# Fields, Alex

## Ban 502

### Module 4 - Assignment 1

#### Task 1 - Q&A

Using CountNA() and vim\_plot I was able to see that we are missing 11 datapoints in our dataset. Final is misssing 4 points, TakeHome is missing 3, Midterm is missing 3, Tutorial is missing 1 and Assignment and Prefix are missing none.

#### Task 2 - Q&A

There does appear to be Systemic Missingness. There are 2 students missing just TakeHome work while there is one student missing TakeHome and Midterm.

#### Task 3 - Q&A

Using Row-Wise deletion their are now 89 observations/rows instead of 99.The 11 NAs from earlier are no longer there

#### Task 4 - Q&A

Using Column-Wise deletion to get rid of missing values leaves me with 2 of the original 6 Columns.

#### Task 5 - Q&A

For this dataset, Row-Wise deletion is the best approach. Since most of the variables have missing data your dataset will not be the same as it was before with Column-Wise Deletion. If those 4 variables deleted are integral to the dataset than you might as well not run the analysis in the first place. Row-Wise will keep “most” of the data in place.

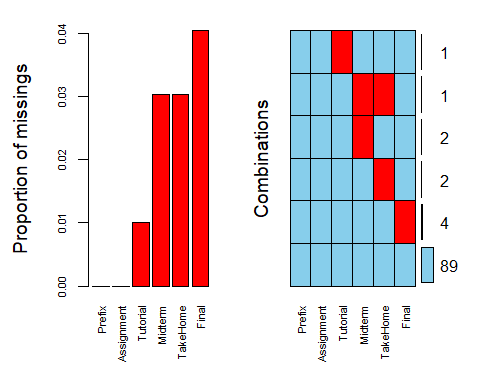
#### Task 7 - Q&A

When looking at the imputation for Tutorial, Midterm, TakeHome and Final we can see that Tutorial is not being plotted.This is because the function only works for 2 or missing values in the variable. When looking at the density plot for TakeHome, it appears like two different densities are being graphed. This shows that the imputed mean calculation is off for this variable and we would not use it. Midterm and Final seem fair to use.

options(tidyverse.quiet = TRUE)  
library(tidyverse)  
library(VIM)  
library(mice)  
  
class\_grades <- read\_csv("class-grades.csv")  
  
countNA(class\_grades)

## [1] 11

vim\_plot = aggr(class\_grades, numbers = TRUE, prop = c(TRUE, FALSE),cex.axis=.7)

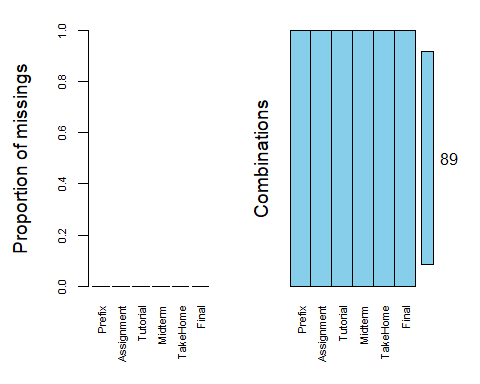


### Row-Wise Deletion

class\_grades\_new = class\_grades %>% drop\_na(Final) %>% drop\_na(TakeHome) %>% drop\_na(Midterm) %>% drop\_na(Tutorial)  
  
countNA(class\_grades\_new)

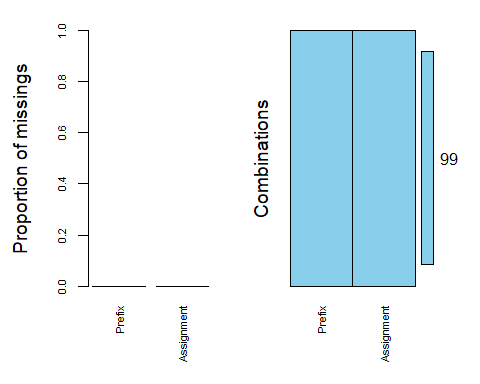
## [1] 0

vim\_plot = aggr(class\_grades\_new, numbers = TRUE, prop = c(TRUE, FALSE),cex.axis=.7)



### Column-Wise Deletion

class\_grades\_col = class\_grades %>% select(-Final) %>% select(-Tutorial) %>% select(-Midterm) %>% select(-TakeHome)   
vim\_plot = aggr(class\_grades\_col, numbers = TRUE, prop = c(TRUE, FALSE),cex.axis=.7)



### impute data

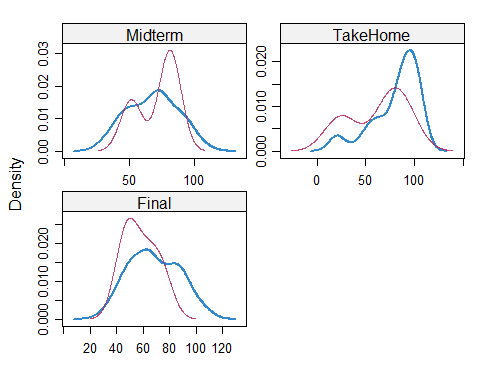
class\_grades\_imp = mice(class\_grades, m=1, method = "pmm", seed = 12345)

##   
## iter imp variable  
## 1 1 Tutorial Midterm TakeHome Final  
## 2 1 Tutorial Midterm TakeHome Final  
## 3 1 Tutorial Midterm TakeHome Final  
## 4 1 Tutorial Midterm TakeHome Final  
## 5 1 Tutorial Midterm TakeHome Final

#in line above: m=1 -> runs one imputation, seed sets the random number seed to get repeatable results  
summary(class\_grades\_imp)

## Class: mids  
## Number of multiple imputations: 1   
## Imputation methods:  
## Prefix Assignment Tutorial Midterm TakeHome Final   
## "" "" "pmm" "pmm" "pmm" "pmm"   
## PredictorMatrix:  
## Prefix Assignment Tutorial Midterm TakeHome Final  
## Prefix 0 1 1 1 1 1  
## Assignment 1 0 1 1 1 1  
## Tutorial 1 1 0 1 1 1  
## Midterm 1 1 1 0 1 1  
## TakeHome 1 1 1 1 0 1  
## Final 1 1 1 1 1 0

densityplot(class\_grades\_imp)



#red imputed, blue original, only shows density plots when more than 1 value the variable was imputed  
#note that the density plots are fairly uninteresting given the small amount of missing data  
class\_grades\_complete = complete(class\_grades\_imp)  
summary(class\_grades\_complete)

## Prefix Assignment Tutorial Midterm   
## Min. :4.000 Min. : 28.14 Min. : 34.09 Min. : 28.12   
## 1st Qu.:7.000 1st Qu.: 80.88 1st Qu.: 84.69 1st Qu.: 52.50   
## Median :8.000 Median : 89.94 Median : 93.10 Median : 69.38   
## Mean :7.313 Mean : 85.49 Mean : 89.76 Mean : 67.80   
## 3rd Qu.:8.000 3rd Qu.: 95.00 3rd Qu.:100.55 3rd Qu.: 81.88   
## Max. :8.000 Max. :100.83 Max. :112.58 Max. :110.00   
## TakeHome Final   
## Min. : 16.91 Min. : 28.06   
## 1st Qu.: 67.96 1st Qu.: 52.09   
## Median : 87.96 Median : 65.56   
## Mean : 80.54 Mean : 67.81   
## 3rd Qu.: 98.42 3rd Qu.: 83.19   
## Max. :108.89 Max. :108.89