

Assignment 1

Nikola Metes

Set up environment

```
# library(caret)
library("Hmisc")
```

```
## Loading required package: lattice
```

```
## Loading required package: survival
```

```
## Loading required package: Formula
```

```
## Loading required package: ggplot2
```

```
##
```

```
## Attaching package: 'Hmisc'
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      format.pval, units
```

```
library(dplyr)
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:Hmisc':
```

```
##
```

```
##      src, summarize
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##      filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

```
Data<-file.path("Data")
```

```
d<-read.csv(file.path(Data,"GermanCredit.csv"))
head(d)
```

```
##      Duration Amount InstallmentRatePercentage ResidenceDuration Age
## 1         6    1169                4                4 67
## 2        48    5951                2                2 22
## 3        12    2096                2                3 49
## 4        42    7882                2                4 45
## 5        24    4870                3                4 53
## 6        36    9055                2                4 35
##      NumberExistingCredits NumberPeopleMaintenance Telephone ForeignWorker
## 1                      2                      1          0          1
## 2                      1                      1          1          1
## 3                      1                      2          1          1
## 4                      1                      2          1          1
## 5                      2                      2          1          1
## 6                      1                      2          0          1
##      Class CheckingAccountStatus.lt.0 CheckingAccountStatus.0.to.200
## 1 Good                      1                      0
## 2 Bad                       0                      1
## 3 Good                      0                      0
## 4 Good                      1                      0
## 5 Bad                       1                      0
## 6 Good                      0                      0
##      CheckingAccountStatus.gt.200 CheckingAccountStatus.none
## 1                      0                      0
## 2                      0                      0
## 3                      0                      1
## 4                      0                      0
## 5                      0                      0
## 6                      0                      1
##      CreditHistory.NoCredit.AllPaid CreditHistory.ThisBank.AllPaid
## 1                      0                      0
## 2                      0                      0
## 3                      0                      0
## 4                      0                      0
## 5                      0                      0
## 6                      0                      0
##      CreditHistory.PaidDuly CreditHistory.Delay CreditHistory.Critical
## 1                      0                      0                      1
## 2                      1                      0                      0
## 3                      0                      0                      1
## 4                      1                      0                      0
## 5                      0                      1                      0
## 6                      1                      0                      0
##      Purpose.NewCar Purpose.UsedCar Purpose.Furniture.Equipment
## 1                      0                      0                      0
## 2                      0                      0                      0
## 3                      0                      0                      0
## 4                      0                      0                      1
## 5                      1                      0                      0
## 6                      0                      0                      0
##      Purpose.Radio.Television Purpose.DomesticAppliance Purpose.Repairs
```

## 1	1	0	0	
## 2	1	0	0	
## 3	0	0	0	
## 4	0	0	0	
## 5	0	0	0	
## 6	0	0	0	
##	Purpose.Education	Purpose.Vacation	Purpose.Retaining	Purpose.Business
## 1	0	0	0	0
## 2	0	0	0	0
## 3	1	0	0	0
## 4	0	0	0	0
## 5	0	0	0	0
## 6	1	0	0	0
##	Purpose.Other	SavingsAccountBonds.lt.100	SavingsAccountBonds.100.to.500	
## 1	0	0	0	
## 2	0	1	0	
## 3	0	1	0	
## 4	0	1	0	
## 5	0	1	0	
## 6	0	0	0	
##	SavingsAccountBonds.500.to.1000	SavingsAccountBonds.gt.1000		
## 1	0	0		
## 2	0	0		
## 3	0	0		
## 4	0	0		
## 5	0	0		
## 6	0	0		
##	SavingsAccountBonds.Unknown	EmploymentDuration.lt.1		
## 1	1	0		
## 2	0	0		
## 3	0	0		
## 4	0	0		
## 5	0	0		
## 6	1	0		
##	EmploymentDuration.1.to.4	EmploymentDuration.4.to.7		
## 1	0	0		
## 2	1	0		
## 3	0	1		
## 4	0	1		
## 5	1	0		
## 6	1	0		
##	EmploymentDuration.gt.7	EmploymentDuration.Unemployed		
## 1	1	0		
## 2	0	0		
## 3	0	0		
## 4	0	0		
## 5	0	0		
## 6	0	0		
##	Personal.Male.Divorced.Seperated	Personal.Female.NotSingle		
## 1	0	0		
## 2	0	1		
## 3	0	0		
## 4	0	0		
## 5	0	0		

## 6		0		0
##	Personal.Male.Single	Personal.Male.Married.Widowed		
## 1	1		0	
## 2	0		0	
## 3	1		0	
## 4	1		0	
## 5	1		0	
## 6	1		0	
##	Personal.Female.Single	OtherDebtorsGuarantors.None		
## 1	0		1	
## 2	0		1	
## 3	0		1	
## 4	0		0	
## 5	0		1	
## 6	0		1	
##	OtherDebtorsGuarantors.CoApplicant	OtherDebtorsGuarantors.Guarantor		
## 1		0		0
## 2		0		0
## 3		0		0
## 4		0		1
## 5		0		0
## 6		0		0
##	Property.RealEstate	Property.Insurance	Property.CarOther	
## 1	1	0	0	
## 2	1	0	0	
## 3	1	0	0	
## 4	0	1	0	
## 5	0	0	0	
## 6	0	0	0	
##	Property.Unknown	OtherInstallmentPlans.Bank	OtherInstallmentPlans.Stores	
## 1	0	0		0
## 2	0	0		0
## 3	0	0		0
## 4	0	0		0
## 5	1	0		0
## 6	1	0		0
##	OtherInstallmentPlans.None	Housing.Rent	Housing.Own	Housing.ForFree
## 1		1	0	1
## 2		1	0	1
## 3		1	0	1
## 4		1	0	0
## 5		1	0	0
## 6		1	0	0
##	Job.UnemployedUnskilled	Job.UnskilledResident	Job.SkilledEmployee	
## 1		0		1
## 2		0		1
## 3		0	1	0
## 4		0	0	1
## 5		0	0	1
## 6		0	1	0
##	Job.Management.SelfEmp.HighlyQualified			
## 1		0		
## 2		0		
## 3		0		

```
## 4 0
## 5 0
## 6 0
```

```
d1<-read.csv(file.path(Data,"GermanCredit_modified_SP19_001.csv"))
head(d1)
```

```
## Duration Amount InstallmentRatePercentage ResidenceDuration Age
## 1 6 1169 4 4 67
## 2 48 5951 2 2 22
## 3 12 2096 2 3 49
## 4 42 7882 2 4 45
## 5 24 4870 3 4 53
## 6 36 9055 2 4 35
## NumberExistingCredits NumberPeopleMaintenance Telephone ForeignWorker
## 1 2 1 0 1
## 2 1 1 1 1
## 3 1 2 1 1
## 4 1 2 1 1
## 5 2 2 1 1
## 6 1 2 0 1
## Class Checking Credit.History Loan.Purpose Savings
## 1 Good lt.0 Critical Radio.Television Unknown
## 2 Bad 0.to.200 PaidDuly Radio.Television lt.100
## 3 Good None Critical Education lt.100
## 4 Good lt.0 PaidDuly Furniture lt.100
## 5 Bad lt.0 Delay NewCar lt.100
## 6 Good None PaidDuly Education Unknown
## Employment.Duration Personal.Status Other.Debtors Property
## 1 gt.7 Single None RealEstate
## 2 1.to.4 NotSingle None RealEstate
## 3 4.to.7 Single None RealEstate
## 4 4.to.7 Single Guarantor Insurance
## 5 1.to.4 Single None Unknown
## 6 1.to.4 Single None Unknown
## OtherInstallmentPlans Housing Job.Type
## 1 None Own SkilledEmployee
## 2 None Own SkilledEmployee
## 3 None Own UnskilledResident
## 4 None ForFree SkilledEmployee
## 5 None ForFree SkilledEmployee
## 6 None ForFree UnskilledResident
```

```
Hmisc::describe(d)
```

```
## d
##
## 62 Variables 1000 Observations
## -----
## Duration
## n missing distinct Info Mean Gmd .05 .10
## 1000 0 33 0.985 20.9 12.98 6 9
## .25 .50 .75 .90 .95
```

```

##          12          18          24          36          48
##
## lowest :  4  5  6  7  8, highest: 47 48 54 60 72
## -----
## Amount
##          n missing distinct      Info      Mean      Gmd      .05      .10
##        1000          0      921          1      3271      2773      709      932
##          .25      .50      .75      .90      .95
##        1366      2320      3972      7179      9163
##
## lowest :   250   276   338   339   343, highest: 15653 15672 15857 15945 18424
## -----
## InstallmentRatePercentage
##          n missing distinct      Info      Mean      Gmd
##        1000          0          4      0.873      2.973      1.2
##
## Value          1      2      3      4
## Frequency      136      231      157      476
## Proportion 0.136 0.231 0.157 0.476
## -----
## ResidenceDuration
##          n missing distinct      Info      Mean      Gmd
##        1000          0          4      0.895      2.845      1.205
##
## Value          1      2      3      4
## Frequency      130      308      149      413
## Proportion 0.130 0.308 0.149 0.413
## -----
## Age
##          n missing distinct      Info      Mean      Gmd      .05      .10
##        1000          0          53      0.999      35.55      12.41      22      23
##          .25      .50      .75      .90      .95
##          27      33      42      52      60
##
## lowest : 19 20 21 22 23, highest: 67 68 70 74 75
## -----
## NumberExistingCredits
##          n missing distinct      Info      Mean      Gmd
##        1000          0          4      0.709      1.407      0.5428
##
## Value          1      2      3      4
## Frequency      633      333      28      6
## Proportion 0.633 0.333 0.028 0.006
## -----
## NumberPeopleMaintenance
##          n missing distinct      Info      Mean      Gmd
##        1000          0          2      0.393      1.155      0.2622
##
## Value          1      2
## Frequency      845      155
## Proportion 0.845 0.155
## -----
## Telephone
##          n missing distinct      Info      Sum      Mean      Gmd

```

```

##      1000      0      2      0.722      596      0.596      0.4821
##
## -----
## ForeignWorker
##      n missing distinct      Info      Sum      Mean      Gmd
##      1000      0      2      0.107      963      0.963      0.07133
##
## -----
## Class
##      n missing distinct
##      1000      0      2
##
## Value      Bad Good
## Frequency  300  700
## Proportion 0.3  0.7
## -----
## CheckingAccountStatus.lt.0
##      n missing distinct      Info      Sum      Mean      Gmd
##      1000      0      2      0.597      274      0.274      0.3982
##
## -----
## CheckingAccountStatus.0.to.200
##      n missing distinct      Info      Sum      Mean      Gmd
##      1000      0      2      0.59      269      0.269      0.3937
##
## -----
## CheckingAccountStatus.gt.200
##      n missing distinct      Info      Sum      Mean      Gmd
##      1000      0      2      0.177      63      0.063      0.1182
##
## -----
## CheckingAccountStatus.none
##      n missing distinct      Info      Sum      Mean      Gmd
##      1000      0      2      0.716      394      0.394      0.478
##
## -----
## CreditHistory.NoCredit.AllPaid
##      n missing distinct      Info      Sum      Mean      Gmd
##      1000      0      2      0.115      40      0.04      0.07688
##
## -----
## CreditHistory.ThisBank.AllPaid
##      n missing distinct      Info      Sum      Mean      Gmd
##      1000      0      2      0.14      49      0.049      0.09329
##
## -----
## CreditHistory.PaidDuly
##      n missing distinct      Info      Sum      Mean      Gmd
##      1000      0      2      0.747      530      0.53      0.4987
##
## -----
## CreditHistory.Delay
##      n missing distinct      Info      Sum      Mean      Gmd
##      1000      0      2      0.241      88      0.088      0.1607

```

```

##
## -----
## CreditHistory.Critical
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000         0         2    0.621     293    0.293    0.4147
##
## -----
## Purpose.NewCar
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000         0         2    0.538     234    0.234    0.3588
##
## -----
## Purpose.UsedCar
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000         0         2    0.277     103    0.103    0.185
##
## -----
## Purpose.Furniture.Equipment
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000         0         2    0.445     181    0.181    0.2968
##
## -----
## Purpose.Radio.Television
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000         0         2    0.605     280    0.28    0.4036
##
## -----
## Purpose.DomesticAppliance
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000         0         2    0.036      12    0.012    0.02374
##
## -----
## Purpose.Repairs
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000         0         2    0.065      22    0.022    0.04308
##
## -----
## Purpose.Education
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000         0         2    0.143      50    0.05    0.0951
##
## -----
## Purpose.Vacation
##      n missing distinct      Info      Mean      Gmd
##    1000         0         1         0         0         0
##
## Value          0
## Frequency 1000
## Proportion    1
## -----
## Purpose.Retaining
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000         0         2    0.027         9    0.009    0.01786
##

```



```

## -----
## Purpose.Business
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000      0        2    0.263      97    0.097    0.1754
##
## -----
## Purpose.Other
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000      0        2    0.036      12    0.012    0.02374
##
## -----
## SavingsAccountBonds.lt.100
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000      0        2    0.718     603    0.603    0.4793
##
## -----
## SavingsAccountBonds.100.to.500
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000      0        2    0.277     103    0.103    0.185
##
## -----
## SavingsAccountBonds.500.to.1000
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000      0        2    0.177      63    0.063    0.1182
##
## -----
## SavingsAccountBonds.gt.1000
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000      0        2    0.137      48    0.048    0.09148
##
## -----
## SavingsAccountBonds.Unknown
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000      0        2    0.449     183    0.183    0.2993
##
## -----
## EmploymentDuration.lt.1
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000      0        2    0.427     172    0.172    0.2851
##
## -----
## EmploymentDuration.1.to.4
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000      0        2    0.672     339    0.339    0.4486
##
## -----
## EmploymentDuration.4.to.7
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000      0        2    0.431     174    0.174    0.2877
##
## -----
## EmploymentDuration.gt.7
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000      0        2    0.567     253    0.253    0.3784

```

```

##
## -----
## EmploymentDuration.Unemployed
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000         0         2    0.174        62    0.062    0.1164
##
## -----
## Personal.Male.Divorced.Seperated
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000         0         2    0.143        50    0.05    0.0951
##
## -----
## Personal.Female.NotSingle
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000         0         2    0.642       310    0.31    0.4282
##
## -----
## Personal.Male.Single
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000         0         2    0.743       548    0.548    0.4959
##
## -----
## Personal.Male.Married.Widowed
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000         0         2    0.251        92    0.092    0.1672
##
## -----
## Personal.Female.Single
##      n missing distinct      Info      Mean      Gmd
##    1000         0         1         0         0         0
##
## Value          0
## Frequency 1000
## Proportion    1
## -----
## OtherDebtorsGuarantors.None
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000         0         2    0.253       907    0.907    0.1689
##
## -----
## OtherDebtorsGuarantors.CoApplicant
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000         0         2    0.118        41    0.041    0.07872
##
## -----
## OtherDebtorsGuarantors.Guarantor
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000         0         2    0.148        52    0.052    0.09869
##
## -----
## Property.RealEstate
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000         0         2    0.607       282    0.282    0.4054
##

```

```

## -----
## Property.Insurance
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000         0         2    0.535     232    0.232    0.3567
##
## -----
## Property.CarOther
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000         0         2    0.665     332    0.332    0.444
##
## -----
## Property.Unknown
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000         0         2    0.391     154    0.154    0.2608
##
## -----
## OtherInstallmentPlans.Bank
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000         0         2    0.359     139    0.139    0.2396
##
## -----
## OtherInstallmentPlans.Stores
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000         0         2    0.134      47    0.047    0.08967
##
## -----
## OtherInstallmentPlans.None
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000         0         2    0.454     814    0.814    0.3031
##
## -----
## Housing.Rent
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000         0         2    0.441     179    0.179    0.2942
##
## -----
## Housing.Own
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000         0         2    0.614     713    0.713    0.4097
##
## -----
## Housing.ForFree
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000         0         2    0.289     108    0.108    0.1929
##
## -----
## Job.UnemployedUnskilled
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000         0         2    0.065      22    0.022    0.04308
##
## -----
## Job.UnskilledResident
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000         0         2    0.48     200     0.2    0.3203

```

```
##
## -----
## Job.SkilledEmployee
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000         0         2    0.699      630      0.63    0.4667
##
## -----
## Job.Management.SelfEmp.HighlyQualified
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000         0         2    0.378      148      0.148    0.2524
##
## -----
```

```
de <- describe(d)
```

```
# p <- plot(d)
# plot(d, which='categorical')
```

Question 1.3.a

This data represents a collection of important variables that a certain financial institution might find important in their processing of financial loans. There are general factors that are mostly represented by numerical values and integers (apart from class) that allow us to

Question 1.3.b

In this data set we have 1000 observations and 62 different variables.

Question 1.3.c

There are 61 numerical variables.

They are the following: Duration, Amount, InstallmentRatePercentage; ResidenceDuration; Age; NumberExistingCredits; NumberPeopleMaintenance; Telephone; ForeignWorker; CheckingAccountStatus.lt.0; CheckingAccountStatus.0.to.200; CheckingAccountStatus.gt.200; CheckingAccountStatus.none; CreditHistory.NoCredit.AllPaid; CreditHistory.ThisBank.AllPaid; CreditHistory.PaidDuly; CreditHistory.Delay; CreditHistory.Critical; Purpose.NewCar; Purpose.UsedCar; Purpose.Furniture.Equipment; Purpose.Radio.Television; Purpose.DomesticAppliance; Purpose.Repairs; Purpose.Education; Purpose.Vacation; Purpose.Retaining; Purpose.Business; Purpose.Other; 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.Separated; Personal.Female.NotSingle; Personal.Male.Single; Personal.Male.Married.Widowed; Personal.Female.Single; OtherDebtorsGuarantors.None; OtherDebtorsGuarantors.CoApplicant; OtherDebtorsGuarantors.Guarantor; Property.RealEstate; Property.Insurance; Property.CarOther; Property.Unknown; OtherInstallmentPlans.Bank; OtherInstallmentPlans.Stores; OtherInstallmentPlans.None; Housing.Rent; Housing.Own; Housing.ForFree; Job.UnemployedUnskilled; Job.UnskilledResident; Job.SkilledEmployee; Job.Management.SelfEmp.HighlyQualified

They represent different numerical values for some (i.e. amount, duration, age), while others are binary (using 1 and 0) displaying if a certain individual has a variable (1) or doesn't have one (0).

Question 1.3.d

There is 1 categorical value. The only categorical value is Class. It measures the customers credit as being either Good or Bad.

Question 1.3.e

We can perform certain calculations and observe patterns and other trends with loan users. This can display different insights (i.e. who is applying for loans, at what age, what is their credit history, their current assets, if they are employed etc.)

```
d %>%  
  ggplot(aes(x=Amount, y=Duration, color=factor(Housing.Own)))+  
  geom_point()+  
  geom_smooth(method="lm")
```

