

High Dimensional Data Analysis - Group Assignment

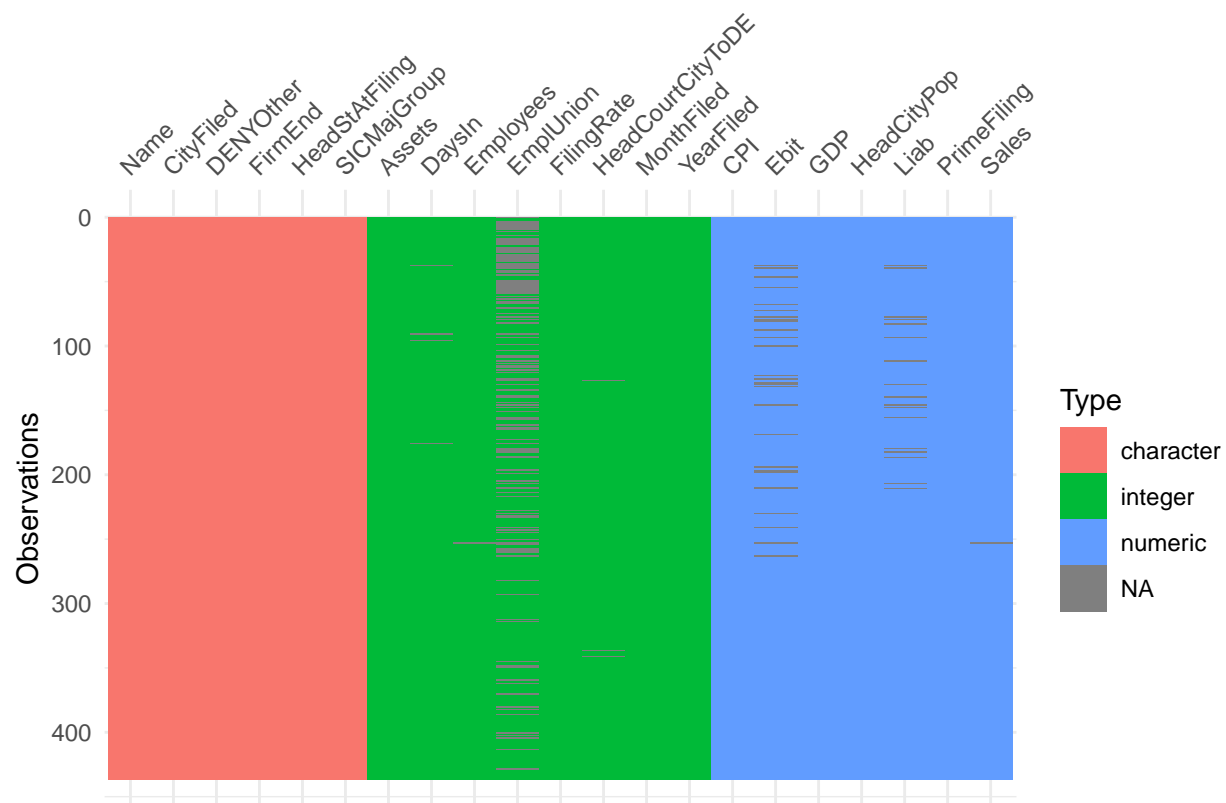
Group 18

08/09/2021

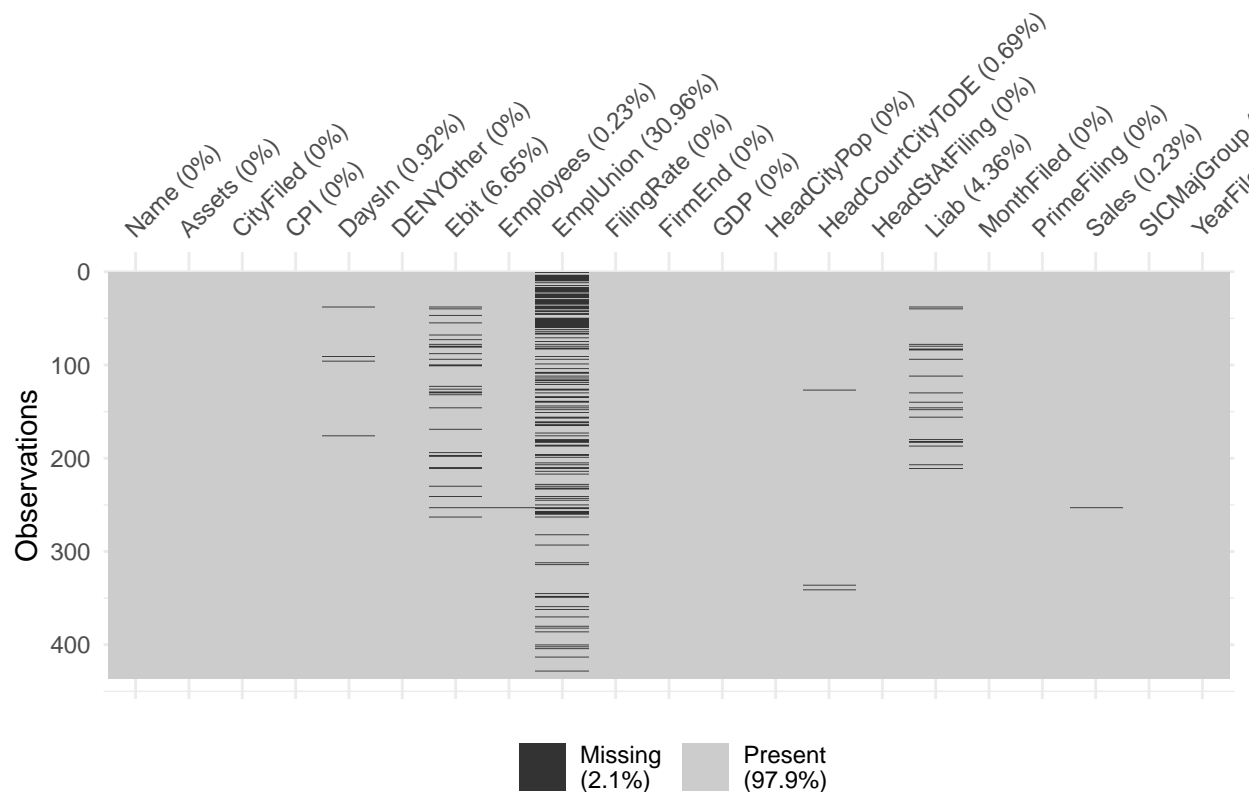
Contents

```
data <- readRDS(here::here("data/Bankruptcy.rds"))
glimpse(data)
```

```
## Rows: 436
## Columns: 21
## $ Name      <chr> "Combustion Equipment Associates, Inc.", "Penn-Dixie~
## $ Assets    <int> 531, 552, 1897, 821, 4097, 1200, 1141, 2628, 1456, 1~
## $ CityFiled <chr> "New York", "New York", "Cleveland", "New York", "Sa~
## $ CPI       <dbl> 84.8, 81.0, 84.0, 93.2, 87.0, 94.0, 93.7, 87.9, 94.9~
## $ DaysIn    <int> 1157, 696, 1170, 1545, 792, 1099, 1343, 2238, 881, 1~
## $ DENYOther <chr> "NY", "NY", "OT", "NY", "OT", "OT", "OT", "NY", "OT"~
## $ Ebit      <dbl> 13.831140, -13.521542, 102.647226, 71.496993, 176.43~
## $ Employees <int> 2400, 4191, 9685, 1116, 1400, 5225, 32000, 1900, 172~
## $ EmplUnion <int> NA, 2975, 5800, NA, NA, NA, NA, NA, NA, 8531, NA~
## $ FilingRate <int> 3, 3, 3, 5, 5, 5, 5, 5, 13, 13, 13, 13, 13, 13, 13, ~
## $ FirmEnd   <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", ~
## $ GDP       <dbl> 42.067, 41.346, 41.296, 43.083, 42.891, 42.613, 42.6~
## $ HeadCityPop <dbl> 7071639, 7071639, 58056, 418532, 683472, 1185802, 11~
## $ HeadCourtCityToDE <int> 106, 106, 435, 241, 2514, 435, 2383, 106, 653, 1238,~
## $ HeadStAtFiling <chr> "NY", "NY", "MI", "PA", "CA", "MI", "CA", "NY", "IL"~
## $ Liab      <dbl> 309.6648, 377.9007, 1201.9985, 751.4130, 4872.2510, ~
## $ MonthFiled <int> 10, 4, 9, 9, 1, 12, 11, 2, 4, 12, 5, 6, 11, 8, 2, 8,~
## $ PrimeFiling <dbl> 14.00, 20.00, 11.50, 19.50, 20.00, 15.75, 16.00, 19.~
## $ Sales     <dbl> 357537044, 900139474, 3662349812, 423926972, 6016918~
## $ SICMajGroup <chr> "38 Measuring, Analyzing and Controlling Instruments~
## $ YearFiled  <int> 1980, 1980, 1980, 1981, 1981, 1981, 1981, 1981, 1982~
vis_dat(data)
```



```
vis_miss(data)
```



```
Data <- data %>% na.omit()
mean_EmplUnion <- mean(Data$EmplUnion)
mean_Ebit <- mean(Data$Ebit)
mean_Liab <- mean(Data$Liab)
mean_DaysIn <- mean(Data$DaysIn)
mean_HeadCourtCityToDE <- mean(Data$HeadCourtCityToDE)

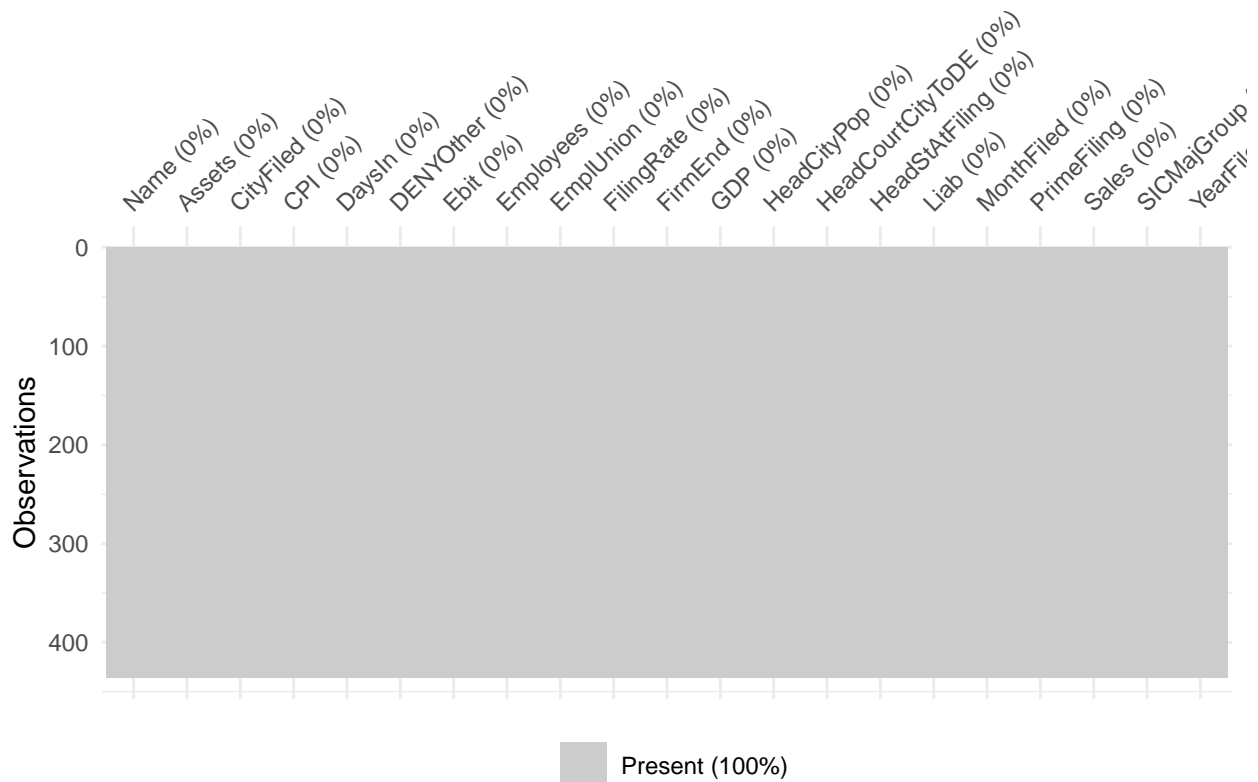
data$EmplUnion <- data %>% pull(EmplUnion) %>% replace_na(mean_EmplUnion)
data$Ebit <- data %>% pull(Ebit) %>% replace_na(mean_Ebit)
data$Liab <- data %>% pull(Liab) %>% replace_na(mean_Liab)
data$DaysIn <- data %>% pull(DaysIn) %>% replace_na(mean_DaysIn)
data$HeadCourtCityToDE <- data %>% pull(HeadCourtCityToDE) %>% replace_na(mean_HeadCourtCityToDE)
Bankruptcy <- data %>% na.omit()

glimpse(Bankruptcy)
```

```
## Rows: 435
## Columns: 21
## $ Name      <chr> "Combustion Equipment Associates, Inc.", "Penn-Dixie~
## $ Assets    <int> 531, 552, 1897, 821, 4097, 1200, 1141, 2628, 1456, 1~
## $ CityFiled <chr> "New York", "New York", "Cleveland", "New York", "Sa~
## $ CPI       <dbl> 84.8, 81.0, 84.0, 93.2, 87.0, 94.0, 93.7, 87.9, 94.9~
## $ DaysIn    <dbl> 1157, 696, 1170, 1545, 792, 1099, 1343, 2238, 881, 1~
## $ DENYOther <chr> "NY", "NY", "OT", "NY", "OT", "OT", "OT", "NY", "OT"~
## $ Ebit      <dbl> 13.831140, -13.521542, 102.647226, 71.496993, 176.43~
## $ Employees <int> 2400, 4191, 9685, 1116, 1400, 5225, 32000, 1900, 172~
```

```
## $ EmplUnion      <dbl> 1183.257, 2975.000, 5800.000, 1183.257, 1183.257, 11~
## $ FilingRate     <int> 3, 3, 3, 5, 5, 5, 5, 5, 13, 13, 13, 13, 13, 13, ~
## $ FirmEnd        <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", ~
## $ GDP            <dbl> 42.067, 41.346, 41.296, 43.083, 42.891, 42.613, 42.6~
## $ HeadCityPop    <dbl> 7071639, 7071639, 58056, 418532, 683472, 1185802, 11~
## $ HeadCourtCityToDE <dbl> 106, 106, 435, 241, 2514, 435, 2383, 106, 653, 1238,~
## $ HeadStAtFiling <chr> "NY", "NY", "MI", "PA", "CA", "MI", "CA", "NY", "IL"~
## $ Liab           <dbl> 309.6648, 377.9007, 1201.9985, 751.4130, 4872.2510, ~
## $ MonthFiled     <int> 10, 4, 9, 9, 1, 12, 11, 2, 4, 12, 5, 6, 11, 8, 2, 8,~
## $ PrimeFiling    <dbl> 14.00, 20.00, 11.50, 19.50, 20.00, 15.75, 16.00, 19.~
## $ Sales          <dbl> 357537044, 900139474, 3662349812, 423926972, 6016918~
## $ SICMajGroup    <chr> "38 Measuring, Analyzing and Controlling Instruments~
## $ YearFiled      <int> 1980, 1980, 1980, 1981, 1981, 1981, 1981, 1981, 1982~
```

```
vis_miss(Bankruptcy)
```



```
vis_dat(Bankruptcy)
```

