

Final Project

German Credit Dataset

kNN Classifier

Vidit Kalani / Abraham Shoun

MSCS 5610 Data Mining

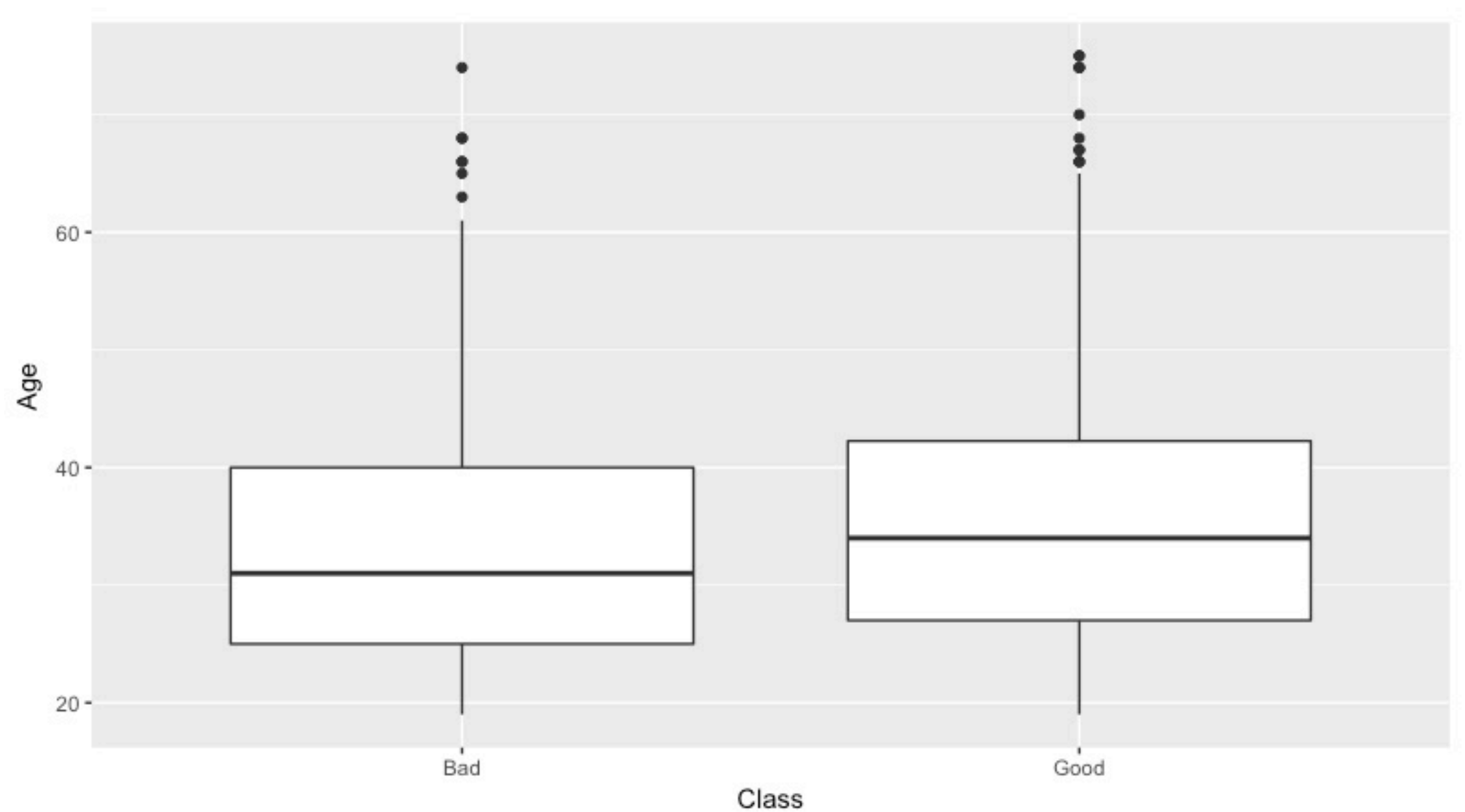
Final Project

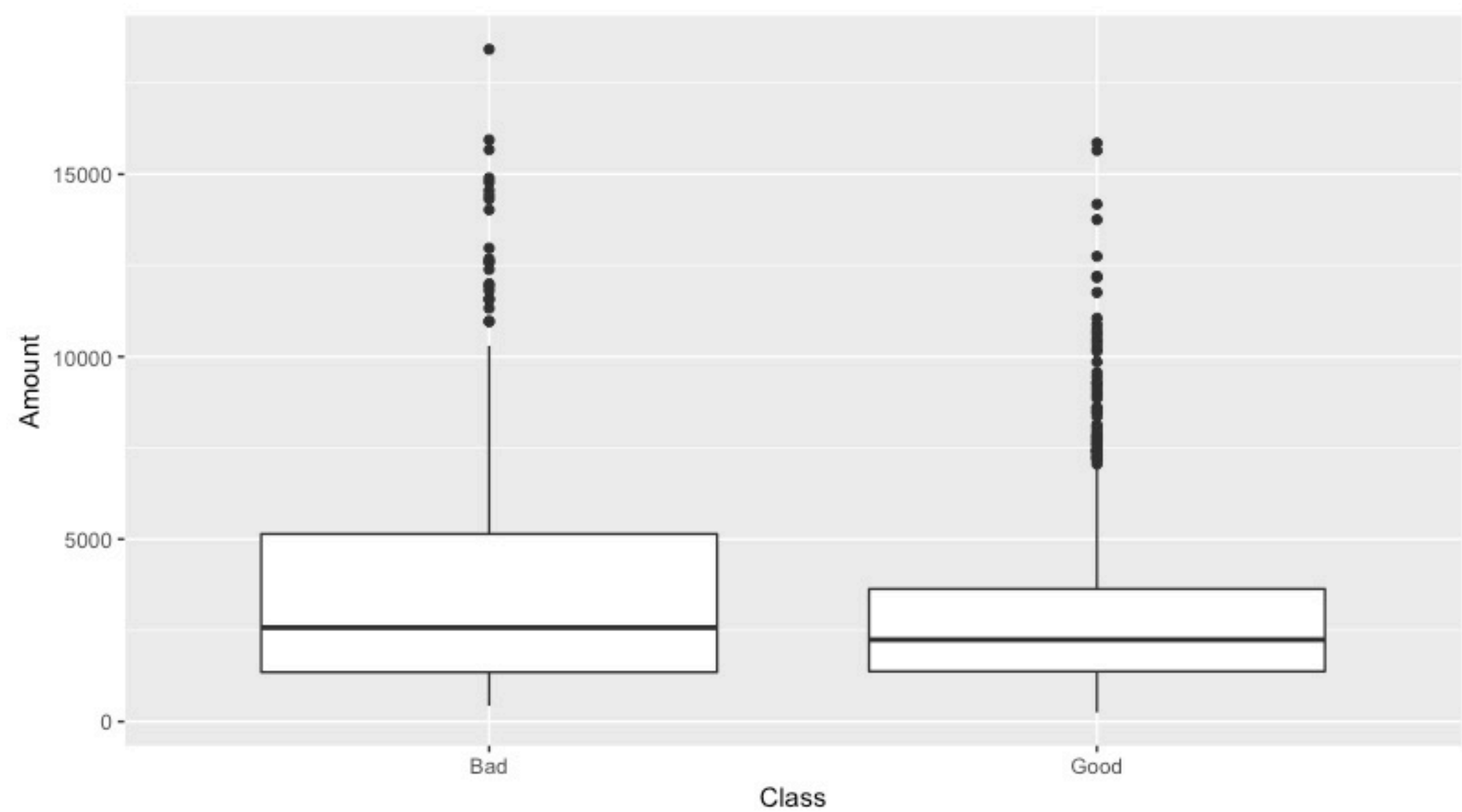
Spring 2018

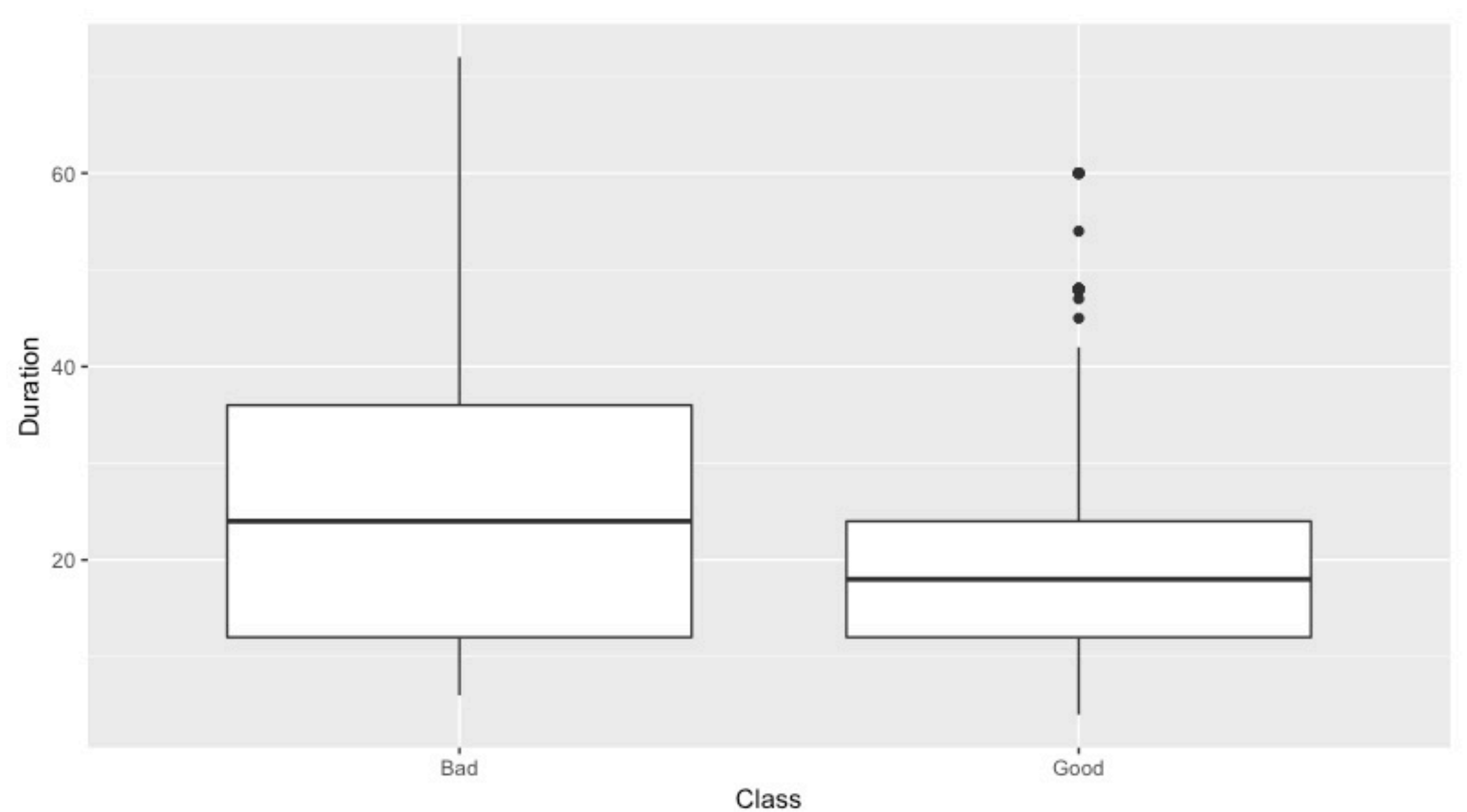
Prof. Dr. Praveen Madiraju

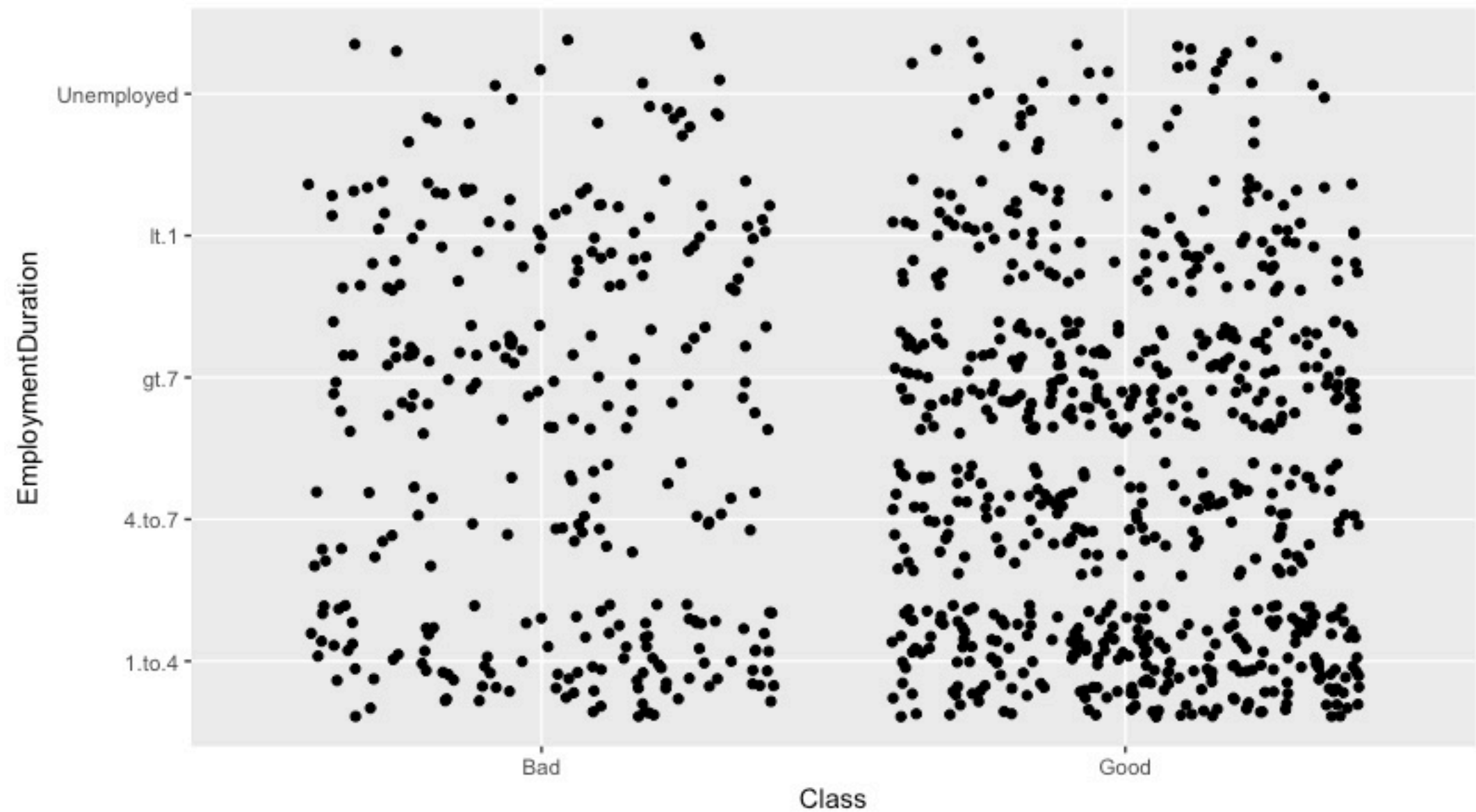
These are my of the features for study in German Credit dataset

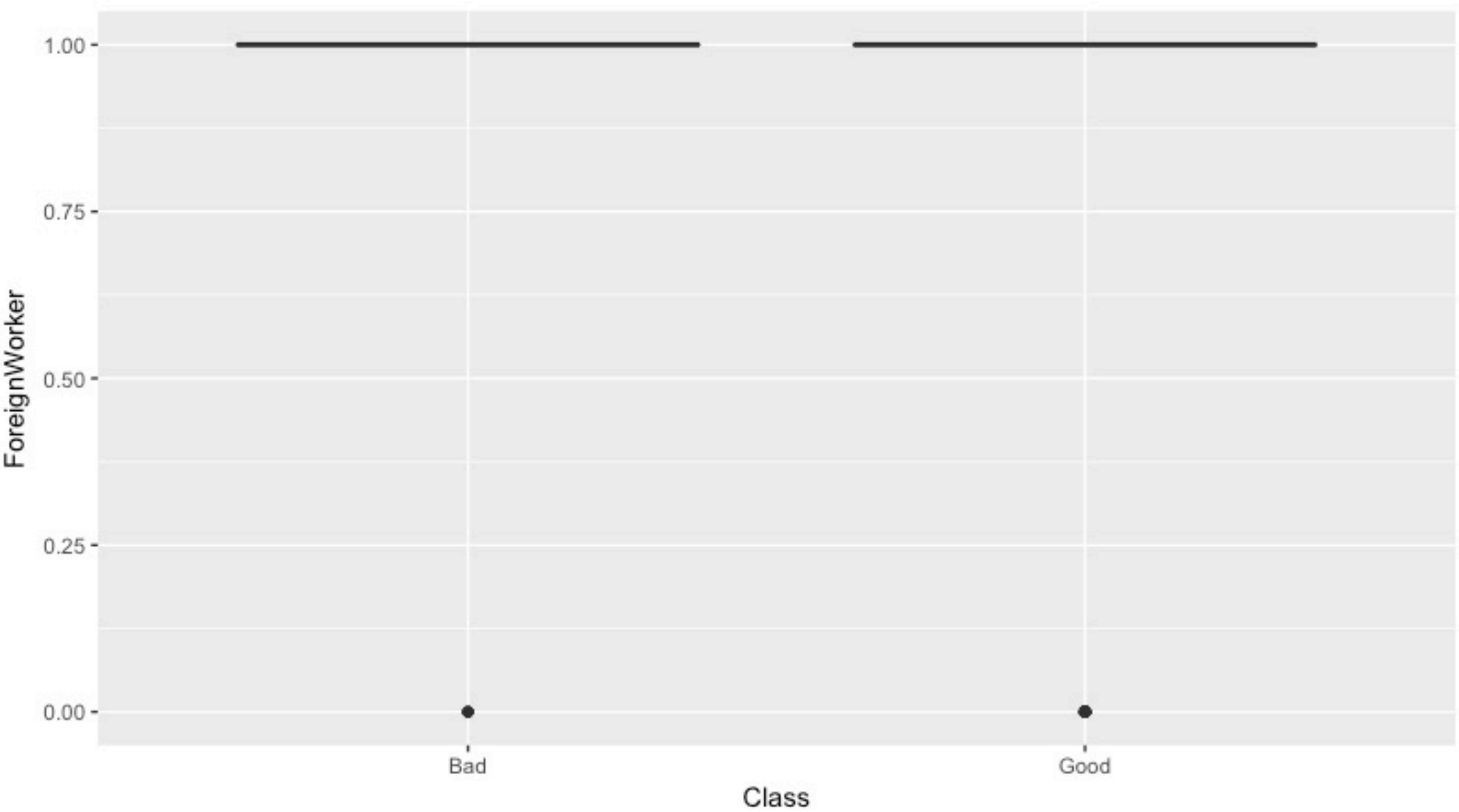
- Duration,
- Amount,
- InstallmentRatePercentage,
- Age,
- NumberExistingCredits,
- NumberPeopleMaintenance,
- Telephone,
- ForeignWorker,
- Class,
- CheckingAccountStatus.lt.0,
- CheckingAccountStatus.0.to.200,
- CheckingAccountStatus.gt.200,
- CheckingAccountStatus.none,
- CreditHistory.NoCredit.AllPaid,
- CreditHistory.ThisBank.AllPaid,
- CreditHistory.PaidDuly,
- CreditHistory.Delay,
- CreditHistory.Critical,
- SavingsAccountBonds.lt.100,
- SavingsAccountBonds.100.to.500,
- SavingsAccountBonds.500.to.1000,
- SavingsAccountBonds.gt.1000,
- SavingsAccountBonds.Unknown,
- EmploymentDuration.lt.1,
- EmploymentDuration.1.to.4,
- EmploymentDuration.4.to.7,
- EmploymentDuration.gt.7,
- EmploymentDuration.Unemployed,
- Personal.Male.Divorced.Seperated,
- Personal.Female.NotSingle,
- Personal.Male.Single,
- Personal.Male.Married.Widowed,
- Personal.Female.Single,
- Property.RealEstate,
- Property.Insurance,
- Property.CarOther,
- Property.Unknown,
- Housing.Rent,
- Housing.Own,
- Housing.ForFree

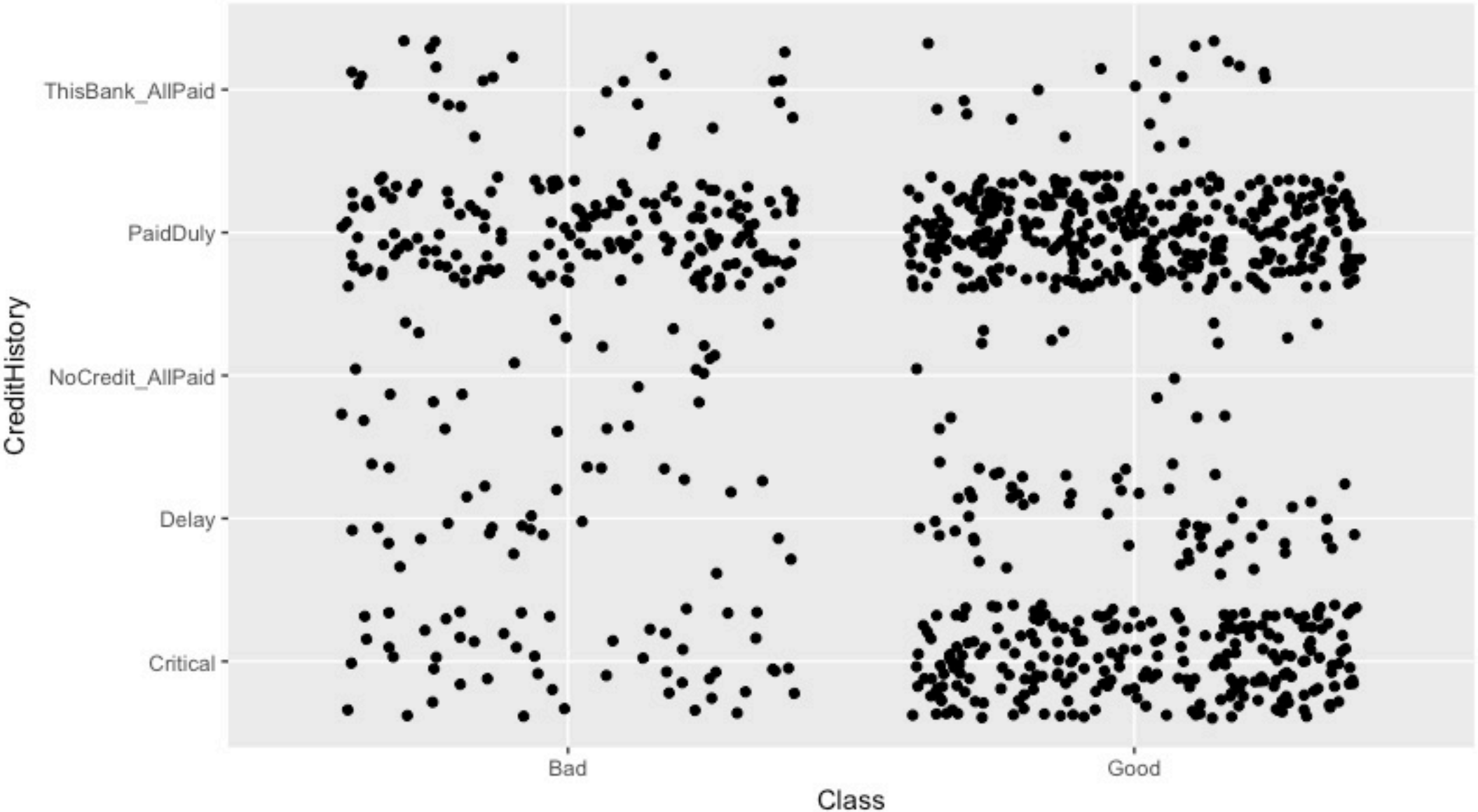


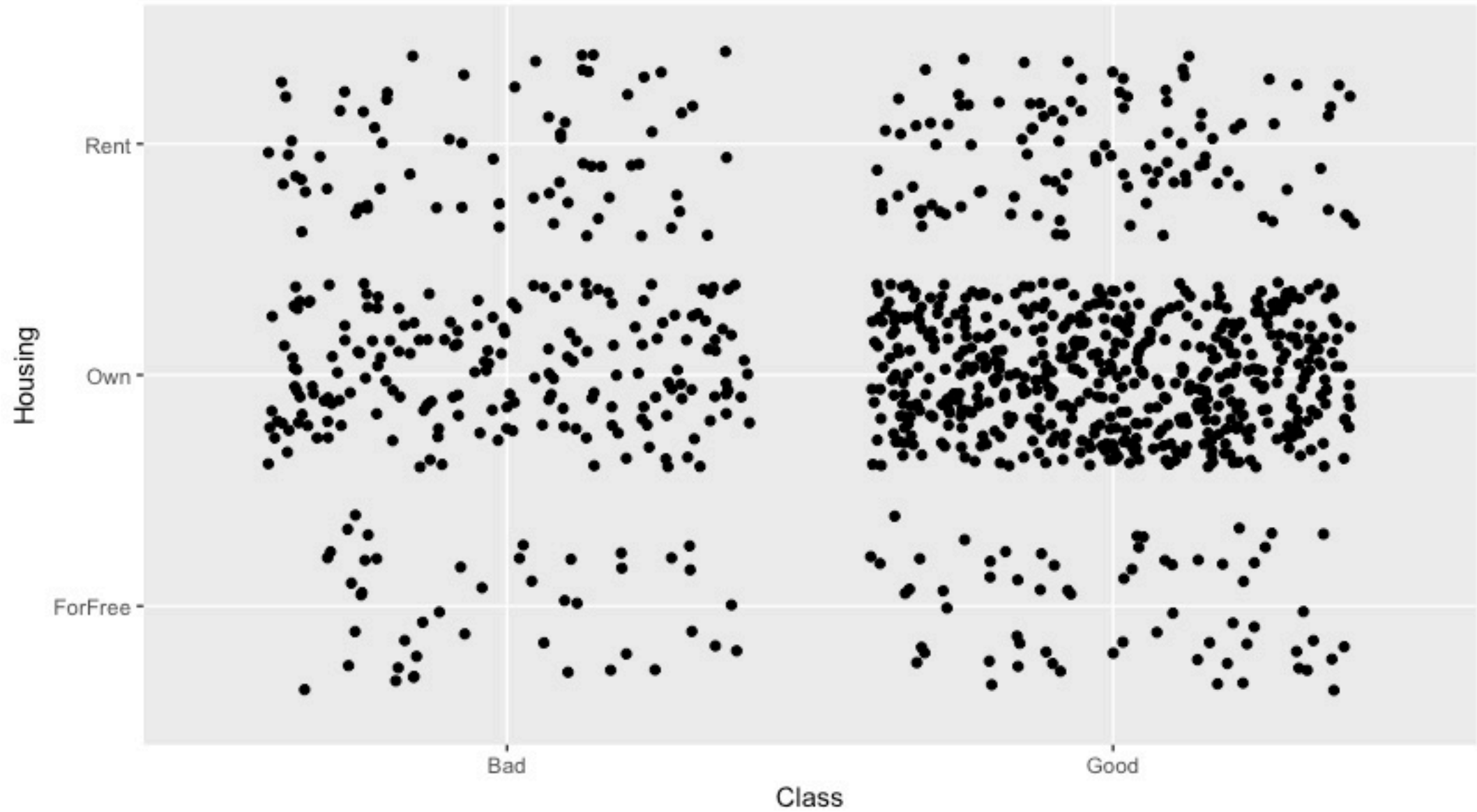


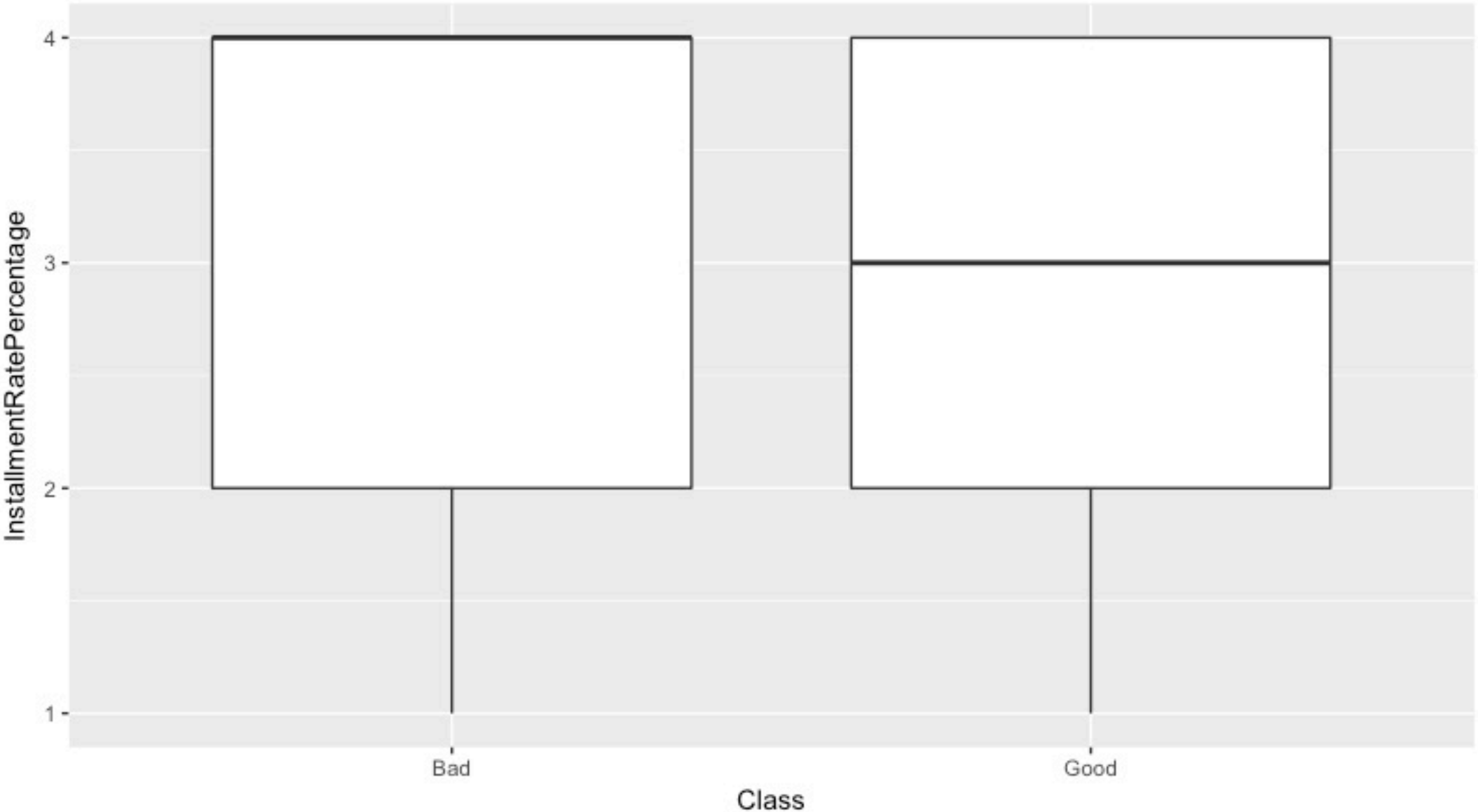


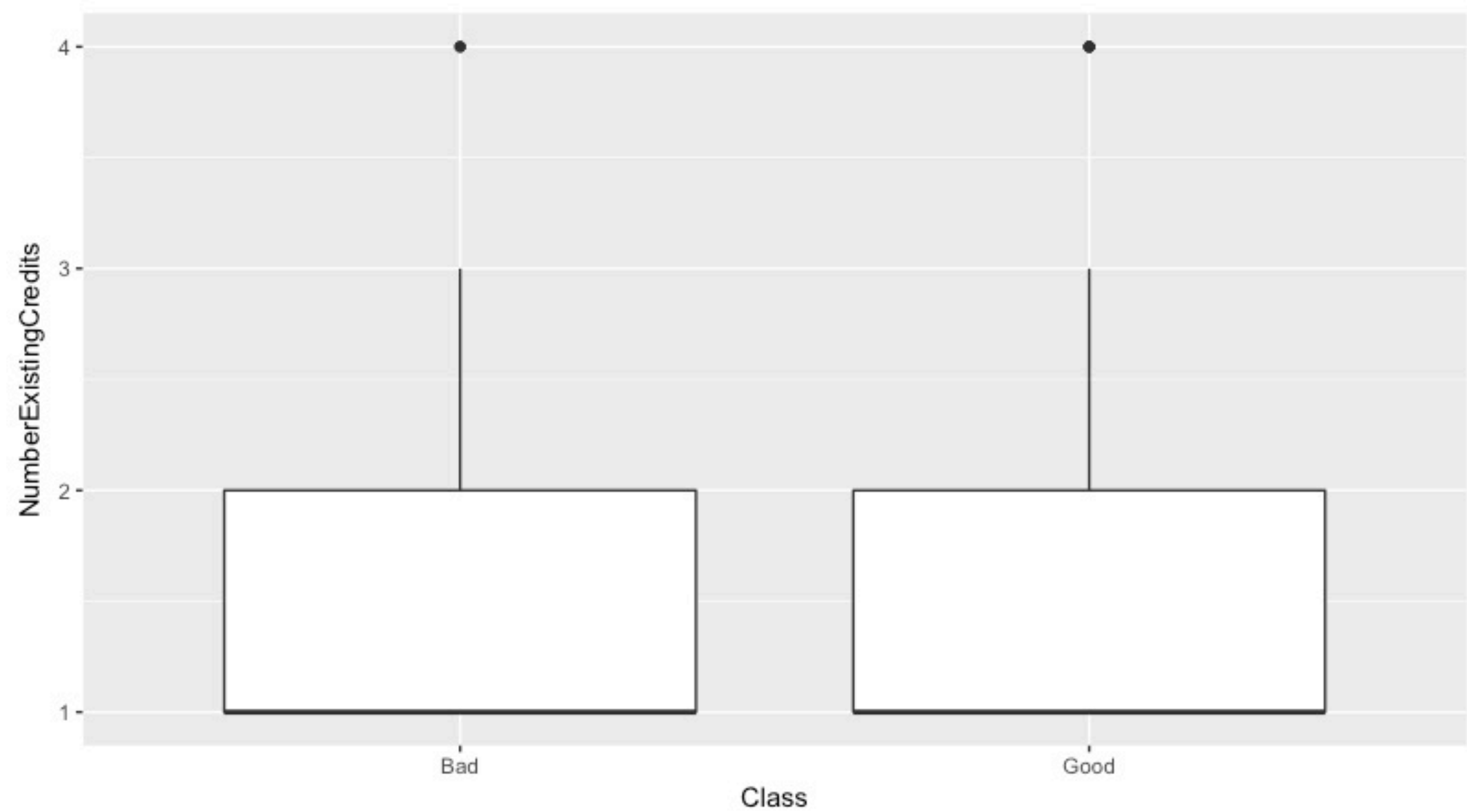


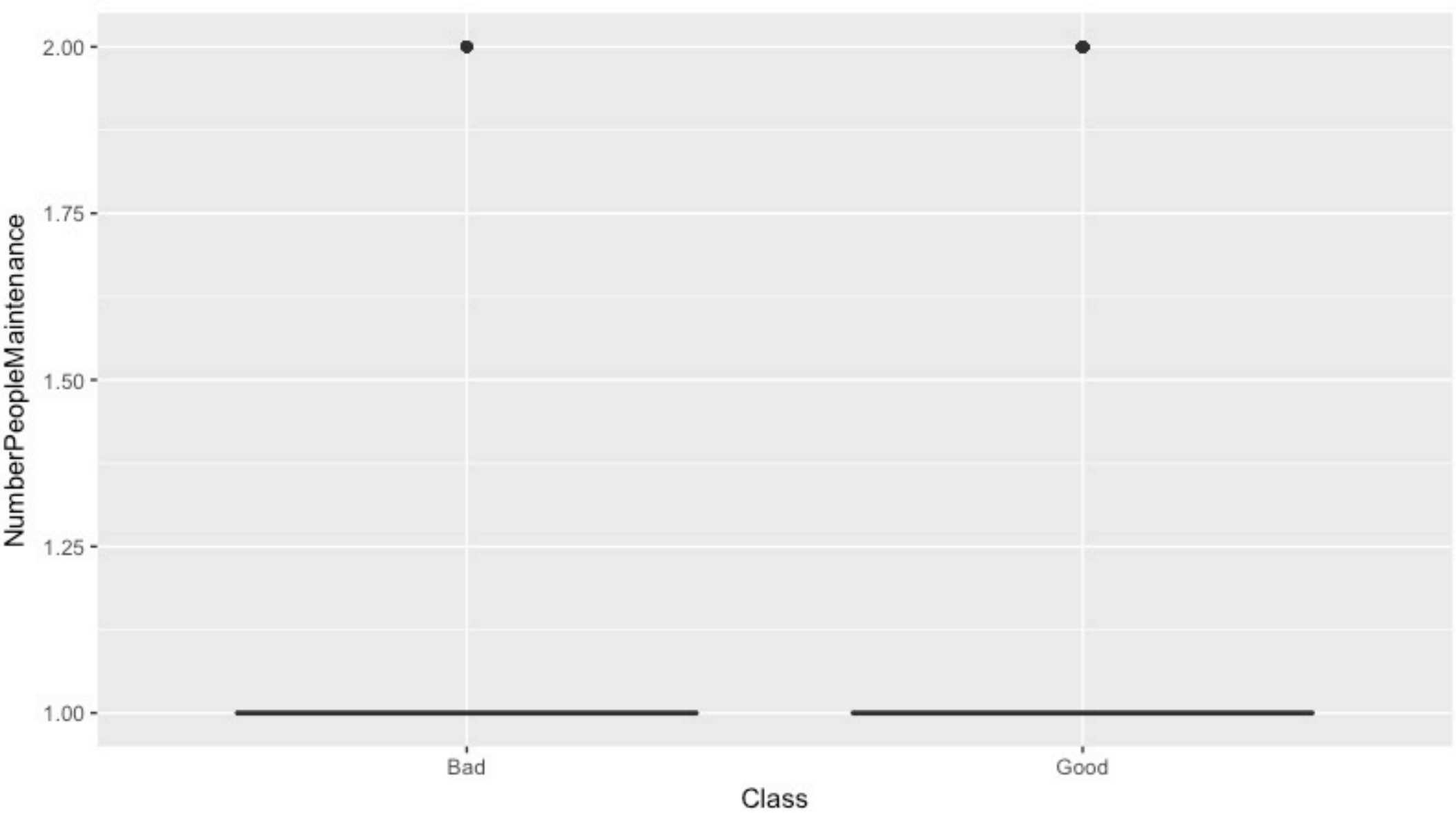


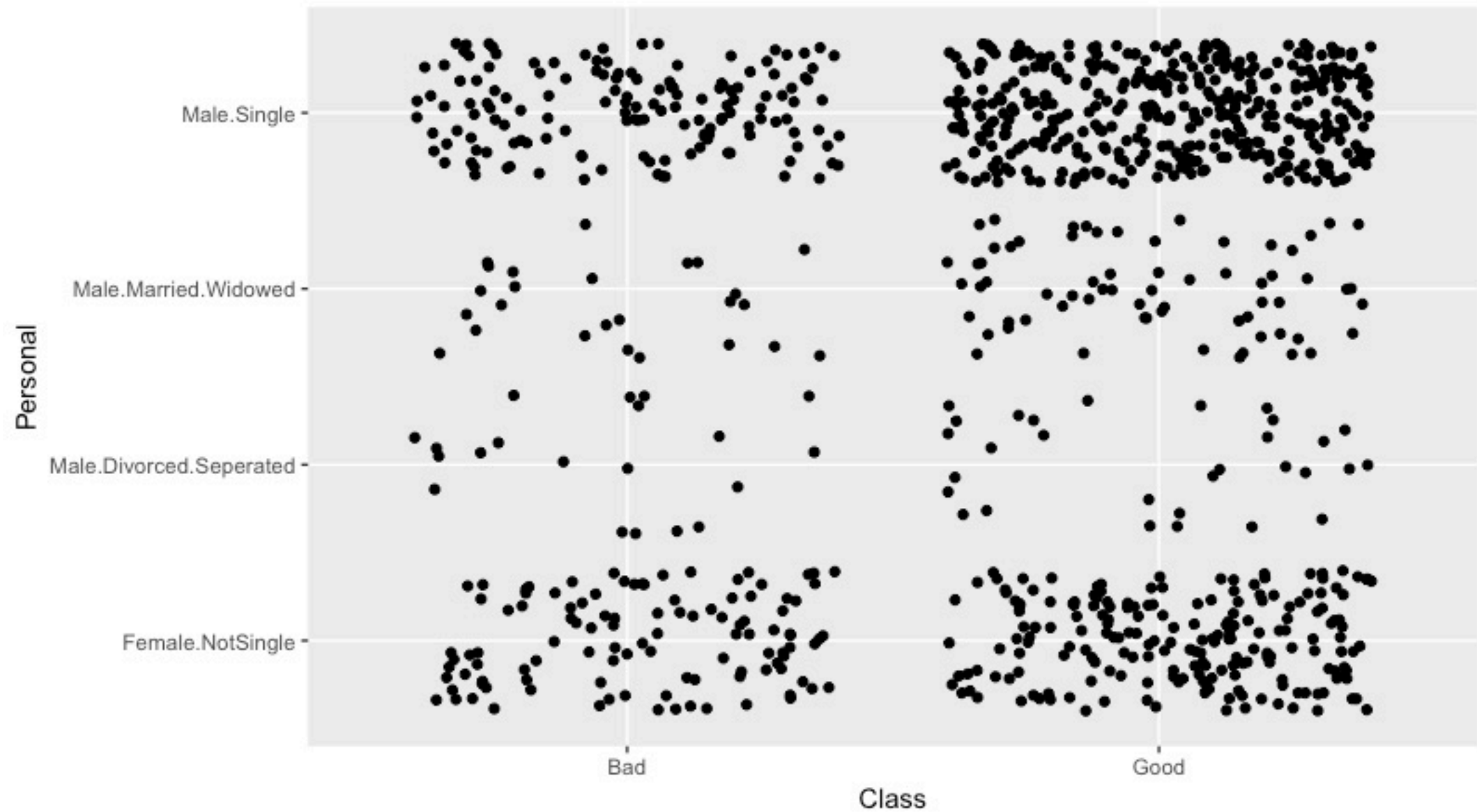


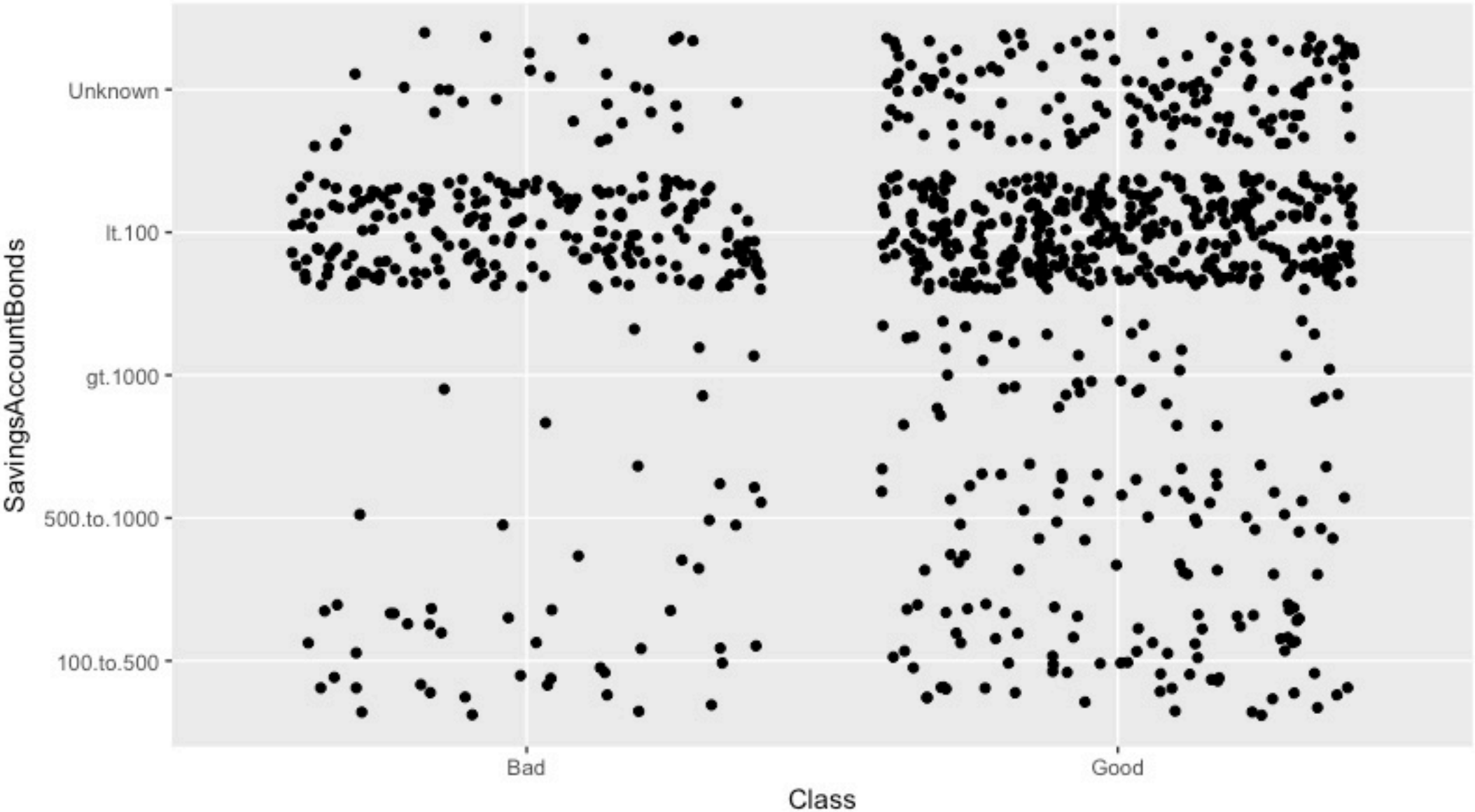




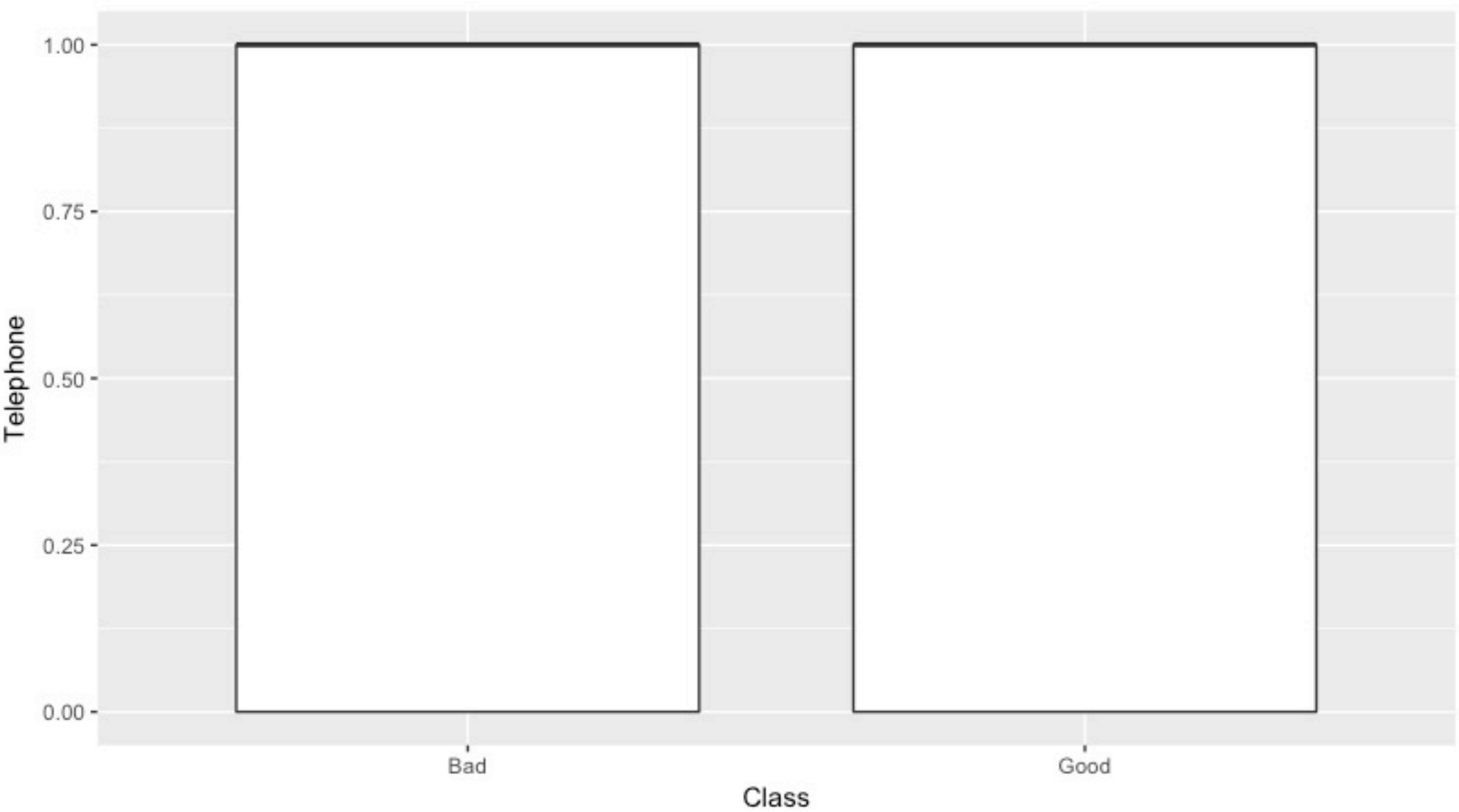












Round: 1	
Features	Accuracies
Duration	0.725
Amount	0.69
InstallmentRatePercentage	0.7
Age	0.685
NumberExistingCredits	0.7
NumberPeopleMaintenance	0.7
Telephone	0.7
ForeignWorker	0.7
CheckingAccountStatus	0.665
CreditHistory	0.7
SavingsAccountBonds	0.7
EmploymentDuratio	0.7
Personal	0.7
Property	0.7
Housing	0.7

Round: 2	
Features	Accuracies
Duration + Amount	0.70
Duration + InstallmentRatePercentage	0.715
Duration + Age	0.715
Duration + NumberExistingCredits	0.725
Duration + NumberPeopleMaintenance	0.735
Duration + Telephone	0.725
Duration + ForeignWorker	0.72
Duration + CheckingAccountStatus	0.71
Duration + CreditHistory	0.71
Duration + SavingsAccountBonds	0.72
Duration + EmploymentDuratio	0.73
Duration + Personal	0.73
Duration + Property	0.70
Duration + Housing	0.7

Round: 3

Features	Accuracies
Duration + NumberPeopleMaintenance + Amount	0.71
Duration + NumberPeopleMaintenance + InstallmentRatePercentage	0.715
Duration + NumberPeopleMaintenance + Age	0.73
Duration + NumberPeopleMaintenance + NumberExistingCredits	0.74
Duration + NumberPeopleMaintenance + Telephone	0.725
Duration + NumberPeopleMaintenance + ForeignWorker	0.74
Duration + NumberPeopleMaintenance + CheckingAccountStatus	0.71
Duration + NumberPeopleMaintenance + CreditHistory	0.735
Duration + NumberPeopleMaintenance + SavingsAccountBonds	0.73
Duration + NumberPeopleMaintenance + EmploymentDuratio	0.715
Duration + NumberPeopleMaintenance + Personal	0.715
Duration + NumberPeopleMaintenance + Property	0.71
Duration + NumberPeopleMaintenance + Housing	0.765

Round: 4

Features	Accuracies
Duration + NumberPeopleMaintenance + Housing + Amount	0.72
Duration + NumberPeopleMaintenance + Housing + InstallmentRatePercentage	0.75
Duration + NumberPeopleMaintenance + Housing + Age	0.735
Duration + NumberPeopleMaintenance + Housing + NumberExistingCredits	0.73
Duration + NumberPeopleMaintenance + Housing + Telephone	0.775
Duration + NumberPeopleMaintenance + Housing + ForeignWorker	0.76
Duration + NumberPeopleMaintenance + Housing + CheckingAccountStatus	0.71
Duration + NumberPeopleMaintenance + Housing + CreditHistory	0.735
Duration + NumberPeopleMaintenance + Housing + SavingsAccountBonds	0.735
Duration + NumberPeopleMaintenance + Housing + EmploymentDuration	0.72
Duration + NumberPeopleMaintenance + Housing + Personal	0.72
Duration + NumberPeopleMaintenance + Housing + Property	0.73

Round: 5

Features	Accuracies
Duration + NumberPeopleMaintenance + Housing + Telephone + Amount	0.735
Duration + NumberPeopleMaintenance + Housing + Telephone + InstallmentRatePercentage	0.71
Duration + NumberPeopleMaintenance + Housing + Telephone + Age	0.775
Duration + NumberPeopleMaintenance + Housing + Telephone + NumberExistingCredits	0.73
Duration + NumberPeopleMaintenance + Housing + Telephone + ForeignWorker	0.74
Duration + NumberPeopleMaintenance + Housing + Telephone + CheckingAccountStatus	0.665
Duration + NumberPeopleMaintenance + Housing + Telephone + CreditHistory	0.705
Duration + NumberPeopleMaintenance + Housing + Telephone + SavingsAccountBonds	0.735
Duration + NumberPeopleMaintenance + Housing + Telephone + EmploymentDuratio	0.705
Duration + NumberPeopleMaintenance + Housing + Telephone + Personal	0.74
Duration + NumberPeopleMaintenance + Housing + Telephone + Property	0.695

- Round 1:- Duration feature is best accuracy.
- Round 2:- Duration + NumberPeopleMaintenance features are best accuracy.
- Round 3:- Duration + NumberPeopleMaintenance + Housing features are best accuracy.
- Round 4:- Duration + NumberPeopleMaintenance + Housing + Telephone features are best accuracy.
- Round 5:- Duration + NumberPeopleMaintenance + Housing + Telephone + Age features are best accuracy.
- But I noticed that after Round 4 the some features accuracy going down compare to first 3 rounds.
- And In Round 5 only one feature's accuracy is going up compare to first 4 rounds.