# David Wells

## Module 4 Assignment 2

library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.3 v purrr 0.3.4  
## v tibble 3.0.6 v dplyr 1.0.3  
## v tidyr 1.1.2 v stringr 1.4.0  
## v readr 1.4.0 v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(VIM)

## Loading required package: colorspace

## Loading required package: grid

## VIM is ready to use.

## Suggestions and bug-reports can be submitted at: https://github.com/statistikat/VIM/issues

##   
## Attaching package: 'VIM'

## The following object is masked from 'package:datasets':  
##   
## sleep

library(mice)

##   
## Attaching package: 'mice'

## The following object is masked from 'package:stats':  
##   
## filter

## The following objects are masked from 'package:base':  
##   
## cbind, rbind

library(skimr)

class\_grades <- read\_csv("C:/Users/wdavi/Documents/BAN 502/Module 4/class-grades.csv")

##   
## -- Column specification --------------------------------------------------------  
## cols(  
## Prefix = col\_double(),  
## Assignment = col\_double(),  
## Tutorial = col\_double(),  
## Midterm = col\_double(),  
## TakeHome = col\_double(),  
## Final = col\_double()  
## )

str(class\_grades)

## spec\_tbl\_df [99 x 6] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ Prefix : num [1:99] 5 8 8 7 8 7 8 7 8 7 ...  
## $ Assignment: num [1:99] 57.1 95 83.7 81.2 91.3 ...  
## $ Tutorial : num [1:99] 34.1 105.5 83.2 96.1 93.6 ...  
## $ Midterm : num [1:99] 64.4 67.5 30 49.4 95 ...  
## $ TakeHome : num [1:99] 51.5 99.1 63.1 105.9 107.4 ...  
## $ Final : num [1:99] 52.5 68.3 48.9 80.6 73.9 ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. Prefix = col\_double(),  
## .. Assignment = col\_double(),  
## .. Tutorial = col\_double(),  
## .. Midterm = col\_double(),  
## .. TakeHome = col\_double(),  
## .. Final = col\_double()  
## .. )

summary(class\_grades)

## 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.: 83.93 1st Qu.: 52.50   
## Median :8.000 Median : 89.94 Median : 93.37 Median : 69.38   
## Mean :7.313 Mean : 85.49 Mean : 89.79 Mean : 67.70   
## 3rd Qu.:8.000 3rd Qu.: 95.00 3rd Qu.:100.56 3rd Qu.: 81.56   
## Max. :8.000 Max. :100.83 Max. :112.58 Max. :110.00   
## NA's :1 NA's :3   
## TakeHome Final   
## Min. : 16.91 Min. : 28.06   
## 1st Qu.: 69.91 1st Qu.: 52.91   
## Median : 88.42 Median : 66.11   
## Mean : 81.12 Mean : 68.23   
## 3rd Qu.: 99.07 3rd Qu.: 83.61   
## Max. :108.89 Max. :108.89   
## NA's :3 NA's :4

skim(class\_grades)

Data summary

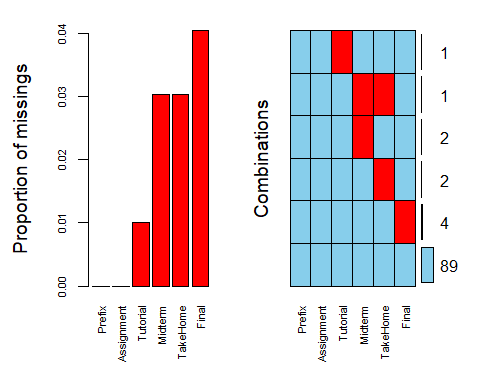
|  |  |
| --- | --- |
| Name | class\_grades |
| Number of rows | 99 |
| Number of columns | 6 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| numeric | 6 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: numeric**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
| Prefix | 0 | 1.00 | 7.31 | 0.93 | 4.00 | 7.00 | 8.00 | 8.00 | 8.00 | ▁▁▁▅▇ |
| Assignment | 0 | 1.00 | 85.49 | 12.60 | 28.14 | 80.88 | 89.94 | 95.00 | 100.83 | ▁▁▁▅▇ |
| Tutorial | 1 | 0.99 | 89.79 | 15.14 | 34.09 | 83.93 | 93.37 | 100.56 | 112.58 | ▁▂▁▇▆ |
| Midterm | 3 | 0.97 | 67.70 | 19.43 | 28.12 | 52.50 | 69.38 | 81.56 | 110.00 | ▃▅▇▅▂ |
| TakeHome | 3 | 0.97 | 81.12 | 23.95 | 16.91 | 69.90 | 88.42 | 99.07 | 108.89 | ▁▁▂▅▇ |
| Final | 4 | 0.96 | 68.23 | 18.82 | 28.06 | 52.92 | 66.11 | 83.61 | 108.89 | ▂▇▇▇▂ |

# \*Tutorial\*, \*MidTerm\*, \*TakeHome\*, and \*Final\* all have missing values

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



# One student is missing both \*Midterm\* and \*TakeHome\*

grades\_rowdel = class\_grades %>% drop\_na()   
skim(grades\_rowdel)

Data summary

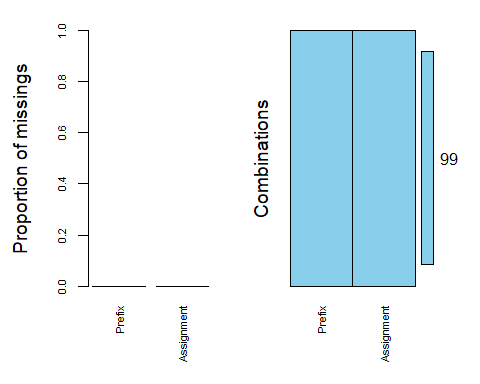
|  |  |
| --- | --- |
| Name | grades\_rowdel |
| Number of rows | 89 |
| Number of columns | 6 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| numeric | 6 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: numeric**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
| Prefix | 0 | 1 | 7.30 | 0.96 | 4.00 | 7.00 | 8.00 | 8.00 | 8.00 | ▁▁▁▅▇ |
| Assignment | 0 | 1 | 85.97 | 12.62 | 28.14 | 82.45 | 89.94 | 95.05 | 100.83 | ▁▁▁▅▇ |
| Tutorial | 0 | 1 | 89.40 | 15.75 | 34.09 | 82.93 | 92.93 | 100.58 | 112.58 | ▁▂▂▇▇ |
| Midterm | 0 | 1 | 68.14 | 19.79 | 28.12 | 52.50 | 70.00 | 83.12 | 110.00 | ▃▅▇▆▂ |
| TakeHome | 0 | 1 | 81.03 | 23.78 | 16.91 | 64.07 | 87.96 | 97.78 | 108.89 | ▁▂▂▅▇ |
| Final | 0 | 1 | 67.97 | 18.82 | 28.06 | 52.50 | 65.56 | 83.06 | 108.89 | ▂▇▇▆▂ |

# 89 rows remain

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



skim(grades\_coldel)

Data summary

|  |  |
| --- | --- |
| Name | grades\_coldel |
| Number of rows | 99 |
| Number of columns | 2 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| numeric | 2 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: numeric**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
| Prefix | 0 | 1 | 7.31 | 0.93 | 4.00 | 7.00 | 8.00 | 8 | 8.00 | ▁▁▁▅▇ |
| Assignment | 0 | 1 | 85.49 | 12.60 | 28.14 | 80.88 | 89.94 | 95 | 100.83 | ▁▁▁▅▇ |

# 2 Variables remain

# Row-wise deletion seems the best course of action. This way we only lose 10 records with incomplete data, rather than entire columns of data that were otherwise useful for analysis.

set.seed(123) #sets seed for random number generator  
imp\_class = mice(class\_grades, m=5, method='pmm', printFlag=FALSE)  
summary(imp\_class)

## Class: mids  
## Number of multiple imputations: 5   
## 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

grades\_complete = complete(imp\_class)   
summary(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.: 52.81   
## Median :8.000 Median : 89.94 Median : 93.10 Median : 70.00   
## Mean :7.313 Mean : 85.49 Mean : 89.76 Mean : 68.31   
## 3rd Qu.:8.000 3rd Qu.: 95.00 3rd Qu.:100.55 3rd Qu.: 82.81   
## 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.: 53.33   
## Median : 88.89 Median : 66.11   
## Mean : 80.80 Mean : 68.33   
## 3rd Qu.: 99.07 3rd Qu.: 83.61   
## Max. :108.89 Max. :108.89

# The average for \*Final\* was 68.33.