MLD

1. Loading data

```
rm(list=ls(all=TRUE))
library(tidyverse)
```

```
── Attaching packages ─────────────────────────────── tidyverse 1.3.2
──
✓ ggplot2 3.4.0      ✓ purrr   1.0.1
✓ tibble  3.1.8      ✓ dplyr   1.0.10
✓ tidyr   1.2.1      ✓ stringr 1.4.1
✓ readr   2.1.3      ✓ forcats 0.5.2
── Conflicts ────────────────────────────────── tidyverse_conflicts()
──
✗ dplyr::filter() masks stats::filter()
✗ dplyr::lag()    masks stats::lag()
```

```
library(vtable)
```

```
Loading required package: kableExtra

Attaching package: 'kableExtra'

The following object is masked from 'package:dplyr':

    group_rows
```

```
library(dplyr)
```

```
MLD <- read.csv("MLD Data File-1.csv", header=TRUE)
```

```
#Checking the data
vtable(MLD, lush = TRUE)
```

MLD

Name	Class	Values	Missing	Summary
MARRIED	character	'.' '0' '1'	0	nuniq: 3
GDLIN	integer	Num: 0 to 666	0	mean: 1.583, sd: 21.087, nuniq: 3
OBRAT	numeric	Num: 0 to 95	0	mean: 32.389, sd: 8.263, nuniq: 288
BLACK	integer	Num: 0 to 1	0	mean: 0.099, sd: 0.299, nuniq: 2
HISPAN	integer	Num: 0 to 1	0	mean: 0.056, sd: 0.23, nuniq: 2

Name	Class	Values	Missing	Summary
MALE	character	'.' '0' '1'	0	nuniq: 3
APPROVE	integer	Num: 0 to 1	0	mean: 0.877, sd: 0.328, nuniq: 2
LOANPRC	numeric	Num: 0.021 to 2.571	0	mean: 0.771, sd: 0.189, nuniq: 1108

```r
#Filtering the data
dat <- MLD %>%
  filter(MARRIED %in% c(0,1) & MALE %in% c(0,1) & LOANPRC <= 1 &
         GDLIN %in% c(0,1))

dat$LOANPRC = dat$LOANPRC *100

#Convert character to integer
dat$MARRIED = as.integer(dat$MARRIED)
dat$MALE = as.integer(dat$MALE)

vtable(dat, lush = TRUE)
```

dat

Name	Class	Values	Missing	Summary
MARRIED	integer	Num: 0 to 1	0	mean: 0.658, sd: 0.474, nuniq: 2
GDLIN	integer	Num: 0 to 1	0	mean: 0.914, sd: 0.281, nuniq: 2
OBRAT	numeric	Num: 0 to 95	0	mean: 32.366, sd: 8.247, nuniq: 286
BLACK	integer	Num: 0 to 1	0	mean: 0.099, sd: 0.299, nuniq: 2
HISPAN	integer	Num: 0 to 1	0	mean: 0.054, sd: 0.225, nuniq: 2
MALE	integer	Num: 0 to 1	0	mean: 0.814, sd: 0.39, nuniq: 2
APPROVE	integer	Num: 0 to 1	0	mean: 0.879, sd: 0.327, nuniq: 2
LOANPRC	numeric	Num: 2.105 to 100	0	mean: 76.075, sd: 16.763, nuniq: 1064

2. EDA for All Races

```r
dat <- dat %>%
  mutate(SINGLE = 1-MARRIED,
         GDLIN_N = 1 - GDLIN,
         WHITE = 1 - BLACK - HISPAN,
         FEMALE = 1 - MALE,
         DISAPPROVE = 1 - APPROVE)

d_mean <- c(apply(dat[ , c('MARRIED', 'SINGLE', 'GDLIN', 'GDLIN_N',
                           'WHITE', 'BLACK','HISPAN', 'MALE',
                           'FEMALE','APPROVE', 'DISAPPROVE')], 2,
                  mean))
d_mean <- round(d_mean*100, 2)
```

```
knitr::kable(d_mean, "pipe")
```

	x
MARRIED	65.82
SINGLE	34.18
GDLIN	91.38
GDLIN_N	8.62
WHITE	84.72
BLACK	9.91
HISPAN	5.37
MALE	81.36
FEMALE	18.64
APPROVE	87.87
DISAPPROVE	12.13

```
d_min <- c(apply(dat[ , c('OBRAT', 'LOANPRC')], 2, min))
d_max <- c(apply(dat[ , c('OBRAT', 'LOANPRC')], 2, max))
d_mean <- c(apply(dat[ , c('OBRAT', 'LOANPRC')], 2, mean))
d_median <- c(apply(dat[ , c('OBRAT', 'LOANPRC')], 2, median))
d_sd <- c(sqrt(apply(dat[ , c('OBRAT', 'LOANPRC')], 2, var)))
d_summary <- round(cbind(d_min, d_max, d_mean, d_median, d_sd),2)
knitr::kable(d_summary, "pipe")
```

	d_min	d_max	d_mean	d_median	d_sd
OBRAT	0.00	95	32.37	33	8.25
LOANPRC	2.11	100	76.08	80	16.76

3. EDA for non-Hispanic white

```
dat_white <- dat %>%
  filter( BLACK == 0 & HISPAN == 0)

d_mean <- c(apply(dat[ , c('MARRIED', 'SINGLE', 'GDLIN', 'GDLIN_N',
                           'WHITE', 'BLACK','HISPAN', 'MALE',
                           'FEMALE','APPROVE', 'DISAPPROVE')], 2,
                mean))
d_mean <- round(d_mean*100, 2)

knitr::kable(d_mean, "pipe")
```

	x
MARRIED	65.82
SINGLE	34.18

	x
GDLIN	91.38
GDLIN_N	8.62
WHITE	84.72
BLACK	9.91
HISPAN	5.37
MALE	81.36
FEMALE	18.64
APPROVE	87.87
DISAPPROVE	12.13

```r
d_min <- c(apply(dat_white[ , c('OBRAT', 'LOANPRC')], 2, min))
d_max <- c(apply(dat_white[ , c('OBRAT', 'LOANPRC')], 2, max))
d_mean <- c(apply(dat_white[ , c('OBRAT', 'LOANPRC')], 2, mean))
d_median <- c(apply(dat_white[ , c('OBRAT', 'LOANPRC')], 2, median))
d_sd <- c(sqrt(apply(dat_white[ , c('OBRAT', 'LOANPRC')], 2, var)))
d_summary <- round(cbind(d_min, d_max, d_mean, d_median, d_sd),2)
knitr::kable(d_summary, "pipe")
```

	d_min	d_max	d_mean	d_median	d_sd
OBRAT	0.00	95	31.99	32.50	8.18
LOANPRC	2.11	100	74.78	79.86	17.17

4. EDA for non-Hispanic black

```r
dat_black <- dat %>%
  filter(BLACK == 1)

d_mean <- c(apply(dat[ , c('MARRIED', 'SINGLE', 'GDLIN', 'GDLIN_N',
                           'WHITE', 'BLACK','HISPAN', 'MALE',
                           'FEMALE','APPROVE', 'DISAPPROVE')], 2,
              mean))
d_mean <- round(d_mean*100, 2)

knitr::kable(d_mean, "pipe")
```

	x
MARRIED	65.82
SINGLE	34.18
GDLIN	91.38
GDLIN_N	8.62
WHITE	84.72
BLACK	9.91

	x
HISPAN	5.37
MALE	81.36
FEMALE	18.64
APPROVE	87.87
DISAPPROVE	12.13

```r
d_min <- c(apply(dat_black[ , c('OBRAT', 'LOANPRC')], 2, min))
d_max <- c(apply(dat_black[ , c('OBRAT', 'LOANPRC')], 2, max))
d_mean <- c(apply(dat_black[ , c('OBRAT', 'LOANPRC')], 2, mean))
d_median <- c(apply(dat_black[ , c('OBRAT', 'LOANPRC')], 2, median))
d_sd <- c(sqrt(apply(dat_black[ , c('OBRAT', 'LOANPRC')], 2, var)))
d_summary <- round(cbind(d_min, d_max, d_mean, d_median, d_sd),2)
knitr::kable(d_summary, "pipe")
```

	d_min	d_max	d_mean	d_median	d_sd
OBRAT	5.60	63	35.03	35.00	8.13
LOANPRC	28.99	100	82.89	86.06	12.63

5. EDA for hispanic

```r
dat_his <- dat %>%
  filter(HISPAN == 1)

d_mean <- c(apply(dat[ , c('MARRIED', 'SINGLE', 'GDLIN', 'GDLIN_N',
                           'WHITE', 'BLACK','HISPAN', 'MALE',
                           'FEMALE','APPROVE', 'DISAPPROVE')], 2,
                mean))
d_mean <- round(d_mean*100, 2)

knitr::kable(d_mean, "pipe")
```

	x
MARRIED	65.82
SINGLE	34.18
GDLIN	91.38
GDLIN_N	8.62
WHITE	84.72
BLACK	9.91
HISPAN	5.37
MALE	81.36
FEMALE	18.64
APPROVE	87.87

	x
DISAPPROVE	12.13

```r
d_min <- c(apply(dat_his[ , c('OBRAT', 'LOANPRC')], 2, min))
d_max <- c(apply(dat_his[ , c('OBRAT', 'LOANPRC')], 2, max))
d_mean <- c(apply(dat_his[ , c('OBRAT', 'LOANPRC')], 2, mean))
d_median <- c(apply(dat_his[ , c('OBRAT', 'LOANPRC')], 2, median))
d_sd <- c(sqrt(apply(dat_his[ , c('OBRAT', 'LOANPRC')], 2, var)))
d_summary <- round(cbind(d_min, d_max, d_mean, d_median, d_sd),2)

knitr::kable(d_summary, "pipe")
```

	d_min	d_max	d_mean	d_median	d_sd
OBRAT	14.60	62	33.32	33.00	8.56
LOANPRC	40.09	100	83.91	89.31	10.92