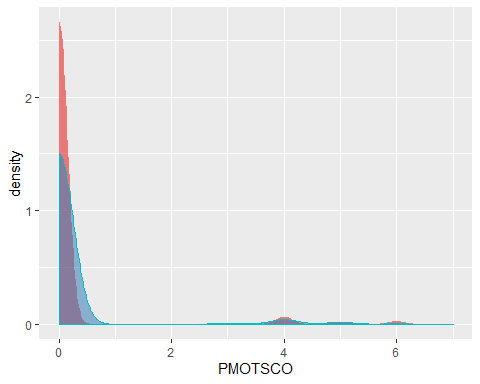
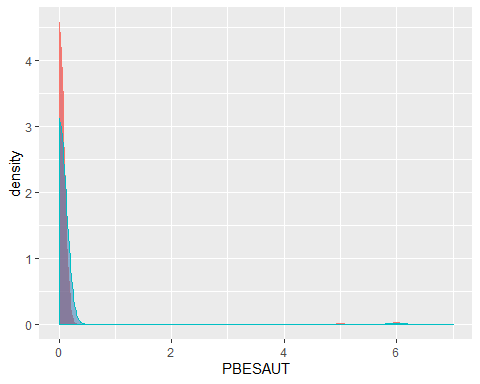
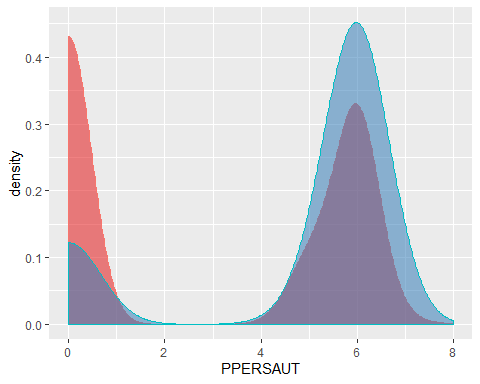
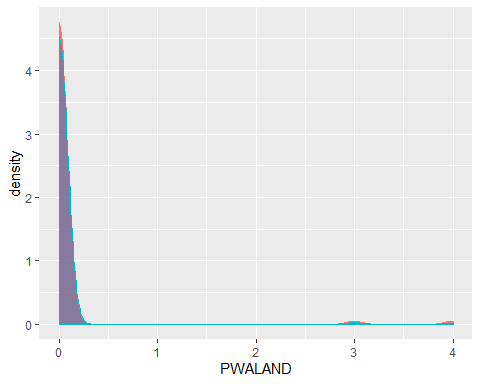
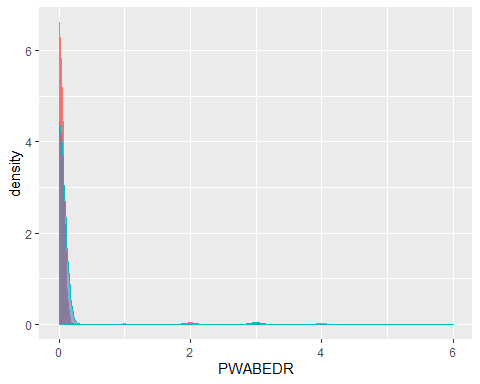
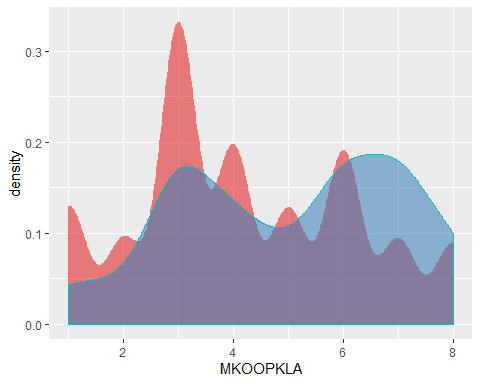
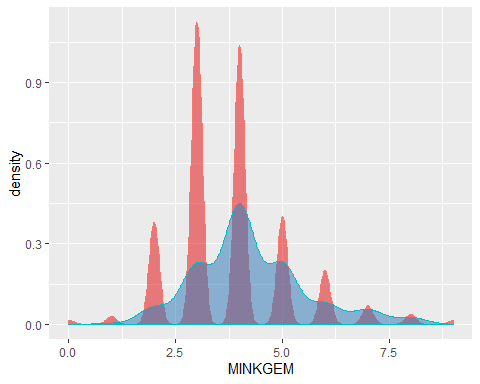
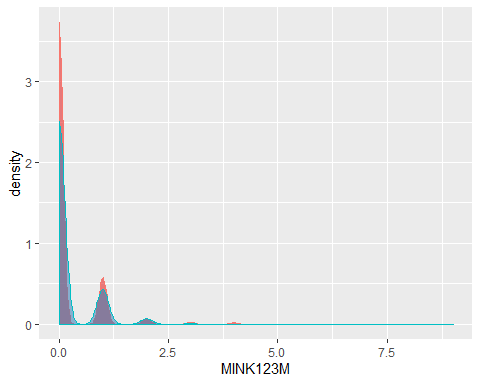
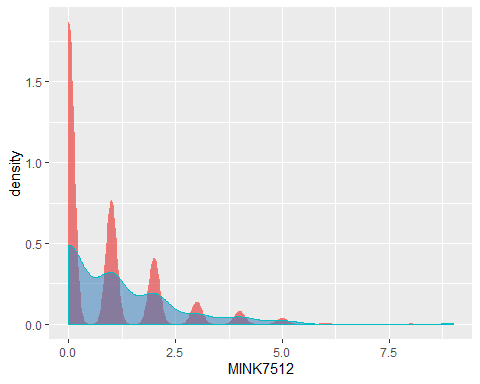
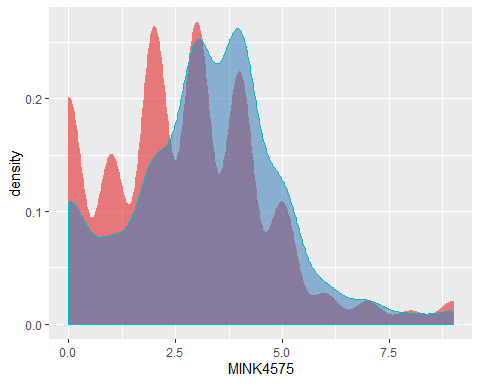
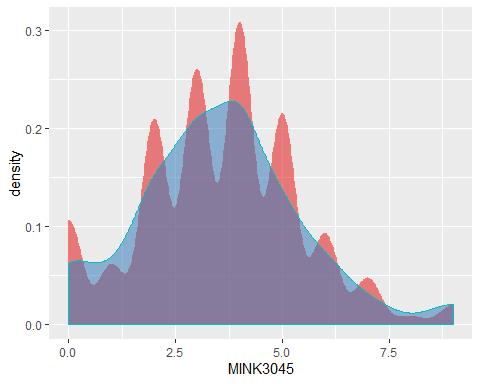
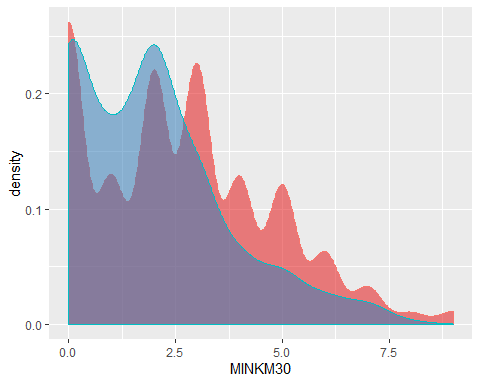
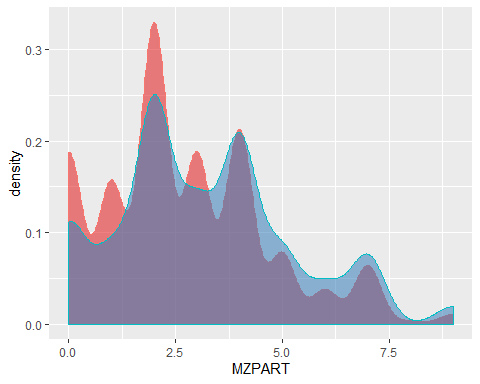
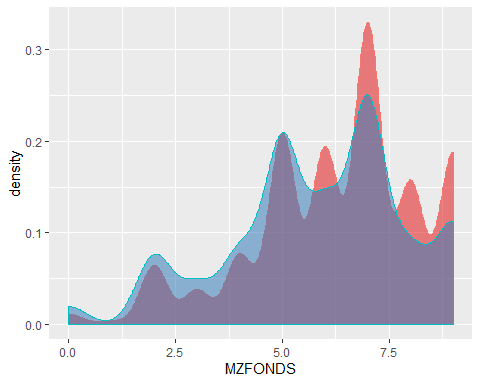
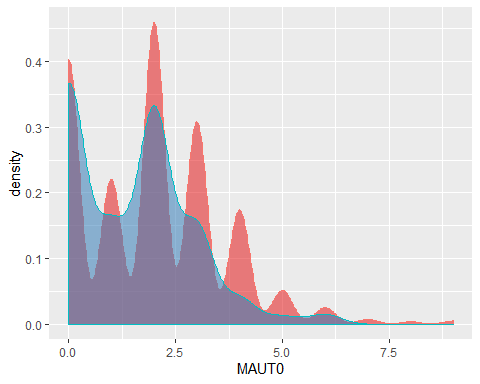
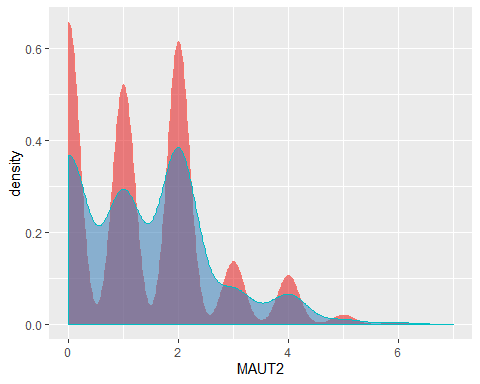
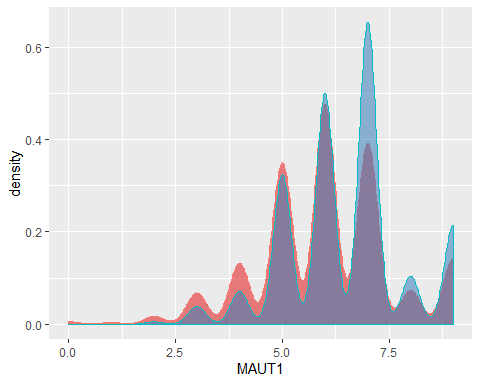
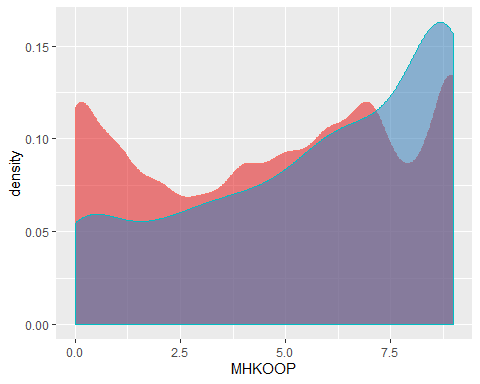
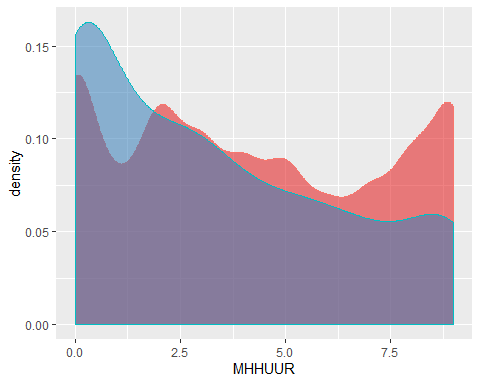
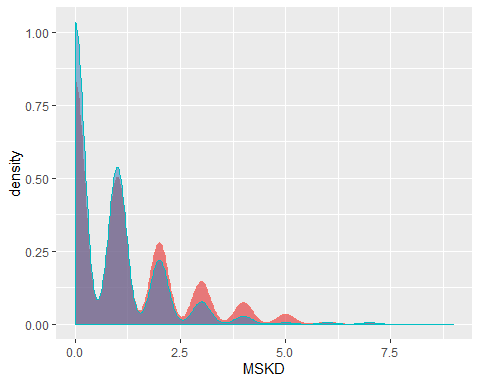
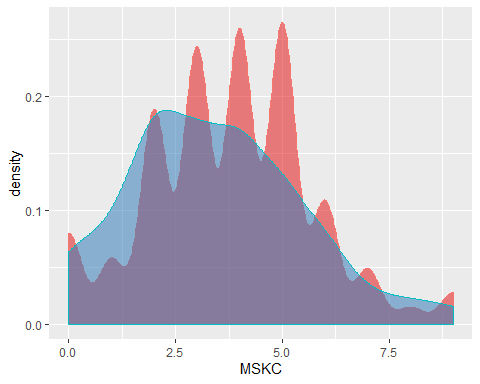
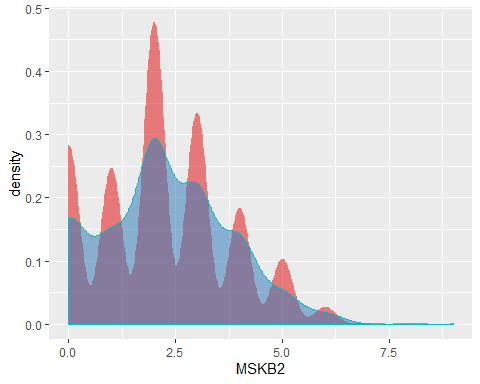
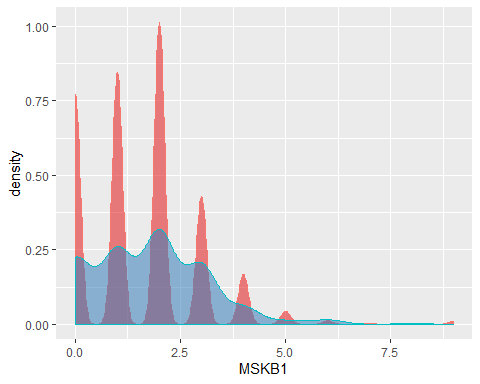
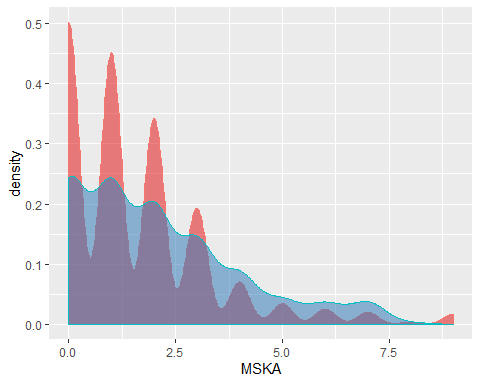
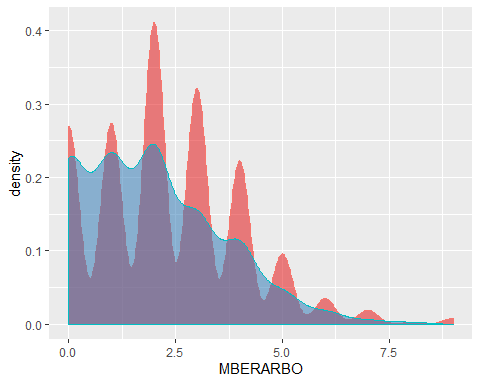
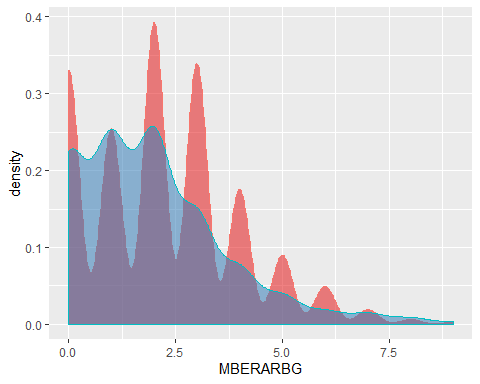
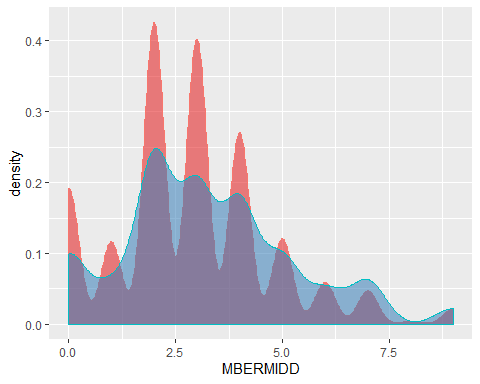
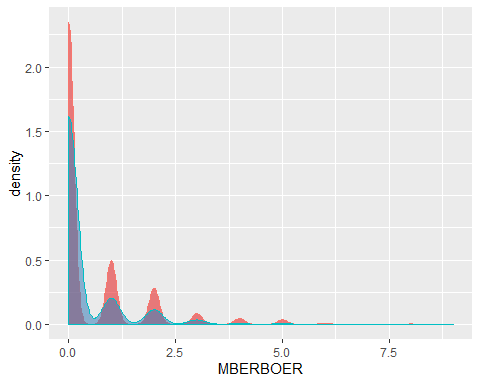
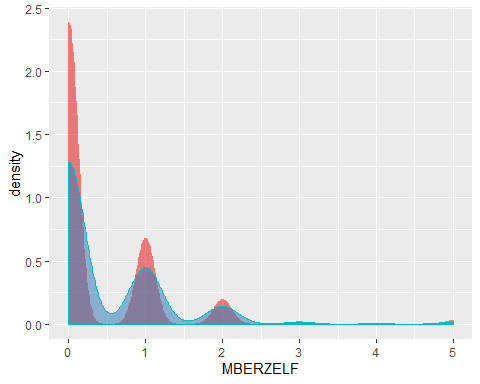
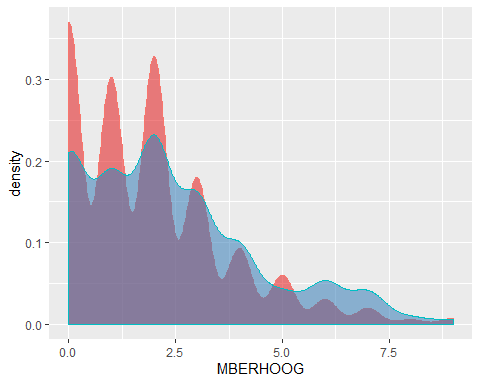
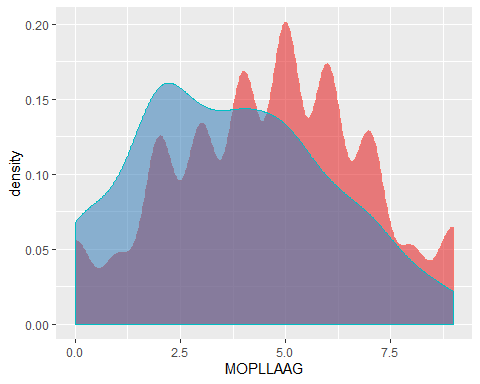
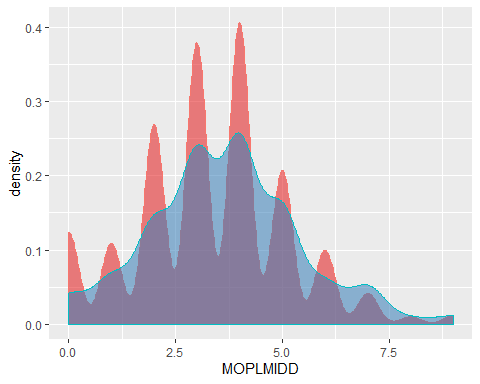
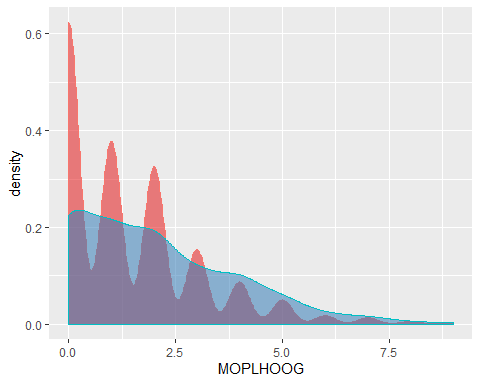
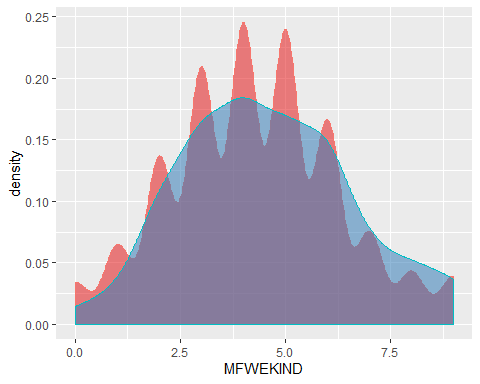
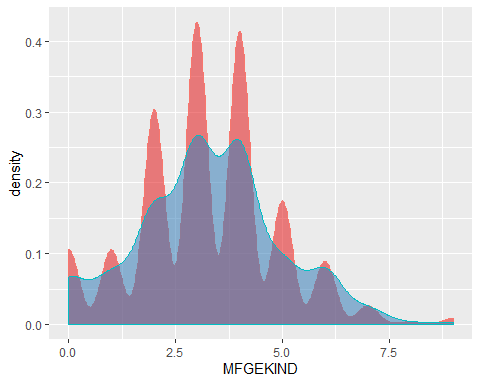
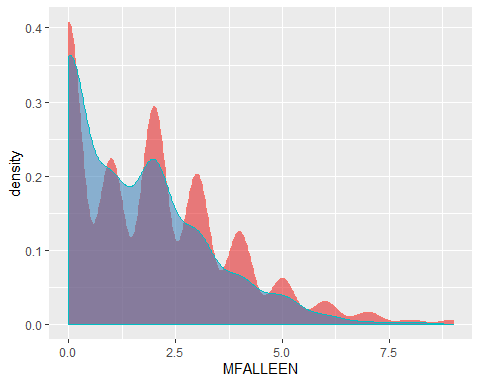
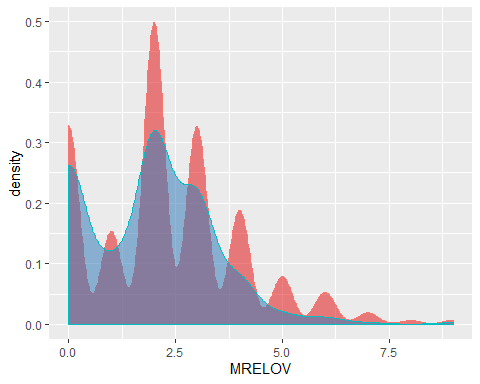
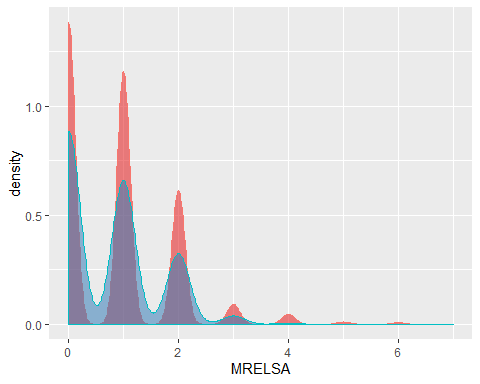
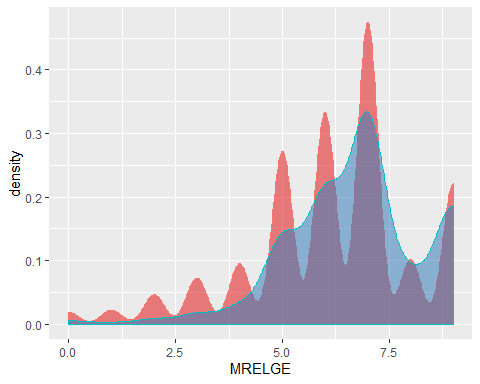
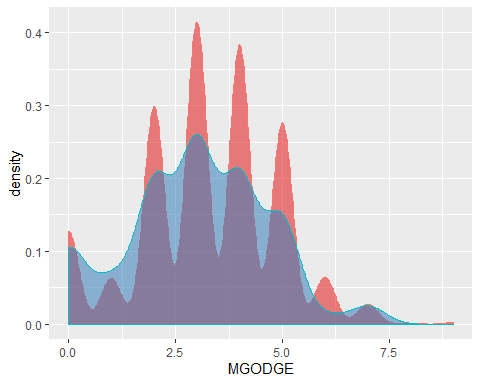
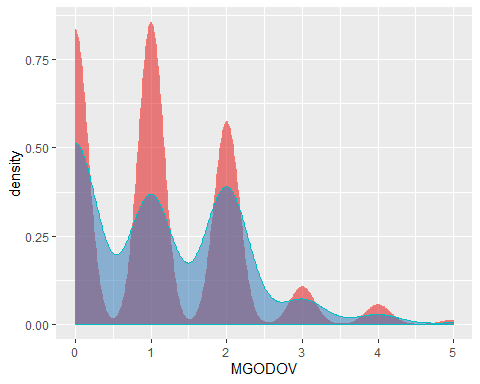
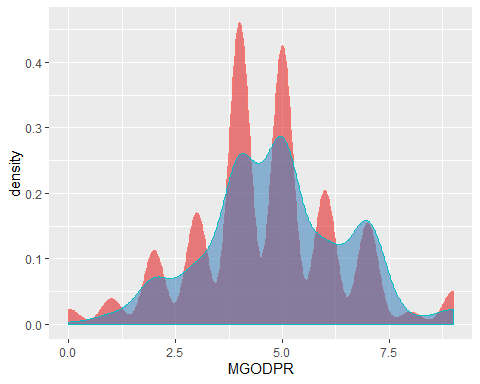
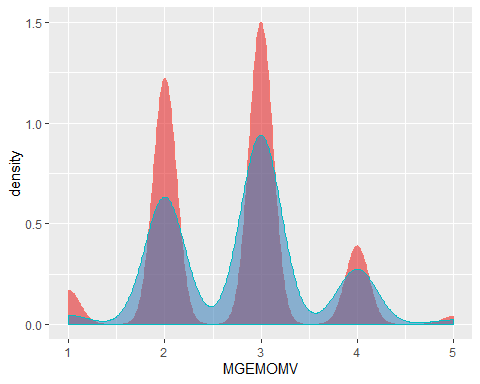
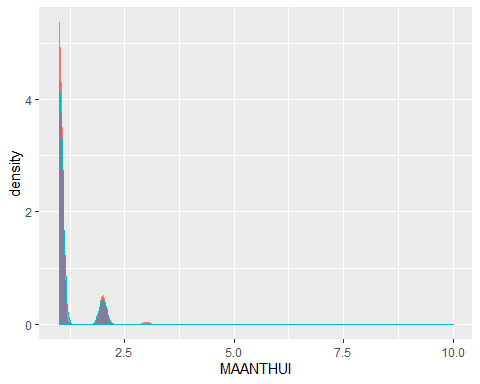
The dataset is unbalanced. There are only 348 cases that have purchased the mobile home insurance policy and 5474 that have not.

## Explanatory Variables

### Categorical



### Numeric



# Appendices

## Data Dictionary

|  |  |
| --- | --- |
| Name | Description |
| MOSTYPE | Customer Subtype |
| MAANTHUI | Number of houses |
| MGEMOMV | Avg size household |
| MGEMLEEF | Avg age |
| MOSHOOFD | Customer main type |
| MGODRK | Roman catholic |
| MGODPR | Protestant |
| MGODOV | Other religion |
| MGODGE | No religion |
| MRELGE | Married |
| MRELSA | Living together |
| MRELOV | Other relation |
| MFALLEEN | Singles |
| MFGEKIND | Household without children |
| MFWEKIND | Household with children |
| MOPLHOOG | High level education |
| MOPLMIDD | Medium level education |
| MOPLLAAG | Lower level education |
| MBERHOOG | High status |
| MBERZELF | Entrepreneur |
| MBERBOER | Farmer |
| MBERMIDD | Middle management |
| MBERARBG | Skilled labourers |
| MBERARBO | Unskilled labourers |
| MSKA | Social class A |
| MSKB1 | Social class B1 |
| MSKB2 | Social class B2 |
| MSKC | Social class C |
| MSKD | Social class D |
| MHHUUR | Rented house |
| MHKOOP | Home owners |
| MAUT1 | 1 car |
| MAUT2 | 2 cars |
| MAUT0 | No car |
| MZFONDS | National Health Service |
| MZPART | Private health insurance |
| MINKM30 | Income < 30.000 |
| MINK3045 | Income 30-45.000 |
| MINK4575 | Income 45-75.000 |
| MINK7512 | Income 75-122.000 |
| MINK123M | Income >123.000 |
| MINKGEM | Average income |
| MKOOPKLA | Purchasing power class |
| PWAPART | Contribution private third party insurance |
| PWABEDR | Contribution third party insurance (firms) |
| PWAAND | Contribution third party insurane (agriculture) |
| PPERSAUT | Contribution car policies |
| PBESAUT | Contribution delivery van policies |
| PMOTSCO | Contribution motorcycle/scooter policies |
| PVRAAUT | Contribution lorry policies |
| PAANHANG | Contribution trailer policies |
| PTRACTOR | Contribution tractor policies |
| PWERKT | Contribution agricultural machines policies |
| PBROM | Contribution moped policies |
| PLEVEN | Contribution life insurances |
| PPERSONG | Contribution private accident insurance policies |
| PGEZONG | Contribution family accidents insurance policies |
| PWAOREG | Contribution disability insurance policies |
| PBRAND | Contribution fire policies |
| PZEILPL | Contribution surfboard policies |
| PPLEZIER | Contribution boat policies |
| PFIETS | Contribution bicycle policies |
| PINBOED | Contribution property insurance policies |
| PBYSTAND | Contribution social security insurance policies |
| AWAPART | Number of private third party insurance |
| AWABEDR | Number of third party insurance (firms) |
| AWALAND | Number of third party insurane (agriculture) |
| APERSAUT | Number of car policies |
| ABESAUT | Number of delivery van policies |
| AMOTSCO | Number of motorcycle/scooter policies |
| AVRAAUT | Number of lorry policies |
| AAANHANG | Number of trailer policies |
| ATRACTOR | Number of tractor policies |
| AWERKT | Number of agricultural machines policies |
| ABROM | Number of moped policies |
| ALEVEN | Number of life insurances |
| APERSONG | Number of private accident insurance policies |
| AGEZONG | Number of family accidents insurance policies |
| AWAOREG | Number of disability insurance policies |
| ABRAND | Number of fire policies |
| AZEILPL | Number of surfboard policies |
| APLEZIER | Number of boat policies |
| AFIETS | Number of bicycle policies |
| AINBOED | Number of property insurance policies |
| ABYSTAND | Number of social security insurance policies |
| CARAVAN | Number of mobile home policy |

## References

Elkan, Charles. 2000. “CoIL Challenge 2000 Entry.” <http://liacs.leidenuniv.nl/~puttenpwhvander/library/cc2000/ELKANP~1.pdf>.

Putten, Peter, Michel Ruiter, and Maarten Someren. 2000. “CoIL Challenge 2000 Tasks and Results: Predicting and Explaining Caravan Policy Ownership.” <http://liacs.leidenuniv.nl/~puttenpwhvander/library/cc2000/PUTTEN~1.pdf>.